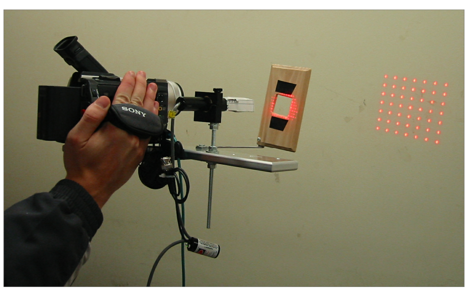
Interactive Scene Modeling from Dense Color and Sparse Depth

Voicu Popescu, Elisha Sacks, Gleb Bahmutov {popescu, eps, bahmutov}@cs.purdue.edu

Scene modeling is a major bottleneck for computer graphics applications. A promising approach is to build the scene model from acquired color and depth data. Systems that build high-quality models are slow (e.g. depth from stereo, laser rangefinding), require expensive and bulky equipment (e.g. laser rangefinding), and require manual intervention (e.g. image-based editing, view morphing, tour into the picture). Systems that do not suffer from these disadvantages build simplistic scene models (e.g. color panoramas) that must be viewed from preferred locations. We present a fast, easy to use, and inexpensive modeling system that builds scene models that support realistic interactive rendering from a wide range of viewing locations. Our system acquires video frames augmented with sparse depth samples. The frames are registered and merged into an evolving model at the rate of five frames per second. The model is displayed continually to provide immediate feedback to the operator. We illustrate the system on an office scene; we obtain meshes of 400,000 triangles in 15 minutes.

1. Acquisition device

Our acquisition device consists of a digital video camera enhanced with a laser system that casts a pattern of 7x7 beams in its field of view. A bright dot is visible in the video frame where a beam intersects a scene surface. The dot is located in the video frame and its 3D position is triangulated by computing the intersection between the video camera ray and the laser beam. We obtain 3 mm depth accuracy at 1m.



Our design provides an even depth sampling of the video frame. The depth data is intrinsically registered with the color data, since depth is inferred from color. This is an advantage over systems that acquire depth and color from separate devices, hence must coregister the data. Dot detection is fast because the dots are confined to fixed epipolar segments in the video frames.

2. Structured scenes

We model scenes in two modes based on their geometric complexity. A scene is called structured when it consists of large smooth surfaces. The laser dots in each frame fall on a few surfaces. These surface patches can be modeled to the required accuracy by fitting polynomials through the dots.

Structured scenes are modeled freehand. The operator sweeps the scene with the camera and the system acquires depth and color, registers this data in the coordinate system of the first frame, and merges it into an evolving scene model. Registration is performed by computing the camera motion between consecutive frames. This motion is computed in three stages: 1) identify the surfaces in each frame; 2) compute a motion that minimizes the distance between the new dots and the old surfaces; and 3) extend the motion to minimize the color difference between selected new rays and the corresponding points on the old surfaces.

The scene is modeled as a collection of depth images that are created on demand. We use depth images because they can be transformed and merged efficiently. The depth images are transformed into texture-mapped triangle meshes that are rendered to provide operator feedback.

3. Unstructured scenes

An unstructured scene consists of many small surfaces. Each surface contains too few laser dots for an accurate polynomial fit. We model such scenes by placing the camera in a bracket that restricts its motion to panning and tilting around its center of projection. The frames are registered using color only. We have developed a real-time stitching algorithm that registers the current frame with respect to the previous frame using a pattern of horizontal and vertical segments in the current frame. The registered frames are merged into a cube map panorama.

The system acquires depth at every registered frame. A triangulation of the acquired dots is obtained in one second. The resulting mesh topology is used to generate 3D triangles. This model, called a depth-enhanced panorama, is an effective representation for unstructured scenes, which are difficult to model. Unlike regular panoramas, depth enhanced panoramas provide motion parallax when the view is translated.

4. Results

We refer the reader to our website http://www.cs.purdue.edu/cgvlab/modelCamera/modelCamera.html for video illustrations of our system and for a gallery of VRML models. Our goal is to model a room in one hour. The remaining research challenges are registration drift and merging depth enhanced panorama. Our system promises to make inside-looking-out modeling an order of magnitude faster and cheaper.

Structured scene.

Depth Enhanced Panorama.