MOTION COMPENSATED TEMPORAL FILTERING WITH OPTIMAL TEMPORAL DISTANCE BETWEEN EACH MOTION COMPENSATION PAIR

Yang Liu[1]*, Z. G. Li*[2] *and Y. C. Soh* [1]

[1] School of EEE Nanyang Technological University
Nanyang Drive, Singapore 639798
eycsoh@ntu.edu.sg
[2]Media Division Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
ezgli@i2r.a-star.org.sg

ABSTRACT

A novel motion compensated temporal filtering (MCTF) scheme is proposed in this paper by properly using forward and backward motion compensations. The mean, the second moment and the maximum value for the temporal distance of all motion compensation pairs (MCPs) in a group of frames (GOF) are minimized such that the number of "unconnected" pixels is minimized. The overall coding efficiency is improved by up to 1.5dB when compared to the scheme provided in [**?**] while the total number of motion estimation remains the same.

1. INTRODUCTION

Reliable transmission of video over heterogeneous networks requires efficient coding, as well as scalability to different client capabilities, system resources, and network conditions. For example, clients may have different display resolutions, systems may have different caching or intermediate storage resources, and networks may have varying bandwidths, loss rates, and best-effort or quality of service (QoS) capabilities. Scalable video coding has been proposed to increase its adaptability to network and client conditions. There are many applications that require scalable and reliable video coding: Internet video, wireless LAN video, mobile wireless video for conversational, video on demand (VOD), live broadcasting purposes, end-to-end Internet/wireless video delivery, multi-channel content production and distribution, storage applications, multi-point surveillance systems, and so on.

Recently, three dimensional subband wavelet coding was proposed as an efficient scalable video coding scheme, especially with the concept of motion compensated temporal filtering (MCTF). The MCTF was proposed by Ohm [**?**] as an efficient tool to remove temporal redundancies in wavelet based video coding schemes. The pixels are classified into "connected" and "unconnected" pixels by the MCTF [**?**]. Typically, there are about 3-5% of the pixels that will be "unconnected" in the MCTF process, and they seriously affect both the overall coding gain and the subjective video quality. It is thus very important to reduce the number of "unconnected" pixels. Ohm [**?**] proposed an interesting method in which motion compensation prediction is performed for the "unconnected" pixels in the previous frame by using the reconstructed frame just before it. Recently, Choi and Woods proposed another novel method in [**?**] to improve the MCTF by making the direction of the motion estimation the same as that of motion compensation. These subband positions are better suited in the case of "unconnected" pixels than that in [**?**].

Obviously, the existing schemes [**?**, **?**] focused on improving the coding efficiency of the high subbands at the first round. This is not enough because the motion-composed 3-D subband coding is usually an iterative process. For example, when the number of frames in a group of frames (GOF) is 16, four rounds of MCTF will be performed. It is thus also very important to improve the coding efficiency of the high subbands in the other rounds by reducing the number of "unconnected" pixels in these rounds.

To minimize the total number of "unconnected" pixels, the mean, the second moment and the maximum value for the temporal distances of all motion compensation pairs (MCPs) in a GOF should be minimized. The MCP is composed of a predicted frame and the corresponding reference frame where the predicted frame preforms motion compensation based on the reference frame. It is with this objective that we propose our MCTF scheme. In our scheme, let M denote the total number of rounds of MCTF. If M is less than 4, the forward and backward motion compensations are alternatively used in all rounds of MCTF. Otherwise, they are alteratively used in all except for the $(M-2)$th round of MCTF, where a combination of forward-backward and

0-7803-8554-3/04/$20.00 ©2004 IEEE.

backward-backward predictions is used. All the means, the second moments and the maximum values for the temporal distances of all MCPs in a GOF are minimized. The overall coding efficiency is improved by up to 1.5dB when compared to that proposed in [**?**]. Meanwhile, the total number of motion estimations remains the same as that of the schemes proposed in [**?**, **?**]. Our scheme is thus very attractive for wavelet based scalable video coding.

There are some other ways to improve the overall coding efficiency. Among these are the lifting-based MCTF schemes proposed in [**?**, **?**] and the unconstrained motion compensated temporal filtering (UMCTF) in [**?**]. It is also possible to take a long tap filter for two-channel subband analysis and cascade it in a tree-structured manner. However, the computation involved in these schemes is much more complex than the schemes proposed in [**?**, **?**] and the associated coding delay would be too long [**?**].

The rest of this paper is organized as follows. The problem formulation and our proposed scheme are given in Section 2. Section 3 contains extensive experimental results to illustrate the effectiveness of our scheme. Finally, concluding remarks are provided in Section 4.

2. A NOVEL MCTF SCHEME

2.1. Problem Formulation

In this section, we shall first present the concept of MCP. An MCP consists of a predicted frame and the corresponding reference frame where the predicted frame preforms motion compensation based on the reference frame.

The process of MCTF together with block matching between an MCP is described as follows. Matched blocks in the reference frame overlap with neighboring blocks except in the case of no motion or pure translational motion. The pixels are classified into "connected" and "unconnected" by their estimated motion vectors [**?**]. The "connected" pixels are filtered along the motion trajectory while for the remaining "unconnected" pixels, the original pixel value of a frame (the predicted frame or the reference frame) is inserted into the temporal low subband and the scaled displaced frame difference (DFD) is inserted into the temporal high subband.

The number of "unconnected pixels" depends heavily on the temporal distance between each MCP. Suppose that the temporal distance between the ith MCP in the jth GOF is $d_{i,j}$ and the corresponding mean for the percentage of "unconnected" pixels is $\beta_{i,j}\%$. The relationship between $\beta_{i,j}$ and $d_{i,j}$ is given by

$$\beta_{i,j} = f(d_{i,j}) \tag{1}$$

Since the "unconnected" pixels seriously affect both the overall coding gain and the subjective video quality, it is very important to minimize the total number of "unconnected" pixels.

Let $\bar{\beta}_j$ denote the average value of $\beta_{i,j}$, and suppose that the total number of MCPs in the jth GOF is N_j. The problem can then be formulated as the following optimization problem.

$$\min_{d_{1,j},\cdots,d_{N_j,j}} \bar{\beta}_j \tag{2}$$

2.2. The Proposed MCTF Scheme

Normally, it is very difficult to solve the optimization problem (**??**) because f is not available in advance. However, the Taylor expansion of f can be used to simplify the problem.

It can be shown from Table 2 in Section 3 that f is a non-decreasing function of $d_{i,j}$, and $f(d_{i,j})$ is well approximated by its second order Taylor series expansion when $d_{i,j}$ is less than 4, i.e.

$$f(d_{i,j}) = f'(0) * d_{i,j} + \frac{f''(0)}{2} d_{i,j}^2 + o(d_{i,j}^3) \ ; \ d_{i,j} \leq 3 \tag{3}$$

where $f'(0)$ and $f''(0)$ are greater than 0.

Let $\bar{d}_j$, $\bar{d_j^2}$ and $d_{j,max}$ denote the mean, the second moment and the maximum value of $d_{i,j}$, respectively. When $d_{j,max}$ is less than 4, we have

$$\bar{\beta}_j \approx f'(0) * \bar{d}_j + \frac{f''(0)}{2} \bar{d_j^2} \tag{4}$$

Therefore, an optimal solution can be achieved by minimizing all $\bar{d}_j$, $\bar{d_j^2}$ and $d_{j,max}$.

Let M denote the total number of rounds of MCTF, the optimal solution can then be obtained as follows:

1. If M is less than 4, the forward and backward predictions are alteratively used in all rounds of MCTF.
2. Otherwise, the forward and backward predictions are alteratively used in all except for the $(M-2)$th round of MCTF, where a combination of forward-backward and backward-backward predictions is used.

After the temporal distance of each MCP is determined, the remaining process is similar to the scheme in [**?**]. The low subband is located at the position of the reference frame while the high subband is situated at the corresponding position in the predicted frame. In other words, the advantage of motion estimation is well used in our scheme in the sense that the high frequency subbands have smaller energy and are compatible with a scalable DFD value for the "unconnected" pixels.

The comparison of our scheme with that in [**?**] is illustrated in Figures **??** and **??** with the size of the GOF set as 16. The solid lines in the second, third, fourth and fifth rows stand for low subbands while dash lines represent high subbands. The solid lines with arrows represent the first round of motion compensation for the original frames while the

dash lines with arrows stand for the second round of motion compensation for an MCP. The temporal distances between each MCP in our scheme and that in the scheme proposed in [**?**] (we denote this by MCTF [**?**]) are listed in Table 1. It can be easily computed that $\bar{d}_j$, $\bar{d}_j^2$ and $d_{j,max}$ are 1.53, 2.53 and 4, respectively for our scheme while they are 2.13, 8 and 8, respectively for the scheme proposed in [**?**].

Table 1. Comparison of temporal distance between each MCP

Round of MCTF	Number of MCPs	Scheme	Temporal distance
1	8	Proposed	1
		MCTF [**?**]	1
2	4	Proposed	1
		MCTF [**?**]	2
3	2	Proposed	3 or 4
		MCTF [**?**]	4
4	1	Proposed	4
		MCTF [**?**]	8

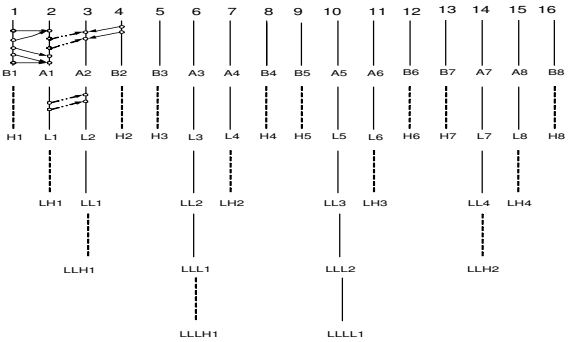

Fig. 1. The proposed MCTF.

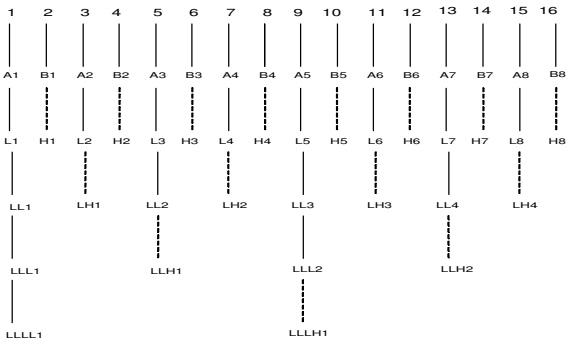

Fig. 2. The MCTF proposed by Choi and Woods [2].

3. EXPERIMENTAL RESULTS

Our experiments are based on the MC-EZBC interframe wavelet coder [**?**]. Two standard video sequences, Flower Garden with SIF size (352×240) and Foreman with CIF size (352×288), are used in our test to compare our proposed scheme with the scheme in [**?**] (we denote this by MCTF [**?**]). The frame rates of these two sequences are set as 20f/s. The bit rate is 1024kb/s. The hierarchical variable size block matching (HVSBM) algorithm provided in [**?**] is used with the maximum search width/height at the lowest resolution in the hierarchical motion estimation set as 1 pixel. The overlapped block motion compensation (OBMC) scheme with 1/8 block overlapped is also employed in our experiments to give better smoothness in the motion vector field.

Table 2. Comparison of the number of unconnected pixels for the sequence Flower Garden at GOFsize=16, half-pixel accuracy

Frames to be filtered	Method	pixels number	
		"connected"	"unconnected"
A1-B1	MCTF [**?**]	81779	2701
	Proposed	81967	2513
A2-B2	MCTF [**?**]	81489	2991
	Proposed	81489	2991
A3-B3	MCTF [**?**]	81562	2918
	Proposed	82089	2391
A4-B4	MCTF [**?**]	81971	2509
	Proposed	81971	2509
A5-B5	MCTF [**?**]	82115	2365
	Proposed	82070	2410
A6-B6	MCTF [**?**]	81831	2649
	Proposed	81831	2649
A7-B7	MCTF [**?**]	82052	2428
	Proposed	82262	2218
A8-B8	MCTF [**?**]	82241	2239
	Proposed	82241	2239
L1-L2	MCTF [**?**]	77700	6780
	Proposed	81165	3315
L3-L4	MCTF [**?**]	77560	6920
	Proposed	80664	3816
L5-L6	MCTF [**?**]	78037	6443
	Proposed	81185	3295
L7-L8	MCTF [**?**]	78724	5756
	Proposed	81546	2934
LL1-LL2	MCTF [**?**]	67384	17096
	Proposed	71062	13418
LL3-LL4	MCTF [**?**]	67368	17112
	Proposed	67461	17019
LLL1-LLL2	MCTF [**?**]	54460	30020
	Proposed	65480	19000

Table 2 shows the number of "connected" and "unconnected" pixels for each MCP in a GOF with size set as 16. At the first round, the difference in the number of "unconnected" pixels is very slight. However, the number of "unconnected" pixels is reduced significantly in the other three rounds. The maximal reduction ratio of the unconnected pixels is more than 50%.

We next compare our scheme with that in [**?**] for different choices of GOF size in term of PSNR gain. The experimental results are illustrated in Table 3. Clearly, we obtain a higher gain with a larger GOF size.

We also compare our scheme with that in [**?**] for different bitrates and different subpixel accuracy. It is shown in

Table 3. Comparison of PSNR gain for different GOF size, half-pixel accuracy

GOF Size	Video Sequence	PSNR Gain(dB)
4	Foreman	0.14
	Flower Garden	0.44
8	Foreman	0.24
	Flower Garden	0.82
16	Foreman	0.57
	Flower Garden	1.24

Table 4. Comparison of the average PSNR for the sequence Foreman at different bitrate, GOFsize=16, half-pixel accuracy

Bitrate (kbps)	PSNR	Method		Gain
		MCTF [**?**]	Propsed	
128	Y (dB)	30.74	31.90	+1.16
	U (dB)	38.10	39.12	+1.02
	V (dB)	39.00	40.48	+1.48
192	Y (dB)	32.79	33.80	+1.01
	U (dB)	39.70	40.44	+0.74
	V (dB)	41.14	42.22	+1.08
256	Y (dB)	33.95	34.99	+1.04
	U (dB)	40.60	41.27	+0.67
	V (dB)	42.06	43.28	+1.22
512	Y (dB)	36.78	37.61	+0.83
	U (dB)	42.90	43.48	+0.58
	V (dB)	44.39	45.38	+0.99
1024	Y (dB)	39.92	40.49	+0.57
	U (dB)	44.90	45.48	+0.58
	V (dB)	46.41	47.23	+0.82

Tables 4 and 5 that our scheme outperforms the result of [**?**] for each bitrate and subpixel accuracy.

Overall, our scheme achieves a higher coding efficiency for any GOF size, bitrate, and subpixel accuracy, while the PSNR variation is comparable. Meanwhile, the number of motion estimation remains the same. Thus, our scheme is very attractive for wavelet based scalable video coding.

Table 5. Comparison of the average PSNR for the sequence Flower Garden with different subpixel accuracy, GOFsize=16

Subpixel	PSNR	Method		Gain
		MCTF [**?**]	Proposed	
Full	Y (dB)	27.10	27.87	+0.77
	U (dB)	30.80	31.53	+0.73
	V (dB)	32.54	33.21	+0.67
Half	Y (dB)	28.11	29.35	+1.24
	U (dB)	32.00	32.98	+0.98
	V (dB)	33.49	34.45	+0.96
Quater	Y (dB)	28.69	30.11	+1.42
	U (dB)	32.50	33.83	+1.33
	V (dB)	33.80	34.97	+1.17
Eighth	Y (dB)	28.77	30.27	+1.50
	U (dB)	32.70	34.24	+1.54
	V (dB)	33.89	35.23	+1.34

4. CONCLUSION

A novel MCTF has been proposed in this paper by minimizing all the means, the second moments and the maximum values for the temporal distances of all motion compensation pairs in a group of frames. The overall coding efficiency is improved by up to 1.5 dB when compared to the existing scheme without any increase in the number of motion estimation. Our scheme is thus very attractive for wavelet based scalable video coding.

5. REFERENCES

[1] J. -R. Ohm. "Three-Dimensional Subband Coding with Motion Compensation". IEEE Trans. on Image Processing, Vol. 3 No. 9, pp. 559-571, Sept. 1994.

[2] Seung-Jong Choi and John W. Woods. "Motion-Compensated 3-D Subband Coding of Video". IEEE Trans. on Image Processing, Vol. 8 No. 2, pp. 155-167, Feb. 1999.

[3] B. Pesquet Popescu and V. Bottreau. "Three-Dimensional Lifting Schemes for Motion Compensated Video Compression". In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1793-1796, May 2001.

[4] A. Secker and D. Tubman. "Motion-Compensated Highly Scalable Video Compression Using an Adaptive 3D Wavelet Transform Based on Lifting". In Proc. IEEE International Conference on Image Processing, pp. 1029-1032, Thessaloniki, Greece, Sep. 2001.

[5] D. S. Turaga and M. van der Schaar. "Unconstrained Motion Compensated Temporal Lifting". MPEG Contribution, M8520, July 2002.

[6] MC-EZBC software packege, ftp://ftp.cipr.rpi.edu/personal/chen/