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

Soheil Kolouri
Carnegie Mellon University
skolouri@andrew.cmu.edu

Gustavo K. Rohde
Carnegie Mellon University
gustavor@cmu.edu

Abstract

Extracting high-resolution information from highly degraded facial images is an important problem with several applications in science and technology. Here we describe a single frame super resolution technique that uses a transport-based formulation of the problem. The method consists of a training and a testing phase. In the training phase, a nonlinear Lagrangian model of high-resolution facial appearance is constructed fully automatically. 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. We test the approach on two face datasets, namely the extended Yale Face Database B and the AR face datasets, and compare it to state of the art methods. The proposed method outperforms existing solutions in problems related to enhancing images of very low resolution.

1. Introduction

Super-resolution (SR) is the process of reconstructing a high-resolution (abbr. *high-res*) image from one or several corresponding low-resolution (abbr. *low-res*) images. SR techniques have been used in a wide variety of applications from satellite and aerial imaging to intelligence surveillance, medical image processing, and finger print enhancement. In particular, the use of SR techniques to infer high-res face images from low-res ones has recently attracted a large amount of interest in the image processing and computer vision communities [2, 15, 27, 28].

Based on the number of low-res images used to reconstruct the corresponding high-res image, SR techniques can be broadly categorized into two major classes [20], namely "multi-frame" SR (MFSR) [4, 10, 9, 19] and "single-frame" SR (SFSR) [2, 11, 16, 27]. Due to the inherent undersampling, most SR problems are inherently ill-posed. Meaning that for a specific low-res image there exist multiple corresponding high-res images. Generally speaking, SR techniques overcome this problem by constraining the space of high-res solutions using either information from multiple low-res images or prior information regarding the class of high-res images. The idea is to restrict the space of solutions to automatically discard irrelevant solutions of the problem.

We begin by noting that the class of face images is a relatively small subset of the entire set of high-res images. This is because of the unique structure of the human face (i.e. eyes, nose, mouth, etc). In SFSR techniques, this prior knowledge is first learned from a set of high-res training images and then it is used to reconstruct a high-res image from a low-res test image. Baker and Kanade [2], for instance, proposed to learn a prior on the spatial distribution of the image gradient for frontal images of faces. Chakrabarti *et al*. [7] proposed to learn a kernel principal component analysis-based prior model for high-res images. More recently, Zou *et al*. [28] proposed a method based on learning the best linear mapping that maps low-res images to their corresponding high-res images.

In this paper we focus on SFSR techniques and describe a method for reconstructing high-res faces from very low-res face images (e.g. 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). The idea is to use the Monge formulation of the optimal transport problem [1, 13], and with it construct a nonlinear model for both the pixel intensities and their locations for facial images. Our model is nonlinear, and Lagrangian (using PDE parlance) in the sense that intensities are not compared using a fixed grid but can also be displaced and transported to other image regions. In short, TB-SFSR first finds diffeomorphisms, in the sense of 'optimal transport', from a reference face to the training faces and then learns a linear subspace that best describes these diffeomorphisms. Next, it constrains the space of high-res images to those that can be obtained by morphing the reference face using an arbitrary diffeomorphism from the learned subspace.

We show that TB-SFSR can be used to recover information from very low-res face images. We test our proposed

method on the extended Yale Face Database B [12, 14] and the AR face dataset [17, 18] and compare our results to those of the methods presented in [7, 27], and [28].

The remainder of this paper is organized as follows. Section 2 describes a few of the main ideas developed in other SFSR works and lays the foundation for our work. In Section 3, we describe our formulation in detail and discuss the idea of Lagrangian modeling using optimal transport. Various experimental results in Section 4 are used to demonstrate the efficacy of the proposed Lagrangian modeling. Finally, in Section 5 we conclude with a discussion and point out future directions.

2. Overview of prior work

Here we describe, in a general sense, some of the main ideas previously used in SFSR problems. Due to space limitations, our goal is to focus on a broad description of the mathematical modeling ideas, citing specific examples, rather than providing an exhaustive review of previously described methods. In SFSR, given a low-res image I_l the goal is to reconstruct the corresponding high-res image I_h. The observed low-res image I_l is a degraded version of I_h. Let $\phi(.)$ be the degradation function in its most general form such that

$$I_l = \phi(I_h). \tag{1}$$

An optimal I_h can be found by maximizing the posterior probability $p(I_h|I_l)$, based on the maximum a posteriori (MAP) criteria,

$$\begin{aligned} I_h^* &= \text{argmax}_{I_h}\, Pr(I_h|I_l) \\ &= \text{argmax}_{I_h}\, ln(Pr(I_l|I_h)) + ln(Pr(I_h)) \end{aligned} \tag{2}$$

where the first term of the above objective function is the log likelihood and the second term is the a priori information on the image, which can be interpreted to represent information about the given class of images (i.e. an image model).

Most commonly, $p(I_l|I_h)$ is modeled by a Gaussian distribution and hence the log likelihood in (2) is written as $ln(Pr(I_l|I_h)) = -\|I_l - \phi(I_h)\|^2$. As for the degradation function ϕ, it is commonly modeled using a low pass blurring filter together with a downsampling operation [20]. The choices for the *a priori* model, on the other hand, are vast throughout the literature. Early SFSR techniques used the assumption that I_h should be smooth, and hence the modeling should enforce the reconstructed image to be piecewise/locally smooth. Markov random fields (MRF) are considered as a useful prior image model [21] to enforce such smoothness. This is equivalent to regularizing the log likelihood by an energy function, $U(I_h)$, derived from the MRF model [21],

$$I_h = \text{argmin}_{I_h}\, \frac{1}{2}\|I_l - \phi(I_h)\|_2^2 + \lambda U(I_h) \tag{3}$$

where λ is the regularization parameter. Local smoothness constraints are ubiquitously used in image reconstruction problems [20]. In SFSR problems, however, they can fail to reconstruct high frequency detail and may produce answers which are overly smooth and suffer from staircasing artifacts (i.e. in TV regularization) [10, 20].

More recent approaches involve constructing a linear subspace model for high-res images, and solve the problem of SFSR by constraining the reconstructed image to be the best approximation to the data this model can produce [6, 24]. Hence, in these methods the log likelihood term is regularized by the projection error of I_h onto the learned subspace, L,

$$I_h = \text{argmin}_{I_h}\, \frac{1}{2}\|I_l - \phi(I_h)\|_2^2 + \lambda\|I_h - P_L(I_h)\|_2^2 \tag{4}$$

where $P_L(I_h)$ is the projection of image I_h onto subspace L. Using similar ideas, Yang *et al.* [27] proposed a model which assumes that the high-res image patches, can be represented as a linear combination of few basis images.

We note that, generally speaking, most SFSR methods previously described are based on a linear model for the high-res images. Meaning that, ultimately, the majority of SFSR models in the literature can be written as

$$I_h(\mathbf{x}) = \sum_i w_i \psi_i(\mathbf{x}), \tag{5}$$

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. For instance, in [24], ψ's are the eigenvectors of the high-res training images, obtained from applying PCA to these images. Chakrabarti *et al.* [7] used KPCA and obtained ψ's to be the eigenvectors of the training images in a determined kernel space. In [27], ψ's are high-res image patches that form the atoms of an RIP matrix learned from the training images. Finally, in [28], ψ's can be thought of as the columns of a linear mapping, which maps low-res images to high-res ones.

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

$$I_h(\mathbf{x}) = det(\mathbf{I} + \sum_i \alpha_i D\mathbf{v}_i(\mathbf{x}))I_0(\mathbf{x} + \sum_i \alpha_i \mathbf{v}_i(\mathbf{x})) \tag{6}$$

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 mapping function

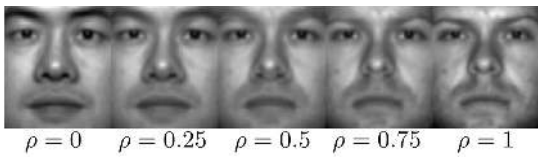

Figure 1. Visualization of the morphing process for two face images by changing ρ from zero to one, each image is calculated from $det(D\mathbf{f}_\rho(\mathbf{x}))I(\mathbf{f}_\rho(\mathbf{x}))$.

$\mathbf{x} + \sum_i \alpha_i \mathbf{v}_i(\mathbf{x})$) our modeling approach can also displace pixels, in addition to change their intensities. In short, rather than learning the linear combination of intensity values (as most SFSR methods do) we seek to learn the mass preserving mappings that can be used to model the high-res training images. In what follows, we formalize the proposed method and show that such nonlinear modeling of images enhances the information recovery process.

3. Transport-based SFSR

TB-SFSR utilizes the mathematics of optimal transport (OT) in combination with subspace learning techniques to learn a nonlinear model for the high-res images in the training set. The OT problem was initially raised in 1781 by G. Monge, as the problem of transporting a given distribution of matter (e.g. pile of sand) into another. The Monge problem is posed as how to minimize the work needed for such transportation. More recently, OT has been used in the image processing and computer vision communities for image registration, image modeling, feature matching, etc [13, 3, 23]. Here we use OT to model the variations in the space of high-res images.

3.1. Training phase

We begin by clarifying that our description of the method is given in continuous domain. The discretization of the model is straightforward and is described subsequently. 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,

$$d_{OT}(I_0, I_i) = \min_{\mathbf{f}} \int_\Omega |\mathbf{f}(\mathbf{x}) - \mathbf{x}|^2 I_i(\mathbf{x}) d\mathbf{x} \quad s.t. \;\; det(D\mathbf{f}(\mathbf{x}))I_0(\mathbf{f}(\mathbf{x})) = I_i(\mathbf{x}) \quad (7)$$

where $\mathbf{f} : \Omega \to \Omega$ is a mass preserving transform from I_i to I_0, and $D\mathbf{f}$ is the Jacobian matrix of $\mathbf{f}$. The work from Brenier et al [5] shows that the optimization problem above is well posed and a unique solution exists. This

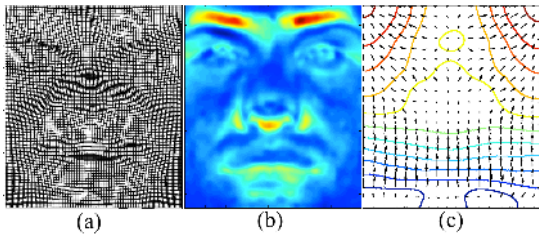

Figure 2. Visualization of the change $\mathbf{f}_i$ applies to the underlying grid of the image (a), the determinant of the Jacobian of $\mathbf{f}_i$ (red corresponds to > 1 and blue corresponds to < 1 values) (b), and the corresponding displacement field $\mathbf{u}_i$ (the contours show the equipotential lines corresponding to this vector field)(c) , for the images depicted in Figure 1.

unique transport function morphs image I_0 to image I_i by $det(D\mathbf{f}_i(\mathbf{x}))I_0(\mathbf{f}_i(\mathbf{x})) = I_i(\mathbf{x})$. Note that, $\mathbf{f}_i$ changes the underlying grid and the intensity values of image I_0 simultaneously, hence it is truly a 'morphing'. In addition, $\mathbf{f}_i$ provides a geodesic on the OT manifold [23] and points on this geodesic can be parametrized by $\rho \in [0, 1]$ as,

$$\mathbf{f}_\rho(\mathbf{x}) = (1 - \rho)\mathbf{x} + \rho\mathbf{f}_i(\mathbf{x}), \quad (8)$$

and the morphing process can be visualized using $det(D\mathbf{f}_\rho(\mathbf{x}))I_0(\mathbf{f}_\rho(\mathbf{x}))$ by changing ρ from zero to one. Figure 1 shows the morphing process between two face images as a function of ρ.

The optimal transport function $\mathbf{f}_i$ is further decomposed into the identity and the optimal displacement function, $\mathbf{u}_i(\mathbf{x}) : \Omega \to \Omega$,

$$\mathbf{f}_i(\mathbf{x}) = \mathbf{x} + \mathbf{u}_i(\mathbf{x}). \quad (9)$$

Note that the optimal displacement function $\mathbf{u}_i$ quantifies the amount of deviation from the identity. To further clarify the concept of the optimal deformation and optimal displacement functions, Figure 2 depicts the change $\mathbf{f}_i$ exerts on the grid of the image, the determinant of the Jacobian, and the corresponding optimal displacement function for the two images depicted in Figure 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, each optimal displacement field, $\mathbf{u}_i$ can be represented as a linear combination of basis displacement fields $\mathbf{v}_j$s via $\mathbf{u}_i = \sum_{j=1}^{M} \alpha_j^i \mathbf{v}_j$. Here, an arbitrary combination of the basis displacement fields can be used to construct an arbitrary deformation field,

$$\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x}) = \mathbf{x} + \sum_{j=1}^{M} \alpha_j \mathbf{v}_j(\mathbf{x}) \quad (10)$$

which can then be used to construct a given image $I_{\boldsymbol{\alpha}}(\mathbf{x}) = det(D\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x}))I_0(\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x}))$. Hence, subspace V provides a gen-

erative model for the high-res face image. In the testing phase, we constrain the space of high-res images to those that are generated by the learned model as $I_{\boldsymbol{\alpha}}(\mathbf{x}) = det(D\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x}))I_0(\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x}))$. As discussed below, numerous techniques for linear (and nonlinear) subspace modeling exist [8, 26]. In the results reported below, we utilized the usual principal component analysis (PCA) technique for this task. That is, in this implementation of the modeling approach, $\mathbf{v}_j$ are the (top) eigenvectors of the covariance matrix given by $S_{i,j} = \int_\Omega (\mathbf{u}_i - \bar{\mathbf{u}})^T(\mathbf{u}_i - \bar{\mathbf{u}})d\mathbf{x}$, where $\bar{\mathbf{u}}$ corresponds to the mean displacement field extracted from the training set. Let $\mathbf{e}_i$ and γ_i correspond to the eigenvectors and eigenvalues of S. The modeling displacement maps are then given by:

$$\mathbf{v}_i = \frac{1}{\sqrt{\gamma_i}} \sum_{k=1}^{N} \mathbf{e}_i[k]\mathbf{u}_k. \tag{11}$$

In our implementation, only the top M eigenvectors corresponding to 99% of the variations in the dataset are extracted during the training procedure.

3.2. Testing phase

Having the displacement space V, we constrain the space of possible high-res solutions to those, which are representable as $I_{\boldsymbol{\alpha}}$ for some $\boldsymbol{\alpha} \in \mathbb{R}^M$. Hence, for a degraded input image, I_l, and assuming that $\phi(.)$ is known and following the MAP criteria we can write,

$$\begin{aligned} I_h^* \quad = \quad & \text{argmin}_{I_h,\boldsymbol{\alpha}} \frac{1}{2}\|I_l - \phi(I_h)\|_2^2 + \\ & \lambda\|I_h - det(D\mathbf{f}_{\boldsymbol{\alpha}})I_0(\mathbf{f}_{\boldsymbol{\alpha}})\|_2^2 \end{aligned} \tag{12}$$

where λ is the regularizer, and $\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x})$ is defined in Eq (10). Letting λ to go to infinity (hard thresholding), the optimization problem above can be written as,

$$\begin{aligned} \boldsymbol{\alpha}^* \quad = \quad & \text{argmin}_{\boldsymbol{\alpha}} \frac{1}{2}\|I_l - \phi(I_{\boldsymbol{\alpha}})\|_2^2 \\ & s.t \;\; I_{\boldsymbol{\alpha}}(\mathbf{x}) = det(D\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x}))I_0(\mathbf{f}_{\boldsymbol{\alpha}}(\mathbf{x})) \end{aligned} \tag{13}$$

Solving (13) with a gradient descent approach leads to a local optima $\boldsymbol{\alpha}^*$. Let $\alpha_i^{(k)}$ denote α_i at k'th iteration of the gradient descent and $I_\alpha^{(k)}(\mathbf{x}) = det(D\mathbf{f}_\alpha^{(k)}(\mathbf{x}))I_0(\mathbf{f}_\alpha^{(k)}(\mathbf{x}))$, then the gradient descent update for α_i can be written as follows,

$$\begin{aligned} \alpha_i^{(k+1)} = \alpha_i^{(k)} - \tau \int_\Omega & (\phi(tr(D\mathbf{v}_i(\mathbf{x})adj(D\mathbf{f}_\alpha^{(k)}(\mathbf{x})))I_0(\mathbf{f}_\alpha^{(k)}(\mathbf{x})) \\ & + det(D\mathbf{f}_\alpha^{(k)}(\mathbf{x}))\langle\nabla I_0(\mathbf{f}_\alpha^{(k)}(\mathbf{x})), \mathbf{v}_i(\mathbf{x})\rangle)) \\ & (\phi(I_\alpha^{(k)}(\mathbf{x})) - I_l(\mathbf{x}))d\mathbf{x} \end{aligned} \tag{14}$$

where τ is the step size, $adj(.)$ denotes the adjoint matrix, ∇ is the gradient operator, $\langle ., . \rangle$ represents the standard inner product, and we assume that $\phi(.)$ is a linear operator. Finally, $I_{\boldsymbol{\alpha}^*}$ represents the reconstructed high-res image.

3.3. Discretization and implementation

In order to solve the underlying (high-res) optimization problem (7) we discretize the equation on the same grid as the high-res image, and utilize the (constrained) gradient descent-based solution described in [13] (details omitted for brevity). In our Matlab [22] based implementation, the average time for morphing two 256×256 images is $4.20 \pm 0.15 sec$. The outcome of the training procedure summarized in equation (11) is thus a set of vector fields each of the same size as the original high-res images. In the testing phase, equation (13) is discretized at the same resolution as the input low-res image with the operation $\phi(I_{\boldsymbol{\alpha}})$ accounting for the operation that transfers the high-res model $I_{\boldsymbol{\alpha}}$ onto the space of images of the same size as the input low-res image. The average time for construction of a high-res image from a low-res input (regardless of the required magnification) is about 4 minutes. The codes were executed on a MacBook pro, with $2.9GHz$ Intel Core $i7$ and $8GB$ $1600MHz$ $DDR3$.

4. Results

In order to evaluate the ability of our TB-SFSR method to reconstruct low-res images, we tested it on two face datasets, namely the extended Yale Face Database B (abbr. *YaleB*) and the AR face dataset. The YaleB face dataset consists of frontal pose images of size 192×168 pixels from 28 human subjects under 64 different illumination conditions. The cropped AR face dataset [18] contains 2600 images of size 160×120 pixels from 100 different subjects under 13 different conditions and with two images for each condition. In the experiments reported below we used 6 of these conditions for which the facial components were clearly visible (unobstructed). This resulted in a dataset consisting of 600 images from 100 subjects. The images are masked to remove background and hair. The degradation function, $\phi(.)$, is chosen to be a low pass filter combined with a downsampling operator as described in [27].

Our results are computed using a standard 'leave one subject out' cross validation procedure. That is, for both datasets all images from one person are left out and our TB-SFSR model (as well as the models to which we are comparing to) is trained on the images from the remaining subjects. We compare our TB-SFSR technique to a variety of techniques that were previously described. In particular, we compare the results of our algorithm to those of a kernel-PCA based SR [7], a sparse representation based SR method [27], and a method based on learning a linear mapping from low-res images to the corresponding high-res images [28], as well as a simple cubic B-spline interpolation (upsampling) procedure. We note that, with the exception of the B-spline interpolation procedure, all comparables utilize the learning-based mathematical framework described

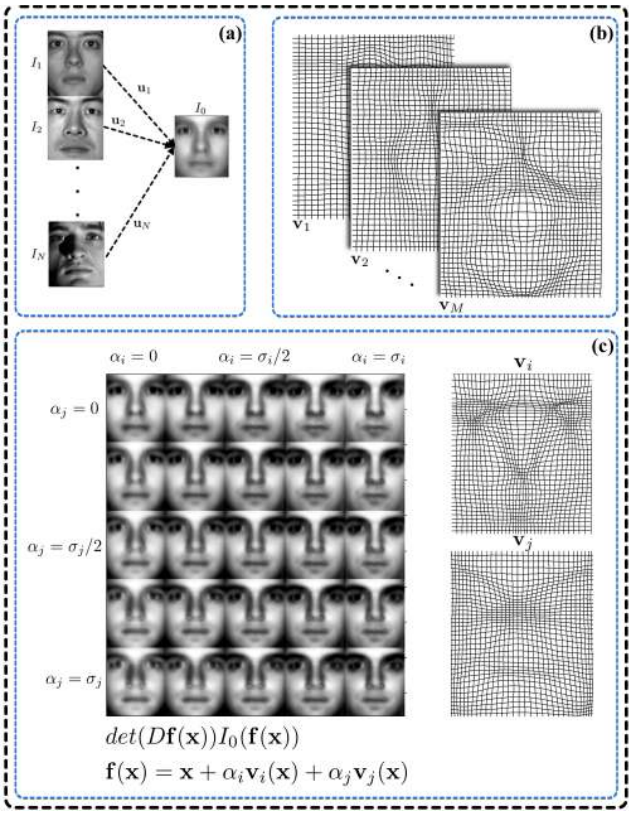

Figure 3. The images in the training set are morphed to the reference image and the optimal displacement fields are calculated for every image (a). The basis displacement fields, $\mathbf{v}_1, \ldots, \mathbf{v}_m$, are calculated as principal components of the optimal displacement fields (b). Demonstration of face modeling using only two of the displacement fields (c). 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.

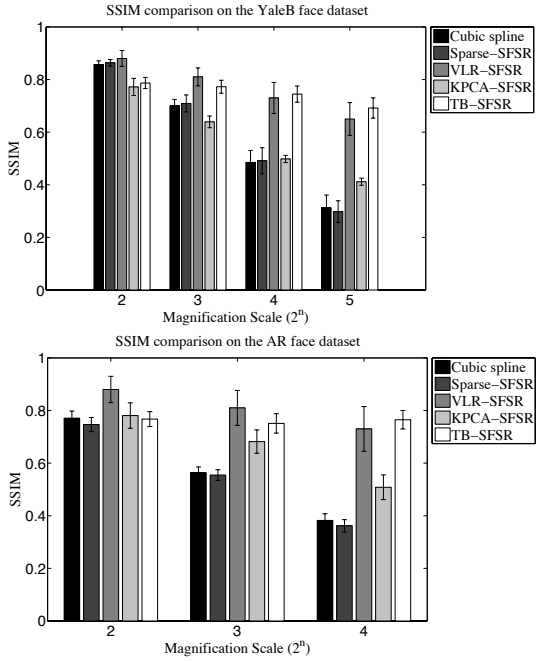

Figure 4. Mean and standard deviation of the structural similarity (SSIM), measured for the Yaleb face dataset (a) and for the AR face dataset (b) for different scales of magnification, between the reconstructed high res image and the original image using cubic spline, the method introduced in [27] (Sparse-SFSR), the method introduced in [28](VLR-SFSR), the method introduced in [7] (KPCA-SFSR), and our method (TB-SFSR).

above. In the experiments presented below, all images pertaining to a subject are removed from the training procedure, and all methods are trained and tested using exactly the same data.

Figure 4 shows the comparison of the mean and standard deviation of the structural similarity (SSIM) index [25] between the original high-res images and the reconstructed images using each method at different scales of magnification, with each scale corresponding to a reduction in size of 2^n, for $n = 2, 3, 4, 5$ (we're seeking to evaluate methods for constructing very low resolution images). The results for 32x magnification ($n = 5$) are not shown for the AR dataset, because the low-res images were of the size 5×4 pixels and all methods failed to reconstruct meaningful high-res images. From Figure 4, it can be seen that our proposed method outperforms the other methods significantly for higher magnification scales (i.e. very low resolution image reconstruction). This is while our method maintains the same reconstruction performance throughout different magnification scales. From the statistical point of view, the improvements provided by TB-SFSR are significant: p-value$<$ 0.01, using t-test statistics). Figures 5 and 6 show the SFSR reconstruction results of these methods for 32x and 16x magnifications and for the YaleB and the AR face datasets, respectively. The performance of our proposed method is comparable to the state of the art methods and outperforms them in the very low resolution setting. The sample images shown in these figures are chosen to have SSIM values close to the average SSIM (of our method) reported in Figure 4 for these datasets.

Figure 7 shows similar result as Figures 5 and 6, as mentioned before, this time the test is done by leaving one instance (one of the 64 face images for a subject) out and repeating the experiment. It is clear that while our method's performance remains the same, performance of the methods introduced in [28] and [7] increases significantly. We note that this is merely because the model has already seen very similar images to the test image. In fact, in the 'leave one instance out' (as opposed to leave one subject out) situation a nearest neighbor search in the high-res training data can provide comparable, if not better, results to those of the mentioned methods. Figure 8 shows the nearest neighbor images in the high-res training dataset for the low-res test

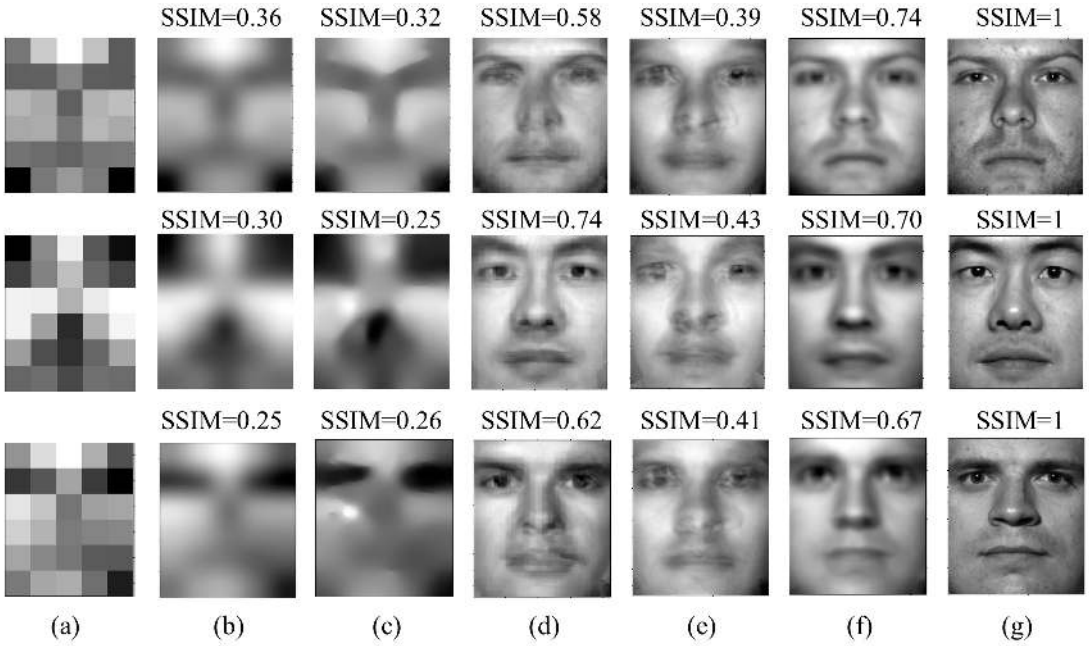

Figure 5. SFSR experiment with 'leave one subject out' training the YaleB face dataset (32x magnification). The degraded image (6×5pixels)(a), high-res image reconstructed using cubic spline (b), the method introduced in [27] (c), the method introduced in [28](d), the method introduced in [7] (e), TB-SFSR (f), and the original high-res image (g).

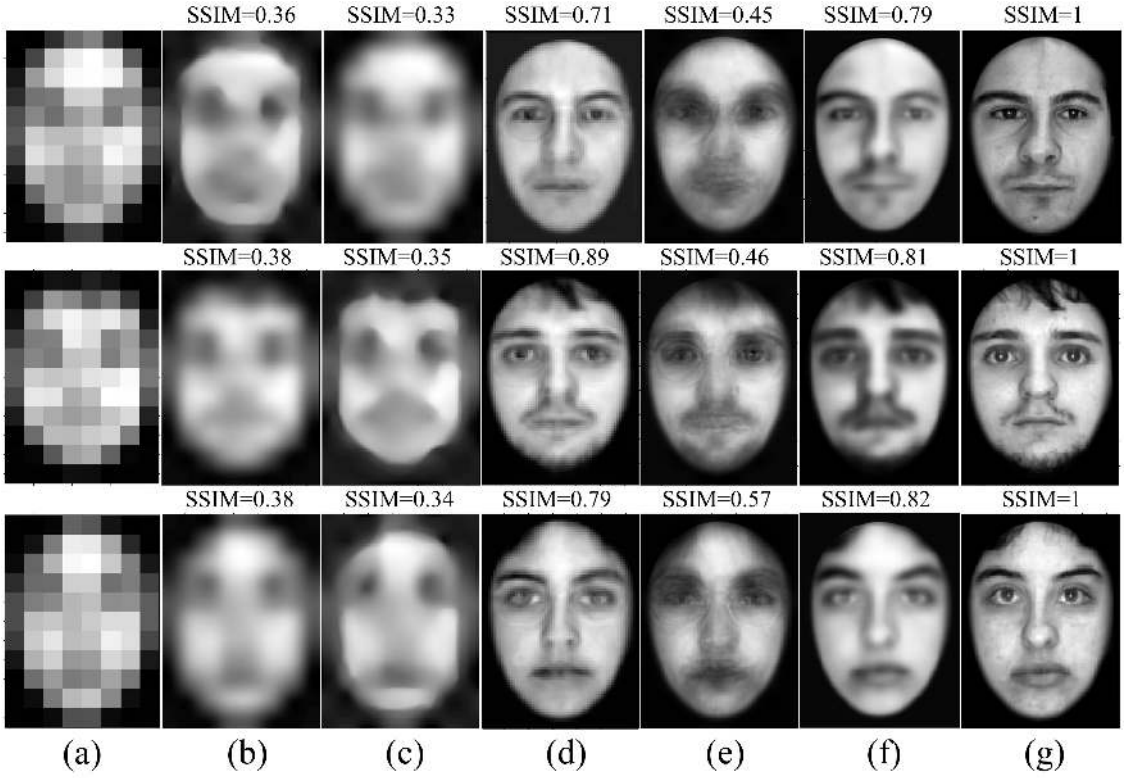

Figure 6. SFSR experiment with 'leave one subject out' training for the AR face dataset (16x magnification). The degraded image (11×8pixels)(a), high res image reconstructed using cubic spline (b), the method introduced in [27] (c), the method introduced in [28](d), the method introduced in [7] (e), TB-SFSR (f), and the original high-res image (g).

images in Figure 7. The reconstructed images are obtained from, $I^* = \text{argmin}_{I_{j=1,\ldots,N}} \frac{1}{2}\|I_l - \phi(I_j)\|_2^2$.

Finally, we propose an intuitive explanation on why the transport-based method is more effective. Unlike all other methods described to date, our transport-based method not only compares intensity values between images in modeling the problem, but also the location of the intensities. 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. This is shown by plotting the cumulative energy content for the principal components of the Euclidean embedding (signal space) and the transport embedding. Figures 9 and 10 show the cumulative energy content of the principal components as a function of the number of principal components in the YaleB and AR datasets, respectively. It can be seen that the variations in the datasets are captured with very few principal components in the transport space.

5. Summary and Discussion

We have described a new learning-based method for reconstructing high resolution estimates from single frame low resolution images. Our method, denoted as transport-based SFSR, employs an optimal transport formulation to derive a facial appearance model from training data, without the need for the definition of correspondence landmarks. In contrast to previously described SFSR methods, which seek to reconstruct a high resolution as a linear combination of 'basis' image patches, our approach utilizes a transport-based mathematical model for the entire facial region of interest. The model is non linear, and Lagrangian (in PDE parlance) in the sense that it compares intensities at different image coordinates. Results computed using two well-known, publicly available, image databases show that the reconstruction capabilities of our transport-based approach, especially for very high magnification tasks (e.g. 8 or 16 times), are comparable and superior to other state of the art methods [28, 7, 27] in unsupervised settings (where the training phase does not include data from test subjects).

We note that the technique described here is closely related to the linear optimal transport framework described in [23]. In fact, our technique can be seen as a PCA-based facial appearance model constructed on the linear optimal transport embeddings produced by the method described in Wang et al. [23]. As such, the model will completely recover, up to interpolation and derivative estimation errors, any image in the training set when all eigenvectors corresponding to nonzero eigenvalues are used in the model reconstruction. The interpolation errors will be introduced given the necessity to differentiate and interpolate data in the reconstructed model 6.

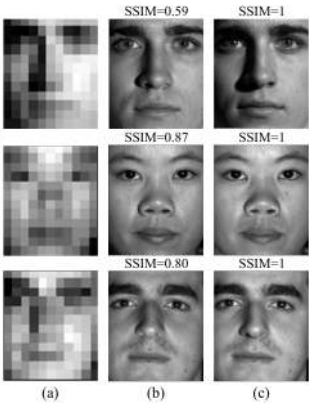

Figure 8. Nearest neighbor reconstruction with a 'leave one instance out' training data for the same images as in Figure 7. The degraded image (a), the nearest neighbor in the high-res training set (b), and the original image (c). We clarify that in these results the training set contains all but one (the test image) of the instances from a particular person.

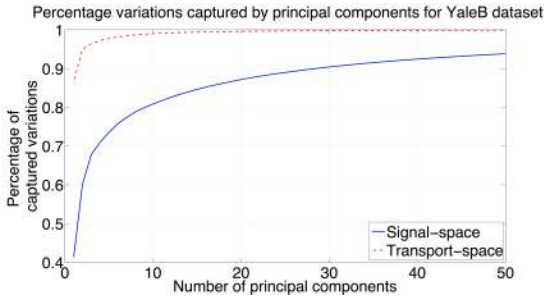

Figure 9. Percentage variations captured by the principal components in the YaleB dataset, in the image space and in the transport-based.

Finally, we mention that the PCA-based modeling procedure described here is one of many subspace learning techniques that can be used for designing a transport-based super-resolution approach. Given the 'localized' nature of the problem a subspace learning model which is more spatially sparse (see for example [26]) could aid the modeling procedure while at the same time simplifying the optimization problem given that the warping of non overlapping parts (e.g eyes) could be computed separately. In the future, we also wish to study the ability of the method to reconstruct facial images which are partially obstructed. To that end, our transport-based approach could be modified to include data from only specified regions of interest, for example. These and other topics will be subject of future work.

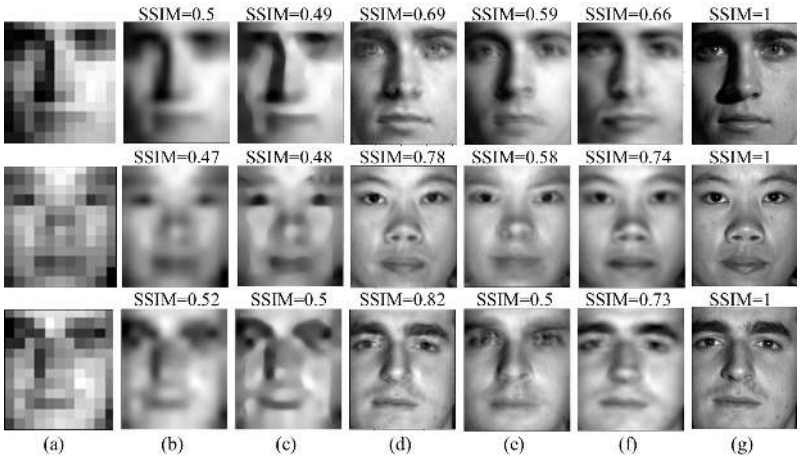

Figure 7. SFSR experiment with 'leave one instance out' training for the YaleB face dataset (16x magnification). The degraded image (12×11pixels)(a), high res image reconstructed using cubic spline (b), the method introduced in [27] (c), the method introduced in [28](d), the method introduced in [7] (e), TB-SFSR (f), and the original high-res image (g). We clarify that in these results the training set contains all but one (the test image) of the instances from a particular person.

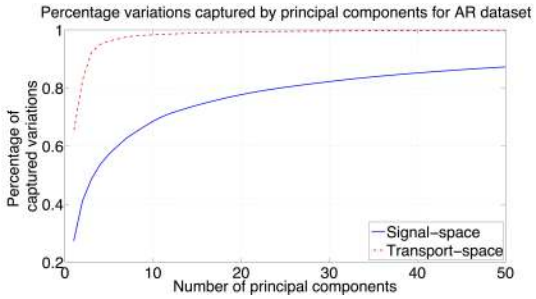

Figure 10. Percentage variations captured by the principal components in the AR dataset, in the image space and in the transport-based.

6. Acknowledgement

This work was financially supported by the National Science Foundation (NSF), grant number 1421502, and the John and Claire Bertucci Graduate Fellowship.

References

[1] L. Ambrosio. Optimal transport maps in monge-kantorovich problem. *arXiv preprint math/0304389*, 2003. 1

[2] S. Baker and T. Kanade. Hallucinating faces. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 83–88. IEEE, 2000. 1

[3] S. Basu, S. Kolouri, and G. K. Rohde. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. *Proceedings of the National Academy of Sciences*, 111(9):3448–3453, 2014. 3

[4] S. Borman and R. L. Stevenson. Super-resolution from image sequences-a review. In *Circuits and Systems, Midwest Symposium on*, pages 374–374. IEEE Computer Society, 1998. 1

[5] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991. 3

[6] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–627. IEEE, 2001. 2

[7] A. Chakrabarti, A. Rajagopalan, and R. Chellappa. Super-resolution of face images using kernel pca-based prior. *IEEE Transactions on Multimedia*, 9(4):888–892, 2007. 1, 2, 4, 5, 6, 7, 8

[8] F. De la Torre. A least-squares framework for component analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(6):1041–1055, 2012. 4

[9] S. Farsiu, M. Elad, and P. Milanfar. Multiframe demosaicing and super-resolution of color images. *Image Processing, IEEE Transactions on*, 15(1):141–159, 2006. 1

[10] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *Image processing, IEEE Transactions on*, 13(10):1327–1344, 2004. 1, 2

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *Computer Graphics and Applications, IEEE*, 22(2):56–65, 2002. 1

[12] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001. 2

[13] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004. 1, 3, 4

[14] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 27(5):684–698, 2005. 2

[15] C. Liu, H.-Y. Shum, and W. T. Freeman. Face hallucination: Theory and practice. *International Journal of Computer Vision*, 75(1):115–134, 2007. 1

[16] A. Marquina and S. J. Osher. Image super-resolution by tv-regularization and bregman iteration. *Journal of Scientific Computing*, 37(3):367–382, 2008. 1

[17] A. M. Martinez. The AR face database. *CVC Technical Report*, 24, 1998. 2

[18] A. M. Martínez and A. C. Kak. Pca versus lda. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):228–233, 2001. 2, 4

[19] H. Nasir, V. Stanković, and S. Marshall. Singular value decomposition based fusion for super-resolution image reconstruction. *Signal Processing: Image Communication*, 27(2):180–191, 2012. 1

[20] K. Nasrollahi and T. B. Moeslund. Super-resolution: A comprehensive survey. *Machine Vision & Applications*, 2014. 1, 2

[21] R. R. Schultz and R. L. Stevenson. A bayesian approach to image expansion for improved definition. *Image Processing, IEEE Transactions on*, 3(3):233–242, 1994. 2

[22] The MathWorks, Inc., Natick, Massachusetts, United States. *MATLAB Release 2013a*. 4

[23] W. Wang, D. Slepcev, S. Basu, J. A. Ozolek, and G. K. Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *International journal of computer vision*, 101(2):254–269, 2013. 3, 7

[24] X. Wang and X. Tang. Hallucinating face by eigentransformation. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 35(3):425–434, 2005. 2

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004. 5

[26] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009. 4, 7

[27] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010. 1, 2, 4, 5, 6, 7, 8

[28] W. W. Zou and P. C. Yuen. Very low resolution face recognition problem. *Image Processing, IEEE Transactions on*, 21(1):327–340, 2012. 1, 2, 4, 5, 6, 7, 8