

In-band adaptive update step based on local content activity

Davide Maestroni, Marco Tagliasacchi and Stefano Tubaro

Dipartimento di Elettronica e Informazione, Politecnico di Milano,
P.zza Leonardo da Vinci, 32, 20133, Milano, Italy;

ABSTRACT

In this paper we present an adaptive version of the update step in the lifting implementation of MCTF (Motion Compensated Temporal Filtering).¹ We explicitly take into account the local image content in order to avoid ghosting artifacts in the updated low-pass frame, tuning the update weighting factor where such artifacts are more likely to be perceived. The proposed solution is integrated into a 2D+t (in-band) wavelet based video codec and improves the subjective quality of sequences reconstructed at reduced frame rate.

Keywords: Update step, scalability, in-band wavelet coding

1. INTRODUCTION

Today's video streaming applications require codecs to provide a bitstream that can be flexibly adapted to the characteristics of the network and the receiving device. Such codecs are expected to fulfill the scalability requirements so that encoding is performed only once, while decoding takes place each time at different spatial resolutions, frame rates and bitrates. Consider for example streaming a video content to TV sets, PDAs and cellphones at the same time. Obviously each device has its own constraints in terms of bandwidth, display resolution and battery life. For this reason it would be useful for the end users to subscribe to a scalable video stream in such a way that a representation of the video content matching the device characteristics can be extracted at decoding time. Wavelet based video codecs have proved to be able to naturally fit this application scenario, by decomposing the video sequence into a plurality of spatio-temporal subbands. Combined with an embedded entropy coding of wavelet coefficients such as JPEG2000,² SPIHT (Set Partitioning in Hierarchical Trees),³ EZBC (Embedded Zero-Block Coding)⁴ or ESCOT (Motion-based Embedded Subband Coding with Optimized Truncation),⁵ it is possible to support spatial, temporal and SNR (quality) scalability. Broadly speaking, two families of wavelet-based video codecs have been described in the literature:

- t+2D schemes⁶⁻⁸: the video sequence is first filtered in the temporal direction along the motion trajectories (MCTF - Motion Compensated Temporal Filtering¹) in order to tackle temporal redundancy. Then, a 2D wavelet transform is carried out in the spatial domain. Motion estimation/compensation takes place in the spatial domain, hence conventional coding tools used in non-scalable video codecs can be easily reused
- 2D+t (or in-band) schemes^{9,10}: each frame of the video sequence is wavelet transformed in the spatial domain, followed by MCTF. Motion estimation/compensation is carried out directly in the wavelet domain.

Due to the non-linear motion warping operator needed in the temporal filtering stage, the order of the transforms does not commute. In fact the wavelet transform is not shift invariant and care has to be taken since the motion estimation/compensation task is performed in the wavelet domain. In the literature several approaches have been used to tackle this issue. Although known under different names (low-band-shift,¹¹ ODWT (Overcomplete Discrete Wavelet Transform),¹² redundant DWT¹⁰), all the solutions present different implementations of the algorithm *a trous*,¹³ that computes an overcomplete wavelet decomposition by omitting the decimators in the fast DWT algorithm and stretching the wavelet filters by inserting zeros. A two level ODWT transform on a 1D signal is illustrated in Figure 1. The extension to 2D signals is straightforward with a separable approach. Despite its higher complexity, a 2D+t scheme comes with the advantage of reducing the impact of blocking

Further author information: (Send correspondence to Marco Tagliasacchi)

Marco Tagliasacchi: E-mail: marco.tagliasacchi@polimi.it, Telephone: +39 031 332 7341

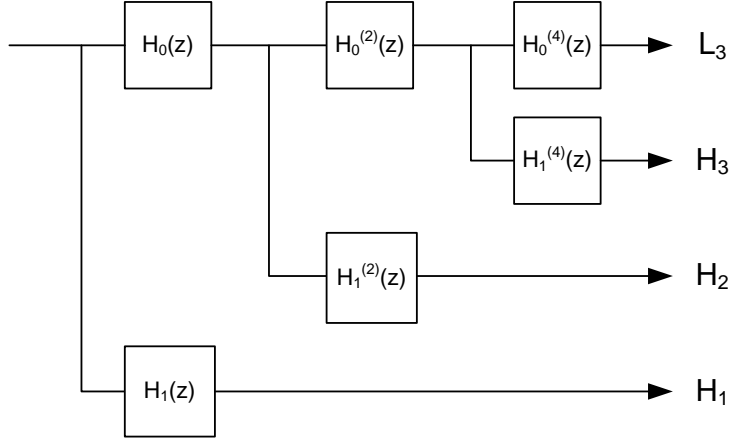


Figure 1. Two level overcomplete DWT (ODWT) of a 1D signal according to the algorithm *a trous* implementation. $H_0(z)$ and $H_1(z)$ are respectively the wavelet low-pass and high-pass filters used in the conventional critically sampled DWT. $H_i^k(z)$ is the dilated version of $H_i(z)$ obtained by inserting $k - 1$ zeros between two consecutive samples.

artefacts caused by the failure of block-based motion models. This is because such artefacts are smeared out by the inverse DWT spatial transform, without the need to adopt some sort of de-blocking filtering. This turns out to greatly enhance the perceptual quality of reconstructed sequences, especially at low-bitrates. Furthermore, as shown in Ref. 14 and,¹⁵ 2D+t approaches naturally fit the spatial scalability requirements providing higher coding efficiency when the sequence is decoded at reduced spatial resolution. This is due to the fact that with in-band motion compensation it is possible to limit the problem of drift that occurs when decoder does not have access to all the wavelet subbands used at the encoder side. Finally, 2D+t schemes naturally support multi-hypothesis motion compensation taking advantage of the redundancy of the ODWT.¹⁰

2. BACKGROUND ON MCTF

Motion Compensated Temporal Filtering (MCTF)¹ has proved to be an effective coding tool in the design of scalable video codecs. Both DCT¹⁶ and wavelet based^{14,15} video coding architectures recently considered by the MPEG AdHoc group on Scalable Video Coding adopt MCTF in order to reduce temporal redundancy. More specifically, wavelet-based filtering along the time axis is usually performed taking advantage of the lifting implementation. This technique enables to split direct wavelet temporal filtering into a sequence of prediction and update steps in such a way that the process is both perfectly invertible and computationally efficient. In the Haar case, the input frames are recursively processed two-by-two, according to the following formulas:

$$H = \frac{1}{\sqrt{2}}[B - W_{A \rightarrow B}(A)] \quad (1)$$

$$L = \sqrt{2}A + W_{B \rightarrow A}(H) \quad (2)$$

Where A and B are two successive frames and $W_{B \rightarrow A}(\cdot)$ is a motion warping operators that warps frame A into the coordinate system of frame B . L and H are respectively the low-pass and high-pass temporal subbands. These two lifting steps are then iterated on the L subbands of the GOP such that for each GOP we end up with only one low-pass subband. The warping operator maps every location (x, y) of the B frame into the location $(x + dx, y + dy)$ of the A frame. Equations (1) and (2) can be rewritten as:

$$H(x, y) = \frac{1}{\sqrt{2}}[B(x, y) - A(x + dx, y + dy)] \quad (3)$$

$$L(x, y) = \sqrt{2}A(x, y) + H(x - dx, y - dy) \quad (4)$$

Equations (1) and (2) refer to the t+2D scheme, where MCTF takes place in the spatial domain. In this paper we are focusing on a 2D+t architecture and the MCTF equations need to be modified to take into consideration the fact that the overcomplete representation of the wavelet transform is needed to tackle the shift variance.

$$H_i(x, y) = \frac{1}{\sqrt{2}}[B_i(x, y) - A_i^O(2^i x + dx, 2^i y + dy)] \quad (5)$$

Where A_i^O is the overcomplete wavelet-transformed reference frame subband at the wavelet decomposition level i and it has the same number of samples as the original frame. The low pass temporal subband L_i can be computed as:

$$L_i(x, y) = \sqrt{2}A_i^O(2^i x, 2^i y) + H_i^O(2^i x - dx, 2^i y - dy) \quad (6)$$

To reduce the frame rate of the decoded sequence by a factor of K , only the first $N_size/2^{K-1}$ temporal subbands need to be sent (N is the GOP size), discarding the proper number of high pass temporal subbands. If the update step is performed, the reconstructed sequence is not equivalent to the frame skipped original sequence, even if the effect of quantization is neglected, but to the low pass temporal subbands. For this reason it is of crucial importance that the update step does not introduce any visual artifacts in the updated low pass temporal subbands, as they would be visible in the sequence reconstructed at lower frame rates. In the rest of this paper we describe the proposed algorithm that allows to adaptively apply the update step only when no artifacts are introduced in the reconstructed sequence.

3. PROPOSED ALGORITHM

In the MCTF framework, the prediction step is the counterpart of motion compensated prediction in conventional closed loop schemes. On the other hand the update step can be thought as a motion compensated averaging along the motion trajectories. There are two reasons that justify the use of the update step:

- the updated frames (L) are freed from temporal aliasing artifacts that characterize frame skipped sequences
- the updated frames (L) requires fewer bits for the same quality than the original frame A because of the motion compensated denoising performed by the update step. The use of lifting update step achieves better compression efficiency, granting a good exploitation of temporal redundancies at all decomposition levels, and results in higher coding efficiency.

In spite of these benefits, many evident artifacts can be generated inside the temporal low-pass subband by the update step. In fact, when the motion model fails, it results not only in a poor compensation of current frame, but also in a mismatch in the motion field inversion, leading to the creation of annoying ghosting artifacts.

Different solutions have been proposed to overcome this problem, based on an adaptive weighting of the update step. The algorithm described in Ref. 17 adds a weighting factor to the update step, in order to reduce its effect when it is likely to add artifacts in the updated frame. The rationale behind this is that when the motion model fails to capture the scene motion high energy coefficients present in the H frame are used to erroneously update the original frame, thus introducing artifacts. The residual coefficients, resulting from the motion compensation of current frame, are used to compute their own weighting factor by means of a quadratic non-linear function. The update step thus becomes, in a t+2D scheme:

$$L = \sqrt{2}A + g(W_{B \rightarrow A}(H))W_{B \rightarrow A}(H) \quad (7)$$

where $g(\cdot)$ is a non-linear weighting function taking values in the interval $[0,1]$ that goes to zero when the energy of the high-pass frame is high (see Figure 2). Although the described technique succeeds in canceling most of the ghosting artifacts, still some irregularities remain visible in some limited areas of the picture, where the texture is particularly smooth. In this paper we suggest to enhance the above adaptive method by inserting also the local activity of reference frame in the weighting function computation. This way we try to skip the update step in those zones where texture irregularities may be more evident to visual inspection. In fact it is known that the

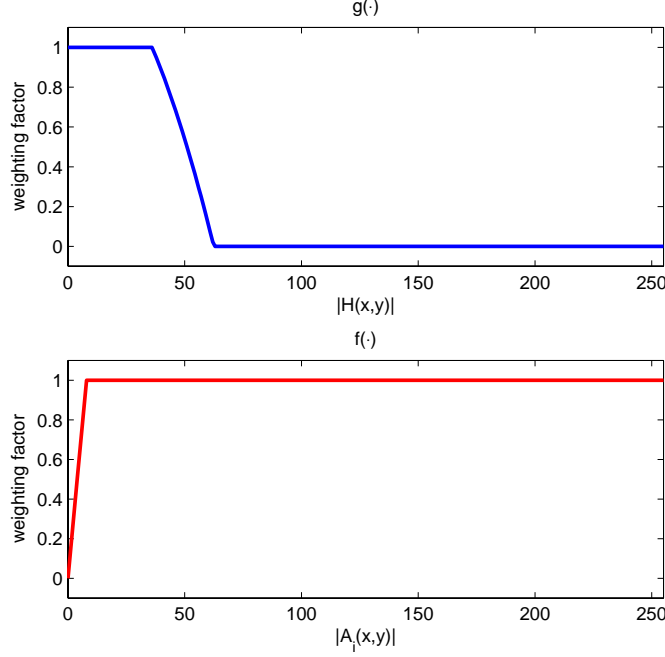


Figure 2. Adaptive update weighting functions $g(\cdot)$ ¹⁸ and $f(\cdot)$

human visual system is more sensitive to noise (artifacts) located in smooth areas. In the t+2D case, the new update step is given by:

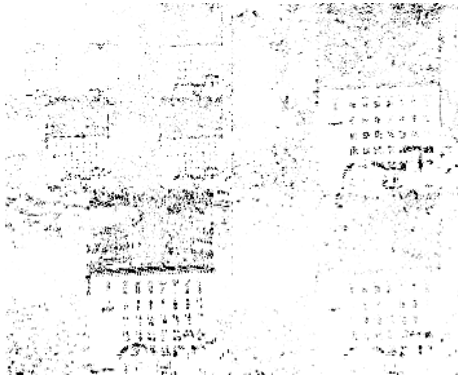
$$L = \sqrt{2}A + f(A)g(W_{B \rightarrow A}(H))W_{B \rightarrow A}(H) \quad (8)$$

In (8) $f(\cdot)$ is a function of the local image activity, that can be estimated as the local sample variance of the pixel values. This function goes to zero when the local activity is low (i.e. in smooth regions) in such a way that the update step is switched off (see Figure 2).

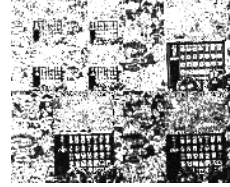
The computation of the local activity for each pixel in the reference frame, even if within a small window, adds complexity at the encoder side. In the 2D+t coding architecture we can take advantage of the fact that the update step takes place in the wavelet domain, therefore the frame to be updated is already available in its wavelet transformed version and it is possible to use this wavelet spatial decomposition to obtain an estimation of the local texture activity. After the wavelet spatial decomposition, in-band motion estimation/compensation is applied to produce the high pass temporal subbands H_i as shown in (5). Then, the computed motion field is inverted to be used in the update lifting step. For the sake of clarity let's take for example a 2-level decomposition (see Figure 4). The update filtering starts from the LL_2 spatial subband, where each coefficient local activity is estimated involving the three other coefficients at the same spatial location in LH_2 , HL_2 and HH_2 . The latter three subbands are then filtered by computing the adaptive weight from the next higher level. Therefore, if we take a coefficient in LH_2 , the variance is the mean square value of the four samples in LH_1 representing the corresponding region at higher resolution, but equal orientation. In formula, the adaptive update step for the HL_i , LH_i and HH_i subbands ($i > 1$) becomes:

$$L_i(x, y) = \sqrt{2}A_i^O(2^i x, 2^i y) + f \left(\sum_{r=0}^1 \sum_{s=0}^1 |A_{i+1}^O(2^{i+1}x + r, 2^{i+1}y + s)|^2 \right) g(H_i^O(2^i x - dx, 2^i y - dy)) H_i^O(2^i x - dx, 2^i y - dy) \quad (9)$$

No activity estimation can be calculated for LH_1 , HL_1 and HH_1 , since no higher subbands exist. Nevertheless they usually contain only little energy and the adaptivity provided by $g(\cdot)$ is sufficient to prevent undesired



Adaptive update as in Ref. 17



Proposed adaptive update

Figure 3. Weighting functions $g(\cdot)$ and $f(\cdot)$ computed for a frame of the sequence *Mobile&Calendar*. White color refers to areas where the update step is turned on. Note that $f(\cdot)$ cannot be computed for the highest frequency wavelet subbands

Table 1. *Mobile&Calendar*, CIF@15fps. Fixed update step, adaptive update¹⁷ (adaptive 1), proposed adaptive update (adaptive 2)

kbps	fixed (a)	adaptive 1 (b)	adaptive 2 (c)	(b) - (a)	(c) - (b)
2048	32.98	34.98	35.37	+2.00	+0.39
1536	32.07	33.36	33.57	+1.29	+0.21
1024	30.61	31.19	31.23	+0.58	+0.04
512	27.44	27.45	27.48	+0.01	+0.03
256	23.96	23.97	23.99	+0.01	+0.02

visual artifacts. Figure 3 shows an example of weighting factors $g(\cdot)$ and $f(\cdot)$ computed for a sample frame of the sequence *Mobile&Calendar*. At the decoder side the update step weight is computed based on the quantized reconstructed low pass temporal subbands L_i instead of A_i . For this reason there might be a mismatch between the analysis and the synthesis stage of the MCTF. The differences tend to be negligible though.

4. EXPERIMENTAL RESULTS

In order to appreciate the performance of the proposed algorithm, we computed a block based motion model using a fast search algorithm, which in general differs from the best motion field computed with a full search approach. Under these conditions, ghosting artifacts tend to be more visible if the update step is not adaptively turned off. Visual inspection of sequences decoded at reduced frame rate shows how the proposed method manages to completely delete visible artifacts inside smooth areas of the picture, thus leading to an overall improved performance. Figure 5 and Figure 6 show the *Mobile&Calendar* sequence decoded at 15fps and 7.5fps in three cases: fixed update step, adaptive update as in,¹⁷ proposed adaptive update. Table 1 and Table 2 report PSNR values of the reconstructed *Mobile&Calendar* sequence whereas Table 3 and Table 4 for the *Foreman* sequence. It is important to interpret these numbers carefully, since changing the update step algorithm directly affects the MCTF decomposition. For this reason, in the computation of the PSNR, we decided to use the original frame skipped sequence as reference. It has been recently shown¹⁹ that there is quite a weak relationship between visual quality and objective PSNR measurements unless the reference stays the same. For this reason we deem visual inspection of reconstructed sequences more meaningful in this case.

5. CONCLUSIONS

Conclusions In this paper we propose a novel adaptive update algorithm that takes into consideration the local content of the frame to be updated in such a way that the update step weighing factor is reduced when the local activity is low, therefore when ghosting artifacts are more likely to be perceived. In order to add minimum complexity at the encoder, we perform this operation in the wavelet domain, exploiting the already computed wavelet coefficient to estimate the local activity.

Table 2. *Mobile&Calendar*, CIF@7.5fps. Fixed update step, adaptive update¹⁷ (adaptive 1), proposed adaptive update (adaptive 2)

kbps	fixed (a)	adaptive 1 (b)	adaptive 2 (c)	(b) - (a)	(c) - (b)
2048	29.81	35.46	36.92	+5.65	+1.46
1536	29.59	34.51	35.48	+4.92	+0.97
1024	28.98	32.62	33.12	+3.64	+0.50
512	27.08	28.71	28.81	+1.63	+0.10
256	24.32	24.86	24.89	+0.54	+0.03
128	21.61	21.78	21.79	+0.17	+0.01

Table 3. *Foreman*, CIF@15fps. Fixed update step, adaptive update¹⁷ (adaptive 1), proposed adaptive update (adaptive 2)

kbps	fixed (a)	adaptive 1 (b)	adaptive 2 (c)	(b) - (a)	(c) - (b)
2048	37.07	39.05	40.28	+1.98	+1.23
1536	36.55	38.31	39.27	+1.76	+0.96
1024	35.61	37.04	37.60	+1.43	+0.56
512	33.68	34.52	34.77	+0.84	+0.25
256	31.40	31.82	31.92	+0.42	+0.10
128	28.95	29.15	29.19	+0.20	+0.04

ACKNOWLEDGMENTS

The authors wish to acknowledge the support provided by the European Network of Excellence VISNET.

REFERENCES

1. J. R. Ohm, "3-d subband coding with motion compensation," *IEEE Transactions on Information Theory* **3**, pp. 559–571, September 1994.
2. D. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Boston, MA, 2002.
3. A. Said and W. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuit and Systems for Video Technology* **6**, pp. 243–250, June 1996.
4. S.-T. Hsiang and J. W. Woods, "Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, **3**, pp. 662–665, (Geneva, Switzerland), May 2000.
5. S. L. Ji-Zheng Xu and Y.-Q. Zhang, "Three-dimensional shape-adaptive discrete wavelet transforms for efficient object-based video coding," in *Visual Communications and Image Processing*, (Perth, Australia), June 2000.
6. P. Chen and J. W. Woods, "Bidirectional mc-ezbc with lifting implementation," *IEEE Transactions on Circuit and Systems for Video Technology* **14**, pp. 1183–1194, October 2004.
7. A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (limat) framework for highly scalable video compression," *IEEE Transactions on Image Processing*, 2004.
8. G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, "Motion compensation and scalability in lifting-based video coding," *EURASIP Signal Processing: Image Communication*, pp. 577–600, August 2004.
9. Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis, and P. Schelkens, "In-band motion compensated temporal filtering," *Signal Processing: Image Communication* **19**, pp. 653–673, August 2004.
10. Y. Wang, S. Cui, , and J. E. Fowler, "3D video coding using redundant-wavelet multihypothesis and motion-compensated temporal filtering," in *Proceedings of the International Conference on Image Processing*, **2**, pp. 755–758, (Barcelona, Spain), September 2003.
11. H.-W. Park and H.-S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Transactions on Image Processing* **9**, pp. 577–587, April 2000.
12. J. C. Ye and M. van der Schaar, "Fully scalable 3-D overcomplete wavelet video coding using adaptive motion compensated temporal filtering," in *Visual Communications and Image Processing*, T. Ebrahimi and T. Sikora, eds., pp. 1169–1180, Proc. SPIE 5150, (Lugano, Switzerland), July 2003.
13. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

Table 4. Foreman, CIF@7.5fps. Fixed update step, adaptive update¹⁷ (adaptive 1), proposed adaptive update (adaptive 2)

kbps	fixed (a)	adaptive 1 (b)	adaptive 2 (c)	(b) - (a)	(c) - (b)
2048	33.03	37.76	40.41	+4.73	+2.65
1536	32.92	37.50	39.84	+4.58	+2.34
1024	32.64	36.86	38.68	+4.22	+1.82
512	31.82	35.16	36.09	+3.34	+0.93
256	30.51	32.76	33.18	+2.25	+0.42
128	28.85	30.24	30.44	+1.39	+0.20
64	26.82	27.61	27.68	+0.79	+0.07
32	24.45	24.77	24.78	+0.32	+0.01

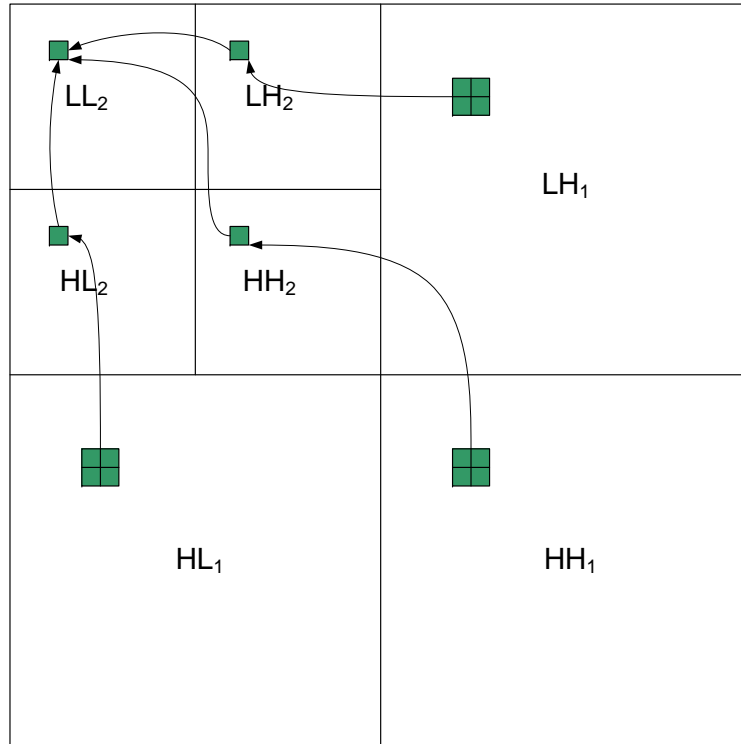


Figure 4. Each arrow connect the wavelet coefficient used to compute the update weight of the coefficient the arrow points to

14. N. Mehrseresht and D. Taubman, "A flexible structure for fully scalable motion compensated 3D-DWT with emphasis on the impact of spatial scalability." Submitted to IEEE Trans. Image Processing.
15. N. Mehrseresht and D. Taubman, "An efficient content-adaptive motion compensated 3D-DWT with enhanced spatial and temporal scalability." Submitted to IEEE Trans. Image Processing.
16. "Scalable Video Model version 3.0." ISO/IEC JTC1/WG11 Doc. N6716, November 2004.
17. N. Mehrseresht and D. Taubman, "Adaptively weighted update steps in motion compensated lifting based on scalable video compression," in *Proceedings of the International Conference on Image Processing*, (Barcelona, Spain), September 2003.
18. D. Taubman, D. Maestroni, R. Mathew, and S. Tubaro, "Svc core experiment 1, description of unsw contribution." ISO/IEC JTC1/SC29/WG11, MPEG2004/M11441, October 2004.
19. U. Benzler and M. Wien, "Results of svc ce3 (quality evaluation)." ISO/IEC JTC1/SC29/WG11, MPEG2004/M10931, July 2004.



(a) fixed update



(b) adaptive update as in Ref. 17



(c) proposed adaptive update

Figure 5. *Mobile&Calendar* CIF@15fps - 1024kbps



(a) fixed update



(b) adaptive update as in Ref. 17



(c) proposed adaptive update

Figure 6. *Mobile&Calendar* CIF@7.5fps