Beyond Principal Components: Deep Boltzmann Machines for Face Modeling

Chi Nhan Duong[1], Khoa Luu[2], Kha Gia Quach[1], Tien D. Bui[1]
[1]Concordia University, Computer Science and Software Engineering, Montréal, Québec, Canada.
[2]Carnegie Mellon University, CyLab Biometrics Center, Pittsburgh, PA, USA.

The "interpretation through synthesis", i.e. Active Appearance Models (AAMs) method [1], has become one of the most successful and popular face modeling approaches over the last two decades. Given a new face image, the purpose is to "describe" that image by generating a new synthesized image that is similar to it as much as possible. This aim can be achieved by an optimization process on the appearance parameters of the model based *apriori* on constrained solutions. Therefore, a subspace model to be suitable and practical must also provide a basis for a broad range of variations that are usually unseen. However, it is quite challenging due to appearance variations of human face images, e.g. facial poses, occlusions, lighting, low resolution, etc. Since these variations are mostly non-linear, it is impossible to represent them in a linear model, such as Principal Component Analysis.

This paper presents a novel Deep Appearance Models (DAMs) approach, an efficient replacement for AAMs, to accurately capture both shape and texture of face images under large variations. In this approach, three crucial components represented in hierarchical layers are modeled using the Deep Boltzmann Machines (DBM) [2] to robustly capture the variations of facial shapes and appearances. DAMs are therefore superior to AAMs in inferring a representation for new face images under various challenging conditions. In addition, DAMs have ability to generate a compact set of parameters in higher level representation that can be used for classification, e.g. face recognition and facial age estimation.

There are three main steps to construct the DAMs as presented in Figure 2. Firstly, facial shape structures and texture variations in DAMs are mathematically modeled using two separated DBMs. Thanks to the nonlinear structure of DBM and the strength of latent variables organized in hidden layers, both shape and texture variations can be efficiently captured. Then the interactions between them are further modeled using a deeper hidden layer. By this way, these interactions can be naturally interpreted and used as a compact set of parameters for further discriminative problems.

Finally, a fitting algorithm is presented in order to synthesize any given new face image as follows. Given a testing face I, the fitting process in DAMs can be formulated as finding an optimal shape $\mathbf{s}$ that maximizes the probability of the shape-free image as in Eqn. (1).

$$\mathbf{s}^* = \arg\max_{\mathbf{s}} P(I(W(r_{\mathcal{D}}, \mathbf{s}))|\mathbf{s};\theta) \tag{1}$$

where $W(r_{\mathcal{D}};\mathbf{s})$ stands for the warping operator and θ denotes the model parameters. The probability of texture $\mathbf{g}$ given hidden units $\mathbf{h}_g^{(1)}$ is computed as

$$P(\mathbf{g}|\mathbf{h}_g^{(1)};\mathbf{s},\theta) = \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{A}) \tag{2}$$

where $\mathbf{A}$ is the identity matrix and σ is the standard-deviation of visible units in the texture model; $\mathbf{W}_\mathbf{g}^{(\mathbf{1})}$ are learned weights of the visible-hidden texture; $\mathbf{c}$ is a bias of this hidden layer $\mathbf{h}_g^{(1)}$; and $\mathbf{m} = \sigma\mathbf{W}_\mathbf{g}^{(\mathbf{1})}\mathbf{h}_g^{(1)} + \mathbf{c}$.

Then the maximum likelihood can be then estimated:

$$\mathbf{s}^* = \arg\min_{\mathbf{s}} \frac{1}{\sigma^2}\sum(I(W(r_{\mathcal{D}}, \mathbf{s})) - \mathbf{m})^2 \tag{3}$$

The forward compositional algorithm can be used to solve the problem (3) by finding the updating parameter $\Delta\mathbf{s}$ that increases the likelihood.

With the deep structured models for shapes and textures, DAMs consist of several attractive properties: (1) providing the capability of generating facial shapes using texture information and vice versa; (2) interpreting both shapes and textures naturally using higher hidden layer; (3) providing a robust feature extraction process even when one of two inputs is missing. (4) being able to deal with facial reconstruction in various challenging conditions, such as: facial occlusions, facial expressions, facial off-angles, etc.

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

Figure 1: A comparison of facial interpretation in real world images between our DAMs approach and the AAMs. The first row: original images; The second row: shape free images; The third row: facial interpretation using PCA-based AAMs; The fourth row: facial interpretation using our proposed DAMs approach.

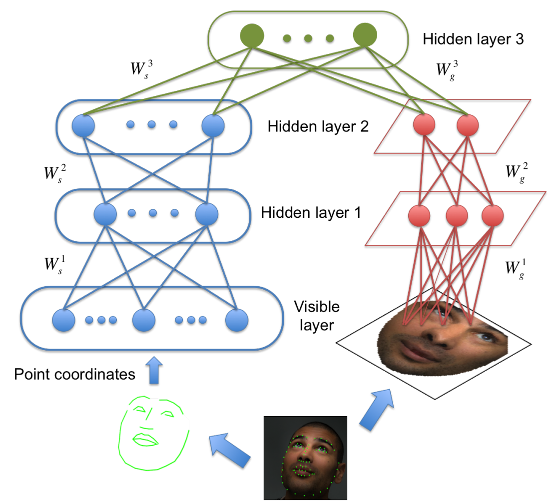

Figure 2: Deep Appearance Models that consists of shape model (left), texture model (right) and the joint representation of shape and texture.

The proposed approach is evaluated in facial image reconstruction, facial super-resolution on two databases, i.e. Labeled Face Parts in the Wild (LFPW) and Helen. It is also evaluated on FG-NET database for the problem of age estimation. DAMs achieve remarkable improvements in both tasks compared to PCA-based AAMs model. Moreover, experimental results in several applications such as facial super-resolution, face off- angle reconstruction, occlusion removal and facial age estimation have shown the potential of the model in dealing with large variations.

[1] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Interpretting Face Images using Active Appearance Models. In *Proc. of the 3rd Intl. Conf. on Automatic Face and Gesture Recognition*, pages 300–305, 1998.

[2] Ruslan Salakhutdinov and Geoffrey E Hinton. Deep boltzmann machines. In *Intl. Conf. on Artificial Intell. and Statistics*, pages 448–455, 2009.