MPEG-Immersive 3DoF+ Standard Work:

Related to 3DoF+ Call for Proposal

Presenter: Eun-Seok Ryu (esryu@gachon.ac.kr)

Jong-Beom Jeong, Dongmin Jang, Ju Hyeong Kim, SoonBin Lee, Eun-Seok Ryu

Multimedia Communications and Systems Lab. (MCSL)

http://mcsl.gachon.ac.kr

Department of Computer Engineering

Gachon University

Test Materials for 3DoF+ - w17726

- Philips
  - ClassroomVideo
- Technicolor
  - TechnicolorMuseum
  - TechnicolorHijack

Class A
ClassroomVideo (4096x2048), 360º x 180º FOV ERP,
30fps, 120frames, 15 source views

Class B
TechnicolorMuseum (2048x2048), 180º x 180º FOV ERP,
30fps, 300frames, 24 source views

Class C
TechnicolorHijack (4096x4096), 180º x 180º FOV ERP,
30fps, 300frames, 10 source views

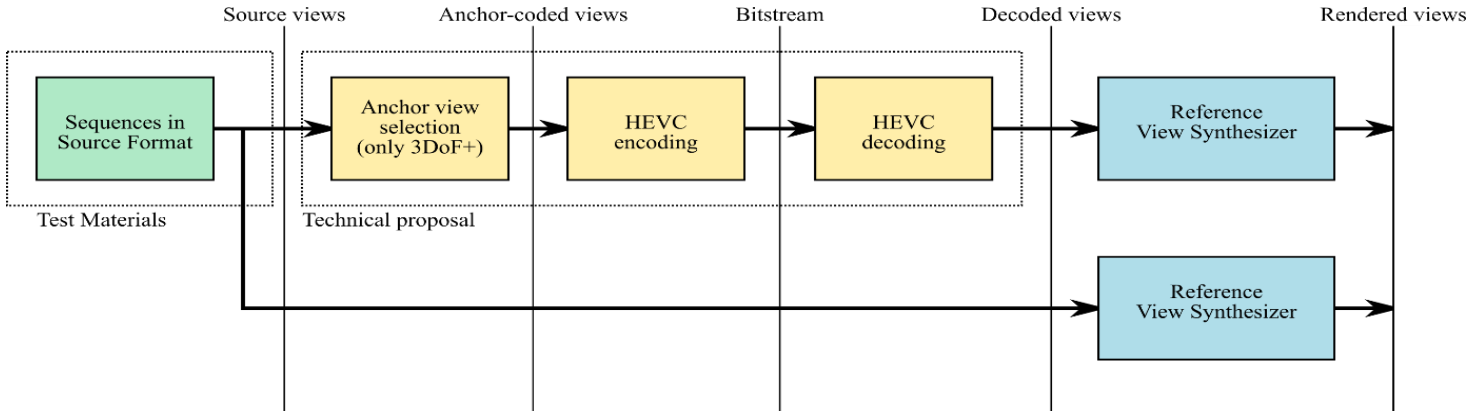
Common Test Conditions (CTC) for 3DoF+ - w17726

- Software

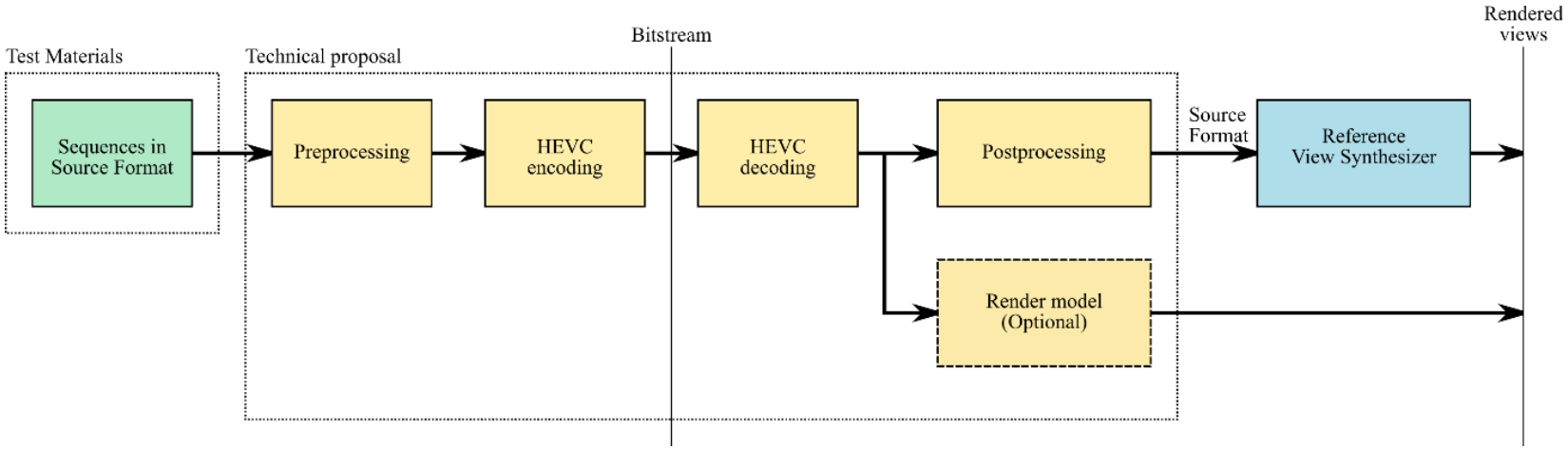
Software name	Location	Tag/branch
RVS	http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/RVS	v2.0.1
ERP-WS-PSNR	http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/ERP_WS-PSNR	v1.0
HDRTools	https://gitlab.com/standards/HDRTools/tree/0.18-dev	v0.18
360Lib	https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/branches/360Lib-5.1-dev	5.1-dev
HM	https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.16	16.16

Common Test Conditions (CTC) for 3DoF+ - w17726

- Anchor definition



Definition of the anchor



Technical proposal with pre- and post-processing

Common Test Conditions (CTC) for 3DoF+ - w17726

❖ Quantization Parameter and class

	QP1	QP2	QP3	QP4	QP5
Depth QP	12	17	22	27	32
Texture QP	22	27	32	37	42

QPs used for depth and texture

Test class	Sequence Name	No. of source views	No. of anchor-coded views	Anchor-coded views
A1	ClassroomVideo	15	15	All
A2	ClassroomVideo	15	9	v0, v7…v14
B1	TechnicolorMuseum	24	24	All
B2	TechnicolorMuseum	24	8	0, 1, 4, 8, 11, 12, 13, 17
C1	TechnicolorHijack	10	10	All
C2	TechnicolorHijack	10	5	1, 4, 5, 8, 9

Anchor-coded views per class

Common Test Conditions (CTC) for 3DoF+ - w17726

- 3DoF+ objective evaluation

Term	Description
Coded view position	The position of a source view which is coded by the anchor
Non coded source view position	The position of a source view which isn't coded by the anchor
Intermediate view position	The position of a view which is out of any source view

Definitions for test class on 3DoF+

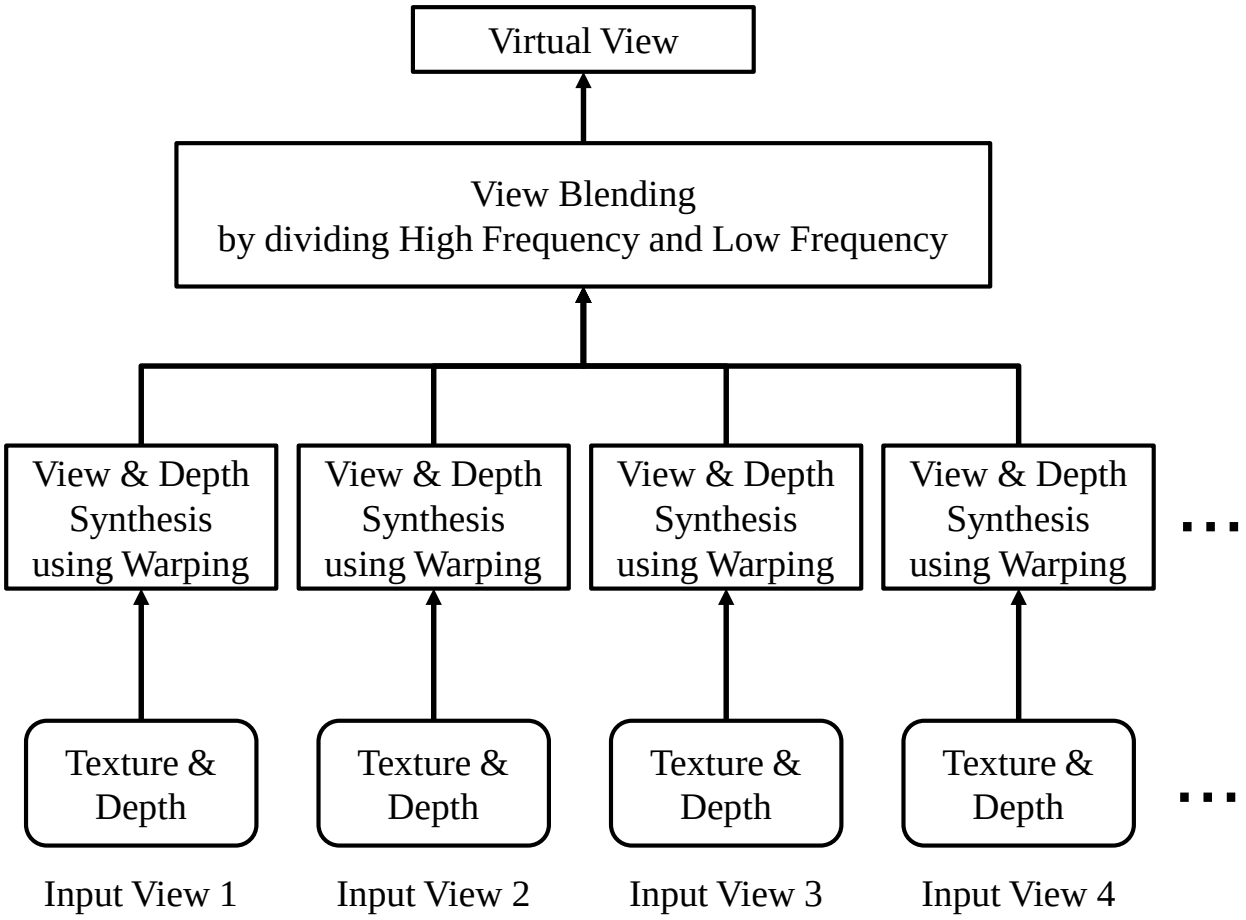
	WS-PSNR	
	For the anchor RD curve	For the proponent's RD curve
Coded BD rate	All frames Coded view & source view	All frames Proponent's view & source view
Non-coded BD rate	All frames Non-coded source view & source view	All frames Proponent's view & source view
Intermediate BD rate	All frames Intermediate view made by decoded views & same view made by all source views with 16-bit depth maps	All frames Proponent's intermediate view & Same view made by all source views with 16-bit depth maps

Requirements for 3DoF+ objective evaluation

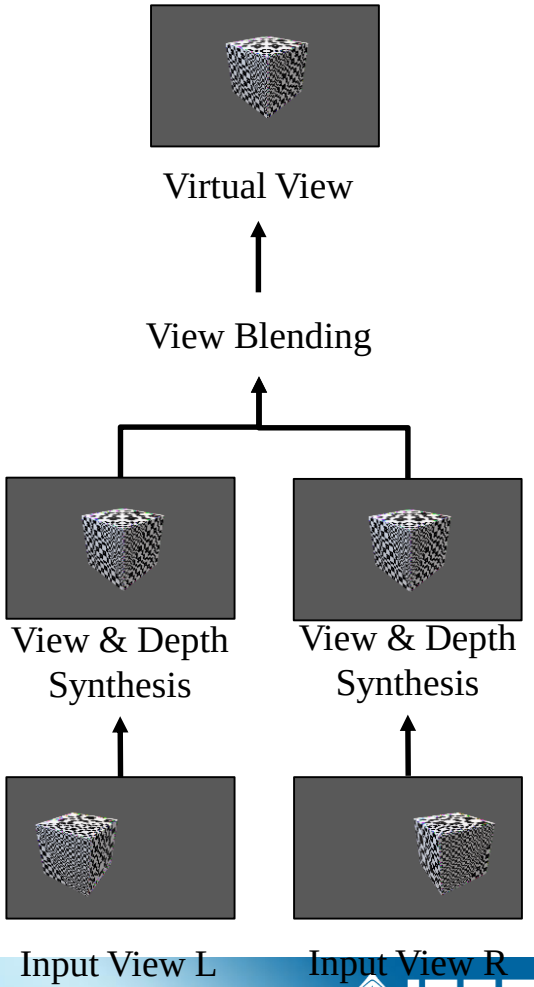
Reference View Synthesizer - w17759

- Reference software for view synthesis of 3DoF+

Conceptual Diagram

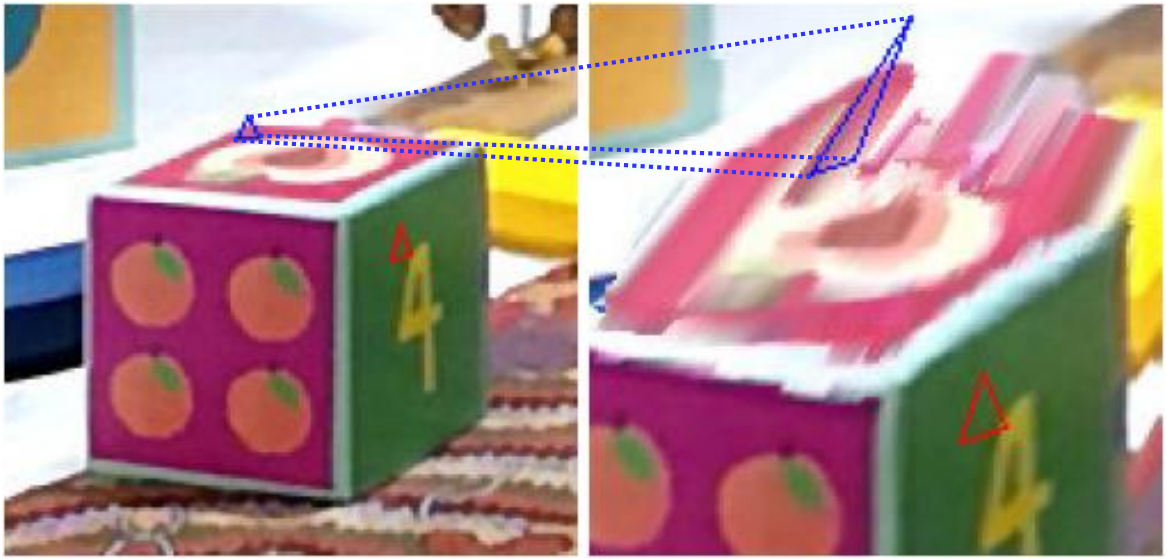


Simple Example



Reference View Synthesizer - w17759

- 3D Warping
  - Pixel movement between viewpoints of the same world coordinate
  - Input view is divided into triangles with the pixels centers as vertices
  - Use affine transformation
  - Being filled with interpolated colors after affine transformation



(a) Input view

(b) Obtained view after synthesizing view and depth

Reference View Synthesizer - w17759

❖ View synthesis

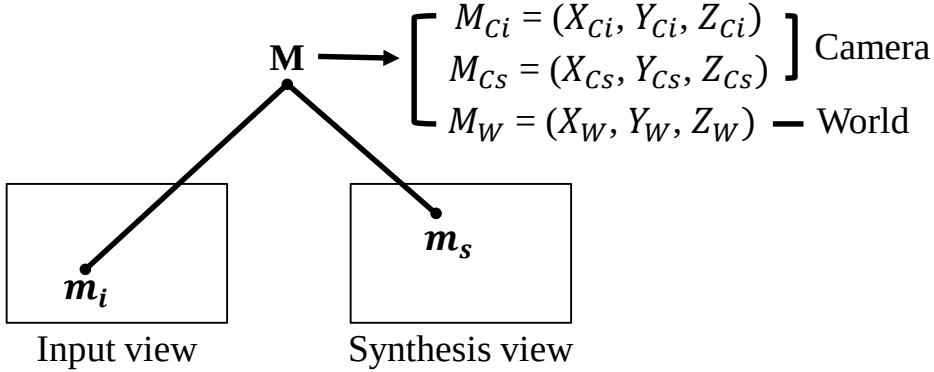


Fig1. Camera and world cordinates of input and synthesis views

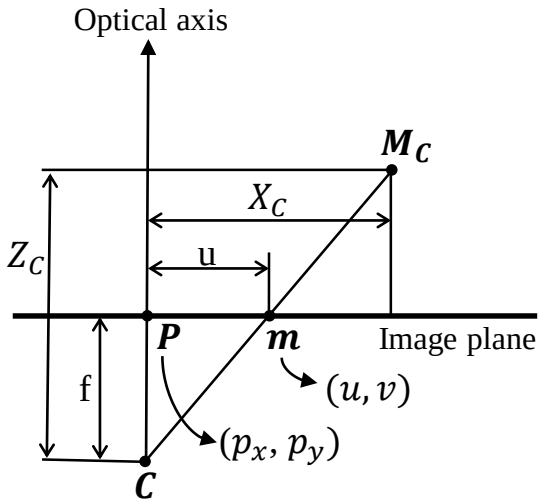


Fig2. Camera coordinate above the image plane

$$Z_C : f = X_C : u$$

$$u = \frac{fX_C}{Z_C} + p_x$$

$$v = \frac{fY_C}{Z_C} + p_y$$

Considering principal point

$$Z_C \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \end{pmatrix} = AM_C$$

$\tilde{m}$

Camera coordinate move by t and rotate by R from world coordinate

$$M_C = RM_w + t \quad \longleftarrow \quad M_C = \tilde{m}A^{-1}$$

$$\tilde{m} = ARM_w + At$$

$$M_w = R_i^{-1}A_i^{-1}\tilde{m}_i - R_i^{-1}t_i \quad \text{(Input view)}$$

$$R_i^{-1}A_i^{-1}\tilde{m}_i - R_i^{-1}t_i = R_s^{-1}A_s^{-1}\tilde{m}_s - R_s^{-1}t_s$$

$$\tilde{m}_s = A_s R_s R_i^{-1} A_i^{-1}\tilde{m}_i - A_s R_s R_i^{-1} t_i + A_s t_s$$

Reference View Synthesizer - w17759

- ❖Good quality : low depth and triangle with a regular shape
- ❖Taking the pixel with the maximal quality would give a sharper result
- ❖Taking the weighted mean is more resistant to errors
- ❖High and low frequencies are separated with a mean blur
- ❖Low frequencies are blended with the weighted mean
- ❖High frequencies are blended by choosing the pixel of highest weight

(a) (b) (c)

(a): Blending by argmax
(b): Weighted mean
(c): Multi-spectral blending: argmax - high frequencies
weighted mean - low frequencies

$$w_i = \left(\frac{q_i}{d_i}\right)^{\alpha}$$

d_i: depth at pixel for synthesized view i

$$C_{low} = \sum_{i=0}^{n} w_i c_i^l \Big/ \sum_{i=0}^{n} w_i$$

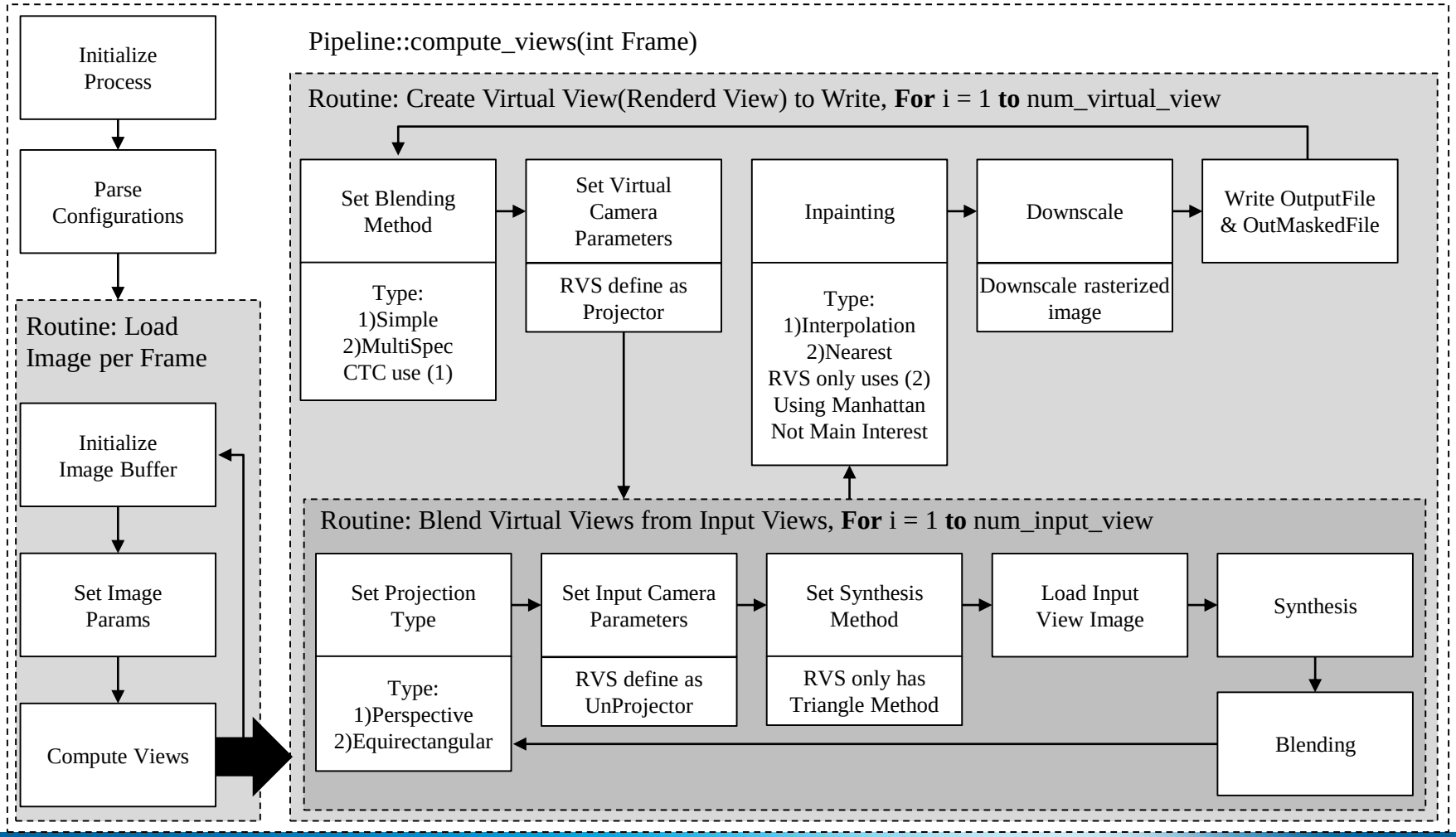
$$C_{high} = argmax_{c_i^h}(w_i)$$

c_i^l , c_i^h: color of the pixel in view i for the low and high frequencies

Reference View Synthesizer - w17759

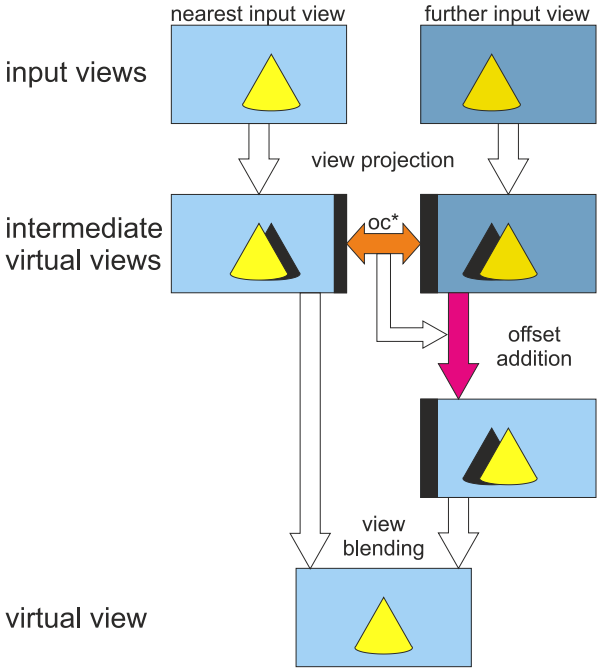
❖ Architecture

Pipeline Object (Pipeline.cpp)



Fast Color Correction for View Synthesis - m43694

- Fast color correction technique for view synthesis
- Reduction of color artifacts in synthesized images
  - Offset calculation
  - Offset addition



View Synthesis with Proposed Color Correction (oc* – offset calculation)

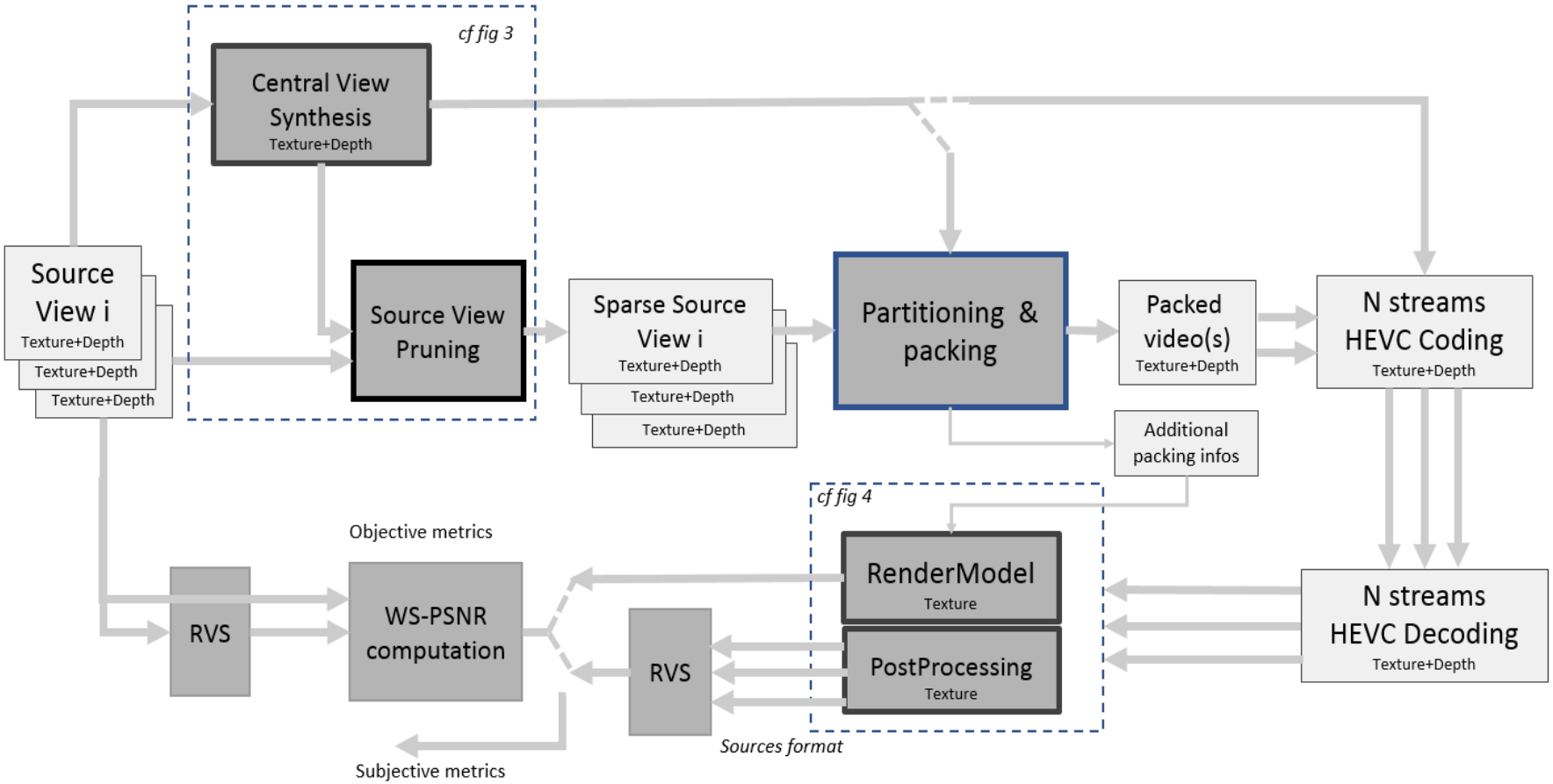
SVS, no color correction (31.7 dB)

SVS, with color correction (32.6 dB)

Test sequence	PSNR [dB]			
	VSRS		SVS	
	No color correction	With color correction	No color correction	With color correction
Ballet	31.45	**31.96**	31.72	**32.55**
Breakdancers	31.89	31.89	32.09	32.16
Poznan_Fencing2	26.14	**28.04**	26.79	**28.50**
Soccer Linear	34.80	34.80	34.92	34.94
BBB Butterfly	32.06	32.06	33.94	33.91
BBB Flowers	22.71	22.71	22.51	22.48

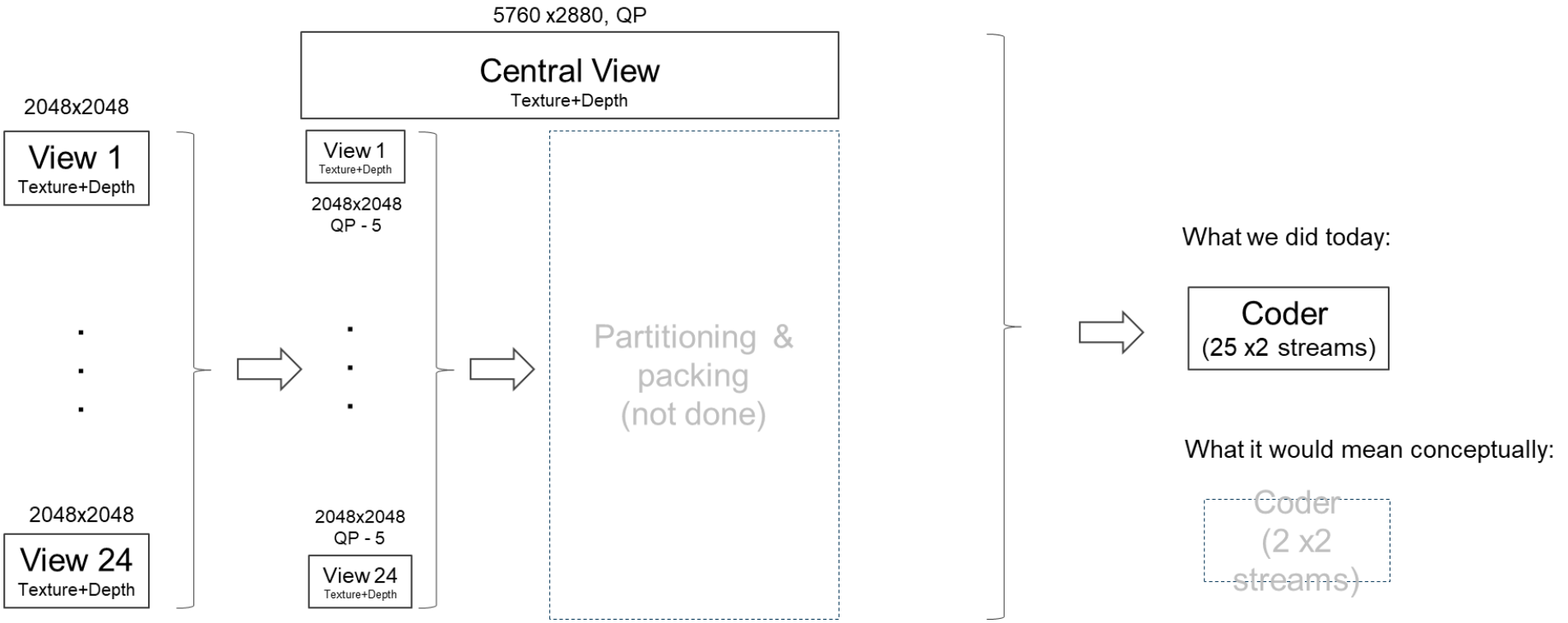
3DoF+ Software Platform Description - w17761

- 3DoF+ Software platform



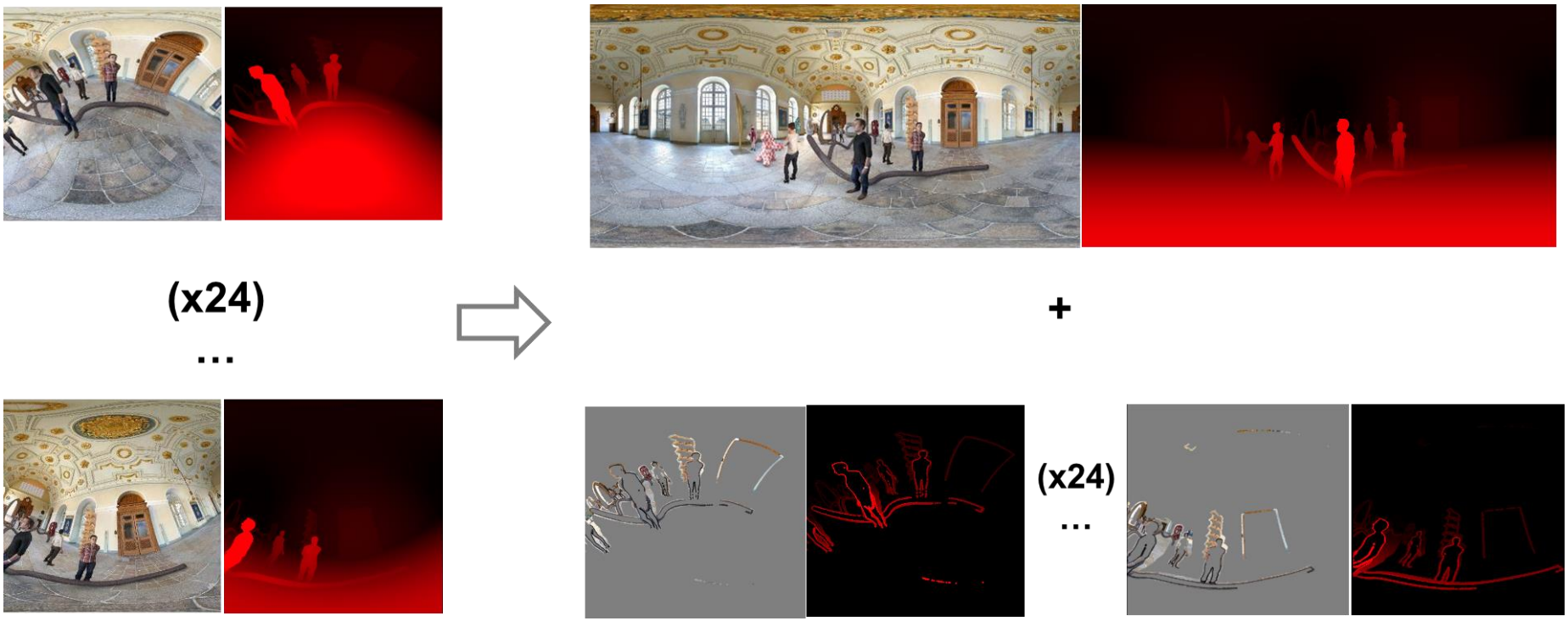
Outperforming 3DoF+ Anchors – m43504

- Removing overlap between anchors reduces data
- Send a central view and sparse views



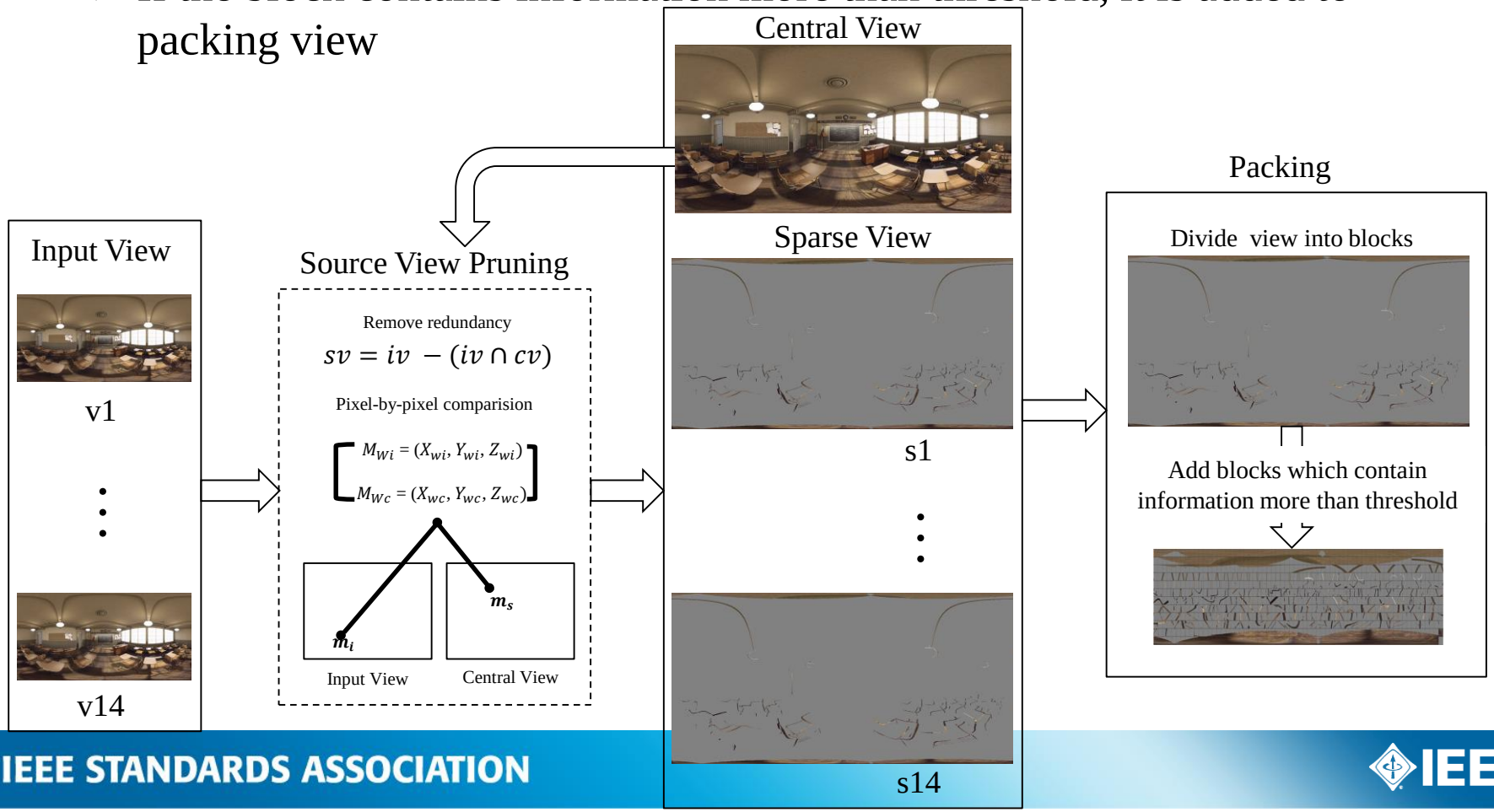
Outperforming 3DoF+ Anchors – m43504

- 1 central ERP + 24 semi ERP views



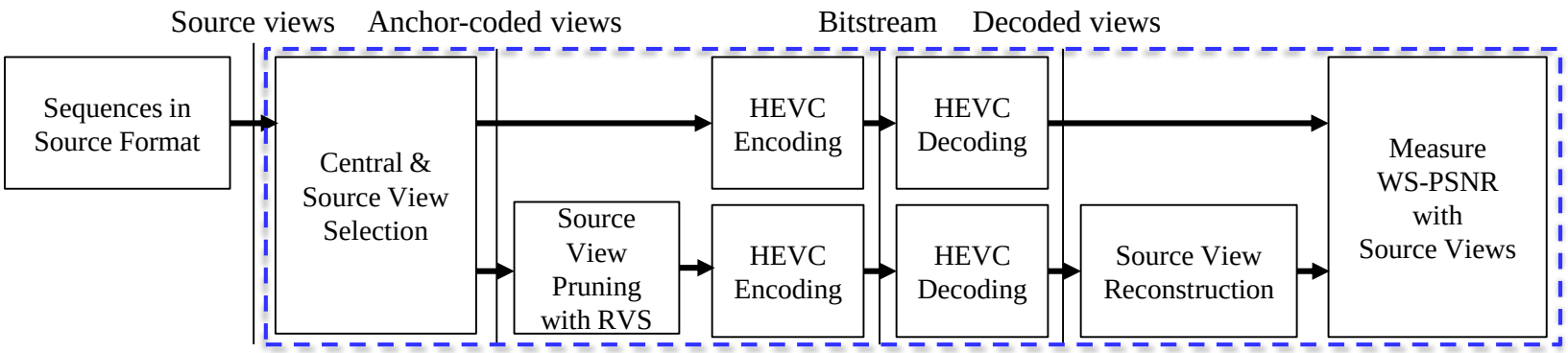
Source View Pruning & Packing

- Based on m43504, build source view pruning module with RVS
- Packing module divides the sparse views into small size blocks
- If the block contains information more than threshold, it is added to packing view



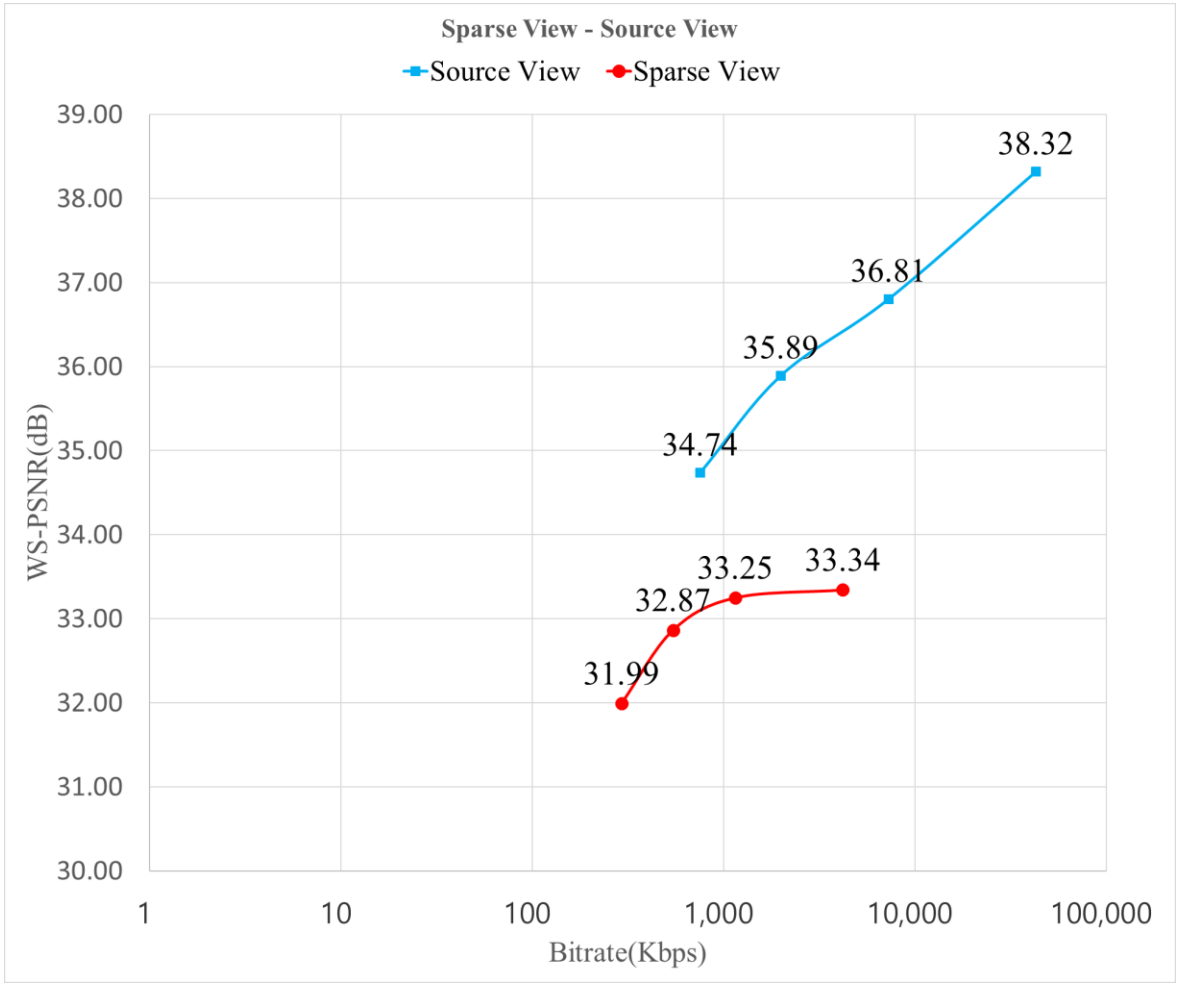
Source View Pruning & Reconstruction

- Architecture



Source View Pruning & Reconstruction

- RD-curve



Conclusion

- Motivation
  - 3DoF+ requires high resolution, large amount of videos
  - Multi-view video transmission needs multiple decoders
- Proposed method
  - Generate sparse view by removing overlap between views with RVS
  - Divide sparse view into blocks and add them to packing view
  - Reconstruct source view with central view and sparse views
- Results
  - Source view pruning saves bitrate with small loss of PSNR
  - Packing reduces the size of view to transmit
- Future work
  - Color correction technique with illumination compensation
  - Extensive experiment for optimal parameters