3D Modeling using Structure from Motion and Shape from Shading

ENEE 731 Final Project

Chethan Mysore Parameshwara

1. Structure from Motion

In this project, two SFM algorithms are tested on a given dataset and results are discussed. The two algorithms are -

1. Orthographic Factorization
2. Projective Factorization with Bundle adjustment

1.1. Orthographic Factorization

The factorization method [1] extracts P feature points from image stream and tracks over F frames. The trajectories of image coordinates $\{(u_{fp}, v_{fp}) | f = 1, \ldots, F, p = 1, \ldots, P\}$. The feature coordinates u_{fp} and v_{fp} are arranged in the form of 2F × P matrix, W.
The registered measurement matrix $\widehat{W}$ can be expressed in a matrix form $\widehat{W}$ = RS, where S is a shape and R is a rotation.

1. Once we have $\widehat{W}$, compute the singular-value decomposition $\widehat{W} = O_1 \Sigma O_2$
2. Define $\widehat{R} = O_1(\Sigma)^{1/2}$ and $\widehat{S} = (\Sigma^{'})^{1/2} O_2^{'}$, where the primes refer to the block partitioning defined in equation (1)

$$O_1 = [\underbrace{O_1'}_{3} \mid \underbrace{O_1''}_{P-3}] \}\ 2F$$

$$\Sigma = \underbrace{\left[\begin{array}{c} \Sigma' \\ 0 \end{array}\right.}_{3} \underbrace{\left.\begin{array}{|c} 0 \\ \Sigma'' \end{array}\right]}_{P-3} \begin{array}{l} \}\ 3 \\ \}\ P-3 \end{array}$$

$$O_2 = \underbrace{\left[\begin{array}{c} O_2' \\ \hline O_2'' \end{array}\right]}_{P} \begin{array}{l} \}\ 3 \\ \}\ P-3 \end{array}$$

$$\mathbf{R} = \hat{\mathbf{R}}\mathbf{Q}$$
$$\mathbf{S} = \mathbf{Q}^{-1}\hat{\mathbf{S}}$$

$$\hat{\mathbf{i}}_f^T QQ^T \hat{\mathbf{i}}_f = 1$$
$$\hat{\mathbf{j}}_f^T QQ^T \hat{\mathbf{j}}_f = 1$$
$$\hat{\mathbf{i}}_f^T QQ^T \hat{\mathbf{j}}_f = 0$$

Equation 1. Equation 2. Equation 3.

3. Compute the matrix Q in equation (2) by imposing the metric constraints (equation (3)).
4. Compute the rotation matrix R and the shape matrix S as $R = \widehat{R}Q$ and $S = Q^{-1}\widehat{S}$.
5. If desired, align the first camera reference system with the world reference system by forming the products RR_0 and $R_0^T S$, where the orthonormal matrix $R_0 = [i_1 j_1 k_1]$ rotates the first camera reference system into the identity matrix.

1.2. Projective Factorization with Bundle adjustment

The process of extracting structure information from motion is usually composed of an initial coarse reconstruction using Projective Factorization which is later refined with Bundle Adjustment.

The 3D reconstruction problem aims at recovering a model of all camera poses and all the m 3D points $X_{i=1...m}$ of a scene from multiple views. Camera projections are written as 3 x 4 matrices $P_{i=1...n}$ and can be decomposed in some metric coordinate frame, as $P_i = K_i(R_i|t_i)$, K_i encodes the intrinsic parameters and (R_i, t_i) represents the orientation and position of the camera in a world coordinate frame.

The measurement matrix W is defined as $W = \frac{1}{\lambda}PX$, where X is set of m stationary 3D points $X = [x_i, y_i, z_i, 1]^T$ and λ is scale factor, commonly called as project depth.

1. Start with an initial estimate of the depths λ using Eight-point algorithm.
2. From the measurement matrix W, find the nearest rank 4 approximation using the SVD and decompose to find the camera matrices and 3D points
3. Reproject the points into each image to obtain new estimates of the depths and repeat from step 2
4. Bundle-adjust the cameras and 3D structure to minimize projection errors

1.3. Observations and Limitations

The hotel[7], castle[4] and medusa[5] datasets were used to test both the algorithms [9][10]. As the hotel dataset was synthetic it worked really well with Orthographic Factorization method as shown in the Figure 1.1. But, the Orthographic Factorization method failed to deliver the results for castle and medusa as both datasets were not synthetic. This was because noisy datasets, castle and medusa, fails to work with the singular value decomposition technique to factor the measurement matrix. However, the method can handle and obtain a full solution from a partially filled-in measurement matrix, which occurs when features appear and disappear in the image sequence due to occlusions or tracking failures.

Further, the castle and medusa datasets were subjected to projective factorization with Bundle adjustment. The results for this method (Figure 1.2. and Figure 1.3.) was better as it uses non-linear optimization method to correct projections. The projective factorization with bundle adjustment has advantages over the orthographic factorization as it can handle large number of views and missing data. However, it requires good initial estimate of depth and solving the large minimization problem is computationally very expensive.

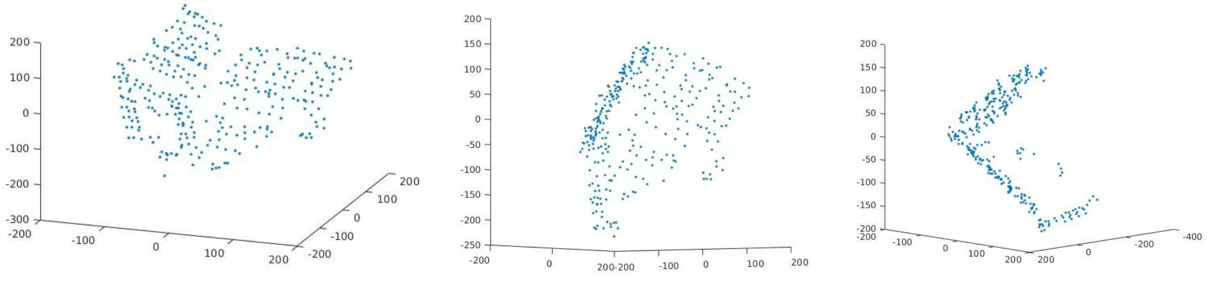

Figure 1.1. Orthographic Factorization method on Hotel data set

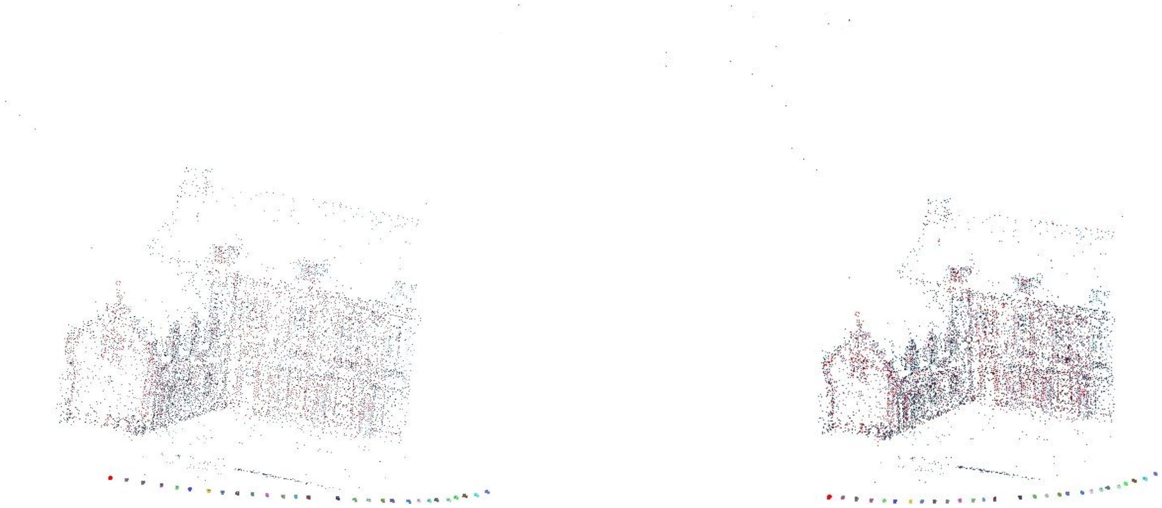

Figure 1.2. Visual SFM on Castle dataset with Bundle Adjustment

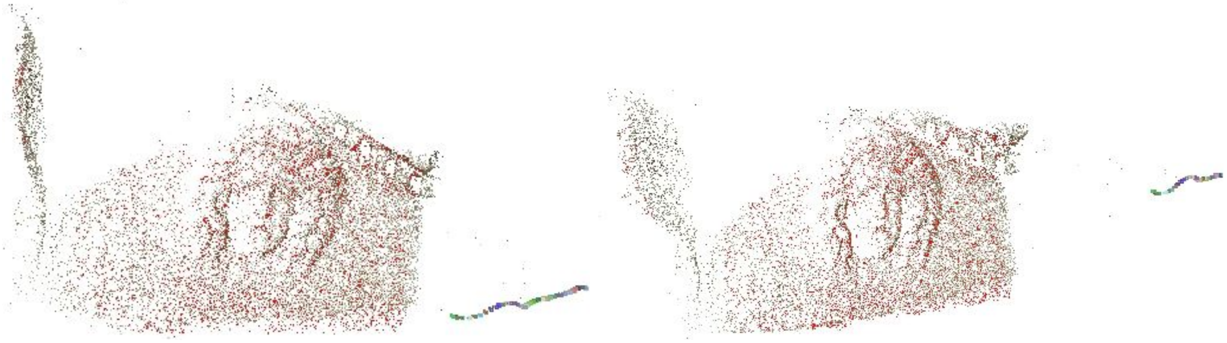

Figure 1.3. Visual SFM on Medusa dataset with Bundle Adjustment

2. Shape from shading

2.1. Introduction

The authors of [2] have formulated the problem of recovering intrinsic scene properties, “shape, illumination, and reflectance”, from shading as an unconstrained optimization problem.

$$min_{Z,L} \quad g(I - S(Z, L)) + f(Z) + h(L)$$

Where R is a log-reflectance image, Z is a depth-map, and L is a spherical-harmonic model of illumination and g(R) , f(Z), and h(L) are cost functions for reflectance, shape, and illumination respectively, which are referred as priors throughout the paper.

2.1.1. Prior on reflectance

1) In assumption of piecewise constancy, the authors have modelled this prior by minimizing the local variation of log-reflectance in a heavy-tailed fashion.
2) The assumption of parsimony of reflectance assumes the palette of colors with which an entire image is painted tends to be small. This is modelled by minimizing the global entropy of log-reflectance
3) An absolute prior on reflectance assumes that some colors are more likely than others. This is modelled by maximizing likelihood under some density model

2.1.2. Prior on shape

1) The assumption of smoothness suggests that shapes tend to bend rarely. The authors have modeled this prior by minimizing the variation of mean curvature.
2) An assumption of isotropy of the orientation of surface normals is modelled by minimizing the slant.

2.1.3. Prior on illuminance

The authors fit a multivariate Gaussian to the spherical-harmonic illuminations through a training set.

2.1.4. Optimization

The authors present the an effective multi-scale optimization technique based on L-BFGS. The optimization involves minimization of a(X), where a( ) is some loss function and X is some signal. The X is rewritten as X = g^TY and then solved using L-BFGS. The naive single-scale optimization for the current problem works poorly.

2.2. Observations and Limitations

The castle[4] and medusa[5] datasets were used to test the algorithm[11]. The algorithm took approximately 30 minutes to converge from initial results to the final results as shown in figures below. Even though the model assumes lot of priors, the results show that the model was able to produce extremely compelling shading and reflectance images, and qualitatively correct illumination. The direction of light was very accurately predicted which is evident in Figure 2.3.

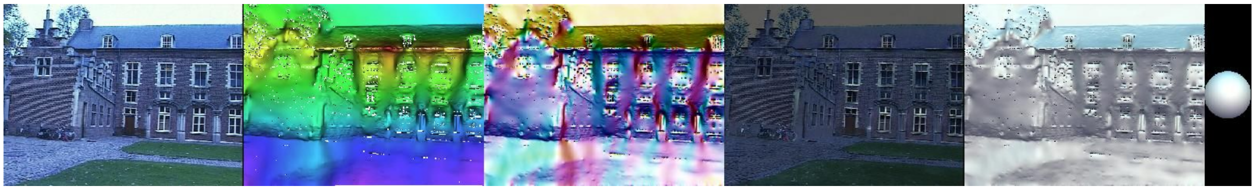

Figure 2.1. Final result of Castle Image (Original, shape, normals,reflectance, shading,light)

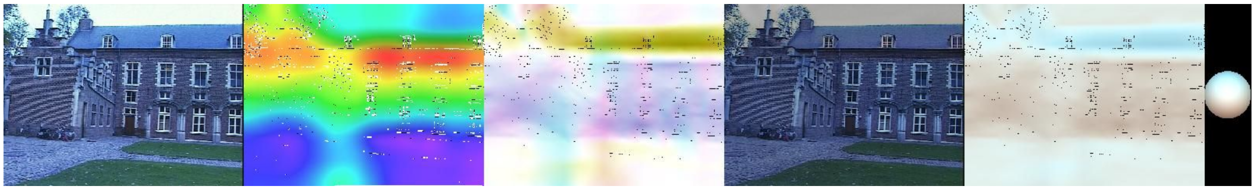

Figure 2.2. Initial result of Castle Image (Original, shape, normals,reflectance, shading,light)

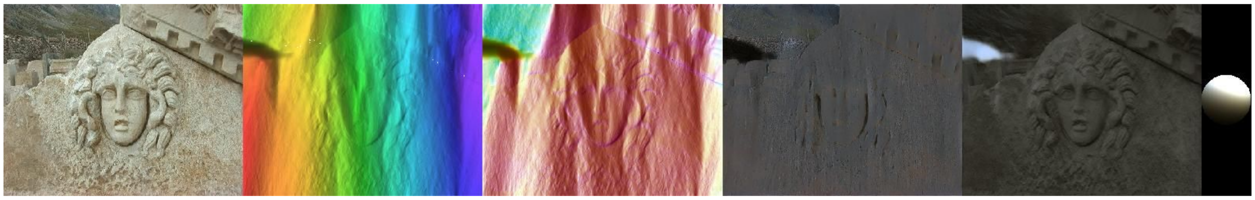

Figure 2.3. Final result of Medusa Image (Original, shape, normals,reflectance, shading,light)

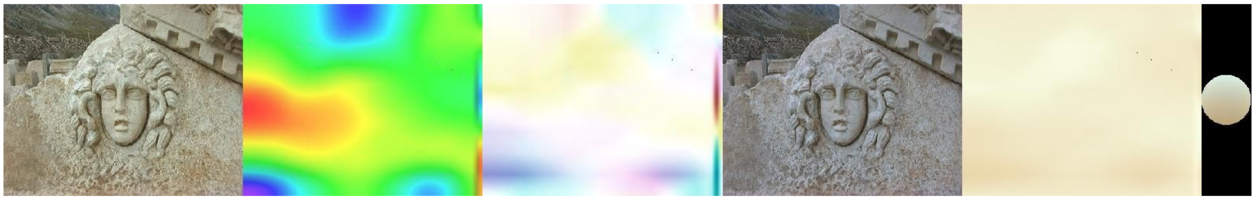

Figure 2.4. Initial result of Medusa Image (Original, shape, normals,reflectance, shading,light)

As shading is an inherently poor cue for low-frequency shape estimation in the case of Castle image, the proposed model often mistakes in coarse shape estimation. Even though authors assumed the lambertian materials and image consisting of single masked object, the model predicted the intrinsic properties close to ground truth. They also assume illumination is global, and ignore illumination issues such as cast shadows, mutual illumination, or other sources of spatially-varying illumination. The priors on shape and reflectance are independent of the category of object present in the scene. This is not really a limitation as authors see this as a strength of their model to generalize across object categories.

3. Reference

[1] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography-- a factorization method," International Journal of Computer Vision, 9(2):137--154, 1992.
[2] J.T. Barron and J.Malik, “Shape, illumination, and reflectance from shading”, IEEE Trans. on Patt. Anal. And Mach. Intel.
[3] Projective Factorization - http://www.umiacs.umd.edu/~ramani/cmsc828d/lecture28.pdf
[4] Leuven castle image sequence http://www.cs.unc.edu/~marc/data/castlejpg.zip
[5] Sagalassos medusa head sequence http://www.cs.unc.edu/~marc/medusa.dv
[6] SIRFS Presentation https://people.eecs.berkeley.edu/~barron/BarronMalikTPAMI2015_presentation.pdf
[7] CMU VASC Image Database: hotel http://vasc.ri.cmu.edu//idb/html/motion/hotel/index.html
[8] Bill Triggs, Factorization methods for projective structure and motion. In Proceeding of 1996 Computer Society Conference on Computer Vision and Pattern Recognition, pages 845–51, San Francisco, CA, USA, 1996. IEEE Comput. Soc. Press. http://citeseer.ist.psu.edu/article/triggs96factorization.html
[9] Orthographic Factorization implementation https://sourceforge.net/p/cvprtoolbox/code/HEAD/tree/
[10] Visual SFM implementation - http://ccwu.me/vsfm/
[11] Shape from shading implementation https://people.eecs.berkeley.edu/~barron/SIRFS_release1.5.zip