M22697
Poznan University of Technology proposal for Call on 3D Video Coding Technology

Olgierd Stankiewicz **(ostank@multimedia.edu.pl),**

T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, J. Stankowski, K. Wegner

team supervisor: ***Marek Domański***

Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology, Poland

November, 2011, Geneva

Overview 1/2

- HEVC-compatible category
  - Compatible substream of the base view
  - Can be applied to MVC
- Multiview + Depth
- Inter-view prediction
  - View-synthesis (DIBR)
  - Disparity compensation (MVC-like)
  - Depth-based motion prediction (DBMP)
- Several other 3D coding tools

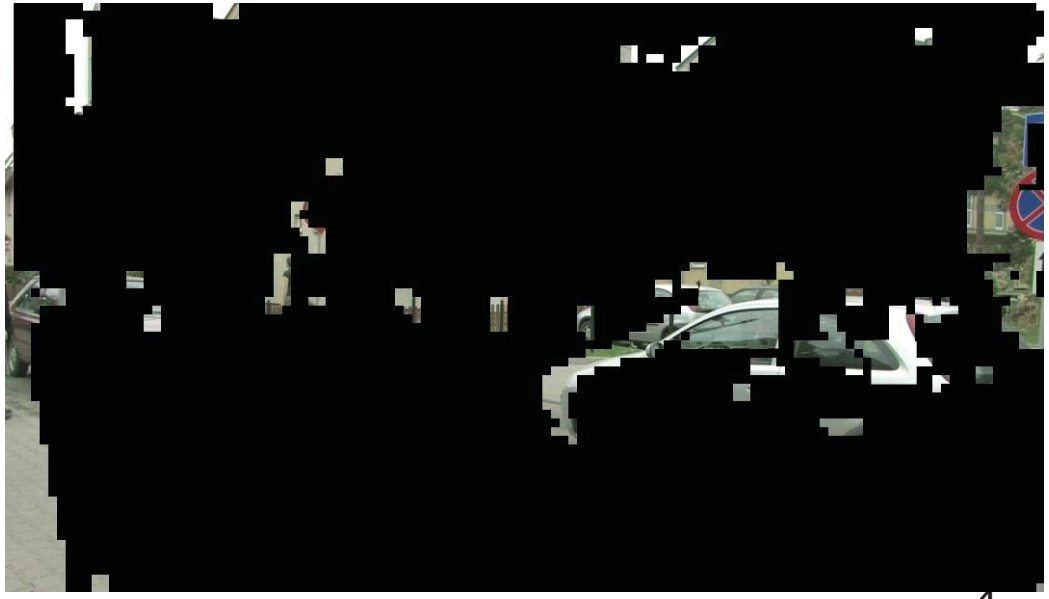
Overview 2/2

- View-synthesis is the primary method
  - Already coded views are used as references for DIBR-based synthesis
  - The base view
    - Always coded directly without synthesis
  - The side views
    - Only disoccluded regions are coded
    - Some coding units (CUs) are skipped
    - Information is derived in the decoder

Disocclusion coding

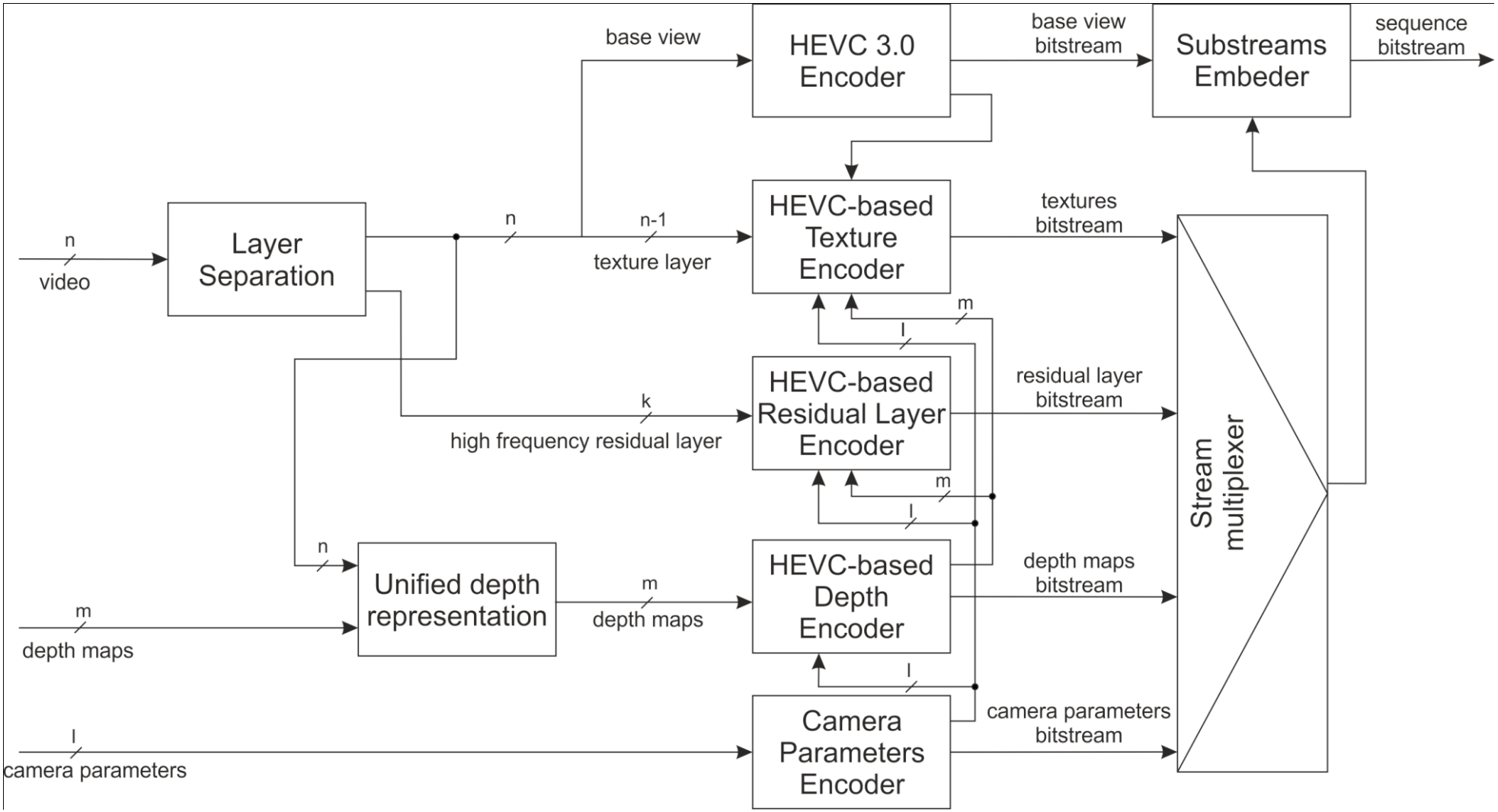
The original side view

Disocclusions in the side view



Coded CU-s in the side view

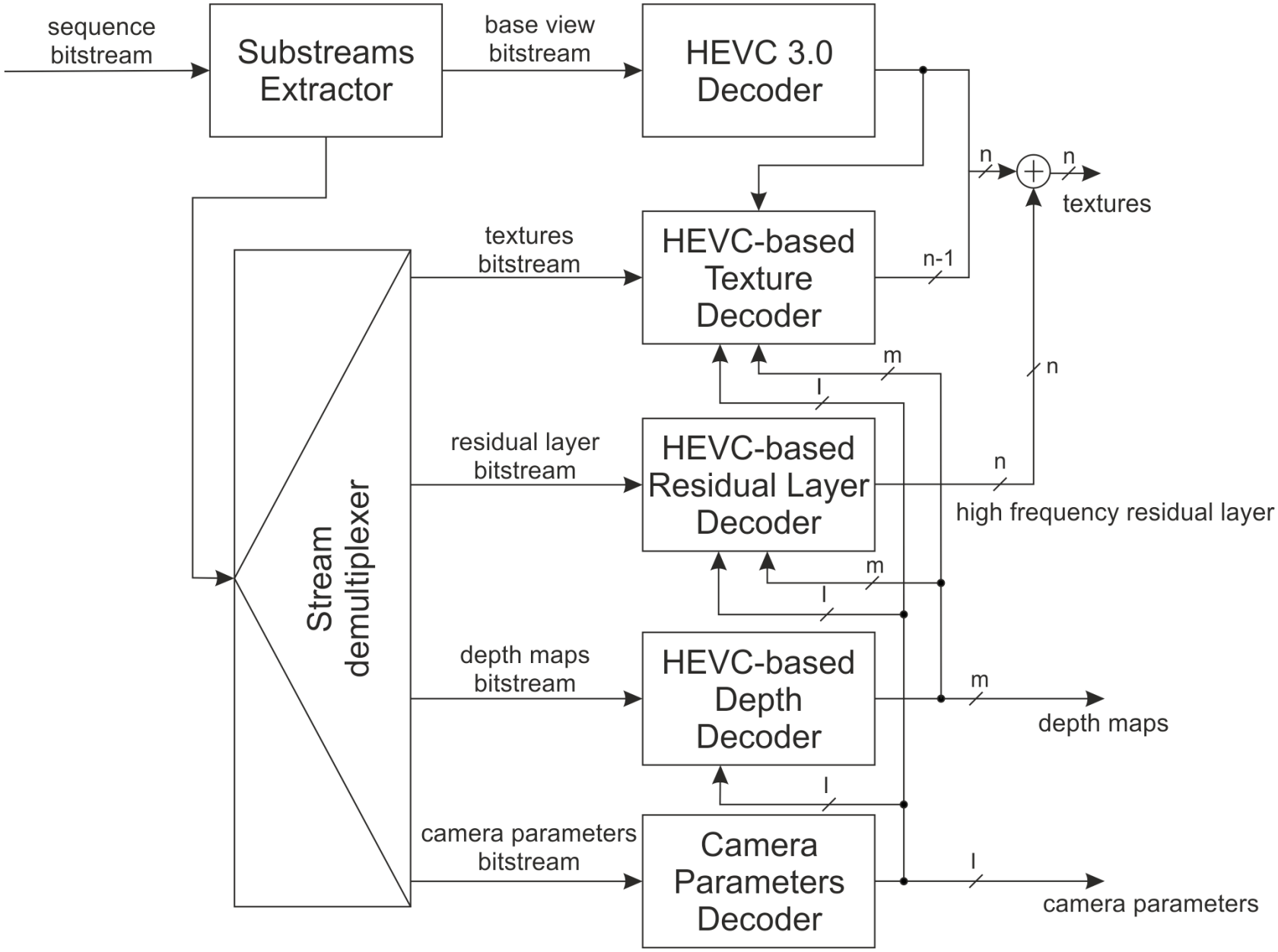
Coder scheme



Transmitted bitstream

- Base view
  - HEVC-compatible syntax
  - Standard NAL units
- Textures (side views), Depths (all views), High frequency residuals
  - HEVC-based syntax
  - Transparently encapsulated in NAL units
- Camera parameters SEI messages
- Texture1,Depth1, Depth0,Texture0, Depth2,Texture2 …

Decoder scheme



View synthesis 1/3

- Used in the side-views
- Based on VSRS
- Used for
  - Content prediction
  - Disocclusion detection
    - Availability map
    - RD-opt choses Coding Units that are actually coded in the side views

Availability map

RD-opt actually coded CUs

View synthesis 2/3

- Improvements
  - Depth-Gradient-based Loopback Filterer (DGLF)
    - Synthesized texture is adaptively filtered basing on gradients in the corresponding depth
    - Better prediction and subjective quality

Regions with high intensity of the corresponding depth-gradients that are filtered

View synthesis 3/3

- Improvements
  - Availability Deblocking Loopback Filter (ADLF)
    - Provides smooth transition between coded (in rectangular CUs) and synthesized regions
    - Better prediction and subjective quality

What we want to encode

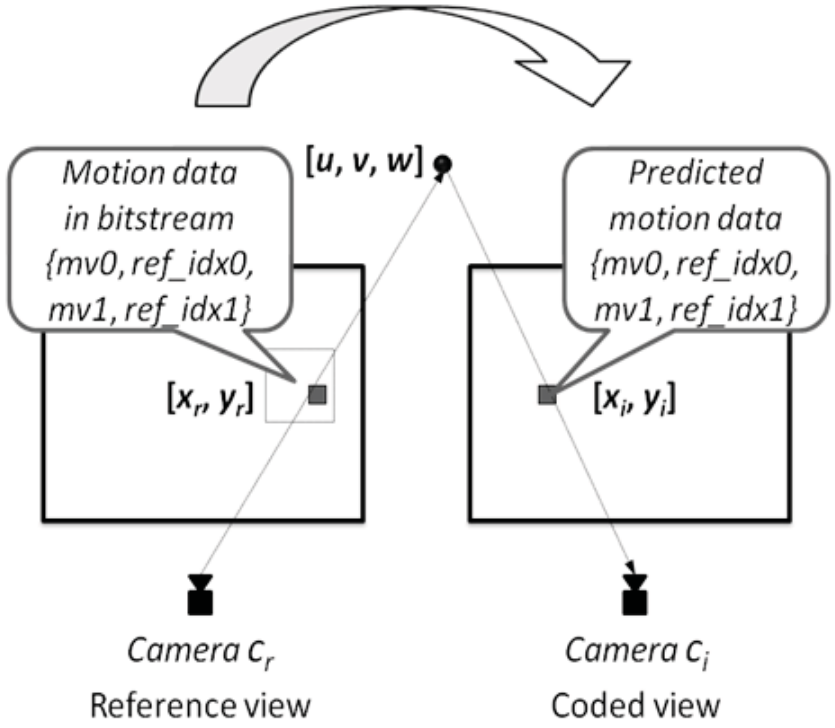
What is actually coded in CUs

MVC toolset in HEVC

- Disparity-compensation
  - Inter-view „motion” vectors
  - Already coded frames from neighbouring views are added on the reference lists

Depth-Based Motion Prediction (DBMP)

- Motion fields:
  - of neighbouring views are highly correlated
  - are inferred from already encoded CUs in the neighbouring views at the same temporal instance
- Additional merge candidate



Texture layer separation

- Input video is split into two layers
  - Texture layer (base)
    - Motion-compensated temporal filtering
  - Residual layer (high frequency)
- Both layers are transmitted
- Approach similar to:
    - SVC or wavelet coding
    - Perceptual Noise Substitution (PNS) in audio coding
- Residual layer can be coded jointly among the views

Unified depth representation

- All depth maps are:
  - Merged into a unified depth representation
  - Then projected back to views
  - Mid-level hypothesis technique

- Views-synthesis based coding
  - Only disoccluded regions in the side views
  - Full depth information only in the base view

Base view coding

- HEVC-compatible
  - Syntax
  - Default NAL units

- GOP size enhancement
  - Normally, only powers of 2 are allowed i.e. 16
  - Any GOP size is alowed,
  - GOP sizes 12 and 15 have been used for coding of test material

Depth coding 1/3

- HEVC-based syntax
- View-synthesis prediction in order to find disoccluded regions
  - Only those regions (CUs) are coded
- Inter-view disparity compensation (MVC-like)
  - Inter-view „motion” vectors
  - Like in MVC
- Inter-view depth-based motion prediction (DBMP)
  - Motion vectors are predicted from neighbouring views with use of depth maps

Depth coding 2/3

- Non-linear depth representation
  - The foreground is subjectively more important than background
    - Artifacts the in the background cannot be nocited
    - Influences only PSNR
  - Power-law is internally used for coding

$$depth\ value\ internal = \left(\frac{depth\ value\ external}{maximum\ value\ external}\right)^{exponent} \cdot maximum\ value\ internal$$

  - Closer objects are represented more accurately than distant objects

Depth coding 3/3

- Z-near – Z-far compensation
  - Range of represented depth (disparities) among frames can be different
  - All frames on the reference list are compensated (scaled) prior to prediction
  - Impose no change if Z-near-Z-far range is equal in all frames

Side view texture coding 1/2

- HEVC-based syntax
- Used only for the side-views
- View-synthesis prediction in order to find disoccluded regions
  - Only those regions (CUs) are coded
  - Size is chosen by RD-opt
- Inter-view disparity compensation (MVC-like)
  - Inter-view „motion” vectors, like in MVC
- Inter-view depth-based motion prediction (DBMP)
  - Motion vectors are predicted from neighbouring views with use of depth maps

Side view texture coding 2/2

- Adjustment of QP parameter for texture
  - Based on the corresponding coded depth
    - Changed simultaneously in coder and decoder
    - No information is transmitted
  - Higher quality for the foreground
  - Lower quality for the background.

Residual layer coding

- Residual of motion-compensated temporal low-pass filtering of the input video
- High-frequency content
- Can be coded jointly among the views
- Modelled as a non-stationary random process
  - Spatial energy distribution
    - Image of energy measured in rectangular blocks
  - Spectral envelope
    - IIR filter coefficients –
      LAR (log-area-ratio) 8-bit representation

Stream multiplexing

- Bitstream is a sequence of standard NAL units
- The base view
  - Compliant with HEVC syntax – default NALUs
- Other streams (3D improved HEVC)
  - Are transparently encapsulated in NAL units
    - Previously marked as undefined in HEVC Working Draft
  - Can be skipped by monoscopic HEVC decoder

Software implementation

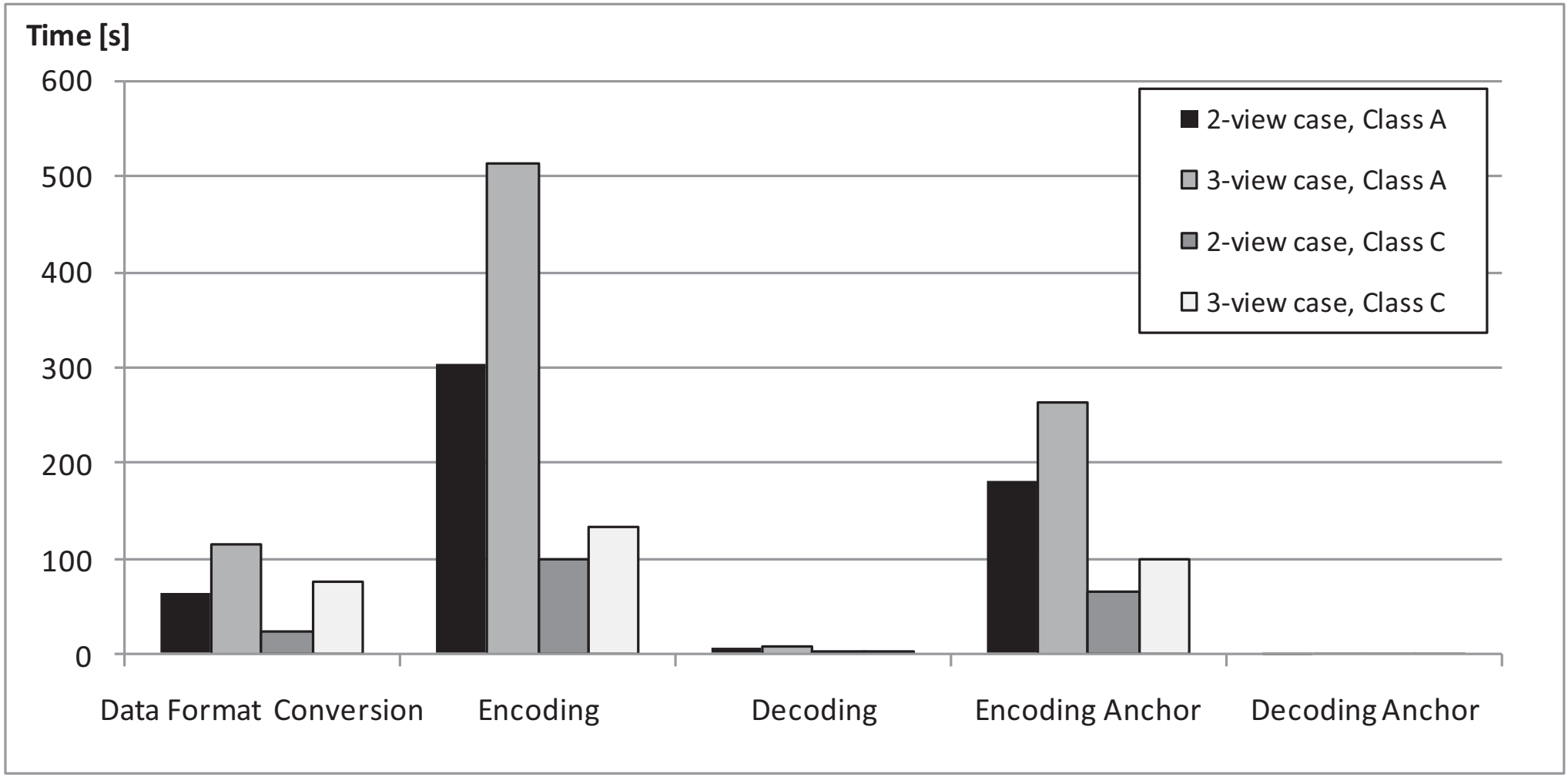
- C++, Microsoft Visual Studio 2005-2010
- Mostly based on HM3.0

- Available solely for scientific purposes for contributors of MPEG-3DV
  - ftp://multimedia.edu.pl/software
  - 3DV # ftvftv

- Ready to be used as a test model for 3DV

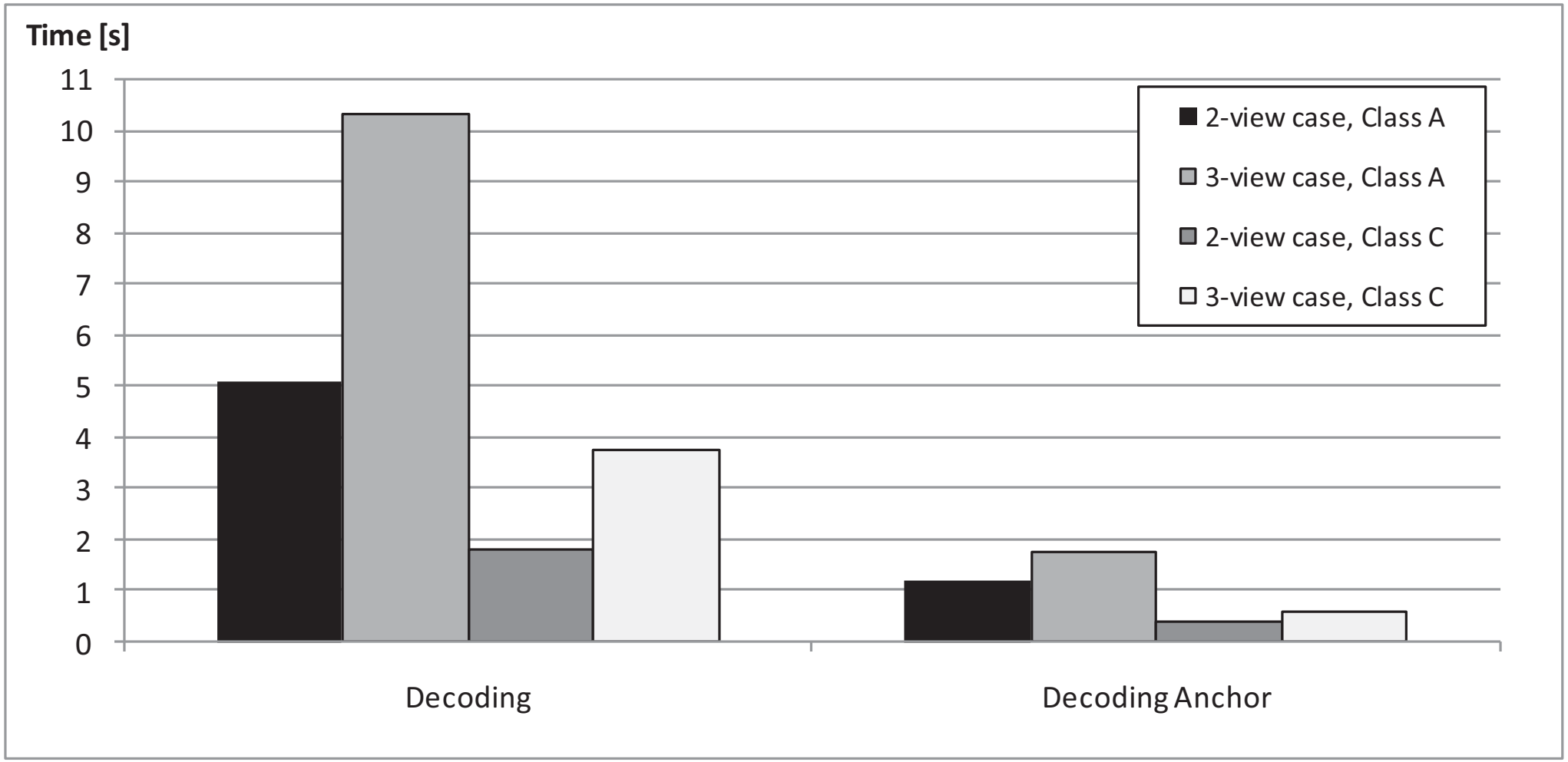
Complexity analysis

Implementation based on HM3.0 without optimizations!! Encoder – about 2,5x of Anchor.

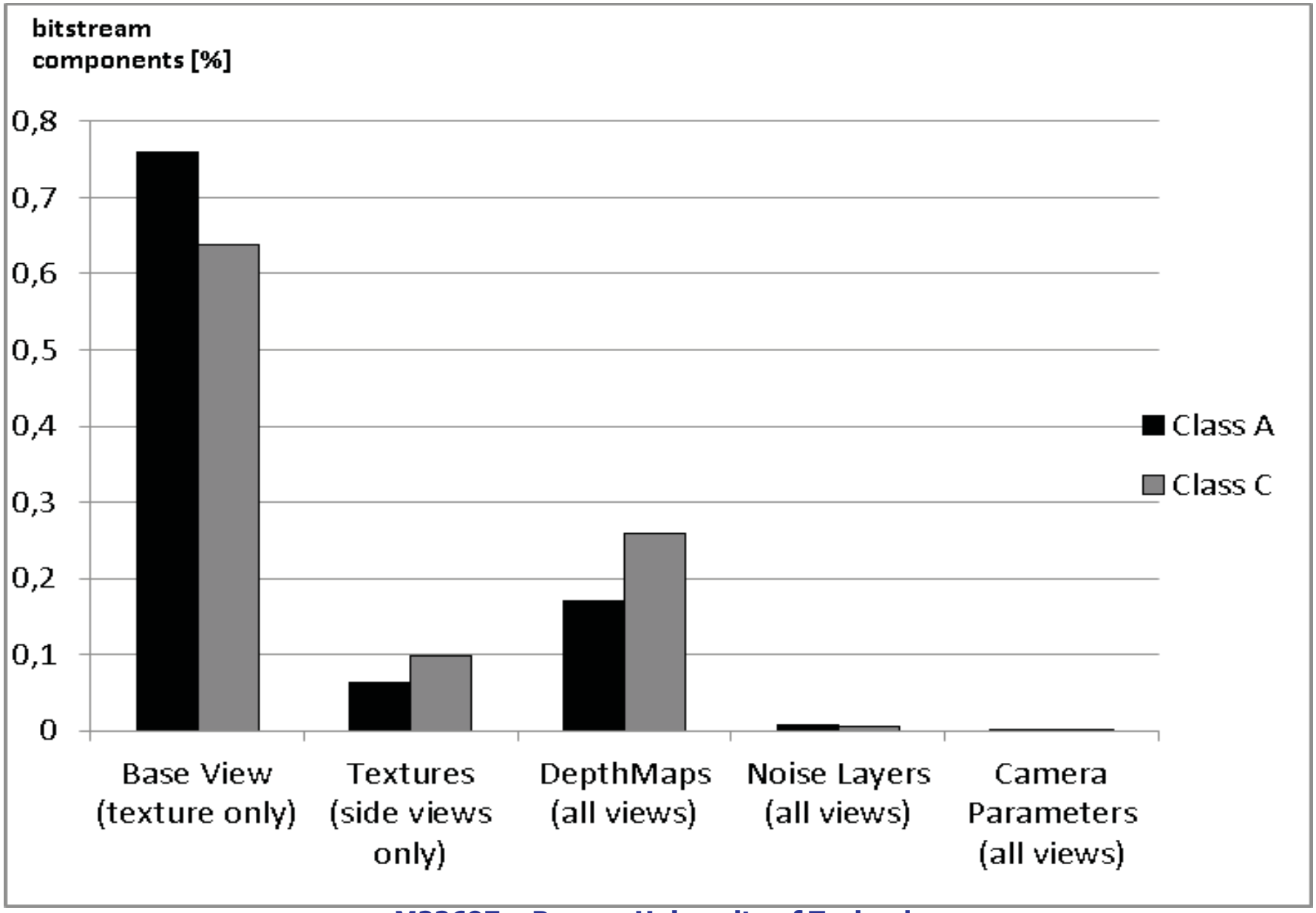


Complexity analysis

Implementation based on HM3.0 without optimizations!! Decoder – about 5,5x of Anchor.



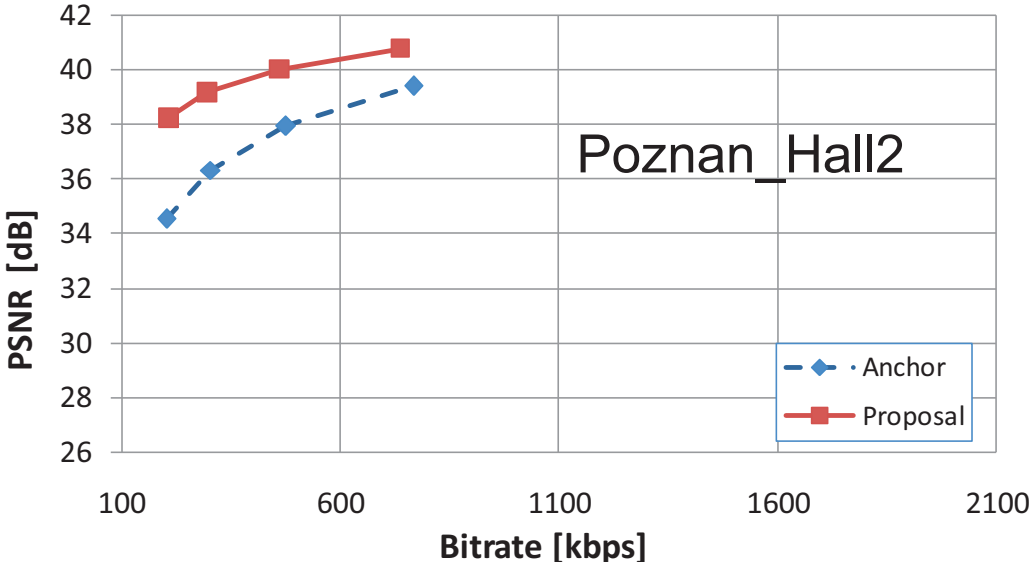
Substreams vs overall bitstream

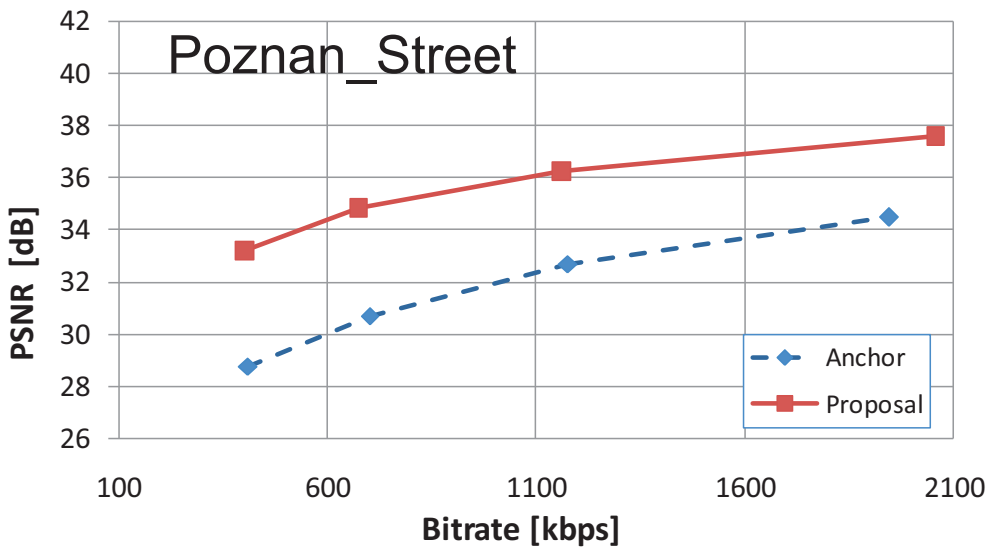


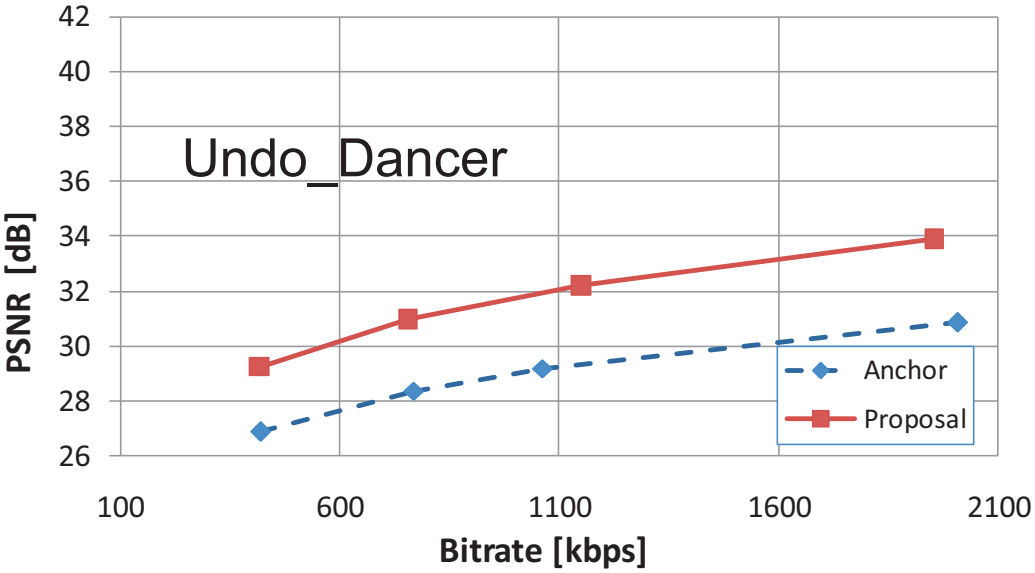
Substreams vs overall bitstream

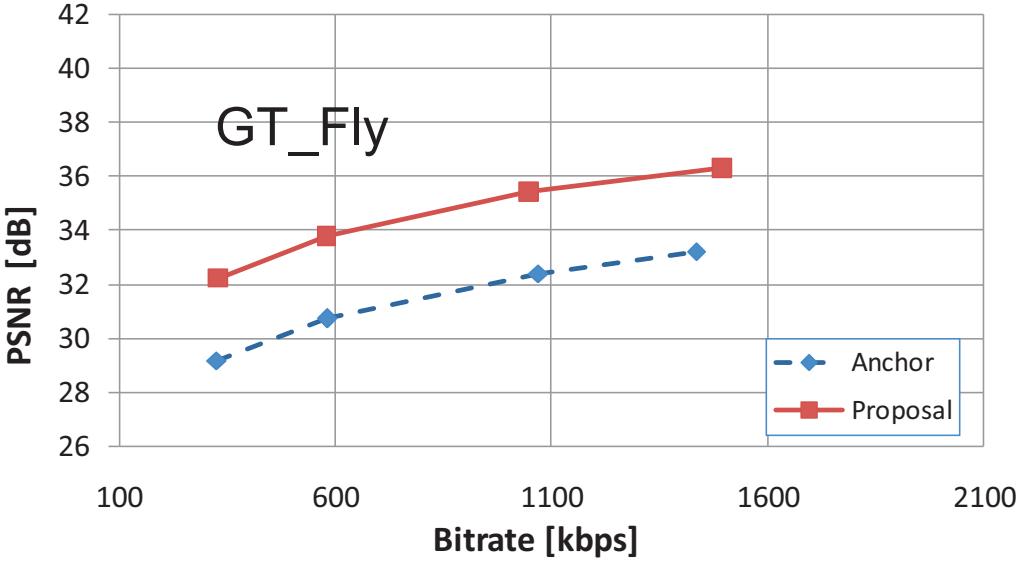
- The base view
  - About 75% of the bitstream in 2-view case
  - About 70% of the bitstream in 3-view case
- Other views, depth maps, residual layer, camera parameters
  - About 25-30% of the bitstream
  - Only vestigial data for disocclusion
  - Mostly reconstructed with use of view synthesis
  - PSNR is not adequate for the side views
  - PSNR of the base view used for objective assessment

Compression performace – 3-view case
PSNR of the base view only

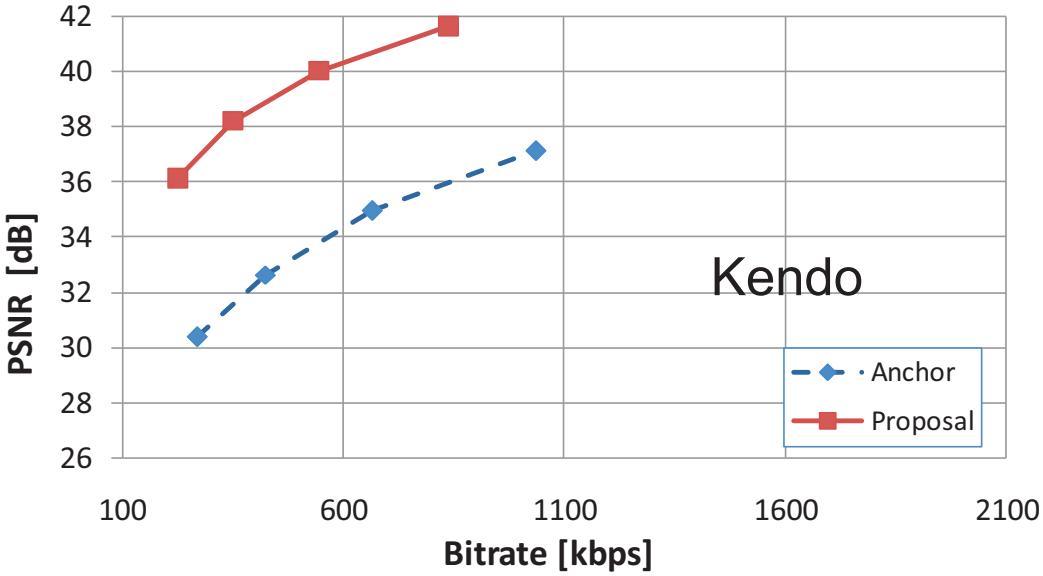


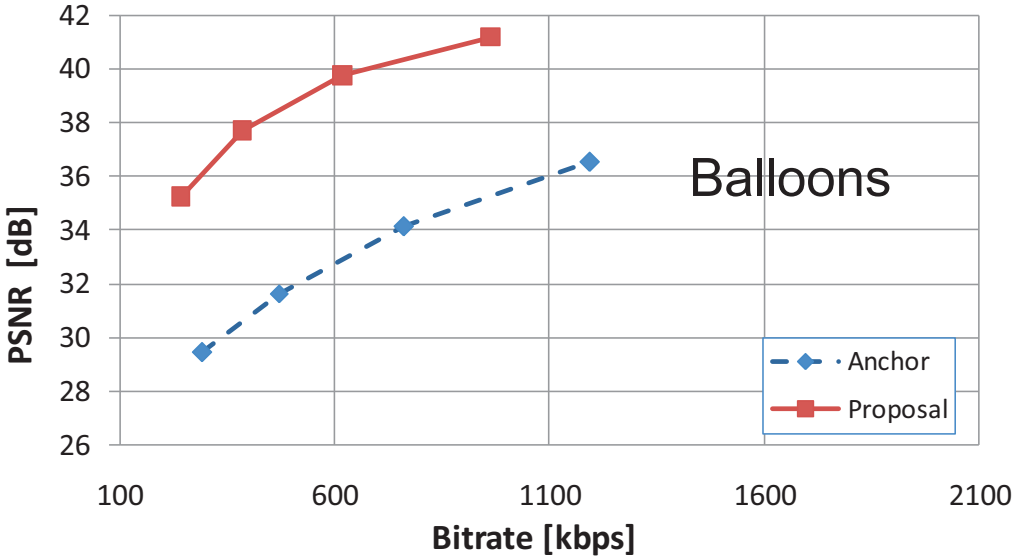


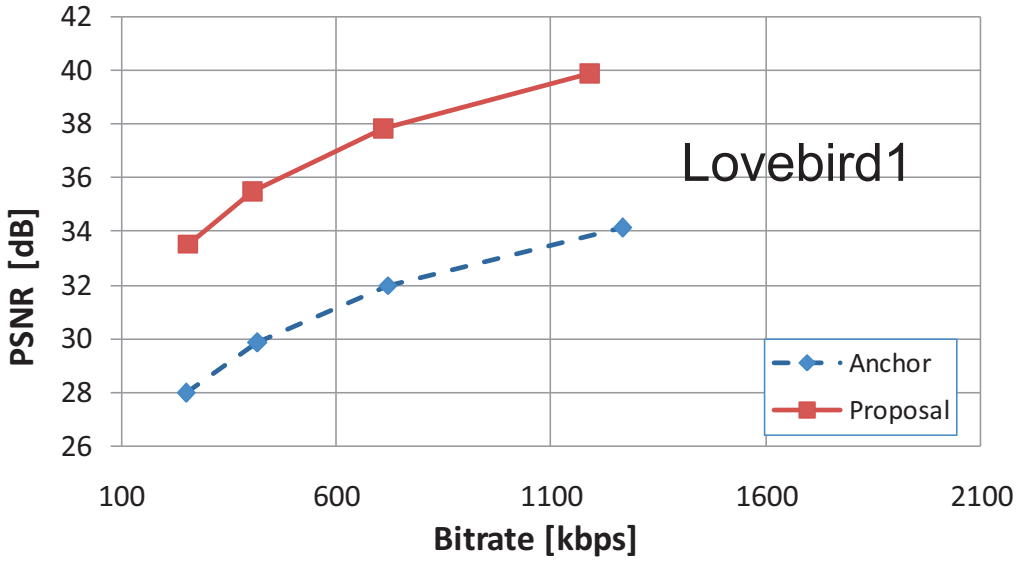


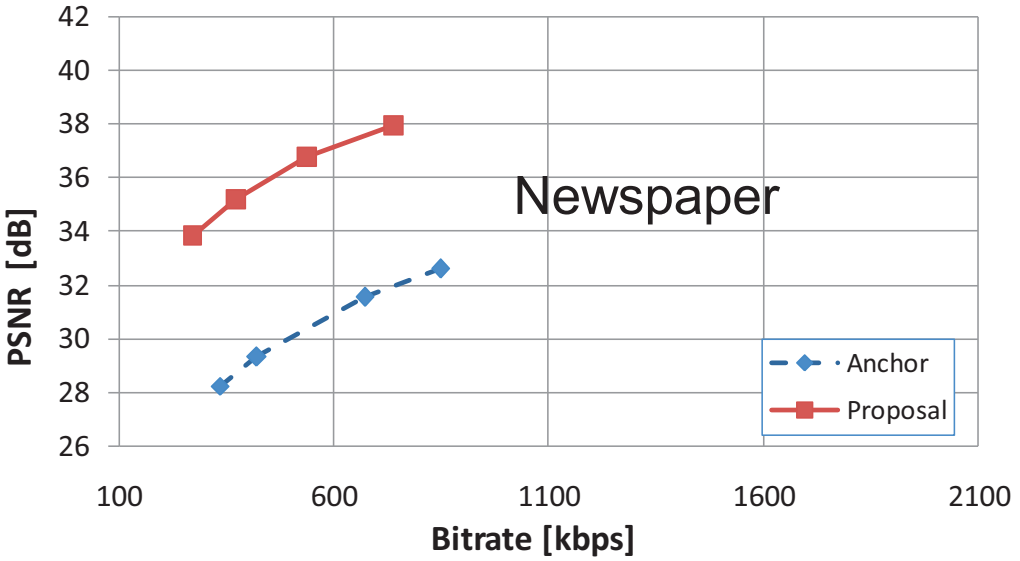


Compression performace – 3-view case
PSNR of the base view only

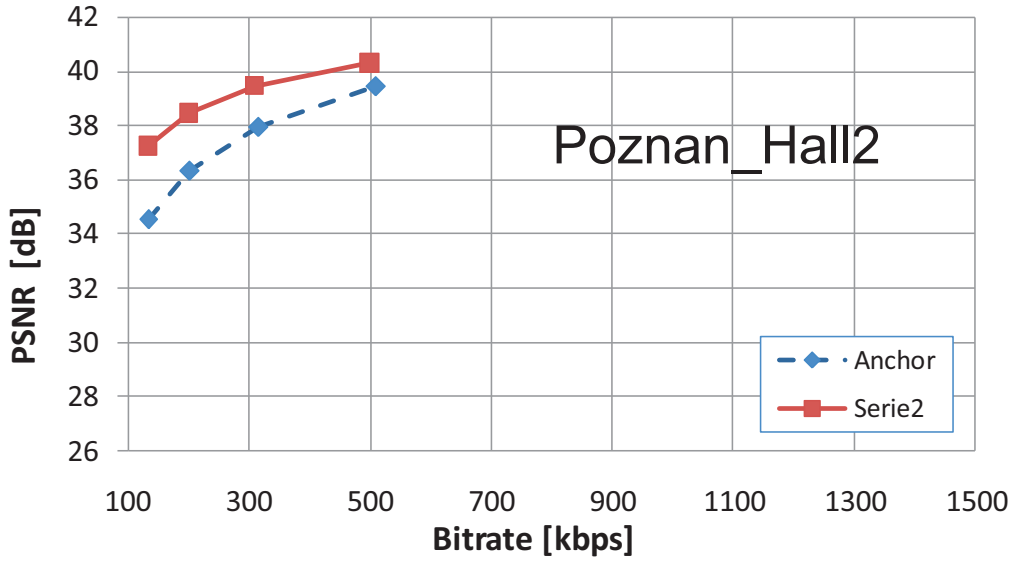


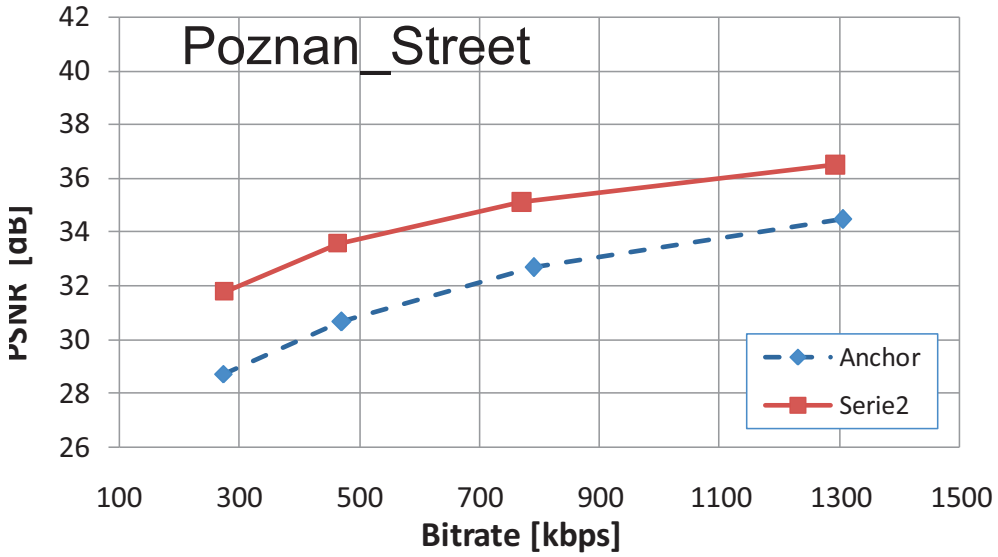


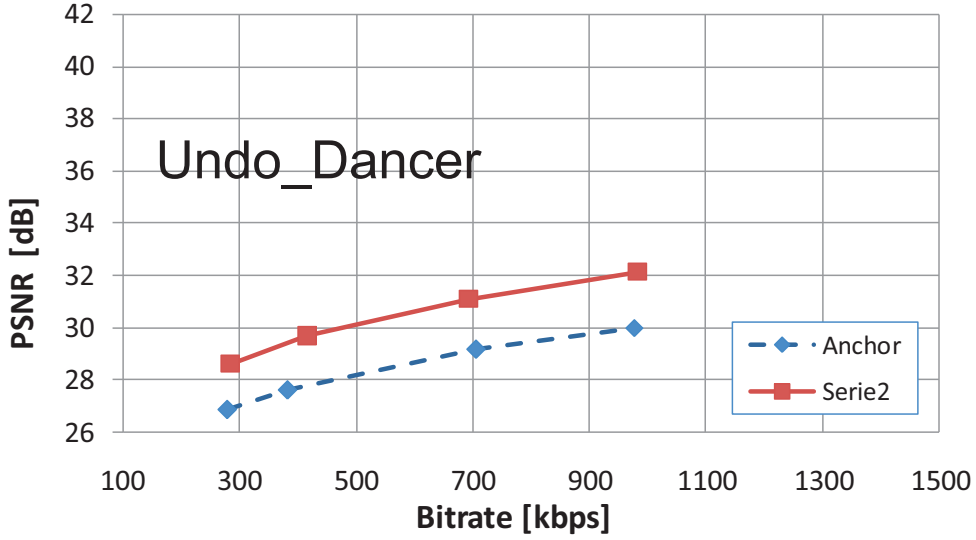


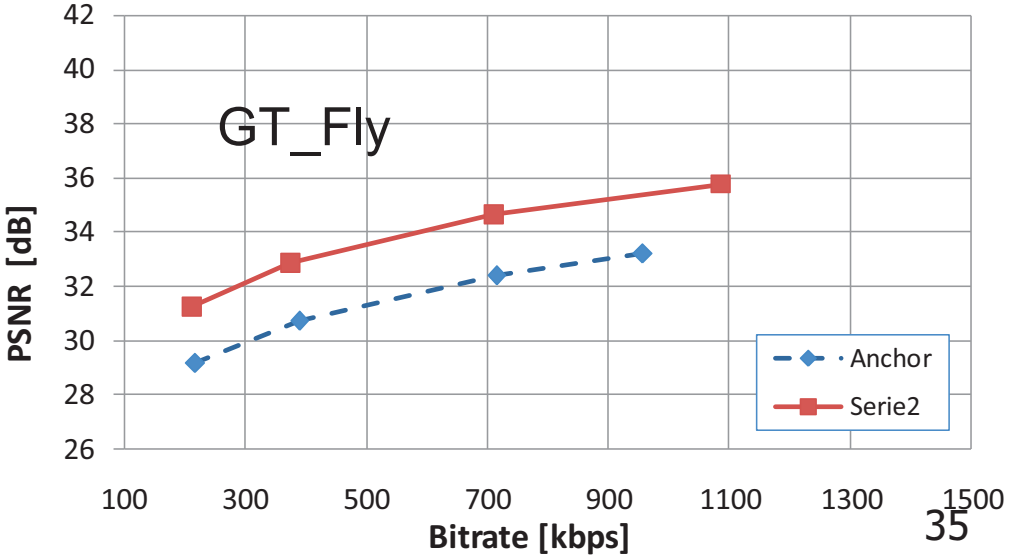


Compression performace – 2-view case
PSNR of the base view only

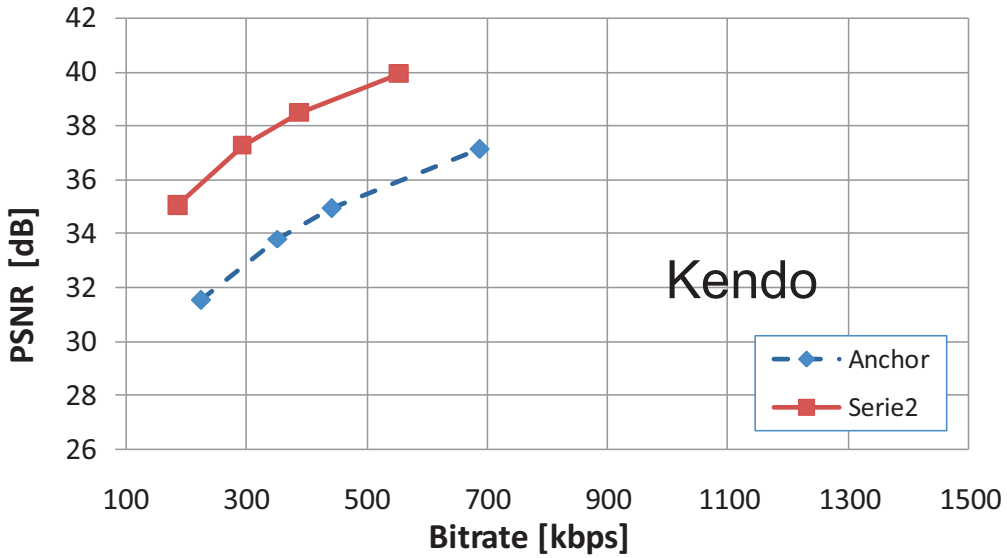


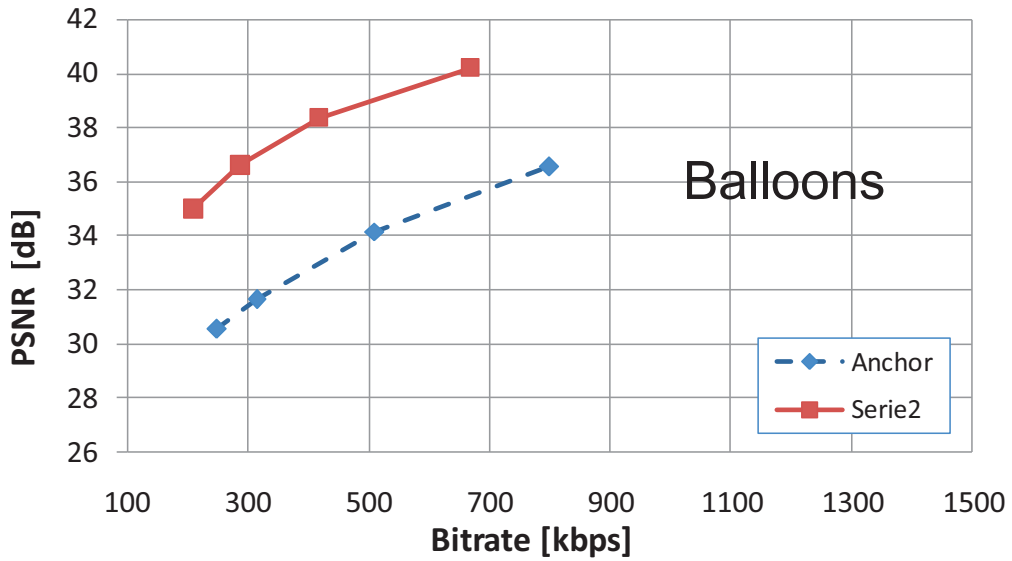


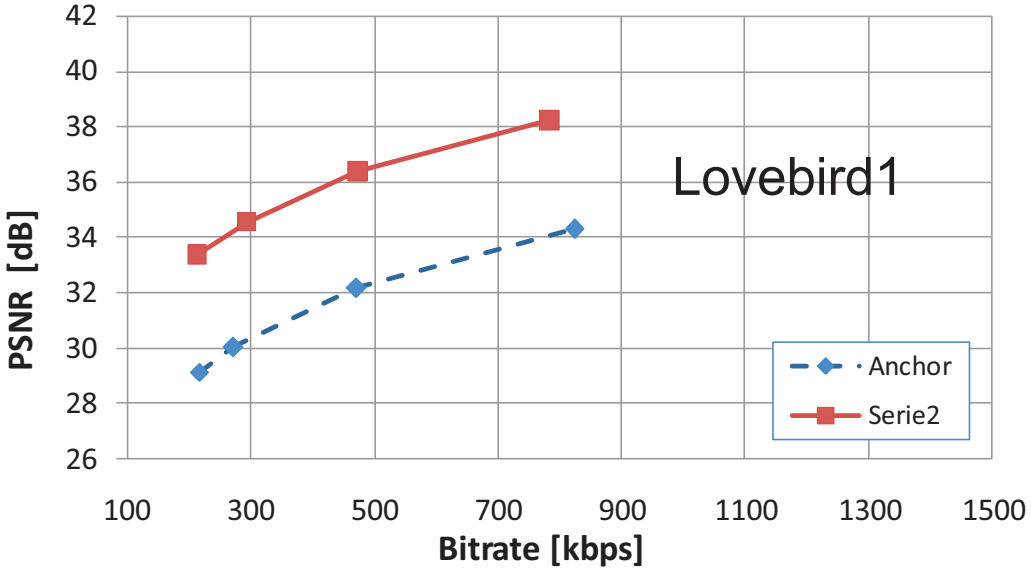


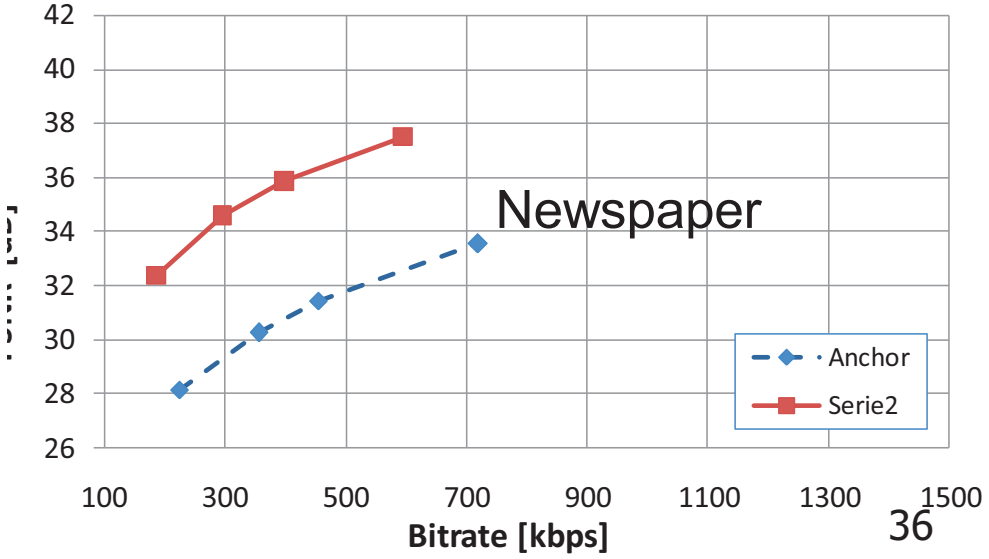


Compression performace – 2-view case
PSNR of the base view only









Bjontegaard – PSNR of the base view only

Sequence	2-view case		3-view case	
	ΔPSNR [dB]	ΔBitrate [%]	ΔPSNR [dB]	ΔBitrate [%]
Poznan_Hall2	1,8	-45,3	2,5	-59,3
Poznan_Street	2,7	-54,6	3,9	-69,6
Undo_Dancer	1,9	-51,0	2,7	-62,6
GT_Fly	2,2	-55,5	3,1	-67,0
Kendo	4,2	-58,8	6,2	-73,3
Balloons	5,2	-64,6	6,7	-74,3
Lovebird1	4,2	-66,7	5,8	-76,4
Newspaper	5,1	-66,5	6,3	-76,1
Average	**3,4**	**-57,9**	**4,7**	**-69,8**

Summary

- HEVC-compatible category
  - Implementation based on HM 3.0
  - The technology can be applied to MVC
- The base view and dissocclusions in the side views
- Several other 3D tools proposed
- PSNR is inadequate
- Questions?