Proceedings of IEEE Computer Vision and Pattern Recognition, December 2001

Calibrated, Registered Images of an Extended Urban Area

Seth Teller Matthew Antone Zachary Bodnar
Michael Bosse Satyan Coorg Manish Jethwa Neel Master
MIT Computer Graphics Group

Abstract

We describe a dataset of several thousand calibrated, geo-referenced, high dynamic range color images, acquired under uncontrolled, variable illumination in an outdoor region spanning hundreds of meters. All image, feature, calibration, and geo-referencing data are available at `http://city.lcs.mit.edu/data`*.*

Calibrated imagery is of fundamental interest in a wide variety of applications. We have made this data available in the belief that researchers in computer graphics, computer vision, photogrammetry and digital cartography will find it useful in several ways: as a test set for their own algorithms; as a calibrated image set for applications such as image-based rendering, metric 3D reconstruction, and appearance recovery; and as controlled imagery for integration into existing GIS systems and applications.

The web-based interface to the data provides interactive viewing of: high-dynamic-range images and mosaics; extracted edge and point features; intrinsic and extrinsic calibration, along with maps of the ground context in which the images were acquired; the spatial adjacency relationships among images; the epipolar geometry relating adjacent images; compass and absolute scale overlays; and quantitative consistency measures for the calibration data.

1. Introduction

This paper describes data produced by a system for calibrated, terrestrial image acquisition in urban areas. The system is end-to-end, in the sense that it acquires uncalibrated images as input, and produces accurately calibrated, geo-referenced images as output, with no human interaction other than operation of the sensor.

Four key ideas distinguish our approach [27] from other methods. Every image is annotated with an absolute, GPS-based position estimate as it is acquired, enabling efficient discovery of adjacent (and likely related) images in subsequent processing. The sensor acquires omni-directional images for more robust recovery and accurate estimation of scene structure and camera motion. Probabilistic, projective uncertainty models are used throughout the system to represent noisy features and camera pose. Finally, all of the system's algorithmic components have asymptotically linear time and space requirements, enabling their application to very large datasets. Detailed descriptions of the system, sensor, and processing components can be found elsewhere [25, 26, 11, 12, 10, 6, 7, 27].

An extensive collection of calibrated, high dynamic range (HDR) image data produced by the system is now available for interactive viewing and download at `http://city.lcs.mit.edu/data`. We envision at least three ways in which the data may be useful to other researchers. First, the uncalibrated imagery (i.e., raw sensor data) can be used as a test dataset by other researchers developing intrinsic or extrinsic calibration methods. Second, the registered imagery (i.e., data produced by our registration algorithms) can be used in a variety of applications including image-based rendering, scene reconstruction and texture estimation. In either context, the scale and extent of the data should pose an interesting collection of challenges. Finally, geo-referenced (Earth-relative) extrinsic calibration enables the imagery to be readily incorporated into a variety of existing GIS and digital cartography applications (e.g. OpenGIS [4], TerraServer [5], and the National Spatial Data Infrastructure [3]).

The paper is organized as follows. Section 2 describes the acquisition and processing stages used to produce the dataset. Section 3 describes several quantitative consistency measures for the image metadata. Section 4 describes a web interface to the dataset. (Information about file formats and organization is deferred to an Appendix, and documented on-line.) Section 5 describes existing acquisition methods for georeferenced imagery. Section 6 summarizes the contributions of the paper.

2. Acquisition and Processing Stages

The sensor, consisting of an actuated pan-tilt head and navigation instrumentation [10], is deployed in sessions typically lasting a few hours each. The operator repeatedly positions the sensor and initiates node acquisition. A node is a set of images with common optical center, acquired under pure rotation of the pan-tilt head from a stationary platform. After each session, the sensor is returned to the lab where the data is uploaded and subjected to a series of processing stages culminating in a geo-referenced image dataset. The remainder of this section describes the steps involved.

2.1. Intrinsic Calibration

We use a fixed lens for each session, but small variations in lens attachment, temperature, etc. can perturb intrinsic parameters, requiring re-calibration. At the start of each acquisition session, several images are taken of a calibration pattern to produce initial estimates of the camera intrinsics and the radial lens distortion parameters. We use a public-domain implementation of Zhang's calibration algorithm [29] to estimate the camera's focal length, aspect ratio, skew, center of projection, and the first and second radial distortion parameters.

2.2. Photometric Calibration

The amount of light entering the camera varies as the reciprocal of f^2, where the f-number is the ratio of focal length to aperture diameter. A second off-line calibration step recovers an absolute radiance level for each HDR pixel value, as a function of shutter aperture, up to a single absolute scale factor [14]. This allows image pixels acquired through different apertures to be compared. (Dependence on exposure time is removed by the HDR calibration process.)

Absolute radiance calibration need be done only once for a particular camera CCD. For calibration we expose it to a bright indoor light source (not the sun) and acquire a HDR image, then set an absolute radiance scale such that the brightest pixel value maps to 1.0 (i.e., zero on a log scale). Dark image calibration is not necessary since the magnitude of heat noise at our longest exposure time (1/10th second) is insignificant. A pixel value of zero is under-saturated and does not have a valid radiance value; such pixels are ignored in further processing. In practice very few such pixels are present in the data set, since images are acquired in daylight.

2.3. Mosaic Tiling Design

A rough estimate of focal length suffices to design the mosaic, or tiling pattern, which covers a portion of the sphere during omni-directional image acquisition. An off-line program generates a series of camera orientations (i.e., azimuth and altitude settings for the sensor pan-tilt head) which tile the sphere while ensuring sufficient overlap for the mosaic generation process (Figure 1). The tile generator also writes the image adjacency graph to an ASCII file for use by the mosaic generation process.

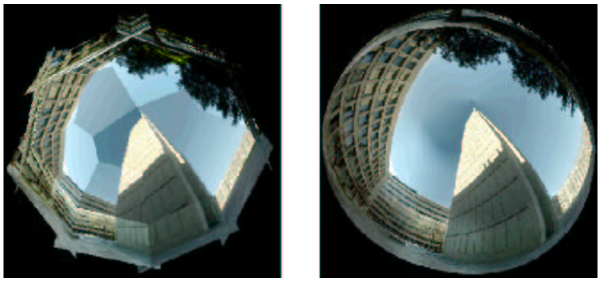


Figure 1. A node tiling, and the resulting mosaic.

2.4. HDR Image and Metadata Capture

The camera is attached to a verniered mount, adjusted so that the pan-tilt head's center of rotation coincides with the camera focal point. At each head position, the camera captures one tile of the image mosaic at multiple exposures. Several frames with the same exposure time are averaged to reduce image noise. The resulting images are combined on the sensor platform to produce a high-dynamic-range (HDR) image [14], which is stored in a log-radiance format [20].

The sensor platform annotates every acquired HDR image with a camera descriptor, date- and time-stamp, camera intrinsics and estimated Earth-relative position and orientation [10]. The absolute pose estimates are typically accurate to a few meters of position and a few degrees of orientation, but can be worse when GPS conditions are poor, for example due to obstructions, multi-path or electromagnetic interference.

2.5. Data Upload and Spatial Indexing

After each acquisition session, the sensor is reconnected to our local network for data upload. Uploaded images are inserted into a spatial index [21] keyed on position and time-stamp. This enables downstream processes to invoke proximity queries on images, to discover, for example, which images were acquired within a specified region or time interval.

2.6. Correction of Radial Distortion

Images are resampled to remove radial lens distortion. This enables downstream computations to use a simple pin-hole camera model. For example, linear features in the world map to straight lines in the image after resampling.

2.7. Image Pyramid Generation

The corrected images are then filtered down to half-, quarter-, and 3/32−resolutions to form a four-level image pyramid. The quarter-, half- and full-resolution images are used in series for multi-resolution mosaic generation, 3D reconstruction, and texture estimation. The 3/32-resolution images are used only for fast visualization.

2.8. Mosaic Generation

The initial rotation estimates from the acquisition rig are not pixel-accurate. However, they are sufficiently accurate to initialize multi-resolution mosaic generation [13], culminating in high-resolution mosaics for each node (cf. Figure 1). The resulting omnidirectional mosaic is used only for visualization; whenever an image sample is needed by any batch processing stage, the system samples directly from the raw images using the refined pose estimate for each image. This avoids the need for any further resampling of the images.

2.9. Sub-Pixel Feature Detection

Our registration algorithms use gradient-based, subpixel edge and point features (Figure 2). Point (corner) features are generated by intersecting edges in image space, when they are separated by at most 2° (or about 40 pixels at our highest image resolution, 1 milliradian per pixel), and form an angle of at least 5°.

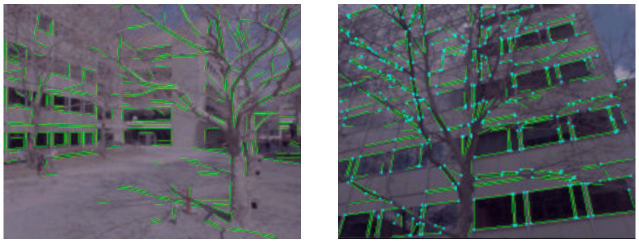


Figure 2. Edge and point features for two images.

2.10. Rotational Registration

The rotational registration stage [6] takes intrinsic calibration information, edge features, and the node adjacency graph as inputs. It groups observed edge features into scene-relative vanishing points (VPs). Each node is assumed to have viewed at least two VPs in common with those VPs observed by its neighbors. Nodes are brought into rotational alignment by registering each to the set of common observed VPs in its vicinity (some nodes observe fewer than two VPs, and cannot be aligned).

2.11. Translational Registration

The translational registration stage [7] takes intrinsic calibration information, point features, the node adjacency graph, and rotationally registered nodes (from the previous stage) as inputs. It produces revised position estimates for every node, valid up to an arbitrary Euclidean transformation (translation, rotation and isotropic scaling).

2.12. Geo-Referencing

The final stage registers the node set to the original GPS (i.e., absolute, geo-referenced) position estimates, determining rigid translation, scaling, and rotational degrees of freedom for the entire dataset [7]. The resulting pose estimates are geo-referenced to absolute (Earth) coordinates and are metrically meaningful. The coordinates are stored, in units of meters, relative to a local tangent plane (LTP) with its origin defined as the location of a nearby GPS base station. Rather than ECEF (Earth Centered Earth Fixed) coordinates, we use LTP coordinates due to their convenience and accuracy for local computations. (The web site documents the transformation between LTP and WGS-84 Geoid ECEF [17] coordinates.)

The resulting absolute exterior orientation (pose) assignments are globally consistent on average to about 5 centimeters of position and about 0.1° of orientation, and epipolar geometry is consistent on average to within four pixels, at 95% confidence (i.e., at containment of 95% of the probability density for each node), for node subsets spanning several hundred meters [7].

3. Self-Consistency Measures

Manual generation of ground truth for this dataset is infeasible due to its complexity and extent. Therefore we have formulated a set of quantitative self-consistency measures for the dataset, each assessing the degree to which separate observations are mutually consistent. These consistency measures take into account both local and global aspects of the dataset, but abstract away low-level errors arising from slight intrinsic mis-calibration and sub-pixel feature localization, the magnitude of which we estimate at one to two pixels.

3.1. Local Consistency Measures

The quality of each node, or mosaic, can be characterized by the following pixel-based and feature-based consistency measures:

- **Mosaic pairwise cross-correlations.** For each node, we log the sum, over all adjacent images within the mosaic [13], of the pixel inverse cross-correlation normalized by the number of overlapping pixels and the pixel dynamic range.

- **Node VP spreads.** Each node typically observes one or more vanishing points (VPs). For truly parallel world lines, perfectly calibrated mosaics, and noise-free edge features, VPs could be localized with no error. In reality, all of these elements exhibit noise, so we represent VPs as projective probability densities [9]. We log the width of the density, in degrees, at 95% confidence – that is, the size of the symmetric region of the density which includes 95% of its probability [6].

3.2. Global Consistency Measures

The self-consistency of extrinsic calibration for a node set can be characterized by several scene-relative consistency measures:

- **Scene VP spreads.** When multiple mosaics observe a common vanishing point, the degree to which they agree on the absolute direction of the VP is a measure of global consistency. We log the width of each consensus VP direction, using the 95% confidence bound described above [6].

- **Node position spreads.** Similarly, we characterize the uncertainty in recovered node positions by evaluating the average and maximum sizes at which 95% confidence bounds are reached for the recovered Gaussian densities [7].

- **Epipolar alignment.** Our translation registration algorithm produces soft (probabilistic) correspondences between observed scene points. For each soft point match with probability greater than a threshold (we use $p \geq 0.8$), we log the mean, maximum, and variance of the distance in image space between the 2D point feature and the epipolar line of its correspondent feature [7].

4. Web Interface

An interactive interface to the dataset is available at `http://city.lcs.mit.edu/data`. The interface depicts acquired nodes overlaid on a geo-referenced map (Figure 3). Each node is color-coded by the type of calibration metadata available for the node (some nodes have no revised position estimates, so are posted only with orientation estimates).

The user may select any individual node for examination, producing a node inspection page in which the full node mosaic and the node's constituent (log-format) images are displayed (Figure 4). This page also presents the omni-directional image mosaic for the node, which can be panned and zoomed interactively, exposure-adjusted, and displayed with overlaid edge and point features, as well as a directional compass. The page includes links to ASCII pose data and consistency information for the node.

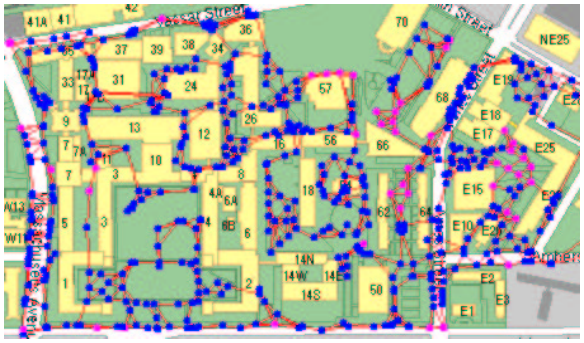


Figure 3. A node set and adjacency graph. Points represent node locations; edges represent node adjacencies. Adjacent nodes are typically 10-30 meters apart.

Near the mosaic viewer is a "mini-map" of node context, showing the node's neighbors in the adjacency graph. Selecting a node from the mini-map brings up the inspection view for the indicated node. Selecting an edge from the mini-map brings up an epipolar geometry view for the implicated node pair (Figure 5). This view depicts each node position as a small cross in the counterpart node. The user can indicate a point in either node, and see the point's epipolar line (ruled with metric tick marks) displayed in the adjacent node. Finally, the user can visualize any available pose information for the nodes, and consistency measures such as inter-node baselines and global vanishing points.

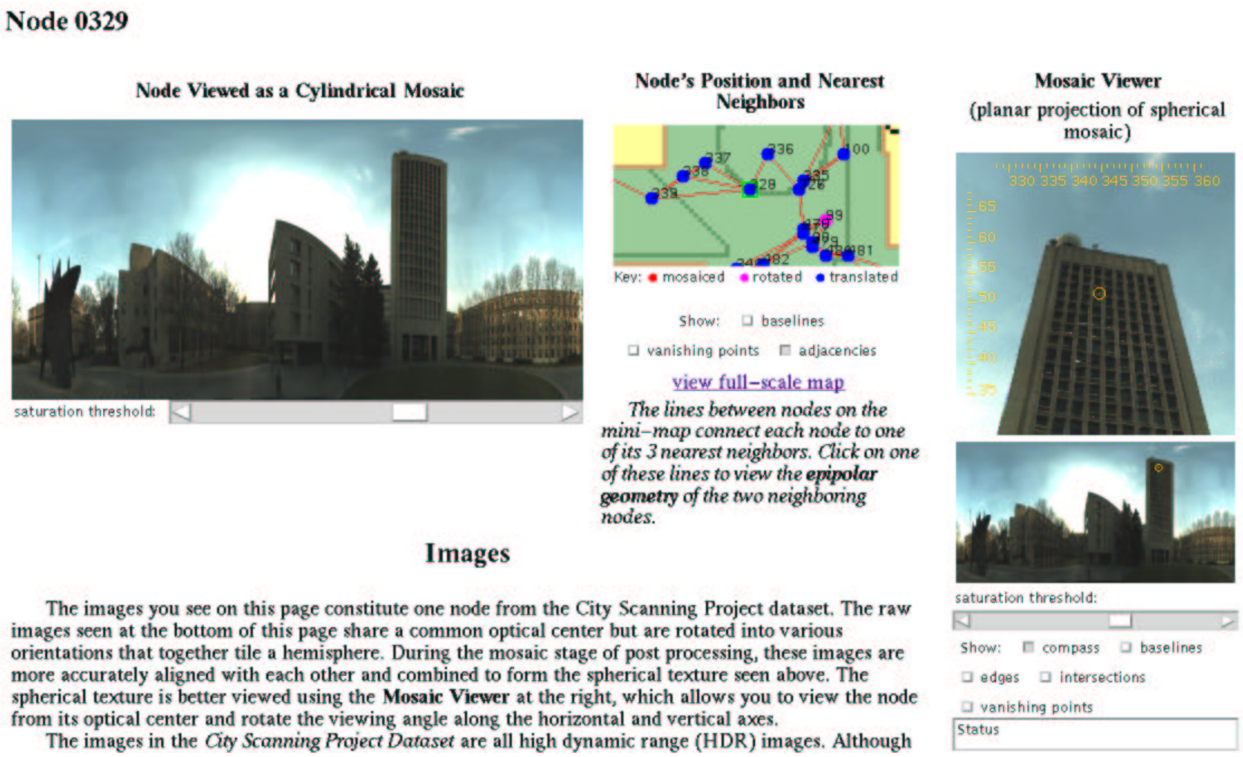
5. Related Work

This section briefly reviews other systems and methods for acquiring geo-referenced image datasets.

5.1. Satellite-Based Acquisition

A number of robotic mapping systems (e.g., [2]) employ satellites to acquire high-altitude imagery of Earth. These systems provide a wealth of data about regions with limited vertical relief (oceans, much natural terrain, etc.). Since detailed information is maintained about each satellite's orbital parameters, acquired images can be geo-referenced fairly accurately by the sensor itself; one data interchange site for publically available imagery [1] states registration to within about twelve meters on the surface of the Earth. Commercial and classified sensors may provide even greater accuracy.

Such systems are less useful for imaging urban canyons. Here, a satellite at a great distance can acquire only near-nadir views (in which case most vertical surfaces are imaged very obliquely) or near-horizon



Node 0329

Node Viewed as a Cylindrical Mosaic

Node's Position and Nearest Neighbors

Mosaic Viewer

(planar projection of spherical mosaic)

The lines between nodes on the mini–map connect each node to one of its 3 nearest neighbors. Click on one of these lines to view the ***epipolar geometry*** *of the two neighboring nodes.*

Images

The images you see on this page constitute one node from the City Scanning Project dataset. The raw images seen at the bottom of this page share a common optical center but are rotated into various orientations that together tile a hemisphere. During the mosaic stage of post processing, these images are more accurately aligned with each other and combined to form the spherical texture seen above. The spherical texture is better viewed using the **Mosaic Viewer** at the right, which allows you to view the node from its optical center and rotate the viewing angle along the horizontal and vertical axes.

The images in the *City Scanning Project Dataset* are all high dynamic range (HDR) images. Although

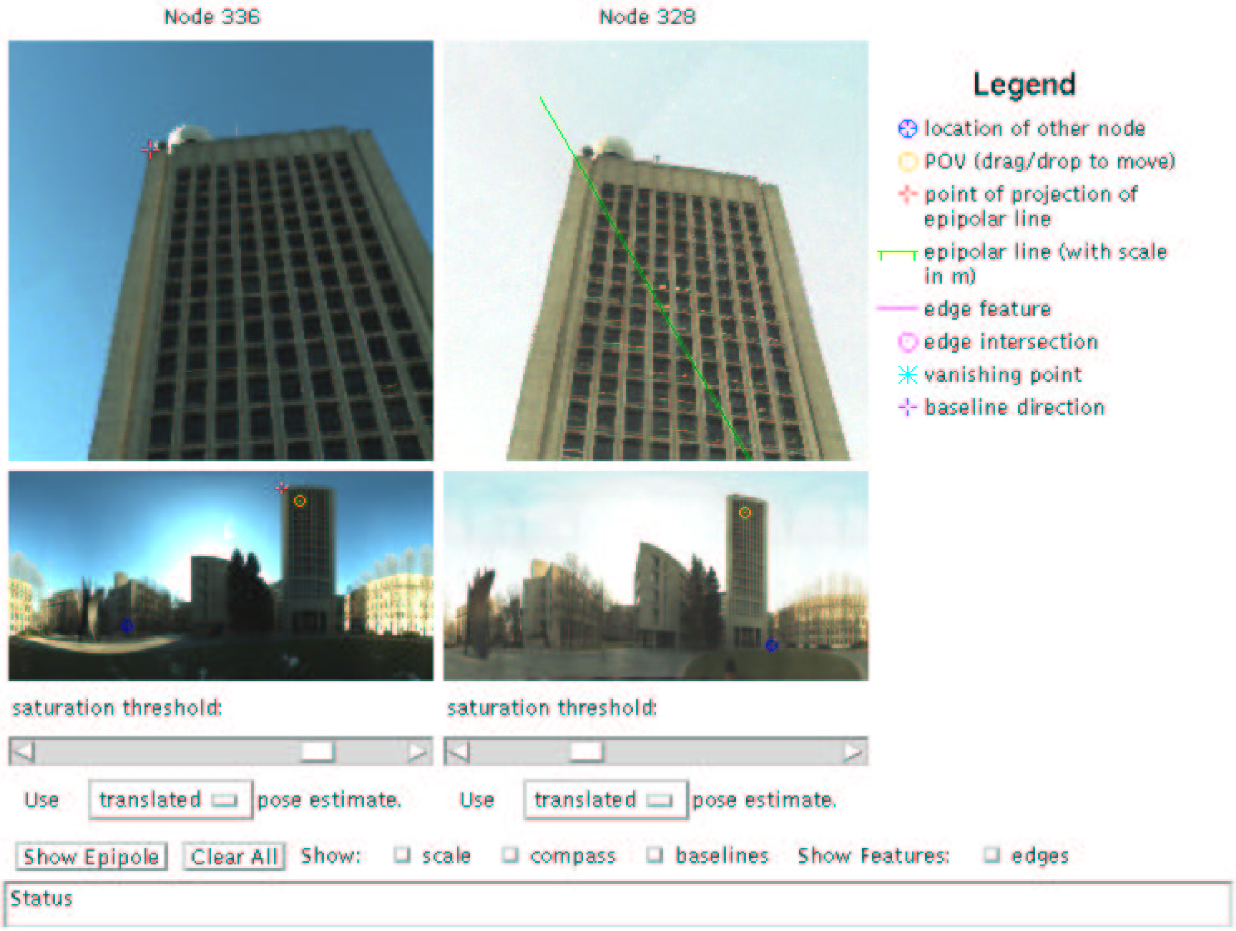
Figure 4. A portion of the node inspection page.

views (in which case most vertical surfaces are occluded). For urban environments, a near-ground sensor is necessary to acquire unoccluded, nearly fronto-parallel views of vertical surfaces. Autonomous low-altitude flying vehicles (e.g. [22]) exist, but have not to date acquired accurately geo-referenced imagery.

5.2. Semi-Automated Methods

An alternative production method for near-ground, geo-referenced imagery is the use of manual interaction. A variety of semi-automated methods have been proposed to control small image sets, in applications including photogrammetry [28, 24, 16], digital mapping, and computer graphics scene modeling [8, 19, 15, 23]. These systems combine automated, semi-automated, or manual feature detection with semi-automated bundle adjustment, in which a human operator indicates or selects corresponding features across multiple images. Sometimes, geo-referenced points (e.g., painted crosses) have been placed in the scene before the sensor is deployed, and are therefore available in the imagery. In this case, the human user can geo-reference the imagery by associating features visible in the image to known features (ground control) in an existing geo-referenced dataset.

Semi-automated methods are fundamentally limited in a number of respects. First, these methods are scale-limited; the number of person-hours required to process a dataset of more than a few hundred images would be prohibitive in most applications. (The L.A. Basin modeling project has expended an estimated 100,000 hours, or fifty person-years, of human effort tying acquired imagery to an extended site model [18].) The modeling process can not be "parallelized" straightforwardly by adding workers, due to the need for coordination among operators. Human operators typically rely on every pair of images overlapping in some fashion, so that common elements can be indicated; in extended datasets, most image pairs have no common elements due to occlusion. Second, interactive methods are vulnerable to human failings: errors and short-cuts. Humans may make errors by indicating incorrect matches in ambiguous situations (for example in the presence of occlusion and visual clutter). Also, human operators tend to specify only as many constraints as are required for nominal convergence of the underlying optimization, rather than entering over-determined constraint sets. In practice, this leads to unstable bundle adjustment. Finally, semi-automated methods do not scale with underlying technology (i.e., CPU speed), but rather have the human operator, whose throughput is essentially fixed, as their bottleneck. Thus, the throughput of semi-automated systems typically improves little over time.



Use translated pose estimate.

Use translated pose estimate.

Figure 5. A portion of the epipolar geometry inspection page.

5.3. Summary

Summarizing, although there are sensor platforms which produce geo-referenced imagery, the data they produce is not suitable for close-range mapping of urban regions. Similarly, although there are interactive techniques for registering images using photogrammetric bundle adjustments, they do not scale to the huge image datasets needed for modeling extended areas. Prior to the acquisition system used to collect the datasets described in this paper, no scalable, automated method had been demonstrated to acquire close-range, accurately geo-referenced imagery of urban areas.

6. Conclusion

This paper described an intrinsically and extrinsically calibrated terrestrial image dataset acquired over a spatially extended urban region. Due to its scale and extent, achieving a comparable dataset using current semi-automated methods would be infeasible.

The dataset is available on-line for interactive browsing and download. We have made this data available to the research community in the hope that it will be useful to researchers and developers of large-scale image self-calibration and registration algorithms, image-based rendering and metric 3D reconstruction from calibrated image datasets, and digital cartography and GIS applications.

A. Appendix: Data Organization

This appendix describes image, calibration (intrinsic and extrinsic), and feature data at each stage of system processing. We also describe the contents of a number of data files used within the system to represent these elements, and the organization of these files in the online repository `http://city.lcs.mit.edu/data`.

A.1. Coordinate Information

A coordinate information file resides at the top of the data hierarchy, specifying an external (Earth-relative [17]) coordinate system for reference, and the dataset's LTP origin and coordinate axes expressed in this coordinate system.

A.2. Image Descriptor Files

All images are stored in a lossless RGB format (the web site supplies Java code to read and display this format). For each image, the sensor generates an ASCII image descriptor file containing all shuttering-specific information except the camera's intrinsic and extrinsic parameters, including:

- A header field;
- Digital camera identifier (make and model);
- Date and time (GMT) of image acquisition;
- Image source (sensor or post-process);
- Image width and height (in pixels);
- Image and pixel types (log-radiance RGB);
- Exposure bracketing and photometric calibration information; and
- Lens type and aperture used.

The image files store eight bits per color channel. (We also retain raw 16-bit log-color images, prior to Bayer color interpolation and radial distortion correction.) We convert each color channel value into a radiance value proportional to the light flux entering the camera as:

$$r \propto \exp\left[\left(\left(p/255 * \left(\mathtt{R_{MAX}} - \mathtt{R_{MIN}}\right)\right) + \mathtt{R_{MIN}}\right) - 2\ln\left(f\right)\right],$$

where $\mathtt{R_{MAX}}$ and $\mathtt{R_{MIN}}$ denote the maximum and minimum observed radiance values for the node, and the `APERTURE` field is used to shift the log-radiance values by $-2\ln f$, effectively scaling by $1/f^2$.

Raw data values of 0 and 255 are used as sentinels to mark regions in the images that are undersaturated and saturated, respectively. There are few such pixels in the data; they occur only when the camera observes very dark areas, or very bright specular reflections, or the sun itself.

A.3. Camera Pose Descriptor Files

For each image, the sensor logs a camera pose descriptor file, capturing what is known of the camera's intrinsic and extrinsic parameters at the instant of shuttering. This file includes fields describing:

- Digital camera identifier (make and model);
- Image source (sensor or post-process);
- Image width and height (in pixels);
- Focal length (in pixels);
- Principal point c_x, c_y (in pixels);
- Skew (assumed zero);
- Camera position in LTP (x, y, z);
- Camera orientation, as a quaternion (q_0, q_1, q_2, q_3).

Subsequent processing stages may append additional fields to the pose descriptor file, describing the outcome of processing. For example, upon successful termination, the `mosaic` process writes a `CONVERGENT` tag and a `RESIDUE` value to each pose file that it revises.

A.4. Node Descriptor Files

Each node has a "node information file," containing the number of images in the node, and the index (usually zero) of the "base image," or first image acquired for the node. Each image is indexed, zero-relative, with respect to the base image. With each node is also associated a "mosaic adjacency graph"; this file lists, for each image, its neighbors in the node tiling.

A.5. Node Adjacency File

Adjacency information among nodes is represented as a list, for each node, of the node's neighbors (cf. Figure 3). Our spatial index produces an adjacency list for any input dataset and specified number of, or maximum distance to, nearest neighbors.

A.6. Feature Descriptors

Per-image edge features and intersection-based point features, both localized to sub-pixel position, are stored in ASCII format.

A.7. Vanishing Point Descriptor Files

The rotational registration stage detects the scene vanishing points (VP) observed by each node. The results are stored in a VP descriptor file, which contains the VP direction (in node coordinates) and variance (i.e., width at 95% confidence) in degrees squared.

After rotational registration of all nodes, a vanishing point descriptor file is produced for the entire dataset. This file describes the union of all vanishing points observed by all nodes in the dataset, and analogous confidence measures. Global VP directions are expressed in absolute, scene-relative coordinates.

A.8. Baseline Descriptor Files

The translational registration stage operates on all pairs of adjacent nodes, establishing a baseline direction (up to scale) for the pair expressed in world coordinates. Each computed baseline is logged to a separate file, along with an estimate of its uncertainty.

A.9. Pose Data Organization

The refined pose data resulting from each processing stage is organized in a series of directories named `initial`, `mosaic`, `rotation`, and `translation`.

The `initial/` directories contain (rough) pose estimates, useful (for example) to establish adjacencies among images, or test image registration or egomotion recovery algorithms. The `mosaic/` directories contain image orientation information sufficient to compose multiple node images into a single mosaic; this is useful for visualizing the field of view around any single sensor position. The `rotation/` pose directories contain accurate node orientations in a single absolute coordinate system, useful (for example) to sample sky illumination or construct far-field environment maps. Finally, the `translation/` directories contain metric, geo-referenced pose estimates, useful for applications requiring projective or Euclidean calibration (e.g., image-based rendering, or metric scene reconstruction), or for integrating the images with existing GIS data or systems.

Acknowledgements

Support for this research was provided in part by the Office of Naval Research under MURI Award SA 1524-2582386, and in part by Intel Corporation.

References

[1] GeoTIFF, `http://www.remotesensing.org/geotiff/geotiff.html`.

[2] NASA's Earth Observing System, `http://eospso.gsfc.nasa.gov/earth_observ.html`.

[3] National Spatial Data Infrastructure, `http://www.fgdc.gov/nsdi/nsdi.html`.

[4] Open GIS Consortium, `http://www.opengis.org`.

[5] TerraServer, the Vertical Portal for Overhead Imagery, `http://www.terraserver.com`.

[6] M. Antone and S. Teller. Automatic recovery of relative camera rotations for urban scenes. In *Proc. CVPR*, pages II–282–289, June 2000.

[7] M. Antone and S. Teller. Scalable, absolute position recovery for omni-directional image networks. In *Proc. CVPR (to appear)*, 2001.

[8] S. Becker and V. M. Bove. Semiautomatic 3-D model extraction from uncalibrated 2-D camera views. In *Proc. Visual Data Exploration and Analysis II, SPIE Vol. 2410*, pages 447–461, 1995.

[9] C. Bingham. An antipodally symmetric distribution on the sphere. *Annals of Statistics*, 2(6):1201–1225, Nov. 1974.

[10] M. Bosse, D. de Couto, and S. Teller. Eyes of argus: Georeferenced imagery in urban environments. *GPS World*, pages 20–30, April 1999.

[11] S. Coorg, N. Master, and S. Teller. Acquisition of a large pose-mosaic dataset. In *CVPR '98*, pages 872–878, 1998.

[12] S. Coorg and S. Teller. Extracting textured vertical facades from controlled close-range imagery. In *Proc. CVPR '99*, pages 625–632, June 1999.

[13] S. Coorg and S. Teller. Spherical mosaics with quaternions and dense correlation. *IJCV*, 37(3):259–273, 2000.

[14] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97 Conference Proceedings*, Aug. 1997.

[15] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *SIGGRAPH '96 Conference Proceedings*, pages 11–20, Aug. 1996.

[16] C. Greeve. *Digital Photogrammetry: an Addendum to the Manual of Photogrammetry.* American Society of Photogrammetry and Remote Sensing, 1997.

[17] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins. *GPS Theory and Practice.* Springer-Wien, 1997.

[18] W. Jepson. Personal communication, Feb 2000.

[19] W. Jepson, R. Liggett, and S. Friedman. Virtual modeling of urban environments. *PRESENCE*, 5.1, March 1996.

[20] G. Larson. LogLuv encoding for full-gamut, high-dynamic range images. *Journal of Graphics Tools*, 3:15–31, 1998.

[21] F. P. Preparata and M. I. Shamos. *Computational Geometry: an Introduction.* Springer-Verlag, 1985.

[22] C. Sanders, P. DeBitetto, E. Feron, H. Vuong, and N. Leveson. Hierarchical control of small autonomous helicopters. In *Proc. 37th IEEE Conference on Decision and Control*, Dec. 1998.

[23] H. Shum, M. Han, and R. Szeliski. Interactive construction of 3-d models from panoramic mosaics. In *Proc. CVPR*, pages 427–433, 1998.

[24] C. Slama. *Manual of Photogrammetry.* American Society of Photogrammetry and Remote Sensing, 1980.

[25] S. Teller. Automatic acquisition of hierarchical, textured 3D geometric models of urban environments: Project plan. In *Proc. the Image Understanding Workshop*, 1997.

[26] S. Teller. Automated urban model acquisition: Project rationale and status. In *Proc. the Image Understanding Workshop*, pages 455–462, Nov. 1998.

[27] S. Teller. Scalable, controlled image capture in urban environments. Technical Report 825, MIT LCS, Sep. 2001.

[28] P. Wolf. *Elements of Photogrammetry.* McGraw-Hill, 1974.

[29] Z. Zhang. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, Dec. 1998.