H.264-Compatible Coding of Background Soccer Video using Temporal Subbands

Xiaohua Lu, Haopeng Li and Markus Flierl
School of Electrical Engineering
KTH Royal Institute of Technology, Stockholm
{*xiaohual, haopeng, mflierl*}*@kth.se*

***Abstract*—This paper presents an H.264-compatible temporal subband coding scheme for static background scenes of soccer video. We utilize orthonormal wavelet transforms to decompose a group of successive frames into temporal subbands. By exploiting the property of energy conservation of orthonormal wavelet transforms, we construct a rate distortion model for optimal bitrate allocation among different subbands. To take advantage of the high efficiency video codec H.264/AVC, we encode each subband with H.264/AVC FRExt intra-coding by assigning optimal bitrates. The experimental results show that our proposed coding scheme outperforms conventional video coding with H.264/AVC for both subjective and objective evaluation.**

***Keywords*- Content-adaptive coding; temporal subband transforms; rate-distortion model.**

I. INTRODUCTION

Content-based coding can be efficient for video coding [1]. In particular, content-adaptive coding for immersive networked experience of soccer games is advantageous [2]. This approach distinguishes among multiple dynamic content items (e.g. player) and a static content item (background). The corresponding sub-sequences are extracted from the input video sequence. The bitrates among static and dynamic content items are optimally allocated due to their different statistical properties. Further, this scheme provides more flexibility by allowing users to access the individual content freely, together with an improved rate-distortion performance.

The areas of the input soccer video depicting the field and the background object are varying slowly over time. As proposed in [2], we reduce the frame rate of the static content to improve the overall rate-distortion performance and encode only the temporal average of the previous input frames. In other words, a temporal transform is applied to the input signal and the mean component is encoded with H.264/AVC.

The static properties of the background scene are such that the energy of the input frames is accumulated in the temporal low-band while the temporal high-bands carry only very small amounts of energy [3]. Due to efficient energy compaction [4][5], temporal subband coding schemes are good candidates for this application. On the other hand, as the background scenes are captured by static cameras, co-located pixels in successive frames usually refer to the same part of an object. Therefore, high coding efficiency at low computational complexity can be achieved without motion compensation.

One example of subband video coding uses 3D Set Partitioning in Hierarchical Trees (3D-SPIHT) [6], which essentially is an extension of SPIHT coding [7]. It fully utilizes the spatial correlation within the current frame as well as the temporal correlation among successive frames. To exploit the spatial correlation, a 2D discrete wavelet transform is applied to the current frame. Then, the Karhunen-Loeve transform (KLT) is applied to each corresponding position in the temporal domain to remove the redundancy among successive temporal subbands. Moreover, it uses vector quantization (VQ) to quantize the cofficients in each subband [8], which improves coding efficiency significantly. However, this increases the computational complexity of this algorithm significantly.

In this paper, we propose a H.264-compatible temporal subband coding scheme for background soccer video. We utilize orthonormal wavelet transforms to decompose groups of successive frames into temporal subbands. To allocate the bitrate among different subbands optimally, we construct a rate distortion model based on energy conservation of orthonormal wavelets [9][10]. Finally, subbands are encoded with the H.264 FRExt intra mode and optimally allocated bitrates.

II. H.264-COMPATIBLE TEMPORAL SUBBAND CODING

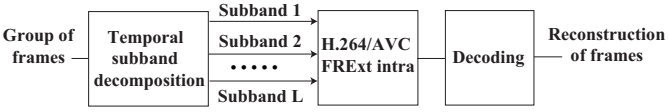


Figure 1. H.264-compatible temporal subband coding.

In this section, we propose an H.264-compatible temporal subband coding algorithm for static background video. As shown in Fig. 1, the background temporal subband coding scheme comprises the subband transform of temporal frames, the optimal bit allocation among subbands, the H.264-compatible coding engine, and an inverse transform unit at the decoder that facilitates the reconstruction of the output image sequence.

A. Temporal Wavelet Transform

To match the properties of the background scenes, we eliminate the temporal redundancy among successive frames by using wavelet transforms. Let the size of one GOP in the original image sequence be denoted by L and the set of pixel positions in one frame by $\mathcal{M} = \{(i,j)\}$. Let one pixel in the k-th frame be denoted by $x(i,j,k)$, where (i,j) indicates the pixel coordinate in the image, with $1 \leq k \leq L$. Furthermore, we assume that L is dyadic, $L = 2^N$, for any non-negative integer N.

To implement the temporal wavelet transform, we define the input vector $X(i,j)$ as the pixels at the same image coordinate (i,j) in one GOP

$$X(i,j) = [x(i,j,1), \ldots, x(i,j,L)]. \tag{1}$$

Each input vector is transformed by a N-level wavelet transform, resulting in L wavelet coefficients

$$[w(i,j,1), \ldots, w(i,j,L)] = W_N[x(i,j,1), \ldots, x(i,j,L)], \tag{2}$$

where W_N denotes the N-level wavelet transform. Note, the 1D wavelet transform is always applied recursively to the lower subband. The selection of the wavelet basis will be addressed in Section III-A.

With above temporal wavelet transform, we obtain L subbands of frequency coefficients by applying (2) to all pixels $x(i,j,k)$. Due to the static properties of the background scenes, the energy of the input frames is accumulated in the temporal low-band while the energy in the temporal high-bands is relatively small. This ensures good temporal scalability since the dependencies among subbands are very weak.

B. Rate-Distortion Model

With the temporal subband transform, the static input video frames are transformed into subbands of wavelet coefficients. Due to the weak dependency among the coefficient subbands, we consider to use intra-coding techniques to encode them. On the other hand, to generate an H.264-compatible output stream, the H.264 FREXt intra-coding method is a good candidate for this purpose.

An efficient coding scheme allocates the bitrate optimally. In particular for H.264/AVC intra-coding, allocation among subbands becomes a crucial problem. To accomplish this, we introduce a rate-distortion model that reflects the trade-off among subbands.

For our rate-distortion model, we use the mean square error to determine the average reconstruction distortion per pixel

$$D_{mse} = \frac{1}{L}\frac{1}{|\mathcal{M}|}\sum_{k=1}^{L}\sum_{i,j\in\mathcal{M}}(x(i,j,k) - \hat{x}(i,j,k))^2, \tag{3}$$

where $\hat{x}(i,j,k)$ is the reconstructed pixel. Exploiting energy conservation of orthonormal wavelet transforms [10][9], the average reconstruction distortion per pixel D_{mse} is equal to the average coding distortion of L coefficient subbands

$$\begin{aligned} D_{mse} &= \frac{1}{L}\frac{1}{|\mathcal{M}|}\sum_{k=1}^{L}\sum_{i,j\in\mathcal{M}}(x(i,j,k) - \hat{x}(i,j,k))^2 \\ &= \frac{1}{L}\frac{1}{|\mathcal{M}|}\sum_{k=1}^{L}\sum_{i,j\in\mathcal{M}}(w(i,j,k) - \hat{w}(i,j,k))^2 \\ &= \frac{1}{L}\sum_{k=1}^{L} D^k_{coef}, \end{aligned} \tag{4}$$

where D^k_{coef} is the average coding distortion of coefficients in the k-th subband, $w(i,j,k)$ are the transformed coefficients and $\hat{w}(i,j,k)$ their reconstructed values.

As mentioned earlier, each subband has different properties due to the energy distribution. Thus, it is reasonable to allocate bitrates depending on the content of each subband. For our model, we allocate average bitrates in bits per pixel R_k for the k-th subband, where $1 \leq k \leq L$.

Combing the property of orthonormal wavelet transforms in (4), we obtain the distortion rate function for our H.264-compatible temporal subband coding scheme as

$$D_{mse}(R_1, R_2, \ldots) = \frac{1}{L}\sum_{k=1}^{L} D^k_{coef}(R_k). \tag{5}$$

C. Optimal Bit Allocation

With above rate-distortion model, we are able to find the optimal rate distortion trade-off for L coefficient subbands. For that, we assume that the individual distortion rate functions $D^k_{coef}(R_k)$ are convex. Further, we impose a bandwidth constraint. Let the constant W be the bandwidth which is allocated to the input image sequence. Let f be the frame rate for the input video.

1) Minimizing Cost Function: The optimal trade-off is obtained by minimizing the average reconstruction distortion, subject to the imposed bandwidth constraint

$$\begin{aligned} \min \quad & D_{mse}(R_1, R_2, \ldots) \\ s.t. \quad & \frac{f}{L}\sum_{k=1}^{L} R_k|\mathcal{M}| \leq W. \end{aligned} \tag{6}$$

The distortion D_{mse} can be rewritten as the average coding distortion of L coefficient subbands

$$\begin{aligned} \min \quad & \sum_{k=1}^{L} D^k_{coef}(R_k) \\ s.t. \quad & \frac{f}{L}\sum_{k=1}^{L} R_k|\mathcal{M}| \leq W. \end{aligned} \tag{7}$$

This constrained problem can be solved by Lagrangian relaxation and leads to the Pareto condition for our H.264-compatible temporal subband coding scheme

$$\frac{dD^1_{coef}}{dR_1} = \frac{dD^k_{coef}}{dR_k} = -\lambda \quad \text{for} \quad k = 1, 2, \ldots, L, \tag{8}$$

where the non-negative λ determines the slope of the distortion rate curve of the individual subbands.

2) Approximation of $QP - \lambda$ Relationship: With the Pareto condition in (8), we are able to calculate the optimal bitrate of each subband for a given λ. However, this requires the knowledge of the rate distortion relationship of individual subbands which may be obtained by pre-encoding each subband at different bitrates.

Since our coding scheme is H.264 compatible, intra-coding is controlled by assigning a quantization parameter QP [11] to each individual subband. This motivates us to find a relationship between QP and λ. By doing so for a given λ, the optimal bitrate can be achieved by simply computing the corresponding QP value, which avoids the heavy computational burden of pre-encoding the subbands.

For intra-frame coding, we assume that the total rate R_k comprises the rate of the intra-frame predictor R_k^p and the rate of the residual encoder R_k^r [12][13]. Therefore, we have the relationship

$$\lambda = -\frac{dD_{coef}^k}{dR_k} = -\frac{\partial D_{coef}^k}{\partial R_k^p} = -\frac{\partial D_{coef}^k}{\partial R_k^r}. \quad (9)$$

A detailed discussion of (9) is given in Appendix A.

We assume the reconstruction error of intra-frame coding to be a memoryless Gaussian signal with distortion-rate function [13]

$$D_{coef}^k(R_k^p, R_k^r) = \sigma^2(R_k^p)2^{-2R_k^r} = \sigma^2(R_k^p)e^{-R_k^r 2ln2}, \quad (10)$$

where $\sigma^2(R_k^p)$ is the variance of the intra prediction error. The partial derivative gives the relation

$$\lambda = D_{coef}^k(R_k^p, R_k^r)2ln2. \quad (11)$$

Compared to the rate of the residual encoder R_k^r, the rate of intra prediction is small in the high rate case. The distortion $D(R_k^p, R_k^r)$ is caused only by the residual encoder. Assuming a high-rate uniform scalar quantizer, we have

$$D_{coef}^k(R_k^p, R_k^r) = \frac{Q^2}{12}, \quad (12)$$

where Q is the step-size of the quantizer. Therefore, we can derive the relationship between λ and Q as

$$\lambda = D_{coef}^k(R_k^p, R_k^r)2ln2 = \frac{ln2}{6}Q^2 \approx 0.1Q^2. \quad (13)$$

The step-size of the residual quantizer Q in H.264/AVC is determined by the quantization parameter QP [11]

$$Q = 0.625 \times 2^{\frac{QP}{6}}, \quad (14)$$

which leads to

$$\lambda = 0.1 \times (0.625 \times 2^{\frac{QP}{6}})^2 = \rho \times 4^{\frac{QP}{6}}, \quad (15)$$

where ρ is approximately 0.04. In our work, we slightly adjust it to $\rho = 0.025$ according to empirical trials. Therefore, we obtain an approximation of the relationship between λ and QP. The optimal bitrate for each subband can be easily achieved by assigning a QP for a given λ. Furthermore, this approximation also reveals that the optimal QP is identical for all subbands in the high rate scenario.

III. Implementation Issues

In this section, we discuss some practical issues regarding the selection of the transform basis and the GOP size for our H.264-compatible temporal subband coding scheme.

A. Transform Basis

To guarantee an efficient orthonormal wavelet transform, we should select an appropriate wavelet basis for our temporal wavelet transform. On the other hand, we also have to consider the computational complexity of the implementation which depends on the length of the wavelet basis. Therefore, we investigate three kinds of transform bases for comparison: Daubechies-1 (Haar), Daubechies-9 and DCT.

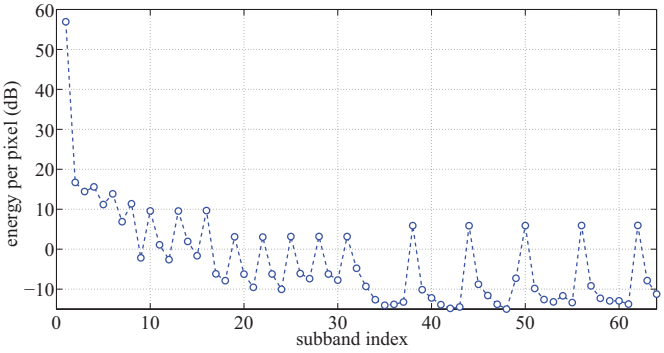
Daubechies wavelets are widely used in image and video coding due to their orthonormal property and high efficiency [9]. Therefore, we consider two Daubechies wavelets with filter length 2 and 18, namely Daubechies-1 (or Haar) and Daubechies-9. Apparently, the length of Daubechies-1 is smaller than that of Daubechies-9, resulting in a lower computational complexity. On the other hand, the DCT is another orthonormal transform which performs quite close to the KLT with complexity $O(N \log N)$, compared to $O(N)$ of the wavelet transform. Note, to implement the DCT transform, the N-level wavelet transform is replaced by the DCT of length L.

We look at the energy compaction to compare the performance of different transform bases. The results are shown in Fig. 2. We observe that the results for different transforms are very similar and more than 99.95% of the energy of the whole GOP is concentrated into the first subband. Therefore, we choose the Daubechies-1 (or Haar) wavelet basis for our coding scheme due to its lower computational complexity.

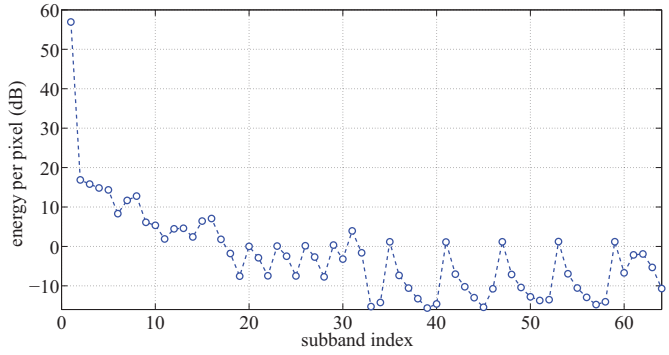
B. GOP Size

The size L of the GOP is also an important factor for our H.264-compatible temporal subband coding scheme. On one hand, a long GOP size accumulates more energy in the lower subbands and improves the overall rate-distortion performance significantly. However, varying lighting conditions are challenging for a longer duration. They may break our static background model, leading to worse energy compaction and, hence, lower the coding efficiency. In addition, long GOP sizes introduce long encoding and decoding delays. On the other hand, short GOP sizes may not be sufficient to exploit the temporal correlation among frames.

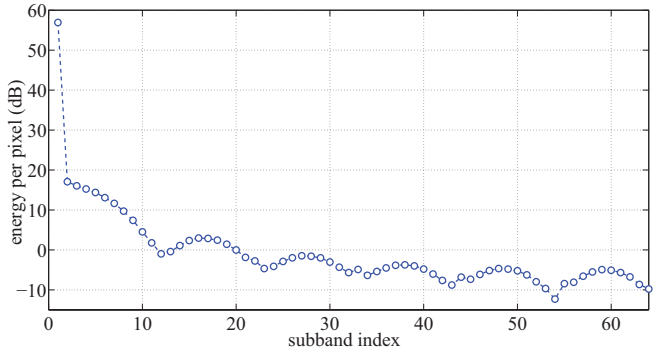
Therefore, we selectively test our coding scheme by setting different GOP sizes, as shown in Fig. 3. The coding



(a) Daubechies-1 wavelet



(b) Daubechies-2 wavelet



(c) DCT

Figure 2. Energy distribution for different transform bases. (64 frames, test sequence $Barca - St.Andreu$)

performance improves when increasing the GOP size, especially at low bitrates. However, considering the coding delay, which is introduced by longer GOP lengths, we choose 64 frames for the GOP size to balance efficiency and delay.

IV. Experimental Results

In this section, we evaluate our H.264-compatible temporal subband coding scheme by comparing it with conventional H.264/AVC predictive coding [14]. We evaluate our H.264-compatible temporal subband coding scheme for the soccer video test sets *Barca-St. Andreu* and *LasPalmas* which are provided by the MEDIAPRO group. The videos are captured by fixed cameras with resolution 1920×1080 at 25 fps.

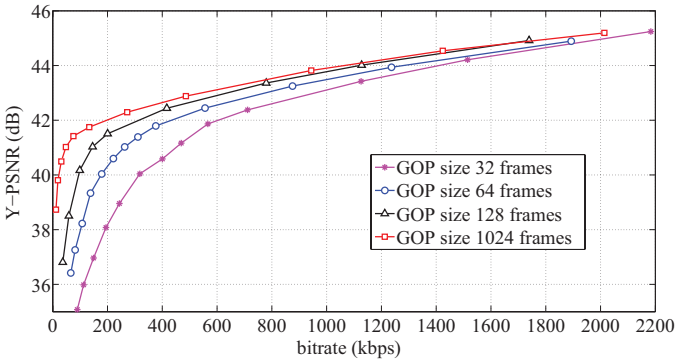


Figure 3. Performance comparison for different GOP sizes. (test sequence $Barca - St.Andreu$, Daubechies-1 wavelet)

The dynamic content items (i.e., players and ball) have been extracted and encoded by using content-adaptive coding [2]. Therefore, the input video sequence for our scenario is only the static background scene. For intra coding of the coefficient subbands, we use the Fidelity Range Extension (FRExt) profile in JM 16.0 [14]. In other words, up to 12 bits are considered for intra coding of subband coefficients.

A. $QP - \lambda$ Relationship

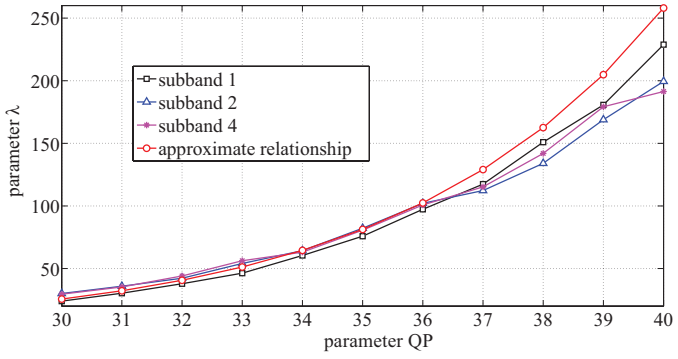
In Section II-C2, we obtained $\lambda = 0.025 \times 4^{\frac{QP}{6}}$ as an approximate relationship between QP and λ under high-rate assumptions. Here, we use experimental results to verify this relationship.

We choose the target bitrate of our coding scheme in a reasonable range from 300 kbps to 2000 kbps. From empirical trials, we obtain a corresponding range of QP values $[QP_{low}, QP_{high}]$ for the first subband. Knowing the individual rate-distortion performance of each subband, we choose $QP_1 \in [QP_{low}, QP_{high}]$ for the first subband and calculate the slope λ. Next, we obtain QP_k for the remaining $L-1$ subbands by finding above slope λ on the corresponding rate-distortion curves. As shown in Fig. 4, the actual $QP - \lambda$ relationship fits the approximate relationship rather well for the high-rate scenario.

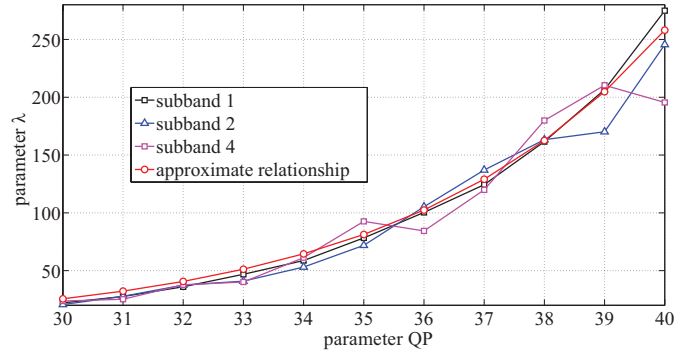
B. Comparison with H.264/AVC

We compare our H.264-compatible temporal subband coding scheme to predictive video coding as implemented by the H.264/AVC codec JM 16.0 [14]. The picture structure for the JM 16.0 is chosen to be "$IPBPBP...$" with a GOP size of 64. We use 64 successive frames from the test sequences. As discussed in Section III-A and III-B, we use the Daubechies-1 (Haar) wavelet basis and a GOP size of 64 for our coding scheme. The optimal QP is computed by our approximation (15).

1) Subjective Results: Here we show a subjective comparison with between two patches in a frame at the same bitrate. For our H.264-compatible temporal subband coding



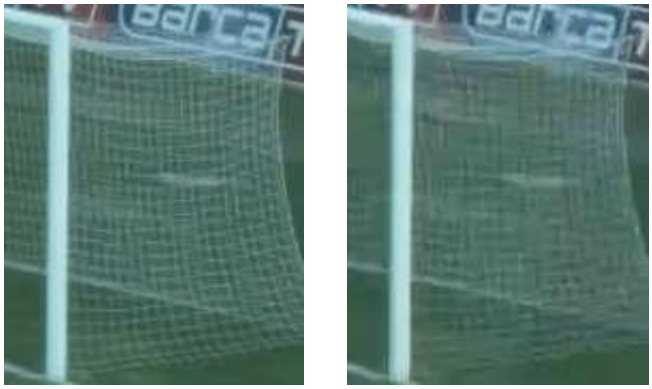
(a) Test sequence *Barca-St. Andreu*



(b) Test sequence *LasPalmas*

Figure 4. $QP - \lambda$ for the first few subbands. (64 frames; Daubechies-1 wavelet)

scheme, we have much better visual quality than the reference scheme. In particular, the net of the goal and the upper right corner of the soccer field are compared in Figs. 5 and 6.



(a) Proposed scheme **(b) JM 16.0**

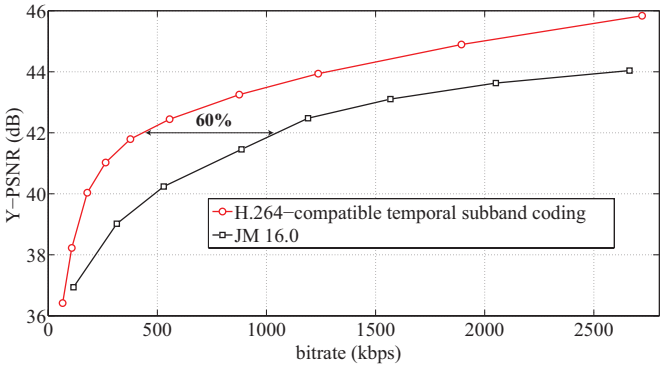
Figure 5. A patch of the reconstructed frame (the net of the goal). (test sequence $Barca - St.Andreu$, bitrate = 500 kbps)

2) Objective Results: To evaluate the performance of the H.264-compatible temporal subband coding scheme, we measure the overall reconstruction quality as luminance PSNR (Y-PSNR) over the total bitrate for our test sequences.

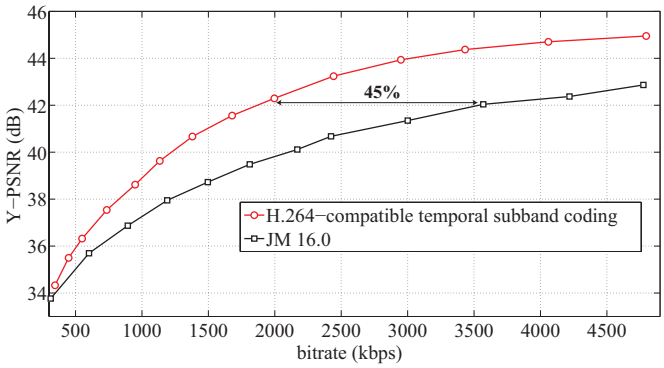
(a) Proposed scheme

(b) JM 16.0

Figure 6. A patch of the reconstructed frame (the upper right corner of the soccer field). (test sequence $LasPalmas$, bitrate = 1200 kbps)



(a) Test sequence *Barca-St. Andreu*



(b) Test sequence *LasPalmas*

Figure 7. Performance comparison between our scheme and JM 16.0.

As shown in Fig. 7, our H.264-compatible temporal subband coding scheme outperforms JM 16.0. At 42 dB Y-PSNR, our scheme saves up to 60% bitrate for the first sequence and 45% bitrate for the second sequence.

V. Conclusions

We discussed an H.264-compatible temporal subband coding scheme for the static background content of soccer video. To match the static properties of the background scenes, we eliminate the temporal redundancy among successive frames by using a temporal wavelet transform. This accumulates efficiently the energy into the temporal low-band while the temporal high-bands show very low amounts of energy. To build an H.264 compatible coding scheme, we encode each subband by using H.264/AVC FRExt intra coding. With energy conservation of orthonormal wavelet transforms, we construct a rate-distortion model for subband coding and obtain the optimal bitrate allocation among subbands. Furthermore, we derive an approximate relationship between the Lagrange multiplier λ and the quantization parameter QP to implement an efficient bitrate allocation. The experimental results show that our H.264-compatible temporal subband coding scheme outperforms conventional predictive coding with H.264/AVC for both subjective and objective evaluation.

VI. Acknowledgments

This work was supported in part by the European Commission in the context of the project ICT-FP7-248020 "FINE – Free-Viewpoint Immersive Networked Experience". We thank the MEDIAPRO group for providing multiview video test data.

Appendix A.
Remarks to Relation (9)

The total bitrate in the k-th subband R_k is the combination of the rate of the intra-predictor R_k^p and the residual encoder R_k^r. We have

$$dD_{coef}^k = \frac{\partial D_{coef}^k}{\partial R_k^p} dR_k^p + \frac{\partial D_{coef}^k}{\partial R_k^r} dR_k^r, \tag{16}$$

$$dR_k = dR_k^p + dR_k^r. \tag{17}$$

With the Pareto condition given in (8), we obtain

$$\begin{aligned} dD_{coef}^k + \lambda dR_k &= (\frac{\partial D_{coef}^k}{\partial R_k^p} + \lambda) dR_k^p \\ &\quad + (\frac{\partial D_{coef}^k}{\partial R_k^r} + \lambda) dR_k^r \\ &= 0. \end{aligned} \tag{18}$$

Therefore,

$$\lambda = -\frac{dD_{coef}^k}{dR_k} = -\frac{\partial D_{coef}^k}{\partial R_k^p} = -\frac{\partial D_{coef}^k}{\partial R_k^r}. \tag{19}$$

References

[1] P. Ndjiki-Nya, T. Hinz, A. Smolic, and T. Wiegand, "A generic and automatic content-based approach for improved H.264/MPEG4-AVC video coding," in *Proc. of the IEEE International Conference on Image Processing*, Sept. 2005, pp. II – 874–7.

[2] H. Li and M. Flierl, "Rate-distortion-optimized content-adaptive coding for immersive networked experience of sports events," in *Proc. of the IEEE International Conference on Image Processing*, Sept. 2011, pp. 3237–3240.

[3] M. Flierl, "Adaptive spatial wavelets for motion-compensated orthogonal video transforms," in *Proc. of the IEEE International Conference on Image Processing*, Nov. 2009, pp. 1045–1048.

[4] J.R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. on Image Processing*, vol. 3, pp. 559–571, Sept. 1994.

[5] M. Flierl and B. Girod, "A motion-compensated orthogonal transform with energy-concentration contraint," in *Proc. of the IEEE Workshop on Multimedia Signal Processing*, Oct. 2006, pp. 391–394.

[6] B. Kim and W. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. of the IEEE Data Compression Conference*, Mar. 1997, pp. 251–260.

[7] A. Said and W. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," in *IEEE Trans. on Circuits and Systems for Video Technology*, June 1996, pp. 243–250.

[8] P. Dragotti, G. Poggi, and A. Ragozini, "Compression of multispectral images by three-dimensional SPIHT algorithm," in *IEEE Trans. on Geoscience and remote sensing*, Jan. 2000, vol. 38.

[9] R. Gonzalez and R. Woods, *Digital Image Processing (3rd Edition)*, Pearson Education, 2007.

[10] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, pp. 909–996, Oct. 1988.

[11] ITU-T and ISO/IEC Joint Video Team, *ITU-T Rec. H.264 – ISO/IEC 14496-10 AVC : Advanced Video Coding for Generic Audiovisual Services*, 2005.

[12] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, 1998.

[13] M. Flierl and B. Girod, *Video Coding with Superimposed Motion-Compensated Signals: Applications to H.264 and Beyond*, Springer, 2010.

[14] "ITU-T and ISO/IEC, H.264/AVC reference software: JM 16.0," http://iphome.hhi.de/suehring/tml/, July 2009.