3D reconstruction using an IMU enabled mobile device

Summer Undergraduate Research Award

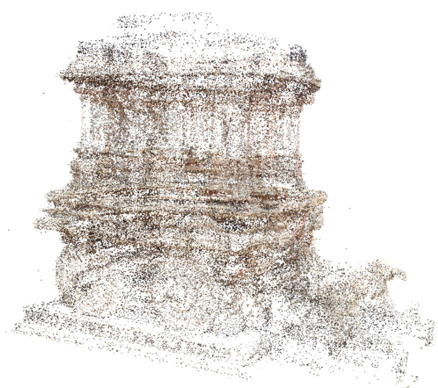
Kartikeya Gupta
2013CS10231
Computer Science
CGPA: 9.78
Mob: 9899134337
cs1130231@iitd.ac.in

Prateek Kumar Verma
2013CS10246
Computer Science
CGPA: 8.87
Mob: 9911577289
cs1130246@iitd.ac.in

Supervisor:-
Subhashis Banerjee
Professor
Department of CSE
suban@cse.iitd.ac.in
IIT Delhi

1 Introduction

3D reconstruction is the process of capturing the shape and appearance of real objects. The conventional methods for 3D reconstruction is based on *multiple images*. In this method, 3D models are constructed from a set of images. It can be considered as a reverse process of obtaining 2D images from 3D scenes.
3D reconstruction has a wide variety of applications. It allows engineers and students to work more efficiently as they can generate copies of 3D objects easily. In orthopedics and joint replacement surgery, the complicated parts to be replaced can be made with high accuracy. Also, it can be used to generate replica of artifacts and fragile objects for further studies, without harming its integrity.



(a) Sparse reconstruction

(b) Dense reconstruction

Figure 1: 3D reconstruction

Figure 1a shows the first step towards 3D construction i.e. *Sparse reconstruction*, which is done for some set of points. Figure 1b is the final reconstructed 3D model of the object. The *traditional pipeline* is shown in Figure 2.

The red-highlighted part of the pipeline is ***computationally expensive***. Thus, our project aims to reduce this computation and perform 3D reconstruction in near real-time.

The processing parts are:

- *Intrinsic and extrinsic parameters*: The camera projection matrix is a 3×4 matrix which represents the pinhole geometry of a camera for mapping 3D points in the world coordinates to 2D points on images. This matrix depends on extrinsic and intrinsic parameters. The intrinsic parameters mainly comprises of focal length, image sensor format, and principal points. The extrinsic parameters define the position of the camera center and the camera's heading in world coordinates in terms of a rigid rotation and translation.

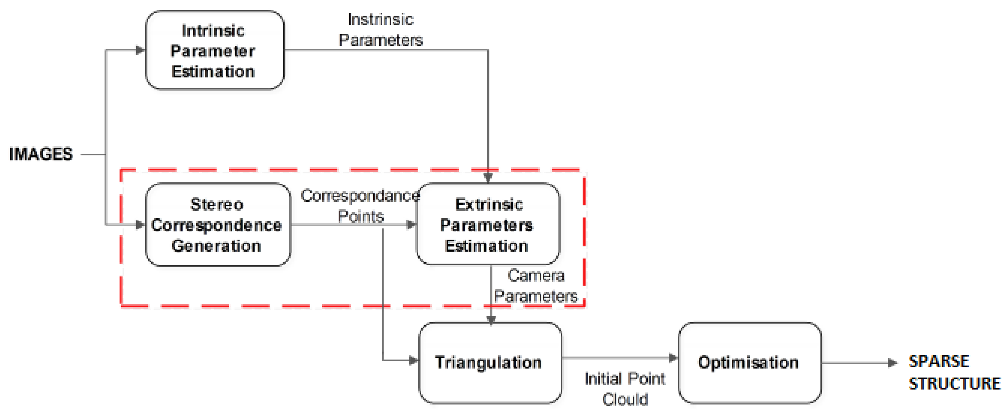


Figure 2: Traditional pipeline

- *Stereo correspondence generation*: Given two or more images of the same 3D scene, taken from different points of view, the correspondence problem refers to the task of finding a set of points in one image which can be identified as the same points in another image. To do this, points or features in one image are matched with the corresponding points or features in another image. The images can be taken from a different point of view, at different times, or with objects in the scene in general motion relative to the camera(s).

- *Triangulation*: Triangulation refers to the process of determining a point in 3D space given, its projections onto two or more images and their corresponding camera projection matrices. This point is found as the intersection of the two or more projection rays formed from the inverse projection of the 2D image points representing that 3D point in space.

- *Initial point cloud and 3D sparse reconstruction*: As the word suggests, *3D sparse construction* is done for only some set of data points in the given coordinate system called *initial point cloud.* Figure 1a illustrates a 3D sparse construction of a chariot. Figure 1b illustrates 3D dense construction of the same initial point cloud.

2 Objectives

Our main objective is to perform 3D reconstruction in near real time using a mobile device. This can be further be subdivided into following points:

1. To get accurate position and orientation estimate based on readings of IMU sensors in smart-phones.

2. To use the camera feed in smart-phones to enhance the position estimate based on visual tracking of objects.

3. To do sparse 3D reconstruction based on sensor fusion data and computer vision techniques.

4. To enhance the quality and efficiency of 3D reconstruction by adding more details and moving towards dense 3D reconstruction.

5. We will ultimately be fusing digital signal processing and computer vision based techniques that will enable us to perform near real time 3D reconstructions on mobile or hand-held devices.

3 Basic Concepts

3.1 Camera calibration

The camera parameters can further be subdivided into intrinsic and extrinsic parameters. **Camera intrinsic parameter** K is dependent on the focal length of the camera and principal point (which in most cases is the center of the image). The **camera extrinsic parameter** is composed of the rotation R and translation t between camera coordinate system and the world coordinate system. Together they form the camera projection matrix P, a 3×4 matrix which describes the mapping of a pinhole camera from 3D points in the world to 2D points in an image.

$$P = K[R|t] \tag{1}$$

3.2 Sparse 3D reconstruction

Given two different images of the same scene from different angles, the position of a 3D point can be found as the intersection of the two projection rays which is commonly referred to as **triangulation**. For this first point correspondences have to be established. Then using this point correspondences a Random Sampling Consensus (RANSAC) based voting framework is used to estimate the camera intrinsic and extrinsic parameters. Finally, a joint non-linear optimization is used to further refine the camera parameters and the 3D points in a **bundle adjustment** framework. This method is computationally very expensive and hence done only for very sparse set of points. This is known as sparse 3D reconstruction.

4 Mobile IMU sensors

IMU (Inertial Measurement Unit) sensors are on-chip devices embedded in most of the smart phones or hand-held devices today. It mainly consists of a series of motion sensors: accelerometer, gyroscope, magnetometer and gravitation sensor. The data from these sensors can be fused to obtain the orientation and the position of the device in the world coordinate system.

5 Conventional versus mobile 3D reconstruction

In the case of a smart-phone or any hand-held device having a camera and IMU sensors, we wish to use the IMU sensors to obtain extrinsic camera parameters in real time. This will help in reducing the load on conventional 3D reconstruction methods and get it in near real time.

6 Approach to the project

First, we shall be using sensor fusion to obtain accurate estimates for camera position and orientation of the mobile device. Then we will move on to 3D reconstruction, which further has two parts: i.e. sparse 3D reconstruction and then use tracking to obtain dense correspondence of points for dense 3D reconstruction.

1. Position and orientation estimation

    (a) Get accelerometer data and orientation data at real time using the IMU sensors like accelerometer, gyroscope, gravity sensor and magnetometer present on the smart phone. This data is highly noisy. Figure 3 shows the position estimate from accelerometer data across various devices.

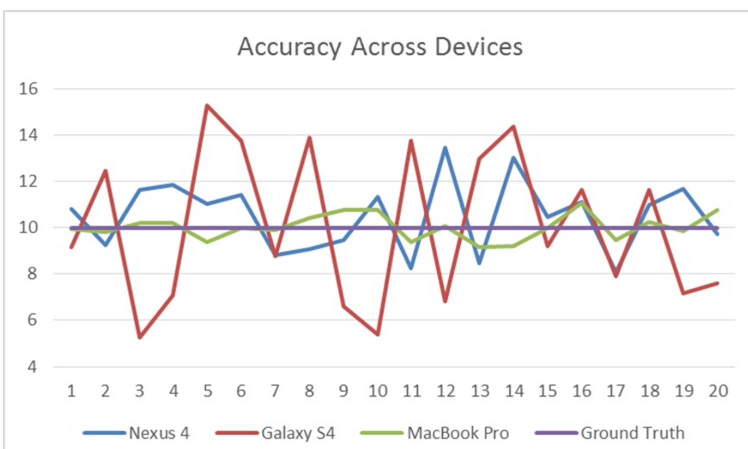


Figure 3: Accuracy of accelerometer data across different devices (scale cm)

    As evident from the graph, this data cannot be directly used for calculation of position and orientation. Figure 4 shows the integration of static accelerometer data to obtain velocity and displacement.

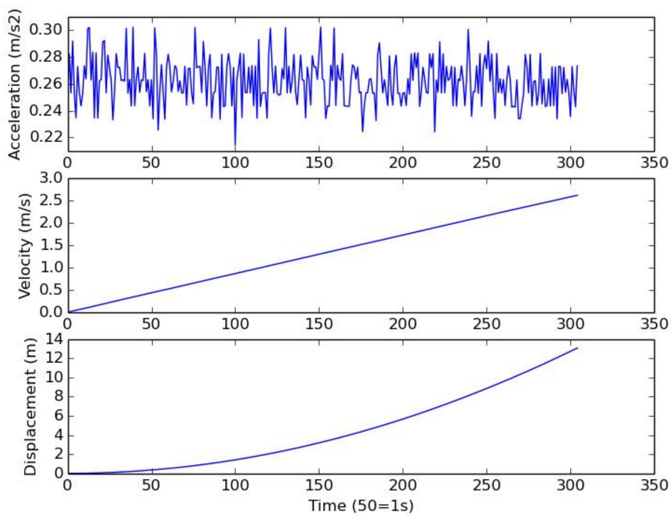


Figure 4: Obtaining velocity and displacement from static accelerometer data

The graph of both velocity and displacement shows significant deviation from the actual value which is zero. Thus, signal processing and smoothening is required to get a better estimate.

(b) Making the orientation data more accurate by infusing the higher frequency components from the gyroscope orientation after drift correction.

(c) Obtaining the displacement and orientation data from the camera feed on the device using visual tracking methods.

(d) A comparative study is to be done between the position estimates obtained by the two methods along with ground truth and fusing the results to obtain an enhanced position and orientation estimate.

2. 3D reconstruction

   (a) Obtain sparse 3D reconstruction based on camera rotation and position parameters obtained previously.

   (b) Use tracking data from different tracking methods like "Good features to track" or "KL tracker" for obtaining dense correspondence of points.

   (c) Use guided matching by indirect computation of fundamental matrix from estimated camera motion from sensors to further enrich the correspondences.

(d) Triangulate the dense correspondences and do a final global refinement.

3. Further possibilities

   - Getting a more detailed texture mapping of the object.
   - Making an object recognition software on the basis of this 3D reconstruction.
   - Improving the algorithm for a quicker and more efficient 3D reconstruction.
   - Releasing applications for Apple, Android and Windows platforms for near real time 3D reconstruction on the device itself.

7 Uses and applications

- Using the device as an accurate measuring device. This can be of particular interest to blind as they will be able to measure distances and angles accurately with great ease.

- Doing real time dense 3D reconstructions on mobile phones and other hand-held devices.

- Allowing the user to generate a 3D printable file on his mobile device. As 3D printers are becoming cheaper and more common, this feature will reduce the need of the person to use a 3D scanner to be able to generate prototypes of objects. This will allow engineers and students to work more efficiently as they can generate copies of 3D objects easily.

- This project can have applications in the field of archeology. It can be used to generate replica of artifacts and fragile objects for further studies, without harming its integrity.

- Our approach can also be applied in the field of medical sciences, especially for orthopedics and joint replacement surgery. The part to be replaced can be made with high accuracy using this project.

- The method can also be used by the astronauts up in space. With the help of our approach the parts to be changed can be easily made using a 3D printer.

- Localization at tourist sites and providing real time directions to landmark locations. This will involve the use of GPS (Global positioning system) as well to get a rough location of the user.

8 Budget, duration and facilities

8.1 Budget

Rs. 25,000 will be needed to purchase an android smart phone having high quality sensors and a high resolution camera.

8.2 Duration

We will try to complete this project by the end of the summer break i.e. the end of July, 2015.

8.3 Facilities

- Access to the vision lab.

9 Report

“To get accurate position and orientation estimate based on readings of IMU sensors in smart-phones.”

- *Initialisation*:
  Started by making an Android application on Android Studio. The following sensor were deployed:
  - Accelerometer
  - Gravity
  - Gyroscope
  - Magnetic Field

  The application consisted of five buttons on the screen with the view from camera on the background
  - Start: To begin the process of data collection from sensors and video capture from the camera
  - Quit: To quit the application
  - Image Click: To take various images of the concerned object after pressing Start
  - Clear: To clear Cache
  - About: About the application

  On pressing the start button, the following data is received from the sensors and stored for further processing. Writing of data is done when the gravity sensor is active since it is the most frequent sensor.
  - Acceleration
  - Magnetic field
  - Gravity
  - Gyroscope
  - Rotation Matrix from inbuilt Android function using gravity and magnetic field (Calculated)
  - Rotation Matrix from gyroscope (Calculated)
  - ImageID corresponding to this data
- *Data Processing*:
  Calculation of Rotation is started in parallel with collection of data. After all the data is collected, the following are done (in order):
  - Acceleration Frame is changed from Ground Frame to Camera Frame

- Gravity is removed from the acceleration data
- *Static Bias Removal*:
  Assuming the device was at rest before the first image was clicked, static bias in the acceleration is removed. First, the mean of acceleration values before the first image was clicked is calculated and then subtracted from all the acceleration values.
- *Smooth Acceleration*:
  Two types of smoothening techniques were tried i.e Gaussian Filter and 'LOESS' local regression. 'LOESS' local regression produced better results and hence was incorporated.
- *Motion Zones Identification*:
  A data point was said to be a motion point if in its neighborhood lied a point having significant acceleration compared to the standard deviation of the acceleration points before the first image was clicked. All the contagious data points were clubbed together and identified as zones. Points outside these zones having non-zero acceleration were declared as rest points and hence noise was removed.
- *Velocity and Distance Calculation*:
  Velocity and distance were calculated taking account only the accelerations in the identified motion zones.

write about the results obtained

"To use the camera feed in smart-phones to enhance the position estimate based on visual tracking of objects."

Exploited the video feed from the camera to generate correspondences through visual tracking of points. It was done in the following way.

- Feature points were identified in the first frame
- These points were tracked in the following video frame using KL Tracker.
- As a result, from the first frame a chain of points was established till the next image was captured and hence correspondences were identified
- Also, to improve tracking, more feature points were added after every 10 frames so that enough points are available.

To identify the FPS required for this method to work, an offline program was run on various FPS. It was seen that FPS above 6-7 was able to produce accurate correspondences.
This process turned out to be extremely fast and efficient.

“To do sparse 3D reconstruction based on sensor fusion data and computer vision techniques.”

After establishing correspondences between different image clicks, Triangulation was done using the previously calculated Rotation and Translation Matrices in the following way

- Converting the Rotation Matrix (Sensor Frame) to Rotation Matrix (Camera Frame)
- Using Triangulation Equation