PAPER

Image Arbitrary-Ratio Down- and Up-Sampling Scheme Exploiting DCT Low Frequency Components and Sparsity in High Frequency Components

Meng ZHANG[†], **Tinghuan CHEN**[†a)], **Xuchao SHI**[††], *Nonmembers, and* **Peng CAO**[†], *Member*

SUMMARY The development of image acquisition technology and display technology provide the base for popularization of high-resolution images. On the other hand, the available bandwidth is not always enough to data stream such high-resolution images. Down- and up-sampling, which decreases the data volume of images and increases back to high-resolution images, is a solution for the transmission of high-resolution images. In this paper, motivated by the observation that the high-frequency DCT components are sparse in the spatial domain, we propose a scheme combined with Discrete Cosine Transform (DCT) and Compressed Sensing (CS) to achieve arbitrary-ratio down-sampling. Our proposed scheme makes use of two properties: First, the energy of a image concentrates on the low-frequency DCT components. Second, the high-frequency DCT components are sparse in the spatial domain. The scheme is able to preserve the most information and avoid absolutely blindly estimating the high-frequency components. Experimental results show that the proposed down- and up-sampling scheme produces better performance compared with some state-of-the-art schemes in terms of peak signal to noise ratio (PSNR), structural similarity index measurement (SSIM) and processing time.
key words: *image, DCT, CS, down- and up-sampling, arbitrary ratio, PSNR, SSIM*

1. Introduction

With advances in image acquisition technology and display technology, high resolution images become widely available in many applications. However, the bandwidth resource is not always enough to the data volume of high resolution images. In such circumstances, many down- and up-sampling schemes were proposed to decrease data volume of high resolution images in transmitting terminal and increase back to high resolution images in users end. This is also called scalable code [1]. Most of the available down- and up-sampling schemes are performed in the spatial domain [2]–[12]. Those spatial-domain-based schemes can be classified into two types: adaptivity [2]–[9] and non-adaptivity [10]–[12]. The non-adaptive schemes have low computational complexity, but the fixed estimators, such as filter, cause unstable performance. Due to training or learning process, adaptive schemes have stable performance. However, the time-consuming training or learning process cannot provide real-time down- and up-sampling.

Recent researches show that down- and up-sampling images in the transform domain, such as the wavelet domain [13], the contourlet domain [14] and the discrete cosine transform (DCT) domain [15]–[26], can provide the better performance. In addition, compared with spatial-domain-based adaptive schemes, the schemes based on the transform domain have lower computational complexity.

Since DCT are often used in image/video coding, such as H.264/AVC, JPEG standard, we pay attention to the down- and up-sampling methods based on DCT. In many DCT-based schemes, the high-frequency DCT coefficients of a image are directly truncated to perform down-sampling [15]–[19], [21]–[28]. Because the energy concerns on the low-frequency components, this down-sampling scheme preserves most of information of the image. Therefore, some works were done to up-sample back the high resolution image from the low-frequency DCT coefficients.

Because of the limitation of computational ability in the previous graphics processing unit, the up-sampling schemes were largely confined to low complexity. The most primitive method is padding zeros in the high-frequency DCT coefficients [15]. However, the scheme causes low peak signal to noise ratio (PSNR) and low structural similarity index measurement (SSIM). Therefore, some works were done to improve PSNR and SSIM. Rakesh *et al* used the symmetry and orthogonality properties of DCT and 8×8 sub-block to improve PSNR and computational efficiency [16]. Mukherjee and Mitra modified Rakesh's scheme by using the low pass truncation to reduce artifacts and improve PSNR [17]. However, this scheme increased computational complexity. In addition, Park and Jeong developed Rakesh's work and proposed a pair of hybrid down- and up-sampling method, where the first row and column coefficients of each 8×8 DCT sub-block were preserved to down-sample a image [20]. However, the weight matrix was singular, which caused the suboptimal results. In [22], the multiplication convolution property of DCT and a large DCT sub-block were used to get finer images and improve computational efficiency. Yong and Park proposed an arbitrary-ratio resizing image method [23]. This method combined inverse and forward DCT to produce finer images and improve PSNR. Tan *et al* extended the Yong's work by

Manuscript received April 15, 2015.
Manuscript revised October 8, 2015.
Manuscript publicized October 30, 2015.
[†]The authors are with the National ASIC Engineering Technology Research Center, School of Electronics Science and Engineering, Southeast University, China.
[††]The author is with the Department of Applied Math and Statistics, Johns Hopkins University, USA.
a) E-mail: plutochen1990@gmail.com
DOI: 10.1587/transinf.2015EDP7141

Copyright © 2016 The Institute of Electronics, Information and Communication Engineers

using the spatial relationship between the DCT block and the sub-block [25]. Liang *et al* adopted forward and inverse DCT of the different size to improve PSNR and SSIM [24]. However, the essential methodology of the works [25] and [24] still based on padding zeros. A computational flexible method was proposed in [19] to realize different scale size according to different reconstructed image qualities.

Recently, scholars proposed some spatial-DCT combined schemes to improve the performance of down- and up-sampling. Typically, Wu and Chen proposed a DCT-Wiener-based scheme, where Wiener filter was adopted to solve minimum mean squares error (MMSE) in the spatial domain [26]. However, the fixed parameters of Wiener filter causes unstable performance. Hung and Siu developed Wu's scheme by using the training images to improve performance [27]. However, only one training image is allowed to apply, which causes limited improvement of performance in PSNR and SSIM. In addition, the features of the training image were similar with that of the test images. They proposed another spatial-DCT combined scheme, where K nearest neighbor (K-NN) MMSE estimation was adopted [28]. Although their scheme improved PSNR and SSIM, the up-sampling process was time-consuming.

According to the literature reviewed above, some drawbacks of the existing methods are listed as follows:

1) The loss of the high-frequency DCT components causes the loss of information, including fast changing and edge areas, which cannot be estimated accurately.

2) Although training-based or learning-based adaptive methods have better performance, such as high PSNR, these methods are time-consuming. And their performance depends on relevancy between training images and test images.

3) The performance of the fixed-parameters-based non-adaptive methods are unstable because these fixed parameters are not suitable for all images.

If up-sampling images need be reconstructed accurately, the high-frequency DCT components are not omitted absolutely. The partial high-frequency coefficients are also transmitted so that the low-frequency components and the high frequency components can be reconstructed accurately.

Compressed Sensing (CS) is applied widely in many fields, including image process and image code [29]–[31]. CS pushes through the limitation of Nyquist theory, which raises our concern. However, the precondition of CS is the sparsity of the sampling signal. Because images are not able to satisfy sparsity in the spatial domain and the DCT domain, CS theory cannot be directly applied to perform down- and up-sampling of images. Fortunately, the high-frequency DCT components meet the condition of sparsity in the spatial domain, so we propose DCT and CS combined method to achieve high performance of the down- and up-sampling.

In this paper, a pair of novel down- and up-sampling method is proposed to achieve better performance in PSNR, SSIM and processing time. Our proposed method makes use of two properties: First, the energy of a image concentrates on the low-frequency DCT components. Second, the high-frequency DCT components are sparse in the spatial domain. The scheme is able to preserve most of the information and avoid absolutely blindly estimating the high-frequency components. Although this is not the first time to combine CS with DCT in the field of image processing, the methodology of our proposed method is essentially different from that of those existing methods [32]–[34] based on DCT and CS. We now summarize our novel contributions in comparison with existing methods:

1) The work [32] employed a sparse basis matrix to provide sparse representations for image, and then adopted a measurement matrix to compressed sample the vector of sparse representation. In [33], image was decomposed in texture and piecewise smooth content. In order to decompose image, the work [34] used dictionary training to find the best overcomplete dictionary and the residual in the DCT domain. While we preserve the low-frequency DCT components and compressed sample the high-frequency DCT components in the spatial domain by using CS.

2) The theme of our paper is essentially different from that of [33], [34]. The aim of the paper [33], [34] was decomposing and representing images. The data volume of decomposition and representation is larger than that of the original image. While our aim is decreasing the data volume in down-sampling process and up-sample back to the original data volume of the image.

The organization of this paper is given as follows. In Sect. 2, we analyze pad-zeros method based on DCT and introduce the CS theory. Then, we prove the sparsity of the high-frequency coefficients in the spatial domain and demonstrate the reason that CS theory can be applied in our proposed method. After that, a pair of down- and up-sampling scheme is introduced. In Sect. 3, we simulate our proposed method, then compare our proposed method with some state-of-the-art and representative methods in PSNR, SSIM and processing time. In Sect. 4, we make some conclusions and give some discussion about future development.

2. DCT-Spatial Domain Scheme for Video Frame Down-Sampling and Up-Sampling Based on CS

As Chen said, sole approach cannot achieve good performance to down- and up-sample high resolution [26]. It motivates us to find a solution to achieve the up-sampled high resolution image. According to DCT theory, most of the energy in the image concentrates upon the low-frequency DCT coefficients. While only sharp regions or fast changing areas in the spatial domain contribute to the high-frequency DCT coefficients. In other words, the high-frequency components are sparse. It promotes us to utilize CS theory to compressed sample these components. Therefore, a pair of DCT-spatial-domain-based scheme of down- and up-sampling naturally is proposed to reach the goal of up-sampling high resolution images after down-sampling.

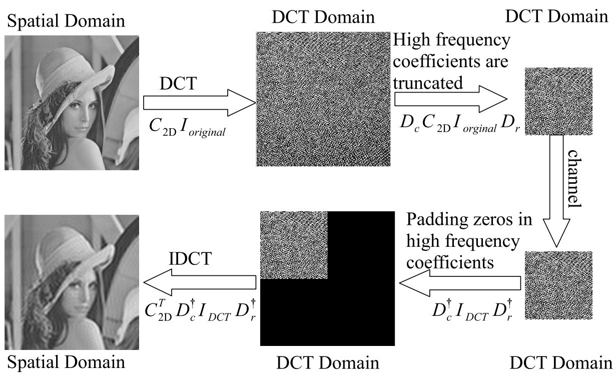

Fig. 1 The pad-zeros method of down-sampling and up-sampling based DCT

2.1 Down-Sampling and Up-Sampling in DCT Domain

To save bandwidth resource, an original image needs to be down-sampled. In universal down-sampling method based on DCT, the low-frequency coefficients are preserved and the high-frequency coefficients are truncated. In this process, 2-D DCT is adopted to process the original image. The mathematical relationship between the matrix of the original image and that of its DCT can be described as:

$$I_{DCT} = D_c mat(C_{2D} vec(I_{original})) D_r \tag{1}$$

where $I_{original} \in R^{M \times N}$ denotes the matrix of the original image, whose size is $M \times N$. C_{2D} is 2-D DCT matrix. The operator $vec(\cdot)$ denotes rearranging a matrix to be a vector. The operator $mat(\cdot)$ denotes rearranged to be a matrix, which is backwards operation of $vec(\cdot)$. It is easy to prove that $C_{2D} = C_{1D} \otimes C_{1D}$ in the Appendix. C_{1D} is 1-D DCT matrix. The operator $\otimes$ denotes Kronecker product. $D_c = [E \quad O] \in R^{M_D \times M}$ is the down-sampling column truncation matrix, correspondingly similarly, $D_r = [E \quad O]^T \in R^{N \times N_D}$ is the down-sampling row truncation matrix. E is identity matrix, M_D and N_D are defined as the column truncation number and the row truncation number, respectively. They indicate the column number and the row number of the low-frequency components after down-sampling, respectively. $I_{DCT} \in R^{M_D \times N_D}$ is the DCT coefficient matrix, whose size is $M_D \times N_D$. We define a concept of down-sampling rate R_D, which is the ratio of the data volume of a image after down-sampling to that of the original image. So down-sampling ratio is $R_D = (M_D \times N_D)/(M \times N)$ in the method of truncation of the high-frequency coefficients.

In up-sampling process, in general, the scheme of padding zeros is adopted. The mathematical relationship of the up-sampling process can be described as follow:

$$I_{re} = mat(C_{2D}^T vec(D_c^{\dagger} I_{DCT} D_r^{\dagger})) \tag{2}$$

where $\cdot^{\dagger}$ denotes pseudo-inverse operation. $\cdot^T$ denotes transposition operation. The pad-zeros method is shown in Fig. 1.

Intuitively, the residual is sparse in the spatial domain when the low-frequency DCT components are removed from an original image. Because the high-frequency DCT coefficients only contribute to fast changing areas and margin areas. The Fig. 2 certifies our conjecture. In Fig. 2, there are only the Lena profile and the hair edge, corresponding to the high-frequency DCT components. And the color of other regions are close to black, which means gray value is close to zero. The image grayscale statistical graph in Fig. 2 also certifies our conjecture. Most of grayscales are nearly to zero, and there is a tiny fraction of grayscales far from zero. Therefore, it prompts us to utilize CS theory to compressed sample the high-frequency DCT components.

2.2 CS in Spatial Domain

To achieve the down- and up-sampling in the spatial domain, a sparse vector needs to be compressed sampled by a random matrix satisfying Restricted Isometric Property(RIP) [35]. When RIP is satisfied, all subsets of K columns taken from measurement matrix are nearly orthogonal. K is sparseness of compressed sampled vector. Therefore, RIP guarantees accurate reconstruction for the compressed sampled vector with high probability [36]. The reconstruction process is solving minimum l_0 norm problem. The CS process and the minimum l_0 norm of CS reconstruction are described as follows:

$$y = \Phi x = \Phi \dot{v}ec(I_{original}) \tag{3}$$

$$\hat{x} = argmin\|x\|_{l_0} \qquad s.t. \quad y = \Phi x \tag{4}$$

where $x \in R^{MN \times 1}$ is an intermediate variable for convenient representation. $y \in R^{M_C N_C \times 1}$ is the compressed sampling result. $\Phi \in R^{M_C N_C \times MN}$ is random matrix. $M_C N_C$ is the compressed sampling number.

It is worth mentioning that a large number of the image data causes time-consuming sampling and reconstruction. In order to avoid that, we adopt the sensing window and the reconstructed window in down-sampling process and up-sampling process, respectively. In the sensing window and the reconstructed window, the image vector with fixed length are compressed sampled and reconstructed, respectively. Empirically, we set the length of the image vector to 128 in order to achieve good performance in processing time and precision.

However, in order to solve Eq. (4), the linear combinations with C_N^K kinds of probability need to be taken into account. Therefore, the numerical computation of Eq. (4) is unstable. So it is named NP-hard [37]. Chen *et al* indicated the solution of minimum l_1 norm is the equivalent of that of minimum l_0 norm in terms of Eq. (4) [38]. Besides, the ways of solving minimum l_1 norm are simpler than that of solving minimum l_0 norm. Therefore, the equation minimum l_0 norm (4) is transformed into the minimum l_1 norm (5) as follow:

$$\hat{x} = argmin\|x\|_{l_1} \qquad s.t. \quad y = \Phi x \tag{5}$$

this is a convex optimization problem. In addition, there are many reconstruction algorithms proposed to solve the

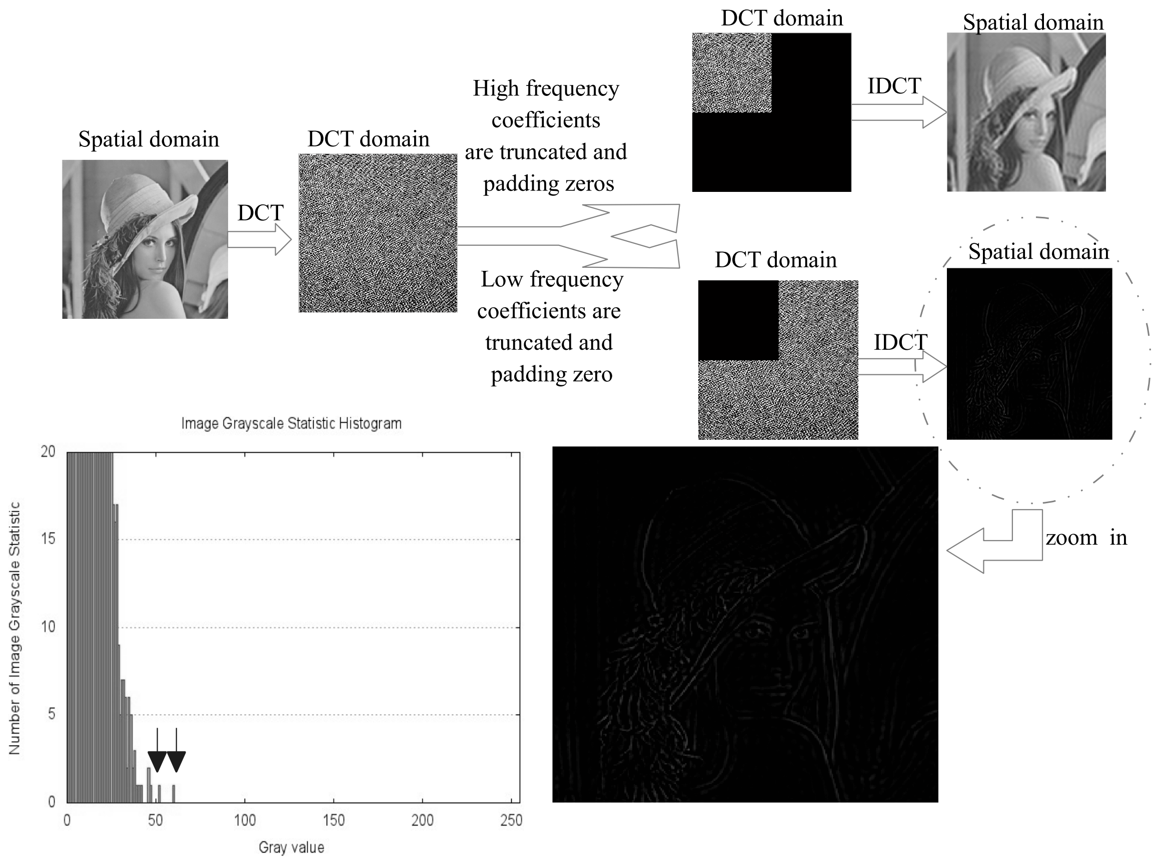

Fig. 2 Image is divided into high frequency components and low frequency components

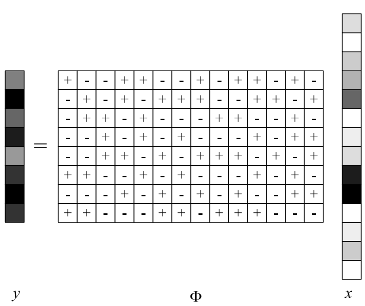

Fig. 3 Compressed Sampling process

problem (5), such as Matching Pursuit (MP) [39], Basis Pursuit (BP) [38], Orthogonal Matching Pursuit (OMP) [40] and Stepwise Subspace Pursuit (SSP) [41]. In this paper, SSP is adopted to solve the problem (5). According to the literature [42], compressed sampling number satisfies $M_C N_C \geq CKlog(MN)$. Φ satisfies RIP, such as Gaussian random matrix or Bernoulli random matrix. Because Bernoulli random matrix, whose elements are 1 or -1, is suitable for implement of hardware. We adopt Bernoulli random matrix as measurement matrix in this paper. The Fig. 3 shows the CS procedure.

2.3 DCT-Spatial Domain Down-Sampling and Up-Sampling Method Based on CS

In traditional down-sampling process, the high-frequency DCT coefficients are simply removed from an original image. This method causes blur in fast changing areas and edge areas. And in other methods, the high-frequency components are estimated from the low-frequency components. The performance of those methods depends on statistical property of the original image. The high-frequency DCT coefficients cannot be estimated from the low-frequency DCT coefficients in up-sampling process because of irrelevance between the high-frequency DCT coefficients and the low-frequency DCT coefficients. It is desired that a scheme is proposed, where the high-frequency DCT components also are down-sampled, transmitted and up-sampled. The reason of adopting CS is that the high-frequency DCT components satisfy the condition of sparsity in the spatial domain. As a matter of convenience, the representation of separating coefficients in the DCT domain in [26] is adopted as follow:

$$mat(C_{2D}vec(I_{original})) = \begin{bmatrix} f_{lf} & f_{hf} \\ f_{hf} & f_{hf} \end{bmatrix} \tag{6}$$

In proposed method, the high-frequency DCT coefficients and the low-frequency DCT coefficients are down-sampled, respectively. As a result, transmitted sequences

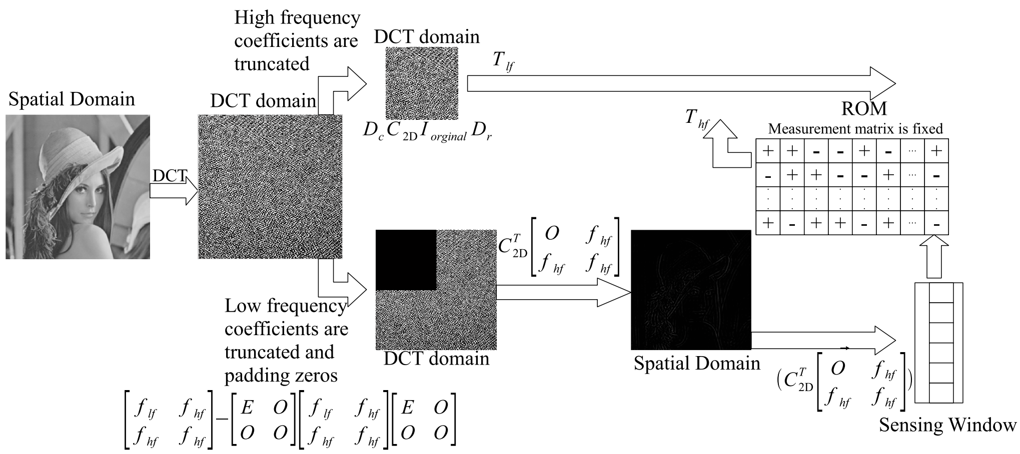

Fig. 4 The proposed down-sampling process

include two parts. The down-sampling process is expressed as:

$$\begin{aligned} T_{lf} &= D_c mat(C_{2D} vec(I_{original}))D_r \\ &= \begin{bmatrix} E & O \end{bmatrix} \begin{bmatrix} f_{lf} & f_{hf} \\ f_{hf} & f_{hf} \end{bmatrix} \begin{bmatrix} E \\ O \end{bmatrix} \\ &= \begin{bmatrix} f_{lf} \end{bmatrix} \end{aligned} \tag{7}$$

$$\begin{aligned} T_{hf} &= \Phi vec(C_{2D}^T(mat(C_{2D}vec(I_{original})) \\ &\quad - D_c^T D_c mat(D_{2D} vec(I_{original}))D_r D_r^T)) \\ &= \Phi vec(C_{2D}^T \begin{bmatrix} O & f_{hf} \\ f_{hf} & f_{hf} \end{bmatrix}) \\ &= \Phi\lambda \end{aligned} \tag{8}$$

where $T_{lf} \in R^{M_D \times N_D}$ and $T_{hf} \in R^{M_C N_C \times 1}$ are transmitted data of the low-frequency components and the high-frequency components, respectively. For the simplicity of discussion, the variable $\lambda \in R^{MN\times 1}$ is defined in (9):

$$\lambda = C_{2D}^T vec(\begin{bmatrix} O & f_{hf} \\ f_{hf} & f_{hf} \end{bmatrix}) \tag{9}$$

It is worth mentioning that there are two concepts we define, which are DCT truncation rate R_T and compressed sampling rate R_C. They like the down-sampling rate. R_T is the ratio of the data volume of the down-sampled image to that of the original image when the high-frequency coefficients are truncated. R_C is the ratio of the data volume of the down-sampling image to that of the original image when the high-frequency coefficients are compressed sampled. In proposed method, DCT truncation rate $R_T = (M_D N_D)/(MN)$, CS rate $R_C = (M_C N_C)/(MN)$, dowmsampling rate $R_D = R_C + R_T = (M_D N_D + M_C N_C)/(MN)$.

From (7) and (8), the low-frequency components are down-sampled in the DCT domain. Besides, the high-frequency components are down-sampled in the spatial domain. It is obvious that information integrity of the image is achieved. The down-sampling process is shown in Fig. 4. In our work, the random compressed sampling matrix is fixed on ROM, whose size is depended on sparsity of the high-frequency components of an image in the spatial domain. When compressed sampling matrix satisfies RIP, compressed sampled vector can be reconstructed with high probability. Besides, the aim of down-sampling is decreasing data volume. It increases the data volume if compressed sampling matrix is randomly generated and transmitted at sending end.

In users end, the up-sampling method matching down-sampling method proposed above needs to be discussed. Up-sampling process also is divided into two parts, the low-frequency components and the high-frequency components, corresponding to the down-sampling process. When users end receives the low-frequency DCT components, up-sampling is performed to gain the primary image by padding zeros and inverse DCT operation. Similarly, when users end receives the compressed sampled high-frequency components, the CS reconstruction algorithm is carried to up-sample the high-frequency components in the spatial domain. Finally, the high-frequency components are added to the primary image in order to get the high resolution image. Since the low-frequency components and the high-frequency components are down- and up-sampled, respectively. The up-sampled image is made of the low-frequency components and the high-frequency components so that the information integrity is achieved. Therefore, the up-sampled image is exactly same with the original image. The up-sampling method is represented as follows:

$$\begin{aligned} I_{relf} &= mat(C_{2D}^T vec(U_c T_{lf} U_r)) \\ &= C_{2D}^T \begin{bmatrix} E \\ O \end{bmatrix} \begin{bmatrix} f_{lf} \end{bmatrix} \begin{bmatrix} E & O \end{bmatrix} \\ &= C_{2D}^T \begin{bmatrix} f_{lf} & O \\ O & O \end{bmatrix} \end{aligned} \tag{10}$$

$$\hat{\lambda} = argmin||\lambda||_{l_1} \qquad s.t. \quad T_{hf} = \Phi\lambda \tag{11}$$

$$I_{rehf} = mat(\hat{\lambda}) \tag{12}$$

where $U_c = D_c^T \in R^{M\times M_D}$ is the column padding zeros matrix. $U_r = D_r^T \in R^{N_D \times N}$ is the row padding zeros matrix.

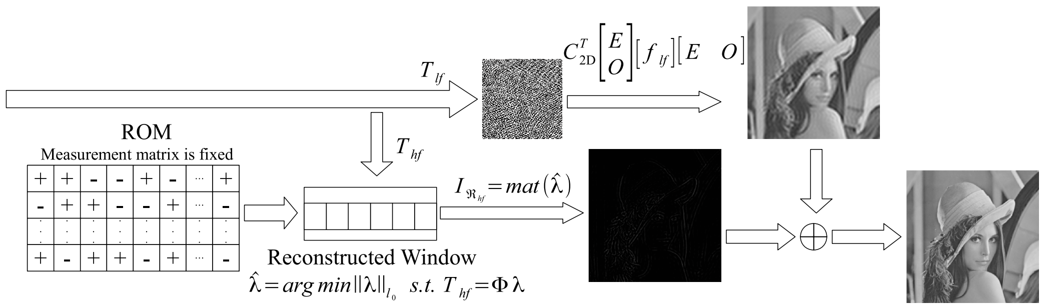

Fig. 5 The proposed up-sampling process

I_{relf} is the up-sampled low-frequency components. I_{rehf} is the up-sampled high-frequency components.

From (10) to (12), we can perceive that the two parts, which are the low-frequency components and the high-frequency components, are up-sampled, respectively. The synthesis of the two parts gains the up-sampling full process. The up-sampling process is shown in Fig. 5.

The error between the original image and the down- and up-sampled image for the high-frequency components is shown in (13)

$$e_{hf} = \sum_{I}(\lambda - \hat{\lambda}) \tag{13}$$

From (7) and (10), it is clear that the down- and up-sampling process of the low-frequency components are linear. However, it is a non-linear process that the high-frequency components are down-sampled and up-sampled. So compared with the high-frequency components, the error of the low-frequency components can be omitted, that is $e_{proposed} \approx e_{hf}$.

3. Experiment Results

Some experimental works have been done to demonstrate our proposed scheme for down- and up-sampling images in this section. A Intel Core i3-2120 personal computer with 3.30 GHz and 2G RAM is used for all simulations in a MATLAB platform. There are thirteen images, whose size is 256×256, to be chosen to perform a series of experiments. The thirteen images include one gray Lena photograph and other twelve images as shown in Fig. 6. Next, we demonstrate the performance of the proposed scheme from: (A) The down- and up-sampling results for the Lena image by using our proposed method; (B) PSNR and SSIM by using our proposed method under different down-sampling rates, including different DCT truncation rates and compressed sampling rates; (C) PSNR and SSIM by using our proposed method and six fixed-ratio state-of-the-art and representative methods for different images as shown in Fig. 6 when down-sampling rate is 0.25; (D) Processing time by using our proposed method and six fixed-ratio state-of-the-art and representative methods for different images as shown in Fig. 6 when down-sampling rate is 0.25; (E) PSNR and SSIM by using our proposed method and two arbitrary-ratio state-of-the-art and representative methods for the gray Lena image under different down-sampling rates.

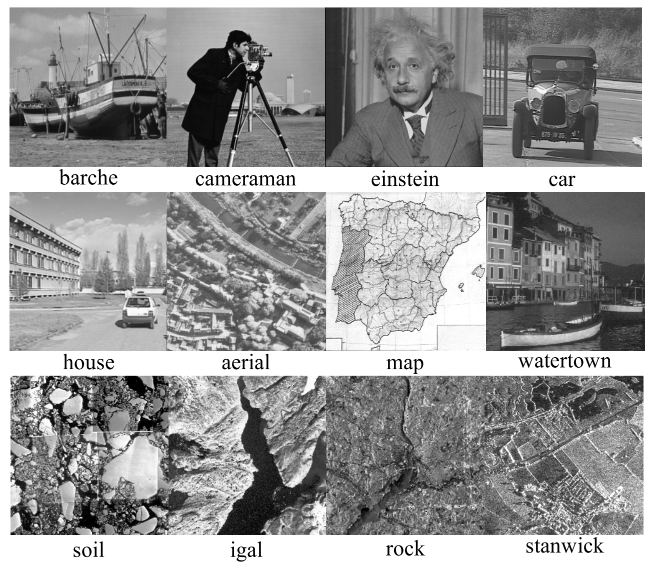

Fig. 6 The test images

3.1 The Down- and Up-Sampling Results for the Lena Image by Using Our Proposed Method

The gray Lena image is adopted to perform the down- and up-sampling by the proposed method. In this example, we set the down-sampling rate to 0.25, in which the compressed sampling rate is 0.1 and the DCT truncation rate is 0.15. Let's analyze the up-sampled image after down-sampling as shown in Fig. 7. The left in Fig. 7 (a) is the original gray Lena image, the right is the up-sampled gray Lena image. The figure shows that our proposed method provides the up-sampled image be closest to the original image. A natural way to analyze detail difference between the original image and the up-sampled image is partitioning the images into small sub-blocks and enlarging them as shown in Fig. 7 (b)-(e). As we can see, each enlarged section of the up-sampled image is closest to that of the original image. However, it is an undeniable fact that there is a shock effect happening in edge neighboring areas where a series of pseudo-edge parallel to them and their amplitude would vanish gradually with the increases of distance from the edge. The shock effects are pointed out by arrows in Fig. 7.

The result shows that the image can be up-sampled accurately, especially the high-frequency components, including edge areas, fast changing areas. However, there is a drawback of shock effect in edge areas.

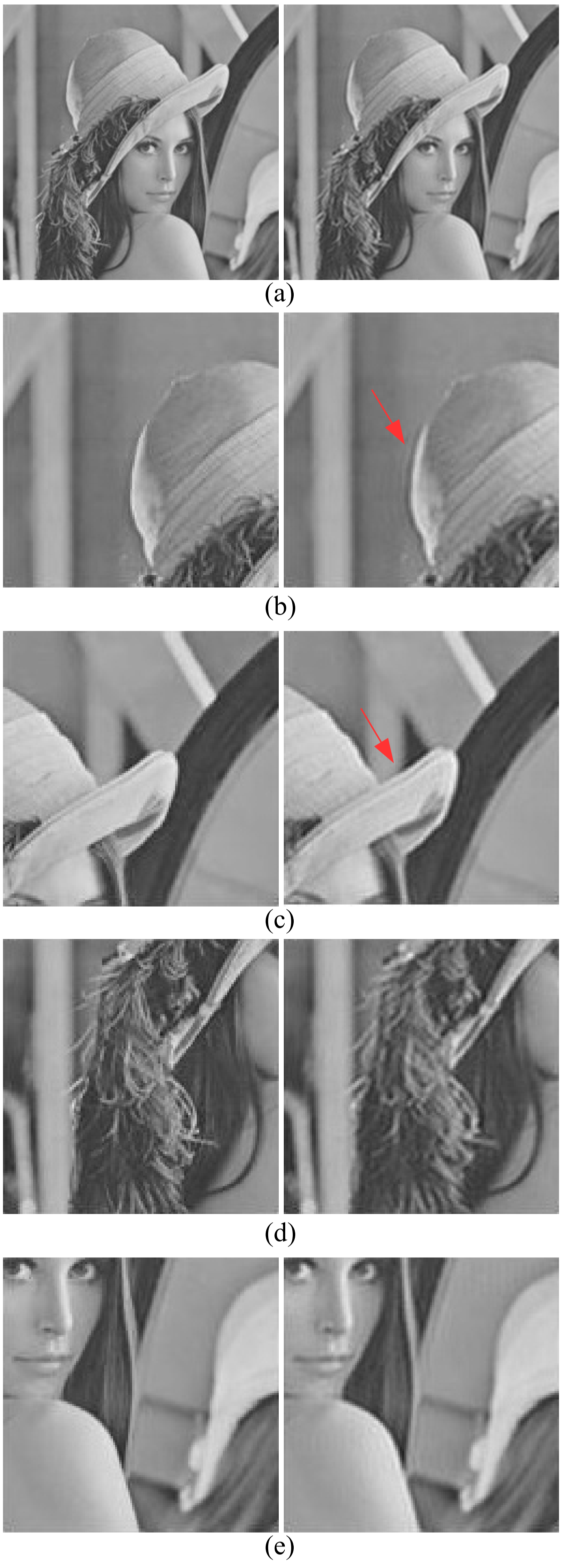

Fig. 7 The original gray Lena image and the up-sampled gray Lena image by using our proposed method

3.2 PSNR and SSIM by Using the Proposed Method under Different Down-Sampling Rates, Including Different DCT Truncation Rates and Compressed Sampling Rates

In this subsection, in order to objectively access the performance of our proposed method, we show PSNR and SSIM between the original Lena image and the down- and up-sampled image under different down-sampling rates. The high-frequency DCT components and the low-frequency DCT components are down-sampled, respectively. Besides, the DCT truncation rate and the compressed sampling rate can be changed. Therefore, our proposed method can achieve arbitrary-ratio down-sampling in terms of DCT truncation rates and compressed sampling rates.

Full reference quality assessment schemes are able to objectively assess the difference between two images. PSNR and SSIM [43] are two common full reference quality assessment schemes. They are adopted to assess the performance of different methods. PSRN is defined as follow:

$$PSNR(dB) = 10log\frac{255^2}{\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}[\hat{I}(i,j) - I(i,j)]^2} \tag{14}$$

where $I_{(i,j)}$ is the original image pixel matrix, $\hat{I}_{(i,j)}$ is the up-sampled image pixel matrix, M and N are the size of image. The larger the value of PSNR is, the higher the similarity between the original image and the up-sampling is. When the value of PSNR would be next to infinite, the two images are absolutely identical. SSIM is defined in (15), where $\bar{I}$ and $\bar{\hat{I}}$ is mean value of the original image pixel and the up-sampled image pixel, respectively. k_1, k_2 are constants, and in general, $0.01 \le k_1, k_2 \le 0.03$. $SSIM \in [0, 1]$. The more nearer SSIM approximates to one, the more similar the two images are in structure.

PSNR and SSIM between the original Lena image and

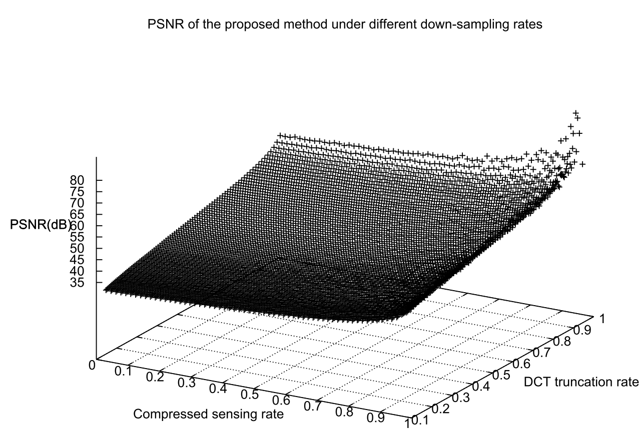

Fig. 8 PSNR between the original image and the up-sampled image by our proposed method under different down-sampling rate

$$
\begin{aligned}
SSIM = {} & \frac{2\bar{I}\bar{\hat{I}} + (255k_1)^2}{\bar{I}^2 + \bar{\hat{I}}^2 + (255k_1)^2} \times \frac{2\sqrt{\frac{1}{MN}\sum_I (I_{i,j} - \bar{I})^2}\sqrt{\frac{1}{MN}\sum_{\hat{I}} (\hat{I}_{i,j} - \bar{\hat{I}})^2} + (255k_2)^2}{\frac{1}{MN}\sum_I (I_{i,j} - \bar{I})^2 + \frac{1}{MN}\sum_{\hat{I}} (\hat{I}_{i,j} - \bar{\hat{I}})^2 + (255k_2)^2} \\
& \times \frac{\frac{1}{MN}\sum_I \ (I_{i,j} - \bar{I})(\hat{I}_{i,j} - \bar{\hat{I}}) + \frac{(255k_2)^2}{2}}{\sqrt{\frac{1}{MN}\sum_I (I_{i,j} - \bar{I})^2}\sqrt{\frac{1}{MN}\sum_{\hat{I}} (\hat{I}_{i,j} - \bar{\hat{I}})^2} + \frac{(255k_2)^2}{2}}
\end{aligned}
\tag{15}
$$

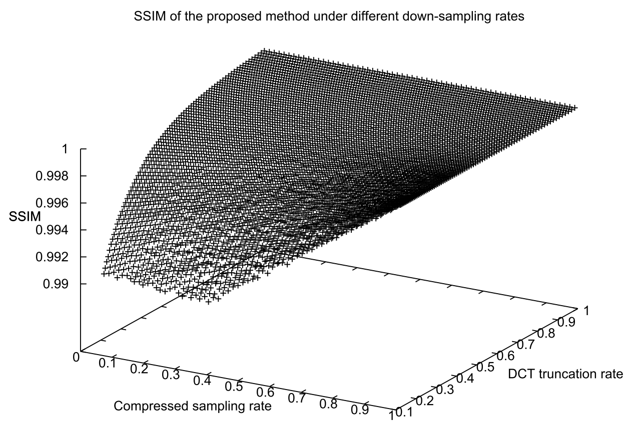

Fig. 9 SSIM between the original image and the up-sampled image by our proposed method under different down-sampling rate

the down-sampled image are shown in Fig. 8 amd Fig. 9, respectively. From Fig. 8 amd Fig. 9, it is observed that with the increasing of DCT truncation rates and compressed sampling rates, the values of PSNR and SSIM also increase. Besides, the rising tendency of PSNR and SSIM by increasing DCT truncation rates is similar with that by increasing compressed sampling rates. Therefore, the high-frequency DCT components and the low-frequency DCT components play the same important role, in terms of PSNR and SSIM. This also illustrates that the traditional methods based on truncation and padding zeros in the high-frequency coefficients have an obvious drawback.

3.3 PSNR and SSIM by Using Our Proposed Method and Six Fixed-Ratio State-of-the-Art and Representative Methods for Different Images as Shown in Fig. 6 When Down-Sampling Rate Is 0.25

In this subsection, we compare the proposed scheme with six fixed-ratio state-of-the-art and representative methods for different images shown in Fig. 6 in terms of PSNR and SSIM. Those methods include an nonadaptive spatial-domain-based method (bilateral filter [11]), an adaptive spatial-domain-based method (roft-decision interpolation [3]), a DCT-based method (hybrid up-sampling [20]) and three spatial-DCT-based methods (DCT learnt Wiener filter [27], K-NN MMSE estimation [28], DCT-Wiener interpolation [26]). Because of the unchangeable down-sampling rate (0.25) and up-sampling (4) rate in those methods, we also fix down-sampling rate on 0.25 and up-sampling rate on 4 to compare performance under the same condition. The twelve images, whose size is 256×256 as shown in Fig. 6, were used to simulate. Our simulation results are shown in Table 1.

From Table 1, the method of K-NN MMSE estimation gains the highest average PSNR and SSIM value. Because the K-NN MMSE estimation method extracts the features of training images to up-sample tested images. Our proposed method achieves the second highest PSNR and SSIM value on average. Besides, the PSNR value and SSIM value of our proposed method improve $0.23dB$ and 0.0213 than that of the method of DCT learnt Wiener filter, which gains the third highest PSNR and SSIM.

3.4 Processing Time by Using Our Proposed Method and Six Fixed-Ratio State-of-the-Art and Representative Methods for Different Images as Shown in Fig. 6 When Down-Sampling Rate Is 0.25

In this subsection, we extend experiments in Sect. 3.3 and list the processing time of those methods mentioned in Sect. 3.3 when down-sampling rate is 0.25. Our simulation results are shown in Table 2. It is worth mentioning that the down-sampling method of DCT-based and DCT-spatial based methods (DCT-Wiener interpolation, DCT-learnt wiener filter and K-NN MMSE estimation) is the truncation of the high-frequency components. So we put together in second column in Table 2. Those down-sampling methods (bilateral filter, soft-decision interpolation) are the spatial decimation using Dirac delta function. Thus we put together in third column in Table 2. From Table 2, the down-sampling processing time of the spatial decimation using Dirac delta function is shorter than other methods, because it performs down-sampling in the spatial domain and does not perform DCT. The method of the second shorter down-sampling processing time is the truncation of the high-frequency DCT components. There is once DCT in this down-sampling method. The down-sampling processing time of our proposed method is the double of that of the truncation based method. Because there are once DCT and once inverse DCT in our proposed method. However, the down-sampling processing time is far shorter than that of the method [20]. So the down-sampling processing time of our proposed method is modest.

Let's pay attention to the up-sampling processing time. There is no doubt that the up-sampling time of the K-NN MMSE estimation is longer than other methods because of training process. Because its training process brings average up-sampling time to 225.6 so that this method cannot

Table 1 PSNR of 12 images shown in Fig. 6 between original image and up-sampling image by using A: DCT-wiener interpolation [26]; B: bilateral filter [11]; C: DCT learnt wiener filter [27]; D: Hybrid upsampling [20]; E: K-NN MMSE estimation [28]; F: soft-decision interpolation [3]; G: our proposed method, when down-sampling rate equals to 0.25

Image	PSNR/dB							SSIM						
	A	B	C	D	E	F	G	A	B	C	D	E	F	G
barche	26.14	25.91	26.81	18.73	28.14	25.93	27.45	0.9672	0.9666	0.9754	0.7508	0.9831	0.9667	0.9767
cameraman	25.99	24.46	26.13	18.34	27.48	24.47	26.02	0.9782	0.9695	0.9816	0.7135	0.9869	0.9695	0.9787
einstein	28.35	26.52	28.81	19.76	30.17	26.48	28.44	0.9641	0.9459	0.9758	0.8077	0.9754	0.9453	0.9657
car	23.94	23.95	24.11	19.74	25.41	23.98	24.77	0.9115	0.9157	0.9241	0.8034	0.9451	0.9162	0.9312
house	26.18	26.04	26.47	17.15	28.39	26.10	27.27	0.9366	0.9380	0.9473	0.7110	0.9550	0.9388	0.9544
aerial	23.17	23.98	23.81	19.49	26.44	24.02	25.99	0.9103	0.9340	0.9284	0.7594	0.9653	0.9343	0.9591
map	18.55	18.62	20.51	15.93	20.35	18.61	19.49	0.7258	0.7513	0.7797	0.6850	0.8548	0.7492	0.7997
watertown	25.49	25.32	26.08	20.16	27.47	25.37	26.66	0.9407	0.9409	0.9517	0.8540	0.9557	0.9414	0.9571
soil	16.98	17.34	18.21	19.35	20.44	17.39	18.52	0.8601	0.8819	0.8842	0.7901	0.9378	0.8820	0.9111
igal	18.71	18.90	20.51	20.49	21.48	18.92	20.23	0.9078	0.9165	0.9152	0.8435	0.9534	0.9164	0.9390
rock	19.62	20.26	21.05	21.48	22.36	20.28	21.21	0.6823	0.7592	0.7258	0.7085	0.8545	0.7582	0.8151
stanwick	16.51	16.74	18.24	20.56	20.91	16.76	17.49	0.6724	0.7198	0.7144	0.7805	0.8559	0.7180	0.7722
average	22.47	22.34	23.40	19.27	24.92	22.36	**23.63**	0.8714	0.8866	0.8920	0.7673	0.9352	0.8863	**0.9133**

Table 2 Processing time by using A: DCT-Wiener interpolation [26]; B: bilateral filter [11]; C: DCT learnt Wiener filter [27]; D: Hybrid upsampling [20]; E: K-NN MMSE estimation [28]; F: soft-decision interpolation [3]; G: our proposed method, when down-sampling rate equals to 0.25

Image	down-sampling processing/s				up-sampling processing/s						
	A, C, E	B, F	D	G	A	B	C	D	E	F	G
barche	0.0110	0.000307	0.2768	0.0233	0.0305	0.0325	0.0161	1.0080	223.0	2.808	0.4719
cameraman	0.0096	0.000346	0.2657	0.0236	0.0439	0.0316	0.0381	0.8715	227.6	2.184	0.4689
einstein	0.0109	0.000326	0.4229	0.0291	0.0383	0.0308	0.0137	1.0511	228.6	2.652	0.4270
car	0.0109	0.000297	0.3135	0.0238	0.0268	0.0327	0.0181	0.7937	221.7	3.089	0.5359
house	0.0099	0.000302	0.5751	0.0308	0.0293	0.0358	0.0149	0.7941	219.8	2.590	0.4583
aerial	0.0118	0.000302	0.2426	0.0275	0.0312	0.0310	0.0150	0.6059	225.0	3.822	0.4539
map	0.0109	0.000306	0.2514	0.0251	0.0361	0.0341	0.0139	0.7852	226.8	3.057	0.4066
watertown	0.0089	0.000304	0.3293	0.0264	0.0242	0.0319	0.0449	0.8700	226.5	2.746	0.5071
soil	0.0108	0.000310	0.4620	0.0243	0.0337	0.0346	0.0142	0.6738	229.3	3.791	0.4433
igal	0.0115	0.000302	0.2450	0.0223	0.0599	0.0340	0.0140	0.6276	227.2	3.962	0.4530
rock	0.0087	0.000305	0.2478	0.0236	0.0238	0.0374	0.0171	0.6737	225.2	4.197	0.4881
stanwick	0.0110	0.000413	0.2604	0.0233	0.0290	0.0307	0.0163	0.6105	226.1	4.212	0.4020
average	0.0105	0.000318	0.3244	**0.0252**	0.0339	0.0331	0.0197	0.7804	225.6	3.259	**0.4597**

achieve real-time process. The up-sampling time of our proposed method is 0.4597 seconds on average, which is longer than that of DCT-Wiener interpolation, bilateral filter and DCT learnt Wiener filter. But compared with the hybrid upsampling, soft-decision interpolation and K-NN MMSE estimation, the up-sampling time of our proposed method is acceptable.

3.5 PSNR and SSIM by Using Our Proposed Method and Two Arbitrary-Ratio State-of-the-Art and Representative Methods for the Gray Lena Image under Different Down-Sampling Rates

As mentioned above, our proposed method can achieve arbitrary-ratio down- and up-sampling. In this subsection, we compare our proposed method with two state-of-the-art arbitrary-ratio methods under different down-sampling rates. The joint arbitrary-ratio resizing method proposed by Chung [24] and the fast arbitrary resizing method proposed by Tan [25] are chosen to down- and up-sample gray Lena image. Fig. 10 and Fig. 11 show PSNR and SSIM between the original image and the down- and up-sampled image by using the joint arbitrary-ratio resizing method, the fast arbitrary resizing method and our proposed method. From Fig. 10 and Fig. 11, the performance of our proposed method is more excellent than that of the arbitrary-ratio resizing method and the fast arbitrary resizing method under different down-sampling rares.

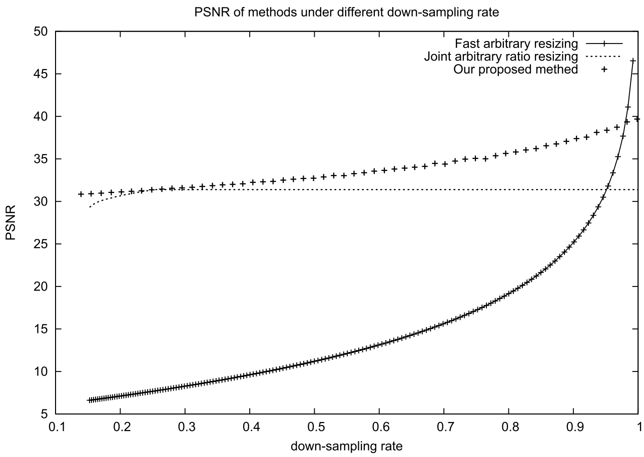

Fig. 10 PSNR between the original image and the up-sampled image by using fast arbitrary resizing [25], arbitrary ratio resizing [24] and our proposed method under different down-sampling rate

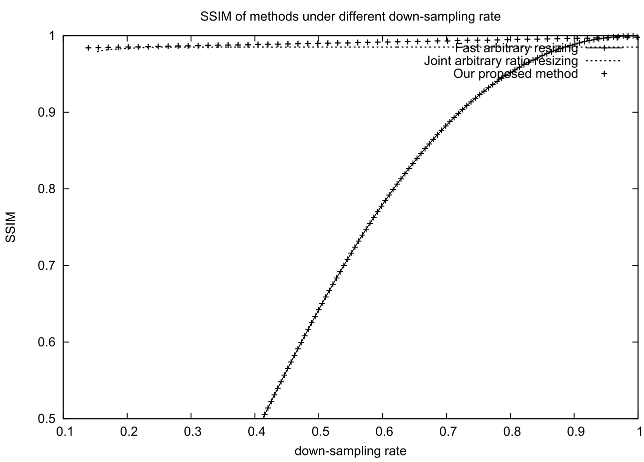

Fig. 11 PSNR between the original image and the up-sampled image by using fast arbitrary resizing [25], arbitrary ratio resizing [24] and our proposed method under different down-sampling rate

4. Conclusion

In this paper, we propose a pair of arbitrary-ratio DCT-CS-combined down-sampling and up-sampling scheme based on the DCT-Spatial domain, which allows us to accurately reconstruct the low-frequency DCT components and the high-frequency DCT components. The high-frequency DCT components are down-sampled in the spatial domain to avoid blindly estimating. In up-sampling process, padding zeros and inverse DCT are adopted to reconstruct the low-frequency coefficients as well as SSP is used to reconstruct the compressed sampled high-frequency coefficients. Then, the whole image is made up with the low-frequency components and the high-frequency components. Experiments demonstrate performance of the proposed scheme obviously outperforms six fixed-ratio and two arbitrary-ratio state-of-the-art and representative methods in terms of PSNR and SSIM. In addition, the down- and up-sampling time in our proposed method is acceptable.

In the future, we plan to improve our proposed method to adaptively choose the down-sampling rate of the low-frequency components and the high-frequency components for each image to achieve high up-sampling image quality. Besides, elimination of the edge shock effect also is our next research goal.

Acknowledgment

This research work was partly supported by the National Natural Science Foundation of China under No. 61550110244 and the National Natural Science Foundation of China under No.61401090. The authors would like to thank the Assistant Professor Yu Bei, from Chinese University of HongKong, and the Associate Professor Li Jing, from Lehigh University, for their comments and suggestions that helped revise this manuscript, both in presentation and in content.

References

[1] Z. Shi, X. Sun, and F. Wu, "Spatially scalable video coding for HEVC," IEEE Trans. Circuits Syst. Video Technol., vol.22, no.12, pp.1813–1826, 2012.
[2] 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.
[3] K.-W. Hung and W.-C. Siu, "Robust soft-decision interpolation using weighted least squares," IEEE Trans. Image Process., vol.21, no.3, pp.1061–1069, 2012.
[4] K.S. Ni and T.Q. Nguyen, "Image superresolution using support vector regression," IEEE Trans. Image Process., vol.16, no.6, pp.1596–1610, 2007.
[5] K.S. Ni and T.Q. Nguyen, "An adaptable *k*-nearest neighbors algorithm for MMSE image interpolation," IEEE Trans. Image Process., vol.18, no.9, pp.1976–1987, 2009.
[6] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single-image super-resolution via a coupled constraint," IEEE Trans. Image Process., vol.21, no.2, pp.469–480, 2012.
[7] S. Ruikar and W. Tushar, "Advanced neighbor embedding for image super resolution," International Journal of Communication Systems, vol.8, no.2, 2013.
[8] Q. Wang and R.K. Ward, "A new orientation-adaptive interpolation method," IEEE Trans. Image Process, vol.16, no.4, pp.889–900, 2007.
[9] X. Zhang and X. Wu, "Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation," IEEE Trans. Image Process., vol.17, no.6, pp.887–896, 2008.
[10] A.-C. Tsai, K. Bharanitharan, J.-F. Wang, and J.-F. Yang, "Classified multifilter up-sampling algorithm in spatial scalability for H.264/SVC encoder," IEEE Trans. Circuits Syst. Video Technol., vol.20, no.6, pp.861–869, 2010.
[11] K.-W. Hung and W.-C. Siu, "Fast image interpolation using the bilateral filter," IET Image Processing, vol.6, no.7, pp.877–890, 2012.
[12] H.A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," IEEE Trans. Image Process., vol.14, no.10, pp.1647–1659, 2005.
[13] H. Demirel and G. Anbarjafari, "Image resolution enhancement by using discrete and stationary wavelet decomposition," IEEE Trans. Image Process., vol.20, no.5, pp.1458–1460, 2011.
[14] Z. Ashouri and S. Shirani, "Video super resolution using contourlet

transform and bilateral total variation filter," IEEE Trans. Consum. Electron., vol.59, no.3, pp.604–609, 2013.
[15] J. Agbinya, "Interpolation using the discrete cosine transform," Electron. Lett., vol.28, no.20, pp.1927–1928, 1992.
[16] R. Dugad and N. Ahuja, "A fast scheme for image size change in the compressed domain," IEEE Trans. Circuits Syst. Video Technol., vol.11, no.4, pp.461–474, 2001.
[17] J. Mukherjee and S.K. Mitra, "Image resizing in the compressed domain using subband DCT," IEEE Trans. Circuits Syst. Video Technol., vol.12, no.7, pp.620–627, 2002.
[18] V. Patil, R. Kumar, and J. Mukherjee, "A fast arbitrary factor video resizing algorithm," IEEE Trans. Circuits Syst. Video Technol., vol.16, no.9, pp.1164–1171, 2006.
[19] C. Salazar and T.D. Tran, "A complexity scalable universal DCT domain image resizing algorithm," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.4, pp.495–499, 2007.
[20] S.-J. Park and J. Jeong, "Hybrid image upsampling method in the discrete cosine transform domain," IEEE Trans. Consum. Electron., vol.56, no.4, pp.2615–2622, 2010.
[21] H. Lim and H. Park, "A ringing-artifact reduction method for block-DCT-based image resizing," IEEE Trans. Circuits Syst. Video Technol., vol.21, no.7, pp.879–889, 2011.
[22] H. Park, Y. Park, and S.K. Oh, "L/M-fold image resizing in block-DCT domain using symmetric convolution," IEEE Trans. Image Process., vol.12, no.9, pp.1016–1034, 2003.
[23] Y.S. Park and H.W. Park, "Arbitrary-ratio image resizing using fast DCT of composite length for DCT-based transcoder," IEEE Trans. Image Process., vol.15, no.2, pp.494–500, 2006.
[24] K.-L. Chung, W.-J. Yang, W.-M. Yan, and C. Fuh, "New joint demosaicing and arbitrary-ratio resizing algorithm for color filter array based on DCT approach," IEEE Trans. Consum. Electron., vol.56, no.2, pp.783–791, 2010.
[25] E.-L. Tan, W.-S. Gan, and S. Mitra, "Fast arbitrary resizing of images in the discrete cosine transform domain," IET Image Processing, vol.5, no.1, pp.73–86, 2011.
[26] Z. Wu, H. Yu, and C.W. Chen, "A new hybrid DCT-wiener-based interpolation scheme for video intra frame up-sampling," IEEE Signal Process. Lett., vol.17, no.10, pp.827–830, 2010.
[27] K.-W. Hung and W.-C. Siu, "Hybrid DCT-Wiener-based interpolation via learnt Wiener filter," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1419–1423, 2013.
[28] K.-W. Hung and W.-C. Siu, "Novel DCT-based image up-sampling using learning-based adaptive k-NN MMSE estimation," IEEE Trans. Circuits Syst. Video Technol., vol.24, no.12, pp.2018–2033, 2014.
[29] J. Ma, G. Plonka, and M.Y. Hussaini, "Compressive video sampling with approximate message passing decoding," IEEE Trans. Circuits Syst. Video Technol., vol.22, no.9, pp.1354–1364, 2012.
[30] Y. Liu, M. Li, and D.A. Pados, "Motion-aware decoding of compressed-sensed video," IEEE Trans. Circuits Syst. Video Technol., vol.23, no.3, pp.438–444, 2013.
[31] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," IEEE Trans. Signal Process., vol.57, no.9, pp.3488–3497, 2009.
[32] L.-W. Kang and C.-S. Lu, "Distributed compressive video sensing," 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1169–1172, 2009.
[33] J.-L. Starck, M. Elad, and D.L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," IEEE Trans. Image Process., vol.14, no.10, pp.1570–1582, 2005.
[34] J.-W. Kang, M. Gabbouj, and C.-C.J. Kuo, "Sparse/DCT (S/DCT) two-layered representation of prediction residuals for video coding," IEEE Trans. Image Process., vol.22, no.7, pp.2711–2722, 2013.
[35] D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, no.4, pp.1289–1306, 2006.
[36] E.J. Candé and M.B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag., vol.25, no.2, pp.21–30, 2008.
[37] R.G. Baraniuk, "Compressive sensing," IEEE Signal Process. Mag., vol.24, no.4, pp.118–121, 2007.
[38] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," SIAM J. Sci. Comput., vol.20, no.1, pp.33–61, 1998.
[39] R. Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuits," IEEE Trans. Circuits Syst. Video Technol., vol.7, no.1, pp.158–171, 1997.
[40] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Trans. Inf. Theory, vol.53, no.12, pp.4655–4666, 2007.
[41] Z. Li, J. Xie, D. Tu, and Y.-J. Choi, "Sparse signal recovery by stepwise subspace pursuit in compressed sensing," International Journal of Distributed Sensor Networks, vol.2013, pp.1–5, 2013.
[42] E.J. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," IEEE Trans. Inf. Theory, vol.52, no.12, pp.5406–5425, 2006.
[43] Z. Wang, 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.

Appendix

One dimension DCT is expressed as follow:

$$X(k) = \sum_{i=0}^{N-1} x(i) cos(\frac{\pi(2i+1)k}{2N}) \tag{A$\cdot$1}$$

where N is the length of the signal. $x(i)$ is the signal. $X(k)$ is the DCT efficient. The matrix of the 1-D DCT is expressed as follow:

$$\begin{aligned}\mathbf{X} &= \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} \\ &= \begin{bmatrix} cos(\frac{\pi(2\times0)\times0}{2N}) & \cdots & cos(\frac{\pi(2\times(N-1))\times0}{2N}) \\ cos(\frac{\pi(2\times0)\times1}{2N}) & \cdots & cos(\frac{\pi(2\times(N-1))\times1}{2N}) \\ \vdots & \ddots & \vdots \\ cos(\frac{\pi(2\times0)\times(N-1)}{2N}) & \cdots & cos(\frac{\pi(2\times(N-1))\times(N-1)}{2N}) \end{bmatrix} \times \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \\ &= C_{1D_N}\mathbf{x}\end{aligned} \tag{A$\cdot$2}$$

Two dimension DCT is expressed as follow:

$$X(k,l) = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} x(i,j) cos(\frac{\pi(2i+1)k}{2M}) cos(\frac{\pi(2j+1)l}{2N}) \tag{A$\cdot$3}$$

where k, l are the coordinates of the DCT coefficients. i, j are the coordinates of the two dimension signal. M and N indicate the length and width of the signal.

$$
\begin{aligned}
\mathbf{X} &= \begin{bmatrix} X(0,0) \\ \vdots \\ X(0,N-1) \\ X(1,0) \\ \vdots \\ \vdots \\ X(M-1,N-1) \end{bmatrix} \\
&= \begin{bmatrix}
cos(\frac{\pi(2\times0)\times0}{2M})cos(\frac{\pi(2\times0)\times0}{2N}) & \cdots & \cdots & \cdots & cos(\frac{\pi(2\times(M-1))\times0}{2M})cos(\frac{\pi(2\times(N-1))\times1}{2N}) \\
cos(\frac{\pi(2\times0)\times0}{2M})cos(\frac{\pi(2\times0)\times1}{2N}) & \cdots & \cdots & \cdots & cos(\frac{\pi(2\times(M-1))\times0}{2M})cos(\frac{\pi(2\times(N-1))\times1}{2N}) \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
cos(\frac{\pi(2\times0)\times0}{2M})cos(\frac{\pi(2\times0)\times(N-1)}{2N}) & \cdots & \cdots & \cdots & cos(\frac{\pi(2\times(M-1))\times0}{2M})cos(\frac{\pi(2\times(N-1))\times(N-1)}{2N}) \\
cos(\frac{\pi(2\times0)\times1}{2M})cos(\frac{\pi(2\times0)\times0}{2N}) & \cdots & \cdots & \cdots & cos(\frac{\pi(2\times(M-1))\times1}{2M})cos(\frac{\pi(2\times(N-1))\times0}{2N}) \\
cos(\frac{\pi(2\times0)\times1}{2M})cos(\frac{\pi(2\times0)\times1}{2N}) & \cdots & \cdots & \cdots & cos(\frac{\pi(2\times(M-1))\times1}{2M})cos(\frac{\pi(2\times(N-1))\times1}{2N}) \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
cos(\frac{\pi(2\times0)\times(M-1)}{2M})cos(\frac{\pi(2\times0)\times(N-1)}{2N}) & \cdots & \cdots & \cdots & cos(\frac{\pi(2\times(M-1))\times(M-1)}{2M})cos(\frac{\pi(2\times(N-1))\times(N-1)}{2N})
\end{bmatrix} \\
&\qquad \times \begin{bmatrix} x(0,0) \\ \vdots \\ x(0,N-1) \\ x(1,0) \\ \vdots \\ \vdots \\ x(M-1,N-1) \end{bmatrix} \\
&= \begin{bmatrix}
C_{1D_N} cos(\frac{\pi(2\times0)\times0}{2M}) & C_{1D_N} cos(\frac{\pi(2\times1)\times0}{2M}) & \cdots & C_{1D_N} cos(\frac{\pi(2\times(M-1))\times0}{2M}) \\
C_{1D_N} cos(\frac{\pi(2\times0)\times1}{2M}) & C_{1D_N} cos(\frac{\pi(2\times1)\times1}{2M}) & \cdots & C_{1D_N} cos(\frac{\pi(2\times(M-1))\times1}{2M}) \\
\vdots & \cdots & \ddots & \vdots \\
C_{1D_N} cos(\frac{\pi(2\times0)\times(M-1)}{2M}) & C_{1D_N} cos(\frac{\pi(2\times1)\times(M-1)}{2M}) & \cdots & C_{1D_N} cos(\frac{\pi(2\times(M-1))\times(M-1)}{2M})
\end{bmatrix}
\begin{bmatrix} x(0,0) \\ \vdots \\ x(0,N-1) \\ x(1,0) \\ \vdots \\ \vdots \\ x(M-1,N-1) \end{bmatrix} \\
&= C_{1D_M} \otimes C_{1D_N} \mathbf{x} \\
&= C_{2D_{MN}} \mathbf{x}
\end{aligned}
\tag{A·4}
$$

Meng Zhang is currently a Professor in National ASIC System Engineering Technology Research Center, Southeast University, Nanjing, China. His research interests include signal and image processing, digital communication and information security, system and network optimization etc.

Tinghuan Chen received the B.S. degree in electricity engineering from the Southeast University of China in 2014, and is current pursuing for Master degree in National ASIC center, Southeast University, China. He is interested in image processing, coding theory, and digital communication.

Xuchao Shi received the B.S. degree in statistics from the Southeast University of China in 2014, and is current pursuing for Master degree in Department of Applied Math and Statistics, Johns Hopkins University. He is interested in analysis of time series and statistical inference.

Peng Cao received the B.S., M.S. and Ph.D. degrees in Information Engineering and Electrical Engineering from Southeast University in 2002, 2005 and 2010 respectively. His research interests mainly include digital signal and image processing, image/video compression.