Perceptual Image Compression via Adaptive Block-Based Super-Resolution Directed Down-Sampling

Lin Ma, Songnan Li, King N. Ngan
Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, NT. Hong Kong
Email: {lma, snli, knngan}@ee.cuhk.edu.hk

Abstract—**In this paper, we propose a novel perceptual image coding scheme via adaptive block-based super-resolution directed down-sampling. At the encoder side, for each macroblock of a given image, Rate Distortion Optimization (RDO) determines whether it is encoded at the original or down-sampled resolution. The down-sampling process is directed by super-resolution, which generates the down-sampled block by minimizing the reconstruction errors between the original macroblock and the one restored by the corresponding super-resolution method. At the decoder side, in order to reduce the complexity, the super-resolution method reconstructs the full-resolution macroblock in the DCT domain together with the inverse DCT. Experimental results have demonstrated that the proposed method can produce higher quality images in terms of both PSNR and visual quality compared with the existing methods.**

I. INTRODUCTION

With the development of the imaging technology, more and more images with high qualities and large spatial resolutions are provided for satisfying people's visual experiences. However, it issues a great challenge to image transmission and storage. Therefore, a more efficient image compression scheme is highly demanded, which can ensure a higher image quality with a smaller quantity of bits for the image representation.

Based on that most images can be obtained via interpolation from sparse pixel data yielded by a signal-sensor camera [9], and natural images exhibit high spatial correlations between neighboring pixels [2], many interpolation-based image coding methods [3]-[7] have been proposed. In [5], 2×2 average operator is employed for decimation before JPEG compression. A replication filter and a Gaussian filter are used for restoring the image from the decimated one. The theoretical down-sampling model is studied and compared in [6]. Tsaig et al. [7] propose to code the filter parameters as the side information for better reconstruction at the decoder side. In [3] [4], the authors suggest coding the down-sampled low-resolution image during encoding and recovering the high frequency components during decoding by interpolating the compact image representation generated by sparse sampling in the spatial domain. Although the predominated smooth regions of an image can be satisfactorily recovered by interpolation, the reconstruction of high frequency components of the edge and texture regions still remains a great challenge. In order to overcome the problem, Wu et al. [1] employ the piecewise autoregressive model to handle the large phase errors during the interpolation of the image edges. However, there is a heavy computation burden at the decoder side due to the optimal block estimation problem driven by the autoregressive model. Moreover, Lin et al. [2] propose a new image coding method based on the adaptive decision of appropriate down-sampling directions/ratios and quantization steps to achieve higher coding quality. The method tries to avoid down-sampling a macroblock along the direction of high spatial variations, which signals the existence of edges and other image features with great impact on the perceptual visual quality. In [2], the down-sampled pixels are obtained by averaging the neighboring pixels of the original resolution image. Although it can somewhat reduce the aliasing artifacts introduced by direct sampling, the blurring artifacts will be introduced. Also as the down-sampling process is independent of the following super-resolution process, the reconstruction errors between the original and the restored macroblock cannot be ensured to be the smallest. More recently, the JPEG2000 [15] and H.264/AVC [16] have been developed for achieving higher compression performances for images.

In order to tackle the aforementioned problems, we propose a novel perceptual image coding scheme via adaptive block-based super-resolution directed down-sampling. For each macroblock of a given image, whether down-sampling or not depends on the contents of the visual signal itself, which will be determined by the Rate Distortion Optimization (RDO) process [14]. And the joint method of down-sampling and super-resolution is proposed to minimize the reconstruction errors between the original and the restored macroblock inferred by the super-resolution method from the down-sampled block. At the decoder side, the super-resolution method performed in DCT domain is employed to recover the full-resolution macroblock for its simplicity.

The paper is organized as follows. Section II will introduce the proposed perceptual image compression framework, as well as the super-resolution method and super-resolution directed down-sampling process. Experimental results in Section III will demonstrate the coding efficiency of the proposed method. Finally, Section IV concludes the paper.

This work was partially supported by a grant from the Chinese University of Hong Kong under the Focused Investment Scheme (Project 1903003).

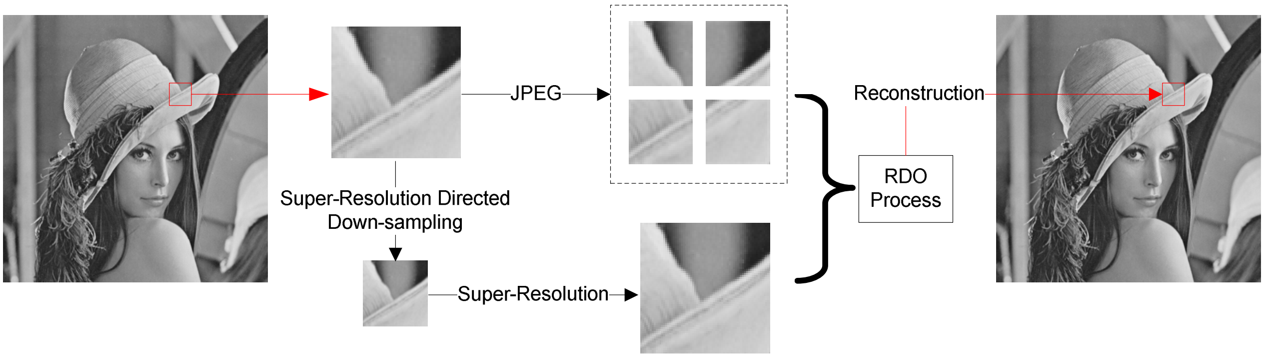

Figure 1. The proposed perceptual image compression framework

II. The Proposed Perceptual Image Compression Framework

The framework of our proposed method is illustrated in Figure 1. For each 16×16 macroblock of a given image, two candidate coding modes are provided. One is the traditional JPEG coding mode. The macroblock is divided into four 8×8 sub-blocks, each of which is processed by transformation, quantization, de-quantization, and the inverse transformation. The other one is the proposed Super-Resolution Directed Down-Sampling (SRDDS) mode. Firstly, an 8×8 low-resolution sub-block is obtained by down-sampling the full-resolution 16×16 macroblock according to the proposed SRDDS. Secondly, the 8×8 sub-block is transformed and quantized (the QP parameter is set as half of the one used in the JPEG mode). Then after de-quantization, the corresponding super-resolution method processed in DCT domain is performed together with the inverse transformation. Finally, the RDO process will determine which mode is employed to process the macroblock. The detailed information of the super-resolution in DCT domain and the SRDDS will be introduced in the following sections.

The proposed method differs with the schemes presented in the prior literatures [1]-[4] [8]. Although the image compression approaches in [1] [3] [4] [8] also employ the interpolation oriented adaptive down-sampling, they are designed to down-sample the whole original image for coding and try to recover the full-resolution image during the decoding process, which makes that the higher frequency components of the local texture and edge regions cannot be faithfully restored. Lin et al. presented an adaptive block-based down-sampling method in [2]. Three down-sampling modes with four different QP settings are employed, which results in high complexity of the encoder. Our experimental results reveal that only one mode is sufficient to improve the coding efficiency. Therefore, some down-sampling modes in [2] are not necessary, which just introduce the overhead information for the coded image. Also the down-sampling process in [2] is not optimized that cannot ensure higher quality reconstructed macroblocks.

A. Super-Resolution in DCT Domain

In order to reduce the computation complexity for the decoding process, the super-resolution performed in DCT domain [10] [11] is employed for generating the full-resolution macroblock from the down-sampled low-resolution sub-block.

In the decoding process, the de-quantization process results in $N\times N$ DCT coefficients $Coef_{N\times N}$ (N is equal to 8). The DCT coefficients are firstly extended into $2N\times 2N$ by inserting the remained positions of the $2N\times 2N$ matrix $Coef_{2N\times 2N}$ with 0, which is defined as:

$$Coef_{2N\times 2N} = \begin{bmatrix} Coef_{N\times N} & 0_{N\times N} \\ 0_{N\times N} & 0_{N\times N} \end{bmatrix}, \tag{1}$$

where $0_{N\times N}$ is $N\times N$ zero matrix. Then the inverse DCT is applied to $Coef_{2N\times 2N}$ for reconstructing the full-resolution macroblock by:

$$P_{2N\times 2N} = D^T_{2N\times 2N} \times (Coef_{2N\times 2N}) \times D_{2N\times 2N}, \tag{2}$$

where $P_{2N\times 2N}$ is the full-resolution macroblock obtained from the super-resolution method, $D_{2N\times 2N}$ denotes the DCT kernel for $2N$ samples, the superscript T denotes the transpose of the matrix. Therefore, (2) can be further expressed as:

$$P_{2N\times 2N}(m,n) = \sum_{p=0}^{2N-1}\sum_{q=0}^{2N-1} \alpha_p\alpha_q \cdot Coef(p,q) \cdot \cos\frac{\pi(2m+1)p}{4N}\cos\frac{\pi(2n+1)q}{4N} \tag{3}$$

where $0 \le \mathrm{m} \le 2\mathrm{N}-1, 0 \le \mathrm{n} \le 2\mathrm{N}-1$,

$$\alpha_\Delta = \begin{cases} 1/\sqrt{2\mathrm{N}}, & \Delta = 0 \\ \sqrt{1/\mathrm{N}}, & 1 \le \Delta \le 2\mathrm{N}-1 \end{cases}$$

and Δ represents p or q. For the super-resolution method in DCT domain, the full-resolution macroblock can be reconstructed during the inverse transformation, which can significantly reduce the complexity of the decoder. Moreover, a fast algorithm of the super-resolution method is presented in [10], which only requires 3.1874 multiplications for each pixel.

B. The Proposed Super-Resolution Directed Down-Sampling (SRDDS)

As aforementioned, the decoder employs the simple super-resolution method performed in DCT domain for up-sampling the low-resolution sub-block to the full-resolution macroblock. Therefore, in order to minimize the reconstruction error, an optimized low-resolution sub-block needs to be generated from the original block by considering the super-resolution process. It can be formulated as:

$$\hat{b}_{N\times N} = \arg\min_b\{\|B_{2N\times 2N} - S(b_{N\times N})\|_2^2\}, \tag{4}$$

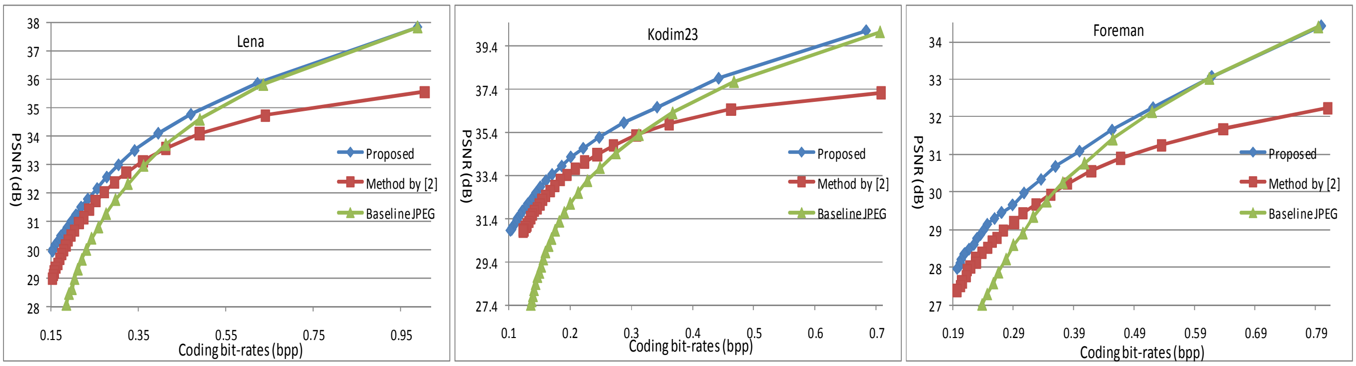

Figure 2. PSNR comparisons of the proposed scheme, baseline JPEG, and the method by [2]

where $S(b_{N\times N})$ is the enlarged $2N\times 2N$ block by the super-resolution method presented in Section II *A*, $B_{2N\times 2N}$ is the original full resolution $2N\times 2N$ macroblock. The solution of (4) is the optimized down-sampled low-resolution sub-block $\hat{b}_{N\times N}$, which yields the smallest reconstruction error.

The super-resolution process (2) can be expressed as:

$$\begin{aligned} P_{2N\times 2N} &= D_{2N\times 2N}^{T} \times \begin{bmatrix} Coef_{N\times N} & 0_{N\times N} \\ 0_{N\times N} & 0_{N\times N} \end{bmatrix} \times D_{2N\times 2N} \\ &= D_{2N\times 2N}^{T} \times \begin{bmatrix} D_{N\times N}\times b_{N\times N}\times D_{N\times N}^{T} & 0_{N\times N} \\ 0_{N\times N} & 0_{N\times N} \end{bmatrix} \times D_{2N\times 2N} \qquad (5) \\ &= (D_{N\times N}^{T}\times D_{N\times 2N})^{T}\times b_{N\times N}\times (D_{N\times N}^{T}\times D_{N\times 2N}) \\ &= V_{2N\times N}\times b_{N\times N}\times H_{N\times 2N} \end{aligned}$$

where $D_{N\times N}$ denotes the DCT kernel for N samples, $D_{N\times 2N}$ represents the upper most N rows of $D_{2N\times 2N}$, $V_{2N\times N}$ and $H_{N\times 2N}$ indicate the vertical and horizontal super-resolution kernels, respectively. And the transpose relationship between their kernels reflects $V_{2N\times N}^{T} = H_{N\times 2N}$. The vertical super-resolution kernel is defined as:

$$V_{2N\times N}(m,n) = \sum_{k=0}^{N-1} \alpha_k \cdot \cos\frac{\pi k(2m+1)}{4N} \cos\frac{\pi k(2n+1)}{2N}$$

where

$$0 \le \mathrm{m} \le 2\mathrm{N}-1, 0 \le \mathrm{n} \le \mathrm{N}-1. \qquad (6)$$

$$\alpha_k = \begin{cases} 1/\mathrm{N}, & \mathrm{k}=0 \\ 2/\mathrm{N}, & 1 \le \mathrm{k} \le \mathrm{N}-1 \end{cases}$$

Therefore, the super-resolution process (2) in DCT domain can be further interpreted as the corresponding up-sampling in spatial domain, as shown in (5). Then the up-sampling can be implemented separately by multiplying the vertical kernel followed by multiplying the horizontal kernel. In the following, $V_{2N\times N}$ and $H_{N\times 2N}$ are denoted as V and H, respectively, for simplicity. The Frobenius norm of the matrix A, with a_{ij} as its component, is employed as the objective function, which is defined according to:

$$\|A\|_F \overset{\text{def}}{=} \left(\sum_{i=0}^{m-1}\sum_{j=0}^{m-1}|a_{ij}|^2\right)^{1/2}. \qquad (7)$$

Then the optimized low-resolution sub-block $\hat{b}$ from the full-resolution macroblock B can be obtained by the minimization problem:

$$\hat{b} = \arg\min_{b}\{\|B - VbH\|_F\}. \qquad (8)$$

And the Kronecker product is employed to solve the problem. (8) can be further expressed as:

$$\hat{b}_v = \arg\min_{b_v}\{\|(V\otimes H^T)b_v - B_v\|_2^2\}, \qquad (9)$$

where b_v, $\hat{b}_v$, and B_v are vectors obtained from the corresponding matrixes, and $\otimes$ is the Kronecker product between two matrixes. Then optimized $\hat{b}_v$ can be obtained according to:

$$\hat{b}_v = (M^T M)^{-1} M^T B_v, \qquad (10)$$

where $M = (V\otimes H^T)$. Then after inverse vectorization, the optimized low-resolution sub-block $\hat{b}$ can be obtained. The sub-block $\hat{b}$ will be processed by transformation, quantization. Finally the RDO process [14] will determine whether the macroblock is coded by the traditional JPEG mode or the SRDDS mode. Therefore, 1-bit flag for each macroblock is encoded and transmitted to indicate which mode is employed.

III. Experimental Results

In order to demonstrate the coding efficiency of the proposed scheme, four typical gray images are employed for experiments: Lena (512×512), Goldhill (512×512), Foreman (352×288), and Kodim23 (768×512) [12]. The baseline JPEG coding method and the down-sampling based image coding scheme [2] are compared with the proposed method.

The images are coded by different coding schemes, with the bit rates ranging from 0.1 bpp to 0.7 bpp. The objective quality of the coded image is evaluated by the Peak Signal-to-Noise Ratio (PSNR). The higher the PSNR, the smaller the difference between the reconstructed image and the original one. Detailed information of PSNR comparisons is illustrated in Figure 2. From the results, it can be observed that the PSNR of the image inferred from our method is significantly higher than the baseline JPEG coded image and the image coded by [2], especially at the low bit-rates. Furthermore, the method by [2] degrades the performance at high bit-rates. The reason is that it employs three down-sampling methods and four QP settings, which results in too many overhead bits to be encoded and transmitted.

Furthermore, in order to demonstrate the perceptual gain of our proposed method, in Figure 3 we have illustrated some images decoded from the baseline JPEG method, and the proposed method, respectively. It can be observed that the

bpp=0.1315, PSNR=26.94dB, SSIM=0.7551 | bpp=0.2285, PSNR=27.78dB, SSIM=0.6564 | bpp=0.2122, PSNR=29.3139dB, SSIM=0.7901

bpp=0.1308, PSNR=32.09dB, SSIM=0.8798 | bpp=0.2264, PSNR=29.02dB, SSIM=0.7498 | bpp=0.2115, PSNR=31.33dB, SSIM=0.8497

Figure 3. Subjective quality comparison. Baseline JPEG images (top) and images by the proposed scheme (bottom).

baseline JPEG decodes images with severe blocking artifacts, which greatly degrades the visual quality. However, the proposed method reconstructs images with better visual quality. In order to further evaluate the image quality, SSIM [13], which is believed to be more consistent to the HVS perception than PSNR, is employed to evaluate the perceptual quality of each reconstructed image. According to its definition, the larger the SSIM value, the better the visual quality of the image. As illustrated from the experimental results in Figure 3, our proposed method generates better visual quality images with higher SSIM values.

IV. CONCLUSION

In this paper, a novel perceptual image coding scheme via adaptive block-based super-resolution directed down-sampling is proposed. The down-sampling method in the encoder is directed by the super-resolution method, which ensures the minimal reconstruction error. As to the decoder, the super-resolution method is implemented in the DCT domain, which can be integrated with the inverse DCT transform process. Therefore, it can significantly reduce the computational complexity. The experimental results have demonstrated that our methods can improve the decoded image quality in terms of both objective and subjective measurements.

REFERENCES

[1] X. Wu, X. Zhang, and X. Wang, "Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion", *IEEE Trans. Image Process.* vol. 18, no. 3, pp. 552-561, Mar. 2009.

[2] W. Lin, and L. Dong, "Adaptive downsampling to improve image compression at low bit rates", *IEEE trans. Image Process.* vol. 15, no. 9, pp. 2513-2521, Sep. 2006.

[3] X. Zhang, X. Wu, and F. Wu, "Image coding on quincunx lattice with adaptive lifting and iterpolation", *DCC* 2007, pp. 193-202, Mar. 2007.

[4] X. Zhang, and X. Wu, "Can lower resolution be better?", *DCC* 2007, pp. 302-311, Aug. 2008.

[5] B. Zeng, and A. N. Venetsanopoulos, "A JPEG-based interpolative image coding scheme", *Proc. ICASSP*, pp. 393-396, 1993.

[6] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down-scaling for better transform compression", *IEEE Trans. Image Process*. vol. 12, no. 9, pp. 1132-1144, Sep. 2003.

[7] Y. Tsaig, M. Elad, G. Golub, and P. Milanfar, "optimal framwork for low bit-rate block coders", *Proc. ICIP*, pp. 219-222, 2003.

[8] Y. Zhang, J. Zhang, R. Xiong, et al., "Low bit-rate image coding via interpolation oriented adaptive down-sampling", *Proc. SPIE*, vol. 7744.

[9] P. Longere, et. al., "Perceptual assessment of demosaicing algorithm performance", *Proc. IEEE*, vol. 90, no. 1, pp. 123-132, Jan. 2002.

[10] I. Shin, and H. W. Park, "Adaptive up-sampling method using DCT for spatial scalability of scalable video coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 206-214. Feb. 2009.

[11] 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. Letters.* vol. 17, no. 10, 827-830. Oct. 2010.

[12] Kodak Lossless True Color Image Suite. Available [online]: http://r0k.us/graphics/kodak/.

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

[14] T. Wiegand, et al. "Rate-Constrained Coder Control and Comparison of Video Coding Standards", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688-703. Jul. 2003.

[15] D. Taubman, and M. Marcellin, JPEG2000: Image Compression Fundamentals, Standards and Practice. Springer, 2001.

[16] T. Wiegand, et. al, "Overview of the H. 264/AVC Video Coding Standard", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576. Jul. 2003