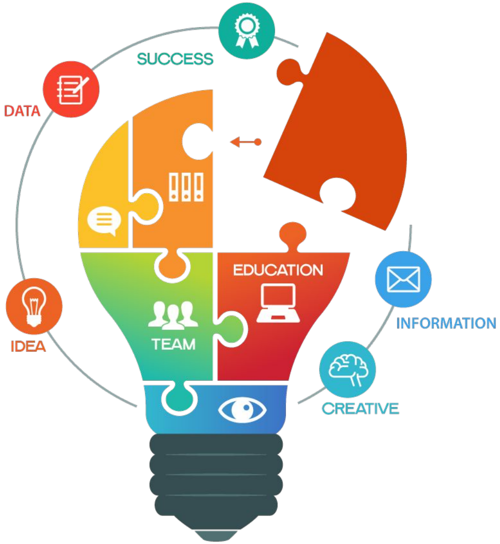
Team: Brown Munde

3D Reconstruction from Accidental Motion



Mentor : Meher Shashwat Nigam

Team : Nonidh Singh (20171203)
Puru Gupta (20171187)
Aayush Goel (20171188)
Priyanshu Madaan (2018101097)

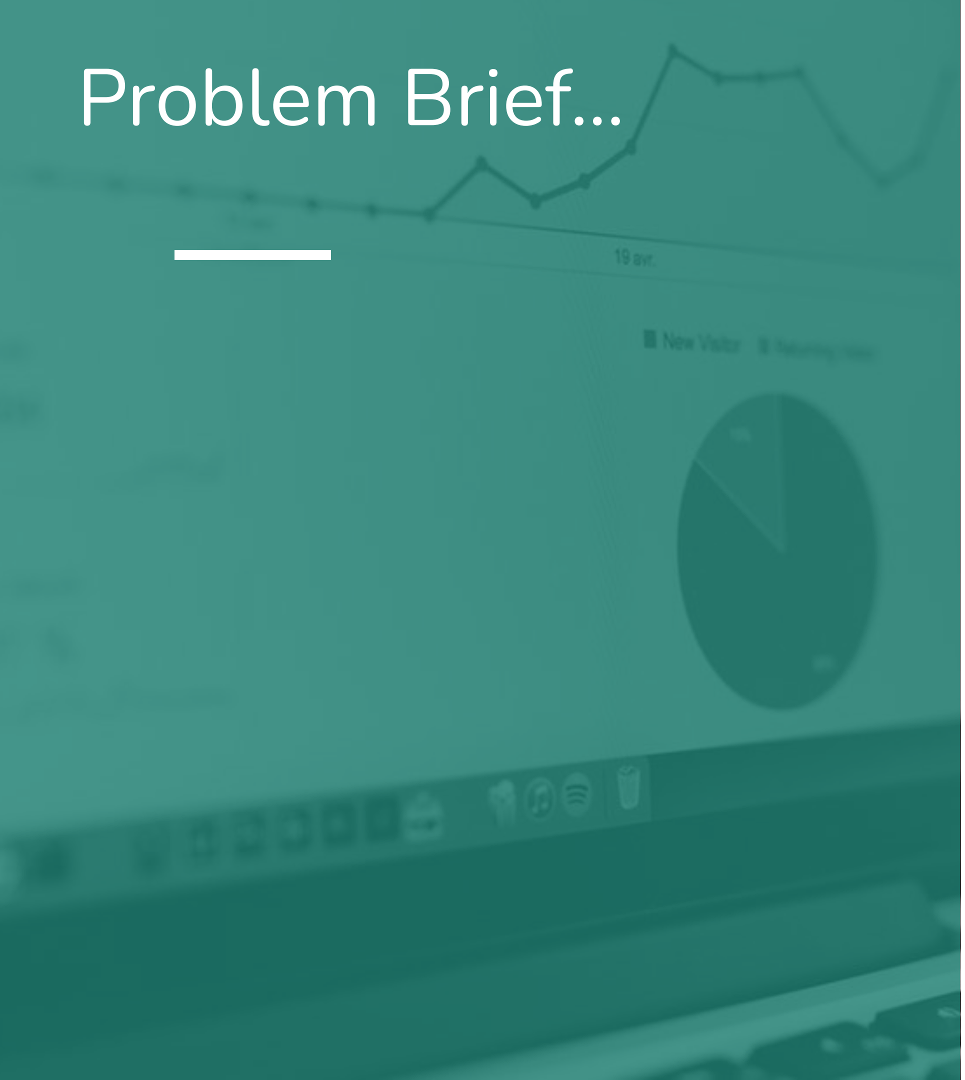
Objective

- ➔ As our project we are going to implement the following paper : [3D Reconstruction from Accidental Motion]
- ➔ Our **aim** is : **"To reconstruct a 3D scene from a set of initial frames of a video capture by exploiting accidental motion"**
- ➔ Our implementation will take the following parameters :
    - ◆ **Input :** Sequence of frames of parts of video.
    - ◆ **Result :** A 3 Dimensional reconstruction depth map of a reference frame

Problem Brief

- ➔ We have an image sequence of N_C images and N_P projections (2D points) of corresponding 3D points as seen from every camera , we try to estimate the world coordinates of the real world points using **Bundle Adjustment**.
  - ◆ **Bundle Adjustment** : It refers to solving the location of pixels for a given estimated initial pose and location of 3D points.
- ➔ We take the first frame as reference frame and parametrize all the 3D points by inverse depth relative to reference frame.
- ➔ We start with a random initialization and then solve for camera poses using Bundle Adjustment.

Problem Brief...



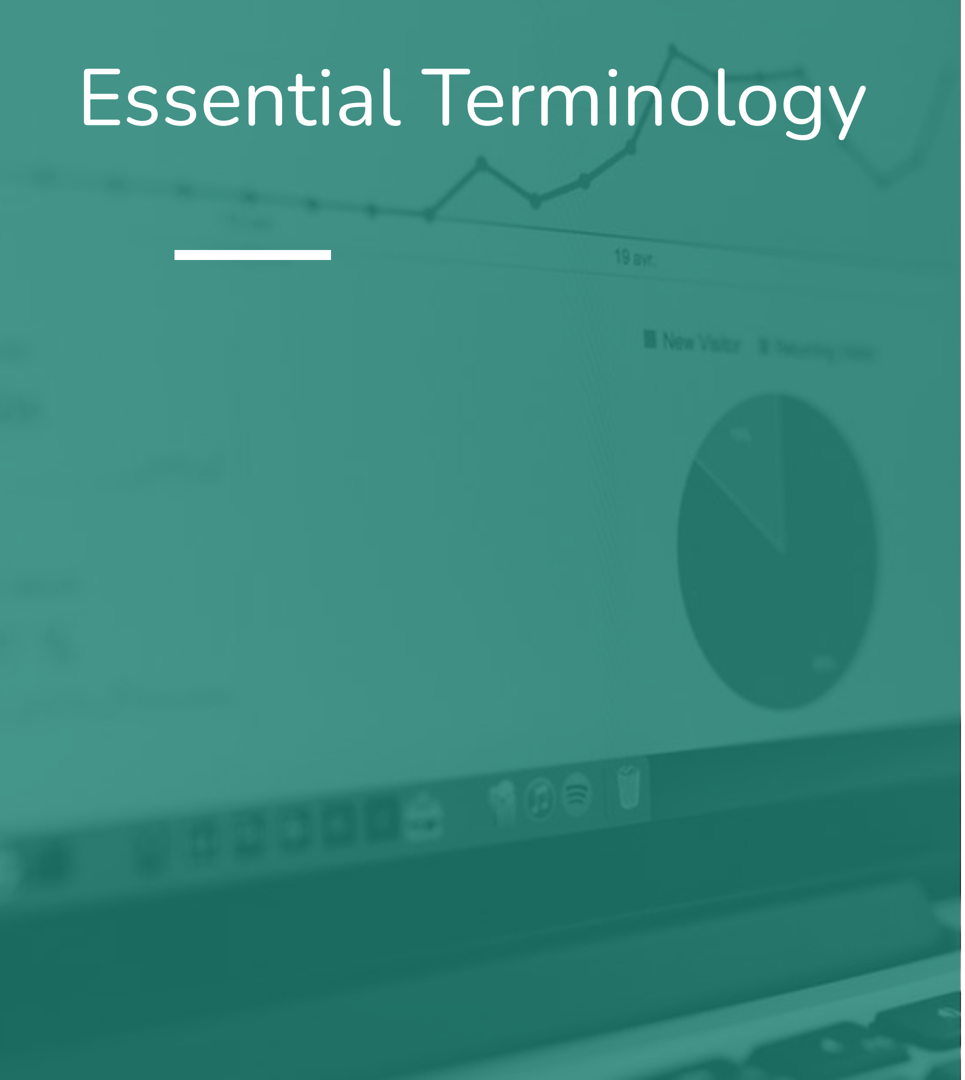
- ➔ We then reconstruct the 3D scene from estimated camera poses. This results in a smooth depth map.
- ➔ We regularize the depth estimation by minimising an energy function.

Method Overview

➔ The paper describes the following pipeline for the 3D reconstruction

- Extract good features using **Shi-Tomasi** method.
- Track the detected features using **Lucas-kanade** method from reference image I_R.
- Now use the tracked features to estimate the 3D structure of scene using bundle adjustment.
- The final result is a dense map reconstructed from the sparse 3D structure using a **CRF model.** This incorporates a photo-consistency and smoothness loss.

Essential Terminology



➔ **Accidental Motion**

- The motion experienced while intending to hold the camera still but inevitably causing motion due to hand shaking or beating of the heart.
- Mostly dominant in cases of light weighed cameras like smartphone, etc.

➔ **Bundle Adjustment:**

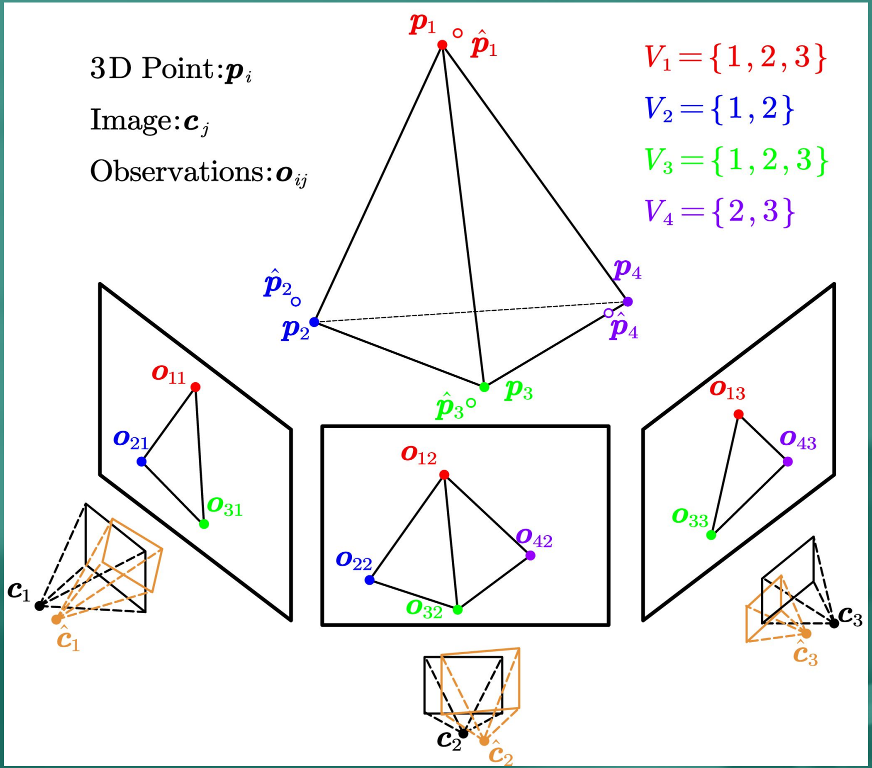
- The problem of using initial estimated initial poses and 3D location points for solving for further location of pixel values and poses.

Sequence Initializer

Optimised initialisation for Bundle adjustment plays an important role :

- ❖ The first image from the sequence of images is kept as a reference image.
- ❖ With respect to the first image, all other images are initialized with zero rotation and translation considering accidental motion to be very small.
- ❖ Kanade-Lucas-Tomasi(KLT) function is used to find the projections for the 3D points over the sequence.
- ❖ Inverse depth is used as the initialisation parameter for these 3D points.

Bundle Adjustment Optimization

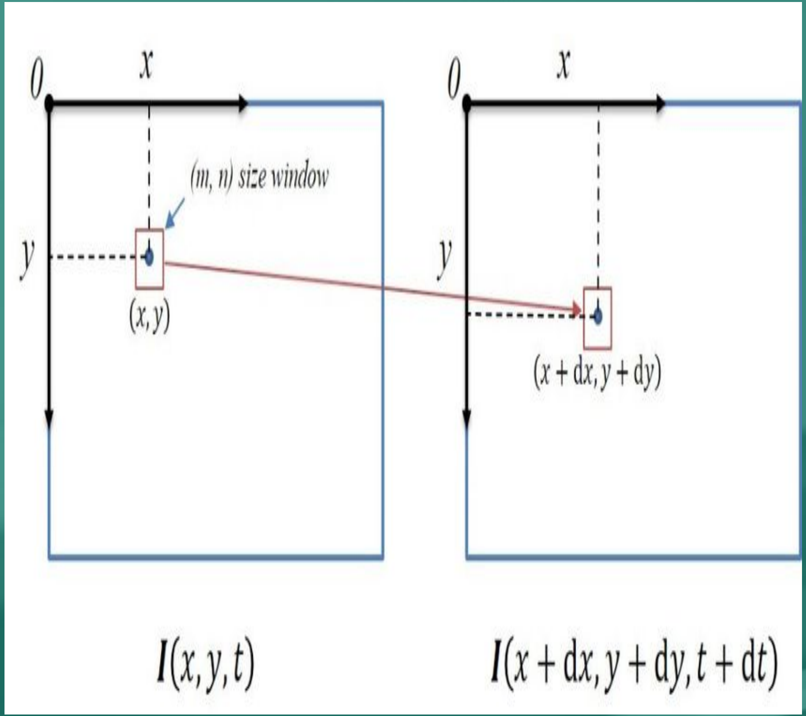


- ❖ L2 norm ($||w||_2$) of 3D points with respect to the pixel values computer by corner pixels tracking is used as the loss function.
- ❖ Ceres solver is used to solve the Bundle Adjustment problem.
- ❖ The cost function is described as follows:

$$F = \sum_{i=1}^{N_c}\sum_{j=1}^{N_p} \|p_{ij} - \pi(R_i P_j + T_i)\|^2,$$
$$= \sum_{i=1}^{N_c}\sum_{j=1}^{N_p} \left(\frac{e_{ij}^x + f_{ij}^x w_j}{c_{ij} + d_{ij} w_j}\right)^2 + \left(\frac{e_{ij}^y + f_{ij}^y w_j}{c_{ij} + d_{ij} w_j}\right)^2,$$

$$\begin{aligned}
\alpha_{ij}^x &= x_j - \theta_i^z y_j + \theta_i^y,\\
b_{ij}^x &= T_i^x,\\
\alpha_{ij}^y &= y_j - \theta_i^x + \theta_i^z x_j,\\
b_{ij}^y &= T_i^y,\\
c_{ij} &= -\theta_i^y x_j + \theta_i^x y_j + 1,\\
d_{ij} &= T_i^z,\\
e_{ij}^x &= p_{ij}^x c_{ij} - \alpha_{ij}^x,\\
f_{ij}^x &= p_{ij}^x d_{ij} - b_{ij}^x,\\
e_{ij}^y &= p_{ij}^y c_{ij} - \alpha_{ij}^y,\\
f_{ij}^y &= p_{ij}^y d_{ij} - b_{ij}^y. \quad (3)
\end{aligned}$$

KLT Tracking



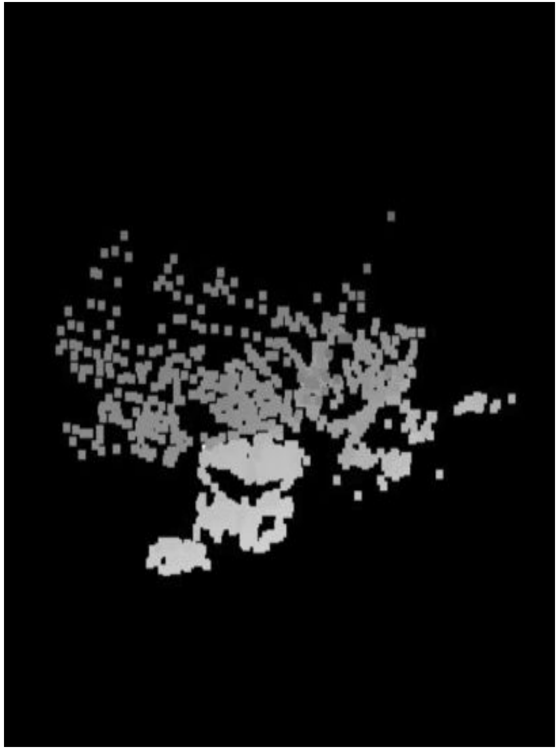
- KLT tracking is used to track features between all the frames.
- Then, find the Shi tomasi corners.
- Major difference between shi tomasi corners and harris corners lie in the change in scoring function.
  - $R = \min(\lambda_1, \lambda_2)$
- Corners can be filtered out by the homography matrix between the reference frame(initial frame) and every other frame in the video sequence.
- Corners that are inliers for more than 95 % of camera frames found by estimating homography matrix are chosen.
- Optical flow over all the images of the sequence is considered.

Results

OPTICAL FLOW

POINT CLOUD

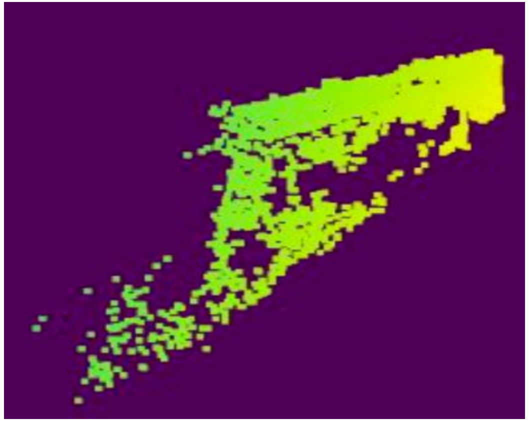
DEPTH MAP



OPTICAL FLOW

POINT CLOUD

DEPTH MAP



Progress Report

➔ We will follow the mentioned deadlines for our project deliverable :

- Extraction of Feature (2 weeks) - DONE
- Tracking of features (2 weeks) - DONE
- Bundle Adjustment (2 weeks) - DONE
- CRF energy minimization (1 week)
- Integration and results (1 week)

➔ We aimed to cover till Bundle adjustment for mid presentation and were successful in the same.

Thank You