

# CSCI 5561: Assignment #5

## Stereo Reconstruction

---

### 1 Submission

- Assignment due: May 10 (11:55pm)
- Individual assignment
- Up to 2 page summary write-up with resulting visualization (more than 2 page assignment will be automatically returned.).
- Submission through Canvas.
- Provided codes and data can be downloaded from:  
<https://www-users.cs.umn.edu/~hspark/csci5561/HW5.zip>. It contains the following three codes and image data:

- `main_stereo.m`
- `ComputeCameraPose.m`
- `WarpImage.m`
- `left.bmp`
- `right.bmp`

- List of submission codes:

- `main_stereo.m`
- `FindMatch.m`
- `ComputeF.m`
- `Triangulation.m`
- `DisambiguatePose.m`
- `ComputeRectification.m`
- `DenseMatch.m`

- A MAT file that contains the following trained weights:

- `stereo.mat`: `x1`, `x2`, `F`, `X`, `H1`, `H2`, `im1_w`, `im2_w`, `disp`

- DO NOT SUBMIT THE PROVIDED IMAGE DATA

- The function that does not comply with its specification will not be graded.
- You are allowed to use MATLAB built-in functions except for the ones in the Computer Vision Toolbox. Please consult with TA if you are not sure about the list of allowed functions.

# CSCI 5561: Assignment #5

## Stereo Reconstruction

---

### 2 Overview

In this assignment, you will implement a stereo reconstruction algorithm given two view images.



(a) Left image



(b) Right image

Figure 1: In this assignment, you will implement a stereo reconstruction algorithm given two images.

You can download the skeletal code and data (`left.bmp` and `right.bmp`) from here:  
<https://www-users.cs.umn.edu/~hspark/csci5561/HW5.zip>

You will fill `main_stereo.m` that takes input images and intrinsic parameters  $\mathbf{K}$ , and produces a stereo disparity map.

# CSCI 5561: Assignment #5

## Stereo Reconstruction

---

### 3 SIFT Feature Matching

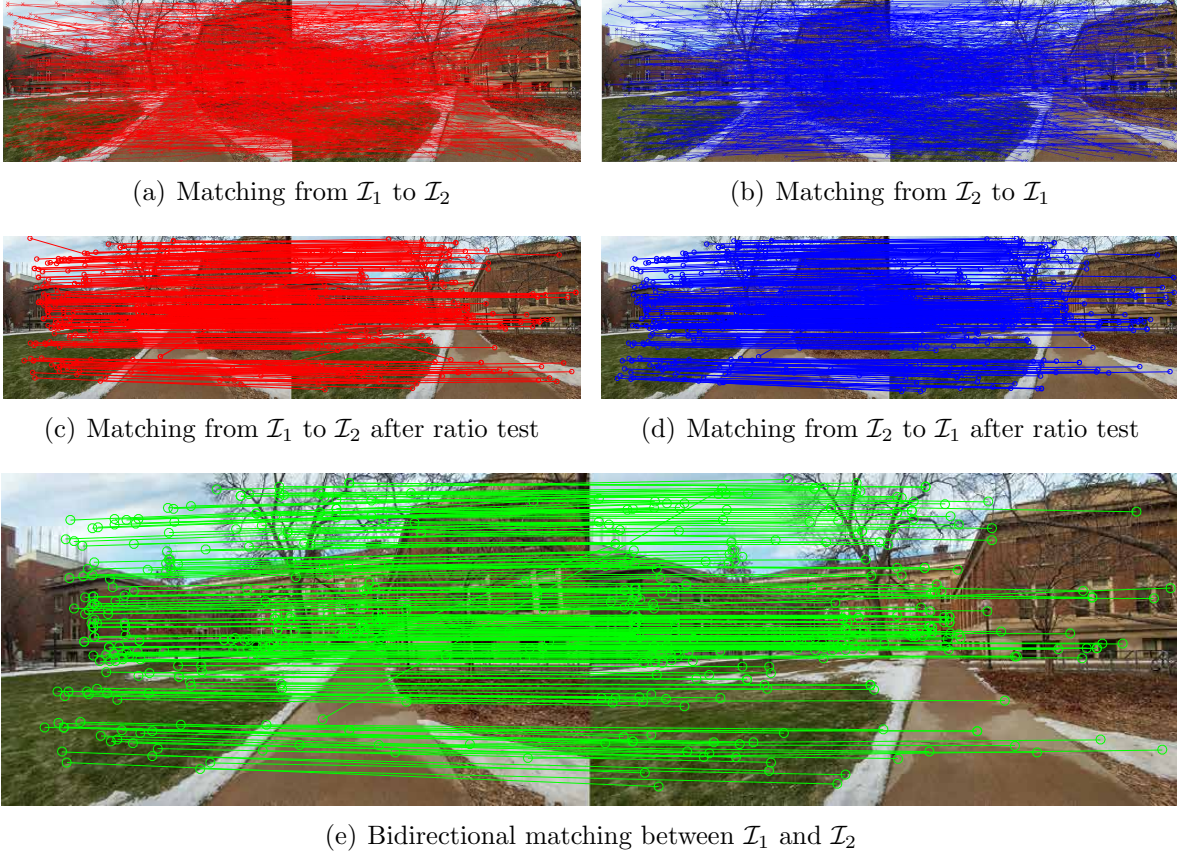


Figure 2: You will match points between  $\mathcal{I}_1$  and  $\mathcal{I}_2$  using SIFT features.

You will use `VLFeat` SIFT to extract keypoints and match between two views using k-nearest neighbor search. The matches will be filtered using the ratio test and bidirectional consistency check.

```
function [x1, x2] = FindMatch(I1, I2)
```

**Input:** two input gray-scale images with `uint8` format.

**Output:** `x1` and `x2` are  $n \times 2$  matrices that specify the correspondence.

**Description:** Each row of `x1` and `x2` contains the  $(x, y)$  coordinate of the point correspondence in  $I_1$  and  $I_2$ , respectively, i.e.,  $\mathbf{x1}(i, :) \leftrightarrow \mathbf{x2}(i, :)$ . This matching function is similar to `HW#2` except for bidirectional consistency check.

(Note) Except for SIFT extraction, you are not allowed to use `VLFeat` functions.

# CSCI 5561: Assignment #5

## Stereo Reconstruction

### 4 Fundamental Matrix Computation



Figure 3: Given matches, you will compute a fundamental matrix to draw epipolar lines.

`function [F] = ComputeF(x1, x2)`

**Input:**  $x1$  and  $x2$  are  $n \times 2$  matrices that specify the correspondence.

**Output:**  $F \in \mathbb{R}^3$  is the fundamental matrix.

**Description:**  $F$  is robustly computed by the 8-point algorithm within RANSAC. Note that the rank of the fundamental matrix needs to be 2 (SVD clean-up should be applied.). You can verify the validity of fundamental matrix by visualizing epipolar line as shown in Figure 3.

(Note) Given the fundamental matrix, you can run the provided function:

`[R1 C1 R2 C2 R3 C3 R4 C4] = ComputeCameraPose(F, K)`

This function computes the four sets of camera poses given the fundamental matrix where  $R1 \ C1 \dots \ R4 \ C4$  are rotation and camera center (represented in the world coordinate system) and  $K$  is the intrinsic parameter. These four configurations can be visualized in 3D as shown in Figure 4.

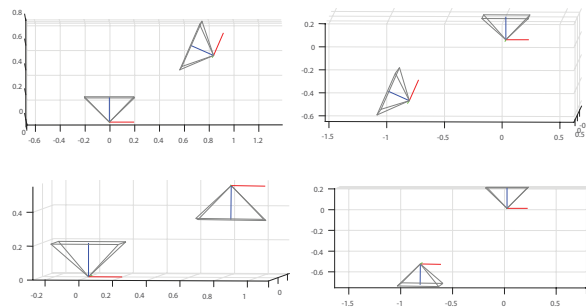


Figure 4: Four configurations of camera pose from a fundamental matrix.

# CSCI 5561: Assignment #5

## Stereo Reconstruction

### 5 Triangulation

Given camera pose and correspondences, you will triangulate to reconstruct 3D points.

`function [X] = Triangulation(P1, P2, x1, x2)`

**Input:** P1 and P2 are two camera projection matrices ( $\mathbb{R}^{3 \times 4}$ ). x1 and x2 are  $n \times 2$  matrices that specify the correspondence.

**Output:** X is  $n \times 3$  where each row specifies the 3D reconstructed point.

**Description:** Use the triangulation method by linear solve, i.e.,

$$\begin{bmatrix} \begin{bmatrix} \mathbf{u} \\ 1 \end{bmatrix} \times \mathbf{P}_1 \\ \begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix} \times \mathbf{P}_2 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = \mathbf{0}$$

(Note) Use `plot3` to visualize them as shown in Figure 5.

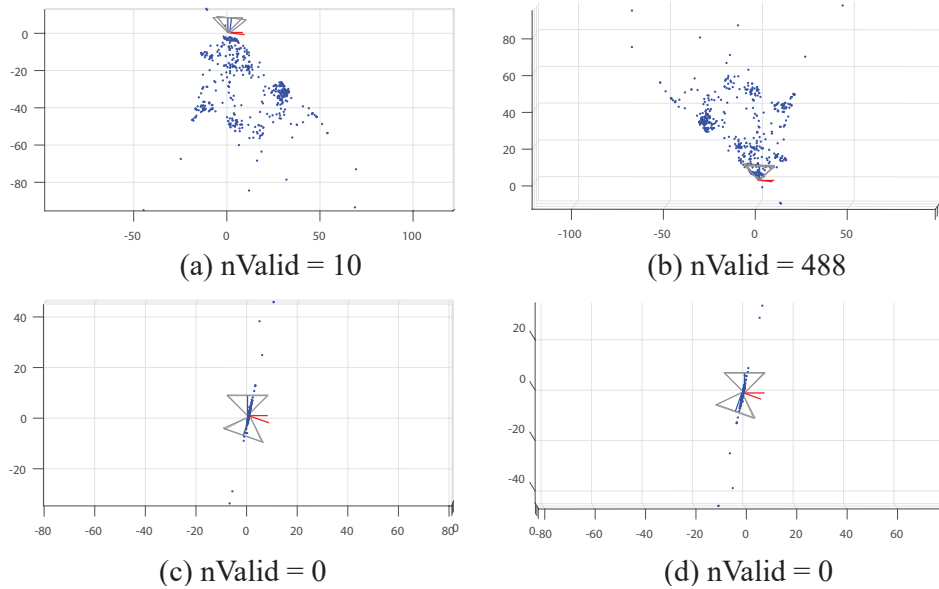


Figure 5: You can visualize four camera pose configurations with point cloud.

# CSCI 5561: Assignment #5

## Stereo Reconstruction

---

### 6 Pose Disambiguation

Given four configurations of relative camera pose and reconstructed points, you will find the best camera pose by verifying through 3D point triangulation.

`function [R,C,X] = DisambiguatePose(R1,C1,X1,R2,C2,X2,R3,C3,X3,R4,C4,X4)`

**Input:** R1, C1, X1  $\cdots$  R4, C4, X4 are four sets of camera rotation, center, and 3D reconstructed points.

**Output:** R, C, X are the best camera rotation, center, and 3D reconstructed points.

**Description:** The 3D point must lie in front of the both cameras, which can be tested by:

$$\mathbf{r}_3^T(\mathbf{X} - \mathbf{C}) > 0 \quad (1)$$

where  $\mathbf{r}_3$  is the 3<sup>rd</sup> row of the rotation matrix. In Figure 5, nValid means the number of points that are in front of both cameras. (b) configuration produces the maximum number of valid points, and therefore the best configuration is (b).



# CSCI 5561: Assignment #5

## Stereo Reconstruction

---

### 7 Stereo



Figure 6: Stereo rectification.

Given the disambiguated camera pose, you will implement dense stereo matching between two views based on dense SIFT of `VLFeat`.

**function** [H1, H2] = ComputeRectification(K, R, C)

**Input:** The relative camera pose ( $\mathbf{R}$  and  $\mathbf{C}$ ) and intrinsic parameter  $\mathbf{K}$ .

**Output:** H1 and H2 are homographies that rectify the left and right images such that the epipoles are at infinity.

**Description:** Given the disambiguated camera pose, you can find the rectification rotation matrix,  $\mathbf{R}_{\text{rect}}$  such that the x-axis of the images aligns with the baseline. Find the rectification homography  $\mathbf{H} = \mathbf{K}\mathbf{R}_{\text{rect}}\mathbf{R}^T\mathbf{K}^{-1}$  where  $\mathbf{R}$  is the rotation matrix of the camera. The rectified images are shown in Figure 6. This rectification sends the epipoles to infinity where the epipolar line becomes horizontal.

(Note) You can use the provided image warping function `im_w = WarpImage(im, H)` to check your result.

# CSCI 5561: Assignment #5

## Stereo Reconstruction

---

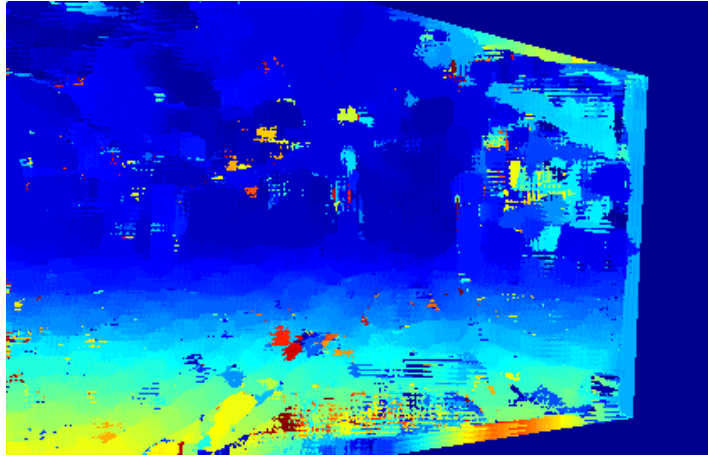


Figure 7: Visualization of stereo match.

```
function [disp] = DenseMatch(im1, im2)
```

**Input:** two gray-scale rectified images with `uint8` format.

**Output:** disparity map  $\text{disp} \in \mathbb{R}^{H \times W}$  where  $H$  and  $W$  are the image height and width.

**Description:** Compute the dense matches across all pixels. Given a pixel,  $\mathbf{u}$  in the left image, sweep along its epipolar line,  $\mathbf{l}_{\mathbf{u}}$ , and find the disparity,  $d$ , that produces the best match, i.e.,

$$d = \arg \min_i \|\mathbf{d}_{\mathbf{u}}^1 - \mathbf{d}_{\mathbf{u}+(i,0)}^2\|^2 \quad \forall i = 0, 1, \dots, N$$

where  $\mathbf{d}_{\mathbf{u}}^1$  is the dense SIFT descriptor at  $\mathbf{u}$  on the left image and  $\mathbf{d}_{\mathbf{u}+(i,0)}^2$  is the SIFT descriptor at  $\mathbf{u} + (i, 0)$  ( $i$  pixel displaced along the x-axis) on the right image. Visualize the disparity map as shown in Figure 7.