Multi-view stereo

Many slides adapted from S. Seitz, Y. Furukawa
Multi-view stereo

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape
Multi-view stereo

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape
- “Images of the same object or scene”
  - Arbitrary number of images (from two to thousands)
  - Arbitrary camera positions (special rig, camera network or video sequence)
  - Calibration may be known or unknown
Multi-view stereo

• Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape

• “Images of the same object or scene”
  • Arbitrary number of images (from two to thousands)
  • Arbitrary camera positions (special rig, camera network or video sequence)
  • Calibration may be known or unknown

• “Representation of 3D shape”
  • Depth maps
  • Meshes
  • Point clouds
  • Patch clouds
  • Volumetric models
  • …
Multi-view stereo: Basic idea

Source: Y. Furukawa
Multi-view stereo: Basic idea

error

depth

Source: Y. Furukawa
Multi-view stereo: Basic idea

Source: Y. Furukawa
Multi-view stereo: Basic idea

Source: Y. Furukawa
Multiple-baseline stereo

Figure 2: An example scene. The grid pattern in the background has ambiguity of matching.

M. Okutomi and T. Kanade, A Multiple-Baseline Stereo System, PAMI 1993
Multiple-baseline stereo

- For each pixel in reference image, simultaneously compute matching scores w.r.t. all the other images

M. Okutomi and T. Kanade, *A Multiple-Baseline Stereo System*, PAMI 1993
Multiple-baseline stereo

Use the sum of SSD scores to find the best depth

Fig. 7. Combining multiple baseline stereo pairs.

M. Okutomi and T. Kanade, *A Multiple-Baseline Stereo System*, PAMI 1993
Multiple-baseline stereo

M. Okutomi and T. Kanade, A Multiple-Baseline Stereo System, PAMI 1993
Plane sweep stereo

- Sweep family of planes at different depths w.r.t. a reference camera
- For each depth, project each input image onto that plane (homography) and compare the resulting stack of images

Plane sweep stereo

- Sweep family of planes at different depths w.r.t. a reference camera.
- For each depth, project each input image onto that plane (homography) and compare the resulting stack of images.

Plane sweep stereo

Scene surface

Sweeping plane

Image 1

Image 2
Plane sweep stereo

- For each depth plane
  - For each pixel in the composite image stack, compute the variance
  - For each pixel, select the depth that gives the lowest variance

- Can be accelerated using graphics hardware

Merging depth maps

- Given a group of images, compute a depth map using each view for reference
- Merge multiple depth maps to a volume or a mesh (see, e.g., Curless and Levoy 96)
Volumetric stereo

**Goal:** Assign RGB values to voxels in V \( \text{photo-consistent} \) with images
Space Carving

Space Carving Algorithm

- Initialize to a volume $V$ containing the true scene
- Choose a voxel on the outside of the volume
- Project to visible input images
- Carve if not photo-consistent
- Repeat until convergence

Space Carving Results

Input Image (1 of 45)

Reconstruction

Source: S. Seitz
Space Carving Results

Input Image
(1 of 100)

Reconstruction

Source: S. Seitz
What reconstruction do you get with space carving?

A photo-consistent scene is a scene that exactly reproduces your input images from the same camera viewpoints.

Space carving gives you the “largest” such scene.

Source: S. Seitz
Reconstruction from Silhouettes

- The case of binary images: a voxel is photo-consistent if it lies inside the object’s silhouette in all views.
Reconstruction from Silhouettes

- The case of binary images: a voxel is photo-consistent if it lies inside the object’s silhouette in all views.

Finding the silhouette-consistent shape (*visual hull*):

- *Backproject* each silhouette
- Intersect backprojected volumes
Volume intersection

Photo-consistency vs. silhouette-consistency

True Scene  Photo Hull  Visual Hull
Patch-based multi-view stereo (PMVS)

1. Detect keypoints
2. Triangulate a sparse set of initial matches
3. Iteratively expand matches to nearby locations
4. Use visibility constraints to filter out false matches
5. Perform surface reconstruction


PMVS software
Patch-based multi-view stereo (PMVS)

PMVS software
Stereo from community photo collections

N. Snavely, S. Seitz, R. Szeliski,
Photo tourism: Exploring photo collections in 3D, SIGGRAPH 2006
Stereo from community photo collections

- Need *structure from motion* to recover unknown camera parameters
- Need *view selection* to find good groups of images on which to run dense stereo
Local view selection

M. Goesele et al., Multi-View Stereo for Community Photo Collections, ICCV 2007
Local view selection

M. Goesele et al., Multi-View Stereo for Community Photo Collections, ICCV 2007
Local view selection

M. Goesele et al., Multi-View Stereo for Community Photo Collections, ICCV 2007
Towards Internet-Scale Multi-View Stereo

Y. Furukawa, B. Curless, S. Seitz and R. Szeliski,
Towards Internet-scale Multi-view Stereo, CVPR 2010.
Towards Internet-Scale Multi-View Stereo

YouTube video, high-quality video, CMVS software

Y. Furukawa, B. Curless, S. Seitz and R. Szeliski,
Towards Internet-scale Multi-view Stereo, CVPR 2010.
Fast stereo for Internet photo collections

- Start with a cluster of registered views
- Obtain a depth map for every view using plane sweeping stereo with normalized cross-correlation

J.-M. Frahm et al., Building Rome on a Cloudless Day, ECCV 2010
Plane sweeping stereo

- Need to register individual depth maps into a single 3D model
- Problem: depth maps are very noisy

J.-M. Frahm et al., Building Rome on a Cloudless Day, ECCV 2010
Robust stereo fusion using a heightmap

- Enforces vertical facades
- One continuous surface, no holes
- Fast to compute, low memory complexity

D. Gallup, M. Pollefeys, J.-M. Frahm,
3D Reconstruction using an n-Layer Heightmap, DAGM 2010
Results

J.-M. Frahm et al., *Building Rome on a Cloudless Day*, ECCV 2010
The Visual Turing Test for Scene Reconstruction

Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. Seitz,
The Visual Turing Test for Scene Reconstruction, 3DV 2013.
Ongoing research directions

Challenging lighting conditions

Indoor modeling

Ground/aerial

Dynamic reconstruction
KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera

Shahram Izadi¹, David Kim¹,³, Omar Hilliges¹, David Molyneaux¹,⁴, Richard Newcombe², Pushmeet Kohli¹, Jamie Shotton¹, Steve Hodges¹, Dustin Freeman¹,⁵, Andrew Davison², Andrew Fitzgibbon¹

¹Microsoft Research Cambridge, UK ²Imperial College London, UK ³Newcastle University, UK ⁴Lancaster University, UK ⁵University of Toronto, Canada

Figure 1: KinectFusion enables real-time detailed 3D reconstructions of indoor scenes using only the depth data from a standard Kinect camera. A) user points Kinect at coffee table scene. B) Phong shaded reconstructed 3D model (the wireframe frustum shows current tracked 3D pose of Kinect). C) 3D model texture mapped using Kinect RGB data with real-time particles simulated on the 3D model as reconstruction occurs. D) Multi-touch interactions performed on any reconstructed surface. E) Real-time segmentation and 3D tracking of a physical object.

Paper link (ACM Symposium on User Interface Software and Technology, October 2011)

YouTube Video