Two-View Stereo

Many slides adapted from Steve Seitz
Problem formulation

• **Given**: stereo pair (assumed calibrated)
• ** Wanted**: dense depth map
Outline

• Motivation and history
• Basic two-view stereo setup
• Local stereo matching algorithm
• Beyond local stereo matching
• Active stereo with structured light
Stereo vision and perception of depth

• What cues tell us about scene depth?
How Two Photographers Unknowingly Shot the Same Millisecond in Time

https://petapixel.com/2018/03/07/two-photographers-unknowingly-shot-millisecond-time/
How Two Photographers Unknowingly Shot the Same Millisecond in Time

https://petapixel.com/2018/03/07/two-photographers-unknowingly-shot-millisecond-time/
How Two Photographers Unknowingly Shot the Same Millisecond in Time

https://petapixel.com/2018/03/07/two-photographers-unknowingly-shot-millisecond-time/
History: Stereograms

- Humans can fuse pairs of images to get a sensation of depth

Stereograms: Invented by Sir Charles Wheatstone, 1838

https://en.wikipedia.org/wiki/Stereoscopy
History: Random dot stereograms

- Invented by Bela Julesz in the mid-20th century
- Demonstration that stereo perception can happen without any monocular cues

History: Random dot stereograms

- Invented by Bela Julesz in the mid-20th century
- Demonstration that stereo perception can happen without any monocular cues

Outline

• Motivation and history
• Basic two-view stereo setup
Problem formulation

- **Given**: stereo pair (assumed calibrated)
- **Wanted**: dense depth map
Basic stereo matching algorithm

- For each pixel in the first image
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match
  - Triangulate the matches to get depth information

- Simplest case: epipolar lines are corresponding scanlines
  - When does this happen?
Parallel images

- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at the same height
- Focal lengths are the same
- Then epipolar lines fall along horizontal scan lines of the images
Essential matrix for parallel images

Epipolar constraint:

\[ x'^T E x = 0, \quad E = [t_x]R \]

\[ R = I \quad t = (t, 0, 0) \]

\[ E = [t_x]R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t \\ 0 & t & 0 \end{bmatrix} \]

\[
\begin{pmatrix} u' & v' & 1 \end{pmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t \\ 0 & t & 0 \end{bmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0
\]

\[
\begin{pmatrix} u' & v' & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -t \\ tv \end{pmatrix} = 0
\]

\[-tv + tv' = 0 \quad v = v'\]

The y-coordinates of corresponding points are the same!
Stereo image rectification

- If the image planes are not parallel, we can find homographies to project each view onto a common plane parallel to the baseline.

If the image planes are not parallel, we can find homographies to project each view onto a common plane parallel to the baseline.
Stereo image rectification

- Before rectification:
Stereo image rectification

• After rectification:
Another rectification example

Unrectified

Rectified
Basic stereo matching algorithm

- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel $x$ in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match $x'$
  - Triangulate the matches to get depth information
Disparity is inversely proportional to depth!
Depth from disparity

\[
\frac{x}{f} = \frac{B_1}{z} \quad \frac{x'}{f} = \frac{B_2}{z}
\]

\[
\frac{x - x'}{f} = \frac{B_1 - B_2}{z}
\]

\[
x - x' = \frac{fB}{z}
\]

\[
z = \frac{fB}{x - x'}
\]
Effect of baseline on stereo results

- Larger baseline
  + Smaller triangulation error
  - Matching is more difficult

- Smaller baseline
  - Higher triangulation error
  + Matching is easier
Basic stereo matching algorithm

- If necessary, rectify the two stereo images to transform epipolar lines into scanlines.
- For each pixel $x$ in the first image
  - Find corresponding epipolar scanline in the right image.
  - Examine all pixels on the scanline and pick the best match $x'$.
  - Compute disparity $x - x'$ and set $\text{depth}(x) = \frac{Bf}{x - x'}$. 
Outline

• Motivation and history
• Basic two-view stereo setup
• Local stereo matching algorithm
Correspondence search

- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation
Correspondence search

Left

Right

scanline

SSD
Correspondence search

Left

scanline

Right

Norm. corr
Effect of window size on correspondence search

- Smaller window:
  - More detail
  - More noise

- Larger window:
  - Smoother disparity maps
  - Less detail
Where will basic window search fail?

- Textureless surfaces
- Occlusions, repetition
- Non-Lambertian surfaces, specularities
Example: Textured neighborhood

Window size: 1 pixel

Source: D. Hoiem
Example: Textured neighborhood

Window size: 7 pixels

Source: D. Hoiem
Example: Smooth neighborhood

Window size: 1 pixel

Source: D. Hoiem
Example: Smooth neighborhood

Window size: 7 pixels

Source: D. Hoiem
Example: Specular highlight

Window size: 1 pixel

Source: D. Hoiem
Example: Specular highlight

Window size: 7 pixels

Source: D. Hoiem
Outline

• Motivation and history
• Basic two-view stereo setup
• Local stereo matching algorithm
• Beyond local stereo matching
Stereo as optimization with non-local constraints

Non-local constraint: Uniqueness

- Each point in one image should match at most one point in the other image.
- Does uniqueness always hold in real life?

Source: J. Johnson and D. Fouhey
Non-local constraint: Uniqueness

- Each point in one image should match at most one point in the other image
- Does uniqueness always hold in real life?

Source: J. Johnson and D. Fouhey
Non-local constraint: Ordering

- Corresponding points should appear in the same order
- Is ordering always preserved in real life?

Source: J. Johnson and D. Fouhey
Non-local constraint: Ordering

- Corresponding points should appear in the same order
- Is ordering always preserved in real life?

Source: J. Johnson and D. Fouhey
Non-local constraint: Smoothness

• We expect disparity values to change slowly (for the most part)
Scanline stereo by dynamic programming

- Match pixels along the entire scanline while preserving uniqueness and ordering
- Different scanlines are still optimized independently

Scanline stereo by dynamic programming


Source: Y. Boykov
Scanline stereo by dynamic programming

- Generates streaking artifacts!
Stereo matching as global optimization

Energy functions of this form can be minimized using \textit{graph cuts}

\begin{equation}
E(D) = \sum_i \left( W_1(i) - W_2(i + D(i)) \right)^2 + \lambda \sum_{(i,j) \in N} \rho(D(i) - D(j))
\end{equation}

- Data term
- Neighborhood smoothness term

Stereo matching with deep networks

L. Lipson et al. **RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching.** arXiv 2021
Self-supervised depth estimation

Stereo datasets

- Middlebury stereo datasets
- KITTI
- Synthetic data
Outline

- Motivation and history
- Basic two-view stereo setup
- Local stereo matching algorithm
- Stereo with non-local optimization
- **Active stereo with structured light**
Active stereo with structured light

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

Active stereo with structured light

Active stereo with structured light

Kinect: Structured infrared light

Example: Book vs. No Book

Source (via D. Hoiem)
Example: Book vs. No Book

Source (via D. Hoiem)
Apple TrueDepth

https://www.cnet.com/news/apple-face-id-truedepth-how-it-works/
Laser scanning

Optical triangulation
- Project a single stripe of laser light
- Scan it across the surface of the object
- This is a very precise version of structured light scanning

Digital Michelangelo Project
Levoy et al.
http://graphics.stanford.edu/projects/mich/

Source: S. Seitz
Laser scanned models

*The Digital Michelangelo Project, Levoy et al.*

Source: S. Seitz
Laser scanned models

*The Digital Michelangelo Project, Levoy et al.*

Source: S. Seitz
Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz
Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz
Laser scanned models

1.0 mm resolution (56 million triangles)

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz
Aligning range images

- A single range scan is not sufficient to capture a complex surface
- Need techniques to register multiple range images

Aligning range images

- A single range scan is not sufficient to capture a complex surface
- Need techniques to register multiple range images

… which brings us to multi-view stereo