L1-L1 NORMS FOR FACE SUPER-RESOLUTION WITH MIXED GAUSSIAN-IMPULSE NOISE

Junjun Jiang [1,2], *Zhongyuan Wang* [3], *Chen Chen* [4], *and Tao Lu* [2]

[1] School of Computer Science, China University of Geosciences, Wuhan 430074, China
[2] Hubei Province Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, 430073, China
[3] School of Computer, Wuhan University, Wuhan 430072, China
[4] Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA

ABSTRACT

In real world surveillance application, the captured faces are often low resolution (LR) and corrupted by mixed Gaussian-impulse noise during the acquisition and transmission processes. In this paper, we propose an effective patch-based face super-resolution method to reconstruct a high resolution (HR) face image given an LR observation that is corrupted by mixed Gaussian-impulse noise. To represent the corrupted image patches, a sparse regularization combined with an ℓ_1 data fitting term is proposed. In the proposed model, both the patch reconstruction term and the regularization term are in the ℓ_1 norm form. As a result, the model is called ℓ_1-ℓ_1 norms. In addition, since image pixels have nonnegative intensities, we further add a nonnegative constraint to the patch representation model. Experimental results demonstrate that the proposed ℓ_1-ℓ_1 norms based method can achieve superior face super-resolution performance over several state-of-the-art approaches based on the objective results in terms of P-SNR, as well as the visual perceptual quality.

Index Terms— Video surveillance, super-resolution, ℓ_1-ℓ_1 norms, mixed Gaussian-impulse noise, sparse representation

1. INTRODUCTION

In criminal investigation, face image, as one of the most informative objects for investigators, is the most direct and important cue for solving a case. However, in real world surveillance application, the captured faces are low resolution (LR) and low quality due to the person of interest is at a distance, bandwidth and storage resources limitation, etc. With many details of facial features lost in image acquisition and transmission, it is very difficult to recognize the person of interest for computers or human. Super-resolution is a technology that can reconstruct a high resolution (HR) image from an LR observation, with the help of an exemplar training set to learn a suitable model [1, 2]. Fig. 1 shows the concept of face super-resolution. Because of the ability of effectively enhancing the resolution of the LR observation and restoring the detailed facial features, face super-resolution is a crucial step for the following face recognition task by human or computers [3, 4].

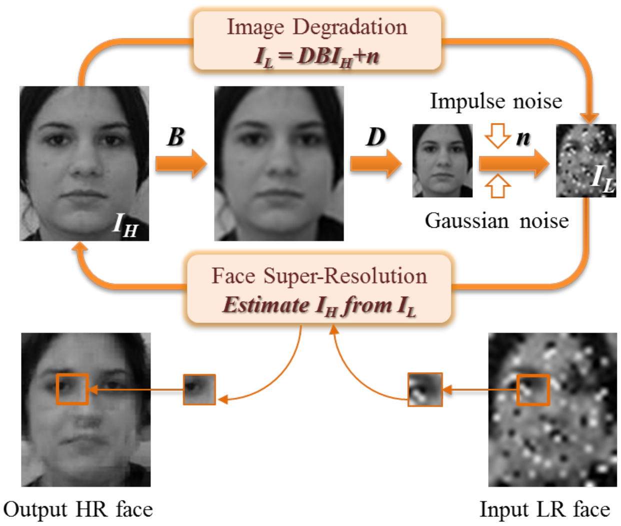

Fig. 1. The image degradation process to be reversed by face super-resolution. B is a blurring filter, D is the decimation operator, and n is the additive noise. Face super-resolution is to solve the ill-posed inverse problem of inferring the target HR face image from the input LR observation.

In [5], Baker and Kanada proposed "*face hallucination*" to model the relationship between LR image patches and HR ones using a Bayesian formulation. Capel and Zisserman [6] presented an algorithm where principal components analy-

The research was supported by the National Natural Science Foundation of China under Grant 61501413, in part by the Open Foundation of Hubei Provincial Key Laboratory of Intelligent Robot under Grant HBIR201404, in part by the National Natural Science Foundation of China under Grant 61502354, and in part by the Natural Science Fund of Hubei Province under Grant 2015CFB406.

sis (PCA) subspace models were used to learn six different regions of the entire faces. Liu *et al.* [7] developed a two-step approach toward super-resolution of faces by integrating a PCA based global parametric model and a Markov random field (MRF) based local nonparametric model. From then on, face super-resolution has attracted growing attention from the image processing community and many algorithms have been developed [8, 9]. In this paper, we mainly focus on the local patch based methods for their strong representation ability.

The most representative work for local patch based super-resolution was proposed by Chang *et al.* [10]. With the assumption that HR image patches and LR image patches share the same geometry structure, they utilized Locally Linear Embedding (LLE) [11] to learn the optimal reconstruction weights of multiple LR base elements to estimate the optimal HR patch representation. Zhang and Cham further extended the LLE algorithm to the discrete cosine transform (DCT) space [12]. Li et al. learned the relationship between the LR and HR patches in the common manifold spaces [13]. In [14], Jiang *et al.* proposed to model the LR and HR patch spaces iteratively. Ma *et al.* [15] proposed a least squares representation (LSR) based local patch face super-resolution method by incorporating the position-patch information. To address the problem of unstable solution of LSR, Yang *et al.* [16] developed a sparse representation (SR) based local patch face super-resolution method by adding a sparsity constraint to the lease squares problem. Most recently, Jiang *et al.* [17, 18] further improved the position-patch based representation method by incorporating the locality constraint instead of the sparsity constraint, and presented a locality-constrained representation (LcR) method. Some recent works [19, 20, 21] that focus on the locality prior and manifold assumption have proposed to improve the performance of [17, 18]. Most recently, many face super-resolution methods that focus on the wild conditions have been proposed [22, 23].

The aforementioned approaches are proposed under the noiseless or the additive white Gaussian noise (AWGN) conditions. However, in practice, especially in video surveillance systems, impulse noise is also a common type of image degradation (as shown in Fig. 1) due to malfunctioning pixel elements in the camera sensors, errors in analog-to-digital conversion, faulty memory locations, and transmission errors [24, 25]. In recent years, many mixed Gaussian-impulse noise removal technologies have been proposed [26, 27]. These methods usually first perform impulse noise pixel detection and then remove the added mixed-noise. The most direct way of face super-resolution with mixed Gaussian-impulse noise is to perform noise removal (or denoising) and face super-resolution reconstruction separately. However, the denoising error may be amplified in the following face super-resolution reconstruction task.

One natural question is that *can we develop a mixed Gaussian-impulse noise robust face super-resolution method which does not perform noise removal and face super-resolution reconstruction separately but conducts the two tasks in the unified framework?* Inspired by the recent works [28, 29], which have verified that the ℓ_1 norm based data-fidelity term can be effectively applied to the mixed-noise removal problem, in this paper we propose an ℓ_1-ℓ_1 norms based patch representation model to super-resolve the LR observation that is corrupted by mixed Gaussian-impulse noise. In our method, the data-fidelity term as well as the regularization term are both modeled using ℓ_1 norm. It does not have an explicit noise removal step and can perform denoising and face super-resolution reconstruction simultaneously. Furthermore, we also incorporate the nonnegative constraint to the proposed ℓ_1-ℓ_1 norms based patch representation model to make the super-resolved results more reasonable.

The rest of this paper is organized as follows. In Section II, we present the implementation details of the proposed method. Section III reports some numerical and visual results with mixed noise inputs on one public face dataset. Finally, the conclusion appears in Section IV.

2. THE PROPOSED ℓ_1-ℓ_1 NORMS BASED PATCH REPRESENTATION METHOD

Following the merits of [10, 15, 16, 30, 18], given an LR observation patch image x_t, the goal of these local patch based face super-resolution is to reconstruct the HR version of the input LR patch image y_t by learning the relationship between the LR and HR training image patch pairs, $X = [x_1, x_2, ..., x_N]$ and $Y = [y_1, y_2, ..., y_N]$. The key issue is how to obtain the optimal representation of x_t with X via the following minimization problem:

$$\hat{w} = \arg\min_{w} f(w; x_t, X) + \lambda\Omega(w), \tag{1}$$

where $f(w; x_t, X)$ is the data-fidelity term, $\Omega(w)$ is an image prior (also called the regularization term), and λ is the regularization parameter that balances the contribution of the patch reinstruction error and the regularization term. In fact, the above regularization based minimization problem can be strictly derived from the maximum a posteriori (MAP) criterion with prior knowledge in image degradation models.

The prior works mainly focused on designing a good regularization term to obtain the optimal patch representation w. For example, collaboration [15], sparsity [16, 30, 31, 32], locality [10] and data-driven locality [17, 33, 18, 19, 20] were introduced to regularize the reconstruction of w. They used the l_2 norm-based linear least squares term, *i.e.*, $f(w; x_t, X) = ||x_t - Xw||_2^2$, for the data-fidelity. The l_2 norm-based linear least squares problem is easy to solve and the solution is robust to AWGN. In the case of AWGN, the encoding model can be generally written as:

$$\hat{w} = \arg\min_{w} ||x_t - Xw||_2^2 + \lambda\Omega(w). \tag{2}$$

With some regularization, *e.g.*, sparsity or locality, the reconstructed weights w are the Bayesian inference for the AWGN noise model. However, for images corrupted by impulse noise, the distribution of noise is generally far from Gaussian and thus the ℓ_2 norm data fidelity term $||x_t - Xw||_2^2$ in Eq. (2) will not lead to an MAP solution for noise removal. This sparsity regularization not only ensures that the under-determined linear least squares problem has an exact solution, but also makes the solution more robust than using the ℓ_2 norm in statistical estimation.

In real video surveillance systems, the LR observations are often degraded by mixed noise (*e.g.*, Gaussian noise mixed with impulse noise). Therefore, the additive noise does not satisfy the Gaussian assumption in this case. Recently, minimizers of cost functions involving ℓ_1 data fidelity have been studied [28, 29], and have shown good performance in image restoration problems. To restore images corrupted by mixed Gaussian-impulse noise, in this paper we also introduce the ℓ_1 data fidelity and propose an effective patch representation model by solving the following minimization problem, which we call the ℓ_1-ℓ_1 problem:

$$\hat{w} = \arg\min_{w} ||x_t - Xw||_1 + \lambda||w||_1, \tag{3}$$

where $||.||_1$ denotes the ℓ_1 norm. In Eq. (3), we use the ℓ_1 norm for both the data fitting and the regularization terms.

Usually, image pixels satisfy $x \geq 0$. By incorporating the nonnegative constraint, it can make the solution more reasonable. Thus, the objective function of our proposed method can be written as follows:

$$\begin{aligned} \hat{w} = \arg\min_{w} ||x_t - Xw||_1 + \lambda||w||_1 \\ \text{s.t.} \quad \mathrm{w} \geq 0. \end{aligned} \tag{4}$$

The problem (4) can be solved by CVX toolbox [34].

Upon acquiring the optimal representation $\hat{w}$ of the input LR patch image with the LR training set, the corresponding HR version can be reconstructed by the same optimal representation with the HR training set.

3. EXPERIMENTAL RESULTS

In this section, we describe the details of experiments performed to evaluate the effectiveness of the proposed ℓ_1-ℓ_1 based patch representation method for face super-resolution. We compare our method with several recent state-of-the-art methods including Wang *et al.*'s eigentransformation method [35], Ma *et al.*'s Least Squares Representation (LSR) method [15], Yang *et al.*'s Sparse Representation (SR) [36] method, and our previously proposed Locality-constrained Representation (LcR) method [18]. The experiments are carried out on the public FEI face database[1] [37]. The face super-resolution performance is quantified by the peak signal-to-noise ratio (PSNR) between the ground truth face images and the super-resolved ones.

Database and Parameter Settings. The FEI face database consists of 200 subjects and each subject has two facial images. Human faces in the database are mainly from 19 to 40 years old with distinct appearances, *e.g.*, hairstyles and adornments. All the images are cropped to 120×100 pixels to form the HR training faces. Note that the face can be aligned by some recently proposed automatic alignment methods [38, 39] and feature points matching methods [40, 41]. The LR images are formed by smoothing (by a 4×4 mean filter) and down-sampling (by a factor of 4 resulting the size of LR face images to be 30×25 pixels) the corresponding HR images, and then the additive white Gaussian noise (with the standard deviation $\sigma = 10$) and impulse noise (by the random value with 10% and the salt-and-pepper impulse noise with 5%) is added. In our experiments, we use 360 images to train the proposed algorithm, leaving the rest 40 images for testing. As reported in [18], we set the size as 12×12 pixels for HR patch and the overlap between neighbor patches as 4 pixels. The corresponding LR patch size is set to 3×3 pixels with an overlap of one pixel.

Comparison Results. From Fig. 2, it can be seen that for face super-resolution with mixed Gaussian-impulse noise, the proposed ℓ_1-ℓ_1 norms based patch representation method could consistently achieve much higher PSNR indices than the LSR [15] and SR [16] methods, and better PSNR performance than the LcR method [18]. Note that the average gain in terms of PSNR of our proposed method over the second best method (*i.e.*, the LcR method [18]) is 1.52 dB.

We also present two groups of visual comparisons of the face super-resolution results using different methods. Fig. 3 shows the super-resolution results on two LR faces. Clearly, the proposed ℓ_1-ℓ_1 norms based patch representation method can better reconstruct edges than all the other competing methods. In particular, Bicubic, LSR [15] and SR [16] methods fail to recover the image structures. LSR [15] and SR [16] cannot well deal with the impulse noise due to the ℓ_2 norm based data fidelity is designed for the Gaussian noise and is not suitable for the mixed noise. Wang *et al.*'s method removes most of the added noise, but also smooths the facial details leading to the super-resolved faces similar to the mean face. Compared to the LcR method [18], our proposed method is able to faithfully reconstruct the edges and detailed texture features (see the regions highlighted by red boxes). The results of the LcR method [18] are smooth and dissimilar to the "ground truth" HR face images.

4. CONCLUSIONS

In this paper, we proposed a novel approach for face super-resolution with mixed Gaussian-impulse noise. By exploring the patch representation prior, we presented an effective

[1] http://fei.edu.br/∼cet/facedatabase.html

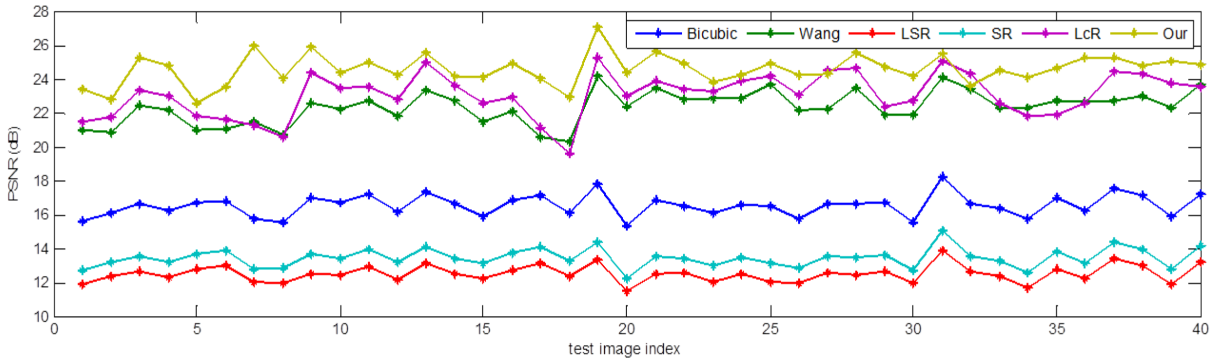

Fig. 2. Face super-resolution results (PSNRs) of different methods. The average PSNR values of the 40 test images for these six comparison methods are 16.54 dB (Bicubic), 22.35 dB (Wang *et al.* [35]), 12.51 dB (LSR [15]), 13.46 dB (SR [16]), 23.07 dB (LcR [18]), and 24.59 dB (our proposed method), respectively.

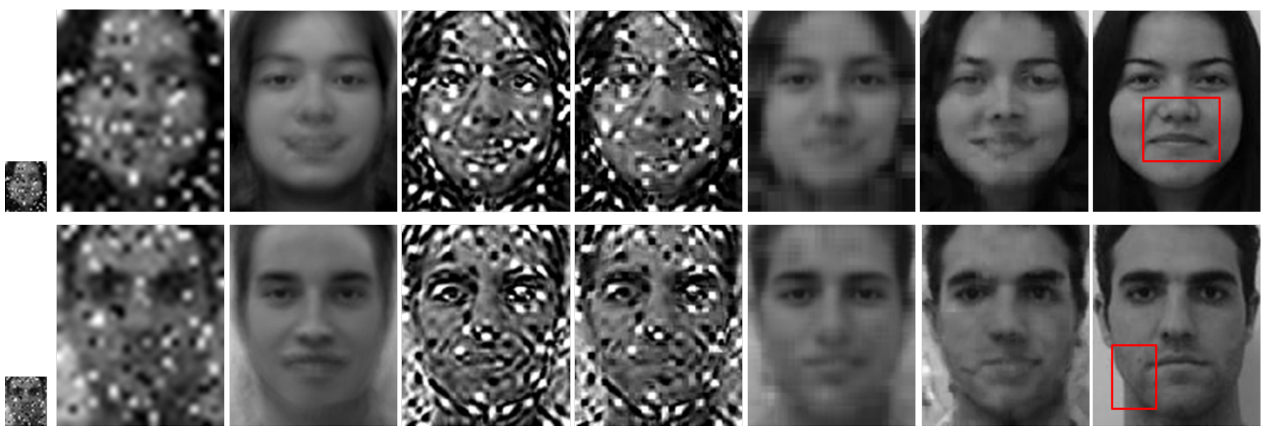

Fig. 3. Face super-resolution results of different methods (from left to right are the input LR images, results of Bicubic interpolation, Wang *et al.*'s method [35], LSR method [15], SR method [16], LcR method [18], our proposed method, and the "ground truth" HR face images.

patch representation with an ℓ_1 based data fidelity and an ℓ_1 based regularization term. The proposed $\ell_1 - \ell_1$ norms is much more suitable for molding the mixed Gaussian-impulse noise. Experimental results on the public FEI dataset verified the robustness of our proposed face super-resolution method to mixed Gaussian-impulse noise. Moreover, the proposed method achieved superior face super-resolution performance over the state-of-the-art algorithms which ignored the influence of data fidelity for the super-resolution reconstruction.

5. REFERENCES

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Process. Magazine*, vol. 20, no. 3, pp. 21–36, may 2003.

[2] Z. Zhu, F. Guo, H. Yu, and C. Chen, "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec. 2014.

[3] J. Jiang, R. Hu, Z. Han, L. Chen, and J. Chen, "Coupled discriminant multi-manifold analysis with application to low-resolution face recognition," in *MultiMedia Modeling*. Springer, 2015, pp. 37–48.

[4] J. Jiang, R. Hu, Z. Wang, and Z. Cai, "Cdmma: Coupled discriminant multi-manifold analysis for matching low-resolution face images," *Signal Processing*, pp. –, 2015.

[5] S. Baker and T. Kanade, "Hallucinating faces," in *FG*, 2000, pp. 83–88.

[6] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *CVPR*, 2001, vol. 2, pp. II–627–II–634 vol.2.

[7] C. Liu, H. Y. Shum, and C. S. Zhang, "A two-step approach to hallucinating faces: global parametric model and local nonparametric model," in *CVPR*, 2001, vol. 1, pp. 192–198.

[8] Kui Jia and Shaogang Gong, "Generalized face super-resolution," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 873–886, Jun 2008.

[9] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *International Journal of Computer Vision*, pp. 1–22, 2013.

[10] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *CVPR*, 2004, vol. 1, pp. 275–282.

[11] Sam T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326.

[12] W. Zhang and W. K. Cham, "Hallucinating face in the dct domain," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2769–2779, Oct 2011.

[13] B. Li, H. Chang, S. Shan, and X. Chen, "Aligning coupled manifolds for face hallucination," *IEEE Signal Proc. Let.*, vol. 16, no. 11, pp. 957–960, 2009.

[14] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, pp. 1–1, 2015.

[15] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recogn.*, vol. 43, no. 6, pp. 2224–2236, 2010.

[16] J. Yang, J. Wright, Thomas H., , and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.

[17] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, "Position-patch based face hallucination via locality-constrained representation," in *ICME*, 2012, pp. 212–217.

[18] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Trans. on Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug 2014.

[19] J. Jiang, R. Hu, Z. Han, Z. Wang, T. Lu, and J. Chen, "Locality-constraint iterative neighbor embedding for face hallucination," in *ICME*, 2013, pp. 1–6.

[20] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4220–4231, Oct 2014.

[21] Reuben F. and Christine G., "Face hallucination using linear models of coupled sparse support," *arXiv preprint arXiv:1512.06009*, 2015.

[22] Erjin Zhou, Haoqiang Fan, Zhimin Cao, Yuning Jiang, and Qi Yin, "Learning face hallucination in the wild," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[23] Y. Jin and C. S. Bouganis, "Robust multi-image based blind face hallucination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5252–5260.

[24] C.L.P. Chen, L. Liu, L. Chen, Y.Y. Tang, and Y. Zhou, "Weighted couple sparse representation with classified regularization for impulse noise removal," *IEEE Trans. on Image Processing*, vol. 24, no. 11, pp. 4014–4026, Nov 2015.

[25] L. Liu, C. L. Philip Chen, Y. Zhou, and X. You, "A new weighted mean filter with a two-phase detector for removing impulse noise," *Inf. Sci.*, vol. 315, pp. 1–16, 2015.

[26] Ming Yan, "Restoration of images corrupted by impulse noise and mixed gaussian impulse noise using blind inpainting," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1227–1245, 2013.

[27] J. Jiang, L. Zhang, and J. Yang, "Mixed noise removal by weighted encoding with sparse nonlocal regularization," *Image Processing, IEEE Transactions on*, vol. 23, no. 6, pp. 2651–2662, 2014.

[28] Mila Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers," *SIAM Journal on Numerical Analysis*, vol. 40, no. 3, pp. 965–994, 2002.

[29] Tommi Kärkkäinen, K Kunisch, and Kirsi Majava, "Denoising of smooth images using l 1-fitting," *Computing*, vol. 74, no. 4, pp. 353–376, 2005.

[30] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," *IEEE Signal Proc. Let.*, vol. 18, no. 6, pp. 367–370, 2011.

[31] Xiang Ma, Hiłp Quang Luong, Wilfried Philips, Huansheng Song, and Hua Cui, "Sparse representation and position prior based face hallucination upon classified over-complete dictionaries," *Signal Processing*, vol. 92, no. 9, pp. 2066 – 2074, 2012.

[32] G. Gao and J. Yang, "A novel sparse representation based framework for face image super-resolution," *Neurocomputing*, vol. 134, pp. 92 – 99, 2014.

[33] Chen C. and James E. Fowler, "Single-image super-resolution using multihypothesis prediction," in *Proceedings of the* 46^{th} *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2012, pp. 608–612.

[34] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[35] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev.*, vol. 35, no. 3, pp. 425–434, 2005.

[36] J. Yang, H. Tang, Y. Ma, and T. Huang, "Face hallucination via sparse coding," in *Proc. IEEE Conf. Image Process.(ICIP)*, 2008, pp. 1264–1267.

[37] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902 – 913, 2010.

[38] J. Ma, J. Zhao, and Alan L Yuille, "Non-rigid point set registration by preserving global and local structures," *Image Processing, IEEE Transactions on*, vol. 25, no. 1, pp. 53–64, 2016.

[39] J. Ma, W. Qiu, J. Zhao, Y. Ma, Alan L Yuille, and Z. Tu, "Robust L_2E estimation of transformation for non-rigid registration," *Signal Processing, IEEE Transactions on*, vol. 63, no. 5, pp. 1115–1129, 2015.

[40] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 12, pp. 6469–6481, 2015.

[41] J. Ma, J. Zhao, J. Tian, Alan L Yuille, and Z. Tu, "Robust point matching via vector field consensus," *Image Processing, IEEE Transactions on*, vol. 23, no. 4, pp. 1706–1721, 2014.