Multi-view stereo

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

- Goal: given several images of the same object or scene, compute a representation of its 3D shape

Source: C. Hernandez, N. Snavely
Multi-view stereo

• Goal: 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)
  • Calibration may be known or unknown
Multi-view stereo

• Goal: 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)
  • Calibration may be known or unknown
• “Representation of 3D shape”
  • Depth maps
  • Meshes
  • Point clouds
  • Patch clouds
  • Volumetric models
  • ....
Outline

• Applications and motivation
• Plane sweep stereo
• Depth map fusion
• Patch-based multi-view stereo (PMVS)
• Stereo from Internet photo collections
• Recent trends
Applications

Source: N. Snavely
Applications

Source: N. Snavely
Applications

Source: N. Snavely
Applications

• Enable inspection in hard to reach areas with drone photos and 3D reconstruction
• Create 3D model from images
• Provide tools to inspect on images and map interactions to 3D

Source: D. Hoiem
Multi-view stereo: Basic idea

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

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

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

Source: Y. Furukawa
Why MVS?

• Different points on the object’s surface will be more clearly visible in some subset of cameras
  • Could have high-res closeups of some regions
  • Some surfaces are foreshortened from certain views
  • Some points may be occluded entirely in certain views

Source: N. Snavely
Cameras 4 and 5 can more clearly see point p.
Cameras 3 and 4 can more clearly see point q.

Source: N. Snavely
Camera 5 can't see point r.

Source: N. Snavely
Camera 1 can’t see point s.

Source: N. Snavely
Why MVS?

- Different points on the object’s surface will be more clearly visible in some subset of cameras
  - Could have high-res closeups of some regions
  - Some surfaces are foreshortened from certain views
  - Some points may be occluded entirely in certain views
- More measurements per point can reduce error

Source: N. Snavely
Estimated points contain some error.

Source: N. Snavely
Estimated points contain some error.

Source: N. Snavely
Estimated points contain some error.

Source: N. Snavely
Additional views reduce error.

Source: N. Snavely
Outline

• Applications and motivation
• Plane sweep stereo
• Depth map fusion
• Patch-based multi-view stereo (PMVS)
• Stereo from Internet photo collections
• Recent trends
Plane sweep stereo

- Sweep plane across a range of 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 plane across a range of 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

R. Collins, A space-sweep approach to true multi-image matching, CVPR 1996
Plane sweep stereo: Key idea

- Scene surface
- Sweeping plane
- Image 1
- Reference image
- Image 2
Plane sweep stereo: Key idea
Plane sweep stereo: Key idea
Plane sweep stereo: Key idea
Plane sweep stereo: Key idea
Plane sweep stereo: Key idea

Image 1

Image 2
Plane sweep stereo: Key idea
Plane sweep stereo: Fast implementation

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

Merging depth maps

• Given a group of images, compute a depth map using each view as a reference
• Merge multiple depth maps into a volume or a mesh (see, e.g., Curless and Levoy, 1996)
Volumetric fusion

Source: N. Snavely
Volumetric fusion

Source: N. Snavely
Fast depth map fusion using height maps

• Start with a cluster of registered views (from SFM on Internet photo collections)

J.-M. Frahm et al., Building Rome on a Cloudless Day, ECCV 2010
D. Gallup et al. 3D Reconstruction using an n-Layer Heightmap, DAGM 2010
Fast depth map fusion using height maps

- Obtain a (noisy) 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
D. Gallup et al. 3D Reconstruction using an n-Layer Heightmap, DAGM 2010
Fast depth map fusion using height maps

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

J.-M. Frahm et al., *Building Rome on a Cloudless Day*, ECCV 2010
D. Gallup et al. *3D Reconstruction using an n-Layer Heightmap*, DAGM 2010
Fast depth map fusion using height maps

YouTube Video

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

• Applications and motivation
• Plane sweep stereo
• Depth map fusion
• Patch-based multi-view stereo (PMVS)
• Stereo from Internet photo collections
• Recent trends
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

- Need \textit{structure from motion} to recover unknown camera parameters
- Need \textit{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
Local view selection

Notre Dame de Paris

653 images
313 photographers

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

Model merged from 72 depth maps

Model from 56 depth maps with laser scan overlaid (90% of points within 0.25% of ground truth)

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, CMVS software

Y. Furukawa, B. Curless, S. Seitz and R. Szeliski, Towards Internet-scale Multi-view Stereo, CVPR 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
Fig. 6. Reference image with filtered depths and normals for crowd-sourced images.


Results video
Outline

- Applications and motivation
- Plane sweep stereo
- Patch-based multi-view stereo (PMVS)
- Stereo from Internet photo collections
- Recent trends
Ongoing research directions

Challenging lighting conditions

Indoor modeling

Ground/aerial

Dynamic reconstruction
Deep learning for MVS

Y. Yao et al. MVSNet: Depth Inference for Unstructured Multi-view Stereo, ECCV 2018
Deep learning for MVS

Y. Yao et al. **MVSNet: Depth Inference for Unstructured Multi-view Stereo**, ECCV 2018
Deep learning for improving SFM

P. Lindenberger et al. *Pixel-Perfect Structure-from-Motion with Featuremetric Refinement*. ICCV 2021