SAPIENZA
UNIVERSITÀ DI ROMA

technische universität
dortmund

AI LAB: Computer Vision & NLP
Final Project

Ashwin Nedungadi
ashwin.nedungadi@tu-dortmund.de

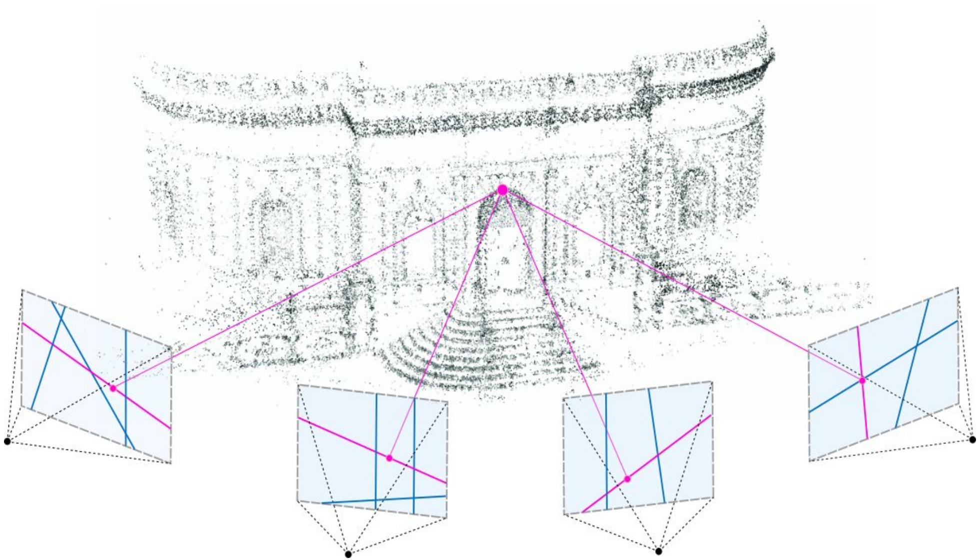
3D Reconstruction of Roman Artifacts Using Structure from Motion (SfM)

Main Contributions

- Design an SfM pipeline for 3D reconstruction of Artifacts.
- Reconstruct famous sculptures & artifacts from Rome
- Demonstrate the feasibility of this technology with real world results
- Open source the work and dataset
- Code and dataset will be available on Github

Broad Use Cases of SfM

- Robotics - for localization and mapping of environments (eg. Self-Driving Cars)
- Archeology - for mapping archaeological sites and documenting artifacts.
- Architecture - for mapping buildings, excavation sites or as-built.
- Space Exploration - Reconstructing geology using rovers (eg. Mars)



SfM on a building

Reconstructed rock from images by Mars 2020 Rover

Motivation

- SfM is cheap and efficient as it only uses multi-view images from a camera (No need of expensive LIDARS)

- It can be implemented in real-time for robotic applications given the right hardware.

- SfM is becoming more feasible in the last decade due to better GPUs and computational power.

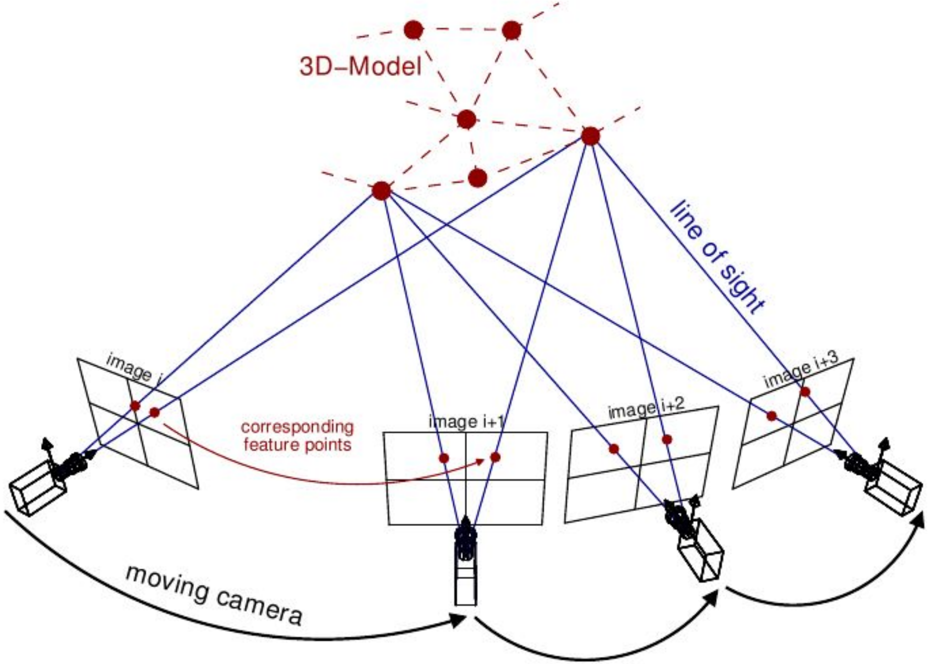
- Reconstructing artifacts is just one of the use cases of SfM, In Rome, there are several artifacts and sculptures to experiment with i.e. Plenty of data sources.

- Having a 3D model of historical artifacts enable researchers from across the world to examine and study them (possibly even in VR).

- It is extremely interesting!

Introduction

Structure from motion is a low cost method that can be used to reconstruct 3D objects just utilizing the movement of the camera i.e. taking photos from multiple views.



The Problem

A camera is represented by a matrix P which transforms a point X in the real world to a pixel x in the image. This relationship is given by the following equation:

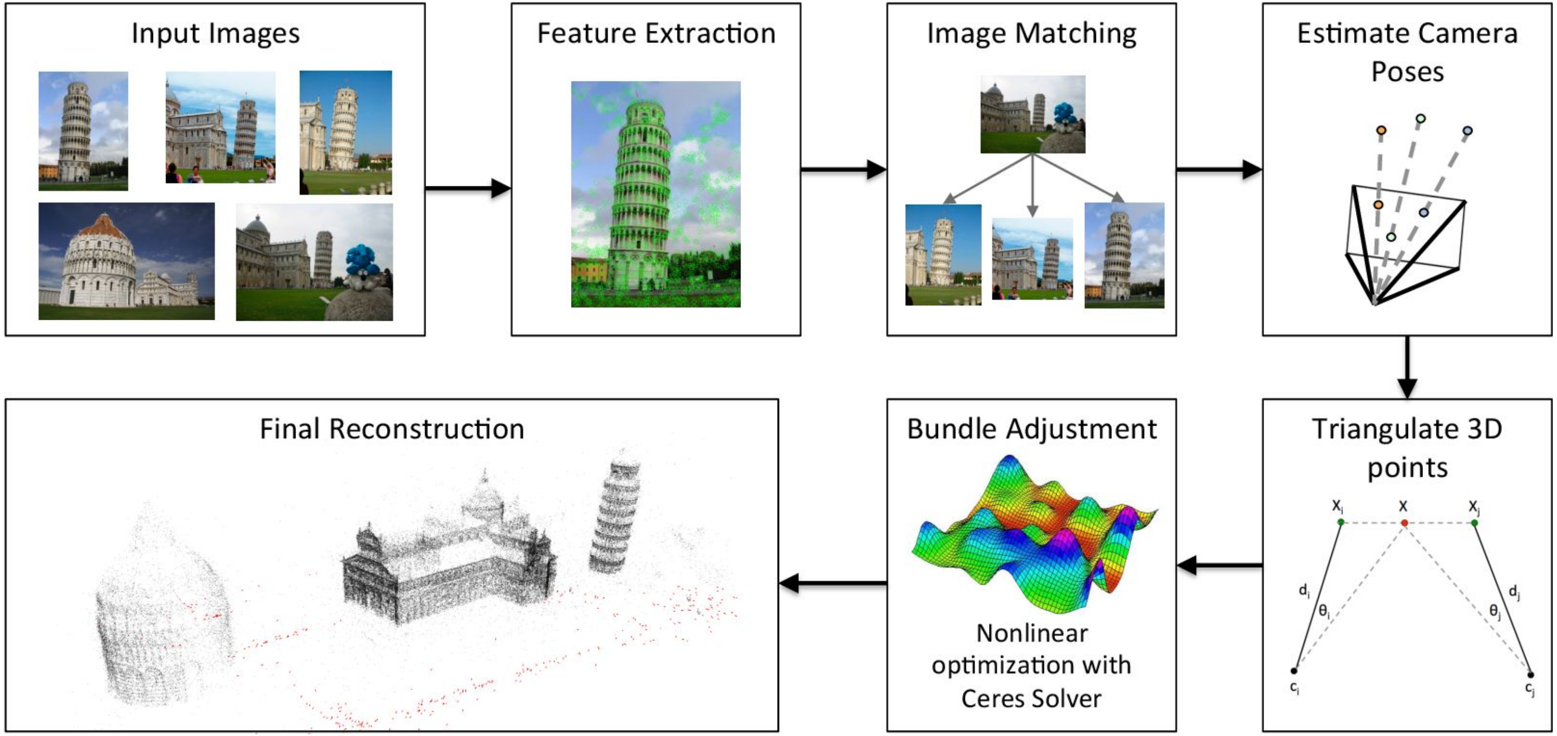
$$x = PX$$

$$X = P^{-1}x$$

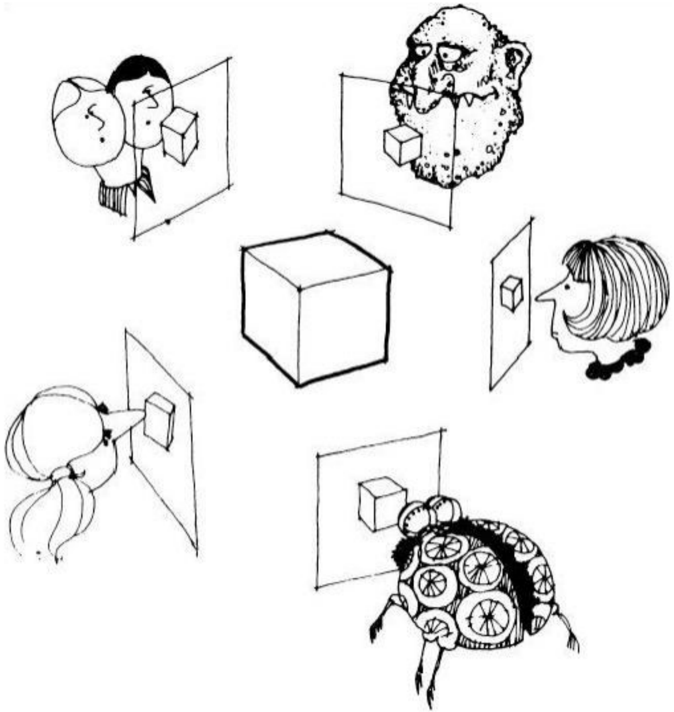
Therefore, finding the camera matrix or the camera pose for each of the images (the 2D snapshots or views) is a part of the process of reconstructing 3D coordinates. SfM, thus, involves estimating the 3D points along with the camera pose from a sequence of images.

The Problem

The SfM problem can be broken down into several subprocesses:



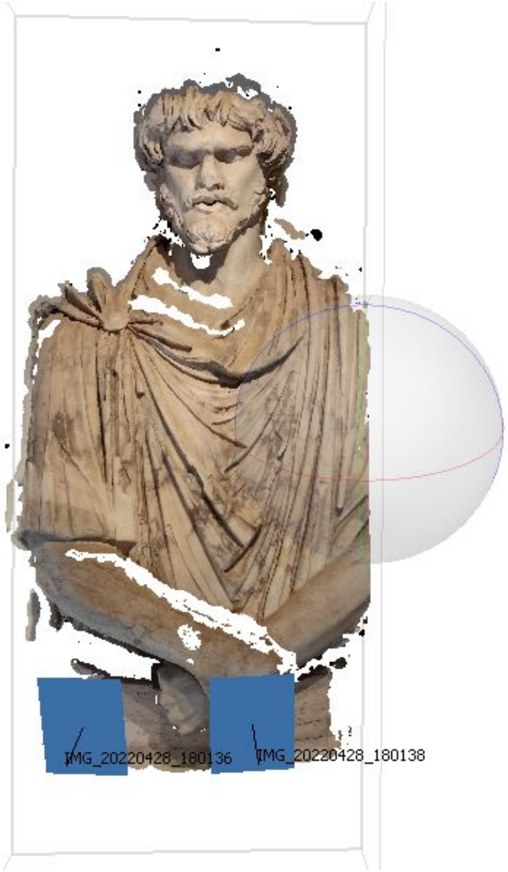
Types of SfM Techniques



Incremental SfM: Incremental SfM is the standard approach that adds on one image at a time to grow the reconstruction.
While this method is robust, it is not scalable because it requires repeated operations of expensive bundle adjustment.
We will use incremental SfM for our task as we do not have a large image input and we can reconstruct an artifact with a maximum of 8 images.

Global SfM: Global SfM is different from incremental SfM in that it considers the entire view graph at the same time instead of incrementally adding more and more images to the Reconstruction. Global SfM methods have been proven to be very fast with comparable or better accuracy to incremental SfM approaches and they are much more readily parallelized.

Types of SfM Techniques

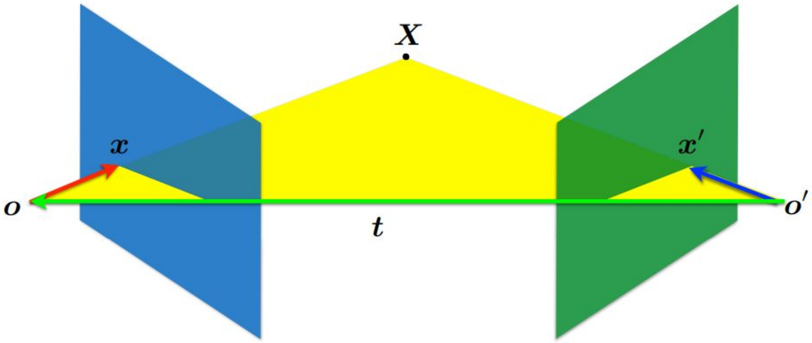


Two-View SfM: Two-view SfM, which means SfM from a pair of images, is the foundation of incremental SfM methods and is quite useful in real applications due to its simplicity and ease of implementation.

The 2-view algorithm aims to get the relative camera motion and 3D structure from two images

Multi-View SfM: Multi-View SfM considers more than two images at a time and can incrementally generate the point cloud or perform a global reconstruction. This type of SfM is best suited for large datasets and requires a lot of computation.

Essential Matrix

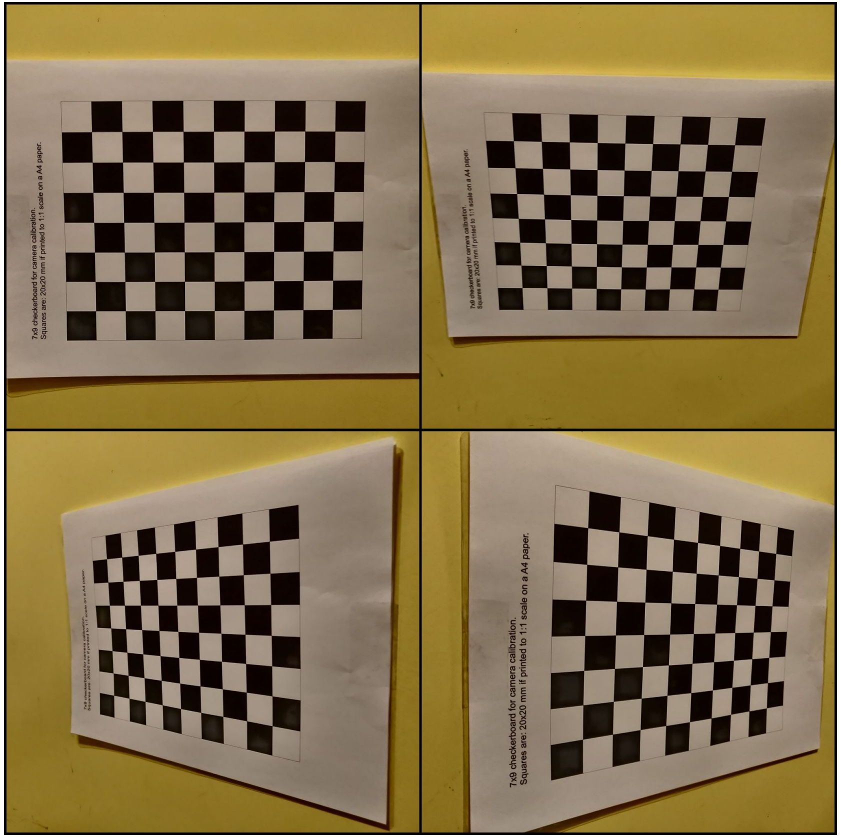


The essential matrix is a 3x3 matrix that relates corresponding points in two images (i.e. encodes epipolar geometry).

Given a point in one image, multiplying by the essential matrix will tell us the epipolar line in the second view.

In MATLAB, we can calibrate the camera using the "Camera Calibrator App" and estimate the essential matrix in the process.

Camera Calibration

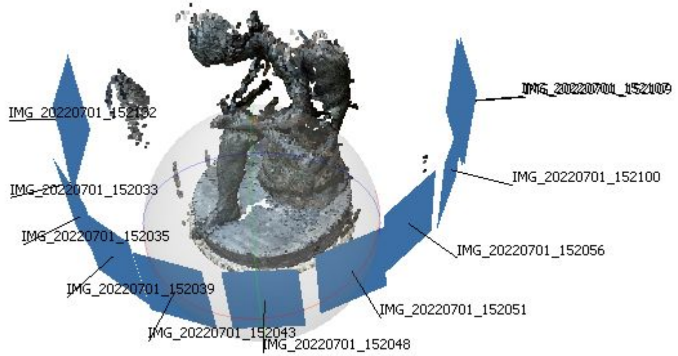


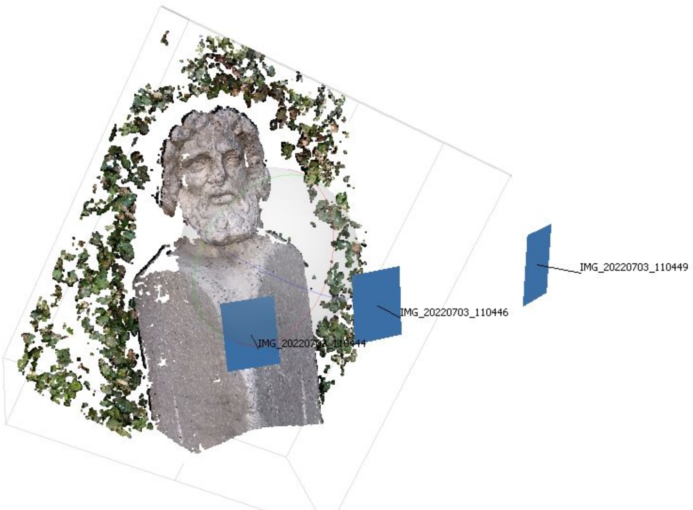
In order to estimate the Essential Matrix **(E)**, we need to calibrate our cameras using a checkerboard pattern with known geometry.

The checkerboard geometry is known and easy to detect. The squares used here are 7x9 and 20mm if printed 1:1 on an A4 paper.

Once we run the camera calibration, we obtain the camera intrinsic matrix and distortion parameters which tell us the relationship of a single image pixel w.r.t a 3D point in the world.

Image Acquisition





Images are acquired with a oneplus 6 mobile smartphone which has a 20 MP sensor.

Images were taken in a consecutive order with slight movement between images resulting in maximal overlap.

After an initial set of images, the camera angles were changed for multi-view images.

It is important to pay attention to the background of the object being imaged as noise is an issue for SfM. Hence, objects with a clear white background was selected for 3D reconstruction.

Around 8 images were taken of each object.

Real World Example

[Statue di Daci](#) - Palatine Museum Collections, Rome, Italy

Multi-View images acquired from a Oneplus6 smartphone.

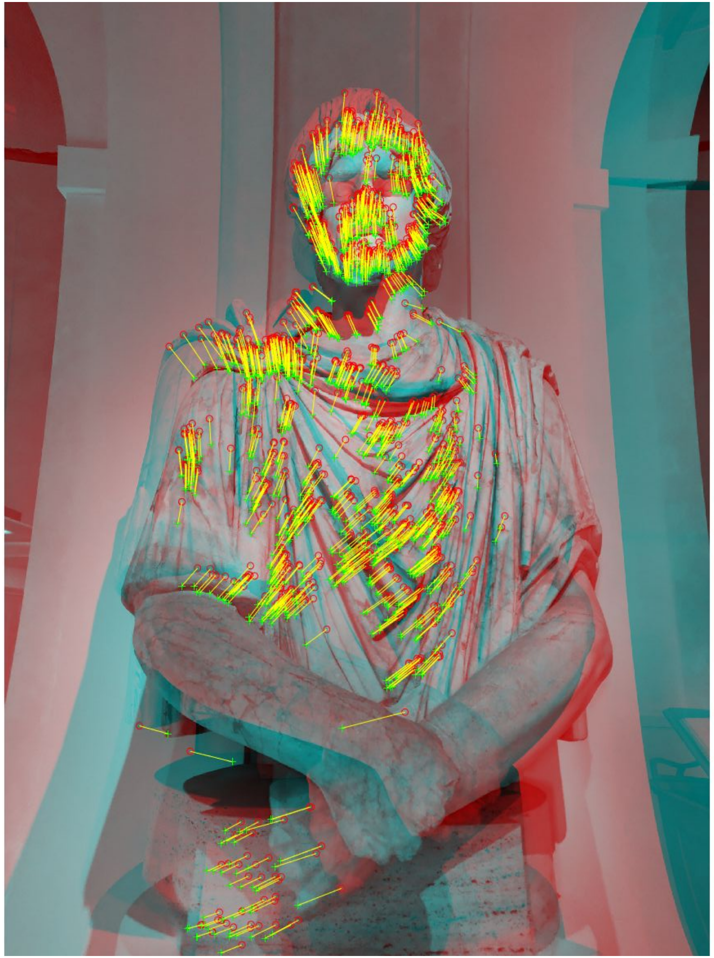
Steps in the SfM Problem - Feature Extraction

For each image pair given as input to the pipeline, a collection of local features is created to describe the points of interest of the image (**key points**).

For feature extraction, different solutions can be used, the choice of the algorithm influences the robustness of the features and the efficiency of the matching phase.

For this particular task, a **SURF (speeded up robust features)** descriptor is used which is based on the famous SIFT method.

Steps in the SfM Problem - Feature Matching



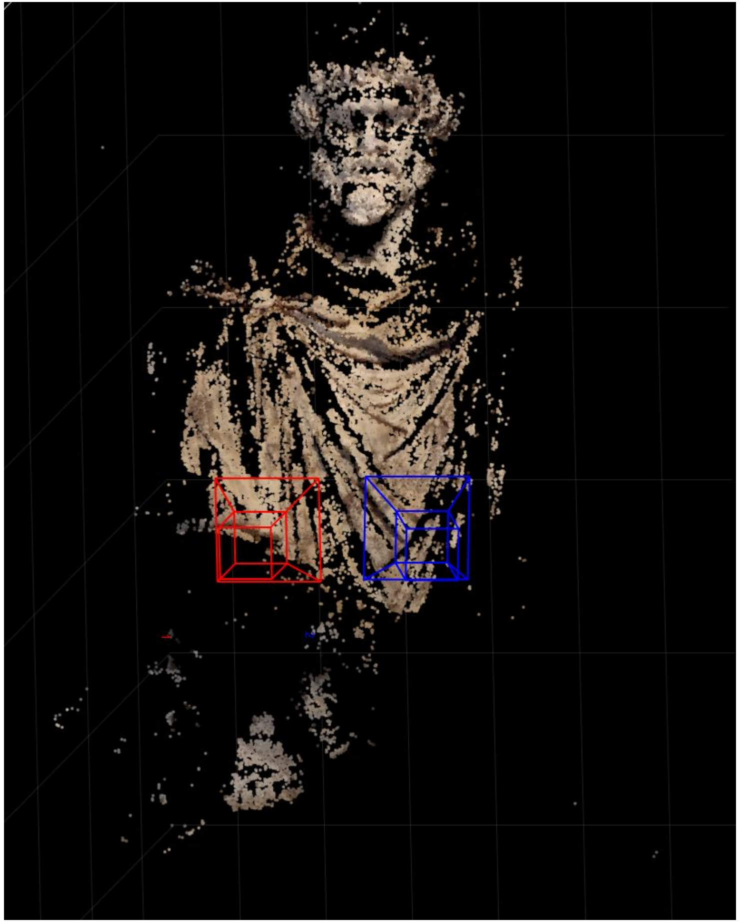
Feature matching is the process of finding corresponding features from two similar datasets based on a search distance.

The key points and features obtained through feature extraction are used to determine which images portray common parts of the scene and are therefore at least partially overlapping.

The output of this process is a set of images overlapping at least in pairs and the set of correspondences between the features.

Here red and blue highlight shows the two overlapping images and the yellow streaks show the tracked features.

Steps in the SfM Problem - Image Registration & Triangulation

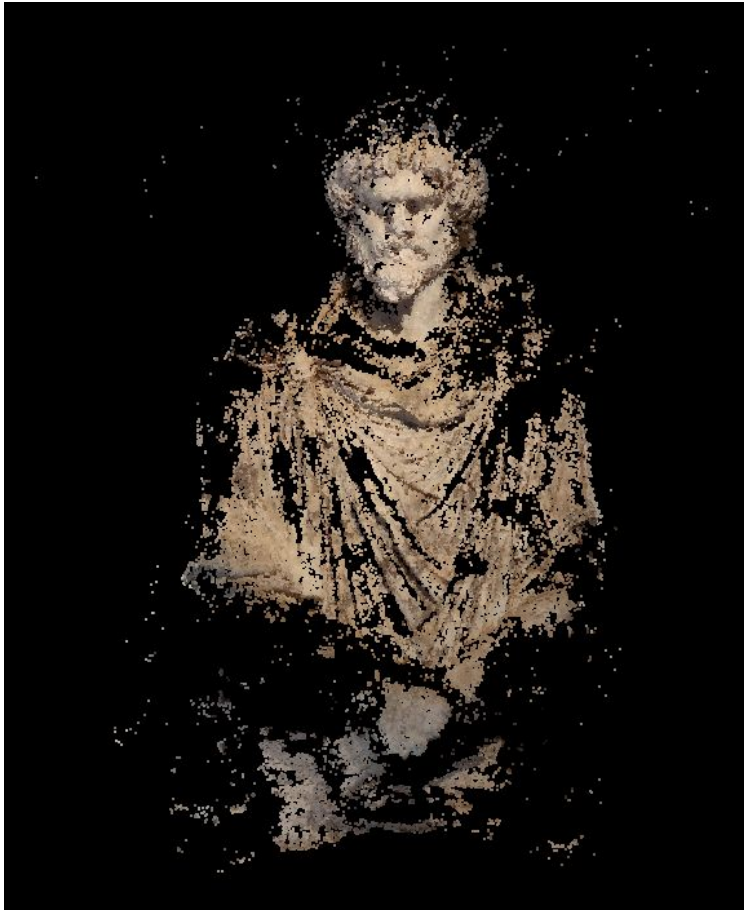


A **triangulation** process is used to define the 3D coordinates of the new points that can be added to the reconstruction and thus generate a more dense point cloud.

The triangulation problem takes a pair of registered images with points in common and the estimate of the respective **camera poses**; then it tries to estimate the 3D coordinates of each point in common between the two images.

However, there may be inaccuracies caused by the previous phase of the pipeline and the points are not completely aligned, we call this the reprojection error which can be solved by techniques such as **bundle adjustment**.

Steps in the SfM Problem - Alignment & Refinement

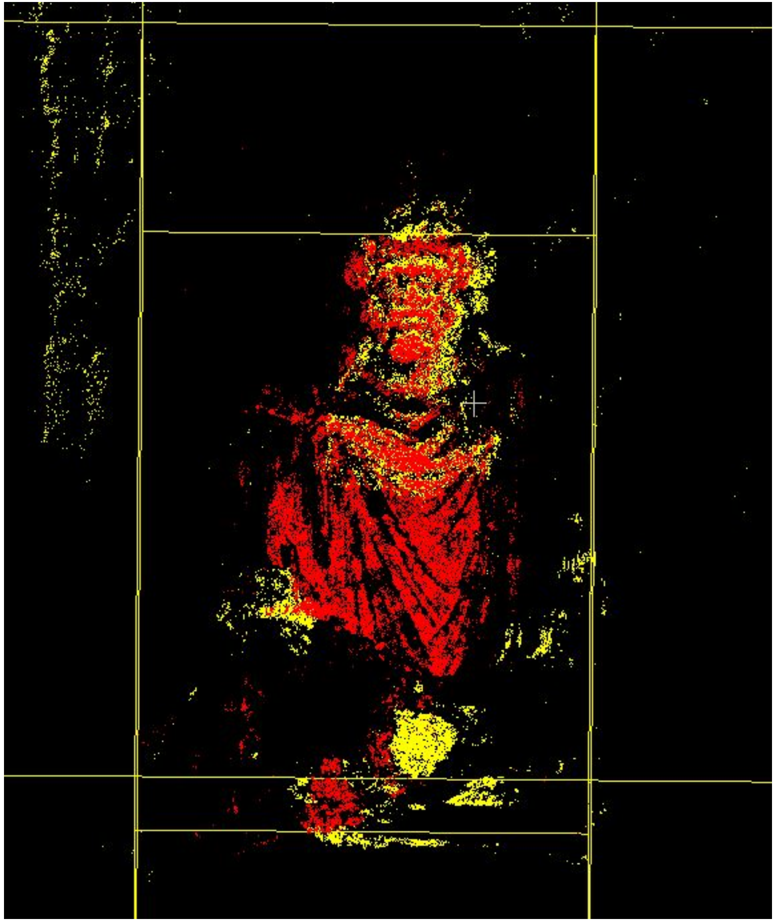


Since we are reconstructing the point clouds only from two images (**two-view SfM**), we need to repeat this process for all image pairs in our dataset.

The resulting point clouds will be partial reconstructions of the final structure which then needs to be filtered, aligned and then stitched.

For this, we will use a process called **Iterative Closest Point (ICP)**. Which is an algorithm employed to minimize the difference between two clouds of points.

Final Step: Iterative Closest Point (ICP)



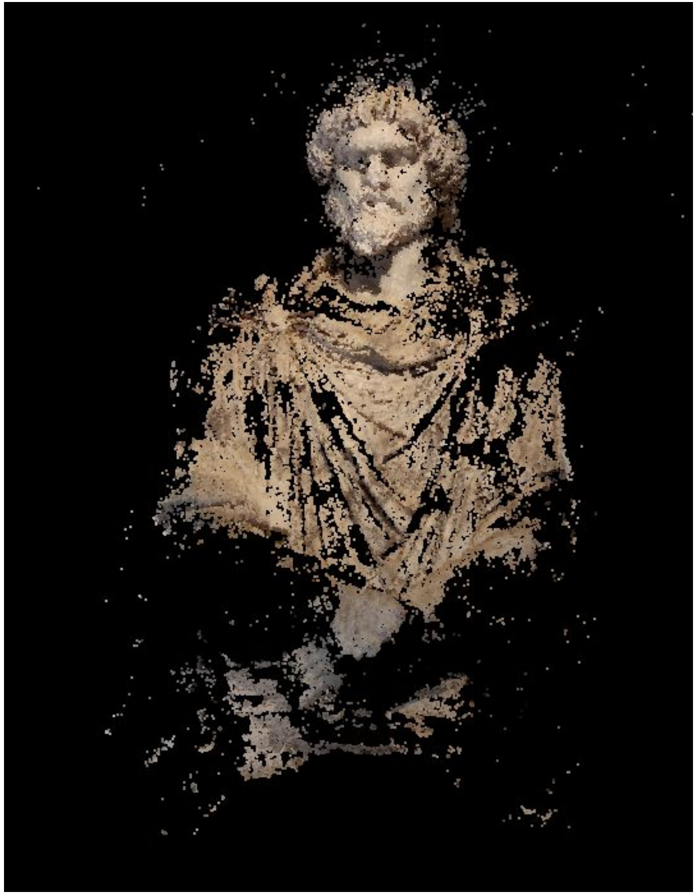
The final step is done in Cloud Compare and requires filtering out the noisy pointcloud with a **statistical outlier removal** (SOR) filter and segmenting out the noisy points.

Then, ICP is performed to align the two pointclouds into a single final pointcloud.

The ICP Algorithm does the following.

1. For each point in the source point cloud (red), match the closest point in the reference point cloud (yellow).
2. Estimate the combination of rotation and translation using a **root mean square** point to point distance metric minimization technique.
3. Transform the source points using the calculated transformation.
4. Iterate (re-associate the points) until a suitable rms error is reached or the pointcloud looks visually stitched.

Final Result



The final result is obtained by combining 4 different pointclouds obtained from using SfM on 4 different **image pairs**.

The pointcloud pairs are aligned using ICP and combined and results in the final pointcloud shown here.

The pointcloud still contains some missing areas. This problem can be solved by taking more overlapping images which will result in a more complete pointcloud.

A total of 8 images capture from a mobile phone camera were used for this statue.

The resulting pointcloud using 30 images from a DSLR camera would be of significantly higher quality.

One Step Further: Meshing & Texturing in Agisoft

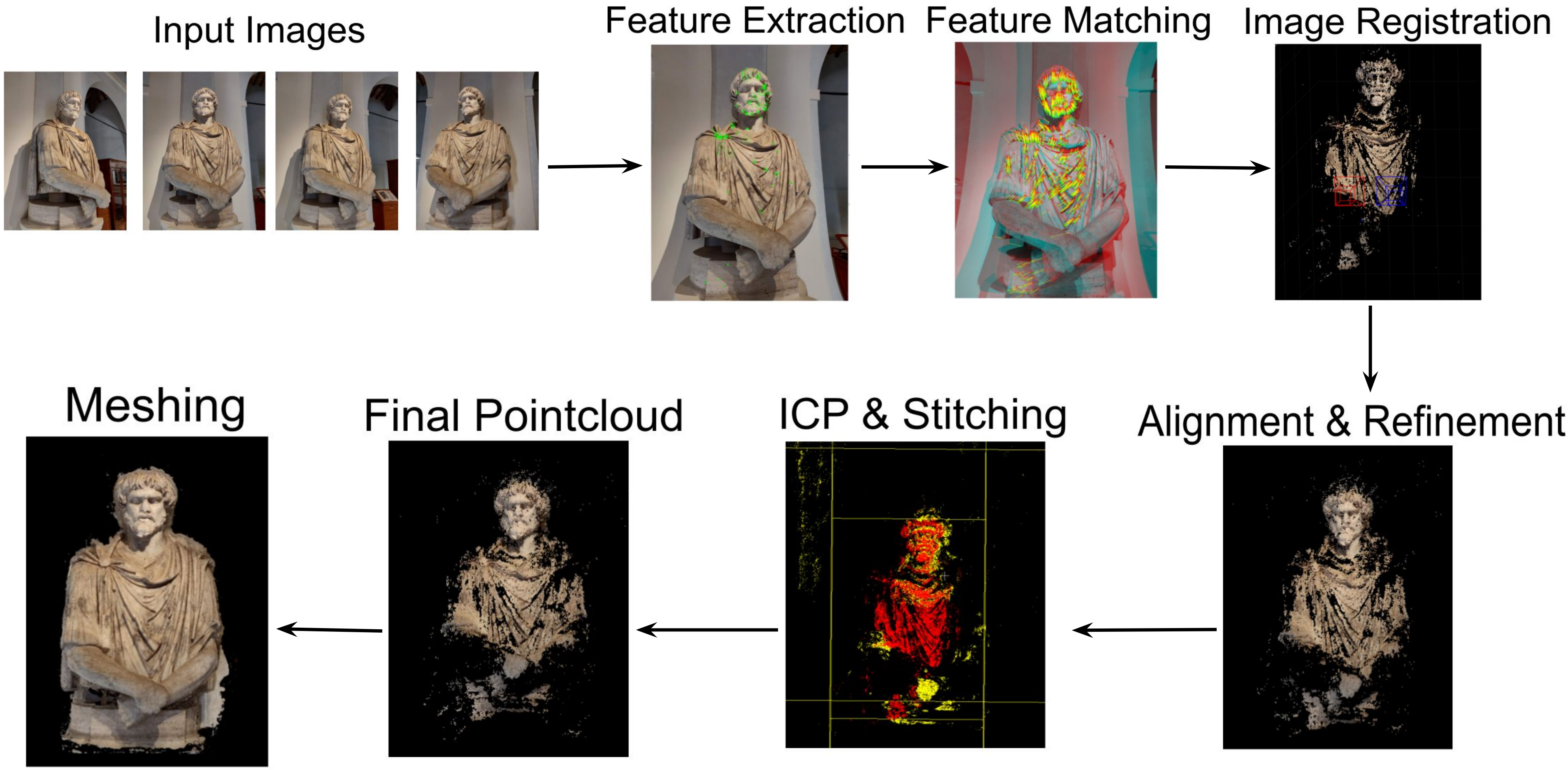
For a 3D model to be useful, we need to have it needs to have a mesh and be textured correctly.

Meshing is a technique which creates polygons from points. It results in a surface and a more complete 3D structure.

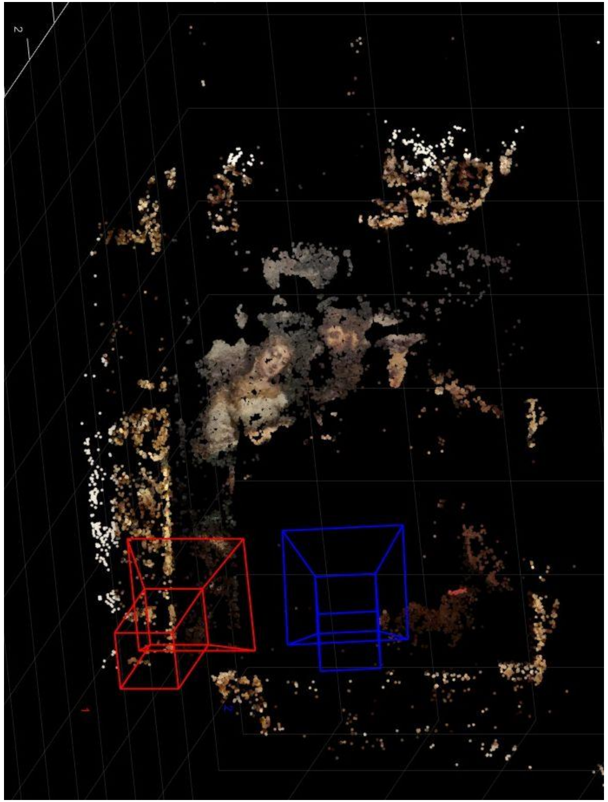
For this we will use a pre-existing software solution called agisoft metashape.

Agisoft is a very powerful tool as it can compute all of the previous discussed process in the click of a button, and hence it's suitable for non-computer scientists who just want the job done.

Complete SfM Pipeline



Matlab Vs. Agisoft



Two-view SfM using MATLAB Code

Global SfM from Agisoft

An example of why Agisoft could be more robust than programming based solution due to its ability to use different SfM pipelines to reconstruct objects with low features.

This renaissance era painting contains a very low amount of features since it’s very two dimensional.
Hence a simple two-view SfM algorithm is not sufficient for it’s reconstruction.

Fully Reconstructed Artifacts



Recommendations for Collecting Data

- Take incremental photos with overlap of the object to reconstruct.

- Avoid changing the perspective of the image (ie. changing zoom by stepping closer).

- Make sure the object is well lit and the background is clear or the algorithm will try to detect features from the background and this results in a noisy pointcloud and less tie-points on the main object.

- Higher quality camera results in higher quality pointclouds.

Softwares Used

Camera Calibration
Structure from Motion

Matlab

Iterative Closest Point Alignment
Filtering

Cloud Compare

Verification
Texturing

Agisoft Metashape

Note: Image acquisition was done completely with an android smartphone with a 12MP camera sensor (Oneplus 6).

Conclusion & Future Outlook

- We have successfully demonstrated a possible application of the Structure from Motion technique which is widely used in robotics.

- The SfM method has a wide range of applications including the possibility of scanning and documenting archeological sites and artifacts.

- Generating high quality digital 3D models for research can enable everyone to have equal access to historical artifacts and allow them to study it in detail.
  Moreover, the original versions can be preserved better and will undergo less destruction due to the changing environment.

- Digital 3D exhibits can also sometime in the future be part of a virtual museum, where tours can be given using VR headsets.

References

1. Brutto ML, Meli P. Computer Vision Tools for 3D Modelling in Archaeology. *International Journal of Heritage in the Digital Era*. 2012;1(1_suppl):1-6. doi:10.1260/2047-4970.1.0.1
2. Willis, M., Koenig, C., Black, S., & Castañeda, A. (2016). Archeological 3D Mapping: The Structure from Motion Revolution. In Index of Texas Archaeology Open Access Grey Literature from the Lone Star State. R.W. Steen Library, SFASU. https://doi.org/10.21112/ita.2016.1.110
3. J.A. Benavides López, G. Aranda Jiménez, et. al. 3D modelling in archaeology: The application of Structure from Motion methods to the study of the megalithic necropolis of Panoria (Granada, Spain), Journal of Archaeological Science. https://doi.org/10.1016/j.jasrep.2016.11.022
4. Historic England. 2017. Photogrammetric Applications for Cultural Heritage. Guidance for Good Practice. Swindon. Historic England. https://historicengland.org.uk/advice/technical-advice/recording-heritage/
5. Bianco, Simone, Ciocca et. al. (2018). Evaluating the Performance of Structure from Motion Pipelines. Journal of Imaging.
6. Gw´ena¨el Caravaca, et. al. , 3D digital outcrop model reconstruction of the Kimberley outcrop, Planetary and Space Science, Volume 182, 2020, 104808, ISSN 0032-0633, https://doi.org/10.1016/j.pss.2019.104808.
7. C. Sweeney et. al. Computing Similarity Transformations from Only Image Correspondences. (CVPR 2015) [SweeneyCVPR2015].