# ENERGY DISTRIBUTED UPDATE STEPS(EDU) IN LIFTING BASED MOTION COMPENSATED VIDEO CODING

*Bo Feng*[*1], *Jizheng Xu*[2], *Feng Wu*[2], *Shiqiang Yang*[1], *Shipeng Li*[2]
[1] Department of Computer Science and Technology, Tsinghua University, Beijing, 100084
[2] Microsoft Research Asia, Beijing, 100080, P. R. CHINA

## ABSTRACT

*Subband video coding is an elegant scheme to fulfill high performance scalable video coding. In this paper, a new update scheme, energy distributed update steps(EDU), is proposed for the temporal transform in lifting based motion compensated video coding. The idea is to update where predict is made by distributing high-pass signals to the low-pass frame. The scheme avoids complex and inaccurate inversion of the motion information that used in the traditional update steps, thus it reduces computations in temporal transform. Experimental results show that the coding performances can be improved up to 0.77dB.*

## 1. INTRODUCTION

Scalable video coding is very helpful for video delivering over Internet or wireless networks. It also gives different devices opportunities to share the same video over heterogonous networks. In various scalable video coding schemes, motion compensated/ aligned subband coding schemes attracted many researches [1][3][4] because of their inherent properties to support frame-rate scalability and SNR scalability. Moreover, these schemes also take advantages of motion models to enhance coding performance, which makes them can achieve high performance while providing scalabilities.

In [1][3][4], pixels in each frames search for their correspondences according to the results of motion estimation. Then temporal subband transform is applied along the motion trajectories. Connecting pixels together along motion trajectory although increases the temporal correlations, it hampers applying advanced motion estimation that yields fractional-pel motion vectors and generates many unconnected pixels. In [2], Daubechies and Swelden proposed a lifting scheme that factors subband transform into elementary steps - predict and update. Lifting-based transform enables elaborate design on each elementary step individually. With such a power tool, many techniques in motion compensated predictive coding can be applied in motion compensated subband coding, for example, fractional-pel motion estimation and compensation[5][6][7], overlapped block motion compensation[8]. By using these techniques, the coding performance of motion compensated subband coding improves significantly. The recent results[8] showed that motion compensated subband coding, with scalability support, can achieve comparable coding performance with the state-of-the-art non-scalable coding standard – H.264, and even sometimes it outperforms H.264.

While many techniques have been investigated on the predict steps to improve the prediction thus the coding performance, relatively few efforts have been spent on the update steps. Actually in the current motion compensated lifting based subband coding systems, motion information used in the update steps is derived from the inversion of the motion information in the predict steps. The motion inversion procedure, on one hand increases the complexity of the coding systems; on the other hand, it brings aliasing of the motion information. In this paper, we propose a new update scheme, named energy distributed update (EDU), in the lifting implementation of the motion compensated lifting based subband video coding. Through update along the motion trajectories, we avoid inversion of the motion information in the update steps. And by distributing the energy updated to corresponding pixels, we show that the new update scheme can make the signals updated smoother, which improves the coding performance.

The paper is organized as follows. In section 2, we introduce temporal transform in the motion compensated lifting based subband coding, and analyze the problems that may occur in the current update steps. Section 3 describes our idea of new update steps and implementation of the scheme. Experimental results are shown in the section 4, and we discuss the results in the same section. Section 5 concludes the paper.

## 2. MOTION COMPENSATED LIFTING BASED TEMPORAL TRANSFORM

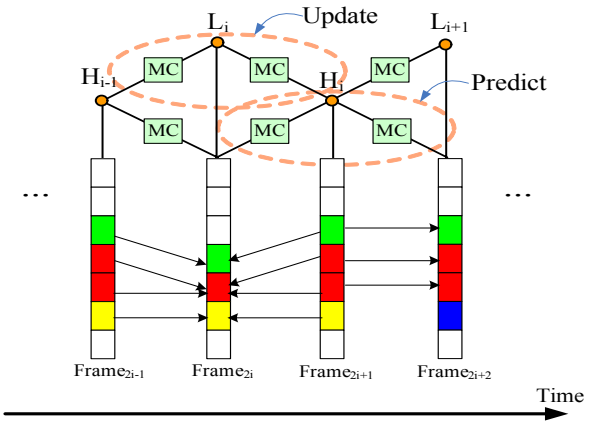



Figure 1. One level temporal transform in lifting based motion compensated video coding

In a common motion compensated lifting based subband video coding system, temporal subband transform is first utilized to analyse the video signals. Then spatial subband transform is applied. Finally the transform coefficients are lossy coded into the bitstream according to the target bitrate required. Lifting based

[*] This work has been done while the author is with Microsoft Research Asia.

transform and motion information are used in the temporal transform stage.

Assume that a video sequence, $I_0, I_1, ..., I_{2n-1}$ are to be processed with temporal transform. Figure 1 shows how the lifting based temporal transform is done with bi-orthogonal 5/3 wavelet filters. The first step is predict, which uses consecutive even frames to predict the odd frame:

$$H_i = I_{2i+1} - \frac{1}{2} MC(I_{2i}, MV_{2i+1\to 2i}) - \frac{1}{2} MC(I_{2i+2}, MV_{2i+1\to 2i+2}) \quad (1)$$

Where $H_i$ is the high-pass frame generated by the predict step. $MV_{2i+1->2i}$ means motion vectors from frame $2i+1$ to frame $2i$, along which the pixel in frame $2i+1$ can find its correspondence in frame $2i$. So does $MV_{2i+1->2i+2}$. And $MC()$ means motion compensation process that generates the current frame's prediction from its consecutive frame. Predict step in temporal 5/3 transform actually process in the same way as B frame's generation in motion compensated predictive coding. When a motion vector has fractional value, interpolation filter is utilized in the even frame to get the odd frame's prediction.

Update step follows the predict step to complete one level 5/3 subband transform, which generates low-pass frames:

$$L_i = I_{2i} + \frac{1}{4} MC(H_{i-1}, MV_{2i\to 2i-1}) + \frac{1}{4} MC(H_i, MV_{2i\to 2i+1}) \quad (2)$$

After one level transform, loss-pass frames can be processed in the next level's 5/3 subband transform.

For Haar filter, corresponding predict and update steps are:

$$H_i = I_{2i+1} - MC(I_{2i}, MV_{2i+1\to 2i}) \quad (3)$$

$$L_i = I_{2i} + \frac{1}{2} MC(H_i, MV_{2i\to 2i+1}) \quad (4)$$

Both in the predict step and update step, to generate one frame's data, two set of motion vectors are needed. In the predict stage, the direction of the motion vectors is from odd frames to even frames, and the direction inverses in the update step. Since both $MV_{2i->2i+1}$ and $MV_{2i+1->2i}$ describes the motion between frame $2i$ and frame $2i+1$, one direction's motion vectors can be derived approximately from the other direction's ones. The procedure is inversion of motion information, which is used in [5][6][7][8].

The inversion procedure, although saving one set of motion vectors' bits, is not very accurate. We say that because the procedure contains ambiguity and may yield results that do not consist with results in predict step. Figure 2 gives two examples. In figure 2(a), the motion displacement pixel $j$ in frame $2i+1$ is $-1/4$, i.e., a quarter pel. Let $p_{2i}(j)$ denote pixel $j$ in frame $2i$. Assume linear interpolation is applied. Then it means that the prediction of $p_{2i+1}(j)$ is weighted average of $p_{2i}(j)$ and $p_{2i}(j+1)$. And in the update step, only $p_{2i+1}(j-1)$ and $p_{2i+1}(j)$ influence $p_{2i}(j)$. There is predict from $p_{2i}(j+1)$ to $p_{2i+1}(j)$ but no update from $p_{2i+1}(j)$ to $p_{2i}(j+1)$. Similarly, there is update from $p_{2i+1}(j-1)$ to $p_{2i}(j)$ but no predict from $p_{2i}(j)$ to $p_{2i+1}(j-1)$. So it is not consistent between the predict step and update step.

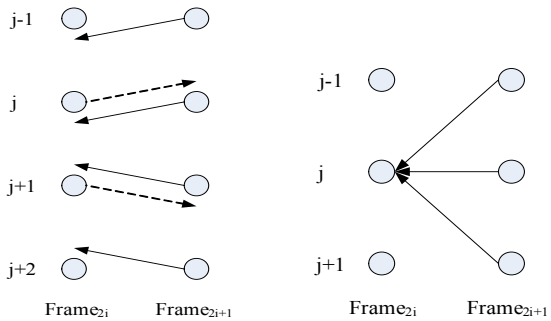



Figure 2. Two examples of motion inversion procedure. Solid arrows denote motion vectors in predict steps and dashed arrows stand for inverse motion vectors.

Figure 2(b) shows multiple-to-one case in the predict step. Motion vectors of $p_{2i+1}(j-1)$, $p_{2i+1}(j)$ and $p_{2i}(j+1)$ make them all point to $p_{2i}(j)$. Then in the motion inversion procedure, it is difficult to assign motion vector for $p_{2i}(j)$. So there is ambiguity in the motion inversion procedure.

Actually due to the complexity of video motion, it is hard to find one-to-one mapping between two consecutive frames along the motion trajectories. Besides multiple-to-one case, there are also cases that pixels in the even frame cannot find their motion vectors. And furthermore, the motion inverse procedure itself may be complex, which makes subband coding complex too.

## 3. ENERGY DISTRIBUTED UPDATE STEPS (EDU)

As described in the above section, the motion inversion procedure involved in the update steps introduces many problems. To solve the problem, we propose Energy Distributed Update steps, dubbed EDU, in motion compensated lifting based temporal transform. The key idea is to update where the predict is made. In the later part we will show that EDU solves the problems we mentioned.

Assume Haar filter is used in temporal transform. Let $h_i(x,y)$ and $l_i(x,y)$ denote position $(x,y)$ 's signals of $H_i$ and $L_i$ respectively. In the predict step, we follow the traditional process. According to (3), there is:

$$h_i(x,y) = p_{2i+1}(x,y) - MC_{(I_{2i}, MV_{2i+1\to 2i})}(x,y) \quad (5)$$

Where $p_{2i+1}(x,y)$ is the pixel $(x,y)$ of $I_{2i+1}$ and $MC_{(I_{2i},MV_{2i+1\to 2i})}(x,y)$ denotes position $(x,y)$ 's value of $MC(I_{2i},MV_{2i+1\to 2i})$ in (3). Despite that different interpolation filters or OBMC may be used in motion compensation, $MC_{(I_{2i},MV_{2i+1\to 2i})}(x,y)$ can be viewed as the weighted sum of pixels' value in $I_{2i}$:

$$MC_{(I_{2i},MV_{2i+1\to 2i})}(x,y)=\sum_{(m,n)}\alpha_{x,y,m,n}p_{2i}(m,n) \qquad (6)$$

$\alpha_{x,y,m,n}$ is the weigh, which determined by $MV_{2i+1\to 2i}$ and methods used in the motion compensation. So the predict step can be expressed as:

$$h_i(x,y)=p_{2i+1}(x,y)-\sum_{(m,n)}\alpha_{x,y,m,n}p_{2i}(m,n) \qquad (7)$$

From (7), we can see that $\alpha_{x,y,m,n}$ actually reflects how $p_{2i}(m,n)$ influences the prediction at position $(x,y)$ of frame $2i+1$. It is natural that $h_i(x,y)$ should also influence the same on $p_{2i}(m,n)$ in the update step. So the update step proposed here can be expressed as:

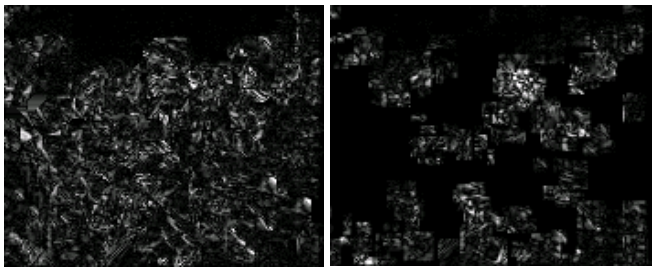
$$l_i(m,n)=p_{2i}(m,n)+\frac{1}{2}\sum_{(x,y)}\alpha_{x,y,m,n}h_i(x,y) \qquad (8)$$

(8) means that the high-pass signals will be added exactly to the positions they are predicted. $h_i(x,y)$ will be distributed to $l_i(m,n)$, that is why we called the method Energy Distributed Update (EDU). Let's go back to the case of figure 2(a). The predict weight from $p_{2i}(j+1)$ to $p_{2i+1}(j)$ is positive. So in the scheme proposed, the update weight from $p_{2i+1}(j)$ to $p_{2i}(j+1)$, which equals the predict weight, is also not zero. Now the predict step and update step are consistent. The new update step avoids deriving inverse motion vectors. Thus on one side, it saves the operations, and on the other side, it does not have ambiguity during update procedure.

Let $U_i$ be the update signals, i.e. $\sum_{(x,y)}\alpha_{x,y,m,n}h_i(x,y)$. To generate $U_i$ for the update step, $\alpha_{x,y,m,n}$ should be calculated. Since the support of the interpolation filter is limited and local, actually the weights can be got easily. For example, if bilinear interpolation filter is used, and the motion vector for pixel $(x,y)$ is $(\Delta x,\Delta y)$, at most four weights ( $\alpha_{x,y,x+\lfloor\Delta x\rfloor,y+\lfloor\Delta y\rfloor}$, $\alpha_{x,y,x+\lfloor\Delta x\rfloor,y+\lfloor\Delta y\rfloor+1}$, $\alpha_{x,y,x+\lfloor\Delta x\rfloor+1,y+\lfloor\Delta y\rfloor}$ and $\alpha_{x,y,x+\lfloor\Delta x\rfloor+1,y+\lfloor\Delta y\rfloor+1}$ ) are not zero. And the values is bilinear weight. So for each pixel $(x,y)$, $h_i(x,y)$ will be added scaled by a weight to at most four positions in $U_i$. When all pixels are processed, $U_i$ is ready and can be added to form the loss-pass frames.

# 4. EXPERIMENTAL RESULTS

Extensive experiments have been done to test the performance of our scheme. Subband coding system described in [8] is used to code Bus(150 frames), Coastguard(300 frames), Football(260 frames), Foreman(300 frames) and Stefan(300 frames) in QCIF format. Fore each sequence, 4-level lifting based temporal transform is applied, which transform the video signals into 5 temporal subbands. Then spatial transform is applied and the transform coefficients are coded and truncated to the target bitrate. In our experiments, we use bilinear interpolation filter to get the weights in the update steps. And 5/3 filter and Haar filter are used respectively.



(a)UIM (b)EDU

Figure 3. Illustration of the update signals.

Figure 3 compares the update signals of the traditional update steps using inverse motion (UIM) and the EDU scheme proposed. They are one frame's update signals in the 4th level temporal transform of Football sequence. Obviously the update signals of UIM are more promiscuous. The update signals in UIM scheme contains more high frequency signals, which will affect the coding performance of low-pass frames.

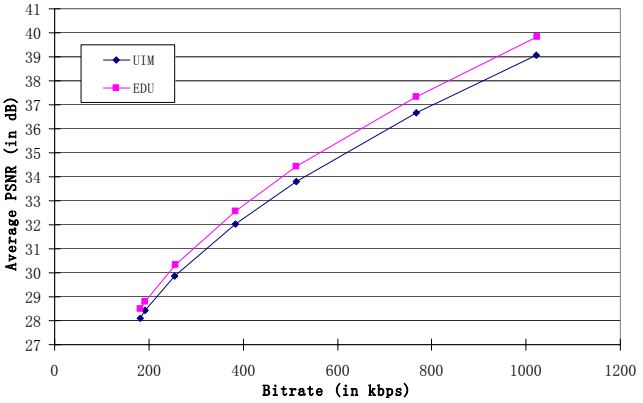



Figure 4. Coding performance comparison of Football sequence.

The coding performance comparisons with 5/3 filter are shown in figure 4. EDU scheme outperforms UIM scheme, no matter it is at low bitrate or high bitrate. Table 1 and table 2 present the coding performance results at different bitrates with 5/3 filter and Haar filter respectively. The coding performances of all sequences improve with EDU, especially when the sequence contains large or complex motions, e.g. Football and Bus sequences. It is consist

with the above analysis that EDU can handle motion better in the update steps.

## 5. CONCLUSIONS

A new update scheme is proposed in the temporal transform for motion compensated lifting based subband video coding. By distributing high-pass energy to low-pass frames along motion trajectories in the predict steps, the update steps proposed avoid motion vector inversion procedure, which may make predict and update be inconsistent and may bring ambiguity. Extensive experimental results validate the effectivity of the new update scheme.

## ACKNOWLEDGEMENT

The authors would like to take this opportunity to thank Mr. Ruiqin Xiong for his help when we did experiments and his many useful discussions.

## REFERENCES

[1] J. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, pp. 559-571, Sep 1994.

[2] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, 4(3):247--269, 1998.

[3] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3D ESCOT)", *Applied and Computational Harmonic Analysis*, vol 10, pp.290-315, 2001.

[4] S.-J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 3, no. 2, pp. 155-167, Feb. 1999.

[5] J. W. Woods and P. Chen, "Improved MC-EZBC with quarter-pixel motion vectors", ISO/IEC JTC1/SC29/WG 11, MPEG doc. M8366, Fairfax, May 2002.

[6] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Trans. Image Processing*, vol. 12, no. 12, Dec 2003.

[7] L. Luo, F. Wu, S. Li, and Z. Zhuang, "Advanced lifting-based Motion-Threading (MTh) techniques for 3D wavelet video coding", *Visual Communications and Image Processing*, Lugano, Switzerland, Jul. 2003.

[8] R. Xiong, F. Wu, S. Li, Z. Xiong and Y.-Q. Zhang, "Exploting temporal correlation with adaptive block-size motion alignment for 3D wavelet coding", *Visual Communications and Image Processing*, San Jose, California, USA, Jan. 2004.

Table 1. Average PSNR(in dB) of the coding results with 5/3 filter.

| Bitrate | 128kbps | | | 256kbps | | | 512kbps | | | 1024kbps | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | UIM | EDU | Gain | UIM | EDU | Gain | UIM | EDU | Gain | UIM | EDU | Gain |
| Bus | 30.12 | 30.32 | **0.20** | 33.85 | 34.10 | **0.25** | 38.22 | 38.44 | **0.22** | 43.56 | 43.71 | **0.15** |
| Coastguard | 34.36 | 34.44 | **0.08** | 37.45 | 37.47 | **0.01** | 41.05 | 41.07 | **0.01** | 45.24 | 45.24 | **0.00** |
| Football | 26.50 | 26.74 | **0.24** | 29.87 | 30.32 | **0.45** | 33.81 | 34.44 | **0.63** | 39.06 | 39.83 | **0.77** |
| Foreman | 36.26 | 36.47 | **0.21** | 40.13 | 40.29 | **0.16** | 43.92 | 44.09 | **0.16** | 47.56 | 47.61 | **0.06** |
| Stefan | 28.25 | 28.55 | **0.30** | 32.25 | 32.49 | **0.24** | 36.28 | 36.56 | **0.27** | 41.12 | 41.42 | **0.30** |

Table 2. Average PSNR(in dB) of the coding results with Haar filter.

| Bitrate | 128kbps | | | 256kbps | | | 512kbps | | | 1024kbps | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | UIM | EDU | Gain | UIM | EDU | Gain | UIM | EDU | Gain | UIM | EDU | Gain |
| Bus | 26.84 | 27.07 | **0.23** | 30.30 | 30.74 | **0.44** | 34.52 | 35.05 | **0.53** | 40.01 | 40.58 | **0.57** |
| Coastguard | 31.63 | 31.85 | **0.22** | 34.39 | 34.60 | **0.20** | 37.65 | 37.87 | **0.23** | 41.99 | 42.26 | **0.27** |
| Football | 24.90 | 24.98 | **0.09** | 28.54 | 28.64 | **0.10** | 32.46 | 32.67 | **0.20** | 37.64 | 37.98 | **0.34** |
| Foreman | 33.46 | 33.69 | **0.22** | 37.11 | 37.35 | **0.25** | 41.06 | 41.40 | **0.34** | 45.50 | 45.75 | **0.25** |
| Stefan | 25.49 | 25.73 | **0.25** | 29.34 | 29.66 | **0.32** | 33.55 | 33.91 | **0.37** | 38.61 | 39.01 | **0.40** |