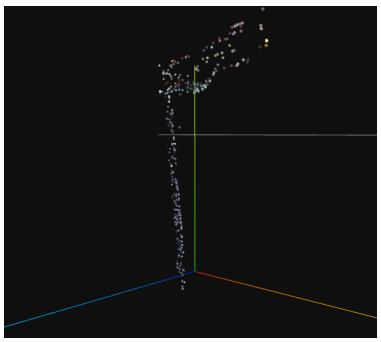
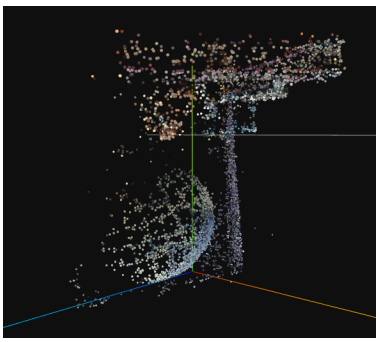
Structure From Motion (PA1)

Student : Mathis Saunier

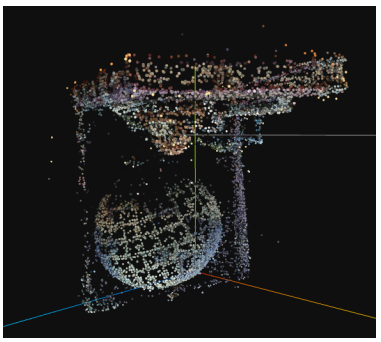
Screen captures : professor dataset



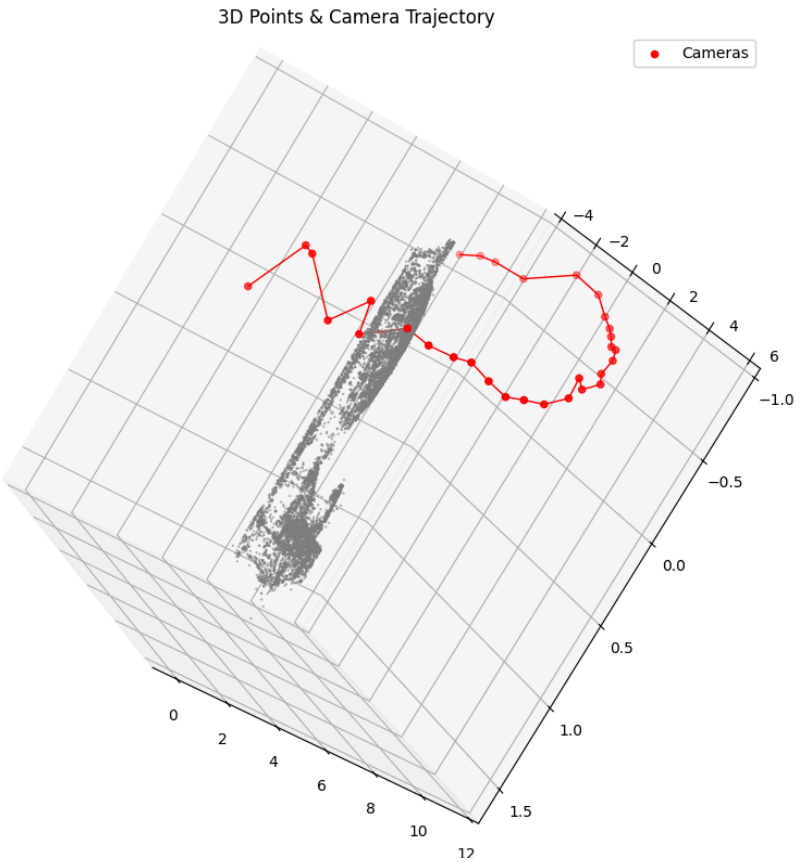
2-view matching



16-view matching

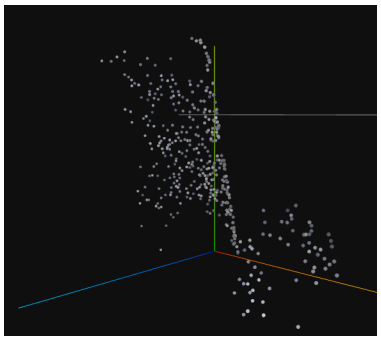


32-view matching

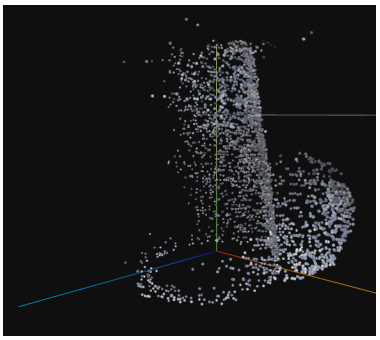


View of 32 camera’s positions

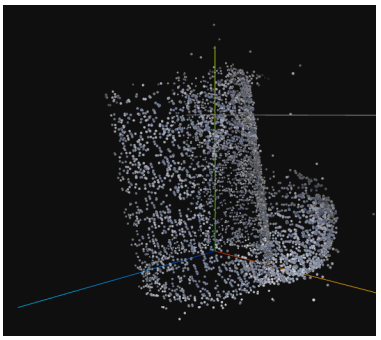
Screen captures : personal dataset



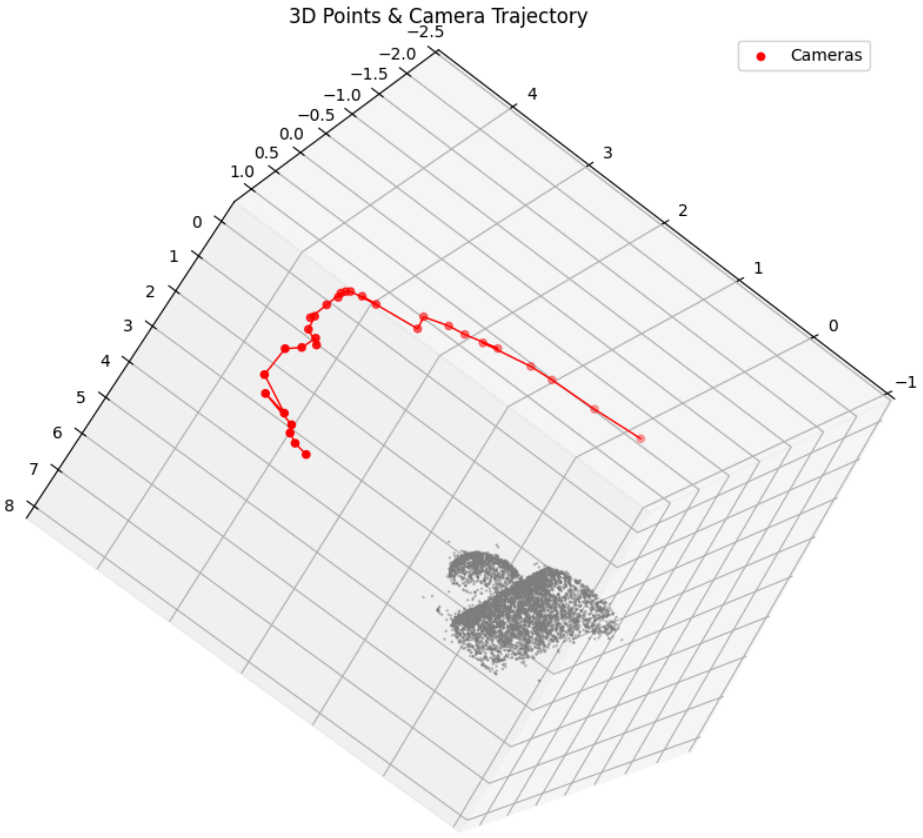
2-view matching



16-view matching



32-view matching



View of 32 camera’s positions

Questions

• (Step II) What is RANSAC? and why should we use RANSAC?

RANSAC stands for Random Sample Consensus. It is used to robustly estimate the Fundamental or Essential matrix by rejecting outliers among feature matches. Because matches between SIFT features is difficult, there are many outliers.

RANSAC will use random subsampling in order to find a robust and consistent result. We used it to estimate the essential matrix and inside the PnP algorithm (Perspective-n-Point).

• (Step III) Why are four camera poses?

To recover rotation (R) and translation (t) between two cameras, the essential matrix E is decomposed using singular value decomposition (SVD). This yields two possible rotations and two possible translations. In total, we have four (R, t) combinations possible, but only one is correct (the one that ensures triangulated 3D points lie in front of both cameras).

$$\mathbf{R} = \mathbf{U}\mathbf{W}\mathbf{V}^{\mathrm{T}} \text{ or } \mathbf{U}\boldsymbol{W}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}$$

$$\mathbf{t} = \boldsymbol{u_3} \text{ or } -\boldsymbol{u_3}$$

- $SVD(E) = Udiag(1,1,0)V^T$
- $W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
- $u_3 = U(0,0,1)^T$: The last column vector of U

Screenshot of assignment (page 8)

• (Step III) Why should we apply W between U and V for R

R is a matrix of rotation which means that the matrix is correct (real) only if $det(R) = 1$.

W is necessary to impose this geometric constraint (which comes from essential matrix's properties).

• (Step V) What are the differences between epipolar geometry and PnP?

Epipolar geometry describes the geometric relationship between two cameras. Thus, as we have seen in previous lectures, we can use the fundamental (F) or essential (E) matrix to find point correspondences between two images. Thanks to epipolar geometry, we can reconstruct relative poses and triangulate some 3D points.

PnP (Perspective-n-Point) is different. With known 3D points (already triangulated) and their 2D projections into a new image, we can deduce the absolute pose (R, t) of that camera (the one which gave us the 2D image). PnP is used in the growing step.

• **(Step VI) What is the meaning of each element in intrinsic matrix K (3x3 matrix)?**

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

We also saw in previous lectures the meaning of the intrinsic matrix K.
f_x and f_y are the focal distances in pixels of our camera.
c_x and c_y are the coordinates of our optic center
The last row symbolizes the homogenization of coordinates.