Improved Face Recognition Using Super-Resolution

Emil Bilgazyev, Boris Efraty, Shishir K. Shah, Ioannis A. Kakadiaris
Dept. of Computer Science
University of Houston
Houston, TX, USA
{emilbek,baefraty,shah}@cs.uh.edu, ioannisk@uh.edu

Abstract

Face recognition is a challenging task, especially when low-resolution images or image sequences are used. A decrease in image resolution typically results in loss of facial component details leading to a decrease in recognition rates. In this paper, we propose a new method for super-resolution by first learning the high-frequency components in the facial data that can be added to a low-resolution input image to create a super-resolved image. Our method is different from conventional methods as we estimate the high-frequency components, that are not used in other methods, to reconstruct a higher-resolution image, rather than studying the direct relationship between the high- and low-resolution images. Quantitative and qualitative results are reported for both synthetic and surveillance facial image databases.

1. Introduction

Recent advances in electronics, sensors, and optics have led to a widespread availability of video-based surveillance and monitoring systems. Many applications, ranging from security to broadcasting, are driving the need for better recognition techniques to recognize individuals from surveillance videos [1]. This poses new challenges to the already difficult task of face recognition (FR). In typical surveillance scenarios, cameras are often at a considerable distance from the subjects [19]. Hence, the captured image typically contains only a small region enclosing the subject's face, often with only a small number of pixels between the eyes. This attribute usually leads to poor face recognition rates. Critical to building a robust face recognition at a distance (FRAD) system is finding a solution to these issues [1]. For many 2D FR systems, image resolution is the critical factor that affects its ability to detect key elements of the facial anatomy (e.g., eyes, lip corners, and facial contours) [19]. When the resolution of the image is lower than anticipated, the performance of the system deteriorates. Figure 1 depicts recognition rates, obtained from databases generated at different resolutions, plotted against the inter-pupillary distance (IPD). From Fig. 1, it is evident that the resolution of the facial image and the recognition performance of the FR system is strongly correlated, especially for IPD lower than 17 pixels.

Many super-resolution (SR) approaches have tried to minimize the mean-squared-error (MSE) or maximize the signal-to-noise ratio (SNR) between the original high-resolution (HR) and the reconstructed SR images [5]. However, in face recognition, these approaches may not perform well, as most FR systems rely on the ability to identify key facial features, typically underscored by the high-frequency content. Obtaining a higher SNR does not necessarily contribute to a higher recognition rate since high-fidelity reconstruction of low-frequency content may dominate the image.

We propose a new approach to obtain a SR image by first learning the high-frequency components of facial images and then applying them to a given low-resolution (LR) image to create the SR image (Fig. 2). In the learning stage (Sec. 3.3), we use a Dual-Tree Complex Wavelet Transform (DT-CWT) to extract the high-frequency components from a database of facial images, and use this information in the reconstruction stage (Sec. 3.4) to create a SR image for a given LR input image (Fig. 2). Selection of the high-frequency components for reconstruction is based on the similarity between the features of the input image and the images from the training database. To accomplish this, we propose to use the manifold learning via sparse representation methodology.

The rest of the paper is organized as follows: Previous work is presented in Section 2. Description of the proposed method is presented in Section 3. The experimental design and the results of the proposed algorithm are presented in Section 4. Finally, in Section 5, we offer our findings.

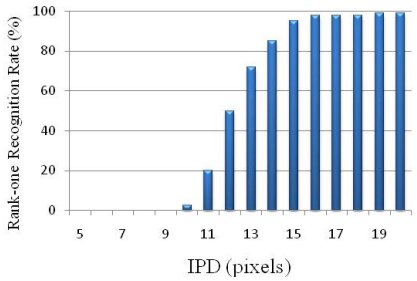


Figure 1. Depiction of rank-one recognition rate for the CMU-PIE database. The CMU-PIE Database with 201 subjects (Section 1 ∩ Section 4) is used for this experiment.

2. Previous Work

In recent years, many methods have been proposed that address the issue of image resolution. Existing SR algorithms may be classified into two classes: multi-frame-based and example-based algorithms. The methods that belong to the first class attempt to compute a HR image from a set of LR images from any domain [5], while the methods in the second class attempt to find a HR counterpart of a single LR image from a known domain [2, 12, 18, 7, 13]. We are interested in domain-specific SR for face recognition. In addition, the SR algorithms for facial images usually differ in the way the *a-priori* knowledge of facial images is applied. Hallucinating LR facial images was proposed by Baker *et al.* [2]. In that approach, an image pyramid was built to learn the prior. Lui *et al.* [12] proposed a two step approach in which principal component analysis (PCA) was used to obtain the global HR image and then, a patch-based non-parametric Markov network was used to learn the residual image. Li *et al.* [10] extended the work of Baker *et al.* [2] to a two step approach where PCA was applied on both the HR and LR images to obtain the global HR image. An MRF prior was used to learn the residual in the second step. Jia *et al.* [9] developed a tensor-based SR method where a tensor of the HR and LR images is built in the training step to reconstruct a SR image. Simultaneous SR and feature extraction for face recognition was proposed by Hennings *et al.* [7] to estimate a SR image by searching for similar features of the LR image in the training set. However, this method uses the relationship between the HR and LR images, and its performance depends on the training dataset. Yang *et al.* [22] introduced a method to reconstruct SR images using a sparse representation of the input LR images. SR reconstruction based on wavelet analysis has been shown to be well suited for reconstruction, denoising and deblurring, and has been used in a variety of application domains including biomedical [8], biometrics [4], and astronomy [20]. In addition, it provides an accurate and sparse representation of images that consist of smooth regions with isolated abrupt changes [15].

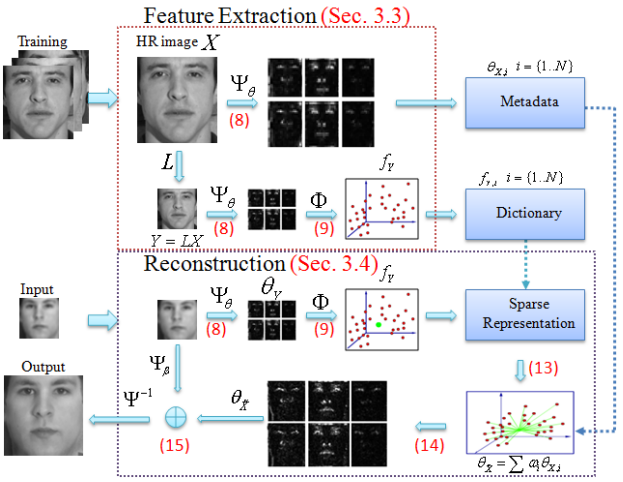


Figure 2. Depiction of the proposed framework for SR reconstruction. The top box (in red dotted lines) illustrates the steps involved in feature extraction. The bottom box (in purple dotted line) illustrates the steps involved in reconstruction. A detailed description of the steps is presented in Sec. 3.

In our method, we propose to take advantage of the wavelet decomposition-based approach in conjunction with compressed sensing techniques to improve face recognition rates. Unlike other methods [2, 12, 22], where the performance is optimized based on the MSE or the qualitative output, the proposed method's performance is optimized with respect to the face recognition task.

3. Methodology

3.1. Notations

The HR and LR images are denoted as X and Y, respectively. We use bold uppercase to denote the transformation matrices (or projection matrices). Specifically, we use $\mathbf{\Psi}$ to denote the dual-tree complex wavelet transform in a matrix form. Bold lowercase letters denote vectors. Plain uppercase letters denote regular matrices (i.e., L is used as a downsampling operation in matrix form). Plain lowercase letters are used as scalars. Superscript t is used to denote $training$.

3.2. Super Resolution Problem Formulation

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 additive i.i.d. Gaussian with zero mean noise parameter. To estimate image X, Eq.(1) can be written as:

$$\mathbf{H}^{\dagger}Y - \mathbf{H}^{\dagger}\eta = \mathbf{H}^{\dagger}\mathbf{H}X \ , \tag{2}$$

where $\mathbf{H}^{\dagger}$ denotes the pseudo-inverse of $\mathbf{H}$. In general, $\mathbf{H}^{\dagger}\mathbf{H}X \neq X$, and therefore:

$$X - \mathbf{H}^{\dagger}\mathbf{H}X = \Gamma_X \ ,$$

where Γ_X is the information that is lost during the pseudo-inverse transformation. Let $\tilde{Y} = \mathbf{H}^{\dagger}Y$, where $\tilde{Y}$ is the inverse transformation of image, Y. Then, Eq. 2 can be written as:

$$X = \tilde{Y} + \mathbf{H}^{\dagger}\eta + \Gamma_X \ .$$

Now, we formulate the relationship between the LR and the HR images using high- and low-frequency components. Let $\mathbf{\Psi} = [\mathbf{\Psi}_{\beta} \ \ \mathbf{\Psi}_{\theta}]$ be an operator that extracts the high- and low-frequency components ($\boldsymbol{\beta}_X = \mathbf{\Psi}_{\beta}X$ and $\boldsymbol{\theta}_X = \mathbf{\Psi}_{\theta}X$, respectively) of an image:

$$[\boldsymbol{\beta}_X \ \ \boldsymbol{\theta}_X]^T = \mathbf{\Psi}X \ .$$

Assuming that $\mathbf{\Psi}$ is orthonormal, then, by applying the inverse of $\mathbf{\Psi}$ on both sides, the original image, X, can be reconstructed. By splitting the high- and low-frequency components as:

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

where $\mathbf{0}$ is a vector with its elements being zero, Eq. 3 can be written as:

$$X = \mathbf{\Psi}^{-1}([\boldsymbol{\beta} \ \ \mathbf{0}]^T) + \mathbf{\Psi}^{-1}([\mathbf{0} \ \ \boldsymbol{\theta}]^T) \ . \tag{4}$$

Let, $\mathbf{\Psi}^{-1}([\boldsymbol{\beta} \ \ \mathbf{0}]^T)$ be denoted as Y^*, and $\mathbf{\Psi}^{-1}([\mathbf{0} \ \ \boldsymbol{\theta}]^T)$ as G. Then Eq. 4 can be written as:

$$X = Y^* + G \ , \tag{5}$$

where Y^* contains the low-frequency components, G contains the high-frequency components of image X, and $Y^* \perp G$. Generally, both Y^* and $\tilde{Y}$ contain almost similar low-frequency information, even though their formation processes are different[1]. Hence, we assume that:

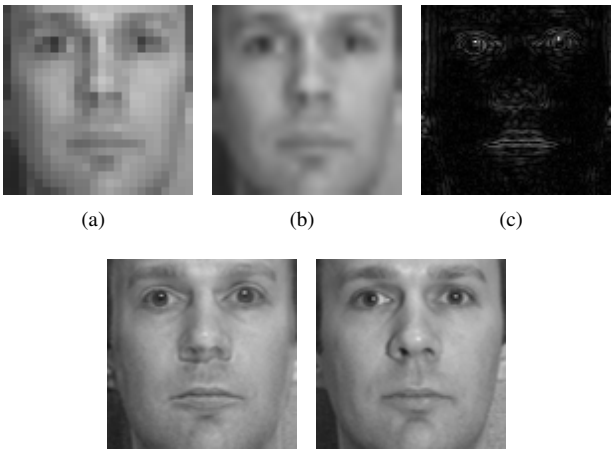
$$\left\| Y^* - \tilde{Y} \right\|_2 \leq \varepsilon \ , \tag{6}$$

where $\varepsilon > 0$ is a small value. Let $Y^* = \tilde{Y} + \Lambda$, where $\|\Lambda\|_2 = \varepsilon$. Then Eq. 5 can be reformulated as:

$$X = \tilde{Y} + G + \Lambda \ . \tag{7}$$

In Eq. 7, G can be estimated using the facial images in the training dataset according to method presented in Secs. 3.3-3.4. Figure 3 depicts the Y, $\tilde{Y}$, G, X, and original images.

[1]In our experiments on 1,000 randomly selected facial images from the CMU-MultiPIE dataset, the average auto-correlation between Y^* and $\tilde{Y}$ was 0.988.



(a) (b) (c) (d) (e)

Figure 3. Super-Resolution performed on facial image. (a) Low-resolution input image (Y), (b) Magnified (4x) input image ($\tilde{Y}$), (c) Estimated high-frequency components image (G), (d) Super-resolved image (X), and (e) Original image.

3.3. Feature Extraction

The Dual-Tree Complex Wavelet Transform (DT-CWT) has been extensively used for face recognition [11]. In our approach, we use the DT-CWT to separate an image into low and high-frequency components. Let us denote $\mathbf{\Psi}$ as the DT-CWT transformation matrix, and $\boldsymbol{\theta}$ as the high-frequency components of DT-CWT. Then, the DT-CWT of images X and Y are represented as:

$$[\boldsymbol{\beta}_X \ \ \boldsymbol{\theta}_X]^T = \mathbf{\Psi}X, \ \text{ and } \ [\boldsymbol{\beta}_{\mathrm{Y}} \ \ \boldsymbol{\theta}_{\mathrm{Y}}]^{\mathrm{T}} = \mathbf{\Psi}\mathrm{LX} = \mathbf{\Psi}\mathrm{Y} \tag{8}$$

where L is the decimation matrix that resizes the training image X to the size of image Y. Naturally, there exists a one-to-one correspondence between $\mathbf{\Psi}X_i$ and $\mathbf{\Psi}Y_i$ for every image X_i in the training set. Let us denote the high-frequency coefficients in Eq. 8 as $\boldsymbol{\theta}^t_{X,i}$ and $\boldsymbol{\theta}^t_{Y,i}$ ($i = 1, ..., N$). The details of how this relationship is modeled to obtain the SR image is presented in Sec. 3.4. The low-frequency components do not contain useful information for the SR task and may be ignored[2].

For any given image Y_i of size $n \times m$ pixels, the corresponding dimension of $\boldsymbol{\theta}_{Y,i}$ is $(6 \times \frac{m}{2} \times \frac{n}{2}) \times 2$. To decrease the dimension of the $\boldsymbol{\theta}_{Y,i}$, it is further converted to a compact descriptor as follows:

$$\boldsymbol{f}_{Y,i} = \mathbf{\Phi}\boldsymbol{\theta}_{Y,i} \ . \tag{9}$$

We propose to employ a supervised manifold learning such as Linear Discriminant Analysis (LDA) [14]. Algorithm 1 summarizes the feature extraction framework.

[2] As the number of decomposition level of the DT-CWT increases, the number of low-frequency components decreases, and each image low-frequency components will be similar to each other.

Algorithm 1 Feature Extraction

Input: X_{ref}, X_i, $i = \{1, \ldots, N\}$
Output: $\boldsymbol{f}^t_{Y,i}$, $\boldsymbol{\theta}^t_{X,i}$, $i = \{1, \ldots, N\}$
1: **for all** $X_i, i = \{1, \ldots, N\}$ **do**
2: Register X_i to X_{ref} (Sec. 3.5)
3: Downsample $Y_i \leftarrow LX_i$
4: Extract $\boldsymbol{\theta}^t_{X,i} \leftarrow \boldsymbol{\Psi} X_i$, $\boldsymbol{\theta}^t_{Y,i} \leftarrow \boldsymbol{\Psi} Y_i$ (Eq. 8)
5: Compute $\boldsymbol{f}^t_{Y,i} \leftarrow \boldsymbol{\Phi}\boldsymbol{\theta}^t_{Y,i}$ (Eq. 9)
6: **end for**

3.4. Reconstruction

In the reconstruction stage, we estimate $\boldsymbol{\theta}_X$ using $\boldsymbol{\theta}_Y$ and $\{\boldsymbol{f}^t_{Y,i}, \boldsymbol{\theta}^t_{X,i}\}$. We assume that $\boldsymbol{\theta}_X$ can be represented as a weighted combination of the training images, such that:

$$\left\| \boldsymbol{\theta}_X - \sum \boldsymbol{\omega}_i \boldsymbol{\theta}^t_{X,i} \right\|_2 \leq \varepsilon_0 \ , \tag{10}$$

where $\boldsymbol{\omega}_i$ is the weight for $\boldsymbol{\theta}^t_{X,i}$ and $\varepsilon > 0$ is a small value, which also satisfies the condition:

$$\left\| \boldsymbol{f}_Y - \sum \boldsymbol{\omega}_i \boldsymbol{f}^t_{Y,i} \right\|_2 \leq \varepsilon_1 \ . \tag{11}$$

Recent research in applying sparse methods for facial image analysis [21] suggests that the appearance of human faces in images may be efficiently modeled using the sparse signal decomposition methodology. This motivates us to incorporate a sparsity constraint in the MAP formulation of signal reconstruction:

$$\min_{\boldsymbol{\omega}} \|\boldsymbol{f}_Y - A\boldsymbol{\omega}\|_2 + \lambda \|\boldsymbol{\omega}\|_0 \ , \tag{12}$$

where the parameter λ balances the sparsity of the solution and the quality of reconstruction, and $A = [\boldsymbol{f}^t_{Y,1} \ldots \boldsymbol{f}^t_{Y,N}]$. The optimization problem in Eq. 12 is known to be NP hard, due to the combinatorial nature of l_0 norm. Prior work in compressed sensing [21] suggests that as long as the underlying solution for $\boldsymbol{\omega}$ is sufficiently sparse, it can be efficiently recovered by minimizing the l_1 norm:

$$\min_{\boldsymbol{\omega}} \|\boldsymbol{f}_Y - A\boldsymbol{\omega}\|_2 + \lambda \|\boldsymbol{\omega}\|_1 \ . \tag{13}$$

The $\boldsymbol{f}^t_Y$ space is different from the $\boldsymbol{\theta}^t_X$ space. Thus, weights ($\boldsymbol{\omega}$) are normalized such that $\sum \boldsymbol{\omega}_i = 1$, and G is estimated as follows:

$$\boldsymbol{\theta}_X = \sum \boldsymbol{\omega}_i \boldsymbol{\theta}^t_{X,i} \ , \tag{14}$$

$$G = \boldsymbol{\Psi}^{-1}([\mathbf{0} \ \ \boldsymbol{\theta}_X]^T) \ .$$

Given the input LR image Y, landmarks are first detected and registered to the reference facial image, and are then up-sampled to generate $\tilde{Y}$. However, $\tilde{Y}$ is not orthogonal to G. Hence, we can modify reconstruction of image X as:

$$X = \boldsymbol{\Psi}^{-1}([\boldsymbol{\beta}_{\tilde{Y}} \ \ \boldsymbol{\theta}_X]^T), \tag{15}$$

where $\boldsymbol{\beta}_{\tilde{Y}}$ are the low-frequency components of image $\tilde{Y}$. The reconstruction framework is summarized in Algorithm 2.

Algorithm 2 Reconstruction

Input: Y, Y_{ref}, $A = [\boldsymbol{f}^t_{Y,1}, \ldots, \boldsymbol{f}^t_{Y,N}]$, $\boldsymbol{\theta}^t_{X,i}$, $(i = \{1, \ldots, N\})$
Output: $\tilde{X}$
1: Register $Y_{ref} \leftarrow Y$ (Sec. 3.5)
2: Compute $[\boldsymbol{\beta}_Y \ \ \boldsymbol{\theta}_Y]^T = \boldsymbol{\Psi} Y$ (Eq. 8)
3: Compute $\boldsymbol{f}_Y \leftarrow \boldsymbol{\theta}_Y$ (Eq. 9)
4: Solve: $\boldsymbol{\omega} \leftarrow \min_{\boldsymbol{\omega}} \|\boldsymbol{f}_Y - A\boldsymbol{\omega}\|_2 + \lambda\|\boldsymbol{\omega}\|_1$
5: Normalize $\boldsymbol{\omega}^* \leftarrow \boldsymbol{\omega}/\|\boldsymbol{\omega}\|_1$
6: Fuse $\boldsymbol{\theta}_X \leftarrow \sum \boldsymbol{\omega}^*_i \boldsymbol{\theta}^t_{X,i}$ (Eq. 14)
7: Extract $\boldsymbol{\beta}_{\tilde{Y}} \leftarrow \boldsymbol{\Psi}_{\beta_{\tilde{Y}}} \tilde{Y}$ (Eq. 8)
8: Compute $X = \boldsymbol{\Psi}^{-1}([\boldsymbol{\beta}_{\tilde{Y}} \ \ \boldsymbol{\theta}_X]^T)$ (Eq. 15)

3.5. Implementation details

Dual-Tree Complex Wavelet Transform: To compute DT-CWT, two wavelet trees are developed in parallel, with the wavelets forming Hilbert pairs. Each of the two trees represent the real and imaginary parts of the wavelet transform. Due to the design of the filter banks of DT-CWT that incorporates the half sample delay, a small translation and scale difference does not affect the reconstruction, and makes it nearly shift invariant. Its other main property is directional selectivity. The DT-CWT coefficients that correspond to the high-frequency information $\boldsymbol{\theta}$ is:

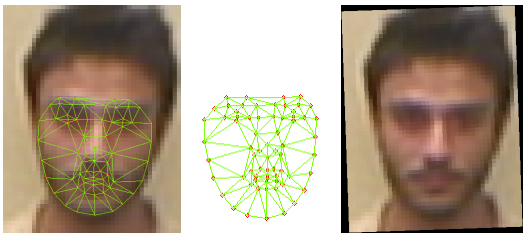
$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}^{+15} & \boldsymbol{\theta}^{+45} & \boldsymbol{\theta}^{+75} \\ \boldsymbol{\theta}^{-15} & \boldsymbol{\theta}^{-45} & \boldsymbol{\theta}^{-75} \end{bmatrix},$$

where $\boldsymbol{\theta}^{\pm 15}, \boldsymbol{\theta}^{\pm 45}, \boldsymbol{\theta}^{\pm 75}$ represent the components corresponding to the specific filters at the respective angles $(\pm 15°, \pm 45°, \pm 75°)$.

Face Alignment: Alignment of the facial images is necessary for SR reconstruction. To register all the images to a reference image, the scale (s), rotation ($\boldsymbol{R}$) and translation ($\mathbf{t}$) parameters should be estimated. Let $\Omega(\cdot)$ be an operator that returns a vector of facial landmarks coordinates $\mathbf{b}_Y = \Omega(Y)$, and $\mathbf{b}_r = \Omega(X_{ref})$ be a vector of facial landmark coordinates of a reference image Y_{ref}. To register the input image Y to X_{ref}, $s, R, \mathbf{t}$ can be estimated as:

$$\min_{s,\boldsymbol{R},t} \|\mathbf{b}_r - (s\boldsymbol{R}\mathbf{b}_Y + \mathbf{t})\|_2 \tag{16}$$

where $\mathbf{b}_Y$ is the facial landmark coordinates of image Y. Using the parameters estimated from Eq. 16, image Y is resized, rotated and translated. We leverage the Active Shape Model (ASM) [3, 17] to detect the key anatomical landmarks in each image. Using the feature points detected by ASM, we register all the images to a reference image (Fig. 4).



(a) (b) (c)

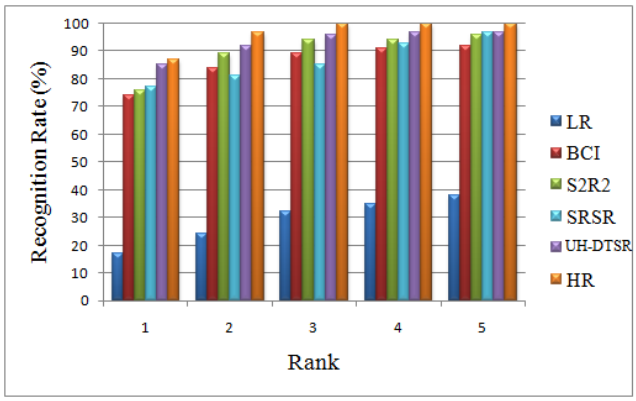
Figure 4. Depiction of registration. (a) Input image with facial landmark coordinates detected by ASM, (b) Landmarks of reference image, and (c) Registered image.

4. Experimental Results

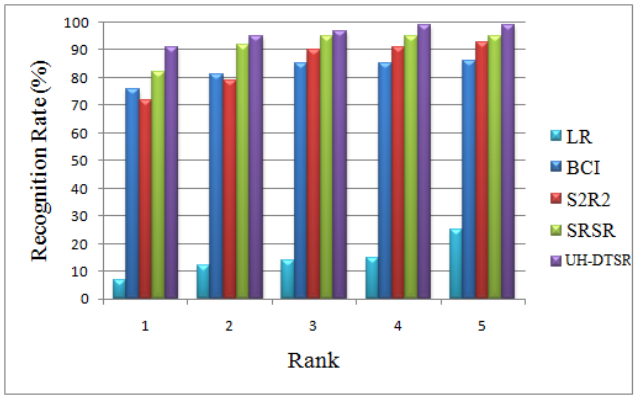
Several experiments were performed to evaluate the performance of the proposed algorithm. We compared the performance of the proposed method, which will be henceforth referred to as UH-DTSR, with other SR algorithms: bicubic interpolation (BCI), Simultaneous Super Resolution and Recognition (S2R2) [7], and Super-Resolution via Sparse Representation (SRSR) [23]. We used the same training set for all the SR methods except BCI. PittPatt SDK 5.0 was used for face recognition [16].

Experiment I (Synthetic Dataset): For the first set of experiments synthetic data were derived from a single HR image. The CMU MultiPIE [6] dataset was used for the this experiment. To create the training dataset we used data from 100 subjects from the first session, and data from 60 subjects from the second session were used to compute $\boldsymbol{f}_Y$ descriptors. Two different resolution datasets were created for the training dataset with IPD = 32 pixels (96x96 pixels) and IPD = 8 pixels (24x24 pixels). We generated LR images from session-four for testing, while images from session two were used as the gallery. Face recognition results of different SR methods and the original HR image are depicted in Fig. 5(a). Note that UH-DTSR has the highest recognition rates and consistently outperformed the other SR methods. In fact, the recognition results for UH-DTSR is almost as good as the original HR image, in terms of recognition rates. Visual outputs of the different SR methods are depicted in Fig. 6. Even though there are some artifacts in the face region (as seen in the image in the third column and last row of Fig. 6), note that the the overall quality of the image has improved.

Experiment II (Surveillance Video): For the second set of experiments, we used LR video sequences obtained from a surveillance system that uses a wide-field camera in an uncontrolled-lighting indoor environment. HR images of 32 subjects were obtained using a DSLR camera in a controlled environment and were used as gallery. Each video had 350-400 frames with a IPD about 8-12 pixels. PittPatt SDK [16] was used to detect faces. ASM may not give promising results for LR facial images, especially for real data captured from surveillance cameras. In this case, we manually annotated the eyes and the nose tip. FR results of different SR methods and the original LR image from surveillance dataset are depicted in Fig. 5(b). Figure 7 depicts images of a face detected in a surveilled environment, in a bounding box, as well as the output of UH-DTSR. These video sequences are very challenging, since they contain variations due to the motion and illumination. Note that in both synthetic and surveillance video experiments, we noticed that UH-DTSR reconstructs the eye regions more accurately than the mouth regions.



(a)



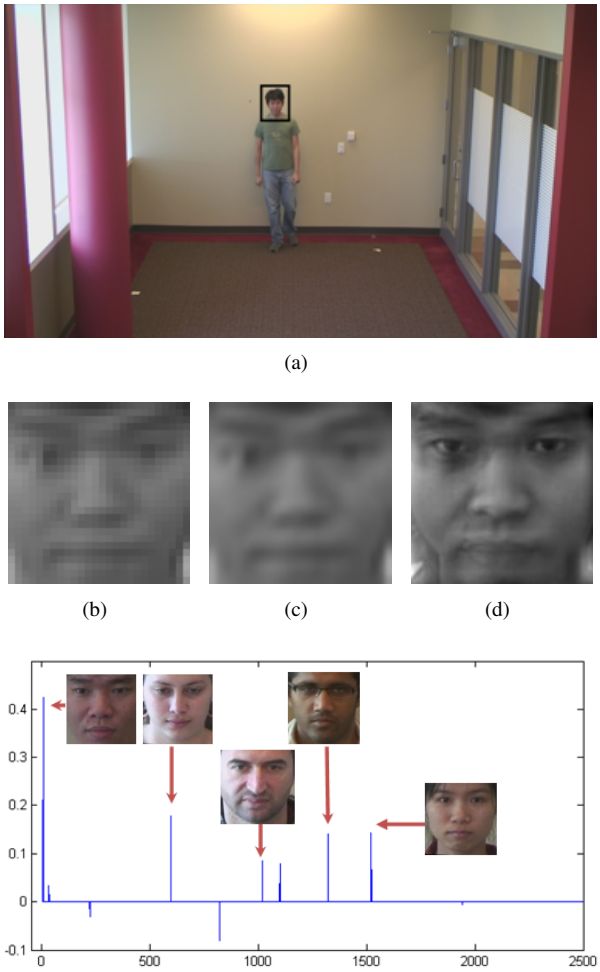
(b)

Figure 5. Rank-one recognition rates for: (a) Synthetic database: LR=17%, HR=87%, BCI=74%, S2R2=76%, SPRS=77% and UH-DTSR=85%. (b) Surveillance video: LR=7%, BCI=76%, S2R2=72%, SPRS=82% and UH-DTSR=91%.

5. Conclusions

We have proposed a method to reconstruct a SR image using DT-CWT and sparse representation, that quantitatively and qualitatively outperforms most of the recently proposed SR methods. Results are shown for synthetic and real datasets. In facial images, DT-CWT responds strongly

Figure 6. Output of the different SR methods. First row depicts the original images, the second row depicts the synthetically generated LR images, the third row depicts the output of S2R2 method, the fourth row depicts the output of the SPRS method, and the bottom row depicts output of UH-DTSR. For the subjects in the first and second columns, we had more than fifteen training images, while for the subject on the third (right) column, we did not have training images.



(a) (b) (c) (d) (e)

Figure 7. Illustration of the surveillance camera output and the SR output. (a) Depiction of a frame acquired by surveillance camera (the black bounding box indicates successful face detection), (b) magnification of the area in the bounding box (IPD $\approx$ 11 pixels), (c) output of BCI, (d) output of the UH-DTSR algorithm, and (e) sparse representation of the LR image.

to the eye regions and hence, the SR images of the eye-regions have better reconstruction when compared to the images of the mouth regions. We empirically demonstrated the advantages of UH-DTSR compared to several state-of-art super-resolution algorithms for the task of face recognition.

6. Acknowledgement

This research was funded in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL) and by the University of Houston (UH) Eckhard Pfeiffer Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, the U.S. Government, or UH.

References

[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. 1

[2] S. Baker and T. Kanade. Hallucinating faces. In *Proc. 4^{th} IEEE International Conference on Automatic Face and Gesture Recognition*, pages 83–88, Grenoble, France, March 28-30 2002. 2

[3] T. Cootes and C. Taylor. Active shape models: smart snakes. In *Proc. British Machine Vision Conference*, pages 266–275, Leeds, UK, Sep. 22-24 1992. 4

[4] A. Elayan, H. Ozkaramanli, and H. Demirel. Complex wavelet transform-based face recognition. *EURASIP Journal on Advances in Signal Processing*, 10(1):1–13, Jan 2008. 2

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

[6] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28:807–813, 2010. 5

[7] P. H. Hennings-Yeomans. *Simultaneous super-resolution and recognition*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2008. 2, 5

[8] J. T. Hsu, C. C. Yen, C. C. Li, M. Sun, B. Tian, and M. Kaygusuz. Application of wavelet-based POCS superresolution for cardiovascular MRI image enhancement. In *Proc.* 3^{rd} *International Conference on Image and Graphics*, pages 572 – 575, Hong Kong, China, Dec. 18-20 2004. 2

[9] K. Jia and S. Gong. Multi-modal tensor face for simultaneous super-resolution and recognition. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1683–1690, San Diego, CA, USA, Jun. 20-26 2005. 2

[10] Y. Li and X. Lin. An improved two-step approach to hallucinating faces. In *Proc.* 3^{rd} *International Conference on Image and Graphics*, pages 298–301, Hong Kong, China, Dec. 18-20 2004. 2

[11] C. Liu and D. Dai. Face recognition using dual tree complex wavelet features. *IEEE Transactions on Image Processing*, 18(11):2593–2599, 2009. 3

[12] C. Liu, H. Y. Shum, and C. S. Zhang. A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 192–198, San Diego, CA, USA, Jun. 20-26 2005. 2

[13] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transaction on Image Processing*, 17(1):53–69, Jan. 2008. 2

[14] A. Martinez and A. Kak. PCA versus LDA. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001. 3

[15] G. Pajares and J. M. Cruz. A wavelet based image fusion tutorial. *Pattern Recognition*, 37(9):1855–1872, Sep. 2004. 2

[16] Pittsburgh Pattern Recognition. PittPatt face tracking & Recognition Software Development Kit 5.1, 2011. 5

[17] M. Rogers and J. Graham. Robust active shape model search. In *Proc. European Conference on Computer Vision*, pages 517–530, London, UK, May 27-31 2002. 4

[18] J. Wang, S. Zhu, and Y. Gong. Resolution enhancement based on learning the sparse association of image patches. *Pattern Recognition Letters*, 31(1):1–10, Jan. 2010. 2

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

[20] R. Willet, I. Jermyn, R. Nowak, and J. Zerubia. Wavelet based super resolution in astronomy. In *Proc. Astronomical Data Analysis Software and Systems*, volume 314, pages 107–116, Strasbourg, France, 2003. 2

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

[22] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, Jun. 23-28 2008. 2

[23] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19:2861 – 2873, 2010. 5