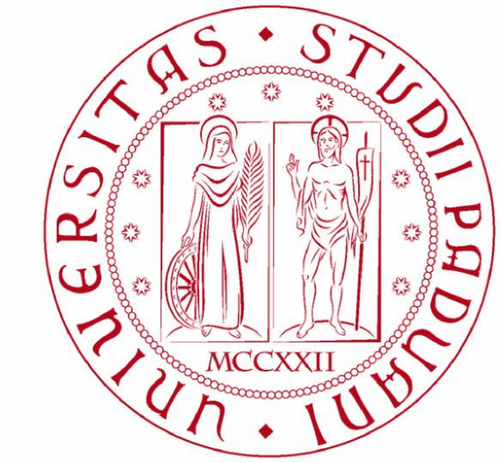
Featuremetric Refined Structure From Motion with a Hand-held Camera and Point Cloud Registration in Urban Scenarios

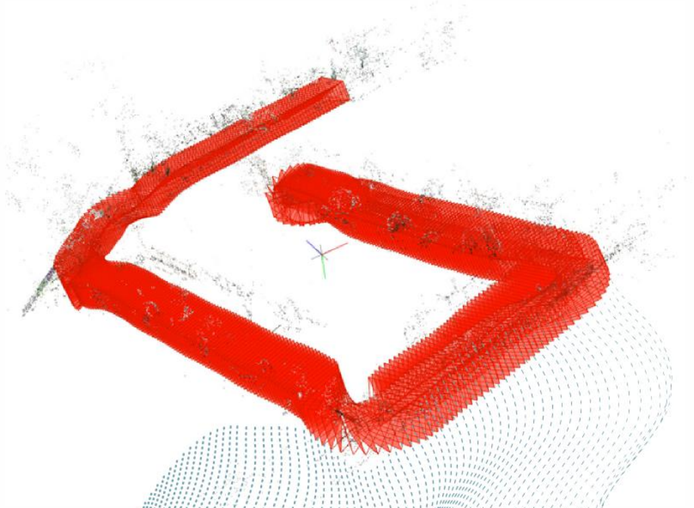
Supervisor: Dr. Alberto Pretto
Student: Kiavash Ghamsari



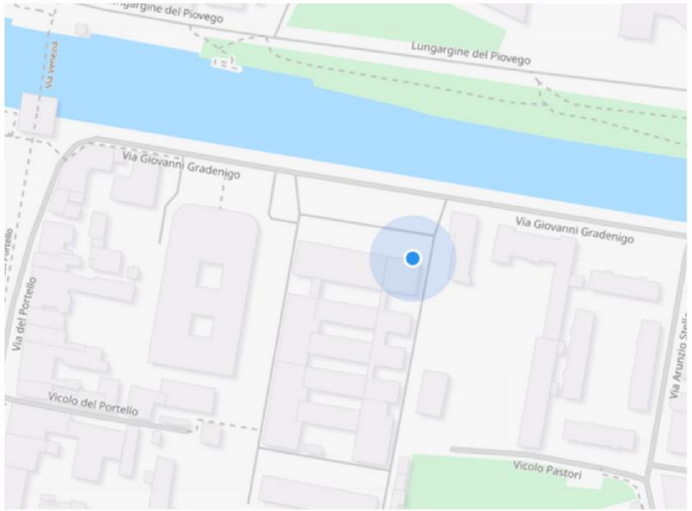
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

OVERVIEW

1. CAPTURE VIDEO AND EXTRACT FRAMES



2. GENERATE 3D POINT CLOUDS USING STRUCTURE FROM MOTION



3. LOCALIZATION USING POINT CLOUD REGISTRATION

MOTIVATION

- Offline Localization using only visual data, e.g. video frames, and no GPS or similar online devices are needed
- The precision is approximately the same as the precision of SfM, which can be within a few centimeters
- The position is updated for every video frame

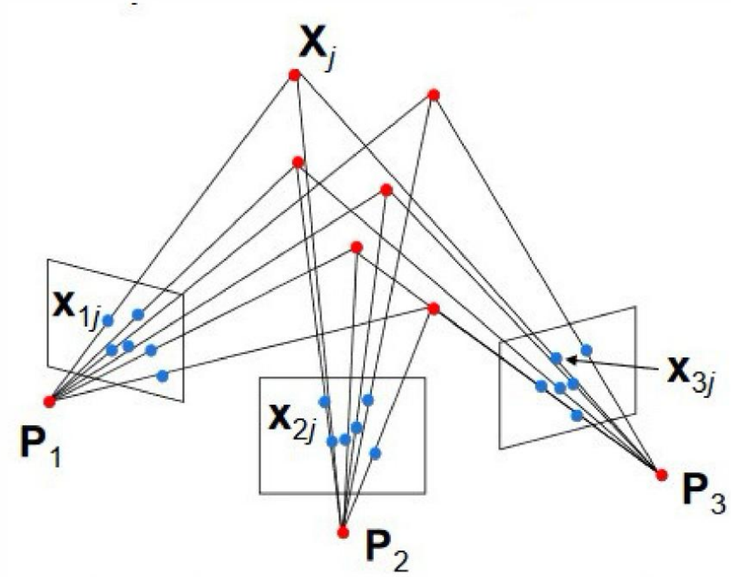
Structure from motion

is the task of calculating the

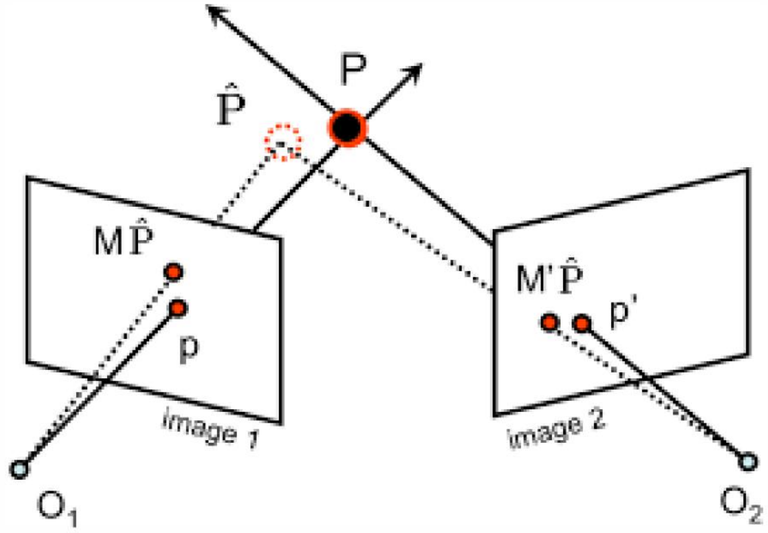
3D structure of a scene and the pose of cameras

from
images in multiple views
or video frames

Feature Detection and Feature Matching



Camera Pose Estimation using Epipolar Geometry



Refining using Bundle Ajdustment

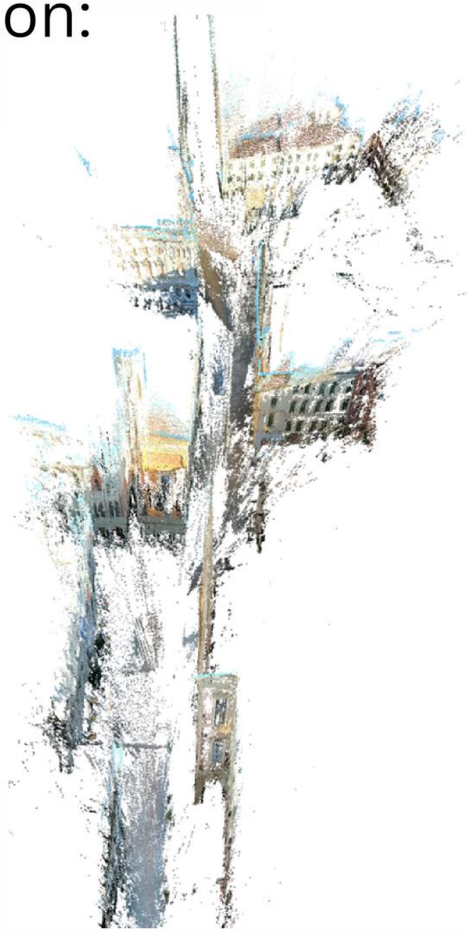
Keypoint refinement using Pixel-Perfect[1] paper

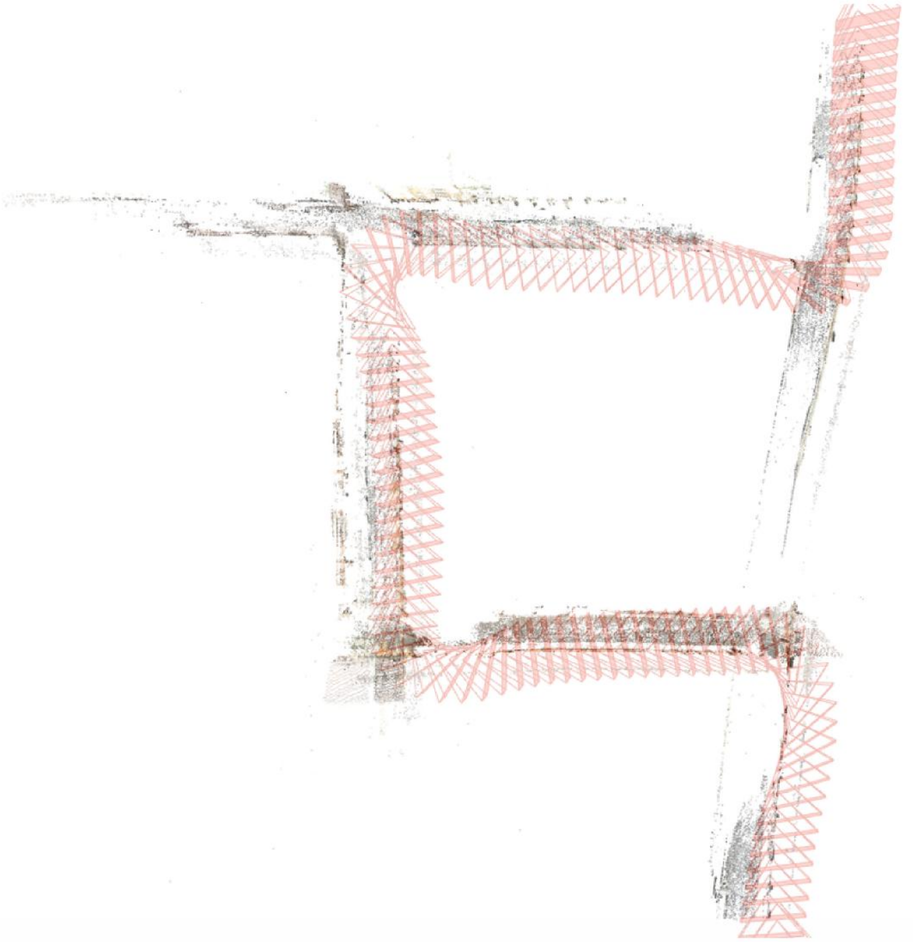
By generating Feature Maps using CNNs
And optimizing non-linear loss function between the matches among all frames

1. Paper: Pixel-Perfect Structure-from-Motion with Featuremetric Refinement. In ICCV, 2021, P. Lindenberger, P.-E. Sarlin, V. Larsson, and M. Pollefeys.

Ground Point Cloud

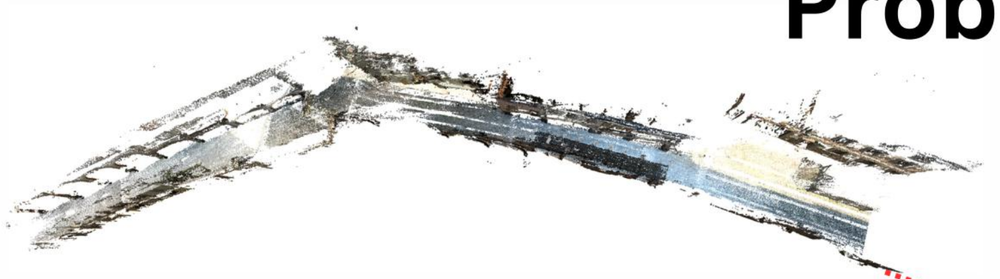
Examples of SfM reconstruction:

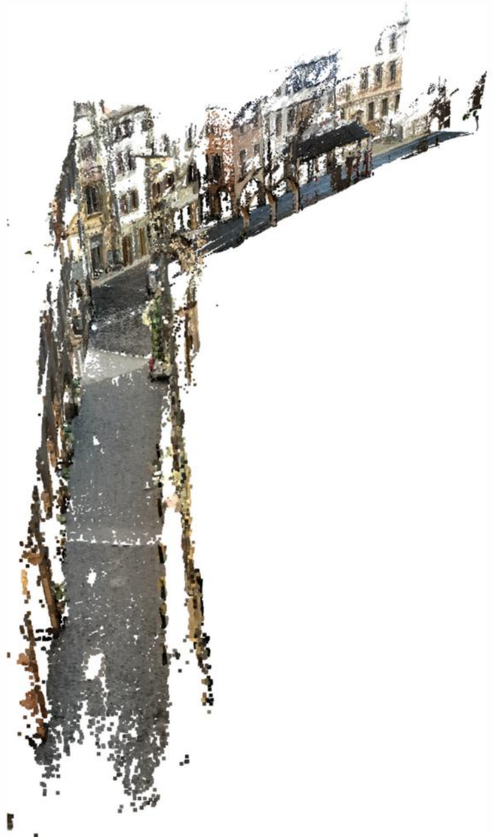


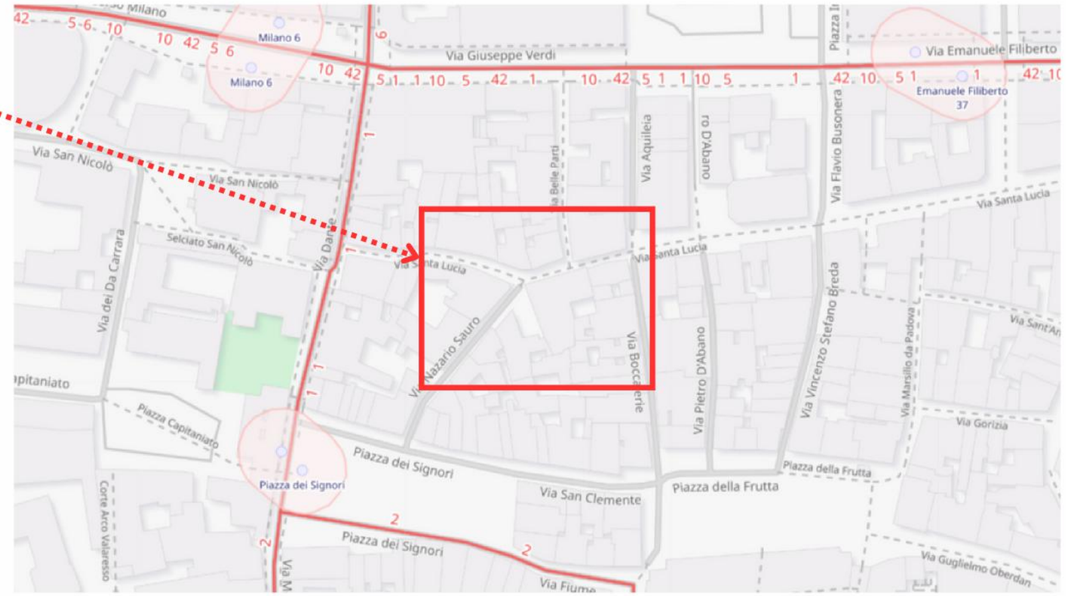


Localization

Problem Definition



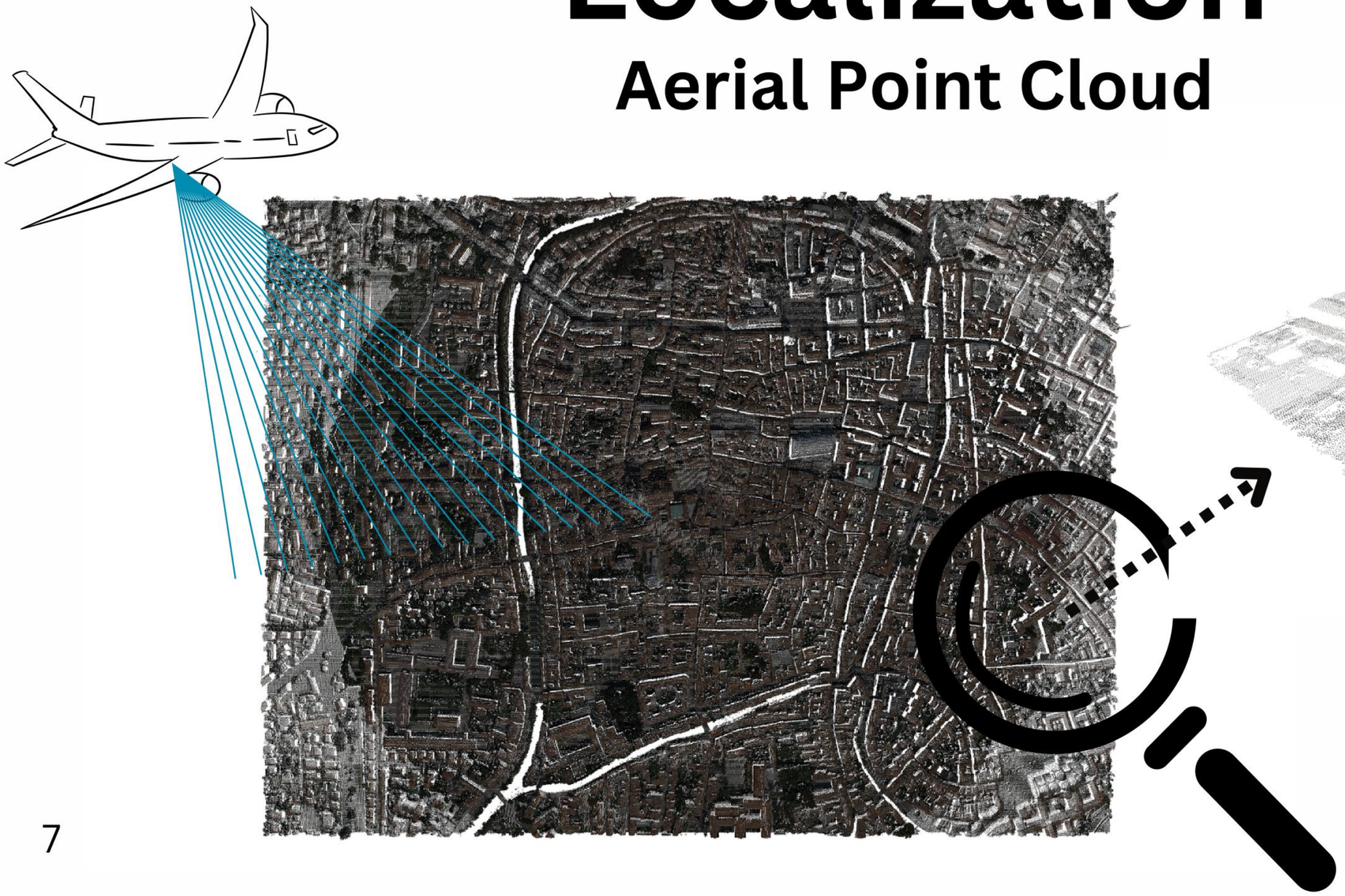




@OpenStreetMap®

Localization

Aerial Point Cloud



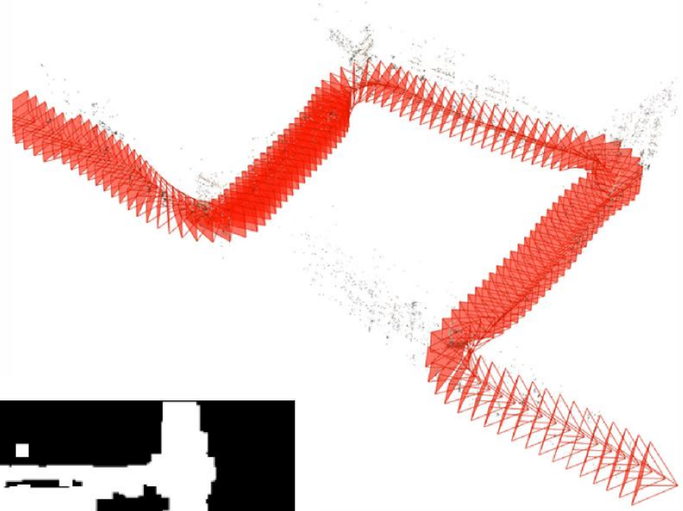
Localization

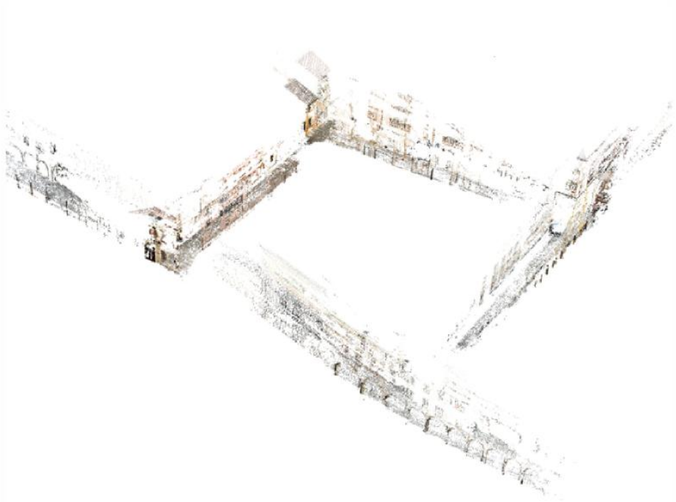
Our Method:

Ground Point Cloud Preprocessing

1

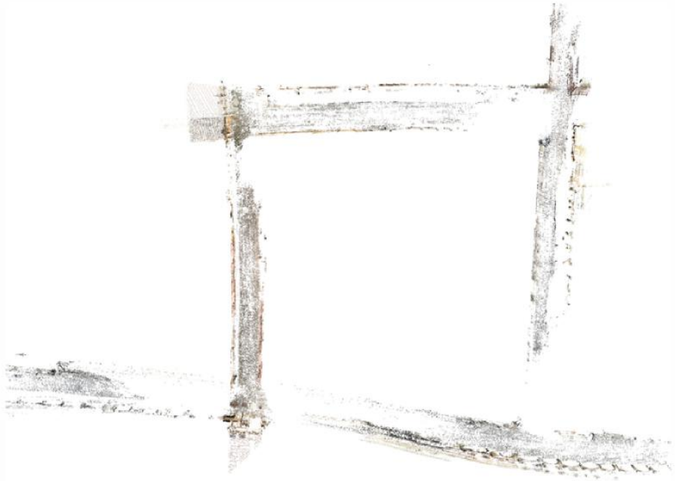
INITIAL RECONSTRUCTION USING SFM





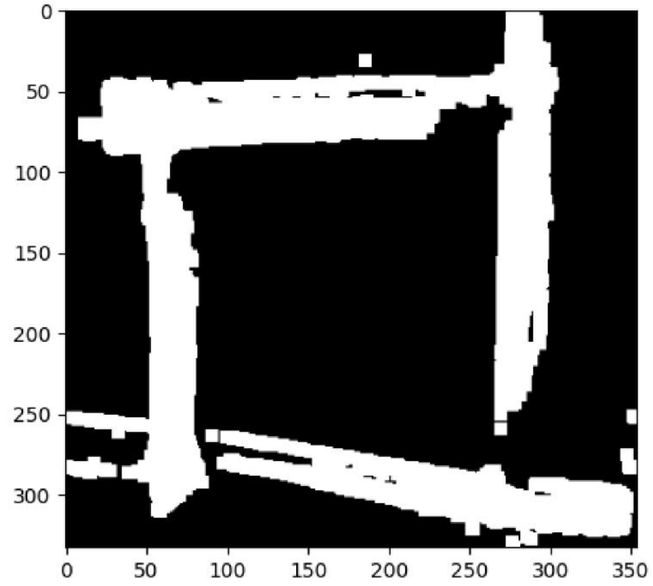
2

REFINEMENT



3

ALIGN WITH Z AXIS, SLICED



4

GENERATE 2D BINARY GRID MAP

Localization

Our Method:

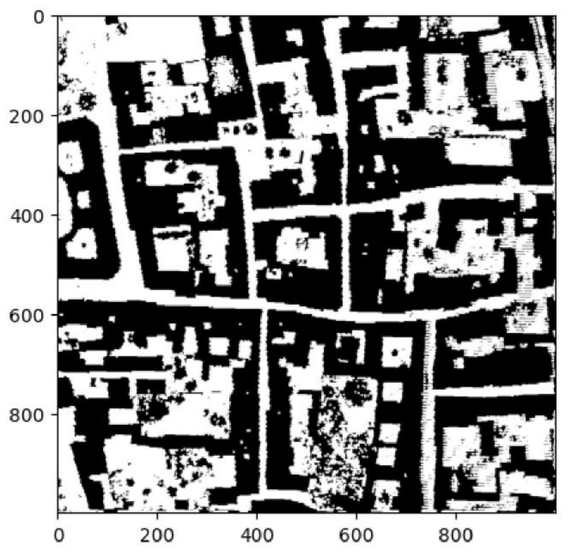
Aerial Point Cloud Preprocessing

1

A WINDOW OF AERIAL POINT GROUND

2

SLICED



3

GENERATE 2D BINARY GRID MAP

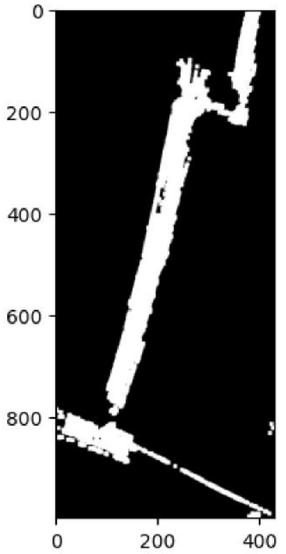
Localization

Our Method:

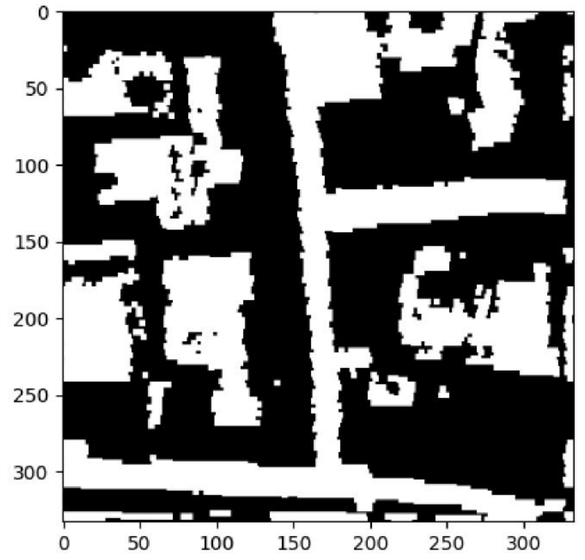
Registration using template matching

The problem is simplified into
a simple template matching problem

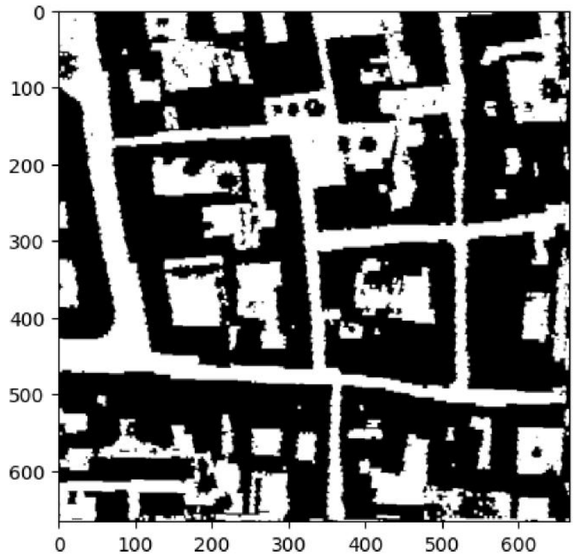
1) Template:



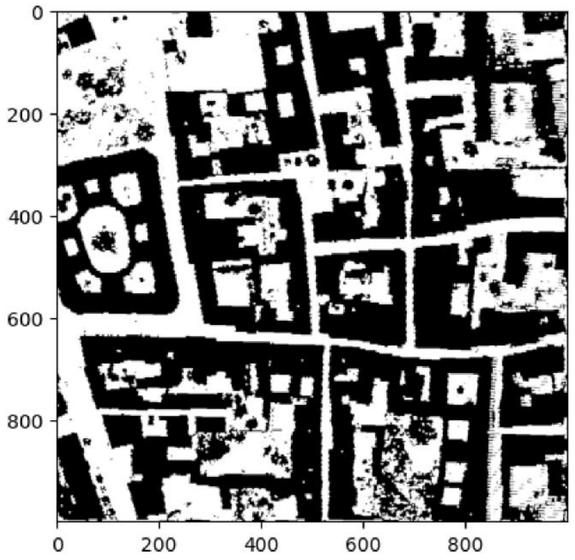
2) Sources:



EASY, 100M



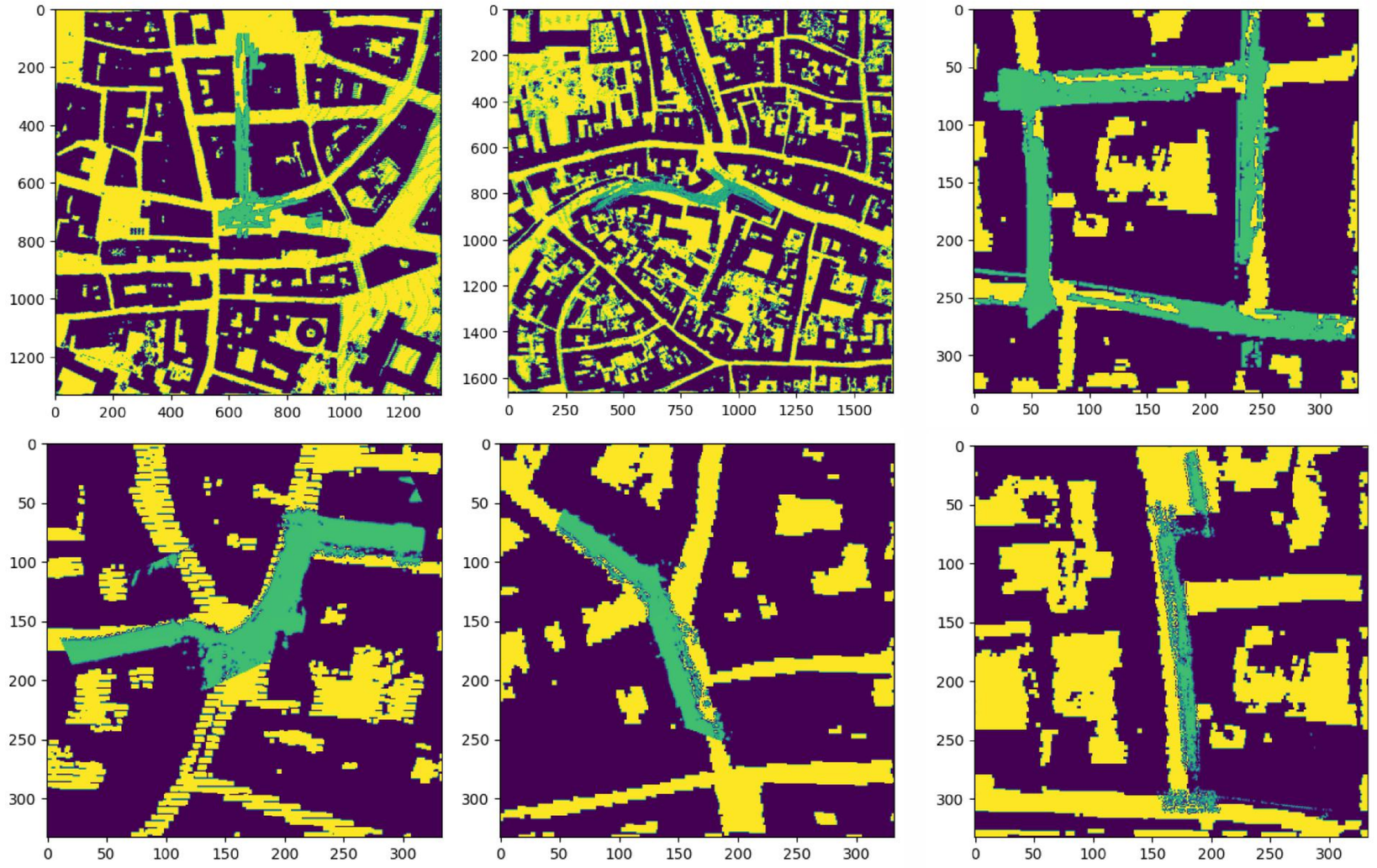
MEDIUM, 150M



HARD, 200M

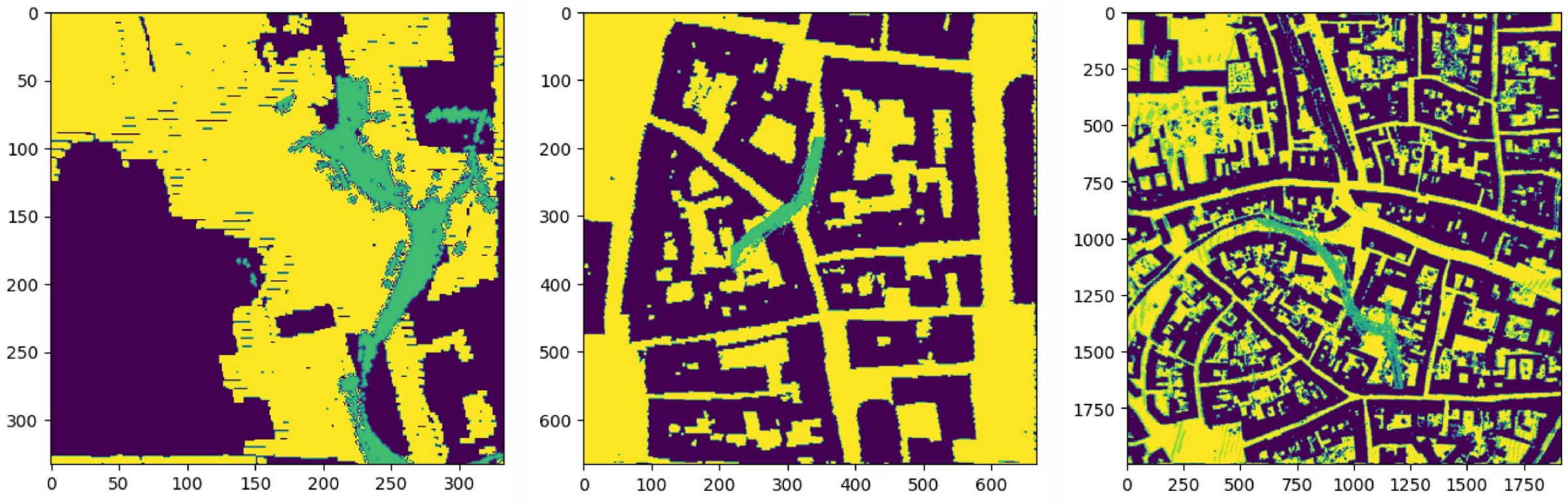
Template matching + Rotation and Scaling

Some of the successful results:



- rich info of geometry of streets:
  - relative angle of streets, crossroads
  - relative width and length
  - no affine transformation, no stretch or squeeze

Some of the failed results:



Reasons for failure could be:

- Open spaces are not discriminant enough
- Straight streets have similar matches with many other streets with different rotations and angles

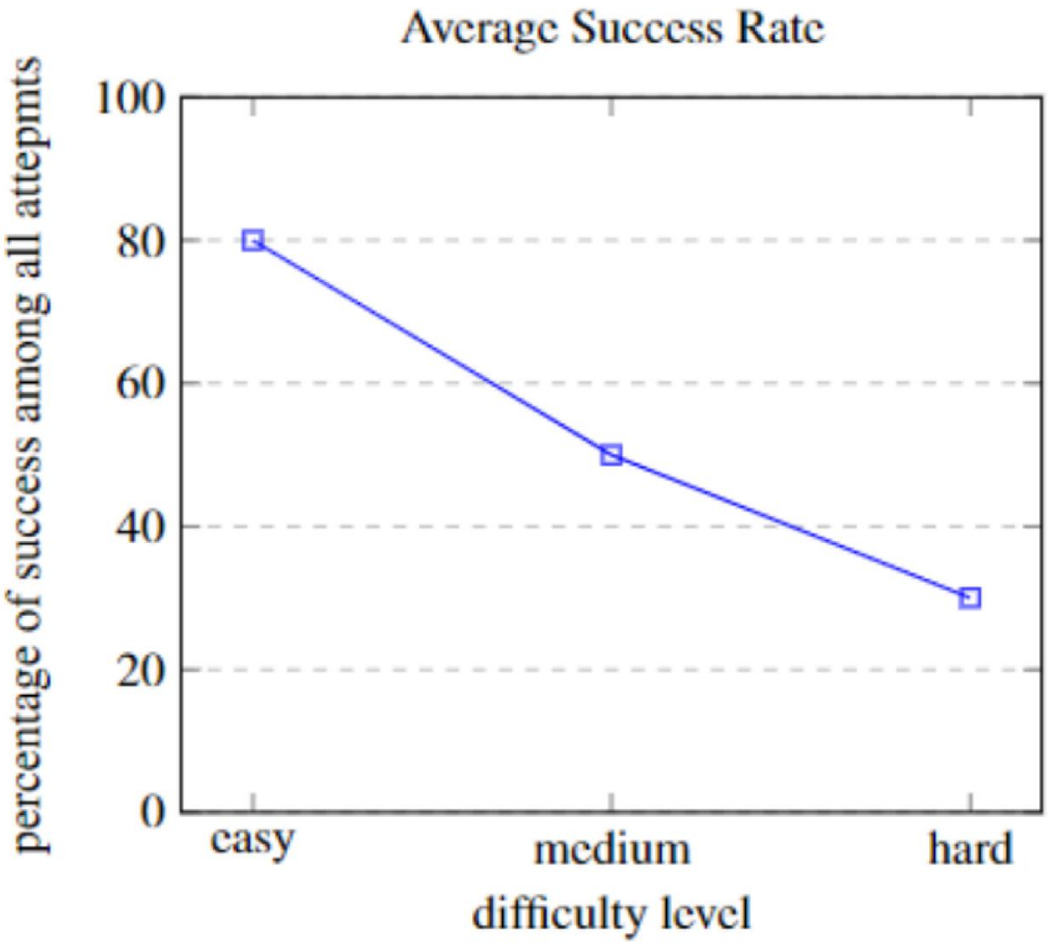
Post processing

1. Transform the Ground point cloud into the Aerial point cloud

2. ICP (Iterative Closest Point) registration

Results and Conclusion

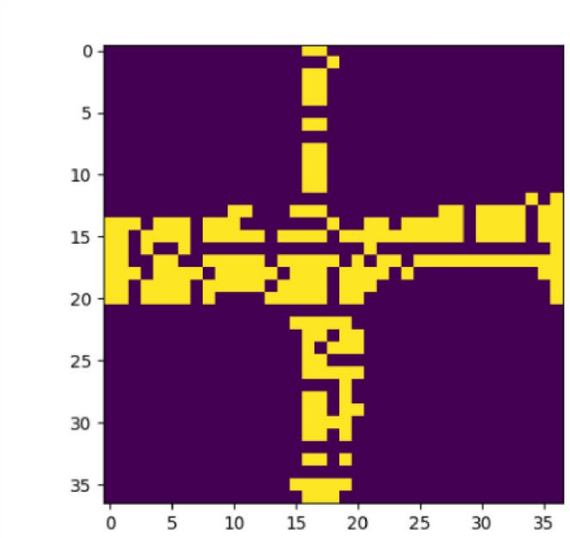
- We are able to register the point clouds using dimensionality reduction(3D ply to 2D binary images)
- Promising results
- Localization can be completely offline
- The error of positioning can be reduced to less than a meter (equal to the error of SfM) in compare to GPS



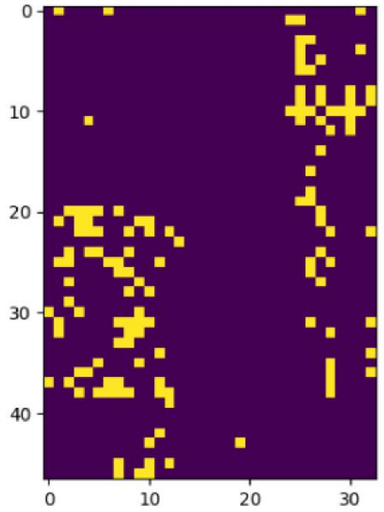
Future Ideas:

Global Registration:

- **CROSSROADS** are discriminant
- Good results by Deep Learning Approach to detect crossroads:
- An example of the inputs of CNN for classification:

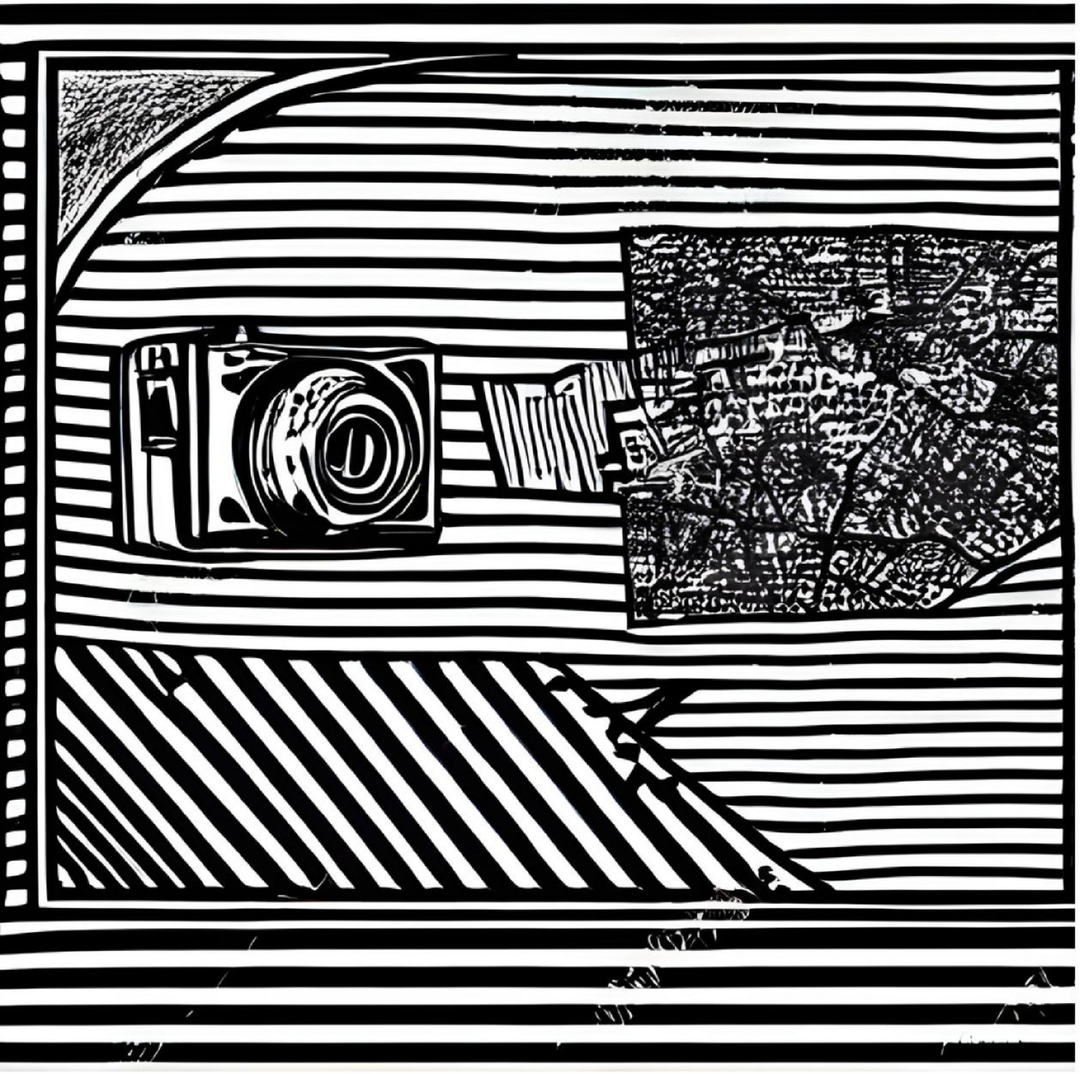


1 - Binary grid map of crossroad



0 - Binary grid map of Non crossroad

A big dataset containing all kinds of crossroads in the Aerial Point Cloud is needed



Featuremetric Refined Structure From Motion with a Hand-held Camera and Point Cloud Registration in Urban Scenarios

Thanks for your attention

Kiavash Ghamsari

13/July/2023