Content Adaptive Wavelet Lifting for Scalable Lossless Video Coding

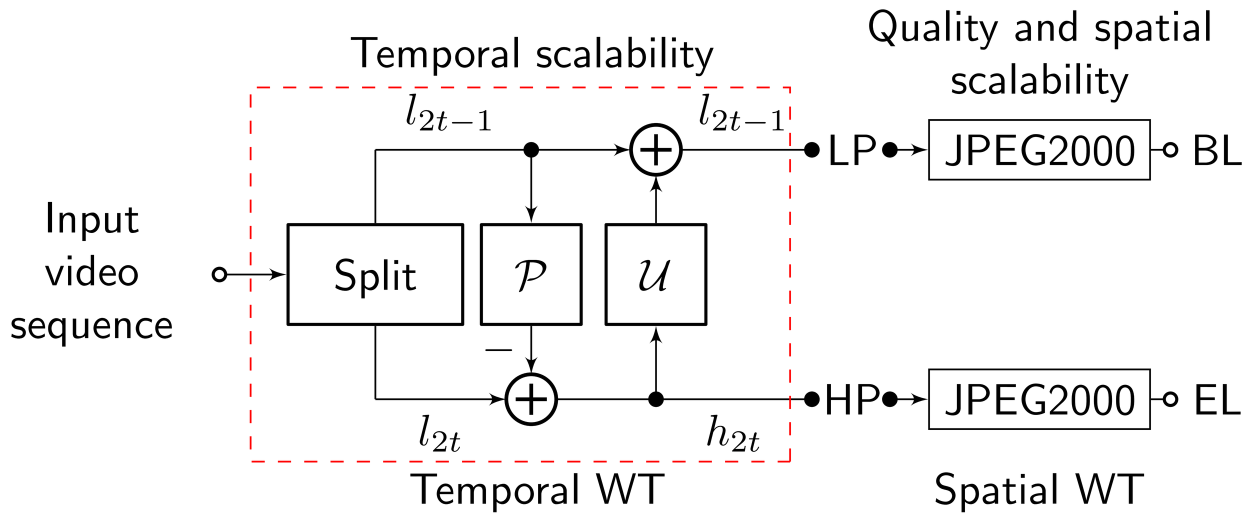
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Daniela Lanz, Christian Herbert, and André Kaup

Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstr. 7, 91058 Erlangen, Germany

1. Introduction

- **Task:** Professional applications often require lossless compression
- **Challenge:** Lossless compression leads to high bit rates
- **Solution:** Scalable lossless video coding based on transmitting a base layer (BL) with coarser quality and one or more enhancement layers (ELs), comprising the residual video data
- **Approach:** 3-D subband coding based on Wavelet Transforms (WT) [1]



- By realizing $\mathcal{P}$ as the warping operator $\mathcal{W}$, Motion Compensated Temporal Filtering (MCTF) is achieved [2]:

$$h_{2t} = l_{2t} - \lfloor \mathcal{W}_{2t-1\to 2t}(l_{2t-1}) \rfloor$$
$$l_{2t-1} = l_{2t-1} + \left\lfloor \frac{1}{2} \mathcal{W}_{2t\to 2t-1}(h_{2t}) \right\rfloor$$

2. Content Adaptive Wavelet Lifting (CA-WL)

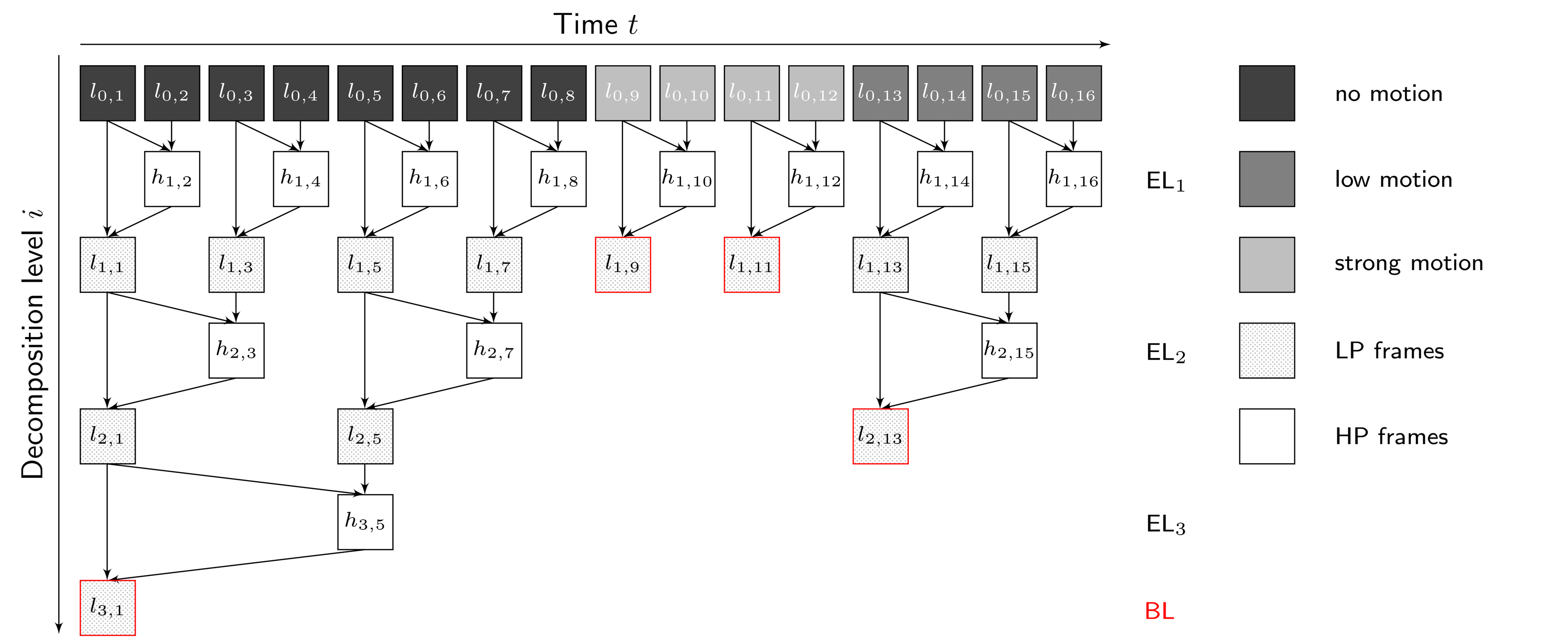
- **Idea:** Adaptive temporal scaling based on significant changes among subsequent frames
- **Stopping Criterion:**
  - Haar WTs can be represented with tree structures
  - With each node a basis vector $\boldsymbol{b}_{i,t}$ and a wavelet coefficient vector $\boldsymbol{c}_{i,t}$ is associated, which is the inner product of the signal $\boldsymbol{s}$ with the basis $\boldsymbol{b}_{i,t}$
  - If combined costs of child nodes exceed costs of parent node, i.e.

$$\mathcal{C}(\boldsymbol{s}, \boldsymbol{b}_{i,[2t-1,2t]}) \leq \mathcal{C}(\boldsymbol{s}, \boldsymbol{b}_{i+1,2t-1} \cup \boldsymbol{b}_{i+1,2t}),$$

  the child nodes shall be pruned from the tree
  - $\mathcal{C}(\cdot)$ describes a Lagrangian cost functional, which represents the coding costs:

$$\mathcal{C}(\boldsymbol{s}, \boldsymbol{b}) = D(\boldsymbol{s}, \boldsymbol{b}) + \lambda R(\boldsymbol{s}, \boldsymbol{b})$$

  - Rate $R(\boldsymbol{s}, \boldsymbol{b})$ is composed of the required rate for lossless coding of the LP and HP frames and, in case of MC, the file size of the motion vectors
  - Distortion $D(\boldsymbol{s}, \boldsymbol{b})$ is calculated by the MSE of the corresponding wavelet coefficients compared to the original signal according to [3]



- **Handling of the Overhead:**
  - Realized by transmitting a vector $\boldsymbol{v}$, whose length equals the number of input frames:

Initialize $\boldsymbol{v}$:	$(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$
$\boldsymbol{v}$ after level i=1 :	$(1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0)$
$\boldsymbol{v}$ after level i=2 :	$(2,0,0,0,2,0,0,0,1,0,1,0,2,0,0,0)$
$\boldsymbol{v}$ after level i=3 :	$(3,0,0,0,0,0,0,0,1,0,1,0,2,0,0,0)$

  - Non-zero entries correspond to the number of applied decomposition levels i
  - Distance d to the corresponding HP frame is given by d=2^{i-1}
  - Encoded using multiple-context adaptive arithmetic coding [4]

3. Experimental Results

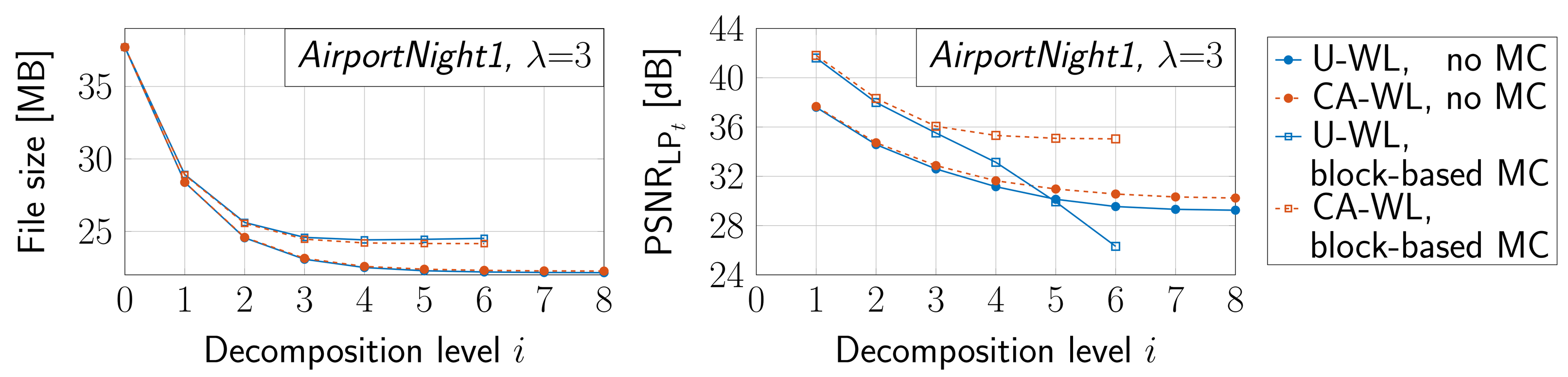
- **Simulation Setup (8 bpp):**

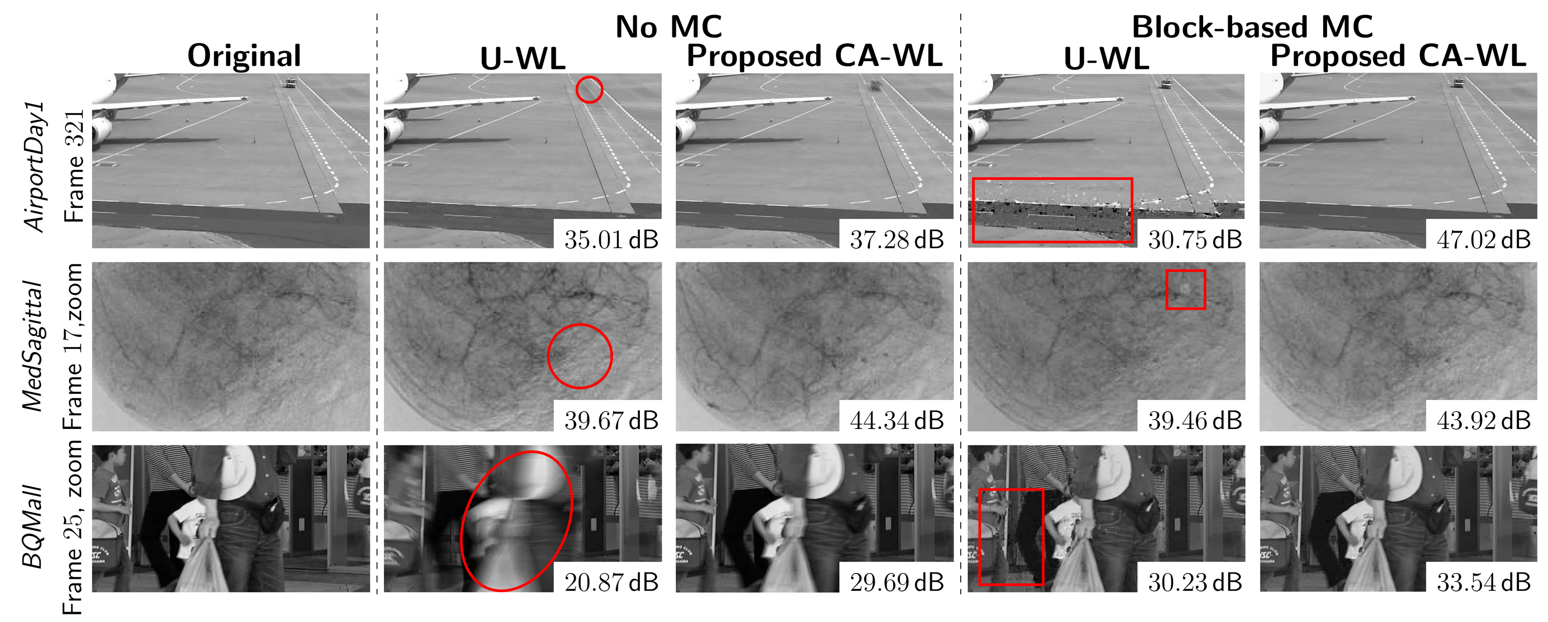
		Spatial resolution	Number of frames
Surv	*AirportNight1*	688 × 352	500
	AirportNight2	688 × 432	500
	AirportNight3	688 × 372	500
	AirportDay1	688 × 432	500
Med	*MedFrontal*	512 × 512	29
	MedSagittal	512 × 512	29
HEVC	*ClassC*	832 × 480	300
	ClassD	416 × 240	300

- **Coding parameters:**
  - LP and HP frames are encoded by JPEG2000 [5]
  - Block-based MC with block size equals 8
  - Search range equals 8 and is doubled for every decomposition level until a maximum size of 64
  - Motion vectors are encoded using the QccPack library [6]

Differences of our proposed CA-WL compared to the uniform WL (U-WL) with and without block-based MC.

		λ	Surv	Med	HEVC	Total average
No MC	Δ PSNR$_{\mathrm{LP}_t}$ [dB]	1	4.12	5.28	15.45	**8.88**
		3	1.64	1.91	8.86	**5.30**
		5	0.97	1.16	6.31	**3.67**
		7	0.65	1.16	6.18	**3.50**
	Δ File size [%]	1	5.99	0.09	10.36	**6.56**
		3	0.80	−0.96	4.15	**2.18**
		5	0.23	−1.29	2.44	**1.08**
		7	0.16	−1.29	1.66	**0.67**
Block-based MC	Δ PSNR$_{\mathrm{LP}_t}$ [dB]	1	9.30	15.56	10.57	**10.98**
		3	8.17	13.89	10.43	**10.28**
		5	7.42	13.89	9.38	**9.47**
		7	7.27	13.89	8.68	**9.02**
	Δ File size [%]	1	0.16	−5.58	4.44	**1.34**
		3	−0.52	−5.64	−0.18	**−1.06**
		5	−0.69	−5.64	−0.66	**−1.38**
		7	−0.80	−5.64	−0.94	**−1.57**





4. Conclusion

- Temporal resolution controlled by recursive application of WT
- Visual quality of BL is degraded by strong motion of underlying video
- CA-WL locally adapts temporal scaling by evaluating a Lagrangian cost functional
- For λ=3 and MC, PSNR$_{\mathrm{LP}_t}$ of BL is increased by 10.28 dB and rate is reduced by 1.06%

References:
[1] G. Karlsson and M. Vetterli, Three dimensional sub-band coding of video, ICASSP, New York City, NY, USA, vol. 2, Apr 1988.
[2] J. R. Ohm, Three-dimensional subband coding with motion compensation, TIP, vol. 3, no. 5, Sep 1994.
[3] D. Lanz, J. Seiler, K. Jaskolka, and A. Kaup, Compression of dynamic medical CT data using motion compensated wavelet lifting with denoised update, PCS, San Francisco, CA, USA, June 2018.
[4] I. H. Witten, R. M. Neal, and J. G. Cleary, Arithmetic coding for data compression, Communications of the ACM, vol. 30, no. 6, June 1987.
[5] ITU-T and ISO/IEC, JPEG 2000 Image Coding System: Core Coding System, in ITU-T Rec. T.800 and ISO/IEC 15444−1:2004, Sep 2004.
[6] J.E. Fowler, Qccpack: An open-source software library for quantization, compression, and coding, App. of Digital Image Proc., San Diego, CA, USA, vol. 4115, Aug 2000.