Fully Automatic Acquisition of Hierarchical, Textured 3D Geometric Models of Urban Environments: PI Report

Seth Teller
MIT Computer Graphics Group
545 Technology Square NE43-208
Cambridge MA 02139
seth@ai.mit.edu, http://www.graphics.lcs.mit.edu

Abstract

This paper presents an overview of a planned system for automatic population of geospatial databases which represent urban exteriors as textured geometric model data. The salient features of the system are that 1) it employs "pose-imagery:" high-quality, digital still images, each timestamped, and annotated with a reliable estimate of the acquiring camera's absolute position and attitude; 2) it exploits both dense and sparse image features by sifting evidence for interpixel correlation across every image; 3) it addresses the combinatorial aspects of large-scale 3D reconstruction from very large number of images; and 4) it has no "human in the loop."

This work involves both engineering and research challenges. The engineering challenges include rapid acquisition of 6-DOF pose-instrumented, high-resolution imagery, and its insertion into a hierarchical three-dimensional data structure. The research challenges include dense reconstruction techniques using thousands of images, and the incremental construction of a textured three-dimensional model from selected, spatially related image subsets.

Also presented are an evaluation plan for the project, as well as a description of the synthetic, hybrid, and actual datasets to be acquired and processed.

1 Introduction

We have proposed the construction of a prototype system whose objective is the fully automatic population of geospatial databases (APGD) for built-up areas. We are developing algorithms to integrate ground-based, low-flyby, and satellite imagery; robustly handle illumination changes and occluding foliage; and gracefully yield multi-resolution models in the presence of data with varying quality. The generated models will have a form suitable for visual simulation, collision detection, change detection, line of sight simulation, and other physically-based simulation operations. The system will also support model amplification (given additional observations) and modification (to incorporate the effects of, e.g., demolition, renovation, or construction activity).

The objective of the system is to fundamentally accelerate and improve the population process for complex geospatial databases, surmounting the scaling problem and human dependencies posed by existing semi-automatic systems, and providing a fast new model acquisition capability for simulation operations. An operational system would also have significant commercial applications: entertainment, embedding of commercial databases, tourism, education, etc.

2 Objectives

The principal goal of the system is to populate textured geospatial databases in a fully automated fashion, and to dramatically reduce the cost and time required to populate such databases as compared to existing semi-automatic (i.e., human-assisted) systems. Using many thousands of close-range images, rather than a few, and annotating them with high-quality pose estimates in an absolute Earth coordinate system, make the "footprint" of this effort quite distinct from those of existing efforts. We hope to learn much and develop a class of useful 3D reconstruction algorithms novel from both research and engineering standpoints.

The fundamental engineering strategy employed by our system is the use of recently available global-positioning systems (GPS) and PC-hosted, high-quality accelerometry which yield continuous estimates of camera position and orientation in an absolute coordinate system. The resulting "pose-camera" will ease some of the most significant hindrances to achieving automated 3D reconstruction. We have developed several novel algorithms that process pose-imagery to: establish sparse (edge) and dense (region) correspondences; identify regions of empty space; and reconstruct globally consistent 3D models (with confidence bounds) from local, occluded observations. Absolute pose estimates are used to avoid combinatorial blowup while processing very large numbers (thousands) of images. The algorithms' salient features are described elsewhere in these proceedings.

Another principal system objective is that the quality and articulation of the reconstructed models should degrade gracefully with the precision and resolution of the acquired pose and image data. In particular, pose imagery arising from "fast" drive-bys or fly-bys will yield crude but consistent 3D models. Higher resolution, or more densely sampled, image data will enable reconstructed geometry and texture of correspondingly higher fidelity (model "amplification" or "intensification"). We plan to validate our algorithms by contrasting their output with the product of traditional (human-effected) site modeling techniques.

3 Project Plan

We are developing both an innovative device (a pose-camera) and innovative algorithms (for 3D data organization and reconstruction) for automated population of a geospatial database. The geospatial database represents the geometric and reflectance characteristics of built structures observed in the physical world, as well as significant trees and vegetation. All data are modeled by templates, that is, canonical parametrized objects fit to multiple observations. The generated 3D model will support line-of-sight computations, physically-based collision detection, and other simulation modes. Associated reflectance information increases realism during simulation by enabling reillumination by synthetic light sources.

We will pursue four specific subtasks in order to demonstrate system feasibility. First, we will design and build a pose-instrumented digital camera. Second, we will deploy the instrument in a known test environment to determine its performance. Third, we will deploy the pose-camera in a complex, unmodeled environment, and develop software algorithms for deriving 3D textured geometry and foliage representations from the acquired pose imagery. Fourth, we will assess the system utility and cost by manually modeling some portion of the automatically-acquired dataset, and comparing traditional techniques with our methods.

We distinguish between three types of pose-imagery data used for algorithm development and testing. Simulated data is synthetic imagery with arbitrary specified pose. Hybrid data is actual imagery, with pose estimates derived *a posteriori* with the help of a semi-automatic photogrammetry system. Actual pose imagery is that produced by the operational pose camera.

4 Engineering and Research Challenges

Engineering challenges in this project include rapid acquisition of 6-DOF pose-instrumented, high-resolution imagery, and its insertion into a hierarchical three-dimensional data structure.

Instrumentation packages for determining absolute position and orientation exist, but (as commercially available) in a largely unintegrated state. We are assembling integrated, PC-hosted instrumentation which will maintain pose estimates through a combination of global positioning, accelerometry, and odometry. This instrumentation will be mounted on a wheeled, human-propelled acquisition platform (Fig. 1), along with a high-resolution digital camera, on-board PC and digital tape storage, and a battery-based power source.

Figure 1: A model of the pose camera.

Raw high-resolution image data demands an enormous amount of storage space. An external-memory spatial data structure mediates storage and processing of both the annotated imagery and the reconstructed model data. Moreover, specular effects and changes in lighting conditions during acquisition will complicate data collection and matching efforts. However, these same factors also make possible estimation of the directional reflectance properties of each reconstructed surface.

Due to the size of the input and output datasets, only the most skeletal global operations, such as insertion of representations of acquiring cameras into a hierarchical spatial data structure, will be possible. Global, pairwise matching strategies are unworkable, since they would result in combinatorial explosion. Instead, local operations will correlate imagery acquired by proximal cameras, or suspected to contain observations of related physical structures. Incremental construction and insertion of reconstructed geometry from overlapping, adjoining imagery subsets must be supported. Associating spatially proximal cameras avoids both falsely negative and falsely positive matches that arise in schemes relying only upon temporal coherence (Fig. 2).

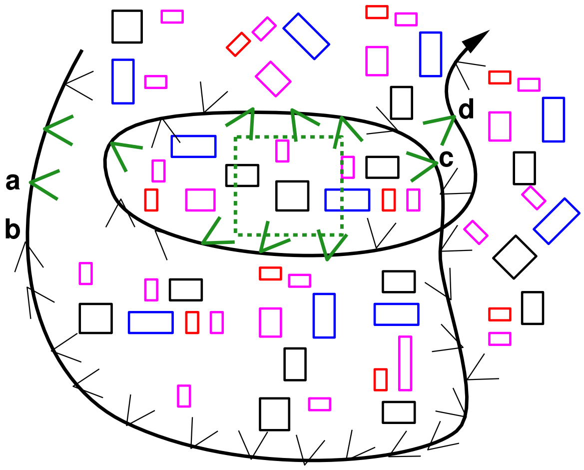

Figure 2: Temporal coherence can wrongly associate images (**a,b**) or fail to associate relevant images (**c,d**).

We augment traditional "sparse" (point- and edge-based) correspondence algorithms with "dense" processing algorithms which correlate every pixel in every image with some other pixel set. The goal, of course, is to find that representation of the 3D world most consistent with the totality of 2D observations (pose images). Conservative error bounds will be maintained for all reconstructed features and structures. Confidence in this representation should increase with more numerous, more finely sampled, or more higher-resolution, acquired imagery.

Foliage imagery can be a significant hindrance to reconstruction of urban exteriors. We address this issue with a model-based foliage reconstruction scheme. Synthetic images of procedurally generated foliage are subjected to filters, and their responses associated with the foliage generator settings. Imagery of actually occurring foliage is then subjected to analogous filtering in order to identify similar procedural foliage. The objective is not to reconstruct foliage directly, but rather produce procedural foliage representations morphologically similar to those observed.

5 Evaluation Plan

We plan to evaluate the pose camera's accuracy by deploying it in a known (manually-surveyed) arena and testing its reported pose against independently derived estimates. We will establish mean and maximum errors in camera position and camera attitude as a function of time and distance from start, both for slow, regular motions, and for rapid translation and rotational acceleration.

We are completing a careful "ground-truth" survey of the Technology Square area, in order to manually align it with NAD-83 survey data for Cambridge. We have chosen fiducial points in and around Technology Square toward which the pose-camera will be directed to generate desired *a priori* poses.

We will evaluate the quality of initial reconstructions by comparison to ground truth in the form of manual survey data, and architectural facilities data maintained by the campus Office of Facilities and Management. We will determine the accuracy and resolution of 3D feature reconstruction, and assess the faithfulness of the reconstructed datasets. We plan also to develop models of degradation of reconstructed data with loss of image resolution or pose accuracy.

During year two, we will attempt to achieve basic, block-based feature and building reconstruction of Technology Square and some portion of the MIT campus (from five to two hundred structures). The pose-camera will be manually moved through the relevant areas. Our initial system implementation will generate reconstructed geometry and texture data. Textures (reflectance maps) for each building facade will be generated by aggregating disparate pose images, to be compared to "ground truth" by reference to axis-aligned photographs of building facades acquired manually under measured, nearly diffuse lighting conditions.

Finally, we plan to evaluate the throughput and cost of the system according to several metrics. The rate of pose-imagery acquisition will be measured. The computational costs of 3D reconstruction will be assessed, along with dollar estimates of system overhead and per-feature cost. Finally, system operation will be compared, for a single large dataset (imagery of several hundred structures), with that of an expert human operating a traditional semi-automatic photogrammetry system.

6 Conclusion

We have described an ambitious, but feasible, vision of a system for fully automatic population of textured geospatial databases representing built-up areas, with no human in the loop except to direct motion of the pose-camera. The system design exploits recent advances in pose instrumentation. Geospatial data entities and associated textures are incrementally deduced from a large collection of images with pose estimates of varying accuracy. The system output is a collection of geometric entities, each with a conservative confidence bound, organized in a hierarchical spatial database suitable for external-memory algorithms such as those employed by real-time simulation systems. Because the acquired dataset is fully three-dimensional, it can be subjected to collision detection, line of sight, arbitrary lighting and atmospheric conditions, and other physically-based or phenomenologically-based simulation operations.