AN EFFECTIVE POSTPROCESSING METHOD FOR LOW BIT RATE BLOCK DCT CODED IMAGES

Deqing Sun and Wai-Kuen Cham

Dept. of EE, the Chinese University of Hong Kong
{dqsun, wkcham}@ee.cuhk.edu.hk

ABSTRACT

Transform coding using the Discrete Cosine Transform (DCT) has been widely used in image and video coding standards, but at low bit rates, the coded images suffer from severe visual distortions which prevent further bit reduction. Postprocessing can reduce these distortions and alleviate the conflict between bit rate reduction and quality preservation. Viewing postprocessing as an inverse problem, we solve it via a Bayesian approach. The distortion caused by coding is modeled as additive, spatially correlated Gaussian noise, while the original image is modeled as a high order Markov random field (MRF) based on the recently proposed Fields of Experts (FoE) framework. Experimental results show that the proposed method, in most cases, achieves higher PSNR gain than other methods and the processed images possess good visual quality.

***Index Terms*—** Postprocessing, Discrete Cosine Transform (DCT), quantization noise, Markov random field (MRF), Fields of Experts (FoE)

1. INTRODUCTION

Image compression aims at reducing the number of bits needed to represent a digital image while preserving image quality. When the compression ratio is very high, the coded images suffer from severe loss in visual quality, as well as decrease in fidelity. Hence there is conflict between bit rate reduction and quality preservation. Postprocessing is a promising solution to this problem because it can improve image quality without the need of changing the encoder structure. Different coding methods require different postprocessing techniques to tackle the different artifacts.

Transform coding using the Discrete Cosine Transform (DCT) has been widely used in image and video coding standards, such as JPEG, MPEG, H.264 etc. The coded images suffer from blocking artifacts and losses around edges. Postprocessing of low bit rate block DCT coded images has attracted a lot of research attention since early 1980s [1, 2, 3, 4]. Yet we have the following observations for most methods in the literature. First, the distortion caused by coding is not accurately modeled until the recent work [5] [6]. Second, the prior model for the original image is usually designed heuristically and captures only coarsely the rich structural information in natural images.

In this paper, postprocessing is formulated as an inverse problem and solved via a Bayesian approach. We use a spatially correlated Gaussian noise model [5] [6] to describe the coding error. The original image is modeled as a high order Markov random field (MRF) based on the Fields of Experts (FoE) framework [7]. The image prior model is more expressive than previously hand crafted models. As a result, we obtain an effective method which, in most cases, achieves higher PSNR gain than other methods and the processed images possess good visual quality.

In Section 2, we first introduce transform coding using the DCT and formulate postprocessing as an inverse problem. Then we explain how to solve it by a Bayesian approach and discuss the noise model and image model in Section 3. Experimental results and comparison with other methods are given in Section 4. Finally we draw conclusions in Section 5.

2. PROBLEM FORMULATION

Transform coding using the DCT first divides an image into non-overlapping blocks, which are 8×8 in case of JPEG. Each block is transformed into the DCT coefficients which are then quantized according to a quantization table and coded losslessly. Quantization is performed on each block independently and the levels and characteristics of the quantization errors may differ from one block to another. As a result, the blocking artifacts arise as abrupt changes across block boundaries and are especially obvious in smooth regions. In addition, edges become blurred and may even contain ringing effects due to the truncation of high frequency DCT coefficients.

The problem of postprocessing can be formulated as this: given the coded image I_q and the quantization table Q, we are to estimate an image $\hat{I}$, using the prior information about both the original image I and the coding process. $\hat{I}$ is expected to be closer to I and of better visual quality than I_q.

3. PROPOSED METHOD

Given a coded image I_q, we hope to obtain a restored image $\hat{I}$ that is most likely the original image I, which corresponds to the use of *maximum a posteriori* (MAP) criterion to estimate the original image

$$\hat{I} = \arg\max_{I} p_{I|I_q}(I|I_q) = \arg\max_{I} p_{I_q|I}(I_q|I)p_I(I). \quad (1)$$

In this expression, $p_{I_q|I}(I_q|I)$ provides a mechanism to incorporate the coded image into the estimation procedure, as it statistically describes the process to obtain I_q from I. Similarly, $p_I(I)$ allows for the integration of prior information about the original image. We shall discuss these two terms in Section 3.1 and then introduce the optimization method in Section 3.2.

3.1. Models

3.1.1. Quantization noise model

The distortion caused by coding can be modeled as adding quantization noise N_q to the original image I

$$I_q = I + N_q. \quad (2)$$

Strictly speaking, once the quantization table Q is given, the coded image I_q is uniquely determined by the original image I and N_q is a deterministic function of I. However, when only I_q is present, explicit information about N_q is lost and common practice is to treat N_q as a random quantity [8].

We use a correlated Gaussian noise model [5] [6] to describe the quantization noise, which makes the following assumptions. First, the quantization noise is assumed to be independent with the original image. Second, the quantization noises for different blocks are assumed to be independent because quantization is performed on each block independently. Third, the quantization noise is assumed to be independent in the DCT domain since quantization is performed independently on the DCT coefficients which are supposed to be uncorrelated [9]. The quantization noise n_q for an 8×8 block is arranged lexicographically into a column vector of length 64 and assumed to be zero mean, jointly Gaussian distributed in the spatial domain

$$n_q \sim \mathcal{N}(0, \Sigma_q), \quad (3)$$

where Σ_q is the noise autocovariance matrix in the spatial domain. It is a 64×64 invertible matrix but not a diagonal matrix due to the correlation of the noise. To calculate Σ_q, we first compute the noise autocovariance matrix in the DCT domain, which is denoted by Σ_{qc}. Σ_{qc} is a diagonal matrix because of the third assumption. In our work, we experimentally set its diagonal elements to be one twelfth of the square of the corresponding quantization intervals as in [5] and [6]. Σ_q is related to Σ_{qc} by the linear DCT transform and can be calculated accordingly.

Now the conditional p.d.f. $p_{I_q|I}(I_q|I)$ is

$$p_{I_q|I}(I_q|I) = \prod_m \frac{1}{(2\pi)^{32}|\Sigma_q|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}n_q^t(m)\Sigma_q^{-1}n_q(m)\right\}, \quad (4)$$

where $n_q(m)$ is the mth block of the noise $N_q = I_q - I$ and is arranged into a column vector of length 64.

3.1.2. Image prior model

An image I can be considered as a 2D function defined on a rectangular grid whose sites are pixels of the image. Let k be an arbitrary pixel in the image and $\mathcal{N}_k$ be a set which contains all the neighboring pixels of k. Markov random field (MRF) assumes the value of a pixel is conditionally dependent only on the values of its neighboring pixels, i.e.

$$p_{I_k|I_{\mathcal{S}-k}}(I_k|I_{\mathcal{S}-k}) = p_{I_k|I_{\mathcal{N}_k}}(I_k|I_{\mathcal{N}_k}), \quad (5)$$

where the set $\mathcal{S}$ contains all the pixels of the image I, the set $\mathcal{S} - k$ contains all the pixels except k, $I_{\mathcal{S}-k}$ denotes values of the pixels in $\mathcal{S} - k$, and $I_{\mathcal{N}_k}$ denotes values of the pixels in $\mathcal{N}_k$.

Whilst MRF models local interactions in an image, it is hard to write the joint p.d.f. of an image from the local conditional p.d.f.. The Hammersley-Clifford theorem [10] establishes that an MRF is equivalent to a Gibbs random field (GRF) and the joint p.d.f. can be written as a Gibbs distribution

$$p_I(I) = \frac{1}{Z} \exp\left\{-\sum_{c \in \mathcal{C}} V_c(I)\right\}, \quad (6)$$

where c, called a clique, is a set whose elements are neighbors to each other, $\mathcal{C}$ is a set which contains all the possible cliques in the image, $V_c(I)$ is a clique potential function defined on the values of all the pixels in c, and Z is a normalization parameter.

Though widely used in image processing applications, MRF exhibits serious limitations because the clique potential function is usually hand crafted and the neighborhood sizes are small. Thus it characterizes natural images only coarsely. Sparse coding, on the other hand, models the complex structural information in natural images in terms of a set of linear filter responses [11]. However, it only focuses on small image patches rather than the whole image. Combining the ideas from sparse coding with MRF model, the Fields of Experts (FoE) [7] defines the local potential function of an MRF with learnt filters. This learnt prior model is very expressive and has obtained success in applications such as image denoising and inpainting.

The FoE uses the following form for the distribution

$$\begin{aligned} p_I(I) &= \frac{1}{Z} \exp\left\{\sum_{k \in \mathcal{S}} \sum_{i=1}^{N} \log \phi_i(\mathcal{J}_i^T I_{c_k}; \alpha_i)\right\} \\ &= \frac{1}{Z} \prod_{k \in \mathcal{S}} \prod_{i=1}^{N} \phi_i(\mathcal{J}_i^T I_{c_k}; \alpha_i), \end{aligned} \quad (7)$$

in which

$$\phi_i(\mathcal{J}_i^T I_{c_k}; \alpha_i) = [1 + \frac{1}{2}(\mathcal{J}_i^T I_{c_k})^2]^{-\alpha_i}, \tag{8}$$

where $\mathcal{J}_i$ is a filter of size $n \times n$, the clique c_k adopted by FoE includes the $n \times n$ pixels with k as their center, $\mathcal{J}_i^T I_{c_k}$ denotes the inner product between the filter and the local image patch, α_i is a parameter associated with $\mathcal{J}_i$, and N is the number of filters used. In our work, we used the twenty four 5×5 filters which have been learnt in [7].

3.2. The optimization problem

Maximizing the *a posteriori* p.d.f. in (1) is equivalent to minimizing its negative log function which will be called the energy function, and the estimated image is

$$\hat{I} = \arg\max_I \exp\{-E(I)\} = \arg\min_I E(I). \tag{9}$$

From (1), (4), and (7), the energy function is

$$\begin{aligned} E(I) &= E_i(I) + \lambda E_n(I) \\ &= -\sum_{k \in \mathcal{S}} \sum_{i=1}^{N} \log \phi_i(\mathcal{J}_i^T I_{c_k}; \alpha_i) \\ &\quad + \lambda \sum_m \frac{1}{2} n_q^t(m) \Sigma_q^{-1} n_q(m)), \end{aligned} \tag{10}$$

where $\lambda \geq 0$ is a regularization parameter. We adopt the conjugate gradient descent method to minimize the energy function. At each iteration, the step size is selected to correspond to the minimum along the search direction. The gradient of the energy function $E(I)$ in (10) is

$$\nabla E(I) = -\sum_{i=1}^{N} J_i^{-1} * \psi_i(J_i * I) + \lambda \nabla E_n(I), \tag{11}$$

where $*$ denotes the convolution operation, J_i^{-1} is obtained by mirroring J_i around its center pixel,

$$\psi_i(y) = \frac{\partial}{\partial y} \log \phi_i(y; \alpha_i), \tag{12}$$

and $\nabla E_n(I)$'s mth block, arranged lexicographically into a column vector of length 64, is

$$-\Sigma_q^{-1} n_q(m). \tag{13}$$

To increase fidelity, the quantization constraint and the range constraint are respectively imposed for the DCT coefficients and the pixel values during the iteration. It is our prior knowledge that the DCT coefficients must lie within the quantization intervals and the pixel values between the range [0-255]. If either of them is violated, the intermediate result is set to the nearest value satisfying the corresponding constraint. When the iteration stops, the narrow quantization constraint set (NQCS) [12] is used for further PSNR gain and the scaling factor was set to be 0.3 in our experiments.

4. EXPERIMENTAL RESULTS

4.1. Parameter setting

We investigated by experiments how the value of λ affects the PSNR performance. In general, $\lambda \in (2, 12)$ produces good results for most images. In our experiments, $\lambda = 6$ was used for it is near optimal for the image set and quantization tables used. Smaller λ results in a smoother image, because it gives less fidelity to the the coded image and the estimated image can be adjusted more freely.

4.2. Results and comparison with other methods

First, we compare the improvement using PSNR. Table 1 summarizes the PSNR results of different methods for the four natural images and the three quantization tables Q1, Q2, and Q3 in [16]. In most cases, the proposed method has the highest PSNR gain except "Barbara" for which Paek et al's method [14] is slightly better. For comparison of visual quality, we show in Fig.1 the coded "Lena" and the processed images generated by the three methods with the highest PSNR gain in Table 1. We found that Liew and Yan's method [16] and the proposed method provide the best visual quality improvement. However, the proposed method is iterative and we are seeking efficient implementation similar to [17].

5. CONCLUSIONS

We have proposed a postprocessing method via a Bayesian approach. The prior models are carefully selected to model accurately both the original image and the distortion caused by coding. Experimental results on standard images and comparison with state-of-the-art methods have demonstrated the effectiveness of the proposed method. In most cases, it achieves higher PSNR gain than other methods and generates recovered images of good visual quality.

6. REFERENCES

[1] Jae S. Lim and Howard C. Reeve, "Reduction of blocking effect in image coding," *Opt. Eng.*, vol. 23, pp. 34–37, 1984.

[2] A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 2, pp. 91–95, 1992.

[3] Zixiang Xiong, M.T. Orchard, and Ya-Qin Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 433–437, April 1997.

[4] T.P. O'Rourke and R.L. Stevenson, "Improved image decompression for reduced transform coding artifacts," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 490–499, Dec. 1995.

[5] M.A. Robertson and R.L. Stevenson, "DCT quantization noise in compressed images," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 27–38, Jan. 2005.

[6] Bahadir K. Gunturk, Yucel Altunbasak, and Russell Mersereau, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Trans. Image Proc.*, vol. 13, pp. 33– 43, 2004.

Table 1. PSNR results(dB)

	LENA			PEPPERS			BARBARA			BABOON		
Quantization table	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Coded Image	30.701	30.091	27.382	30.689	30.141	27.641	25.939	25.591	24.028	24.320	24.143	22.133
$\mathrm{WT_X}$[3]	31.215	30.758	28.315	31.335	30.922	28.698	25.226	25.070	24.100	24.240	24.125	22.476
$\mathrm{MPEG_4}$[13]	31.211	30.694	28.095	31.312	30.842	28.557	26.092	25.774	24.367	24.451	24.293	22.401
$\mathrm{POCS_P}$[14]	31.629	31.020	28.513	31.499	31.009	28.848	**26.689**	**26.321**	24.746	24.631	24.469	22.522
$\mathrm{POCS_Y}$[15]	31.313	30.739	28.292	31.232	30.747	28.567	26.400	26.052	24.453	24.545	24.387	22.415
$\mathrm{MAP_R}$[5]	31.592	31.128	28.642	31.841	31.378	29.131	26.125	25.860	24.478	24.504	24.429	22.573
$\mathrm{WT_L}$[16]	31.612	31.187	28.654	31.599	31.305	29.033	26.374	26.043	24.660	24.591	24.450	22.558
Our method	**31.963**	**31.435**	**28.806**	**32.049**	**31.610**	**29.358**	26.655	26.320	**24.869**	**24.774**	**24.623**	**22.618**

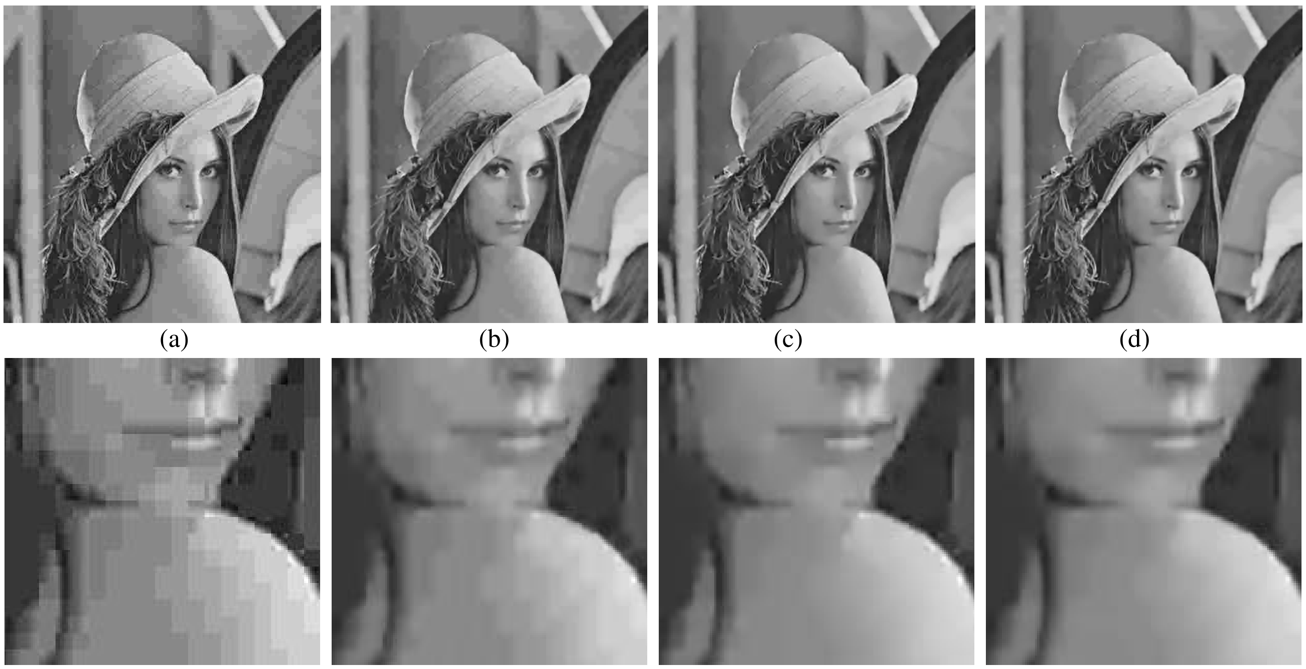

Fig. 1. Results for “Lena” (a) Coded image using Q2 (b) Robertson’s method [5] (c) Liew’s method [16] (d) The proposed method; Regions around the shoulder of “Lena”(e) Coded image using Q3 (f) Robertson’s method [5] (g) Liew’s method [16] (h) The proposed method

[7] S. Roth and M.J. Black, “Fields of experts: a framework for learning image priors,” in *IEEE Int. Conf. on Conf. Comp. Vision & Pattern Recognition*, 2005, vol. 2, pp. 860–867.

[8] B. Widrow, I. Kollar, and Ming-Chang Liu, “Statistical theory of quantization,” *IEEE Trans. on Instrumentation and Measurement*, vol. 45, pp. 353–361, 1996.

[9] O.G. Guleryuz, “Linear, worst-case estimators for denoising quantization noise in transform coded images,” *IEEE Trans. Image Proc.*, vol. 15, no. 10, pp. 2967–2986, Oct. 2006.

[10] J. Besag, “Spatial interaction and the statistical analysis of lattice systems (with discussion),” *J. Royal Stat. Soc., B*, vol. 36, pp. 192–236, 1973.

[11] M. Welling, G. Hinton, and S. Osindero., “Learning sparse topographic representations with products of student-t distributions,” in *Neural Information Processing Systems*, 2003, pp. 1359–1366.

[12] Seop Hyeong Park and Dong Sik Kim, “Theory of projection onto the narrow quantization constraint set and its application,” *IEEE Trans. Image Proc.*, vol. 8, pp. 1361–1373, 1999.

[13] *MPEG4 Verification Model, VM 18.0, 2001*, pp. 271 – 275.

[14] Hoon Paek, Rin-Chul Kim, and Sang-Uk Lee, “On the POCS-based postprocessing technique to reduce the blocking artifacts in transform coded images,” *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 3, no. 3, pp. 358–367, June 1998.

[15] Yongyi Yang, N.P Galatsanos, and A.K. Katsaggelos, “Projection-based spatially adaptive reconstruction of block-transform compressed images,” *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 4, no. 7, pp. 896–908, July 1995.

[16] A.W.C. Liew and H Yan, “Blocking artifacts suppression in block-coded images using overcomplete wavelet representation,” *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 14, no. 4, pp. 450–461, April 2004.

[17] M.A. Robertson and R.L. Stevenson, “Reduced-complexity iterative post-filtering of video,” *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 11, no. 10, pp. 1121–1127, Oct. 2001.