Hierarchical-PEP Model for Real-world Face Recognition

Haoxiang Li, Gang Hua
Stevens Institute of Technology

Pose variation remains one of the major factors adversely affect the accuracy of real-world face recognition systems. The same face in different poses can look drastically different to each other. Belhumeur *et al.* [1] empirically demonstrate that frontal faces can be projected to a low-dimensional subspace invariant to variation in illumination and facial expressions. This observation highlights the importance of addressing pose variation because it can greatly help relieve the adverse effects of the other visual variations.

A set of methods build pose-invariant face representations by locating the facial landmarks. For example, Chen *et al.* [2] concatenate dense features around the facial landmarks to build the face representation. The pose-invariance is achieved in this way, because it always extracts features from the face part surrounded around the facial landmarks regardless of their locations in the image. The elastic matching methods [5] generalize this design with a probabilistic elastic part (PEP) model unsupervisedly learned from face image patches.

While this procedure – locating the face parts and stacking the face part features to build face representation – is empirically demonstrated to be effective by both Chen *et al.* [2] and Li *et al.* [5], we argue that directly describing the face parts with naive dense extraction of low-level features may not be optimal.

In this work, we propose to build a better face part model to construct an improved face representation. Inspired by the probabilistic elastic part (PEP) model and the success of the deep hierarchical architecture in a number of visual tasks, we propose the Hierarchical-PEP model to approach the unconstrained face recognition problem.

As shown in Figure 1, we apply the PEP model hierarchically to decompose a face image into face parts at different levels of details to build pose-invariant part-based face representations. Following the hierarchy from bottom-up, we stack the face part representations at each layer, discriminatively reduce its dimensionality, and hence aggregate the face part representations layer-by-layer to build a compact and invariant face representation. The Hierarchical-PEP model exploits the fine-grained structures of the face parts at different levels of details to address the pose variations. It is also guided by supervised information in constructing the face part/face representations.

We empirically verify the Hierarchical-PEP model on two public benchmarks and a face recognition challenge for image-based and video-based face verification. The state-of-the-art performance demonstrates the potential of our method. We show the performance comparison on the YouTube faces dataset [9] in Table 1.

Table 1: Performance comparison on YouTube Faces dataset under the restricted with no outside data protocol.

Algorithm	Accuracy $\pm$ Error(%)
MBGS [9]	76.4 ± 1.8
MBGS+SVM- [8]	78.9 ± 1.9
STFRD+PMML [10]	79.5 ± 2.5
VF^2 [7]	84.7 ± 1.4
DDML (combined) [3]	82.3 ± 1.5
Eigen-PEP [6]	84.8 ± 1.4
LM3L [4]	81.3 ± 1.2
Hierarchical-PEP (layers fusion)	$\mathbf{87.00 \pm 1.50}$

[1] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *PAMI*, 1997.

[2] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun. Blessing of dimensionality: High dimensional feature and its efficient compression for face verification. In *CVPR*, 2013.

[3] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Discriminative deep metric learning for face verification in the wild. In *CVPR*, 2014.

[4] Junlin Hu, Jiwen Lu, Junsong Yuan, and Yap-Peng Tan. Large margin multi-metric learning for face and kinship verification in the wild. In *ACCV*, 2014.

[5] Haoxiang Li, Gang Hua, Zhe Lin, Jonathan Brandt, and Jianchao Yang. Probabilistic elastic matching for pose variant face verification. In *CVPR*, 2013.

[6] Haoxiang Li, Gang Hua, Xiaohui Shen, Zhe Lin, and Jonathan Brandt. Eigen-pep for video face recognition. In *ACCV*, 2014.

[7] O. M. Parkhi, K. Simonyan, A. Vedaldi, and A. Zisserman. A compact and discriminative face track descriptor. In *CVPR*, 2014.

[8] Lior Wolf and Noga Levy. The svm-minus similarity score for video face recognition. In *CVPR*, 2013.

[9] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR*, 2011.

[10] Cui Zhen, Wen Li, Dong Xu, Shiguang Shan, and Xilin Chen. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In *CVPR*, 2013.

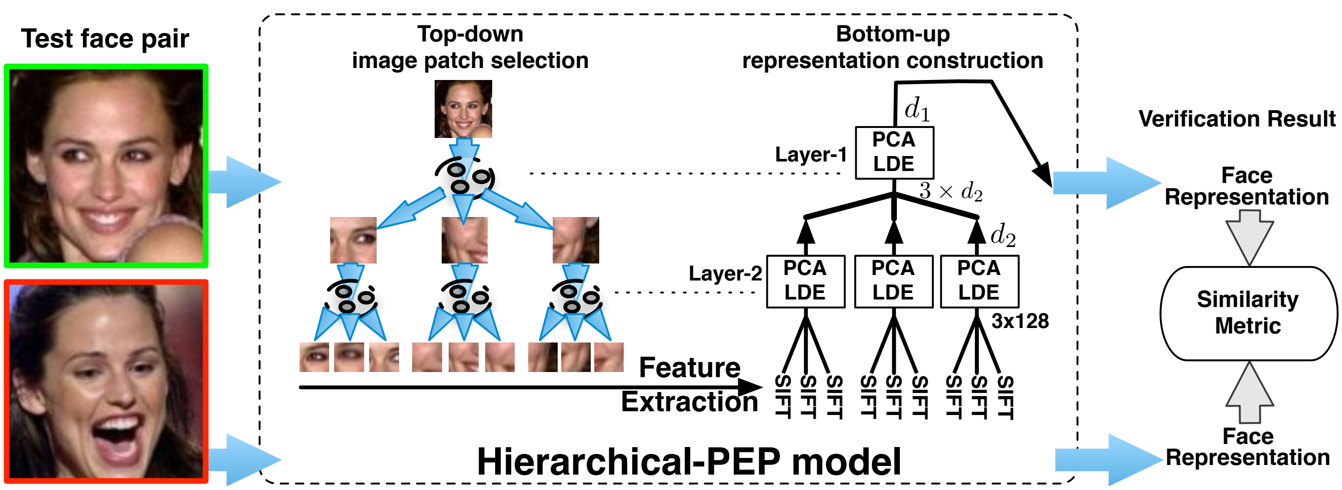

Figure 1: Construction of the face representation with an example 2-layer Hierarchical-PEP model: PCA at layer t keeps d_t dimensions.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.