Robust Camera Pose Recovery Using Stochastic Geometry

Matthew E. Antone & Seth Teller

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: The objective of 3-D machine vision is to infer geometric properties (e.g. shape and size) and photometric attributes (e.g. color, texture, reflectance) from a set of 2-D images. Every such vision task relies on accurate camera calibration, that is, knowledge of the camera's intrinsic parameters (focal length, lens distortion, etc.) and extrinsic parameters—orientation, position, and scale relative to a fixed frame of reference. This research concerns the automatic recovery of precise extrinsic pose among a large set of images, assuming that accurate intrinsic parameters and rough estimates of extrinsic parameters are available. This work also investigates models of geometric uncertainty and the application of projective inference techniques.

Motivation: The MIT City Project [2] employs a set of unique algorithms to produce accurate, metric reconstruction of urban landscapes from *pose imagery*, or images annotated with approximate geo-referenced position and orientation. All components of the system are completely automatic except for extrinsic camera registration, which is obtained via manual correspondence of point features across images followed by global bundle adjustment. The acquired data consists of a large set (tens of thousands) of images, making the the human-assisted component quite tedious. Thus the primary goal of this work is to provide fully automatic, scalable external registration.

Previous Work: There exists an enormous body of work in automatic camera registration, which is most often coupled with the recovery of 3-D structure. Some techniques use known calibration targets to determine camera pose [1]. Others track point features over image sequences, assuming small spatial/temporal separation, and then use classical structure from motion [5] or projective reconstruction methods [4]. Still others determine the most probable pose and structure over all possible correspondence sets [3], circumventing the need for explicit correspondence. However, existing technique possess neither the scalability nor the robustness required for automatic pose recovery in a large, wide-baseline data set.

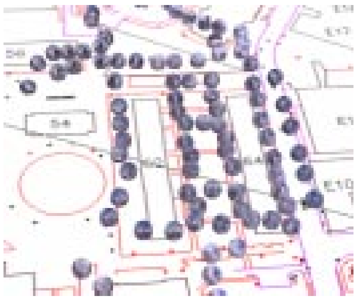
Approach: Orientation can be recovered independently of position via the detection and registration of translation-invariant image features (*vanishing points*). An expectation maximization (EM) technique initialized by a Hough transform produces accurate vanishing point estimates in each image from a set of unclassified line features; these estimates are aligned by another EM algorithm which produces optimal rotational pose and uncertainty without explicit correspondence between vanishing points.

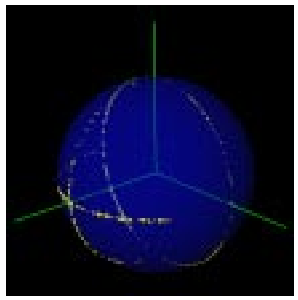
Known orientation constrains the epipolar geometry of point features, reducing the determination of local translational offset to a simple linear form. The baseline (direction of translation) between each pair of adjacent images is estimated using a Hough transform coupled with various geometric constraints, and refined by a Monte Carlo EM algorithm that finds the best motion direction over all possible correspondence sets. Local translation estimates are then assembled into a global set of constraints to determine camera positions with respect to a common coordinate frame. The final result is an assignment of accurate pose and uncertainty to every relevant camera.

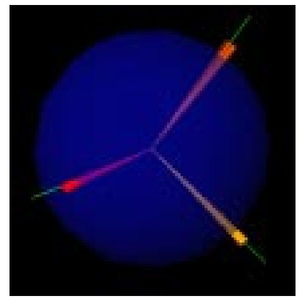
Difficulty: Several challenges must be overcome in order to achieve completely automated pose registration in a real-world system. Viewing conditions vary widely among images due to changes in weather, time of day, and baselines, causing significant variation in 2-D structure and texture. Unwanted objects such as trees occlude objects of interest, cluttering images with spurious features and obstructing desired features. The features themselves are corrupted by sensor noise and detection artifacts. Nothing is known in advance about 3-D structure or correspondence, so certain regular structures (e.g. window grids) give rise to pose ambiguity. Finally, the large size of the data set necessitates scalable algorithms.

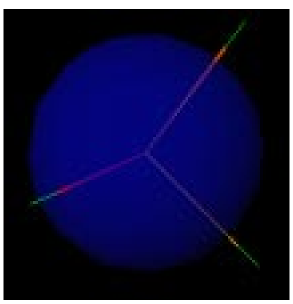
Impact: This system exhibits properties not possessed by other vision systems, such as scalability and robustness, which allow for significant variation in viewing conditions among a very large set of images. Recovery of precise camera pose and uncertainty facilitates accurate, detailed 3-D reconstruction via a host of algorithms. Although the primary application context is the City Project, few domain-specific restrictions are required or imposed, allowing these techniques to be applied to more general situations.

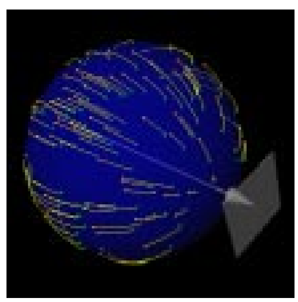
Future Work: Many extensions can be made to the geometric models and methods proposed in this work. Vanishing points can be used to significantly reduce unwanted image and feature clutter in various subsequent tasks. Domain-specific assumptions such as coplanarity of points and orthogonality of lines can assist in the recovery of first-order intrinsic camera parameters and further constrain epipolar geometry. Translational registration can be improved by using lines, which are detected more reliably than points. Finally, explicit feature correspondence can be inferred from pair-wise baseline estimates and propagated, allowing for global bundle-adjustment.

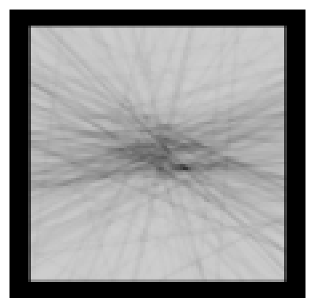

Camera configuration


Edge clusters


Raw vanishing points


Refined vanishing points


Epipolar geometry


Hough transform of correspondences

Research Support: This work is supported by the Advanced Research Projects Agency (under contract DACA76-97-K-0002), the Office of Naval Research (under MURI award SA 1524-2582386), and Intel Corporation.

References:

[1] X. Chen, J. Davis, and P. Slusallek "Wide Area Camera Calibration Using Virtual Calibration Objects". In *Proceedings of CVPR 2000*, Vol. 2, pp. 520-527.

[2] S. Coorg, N. Master, and S. Teller. "Acquisition of a Large Pose-Mosaic Dataset". In *Proceedings of CVPR 1998*, pp. 282-289, June 1998.

[3] F. Dellaert, S. Seitz, C. Thorpe, S. Thrun "Structure from Motion without Correspondence". In *Proceedings of CVPR 2000*, pp. 557-564, June 2000.

[4] A. Fitzgibbon and A. Zisserman. "Automatic Camera Recovery for Closed or Open Image Sequences". In *Proceedings of ECCV 1998*, pp. 1261-1269.

[5] H. Shum, Q. Ke, Z. Zhang "Efficient Bundle Adjustment with Virtual Key Frames: A Hierarchical Approach to Multi-Frame Structure from Motion". In *Proceedings of CVPR 1999*, Vol. 2, pp. 538-543.