3D Scene Reconstruction Using Multiple 2D Images

02504 Computer Vision - Spring 21

s201850 - Anubhav Goel, **s206641** - Madhur Sudarshan, **s205912** - Lucia Montesinos

Introduction and Motivation

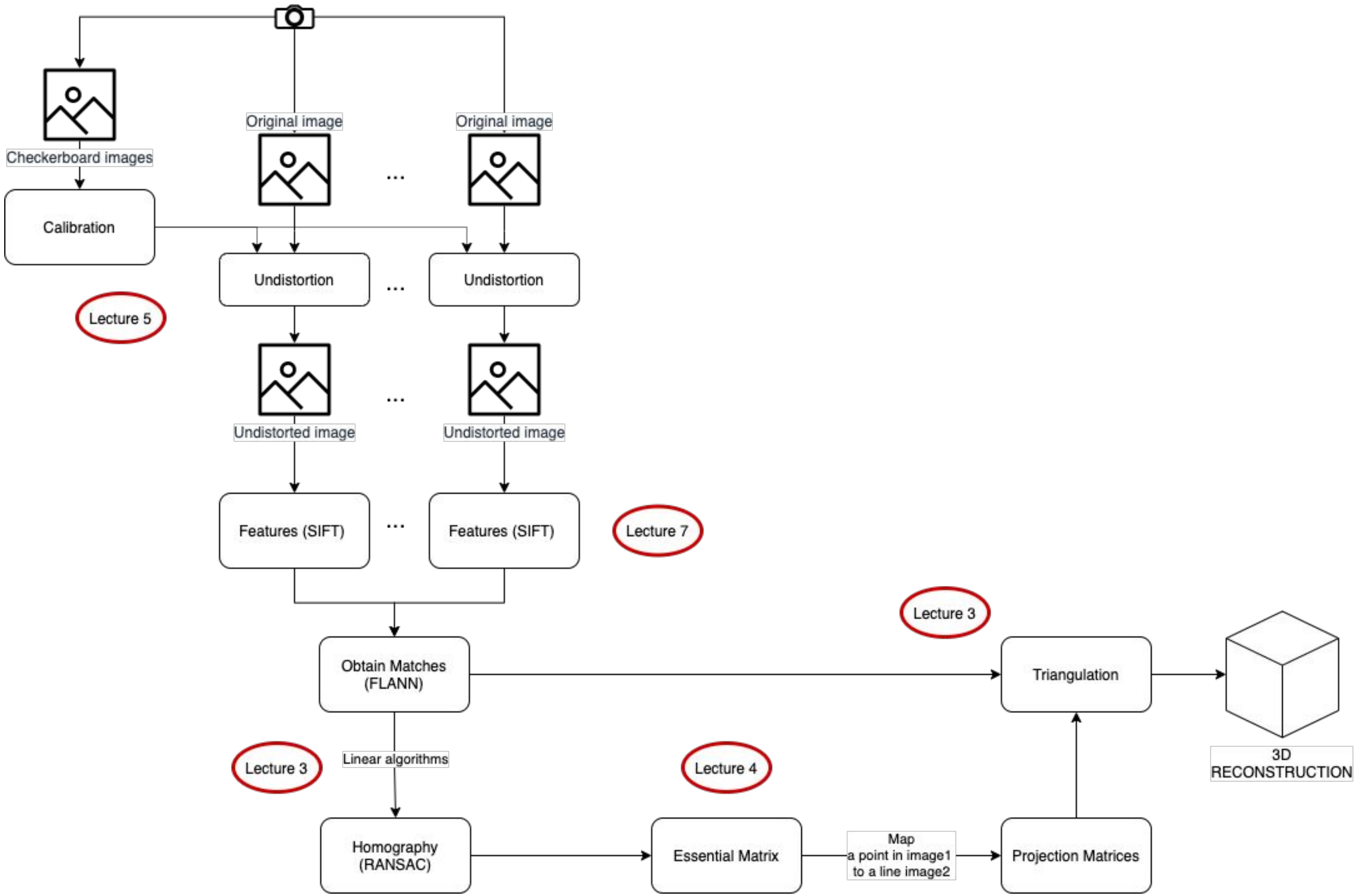
What 3D reconstruction can be thought of as the process of capturing the shape and appearance of real objects.

Why Important demand and core technology for numerous fields.

Many existing systems for reconstructing 3D models are built around specialized hardware resulting in a high cost and they may not also satisfy the requirements of the systems

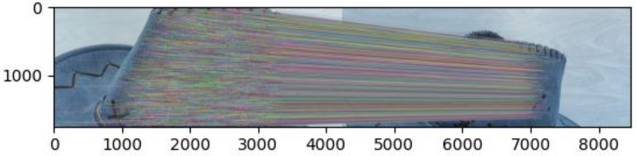
How By taking multiple images of the same scene, identifying and matching features between them, triangulating and obtaining a sparse or point 3D reconstruction of the object

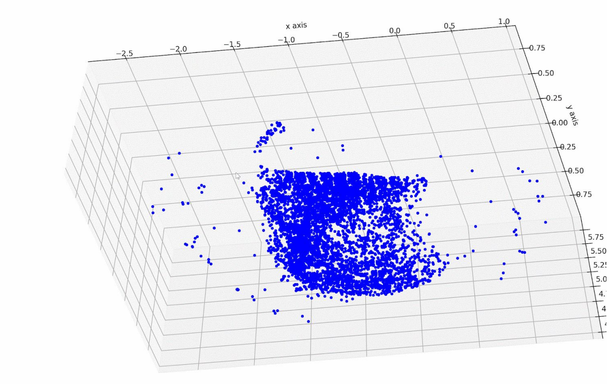
Pipeline

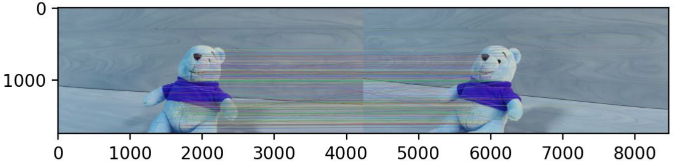


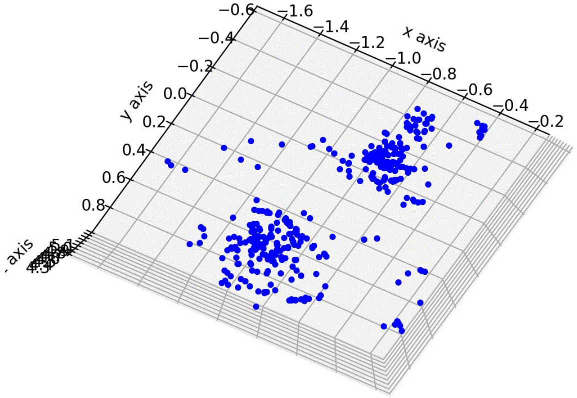
Checkerboard images
Original image
Original image
...
Calibration
Undistortion
...
Undistortion
Lecture 5
Undistorted image
Undistorted image
...
Features (SIFT)
...
Features (SIFT)
Lecture 7
Lecture 3
Obtain Matches (FLANN)
Triangulation
3D RECONSTRUCTION
Lecture 3
Linear algorithms
Lecture 4
Homography (RANSAC)
Essential Matrix
Map a point in image1 to a line image2
Projection Matrices

Results: 3D Reconstruction of a Hat and a Plush Toy





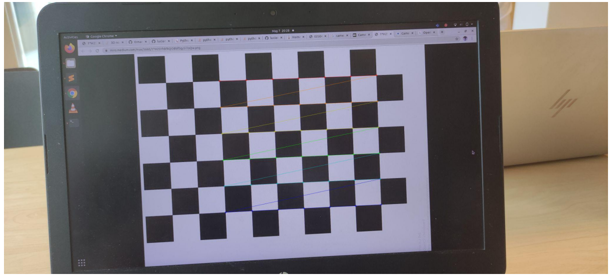




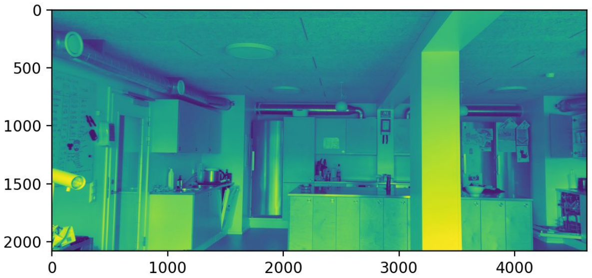
Problems faced: Camera Calibration

We used a phone camera (Poco F2 pro) and used images of a checkerboard to calibrate the camera and obtain the camera matrix

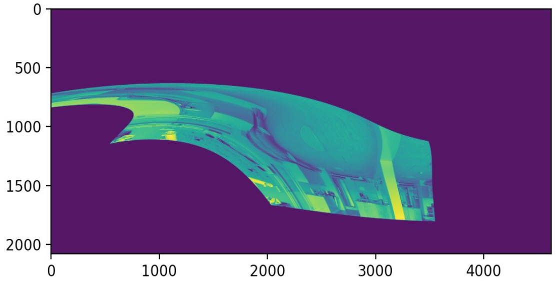
We initially use pictures of a checkerboard printed out on a paper but we eventually switch to images taken from a laptop screen as the former gives unsatisfactory results



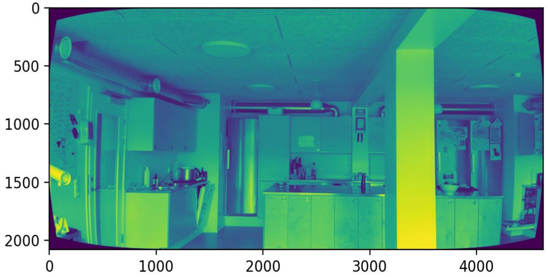
Example of image used to calibrate the camera



Original image



Poorly undistorted image



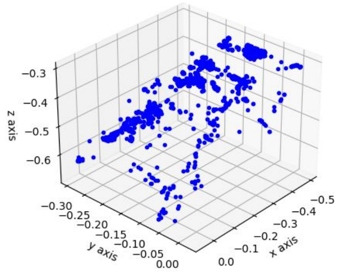
Correctly undistorted image

Problems faced: Dataset

We look at numerous datasets to evaluate the performance of our model across datasets as the choice of the dataset heavily influences the results of the reconstruction:

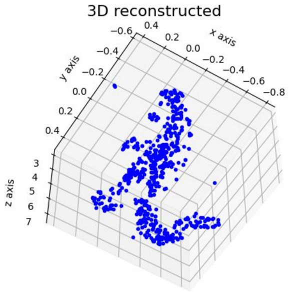
- Feature-detection does not work well on textured backgrounds or plain foregrounds → generates noise in the final reconstruction
- Objects with recognizable shapes are easier to recognize in a reconstruction

Sample of the initial dataset



Noisy 3D reconstruction from a pair of images from the initial dataset

Sample Dataset Image



3D reconstruction (with less noise)

Problems faced: Approach

We started with measurements of rotation and translation and calculation of the projection matrix manually

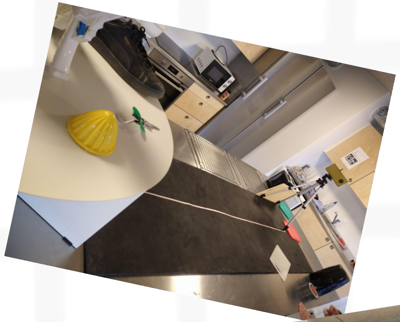
But we then switched to the essential matrix approach because

1. Inaccuracy in translation which we were measuring

2. Inability to measure rotation accurately and rotation is required to capture depth info sometimes that just translation might not be able to capture

3. Possibility of working with datasets without knowing the extrinsic camera parameters

Thank You



References

https://web.stanford.edu/class/cs231a/prev_projects_2016/xuerong_ltao2_report.pdf

https://docs.opencv.org/master/d6/d00/tutorial_py_root.html

https://github.com/alyssaq/3Dreconstruction

https://github.com/filchy/slam-python

https://www.robots.ox.ac.uk/~vgg/data/mview/

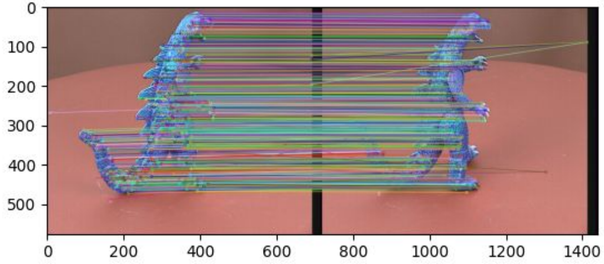
Appendix I: Future work

- Use of multiple data points obtained from more than one pair of images of the same objects to denoise the results
- Use of clustering algorithms (K-means clustering) to obtain a single 3D point instead of obtaining point clouds
- This also helps get rid of outliers that are generated due to random textures sometime present in the background
- Use of CNNs as an end-to-end pipeline to perform the entire task of predicting 3D coordinates from the multiple 2D images

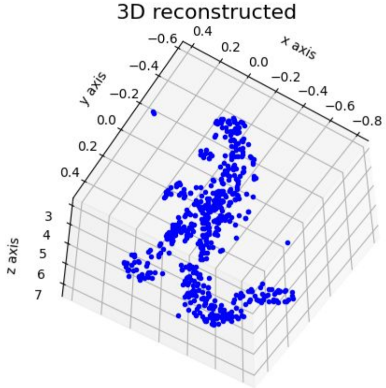
Appendix III: Dataset

- We move on to using the Dinosaur dataset which contains 36 images of a toy dinosaur taken from different which has a clear foreground with high texture (several corners) and a plain background
- Since we are not aware of the camera rotation and translation values from which the images were taken, we make use of the essential matrix approach
- We get definite improvements in our results with this dataset

Sample Dataset Image



A matching shown between two images of the dataset



The result of the 3D reconstruction