

Depth-aware Mixed Reality: Capture the AR-Flag

Students:

Ihsan Berkan Balaban
Muhammad Faizan
Jeremias Neth

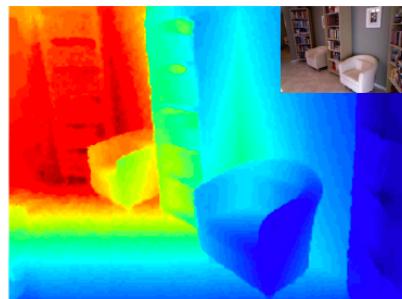
Tutors:

Patrick Ruhkamp
Benjamin Busam
Hyunjun Junq

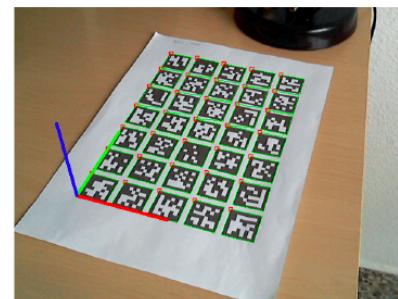
Project Overview

- **Objective:** Interactively manipulate a scene, whose 3D structure is estimated with an RGB-D camera. Realistic object augmentation and occlusion awareness is explored, and the results are implemented in a simple AR game.
- Three main parts:
 1. Depth Estimation and Camera Pose
 2. User Interaction and Scene Manipulation
 3. Realism and Embedding
 4. Capture the Flag AR (Optional)

Depth Estimation



Pose Tracking



Augmentation





Depth Estimation and Camera Pose

Depth Estimation

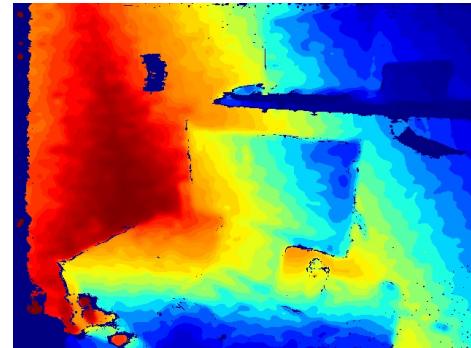
- Intel RealSense Camera.



IR Left Image



IR Right Image



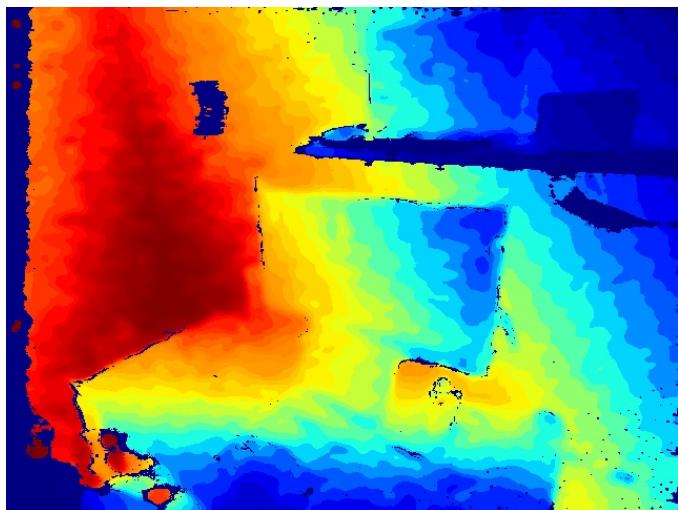
Depth Map



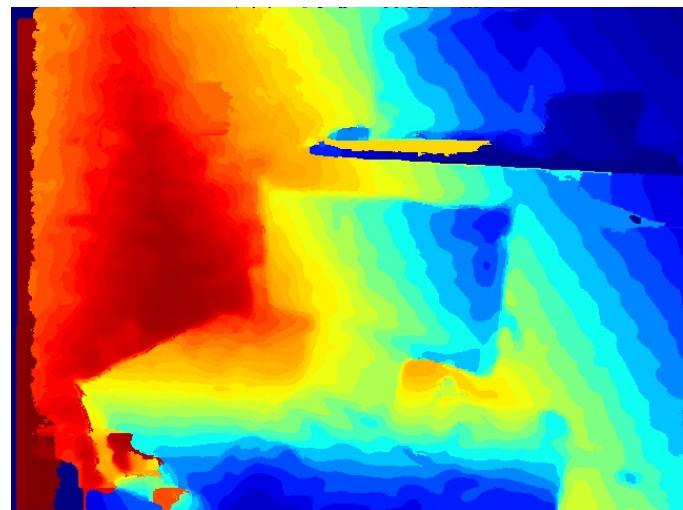
[1] Intel RealSense D435: <https://www.intelrealsense.com/depth-camera-d435/>

Depth Estimation

- Post-processing on depth map.
- Filters:
 - Spatial Edge-Preserving filter [2]
 - Temporal filter
 - Holes Filling filter



Depth Map (Without filtering)



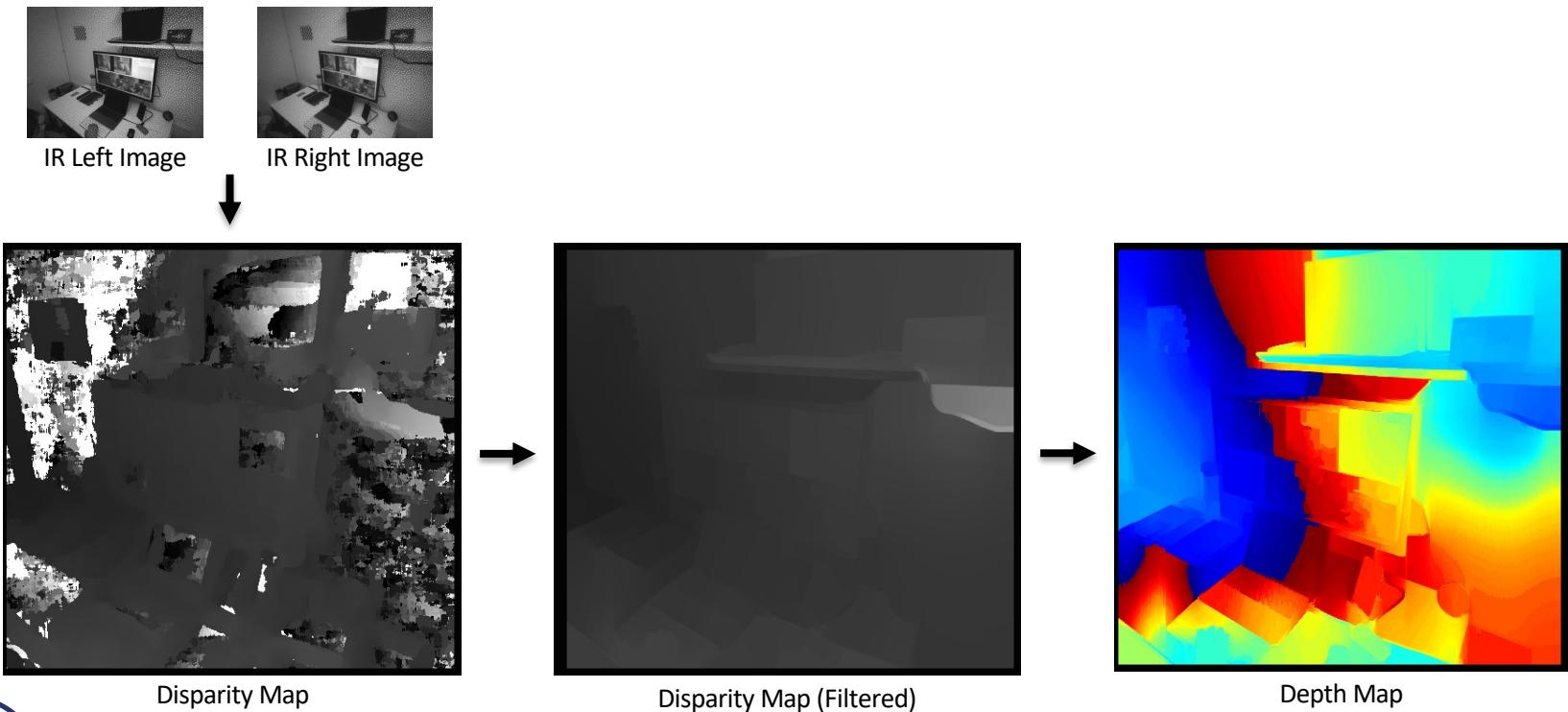
Depth Map (With filtering)

[2] Domain Transform for Edge-Aware Image & Video Processing, Eduardo S.L. Gastal, Manuel M. Oliveira (<https://www.inf.ufrgs.br/~eslgastal/DomainTransform/>)



Depth Estimation

- Stereo matching
- Filters
 - WLS Disparity Filter

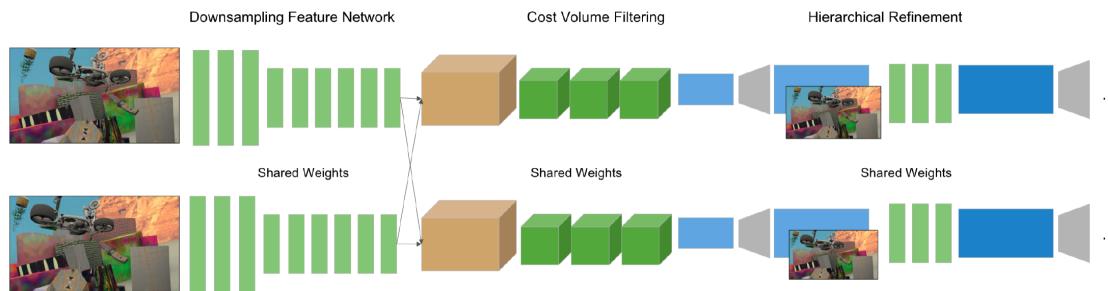


[3] Disparity map post-filtering: https://docs.opencv.org/trunk/d3/d14/tutorial_ximgproc_disparity_filtering.html

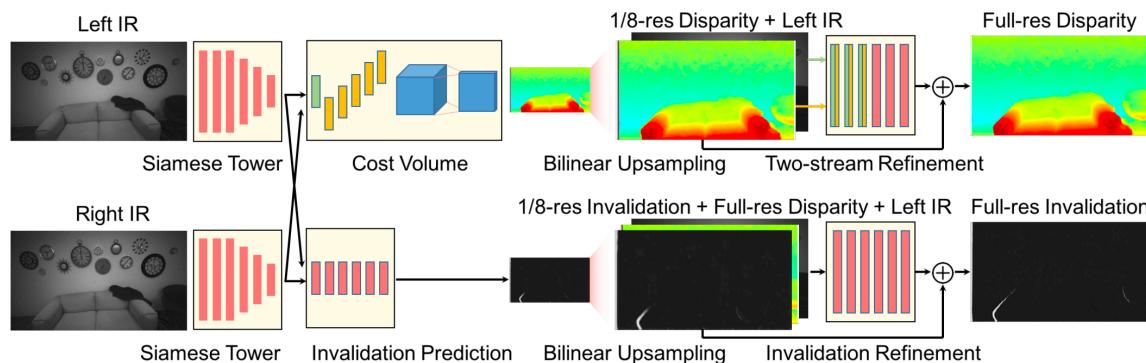


[Next] Depth Estimation using Deep NNs

- StereoNet:



- ActiveStereoNet:



[4] StereoNet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction, Sameh Khamis, Sean Fanello, Christoph Rhemann, Adarsh Kowdle, Julien Valentin and Shahram Izadi, 2018, 1807.08865, arXiv

[5] ActiveStereoNet: End-to-End Self-Supervised Learning for Active Stereo Systems, Yinda Zhang, Sameh Khamis, Christoph Rhemann, Julien Valentin, Adarsh Kowdle, Vladimir Tankovich, Michael Schoenberg, Shahram Izadi, Thomas Funkhouser and Sean Fanello, 2018, 1807.06009, arXiv





User Interaction and Scene Manipulation





Realism and Embedding

Blending Object with the Scene

- The depth map is used to blend the augmented object with occlusion awareness, given their pose.
- Triangular mesh of the object is first rendered using Pyrender library in order to get the color and depth maps.

Triangular Mesh Represented in Meshlab



Blending Object with the Scene

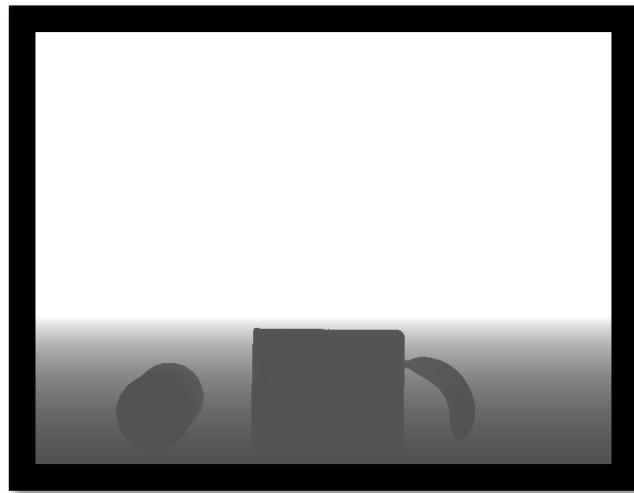
- ($w, h, 3$) channel floating-point color image from Pyrender is on the left and (w, h) floating-point depth image from Pyrender is on the right. Camera parameters (intrinsic and pose) and triangular mesh as well as color and depth image of scene has been found online.



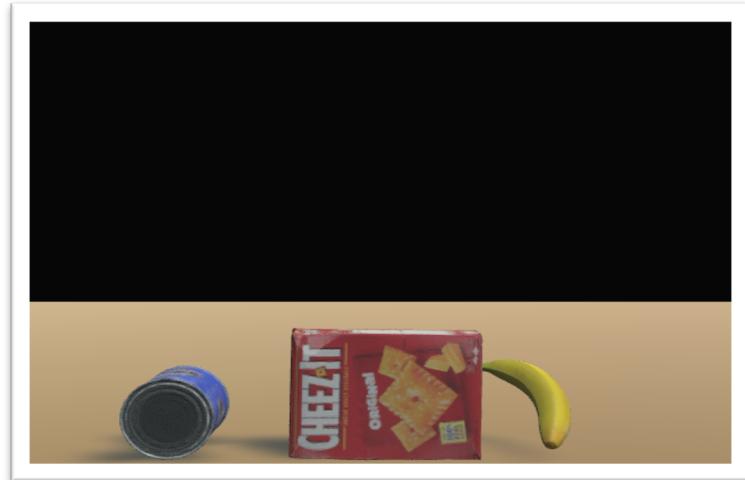
Blending Object with the Scene

- The object is blended into the scene using its color and depth image. For this, the depth and color images of the scene is used.

Depth image of the scene



RGB Color image of the scene



Blending Object with the Scene

Ground Truth Blending



Our blending output



x=275.344 y=302.617 [202, 176, 136]



Time Plan

19.06.2020 – 24.07.2020	Week 1	Week 2	Intermediate Presentation	Week 4	Week 5
			Final Presentation		
WP1: Depth Estimation and Camera Pose	<ul style="list-style-type: none"> Get familiar with topics and read papers. Getting started with Intel RealSense SDK. 	<ul style="list-style-type: none"> Post-Processing filters on Depth map from RGB-D camera. Depth improvement using WLS filters and StereoBM. 	<ul style="list-style-type: none"> Deep learning approaches for depth estimation (StereoNet, ActiveStereoNet). 	<ul style="list-style-type: none"> Depth map fusion. Integration with WP2. 	<ul style="list-style-type: none"> Performance and quality improvements.
WP2: User Interaction and Scene Manipulation					
WP3: Realism and Embedding	Get familiar with topics and read papers.	<p>Learning how to use Pyrender.</p> <p>Blending the object into the scene with some occlusion.</p>	Working on occlusion awareness while blending.	Working on image harmonization to improve the realism of the scene.	<p>Integration with other WP's.</p> <p>Improving performance.</p>





Thank you very much ☺
Any question?

REFERENCES

- [1] Intel RealSense D435: <https://www.intelrealsense.com/depth-camera-d435/>
- [2] Domain Transform for Edge-Aware Image & Video Processing, Eduardo S.L. Gastal, Manuel M. Oliveira (<https://www.inf.ufrgs.br/~eslgastal/DomainTransform/>)
- [3] Disparity map post-filtering: https://docs.opencv.org/trunk/d3/d14/tutorial_ximgproc_disparity_filtering.html
- [4] StereoNet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction, Sameh Khamis, Sean Fanello, Christoph Rhemann, Adarsh Kowdle, Julien Valentin and Shahram Izadi, 2018, 1807.08865, arXiv
- [5] ActiveStereoNet: End-to-End Self-Supervised Learning for Active Stereo Systems, Yinda Zhang, Sameh Khamis, Christoph Rhemann, Julien Valentin, Adarsh Kowdle, Vladimir Tankovich, Michael Schoenberg, Shahram Izadi, Thomas Funkhouser and Sean Fanello, 2018, 1807.06009, arXiv

