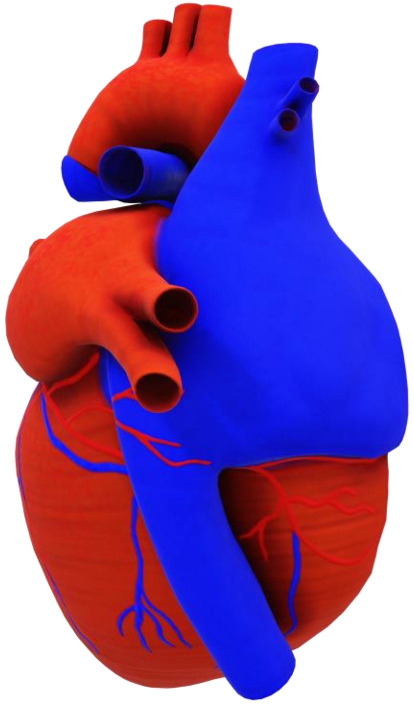
3D Scanning and Motion Capturing

Stereo Reconstruction

Motivation and Idea

Various Applications



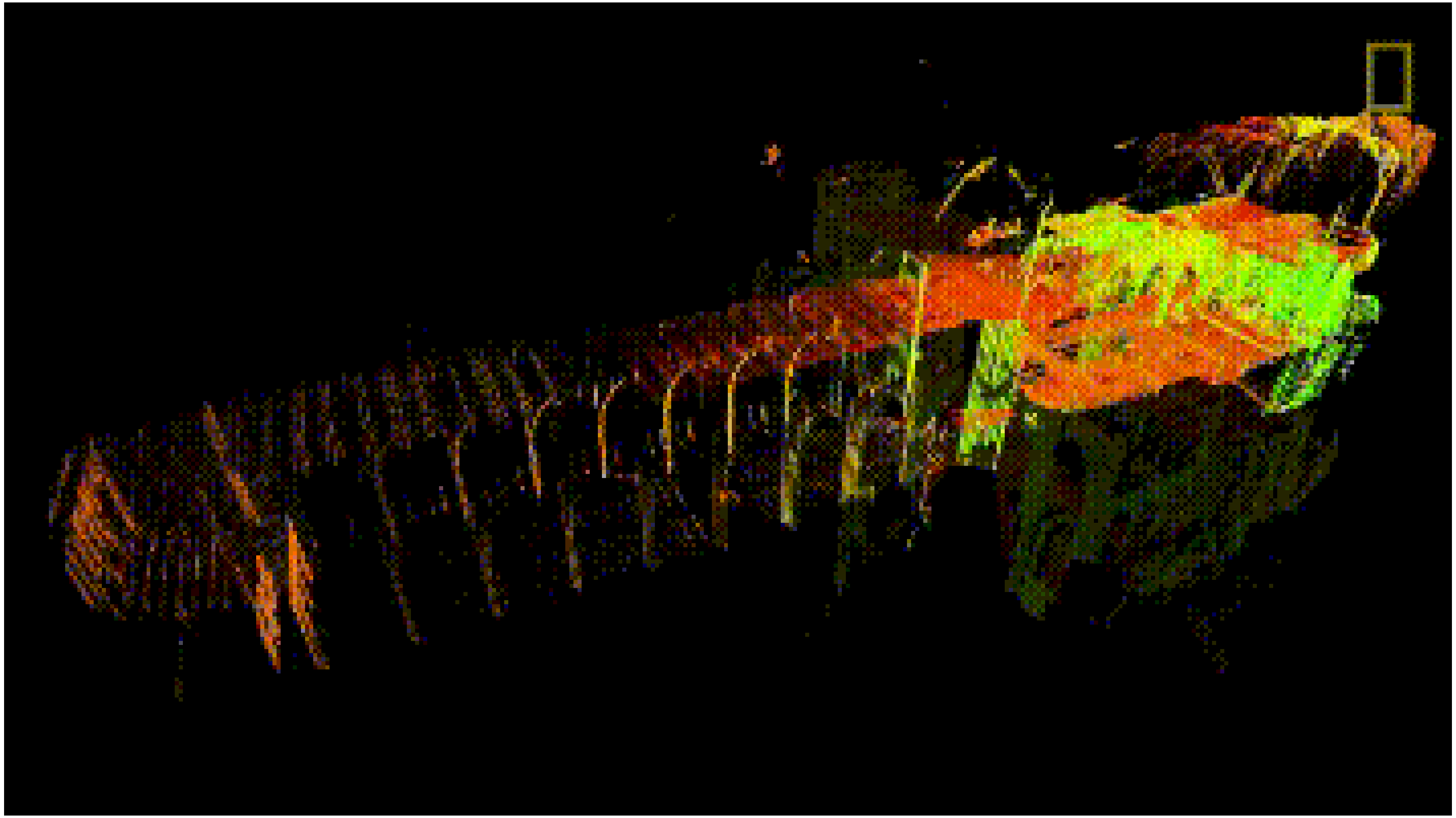
3D models of human organs for surgical planning

Reconstruct historical architectures

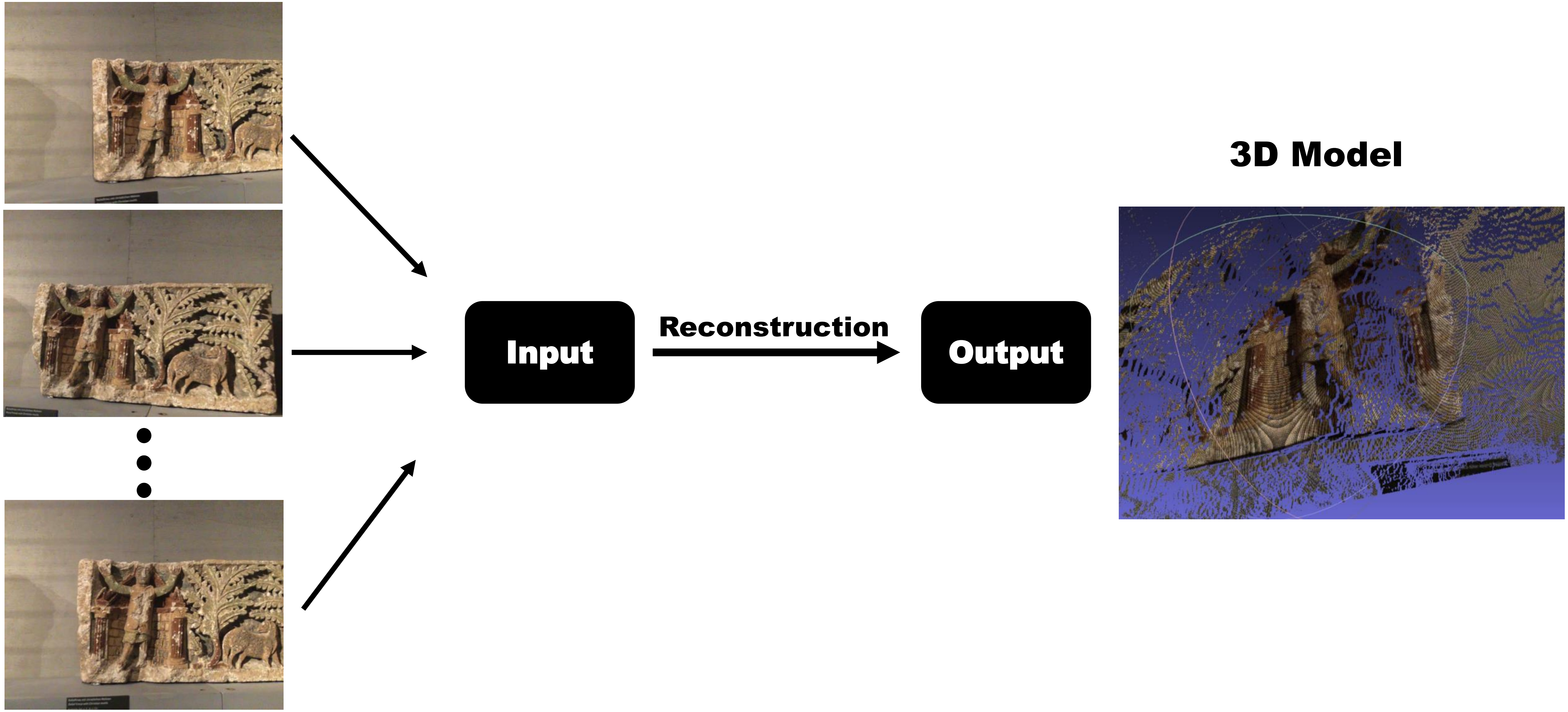
Reconstruct environment for robots navigation

Various Applications

Reconstruction of Notre Dame



Our Idea

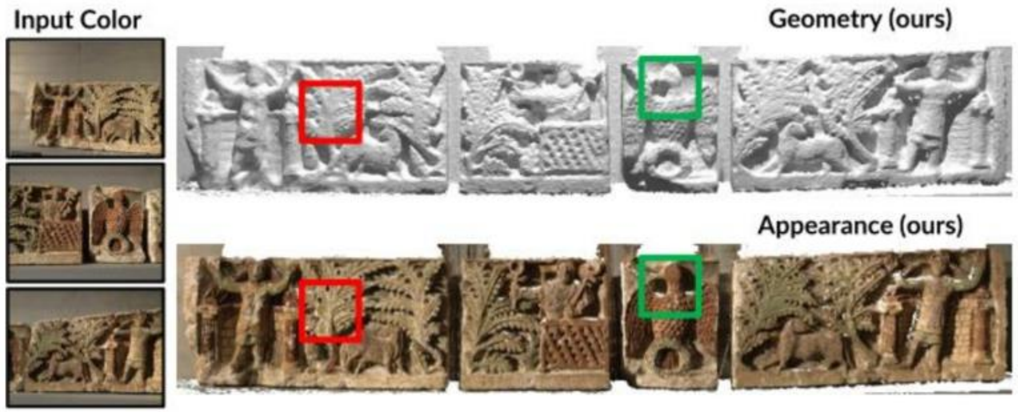


Related Work

Choose Datasets

TUM Intrinsic3D Bricks RGBD

BRICKS



bricks-preview.jpg

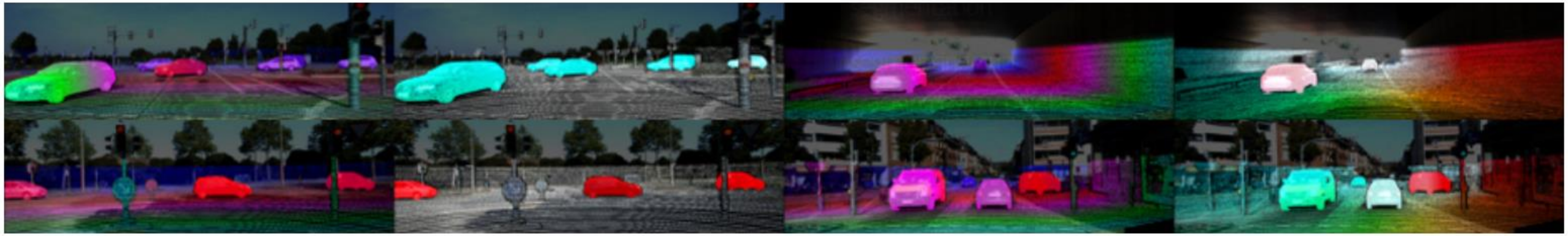
Data Provided

- Color and depth images
- Color and depth camera intrinsics
- Camera pose matrix
- Multiview of single object

Choose Datasets

Kitti Stereo 2015

Stereo Evaluation 2015



Data Provided

- **disparity maps**
- **color images**
- **camera distortion coefficient**
- **camera extrinsic**
- **images of different scenarios**

Dataset Comparisons

	TUM Intrinsic3D	Kitti Stereo 2015
Pros	• No need to consider camera distortion • Able to apply ICP • Provide ground truth depth	• camera parameters can be easily load • Ground truth disparity data can be directly used for evaluation
Cons	• No ground truth disparity	• Images of a single scenario not sufficient

Libraries

OpenCV

Computer vision algorithms

Eigen

Matrix compute

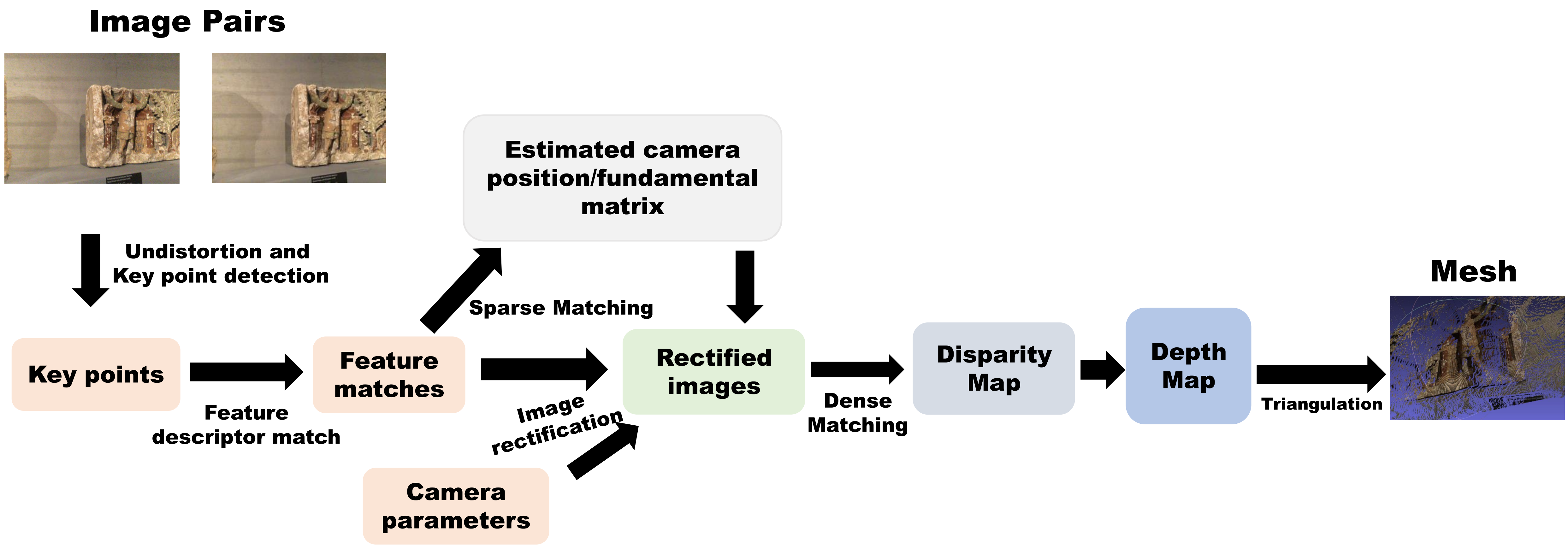
- **FLANN**

Nearest neighbor search

- **Ceres Solver**

Non-linear optimization

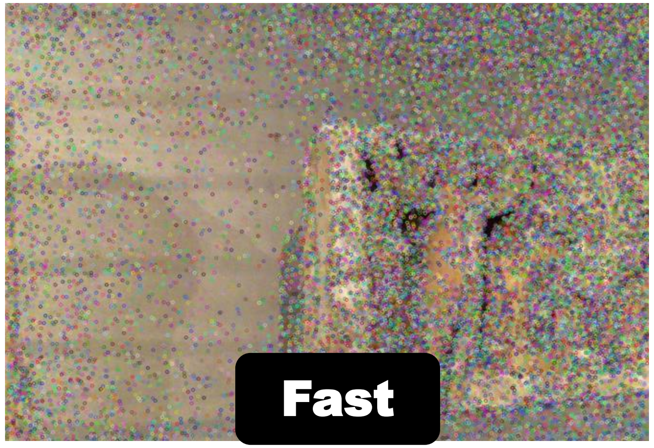
Reconstruction Pipeline

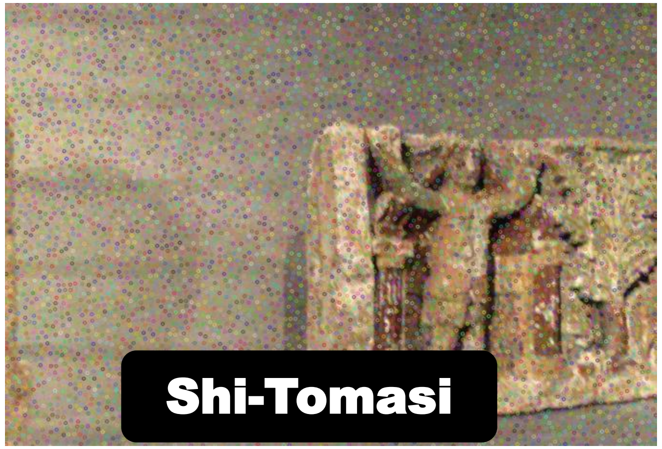


Methods

Sparse Matching

Key Point Detector





Key point detector and feature descriptor

Dataset Name	Measurements	brisk	orb	shi-tomasi	sift	surf	fast
TUM Intrinsic Dataset	Average rotation error	0.0616	0.1241	0.1053	0.1014	0.0838	0.0619
	Abnormal fundamental matrix number	20	40	33	33	27	20
	Average translation error	0.9993	0.9998	0.9993	0.9994	0.9993	0.9994
	time to process 10 pairs images/s	22.5254	57.269	62.391	430.32	44.541	670.31
KITTI Raw Dataset	Average rotation error	0.0125	0.0318	0.0097	0.0078		0.0111
	Abnormal fundamental matrix number	0	0	0	0		0
	Average translation error	0.7645	1.03877	0.5479	0.5403		0.6106

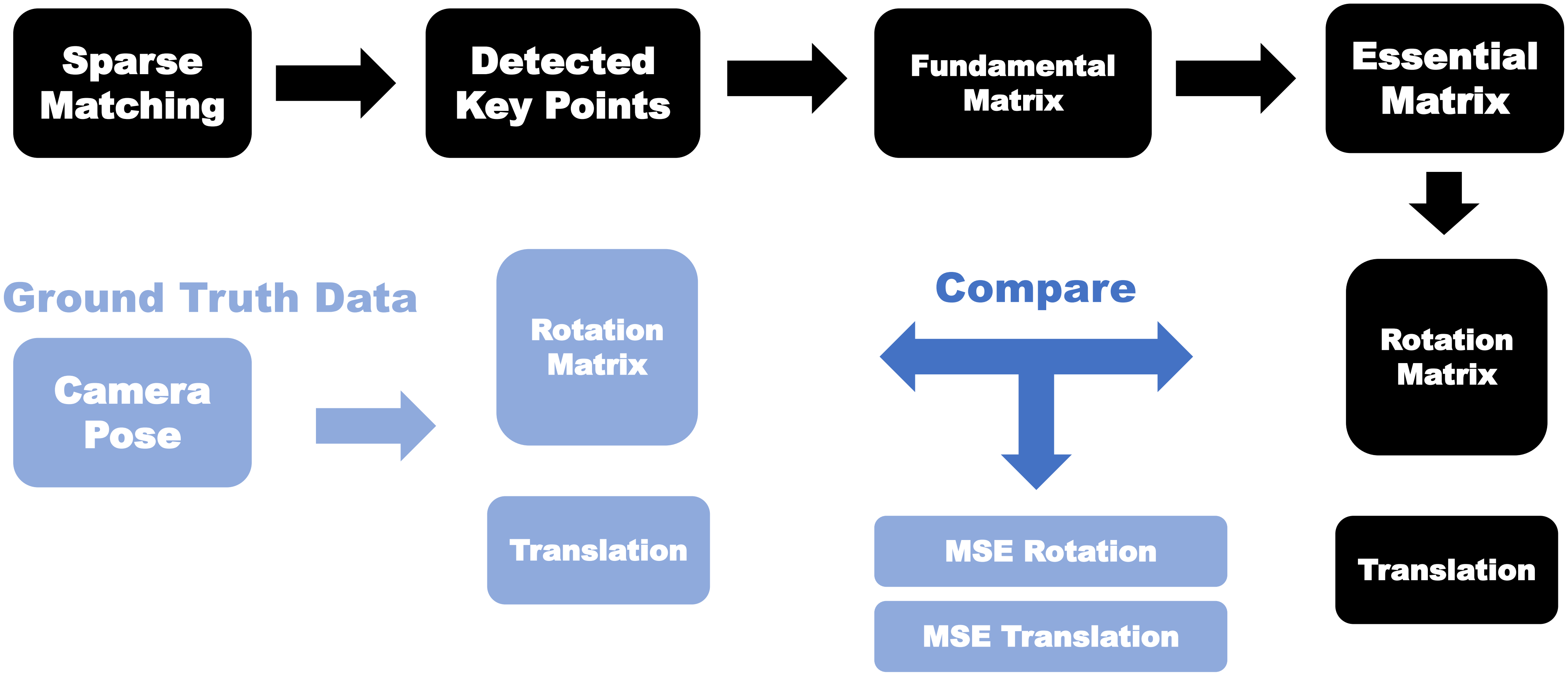
Performance of different methods on two datasets, using RANSAC as fundamental matrix calculation method and FLANN as descriptor matching method

 # translation error is too high in all methods, so we use ground truth translation for further steps

Fundamental Matrix Calculation

Fundamental matrix estimation

Evaluating sparse matching method



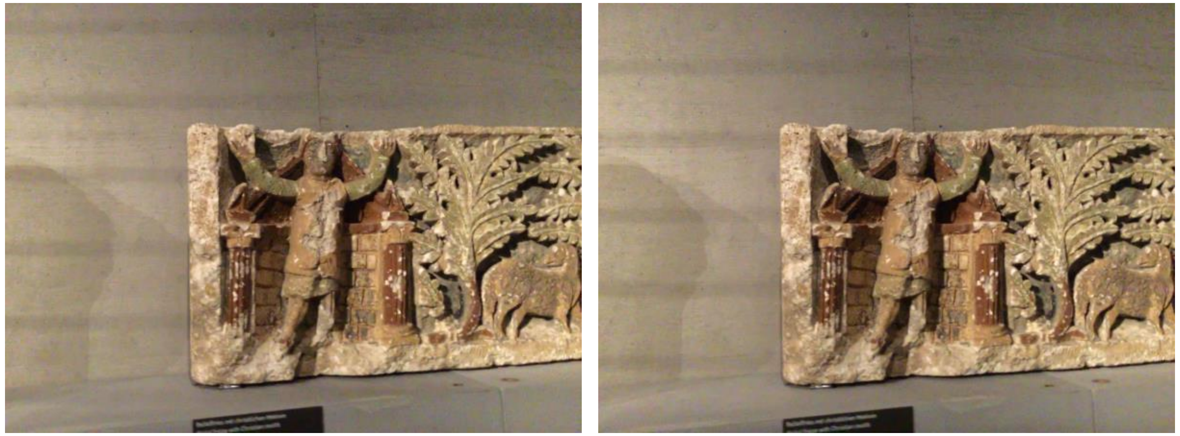
Fundamental matrix estimation

Dataset Name	Measurements	RANSAC	LMEDS	7 Point	8 Point
TUM Intrinsic Dataset	Average rotation error	0.0616	0.1813	0.1813	0.0296
	Abnormal fundamental matrix number	20	60	60	9
	Average translation error	0.9993	0.9998	0.9998	0.9995
KITTI Raw Dataset	Average rotation error	0.0125	0.0055	0.0055	0.0199
	Abnormal fundamental matrix number	0	0	0	0
	Average translation error	0.7645	0.5275	0.5275	0.9225

Performance of different methods on two datasets, using BRISK as key point detector and feature descriptor and FLANN as descriptor matching method

 # translation error is too high in all methods, so we use ground truth translation for further steps

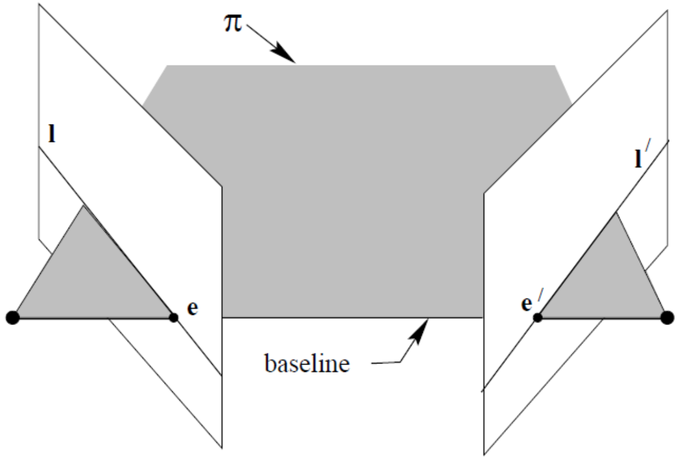
Image Rectification



Rectify

Rectify

Image Rectification



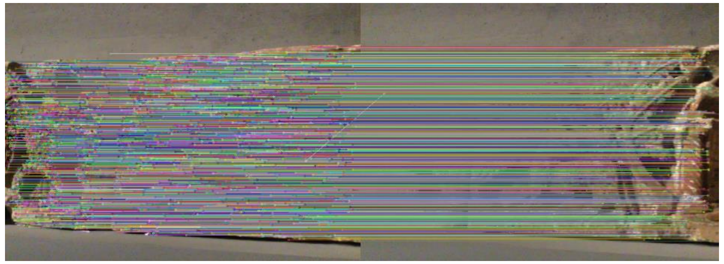
Fundamental Matrix

↕

Epipolar Lines

Dense Matching

Reconstruction Pipeline

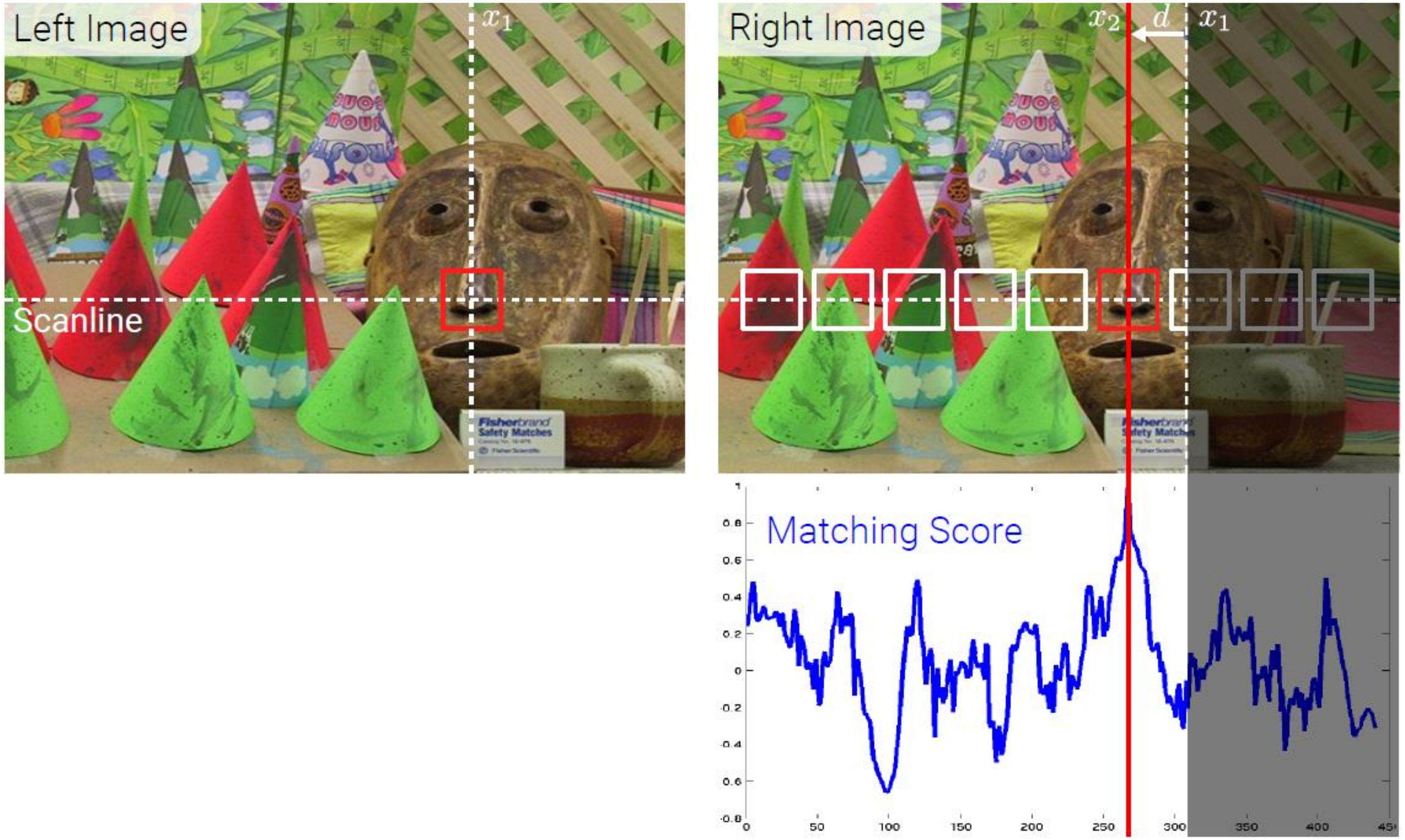


- Fundamental Matrix
- Essential Matrix
- Translation Matrix
- Rotation Matrix

Calculate Disparity

Block Matching

- Once correspondences are on parallel epipolar lines, we can scan the left image line by line
- Need to find "similar" pixels in the right image
- Individual pixels are hard to match, so let's take a "block" of flattened pixels of size **K×K**
- Subtract the difference in x-axis to get disparity



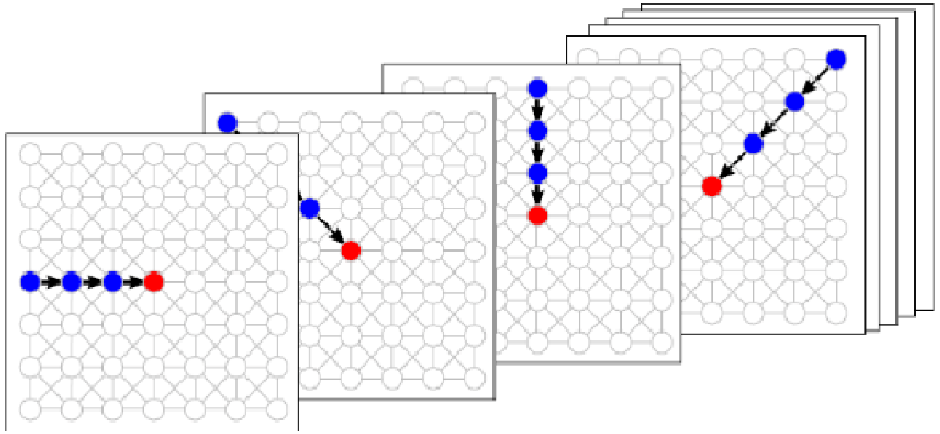
Block Matching

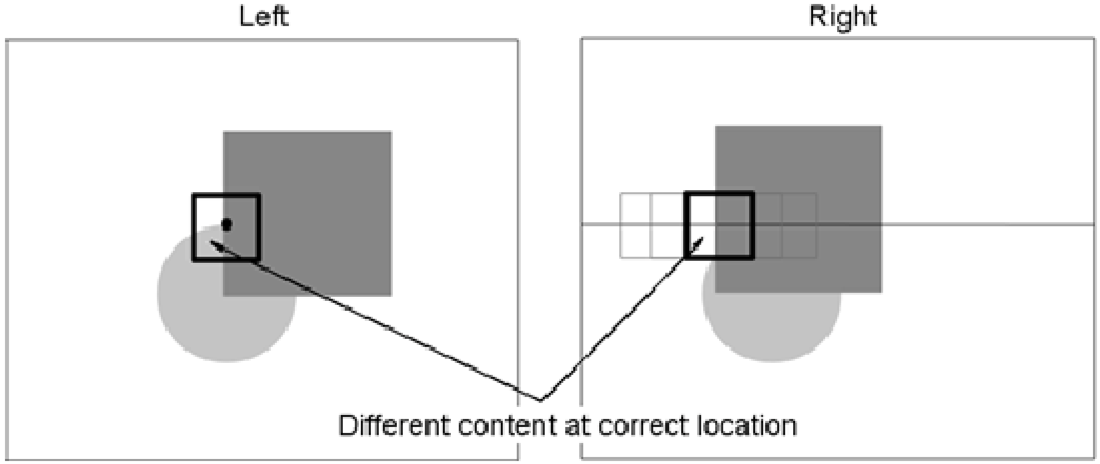
- How do we quantify “similar”?
- Consider two **K×K** windows of pixels flattened to vectors $\mathbf{w}_L$, $\mathbf{w}_R$
- Sum of absolute differences (SAD):

$$\text{SAD}(x, y, d) = ||\mathbf{w}_L(x, y) - \mathbf{w}_R(x - d, y)||$$

Semi-global Block Matching

- "Global" because it considers information available in the whole image
- Does so by considering the neighbouring pixels in multiple directions to calculate the disparity
- Again, we take a "block" of size **K×K**, and consider 8 directions around it
- The direction with the min. aggregated matching cost is used to calculate the disparity





Semi-global Block Matching

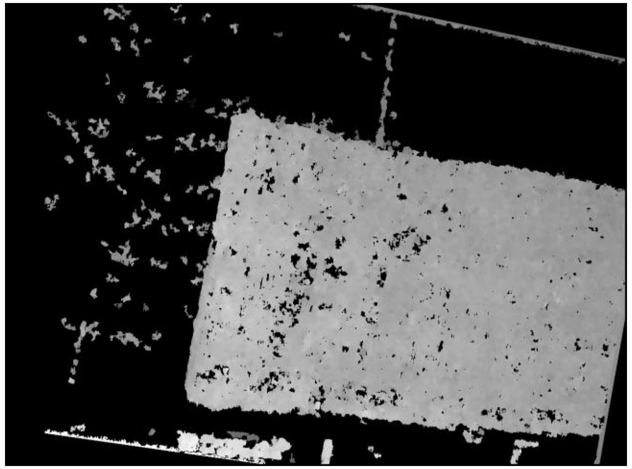
- Uses Birchfield-Tomasi dissimilarity to find “similar” pixels
- For left and right images, with x_l and x_r columns along the same scanline, we can define symmetric functions:

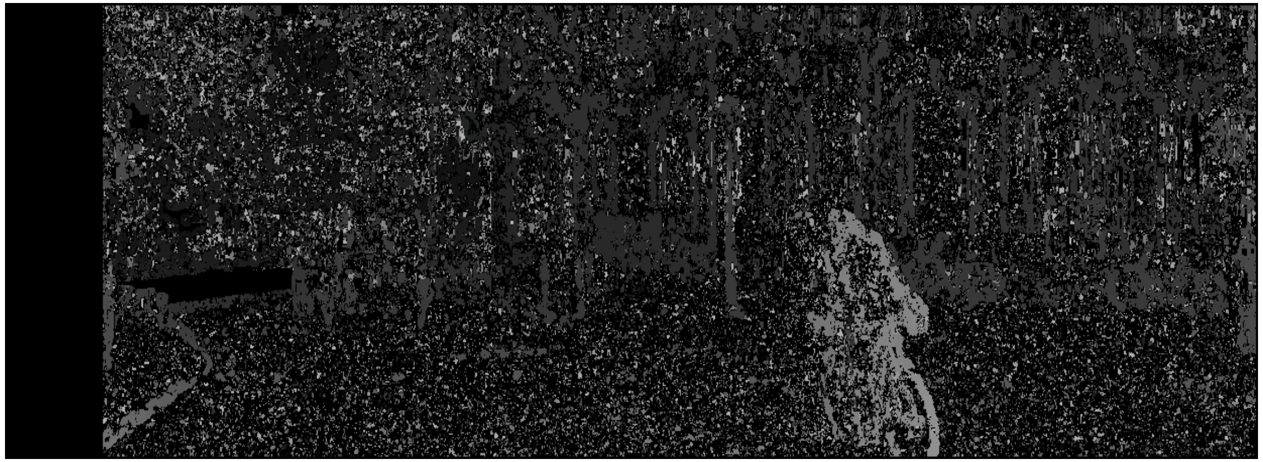
- Here $I^{(hat)}_l$ and $I^{(hat)}_r$ are the linear interpolation functions of the left and right image intensity I_l and I_r
- The min. of the symmetric functions gives the dissimilarity value

- Now D is the dissimilarity value, and R is the regularisation term computed from neighbouring pixels. Total disparity is:

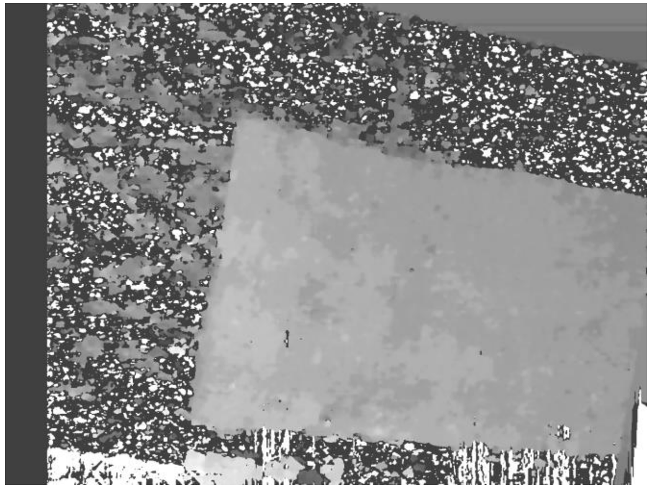
Generated Disparity Map **without** Post Filtering

BM





SGBM

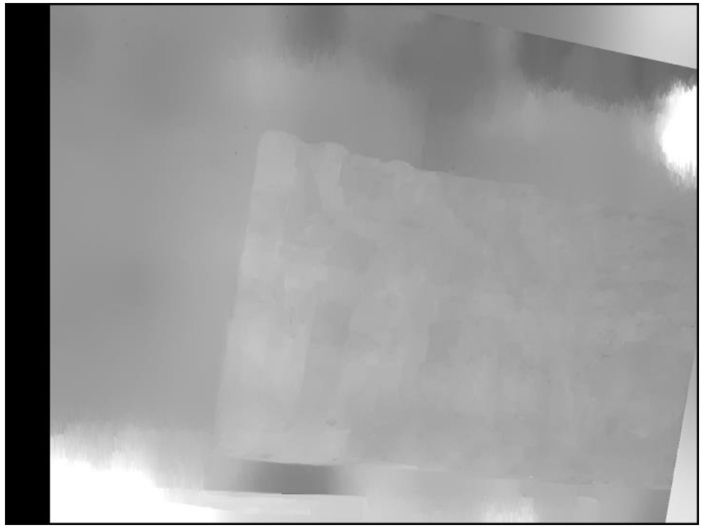


TUM Dataset

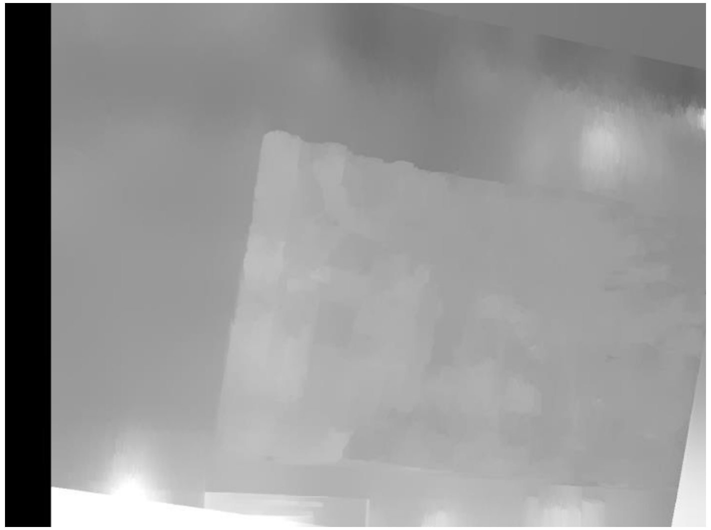
KITTI Raw

Generated Disparity Map with Post Filtering

BM



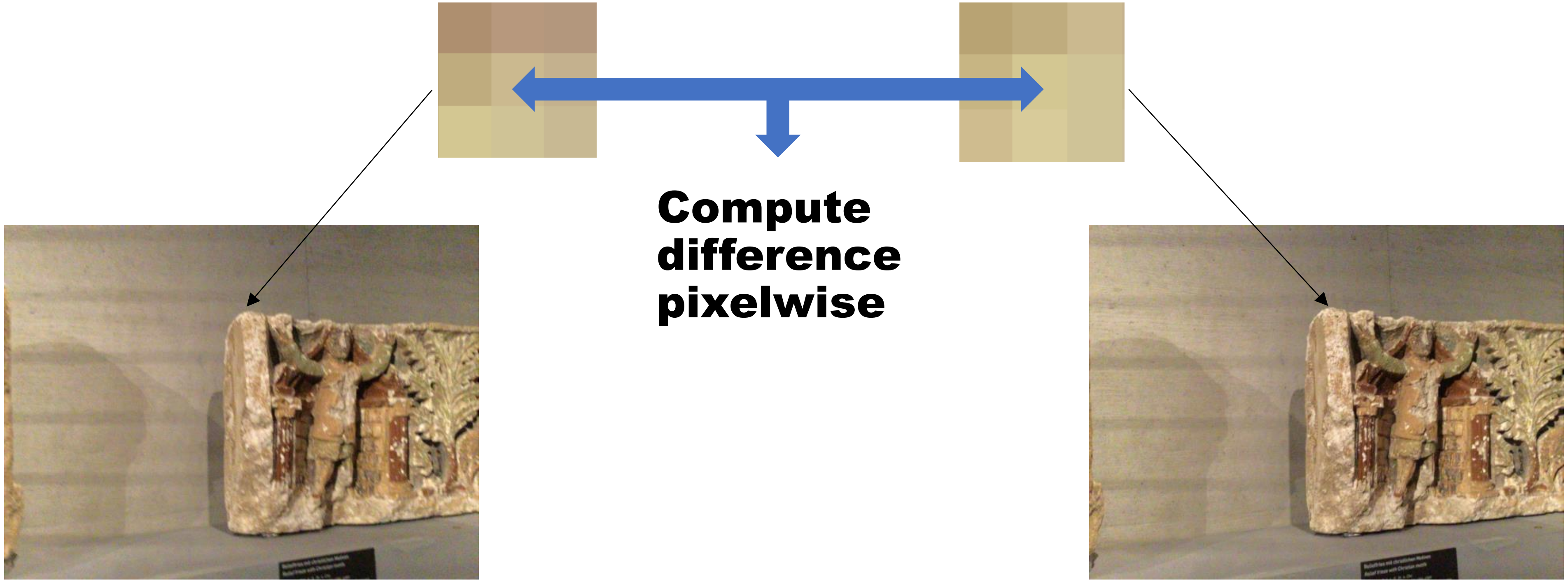
SGBM



TUM Dataset

KITTI Raw

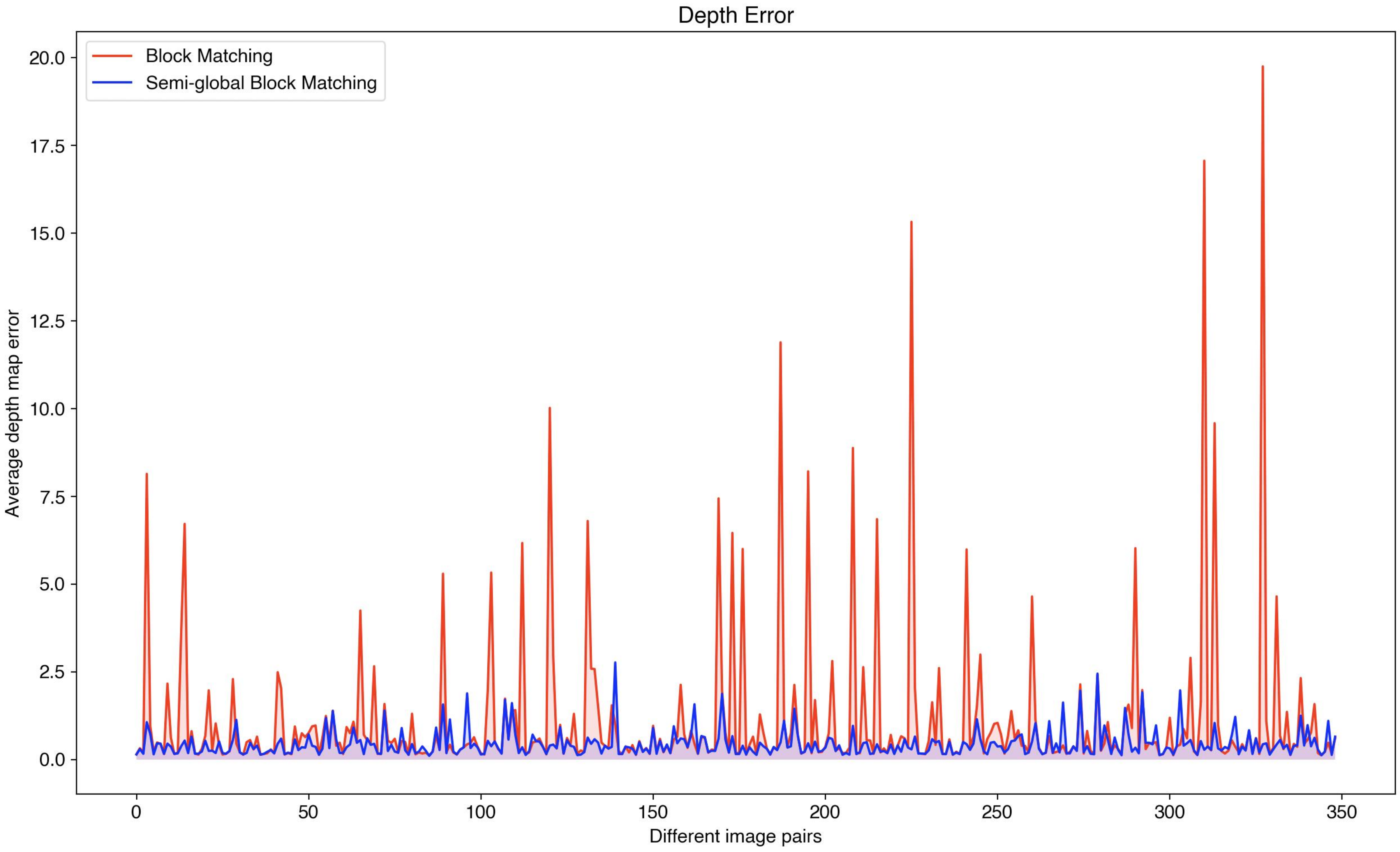
Dense Matching Method



Generated depth map **Ground truth depth map**

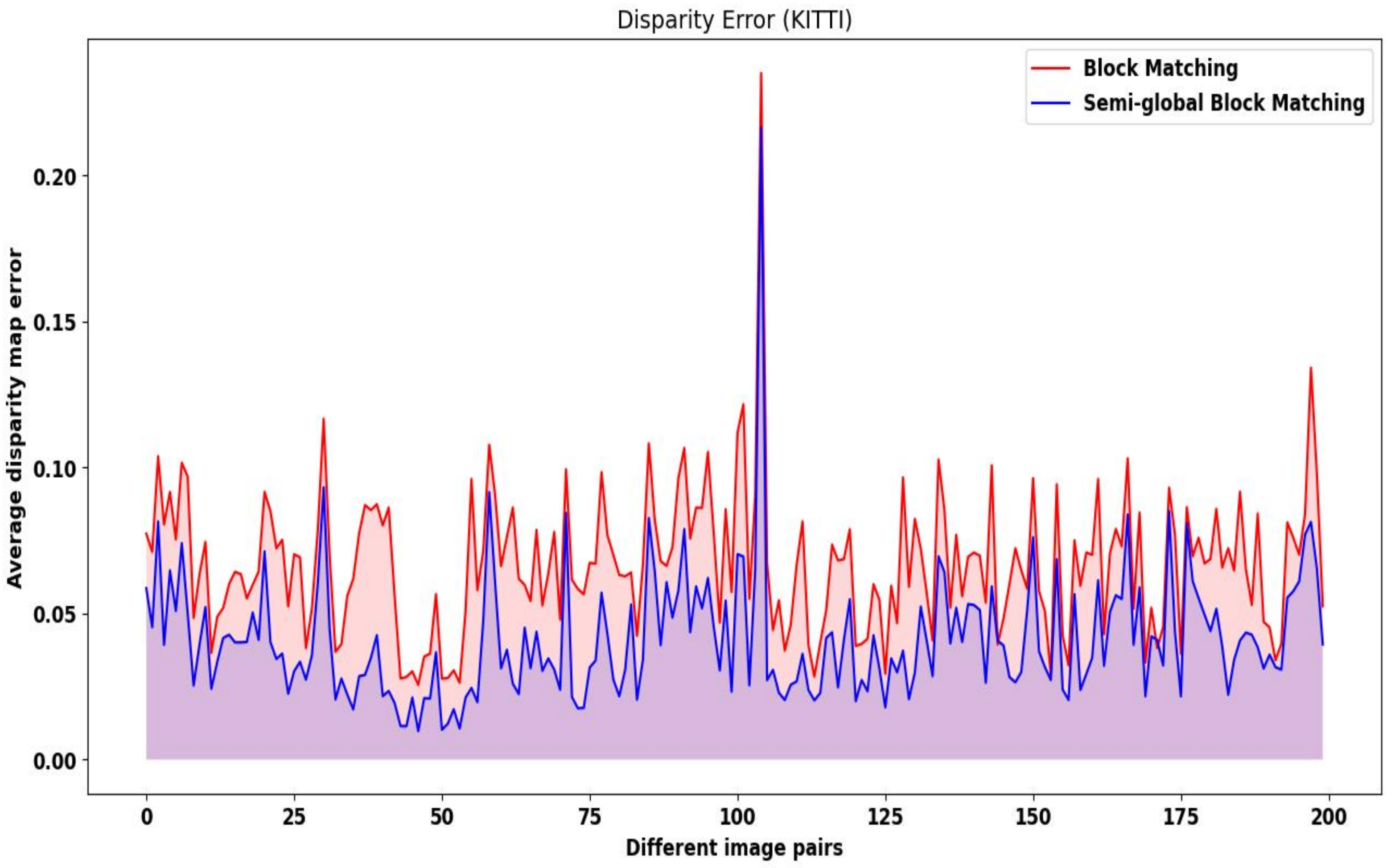
$$\text{Error} = \frac{\text{count}(d_{ij} - d'_{ij} > 3px \ \& \ d_{ij} - d'_{ij} > 0.05)}{\text{total valid pixels}}$$

Dense Matching Method



 # the comparison is between the ground truth depth map and the calculated depth map

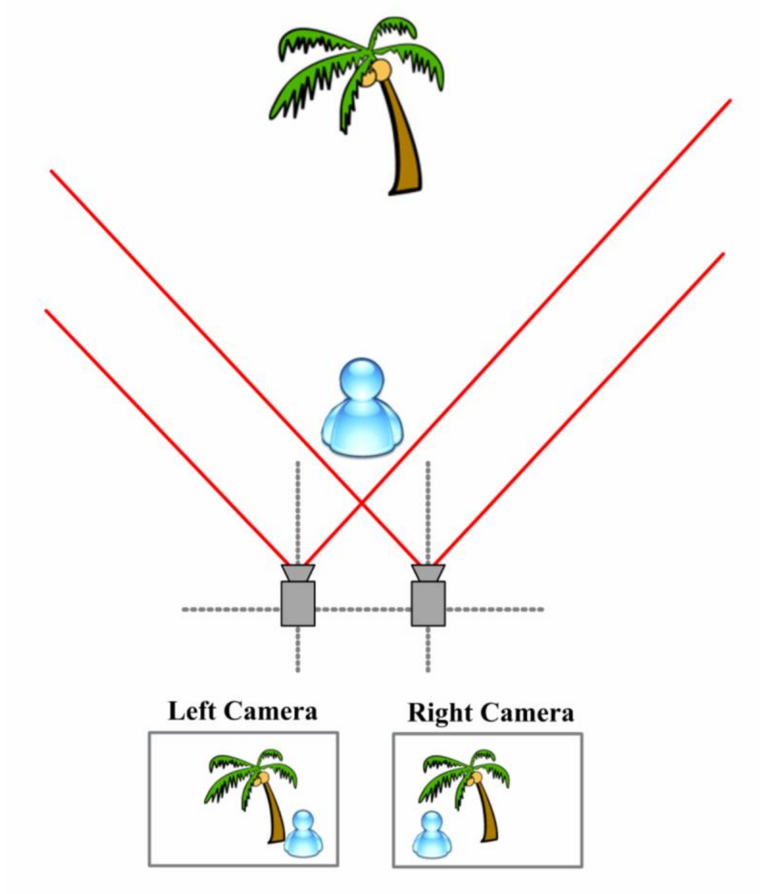
Dense Matching Method



 # the comparison is between the ground truth disparity map and the calculated disparity map

3D Model Generation

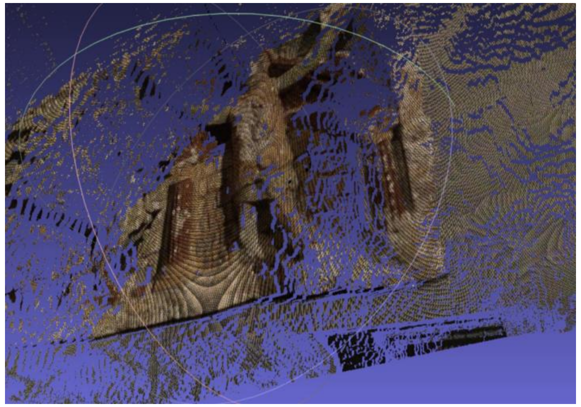
Disparity map to point cloud

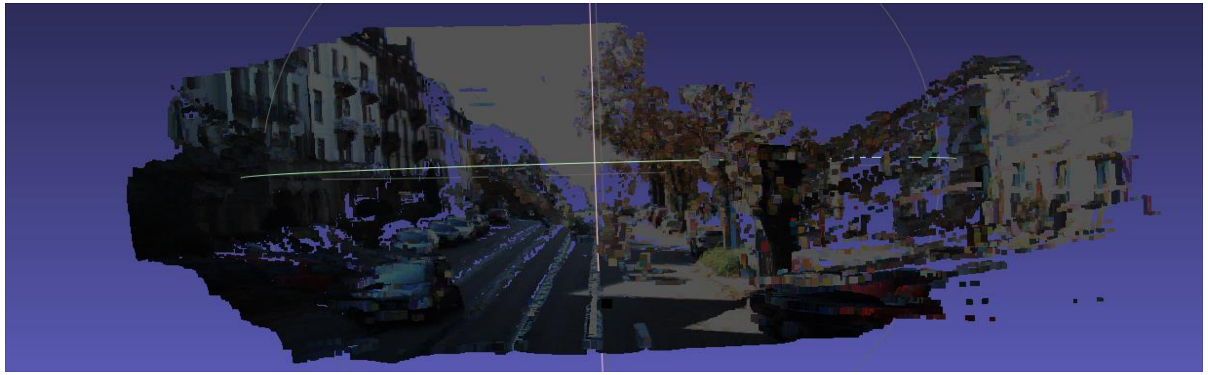


Disparity map to 3D model

Disparity Map

3D Model

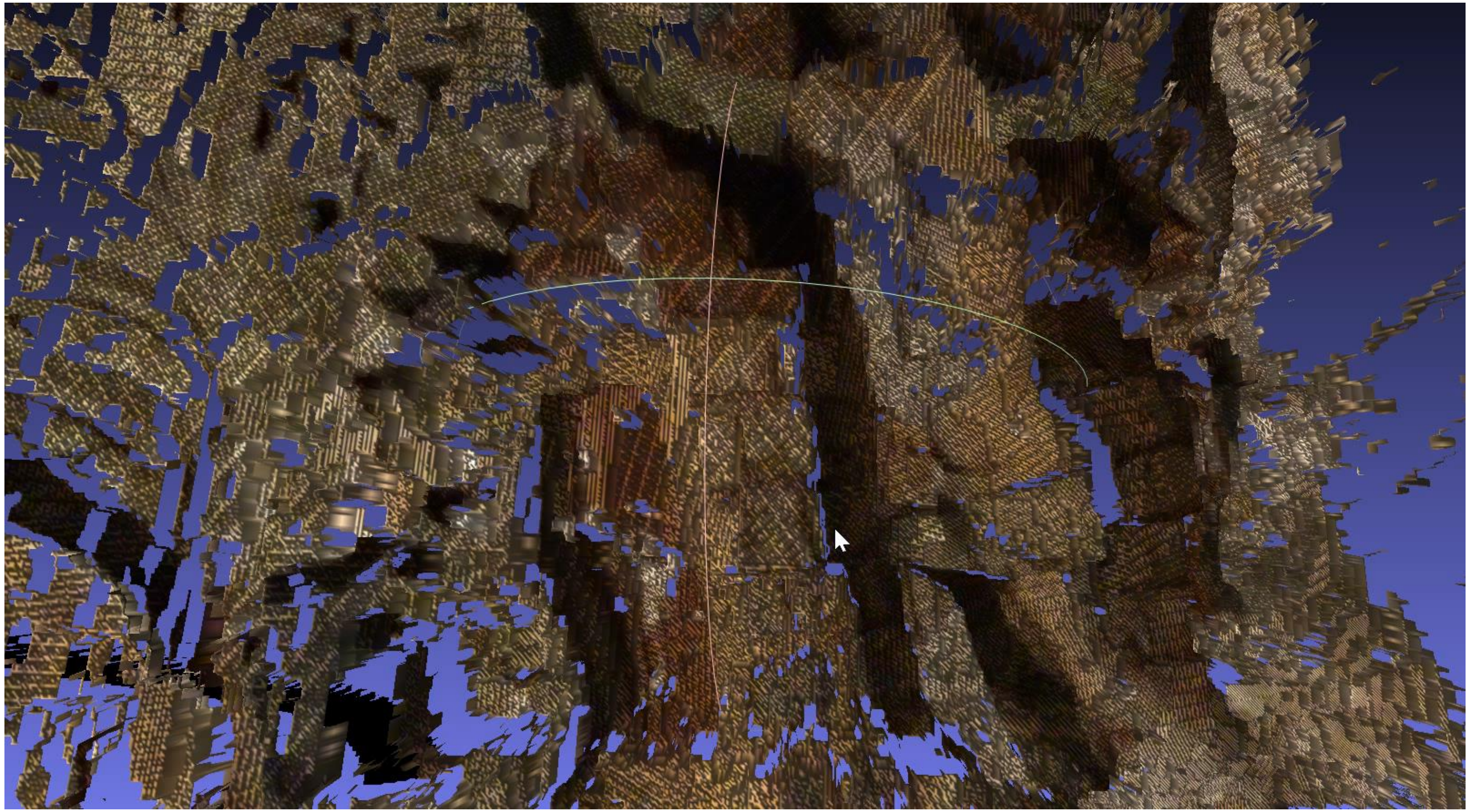




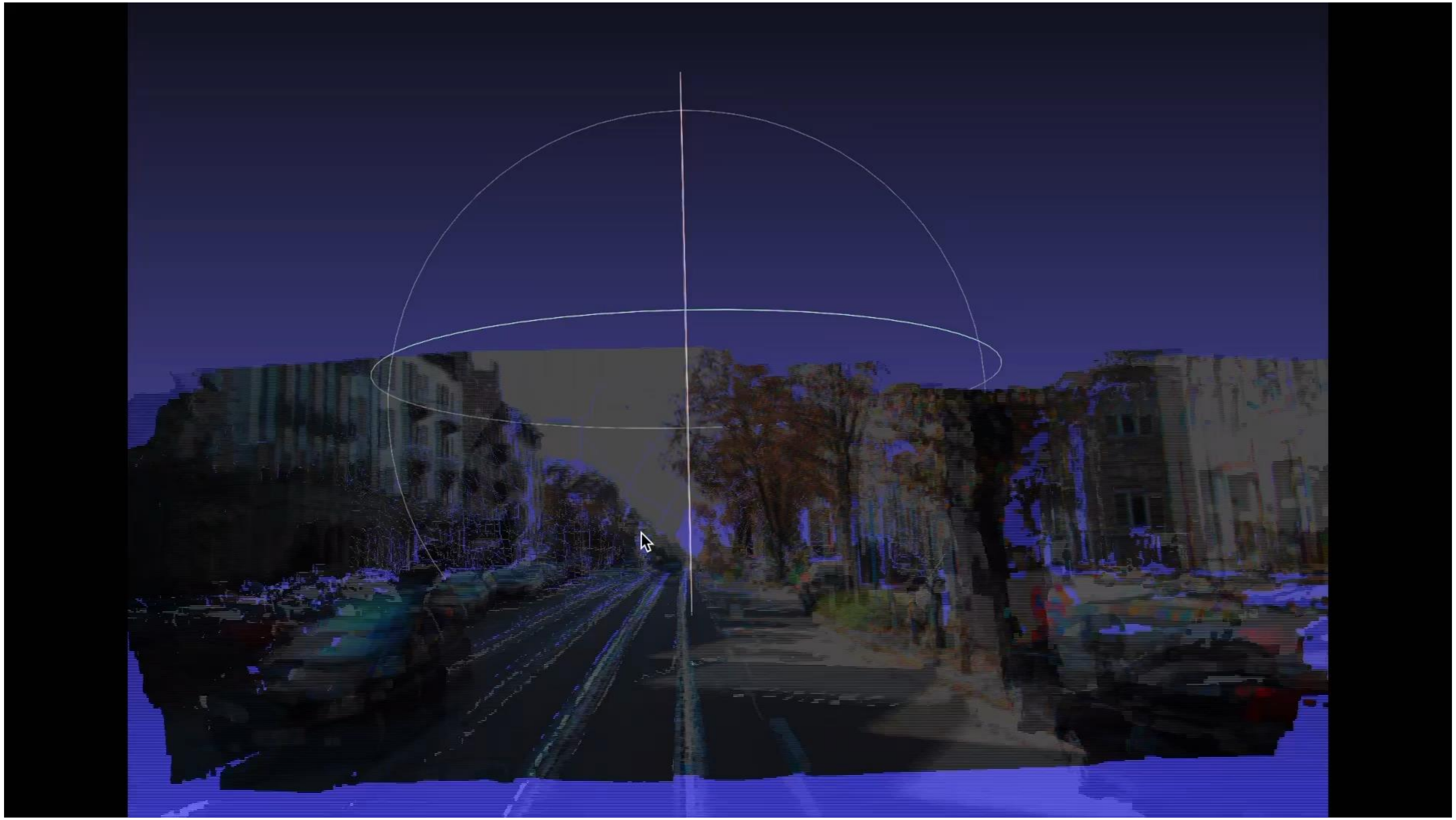
TUM Dataset

KITTI Raw

Generated Mesh (TUM Intrinsic3D)



Generated Point Cloud (KITTI)



Conclusion & Discussion

Conclusion

1. Implemented stereo reconstruction pipeline

2. Key steps such as calibration, keypoint calculation, and 3D model construction

3. Results show the ease and efficiency of reconstructing 3D models from stereo images

4. Generated 3D models have potential applications in scene understanding, object recognition, and navigation

Discussion

1.The project serves as a starting point for further research and development

2.Possibility of incorporating ICP methods and other data for improved 3D models

3.Potential for exploring new applications in fields such as robotics and autonomous vehicles

4.Opportunity to improve the accuracy and robustness of 3D models

5.Possibility of integrating new algorithms and techniques for enhanced performance