3D Exploitation of Large 2D Urban Photo Archives*

Peter Cho and Ross Anderson, MIT Lincoln Lab, cho,ross@ll.mit.edu
Noah Snavely, Cornell University, snavely@cs.cornell.edu.

Introduction

The quantity, quality and availability of urban imagery are rapidly increasing over time. Millions of photos shot by inexpensive digital cameras in cities can now be accessed via the web. But usually no connection exists between retrieved urban thumbnails other than their having been collected in a common metropolitan area. Some organizing principle is therefore needed to enable intuitive navigating and efficient searching of vast imagery archives.

Fortunately, three-dimensional geometry provides such an organizing principle for images collected at different times, places and resolutions. For example, a set of digital ground photos represents 2D projections of 3D subvolumes onto a variety of image planes. If world-space geometry is captured in a 3D urban map, it can be used to mathematically connect a pixel in one picture onto its counterparts in other ground photos. Moreover, the map acts as a repository for knowledge distilled from multiple cameras. Ultimately, urban situational awareness comes much more directly from high-level knowledge encoded in the map than from low-level data on which it is based.

In this talk, we present a 3D approach to exploiting 2D urban imagery following the algorithm flow diagrammed in figure 1. We focus upon implementing the most computationally intensive step in this flow on a parallelized cluster. Current cluster performance results indicate reconstruction of entire cities lies within reach. We close with future military and civilian applications of this work which include augmented reality systems, city mission planning and robotic navigation.

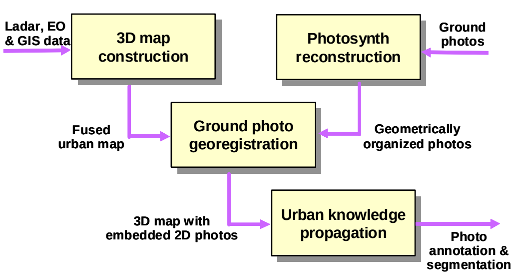

Figure 1: Algorithm flow diagram

3D map construction

Urban environments exhibit complex patterns in geometry and color. To mathematically capture these patterns, we first fuse satellite imagery with ladar height data. We also introduce Geographic Information System layers into the 3D map that generally contain points, curves and regions. Figure 2 displays one interesting example of a fused map for New York City. It serves as a global backdrop into which other sensor data localized in space and/or time such as ground digital photos may be organized.

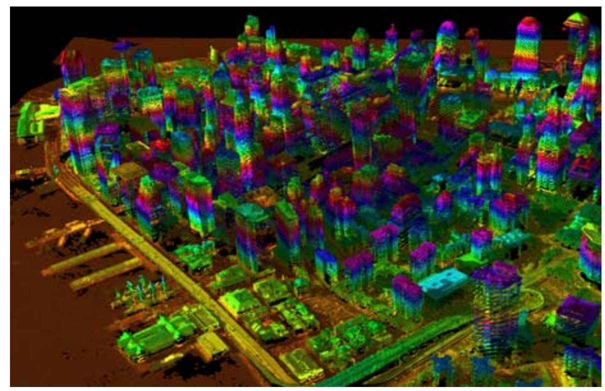

Figure 2: Fused 3D NYC map

Photosynth reconstruction

It is natural to regard a digital photo as a 2D array of RGB pixels. But from a geometrical standpoint, a photograph represents an angle-angle projection onto an image plane, and its snapping is modeled by a 4×3 homogenous matrix. Computing the entries in this projection matrix is tantamount to calibrating the camera and reconstructing its 3D viewing frustum. In order to reconstruct several hundreds and eventually many thousands of images, we are implementing "Photosynth" algorithms developed by Snavely et al [1] on Lincoln Lab's "LLGrid" cluster [2].

The reconstruction codes first extract Scale Invariant Feature Transform (SIFT) features [3] from a large input set of photos which must have some non-negligible degree of overlap. They next form tracks for corresponding features running throughout the photo set. The tracks become inputs to structure from motion algorithms which return the 3D positions and orientations of all input cameras as well as the 3D locations of target points corresponding to SIFT features. After fully parallelizing feature extraction and matching, we find LLGrid performance exceeds that of a previously employed, smaller cluster by 400%.

Photosynth generates graph connections among *a priori* unstructured sets of digital photos (see figure 3). To date, we have processed 1200 images on LLGrid. We are currently working towards applying the cluster to multi-thousand photo collections and look forward to crossing the 10,000 image mark in a few weeks.

* This work was sponsored by the Department of the Air Force under Air Force Contract No. FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

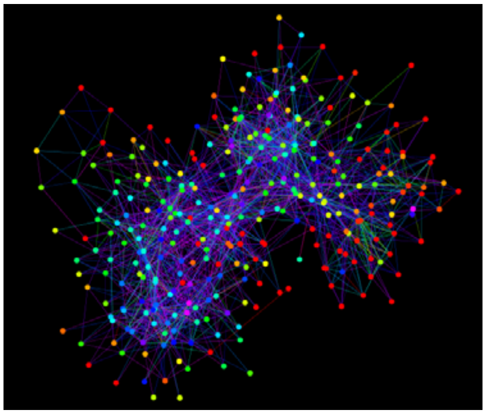

Figure 3: Graph of reconstructed photos whose edges indicate overlapping feature content between image nodes.

Ground photo georegistration

Computer vision techniques alone cannot determine the absolute position, orientation and scale for cameras that shot some set of photos. But after relative 3D camera output from Photosynth is combined with corresponding world space information from a 3D urban map, the projection matrices for large numbers of ground-level city photos can be fixed. Figure 4 displays reconstruction results for a preliminary set of 318 photos georegistered with our NYC map.

Navigating through geo-organized images is much more natural than through random urban thumbnails. In computer graphics rendering programs, the virtual camera can assume the same position and orientation as some real camera which took a picture. It may then peer through the 2D image and observe its 3D geometry progenitor (see figure 5). Moreover, the virtual camera can tilt past the sides of the photo to observe the surrounding urban environment and thereby enhance contextual understanding. The virtual camera may also smoothly transition to another photo that displays some target of interest in greater detail or from a different angle. Conducting such virtual city tours significantly enhances urban situational awareness.

Figure 4: Reconstructed photos georegistered with the 3D NYC map

Urban knowledge propagation

Once photographs have been geoaligned, high-level information can project from 3D world-space onto their 2D image planes. For instance, GIS points-of-interest are tied to specific longitude, latitude and altitude coordinates. When those geopoints are projected into photos, their associated semantic tags follow with them. Names for buildings and streets can automatically appear in photos based upon their 3D → 2D correspondences. Moreover, we demonstrate how city structures' absolute ranges and heights can be measured from pictures registered with the fused map. And we illustrate how abstract knowledge can flow from one 2D photo to another via 3D geometry.

We conclude with the provocative claim that image-based querying of 3D organized photos will rival Google text-based search for information mining in the next decade.

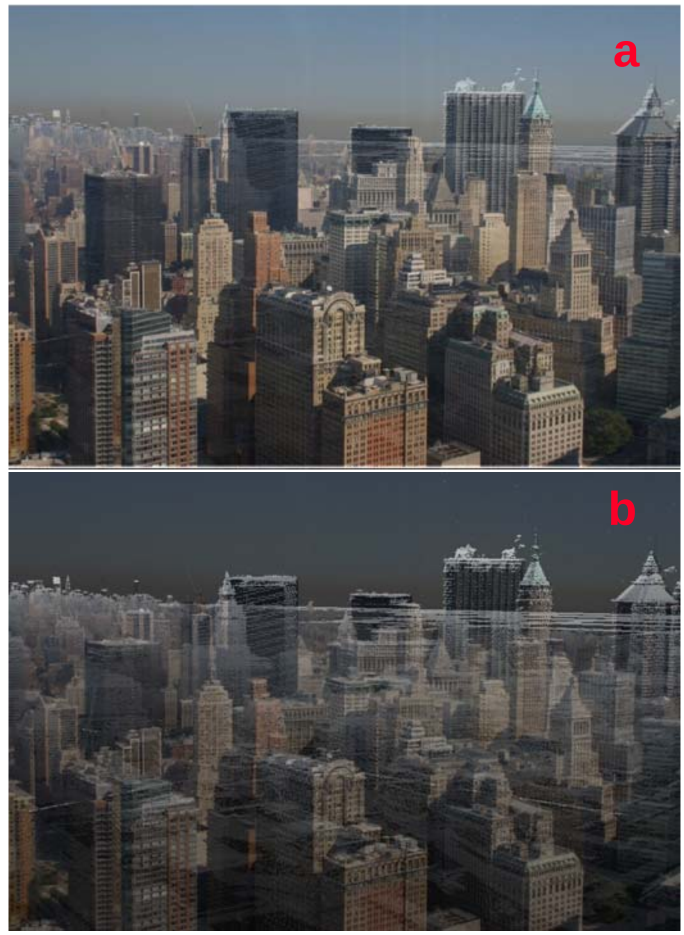

Figure 5: Alignment between the 3D map and a reconstructed photo with (a) 33% and (b) 66% alpha blending.

References

[1]. N. Snavely, S.M. Seitz and R. Szeliski, *Photo Tourism: Exploring Photo Collections in 3D*, ACM Transactions on Graphics, **25** (2006) 835.

[2]. N. Travin-Bliss, R. Bond, J. Kepner, H. Kim & A. Reuther, *Interactive Grid Computing at Lincoln Laboratory*, LL Journal, **16** (2006) 165.

[3]. D.G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision **60** (2004) 91.