Low Bit-rate Image Coding via Interpolation Oriented Adaptive Down-sampling

Yongbing Zhang*[a], Jian Zhang[a], Ruiqin Xiong[b], Debin Zhao[a], Siwei Ma[b]
[a]Department of Computer Science, Harbin Institute of Technology, Harbin, 150001, China;
[b]Institute of Digital Media, Peking University, Beijing, 100871, China

ABSTRACT

An interpolation oriented adaptive down-sampling algorithm is proposed for low bit-rate image coding in this paper. Given an image, the proposed algorithm is able to obtain a low resolution image, from which a high quality image with the same resolution as the input image can be interpolated. Different from the traditional down-sampling algorithms, which are independent from the interpolation process, the proposed down-sampling algorithm hinges the down-sampling to the interpolation process. Consequently, the proposed down-sampling algorithm is able to maintain the original information of the input image to the largest extent. The down-sampled image is then fed into JPEG. A total variation (TV) based post processing is then applied to the decompressed low resolution image. Ultimately, the processed image is interpolated to maintain the original resolution of the input image. Experimental results verify that utilizing the down-sampled image by the proposed algorithm, an interpolated image with much higher quality can be achieved. Besides, the proposed algorithm is able to achieve superior performance than JPEG for low bit rate image coding.

Keywords: Adaptive down-sampling, image interpolation, total variation, image coding

1. INTRODUCTION

With the accelerative growth of high-performance computers and source-channel coding, a great progress of image and video coding standards have been recently achieved. Both image and video coding first transform the spatial signal into a space, which has a sparse representation (e.g., DCT or wavelet domain), and then the transformed coefficients are entropy encoded. In the recent years, many algorithms have been proposed to improve the image and video compression performances, during which one typical way is to sparsely sample the signal and then obtain the original signal representation [1]. This method highly coincides with the emerging compressive sensing (CS) theory [2], which shows that it is possible, at least theoretically, to obtain compact signal representation by a greatly reduced number of random samples.

Inspired by the CS theory, we investigate the problem of obtaining compact image representation via sparse sampling in the spatial domain. As indicated in [1], the fact that most natural images have an exponentially decaying power spectrum suggests the possibility of interpolation-based compact representation of images. Obviously, a scene contains predominate smooth regions has a high possibility of obtaining satisfactory recovery by interpolating from the sparsely down-sampled low resolution image. The difficulty resides in the reconstruction of high frequency contents (e.g., edge regions). In terms of image and video coding, faithfully reconstructing edges without large phase errors is of particular importance, since it plays a key role to the perceptual quality of decoded image.

As we know, down-sampling is a statistical process to make a digital image smaller by removing pixels. On the contrary, up-sampling (termed as image interpolation in this paper) is a statistical process to make a digital image larger by adding pixels. Generally speaking, down-sampling can be divided into direct down-sample and adaptive down-sample. The former is performed by directly retaining the corresponding pixels within the input image, while the latter obtains the down-sampled pixels according to the image content as well as sampling ratio. The direct down-sampling is much easier compared with the adaptive one. However, aliasing artifacts are inevitable around the edge regions due to the evenly pixels fetching operations. To alleviate the aliasing artifacts, many adaptive down-sampling methods can be applied. Wavelet transform [3] and Discrete Cosine transform [4] are two common methods, where the down-sampling can be achieved by fetching the low frequency of the transformed coefficients. Besides, pre-filtering can also be applied before uniform spatial down sampling. For example, a spatially varying, directional low-pass pre-filtering is applied

*ybzhang@jdl.ac.cn

before down-sampling in [1]. Although adaptive down-sampling methods is able to remedy the disadvantage of direct down-sampling, the up-sampled image by phase invariant interpolation can not guarantee good quality in terms of objective criterion. Take 1:2 interpolation for example, direct down-sampling method can keep 1/4 pixels intact compared with the input image. Consequently, interpolating the direct down-sampled image may generate an image with higher PSNR values compared to that generated by interpolating the adaptive down-sampled method, although the latter may have better visual quality than the former.

To answer the above mentioned challenges, we proposed an interpolation oriented adaptive down-sampling algorithm and apply it in the low bit-rate image coding in this paper. The block diagram of the proposed image compression is illustrated in Fig. 1, which consists of four components: the proposed interpolation adaptive down-sampling, JPEG [5] or encoder, total variation (TV) [6] based post-processing and interpolation. The contribution of this paper concentrates in interpolation oriented adaptive down-sampling and TV based post-processing. Different from the traditional down-sampling methods, which are independent from interpolation process, the proposed algorithm hinges the down-sampling to the interpolation process. That is to say, when performing down-sampling, the influences it brought to the interpolation process should be jointly considered. The aim of the proposed down-sampling is to maintain the original information of the input image to the largest extent. Specially, for each pixel $X_{i,j}$ to be down-sampled, all the neighboring pixels utilizing $X_{i,j}$ during the interpolation process are identified. The least-squares algorithm is then employed to down-sample the pixels via minimizing the sum of square errors between the original and the interpolated intensity values within all the identified neighboring pixels. The down-sampled image is then fed into JPEG. The decompressed low resolution image is then post-processed by a TV method. After post-processing, the low resolution image is interpolated to the original resolution of the input image. Experimental results verify that utilizing the down-sampled image by the proposed algorithm, an interpolated image with much higher quality can be achieved. Besides, the proposed algorithm is able to achieve superior performance than JPEG for low bit rate image coding.

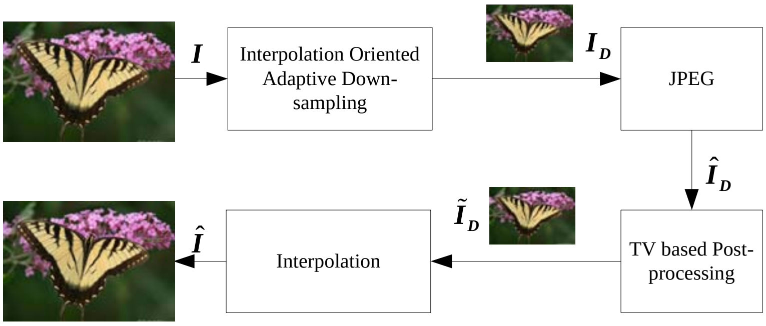


Fig. 1: Block diagram of the proposed image compression system

The organization of this paper is as follows. Section 2 presents the detail description of the proposed interpolation oriented adaptive down-sampling. Section 3 provides the TV based post-processing. Section 4 gives the experimental results, and finally this paper is concluded in Section 5.

2. INTEROLATION ORIENTED ADAPTIVE DOWN-SAMPLING

In this section, we will give the detail description of the proposed interpolation oriented adaptive down-sampling. For the convenience of expression, we will take the 2:1 down-sampling processing as an example in this paper, and it can be easily extended to other down-sampling ratios. The sketch map of the proposed interpolation oriented adaptive down-sampling is illustrated in Fig .2. We hinge the down-sampling process to the interpolation process, i.e, for each pixel $X_{i,j}$ to be down-sampled, all the interpolated pixels utilizing $X_{i,j}$ during the interpolation are identified and the sum of the square errors between the input image and the identified interpolated pixels are minimized. As illustrated in Fig. 2, for the pixel $X_{i,j}$, located at the center of the nine down-sampled pixels (black circles), all the pixels, indicated by the gray circles, are identified. The optimization target of the proposed down-sampling algorithm is to minimize the sum of the square errors between the identified pixels and the original pixels within the input image. It is noted that bilinear interpolation is utilized in Fig. 2, and it can be easily extended to arbitrary interpolation methods. Let $N\left(X_{i,j}\right)$ denote the identified pixels, which will involve $X_{i,j}$ during interpolation, the optimal down-sampled value of $X_{i,j}$ should be

$$X_{i,j}^{*} = \min_{\hat{X}_{i,j}} \left\{ \left\| X_{i,j} - \hat{X}_{i,j} \right\|^{2} + \left\| N\left(X_{i,j}\right) - N\left(\hat{X}_{i,j}\right) \right\|^{2} \right\}, \tag{1}$$

where $X_{i,j}^{*}$ represents the optimal down-sampled value of pixel $X_{i,j}$, and $N\left(\hat{X}_{i,j}\right)$ represents the interpolated values involving $X_{i,j}$ during interpolation process.

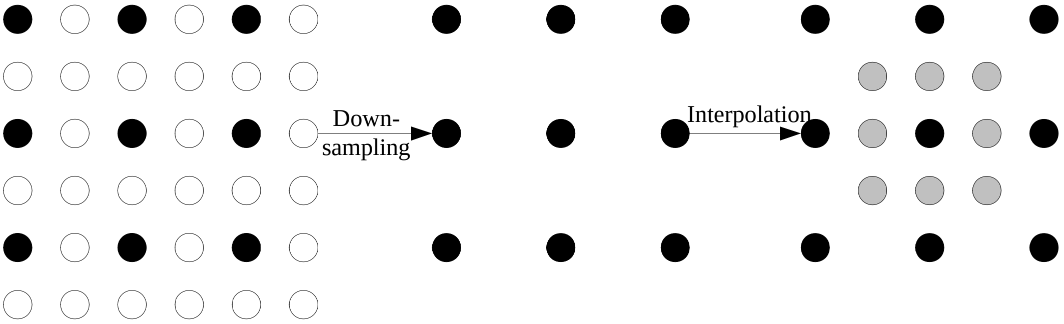


Fig. 2: Sketch map of the proposed interpolation oriented adaptive down-sampling

It should be noted that in Fig. 2, the value of interpolated pixel, indicated by the gray circle, depends on not only the center down-sampled values, but also other adjacent down-sampled pixels. To obtain the optimal values of all the down-sampled pixels, we should consider all the down-sampled values simultaneously.

Assume the input image I is of $M \times N$ dimension, and define the down-sampled image as a column vector with the form of $\mathbf{X} = \left[X_{0,0}, X_{0,2}, ..., X_{0,N-2,} X_{2,0}, X_{2,2}, ..., X_{2,N-2,} ..., X_{M-2,0}, X_{M-2,2}, ..., X_{M-2,N-2,} \right]^{T}$, the interpolation matrix, which is composed of interpolation weights, can be expressed as

$$\mathbf{H} = \begin{bmatrix} h_{0,0} & h_{0,1} & .. & h_{0,\frac{M}{2}\times\frac{N}{2}-1} \\ h_{1,0} & h_{1,1} & .. & h_{1,\frac{M}{2}\times\frac{N}{2}-1} \\ h_{2,0} & h_{2,1} & .. & h_{2,\frac{M}{2}\times\frac{N}{2}-1} \\ .. & .. & .. & .. \\ h_{M\times N-1,0} & h_{M\times N-1,1} & .. & h_{M\times N-1,\frac{M}{2}\times\frac{N}{2}-1} \end{bmatrix}, \tag{2}$$

where $h_{m,n}$ represents the interpolation weight contributed by the *n*th down-sampled pixel during the interpolation of the *m*th pixel, with $0 \le m \le M \times N - 1$ and $0 \le n \le \left(M \times N\right)/4 - 1$. It is noted that here the interpolated pixel is represented in a concatenated and lexicographically order. Consequently, the objective function to derive the optimal down-sampled image is to minimize

$$J = \left(\mathbf{HX} - \mathbf{I}\right)^{T} \left(\mathbf{HX} - \mathbf{I}\right), \tag{3}$$

with $\mathbf{I}=\left[I_{0,0},I_{0,1},\dots,I_{0,N-1,}I_{1,0},I_{1,1},\dots,I_{1,N-1,}\dots,I_{M-1,0},I_{M-1,1},\dots,I_{M-1,N-1,}\right]^T$. Set the Partial derivative of J to be zero, that is to say

$$\frac{\partial J}{\partial \mathbf{X}}=\mathbf{H}^T\left(\mathbf{HX}-\mathbf{I}\right)=0\,, \tag{4}$$

the optimal down-sampled image can be expressed as

$$\mathbf{X}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{I}\,. \tag{5}$$

It should be noted that the dimension of $\mathbf{H}$ is $(MN)\times(MN/4)$, which is of too heavy space complexity. To reduce the space complexity, we perform down-sampling block by block. Obviously, as long as the interpolation weights are known, the proposed down-sampling algorithm is able to alleviate the aliasing artifacts while ensure the interpolated image with the best quality.

3. TV BASED POST-PROCESSING

The down-sampled image is directly fed into JPEG, which will inevitably bring distortions. To make the decompressed image much smoother, a TV based post processing method is introduced. Denote $\hat{\mathbf{X}}$ as the decompressed image by JPEG, the objective of TV based post processing method is to derive $\mathbf{Y}$, which satisfies

$$\mathbf{Y}=\min_{\mathbf{Y}}\left\|\mathbf{Y}-\hat{\mathbf{X}}\right\|^2+2\lambda\left\|\mathbf{Y}\right\|_{TV}\,, \tag{6}$$

where $\mathbf{Y}$ is the desired unknown image to be recovered, $\left\|\bullet\right\|_{TV}$ is a discrete TV (semi)-norm. In this paper, the isotropic TV defined by (see [6]) is adopted as

$$\mathbf{Y}\in\mathbf{R}^{m\times n}, TV(\mathbf{Y})=\sum_{i=0}^{M-2}\sum_{j=0}^{N-2}\sqrt{\left(y_{i,j}-y_{i+1,j}\right)^2+\left(y_{i,j}-y_{i,j+1}\right)^2}+\sum_{i=0}^{M-2}\left|y_{i,N-1}-y_{i+1,N-1}\right|+\sum_{j=0}^{N-2}\left|y_{M-1,j}-y_{M-1,j+1}\right|, \tag{7}$$

where in the above formula we assume the (standard) reflexive boundary conditions $y_{M-1,j}-y_{M-2,j}=0,\ \forall j$ and $y_{i,N-1}-y_{i,N-2}=0,\ \forall i$. The method in [7] is adopted to derive the optimal solution of Eq. (6). After the post-processing, the low-resolution image is interpolated to maintain the original resolution of the input image.

4. EXPERIMENTAL RESULTS

In this section, we will provide extensive experimental results to verify the superiority of the proposed image compression system. Four images: *Lena* (512x512), *Peppers* (512x512), *Boat* (512x512) and *Splash* (512x512) were selected to carry out the experiments. In the experiment, the block size of the adaptive down-sampling is set to be 16x16, and the interpolation method utilized during the adaptive down-sampling process is bilinear interpolation. In table 1, we give the PSNR values of the interpolated images, where the down-sampled images are interpolated directly (without compression). It can be observed that under the same down-sampling algorithms the method in [7] always outperforms the bilinear method. However, the gap between the bilinear method under the proposed down-sampled method (the third column) and the method in [7] under the directly down-sampled image (the second column) decreases greatly. Bilinear in the third column even outperforms the method in [7] in the second column for *Peppers* and *Splash*. Another observation is that the performances of both bilinear and the method in [7] get improved under the proposed down-sampling algorithm compared to those under the direct down-sampling method. This is greatly attributed to the property that the proposed down-sampling algorithm is able to maintain as more information as possible compared with the direct down-sampling.

We also present the comparison results of the proposed algorithm, JPEG and other methods. For each test image, we first use direct down-sampling and adaptive down-sampling algorithms to obtain the low resolution image, and then the low resolution image is compressed by JPEG. Next, the decompressed low resolution image is fed into the proposed TV based post-processing. Ultimately, the processed images are interpolated by the method in [7], which is reported to be the best interpolation method. It is noted that since [7] sometimes fails for some flat regions, i.e., the ghost artifacts in Fig. 3. To remove the ghost artifacts, we compare the interpolated images by the method in [7] and bilinear method, for each pixel location, if the absolute error between the two images is above 50, we replace the pixel values in [7] by those generated by bilinear.

Table 1. PSNR comparisons of different images by different down-sampling and interpolation methods

Images	Direct down-sampling		Proposed adaptive down-sampling	
	bilinear	Method in [7]	bilinear	Method in [7]
Lena	33.442	34.763	34.402	35.296
Peppers	31.602	31.985	32.917	33.321
Boat	29.152	30.249	29.708	30.942
Splash	33.473	33.544	34.576	34.662

Fig. 3: ghost artifacts generated by the method in [7]

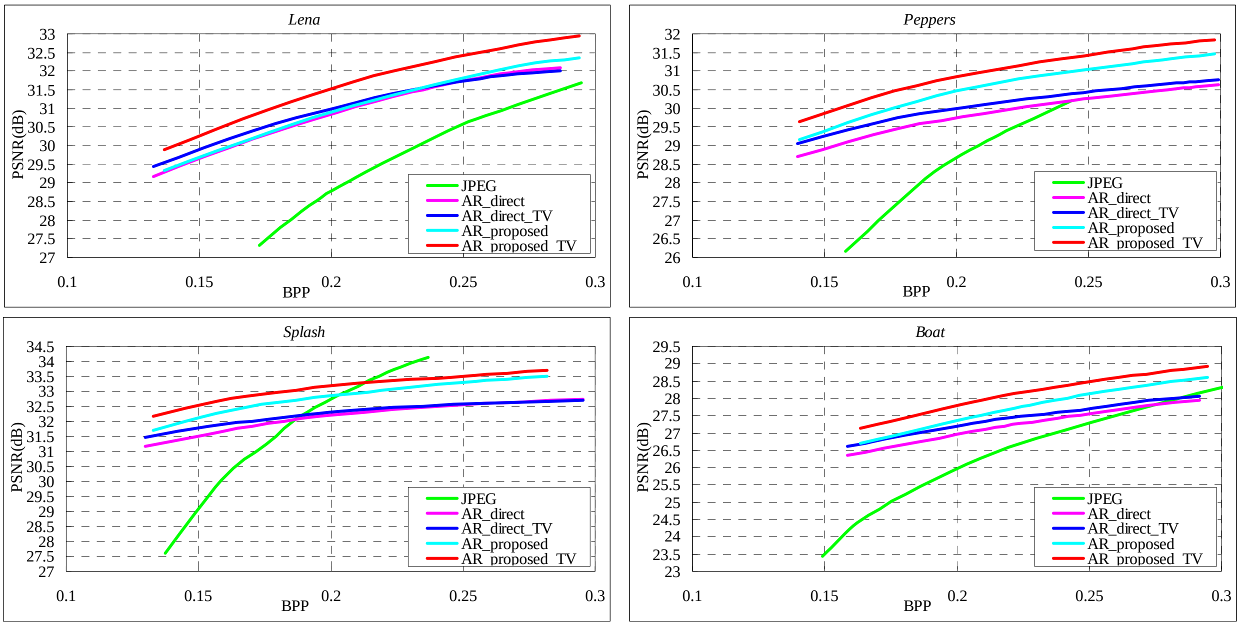
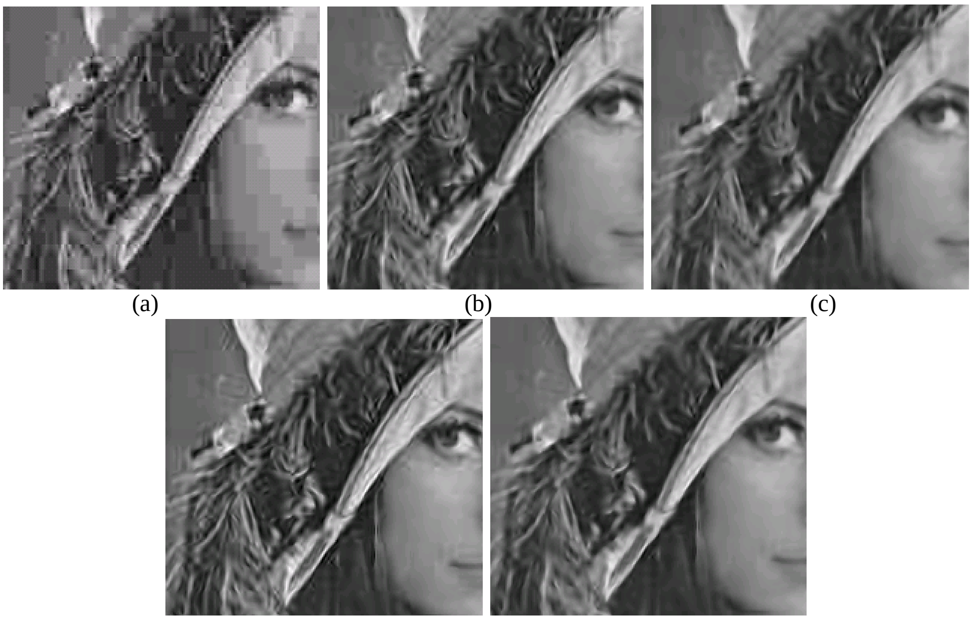


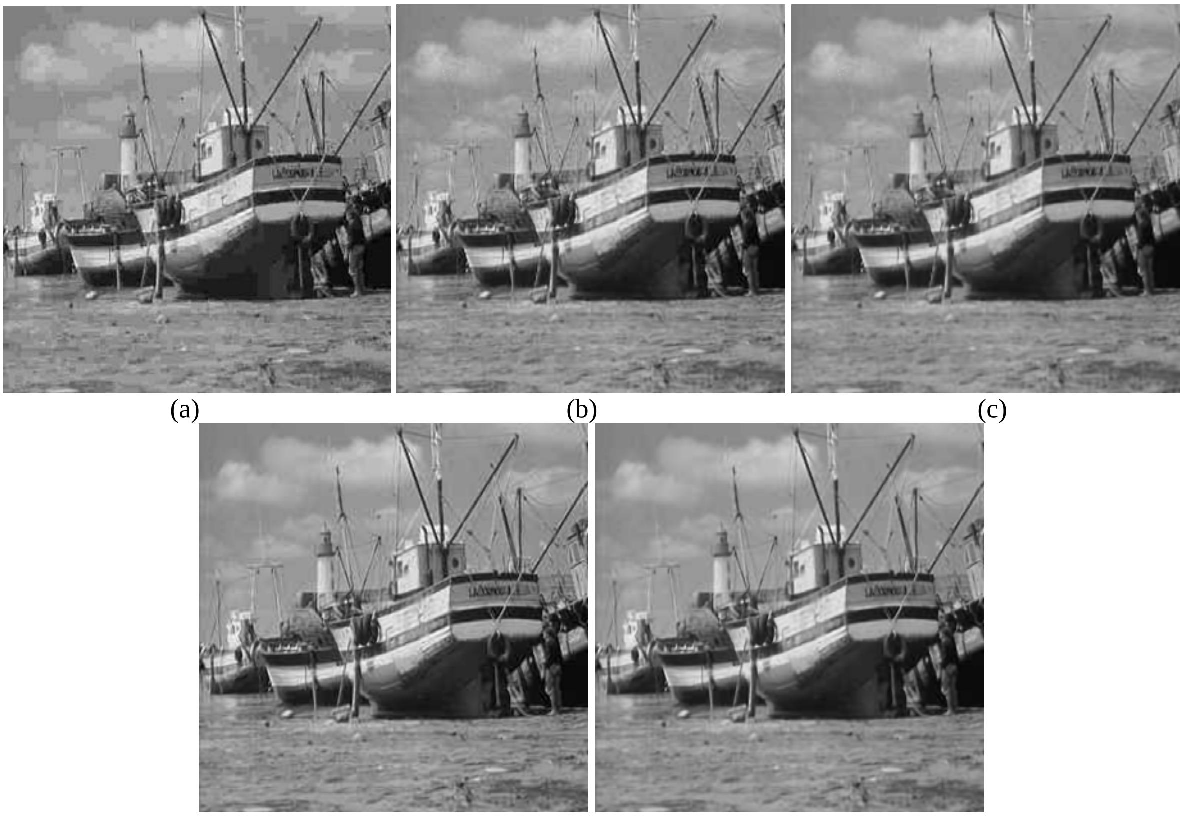
Fig. 4: Comparisons of different methods at various bpps



(a) (b) (c)

(d) (e)

Fig. 5: Comparisons of different methods at 0.2 bpp for *Lena*. (a) JPEG, 28.892dB; (b) AR_direct, 30.874dB; (c) AR_direct_TV, 31.015dB ; (d) AR_adaptive, 31.077dB ; (e) AR_adaptive_TV, 31.717dB



(a) (b) (c)

(d) (e)

Fig. 6: Comparisons of different methods at 0.25 bpp for *Boat*. (a) JPEG, 27.317dB; (b) AR_direct, 27.567dB; (c) AR_direct_TV, 27.722dB; (d) AR_adaptive, 28.191dB; (e) AR_adaptive_TV, 28.544dB

The comparisons of different methods at various bpps are provided in Fig. 4, where AR_direct represents the direct down-sampling with the interpolation method in [7], AR_direct_TV represents the direct down-sampling with the interpolation method in [7] and the TV based post-processing, AR_proposed represents the proposed adaptive down-sampling with the interpolation method in [7], and AR_proposed_TV represents the proposed adaptive down-sampling with the interpolation method in [7] and TV based post-processing. It is observed that at the low bpps, downsampling-JPEG-interpolation method is able to achieve superior performance than JPEG. Another observation is that AR_proposed and AR_proposed_TV have higher PSNR values than AR_direct and AR_direct_TV do, which verify that the proposed down-sampling algorithm is able to improve the quality of interpolated image under the same interpolation method. In addition, applying the TV based post-processing the performance gets improved for both the direct down-sampling and the proposed adaptive down-sampling. This is because the TV based post-processing has a desirable ability to remove the blocking artifacts produced by JPEG, which will improve the quality of the decompressed low resolution image.

We also represent the visual quality comparisons for *Lena* and *Boat* in Figs 5 and 6, respectively. In Fig. 5 (a), there exhibits significant blocking artifacts in the JPEG compressed image, however the blocking artifacts are greatly removed in the images generated by other four methods. Compared with the direct down-sampling results, the proposed method achieves better visual quality. For example, in Fig. 5(d) and Fig. 5(e) the rims of the hat in the top right corner can be more easily observed than those in Fig. 5(b) and Fig. 5(c). Similarly, in Fig. 6(a), significant blocking artifacts can be observed, whereas no blocking artifacts can be perceived in Fig. 6(b)-(e). The proposed adaptive down-sampling algorithm still provides good visual quality than the direct down-sampling methods. For example, the masts in Fig. 6(d) and (e) can be more clearly observed than those in Fig. 6(b) and Fig. 6(c).

5. CONCLUSION

We have introduced an interpolation oriented adaptive down-sampling algorithm for low bit-rate image coding. Different from other down-sampling process, the proposed algorithm hinges the down-sampling process to the interpolation process. Consequently, for each input image, the proposed down-sampling algorithm is able to generate a low resolution image, from which a high quality image with the same resolution as the input image can be interpolated. The generated low resolution image is fed into JPEG. A TV based post processing is then applied to the decompressed low resolution image. Ultimately, the processed image is interpolated to maintain the original resolution of the input image. Experimental results verify that the proposed algorithm is able to outperform JPEG for low bit rate image coding. Besides, it evens achieves superior performance than the other existing directly down-sampling and interpolation methods.

REFERENCES

1. Wu, X., Zhang, X. and Wang, X., “Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion” IEEE Trans. on Image process, 18(3), 552-561 (2009).
2. Cands, E., “Compressive sampling,” Proc. Int. Congr. Mathematics, 1433-1452 (2006).
3. Vetterli, M. and Kovacevic, J., [Wavelets and Subband Coding], Prentice Hall, (1995).
4. Zeng, Y., Cheng, L., Bi, G. and Kot, A. C., “Integer DCTs and fast algorithms” Signal Processing, IEEE Transactions on Acoustics, Speech, and Signal Processing, 49 (11), 2774-2782 (2001).
5. Neelamani, R., de Queiroz, R, Z. Fan, Dash, S. and Baraniuk, R.G., “JPEG compression history estimation for color images” IEEE Trans. on Image process, 15 (6), 1365-1378 (2006).
6. Beck, A. and Teboulle, M., “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems” IEEE Trans. on Image process, 18 (11), 2419-2434 (2009).
7. Zhang, X. and Wu, X., “Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation” IEEE Trans. on Image process, 17 (6), 887-896 (2008).