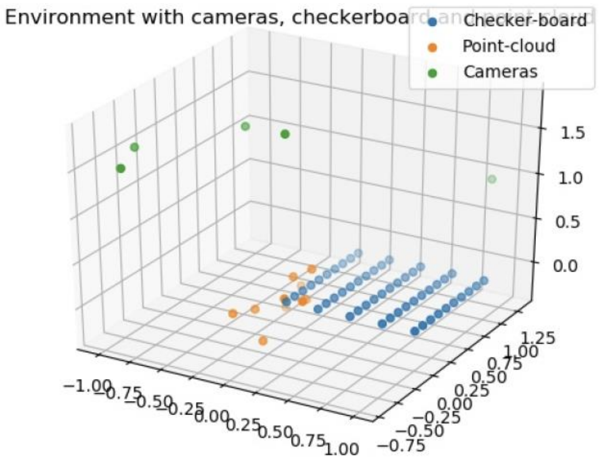
Classical Algorithm Engineer
Assignment Report

Qualcomm

Ashish Garg
Research Engineer – Perception @ OLA

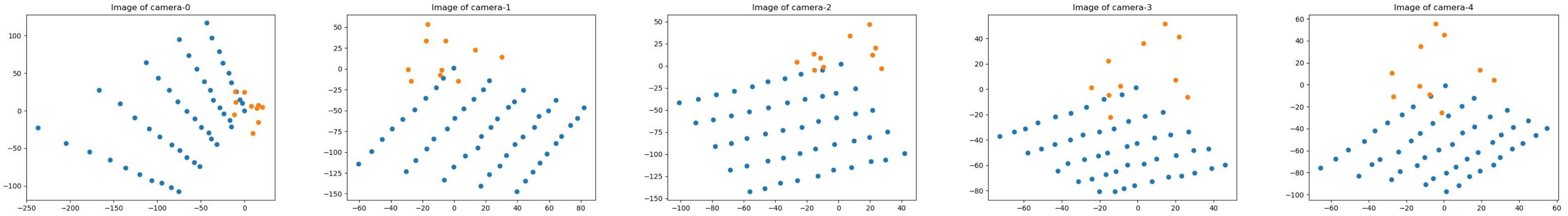
Problem statement

Given configuration of environment



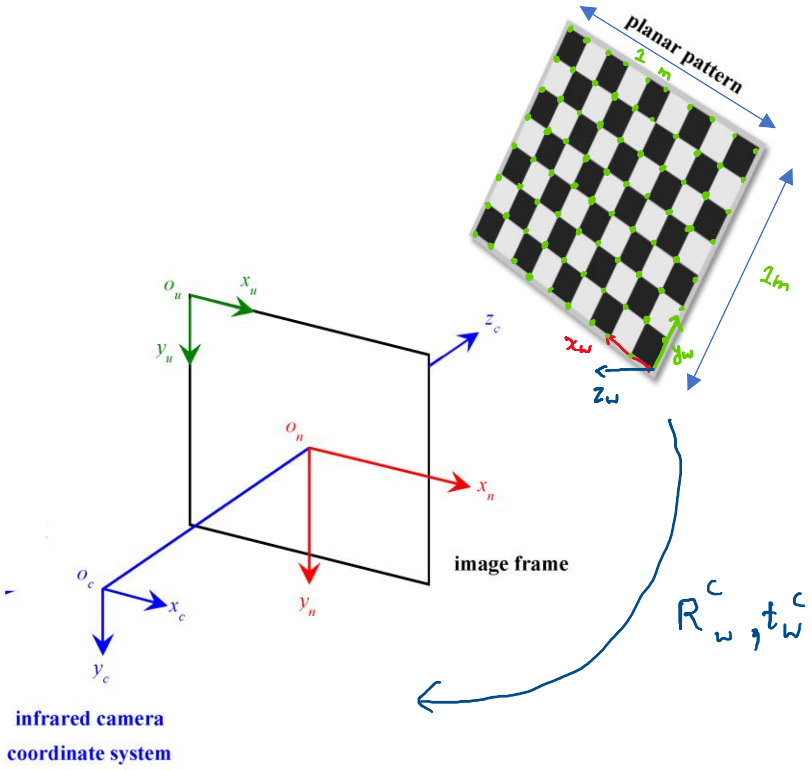
- **Camera Specifications:**
  - Focal length: 200
  - Principle point: (0,0)
  - No distortion
- **Checkerboard Details:**
  - Size: 1m x 1m
  - Features: 50 identifiable corners
- **Point Cloud:**
  - Location: Fixed within the environment
- **Imaging Process:**
  - The camera captures the checkerboard and point cloud from five different angles.
- **Objective:**
  - Determine the most accurate 3D positions for the points in the point cloud using multi-view geometry.

Projection of points in each camera



Homography

Homography Matrix



$$(x,y,1)^T \sim x \cdot PX = K[r_1\ r_2\ \cancel{r_3}\ t]\begin{bmatrix}X\\Y\\\cancel{Z}\\1\end{bmatrix} = H\begin{bmatrix}X\\Y\\1\end{bmatrix}$$

$Z = 0$

$$H = \lambda K[r_1\ r_2\ t]$$

$$K^{-1}H = \lambda[r_1\ r_2\ t]$$

$$P = K[r_1\ r_2\ (r_1 \times r_2)\ t]$$

- Ensures that the rotation matrix columns have a unit norm
- The cross product $r1 \times r2$ yields a vector orthogonal to both $r1$ and r2,
- Aligns the scale of the translation with the scale of the rotation column

$$R|t = K^{-1}H = \begin{bmatrix}r_1 & r_2 & t\end{bmatrix}$$

$$r_1 = \frac{r_1}{\|r_1\|}$$

$$r_2 = \frac{r_2}{\|r_2\|}$$

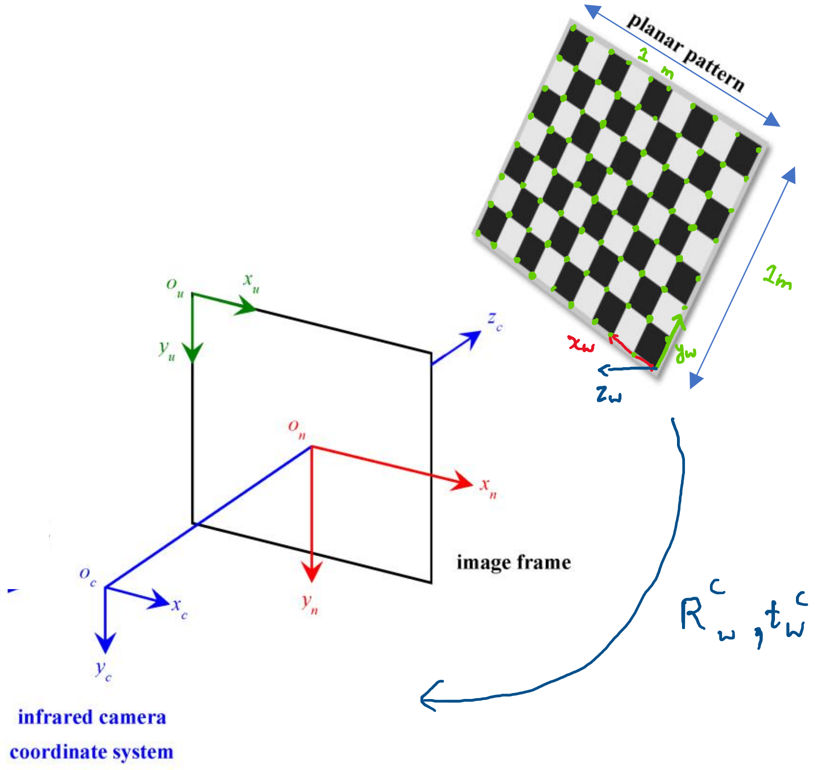
$$r_3 = r_1 \times r_2$$

$$t = \frac{t}{\|r_1\|}$$

Reference:-

1. https://docs.opencv.org/4.x/d9/dab/tutorial_homography.html
2. https://web.archive.org/web/20171226115739/https://ags.cs.uni-kl.de/fileadmin/inf_ags/3dcv-ws11-12/3DCV_WS11-12_lec04.pdf

Compute the optimal rotation matrix



$$\hat{R} = U\Sigma V^T$$

where U and V are orthogonal matrices from the SVD of $\hat{R}$, and Σ is a diagonal matrix. To ensure that Σ is also orthogonal and thus R is a proper rotation matrix with $\det(R) = 1$, we modify Σ as follows:

$$d = \text{sign}(\det(VU^T))$$

$$\Sigma' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d \end{bmatrix}$$

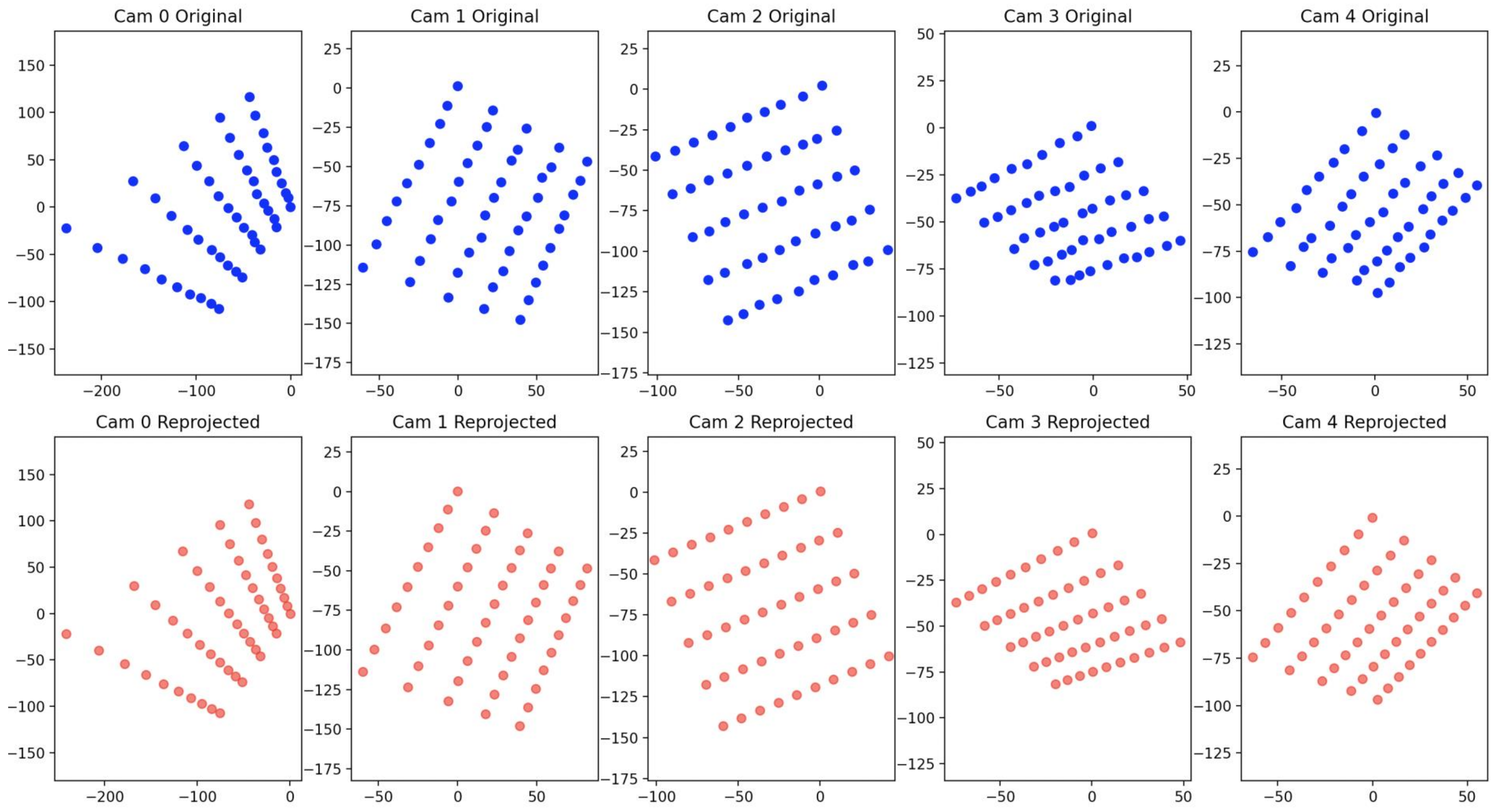
$$R = V\Sigma' U^T$$

This adjustment guarantees that R is orthogonal since V and U are orthogonal and Σ' is chosen such that R has a determinant of $+1$, which is a necessary condition for a rotation matrix.

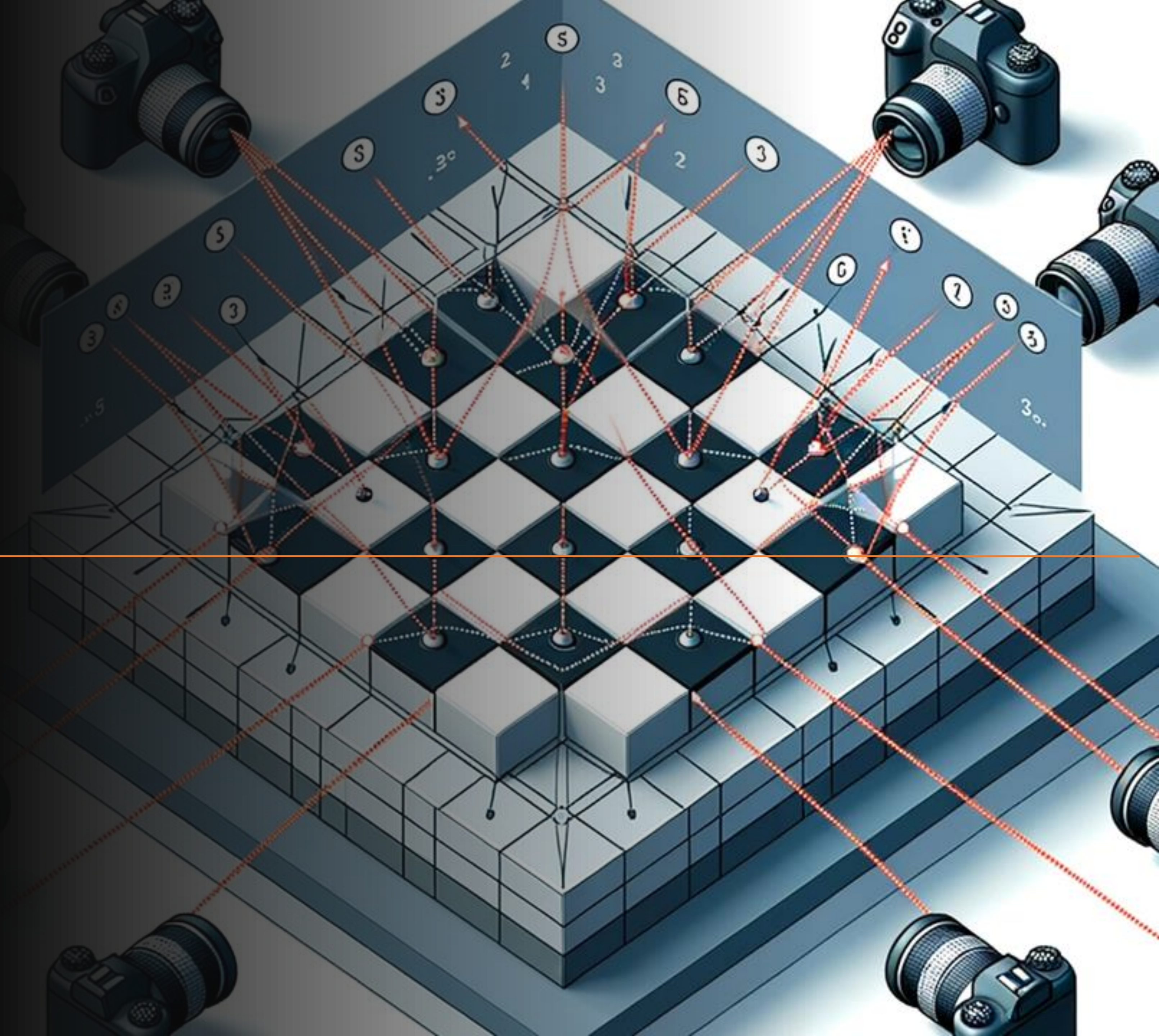
Reference:-

1. https://en.wikipedia.org/wiki/Kabsch_algorithm#Computation_of_the_optimal_rotation_matrix

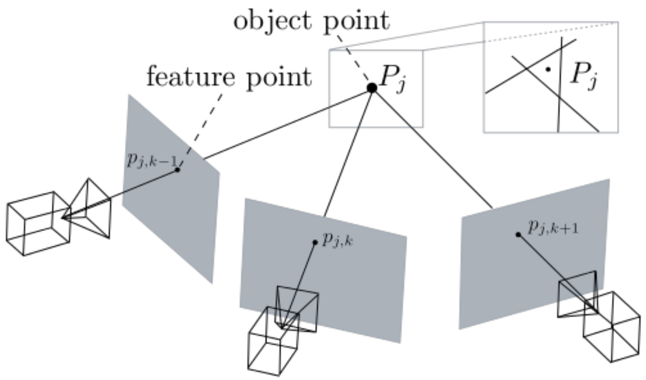
Reprojection based on [R|t] (Homography)



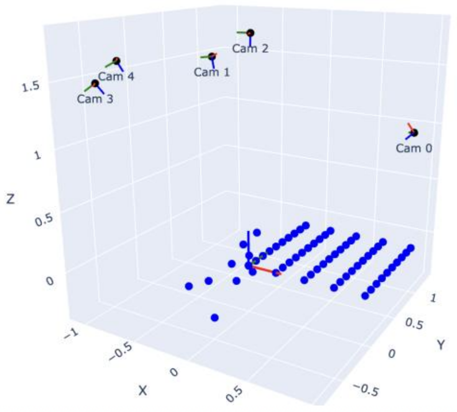
Multi-View Triangulation



Multi-View Stereo Triangulation



Resulting configuration after triangulation



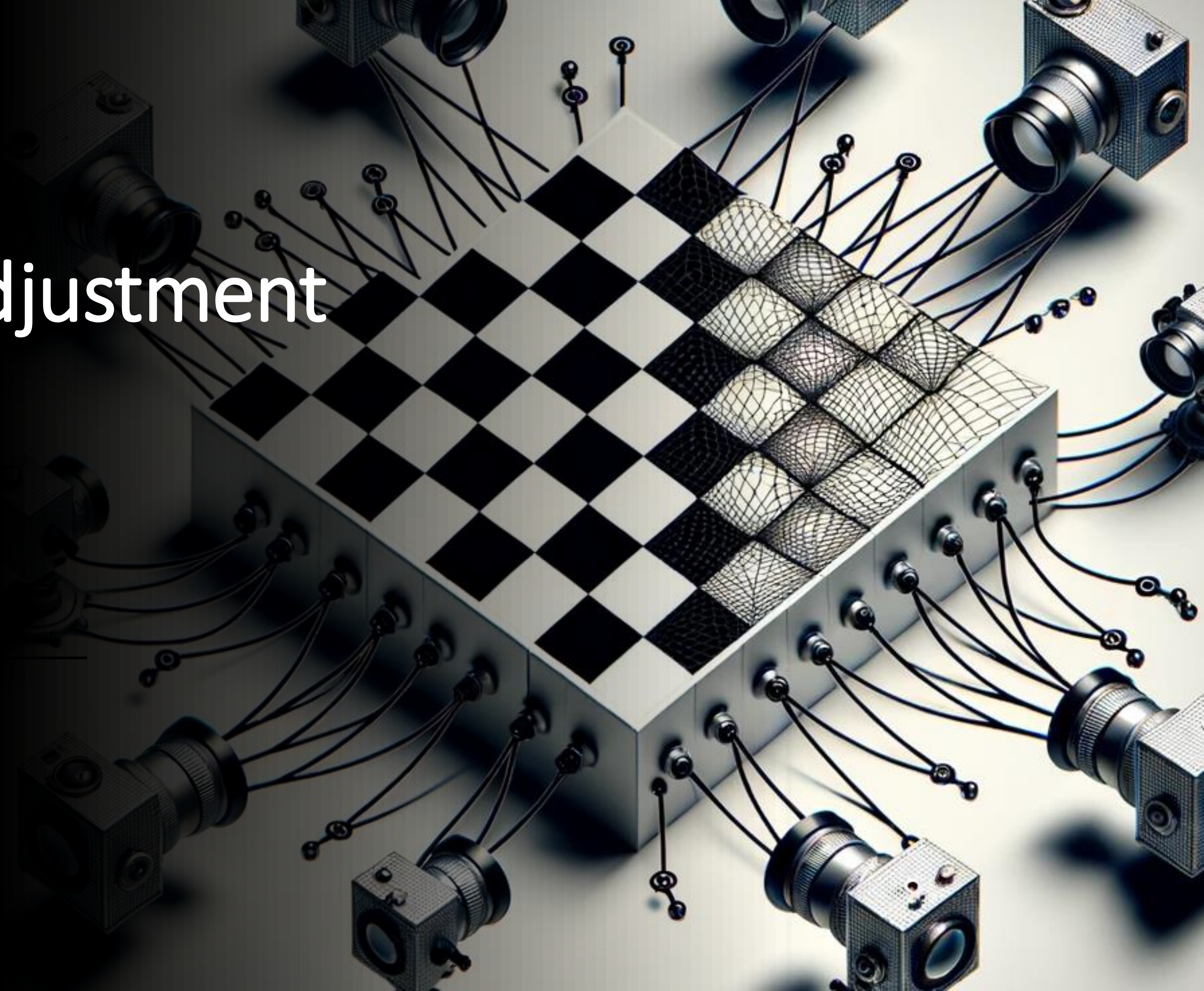
3D Visualization of World Frame, Camera Frames, and Checkerboard Points

1. **Representation of Rays**:
   - $P_i = C_i + \lambda_i d_i$
   - P_i: Point on the ray from the i^{th} camera
   - C_i: Camera center in world coordinates
   - λ_i: Scalar representing distance along the ray
   - d_i: Direction vector of the ray from the i^{th} camera
2. **Ideal Case (No Noise)**:
   - $\lambda_i d_i \approx X - C_i$
   - Assumes the 3D point X lies exactly on the ray from C_i in the absence of noise
3. **Ray Equation and Residual**:
   - $r = X - C_i - d_i d_i^T (X - C_i)$
   - r: Residual vector, component perpendicular to d_i
   - $d_i^T (X - C_i)$: Projection of $X - C_i$ onto d_i
4. **Loss Function**:
   - $L = \sum_{i=1}^{N} ((I - d_i d_i^T)(X - C_i))^2$
   - L: Sum of squares of residuals for all cameras
   - I: Identity matrix
5. **Minimization and Derivative**:
   - $\frac{\partial L}{\partial X} = 2 \sum_{i=1}^{N} (I - d_i d_i^T)^2 (X - C_i) = 0$
   - Derivative of L with respect to X set to zero
6. **Matrix Formulation and Solution**:
   - $A_i = (I - d_i d_i^T)$
   - $X = (\sum_{i=1}^{N} A_i^T A_i)^{-1} \sum_{i=1}^{N} A_i^T C_i$
   - A_i: Matrix to find orthogonal component of $X - C_i$ to d_i

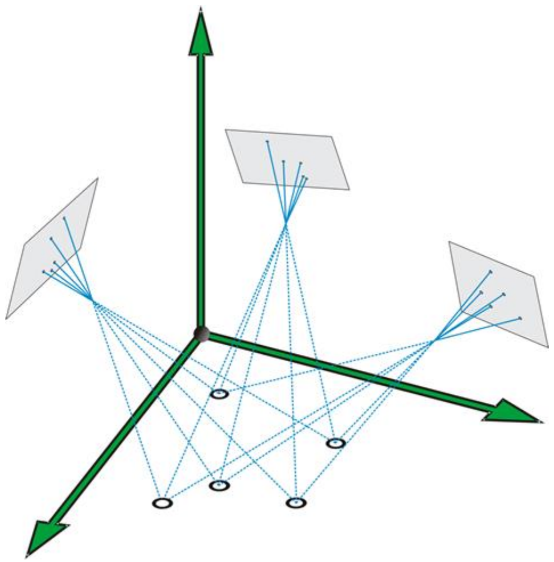
Reference:-

1. https://3d.bk.tudelft.nl/nail/stereo/

Bundle Adjustment



Bundle Adjustment



The reprojection error equation in bundle adjustment is formulated as:

$$e_{ij} = x_{ij}^{\text{obs}} - \pi(K_i, P_i, X_j) \qquad \min_{X,P,K} \sum_{i=1}^{n} \sum_{j=1}^{m} \|e_{ij}\|^2$$

where:

- e_{ij} is the reprojection error for the jth point in the ith image.
- x_{ij}^{obs} is the observed image point.
- π represents the projection function that maps 3D points X_j into 2D image points using the camera intrinsic parameters K_i and the extrinsic parameters $P_i = (R_i, t_i)$.
- K_i are the intrinsic parameters of the ith camera.
- $P_i = (R_i, t_i)$ represents the pose (rotation R_i and translation t_i) of the ith camera.
- X_j is the jth 3D point in the world coordinate system.

The goal is to find the parameters $\theta = \{X_j, P_i, K_i\}$ that minimize the sum of squared reprojection errors for all observations. The LM optimization problem can be formulated as:

$$\min_\theta \sum_{i=1}^{n} \sum_{j=1}^{m} \|e_{ij}(\theta)\|^2$$

The update rule in the LM method is given by:

$$\Delta\theta = -(J^T J + \lambda \cdot \text{diag}(J^T J))^{-1} J^T e$$

where:

- J is the Jacobian matrix of partial derivatives of the error terms with respect to the parameters,
- e is the vector of all reprojection errors,
- λ is the damping factor controlling the step size and the mix between the gradient descent and Gauss-Newton method,
- $\text{diag}(J^T J)$ represents the diagonal matrix of $J^T J$.

The parameters are updated iteratively:

$$\theta^{(k+1)} = \theta^{(k)} + \Delta\theta^{(k)}$$

Note -

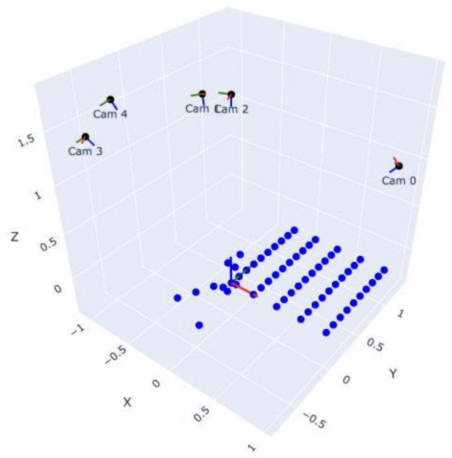
- Used both Axis – Angle and quaternion for Rotation representation
- Fixed the checkboard 3d point to fix the structure
- Used sparse Jacobian method to speed up the bundle adjustment

Reference:-

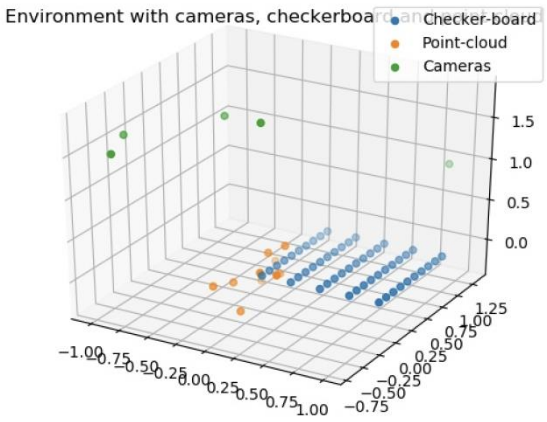
1. https://www.telesens.co/2016/10/25/bundle-adjustment-part-3/

Bundle Adjustment

Resulting configuration after Bundle Adjustment



Given configuration



Resulting point cloud location

Point #	X Coordinate	Y Coordinate	Z Coordinate
1	0.13	-0.21	0.12
2	-0.13	0.21	-0.02
3	-0.34	-0.36	-0.11
4	-0.02	0.10	-0.10
5	-0.02	-0.21	0.10
6	-0.26	-0.19	-0.11
7	-0.11	-0.33	-0.34
8	-0.00	0.14	0.24
9	-0.18	0.10	-0.22
10	-0.09	0.07	0.14