CS676: 3D Reconstruction from Several Images(Structure from Motion)

Report

Anshu Avinash 10327122
Pranjal Singh 10327511
Sunil Kumar 10327742
Indian Institute of Technology,Kanpur
Advisor: Prof. Vinay P. Namboodiri

April 17, 2014

Abstract

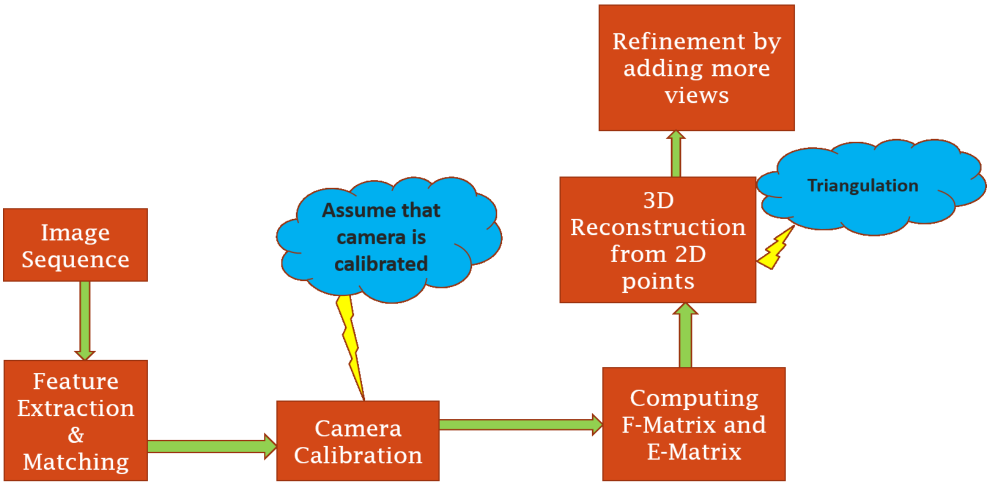
In this work we show how 3D reconstruction, basically structure from motion, from point correspondences of multiple images can be achieved. The original contributions with respect to related works in the field are mainly a direct global method for relative 3D reconstruction and a geometric method to select a correct set of reference points among all image points. Experimental results from real image sequences are presented, and algorithm has been discussed in detail.

1 Introduction

3D reconstruction from multiple images is the creation of three-dimensional models from a set of images. It is the reverse process of obtaining 2D images from 3D scenes.
Structure from motion (SfM) is a range imaging technique; it refers to the process of estimating three-dimensional structures from two-dimensional image sequences which may be coupled with local motion signals.
We humans perceive information about 3D structure in the environment by moving through it. When the observer moves and the objects around him move, information is obtained from images sensed over time.
We expect that our future will be full of robots, so these robots must have ability to perceive 3D structure like us. Hence, this work is motivated by this observation. The problem of SFM can be mathematically described as follows:
Given many points in correspondence across several images, $\{(u_{ij}, v_{ij})\}$, simultaneously compute the 3D location x_i and camera (or motion) parameters (K, R_j, t_j)
$\bar{u_{ij}} = f(K, R_j, t_j, x_i)$
$\bar{v_{ij}} = g(K, R_j, t_j, x_i)$

2 Algorithm

The basic framework and algorithm of this work can be understood by the following figure:


Refinement by adding more views
Triangulation
3D Reconstruction from 2D points
Assume that camera is calibrated
Image Sequence
Feature Extraction & Matching
Camera Calibration
Computing F-Matrix and E-Matrix

Now we will discuss each step in some detail.

2.1 Feature Matching

Feature matching is an integral part of SFM because we need to know the exact location of each point in every image sequence that we have so that we can project the same in 3D.
The geometrical theory of structure from motion assumes that one is able to solve the correspondence problem, which is to identify points in two or more views that are the projections of the same point in space.
One solution is to identify corresponding points interactively in each view. An important advantage is that surfaces can be defined simultaneously with correspondences. Feature matching works by detecting interest points in the images.
We have used **SIFT**(Scale Invariant Feature Transform) for feature extraction and k-NN for matching.
Optical flow is another alternative to get better matching. It is the process of matching selected points from one image to another, assuming both images are part of a sequence and relatively close to one another. Its advantageis that it is usually faster and can accommodate matching many more points, making the reconstruction denser. We have also experimented with Optical flow technique.

2.2 Camera Calibration

Camera intrinsic and extrinsic parameters can be determined for a particular camera and lens combination by photographing a controlled scene. We have assumed that our camera is calibrated, hence we know its intrinsic parameters such as focal length, optical center, aspect ration, etc.

2.3 Fundamental Matrix-F

It records motion between cameras and we have used RANSAC for calculation of F. The mathematical condition for determining F is $x'Fx = 0$.
It is used to calculate Essential Matrix which is described below.

2.4 Essential Matrix-E

$E = K^T \times F \times K$
It is 3×3 sized matrix which can be used to recover projection matrix for each camera. It imposes a constraint between a point in one image and a point in the other image with $x' \times E \times x = 0$
where x is a point in image one and x' is the corresponding point in image two.

2.5 Projection Matrix-P

It is 3×4 sized matrix which contains rotation and translation parameters of the camera. It is derived by doing *Singular Value Decomposition* of the *Essential Matrix.*

2.6 Triangulation

Given projection matrices, 3D points can be computed from their measured image positions in two or more views. This is triangulation. Ideally, 3D points should lie at the point of intersection of the back-projected rays. However, because of measurement noise, back-projected rays will not generally intersect. Thus 3D points must be chosen in such a way as to minimize an appropriate error metric.
We have used *Linear Method* for triangulation. Two key equations arising from the 2D point matching and P matrices are:
$x = PX$ and $x' = P'X$
where x and x' are matching 2D points and X is a real world 3D point.
If we rewrite the equations, we can formulate a system of linear equations that can be solved for the value of X, which is what we desire to find.

2.7 Reconstruction from Many Views

Now that we know how to recover the motion and scene geometry from two cameras, it would seem trivial to get the parameters of additional cameras and more scene points simply by applying the same process. This matter is in fact not so simple as we can only get a reconstruction that is up-to-scale, and each pair of pictures gives us a different scale.
There are a number of ways to correctly reconstruct the 3D scene data from multiple views. One way is of resection or camera pose estimation, also known as Perspective N-Point(PNP), where we try to solve for the position of a new camera using the scene points we have already found.
We have used $SolvePnPRansacfunction$ available in *OpenCV.*

2.8 Refinement of Reconstruction

One of the most important parts of an SfM method is refining and optimizing the reconstructed scene, also known as the process of *Bundle Adjustment.* This is an optimizing step where all the data we gathered is fitted to a monolithic model. Both the position of the 3D points and the positions of cameras are optimized, so re-projection errors are minimized (that is, approximated 3D points are projected on the image close to the position of originating 2D points).
We have used SSBA(Simple Sparse Bundle Adjustment) Library.

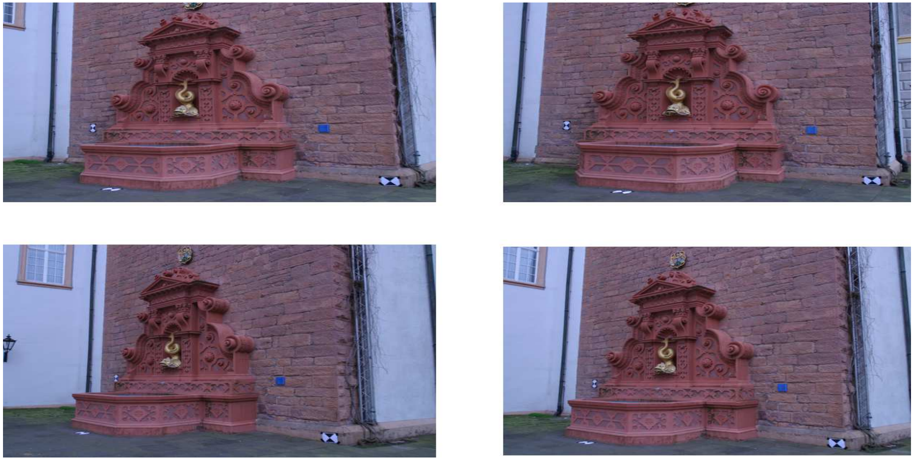
2.9 Visualization

We have used Point Cloud Library (PCL) for visualization of 3D points.

3 Dataset Used

We have used two datasets of CVlab of EPFL which are:

- Fountain P-11 Sequence *http://cvlabwww.epfl.ch/s̃trecha/multiview/denseMVS.html*

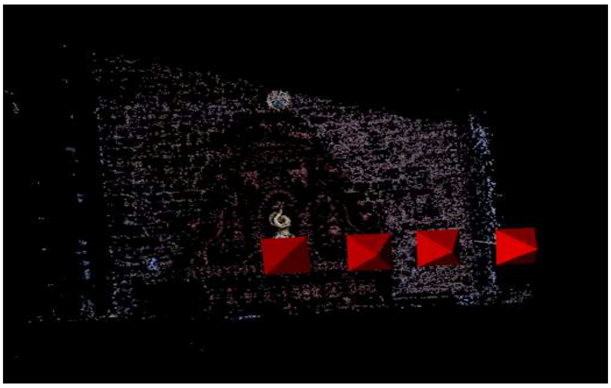


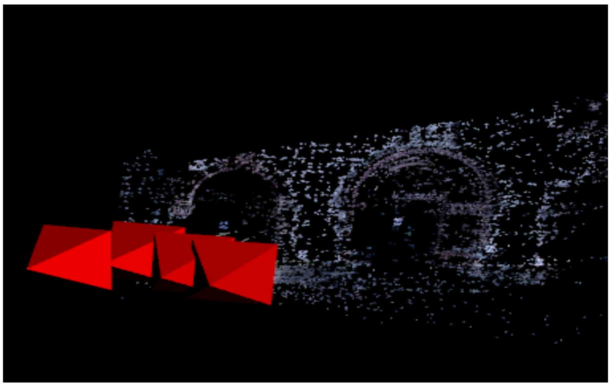
- Herz-Jesu-P8 *http://cvlabwww.epfl.ch/s̃trecha/multiview/herzjesu_dense.html*

4 Result

The following images shows our output on the two datasets of CVlab of EPFL:

- Fountain P-11 Sequence



- Herz-Jesu-P8 

5 Conclusions and Future Work

In this paper we have presented a novel tool, which enables us to solve the structure from motion problem without a priori correspondence information. The final algorithm is simple and easy to implement and fast.
The quality of results can be easily improved if we use more images from different views though it slows down the rate of computing. We can also improve if we have better feature correspondences in all images, implying that we can try with other feature extraction techniques.
The triangulation method that we have used is a linear method, we can experiment with other triangulation methods and then compare our results. There are also many state-of-the-art techniques available such as Bundler, VisualSFM,LibV. We can try to incorporate their techniques into our system.

References

[1] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[2] Richard Hartley and Peter Sturm. Triangulation, 1996.

[3] M. Johnson-Roberson, O. Pizarro, S. Williams, and I. Mahon. Large scale 3d reconstruction and visualization of stereo surveys. In *Marine Geological and Biological Habitat Mapping (GeoHab) Video Workshop*, 2009.