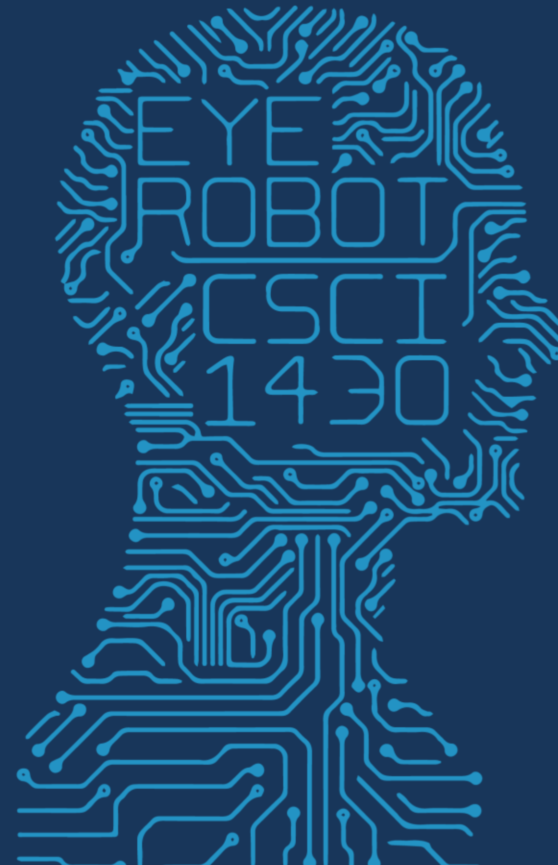
1950

Future Vision



15 April 2019

Computer Vision

Notre-Dame cathedral: Firefighters tackle blaze in Paris

8 minutes ago

Share

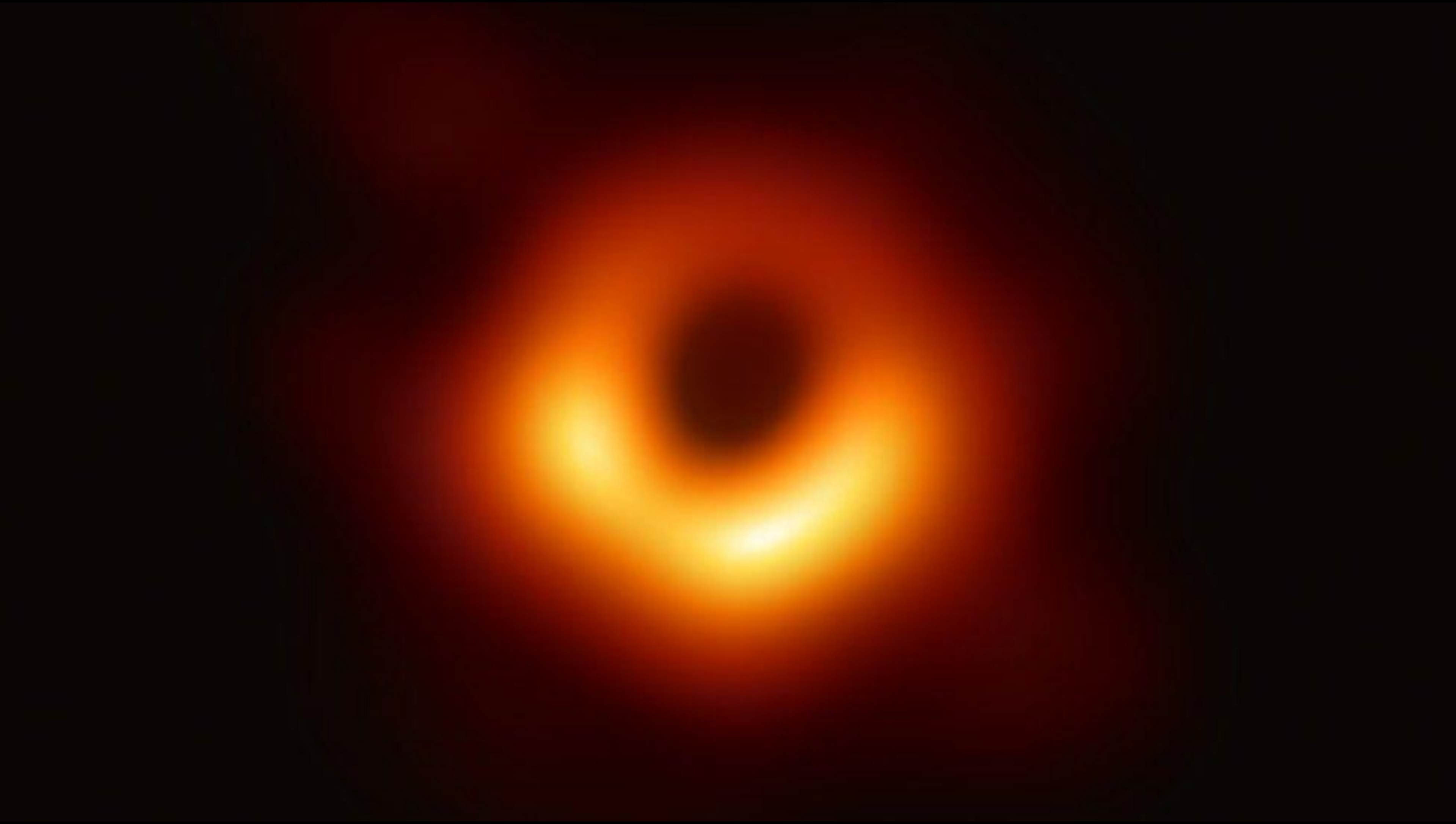
The fire department said a major operation was under way

Top Stories

Notre-Dame cathedral on fire in Paris
A major operation is under way to tackle the blaze at the 850-year-old Gothic building.
8 minutes ago

Sudan protesters demand end of 'deep state'
5 hours ago

VW ex-boss charged over emissions scandal
5 hours ago



Computational Imaging for VLBI Image Reconstruction

Katherine L. Bouman[1] Michael D. Johnson[2] Daniel Zoran[1] Vincent L. Fish[3]
Sherperd S. Doeleman[2,3] William T. Freeman[1,4]

[1]Massachusetts Institute of Technology, CSAIL [2]Harvard, Center for Astrophysics [3]MIT Haystack Observatory [4]Google

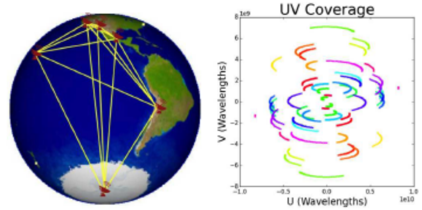
Abstract

Very long baseline interferometry (VLBI) is a technique for imaging celestial radio emissions by simultaneously observing a source from telescopes distributed across Earth. The challenges in reconstructing images from fine angular resolution VLBI data are immense. The data is extremely sparse and noisy, thus requiring statistical image models such as those designed in the computer vision community. In this paper we present a novel Bayesian approach for VLBI image reconstruction. While other methods often require careful tuning and parameter selection for different types of data, our method (CHIRP) produces good results under different settings such as low SNR or extended emission. The success of our method is demonstrated on realistic synthetic experiments as well as publicly available real data. We present this problem in a way that is accessible to members of the community, and provide a dataset website (vlbiimaging.csail.mit.edu) that facilitates controlled comparisons across algorithms.

1. Introduction

High resolution celestial imaging is essential for progress in astronomy and physics. For example, imaging the plasma surrounding a black hole's event horizon at high resolution could help answer many important questions; most notably, it may substantiate the existence of black holes [10] as well as verify and test the effects of general relativity [22]. Recently, there has been an international effort to create an Event Horizon Telescope (EHT) capable of imaging a black hole's event horizon for the first time [12, 13]. The angular resolution necessary for this observation is at least an order of magnitude smaller than has been previously used to image radio sources [24]. As measurements from the EHT become available, robust algorithms able to reconstruct images in this fine angular resolution regime will be necessary.

Although billions of dollars are spent on astronomical imaging systems to acquire the best images, current reconstruction techniques suffer from unsophisticated priors and



(a) Telescope Locations (b) Spatial Frequency Coverage

Figure 1. **Frequency Coverage:** (A) A sample of the telescope locations in the EHT. By observing a source over the course of a day, we obtain measurements corresponding to elliptical tracks in the source image's spatial frequency plane (B). These frequencies, (u, v), are the projected baseline lengths orthogonal to a telescope pair's light of sight. Points of the same color correspond to measurements from the same telescope pair.

a lack of inverse modeling [36], resulting in sub-optimal images. Image processing, restoration, sophisticated inference algorithms, and the study of non-standard cameras are all active areas of computer vision. The computer vision community's extensive work in these areas are invaluable to the success of these reconstruction methods and can help push the limits of celestial imaging. [16, 17, 27, 43].

Imaging distant celestial sources with high resolving power (i.e. fine angular resolution) requires single-dish telescopes with prohibitively large diameters due to the inverse relationship between angular resolution and telescope diameter [41]. For example, it is predicted that emission surrounding the black hole at the center of the Milky Way subtends $\approx 2.5 \times 10^{-10}$ radians [15]. Imaging this emission with a 10^{-10} radian resolution at a 1.3 mm wavelength would require a telescope with a 13000 km diameter. Although a single telescope this large is unrealizable, by simultaneously collecting data from an array of telescopes located around the Earth, it is possible to emulate samples from a single telescope with a diameter equal to the maximum distance between telescopes in the array. Using multiple telescopes in this manner is referred to as very long baseline interferometry (VLBI) [41]. Refer to Figure 1a.

People in the computer vision community integral to effort!

Paper describing technique @ CVPR 2016

Had 11 citations up to last week…

How does it work?

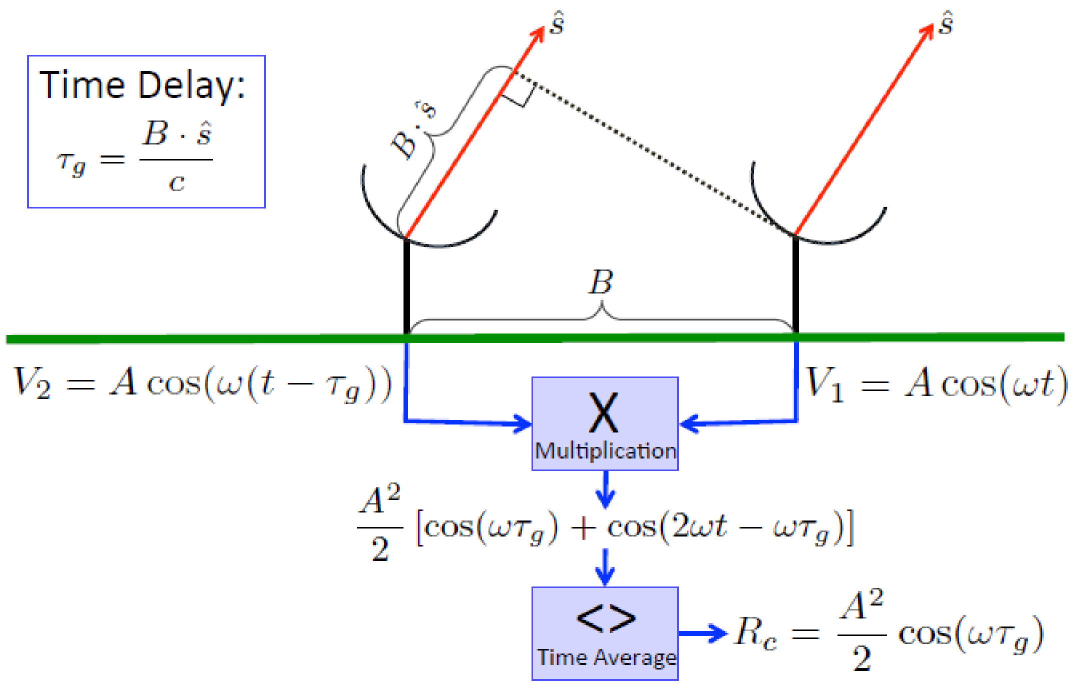


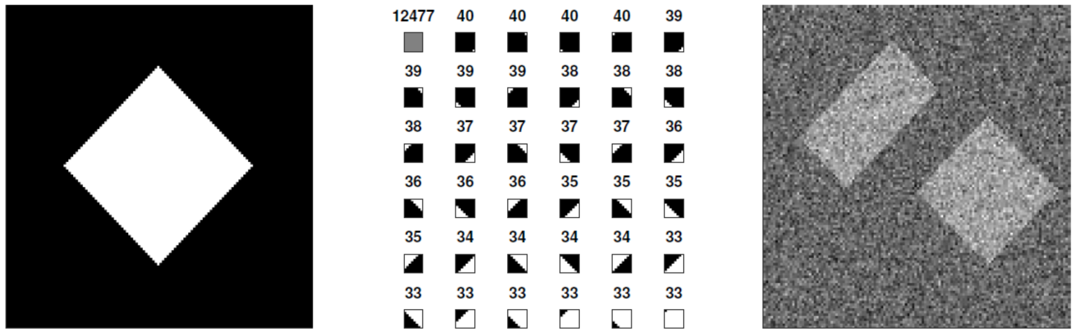
Figure 2. **Simplified Interferometry Diagram:** Light is emitted from a distant source and arrives at the telescopes as a plane wave in the direction $\hat{s}$. An additional distance of $B \cdot \hat{s}$ is necessary for the light to travel to the farther telescope, introducing a time delay between the received signals that varies depending on the source's location in the sky. The time-averaged correlation of these signals is a sinusoidal function related to the location of the source. This insight is generalized to extended emissions in the van Cittert-Zernike Thm. and used to relate the time-averaged correlation to a Fourier component of the emission image in the direction $\hat{s}$.

Problems:

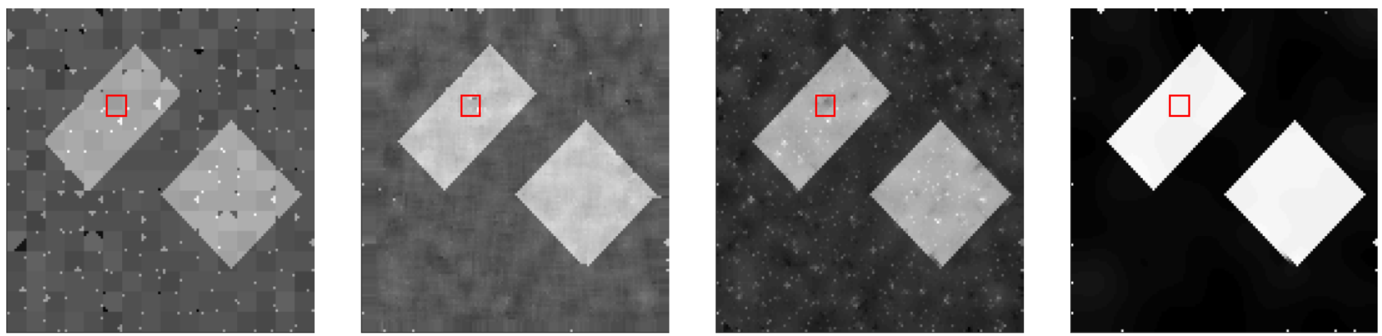
- Very noisy data!
- Significant phase shifts

Computer vision to the rescue!

- Image modeling in Fourier domain -> represent phase
- Data-driven priors using Gaussian mixture models on patches of signal -> represent expected statistics!
- Expected Patch Log Likelihood – Zoran and Weiss ICCV 2011 -> generic prior method for image reconstruction
- Simulated data to build prior



(a) Training Image (b) Prior Learned (c) Noisy Image

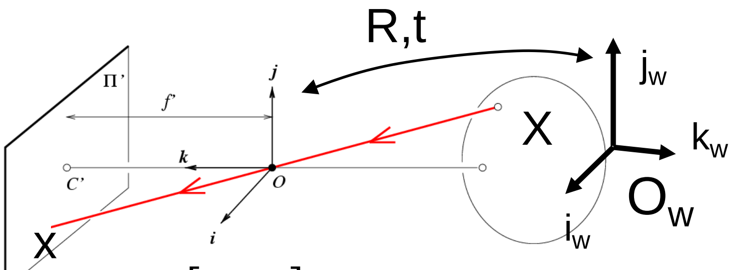


(d) Non Overlapping (e) Center Pixel (f) Averaged Overlapping (g) Our Method

Figure 2: The intuition behind our method. 2a A training image. 2b The prior learned from the image, only the 36 most frequent patches are shown with their corresponding count above the patch - flat patches are the most likely ones, followed by edges with 1 pixel etc. 2c A noisy image we wish to restore. 2d Restoring using non-overlapping patches - note the severe artifacts at patch borders and around the image. 2e Taking the center pixels from each patch. 2f Better results are obtained by restoring all overlapping patches, averaging the results - artifacts are still visible, and a lot of the patches in the resulting image are unlikely under the prior. 2g Result using the proposed method - note that there are very few artifacts, and most patches are very likely under our prior.

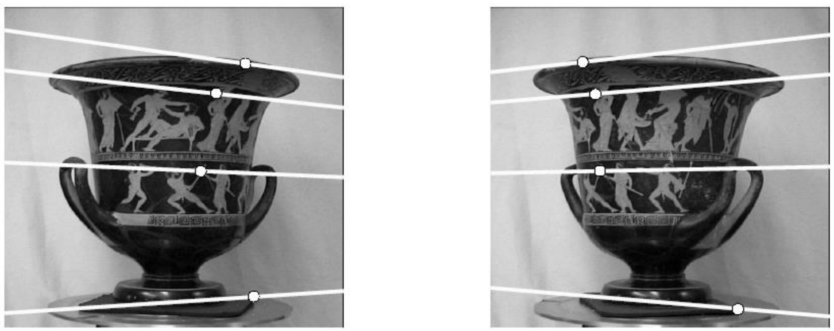
[Zoran and Weiss, 2011]

Multiple view geometry



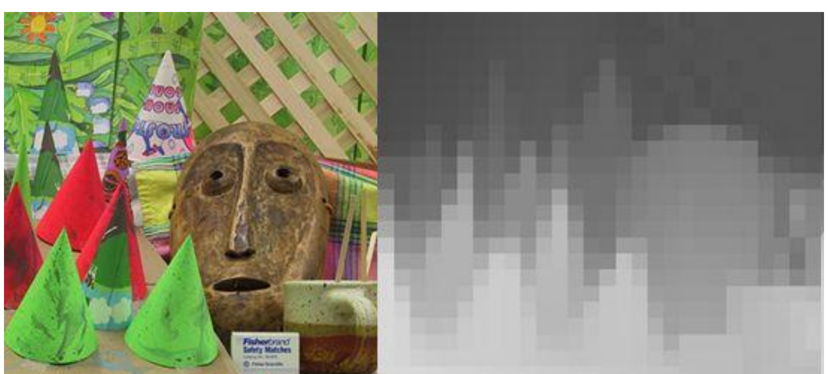
$$\mathbf{x} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}]\,\mathbf{X}$$

Camera calibration



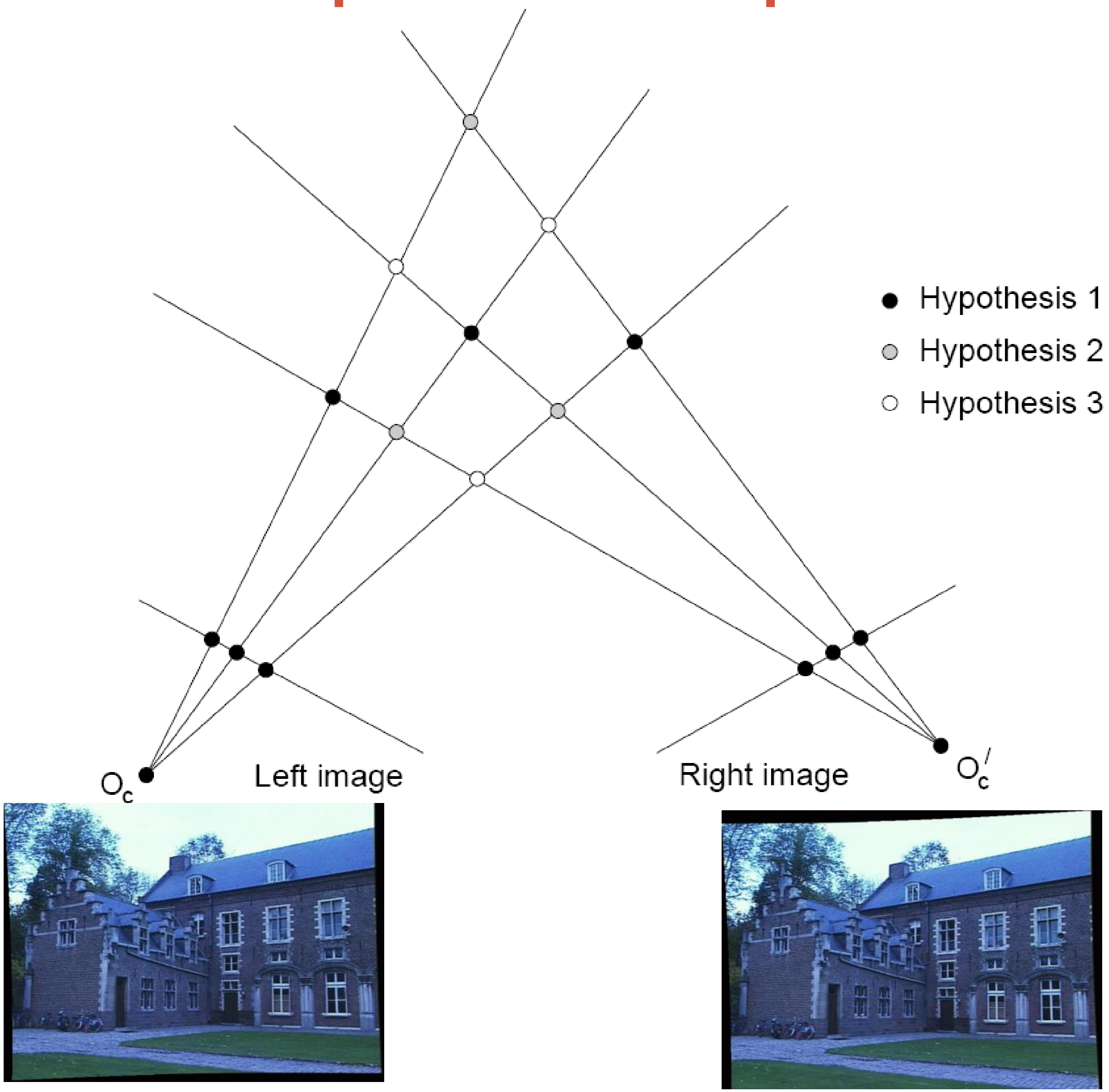
Hartley and Zisserman

Epipolar geometry



Dense depth
map estimation

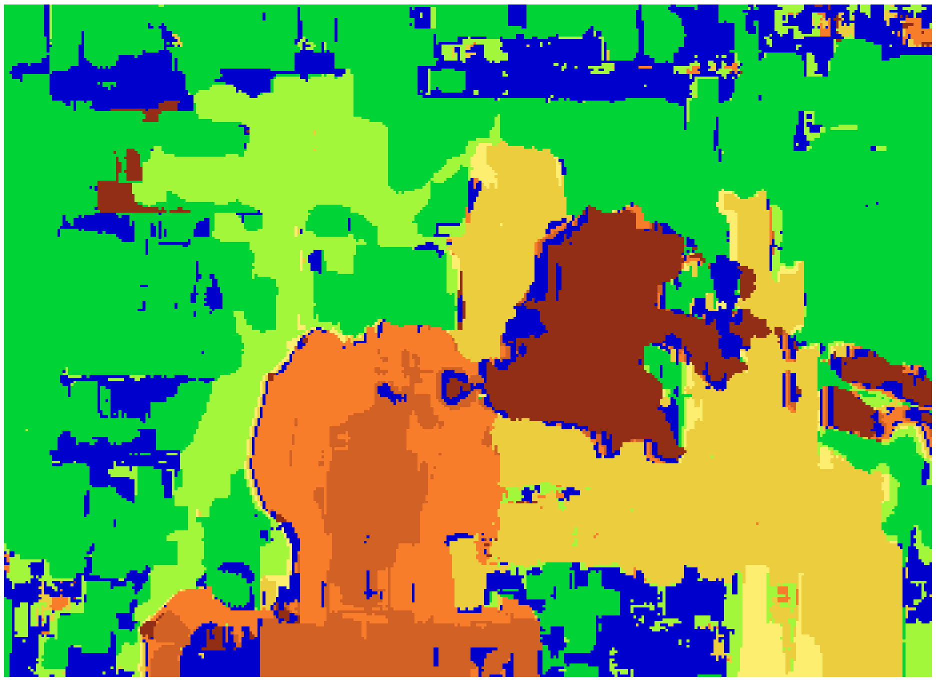
Correspondence problem



Multiple match hypotheses satisfy epipolar constraint, but which is correct?

Figure from Gee & Cipolla 1999

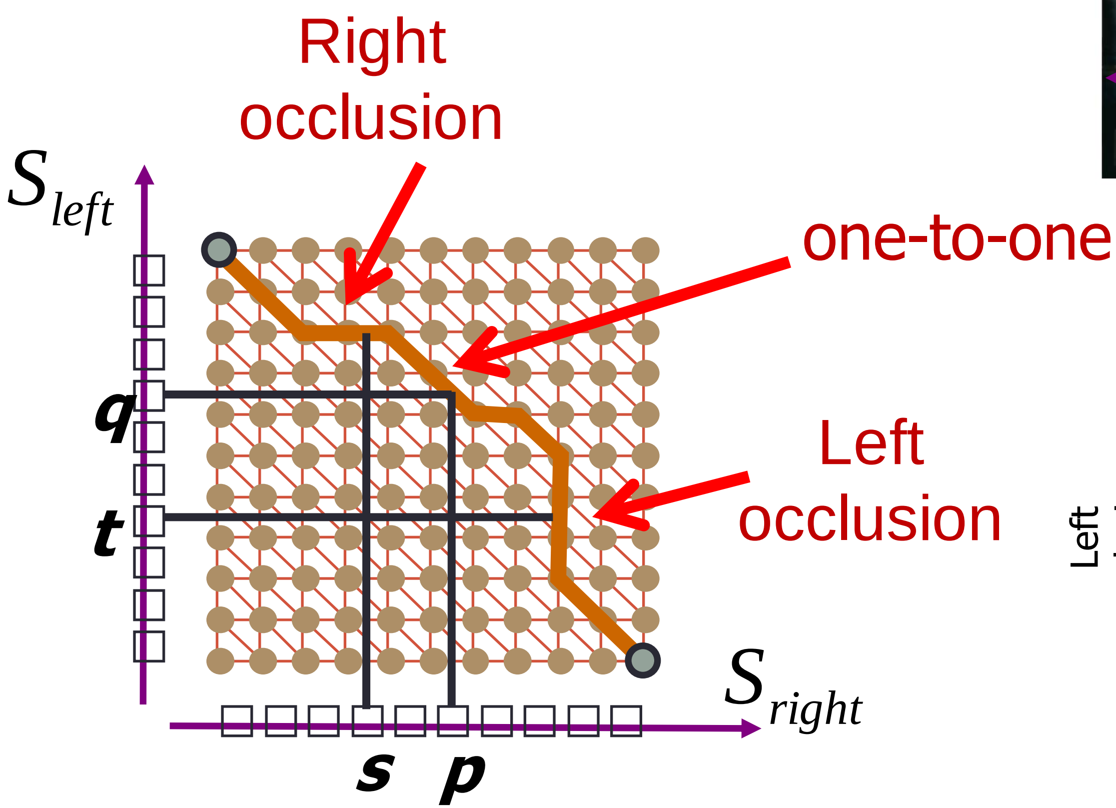
Results with window search

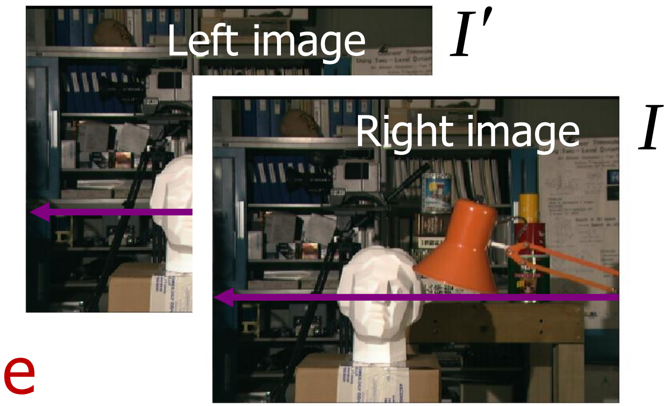


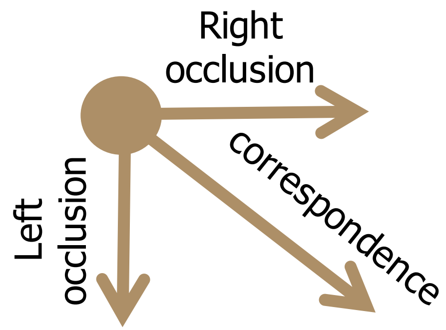
Window-based matching
(best window size)

‘Ground truth’

“Shortest paths” for scan-line stereo







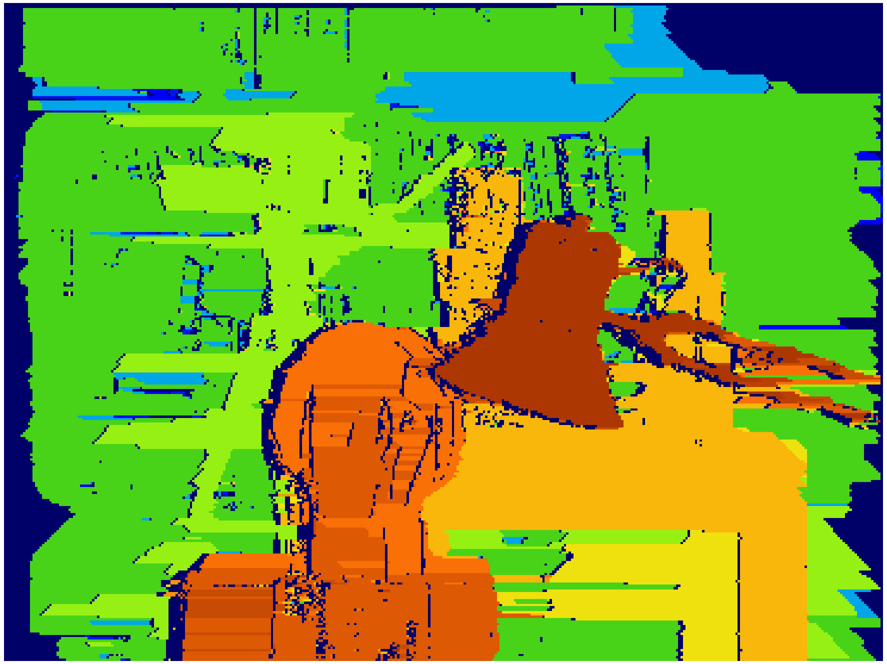
Can be implemented with dynamic programming

Ohta & Kanade ’85, Cox et al. ’96, Intille & Bobick, ‘01

Slide credit: Y. Boykov

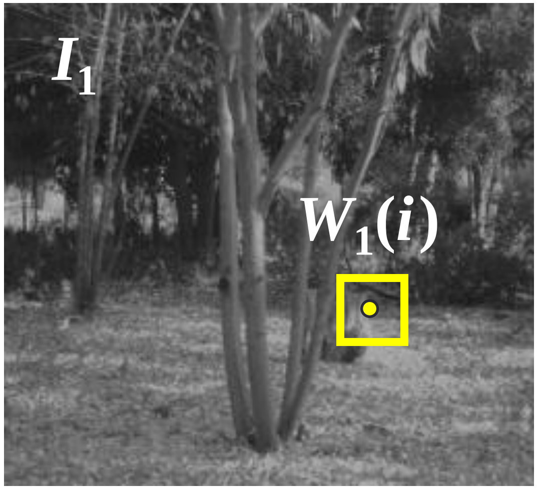
Coherent stereo on 2D grid

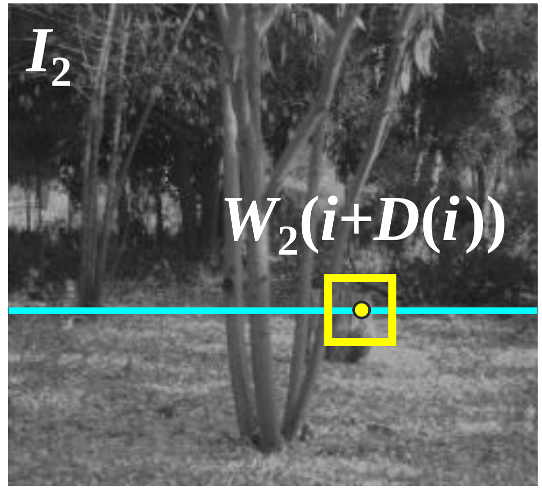
- Scanline stereo generates streaking artifacts

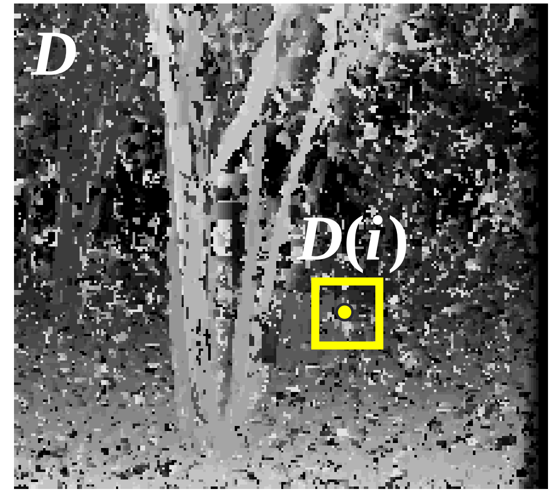


- Can’t use dynamic programming to find spatially coherent disparities/ correspondences on a 2D grid

Stereo matching as energy minimization







$$\boxed{E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)}$$

$$\boxed{E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2}$$

$$\boxed{E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)}$$

Energy functions of this form can be minimized using *graph cuts.*

Y. Boykov, O. Veksler, and R. Zabih, [Fast Approximate Energy Minimization via Graph Cuts](), PAMI 2001

Source: Steve Seitz

Better results…

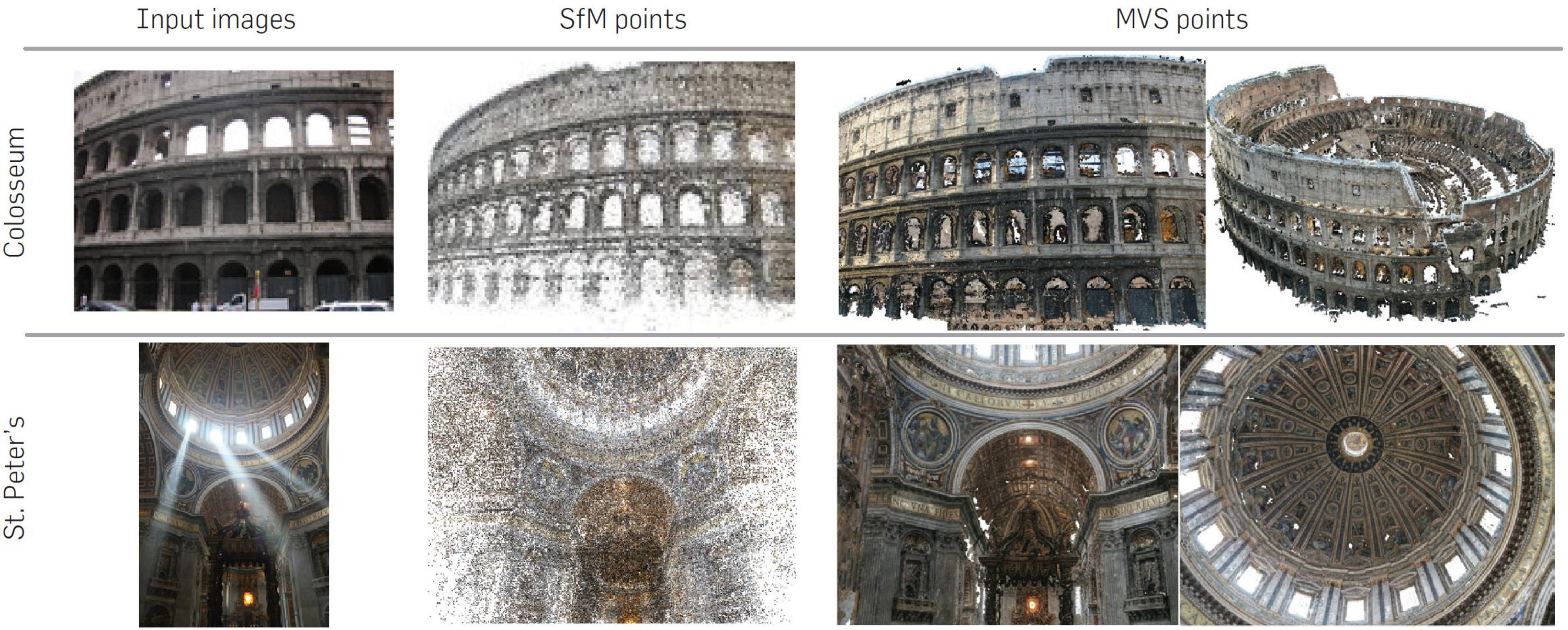
Graph cut method

Ground truth

Boykov et al., [Fast Approximate Energy Minimization via Graph Cuts](), International Conference on Computer Vision, September 1999.

For the latest and greatest: http://www.middlebury.edu/stereo/

SIFT + Fundamental Matrix + RANSAC + dense correspondence



Building Rome in a Day

By Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, Richard Szeliski
Communications of the ACM, Vol. 54 No. 10, Pages 105-112

SIFT + Fundamental Matrix + RANSAC + dense correspondence

The Visual Turing Test for Scene Reconstruction
Supplementary Video

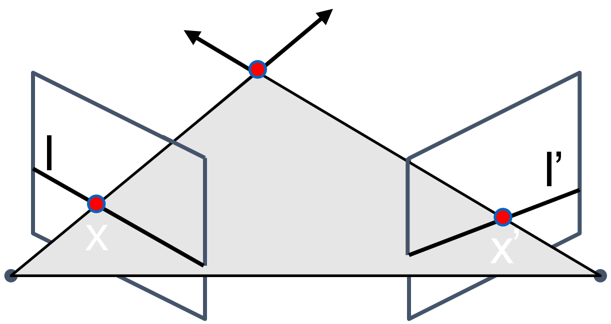
Qi Shan[+] Riley Adams[+] Brian Curless[+]
Yasutaka Furukawa[*] Steve Seitz[+*]

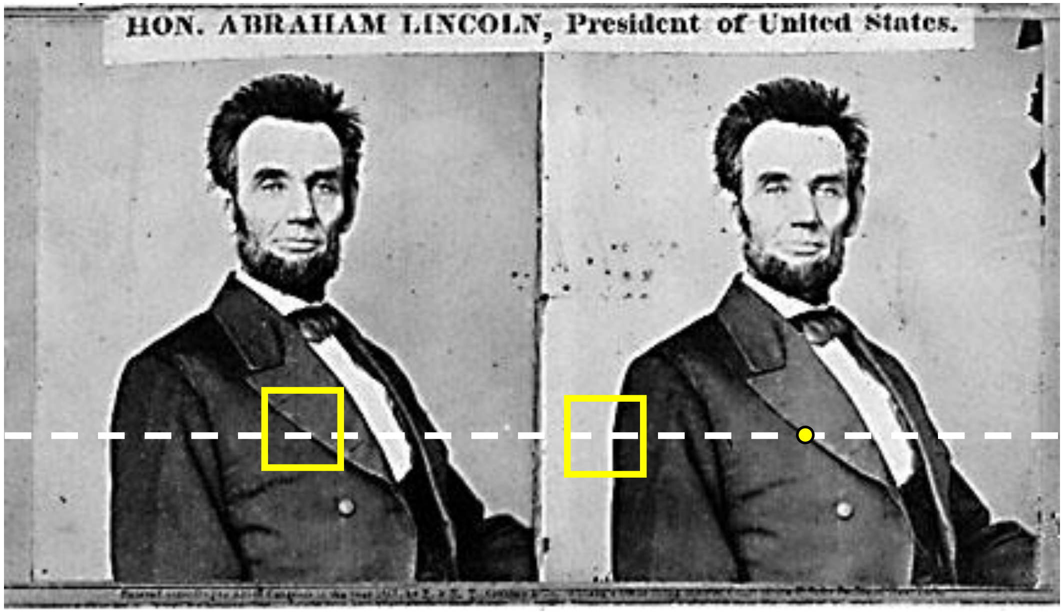
[+]University of Washington [*]Google

3DV 2013

Stereo correspondence

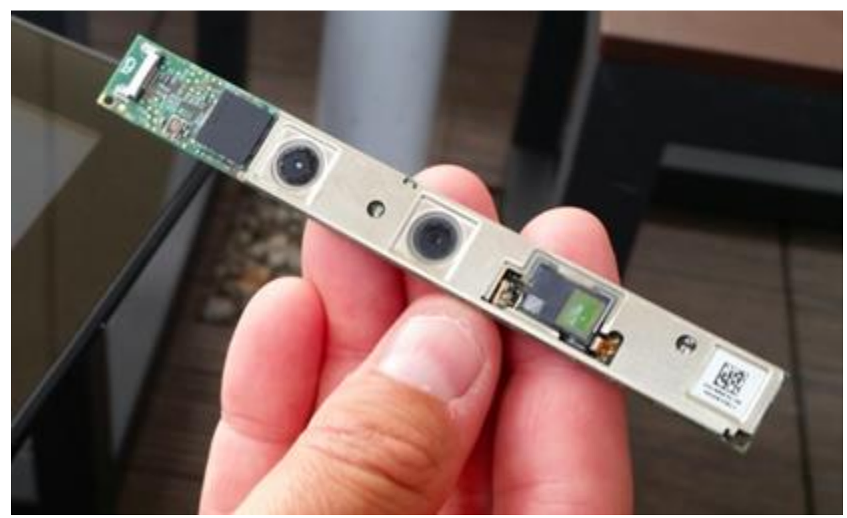
- Let *x* be a point in left image, *x'* in right image
- Epipolar relation
  - *x* maps to epipolar line *l'*
  - *x'* maps to epipolar line *l*





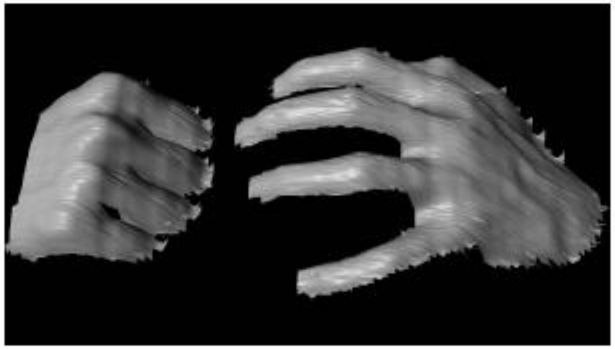
How does a depth camera work?

Microsoft Kinect v1

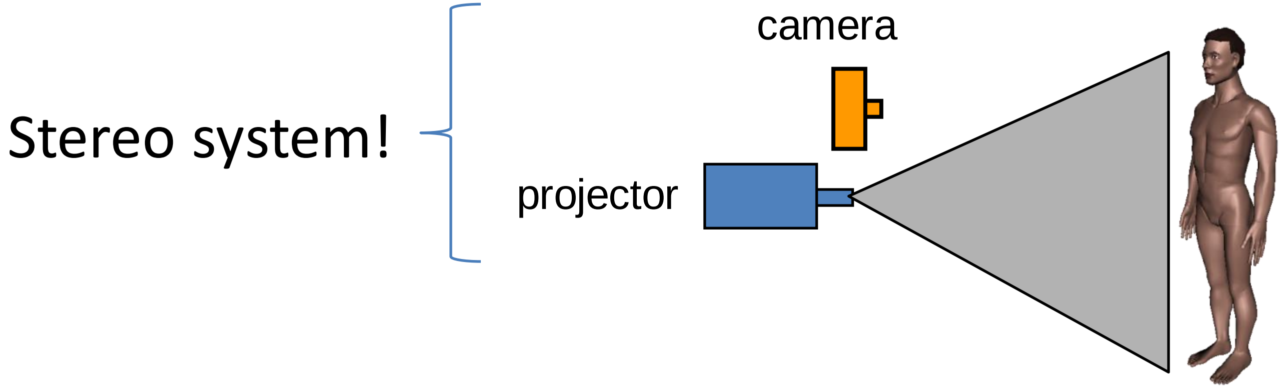


Intel laptop depth camera

Active stereo with structured light



- Project "structured" light patterns onto the object
  - Simplifies the correspondence problem
  - Allows us to use only one camera



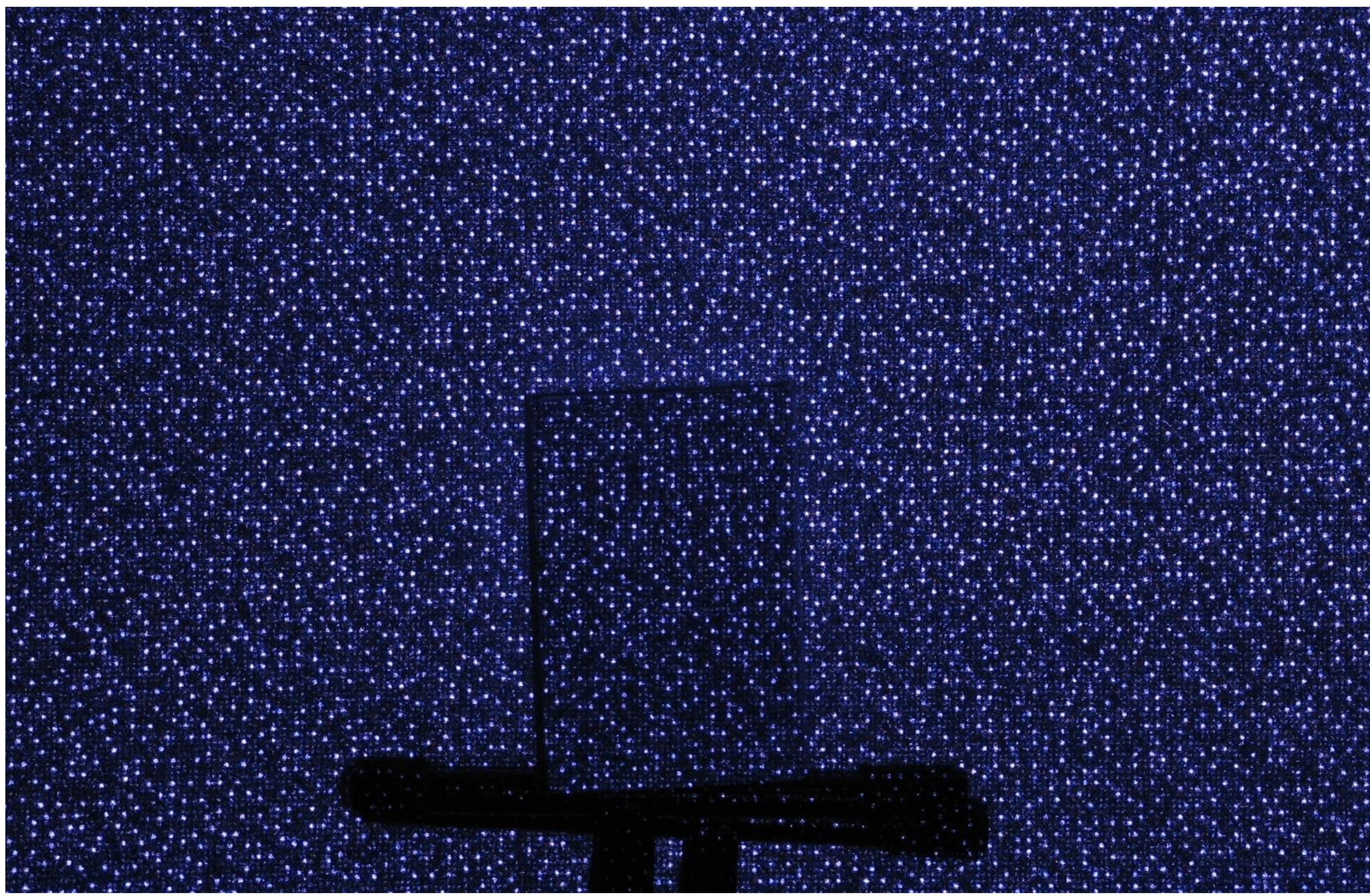
L. Zhang, B. Curless, and S. M. Seitz. [Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.]() *3DPVT* 2002

Kinect: Structured infrared light

http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/

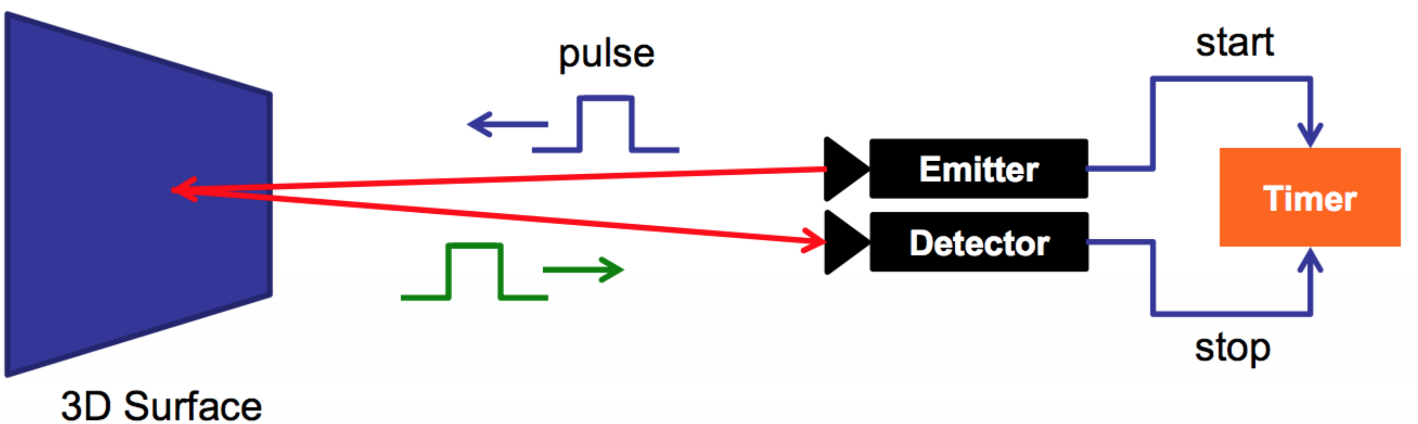
How does a depth camera work?

Stereo in infrared.



Time of Flight (Kinect V2)

- Depth cameras in HoloLens use *time of flight*
  - “SONAR for light”
  - Emit light of a known wavelength, and time how long it takes for it to come back



With either technique…

…I gain depth maps over time.

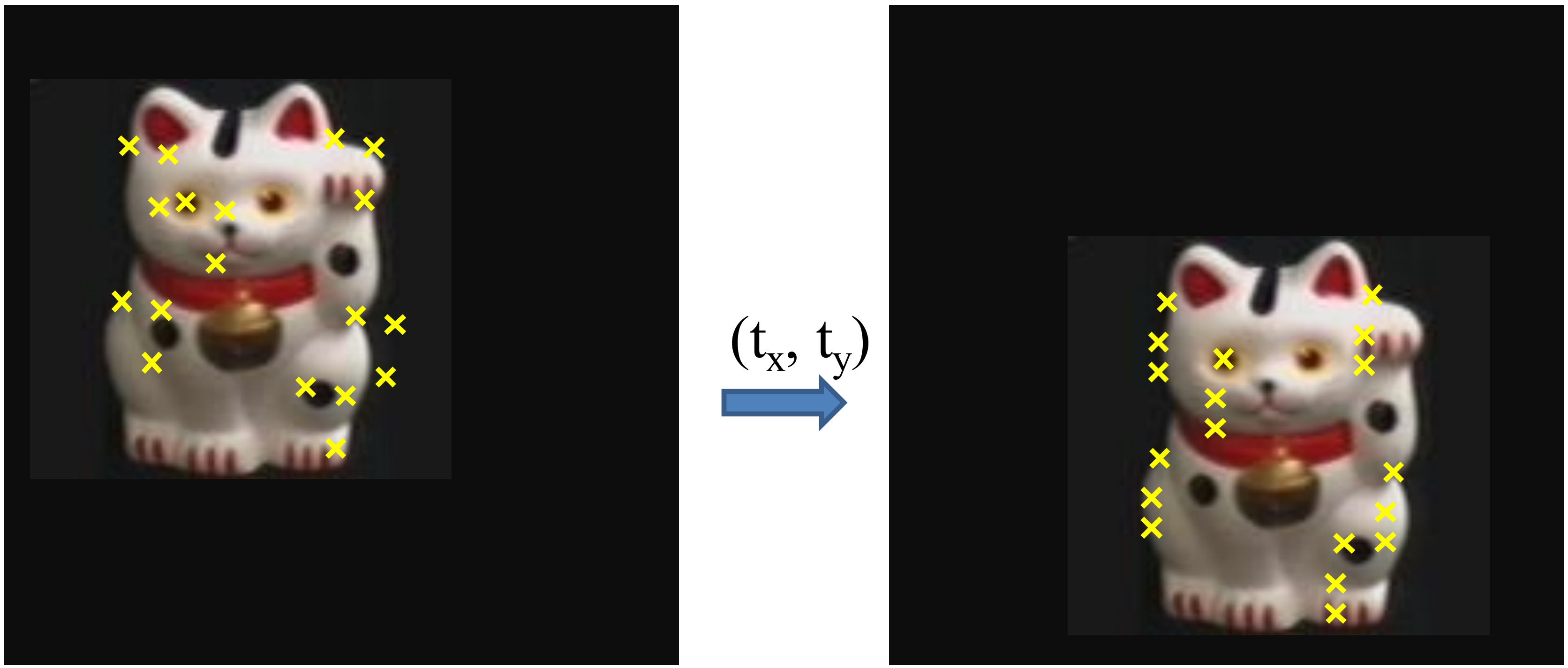
Optex Depth Camera Based on Canesta Solution

Once I have my depth map,
what can I do with it?

Measure.
Combine! (Reorganize?)

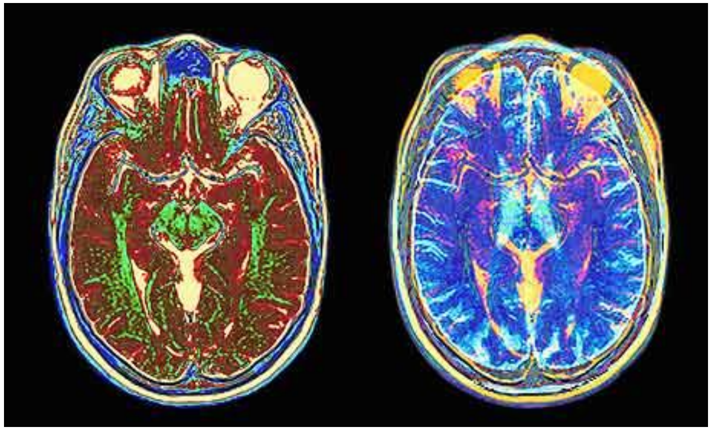
What if we want to align...
but we have no matched pairs?

- Hough transform and RANSAC not applicable

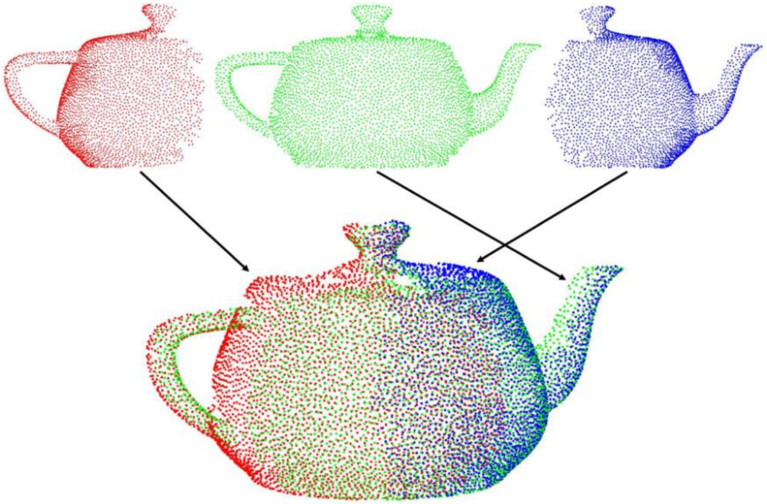


Problem: no initial guesses for correspondence

Applications



Medical imaging: match brain scans or contours



Vision/Robotics: match point clouds

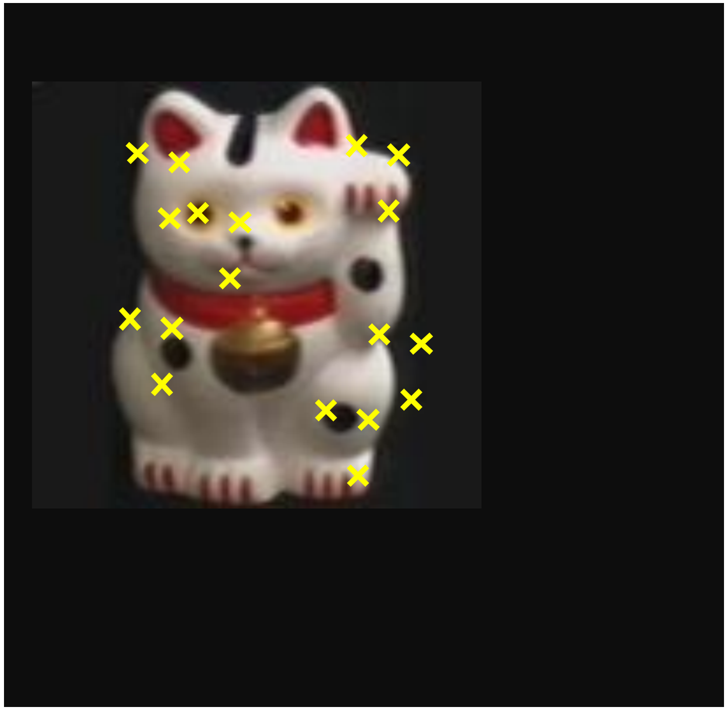
Iterative Closest Points (ICP) Algorithm

Goal:

Estimate transform between two dense point sets S_1 and S_2

1. **Initialize** transformation
   - Compute difference in mean positions, subtract
   - Compute difference in scales, normalize
2. **Assign** each point in S_1 to its nearest neighbor in S_2
3. **Estimate** transformation parameters T
   - Least squares or robust least squares, e.g., rigid transform
4. **Transform** the points in S_1 using estimated parameters T
5. **Repeat** steps 2-4 until change is very small (convergence)

Example: solving for translation



(t_x, t_y)

Problem: no initial guesses for correspondence

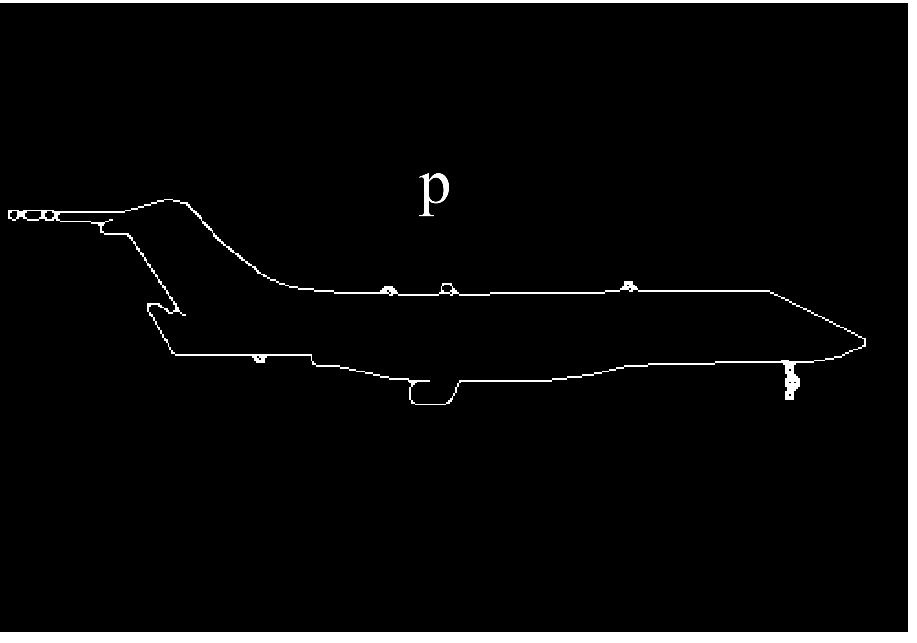
ICP solution

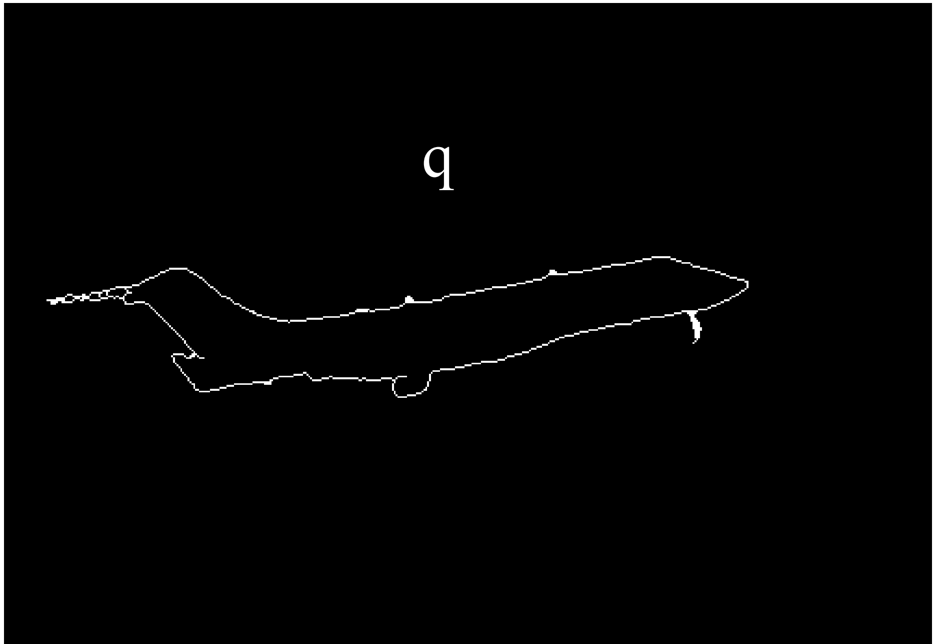
1. Initialize t by mean point translation
2. Find nearest neighbors for each point
3. Compute transform using matches
4. Move points using transform
5. Repeat steps 2-4 until convergence

$$\begin{bmatrix} x_i^B \\ y_i^B \end{bmatrix} = \begin{bmatrix} x_i^A \\ y_i^A \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

Example: aligning boundaries

1. Extract edge pixels $p_1..p_n$ and $q_1..q_m$
2. Compute initial transformation (e.g., compute translation and scaling by center of mass, variance within each image)
3. Get nearest neighbors: for each point p_i find corresponding $\text{match}(\text{i}) = \underset{j}{\text{argmin}}\, dist(pi, qj)$
4. Compute transformation ***T*** based on matches
5. Transform points ***p*** according to ***T***
6. Repeat 3-5 until convergence





ICP demonstration

Time = iterations of ICP

Bouaziz et al.

Sparse ICP

Sofien Bouaziz   *Andrea Tagliasacchi*   *Mark Pauly*

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

BundleFusion: Real-time Globally Consistent 3D Reconstruction using Online Surface Re-integration

Angela Dai[1] *Matthias Nießner*[1]
Michael Zollhöfer[2] *Shahram Izadi*[3]
Christian Theobalt[2]

[1]Stanford University
[2]Max Planck Institute for Informatics
[3]Microsoft Research

(*contains audio*)

ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes

Angela Dai   Angel X. Chang   Manolis Savva   Maciej Halber
Thomas Funkhouser   Matthias Nießner

Stanford University
Princeton University
Technical University of Munich

CVPR 2017 (Spotlight)