Computer Vision and Image Understanding 122 (2014) 143–154

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Exploiting relationship between attributes for improved face verification ☆

Fengyi Song, Xiaoyang Tan *, Songcan Chen

Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China

ARTICLE INFO

Article history:
Received 27 March 2013
Accepted 19 February 2014
Available online 1 March 2014

Keywords:
Attribute relationship graph
Attribute-graph regularized SVM
Face verification

ABSTRACT

Recent work has shown the advantages of using high level representation such as attribute-based descriptors over low-level feature sets in face verification. However, in most work each attribute is coded with extremely short information length (*e.g.*, "is Male", "has Beard") and all the attributes belonging to the same object are assumed to be independent of each other when using them for prediction. To address the above two problems, we propose a discriminative distributed-representation for attribute description; on the basis of this description, we present a novel method to model the relationship between attributes and exploit such relationship to improve the performance of face verification, in the meantime taking uncertainty in attribute responses into account. Specifically, inspired by the vector representation of words in the literature of text categorization, we first represent the meaning of each attribute as a high-dimensional vector in the subject space, then construct an attribute-relationship graph based on the distribution of attributes in that space. With this graph, we are able to explicitly constrain the searching space of parameter values of a discriminative classifier to avoid over-fitting. The effectiveness of the proposed method is verified on two challenging face databases (*i.e.*, LFW and PubFig) and the a-Pascal object dataset. Furthermore, we extend the proposed method to the case with continuous attributes with promising results.

1. Introduction

Recently, there has been growing interest in using middle-to-high level feature descriptors for face representation. One typical example is the attribute descriptors [1–9]. N.Kumar et al. [3,4] have recently shown that using the outputs of a series of component classifiers with each tailored to some particular aspects of the human face images, called visual attributes, they are able to achieve close to state-of-the-art performance of face verification on the challenging Labeled Faces in the Wild (LFW) [10]. This result is interesting in several aspects. Firstly, the number of features used in their work is very small (*i.e.*, only 73 attributes), which means that it provides a very economical but powerful way to describe faces. This is in sharp contrast with the commonly used low-level features in image description, such as pixel values, gradient directions, scale-invariant feature transform (SIFT) [11], where usually thousands of features are needed. Secondly, the attribute descriptor is user-friendly in that its meaning is understandable to human beings (everyone knows what "white male" means) while the meaning of most previously mentioned low-level features is less intuitive to us. Last but not least, such a descriptor is generalizable and sharable, which makes it particularly suitable for such problems as zero-shot learning [12,13] or between-class transfer learning [2,14].

However, in most work each attribute is coded with extremely short information length (*e.g.*, using binary code such as "is Male", "has Beard") and all the attributes belonging to the same object are assumed to be independent of each other when using them for prediction. The one-bit information length of attribute coding makes the representation less stable, and could bring trouble to many interesting subsequent processing tasks, such as modeling the similarity between attributes. Actually, research in the field of cognitive discovery has shown the usefulness of the relationship between feature sets. For example, Bhatt and Rovee-Collier [15] experimentally showed that infants as young as three months of age gain the capability to encode the relations among object features, and use such a feature configuration for general object recognition. However, traditionally one of the major challenges in modeling the feature configurations lies in the huge number of low-level features (*e.g.*, the dimension of a 100×100 face image is as high as $10,000$ using the gray-value features). In addition, it

☆ This paper has been recommended for acceptance by Aleix M Martinez.
* Corresponding author. Fax: +86 25 8489 2452.
E-mail addresses: f.song@nuaa.edu.cn (F. Song), x.tan@nuaa.edu.cn (X. Tan), s.chen@nuaa.edu.cn (S. Chen).

http://dx.doi.org/10.1016/j.cviu.2014.02.010
1077-3142/

is very difficult for a human being to understand what exactly such a big feature configuration mean. Fortunately, both aforementioned problems can be addressed by the attribute descriptors due to its high level and compactness in object description. Indeed, despite the partial success of using attribute descriptors by treating them statistically independent of each other [1,3,4,16] or conditionally independent given the class label [2], recent work has shown that it is beneficial to exploit the relationship between attributes under various contexts [5,17–19]. Some of them will be discussed in the next section.

In this work, we propose a discriminative distributed-representation for attribute description; on the basis of this description, we investigate how to model the similarity relationship between attributes and how such relationship could be exploited to improve the performance of face verification. The idea of distributed representation was first introduced by Hinton [20], and successfully applied in statistical language modeling [21]. In this work, we develop a new distributed representation for each individual attribute by taking the information of subject identification into account. The method is inspired by the vector representation of words in the literature of text categorization, and the meaning of each attribute is embedded into a high-dimensional vector in the subject space (*cf.* Fig. 1). Such a representation allows us to model the similarity between attributes in a much stable and reliable way. In particular, we construct an attribute-relationship graph based on the distribution of attributes in the subject space, which effectively encodes the pairwise closeness relationship between any two attributes. For example, a "male" attribute is highly related to such attributes as "wearing necktie", "bushy eyebrows", "beard", and so on (*cf.* Fig. 9). To exploit such information for prediction, we integrate the attribute-relationship graph into a linear classifier to constrain the searching space of its parameters, based on the assumption that similar attributes should have similar weights. This is helpful to avoid over-fitting and improve the generalization capability of the learned classifier. The uncertainty in attributes responses is also taken into account in the final model.

This journal paper builds on the earlier conference work [22]. In this extended version, we extend above ideas and merge them into a single framework, which works for both discrete and continuous attributes. The effectiveness of the proposed methodology is empirically verified with encouraging results on two large-scale face databases, one object classification dataset and several UCI data sets. In what follows, we first review the related work in Section 2 and then present the proposed method in Section 3. Extensive experimental results are given in Section 4. Finally, we conclude this work in Section 5.

2. Related work

Recently, attribute-based representation has been extensively researched in and beyond the field of face recognition [3,4], including object recognition [5,9,23], scene understanding [18,24], image retrieval [25,26], activity analysis [27–29], and shows special advantages in active learning [30–32], transfer learning [2,12,14] and zero-shot learning [13]. Since this work is mainly about attribute representation and modeling their relationship, in what follows, we will not go into the details on how to extract attributes and apply them in various applications, but first give a brief discussion on how to define attributes and then focus on the related work on building the relationship between attributes.

2.1. Attribute definition

To use the attributes, we have to define them firstly. Attribute definition is the process of deciding which visual qualities should be used for depicting the objects or events. Most attributes are manually specified with respect to different application scenarios. These attributes are usually semantically understandable and can be seen as concepts in natural language. The specified attributes are then extracted from the images based on some low-level features. In this way, attribute can be thought of as a high-level representation which incorporates human understandable concepts into the machine learning process in a reasonable way.

Although attribute description is mostly intuitive, building a suitable taxonomy of attributes for a particular task is not easy. In [3], the authors proposed to describe each face with 73 attributes (*cf.* Fig. 2), which can be roughly categorized into four types: (1) appearance description of key facial parts, such as the shape, size and style of the nose, mouth, eyes, eyebrow, jaw, and hair; (2) high-level semantic features like gender, age, and ethnicity; (3) specification about imaging conditions, *e.g.*, lighting, expression, posture, accessory, and the environment; and (4) personal specific traits like bald, goatee, and attractiveness. In [9], by surveying multiple online cataloges, the authors produced 26 common attributes to describe clothing, covering 6 patterns, 11 colors, and 6 miscellaneous characteristics such as wearing the necktie or the scarf, and the collar or the placket presence. Patterson and Hays [33] gave a comprehensive discussion on attribute definition, discrimination and predictive power of attribute in the context of scenes description.

2.2. Attribute relationship exploitation

We now review how to model and exploit the relationship between attributes. As mentioned before, this is not trivial because an attribute is usually simply represented as a binary bit to denote its presence/absence. Despite this, there is some work which exploit various types of attribute relationship in different contexts to improve the performance of prediction.

In [19], the concept of binary attribute was introduced to describe the spatial relationship between a pair of attributes corresponding to two image segments respectively. Such relationship was shown to be very effective in describing simple geometric

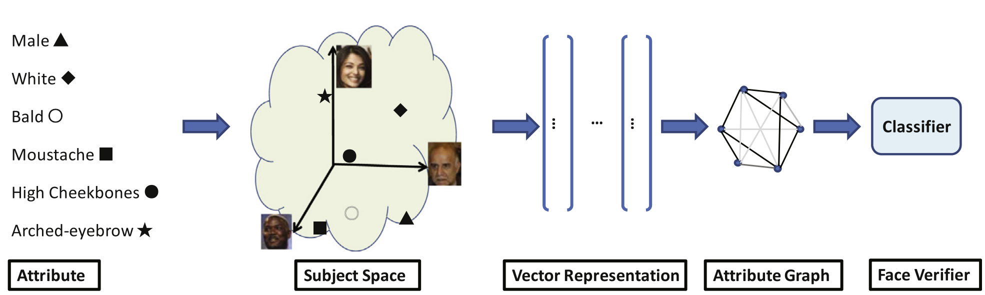

Fig. 1. The overall pipeline of the proposed algorithm. Each attribute descriptor is first projected into a common subject space to obtain a high-dimensional vector representation, which are then used to construct an attribute graph. The graph is finally exploited to regularize the objective of a linear SVM-based face verifier.