3D Reconstruction using smart-phone sensors

Suvam Patra
Under supervision of
Prof. Subhashis Banerjee and Prem Kalra

Indian Institute of Technology, Delhi

suvam@cse.iitd.ac.in

December 9, 2014

Overview

Introduction

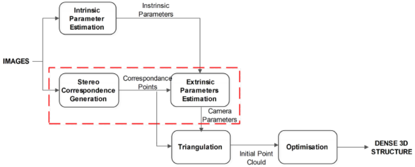
- Calibration of camera is an essential part for recovering 3D geometry from multi-view.
- Two kind of calibration are needed.
  - Internal calibration, which estimate camera focal length, and principle point.
  - External calibration, which estimate the camera rotation and position with respect to some world coordinate.
- Internal calibration matrix K is obtained using traditional method like Zhang's calibration.

Conventional calibration and structure estimation

- Conventional method for computing external calibration:
  - In a multi-view set-up essential matrices (E) are computed between pair of views in a RANSAC framework.
  - Essential matrix is further decomposed into rotation (R) and translation (t) using relationship

$$E = R[t]_x$$

- Use the estimated internal and external calibration matrices to triangulate the refined 2D image correspondences.
- Simultaneously optimise the camera calibration and 3D structure in a Bundle adjustment framework.



Motivation

- Pairwise epipolar geometry, i.e estimating external calibration takes most of the time needed for 3D reconstruction in multi view set-up.
- The time could be minimised if some prior regarding the rotation and translation is available during image capture.
- Smart phones are equipped with three additional inertial sensors:
  - Accelerometer
  - Gyroscope
  - Magnetometer.

Motivation

- The IMU sensors can be used for real time estimation of camera calibration parameters.
- **Accelerometer** provides real time feed on acceleration of the device once a thrust is given.
- **Gyroscope** provides the direction of the Earth's gravitational field.
- **Magnetometer** provides the direction of the Magnetic North.

Sensor Fusion

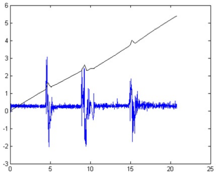
- All three sensors are synchronised to a single clock for obtaining synchronous data.
- The accelerometer in conjunction with the gyroscope provides the unit gravitational vector $\mathbf{r}_z$ which is considered as Z axis.
- The Magnetic North direction $\mathbf{r}_y$ is considered as Y axis.
- The X axis direction $\mathbf{r}_x$ is computed using cross product $\mathbf{r}_x = \mathbf{r}_y \times \mathbf{r}_z$

If $\mathbf{r}_x = [v_{11}\ v_{12}\ v13]$; $\mathbf{r}_y = [v_{21}\ v_{22}\ v_{23}]$; $\mathbf{r}_z = [v_{31}\ v_{32}\ v_{33}]$, rotation matrix of the camera is determined as:

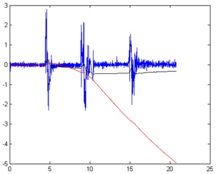
$$R = (\mathbf{r}_x^T\ \mathbf{r}_y^T\ \mathbf{r}_z^T)$$

Refinement of acceleration for translation estimation I

- Smart-phones accelerometer works on thrust/force applied on the device.
- Device provides acceleration in all three dimensions ($\mathbf{f}_x$ $\mathbf{f}_y$ $\mathbf{f}_z$).
- Due to sensor errors, the device does not produce zero acceleration even if in static condition.
- A static bias B is computed from the device keeping it in rest condition and is subtracted from acceleration.

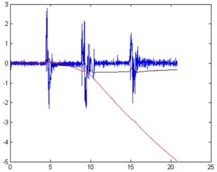


(b) Without static bias correction.

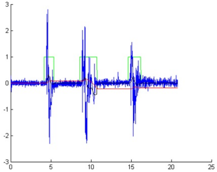


(c) After static bias correction.

- As smart phones provides continuous data for acceleration, we need to identify the region of actual movement in acceleration data to estimate position correctly.
- Without detection of actual movement the position drift a lot from its actual values.



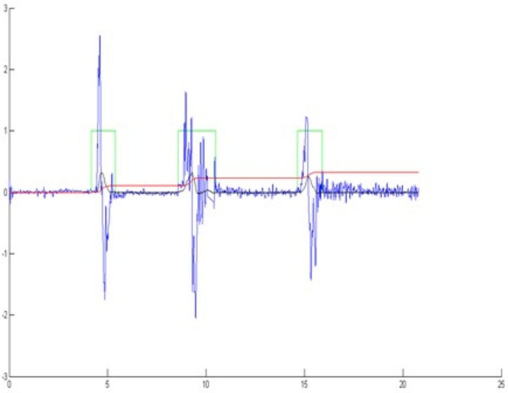
(d) Before activity region detection



(e) After activity region detection (in green)

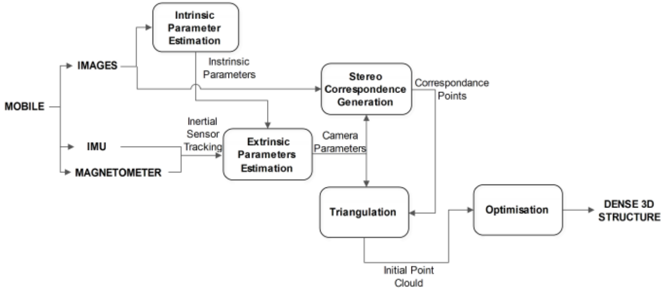
Refinement of acceleration for translation estimation III

- The acceleration data is now smoothed using median filter and position is estimated using Newton's law of motion.



(f) Position estimation in red after refinement.

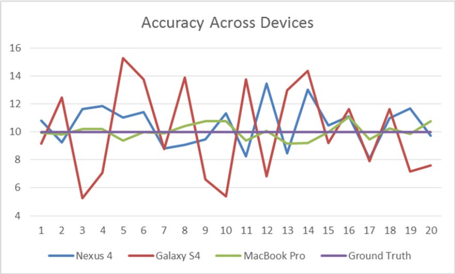
- Use the point correspondences from tracking like KLT.
- Use the K, and pairwise rotations R and translations $\mathbf{t}$ estimated from the sensors to calculate the fundamental matrix $F = K^{-T}[\mathbf{t}]_{\times}RK^{-1}$
- Refine point correspondences $\mathbf{x}_i$ and $\mathbf{x}_j$ s.t. $\mathbf{x}_i^T F \mathbf{x}_j < \epsilon$.
- Triangulate to obtain structure.



(g) Our proposed pipeline.

Position estimate validation on different devices

- Measurements across different devices against a ground truth distance of 10 cms have been shown below.



(h) Measurement across devices.

- Use the 3D probe in our lab to find the real time 3D location of the camera each time a picture is clicked by touching the same location on the mobile.
- The analysis suggests a standard deviation of 7 - 8 cms with our estimated positional data.

What next?

- The error analysis suggests refinement in camera position before triangulation.
- Also a global refine of structure and position will be necessary to further refine structure.

Thank You