3D RECONSTRUCTION FROM ACCIDENTAL MOTION

PROJECT REPORT - COMPUTER VISION

AadilMehdi Sanchawala
aadilmehdi.s@students.iiit.ac.in

Rahul Sajnani
rahul.sajnani@research.iiit.ac.in

Rohan Chacko
rohan.chacko@students.iiit.ac.in

ABSTRACT

We tackle the problem of 3D Reconstruction from *accidental motion* of the photographer. Given the initial few frames of a video or series of burst photos, we aim to reconstruct a dense depth map of the scene from bundle adjustment. We further apply a CRF model to regularize the depth to provide a smooth depth map from a reference view. We demonstrate the results of bundle adjustment for a few scenes.

Keywords 3D Reconstruction · Bundle Adjustment · Multi-View Stereo · CRFs

1 Brief Overview

We implement the paper *3D Reconstruction from accidental motion*[1] which aims to reconstruct a 3D scene from the accidental motions of a photographer. Accidental motion is defined as the inevitable motion that occurs when trying to hold a camera still. Given a series of N images, we consider the first image N_0 as the reference image. The final depth map is given w.r.t. the reference view. The paper implements the following pipeline to perform the above task :

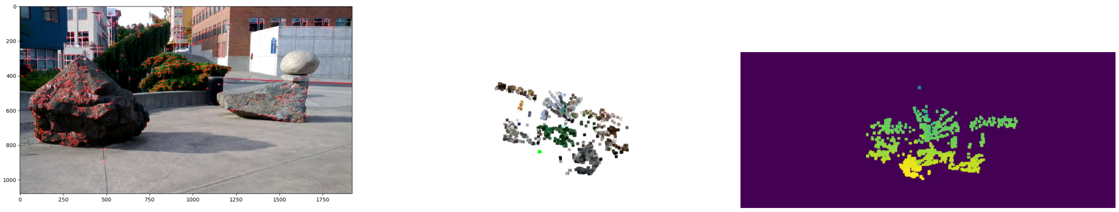
- Extract good features using the *Tomasi-Kanade*[2] method.
- Track the detected features using the *Lucas-Kanade*[3] method w.r.t the reference image N_0
- We use these tracked features to perform bundle adjustment on the set of N frames to estimate the 3D structure of the scene
- A dense map is reconstructed from the sparse 3D structure using a CRF model [4] which incorporates a *photo-consistency* loss and a *smoothness* loss. The final output is the depth map from the reference view.

2 Detect & Track Features

We use the KLT Tracker to track features across the $N-1$ frames w.r.t to the reference frame. This step can be broken into two steps: (i) Feature Detection using [2] (ii) Feature Tracking using [3].

Tomasi-Kanade method for feature detection uses the eigenvalues (λ_i's) of the Hessian matrix. The Hessian matrix considers the image intensities around a small square patch in the image. Based on the values of λ_1 and λ_2, we choose whether to consider the point or not as $\min(\lambda_1, \lambda_2) > \lambda$.

Lucas-Kanade method for feature tracking uses optical flow to estimate the displacement of the point from the reference image to the other frames. The KLT algorithm requires that all features be tracked to all the non-reference images. Another method to filter the patches is to choose only those patches that have a maximum color gradient difference per pixel below a threshold w.r.t. the reference image.



(a) (b) (c)

Figure 1: 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.

3 Bundle Adjustment

Given a set of images depicting a number of 3D points from different viewpoints, bundle adjustment can be defined as the problem of simultaneously refining the 3D coordinates describing the scene geometry, the parameters of the relative motion, and the optical characteristics of the camera(s) employed to acquire the images, according to an optimality criterion involving the corresponding image projections of all points.

3.1 Modelling bundle adjustment as an Optimisation Problem

Bundle adjustment boils down to minimizing the re-projection error between the image locations of observed and predicted image points, which is expressed as the sum of squares of a large number of nonlinear, real-valued functions. Thus, the minimization is achieved using nonlinear least-squares algorithms. Of these, Levenberg–Marquardt has proven to be one of the most successful due to its ease of implementation and its use of an effective damping strategy that lends it the ability to converge quickly from a wide range of initial guesses.

By iteratively linearizing the function to be minimized in the neighborhood of the current estimate, the Levenberg–Marquardt algorithm involves the solution of linear systems termed the normal equations. When solving the minimization problems arising in the framework of bundle adjustment, the normal equations have a sparse block structure owing to the lack of interaction among parameters for different 3D points and cameras.

3.2 Mathematical Modelling

The reprojection error minimisation for the problem can be written as shown in Figure 2. The Jacobian for the residuals is taken be differentiating the residuals with the rotation, and the translation parameters of the camera and the 3D position of the point. The Jacobian has the following structures as shown in Figure 3.

3.3 Problem formulation particular to our use case

As per the authors, we model the bundle adjustment problem by the following initialisation,

- The Rotation parameters for each view is set to Identity matrix.
- The Translation parameters for each view is set to the origin that is $\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$ in homogeneous coordinates.
- The 3D world points are initialised with their X and Y coordinates as the reference view's (In our case the first image) pixel coordinates u and v respectively. The depth for the points is randomly initialised between 2 to 4 meters. The points are parameterised by their inverse depth.

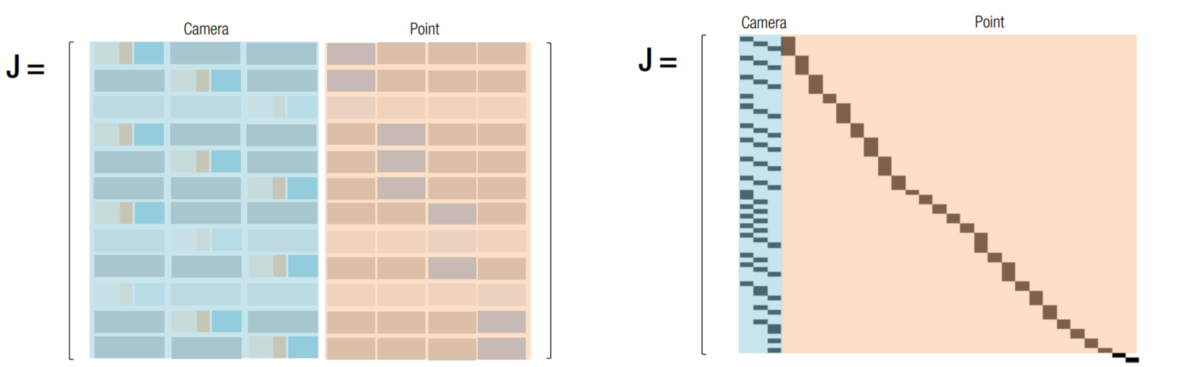
As for the residuals, we obtain them by tracking the optical flow of the obtained feature points as described in Section 2. Therefore we obtain, the location of the feature points in all the views. Using that we initialise the Jacobian's residuals.

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

where

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

Figure 2: Cost function for Bundle Adjustment



(a) (b)

Figure 3: Jacobian structure

$$\mathbf{R}_i = \begin{bmatrix} 1 & -\theta_i^z & \theta_i^y \\ \theta_i^z & 1 & -\theta_i^x \\ -\theta_i^y & \theta_i^x & 1 \end{bmatrix} \quad \mathbf{P}_j = \frac{1}{w_j}[x_j, y_j, 1]^T, \text{ where } (x_j, y_j) \text{ is the projection}$$

(a) (b) (c)

Figure 4: (a) Initialization of Rotation matrix (b) Inverse depth point initialization

(a) (b)

Figure 5: Results of point-cloud with its corresponding original image overlaid with optical flow.

(a) (b)

Figure 6: Results of point-cloud with its corresponding original image overlaid with optical flow.

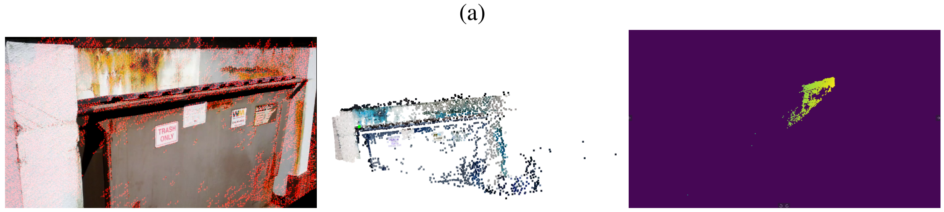
3.4 Solving the Bundle Adjustment Problem

The cost function of bundle adjustment with the assumption of small motion (both rotation and translation) yields a convex problem. When the camera poses are fixed, it is convex to get the depth of a feature relative to a reference view. Also, it is convex to optimize the rotation for the points at infinity when an approximation is used.

We optimize the cost function of bundle adjustment in Figure 2 with Ceres Solver. In the following subsection we show the results obtained after solving the BA problem.

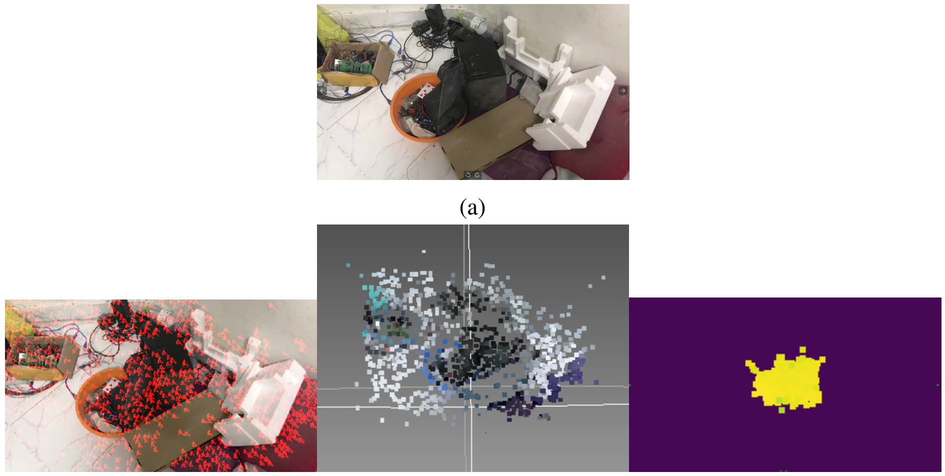
3.5 Results

(a)



(b) (c) (d)

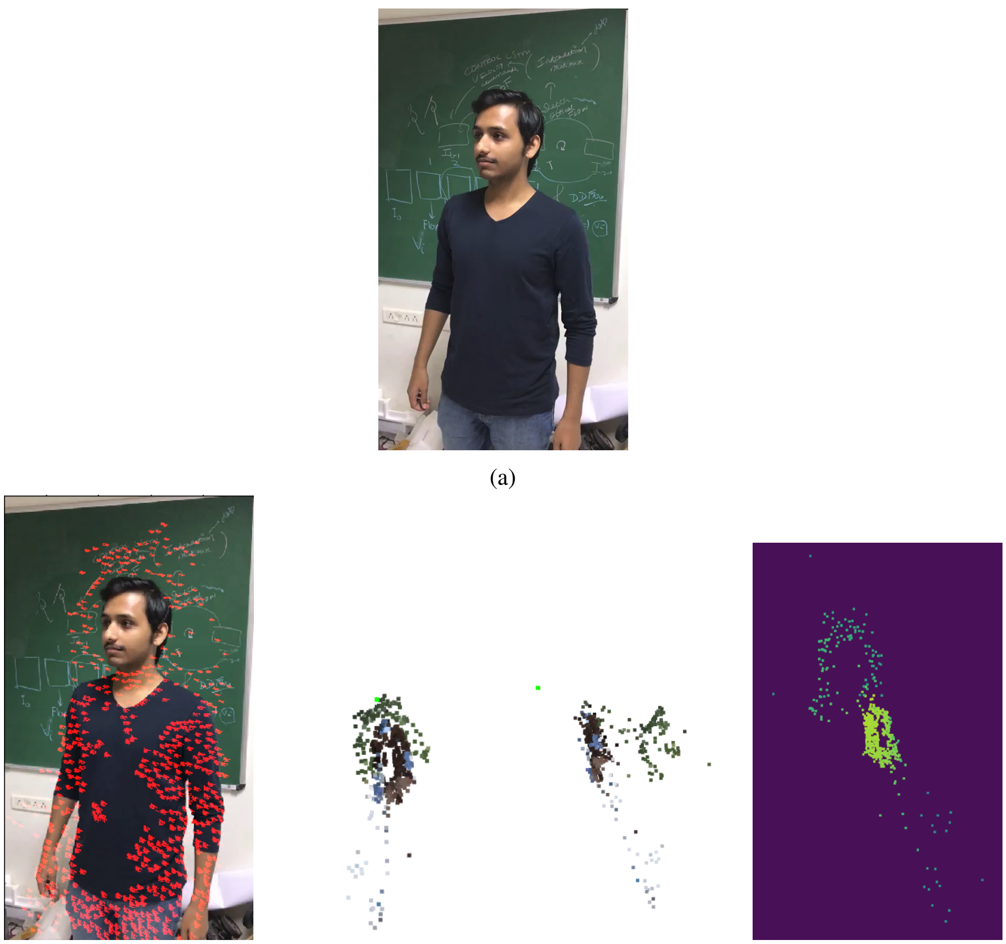
Figure 7: More results. (a) Original image (b) Optical Flow (c)Point-cloud (e) Generated depth-map (colorized)



(a)

(b) (c) (d)

Figure 8: More results. (a) Original image (b) Optical Flow (c),(d) Point-cloud (e) Generated depth-map (colorized)



(a)

(b) (c) (d) (e)

Figure 9: More results. (a) Original image (b) Optical Flow (c),(d) Point-cloud (e) Generated depth-map (colorized)

References

[1] Fisher Yu and David Gallup. 3d reconstruction from accidental motion. In *27th IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[2] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. 1991.

[3] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, page 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[4] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS'11, page 109–117, Red Hook, NY, USA, 2011. Curran Associates Inc.