Adaptive DCT-Domain Down-Sampling and Learning Based Mapping for Low Bit-Rate Image Compression

Zhongbo Shi[1,*], Xiaoyan Sun[2], and Feng Wu[2]

[1] University of Science and Technology of China, Hefei 230027, China
[2] Microsoft Research Asia, Beijing 100190, China
stoneshi@mail.ustc.edu.cn, xysun@microsoft.com,
fengwu@microsoft.com

Abstract. This paper proposes a new image compression approach. The high frequency information in an image is adaptively decreased to facilitate compression, which is then compensated through learning-based mapping. Specifically, an adaptive DCT-domain down-sampling method is presented to reduce the high frequencies in images and the visual pattern based learning is introduced to recover the removed high frequencies. The collaboration of the adaptive DCT-domain down-sampling and learning-based mapping presents a new coding mode for image compression (namely D-mode) which can be readily integrated into the mainstream coding schemes. Experimental results demonstrate that our proposed coding scheme achieves better coding performance in terms of both objective and subjective quality in comparison with H.264 intra coding and JPEG 2000, especially at low bit rates.

Keywords: Image compression, low bit-rate coding, down-sampling, mode decision.

1 Introduction

It is well known that the conventional block-based image compression methods (e.g. JPEG and JPEG2000) usually suffer from several kinds of artifacts, such as blocking and ringing, at low bit rates. It brings difficulties to the applications of low bit-rate images, especially to mobile multimedia applications. How to improve the performance of low bit-rate image compression is still a problem to be solved.

Researches in [1]-[3] have shown that down-sampling (DS) before encoding and up-sampling after decoding can improve the quality of coded image at low bit rates. Furthermore, compression via an adaptive DS method is proposed in [4], in which the DS ratio, DS direction, and quantization parameter (QP) are adjusted according to the local variance of images. It improves the visual quality of the reconstructed images in comparison with JPEG at low bit rates.

---

* This work was done during Zhongbo Shi's internship in Microsoft Research Asia.

P. Muneesawang et al. (Eds.): PCM 2009, LNCS 5879, pp. 222–231, 2009.

Instead of focusing on the DS process, proposals [5][6] introduce learning-based approaches to compensate the high frequencies lost in DS. Reported results show that these schemes outperform the current standard image compression schemes (e.g. JPEG and JPEG 2000) at low bit rates. However, the uniform DS method adopted in these works may drop the coding performance in certain regions that are not suitable for DS as well as learning-based mapping. Furthermore, the performance of DS in the spatial domain is usually not as good as that in the frequency domain [7].

In this paper, we present an image coding scheme by adaptive DCT-domain DS and learning-based mapping. The key contribution of our work is twofold. First, the DCT-domain DS is adopted to reduce the high frequency information in images during which the DS regions are adaptively selected in a rate-distortion (R-D) optimal way. Second, the visual pattern based learning is used to explore the distortion of DCT-domain DS for recovering the reduced high frequencies. The collaboration of the adaptive DCT-domain DS and learning-based mapping presents a new coding mode for image compression, namely D-mode. Experimental results demonstrate that our compression scheme integrated with the D-mode outperforms the H.264 intra coding and JPEG2000 at low bit rates in terms of both objective and subjective qualities.

The rest of this paper is organized as follows. Section 2 provides an overview of our compression framework. The D-mode is introduced in detail in Section 3, including the DCT-domain DS, learning-based mapping and the R-D selection. Section 4 presents the experimental results. Conclusions are drawn in Section 5.

2 Overview of Our Compression Scheme

Fig. 1 illustrates the framework of our compression scheme. Given an input image *I*, it is first divided into non-overlapped blocks $\{b_o^i\}$. For each original block b_o, it is coded by either conventional coding modes or our proposed D-mode (denoted by the dashed box in Fig.1). A mode decision module is adopted here to select the optimal compression mode between our D-mode and the other conventional compression modes such as the intra modes of H.264.

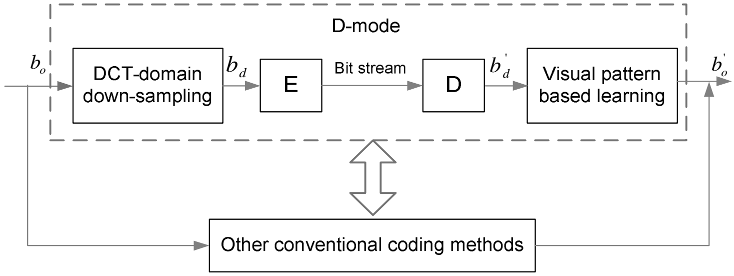


Fig. 1. Illustration of our proposed image compression scheme

In our proposed D-mode, each input block b_o is first down-sampled by the DCT-domain DS module by a factor of $1/s$. The resulting block b_d is then coded normally by the encoding module (E module in Fig.1). The E module here usually consists of intra prediction, transform, quantization and entropy coding. The corresponding process for decoding is denoted as D module in Fig. 1. During decoding, the decompressed block b'_d is generated by the decoding module (D module in Fig.1). Then the final reconstructed block b'_o is deduced from b'_d by visual pattern based learning.

3 D-Mode

The basic idea of our proposed D-mode coding is to decrease the high frequencies of images by a DCT-domain DS while compensated them by learning-based mapping. In this section, we will introduce the D-mode coding in detail. There are three main components involved, the DCT-domain DS, learning-based mapping and the corresponding R-D optimal mode decision.

3.1 DCT-Domain Down-Sampling

In our D-mode DS, we propose to reduce the high frequencies by dropping the high frequency coefficients at sub-block level instead of at block level. Fig.2 exhibits the DCT-domain DS method used in our D-mode, which is similar to that proposed in [7].

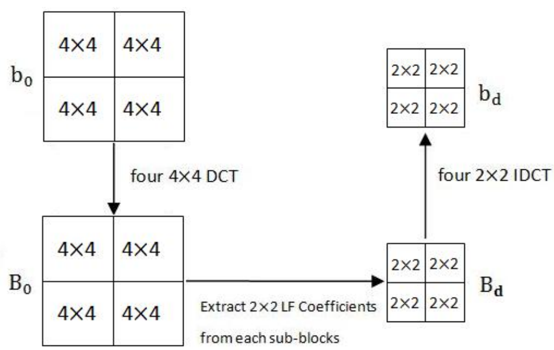


Fig. 2. DCT-domain DS used in our scheme

Taking an 8×8 image block b_o as example, its down-sampled version b_d is generated as follows.

1) Divide b_o into four 4×4 sub-blocks;
2) Perform 4×4 DCT transform on each 4×4 sub-block, resulting B_o;
3) Extract the 2×2 low-frequency (LF) coefficients from each sub-block, resulting B_d;
4) Perform 2×2 IDCT transform on each 2×2 DCT coefficients of B_d, resulting b_d.

The corresponding up-sampling method used in our scheme is simple, in which all the removed high frequency coefficients are filled by zero.

3.2 Visual Pattern Based Learning and Mapping

Inspired by [5], we propose to recover the removed high frequencies by visual pattern based learning and mapping. Similar to [6], visual pattern in our scheme not only describes the intensity variation of image blocks but also implicates geometric edge information. A visual pattern pair includes two corresponding visual pattern patches - a blurred visual pattern patch P_{b} and the corresponding high-quality visual pattern patch P_{h}. The correlation between them is utilized to reduce the distortion caused by DS. Here, patches refer to overlapped N×N blocks.

As shown in Fig. 3, a database consisting of visual pattern pairs is introduced in the D-mode for the recovery of high frequencies. Given a training image I_{o}, it is first down-sampled and then up-sampled in the DCT domain to form I_o'. Afterwards, $P_{\mathrm{b}}^{\mathrm{j}}$ is extracted from I_{h}, which is a high-pass filtered version of I_o' , at each edge pixel. Meanwhile, the corresponding $P_{\mathrm{h}}^{\mathrm{j}}$ is extracted at the same edge position from the differential image between I_{o} and I_o'.

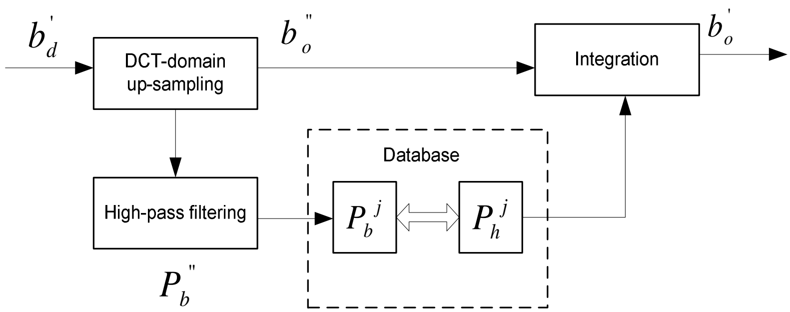


Fig. 3. Recovery of the high frequency information by visual pattern based mapping

Fig.3 shows the detailed recovery process based on the learned database. The decoded down-sampled block b_d' is first up-sampled to b_o'' in the DCT domain, in which all the removed high frequency coefficients are replaced by zero. After high-pass filtering, the blurred patch P_{b}'' is extracted at each edge pixel of b_o''. From the database, the most similar visual pattern patch $P_{\mathrm{b}}^{\mathrm{k}}$ to P_{b}'' in terms of Euclidean distance can be found by an approximate nearest neighbor search. The corresponding high-quality visual pattern patch $P_{\mathrm{h}}^{\mathrm{k}}$ is then used to enhance the quality of b_o'' in the edge regions and finally results in the reconstructed block b_o'.

3.3 R-D Optimization

In the D-mode, the down-sampled block B_d rather than B_o is coded, as shown in Fig. 2. At the same bit rate, the coding performance of B_d outperforms that of B_o if

$$d_{ds}(B_o, QP_d) < d(B_o, QP_o) \quad (1)$$

where $d_{ds}(B_o, QP_d)$ is the overall distortion of coding B_o via D-mode DS, $d(B_o, QP_o)$ is the distortion of direct coding B_o, and QP_d and QP_o are the quantization parameters used in the D-mode and the direct coding, respectively.

Fig.4 illustrates the relationship between $d(B_o, QP_o)$ and $d_{ds}(B_o, QP_d)$. Regarding B_o as an integration of B_d and $\tilde{B}_d$, where $\tilde{B}_d$ represents the removed high frequencies during DS, we have

$$d(B_o, QP_o) = d(B_d, QP_o) + d(\tilde{B}_d, QP_o).$$
$$d_{ds}(B_o, QP_d) = d(B_d, QP_d) + d_{\downarrow}(\tilde{B}_d). \quad (2)$$

where $d_{\downarrow}(\tilde{B}_d)$ denotes the distortion of removing $\tilde{B}_d$. Then (1) can be written as

$$d(B_d, QP_o) - d(B_d, QP_d) > d_{\downarrow}(\tilde{B}_d) - d(\tilde{B}_d, QP_o) \quad (3)$$

In our D-mode, since some high-frequency coefficients have been removed during DS, QP_d is normally smaller than QP_o at the same bit rate. In this case, we have $d(B_d, QP_o) - d(B_d, QP_d) \geq 0$. Moreover, $d(\tilde{B}_d, QP_o)$ approaches $d_{\downarrow}(\tilde{B}_d)$ with the increase of QP_o (the decrease of bit rate), which means that $d_{\downarrow}(\tilde{B}_d) - d(\tilde{B}_d, QP_o)$ tends to be 0 at low bit rates. Hence, (3) can be satisfied when QP_o increases (or bit rate decreases) to a certain level. On the other hand, (3) is hard to be true at high bit rates because of $d(B_o, QP_o) < d_{\downarrow}(\tilde{B}_d)$.

Accordingly, in our coding scheme, the D-mode is adaptively selected based on the joint rate-distortion optimization shown in (4).

$$J_{mode}(I_k | QP_o, \lambda) = D_{rec}(I_k | QP_o) + \lambda \times R_{rec}(I_k | QP_o) \quad (4)$$

where I_k is the kth candidate mode, $D_{rec}(I_k | QP_o)$ denotes the sum of the squared differences between the reconstructed and the original block pixels with a given QP_o, $R_{rec}(I_k | QP_o)$ denotes the rate with a given QP_o, and λ is set to $0.85 \times 2^{(QP_o - 12)/3}$ as suggested in H.264 [8].

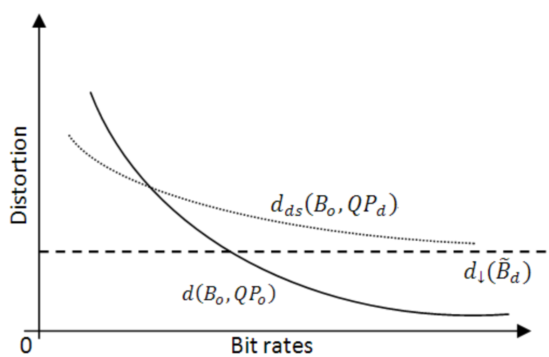


Fig. 4. Illustration of $d_{ds}(B_o, QP_d)$ and $d(B_o, QP_o)$

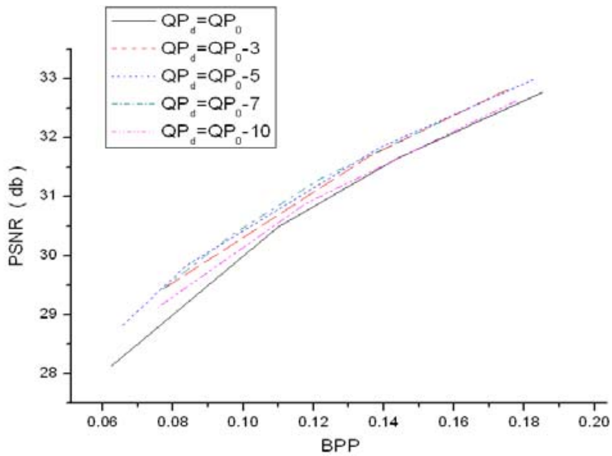


Fig. 5. Rate-distortion curve of different QP_d

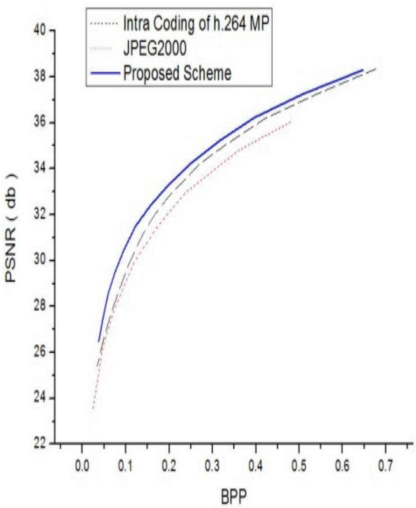
Furthermore, we would like to point out that we adjust the Lagrange parameter λ in (4) for the D-mode selection as the relationship between the rate and distortion differs from the traditional one due to the distortion introduced by the DS in the DCT domain. As mentioned before, at the same bit rate, $\mathrm{QP_d}$ should be smaller than $\mathrm{QP_o}$. Fig.5 shows some joint rate-distortion curves of setting $\mathrm{QP_d}$ between $\mathrm{QP_o}$ and $\mathrm{QP_o} - 10$. It can be observed that similar performance can be achieved during $\mathrm{QP_o} - 3$ and $\mathrm{QP_o} - 7$, which is better than that of using $\mathrm{QP_o}$ directly. In our current scheme, we set $\lambda = 0.85 \times 2^{(\mathrm{QP_d}-5)/3}$, where $\mathrm{QP_d} = \mathrm{QP_o} - 7$.

4 Experimental Results

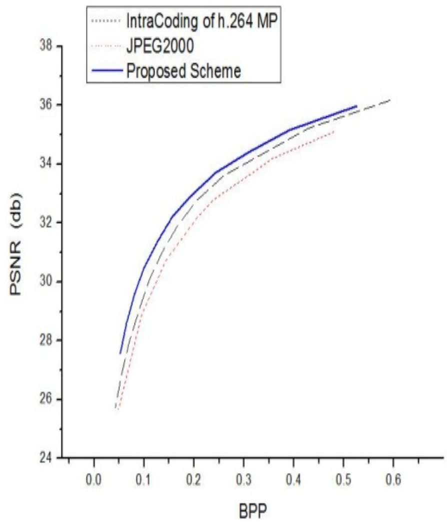
The performance of our proposed coding scheme is evaluated in comparison with that of H.264 intra coding [9] and JPEG 2000 [10]. In our coding scheme, the D-mode is incorporated into the intra coding of H.264 main profile (MP) and the DS ratio is set to 2:1 at orthogonal lattice. The training set in the test consists of two Kodak images,

Fig. 6. Training images

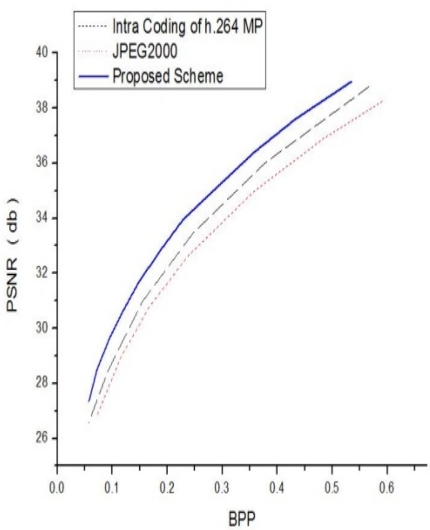
as shown in Fig.6. There are 2.4×10^5 visual pattern pairs extracted from the two images. In our experiments, the size of each visual pattern patch is 9 by 9. Four test gray images, Lena, Pepper, Lily and Kodim07, are involved in the tests.



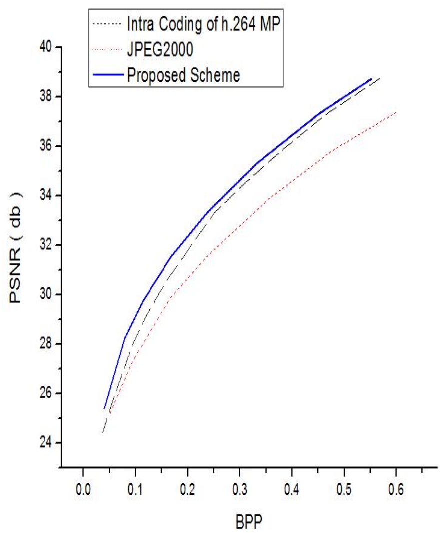
(a) *Lena* (512×512)



(b) *Pepper* (512×512)



(c) *Lily* (496×320)



(d) *Kodim07*(768×512)

Fig. 7. PSNR comparison of four test images

Fig.7 shows the PSNR results of the four test images. It can be observed that our proposed coding scheme always outperforms H.264 intra coding as well as JPEG2000. The improvement of our proposed algorithm over H.264 intra coding can be 1.0 dB on average and up to 1.5 dB gain can be achieved in comparison with JPEG2000.

Fig.8 and Fig.9 demonstrate the improvement of our scheme in terms of perceptual quality. Fig.8 shows the visual quality comparisons of test image Lena. One can notice that the obvious ringing and block artifacts in Fig.8 (a) and (b) have been effectively deduced in (c). Furthermore, the eyes, check and shoulder in (c) are much vivid and clean in our result. Similar improvement can be observed in Fig. (9), especially at the edge regions. Hence, the better perceptual quality is achieved by the collaboration of DCT-domain DS and the visual pattern based learning in our proposed D-mode.

Fig.10 shows the statistical result of our D-mode and intra modes of H.264 for Lena image coding. The percentage of our proposed D-mode increases as bit rate decreases and the D-mode becomes the major mode when bit rate decreases to a certain level. Clearly, the proposed D-mode is more efficient than the other intra modes at low bit rates.

(a) (b) (c)

Fig. 8. Lena at 0.12bpp. (a) JPEG2000 (29.93dB), (b) Intra coding of H.264 MP (30.43dB), (c) The proposed scheme (31.44dB).

(a) (b) (c)

Fig. 9. Pepper at 0.11bpp. (a) JPEG2000 (29.73 dB), (b) Intra coding of H.264 MP (30.02dB), (c) The proposed scheme (30.96 dB).

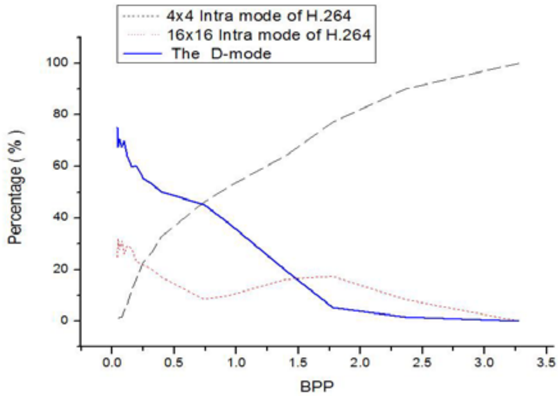


Fig. 10. The percentage of D-mode and other intra modes

5 Conclusion

This paper proposes an adaptive D-mode which integrates DCT-domain DS and visual pattern based learning to improve the coding performance of current mainstream image compression methods at low bit rates. The high frequencies removed by the adaptive DS in the DCT domain are recovered by visual pattern based learning. Experimental results demonstrate the advantages of the proposed scheme compared with the intra coding of H.264 and JPEG2000 at low bit rates.

Acknowledgment

The authors are grateful to Zhiwei Xiong for his valuable suggestions. The discussion among us is really helpful.

References

1. Zeng, B., Venetsanopoulos, A.N.: A JPEG-based interpolative image coding scheme. In: Proc. IEEE ICASSP, vol. V, pp. 393–396 (1993)
2. Bruckstein, A.M., Elad, M., Kimmel, R.: Down-scaling for better transform compression. IEEE Trans. Image Process. 12(9), 1132–1144 (2003)
3. Tsaig, Y., Elad, M., Golub, G., Milanfar, P.: Optimal framework for low bit-rate block coders. In: Proc. IEEE Int. Conf. Image Processing, 2003, vol. 2, pp. 219–222 (2003)
4. Lin, W., Dong, L.: Adaptive downsampling to improve image compression at low bit rates. IEEE Trans. Image Processing 15(9) (September 2006)
5. Li, Y., Sun, X., Xiong, H., Wu, F.: Incorporating primal sketch based learning into low bit-rate image compression. In: Proc. IEEE Int. Conf. on Image Processing, 2007, vol. 3, pp. 173–176 (2007)

6. Wu, F., Sun, X.: Image compression by visual pattern vector quantization (VPVQ). In: Data Compression Conference (2008)
7. Dugad, R., Ahuja, N.: A fast scheme for image size change in the compressed domain. IEEE Trans. Circuits and Systems for Video Technology 11, 461–474 (2001)
8. Wiegand, T., Schwarz, H., Joch, A., Kossentini, F., Sullivan, G.J.: Rate-constrained coder control and comparison of video coding standards. IEEE Trans. Circuits and Systems for Video Technology 13, 657–673 (2003)
9. `http://iphome.hhi.de/suehring/tml/download/`
10. `http://www.ece.uvic.ca/~mdadams/jasper/`