GHENT UNIVERSITY

imec

Immersive Video

IDLab – MEDIA

HTTP://MEDIA.IDLAB.UGENT.BE

Glenn Van Wallendael | glenn.vanwallendael@ugent.be

Bert Ramlot | bert.ramlot@ugent.be

FACULTY OF ENGINEERING AND ARCHITECTURE

IDLab INTERNET & DATA LAB

Challenges

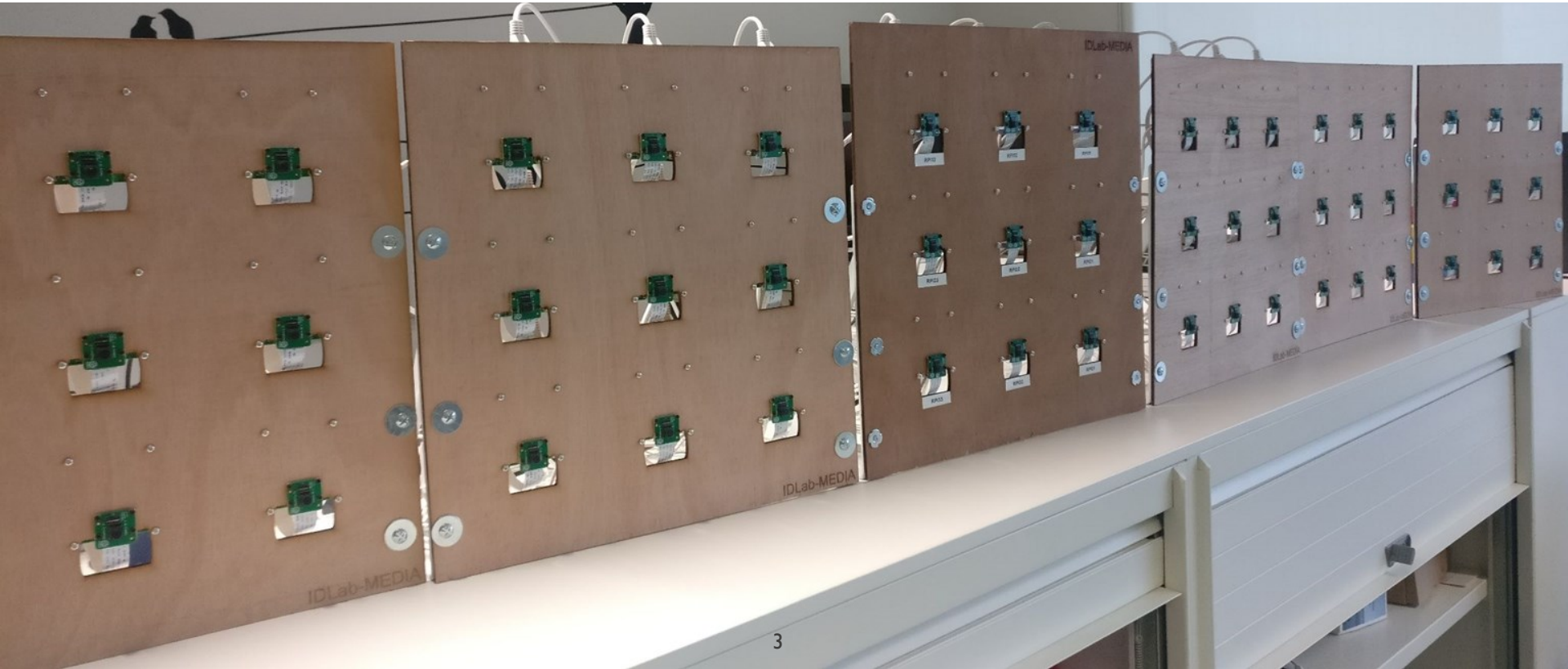
Direction-dependent and complex light phenomena

- refractions
- smoke
- fire
- complex reflections

Challenges

Camera captured

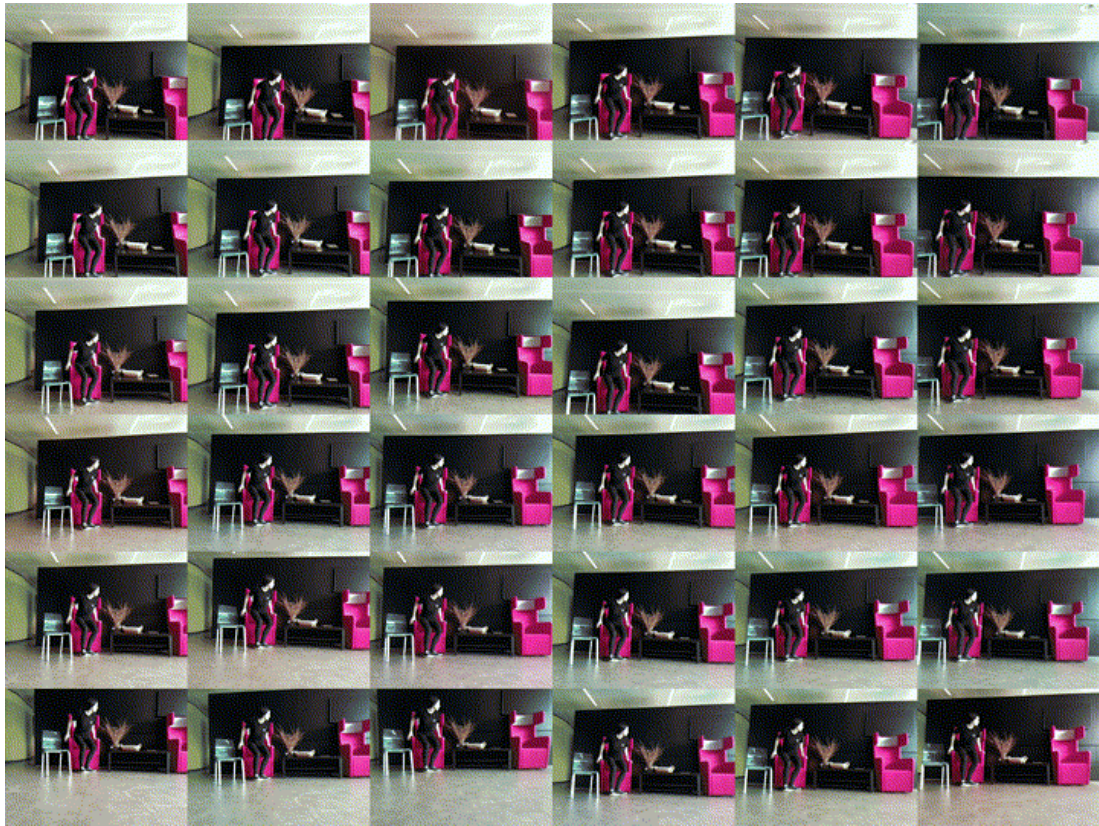
- accurate depth information
- camera calibration
- color matching



Challenges

Lack of dynamic datasets

We currently have a 6x9 PiCameras to shoot videos synchronously.
https://media.idlab.ugent.be/multi-camera-setup



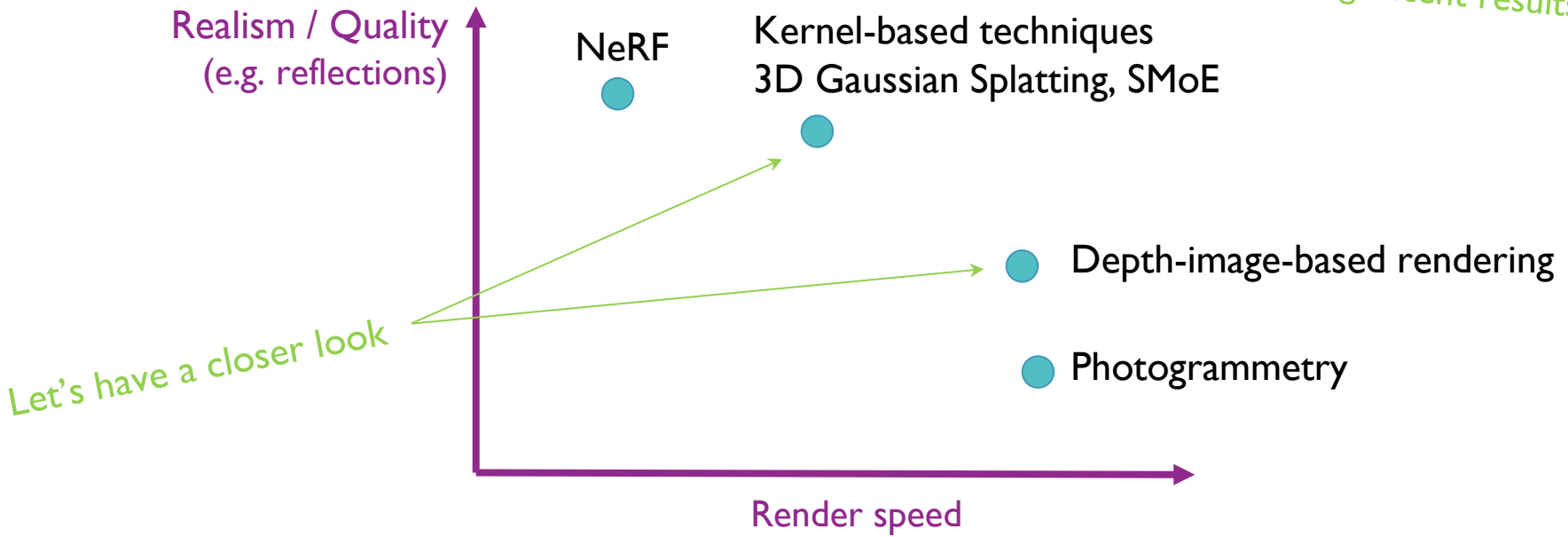
Do you have video captures from different viewpoints of a scene? Great, let us know!

GHENT UNIVERSITY

Immersive technologies

What technology should I use to immersive myself in a life-like environment?

Good news: for *static* environments (no moving objects), the research has already produced magnificent results!



Gaussian Splatting

What is a "**splat**"?

Attribute class	# attributes
Position	3
Scale	3
Rotation	4
Opacity	1
Color	48
Total	59


Remote Viewer Settings (Point view)
Show Input Points
Show Input Points during Motion
Train
SHs Python
Rot-Scale Python
Keep model alive (after training)
1.000
Scaling Mod
Metrics
Camera Point view
FPS
Mode
Load camera
Save camera (bin)
Snap to closest
0
Snap to
50.154263
Fov Y
0.009000
Near
1100.000000
Far
Key cameras: 0
Add key
Save key cameras...
Play
Play (No Interp)
Record
Stop
1.000000
Speed
Load path
Save path
Save video (from playing)
Save frames (from playing)
Acceleration
0.300000
Speed
1.000000
Rot. speed

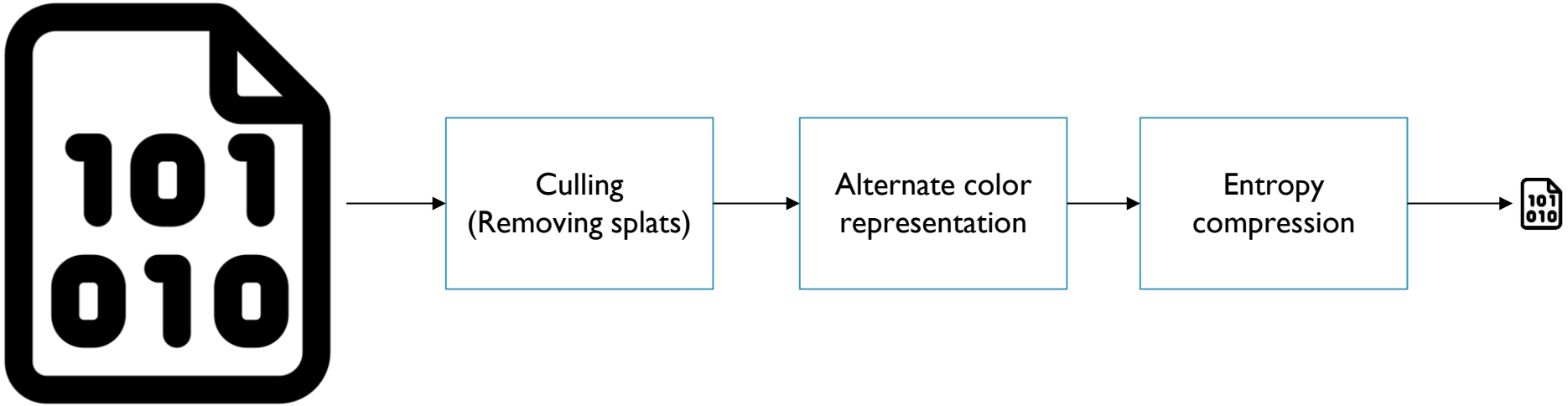
Gaussian Splatting

Active development/research areas.

- Practical
  - Game engine integration
  - Editing
- Research topics
  - Large-Scale
  - Sparse
  - Object detection
  - Dynamic splats
  - Mesh extraction
  - Compression

Gaussian Splatting

Compression: Pipeline



Gaussian Splatting

Compression: Culling

(a) 0% removed (PSNR=26.66)

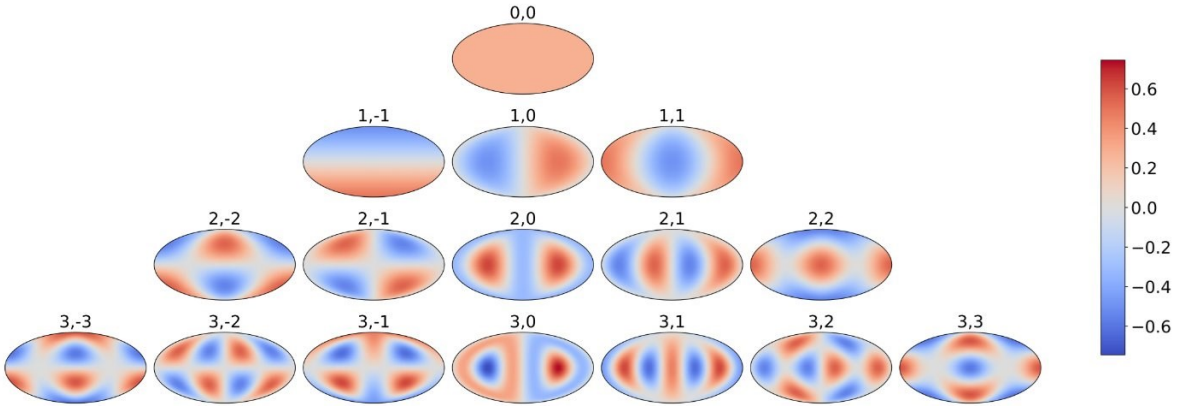
(c) 90% removed (PSNR=24.58)

Gaussian Splatting

Compression: Spherical harmonics energy compaction

- A splat's color is dependent on the angle it is looked at ⇒ spherical function
- A spherical function can be decomposed into spherical harmonics (c.f. Fourier)
- 48 lighting coefficients per splat (!)

Idea: Find a 'cheaper' set of lighting coefficients.

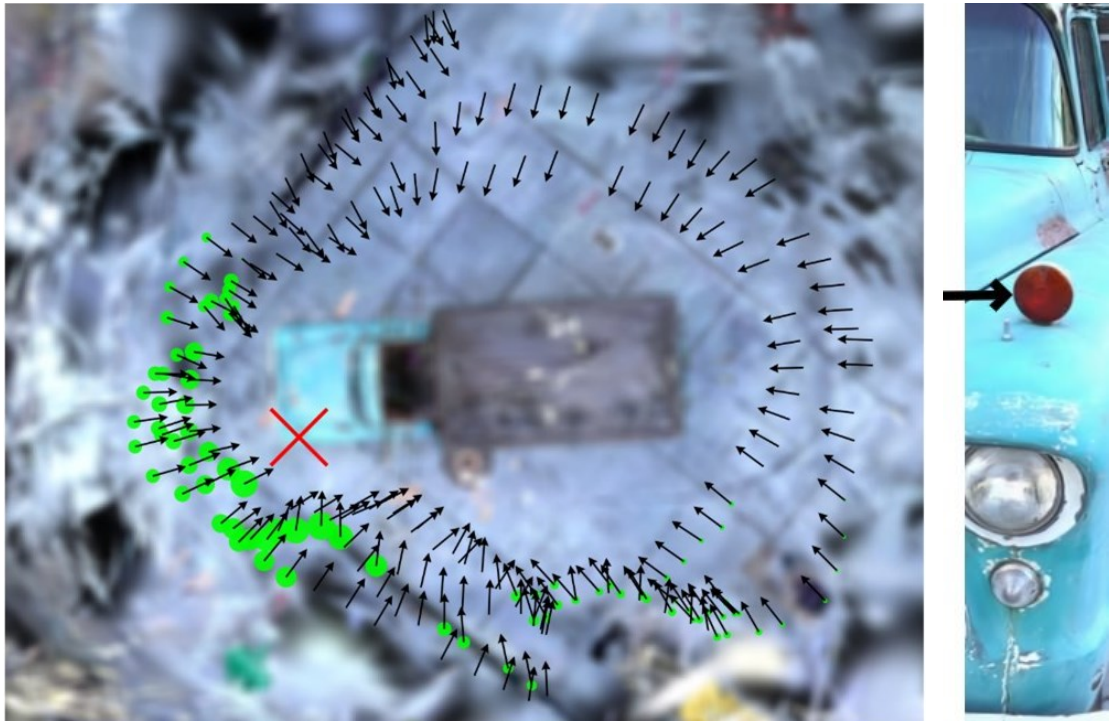


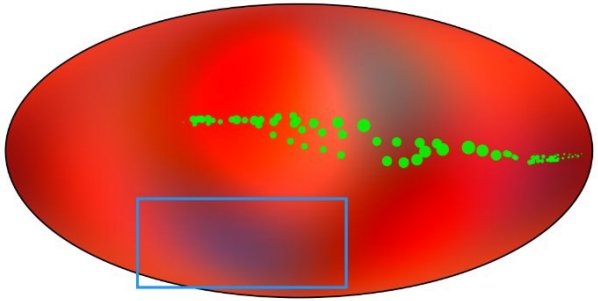
GHENT UNIVERSITY

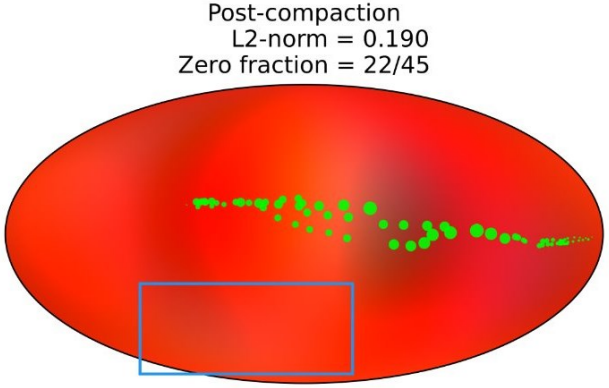
imec

Gaussian Splatting

Compression: Spherical harmonics energy compaction

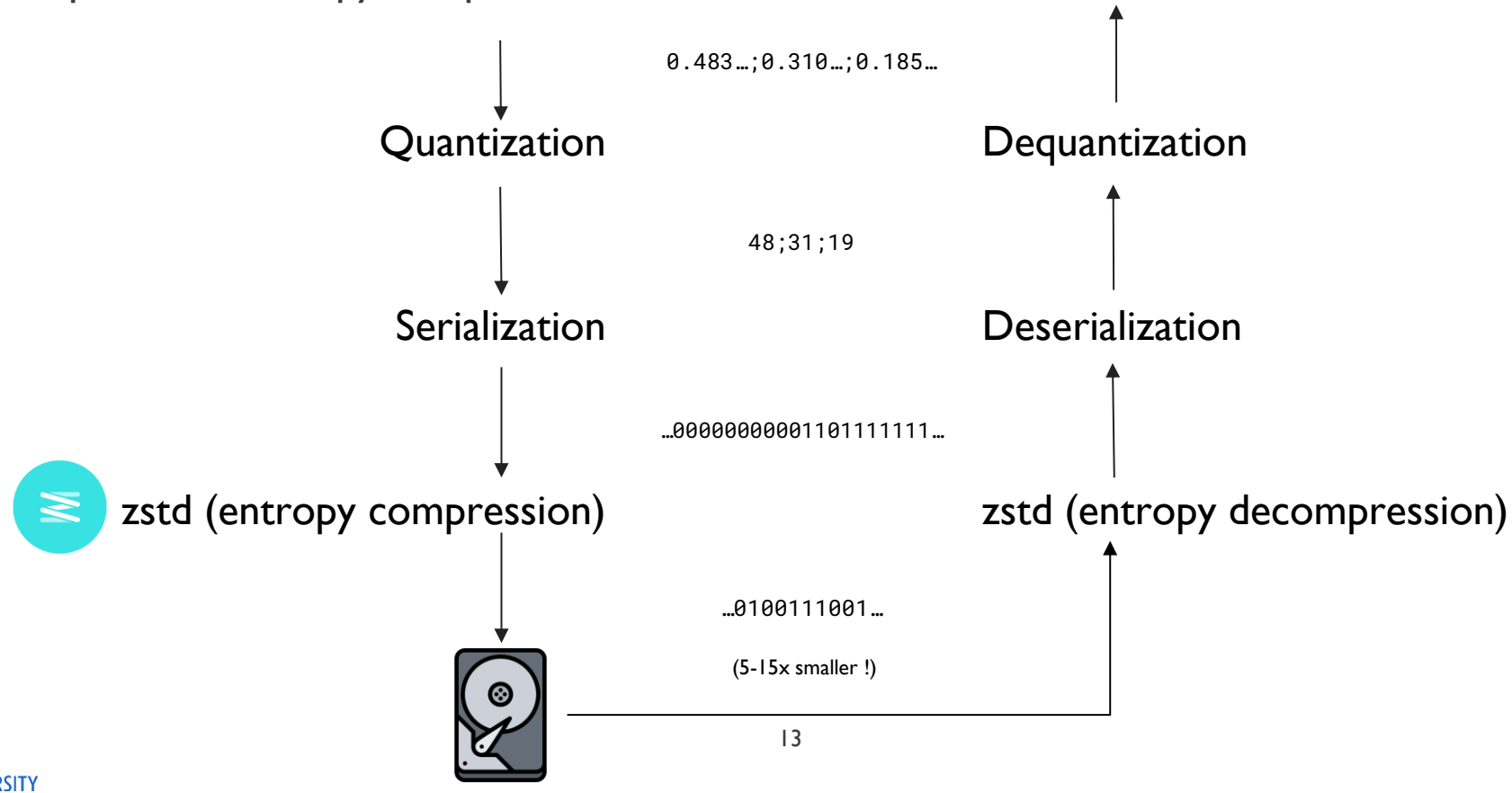






Gaussian Splatting

Compression: Entropy compression



GHENT UNIVERSITY

imec

Gaussian Splatting

Compression: results

BEFORE compression

PSNR = 24.94
Size = 630 MB
FPS = ~100

imec

Gaussian Splatting

Compression: results

AFTER compression

PSNR = 24.76
Size = 10 MB
FPS = ~200

General compression effects

+ 35-100x smaller
+ 1.5-3x faster rendering
- Slightly lower quality

imec

Steered Mixture of Experts (SMoE)

Our kernel-based technique

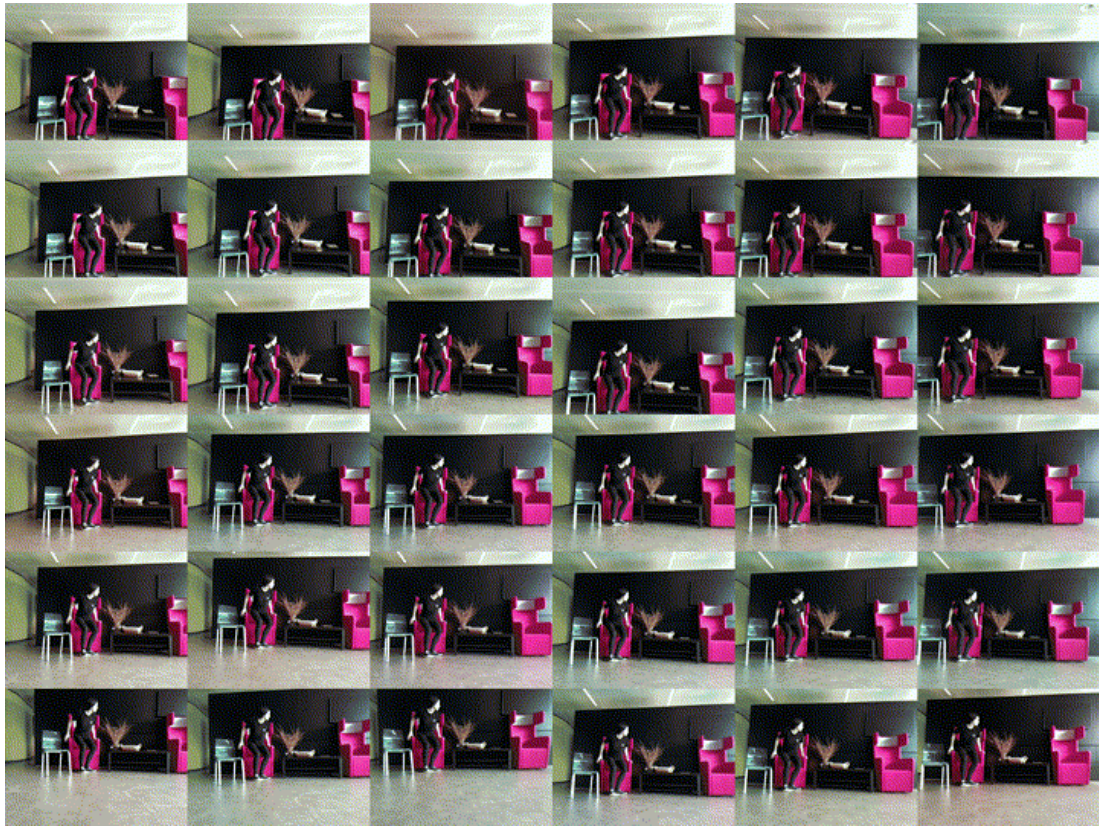
Light Field

Pixel position:

x,y

Camera position:

s, t



Demo Steered Mixture of Experts (SMoE)

It is currently a lot harder to build applications for *dynamic* content.

You will have to sacrifice speed or visual quality. For example:

- 4D Gaussian Splatting (= dynamic): RTX 3090 GPU, 800x800: 82 FPS
- dynamic NeRF: not well established, likely not viable on current hardware architectures

What does state-of-the-art real-time dynamic view synthesis look like?

→ Depth-Image-Based rendering (DIBR)

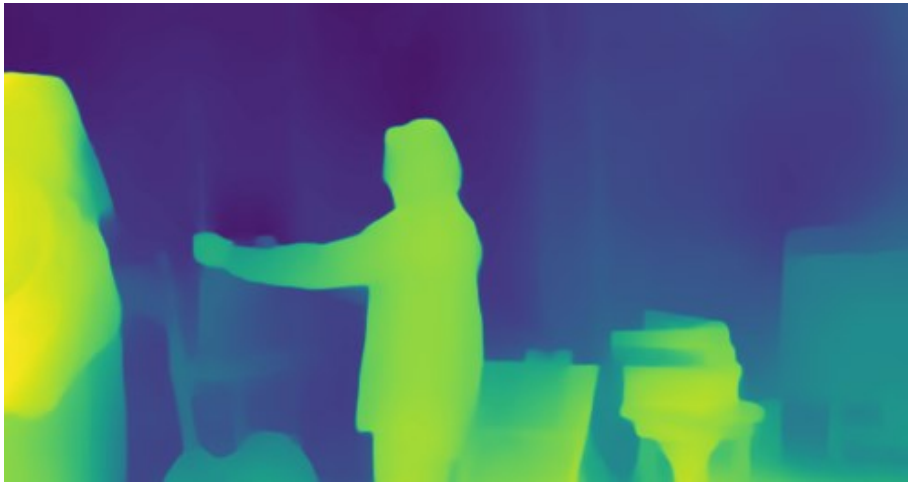
Depth Image Based Rendering (DIBR)

Film the scene from multiple viewpoints

Estimate the depth for each frame of each view

Depth estimation

Monocular vs Multi-view stereo



Monocular

+ fast
+ handles low-texture areas well
+ single camera
- low resolution
- depth is a guess!

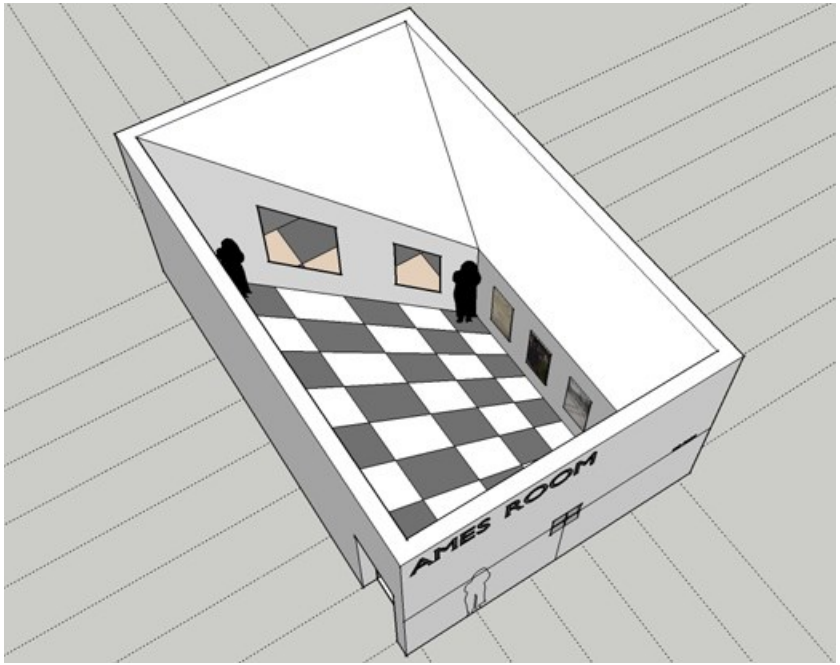
Multi-view stereo

+ fast
- doesn't handle low-texture areas well
- multiple cameras
+ high resolution
+ depth is correct

We're currently combining both techniques to "get the best of both worlds"

Depth estimation

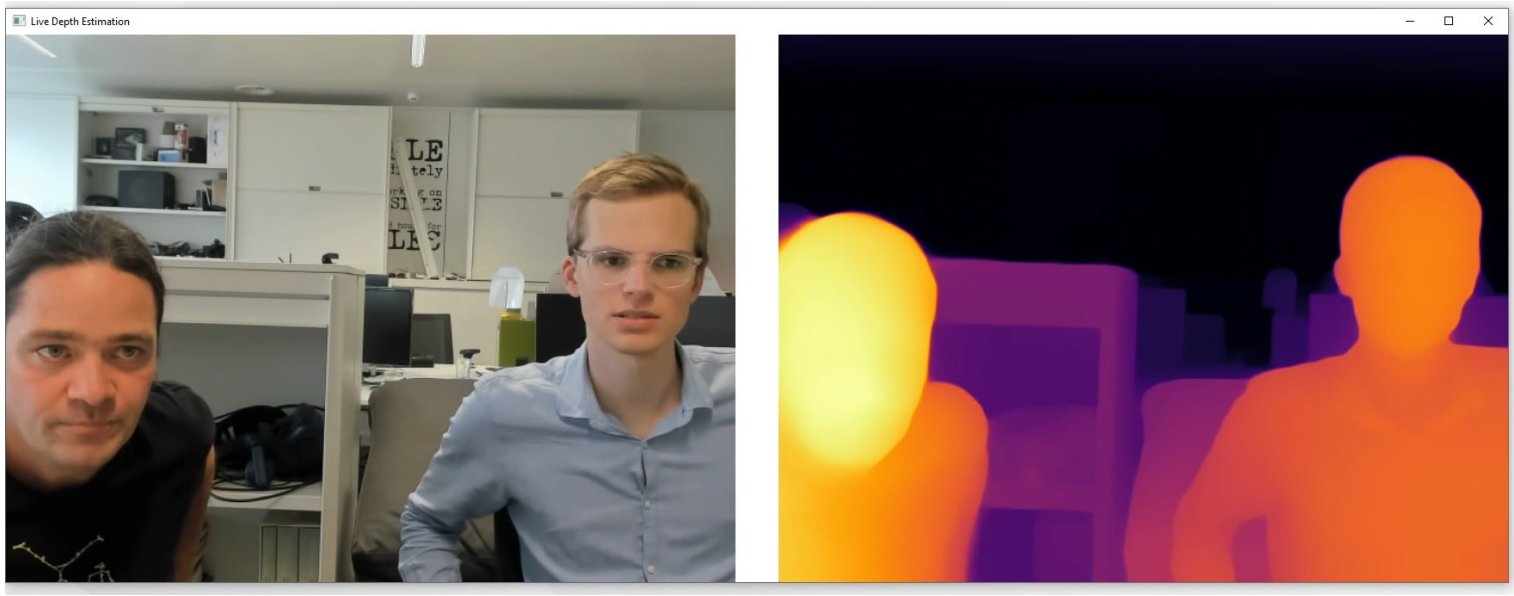
Problem with Monocular depth estimation: Ames room illusion



Monocular depth estimation

State-of-the-art

Monocular depth estimation demo



Depth Image Based Rendering (DIBR)

The more correct the estimated depth maps, the closer to “real” the end result.

Noisy depth, some wrong spots

synthetic scene ⇒ perfect depth ⇒ looks great

Dynamic view synthesis: DIBR

Triangle meshes of 16 views blended together:

VR:
90Hz refresh rate for
2448x2448 (per-eye)

with some optimizations:

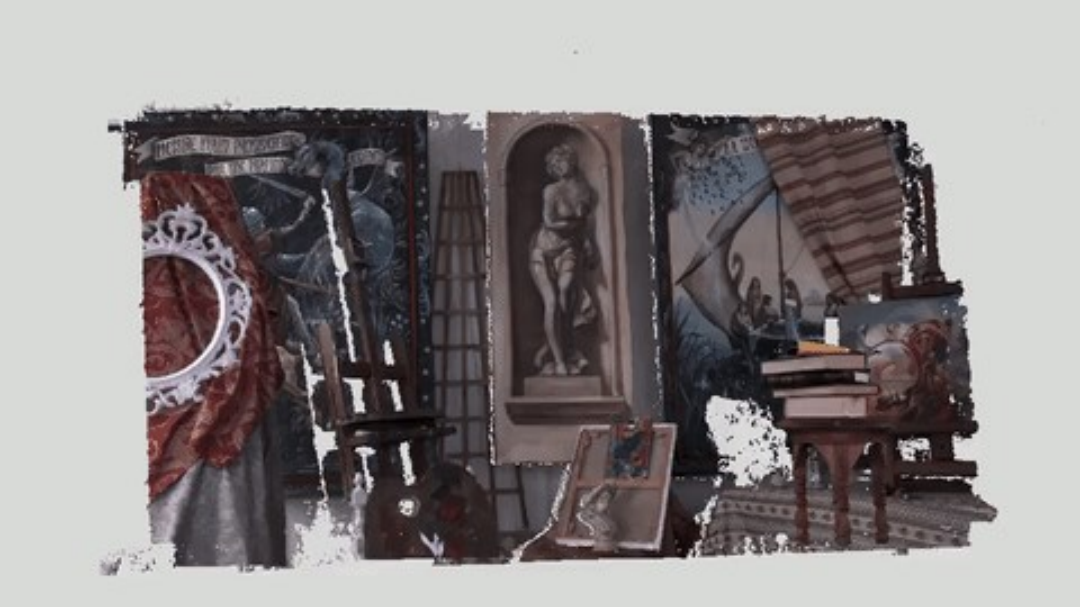
- NVIDIA hardware accelerated video decoding (for color and depth videos)
- OpenGL for mesh rendering
- Only render the 4 closest views

https://media.idlab.ugent.be/opendibr

GHENT UNIVERSITY

Immersive Video

imec



HTTP://MEDIA.IDLAB.UGENT.BE

Glenn Van Wallendael | glenn.vanwallendael@ugent.be
Bert Ramlot | bert.ramlot@ugent.be

FACULTY OF ENGINEERING AND ARCHITECTURE

IDLab
INTERNET & DATA LAB