Multi-view Stereo

Ziling Cheng
Sarah Al Taleb
Tad Wu

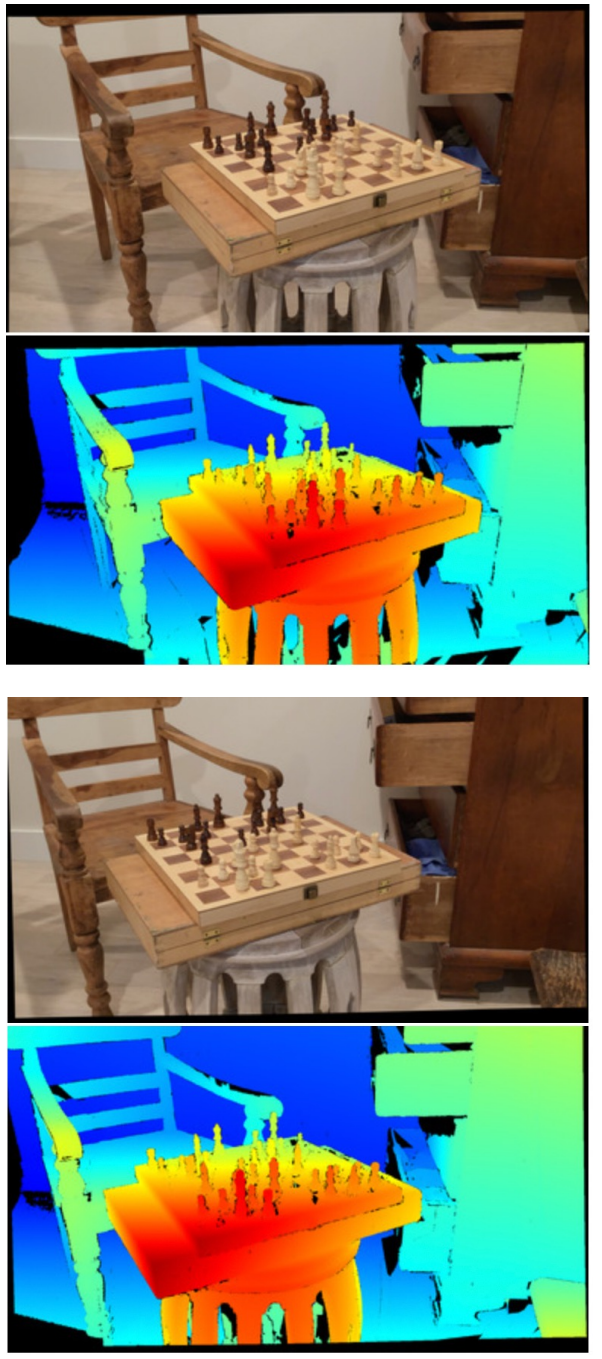
Multi-view Stereo

Goal: recover the disparity/depth map of the scene from two views

- Step 1: Camera calibration
  - Get the calibration matrix K
- Step 2: Estimation of the fundamental matrix F (sparse stereo matching)
  - 8-point algorithm with RANSAC
  - Extract the relative rotation and translation between two cameras
- Step 3: Dense stereo matching (window-based 1D search)
  - Given a point in one image, the corresponding point in the other image must lie on the **epipolar line** -> 1D search through the epipolar line
- Step 4: Depth map reconstruction
  - For each point in the reference image, compute the disparity between two images and the the depth is inversely proportional to the disparity

Middlebury Stereo Dataset

- A set of rectified two-view image pairs
- For each pair, we have
  - Left and right images
  - Ground truth disparity maps for both images
  - Calibration information including
    - **Intrinsic parameters**: the calibration matrix *K*
    - **Extrinsic parameters***:* the baseline *d*
    - **Image size:** width, height
    - A **conservative bound** on the number of **disparity levels**

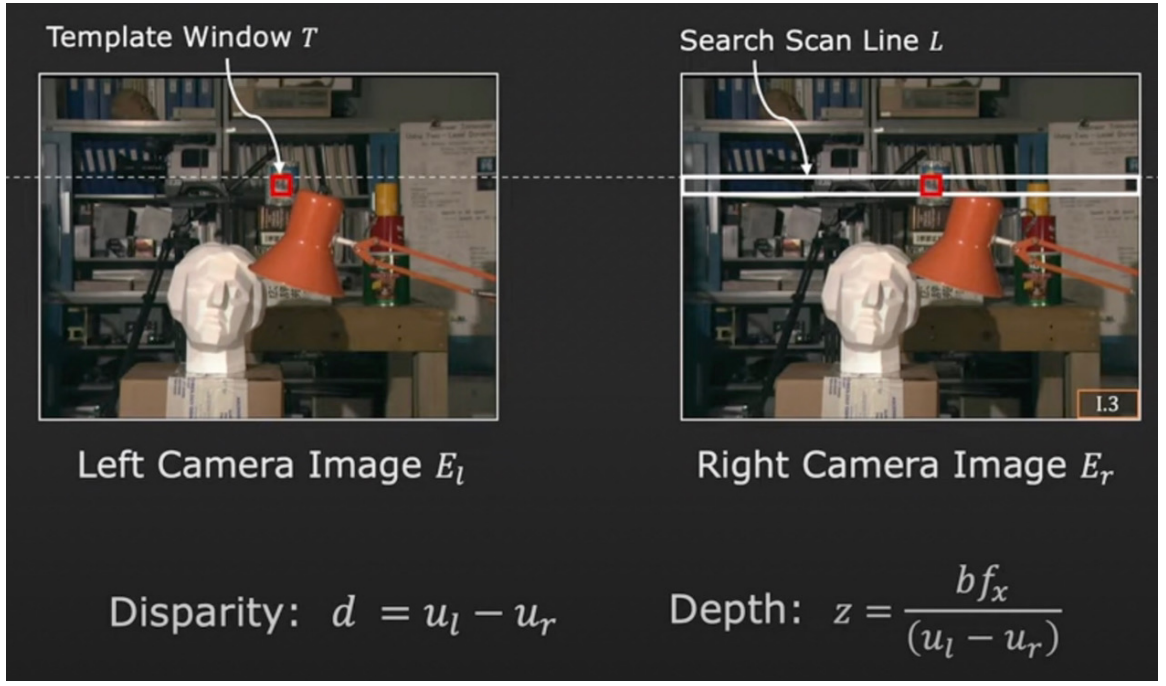


https://vision.middlebury.edu/stereo/data/scenes2021/

Multi-view Stereo

Goal: recover the disparity/depth map of the scene from two views

- Step 1: Camera calibration ☑
- Step 2: Estimation of the fundamental matrix F ☑
- Step 3: Dense stereo matching
  - 1D search through the **horizontal line**
- Step 4: Depth map reconstruction

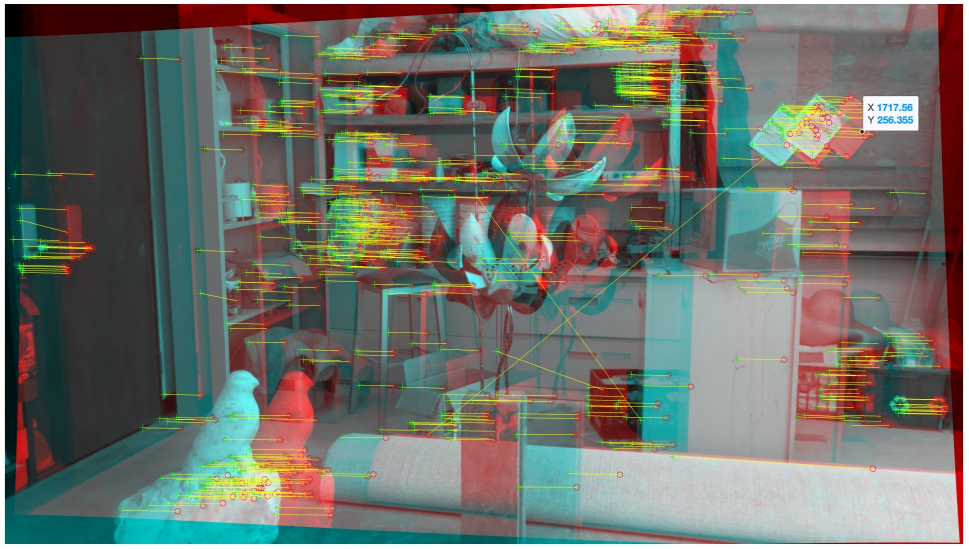


Estimation of the Fundamental Matrix

- Step 1: Find a set of corresponding features in left and right images using SIFT $\{(x_i, y_i)\}$

I_left

I_right



Correspondence Visualization

Estimation of the Fundamental Matrix

- Step 2:

  Repeat n times {

  - Randomly choose 8 matching pairs and solve for the following equation exactly by computing the null space of A
  - Find the consensus set for the fundamental matrix in the set $\{(x_i, y_i)\}$.
  - If the consensus set is sufficiently large, then we could refit the model using all points in the consensus set.
  - Save the best model}

$$\begin{bmatrix} x_1^1x_2^1 & y_1^1x_2^1 & x_2^1 & x_1^1y_2^1 & y_1^1y_2^1 & y_2^1 & x_1^1 & y_1^1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x_1^ix_2^i & y_1^ix_2^i & x_2^i & x_1^iy_2^i & y_1^iy_2^i & y_2^i & x_1^i & y_1^i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x_1^Nx_2^N & y_1^Nx_2^N & x_2^N & x_1^Ny_2^N & y_1^Ny_2^N & y_2^N & x_1^N & y_1^N & 1 \end{bmatrix} \begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Our estimation of F:

$$\begin{matrix} 0.0000 & 0.0000 & 0.0009 \\ -0.0000 & 0.0000 & 0.3335 \\ -0.0010 & -0.3337 & 0.8887 \end{matrix}$$

Stereo Calibration

- Goal: extract the translation matrix T and rotation matrix R that represent the relative geometry between two cameras given the fundamental matrix F

$$E = [T]_X R$$

- Theoretically, in our dataset, we only have a pure translation that is parallel to the x-axis (say, $t = (T, 0,0)$) then $[T]_X = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{matrix}$ and R = I = $\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$
- The rotation matrix R and translation matrix T extracted from our estimated fundamental matrix using SVD:

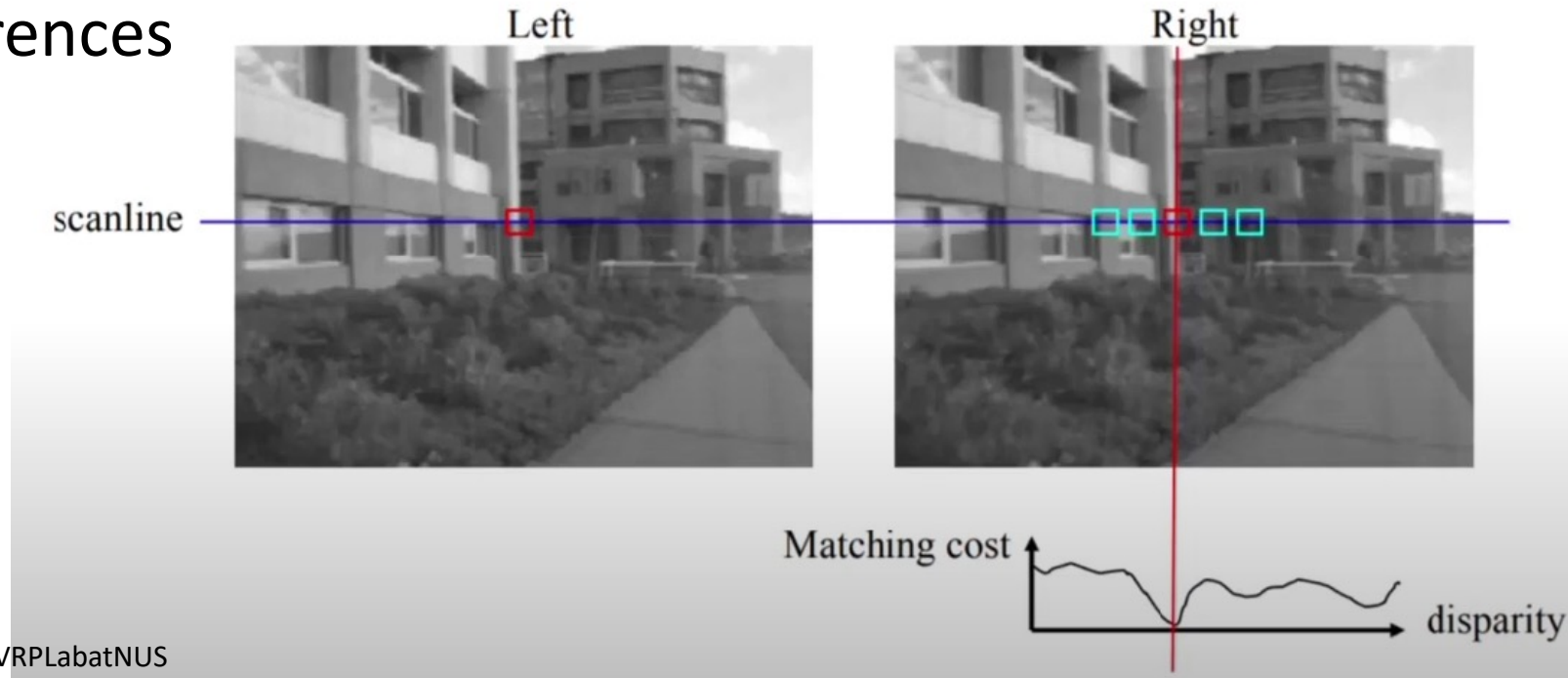
$$R = \begin{matrix} 1.000 & 0.0002 & 0.0000 \\ -0.0002 & 1.0000 & 0.0014 \\ -0.0000 & -0.0014 & 1.0000 \end{matrix}$$

$$T = \begin{matrix} 0.9996 & -0.0114 & 0.0274 \end{matrix}$$

Correspondence Search

Methods to compute the matching cost:

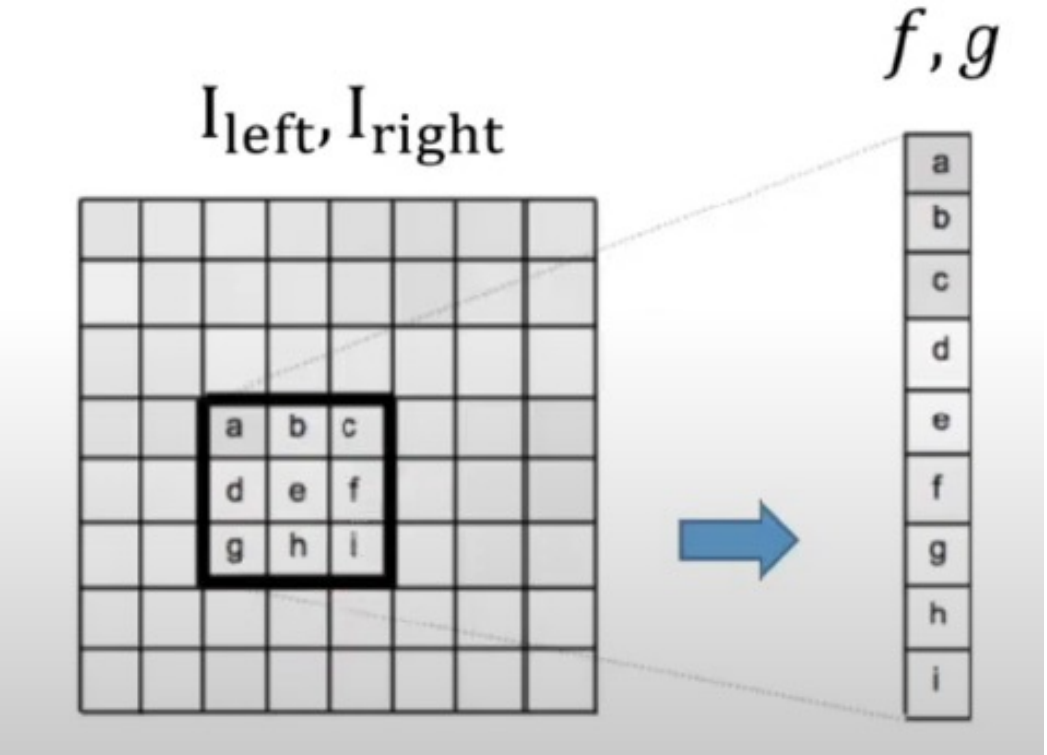
- Normalized Cross Correlation
- Sum of Squared Differences
- Sum of Absolute Differences
- Mutual Information

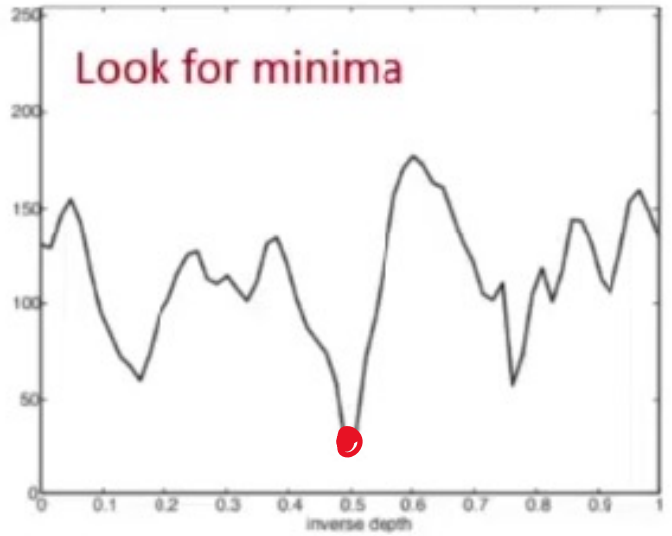


https://www.youtube.com/watch?v=N_iECTqZtQQ&ab_channel=CVRPLabatNUS

Sum of Squared Differences

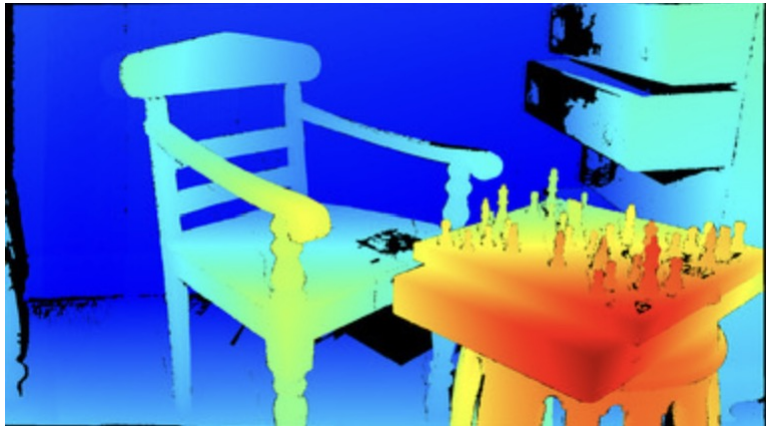
$$\rho_{SSD}(f,g) = ||f-g||^2$$





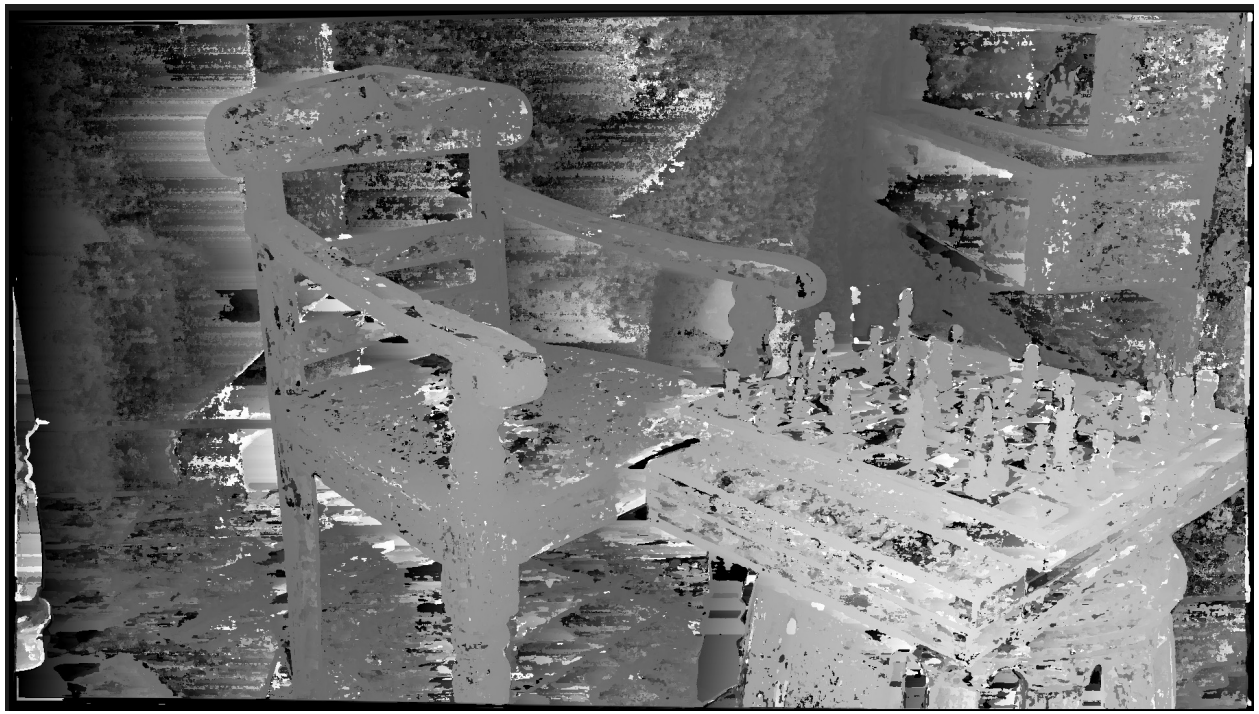
https://www.youtube.com/watch?v=hfQR7X8eH9s&t=1752s&ab_channel=CVRPLabatNUS

Chess room



Conservative estimation : 310

window size: 3



window size: 6

OpenCV depth reconstruction

Presented by Tad

StereoBM class and parameters

- Calculates a disparity map for a pair of rectified stereo images
- SAD as similarity metric
- Most important parameters
  - numDisparities
  - blockSize
- Demo