INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2009/M16977
October 2009, Xian, China

Source **Ericsson**
Status **Report**
Title **3DTV Exploration Experiments (EE1 & EE4) on the *Lovebird 1* data set**
Author **Apostolos Georgakis (apostolos.georgakis@ericsson.com)**
Pravin Kumar Rana (prara@kth.se)
Ivana Radulovic (ivana.radulovic@ericsson.com)

Introduction

This contribution describes the results to two sets of 3DTV exploration experiments undertaken by Ericsson using the *Lovebird 1* sequence defined in the last MPEG meeting in London (see w10720). These sets cover both EE1 for depth estimation and view synthesis and EE4 for coding efficiency.

Experimental set up

The exploration experiments describe a 2-view (views 6 and 8) and a 3-view (views 4, 6 and 8) configuration. In some cases a narrow baseline is used whereas in some other cases a median baseline is required. Specifically, in the 2-view case for depth estimation views 4, 6 and 8 generate depth 6 and view 6, 8 and 10 generate depth 8 (median baseline). Subsequently the generated depths and the corresponding views are used to synthesize view 7 (narrow baseline). The 3-views case is similar with the exception that multiple intermediate views are requested, synthesized views 4.5, 5, 5.5, 6.5, 7 and 7.5. Figure 1 depicts the 2-view EE1 experiment.

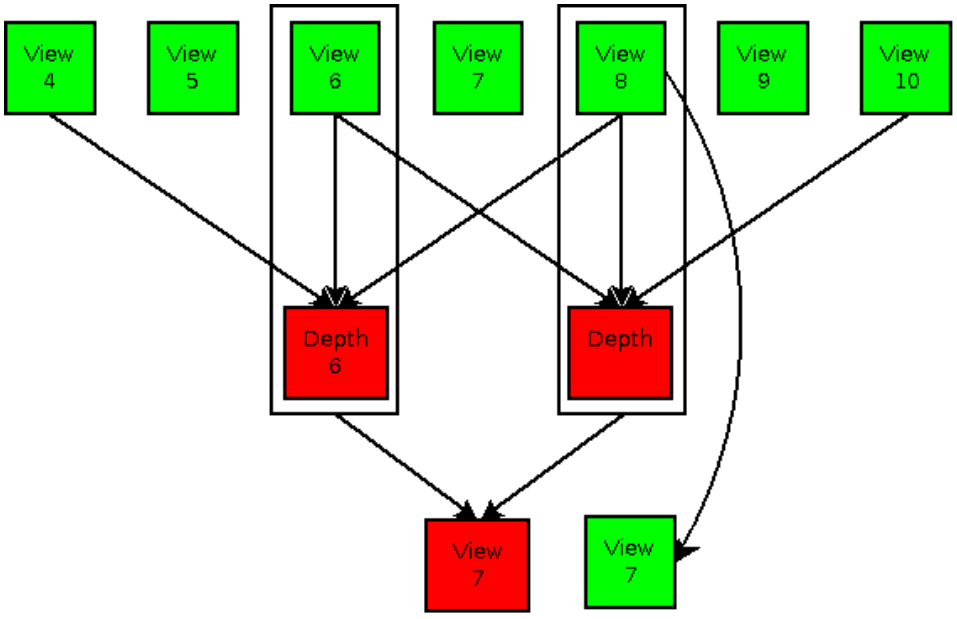

Figure 1: EE1 for the 2-view configuration.

The way the experiments were conducted is the following:

- Experimental Set 1 (ES1): Certain parameters for DE, VS and MVC were hand picked and both EE1 and EE4 were conducted using these parameters.
- Experimental Set 2 (ES2): A wide range of parameters were selected for DE, VS and MVC and an exhaustive search (brute-force approach) has been implemented.

What follows is the results for both ES1 and ES2.

Experimental Set 1 (ES1)

EE1: Depth map estimation

EE1 studied how the quality of estimated depth maps is affected by the "matching methods" in DERS4.9. In Set 1 we tested both with depth estimation mode *on* and *off*. First we will present the results without depth estimation mode and then we will proceed with each one of these modes (depth estimation mode = 1, 2 and 3). Table 1 shows the selected parameters along with the average PSNR values and Fig. 2 shows the PSNR over each frame. It should be noted here that in Tables 1-4 the blue line correspond to the best results from the perspective of PSNR and in Fig. 2-5 the curve denoted with a star (*) corresponds to that particular configuration.

Set	Smoothing Coefficients	Baseline Basis	Precision	Search Level	Filter	Matching Methods	Temporal Enhancement (ON) / Threshold	Depth Estimation Mode	Average Y-PSNR (dB)
1	**5.5**	**0**	**1**	**1**	**0**	**1**	**1.50**	**0**	**30.7307**
2	5.5	0	2	2	0	0	1.50	0	30.6321
3	5.5	0	2	2	0	1	1.50	0	30.6708
4	5.5	1	1	1	1	0	1.50	0	30.1099
5	5.5	1	1	1	1	1	1.50	0	30.1099
6	5.5	1	2	2	1	0	1.50	0	30.1099
7	5.5	1	2	2	1	1	1.50	0	30.1113
8	5.5	2	1	1	2	0	1.50	0	29.9214
9	5.5	2	1	1	2	1	1.50	0	29.9191
10	5.5	2	2	2	2	1	1.50	0	29.9503

Table 1: Depth estimation mode 0 (off)

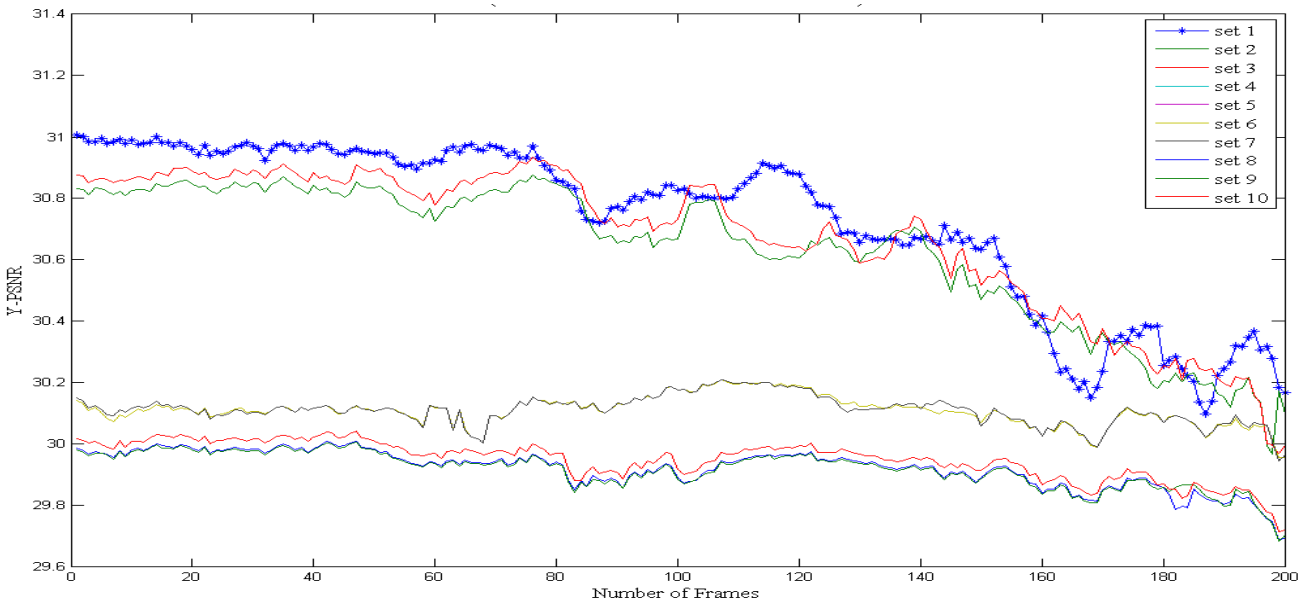

Figure 2: PSNR values of the synth view 7 without depth estimation

Table 2 shows the parameters used for depth map estimation using mode 1 and Fig. 3 shows the PSNR values for the synthesized view 7 again for depth estimation mode 1.

Set	Smoothing Coefficients	Baseline Basis	Precision	Search level	Filter	Matching Methods	Soft Segmentation				Depth Estimation Mode	Average Y-PSNR
							Soft Distance Coeff	Soft Color Coeff	Soft Block Width	Soft Block Height		
1	5.5	0	1	1	0	3	05	10	05	05	1	30.6945
2	**5.5**	**0**	**1**	**1**	**0**	**3**	**05**	**10**	**10**	**10**	**1**	**30.7409**
3	5.5	0	1	1	0	3	10	10	05	05	1	30.7111
4	5.5	0	1	1	0	3	10	10	10	10	1	30.6638
5	5.5	0	1	1	0	1	Ignored for Matching methods 0 and 1				1	30.6757
6	5.5	0	1	1	0	0					1	30.5990
7	5.5	2	1	1	2	1					1	30.0406

Table 2: Depth estimation mode 1

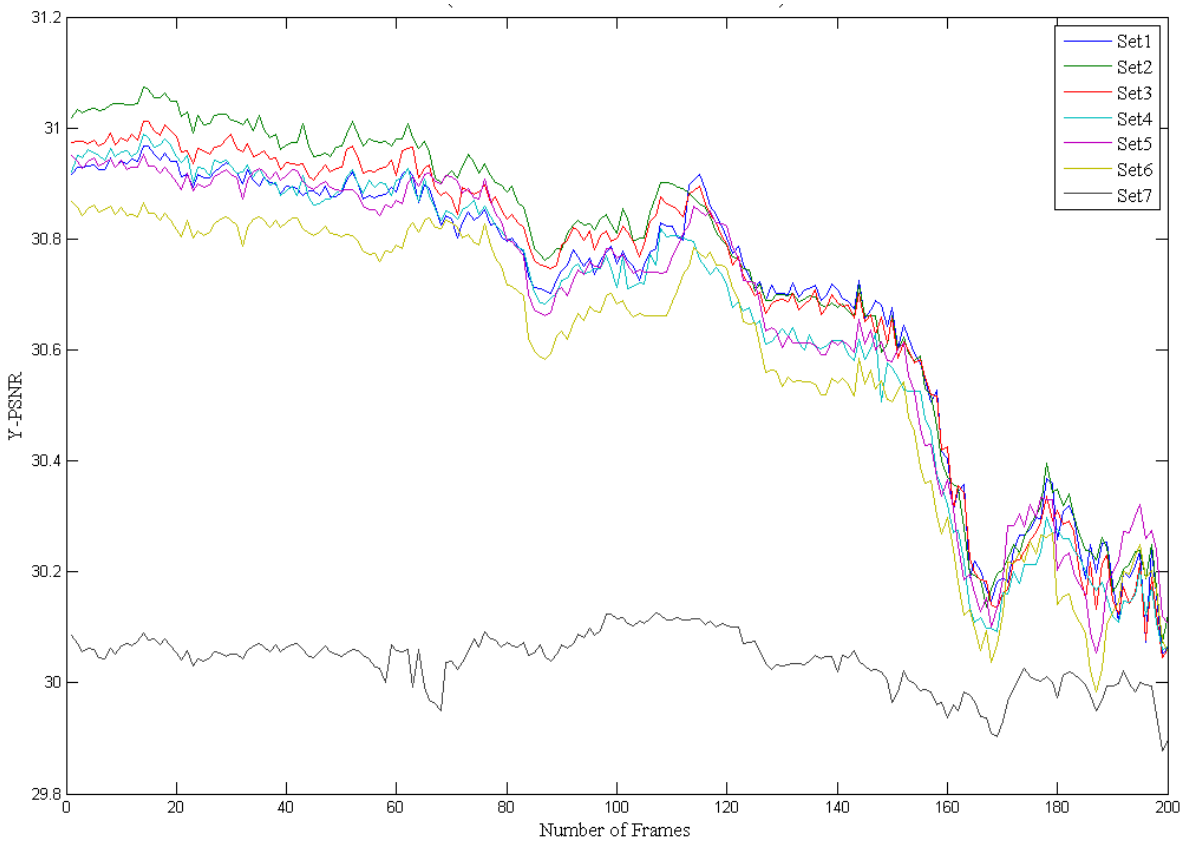

Figure 3: PSNR values of the synth view 7 and depth estimation mode 1

Following we present the results for depth estimation mode 2 (Table 3) along with the PSNR progression over all 200 frames of the sequence in Fig. 4.

Set	Smoothing Coefficients	Matching Methods	**Depth Estimation Mode = 2**			Average Y-PSNR (dB)
			Threshold Of Depth Difference	Moving Objects BSize	Motion Search BSize	
1	5.5	0	10	0	0	30.2520
2	5.5	0	10	0	1	30.2440
3	5.5	0	10	1	0	30.2524
4	5.5	0	10	1	1	30.2443
5	5.5	1	10	0	0	30.2538
6	5.5	1	10	0	1	30.2452
7	**5.5**	**1**	**10**	**1**	**0**	**30.2547**
8	5.5	1	10	1	1	30.2457
9	5.5	3	10	0	0	30.1378

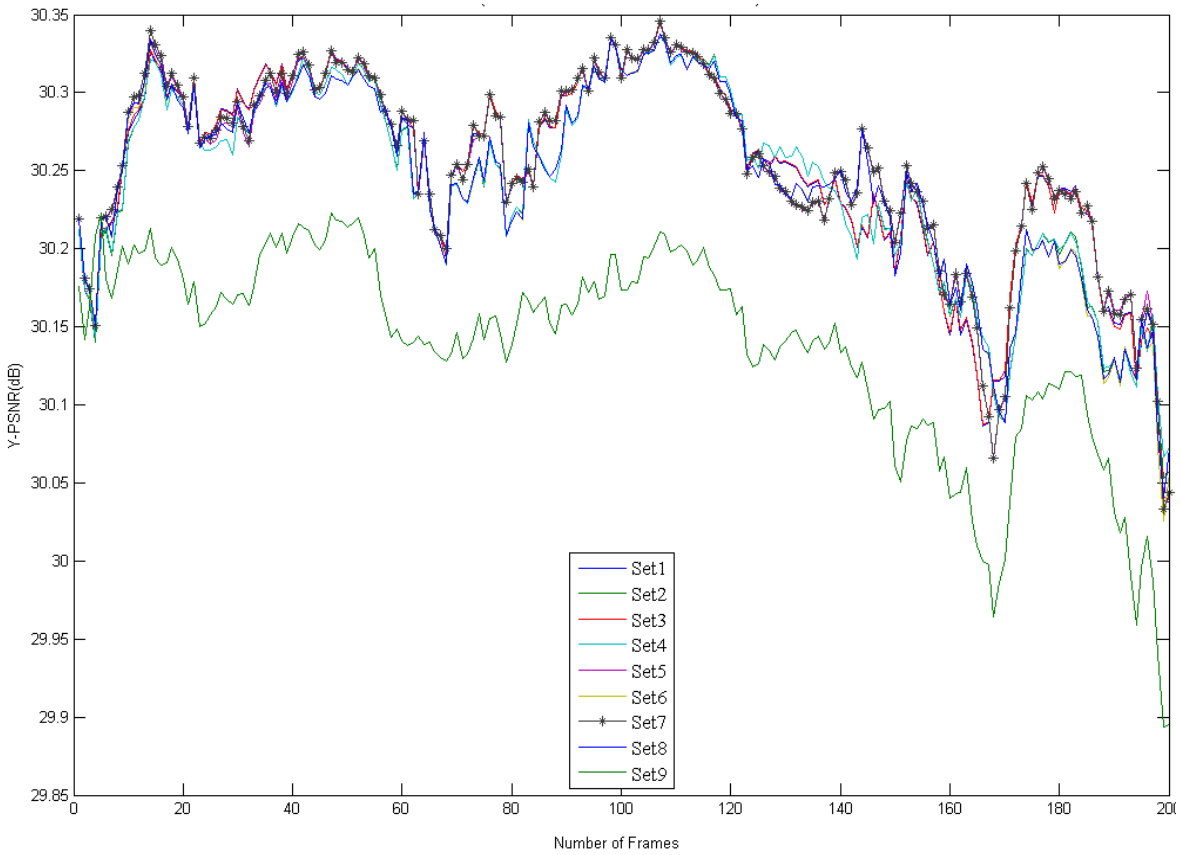

Figure 4: PSNR values of the synth view 7 and depth estimation mode 2

Finally we present results for depth estimation mode 3 using the supplementary data consisting of reference depth maps and camera position. The parameters selected and the average PSNR can be seen in Table 4 whereas the PSNR progression over all the frames are depicted in Fig. 5.

Set	Smoothing Coefficients	Baseline Basis	Precision	Search level	Filter	Matching Methods	Soft Segmentation				Depth Estimation Mode	Average Y-PSNR (dB)
							Soft Distance Coeff	Soft Color Coeff	Soft Block Width	Soft Block Height		
1	5.5	0	1	1	0	0	Ignored for Matching methods 0, 1, and 2				3	30.66601
2	5.5	1	1	1	1	0					3	30.75930
3	**5.5**	**2**	**1**	**1**	**2**	**0**					**3**	**30.76841**
4	5.5	0	1	1	0	1					3	30.67342
5	5.5	1	1	1	1	1					3	30.04774
6	5.5	2	1	1	2	1					3	30.04926
7	5.5	0	1	1	0	2					3	30.09371
8	5.5	1	1	1	1	2					3	30.02198
9	5.5	2	1	1	2	2					3	30.04774
10	5.5	0	1	1	0	3	10	20	11	11	3	30.04926
11	5.5	1	1	1	1	3	10	20	11	11	3	30.09371
12	5.5	2	1	1	2	3	10	20	11	11	3	30.09371

Table 4: Depth estimation mode 3

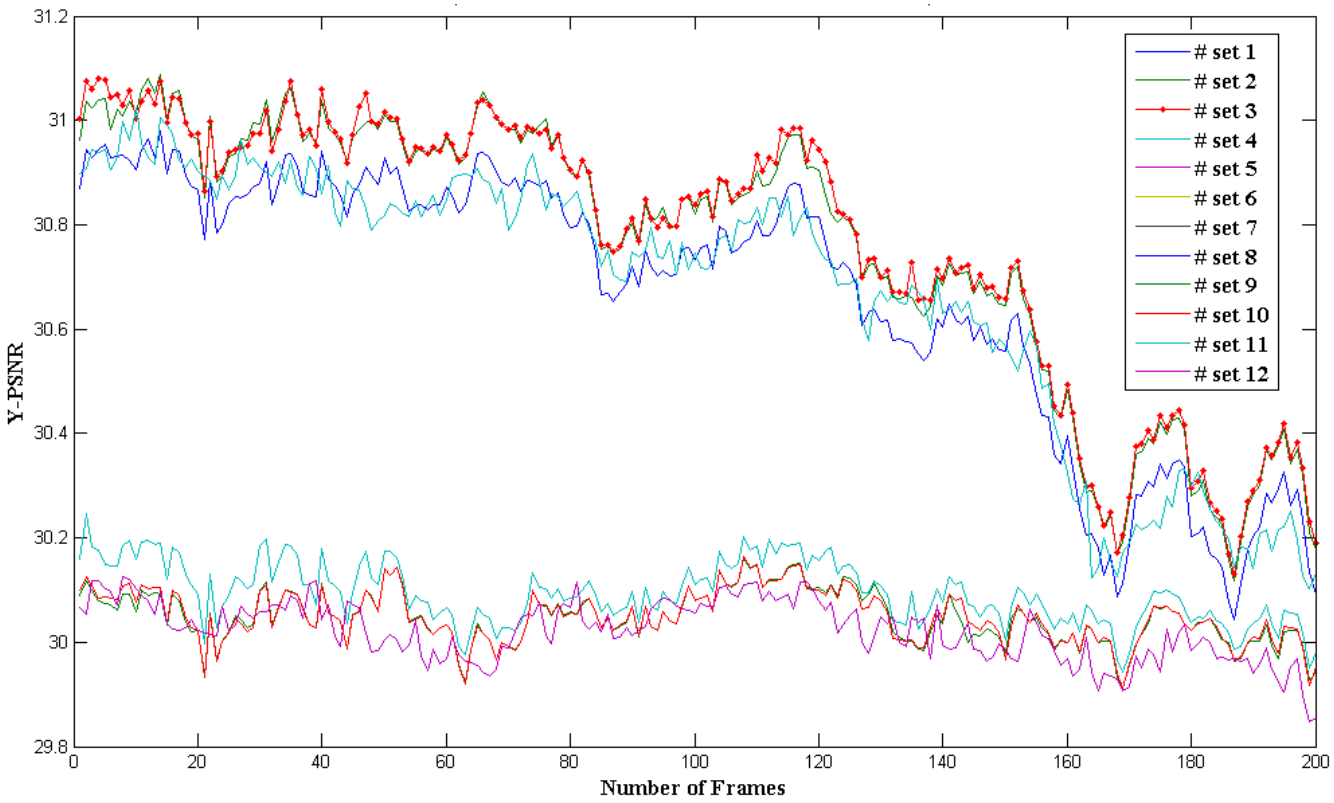

Figure 5: PSNR values of the synth view 7 and depth estimation mode 3

So in order to summarize the results for ES1 and EE1 we have:

Depth Estimation Mode	Set	Average Y-PSNR(dB)
0	1	**30.73070**
1	2	**30.74090**
2	7	**30.25470**
3	3	**30.76841**

EE1: Multi view coding

The goal of EE4 is to get insights on how the depth map coding affects the quality of synthesized views. In ES1 we used QPs {22, 24, 26, 28, 32, 34, 36, 38} for texture. For depth map coding, we use an extended QP range {22, 24, 26, 28, **30**, 32, 34, 36, 38}. We are encoding both texture and depth for the 2-view configuration. For depth maps we used the best overall results from the previous experiments, that is set 3 from depth estimation mode 3 (see Table 4). The following 3 tables (Table 5-7) will present our findings.

QP	View 6		View 8		Depth 6		Depth 8	
	Y-PSNR (dB)	Rate (kbps)	PSNR (dB)	Rate (kbps)	PSNR (dB)	Rate (kbps)	PSNR (dB)	Rate (kbps)
22	42.6576	3627.5616	41.9686	2854.0752	56.2562	153.5472	56.6704	153.5100
24	41.5154	2798.4384	40.9611	2005.9140	54.9718	119.2080	55.2945	118.7928
26	40.3647	2158.5408	39.8740	1405.7892	53.7112	90.4212	53.8529	89.7180
28	39.2506	1687.2156	38.7726	988.1952	52.5718	71.1720	52.0559	69.2580
30	-	-	-	-	51.4473	55.2960	51.3485	54.3648

32	36.9348	1022.8764	36.2147	449.4972	50.2149	44.1108	47.7454	43.6872
34	35.8620	810.3900	35.0653	305.3232	49.0041	37.1928	46.7678	35.3616
36	34.7115	613.2084	33.9686	205.9908	48.7988	31.9812	43.5245	30.4392
38	33.5162	476.1720	32.8927	146.1684	45.4275	26.7552	41.9470	24.5592

Table 5: Y-PSNR values and bitrates for various QPs

Depth QP	Video (QP22)	Video (QP24)	Video (QP26)	Video (QP28)	Video (QP32)	Video (QP34)	Video (QP36)	Video (QP38)
22	6788.6940	5111.4096	3871.3872	2982.4680	1779.4308	1467.7704	1126.25564	929.3976
24	6719.6376	5042.3532	3802.3308	2913.4116	1710.3744	1398.7140	1057.2000	860.3412
26	6661.7760	4984.4916	3744.4692	2855.5500	1652.5128	1340.8524	999.3384	802.4796
28	6622.0668	4944.7824	3704.7600	2815.8408	1612.8036	1301.1432	959.6292	762.7704
30	6591.2976	4914.0132	3673.9908	2785.0716	1582.0344	1270.3740	928.8600	732.0012
32	6569.4348	4892.1504	3652.1280	2763.2088	1560.1716	1248.5112	906.9972	710.1384
34	6554.1912	4876.9068	3636.8844	2747.9652	1544.9280	1233.2676	891.7536	694.8948
36	6544.0572	4866.7728	3626.7504	2737.8312	1534.7940	1223.1336	881.6196	684.7608
38	6532.9512	4855.6668	3615.6444	2726.7252	1523.6880	1212.0276	870.5136	673.6548

Table 6: Total bit rate spent on encoding video and depth for views 6, 8

Depth QP	Video (QP22)	Video (QP24)	Video (QP26)	Video (QP28)	Video (QP32)	Video (QP34)	Video (QP36)	Video (QP38)
22	31.0369	31.0240	30.9998	30.9530	30.7753	30.6313	30.4215	30.1152
24	31.0585	31.0413	31.0175	30.9718	30.7944	30.6478	30.4393	30.1357
26	31.0286	31.0167	30.9887	30.9432	30.7688	30.6351	-	30.1346
28	31.0027	30.9893	30.9627	30.9205	30.7463	30.6146	30.4147	30.1154
30	30.8884	30.8747	30.8534	30.8114	30.6455	30.5186	30.3260	30.0362
32	30.6958	30.6841	30.6649	30.6273	30.4772	30.3672	30.1925	29.9270
34	30.5125	30.5030	30.4865	30.4543	30.3102	30.2125	30.0369	29.7798
36	30.3615	30.3523	30.3345	30.3040	30.1759	30.0695	29.9217	29.6723
38	30.2417	30.2316	30.2185	30.1875	30.0647	29.9640	29.8207	29.5801

Table 7: Y-PSNR of view 7 reconstructed from the decoded texture and depth

Experimental Set 2 (ES2)

EE1: Depth map estimation

For the depth map generation in EE1 and ES2 again DERS was used. A series of parameters were tested using the 2-view setting. In ES2 the parameters were tested exhaustively (brute force). The parameters tested will be shown further down. It must be noted that due to very high computational complexity for such an approach the initial tests were run on a 10 frame sub-sampled version of the sequence and for both the 4-6 and 6-8 view combination and then the best parameters were used for the rest of the experiments. Sampling the sequence was selected as a course of action due to time constraints.

An example of a configuration file for DE for view 4 can be seen below:

```
DepthType                              1
SourceWidth                            1024
SourceHeight                           768
StartFrame                             0
TotalNumberOfFrames                    10
LeftCameraName                         param_loveb1_002
CenterCameraName                       param_loveb1_004
RightCameraName                        param_loveb1_006
MinimumValueOfDisparitySearchRange     1
MaximumValueOfDisparitySearchRange     70
MinimumValueOfDisparityRange           1
MaximumValueOfDisparityRange           70
SmoothingCoefficient                   1 #####
FileLeftViewImage                      lovebird1_cam02.yuv
FileCenterViewImage                    lovebird1_cam04.yuv
FileRightViewImage                     lovebird1_cam06.yuv
FileOutputDepthMapImage      lovebird1_cam04_depth_S1_M0_Mm0_T0_I0_Is0_SA0.yuv
FileCameraParameter                    cam_param_lovebird1.txt
BaselineBasis                          1
Precision                              4
SearchLevel                            2
Filter                                 1
MatchingMethod                         0 #####
TemporalEnhancement                    0 #####
Threshold                              1.5
MatchingBlock                          3
SoftDistanceCoeff                      0 #####
SoftColorCoeff                         0 #####
SoftBlockWidth                         11
SoftBlockHeight                        11
ImageSegmentation                      0 #####
SmoothingCoefficient2                  0 #####
SegmentationMethod                     3
MaxCluster                             32
DepthEstimationMode                    0 #####
FileCenterManual
ThresholdOfDepthDifference             10
MovingObjectsBSize                     1
MotionSearchBSize                      0
RefDepthCameraName
RefDepthFile
```

In the above configuration file the parameters that were tested are the following (marked with ##### in the configuration file):

Parameter name	***Variable name***	***Value***
SmoothingCoefficient	S	1, 2
MatchingMethod	M	0, 1, 2, 3
TemporalEnhancement	T	0, 1
SoftDistanceCoeff	Mm	0(=not used), 1 (10), 2 (20)
SoftColorCoeff	Mm	0(=not used), 1 (20), 2 (40)
ImageSegmentation	I	0, 1

SmoothingCoefficient2	Is	0, 2
DepthEstimationMode	SA	0, 1, 2, 3

The term "*variable name*" in the above matrix describes the codification for these parameters throughout all the experiments (both EE1 and EE4 for ES2). For example the depth map with file name: lovebird1_cam04_depth_S1_M0_Mm0_T0_I0_Is0_SA0.yuv (as it is mentioned in the above configuration file) corresponds to SmoothingCoefficient 1, MatchingMethod 0, etc. The presented results henceforth will make heavy use of the above codification without any more explanation. It should also be mentioned that there are some specific combinations of parameters that are not compatible with the available software tools. In such cases of parameter combinations the experiments were simply not performed. These combinations wont be mentioned here but can easily be deduced from the corresponding software manuals.

It is important to note here that for EE1 with view 4 the Semi-Automatic Depth Estimation (DepthEstimationMode; SA) was set to zero due to lack of manual input files for view 4 whereas all three values (SA0, SA1 and SA2) were tested for views 6 and 8.

The following figures (Fig. 8-12) will present depth maps for view 4 and frame #1.

Mm0 T0 I0 Is0

Mm0 T1 I0 Is0

Mm0 T1 I1 Is2

Figure 8: Depth maps for S1 and M0

Mm0 T0 I0 Is0

Mm0 T0 I1 Is2

Mm0 T1 I0 Is0

Mm0 T1 I1 Is2

Figure 9: Depth maps for S1 and M1

Mm0 T0 I0 Is0

Mm0 T0 I1 Is2

Mm0 T1 I0 Is0

Figure 10: Depth maps for S1 and M2

Figure 11: Depth maps for S1, M2 and Mm1

Figure 12: Depth maps for S1, M2 and Mm2

Similar results will be added here for views 6-8.

EE1: View synthesis

For the view synthesis part of EE1 VSRS was used. A sample configuration file for synth view 5 can be seen in the following:

```
DepthType                1
SourceWidth              1024
SourceHeight             768
StartFrame               0
TotalNumberOfFrames      10
LeftNearestDepthValue    9346.102154
LeftFarthestDepthValue   136662.003208
RightNearestDepthValue   9346.102154
RightFarthestDepthValue  136662.003208
CameraParameterFile      cam_param_lovebird1_original.txt
LeftCameraName           param_loveb1_004
VirtualCameraName        param_loveb1_005
RightCameraName          param_loveb1_006
LeftViewImageName        lovebird1_cam04.yuv
RightViewImageName       lovebird1_cam06.yuv
LeftDepthMapName         lovebird1_cam04_depth_S1_M0_Mm0_T0_I0_Is0_SA0.yuv
RightDepthMapName        lovebird1_cam06_depth_S1_M0_Mm0_T0_I0_Is0_SA0.yuv
OutputVirtualViewImageName 5f46_VSS0_VSB1_S1_M0_Mm0_T0_I0_Is0_SA0.yuv
SynthesisMode            0 #####
ColorSpace               0
Precision                2
Filter                   1
BoundaryNoiseRemoval     1 #####
ViewBlending             1
#SplattingOption         2
#BoundaryGrowth          40
#MergingOption           2
#DepthThreshold          75
#HoleCountThreshold      30
#TemporalImprovementOption 1
#WarpEnhancementOption   1
#CleanNoiseOption        1
```

In a similar manner to depth estimation, some parameters in the VSRS configuration files were tested for view synthesis. These are marked also with ##### and the actual values tested are the following:

Parameter name	***Variable name***	***Value***
SynthesisMode	VSS	0, 1
BoundaryNoiseRemoval	VSB	1, 2

After the synthesis of virtual views PSNR values between the original and the synthetic views were estimated. The following figures will present the PSNR values under different parameter configurations in an effort to isolate the most prominent parameter(s). It must be noted before we present the results that the best overall PSNR value was 30.8527dB and it was achieved for the following configuration: VSS0_VSB1_S2_M2_Mm0_T1_I1_Is2_SA0 which translates to:

Depth estimation:

- SmoothingCoefficient 2
- MatchingMethod 2
- TemporalEnhancement 1
- SoftDistanceCoeff not used
- SoftColorCoeff not used
- ImageSegmentation 1
- SmoothingCoefficient2 2
- DepthEstimationMode 0

View synthesis:

- SynthesisMode 0
- BoundaryNoiseRemoval 1

For the four possible combination of parameters in VSRS the average PSNR over all the resulted synthesized views are the following:

Mean VSS0_VSB1: 29.6694dB
Mean VSS1_VSB1: 21.4550dB
Mean VSS0_VSB2: 29.6694dB
Mean VSS1_VSB2: 29.6694dB

From the above results we can see that the parameters VSS1 and VSB1 result in very bad reconstructed views whereas the other three combinations don't have any statistically significant effect. It is also interesting to notice that the values are the same which is kind of curious.

Moreover, it is interesting to quantify the effect of various depth estimation parameters in the PSNR of the synthetic view. For that we will present average PSNR values for the parameters that were used in depth estimation. Since VSS1_VSB1 gave bad PSNR values we will exclude them from any further investigation. The following matrix contains mean and variance values for PSNR when a specific parameter is kept constant. We should note that M2 achieve the highest average PSNR and at the same time it delivered the lowest variance among all cases.

Parameter	Mean value (in dB)	Variance
S1	29.5594	0.3035
S2	29.8633	0.3981
M0	29.3083	0.1903
M1	29.3653	0.1727
M2	30.5656	0.0817
M3	29.5640	0.1481
Mm0	29.7317	0.4741
Mm1	29.4644	0.1498
Mm2	29.7883	0.0782
T0	29.7571	0.3802
T1	29.5765	0.3225
Is0	29.2846	0.3420
Is2	29.9260	0.2055
SA0	29.6694	0.3570

EE4 – MVC

During the MVC part of EE4 the best synthesized in terms of PSNR was used for coding. The following two figure (Fig. 13 and 14) depict the PSNR errors for different QP. The QP selected range from 20 until 40 in an interval of 2.

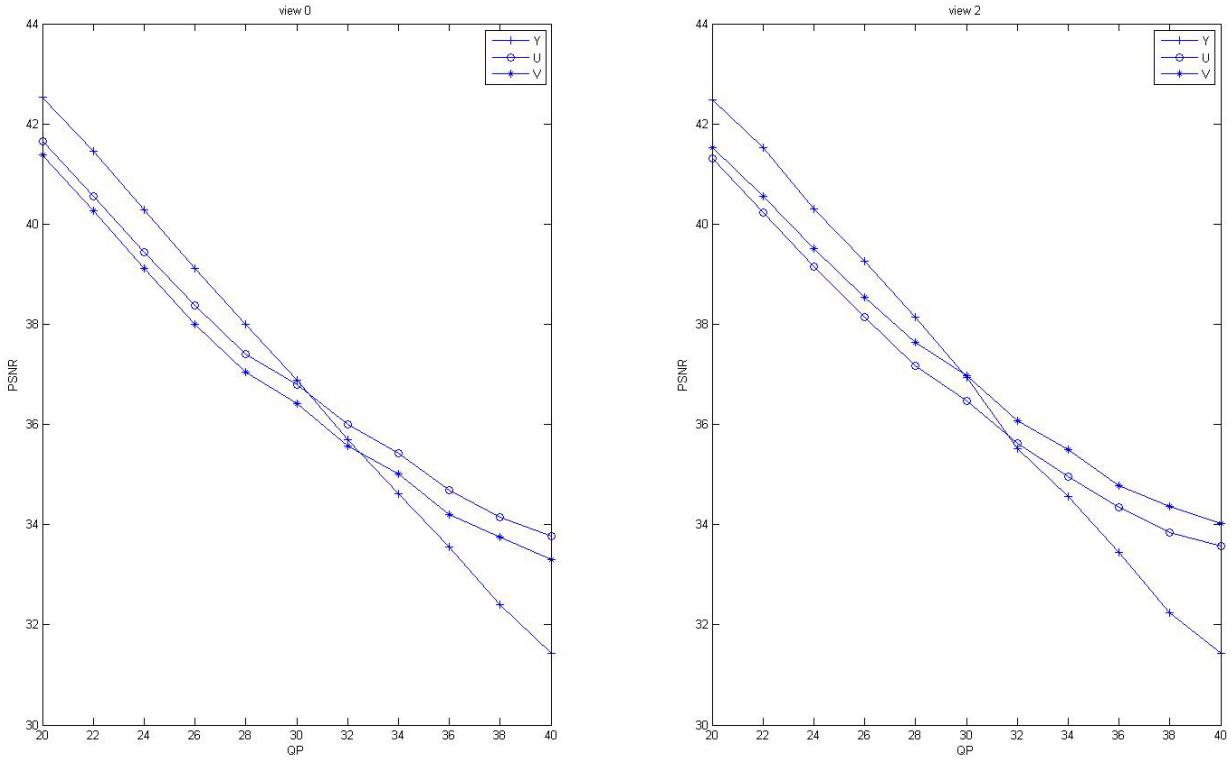

Figure 13: Average PSNR curve per channel and view for the 2 view configuration.

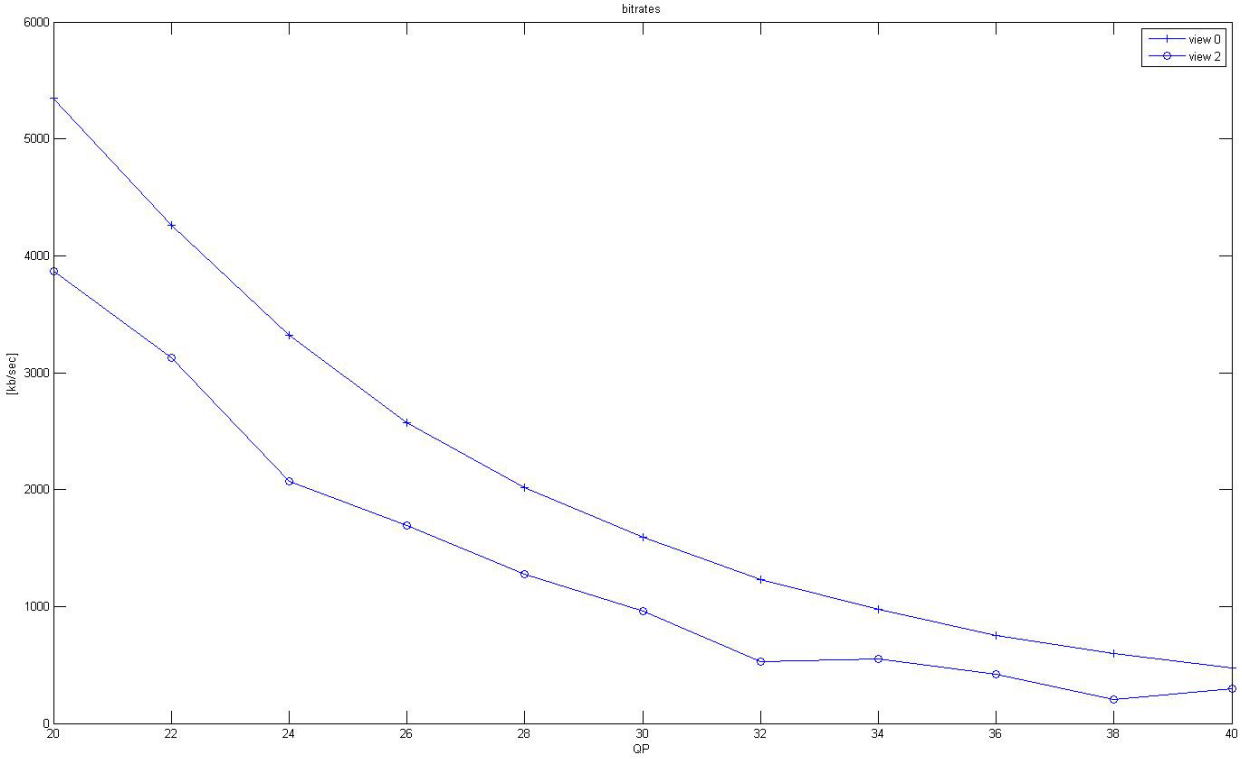

Figure 4: Average bit rate curve per view for the 2 view configuration.