J. Vis. Commun. Image R. 40 (2016) 237–250

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Multiple clusters parts-based sparse representation for single example face identification ☆

Bin Wang [a], Cungang Wang [b,*], Jifeng Huang [a]

[a] College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China
[b] School of Computer Science, Liaocheng University, Liaocheng 252400, China

ARTICLE INFO

Article history:
Received 18 January 2016
Revised 19 June 2016
Accepted 22 June 2016
Available online 23 June 2016

Keywords:
Parts-based
Single example
Face identification
Sparse representation

ABSTRACT

While great progress has been made on face identification, it is still highly challenging under real-world conditions, e.g., illumination variation and occlusion with insufficient labeled examples. In this paper, we proposed a probabilistic generative model and a face identification approach based on the model to attack the above challenge. The proposed model learns multiple clusters parts-based representation built upon three concepts: part, component and cluster. Face parts, represented by the linear combination of a set of components, refer to regions like eye and mouth. Cluster refers to a group of faces with similar appearance. A face is the summation of all parts. Based upon the learned face representation, we derived a similarity measure over faces for single example face identification. The proposed approach is able to fully exploit unlabeled data and face representation can provide informative features. We validate the effectiveness of the proposed approach through a set of experiments.

1. Introduction

Motivated by the wide range of real-word applications [1–5] in forensics and security, face identification has been a popular research topic in computational intelligence community over the past decades. While great progress has been made on face identification techniques towards a number of difficult obstacles, it is still highly challenging to identify faces under real-word conditions, most typically, variant illumination, occlusion with insufficient labeled examples. In this paper, we seek to attack the above challenging task by building an effective face representation which is informative under insufficient labeled examples and simultaneously robust to illumination and occlusion. We consider the most frequent case in real-word applications, one labeled example per subject with variant illumination and partial occlusion.

To find an effective face representation, we start from the methodology of generating face images. As a general observation, human faces can be viewed as the assembly of a certain number of typical parts, which suggests that the face images can be generated through a part-by-part fashion. Usually, the generating process can be implemented by appropriately designed probabilistic generative models, resulting in parts-based representation. The parts correspond to face organs, e.g., mouth, eyes and nose. Given a set of human face images, learning parts-based representation is to find the typical shapes of parts, the common structure among parts, the representation of parts, such that the generative models can fit the training images with least error.

Several models [6–9] have been proposed to learn parts-based face representation. Among them, Multiple Cause Vector Quantization (MCVQ) [6] and Structured Sparse Principle Component Analysis (SSPCA) [7] show attracting abilities. They are able to learn the face parts unsupervisedly, without needing to labeled examples, and the learned parts are flexible in representation. However, they exhibit two limitations in face representation. First, the parts learned by MCVQ might be fragmentary while the parts learned by SSPCA are limited to convex shapes which might not always match the shapes of face parts perfectly. See Fig. 9 for examples. Second, these models are not always capable enough to model the complexly distributed data and fully capture the discriminative information for identification. Fig. 5 presents an illustrative example. Mixture MCVQ (MMCVQ) [10] is an extension of MCVQ and achieves significantly improvement. However, its capability is still limited due to its discrete state representation. Motivated by these observations, we propose multiple clusters parts-based sparse representation (MCSR) to increase the discrimination power of face representation and subsequently improve the performance of face identification.

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: cungwang@163.com (C. Wang).

http://dx.doi.org/10.1016/j.jvcir.2016.06.019
1047-3203/

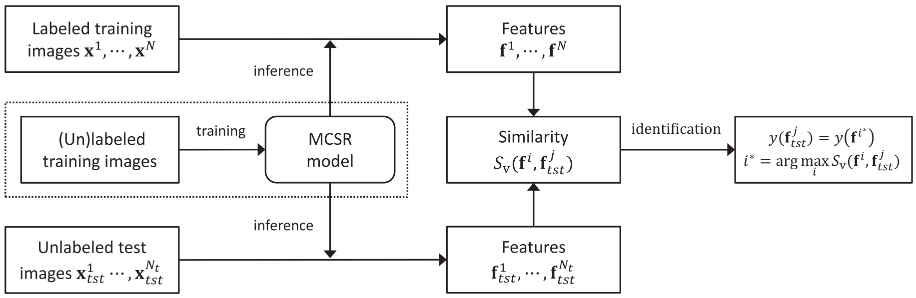

Fig. 1. The framework of the proposed single example face recognition approach.

MCSR is built upon the concepts of three levels, i.e., cluster, part and component. It generates face images pixel-by-pixel. To generate a pixel, it first randomly chooses a cluster and a part, and then chooses a set of combination coefficients. Given the chosen cluster, part and coefficients, the pixel is at last chosen from a sparse representation model. For MCSR, we proposed a learning method based on variational Expectation-Maximization (EM) algorithm [11]. Further, at the inference step, we use a variant of the posterior regularization [12,13,3] to implement regularized inference to ensure the learned parts being continuous. Having the learned parts-based representation by MCSR, we then construct the similarity measure with which the single face identification problem becomes, for a test face image, assigning the label of its nearest labeled example to it. The framework of our approach is shown in Fig. 1.

The contributions of this paper are threefold:

1. We propose a probabilistic generative model MCSR to learn the parts-based face representation and derive the similarity measure based on the model for single example face identification.
2. We propose a regularized inference approach based on posterior regularization to learn better face parts.
3. The multiple clusters representation introduced in MCSR is able to improve the capacity of modeling complex data and provide similarity measure with stronger discriminative power.

The remainder of this paper is organized as follows. We first review the related works in Section 2. Section 3 presents the multiple clusters parts-based sparse representation (MCSR), and Section 4 derives the inference and parameter estimation recipes. Section 6 evaluates the proposed method on four experiments. Section 7 draws a conclusion.

2. Related works

In this section, we will briefly review those approaches and models mostly related to our approach, i.e., the approaches for single example face identification, and models for parts-based representation which shows promising performance in single example face identification.

2.1. One example face identification

A number of works [1,2] on face identification have been published over the past decades. Among them, many efforts were devoted to improve the identification performance under distinct imaging conditions. With sufficient number of labeled training samples taken under well controlled imaging conditions, some current approaches can be robust against to varying imaging conditions. However, in real-world applications, the cameras can only scan the faces of subjects within limited time and under uncontrolled imaging conditions, which indicates that the number of labeled examples per subject is quite few and the obtained images are distributed with large variance. Subsequently, most traditional face identification systems mentioned above will suffer from insufficient labeled training examples [14–16], considering the illumination variation and occlusion.

One example face identification is the extreme case of face identification with insufficient labeled training examples, where only one labeled training example per subject is available. A comprehensive survey on this problem can be found in [17]. The primary challenge of single example face identification is that the identification model is difficult to cover the variation of imaging conditions using only one labeled training example per subject. To attack this problem, a branch of approaches seek to exploit the knowledge or the distribution of face images. For instance, [18,19] built models to deal with the variations in lighting conditions. [20,21] proposed to learn face representation using semi-supervised subspace learning.

Identifying an object by first recognizing its parts [6,22] is a widely used methodology in computer vision and pattern recognition, supported by the observation that most objects in real-world can be viewed as the assembly of a certain number of typical parts. A representative example is the method of Recognition-by-Components[1] proposed in [23]. This methodology directly leads to the parts-based representation. In [24,25], face part detectors were constructed based on discriminative classifiers, in the fashion of supervised learning, and are trained using the labeled samples of the manually chosen parts. The detectors perform part detection by scanning the image and return the windows with strong response. In [26], the face part detectors were built into a hierarchical model and perform detection by sliding over the window, also in supervised fashion. The performance of using parts for face identification was evaluated in [27]. However, learning these detectors needs a number of labeled part samples manually collected form training images. Different from the above methods based on supervised learning, the method proposed in [28] discovered face parts automatically rather than chooses the parts manually. In [6], it learned parts-based representation in an unsupervised fashion, where the pixels are assigned to parts through Bayesian inference such that the resulted parts can explain the given data with least error.

2.2. Multiple cause vector quantization

MCVQ [6] is a probabilistic generative model designed to learn parts-based representation, with latent variables introduced, where cause and factor refer to image part and quantization state respectively. MCVQ models the image vector $\mathbf{x} \in \mathbb{R}^D$ as the output of a multiple cause model, and generates an image pixel-by-pixel.

[1] The notation 'component' corresponds to our natation 'part'.