Dense 3D Geometry Estimation

Enrique Dunn

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY


ETH zürich Microsoft URCV

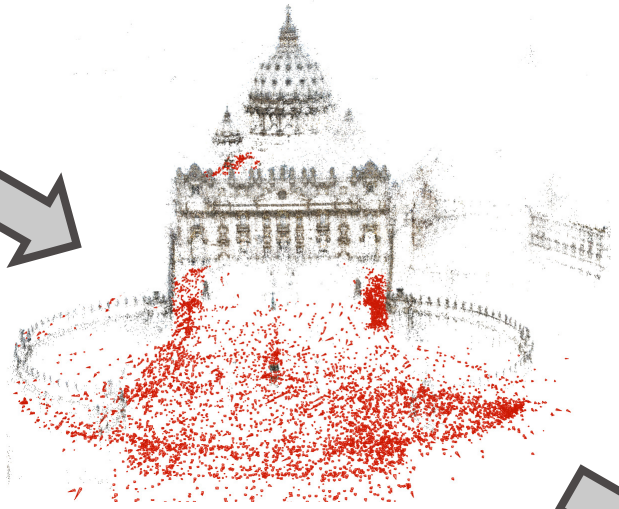
Dense 3D Modeling

Unstructured Images

- A generalization of sparse 3D reconstruction

Sparse Model

SfM



Dense Model

MVS

COLMAP - 3D reconstruction pipeline:
https://github.com/colmap/colmap

Dense 3D Modeling

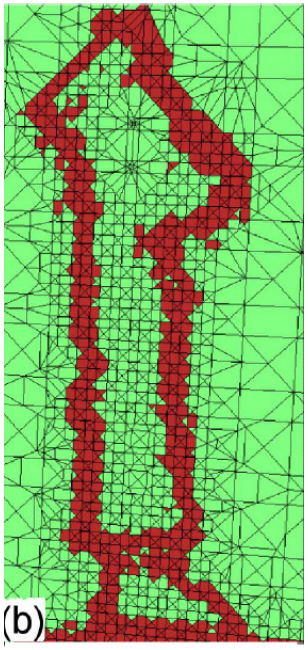
- Goal: Reconstruct all observed scene surfaces that are compliant with available geometric SfM estimates and local image appearance

3D Modeling Frameworks

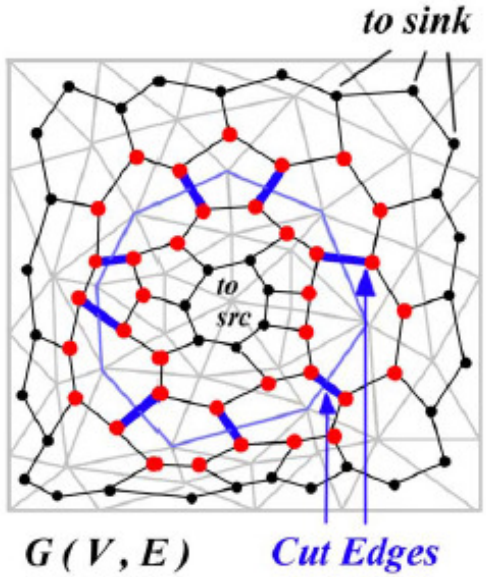
- Globally Parameterized 3D Structure

S. Sinha, P. Mordohai, M. Pollefeys, Multi-View Stereo via Graph Cuts on the Dual of an Adaptive Tetrahedral Mesh, ICCV'07

Sample Input Imagery



Tetrahedral Lattice



Graph Cut Optimization

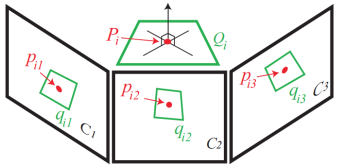
Surface Model

3D Modeling Frameworks

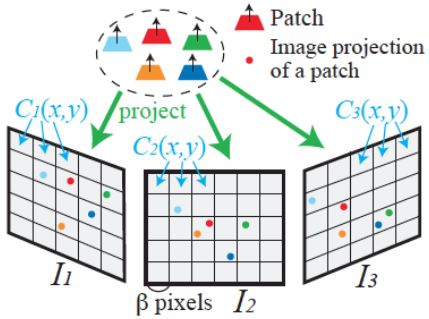
- Globally Parameterized 3D Structure
- Locally Parameterized 3D Structure

Y Furukawa, J Ponce, Accurate, dense, and robust multiview stereopsis. PAMI 2010.

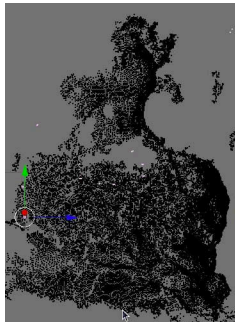
Sample
Input Imagery



Multi-view
Surface Patches



Procedural Verification
& Propagation

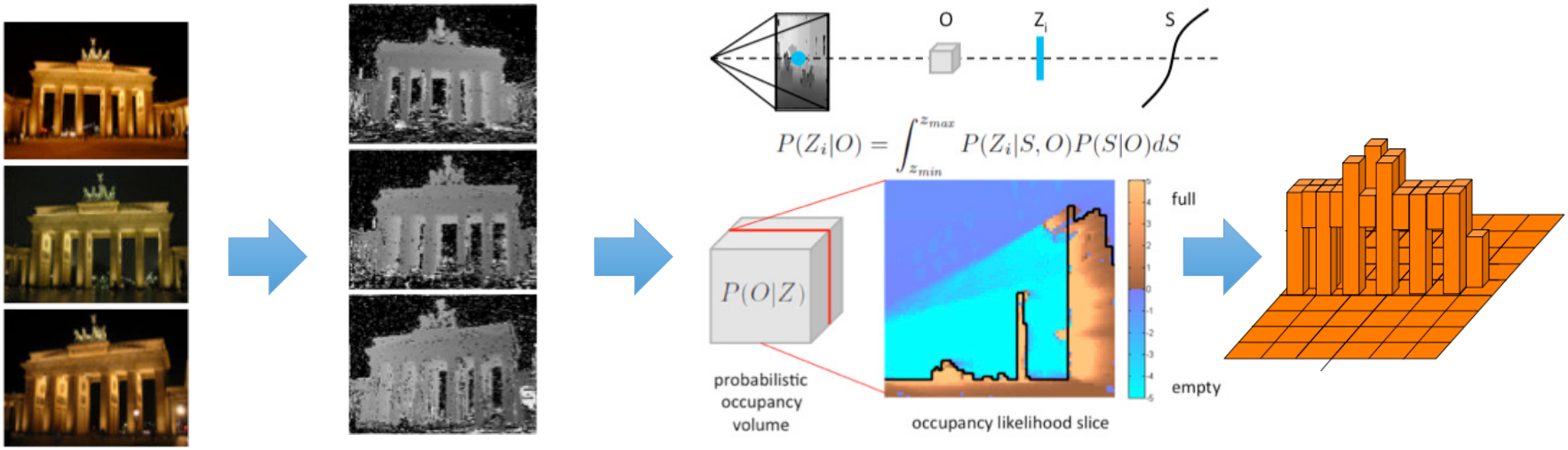


Oriented
Point Cloud

3D Modeling Frameworks

- Globally Parameterized 3D Structure
- Locally Parameterized 3D Structure
- Mixed Parameterization

D Gallup, M Pollefeys, JM Frahm . 3d reconstruction using an n-layer heightmap. Joint Pattern Recognition Symposium, 2010



Sample Input Imagery

Multiple Depthmaps

Probabilistic Occupancy Analysis

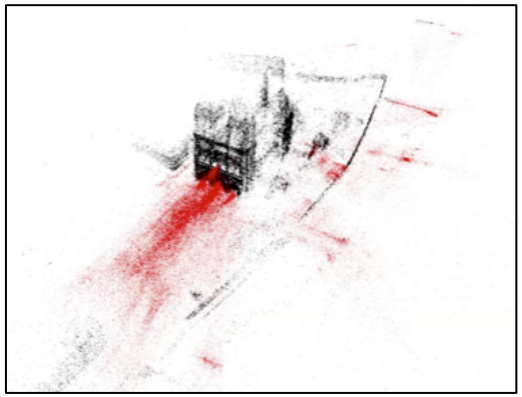
Multi-Layer Heightmap

Crowd-Sourced Dense 3D

- Data Scale
  - Total Imagery

Herz-Jesu-P8 images

vs.



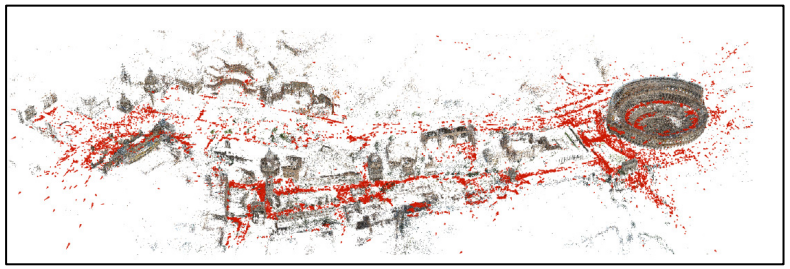
126K Registered images

Crowd-Sourced Dense 3D

- Data Scale
  - Total Imagery
  - Spatial extent

Herz-Jesu-P8 images

vs.



Towards City-Scale Registration

Crowd-Sourced Dense 3D

- Data Scale
- Content Diversity
  - Indoor/outdoor

Inside Sistine Chapel
Vatican City

Outside Sistine Chapel
Vatican City

Crowd-Sourced Dense 3D

- Data Scale
- Content Diversity
  - Indoor/outdoor
  - Static/Dynamic

Crowd-Sourced Dense 3D

- Data Scale
- Content Diversity
- Heterogeneous Capture
  - Illumination
  - Viewpoint Bias
  - Resolution
  - Field of View

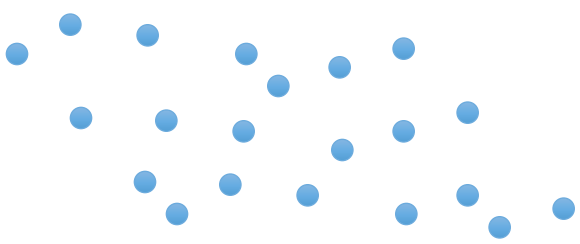
Crowd-Sourced Dense 3D

- Data Scale **Efficiency**
- Content Diversity **Generality**
- Heterogeneous Capture **Robustness**

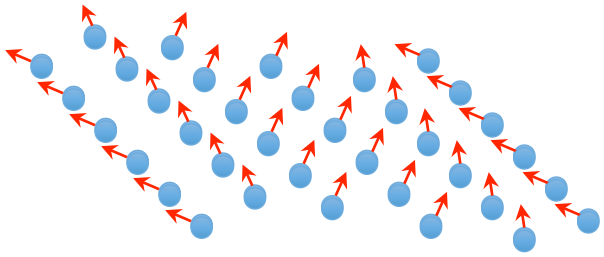
Dense 3D Modeling

Multiview Stereo

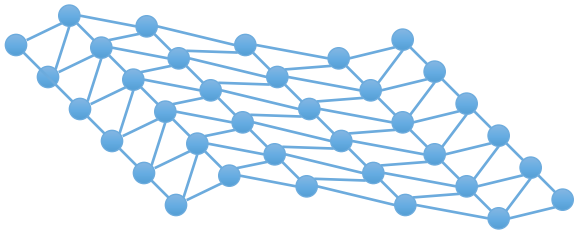
Depth Fusion



Initial
Reconstructed 3D Points



Densified
3D Points + Normals



Surface Approx.
Triangular 3D Mesh

Outline

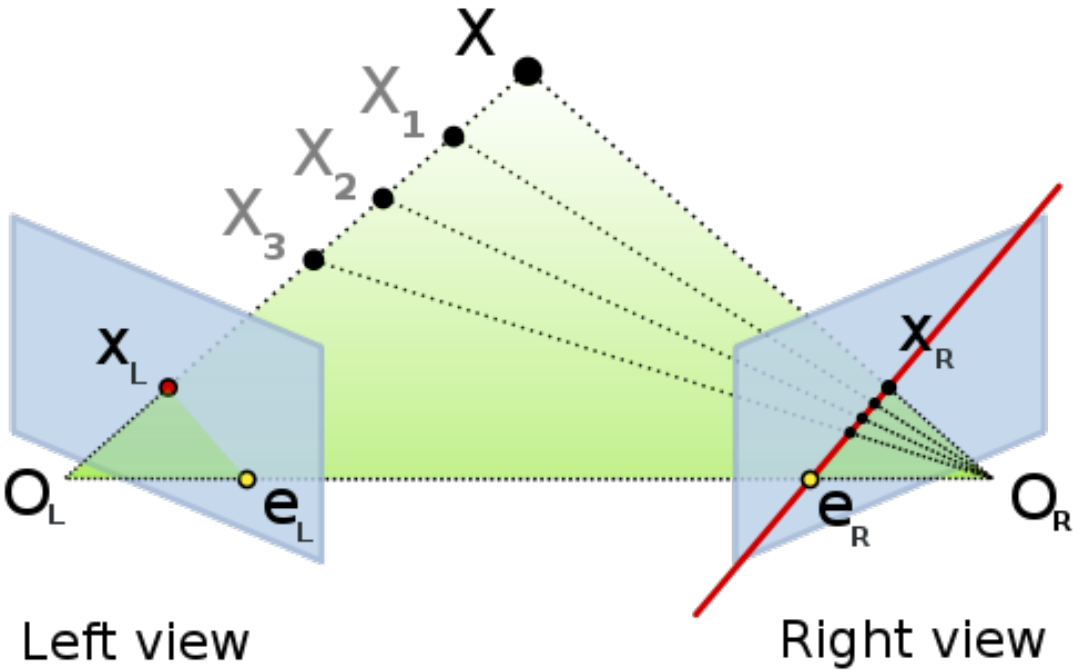
- Multi-View Stereo
- Robust Depthmap Estimation
- Fusion and Surface Estimation

Outline

- Multi-View Stereo
  - Building a Cost Volume
  - Plane Sweep
  - Patch Match
- Robust Depthmap Estimation
- Fusion and Surface Estimation

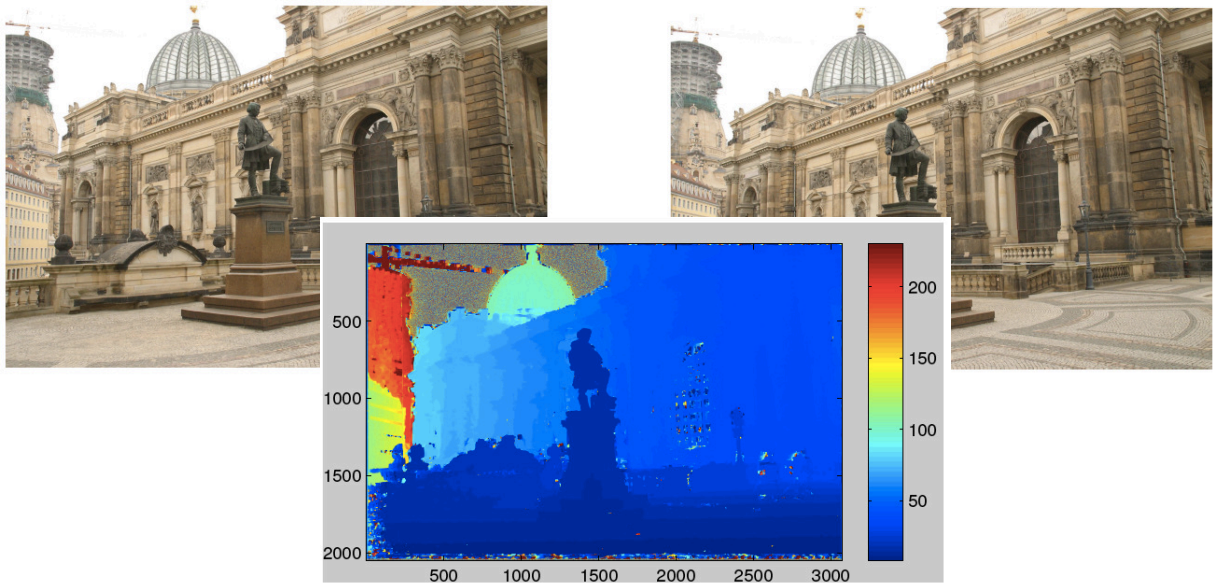
Stereo Depth Estimation

- Basic Insight: Given a known imaging configuration, geometry estimation is reduced to pixel correspondence search.

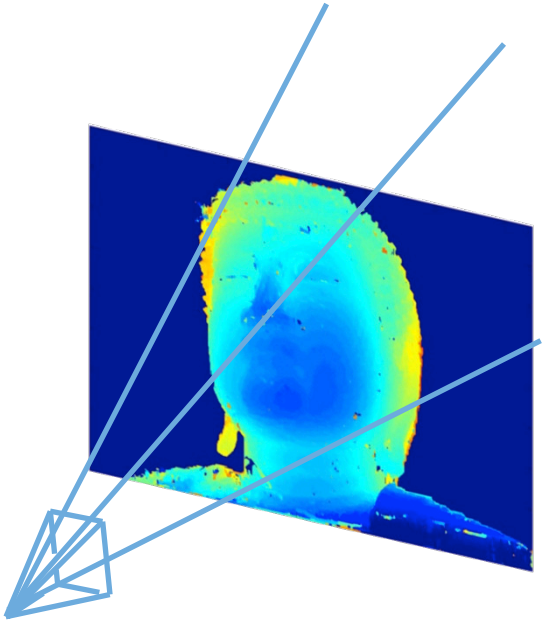


Depthmap Estimation

- Goal: Determine a view-dependent depthfield (*i.e.* pixel-wise estimates) by leveraging the photo-consistency of a set of overlapping images.



Depthmap Estimation

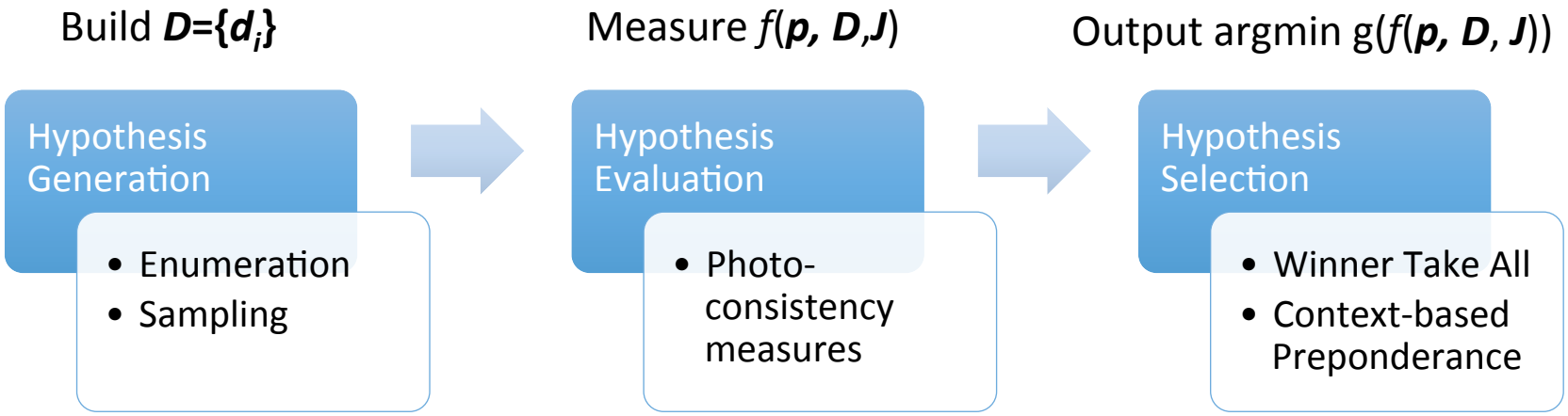


Projection into 3D space

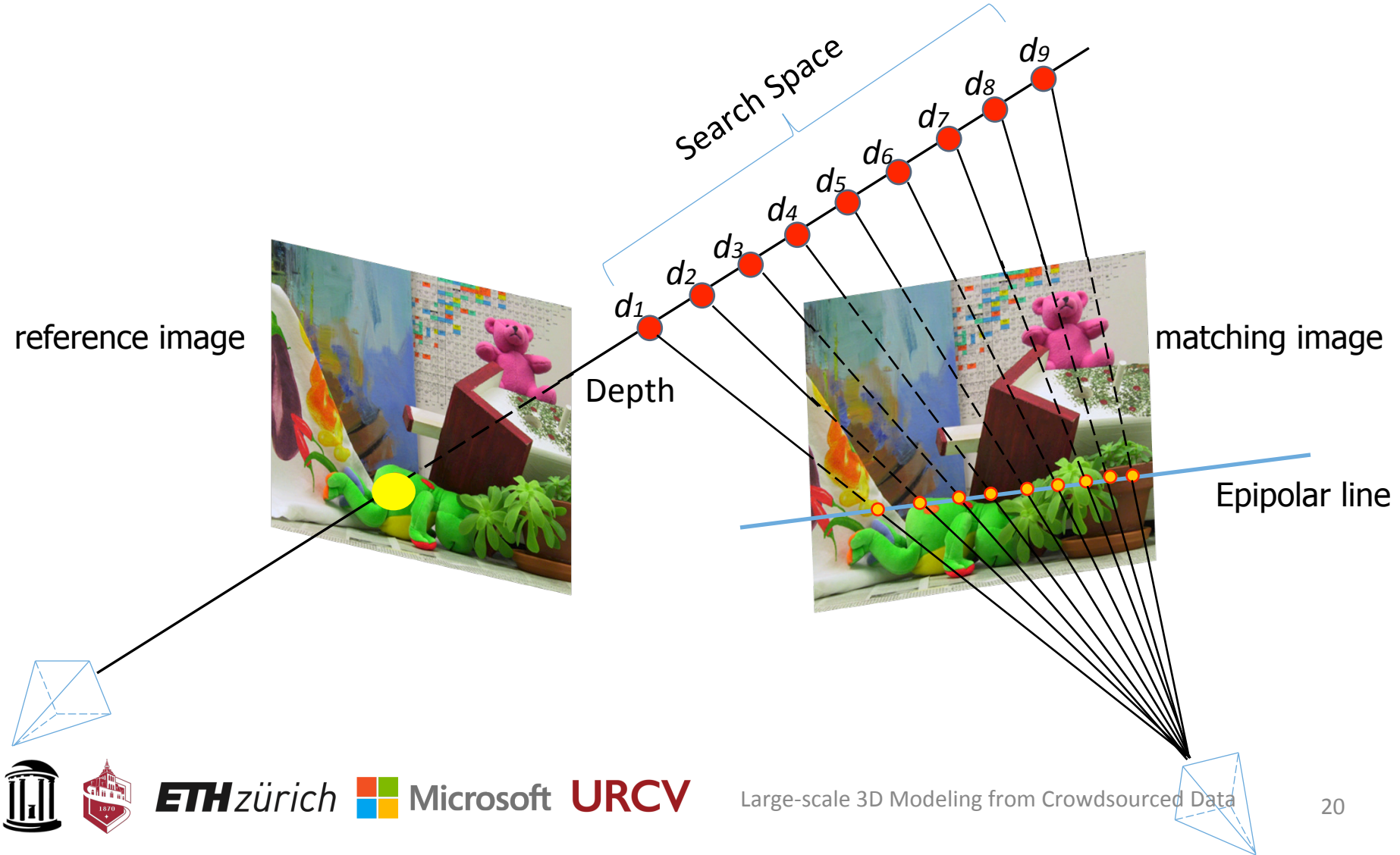
3D point cloud

Depthmap Estimation

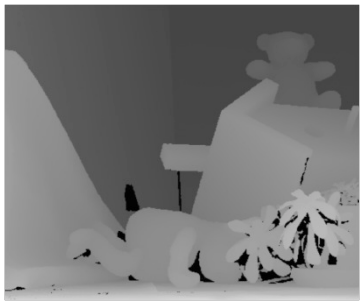
- For each pixel ***p*** in image ***I*** determine a depth value, among a set ***D*** of depth hypotheses, most congruent with a set ***J*** of source images.



The correspondence cost profile



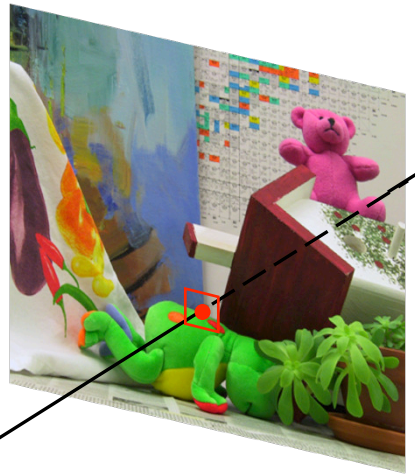
Pixel-wise Cost Evaluation



Depth Map

Ground Truth

Pixel Matching

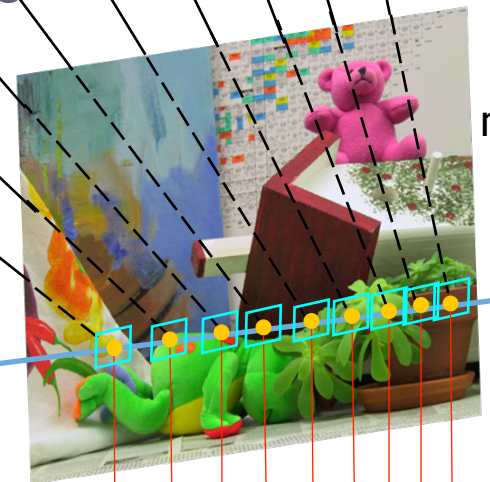


reference image

Search Space

d_1 d_2 d_3 d_4 d_5 d_6 d_7 d_8 d_9

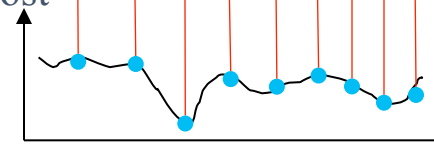
Depth



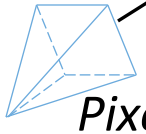
matching image

Epipolar line

Matching Cost

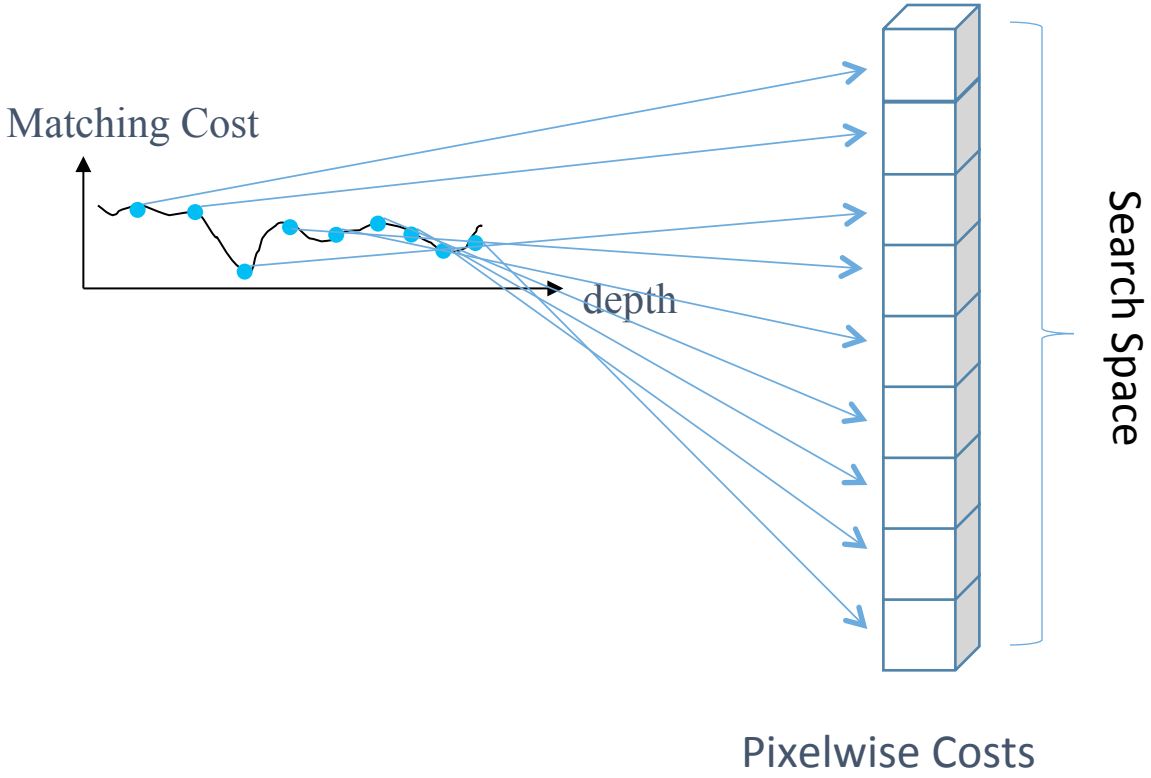


depth

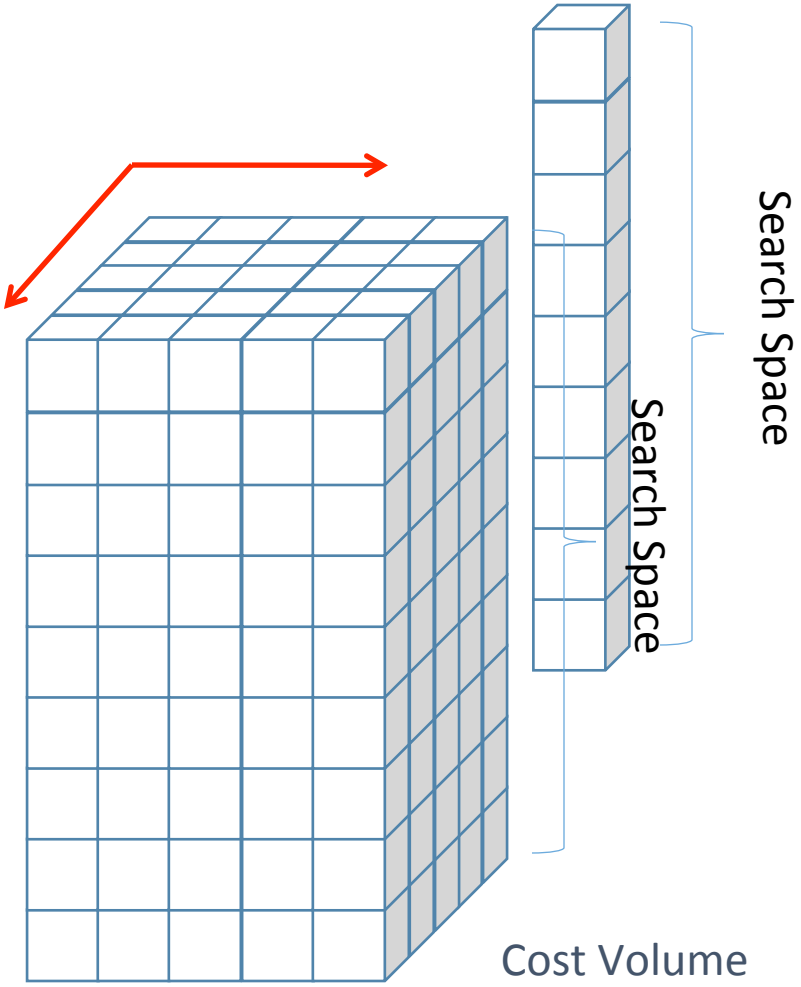


Pixel similarity: measured by color differences

Building a Cost Volume

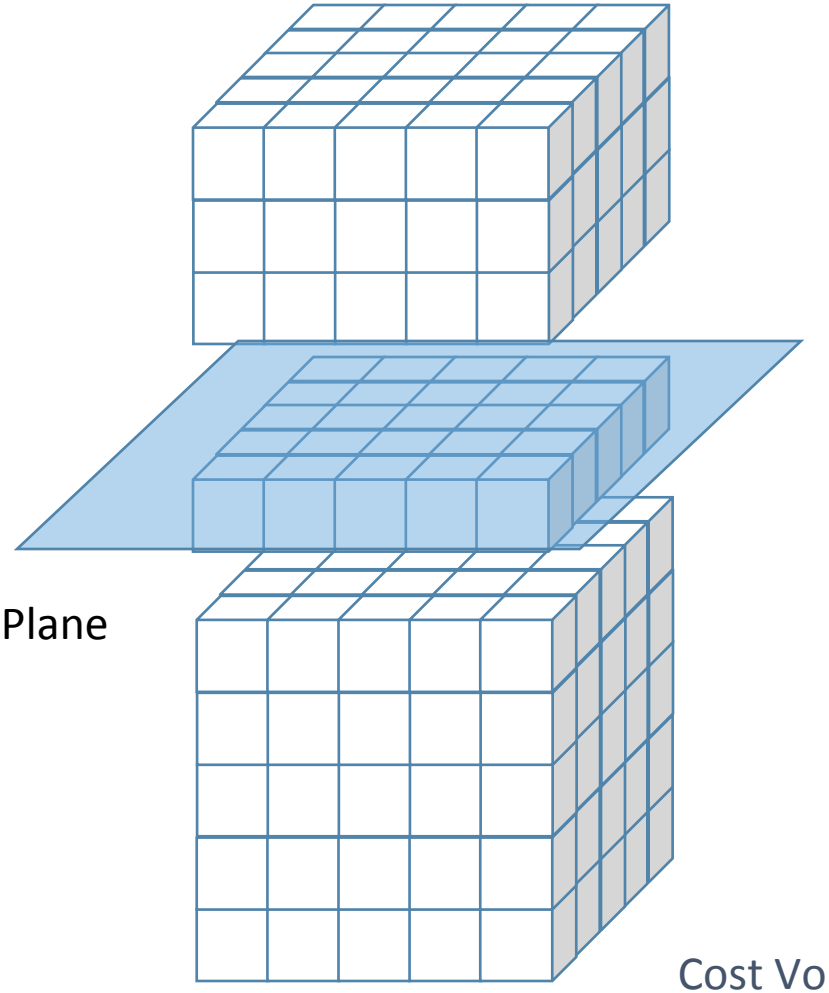


Building a Cost Volume



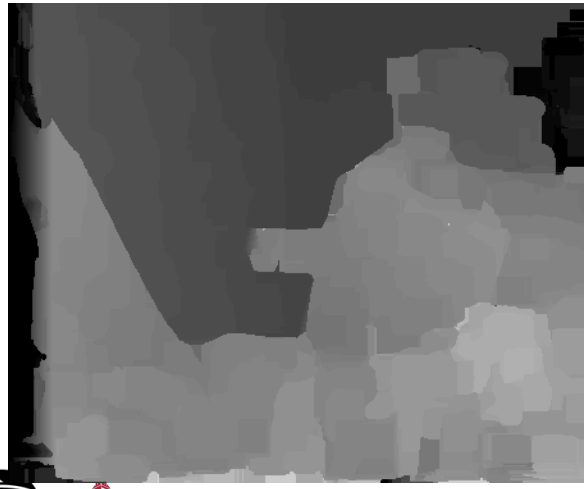
Cost aggregation:
cutting the cost volume.

Window-based Cost Evaluation

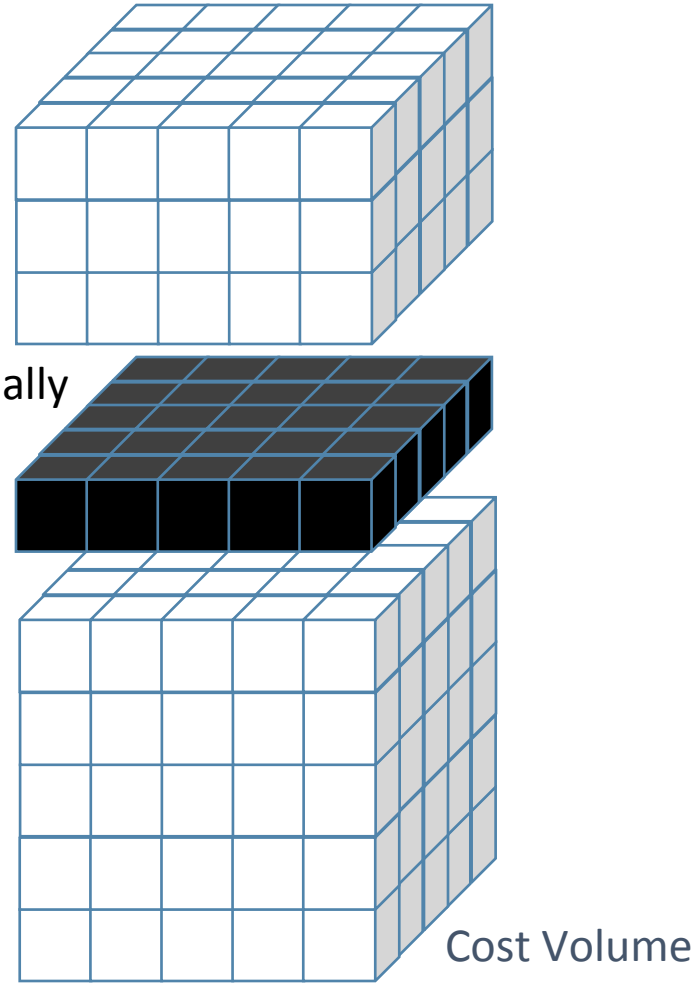


Window-based Evaluation

Treat neighbors equally

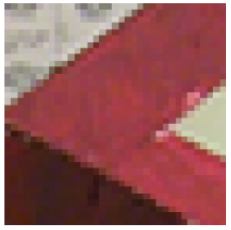


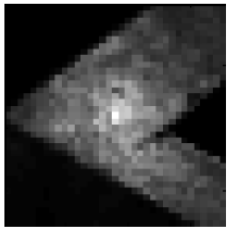
Sum of Absolute Differences (SAD)



Cost Volume

Window-based Evaluation

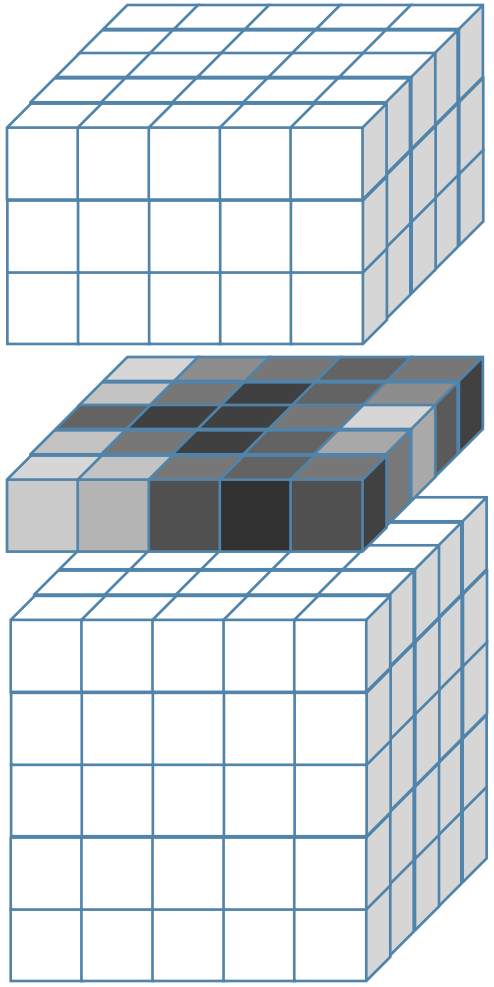




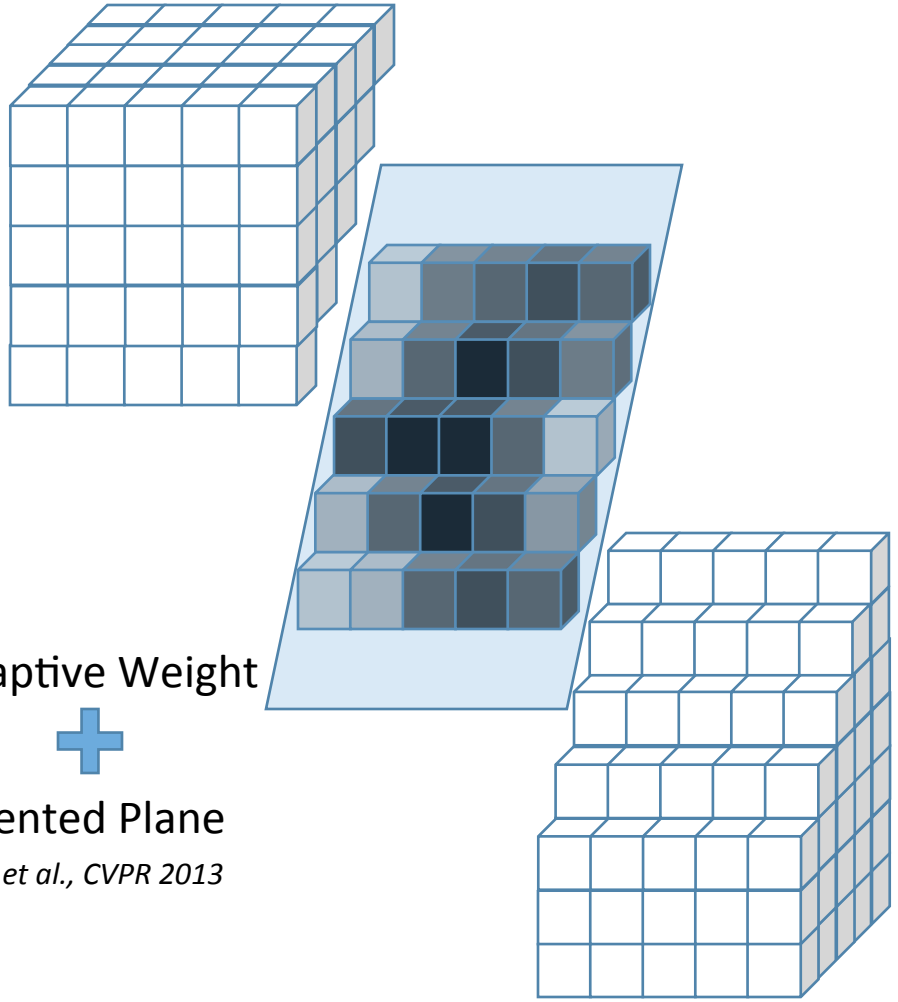
- Color differences
- Spatial distances

Adaptive Weight

Yoon and Kweon, PAMI 2006



Window-based Evaluation



Adaptive Weight

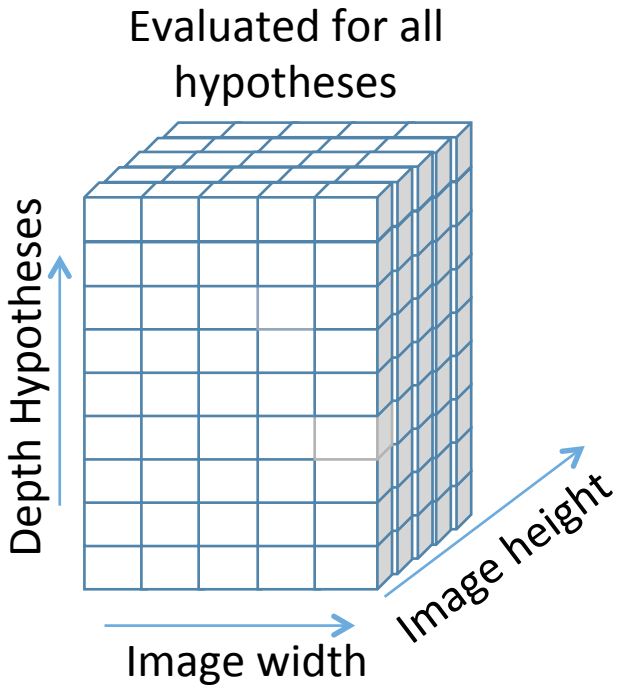
+

Oriented Plane

Lu et al., CVPR 2013

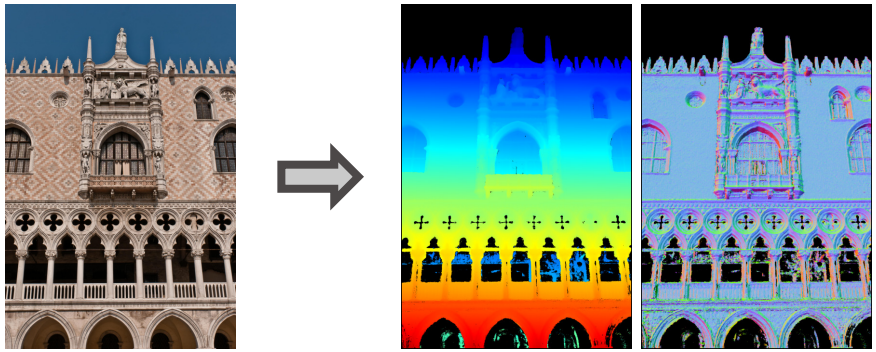
Aggregated Photo-consistency

- Sum of Squared Differences (SSD)
- Normalized Cross Correlation (NCC)
- CENSUS Transform
- Rank Transform
- Mutual Information
- Learning-based Methods



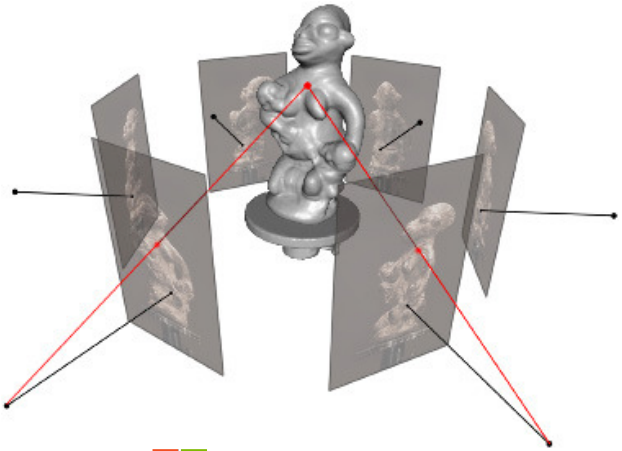
Binocular vs. Multi-view Depthmaps

- Dense depth/normal reconstruction



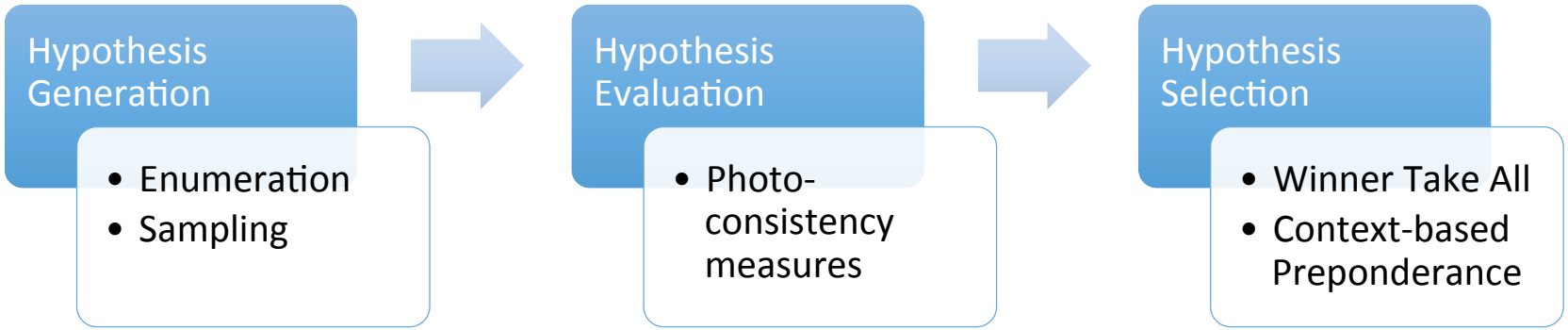
Problem/Solution Representation

- Multi-view correspondence search



Data Association Strategies

Different Implementations

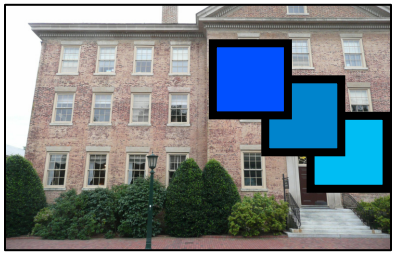


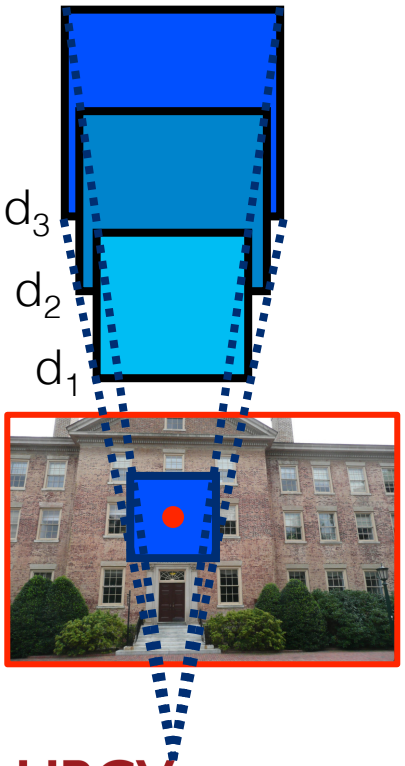
Plane Sweep

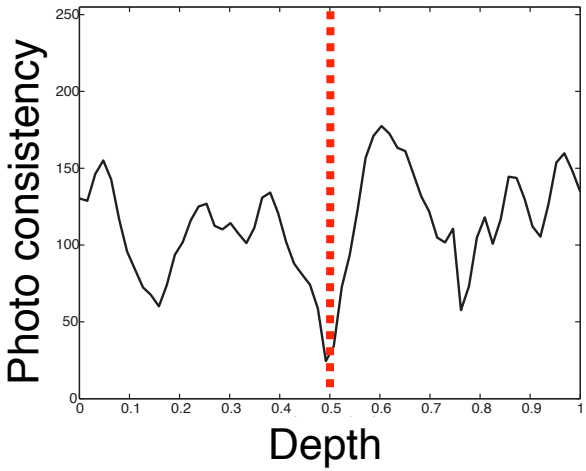
Patch Match

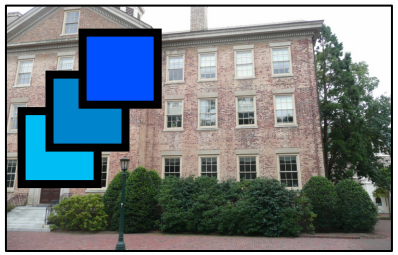
Plane Sweep Framework

- Hypothesis space is explored by “sweeping” a plane across the scene.







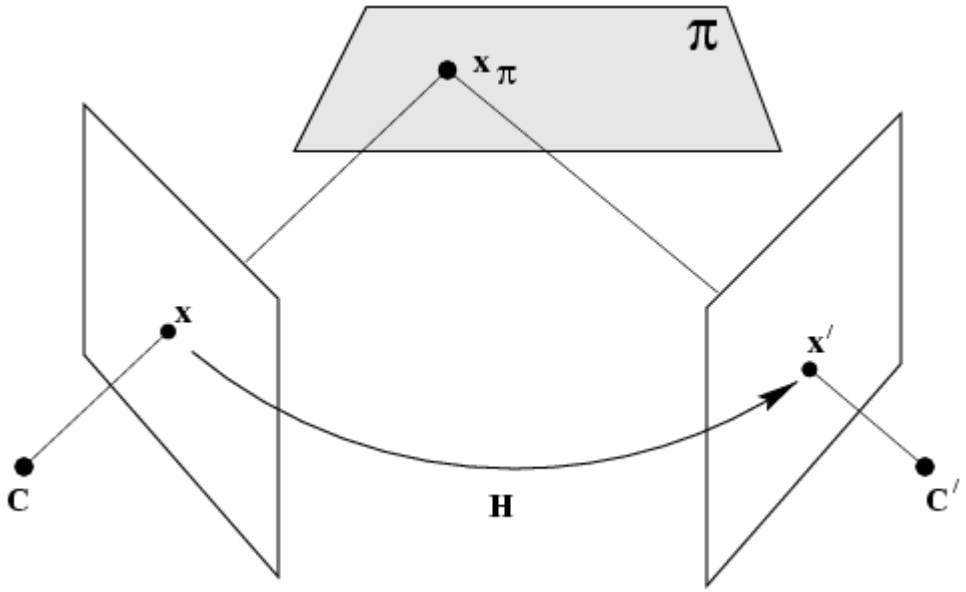


Plane Sweep: Plane-Induced Homography

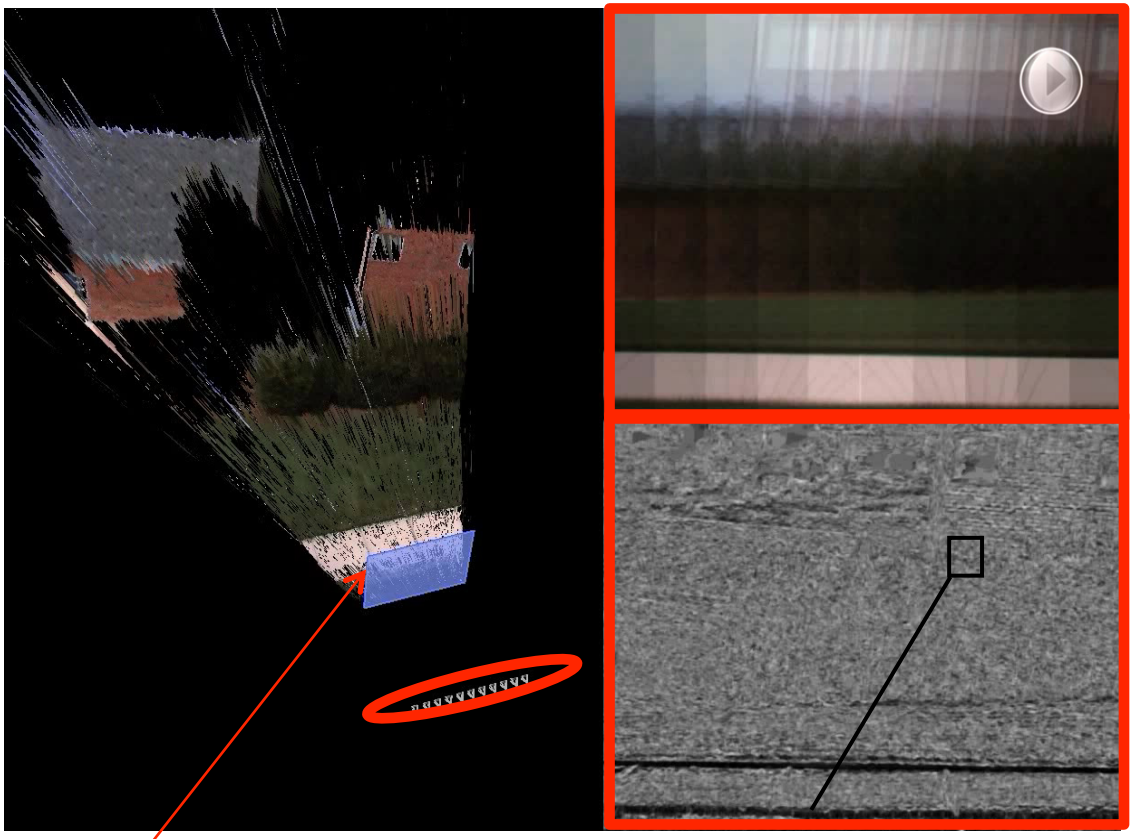
Given: two cameras $P=K[I\ |0]$ and $P' = K'[R\ |\ t]$, and
A plane $\pi=(n^T,d)^T$

The homography $x'=Hx$ is defined as

$$H = K'(R - tn^T/d)K^{-1}$$

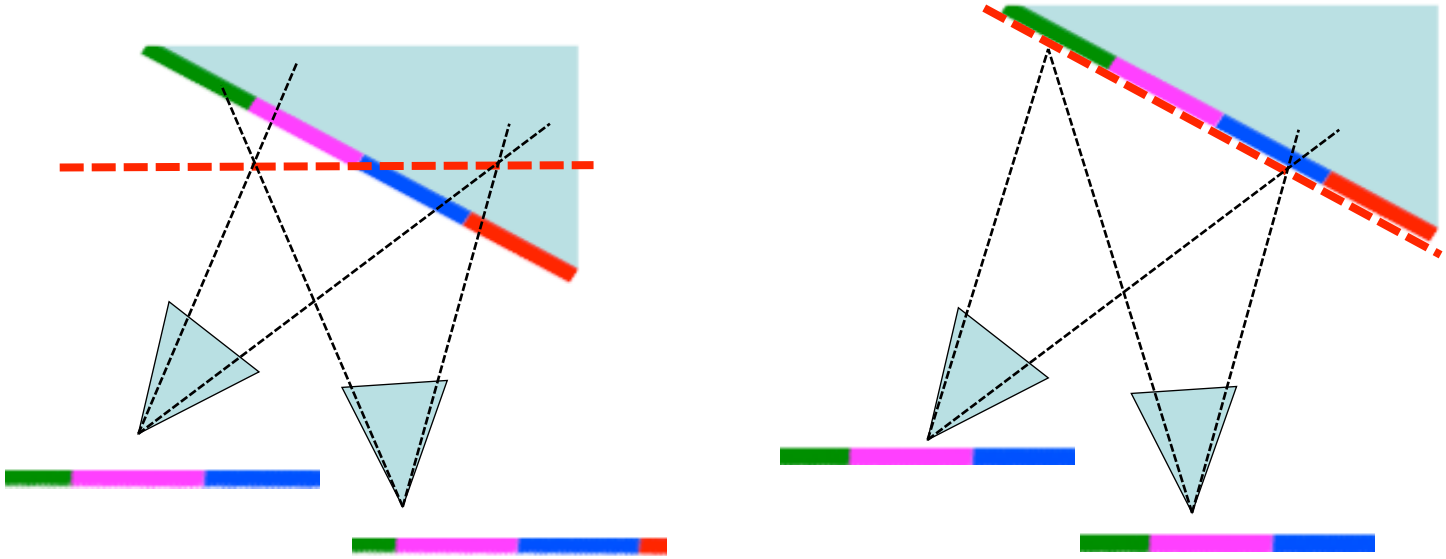


Plane Sweep: Operation



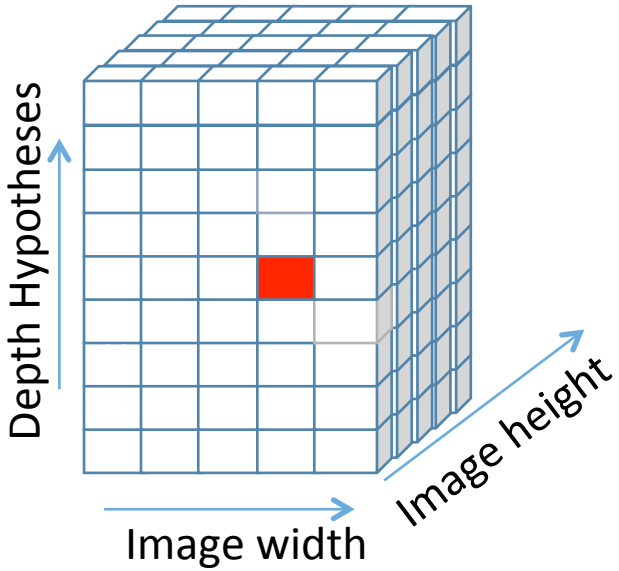
Plane Sweep: Enhanced Robustness

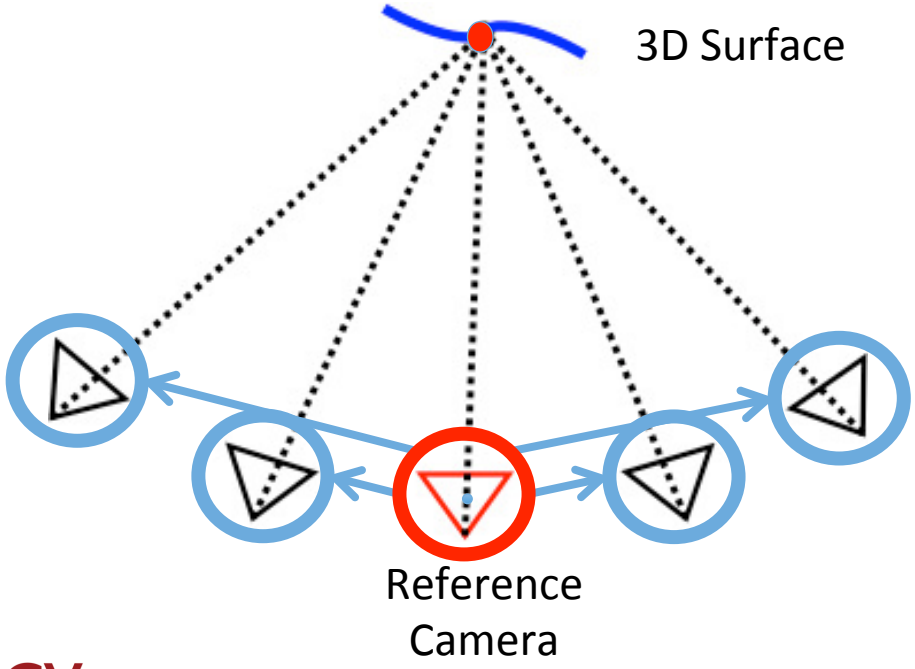
- Aligning sweeping direction to surface orientation reduces photometric inconsistencies at correct depth



Plane Sweep: Computational Burden

```
Initialize Similarity Cost
For Each Pixel
    For Each Depth
        For Each Source Image
            Accumulate Similarity Cost
```





Plane Sweep Overview

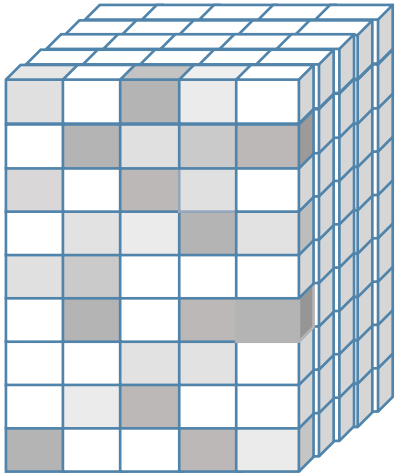
- Pros:
  + Flexible homography-based warping
  + Controllable search resolution

- Cons:
  - Exhaustive Approach
  - Local/Greedy Search

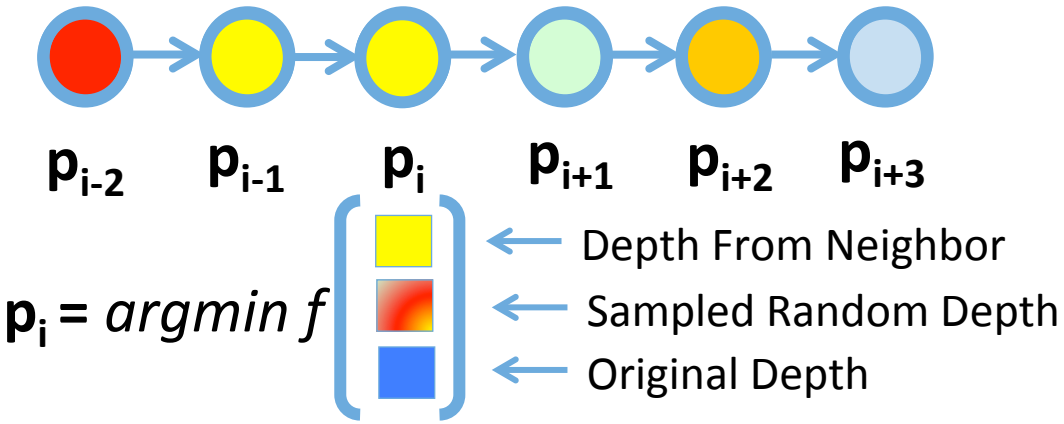
Patchmatch Stereo

- Randomized correspondence search

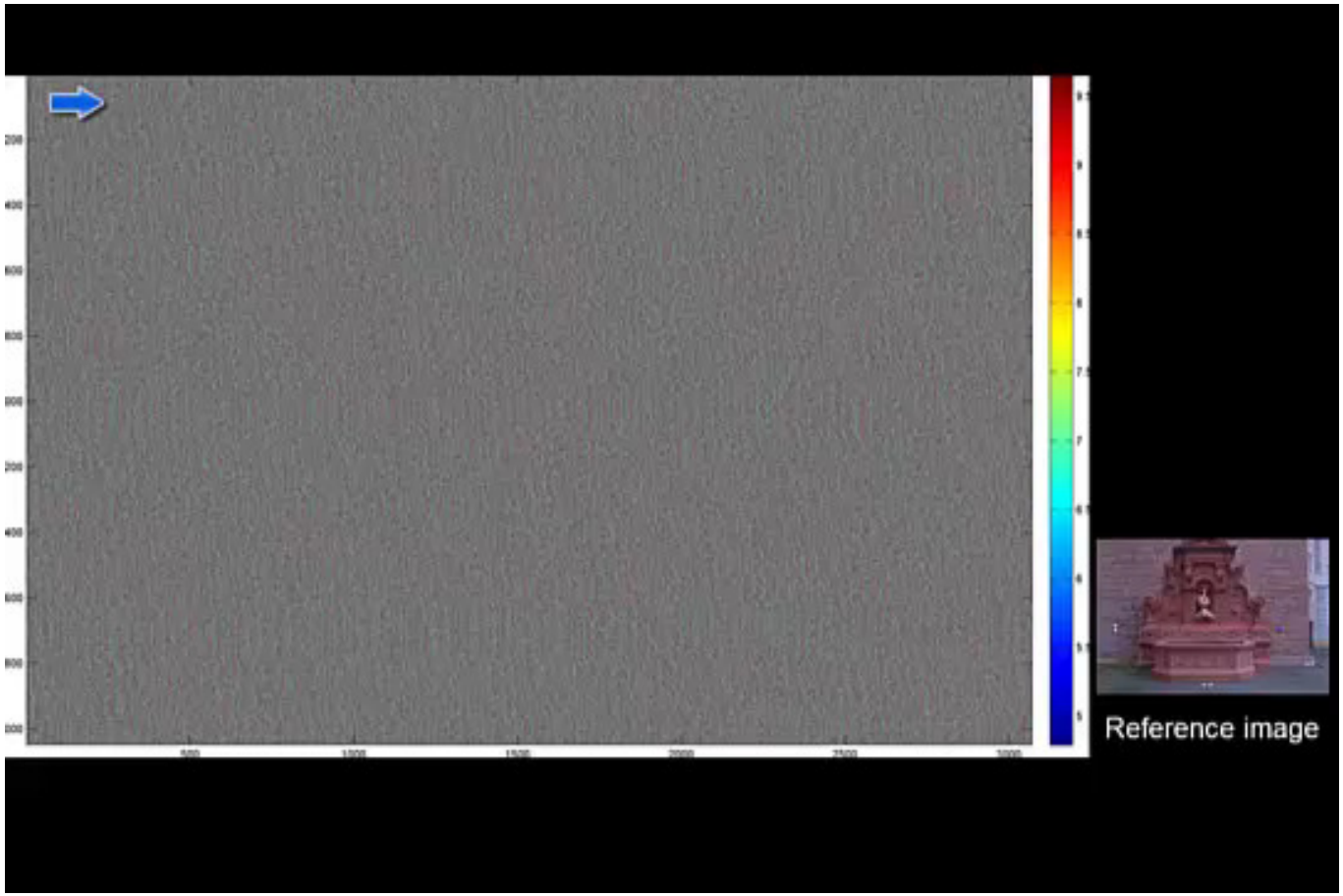
```
For Each Pixel
   Assign RandomDepth
For N iterations
   For Each Pixel
      Propagate Depth From Neighbor
      Sample new RandomDepth
      Update Depth
```



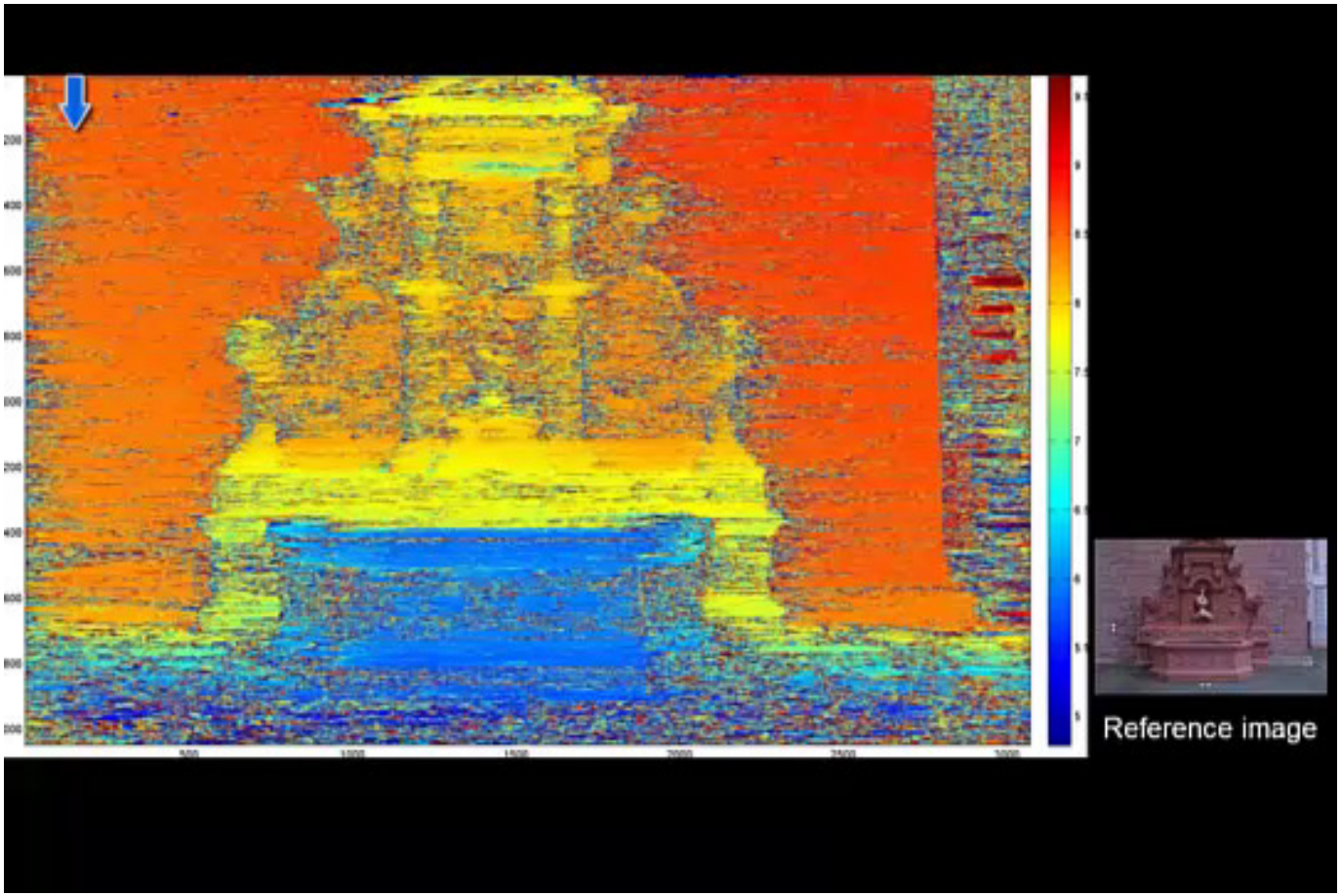
Sparse Search



Propagation Scheme



Propagation Scheme



PatchMatch Overview

- Pros:
  - + Efficient Search Space Exploration/Sampling
    - Example required only 8N+1 evaluations per pixel!!
  - + Good Empirical Convergence
- Cons:
  - Implicit regularization tends to over-smooth
  - Stochastic non-local search

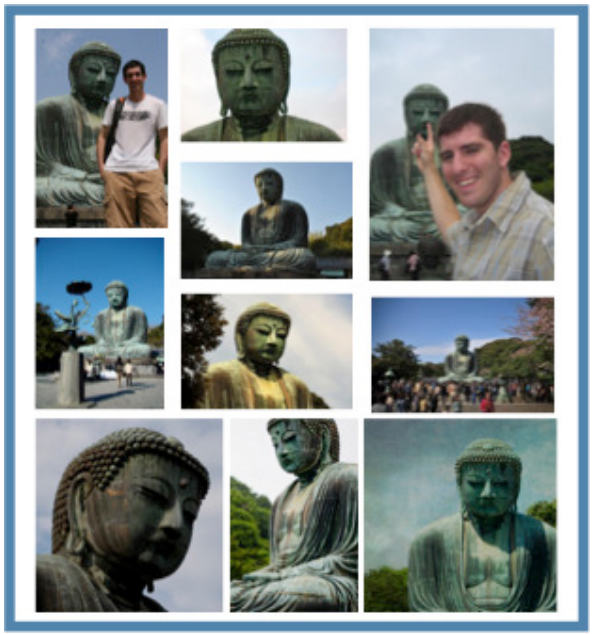
Outline

- Multi-View Stereo
- Robust Depthmap Estimation
  - Data Association
  - Estimating Surface normals
  - Joint Inference
- Fusion and Surface Estimation

Multi-view Data Association

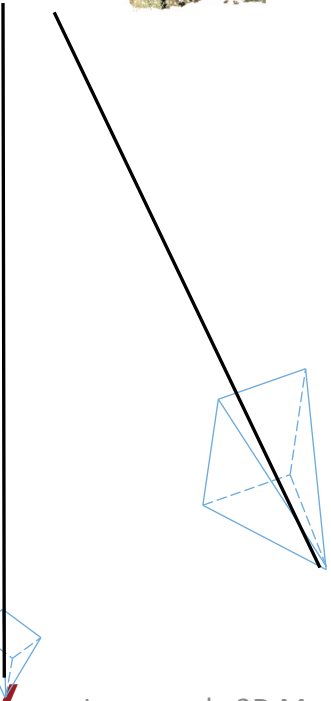
How to robustly integrate photo-consistency measurements from multiple views?

Reference image



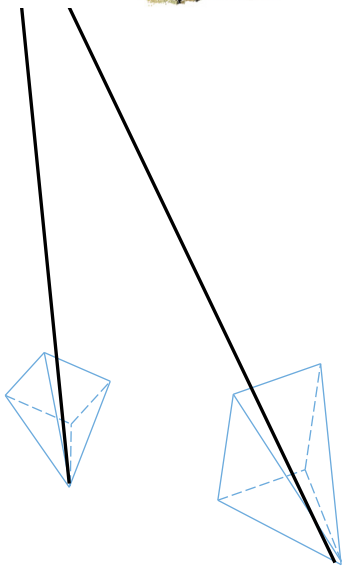
211 Source images (only 10 are shown)

Multi-view Data Association



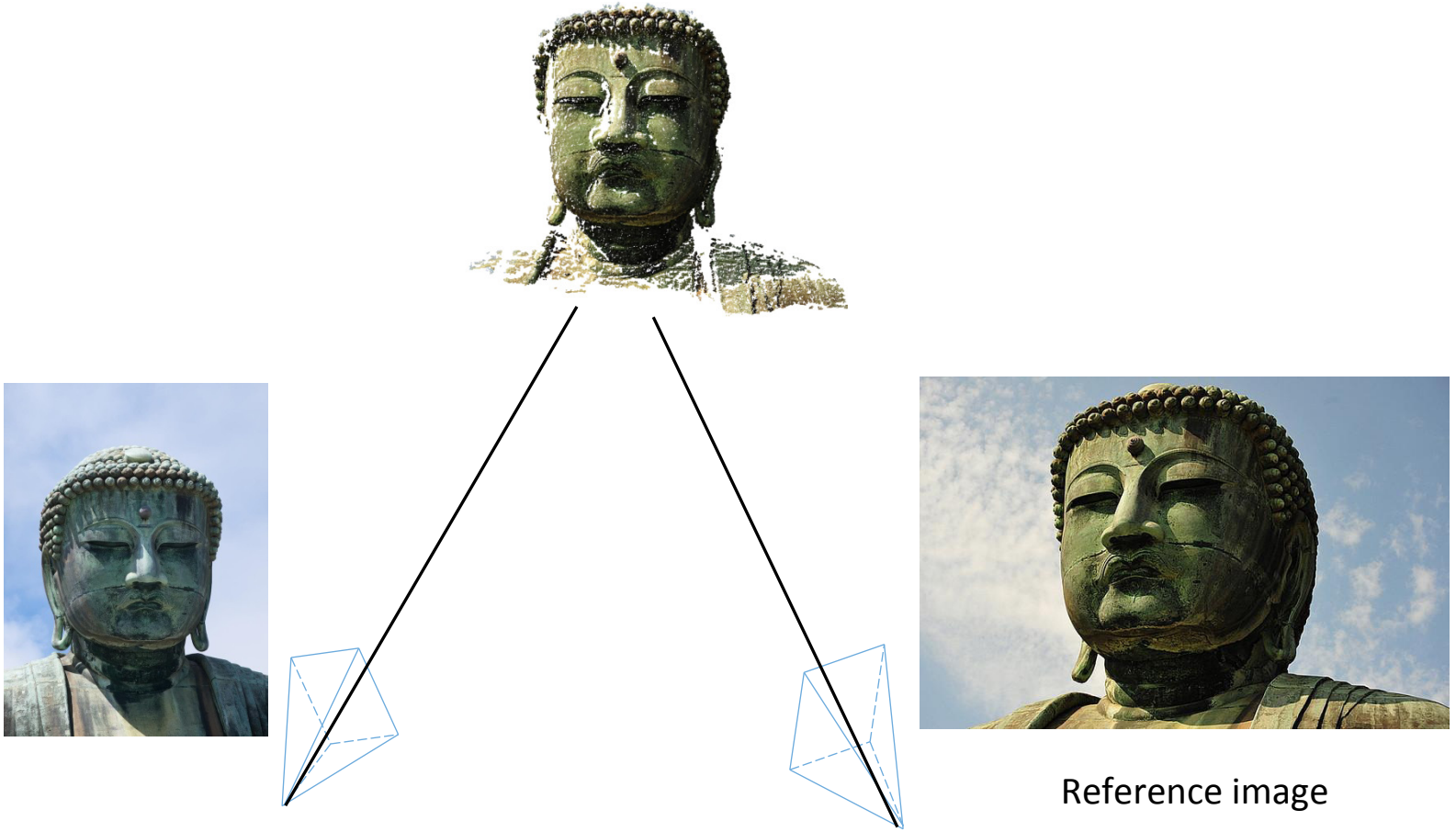
Reference image

Multi-view Data Association



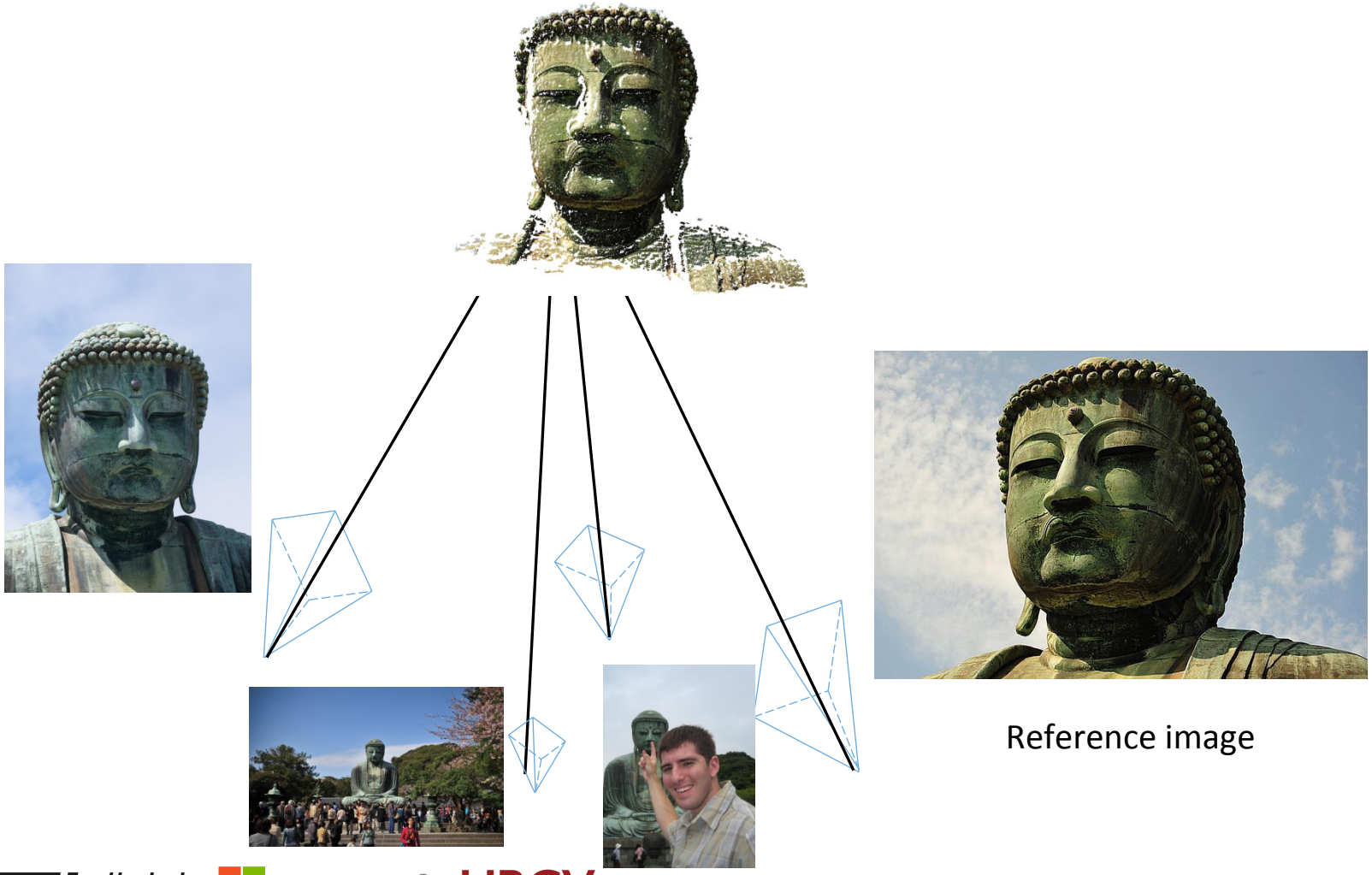
Reference image

Multi-view Data Association



Reference image

Multi-view Data Association

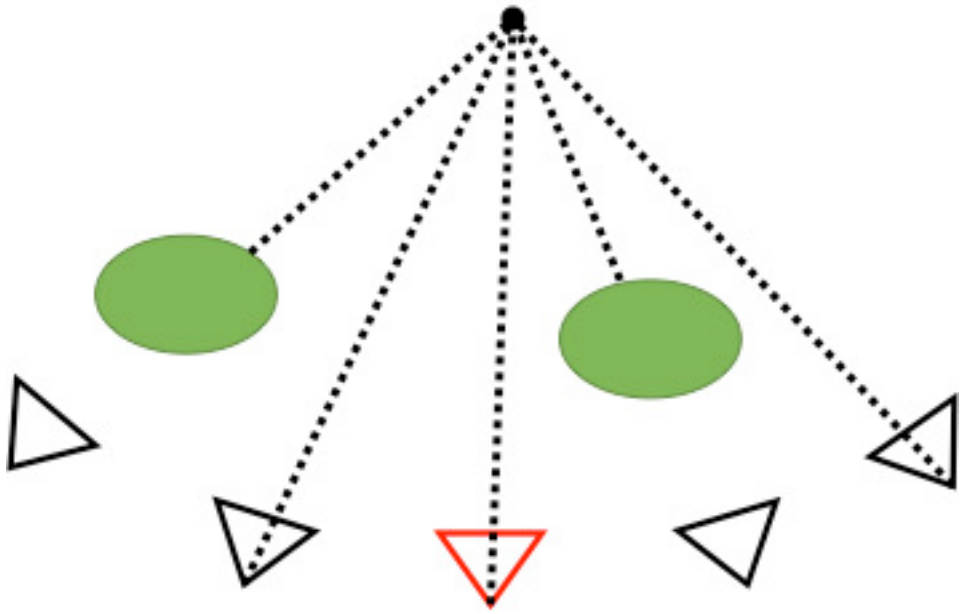


Reference image

Multi-view Data Association Strategies

How to robustly integrate photo-consistency measurements from multiple views?

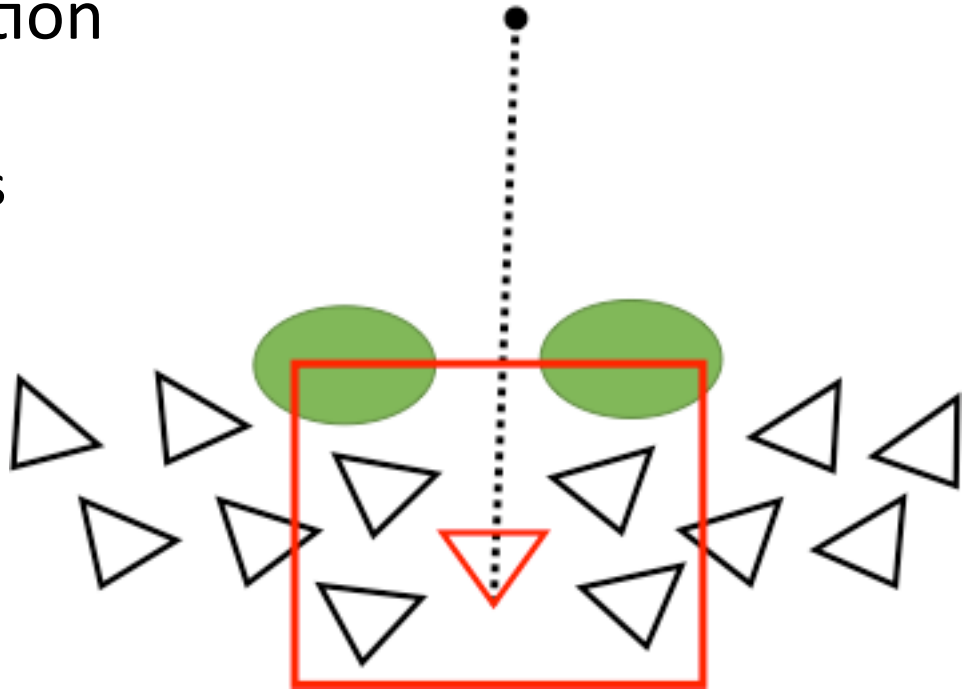
- Fixed Selection
  - Thresholding
  - Summary Statistics
    (e.g. mean, median)
  - Rank Statistics
    (e.g. Top-K)



Multi-view Data Association

How to robustly integrate photo-consistency measurements from multiple views?

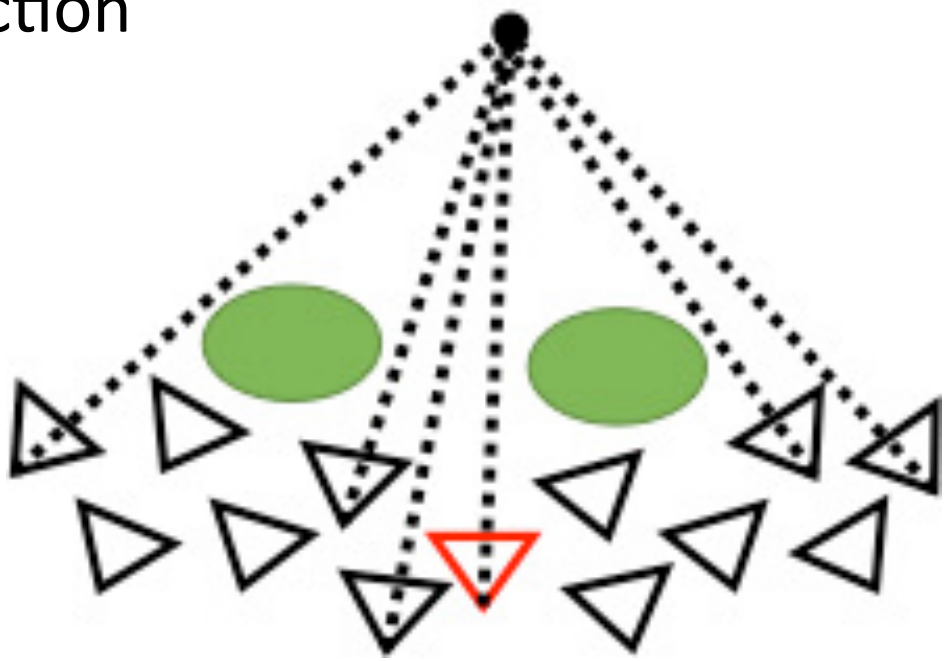
- Geometry-based Selection
  - Camera Proximity
  - Shared Sparse Features



Multi-view Data Association

How to robustly integrate photo-consistency measurements from multiple views?

- Appearance-Based Selection
  - Global or Per-Pixel?



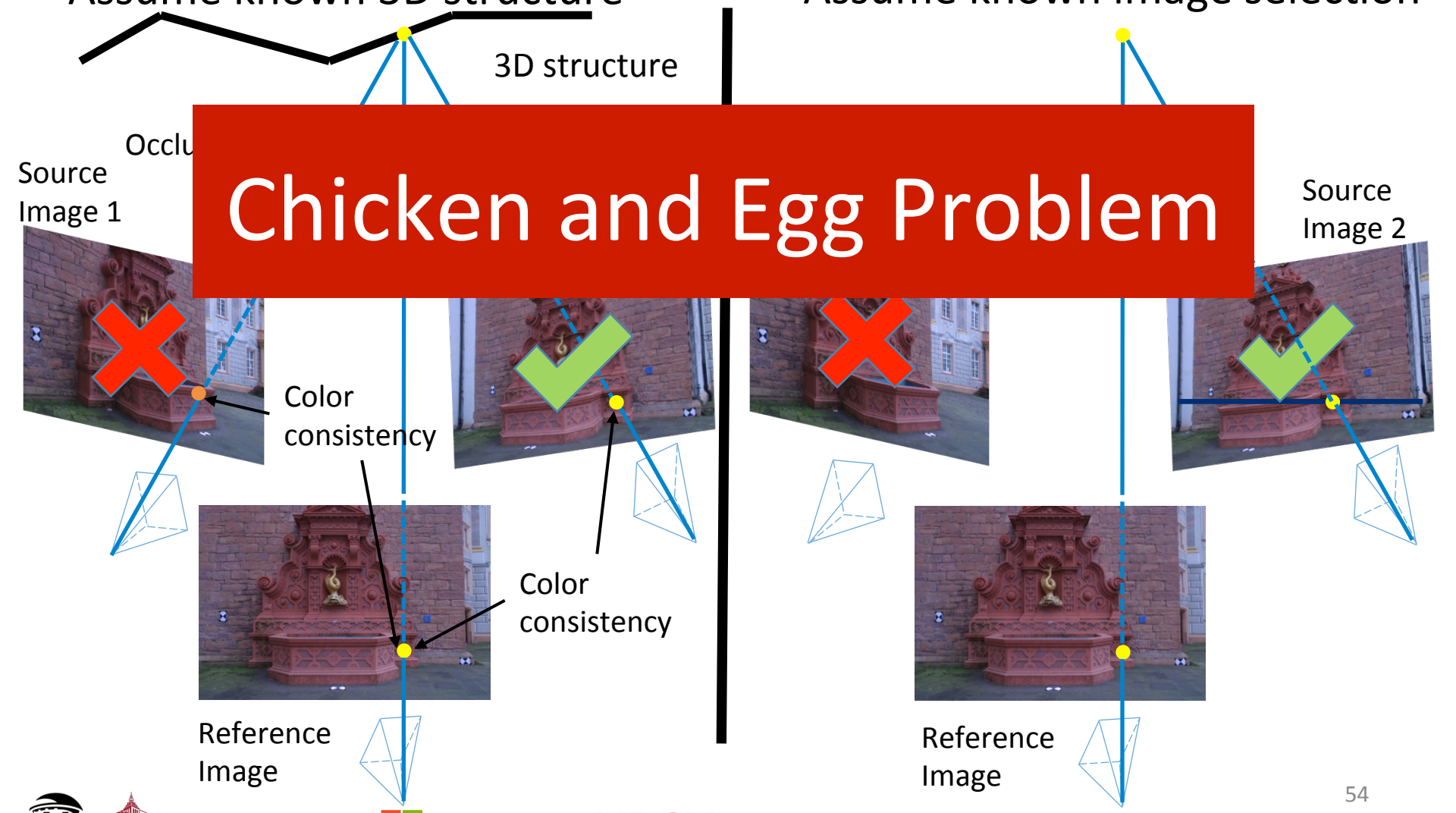
Pixel-level image selection

Pixel-level image selection

Pixel-level image selection

Pixel-level image selection

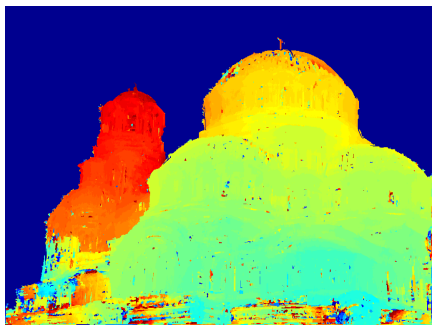
Image Selection vs Depth Estimation

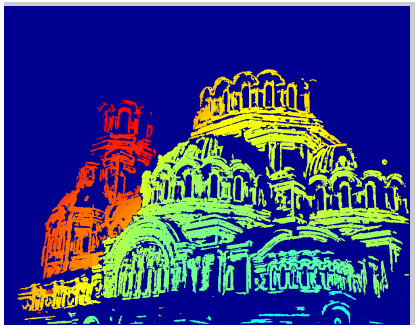


Robustness of pixel-wise selection



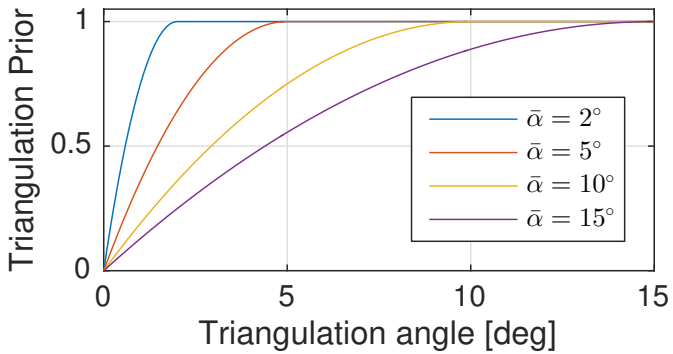


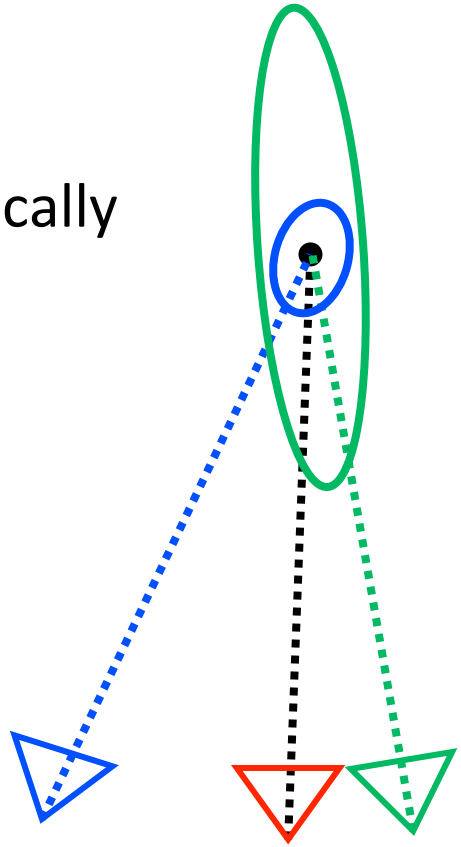


Multi-View Stereo

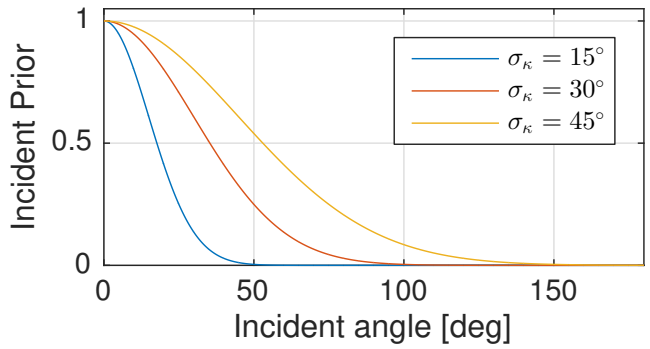
- Pixelwise Monte-Carlo view sampling
  - Visibility probability derived photometrically
  - Incorporate geometric selection priors
    - Triangulation angle prior

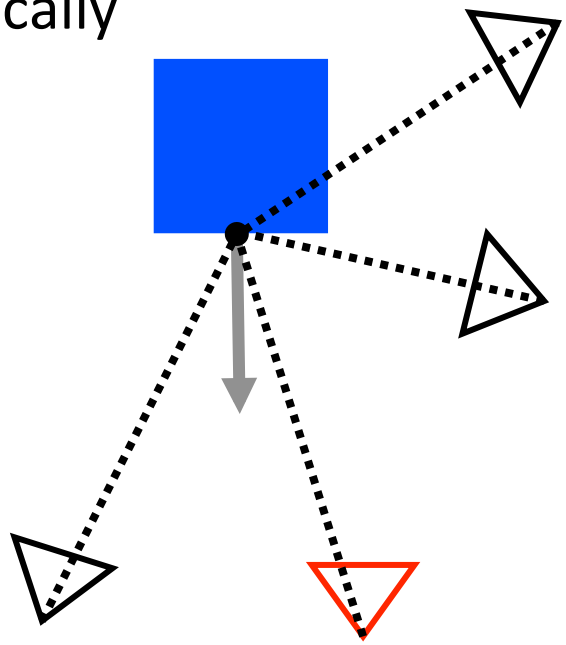




Multi-View Stereo

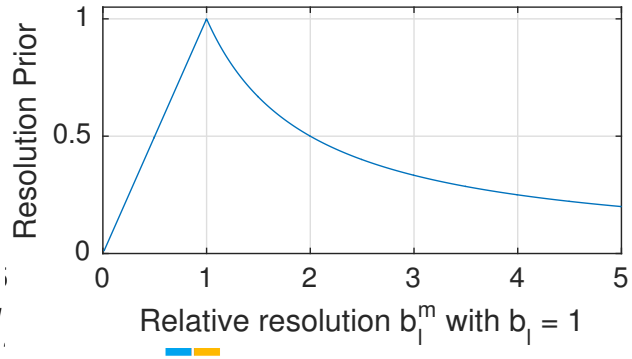
- Pixelwise Monte-Carlo view sampling
  - Visibility probability derived photometrically
  - Incorporate geometric selection priors
    - Triangulation angle prior
    - Incident angle prior

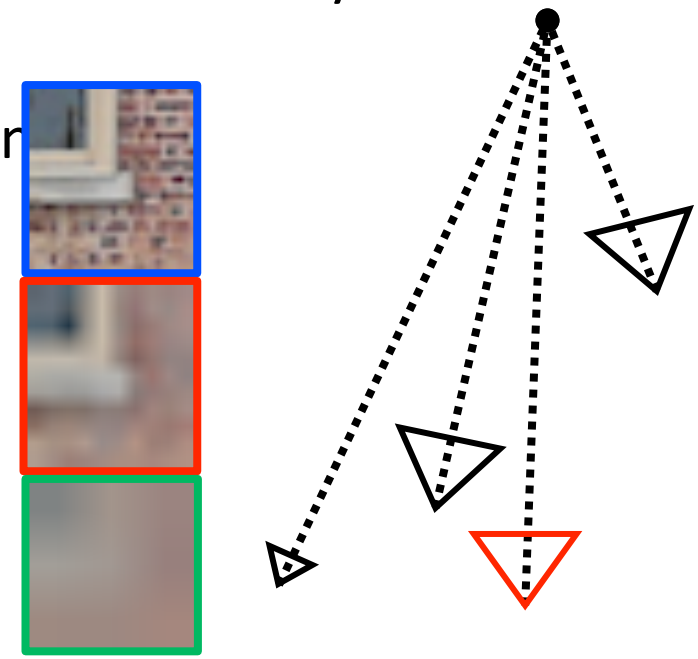




Multi-View Stereo

- Pixelwise Monte-Carlo view sampling
  - Visibility probability derived photometrically
  - Incorporate geometric selection
    - Triangulation angle prior
    - Incident angle prior
    - Resolution prior





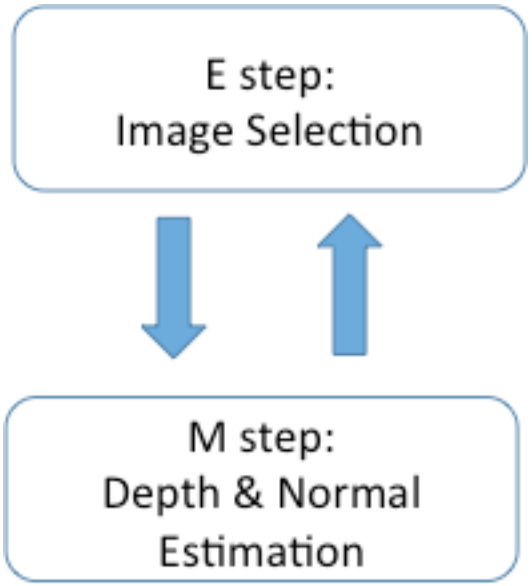
Joint likelihood Estimation

$$P(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{N})$$

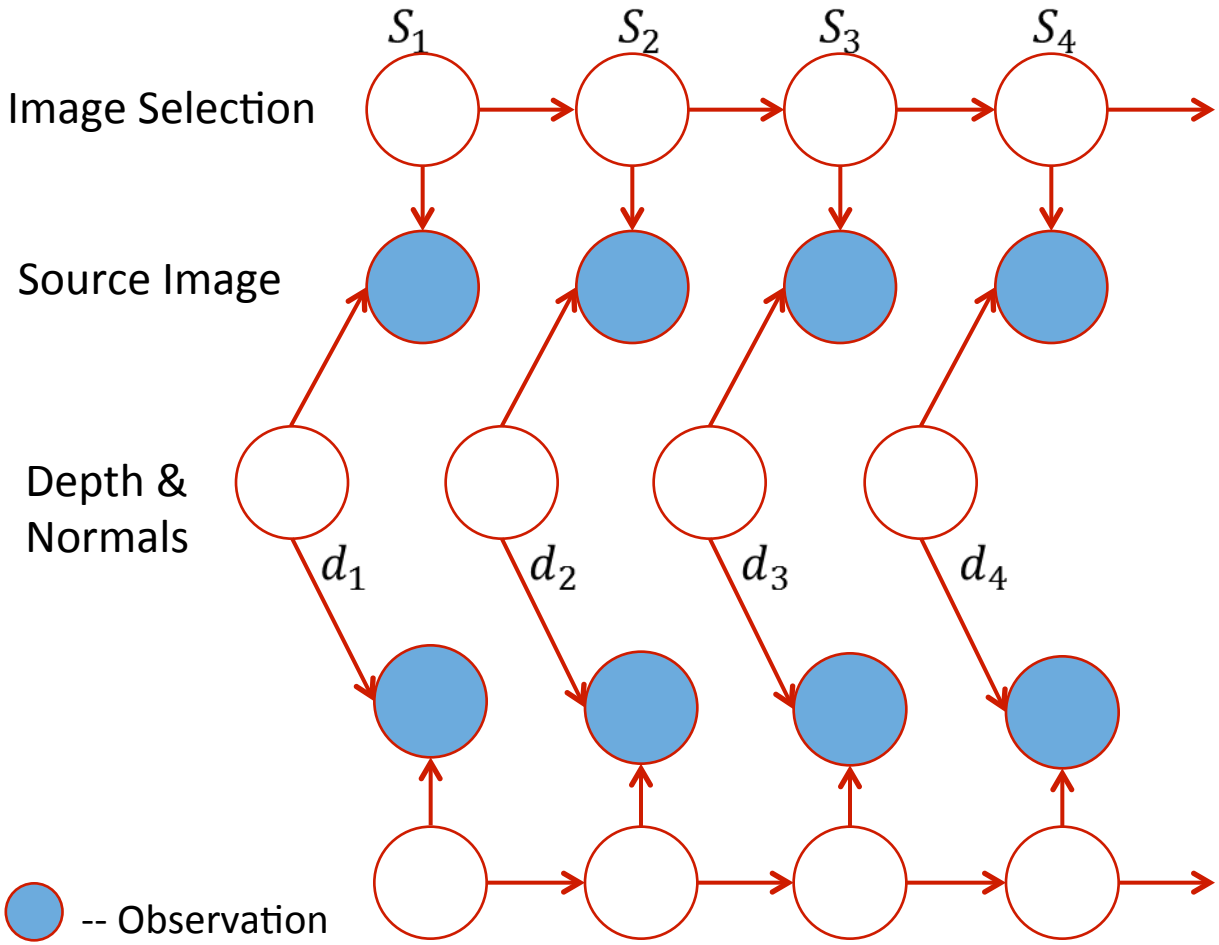
Normals
Depth
Occlusion
Images

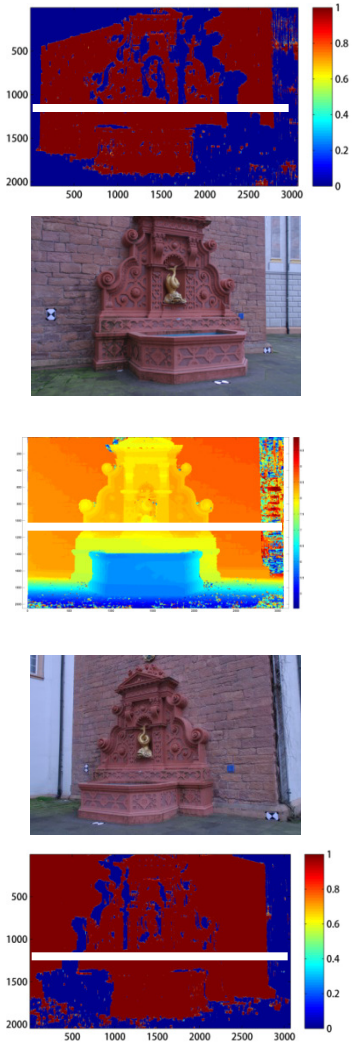
JL. Schönberger, E Zheng, M Pollefeys, J-M Frahm "Pixelwise View Selection for Unstructured Multi-View Stereo" ECCV 2016

- Generalized Expectation Maximization
  - E-Step
    - Infer $\boldsymbol{Z}$ using variational inference
  - M-Step
    - Infer $\boldsymbol{\theta}$, $\boldsymbol{N}$ using PatchMatch sampling



Joint Inference

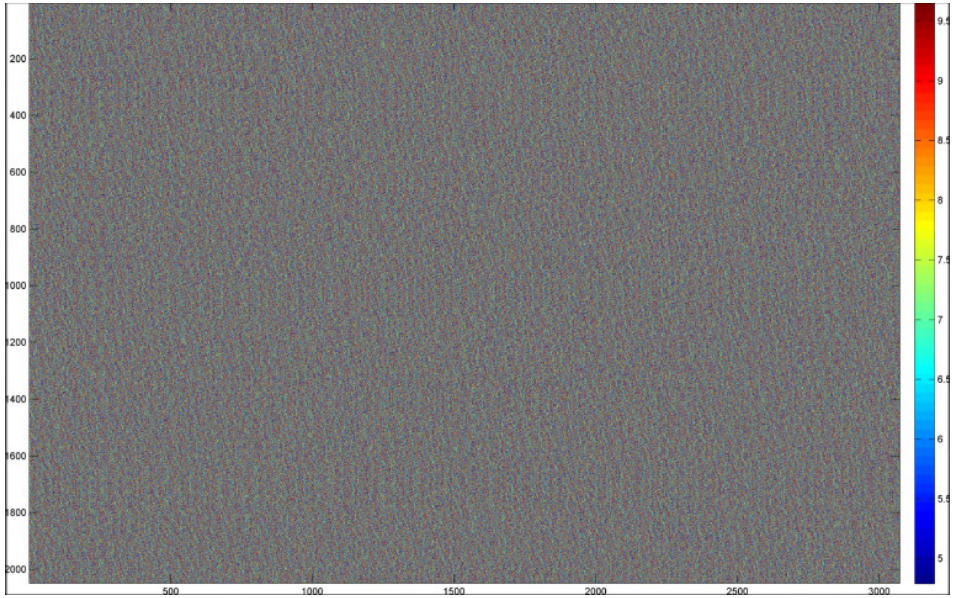




Joint Inference

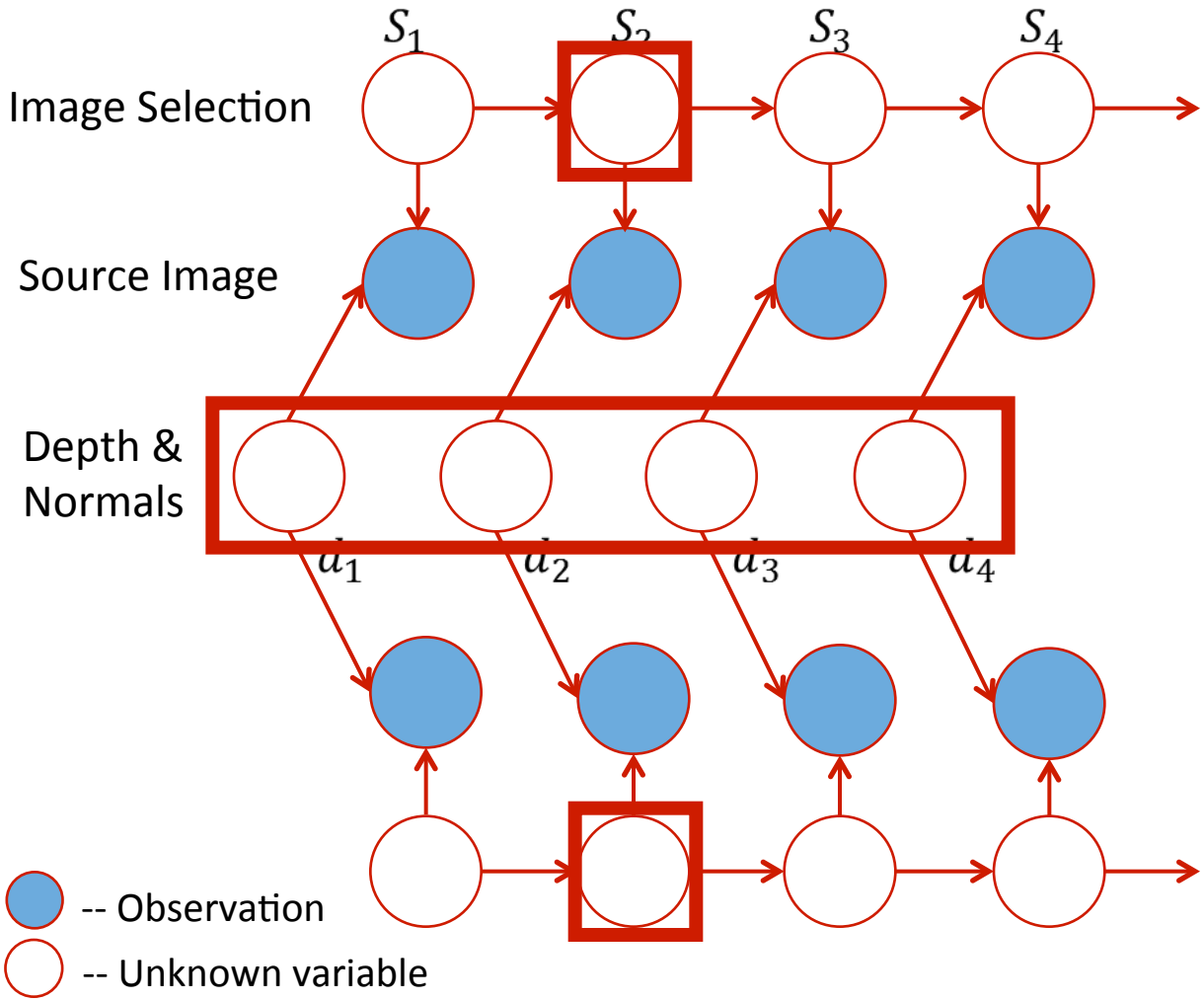
- Random depthmap initialization

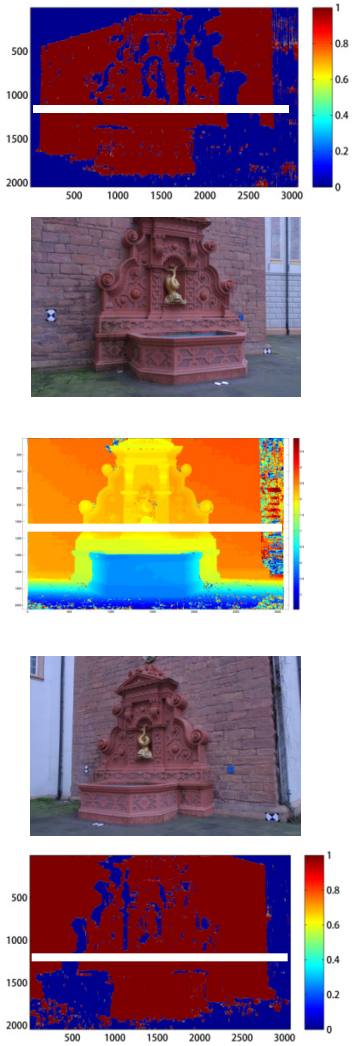
Reference image



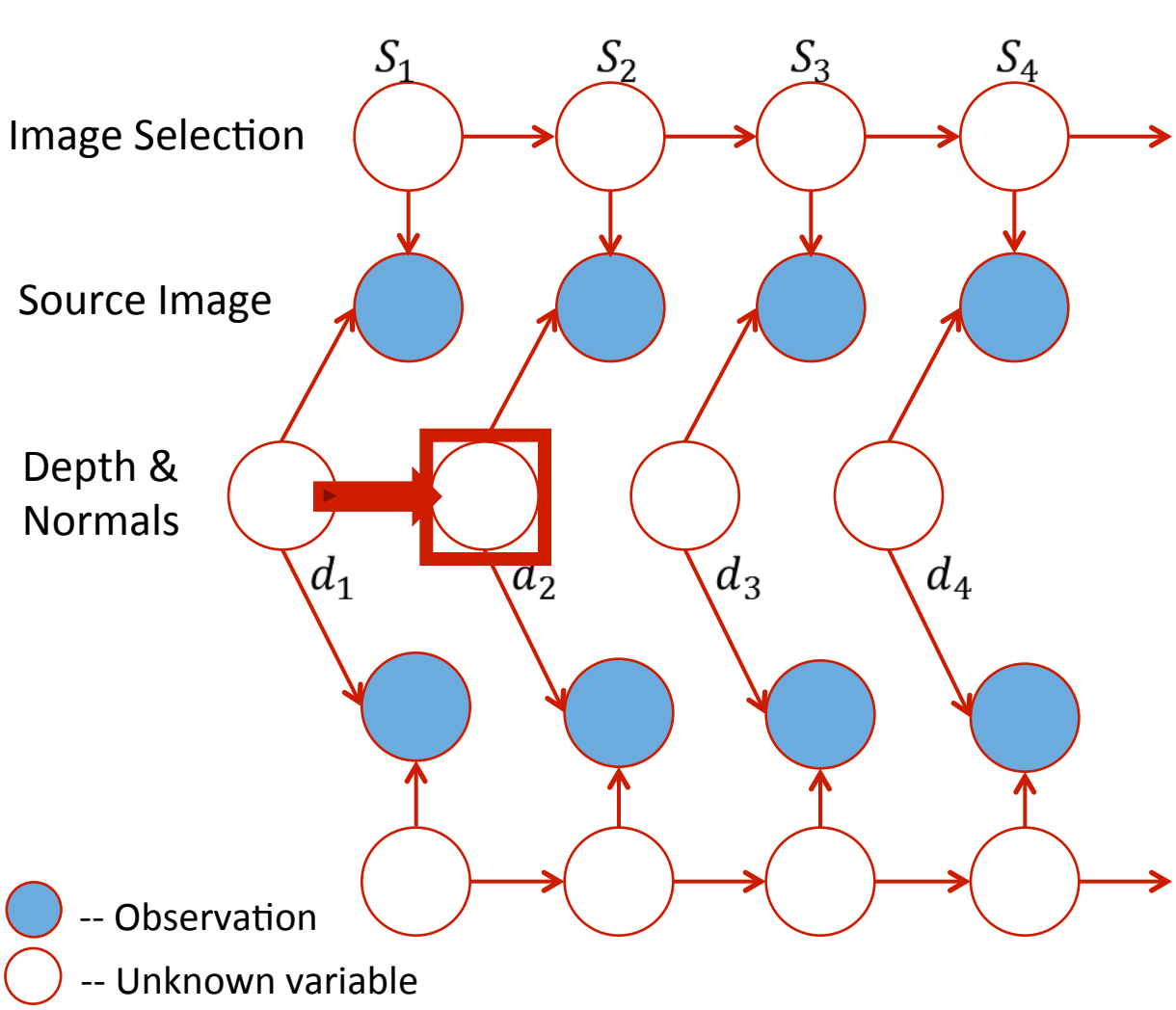
Random depthmap

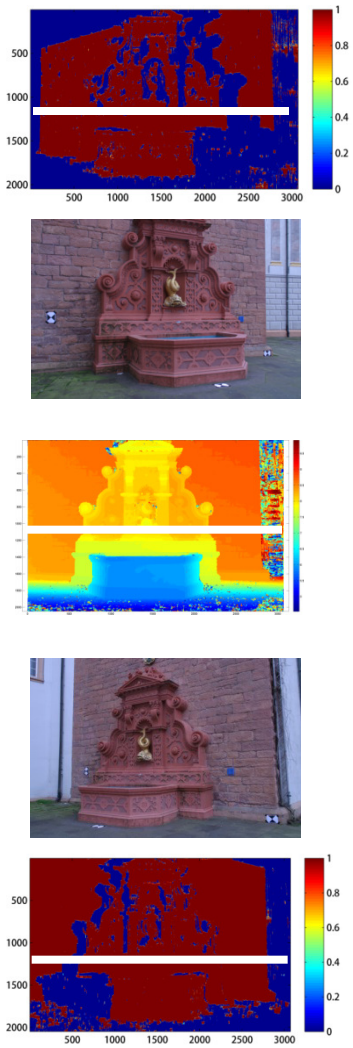
Joint Inference



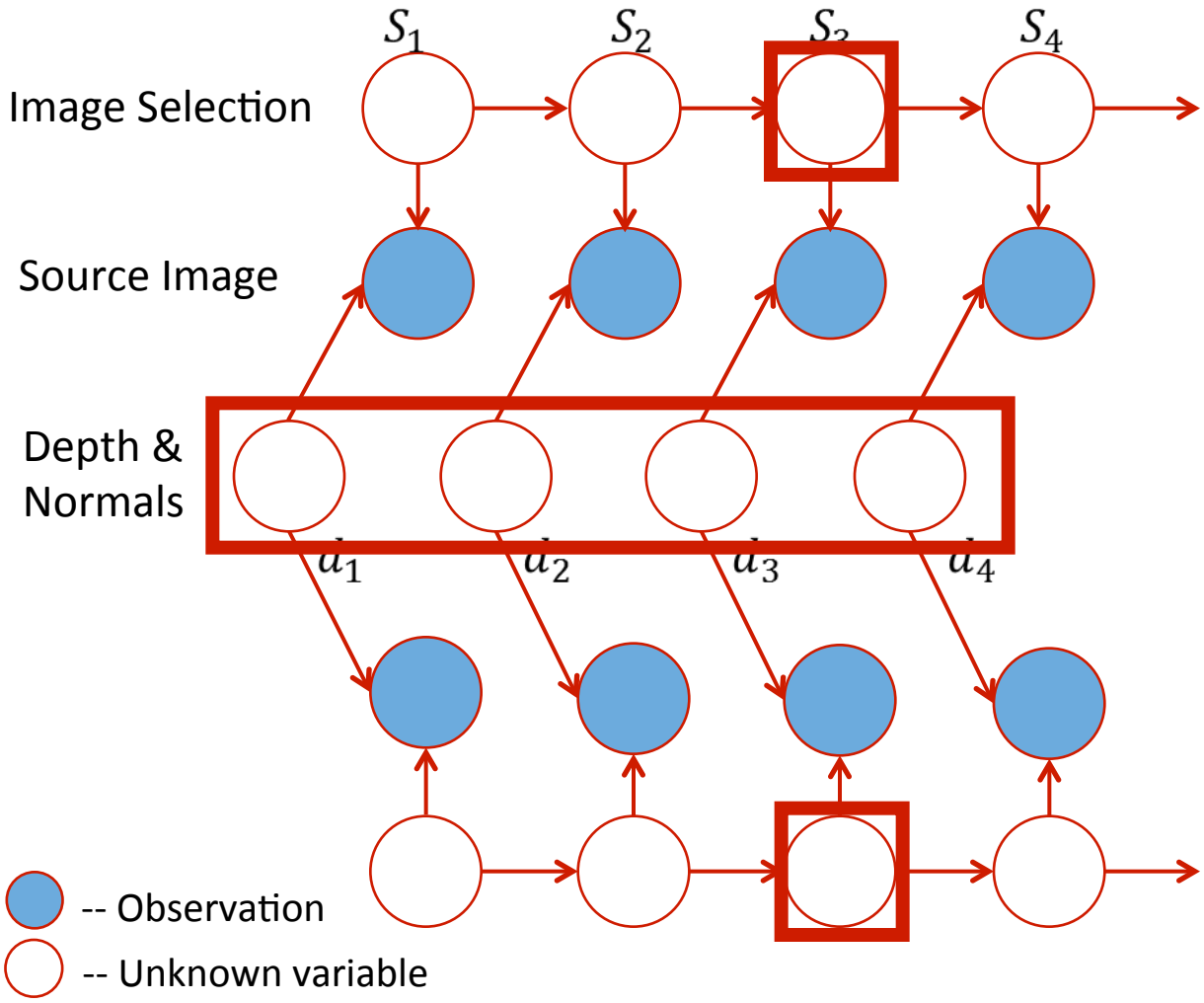


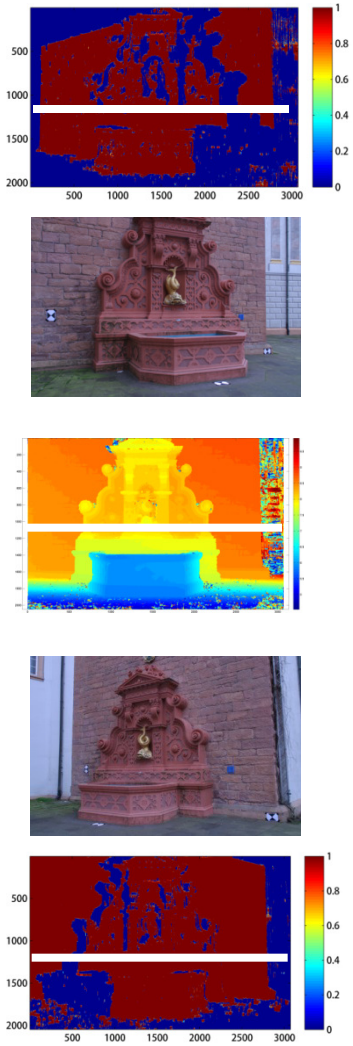
Joint Inference



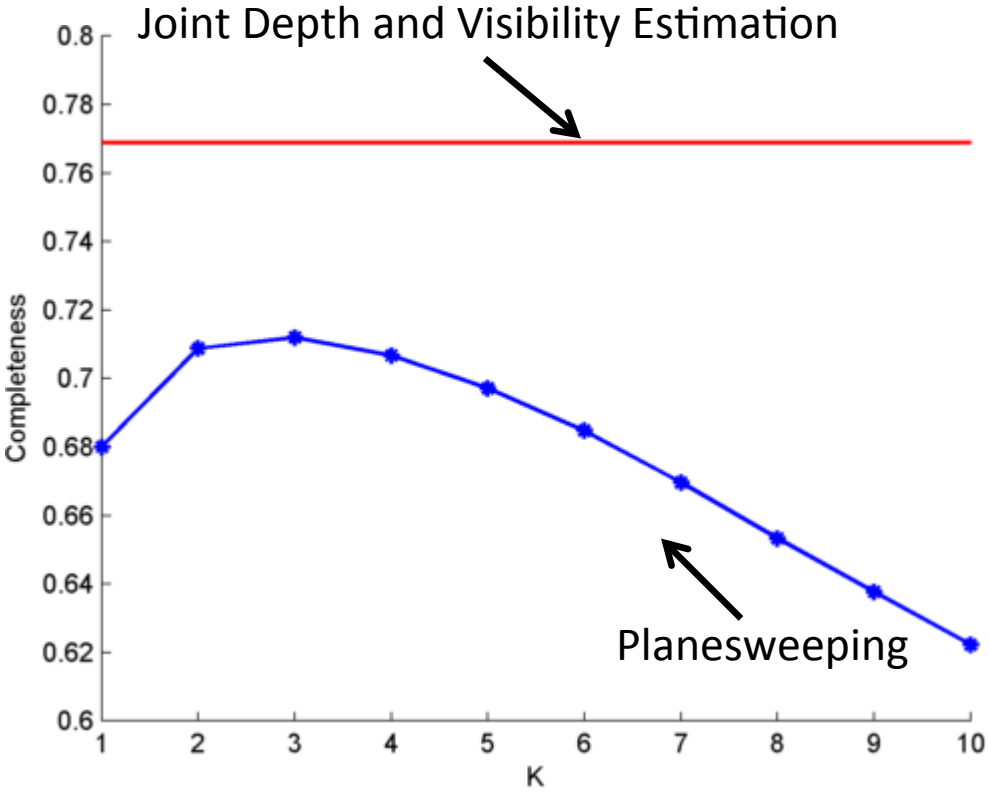


Joint Inference

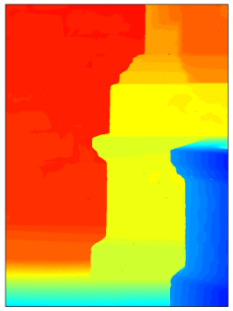




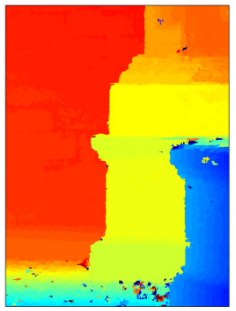
Results vs. best-K Stereo



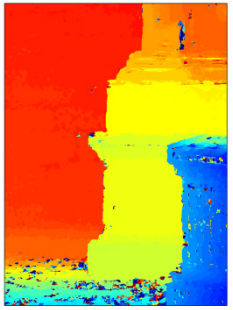
Completeness: percentage of pixels with errors less than 2 cm



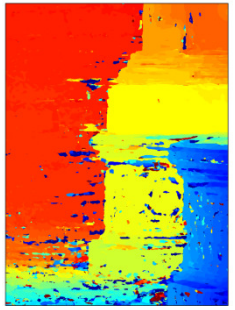
Ground truth



Joint Estimation

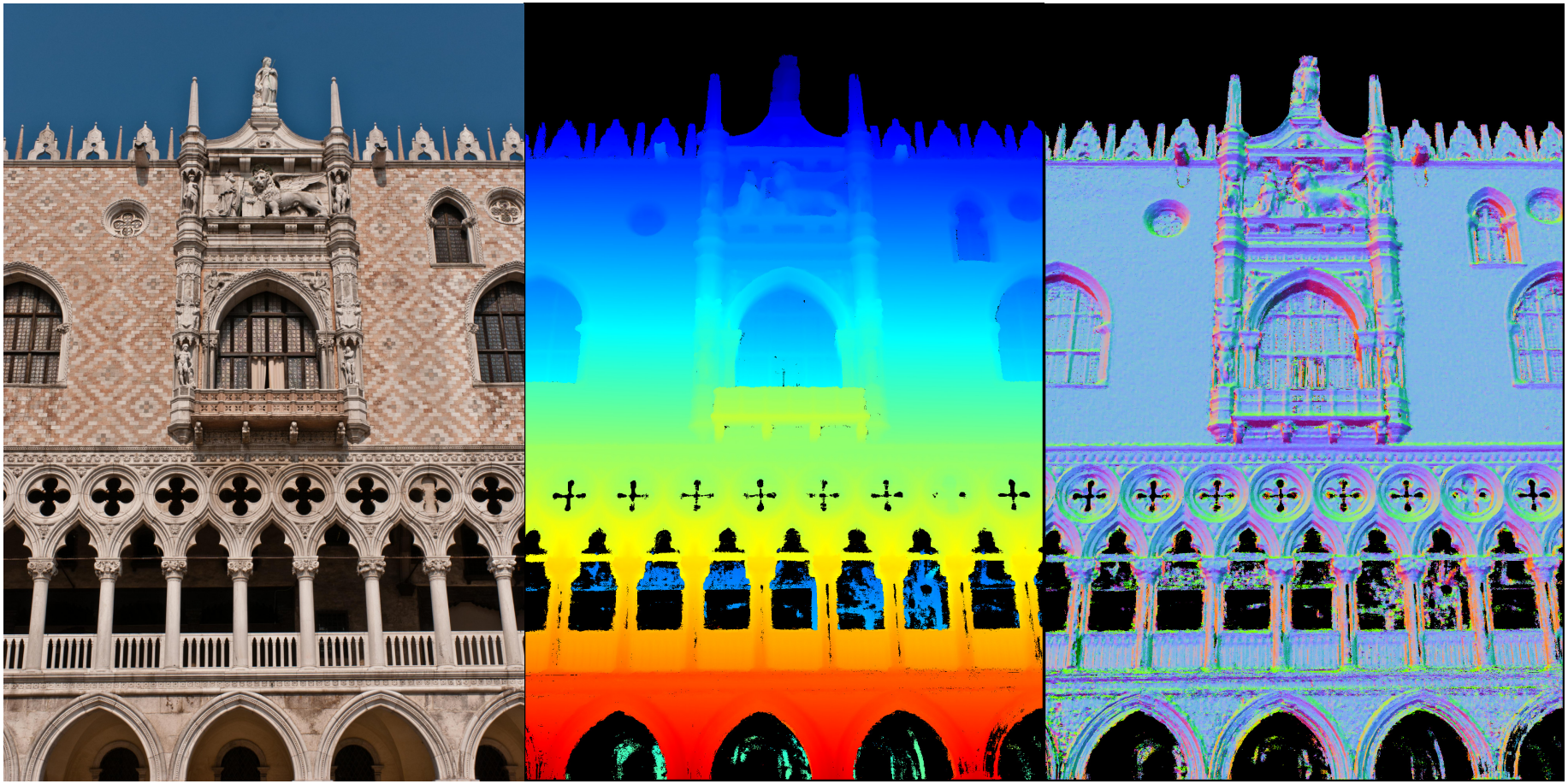


K=3

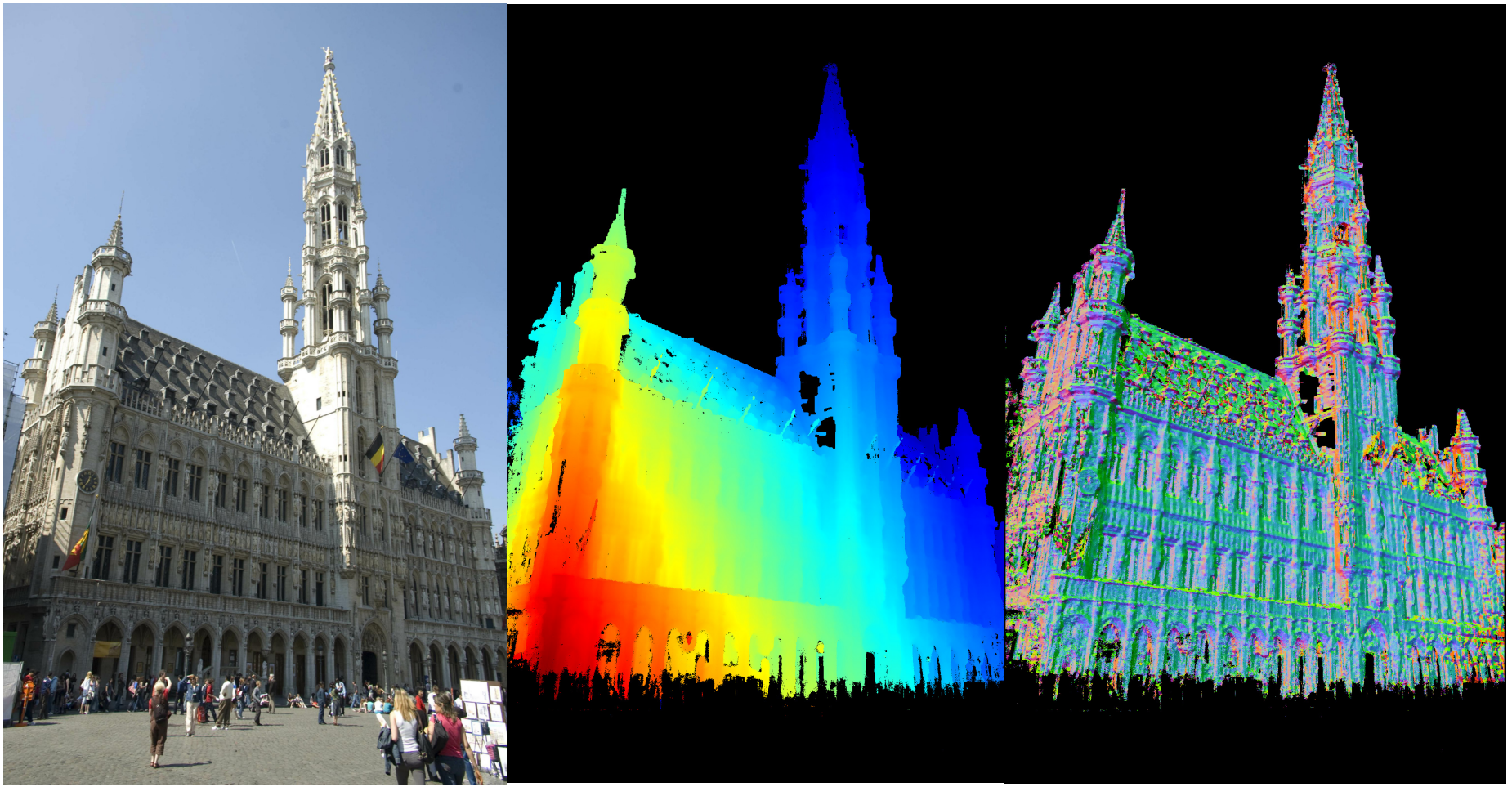


K=10

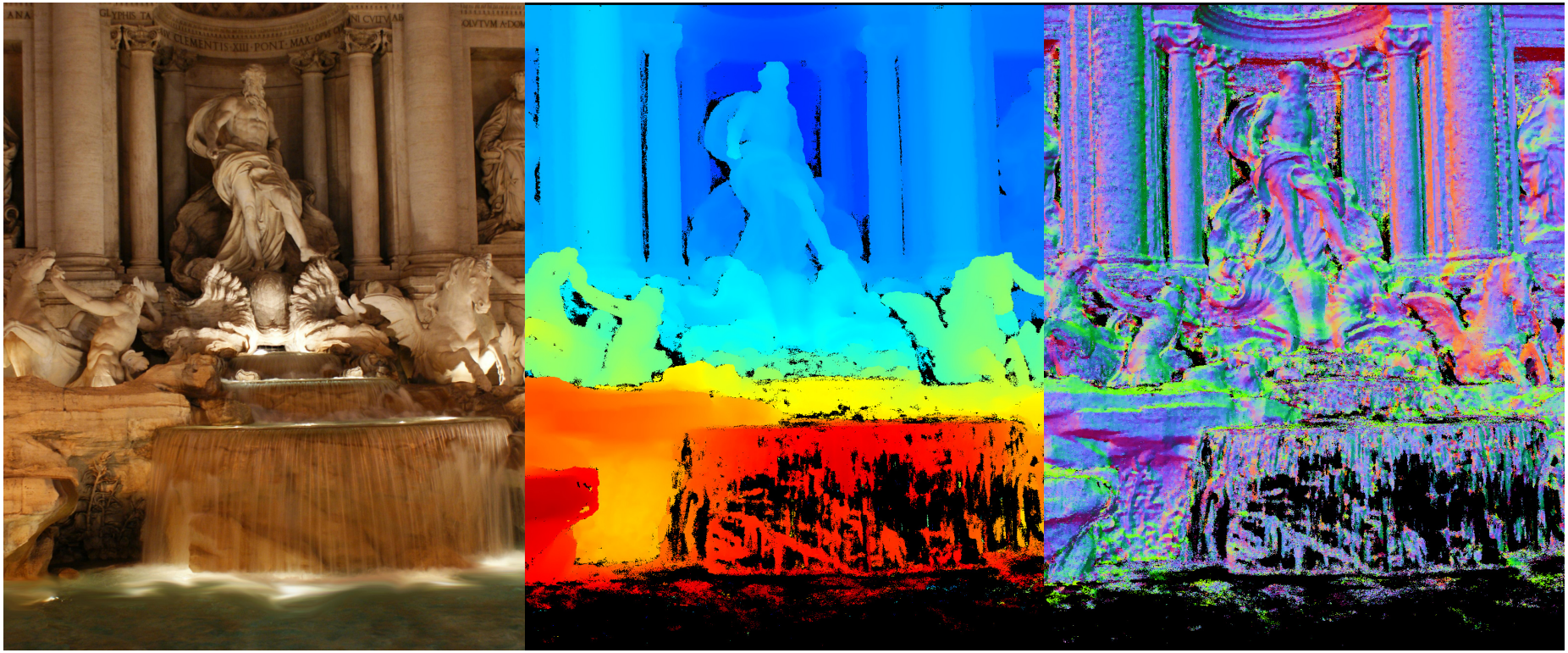
Results



Results



Results

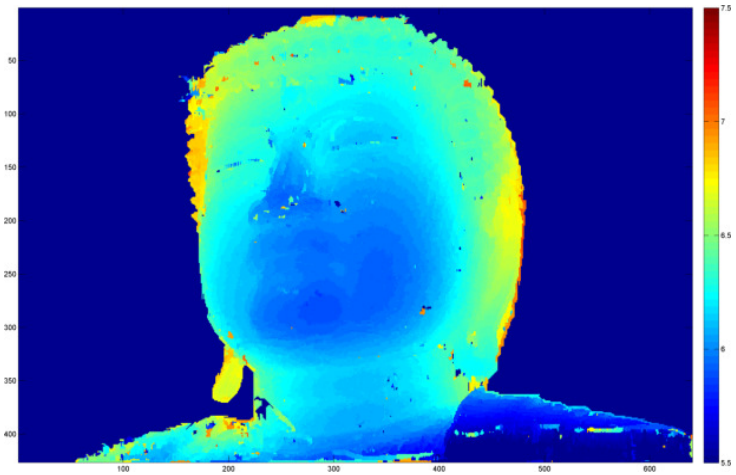


Outline

- Multi-View Stereo
- Robust Depthmap Estimation
- Fusion and Surface Estimation
  - Depthmap integration
  - Surface Approximation

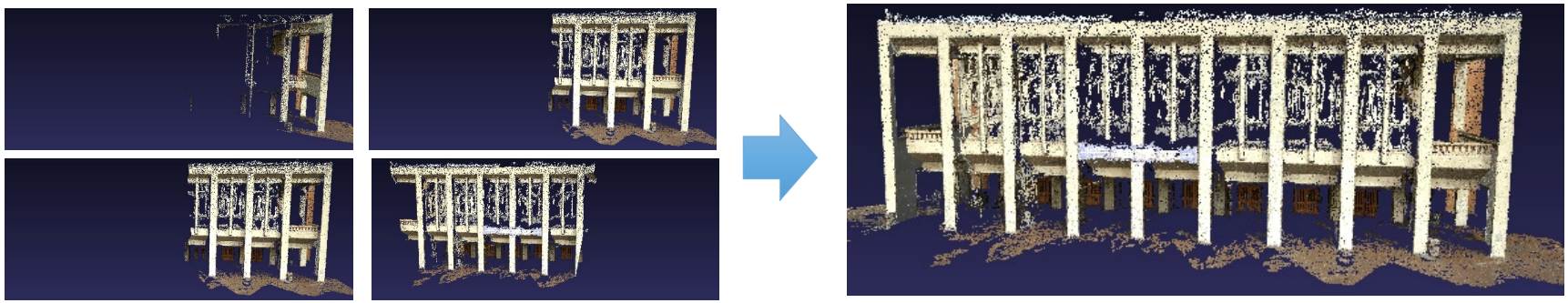
Depthmap Fusion

- Depthmaps as point clouds



Depthmap Fusion

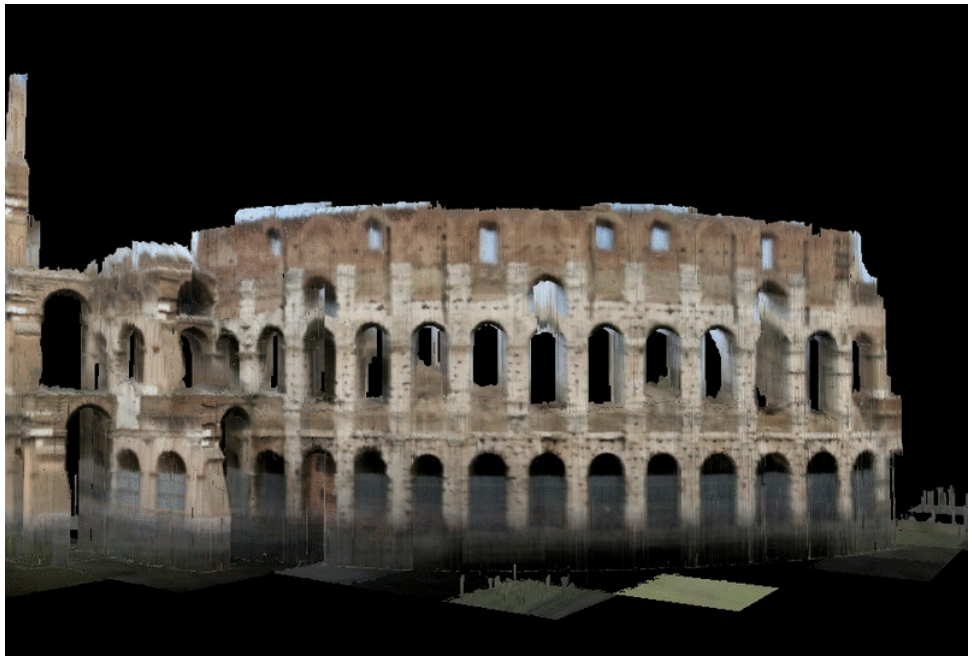
- Direct Aggregation



- Challenges
  - Error/inconsistency mitigation
  - Completeness

Depthmap Fusion

- Goal: Mitigate depthmap measurement errors by leveraging observation redundancies, to provide an aggregated and augmented 3D representation.



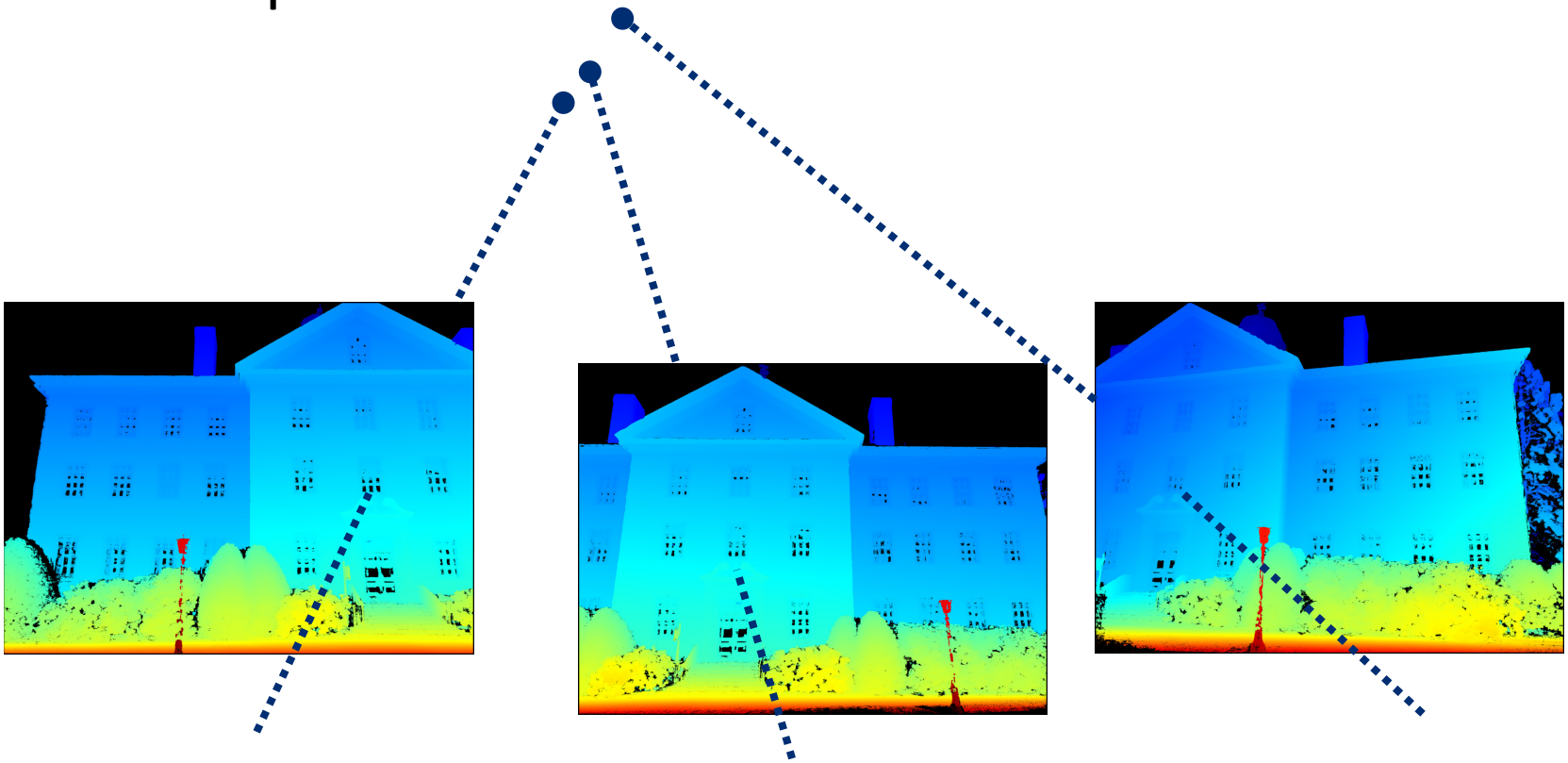
Depthmap Fusion Approaches

- Depthmap Refinement:
  - Local analysis of relationships among 3D points
  - Filter cross-depthmap inconsistencies

- Depthmap Integration:
  - Global analysis based on auxiliary geometric structures
  - Robustly Estimate Surface Parameters

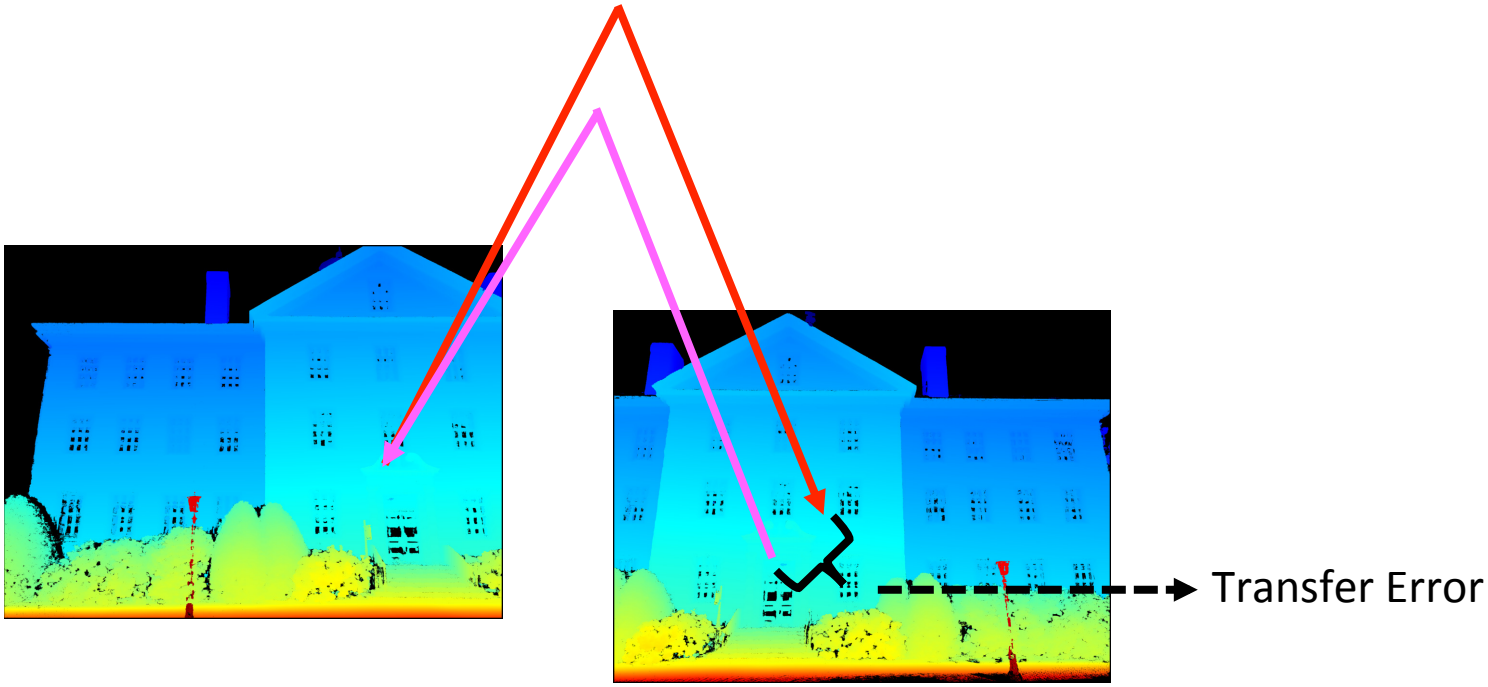
Enforcing Depthmap Consistency

- In 3D Space



Enforcing Depthmap Consistency

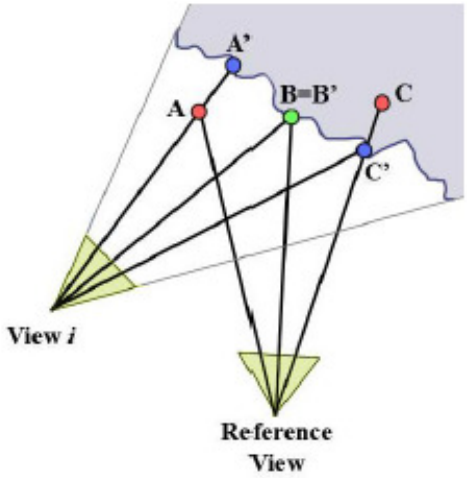
- Two-view Point Transfer



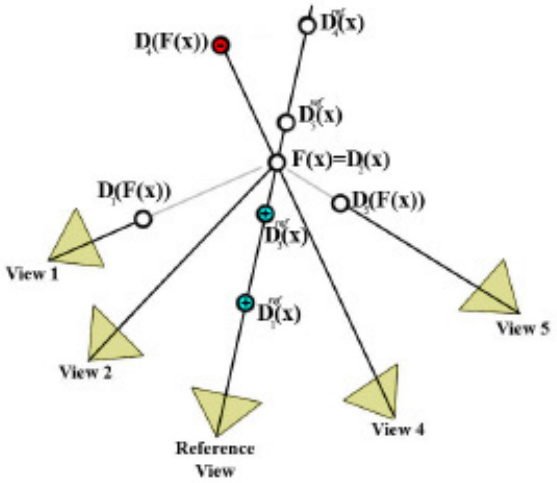
Enforcing Visibility Constraints

- Identify and penalize free space violations

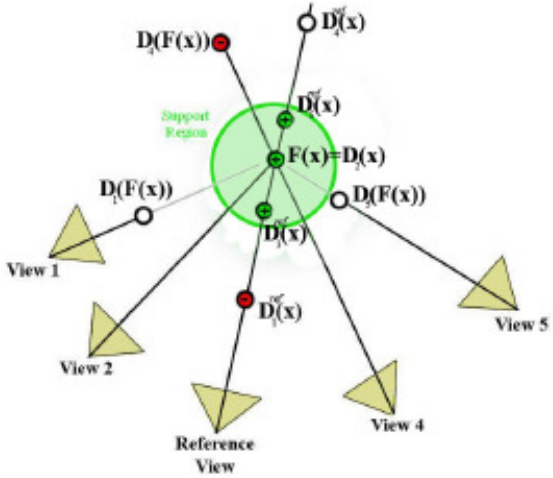
Merrel et al. ICCV 2007



(a) Visibility relations between points



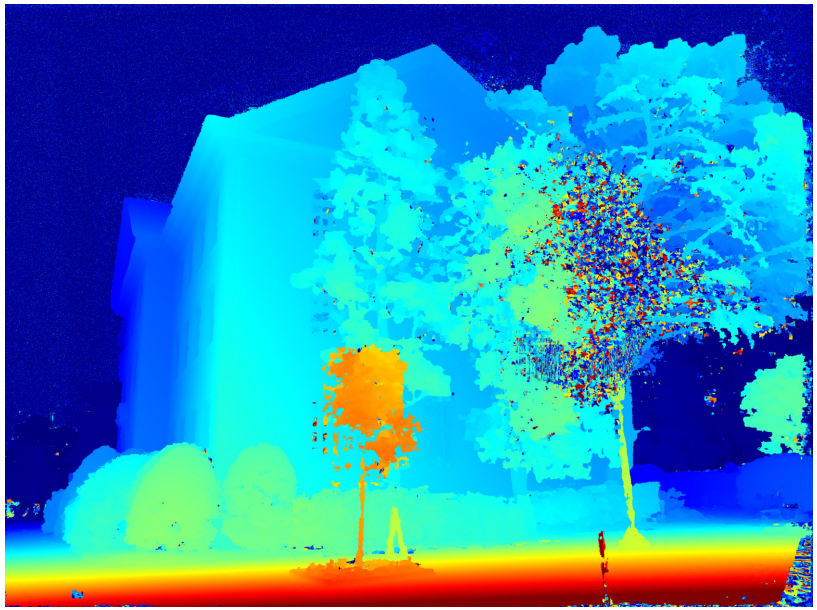
(b) Stability calculation

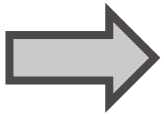


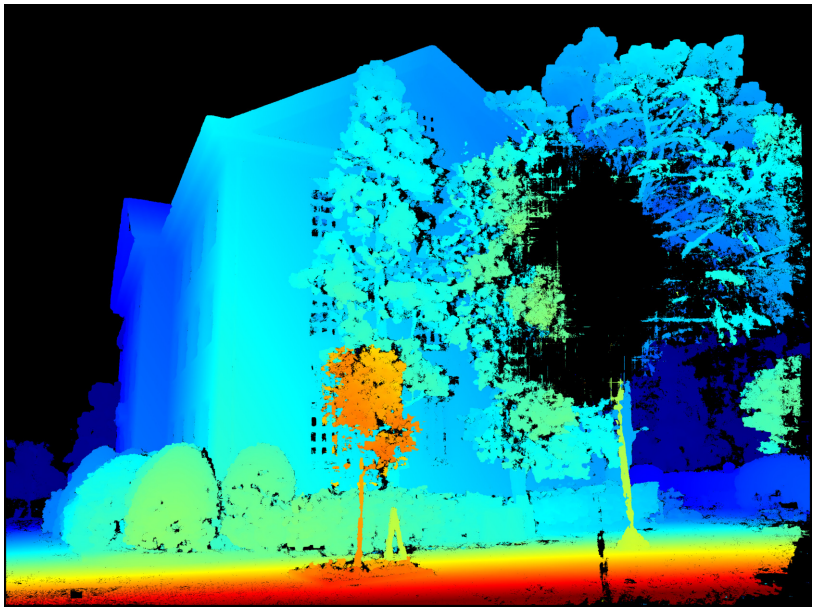
(c) Support estimation

Depthmap Filtering

- Filtering Result







Results





Results

Results

Poisson Surface Reconstruction

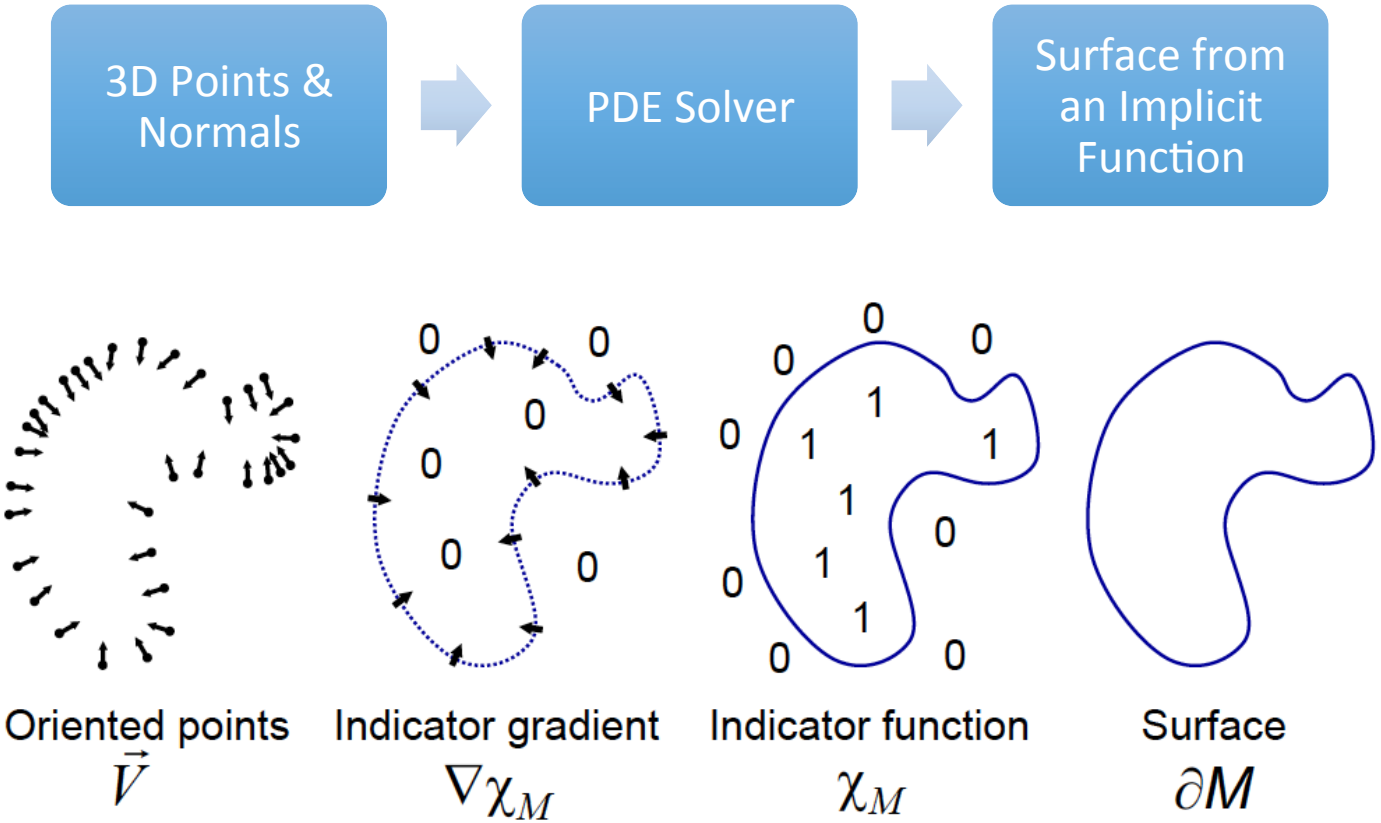
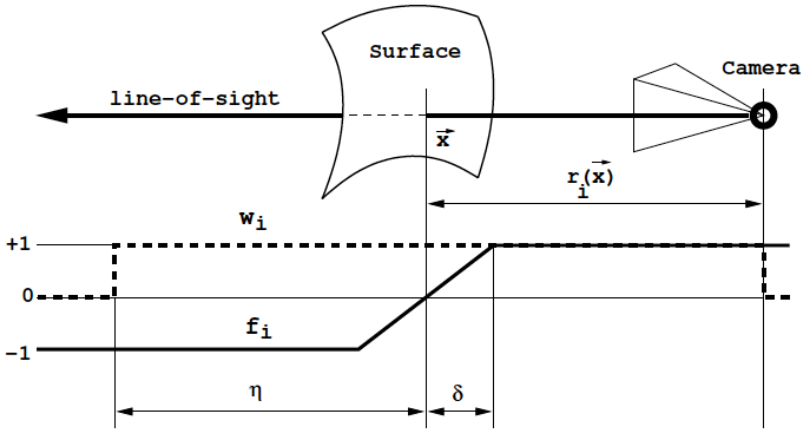
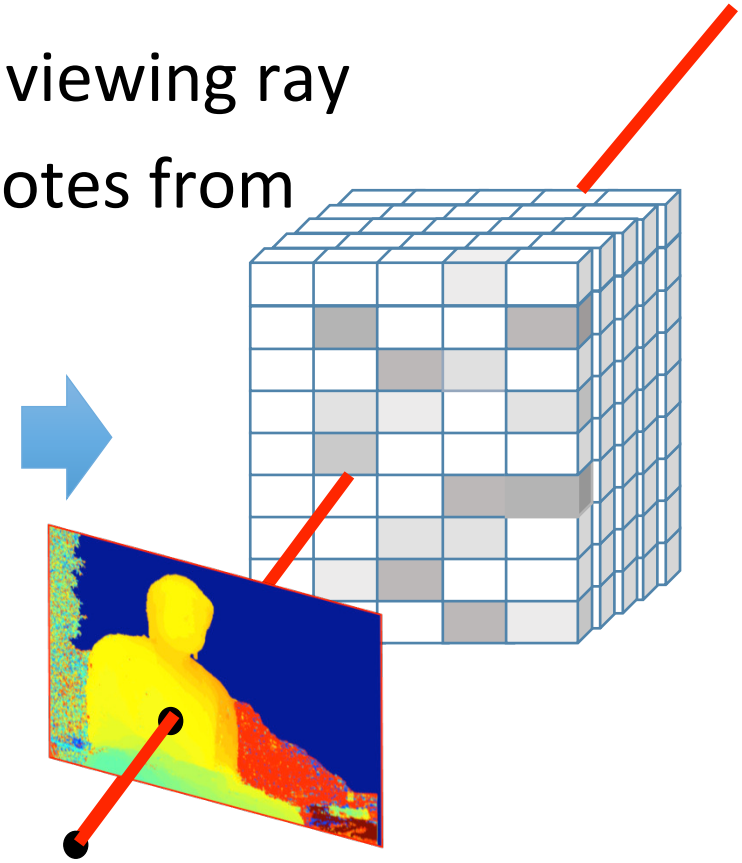


Figure from: Kazhdan et al. / Poisson Surface Reconstruction

Spatial Aggregation

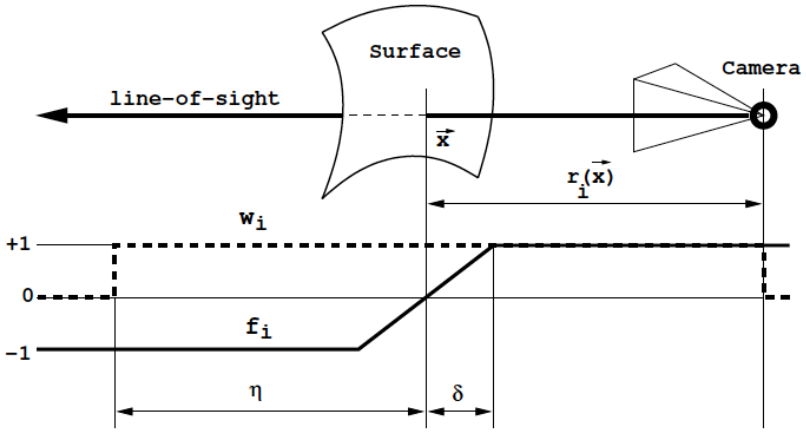
- Volumetric Scene Representation
- Each pixel votes along its viewing ray
- Each voxel accumulates votes from all intersecting rays

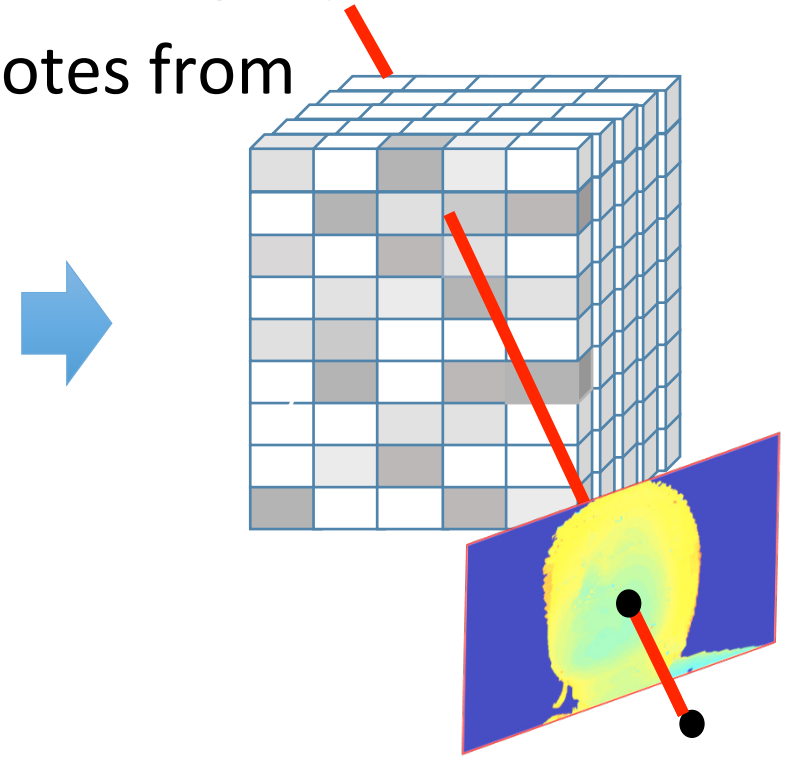




Spatial Aggregation

- Volumetric Scene Representation
- Each pixel votes along its viewing ray
- Each voxel accumulates votes from all intersecting rays



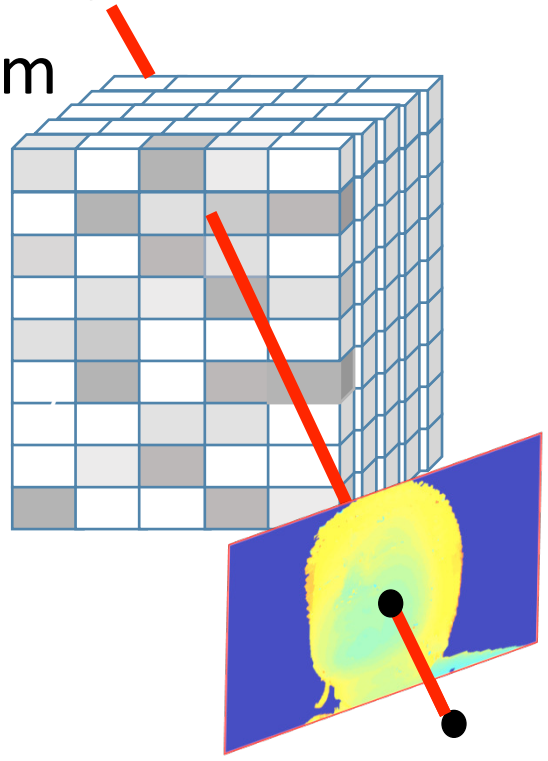


Spatial Aggregation

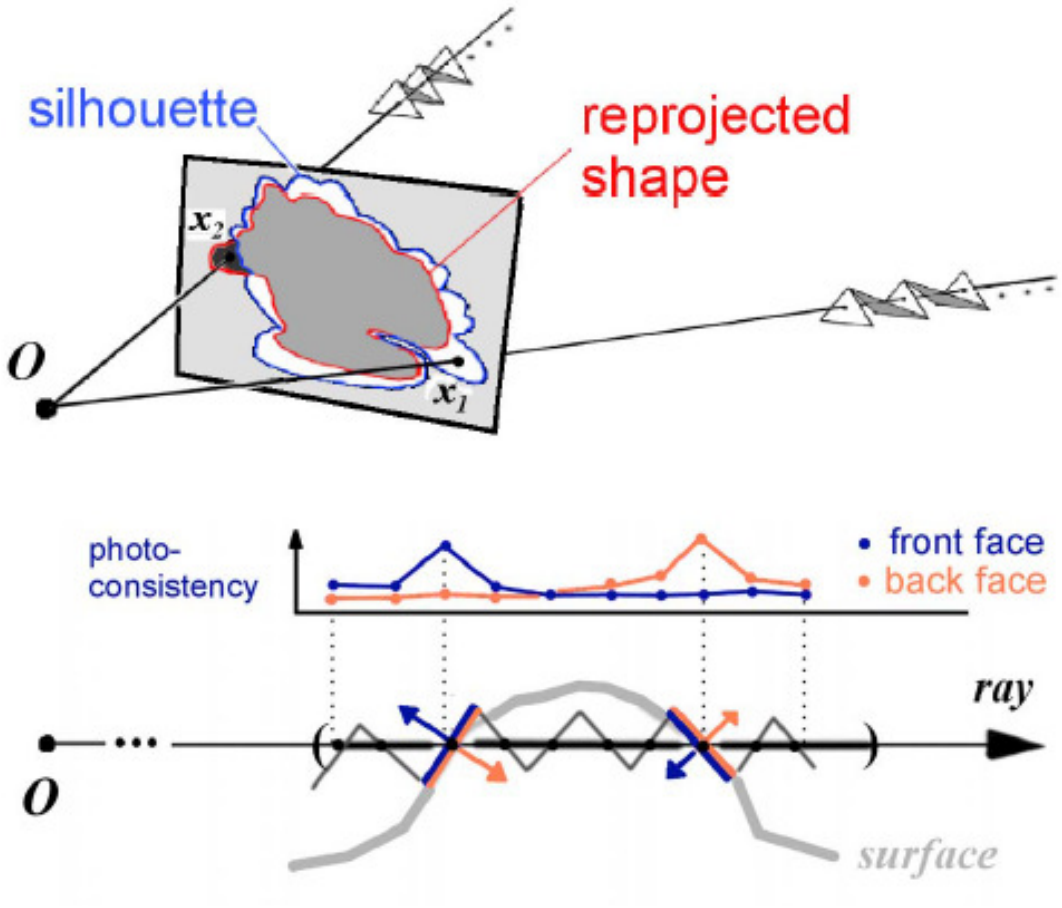
- Volumetric Scene Representation
- Each pixel votes along its viewing ray
- Each voxel accumulates votes from all intersecting rays
- Binary Labeling Problem

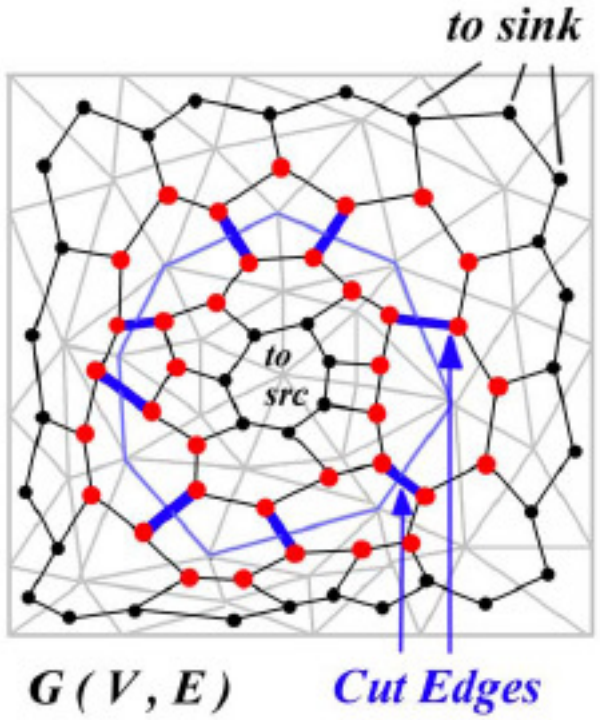
  Graph Cuts

  Variational Methods



Analyzing Adaptive Triangulations

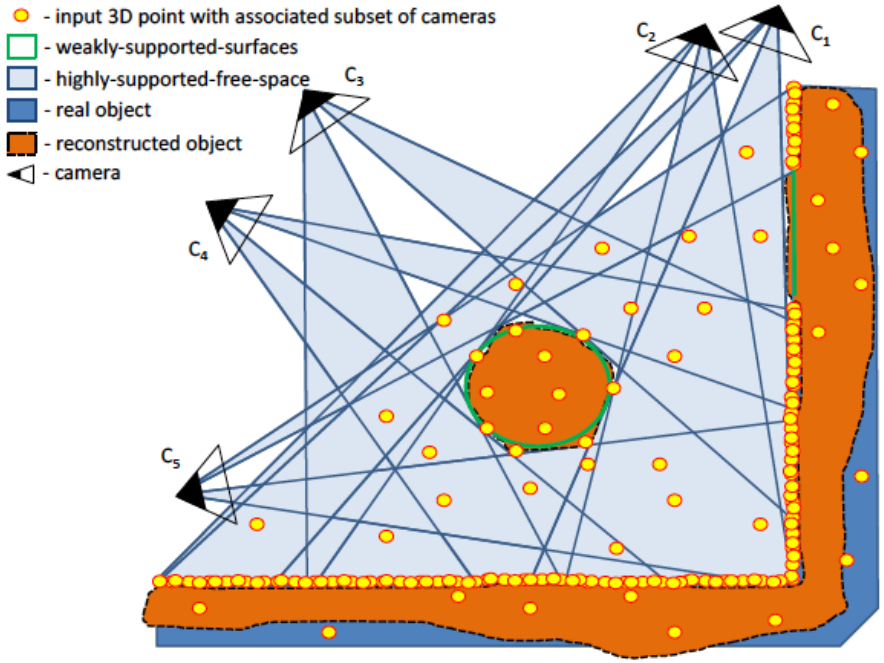


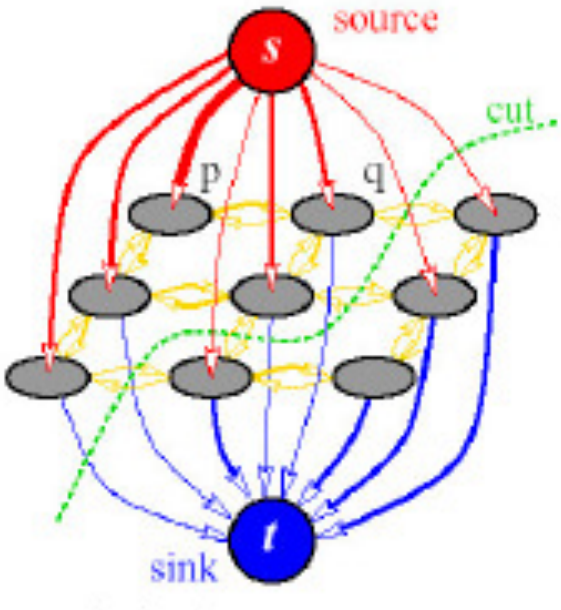


Weakly Supported Surfaces

- Account for proximity to open-space boundaries

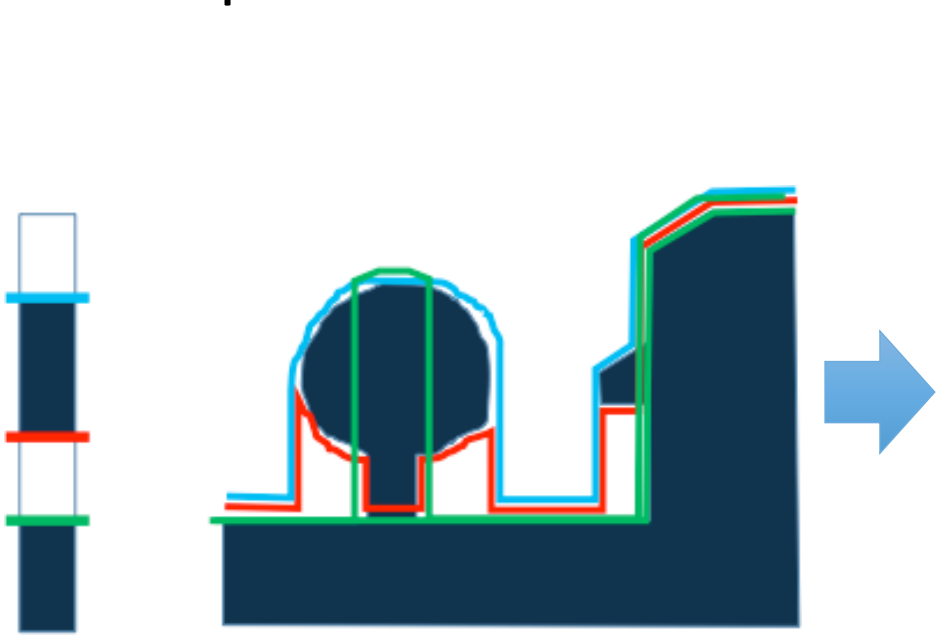
M Jancosek, T Pajdla
Multi-view reconstruction preserving weakly-supported surfaces. CVPR 2011

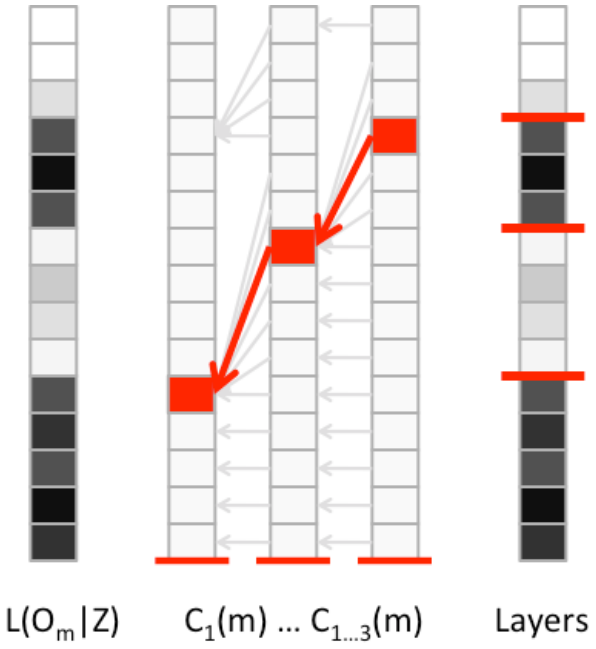




Multi-Layer Heightmaps

- Model transitions in space occupancy along independent 1D domains

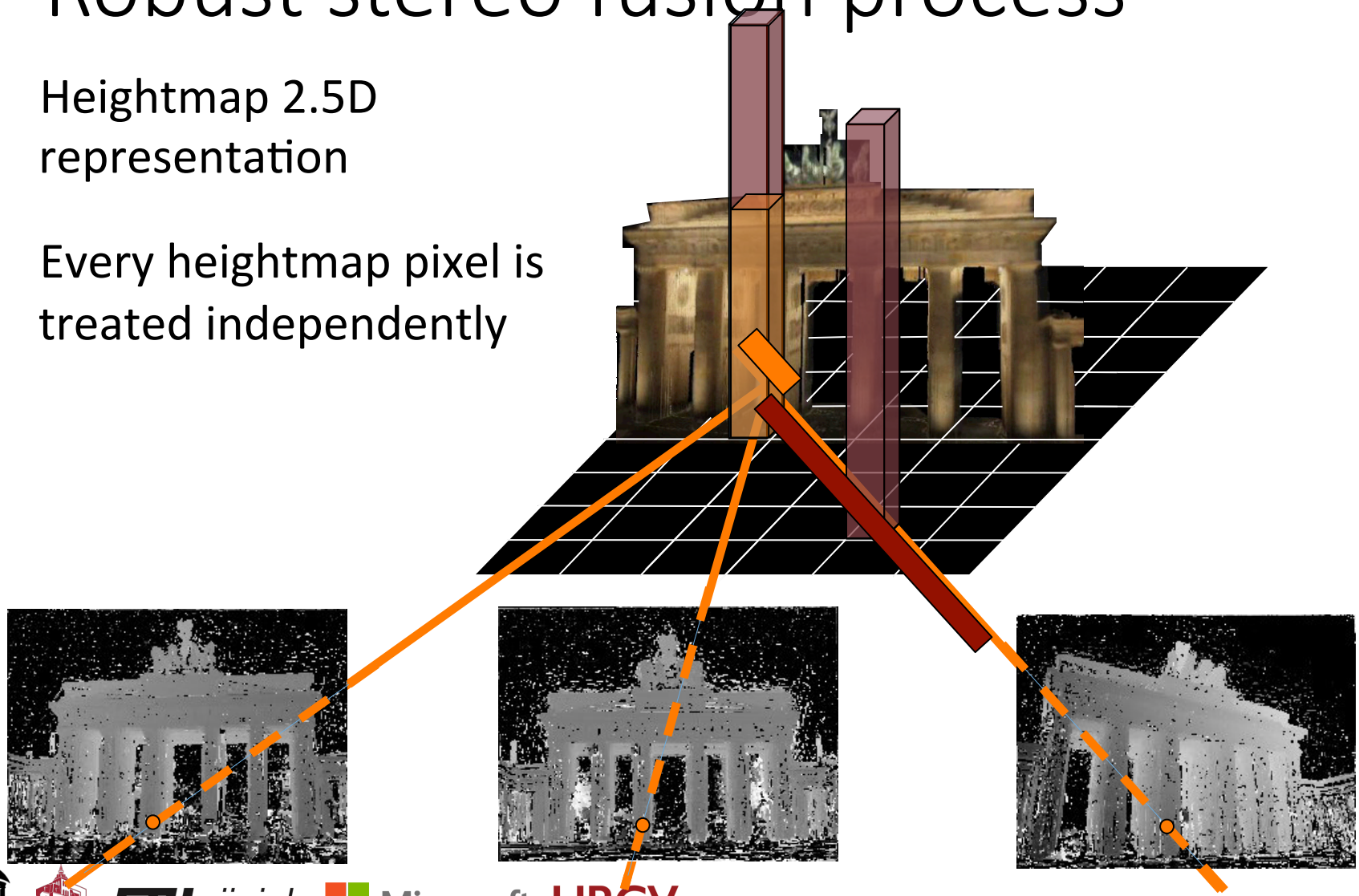




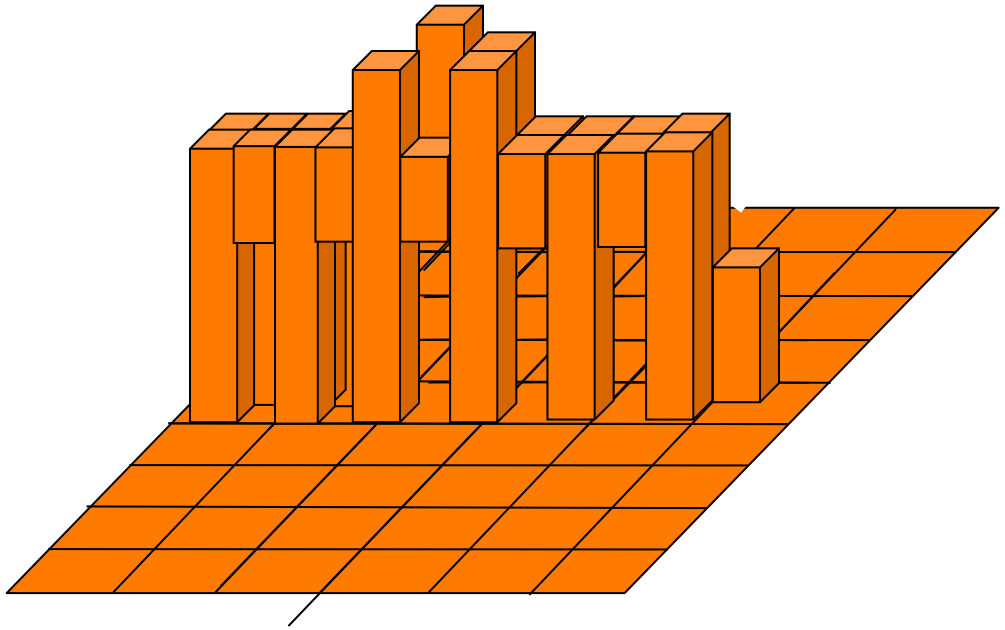
Bayesian Inference
Dynamic Programming

Robust stereo fusion process

- Heightmap 2.5D representation
- Every heightmap pixel is treated independently

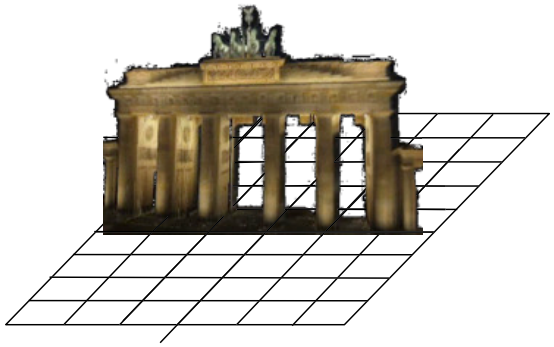


Robust stereo fusion process



- Enforces vertical facades
- One continuous surface, no holes
- Fast to compute, easy to store $O(n^2)$ instead of $O(n^3)$

Layout



- Vertical direction
  - vanishing point
  - camera constraints (Szeliski 2005, Snavely et al. 2006)

Video

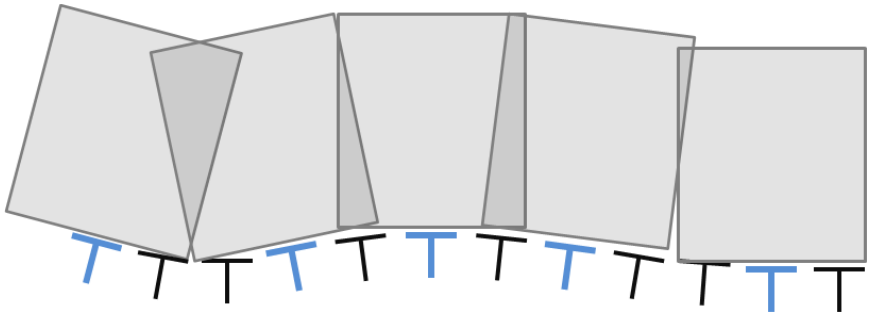
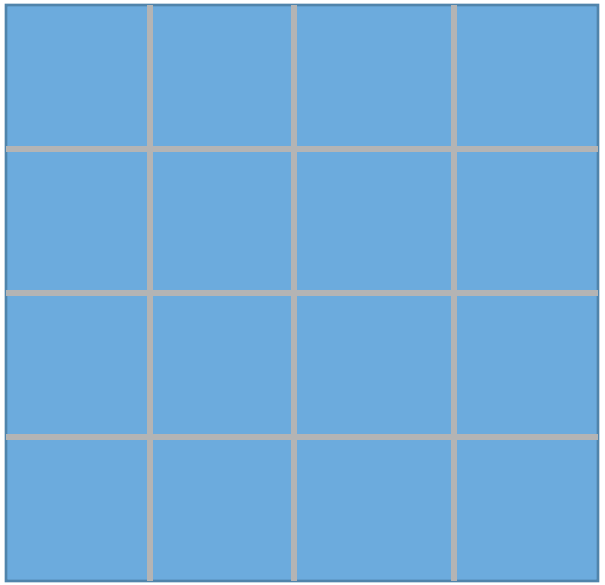


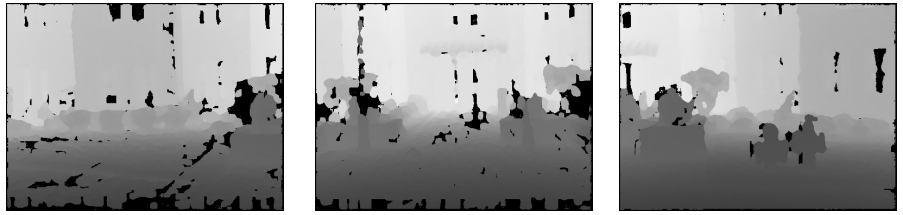
Photo Collections



Vertical Surfaces

Gallup et al. 3DPVT 2010

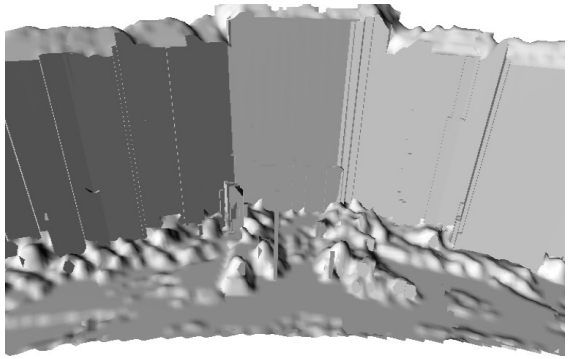
Input Images



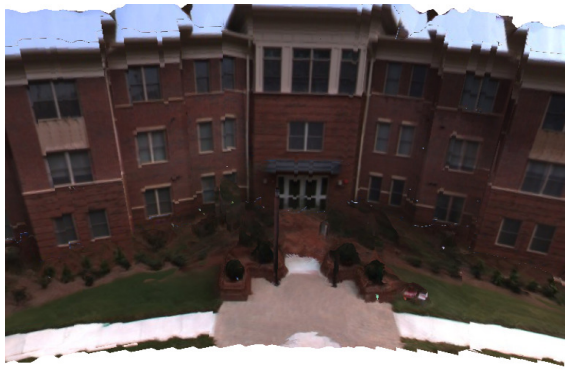
Input Depthmaps

Heightmap
(Top-Down View)
the brighter the higher

3D Model
Geometry

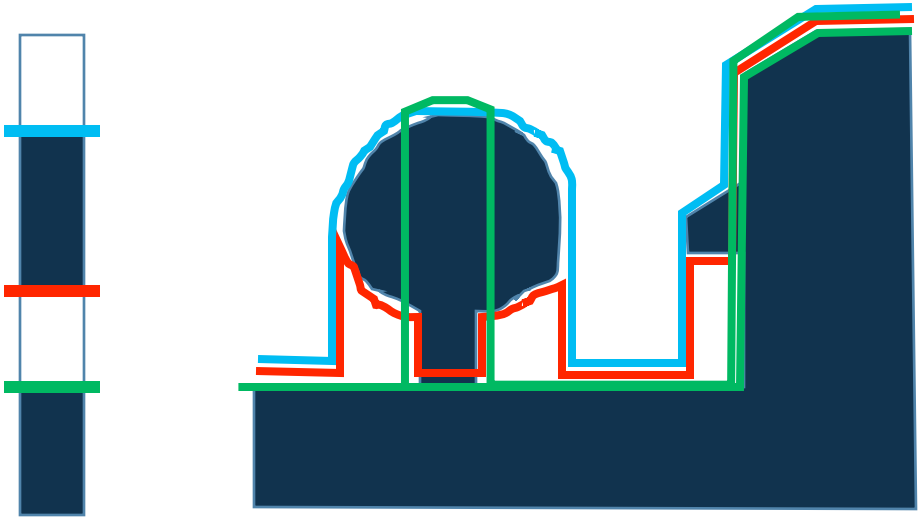


Textured
3D Model



N-Layer Heightmap

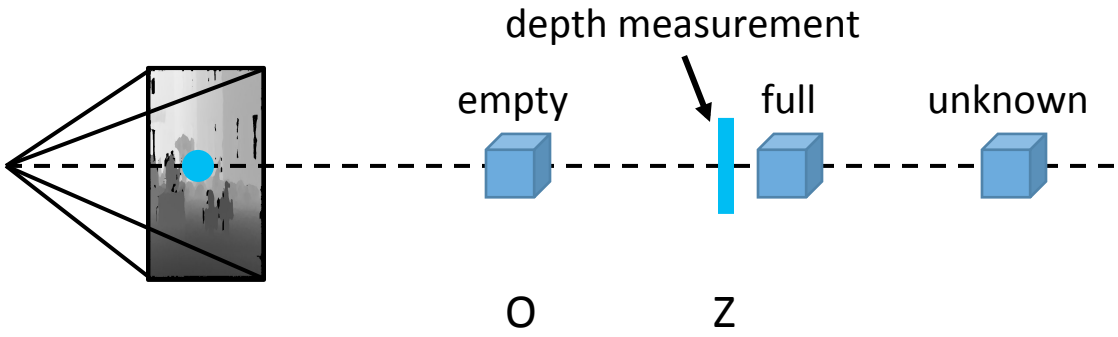
- Generalize to n-layer heightmap
- Each layer is a transition from full/empty or empty/full



- Compute layer positions with dynamic programming
- Use model selection (BIC) to determine number of layers

Probabilistic Occupancy

Introduced in robotics: Margaritis and Thrun 1998 Follows derivation of Guan et al. 2008



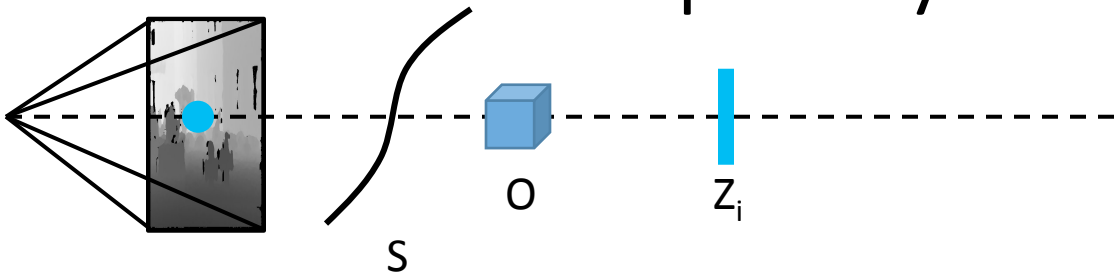
Random Variables

O – binary, occupancy of voxel

$Z=\{Z_i\}$ – continuous, depth measurements for i=1…k depth maps

$$P(O|Z)$$

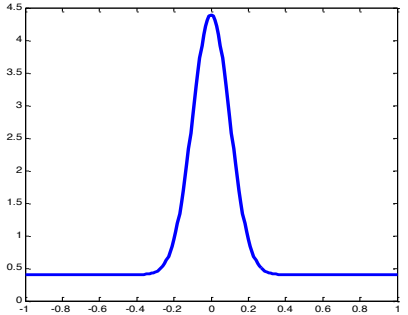
Probabilistic Occupancy



$$P(Z_i|O)$$

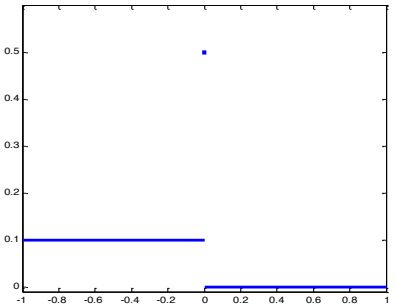
Measurement Model

$$P(Z_i|S,O) = P(Z_i|S) = \begin{cases} \mathcal{N}(S,\sigma)|_{Z_i} & \text{if inliner} \\ \mathcal{U}(z_{min}, z_{max})|_{Z_i} & \text{if outlier} \end{cases}$$
$$= \rho\mathcal{N}(S,\sigma)|_{Z_i} + (1-\rho)\mathcal{U}(z_{min}, z_{max})|_{Z_i}$$

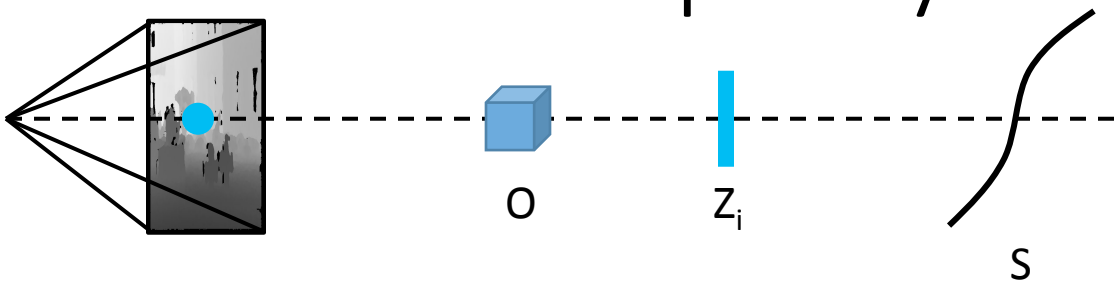


Surface Formation Model

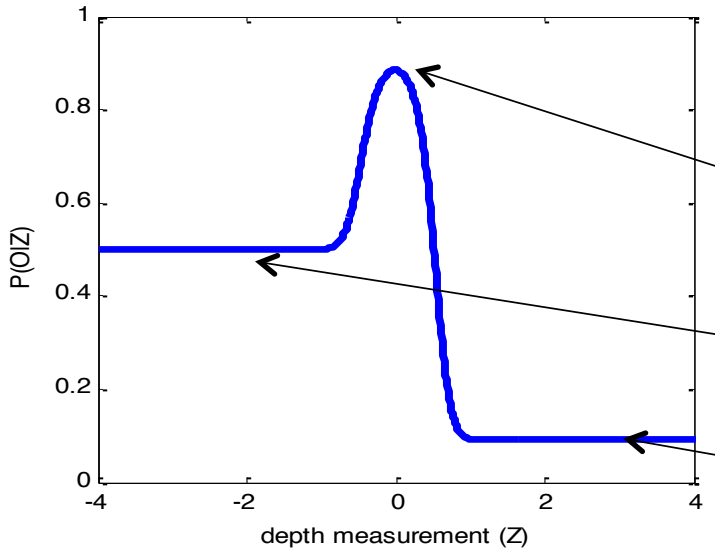
$$P(S|O) = \begin{cases} 1/(z_{max} - z_{min}) & \text{if } S < z_p - \epsilon \\ (1 - z_p/(z_{max} - z_{min})/\epsilon & \text{if } z_p - \epsilon \leq S \leq z_p \\ 0 & \text{if } S > z_p \end{cases}$$



Probabilistic Occupancy



$$P(Z_i|O) = \int_{z_{min}}^{z_{max}} P(Z_i|S,O)P(S|O)dS$$



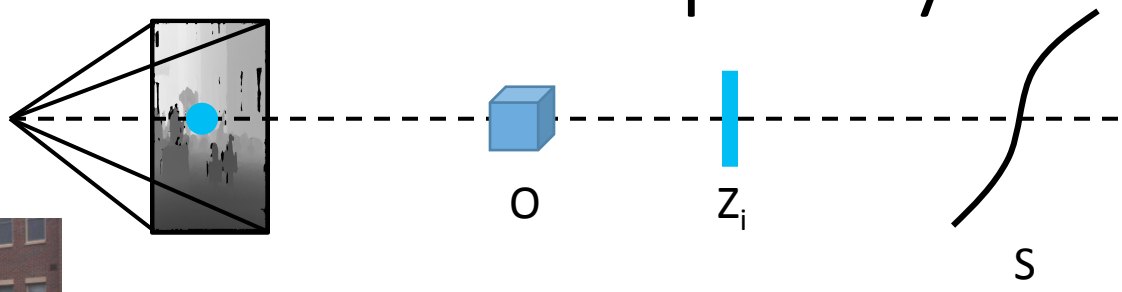
$P(O|Z_i)$ for different values of Z_i

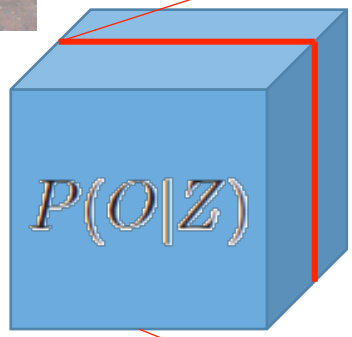
almost certainly occupied at depth

unknown occupancy behind depth

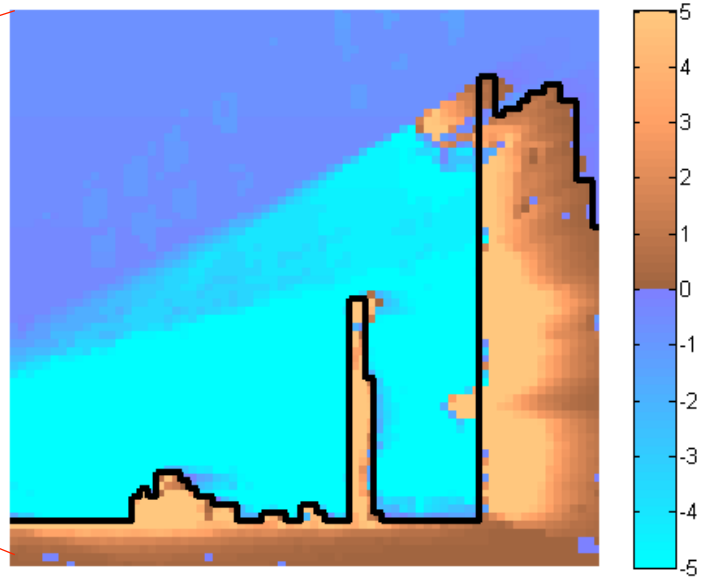
unlikely occupancy in front of depth

Probabilistic Occupancy





probabilistic occupancy volume

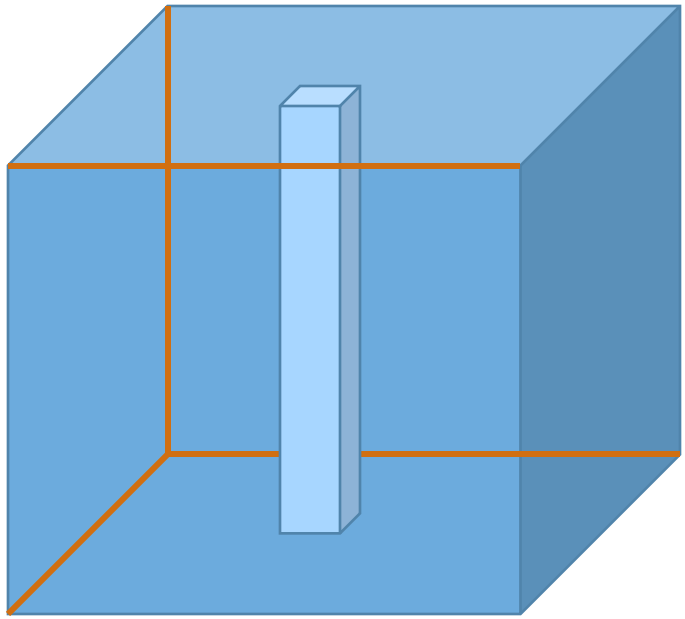


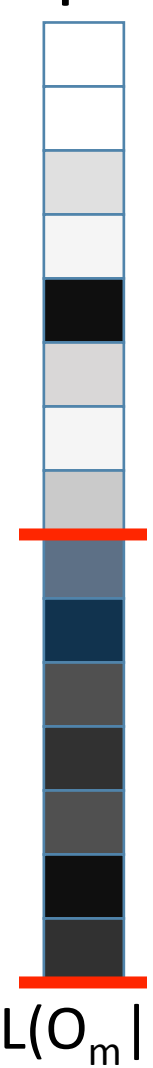
full

empty

occupancy likelihood slice

Heightmap Computation



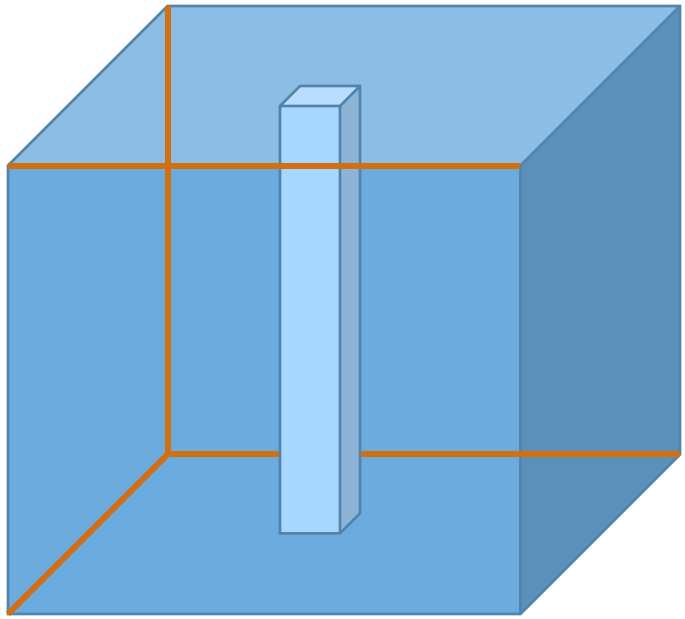


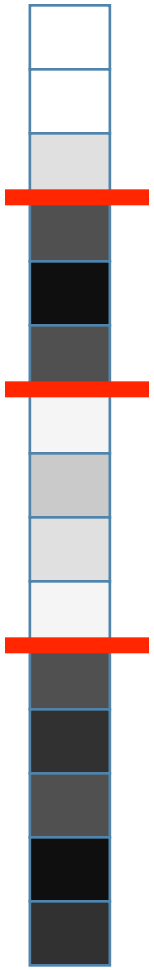
cost of voxels 1…m being full

$$C(m) = \sum_{1}^{m} L(Z \mid O_m) + \sum_{m+1}^{M} L(Z \mid \neg O_m)$$

cost of voxels m+1…M being empty

Heightmap Computation





hold fixed

independent

$C_1(m)$: cost of 1st layer at height m

$C_{1\ldots k}(m) = C_k(m) + \min C_{1\ldots k-1}(m')$:

optimal cost of layers 1…k with layer k fixed at height m

$L(O_m|Z)$

Heightmap Results

Results

References

- Extensive Tutorial on MVS
  - Yasutaka Furukawa and Carlos Hernández (2015), "Multi-View Stereo: A Tutorial", Foundations and Trends® in Computer Graphics and Vision: Vol. 9: No. 1-2, pp 1-148. http://dx.doi.org/10.1561/0600000052
- MVS Benchmarks
  - Middlebury MVS
    Steve Seitz • Brian Curless • James Diebel • Daniel Scharstein • Richard Szeliski
    http://vision.middlebury.edu/mview/
  - The KITTI Vision Benchmark Suite
    Andreas Geiger | Philip Lenz | Christoph Stiller | Raquel Urtasun
    http://www.cvlibs.net/datasets/kitti/

Bibliography

- D Gallup, M Pollefeys, JM Frahm . 3d reconstruction using an n-layer heightmap. Joint Pattern Recognition Symposium, 2010
- Y Furukawa, J Ponce, Accurate, dense, and robust multiview stereopsis. PAMI 2010.
- SN Sinha, M Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. ICCV 2005
- Multi-view reconstruction preserving weakly-supported surfaces M Jancosek, T Pajdla Computer Vision and Pattern Recognition (CVPR), 2011
- Johannes L. Schönberger, Enliang Zheng, Marc Pollefeys, Jan-Michael Frahm, " Pixelwise View Selection for Unstructured Multi-View Stereo", ECCV 2016
- Jared Heinly, Johannes L. Schönberger, Enrique Dunn, Jan-Michael Frahm, "Reconstructing the World* in Six Days *(As Captured by the Yahoo 100 Million Image Dataset)", CVPR 2015
- Johannes L. Schönberger, Jan-Michael Frahm, "Structure-from-Motion Revisited", CVPR 2016

Thanks

Questions?