Postprocessing of Low Bit-Rate Block DCT Coded Images Based on a Fields of Experts Prior

Deqing Sun, *Student Member, IEEE*, and Wai-Kuen Cham, *Student Member, IEEE*

***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 propose to solve it by the *maximum a posteriori* criterion. 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 based on the fields of experts 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. In addition, we examine the noise model used and its parameter setting. The noise model assumes that the DCT coefficients and their quantization errors are independent. This assumption is no longer valid when the coefficients are truncated. We explain how this problem can be rectified using the current parameter setting.**

***Index Terms*—Discrete cosine transform (DCT), fields of experts (FoE), Markov random field (MRF), postprocessing, quantization noise.**

I. 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 DCT has been widely used in image and video coding standards, such as JPEG, MPEG, and H.263. 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.

Viewing the blocking artifacts as artificial high frequency components around block boundaries, Lim and Reeve [1] performed low pass filtering on the boundary pixels to reduce them. This method sometimes blurs true edges of the image and so adaptive filtering techniques were proposed to tackle this problem. Ramamurthi and Gersho [2] classified the blocks in the coded image and performed filtering parallel to the edges. The loop filtering [3] in H.264/AVC, the recent video coding standard, alternates several filters according to the local activity of the coded image. These filtering methods are from the enhancement angle and consider the artifacts as irregularities to be smoothed for visual improvement [4].

Viewing the problem as reducing noise with certain structure, some researchers adopted the wavelet thresholding technique. Xiong *et al.* [5] used thresholding by the overcomplete wavelet transform and they assumed the blocking artifacts mainly concentrated around the block boundaries. Liew and Yan [6] analyzed the block discontinuities caused by coding to derive more accurate thresholds at different wavelet scales. They also classified the blocks and performed edge detection to preserve textures and edges.

On the other hand, many researchers viewed the compression operation as a distortion process and proposed restoration techniques to recover the original image. For example, the projection onto convex sets (POCS) based methods [7]–[11] represent the prior information about the original image as convex sets and, by iterating projections onto these sets, they converge in the intersection of all the sets. Therefore, the final result is consistent with all the prior information we have about the original image. One commonly used convex set is the quantization constraint set (QCS) whose elements after quantization become the coded image. Park and Kim [12] narrowed down the QCS to form the narrow quantization constraint set (NQCS) which can result in recovered images of higher PSNR. Other constraint sets usually impose spatial domain smoothness on the recovered image. A novel smoothness constraint set has been proposed in the DCT domain using the Wiener filtering concept [13]. Some other smoothness constraint sets are designed for images of particular types, for example, graphic images [14] and images mainly with homogeneous regions [15].

The POCS-based methods are effective for suppressing blocking artifacts because it is easy to impose smoothness constraint around block boundaries. Losses around edges, however, have no fixed positions, and it is relatively complicated for the POCS-based methods to construct convex sets to reduce the artifacts around edges [16]. Fan and Cham [17], [18] proposed methods using an edge model to tackle losses around edges caused by wavelet coding. The methods can suppress ringing effects and also sharpen the blurred edges with low computation requirement.

Generally speaking, postprocessing, or restoration, is a typical inverse problem. The most general and simple theory

Manuscript received February 1, 2007; revised June 27, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Elad.

The authors are with the Chinese University of Hong Kong, Shatin N.T., Hong Kong (e-mail: dqsun@ee.cuhk.edu.hk; wkcham@ee.cuhk.edu.hk).

Digital Object Identifier 10.1109/TIP.2007.904969

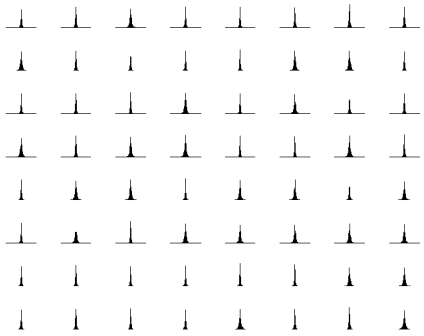


Fig. 1. Histogram of the quantization noise for each position of the 8×8 block in the spatial domain, obtained from over 80 000 obervations using the quantization table Q2 in Table II.

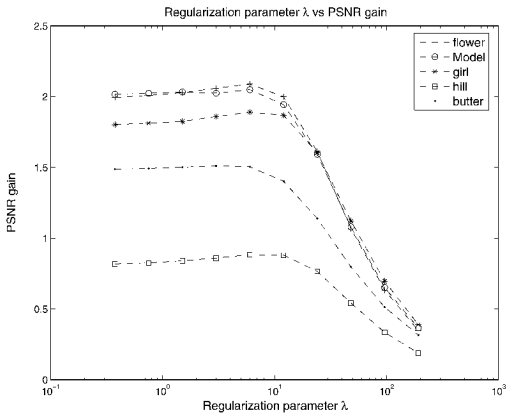


Fig. 2. Determination of λ for MAP estimation; $\lambda = 6$ is chosen for it produces near optimal result and slightly changing λ for a particular image may bring a slight PSNR gain. This curve is obtained using quantization table Q2, but the trend applies to Q1 and Q3.

for inverse problems is from the probabilistic point of view. From this angle, all prior information is represented in the form of *a priori* distributions. Thus, all the assumptions are made explicit and easy to examine [19]. O'Rourke and Stevenson [20] modeled the original image as a Huber Markov random field (MRF) and adjusted the coded image according to the model within the QCS. By doing so, they implicitly assumed the coded image was corrupted by uniform noise in the DCT domain, while Meier *et al.* [21] modeled the coding error as white Gaussian noise (WGN) in the spatial domain, but neither the uniform noise model nor the WGN model characterizes the coding error well. Robertson and Stevenson [22] found that a correlated Gaussian noise model in the spatial domain is more accurate and the use of this model can produce recovered images of higher PSNR. Gunturk *et al.* [23] independently used the same noise model in the superresolution reconstruction of compressed videos. For most of the methods described above, certain parameters are either chosen by users or empirically estimated from the data, Mateos *et al.* [24] proposed to estimate iteratively both the original image and the required parameters within the hierarchical Bayesian paradigm.

In this paper, postprocessing is treated as an inverse problem and solved using the *maximum a posteriori* (MAP) estimation. We use the noise model [22], [23] to describe the distortion caused by coding. The original image is modeled as a high order MRF based on the fields of experts (FoE) framework [25]. 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 generates images of good visual quality.

In Section II, we first formulate postprocessing as an inverse problem and explain how to solve it using the MAP criterion. We then describe the noise model and the image model separately. Experimental results are given in Section III, where we also examine the noise model used. Finally, we draw conclusions in Section IV.

II. Problem Formulation and the Proposed Method

Transform coding using the DCT first divides an image into nonoverlapping 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 both closer to I and of better visual quality than I_q. Here, I_q and I are assumed to be random vectors. This problem is ill-posed, since quantization is a many-to-one mapping. Then it is essential to model accurately both the original image and the coding process in conducting the estimation.

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 the *maximum a posteriori* (MAP) criterion to estimate the original image

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

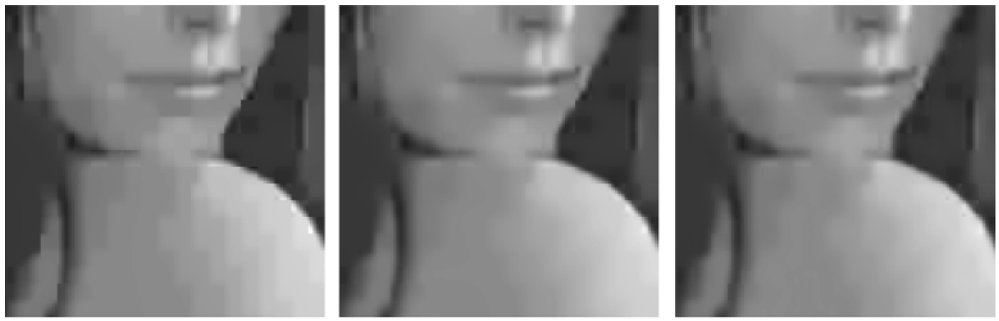
By Bayes's rule, (1) can be rewritten as

$$\begin{aligned} \hat{I} &= \arg\max_{I} \frac{p_{I_q|I}(I_q|I)p_I(I)}{p_{I_q}(I_q)} \\ &= \arg\max_{I} p_{I_q|I}(I_q|I)p_I(I). \end{aligned} \tag{2}$$

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 II-A and then introduce the optimization method in Section II-B.

A. Models and Assumptions

1) Quantization Noise Model: We assume there is no channel error and only quantization introduces distortions.



(a) (b) (c)

Fig. 3. Region around the shoulder of "Lena" coded by Q3, processed by the proposed method using FoE filters of different sizes. (a) 3×3, (b) 5×5, and (c) 7×7.

TABLE I
PSNR RESULTS (DECIBELS) FOR "LENA" CODED BY Q3, PROCESSED BY THE PROPOSED METHOD USING FOE FILTERS OF DIFFERENT SIZES

	LENA			PEPPERS			BARBARA			BABOON		
Quantization table	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Coded Image	30.702	30.091	27.382	30.689	30.141	27.641	25.939	25.591	24.028	24.320	24.143	22.133
3×3	31.619	31.125	28.496	31.762	31.321	28.982	26.454	26.134	24.683	24.648	24.491	22.546
5×5	**31.963**	**31.435**	**28.806**	**32.049**	**31.610**	**29.358**	**26.655**	**26.320**	**24.869**	**24.774**	**24.623**	**22.618**
7×7	31.921	31.386	28.757	32.016	31.554	29.281	26.597	26.254	24.823	24.763	24.605	22.610

Then the compression operation can be modeled as a distortion process that adds quantization noise N_q to the original image I

$$I_q = I + N_q \tag{3}$$

where the quantization noise N_q is assumed to be a random vector. Strictly speaking, once the quantization table Q is given, the coded image I_q is uniquely determined by the original image I and so N_q may be regarded as 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 [26]. Hence

$$p_{I_q|I}(I_q|I) = p_{I_q|I}(I + N_q|I) = p_{N_q|I}(N_q|I). \tag{4}$$

Note that the I in $p_{I_q|I}$ is given and so not a random quantity.

We need to understand the behavior of N_q. Empirically, it has uneven variances at different positions and the correlation is high among different positions within a block. Fig. 1 shows the histogram of the quantization noise at each position of the 8×8 block in the spatial domain. It appears to be centered distributed and so can be approximated by Gaussian distribution at each pixel position. As a result, we use a correlated Gaussian noise model [22], [23] to describe the quantization noise.

The following assumptions are made in [22] and [23]. First, the quantization noise N_q and the original image I are assumed to be independent. Hence, the conditional p.d.f. of the coded image I_q given I can be obtained from the p.d.f. of N_q

$$p_{I_q|I}(I_q|I) = p_{N_q|I}(N_q|I) = p_{N_q}(N_q). \tag{5}$$

Second, the quantization noises for different blocks are assumed to be independent because quantization is performed on each block independently. Then the p.d.f. of N_q can be expressed by the p.d.f. of the quantization noises for the individual blocks

$$p_{N_q}(N_q) = \prod_m p_{\mathbf{n}_q}(\mathbf{n}_q(m)) = \prod_m p_{\mathbf{n}_q}(\mathbf{i}_q(m) - \mathbf{i}(m)) \tag{6}$$

where m is a block index, and $\mathbf{n}_q(m)$, $\mathbf{i}_q(m)$, and $\mathbf{i}(m)$ are, respectively, the mth block of the quantization noise, the coded image, and the original image. Third, the quantization noise is assumed to be independent in the DCT domain. The assumption is because quantization is performed independently on the DCT coefficients which are supposed to be uncorrelated [27]. When the DCT domain noise variances $\sigma^2_{qc}(u, v)$ are known, the noise distribution is determined. Fourth, the noise for a block, arranged lexicographically into a column vector $\mathbf{n}_q(m)$ of length 64, is assumed to be zero mean, jointly Gaussian distributed in the spatial domain

$$\mathbf{n}_q(m) \sim \mathcal{N}(0, \Sigma_q) \tag{7}$$

where Σ_q is a 64×64 invertible matrix but not a diagonal matrix due to the correlation of the quantization noise in the spatial domain. It can be determined from the DCT domain noise variances $\sigma^2_{qc}(u, v)$, Setting of which will be discussed in Section III.

From (5)–(7), the conditional p.d.f. of the coded image I_q given the original image I is

$$p_{I_q|I}(I_q|I) = \prod_m \frac{1}{(2\pi)^{32}|\Sigma_q|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}\mathbf{n}_q^t(m)\Sigma_q^{-1}\mathbf{n}_q(m)\right\} \tag{8}$$

where $\mathbf{n}_q(m) = \mathbf{i}_q(m) - \mathbf{i}(m)$ and has been arranged lexicographically into a column vector of length 64.

Fig. 4. Four original images of size 512 by 512: "Lena" and "Peppers" mainly contain smooth regions and major edges, while "Barbara" and "Baboon" are full of textures. (a) "Lena," (b) "Peppers," (c) "Barbara," and (d) "Baboon."

TABLE II
QUANTIZATION TABLES; Q1, Q2, AND Q3 CORRESPOND TO 0.24, 0.189, AND 0.15 BPP COMPRESSION FOR "LENA," RESPECTIVELY

Q1	Q2	Q3
$\rightarrow v$	$\rightarrow v$	$\rightarrow v$
050 060 070 070 090 120 255 255	086 059 054 086 129 216 255 255	110 130 150 192 255 255 255 255
060 060 070 096 130 255 255 255	064 064 075 102 140 255 255 255	130 150 192 255 255 255 255 255
070 070 080 120 200 255 255 255	075 070 086 129 216 255 255 255	150 192 255 255 255 255 255 255
$\downarrow$ 070 096 120 145 255 255 255 255	$\downarrow$ 075 091 118 156 255 255 255 255	$\downarrow$ 192 255 255 255 255 255 255 255
u 090 130 200 255 255 255 255 255	u 097 118 199 255 255 255 255 255	u 255 255 255 255 255 255 255 255
120 255 255 255 255 255 255 255	129 189 255 255 255 255 255 255	255 255 255 255 255 255 255 255
255 255 255 255 255 255 255 255	255 255 255 255 255 255 255 255	255 255 255 255 255 255 255 255
255 255 255 255 255 255 255 255	255 255 255 255 255 255 255 255	255 255 255 255 255 255 255 255

TABLE III
PSNR RESULTS (DECIBEL) FOR IMAGES IN FIG. 4 USING QUANTIZATION TABLES Q1, Q2, AND Q3 IN TABLE II

	LENA			PEPPERS			BARBARA			BABOON		
Quantization table	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Coded Image	30.702	30.091	27.382	30.689	30.141	27.641	25.939	25.591	24.028	24.320	24.143	22.133
WT_X [5]	31.215	30.758	28.315	31.335	30.922	28.698	25.226	25.070	24.100	24.240	24.125	22.476
$MPEG_4$ [31]	31.211	30.694	28.095	31.312	30.842	28.557	26.092	25.774	24.367	24.451	24.293	22.401
$POCS_P$ [11]	31.629	31.020	28.513	31.499	31.009	28.848	**26.689**	**26.321**	24.746	24.631	24.469	22.522
$POCS_Y$ [10]	31.313	30.739	28.292	31.232	30.747	28.567	26.400	26.052	24.453	24.545	24.387	22.415
MAP_R [22]	31.592	31.128	28.642	31.841	31.378	29.131	26.125	25.860	24.478	24.504	24.429	22.573
WT_L [6]	31.612	31.187	28.654	31.599	31.305	29.033	26.374	26.043	24.660	24.591	24.450	22.558
The proposed 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**

2) Image Prior Model: An image I can be considered as a 2 D 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. The Markov random field (MRF) assumes the value of a pixel is conditionally dependent only on the values of its neighboring pixels, i.e.,

$$p_{I_k|I_{\mathcal{S}-k}}(I_k|I_{\mathcal{S}-k}) = p_{I_k|I_{\mathcal{N}_k}}(I_k|I_{\mathcal{N}_k}) \tag{9}$$

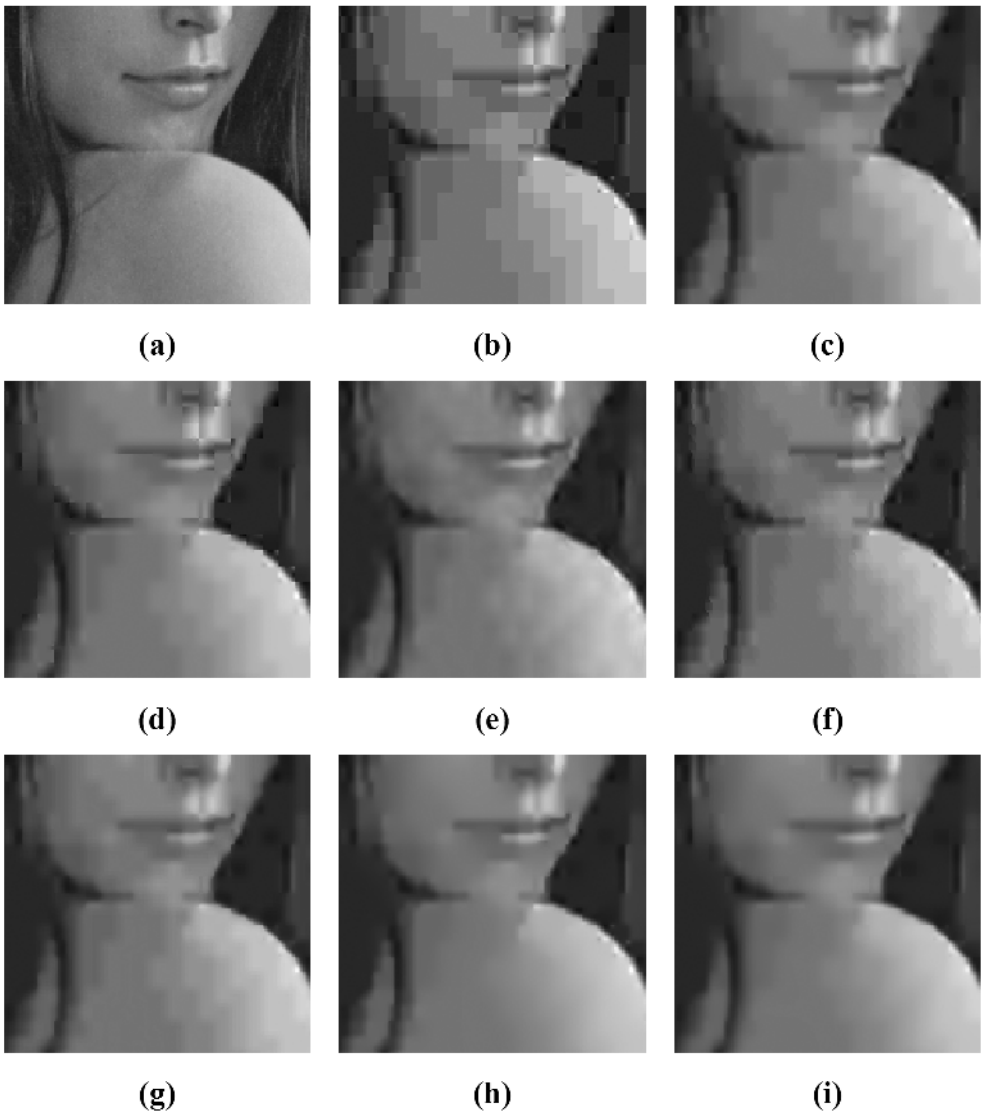


Fig. 5. Postprocessing results around the shoulder of "Lena", which tests the "deblocking" ability of the methods; (h) and (i) suppress the blocking artifacts effectively. (a) Original image, (b) coded image, (c) $\mathrm{WT_X}$ [5], (d) $\mathrm{MPEG_A}$ [31], (e) $\mathrm{POCS_P}$ [11], (f) $\mathrm{POCS_Y}$ [10], (g) $\mathrm{MAP_R}$ [22], (h) $\mathrm{WT_L}$ [6], (i) proposed method.

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 $\mathcal{N}_k$.

Whilse 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 [28] 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

$$p_I(I) = \frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} V_c(I)\right\} \tag{10}$$

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 functions are usually hand crafted and the neighborhood systems are small. Hence, 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 [29]. However, it only focuses on small image patches rather than the whole image. Combining the ideas from sparse coding with the MRF model, FoE [25] defines the local potential function of an MRF with learned filters. This learned prior model is very expressive and has obtained success in applications such as image denoising and image inpainting.

FoE uses the following form for the distribution:

$$\begin{aligned} p_I(I) &= \frac{1}{Z}\exp\left\{\sum_{k\in\mathcal{S}'}\sum_{i=1}^{N}\log\phi\left(\mathcal{J}_i^T I_{c_k};\alpha_i\right)\right\} \\ &= \frac{1}{Z}\prod_{k\in\mathcal{S}'}\prod_{i=1}^{N}\phi\left(\mathcal{J}_i^T I_{c_k};\alpha_i\right) \end{aligned} \tag{11}$$

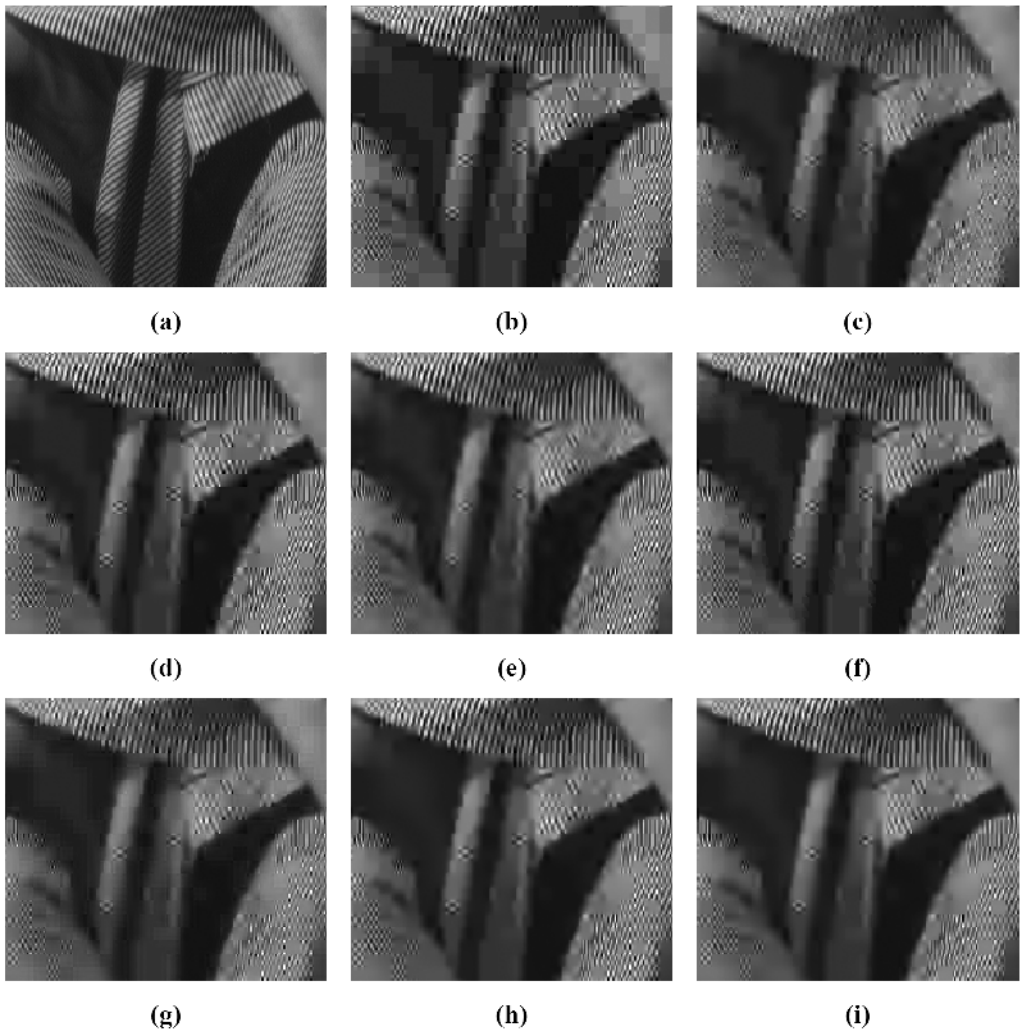


Fig. 6. Postprocessing results around the leg of "Barbara", which tests both the "deblocking" and the detail-preserving ability of the methods; (e), (h), and (i) preserve the details well while suppressing the blocking artifacts. (a) Original image, (b) coded image, (c) $\mathrm{WT_X}$ [5], (d) $\mathrm{MPEG_4}$ [31], (e) $\mathrm{POCS_P}$ [11], (f)$\mathrm{POCS_V}$ [10], (g) $\mathrm{MAP_R}$ [22], (h) $\mathrm{WT_L}$ [6], (i) proposed method.

where

$$\phi\left(\mathcal{J}_i^T I_{c_k};\alpha_i\right)=\left[1+\frac{1}{2}\left(\mathcal{J}_i^T I_{c_k}\right)^2\right]^{-\alpha_i} \tag{12}$$

$\mathcal{J}_i$ 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, $\mathcal{S}'$ contains the center pixels of all the $n_1\times n_1$ cliques that fully overlap with the support of the image, α_i is a parameter associated with $\mathcal{J}_i$, and N is the number of filters used. The performance of different-sized FoE filters will be examined in Section III-A3.

FoE builds the distribution of an image in terms of its responses to a set of filters. The product form in (11) implicitly assumes the responses to different filters are statistically independent, and to the same filter also independent at different pixel positions. Equation (12), if properly normalized, is student-t distribution which is featured by its heavy tails. It has been observed that, for a wide variety of filters, the response of an image has only a few large coefficients, with the left very small. Such statistics can be fitted well by the student-t distribution. Then the parameter α_i associated with each filter controls the width of the distribution of the filter response, and is positive to make the ϕ_i proper distributions.

B. Optimization Problem

Maximizing the objective function in (2) is equivalent to minimizing its negative log function which will be called the energy function $E(I)$, and the estimated image is

$$\hat{I}=\arg\max_I \exp\{-E(I)\}=\arg\min_I E(I). \tag{13}$$

From (2), (8), and (11), the energy function is

$$\begin{aligned}E(I)&=E_i(I)+\lambda E_n(I)\\&=-\sum_{k\in\mathcal{S}'}\sum_{i=1}^{N}\log\phi\left(\mathcal{J}_i^T I_{c_k};\alpha_i\right)\\&\quad+\lambda\sum_m\frac{1}{2}\mathbf{n}_q^t(m)\Sigma_q^{-1}\mathbf{n}_q(m)\end{aligned} \tag{14}$$

TABLE IV
QUANTIZATION NOISE VARIANCES IN THE DCT DOMAIN FOR "LENA" CODED USING Q2 IN TABLE II

$\frac{Q^2(u,v)}{12}$								Actual variances							
			$\rightarrow v$								$\rightarrow v$				
616.33	290.08	243.00	616.33	386.80	3888.0	5418.8	5418.8	633.61	229.30	130.43	166.28	161.44	95.824	44.193	25.538
341.33	341.33	468.75	867.00	1633.3	5418.8	5418.8	5418.8	206.49	161.89	153.41	157.09	107.67	60.753	33.899	19.217
468.75	408.33	616.33	1386.8	3888.0	5418.8	5418.8	5418.8	133.94	117.87	135.53	143.09	84.071	48.610	22.906	15.028
↓ 468.75	690.08	1160.3	2028.0	5418.8	5418.8	5418.8	5418.8	↓ 75.917	93.382	108.45	85.617	52.711	29.081	18.626	11.491
u 784.08	1160.3	3300.1	5418.8	5418.8	5418.8	5418.8	5418.8	u 42.536	45.917	44.543	43.304	28.821	18.405	11.806	9.7435
1386.8	2976.8	5418.8	5418.8	5418.8	5418.8	5418.8	5418.8	21.151	20.747	21.071	19.519	15.437	11.447	9.4009	8.2653
5418.8	5418.8	5418.8	5418.8	5418.8	5418.8	5418.8	5418.8	12.089	11.747	10.588	11.060	10.205	8.4466	7.0938	6.5371
5418.8	5418.8	5418.8	5418.8	5418.8	5418.8	5418.8	5418.8	9.1657	8.6195	7.6928	7.4576	7.4224	7.0370	6.1908	5.6512

TABLE V
PSNR RESULTS (DECIBELS FOR IMAGES IN FIG. 4 USING QUANTIZATION TABLES Q1, Q2, AND Q3 IN TABLE II

	LENA			PEPPERS			BARBARA			BABOON		
Quantization table	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Coded Image	30.702	30.091	27.382	30.689	30.141	27.641	25.939	25.591	24.028	24.320	24.143	22.133
Recovered image using $\sigma^2_{qc}(u,v) = \frac{Q^2(u,v)}{12}$	**31.963**	**31.435**	28.806	32.049	**31.610**	29.358	**26.655**	**26.320**	**24.869**	**24.774**	**24.623**	**22.618**
Recovered image using actual $\sigma^2_{qc}(u,v)$	31.839	31.340	28.759	31.928	31.524	29.292	26.607	26.283	24.861	24.719	24.574	22.606
Recovered image with truncation considered	31.961	31.409	**28.813**	**32.070**	31.608	**29.385**	26.617	26.281	24.868	24.736	**24.623**	22.610

where $\lambda \geq 0$ is a regularization parameter. It balances the constraints from the image model and the noise model. Smaller λ gives less fidelity to the coded image and generates smoother images. The setting of λ will be discussed in Section III.

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 (14) is

$$\nabla E(I) = -\sum_{i=1}^{N} J_i * \psi\left(J_i^{-1} * I; \alpha_i\right) + \lambda \nabla E_n(I) \qquad (15)$$

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

$$\psi(y;\alpha_i) = \frac{\partial}{\partial y}\log\phi(y;\alpha_i) == -\frac{\alpha_i y}{1+\frac{1}{2}y^2} \qquad (16)$$

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

$$-\Sigma_q^{-1}(\mathbf{i}_q(m) - \mathbf{i}(m)). \qquad (17)$$

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 original DCT coefficients must lie within the quantization intervals and the pixel values between 0 and 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 coefficients were set to be 0.3 in our experiments.

III. EXPERIMENTAL RESULTS

In this section, we first describe the parameter setting for the proposed method and then give the experimental results. We also examine the quantization noise model used and discuss some problems found.

A. Setting Algorithm Parameters

1) Noise Variances: In our experiments, the noise variances $\sigma^2_{qc}(u,v)$ were set as in [22], which are

$$\sigma^2_{qc}(u,v) = \frac{Q^2(u,v)}{12}. \qquad (18)$$

Robertson and Stevenson [22] chose $Q^2(u,v)/12$ because they assumed the quantization noise in the DCT domain is uniformly distributed within the corresponding quantization interval. We will discuss this setting in Section III-C in detail.

2) Regularization Parameter λ: We investigated by experiments how the value of λ affects the PSNR performance. Five 512×512 images, coded using quantization tables Q1, Q2, and Q3 in Table II, were processed by the proposed method with different λ. The results, as shown in Fig. 2, show that the PSNR varies little for λ less than 10 and then drops quickly for λ larger than 10. 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 this image set and the three quantization tables in Table II.

3) FoE Filter Size: We compared three groups of FoE filters of different sizes, including 3×3, 5×5, and 7×7.[1] The numbers of filters for the three groups are respectively 8, 24, and 48. These filters were obtained using a subset of the 200 training images of the Berkeley Segmentation Database [30]. In all the experiments, λ was fixed to be 6. Table I summarizes the PSNR results and Fig. 3 shows the processed regions around the shoulder of "Lena" coded by Q3. The 5×5 group has about 0.2-dB gain over the 3×3 group and also produces smoother images than the latter. In addition, it gives results similar to, or slightly better than, the 7×7 group. The FoE filters of larger size are expected to give better results. However, the 7×7 group does not produce better results than the 5×5 group in our experiments. This implies that filters of size 5×5 are sufficient to capture the complex structural information in natural images,

[1]The first two groups and their associated parameters are available at http://www.cs.brown.edu/~roth/research/foe/downloads.html, and the last group was provided by Dr. Stefan Roth at Brown University.

and a larger size helps little. In the following experiments, the 5×5 group was used for its good PSNR performance.

B. Results

We tested the proposed method on twenty three images of size 512×512. Detailed experimental results on four images shown in Fig. 4, using the three quantization tables in Table III, are given here. Blocking artifacts are more prominent in smooth regions. "Lena" and "Peppers" which contain large smooth regions are selected to examine the suppression of the blocking artifacts by the proposed method. On the other hand, a postprocessing method should not over-smooth details. Thus, "Barbara" and "Baboon" which have a lot of textures are selected to reveal the detail-preserving ability of the proposed method. The quantization tables Q1, Q2, and Q3 in Table II correspond to 0.24, 0.189, and 0.15 bits per pixel (bpp) compression for "Lena."

The proposed method is compared to a few popular postprocessing methods which include Xiong's wavelet-based method $(\mathrm{WT_X})$ [5], the MPEG4-VM postfiltering $(\mathrm{MPEG_4})$ [31], Paek's POCS-based method $(\mathrm{POCS_P})$ [11], Yang's POCS-based method $(\mathrm{POCS_Y})$ [10], Robertson's method $(\mathrm{MAP_R})$ [22], and Liew's wavelet-based method $(\mathrm{WT_L})$ [6]. Table III summarizes the PSNR results of these methods on the four images in Fig. 4 using the three quantization tables in Table III. In most cases, the proposed method has the highest PSNR gain except "Barbara" for which Paek's POCS-based method is slightly better. Based on the twenty three images tested, it achieves about 0.3 and 0.4 dB PSNR gain on average over Liew's wavelet-based method and Paek's POCS-based method, respectively.

For comparison of visual quality, we show in Figs. 5 and 6 the processed results around the shoulder of "Lena" and the leg of "Barbara" respectively. We found that Liew's wavelet-based method and the proposed method provide the best visual quality improvement. Both methods suppress blocking artifacts effectively while preserving the details well. However, the proposed method is computationally expensive, due to the use of iteration. At present, we are seeking efficient implementation, following the approach in [32].

C. Investigation on the Quantization Noise Model

In the experiments above, the noise variances for the proposed method were set to be one twelfth of the square of the corresponding quantization step sizes, as in [22] and [23]. To examine the correctness of this model, we estimated the actual noise variances using the original images and the coded images. For "Lena" coded using Q2, the actual noise variances, as shown in Table IV, do not deviate much from the predefined values for the low frequency coefficients. However, they are much smaller than the predefined values for the high frequency coefficients. We then found the optimal λ and performed the MAP estimation using the actual variances. Strangely, the images estimated using the actual variances have lower PSNR than those estimated using $\sigma^2_{qc}(u,v) = (Q^2(u,v))/(12)$, as shown in Table V.

We believe this apparently strange result is due to the independent quantization noise assumption made in (5), which is severely violated by the high frequency coefficients. Widrow *et al.* [26] have shown that, under certain conditions, the input signal of a uniform quantizer and the quantization error are uncorrelated, despite their deterministic relationship. If the standard deviation of the input signal is no less than the quantization step size, the conditions are approximately satisfied. Under such conditions, it is reasonable to assume the input signal and the quantization noise are independent, when only the quantized signal is available. Table VI shows the standard deviations of the DCT coefficients of "Lena." The standard deviations of the high frequency coefficients are much smaller than the corresponding quantization step sizes in Table II. Nearly all these high frequency coefficients are truncated during quantization, and the original coefficients and the quantization noise are of the same magnitude and opposite sign. As a result

TABLE VI
STANDARD DEVIATIONS OF THE ORIGINAL DCT COEFFICIENTS FOR "LENA" IN FIG. 4

	$\rightarrow v$							
	359.41	85.999	40.041	21.729	14.571	9.7927	6.6478	5.0535
	53.403	37.014	24.905	17.367	10.706	7.7944	5.8223	4.3838
	21.175	21.436	18.848	13.191	9.1690	6.9721	4.7860	3.8766
$\downarrow$	11.664	11.548	11.125	9.2660	7.2602	5.3927	4.3158	3.3898
u	6.8718	6.7918	6.6741	6.5806	5.3685	4.2901	3.4359	3.1215
	4.5990	4.5549	4.5904	4.4180	3.9290	3.3833	3.0661	2.8749
	3.4770	3.4274	3.2540	3.3256	3.1945	2.9063	2.6634	2.5568
	3.0275	2.9359	2.7736	2.7309	2.7244	2.6527	2.4881	2.3772

$$p_{n_{qc}(u,v)|C(u,v)}(n_{qc}(u,v)) \,|\, C(u,v)) \neq p_{n_{qc}(u,v)}(n_{qc}(u,v)). \quad (19)$$

Now we rewrite the term involving the noise model in (14) in the DCT domain as

$$E_n(I) = \sum_m \sum_{u,v=0}^{7} \frac{(C_q(m,u,v) - C(m,u,v))^2}{2\sigma^2_{qc}(u,v)} \quad (20)$$

where $C_q(m,u,v)$ and $C(m,u,v)$ denote, respectively, the (u,v)th DCT coefficients of the mth block of I_q and I. If $\sigma^2_{qc}(u,v)$ is set to be very big for large u and v, the corresponding term in (20) becomes insignificant and the influence of the inaccurate assumption is reduced. The strategy is to give the noise model less weights, when its assumption is severely violated. We should rely more, or solely, on the image model to estimate the truncated high frequency coefficients.

In another experiment, we used only the image prior model to estimate the truncated coefficients. In the implementation, the terms involving the image prior model in (14) and (15) were computed as before. $E_n(I)$ and $\nabla E_n(I)$ were calculated in the DCT domain according to (20). If the quantized coefficients were zero, we set the corresponding terms to be zero. As shown in Table V, the recovered images with the effect of coefficient truncation considered have comparable PSNR to those recovered with $\sigma^2_{qc}(u,v) = (Q^2(u,v))/(12)$ for all the coefficients.

IV. CONCLUSION

We have proposed a postprocessing method according to the MAP criterion. The prior models are carefully selected to model accurately both the original image and the distortions caused by coding. Experimental results on standard images and comparison with other methods have demonstrated the effectiveness of

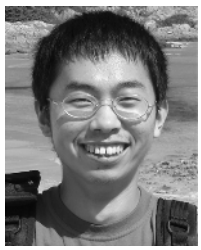
the proposed method. In most cases, it achieves higher PSNR gain than other methods and generates recovered images of good visual quality. We also examine the quantization noise model adopted by some state-of-the-art methods. We identify some problems in the noise model and explain why it still works with the current parameter setting.

ACKNOWLEDGMENT

The authors would like to thank all the anonymous reviewers for the constructive comments and useful suggestions that have led to the improvements in the quality, presentation, and organization of this paper.

REFERENCES

[1] J. S. Lim and H. C. Reeve, "Reduction of blocking effect in image coding," *Opt. Eng.*, vol. 23, pp. 34–37, 1984.
[2] B. Ramamurthi and A. Gersho, "Nonlinear space-variant postprocessing of block coded images," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1258–1268, Oct. 1986.
[3] P. List, A. Joch, J. Lainema, G. Bjntegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 614–619, Jul. 2003.
[4] M.-Y. Shen and C.-C. J. Kuo, "Review of postprocessing techniques for compression artifact removal," *J. Vis. Commun. Imag. Represent.*, vol. 9, no. 1, pp. 2–14, Mar. 1998.
[5] Z. Xiong, M. Orchard, and Y.-Q. Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 433–437, Apr. 1997.
[6] A. W. C. Liew and H. Yan, "Blocking artifacts suppression in block-coded images using overcomplete wavelet representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 4, pp. 450–461, Apr. 2004.
[7] R. E. Rosenholtz and A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," in *Proc. SPIE Image Processing Algorithms and Techniques II*, M. R. Civanlar, S. K. Mitra, and R. J. Moorhead, Eds., 1991, vol. 1452, pp. 116–126.
[8] A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, no. 1, pp. 91–95, Mar. 1992.
[9] Y. Yang, N. Galatsanos, and A. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 6, pp. 421–432, Dec. 1993.
[10] Y. Yang, N. Galatsanos, and A. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 896–908, Jul. 1995.
[11] H. Paek, R.-C. Kim, and S.-U. Lee, "On the POCS-based postprocessing technique to reduce the blocking artifacts in transform coded images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 3, pp. 358–367, Jun. 1998.
[12] S. H. Park and D. S. Kim, "Theory of projection onto the narrow quantization constraint set and its application," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1361–1373, Oct. 1999.
[13] X. Gan, A. W. C. Liew, and H. Yan, "A smoothness constraint set based on local statistics of BDCT coefficients for image postprocessing," *Image Vis. Comput.*, vol. 23, pp. 731–737, 2005.
[14] A. W. C. Liew, H. Yan, and N.-F. Law, "POCS-based blocking artifacts suppression using a smoothness constraint set with explicit region modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 4, pp. 795–800, Apr. 2005.
[15] C. Weerasinghe, A. W. C. Liew, and H. Yan, "Artifact reduction in compressed images based on region homogeneity constraints using the projection onto convex sets algorithm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 891–897, Oct. 2002.
[16] Y. Yang and N. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," *IEEE Trans. Image Process.*, vol. 6, no. 10, pp. 1345–1357, Oct. 1997.
[17] G. Fan and W.-K. Cham, "Model-based edge reconstruction for low bit-rate wavelet-compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 1, pp. 120–132, Feb. 2000.
[18] G. Fan and W.-K. Cham, "Postprocessing of low bit-rate wavelet-based image coding using multiscale edge characterization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1263–1272, Dec. 2001.
[19] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM, 2005.
[20] T. O'Rourke and R. Stevenson, "Improved image decompression for reduced transform coding artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 490–499, Dec. 1995.
[21] T. Meier, K. Ngan, and G. Crebbin, "Reduction of blocking artifacts in image and video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 3, pp. 490–500, Apr. 1999.
[22] M. Robertson and R. Stevenson, "DCT quantization noise in compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 27–38, Jan. 2005.
[23] B. K. Gunturk, Y. Altunbasak, and R. Mersereau, "Superresolution reconstruction of compressed video using transform-domain statistics," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 33–43, Jan. 2004.
[24] J. Mateos, A. Katsaggelos, and R. Molina, "A Bayesian approach for the estimation and transmission of regularization parameters for reducing blocking artifacts," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1200–1215, Jul. 2000.
[25] S. Roth and M. Black, "Fields of experts: A framework for learning image priors," in *Proc. IEEE Int. Conf. Computer Vision Pattern Recognition*, 2005, vol. 2, pp. 860–867.
[26] B. Widrow, I. Kollar, and M.-C. Liu, "Statistical theory of quantization," *IEEE Trans. Instrum. Meas.*, vol. 45, no. 2, pp. 353–361, Apr. 1996.
[27] O. Guleryuz, "Linear, worst-case estimators for denoising quantization noise in transform coded images," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2967–2986, Oct. 2006.
[28] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. B*, vol. 36, pp. 192–236, 1973.
[29] M. Welling, G. Hinton, and S. Osindero, "Learning sparse topographic representations with products of student-t distributions," *Neural Inf. Process. Syst.*, pp. 1359–1366, 2003.
[30] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application toevaluating segmentation algorithms and measuring ecological statistics," presented at the IEEE Int. Conf. Comp. Vision, 2001.
[31] MPEG4 Verification Model, VM 18.0 2001, pp. 271–275.
[32] M. Robertson and R. Stevenson, "Reduced-complexity iterative postfiltering of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 10, pp. 1121–1127, Oct. 2001.



Deqing Sun (S'05) received the B.Eng. degree from the Harbin Institute of Technology, China. He is currently pursuing the M.Phil. degree at the Department of Electronic Engineering, Chinese University of Hong Kong.

His research interests include signal processing, image processing, and computer vision.

Wai-Kuen Cham (S'77–M'79–SM'91) graduated from the Chinese University of Hong Kong in 1979 in electronics. He received the M.Sc. and Ph.D. degrees from Loughborough University of Technology, U.K., in 1980 and 1983, respectively.

From 1984 to 1985, he was a Senior Engineer with Datacraft Hong Kong Limited and a Lecturer in the Department of Electronic Engineering, Hong Kong Polytechnic. Since May 1985, he has been with the Department of Electronic Engineering, the Chinese University of Hong Kong. His research interests include image coding, image processing, and video coding.