Pattern Recognition 67 (2017) 353–365

Contents lists available at ScienceDirect

Pattern Recognition

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

Discriminative pose-free descriptors for face and object matching

Soubhik Sanyal, Sivaram Prasad Mudunuri, Soma Biswas*

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

ARTICLE INFO

Article history:
Received 25 July 2016
Revised 25 December 2016
Accepted 10 February 2017
Available online 17 February 2017

Keywords:
Face recognition
Object recognition
Pose invariant matching
Metric learning
Canonical correlation
Subspace to point representation.

ABSTRACT

Pose invariant matching is a very important problem with various applications like recognizing faces in uncontrolled scenarios in which the facial images appear in wide variety of pose and illumination conditions along with low resolution. Here we propose two discriminative pose-free descriptors, Subspace Point Representation (DPF-SPR) and Layered Canonical Correlated (DPF-LCC) descriptor, for matching faces and objects across pose. Training examples at very few poses are used to generate virtual intermediate pose subspaces. An image is represented by a feature set obtained by projecting its low-level feature on these subspaces and a discriminative transform is applied to make this feature set suitable for recognition. We represent this discriminative feature set by two novel descriptors. In one approach, we transform it to a vector by using subspace to point representation technique. In the second approach, a layered structure of canonical correlated subspaces are formed, onto which the feature set is projected. Experiments on recognizing faces and objects across pose and comparisons with state-of-the-art show the effectiveness of the proposed approach.

1. Introduction

Matching faces (or objects) across wide variety of poses is a very important area of research in the field of computer vision with many applications. For example, in surveillance setting, the face of a person captured by the overhead cameras may be in any arbitrary pose and poor resolution as opposed to the frontal image under high resolution that is typically captured during enrolment (Fig. 1, column 1 and 2). For object matching, the images captured during testing can be taken from a different view-point compared to the images stored in the database which again requires comparing objects in different poses (Fig. 1, column 3–6). The aforesaid tasks are challenging because the appearance of the images to be matched can be very different due to significant pose variations.

In this paper, we propose two discriminative pose-free descriptors, Subspace Point Representation (*DPF-SPR*) descriptor (which is also termed as DPFD in [1]) and Layered Canonical Correlated (*DPF-LCC*) descriptor, for matching faces and objects across different poses. During training phase, images from a few poses (two to three) are used to generate virtual subspaces for the intermediate poses. We generate the virtual intermediate subspaces by treating the subspaces generated by the training data as points on the Grassmann manifold and sampling the shortest geodesic path between those points. Then, we represent an image (or image region depending on application) by a set of features, computed by projecting its low level feature vector onto all the intermediate subspaces, which will ensure that at least one or more of the features from the entire feature set will match if the images with different poses are compared. Since our final goal is recognition, we use a discriminative transform learned using the class labels of the training data to transform the feature set. Then *DPF-SPR* or *DPF-LCC* descriptor is computed from the feature set which can be directly used for matching. In this paper, our focus is on the following challenging tasks:

1. Unconstrained face recognition, where the gallery consists of frontal images captured during enrolment and the probe images can be in any arbitrary pose. We also address the problem where, in addition to non-frontal pose, the probe images also have low-resolution as is usually the case in surveillance setting when the images are taken from a large distance from the subject. We perform extensive experiments on the CMU PIE [2], Multi-PIE [3] and the SCface database [4].
2. Object recognition across pose, where the objects of different poses are to be matched. For this purpose we evaluate the proposed approach on COIL-20 [5] and RGB-D object datasets [6]. We also consider the task of matching depth images of objects across pose to test the generalizability of the proposed approach.

We compare the proposed approach with state-of-the-art metric learning, cross-pose methods, domain adaptation and coupled dictionary learning approaches to show the effectiveness of the two descriptors. The novelty/contribution of the work is as follows:

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

http://dx.doi.org/10.1016/j.patcog.2017.02.016

Fig. 1. Applications requiring matching across pose variations. Column 1,2: Face recognition in uncontrolled setting; and Column 3–6: Object recognition across viewpoint.

- Two novel discriminative pose-free descriptors, Subspace Point Representation (*DPF-SPR*) and Layered Canonical Correlated (*DPF-LCC*) descriptor, for matching faces and objects across different views are proposed.
- The approach does not require separate training for different probe poses/view points. This is an advantage over many other approaches which work well when separate training is performed for different poses encountered during testing.
- Very few poses (as little as two or three) are required during the training phase and the method can generalize well to unseen poses.
- Extensive experiments illustrate the applicability of the proposed approach in diverse domains like face and object recognition.

A preliminary version of this work appeared in [1]. The rest of the paper is organized as follows. Section 2 describes the related literature. Details of the proposed approach are given in Section 3. The experimental results and analysis of the descriptors are reported in Section 4. The paper concludes with a brief discussion section.

2. Related work

In this section, we provide pointers to some of the related work in the area of recognizing faces and objects across pose. Recognizing faces across pose is an important research area. Li et al. [7] propose maximal likelihood correspondence estimation after encoding face specific structure information of semantic correspondence. Ding et al. [8] learn a transformation dictionary which transforms the features of different poses into a discriminative subspace where face matching is performed at patch level rather than at holistic level. Ding et al. [9] extract multi-directional multi-level dual cross patterns as pose invariant face descriptor. Zhang et al. [10] propose a mixed norm approach which is achieved by a trade-off between sparse representation classification and joint sparse representation classification. Castillo and Jacobs [11] propose a method to compute stereo matching cost between two facial images by using epipolar geometry. Wang et al. [12] formulate a general representation of kernel collaborative framework and develop an l_2 regularized algorithm within it. Cament et al. [13] extract Gabor features from modified grids using mesh to model face deformations produced by varying pose. Li et al. [14] train a Gaussian mixture model to capture the spatial appearance distribution of all face images in the training corpus. Yin et al. [15] design a model for recognizing faces under large variations which can predict the appearance and likelihood of the given query face against the collected generic identities. Arandjelovic et al. [16] propose a framework that can improve the performance of any baseline face retrieval algorithm by leveraging the structure of the database.

Recently, matching of low resolution facial images has gained considerable attention [17,18]. Bhatt et al. [19] propose using a combination of transfer learning and co-training paradigms for cross resolution matching. Ren et al. [20] propose a method of extracting and representing discriminant feature from faces and then alternatively optimize across different data domains. Al-Maadeed et al. [21] learn a pairwise dictionary and utilize a random pooling strategy to select a subset of visual words. Zhao et al. [22] combine forward and backward sparse representation for robust face recognition. A piecewise linear regression model is developed to learn the relationship between the high resolution (HR) image space and the low resolution (LR) image space for face super resolution in [23]. Metric learning approaches have shown a lot of promise for matching faces in unconstrained environments. Kostinger et al. [24] propose a method that learns a distance metric from the co-variance matrices of similar and dissimilar pairs. Moutafis and Kakadiaris [25] propose an algorithm that can match HR and LR facial images by learning individual basis for optimal representation and coupled distance metrics to enhance the classification. Domain adaptation techniques have also been successfully used for matching face images across pose, illumination, blur, etc. [26]. In [27], dictionary learning is used to interpolate subspaces to link the source and target domains. The main drawback of most of these approaches is that they perform well only if the test faces have the same pose as those of the faces used for training. This is the key motivation in developing our algorithm, so that the training can be done using a few representative poses, and the testing image can have a different pose than those used for training.

There has been a lot of progress in the field of deep learning in face and object recognition tasks [28] in recent times. Taigman et al. [29] apply a piecewise affine transformation for 3D modeling of faces using deep convolutional networks. Schroff et al. [30] propose to learn a deep convolutional network by mapping from face images to an Euclidean space. Chan et al. [31] build a deep learning network with the help of cascade principle component analysis, binary hashing and block-wise histograms. Though the deep learning methods have shown better performance under wide range of pose variations, the performance degrades when the resolution is low [32,33]. Handling the low resolution problem with deep learning approaches requires sufficient amount of training samples captured under poor resolution conditions for learning the model.

There has also been a lot of research in the area of general object recognition across different viewpoints [34]. Hsiao and Hebert [35] model occlusions by reasoning about 3D interactions of object. Rubio et al. [36] use generative non-negative matrix factorization to find out relevant parts for training instances. Wu et al. [37] propose a query-expanded collaborative representation based classifier with class-specific prototypes. A model that separates a view-invariant category representation from category-invariant pose representation is proposed in [38]. He et al. [39] use spatial pyramid pooling strategy to eliminate the need of fixed size input image for a convolutional neural network for object recognition.

One of the proposed descriptors is inspired by [40] in which images are matched across varying scales. Features at different scales can be computed from the same image itself, unlike features at different poses which is the focus of our work. Generating intermediate subspaces by sampling the Grassmann manifold has also been exploited by [26], and then the projections on these subspaces are used to train discriminative classifiers for each object. Instead, using the intermediate subspaces, we form a discriminative feature vector which can directly be used for matching. Our approach is more suitable for applications like face recognition, where there may not be any overlap between the training and testing subjects.

3. Proposed approach

In this section, we describe in detail the computation of the two discriminative pose-free descriptors, namely, *DPF-SPR* and *DPF-LCC*. The different steps required during training are: