Graphics and Imaging Lab

Adobe

IEEE VR 2019 OSAKA

Journal Paper

# Motion parallax for 360º RGBD video

Ana Serrano[1] Incheol Kim[1] Stephen DiVerdi[2] Diego Gutierrez[1]
Aaron Hertzmann[2] Belen Masia[1]

[1]Universidad de Zaragoza [2]Adobe Research

# Introduction

IEEE VR 2019 OSAKA


*Miyubi – Felix & Paul Studios*

*SuperHOT VR*

Recorded with a fixed camera
3-DoF (only rotation)

CG content
6-DoF (rotation and translation)

# Introduction

IEEE VR 2019
OSAKA


*Miyubi – Felix & Paul Studios*


*SuperHOT VR*

Recorded with a fixed camera
3-DoF (only rotation)

CG content
6-DoF (rotation and translation)

# Introduction



Videos recorded from fixed camera position

→ How to render the scene from different head positions?


Scene recorded from a fixed camera position

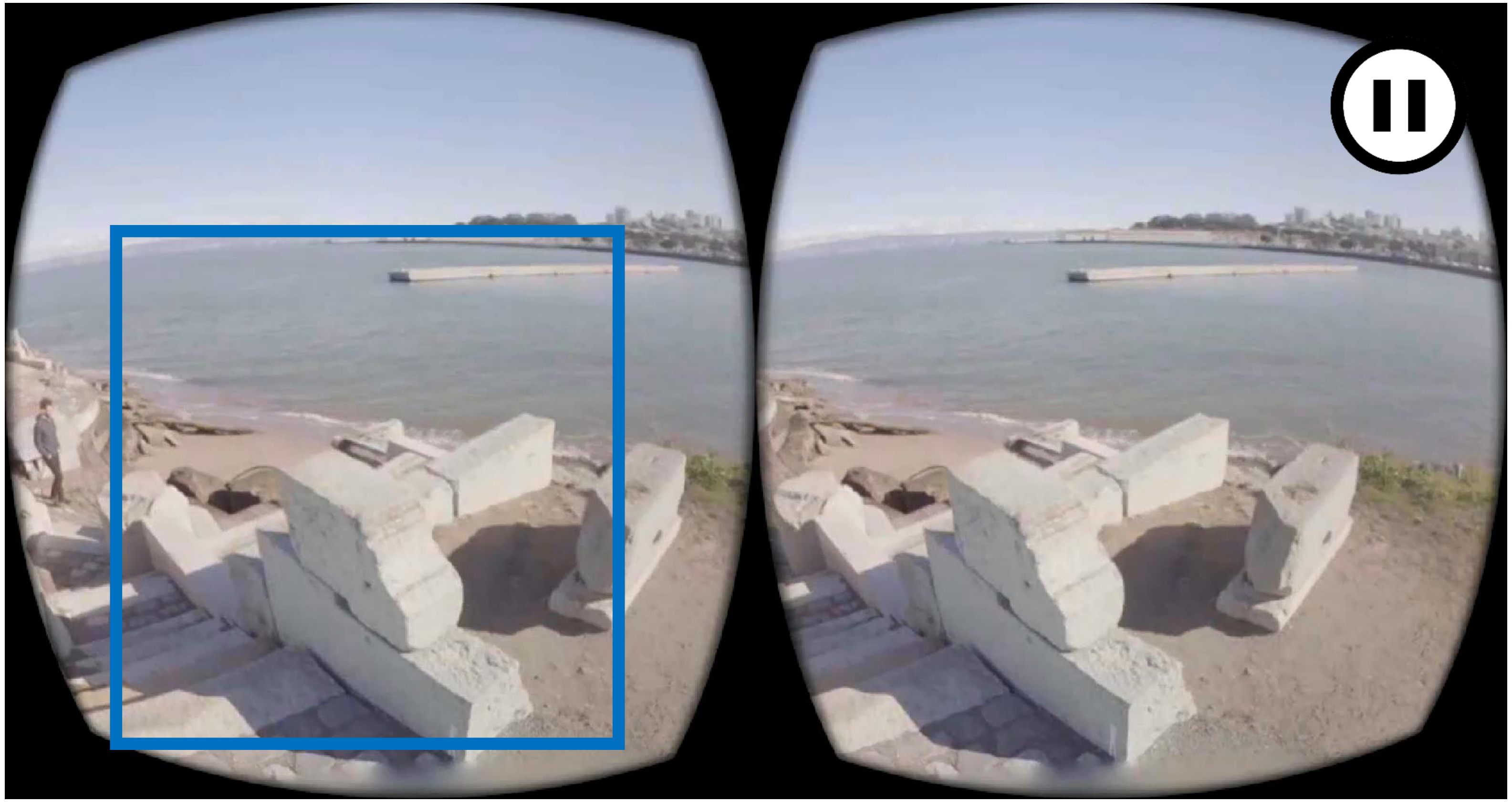

New camera view to show to the user

# 6-DoF for 360 video



Close-up

VR view (stereo)



# Related works



## Image-based rendering

- Methods that use **implicit geometry** (image correspondences)
  [Lipski2010], [Mahajan 2009], [Stich 2011], [Huang 2017]…

- Methods that use **explicit geometry** (depth maps or other geometry proxy)
  [Debevec 1998], [Chaurasia 2013], [Eiseman 2008]…

# Related works



## Image-based rendering

- Methods that use **implicit geometry** (image correspondences)
  [Lipski2010], [Mahajan 2009], [Stich 2011], [Huang 2017]...

- Methods that use **explicit geometry** (depth maps or other geometry proxy)
  [Debevec 1998], [Chaurasia 2013], [Eiseman 2008]...

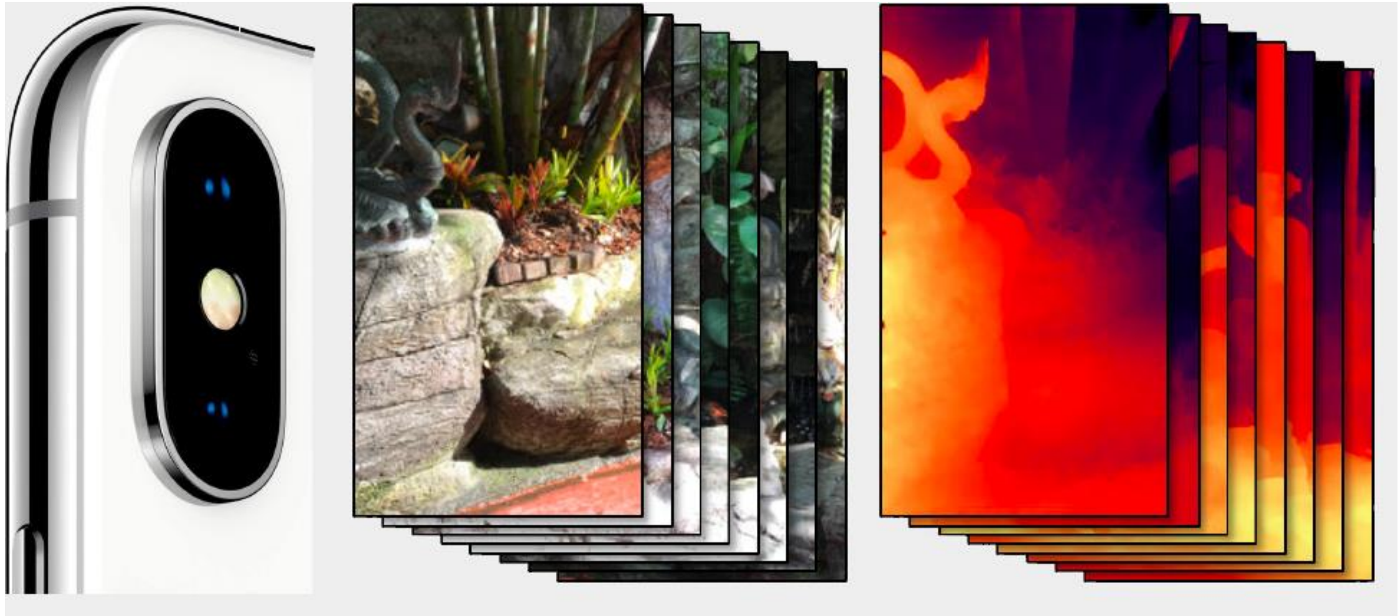
In contrast with these works:

- Our starting point is just a RGBD panorama (with a very narrow baseline)
- We generate novel unseen viewpoints (rather than interpolating between existing ones)

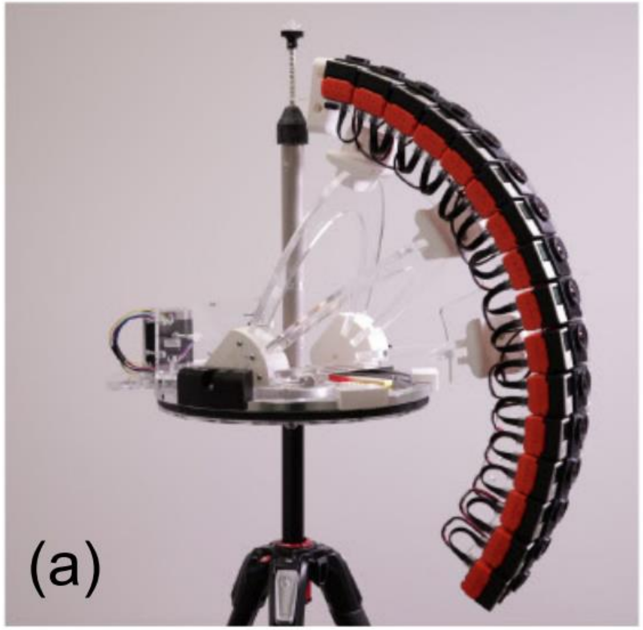
# Related works



[Hedman and Kopf 2018] Instant 3D Photography



[Overbeck et al. 2018] Welcome to lightfields




High-fidelity static 3D scenes → not suitable for dynamic scenes nor video

# Our approach

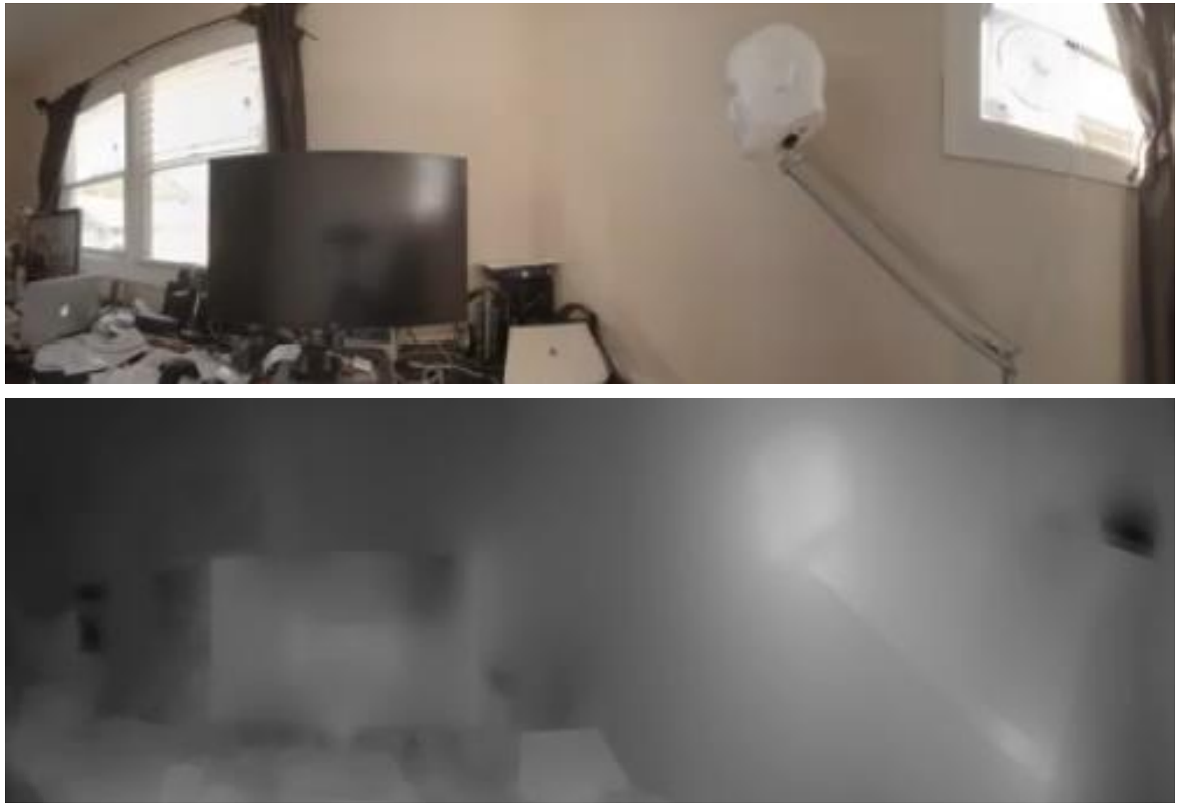
Commercially available cameras

→ Most of them with a narrow baseline


# Our approach

IEEE VR 2019 OSAKA

Our input: RGBD video panoramas



Yi Halo (Google Jump Manager)

Facebook x24 (Facebook Surround 360)

# Our approach

IEEE VR 2019 OSAKA

Input : RGBD 360 video

Output: Novel views from different camera positions

(1) Layered video representation
(2) Depth improvement optimization

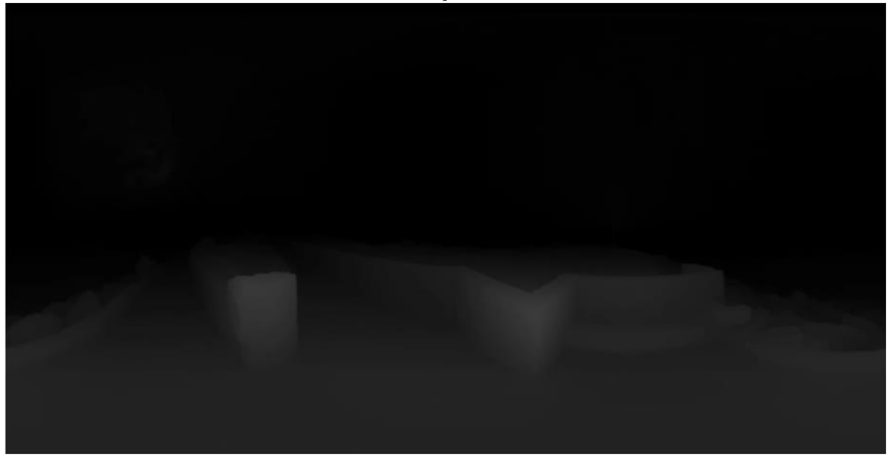
# Layered video representation

# Mesh-based depth reprojection

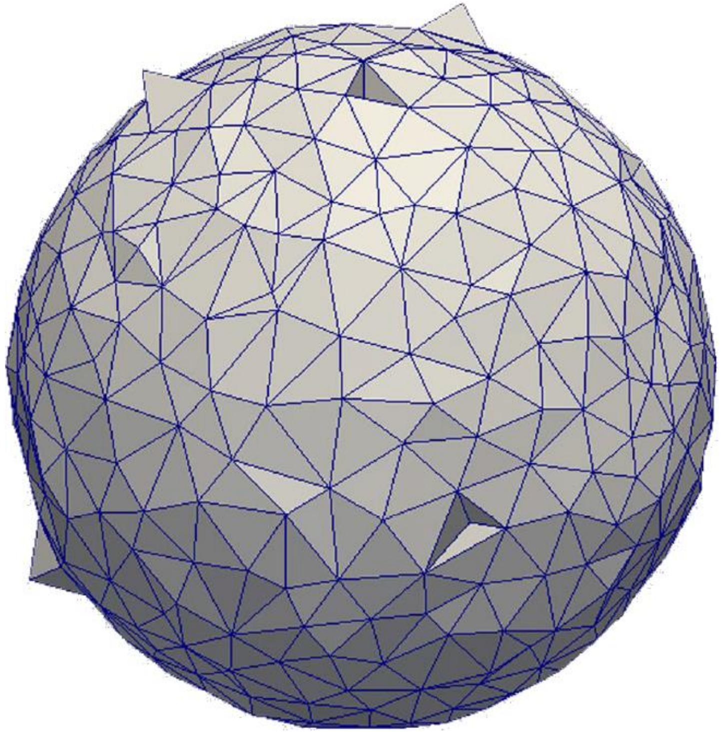
IEEE VR 2019 OSAKA

RGB

Depth



Depth-distorted mesh



# Mesh-based depth reprojection

IEEE VR 2019 OSAKA

HMD original view

Displaced view

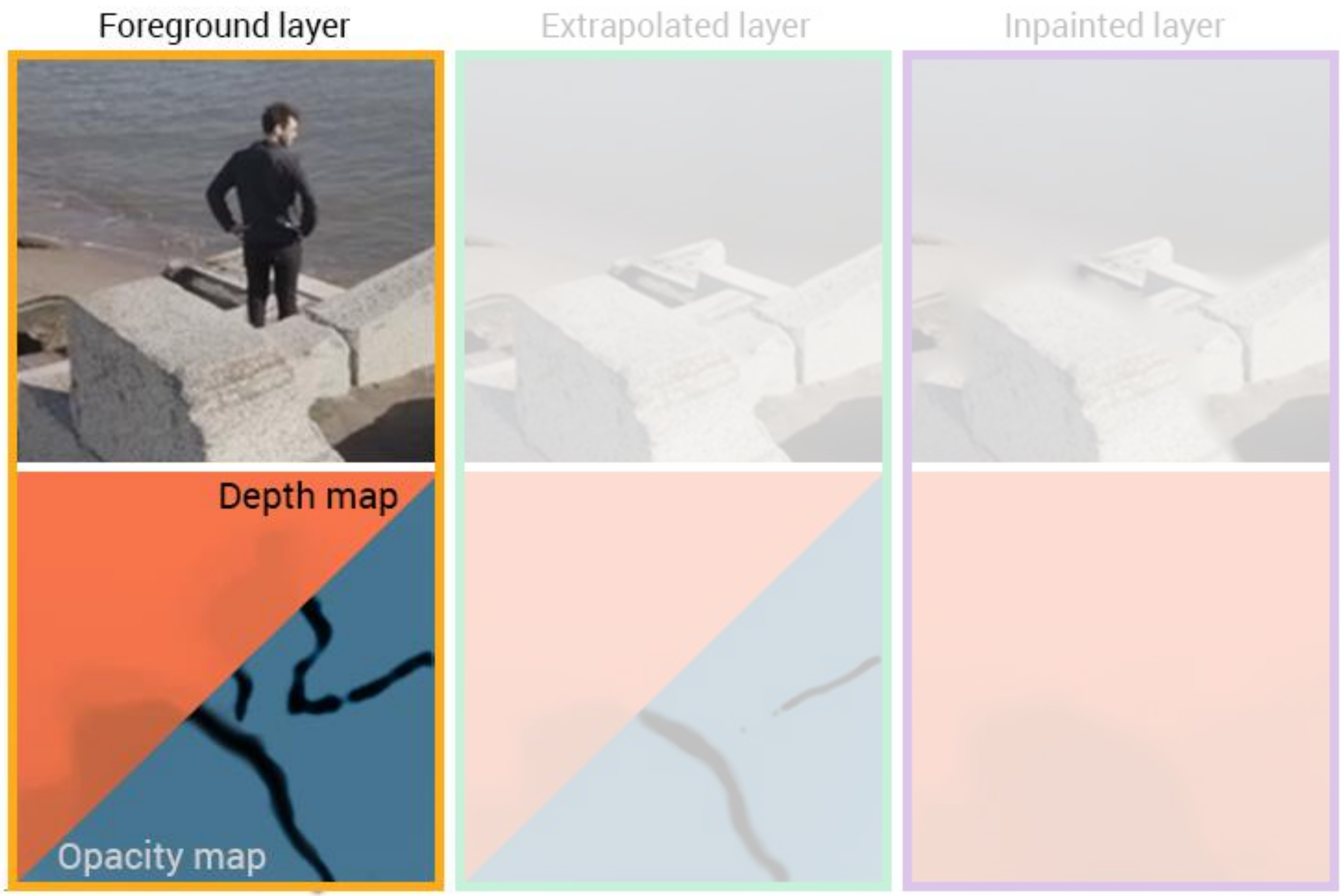
# Layered video representation

IEEE VR 2019 OSAKA

Mesh-based approach

Three layers

- Foreground layer
- Extrapolated layer
- Inpainted layer



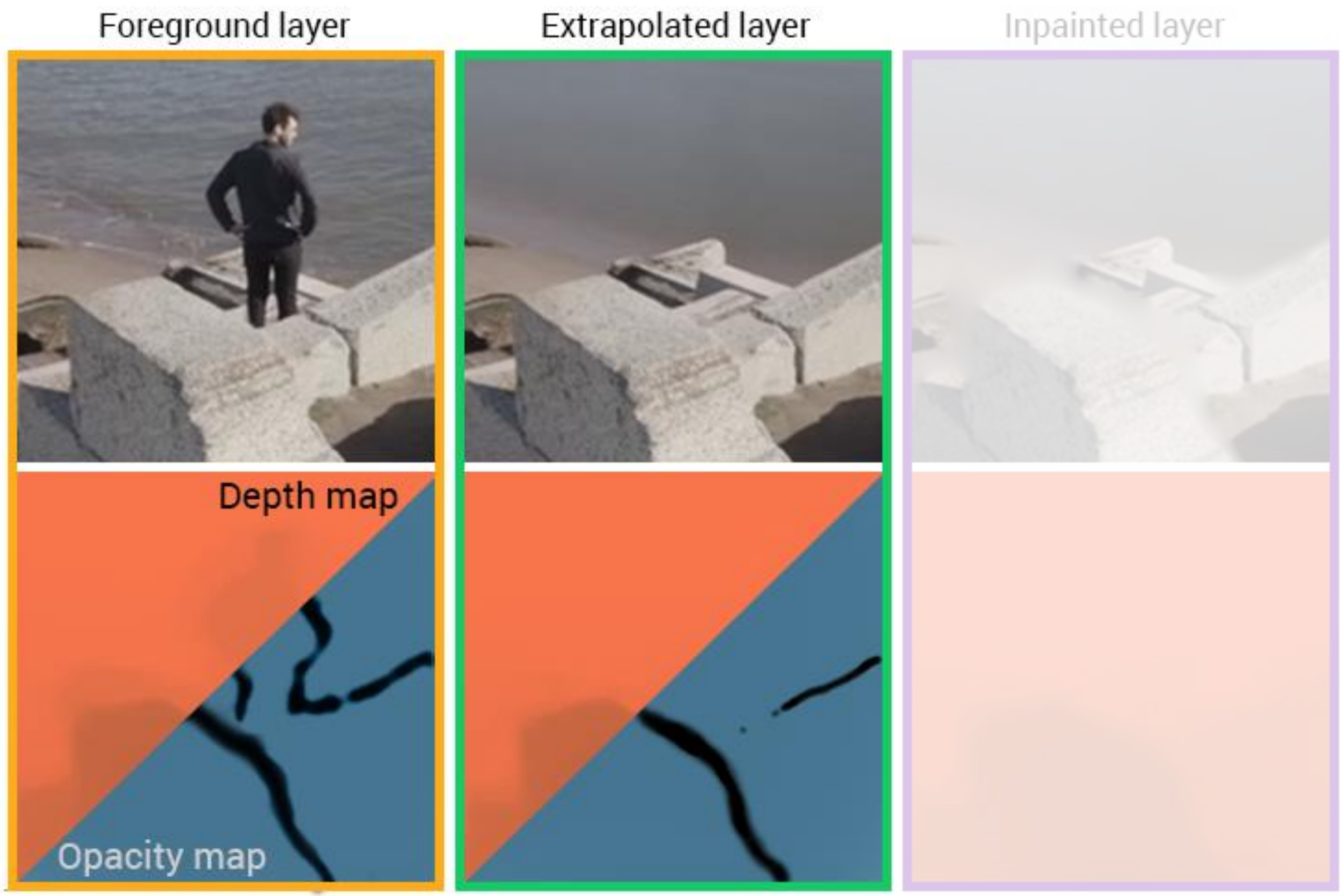
# Layered video representation



Mesh-based approach

**Three layers**

- Foreground layer
- Extrapolated layer
- Inpainted layer



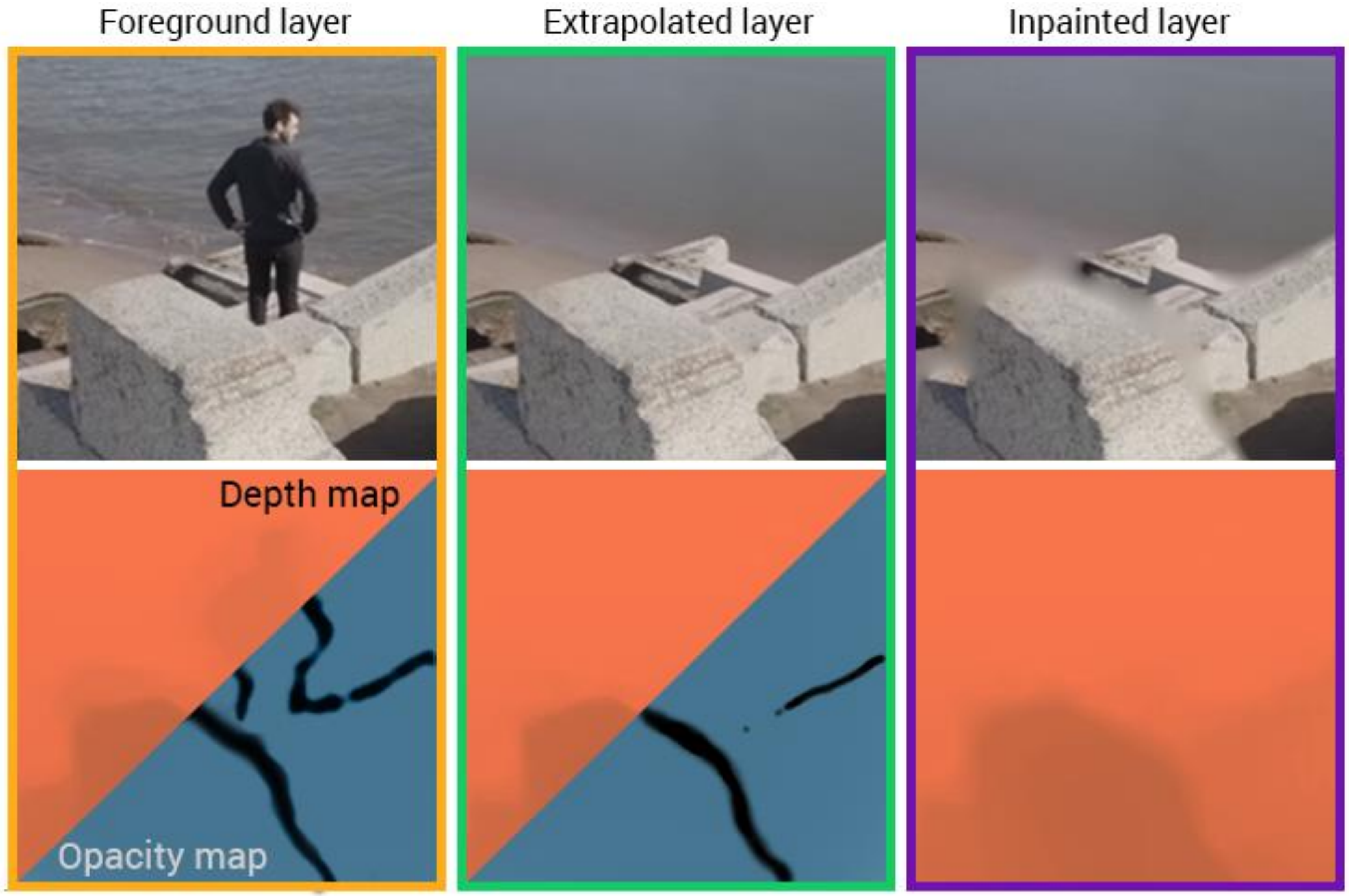
# Layered video representation

IEEE VR 2019 OSAKA

Mesh-based approach

Three layers

- Foreground layer
- Extrapolated layer
- Inpainted layer



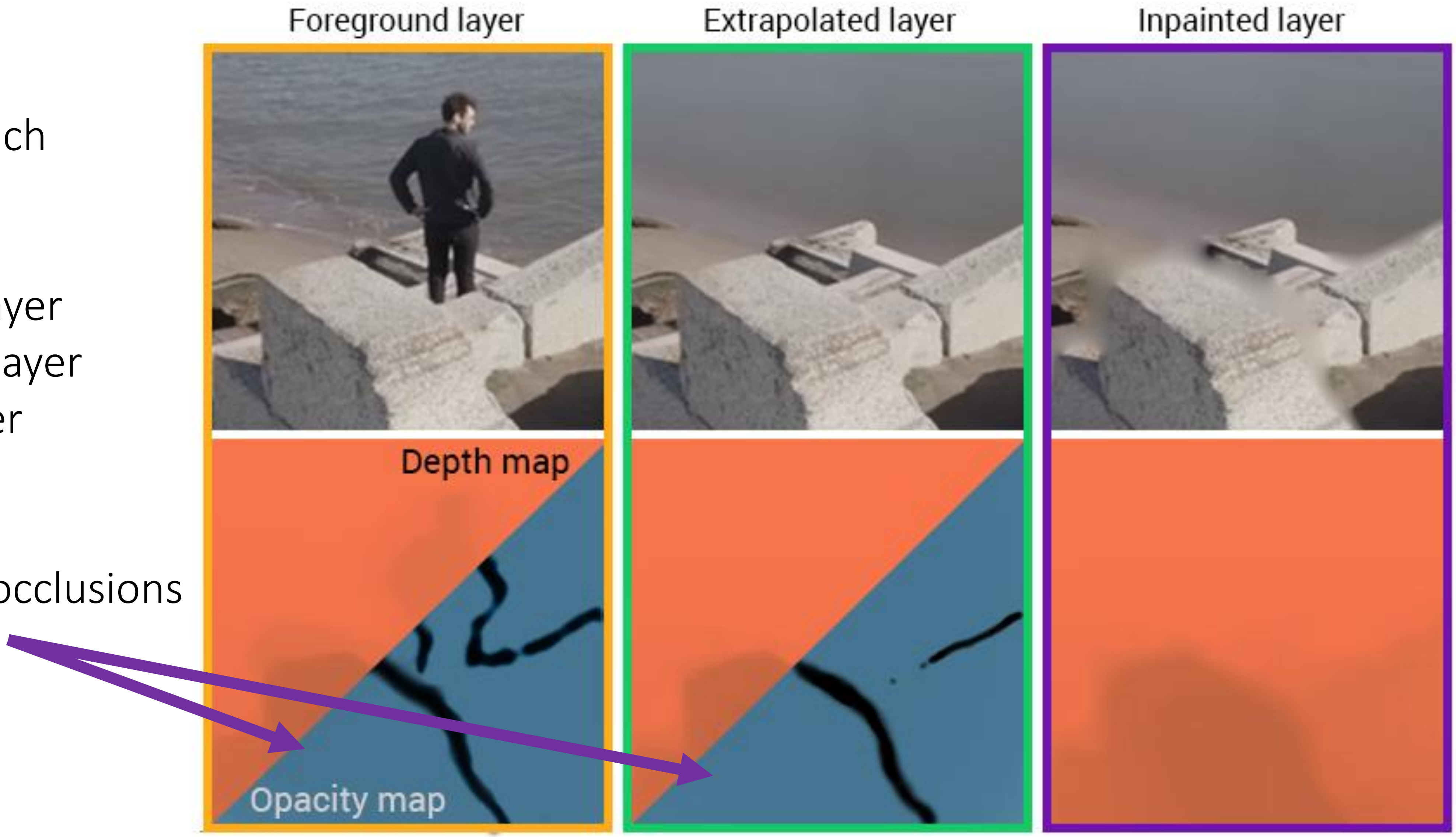
# Layered video representation



Mesh-based approach

Three layers
- Foreground layer
- Extrapolated layer
- Inpainted layer

Opacity map
Transparency at disocclusions



# Depth improvement optimization

# Depth improvement

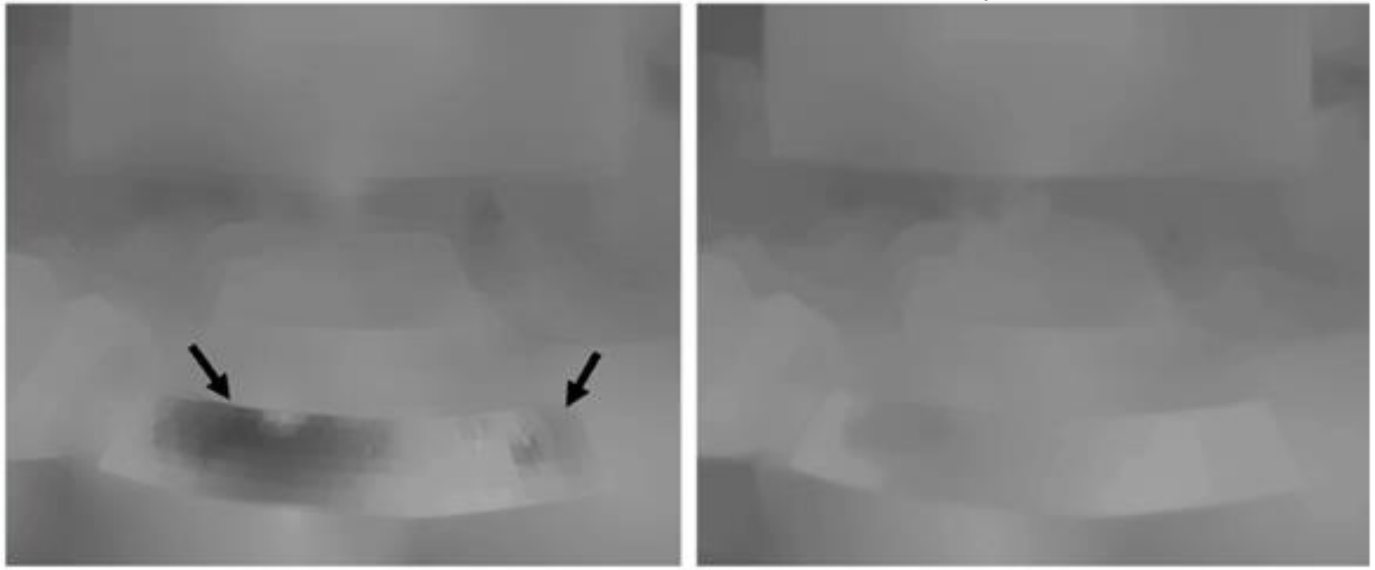


# Depth improvement



Input Improved

Input Improved

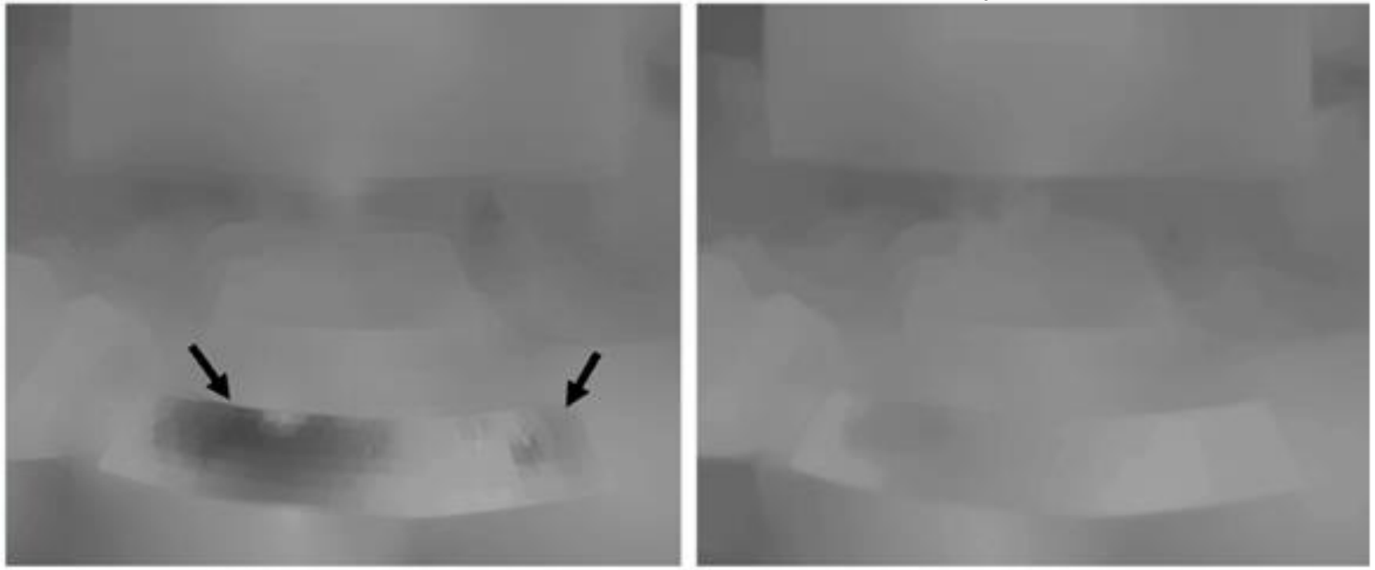


# Depth improvement



Input Improved

Input Improved



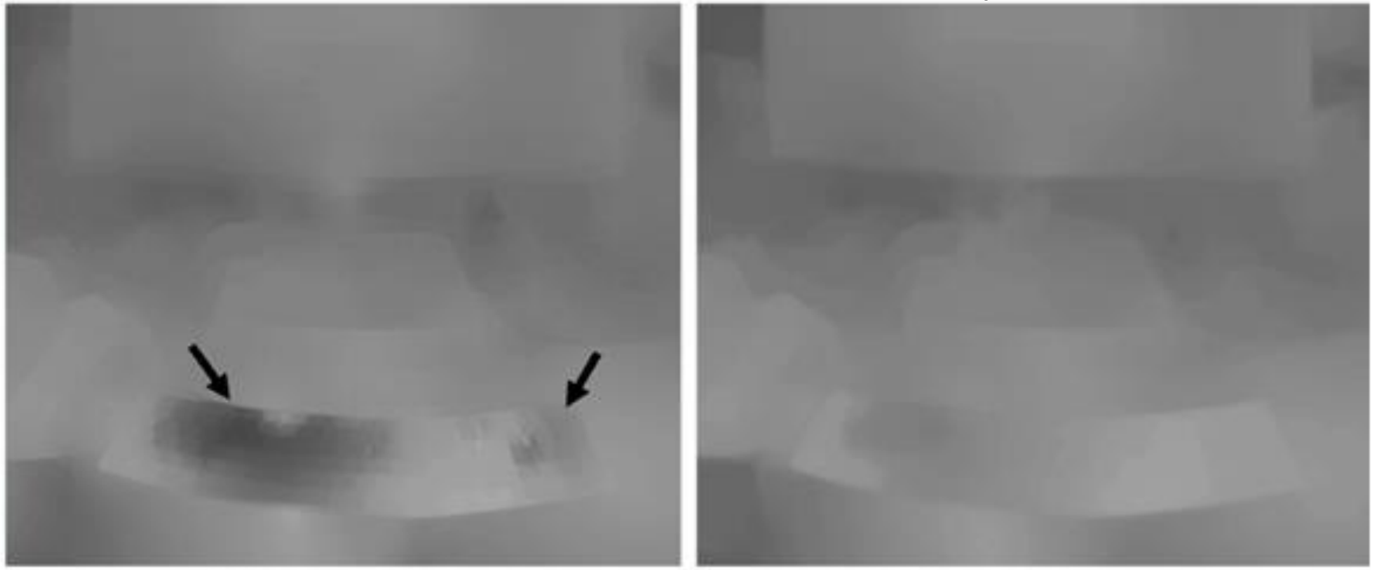
$$\underset{d}{\operatorname{argmin}} \; \lambda_{data} E_{data} + \lambda_e E_e + \lambda_{sm} E_{sm} + \lambda_t E_t$$

Data term

# Depth improvement

Input Improved

Input Improved



Edge preservation

$$\underset{d}{\operatorname{argmin}}\ \lambda_{data}E_{data} + \lambda_e E_e + \lambda_{sm}E_{sm} + \lambda_t E_t$$

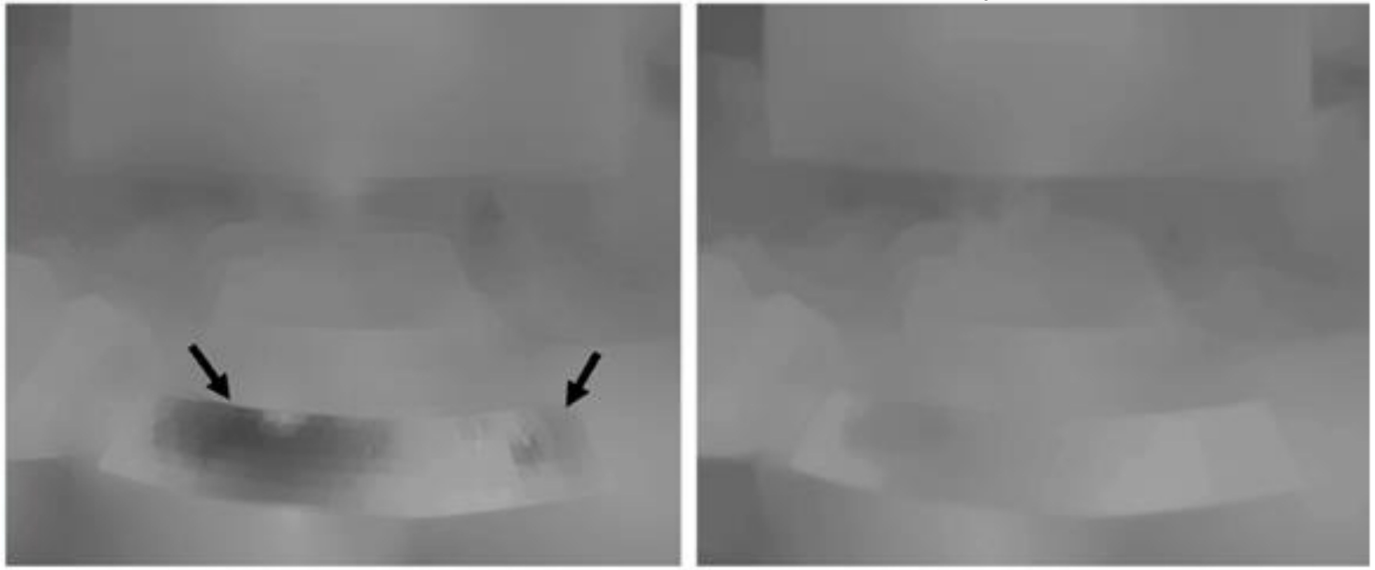
Data term

# Depth improvement



Input Improved

Input Improved



Edge preservation

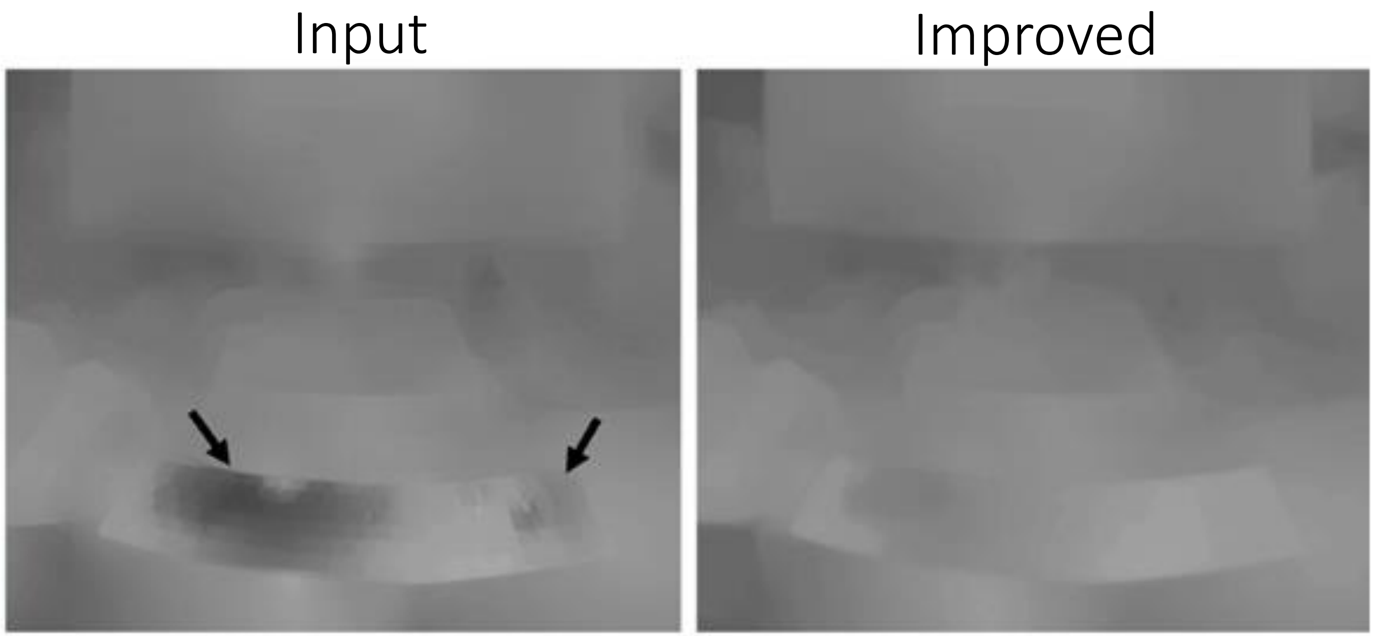
$$\underset{d}{\operatorname{argmin}}\ \lambda_{data} E_{data} + \lambda_e E_e + \lambda_{sm} E_{sm} + \lambda_t E_t$$

Data term

Spatial smoothness

# Depth improvement

IEEE VR 2019 OSAKA


Edge preservation

Temporal consistency

$$\operatorname*{argmin}_{d} \; \lambda_{data} E_{data} + \lambda_e E_e + \lambda_{sm} E_{sm} + \lambda_t E_t$$

Data term

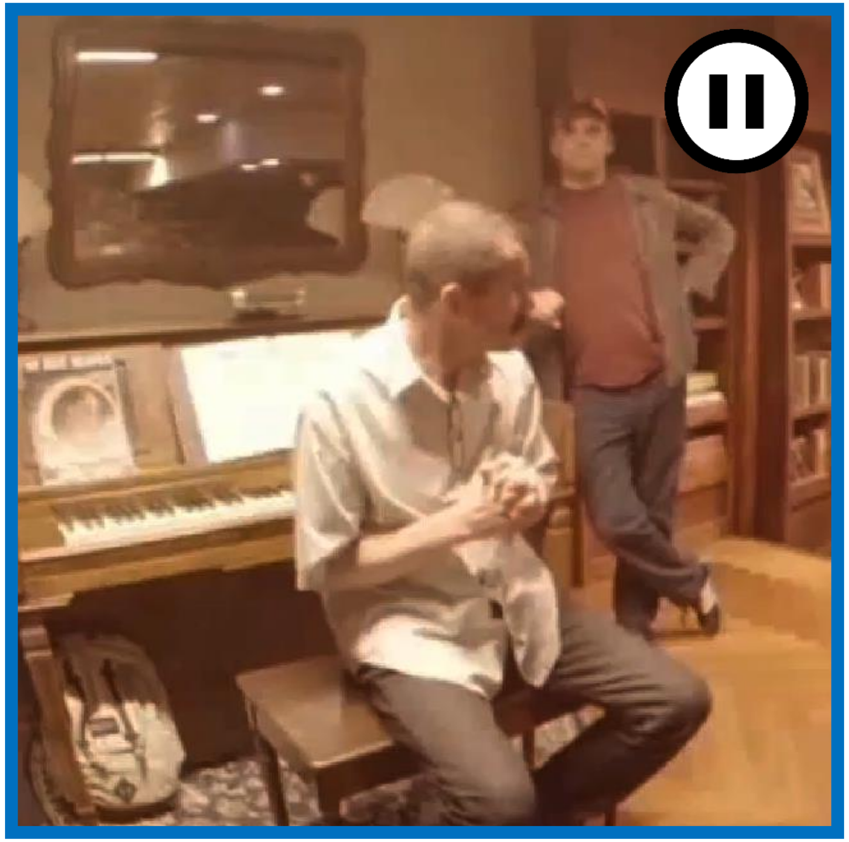
Spatial smoothness

# Results

# Results



Close-up



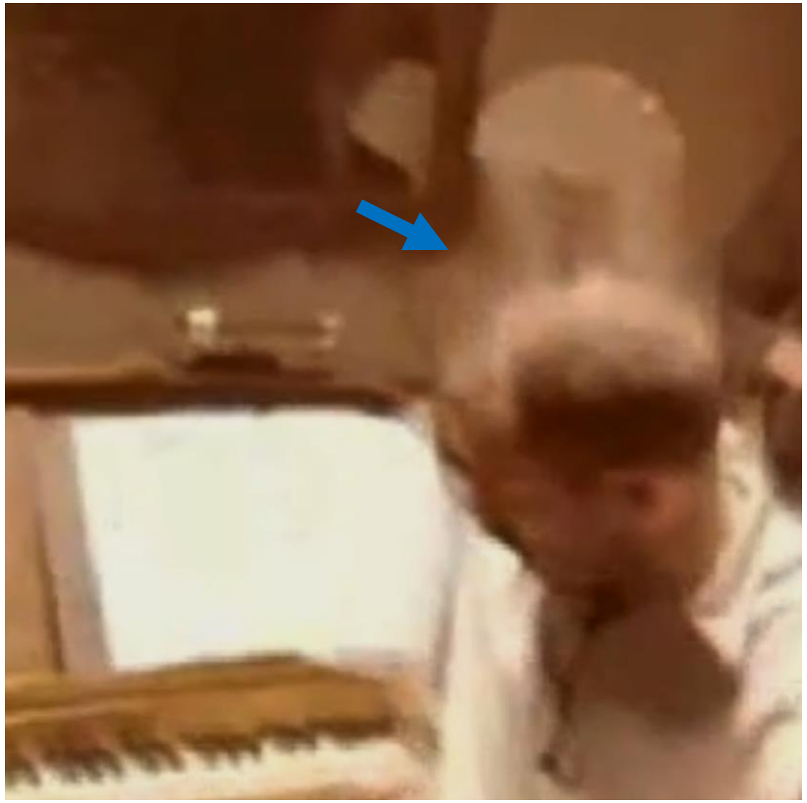
VR view (stereo)

# Results



Naive reprojection

Layered – raw depth

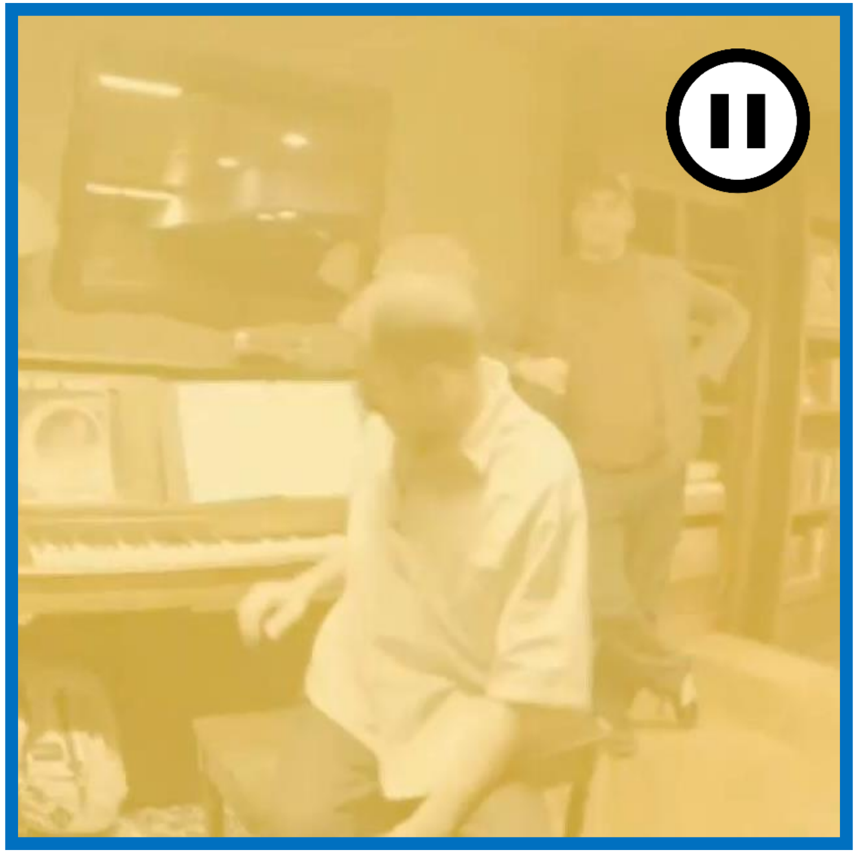


Ours – improved depth

# Results



Close-up



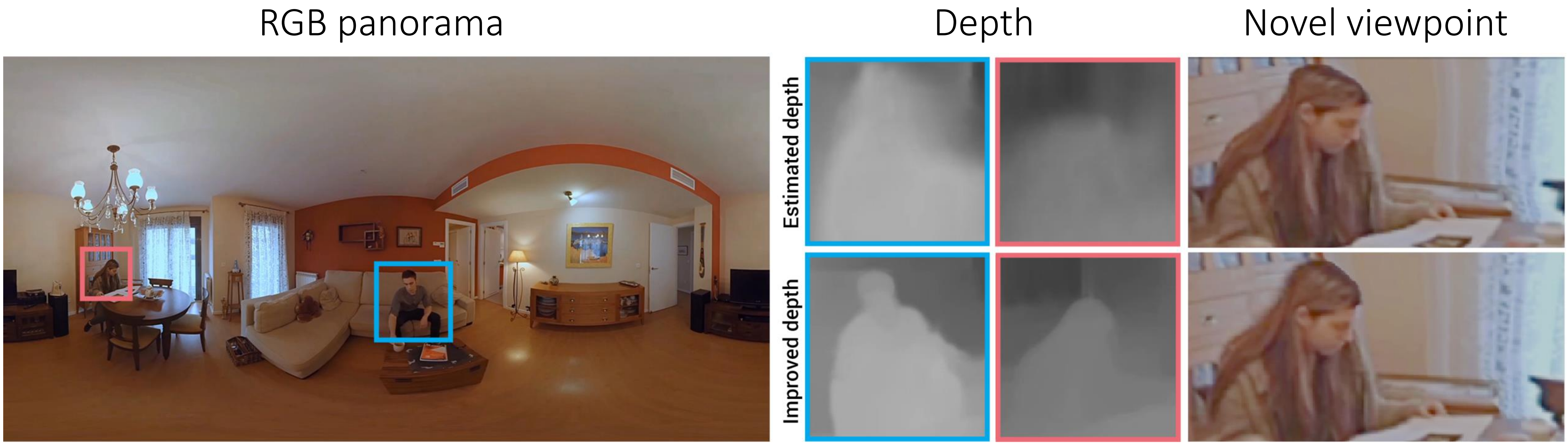
VR view (stereo)

Foreground layer
Extrapolated layer
In-painted layer

# Results using monocular video



Depth estimation: DNN-based approach [Godard et al. 2017]




Source: [Serrano et al. 2017]

# Evaluation



**Experiment #1: Preference (blind)**

- Videos with added parallax using our method were preferred in six out of the seven cases

**Experiment #2: Sickness**

- 3-DoF: 17 out of 24 participants reported symptoms of sickness, dizziness, and/or vertigo
- Ours: 5 out of 24 reported these symptoms

**Experiment #3: Preference (non-blind)**

- Our method was strongly preferred for five out of the six videos, with no clear preference for the sixth

# Evaluation



**Experiment #1: Preference (blind)**

- Videos with added parallax using our method were preferred in six out of the seven cases

**Experiment #2: Sickness**

- 3-DoF: 17 out of 24 participants reported symptoms of sickness, dizziness, and/or vertigo
- Ours: 5 out of 24 reported these symptoms

**Experiment #3: Preference (non-blind)**

- Our method was strongly preferred for five out of the six videos, with no clear preference for the sixth

# Evaluation



**Experiment #1: Preference (blind)**

- Videos with added parallax using our method were preferred in six out of the seven cases

**Experiment #2: Sickness**

- 3-DoF: 17 out of 24 participants reported symptoms of sickness, dizziness, and/or vertigo
- Ours: 5 out of 24 reported these symptoms

**Experiment #3: Preference (non-blind)**

- Our method was strongly preferred for five out of the six videos, with no clear preference for the sixth

# Evaluation



**Experiment #1: Preference (blind)**

- Videos with added parallax using our method were preferred in six out of the seven cases

**Experiment #2: Sickness**

- 3-DoF: 17 out of 24 participants reported symptoms of sickness, dizziness, and/or vertigo
- Ours: 5 out of 24 reported these symptoms

**Experiment #3: Preference (non-blind)**

- Our method was strongly preferred for five out of the six videos, with no clear preference for the sixth

# Conclusions



- Novel approach to enable head motion parallax in 360 video

- Independent of a specific hardware, camera setup, or recorded baseline

- Requires only RGBD 360 video as input
    - Robust to depth inaccuracies
    - Can deal with 360 monocular video (with depth estimation)

- Our user studies confirm that our method provides a more compelling viewing experience, while reducing discomfort and sickness.

# Limitations and future work



- Static camera assumption
  - Large amount of 360 content shot with static cameras
  - Manufacturers typically recommend static cameras
- Number of layers in the layered representation
- Our method relies on the quality of the input depth map
  - Combining ideas from our work and the works by Overbeck et al. and Hedman et al. could lead to higher-quality 6-DoF capture

# Motion parallax for 360º RGBD video

Project page (more results, demo):

http://webdiis.unizar.es/~aserrano/projects/VR-6dof

Adobe

Graphics and Imaging Lab

IEEE VR 2019 OSAKA

# Layered video representation: opacity maps




$$\hat{\alpha}^F = S(G \circledast (\tau(O^F \bullet K)))$$

- Logistic function → $S$
- Thresholding → $\tau$
- Closing (disk kernel) → $K$
- Gaussian blur → $G$
- Original orientations → $O^F$

# Naïve handling of disocclusions

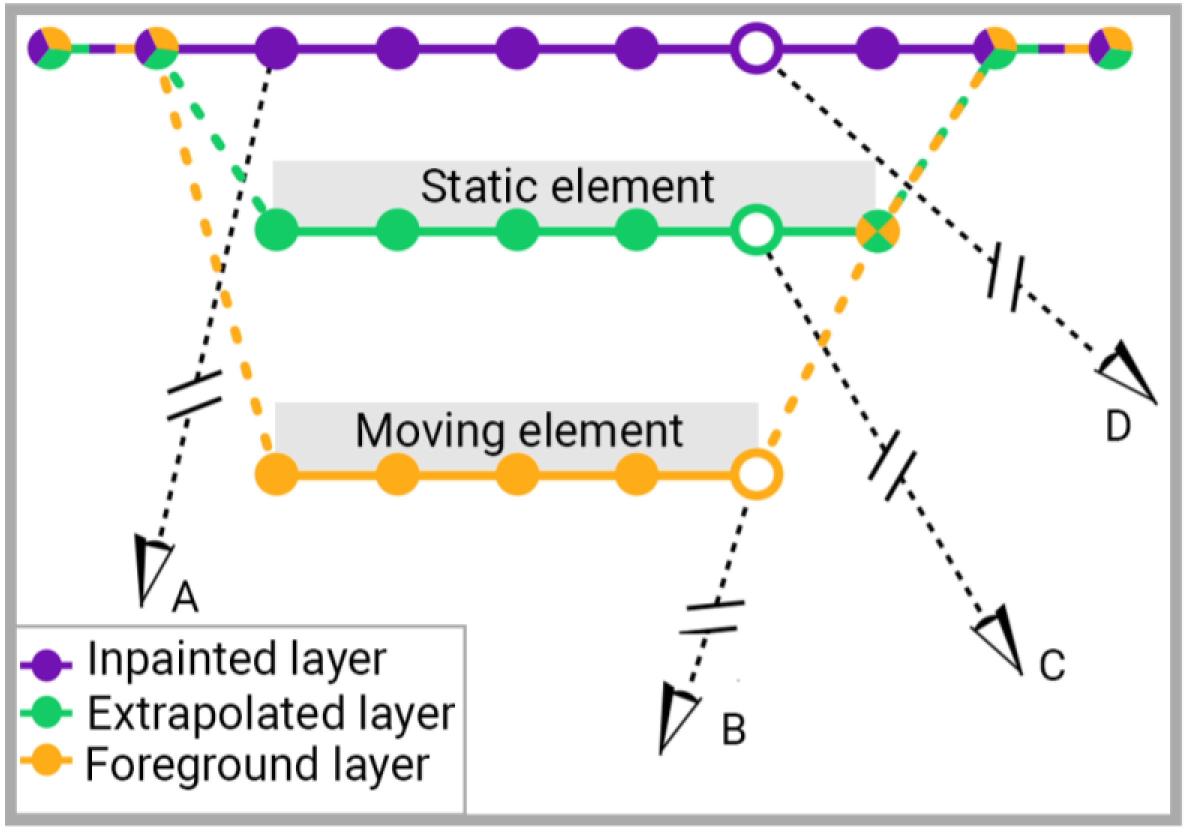


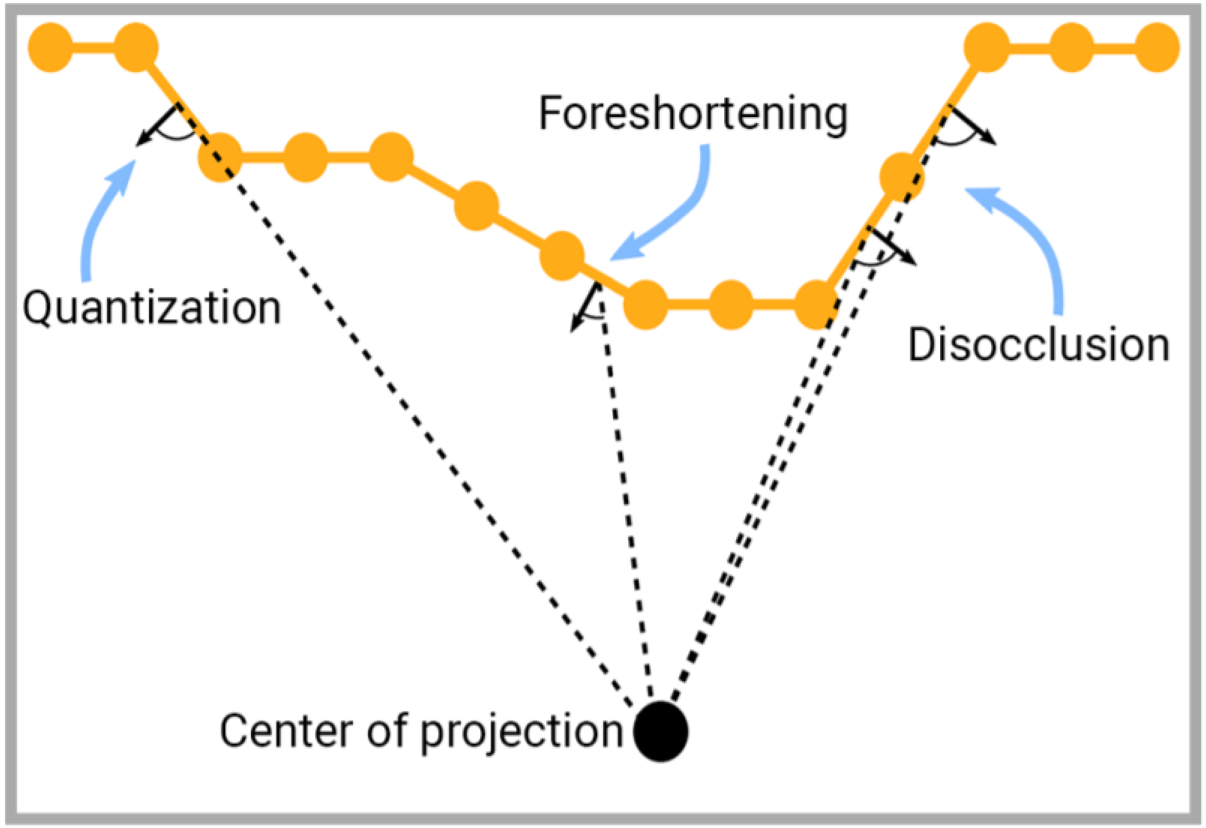
HMD original view

Displaced view

# Layered representation

IEEE VR 2019 OSAKA



# Depth improvement optimization



$$E_{data}(i) = \sum_i w_d(i)\left(d(i) - \hat{d}(i)\right)^2,$$

$$E_e(i) = \sum_i \left( d(i) - \sum_{j \in \mathcal{N}(i)} w_e(i,j)\, d(j) \right)^2$$

$$E_{sm}(i) = \sum_i \sum_{j \in \mathcal{N}(i)} w_{sm}(i)\left(d(i) - d(j)\right)^2$$

$$E_t(i) = \sum_i w_t(i)\left(d(i) - \psi_{prev \to cur}\left(d_{prev}(i)\right)\right)^2$$