# BAYESIAN POSTPROCESSING ALGORITHM FOR DWT-BASED COMPRESSED IMAGE

*Wei Wen, Zhiyun Xiao and Silong Peng*

National ASIC Design Engineering Center, Institute of Automation,
Chinese Academy of Sciences, Beijing 100080
Email: wei.wen@mail.ia.ac.cn

**ABSTRACT**

The perceived quality of compressed images is severely degraded especially when the bit rate becomes very low. The traditional postprocess methods will lose their effect in dealing with the DWT-based compressed image at very low bit rate because they do not consider the blurring effect in quantization process. In this paper, we propose a new model for the postprocess by incorporating a blur kernel into it, which is used to deblur. Median filter is used to detect and penalize the quantization noise. Under Bayesian analysis, MAP estimation is given. Alternate iteration method is proposed to solve this problem. Numerical experiments show that the subjective perceived quality as well as objective evaluation is improved.

## 1. INTRODUCTION

Postprocessing algorithms can be classified into two categories according to the transform used in compression: BDCT and DWT. The most noticeable artifacts for BDCT-based images or videos are the blocking artifacts. A lot of works have been done to deal with removal of the blocking artifacts [1]~[4]. For the DWT-based compressed image, ringing artifacts appears as small ripples around the edge when the bit rate becomes low. There are also some researches dealing with the postprocess on this kind of images [5]~[7]. But this area is not active compared with the area on the research of removal blocking artifacts.

From the technique viewpoint used to process the images, the postprocessing algorithms can be classified into two categories: image enhancement and image restoration. For algorithms based on image enhancement, the goal is to improve the perceived quality subjectively. For the image restoration approach, the postprocessing is viewed as an image recovery problem. Some classical image restoration techniques, including constrained least squares (CLS)[1], projection onto convex sets (POCS) [2][3], and maximum a posteriori (MAP) [4] restoration have been used to alleviate compression artifacts.

In this paper, we propose a new algorithm that postprocesses DWT-based compressed image at very low bit rate using the restoration technique. The motivation is that almost all the algorithms used on BDCT-based compressed images cannot be applied directly to the DWT-based compressed image because the boundary positions used as prior knowledge in the BDCT-based compressed images do not exist for DWT-based compressed images. And some other existent algorithms using either image restoration [4] or image enhancement [5] techniques on DWT-based compressed images have little effect when the compression bit rate is very low due to all of them do not consider the blurring effect in quantization process.

## 2. POSTPROCESSING MODEL

A typical image degradation model for wavelet-based image compression system has the form $y = W^{-1}Q^{-1}QWx$, where $y$ represents the degraded image, $x$ original image, $W$ wavelet transform, $Q$ quantization, $Q^{-1}$ inverse quantization and $W^{-1}$ inverse wavelet transform respectively. Thomas P. O'Rourke [4] uses this model and gives the MAP estimation directly. Because the conditional probability $P_r(y|x)$ is a constant, under the MAP estimation, the optimization is only on the prior knowledge $P_r(x)$. But only using this kind of prior is too week to deal with the postprocessing task especially when the bit rate is very low. From figure 1.a), we can see that the degraded image suffers from severe quantization noise as well as annoying blurring. For the purpose of removing quantization noise at the same time debluring, we present a new model for postprocessing

$$y = Dx + \xi , \tag{1}$$

where $D$ represents blur kernel and $\xi$ quantization noise respectively. Here we regard that the original image is firstly blurred by $D$ and then quantization noise is added before the degraded image $y$ is obtained.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

## 3. MAP ESTIMATION

Using the model (1) to postprocess the compressed image, a MAP estimation technique is proposed. The estimation of $D$ and $x$ is given by

$$(\hat{x}, \hat{D}) = \arg\max_{x,D} \{\log P_r(x, D|y)\}, \quad (2)$$

where $P_r(x, D|y)$ is the posteriori distribution which is a measure of how likely the $D$ and $x$ result in $y$. Using Bayes rule, (2) becomes

$$(\hat{x}, \hat{D}) = \arg\max_{x,D} \{\log P_r(y|x, D) + \log P_r(x|D) + \log P_r(D) - \log P_r(y)\}, \quad (3)$$

where the PDF $P_r(y)$ of the observed image is a constant which will be omitted in optimization. Considering that the blur kernel is only related with the compression process, or precisely related with the bit rate, we can regard that $D$ and $x$ are independent: $P_r(x|D) = P_r(x)$. So (3) can be simplified as

$$(\hat{x}, \hat{D}) = \arg\max_{x,D} \{\log P_r(y|x, D) + \log P_r(x) + \log P_r(D)\}. \quad (4)$$

### 3.1. Image Priors

Commonly, a useful model for the image prior is based on the stationary Gaussian zero-mean probability density function

$$P_r(x) = \frac{1}{\sqrt{\det(2\pi R_x)}} \exp\left\{-\frac{1}{2} x^T R_x^{-1} x\right\}, \quad (5)$$

where $R_x$ is the covariance matrix of image $x$. When $R_x^{-1} = \alpha' C^T C$, (5) becomes

$$P_r(x) = \frac{1}{Z} \exp\left\{-\frac{\alpha'}{2}\|Cx\|^2\right\}, \quad (6)$$

where $Z$ is a normalized constant and $\alpha'$ is a positive unknown parameter that controls the smoothness of the image. $C$ represents a highpass filter. From the analysis of section 2, we know that only using this kind of smoothness constraint is not enough for the postprocess task. Other smoothness constraint dealing with quantization noise must be considered which will play the role of suppressing the ringing effects. Here we use the 2D median filter and incorporate it into the PDF (6)

$$P_r(x) = \frac{1}{Z} \exp\left\{-\left(\frac{\alpha'}{2}\|Cx\|^2 + \frac{\beta'}{2}\|x - Med(x)\|^2\right)\right\}, \quad (7)$$

$Med(x)$ represent the median filtering process and $\|x - Mid(x)\|^2$ is used to punish the quantization noise. $\alpha'$ and $\beta'$ control the influence of the two norms.

### 3.2. The likelihood

We use Gaussian assumption for the noise $\xi$. It is straightforward to see that the likelihood is give by

$$P_r(y|x, D) = \frac{1}{\sqrt{\det(2\pi R_\xi)}} \exp\left\{-\frac{1}{2}(y - Dx)^T R_\xi^{-1}(y - Dx)\right\},$$

where $R_\xi$ is the covariance matrix of noise $\xi$. If $\xi$ is modeled as a zero-mean IID Gaussian process, we have $R_\xi = \sigma^2 I$. Then it becomes

$$P_r(y|x, D) = \frac{1}{Z} \exp\left\{-\frac{1}{2\sigma^2}\|y - Dx\|^2\right\}, \quad (8)$$

where $Z$ is a normalized constant too.

### 3.3. Blur Kernel Prior

There is little prior knowledge about the blur kernel prior, but we know that it is a lowpass filter and satisfied the normal lowpass filter constraint. Like the distribution in (6), we consider that the blur kernel is smooth too due to its lowpass property. So the prior is give by

$$P_r(D) = \frac{1}{Z} \exp\left\{-\frac{\gamma'}{2}\|AD\|^2\right\}, \quad (9)$$

where $A$ is a highpass filter and $\|AD\|^2$ is used to penalize the acute change of the kernel's coefficients.

## 4. SOLUTION OF MAP ESTIMATION

Substituting (7), (8) and (9) into (4), we get

$$(\hat{x}, \hat{D}) = \arg\max_{x,D} \left\{-\left(\frac{1}{2\sigma^2}\|y - Dx\|^2\right) - \left(\frac{\alpha'}{2}\|Cx\|^2 + \frac{\beta'}{2}\|x - Med(x)\|^2\right) - \left(\frac{\gamma'}{2}\|AD\|^2\right)\right\}, \quad (10)$$

Maximizing (10) is equal to minimizing following equation

$$(\hat{x}, \hat{D}) = \arg\min_{x,D} \{\|y - Dx\|^2 + \alpha\|Cx\|^2 + \beta\|x - Med(x)\|^2 + \gamma\|AD\|^2\}, \quad (11)$$

where $\alpha = \sigma^2\alpha'$, $\beta = \sigma^2\beta'$, $\gamma = \sigma^2\gamma'$. Let $J(x, D) = \|y - Dx\|^2 + \alpha\|Cx\|^2 + \beta\|x - Med(x)\|^2 + \gamma\|AD\|^2$, then the above MAP estimation problem can be developed into a constrained optimization problem

$$\min_{x,D} J(x, D) \quad s.t. \quad x \in C_x \quad D \in C_D,$$

where $C_x$ is a constraint set of $x$ and $C_D$ is a constraint set of $D$.

The gradient of $J(x, D)$ can be expressed as

$$p(x)=\frac{\partial J(x,D)}{\partial x}=2(D^T Dx-D^T y)+2\alpha C^T Cx+2\beta(x-M),$$

$$q(D)=\frac{\partial J(x,D)}{\partial D}=2(x^T xD-x^T y)+2\gamma A^T AD,$$

where $M=Med(x)$. Because the gradient of $Med(x)$ is difficult to get, we can regard it as a constant approximately when computing the gradient of $J(x,D)$. The minimum reaches when the gradient is equal to zero, then the above equations become

$$(D^T D+\alpha C^T C+\beta I)x=D^T y+\beta M,$$

$$(x^T x+\gamma A^T A)D=x^T y.$$

Here we use the Van Cittert [10] iteration method to solve them

$$\hat{x}_{k+1}=\hat{x}_k+\lambda_1[D^T y+\beta M-(D^T D+\alpha C^T C+\beta I)\hat{x}_k], \quad (12)$$

$$\hat{D}_{k+1}=\hat{D}_k+\lambda_2[x^T y-(x^T x+\gamma A^T A)\hat{D}_k],$$

(13)

where $\lambda_1$ and $\lambda_2$ are the step size of iteration. The iteration must be done alternatively, because when iterating one variation the other must be regarded as a determined one. The algorithm is given as following

1) Initialize $x$ and $D$:

$\hat{D}_0=ones(size)/\|ones(size)\|_1$, $\hat{x}_0=\lambda_1\hat{D}_0{}^T y$

2) At the $k$ th step:

a. Update $x$: $\bar{x}_{k+1}=\hat{x}_k+\lambda_1[\hat{D}_k{}^T y+\beta Med(\hat{x}_k)-(\hat{D}_k{}^T\hat{D}_k+\alpha C^T C+\beta I)\hat{x}_k]$

b. Update $D$: $\overline{D}_{k+1}=\hat{D}_k+\lambda_2[\bar{x}_{k+1}{}^T y-(\bar{x}_{k+1}{}^T\bar{x}_{k+1}+\gamma A^T A)\hat{D}_k]$

c. Project on the constraint set:

$\hat{x}_{k+1}=P_{C_x}(\bar{x}_{k+1})$, $\hat{D}_{k+1}=P_{C_D}(\overline{D}_{k+1})$

3) Stop when converged, else $k=k+1$ and go to 2)

In the theory of POCS [2][3], the constraint set has the form $C_x=\{x_k \mid F_i^{\min}\le(Wx_k)_i\le F_i^{\max},\forall i=1,2,\cdots N\}$, where $W$ represent the wavelet transform, $F_i^{\min}$ and $F_i^{\max}$ represent the quantizer's boundary at location $i$. The projection $P_{C_x}(x_k)$ of $x_k$ onto $C_x$ can be expressed as

$$P_{C_x}(x_k)=\begin{cases}F_i^{\min} & ,if\,(Wx_k)_i<F_i^{\min},\\ F_i^{\max} & ,if\,(Wx_k)_i>F_i^{\max},\\ (Wx_k)_i & ,otherwise\end{cases}.$$

For the constraint of blur kennel $D$:

$$D(i,j)\ge 0,\quad 1\le i,j\le m \text{ and } \sum_i\sum_j D(i,j)=1$$

the projection $P_{C_D}(D_k)$ can be expressed as

$$P_{C_D}(D_k)=\frac{D'_k}{\sum_i\sum_j D'_k(i,j)},\quad D'_k=\begin{cases}D_k & ,D_k\ge 0\\ 0 & ,D_k<0\end{cases}.$$

## 5. NUMERICAL EXPERIMENTS

The proposed method is applied to image that is compressed using EZW [8] algorithm at 0.07 bpp. In our experiment, the size of $D$ is selected as $7\times 7$, because the small size of $D$ has little effect to debluring while large size introduces ring effect in the restoration. Here we adopt the method similar to [9] to select the regularization parameter $\alpha$, $\gamma$,

$$\alpha=\frac{(C^T Cx)^T\cdot[-\beta(x-Med(x))-(D^T Dx-D^T y)]}{\|C^T Cx\|^2},$$

$$\gamma=\frac{(A^T AD)^T\cdot(x^T y-x^T xD)}{\|A^T AD\|^2}.$$

The regularization parameter $\beta$ is very sensitive to the selection in our experiment. This is easy to understand. Because that $\beta$ acts on the term $(x-Med(x))$, any value much larger than 1 will introduce a lager term $\beta x$, which will lead to unstableness in iteration. To ensure the robustness, we select that the value of $\beta$ will not exceed 1.1. Here $\beta$ is selected adaptively in the iteration process, which means a larger value of $\beta$ is selected at the beginning and a smaller one is used when the iteration is close to end. This will result in a better performance.

In [10], we know that small step size $\lambda_1$ and $\lambda_2$ should be selected to ensure the convergence. But proper scale between $\lambda_1$ and $\lambda_2$ is very important too because $\lambda_1$ acts on the iteration of $x$ and $\lambda_2$ acts on $D$. Instead of the method in [9], here we give a different explanation of how to choose the proportion. Analyze (15) and change it into

$$\hat{D}_{k+1}=\hat{D}_k+\lambda_2 x^T(y-x\hat{D}_k)-\lambda_2\gamma A^T A\hat{D}_k.$$

The second term $\lambda_2 x^T(y-x\hat{D}_k)$ has the largest energy. Its energy is almost $\sum x$ times larger than $(y-x\hat{D}_k)$, because of the convolution with $x^T$. Considering that the value of $D$ is not larger than 1 because that $\sum_i\sum_j D_{i,j}=1$, the energy of $(y-x\hat{D}_k)$ is at least 255 times larger than $D$. So $\lambda_2$ must be small enough. Here we select $\lambda_2=\frac{\lambda_1}{255\sum\hat{x}_k}$.

We compare our method with other method [6][7] and the results are presented in fig.1. The identified blur kernel after 30 times iteration is shown in fig.2. We use PSNR to assess the quality of image. In our experiment, the PSNR of compressed image is 26.29db. After the postprocess proposed by us, the PSNR is improved to 27.42db, more than 1db improvement is achieved. The PSNR results are shown in table 1.


(a) (b) (c) (d)

Fig. 1 The result: (a) Compressed image at 0.07 bpp; (b) Filtering method used in [7]; (c) Aria Nosratinia's method in [6]; (d) the proposed method in this article

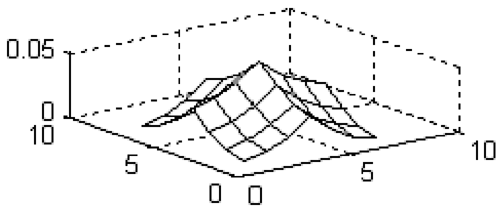

Fig. 2 The identified blur kernel $D$ after 30 times iteration

| Image | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| PSNR(db) | 26.29 | 25.57 | 24.95 | 27.42 |

Table 1. PSNR comparison

## 6. CONCLUTION

In this paper, a new compressed image postprocessing algorithm based on restoration theory is proposed. Under the Bayesian analysis, MAP estimation is given. We use the blur kernel to model the blurring effect in compressed image at the same time use median filter to detect the quantization noise and penalize it in the optimization process. Results show both the subjective perceived quality and objective evaluation are improved.

## 7. REFERENCES


[1] Y. Yang, N. Galatsanos, and A.K. Katsaggelos, "Regularized reconstruction to reduce blocking discrete cosine transform compressed images," IEEE Trans. Circuits Syst. Video Technol., vol. 3, no. 6, pp. 421-432, 1993

[2] Y.Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images," IEEE Trans. Image Processing, vol. 4, no. 1, pp. 896–908, 1995

[3] Y.Yang and N.P. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," IEEE Trans. Image Processing, vol. 6, no. 10, pp. 1345-1357, 1997

[4] Thomas P.O`Rourke, Robert L.Stevenson, "Improved Image Decompression for Reduced Transform Coding Artifacts," IEEE 1995

[5] Seungjoon Yang, Yu-Hen Hu, Truong Q. Nguyen, *et al.,* "Maximum-Likelihood Parameter Estimation for Image Ringing-Artifact Removal," IEEE Trans. Circuits Syst. Video Technol., vol. 11, no. 8, pp. 963-973, 2001

[6] Aria Nosratinia, "Post-Processing of JPEG-2000 Images to Remove Compression Artifacts," IEEE Signal Processing Letters, vol. 10, no. 10, pp. 296-299, 2003

[7] Wen Wei, Han Hua, Peng Silong, "DWT-based compressed image postprocessing algorithm using multiscale adaptive 2D filtering," Journal of Computer-Aided Design & Computer Graphics, to appear

[8] Jerome M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," IEEE Trans. Signal Processing, vol. 41, no. 12, pp. 3445-3462, 1993

[9] Yu-Li and M. Kaveh, "A Regularization Approach to Joint Blur Identification and Image Restoration," IEEE Trans. Image Processing, vol. 5, no. 3, pp. 416-428, 1996

[10] Zou Mouyan, Deconvolution and Signal Recovery, National Defence Industry Press, Beijing, 2001