Two-View Stereo
What do you see in this image?

Autostereograms: www.magiceye.com

Slide from L. Lazebnik.
Stereo

• What cues tell us about scene depth?
Stereograms

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

Slide from L. Lazebnik.
Stereograms

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

Autostereograms: [www.magiceye.com](http://www.magiceye.com)

Slide from L. Lazebnik.
Stereograms

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

Autostereograms: www.magiceye.com

Slide from L. Lazebnik.
Problem formulation

- Given a calibrated binocular stereo pair, fuse it to produce a depth image

image 1

image 2

Dense depth map

Slide from L. Lazebnik.
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?

Slide from L. Lazebnik.
Simplest Case: Parallel images

- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same

Slide from L. Lazebnik.
Simplest Case: Parallel images

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

Slide from L. Lazebnik.
Essential matrix for parallel images

Epipolar constraint:

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

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

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

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

\[
\begin{pmatrix} u' & v' & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -T \\ Tv \end{pmatrix} = 0 \quad Tv' = Tv
\]

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

Slide from L. Lazebnik.
Stereo image rectification

- Reproject image planes onto a common plane parallel to the line between optical centers

Rectification example

Slide from L. Lazebnik.
Another rectification example

Unrectified

Rectified

Slide from L. Lazebnik.
Basic stereo matching algorithm

- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- 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
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

Slide from L. Lazebnik.
Correspondence search

Left

Right

scanline

SSD

Slide from L. Lazebnik.
Correspondence search

Left

Right

scanline

Norm. corr

Slide from L. Lazebnik.
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

Slide from L. Lazebnik.
From Wikipedia: Gemma Frisius's 1533 diagram introducing the idea of triangulation into the science of surveying. Having established a baseline, e.g. the cities of Brussels and Antwerp, the location of other cities, e.g. Middelburg, Ghent etc., can be found by taking a compass direction from each end of the baseline, and plotting where the two directions cross. This was only a theoretical presentation of the concept — due to topographical restrictions, it is impossible to see Middelburg from either Brussels or Antwerp. Nevertheless, the figure soon became well known all across Europe.
Disparity is inversely proportional to depth!
Depth from disparity

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

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

\[
\text{disparity} = x - x' = \frac{B \cdot f}{z}
\]

Slide from L. Lazebnik.
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 $\text{depth}(x) = B*f/(x-x')$.
Failures of correspondence search

Textureless surfaces

Occlusions, repetition

Non-Lambertian surfaces, specularities

Slide from L. Lazebnik.
Effect of window size

- Smaller window
  + More detail
  - More noise

- Larger window
  + Smoother disparity maps
  - Less detail

Slide from L. Lazebnik.
Results with window search

Slide from L. Lazebnik.
Better methods exist...

Graph cuts

Ground truth


For the latest and greatest: [http://www.middlebury.edu/stereo/](http://www.middlebury.edu/stereo/)

Slide from L. Lazebnik.
How can we improve window-based matching?

- The similarity constraint is **local** (each reference window is matched independently)
- Need to enforce **non-local** correspondence constraints
Non-local constraints

- **Uniqueness**
  - For any point in one image, there should be at most one matching point in the other image
Non-local constraints

• Uniqueness
  • For any point in one image, there should be at most one matching point in the other image

• Ordering
  • Corresponding points should be in the same order in both views
Non-local constraints

- **Uniqueness**
  - For any point in one image, there should be at most one matching point in the other image

- **Ordering**
  - Corresponding points should be in the same order in both views

Slide from L. Lazebnik.
Non-local constraints

- **Uniqueness**
  - For any point in one image, there should be at most one matching point in the other image

- **Ordering**
  - Corresponding points should be in the same order in both views

- **Smoothness**
  - We expect disparity values to change slowly (for the most part)
Scanline stereo

- Try to coherently match pixels on the entire scanline
- Different scanlines are still optimized independently
Stereo matching as global optimization

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


Slide from L. Lazebnik.
Stereo matching as a prediction problem


Slide from L. Lazebnik.
Review: Basic stereo matching algorithm

- For each pixel \( x \) in the reference image
  - Find corresponding epipolar scanline in the other image
  - Examine all pixels on the scanline and pick the best match \( x' \)
  - Compute disparity \( x - x' \) and set depth(\( x \)) = \( B \times f \times (x - x') \)
Depth from Triangulation

Passive Stereopsis

Active Stereopsis

Active sensing simplifies the problem of estimating point correspondences
Kinect: Structured infrared light


Slide from L. Lazebnik.
Apple TrueDepth

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

Slide from L. Lazebnik.
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
**Stereo error(distance)**

Error in distance estimate increases quadratically with the distance

\[
\begin{align*}
    Z &= \text{distance} \\
    d &= \text{disparity} \\
    Z &= \frac{C}{d} \\
    \delta Z &= \frac{-Z^2}{C} \delta d \\
    |\delta Z| &= \frac{Z^2}{C} |\delta d| \\
    \text{error} \propto \text{distance}^2
\end{align*}
\]

![Graph showing the quadratic relationship between distance and quantization error.](image_url)
Multi-view stereo
Multi-view stereo: Basic idea

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

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

Source: Y. Furukawa
Towards Internet-Scale Multi-View Stereo

YouTube video, CMVS software

Applications

Source: N. Snavely
Latest and greatest: NeRF

Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV 2020.
Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng.