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

Abstract

Face recognition is a challenging task, especially when low-resolution images or image sequences are used. A decrease in image resolution results in a loss of facial high frequency components leading to a decrease in recognition rates. In this paper, we propose a new method for super-resolution by building a dictionary of high-frequency components in the facial data, which are added to a low-resolution input image to create a super-resolved image. Our method is different from existing methods as we estimate the high-frequency components, 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 the widespread availability of video-based surveillance and monitoring systems. Many applications, ranging from security to broadcasting, are driving the need for better recognition of individuals from surveillance videos [1]. This requirement poses new challenges to the already difficult task of face recognition. 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 surrounding the subject's face, often characterized by a small interpupillary distance (IPD). This attribute usually leads to poor recognition rates. 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]. For many 2D face recognition (FR) systems, image resolution is the critical factor that affects its ability to detect key elements of facial anatomy (e.g., eyes, lip corners, facial contours) [19].

Many super-resolution (SR) approaches have attempted to minimize the mean-squared-error (MSE) or maximize the signal-to-noise ratio (SNR) between the original high-resolution
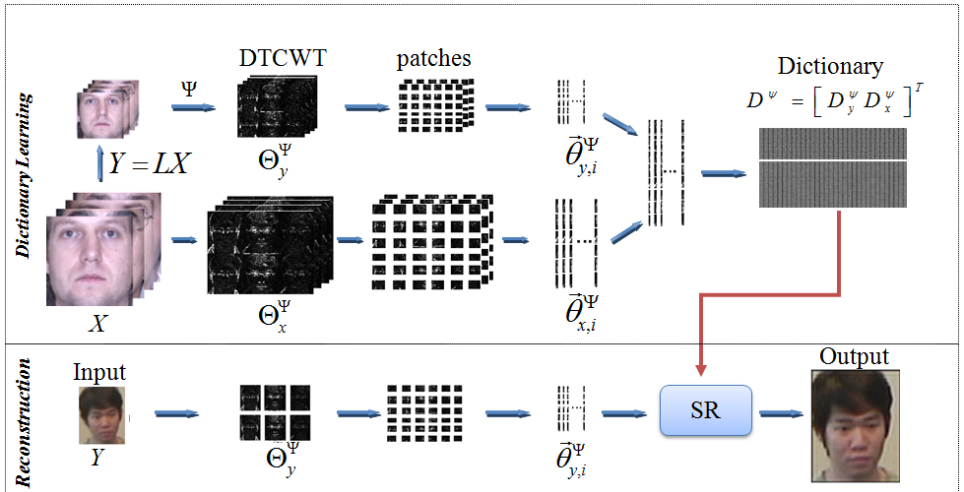
Figure 1: Depiction of the proposed framework for SR reconstruction. In the dictionary-learning stage, we extract the high-frequency components from a database of HR images and synthetically generated LR images, and build dictionaries for the LR and HR datasets (Sec. 3.3). In the reconstruction stage, the high-frequency components of the input LR image are used to compute a sparse representation with respect to the dictionary and use this representation to reconstruct a SR image (Sec. 3.4).

(HR) and reconstructed SR images [4]. However, in face recognition, these SR approaches may not perform well, as most FR systems rely on the ability to identify key facial features, typically captured 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 the HR facial images and then applying them to a given low-resolution (LR) image to create the super-resolved image (Fig. 1). In the training stage, we use a Dual Tree Complex Wavelet Transform (DT-CWT) [13] to extract the high-frequency components from a database of HR facial images and synthetically generate the LR images. A dictionary is built with the high-frequency components for each of the databases, both 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 an SR image. Instead of using the whole facial image, we divide it into patches which overlap to avoid the "block effect" artifacts during reconstruction.

Our contributions are: (i) we have introduced the use of DT-CWT in a sparse representation framework for SR reconstruction, and (ii) we have introduced a framework to reconstruct a SR facial image using high-frequency components (even if images of the subject are not available in the training set).

The rest of this paper is organized as follows: Section 2 describes previous work, while in Section 3 we provide a detailed description of the proposed method (UHSR). In Section 4, we evaluate the proposed method on facial databases using extensive experiments and

compare its performance with other state-of-the-art SR methods. Section 5 summarizes our findings.

2 Previous Work

In recent years, many methods have been proposed that address the issue of image resolution. Existing SR algorithms may be roughly classified into two classes: multi-frame-based and example-based algorithms. The methods in the first class attempt to compute a HR image from a set of LR images from any domain [4], while the methods in the second class attempt to find a HR counterpart of a single LR image from a known domain [2, 6, 12, 14, 18]. 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. Liu *et al*. [12] proposed a two step approach in which principal component analysis (PCA) was first used to obtain the global HR image and a patch-based non-parametric Markov network was then used to learn the residual image. Li *et al*. [11] 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 a 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 super-resolved image. Simultaneous SR and feature extraction for FR was proposed by Hennings [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 image, and its performance depends on the training dataset. Yang *et al*. [21] 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 [3], and astronomy [20]. In addition, it may be used to obtain a sparse representation of images that have smooth regions with isolated abrupt changes [15].

In our method, we propose to take advantage of the wavelet decomposition-based rules in conjunction with sparse representation and dictionary learning techniques to improve face recognition rates. Unlike the other methods [2, 12, 21], the performance of UHSR is optimized with respect to the FR task, rather than with respect to a visually plausible output or a lower mean-squared-error (MSE).

3 Methodology

3.1 Overview of Super-Resolution Reconstruction based on Multi-Resolution Decomposition

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 the additive i.i.d. Gaussian with zero mean noise. To estimate image X,

Equation 1 can be written as:

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

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

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

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

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

Next, we formulate the relationship between the LR- and HR-images using the high- and low-frequency components. Let $\Psi = [\Psi_{\beta} \ \Psi_{\theta}]$ be an operator that extracts a high- and low-frequency components ($\beta_X = \Psi_{\beta}X$ and $\theta_X = \Psi_{\theta}X$, respectively) of an image; i.e.,

$$[\beta_X \ \theta_X]^T = \Psi X \quad.$$

Assuming that Ψ is orthonormal, by applying inverse of Ψ on both sides, the original image, X, can be reconstructed. By splitting the high- and low-frequency components as:

$$\Psi^{-1}([\beta_X \ \mathbf{0}]^T + [\mathbf{0} \ \theta_X]^T) = X \quad, \tag{3}$$

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

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

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

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

where Y^* contains the low-frequency components of image X, 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

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

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

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

In Eq. 7, G can be estimated using the training set of facial images.

3.2   Complex Wavelet Transform for Super-Resolution

The Dual-Tree Complex Wavelet Transform (DT-CWT) has been extensively used for face recognition [13]. In our approach, we use the DT-CWT to separate an image into low- and high-frequency components. To compute DT-CWT, two wavelet trees are developed in

---

[1]In our experiments on 1,000 randomly selected facial images from the CMU-MultiPIE dataset, we computed the average auto-correlation between Y^* and $\tilde{Y}$ to be equal to 0.988.
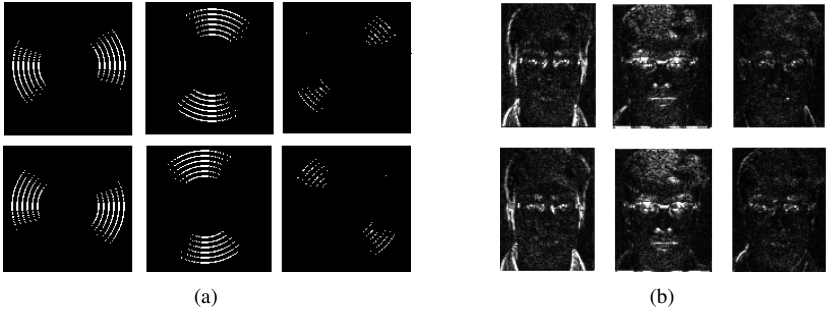
Figure 2: Depiction of DT-CWT. (a) The six images are the detail coefficients of the DT-CWT transformation of a synthetically generated binary image with five circles of different radii. (b) The coefficients of the DT-CWT transformation of the facial image. Grayscale intensity values are used to represent the magnitude of the complex wavelet coefficients.

parallel, with the wavelets forming Hilbert pairs. The trees represent the real and imaginary parts of the wavelet transform. Due to the design of the filter banks that incorporates the half sample delay, DT-CWT is nearly shift invariant, due to its slight translated filter designs. Its other main property is directional selectivity (Fig. 2). Let us denote Ψ as the DT-CWT transformation matrix and Θ as the coefficients of DT-CWT. Then, the DT-CWT of an image X is represented as:

$$\Theta_h = \Psi X \quad \text{and} \quad \Theta_l = \Psi L X, \tag{8}$$

where L is the decimation matrix that resizes the training image X, to the size of image Y. Since Θ includes both the low- and high-frequency components of an image, we denote

$$\Theta = [\Theta^\phi \ \Theta^\psi] \quad \text{and} \quad \Theta^\psi = \begin{bmatrix} \Theta^{+15} & \Theta^{+45} & \Theta^{+75} \\ \Theta^{-15} & \Theta^{-45} & \Theta^{-75} \end{bmatrix}, \tag{9}$$

where $\Theta^{\pm15}, \Theta^{\pm45}, \Theta^{\pm75}$ represent the high-frequency components corresponding to the specific filters at the respective angles ($\pm15°, \pm45°, \pm75°$), and Θ^ϕ are the low frequency components. The result of applying the DT-CWT on sample images is depicted in Fig. 2. It can be observed that each wavelet subband responds strongly to a specific orientation, and 2D DT-CWT can discriminate between positive and negative orientations, which is one of the important features that makes it suitable for FR.

Let $\{x_1, \ldots, x_n\} \in X$ be a set of n overlapping square patches defined in image X, and $\{y_1, \ldots, y_n\} \in Y$ the set of corresponding patches in Y. The wavelet coefficients in Θ are arranged in blocks according to their spatial correspondence in the original image. Let us denote the coefficients associated with x_i and y_i as $\vec{\theta}_{x,i}^{\psi}$ and $\vec{\theta}_{y,i}^{\psi}$, respectively. Each such structure includes six high-frequency components, corresponding to the six different types of filters (Eq. 9). We exclude coefficients in Θ^ϕ from any further analysis, since they carry almost no additional information needed for SR reconstruction. The elements of $\vec{\theta}_{x,i}^{\psi}$ and $\vec{\theta}_{y,i}^{\psi}$ are concatenated in column vectors denoted as, $\vec{\theta}_i^{\psi} = [\vec{\theta}_{x,i}^{\psi} \ \vec{\theta}_{y,i}^{\psi}]^T$, which contain high-frequency components of the LR and HR images. Our next step is to learn a dictionary for an efficient sparse representation of the patch descriptors $\vec{\theta}_i^{\psi}$.

3.3 Learning a Coupled Dictionary

Using $\vec{\theta}_i^{\psi}$, where $\{\vec{\theta}_1^{\psi},\ldots,\vec{\theta}_n^{\psi}\} \in \vec{\Theta}^{\psi}$, we build a dictionary D^{ψ}, which assures an results in sparse-reconstruction of the images in the training set. The size of the dictionary will be very big if all the patches are used to build it, and will be computationally expensive for reconstruction. To reduce the computation time, Lee *et al.* [10] proposed a method to build compact dictionaries. Note that there is a trade off between the computation time for the reconstruction, the dictionary size, and the reconstruction error.

Specifically, the dictionary is learnt from a paired input vector, $\vec{\theta}_i^{\psi}$, and should satisfy the following condition:

$$D^{\psi} = \underset{D^{\psi},\vec{\alpha}}{\arg\min} \left\| \vec{\theta}_i^{\psi} - D^{\psi}\alpha_i \right\|_2 + \lambda \left\| \alpha_i \right\|_1 \text{ s.t. } \left\| D_j^{\psi} \right\|_2 = 1, \quad i = 1,2..n, \;\; j = 1,2..K \quad (10)$$

where K is the number of atoms in dictionary D^{ψ} (K is typically much smaller than the total number of patches), α_i is the representation of $\vec{\theta}_i^{\psi}$, $\|.\|_1$ is an l_1 norm that enforces sparsity, and l_2 normalizes the atoms in the dictionary. The dictionary D^{ψ} and coefficients α_i cannot be simultaneously estimated using convex optimization methods. However, a solution can be found if either D^{ψ} or α_i is fixed. This is accomplished by estimating D^{ψ} iteratively as presented in Algorithm 1.

Algorithm 1 Dictionary Learning

Input: $\vec{\Theta}^{\psi} = \{\vec{\theta}_1^{\psi},\ldots,\vec{\theta}_n^{\psi}\}$
Output: D^{ψ}
 1: Init D^{ψ} using a Random Matrix
 2: **for all** α_i, $i = \{1,\ldots,n\}$ **do**
 3:

$$\alpha_i = \underset{\alpha_i}{\arg\min} \left\| \vec{\theta}_i^{\psi} - D^{\psi}\alpha_i \right\|_2 + \lambda \left\| \alpha_i \right\|_1 \quad (11)$$

 4: **end for**
 5:

$$D^{\psi} = \underset{D^{\psi},\alpha_i}{\arg\min} \left\| \vec{\Theta}^{\psi} - D^{\psi}\alpha_i \right\|_2 \quad \text{s.t } \left\| D_j^{\psi} \right\|_2 \leq 1, \;\; i = 1,2..,n, \;\; j = 1,2,..K \quad (12)$$

 6: Repeat steps 2 and 5 until convergence.

The vector $\vec{\theta}_i^{\psi}$ is a paired input (i.e., $\vec{\theta}_i^{\psi} = [\vec{\theta}_{x,i}^{\psi} \; \vec{\theta}_{y,i}^{\psi}]^T$). Thus, the output $D^{\psi} = [D_x^{\psi} \; D_y^{\psi}]^T$ is also paired.

Let D_h^{ψ} be an overcomplete dictionary that contains the high-frequency components of X. Then, $\theta_{x,i}^{\psi}$ can be represented as a sparse and linear combination of the atoms in dictionary D_x^{ψ}. Specifically,

$$\hat{\theta}_{x,i}^{\psi} = D_x^{\psi}\alpha_i, \quad (13)$$

where α is a vector of coefficients with very few non-zero entries. Similarly, $\theta_{y,i}^{\psi}$ can be represented as a sparse linear combination of atoms in $D_{y,i}^{\psi}$ as follows:

$$\hat{\theta}_{y,i}^{\psi} = D_y^{\psi}\alpha_i. \quad (14)$$

Two different dictionaries are used to reconstruct the high-frequency components of SR patch, $\hat{\theta}_{h,i}^{\psi}$, where $\{\hat{\theta}_{x,1}^{\psi}, \ldots, \hat{\theta}_{x,n}^{\psi}\} \in \hat{\Theta}_x^{\psi}$. The resulting dictionary contains both the dictionaries, D_x^{ψ} for HR images and D_y^{ψ} for LR images and hence, is coupled. This coupling, which enforces the relationship between high- and low-resolution patches, x_i and y_i, is used in the reconstruction step, as explained in the following section.

3.4 Super-resolution reconstruction via Sparse Representation

We propose a method that is independent of the relationship between the HR and LR images, and is nearly shift invariant due to the shift-invariance in feature extraction. Similar to patch-based methods, our method also reconstructs super-resolved patches from selected LR patches and a global method is used to remove the artifacts and discontinuities.

Given the high-frequency component of the patch descriptor, $\vec{\theta}_{y,i}^{\psi}$, its sparse-representation, α_i, is obtained by minimizing

$$\min \|\alpha_i\|_1 \quad \text{s.t} \quad \left\| \vec{\theta}_{y,i}^{\psi} - D_y^{\psi} \alpha \right\|_2 \leq \varepsilon_0 \;, \tag{15}$$

where $\varepsilon_0 > 0$ is a small value, which represents trade-off between reconstruction quality and sparsity. Reconstructing SR patches individually, using Eq. 15, will introduce discontinuity artifacts between neighboring patches. To overcome this problem and to enforce the compatibility in the neighboring patches, we use an approach similar to the one proposed by Yang *et al*. [22]:

$$\min \|\alpha_i\|_1 \quad \text{s.t.} \quad \left\| \vec{\theta}_{y,i}^{\psi} - D_y^{\psi} \alpha_i \right\|_2 \leq \varepsilon_0 \;\; \text{and} \;\; \left\| \Omega \hat{\theta}_{x,i}^{\psi} - \Omega D_x^{\psi} \alpha_i \right\|_2 \leq \varepsilon_1, \tag{16}$$

where the operator Ω extracts the regions of overlap between a previously reconstructed SR patch and the current patch. Using a Lagrange multiplier, Eq. 16 can be written as:

$$\alpha_i^* = \arg\min_{\alpha_i} \left\| \vec{\theta}_{y,i}^{\psi} - D_y^{\psi} \alpha_i \right\|_2 + \left\| \Omega \vec{\theta}_{x,i}^{\psi} - \Omega D_x^{\psi} \alpha_i \right\|_2 + \lambda \|\alpha_i\|_1 \tag{17}$$

The SR patch can be recovered as:

$$\hat{\theta}_{x,i}^{\psi} = D_x^{\psi} \alpha_i^* \\ \hat{X} = \Psi^{-1}(\Psi L^{-1} Y + \hat{\Theta}^{\psi}), \;\; \hat{\Theta}^{\psi} = \{\hat{\theta}_{x,1}^{\psi} \ldots \hat{\theta}_{x,n}^{\psi}\} \tag{18}$$

where L^{-1} is the upsampling operator.

4 Experimental Results

In order to evaluate the performance of the proposed algorithm, two experiments were performed. We compared the proposed DT-CWT-based SR method (UHSR) with the following SR algorithms: bicubic interpolation (BCI), Simultaneous Super Resolution and Recognition (S2R2) [6], and SR via Sparse Representation (SRSR) [22].

For the first set of experiments synthetic data were derived from a single HR image. For the second test we used a LR video sequence obtained from a surveillance camera installed in an indoor facility. For face recognition, the PittPatt SDK [16] was used. In both experiments, we have used only nearly-frontal facial images.
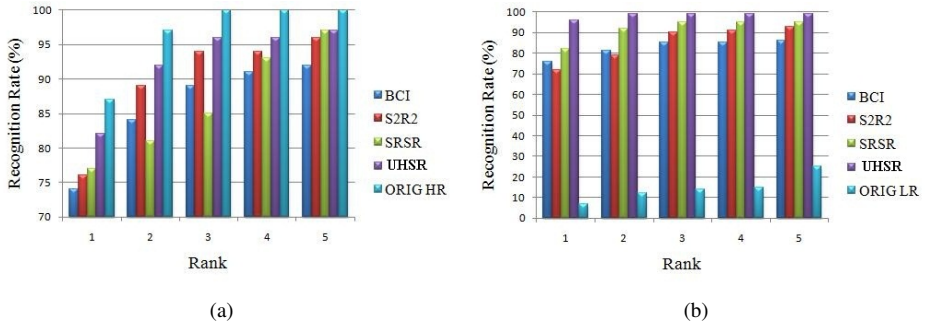
(a)   (b)

Figure 3:   a) The rank-one recognition rates for different SR methods for the synthetic dataset are: 74% for BCI, 76% for S2R2, 77% for SRSR, 82% for UHSR, and 87% for the original HR image. b) The rank-one recognition rates for the surveillance dataset are: 76% for BCI, 72% for S2R2, 82% for SRSR, 96% for UHSR, and only 7% for the original LR image.

Experimental Settings: We leverage the Active Shape Model (ASM) [17] methodology to detect key anatomical landmarks in each image. ASM may not provide promising results for LR facial images, especially for real data captured from surveillance cameras. In this case, we manually annotate the two eye centers. Using the landmark coordinates, we register all images to a reference image. We used the IPD to define the resolution of the facial image. The target facial region size of the low resolution images is set to be 24×24 pixels, and the size for HR image is set to 96×96 pixels, which corresponds to IPD ≈ 8 pixels and IPD ≈ 32 pixels, respectively.

Experiment 1 - Synthetic LR images: The first set of experiments involved the use of synthetic LR data derived from a HR image. The CMU MultiPIE [5] database was used for the this experiment. To create the training dataset we used data from 100 subjects from the first session. Two different resolution datasets were created for the training dataset with IPD = 32 pixels and IPD = 8 pixels. The LR images that were generated from session four were used for testing while the images from second session were used as a gallery. We compared the recognition rates of BCI, S2R2, SRSR, and UHSR (Fig. 3(a)). The proposed method has the highest recognition rate and consistently outperformed the other SR methods (Fig. 3(a)). Rank-r recognition rate in Fig. 3(a) stands for the probability of the correct subject to be present among r most successful candidates.

Experiment 2 - Recognition at a distance from 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 13 subjects were obtained using a DSLR camera in a controlled environment and were used as gallery. From each video, we selected six frames which were furthest from the surveillance camera. Figure 3(b) depicts the recognition rates of SR algorithms. The proposed method exhibited the highest recognition rate and outperformed the other methods in our evaluation. Figure 4 shows one frame from the surveilled environment and the bounding box denotes the detected face using PittPatt SDK. These video sequences are very challenging, due to the presence of motion blur and illumination challenges (Fig. 4(a)). In addition, the training database contains images obtained at different camera settings, which caused a severe drop in the performance of S2R2 and SRSR. Therefore, we did not include their results in Fig. 4(d).
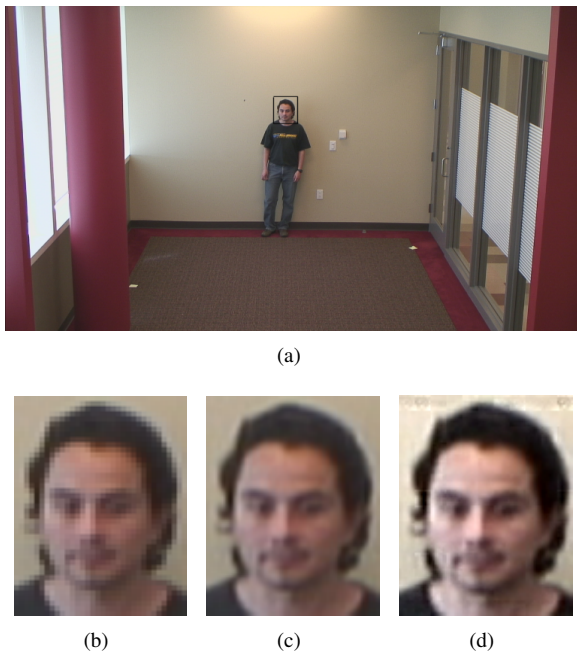
Figure 4: Illustration of the surveillance camera output and the SR output. (a) Depiction of a frame acquired by surveillance camera (the black bounding box indicates successful face detection), (b) magnification of the area in the bounding box (IPD $\approx$ 11 pixels), (c) output of BCI, and (d) output of the UHSR algorithm.

.

Figure 4(b) shows the detected face, and Figs. 4(c,d) depict the output of the BCI, and the UHSR, respectively. For better visualization, we converted the input RGB image into a YCbCr image, and used the luminance component to create the SR image using UHSR. After reconstruction we added the chrominance components back to the SR image. Thus, in Fig. 4(d), there is an illumination difference between BCI and UHSR.

5   Conclusions

This paper introduces a novel method for the super-resolution reconstruction of face images. The proposed framework combines together two powerful tools: wavelet decomposition and sparse representations via dictionary learning. The coupled dictionary enforces the relationships between the low- and the high-frequency components of the patch descriptors, and the wavelet decomposition allows the DTSR to be selective in determining the frequencies and orientations, thereby optimizing the FR task, specifically. The proposed algorithm can provide a qualitative output even when the test image is affected by the illumination and has blur effects due to poor focus or optics. We empirically demonstrated the advantage of the proposed method compared to several state-of-art super-resolution algorithms for the task of face recognition.

6 Acknowledgement

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.

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

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

[4] 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.

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

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

[7] P.H. Hennings-Yeomans, S. Baker, and V. Bhagavatula. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, Jun. 24-26 2008.

[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. 3rd International Conference on Image and Graphics*, pages 572 – 575, Hong Kong, China, Dec. 18-20 2004.

[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.

[10] H. Lee, A. Battle, R. Raina, and A. Ng. Efficient sparse coding algorithms. *Neural Information Processing Systems*, pages 801–808, 2006.

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

[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, June 20-26 2005.

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

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

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

[16] Pittsburgh Pattern Recognition. PittPatt face tracking & Recognition Software Development Kit 5.1, 2011. URL http://www.pittpatt.com/products/ftr_sdk/.

[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.

[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.

[19] 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.

[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.

[21] 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.

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