Table 1: For the evaluation, sequences from surveillance systems, medical applications, and the HEVC test data set are employed. All sequences are used in 4:0:0 color sub-sampling format.

		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

functional for a signal s

$$C(s) = D(s) + \lambda R(s). \qquad (4)$$

To determine the distortion $D(s)$ of the resulting LP frame, we calculate the MSE of the corresponding wavelet coefficients compared to the original signal according to [13]. In this work, not only the similarity of the LP frame to the odd-indexed frame, but also the similarity to the even indexed frame is considered, which is a very important aspect for many professional applications. The rate $R(s)$ 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. Then, the current decomposition can be evaluated locally by comparing the R-D costs of the children to the costs of the parent node for a given value λ. If the R-D costs of the child nodes exceed the costs of the parent node, thus if the following inequality

$$D(l_{i,[2t-1,2t]}) + \lambda R(l_{i,[2t-1,2t]}) \leq \qquad (5)$$
$$(D(l_{i+1,2t-1}) + D(h_{i+1,2t})) + \lambda (R(l_{i+1,2t-1}) + R(h_{i+1,2t}))$$

holds, then a further decomposition is prevented. The Lagrange multiplier λ can be any positive value. By choosing large values for λ, the rate is decreased, while for small values the distortion is decreased.

3.2. Handling of the Overhead

For lossless reconstruction, the decomposition depth for every part of the input sequence has to be transmitted additionally. Therefore, a vector $\boldsymbol{v}$ is generated, whose length corresponds to T. This vector is initialized with zeros and gets an increment of 1 at every temporal position of a LP frame after one decomposition level. The position to the corresponding HP frame is set to zero. Consequently, the non-zero entries correspond to the number of applied decomposition levels i for every temporal position of a LP frame, while the distance d to the corresponding HP frame is given by $d=2^{i-1}$. For the schematic video sequence in Fig. 2, vector $\boldsymbol{v}$ is generated as follows:

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 $i{=}3$:	$(3,0,0,0,0,0,0,0,1,0,1,0,2,0,0,0)$

The entire vector $\boldsymbol{v}$ is encoded using Context Adaptive Binary Arithmetic Coding (CABAC) [14] and is transmitted to the decoder side. Then, for lossless reconstruction of the previous stage, the decoder can easily determine the decomposition level and the temporal positions of the LP and HP frames regarding the original video sequence.

Table 2: Absolute $\text{PSNR}_{\text{LP}_t}$ [dB] and relative rate [%] differences of our proposed CA-WL compared to the U-WL with (bottom) and without (top) block-based MC. Positive numbers denote a better visual quality and a higher rate of our proposed CA-WL and vice versa.

		λ	Surv	Med	ClassC	ClassD	Total average
No MC	$\Delta\,\text{PSNR}_{\text{LP}_t}$	1	4.12	5.28	12.83	18.07	**8.88**
		3	1.64	1.91	6.32	11.4	**5.3**
		5	0.97	1.16	3.73	8.89	**3.67**
		7	0.65	1.16	4.09	8.27	**3.5**
	Δ File size	1	5.99	0.09	11.64	9.07	**6.56**
		3	0.8	-0.96	2.53	5.77	**2.18**
		5	0.23	-1.29	0.84	4.04	**1.08**
		7	0.16	-1.29	0.25	3.07	**0.67**
Block-based MC	$\Delta\,\text{PSNR}_{\text{LP}_t}$	1	9.3	15.56	6.99	14.15	**10.98**
		3	8.17	13.89	9.6	11.26	**10.28**
		5	7.42	13.89	9.21	9.55	**9.47**
		7	7.27	13.89	8.95	8.42	**9.02**
	Δ File size	1	0.16	-5.58	1.98	6.9	**1.34**
		3	-0.52	-5.64	-1.7	1.35	**−1.06**
		5	-0.69	-5.64	-1.96	0.65	**−1.38**
		7	-0.8	-5.64	-2.18	0.29	**−1.57**

4. EXPERIMENTAL RESULTS

Our simulation setup comprises surveillance videos, medical sequences with contrast medium, and natural sequences from the HEVC test data set [15]. The dimensions are summarized in Table 1. The bit depth for all sequences constitutes 8 bits per sample. All surveillance sequences are characterized by a static background and some moving objects in the foreground. The medical sequences origin from Digital Subtraction Angiography (DSA), showing the inflow of a contrast medium into a human cranium in frontal and sagittal perspectives.

In the following, we will compare our proposed CA-WL to a uniform wavelet lifting (U-WL) with the same number of total decomposition levels. The single frames of each subband are encoded by JPEG 2000, using the OpenJPEG [16] implementation with four spatial wavelet decomposition steps in xy-direction. Further, we evaluate the CA-WL and the U-WL with and without a block-based MC, respectively. For block-based MC, the block size is set to 8, while the search range starts with a size of 8 and is doubled for every decomposition level until a maximum size of 64. The increasing search range is important, since the input frames of higher decomposition levels have a larger temporal distance, which has to be covered. To keep the computational effort realistic, we limit the increment of the search range by 64 pixels. The resulting motion vectors are encoded using the QccPack library [17]. Then, the entire file size is composed of the rate resulting from each subband, the required motion vectors and the coding costs for transmitting vector $\boldsymbol{v}$. The visual quality of the resulting LP subband is measured by the same metric as already used in Section 3.1, but in terms of $\text{PSNR}_{\text{LP}_t}$ [13].

Table 2 gives the differences regarding $\text{PSNR}_{\text{LP}_t}$ in [dB] and the entire file size in [%] of our proposed method compared to the U-WL for all data sets with and without the application of MC and for different values of λ. As can be seen in the right column, our method always results in a better visual quality compared to the U-WL. By increasing λ, the file size is reduced, while the $\text{PSNR}_{\text{LP}_t}$ gains are also decreased. However, for $\lambda{=}7$, the $\text{PSNR}_{\text{LP}_t}$ gains are still positive. By including MC into both methods, we are able to get a lower rate than for the U-WL, resulting in positive $\text{PSNR}_{\text{LP}_t}$ gains at the same time. For $\lambda{=}3$, the file size can be reduced by up