INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG4
MPEG VIDEO CODING

ISO/IEC JTC1/SC29/WG4 MPEG/M54895
June 2020, Online

Source **Poznań University of Technology (PUT), Poznań, Poland**
Status **Input**
Title **Noise modelling in TMIV**
Authors Adrian Dziembowski, Dawid Mieloch, Olgierd Stankiewicz, Jakub Nikonowicz, Marek Domański

1 Introduction

This document presents a technical description of the PUT experiment on noise handling in TMIV. In the proposed solution, input views are being denoised at the encoder side, while in the decoder, the resynthesized additive Gaussian noise is added back to recovered pruned input views. Characteristics of the noise for each input view are being sent within metadata.

2 Proposed technique

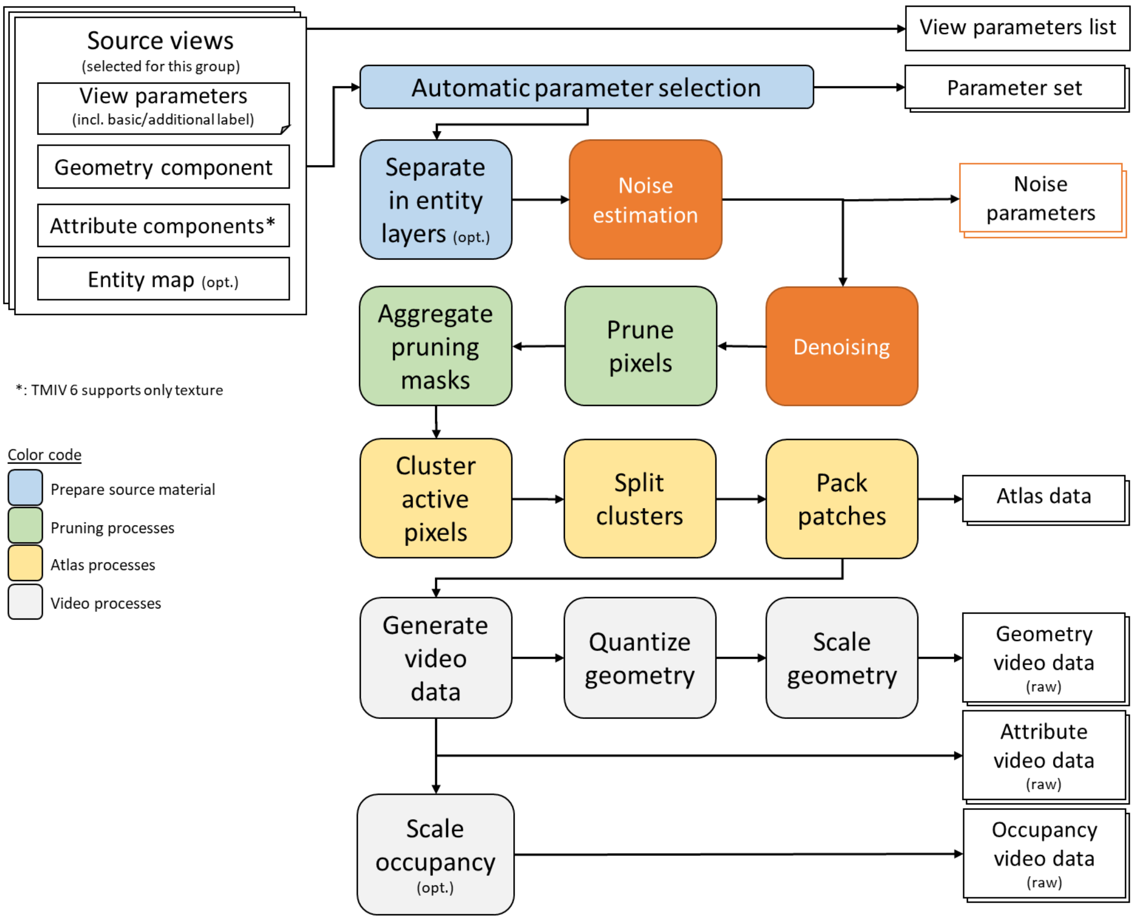

Fig. 1. TMIV encoder with proposed noise modelling.

The proposed approach consists of three main steps: noise estimation and denoising at the encoder side (Fig. 1) and renoising (noise regeneration) in the decoder (Fig. 2). Parameters of the noise are sent within metadata. It is assumed, that noise in the video is additive Gaussian noise.

Noise is analyzed and removed independently for each input view in the encoder and added independently for each pruned reconstructed view in the decoder.

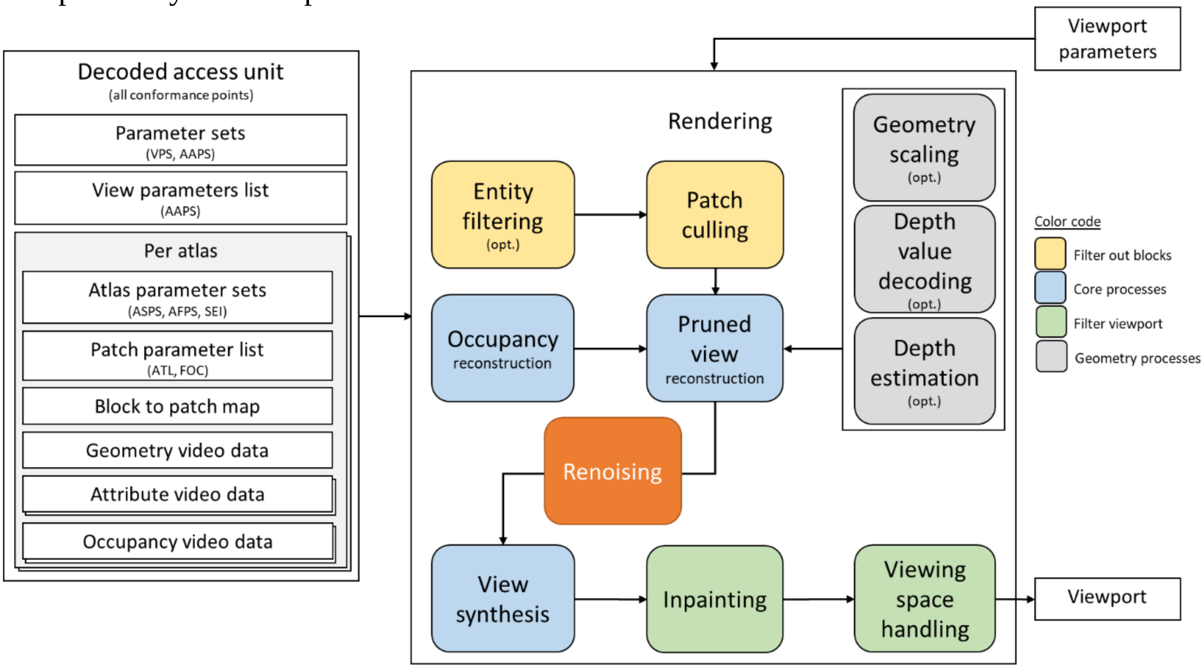

Fig. 2. TMIV decoder with proposed noise modelling.

In order to estimate noise parameters (i.e. standard deviation of the noise of each color component), the input view is analyzed in the temporal domain. For each pixel, which has the same value for frame f and $f-2$, the difference between color component value in frames f and $f-1$ is calculated: $D(f,x,y) = Y(f,x,y) - Y(f-1,x,y)$. The same operation is performed for all possible frames in the GOP (i.e. for 30 of 32 frames). Then, the standard deviation is calculated for all calculated differences $D(f,x,y)$.

Denoising is performed spatially. For every pixel, a 3×3 neighborhood is analyzed. All pixels which have color component value similar to the center's one (difference smaller than two standard deviations) are being averaged. This average substitutes the original value of central pixel within a block (Fig. 3. presents an example with a standard deviation equal to 1: all values from the range [7,9] within a 3×3 block are averaged).

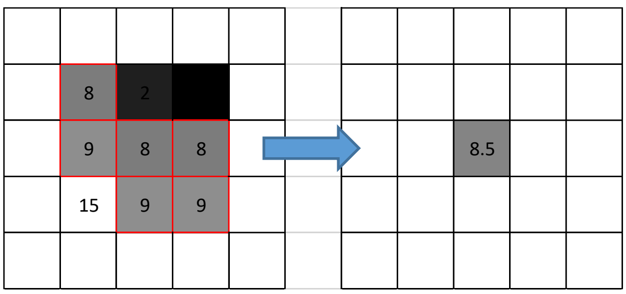

Fig. 3. Example of denoising, block 3×3, central value: 8, stdev: 1.

The standard deviations for all 3 color components for each view are sent within metadata.

At the decoder side, the noise is added to the recovered pruned views. Values of the noise are generated pseudo-randomly with assumed Gaussian distribution, mean value equal to 0, and proper standard deviation.

3 Experimental results

Denoising of input views allows significantly reducing the bitrate of attribute atlases for lower compression (QP1, QP2) for natural sequences and SA. For other CG sequences, this difference is negligible.

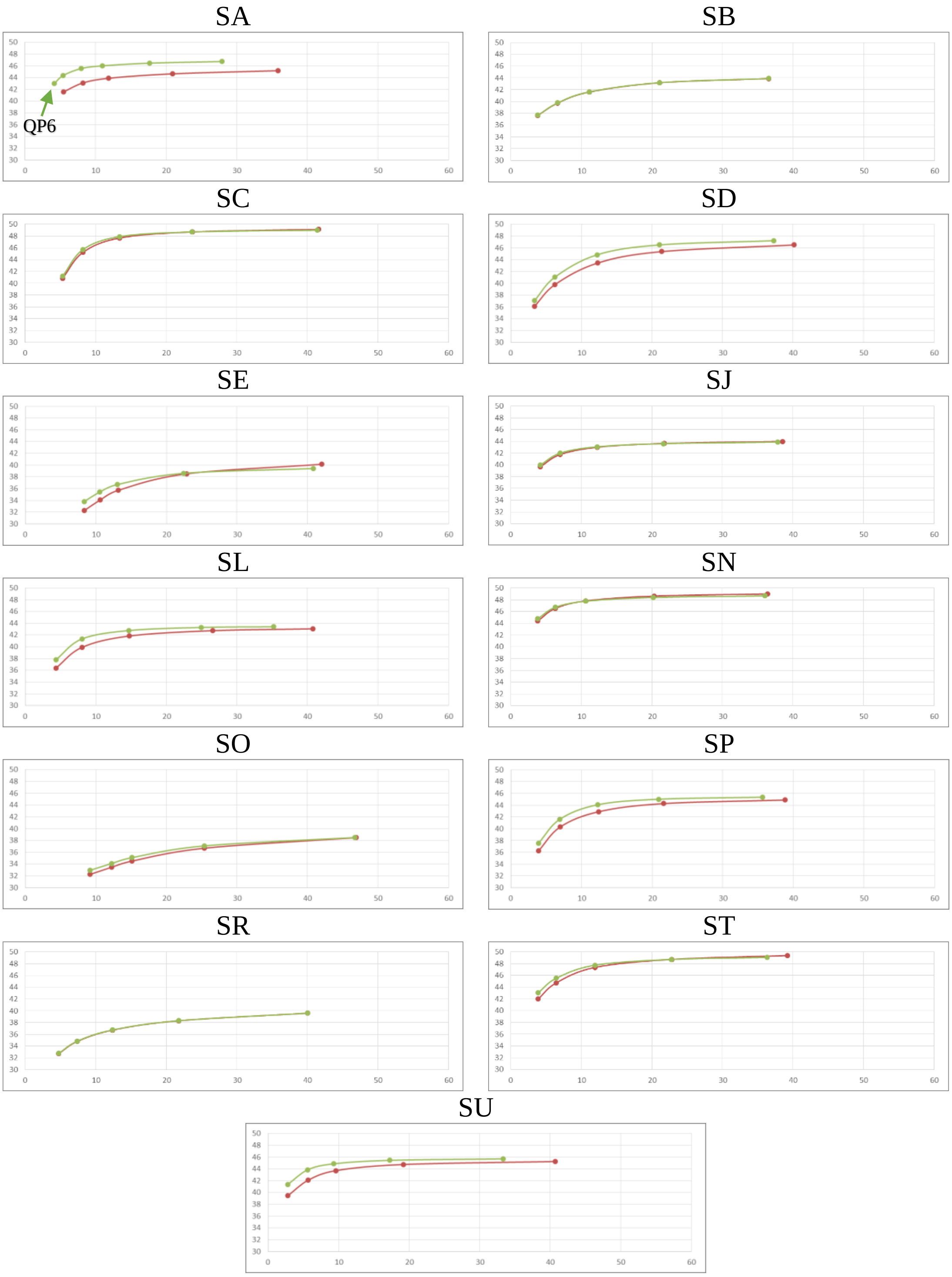

Fig. 4. IV-PSNR RD-curves: anchor (red) vs. proposed (green).

The proposed approach also significantly increases the **subjective quality** and IV-PSNR values. Other objective quality metrics show the opposite: VMAF values are slightly lower and WS-PSNR is significantly lower than for the anchor.

Table 1. Objective evaluation of the proposed technique.

Mandatory content - Proposal vs. Low/High-bitrate Anchors

Sequence		**High-BR BD rate Y-PSNR**	**Low-BR BD rate Y-PSNR**	**Max delta Y-PSNR**	**High-BR BD rate VMAF**	**Low-BR BD rate VMAF**	**High-BR BD rate IV-PSNR**	**Low-BR BD rate IV-PSNR**
ClassroomVideo	SA	---	---	1.05	33.7%	10.8%	-75.3%	-59.2%
Museum	SB	0.5%	0.3%	16.64	0.2%	0.1%	-1.0%	-1.0%
Hijack	SC	8.0%	2.9%	7.55	0.9%	0.2%	-5.9%	-6.1%
Chess	SN	11.1%	4.9%	13.95	4.9%	1.2%	7.7%	-3.0%
Kitchen	SJ	9.9%	4.4%	14.47	6.5%	2.0%	-3.4%	-6.8%
Painter	SD	20.0%	10.7%	4.99	2.7%	0.5%	-29.9%	-21.8%
Frog	SE	38.8%	22.5%	2.98	1.6%	-0.2%	-9.4%	-15.1%
Carpark	SP	41.7%	17.5%	4.57	5.5%	1.8%	-34.4%	-23.7%
MIV		**---**	**---**	**8.27**	**7.0%**	**2.0%**	**-19.0%**	**-17.1%**

← QP2 – QP6 (ClassroomVideo row)

Optional content - Proposal vs. Low/High-bitrate Anchors

Fencing	SL	80.3%	30.2%	8.30	2.6%	0.7%	-42.2%	-31.2%
Hall	ST	39.0%	18.7%	8.16	3.8%	1.4%	-11.4%	-14.0%
Street	SU	199.4%	52.5%	6.67	8.8%	2.7%	-54.9%	-45.3%
Group	SR	0.5%	0.3%	11.78	0.1%	-0.0%	-0.8%	-1.0%
Fan	SO	10.7%	7.6%	7.28	0.3%	-0.3%	-10.4%	-12.3%
MIV		**66.0%**	**21.9%**	**8.44**	**3.1%**	**0.9%**	**-23.9%**	**-20.8%**

For noisy SA sequence, the difference between WS-PSNR and IV-PSNR is so high, that RD-curves do not overlap but the proposed (green) curve is at once above (IV-PSNR) and below (WS-PSNR) the red anchor curve (Fig. 5).

Sequence		**High-BR BD rate Y-PSNR**	**Low-BR BD rate Y-PSNR**	**Max delta Y-PSNR**	**High-BR BD rate VMAF**	**Low-BR BD rate VMAF**	**High-BR BD rate IV-PSNR**	**Low-BR BD rate IV-PSNR**
ClassroomVideo	SA	---	---	1.05	42.1%	12.8%	---	-70.3%

too big loss too big gain

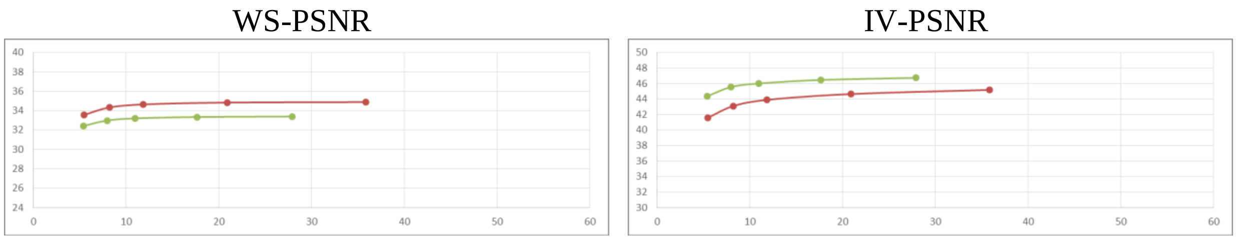

Fig. 5. WS-PSNR vs. IV-PSNR for SA (QP1 – QP5); red curve – anchor, green curve – proposed.

Changing the QP range to QP2 – QP6 allowed to estimate BD-rates for IV-PSNR (Table 1), while additionally reducing curves overlap for WS-PSNR. Calculation of all the BD-rate values was impossible under current CTC.

In Figs. 6 and 7, the subjective quality of synthesized views is compared. In general, denoising at the encoder side allows preserving more details (e.g. name and nationality of the fencer) and reduce coding artifacts (e.g. ringing at the saber or close to the windows) – Fig. 6.

Reference Anchor (QP1) Proposed (QP1)

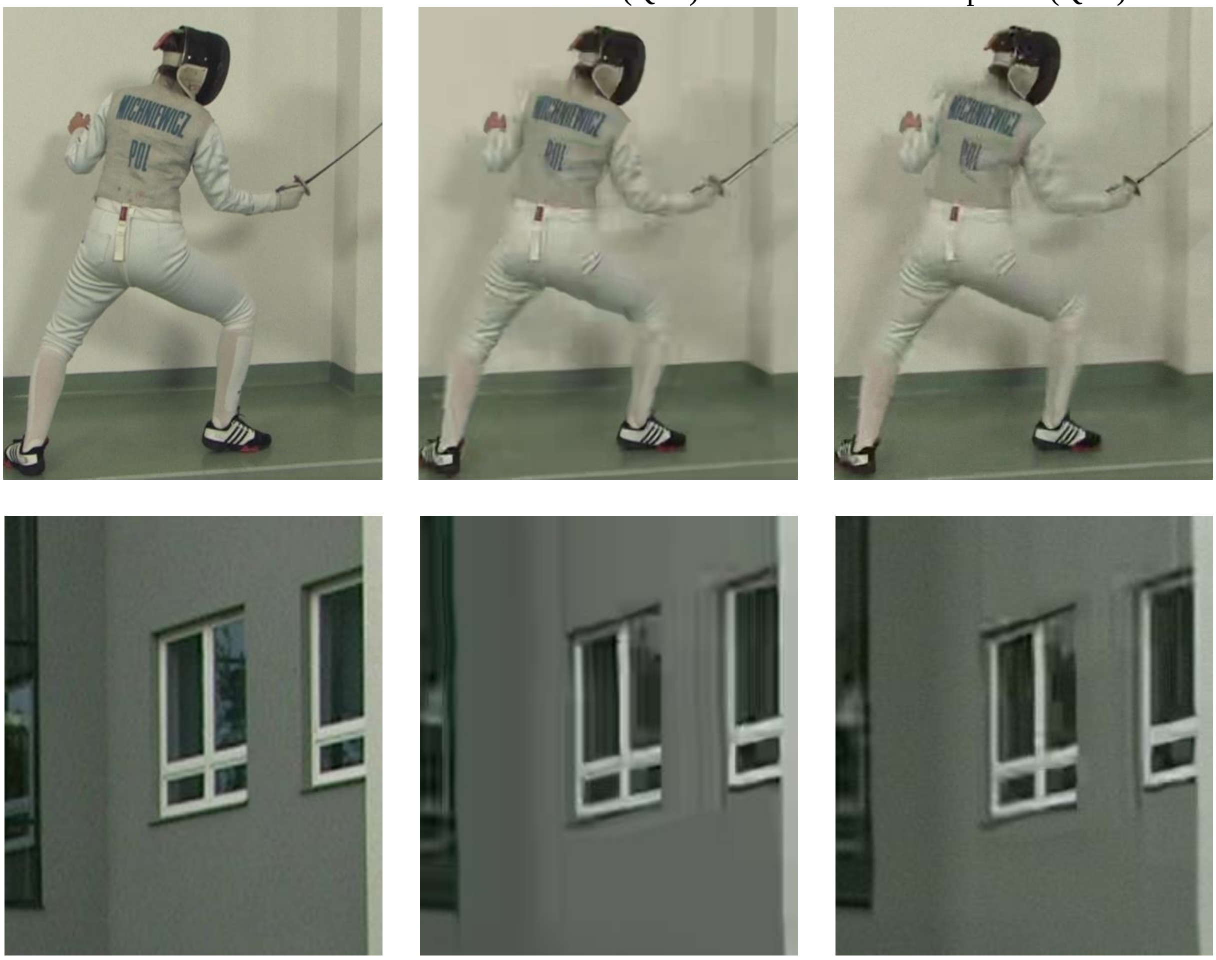

Fig. 6. Fragments of synthesized input views (SL, v3, QP4 and SP, v5, QP5).

The amplitude of the regenerated noise is similar to the original noise (Fig. 7). Of course, noise in the proposed approach is generated based on noise characteristics – therefore, it is not correlated with original noise. Such a lack of correlation significantly decreases pixel-wise metric (WS-PSNR).

In terms of IV-PSNR, where each pixel is being compared with a colocated 5×5 block of pixels, the image with added noise (with the same characteristics) is more similar to the original one than the image without noise.

The proposed approach also increases the temporal consistency of noise characteristics, especially for lower compression (Fig. 8). In anchor, there is noticeable noise in the I-frames, but there is no noise in B-frames. In the proposed approach, characteristics of the noise are the same for the entire GOP.

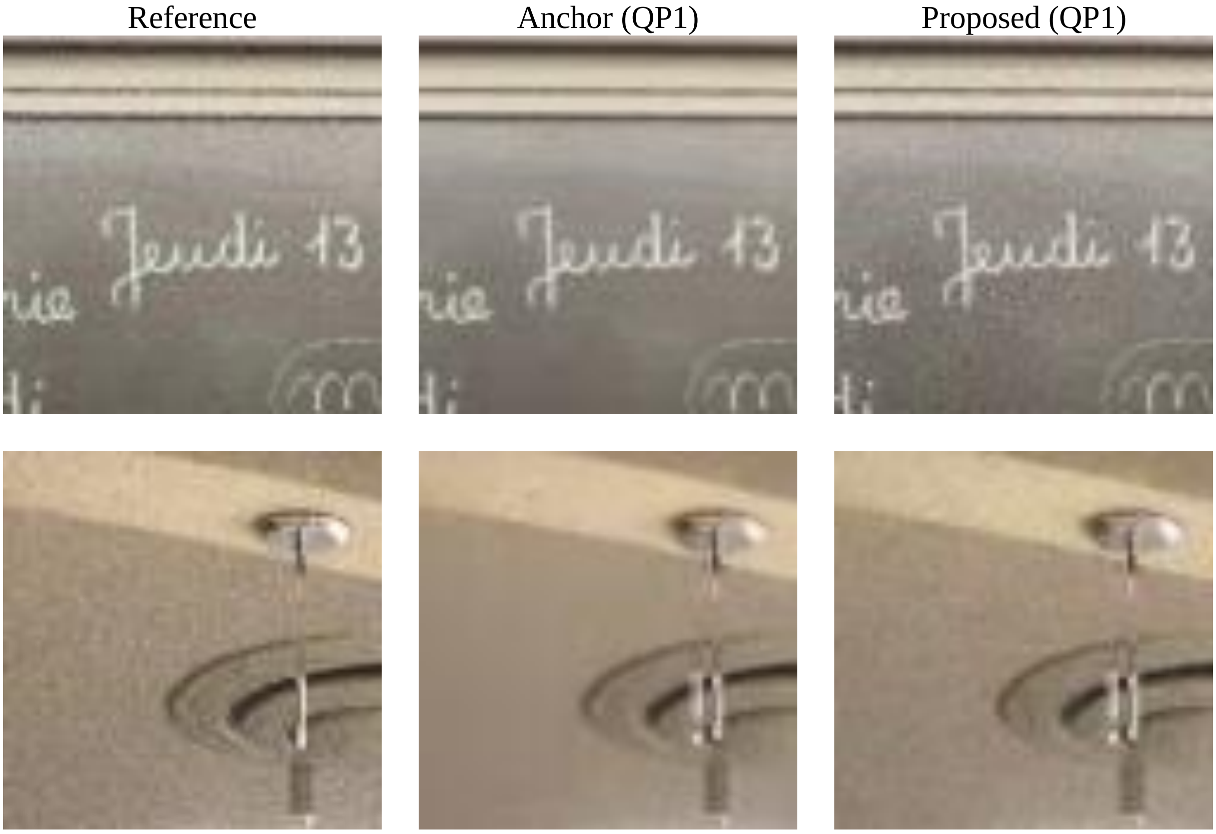

Fig. 7. Fragments of synthesized input view (SA, v3, QP1).

Fig. 8. Fragment of synthesized input view (SA, v0, QP1).

Average standard deviations of the noise for each sequence are presented in Table 2. Standard deviation values for each view are presented in Tables 3 and 4, separately for CG sequences (Table 3) and NC sequences including SA (Table 4).

Table 2. Calculated noise standard deviation values per sequence.

Sequence	σ_Y	σ_U	σ_V
SA	14.49	2.30	1.92
SD	10.17	3.89	3.89
SE	26.77	5.59	5.62
SL	12.83	3.38	3.83
SP	10.29	4.23	4.57
ST	3.86	1.89	1.95
SU	9.85	3.84	4.21
SB	1.14	0.56	0.57
SC	3.25	0.83	0.84
SJ	3.98	0.72	0.54
SN	3.26	0.63	0.51
SO	14.53	1.69	2.51
SR	1.77	0.76	0.79

Table 3. Calculated noise standard deviation values per view – CG sequences.

	SB			SC			SJ			SN			SO			SR		
	σ_Y	σ_U	σ_V	σ_Y	σ_U	σ_V	σ_Y	σ_U	σ_V	σ_Y	σ_U	σ_V	σ_Y	σ_U	σ_V	σ_Y	σ_U	σ_V
v0	1.60	0.73	0.73	3.54	0.90	0.91	3.90	0.74	0.51	3.89	0.61	0.35	14.38	1.65	2.51	1.73	0.74	0.78
v1	1.08	0.57	0.59	3.12	0.83	0.84	3.81	0.70	0.51	1.89	0.47	0.39	14.49	1.68	2.52	1.58	0.73	0.77
v2	1.96	0.67	0.69	2.93	0.77	0.78	3.75	0.67	0.52	1.81	0.49	0.39	14.64	1.70	2.48	1.79	0.75	0.79
v3	1.50	0.69	0.70	3.14	0.79	0.79	3.84	0.67	0.52	1.84	0.52	0.46	14.59	1.72	2.47	1.83	0.75	0.79
v4	1.01	0.49	0.51	3.49	0.85	0.87	3.83	0.66	0.53	4.26	0.70	0.54	14.11	1.72	2.44	1.69	0.74	0.79
v5	1.92	0.76	0.79	3.12	0.81	0.81	4.07	0.76	0.52	5.16	0.78	0.63	14.36	1.65	2.56	1.69	0.73	0.78
v6	1.80	0.76	0.78	3.40	0.88	0.88	4.01	0.73	0.52	4.95	0.79	0.70	14.46	1.71	2.60	1.76	0.74	0.78
v7	1.21	0.65	0.66	3.11	0.79	0.80	3.88	0.70	0.52	3.84	0.74	0.59	14.65	1.71	2.56	1.80	0.74	0.78
v8	1.77	0.73	0.74	3.39	0.86	0.87	3.97	0.69	0.53	1.98	0.54	0.46	14.73	1.73	2.52	1.60	0.72	0.74
v9	2.41	0.66	0.69	3.26	0.83	0.84	3.87	0.68	0.53	2.93	0.66	0.59	14.46	1.73	2.46	1.82	0.75	0.78
v10	1.38	0.59	0.61				4.19	0.78	0.53				14.38	1.63	2.58	1.93	0.77	0.80
v11	2.02	0.71	0.73				4.11	0.75	0.53				14.59	1.66	2.59	1.95	0.76	0.81
v12	0.35	0.32	0.32				4.15	0.71	0.53				14.77	1.65	2.53	1.80	0.75	0.79
v13	0.53	0.32	0.34				3.97	0.71	0.53				14.81	1.67	2.46	1.79	0.76	0.79
v14	0.73	0.46	0.47				3.90	0.69	0.53				14.49	1.66	2.36	1.85	0.77	0.80
v15	0.94	0.62	0.63				4.21	0.79	0.55							1.89	0.78	0.82
v16	0.59	0.51	0.51				4.07	0.76	0.55							1.80	0.78	0.81
v17	0.79	0.53	0.53				3.95	0.72	0.55							1.92	0.79	0.84
v18	0.71	0.42	0.43				3.94	0.71	0.54							1.61	0.77	0.80
v19	0.58	0.49	0.49				3.90	0.70	0.54							1.65	0.77	0.80
v20	0.78	0.62	0.61				4.22	0.79	0.57							1.68	0.79	0.83
v21	0.40	0.37	0.37				4.11	0.76	0.57									
v22	0.27	0.26	0.26				3.99	0.73	0.57									
v23	1.07	0.62	0.62				3.99	0.71	0.56									
v24							3.99	0.71	0.56									

Table 3. Calculated noise standard deviation values per view – natural sequences + SA.

	SA			**SD**			**SE**			**SL**			**SP**			**ST**			**SU**		
	σ_Y	**σ_U**	**σ_V**	**σ_Y**	**σ_U**	**σ_V**	**σ_Y**	**σ_U**	**σ_V**	**σ_Y**	**σ_U**	**σ_V**	**σ_Y**	**σ_U**	**σ_V**	**σ_Y**	**σ_U**	**σ_V**	**σ_Y**	**σ_U**	**σ_V**
v0	17.24	2.50	2.04	9.82	3.75	3.80				12.40	3.45	3.74	10.52	4.05	4.27	3.77	1.86	1.92	9.72	3.75	4.25
v1	13.93	2.26	1.90	9.25	3.49	3.55	24.70	6.04	5.88	12.08	3.47	4.20	10.53	4.41	4.75	3.74	1.85	1.98	9.99	4.03	4.27
v2	14.35	2.29	1.91	9.51	3.58	3.66	26.07	5.89	5.80	12.83	3.43	4.21	10.46	4.16	4.61	3.81	1.87	1.93	10.03	3.97	4.39
v3	14.24	2.28	1.90	9.37	3.55	3.58	25.55	5.91	5.87	12.73	3.46	3.62	10.42	4.46	4.82	3.79	1.93	2.05	9.86	3.82	4.38
v4	14.43	2.29	1.91	10.14	3.83	3.86	25.63	5.81	5.85	13.04	3.36	3.73	10.05	4.17	4.67	3.75	1.92	1.96	9.78	4.11	4.46
v5	14.30	2.28	1.90	10.11	3.78	3.86	24.39	5.61	5.68	12.69	3.06	3.59	10.04	4.28	4.64	3.83	1.88	2.00	9.62	3.56	4.00
v6	14.39	2.29	1.91	9.77	3.69	3.72	24.32	5.42	5.47	13.56	3.16	3.91	10.36	3.90	4.14	3.89	1.86	1.87	9.90	3.87	3.99
v7	14.35	2.29	1.91	10.59	4.35	4.16	24.57	5.47	5.48	13.12	3.21	3.50	10.28	4.36	4.65	4.04	1.89	1.89	9.96	3.67	4.17
v8	14.31	2.28	1.90	10.30	4.08	4.01	24.40	5.33	5.56	12.76	3.79	4.02	9.97	4.26	4.56	4.15	1.93	1.97	9.81	3.79	3.96
v9	14.29	2.29	1.91	10.71	4.05	4.07	28.75	5.39	5.46	13.09	3.41	3.78									
v10	14.09	2.27	1.90	10.09	3.81	3.76	29.11	5.39	5.42												
v11	14.45	2.30	1.92	10.16	3.80	3.86	32.73	5.38	5.54												
v12	14.21	2.28	1.90	10.20	3.79	3.84	30.64	5.44	5.50												
v13	14.49	2.30	1.91	10.74	4.09	4.10	27.21	5.63	5.50												
v14	14.33	2.28	1.90	10.49	4.04	3.99															
v15				11.49	4.50	4.39															

4 Syntax and semantics

4.1 Modify §7.3.6.13.1:

common_atlas_frame_rbsp( ) {	**Descriptor**
caf_atlas_adaptation_parameter_set_id	ue(v)
caf_frm_order_cnt_lsb	u(v)
caf_miv_view_params_list_update_mode	u(3)
if(caf_miv_view_params_list_update_mode == VPL_INITLIST)	
miv_view_params_list( )	
else if(caf_miv_view_params_list_update_mode == VPL_UPD_EXT)	
miv_view_params_update_extrinsics( )	
else if(caf_miv_view_params_list_update_mode == VPL_UPD_INT)	
miv_view_params_update_intrinsics( )	
else if(caf_miv_view_params_list_update_mode == VPL_UPD_DQ)	
miv_view_params_update_depth_quantization( )	
else if(caf_miv_view_params_list_update_mode == VPL_UPD_NCH)	
miv_view_params_update_noise_characteristics( )	
else if(caf_miv_view_params_list_update_mode == VPL_ALL) {	
miv_view_params_update_extrinsics( )	
miv_view_params_update_intrinsics( )	
miv_view_params_update_depth_quantization( )	
miv_view_params_update_noise_characteristics( )	
}	

caf_extension_flag	u(1)
if(caf_extension_flag)	
caf_extension_8bits	u(6)
if(caf_extension_8bits)	
while(more_rbsp_data( ))	
caf_extension_data_flag	u(1)
rbsp_trailing_bits( )	
}	

4.2 *Add §7.3.6.13.10:*

noise_characteristics(v) {	**Descriptor**
nch_stdev_Y[v]	fl(32)
nch_stdev_U[v]	fl(32)
nch_stdev_V[v]	fl(32)
}	

4.3 *Modify §7.4.6.13.1:*

Table 4-1 – Updating modes for view parameters list update

caf_miv_view_params_list_update_mode	**Identifier**	**Description**
0	VPL_INITLIST	a new initialized view parameters list is present
1	VPL_UPD_EXT	extrinsic parameters are updated for sub-set of existing views
2	VPL_UPD_INT	intrinsic parameters are updated for sub-set of existing views
3	VPL_UPD_DQ	depth quantization parameters are updated for a sub-set of existing views
4	VPL_UPD_NCH	noise characteristics parameters are updated for a sub-set of existing views
5	VPL_ALL	extrinsic, intrinsic, ~~and~~ depth quantization and noise characteristics parameters are updated for a sub-set of existing views
6 .. 7	Reserved	Reserved for future use by ISO/EC

4.4 *Add §7.4.6.13.10:*

nch_stdev_Y[v] specifies the standard deviation of luma component's noise of the v-th view as floating point.

nch_stdev_U[v] specifies the standard deviation of first chroma component's noise of the v-th view as floating point.
nch_stdev_V[v] specifies the standard deviation of second component's noise of the v-th view as floating point.

5 Acknowledgement

This work was supported by the Ministry of Science and Higher Education.

6 Recommendations

We recommend:

- to include the proposed technique into TMIV7,
- to adopt proposed syntax and semantics.