# 3D modeling using Structure from Motion and Shape from Shading

Rohan Chandra

*(CMSC 828G: Project 2))*

**Abstract**

The structure from motion - recovering scene geometry and camera motion from a sequence of images - is an important task and has wide applicability in many tasks, such as navigation and robot manipulation. Tomasi and Kanade first developed a factorization method to recover shape and motion. Factorisation allows for de-noising the data and for extracting leading components. We further extract shape, illuminance and reflectance information from shading by implementing the work done by Barron and Malik. In this project, we implemented the factorization method and shape from shading algorithm, and comparisons are shown.

*Keywords:* Factorisation, SfM, SfS

## 1. Introduction

This project was implemented as part of coursework for CMSC 828G under the guidance of Prof. Rama Chellappa. The main aim of this work is to explore classical and recent techniques for structure from motion (SfM) and shape from shading (SfS). Humans perceive a lot of information about the three-dimensional structure in their environment by moving through it. When the observer moves and the objects around the observer move, information is obtained from images sensed over time.

### Structure From Motion

Finding structure from motion presents a similar problem to finding structure from stereo vision. In both instances, the correspondence between images and the reconstruction of 3D ob-

ject needs to be found. To find correspondence between images, features such as corner points (edges with gradients in multiple directions) are tracked from one image to the next. One of the most widely used feature detectors is the scale-invariant feature transform (SIFT). It uses the maxima from a difference-of-Gaussians (DOG) pyramid as features. The first step in SIFT is finding a dominant gradient direction. To make it rotation-invariant, the descriptor is rotated to fit this orientation.[2] Another common feature detector is the SURF (Speeded Up Robust Features).[3] In SURF, the DOG is replaced with a Hessian matrix-based blob detector. Also, instead of evaluating the gradient histograms, SURF computes for the sums of gradient components and the sums of their absolute values.[4] The features detected from all the images will then be matched. One of the matching algorithms that track features from one image to another is the LukasKanade tracker.

Sometimes some of the matched features are incorrectly matched. This is why the matches should also be filtered. RANSAC (Random Sample Consensus) is the algorithm that is usually used to remove the outlier correspondences. In the paper of Fischler and Bolles, RANSAC is used to solve the Location Determination Problem (LDP), where the objective is to determine the points in space that project onto an image into a set of landmarks with known locations.

The feature trajectories over time are then used to reconstruct their 3D positions and the camera's motion. An alternative is given by so-called direct approaches, where geometric information (3D structure and camera motion) is directly es-

timated from the images, without intermediate abstraction to features or corners.

There are several approaches to structure from motion. In incremental SFM, camera poses are solved for and added one by one to the collection. In global SFM, the poses of all cameras are solved for at the same time. A somewhat intermediate approach is out-of-core SFM, where several partial reconstructions are computed that are then integrated into a global solution. In this work, we use the factorisation technique by Tomasi and Kanade where they elegantly decompose the observation matrix by SVD and also perform de-noising of the data.

### Shape From Shading

Photometric stereo is a technique in computer vision for estimating the surface normals of objects by observing that object under different lighting conditions. It is based on the fact that the amount of light reflected by a surface is dependent on the orientation of the surface in relation to the light source and the observer. By measuring the amount of light reflected into a camera, the space of possible surface orientations is limited. Given enough light sources from different angles, the surface orientation may be constrained to a single orientation or even overconstrained.

The technique was originally introduced by Woodham in 1980.[2] The special case where the data is a single image is known as shape from shading, and was analyzed by B. K. P. Horn in 1989.[3] Photometric stereo has since been generalized to many other situations, including extended light sources and non-Lambertian surface finishes.[4] Current research aims to make

the method work in the presence of projected shadows, highlights, and non-uniform lighting. In this work, we attempt to implement the algorithm by Barron and Malik where they extract shape, illuminance and reflectance information from a single image.

**2. Experiments**

**3. Results**

**4. References**