This CVPR2015 extended abstract is the Open Access version, provided by the Computer Vision Foundation.

Transport-Based Single Frame Super Resolution of Very Low Resolution Face Images

Soheil Kolouri[1], Gustavo K. Rohde[1,2]
[1]Department of Biomedical Engineering, Carnegie Mellon University. [2]Department of Electrical and Computer Engineering, Carnegie Mellon University.

We describe a single-frame super-resolution method for reconstructing high-resolution (abbr. *high-res*) faces from very low-resolution (abbr. *low-res*) face images (e.g. smaller than 16×16 pixels) by learning a nonlinear Lagrangian model for the high-res face images. Our technique is based on the mathematics of optimal transport, and hence we denote it as transport-based SFSR (TB-SFSR). In the training phase, a nonlinear model of high-res facial images is constructed based on transport maps that morph a reference image into the training face images. In the testing phase, the resolution of a degraded image is enhanced by finding the model parameters that best fit the given low resolution data.

Generally speaking, most SFSR methods [2, 3, 4, 5] are based on a linear model for the high-res images. Hence, ultimately, the majority of SFSR models in the literature can be written as, $I_h(\mathbf{x}) = \sum_i w_i \psi_i(\mathbf{x})$, where I_h is a high-res image or a high-res image patch, w's are weight coefficients, and ψ's are high-res images (or image patches), which are learned from the training images using a specific model. Here we propose a fundamentally different approach toward modeling high-res images. In our approach the high-res image is modeled as a mass preserving mapping of a high-res template image, I_0, as follows

$$I_h(\mathbf{x}) = det(\mathbf{I} + \sum_i \alpha_i D\mathbf{v}_i(\mathbf{x})) I_0(\mathbf{x} + \sum_i \alpha_i \mathbf{v}_i(\mathbf{x})), \tag{1}$$

where $\mathbf{I}$ is the identity matrix, α_i is the weight coefficient of displacement field $\mathbf{v}_i$ (i.e. a smooth vector field), and $D\mathbf{v}_i(\mathbf{x})$ is the Jacobian matrix of the displacement field $\mathbf{v}_i$, evaluated at $\mathbf{x}$. The proposed method can be viewed as a linear modeling in the space of mass-preserving mappings, which corresponds to a non-linear model in the image space. Thus (through the use of the optimal mapping function $\mathbf{f}(\mathbf{x}) = \mathbf{x} + \sum_i \alpha_i \mathbf{v}_i(\mathbf{x})$) our modeling approach can also displace pixels, in addition to changing their intensities.

Given a training set of high-res face images, $I_1, ..., I_N : \Omega \to \mathbb{R}$ with $\Omega = [0,1]^2$ the image intensities are first normalized to integrate to 1. This is done so the images can be treated as distributions of a fixed amount of intensity values (i.e. fixed amount of mass). Next, the reference face is defined to be the average image, $I_0 = \frac{1}{N}\sum_{i=1}^N I_i$, and the optimal transport distance between the reference image and the i'th training image, I_i, is defined to be,

$$\begin{aligned} d_{OT}(I_0, I_i) &= \min_{\mathbf{u}_i} \int_\Omega |\mathbf{u}_i(\mathbf{x})|^2 I_i(\mathbf{x}) d\mathbf{x} \\ & \quad s.t.\ \ det(\mathbf{I} + D\mathbf{u}_i(\mathbf{x})) I_0(\mathbf{x} + \mathbf{u}_i(\mathbf{x})) = I_i(\mathbf{x}) \end{aligned} \tag{2}$$

where $(\mathbf{f}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})) : \Omega \to \Omega$ is a mass preserving transform from I_i to I_0, $\mathbf{u}$ is the optimal displacement field, and $D\mathbf{u}_i$ is the Jacobian matrix of $\mathbf{u}$. The optimization problem above is well posed and has a unique minimizer [1]. Having optimal displacement fields $\mathbf{u}_i$ for $i = 1, \ldots, N$ a subspace, V, is learned for these displacement fields. Let $\mathbf{v}_j$ for $j = 1, ..., M$ be a basis for subspace V. Then, any combination of the basis displacement fields can be used to construct an arbitrary deformation field, $\mathbf{f}_\alpha(\mathbf{x}) = \mathbf{x} + \sum_{j=1}^M \alpha_j \mathbf{v}_j(\mathbf{x})$, which can then be used to construct a given image $I_\alpha(\mathbf{x}) = det(D\mathbf{f}_\alpha(\mathbf{x})) I_0(\mathbf{f}_\alpha(\mathbf{x}))$. Hence, subspace V provides a generative model for the high-res face image. In the testing phase, we constrain the space of possible high-res solutions to those, which are representable as I_α for some $\alpha \in \mathbb{R}^M$. Hence, for a degraded input image, I_l, and assuming that $\phi(.)$ is known and following the MAP criteria we can write,

$$\begin{aligned} \alpha^* &= \text{argmin}_\alpha \frac{1}{2} \|I_l - \phi(I_\alpha)\|_2^2 \\ & \quad s.t\ \ I_\alpha(\mathbf{x}) = det(D\mathbf{f}_\alpha(\mathbf{x})) I_0(\mathbf{f}_\alpha(\mathbf{x})) \end{aligned} \tag{3}$$

where a gradient descent approach is used to obtain a local optima α^*. Note that, images of faces (and other deformable objects) differ from each other not only due to differences in appearance (i.e. tone and texture) of their parts, but also due to the different locations of these parts for different individuals. Hence, trying to model the displacement of parts by only taking the co-variance structure of intensities on a fixed grid would lead to high variances at each pixel. Therefore, the nonlinear model we use is more effective in capturing the real variations in appearance of the data.

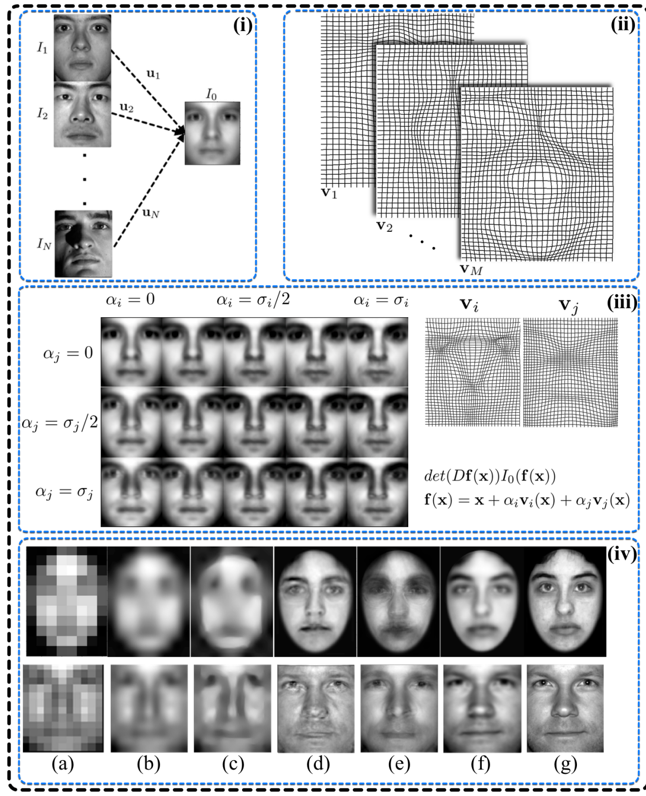

Figure 1: Training images are morphed to a reference image and the optimal displacement fields are calculated for every image (i). The principal components of optimal displacement fields are calculated, $\mathbf{v}_i$ (ii). Demonstration of face modeling using only two of the displacement fields (iii), where, σ_i and σ_j are the standard deviation of the projected training displacement fields onto $\mathbf{v}_i$ and $\mathbf{v}_j$, respectively. Sample results are shown in (iv), where (a) and (g) are the low-res and high-res images (b) is the cubic spline interpolated image, (c), (d), and (e) are obtained using the methods in [2, 4, 5], respectively, and (f) is obtained using our proposed method.

[1] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[2] Ayan Chakrabarti, AN Rajagopalan, and Rama Chellappa. Super-resolution of face images using kernel pca-based prior. *IEEE Transactions on Multimedia*, 9(4):888–892, 2007.

[3] Xiaogang Wang and Xiaoou Tang. Hallucinating face by eigentransformation. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 35(3):425–434, 2005.

[4] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010.

[5] Wilman WW Zou and Pong C Yuen. Very low resolution face recognition problem. *Image Processing, IEEE Transactions on*, 21(1):327–340, 2012.

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