3D Scene Acquisition and Modeling using an Uncalibrated View Network

Daniel G. Aliaga
Ji Zhang
Mimi Boutin

Department of Computer Science
Purdue University

Why the heck is there a living room in the graphics lab?

Daniel G. Aliaga
Ji Zhang
Mimi Boutin

Department of Computer Science
Purdue University

Motivation

- Capturing and modeling 3D scenes is an important goal for several applications in computer graphics, computer vision, and geometric modeling
  - e.g., telepresence, gaming and simulations, and several forms of virtual reality

Approaches

- Single viewpoint acquisition
  - Time-of-flight (TOF) systems
  - Advantages
    - Simplicity: external (relative) information to other devices/cameras is not needed nor exists
  - Disadvantages
    - Only sees surfaces visible from one viewpoint

Approaches

- (Traditional) Multi-viewpoint acquisition
  - TOF or triangulation systems
  - Advantages
    - Can acquire surfaces visible from more than one viewpoint
  - Disadvantages
    - For TOF and triangulation
      - All captures must be relatively calibrated
    - For triangulation
      - Low error requires large baseline
      - Large baseline reduces amount of mutually visible surfaces (to less than that visible from a single viewpoint)
      - Reduction of mutually visible surfaces necessitates more captures (to “fill in the gaps”)

Challenge

- We would like the simplicity of single viewpoint acquisition but also the additionally visible surfaces of multi-viewpoint acquisition
- So what can we do?

Observation

- The structure of the 3D scene is inherent to it and does not depend on from where the triangulation or TOF system acquired information
- Thus, let's make the capture process independent of the relative locations of the acquisition device

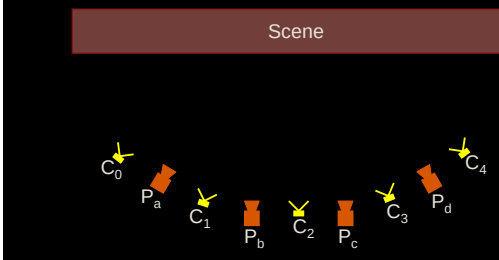
Our Approach

- We introduce a new multi-viewpoint acquisition method for 3D scenes of arbitrary size where we can combine captures without having to know the relative positions of the capture device within the scene
  - Acquisition consists of merely "taking a set of pictures"
  - Multiple captures can be easily refined and combined
  - This also enables triangulation systems to use wide-baselines

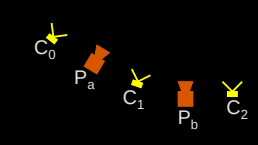
Our Approach

- Acquisition consists of an alternating sequence of taking pictures and establishing correspondences
  - We take pictures using an internally calibrated camera pair
  - We establish correspondences using an uncalibrated projector
    - (note: using feature tracking is an option too)
- We sample and reconstruct the scene surfaces by creating a network of views and correspondences without any knowledge of the location of the camera-pair or of the projector

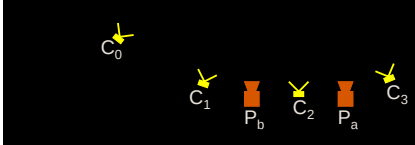
Example Acquisition Sequence



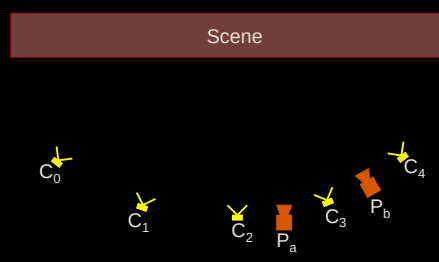
Example Acquisition Sequence



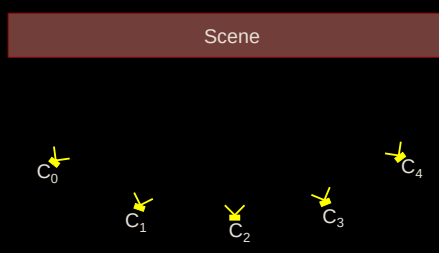
Example Acquisition Sequence



Example Acquisition Sequence



Example Acquisition Sequence



Acquisition Sequence Details

- $C_0 \ldots C_{n-1}$ is each really a cluster of (6) pictures
- Many different "alternating camera-projector patterns" exists (and should be explored)
  - But at least 2 of 6 pictures must observe both neighboring projectors in order to yield a unique answer (e.g., like stereo)

Live Demo!

Many more details coming soon…

- Major pending items
  - Choose optimal configuration of baseline and FOV
  - Improve the sample merging
  - Add texture-mapping
  - Remove indirect lighting from structured-light depth estimation