A Statistical Model for Disocclusions In Depth-based Novel View Synthesis

Jayant Thatte (*jayantt@stanford.edu*) and Bernd Girod
Department of Electrical Engineering, Stanford University

Stanford | ENGINEERING
Electrical Engineering

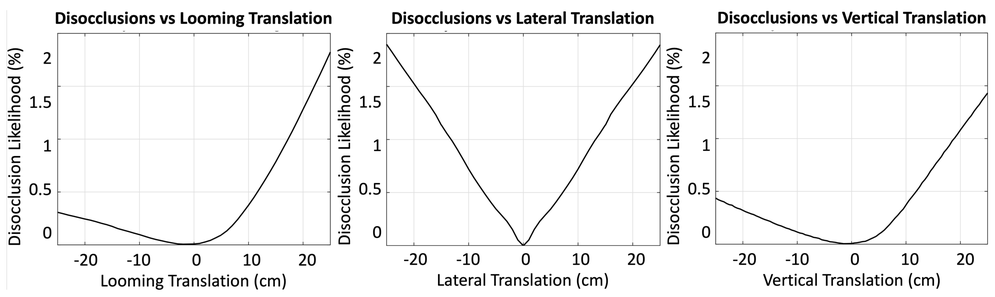
Abstract

- *Disocclusions are a critical issue in depth-based novel view synthesis; Inpainting algorithms are slow and yield unconvincing results*
- *Understanding the likelihood of the occurrence of disocclusions can help in designing better data representations and camera systems*
- *We propose a statistical model that predicts the likelihood of holes in the synthesized views as a function of the viewpoint translation*
- *Scene-dependent model parameters are estimated using shift/scale transformations on input depth, without needing view synthesis*

Disocclusion Model

Observations: For small viewpoint translations, the disocclusion likelihood

- Grows linearly with small viewpoint translation
- Is additive over small translations
- Is affine: no translation → no occlusions, i.e. $p(0,0,0) = 0$



Under random viewing direction and scene geometry, the indicator variable for occlusion A can be modeled as:

$$A = \begin{cases} 1 & with\ probability \quad p(x,y,z) \\ 0 & with\ probability \quad 1 - p(x,y,z) \end{cases}$$

Where x, y, z are the looming, lateral, and vertical components of the viewpoint translation from the source vantage point. And $p(x,y,z)$ is piecewise linear

$$p(x,y,z) = ax + by + cz$$

$$a = \begin{cases} a_+ & if\ x > 0 \\ a_- & else \end{cases} \qquad b = \begin{cases} b_+ & if\ y > 0 \\ b_- & else \end{cases} \qquad c = \begin{cases} c_+ & if\ z > 0 \\ c_- & else \end{cases}$$

$(a_+\ a_-\ b_+\ b_-\ c_+\ c_-)$ are model parameters

Estimating Model Parameters

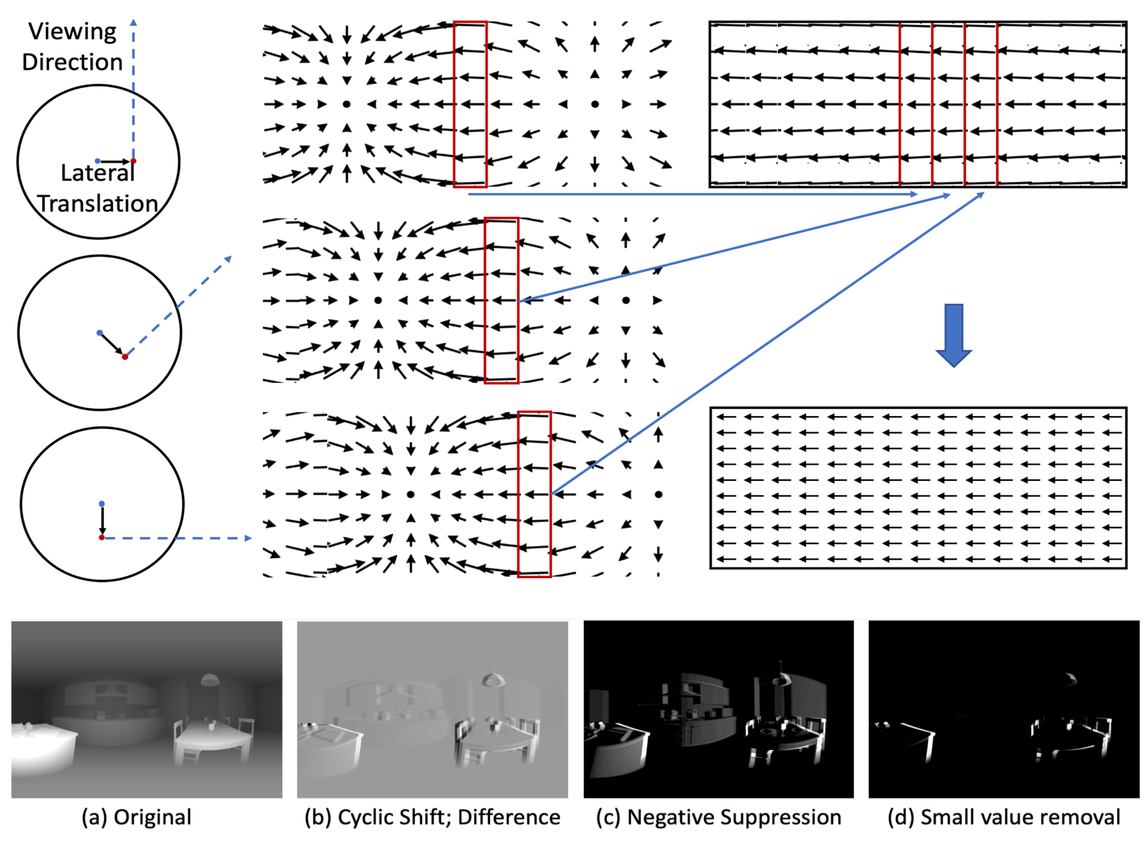
Ego motion induced flow fields + Depth discontinuities = Disocclusions

Approximated using simple shift/scale image transformations

Known from source depth map

E.g. Lateral viewpoint translation ≈ Cyclic horizontal shift of source panorama

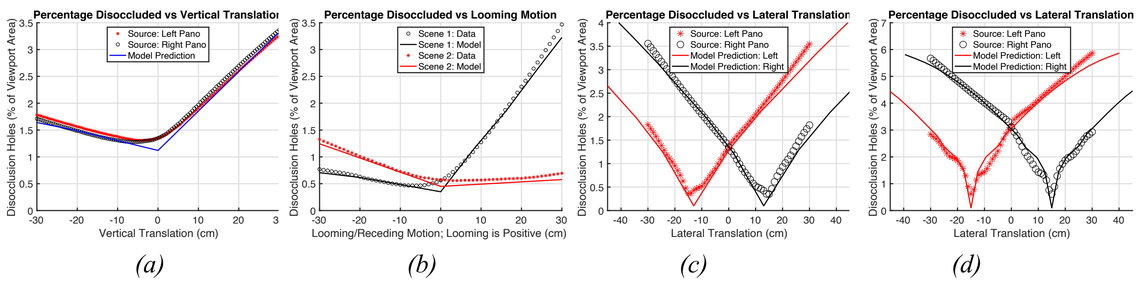
1. Apply shift/scale transformation to source depth (in Diopters)
2. Subtract from original; positive differences indicate disocclusion holes
3. Compute effective translational baseline that would give rise to each hole
4. Repeat with varying scales/shifts and accumulate holes for each baseline



Applications and Results

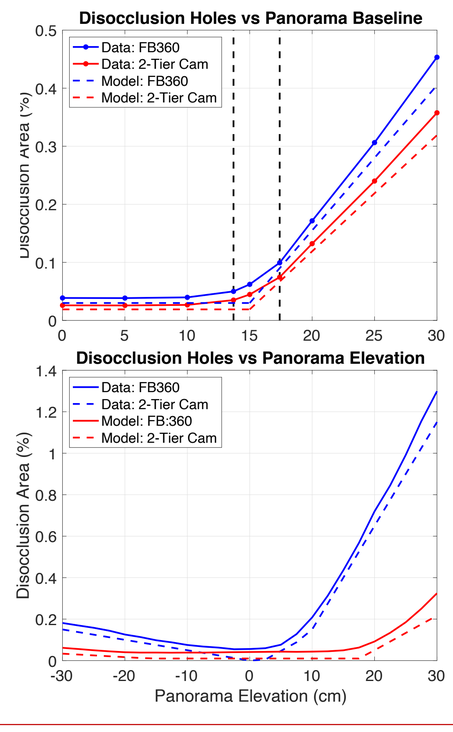
A. Views Synthesized from a Single Texture-plus-depth Source

Our model can correctly predict disocclusion trends including the nonlinear behavior of thin structures in the scene (d)



B. Middlebury Stereo Datasets

Over 27 scenes and 21 translations along each axis, the disocclusion likelihood predicted by the model and the ground truth have a Pearson correlation of – lateral motion: 0.977, vertical: 0.984, and looming: 0.910

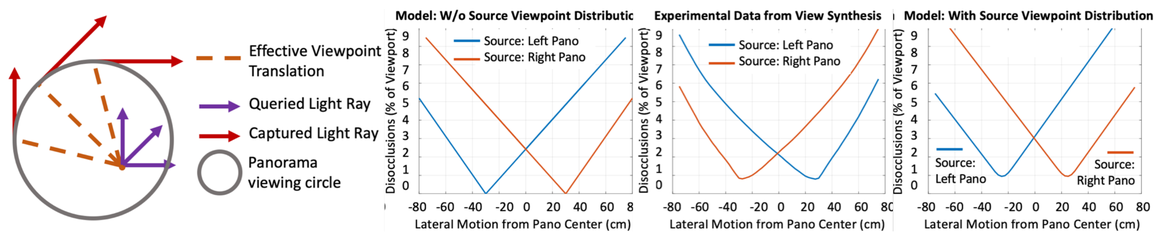


E. Analyzing Camera Geometries

- DIBR is commonly used to synthesize panoramas from camera systems
- We model how the disocclusion hole correlate in novel views synthesized using individual camera images and predict disocclusion behavior

C. Accounting for Per-Pixel Viewpoint Translation

Different pixels in a stereo panorama are captured from different viewpoints
Not accounting for this gives wrong prediction (left), improved model (right)



D. Views Synthesized from a Pair of Texture-plus-depth Sources

- Using multiple source images reduces disocclusions in novel views
- Disocclusion holes in views synthesized from multiple source images can be predicted if we know how the individual holes correlate with each other