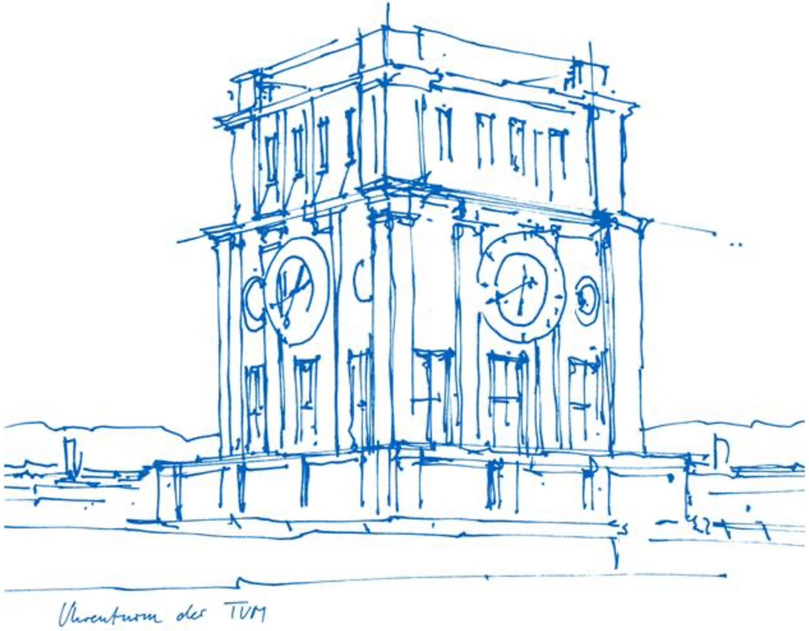
3D Stereo Reconstruction

Advisor: Yuchen Rao, Andrei Burov, Angela Dai

Students: Barry (Shichen) Hu, Ran Ding, Jiongyan Zhang



Agenda

- Introduction
- Method
- Quantitative Evaluation
- Qualitative Evaluation
- Conclusion

Introduction

- Goal for this project:

Left view

Right view

Pointcloud
Reconstruction

- Dataset: Middlebury 2021 Mobile dataset

https://vision.middlebury.edu/stereo/data/
D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth.
In *German Conference on Pattern Recognition (GCPR 2014), Münster, Germany,* September 2014.

Method

- Structure from Motion (SfM)
- Bundle Adjustment
- Multiview Stereo (MVS)

Method - 1: Structure from Motion

Target: Recover the extrinsic matrix and sparse points

How?

1. Keypoint extraction: SIFT, ORB, SURF

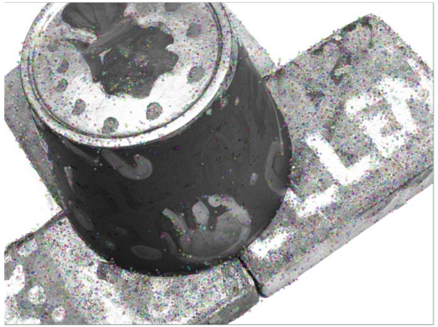
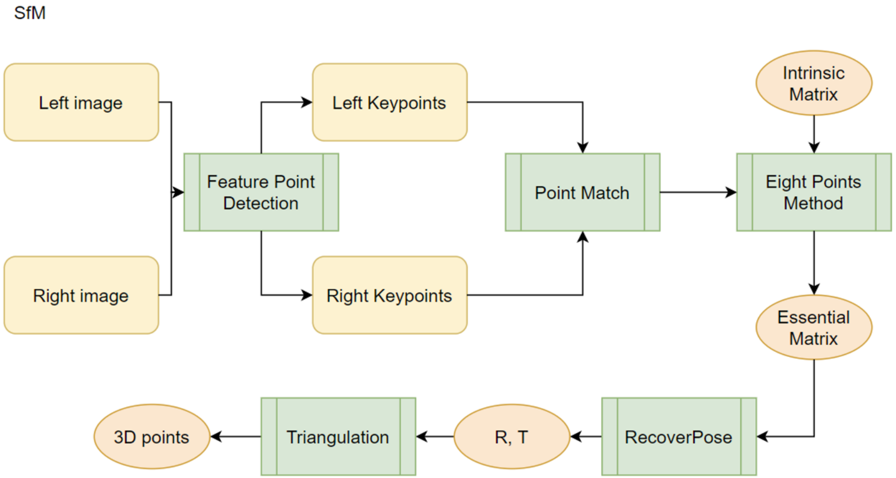


Figure 1: Keypoints (SURF)



Method - 1: Structure from Motion

Target: Recover the extrinsic matrix and sparse points

How?

2. Matching: Brute force or Flann

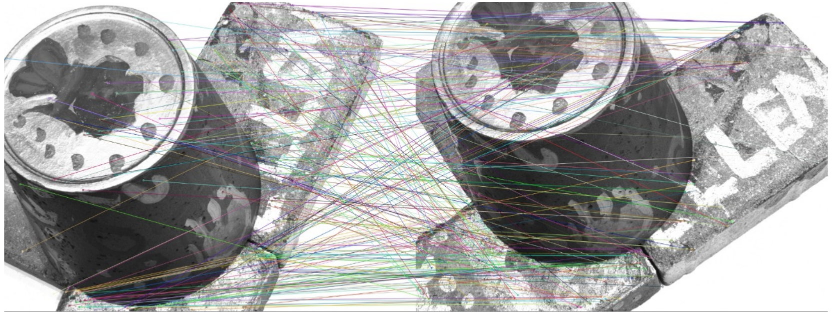


Figure 3: Keypoints matching

3. Filtering: Lowe’s ratio test

Method - 1: Structure from Motion

Target: Recover the extrinsic matrix and sparse points

How?

4. Fundamental matrix:
    - 8 points algorithm
    - RANSAC algorithm ✅

5. Recover Essential matrix
6. Recover R and T

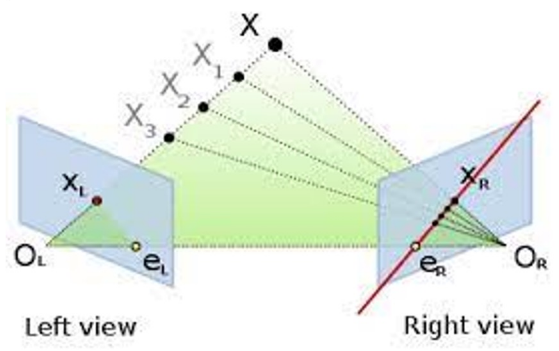


Figure 4: Epipolar geometry, from https://en.wikipedia.org/wiki/Epipolar_geometry

Method - 1: Structure from Motion

Target: Recover the extrinsic matrix and sparse points

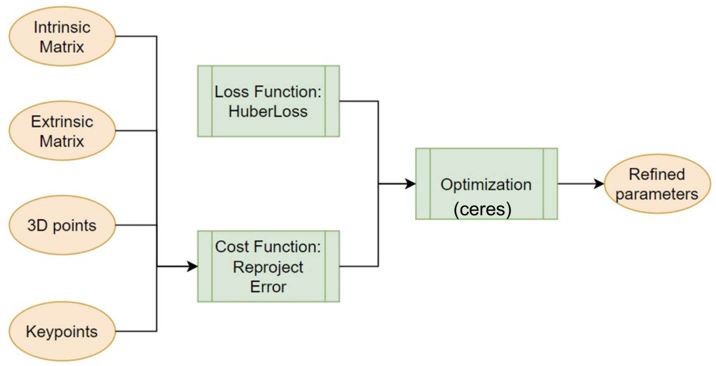
How?

7. Triangulation

Figure 5: recovered sparse points

Method - 2: Bundle Adjustment

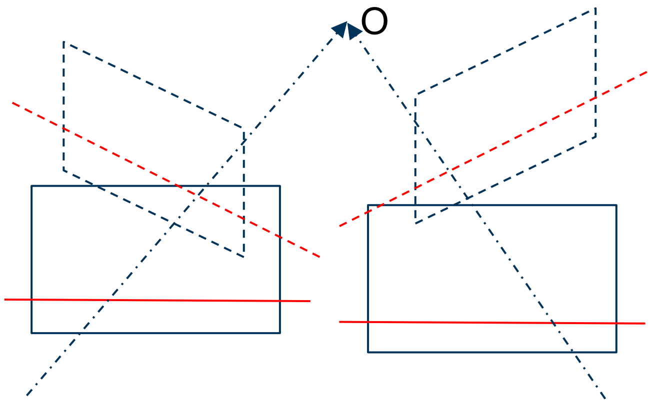
- Optimising the reprojection error objective.
- **Huber loss** is used for its robustness.
- **Levenberg-Marquardt**
- Ceres (RMSE)




```
Bundle Adjustment statistics (approximated RMSE):
 #residuals: 19962
 Initial RMSE: 2.36064
 Final RMSE: 1.88907
 Time (s): 31.7501
```

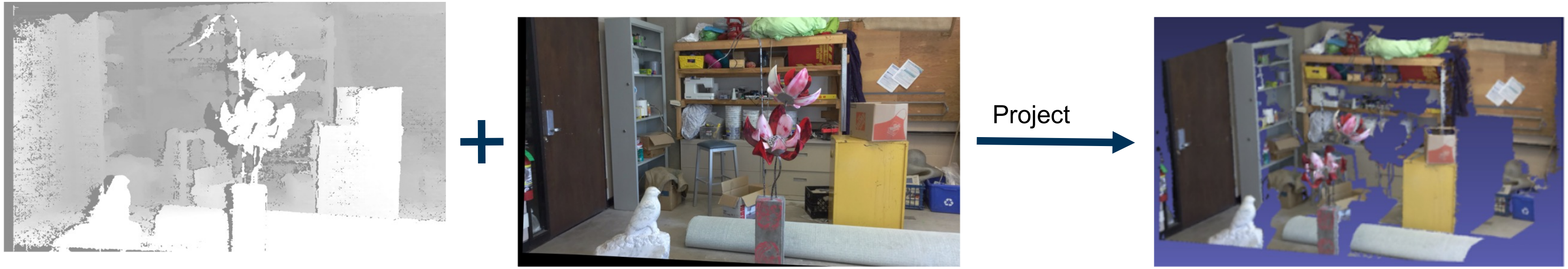
Method - 3: Multiview Stereo

- **Image rectify:** project images onto a common image plane.



Method - 3: Multiview Stereo

- **StereoBM**
- Semi Global Block Matching (**SGBM**)
- Depth map -> Point cloud



Depth Map RGB Info Point Cloud

Method - 3: Multiview Stereo

The semi-global algorithm attempts to establish a global Markov energy equation by constraining the one-dimensional path in multiple directions on the image. The final matching cost of each pixel is the superposition of all path information:

$$L_r(p,d) = c(p,d) + \min \left\{ \begin{array}{c} L_r(p-r,d) \\ L_r(p-r,d\pm 1) + P_1 \\ \min_{i=d_{\min},\ldots,d_{\max}} L_r(p-r,i) + P_2 \end{array} \right\} - \min_{i=d_{\min},\ldots,d_{\max}} L_r(p-r,i)$$

The output is the depth map. We can use the projection matrix Q to obtain the 3D points.

The color information can be obtained from the original RGB image.

Quantitative Results

- Keypoints detection

Method	Avg number of detected points*
ORB	500
SIFT	3850
SURF	**7708**

*averaged over 24 scenes in Middlebury 2021 Mobile Dataset

https://vision.middlebury.edu/stereo/data/
D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition (GCPR 2014), Münster, Germany,* September 2014.

Quantitative Results

- Keypoints Matching

Detection	Matching	Outlier Ratio[1]	Avg processing time(s)
SIFT	Flann (KDTree)	13.77	0.634
	BF[2]	**12.94**	0.608
ORB	Flann (LSH)	17.64	0.111
	BF	14.39	**0.104**
SURF	Flann (KDTree)	21.46	0.660
	BF	20.23	0.795

[1]Outliers ratios calculated from the RANSAC algorithms.
[2]Brutal Force

Quantitative Results

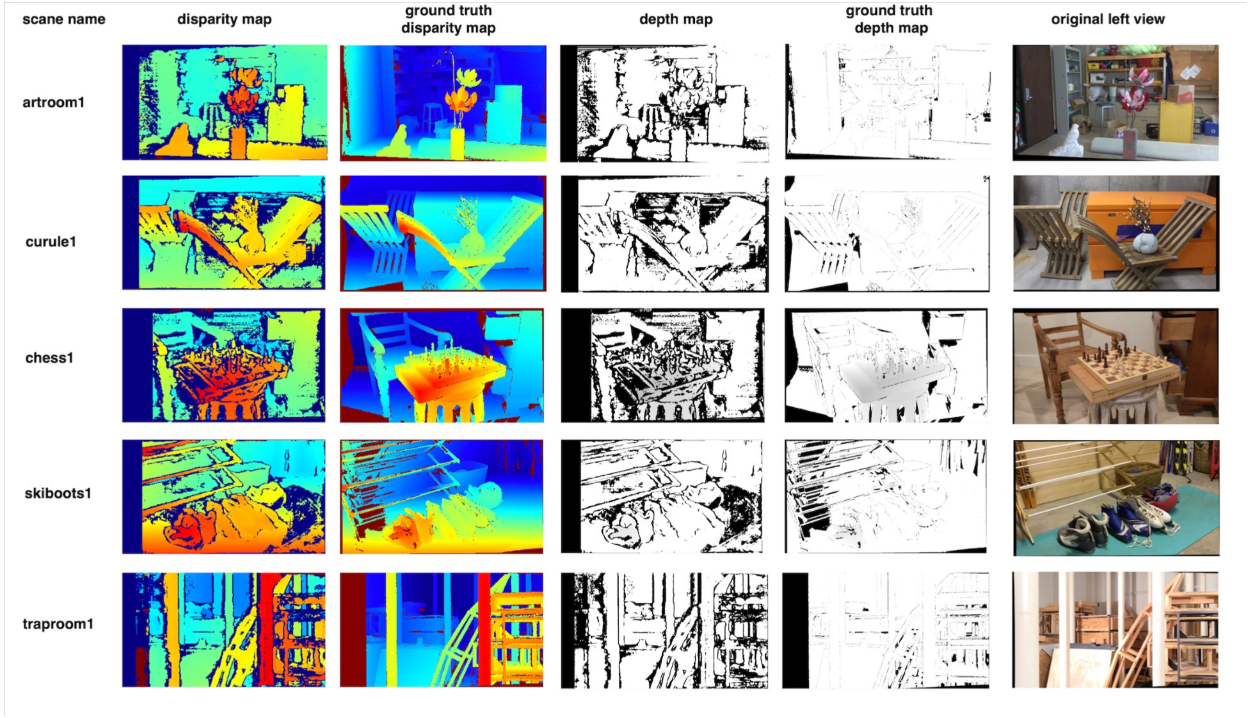
- Disparity map generation

Method	Avg Bad2.0* score	Avg processing time(s)
SGBM	**46.633**	0.941
StereoBM	65.213	**0.216**

*Bad2.0: the percentage of the bad pixels with disparity error larger than 2 pixels

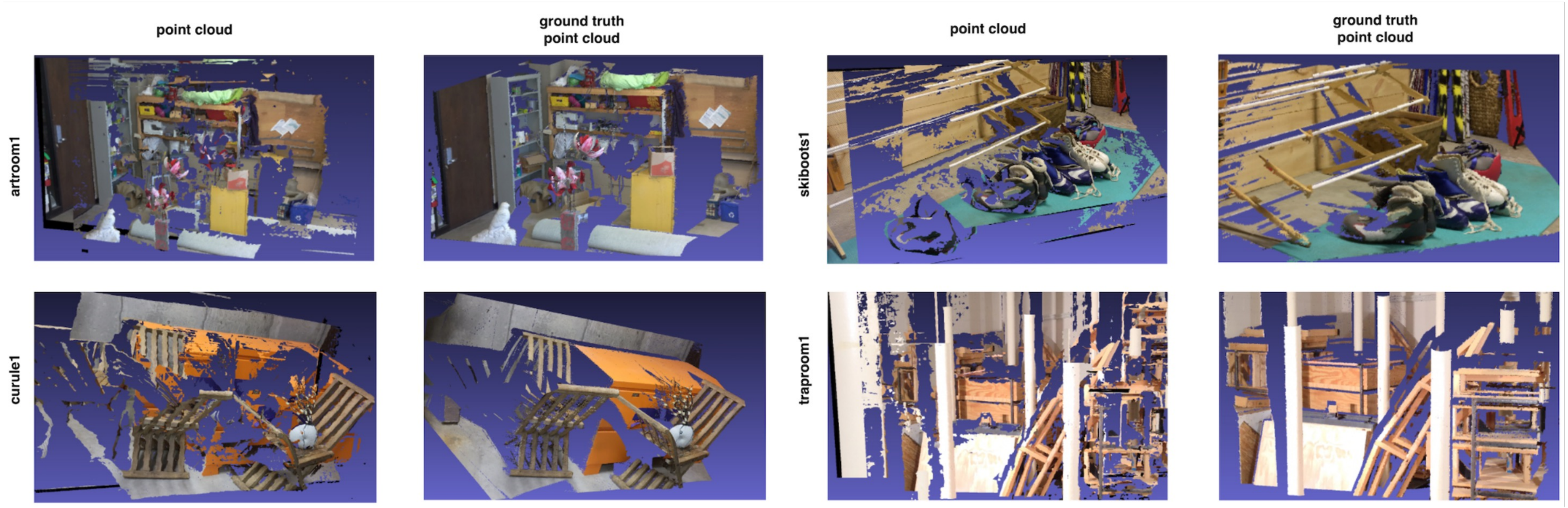
https://vision.middlebury.edu/stereo/data/
D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition (GCPR 2014), Münster, Germany,* September 2014.

Qualitative Results



https://vision.middlebury.edu/stereo/data/
D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition (GCPR 2014), Münster, Germany,* September 2014.

Qualitative Results



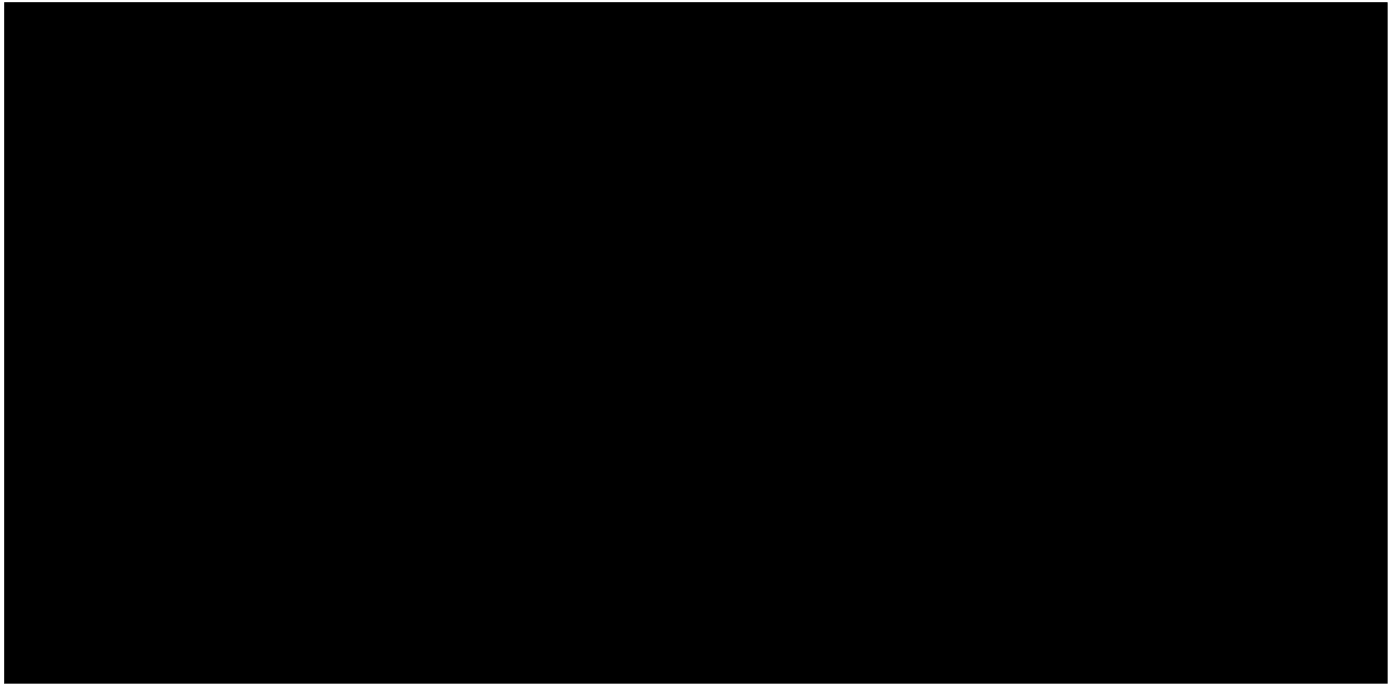
https://vision.middlebury.edu/stereo/data/

D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition (GCPR 2014), Münster, Germany,* September 2014.

Qualitative Results - Artroom scene

https://vision.middlebury.edu/stereo/data/
D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition (GCPR 2014), Münster, Germany,* September 2014.

Qualitative Results - Skiboots scene



https://vision.middlebury.edu/stereo/data/
D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth.
In *German Conference on Pattern Recognition (GCPR 2014), Münster, Germany,* September 2014.

Conclusion

- Experimented with different detection and matching methods. compare them:
    - Quantitatively
    - Qualitatively
- Reconstruction is bad in constant-intensity areas

Any Questions?