

ADVANCED MOTION COMPENSATION TECHNIQUES FOR BLOCKING ARTIFACTS REDUCTION IN 3-D VIDEO CODING SYSTEMS

Cho-Chun Cheng[†], Wen-Liang Hwang[†], Zuowei Shen[‡], and Tao Xia^{}*

[†] Institute of Information Science, Academia Sinica, Taipei, Taiwan

[‡] Department of Mathematics, National University of Singapore, Singapore

^{*} Center for Wavelet, Approximation and Information Processing (CWAIP), National University of Singapore, Singapore

ABSTRACT

This paper describes a new 3-D framework for the construction of a hybrid motion compensation model to reduce blocking artifacts in a highly scalable video coding system. Previous works have focused primarily on using control grid interpolation (CGI) or overlapped block motion compensation (OBMC) to reduce blocking artifacts. However, both methods generate distinct side effects, such as PSNR degradation or poor de-blocking results, during the de-blocking process. The main objective of our proposed model is to achieve better visual quality and maintain satisfactory coding efficiency simultaneously by integrating the advantages of CGI and OBMC. The introduction of the dual mode selection between CGI and OBMC provides a more flexible mechanism to adjust our objective function and thereby achieves higher objective and subjective coding gains. Experimental results show that our proposed model achieves improved coding efficiency in terms of both PSNR and visual quality.

1. INTRODUCTION

Recent advances in motion-compensated temporal filtering (MCTF), which are based on discrete wavelet transform (DWT), make it possible to implement highly efficient video codecs in a 3-D structure with spatial, temporal, and SNR scalability [2, 7, 8, 11]. As implemented in most 3-D video coding systems, the block matching algorithm (BMA) is still the predominant method for reducing temporal redundancies due to its relatively low computational complexity. However, it is widely recognized that the discontinuous motion field generated by BMA may cause block-wise varying coefficient statistics and result in unpleasant blocking artifacts that degrade the visual quality of a video [1]. Previous research has suggested that many advanced motion compensation techniques, such as CGI or OBMC, can effectively improve the visual quality of a decompressed video. An early attempt at 3-D subband coding using CGI was reported in [7]. Meanwhile, Hanke et al. [3] reduced the blocking artifacts by introducing a smooth transition filtering method to blur the areas where the motion is inconsistent, and Wu et al. [12] further extended OBMC to the 3-D video coder.

In this work, the temporal filtering schemes based on advanced motion compensation techniques are described first, including CGI and OBMC. From the preliminary coding results of 3-D video coding systems based on CGI and OBMC, we observe that both methods generate distinct side effects during the de-blocking process. For CGI, although the decompressed sequences are free of blocking artifacts, undesirable blurred edges are often perceived due to

the stronger smoothing effect resulting from a completely smooth motion field. In contrast, poor de-blocking results often appear in those areas with high motion activities after applying OBMC. In view of these problems, we have developed a novel motion compensation model (OBCG) that can integrate the advantages of CGI and OBMC to achieve better visual quality, and maintain satisfactory coding efficiency simultaneously. Moreover, because OBCG creates inter-dependency relations between neighboring motion vectors (MVs), an iterative motion refinement algorithm based on dynamic programming is proposed for estimating the dependent motion field.

This paper is organized as follows. In Section 2, the frameworks of temporal filtering based on CGI and OBMC are introduced. In Section 3, a novel motion compensation model is presented, and our proposed iterative motion refinement algorithm is formulated. The simulation results are reported and compared with those of conventional methods objectively and subjectively in Section 4. Finally, in Section 5, we present our conclusions.

2. ALTERNATIVE MOTION COMPENSATED TEMPORAL FILTERING

Under a conventional MCTF framework, motion compensation is implemented between a pair of successive frames $\mathbf{F} = (f_1, f_2)$. The highpass subband frame H corresponds to the temporal location of frame f_1 , while the lowpass subband has the same temporal position as f_2 . During motion compensation, f_1 is partitioned into a group of non-overlapped square blocks, $\mathbf{P} = (P_{0,0}, \dots, P_{N_1-1, N_2-1})$. Each block then finds its corresponding block with the least distortion measurement in f_2 during the motion estimation operation. Let $f_1(\vec{s})$ and $f_2(\vec{s})$ denote the pixel intensity at spatial index $\vec{s} = (x, y)$ of frames f_1 and f_2 respectively. Also, let $\hat{f}_1(\vec{s})$ denote the predicted pixel intensity at \vec{s} . For the Haar wavelet, the temporal filtering scheme is given by

$$H(\vec{s}) = \frac{1}{\sqrt{2}}[f_1(\vec{s}) - \hat{f}_1^{MC}(\vec{s})], \quad \vec{s} \in P_{i,j} \quad (1)$$

$$L(\vec{s}) = \sqrt{2}f_2(\vec{s}) + H(\vec{s} + \vec{v}^{MC}), \quad \vec{s} \in P_{i,j} \quad (2)$$

The MC in Equations 1 and 2 represents the motion compensation technique adopted in temporal filtering. Although CGI and OBMC both try to eliminate unexpected discontinuities in motion-compensated frames, they adopt different concepts. CGI produces a smooth motion field in order to preserve continuity and connectivity in the predicted frame. As shown in Fig. 1, the vertices of

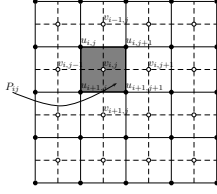


Fig. 1. Configuration of CGI and OBMC. The black nodes on the vertices of each block represent the control points in CGI. The white node in the center of each block represents the MV of each block in OBMC.

each block are chosen as control points, with each point assigned to an MV in the set $\mathbf{U} = (\vec{u}_{0,0}, \dots, \vec{u}_{N_1,N_2})$. Let $\vec{u}_{i+p,j+q}$ be the p, q th neighboring MV associated with $P_{i,j}$. For each interior pixel in $P_{i,j}$, CGI computes an individual MV by interpolating its neighboring control points. $w_{p,q}(\vec{s})$ denotes the scalar interpolation coefficient, and depends on the desired contribution of the p, q th control point to \vec{s} . Therefore, the temporal filtering scheme of CGI can be formulated by replacing $\hat{f}_1^{MC}(\vec{s})$ and \vec{v}^{MC} in Equations 1 and 2 with $\hat{f}_1^{CGI}(\vec{s})$ and \vec{v}^{CGI} respectively, which are given as

$$\begin{aligned} \hat{f}_1^{CGI}(\vec{s}) &= f_2(\vec{s} - \vec{v}^{CGI}) \\ \vec{v}^{CGI} &= \sum_{p=0}^1 \sum_{q=0}^1 w_{p,q}(\vec{s}) \vec{u}_{i+p,j+q}, \end{aligned}$$

where

$$\sum_{p=0}^1 \sum_{q=0}^1 w_{p,q}(\vec{s}) = 1, \quad \text{and} \quad \vec{s} \in P_{i,j}.$$

On the other hand, OBMC reduces the abrupt changes of pixel intensities between neighboring blocks by using a weighted sum at block boundaries. As shown in Fig. 1, instead of assigning four MVs to every vertex in the CGI scheme, OBMC assigns a single MV to each block, and the MV set is denoted as $\mathbf{V} = (\vec{v}_{0,0}, \dots, \vec{v}_{N_1-1,N_2-1})$. Let $\alpha_{p,q}(\vec{s})$ denote the weighting factor for the p, q th neighboring MV at \vec{s} . Then, the temporal filtering scheme of OBMC can be formulated by replacing $\hat{f}_1^{MC}(\vec{s})$ and \vec{v}^{MC} in Equations 1 and 2 with $\hat{f}_1^{OBMC}(\vec{s})$ and \vec{v}^{OBMC} respectively, which are given as

$$\begin{aligned} \hat{f}_1^{OBMC}(\vec{s}) &= \sum_{p=-1}^1 \sum_{q=-1}^1 \alpha_{p,q}(\vec{s}) f_2(\vec{s} - \vec{v}_{i+p,j+q}) \\ \vec{v}^{OBMC} &= \vec{v}_{i,j}, \end{aligned}$$

where

$$\sum_{p=-1}^1 \sum_{q=-1}^1 \alpha_{p,q}(\vec{s}) = 1, \quad |p| + |q| < 2, \quad \text{and} \quad \vec{s} \in P_{i,j}.$$

3. A NOVEL MOTION COMPENSATION MODEL

In this work, we develop a novel motion compensation model that integrates the advantages of CGI and OBMC to achieve better vi-

sual quality and maintain satisfactory coding efficiency simultaneously. The major operation involved in the de-blocking process is smoothing. CGI, which is regarded as a stronger smoothing operation, provides a completely smooth motion field that includes those areas around and within a block. On the other hand, OBMC, which is considered as a moderate smoothing operation, only sums up the weighted intensities around the block boundaries. To select a proper mode between CGI and OBMC, local image characteristics must be examined. When the motion activities between neighboring blocks are relatively high, the stronger smoothing operation provided by CGI inside a block, as well as at its boundaries, has a better de-blocking effect. In contrast, when the motion activities are relatively low, using OBMC to smooth a few pixels around a block's boundary is enough to achieve the desired de-blocking effect and retain a high PSNR performance without losing most of the edge properties. In our proposed coding scheme, the motion activities between neighboring areas are examined by using the following measurement:

$$\phi_{P_{i,j}}(\mathbf{v}_{P_{i,j}}) = \begin{cases} 1, & \text{if } \max |(\mathbf{v}_{P_{i,j}} - \vec{v}_{i,j})| \geq \text{Threshold}, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where

$$\mathbf{v}_{P_{i,j}} = \{\vec{v}_{i+p,j+q} \mid -1 \leq p \leq 1; -1 \leq q \leq 1; p, q \in \mathbb{Z}\}.$$

$\mathbf{v}_{P_{i,j}}$ represents the set of neighboring MVs associated with $P_{i,j}$. If the difference between the target block's MV ($\vec{v}_{i,j}$) and any of its neighboring block's MV ($\vec{v}_{i+p,j+q}$) is larger than a certain threshold, $\phi_{P_{i,j}}$ is set to 1, and CGI is adopted in the target block. Otherwise, $\phi_{P_{i,j}}$ is set to 0 and OBMC is adopted in the target block.

OBCG utilizes most of the settings of OBMC. However, additional motion compensation choices are added to the original OBMC framework. The major feature of our proposed motion compensation model is that it allows each block to be compensated by either CGI or OBMC in accordance with the surrounding motion activities. Therefore, the dependent motion estimation optimization problem of OBCG can be formulated as

$$\begin{aligned} \min J_{OBCG}(\mathbf{V}) \\ = \min_{\mathbf{V}} \left\{ \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} [D_{OBMC}(P_{i,j}, \mathbf{v}_{P_{i,j}}) \cdot (1 - \phi_{P_{i,j}}) \right. \\ \left. + D_{CGI}(P_{i,j}, \mathbf{v}_{P_{i,j}}) \cdot \phi_{P_{i,j}}] \right\}, \end{aligned} \quad (4)$$

where

$$D_{CGI}(P_{i,j}, \mathbf{v}_{P_{i,j}}) = \frac{1}{N_p} \sum_{\vec{s} \in P_{i,j}} |f_1(\vec{s}) - \hat{f}_1^{CGI}(\vec{s})|$$

$$D_{OBMC}(P_{i,j}, \mathbf{v}_{P_{i,j}}) = \frac{1}{N_p} \sum_{\vec{s} \in P_{i,j}} |f_1(\vec{s}) - \hat{f}_1^{OBMC}(\vec{s})|.$$

$D_{CGI}(P_{i,j}, \mathbf{v}_{P_{i,j}})$ and $D_{OBMC}(P_{i,j}, \mathbf{v}_{P_{i,j}})$ denote the prediction error of block $P_{i,j}$ using CGI and OBMC respectively, and N_p represents the number of pixels in $P_{i,j}$. It is worth noting that the motion field of the proposed model can not be solved by optimizing each block's prediction error independently, because each

adjacent block pair shares more than one MV. Optimizing each block's prediction error independently leads to considerable performance degradation. Therefore, an iterative motion refinement scheme based on dynamic programming is proposed to solve the dependent optimization problem. First, the original 2-D dependent motion estimation optimization problem of OBCG is decomposed into a series of 1-D problems; then, the iterative motion refinement algorithm of OBCG is performed as follows:

Step 0. Initialize the MV set \mathbf{V} of each block using BMA, and split \mathbf{V} into two subsets $\{\mathbf{vo}, \mathbf{ve}\}$, where \mathbf{vo} and \mathbf{ve} denote the MV subset of odd rows and even rows respectively. Moreover, for each $i \in \{i = 1 + 2x | 0 \leq x < \frac{N_1-1}{2}, x \in \mathbb{Z}\}$, let \mathbf{vo}_i denote the MV subset of the i th odd row, and for each $i \in \{i = 2x | 0 \leq x \leq \frac{N_1-1}{2}, x \in \mathbb{Z}\}$, let \mathbf{ve}_i denote the MV subset of the i th even row.

Step 1. Set optimization cycle $k = 0$ to $\mathbf{vo}^k = \{\mathbf{vo}_1^k, \mathbf{vo}_3^k, \dots\}$ and $\mathbf{ve}^k = \{\mathbf{ve}_0^k, \mathbf{ve}_2^k, \dots\}$ for the initial condition.

Step 2. Use \mathbf{V} to calculate ϕ , which classifies the motion compensation technique adopted by each block.

Step 3. Odd row processing:

- Fix \mathbf{ve}^k to provide the neighboring condition for odd row optimization, and apply Equation 5, shown below, for each i th row.

$$\mathbf{vo}_i^{k+1} = \arg \min_{\mathbf{vo}_i^{k+1}} \left(\sum_{j=0}^{N_2-1} D_{OBCG}(A_{i,j}, \mathbf{v}_{A_{i,j}}) \right) \quad (5)$$

for $i \in \{i = 1, 3, \dots\}$.

- Update $\mathbf{V} = \{\mathbf{vo}^{k+1}, \mathbf{ve}^k\}$, and increase $k = k + 1$.

Step 4. Even row processing:

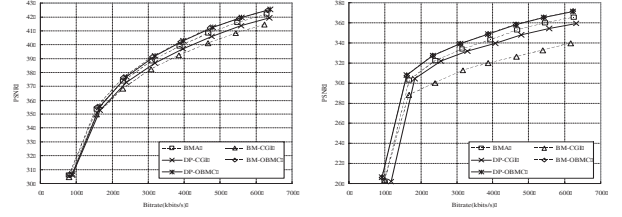
- Fix \mathbf{vo}^k , and apply even row optimization to find \mathbf{ve}^{k+1} by replacing the \mathbf{vo} with \mathbf{ve} in Equation 5, for $i \in \{i = 0, 2, \dots\}$.
- Update $\mathbf{V} = \{\mathbf{vo}^k, \mathbf{ve}^{k+1}\}$, and increase $k = k + 1$.

Step 5. Iterate Steps 2-4 until J_{OBCG} converges to a minimum.

When calculating Equation 5 for the possible paths in the trellis, instead of calculating the prediction error of each non-overlapped $P_{i,j}$, different corresponding areas $A_{i,j}$ should be considered. The corresponding areas are only those two overlapped regions affected while refining consecutive MVs. $\mathbf{v}_{A_{i,j}}$ represents the set of neighboring MVs needed to compensate for $A_{i,j}$. Note that the same classification ϕ is adopted at the decoder side, so our proposed dual mode selection model can be perfectly reconstructed without sending any additional overhead.

4. PERFORMANCE EVALUATION

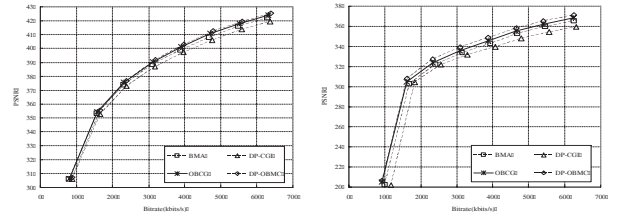
In this section, we perform comparative studies of alternative 3-D subband coding schemes in terms of both objective and subjective evaluations. Conventional motion compensation based on BMA is implemented using full-search with integer pel accuracy, and a $[-16, 15]$ search range is set for both the horizontal and vertical dimensions. The block size is set at 16×16 . With regard to the proposed motion compensation schemes based on OBCG,



(a) CIF, Mother and Daughter

(b) CIF, Dancer

Fig. 2. The PSNR performance of BMA, BM-CGI, DP-CGI, BM-OBMC, and DP-OBMC at various bit rates.



(a) CIF, Mother and Daughter

(b) CIF, Dancer

Fig. 3. The PSNR performance of BMA, DP-CGI, DP-OBMC, and OBCG at various bit rates.

the initial motion field for each block is estimated by a full-search BMA. The motion field is then updated using the iterative motion refinement scheme proposed in this work. The bilinear transform is adopted to compute the scalar interpolation coefficients w of CGI, and the 24×24 weighting window proposed in [9] is chosen as the α for our OBMC. First, all of the video signals are applied with three levels of temporal filtering using the Haar filter, and a 3-level spatial DWT with the $9 - 7$ filter is then applied. Finally, the 3-D SPIHT (Set Partitioning in Hierarchical Trees) [5] algorithm is used to encode the wavelet coefficients of each subband to generate a scalable bitstream.

Because both CGI and OBMC also create inter-dependency relations between neighboring MVs, the proposed iterative motion refinement algorithm is adopted to improve the coding efficiency. As shown in Fig. 2, the improvement achieved by using the proposed iterative motion refinement scheme based on dynamic programming for CGI (DP-CGI) is significant. The scheme can usually achieve a 0.4 - 1.9 dB gain over the result of CGI, which uses the motion field obtained by BMA(BM-CGI). For OBMC, both BM-OBMC and DP-OBMC achieve considerable performance improvements over the conventional coding scheme based on BMA. The coding gain is about 0.3 - 0.7 dB, on average. The PSNR performance of OBCG for different video sequences at various bit rates is shown in Fig. 3. Because DP-CGI and DP-OBMC represent two special cases of OBCG, the PSNR performance of OBCG usually lies between these two methods. However, OBCG still maintains a 0.3 - 0.4 dB gain over the conventional coding scheme based on BMA.

For subjective evaluation, the decompressed frames in Fig. 4 show the visual quality of different schemes. It is clear that our proposed 3-D video coding scheme adapts to various frame features successfully, and removes noticeable blocking artifacts without degrading details. To compare it with conventional de-



Fig. 4. Perceptual quality: (a),(b) frame 76 of CIF Mother and Daughter at 95 kbits/sec; and (c),(d) frame 50 of CIF Dancer at 200 kbits/sec.

blocking techniques, more sophisticated subjective tests were conducted according to the requirements of [4], and the DSIS (Double Stimulus Impairment Scale) was evaluated. While evaluating DSIS, the testers were required to vote using a five-grade impairment scale from 5 (imperceptible) to 1 (very annoying) after being presented with the test materials. The comparative results are shown in Fig. 5. We observe that better subjective performances are consistently maintained through different types of video sequences. Although the coding results of DP-CGI and DP-OBMC show the performance bounds of our proposed model in terms of PSNR, the results of subjective evaluation show that the perceptual quality of our proposed model is not necessarily between them. The perceptual quality of most video sequences compressed by OBCG can also outperform both DP-CGI and DP-OBMC.

5. CONCLUSIONS

We propose a novel motion compensation model (OBCG) that integrates the advantages of CGI and OBMC. Moreover, an iterative motion refinement algorithm is proposed for estimating the dependent motion field and improving the coding efficiency of OBCG. Experiment results show that our proposed model adapts to various frame features successfully, and leads to improved PSNR and better perceptual quality over the BMA model. Compared to conventional de-blocking techniques, although DP-CGI and DP-OBMC have shown the performance bounds of our proposed model in terms of PSNR, the results of subjective evaluation show that the perceptual quality of our proposed model can outperform them in most cases.

6. REFERENCES

[1] M.C. Chen and A.N. Willson, Jr., "Motion-vector optimization of control grid interpolation and overlapped block motion

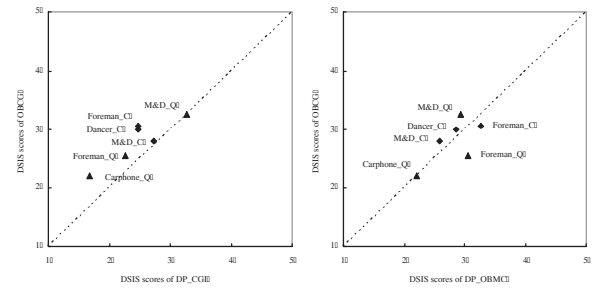


Fig. 5. Perceptual comparison of 3-D video coding systems using DP-CGI, DP-OBMC, and OBCG for various video sequences at low bit rates. (Q: QCIF, and C: CIF).

compensation using iterated dynamic programming," *IEEE Trans. on Image Processing*, vol. 9, no. 7, pp. 1145-1157, July 2000.

- [2] S.-J. Choi and J.W. Woods, "Motion-compensated 3-D sub-band coding of video," *IEEE Trans. on Image Processing*, vol. 8, no. 2, pp. 155-167, February 1999.
- [3] K. Hanke, T. Ruster, and J.-R. Ohm, "Motion-compensated 3D video coding using smooth transitions," *Proc. SPIE Visual Communications Image Processing*, vol. 5022, pp. 933-940, 2003.
- [4] ITU-R, "ITU-R Recommendation BT.500-11: Methodology for the subjective assessment of the quality of television pictures," *The International Telecommunication Union*, 2002.
- [5] B.-J. Kim, Z. Xiong, and W.A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374-1387, December 2000.
- [6] M.T. Orchard and G.J. Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach," *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 693-699, September 1994.
- [7] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. on Image Processing*, vol. 4, no. 3, pp. 339-367, 1994.
- [8] A. Secker and D. Taubman, "Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation," *Proc. IEEE Int. Conf. Image Processing*, pp. 1029-1032, 2001.
- [9] K. Shen and E.J. Delp, "Wavelet based rate scalable video compression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 109-122, February 1999.
- [10] G.J. Sullivan and R.L. Baker, "Motion compensation for video compression using control grid interpolation," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, pp. 2713-2716, 1991.
- [11] J.W. Woods and P. Chen, "Improved MC-EZBC with quarter-pixel motion vectors," *ISO/IEC JTC1/SC29/WG11*, MPEG doc., M8366, Fairfax, May 2002.
- [12] Y. Wu, R.A. Cohen, and J.W. Woods, "An overlapped block motion estimation for MC-EZBC," *ISO/IEC JTC1/SC29/WG11*, MPEG doc., M10158, Brisbane, October 2003.