Volumetric Scene Reconstruction from Multiple Views

Chuck Dyer
University of Wisconsin
dyer@cs.wisc.edu
www.cs.wisc.edu/~dyer

Image-Based Scene Reconstruction

Images [Levoy and Hanrahan, 1996]
A range of viewpoints represented by a set of light fields.

Light Fields

Two General Approaches

World Representation
• World centered: Recover a complete 3D geometric model of the scene and possibly photometric information from motion, model fitting, calibration, structure from motion, and tracking.
• Operations: feature correspondence, tracking, calibration, structure from motion, model fitting, ...

Plenoptic Function Representation
• Camera centered: Integration of images which sample scene geometry
• E.g., panoramas, light fields, LDIs
• Operations: Image segmentation, registration, warping, compositing, interpolation, ...
• World centered: Recovery of a complete 3D geometric model of scene

Applications
• Image-based modeling, 3D photogrammetry
• Image-based virtual environments, video-based models of a scene from multiple images taken from a set of arbitrary viewpoints
• Automatic construction of photo-realistic 3D models of a scene
• Interactive visualization of remote environments or objects

Applications

Light Fields
A range of viewpoints represented by a set of images
[Levoy and Hanrahan, 1996]
Standard Approach: Multiple View Stereo

Weaknesses of the Standard Approach

- Views must be close together in order to obtain point correspondences.
- Point correspondences must be tracked over many consecutive frames.
- Many partial models must be fused.
- Must fit a parameterized surface model to point features.
- No explicit handling of occlusion differences between views.

Our Approach: Volumetric Scene Modeling

Goal: Determine transparency and radiance of points in a volume.

3D Scene Reconstruction from Multiple Views

Input Images

Camera calibration

Scene Volume

(Fitzgibbon and Zisserman, 1998)
Discrete Formulation:

Voxel Space

Goal: Assign RGBA values to voxels in $V$ that are consistent with all input images

$S \subseteq \mathcal{C}$
$S$ = true scene (not computable)
$P = \text{space of all photo-consistent colorings (computable)}$
$G = \text{space of all colorings (computable)}$

$G \subseteq \mathcal{C}$
$N$ = number of voxels
$N \times V$ = scene volume

Complexity and Computability

1. Shape from Silhouettes
   - Volume intersection
   - [Martin & Aggarwal, 1983]

2. Shape from Photo-Consistency
   - Photo-consistency with all input images
   - [Seitz & Dyer, 1997]
   - [Kutulakos & Seitz, 1999]

Reconstruction from Silhouettes

Approach:

- Backproject each silhouette
- Intersect backprojected generalized cone volumes

Voxel-based Scene Reconstruction Methods

1. Shape from Silhouettes
2. Shape from Photo-Consistency
Volume Intersection contains the true scene.

Best case (infinite # views):

- Visual hull
- Convex hull
- Hyperbolic regions

Shape from Silhouettes

Image-based Visual Hulls

Voxel Algorithm for Volume Intersection

Don't have to search $2^N$ possible scenes.

- $O(MN^3)$ time for M images, N voxels
- Color voxel black if in silhouette in every image

Reconstruction:
- Object + concavities + points not visible

Reconstruction contains the true scene

Visible = object + concavities + points not visible
Shape from 49 Silhouettes

Surface model constructed using Marching Cubes algorithm

Properties of Volume Intersection

Pros
- Easy to implement
- Accelerated via octrees

Cons
- Concavities are not reconstructed
- Reconstruction does not use photometric properties in each image
- Requires image segmentation to extract silhouettes
- Reconstruction does not use photometric properties

Texture mapped and sound synthesized from 6 sources

CMU's Virtualized Reality System

Virtual Camera Fly-By
1. **Shape from Silhouettes**
   - The Global Visibility Problem
   - Based Scene Reconstruction Methods

   Which points are visible in which images?

   **Voxel Coloring Approach**

2. **Shape from Photo**
   - Consistency

   The Global Visibility Problem

   In which images is each voxel visible?

   - Forward Visibility
   - Inverse Visibility

   Which points are visible in which images?

   **Voxel Coloring Approach**

   **Depth Ordering:** Visit Occluders First

   **Visibility Problem:** In which images is each voxel visible?

   1. Choose voxel
   2. Project and correlate
   3. Color the photo-consistent

   **Voxel-based Scene Reconstruction Methods**

   - Space carving (Kim and Szeliski, 1999)
   - Voxel coloring (Szeliski, 1997)
   - Shape from Photo-consistency
   - Volume intersection (Wiman, 1993)
   - Shape from Silhouettes
What is a View-Independent Depth Order?

A function \( f \) over a scene \( S \) and a camera space \( C \) such that for all \( p \) and \( q \) in \( S \), \( v \) in \( C \):

\[
f(p) < f(q) \quad \text{if} \quad p \text{ occludes } q \text{ from } v.\]

Example: 2D Scene and Line of Cameras

Panoramic Depth Ordering

- Layers radially outwards from cameras
- Plane sweep order

Panoramic Depth Ordering

- Cameras oriented in many different directions
- Planar depth ordering does not apply

Example: 2D Scene and Line of Cameras
Panoramic Layering

Layers radiate outwards from cameras

Calibrated Image Acquisition

Compatible Camera Configurations

• Scene outside convex hull of camera centers
• Scene inside convex hull of camera centers

Calibrated Turntable
360° rotation (21 images)

Selected Dinosaur Images

Selected Flower Images

Cameras inside scene

Outward-looking

Cameras above scene

Inward-looking

Depth-Order Constraint
Layered Scene Traversal

Results: Dinosaur

7.6M voxels trained
7.0K voxels colored
21 input images
5K voxels colored
21 input images
1K voxels colored
21 input images

Results: Rose

1 of 21 images
3 synthesized views

Layered Scene Traversal

Results: Dinosaur

21 input images
spanning 360° rotation

Results: Rose
### Scaling Up Voxel Coloring

- **Time complexity**: \( \# \text{voxels} \times \# \text{images} \)
- **Coarse-to-Fine Voxel Coloring**: Octrees
  - Determine colored voxels at current level
  - Spatial coherence and neighboring voxels
  - Decompose colored voxels into octants; repeat

### Volumetric Warping

- **Enhancements**
  - Texture mapping: use hardware to project images to each layer of voxels
  - Variable voxel resolution: use octrees and coarse-to-fine processing
  - Variable voxel resolution: use octrees and coarse-to-fine processing
  - Texture mapping: use hardware to project images to each layer of voxels

- **G. Slabaugh, T. Malzbender, B. Culbertson, 2000**

### Results

- Initial domain: Volumetric warping – warp voxel space to extend to an infinite domain
- Time complexity: \( \# \text{voxels} \times \# \text{images} \)
- Too many voxels: in large, high-resolution scenes
- Volumetric warping – use octrees and coarse-to-fine processing
- Inadequate voxel resolution – use octrees and coarse-to-fine processing
Voxel Coloring for Dynamic Scenes

Given: Video sequences from multiple cameras
Goal: Interactive, real-time fly-by of dynamic scene

Reconstruction for One Time Instant

Voxel Coloring: Input Views

Sequence of Reconstructions
Voxel Coloring for Dynamic Scenes

• Coarse-to-fine recursive decomposition focuses
• Exploit temporal coherence
• Trace rays from changed pixels only
• Voxel Coloring [Seitz & Dyer, 1997]
• Voxel Coloring [Martin & Aggarwal, 1983]

Space Carving Algorithm

Step 1: Initialize V to volume containing true scene with all voxels marked opaque

Step 2: For every voxel on surface of V
• Test photo-consistency of voxel with those cameras that are “in front of” it
• If voxel is inconsistent, carve it (i.e., mark it transparent)

Step 3: Repeat Step 2 until all voxels consistent

Volume Intersection [Kutulakos & Seitz, 1999]

Voxel-based Scene Reconstruction Methods

1. Shape from Silhouettes
   • Volume Intersection [Martin & Aggarwal, 1983]
   • Shape from Photo-Consistency

2. Shape from Photo
   • Consistency
   • Voxel Coloring

Limitations of Depth Ordering

A view-independent depth order may not exist:

Unconstrained camera positions
Unconstrained scene geometry and topology

Need more general algorithm

1. Shape from Silhouettes

Limitations of Depth Ordering

A view-independent depth order may not exist:

Unconstrained camera positions
Unconstrained scene geometry and topology

Need more general algorithm
This property ensures that carving converges
\[ p \in S \iff p \notin S \]

Visibility Property

Multi-Pass Plane Sweep Algorithm

- Easy to implement
- Converges quickly in practice
- Efficient: can use texture-mapping hardware

Optimal algorithm is unwieldy

Space Carving Algorithm

- Guaranteed convergence to the photo hull
- Union of all photo-consistent scenes

Complex visibility update procedure

Worst case # consistency checks:
\[ \text{# cameras} \times \text{# voxels} \]

Optimal algorithm is unwieldy

Multi-Pass Plane Sweep Algorithm

- Repeat until convergence
- Consider cameras on only one side of plane
- Sweep plane in each of 6 principle directions

Optimal algorithm is unwieldy

Multi-Pass Plane Sweep Algorithm

- Sweep plane in each of 6 principle directions
- Consider cameras on only one side of plane
- Repeat until convergence

True Scene
Reconstruction

False Scene
Reconstruction

This property ensures that carving converges
\[ p \in S \iff p \notin S \]

Visibility Property
Multi-pass Plane Sweep

- Sweep plane in each of 6 principle directions
- Consider cameras on only one side of plane
- Repeat until convergence
Multi-pass Plane Sweep

- Sweep plane in each of 6 principle directions
- Consider cameras on only one side of plane
- Repeat until convergence
Repeat until convergence
Consider cameras on only one side of plane
Sweep plane in each of 6 principle directions

Multi-Pass Plane Sweep

Repeat until convergence
Consider cameras on only one side of plane
Sweep plane in each of 6 principle directions

Multi-Pass Plane Sweep

Repeat until convergence
Consider cameras on only one side of plane
Sweep plane in each of 6 principle directions

Multi-Pass Plane Sweep

Repeat until convergence
Consider cameras on only one side of plane
Sweep plane in each of 6 principle directions

Multi-Pass Plane Sweep

Repeat until convergence
Consider cameras on only one side of plane
Sweep plane in each of 6 principle directions

Multi-Pass Plane Sweep
Multi-Pass Plane Sweep

- Repeat until convergence
- Consider cameras on only one side of plane
- Sweep plane in each of 6 principle directions
- Repeat until convergence
Multi-Pass Plane Sweep

- Repeat until convergence
- Consider cameras on only one side of plane
- Sweep plane in each of 6 principle directions

Repeat until convergence
- Consider cameras on only one side of plane
- Sweep plane in each of 6 principle directions

Repeat until convergence
- Consider cameras on only one side of plane
- Sweep plane in each of 6 principle directions

Repeat until convergence
- Consider cameras on only one side of plane
- Sweep plane in each of 6 principle directions
Multi-Pass Plane Sweep

• Sweep plane in each of 6 principle directions
• Consider cameras on only one side of plane
• Repeat until convergence

Results: African Violet

Results: Hand

Texture Effects on Voxel Coloring

Multi-Pass Plane Sweep

- Repeat until convergence
- Consider cameras on only one side of plane
- Sweep plane in each of 6 principle directions
Voxel Coloring / Space Carving Summary

Effects of Voxel Resolution

Other Extensions

The more the marble wastes, the more the statue grows.
— Michelangelo

Collaborators
• Steve Seitz, Andrew Prock, Kyros Kutulakos

Pros
• Guaranteed convergence
• Camera positions unconstrained
• Can model arbitrary geometry and topology

Cons
• Guaranteed convergence
• Camera positions unconstrained
• Can model arbitrary geometry and topology

Other Extensions

Dealing with Calibration Errors

• Kutulakos, 2000

• Partly transparent scenes

• DeBonet and Viola, 1999

• Construct approximate photo hull defined by weakening the definition of photo-consistency so that it requires only that there exists a photo-consistent pixel within distance $r$ of the ideal position

Pros
• Non-parametric
• Can model arbitrary geometry and topology

Cons
• Expensive to process high-resolution voxel grids
• Carving stops at first consistent voxel, not best
• Expensive to process high-resolution voxel grids

Effects of Noise

Collaborators
• Steven Seitz, Andrew Prock, Kyros Kutulakos

Pros
• Guaranteed convergence
• Camera positions unconstrained
• Can model arbitrary geometry and topology

Cons
• Guaranteed convergence
• Camera positions unconstrained
• Can model arbitrary geometry and topology

Voxel Coloring / Space Carving Summary

Effects of Voxel Resolution

Other Extensions

Dealing with Calibration Errors

• Kutulakos, 2000

• Partly transparent scenes

• DeBonet and Viola, 1999

• Construct approximate photo hull defined by weakening the definition of photo-consistency so that it requires only that there exists a photo-consistent pixel within distance $r$ of the ideal position

Pros
• Non-parametric
• Can model arbitrary geometry and topology

Cons
• Expensive to process high-resolution voxel grids
• Carving stops at first consistent voxel, not best
• Expensive to process high-resolution voxel grids

Effects of Noise
Current Work

• BRDF estimation from multiple views
• Modeling is more than geometry – need to simultaneously recover surface reflectance and geometry
• Calibration from multiple moving objects
• Wide-baseline feature point correspondence
• Metric self-calibration from static scenes