A New Update Step for Reduction of PSNR Fluctuations in Motion-Compensated Lifted Wavelet Video Coding

Aditya Mavlankar, Sang-eun Han, Chuo-Ling Chang and Bernd Girod
Information Systems Laboratory
Department of Electrical Engineering, Stanford University
Stanford, CA, 94305, U.S.A.
{maditya, sehan, chuoling, bgirod}@stanford.edu

Abstract—In wavelet video coding, due to motion-compensation entwined in the temporal wavelet transform, the distortion in the temporal subbands propagates in an uneven manner into the reconstructed frames of a group of pictures. This leads to fluctuation of reconstruction quality in time, which can be visually displeasing. We propose a new update step which is derived by taking into account a Lagrangian term for even distribution of distortion among reconstructed frames in addition to minimizing total distortion. Additionally some heuristics for the implementation of this new update step are proposed. Experimental results show reduction of quality fluctuation compared to the conventional update step.

Index Terms—wavelet; scalable video coding; motion compensated lifted;

I. INTRODUCTION

The recent approach of motion-compensated lifted wavelet transform for video compression provides a signal decomposition ([1], [2], [3], [4]) which is invertible and can employ arbitrary motion compensation (MC) techniques. This approach provides inherent scalability of bit-rate, spatial resolution and temporal resolution while intending to approach the coding efficiency of traditional hybrid coding schemes. The analysis in [5] implies that wavelet video coding might reach the same or even better coding efficiency than traditional hybrid coding schemes and at the same time provide the additional benefits of scalability. Due to their DPCM structure, hybrid video codecs have to sacrifice some coding efficiency in order to provide scalability.

The pyramidal type temporal decomposition of a group of 4 pictures using MC lifted Haar wavelet as the basic decomposition unit is shown in Fig. 1. MC lifted Haar wavelet decomposition with input frames X and Y is shown in Fig. 2. Note that transform in Fig. 2 is never orthonormal since frame L should be scaled by $\sqrt{2}$ and frame H by $\frac{1}{\sqrt{2}}$. Even after including these scaling factors, due to the inclusion of motion compensation, the temporal wavelet transform is no longer orthonormal, and typically frame Y is penalized more than frame X when quantization of the temporal subbands is performed [6]. Since the basic decomposition unit is used in a pyramidal fashion within a group of pictures (GOP) this leads to a PSNR fluctuation pattern. In Section II we derive the new update step, which takes into account a Lagrangian term for even distribution of distortion in addition to minimizing total distortion in the reconstructed frames. Section III proposes some heuristics for implementation. Section IV shows experimental results.

This work was supported, in part, by the Max Planck Center for Visual Computing and Communication.
from the point of view of minimizing total distortion in the
directly stated here:
\[ E \]

The term weighted by the Lagrangian multiplier
The distortion terms in the above equation can be expressed
where
\[ \alpha \]

denotes trace. It is assumed that
\[ \lambda \]

The expected total distortion in the reconstructed frames
We formulate an extended cost function to be minimized as
\[ \Delta = \text{zero-mean uncorrelated stochastic processes. Similarly} \]
\[ \lambda \]

The new update step is obtained by setting the derivative of.
\[ \alpha \]

The second pair consists of frame number 1 and number 5.
\[ \lambda \]

Equation (2) entails a large amount of complexity and storage
for implementation. Hence we propose some heuristics which
can either exactly or approximately implement (2) for a given
value of \( \alpha \).

A. Scaling the Conventional Update Step

By performing an eigenvalue decomposition of \( P^T P \) and
denoting the eigenvalues by \( \lambda_1, \lambda_2, \ldots, \lambda_n \) in both (1) and (2)
we can write:
\[ U_{opt} = T^{-1} \text{diag} \left( \frac{1}{1 + \lambda_1}, \frac{1}{1 + \lambda_2}, \ldots, \frac{1}{1 + \lambda_n} \right) T P^T \] (3)
\[ U_{m} = T^{-1} \text{diag} \left( \frac{1}{\alpha + \lambda_1}, \frac{1}{\alpha + \lambda_2}, \ldots, \frac{1}{\alpha + \lambda_n} \right) T P^T \] (4)

All the \( \lambda_i \) above are positive and for simple motion are quite
close to each other. For the extreme case of all \( \lambda_i \) constant, (4) is just a scaled version of (3). But from [7] we know that \( U_{opt} \)
is very close to the conventional update step \( U_{conv} \). This leads us
to our first heuristic: \( U_m \approx \beta U_{conv} \), i.e. a scaled version of
the conventional update step.

B. Modified Barbell Update

Barbell lifting [8] implies the inversion of a many-to-many
mapping. The weights used in this inversion are the same as
the pixel-connection weights in the forward mapping. This
idea suffers from the fact that if a pixel in frame \( X \) is M-
connected then it receives an update component from several
pixels in the update step and this might lead to inappropriate
amounts of energy. This problem is mitigated by using the
attenuation factor \( \frac{1}{M+2} \) in addition to the weight of bilinear
interpolation which connects the particular pixel pair from
frame \( X \) and frame \( Y \). The attenuation factor \( \frac{1}{M+2} \) is a
characteristic of the pixel in frame \( X \), while the weight is a
characteristic of the connection of the pixel-pair. Fig. 3 shows
an example with half-pel MC. In case of full-pel MC the pixel-
pair connection weight is always 1 for any connection and
modified Barbell update implements the solution (2) exactly
for any given value of \( \alpha \). This jibes with the implementation
rule for \( U_{opt} \) for full-pel MC given in [7] which is stated
as follows: if a pixel in frame \( X \) is M-connected, all the
connected pixels in the highband \( H \) are added to the pixel
with a weight of \( \frac{1}{M+1} \). 1-connected pixels are included as the
special case \( M = 1 \). If a pixel in frame \( X \) is unconnected then
it is simply copied to the respective position in the lowband
\( L \). The exact implementation of \( U_m \) for full-pel MC can now
be looked upon as the more general case of the rule in [7]
with \( \frac{1}{M+2} \) for any given \( \alpha \).

Fig. 4 shows the reconstruction PSNR for 2 different pairs
of frames, \( (X_1, Y_1) \) and \( (X_2, Y_2) \). The first pair consists of
frame number 1 and number 2 of the Foreman sequence (CIF).
The second pair consists of frame number 1 and number 5.

III. Heuristic Rules for Implementing the
Modified Update Step

Equation (2) entails a large amount of complexity and storage
for implementation. Hence we propose some heuristics which
can either exactly or approximately implement (2) for a given
value of \( \alpha \).
The two frames in the second pair are further apart temporally compared to the two frames in the first pair. In this case there are more unconnected pixels and also the efficiency of the motion compensation is lower. Without any consideration for reducing fluctuations ($\alpha = 1$) frame $Y$ is penalized more than frame $X$. As we reduce $\alpha$ we see that the distortion in the 2 frames gets closer. In this particular experiment, for the first pair the optimal value of $\alpha$ is around 0.76 and for the second pair it is around 0.62. If $\alpha$ is reduced below these values then the distortion pattern gets reversed and frame $X$ gets more distortion than frame $Y$. Notice that with more unconnected pixels in general we find the required balance for a lower value of $\alpha$. In this experiment we added exactly the same noise, depicting quantization noise of the temporal subbands, for every trial with a different $\alpha$. The noise added in the two temporal subbands $L$ and $H$ is zero-mean with no cross-correlation but having the same variance. This additive noise chosen from a uniform distribution is referred to as quantization noise in the rest of the paper. In this experiment we also see that there is no significant drop in the average PSNR. This is also observed when the experiment is extended to a larger GOP size, i.e. more temporal decomposition levels.

IV. PERFORMANCE OF THE HEURISTICS

All the results in this section are plotted for 8 frames in a GOP which amounts to 3 levels of temporal decomposition with the Haar wavelet. The quantization noise added in the 8 temporal subbands is zero-mean with no cross-correlation but having the same variance. Comparison of various schemes is performed by adding exactly the same quantization noise. Fig. 5 shows the results for quarter-pel MC and scaling the conventional update step. The scaling $\beta$ selected here is determined empirically from a few trials and is not guaranteed to be the global optimum. Also a constant value of $\beta$ is used for the entire sequence whereas ideally $\beta$ should be optimized for every single decomposition unit. Even while searching for the optimal value of a parameter (like $\beta$) for any scheme, the same quantization noise is added for every trial. It can be seen that the range of fluctuations is reduced by half.

Fig. 6 shows the results for quarter-pel MC and modified Barbell update. In this case a different value of $\alpha$ is used for every temporal level, though it is kept constant for the entire sequence. These values are also obtained from a few trials with the same sequence. This heuristic performs better than scaling
which leads to fluctuation of PSNR of the reconstructed frames. To mitigate this problem we design a new update step as well as propose some heuristics for implementation. This can be looked upon as a solution which is employed early during the temporal decomposition, as compared to other approaches which do not alter the temporal decomposition but instead change the rate allocation among the temporal subbands to mitigate the fluctuation problem as much as possible. We believe that these two solutions should be combined. I.e. the rate allocation among the temporal subbands following our new update step should not only aim to minimize total distortion in a group of pictures but instead minimize the maximum distortion in the reconstructed frames of the group of pictures. This requires the modeling of the distortion propagation from the temporal subbands into the reconstructed frames in a flexible way for this new update step with any given $\alpha$. Thus we propose to tackle the fluctuation problem in two consecutive stages, temporal decomposition and spatial encoding of temporal subbands. This will be part of future research.

V. CONCLUSION

Due to motion compensation in the lifted temporal wavelet transform, orthonormality is sacrificed for better decorrelation

REFERENCES