MARKOV RANDOM FIELD ESTIMATION OF LOST DCT COEFFICIENTS IN JPEG DUE TO PACKET ERRORS

Jinwha Yang and Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering, Purdue University
West Lafayette, Indiana, U.S.A.

ABSTRACT

Interleaving is used before the encoding of source symbols in JPEG to reduce visual artifacts due to lost packets because interleaving distributes the locations of errors. The recovery of lost DCT coefficients in interleaved image compression is investigated in this paper. To restore the lost coefficients, an *Maximum a Posteriori* (MAP) estimate for the DCT coefficients is proposed. Under the assumption of a *Gauss-Markov Random Field* (GMRF) model in the pixel domain, the MAP estimate for the lost DCT coefficients is derived.

1. INTRODUCTION

In packet-based communication networks, data may be damaged by uncorrected errors and packets that were dropped due to network congestion. Retransmission strategies between end-to-end at the network layer and Automatic Repeat Request (ARQ) at the link layer are appropriate to deal with packet loss for non-real time applications such as file transfer. However, the end-to-end delay due to the retransmission of lost packets may not be acceptable in real-time applications such as media streaming.

When packet loss occurs in bursts during the delivery of compressed images or video, the packet loss manifests itself as a large damaged area in the image or video. In this case, *Error Concealment* (EC) techniques are used to recover the lost information, which involves processing at the decoder based only on a prior knowledge of the image or video [1, 2, 3]. The EC techniques can effectively reduce the visibility of transmission errors if the area of damaged pixels is not large. However, these EC techniques cannot effectively reduce the visibility of errors when the area of damaged pixels is large. Also, the visual quality of the restored region is not uniform compared to that of the undamaged neighboring region.

If the EC at the decoder is combined with pre-processing at the encoder such as interleaving, the visibility of the errors can be reduced. We propose a reconstruction algorithm for lost DCT (*discrete cosine transform*) coefficients in blocks of images compressed by JPEG assuming that the encoder interleaves the DCT coefficient of each block before entropy coding[1]. In this paper, we consider the concealment of packet loss errors in the DCT coefficients after interleaving. We are not concerned about how the coefficients are interleaved.

This work was supported by a grant from the Indiana 21^{st} Century Research and Technology Fund. Address all correspondence to E. J. Delp, ace@ecn.purdue.edu.

[1] It is assumed that all the DCT coefficients in each entropy coded block are from different blocks after interleaving. The interleaving process permutes the DCT coefficients of the image such that each of $m \times m$ coefficients appearing in each block after interleaving originated from different blocks. The inverse permutation is used by the decoder (deinterleaving) after missing coefficients are restored to reconstruct the image.

2. RESTORATION OF MISSING COEFFICIENTS

Let X be a decoded JPEG image of width W and height H after deinterleaving at the decoder. Assume that the compressed image data is error-free and that W and H are integral multiples of block size m. Let b_i be the i^{th} block of X in the raster scan ordering of blocks, where each block is $m \times m$ pixels in size. Let C_{b_i} correspond to the $m \times m$ DCT coefficients of b_i. Within b_i, x_k denotes the pixel at $(k/m, k \bmod m)$ coordinate relative to the top left corner of b_i. Similarly to the representation of x_k, c_p is the reordered 2-D DCT coefficients in C_{b_i} after deinterleaving. Then the DCT coefficients vector C_{b_i} and the pixel vector X_{b_i} have the following transform,

$$\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ \vdots \\ c_{m^2-1} \end{bmatrix} = \begin{bmatrix} t_{0,0} & t_{0,1} \dots & t_{0,m^2-1} \\ & \vdots & \\ t_{p,0} & t_{p,1} \dots & t_{p,m^2-1} \\ & \vdots & \\ t_{m^2-1,0} & \dots & t_{m^2-1,m^2-1} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ \vdots \\ x_{m^2-1} \end{bmatrix}, \quad (1)$$

$$C_{b_i} = T_b \, X_{b_i}, \quad (2)$$

where the $t_{p,k}$ is the term at p^{th} row and k^{th} column of 2-D DCT transform matrix T_b. Since $X^t = [X^t_{b_1}, X^t_{b_2}, \cdots, X^t_{b_N}]$ and $C^t = [C^t_{b_1}, \cdots, C^t_{b_N}]$ are block vectors and $T = \text{diag}[T_b, \cdots, T_b]$ are block matrices, the DCT transform for the entire image becomes $C = T\,X$.

Now, we will assume that blocks in packets may be lost. The packet indices of the interleaved blocks lost are assumed to be known during reception and thus the indices of the lost interleaved blocks are deduced at the receiver. After deinterleaving, each DCT coefficient block is formed as $C_{b_i} = C^R_{b_i} + C^L_{b_i}$, where $C^R_{b_i}$ and $C^L_{b_i}$ correspond to the received and the lost coefficients in the block b_i, respectively. Once indices of lost block and the interleaving permutation are known, the locations of $C^L_{b_i}$ in $m^2 \times 1$ vector are deduced and the map M_{b_i} of lost coefficient locations in C_{b_i} can be obtained. M_{b_i} is a $m^2 \times m^2$ diagonal matrix diag$[m_{ii}]$ such that $m_{ii} = 1$ for the received coefficients and $m_{ii} = 0$ for the lost coefficients. Since $C^R_{b_i} = M_{b_i} C_{b_i}$, the received image block Y_{b_i} and image Y are expressed as

$$Y_{b_i} = T^t_b C^R_{b_i} = T^t_b M_{b_i}(T_b X_{b_i}), \quad (3)$$

$$Y = T^t \, C^R = T^t \, M(T\,X), \quad (4)$$

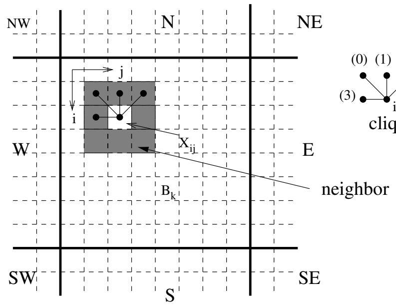
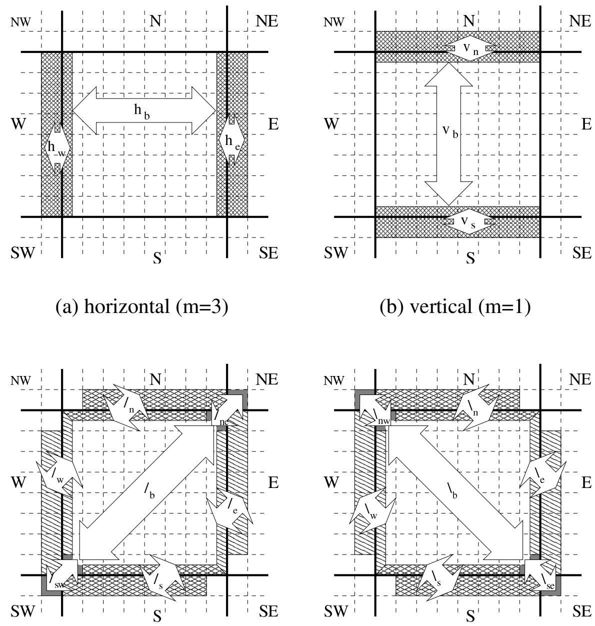


Fig. 1. Clique at pixel $X_{i,j}$ in block B_k. Neighbor blocks are designated according to azimuth direction.



(a) horizontal (m=3) (b) vertical (m=1)

(c) diagonal left (m=2) (d) diagonal right (m=0)

Fig. 2. Potential energy contribution of each group of cliques according to index m.

where Y and C^R are block vectors, and $M = \text{diag}[M_{b_1}, \cdots, M_{b_N}]$.

To estimate the lost coefficients C^L for each block, we use Bayesian MAP estimation assuming that the decoded JPEG image X without error is modeled as a Markov random field (*MRF*). Let $g(x)$ be the a priori distribution for X. Then the MAP estimate $\widehat{X}$ in the pixel domain is expressed as [4]

$$\begin{aligned} \widehat{X} &= \arg \max_{X|Y=T^tMTX} p(x|y) \\ &= \arg \max_{X|Y=T^tMTX} \{\log p(y|x) + \log g(x)\}. \end{aligned} \quad (5)$$

In (5), $\log p(y|x)$ becomes constant when the probability distribution of packet loss is considered to be uniform. Then the MAP estimate reduces to $\widehat{X} = \arg \max_{X|Y=T^tMTX} \log g(x)$, which includes only the a priori probability term. For the MRF, the prior probability distribution of X follows the Gibbs distribution [5, 6, 4] given by $f(x) = \frac{1}{Z} \exp\left\{-\sum_{c\in C} V_c(x)\right\}$, where Z is a normalizing constant, $V_c(\cdot)$ is a potential function of a local group of pixel configuration c defined as a clique [4, 3], and C is the set of all such configurations [7, 5]. The MAP estimate in (5) reduces to $\widehat{X} = \arg \min_{X|Y=T^tMTX} \left\{\sum_{c\in C} V_c(x)\right\}$.

The general form of the potential function with a second order clique system is

$$\sum_{c\in C} V_c(X) = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} \sum_{m=0}^{3} w_{i,j}^{(m)} \rho\left(\frac{D_m(X_{i,j})}{\sigma}\right), \quad (6)$$

where $w_{i,j}^{(m)}$ denotes the weight, and $D_m(X_{i,j}) = X_{i,j}^{(m)} - X_{i,j}$ denotes the pixel value difference, respectively depending on the direction of a clique indexed by m. σ is a scaling factor. For simplicity, we have chosen the Gaussian MRF prior as the potential function and typical second order clique system. Hence $\rho(\cdot) = (\cdot)^2$, $w_{i,j}^{(m)} = 1$ and the scaling factor $\sigma = 1$. The set of cliques consists of northwest(0), north(1), northeast(2) and west directions(3) as shown in Figure 1.

2.1. Transformation of the Cost Function

The a prior probability, $\log g(x)$, of pixels in (6) is converted in terms of DCT coefficients [8]. Equivalently, we can convert the potential function to be expressed with DCT coefficients. Using the DCT transform relation (1), MAP estimate for lost DCT coefficient becomes

$$\widehat{C_b} = \arg \min_{C_b^R = M_b C_b} \frac{1}{|J|} \sum_{c\in C} V_c((T^t C_b)_x),$$

where $|J|$ is the Jacobian, and we have $|J| = 1$ since the DCT used in JPEG is a unitary transform.

For the potential energy between the neighbors in a block, we can regroup the energy according to clique directions by rearranging the summation order in (6). As in Figure 2, the horizontal ($h_{(d)}$) and the vertical ($v_{(d)}$) contributions are divided into three groups, respectively, where $d \in \{n, w, e, s, nw, ne, sw, se, b\}$ and b stands for the center block as in Figure 1. The two diagonal ($\backslash_{(d)}$, $/_{(d)}$) contributions are further classified into seven groups. For an example, $\backslash_{nw}$ corresponds to the contribution between the left-upper corner pixel in a block and the right-lower corner pixel in the north west boundary block. The pixel and the DCT locations are indexed as a one dimensional index k as in (1). The entire neighbor blocks are represented as $m^2 \times 1$ column vector, $\mathbf{C}_{(d)}$, scanned in raster order in the block.

Summing up the above contributions, the energy for a block becomes $\sum_{c\in block} V_c = [h_b + h_w + h_e + v_b + v_n + v_s + /_b + /_w + /_s + /_{sw} + /_e + /_n + /_{ne} + \backslash_b + \backslash_w + \backslash_s + \backslash_{se} + \backslash_e + \backslash_n + \backslash_{nw}]$. Since $\rho(\cdot) = (\cdot)^2$ as mentioned in section 2, the energy contribution $h_b = \mathbf{C_b^t B_1^{b^t} B_1^b C_b}$, where $\mathbf{B_1^b} = [t_{0,mi+j} - t_{0,mi+j+1}, \cdots, t_{m^2-1,mi+j} - t_{m^2-1,mi+j+1}]_{\substack{i=[0,m-1]\\ j=[0,m-2]}}$ is a matrix with $m \times (m-1)$ rows. $\mathbf{B_1^b}$ is obtained by linear combination of DCT transform basis on selected i, j within the given boundary as in Figure 2. Similarly the remaining energy contributions h_w to $\backslash_{nw}$

are converted into matrix form $\mathbf{B_i^{(d)}}$. Finally, the energy function for a block can be written collectively using matrix notation $\sum_{c \in block} V_c = \mathbf{C^t}\,\mathbf{A^t}\,\mathbf{A}\,\mathbf{C}$, where $\mathbf{C}$ and $\mathbf{A}$ are given in (7). $\mathbf{C}$ is a $9m^2 \times 1$ column vector and $\mathbf{A}$ is a $(4m^2+6m-2) \times 9m^2$ matrix with rank $(m+2)^2-1$.

$$
\mathbf{A}\,\mathbf{C} =
\begin{bmatrix}
0 & 0 & 0 & 0 & \mathbf{B_1^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mathbf{B_2^w} & -\mathbf{B_2^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_3^b} & \mathbf{B_3^e} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_4^b} & 0 & 0 & 0 & 0 \\
0 & \mathbf{B_5^n} & 0 & 0 & -\mathbf{B_5^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_6^b} & 0 & 0 & -\mathbf{B_6^s} & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_7^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mathbf{B_8^w} & -\mathbf{B_8^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_9^b} & 0 & 0 & -\mathbf{B_9^s} & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_{10}^b} & \mathbf{B_{10}^e} & 0 & 0 & 0 \\
0 & \mathbf{B_{11}^n} & 0 & 0 & -\mathbf{B_{11}^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_{12}^b} & 0 & 0 & 0 & -\mathbf{B_{12}^{se}} \\
\mathbf{B_{13}^{nw}} & 0 & 0 & 0 & -\mathbf{B_{13}^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_{14}^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\mathbf{B_{15}^w} & \mathbf{B_{15}^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_{16}^b} & 0 & 0 & -\mathbf{B_{16}^s} & 0 \\
0 & 0 & 0 & 0 & -\mathbf{B_{17}^b} & \mathbf{B_{17}^e} & 0 & 0 & 0 \\
0 & \mathbf{B_{18}^n} & 0 & 0 & -\mathbf{B_{18}^b} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{B_{19}^b} & 0 & -\mathbf{B_{19}^{sw}} & 0 & 0 \\
0 & 0 & \mathbf{B_{20}^{ne}} & 0 & -\mathbf{B_{20}^b} & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\mathbf{C_{nw}^t} \\ \mathbf{C_n^t} \\ \mathbf{C_{ne}^t} \\ \mathbf{C_w^t} \\ \mathbf{C_b^t} \\ \mathbf{C_e^t} \\ \mathbf{C_{sw}^t} \\ \mathbf{C_s^t} \\ \mathbf{C_{se}^t}
\end{bmatrix}
. \quad (7)
$$

2.2. Minimization of the Cost Function

Using the matrix notation, the MAP estimation is formulated as Lagrangian [9] minimization problem:

$$
\begin{aligned}
\text{minimize} \quad J(\mathbf{C_b}) &= \frac{1}{2}\mathbf{C^t A^t A C} \qquad (8)\\
\text{subject to} \quad \mathbf{SC_b} &= \mathbf{R_b}, \qquad (9)
\end{aligned}
$$

where $\mathbf{S}$ is the matrix excluding all the zero rows in M_{b_i} given in (4), $\mathbf{C_b}$ is a column vector of the center DCT block, and $\mathbf{R_b}$ is a compact vector composed of only the received coefficient in the block $\mathbf{C_b}$. With respect to the contribution of neighbor blocks and of transform kernel, the objective function $J(\mathbf{C_b})$ is rearranged as

$$
J(\mathbf{C_b}) = \mathbf{\frac{1}{2}}\left[\mathbf{C_b^t}\psi^{\mathbf{t}}\psi\mathbf{C_b} - \mathbf{2N_c^t}\psi\mathbf{C_b} + \mathbf{N_c^t N_c}\right].
$$

ψ^t is the matrix written as $[\mathbf{B_1^b}^{\mathbf{t}}, \mathbf{B_2^b}^{\mathbf{t}}, \cdots, \mathbf{B_{20}^b}^{\mathbf{t}}]$ and $\mathbf{N_c^t}$ is the row vector composed of only neighbor blocks [0^t, $\mathbf{C_w^t B_2^{w\,t}}$, $\mathbf{C_e^t B_3^{e\,t}}$, 0^t, $\mathbf{C_n^t B_5^{n\,t}}$, $\mathbf{C_s^t B_6^{s\,t}}$, 0^t, $\mathbf{C_w^t B_8^{w\,t}}$, $\mathbf{C_s^t B_9^{s\,t}}$, $\mathbf{C_e^t B_{10}^{e\,t}}$, $\mathbf{C_n^t B_{11}^{n\,t}}$, $\mathbf{C_{se}^t B_{12}^{se\,t}}$, $\mathbf{C_{nw}^t B_{13}^{nw\,t}}$, 0^t, $\mathbf{C_w^t B_{15}^{w\,t}}$, $\mathbf{C_s^t B_{16}^{s\,t}}$, $\mathbf{C_e^t B_{17}^{e\,t}}$, $\mathbf{C_n^t B_{18}^{n\,t}}$, $\mathbf{C_{sw}^t B_{19}^{sw\,t}}$, $\mathbf{C_{ne}^t B_{20}^{ne\,t}}$]. To solve the Lagrangian minimization, let's define new objective function $l(\mathbf{C_b}, \lambda) = \mathbf{J}(\mathbf{C_b}) + \lambda^{\mathbf{t}}(\mathbf{SC_b} - \mathbf{R_b})$. Since the objective function is quadratic, we expect to find a minimizer $\widehat{\mathbf{C}}_b$ by solving the Lagrange conditions

$$
\left\{
\begin{aligned}
D_{C_b} l(\mathbf{C_b}, \lambda) &= \mathbf{0^t} \\
D_\lambda l(\mathbf{C_b}, \lambda) &= \mathbf{0^t},
\end{aligned}
\right. \qquad (10)
$$

where D_x implies differentiation with respect to variable x. Let $\mathbf{K} = \psi^{\mathbf{t}}\psi$ and $\mathbf{C}_* = \mathbf{N_c^t}\psi$. The solution of (10) becomes the minimizer

$$
\widehat{\mathbf{C}}_b = \mathbf{K^{-1}C_*}^{\mathbf{t}} - \mathbf{K^{-1}S^t}(\mathbf{SK^{-1}S^t})^{-1}(\mathbf{SK^{-1}C_*}^{\mathbf{t}} - \mathbf{R_b}).
$$

This is the local minimizer for a given block. To achieve a global minimum throughout the entire image, we need to perform the block-wise estimation repeatedly until the final estimator converges to a global minimum similar to the *Iterative Conditional Mode* method [7].

3. SIMULATION

Four test images were DCT transformed and quantized using JPEG. After the DCT coefficients in each block are interleaved, each block is coded with JPEG's Huffman code. We assumed that DC is not predictively encoded. Block loss is simulated using an uniform error distribution, with ranges from 5 to 50 % in steps of 5 % increase. The general circular boundary condition[2] was assumed for the MAP estimate of boundary blocks.

Since the proposed method depends on iterations, the convergence rate is illustrated in Figure 4. It shows that as the loss rate becomes larger, the more iterations are required to achieve convergence. However, even for the 50 % block loss rate, the estimate converges to a global minimum only after about 20 iterations. For the case of simple recovery, the lost DCT coefficients were set to zero after deinterleaving. The decoded images using the *barbara* test image with this simple recovery are shown in the first row of Figure 3 from 10% to 50% loss. The restored images using our MAP estimation are shown in the second row of Figure 3.

Due to the circular symmetric boundary assumption, the restoration for blocks near the top and the bottom edges of the images seem rather poor when compared with the restored blocks in the middle region of the images. Among the test images, the *text* image shows steeper PSNR degradation as the block loss rate increases, than the other images. This is because the choice of the $\rho(\cdot)$ energy cost function in (6) was derived from the GMRF, which is reported to work poorly for images with discontinuities [4]. On the other hand, the GMRF provides an analytic solution for local minimum and fast convergence.

4. CONCLUSION

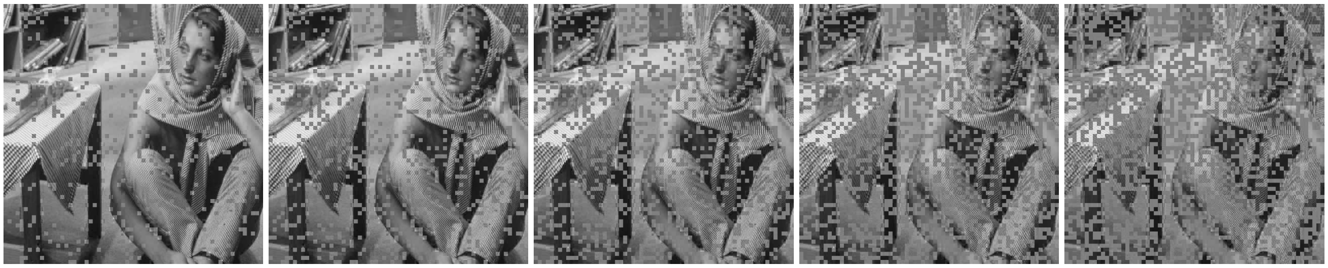
The new estimator converges to a global minimum very quickly. When the DCT interleaving and the MAP estimation operate together, the reconstructed images with the block loss due to burst errors showed graceful degradations with little unevenness between blocks.

As further research, the DCT domain interleaving on compressed video is being investigated. For predicted error frames, the description of a prior probability using a 3D MRF model is being studied.

5. REFERENCES

[1] Yao Wang and Qin-Fan Zhu, "Error control and concealment for video communication: A review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

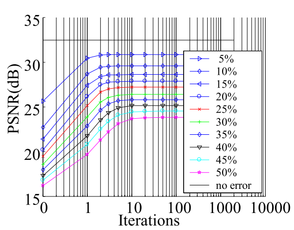
[2] This corresponds to deforming an image plane into a doughnut shaped ring. First a cylinder is formed by rolling the plane to meet the top and the bottom edge. Then a ring is formed by bending the cylinder for the right and the left edges to meet together.



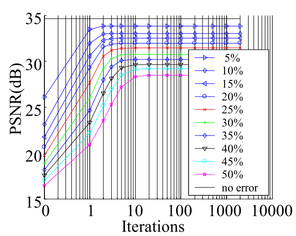
(a) 10% packet loss (b) 20% packet loss (c) 30% packet loss (d) 40% packet loss (e) 50% packet loss

(f) 10% loss recovery (g) 20% loss recovery (h) 30% loss recovery (i) 40% loss recovery (j) 50% loss recovery

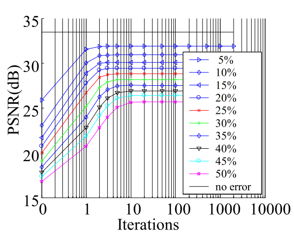
Fig. 3. Comparison of restored barbara images at various packet loss rates. The upper row shows the decoded JPEG images by setting all the lost DCT coefficients as zero. The bottom row shows the JPEG images decoded after our MAP estimation.



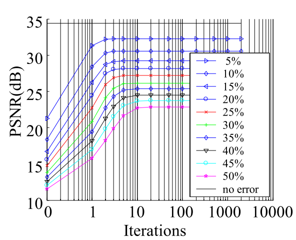
(a) barbara



(b) girls



(c) goldhill



(d) text

Fig. 4. PSNR convergence with respect to iteration step for the estimation. Convergence of PSNR in four different images are simulated from 5% packet loss to 50% loss.

[2] John D. Villasenor, Ya-Qin Zhang, and Jiangtao Wen, "Robust video coding algorithms and systems," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1724–1733, October 1999.

[3] Paul Salama, Ness B. Shroff, and Edward J. Delp, "Error concealment in encoded video streams," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1129–1144, June 2000.

[4] Charles Bouman and Ken Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[5] Stuart Geman and Donald Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, November 1984.

[6] Donald Geman and George Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Transactions on Pattern and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, March 1992.

[7] Julian Besag, "On the statistical analysis of dirty pictures," *Journal of Royal Statistical Society*, vol. 48, no. 3, pp. 259–302, 1986.

[8] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.

[9] Edwin K. P. Chong and Stanislaw H. Zak, *An Introduction to Optimization*, John Wiley & Sons, Inc., New York, 1996.