

# CSE 190A Project Proposal: 3D Photography

Kristen Kho

Department of Computer Science and Engineering  
University of California, San Diego  
kmkho@ucsd.edu

## Abstract

*This paper presents a research proposal for CSE 190A: Projects in Vision and Learning, Winter 2007, on the subject of 3D photography. It addresses the objectives, datasets, milestones, and student qualifications for the project.*

## 1 Introduction

3D photography is the use of technology to capture the shape and appearance of real world objects and reconstruct them in the digital world. The digitization of an object can be viewed from novel viewpoints that may be difficult or impossible to achieve in the real world, and does not require one to physically traverse great distances. Due to the detail and realism of the models generated by this process, 3D photography has many practical applications and has been employed across many different fields.

There are two main approaches to 3D photography: passive and active. Active techniques require tightly controlled conditions when capturing a scene, often utilizing projected light and precisely calibrated cameras or scanners. The results are extremely accurate and offer an unparalleled amount of detail. However, this approach is usually not practical for distant or fast moving objects. In addition, the process is intrusive and requires expensive equipment.

The other approach, and the one to be used in this project, is passive 3D photography. Passive techniques use existing light in the

scene and can utilize an unrigged camera. The resulting models do not have the same level of detail and accuracy as those produced with active techniques, but they are a practical alternative. There are several passive approaches: stereo, structure from motion, shape from shading, and photometric stereo [3]. This project will take the stereo approach.

The stereo approach uses images taken from two different viewpoints to extract depth information from the scene. Once the correspondence between points from the two images is found, the amount of displacement can be used to compute the distance from the camera. Accuracy of stereo techniques often depends critically on camera calibration: knowledge of the camera's position, orientation, and internal parameters. However, as demonstrated by Pollefeys et al [2], it is possible to reconstruct a complex scene without this information.

## 2 Project Objectives

The main objective of this project is to use images to reconstruct digital models of sculptures around campus using 3D photography techniques. However, the same system could also be used by students to create 3D models for use in class projects like CSE 125.<sup>a</sup> Such an application would allow them to quickly and easily set up convincing virtual environments and be able to concentrate on the more difficult parts of their games. Therefore, this project will have an emphasis on model acquisition for gaming.

---

<sup>a</sup> CSE 125 class webpage: <http://pisa.ucsd.edu/cse125/>

A general pipeline for reconstructing a 3D model from a set of images consists of the following steps: image acquisition, feature selection, feature correspondence, reconstruction, and visualization.

## 2.1 Image acquisition

The method of acquisition not only determines the usability of the system but defines the subsequent steps that must be taken in order to derive information from the images. Since this project will take a passive approach, no setup is required when acquiring images. Photos will be taken by freely moving the camera around the object, making image acquisition relatively painless. The tradeoff is that without camera calibration information, the reconstruction process becomes more difficult.

## 2.2 Feature selection

It is not practical to compare every pixel of every image; therefore it is necessary to select features or interest points that are distinct and easy to match across images. The Harris edge and corner detector [7] will be used to automatically detect these points.

## 2.3 Feature correspondence

Once the interest points have been selected, a correspondence between them must be established across the different images. RANSAC (RANdom SAMpling Consensus), as proposed by Fischler and Bolles [8], is one of the more common and robust approaches and will be the one implemented by this project.

## 2.4 Reconstruction

The next task is to establish the 3D position of the interest points. According to Ma, Soatto, Košecák, and Sastry [1], a projective

transformation is the best one can do in the absence of calibration information. The process involves using two views to initialize the projection and then adding subsequent views one at a time to achieve a multiple-view reconstruction.

## 2.5 Visualization

Once the correspondence of the interest points has been established, the 3-D position for many if not most pixels in each image can be recovered. The resulting point cloud can be used to generate a dense, highly irregular polygonal mesh model of the scene.

There are several existing algorithms for creating surfaces from point clouds, including alpha shapes [4], Delaunay triangulations [5], and the crust algorithm [6]. Also, existing software can be used to complete this task and may be considered if necessary to meet the time constraints for this project.

The same images that were used to extract the geometry can be used as texture-maps to enhance the realism of the model. Since the pixel locations of the textures correspond to the interest point locations, this task should not be too hard to complete. The images can also be combined to create one highly detailed texture map.

Since the models must also be able to be used in an actual game environment, the final step will be saving the model information to a usable file format. The Microsoft DirectX Framework requires the X file format, but the Wavefront OBJ format may be more universally recognized across applications. Since many programs/plugin already exist to convert between the myriad 3D file formats available, OBJ will be the target file format.

## 3 Software and Datasets

Matlab will be used to test theorems and algorithms, but the actual implementation will be an OpenCV-based project in Visual C++,

using Visual Studio .Net 2003 as the development environment.

The dataset will consist of images taken with an off-the-shelf consumer digital camera. The campus sculptures that have been selected for the dataset include: Tim Hawkinson's Bear, Niki de Saint Phalle's Sun God, and Elizabeth Murray's Red Shoe.<sup>b</sup> These sculptures will be photographed from many different angles to allow for the most accurate and complete digital representation possible, but the use of a digital camera allows for this task to be completed relatively quickly.

Also, simple training sets will be used to help with debugging the various stages of the project. One training set will consist of screenshots of a synthetic scene (i.e. rendered with a modeling program), which allows total control over the angle, lighting, texture, etc. Another training set will consist of photos of a small, simple object, such as a soda can.

## **4 Milestones**

### **4.1 Image acquisition completed**

Deadline: Jan. 8

I plan to have all my datasets completed before the first meeting of the class. Also, I will have my development environment set up and ready to work with OpenCV.

### **4.2 Feature selection implemented**

Deadline: Jan. 15

### **4.3 Feature correspondence implemented**

Deadline: Jan. 29

### **4.4 Reconstruction implemented**

Deadline: Feb. 12

### **4.5 Visualization implemented**

Deadline: Feb. 26

### **4.6 Rough draft of final report completed**

Deadline: Mar. 5

### **4.7 Final project presentation**

Deadline: Mar. 14/16

A short 3D computer game may be created (if time allows) that incorporates the models constructed during this project. It would be demo-ed as part of the final project presentation.

### **4.8 Final report completed**

Deadline: Mar. 19

## **5 Qualifications**

I am currently taking CSE 166: Image Processing (Fall 2006), my first computer vision course. Other relevant courses I have taken include CSE 167: Computer Graphics (Fall 2004) and CSE 125: Software System Design and Implementation (Spring 2006). In CSE 167 I learned basic computer graphics concepts by completing several projects in OpenGL. In CSE 125, I was responsible for my team's graphics engine as well as most of the 3D modeling and animation. We used the DirectX API to implement our project.

---

<sup>b</sup> Stuart Collection webpage: <http://stuartcollection.ucsd.edu>

## References

- [1] Ma, Soatto, Košecká, and Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer-Verlag New York, Inc., 2004, pp. 375-411.
- [2] Pollefeys, Koch, Vergauwen, Deknuydt, and Gool. “Three-dimensional scene reconstruction from images.” *SPIE Electronic Imaging*, 2000.
- [3] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. “Modeling and Rendering Architecture from Photographs.” *SIGGRAPH '96*, August 1996, pp. 11-20.
- [4] H. Edelsbrunner and E. P. Mucke. “Three-dimensional alpha shapes.” *ACM Trans. Graphics* 13, 1994, 43-72.
- [5] J.-D. Boissonnat. “Geometric structures for three-dimensional shape representation.” *ACM Trans. Graphics*, 1984, 3(4):266-286.
- [6] Amenta, N., Bern, M., and Kamvysselis, M., “A new voronoi-based surface reconstruction algorithm,” *SIGGRAPH '98*. Orlando, FL, USA, 19-24 July 1992. pp. 415-421.
- [7] C. Harris and M. Stephens. “A combined corner and edge detector.” Fourth Alvey Vision Conference, 1988, pp.147-151.
- [8] M. Fischler and R. Bolles. “RANDOM SAMPLING CONSENSUS: a paradigm for model fitting with application to image analysis and automated cartography.” *Commun. Assoc. Comp. Mach.*, 24, pp.381-95, 1981.