IMAGE SUPER-RESOLUTION BASED ON DICTIONARY LEARNING AND ANCHORED NEIGHBORHOOD REGRESSION WITH MUTUAL INCOHERENCE

Yulun Zhang[1]*, Kaiyu Gu*[2]*, Yongbing Zhang*[1]*, Jian Zhang*[3]*, and Qionghai Dai*[1,4]

[1]Shenzhen Key Lab of Broadband Network and Multimedia,
Graduate School at Shenzhen, Tsinghua University, China
[2]Ningbo Vision 3D Display Technology Co., Ltd, China
[3]Institute of Digital Media, Peking University, China
[4]Department of Automation, Tsinghua University, China

ABSTRACT

In this paper, we employ unified mutual coherence between the dictionary atoms and atoms/samples when learning the dictionary and sampling anchored neighborhoods respectively for image super-resolution (SR) application algorithm. On one hand, an incoherence promoting term in dictionary learning for SR is introduced to encourage dictionary atoms, associated to different anchored regressors, to be as independent as possible, while still allowing for different regressors to share same samples. On the other hand, a unified form with mutual coherence between dictionary atoms and training samples is proposed when we group neighborhoods of samples centered on each atom and find the nearest neighbors for input samples in image super-resolution. Extensive experimental results on commonly used datasets demonstrate that our method outperforms state-of-the-art methods by obtaining compelling results with improved quality, such as sharper edges, finer textures and higher structural similarity.

Index Terms— Dictionary learning, mutual incoherence, neighbor embedding, super-resolution

1. INTRODUCTION

Single image super-resolution (SR), a classical and important task in computer vision, is the process to recover a high-resolution (HR) image from its down-scaled low-resolution (LR) version, while minimizing visual artifacts as much as possible.

During past decades, there are various SR algorithms, among which the learning-based SR is one of the main streams and represents the state-of-the-art methods recently. Neighbor embedding (NE) approach proposed by Chang *et al.* [1] assumes that the LR and HR patches have locally similar geometry on low-dimensional nonlinear manifolds, allowing LR input patch to be reconstructed by a linear combination of its nearest neighbors approximately. The same interpolation coefficients employed in LR patch reconstruction can be adopted to estimate an output patch in HR space. Since the mapping from HR to LR patches is many to one, the manifold assumption is not always true for single image SR. Yang *et al.* [2, 3] conducted SR via sparse coding (SC), where HR patch can be reconstructed as a sparse linear combination of the learned dictionary atoms based on the assumption that LR and HR patches share the same reconstruction sparse coefficients. This work was further developed by Zeyde *et al.* [4], where PCA and K-SVD [5] were used to reduce the dimension of the features within the LR patches and learn dictionary respectively for efficient learning. Yang *et al.* [6] learned simple functions from LR to HR patches by clustering a large number of natural image patches into a relatively small amount of subspace. Dong *et al.* [7] learned a deep convolutional neural network for image SR, which further improved the restoration quality comparing with state-of-the-art example-based methods. Timofte *et al.* [8] proposed a promising SR method named Adjusted Anchored Neighborhood Regression (A+) [8] by combining neighbor embedding and sparse coding SR.

In this paper, we further extend and improve the method in [8] by utilizing unified form of mutual coherence between dictionary atoms and atoms or training samples in the dictionary learning phase and sampling anchored neighborhoods in a unified optimized framework. We first introduce the mutual coherence of dictionary into the dictionary training phase using MI-KSVD [9] for SR in a more reasonable and optimized way. Then we use the mutual coherence rather than the Euclidean distance used in [8] between dictionary atoms and training samples to group anchored neighborhoods with less training time and better reconstruction quality. Experimental results on the commonly used datasets (e.g. **Set**5 [10] , **Set**14 [4] and **B**100 [11]) show that our proposed method obtains not only sharper edges, finer textures but also higher structural similarity comparing with state-of-the-art SR methods.

The remainder of the paper is organized as follows: In Section 2, we give a close look at recent works for the combination of neighbor embedding and sparse coding for SR and

dictionary learning with mutual incoherence. In Section 3, we present our proposed unified optimized SR framework with mutual incoherence. We report the experimental results and conclude the paper in Section 4 and Section 5 successively.

2. RELATED WORK

As we have briefly stated the NE and SC for SR, here we mainly focus on the combination of these two SR strategies. Moreover, we present the usage of mutual incoherence in dictionary learning.

The combination of NE and SC for SR was first used in Anchored Neighborhood Regression (ANR) [12], which is to anchor the neighborhood embedding of an LR patch to the nearest dictionary atom. For each atom, ANR [12] obtained projective matrix by computing its *K* nearest neighbors from other atoms in the same dictionary. This preprocessing improved SR speed significantly while the quality of the recovered image was no better than that of Zeyde *et al.* [4]. Soon after, Timofte proposed A+ [8], an improved version of ANR [12], with the quality being also improved distinctly. Different from ANR [12] learning the regressors on the dictionary, A+ [8] used the full training samples, from which Euclidean distance between dictionary atom and the training samples was used to compute *K* nearest neighbors for each dictionary atom. Both ANR [12] and A+ [8] learned dictionary based on Zeyde *et al.*'s algorithm, neglecting the mutual incoherence among atoms in the dictionary.

Dictionary learning with mutual incoherence is often used in classification and clustering. An incoherence promoting term in dictionary training encourages dictionary atoms to be as independent as possible. Ramirez *et al.* [13] proposed dictionary learning algorithms using L_2 norm to measure mutual incoherence and L_1 norm to enforce sparsity. Recently, Bo *et al.* [9] proposed MI-KSVD by adapting the well-known K-SVD [5] with the mutual incoherence of the dictionary being included in the reconstruction error.

Different from previous works this paper combines the advantages of NE, SC and MI-KSVD in a unified optimized SR framework. The distance measure we use is the absolute value of the inner product between dictionary atoms and atoms/samples all along the SR pipeline.

3. PROPOSED APPROACH

In this section, we first introduce dictionary learning with mutual incoherence, followed by sampling anchored neighborhoods based on the unified form of mutual coherence. Finally, a unified optimized SR framework is given in details.

3.1. Dictionary Learning with Mutual Incoherence

From randomly selected training patches $\boldsymbol{Y} = \left[\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N\right] \in \mathbb{R}^{S\times N}$, we learn the corresponding sparse coefficients $\boldsymbol{X} = \left[\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\right] \in \mathbb{R}^{M\times N}$ as well as the dictionary $\boldsymbol{D} = \left[\boldsymbol{d}_1, \cdots, \boldsymbol{d}_M\right] \in \mathbb{R}^{S\times M}$. The commonly used dictionary learning approach is to minimize the reconstruction error

$$\begin{aligned} &\min_{\boldsymbol{D},\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{D}\cdot\boldsymbol{X}\|_F^2 \\ &s.t.\ \forall m,\ \|\boldsymbol{d}_m\|_2 = 1\ and\ \forall n,\ \|\boldsymbol{x}_n\|_0 \le L, \end{aligned} \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, the L_0 norm counts the number of non-zero elements in the sparse coefficients $\boldsymbol{x}_n$, and L controls the sparsity level.

Maintaining large mutual incoherence helps to balance the roles of different types of image patches. Moreover, theoretical results from Candes *et al.* [14] have also indicted that it is much easier to recover the underlying sparse coefficients when the mutual incoherence of the dictionary is large.

Since we use mutual coherence between dictionary atoms and atoms or samples when grouping the anchored neighborhood regressors or finding the nearest neighbors, we balance the reconstruction error and the mutual coherence in the learning stage as follows:

$$\begin{aligned} &\min_{\boldsymbol{D},\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{D}\cdot\boldsymbol{X}\|_F^2 + \lambda_1 \sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M} \left|\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{d}_j\right| \\ &s.t.\ \forall m,\ \|\boldsymbol{d}_m\|_2 = 1\ and\ \forall n,\ \|\boldsymbol{x}_n\|_0 \le L. \end{aligned} \tag{2}$$

Here, $\lambda_1 \sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M} \left|\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{d}_j\right|$ is the mutual coherence to encourage large mutual incoherence of the learned dictionary and $\lambda_1 \ge 0$ is a weighting parameter. We solve the above optimization problem (2) via MI-KSVD proposed by Bo *et al.* [9].

3.2. Sampling Anchored Neighborhoods

Different from A+ [8] using Euclidean distance between the anchored dictionary atoms and the training samples when computing the corresponding regressors, we use the mutual coherence as distance measure with the same form mentioned above.

For a collection of training patches $\boldsymbol{Y} = \left[\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N\right] \in \mathbb{R}^{S\times N}$ and a learned dictionary $\boldsymbol{D} = \left[\boldsymbol{d}_1, \cdots, \boldsymbol{d}_M\right] \in \mathbb{R}^{S\times M}$, we sample anchored neighborhoods for each dictionary atom $\boldsymbol{d}_m$ by computing the coherence

$$\boldsymbol{C}_m = \left|\boldsymbol{Y}^{\mathrm{T}}\cdot\boldsymbol{d}_m\right| = \left[\left|\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{d}_m\right|, \cdots, \left|\boldsymbol{y}_N^{\mathrm{T}}\boldsymbol{d}_m\right|\right], \tag{3}$$

and selecting K nearest patches $\boldsymbol{N}_m = \left[\boldsymbol{y}_{m_1}, \cdots, \boldsymbol{y}_{m_K}\right]$ as its neighborhoods.

By using the mutual coherence between atoms and samples, we not only comply with the optimized way in dictionary learning but also speed up about 25% in computational time compared with that in A+ [8].

Table 1. PSNR and SSIM comparisons for the reconstructed images. PSNR and SSIM is measured for 3 different magnification factors ($\times 2$, $\times 3$, $\times 4$). The best result for each image and SR factor is highlighted.

		Bicubic		K-SVD [4]		ANR [12]		A+ [8]		SRCNN [7]		Ours	
Images	scale	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
butterfly	$\times 2$	27.44	0.9154	30.62	0.9527	30.42	0.9506	32.07	0.9647	32.21	0.9605	**32.24**	**0.9657**
foreman	$\times 2$	34.20	0.9519	36.46	0.9668	36.37	0.9668	37.09	0.9708	36.28	0.9679	**37.22**	**0.9711**
monarch	$\times 2$	32.94	0.9601	35.72	0.9728	35.66	0.9727	37.01	0.9767	**37.18**	0.9756	37.16	**0.9769**
ppt3	$\times 2$	27.10	0.9493	29.71	0.9735	29.39	0.9698	30.48	0.9796	**30.74**	0.9770	30.63	**0.9804**
crow	$\times 2$	31.62	0.9529	34.58	0.9700	33.99	0.9666	36.01	0.9759	36.03	0.9738	**36.16**	**0.9763**
fish	$\times 2$	35.09	0.9634	38.84	0.9795	38.86	0.9797	41.09	0.9849	40.33	0.9811	**41.28**	**0.9851**
plane	$\times 2$	39.62	0.9836	42.55	0.9885	42.37	0.9883	43.96	0.9901	43.75	0.9896	**44.07**	**0.9902**
Avg.	$\times 2$	32.57	0.9538	35.50	0.9720	35.30	0.9706	36.82	0.9775	36.65	0.9751	**36.97**	**0.9780**
butterfly	$\times 3$	24.04	0.8217	26.02	0.8782	25.87	0.8708	27.31	0.9094	**27.59**	0.9012	27.43	**0.9117**
foreman	$\times 3$	31.22	0.9059	33.23	0.9300	33.26	0.9299	34.37	0.9402	33.41	0.9322	**34.46**	**0.9403**
monarch	$\times 3$	29.43	0.9198	31.16	0.9385	31.09	0.9375	32.19	0.9474	**32.39**	0.9450	32.29	**0.9478**
ppt3	$\times 3$	23.90	0.8838	25.52	0.9209	25.32	0.9131	26.38	0.9399	26.32	0.9290	**26.48**	**0.9413**
crow	$\times 3$	28.62	0.9113	30.57	0.9349	30.16	0.9277	31.85	0.9483	31.62	0.9406	**32.02**	**0.9496**
fish	$\times 3$	31.34	0.9134	33.85	0.9428	33.98	0.9436	36.12	0.9614	34.86	0.9463	**36.31**	**0.9623**
plane	$\times 3$	36.66	0.9709	38.72	0.9777	38.52	0.9767	40.19	0.9816	39.79	0.9797	**40.38**	**0.9819**
Avg.	$\times 3$	29.32	0.9038	31.29	0.9318	31.17	0.9285	32.63	0.9469	32.28	0.9392	**32.77**	**0.9478**
butterfly	$\times 4$	22.10	0.7370	23.62	0.7977	23.53	0.7887	24.46	0.8397	**25.08**	0.8417	24.64	**0.8460**
foreman	$\times 4$	29.43	0.8666	31.14	0.8926	30.85	0.8877	32.25	0.9083	31.50	0.8973	**32.37**	**0.9107**
monarch	$\times 4$	27.46	0.8808	28.75	0.9013	28.70	0.9001	29.43	0.9126	**29.89**	0.9126	29.51	**0.9138**
ppt3	$\times 4$	22.14	0.8248	23.31	0.8636	23.10	0.8544	23.94	0.8870	24.06	0.8762	**24.07**	**0.8902**
crow	$\times 4$	26.99	0.8785	28.41	0.9017	28.16	0.8941	29.34	0.9192	29.19	0.9089	**29.49**	**0.9217**
fish	$\times 4$	29.16	0.8592	31.00	0.8944	31.13	0.8961	32.83	0.9241	31.58	0.8961	**32.98**	**0.9260**
plane	$\times 4$	34.89	0.9601	36.41	0.9675	36.22	0.9660	37.37	0.9723	37.27	0.9705	**37.62**	**0.9733**
Avg.	$\times 4$	27.45	0.8581	28.95	0.8884	28.81	0.8839	29.95	0.9090	29.80	0.9005	**30.10**	**0.9117**

3.3. A Unified Optimized SR Framework

Here, we present a unified optimized SR framework including several core steps in the SR pipeline. We follow the same preprocessing stage as that in Zeyde *et al.*'s algorithm, such as HR/LR image patch extraction, low-frequencies removal from HR patches, feature extraction and dimensionality reduction from LR patches. After accomplishing that, we obtain the training patch pairs $\boldsymbol{Y}_H = \left[\boldsymbol{y}_{H,1}, \cdots, \boldsymbol{y}_{H,N}\right]$ and $\boldsymbol{Y}_L = \left[\boldsymbol{y}_{L,1}, \cdots, \boldsymbol{y}_{L,N}\right]$.

$\boldsymbol{Y}_L$ is then used in formula (2) to train its corresponding LR dictionary $\boldsymbol{D}_L$ via MI-KSVD, minimizing the reconstruction error and pursuing large mutual incoherence of $\boldsymbol{D}_L$,which is reasonable and optimized for the next anchored neighborhood regression step.

For each LR dictionary atom $\boldsymbol{d}_m$, we select its K nearest LR patches from $\boldsymbol{Y}_L$ via formula (3) to group its LR neighborhoods $\boldsymbol{N}_{L,m}$. K HR patches with same positions from $\boldsymbol{Y}_H$ are selected to group the corresponding HR neighborhoods $\boldsymbol{N}_{H,m}$. These HR/LR neighborhoods are obtained with more consideration of structured mutual incoherence while still allowing some same patch pairs to be shared among different atom centered neighborhoods.

Same as Timofte *et al.*'s work about global regression [12], we use Collaborative Representation [15] to obtain the projection matrix from LR patch $\boldsymbol{y}_L$ to its HR patch $\boldsymbol{y}_H$ by first solving

$$\min_{\boldsymbol{x}} \|\boldsymbol{y}_L - \boldsymbol{N}_{L,m} \cdot \boldsymbol{x}\|_2^2 + \lambda_2 \|\boldsymbol{x}\|_2 , \tag{4}$$

where $\boldsymbol{N}_{L,m}$ is the LR neighborhood corresponding to the LR dictionary atom $\boldsymbol{d}_m$. With corresponding HR neighborhood $\boldsymbol{N}_{H,m}$ and the closed-form resolution of formula (4), the HR patch $\boldsymbol{y}_H$ can then be recovered via

$$\boldsymbol{y}_H = \boldsymbol{N}_{H,m} \cdot \boldsymbol{x} = \boldsymbol{P}_m \cdot \boldsymbol{y}_L, \tag{5}$$

where the corresponding projection matrix

$$\boldsymbol{P}_m = \boldsymbol{N}_{H,m}(\boldsymbol{N}_{L,m}^{\mathrm{T}}\boldsymbol{N}_{L,m} + \lambda_2 \boldsymbol{I})^{-1}\boldsymbol{N}_{L,m}^{\mathrm{T}} \tag{6}$$

can be computed offline.

Finally, for each input LR patch $\boldsymbol{y}_L$, the SR problem can be solved by searching its nearest atom $\boldsymbol{d}_{L,m}$ with the highest coherence $\left|\boldsymbol{y}_L^{\mathrm{T}} \cdot \boldsymbol{d}_{L,m}\right|$, followed by recovering the HR patch $\boldsymbol{y}_H$ via formula (5). A+ [8] used two different distance measures, Euclidean distance and inner product, in the process of neighborhoods sampling and nearest neighbors searching respectively and also failed to optimize the mutual incoherence of dictionary in the training phase, while our united optimized SR framework revises these deficiencies.

Fig. 1. Visual quality comparisons. "butterfly" image from **Set**5 with upscaling $\times$**2**; "ppt3" image from **Set**14 with upscaling $\times$**3**; "plane" image from **B**100 with upscaling $\times$**4**. (Zoom in for better view.)

4. EXPERIMENTAL RESULTS

For a fair comparison with traditional learning-based SR approaches, we use the same training set, testing sets and the standard parameters setting as in [8]. More specifically, the training set of 91 images as proposed by Yang *et al.* [2] is used. 3 datasets (**Set**5 [10] , **Set**14 [4] and **B**100 [11]) are used to evaluate the performance of upscaling factors 2, 3 and 4. A dictionary size of 1024 and a neighborhood size of 2048 are used to compute 1024 projection matrix from 5 million training samples. We set $L = 3$, $\lambda_1 = 0.03$ and $\lambda_2 = 0.1$ throughout the experiment. Extensive experiments are conducted to demonstrate the superiority of our proposed method, comparing with other state-of-the-art competing methods [4, 12, 8, 7] quantitatively and visually.

Table 1 shows numerical evaluations in terms of PSNR and SSIM values obtained by different methods. Due to space limitation, we present comparisons on 8 images, where "butterfly" image is from **Set**5, images "foreman", "monarch" and "ppt3" are from **Set**14, the rest images "crow", "fish", "fox" and "plane" with original names "42049", "210088", "167062" and "3096" respectively are from **B**100. We can observe that our method gets the highest values in SSIM than others, which means our reconstructed results achieve best structural similarity with the original HR images. In terms of PSNR, our method outperforms most of state-of-the-art ones and obtains best performance on average. The newly developed SR method SRCNN [7] using a deep convolutional network achieved highest PSNR values on some testing images, while it failed to enhance the structural similarity as well. When compared with ANR [12] and A+ [8], our method shows superiors quantitative performance with different upscaling factors 2, 3 and 4, which demonstrates the effectiveness and robustness of our unified optimized SR framework.

To further demonstrate the effectiveness of our SR method, we compare our visual results with other state-of-the-art methods with upscaling factors 2, 3 and 4 in Fig. 1. In the 3 testing images, obvious ringing artifacts along edges and annoying details appear in the results by the methods [4, 12, 7]. In the "plane" image, all other methods [4, 12, 8, 7] suffer from aliasing artifacts in the empennage of the plane. By contrast, our method recovers HR images more faithful to the ground truth with sharper edges, finer details and higher structural similarity.

5. CONCLUSION

In this paper, we propose a unified optimized SR framework using the same form of mutual coherence both in dictionary learning and anchored neighborhoods sampling. The mutual coherence used in our method is the absolute value of the inner product between dictionary atoms and atoms or training samples, which is not only conducted by an optimized way, but also save training time compared to A+ [8]. Experimental results validate the effectiveness and robustness of our approach quantitatively and visually.

6. ACKNOWLEDGEMENT

This work was partially supported by the National Natural Science Foundation of China under Grant 61170195, U1201255&U1301257.

7. REFERENCES

[1] H. Chang, D.Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *CVPR*, 2004.

[2] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *CVPR*, 2008.

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

[4] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, 2010.

[5] M. Aharon, M. Elad, and A.M. Bruckstein, "The k-svd: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[6] C.Y. Yang and M.H. Yang, "Fast direct super-resolution by simple functions," in *ICCV*, 2013.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*, 2014.

[8] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *ACCV*, 2014.

[9] L. Bo, X. Ren, and D. Fox, "Multipath sparse coding using hierarchical matching pursuit," in *CVPR*, 2013.

[10] M. Bevilacqua, A. Roumy, C. Guillemot, and M.L. Alberi Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *BMVC*, 2012.

[11] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *ICCV*, 2001.

[12] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *ICCV*, 2013.

[13] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *CVPR*, 2010.

[14] E. J. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[15] R. Timofte and L. Van Gool, "Adaptive and weighted collaborative representations for image classification," *Pattern Recognition Letters*, vol. 43, 2014.