Exploiting Relationship between Attributes for Improved Face Verification

Fengyi Song
f.song@nuaa.edu.cn
Xiaoyang Tan
x.tan@nuaa.edu.cn
Songcan Chen
s.chen@nuaa.edu.cn

Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, P.R. China

Recently work has shown the advantages of using attribute representation over low-level feature descriptors in face verification [6], due to its capability to explicitly encode high-level semantic meaning with economical coding bits. These advantages allow us to explicitly investigate the similarity relationship between attributes and to see how such relationship could be exploited to improve the performance of face verification. Actually, research in the field of cognitive discovery has shown that infants as young as 3 months of age gain the capability to encode the relations among object features, and use such feature configuration for general object recognition [2]. Indeed, despite of the partial success of using attribute descriptors by treating them statistically independent to each other [4, 6], recent work has shown that it is beneficial to exploit the relationship between attributes under various contexts [7][3][5]

In this paper we proposed a novel method to model the relationship between attributes and exploit such information to improve face verification. In particular, we first represent the meaning of each attribute as a high-dimensional vector in the subject space, which enable us to conveniently construct the corresponding attribute-relationship graph based on the distribution of attributes in that space (c.f., Fig.1). The resulting attribute-relationship encode the pairwise closeness relationship between any two attributes, which are further integrated into a linear classifier to constrain the searching space of its parameters, based on the idea that similar attributes should have similar weights. This is helpful to avoid overfitting and improve the generalization capability of the learned classifier. We also extend the model to handle uncertainty in attribute responses.

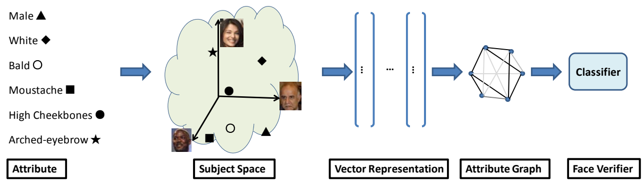


Figure 1: The overall pipeline of the proposed algorithm. Each attribute descriptor is first projected into a common subject space to obtain a high-dimensional vector representation, which are then used to construct an attribute graph. The graph is finally exploited to regularize the objective of a linear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

The overall pipeline of our algorithm is presented in Fig.1, with the following major steps:

Subject space projection. Although the meaning of each attribute is clear to human beings, the way to represent each attribute as a real number is too simple to estimate their correlation. Therefore, we need to represent each attribute in a richer manner to support more advanced inference. Our method is inspired by the vector representation of words in the literature of text categorization. Assuming that we are given a set of M attribute descriptors $A = \{A_i \in R\}_{i=1}^{M}$ for each face image. We use the subjects available in the training set and call the space spanned by these subjects subject space (see Fig.1). Hence for K subjects, we have a subject space with K-dimensions and the meaning of each attribute is represented as a high-dimensional vector in the subject space, with each entry representing whether the corresponding subject owns such attribute.

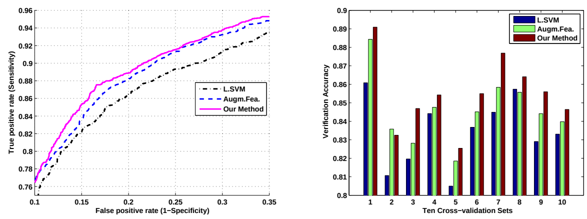
Attribute graph building. Through projecting all the attributes into the subject space, we may model their relationship based on the distribution of each attribute in an information theory framework. In particular, we first compute the point-wise mutual information $I(A_i, y_j)$ of each attribute A_i with each subject with label y_j. After this, correlated information encoded by M attributes and K subjects is organized as the following matrix, based on which, the attribute graph can be constructed by treating each row as a node.

$$\begin{array}{c} \\ A_1 \\ \vdots \\ A_i \\ \vdots \\ A_M \end{array} \left(\begin{array}{ccccc} y_1 & \cdots & y_j & \cdots & y_K \\ I(A_1, y_1) & \cdots & I(A_1, y_j) & \cdots & I(A_1, y_K) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I(A_i, y_1) & \cdots & I(A_i, y_j) & \cdots & I(A_i, y_K) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I(A_M, y_1) & \cdots & I(A_M, y_j) & \cdots & I(A_M, y_K) \end{array} \right) \tag{1}$$

Attribute graph constraints. After the weighted attribute graph is built, its Laplacian L is then constructed. We add this as an extra regulariser into the standard SVM objective function and use it to regularize the identity prediction for face images, as follows,

$$\min_{w} \sum_{i=1}^{N} \max\{0, 1 - y_i(w^T x_i + b)\} + \frac{\lambda_1}{2} w^T w + \frac{\lambda_2}{2} w^T L w \tag{2}$$

The above formulation is similar to that of Laplacian SVM [1], but instead of constructing an instance-graph, we build an attribute-relationship graph. One advantage of attribute-graph is that its complexity is controllable since its size will not grow with the number of instances as in [1] but only with the number of attributes. Furthermore, our graph is not meant to constrain the output space of instances but the searching space of model parameters, based on the simple idea that similar attributes should play similar roles in the learnt classifier.



(a) (b)

Figure 2: Comparison of our method with [6][7] on the LFW dataset: a) overall ROC curve, b) detailed results on 10 cross-validation test sets.

Detailed implementations are described in the paper. Fig.2 gives the major experimental results, which indicates that the performance of the attribute-based face verification method can be improved by regularizing it with attribute relationships graph induced from subject space.

[1] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *JMLR*, 7:2399–2434, 2006.

[2] R.S. Bhatt and C. Rovee-Collier. Infants' forgetting of correlated attributes and object recognition. *Child development*, 67(1):172–187, 1996.

[3] Lubomir Bourdev, Subhransu Maji, and Jitendra Malik. Describing people: Poselet-based attribute classification. In *ICCV'11*, 2011.

[4] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *CVPR'09*, pages 1778–1785, 2009.

[5] V. Ferrari and A. Zisserman. Learning visual attributes. In *Advances in Neural Information Processing Systems*, 2007.

[6] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Describable visual attributes for face verification and image search. *PAMI*, (99):1–1, 2011.

[7] Y. Wang and G. Mori. A discriminative latent model of object classes and attributes. *ECCV'10*, pages 155–168, 2010.