Accepted and to appear in IEEE MMSP 2013

Self-Learning-Based Low-Quality Single Image Super-Resolution

Li-Wei Kang[1], Boqi Zhuang[2], Chih-Chung Hsu[2], Chia-Wen Lin[2,*], and Chia-Hung Yeh[3]

[1]*Department of Computer Science and Information Engineering, National Yunlin University of Science & Technology, Taiwan*

[2]*Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan*

[3]*Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan*

[*]cwlin@ee.nthu.edu.tw

Abstract—**Low-quality images are usually not only with low-resolution, but also suffer from compression artifacts, e.g., blocking artifacts. Directly performing image super-resolution (SR) to a low-quality image would also simultaneously magnify the blocking artifacts, resulting in unpleasing visual quality. In this paper, we propose a self-learning-based SR framework to simultaneously achieve low-quality single-image SR and compression artifact removal (deblocking is treated as an example in this work). We argue that individually performing deblocking first, followed by SR to an image, would usually inevitably lose some image details induced by deblocking, which may be useful for SR, resulting in worse SR result. In our method, we propose to self-learn image sparse representation for modeling the relationship between low and high-resolution image patches in terms of the learned dictionaries, respectively, for image patches with and without blocking artifacts. As a result, image SR and deblocking can be simultaneously achieved via sparse representation and MCA (morphological component analysis)-based image decomposition. Experimental results demonstrate the efficacy of the proposed algorithm.**

I. Introduction

With the rapid development of multimedia and network technologies, delivering and sharing multimedia contents through the Internet and heterogeneous devices has become more and more popular. However, limited by the channel bandwidth and storage capability, most images distributed over the Internet exist in low quality versions degraded from the sources. The most common image degradations are downscaling and compression. Downscaling exploits the spatial redundancy in an image while compression further exploits the correlation in the frequency and temporal (for video case) domains. Even if the quality degradation greatly reduces the required bandwidth and storage for images, it would also leads to significant information loss and unpleasing visual artifacts, which usually behaves as blocking, ringing, or blurring [1].

There has been a great demand for improving the perceptual quality of images in terms of the spatial resolution enhancement of an image, denoted by image super-resolution (SR). The goal of image SR is to recover a high-resolution (HR) image from one or multiple low-resolution (LR) input images, which is essentially an inverse and ill-posed problem. There are mainly two categories of approach for image SR: (i) classical approaches and (ii) example/learning-based approaches. In the classical approaches, one sub-category is reconstruction-based schemes, where a set of LR images of the same scene are aligned with sub-pixel accuracy to generate an HR image [2]. Such kind of approaches is usually time-consuming and impractical due to multiple input LR images are required. The other sub-category of the classical is frame interpolation [3], which has been shown to generate overly smooth images with ringing and jagged artifacts.

The example/learning-based methods [4]–[10] hallucinate the high frequency details of an LR image based on the co-occurrence prior between LR and HR image patches in a training set, which has been shown to outperform the classical approach. More specifically, for an LR input, example-based methods [4]–[6] search for similar image patches from a pre-collected training LR image dataset or the same image itself based on self-examples, and use their corresponding HR versions to produce the final SR output. Nevertheless, the HR details hallucinated by such kind of approaches cannot be guaranteed to provide the true HR details. Hence, the performance of this approach highly relies on the similarity between the training set and test set or the self-similarity in the image itself. Moreover, learning-based SR approaches [7]–[10] focus on modelling the relationship between different resolutions of images. For example, Yang *et al.* [7] proposed to apply sparse coding techniques to learn a compact representation for HR/LR patch pairs for SR based on pre-collected HR/LR image pairs. Then, Yang *et al.* [8] advances [7] to propose a coupled dictionary training approach for SR based on patch-wise sparse recovery, where the learned couple dictionaries relate the HR/LR image patch spaces via sparse representation. It is guaranteed that the sparse representation of an LR image patch can well reconstruct its underlying HR image patch, which cannot be guaranteed in [7]. In addition, self-learning frameworks based on self-similarity of an image were introduced for SR in [9], [10].

However, the above-mentioned SR approaches only consider that an input LR image is only degraded by down-scaling or at most an additional blurring operation (with known or well-estimated blurring kernel). It is not always practical in a network environment, where image compression is usually adopted, which yields additional compression artifacts such as blocking and ringing. For image search engines, compression helps reduce the image size by up to 50% without obvious perceptual quality loss presented in the LR form of an image. Nevertheless, if SR is directly performed on the LR image, compression artifacts will be simultaneously magnified and therefore the perceptual quality of resulting HR image would be poor [11]. Hence, a high-performance SR scheme for low-quality image is desirable for enhancing the resolution of image/video degraded by both down-scaling and compression. In [11], a unified framework was proposed to simultaneously improve the resolution and perceptual quality of low-quality web image/video degraded by down-scaling and compression. This approach combines adaptive regularization and learning-based SR, where the regularization strength is determined by the JPEG compression quality factor (QF) of an input image.

On the other hand, an example-based SR algorithm of compressed videos in DCT (discrete cosine transform) domain was proposed in [12]. Input to the system is a compressed LR video together with a HR still image of similar content which is assumed to be available in advance. In addition, a unified framework achieving simultaneous denoising and SR for noisy video was proposed in [13]. This approach assesses the visual quality with respect to fidelity preserving, detail preserving, and spatial-temporal smoothness.

Existing low-quality image/video SR approaches [11]–[13], however, are all designed for some special purposes. For example, the framework proposed in [11] is mainly designed for JPEG compressed web image with known QF. Moreover, most other related approaches are designed for video [12], [13]. In addition, strictly speaking, existing low-quality image SR methods usually rely on two cascading operations, namely, a denoising-like operation, followed by SR operation. For example, in [11], an input LR image is first regularized via PDE (partial differential equations) regularization (similar to denoising), followed by bicubic interpolation and then, enhanced by learning-based pair matching. It is therefore essentially equivalent to separately performing denoising first, followed by SR/interpolation. We find that such cascaded operation (i.e., desnoising followed by SR), would usually inevitably lose some image details induced by desoising (e.g., deblocking), thereby degrading the performance of SR.

In this paper, we propose a self-learning-based sparse representation framework to simultaneously achieve low-quality single image SR and blocking artifact removal. Blocking artifact is one of the major visually unpleasing effects induces by transform block-based compression and hence, deblocking has been extensively studied in the literature [1], [14]–[16]. The main contribution of this paper is three-fold: i) to the best of our knowledge, we are among the first to propose a unified framework to simultaneously remove blocking artifacts and perform single image SR via sparse coding-based image decomposition; ii) the proposed framework can be adapted to any block transform-based compressed image/video without the need of any prior knowledge from data source, coding bitrates, and compression algorithms; and iii) the learning process in the proposed framework is self-contained, where no extra training samples are required. The rest of this paper is organized as follows. In Sec. II, we briefly review some related works. Sec. III presents the proposed single image SR framework for low-quality image. In Sec. IV, experimental results are demonstrated. Finally, Sec. V concludes this paper.

II. Related Works

A. MCA-based Image Decomposition

Suppose an image I of N pixels is a superposition of S layers (called morphological components), denoted by $I=\sum_{s=1}^{S} I_s$, where I_s denotes the s-th component, such as the geometric or textural component of I. To decompose the image I into $\{I_s\}_{s=1}^{S}$, the MCA algorithms [16]-[18] iteratively minimize the following energy function:

$$E(\{I_s\}_{s=1}^{S},\{\theta_s\}_{s=1}^{S})=\frac{1}{2}\left\|I-\sum_{s=1}^{S} I_s\right\|_2^2+\tau\sum_{s=1}^{S} E_s(I_s,\theta_s),\quad (1)$$

where θ_s denotes the sparse coefficients corresponding to I_s with respect to the dictionary D_s, τ is a regularization parameter, and E_s is the energy function defined according to the type of D_s (global or local dictionary). In this work, MCA-based image decomposition is integrated in the proposed SR framework mainly for image deblocking.

B. Sparse Representation and Dictionary Learning

Sparse coding [7], [8] is a technique of finding a sparse representation for a signal with a small number of nonzero or significant coefficients corresponding to the atoms in a dictionary. To construct a dictionary D_s to sparsely represent each patch b_s^k extracted from the component I_s of the image I, we may use a set of available training exemplars $y^k, k=1,2,\dots,P$, to learn a dictionary D_s sparsifying y^k by solving the following optimization problem:

$$\min_{D_s,\theta^k}\sum_{k=1}^{P}\left(\frac{1}{2}\|y^k-D_s\theta^k\|_2^2+\lambda\|\theta^k\|_1\right),\qquad (2)$$

where θ^k denotes the sparse coefficient vector of y^k with respect to D_s and λ is a regularization parameter.

III. Proposed Single-Image-based Super-Resolution Framework for Low-Quality Images

Fig. 1 depicts the proposed self-learning-based SR framework for low-quality single image, formulated as an MCA-based image decomposition problem via sparse representation. In our method, an input LR image I [see Fig. 1(a)] with blocking artifacts and its down-scaled version I^d [see Fig. 1(b)] are first roughly decomposed into the corresponding low-frequency (LF) parts, I_{LF} and I_{LF}^d, and the high-frequency (HF) parts, I_{HF} and I_{HF}^d, using the BM3D (block-matching and 3D filtering) algorithm [19], respectively, where the respective most basic information will be retained in the LF parts while the blocking artifacts and the other edge/texture details will be included in the HF parts of the images, as illustrated in Figs. 1(c) and 1(d) and Figs. 1(e) and 1(f), respectively. Then, we classify all of the patches (denoted by $x_i\in\mathcal{X}$, $\mathcal{X}$ is the HR patch set) in I_{HF} together with their corresponding patches (denoted by $y_i\in\mathcal{Y}$, $\mathcal{Y}$ is the LR patch set) in I_{HF}^d into two clusters of "blocking" and "non-blocking" HR/LR patch pairs, denoted by $\{\mathcal{X}^B,\mathcal{Y}^B\}$ and $\{\mathcal{X}^N,\mathcal{Y}^N\}$, respectively, based on the blocking artifact detection algorithm proposed in [14]. Based on the two training sets of patch pairs extracted from the input image

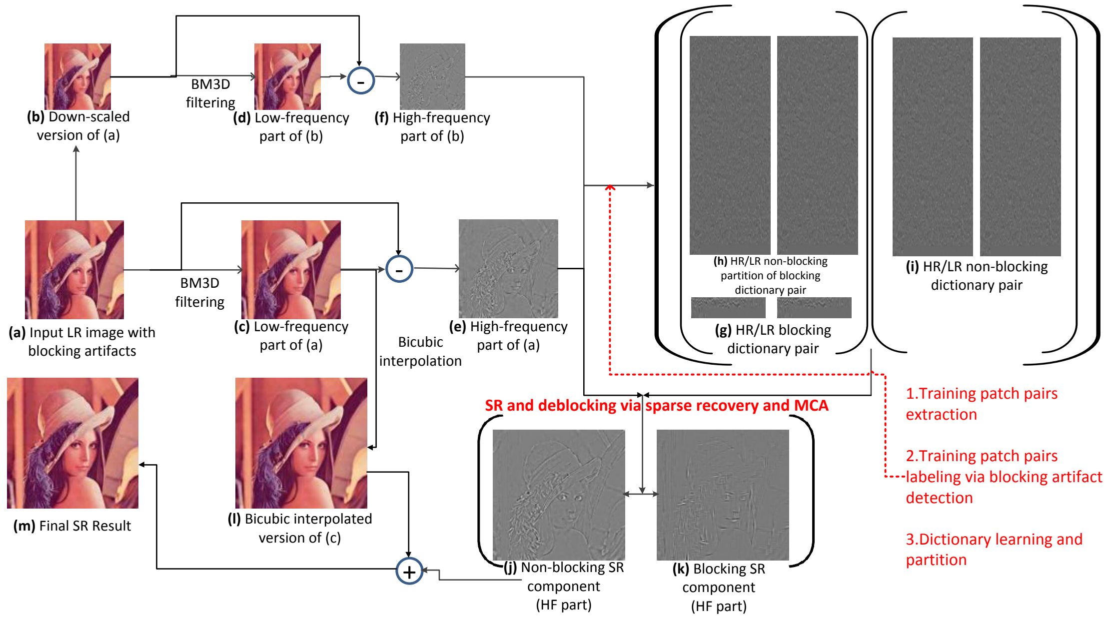

Fig. 1. Flowchart of the proposed super-resolution framework for low-quality single image via sparse representation and MCA-based image decomposition.

itself, we apply the coupled dictionary training method proposed in [8] to learn two sets of coupled dictionaries, D_B and D_N, used for SR of blocking and non-blocking patches, respectively, as illustrated in Figs. 1(g) and 1(h), and Fig. 1(i).

To achieve the SR of I_{HF}, we perform patch-wise sparse reconstruction with the coupled dictionary set D_N, consisting of a pair of dictionaries, D_N^{HR} and D_N^{LR}, of corresponding HR/LR atoms, respectively [see Fig. 1(i)], for each patch without blocking artifacts in I_{HF}. For each patch with blocking artifacts in I_{HF}, we perform SR reconstruction with D_B, consisting of D_B^{HR} and D_B^{LR}, of corresponding HR/LR atoms, respectively, and MCA-based image decomposition to obtain the underlying HR patch. Based on MCA-based dictionary partitioning [16], [18], D_B^{HR} and D_B^{LR} can be further divided into $D_{B\_N}^{\mathrm{HR}}$ and $D_{B\_N}^{\mathrm{LR}}$, and $D_{B\_B}^{\mathrm{HR}}$ and $D_{B\_B}^{\mathrm{LR}}$, respectively. The former pair is non-blocking dictionaries consisting of corresponding HR/LR atoms [see Fig. 1(h)], while the latter one is blocking dictionaries [see Fig. 1(g)]. Then, I_{HF} can be simultaneously enlarged and decomposed into HR non-blocking and HR blocking components, denoted by I_{HF}^N and I_{HF}^B, respectively [see Figs. 1(j) and 1(k)]. We then integrate I_{HF}^N and the bicubic-interpolated I_{LF} [see Fig. 1(l)] to obtain the final SR result of I, as illustrated in Fig. 1(m). The detailed method will be elaborated below.

A. Preprocessing and Problem Formulation

Without the need of pre-collecting enormous extra training patches for SR, the proposed method intends to extract training patches from an input LR image itself. Moreover, to simultaneously achieve SR and deblocking, similar to [16], we convert the problem into the high-frequency domain of the input image and conduct the following preprocessing tasks. To model the relationship between LR and HR image patches, for an input LR image I with blocking artifacts, we first down-scale I to obtain its down-scaled version I^d, and then apply the BM3D algorithm [19] to decompose I into the LF part (I_{LF}) and HF part (I_{HF}), while decomposing I^d into I_{LF}^d and I_{HF}^d, i.e., $I = I_{LF} + I_{HF}$, and $I^d = I_{LF}^d + I_{HF}^d$. Then, we identify two sets of HR/LR patch pairs as the training samples for learning the dictionaries used for SR and deblocking, where we extract each patch ($x_i \in \mathcal{X}$) in the higher scale (I_{HF}) and its corresponding patch ($y_i \in \mathcal{Y}$) in the lower scale (I_{HF}^d) with a certain magnification factor to form a coupled training patch pair. Then, we perform blocking artifact detection [14] to the HR part of each coupled patch pair (e.g., patch from I_{HF}) and classify all of the coupled patch pairs into two sets of "blocking" and "non-blocking" patch pairs: $\{\mathcal{X}^B, \mathcal{Y}^B\}$ and $\{\mathcal{X}^N, \mathcal{Y}^N\}$.

Based on the two sets of training samples, we propose to learn two sets of dictionaries, respectively, for SR of non-blocking patches and both SR and deblocking of blocking patches, as detailed in Sec. III.B and Sec. III.C, respectively. Then, we formulate the SR of each input LR non-blocking patch b_p (in I_{HF}) as a sparse representation problem:

$$\hat{\theta}_p = arg\,\min_{\theta_p} \left\| b_p - D_N^{\mathrm{LR}} \theta_p \right\|_2^2 + \lambda \left\| \theta_p \right\|_1, \qquad (3)$$

where D_N^{LR} denotes the learned LR dictionary of non-blocking atoms, $\hat{\theta}_p$ is the sparse presentation of b_p with respect to D_N^{LR}, and λ is a parameter controlling the sparsity penalty and representation fidelity. As a result, the SR result B_p of b_p can be obtained as follows:

$$B_p = D_N^{\mathrm{HR}} \times \hat{\theta}_p, \qquad (4)$$

where D_N^{HR} is the learned HR dictionary of non-blocking atoms, which can produce the same sparse representation $\hat{\theta}_p$ for an HR patch as the one for its corresponding LR patch with respect to D_N^{LR}.

Moreover, we formulate the SR and deblocking for each input LR blocking patch b_p (in I_{HF}) as a MCA-based image decomposition problem via sparse representation:

$$\hat{\theta}_p = arg \min_{\theta_p} \left\| b_p - D_B^{\mathrm{LR}} \theta_p \right\|_2^2 + \lambda \left\| \theta_p \right\|_1, \tag{5}$$

where $D_B^{\mathrm{LR}} = \left[D_{B\_N}^{\mathrm{LR}} | D_{B\_B}^{\mathrm{LR}} \right]$ is the learned LR dictionary based on the training samples with blocking artifacts, which can be further partitioned into $D_{B\_N}^{\mathrm{LR}}$ and $D_{B\_B}^{\mathrm{LR}}$, including the non-blocking and blocking atoms, respectively. As a result, the SR and deblocking result B_p of b_p can be obtained via:

$$B_p = D_{B\_N}^{\mathrm{HR}} \times \hat{\theta}_{p\_N}, \tag{6}$$

where $D_{B\_N}^{\mathrm{HR}}$ is obtained from the partition of the learned HR dictionary D_B^{HR} based on the training samples with blocking artifacts, which can be further partitioned into $D_{B\_N}^{\mathrm{HR}}$ and $D_{B\_B}^{\mathrm{HR}}$, including the non-blocking and blocking atoms, respectively, and $\hat{\theta}_{p\_N}$ is the sparse representation of b_p obtained by solving (5) with the coefficients corresponding to the atoms in $D_{B\_B}^{\mathrm{LR}}$, being set to zeros.

B. Dictionary Learning for Single Image Super Resolution

Based on the extracted HR/LR training patch pairs without blocking artifacts ($\{\mathcal{X}^N, \mathcal{Y}^N\}$) from I_{HF} itself, we intend to learn a couple of dictionaries (D_N^{HR} and D_N^{LR}) to model the relationships between HR and LR image patches. Similar to the coupled dictionary training method proposed in [8], we treat the set $\mathcal{Y}^N$ of LR training patches as the observation space, while the set $\mathcal{X}^N$ of HR training patches as the latent space, where the patches have sparse representations with respect to certain dictionaries. Patches (LR) in $\mathcal{Y}^N$ are observable, while patches (HR) in $\mathcal{X}^N$ are what we want to recover. The problem is to find a coupled dictionary pair D_N^{HR} and D_N^{LR} for spaces $\mathcal{X}^N$ and $\mathcal{Y}^N$, respectively, such that given any input LR patch $y_i \in \mathcal{Y}^N$, we can use its sparse representation with respect to D_N^{LR} to reconstruct the corresponding latent HR patch $x_i \in \mathcal{X}^N$ with respect to D_N^{HR}. Hence, for any coupled patch pair $\{x_i, y_i\}$, a desired pair of coupled dictionaries D_N^{HR} and D_N^{LR} should satisfy:

$$\hat{\theta}_i = arg \min_{\theta_i} \|y_i - D_N^{LR} \theta_i\|_2^2 + \lambda \|\theta_i\|_1, \tag{7}$$

$$\hat{\theta}_i = arg \min_{\theta_i} \|x_i - D_N^{HR} \theta_i\|_2^2, \tag{8}$$

where $\hat{\theta}_i$ is the sparse representation of y_i with respect to D_N^{LR}, and the sparse representation of x_i for accurate recovery of x_i with respect to D_N^{HR}.

To learn a coupled dictionary pair D_N^{HR} and D_N^{LR} for accurately reconstructing x_i with respect to D_N^{HR} based on the input y_i, according to [8], the problem can be formulated as:

$$\min_{D_N^{HR}, D_N^{LR}} \frac{1}{N} \sum_{i=1}^{N} L(D_N^{\mathrm{HR}}, D_N^{\mathrm{LR}}, x_i, y_i) \tag{9}$$

s.t. $\hat{\theta}_i = arg \min_{\theta_i} \left\| y_i - D_N^{\mathrm{LR}} \theta_i \right\|_2^2 + \lambda \|\theta_i\|_1, i = 1, 2, \cdots, N,$

where N is the number of training patch pairs, and L is a loss function for ensuring well representation of y_i with respect to D_N^{LR}, and minimizing the reconstruction error of x_i with respect to D_N^{HR}, which is defined as:

$$L = \frac{1}{2} \left[\alpha \left\| D_N^{\mathrm{HR}} \hat{\theta}_i - x_i \right\|_2^2 + (1 - \alpha) \left\| D_N^{\mathrm{LR}} \hat{\theta}_i - y_i \right\|_2^2 \right], \tag{10}$$

where α ($0 < \alpha \leq 1$) balances the two reconstruction errors. The objective function in (9) can be minimized by alternatively optimizing over D_N^{HR} and D_N^{LR}, while keeping the other fixed. More details about the coupled dictionary learning can be found in [8].

C. Dictionary Learning for Simultaneous Super Resolution and Deblocking of Single Image

For the extracted HR/LR training patch pairs with blocking artifacts, we need to not only learn a coupled dictionary pair D_B^{HR} and D_B^{LR} for SR purpose, but also identify the "blocking/non-blocking atoms" in the two dictionaries for achieving MCA-based image deblocking. Similar to the employed coupled dictionary learning technique described in Sec. III.B, we first learn D_B^{HR} and D_B^{LR} based on solving (9), where the parameters are replaced accordingly. Then, we analyze the atoms constituting D_B^{HR}, and find that these atoms can be roughly divided into two clusters (sub-dictionaries) for representing the non-blocking and blocking components of I_{HF}, respectively. That is, even if in this case, the training patches extracted from I_{HF} are determined to be with blocking artifacts based on [14], two different kinds of atoms in terms of blocking and non-blocking artifacts can be still analyzed after dictionary learning process.

Based on [16], we utilize and modify the HOG (histograms of oriented gradients) descriptor [20] to describe each atom in D_B^{HR} (the HOG features of the atoms in the dictionary learned from HR patches should be more significant than those from LR patches). Based on the fact that blocking artifacts are characterized by visually noticeable changes in pixel values along block boundaries, it is only required to consider the horizontal and vertical gradients in each dictionary atom. Hence, we modify the original HOG to calculate only the histogram over the intervals of angles around 0°, 180°, 90°, and 270°. After extracting the HOG feature for each atom in D_B^{HR}, we first apply the *K*-means algorithm to classify all of the atoms in D_B^{HR} into two clusters, D_1 and D_2, based on their horizontal HOG feature descriptors. Then, we calculate the variance of gradient directions for each atom d_{ij} in cluster D_i, as VG_{ij}, $i = 1, 2$. We then calculate the mean of VG_{ij} for each cluster D_i as MVG_i. Based on the fact that the edge directions of the samples of an atom with horizontal (or vertical) blocking artifacts should be consistent, *i.e.*, the variance of gradient directions for a "horizontal (or vertical) blocking" atom should be smaller than those of the other atoms with no remarkably dominating edge direction, we identify the cluster with the smaller MVG_i as horizontal blocking sub-dictionary. Then, the other cluster can be further similarly classified again to obtain the "vertical blocking" sub-dictionary and non-blocking sub-dictionary $D_{B\_N}^{\mathrm{HR}}$. In addition, the identified horizontal and vertical blocking sub-dictionaries constitute the blocking sub-dictionary $D_{B\_B}^{\mathrm{HR}}$. That is, $D_B^{\mathrm{HR}} = \left[D_{B\_N}^{\mathrm{HR}} | D_{B\_B}^{\mathrm{HR}} \right]$. Meanwhile, the atoms in D_B^{LR} can be also classified into two clusters according to their corresponding atoms in D_B^{HR} to form the blocking and non-blocking dictionaries, $D_{B\_N}^{\mathrm{LR}}$ and $D_{B\_B}^{\mathrm{LR}}$, respectively, of LR atoms, *i.e.*, $D_B^{\mathrm{LR}} = \left[D_{B\_N}^{\mathrm{LR}} | D_{B\_B}^{\mathrm{LR}} \right]$.

D. Image SR and Deblocking via Sparse Reconstruction

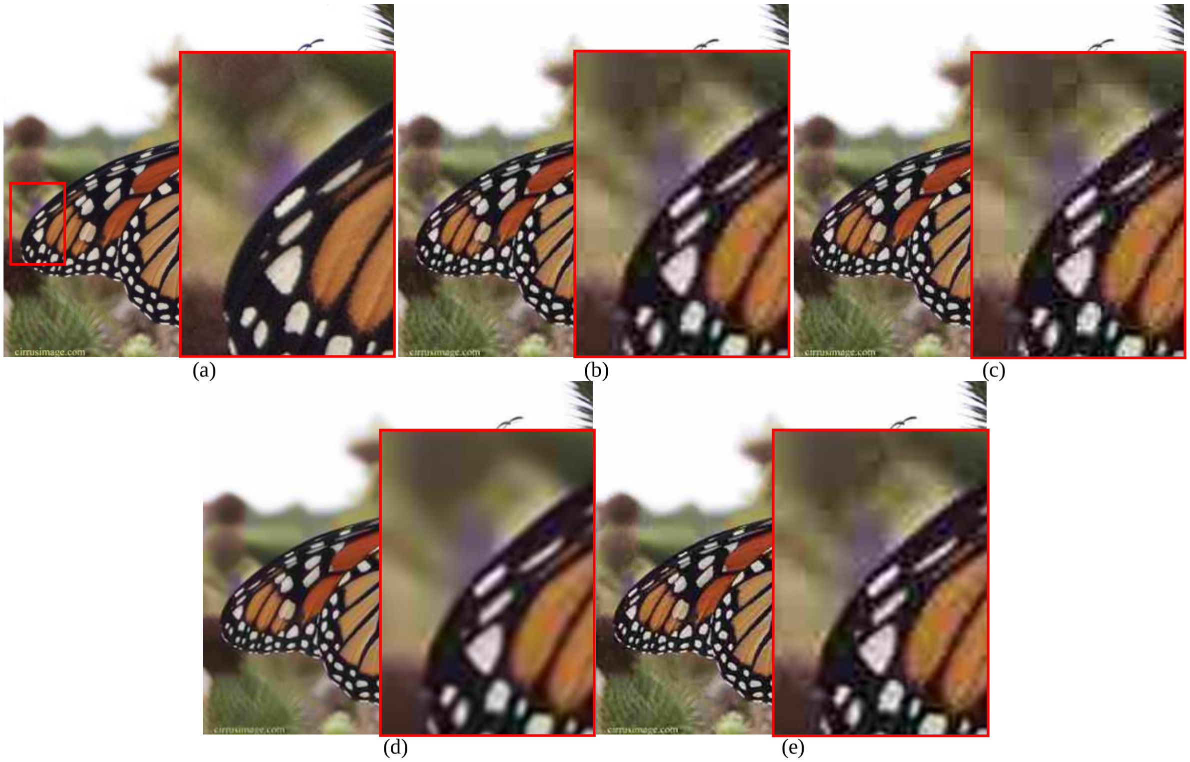

Fig. 2. SR results: **(a)** the original HR image; and the SR versions of the input LR image (LR and compressed version of (a), QF = 15) via the: **(b)** bicubic [3]; **(c)** ScSR [7]; **(d)**; and **(e)** proposed methods.

After learning the six dictionaries, D_N^{HR}, D_N^{LR}, $D_{B\_N}^{\mathrm{HR}}$, $D_{B\_B}^{\mathrm{HR}}$, $D_{B\_N}^{\mathrm{LR}}$, and $D_{B\_B}^{\mathrm{LR}}$, based on the training patches extracted from I_{HF}, simultaneous SR and deblocking of I_{HF} can be efficiently achieved via patch-wise sparse recovery as follows.

1) SR for Patches without Blocking Artifacts: In this case, only SR is required to be performed for each patch. For each input LR non-blocking patch b_p in I_{HF}, we first calculate its sparse representation $\hat{\theta}_p$ with respect to D_N^{LR} via (3). Based on the employed coupled dictionary learning algorithm described in Sec. III-B, the spare representation can be also used to recover its HR version B_p with respect to D_N^{HR} via (4).

2) Simultaneous SR and Deblocking for Patches with Blocking Artifacts: In this case, both SR and deblocking are required to be simultaneously performed to each patch, For each input LR blocking patch b_p in I_{HF}, we first perform MCA-based image decomposition via sparse representation with respect to D_B^{LR} (=$\left[D_{B\_N}^{\mathrm{LR}} | D_{B\_B}^{\mathrm{LR}}\right]$) to find the atoms which actually contribute for representing the non-blocking part of b_p via (5). We then set the coefficients corresponding to the atoms in $D_{B\_B}^{LR}$ in $\hat{\theta}_p$ (the solved sparse representation of b_p) to zeros to obtain $\hat{\theta}_{p\_N}$ (the estimated sparse representation of the non-blocking component of b_p). As a result, the SR and deblocking for b_p can be achieved by the sparse representation $\hat{\theta}_{p\_N}$ with respect to $D_{B\_N}^{\mathrm{HR}}$ via (6).

Then, the recovered and deblocked HR patches are tiled together to reconstruct the HR version I_{HF}^N of I_{HF}, where the average of multiple reconstructs is taken for each pixel in the overlapping region as its final recovery. Finally, we integrate I_{HF}^N and the bicubic-interpolated I_{LF} to obtain the final SR result of I.

IV. Experimental Results

Several JPEG LR images with blocking artifacts were used to evaluate the performance of the proposed low-quality single image SR algorithm subjectively. In our experiments, all of the test LR images were compressed by JPEG with QF ranged from 15 to 25. Different from [11], where the JPEG compression QF is required to be known in advance, in our approach, it is not required to know any prior knowledge (including QF) about an input LR image. The parameter settings of the proposed method are described as follows. For each test LR image of size ranged from 140×140 to 442×400, the magnification factor, HR/LR patch sizes, and size (number of atoms) of learned dictionaries (including D_N^{LR}, D_N^{HR}, D_B^{LR}, and D_B^{HR}) are set to 2, 16×16/8×8, and 1024, respectively.

To highlight the advantage of the proposed joint SR and deblocking approach for single image, we compared our method with pure SR/interpolation approaches, including the bicubic interpolation [3], sparse coding-based SR with extra training samples (denoted by ScSR) [7], and a cascading approach integrating the sparse representation-based deblocking method [16], followed by the self-learning-based SR proposed in this paper (denoted by Cascading-based). Figs. 2 and 3 show some SR results obtained by the proposed method and the existing approaches used for comparisons. It can be observed from Figs. 2 and 3 that the visual quality of our SR results can outperform the results obtained by these

Fig. 3. SR results: (a) the original HR image; and the SR versions of the input LR image (LR and compressed version of (a), QF = 25) via the: (b) bicubic [3]; (c) ScSR [9]; (d) Cascading-based; and (e) proposed methods.

pure SR approaches used for comparisons, where blocking artifacts are significantly magnified while directly enlarging the images. Moreover, our SR results also outperform the Cascading-based approach based on the fact that first apply deblocking would inevitably lose some image details, which may be useful for SR and hard to be recovered in the following SR stage.

V. CONCLUSION

In this paper, we have proposed a self-learning-based SR framework to simultaneously achieve low-quality single-image SR and blocking artifact removal. Our method self-learns image sparse representation for modeling the relationship between LR and HR image patches in terms of the learned dictionaries, respectively, for image patches with and without blocking artifacts. As a result, image SR and deblocking can be simultaneously achieved via sparse representation and MCA-based image decomposition. Our experimental results have demonstrated the efficacy of the proposed algorithm.

REFERENCES

[1] M. Y. Shen and C. C. J. Kuo, “Review of postprocessing techniques for compression artifacts removal,” *J. Vis. Commun. Image Rep.*, 1998.

[2] S. Farsiu et al., “Fast and robust multiframe super resolution,” *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[3] H. S. Hou and H. C. Andrews, “Cubic splines for image interpolation and digital filtering,” *IEEE Trans. Signal Process.*, vol. 26, no. 6, 1978.

[4] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super-resolution,” *IEEE Computer Graphics and Applications*, vol. 22, 2002.

[5] D. Glasner, S. Bagon, and M. Irani, “Super-resolution from a single image,” *Proc. IEEE Int. Conf. Comput. Vis.*, Sept. 2009, pp. 349-356.

[6] G. Freedman and R. Fattal, “Image and video upscaling from local self-examples,” *ACM Trans. Graph.*, vol. 30, no. 2, 2011.

[7] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE Trans. Image Process.*, vol. 19, 2010.

[8] J. Yang et al., “Coupled dictionary training for image super-resolution,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467-3478, Aug. 2012.

[9] C.-Y. Yang et al., “Exploiting self-similarities for single frame super-resolution,” *Proc. Asian Conf. Comput. Vis.*, 2010.

[10] M.-C. Yang and Y.-C. F. Wang, “A self-learning approach to single image super-resolution,” *IEEE Trans. Multimedia*, vol. 15, no. 3, 2013.

[11] Z. Xiong, X. Sun, and F. Wu, “Robust web image/video super-resolution,” *IEEE Trans. Image Process.*, vol. 19, no. 8, Aug. 2010.

[12] T. Q. Pham et al., “Resolution enhancement of low quality videos using a high-resolution frame,” *Proc. SPIE*, vol. 6077, 2006.

[13] F. Liu, J. Wang, S. Zhu, M. Gleicher, and Y. Gong, “Noisy video super-resolution,” *Proc. ACM Multimedia*, 2008, pp. 713-716.

[14] G. A. Triantafyllidis *et al.*, “Blocking artifact detection and reduction in compressed data,” *IEEE Trans. Circuits Syst. Video Technol.*, 2002.

[15] P. List et al., “Adaptive deblocking filter,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 614−619, July 2003.

[16] Y.-W. Chiou *et al.*, “Efficient image/video deblocking via sparse representation,” *Proc. IEEE Visual Commun. Image Process.*, 2012.

[17] J. M. Fadili et al., “Image decomposition and separation using sparse representations: an overview,” *Proc. IEEE*, vol. 98, no. 6, June 2010.

[18] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, “Automatic single-image-based rain streaks removal via image decomposition,” *IEEE Trans. Image Process.*, vol. 21, 2012.

[19] K. Dabov et al., “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, 2007.

[20] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005.