SINGLE FACE IMAGE SUPER-RESOLUTION VIA SOLO DICTIONARY LEARNING

Felix Juefei-Xu and Marios Savvides

Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

ABSTRACT

In this work, we have proposed a single face image super-resolution approach based on solo dictionary learning. The core idea of the proposed method is to recast the super-resolution task as a missing pixel problem, where the low-resolution image is considered as its high-resolution counterpart with many pixels missing in a structured manner. A single dictionary is therefore sufficient for recovering the super-resolved image by filling the missing pixels. In order to fill in 93.75% of the missing pixels when super-resolving a 16×16 low-resolution image to a 64×64 one, we adopt a whole image-based solo dictionary learning scheme. The proposed procedure can be easily extended to low-resolution input images with arbitrary dimensions, as well as high-resolution recovery images of arbitrary dimensions. Also, for a fixed desired super-resolution dimension, there is no need to retrain the dictionary when the input low-resolution image has arbitrary zooming factors. Based on a large-scale fidelity experiment on the FRGC ver2 database, our proposed method has outperformed other well established interpolation methods as well as the coupled dictionary learning approach.

***Index Terms*—** Super-Resolution, Dictionary Learning, Missing Pixel Problem

1. INTRODUCTION

Image super-resolutions aims to recover the high-resolution representation of a low-resolution input image. There are three major thrusts in tackling this problem in the literature. First, methods that are based on self similarities. Through searching similar patches within the image itself across different scales, the algorithm gives rise to the super-resolution such as in [1, 2]. Second, methods that learn a mapping function from low-resolution images to high-resolution ones such as [3, 4]. One recent paper by Dong et al. [5] is worth mentioning. The authors learn the mapping between low- and high-resolution images through a deep convolutional neural network (CNN). With reasonable parameter tuning and days of training, the model yields good performance. This is the very first attempt to use deep learning tools for the super-resolution task. Third, methods based on over-complete dictionary learning and sparse representation represented by the seminal work from Yang et al. [6] and others [7, 8, 9, 10]. These methods initiate a coupled dictionary learning paradigm where the low-resolution dictionary and high-resolution dictionary are jointly learned. By enforcing the sparse coefficients to be shared between the low-resolution query image and the high-resolution recovered image, the coupled dictionary methods allow an implicit mapping between the low- and high-resolution images. However, all the aforementioned dictionary learning approaches require learning a dictionary pair. For a low-resolution dictionary whose atom is of dimension d_L and a high-resolution dictionary whose atom is of dimension d_H, one of the mostly widely adopted approaches is to concatenate the training instances of dimension $d_L + d_H$ to learn one joint dictionary, where each dictionary atom is therefore of dimension $d_L + d_H$, and then to split the joint dictionary into low- and high-resolution dictionaries. Common dictionary learning approach includes K-SVD [11]. K-SVD is a greedy iterative ℓ_0-based method whose computation is greatly affected by the dimension of the dictionary atom and the number of atoms to be learned in the dictionary. An over-complete dictionary whose atoms are of dimension d means that there are $N \gg d$ atoms to be learned.

In this work, we propose to achieve single image super-resolution through a solo dictionary learning scheme such that there is no need to retrain a new coupled dictionary whenever the zooming factor has changed. This is made possible by recasting the super-resolution task as a missing pixel problem where the low-resolution image is considered as the high-resolution counterpart with many pixels missing, in a structured manner. Figure 1 depicts the flow chart of the proposed super-resolution method based on solo dictionary learning.

2. PROPOSED METHOD

In this section, we detail our proposed pipeline for super-resolution via solo dictionary learning.

In our setting, the input low-resolution image is of size 16×16 as shown in the second column in Figure 1. The desired super-resolved output is of size 64×64, and therefore, the magnification factor is 4, yielding a $16\times$ larger image. The proposed method can, of course, deal with other input dimensions and desired output dimensions. However, for the sake of brevity, we will stick to super-resolving a 16×16 image to a 64×64 one throughout the rest of the paper.

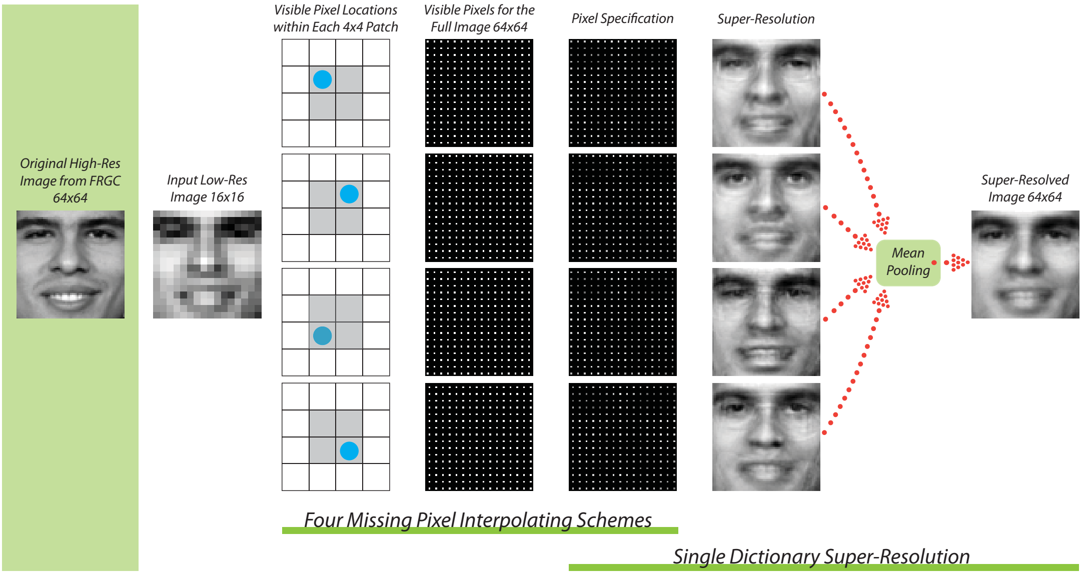
Original High-Res Image from FRGC 64x64
Input Low-Res Image 16x16
Visible Pixel Locations within Each 4x4 Patch
Visible Pixels for the Full Image 64x64
Pixel Specification
Super-Resolution
Mean Pooling
Super-Resolved Image 64x64
Four Missing Pixel Interpolating Schemes
Single Dictionary Super-Resolution

Fig. 1. Flow chart of the proposed single face image super-resolution through solo dictionary learning. The core idea of the proposed method is that, we recast the super-resolution task as a missing pixel problem, where the low-resolution image is considered as its high-resolution counterpart with many pixels missing in a structured manner.

The core idea of the proposed method is that, we recast the super-resolution task as a missing pixel problem, where the low-resolution image is considered as its high-resolution counterpart with many pixels missing in a structured manner. In other words, we want to evenly blow up the low-resolution image by interpolating with zero-valued pixels.

By doing so, we can see that for each of the 4×4 sub-image of the blown-up image, one of the 16 pixels is represented by the spatially corresponding pixel in the low-resolution image, and the remaining 15 pixels in the 4×4 sub-image are considered as missing pixels. Apparently, there are 16 possible ways to land the pixels from the low-resolution image to the interpolated one. In our approach, we choose to allow four possible locations for the low-resolution image pixels to land in each 4×4 sub-image, as indicated by blue dots in Figure 1.

After we have specified all the locations where the pixels from the low-resolution image should be in the interpolated image, we can specify the pixel intensities accordingly. That way, we can obtain four interpolated 64×64 images with many missing pixels, as a result of four choices of visible pixel locations within each 4×4 sub-image. In this case, 93.75% of the pixels are missing in each interpolated image. Our goal is to fill these missing pixels and recover the super-resolved image via a single dictionary. The final output is a mean-pooling of the four recovery results to account for small artifact caused by the shifts during pixel specification.

The solo dictionary learning is carried out in a standard way using K-SVD [11], with the same dimension as the high-resolution images which is $4,096 = 64 \times 64$. For the sake of over-completeness, our trained dictionary is of size $4,096 \times 12,000$. Some atoms of the pre-trained solo dictionary are shown in Figure 2.

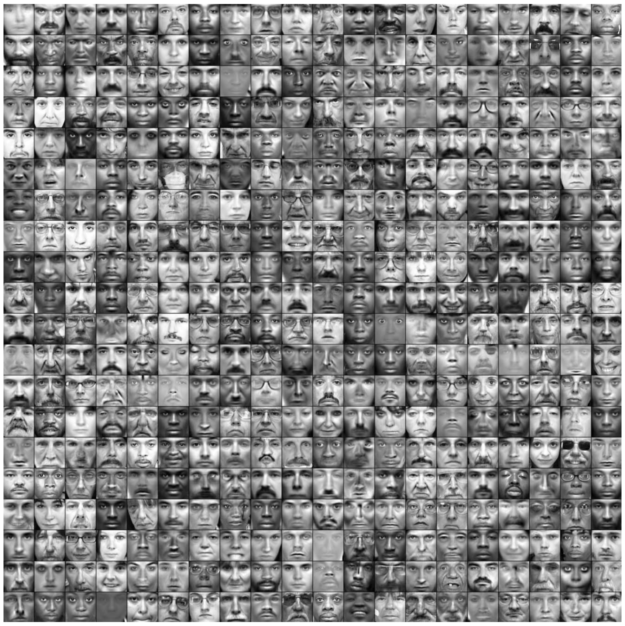

Fig. 2. 400 out of 12,000 atoms of the trained solo dictionary.

Let $\mathbf{D} \in \mathbb{R}^{d_H \times N}$ be the trained dictionary, let $\mathbf{y}_L \in \mathbb{R}^{d_L}$ be the low-resolution image, and $\mathbf{y}_H \in \mathbb{R}^{d_H}$ be the super-resolve image. The missing pixel interpolation is obtained by the mapping f. Also, we make use of the interpolation mask $\mathbf{m} \in \mathbb{R}^{d_H}$ to zero out certain dimensions (rows) of the dictionary $\mathbf{D}$ corresponding to the missing pixel locations and obtain a masked dictionary $\mathbf{D_m}$. Using Matlab notation, the masked dictionary is obtained by: $\mathbf{D_m} = \text{diag}\left(\mathbf{m}\right) * \mathbf{D}$.

Next, we use the masked dictionary for sparse coding the interpolated image $f(\mathbf{y}_L)$ as follows:

$$\mathbf{x} = \arg\min_{\mathbf{x}} \|f(\mathbf{y}_L) - \mathbf{D_m}\mathbf{x}\|_2^2 \text{ subject to } \|\mathbf{x}\|_0 \le \kappa \quad (1)$$

where the sparsity level is captured by parameter κ. This can be done by any pursuit algorithm such as the orthogonal matching pursuit (OMP) [12]. Here the coefficient vector $\mathbf{x}$ is then used for obtaining the super-resolved image $\mathbf{y}_H$ by projecting onto the original dictionary $\mathbf{D}$ as follows:

$$\mathbf{y}_H = \mathbf{D}\mathbf{x} \quad (2)$$

Following this, we are able to recover the high-resolution image using only one dictionary, preventing retraining a new coupled dictionary when the zooming factor changes.

3. EXPERIMENTS

3.1. Database

The database we use for the following reconstruction fidelity experiments is a large-scale NIST's FRGC ver2 database [13] which has the following three components: First, the generic **training** set contains both controlled and uncontrolled images of 222 subjects, and a total of $12,776$ images. Second, the **target** set represents the people that we want to find. It has 466 different subjects, a total of $16,028$ images. Last, the **probe** set represents the unknown images that we need to match against the **target** set. It contains the same 466 subjects as in target set, with half as many images for each person, bringing the total number of probe images to $8,014$. In the following experiments, we will use the entire **target** set with $16,028$ images. FRGC database has been widely used in unconstrained face recognition tasks [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. In addition, we use a separate large-scale mugshot type database for pre-training the solo dictionary to be used in our super-resolution recovery.

3.2. Experimental Setup and Results

Each image in the FRGC target set is registered by the eye location and cropped into a square face of size 64×64, which are served as the ground truth images to be compared with the super-resolved images. Examples of original face images are shown in the first column in Figure 4. Next, the original image is Gaussian blurred and down-sampled to the size of 16×16. The goal of the super-resolution task is to recover a 64×64 high-resolution representation of the low-resolution image, yielding a $4\times$ magnification, or equivalently, a $16\times$ larger image. The success of an super-resolution algorithm is judged by how close the super-resolved image is to the original high-resolution one. The procedure can be easily extended to low-resolution input images with arbitrary dimensions, as well as high-resolution recovery images of arbitrary dimensions. The only need to mask out certain dimensions in the trained dictionary during super-resolution as previously discussed. Also, for a fixed desired super-resolution dimension, there is no need to retrain the dictionary when the input low-resolution image has arbitrary zooming factors.

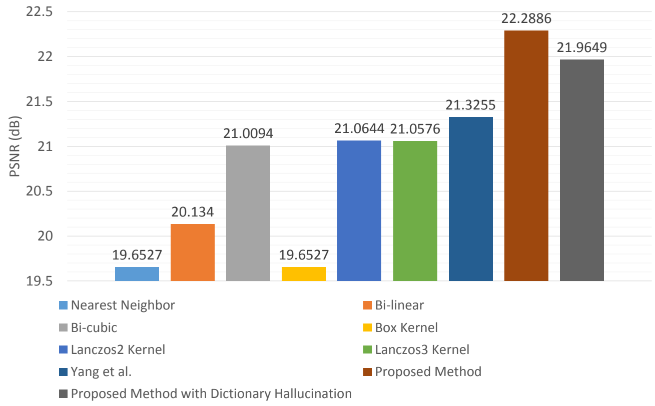

Fig. 3. Average PSNR on FRGC target set for various super-resolution methods.

Here, we adopt the broadly used peak signal-to-noise ratio (PSNR) as the fidelity measurement [27, 28, 29]. We benchmark our proposed method against some well established interpolation techniques for super-resolution including nearest neighbor interpolation, bi-linear interpolation, bi-cubic interpolation, interpolation using box kernel, interpolation using Lanczos2 and Lanczos3 kernels, as well as the work of Yang et al. [6] using coupled dictionary learning. We report the average PSNR on the FRGC target set ($16,028$ images) for each of the algorithms in Figure 3. In addition, some visual results of the proposed method along with other competing methods are shown in Figure 4.

3.3. Discussion

It can be observed that our proposed super-resolution method yields the highest average PSNR compared to 7 other methods on this particular dataset. Also, from Figure 4, we can see that the super-resolved images using our proposed method (last column) are also of high fidelity with the original ones (first column). Linear interpolating methods using various kernels tend to produce overly smooth super-resolution results, while ℓ_0-based dictionary learning methods like Yang et al. [6] and our proposed method are able to reconstruct detailed facial features. The advantage over Yang et al. [6] is that only solo dictionary learning is required in our method, while obtaining better results under our experimental settings. There is a related work by Mu et al. [30] which is worth noticing.

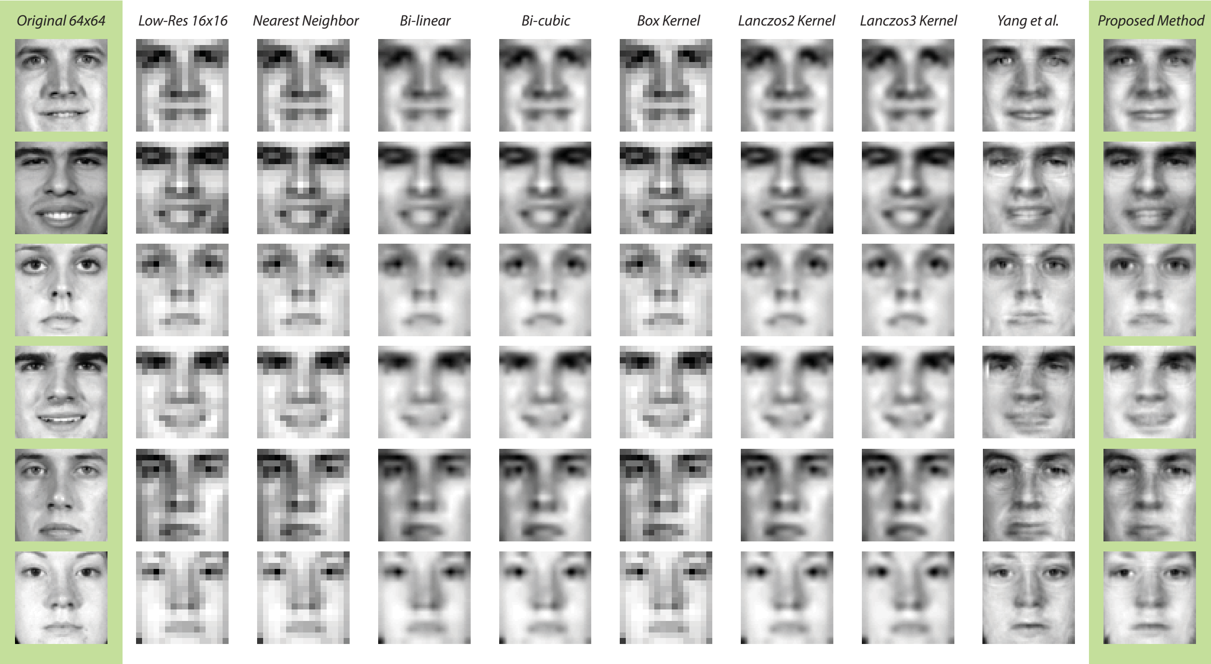

Fig. 4. Visual results for face super-resolution using various algorithms.

The authors also proposed a super-resolution method based on a single high-resolution dictionary, however, their work is fundamentally different from ours. [30] is a patch-based dictionary approach which relies on energy function to model local similarities and the recovery needs an iterative fusion step. The gist is that one pixel in the image can be estimated by a weighted average of the similarity pixels in the same image. While our method is a whole image-based dictionary learning approach, and by recasting the super-resolution task as a missing pixel problem, our method can super-resolve the image in one shot, with high fidelity.

3.4. Dictionary Hallucination

In the cases when high-resolution dictionary or high-resolution dictionary training images are not available, we propose to hallucinate the high-resolution dictionary directly from the low-resolution one. Our experiments have shown that even with the simplest bi-cubic interpolation, the hallucinated high-resolution dictionary can be used for the same aforementioned super-resolution task with the solo hallucinated dictionary. The performance is almost as good as directly training the high-resolution dictionary as shown in Figure 3.

4. CONCLUSION

In this work, we have proposed a single face image super-resolution approach based on solo dictionary learning. The core idea of the proposed method is to recast the super-resolution task as a missing pixel problem, where the low-resolution image is considered as its high-resolution counterpart with many pixels missing in a structured manner. A single dictionary is therefore sufficient for recovering the super-resolved image by filling the missing pixels. In order to fill in 93.75% of the missing pixels when super-resolving a 16 × 16 low-resolution image to a 64 × 64 one, we adopt a whole image-based solo dictionary learning scheme. The proposed procedure can be easily extended to low-resolution input images with arbitrary dimensions, as well as high-resolution recovery images of arbitrary dimensions. Also, for a fixed desired super-resolution dimension, there is no need to retrain the dictionary when the input low-resolution image has arbitrary zooming factors. Based on a large-scale fidelity experiment on the FRGC ver2 database, our proposed method has outperformed other well established interpolation methods as well as the coupled dictionary learning approach.

5. REFERENCES

[1] D. Glasner, S. Bagon, and M. Irani, “Super-resolution from a single image,” in *ICCV*, Sept 2009, pp. 349–356.

[2] G. Freedman and R. Fattal, “Image and video upscaling from local self-examples,” *ACM Trans. Graph.*, vol. 30, no. 2, pp. 12:1–12:11, Apr. 2011.

[3] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Morel, “Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding,” in *BMVC*, 2012.

[4] H. Chang, D. Yeung, and Y. Xiong, "Super-Resolution through Neighbor Embedding," in *CVPR*, June 2004.

[5] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*, Sept 2014, pp. 184–199.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *TIP*, vol. 19, no. 11, pp. 2861–2873, Nov 2010.

[7] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *TIP*, vol. 21, no. 8, pp. 3467–3478, Aug 2012.

[8] H. Huang, J. Yu, and W. Sun, "Super-resolution mapping via multi-dictionary based sparse representation," in *ICASSP*, May 2014, pp. 3523–3527.

[9] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *CVPR*, June 2013, pp. 345–352.

[10] R. Walha, F. Drira, F. Lebourgeois, C. Garcia, and A.M. Alimi, "Sparse coding with a coupled dictionary learning approach for textual image super-resolution," in *ICPR*, Aug 2014, pp. 4459–4464.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *TSP*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[12] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Application to Wavelet Decomposition," in *Asilomar Conf. on Signals, Systems and Comput.*, 1993.

[13] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek, "Overview of the face recognition grand challenge," in *CVPR*, jun 2005, vol. 1, pp. 947–954.

[14] F. Juefei-Xu and M. Savvides, "Pareto-optimal Discriminant Analysis," in *ICIP*, Sept 2015.

[15] N. Zehngut, F. Juefei-Xu, R. Bardia, D. K. Pal, C. Bhagavatula, and M. Savvides, "Investigating the Feasibility of Image-Based Nose Biometrics," in *ICIP*, Sept 2015.

[16] F. Juefei-Xu and M. Savvides, "Weight-Optimal Local Binary Patterns," in *ECCVW*, 2015, pp. 148–159.

[17] F. Juefei-Xu and M. Savvides, "Subspace Based Discrete Transform Encoded Local Binary Patterns Representations for Robust Periocular Matching on NIST's Face Recognition Grand Challenge," *TIP*, vol. 23, no. 8, pp. 3490–3505, aug 2014.

[18] F. Juefei-Xu and M. Savvides, "An Image Statistics Approach towards Efficient and Robust Refinement for Landmarks on Facial Boundary," in *BTAS*, Sept 2013.

[19] F. Juefei-Xu and M. Savvides, "An Augmented Linear Discriminant Analysis Approach for Identifying Identical Twins with the Aid of Facial Asymmetry Features," in *CVPRW*, June 2013, pp. 56–63.

[20] F. Juefei-Xu and M. Savvides, "Unconstrained Periocular Biometric Acquisition and Recognition Using COTS PTZ Camera for Uncooperative and Non-cooperative Subjects," in *WACV*, Jan 2012, pp. 201–208.

[21] F. Juefei-Xu, K. Luu, M. Savvides, T. D. Bui, and C. Y. Suen, "Investigating Age Invariant Face Recognition Based on Periocular Biometrics," in *IJCB*, Oct 2011, pp. 1–7.

[22] F. Juefei-Xu and M. Savvides, "Can Your Eyebrows Tell Me Who You Are?," in *ICSPCS*, Dec 2011, pp. 1–8.

[23] F. Juefei-Xu, M. Cha, M. Savvides, S. Bedros, and J. Trojanova, "Robust Periocular Biometric Recognition Using Multi-level Fusion of Various Local Feature Extraction Techniques," in *DSP*, 2011.

[24] F. Juefei-Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust Local Binary Pattern Feature Sets for Periocular Biometric Identification," in *BTAS*, sep 2010, pp. 1–8.

[25] F. Juefei-Xu and M. Savvides, "Image Matching Using Subspace-Based Discrete Transform Encoded Local Binary Patterns," Sept. 2013, US Patent 2014/0212044.

[26] F. Juefei-Xu and M. Savvides, "Facial Ethnic Appearance Synthesis," in *ECCVW*, 2015, pp. 825–840.

[27] F. Juefei-Xu and M. Savvides, "Encoding and Decoding Local Binary Patterns for Harsh Face Illumination Normalization," in *ICIP*, Sept 2015.

[28] F. Juefei-Xu, D. K. Pal, and M. Savvides, "NIR-VIS Heterogeneous Face Recognition via Cross-Spectral Joint Dictionary Learning and Reconstruction," in *CVPRW*, June 2015.

[29] F. Juefei-Xu, Dipan K. Pal, and M. Savvides, "Hallucinating the Full Face from the Periocular Region via Dimensionally Weighted K-SVD," in *CVPRW*, June 2014.

[30] G. Mu, X. Gao, K. Zhang, X. Li, and D. Tao, "Single image super resolution with high resolution dictionary," in *ICIP*, Sept 2011, pp. 1141–1144.