Minimal Parameterization of Multiple View Geometry

Raquel Romano

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: The automatic extraction of the unknown geometric relations among a set of cameras capturing images of a common scene is a fundamental problem in computer vision. This work addresses the problem of estimating the configuration of arbitrarily many cameras using image points matched across multiple frames, without requiring a full reconstruction of the viewed scene points.

Motivation: Fast and inexpensive processors, cameras, and storage are making feasible a variety of applications involving image sequences and collections, ranging from visually intelligent automobiles to enhanced sports viewing to home video manipulation. A prerequisite for many of these applications is knowledge of the physical positions of the cameras that captured the original images, whether it be a collection of stationary cameras placed around a common scene, or a single video camera capturing frames as it moves through an environment. Accurate estimation of the geometry of multiple views is the bedrock of such applications.

Previous Work: Projective geometric models of multiple views give rise to algebraic relations among the locations of points in uncalibrated images that are 2D projections of rigidly moving points in the scene. These algebraic relations are multilinear in the image coordinates and their forms depend on the unknown projective parameters relating the cameras. The projective parameters in turn depend on the external and internal camera parameters, which may be directly recovered without having to estimate the 3D structure of the scene points[3].

There exist 2-, 3-, and 4-view projective multilinear constraints, expressed via the fundamental matrix, the trifocal tensor, and the quadrilinear tensor, respectively, that have been well-studied and used in many practical situations, while their N-view analogs have been explored theoretically[5, 6]. Since the projective relationships among more than 4 cameras in general position are completely described by the collection of all 2-view relationships, a single N-view model is not necessarily required. Moreover, the projective multilinearities that simultaneously describe how N matching points must be related have exponentially many linear parameters that are highly coupled through their dependence on the camera properties and positions. Enforcing these parameter dependencies is a difficult and active area of research [2] and an important consideration for modeling the geometry of a large collection of views, such as all the frames in a video sequence.

Approach: We present an N-focal projective model of camera geometry that is minimally parameterized and, by construction, inherently geometrically consistent. Using the fact that all multifocal constraints of a collection of cameras in general position may be reduced to a collection of bifocal constraints, we build an N-view model from a sparse set bifocal parameters that exploit the dependencies among the pairwise geometric relationships.

For each new view added to a camera collection with known configuration, its relation to the entire collection is completely defined by its bifocal relation to two other views, via the positions of the epipoles in each view, and the epipolar collineation that maps epipolar lines from one view to the other[1]. For an isolated view pair, these quantities have a total of 7 degrees of freedom, but once the new view's relation to a single existing view is fixed, its relation to a second view in the collection is constrained to only 4 degrees of freedom. We explicitly model these dependent bifocal relations, and show how for the dependent view pair, each epipole's variation is constrained to a line (leaving 1 DOF each), and the epipolar collineation is constrained by one line match (leaving 2 DOF). We compute this minimal parameter set from point matches using a Levenberg-Marquardt minimization, initialized from the linear estimation of the fundamental matrix [4].

We test the minimally parameterized model on a simulated camera configuration viewing a block of points at the scene's origin. Two algorithms are compared: the nonlinear 7-parameter optimization of independent fundamental matrices, and the nonlinear optimization of dependent 4-parameter bifocal models. Since the projective models

are defined only up to an arbitrary projective transformation, we use the intrinsic parameters of the simulated cameras to recover the Euclidean positions from the bifocal parameters and display the solutions. Figure 1 shows the recovered camera positions overlaid on the ground truth camera positions, and Table 2 shows the improved accuracy in camera position gained by enforcing bifocal dependencies.

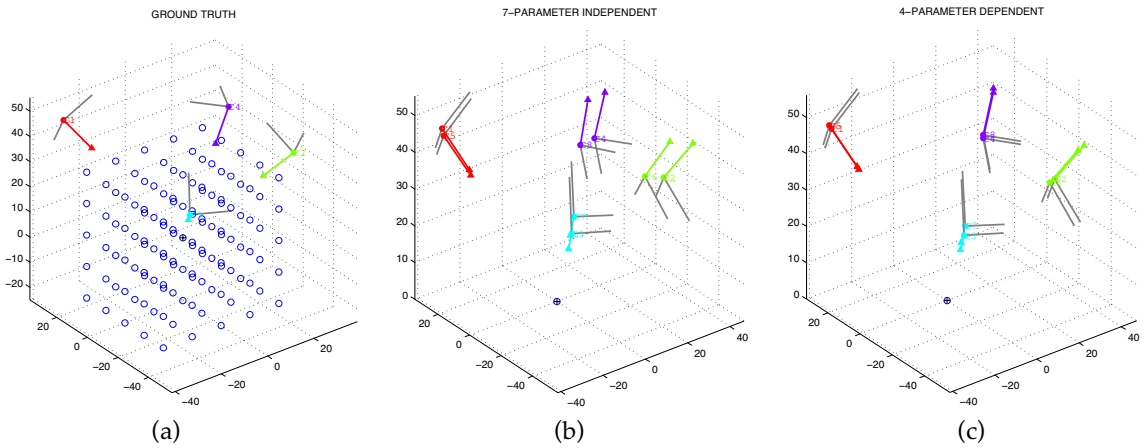

Figure 1: Estimates of four camera positions. (a) Ground truth. camera positions and viewed points. (b) Independent 7-parameter fundamental matrices. (c) Dependent 4-parameter bifocal parameters.

Camera	1	2	3	4
(a) Nonlinear Independent	2.2068	2.2068	2.2068	2.2068
(b) Nonlinear Dependent	0.1593	0.1593	0.1593	0.1593

Figure 2: Translation errors for recovery of camera positions from bifocal parameters. (a) Independent 7-parameter estimates. (b) Dependent 4-parameter estimates.

Impact: While camera motion estimation is a mature problem in the field of computer vision, it is far from being solved. A geometrically consistent model of arbitrarily many views that can be automatically recovered from matching points is essential to any application whose performance depends on first establishing the relative positions of cameras.

Future Work: A difficult and pervasive issue is the degeneracy of camera configurations, particularly in videos sequences taken with approximately linear camera motions. If a model has too many parameters, such a sequence will be overfit, while if a model has too few parameters, the complexity of some sequences will be too rich to describe. It would be useful to use information criteria to determine which subsets of views are degenerate and thus automatically choose appropriate view pairs to model.

Research Support: This research is supported in part by DARPA under ONR grant N00014-97-0363.

References:

[1] Gabriella Csurka, Cyril Zeller, Zhengyou Zhang, and Olivier D. Faugeras. Characterizing the uncertainty of the fundamental matrix. *Computer Vision and Image Understanding*, 68(1):18–36, October 1997.

[2] O. Faugeras and T. Papadopoulo. A nonlinear method for estimating the projective geometry of three views. In *International Conference on Computer Vision*, pages 477–484, 1998.

[3] David Forsyth and Jean Ponce. *Computer Vision–A Modern Approach*. Unpublished (Third Draft), 2001.

[4] R. Hartley. In defence of the eight-point algorithm. In *International Conference on Computer Vision*, pages 1064–1070, 1995.

[5] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[6] Bill Triggs. Matching constraints and the joint image. In *International Conference on Computer Vision*, pages 338–343, 1995.