A MULTI-FRAME BLOCKING ARTIFACT REDUCTION METHOD FOR TRANSFORM-CODED VIDEO

Bahadir K. Gunturk, Yucel Altunbasak, and Russell Mersereau

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, GA 30332-0250

ABSTRACT

A major drawback of block-based still image or video compression methods at low rates are the visible block boundaries that are also known as blocking artifacts. Several methods have been proposed in the literature to reduce these artifacts for video sequences. However, most are simply adaptations of still image blocking artifact reduction methods, which do not exploit temporal information. In this paper, we propose a novel multi-frame blocking artifact reduction method that incorporates temporal information effectively. This method uses the spatial correlations that exist between the successive frames to define constraint sets *at multiple frames* and provides a Projections Onto Convex Sets (POCS) solution. The proposed method operates solely on transform domain (DCT) data, and hence provides a solution that is compatible with the observed video. It does not need to make any spatial smoothness assumptions, which are typical with blocking artifact reduction algorithms for still images.

1. INTRODUCTION

Transform coding is a popular and effective compression method for both still images and video sequences, as evidenced by its widespread use in international media coding standards such as MPEG, H.263 and JPEG. The motion-compensated image (or the image itself) is divided into blocks and each block is independently transformed by a 2-D orthogonal transform to achieve energy compaction. The most commonly used transform is the Discrete Cosine Transform (DCT). After the block transform, the transform coefficients undergo a quantization step. At low bit-rates, the DCT coefficients are coarsely quantized. This coarse quantization along with independent quantization of neighboring blocks gives rise to blocking artifacts—visible block boundaries.

Several blocking artifact reduction methods have been proposed in the literature. Spatial filtering [1, 2], iterative reconstruction techniques [3, 4, 5], and stochastic reconstruction techniques [6, 7, 8] are among the blocking artifact reduction methods that have been proposed for still images. Temporal information adds another dimension to these methods for video sequences. Ironically, this information is not used effectively, or not used at all, for blocking artifact reduction in video. One method that uses temporal information was proposed by Park and Lee [9]. It makes use of motion vectors to extract the blocking semaphores and employs adaptive spatial filtering to remove the artifacts. Another method [10] uses space-varying spatial filtering followed by a motion compensated nonlinear filter.

In this paper, we propose a multi-frame restoration-based method that makes use of the spatial correlations between successive frames effectively. The proposed method constructs convex constraint sets *at each frame within a neighborhood of the frame of interest, using the motion between the frames and the quantization information extracted from the bit stream*. The method is based on the fact that, although the exact value of the quantization noise added to each DCT coefficient is not known, the range within which it lies can be determined from the bit stream [3]. Incorporating the motion between the frames, we can define constraint sets not only at the current frame, but also at each frame within a small neighborhood of the current frame. By projecting the initial "blocky" frame onto these constraint sets successively, we can reconstruct a better estimate of the "original" frame—the one before the quantization step.

In Section II, the constraint sets are defined. The proposed iterative algorithm is explained in Section III. Section IV presents the experimental results. Finally conclusions are provided in Section V.

2. DERIVATION OF CONSTRAINT SETS

In this section we relate the pixel intensities at a particular frame to the DCT coefficients of the neighboring frames. This relation will enable us define constraint sets on an arbitrary video frame in order to reconstruct the original blocking-artifact-free image.

We start with the intensity conservation assumption along the motion trajectories, and then employ MPEG compression stages. Let $f(\mathbf{x}, t)$ denote the intensity of the continuous spatio-temporal video signal at spatial coordinate $\mathbf{x} \equiv [x_1, x_2]$ at time t. Throughout the paper, we will use the vector notation $\mathbf{x}$ and explicit spatial coordinates $[x_1, x_2]$ interchangeably. Pixel intensities of any two video frames can be related to each other through the motion vectors. Denoting $\mathbf{M} \equiv [M_1(\mathbf{x}, t_k; t_j), M_2(\mathbf{x}, t_k; t_j)]$ as the motion mapping between the frames at times t_k and t_j, we can write:

$$f(\mathbf{x}, t_k) = f(\mathbf{x} - \mathbf{M}, t_j), \tag{1}$$

where we dropped the $(\mathbf{x}, t_k; t_j)$ dependency for simplicity. We now proceed by relating the spatially continuous video frame at time t_j to the corresponding discrete frame. Denoting $f(\mathbf{n}, j)$ as the intensity of the j^{th} discrete frame at the coordinate $\mathbf{n} \equiv$

$[n_1, n_2]$, we can write the spatially-continuous reconstruction as:

$$\begin{aligned} f(\mathbf{x}, t_j) &= \left[\sum_{\mathbf{n}} f(\mathbf{n}, j)\delta(\mathbf{x} - \mathbf{n})\right] * h_r(\mathbf{x}) \\ &= \int \sum_{\mathbf{n}} f(\mathbf{n}, j)\delta(\xi - \mathbf{n}) h_r(\mathbf{x} - \xi) d\xi \\ &= \sum_{\mathbf{n}} f(\mathbf{n}, j) h_r(\mathbf{x} - \mathbf{n}), \end{aligned} \tag{2}$$

where $h_r(\mathbf{x})$ is the reconstruction filter. Substituting Equation (2) into Equation (1), we get:

$$f(\mathbf{x}, t_k) = \sum_{\mathbf{n}} f(\mathbf{n}, j) h_r(\mathbf{x} - \mathbf{M} - \mathbf{n}). \tag{3}$$

Since we are only dealing with digital video, we evaluate $f(\mathbf{x}, t_k)$ at integer locations $\mathbf{m} \equiv [m_1, m_2]$. The discrete k^{th} frame is then written as:

$$f(\mathbf{m}, k) = \sum_{\mathbf{n}} f(\mathbf{n}, j) h_r(\mathbf{m} - \mathbf{M} - \mathbf{n}). \tag{4}$$

In order to emphasize the point that $\mathbf{M}$ is a function of frames k and j, corresponding to times t_k and t_j, respectively, we define $h(\mathbf{m}, k; \mathbf{n}, j) \equiv h_r(\mathbf{m} - \mathbf{M} - \mathbf{n})$, and write Equation (4) as:

$$f(\mathbf{m}, k) = \sum_{\mathbf{n}} h(\mathbf{m}, k; \mathbf{n}, j) f(\mathbf{n}, j). \tag{5}$$

Now we model the operations that take place in the process of MPEG compression (*i.e.,* motion compensation, block-DCT calculation, and quantization) to the k^{th} frame. Motion compensation is simply the subtraction of an offset value from $f(\mathbf{m}, k)$. Denoting $f_m(\mathbf{m}, k)$ as the motion compensated frame and $\hat{f}(\mathbf{m}, k)$ as the predicted frame, we write:

$$f_m(\mathbf{m}, k) = f(\mathbf{m}, k) - \hat{f}(\mathbf{m}, k). \tag{6}$$

After taking the 8×8 block-DCTs of the residual image, $f_m(\mathbf{m}, k)$, and quantizing the resulting DCT coefficients $d(\mathbf{m}, k)$, we end up with the quantized DCT coefficients $d_q(\mathbf{m}, k)$:

$$d_q(\mathbf{m}, k) = F(\mathbf{m}, k) - \hat{F}(\mathbf{m}, k) + Q(\mathbf{m}, k), \tag{7}$$

where $F(\mathbf{m}, k)$ and $\hat{F}(\mathbf{m}, k)$ are the block-DCT values of $f(\mathbf{m}, k)$ and $\hat{f}(\mathbf{m}, k)$, respectively, and $Q(\mathbf{m}, k)$ is the quantization noise that is introduced. We can explicitly write the DCT of $f(\mathbf{m}, k)$ as follows:

$$F(\mathbf{m}, k) = \sum_{\mathbf{l}=L(\mathbf{m})}^{L(\mathbf{m})+7} K\left[((\mathbf{m}))_8; \mathbf{l}\right] f(\mathbf{l}, k), \tag{8}$$

where the summation is actually a double summation over $\mathbf{l} \equiv [l_1, l_2]$. The limit function $L(\cdot)$ is defined by $L(m) = 8\lfloor m/8 \rfloor$, $((\cdot))_8$ denotes the $modulo$ 8 operator, and the DCT kernel K is given by:

$$K(\mathbf{m}; \mathbf{l}) = k_{l_1} k_{l_2} \cos\left(\frac{(2m_1 + 1)l_1\pi}{16}\right) \cos\left(\frac{(2m_2 + 1)l_2\pi}{16}\right), \tag{9}$$

with k_{l_1} and k_{l_2} being the normalization constants:

$$k_{l_{1,2}} = \begin{cases} \frac{1}{4\sqrt{2}} & ; l_{1,2} = 0 \\ \frac{1}{4} & ; l_{1,2} \neq 0 \end{cases}. \tag{10}$$

Substituting Equation (5) into Equation (8), and changing the order of summations gives:

$$F(\mathbf{m}, k) = \sum_{\mathbf{n}} \sum_{\mathbf{l}=L(\mathbf{m})}^{L(\mathbf{m})+7} K\left[((\mathbf{m}))_8; \mathbf{l}\right] h(\mathbf{l}, k; \mathbf{n}, j) f(\mathbf{n}, j). \tag{11}$$

Defining

$$h_K(\mathbf{m}, k; \mathbf{n}, j) \equiv \sum_{\mathbf{l}=L(\mathbf{m})}^{L(\mathbf{m})+7} K\left[((\mathbf{m}))_8; \mathbf{l}\right] h(\mathbf{l}, k; \mathbf{n}, j), \tag{12}$$

we can write Equation (11) as:

$$F(\mathbf{m}, k) = \sum_{\mathbf{n}} h_K(\mathbf{m}, k; \mathbf{n}, j) f(\mathbf{n}, j). \tag{13}$$

Incorporating Equation (13), we will repeat Equation (7) since this equation forms the basis of our algorithm:

$$\begin{aligned} d_q(\mathbf{m}, k) = &\sum_{\mathbf{n}} h_K(\mathbf{m}, k; \mathbf{n}, j) f(\mathbf{n}, j) \\ &- \hat{F}(\mathbf{m}, k) + Q(\mathbf{m}, k) \end{aligned} \tag{14}$$

Although the exact value of $Q(\mathbf{m}, k)$ is not known, the range within which the DCT coefficient $d(\mathbf{m}, k)$ lies can be extracted from the MPEG bit stream. Based on this fact we define constraint sets $C(\mathbf{m}, k)$ on frame $f(\mathbf{n}, j)$. Defining $b_l(\mathbf{m}, k)$ and $b_u(\mathbf{m}, k)$ as the lower and upper bounds of the DCT coefficient at spatio-temporal location $(\mathbf{m}, k)$, $C(\mathbf{m}, k)$ can be written as:

$$C(\mathbf{m}, k) = \left\{ \begin{array}{l} f(\mathbf{n}, j) : \left[\sum_{\mathbf{n}} h_K(\mathbf{m}, k; \mathbf{n}, j) f(\mathbf{n}, j) - \hat{F}(\mathbf{m}, k)\right] \\ \in [b_l(\mathbf{m}, k), b_u(\mathbf{m}, k)] \end{array} \right\}. \tag{15}$$

This equation shows how to impose constraint sets on any frame j using the quantization information on another frame k. By projecting the "blocky" frame onto these constraint sets, the blocking artifacts can be reduced significantly. The next section explains how the proposed method works in detail.

3. MULTIFRAME RECONSTRUCTION ALGORITHM

In order to construct the constraint sets on a reference frame j, as in Equation (15), we first compute the transfer function $h_K(\mathbf{m}, k; \mathbf{n}, j)$ between the reference frame j and another arbitrary frame k using Equation (12). As explained in the previous section, this requires accurate motion estimation. Once the transfer function is computed, the reference frame is projected onto the constraint sets using the projection operator $P_{C(\mathbf{m},k)}[\cdot]$ as follows:

$$P_{C(\mathbf{m},k)}\left[f(\mathbf{n}, j)\right] = \left\{ \begin{array}{l} f(\mathbf{n}, j) + \frac{(b_l + \hat{F} - f \bullet h_K) h_K(\mathbf{m}, k; \mathbf{n}, j)}{\sum_{\mathbf{n}} |h_K(\mathbf{m}, k; \mathbf{n}, j)|^2}; f \bullet h_K < b_l + \hat{F} \\ f(\mathbf{n}, j) + \frac{(b_u + \hat{F} - f \bullet h_K) h_K(\mathbf{m}, k; \mathbf{n}, j)}{\sum_{\mathbf{n}} |h_K(\mathbf{m}, k; \mathbf{n}, j)|^2}; f \bullet h_K > b_u + \hat{F} \\ f(\mathbf{n}, j); elsewhere \end{array} \right., \tag{16}$$

where the transfer relation given in Equation (12) is denoted by $f \bullet h_K$, and the dependencies on $(\mathbf{m}, k)$ are dropped for convenience. Since this operation is valid for any k we can construct an arbitrary number of sets $C(\mathbf{m}, k)$ to constraint the solution space.

In order to visualize how the algorithm works, we consider a video sequence with frames having two pixels only. Figure 1 depicts a single frame $f^{(0)}(\mathbf{n}, j)$ represented by a point with pixel intensities given by the distances to the p_1 and p_2 axes. The shaded region shows the constraint set $C(\mathbf{m}, k)$ bounded by the hyperplanes $\hat{F}(\mathbf{m}, k) + b_l(\mathbf{m}, k)$ and $\hat{F}(\mathbf{m}, k) + b_u(\mathbf{m}, k)$.

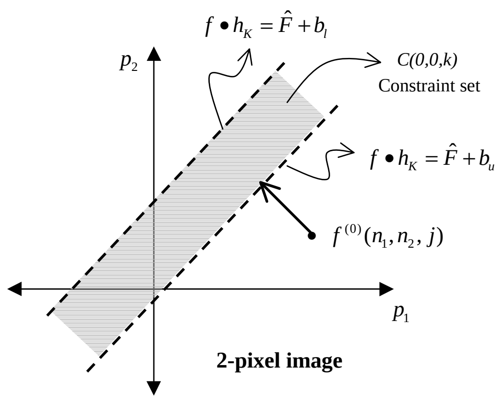

Fig. 1. One constraint set

This constraint set is defined simply by using the quantization bounds $[b_l(\mathbf{m}), b_u(\mathbf{m})]$ of one of the DCT coefficients as explained in the previous section. Projection of an image onto this constraint set amounts to simply finding the closest point on $C(\mathbf{m}, k)$ using Equation (16).

Although this projection is likely to reduce the blocking artifacts, it does not guarantee a significant improvement since the "original" blocking-artifact-free image could be anywhere in the shaded region. Defining another constraint set could improve the quality significantly. As depicted in Figure 2, the second constraint set defined with help of the neighboring frame $(k + 1)$ reduces the region where the "original" image lies. Projecting the initial frame onto these convex sets successively produces a better result. By using additional frames we can impose more constraint sets onto the reconstructed frame and reduce the blocking artifacts further.

It should be noted that the performance of the method depends on the accuracy of the motion estimation. Using inaccurate motion estimates means imposing "wrong" constraints on the reconstructed frame. In our experiments we used the hierarchical block matching (HBM) technique of Bierling [11], to compute the motion vectors from the previous enhanced frame and only used the constraint sets defined at those locations where the motion-compensated frame difference was below a predetermined threshold. We did *not* use the motion vectors that were included in the MPEG bit stream.

4. RESULTS

The proposed algorithm is successfully applied to several real video sequences including the "Susie" and "Flower Garden" sequences. The "Susie" sequence has smooth textures and the motion is relatively slow. On the other hand, the "Flower Garden" sequence has high spatial frequency components and the motion is much faster.

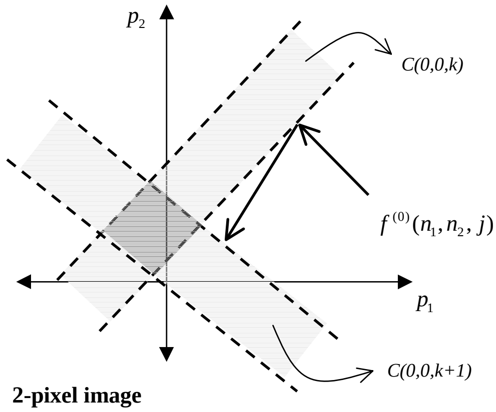

Fig. 2. Two constraint sets

Figure 3 shows a frame from the "Susie" MPEG-1 compressed at 112 kbits/s. Figures 4 and 5 show the reconstructed frames by the proposed algorithm using 2 and 4 frames, respectively. Figure 6 shows a frame from the "Flower Garden" sequence encoded at 112 kbits/s. The effectiveness of the proposed algorithm for the "Susie" sequence is clearly observed in Figures 4 and 5. The "Flower Garden" sequence is zoomed in to show its performance. It is seen in Figure 7 that the algorithm removed the blocking artifacts without smoothing out the frame.

5. DISCUSSION

The proposed multi-frame blocking artifact reduction method exploits temporal information by means of the motion between frames and the quantization bounds available in the video bit stream. It is a general framework in the sense that additional constraint sets could be incorporated into the model easily. However, the method requires accurate motion estimation and has high computational complexity. This is not a significant drawback for offline applications where video quality is the main concern. It could be also used in real time applications by implementing the algorithm as part of a hardware solution.

6. REFERENCES

[1] H.C.Reeves and J.S.Lim, "Reduction of blocking effects in image coding," *Opt. Eng.*, vol. 23, pp. 34–37, 1984.

[2] G.Ramamurthi and A.Gersho, "Nonlinear space-variant postprocessing of block coded images," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-34, pp. 1258–1269, 1986.

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

[4] Y.Yang, N.P.Galatsanos, and A.K.Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Sys. Video Technology*, vol. 3, pp. 421–432, 1993.

[5] H.Paek, R.Kim, and S.Lee, "On the pocs-based postprocessing technique to reduce blocking artifacts in transform coded

Fig. 3. “Susie” MPEG-1 compressed (112 kbits/s)

Fig. 4. Reconstructed using 2 frames

images,” *IEEE Trans. Circuits Sys. Video Technology*, vol. 8, pp. 358–367, 1998.

[6] R.L.Stevenson, “Reduction of coding artifacts in transform image coding,” *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 401–404, 1993.

[7] J.Luo, C.W.Chen, K.J.Parker, and T.S.Huang, “Artifact reduction in low bit-rate dct-based image compression,” *IEEE Trans. Image Processing*, vol. 5, pp. 1363–1368, 1996.

[8] T.Ozcelik, J.C.Brailean, and A.K.Katsaggelos, “Image and video compression algorithms based on recovery techniques using mean field annealing,” *Proc. IEEE*, vol. 83, pp. 304–316, 1995.

[9] H.W.Park and Y.L.Lee, “A postprocessing method for reducing quantization effects in low bit-rate moving picture coding,” *IEEE Trans. Circuits Sys. Video Technology*, vol. 9, pp. 161–171, 1999.

[10] C.Derviaux, F.-X.Coudoux, M.G.Gazalet, and P.Corlay, “Blocking artifact reduction of dct coded image sequences using visually adaptive postprocessing,” *International Conference on Image Processing’96*, vol. 2, pp. 5–8, 1996.

[11] M.Bierling, “Displacement estimation by hierarchical block-matching,” *Proc. SPIE Visual Communications and Image Processing ’88*, pp. 942–951, 1988.

Fig. 5. Reconstructed using 4 frames

Fig. 6. “Flower Garden” MPEG-1 compressed (112 kbits/s)

Fig. 7. Reconstructed using 4 frames