2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

IMAGE SUPER-RESOLUTION VIA KERNEL REGRESSION OF SPARSE COEFFICIENTS

Tingrong Yuan, Fei Zhou, Wenming Yang, and Qingmin Liao*

Visual Information Processing Laboratory/Shenzhen Key Laboratory of Science and Technology
Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China

ABSTRACT

In this paper, we present a sparse coding (SC) inspired method to reconstruct a high-resolution (HR) image from one single low-resolution (LR) image. Instead of restricting the coding coefficients of LR and HR image patches to be equal or linearly mapped, we introduce kernel regression to nonlinearly relate the coding coefficients of LR patches and those of corresponding HR ones in an implicit fashion. Meanwhile, principal component analysis (PCA) is employed to train independent dictionaries which can well express image geometrical structure and ensure image sparse property. Experimental results show that the proposed method can effectively reconstruct image details and outperforms state-of-the-art algorithms in both quantitative and visual comparisons.

Index Terms— super-resolution (SR), sparse coding (SC), kernel regression, nonlinear mapping

1. INTRODUCTION

Single-frame image super-resolution (SR) aims at inferring high-resolution (HR) image from one low-resolution (LR) image, which is severely in demand in many applications such as showing LR images on high definition displays of digital devices.

It is well known that SR is an ill-posed inverse problem, which can be typically modeled as: $\mathbf{y} = \mathbf{DHx} + \mathbf{n}$, where $\mathbf{y}$ and $\mathbf{x}$ denote LR image patch and HR image patch in lexicographic order respectively, $\mathbf{H}$ represents blurring matrix, $\mathbf{D}$ is the downsampling matrix, and $\mathbf{n}$ is additive noise. Obviously, the feature dimension of $\mathbf{x}$ is much higher than $\mathbf{y}$. To deal with the severely ill-posed problem, many methods have been proposed in the past decades. Classical methods based on interpolation, such as bi-linear interpolation and bi-cubic interpolation, are widely used in image or video processing software or hardware products. The advantage of these methods is their simplicity. However, they tend to cause jaggies and ringing artifacts since they cannot adapt to varying image structures. Edge-guided interpolation [1] improves visual quality by performing interpolation in a chosen direction to preserve image edge structures. This approach is further evolved in [2] by using sparse mixing estimators.

Some prior knowledge has also been applied to regularize the SR problem due to its ill-posed nature. One popular regularization is the total variation [3], which assumes that natural images have small first-order derivatives. This assumption means that images have piecewise constant structure. Accordingly, this method tends to smooth image details. Sun *et al.*[4] propose the gradient profile prior for local image structures which is effective in preserving image edges. However, these approaches are limited in modeling the visual complexity of the natural images [5].

In recent years, learning-based methods show great potentiality in dealing with SR problem. These methods assume that the lost details in an LR image can be predicted by the learned information from a specified database. Freeman *et al.* propose an example-based learning approach using a Markov Random Field (MRF) with belief propagation in [6]. However, the learning stage is time-consuming. In [7], Chang *et al.* adopt the theory of manifold learning, assuming that the manifolds of LR image and corresponding HR image are located in similar geometrical patterns. Hence, neighbor embedding is proposed to estimate HR patch as a linear combination of neighbors. Nevertheless, this method often results in blurring effects. In [8], a SR algorithm using support vector regression (SVR) learning strategy in spatial and DCT domain is proposed. In [9], upscaling is achieved by kernel regression. However, all these methods are limited in revealing the intrinsic and complex relation between the HR and LR images.

2. RELATED WORKS

Recently, the theory of sparse representation has been successfully applied to the SR problem. In [5], Yang *et al.* employ L_1 norm sparsity regularization, utilizing the prior knowledge that image patches can be coded sparsely with respect to trained dictionary. Moreover, the sparse coefficients of LR image patches are assumed to be identical to the corresponding HR patches. Essentially, the method transforms the SR problem into seeking the sparsest solution of an L_1 norm optimization problem. This method has been improved in [10], where a bilevel sparse coding model is proposed to en-

This work was supported in part by the Natural Science Foundation of China (Grant No. 61271393 and 61301183) and China Postdoctoral Science Foundation under Grant 2013M540947, and in part by the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under Grant JCYJ20120619151228556.

*Corresponding Author, E-mail: yangelwm@163.com.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

hance the degree of coupling between the LR and HR feature spaces. However, the fully coupled dictionary learning model is inflexible to express image structures of HR and LR image patches. In [11], A semi-coupled dictionary learning (SCDL) model is used. In the SCDL model, the LR and HR dictionaries are not fully coupled. A linear mapping is learned to relate the coding coefficients of LR and HR image patches. In [12], a nonlocal autoregressive model (NARM) is incorporated into the spare coding (SC) based SR framework. The method improves the ability of reconstructing edge structures. However, the identical or linear mapping can not well capture the intrinsic relation between LR and HR coding coefficients.

Inspired by these work, in this paper, we focus on the SC based methods for SR problem. Instead of fully coupled or semi-coupled dictionaries as in the methods mentioned above, we train the LR dictionary and corresponding HR dictionary independently by using PCA. Furthermore, we propose a strategy aimed at discovering the intrinsic and nonlinear relation between the sparse coefficients using kernel regression.

3. SR METHODOLOGY

In this section, the classical SC-based method and its enhanced version are first briefly reviewed. Next, the proposed method will be discussed in detail.

3.1. Sparse Coding

According to [5], image patches can be coded sparsely with respect to trained dictionary, *i.e.*,

$$\begin{aligned} \mathbf{x} &= \mathbf{D}_h\boldsymbol{\alpha} \\ \mathbf{y} &= \mathbf{D}_l\boldsymbol{\alpha} \quad with \ \|\boldsymbol{\alpha}\|_0 \ll N, \end{aligned} \tag{1}$$

where $\mathbf{D}_h \in \mathbb{R}^{d_h \times N}$ is the over-complete HR dictionary, $\mathbf{D}_l \in \mathbb{R}^{d_l \times N}$ is the corresponding LR dictionary, N is the atom number of the dictionary, $\boldsymbol{\alpha}$ is the coding coefficients, $\|\cdot\|_0$ denotes L_0 norm. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$, $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$ be the HR and LR data matrices respectively, $\boldsymbol{\Lambda} = [\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_n]$ be the coefficient matrix. The coupled dictionary can be obtained by solving the following L_1 and L_2 norm mixing optimization problem:

$$\min_{\{\mathbf{D}_h,\mathbf{D}_l,\boldsymbol{\Lambda}\}} \|\mathbf{X} - \mathbf{D}_h\boldsymbol{\Lambda}\|_F^2 + \|\mathbf{Y} - \mathbf{D}_l\boldsymbol{\Lambda}\|_F^2 + \gamma\|\boldsymbol{\Lambda}\|_1, \tag{2}$$

where γ is the regularization parameter, $\|\cdot\|_F$ and $\|\cdot\|_1$ denote Frobenius norm and L_1 norm respectively. In (2) the first and second terms represent data fidelity, and the third one is sparse regularization term. When reconstructing an HR patch $\mathbf{x}$ from LR patch $\mathbf{y}$, the sparse coefficient vector $\boldsymbol{\alpha}$ can be estimated as:

$$\tilde{\boldsymbol{\alpha}} = arg\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}_l\alpha\|_2^2 + \gamma\|\boldsymbol{\alpha}\|_1. \tag{3}$$

Then, HR image patch is reconstructed by using the coefficients calculated in (3): $\mathbf{x} = \mathbf{D}_h\tilde{\boldsymbol{\alpha}}$.

From the objective function in (2), we can find that sparse representation of HR patch is the same as the corresponding LR one. Actually, the coding coefficients may not be strictly equal. To express image structures in a more flexible way, Wang *et al.* introduce a semi-coupled dictionary learning strategy [11] :

$$\begin{aligned} &\min_{\{\mathbf{D}_h,\mathbf{D}_l,\mathbf{W}\}} \|\mathbf{X} - \mathbf{D}_h\boldsymbol{\Lambda}_h\|_F^2 + \|\mathbf{Y} - \mathbf{D}_l\boldsymbol{\Lambda}_l\|_F^2 \\ &+ \gamma\|\Lambda_h - \mathbf{W}\boldsymbol{\Lambda}_l\|_F^2 + \lambda_l\|\boldsymbol{\Lambda}_l\|_1 + \boldsymbol{\lambda}_h\|\boldsymbol{\Lambda}_h\|_1 + \boldsymbol{\lambda}_W\|\mathbf{W}\|_F^2 \end{aligned} \tag{4}$$

where λ_1, λ_h, λ_W are regularization parameters. The cost function in (4) shows that the coding coefficients of HR and LR patches have a linear mapping. And the linear mapping matrix $\mathbf{W}$ is pre-learned during training stage. The SCDL model in [11] has improved the ability to express image structure by adopting the linear mapping. However, for complex real-word images, this model is also limited in describing the relation between LR coefficients $\boldsymbol{\Lambda}_l$ and HR coefficients $\boldsymbol{\Lambda}_h$.

3.2. Our model

To overcome the drawbacks of the models analyzed above, we propose a new model aimed at finding the intrinsic and implicit relationship between the sparse coefficients. An independent dictionary learning strategy and kernel regression are used to achieve this goal.

Independent Dictionary Learning

In the first stage of learning, we train the HR and LR dictionaries separately. Before learning, we collect thousands of image patch pairs from several HR natural images. To better characterize the image local structures, we follow the suggestions of [12] [13] to classify the patches into K subsets : $\mathbf{S} = [\mathbf{S}_1, ..., \mathbf{S}_j, ...\mathbf{S}_K]$, where $\mathbf{S}_j$ represents the j-th cluster. Then we perform PCA on each subset to obtain the local sub-dictionary $\boldsymbol{D}_j$ $(j = 1, ..., K)$, where $\boldsymbol{D}_j = [\boldsymbol{p}_1, ..., \boldsymbol{p}_i, ..., \boldsymbol{p}_r]$, $\boldsymbol{p}_i$ is i-th eigenvector. To better approximate the sparse property, the value of r is determined by

$$\min_r \|\boldsymbol{S}_k - \mathbf{D}_{kr}\boldsymbol{\Lambda}_{kr}\|_2^2 + \gamma\|\boldsymbol{\Lambda}_{kr}\|_1, \tag{5}$$

where $\mathbf{D}_{kr} = \boldsymbol{D}_k$ is the sub-dictionary of k-th cluster, $\boldsymbol{\Lambda}_{kr} = \boldsymbol{D}_k^T\mathbf{S}_k$ represents coding coefficients. Obviously, the first term of (5), *i.e.*, reconstruct error, will decrease if the number of principal components r increases, however, the sparse regularization term will increase. Therefore, an appropriate value can be obtained via solving (5). Those sub-dictionaries ultimately form a big over-complete dictionary $\mathbf{D} = [\mathbf{D}_1, ..., \mathbf{D}_K]$ for HR image patches and LR patches respectively.

Kernel Regression Training

In the second stage of learning, Kernel Support Vector Regression (K-SVR) [14] is adopted to learn the relation between the coding coefficients. For all the training patch pairs

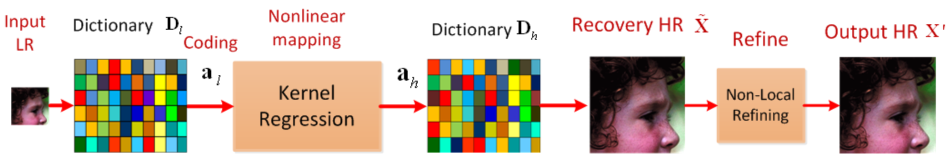

Fig. 1: Flowchart for reconstructing HR images.

belong to cluster k, we code these patches as:

$$\begin{aligned} \mathbf{\Lambda}_{lk} &= \boldsymbol{D}_{lk}^T \mathbf{S}_{lk} = [\boldsymbol{\alpha}_{l,1}, \ldots, \boldsymbol{\alpha}_{l,M}] \\ \mathbf{\Lambda}_{hk} &= \boldsymbol{D}_{hk}^T \mathbf{S}_{hk} = [\boldsymbol{\alpha}_{h,1}, \ldots, \boldsymbol{\alpha}_{h,M}] \\ \boldsymbol{\alpha}_{l,i} &\in \mathbb{R}^{N_l \times 1}, \quad \boldsymbol{\alpha}_{h,i} \in \mathbb{R}^{N_h \times 1}, \quad i = 1, \ldots, M \ , \end{aligned} \tag{6}$$

where $\mathbf{S}_{lk}$ and $\mathbf{S}_{hk}$ are training LR and HR patches belong to cluster k ($k = 1, \ldots, K$), M is the number of the patches, $\boldsymbol{D}_{lk}$ and $\boldsymbol{D}_{hk}$ are corresponding sub-dictionaries, and Λ_{lk}, Λ_{hk} are LR and HR coding coefficients matrices belong to cluster k. N_h, N_l are atoms number of HR and LR dictionary respectively. Given training set with M input-output pairs as:

$$\Omega = \{(\boldsymbol{\alpha}_{l,1}, \alpha_{h,1}^j), \ldots, (\boldsymbol{\alpha}_{l,i}, \alpha_{h,i}^j), \ldots, (\boldsymbol{\alpha}_{l,M}, \alpha_{h,M}^j)\} \tag{7}$$

where $j \in [1, N_h]$, $i \in [1, M]$, $\boldsymbol{\alpha}_{l,i}$ is input coding vector, $\alpha_{h,i}^j$ is the associated output value, *i.e.*, the j-th dimensional value of HR patch coefficients. We try to estimate the function $f : \boldsymbol{\alpha}_l \to \alpha_h^j$. In SVR, the nonlinearity is introduced by mapping the data into a high-dimensional feature space $\mathcal{F}$ using a nonlinear mapping $\Phi : \mathbb{R}^{N_l} \to \mathcal{F}$. By introducing kernel function, inner product in feature space can be calculated without explicitly computing the mapping. In this work, the Gaussian function of width $\sigma > 0$ is adopted as the kernel function,

$$\mathcal{K}(\mathbf{x}, \mathbf{y}) = exp(-\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{2\sigma^2}). \tag{8}$$

In SVR, regression function f can be obtained by solving the following optimization problem :

$$\begin{aligned} &\min_{\boldsymbol{w}, b, \varepsilon, \varepsilon^*} \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + C \sum_{i=1}^{M} (\varepsilon_i + \varepsilon_i^*) \\ &s.t. \ \ \alpha_{h,i}^j - (\langle \boldsymbol{w}, \phi(\boldsymbol{\alpha}_{l,i}) \rangle + b) \le \varepsilon + \varepsilon_i \\ &\quad (\langle \boldsymbol{w}, \phi(\boldsymbol{\alpha}_{l,i}) \rangle + b) - \alpha_{h,i}^j \le \varepsilon + \varepsilon_i^* \\ &\quad \varepsilon_i, \ \ \varepsilon_i^* \ge 0 \end{aligned} \tag{9}$$

where $\boldsymbol{w} \in \mathcal{F}$, $j \in [1, N_h]$, C is a parameter to tradeoff between the flatness of the regression function and the upper and lower bounds of training errors ε_i and ε_i^*, subject to a margin ε. The optimization problem (9) can be transformed into a dual maximization problem, which is easier to solve. And the regression function f takes the form

$$f_j(\boldsymbol{\alpha}_l) = \sum_{i=1}^{M} (\lambda_i - \lambda_i^*) \mathcal{K}(\boldsymbol{\alpha}_l, \boldsymbol{\alpha}_{l,i}) + b \tag{10}$$

where λ, λ^* are the Lagrange multipliers. Refer to [14] for more details. Note that this regression f_j aims at inferring the j-th dimensional value of HR patch coefficients $\boldsymbol{\alpha}_h$, so we need to learn a series of regression functions covering all the dimensional value for the cluster.

Reconstructing HR Image

In the synthesis stage, given an LR image, the HR image can be reconstructed via dictionary coding and K-SVR. Fig.1 shows the flowchart of synthesis. Firstly, for a given LR patch $\mathbf{y}_i$ to be coded, we select the subdictionary of the closest cluster, say $\mathbf{D}_{l,k}$, to code it, *i.e.*, $\boldsymbol{\alpha}_l = \mathbf{D}_{l,k}^T \mathbf{y}_i$, the remaining coding coefficients over other sub-dictionaries is set to $\mathbf{0}$. Thus a very sparse representation of $\mathbf{y}_i$ is obtained. Secondly, Map the sparse coefficients $\boldsymbol{\alpha}_l$ to $\boldsymbol{\alpha}_h$ using the learned K-SVR. Then HR patch is recovered by : $\tilde{\mathbf{x}}_i = \mathbf{D}_{h,k} \boldsymbol{\alpha}_h$. The HR image $\tilde{\mathbf{X}}$ can be initially reconstructed by merging all the HR patches and averaging the overlapping regions between the adjacent patches. Finally, a non-local constraint is introduced to further improve the SR performance,

$$\begin{aligned} &\mathbf{X}' = arg \min_{\mathbf{X}} \|\mathbf{X} - \tilde{\mathbf{X}}\|^2 \\ &s.t. \ \ \ \|\mathbf{x}_i - \sum_{m=1}^{L} w^m \tilde{\mathbf{x}}_i^m\|^2 \le \varepsilon, \end{aligned} \tag{11}$$

where L is the number of similar patches selected, $\mathbf{x}_i$, $\tilde{\mathbf{x}}_i$ are patches in $\mathbf{X}$ and $\tilde{\mathbf{X}}$, $\tilde{\mathbf{x}}_i^m$ is the m-th most similar patch to $\mathbf{x}_i$, and w^m is the non-local weight as defined in [15].

4. EXPERIMENTS

In this section, we provide experimental results which demonstrate the effectiveness of our method.

4.1. Experimental settings

For the HR reconstruction, zooming factor of 3 is conducted. HR test images are down-sampled to produce the corresponding LR images. We randomly sample 80000 HR and LR patch pairs from five training images used in [12]. The size of LR patch is 3×3, which is up-sampled to 6×6 when training, and HR patch size is 9×9. Each patch has been subtracted by its mean value in the feature space. We use K-means to divide these patches into 39 clusters. Patches in each cluster have similar structural pattern. For each cluster, we apply PCA to train sub-dictionary. And we set $\lambda = 0.1$ in (5) to get the number of principal components.

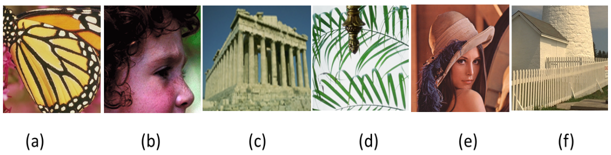

Fig. 2: Images for test, from (a) to (f): *Butterfly, Girl, Parthenon, Leaves, Lena, Fence*

After dictionary learning, we train a series of SVRs for each cluster. The parameters C, ξ in (9) and σ in (8) are slightly different between each cluster. Cross-validation is adopted to determine these parameters. As human visual system is more sensitive to the change of luminance, we only apply the SR methods to the luminance component and use simple bi-cubic interpolator for the chromatic components. During synthesis stage, an overlap of 1 pixel between adjacent LR patches is adopted.

4.2. Results and discussions

To demonstrate the superiority of our method, we compare it with methods including bi-cubic, SC [5], SCDL [11] and NARM [12]. Six images shown in Fig.2 are tested. And three crita, peak signal-to-noise (PSNR), multi-scale structural similarity (MS-SSIM) [16] and visual information fidelity(VIF) [17] are adopted to measure the SR reconstruction performance.

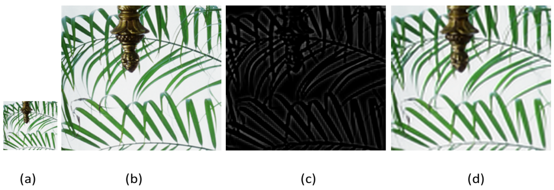

Fig. 3: Reconstructed results of *Leaves*, (a)Input LR image, (b)Origin HR image, (c)Edge profile inferred, (d)Reconstructed HR images of our method

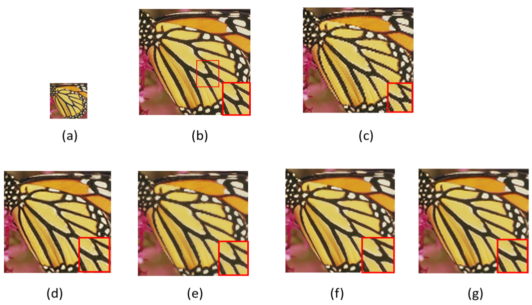

Fig. 4: Reconstructed results of *Butterfly* by different SR methods. (a)Input LR image, (b)Origin HR image, (c)Bi-cubic, (d)SC [5], (e)SCDL [11], (f)NARM [12], (g)Ours.

Table 1: Quantitative comparison on PSNR, MS-SSIM and VIF. For each block, the first row is PSNR, the second is MS-SSIM, the third is VIF.

images	bi-cubic	SC	SCDL	NARM	Ours
B.fly	23.24	23.84	24.42	25.46	**25.95**
	0.9597	0.9729	0.9685	0.9788	**0.9812**
	0.3588	0.4338	0.4125	0.4753	**0.4868**
Girl	30.19	31.10	31.82	31.24	**33.18**
	0.9417	0.9578	0.9480	0.9420	**0.9717**
	0.3400	0.4256	0.3535	0.3491	**0.4387**
Leaves	22.87	22.93	23.67	24.61	**25.26**
	0.9574	0.9715	0.9584	0.9705	**0.9775**
	0.3714	0.4311	0.4204	0.4760	**0.4919**
Lena	29.01	30.00	29.62	30.18	**31.03**
	0.9477	0.9470	0.9531	0.9536	**0.9753**
	0.4187	0.4197	0.4407	0.5086	**0.5181**
Parth.	24.12	24.06	24.71	24.99	**26.46**
	0.8887	0.9453	0.8975	0.8996	**0.9507**
	0.2181	0.2359	0.2359	0.2732	**0.3560**
Fence	20.57	20.38	20.91	20.91	**22.41**
	0.7502	0.8819	0.7597	0.7597	**0.8887**
	0.1588	0.2011	0.2009	0.2009	**0.2580**

For visual illustration, in Fig.3, the edge profile of *leaves* learned via coefficients kernel regression is displayed, which can be observed that both large-scale and fine-scale edges are well constructed. In Fig.4, the reconstruction results of *butterfly* by different methods are displayed. From these figures, we can see that SC method introduces unexpected noisy details, SCDL method leads to severe ringing artifacts. NARM method causes over-smoothed reconstruction results and loses image details, and this phenomenon is more severe in the other test images. However, our method reduces ringing and zipper artifacts and obtains better visual quality.

The results of quantitative comparison are listed in Table 1. The SC-based SR methods [5] [11] [12] perform better than bi-cubic interpolation, since the over-complete dictionary contains high-frequency information pre-learned. And our method surpasses all the other methods in terms of PSNR, MS-SSIM and VIF. The performance of our method demonstrates the effectiveness of nonlinear mapping between sparse coefficients and independent dictionary learning strategy.

5. CONCLUSIONS

In this paper, we propose a new learning-based SR method. Instead of coupled or semi-coupled dictionaries, we train independent ones via PCA, aimed at ensuring image sparsity property and increasing the flexibility of dictionaries to express image geometrical structures. More importantly, the intrinsic relation between the sparse coefficients of LR and HR patches is obtained through kernel regression. Furthermore, we introduce image nonlocal similarity to exploit image redundancies. Experimental results show that our method outperforms the state-of-the-art algorithms.

6. REFERENCES

[1] H. Shi and R. Ward, "Canny edge based image expansion," in *IEEE International Symposium on Circuits and Systems, ISCAS*. IEEE, 2002, vol. 1, pp. I–785.

[2] S. Mallat and G. Yu, "Super-resolution with sparse mixing estimators," *IEEE Trans. on Image Processing*, vol. 19, no. 11, pp. 2889–2900, 2010.

[3] A. Marquina and S. Osher, "Image super-resolution by tv-regularization and bregman iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, 2008.

[4] J. Sun, Z. Xu, and H. Shum, "Image super-resolution using gradient profile prior," in *IEEE Conference on Computer Vision and Pattern Recognition CVPR*. IEEE, 2008, pp. 1–8.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[6] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *Computer Graphics and Applications, IEEE*, vol. 22, no. 2, pp. 56–65, 2002.

[7] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*. IEEE, 2004, vol. 1, pp. I–275.

[8] K. Ni and T. Nguyen, "Image superresolution using support vector regression," *IEEE Trans. on Image Processing*, vol. 16, no. 6, pp. 1596–1610, 2007.

[9] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.

[10] J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang, "Bilevel sparse coding for coupled feature spaces," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2360–2367.

[11] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2216–2223.

[12] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. on Image Processing*, 2013.

[13] F. Zhou, W. Yang, and Q. Liao, "Single image super-resolution using incoherent sub-dictionaries learning," *IEEE Trans. on Consumer Electronics*, vol. 58, no. 3, pp. 891–897, 2012.

[14] S. Gunn, "Support vector machines for classification and regression," *ISIS technical report*, vol. 14, 1998.

[15] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2005, vol. 2, pp. 60–65.

[16] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*. IEEE, 2003, vol. 2, pp. 1398–1402.

[17] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.