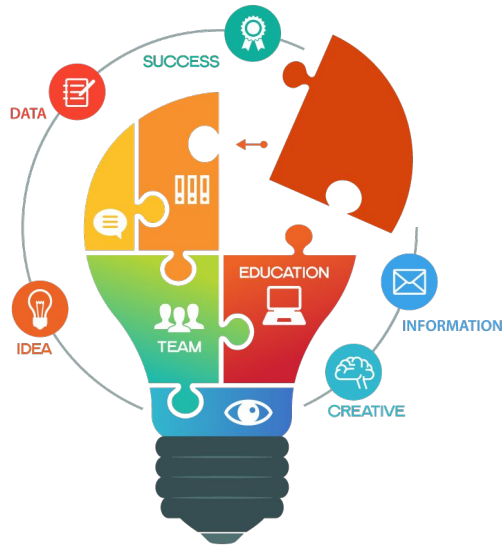


Team: Brown Munde

3D Reconstruction from Accidental Motion



Mentor : Meher Shashwat Nigam

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

Objective

We have implemented the following paper : [3D Reconstruction from Accidental Motion](#)

Our implementation takes the following parameters :

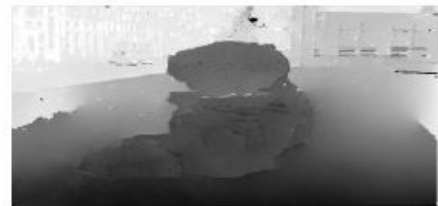
- **Input** : Sequence of frames of parts of video.
- **Result** : A 3 Dimensional reconstruction depth map of a reference frame



(a) Input image sequence



(b) Foreground points of SfM



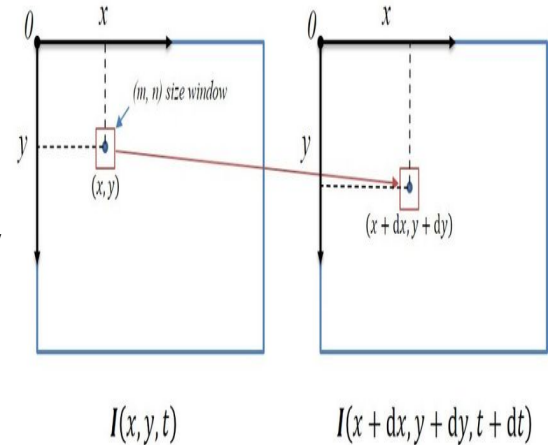
(c) Dense reconstruction

Problem Brief

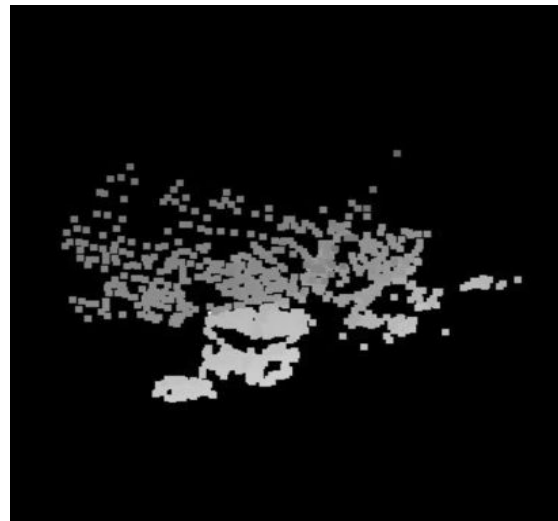
- ❖ We have an image sequence of N_c images and N_p projections (2D points) of corresponding 3D points as seen from every camera, we try to estimate the world coordinates of the real world points using **Bundle Adjustment**.
 - **Bundle Adjustment** : It refers to solving the location of pixels for a given estimated initial pose and location of 3D points.
- ❖ Using the estimated camera parameters, the 3D scene is densely reconstructed as a single depth map. A **Conditional Random Field Model** is used to minimize an energy function using plane-sweep approach and mean-field.

Method Overview : Tracking Features

- ❖ **KLT** tracking is used to track features between all the frames.
- ❖ The next step is to find the **Shi Tomasi Corners**.
- ❖ Major difference between **Shi Tomasi Corners** and **Harris Corners** lie in the change in scoring function : $\mathbf{R} = \min(\lambda_1, \lambda_2)$
- ❖ 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 90 % of camera frames found by estimating homography matrix are chosen.
- ❖ Optical flow over all the images of the sequence is considered.



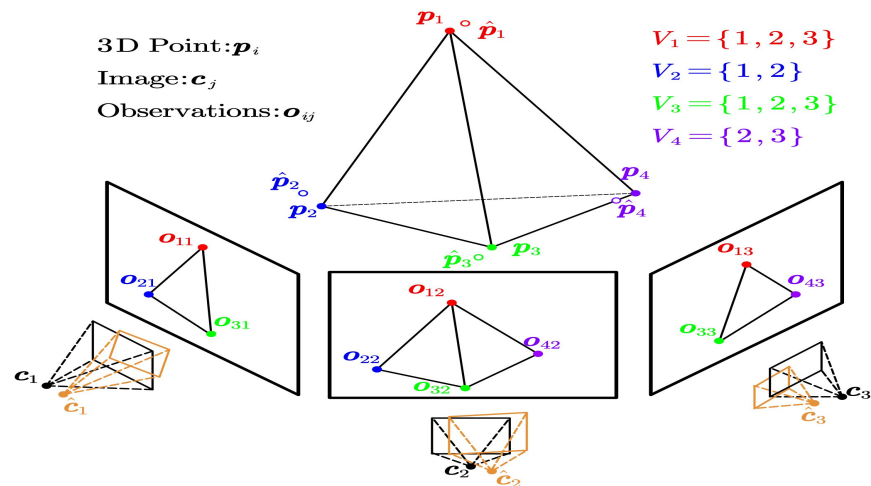
Method Overview : Tracking Features..



The point-cloud and the corresponding depth map of the image. The detected features are tracked across all non-reference images. We show here the trajectory of each feature across the set of frames.

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:



$$F = \sum_{i=1}^{N_c} \sum_{j=1}^{N_p} \|p_{ij} - \pi(R_i P_j + T_i)\|^2,$$

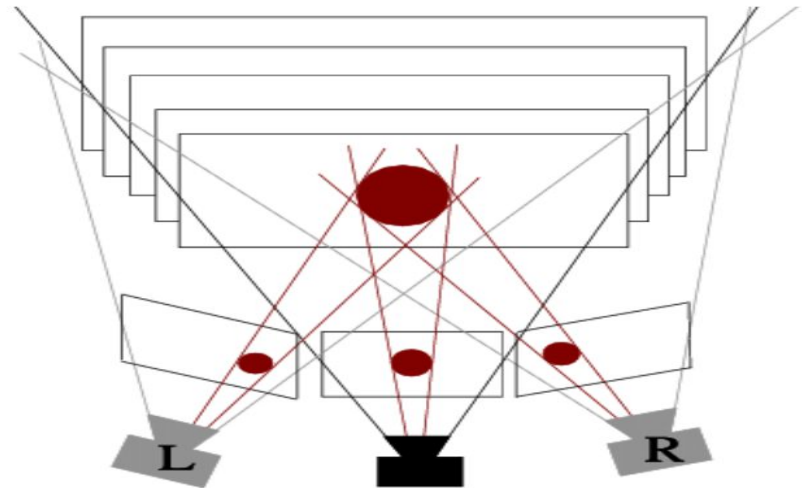
$$= \sum_{i=1}^{N_c} \sum_{j=1}^{N_p} \left(\frac{e_{ij}^x + f_{ij}^x w_j}{c_{ij} + d_{ij} w_j} \right)^2 + \left(\frac{e_{ij}^y + f_{ij}^y w_j}{c_{ij} + d_{ij} w_j} \right)^2,$$

where

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

Dense Conditional Random Field

- ❖ After estimating the sparse 3D structure of the scene along with camera extrinsics, we want to construct a **dense depth map** of the 3D scene using a **conditional random field**.
- ❖ We use the **Plane-Sweeping** along with a CRF framework to solve for a dense depth map. The plane-sweeping method generates unary potentials for the CRF model. Using a fully-connected CRF model allows pixel connections with longer range so that the **photo-consistency** measurement can be effectively aggregated from an area to a pixel in it.



Unary Potential Calculations

- ❖ We first sample different depth values in the acceptable depth range of the 3D scene. To calculate the unary potentials we compute a homography between the reference image and every other image. The homography matrix between the reference image and jth camera frame at these different depths D is computed as follows:

$$H_j^{ref}(D) = D * K * {}^C_W R_{ref} * {}^C_W R_j^{-1} * K^{-1}$$

$$H_j^{ref}(D)[:, 2] = K * {}^C_W R_{ref} * (C_j - C_{ref})$$

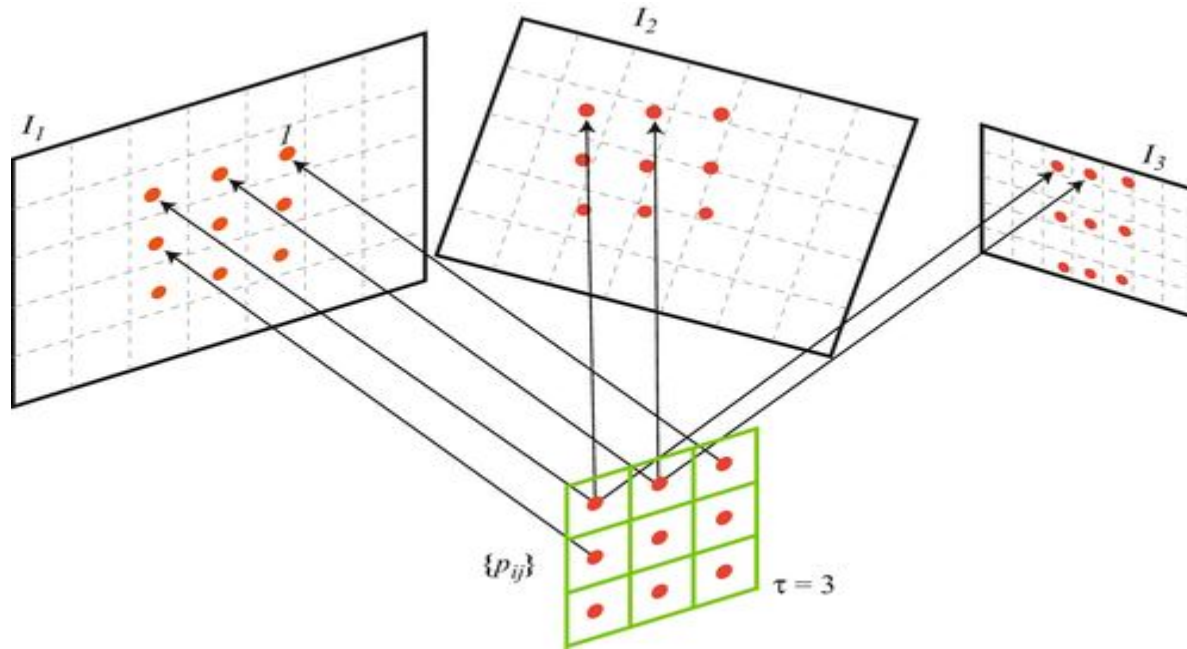
Here D, R_{ref} , K, and C are the depth, rotation matrix, camera intrinsic matrix, and camera centers respectively.

- ❖ E_p is the photo-consistency term defined as an L1 Loss between small patches in the reference image and the warped image from another viewpoint. It can be expressed as:

$$E_p(D) = \sum_j \sum_i \left| p_{i,ref} - H_j^{ref}(D) * p_{i,j} \right|$$

$p_{i,ref}$ and $p_{i,j}$ are the gray-scale patches of the reference camera and j th camera respectively.

Unary Potential Calculations



Pair-wise Potential Calculations

- ❖ The pairwise potential is a Gaussian kernel in arbitrary feature space.
- ❖ It has a spatial term such that the depth within a small neighbourhood is consistent.
- ❖ It has an intensity term such that pixels within an area with similar colours have consistent depth, since they are likely to belong to the same object.

$$E_s(D) = \sum_{i \in \mathcal{I}, j \in \mathcal{I}, i \neq j} C(i, j, I, L, D) \quad C(i, j, I, L, D) = \rho_c(D(i), D(j)) \times \exp \left(- \underbrace{\frac{\|I(i) - I(j)\|^2}{\theta_c}}_{\text{Intensity term}} - \underbrace{\frac{\|L(i) - L(j)\|^2}{\theta_p}}_{\text{Spatial term}} \right)$$

Here $\rho_c(\cdot)$ is the truncated linear function defined as **pc = min(t, |D(i) - D(j)|)** with some threshold t . The pairwise term has a spatial term such that depth within a small neighborhood is consistent as well as a intensity term such that pixels within an area with similar colors have consistent depth, since they are more likely to belong to the same object.

Pair-wise Potential Calculations

Bilateral Filter

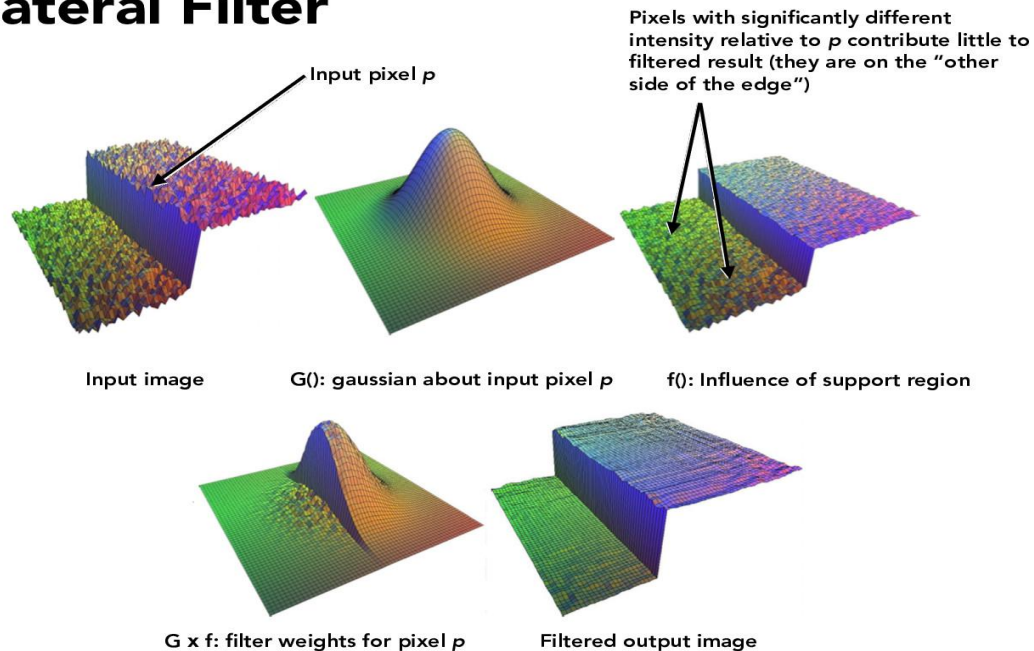
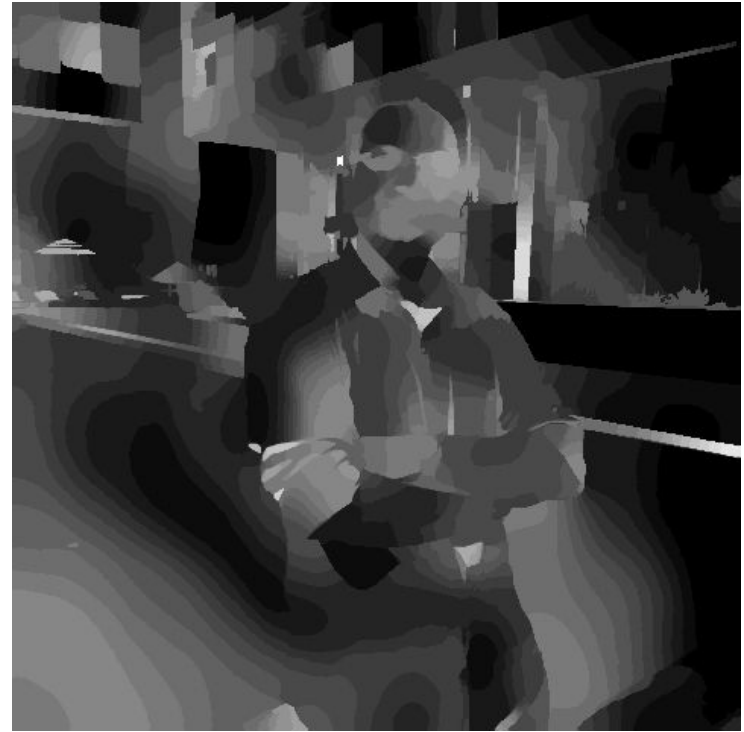


Figure credit: Durand and Dorsey, "Fast Bilateral Filtering for the Display of High-Dynamic-Range Images", SIGGRAPH 2002

Results



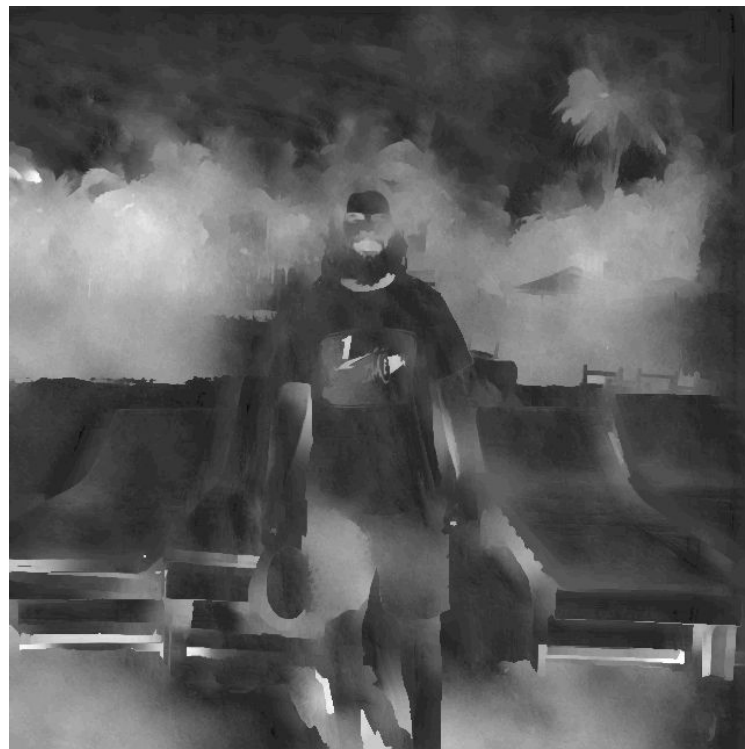
Optical Flow



Dense Map



Optical Flow



Dense Map



Optical Flow



Dense Map

Experiments

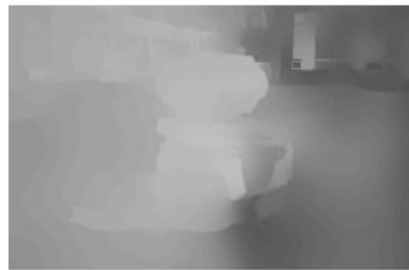
Experiment 1 : Number of Images



(a) Optical flow



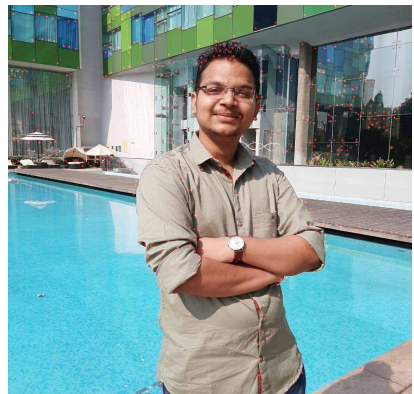
(b) 30 frames



(c) 50 frames



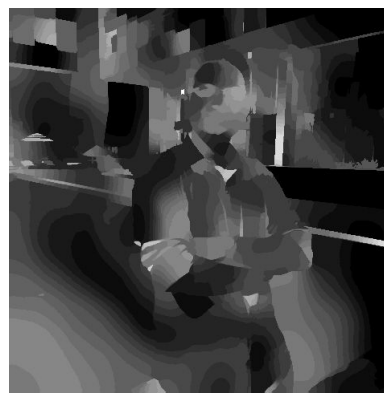
(d) 100 frames



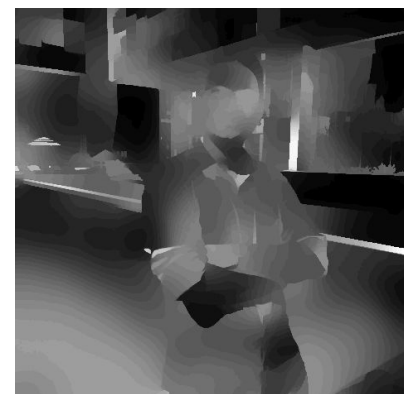
(a) Optical flow



(b) 30 frames



(c) 50 frames



(d) 100 frames

Experiment 2 : Weight for CRF



(a) Optical flow



(b) $w = .5$



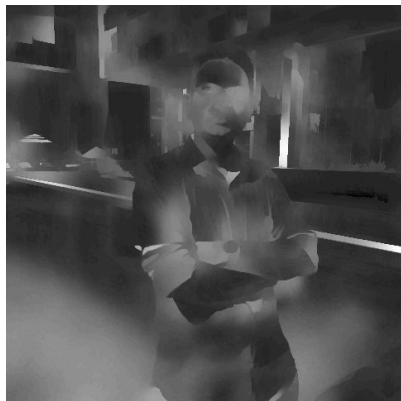
(c) $w = 1$



(d) $w = 2$



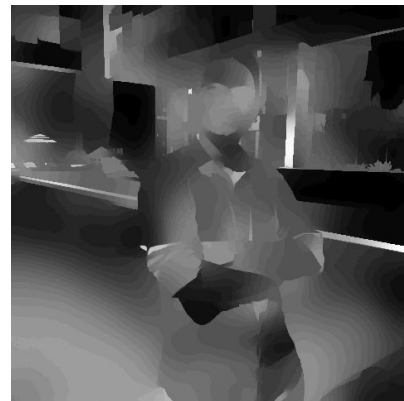
(a) Optical flow



(b) $w = .5$



(c) $w = 1$



(d) $w = 2$

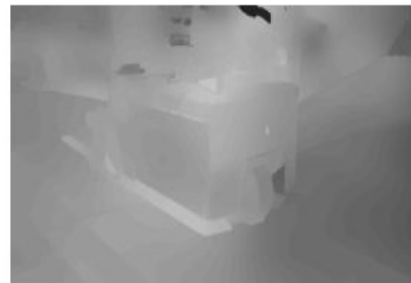
Experiment 3 : Intensity STD for potential



(a) Optical flow



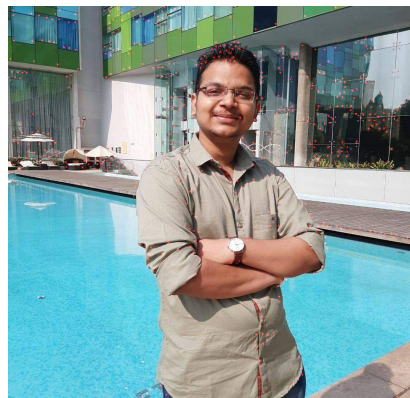
(b) STD = 10



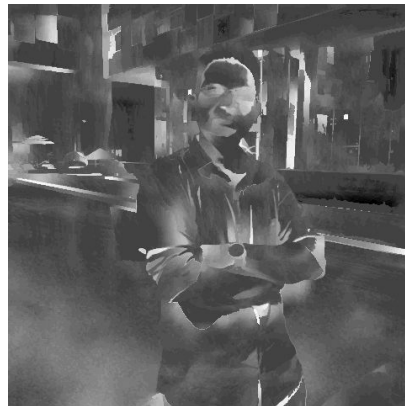
(c) STD = 20



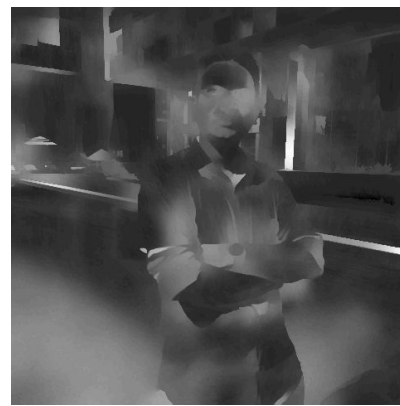
(d) STD = 30



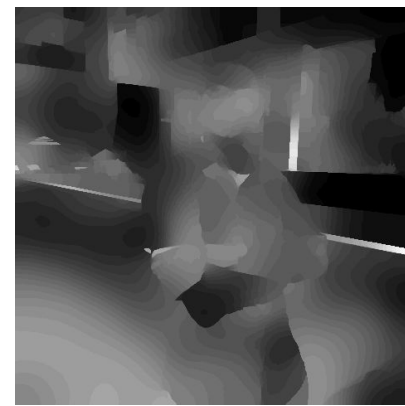
(a) Optical flow



(b) STD = 10



(c) STD = 20

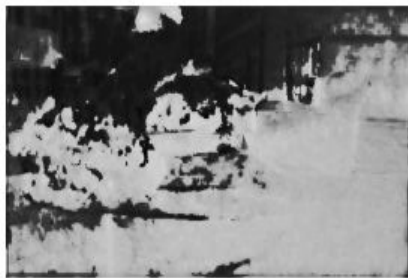


(d) STD = 30

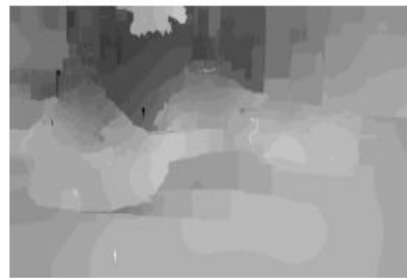
Experiment 4 : Max Penalty for CRF



(a) Optical flow



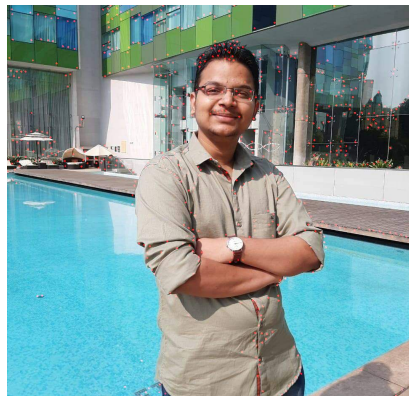
(b) Max_P = .5



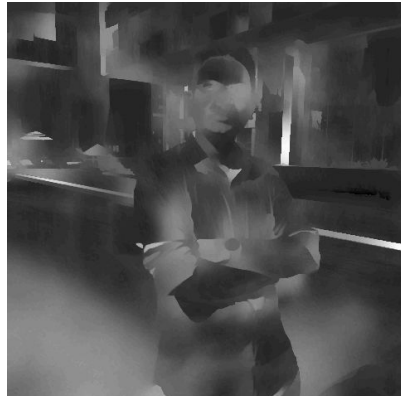
(c) Max_P = .1



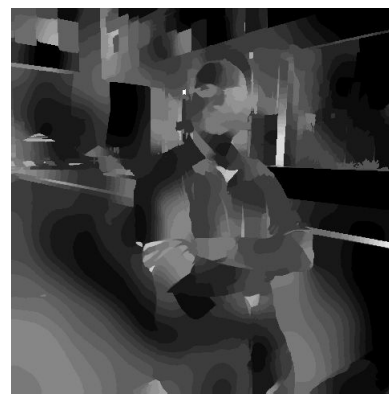
(d) Max_P = .2



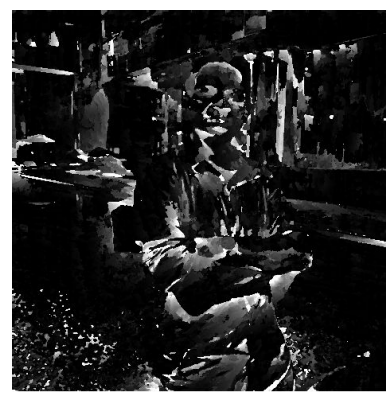
(a) Optical flow



(b) Max_P = .5



(c) Max_P = .1



(d) Max_P = .2

Experiment 5 : Patch Radius for Unary



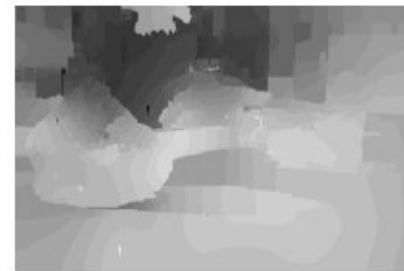
(a) Optical flow



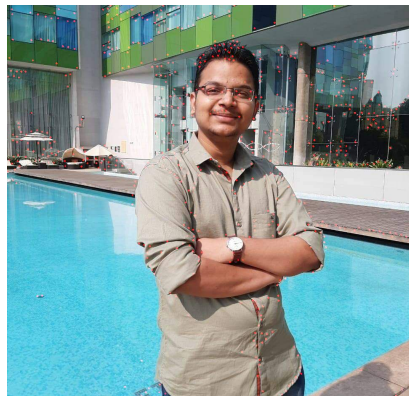
(b) $R = 1$



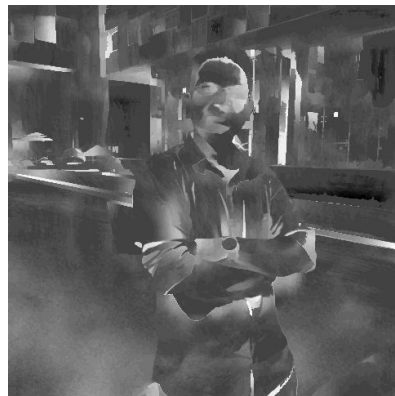
(c) $R = 2$



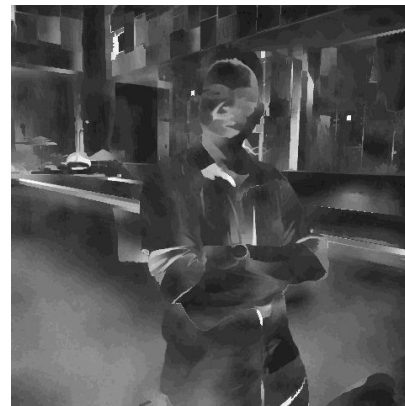
(d) $R = 3$



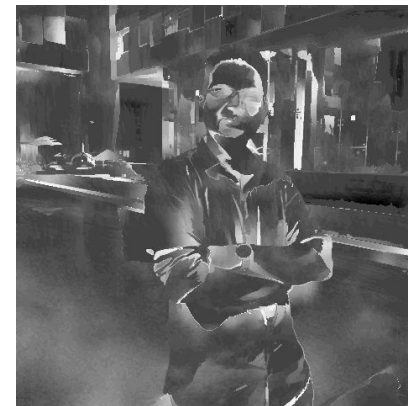
(a) Optical flow



(b) $R = 1$



(c) $R = 2$



(d) $R = 3$

Drawbacks

- ❖ More theoretical analysis required to work with ambiguities of multiview reconstruction with small motion.
- ❖ More information about the method of capture required.
- ❖ Success rate to be analysed for usage under different conditions.
- ❖ Exact conditions for the optimal working of the method to be worked out.

Thank You