Scalable Rotational Registration of Omni-Directional Image Networks

Matthew Antone Seth Teller
MIT Computer Graphics Group

Abstract

We describe a linear-time algorithm that recovers absolute, scene-relative camera orientations for networks of thousands of terrestrial images spanning hundreds of meters, in outdoor urban scenes, under uncontrolled lighting. The algorithm requires no human input or interaction, and produces orientation estimates globally consistent to roughly 0.1° (2 milliradians) on average, or about two pixels of epipolar alignment, expending a few CPU-minutes of computation on a 250MHz processor.

Ordinary planar images are grouped by optical center into "omni-directional" (wide-FOV) nodes. The algorithm then estimates a local coordinate frame for each node by classifying and combining thousands of line features into a few robust vanishing points (VPs), modeling features and VPs as projective probability densities. Scene-relative 3-D line directions are hypothesized based on all local VPs, and the entire dataset is finally registered to these directions. As output, the algorithm produces an accurate absolute orientation assignment, and an associated uncertainty estimate, for every camera.

The algorithm takes accurate intrinsic parameters, approximate extrinsic orientations, and a connected node adjacency graph as input. It requires that at least two distinct VPs are visible in each node, but makes no assumption about the orientation of any single VP (e.g. that it is horizonal or vertical); nor does it make any assumption about the relationship between any pair of VPs (e.g., that they are orthogonal).

This paper's principal contributions include: the extension of classical multi-camera alignment methods to a probabilistic framework; a new, limited (but more robust and accurate) use of the Hough Transform to initialize an iterative algorithm; and strong quantitative evidence of the superior utility of omni-directional (wide-FOV) images for extrinsic calibration.

We assess the algorithm's performance on synthetic and real data, and draw several conclusions. First, registration of wide-FOV images is fundamentally more robust against failure, and more accurate, than is registration of ordinary imagery. Second, we show that by fusing thousands of gradient-based line features into a few ensemble projective features (vanishing points), the algorithm achieves accurate registration even in the face of significant lighting variations, low-level feature noise, and error in initial rotation estimates. Third, the system surmounts the usual tradeoff between speed and accuracy; it is both faster and more accurate than manual bundle adjustment.

1 Introduction

Intrinsically and extrinsically calibrated imagery is of fundamental interest in a variety of computer vision and graphics applications, including sensor fusion, 3D reconstruction for model capture, and image-based rendering for visual simulation of realistic scenes. Algorithms exist to recover parameters for a fixed camera with little or no manual input [42]. In practice, applications requiring fully calibrated imagery depend on substantial manual effort, for example specification of tie points across multiple images that allow an automated bundle adjustment algorithm to recover extrinsic parameters. Even for small datasets, this manual component can demand many hours of human effort, and is difficult or impossible to partition among several workers.

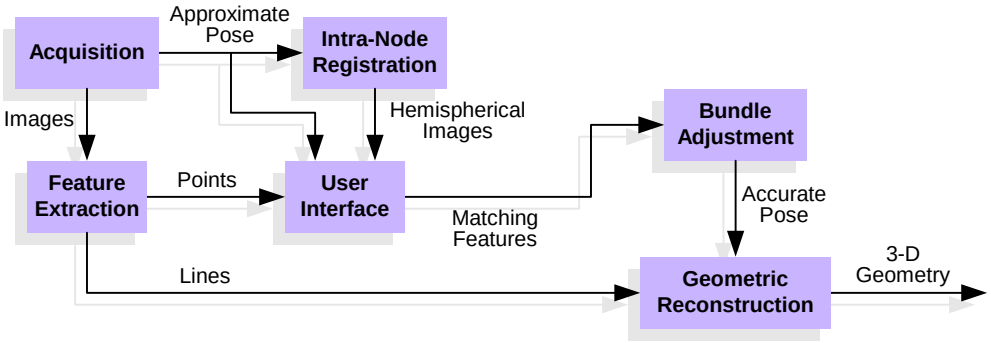


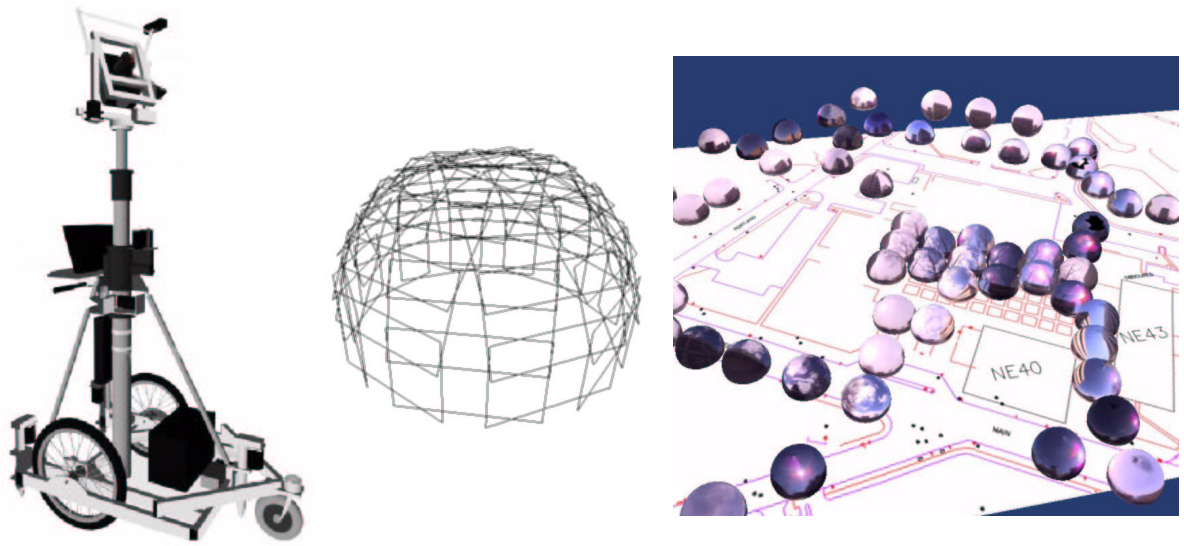
Figure 1: Motivation: An Automated Model Capture System

Images and approximate camera pose are acquired and grouped by optical center into hemispherical images (nodes). An automated process recovers intrinsic and extrinsic camera parameters, after which detailed scene geometry and texture are estimated.

We have developed two automated camera registration algorithms as part of a system (Fig. 1) for automated model capture in extended urban environments [41, 13]. In our system, a human operator moves a sensor [8] to many viewing positions in and around the scene of interest. At each position, the sensor acquires a high-resolution image of some portion of the scene, along with a rough estimate of the acquiring camera's position and orientation, in absolute (Earth) coordinates. The result is a set of images, each with a hemispherical or greater field of view, acquired 15 to 20 meters apart (Fig. 2).

Images are grouped by optical center into single, wide-FOV mosaics called "nodes" [13]. Each node is subsequently treated as a rigid, super-hemispherical image with a single pose. The use of wide-FOV imagery provides a significant advantage in practice, by reducing the number of optimization parameters, and by eliminating classical bias and ambiguities in camera motion estimation [19, 13, 17].

The initial pose estimates are not sufficiently accurate for 3D reconstruction. Under ideal acquisition conditions, they are accurate to a meter of position and a degree of orientation; however, in our urban environment these confidence bounds degrade to several meters and several degrees. Thus one critical component of our system is the refinement of the sensor's initial camera pose estimates to bring all cameras into registration. The scale of the dataset



(a) (b) (c)

Figure 2: Pose-Image Acquisition

(a) Our prototype sensor, which acquires geo-referenced images. (b) A typical omni-directional image configuration. (c) Node locations registered with a ground map.

rules out interactive techniques; thus pose recovery must be fully automated. Solving the general registration problem requires determining six parameters for each camera: three of rotation and three of position. Our approach decouples the 6-DOF problem into a pure rotation (3-DOF) and pure translation (3-DOF) component. This paper addresses only rotational recovery; a companion paper [2] addresses the recovery of absolute positions.

1.1 Algorithm Overview

The goal of our algorithm is to accurately register camera (node) orientations to a single coordinate system. Intuitively, the algorithm detects common scene structure observed by clusters of nearby nodes. Each node is then aligned to this locally observed structure. Since nearby nodes typically view the same or overlapping scene geometry, this process brings proximal nodes into alignment with each other. A global formulation then brings all nodes into alignment in a common coordinate system.

More formally, the algorithm proceeds as follows. Lines in each image are classified into parallel sets and assembled to construct position-invariant features (vanishing points) by exploiting projective duality and inference techniques. An expectation maximization (EM) algorithm, based on a projective mixture model and initialized by a Hough transform, simultaneously estimates multiple vanishing points in each node. Another EM formulation probabilistically correlates vanishing points with scene-relative directions and recovers global orientation and uncertainty for each node, independently of position.

1.2 Input Requirements and Assumptions

To register a set of images, our algorithm requires the following inputs:

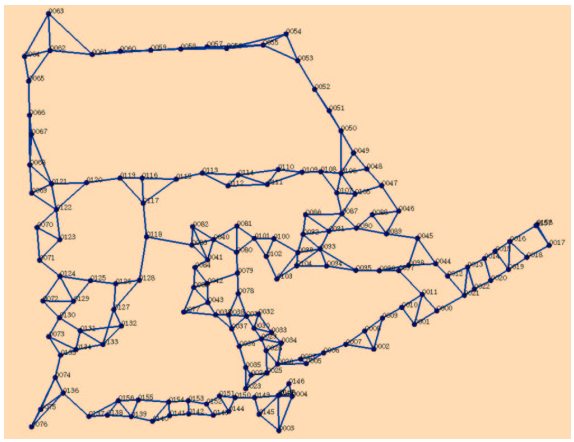


Figure 3: Camera Adjacency
An image network: points represent node locations; edges represent node adjacency.

- ***Accurate intrinsic calibration.*** Non-linear image distortion, if present, is corrected offline, and pin-hole parameters (focal length, principal point, and skew) are supplied.
- ***Rough extrinsic rotations.*** Approximate orientation estimates are supplied for each node, enabling false minima to be avoided during optimization.
- ***Camera adjacency.*** For each node, a list of the node's neighbors is given, identifying cameras likely to have viewed overlapping portions of the scene. This adjacency information is extracted from the sensor's GPS-based position estimates [8].
- ***Line features.*** For each image, sub-pixel gradient based line features are specified, as generate by a modified Canny edge detector [9].

In practice, registration succeeds when the following conditions hold:

- ***Visible vanishing points.*** At least two distinct vanishing points (VPs) are visible in each wide-FOV node (though not necessarily in each narrow-FOV image). These VPs provide a local, translation-invariant coordinate frame for each node.
- ***Overlapping scene geometry.*** Nodes are acquired with sufficient density such that adjacent nodes observe overlapping scene geometry (in this case, two or more common vanishing points). The inter-node distance in our datasets is about fifteen meters.
- ***Wide-FOV nodes.*** Our algorithm is applicable to images with any FOV. However, wide-FOV images are fundamentally more powerful than conventional imagery; they provide maximal observations of surrounding structure, disambiguate small rotations from small translations, reduce bias in inference, and in general enable more reliable convergence and higher accuracy.

1.3 Paper Overview

The remainder of the paper is structured as follows. Section 2 reviews projective feature representations and geometric probability. Section 3 describes the orientation refinement algorithm, and Section 4 reports the result of applying the algorithm to several synthetic and real datasets. Finally, Section 6 reviews previous work on rotational image registration, and Section 7 summarizes our contributions and results.

2 Preliminaries

This section reviews the representations of coordinate transformations and uncertain projective features used by the registration algorithm.

2.1 Extrinsic Pose

A rigid transformation, consisting of a 3×1 translation $\boldsymbol{t}$ and orthonormal rotation $\boldsymbol{R}$, expresses points $\boldsymbol{p}^w$ in world space as points $\boldsymbol{p}^c$ in camera space. Its inverse specifies the orientation and position of the camera with respect to the scene coordinate system. Formally,

$$\boldsymbol{p}^c = \boldsymbol{R}^\top(\boldsymbol{p}^w - \boldsymbol{t}) \quad (1)$$

$$\boldsymbol{p}^w = \boldsymbol{R}\boldsymbol{p}^c + \boldsymbol{t} \quad (2)$$

where $\boldsymbol{t}$ is the position of the focal point, and the columns of $\boldsymbol{R}$ are the principal axes of the camera coordinate system, both expressed in scene coordinates. These two quantities thus summarize the external pose of the camera. We represent rotations by unit quaternions.

2.2 Projective Lines

We represent infinite lines in the (u, v) image plane as

$$au + bv + c = 0 \quad (3)$$

or equivalently

$$\boldsymbol{p} \cdot \boldsymbol{l} = 0 \quad (4)$$

where $\boldsymbol{p} = (u, v, 1)^\top$ and $\boldsymbol{l} = (a, b, c)^\top$. Although the Euclidean plane is a convenient space for feature detection, it is not ideal for feature representation: it implies non-uniform sampling with respect to the focal point, and is unstable for rays nearly parallel to the image plane. This leads to instability and poor conditioning in inference tasks.

Thus to represent line features we use the *projective plane* $\mathbb{P}^2$, a closed topological manifold containing the set of all 3-D lines through the focal point. Points along any given 3-D line, except the focal point itself, constitute an equivalence class $\sim$:

$$\boldsymbol{p} \sim \boldsymbol{r} \quad \Leftrightarrow \quad \boldsymbol{p} = \alpha\boldsymbol{r}, \quad (5)$$

where α is a real nonzero scalar value. Because of the relationship in Eqn. (5), the projective plane is a *quotient space* on $\mathbb{R}^3$ (minus the focal point) and also on the surface of the unit sphere $\mathbb{S}^2$, sometimes referred to in the literature as the *Gaussian sphere* [4]. The sphere's surface is an ideal space for representation of projective features, just as it is an ideal space for image projection: it is closed, compact, and symmetric, and it provides uniform treatment of rays from all directions.

Points in the Euclidean image plane can be transformed to the sphere by augmentation to homogeneous (projective) coordinates and normalizing:

$$(u, v) \quad \rightarrow \quad \boldsymbol{p} = (u, v, 1)^\top \quad \rightarrow \quad \frac{\boldsymbol{p}}{\|\boldsymbol{p}\|}. \tag{6}$$

All collinear points $\boldsymbol{p}_i$ in the image must satisfy the orthogonality constraint of Eqn. (4), which implies two important facts. First, the line parameters $\boldsymbol{l}$ are unique only up to an arbitrary non-zero scale; $\boldsymbol{l}$ itself can thus be viewed as a projective point. Second, $\boldsymbol{p}$ and $\boldsymbol{l}$ are symmetrically related: the roles of the two quantities can be interchanged without altering the constraint. These two facts imply a simple *projective duality* between points and lines, which states that a line $\boldsymbol{l}$ can be represented as a unit direction $\frac{\boldsymbol{l}}{\|\boldsymbol{l}\|}$ on the sphere to which any projective point lying on the line is orthogonal; the set of all such points traces a great circle on the sphere.

Similarly, a given image point $\boldsymbol{p}$ can be viewed as a *pencil* of image lines, all of which contain, and thus intersect at $\boldsymbol{p}$. The parameterizations $\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots$ of such lines must satisfy Eqn. (4), and thus the set of all line duals through $\boldsymbol{p}$ trace a great circle on the sphere orthogonal to $\boldsymbol{p}$ (Figure Fig. 4). We will return to these relationships in considering projective inference and data fusion for the formation of line features and vanishing points.

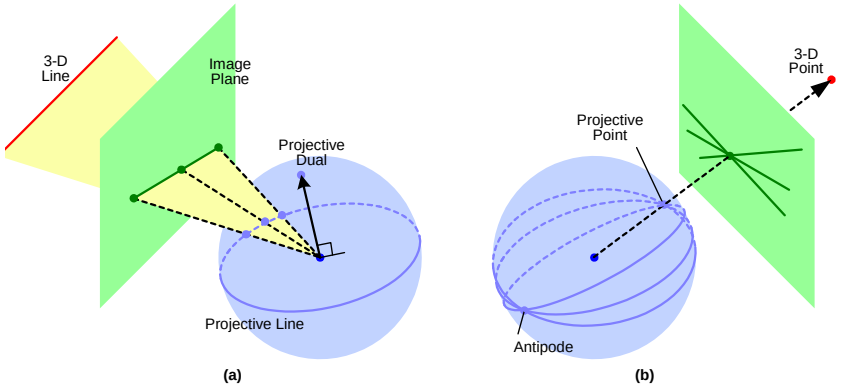


Figure 4: Projective Image Features

(a) A 3-D line can be represented by a 2-D line in planar projection or a great circle in spherical projection. Any point on the line must be orthogonal to the line's dual representation. (b) A 3-D point can be represented as a unit vector on the sphere, or as a pencil of lines passing through its projection.

2.3 Bingham's Distribution

Features viewed by a single camera are inherently projective, since no depth information is available. We wish to represent projective features with suitable spherical probability distributions.

Exponential distributions are useful for inference tasks [6], but the most commonly used multi-variate Gaussian density is a Euclidean probability measure and is therefore not suitable for projective variables. Conditioning a zero-mean Gaussian variable $\boldsymbol{x} \in \mathcal{R}^3$) on the event that $\|\boldsymbol{x}\| = 1$ results in *Bingham's distribution*, a flexible exponential density defined on the unit sphere [7, 23, 44].

This distribution can be generalized to arbitrary dimension, and is parameterized by a symmetric $n \times n$ matrix $\boldsymbol{M}$, and diagonalized into the product $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\kappa}\boldsymbol{U}^\top$, where $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ is a real unitary matrix whose columns $\boldsymbol{u}_i$ represent the principal directions of the distribution and $\boldsymbol{\kappa} \in \mathbb{R}^{n\times n}$ is a diagonal matrix of n concentration parameters κ_i. The density is given by

$$\begin{aligned} p(\boldsymbol{x}) &= \frac{1}{c(\boldsymbol{\kappa})} \exp(\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x}) \\ &= \frac{1}{c(\boldsymbol{\kappa})} \exp\left(\sum_{i=1}^{n} \kappa_i (\boldsymbol{u}_i^\top \boldsymbol{x})^2\right) \end{aligned} \tag{7}$$

where $c(\boldsymbol{\kappa})$ is a normalizing coefficient that depends only on the concentration parameters. We denote this density by $\mathcal{B}_n(\boldsymbol{x}; \boldsymbol{\kappa}, \boldsymbol{U})$, or simply $\mathcal{B}_n(\boldsymbol{x}; \boldsymbol{M})$, with the subscript n denoting the dimension of the space. The matrix $\boldsymbol{M}$ is analogous to the *information matrix* (inverse of the covariance) of a zero-mean Gaussian distribution [35].

The Bingham density is antipodally symmetric, or *axial*: the probability of any point $\boldsymbol{x}$ is identical to that of $-\boldsymbol{x}$. It is closed under rotations: if $\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}$, where $\boldsymbol{R}$ is a rotation matrix and $\boldsymbol{x}$ has Bingham distribution $\mathcal{B}_n(\boldsymbol{x}; \boldsymbol{\kappa}, \boldsymbol{U})$, then $\boldsymbol{y}$ also has a Bingham distribution given by $\mathcal{B}_n(\boldsymbol{y}; \boldsymbol{\kappa}, \boldsymbol{R}\boldsymbol{U})$. Finally, the Bingham representation is expressive: the concentration parameters can describe a wide variety of distributions, including uniform, bipolar, and equatorial (Fig. 5).

The set of concentration parameters is unique only up to an additive shift; in other words, the density is unchanged if a single constant is added to all parameters. By convention, the parameters (along with their corresponding modal directions $\boldsymbol{u}_i$) are ordered from smallest to largest, and shifted by an additive constant so that

$$\kappa_1 \leq \kappa_2 \leq \ldots \leq \kappa_n = 0.$$

Jupp and Mardia [23] have shown that the maximum likelihood estimates of Bingham parameters given a set of deterministic unit-length data points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ is related to the sample second moment matrix

$$\boldsymbol{S}_{\boldsymbol{x}} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{x}_i \boldsymbol{x}_i^\top. \tag{8}$$

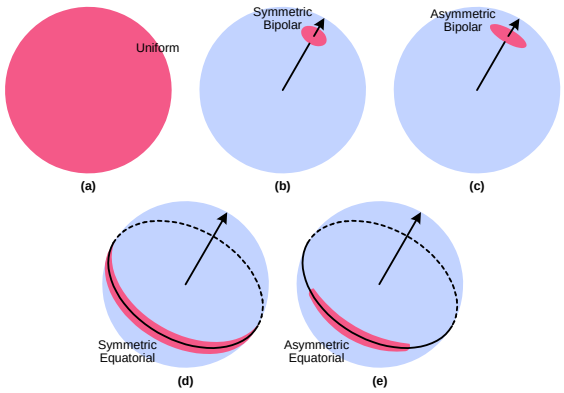


Figure 5: Bingham's Distribution on the Sphere

The shape of iso-density contours on Bingham's distribution depends on the concentration parameters. Examples are shown for $\mathcal{B}_3$. (a) A uniform distribution ($\kappa_1 = \kappa_2 = 0$). (b) A symmetric bipolar distribution ($\kappa_1 = \kappa_2 \ll 0$). (c) An asymmetric bipolar distribution ($\kappa_1 < \kappa_2 \ll 0$). (d) A symmetric equatorial distribution ($\kappa_1 \ll \kappa_2 = 0$). (e) An asymmetric equatorial distribution ($\kappa_1 \ll \kappa_2 < 0$).

If the matrix is diagonalized into $\boldsymbol{S_x} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of the eigenvalues, then $\boldsymbol{U} = \boldsymbol{V}$ (that is, the principal directions of the Bingham distribution are exactly the eigenvectors of $\boldsymbol{S_x}$), and the concentration matrix $\boldsymbol{\kappa}$ is a function of $\boldsymbol{\Lambda}$. It is thus possible to transform the Euclidean sample covariance into a spherical Bingham parameter matrix and vice versa. The unit-length constraint on the $\boldsymbol{x}_i$ ensures that $\text{trace}(\boldsymbol{S_x}) = 1$.

We describe all projective image features on $\mathbb{S}^2$ as Bingham variables $\mathcal{B}_3(\cdot)$. In addition, since unit quaternions are antipodally symmetric and defined on the surface of the unit hypersphere $\mathbb{S}^3$, we represent rotational uncertainty by the Bingham variables $\mathcal{B}_4(\cdot)$.

3 Orientation Recovery Algorithm

This section describes the orientation recovery algorithm (Fig. 6). Section 3.1 reviews the geometry of vanishing points (VPs) on the sphere. Section 3.2 presents a novel method which robustly detects, then accurately estimates, multiple vanishing points in a single node. Section 3.3 extends a classical, deterministic algorithm for rotationally registering two cameras to account for (input) feature and (output) orientation uncertainty. Finally, Section 3.4 uses EM to classify vanishing points, estimate global scene-relative line directions, and refine rotations for an arbitrary number of cameras.

3.1 Vanishing Point Geometry

Parallel 3-D lines viewed under perspective converge to an apparent point of intersection known as a *vanishing point* or VP. VPs have long been used in vision to extract information about scene geometry [4] or egomotion or both.

Consider a 3-D line parallel to some unit direction $\boldsymbol{v}$, and its 2-D projection on the image surface (Fig. 7). The two quantities are projectively equivalent; that is, any projective ray

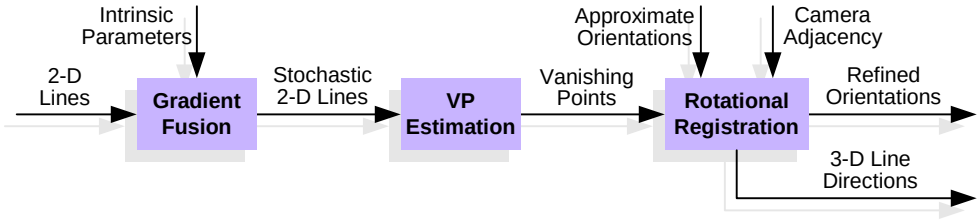


Figure 6: Rotational Registration

Line features are detected in individual nodes and fused into VPs, which are then used to register adjacent image pairs and finally the entire node set.

$\boldsymbol{p}$ that intersects the image line also intersects the scene line. The set of all such rays thus forms a plane $\mathcal{P}$ that includes the focal point, the 2-D line, and the original 3-D line. Let $\boldsymbol{l}$ represent the projective dual of the line, that is the direction on the sphere orthogonal to all rays $\boldsymbol{p}$ through the image line. Since by construction $\boldsymbol{l}$ is orthogonal to $\mathcal{P}$, it must also be orthogonal to the 3-D line; that is, $\boldsymbol{l} \cdot \boldsymbol{v} = 0$.

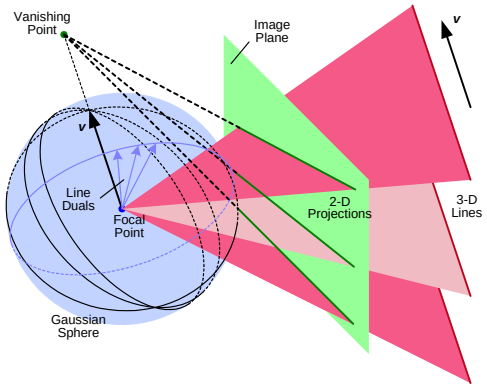


Figure 7: Geometry of Vanishing Points

Projections of parallel 3-D lines converge to an apparent vanishing point in the plane. Projectively, the vanishing point represents the intersection of a pencil formed by observations of the lines.

Similarly, *any* 3-D line parallel to $\boldsymbol{v}$ has a projective dual representation $\boldsymbol{l}_i$ for which $\boldsymbol{l}_i \cdot \boldsymbol{v} = 0$. The direction $\boldsymbol{v}$ is thus the normal to a plane containing all such dual rays $\boldsymbol{l}_i$ (Fig. 7). Because of the projective equivalence between scene lines and image lines, 2-D observations alone suffice for this construction; thus, $\boldsymbol{v}$ can be recovered from a set of image lines if their associated 3-D lines are known to be parallel to $\boldsymbol{v}$. For example, an error expression such as $\sum_{i=1}^{k} \boldsymbol{l}_i \cdot \boldsymbol{v}$ can be minimized, or a set of noisy line measurements can be fused to form a projective distribution on $\boldsymbol{v}$.

The intersection of two or more projective lines can be represented as a point on the sphere to which the lines' dual representations are orthogonal. This is the relationship stated by $\boldsymbol{l}_i \cdot \boldsymbol{v} = 0$, so the projection of the recovered 3-D line direction $\boldsymbol{v}$ onto the original image is precisely the apparent intersection of all the 2-D lines. That is, the intersection and the 3-D line direction represent the same entity, the vanishing point. Since vanishing points lie at infinity, they are invariant to local camera translations.

3.2 Detecting VPs in One Node

Image lines, represented by projective random variables $\boldsymbol{x}_i$, serve as the primary features for vanishing point recovery. However, the collection of lines in a given image is initially unclassified; that is, lines are not grouped into parallel sets, and random outliers (arising from organic objects and visual clutter such as foliage, cars, and people) are mixed with the true data. The problem of vanishing point estimation thus has three components. First, the number of groups J (that is, the number of prominent 3-D line directions) must be established. Next, lines $\boldsymbol{x}_i$ must be classified according to their corresponding 3-D direction or discarded as outliers. Finally, the vanishing point $\boldsymbol{v}_j$ for each group must be estimated.

These three problems are tightly coupled, in that given a deterministic classification of all line features, the estimation problem reduces to a collection of J isolated, straightforward projective inference tasks, one for each line group; and similarly, given knowledge of all 3-D directions $\boldsymbol{v}_j$, line classification is reduced to a similarity metric between lines and directions.

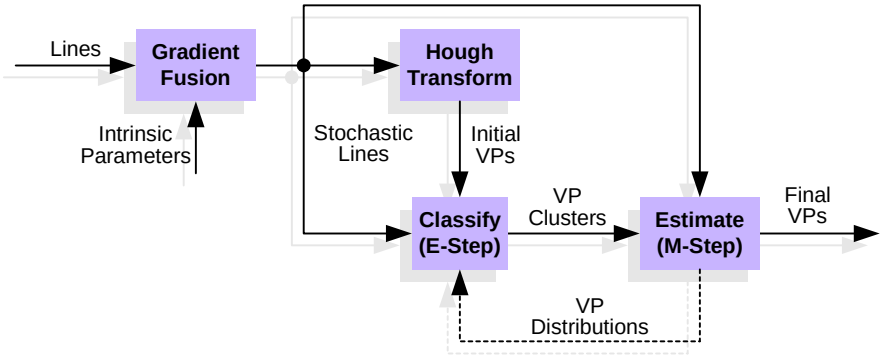


Figure 8: Vanishing Point Estimation

Stochastic line features from a single image are obtained from the fusion of gradient pixel distributions. These lines are used in a Hough transform, which finds prominent 3-D line directions to initialize an expectation maximization algorithm. The result is a set of accurate vanishing points and a line classification.

The expectation maximization (EM) algorithm [16] is a powerful tool for parameter estimation from incomplete or unclassified data. This section presents an EM formulation for simultaneous line classification and vanishing point estimation as inference problems on the sphere. For the moment, it is assumed that the algorithm is appropriately initialized; that is, the number of prominent 3-D directions J is known, and their approximate directions are available. Section 3.2.4 describes an efficient Hough transform technique which determines these quantities. The overall system is summarized in Fig. 8.

3.2.1 Mixture Model for Vanishing Points

Fig. 7 shows that vanishing points are projective quantities constrained by the dual points of contributing projective lines. Thus each of the J observed vanishing points $\boldsymbol{v}_j$ is treated as a Bingham random variable with unknown parameter matrix $\boldsymbol{M}_j^{\boldsymbol{v}}$, formed by fusion of the appropriate uncertain line features. The entire data set $\mathcal{X}$ is a collection of unclassified samples from the set of random variables $\mathcal{V} = \{\boldsymbol{v}_0 \ldots, \boldsymbol{v}_J\}$, where $\boldsymbol{v}_0$ represents an unknown

outlier distribution; thus, $\mathcal{X}$ is described by a mixture of $J+1$ Bingham densities $p(\boldsymbol{x}_i|j,\mathcal{V})$, so that

$$p(\boldsymbol{x}_i|\mathcal{V}) = \sum_{j=0}^{J} p(\boldsymbol{x}_i|j,\mathcal{V})p(j|\mathcal{V}), \tag{9}$$

where $p(j|\mathcal{V})$ is a prior probability representing the fraction of data generated by $\boldsymbol{v}_j$. Each observation $\boldsymbol{x}_i$ represents an uncertain line feature with known equatorial Bingham distribution $\mathcal{B}_3(\boldsymbol{x}_i;\boldsymbol{M}_i)$. The parameter matrices $\boldsymbol{M}_i$ are maximum likelihood estimates obtained from constituent gradient pixels when the line feature is detected.

With the form of the underlying distributions specified, the E-step, in which likelihoods are calculated, and the M-step, in which parameters of interest are estimated by maximum likelihood, can proceed in alternation. These steps are described in the next two sections.

3.2.2 The E-Step

In the E-step of the EM algorithm, a set of posterior probabilities α_{ij} is computed which effectively "weigh" each observation $\boldsymbol{x}_i$ during the estimation of the parameters $\boldsymbol{M}_j^{\boldsymbol{v}}$ for distribution j in the subsequent M-step. The weights are given by

$$\alpha_{ij} = p(j|\boldsymbol{x}_i,\tilde{\mathcal{V}}) = \frac{p(\boldsymbol{x}_i|j,\tilde{\mathcal{V}})p(j|\mathcal{V})}{\sum_{m=1}^{J} p(\boldsymbol{x}_i|m,\tilde{\mathcal{V}})p(m|\mathcal{V})}, \tag{10}$$

where $\tilde{\mathcal{V}}$ represents the vanishing point distributions as computed from the previous M-step. Assuming the prior probabilities $p(j|\mathcal{V})$ and current parameter estimates $\boldsymbol{M}_j^{\boldsymbol{v}}$ are known (either from the previous step or from initialization), all that remains is to calculate the mixture component probabilities $p(\boldsymbol{x}_i|j,\mathcal{V})$.

Intuitively, $p(\boldsymbol{x}_i|j,\mathcal{V})$ represents the likelihood of the line $\boldsymbol{x}_i$ given that it belongs to vanishing point $\boldsymbol{v}_j$. If the line observation were deterministic, this likelihood would simply be given by $\mathcal{B}_3(\boldsymbol{x}_i;\boldsymbol{M}_j^{\boldsymbol{v}})$. However, $\boldsymbol{x}_i$ is a stochastic measurement which is itself represented by a probability distribution.

Bayesian arguments can therefore be used to determine the likelihood. Let $\boldsymbol{x}_i^0$ represent a particular measurement from the distribution of random variable $\boldsymbol{x}_i$; then

$$p(\boldsymbol{x}_i|j,\mathcal{V},\boldsymbol{x}_i^0) = \frac{1}{c(\boldsymbol{M}_j^{\boldsymbol{v}})}\exp((\boldsymbol{x}_i^0)^\top \boldsymbol{M}_j^{\boldsymbol{v}}(\boldsymbol{x}_i^0)). \tag{11}$$

To eliminate the dependence on the particular value of the random variable, the joint likeli-

hood is integrated over all possible measurement values:

$$\begin{aligned} p(\boldsymbol{x}_i|j,\mathcal{V}) &= \int p(\boldsymbol{x}_i|j,\mathcal{V},\boldsymbol{x}_i^0)p(\boldsymbol{x}_i^0)\mathrm{d}\boldsymbol{x}_i^0 \\ &= \int \frac{1}{c(\boldsymbol{M}_j^{\boldsymbol{v}})}\exp\left[(\boldsymbol{x}_i^0)^\top \boldsymbol{M}_j^{\boldsymbol{v}}(\boldsymbol{x}_i^0)\right]\frac{1}{c(\boldsymbol{M}_i)}\exp\left[(\boldsymbol{x}_i^0)^\top \boldsymbol{M}_i(\boldsymbol{x}_i^0)\right]\mathrm{d}\boldsymbol{x}_i^0 \\ &= \frac{1}{c(\boldsymbol{M}_j^{\boldsymbol{v}})c(\boldsymbol{M}_i)}\int \exp[(\boldsymbol{x}_i^0)^\top(\boldsymbol{M}_j^{\boldsymbol{v}}+\boldsymbol{M}_i)(\boldsymbol{x}_i^0)]\mathrm{d}\boldsymbol{x}_i^0 \\ &= \frac{c(\boldsymbol{M}_j^{\boldsymbol{v}}+\boldsymbol{M}_i)}{c(\boldsymbol{M}_j^{\boldsymbol{v}})c(\boldsymbol{M}_i)}. \end{aligned} \tag{12}$$

Thus $p(\boldsymbol{x}_i|j,\mathcal{V})$ can be calculated as a ratio of normalizing coefficients from three different Bingham densities.

3.2.3 The M-Step

Once the weights are known, the Bingham parameter matrices $\boldsymbol{M}_j^{\boldsymbol{v}}$ of each vanishing point distribution can be estimated by maximizing the log likelihood function

$$\sum_{i=1}^{k}\sum_{j=1}^{J}\alpha_{ij}\log[p(\boldsymbol{x}_i|j,\mathcal{V})p(j|\mathcal{V})] + \log p(\mathcal{V}) \tag{13}$$

where $p(\mathcal{V})$ is a prior distribution on the vanishing points. The exponential form of the Bingham distribution facilitates calculation of the log likelihood. Every parameter matrix $\boldsymbol{M}_j^{\boldsymbol{v}}$ is computed independently by fusing all k observations $\boldsymbol{x}_i$, each weighted by the α_{ij} from the E-step. Using Eqn. (13) and pooling all data information matrices $\boldsymbol{M}_i$ yields

$$\boldsymbol{M}_j^{\boldsymbol{v}} = \sum_{i=1}^{k}\alpha_{ij}\boldsymbol{M}_i + \boldsymbol{M}_j^{\boldsymbol{0}} \tag{14}$$

where $\boldsymbol{M}_j^{\boldsymbol{0}}$ represents the parameters of a prior distribution on $\boldsymbol{v}_j$ (see Section 3.2.4).

3.2.4 EM Initialization Using HT

Proper formulation and implementation of the EM algorithm described above relies on knowledge of the number of vanishing points J. In addition, guaranteed convergence to the correct solution (i.e. avoidance of local optima) requires the availability of reasonably accurate parameter estimates. Both required quantities can be obtained using a Hough transform [22, 4, 25]. We circumvent the practical difficulties of accuracy and parameterization involved in implementing the HT by using it only to initialize the EM algorithm and to generate a strong prior on the vanishing point estimates.

The HT parameter space is $\mathbb{S}^2$ (i.e. the space of all 3-D line directions), and constraints take the form

$$f(\boldsymbol{x}_i, \boldsymbol{v}_j) = \boldsymbol{x}_i \cdot \boldsymbol{v}_j = 0 \tag{15}$$

where here the $\boldsymbol{x}_i$ are the polar (dual) directions of the input lines. The whole of $\mathbb{S}^2$ is discretized using a cubic parameterization, as in [43]. Geometrically, each constraint $f(\cdot)$ represents a plane through the focal point with normal $\boldsymbol{x}_i$; intersection of this plane with three faces of the unit cube results in a set of at most three lines, which are easily discretized using standard line clipping and drawing algorithms.

When all data has been accumulated, peaks in the accumulation space, which represent likely vanishing points directions, are identified. The number of statistically significant peaks is used as the number of mixture components J, and the peak directions (i.e. the vectors from the center of the cube through each peak) are used to initialize the EM algorithm.

Peak directions also serve as prior densities $p(\boldsymbol{v}_j)$, each of which is formulated as a bipolar Bingham density ($\kappa_1 \leq \kappa_2 \ll 0$) whose modal axis is aligned with the peak direction. The parameter matrix $\boldsymbol{M}_j^0$ for the prior density can be determined by forming a scatter matrix from accumulation values in a region around the peak.

3.3 Registering Adjacent Node Pairs

The relative rotation bringing a given pair of cameras into registration can be determined by aligning two or more distinct vanishing points viewed by both cameras. Thus, once vanishing points have been estimated for each node, relative orientations can be found for each node adjacency.

This section first reviews the classical, deterministic formulation for two-camera registration (Section 3.3.1), then introduces two novel extensions. Section 3.3.2 describes a model for uncertainty in the resulting rotations as the fusion of deterministic samples from a Bingham distribution on $\mathbb{S}^3$, and proceeds by considering how uncertainty in the vanishing points themselves affects the distribution of the resulting rotation.

The estimation methods assume that correspondence between vanishing points in different cameras is known, which is generally not the case. Sections 3.3.3 and 3.3.4 address the correspondence problem for the two-camera case, and ambiguities that arise in practice.

3.3.1 Deterministic Pair Registration

Consider two cameras $\mathcal{A}$ and $\mathcal{B}$, each of which views a common set of J vanishing points. Let $\boldsymbol{v}_j^{\mathcal{A}}$ and $\boldsymbol{v}_j^{\mathcal{B}}$ denote the directions of a particular line direction $\boldsymbol{d}_j$ as seen by each camera, and further assume that camera $\mathcal{B}$ is free to rotate while camera $\mathcal{A}$ is held fixed. We wish to estimate a single rotation in the form of a quaternion $\boldsymbol{q}$ which, when applied to camera $\mathcal{B}$ and its vanishing points, produces the best alignment between the $\boldsymbol{v}_j^{\mathcal{A}}$ and the $\boldsymbol{v}_j^{\mathcal{B}}$. Note that a single common VP does not uniquely determine the rotation relating the cameras; two distinct VPs are needed for uniqueness (i.e. we require that $J \geq 2$).

This problem has been studied in various contexts. The most relevant derives the optimal $\boldsymbol{q}$ as part of a more general 3-D to 3-D correspondence problem [21]. The objective is to determine

$$\text{argmin}_{\boldsymbol{q}} \sum_{j=1}^{J} \|\boldsymbol{v}_j^{\mathcal{A}} - \boldsymbol{R}(\boldsymbol{q})\boldsymbol{v}_j^{\mathcal{B}}\|^2 \tag{16}$$

$$= \text{argmin}_{\boldsymbol{q}} \left[\boldsymbol{q}^\top \sum_{j=1}^{J} \boldsymbol{A}_j^\top \boldsymbol{A}_j \boldsymbol{q}\right] \tag{17}$$

$$= \text{argmin}_{\boldsymbol{q}} \boldsymbol{q}^\top \boldsymbol{A} \boldsymbol{q} \tag{18}$$

i.e. the $\boldsymbol{q}$ that minimizes a quadratic error function. Each 4×4 matrix $\boldsymbol{A}_j$ is constructed as a linear function of its constituent vanishing points $\boldsymbol{v}_j^{\mathcal{A}}$ and $\boldsymbol{v}_j^{\mathcal{B}}$. The solution to Eqn. (18) is the eigenvector corresponding to the minimum eigenvalue of the symmetric 4×4 matrix $\boldsymbol{A}$.

This method produces optimal results using the error metric in Eqn. (16) but, aside from the scalar error residual, produces no notion of uncertainty in the result $\boldsymbol{q}$; nor does it treat uncertainty in the measurements themselves. The next section addresses these issues in more detail, proposing methods for obtaining descriptive error measures on the estimated rotations by incorporation of uncertainty in the underlying data.

3.3.2 Stochastic Pair Registration

Recall from Section 2.3 that rotational uncertainty can be described as a Bingham distribution on $\mathbb{S}^3$ characterized by a 4×4 matrix of parameters $\boldsymbol{M}_{\boldsymbol{q}}$. The matrix $\boldsymbol{A}$ obtained in Eqn. (18), when properly normalized, is analogous to a sample second moment matrix: it is symmetric and positive semidefinite, and its eigenvalues sum to unity. $\boldsymbol{A}$ is composed of a sum of J matrices $\boldsymbol{A}_j^\top \boldsymbol{A}_j$ that also possess these properties. Each of these constituent matrices can be viewed as the squared contribution of a "sample" $\boldsymbol{q}_j$ formed from an individual vanishing point correspondence. Thus, a parameter matrix $\boldsymbol{M}_{\boldsymbol{q}}$ for the distribution on the resulting quaternion $\boldsymbol{q}$ can be obtained directly from $\boldsymbol{A}$ using the ML method mentioned in Section 2.3.

This method can be used only for data sets that give each measurement equal weight. Extension to scalar-weighted data is straightforward, involving a weighted sum of constituent sample matrices normalized by the total weight. In the general case, however, where vanishing points are described by Bingham-distributed uncertainty, the Bingham distribution on $\mathbb{S}^3$ induced by each correspondence must be computed.

Every matrix $\boldsymbol{A}_j$ is a function of the vanishing point directions in its constituent correspondence. Thus, the parameters of the Bingham distribution associated with $\boldsymbol{A}_j$ can also be expressed as a function of these directions. Given particular sample values of vanishing point distributions $\boldsymbol{v}_j^{\mathcal{A}}$ and $\boldsymbol{v}_j^{\mathcal{B}}$, define $\boldsymbol{M}(\boldsymbol{v}_j^{\mathcal{A}}, \boldsymbol{v}_j^{\mathcal{B}})$ as the parameter matrix of the associated distribution. The contribution of correspondence j can then be obtained by Bayesian

integration over all possible sample values of the two constituent vanishing points:

$$\begin{aligned} p(\boldsymbol{q}_j) &= \int_{\mathbb{S}^2}\int_{\mathbb{S}^2} p(\boldsymbol{q}_j|\boldsymbol{v}_j^{\mathcal{A}},\boldsymbol{v}_j^{\mathcal{B}})p(\boldsymbol{v}_j^{\mathcal{A}})p(\boldsymbol{v}_j^{\mathcal{B}})\mathrm{d}\boldsymbol{v}_j^{\mathcal{A}}\mathrm{d}\boldsymbol{v}_j^{\mathcal{B}} \\ &= \int_{\mathbb{S}^2}\int_{\mathbb{S}^2} \mathcal{B}_4(\boldsymbol{q}_j;\boldsymbol{M}(\boldsymbol{v}_j^{\mathcal{A}},\boldsymbol{v}_j^{\mathcal{B}}))\mathcal{B}_3(\boldsymbol{v}_j^{\mathcal{A}};\boldsymbol{M}^{\mathcal{A}})\mathcal{B}_3(\boldsymbol{v}_j^{\mathcal{B}};\boldsymbol{M}^{\mathcal{B}})\mathrm{d}\boldsymbol{v}_j^{\mathcal{A}}\mathrm{d}\boldsymbol{v}_j^{\mathcal{B}}. \end{aligned} \tag{19}$$

This quantity can be approximated by a Bingham distribution on $\mathbb{S}^3$ with parameter matrix $\boldsymbol{M}_j$. Once distributions have been determined for each correspondence $\boldsymbol{q}_j$, the final aggregate distribution is described simply by the parameters

$$\boldsymbol{M}_{\boldsymbol{q}} = \sum_{j=1}^{J} \boldsymbol{M}_j. \tag{20}$$

3.3.3 Matching VPs Across a Node Pair

The registration methods above assume that one-to-one correspondence has been established between vanishing points detected in a given pair of images. Determination of correspondence is generally a difficult task without additional information; however, if the two relevant cameras view a significant portion of common scene geometry, then the assumption of approximately known initial pose is typically sufficient to establish consistent correspondence. This section presents a few heuristic methods to determine local (i.e. two-camera) correspondence used to initialize the global technique described in Section 3.4.3.

If two cameras view overlapping scene geometry, then the sets of vanishing points detected in each camera are likely to contain common members. In this case it is assumed that cameras $\mathcal{A}$ and $\mathcal{B}$ have in common a set of vanishing points related by a single rotation $\boldsymbol{q}$ which preserves the relative (intra-camera) angles between them.

Since a minimum of two correspondences is needed to find a unique rotation relating the two cameras, relative angles between pairs of vanishing points in each camera can be used as a matching criterion. For example, if the angle between $\boldsymbol{v}_1^{\mathcal{A}}$ and $\boldsymbol{v}_2^{\mathcal{A}}$ differs significantly from that between $\boldsymbol{v}_1^{\mathcal{B}}$ and $\boldsymbol{v}_2^{\mathcal{B}}$, then these two pairs (which constitute a *pair couplet*) cannot possibly match. Thus, only those pair couplets are considered whose relative angles are within a small threshold of each other. Angular thresholds are related to the Bingham parameters of the respective vanishing point distributions; highly concentrated distributions thus have tighter thresholds than do distributions with more spread. It should also be noted that, since vanishing points are axial, there are two supplementary angles to consider; the minimum of the two is always chosen for angle comparison (Fig. 9a).

A set of scores $\mathcal{S} = \{s_1, \ldots, s_k\}$ is computed, one for each pair couplet meeting the relative angle criterion above as follows. First, the pair from camera $\mathcal{B}$ is rotated to the pair from camera $\mathcal{A}$ by $\boldsymbol{q}$ using the deterministic pair registration technique from Section 3.3.1; the direction of a given vanishing point is taken as the major axis of its associated Bingham distribution. The angle of rotation θ_i required to align the two pairs is noted, and the remaining vanishing points from camera $\mathcal{B}$ are then rotated by $\boldsymbol{q}$ and compared with

each vanishing point from camera $\mathcal{A}$. The total number N_i of vanishing points that align to counterparts in camera $\mathcal{A}$ within a threshold angle, including the original pair, is also noted.

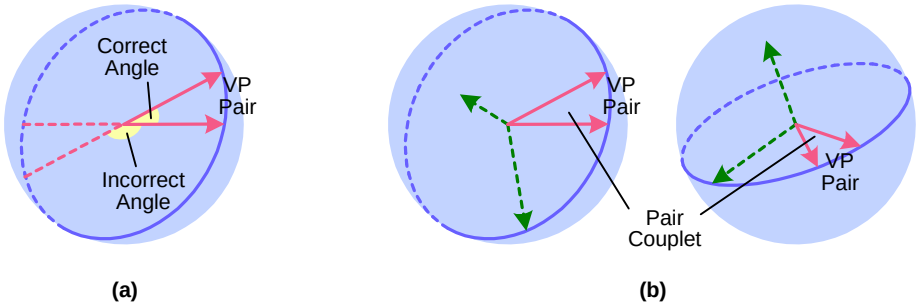


Figure 9: Pair Couplets

(a) There are two possible choices when comparing relative angles in axial quantities. By convention, the smaller of the two is chosen. (b) A matching pair couplet is depicted. Relative angles between the pairs are identical despite the fact that the cameras are not rotationally aligned.

Each score is then computed as $s_i = N_i/\theta_i$. This score emphasizes correspondence sets containing many matches, while preserving the assumption that the relative rotations are already known to reasonable accuracy. The correspondence set with the highest score is chosen as the "correct" set, for later use in global rotational alignment (Section 3.4.3).

Let J represent the number of vanishing points viewed by each camera. Then enumeration of all possible vanishing point pairs per camera is $\mathcal{O}(J^2)$, and enumeration of all possible pair couplets is $\mathcal{O}(J^4)$. Computation of correspondence sets for each couplet is $\mathcal{O}(J^2)$, raising the overall work required to $\mathcal{O}(J^6)$. The order of this rather brute-force technique is high, but in practice the value of J is small—typically less than 6.

3.3.4 Correspondence Ambiguities

Rotation as presented in Section 3.3 requires correspondence between signed directions, but vanishing points are axial (i.e. undirected) quantities. In truth, for each pair couplet meeting the relative angle criterion, *two* different rotations (and associated scores) must be computed, one for each combination of sign that maintains relative angle consistency (Fig. 10). Evaluation otherwise proceeds as described above.

Other ambiguities can also arise that are not so easily resolved, especially in urban scenes consisting of mutually orthogonal lines (Fig. 11). Since relative angles between multiple pairs of vanishing points can be identical within a single image, there may exist several plausible match configurations. The matching algorithm must thus rely on prior knowledge, such as the fact that since approximate pose is known, the rotational discrepancy between any two cameras should be relatively small; solutions implying large rotation are unlikely (hence the score criteria in Section 3.3.3). Another assumption is that nearly all urban scenes contain vertical lines, so the vanishing point directions closest to "up" in each image can be assumed to match, thus constraining correspondence sets and reducing computation to $\mathcal{O}(J^4)$.

If there is significant error in initial rotational pose and if correspondence ambiguities

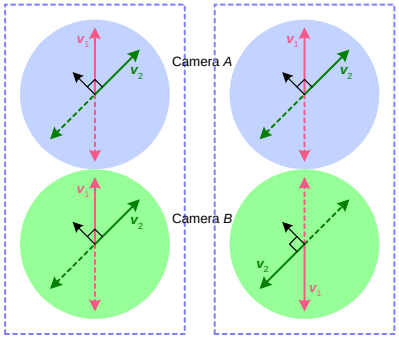


Figure 10: Two Solutions
There are two optimal rotations, differing by 180°, that align axial (i.e., antipodal) features.

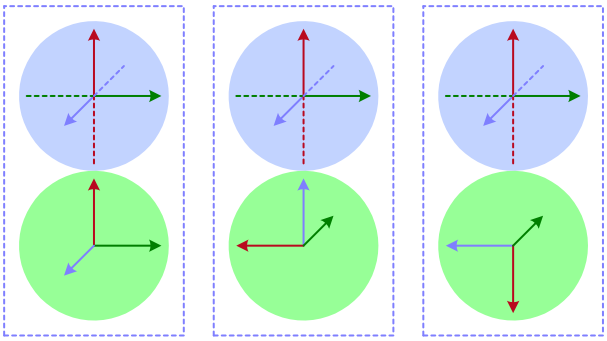


Figure 11: Vanishing Point Correspondence Ambiguity
An example of ambiguities that can arise in symmetric configurations. Without additional information, there is no way to differentiate between match candidates.

exist, the matching algorithm can fail, finding a plausible but incorrect match assignment. In our data, initial camera pose is sufficiently accurate to avoid this problem.

3.4 Registering Clusters of Multiple Nodes

The above treatment of rotational registration is deficient in two main respects. First, it determines explicit or "hard" correspondence among vanishing points rather than stochastic correspondence; and second, it considers only two cameras at a time. This section presents a multi-camera extension for rotational registration which addresses the above concerns and produces a globally optimal set of camera orientations along with their associated uncertainty.

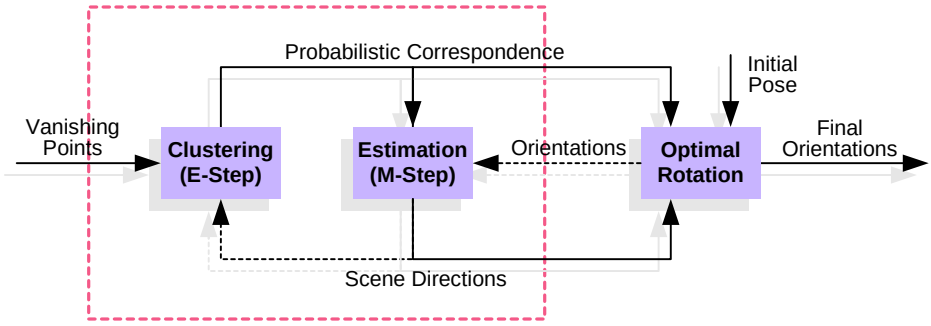


Figure 12: Global Orientation Recovery
Vanishing points are aligned with one another to determine camera orientations. A two-level feedback hierarchy is used: the top level estimates rotations and scene-relative line directions in alternation; the bottom level (outlined) classifies VPs, estimates line directions, and provides feedback.

It may seem at first that the two-camera method can be directly extended to handle multiple cameras in a sequential fashion. This type of sequential approach is often used when data consists of image streams, such as video [18]; images are locally registered a pair

or triple at a time as the stream progresses. Such purely local methods invariably propagate and accumulate error, however, since pose estimates from early images feed registration of subsequent images. To minimize this effect and distribute error equally among all cameras, pose must be recovered by simultaneously considering all available data.

As is typical in 3-D vision problems, pose recovery consists of the two coupled subproblems of correspondence and registration. In this context, given a grouping of vanishing points into sets, where each set represents observations of a true scene-relative line direction, estimation of relative rotations becomes simpler; and, conversely, given a set of accurate camera orientations, determination of correspondence is simplified. These facts, at a high level, suggest an iterative bundle-adjustment scheme that alternately estimates orientations given correspondence, then establishes correspondence given orientations.

The basic idea is shown in Fig. 12. Rotations and correspondence are initially produced by exhaustive search (Section 3.4.3). Global (scene-relative) line directions are estimated based on vanishing point clusters; each camera is then rotated until its vanishing points optimally align with these global directions, and the process repeats. There are two levels of feedback in the process, one at the high level of rotational bundle adjustment and the other in the estimation of global line directions, which alternates between determination of probabilistic correspondence and estimation of directional distributions.

3.4.1 EM for Multi-Camera Registration

This alternation between classification and estimation suggests application of an EM algorithm, which would also circumvent the need for explicit correspondence and provide an adequate probabilistic estimation framework. At its core, the problem is to determine the probability distributions of a set of rotations in the form of quaternions, $\mathcal{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_M\}$, based solely on a data set $\mathcal{V} = \{\mathcal{V}^1, \ldots, \mathcal{V}^M\}$, where $\mathcal{V}^i$ is the set of vanishing points $\boldsymbol{v}_j^i$ detected in image i. Probabilistically, this can be written as

$$\text{argmax}_{\mathcal{Q}}[p(\mathcal{Q}|\mathcal{V})]. \tag{21}$$

However, the rotations depend on scene-relative line directions $\mathcal{D}$, as well as correspondence $\mathcal{C}$ between these directions and the vanishing points in each individual image. Using Bayes' rule, the likelihood to be maximized can thus be rewritten as

$$\begin{aligned} p(\mathcal{Q}|\mathcal{V}) &= \int_{\mathcal{D}} p(\mathcal{Q}|\mathcal{D},\mathcal{V}) p(\mathcal{D}|\mathcal{V}) \mathrm{d}\mathcal{D} \\ &= \int_{\mathcal{D}} \sum_{\mathcal{C}} p(\mathcal{Q}|\mathcal{C},\mathcal{D},\mathcal{V}) p(\mathcal{C}|\mathcal{D},\mathcal{V}) p(\mathcal{D}|\mathcal{V}) \mathrm{d}\mathcal{D}. \end{aligned} \tag{22}$$

Note that a sum is taken over $\mathcal{C}$ rather than an integral, because the set of correspondence configurations is discrete. The quantity $p(\mathcal{D}|\mathcal{V})$ represents the prior distribution on global line directions given only the vanishing point data, and is taken to be uniform, since in the absence of rotational pose, nothing is known about this distribution. The quantity

$p(\mathcal{C}|\mathcal{D},\mathcal{V})$ is the prior distribution on correspondence given only global line directions and vanishing points (not rotations). This distribution can be approximated from the pair-wise correspondences established in Section 3.3.3.

The high-level EM algorithm alternates between two steps: first, compute the likelihoods $p(\mathcal{C},\mathcal{D}|\mathcal{Q},\mathcal{V})$; next, maximize the expression

$$\int_{\mathcal{D}} \sum_{\mathcal{C}} p(\mathcal{C},\mathcal{D}|\mathcal{Q},\mathcal{V}) \log p(\mathcal{Q}|\mathcal{C},\mathcal{D},\mathcal{V}) \mathrm{d}\mathcal{D} \tag{23}$$

The likelihoods computed in the E-step thus serve as weights on the conditional log-likelihood to be maximized in the M-step. Conditioned on line directions and correspondence, the quaternions are independent of one another because vanishing points in each camera can be rotated in isolation to optimally align with the global line directions. Thus,

$$\begin{aligned} \log p(\mathcal{Q}|\mathcal{C},\mathcal{D},\mathcal{V}) &= \log \prod_{i=1}^{M} p(\boldsymbol{q}_i|\mathcal{C},\mathcal{D},\mathcal{V}) \\ &= \sum_{i=1}^{M} \log p(\boldsymbol{q}_i|\mathcal{C},\mathcal{D},\mathcal{V}) \end{aligned} \tag{24}$$

and each quaternion can be estimated independently. Maximization proceeds as described in Section 3.3.2, with the Bingham distribution of orientation $\boldsymbol{q}_i$ specified by the parameter matrix $\boldsymbol{M}_i^{\boldsymbol{q}}$, which represents the weighted sum of correspondence matrices of the form in Eqn. (20).

3.4.2 EM for Multi-Camera Correspondence

The above formulation solves the M-step of the bundle adjustment, but the E-step still remains—the likelihoods $p(\mathcal{C},\mathcal{D}|\mathcal{Q},\mathcal{V})$ must be computed. Intuitively, these likelihoods represent distributions on correspondence $\mathcal{C}$ and scene-relative line directions $\mathcal{D}$ given the current set of orientation estimates $\mathcal{Q}$. $\mathcal{C}$ and $\mathcal{D}$ are coupled, however; knowledge of the line directions influences the groupings, and vice versa.

Let $\tilde{\boldsymbol{v}}_j^i$ represent vanishing point j in image i after rotation by $\boldsymbol{q}_i$; the set of all such directions serves as the pool of data to be grouped. Further, let $\boldsymbol{d}_k$ represent a particular scene-relative 3-D line direction. The problem then becomes to simultaneously estimate the $\boldsymbol{d}_k$ and classify the $\tilde{\boldsymbol{v}}_j^i$.

This formulation is identical to the vanishing point estimation problem posed in Section 3.2. The collective data set $\tilde{\mathcal{V}}$ is drawn from a weighted mixture of Bingham distributions of $\boldsymbol{d}_k$; the only difference is that the underlying samples are now bipolar rather than equatorial. Application of the lower-level EM algorithm results in a set of parameters describing the line direction distributions, as well as a probabilistic assignment of individual vanishing points to each global line direction. After convergence, these results are fed back into the M-step of Section 3.4.1.

3.4.3 Initializating Multi-Camera Registration

As mentioned previously, expectation maximization algorithms are effective only when properly initialized. The number of mixtures (in this case J, the number of 3-D line directions) must be known, and the algorithm must begin with a reasonable set of initial values (rotations and correspondence). If possible, prior distributions should also be supplied. This section outlines an algorithm that provides adequate initialization for the EM techniques described above.

Camera adjacency is supplied with the input, as a graph whose vertices represent image nodes, and whose edges indicate proximal camera pairs. The two-camera correspondence technique of Section 3.3.3 is then applied to each pair (i.e., to each graph edge), and unique vanishing point matches are extracted. A list of global line directions is constructed, each containing a set of references to its constituent vanishing points. The algorithm proceeds as follows:

```
Clear list of global line directions
For each camera pair in adjacency graph
    Apply two-camera VP correspondence
        For each VP pair matched
            If neither VP exists in any global line direction then
                Create new global line direction and add to list
                Link both constituent VPs to this new direction
            Else if one VP exists then
                Find its global line direction
                Link other VP to this direction
            Else if both VPs exist then
                If associated with different global line directions then
                    Merge two global line directions into one
```

This algorithm produces a list of vanishing point sets, each of which represents observations of a single scene-relative 3-D line direction. These sets are the components of the mixture model in Section 3.4.2, and correspondence weights can be initialized to binary values according to the grouping produced above. Any camera not having at least two vanishing point entries in the list of line directions is tagged as unalignable, since at least two correspondences are needed for unique rotational registration. In practice, only about 5% of the nodes in a given configuration observe insufficient vanishing points for alignment.

3.4.4 Merging and Separating Clusters

As the EM algorithm proceeds, separate vanishing point clusters that truly represent the same 3-D direction, or single clusters that represent multiple directions, may arise. The latter misclassifications can result from distinct 3-D lines having nearly identical directions

that are fused due to noisy observations; the former usually results from the graph traversal described in the previous section.

After each rotation step of the EM algorithm, all pairs of cluster distributions are compared, and if two sufficiently overlap (e.g. with 95% probability) they are merged into one, decrementing J by one. Similarly, clusters containing vanishing points significantly different from their respective modal directions are split, incrementing J by one.

3.5 Limitations

The algorithm has several limitations. It requires useable line features from a feature detector. The algorithm's assumption that nearby nodes are likely to have observed overlapping scene structure may be faulty, for example when two nodes are on opposite sides of a thin building. Very poor initial position estimates may cause the overlap assumption to fail, since nodes classified as "adjacent" may have in fact been acquired far apart. Due to curvature of the Earth's surface, the scope of orientation estimation is inherently limited to relatively small geographic areas (areas spanning more than about ten kilometers will have a local vertical that varies by more than 0.1°, our current implementation's error floor).

3.6 Asymptotic Performance

This section characterizes the running time of the above algorithms as a function of the number of input images and input features. All methods scale linearly with the number of cameras and the number of 2-D line features, and incorporate all available data to automatically and robustly produce globally optimal orientations. Uncertainty in image line features and vanishing points, and in the orientations themselves, are carefully modeled and estimated using projective inference methods.

4 Experiments

We implemented the registration algorithm in roughly 5,000 lines of C++ code. This section assesses the algorithm's end-to-end performance using several objective metrics, on both synthetic and real data. We use notions of self-consistency (e.g., from [26]) and a variety of application-specific consistency measures.

4.1 Synthetic Data

We evaluated the localization of single vanishing points (3.2) by generating sets of parallel, ideal 3-D lines, then projecting them onto the unit sphere. The projections were then perturbed at specified noise levels. We also generated a controlled percentage of random outlier lines, uniformly distributed on the sphere. Finally, we varied the number of distinct 3-D directions (denoted by J in the notation of 3).

The Hough transform method for EM initialization was applied to 50 data sets, each containing a mixture of 500 points and outliers. The percentage of true peaks detected, as well as the angular deviation of the peaks from the true 3-D line directions, were examined as a function of measurement noise, outlier percentage, and J (Fig. 13). Cells were sized so that the maximum angular coverage was 1°, and a window size of 5 cells (roughly 5°) was used for peak detection.

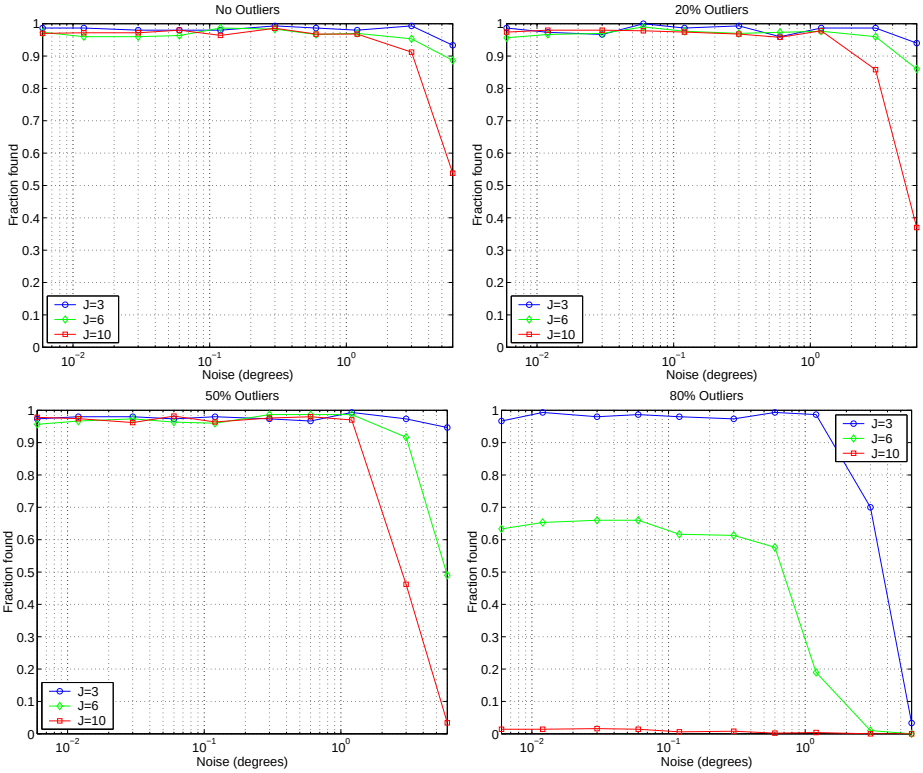


Figure 13: Percentage of Vanishing Points Detected

The percentage of true vanishing points detected as a function of point projection error (x axis), number of true line directions, ($J = 3, 6, 10$), and percentage of outlier features (zero, 20, 50, and 80 percent from upper left).

Successfully detected vanishing points were consistently within about 1° of the true directions. A small number of false peaks were identified (about 2%), but only when feature noise exceeded several degrees. Performance of the expectation maximization algorithm (as initialized by the Hough transform) was also assessed, varying the same parameters as above. Angular deviations of the estimates from the true directions are shown in Fig. 14.

Vanishing point estimation has proven to be quite robust. The Hough transform and peak detection methods provide sound initial estimates, with performance degradation occurring only at very high levels of feature noise; for moderate feature noise, all correct vanishing points are detected even with an outlier-to-data ratio of 4:1. Initialization and prior distributions provided by the Hough transform make the EM algorithm robust against outliers. Performance degrades as the number of contributing vanishing points increases, because fea-

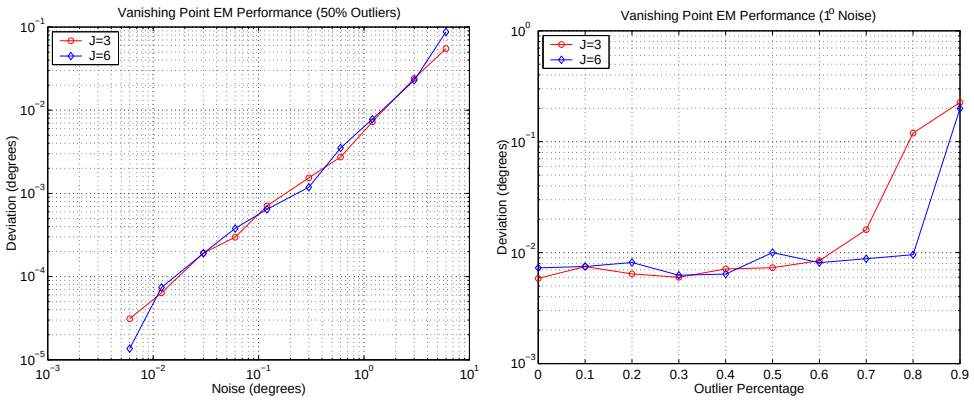


Figure 14: EM Vanishing Point Error

Average error in vanishing points estimated by the EM algorithm are plotted as a function of line feature noise with 50% outliers (left) and outlier percentage with 1° feature noise (right).

tures tend to crowd the closed projective space and vanishing point clusters "interfere" with one another. However, since there are only two to six prominent line directions in typical real-world data (e.g. urban scenes), interference effects are negligible in practice.

4.1.1 Two-Camera Rotational Pose

To assess the two-camera rotation method, a set of 4 randomly generated 3-D line directions was viewed by two cameras and perturbed by controllable noise. Outlier directions were also added to each camera, and the stochastic two-camera registration method described in 3.3 was applied to 50 such data sets.

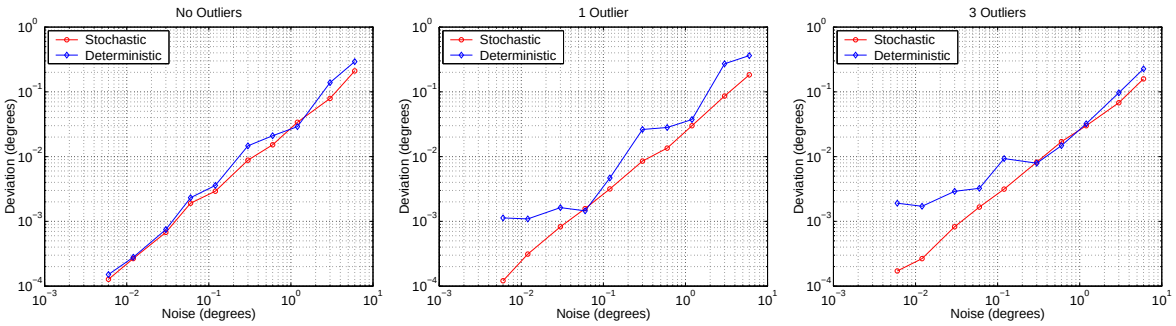


Figure 15: Comparison of Two-Camera Rotation Methods

The stochastic two-camera rotational registration technique is compared with the classical deterministic technique with 4 vanishing points. The plots show relative pose error as a function of vanishing point noise with 0, 1, and 2 outlier directions introduced. Behavior of the stochastic method exhibits more stability and consistency.

Incorporation of stochastic correspondence and vanishing point uncertainty improves rotational registration. Side-by-side comparison of deterministic and stochastic registration methods in Fig. 15 shows the novel method to be more stable and consistent than the deterministic rotation technique of 3.3.1, and to produce more accurate estimates.

4.1.2 Multi-Camera Orientation

End-to-end rotational pose recovery was examined by generating parallel 3-D lines and outliers as above, and projecting this geometry onto randomly situated cameras with controllable pose perturbations. As long as correspondence was unambiguous (3.3.4), correct orientations were recovered for arbitrary initial rotational error, even up to 180°. Fig. 16 shows error in orientation estimates.

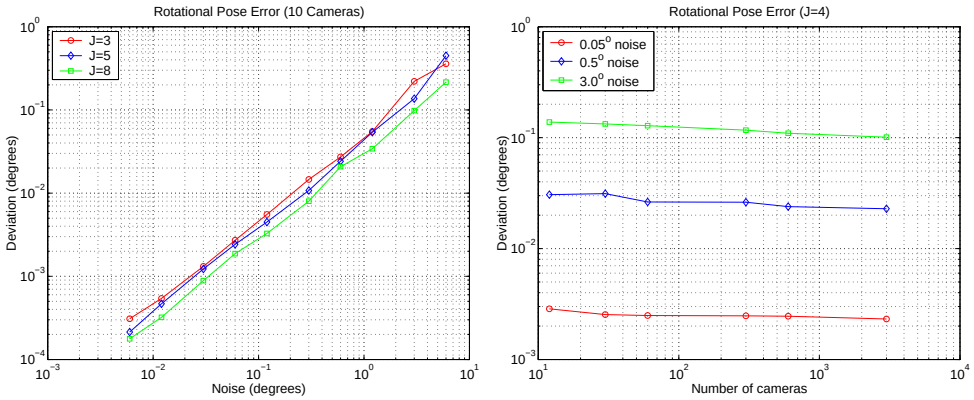


Figure 16: Multi-Camera Orientation Performance

Average orientation error (left) as a function of vanishing point noise for 10 cameras viewing varying numbers of 3-D line directions. Orientation error (right) plotted as a function of the number of cameras in the configuration with varying degrees of noise in 4 vanishing points.

As expected, accuracy of multi-camera rotational alignment increases (though slightly) with the number of vanishing points, since estimates generally improve with more observations. It is unclear why estimation error does not decrease more quickly as the number of cameras increases; one would expect that more observations of a single entity (namely a given 3-D line direction) should increase the certainty with which that entity is estimated, and consequently the accuracy of rotational estimates.

4.2 Real Data

We applied the registration algorithm to several outdoor data sets. Although it is difficult to quantify accuracy when no metric ground truth is available, we formulated several quantitative consistency metrics. For each data set, we report these quantities:

- ***Data Size and Extent.*** We report the dimensions of the acquisition area in meters, the average inter-node baseline (i.e., the average distance between a node and its neighbors), the total number of images ("Images"), the total and average number of line features ("Lines"), the number of omni-directional images ("Nodes"), the number of adjacent camera pairs ("Pairs"), and the number of detected VPs ("VPs").

- ***Computation Time.*** We report average and total running times for each stage of the algorithm, excluding file I/O, on a 250 MHz SGI Octane with 1.5 gigabytes of memory.

"VP Hough" reports time to initialize the HT data structure by scan-converting the dual of each line feature. "VP EM" reports time to detect VPs in each node. "Rotation EM" reports time to register VPs across pairs, and across the dataset.

- ***Rotational Offsets.*** Average and maximum deviations from the initial orientations are reported as "Rot Offset". These quantities characterize both the error in the initial pose, and the degree of robustness exhibited by our registration method.

- ***Consistency Measures.*** We report average and maximum probability density parameters of vanishing points and global orientations at 95% confidence bounds ("VP Bound" and "Rot Bound" respectively). When two VPS arise from scene VPS thought to be orthogonal, we report the discrepancy between the angle they form and 90 degrees as "VP Ortho Error."

4.2.1 Tech Square Data, and Manual Bundle-Adjustment

The "Tech Square" dataset consists of 81 nodes spanning an area of roughly 285 by 375 meters. The average inter-node baseline was 30.88 meters.

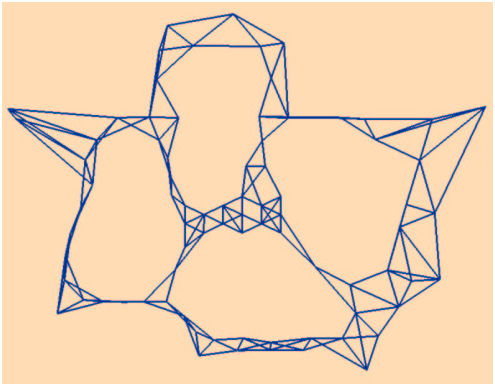


Figure 17: TechSquare Node Configuration
A top-down map view of the node configuration and adjacencies for the TechSquare data set. The average neighbor baseline was 30.88 meters.

Of the 81 nodes in the set, 75, or roughly 92%, were successfully registered; 6 of the nodes were not refined, due to insufficient vanishing point information. Our algorithm corrected initial orientation errors of over 17°. It recovered global orientation consistent on average to within 0.072°. The maximum rotation error reported was 0.098°, or roughly two pixels at our image resolution.

We also compared the orientations reported by the algorithm to those produced by a manual, 6-DOF bundle-adjustment process [13]. Interactive inspection of VPs in the manually registered dataset reveal that it does not represent ideal ground truth. Because the number of nodes and adjacencies were so large, the student operators of the bundle adjustment process naturally specified as few constraints as possible to allow convergence of the

Data Type	Per Image	Per Node	Total
Images	—	48	3899
Line Features	218	10,516	851,819
Nodes	—	—	81
Node Adjacencies	—	—	189
VPs	0.69	3.6	9

Stage	Per Node	Total
VP Hough	. 0.19 s	0 m 15 s
VP EM	. 6.68 s	7 m 54 s
Rotation EM	. —	0 m 46 s
Total	6.87 s	8 m 55 s

Measure	Avg	Max
Rot Offset	. 1.53°	. 17.18°
VP Bound	. 0.18°	. 0.80°
Rot Bound	. 0.072°	. 0.098°
VP Ortho Error	. 0.056°	. 0.09°

Table 1: Tech Square Data; Computation Times by Stage; Consistency Measures

underlying optimization. This produced unstable constraint sets, and a rather poor global pose assignment overall in the manual case.

4.2.2 Robustness Against Noisy Initial Pose

We studied the algorithm's robustness against noisy initial pose by applying it to the "Green Building" dataset, a set of thirty nodes with rotations and translation accurate only to roughly seven degrees and ten meters. The nodes spanned an area of roughly 80 by 115 meters, with an average inter-node baseline of 15.61 meters.

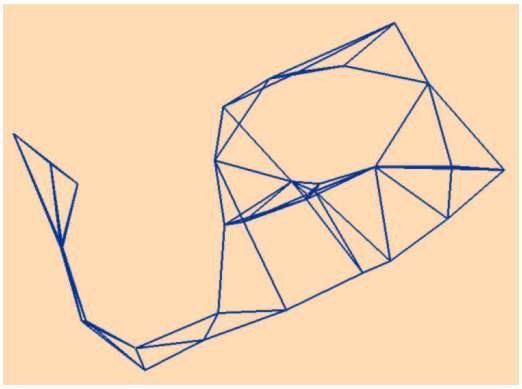


Figure 18: GreenBuilding Node Configuration

A top-down map view of the node configuration and adjacencies for the GreenBuilding data set. The average neighbor baseline was 15.61 meters.

With initial rotation error of 6.83°, our algorithm successfully registered all nodes, recovering orientations consistent on average to within 0.067°, and in the worst case to within 0.12°, again about two pixels.

4.2.3 An Extended Acquisition Region

The AmesCourt data set includes 100 nodes, and spans an area of 315 by 380 meters, with an average inter-node baseline of 23.53 meters. This data set represents a larger portion of geography and a larger number of camera sites. Our algorithm registered 95% successfully.

Data Type	Per Image	Per Node	Total
Images	—	23	695
Line Features	237	5,498	164,945
Nodes	—	—	30
Node Adjacencies	—	—	80
VPs	0.35	3.3	5

Stage	Per Node	Total
VP Hough	. 0.11 s	0 03 m
VP EM	. 2.93 s	1 28 m
Rotation EM	. —	0 18 m
Total	3.04 s	1 49 m

Measure	Avg	Max
Rot Offset	. 2.95°	. 6.83°
VP Bound	. 0.092°	. 0.52°
Rot Bound	. 0.067°	. 0.12°
VP Ortho Error	. 0.047°	. 0.11°

Table 2: Green Building Data; Computation Times by Stage; Consistency Measures

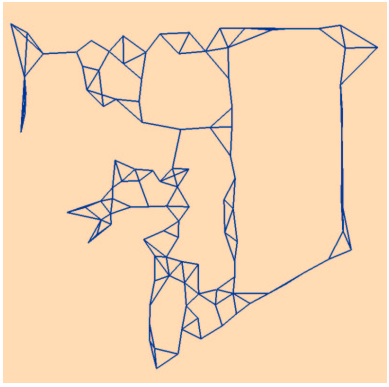


Figure 19: AmesCourt Node Configuration
A top-down map view of the node configuration and adjacencies for the AmesCourt data set. The average neighbor baseline was 23.53 meters.

Initial pose was corrected by 5.59°, achieving average consistency of 0.095°. Errors did not exceed 0.21°.

	Per Image	Per Node	Total
Images	—	20	2,000
Line Features	228	4,562	456,246
Nodes	—	—	100
Node Adjacencies	—	—	232
VPs	0.43	3.2	8

Stage	Per Node	Total
VP Hough	. 0.09 s	10 s
VP EM	. 2.55 s	4 m 22 s
Rotation EM	—	33 s
Total	2.64 s	5 m 05 s

Measure	Avg	Max
Rot Offset	. 2.83°	. 5.59°
VP Bound	. 0.23°	. 0.74°
Rot Bound	. 0.095°	. 0.21°
VP Ortho Error	. 0.043°	. 0.09°

Table 3: Ames Court Data; Computation Times by Stage; Consistency Measures

4.2.4 Estimation Accuracy vs. Field of View

We studied the effect of increasing field of view on the robustness and accuracy of VP localization, for a fixed sensor resolution. We found that both robustness and accuracy were strongly dependent on field of view. We varied the effective field of view by varying the number of constituent images used from the mosaic's underlying hemispherical tiling.

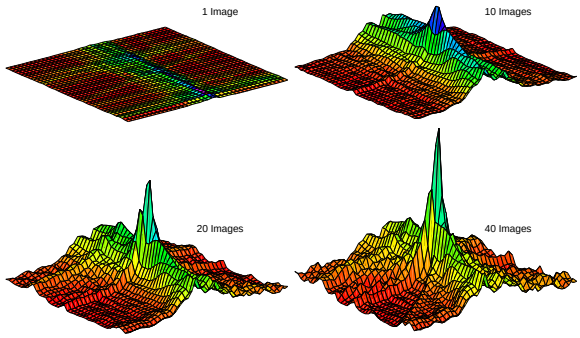


Figure 20: Hough Transform Peak Coherence

Illustrations of the dependence of Hough transform peak coherence on field of view for nodes containing 47 images. Peaks are shown for a vanishing point for increasing numbers of images in the hemispherical tiling.

We selected a particular vanishing point from a single node, and applied the Hough transform estimation process. The resulting HT values are plotted in Fig. 20. As expected, the more images included (i.e., the wider the FOV), the more accurate the VP estimation became, and the more VPs were visible.

5 Low-Level Error Sources

The algorithm described in this paper operates as one of a series of automated processing stages forming part of a large-scale system for outdoor model capture capability. Other system stages unavoidably contribute error. Some types of error (e.g., feature detection error) can be overcome in part by data fusion. Other types (e.g., in intrinsic camera calibration and image formation) seem less avoidable. Our sensor has a resolution of roughly 1 milliradian (mrad), or 0.05°, per pixel. The radial distortion correction, mosaic, and pinhole model calibration processes are each accurate to roughly one half or one pixel. Line features are detected to roughly half a pixel of accuracy. Therefore the overall error in projective feature localization can be estimated at about two pixels. Thus it appears that our global registration algorithm attains accuracy within a small constant factor (one to three) of the low-level system noise.

6 Related Work

There is a great deal of literature concerning the estimation of vanishing points and their use in recovering inter-camera rotations or more general egomotion. Here we limit our review to

methods which make explicit use of vanishing points, which recover only rotations, or which treat uncertainty of detected features or derived orientations. Our companion paper reviews previous work on the full 6-DOF camera registration problem.

6.1 Use of Vanishing Points

Some interactive systems use vanishing points for orientation recovery [40, 5, 15, 38]. In some cases the user can also specify constraints which must be met, for example that a particular pair of VPs must be orthogonal. In these systems, a human operator must perform the VP detection and classification tasks by indicating matching lines or higher-order features in multiple images. The drawback, of course, is that manual interaction becomes cumbersome and eventually infeasible as the problem size grows. Moreover, since humans make errors, and generally do as little work as they have to, the constraint sets produced by interactive systems are often unstable.

Researchers have proposed algorithmic (i.e., non-interactive) methods for matching VPs across multiple images. Shigang uses VPs for egomotion recovery [36], but the method assumes three mutually orthogonal line directions. Leung describes a graph algorithm for matching VPs across multiple images [27], but reports results only for image pairs, and does not treat uncertainty.

6.2 Discrete vs. Continuous Methods

The majority of VP estimation techniques rely exclusively on the Hough transform (e.g., [29, 37]). However, the accuracy of Hough transform techniques is inherently limited by discretization artifacts, and uncertainty in the estimates is difficult to characterize [12].

Other researchers have pursued continuous space approaches. For example, VPs can be localized by intersecting all pairs of lines in the image plane and searching for high-incidence regions [32, 28]. These approaches suffer from instability and degeneracies when lines are nearly parallel in the image plane. Formulating the intersection on the sphere addresses this problem [30]. However, the approach proposed in [30] expends quadratic time in the number of line features. Clustering nearby line intersections, for example by deterministic k-means algorithms, works well when the number of outliers is small, but is not robust to noise.

Collins [11] proposes an elegant use of the HT for reliable detection and clustering, followed by a more careful projective inference approach for accurate estimation of each VP. His overall clustering approach is deterministic, however, and uses a hard threshold to reject outliers, biasing the resulting estimates. Moreover, although his approach is formulated on the sphere, it is demonstrated only for narrow-FOV planar images.

6.3 Geometric Uncertainty

Since physical sensors are noisy, modeling uncertainty is crucial in real-world applications. Sophisticated noise-handling procedures exist for many least-squares [20] and computer vi-

sion problems [39, 1, 31]. Nearly all of these formulations assume additive Gaussian noise, arguably the most convenient error model. This model may not be appropriate for all representations; for example, when it is used for recovery of the covariance of the fundamental matrix [14] and [45], its physical meaning (i.e., its units) is unclear.

Uncertainty in image features is often treated in the Euclidean image plane, which is poorly suited to projective features. Some investigators have used axial probability distributions on the sphere to represent measurement error and perform inference [12]. Others have used a hybrid of projective and Euclidean distributions to model 6-DOF pose uncertainty [33].

Kanatani [24] and Antone [3] propose scalar weights that value VP correspondences by the certainty of their constituent VP pairs, but this weighting scheme allows only a limited description of randomness in the underlying observations. Chang [10] and Prentice [34] present errors-in-variables models for so-called *spherical regression*, but consider only extreme cases in which data points have tightly concentrated symmetric bipolar distributions.

There are several significant differences between our formulation and previous approaches. First, all vanishing points are estimated simultaneously as a mixture of distributions with soft classification rather than one at a time by hard classification. The mixture model reduces bias and other artifacts caused by somewhat arbitrary thresholding. Second, vanishing points are estimated as equatorial probability *distributions* rather than as deterministic bipolar vectors that maximize a likelihood function. Finally, the data fusion technique uses the full description of uncertainty in each measurement, as opposed to scalar weighting by line length or other heuristic criteria.

Regardless of method, no tracking or registration technique has been previously demonstrated that, given outdoor terrestrial imagery, can register a collection of cameras extending all the way around even a single building. Thus there is a significant gap between the sophisticated techniques reviewed above, and the fundamental "end-to-end" camera registration capability.

7 Conclusions

The algorithm described in this paper builds on a number of existing ideas, including: the use of gradient-based (line) features for robustness against lighting variations and strong perspective; decoupling the 6-DOF pose estimation problem into two pure 3-DOF problems; and probabilistic inference techniques on the sphere.

7.1 Contributions

This paper makes several contributions to the class of techniques for recovering the rigid orientation for large collections of cameras.

First, we propose the use of *a priori absolute pose estimates* and an *image adjacency graph* derived from sensor observations. The adjacency graph, supplied as input to our algorithm, enables it to search for common structure only among images known to be near each other in

the input. For a dataset with roughly constant density (number of images per unit volume), this implies that each image will have only a constant number of neighbors on average. Thus our algorithm expends total time which is linear in the number of input images, rather than quadratic as with previous methods. This asymptotic improvement is a significant advantage as the number of images grows. As an engineering advantage, the use of absolute initial pose estimates enables images to be registered to absolute (Earth) coordinates automatically, without the need for a human operator to indicate photogrammetric tie points or fiducial structures in the images.

Second, we show quantitative evidence that wide-FOV (omnidirectional) images are *fundamentally more powerful observations* than are narrow-FOV (planar) images for the recovery of inter-image and global orientations. Intuitively, the use of wide-FOV images ameliorates the aperture problem and reduces fundamental motion ambiguities. Experimental evidence shows that scene vanishing points can be detected more reliably, and estimated more robustly, with wide-FOV images than with conventional imagery. This is a clear engineering advantage in practice. We also observe that using a spherical image acquisition device decouples the choice of acquisition subject from the choice of reconstruction subject, by removing the need to orient the camera during acquisition.

Third, we extend classical deterministic methods for pair-wise and multi-camera alignment to a *stochastic framework* in which noisy input features, and derived camera orientations, are treated as *projective probability distributions*. This approach leads to robust performance in practice, both by allowing fusion of many noisy observations into a few accurate ensemble quantities (VPs), and by deferring determinative decisions until global information, namely the number of VPs, and an absolute orientation estimate and uncertainty for each, is available. Finally, this probabilistic model eschews the minimization of screen-space quantities in favor of angular error, a more natural measure inherently arising from the metric structure of the scene.

Fourth, we *use the Hough transform only for initialization* of an EM algorithm. This obviates a fundamental problem with classical HT methods, the choice of bucket size. In our approach, the HT is used only to determine the number and approximate location of peaks, so the discrete bucket can be chosen conservatively large. The ensuing continuous-space EM stage then converges to accurate estimates of the derived scene quantities.

Finally, the algorithm described in this paper uses *linear time and space resources* in the number of input images, rather than quadratically or worse as in many previous methods. This removes a fundamental barrier to the development of automated registration techniques for very large numbers of images. In practice, we demonstrated the algorithm's performance on datasets containing roughly one, two, and four thousand images, complexities which can not be attained with any other automated method, and which would be difficult or impossible in an interactive system.

One perhaps unexpected advantage of working at this scaling regime is that of *overdetermination and data fusion to reduce uncertainty*; our algorithms register images to within two pixels of rotational error, on average, outperforming manual bundle-adjustment due to the human operator's specification of insufficient constraints. We emphasize that the image

datasets for which we report performance were acquired outdoors, over wide baselines, under uncontrolled and varying lighting conditions, and in the presence of significant visual clutter.

7.2 Summary

This paper presented a scalable algorithm which registers thousands of images using hundreds of thousands of noisy line features. The algorithm outperforms manual bundle adjustment in both speed and accuracy. We emphasize that the image datasets for which we report performance were acquired outdoors, over wide baselines, under uncontrolled and varying lighting conditions, and in the presence of significant occlusion and visual clutter. The datasets produced by the algorithm are consistent to within a tenth of a degree, or about two pixels, across acquisition regions spanning hundreds of meters. Considered together, the algorithms presented here and in the companion paper [2] represent a new end-to-end capability for automated, absolute registration of terrestrial images.

8 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 the NTT Corporation under Award MIT9904-20.

References

[1] Yasuo Amemiya and Wayne A. Fuller. Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. *Annals of Statistics*, 12(2):497–509, June 1984.

[2] M. Antone and S. Teller. Automatic recovery of camera positions in urban scenes. Technical Report 814, MIT LCS, Dec. 2000.

[3] Matthew E. Antone and Seth Teller. Automatic recovery of relative camera rotations for urban scenes. In *Proc. CVPR*, volume 2, pages 282–289, June 2000.

[4] Stephen T. Barnard. Methods for interpreting perspective images. *Artificial Intelligence*, 21:435–462, 1983.

[5] Shawn Becker and V. Michael 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.

[6] Rudolph Beran. Exponential models for directional data. *Annals of Statistics*, 7(6):1162–1178, Nov. 1979.

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

[8] Michael Bosse, Douglas de Couto, and Seth Teller. Eyes of argus: Georeferenced imagery in urban environments. *GPS World*, pages 20–30, April 1999.

[9] John F. Canny. A computational approach to edge detection. *PAMI*, 8(6):679–698, Nov. 1986.

[10] Ted Chang. Spherical regression with errors in variables. *Annals of Statistics*, 17(1):293–306, March 1989.

[11] Robert T. Collins. *Model Acquisition using Stochastic Projective Geometry*. PhD thesis, UMASS, Sep. 1993.

[12] Robert T. Collins and R. Weiss. Vanishing point calculation as statistical inference on the unit sphere. In *Proc. ICCV*, pages 400–403, Dec. 1990.

[13] Satyan Coorg, Neel Master, and Seth Teller. Acquisition of a large pose-mosaic dataset. In *Proc. CVPR*, pages 872–878, June 1998.

[14] G. Csurka, C. Zeller, Z. Zhang, and O. Faugeras. Characterizing the uncertainty of the fundamental matrix. *Computer Vision and Image Understanding*, 68(1):18–36, Oct. 1997.

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

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[17] Cornelia Fermüller and Yiannis Aloimonos. Ambiguity in structure from motion: Sphere versus plane. *Int'l Journal of Computer Vision*, 28(2):137–154, 1998.

[18] A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *Proc. ECCV*, pages 311–326, June 1998.

[19] Joshua Gluckman and Shree Nayar. Ego-motion and omnidirectional cameras. In *ICCV*, pages 35–42, 1998.

[20] Gene H. Golub and Charles F. Van Loan. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6):883–893, Dec. 1980.

[21] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4(4):629–642, April 1987.

[22] P. V. C. Hough. A method and means for recognizing complex patterns. U. S. Patent No. 3,069,654, 1962.

[23] P. E. Jupp and K. V. Mardia. Maximum likelihood estimators for the matrix von Mises-Fisher and Bingham distributions. *Annals of Statistics*, 7(3):599–606, May 1979.

[24] Kenichi Kanatani. Analysis of 3-D rotation fitting. *PAMI*, 16(5):543–549, May 1994.

[25] A. A. Kassim, T. Tan, and K. H. Tan. A comparative study of efficient generalised Hough transform techniques. *Image and Vision Computing*, 17:737–748, 1999.

[26] Y. H. Leclerc, Q. T. Luong, and P. Fua. Self-consistency: A novel approach to characterizing the accuracy and reliability of point correspondence algorithms. In *Proc. the Image Understanding Workshop*, pages 793–807, Nov. 1998.

[27] John C. H. Leung and Gerard F. McLean. Vanishing point matching. In *Proc. ICIP*, volume 2, pages 305–308, 1996.

[28] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proc. CVPR*, pages 482–488, June 1998.

[29] Evelyne Lutton, Henri Maître, and Jaime Lopez-Krahe. Contribution to the determination of vanishing points using Hough transform. *PAMI*, 16(4):430–438, April 1994.

[30] M. J. Magee and J. K. Aggarwal. Determining vanishing points from perspective images. *Computer Vision, Graphics and Image Processing*, 26(2):256–267, May 1984.

[31] Bogdan Matei and Peter Meer. A general method for errors-in-variables problems in computer vision. In *Proc. CVPR*, volume 2, pages 18–25, June 2000.

[32] G. F. McLean and D. Kotturi. Vanishing point detection by line clustering. *PAMI*, 17(11):1090–1095, Nov. 1995.

[33] Keith E. Nicewarner and A. C. Sanderson. A general representation for orientation uncertainty using random unit quaternions. In *Proc. IEEE International Conference on Robotics and Automation*, volume 2, pages 1161–1168, May 1994.

[34] Michael J. Prentice. Spherical regression on matched pairs of orientation statistics. *Journal of the Royal Statistical Society, Series B*, 51(2):241–248, 1989.

[35] Louis-Paul Rivest. On the information matrix for symmetric distributions on the unit hypersphere. *Annals of Statistics*, 12(3):1085–1089, Sep. 1984.

[36] Li Shigang, Saburo Tsuji, and Masakazu Imai. Determining of camera rotation from vanishing points of lines on horizontal planes. In *Proc. ICCV*, pages 499–502, 1990.

[37] Jefferey A. Shufelt. Performance evaluation and analysis of vanishing point detection techniques. *PAMI*, 21(3):282–288, March 1999.

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

[39] Leonard A. Stefanski. The effects of measurement error on parameter estimation. *Biometrika*, 72(3):583–592, Dec. 1985.

[40] Camillo J. Taylor and David J. Kriegman. Structure and motion from line segments in multiple images. In *Proc. IEEE International Conference on Robotics and Automation*, pages 1615–1620, May 1992.

[41] Seth Teller. Toward urban model acquisition from geo-located images. In *Proc. Pacific Graphics*, pages 45–51, Oct. 1998.

[42] Roger Y. Tsai. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.

[43] Tinne Tuytelaars, Marc Proesmans, and Luc Van Gool. The cascaded Hough transform. In *Proc. ICIP*, volume 2, pages 736–739, 1997.

[44] G. S. Watson. *Statistics on Spheres*. John Wiley and Sons, New York, NY, 1983.

[45] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. *IJCV*, 27(2):161–195, 1998.