MANIFOLD REGULARIZED SPARSE SUPPORT REGRESSION FOR SINGLE IMAGE SUPER-RESOLUTION

Junjun Jiang, Ruimin Hu, Zhongyuan Wang, Zhen Han, and Shi Dong
National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, 430072, China
Email: {jiangjunjun, hrm1964, wzy_hope, hanzhen, dongshi}@whu.edu.cn

ABSTRACT

In this paper, we present a novel single image super-resolution method. To simultaneously improve the resolution and perceptual image quality, we bring forward a practical solution combining *manifold regularization* and *sparse support regression*. The main contribution of this paper is twofold. Firstly, a mapping function from low resolution (LR) patches to high-resolution (HR) patches will be learned by a local regression algorithm called sparse support regression, which can be constructed from the support bases of the LR-HR dictionary. Secondly, we propose to preserve the geometrical structure of the image patch dictionary, which is critical for reducing the artifacts and obtaining better visual quality. Experimental results demonstrate that the proposed method produces high quality results both quantitatively and perceptually.

***Index Terms*—**image enhancement, super-resolution, support regression, manifold learning, sparse representation.

1. INTRODUCTION

With the development of computer network and the rapid progress in hand-held photographic mobile devices, images and videos are becoming more and more popular on the web, due to their rich content and easy perception. However, limited by the network bandwidth and server storage, most images exist as low resolution (LR) and low quality versions degraded from the source. There is a huge need for improving the perceptual image quality, among which the resolution enhancement technology is called *super-resolution*. Instead of imposing higher requirements on hardware devices and sensors, it can offer us high-resolution (HR) and high-quality images with more details economically. In this paper, we focus on the *Single Image Super-Resolution* (SISR) problem because of its potential usefulness and flexibility for different applications.

Since SISR is inherently *ill-posed* as there are generally multiple HR images corresponding to the same LR image, accordingly, one has to rely on strong prior information, which is available either in the explicit form of a distribution or energy functional (*e.g.*., Tikhonov regularization [1] and Total Variation regularization [2]), and/or in the implicit form of training images which leads to learning-based super-resolution. A few representative methods of such kind are summarized as follows.

A manifold assumption based on *Locally Linear Embedding* (LLE) [3] is proposed by Chang *et al*. [4], and they assume that image patches in LR patch space and the corresponding HR one are located at two similar local geometries, and the HR patch could be generated as a linear combination of its *K* neighbor HR patches found in the training database. Recently, in [5] and [6], Yang *et al*. employ *sparse coding* to perform image super-resolution, which enforces corresponding LR and HR patches to share the *same sparse representations*. In their works, by enforcing sparsity regularization, LR patches are coded with respect to an over-complete LR dictionary, and the coefficients (*i.e.,* the outcome of the sparse coding process) are obtained to linearly combine corresponding HR counterparts to perform image super-resolution reconstruction. However, the constraint of "same sparse representation" in their approach is too strong to achieve in practice [7].

In this paper, we present a manifold regularized regression framework for super-resolution as shown in Fig. 1. The "same sparse representation" is relaxed for LR-HR sparse support domain regression, which is flexible in using the information of local training samples. Note that image patches have regular structures where accurate estimation of pixel values via regression is possible. Accordingly, the proposed method has more power and flexibility to describe different image patterns. In addition, the proposed method simultaneously considers the manifold regularization, thus capturing the intrinsic geometrical structure of the dictionary.

2. PROPOSED MANIFOLD REGULARIZED SPARSE SUPPORT REGRESSION METHOD FOR SUPER-RESOLUTION

This section presents the formulation of the proposed *Manifold regularized Sparse Support Regression* (MSSRt) for super-resolution. It then describes the optimization algorithm.

2.1. Formulation

Given a set of LR and HR training image patch pairs, $\{(x_1,y_1),\cdots,(x_N,y_N)\}\subset\mathbb{R}^d\times\mathbb{R}^D$, d and D are the dimensions of one LR and one HR patch respectively. Define $X=[x_1,\cdots,x_N]$ and $Y=[y_1,\cdots,y_N]$, each column of which is a patch sample. Thus the matrixes X and Y can be viewed as the LR and HR patch dictionaries respectively.

Considering that the manifold assumption (two manifolds spanned by the feature spaces of the LR and HR patches are locally similar) may not be tenable, we learn a much more stable LR-HR mapping in the support domain for super-resolution. Thus it can be transformed to a regression problem.

Our another important goal is to encode the geometry of the HR patch manifold, which is much more credible and discriminated compared with that of the LR one [13], and preserve the geometry for the reconstructed HR patch space. This will ensure that the local geometric structure of the reconstructed

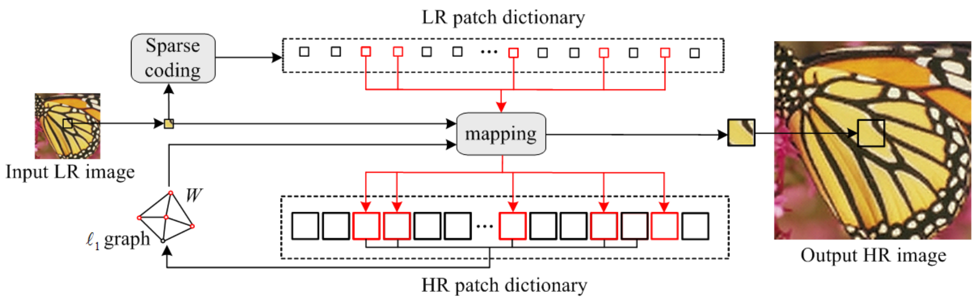

Fig. 1. Flowchart of the proposed method. Note that the red patches denote the sparse support domain of the input LR patch on the dictionary, and we use the sparse graph of the HR sparse support domain to guide the construction of the mapping function.

HR patch manifold is consistent with that of the original HR one.

Based on the above discussions, our MSSR algorithm for image super-resolution should be equipped with two properties:

i) The shared support of each LR patch and HR patch has an explicit regression relationship;

ii) The local geometrical information on the original HR patch dictionary is preserved.

In the following part, we will describe how we formulate MSSR with these two desired properties.

2.2. Sparse Support Regression

Instead of assuming that each pair of HR and LR patches has the same sparse representation, in our proposed MSSR method, this strong regularization of "same sparse representation" is relaxed for sparse support regression, and the sparse coefficient vectors of one LR and HR patch pair share the "same support", i.e., the same indices of nonzero elements.

Given a set of LR and HR training patches (dictionary pairs), $\{(x_1,y_1),\cdots,(x_N,y_N)\}$, for an unseen LR patch x_t, we try to learn a mapping function $f(x,P)=Px$, from the LR patch to the HR one to minimize the following regularized cost function for the regression

$$\varepsilon(P)=\sum_{i\in S}(Px_i-y_i)^2+\alpha\left\|P\right\|_H^2, \qquad (1)$$

where α is a regularization parameter, P is a $D\times d$ matrix to be learned, $\left\|P\right\|_H^2$ is the induced norm of f in the *reproducing kernel Hilbert space* (RKHS) space H, and S is the support of the coding coefficients $\hat{\theta}$ of the unseen patch x_t on LR training patches X:

$$\hat{\theta}=\arg\min_{\theta}\left\|x_t-X\theta\right\|_2+\lambda_1\left\|\theta\right\|_1 \quad . \qquad (2)$$

Thus, $S=\text{support}(\hat{\theta})$. In Eq. (2), $\left\|\theta\right\|_1$ denotes the ℓ_1 norm of θ, and the parameter λ_1 balances the coding error of x_t and the sparsity of θ. The solution of Eq. (2) can be achieved by convex optimization methods referring to [9].

The support of one vector is referring to the indices of nonzero elements in the vector. Defining X_S and Y_S as $X_S=\{x_i \mid i\in S\}$ and $Y_S=\{y_i \mid i\in S\}$ respectively and using Fibonacci norm to represent the smoothness of H, we can rewrite Eq. (1) as the following matrix form:

$$\varepsilon(P)=\left\|PX_S-Y_S\right\|_F^2+\alpha\left\|P\right\|_F^2 . \qquad (3)$$

2.3. Mining the Geometry on HR Patch Dictionary

This section targets on the second property, which is to preserve the local geometrical information on the HR patch dictionary. Note that the neighborhood relation, which guides the formulation of sparseness, is defined on the manifold rather than the Euclidean space.

Researchers have proposed various methods to measure the similarity between data points [10, 11], e.g., pair-wise distance based similarity and reconstruction coefficient based similarity. Since the former is suitable for discriminant analysis problems, such as recognition and clustering. Alternatively, reconstruction coefficient based similarity is datum-adaptive, and thus more suitable for image super-resolution. LLE is one of the representative works for reconstruction coefficient similarity estimation. It calculates the coefficient for each data through k-NN searching, thus k sparsity. The performance of LLE graph will decrease rapidly when the datas are non-uniformly sampled from underlying manifold, and this situation is very common in practice.

Recently, some researchers have demonstrated that the sparse structure of one manifold can be explored by the ℓ_1 graph [11], resulting in many benefits for machine learning and image processing problem. Let y_i be the i-th HR patch , which is under consideration now. We want to identify its neighbors on the smooth manifold rather than the entire Euclidean space. On the smooth patch manifold space, the patch can be well sparsely approximated by a linear combination of a few nearby patches. Thus, it has a sparse representation over the support domain Y_S. For any HR patch y_i, it can be sparsely approximate by the data matrix Y_S except y_i:

$$\begin{aligned}\hat{W}_{\bullet i}&=\arg\min_{W_{\bullet i}}\left\|y_i-Y_SW_{\bullet i}\right\|_2+\lambda_2\left\|W_{\bullet i}\right\|_1,\\ &\quad\text{s.t.}\quad W_{ii}=0\end{aligned} \qquad (4)$$

where $W_{\bullet i}$ denotes the i-the column of the matrix W whose diagonal elements are zeros, and λ_2 is the parameter balancing the coding error of y_i and the sparsity of $W_{\bullet i}$.

2.4. MSSR Objective Function and Optimization

We preserve the geometry relation represented by W for the reconstructed HR patch manifold. When LR patch is transformed to the HR patch, we try to preserve geometry constraint from W for $f(X_S, P)$. It can be gained by minimizing

$$\sum_{i\in S}\left\|Px_i - PX_S W_{\bullet i}\right\|_2^2 = \left\|PX_S - PX_S W\right\|_F^2 = \left\|PX_S(I-W)\right\|_F^2, \quad (5)$$

where I is an identity matrix.

Considering both of the two properties we want to engage, the objective function of our proposed MSSR is defined as:

$$O_{MSSR} = \left\|PX_S - Y_S\right\|_F^2 + \alpha\left\|P\right\|_F^2 + \beta\left\|PX_S(I-W)\right\|_F^2, \quad (6)$$

where β is a regularization parameter.

Using matrix properties $tr(AB) = tr(BA)$, $\left\|A\right\|^2 = tr(AA^T)$, and $tr(A) = tr(A^T)$, we have

$$\begin{aligned} O_{MSSR} &= tr\left\{(PX_S - Y_S)(PX_S - Y_S)^T\right\} + \alpha tr(PP^T) \\ &\quad + \beta tr(PX_S(I-W)(I-W)^T X_S^T P^T) \\ &= tr(PX_S X_S^T P^T - PX_S Y_S^T - Y_S X_S^T P^T + Y_S Y_S^T) \\ &\quad + \alpha tr(PP^T) + \beta tr(PX_S G X_S^T P^T), \end{aligned} \quad (7)$$

where $G = (I-W)(I-W)^T$.

In order to minimize the objective Eq. (7), we would like to take the derivative of O_{MSSR} with respect to P and set it to zero, i.e., we have the following equation

$$\begin{aligned} &\frac{\partial O_{MSSR}}{\partial P} = 2PX_S X_S^T - 2Y_S X_S^T + 2\alpha P + 2\beta PX_S G X_S^T = 0 \\ &\Rightarrow P(X_S X_S^T + \alpha I + \beta X_S G X_S^T) = Y_S X_S^T \\ &\Rightarrow P = Y_S X_S^T (X_S X_S^T + \alpha I + \beta X_S G X_S^T)^{-1}. \end{aligned} \quad (8)$$

Following [6], we perform a back projection for the super-resolved HR image of the proposed MSSR method to satisfy the global reconstruction constraint.

3. RELATION TO PRIOR WORK

Note that our method is similar to the *Local Learning based Regression* (LLR) method proposed in [8], which is also trying to learn a mapping between LR and HR patches. However, there are essential differences between LLR and the proposed method:

i) LLR learns the LR-HR mapping in the local space of *K* nearest neighbors, which uses a fixed number of nearest neighbors through the feature space, while the proposed method adaptively selects the neighbors without any predefined neighborhood size, and reveals the mapping relationship between the LR and HR patch in the sparse support domains;

ii) LLR does not take into account the geometric structure of the patch manifold that plays an important role in the choice of example patches, while the proposed method aims to preserve the geometric structure of the original HR patch manifold space for that of the reconstructed HR one, thus well revealing the similar local geometric structure manifold of LR and HR patch spaces and enhancing the learning performance.

4. EXPERIMENTAL RESULTS

In this section, we verify the performance of the proposed MSSR method. We conduct experiments on five widely used test images as shown in Fig. 2. Several state-of-the-art methods, such as Bicubic interpolation, *Neighbor Embedding* (NE) [4], *Sparse Coding* (SC) [6], and *Local Learning based Regression* (LLR) [8] are used as comparison baselines. *Peak Signal to Noise Ratio* (PSNR), *Root Mean Square Error* (RMSE), and *Structural Similarity* (SSIM) [12] indices are adopted to evaluate the objective quality of the super-resolved results. Since human eyes are more sensible to the change of the luminance, hence, the super-resolution reconstruction is only performed on the luminance component, and the simple Bicubic interpolator is used for the chromatic components.

Fig. 2. Gallery of test images used in our experiments. From left to right, they are named "*barbara*", "*foreman*", "*house*", "*lenna*", and "*zebra*" respectively.

To extract the high frequency information of LR images, 4 directions of gradients (2 horizontal directions and 2 vertical directions) are used as input features in all super-resolution algorithms. In the following experiments, the magnification factor is 3, the size of the LR patches is set to 3 × 3, and the size of HR patches is set to 9 × 9. 50,000 LR and HR training patch pairs are randomly chosen from the training images used in [6], which has no relation with the test images used in our experiments, for training neighbor embedding [4] and local learning [8] and the coupled dictionaries with 1024 elements [6] respectively. The neighborhood number of NE [4] is set to 10 and the sparsity parameter of SC is set to 0.1. For the sake of fairness, we use the some trained dictionary for SC [6] and MSSR. For MSSR, the regularization parameters λ_1, λ_2, α and β are empirically set to 0.1 , 0.15, 0.3 and 10 respectively.

The PSNR (dB), RMSE and SSIM of all five different test images are reported in Table I. It can be seen from Table I that the proposed MSSR method achieves the best in terms of PSNR, RMSE and SSIM. MSSR outperforms Bicubic interpolation, NE [4], SC [6], and LLR [8] in all cases, which validates the necessity and effectivity of sparse support regression and manifold geometric preservation.

Fig. 3 shows the visual results of different super-resolution algorithms. All of the learning-based super-resolution methods outperformed the Bicubic interpolation in terms of visual plausibility. Note that the proposed algorithm performs visually much better than Bicubic interpolation, having less visual artifacts and producing sharper results. Compared with other learning-based super-resolution methods, the proposed algorithm provides more image details with improved objective values. The results of NE [4] method are sharp in the textures. However, unpleasant artifacts and tiny block effects are also introduced as shown in "*foreman*" and "*lenna*". SC [6] method uses the sparseness prior to regularize the HR image, which suppresses the high frequency details in the texture region but introduces some noise as shown in "*barbara*", "*house*", and "*zebra*" images. This is mainly due to the difficulty to learn a universal coupled LR and HR dictionary that can represent various LR and HR structure pairs. The result of locality prior method (LLR [8]) shown in the fifth column is sharp along salient edges. However, the texture detail is blurry and there are some jaggy artifacts and ringing artifacts. Our result in the sixth column is sharp both along edges and in the textural regions. We owe the superiority of the proposed method to manifold constrained local sparse regression, which is more powerful and flexible to describe different image patterns.

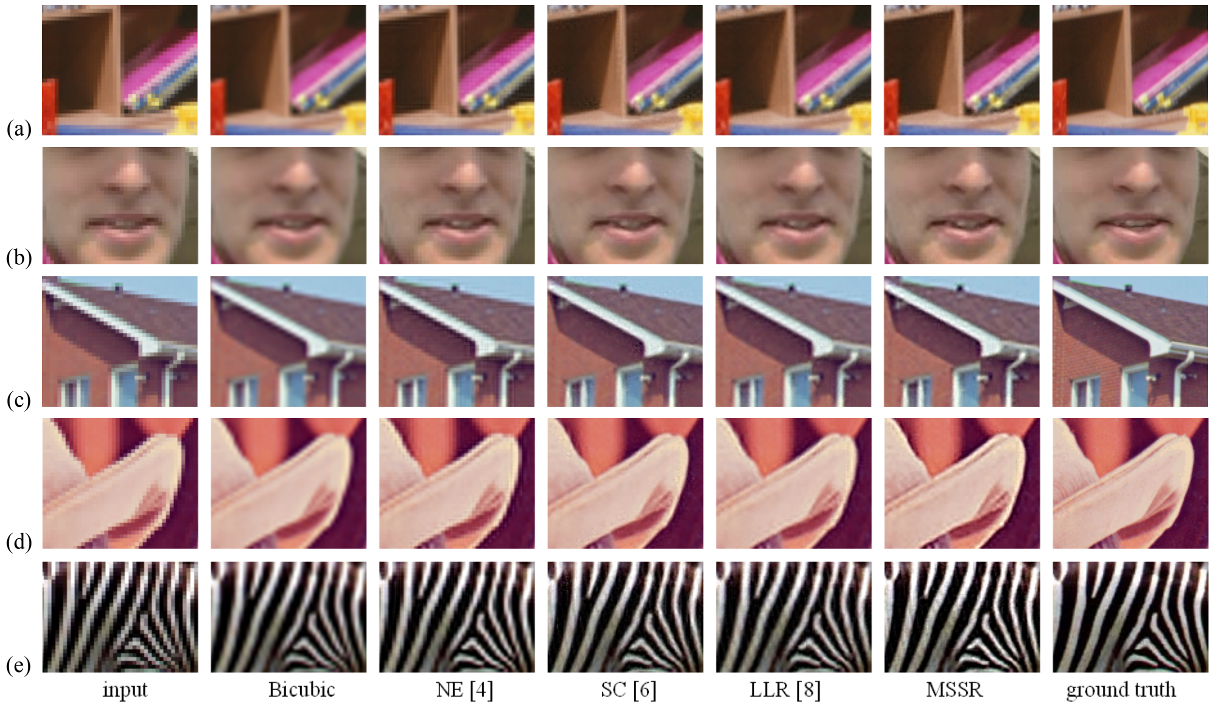

Fig. 3. Super-resolution results of different methods. (a)-(e) are the local magnification of "*barbara*", "*foreman*", "*house*", "*lenna*" and "*zebra*" respectively.

Table I. PSNR (dB), RMSE, and SSIM comparisons of different super-resolution methods.

Images	**Bicubic**	**NE [4]**	**SC [6]**	**LLR [8]**	**MSSR**
barbara	26.20	26.49	26.31	26.51	26.58
	12.49	12.08	12.33	12.05	11.96
	0.7543	0.7651	0.7689	0.7751	0.7810
foreman	29.64	29.84	30.59	30.55	31.00
	8.41	8.21	7.54	7.57	7.19
	0.9022	0.8874	0.9096	0.9083	0.9211
house	29.54	29.70	30.33	30.25	30.60
	8.51	8.35	7.77	7.83	7.53
	0.8564	0.8529	0.8586	0.8633	0.8756
lenna	31.73	31.75	32.85	32.66	33.07
	6.61	6.59	5.81	5.93	5.66
	0.8587	0.8555	0.8710	0.8710	0.8782
zebra	26.69	27.31	28.06	28.08	28.49
	11.80	10.99	10.08	10.05	9.60
	0.7946	0.8210	0.8250	0.8318	0.8417
Average	28.76	29.02	29.63	29.61	29.95
	9.56	9.24	8.70	8.69	8.39
	0.8332	0.8364	0.8466	0.8499	0.8595
Improve-ment	**1.19**	**0.93**	**0.32**	**0.34**	—
	1.17	**0.85**	**0.31**	**0.30**	—
	0.0263	**0.0231**	**0.0129**	**0.0096**	—

5. MAIN FINDINGS AND FUTURE DIRECTIONS

This paper propose a novel single image super-resolution method, namely Manifold regularized Sparse Support Regression (MSSR), which simultaneously considers the manifold geometrical structure of the patch manifold space and the support of the corresponding sparse coefficients. The support information as well as the geometrical structure information of the data manifold are incorporated into the MSSR model. We design a novel sparse regression algorithm for having both reconstruction and generalization properties, which can enhance the learning performance. It is experimentally shown that the proposed MSSR methods can produce more faithful details and higher objective quality in comparison to the other state-of-the-art super-resolution approaches. Extending the current linear model to the non-linear case will be our further work. In addition, we may introduce some reasonable prior [14, 15, 16] to suppress the artifact of the super-resolved image.

6. ACKNOWLEDGMENT

The research was supported by the major national science and technology special projects (2010ZX03004-003-03, 2010ZX03004-001-03), the National Basic Research Program of China (973 Program) (2009CB320906), the National Natural Science Foundation of China (61231015, 61172173, 61070080, 61003184, 61170023), the Fundamental Research Funds for the Central Universities (201121102020009), the science and technology Program of Wuhan (201271031366).

7. REFERENCE

[1] N. Tikhonov and V. Y. Arsenin , “Solutions of ill-posed problems,” New York: Winston, 1977.

[2] L. Rudin, S. Osher, and E. Fatemi, “ Nonlinear total variation based noise removal algorithms,” *Phys. D: Nonlinear Phenomena*, vol. 60, no.1–4, pp. 259–268, 1992.

[3] S.T. Roweis and L.K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[4] H. Chang, D. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit.*, *(CVPR)*, pp. 275–282, 2004.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR)*, pp.1-8, 2008.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma. “Image super-resolution via sparse representation,” *IEEE Trans. Image Process.,*, vol.19, no.11, pp.2861–2873, 2010.

[7] S. Wang, L. Zhang, Y. Liang and Q. Pan, “Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch image synthesis,” in *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2216-2223, 2012.

[8] Y. Tang, P. Yan, Y. Yuan, and X. Li, “Single-image super-resolution via local learning,” *Int. J. Mach. Learn. & Cyber.*,vol. 2, pp. 15–23, 2011.

[9] H. Lee, A. Battle, R. Raina, and A. Y. Ng, “Efficient sparse coding algorithms,” in *Advances in Neural Information Processing Systems (NIPS)*, pp. 801-808, 2007.

[10] B. Cheng, J. Yang , S. Yan , Y. Fu , T. Huang, “Learning with ℓ^1-graph for image analysis,” *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858-866, 2010.

[11] B. Shen and L. Si, “Nonnegative matrix factorization clustering on multiple manifolds,” in *Proc. of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, pp. 575-580, 2010.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error measurement to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp.600-612, 2004.

[13] J. Jiang, R. Hu, Z. Han, K. Huang, and T. Lu, “Efficient single image super-resolution via graph embedding,” in *Proc. IEEE Conf. on Multimedia and Expo (ICME)*, pp. 610-615, Jul 2012.

[14] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin, “Super resolution using edge prior and single image detail synthesis,” in *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2400-2407, 2010.

[15] J. Sun, Z. Xu, and H. Shum, “Image super-resolution using gradient profile prior,” in *Proc. on IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1-8, 2008.

[16] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, “Position-Patch Based Face Hallucination via Locality-constrained Representation,” in *Proc. IEEE Conf. on Multimedia and Expo (ICME)*, pp. 212-217, Jul 2012.