Content adaptive update for lifting-based motion-compensated temporal filtering

L. Song, J. Xu, H. Xiong and F. Wu

Content adaptive update steps based on the property of the human vision system are presented for lifting-based motion-compensated temporal filtering (MCTF). Simulations confirm that the proposed method improves both the rate-distortion performance for MCFT and the visual quality significantly.

Introduction: Lifting-based motion-compensated temporal filtering (MCTF) is used widely as the preferred temporal transformation technique in state-of-the-art three-dimensional subband video coding. It involves two basic stages: the prediction stage that takes the original input frames to generate the highpass frames and the update stage that uses the available highpass frames and even frames to generate the lowpass frames. Each highpass frame is essentially the residual from motion-compensated prediction of the relevant odd indexed original video frames. Its energy depends only upon the success of the motion models. When the motion model fails, object features may not be aligned, resulting in multiple edges and increased energy in the highpass subband frames, which not only reduce the compression performance, but also add ghosting to lowpass temporal frames, significantly reducing their visual quality. Turaga [1] proposed omitting the update steps based on quality of motion estimation and the nature of the motion in the sequence. Mehrseresht [2] showed that ignoring the update not only incurs the loss of compression performance, but also increases the fluctuation in PSNR values. Therefore, we should mitigate adaptively the ghosting artefacts of lowpass temporal frames, while preserving the update steps in general.

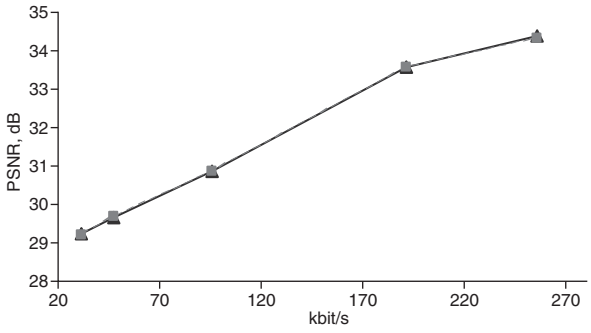


Fig. 1 *Performance of content adaptive update operator for 'foreman' sequence*

classical update steps
content adaptive update steps

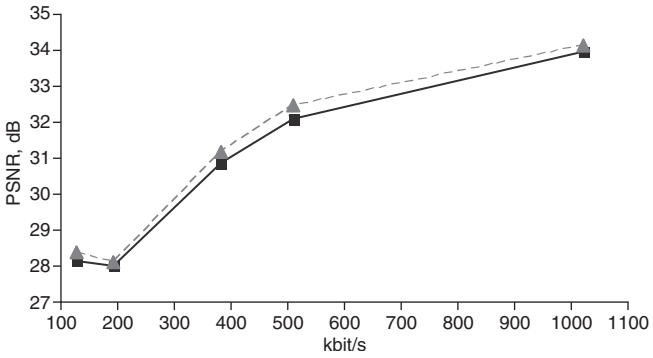


Fig. 2 *Performance of content adaptive update operator for 'football' sequence*

classical update steps
content adaptive update steps

Content adaptive update steps: Assume that a video sequence, $s_0, s_1, \ldots, s_{2n-1}$ is to be processed with (5, 3) lifting-based MCTF. The classical predict and update steps can be formulated as follows [3]:

$$H[i] = s[2i+1] - P(s[2i+1])$$

with

$$P(s[2i+1]) = \frac{1}{2}[MC(s[2i+1], MV_{2i+1\to 2i}) + MC(s[2i+1], MV_{2i+1\to 2i+2})] \quad (1)$$

$$L[i] = s[2i] + U(s[2i]) \quad \text{with}$$

$$P(s[2i]) = \frac{1}{4}[MC(H[i-1], MV_{2i\to 2i-1}) + MC(H[i], MV_{2i\to 2i+1})] \quad (2)$$

where $H[i]$ is the highpass frame generated by the predict step, P is the motion-compensated prediction of relevant frames, $MV_{2i+1\to 2i}$ means motion vectors from frame $2i+1$ to frame $2i$, so does $MV_{2i+1\to 2i}$, and $MC(\,)$ means motion compensation process. $L[i]$ is the lowpass frame, and U is a residual to be added to even frames. Since the predict step strives to minimise the bit-rate required to encode highpass frame along with motion vectors used for prediction, $L[i]$ can be treated as a contaminated frame if we regard U as 'noise' added to the 'original' even indexed frame s_{2i}. Based on this analysis, the proposed update step is generalised as follows:

$$L[i] = s[2i] + f(U(s[2i])) \quad (3)$$

where $f(\,)$ is the proposed adaptation function. The function introduced here takes advantage of the results and developments of the human visual model (HVS) in computer vision. Among numerous computing models of the HVS, the just noticeable difference (*JND*) is used widely in perceptual coding. It is referred to as visibility thresholds that are defined as functions of the amplitude of luminance edge in which perturbation is increased until it becomes just discernible [4]. The *JND* thresholds are image dependent, and as long as the update information remains below these thresholds, we achieve 'update residual' transparency. Therefore, the *JND* matches very well with the subject of update steps addressed before. In this Letter, we define the following *JND* models:

$$JND_s(m, n) = 1 - \frac{1}{1 + \theta\sigma_s^2(m, n)} \quad (4)$$

where $\sigma_s^2(m, n)$ denotes the local variance of the image s in a window (W) centred on the pixel with co-ordinates (m, n). θ is a tuning parameter that can be computed as follows:

$$\theta = \frac{D}{\sigma_{s_{\max}}^2} \quad (5)$$

$\sigma_{s_{\max}}^2$ is the maximum local variance for a given image, and $D \in [50,100]$ is an experimentally-determined parameter. The second item of (4) is the same as the noise visibility function in image watermarking supposing that the image is a non-stationary Gaussian process [5]. It can be seen that *JND* reflects the texture masking property of HVS; the noise is more visible in flat or textureless areas and less visible in regions with edges and textures.

a *b*

Fig. 3 *Demonstration of visual quality improvement of content adaptive update operator for 'foreman' sequence, frame 38*

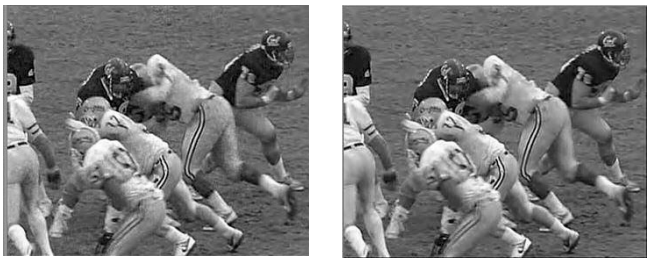
a CIF video at 256 kbit/s, 30 fps classical update steps
b CIF video at 256 kbit/s, 30 fps content adaptive update steps

Based on the above analysis, the content adaptive update function is defined as:

$$f(U(s[2i])) = \begin{cases} U(s[2i]) & |U(s[2i])| < JND_{s[2i]}S \\ JND_{s[2i]}S & U(s[2i]) \geq JND_{s[2i]}S \\ -JND_{s[2i]}S & U(s[2i]) \leq -JND_{s[2i]}S \end{cases} \quad (6)$$

S denotes the strength factor. The encoder and the decoder use the same *JND* metric and have no overhead. Although they operate on different images, our experiments show that the resulting *JND* masks are very similar on both sides. Therefore, they have negligible mismatch.

Results of simulation: Extensive experiments have been performed on several standard test sequences to show the performance of the proposed method. The four-level temporal transform was adopted and a five-layered bit stream with different rate and scalability was tested with the following parameters: $S = 12.5$, $D = 100$ and $W = 3 \times 3$. Figs. 1 and 2 show the performance of content adaptive update operator for '*foreman*' and '*football*' sequences. They are also compared with results of classical update steps. Figs. 1 and 2 demonstrate that the proposed content adaptive update operator can always catch or exceed the classical update steps, both at the low bit rate end and at the high bit rate end. Figs. 3 and 4 present a visual quality comparison of our method and the classical one, using the same parameters as in Figs. 1 and 2. We assume a bit rate of 256 kbit/s for the '*foreman*' sequence, which has the same frame rate as the original sequence, and a bit rate of 512 kbit/s for the '*football*' sequence which has half the frame rate of the original sequence. It is obvious that the visual quality of our method is much better, especially in smooth areas.



a *b*

Fig. 4 *Demonstration of visual quality improvement of content adaptive update operator for 'football' sequence, frame 9*

a CIF video at 512 kbit/s, 15 fps classical update steps
b CIF video at 512 kbit/s, 15 fps content adaptive update steps

4 September 2004
Electronics Letters online no: 20056525
doi: 10.1049/el:20056525

L. Song and H. Xiong (*Institute of Image Communication & Information Processing, Shanghai Jiaotong University, Huashan Road, No.1954, Shanghai, People's Republic of China*)

J. Xu and F. Wu (*Microsoft Research Asia, 3/F, Beijing Sigma Center, No. 49, Zhichun Road, Hai Dian District, Beijing, People's Republic of China*)

E-mail: fengwu@microsoft.com

References

1 Turaga, D.S., and van der Schaar, M.: 'Content-adaptive filtering in the UMCTF framework'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP2003), Hong Kong, April 2003, Vol. 3, pp. 6–10
2 Mehrseresht, N., and Taubman, D.: 'Adaptively weighted update steps in motion compensated lifting based on scalable video compression'. IEEE Int. Conf. on Image Processing (ICIP2003), Barcelona, September 2003, Vol. 2, pp. 771–774
3 Luo, L., Li, S., Zhuang, Z., and Zhang, Y.-Q.: 'Motion compensated lifting wavelet and its application in video coding'. IEEE Int. Conf. on Multimedia and Expo (ICME2001), Tokyo, August 2001, pp. 365–368
4 Netravali, A.N., and Prasada, B.: 'Adaptive quantization of picture signals using spatial masking', *Proc. IEEE*, 1977, **65**, pp. 536–548
5 Voloshynovskiy, S., Herrigel, A., Baumgärtner, N., and Pun, T.: 'A stochastic approach to content adaptive digital image watermarking'. International Workshop on Information Hiding, Vol. LNCS1768 of Lecture Notes in Computer Science, Dresden, October 1999, pp. 212–236