MAP-BASED POST PROCESSING OF VIDEO SEQUENCES USING 3-D HUBER-MARKOV RANDOM FIELD MODEL

Zhen Li and Edward J Delp[†]

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue Univ., West Lafayette, IN 47907-1285 USA

ABSTRACT

Block DCT (BDCT) is by far one of the most popular transforms used in image and video coding. However, it introduces a noticeable blocking artifact at low data rates. A great deal of work has been done to remove the artifact with information extracted from the spatial and frequency domains. In this paper we address the video sequence restoration problem as a 3-D Huber-Markov Random Field model and derive the temporal extension to traditional maximum *a posteriori* (MAP)-based methods. Two schemes, we call Temporal MAP (TMAP) and Motion Compensated TMAP (MC-TMAP) respectively, are presented. We test our methods on MPEG-2 compressed sequences and evaluate their performances with traditional MAP restoration. Experiment results confirm that our schemes can significantly improve the visual quality of the reconstructed sequences.

INDEX TERMS - BDCT, MAP, Huber-Markov Random Field, Motion Compensation

1. INTRODUCTION

Transform techniques have proved to be very effective in image and video coding over the last two decades. Among them, the block-based Discrete Cosine Transform (BDCT) is most widely used due to its sub optimality to the Karhunen-Loeve Transform (KLT) and availability of fast algorithms.

However, it is also known that BDCT will introduce blocking artifacts at low data rates. The blocking artifact manifests itself as an annoying discontinuity between adjacent blocks. This is a direct result of independent transformation and quantization of each block, which fails to take into account the inter-block correlation. The artifact is propagated and increased in video coding due to the use of motion compensation.

A great deal of research has been done to remove the blocking artifact. Post-processing approaches, such as low-pass post filtering [1], iterative restoration with projection onto convex sets (POCS) [2], over-complete wavelet representation [3] and maximum *a posteriori* (MAP) restoration [4], have been popular since they maintain the compatibility with the original decoder structure. In this paper we base our work on MAP estimation, in particular, by modeling the video sequence as a 3-D Huber Markov Random Field (HMRF) [5].

When used in video post processing, these techniques generally treat each frame as a single image and perform post-processing independently. However, the blocking artifact in video sequences comes from two sources: (a) The quantization noise at each block boundary in the reference or residue frames, which is the same as in the case of still image; (b) The propagation of blocking artifact from previous frames due to motion compensation. Also, as the adjacent video frames are not independent, lack of temporal information leads to either inaccuracy of determining the block location or unnecessary computation.

In this paper we aim to exploit the temporal dependence among frames by using a new 3-D HMRF model. Based on this, we propose a temporal extension, known as TMAP to traditional MAP estimation. We also further consider the effect of motion compensation and describe our Motion-Compensated TMAP (MC-TMAP) scheme. We test our schemes on MPEG-2 bit streams and demonstrate significant improvements of the reconstructed video sequences.

2. MAP RESTORATION OF STILL IMAGES

Maximum *A Posteriori* (MAP) restoration models image restoration as a Bayesian estimation problem, i.e. assume

[†] This work was supported by the Indiana 21st Century Research and Technology Fund.
Address all correspondence to E J Delp, ace@ecn.purdue.edu

0-7803-7304-9/02/$17.00 C2002 IEEE

$$Y = X + N \quad (1)$$

where Y is the observed noisy image, X is the original image and N is the noise. Then the restored image is

$$\tilde{X} = \arg\max_X \{\log \Pr(Y \mid X) + \log \Pr(X)\} \quad (2)$$

Moreover, the image X can be regarded as a Markov Random Field represented by a Gibbs Distribution

$$\Pr(x) = \frac{1}{Z}\exp\{-\sum_{c \in C} V_c(x_c)\} \quad (3)$$

where

x_c - The value of X at the points in clique c

$V_c(x_c)$ - A potential function of x_c

C - The set of all cliques

Z - The normalizing constant for the density

$U(x) = \sum_{c \in C} V_c(x_c)$ is known as energy function, and

$$U(x) = \sum_{x_n \in N_x} \rho_T(x - x_n) \quad (4)$$

Where N_x is the neighborhood of x, which is the eight nearest pixel around x, and $\rho_T(.)$ is a potential function.

With (4), (2) is further simplified to

$$\tilde{X} = \arg\min_X \sum_{x \in X} \sum_{x_c \in N_x} \rho_T(x - x_n) \quad (5)$$

By taking $\rho_T(.)$ as the Huber cost function [6]

$$\rho_T(u) = \begin{cases} u^2, |u| \le T \\ T^2 + 2T(|u| - T), |u| > T \end{cases} \quad (6)$$

where T is a threshold value, we construct the 2-D Huber-Markov Random Field model (2-D HMRF) model. Due to the convexity of the Huber function, the optimization problem in (5) is a constrained minimization problem, which can be solved by iterative method, such as Golden Section Search method. A detailed discussion can be found in [4].

Basically there are two parts in a MAP restoration scheme. One is to penalize the discontinuities and encourage smoothness, denoted as the "smoothing term"; the other is to encourage the fidelity of the image based on some prior knowledge, denoted as the "fidelity term". HMRF model enforces the "smoothing term" in MAP, which tend to punish large energy value. Fidelity is generally ensured by the Quantization Constraint Set (QCS). It requires that the DCT coefficients of the processed image fall into the same DCT quantization range as the unprocessed image.

3. TEMPORAL EXTENSION: 3-D HMRF

While most traditional MAP-based video restoration simply use still image techniques mentioned in Section 2 to each frame independently, some research has paid attention to temporal correlation [7]. In [8], the adjacent frames are taken into account and the MAP problem is formulated as:

$$\tilde{X}_k = \arg\max_{X_k} \Pr(X_k \mid Y_l, l = k-m, \ldots, k+m) \quad (7)$$

where X_k is the original $k-th$ frame and Y_k is the received $k-th$ frame.

In [8], the "smoothing" term is the same as the still image case. The temporal correlation is considered in the "fidelity" term, where a multi-frame Gaussian noise model, rather than the QCS, is used. Our research differs from [8] in that we also exploit the temporal correlation in the "smoothing" term, which we believe is important in video restoration. Also we take as prior knowledge the processed adjacent frames, instead of the received adjacent frames, to provide more accurate "smoothing" information. The MAP problem here is formulated as

$$\tilde{\underline{X}} = \arg\max_{\underline{X}} \Pr(\underline{X} \mid \underline{Y}) \quad (8)$$

where $\underline{X} = [X_{k-m}, \ldots, X_k, \ldots, X_{k+m}]$ is the original frame vector and $\underline{Y} = [Y_{k-m}, \ldots, Y_k, \ldots, Y_{k+m}]$ is the received frame vector. The iterative MAP solution to (8) is

$$\tilde{X}_k^{(n+1)} = \arg\max_{X_k}\{\log \Pr(X_{k-m}^{(n+1)}, \ldots, X_{k-1}^{(n+1)}, X_k^n, \ldots, X_{k+m}^n \mid X) + \log \Pr(X_k)\} \quad (9)$$

Where $X_k^{(n)}$ is the post-processed $k-th$ frame after n iterative steps.

Our 3-D Huber-Markov Random Field model is based on the generalization of the 2-D HMRF. The neighborhood system N_x and cliques is extended to include the adjacent frames. It includes not only the eight nearest pixels around x in the current frame, which we denote as $N_{x,k}$, but also the nearest pixels in the adjacent frames, which we denote as $N_{\hat{x}_{k+l}}$, $-m \le l \le m$, i.e.,

$$N_x = \bigcup_{l=-m}^{m} N_{\hat{x}_{k+l}} \quad (10)$$

where $\hat{x}_{k+l}$ is the "corresponding pixel " of x in $X_{k+l}^{(n)}$. The corresponding pixel is either in the same the spatial position in the frame as x, or determined by a searching algorithms described later.

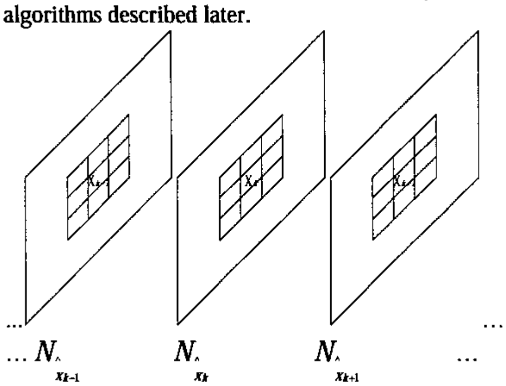

Fig. 1 Neighborhood system in 3-D HMRF

With (10), the energy function is

$$U(x) = K\sum_{l=-m}^{m} |\lambda_{k,k+l}(\sum_{x_n \in N_{\hat{x}_{k+l}}} \rho_T(x-x_n))| \quad (11)$$

where $\lambda_{k,k+l}$ is the correlation coefficient between N_{x_k} and $N_{\hat{x}_{k+l}}$,which is determined later, and

$$K = 1/(\sum_{l=-m}^{m} \lambda_{k,k+l}) \quad (12)$$

is the normalizing factor . By adjusting m , we can control the influence of temporal knowledge. When $m = 0$, it is the traditional MAP restoration scheme.

Using the Huber function defined in (6) as potential function, we construct the 3-D Huber Markov Random Field Model (3-D HMRF). Since (12) is also convex, the solution can be obtained from the similar optimization methods [9] described in Section 2.

The correlation coefficient λ should meet the following four conditions:

i. $0 \le \lambda \le 1$;

ii. $\lambda = 1$ if two frames are fully correlated;

iii. $\lambda = 0$ if two frames are independent;

iv. λ vanishes with the distance between the frames.

We propose the following way to determine λ

$$\lambda_{k,k+l} = \frac{< N_{\hat{x}_k} - avg(N_{\hat{x}_k}), N_{\hat{x}_{k+l}} - avg(N_{\hat{x}_{k+l}}) >}{| N_{\hat{x}_k} - avg(N_{\hat{x}_k}) || N_{\hat{x}_{k+l}} - avg(N_{\hat{x}_{k+l}}) |} \quad (13)$$

Where $< A, B >$ is the inner product of matrix A and B, $| A |= \sqrt{< A, A >}$ is the norm of A and $avg(A)$ is the matrix whose entries are the average of all entries of A.

We can also look at the penalty value in (11) from the point view of parameter prediction, where

$$U_k(x) = |\sum_{l=-m}^{m} \lambda_{k,k+l} U_{k+l}(x)| / |\sum_{l=-m}^{m} \lambda_{k,k+l}| \quad (14)$$

i.e., the energy value in current frame is the output of an IIR filter which take the energy values of adjacent frames as input. Hence we allow large "energy value" if previous energy values are large, and we also penalize small energy value, which may lead to blurring, if it is much larger (by some decision rule) compared with previous values. Hence 3-D HMRF model can provide more accurate blocking decision and edge-preservation capability. On the other hand, the traditional MAP model makes penalty decision by only comparing the energy value with some fixed threshold T_{thr}, which determines the strength of de-blocking. It should be noted that when T_{thr} is set to ∞, it is the original reconstructed sequences at decoder.

Based on different approaches to determine the "corresponding pixel", we propose two temporal extensions to 2-D HMRF MAP restoration.

3.1. Direct Temporal Extension to MAP (TMAP)

In this scheme we take the pixel at the same spatial position in the adjacent frames as the corresponding pixel, i.e., $\hat{x}_{k+l}(m,n) = x_{k+l}(m,n)$, where m, n denote the spatial position of the pixel.

3.2. Motion Compensated Temporal Extension to MAP (MC-TMAP)

In TMAP we implicitly assume that the adjacent frames are stationary. However, this is generally not true in video sequences. To get a better restoration, we need to consider the effect of motion compensation.

While there are various approaches to search the motion vectors in reconstructed video sequences, we use the criterion that finds the corresponding pixel that maximizes λ in a N by N search window.

4. EXPERIMENTAL RESULTS

Both TMAP and MC-TMAP are used to restore MPEG-2 decoded video sequences. Examples are shown in figure2 - figure 6(only a part of the frame is shown to demonstrate the effect). The CCIR 601 *flower garden* (720×486) sequences are compressed at 15 frames/second with a data rate of 2.0 Mbps. Although larger window size m can improve the accuracy of the determination of the blocking area, we fix $m = 2$ due to computation and frame buffer limit.

Figure 3 shows that the reconstructed image is visibly blocky. The traditional MAP restoration can greatly remove the blocky artifact, as shown in figure 4, but introduce a blurring effect, this is especially true in those regions with little motion (e.g. the details on the roof are totally blurred). This is because the traditional approach tends to penalize abrupt discontinuities in the image. The result in figure 5 shows that such blurring is avoided in the "slow-motion" area with TMAP by exploiting temporal knowledge. The result with MC-TMAP in figure 6 further improves the visual quality by taking motion compensation into account.

In each iterative step, the DCT and IDCT pair is still only applied once, so the computation cost of each iterative step is nearly the same in traditional MAP and TMAP. The only additional cost is the increase of frame buffer. There is, however, some additional computation cost in each step at MC-TMAP due to searching the "corresponding pixel".

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a temporal extension to traditional MAP restoration based on 3-D Huber-Markov Random Field. Adjacent frames are considered to improve the restored quality. Experiment results demonstrate the efficiency of our schemes.

In the experiments we notice that abrupt scene changes may affect the efficiency of the 3-D HMRF model, since our model outperforms traditional 2-D model mostly as a result of exploiting correlations among adjacent frames. Currently we are considering a further improvement with scene change detection. Also more work need to be done on buffer control. A complexity adaptive model should also be investigated for the real time applications.

REFERENCES

[1] ITU-T, "ITU-T Recommendation H.263: Video Coding for Low Bit Rate Communication", Sept, 1997

[2] A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," IEEE Trans. CSVT., vol.2,p91-95, Mar. 1992

[3] Z. Xiong, M.T. Orchard and Y-Q. Zhang, "A simple deblocking algorithm for JPEG compressed images using overcomplete wavelet representations," IEEE Trans. CSVT, Vol. 7, p433-437, April 1997

[4] T.O'Rourke and R. Stevenson, "Improved image decompression for reduced transform coding artifacts," IEEE Trans. CSVT, vol.5, pp.490-499, Dec 1995

[5] R. Stevenson and E. J. Delp, "Fitting curves with discontinuities", Proc. Of the 1st International Workshop on Robust Computer Vision, P127-136, Oct, 1990

[6] P. Huber, Robust Statistics. New York: Wiley, 1981

[7] J Brailean and R. Lagendijk, etc, "Noise Reduction Filters for Dynamic Image Sequences: A Review", Proc. Of the IEEE, vol. 83, No. 9, p1272 – p1292, Sept. 1995

[8] Mark Robertson and Robert Stevenson, "Restoration of compressed video using temporal information" , Proc. Of VCIP 2001, p21-29, San Jose, Jan 2001

[9] L. Hong and D. Brzakovic, "Bayesian restoration of image sequences using 3-D Markov random fields,", Proc. Of ICASSP 1989, vol. M1.5, pp. 1413-1416, Glasgow, UK, May 1989

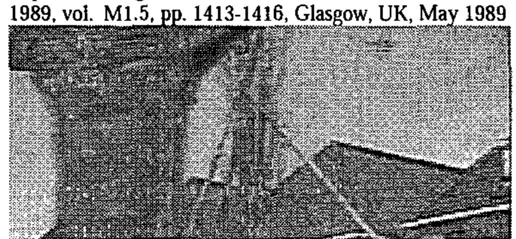

Fig. 2 Original 4th Frame (Part)

Fig. 3 Reconstruct 4th Frame (Part)

Fig.4 Traditional MAP 4th Frame (Part)

Fig. 5 TMAP 4th Frame (Part)

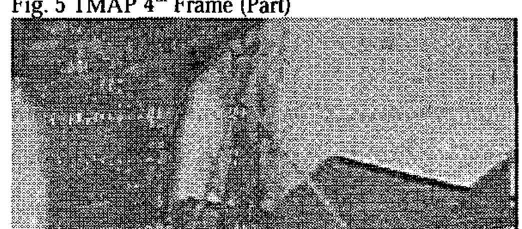

Fig. 6 MC-TMAP 4th Frame (Part)