PRINCETON VISION & ROBOTICS

Google

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

ActiveStereoNet

End-to-End Self-Supervised Learning for Active Stereo Systems

Presenter: Yinda Zhang

Yinda Zhang[1,2], Sameh Khamis[1], Christoph Rhemann[1], Julien Valentin[1], Adarsh Kowdle[1], Vladimir Tankovich[1], Michael Schoenberg[1], Shahram Izadi[1], Thomas Funkhouser[1,2], Sean Fanello[1]

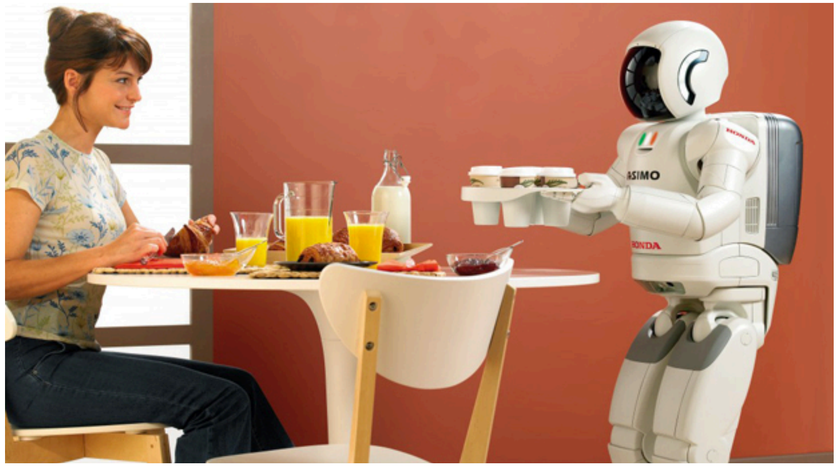
Google Inc.[1] Princeton University[2]

Project Webpage: http://asn.cs.princeton.edu/

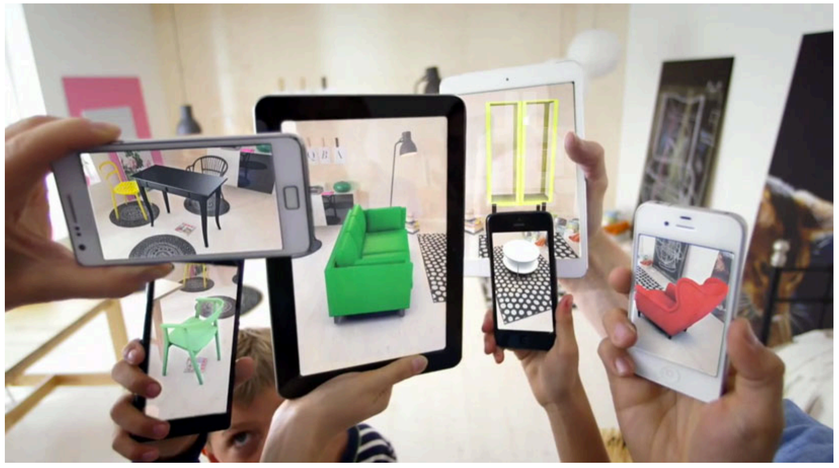
Goal: Sensing 3D Geometry

Biometric authentication

Autonomous driving



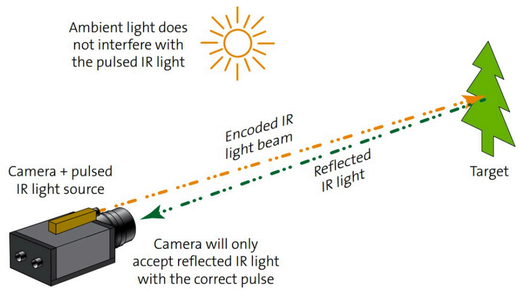
Indoor Robotics



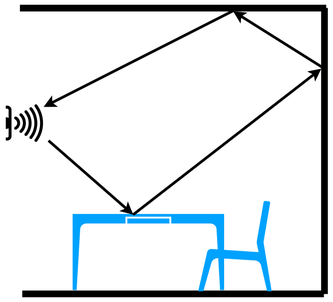
Augmented Reality

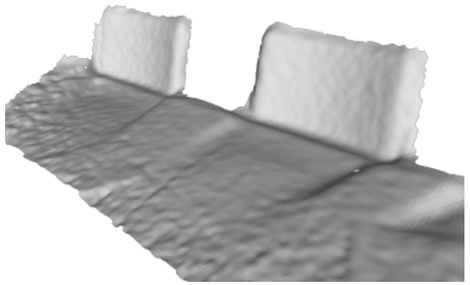
Active Depth Sensing

- Time of Flight
  - ✗ Fast motion
  - ✗ Multi-Path



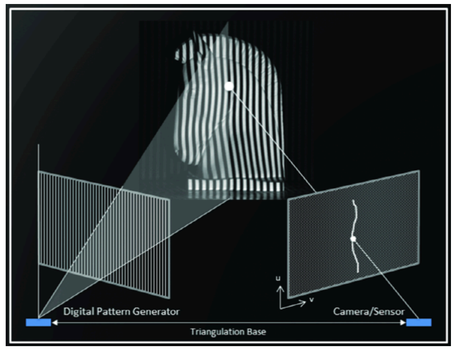
Time of Flight



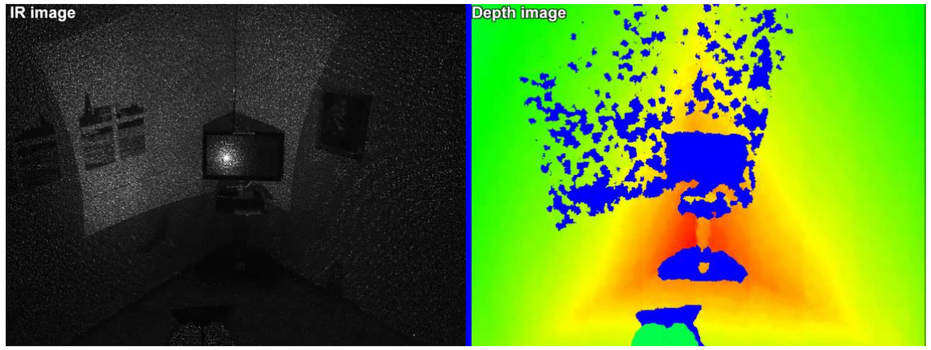


Multi-Path Interference

- Structured Light
  - ✗ Calibration
  - ✗ Multi-Device



Structured Light



Multi-Device Interference

Passive Depth Sensing

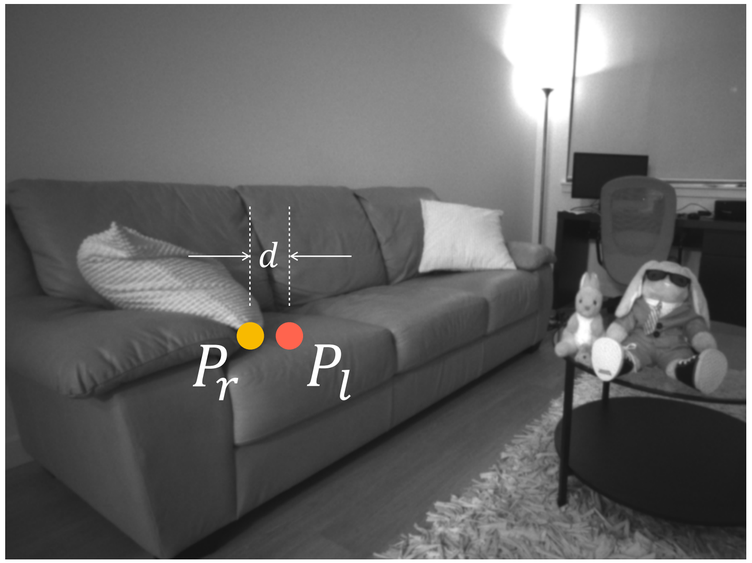
- Stereo Matching

✗ Texture-less Region

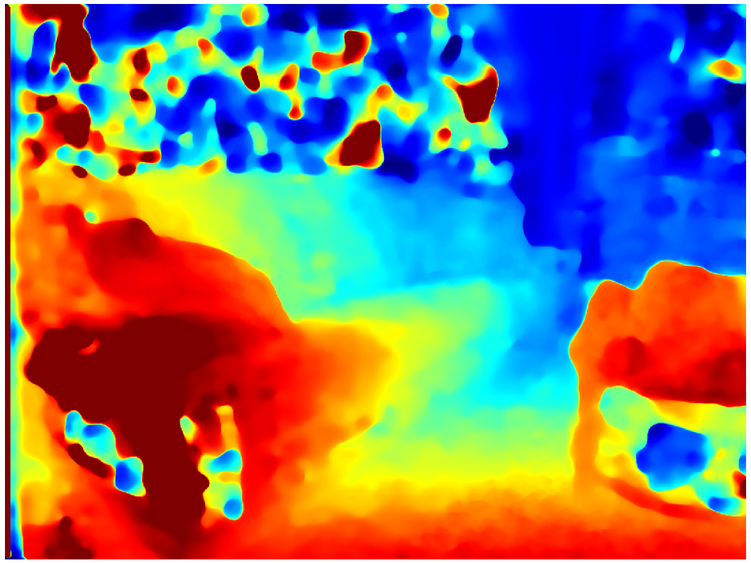
$$Z = \frac{bf}{d}$$

b: *distance between camera centers*
f: *camera focal length*
d: *disparity*
Z: *depth*

Left View



Right View



Disparity

Active Stereo System

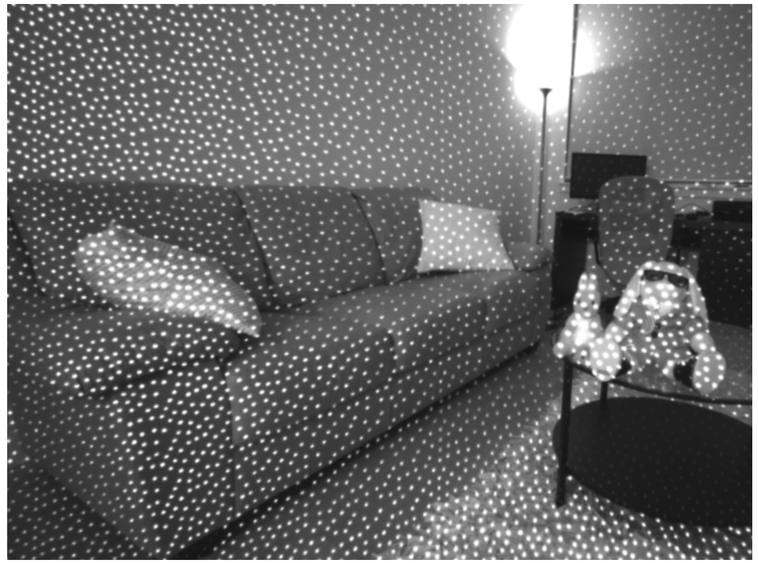
- Stereo Matching

~~✗ Texture-less Region~~

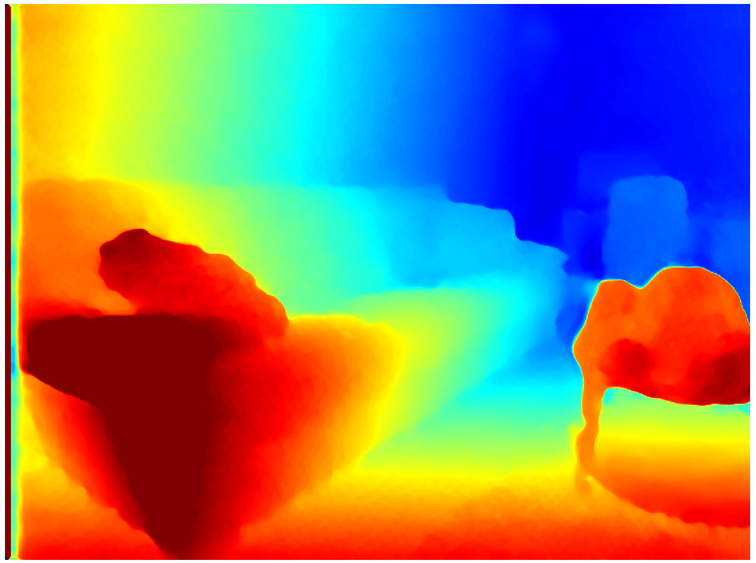
$$Z = \frac{bf}{d}$$

b: *distance between camera centers*
f: *camera focal length*
d: *disparity*
Z: *depth*

Left View

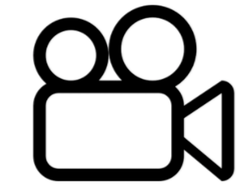


Right View

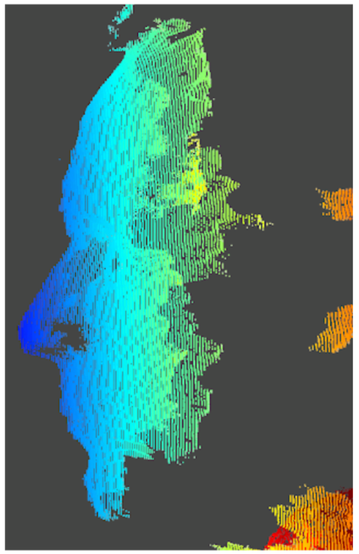


Disparity

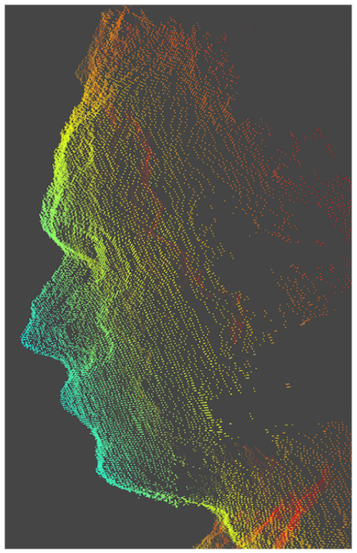
Active Stereo System



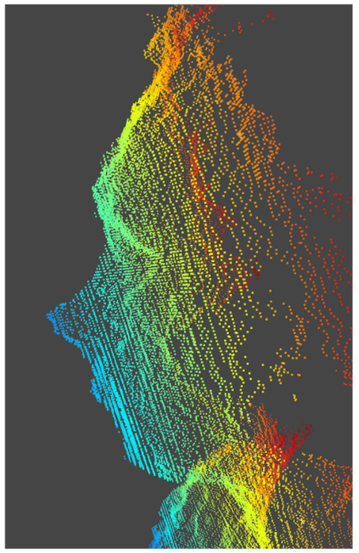
Depth Sensor



300mm



500mm



700mm

2500mm

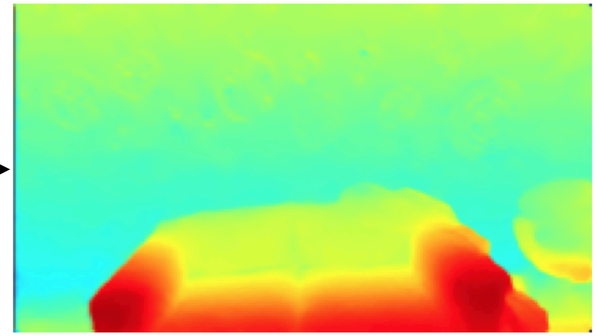
Use deep learning!

No ground truth…

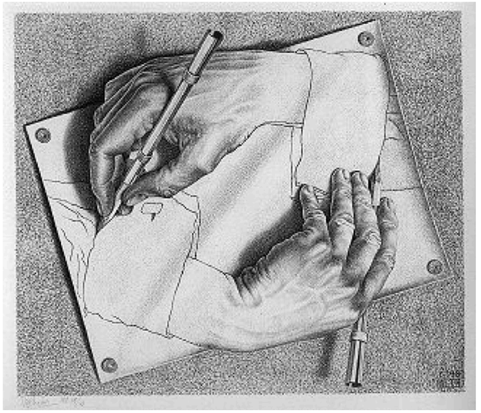
ActiveStereoNet

Input: Left/Right View

End-to-End System



Output: Disparity

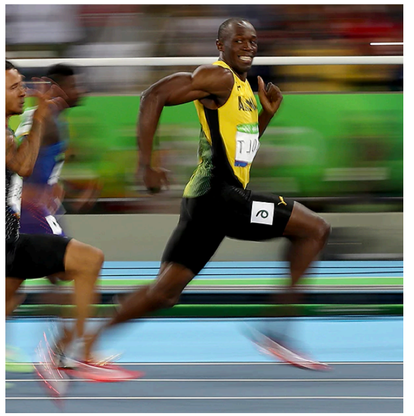


Self-supervised Learning

=

Annotation

Supervision



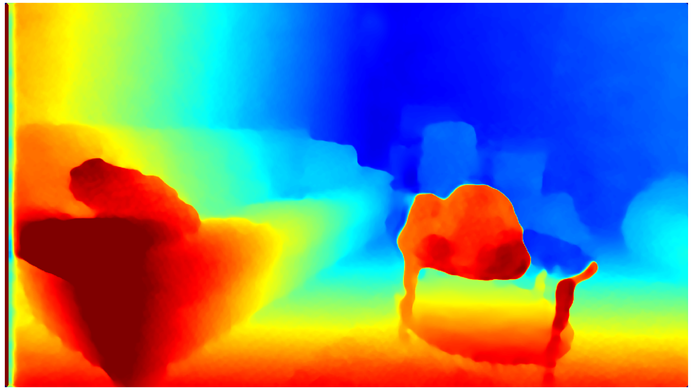
Just keep running…

Self-Supervised Learning

Left View

Right View

Neural Network



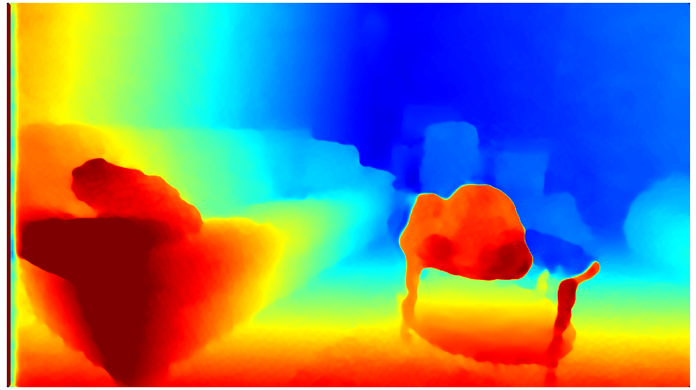
Estimated Disparity

Self-Supervised Learning

Left View

Right View

Neural Network



Estimated Disparity

Left View

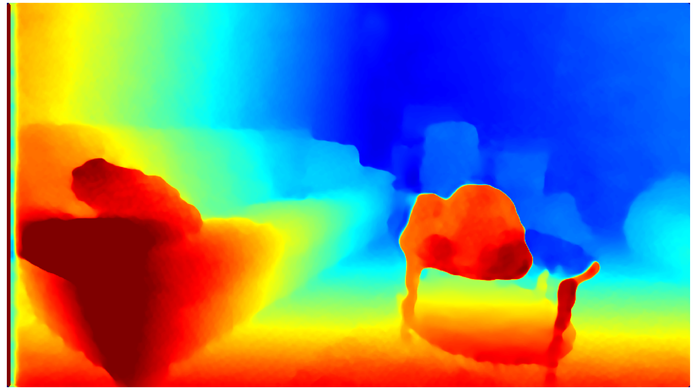
Right View

Self-Supervised Learning

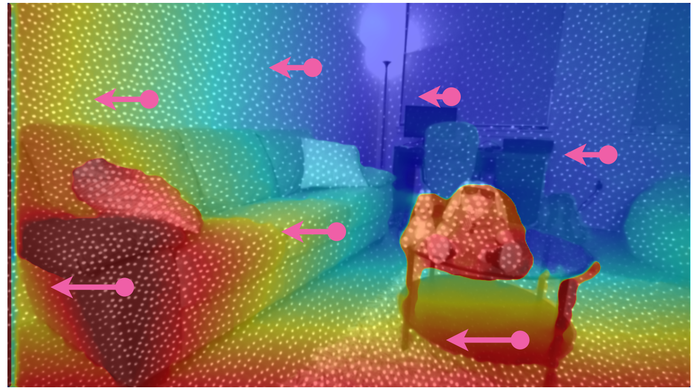
Left View

Right View

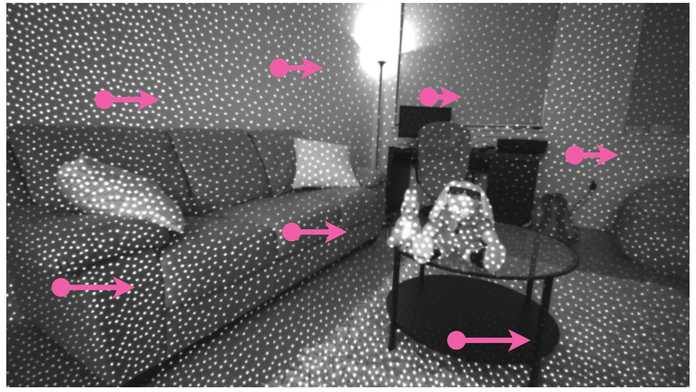
Neural Network



Estimated Disparity



Left View



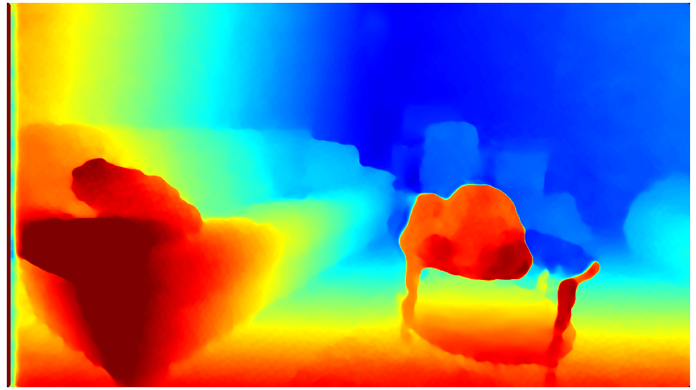
Right View

Self-Supervised Learning

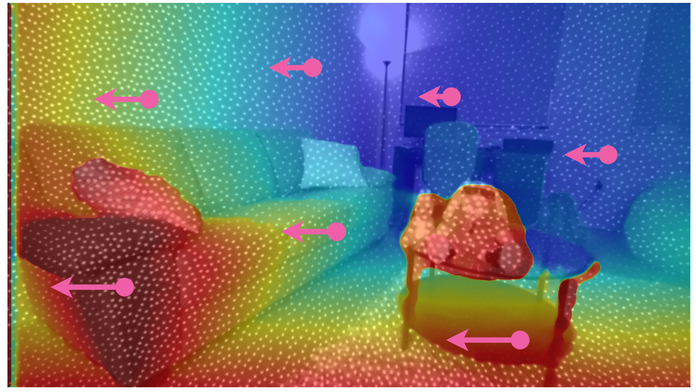
Left View

Right View

Neural Network



Estimated Disparity



Left View

Reconstructed Left View

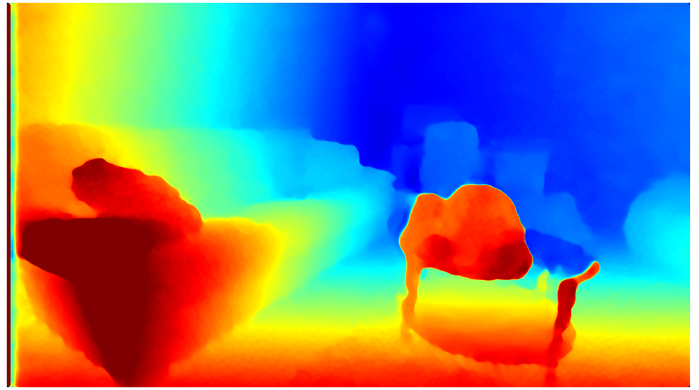
Self-Supervised Learning

Photometric Loss = | Left View - Reconstructed Left View |

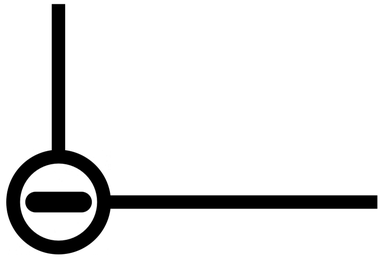
Left View

Right View

Neural Network

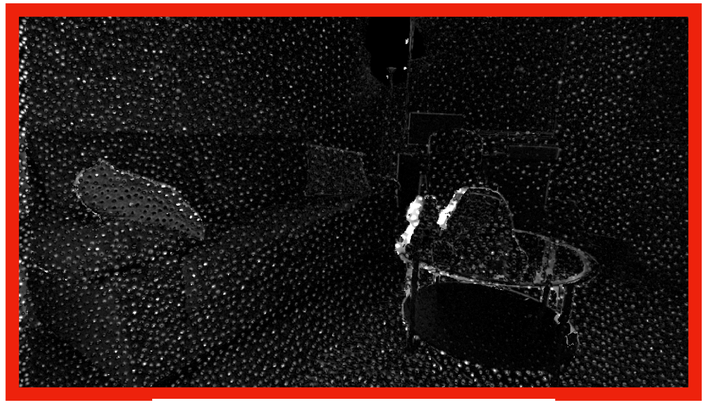


Estimated Disparity



Reconstructed Left View

=



Photometric Loss

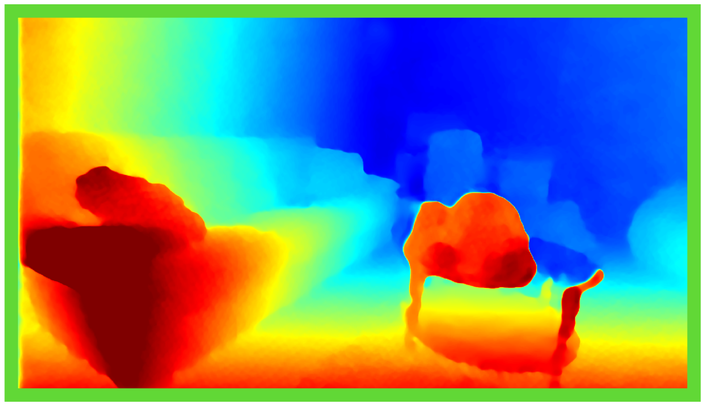
Self-Supervised Learning

Photometric Loss = | Left View - Warping(Right View, Left Disparity) |

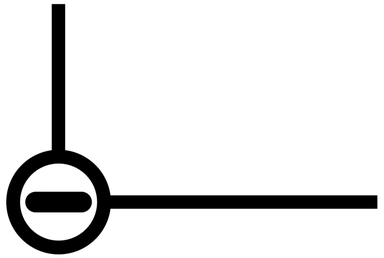
Left View

Right View

Neural Network

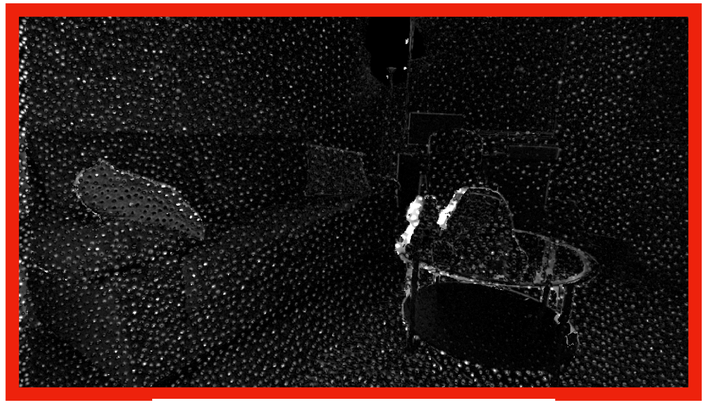


Estimated Disparity



Reconstructed Left View

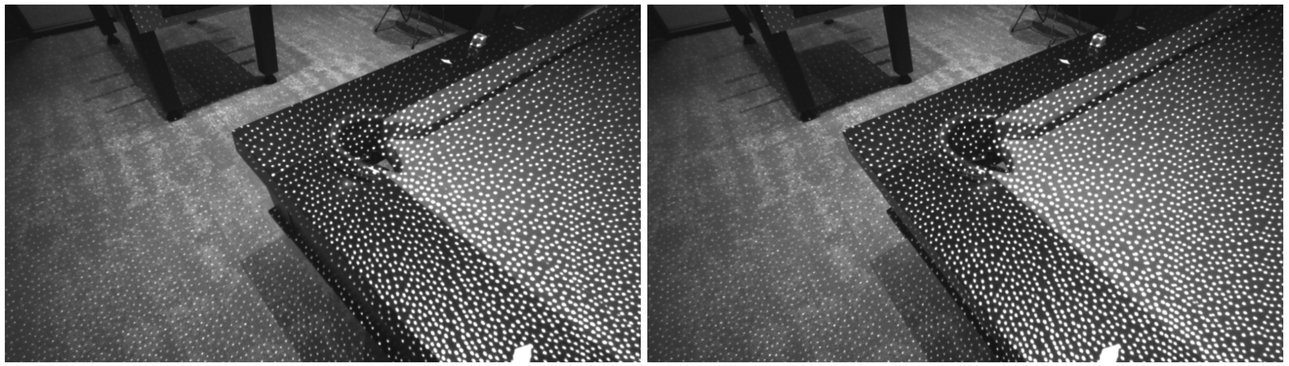
=



Photometric Loss

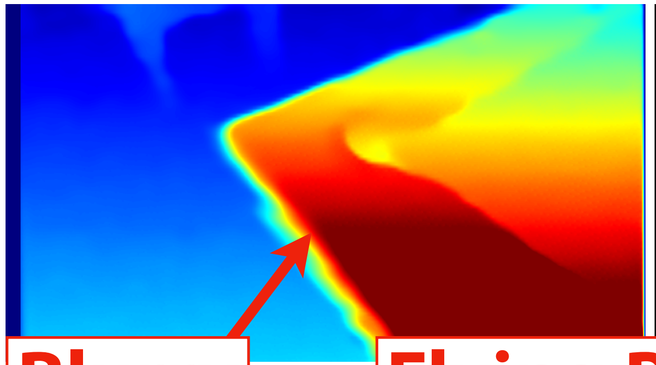
However…

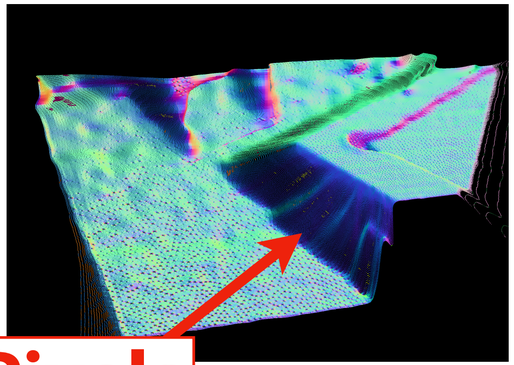
Photometric Loss = | Left View - Warping(Right View, Left Disparity) |



Left View

Right View





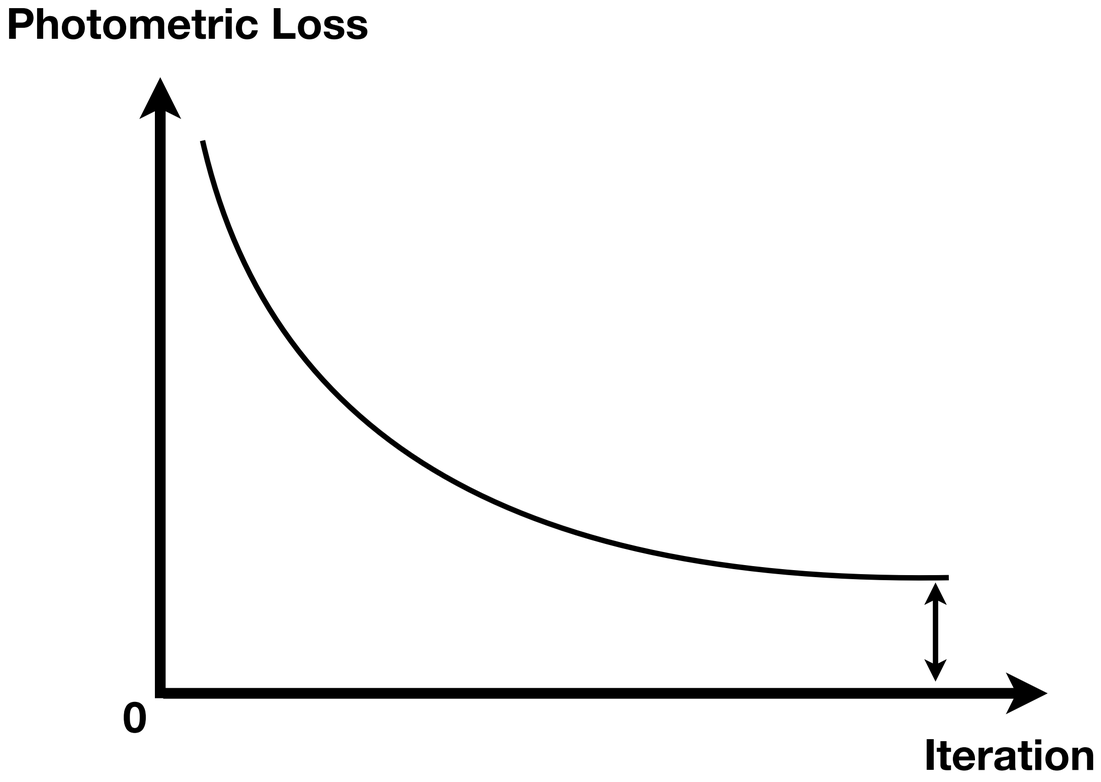
Estimated Disparity

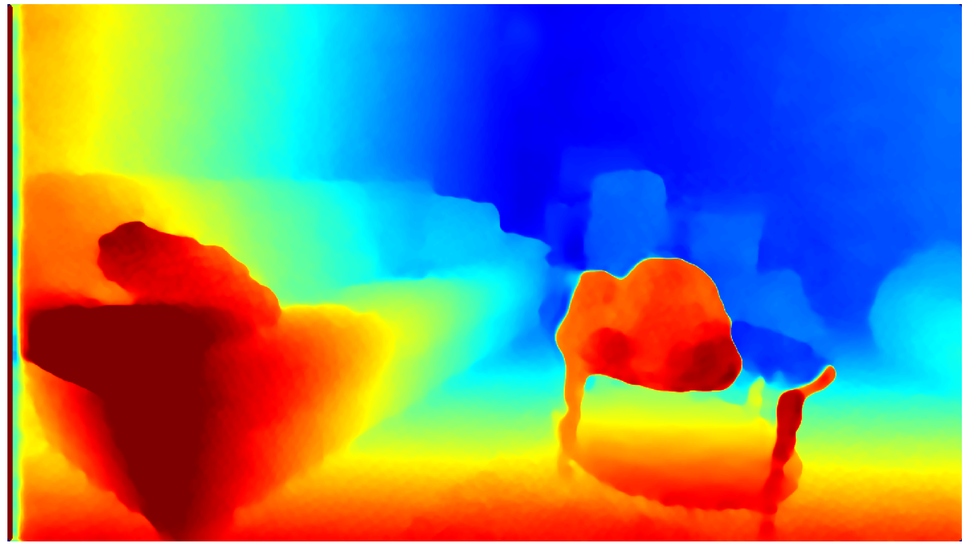
Visualize in 3D

Godard et.al., Unsupervised Monocular Depth Estimation with Left-Right Consistency, CVPR 2017

How to fix the problem?

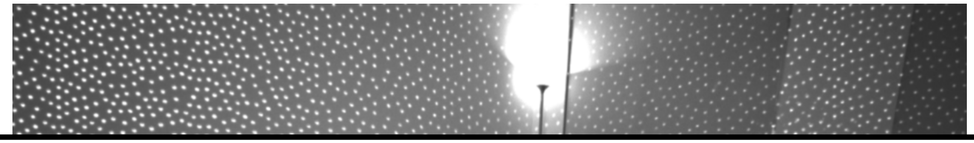
$$\text{Disparity} = \text{argmin}(\text{Photometric Loss})$$





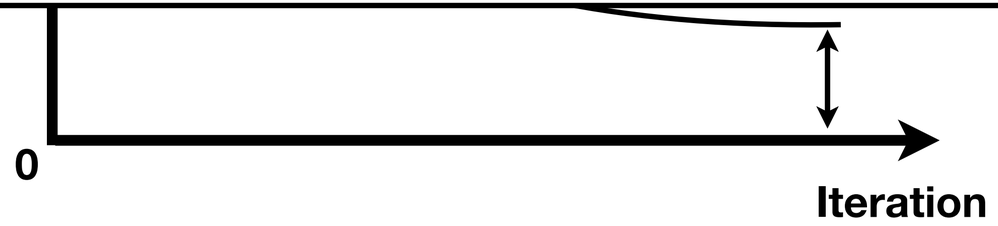
How to fix the problem?

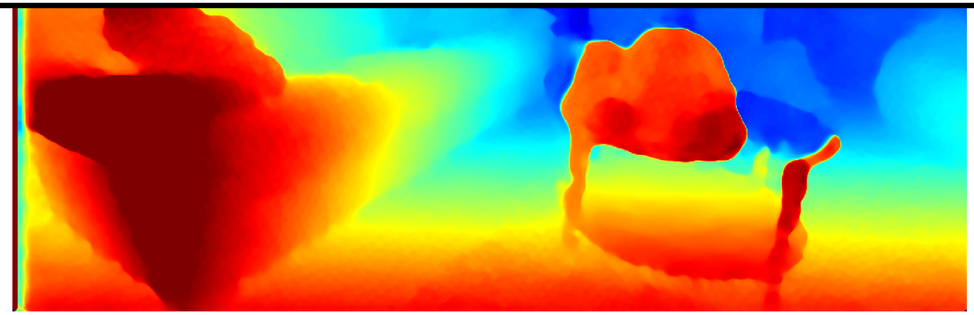
$$\text{Disparity} = \text{argmin}(\text{Photometric Loss})$$



Lower Loss → Better Solution?

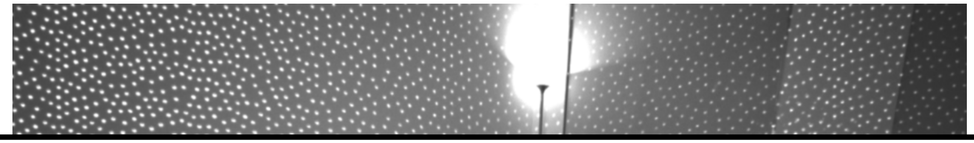
- ✅ Entropy Loss (Image Classification, Semantic Segmentation)
- ✅ L1/L2 Loss (Depth estimation, Colorization)
- ❌ Photometric Loss for Stereo Matching





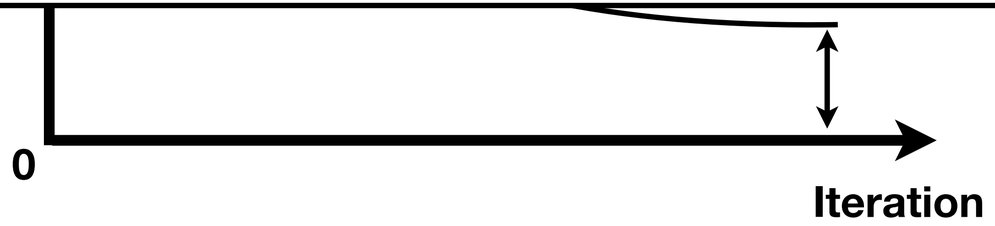
How to fix the problem?

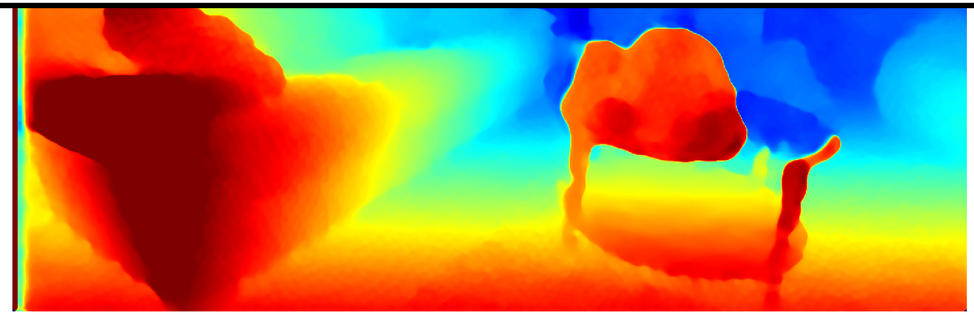
$$\text{Disparity} = \text{argmin}(\text{Photometric Loss})$$



Lower Loss → Better Solution?

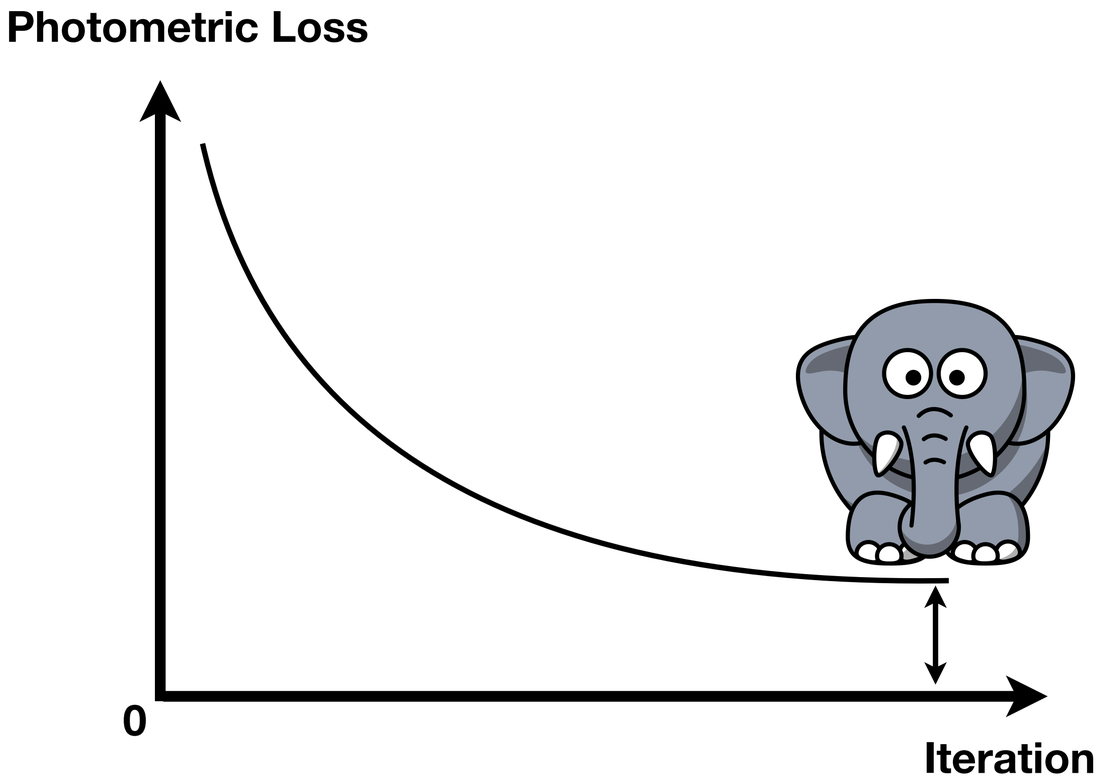
- Photometric Loss for Stereo Matching is not ideally zero!
  - Different Exposure
  - Occlusion
  - …





How to fix the problem?

$$\text{Disparity} = \text{argmin}(\text{Photometric Loss})$$



Unnecessary over optimization hurts performance!

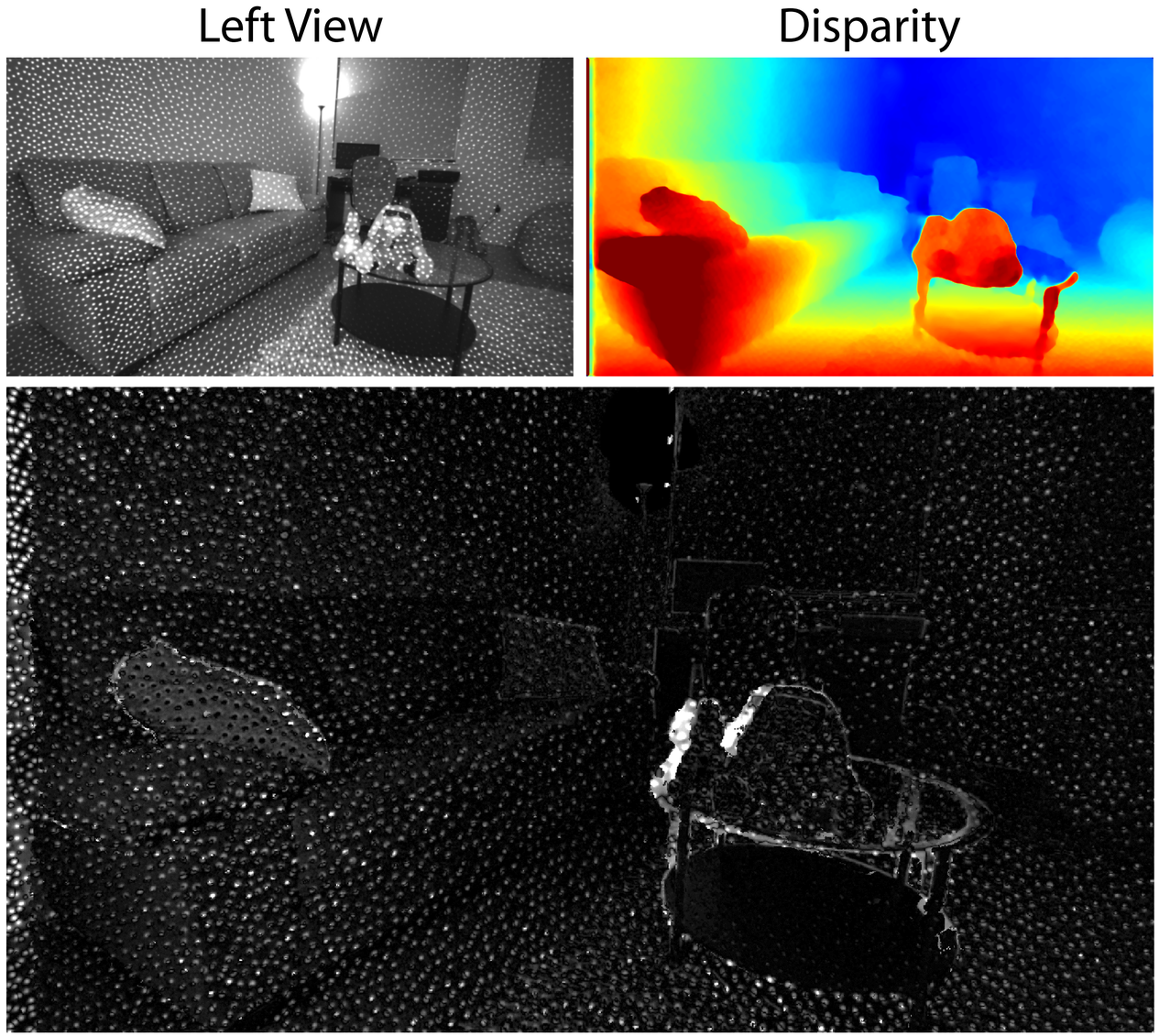
Fix the loss!

Improving Photometric Loss

1. Remove Unnecessary Dependence.

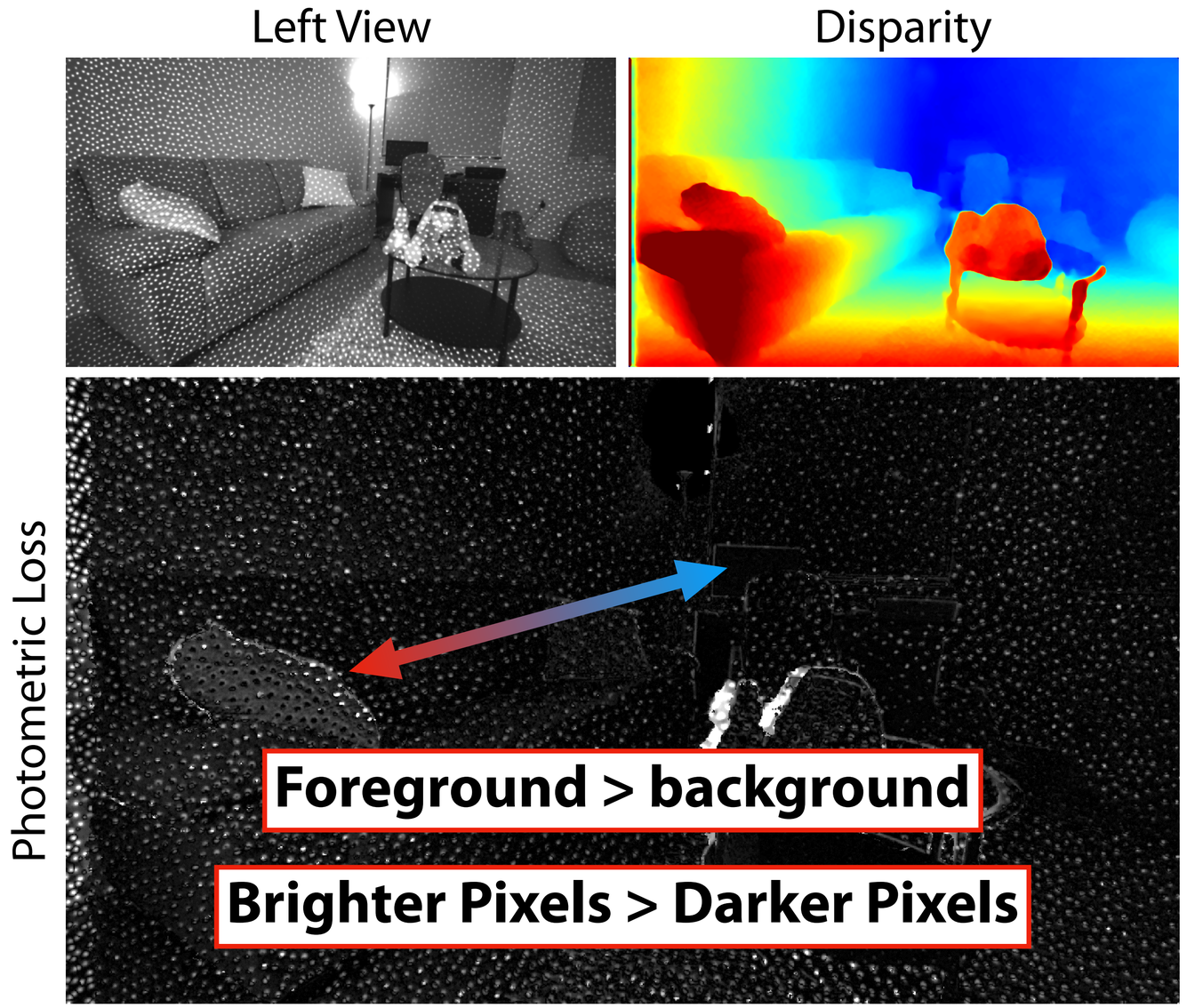
2. Remove Unexplainable Region.

3. Remove Bad Local Optima.



Improving Photometric Loss

1. Remove Unnecessary Dependence.



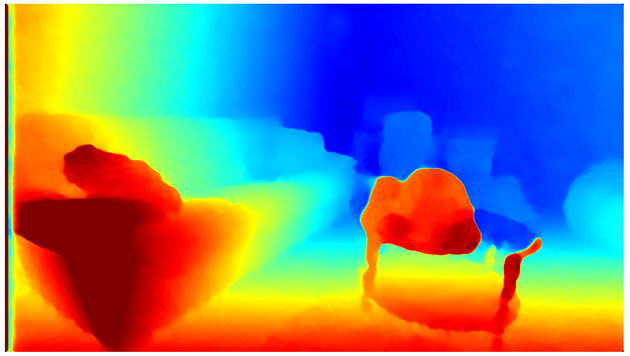
Remove Dependence

1. Remove Unnecessary Dependence.
   - Local Contrast Normalization

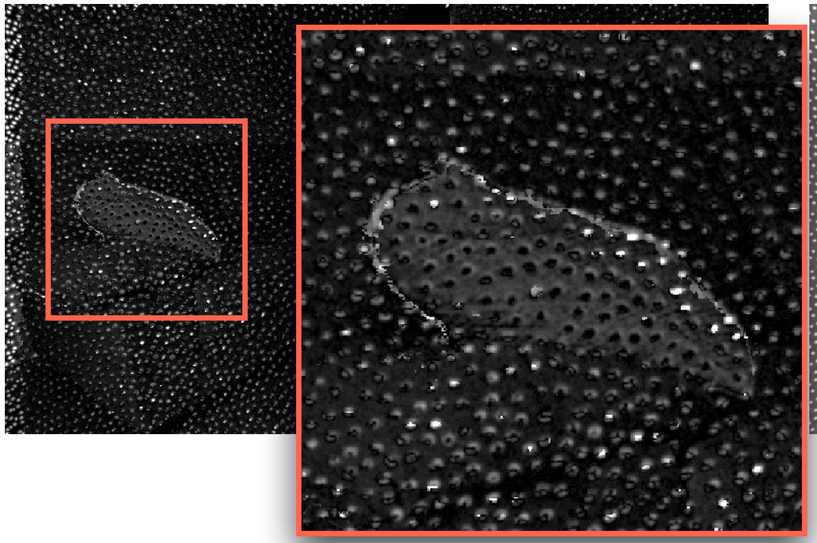
$$I_{LCN} = \frac{I - \mu}{\sigma + \eta}$$

Left View

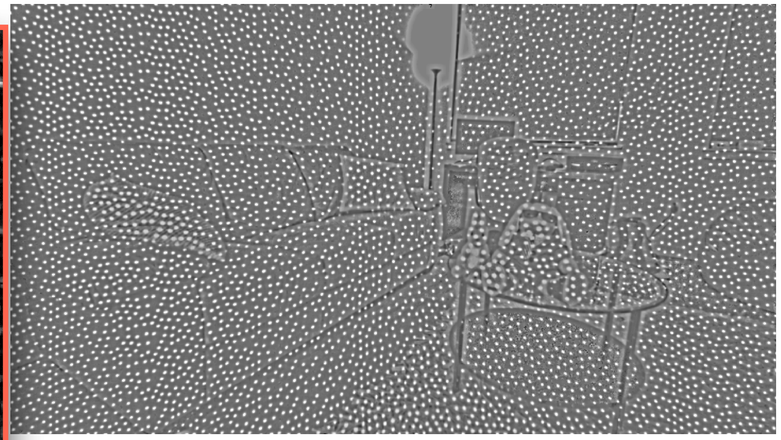
Disparity



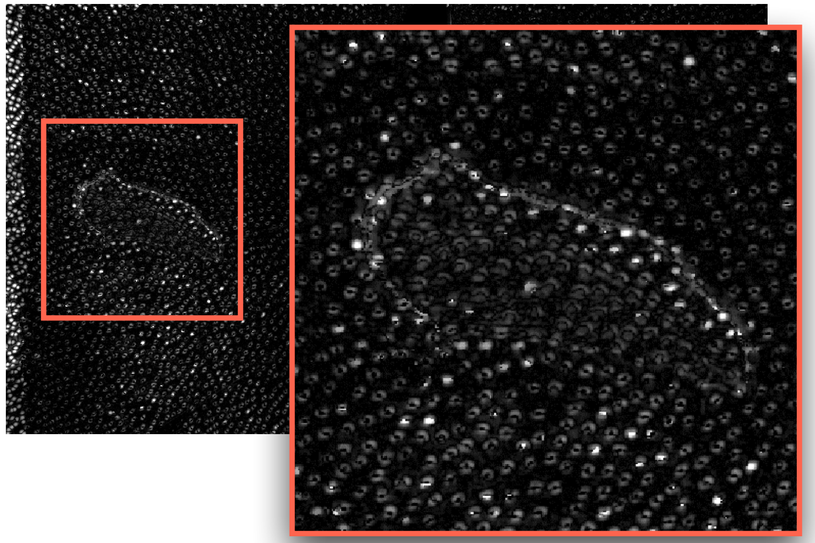
Loss on Raw IR



Local Contrast Normalization

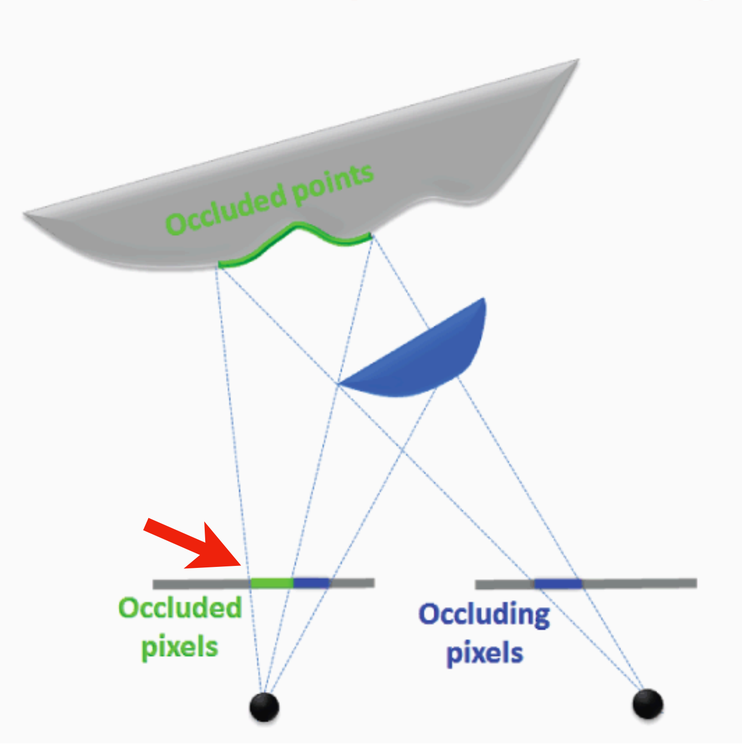


Loss on Normalized IR



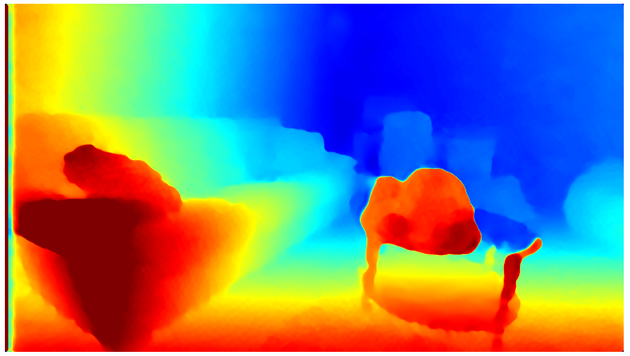
Improving Photometric Loss

1. Remove Unnecessary Dependence.
    - Local Contrast Normalization
2. Remove Unexplainable Region.



Left View

Disparity



Photometric Loss

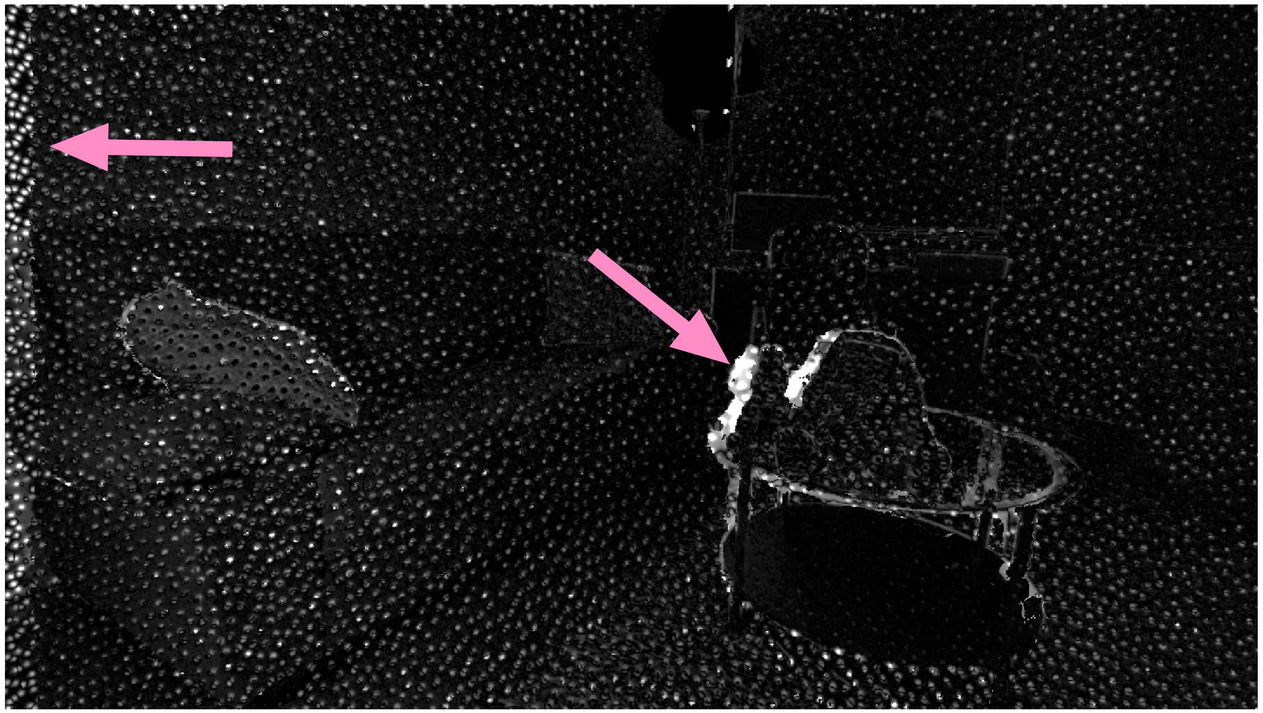
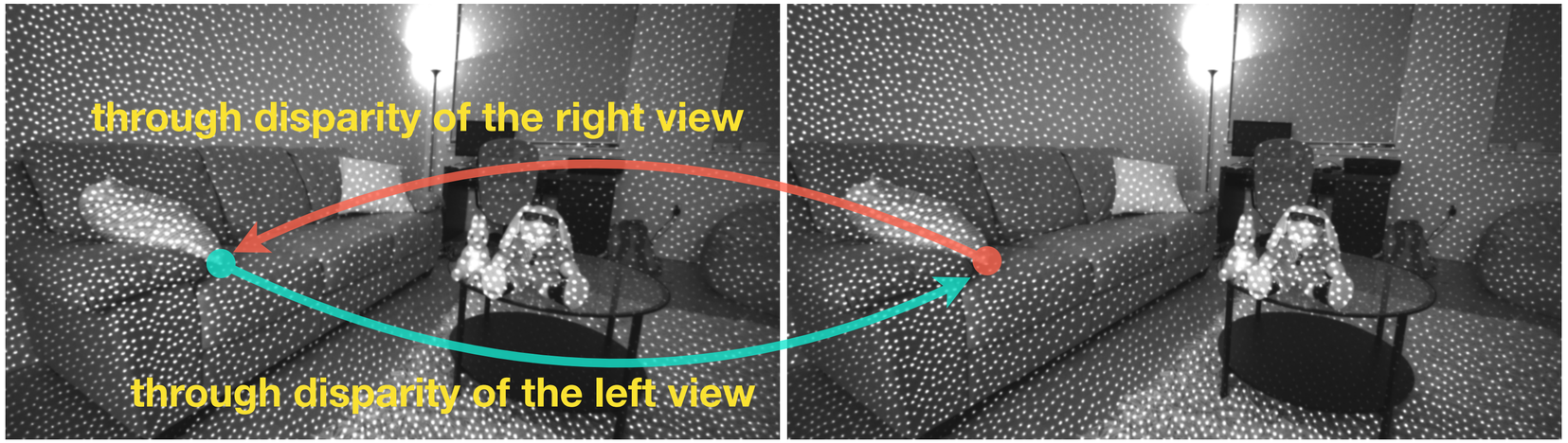


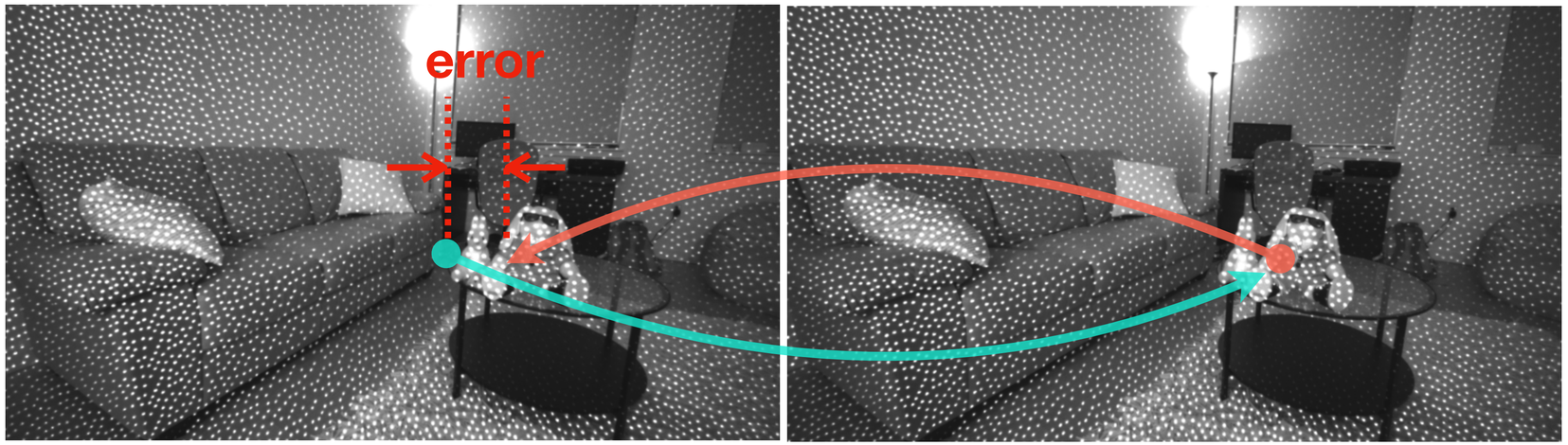
Photo from: Basha et.al., Stereo Sean Carving a Geometrically Consistent Approach, PAMI 2013, vol. 35

How to find occlusion?

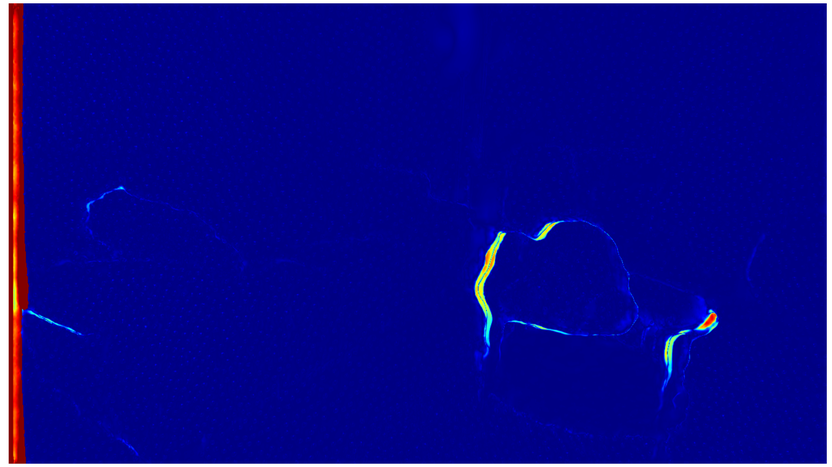
Left View

Right View



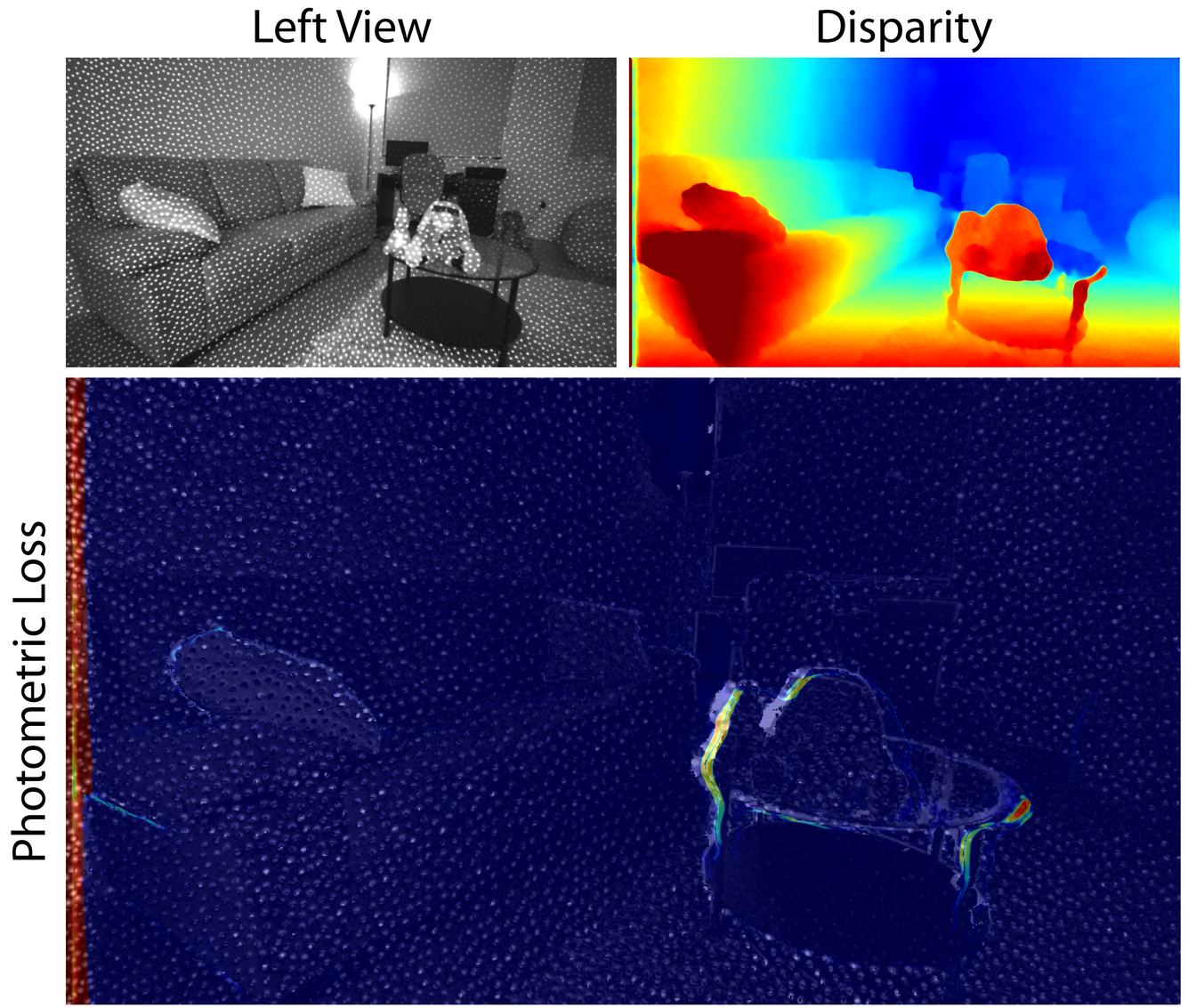


Remove the pixel loss if error passes threshold.



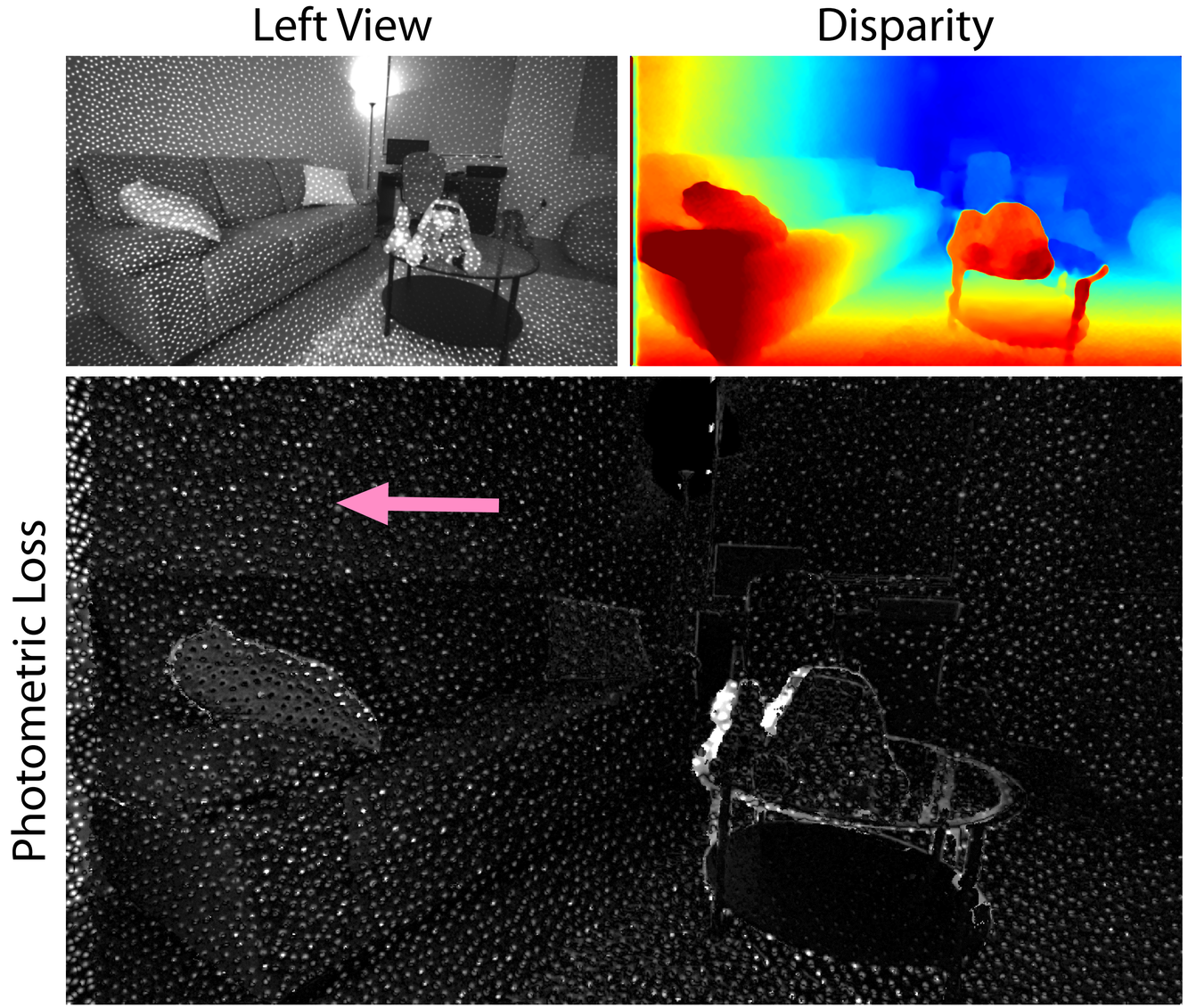
Improving Photometric Loss

1. Remove Unnecessary Dependence.
    - Local Contrast Normalization
2. Remove Unexplainable Region.
    - Loop-based Check



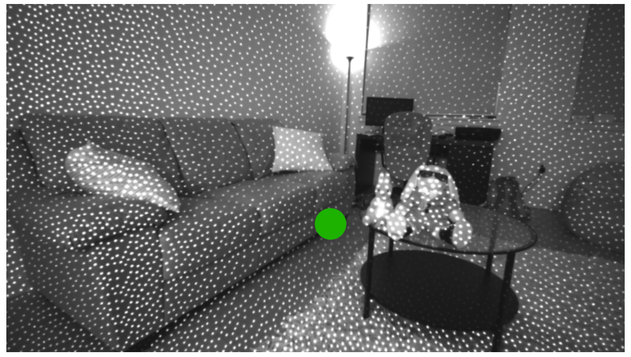
Improving Photometric Loss

1. Remove Unnecessary Dependence.
    - Local Contrast Normalization
2. Remove Unexplainable Region.
    - Loop-based Check
3. Remove Bad Local Optima.

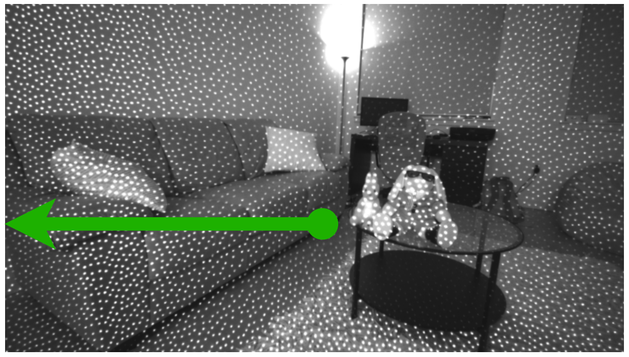


Remove Bad Optima

Left View



Right View



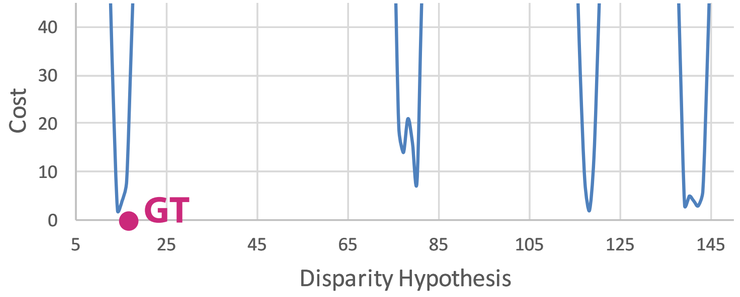
Adaptive Support Weight

$$C_{i,j} = \left\| \sigma_{i,j} (I_{LCN\ i,j} - \hat{I}_{LCN\ i,j}) \right\|_1$$

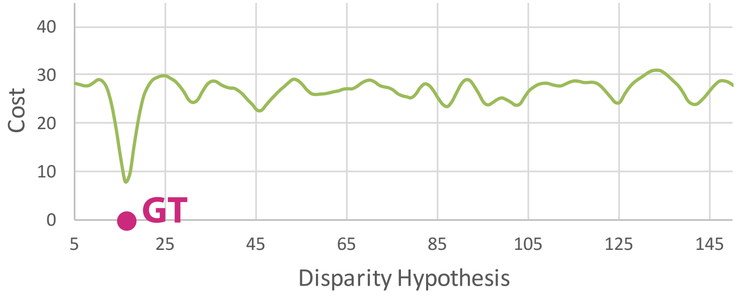
$$\hat{C}_{i,j} = \frac{\sum_{x=i-k}^{i+k-1} \sum_{y=j-k}^{j+k-1} \omega_{x,y} C_{x,y}}{\sum_{x=i-k}^{i+k-1} \sum_{y=j-k}^{j+k-1} \omega_{x,y}}$$

$$\omega_{x,y} = \exp\left(-\frac{\left| I_{i,j} - I_{x,y} \right|}{\sigma_\omega}\right)$$

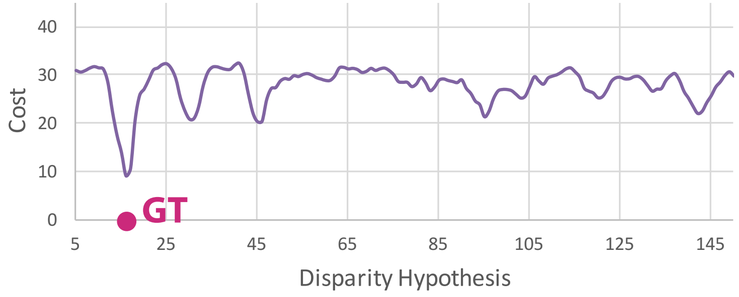
1x1 Window

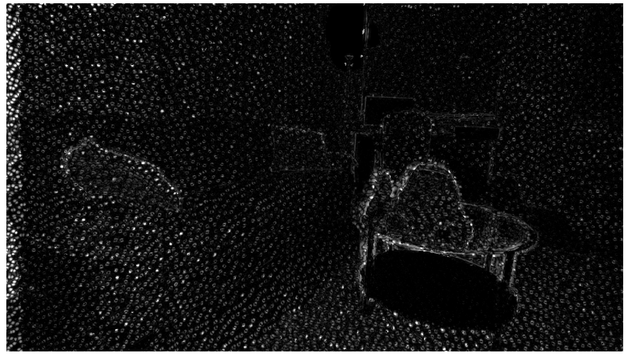


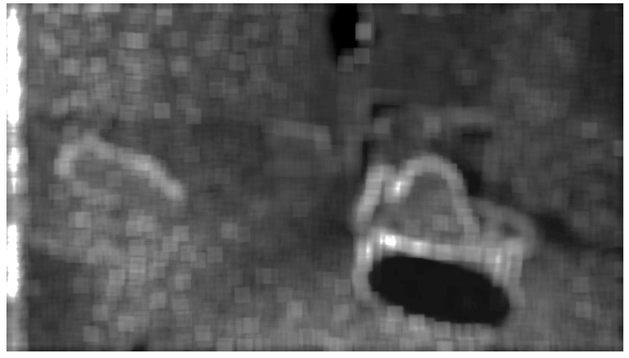
32x32 Window

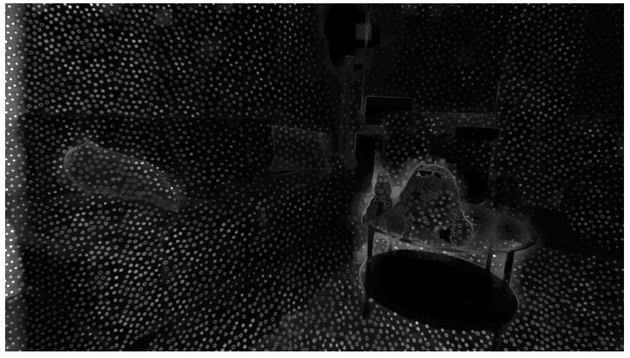


32x32 Window with ASW







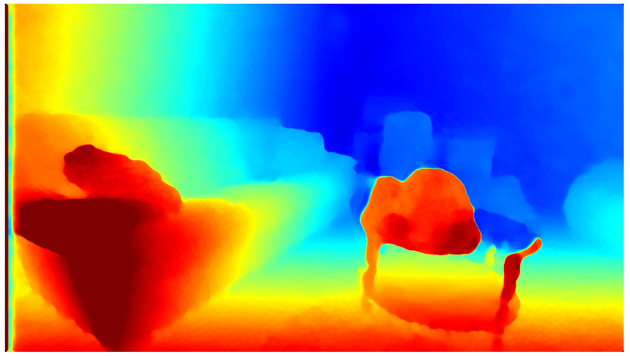


Improving Photometric Loss

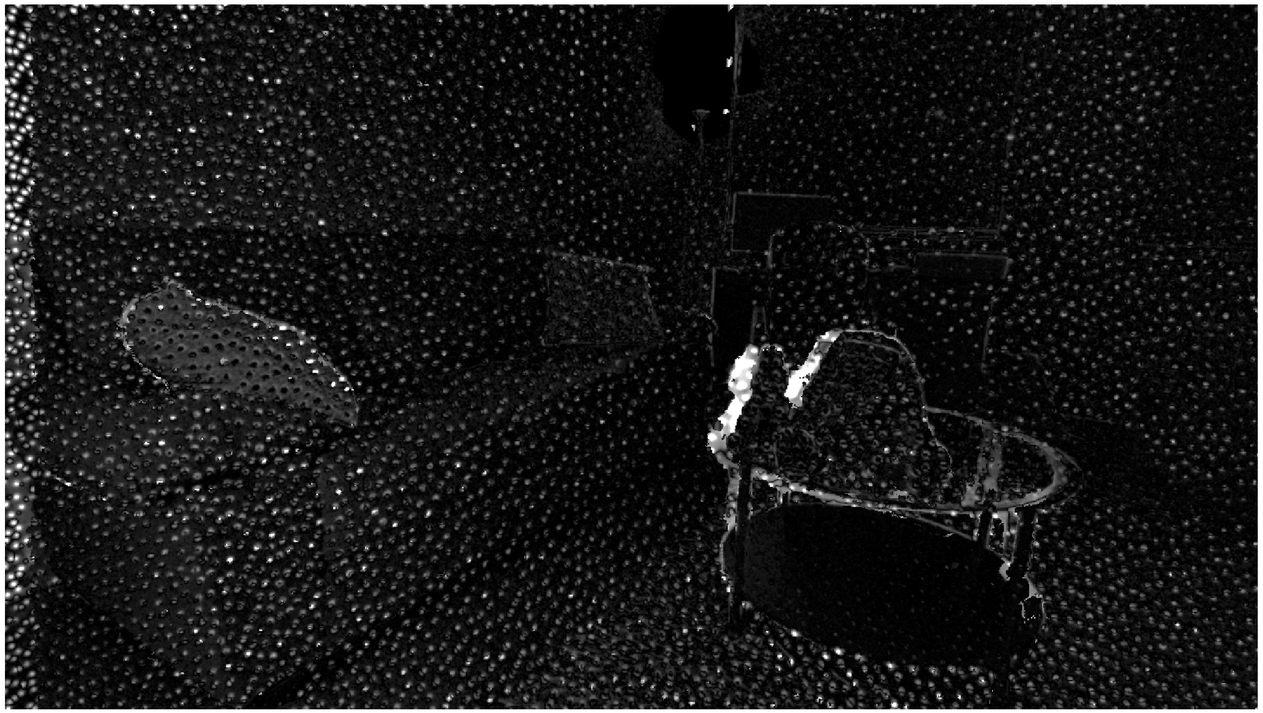
1. Remove Unnecessary Dependence.
    - Local Contrast Normalization
2. Remove Unexplainable Region.
    - Loop-based Check
3. Remove Bad Local Optima.
    - Window aggregation with ASW

Left View

Disparity



Photometric Loss



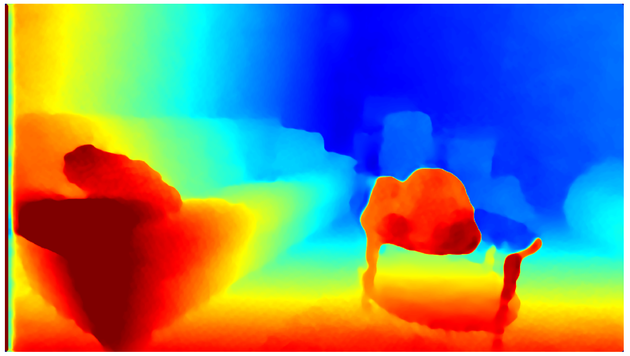
Improving Photometric Loss

1. Remove Unnecessary Dependence.
    - Local Contrast Normalization
2. Remove Unexplainable Region.
    - Loop-based Check
3. Remove Bad Local Optima.
    - Window aggregation with ASW

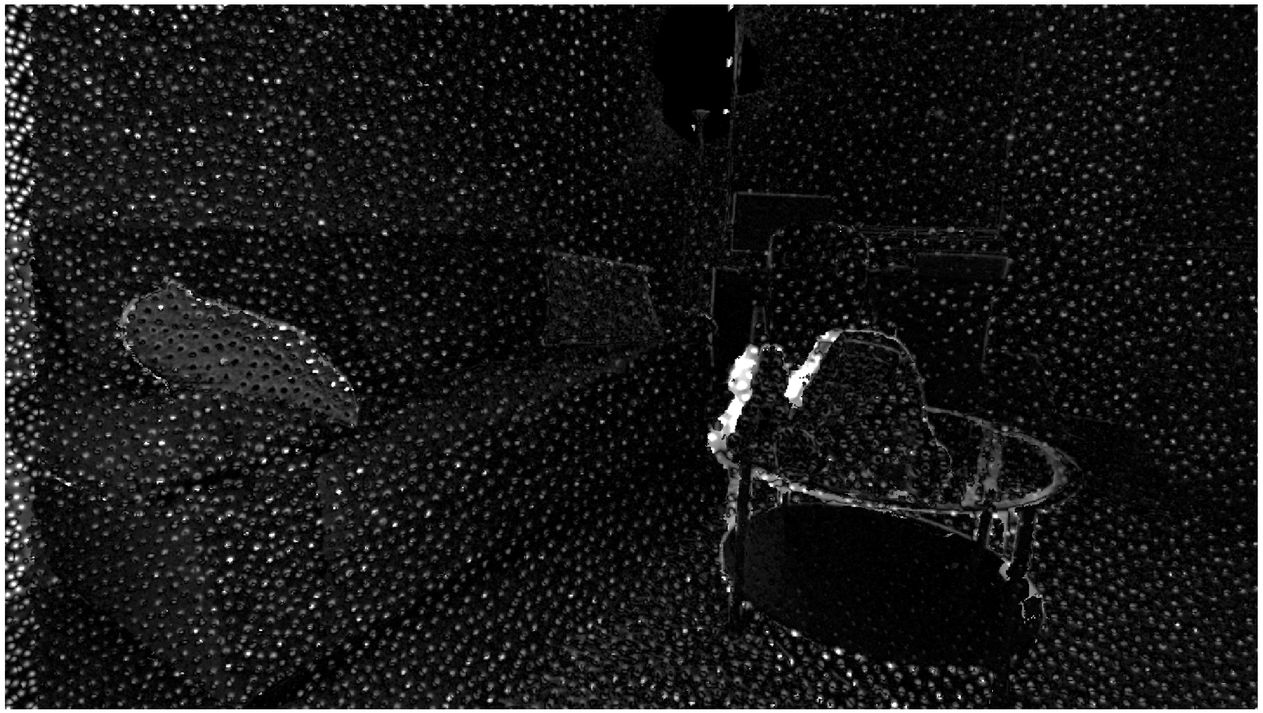
$$L = \sum_{p \notin Inv} \hat{C}_p$$

Left View

Disparity

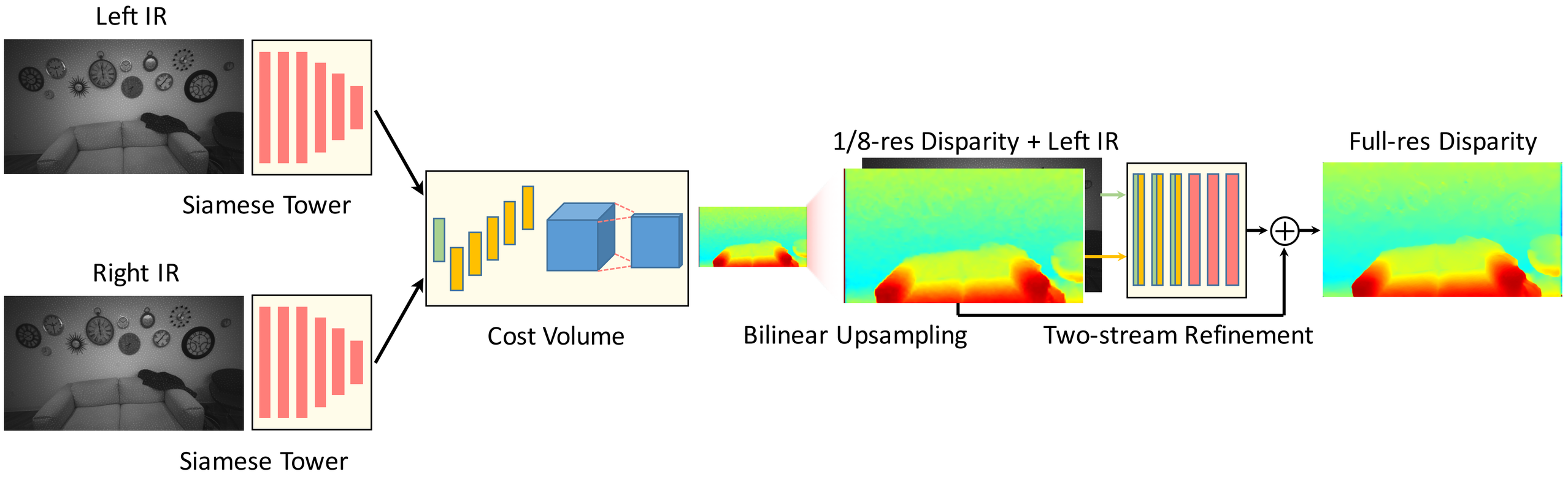


Photometric Loss



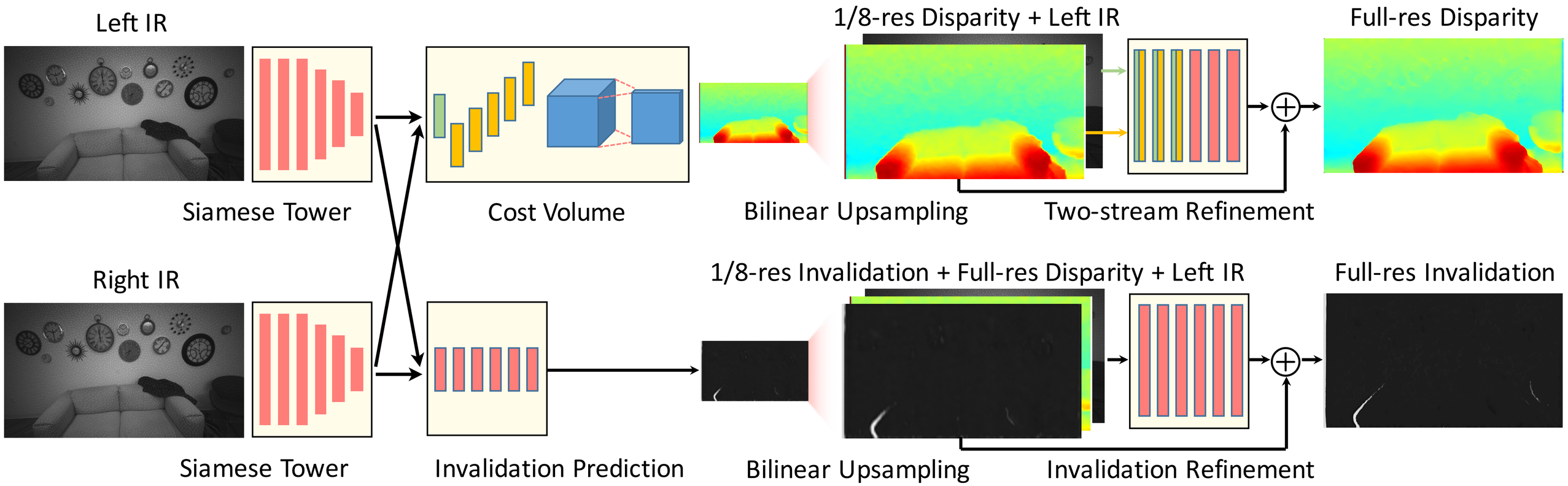
Network Architecture

Network Architecture



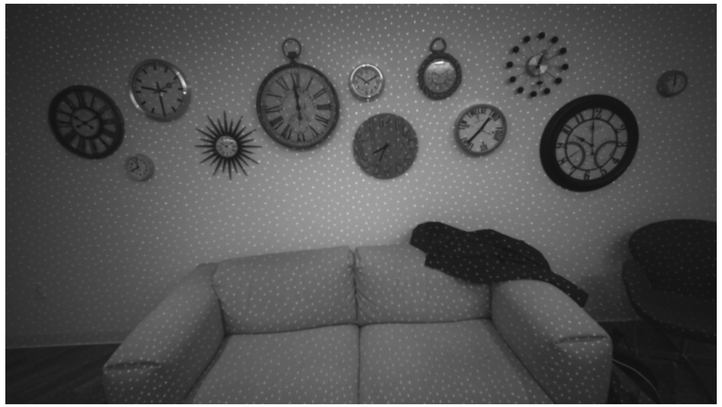
Khamis et.al., StereoNet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction, ECCV2018

Network Architecture



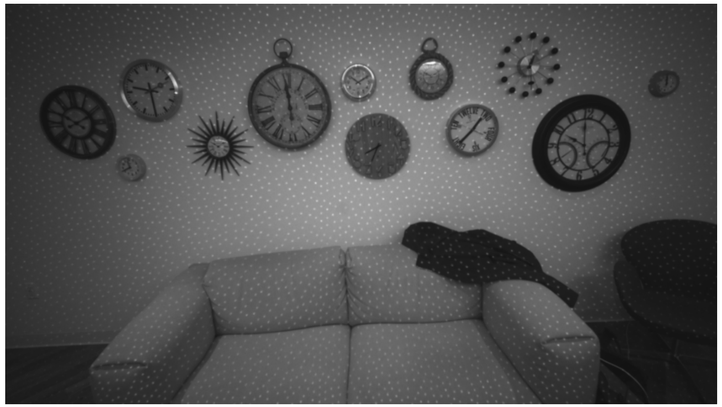
Experiments

Experiments



Left IR Image

Intel RealSense D435

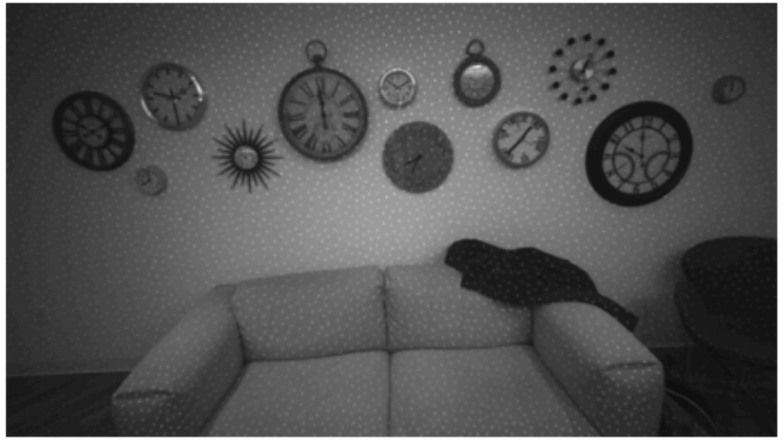


Right IR Image

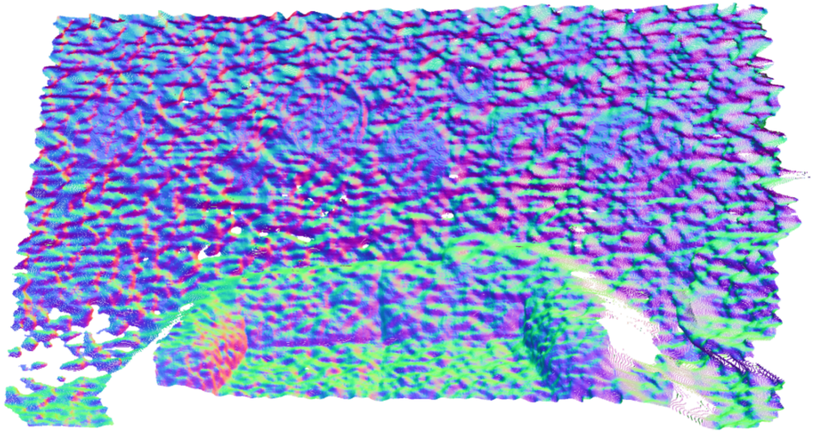
Color Image

Experiments

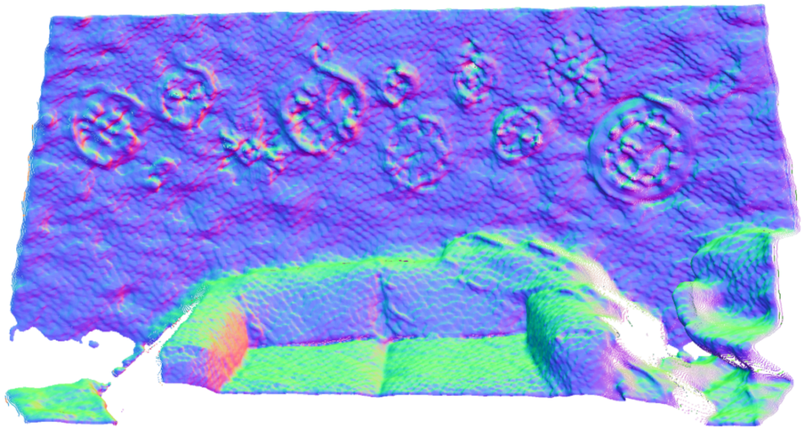
Intel RealSense D435



Left View



Intel RealSense D435



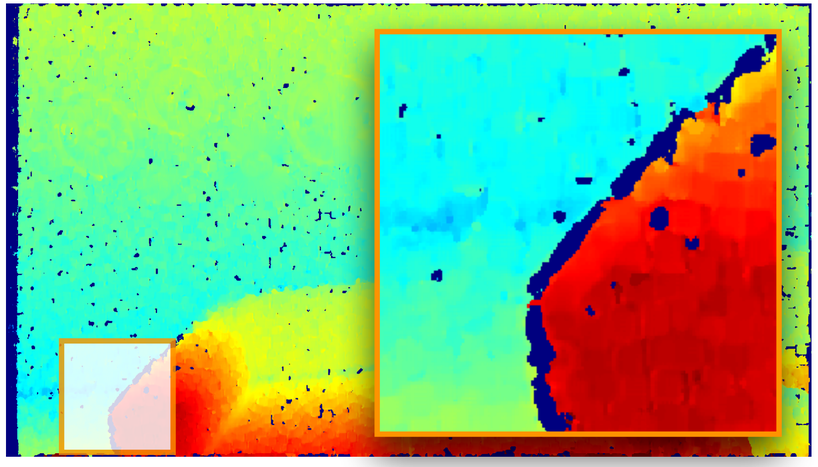
ActiveStereoNet

Color Image

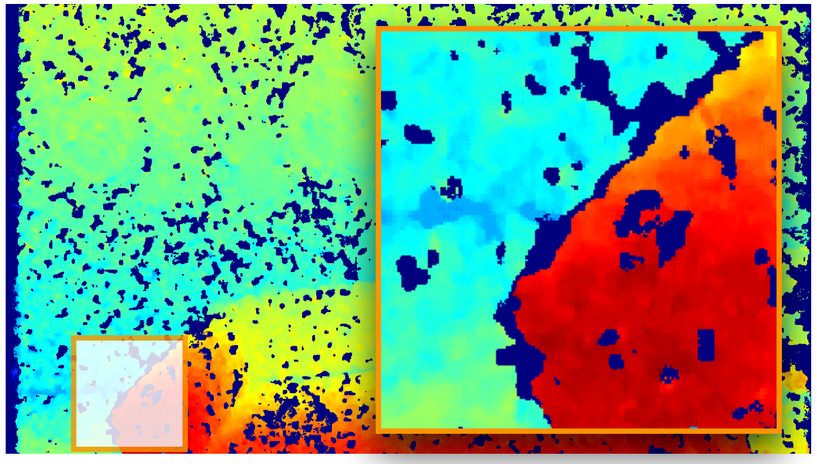
Disparity Qualitative Result

IR Left Input

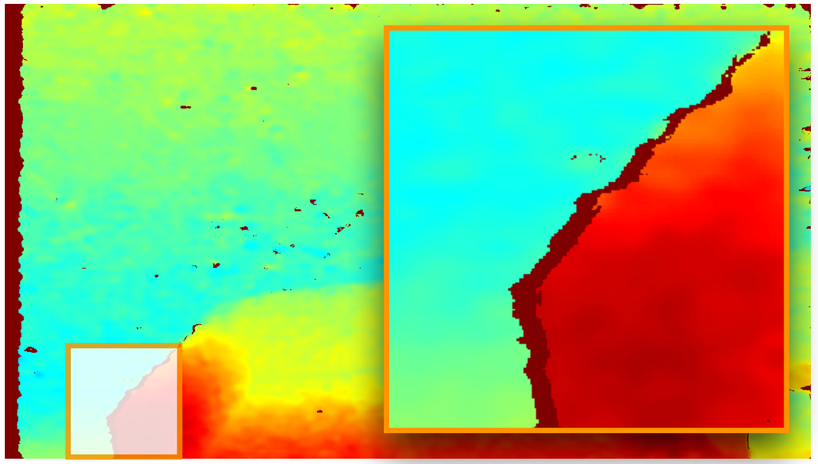
PatchMatch Stereo



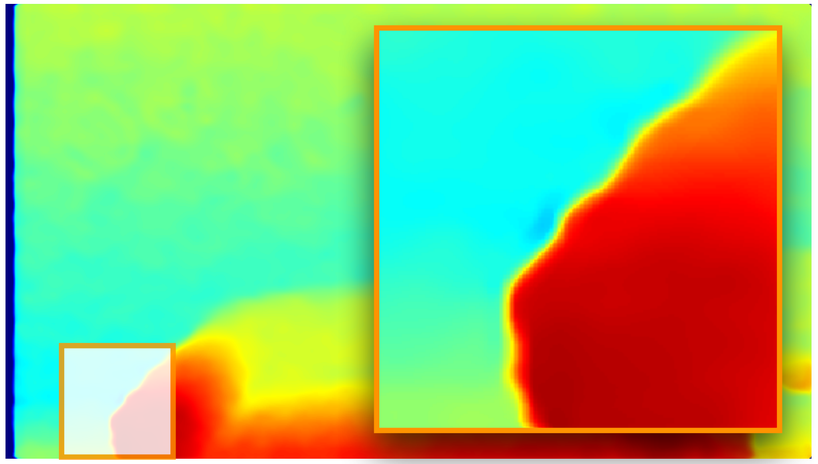
HashMatch Stereo



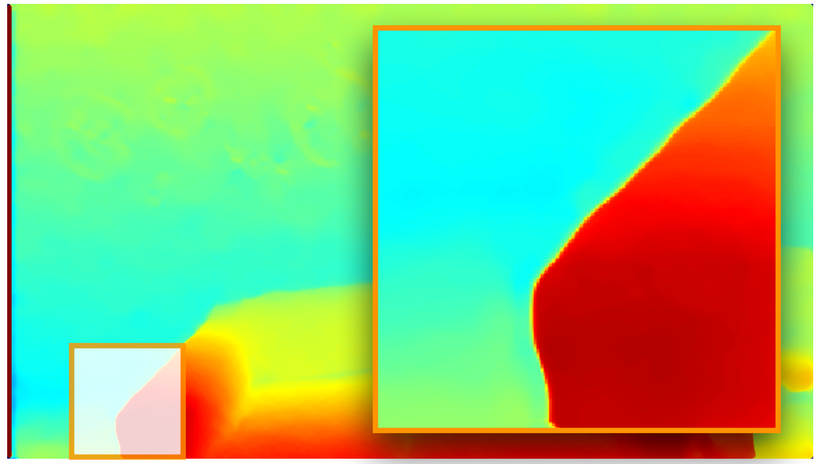
Sensor Output



ASN Semi Supervised (ours)

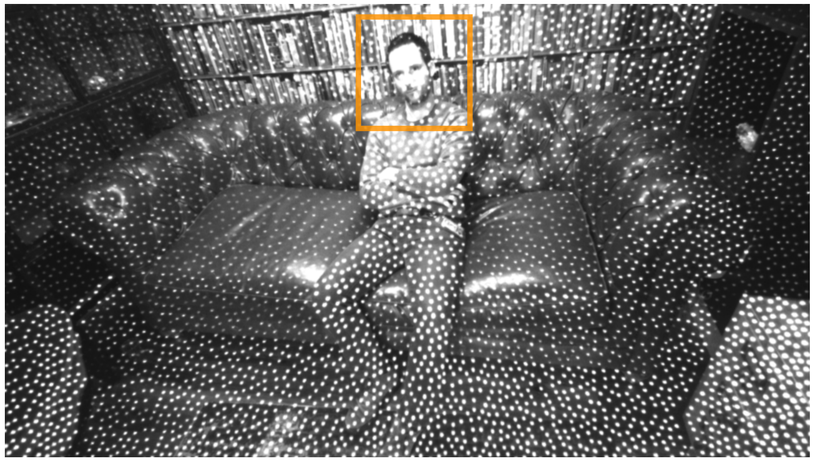


ASN Self-Supervised (ours)

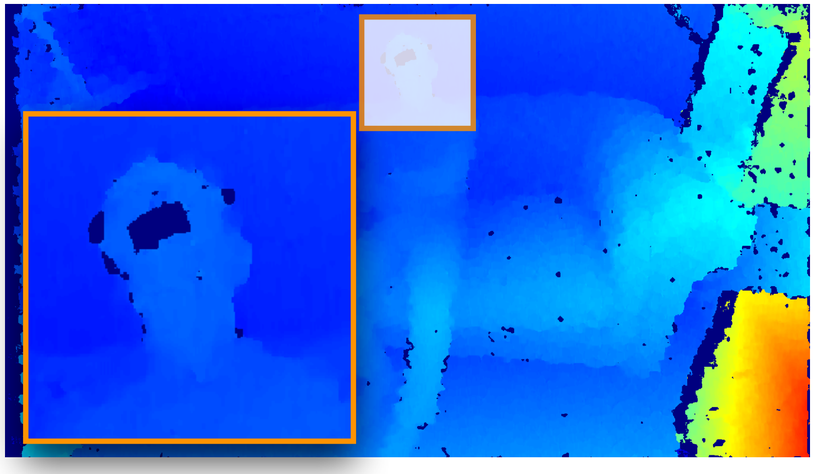


Disparity Qualitative Result

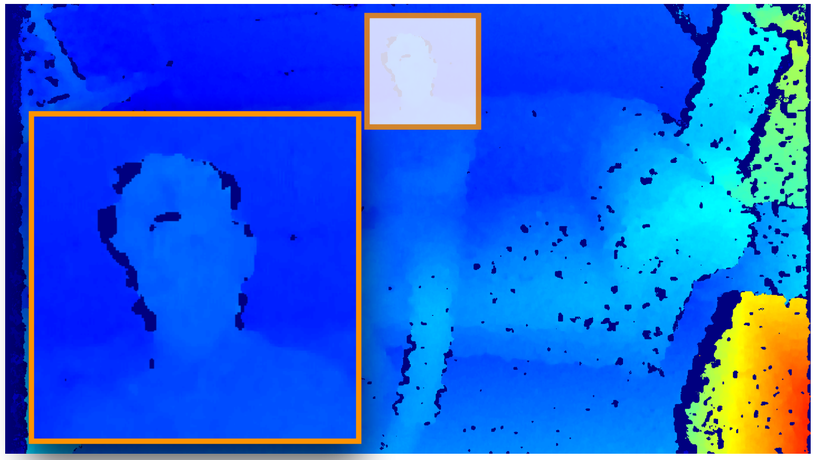
IR Left Input



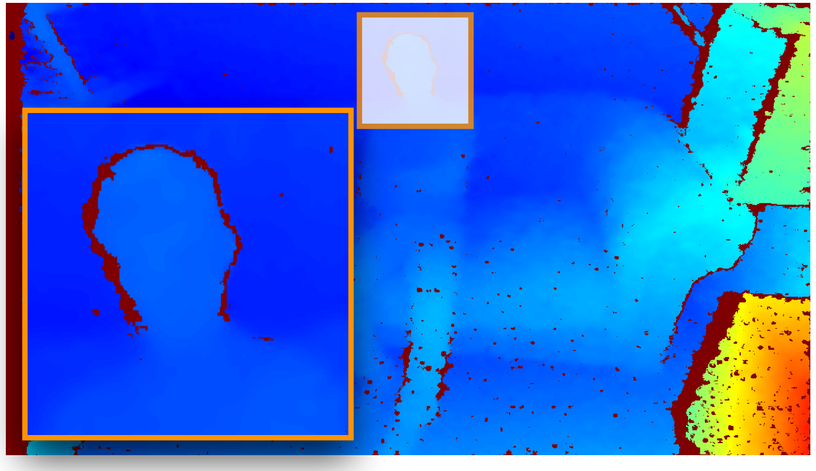
PatchMatch Stereo



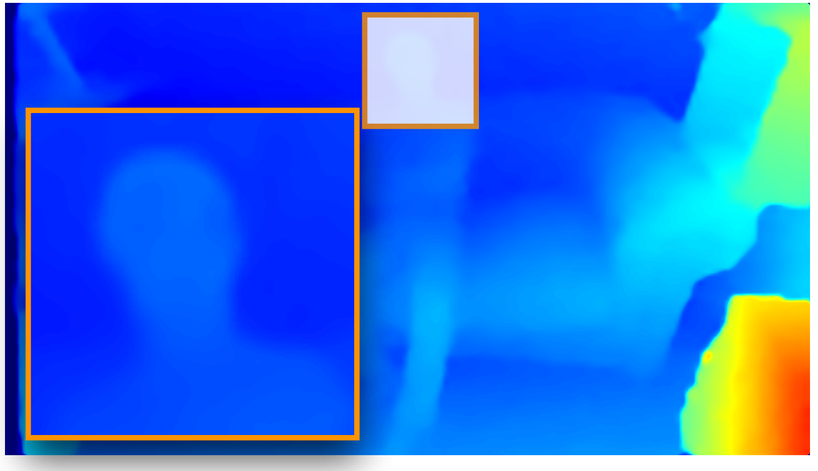
HashMatch Stereo



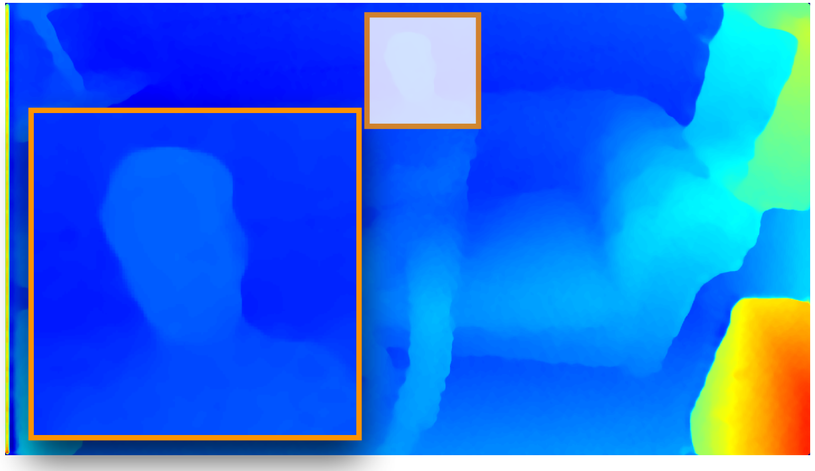
Sensor Output



ASN Semi Supervised (ours)

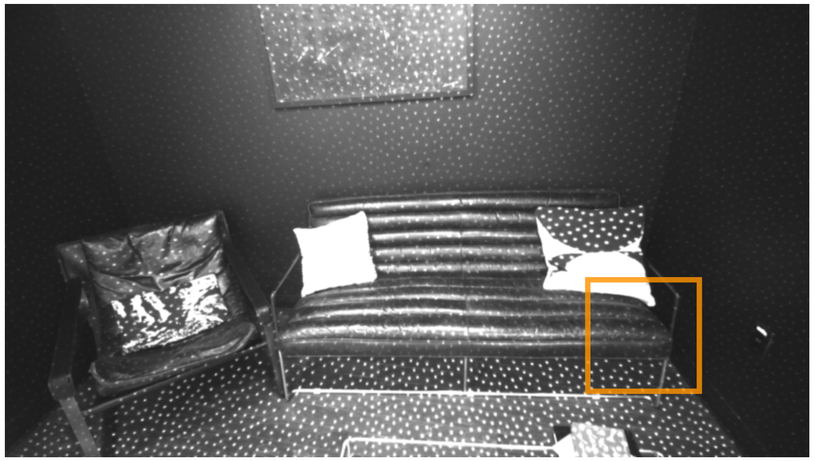


ASN Self-Supervised (ours)

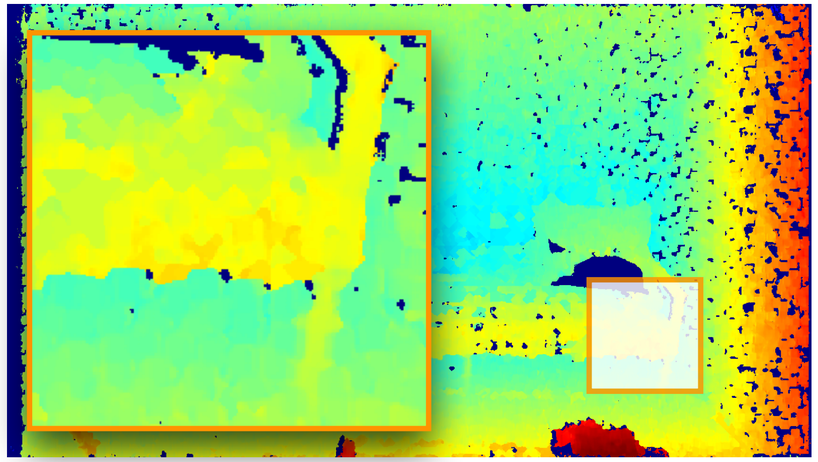


Disparity Qualitative Result

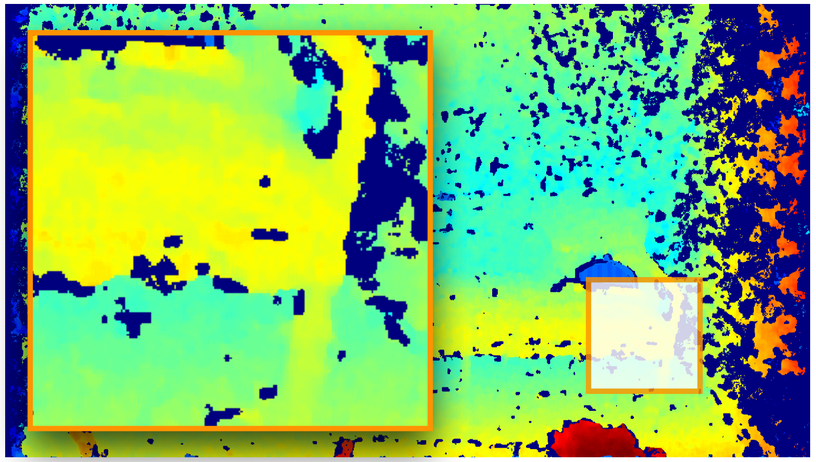
IR Left Input



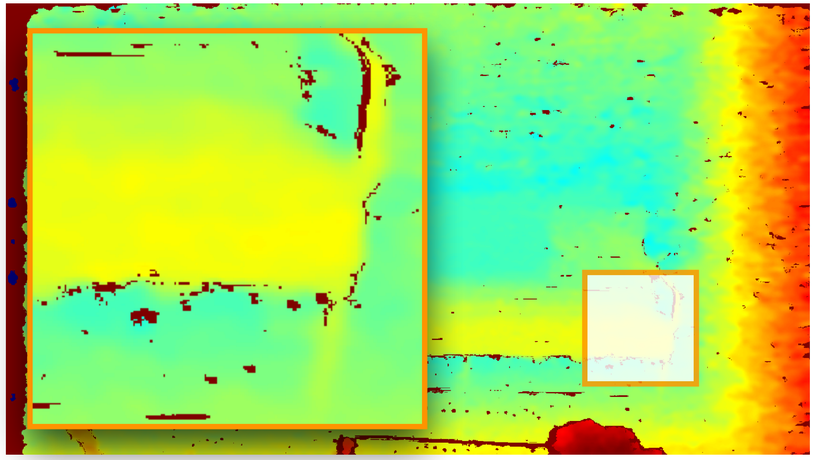
PatchMatch Stereo



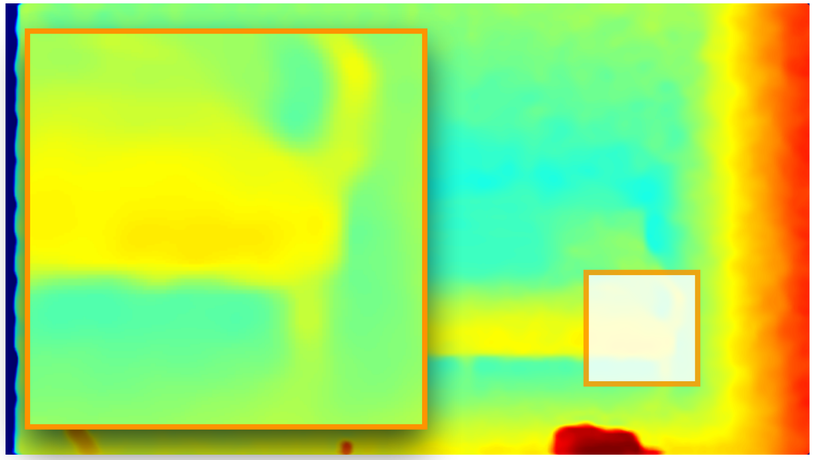
HashMatch Stereo



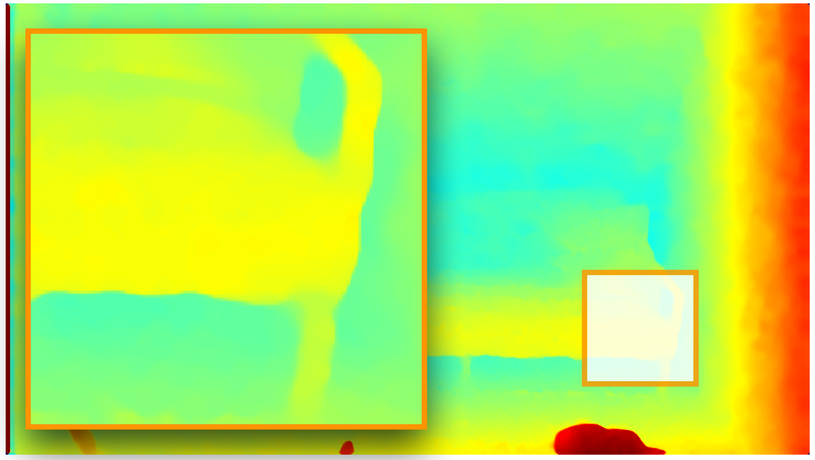
Sensor Output



ASN Semi Supervised (ours)



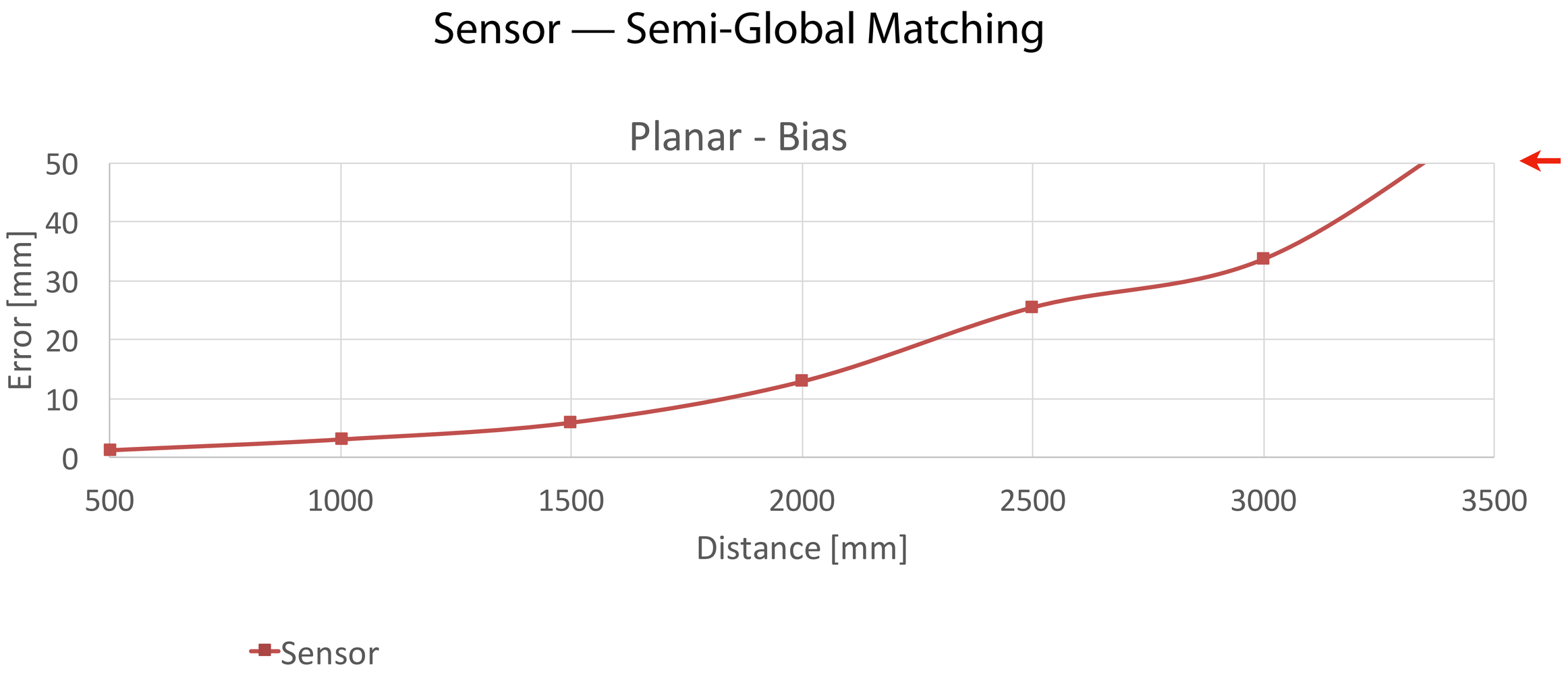
ASN Self-Supervised (ours)



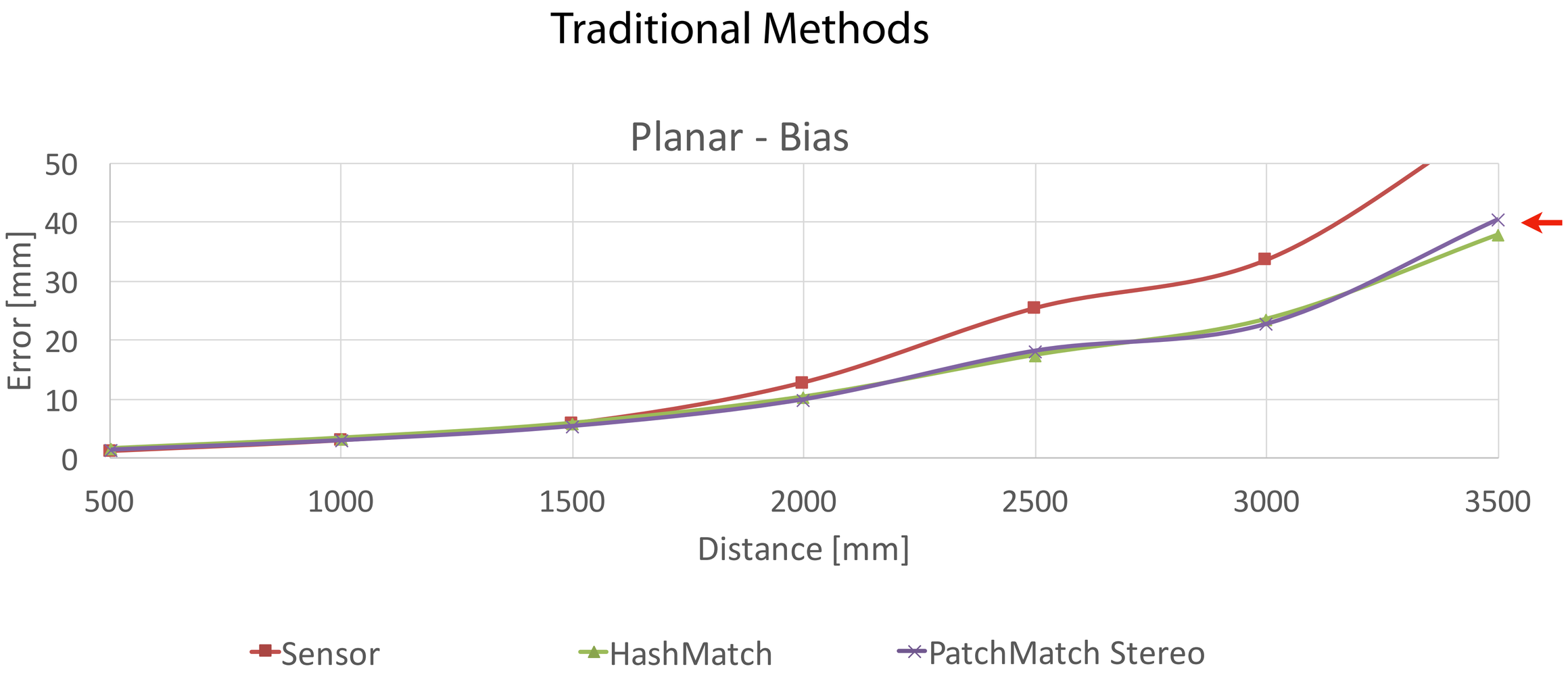
Disparity Quantitative Result

Fit plane on planar-wise scene as ground truth.

Disparity Quantitative Result

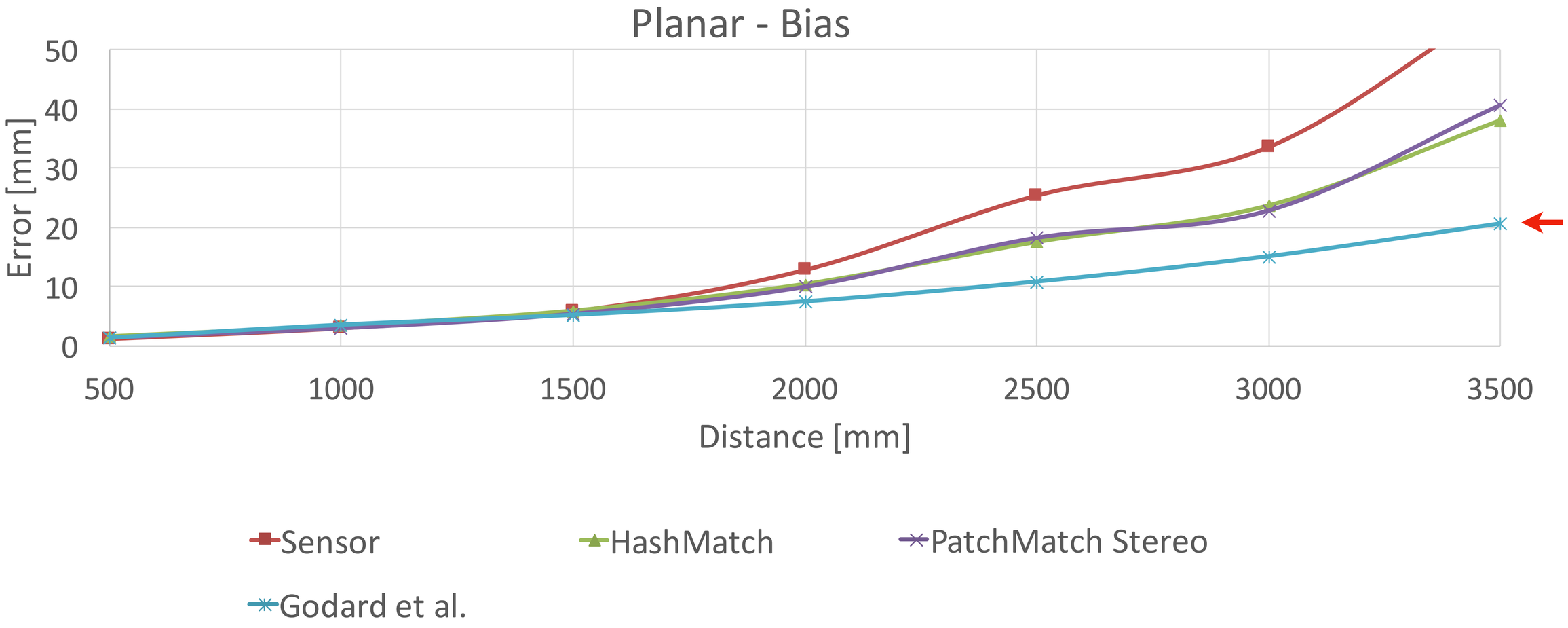


Disparity Quantitative Result

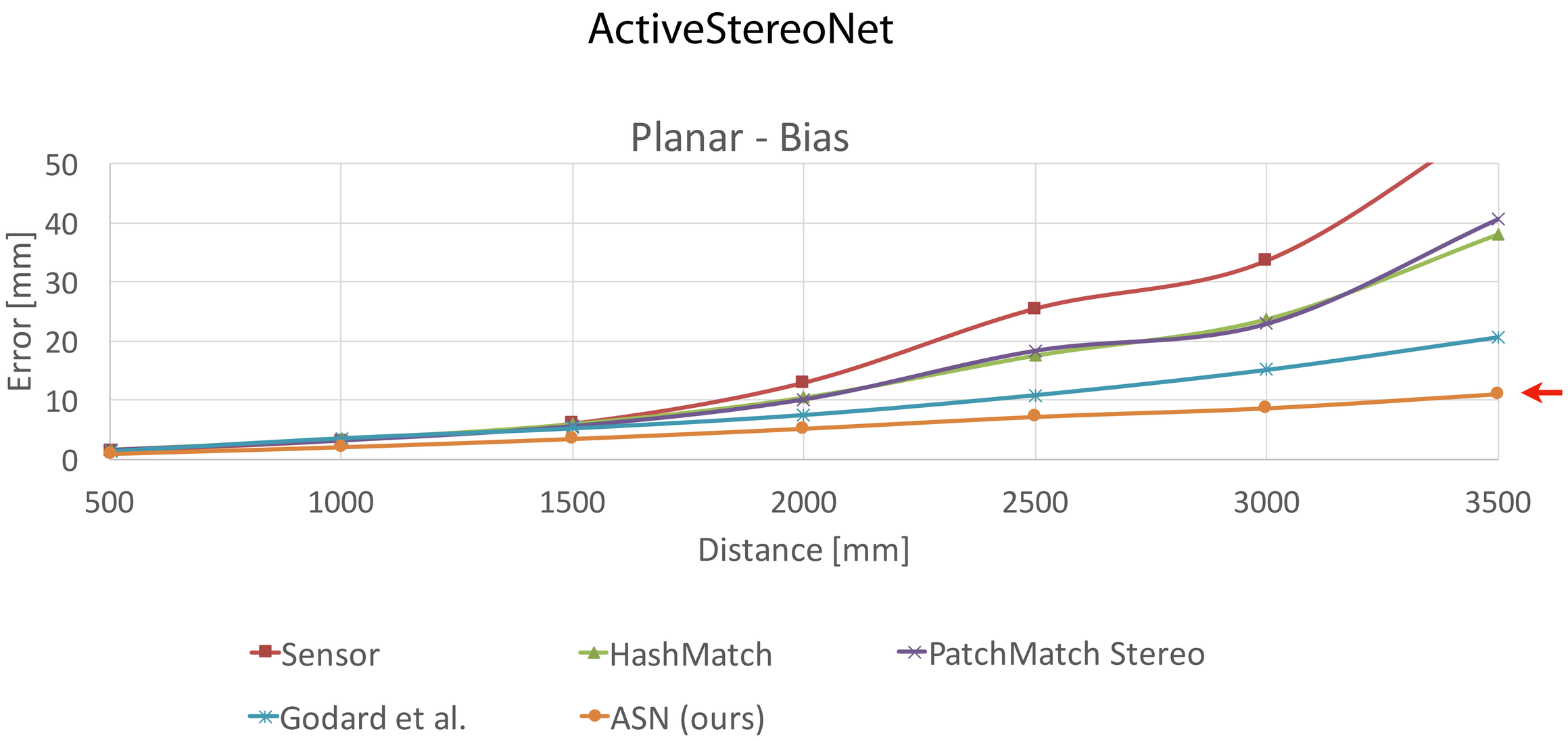


Disparity Quantitative Result

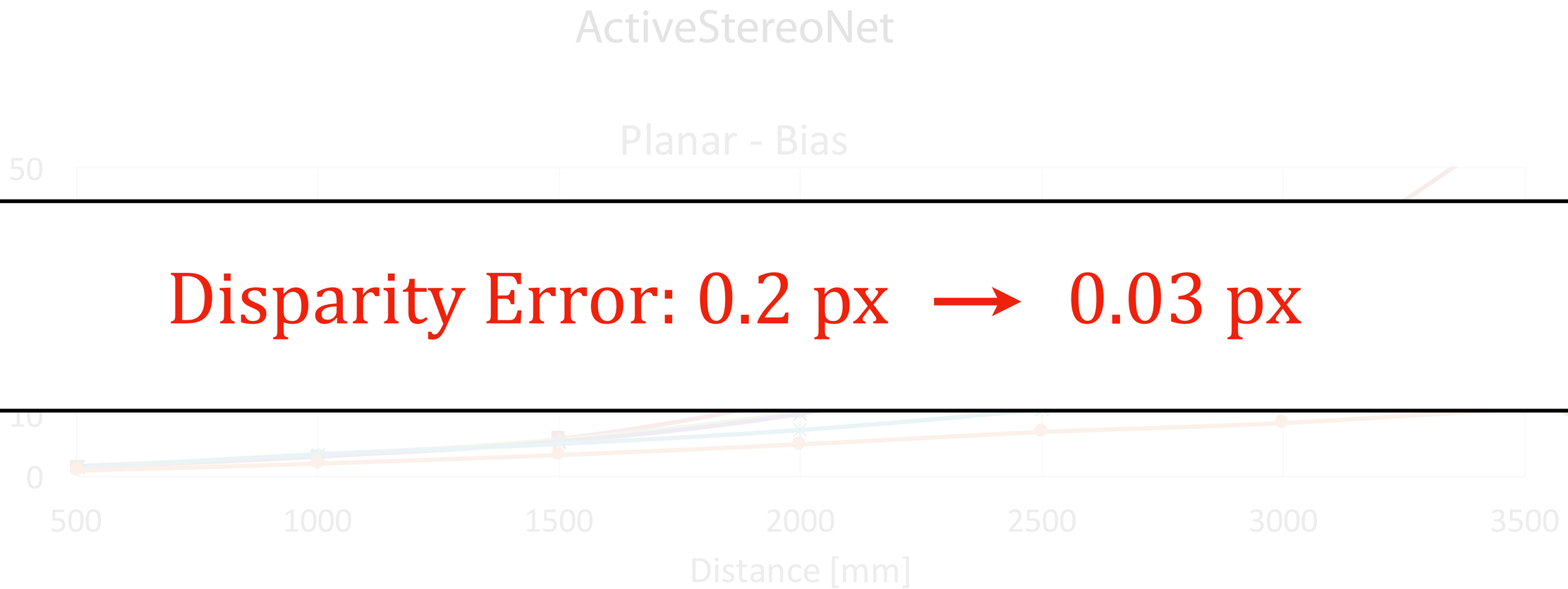
Previous Self-Supervised Method



Disparity Quantitative Result



Disparity Quantitative Result



Disparity Error: 0.2 px → 0.03 px

Ablation Study

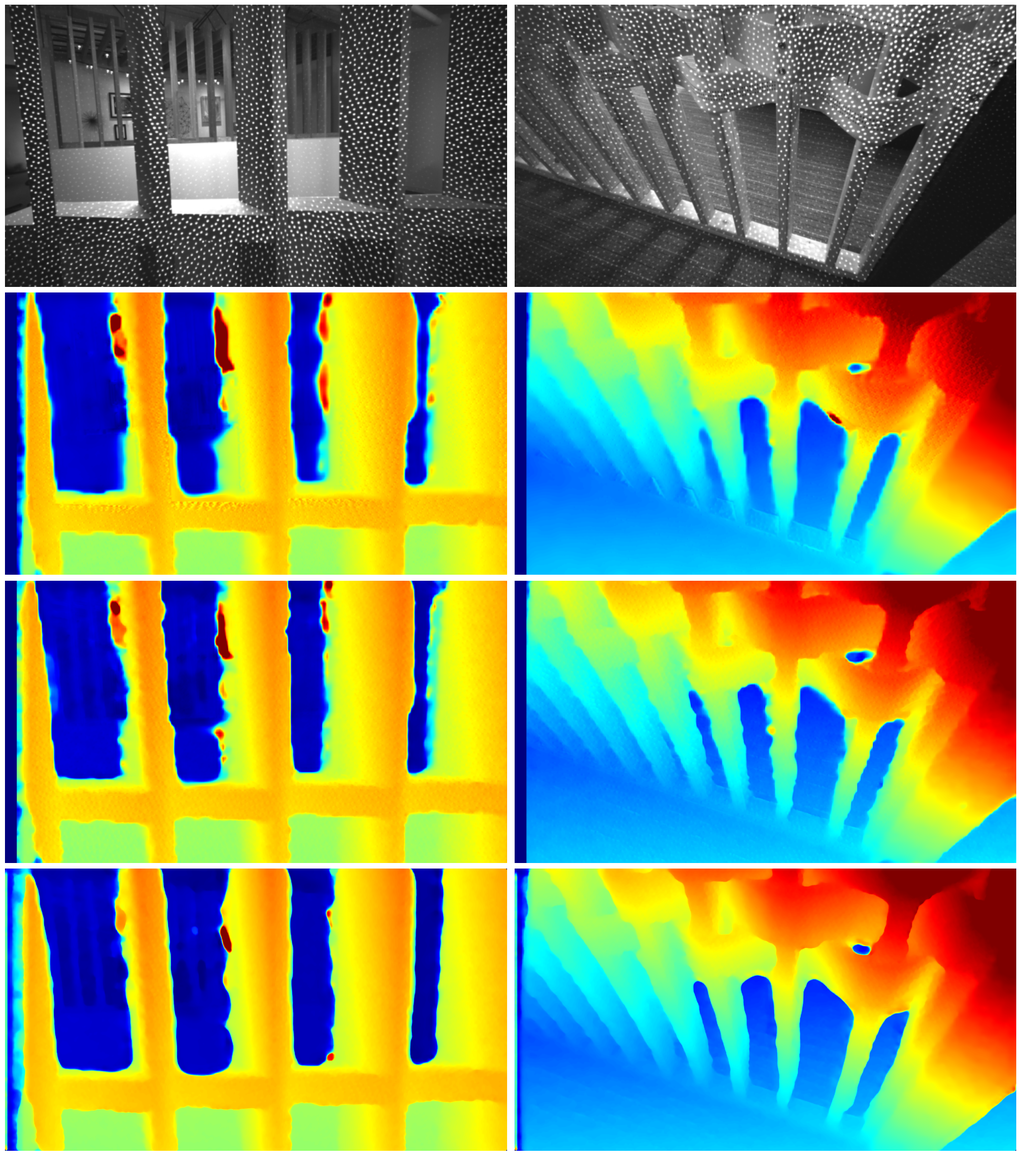
Does LCN Matter?

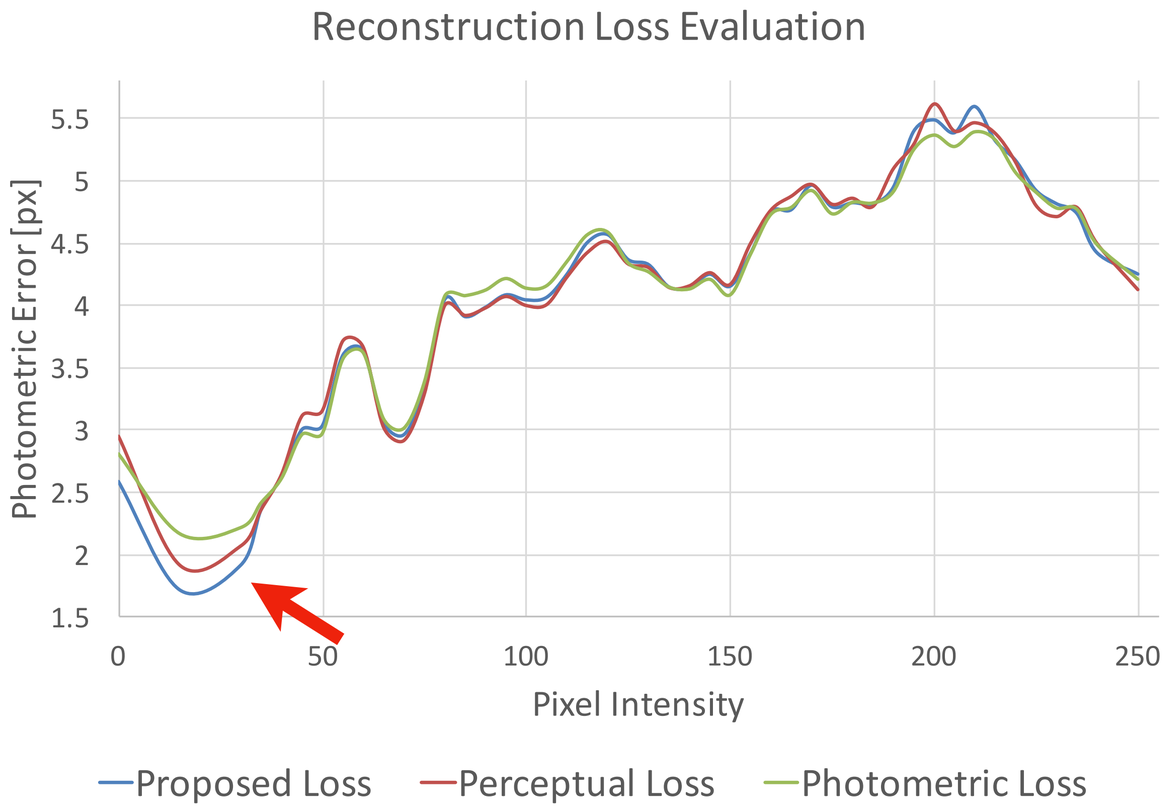
IR Left Input

Traditional Photometric Loss

Perceptual Loss

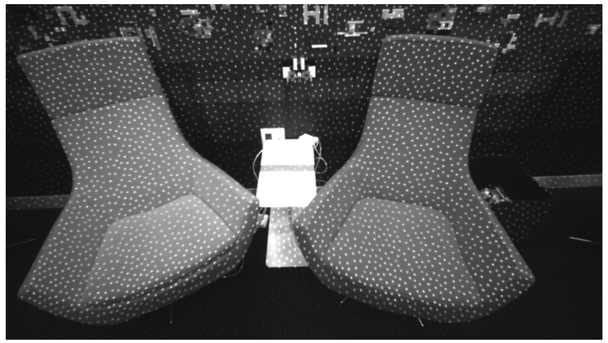
LCN Loss



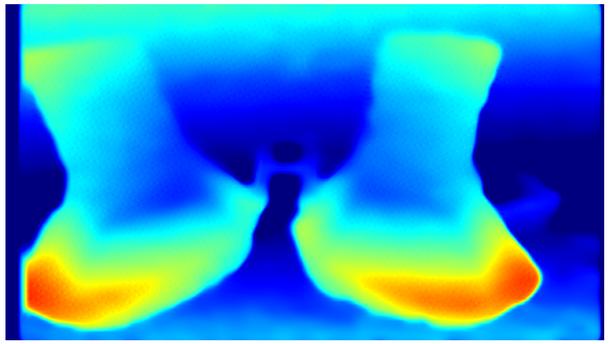
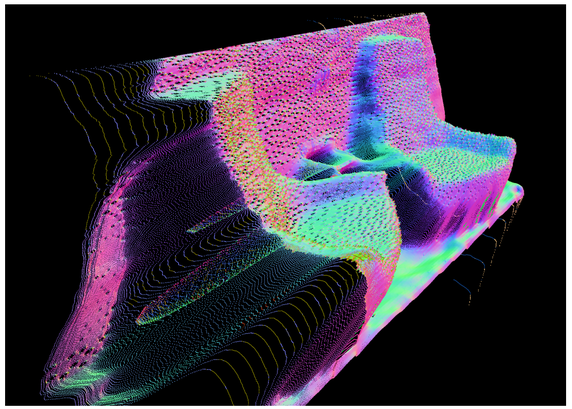


LCN: Local Contrast Normalization

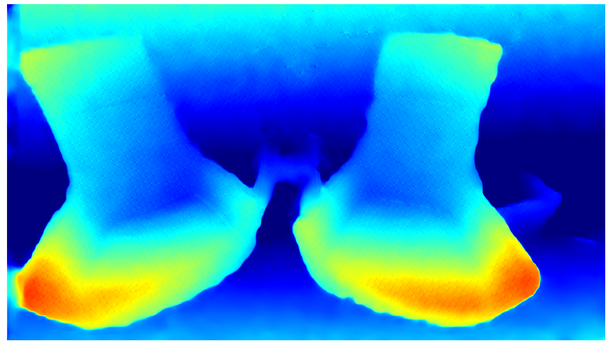
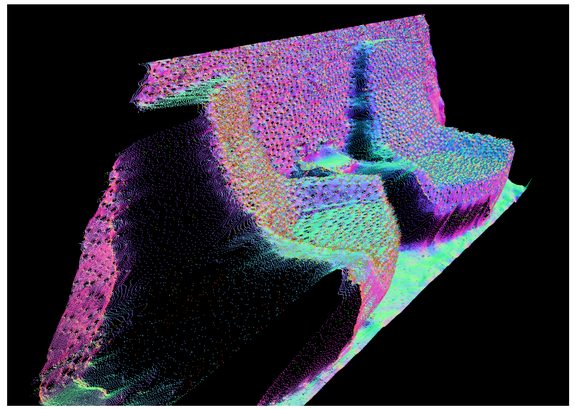
Does Invalidation Matter?



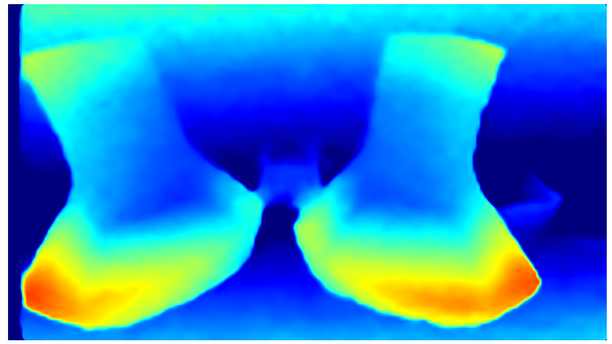
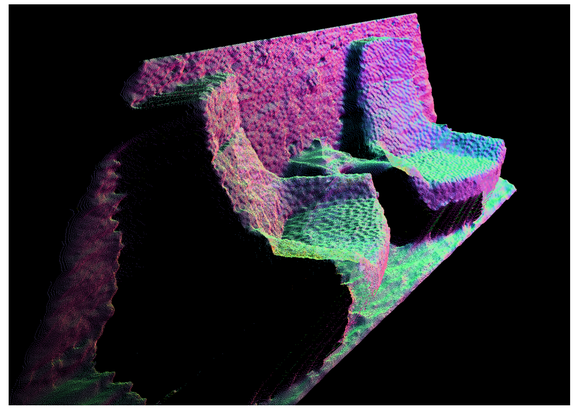

Godard et al.

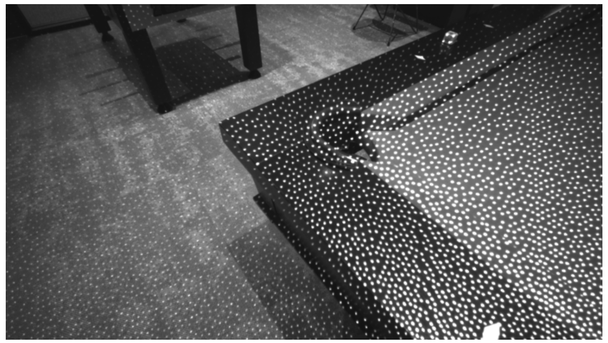
ASN No Occlusion Mask

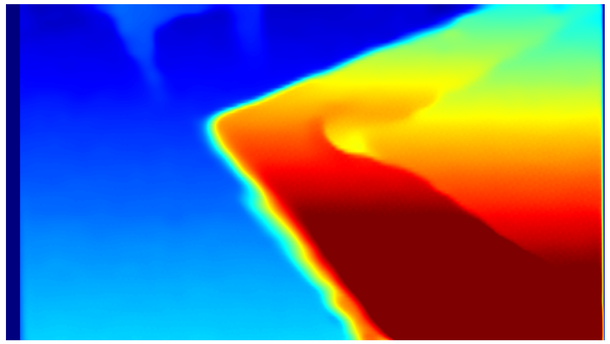
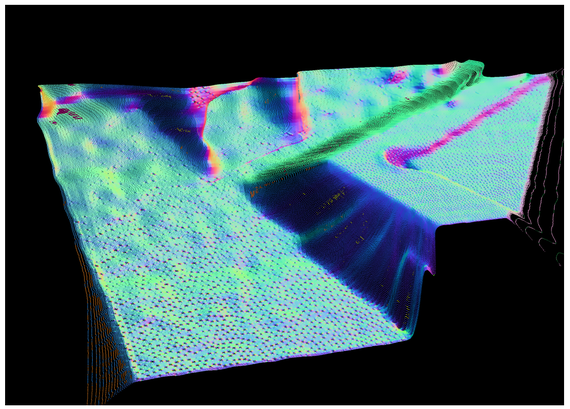
ASN Full

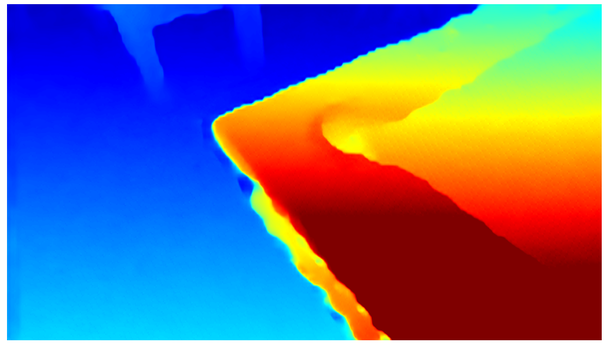
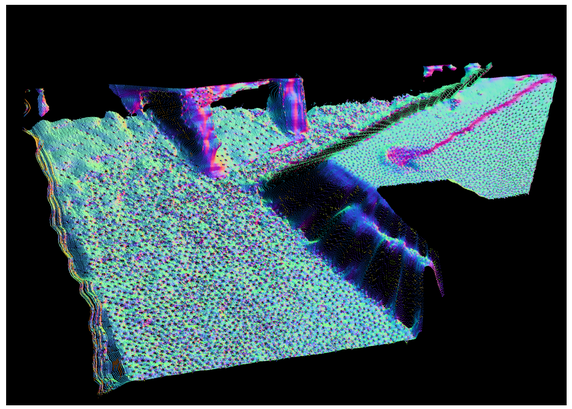
Does Invalidation Matter?



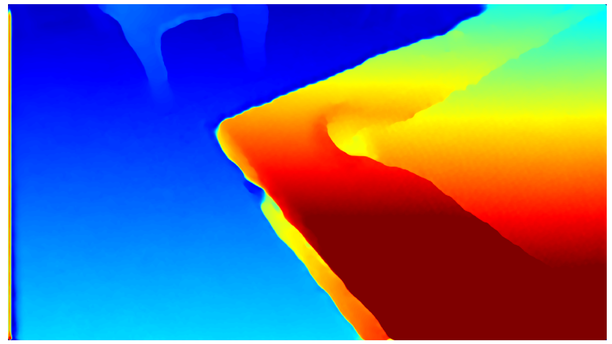
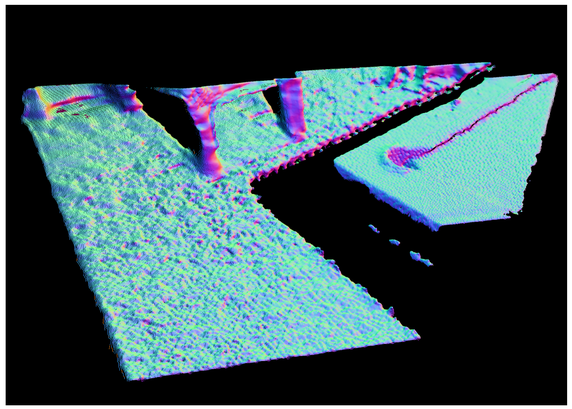

Godard et al.

ASN No Occlusion Mask

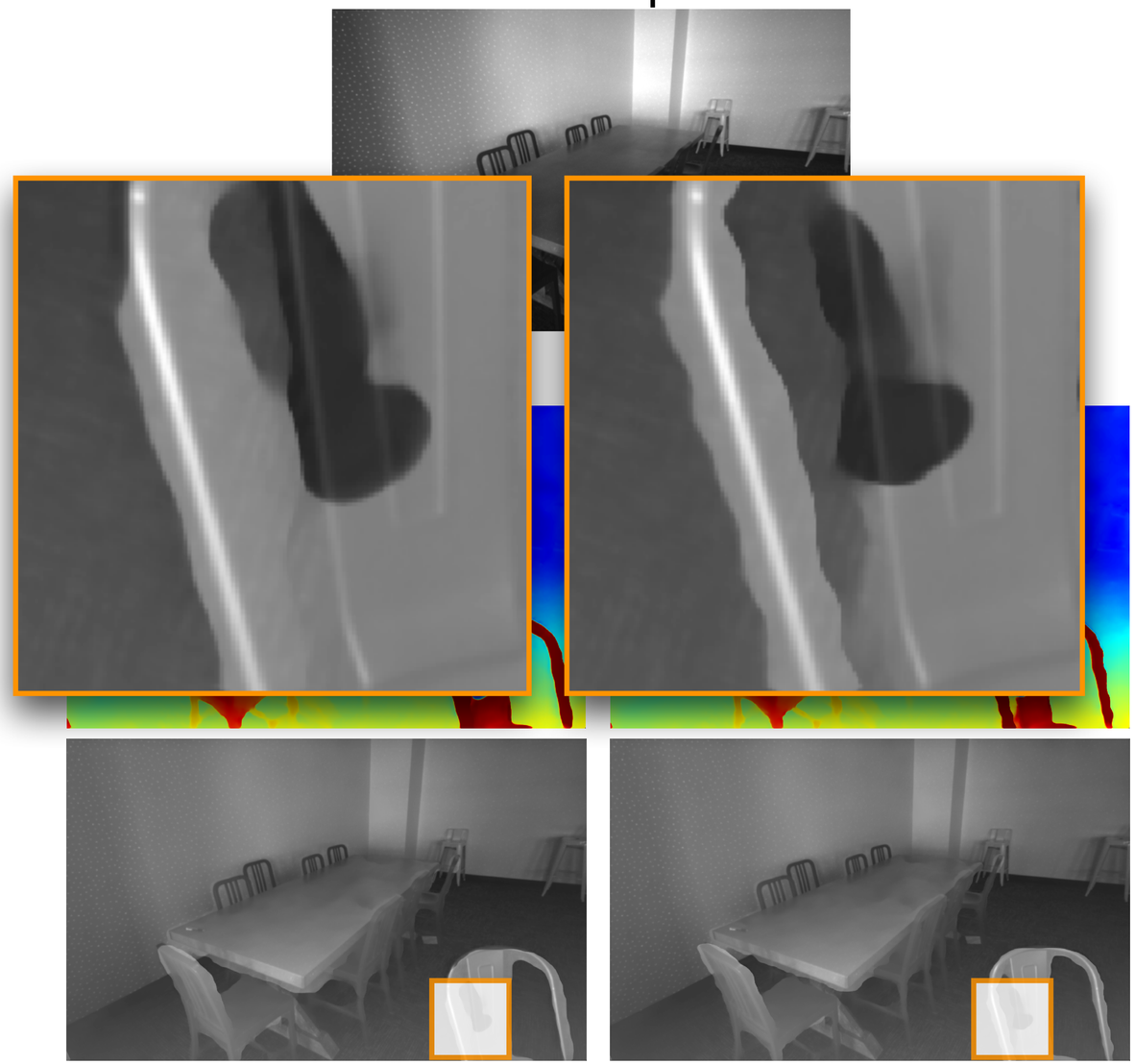



ASN Full

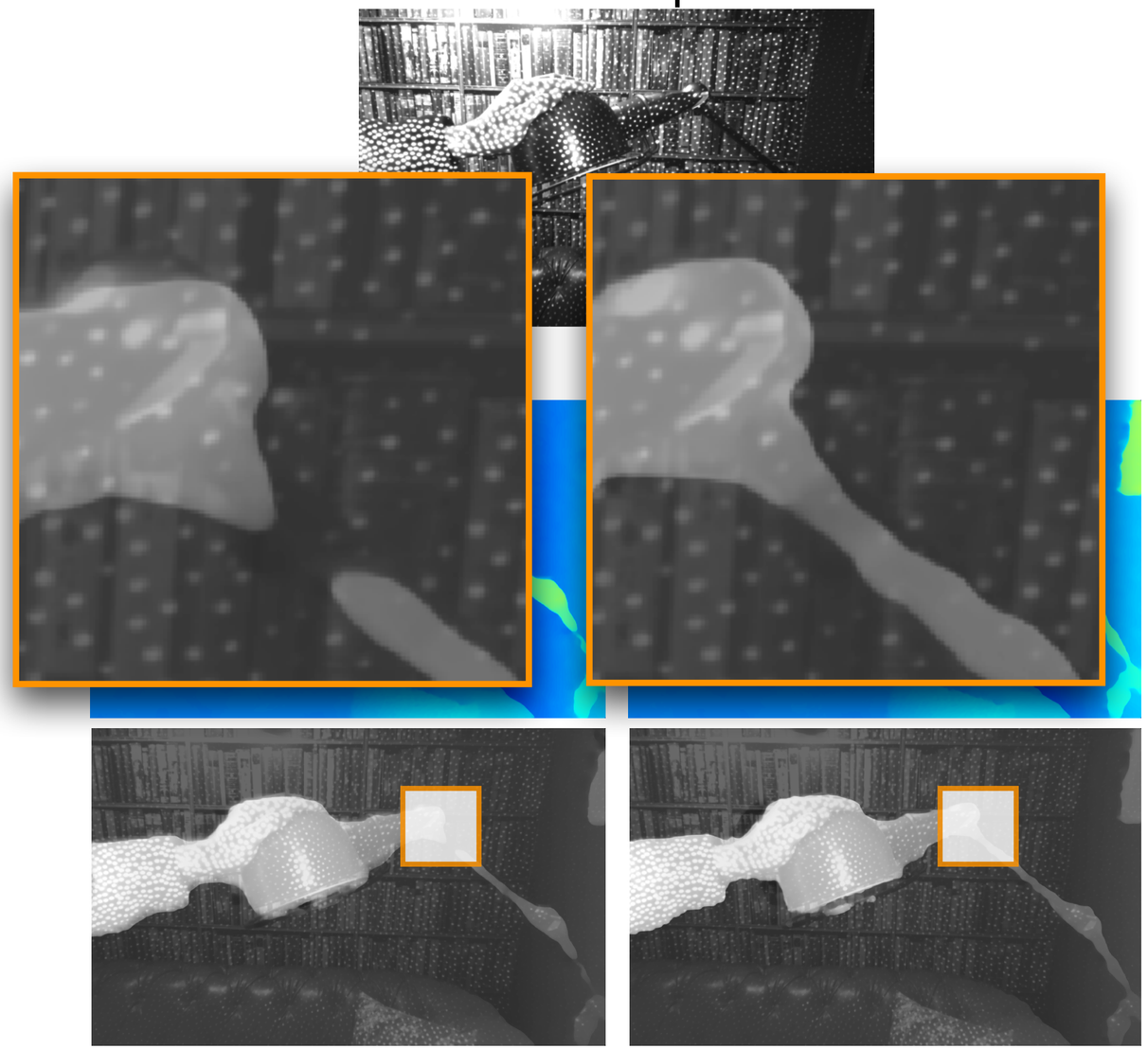



Does Window-based Matter?

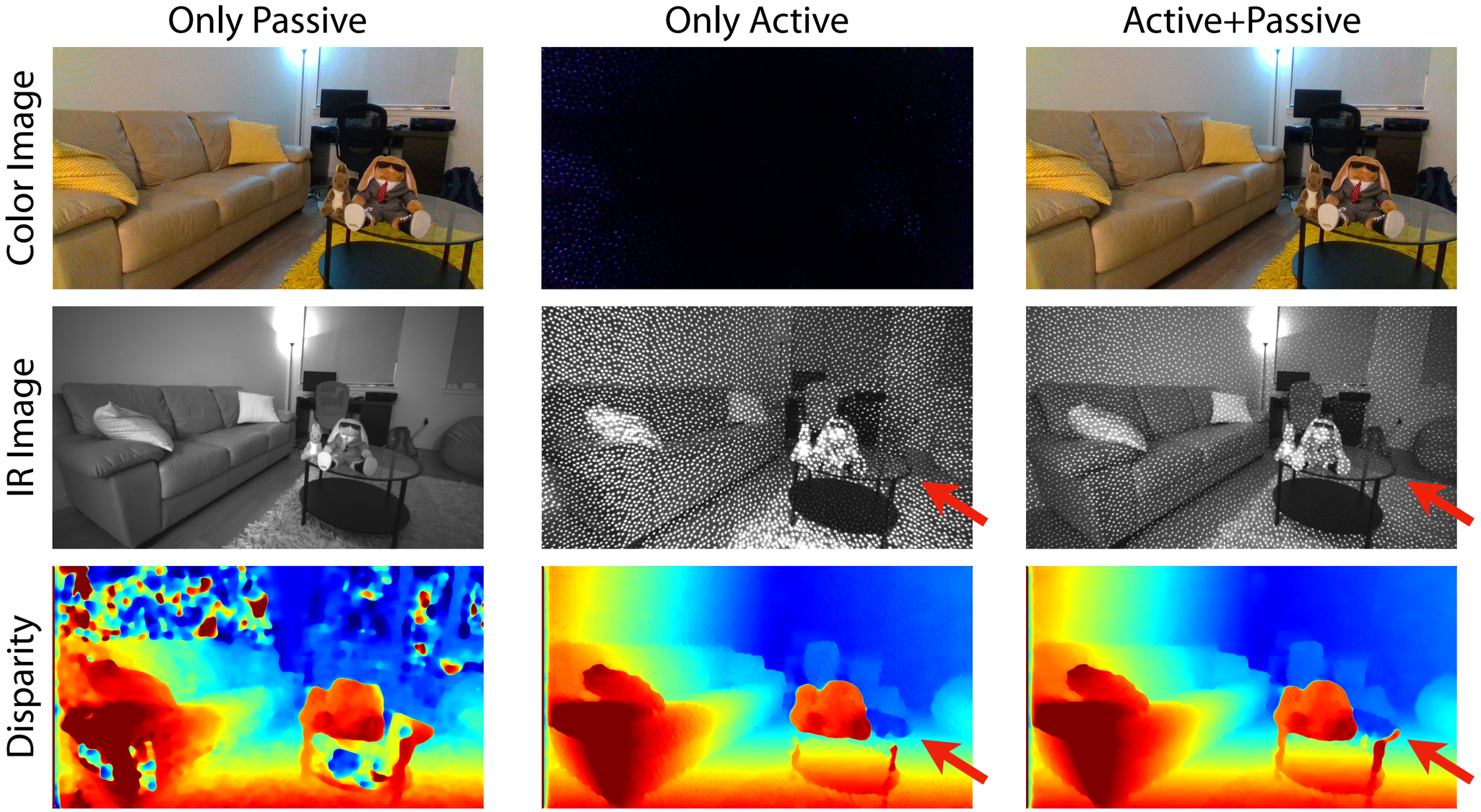
Left IR Input



Left IR Input

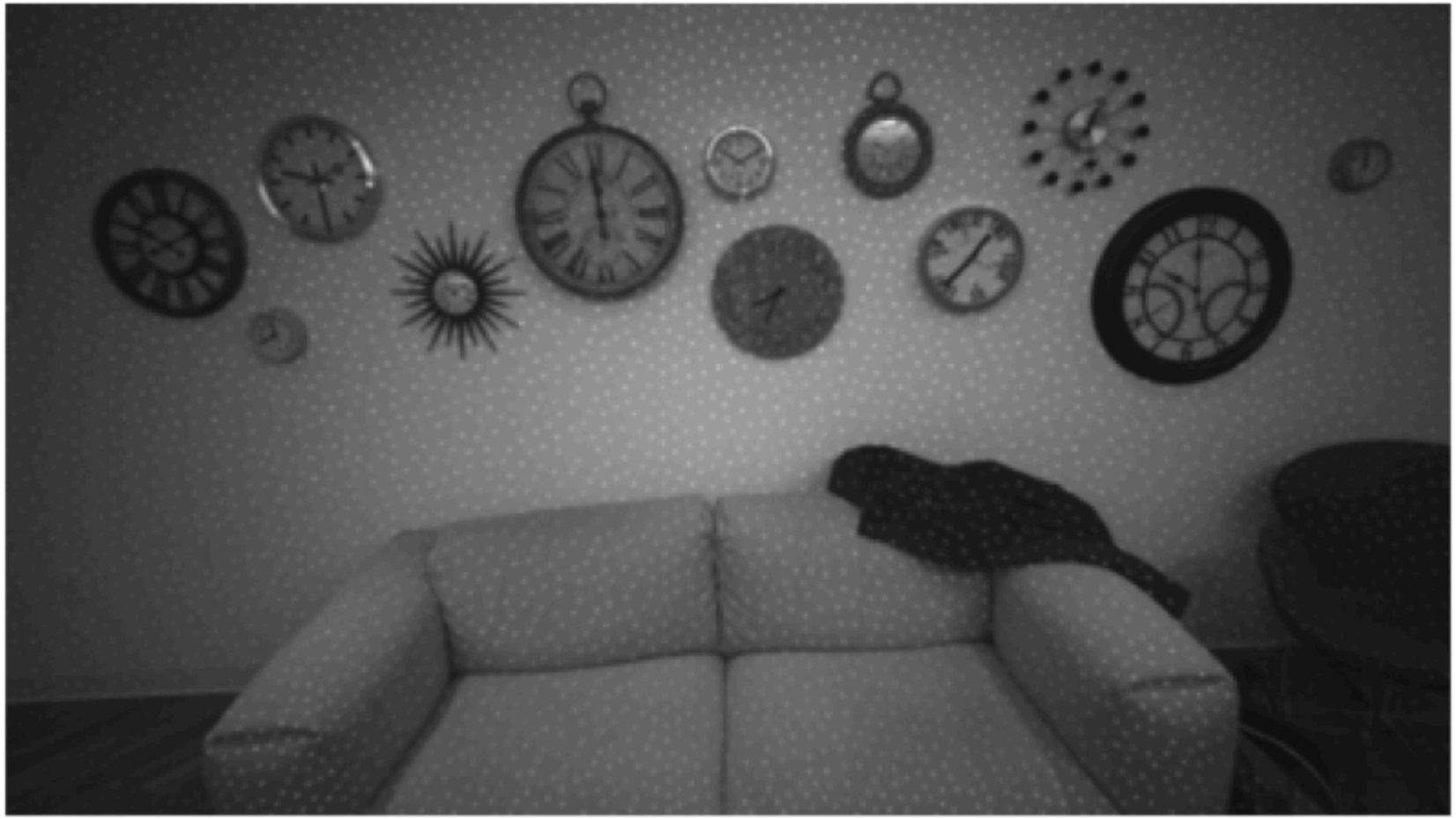


Why Active+Passive?



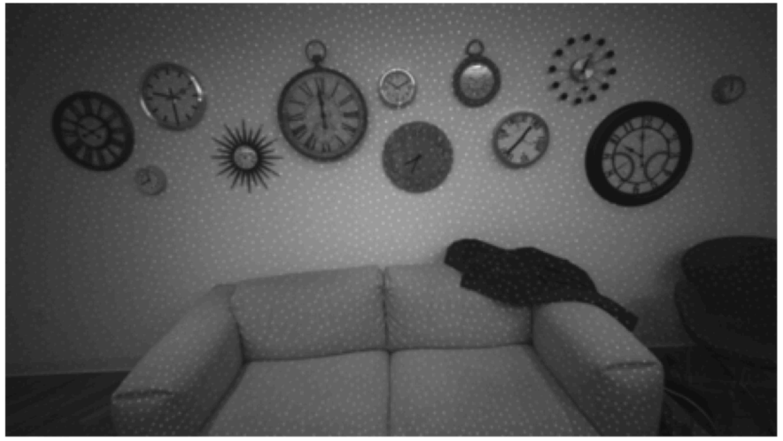
Conclusion

- We investigate active stereo with deep learning.

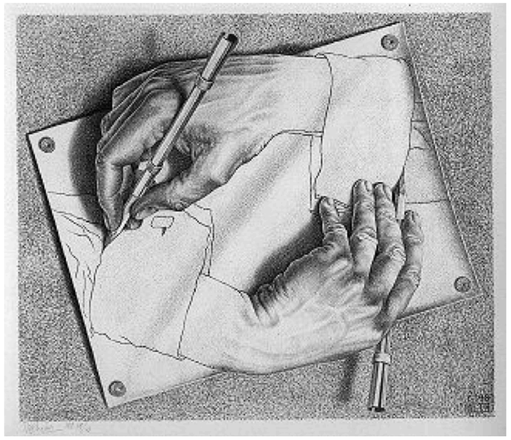


Conclusion

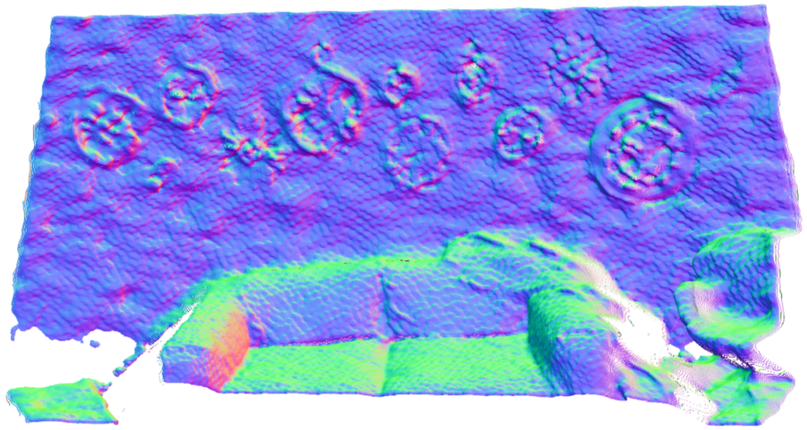
- We investigate active stereo with deep learning.
- We propose self-supervised learning for active stereo system using the improved photometric loss.
- We reduce the disparity error from 0.2 px to 0.03 px.



Active Stereo



Self-supervised Learning



Disparity Error: 0.03px

PRINCETON VISION & ROBOTICS

Google

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

ActiveStereoNet

End-to-End Self-Supervised Learning for Active Stereo Systems

Presenter: Yinda Zhang

Yinda Zhang[1,2], Sameh Khamis[1], Christoph Rhemann[1], Julien Valentin[1], Adarsh Kowdle[1], Vladimir Tankovich[1], Michael Schoenberg[1], Shahram Izadi[1], Thomas Funkhouser[1,2], Sean Fanello[1]

Google Inc.[1] Princeton University[2]

Project Webpage: http://asn.cs.princeton.edu/