A Maximum Entropy Feature Descriptor for Age Invariant Face Recognition

Dihong Gong[1], Zhifeng Li[1], Dacheng Tao[2], Jianzhuang Liu[3,4], Xuelong Li[5]

[1]Shenzhen Key Lab of Computer Vision and Pattern Recognition
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[2]Centre for Quantum Computation & Intelligent Systems, Faculty of Engineering and IT, University of Technology, Sydney, NSW 2007, Australia
[3]Dept. of Information Engineering, the Chinese University of Hong Kong
[4]Media Lab, Huawei Technologies Co. Ltd., China
[5]Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences

Age invariant face recognition (AIFR) is an important but challenging area of face recognition research. Most existing works focus on age estimation [1,2] and aging simulation [3,4]; only a very limited number of studies tackle AIFR [5]. A typical AIFR approach is to use face modeling to synthesize and render face images to the same age as the gallery image prior to recognition [2,4]. However, due to strong parametric assumptions and the complexity of the algorithm, these methods are computationally expensive and the results are often unstable for real-world face recognition.

In this paper, we propose a new two-step AIFR approach. First, a new feature descriptor called the maximum entropy feature descriptor (MEFD) is used to extract expressive and informative features. Unlike existing feature descriptors, MEFD can maximize the expressive power in terms of maximum entropy, which is highly beneficial to classification. Second, a factor analysis-based matching framework termed identity factor analysis (IFA) is developed to further improve recognition performance.

Specifically, the proposed encoder is trained using a set of training face images such that the frequency of output codes distributes as evenly as possible; this maximizes the discriminative ability in terms of maximum entropy. As illustrated in Figure 1, the pattern space is quantized using a decision tree. For instance, suppose we have a training dataset $X = \{x_i | x_i \in R^{d\times 1}, i = 1, \dots, N\}$; the goal of the decision tree is to build a partition-based model that assigns each point x_i with a code $y_i \in \{1,2,\dots,K\}$, where the probability mass function over the set of codes is as close to a uniform distribution as possible. We grow the decision tree in a greedy manner such that, at each split step, it extends the *best* node to maximize the entropy of the code distribution.

$$E^{(K+1)} = -\left(\sum_{k=1}^{i-1} p(k)\log p(k) + \sum_{k=i+1}^{K} p(k)\log p(k) + p1\log p1 + p2\log p2\right), \quad (1)$$

where *p1* and *p2* represent the probabilities of a pattern falling into these new partitions. The $E^{(K+1)}$ can be rewritten as:

$$E^{(K+1)} = E^{(k)} + p(i)\log(i) - \left(p1\log p1 + p2\log p2\right). \quad (2)$$

In order to optimize (2), we maximize the information gain:

$$G(i) = p(i)\log(i) - \left(p1\log p1 + p2\log p2\right). \quad (3)$$

For face matching, we develop an effective matching framework called identity factor analysis (IFA) for feature classification. It has been shown that the observable face features $\vec{t}$ (in the presence of aging variations) can be decomposed into the following four terms: the mean component ($\vec{\beta}$), the identity-related component ($U\vec{x}$), the age-related component ($V\vec{y}$), and the noise component ($\vec{\varepsilon}$). This decomposition model can be formulated as:

$$\vec{t} = \vec{\beta} + U\vec{x} + V\vec{y} + \vec{\varepsilon}. \quad (4)$$

The model parameters $\{\vec{\beta}, U, V, \vec{\varepsilon}\}$ can be estimated using the EM algorithm.

Denote $\vec{t}_g$ the gallery sample and $\vec{t}_p$ the probe sample. If $\vec{t}_g$ and $\vec{t}_p$ are from the same identity, according to (4) we have

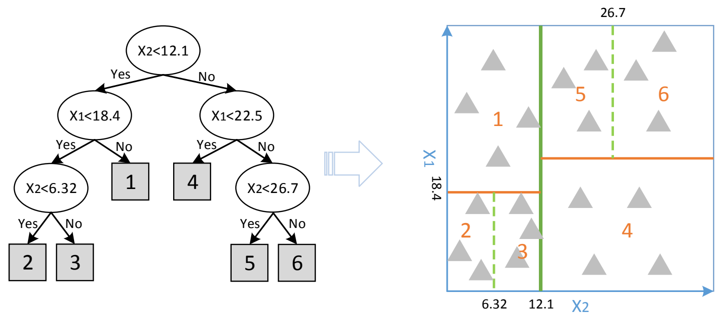

Figure 1. The decision tree based encoding scheme. The internal nodes are decision rules, while the leaf nodes represent code assignment. The decision tree is trained in a greedy manner such that each leaf node hosts a similar number of training samples, as illustrated on the right. X1 and X2 represent two different attributes.

$$\begin{bmatrix} \vec{t}_g \\ \vec{t}_p \end{bmatrix} = \begin{bmatrix} \vec{\beta} \\ \vec{\beta} \end{bmatrix} + \begin{bmatrix} U & V & 0 \\ U & 0 & V \end{bmatrix} \begin{bmatrix} \vec{x}_{g,p} \\ \vec{y}_g \\ \vec{y}_p \end{bmatrix} + \begin{bmatrix} \vec{\varepsilon}_g \\ \vec{\varepsilon}_p \end{bmatrix}.$$

However, if $\vec{t}_g$ and $\vec{t}_p$ are from different identities, then we have

$$\begin{bmatrix} \vec{t}_g \\ \vec{t}_p \end{bmatrix} = \begin{bmatrix} \vec{\beta} \\ \vec{\beta} \end{bmatrix} + \begin{bmatrix} U & 0 & V & 0 \\ 0 & U & 0 & V \end{bmatrix} \begin{bmatrix} \vec{x}_g \\ \vec{x}_p \\ \vec{y}_g \\ \vec{y}_p \end{bmatrix} + \begin{bmatrix} \vec{\varepsilon}_g \\ \vec{\varepsilon}_p \end{bmatrix}.$$

The matching score is then calculated as the ratio between the assumption that $\vec{t}_g$ and $\vec{t}_p$ from the same or different identity, as given by

$$LR(\vec{t}_g, \vec{t}_p) = \log \frac{N\left(\begin{bmatrix} \vec{t}_g \\ \vec{t}_p \end{bmatrix} \middle| \begin{bmatrix} \vec{\beta} \\ \vec{\beta} \end{bmatrix}, \begin{bmatrix} \Sigma_{tot} & \Sigma_{ac} \\ \Sigma_{ac} & \Sigma_{tot} \end{bmatrix}\right)}{N\left(\begin{bmatrix} \vec{t}_g \\ \vec{t}_p \end{bmatrix} \middle| \begin{bmatrix} \vec{\beta} \\ \vec{\beta} \end{bmatrix}, \begin{bmatrix} \Sigma_{tot} & 0 \\ 0 & \Sigma_{tot} \end{bmatrix}\right)}$$
$$= const + 0.5\vec{t}_g^{\,T} Q\vec{t}_g + 0.5\vec{t}_p^{\,T} Q\vec{t}_p + \vec{t}_g^{\,T} P\vec{t}_p$$

[1] Yun Fu and Thomas S. Huang, “Human age estimation with regression on discriminative aging manifold,” IEEE Transactions on Multimedia, vol. 10, no. 4, pp. 578–584, 2008.

[2] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles, “Automatic age estimation based on facial aging patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 12, pp. 2234–2240, 2007.

[3] Andreas Lanitis, Christopher J. Taylor, and Timothy F. Cootes, “Toward automatic simulation of aging effects on face images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 4, pp. 442–455, 2002.

[4] Jin-Li Suo, Song Chun Zhu, Shiguang Shan, and Xilin Chen, “A compositional and dynamic model for face aging,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 3, pp. 385–401, 2010.

[5] Haibin Ling, Stefano Soatto, Narayanan Ramanathan, and David W. Jacobs, “Face verification across age progression using discriminative methods,” IEEE Transactions on Information Forensics and Security, vol. 5, no. 1, pp. 82–91, 2010.