Pattern Recognition Letters 71 (2016) 38–44

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A coupled discriminative dictionary and transformation learning approach with applications to cross domain matching☆

Sivaram Prasad Mudunuri, Soma Biswas*

Department of Electrical Engineering, Indian Institute of Science, Bangalore, Karnataka 560012, India

ARTICLE INFO

Article history:
Received 19 July 2015
Available online 18 December 2015

Keywords:
Face recognition
Activity recognition
Dictionary learning
Metric learning

ABSTRACT

Cross domain and cross-modal matching has many applications in the field of computer vision and pattern recognition. A few examples are heterogeneous face recognition, cross view action recognition, etc. This is a very challenging task since the data in two domains can differ significantly. In this work, we propose a coupled dictionary and transformation learning approach that models the relationship between the data in both domains. The approach learns a pair of transformation matrices that map the data in the two domains in such a manner that they share common sparse representations with respect to their own dictionaries in the transformed space. The dictionaries for the two domains are learnt in a coupled manner with an additional discriminative term to ensure improved recognition performance. The dictionaries and the transformation matrices are jointly updated in an iterative manner. The applicability of the proposed approach is illustrated by evaluating its performance on different challenging tasks: face recognition across pose, illumination and resolution, heterogeneous face recognition and cross view action recognition. Extensive experiments on five datasets namely, CMU-PIE, Multi-PIE, ChokePoint, HFB and IXMAS datasets and comparisons with several state-of-the-art approaches show the effectiveness of the proposed approach.

1. Introduction

Many applications require matching data coming from different domains or modalities, for example, we may want to compare a low-resolution uncontrolled face image captured using surveillance camera with a high-resolution image that is stored in the database, or given a text query, we may want to retrieve images (Fig. 1). But cross domain or cross modality matching (everything referred as cross-domain matching from now) is very challenging, since the data coming from different domains is usually very different. This is in addition to the intra-class variability that is present in the data of both the domains. Recently, there have been significant research efforts in addressing this problem and many approaches have been proposed. Among other methods, coupled dictionary learning has emerged as a powerful method for matching images coming from different domains. Though initially proposed for reconstruction applications, recent approaches have shown excellent performance in both reconstruction and classification problems [14].

In this work, we build upon the success of the coupled dictionary learning approaches and propose a coupled discriminative dictionary and transformation learning approach specifically designed for the classification/recognition tasks. First, data samples in the two domains are transformed in such a manner that they share common sparse representations with respect to their own dictionaries in the transformed space. For improving discriminability for recognition tasks, the sparse coefficients are further mapped such that the k-nearest neighbors of the same class move closer and those of different classes are pushed far apart. This ensures that the final sparse coefficients obtained using the proposed approach are discriminative enough to distinguish between features from different classes. The dictionaries, sparse coefficients and all the transformation matrices are jointly updated in an iterative manner. During testing, the features from the two domains are first transformed using the learnt feature transformation matrices before computing the sparse coefficients. These coefficients are further transformed using the learnt discriminative mapping and then used for matching. The applicability of the proposed approach is illustrated by evaluating its performance on different challenging tasks: face recognition across pose, illumination and resolution, heterogeneous face recognition (matching visible with NIR images) and cross view action recognition. Extensive experiments on five datasets namely, CMU-PIE, Multi-PIE, ChokePoint, HFB and IXMAS

☆ This paper has been recommended for acceptance by Dr. L. Yin.

* Corresponding author. Tel.: +91 7204012320.
E-mail address: soma.biswas@ee.iisc.ernet.in (S. Biswas).

http://dx.doi.org/10.1016/j.patrec.2015.12.003

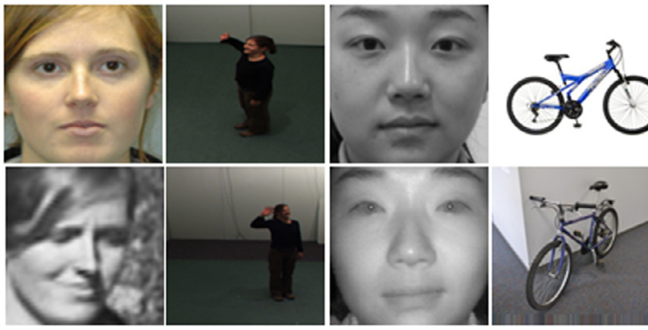

Fig. 1. Applications of cross domain matching. (a) high-resolution controlled image with low-resolution uncontrolled image as in surveillance scenario; (b) cross view action recognition; (c) NIR images with visible light images; (d) Object recognition from retailer sites and consumer images.

datasets and comparisons with several state-of-the-art approaches show the effectiveness of the proposed approach. The proposed formulation shares similarities with the seminal work by Huang and Wang [14]. The differences of the proposed approach over [14] are as follows:

- The original features are transformed using the feature transformation matrices.
- Both the domains in the transformed space have the same sparse coefficients.
- Learning a classifier on the sparse coefficients is not always possible where the test classes are not the same as the training classes as in face recognition applications as done in [14].
- There is an explicit discriminative term which significantly improves the performance of coupled dictionary approaches for recognition and classification tasks as shown by the experimental results.

The rest of the paper is organized as follows. Section 2 discusses the related work. The proposed approach is presented in Section 3 and the experimental results are presented in Section 4. The paper concludes with a discussion and conclusion section.

2. Related work

In this section, we discuss the papers in the literature which are closely related to the proposed approach. Dictionary learning-based algorithms have recently been successfully applied in many applications. Wright et al. [35] propose a sparse representation based classification approach for robust face recognition. Since then, several approaches have been proposed to learn discriminative dictionaries which can help in distinguishing between different classes [15,25,39,44] and [38]. A sparse modeling and dictionary learning based method for clustering and classification of different classes is presented in [27]. Wang et al. [33] propose a semi-coupled dictionary learning method that aims at reducing the distance between the sparse coefficients of the same subject belonging to different domains by transforming one of the domain features to the new space. Huang and Wang [14] propose a coupled dictionary and feature space learning algorithm that iteratively updates two separate dictionaries and transformation matrices to transform the sparse coefficients. Matching cross-domain data using joint dimensionality reduction techniques like Canonical Correlation Analysis [12], Partial Least Squares [28], etc. have also been quite successful. In [29], the covariance between the sets is jointly optimized and the classes are also separated in their respective feature spaces. Multi-view Discriminant Analysis [17] finds a discriminative common space for all the views by jointly learning multiple view-specific linear transforms. Domain adaptation methods like [21] have also been successfully applied for cross-domain matching tasks.

Cross-domain matching has several applications. In this work, we have focused mainly on face recognition across pose and resolution, NIR-vs-Visible face matching and cross-view action recognition and we will provide pointers to some related works in these areas. Recognizing faces across multiple variations like illumination, resolution, pose, etc. has received considerable attention [26,42,43] and [7]. A co-transfer learning framework, which combines transfer learning with co-training for matching faces across resolutions is proposed in [3]. Zou and Yuen [47] propose an algorithm to perform matching of low resolution facial images by learning the relationship between high-resolution gallery and the low-resolution probe images. For matching near infra-red and visible facial images, Zhu et al. [46] propose a transductive heterogeneous face matching approach that can reduce the modality gap by extracting the domain invariant and target-related discriminative features. Hou et al. [13] propose an approach to derive a common space which can relate and represent facial images of different modalities. Jin et al. [16] propose a method that can learn several image filters to simultaneously utilize discriminative information and reduce the appearance difference of facial images captured across different modalities. Lu et al. [23] propose a compact binary face descriptor (CBFD) for matching facial images of different domain. Lei et al. [19] propose a discriminant face descriptor (DFD) that can enhance the discriminative capacity of face representation. The method also formulates a coupled DFD feature that can further improve the performance of matching across different modalities.

For cross-view action recognition, Wang et al. [31] propose a method to learn action units using the graph regularized nonnegative matrix factorization from the extracted novel spatial-temporal descriptors. A multiview spatio-temporal representation based approach to handle the problem of cross view action representation is discussed in [32] and [22]. An approach to learn view-invariant sparse representations to perform cross-view action matching is described in [45]. An approach that constructs animated pose templates to detect short-term, long-term, and contextual actions from cluttered scenes in videos can be found in [40]. Wu et al. [36] propose an approach to construct a common feature space to link source view and target view for transferring knowledge between them. Yeh et al. [41] propose a method that can exploit the domain transfer ability in the correlation subspace.

3. Proposed approach

In this section, we present the proposed joint dictionary and transformation learning algorithm for matching data from two different domains.

3.1. Problem formulation

Let $\mathbf{X}_1 \in \mathbf{R}^{d_1 \times N}$ and $\mathbf{X}_2 \in \mathbf{R}^{d_2 \times N}$ be the two matrices that represent features computed from the two domains. Here, d_1 and d_2 are the length of the feature vectors and N denotes the number of training images. The goal is to learn transformation matrices, dictionaries and a mapping function such that the following two criteria are satisfied:

- The transformation matrices should transform the input features from the different domains such that they have the same sparse representation with respect to their own dictionaries in the transformed space.
- The mapping function should be capable of moving the k-nearest sparse coefficient vectors of the same class closer and simultaneously those of different classes apart.