Information Sciences 355–356 (2016) 74–89

Contents lists available at ScienceDirect

Information Sciences

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

# Sparse representation matching for person re-identification

Le An[a], Xiaojing Chen[b,*], Songfan Yang[c], Bir Bhanu[d]

[a] National Key Laboratory of Science and Technology on Multi-spectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA
[c] College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China
[d] Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

ARTICLE INFO

*Article history:*
Received 28 September 2015
Revised 15 January 2016
Accepted 25 February 2016
Available online 22 March 2016

*Keywords:*
Person re-identification
Surveillance
Subspace learning
Sparse representation

ABSTRACT

The need for recognizing people across distributed surveillance cameras leads to the growth of recent research interest in person re-identification. Person re-identification aims at matching people in non-overlapping cameras at different time and locations. It is a difficult pattern matching task due to significant appearance variations in pose, illumination, or occlusion in different camera views. To address this multi-view matching problem, we first learn a subspace using canonical correlation analysis (CCA) in which the goal is to maximize the correlation between data from different cameras but corresponding to the same people. Given a probe from one camera view, we represent it using a sparse representation from a jointly learned coupled dictionary in the CCA subspace. The $\ell_1$ induced sparse representation are regularized by an $\ell_2$ regularization term. The introduction of $\ell_2$ regularization allows learning a sparse representation while maintaining the stability of the sparse coefficients. To compute the matching scores between probe and gallery, their $\ell_2$ regularized sparse representations are matched using a modified cosine similarity measure. Experimental results with extensive comparisons on challenging datasets demonstrate that the proposed method outperforms the state-of-the-art methods and using $\ell_2$ regularized sparse representation ($\ell_1 + \ell_2$) is more accurate compared to use a single $\ell_1$ or $\ell_2$ regularization term.

## 1. Introduction

The vast deployment of video surveillance cameras in public venues drives the needs for automated surveillance applications such as people tracking [21], anomaly detection in crowd [31], *etc*. Many of these applications require the ability to determine the ID of the subject in different camera views, which is a problem referred as person re-identification that is gaining more attentions in the literature recently [2,12,14,19,32,56,59,67,74]. Specifically, the goal of person re-identification is to accurately match individuals across non-overlapping cameras at different time and locations. The results of person re-identification can be readily used in further processing tasks such as tracklet association for multi-camera people tracking [7].

Despite of the plethora of advanced pattern recognition techniques developed in the past few years, the performance of person re-identification is still not robust enough to warrant high accuracy in practice. The difficulties for person re-identification involve the following aspects:

[*] Corresponding author. Tel.: +1 9516404692.
*E-mail address:* xchen010@ucr.edu (X. Chen).

http://dx.doi.org/10.1016/j.ins.2016.02.055
0020-0255/© 2016 Elsevier Inc. All rights reserved.

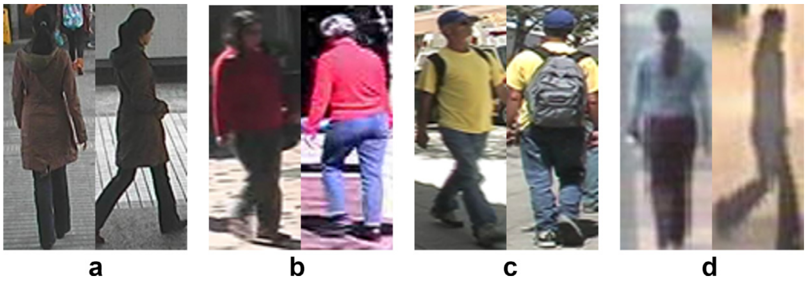

**Fig. 1.** Samples of image pairs of the same person in different camera views, showing (a) pose variation, (b) illumination change, (c) occlusion, and (d) low image quality, which make re-identification of people in different cameras a challenging problem.

1. *Pose variation*. In different camera views, a subject may have arbitrary poses (Fig. 1(a)),
2. *Illumination change*. The lighting condition is usually not constant in different camera views. As a result, the appearance of the same subject may vary significantly due to changing illumination (Fig. 1(b)),
3. *Occlusion*. A subject in one camera view may be fully or partially occluded by other subject or carrying items such as a backpack (Fig. 1(c)),
4. *Low image quality*. The captured image of a subject may suffer from low resolution, noise, or blur due to limited imaging quality of surveillance cameras (Fig. 1(d)).

In a person re-identification system usually two steps are involved: (1) extracting feature representations from person detections, and (2) establishing the correspondence between feature representations of probe and gallery. A gallery is a dataset composed of images of people with known IDs. A probe is the detection of a person from a different camera. Although other forms of biometrics such as face and gait [17,76] can be used to recognize people, however, acquiring such biometrics is difficult in uncontrolled low-resolution videos. For person re-identification, most of the existing approaches are appearance-based.

With the availability of tools for person detections, most of the previous work on person re-identification can be categorized into two groups:

1. Extracting feature representations which are robust against pose or illumination change, *e.g.*, [2,12,14,59].
2. Developing new matching methods using metric learning or ranking classifiers, *e.g.*, [19,26,53,74].

For the first group, discriminative appearance features are desired. Normally color and texture based features are widely used [19,28]. However, color or texture feature representations are sensitive to pose and illumination change, which may result in larger intra-person variation (difference between features of same person) than inter-person variation (difference between features of different persons). Besides low level image features, attribute or shape information has been applied in conjunction with color or texture features to improve the recognition accuracy [59].

To pursue more reliable matching, feature transformations or distance metrics are learned such that the distance between feature representations of the same person from different cameras is reduced while the distance between feature representations from different persons is increased [9,16,26,61]. SVM with ranking [53] and transfer learning [73] have also been proposed to obtain better matching correspondence.

In this paper, we propose a novel feature representation for person re-identification based on sparse coding. Inspired by coherent subspace learning to handle cross-type image synthesis [58] and face image super-resolution [1], we first learn a transformation to project the original image features into a subspace using canonical correlation analysis (CCA). In this learned subspace, the correlation between the features of the same people from different camera views is maximized. Then, two dictionaries for two camera views are jointly learned using training data in the coherent subspace. Given an image in the gallery, its image features are first projected into the CCA subspace and the sparse coefficients of this gallery subject are obtained using the learned dictionary with $\ell_2$ regularization. These coefficients become the new feature representation for this gallery instance. During re-identification, given a probe, its sparse representation is obtained in the same way using the corresponding dictionary. Fig. 2 illustrates the outline of the proposed method for generating the sparse representation. The matching is then performed by computing the similarity between the sparse representations of the probe and gallery.

A related work for person re-identification was introduced in which a sparse representation was directly learned using a dictionary [25]. The dictionary was composed of existing data without any learning and the identity of the probe was determined through the non-zero coefficients by majority voting rule. In contrast to our approach, the sparse representations in [25] were used for determining the identity of the probe, while in our method the sparse representations are used as new feature representations for matching. Another related method was proposed in [41], in which coupled dictionary learning was used. Our method is different from [41] in the following aspects: (1) the learning methodology is different. In [41], both labeled and unlabeled data are required to learn the coupled dictionaries, which is a semi-supervised framework. The unlabeled data are used to exploit the geometry of the data distribution. On the other hand, our framework is supervised and we do not require extra unlabeled training data to carry out learning; (2) The dictionaries are learned in different