多视图立体匹配与三维重建

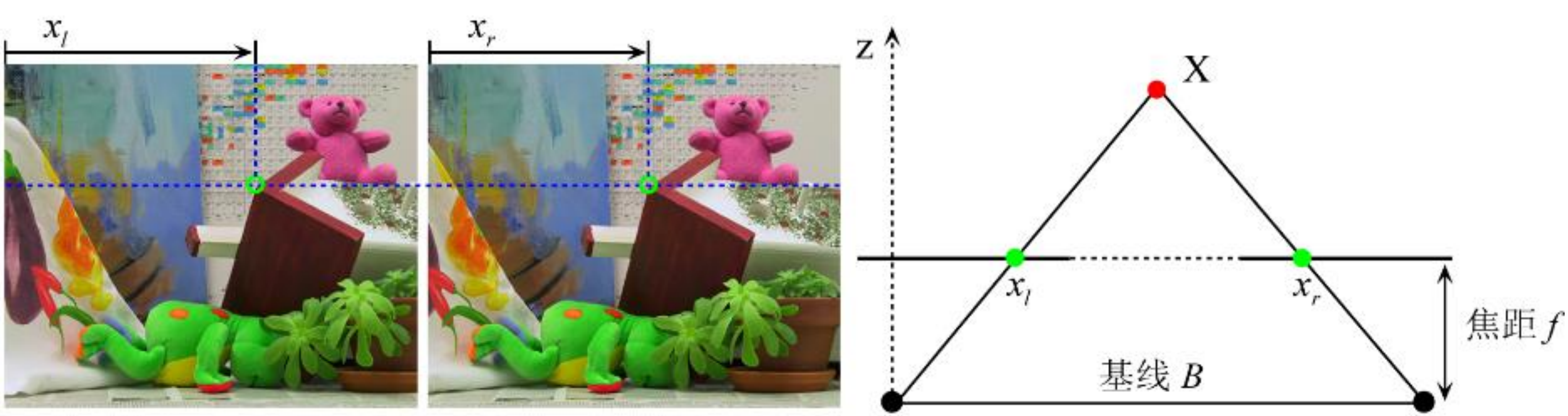
章国锋

浙江大学CAD&CG国家重点实验室

深度恢复技术

Overview

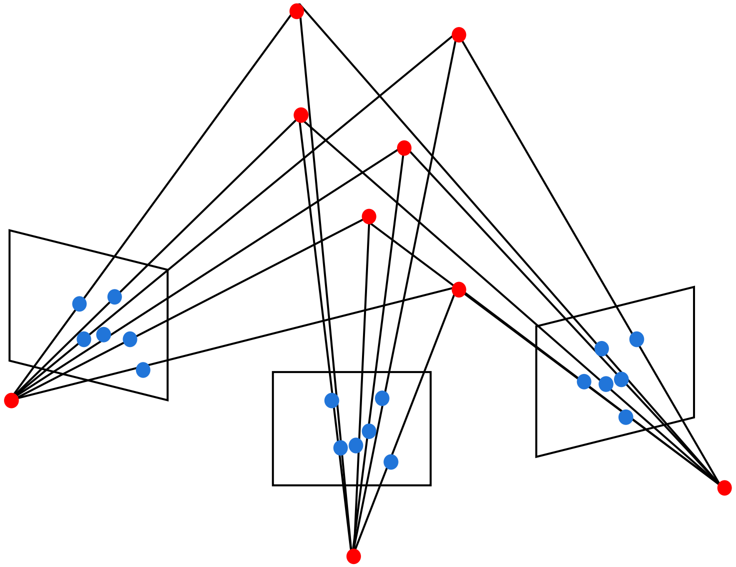
- 双视图立体匹配



视差 $d = x_l - x_r$　　深度 $z = Bf / d$

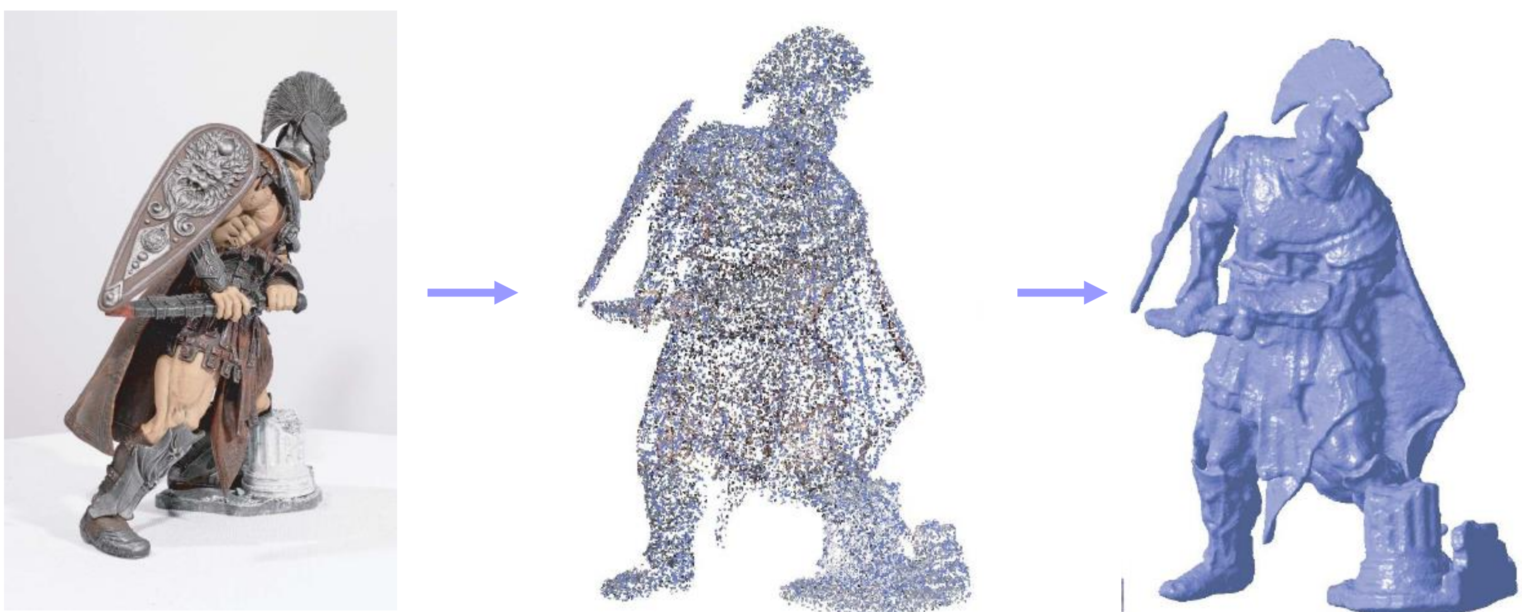
Overview

- 双视图立体匹配
- 多视图立体匹配



Overview

- 双视图立体匹配
- 多视图立体匹配
- 三维几何重建

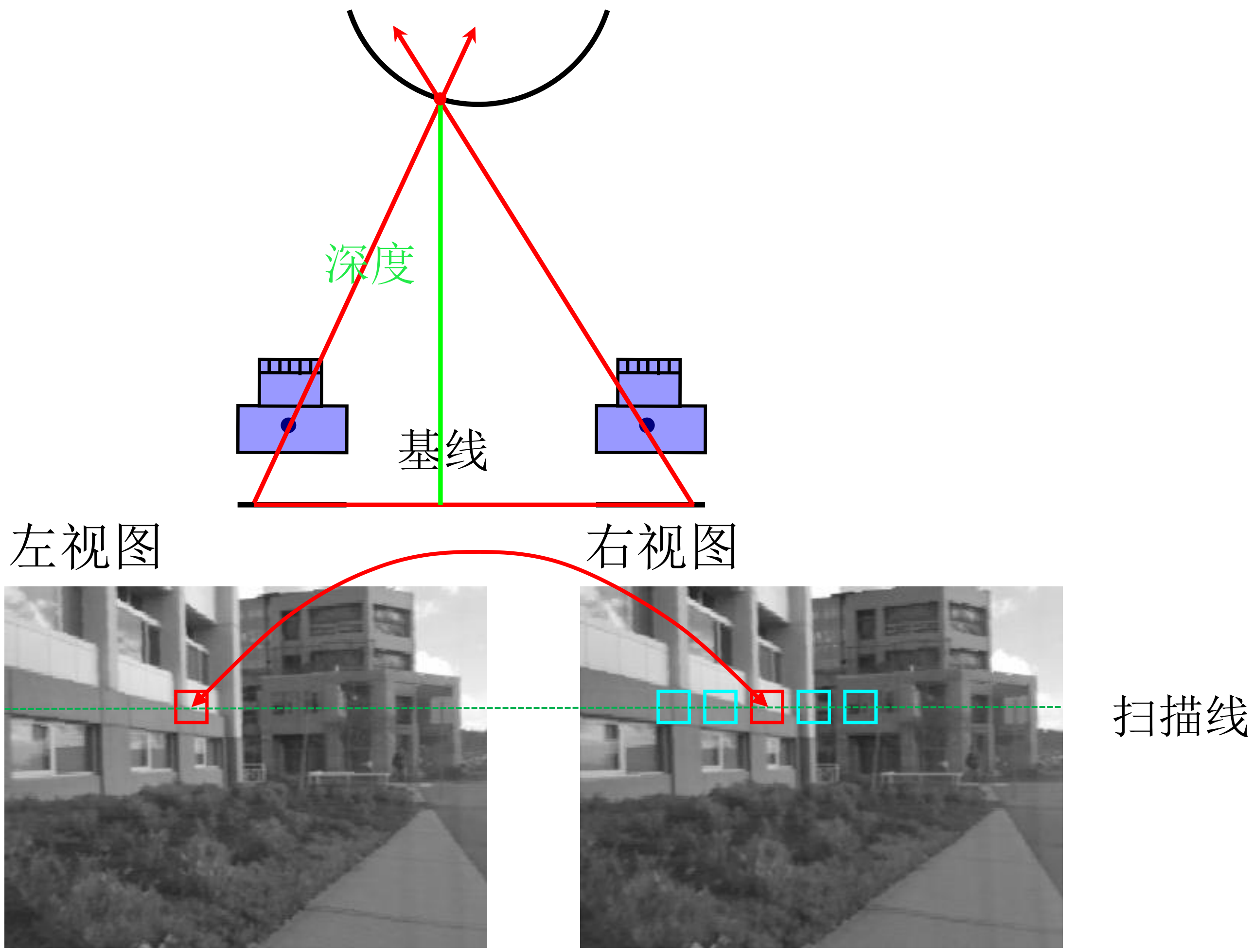


立体视觉

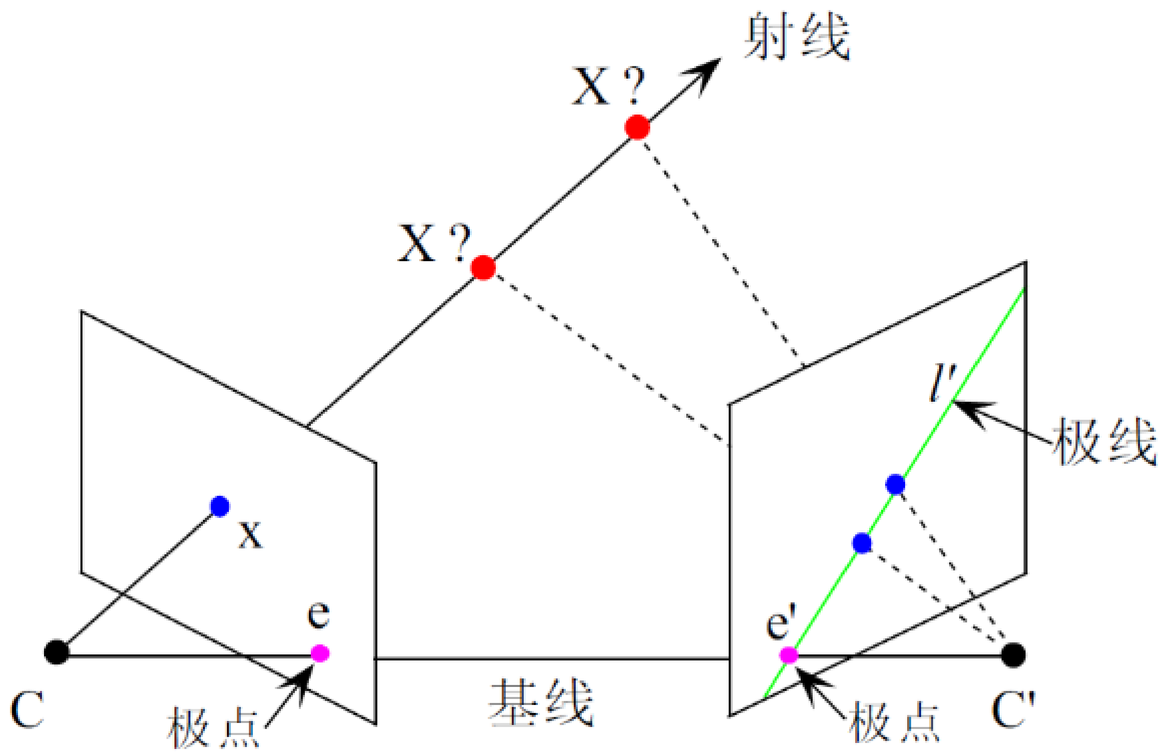
- 立体匹配
  - 从两幅或多副图像中恢复出稠密的深度信息

- 通常的流程
  - 运动推断结构：恢复出摄像机参数
  - 逐像素匹配
  - 计算深度

立体视觉



极线几何约束

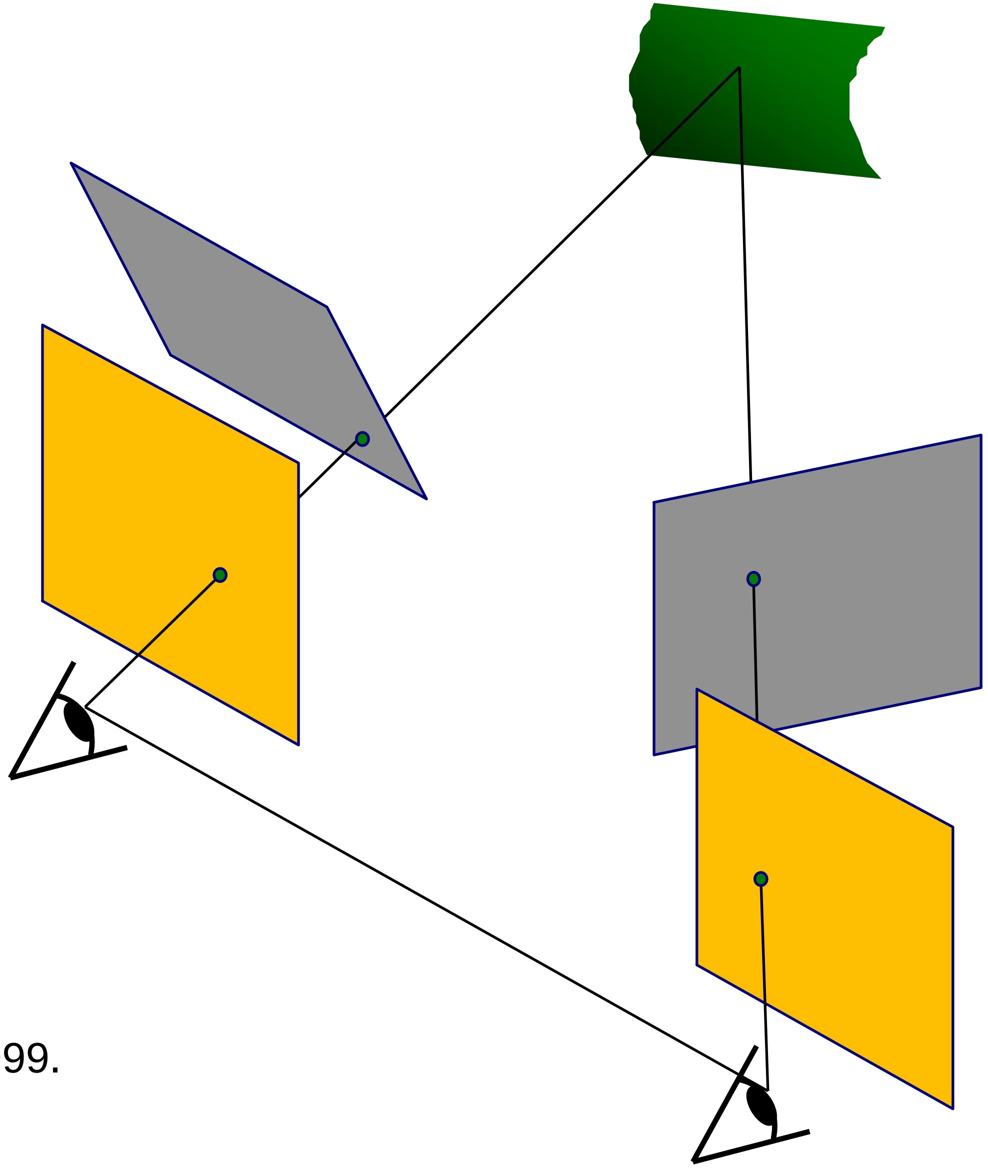


- 只需要在极线上进行匹配
- 二维搜索变成一维搜索
- 极大地减小匹配代价

图像矫正

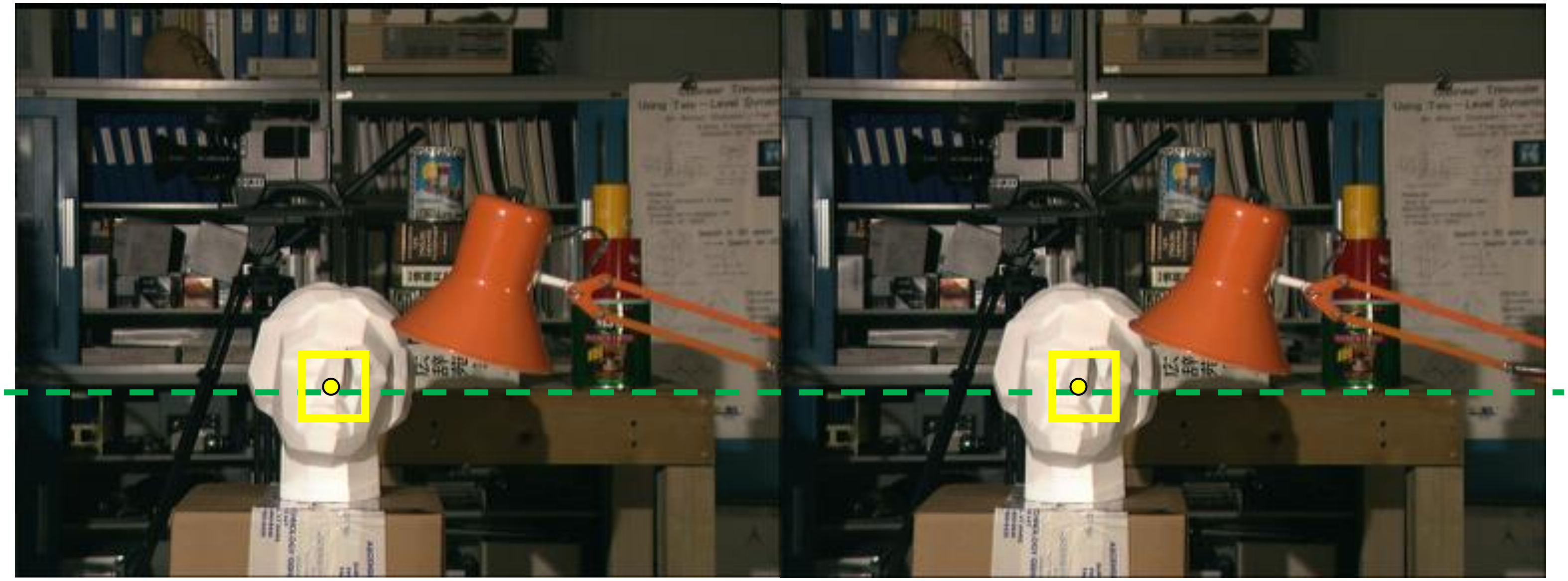
- 标准的配置
  - 左右相机的方向一致且它们的中心连线垂直
- 矫正方法
  - 将左右视图投影到一个公共平面上

C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.



(Seitz)

像素匹配



每一条扫描线：

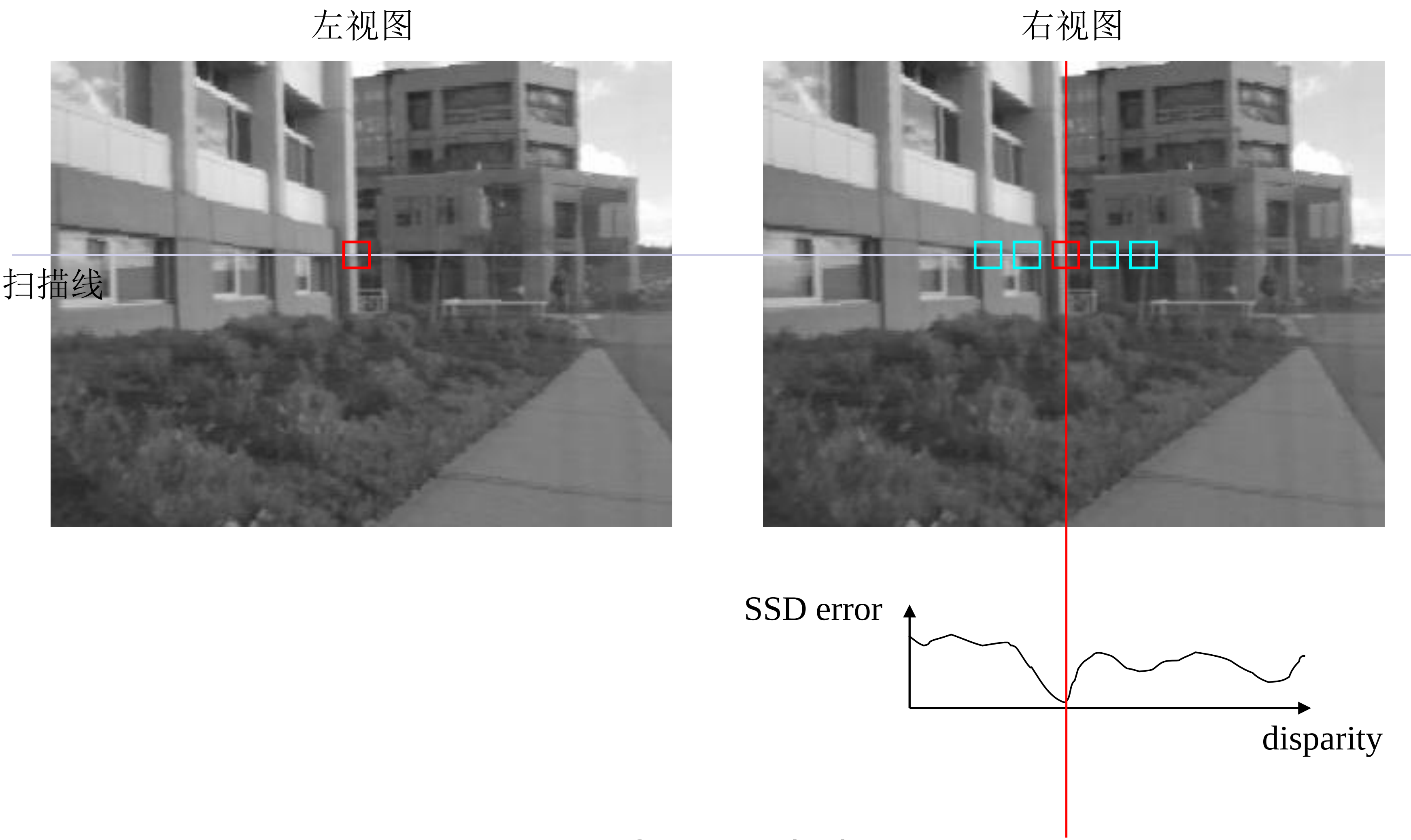
左视图上的每一个像素：

- 跟右视图上的同扫描线上的每一个像素进行比较
- 选出颜色最相似的像素作为匹配点

单像素匹配很容易受噪声影响

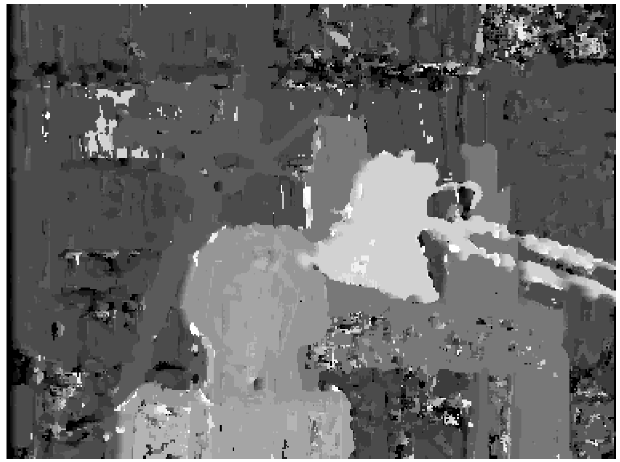
改进：基于窗口的匹配

基于窗口的匹配

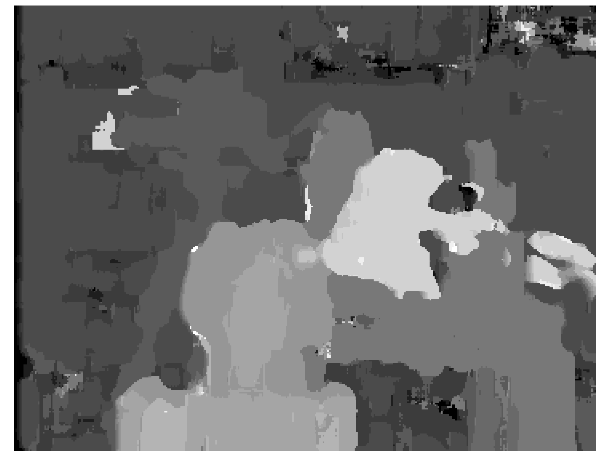


SSD: Sum of Squared Distance

窗口大小的选择



W= 5*5



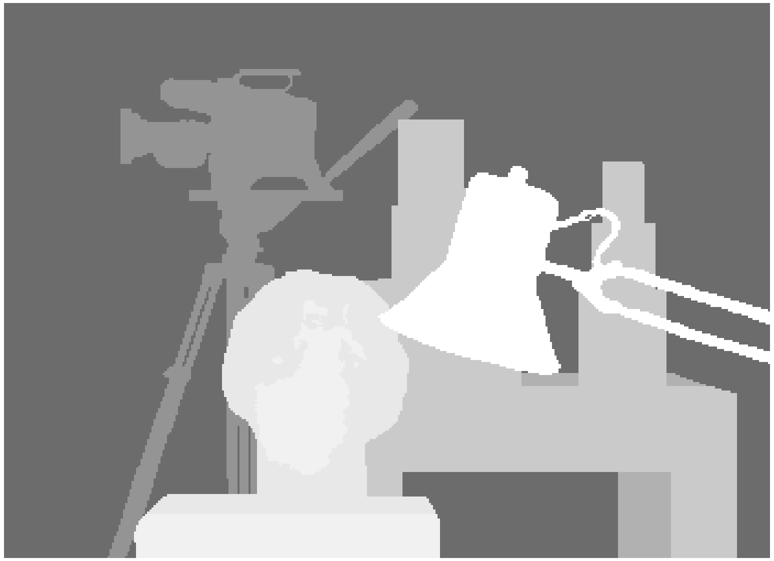
W= 11*11

- 不同窗口大小的影响
  - 匹配窗口很小时，起不到很好的平滑作用
  - 窗口取得比较大的时候，一些细小结构和不连续边界附近的深度会变得不准确
- 自适应的窗口匹配方法

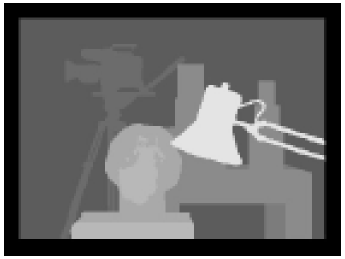
立体匹配方法的评测

D. Scharstein and R. Szeliski. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision,* **47** (2002), pp. 7-42.

Scene



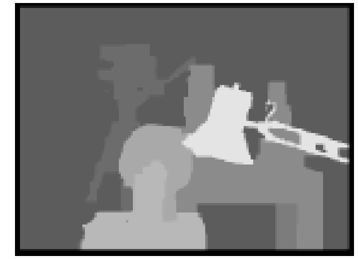
Ground truth



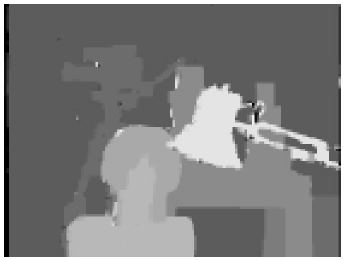
True disparities

19 – Belief propagation

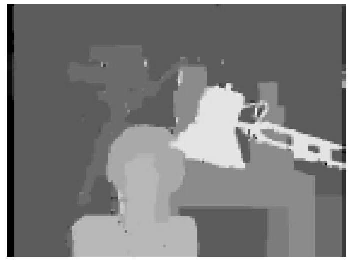
11 – GC + occlusions



20 – Layered stereo

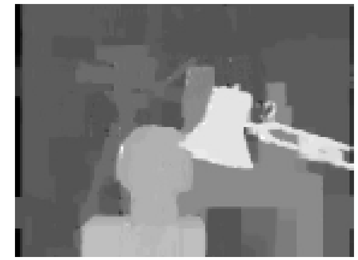


10 – Graph cuts



*4 – Graph cuts

13 – Genetic algorithm



6 – Max flow

12 – Compact windows

9 – Cooperative alg.

15 – Stochastic diffusion

*2 – Dynamic progr.

14 – Realtime SAD

*3 – Scanline opt.

7 – Pixel-to-pixel stereo

*1 – SSD+MF

Scharstein and Szeliski

全局优化方法

- Energy Function

$$E(d) = E_d(d) + \lambda E_s(d)$$

  - Data Term

$$E_d(d) = \sum_{x \in I} \phi(x, d)$$

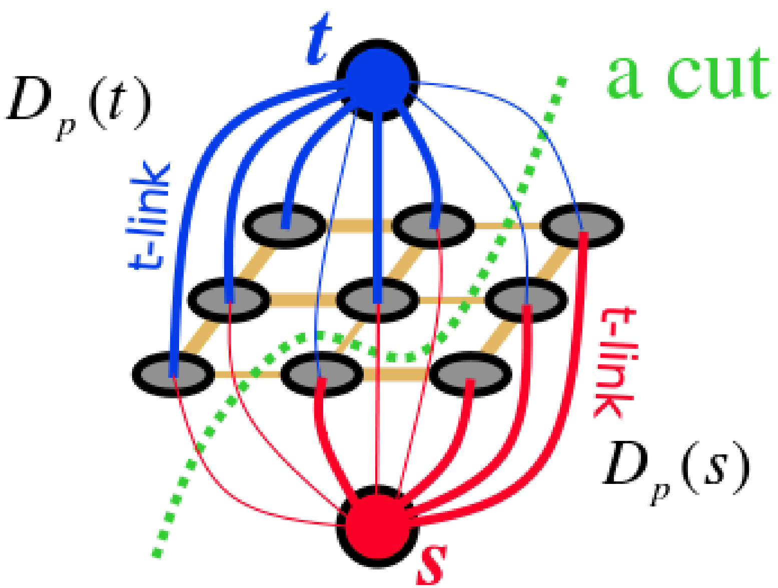
  - Smoothness Term

$$E_s(d) = \sum_{x \in I} \sum_{y \in N(x)} \psi(d_x, d_y)$$

- Optimization
  - Graph Cuts
  - Belief Propagation

Graph Cuts

- 定义
  - 对于一个图 $G = (V, E)$，其中 V 为节点集合，包括源点s和终点t、以及其他诸多中间节点集合V'，E 为连接这些节点的边，每条边附有容量$c(u, v)$ 代表节点u通过这条边流向节点v所能承受的最大流量。
  - Graph cuts的目的在于找到图的Min-cut，Cut将 V' 分割为两个部分，去掉这些边将使舍得图中的任意一个节点只与s或t相连通，而Min-cut是所有cut中边的能量值总和最小的一个。



$$E(f)=\sum_{p\in P} D_p(f_p)+\sum_{p,q\in N} V_{p,q}(f_p,f_q)$$

Graph Cuts

- Recommended Paper
  - Yuri Boykov, Olga Veksler, Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23(11): 1222-1239, 2001.
- Graph Cuts Home Page
  - http://www.cs.cornell.edu/~rdz/graphcuts.html
- Source code:

http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/software.html

Multi-Label Graph-Cuts

- α-β Swap
  - Semi-metric

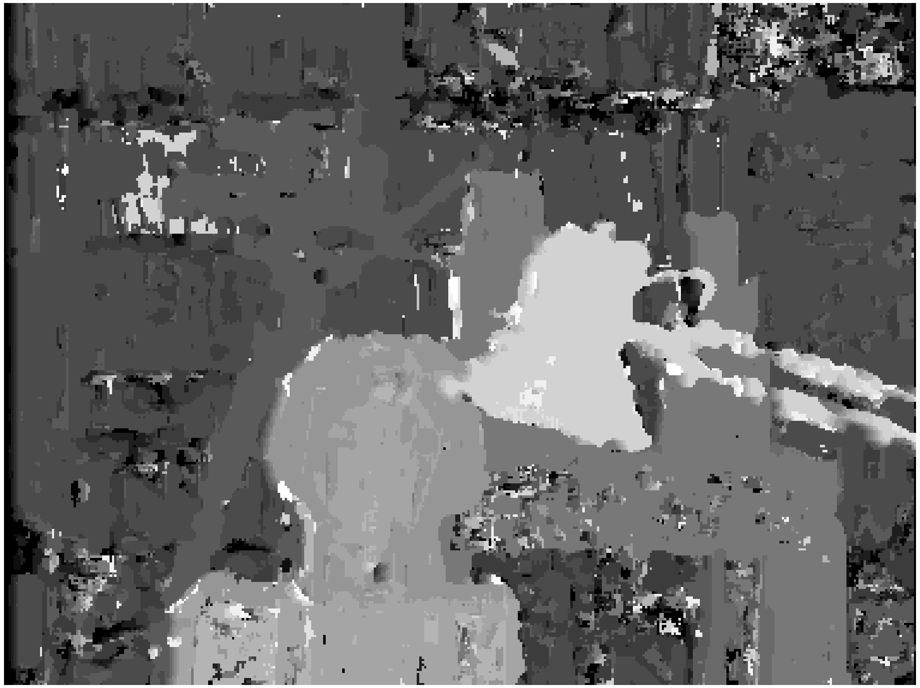
$$V(\alpha,\beta) = V(\beta,\alpha) \geq 0 \ \text{ and } \ V(\alpha,\beta) = 0 \ \Leftrightarrow \ \alpha = \beta.$$

- α-expan
  - Metric

If V also satisfies the triangle inequality

$$V(\alpha,\beta) \leq V(\alpha,\gamma) + V(\gamma,\beta)$$

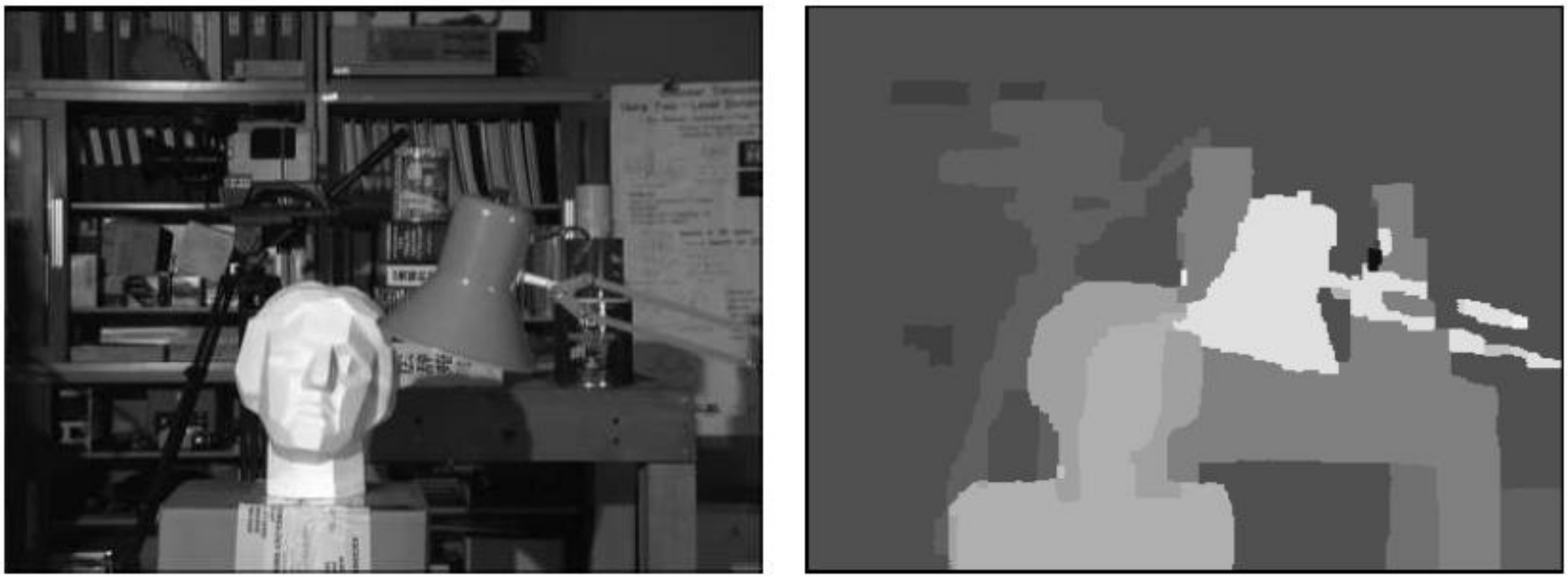
Comparison



The Result with Window-based Matching

The Result with Graph Cuts

Stereo Matching with Belief Propagation



Stereo results for the Tsukuba image pair

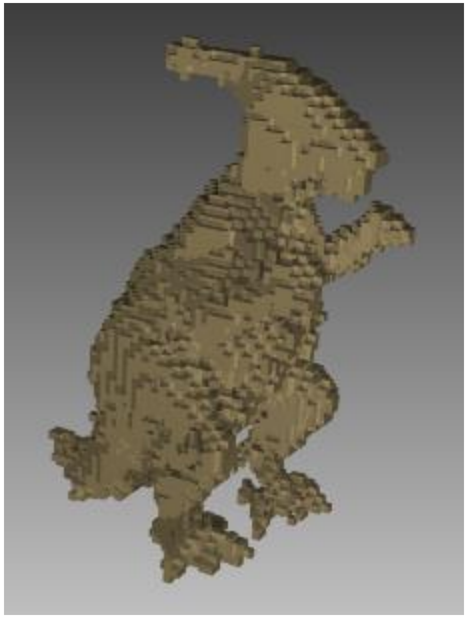
Belief Propagation

- Recommended Paper:
  - Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient Belief Propagation for Early Vision. International Journal of Computer Vision, Vol. 70, No. 1, October 2006.
- Source Code:
  - http://people.cs.uchicago.edu/~pff/bp/

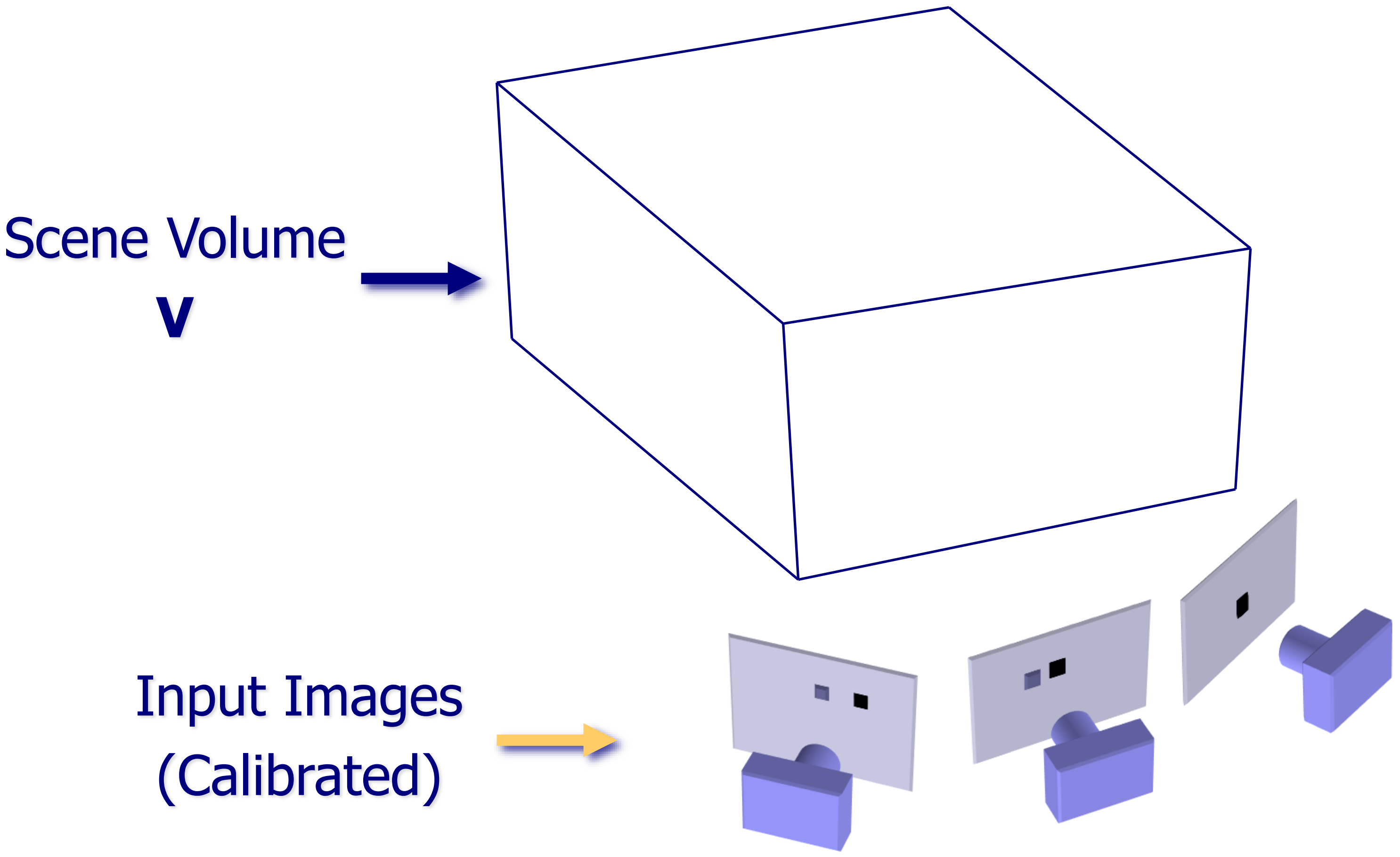
Multi-View Stereo

A Brief Review

- Voxel-based Approaches
  - Voxel Coloring [Seitz & Dyer 97], Space carving [Kutulakos & Seitz 98], Faugeras & Keriven 98, Paris et al. 04, Pons et al. 05, Tran & Davis 06, Vogiatzis et al. 05, …

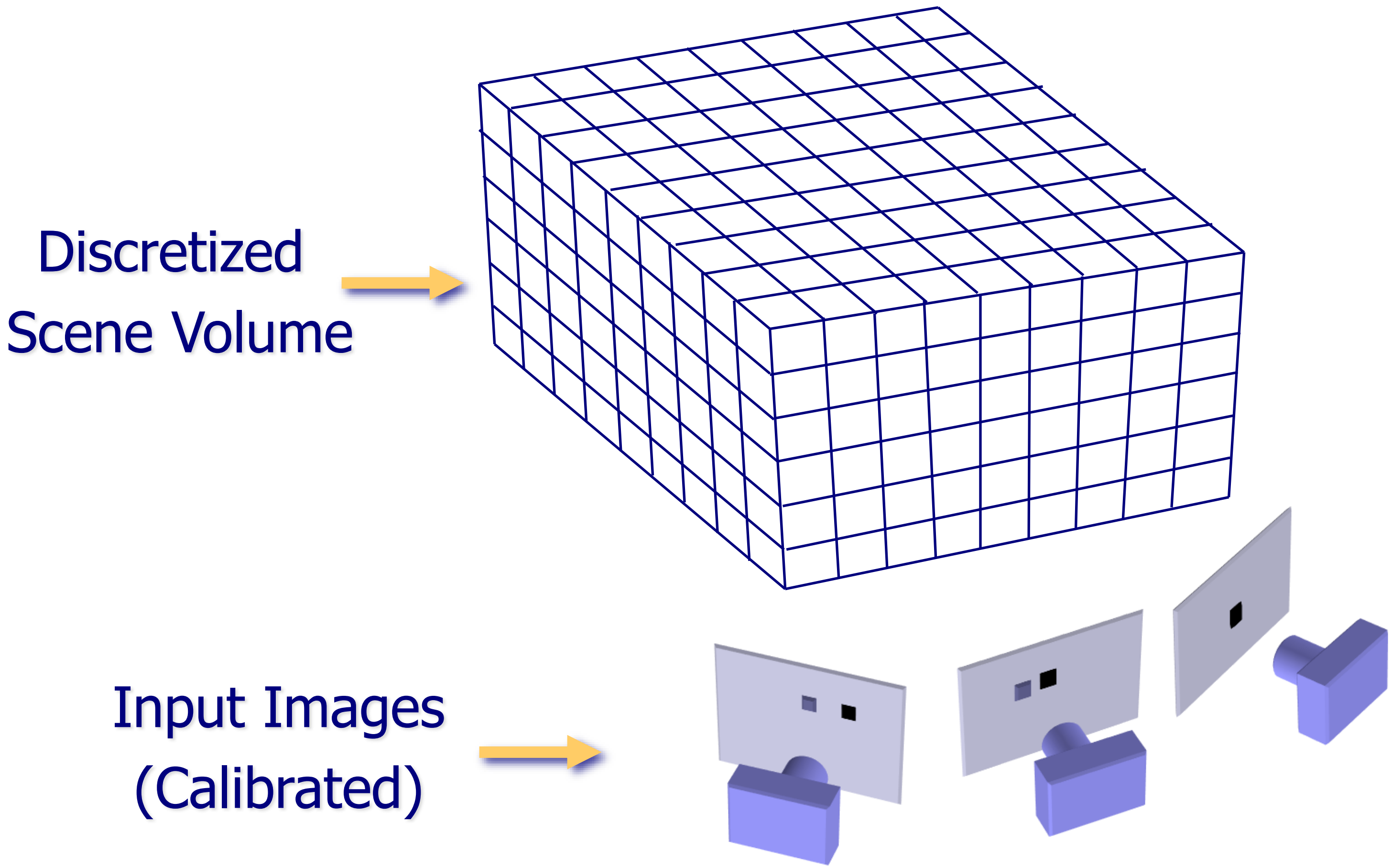


Volumetric Stereo



(Alexei Efros)

Discrete Formulation

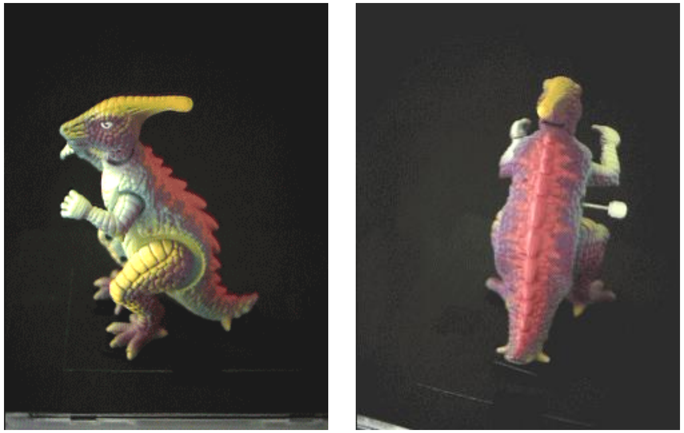


(Alexei Efros)

Calibrated Image Acquisition

- *Calibrated Turntable*

360° rotation (21 images)



Selected Dinosaur Images

Selected Flower Images

(Alexei Efros)

Voxel Coloring Results

Seitz and Dyer 97

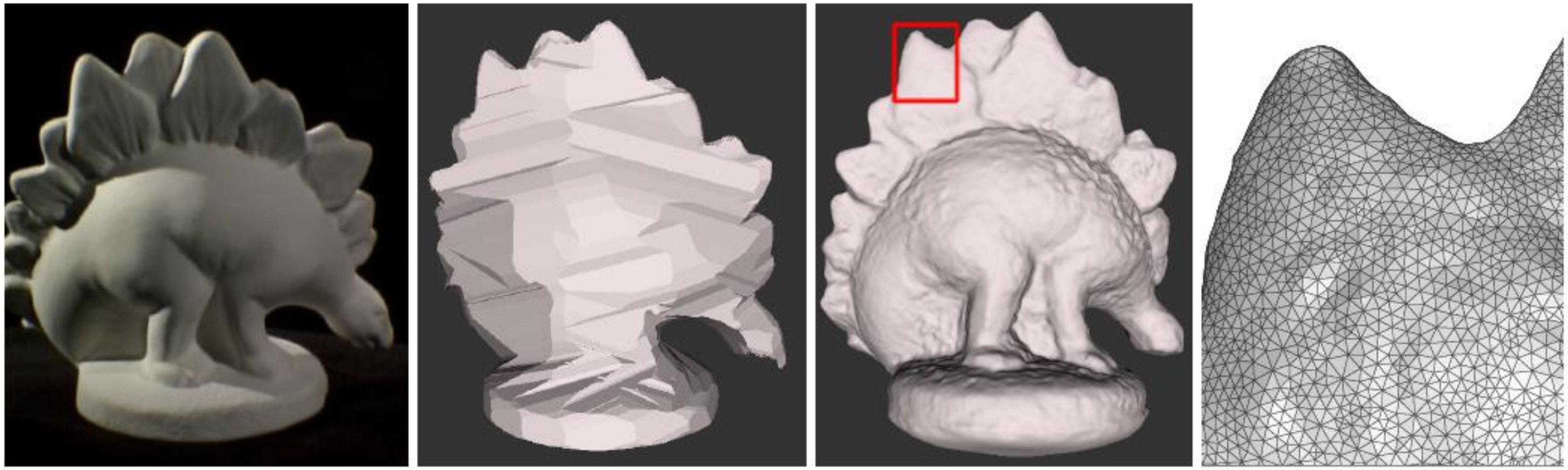
Dinosaur Reconstruction
72 K voxels colored
7.6 M voxels tested
7 min. to compute
on a 250MHz SGI

Flower Reconstruction
70 K voxels colored
7.6 M voxels tested
7 min. to compute
on a 250MHz SGI

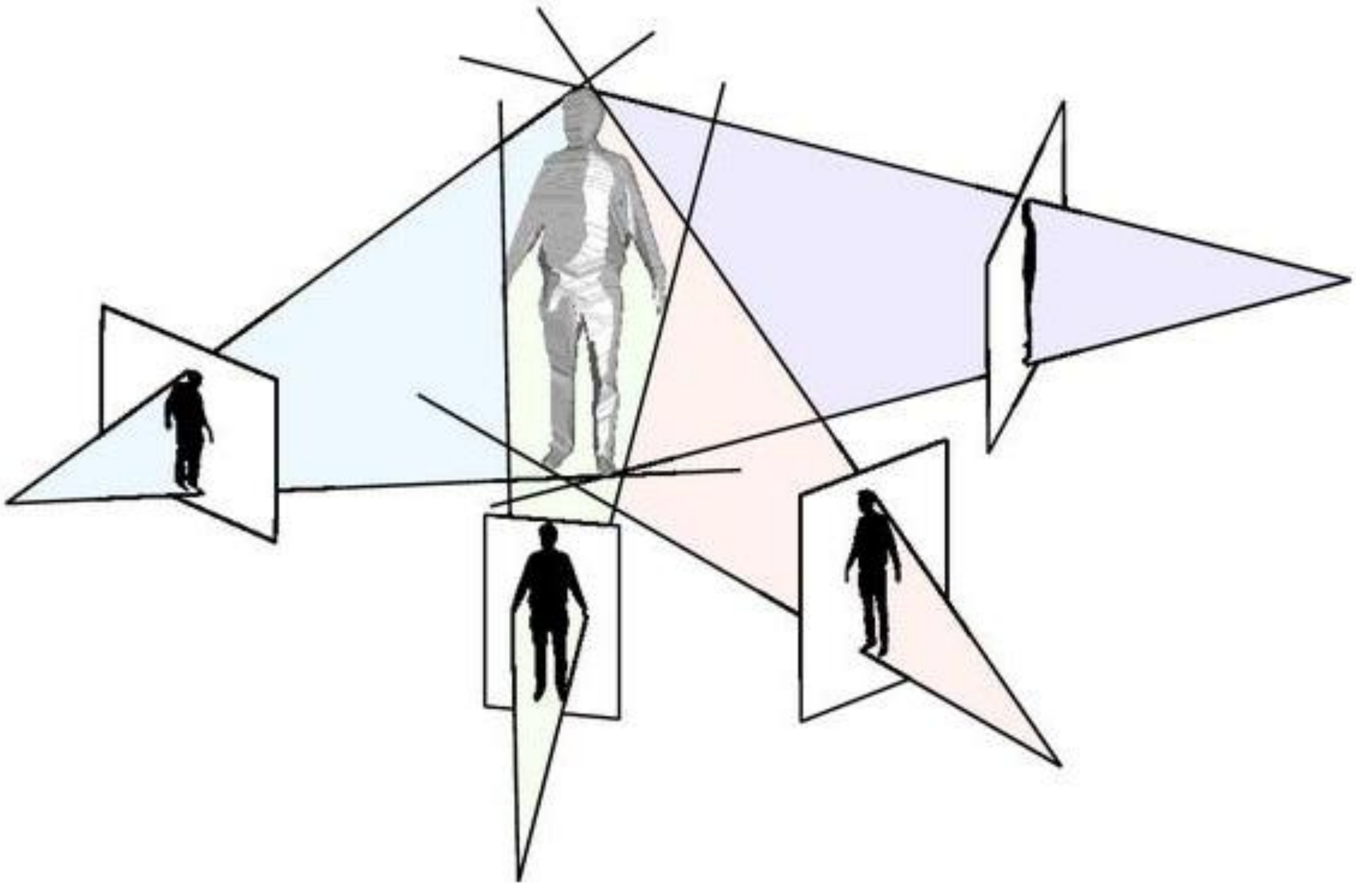
(Alexei Efros)

A Brief Review

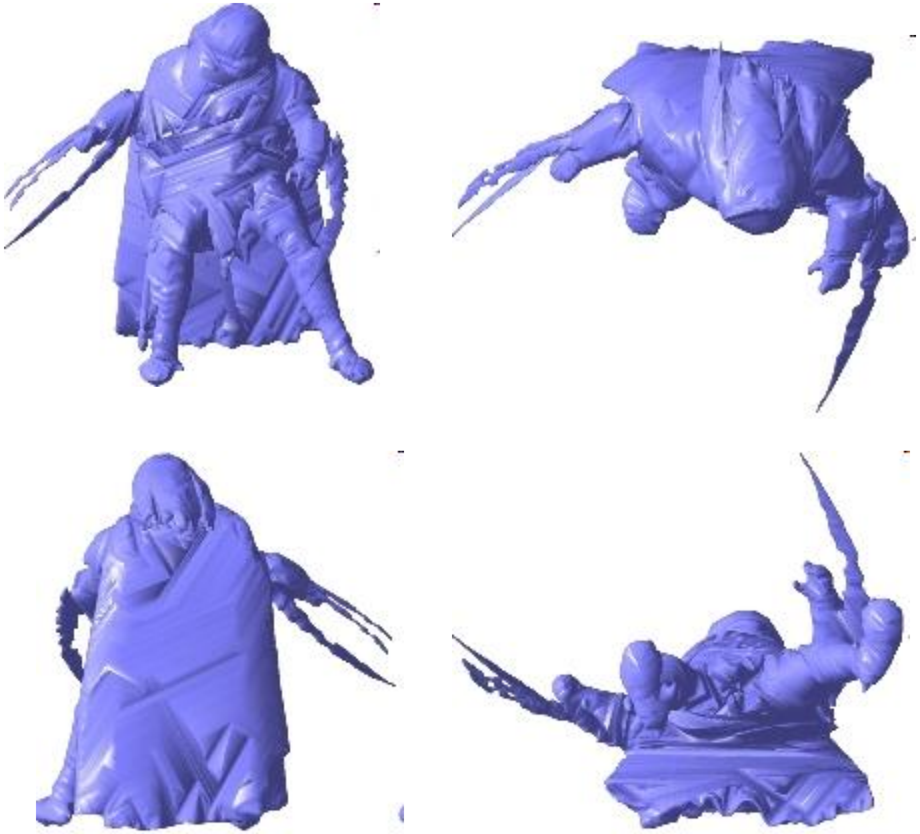
- Approaches based on Deformable Polygonal Meshes
  - Esteban & Schmitt 04, Zaharescu et al. 07, Furukawa & Ponce 08, …



Visual Hull



A Result of Visual Hull



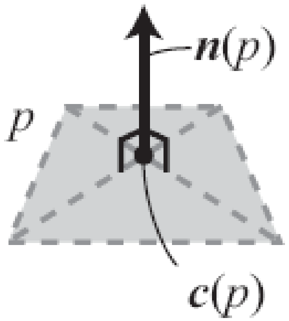
http://www.cs.washington.edu/homes/furukawa/research/visual_hull/index.html

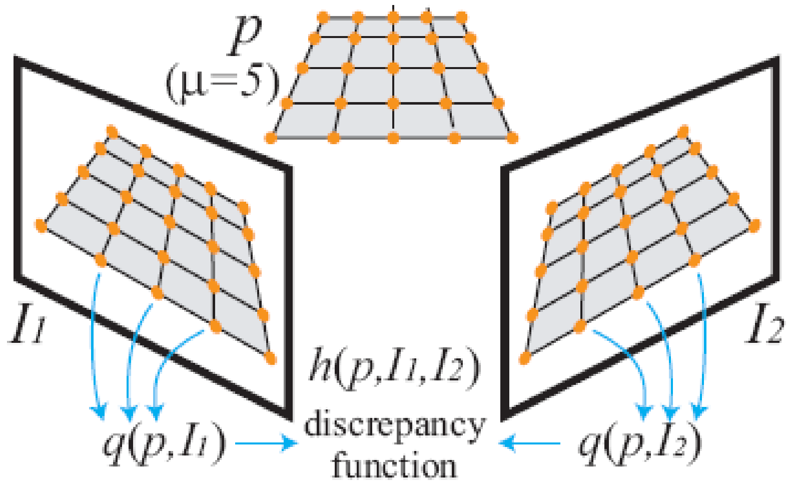
A Brief Review

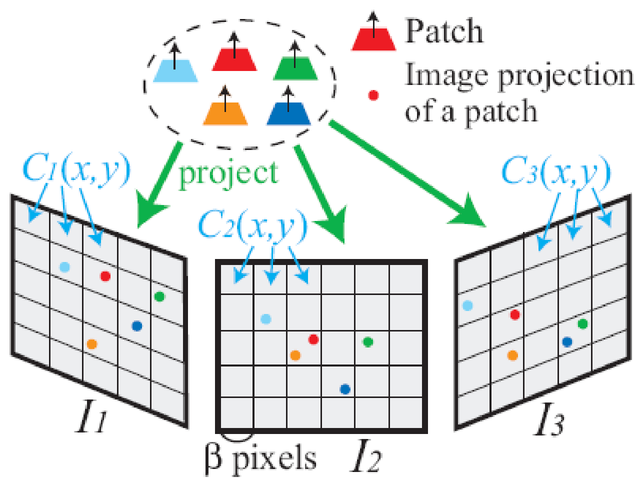
- Patch-based Method
  - Furukawa & Ponce 07, 10

Yasutaka Furukawa, Jean Ponce: Accurate, Dense, and Robust Multiview Stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. 32(8): 1362-1376 (2010)

Patch-Based Multi-View Stereo







Furukawa & Ponce 2010

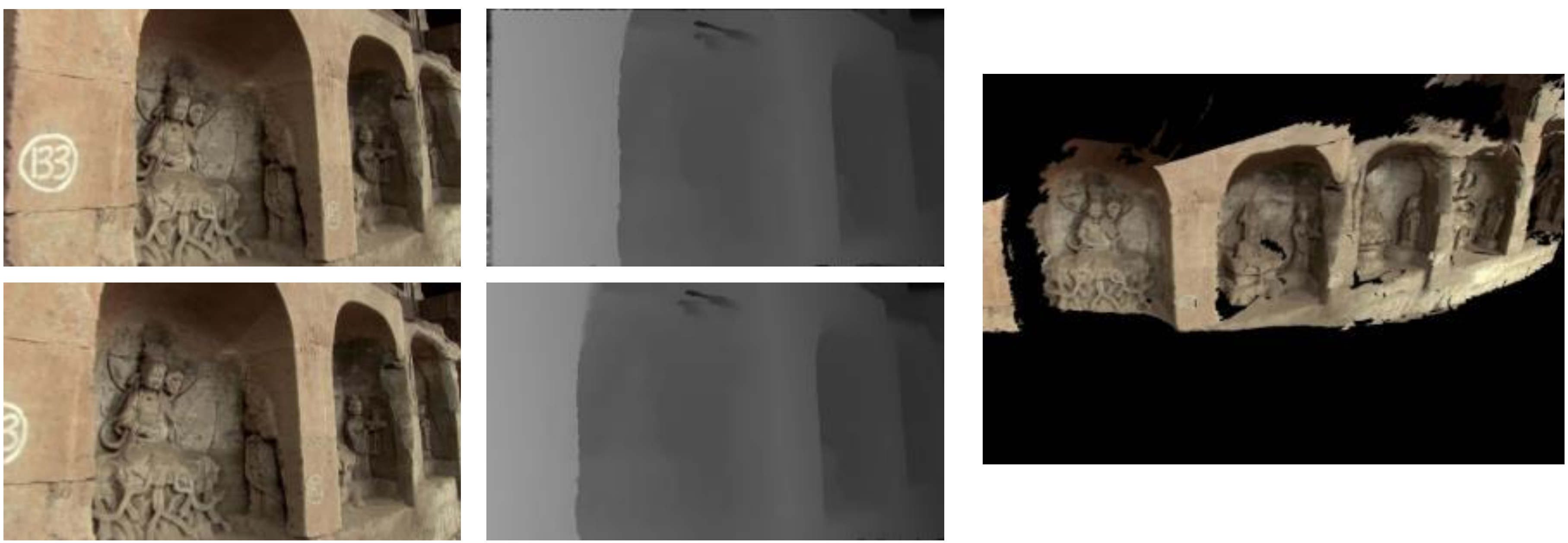
Patch-based Multi-view Stereo Software
(PMVS - Version 2)

- http://grail.cs.washington.edu/software/pmvs/

Software developped and distributed by Yasutaka Furukawa - University of Washington
Jean Ponce - Ecole Normale Supérieure

A Brief Review

- Approaches based on Multiple Depth Maps
  - Goesele et al. 06, Strecha et al. 06, Bradley et al. 08, Zhang et al. 09, ...



Consistent Depth Recovery

- Key Issues
  - Image noise
  - Textureless regions
  - Occlusions

Typical Solutions

- Cost Aggregation
- Smoothness Constraint
- Incorporating Segmentation
- Visibility Labeling

Image noise

Textureless regions

Occlusions

Typical Solutions

- Window-based Aggregation
  - Make estimated depths less accurate
  - Introduce artifacts around boundaries

- Smoothness Constraint
  - Make optimization complex
  - Require a good starting point

Typical Solutions

- Segmentation-based Approaches
  - Less accurate in textured region
  - Introduce segmentation errors

- Binary Visibility Labeling
  - Hand tune the threshold
  - Difficult to distinguish between noise & occlusions
  - Make optimization difficult

Our Key Idea

Image noise

Occlusions

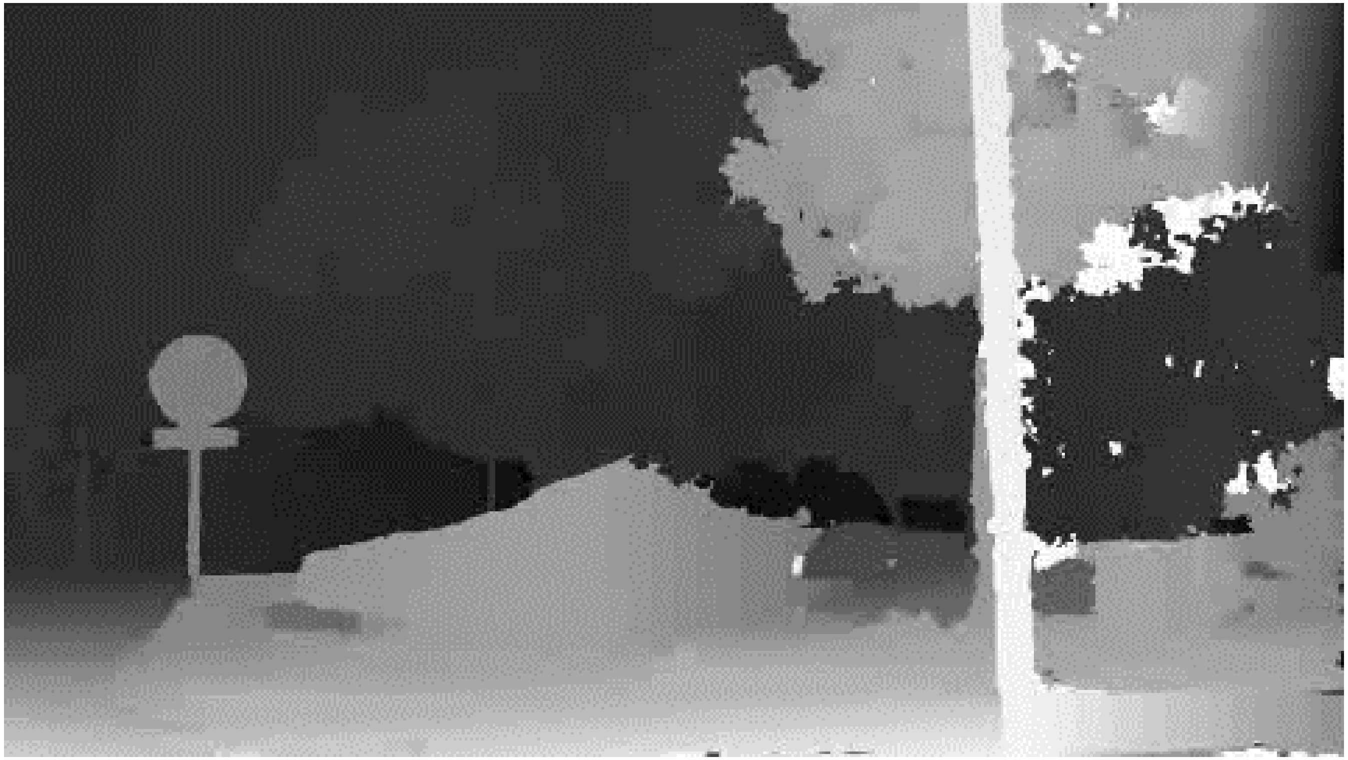
Estimation Error

Our Key Idea

Image noise

Occlusions

Estimation Error



Our Key Idea

Image noise

Occlusions

Estimation Error

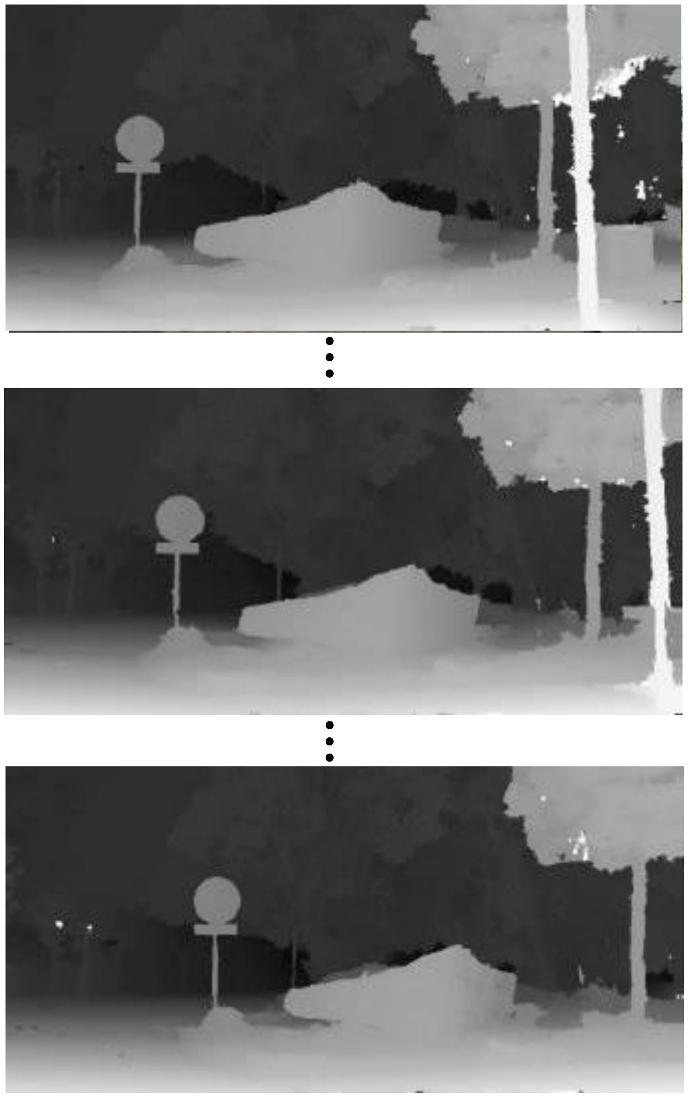
$\underbrace{\hspace{10em}}$

All can be regarded as **temporal noise** !

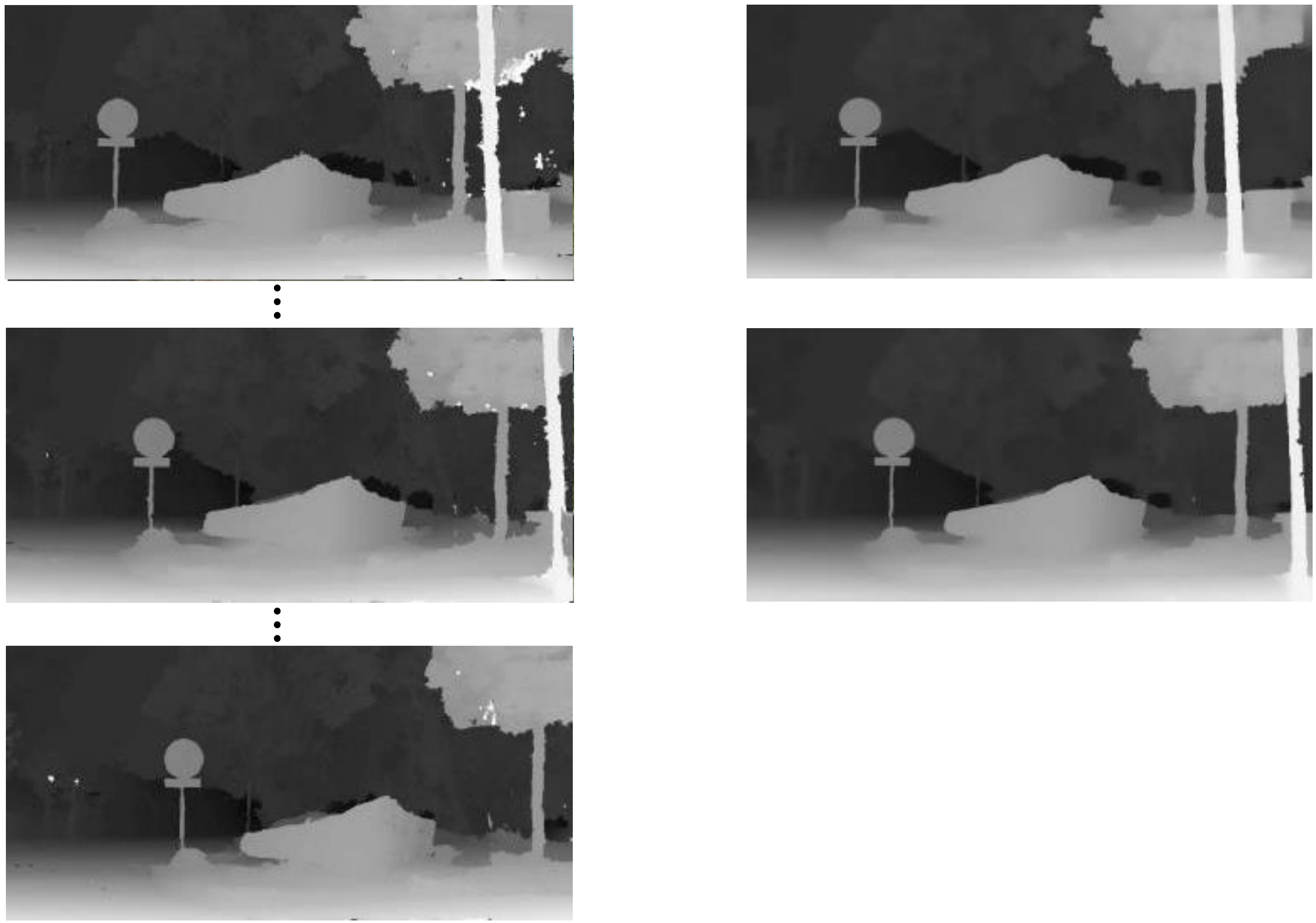
A **unified** framework for handling them

?

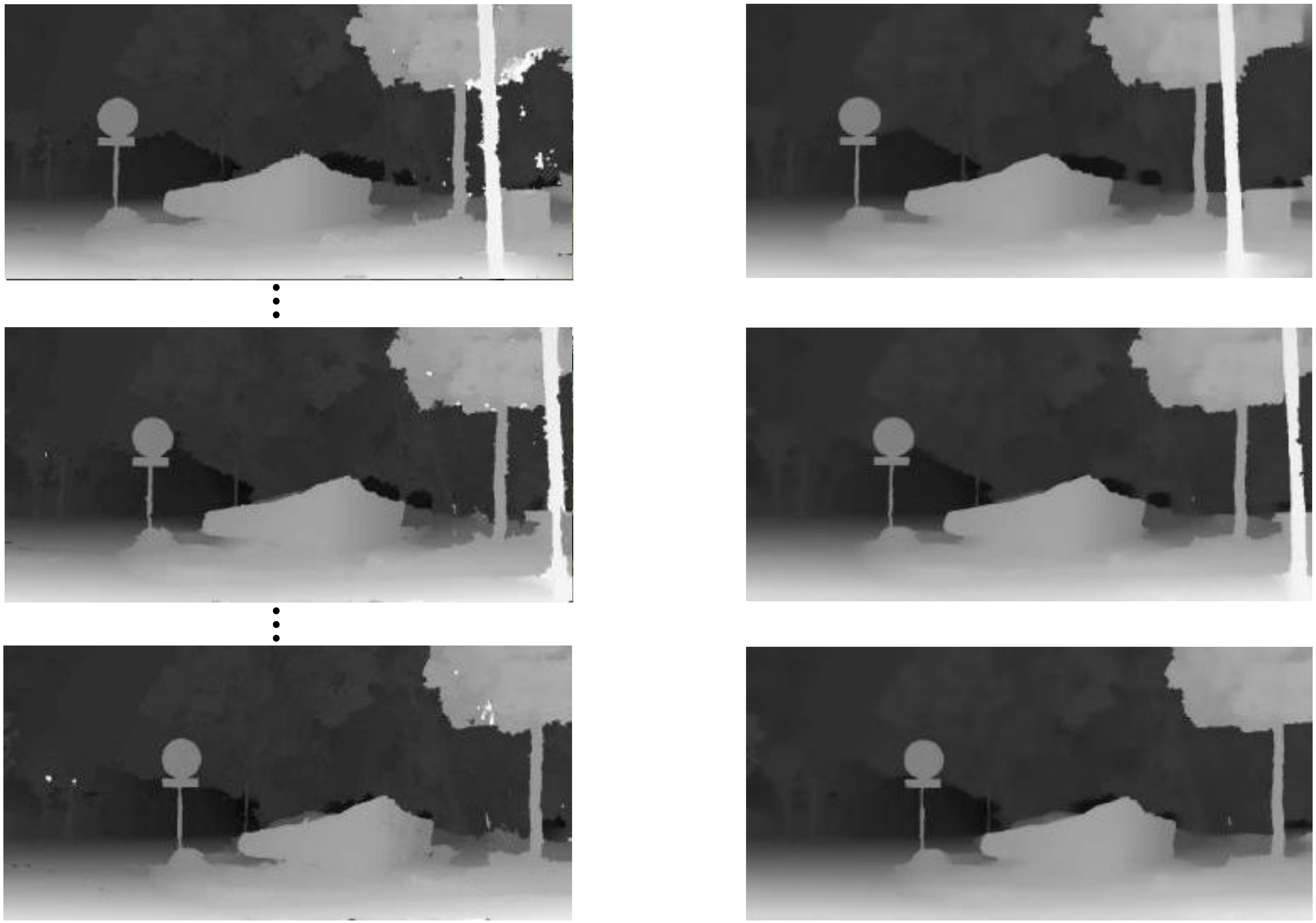
Our Key Idea



Our Key Idea

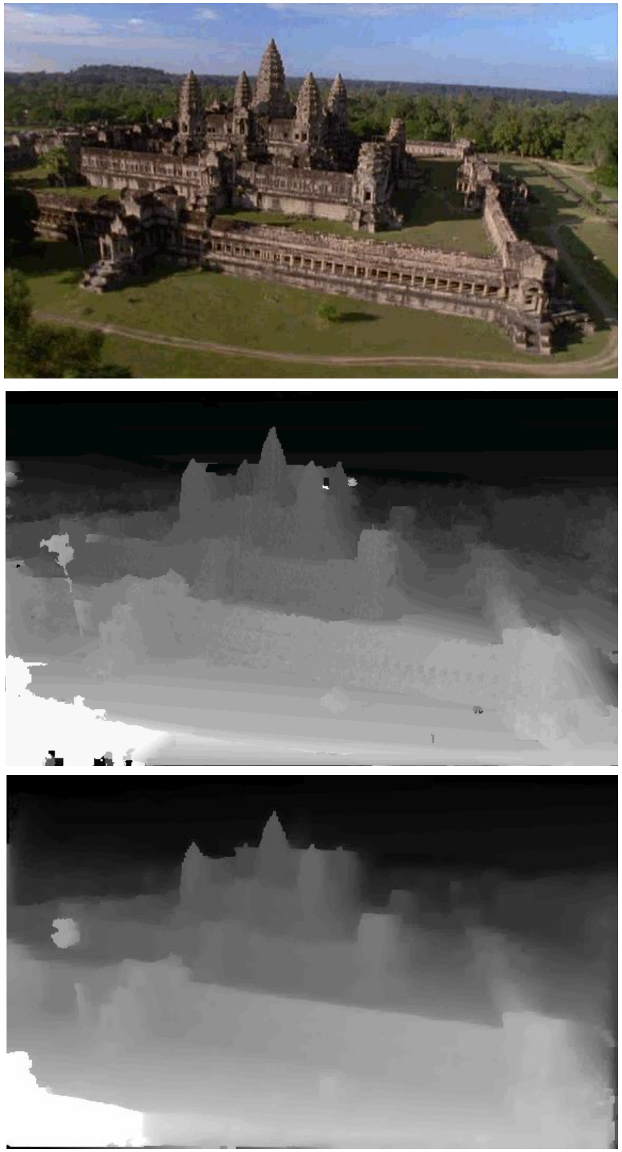


Our Key Idea



Framework Overview

- **Structure from Motion**
  - Recover the camera parameters
- **Depth Initialization**
  - Initialize depths without using segmentation
  - Refine the initialized depths with segmentation
- **Bundle Optimization**



Bundle Optimization

$$E(\hat{D};\hat{I}) = \sum_{t=1}^{n} \left(\boxed{E_d(D_t;\hat{I},\hat{D}\backslash D_t)} + \boxed{E_s(D_t)} \right)$$

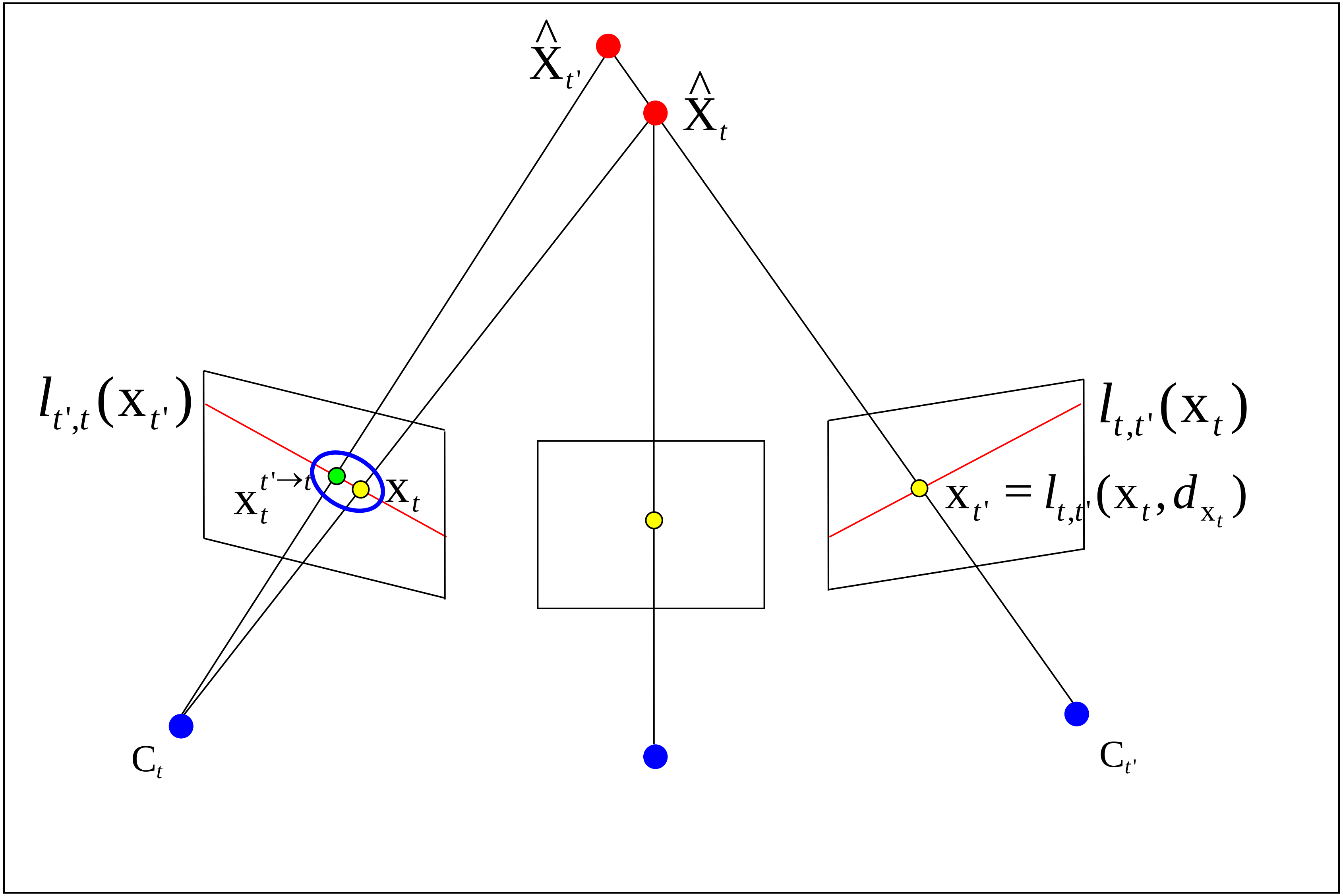
- E_d : Data Term
  - Color constancy constraint
  - Geometric coherence constraint
- E_s : Smoothness Term
  - Encodes the spatial smoothness

Bundle Optimization

$$E(\hat{D};\hat{I}) = \sum_{t=1}^{n} \left(E_d(D_t;\hat{I},\hat{D}\backslash D_t) + E_s(D_t)\right)$$

- E_d : Data Term
  - Color constancy constraint
  - Geometric coherence constraint
- E_s : Smoothness Term
  - Encodes the spatial smoothness

Geometric Coherence Constraint



Data Term Definition

- **Essential role** in energy minimization
  - Unreliable cost makes optimization problematic!
- The disparity likelihood
  - Combining color and geometry constraints
  - Complement each other

$$L(\mathbf{x}, d) = \sum_{t'} p_c(\mathbf{x}, d, I_t, I_{t'}) \cdot p_v(\mathbf{x}, d, D_{t'})$$

color constancy geometric coherence

Energy Definition

- The Complete Data Term

$$E_d(D_t; \hat{I}, \hat{D} \backslash D_t) = \sum_{\mathbf{x}} 1 - u(\mathbf{x}) \cdot L(\mathbf{x}, D_t(\mathbf{x}))$$

- The Smoothness Term

$$E_s(D_t) = \sum_{\mathbf{x}} \sum_{\mathbf{y} \in N(\mathbf{x})} \lambda(\mathbf{x}, \mathbf{y}) \cdot \min\{|D_t(\mathbf{x}) - D_t(\mathbf{y})|, \eta\}$$

How to Solve it?

$$E(\hat{D};\hat{I}) = \sum_{t=1}^{n} \left(E_d(D_t;\hat{I},\hat{D}\backslash D_t) + E_s(D_t)\right)$$

- Directly Solving the energy is intractable
  - Require Initial depth maps
- Iterative Optimization Scheme
  - Initialization
  - Iterative Refinement

Initialization

- Remove geometric coherence constraint
  - The disparity likelihood is reformed

$$L_{init}(\mathbf{x}, D_t(\mathbf{x})) = \sum_{t'} p_c(\mathbf{x}, D_t(\mathbf{x}), I_t, I_{t'})$$

- Estimate each frame independently

$$E_{init}^{t}(D_t; \hat{I}) = \sum_{\mathbf{x}} \Big(1 - u(\mathbf{x}) \cdot L_{init}(\mathbf{x}, D_t(\mathbf{x})) + \sum_{\mathbf{y} \in N(\mathbf{x})} \lambda(\mathbf{x}, \mathbf{y}) \cdot \rho(D_t(\mathbf{x}), D_t(\mathbf{y})) \Big)$$

- Incorporate segmentation
  - Improve disparity values in textureless regions

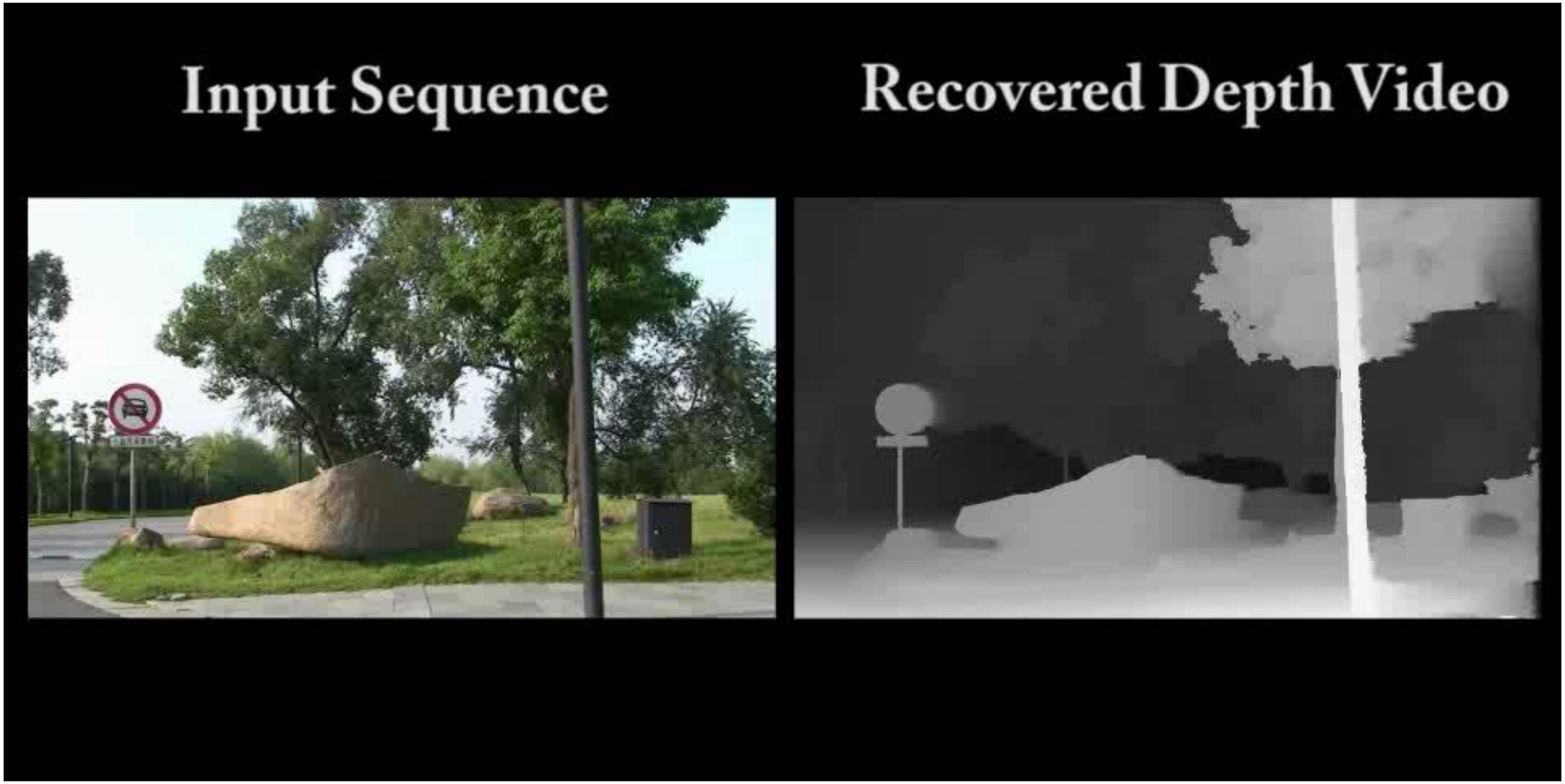
Iterative Optimization

- Solve the energy
  - Using loopy belief propagation.

$$E(\hat{D};\hat{I}) = \sum_{t=1}^{n} \left(E_d(D_t;\hat{I},\hat{D}\backslash D_t) + E_s(D_t) \right)$$

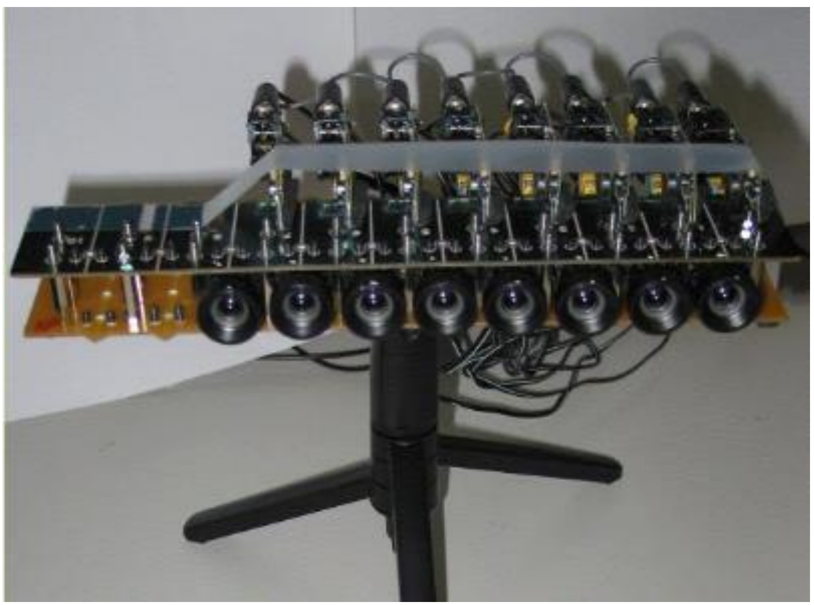
- Process frames from 1 to n:
  - For each frame *t*, fix disparities in other frames and refine D_{t}.
- Repeat the above step for 2~3 passes.

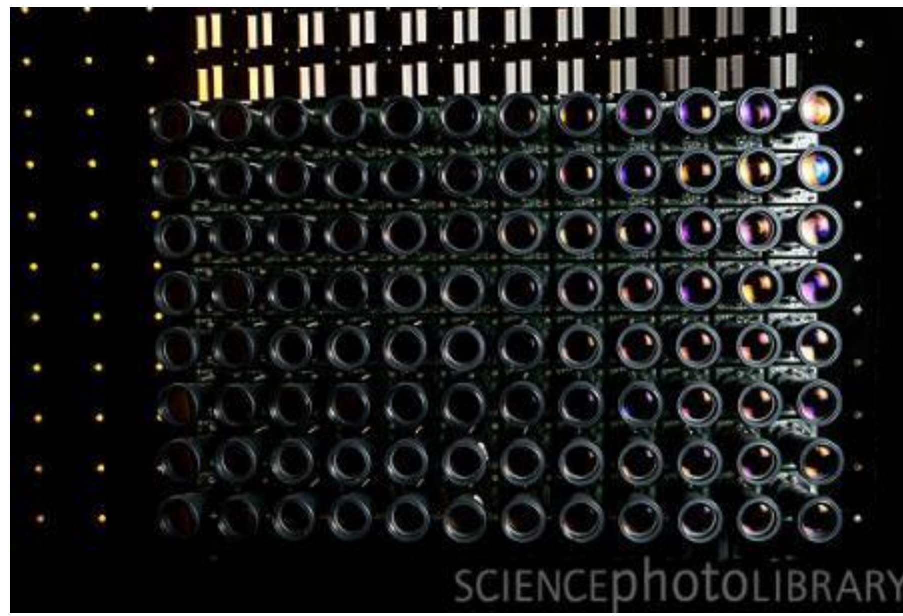
Results



High-Quality Depth Recovery of Dynamic Scenes

- Traditional methods require quite a number of synchronized cameras





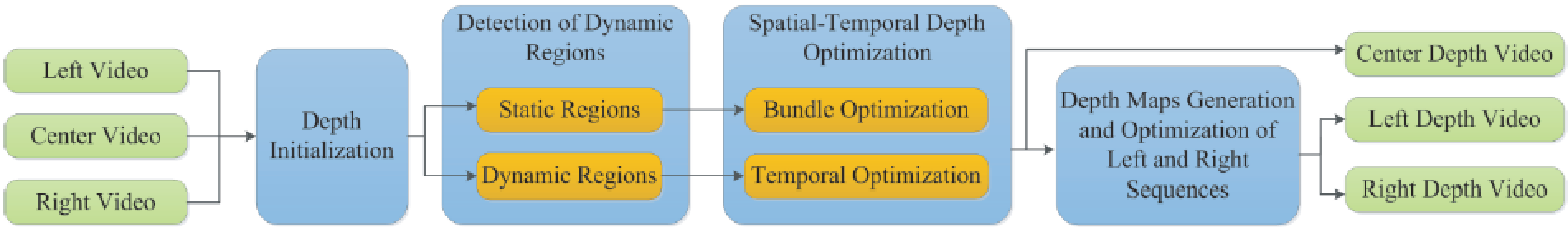
High-Quality Depth Recovery of Dynamic Scenes

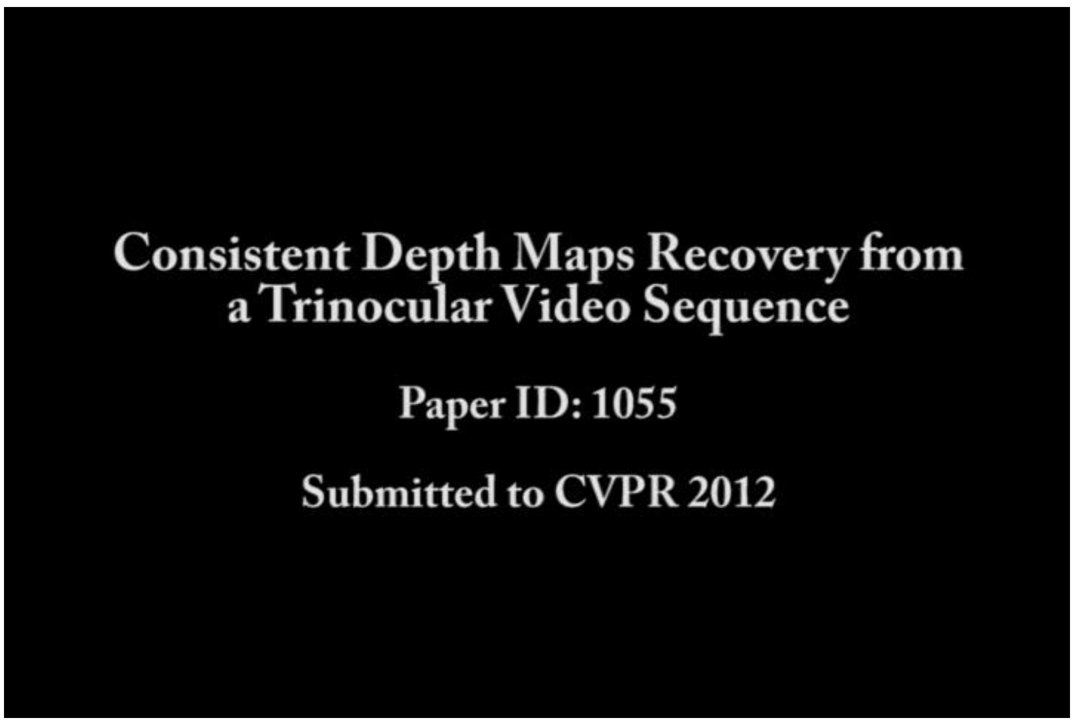
- Our methods
  - Trinocular Cameras

or

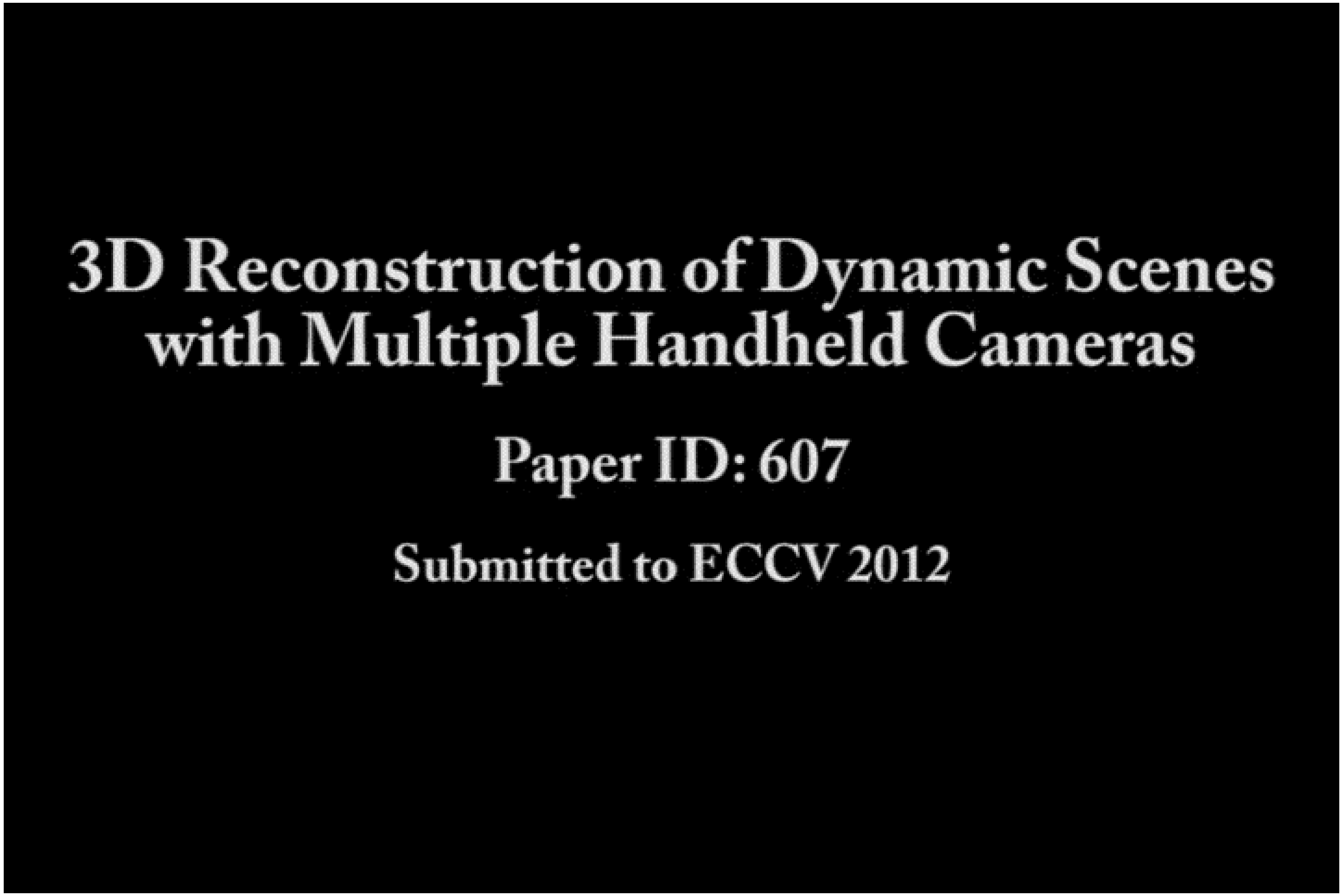
  - Few Handheld Cameras

Consistent Depth Maps Recovery from a Trinocular Video Sequence



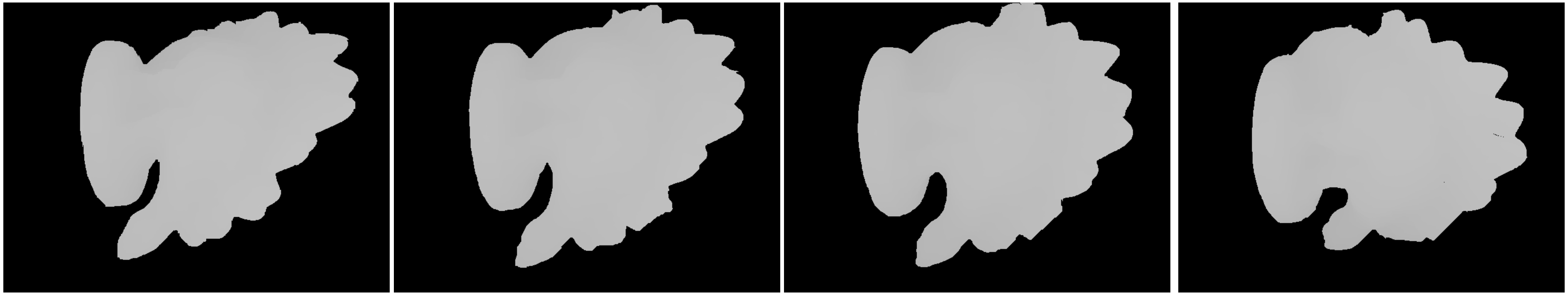


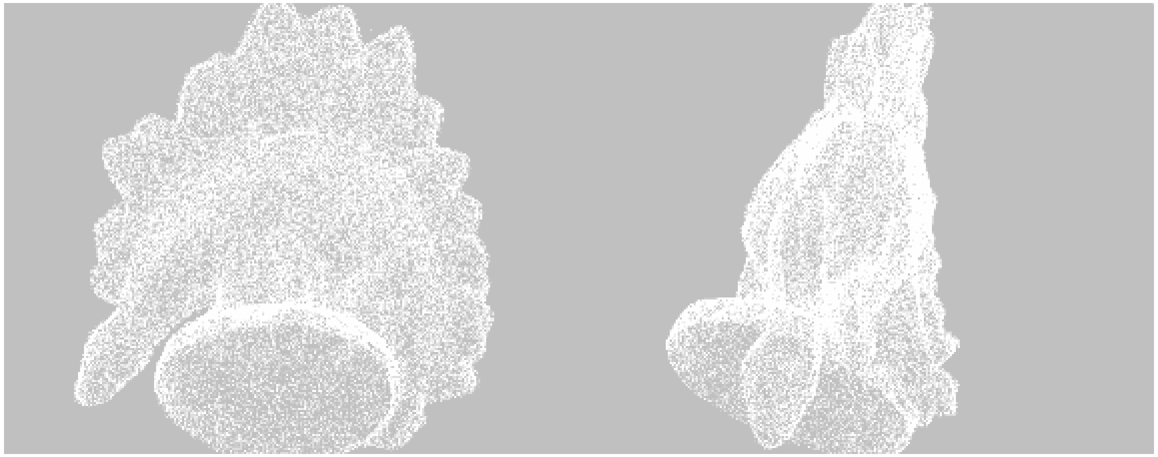
3D Reconstruction of Dynamic Scenes with Multiple Handheld Cameras



3D Reconstruction

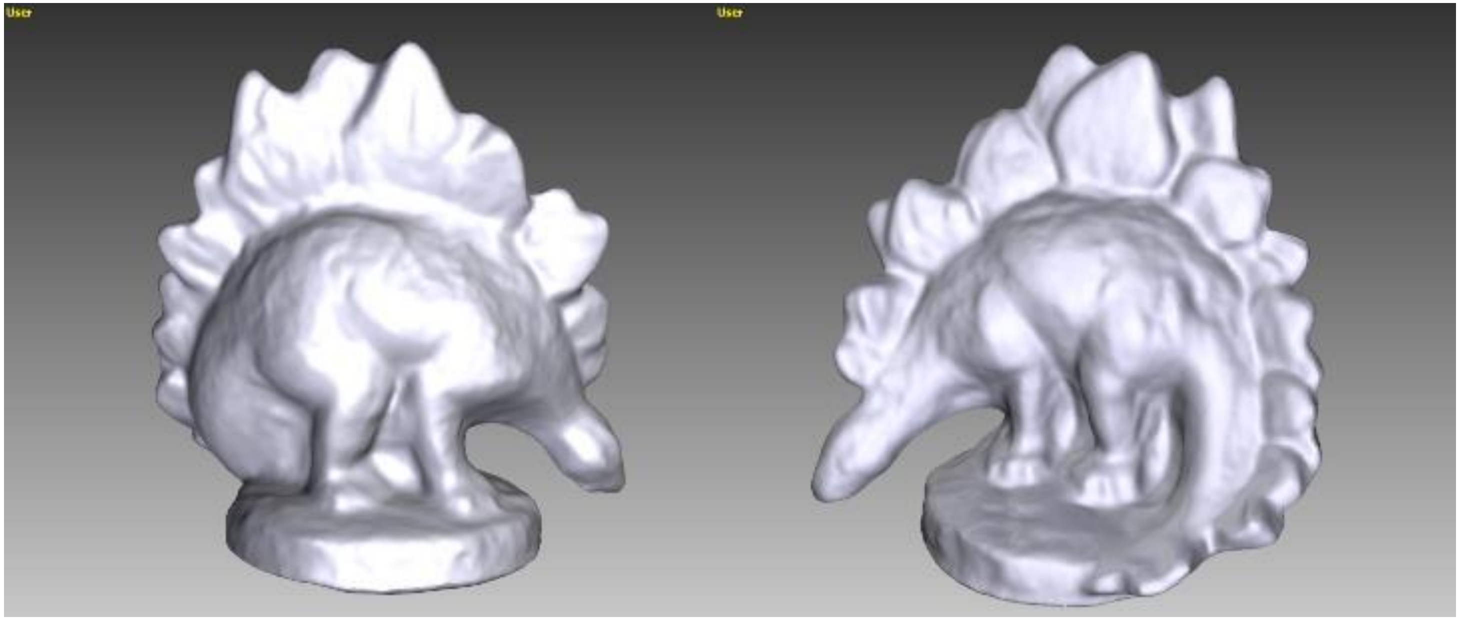
- Generate Point Samples from Depth Maps





3D Reconstruction

- Reconstructing 3D Surfaces from Point Samples

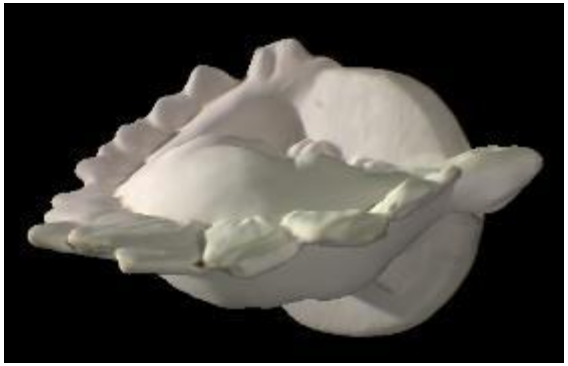


Poisson Surface Reconstruction
http://www.cs.jhu.edu/~misha/Code/PoissonRecon/

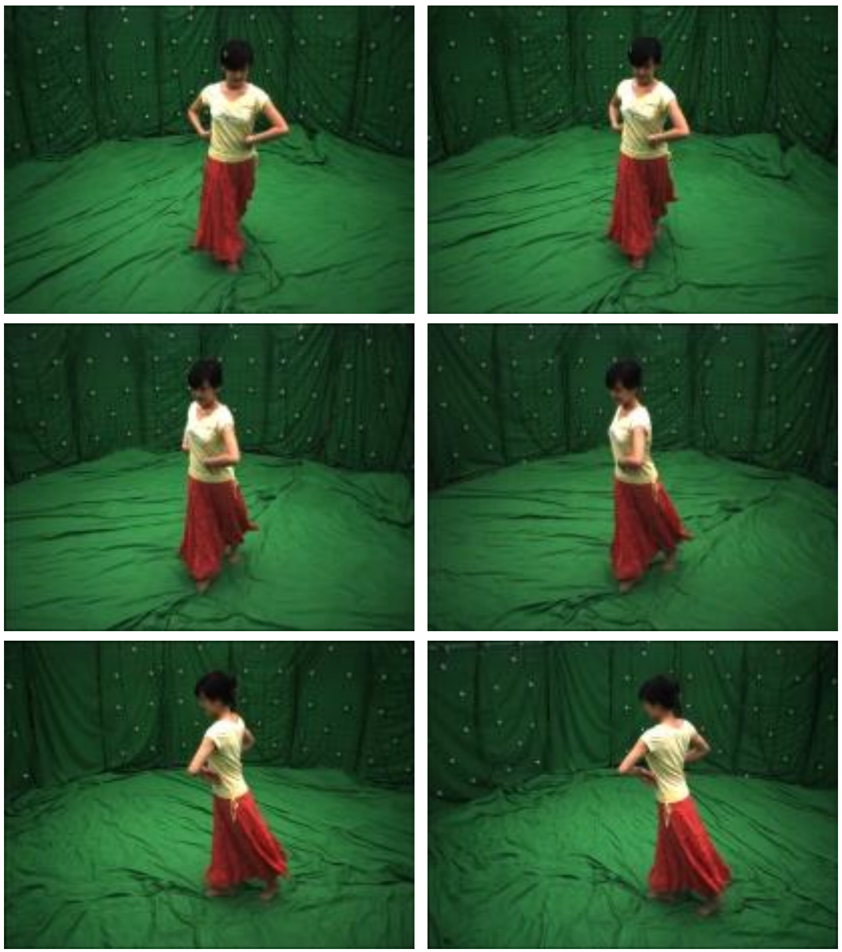
3D Reconstruction

- Texture Mapping

More Result



...

More Result

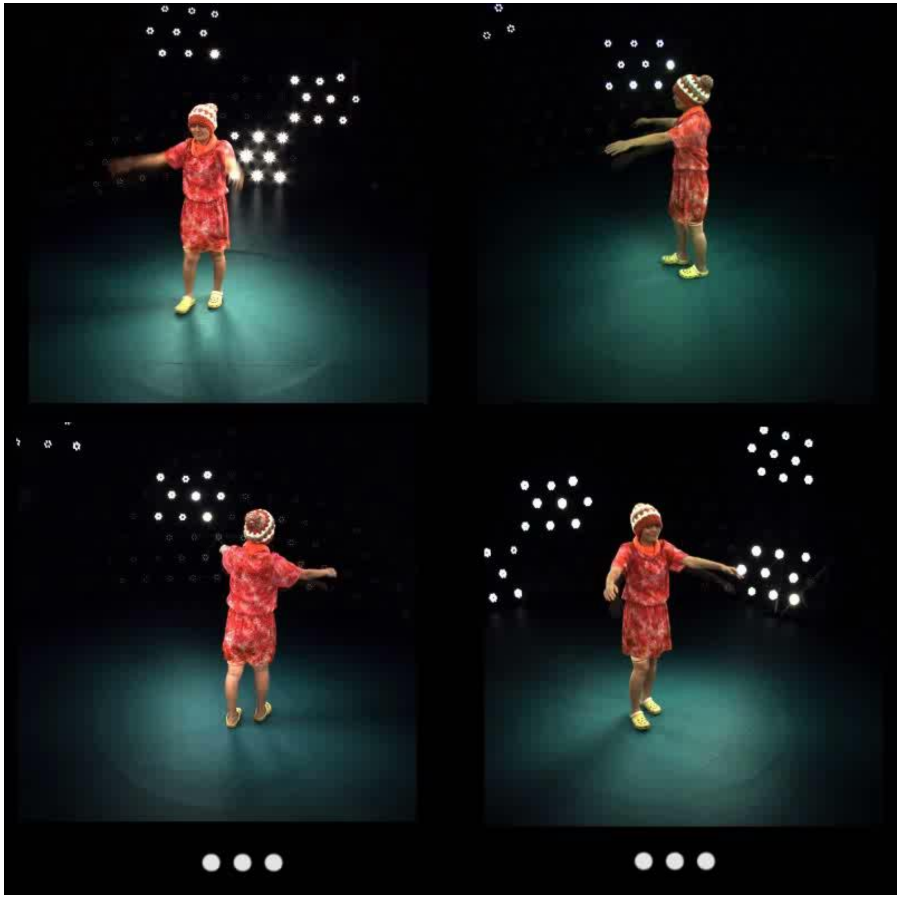
Applications

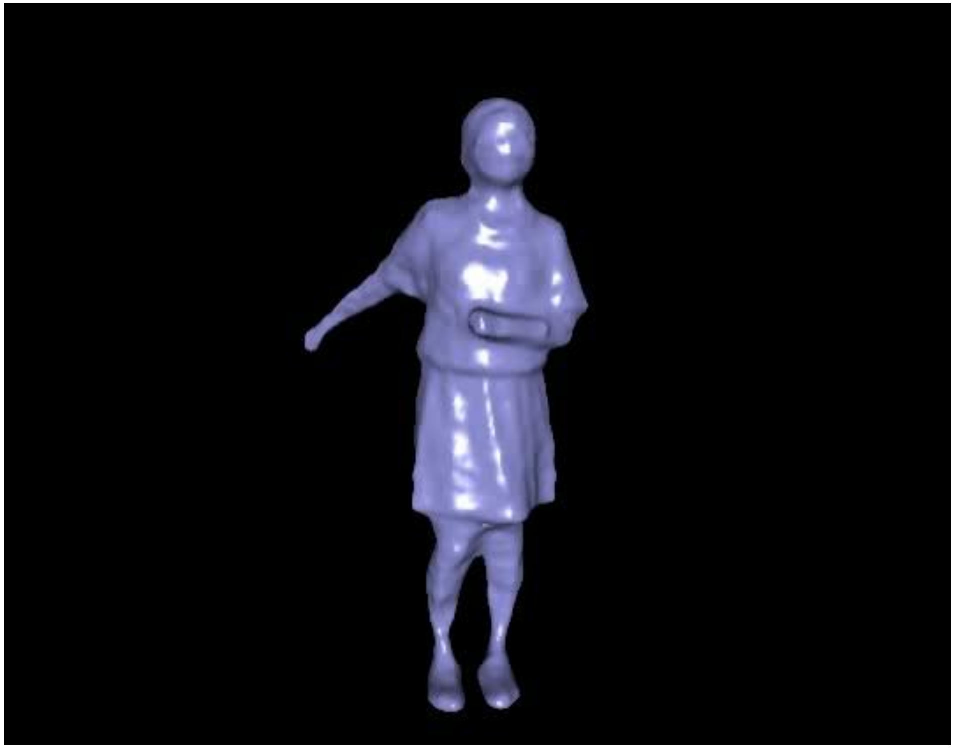
- Motion Capture
- Video Editing
- Spatio-Temporal Segmentation
- Motion Retargeting

Applications

- Motion Capture
- Video Editing
- Spatio-Temporal Segmentation
- Motion Retargeting

Motion Capture





Applications

- Motion Capture
- Video Editing
- Spatio-Temporal Segmentation
- Motion Retargeting

Refilming with Depth-Inferred Videos

More Results

Applications

- Motion Capture
- Video Editing
- Spatio-Temporal Segmentation
- Motion Retargeting

Spatio-Temporal Segmentation

Applications

- Motion Capture
- Video Editing
- Spatio-Temporal Segmentation
- Motion Retargeting

Motion Retargeting

Conclusions

- Facilitate many applications
  - 3D modeling, augmented reality, video editing, video segmentation,…
  - Also may be applicable to other fields:
    - video compression, analysis, and understanding

Thank you!