Sparse Representation-based Super-Resolution for Face Recognition At a Distance

E. Bilgazyev
emilbek@cs.uh.edu

B. Efraty
baefraty@uh.edu

S. K. Shah
shah@cs.uh.edu

I. A. Kakadiaris
ioannisk@uh.edu

Dept. of Computer Science
University of Houston
Houston, TX, 77204-3010
USA

Face recognition is a challenging task, especially when low-resolution images or image sequences are used. In typical surveillance scenarios, cameras are often at a considerable distance from the subjects [2]. Hence, the captured image typically contains only a small region surrounding the subject's face, often characterized by a small interpupillary distance (IPD). This decrease in image resolution results in the loss of facial high-frequency components leading to a decrease in recognition rates. Therefore, in order to maintain the robustness of face recognition at a distance (FRAD) systems, it is important to find a solution to this difficult task [1].

In this paper, we propose a new approach to obtain a super-resolved (SR) image by learning the high-frequency components of high-resolution (HR) facial images and applying them to a given low-resolution (LR) image to create the SR image (Fig. 1). In the training stage, we use a Dual Tree Complex Wavelet Transform (DT-CWT) to extract the high-frequency components from a database of HR facial images and synthetically generate LR images. A dictionary is built with the high-frequency components for each of the two databases (HR and LR). In the reconstruction stage, we compute a sparse representation of the input LR image using the dictionary built for LR images and estimate the HR high-frequency components using that sparse representation with respect to the HR dictionary. The estimated high-frequency components of the HR image are then added to the LR input image to create a SR image. Instead of using the whole facial image, we divide it into patches, which overlap to avoid "block effect" artifacts during reconstruction.

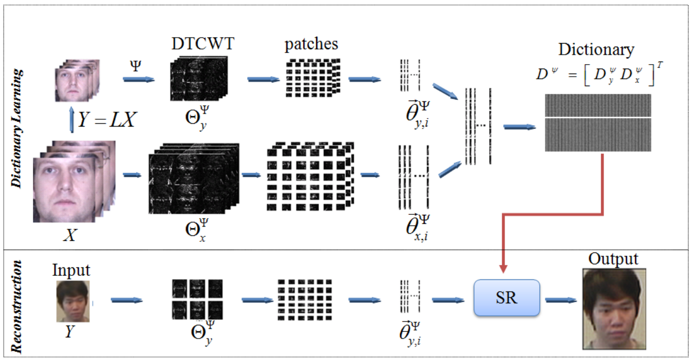

Figure 1: Depiction of the proposed framework for super-resolution reconstruction.

The relationship between a degraded LR image, Y, and the HR image, X, can be described as:

$$Y = \mathbf{H}X + \eta \ , \tag{1}$$

where $\mathbf{H}$ is the linear transformation matrix, that downsamples, blurs and transforms image X, and η represents and additive i.i.d. Gaussian with zero mean noise.

To estimate image X, Eq. 1 can be re-written as:

$$X = \tilde{Y} + \mathbf{H}^{\dagger}\eta + \Gamma_X \ , \tag{2}$$

where, $\mathbf{H}^{\dagger}$ denotes the pseudo-inverse of $\mathbf{H}$, $\Gamma_X = X - \mathbf{H}^{\dagger}\mathbf{H}X$ is the information loss, and $\tilde{Y}$ is the upsampled version of the input LR image. Let Ψ be an operator that extracts the high- and low-frequency components of an image. Then, reconstruction of X can be written as:

$$X = \Psi^{-1}([\beta \quad \mathbf{0}]^T) + \Psi^{-1}([\mathbf{0} \quad \theta]^T) \ . \tag{3}$$

Combining Eqs. 2-3, the HR image X can be estimated as:

$$X \approx \Psi^{-1}(\Psi L^{-1}Y + \hat{\Theta}^{\Psi}), \quad \hat{\Theta}^{\Psi} = \{\hat{\theta}^{\Psi}_{x,1} \dots \hat{\theta}^{\Psi}_{x,n}\}, \tag{4}$$

Figure 2: Illustration of the surveillance camera output and the SR output. (a) Depiction of a frame acquired by surveillance camera (the black bounding box output indicates successful face detection), (b) magnification of the area in the bounding box (IPD≈ 11 pixels), (c) output of BCI, and (d) output of the proposed (UHSR) algorithm.

.

where $\hat{\Theta}^{\Psi}$ contains the high-frequency components of image X, and L^{-1} is an upsampling operator. To estimate X in Eq. 4, we need to first estimate $\hat{\Theta}^{\Psi}$. We estimate it by learning the high-frequency components of the LR and HR images in the training dataset. Let $\{x_1, \dots, x_n\} \in X$ be a set of n overlapping square patches of the HR image, and $\{y_1, \dots, y_n\} \in Y$ be the set of corresponding patches of the LR image. Let us denote the high-frequency coefficients associated with x_i and y_i as $\vec{\theta}^{\Psi}_{x,i}$ and $\vec{\theta}^{\Psi}_{y,i}$, respectively.

Using $\vec{\theta}^{\Psi}_i = [\vec{\theta}^{\Psi}_{x,i} \quad \vec{\theta}^{\Psi}_{y,i}]^T$, where $\{\vec{\theta}^{\Psi}_i, \dots, \vec{\theta}^{\Psi}_n\} \in \vec{\Theta}^{\Psi}$, we need to build a dictionary $\mathbf{D}^{\Psi}$, which results in an accurate and sparse-reconstruction of the images in the training set. Specifically, the dictionary is learnt from a paired input vector, $\vec{\theta}^{\Psi}_i$, and should satisfy the following condition:

$$D^{\Psi} = \underset{D^{\Psi},\vec{\alpha}}{\arg\min} \left\| \vec{\theta}^{\Psi}_i - D^{\Psi}\alpha_i \right\|_2 + \lambda \|\alpha_i\|_1 . \tag{5}$$

In the reconstruction step, given the high-frequency component of the patch descriptor of the input LR image, $\vec{\theta}^{\Psi}_{y,i}$, its sparse-representation, α_i, is obtained by minimizing

$$\alpha^*_i = \underset{\alpha_i}{\arg\min} \left\| \vec{\theta}^{\Psi}_{y,i} - D^{\Psi}_y \alpha_i \right\|_2 + \lambda \|\alpha_i\|_1 . \tag{6}$$

The SR patch is recovered as:

$$\hat{\theta}^{\Psi}_{x,i} = D^{\Psi}_x \alpha^*_i , \tag{7}$$

where $\hat{\theta}^{\Psi}_{x,i}$ is used in Eq. 4 to reconstruct the SR image. We compared the proposed DT-CWT-based SR method (UHSR) with other SR algorithms and empirically demonstrated the advantage of the proposed method compared to several state-of-art super-resolution algorithms for the task of face recognition.

[1] M. Ao, D. Yi, Z. Lei, and S. Z. Li. *Handbook of remote biometrics*, chapter Face Recognition at a Distance: System Issues, pages 155–167. Springer London, 2009.

[2] F.W. Wheeler, X.M. Liu, and P.H. Tu. Multi-frame super-resolution for face recognition. In *Proc. 1st International Conference on Biometrics Theory, Applications and Systems*, pages 1–6, Washington D.C, Sep. 27-29 2007.