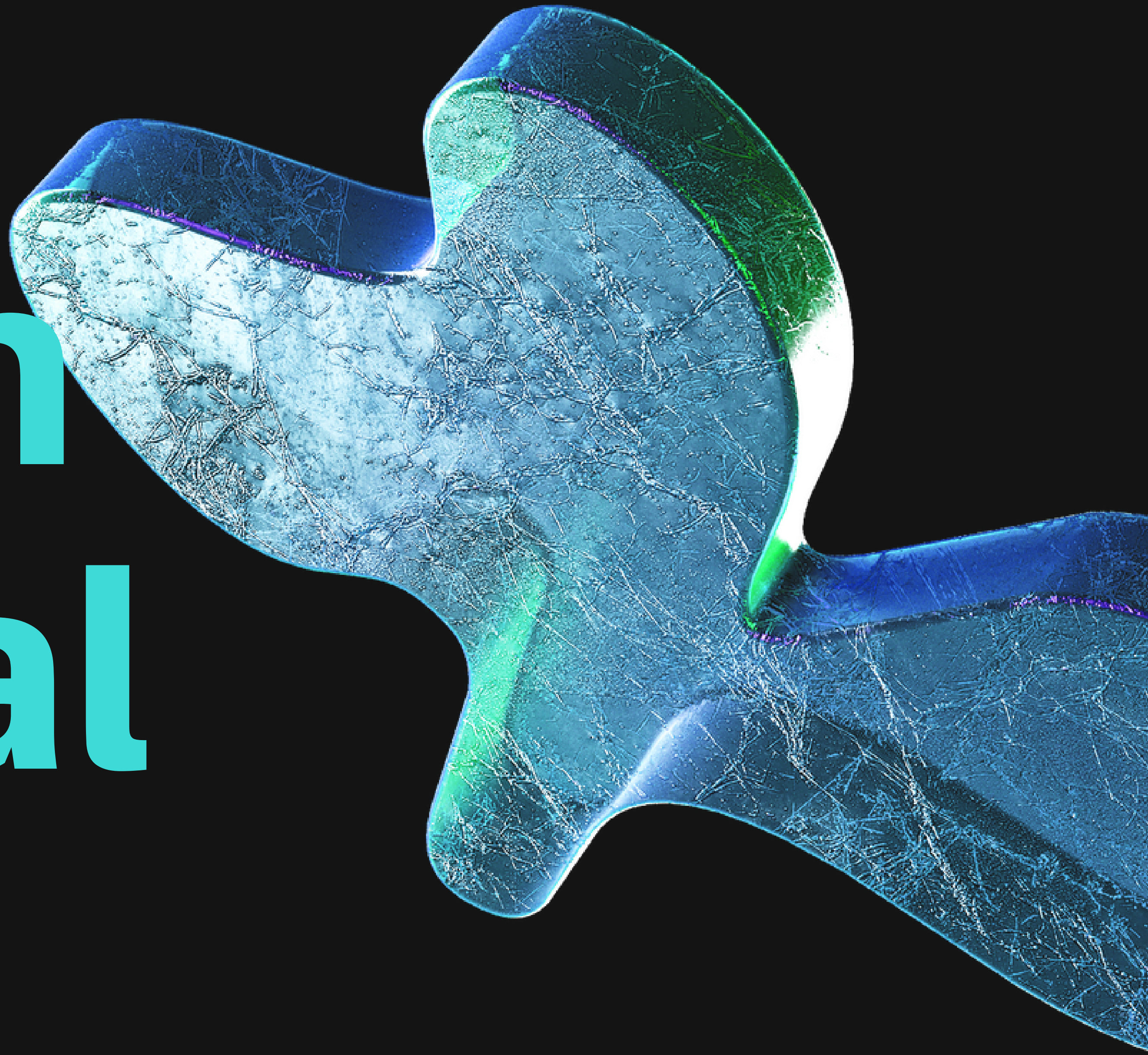
3D Reconstruction from Accidental Motion



TEAM 14

3D Reconstruction

SFM + Bundle Adjustment

- It usually takes a couple of seconds to click a picture

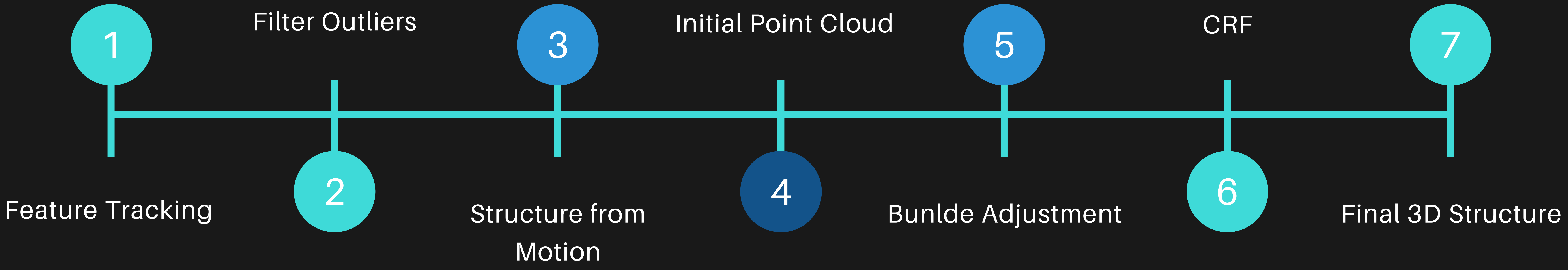
- During this time there is inevitable motion due to
  - Hand Shaking
  - Heart Beat

Problem Formulation

- Given Image Sequence of Nc Images, We apply KLT Tracking method to track the set of Np points. We assume each image is captured by a different camera with unique extrinsic parameters.
- We then use the Bundle Adjustment to estimate the ground-truth 3D world coordinates along with the extrinsic parameters of every camera using the initially tracked points.
- Using the estimated camera poses, the 3D scene is densely reconstructed as a single smooth depth map. A Conditional Random Field is used to minimise an energy function using plane-sweep approach and mean-field.

3D Reconstruction from Accidental Motion

Project Flow



Feature Tracking

We first track features between all the image frames using KLT tracking method.

Detect Shi-Tomasi features in the reference image using the scoring function:

$$R = min(\lambda_1, \lambda_2)$$

Estimate Lucas-Kanade optical flow over all the images in the sequence to find out good trackable features across all images.

Filter out corners by estimating the homography between the reference frame and every other frame in the image sequence.

Select those corners that are inliers for more than 95 % of images found by estimating homography

Bundle Adjustment

Given a set of features from the previous step, bundle adjustment is applied to simultaneously refine the 3D coordinates, the parameters of the relative motion, and the optical characteristics of the cameras using the reprojection error as a minimization criterion.

Reprojection error is the error between the projected 3D point on the image frame and the observed pixel.

Bundle adjustment optimizes for both 3D point locations and camera poses.

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

Dense Reconstruction

Each node is connected to every other node in Dense CRF. Each pixel here is a node of network.

Given an initial estimate of the 3D scene from the previous step, we construct a dense depth map of the scene from the reference view.

Since the depth signal at each pixel tends to be noisy in our case, we adopt plane sweeping together with CRF framework to solve a smooth depth map. Using this plane sweeping algo, we obtain photo consistency score.

A Dense CRF model is constructed with the photo-consistency term as the unary potentials and use a Gaussian kernel in an arbitrary feature space (spatial & intensity) as pairwise edge potentials.

In a Dense CRF, each node is connected to every other node. Each pixel here is a node of the network.

CRF : Conditional Random Feilds

- The energy function we need to minimise,

$$E(D) = Ep(D) + a * Es(D)$$

  - Ep(D): unary (photo-consistency) term
  - Es(D): Pairwise () term
  - D: Dense Depth Map that is optimised over.
- We formulate this problem of obtaining a dense depth map as a labelling problem where the depth values are the labels and each pixel is a node. The unary potential of every node gives an initial estimate of the depth locations.

Unary term

$$E_p(\mathrm{D}) = \sum_{i \in \mathcal{I}} \mathrm{P}(i, \mathrm{D}(i)),$$

- P(i,d) is the photo consistency score of the i th pixel at distance d.
- It can be obtained by the plane sweeping algorithm.

Pairwise Potential

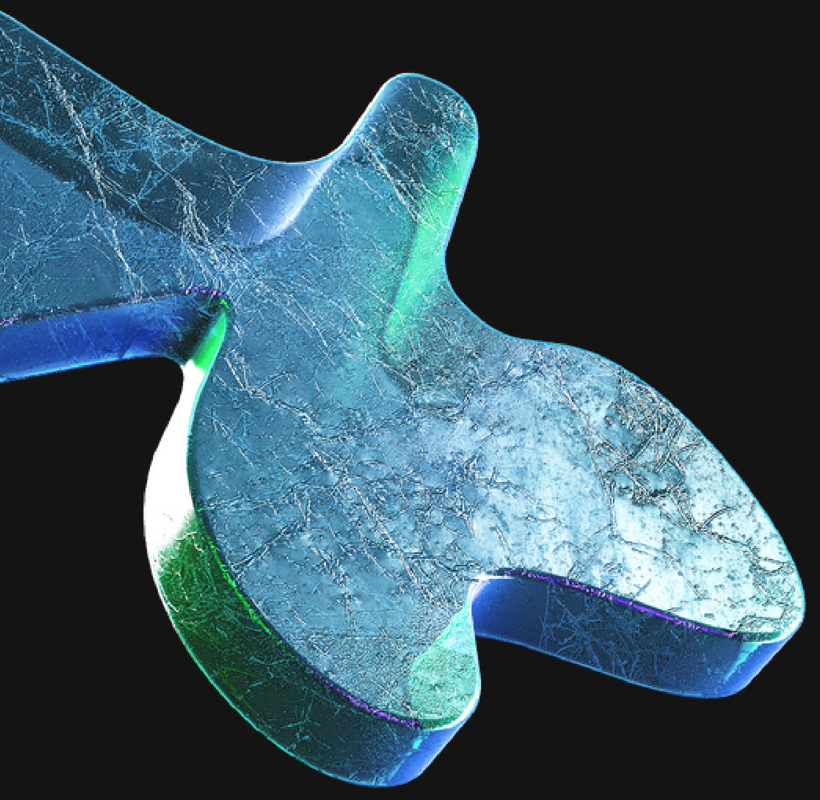
$$E_s(\mathrm{D}) = \sum_{i \in \mathcal{I}, j \in \mathcal{I}, i \neq j} \mathrm{C}(i, j, \mathrm{I}, \mathrm{L}, \mathrm{D}),$$

and

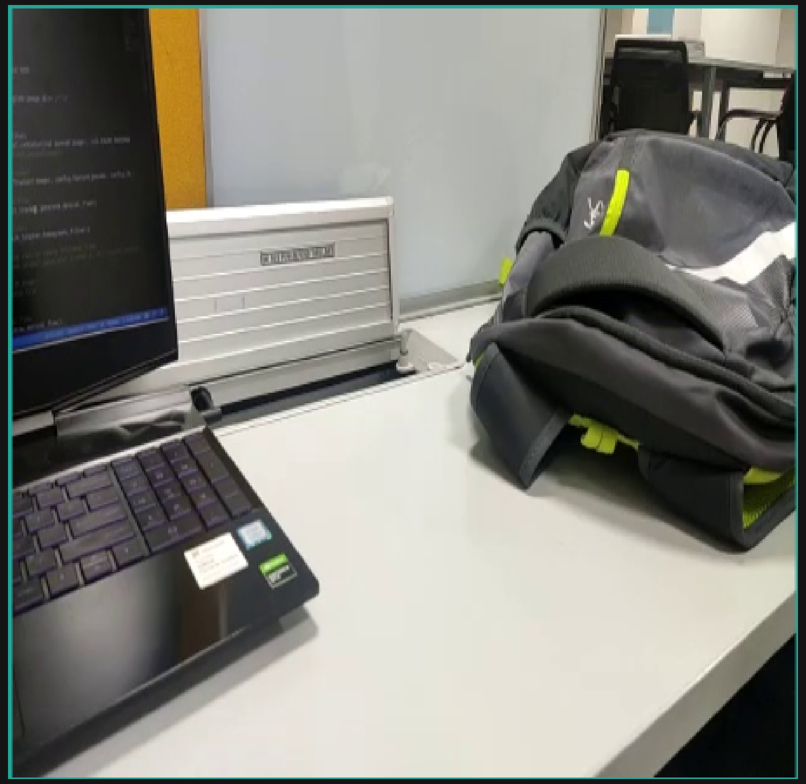
$$\mathrm{C}(i, j, \mathrm{I}, \mathrm{L}, \mathrm{D}) = \rho_c(\mathrm{D}(i), \mathrm{D}(j)) \times \exp\left(-\frac{||\mathrm{I}(i) - \mathrm{I}(j)||^2}{\theta_c} - \frac{||L(i) - L(j)||^2}{\theta_p}\right),$$

- C(i,j,I,L,D) gives a score for depth assignment of the i-th and j-th pixels based on the color intensities and locations in the reference images.
- pc is robust measurement of the depth difference. it is a truncated linear function

Results

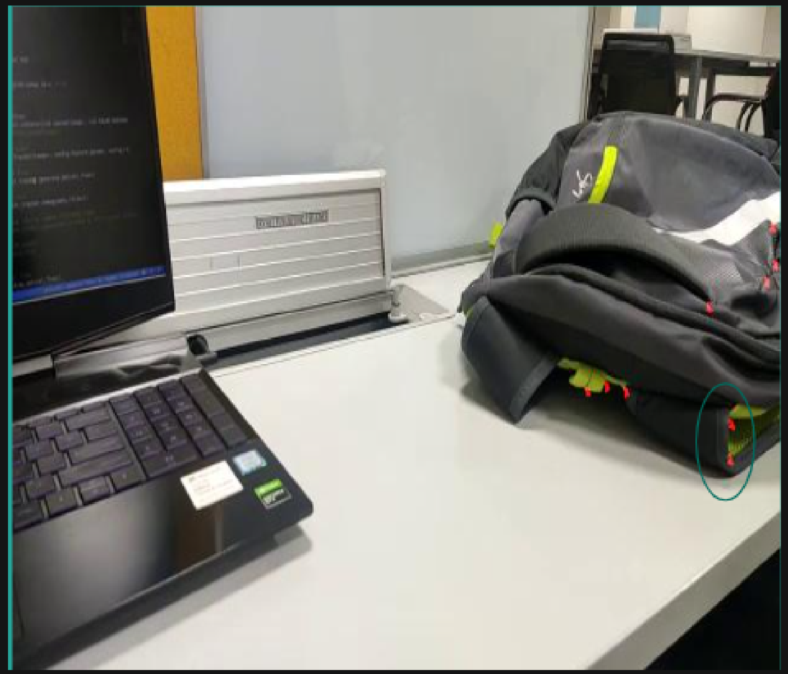


Original Image



Original Corners

Original outliers



Original Corner tracking

Ref Image

Ref Image depth map

Image

Image
30 frames

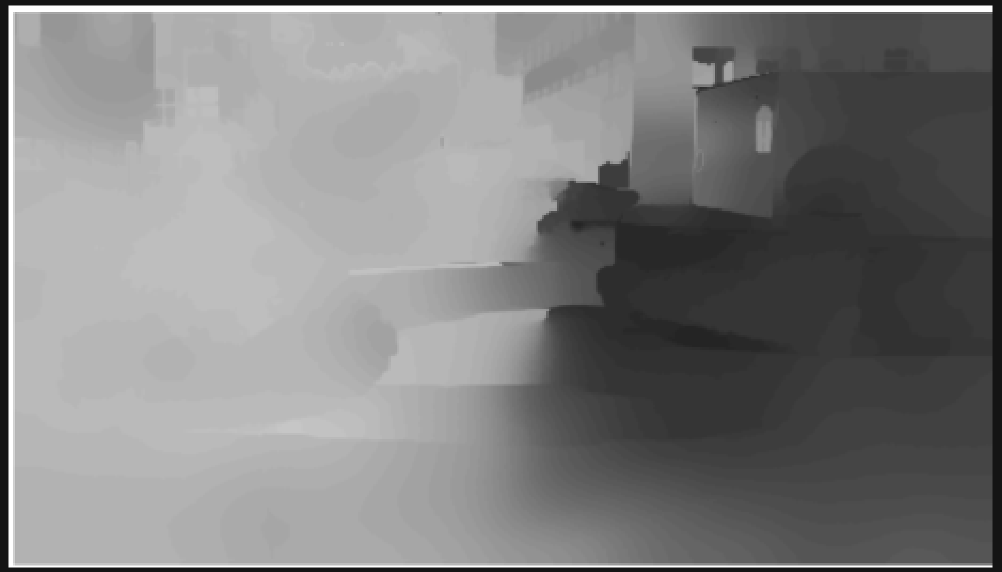


Image
50 frames

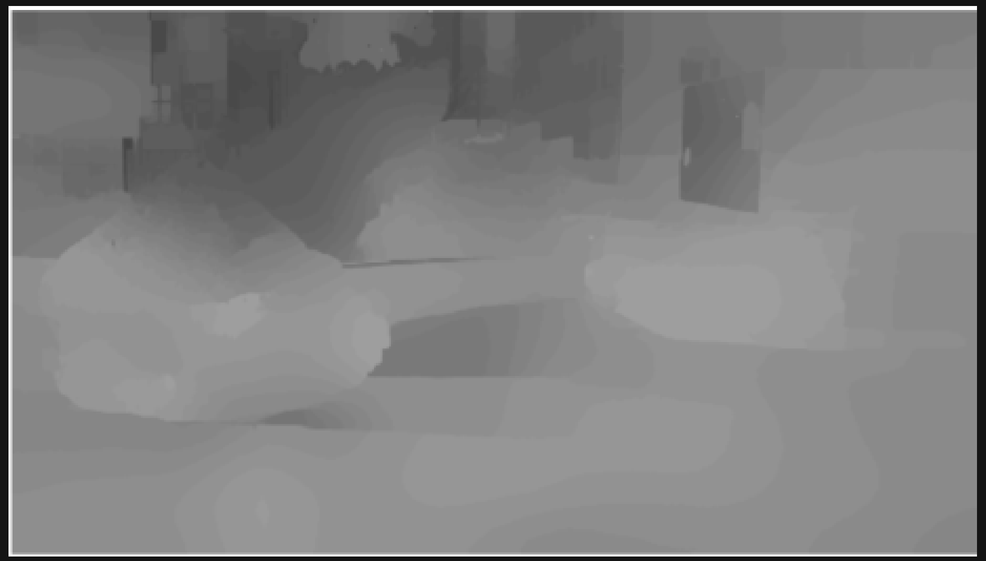
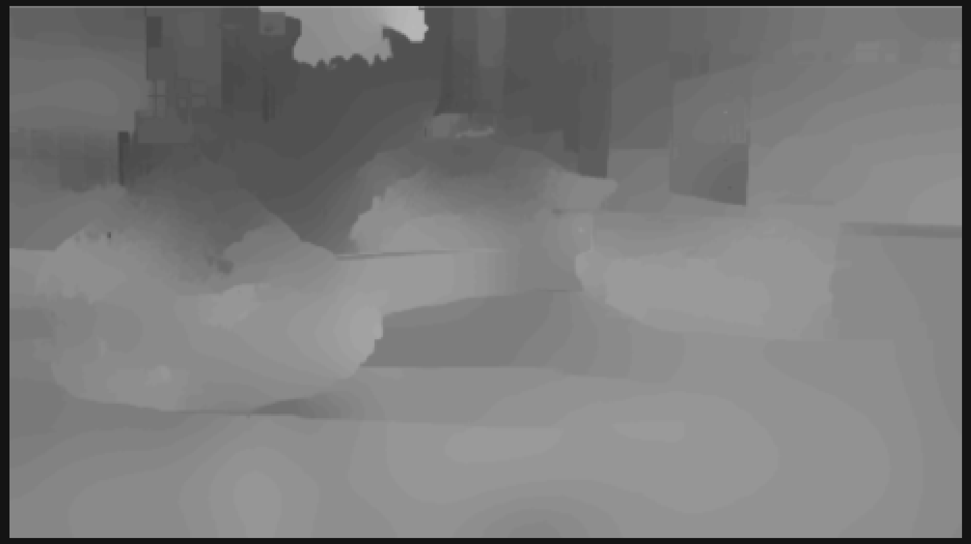


Image
100 frames



Image

Image
30 frames

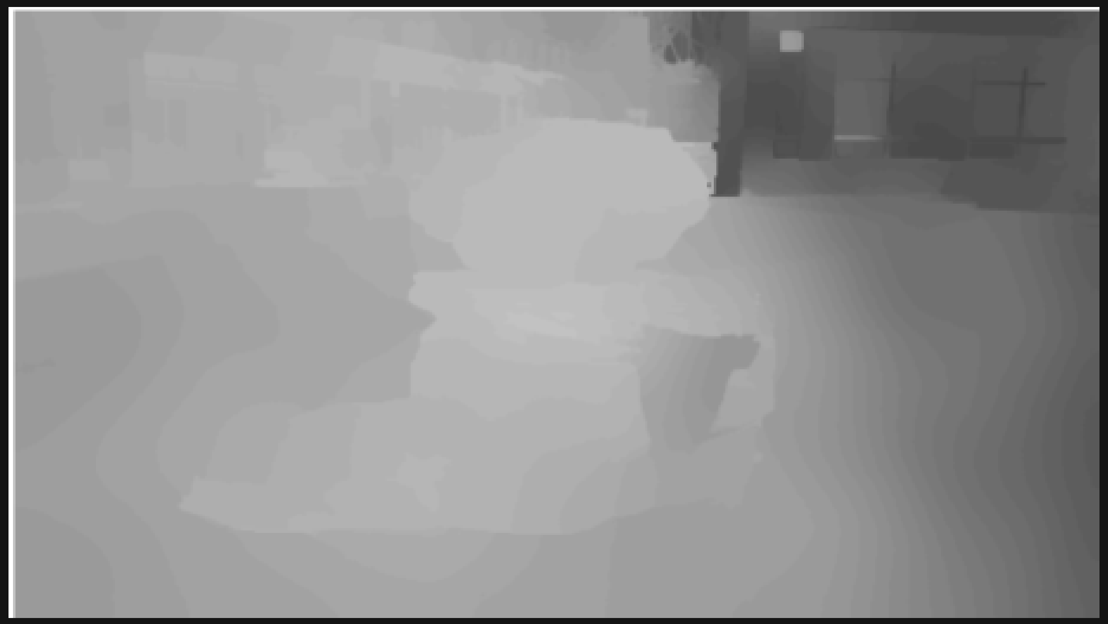


Image
50 frames

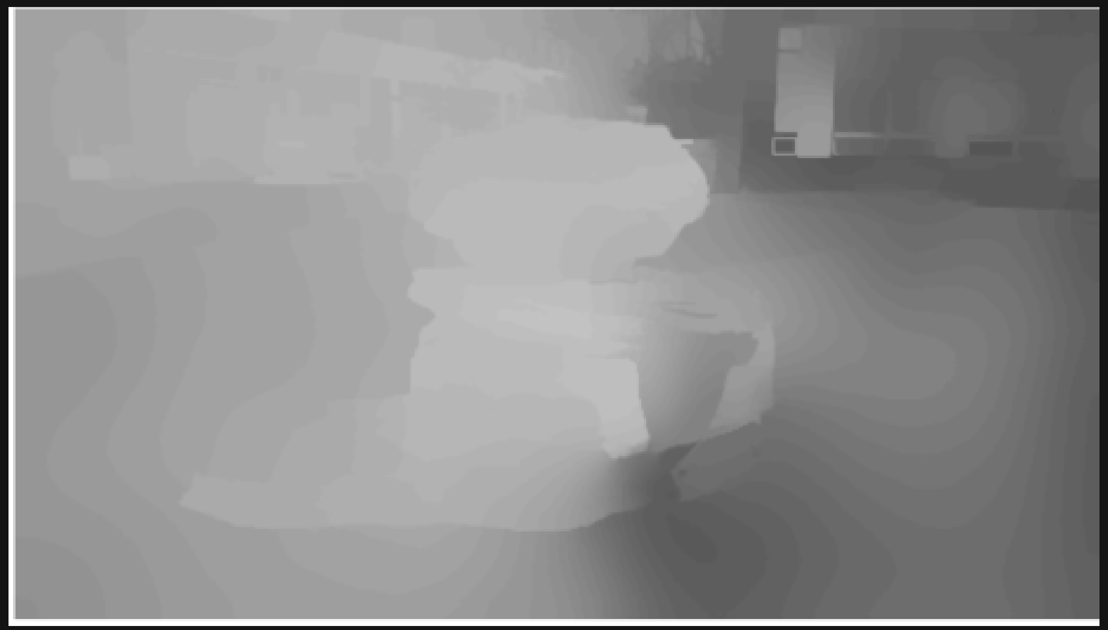


Image
100 frames