# SUPER-RESOLUTION USING MULTIPLE STRUCTURED DICTIONARIES BASED ON THE GRADIENT OPERATOR

*Faezeh Yeganli, Mahmoud Nazzal and Huseyin Ozkaramanli*

Department of Electrical and Electronic Engineering
Eastern Mediterranean University
Famagusta, Mersin10, Turkey

## ABSTRACT

A new single-image super-resolution algorithm based on selective sparse representation over a set of coupled low and high resolution cluster dictionary pairs is proposed. Patch clustering and sparse model selection are carried out using the magnitude and phase of the patch gradient operator. Acceptable degrees of scale-invariance of these criteria are presented. A compact dictionary pair is learned for each cluster. A low resolution patch is classified into one of the clusters using these two criteria. A high resolution patch is reconstructed using the high resolution cluster dictionary, and the spare representation coefficients of its low resolution counterpart over the low resolution cluster dictionary. The proposed algorithm is shown to be competitive to the leading super-resolution algorithms in terms of PSNR, SSIM and computational cost.

***Index Terms***— Super-resolution, sparse representation, dictionary learning, clustering, gradient operator.

## 1. INTRODUCTION

Reconstructing a high resolution (HR) image estimate based on a low resolution (LR) image of the same scene is referred to as the single-image super-resolution (SR) problem. Various approaches exist in the literature [1] to solve for this ill-posed inverse problem.

Sparse representation over learned dictionaries has been widely used in several signal and image processing application areas. Sparsity is effectively employed as regularizer in the context of the sparse-representation SR framework. Given a (vector) signal $x \in R^n$ and an over-complete dictionary $D \in R^{n \times k}$, where $n$ is the dimension of the signal space and $k$ is the number of atoms in $D$, the sparse representation problem can be formulated as

$$\arg\min_{\alpha} \|x - D\alpha\|_2 \;\; subject\,to \;\; \|\alpha\|_0 < S, \qquad (1)$$

where $S$ denotes sparsity, and $\| \;\|_2$ denotes the vector Euclidean norm and $\| \;\|_0$ the number of nonzero elements in a vector. It is customary to obtain a dictionary $D$ by training over example signals in the dictionary learning (DL) process.

The representation power of a learned dictionary depends on its redundancy. Though, it has been reported that high redundancy tends to cause instabilities and degrade the representation quality [2]. From these observations learning class-dependent dictionaries has been considered recently. This includes the work of Dong et al. [3] where they applied K-means clustering to divide the training set into clusters, and learned compact cluster dictionaries. In [4], Feng et al. used K-subspace clustering to divide the training set into subspaces and obtained the dictionaries by extracting the shared bases in these subspaces. More recently, designing structural dictionaries like the work of Yu et al. [5] is considered. In [5], a structural dictionary is constructed as a composition of learned orthogonal bases. Yang et al. in [6] proposed a multiple geometric dictionary-based clustered sparse coding scheme. In this scheme, geometric dictionaries of geometric clusters are learned and image patches are sparsely coded over different geometric dictionaries.

Representation of image regions with sharp edges, corners and texture is crucial and most SR algorithms fail at faithfully reconstructing such salient features. These features are characterized by high gradient magnitudes. The gradient operator's magnitude is an important discriminating property for image patches [7]. The early usage of the gradient is related to the work by Rudin et al. in [8]. Recently, the invariance of the gradient histogram is argued and it is thus applied as a prior for inverse problems. An example is the work conducted by Sun et al. [9, 10] where a scale-invariant gradient profile prior is defined and used as a generic image prior for SR.

In this work, two natural patch measures are shown to be approximately scale-invariant, and thus used as clustering criteria. These are the sharpness measure (SM) and the dominant phase angle (DPA) defined via the magnitude and phase of the gradient operator, respectively. Training patches are first clustered using SM based on their spatial sharpness. Then, patches in each SM cluster are further clustered with the DPA measure based on their directionality. For each cluster, a pair of structured and compact LR and HR dictionaries is learned. In the reconstruction stage, each LR patch is classified into a certain cluster using the SM and DPA measures. The corre-

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

sponding HR patch is reconstructed by multiplying the cluster HR dictionary with the sparse representation coefficients of the underlying LR patch as coded over the LR dictionary of the same cluster. This clustering paradigm allows for super-resolving patches of insignificant high frequency components with bicubic interpolation, reducing the computational cost of the proposed algorithm. Simulation results in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [11] validate a competitive performance of the proposed algorithm compared to the leading SR algorithms.

This paper is organized as follows. Single-image SR via sparse representation is described in Section 2. Then, Section 3 details the proposed SR algorithm. Simulation results are demonstrated in Section 4 with the conclusions in Section 5.

## 2. SINGLE-IMAGE SUPER-RESOLUTION VIA SPARSE REPRESENTATION

With the availability of a HR dictionary $D_H$, a sparse approximation to a HR image patch $x_H$ can be written as

$$x_H \approx D_H \alpha_H, \tag{2}$$

where $\alpha_H$ is the sparse representation coefficient vector of $x_H$. Correspondingly, a LR patch of the same scene $x_L$ can have a sparse approximation with the availability of a LR dictionary $D_L$, as follows

$$x_L \approx D_L \alpha_L, \tag{3}$$

where $\alpha_L$ is the sparse representation coefficient vector of $x_L$. Considering $\Psi$ as a blurring and downsampling operator, $x_L$ can be related to $x_H$ as $x_L \approx \Psi x_H$. It can be further assumed that $D_L \approx \Psi D_H$ if $D_L$ and $D_H$ are learned in a coupled manner. From the above assumptions, one can obtain

$$x_L \approx \Psi x_H \approx \Psi D_H \alpha_H \approx D_L \alpha_H. \tag{4}$$

In view of (4), it can be concluded that $\alpha_H \approx \alpha_L$ and this forms the foundation for the HR patch reconstruction.

## 3. THE PROPOSED SUPER-RESOLUTION ALGORITHM

In this section, we present a 2-level clustering scheme using SM and DPA. First, patches are clustered into three main SM clusters ($C_1$, $C_2$ and $C_3$) corresponding to SM interval of [0, 10], [10, 20] and [20, 255]. It is found empirically that such intervals classify patches into un-sharp, moderately sharp and very sharp. Second, patches in each SM cluster are clustered with the DPA measure into five clusters $C^0$, $C^{45}$, $C^{90}$, $C^{135}$ and $C^{nd}$. $C^0$ through $C^{135}$ correspond to DPA values of 0°, 45°, 90° and 135°, respectively, and $C^{nd}$ is non-directional. Totally, fifteen clusters $C_1^0$ through $C_3^{nd}$ are obtained as depicted in Fig. 1. In this notation, subscripts denote the SM cluster, and the superscript represents the DPA cluster. This setting is found to work successfully at a reasonable computational cost. Still, more SM and DPA clusters can be considered, covering more sharpness levels and orientations.

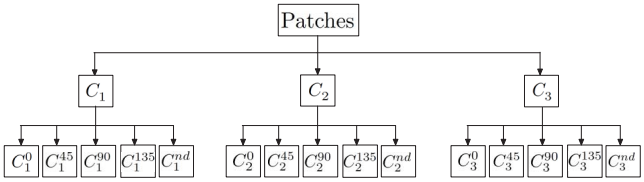


**Fig. 1**: The proposed SM and DPA 2-level clustering scheme.

### 3.1. Approximate scale-invariance of the image patch sharpness and dominant gradient phase angle measures

SM can be defined via the magnitude of the gradient operator [12] as

$$SM = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sqrt{\mid G_i^h \mid + \mid G_j^v \mid}, \tag{5}$$

where $G^h$ and $G^v$ are the horizontal and vertical gradients and $N_1$ and $N_2$ are the patch dimensions, respectively. SM is used to classify edges, corners and textures in images based on their spatial intensity variations.

Sun et al. in [9, 10] defined a gradient profile prior and studied its behavior with respect to image scale. Based on their studies, the edge sharpness of images is independent of image resolution.

It is well-known that the phase of a quantity is more informative than its magnitude. The phase matrix of the gradient operator is calculated based on the horizontal $G^h$ and vertical $G^v$ gradients as follows

$$\Phi = \arctan(\frac{G^v}{G^h}). \tag{6}$$

Intuitively, patch directionality can be characterized in terms of the angles in the phase matrix of the gradient operator. In this work, we quantize the angles in the gradient's phase matrix into 0°, 45°, 90° and 135°. Other values may still be considered. If a value is repeated more than half of the number of angles in the phase matrix, then this value is determined as the patch's DPA. Otherwise, the patch is said to be non-directional.

To investigate the impact of scale on SM and DPA, the following experiment is conducted on each of the test images shown in Fig. 2. Each HR image is divided into non-overlapping 6×6 patches. Patches are extracted without overlapping to maintain the correspondence between each HR patch and its LR counterpart. A LR image is obtained by applying a bicubic filter on the HR image and downsampling it with a scale factor of 2. Interpolating the LR image by a scale factor of 2, gives the middle resolution (MR) image.

| Image | $C_1$ | | | | | $C_2$ | | | | | $C_3$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $C_1^0$ | $C_1^{45}$ | $C_1^{90}$ | $C_1^{135}$ | $C_1^{nd}$ | $C_2^0$ | $C_2^{45}$ | $C_2^{90}$ | $C_2^{135}$ | $C_2^{nd}$ | $C_3^0$ | $C_3^{45}$ | $C_3^{90}$ | $C_3^{135}$ | $C_3^{nd}$ |
| Barbara | 421 | 175 | 252 | 167 | **2671** | 528 | 214 | 150 | 128 | 563 | 870 | 191 | 64 | 213 | 618 |
| | 61.28 | 72.00 | 88.89 | 77.84 | **76.19** | 40.91 | 21.96 | 61.33 | 59.38 | 13.50 | 4.60 | 1.57 | 57.81 | 5.16 | 19.74 |
| Butterfly | 29 | 57 | 54 | 10 | **683** | 21 | 48 | 76 | 11 | 173 | 76 | 119 | 139 | 26 | 242 |
| | 62.07 | 68.42 | 83.33 | 60.00 | 73.94 | 33.33 | 60.42 | 65.79 | 54.55 | 36.42 | **81.58** | 68.07 | 76.26 | 73.08 | 31.82 |
| Foreman | 78 | 130 | 185 | 62 | **1039** | 12 | 52 | 66 | 13 | 63 | 0 | 33 | 11 | 5 | 15 |
| | 65.38 | 85.38 | 79.46 | 82.26 | **81.14** | 25.00 | 65.38 | 48.48 | 69.23 | 20.63 | N.A. | 36.36 | 36.36 | 0.00 | 0.00 |
| house | 12 | 14 | 175 | 10 | **1147** | 41 | 12 | 109 | 12 | 66 | 38 | 20 | 65 | 16 | 27 |
| | 66.67 | 50.00 | 24.57 | 50.00 | **86.84** | 73.17 | 75.00 | 66.06 | 41.67 | 30.30 | 50.00 | 90.00 | 24.62 | 50.00 | 25.93 |
| Lena | 489 | 74 | 100 | 193 | **4896** | 298 | 65 | 49 | 195 | 507 | 107 | 32 | 8 | 86 | 126 |
| | 82.21 | 59.46 | 68.00 | 75.13 | **80.60** | 58.39 | 55.38 | 57.14 | 63.08 | 18.15 | 24.30 | 46.88 | 50.00 | 48.84 | 7.94 |
| ppt3 | 88 | 213 | 244 | 151 | **6331** | 169 | 66 | 238 | 47 | 430 | 379 | 67 | 332 | 101 | 736 |
| | 38.64 | 49.77 | 63.52 | 43.71 | **91.31** | 73.37 | 75.76 | 60.08 | 55.32 | 34.42 | 60.69 | 67.16 | 74.40 | 61.39 | 41.85 |
| Text Image 1 | 0 | 0 | 0 | 0 | 5188 | 0 | 0 | 0 | 0 | 490 | 288 | 1 | 8 | 3 | **5358** |
| | N.A. | N.A. | N.A. | N.A. | 47.96 | N.A. | N.A. | N.A. | N.A. | 12.65 | 11.46 | 0.00 | 12.50 | 0.00 | **18.51** |
| Texture | 48 | 12 | 2 | 0 | 412 | 272 | 59 | 0 | 4 | 515 | **6282** | 1248 | 72 | 93 | 2217 |
| | 47.92 | 41.67 | 0.00 | N.A. | 52.18 | 44.49 | 35.59 | N.A. | 0.00 | 25.83 | **68.48** | 74.12 | 80.56 | 70.97 | 41.00 |
| Average | 56.51 | 61.89 | 68.69 | 66.42 | 83.09 | 48.08 | 43.84 | 38.24 | 62.83 | 20.31 | 50.23 | 59.40 | 39.70 | 46.21 | 28.68 |

**Table 1**: Number of HR patches in each cluster (top) and scale-invariance ratios of SM and DPA (bottom). The largest number of patches in a cluster with the corresponding percentage is in bold face.

Clustering is done with SM and DPA in two levels as shown in Fig. 1. The scale-invariance of SM and DPA is defined as the ratio between the total number of MR patches correctly classified into a certain cluster and total number of HR patches in that cluster. Invariance ratios are listed in Table 1. In view of Table 1, for images with sharp details such as textures and edges, SM and DPA are strongly scale-invariant in the clusters containing the apriority of patches. For images such as Lena and Foreman, scale-invariance is not as bold.


**Fig. 2**: Test images: Barbara, Butterfly, Foreman, House, Lena, ppt3, Text Image 1 and Texture.

### 3.2. Clustering and sparse model selection with the patch sharpness and dominant gradient phase angle

The proposed algorithm consists of two stages of training and reconstruction. In the training stage, a set of coupled cluster dictionaries is learned and in the reconstruction stage the best dictionary pair is selected to reconstruct HR patches from the corresponding LR patches.

Based on its SM and DPA values, each LR patch is classified into an appropriate cluster and the corresponding HR patch is placed into the same cluster. The classified LR and HR patch pairs are used to learn a coupled LR and HR dictionary pair using the method proposed in [1]. Algorithm 1 summarizes the main steps of the training stage.

In the reconstruction stage, the SM and DPA values of each MR patch are calculated to cluster it. The dictionary pair of the identified cluster is employed to calculate the sparse representation coefficients of the corresponding MR patch over the cluster LR dictionary. Then, by right multiplying the cluster HR dictionary with the sparse representation coefficient vector of the LR patch, the HR patch is reconstructed. Finally all reconstructed HR patches are reshaped into the two-dimensional form and merged to constitute a HR image estimate. The proposed reconstruction algorithm is summarized in Algorithm 2.

**Algorithm 1** The Proposed Cluster DL Algorithm.

1: **INPUT:** HR Training Image Set.
2: **OUTPUT:** Directional Cluster Dictionary Pairs with varying sharpness.
3: Divide each HR image into patches and subtracting the mean value.
4: Vectorize patches and column-stack them to form a HR training array.
5: Blur and downsample each HR image to generate a LR image.
6: Divide each LR image into patches.
7: Upsample each LR image to the MR level.
8: Apply feature extraction filters (first and second order gradient) on each MR image.
9: Divide the extracted features into patches and vectorize them.
10: Combine the features column-wise to form the LR training array.
11: **for** Each patch in the LR training array, **do**
12: Calculate SM and DPA of the MR patch, and find the cluster number.
13: Add the MR patch to the LR training set of this cluster.
14: Add the respective HR patch to the HR training set of this cluster.
15: **end for**
16: For each cluster, learn a pair of coupled dictionaries.

## 4. EXPERIMENTAL VALIDATION

In this section, the proposed algorithm is compared to the SR algorithms of Yang et al. [1], Peleg et al. [13] and He et al. [14]. These algorithms are simulated with source codes provided by the authors.

Dictionaries of the algorithms of Peleg et al. [13] and He

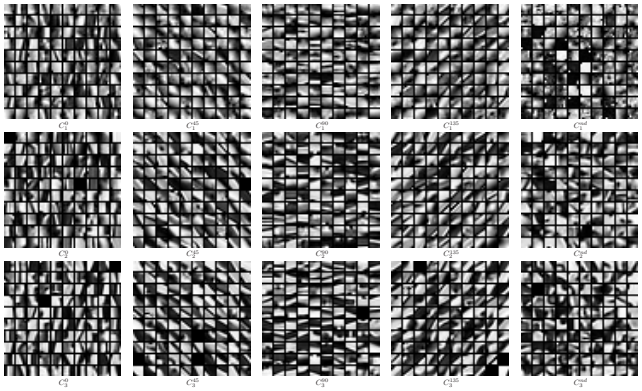

**Fig. 3**: Reshaped example atoms of the fifteen designed HR cluster dictionaries.

**Algorithm 2** The Proposed Single-Image SR Algorithm.

1: **INPUT:** A LR Test Image, Cluster Dictionary Pairs.
2: **OUTPUT:** A HR Image Estimate
3: Divide the LR image into overlapping patches.
4: Upsample the LR image to the MR resolution level.
5: Apply feature extraction filters (first and second order gradient) on the MR image.
6: Divide the extracted features into patches and vectorize them.
7: **for** Each MR patch **do**
8: Calculate SM and DPA of the MR patch and determine the cluster.
9: Sparsely code the MR patch features over the cluster LR dictionary.
10: Reconstruct the corresponding HR patch by multiplying the HR dictionary of the same cluster with the sparse codes of the MR features.
11: **end for**
12: Merge overlapping patches to obtain a HR image estimate.

et al. [14] are used as provided by the authors. 40,000 patch pairs are sampled from the Flicker dataset [15] and used to learn a single 1000-atom dictionary pair for the algorithm of Yang et al. [1]. 40,000 patch pairs are randomly sampled from the same dataset in each cluster and used to learn a cluster dictionary pair for the proposed algorithm. The dictionary size for each cluster is set to $36\times600$. This size is empirically found to give good results at suitable computational cost. Fig. 3 shows example reshaped atoms from the designed cluster doctrinaires. It is notable that the cluster dictionaries inherit the sharpness levels of their SM clusters and the directional natures of their DPA clusters.

The images used in this test are shown in Fig. 2. It is noted that SR is only applied to the luminance color component in case of color images, while the chrominance components are super-resolved with bicubic interpolation, in accordance with the common practice in the literature. Comparisons are conducted in terms of the PSNR and SSIM measures. For the case of color images, PSNR is calculated between the luminance components of the images, while SSIM is calculated as the average of its values calculated between each of the luminance and chrominance components.

Table 2 lists the PSNR (top) and SSIM (bottom) values of the aforementioned algorithms along with bicubic interpolation. The proposed algorithm's performance is denoted by ($P$). It is noted that the proposed algorithm is superior to or at least competitive with the others. It has average PSNR improvements of 0.3 dB, 0.4 dB and 0.39 dB over the algorithms of He et al. [14], Peleg et al. [13] and Yang et al. [1], respectively. The proposed algorithm uses a total of fifteen 600-atom cluster dictionary pairs. Compared to the algorithm of Yang et al. [1] which uses a single 1000-atom dictionary pair, using multiple dictionaries of smaller size is promising to reduce the computational complexities of dictionary learning and sparse representation, as compared to the case of a single highly redundant dictionary.

| Image | Bicubic | Yang et al. [1] | Peleg et al. [13] | He et al. [14] | $P$ | $P_{bic}$ |
|---|---|---|---|---|---|---|
| Barbara | 25.35 | 25.86 | 25.76 | 25.84 | 25.86 | **25.87** |
| | 0.7930 | 0.8357 | 0.8359 | **0.8372** | 0.8353 | 0.8347 |
| Butterfly | 27.46 | 31.26 | 30.96 | 31.44 | **32.07** | 31.76 |
| | 0.8985 | 0.9457 | 0.9227 | 0.9463 | **0.9511** | 0.9461 |
| Foreman | 35.36 | 38.50 | 38.20 | 38.45 | **38.66** | 38.27 |
| | 0.9469 | 0.9639 | 0.9644 | 0.9641 | **0.9642** | 0.9617 |
| house | 32.76 | 34.71 | 34.49 | 34.84 | **35.17** | 34.99 |
| | 0.8928 | 0.9129 | 0.9136 | 0.9145 | **0.9154** | 0.9136 |
| Lena | 34.71 | 36.36 | **36.59** | 36.58 | 36.44 | 36.34 |
| | 0.8507 | 0.8631 | 0.8387 | **0.8647** | 0.8633 | 0.8628 |
| ppt3 | 26.85 | 29.68 | 29.71 | 29.79 | **30.51** | 30.35 |
| | 0.9372 | 0.9604 | 0.9494 | 0.9621 | **0.9646** | 0.9618 |
| Text Image 1 | 17.52 | 18.58 | 18.73 | 18.53 | **18.89** | 18.77 |
| | 0.7246 | 0.7974 | 0.8118 | 0.7950 | **0.8124** | 0.8012 |
| Texture | 20.64 | 22.55 | 22.97 | 22.78 | **23.05** | 23.02 |
| | 0.8272 | 0.8939 | 0.9032 | 0.8997 | **0.9040** | 0.9029 |
| Average | 27.58 | 29.69 | 29.68 | 29.78 | **30.08** | 29.92 |
| | 0.8589 | 0.8966 | 0.8925 | 0.8980 | **0.9013** | 0.8981 |

**Table 2**: PSNR and SSIM comparisons. The best PSNR and SSIM results are in bold.

The computational complexity of the proposed algorithm can be further reduced by using bicubic interpolation to super-resolve un-sharp patches. First, one can notice that the SM cluster $C_1$ contains un-sharp patches with insignificant high frequency components. This suggests the possibility of using bicubic interpolation to super-resolve such patches. Following the same logic, patches in $C_2^{nd}$ are non-directional and can therefore be effectively super-resolved with bicubic interpolation. In contrast, $C_3^{nd}$ is located in the sharpest SM cluster $C_3$, and its patches have significant high frequency components. So, bicubic interpolation will not effectively reconstruct patches in $C_3^{nd}$.

| Image | Barbara | Butterfly | Foreman | House | Lena | ppt3 | Text image 1 | Texture | Average |
|---|---|---|---|---|---|---|---|---|---|
| Perc. | 15.3 % | 41.3 % | 9.2 % | 16.6% | 10.3% | 21.7% | 40.4% | 84.3% | 29.89% |
| $P$ Run Time | 569.40 | 133.20 | 131.80 | 126.50 | 531.70 | 542.00 | 692.40 | 799.30 | 440.79 |
| $P_{bic}$ Run Time | 216.20 | 72.10 | 50.40 | 56.40 | 193.40 | 300.50 | 444.00 | 632.40 | 245.68 |

**Table 3**: Percentage of patches handled with sparse representation in scenario $P_{bic}$, and SR run times of $P$ and $P_{bic}$.

In view of Table 1, $C_1$ contains a high percentage of image patches. Besides, the scale-invariance of SM and DPA in $C_2^{nd}$ is particularly weaker than in the other clusters. Therefore, applying bicubic interpolation to patches in $C_1$ and $C_2^{nd}$, while the sparse representation-based SR is only applied to clusters $C_2^0$ though $C_2^{135}$ and $C_3$ is expected to substantially reduce the SR computational complexity without sacrificing the reconstruction quality. This scenario is also studied and

denoted by ($P_{bic}$) in Table 2. With this scenario, the proposed algorithm is still generally superior to the other algorithms with PSNR improvements of 0.14 dB, 0.24 dB and 0.23 dB over the algorithms of He et al. [14], Peleg et al. [13] and Yang et al. [1], respectively. SSIM values come inline with the PSNR performances.

To analyze the reduction in computational complexity in scenario $P_{bic}$ compared to scenario $P$, Table 3 shows the percentage of patches for which the sparse representation-based SR is applied out of the total number of patches for each image, as denoted by ($Perc.$). On average, only 29.89 % of patches are super-resolved with the sparse representation framework while the others are handled with bicubic interpolation. Table 3 also shows the super-resolution execution times for each image with scenarios $P$ and $P_{bic}$. Simulations are conducted on an Intel Core i3-3220 3.30 GHz PC under Matlab R2013a environment. The average execution time for scenario $P$ is 440.79 seconds, while it is 245.68 seconds for scenario $P_{bic}$.

The proposed algorithm can be extended to handle noisy images. However, the presence of noise will affect the estimation of the SM and DPA measures. Therefore, this extension requires further research.

## 5. CONCLUSION

A new single-image super-resolution algorithm based on multiple structured cluster dictionary pairs is proposed. Clustering is done with the magnitude and the phase of the gradient operator of image patches. This is done by classifying patches based on their spatial sharpness using the magnitude and then sub-clustering these patches based on their directionality using the phase of the gradient operator. For each cluster, a pair of compact dictionaries is designed. For each low resolution patch, a cluster is selected using these measures, and the cluster dictionary pair is used to reconstruct a high resolution patch. The proposed clustering allows for applying bicubic interpolation to super-resolve patches of insignificant high frequency components, thereby reducing the computational complexity of the proposed algorithm. Besides, one can afford to design compact cluster dictionaries. Numerical experiments validate that the proposed algorithm is competitive with the leading super-resolution algorithms in terms of PSNR and SSIM, at a competitive computational complexity.

## REFERENCES


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

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

[3] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838-1857, 2011.

[4] J. Feng, L. Song, X. Yang, and W. Zhang, "Learning dictionary via subspace segmentation for sparse representation," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 1245-1248, 2011.

[5] G. Yu, G. Sapiro, and S. Mallat S., "Image modeling and enhancement via structured sparse model selection," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 1641-1644, 2010.

[6] S. Yang, M. Wang, Y. Chen, and Y. Sun, "Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding," *IEEE Trans. Image Process.*, vol.21, no. 9, 4016-4028, 2012.

[7] J. Kumar, F. Chen, and D. Doermann, "Sharpness estimation for document and scene images," in *Proc. Int. Conf. Pattern Recog (ICPR)*, pp. 3292-3295, 2012.

[8] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259-268, 1992.

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

[10] J. Sun J., Z. Xu, and H. Y. Shum, "Gradient profile prior and its applications in image super-resolution and enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1529-1542, 2011.

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

[12] R. C. Gonzalez, and R. E. Woods, *Digital image processing*, 2nd edn. Prentice-Hall, Inc., Englewood Cliffs, NJ, 2002.

[13] T. Peleg, and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2569-2582, 2014.

[14] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 345-352, 2013.

[15] http://see.xidian.edu.cn/faculty/wsdong/wsdong_downloads.htm