PHILIPS

www.philips.com

360° and 3DoF+ video

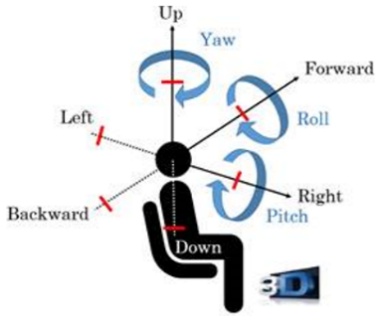
Workshop on Coding Technologies for Immersive Audio/Visual Experiences

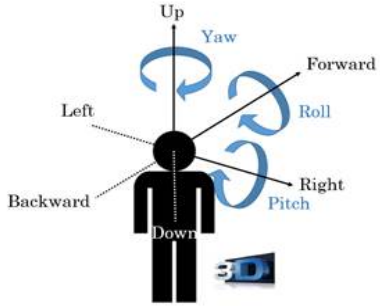
Bart Kroon
Philips Research Eindhoven
July 10, 2019

innovation you

Introduction

- 360° video: ability to look around (regular or stereo)
- 3DoF+ video: ability to look around and move head while standing or sitting on a chair
- 6DoF video: ability to look around and walk a few steps





What is OMAF?

It is a **systems standard**
developed **by MPEG**
that defines **a media format**
that enables **omnidirectional media applications,**
focusing on **360° video**, images, and audio, as well as
associated timed text.

NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

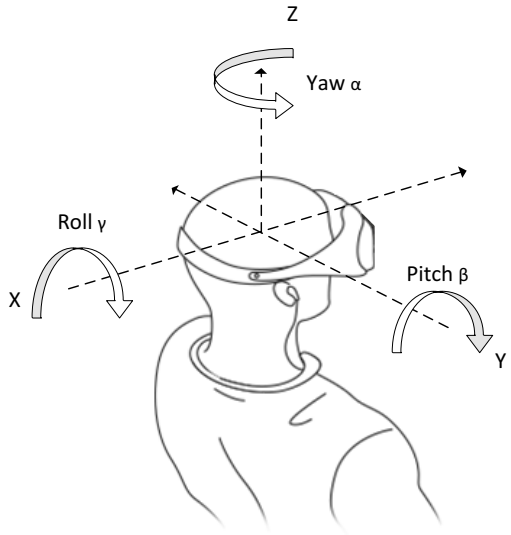
What is 360º video?

is a **simple version**

of **virtual reality (VR)** where only

3 degrees of freedom (3DOF)

is supported



The user's viewing perspective is from the center of the sphere looking outward towards the inside surface of the sphere.

Purely translational movement of the user would not result in different omnidirectional media being rendered to the user.

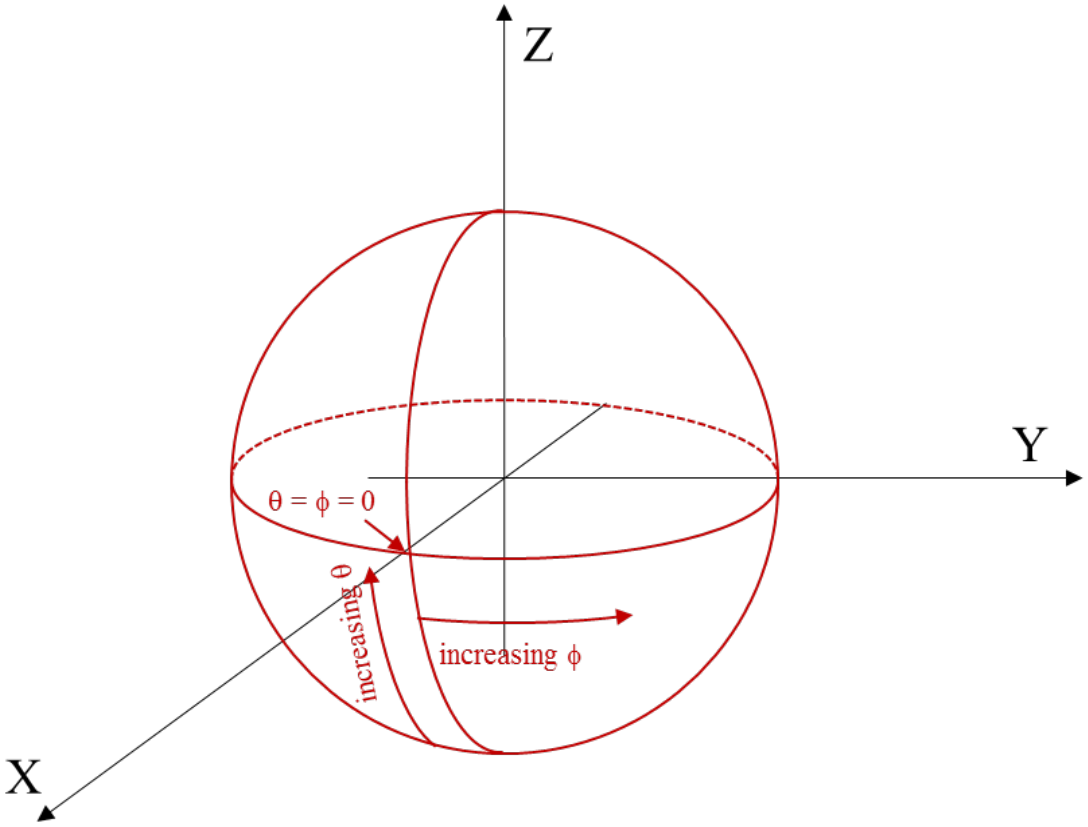
NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

OMAF – what

- Scope: 360º video, images, audio, and associated timed text, 3 DOF only
- Specifies
  - A coordinate system
    - that consists of a unit sphere and three coordinate axes, namely the x (back-to-front) axis, the y (lateral, side-to-side) axis, and the z (vertical, up) axis
  - Projection and rectangular region-wise packing methods
    - that may be used for conversion of a spherical video sequence or image into a two-dimensional rectangular video sequence or image, respectively
      - The sphere signal is the result of stitching of video signals captured by multiple cameras
      - A special case: fisheye video
  - Storage of omnidirectional media and the associated metadata using ISOBMFF
  - Encapsulation, signalling, and streaming of omnidirectional media in DASH and MMT
  - Media profiles and presentation profiles
    - that provide interoperable and conformance points for media codecs as well as media coding and encapsulation configurations that may be used for compression, streaming, and playback of the omnidirectional media content
- Provides some informative viewport-dependent 360º video processing approaches

NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

The coordinate system



Consists of a unit sphere and three coordinate axes

X: back-to-front
Y: lateral, side-to-side
Z: vertical, up

A location on the sphere: (azimuth, elevation), (ϕ, θ)

The user looks from the sphere center outward towards the inside surface of the sphere

NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

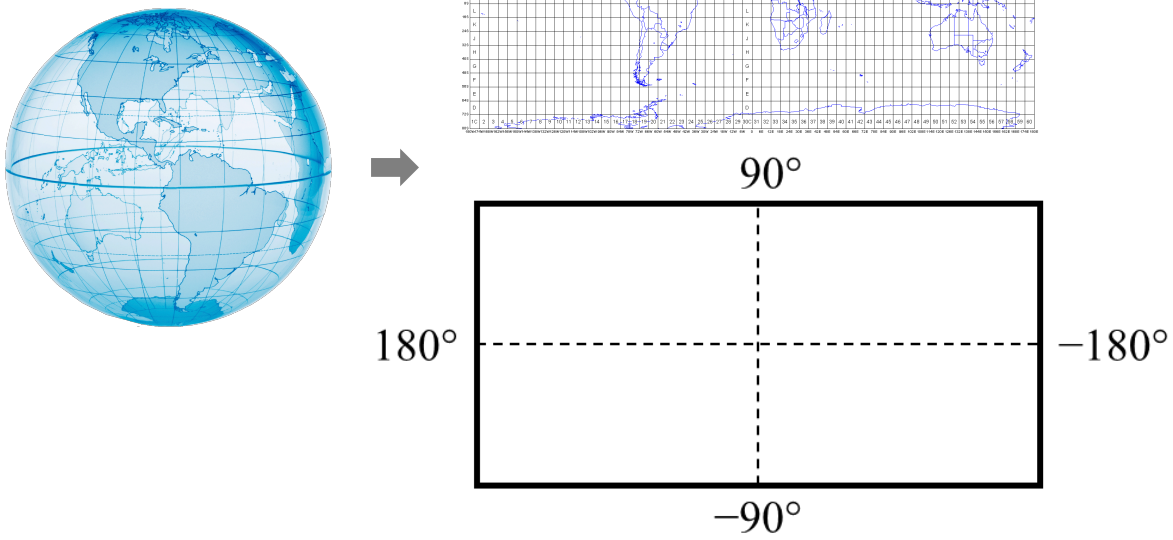
Projection

- Projection is a fundamental processing step in 360^{o} video

- OMAF supports two projection types:
  1. Equirectangular and
  2. Cubemap
     - Descriptions of more projection types can be found in JVET-H1004

NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

1. Equirectangular projection (ERP)



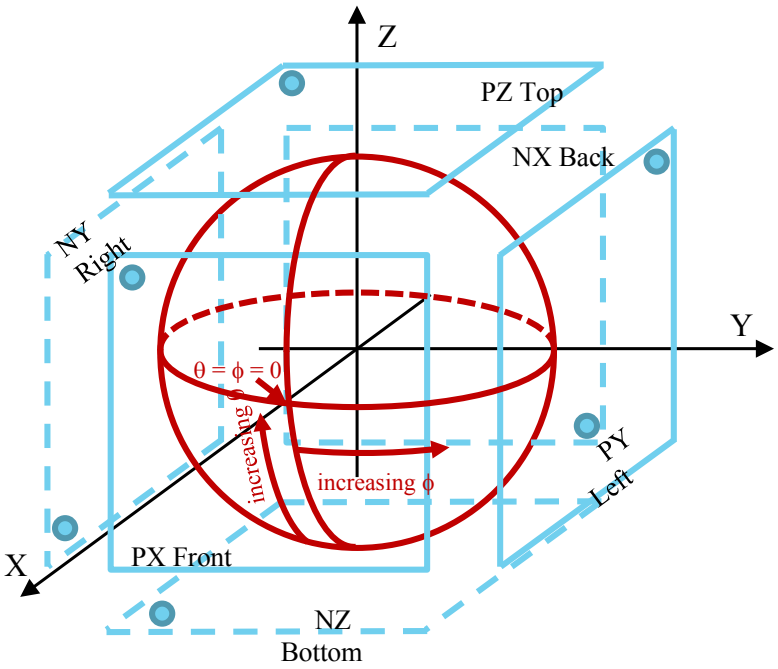
The ERP projection process is close to how a world map is generated, but with the left-hand side being the east instead of the west, as the viewing perspective is opposite.

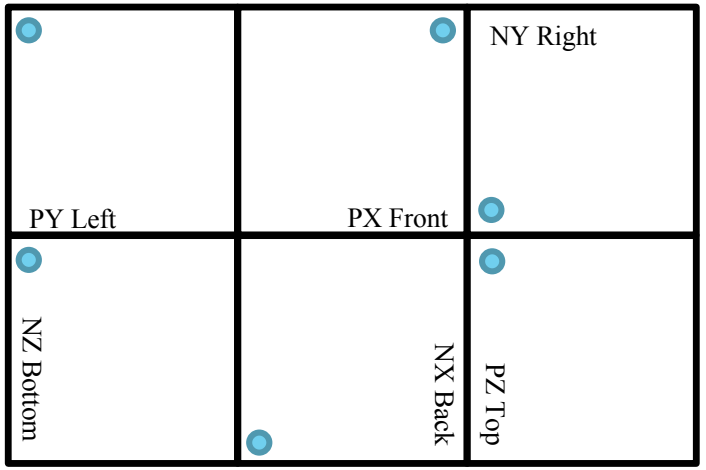
In ERP, the user looks from the sphere center outward towards the inside surface of the sphere.

While for a world map, the user looks from outside the sphere towards the outside surface of the sphere.

NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

2. Cubemap projection (CMP)





Six square faces

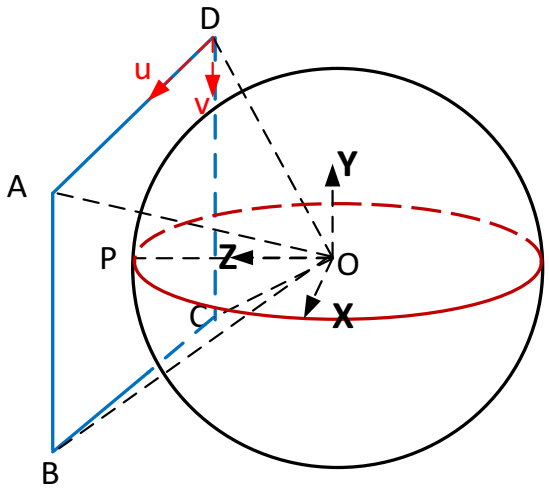
3x2 layout

Some faces rotated to maximize face edge continuity

NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

Rendering

- The rendering process typically involves generation of a viewport
  - Using the rectilinear projection



- In implementations, the viewport can also be directly generated from the decoded picture
  - Where the geometric processing steps like de-packing, inverse of projection, etc. are combined in an optimized manner

NOTE: OMAF slides taken from An Overview of Omnidirectional MediA Format (OMAF) by Ye-Kui Wang [MPEG/m41993]

3DoF+

- Problems with 360° video:
  - Objects for monoscopic 360° video have a size conflict due to lack of parallax
  - Head rotation for stereo 360° causes visual discomfort due to vertical disparities
  - Head motion is not reflected (breaks immersion)
- Benefits of 3DoF+:
  - Look around effect (more immersion)
  - 3D effect (nearby objects are rendered correctly)
  - More comfortable watching (no projection errors)
- Extra cost:
  - More cameras and a larger synthetic camera aperture
  - Higher bitrate and pixel rate for transmission
- Difference with envisioned 6DoF application: size of viewing zone
- Difference with envisioned 6DoF standard: HEVC + metadata vs. VVC amendment

Applications for 3DoF+

- Sports broadcast
- News broadcast
- Entertainment (VR movies)
- Telecommunication (video chat)
- Professional use (coaching, training)
- Education

3DoF+ timeline

• MPEG 126	WD 1	(March 2019)
• MPEG 127	WD 2	(July 2019)
• **MPEG 128**	**CD**	**(October 2019)**
• MPEG 129	DIS	(January 2020)
• **MPEG 131**	**FDIS**	**(July 2020)**

- CfP responses:
  - m47372 Nokia
  - m47179 Philips
  - m47407 PUT/ETRI
  - m47445 Technicolor/Intel
  - m47684 ZJU

CfP responses

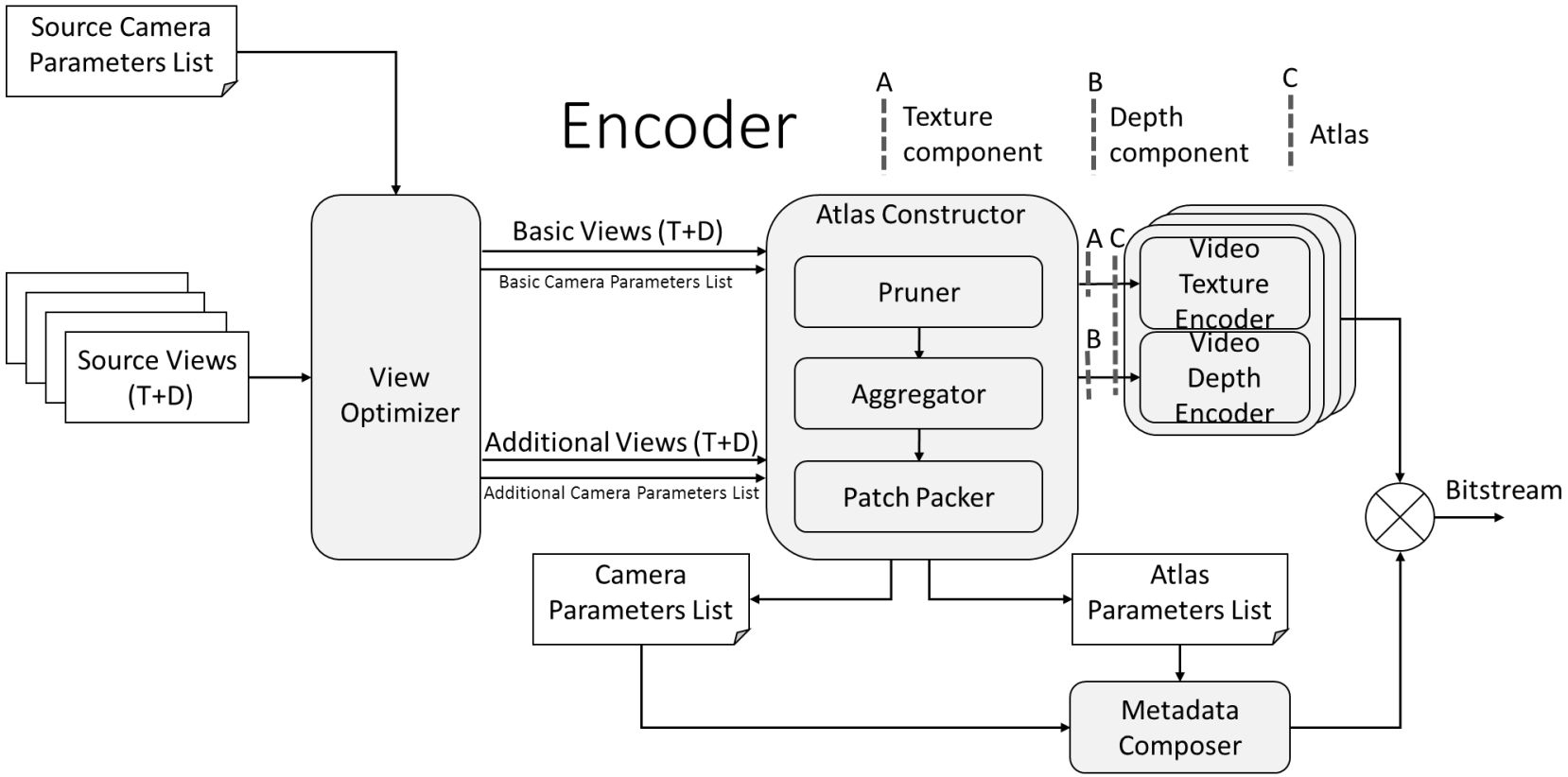
Large differences but common architecture identified

Depth/color refinement	View optimization	Prune pixels	Aggregate masks	Pack patches	Encode metadata	Encode depth	Encode occupancy	Render
Absent (3x)	**Select reference views (3x)**	Absent	Absent (2x)	Absent (2x)	**High frequency residual layer**	Same as source (3x)	Full rectangles (3x)	RVS
Depth	**Equirectangular reprojection**	**Crop views**	**OR masks per intra period (2x)**	**Largest first in scanning order**	**(Rotated) rectangles w. zlib (3x)**	**Optimized mapping (2x)**	**Pixel-based enc. in depth map**	**RVS + improvements**
Depth and color	**Map surfaces (Orthographic reprojection)**	**Point reprojection**	**Sum weights per intra period**	**MaxRect with Picture in Picture**	**Block tree w. CABAC**		**Block-based enc. in metadata**	**Internal (3x)**
		View synthesis (2x)		**Block tree transfer**	**+**			
					Camera parameters (5x)			

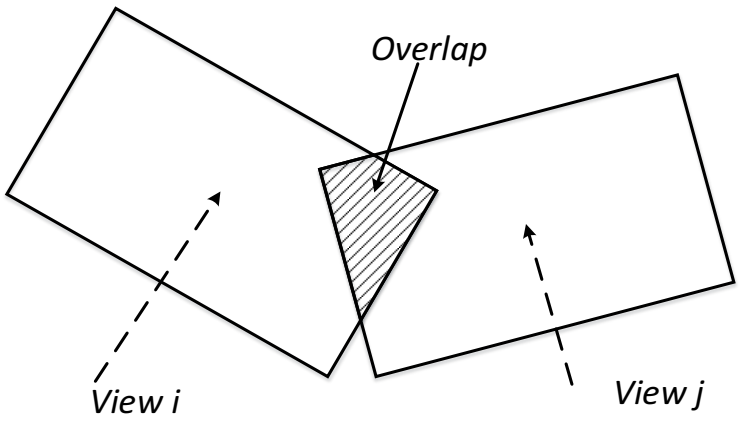
Forming a test model

- All proposals share a common architecture
- It was decided to create a single test model
- TMIV 1.0 constructed with parts from Technicolor, Philips, ZJU, Intel, PUT/ETRI

Encoder model

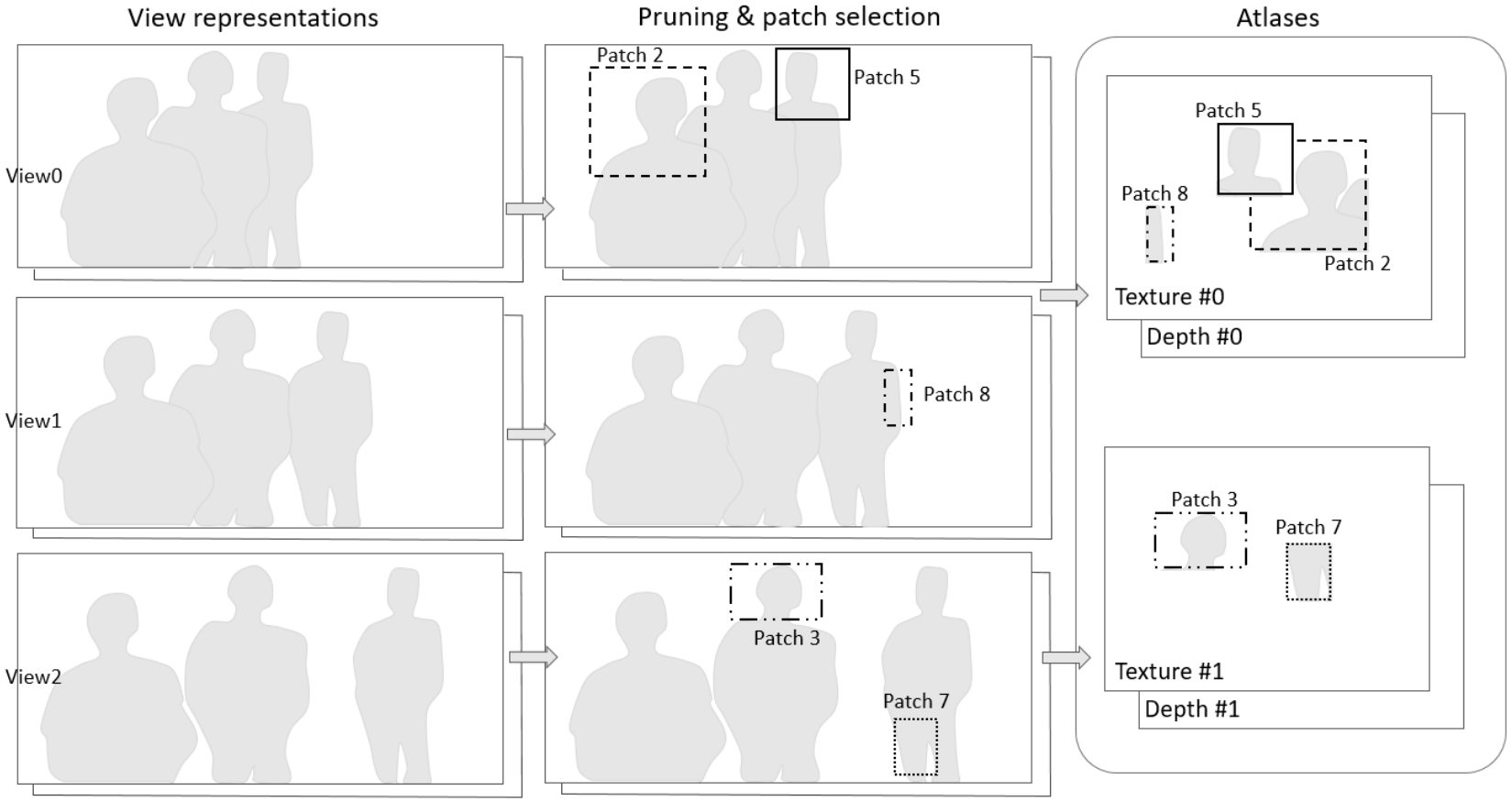


View optimization



- View optimizer:
  - Reproject to reduce pixel rate
  - Provide ***basic views*** to be fully transmitted
  - Provide ***additional views*** for extracting patches
- View reducer (TMIV 1.0):
  - No reprojection of the source views
  - Select 1 or 2 views as basic views based on overlap
  - All other source views are additional views

Encoder – Atlas Constructor



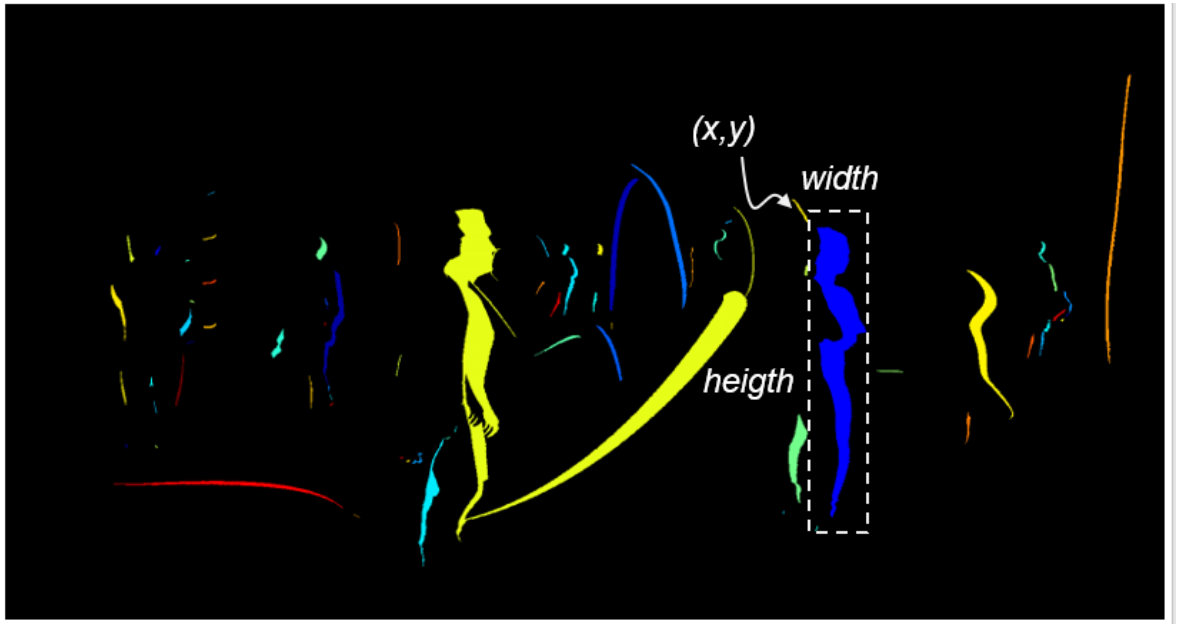
Mask aggregation

AggregatedMask @ frame i

AggregatedMask @ frame i + k

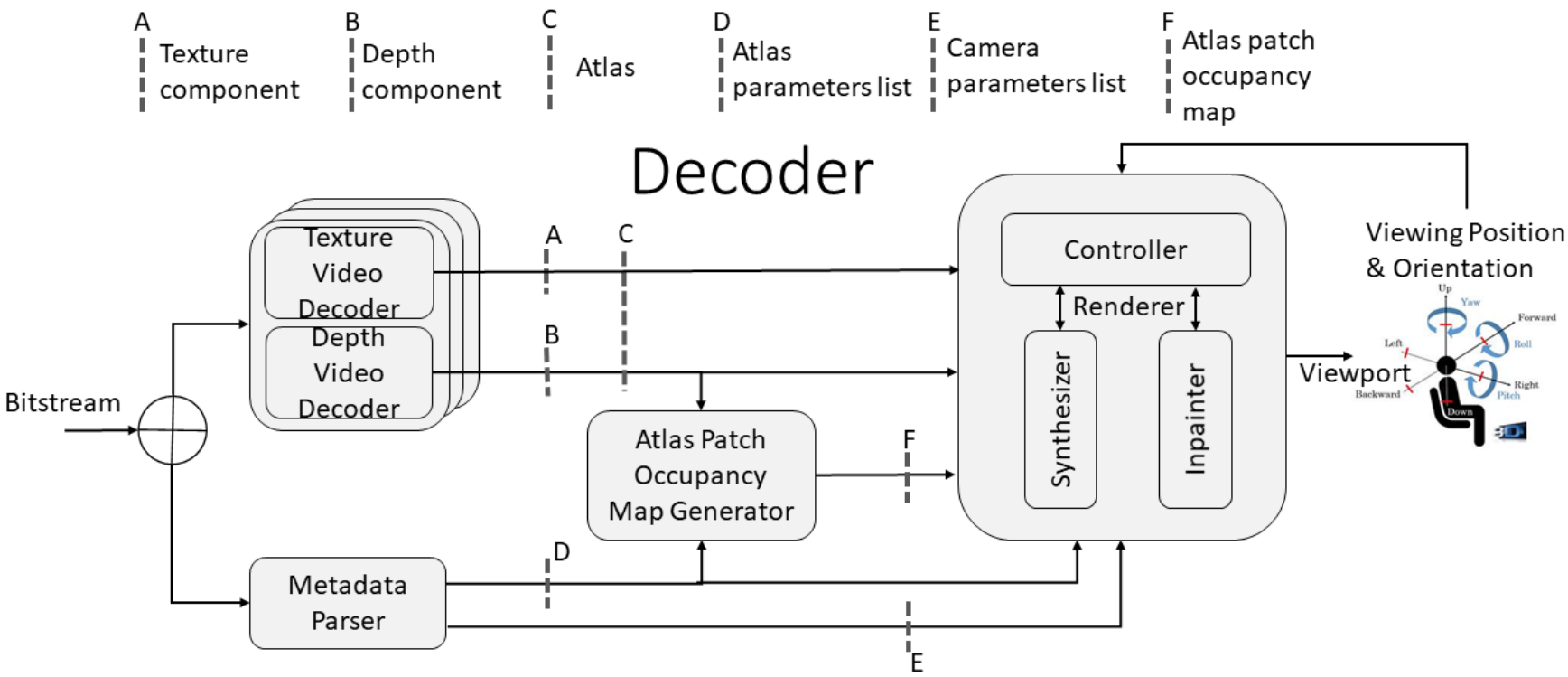
- The packing is updated only at IRAP frames.
- Mask aggregation combines the masks within an intra period to form a single mask per view.
- TMIV 1.0 uses an “OR” operation.

Patch packing

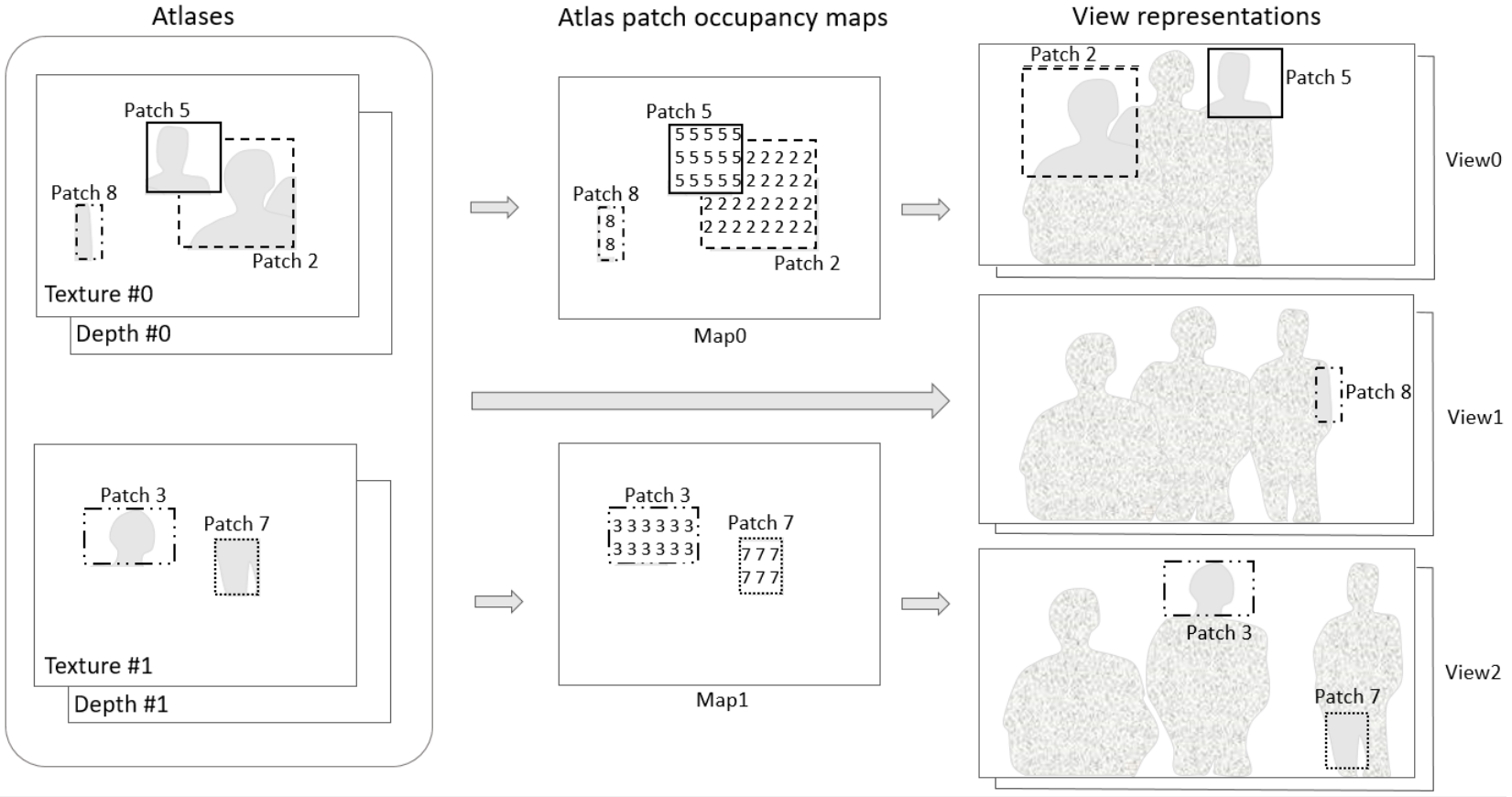


- The patch packer generates patches based on the aggregated masks, and fits them in one of the atlases.
- Patches are rectangular with occupancy signaled in the depth maps.
- Patches can be split or rotated to make them fit better.
- TMIV 1.0 uses the MaxRect algorithm with Patch-in-Patch improvement, but no direct occupancy map.

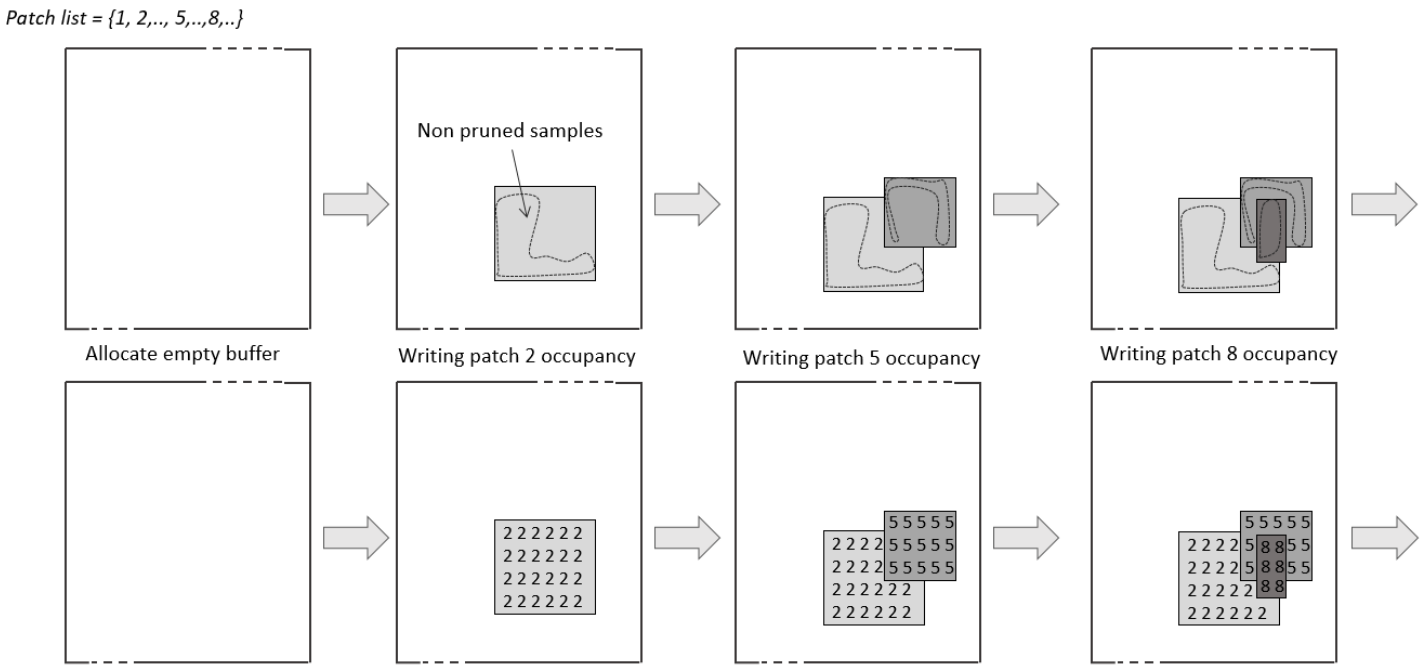
Decoder model



Decoder – Atlas Patch Occupancy Map Generator & Renderer

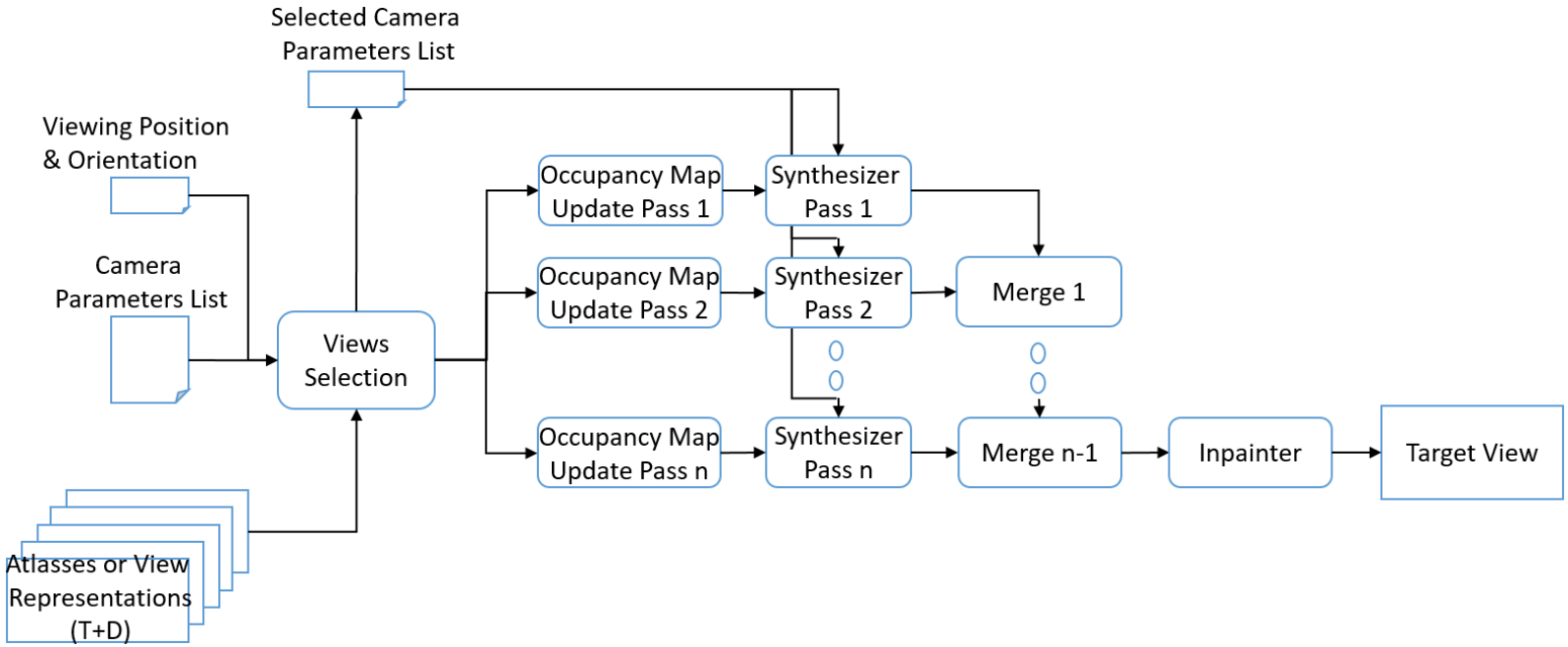


Atlas patch occupancy map generator



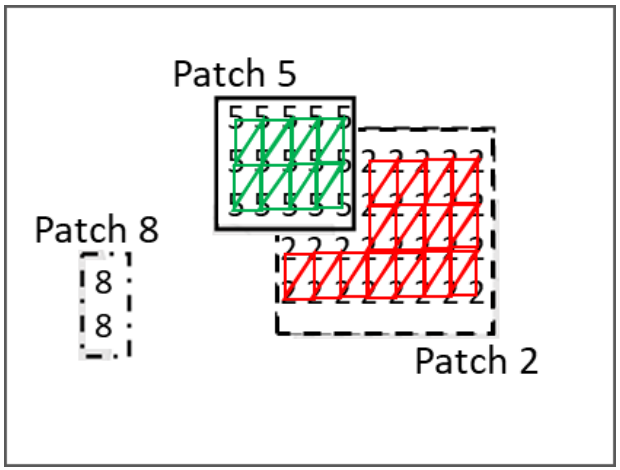
Multi pass renderer

- Give more weight to (patches from) nearby views
- TMIV 1.0 uses multi pass rendering for full views and single pass rendering for patch atlases.



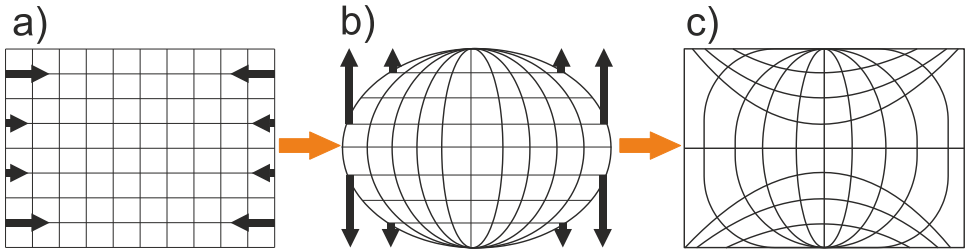
View synthesizer

- The view synthesizer and blender renders directly from the atlases using a fixed triangular mesh.
- Only when all pixels in a triangle have the same patch ID, that triangle is projected to the target view.
- Rasterization blends pixels based on:
  - Camera ray angle
  - Triangle stretching
  - Depth ordering.
- Triangles that stretch too much are not rastered.



Inpainter

- The synthesis result may have missing pixels due to viewports and disocclusions.
- The task of the inpainter is to produce a full output.
- TMIV 1.0 has a 2-way inpainter:
  - Search left & right for available pixel
  - Prefer pixel with larger depth
  - Blend when similar depth
- For ERP → perspective the nearest point is searched within a reprojected image:



Core experiments

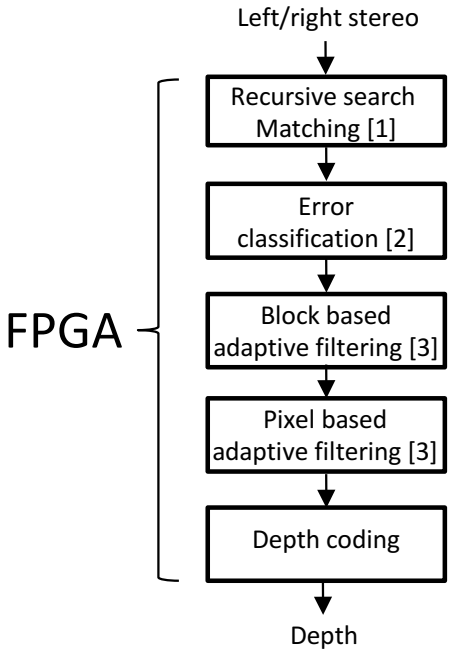
CE	Description	Intel	PUT/ETRI	Technicolor	Nokia	ZJU	Philips
CE-1	View optimization	P	P	P	O P		
CE-2	Pruning and temporal aggregation		P		O P	P	P
CE-3	Packing			O	P		P
CE-4	Rendering		P				O P
CE-5	Depth and color refinement		O P				P

O = coordinator, P = participant & cross checker

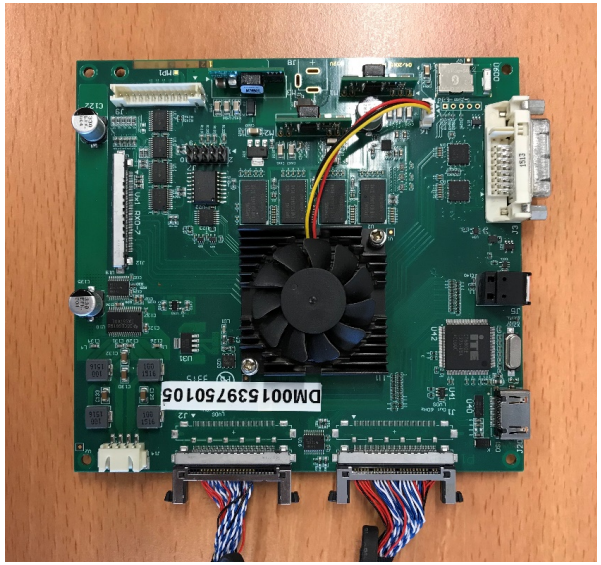
Future

- What about live transmission?
- Expensive operations are:
  - Depth estimation (and refinement)
  - Pruning
  - Video encoding
- Possible but to be demonstrated

Real-time depth estimation (1/2)

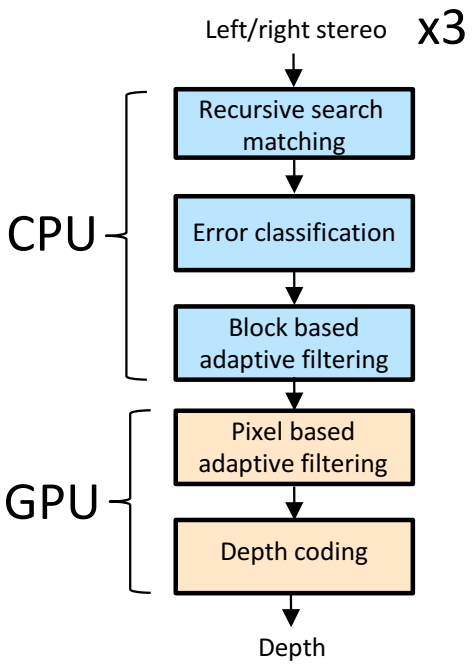


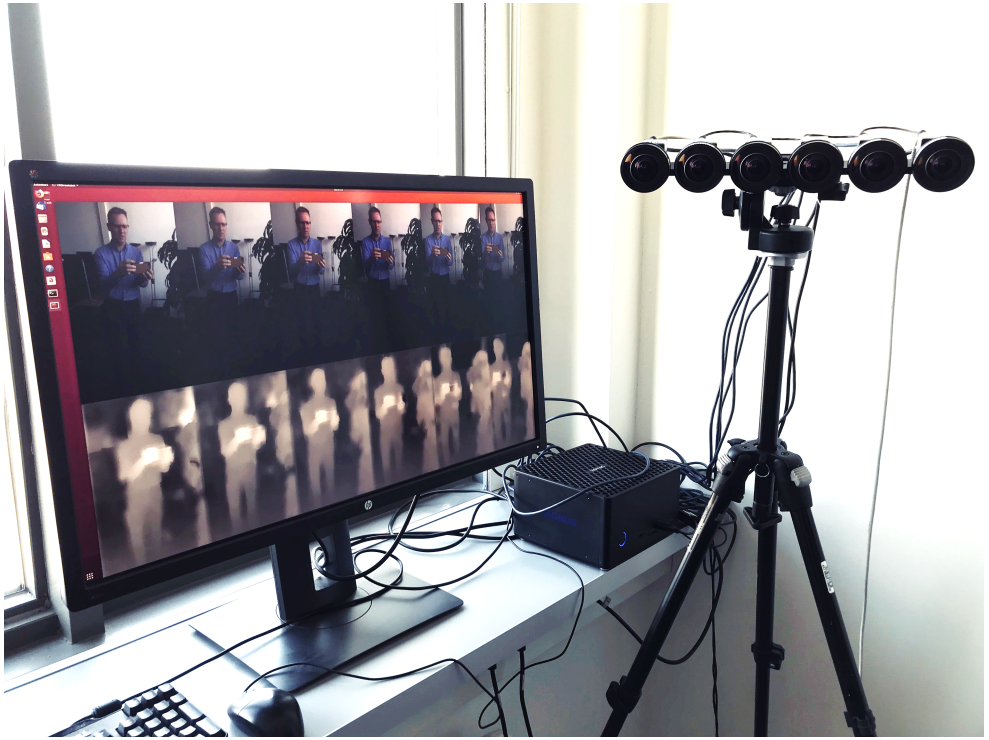
1080p @ 60Hz on TV board
FPGA: Altera Arria V device



[1] G. de Haan, et al. True-motion estimation with 3-D recursive search block matching. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 5, October 1993.
[2] C. Varekamp, et al. Detection and correction of disparity estimation errors via supervised learning. *International Conference on 3D Imaging*, 3-5 Dec. 2013.
[3] L. Vosters, et al. Overview of efficient high-quality state-of-the-art depth enhancement methods by thorough (…). *Journal of Real-Time Image Processing*, pp. 1–21, 2015.
[4] C. Varekamp, *Dynamic 6DoF VR*, AWE 2018, url: https://www.youtube.com/watch?v=Uj3B9kBqhGo

Real-time depth estimation (2/2)





Paper to be published at IBC 2019


PHILIPS