

Image Super-Resolution via Nonlocal Similarity and Group Structured Sparse Representation

Wenhan Yang, Jiaying Liu, Saboya Yang and Zongming Guo*

Institute of Computer Science and Technology, Peking University, Beijing, P.R. China, 100871

Abstract—Sparse prior provides an effective tool for the image reconstruction. However, the sparse coding for independent patches leads to the unstable sparse decomposition. In this paper, we propose a group structured sparse representation model by considering the nonlocal similarity. The nonlocal similar patches are collected and classified into groups. Patches in the same group are reconstructed based the same basis of dictionaries. The dictionary is organized as the combination of many orthogonal sub-dictionaries. To provide the redundancy, the dictionary used for the sparse coding is generated online with several sub-dictionaries, thus it is over-complete. We apply the proposed model into a gradual SR framework. The framework enlarges LR to HR by a patch enhancement and an alternative sparse reconstruction on the patch and group. Objective quality evaluation shows that our proposed SR method achieves highest PSNR results comparing with the state-of-the-art methods. And subjective results demonstrate the proposed method reduces artifacts and preserves more details.

Index Terms—Sparse representation, super-resolution, nonlocal similarity, group sparse, structured sparse

I. INTRODUCTION

Image super-resolution (SR) tries to estimate a high-resolution (HR) image by one or more low-resolution (LR) images. Because of the information loss in the degradation, the recovery from LR to HR is under-determined. In order to constrain the recovery, different kinds of priors are put forward. Based on the form of priors, SR methods are classified into two categories: example-based and reconstruction-based.

Example-based methods [1] utilize patch pairs to build the connection between the LR and HR space. According to the nearest neighbors in the LR space, the HR patch is estimated by the combination of corresponding HR neighbors. These methods provide abundant high frequency details, whereas the geometry properties are hard to recover, leading to implausible visual effects.

Reconstruction-based methods impose the regularization as the prior to lead the recovery. The typical regularizations include gradient priors [2], nonlocal self-similarity priors [3] and sparsity priors [4], [5]. They characterize various aspects of natural image properties and recover high frequency information while preserving the intrinsic geometry properties.

The sparsity prior is one of the most important priors for the image reconstruction. It suggests that natural signals can be compactly expressed as a linear combination of per-specified atoms and the majority of the linear coefficients are zero. All the atoms form a basis signal set, which is called dictionary.

In the sparse-based methods, there are two fundamental problems to be considered: the design of the dictionary and the method to perform sparse coding. The structure of the dictionary plays a key role in its expressivity. The designed dictionaries, such as

DCT and wavelet, are equals to orthogonal transformations. Their coding and reconstruction process are simple and fast while they lack the adaptivity for the local structures and geometries. Learned dictionaries [4], [6], whose elements are selected based on the reconstructed performance on the training set, increasing the local adaptiveness and modeling capacity. The structure regularity between the items inside the dictionary are ignored and universal dictionaries are not adaptive to local image properties. Then, structured dictionary learning algorithms based on patch clusters [7], [8], [9] are proposed. The training patches are clustered first, then sub-dictionaries are learned based on patch clusters. Sparse decomposition on one patch is carried out with the corresponding sub-dictionary, thus highly adapted to local structures. However, in these methods, sub-dictionaries are orthogonal and the redundancy, which is proved to be an effective tool to model and reconstruct the image, is limited.

As for sparse coding methods, there many efficient L1-minimization techniques, such as iterative thresholding [10] and Bregman split algorithms [11]. In the coding, they regard patches as independent and uncorrelated ones. Therefore, the sparse pattern selection turns unstable and visual artifacts are easy to generate. This leads us to pay attention to the potential of sparse coding methods based on patch groups [12], [13].

Recently, nonlocal based image methods [5], [7] showed the outstanding ability for the image reconstruction. A large number of self-similarity recurrences deliver useful information for modeling image structures. Motivated by the nonlocal self-similarity, in order to provide an online over-complete dictionary that is adapted to local structures, we propose a group structured sparse representation (GSSR) model. The nonlocal similar patches are collected as a group and the sparse coding is carried out over the patch group. The sparse patterns selected for the patches in one group keep the same, making the sparse coding more stable. For the efficiency, the dictionary is organized as the combination of many orthogonal sub-dictionaries. To provide the redundancy, the dictionary used for the coding is generated by combining several sub-dictionaries and it turns over-complete. Furthermore, We apply GSSR to a gradual SR framework. High frequency details are introduced by a patch enhancement with nonlocal patches and external patches. And an alternative sparse coding over a single patch and a patch group is performed, to provide both the adaptivity and consistency.

The rest of the paper is organized as follows: Section II presents GSSR model. Section III suggests a gradual enlargement framework embedded with a patch enhancement scheme and an alternative sparse reconstruction on the patch and patch group. Experimental results and analysis are presented in Section IV. Finally, concluding remarks are given in Section V.

II. GROUP STRUCTURED SPARSE REPRESENTATION

A. Patch-based sparse representation

In SR scenarios, the patch-based sparse representation model is built on the independent patch. The LR image x is first cropped into overlapped patches $\{x_k | k = 1, 2, \dots, n\}$, where n is the number of

*Corresponding author

This work was supported by National Natural Science Foundation of China under contract No. 61472011 and National High-tech Technology R&D Program (863 Program) of China under Grant 2014AA015205.

patches and k is the variable related to the locations of patches. We use the extraction operator $R_k(\cdot)$ to build the connection between x and x_k . It is defined as: c

$$x_k = R_k(x) \quad (1)$$

Then, the high-resolution image patch x_k is represented with a sparse coefficient vector α_k by solving the following minimization problem:

$$\hat{\alpha}_k = \arg \min_{\alpha_k} \{ \|y_k - HD\Phi \cdot \alpha_k\|_2^2 + \lambda \|\alpha_k\|_1 \}, \quad (2)$$

where y_k is the degraded observation, H and D are the blur kernel and down-sampling operator in the degradation, respectively. Φ is the dictionary that provides prior information for the reconstruction. The Lagrange multiplier λ is a parameter that balances the tradeoff between the fidelity term and sparsity priors. Then, the reconstructed HR image patch \hat{x}_k is obtained by:

$$\hat{x}_k = \Phi \cdot \hat{\alpha}_k. \quad (3)$$

And the estimated HR image x is reconstructed by:

$$\hat{x} = \Phi \cdot \hat{\alpha} = \left(\sum_{i=1}^n R_i^T R_i \right)^{-1} \sum_{i=1}^n R_i^T \Phi \hat{\alpha}_i, \quad (4)$$

B. group-based sparse representation

In the traditional single patch-based method, due to the suboptimality of the sparse pattern selection, the sparse coding α_k on a single patch x_k is unstable. In order to obtain a precise sparse pattern P_k and get an accurate α_k , we add the nonlocal similarity redundancy into the sparse representation model to derive a group sparse representation model, improving the accuracy of the sparse coding.

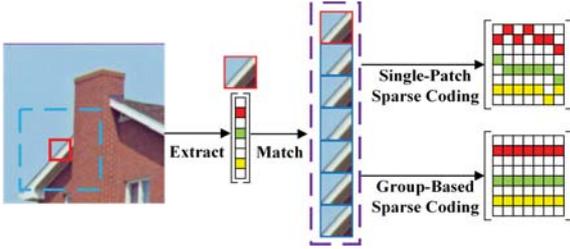


Fig. 1. The group-based sparse coding makes the representation coefficient estimation more consistent and stable.

First, for a single patch x_k , its nonlocal similar patches $\{x_k^1, x_k^2, \dots, x_k^m\}$ are collected, where m is the number of similar patches including x_k and x_k^1 is exactly x_k . The similarity is measured by the mean square error between patches $\|x_k - x_i\|_2^2$, where i is in the nonlocal scope of k . $\{x_k^1, x_k^2, \dots, x_k^m\}$ are arranged as a group x_k^g . Then, we perform the sparse coding on x_k^g instead of x_k . The sparse pattern P_k used for the coding stays the same. Thus, the unstable sparse pattern selection is avoided.

In the group-based model, the extraction operator is defined as $R_k^g = [R_{k_1}, R_{k_2}, R_{k_3}, \dots, R_{k_m}]$. And the sparse representation coefficients are defined as $\alpha_k^g = [\alpha_{k_1}, \alpha_{k_2}, \dots, \alpha_{k_m}]$. Thus, Equ. (4) turns to:

$$\begin{aligned} \hat{\alpha}_k^g &= \arg \min_{\alpha_k^g} \{ \|y_k^g - HD\Phi \cdot \alpha_k^g\|_2^2 + \lambda \|\alpha_k^g\|_1 \}, \\ &= \arg \min_{\alpha_k^g} \left\{ \sum_{l=1}^m \|y_k^l - HD\Phi \cdot \alpha_k^l\|_2^2 + \sum_{l=1}^m \lambda \|\alpha_k^l\|_1 \right\}, \end{aligned} \quad (5)$$

where $\{\alpha_k^l | l = 1, 2, \dots, m\}$ is under the constraint of the same sparse pattern. It means the nonzero items in α_k^l locate in the same dimension. Equ. (5) incorporates the nonlocal similarity into the sparse representation model. It imposes similar patches to have the same sparse decomposition. The problem in Equ. (5) is like the form of simultaneous sparse coding [12], [13] Thus, we convert it to a

simultaneous orthogonal matching pursuit problem by change the norm type in the fidelity term and sparsity constraint as follows:

$$\begin{aligned} \hat{\alpha}_k^g &= \arg \min_{\alpha_k^g} \{ \|y_k^g - HD\Phi \cdot \alpha_k^g\|_F^2 + \lambda \|\alpha_k^g\|_0 \}, \\ &= \arg \min_{\alpha_k^g} \left\{ \sum_{l=1}^m \|y_k^l - HD\Phi \cdot \alpha_k^l\|_F^2 + \sum_{l=1}^m \lambda \|\alpha_k^l\|_0 \right\}, \end{aligned} \quad (6)$$

where F is the Frobenius norm.

Then, We solve the sparse coding by simultaneous orthogonal matching pursuit with SPAMS [14]. After the sparse decomposition, the entire image x is represented in the sparse domain with $\{\alpha_k^g | k = 1, 2, \dots, n\}$. Then, x is rebuilt by the first reconstructed patch in each group x_k^g :

$$\hat{x} = \left(\sum_{i=1}^n R_{i_1}^T R_{i_1} \right)^{-1} \sum_{i=1}^n R_{i_1}^T \Phi \hat{\alpha}_i, \quad (7)$$

C. Group Structured Dictionary Learning

One key factor for the sparse representation modeling is the construction of the dictionary D . In general, there are two kinds of dictionaries: orthogonal and over-complete. Traditional human designed dictionaries, such as DCT and wavelet, are orthogonal. They are very simple and their corresponding dictionary learning and sparse coding have low computational complexities. However, due to limiting the redundancy, they cannot characterize the complex natural image signal. Learned dictionaries form a basis signal set to represent the image signal according to the reconstruction performance on a natural image training set. And their structures can be set as over-complete, thus they can provide the redundancy to depict complex image signals.

We designed an online over-complete dictionary generation algorithm. It forms over-complete dictionaries by combing several orthogonal sub-dictionaries, to make use of the efficiency of the orthogonal dictionary and the effectiveness of the over-complete dictionary. A series of orthogonal sub-dictionaries are trained based on patch clusters. In the stage of the sparse coding, the nearest several sub-dictionaries are chosen to form an over-complete dictionary. Then, the sparse coding is performed on the online dictionary by SOMP. The whole process is shown in Fig. 2.

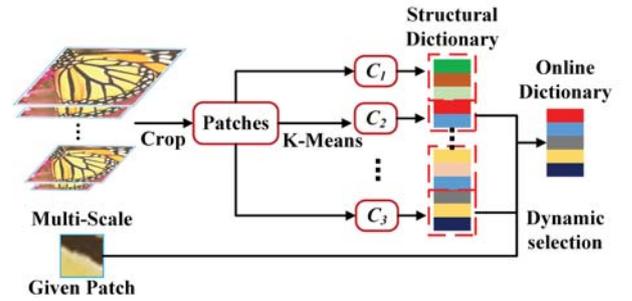


Fig. 2. The process of the group structured dictionary learning and online dictionary generation in GSSR.

All patches used for the dictionary training are cropped from the multi-scale LR images. Over-smooth patches are discarded by the condition of $var(p_i) < c$, where $var(\cdot)$ is the variance of the given patch and c is the given threshold. Then, we acquire a training set $T = \{p_1, p_2, \dots, p_N\}$ where n is the number of patches in T . In order to get meaningful features, we transform these patches into their high-frequency versions $T^h = \{p_1^h, p_2^h, \dots, p_N^h\}$ by the DoG operator. The K-means algorithm is carried out to partition T^h into K partitions $\{T_1^h, T_2^h, \dots, T_K^h\}$.

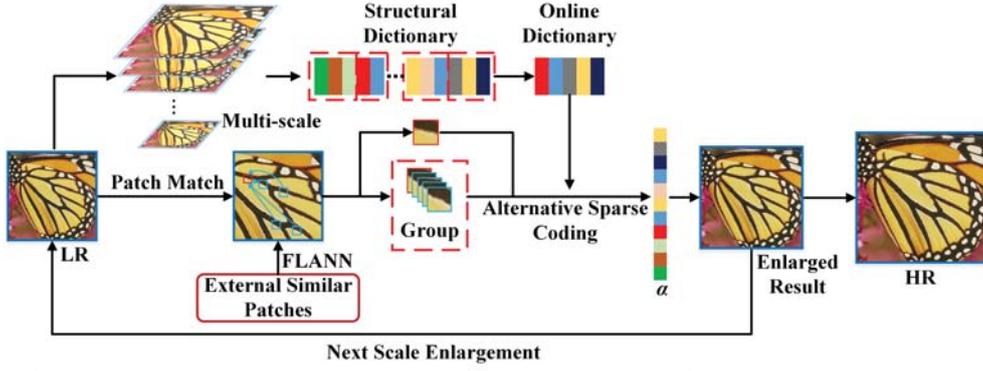


Fig. 3. The flow chart of the proposed super-resolution algorithm, including a gradual enlargement framework with the patch enhancement and the alternative sparse coding.

The traditional dictionary learning problem is formulated like the form of sparse coding problem in Equ. (2). However, for the efficiency, we do not directly optimize it. Instead, we obtain the dictionary utilizing the effective PCA transformation over each cluster. For each cluster, let Ω_k be the covariance matrix of the k -th partition T_k^h . By applying PCA to Ω_k , we get an orthogonal transform F_k . And the representation coefficients are $Z_k = F_k^T T_k^h$. For better modeling and avoiding overfitting, only parts of eigenvectors are used to form F_k . Thus, in the limit that the number of eigenvectors is less than r , we define $F_{k,r}$ and α_r as the transform matrix and representation coefficients, respectively. We choose a proper r as the optimal number of the involved eigenvectors in each cluster as follows:

$$\hat{r} = \arg \min_r \left\{ \|T_k^h - F_{k,r} \alpha_r\|_F^2 + \lambda \|\alpha_r\|_1 \right\}, \quad (8)$$

where $\|\cdot\|_F$ is the Frobenius norm.

III. GRADUAL SUPER-RESOLUTION FRAMEWORK

The framework of our method is shown in Fig. 3. For a given input LR image, we aim to get a good HR estimator, recovering the high frequency details while preserving the intrinsic geometry properties. Therefore, our method consists of two stages, the gradual magnification with patch enhancement and alternative sparse coding based on the patch and patch group.

A. Gradual Enlargement With Patch Enhancement

The image degradation leads to the high frequency detail loss. To restore it, we need to build the connection between LR and HR space. Utilizing the external similarity and self-similarity, we adopt an example-based high-frequency enhancement method. Because in the condition that scaling factor is small, the given LR is more similar to its HR version, we enlarge the LR to HR gradually. Each enlargement is carried out with a small scaling factor. For patch x_k , we search similar patches $\{x_{k_i}^n | i = 1, 2, \dots, t_1\}$, $\{x_{k_i}^e | i = 1, 2, \dots, t_2\}$ from the multi-scale LR images and external images, respectively. Then, x_k is reconstructed by a weighted combination of $\{x_{k_i}^n\}$ and high-frequency version of external similar patches $\{y_{k_i}^e\}$:

$$\hat{x}_k = \sum_{i=1}^{t_1} w_i^n x_{k_i}^n + \sum_{i=1}^{t_2} w_i^e y_{k_i}^e \quad (9)$$

$$w_i^n = \frac{1}{W_1} \exp(-\|x_k - x_{k_i}^n\|_2^2 / h_1), \quad (10)$$

$$w_i^e = \frac{1}{W_2} \exp(-\|x_k - x_{k_i}^e\|_2^2 / h_2), \quad (11)$$

where W_1 and W_2 are normalization factors, h_1 and h_2 are pre-determined scalars.

B. Alternative Sparse Coding

After introducing high frequency details in the patch enhancement, some artifacts may be presented. Thus, we then perform the sparse reconstruction to depress artifacts while preserving intrinsic geometric structures. Because the reconstruction is based on the gradual enlargement, a small deviation in the intermediate result may lead to a considerable error in the final result. Therefore, we hope the sparse coding stage can provide more diversified information, in order that we can balance the tradeoff between the adaptivity and consistency to pursue a better performance. Based on this consideration, we use an alternative sparse coding. Two kinds of sparse coding schemes are involved: ASDS [5] and our GSSP. The orthogonal sparse coding and recovery in ASDS provide adaptivity and conciseness while the group sparse coding and recovery in GSSP provide consistency and richness. Combing them, better performance is obtained and the algorithm achieves a good tradeoff between the adaptivity and consistency, conciseness and richness, respectively. From the point of solution process, single sparse coding method is easy to drop into a local minimum, while in the alternative sparse coding, two independent powers guide the solving process and the probability of dropping into local minimum turn smaller.

IV. EXPERIMENTAL RESULTS

To verify the effectiveness of our method, we conduct extensive experiments on the image super-resolution. We set the basic parameter setting as follows: 12 external similar patches and 12 nonlocal similar patches are involved to reconstruct every LR patch in the patch enhancement, the patch size is 6×6 , the overlap width is equal to 2. The initial cluster number of the group structured dictionary $K = 64$.

TABLE I
PSNR(DB) RESULTS IN $3 \times$ ENLARGEMENT.

	Bicubic	ScSR	ASDS	NCSR	Proposed
<i>Bike</i>	20.80	23.36	23.56	24.72	25.08
<i>Butterfly</i>	20.78	25.14	25.51	28.09	28.86
<i>Girl</i>	29.95	32.91	32.38	33.66	33.61
<i>Hat</i>	27.20	29.86	29.60	31.28	31.53
<i>Parrots</i>	25.58	28.85	28.59	30.49	30.76
<i>Parthenon</i>	24.12	26.35	26.17	27.17	27.35
<i>Plants</i>	27.83	31.96	31.84	34.04	34.30
<i>Raccoon</i>	26.38	28.50	28.11	29.27	29.32
Average	25.33	28.37	28.22	29.84	30.10
Gain	*	3.03	2.89	4.51	4.77

We conduct the qualitative and quantitative evaluations on our method in comparison with Bicubic interpolation method, ScSR [4], ASDS [5] and NCSR [7]. For ScSR, it is designed only for the enlargement without the deblurring. Thus, for a fair comparison, an

TABLE II
SSIM RESULTS IN $3\times$ ENLARGEMENT.

	Bicubic	ScSR	ASDS	NCSR	Proposed
<i>bike</i>	0.5755	0.7400	0.7617	0.8025	0.8152
<i>Butterfly</i>	0.7172	0.8543	0.8702	0.9158	0.9292
<i>girl</i>	0.7327	0.8066	0.8126	0.8276	0.8249
<i>hat</i>	0.7769	0.8449	0.8526	0.8705	0.8737
<i>Parrots</i>	0.8253	0.8943	0.8993	0.9148	0.9201
<i>Parthenon</i>	0.6202	0.7199	0.7264	0.7507	0.7567
<i>plants</i>	0.7867	0.8855	0.8969	0.9191	0.9223
<i>raccoon</i>	0.6276	0.7334	0.7426	0.7708	0.7669
Average	0.7078	0.8099	0.8203	0.8465	0.8511
Gain	*	0.1021	0.1125	0.1387	0.1434

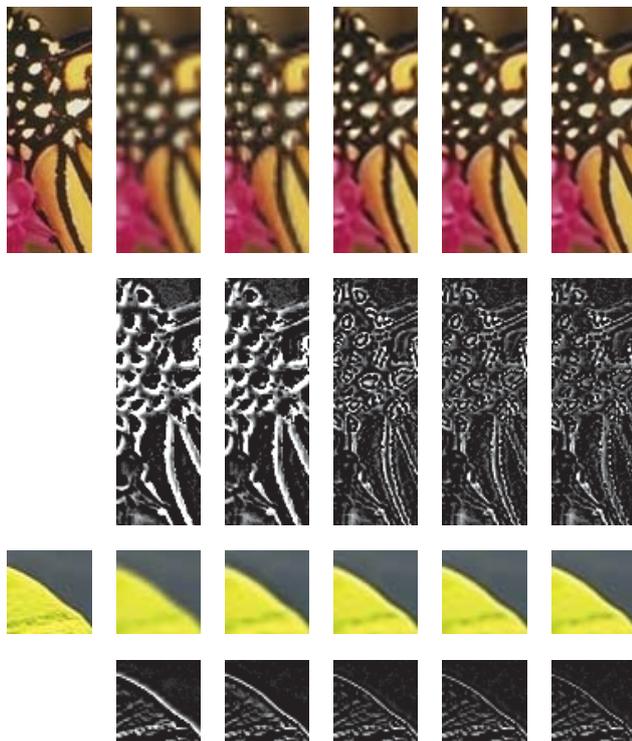


Fig. 4. Visual comparison between different algorithm results for the local images and the difference images. From up to down: *Butterfly*, *Hat*. From left to right: HR image, Bicubic, ScSR, ASDS, NCSR, Proposed method.

iterative back-projection is carried out for the image deblurring before the SR.

For simulating the image degradation process, we follow the similar operations in [5], [7]. The LR images is generated from HR images by a blurring and down-sampling operator. The blurring kernel is set as a 7×7 Gaussian kernel and its standard deviation is 1.6. For colorful images, the SR operator is only applied to the luminance component, whereas the chromatic components are enlarged by the Bicubic interpolation. To evaluate the quality of SR results, the Peak Signal-to-Noise Ratio (PSNR) and the perceptual quality metric Structural SIMilarity (SSIM) are chosen as the evaluation criteria.

Table I and Table II list image SR results of our method and four comparison methods with scaling factors 3 for the objective quality evaluation. Our method outperforms the other SR methods for the majority of test images. In $3\times$ enlargement, our method achieves the best SR performance at 30.10 dB (PSNR) and 0.8511 (SSIM) on average over 8 test images, obtaining a gain of 0.26 dB in PSNR and 0.0046 in SSIM over the average results (29.84 dB and 0.8465) of the second best method NCSR [7].

Fig. 4 demonstrate the super-resolution results by $3\times$ on *Butterfly*, *Hat*. As shown in figures, the Bicubic generates blurred results.

The ScSR preserves the majority of edges though there is a little blurring around them. The ASDS generates more natural edges and textures, but it is hard to avoid the blurring and artifacts. The NCSR recovers key structures, however, it still brings in little blurring and slight but noticeable artifacts around the edges. By comparison, due to incorporating the nonlocal similarity and consistency into the sparse representation model and introducing external high frequency information, our method preserves the edge better and generates more natural textures. Our method presents the desirable local results and obtains the darkest difference image, which means the results of our method contains fewer errors and more similar to the corresponding HR image.

V. CONCLUSION

In this paper, we propose a group sparse representation model by considering the nonlocal similarity. The nonlocal similar patches are clustered as a group. The sparse coding for patches in a group are performed jointly with the same sparse pattern. The dictionary is organized as the combination of many orthogonal sub-dictionaries. When performing the sparse coding, the dictionary is generated online with several nearest sub-dictionaries to the given patch and turns to be over-complete. We apply our GSSR model into a gradual enlargement framework. The framework gradually enlarge the LR with patch enhancement and alternative sparse coding. The objective and subjective quality evaluation demonstrate the effectiveness of our method.

REFERENCES

- [1] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *CGA*, 2002.
- [2] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Gradient Profile Prior and Its Applications in Image Super-Resolution and Enhancement," *TIP*, 2011.
- [3] V. Katkovich, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *IJCV*, 2010.
- [4] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image Super-Resolution Via Sparse Representation," *TIP*, 2010.
- [5] W. Dong, D. Zhang, G. Shi, and X. Wu, "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization," *TIP*, 2011.
- [6] S. Wang, D. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 2216–2223.
- [7] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally Centralized Sparse Representation for Image Restoration," *TIP*, 2013.
- [8] J. Zhang, D. Zhao, F. Jiang, and W. Gao, "Structural Group Sparse Representation for Image Compressive Sensing Recovery," in *DCC*, 2013.
- [9] J. Zhang, D. Zhao, and W. Gao, "Group-Based Sparse Representation for Image Restoration," *TIP*, 2014.
- [10] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [11] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized non-local regularization for deconvolution and sparse reconstruction," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 253–276, 2010.
- [12] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation: Part i: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [13] J. A. Tropp, A. C. Gilbert, Martin, J. Strauss, J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," 2004.
- [14] Spams: a sparse modeling software. [Online]. Available: <http://spams-devel.gforge.inria.fr/doc/html/index.html>