Simultaneous Mapping and Stereo Extrinsic Parameter Calibration Using GPS Measurements

Jonathan Kelly[†], Larry H. Matthies[*] and Gaurav S. Sukhatme[†]

***Abstract*—Stereo vision is useful for a variety of robotics tasks, such as navigation and obstacle avoidance. However, recovery of valid range data from stereo depends on accurate calibration of the extrinsic parameters of the stereo rig, i.e., the 6-DOF transform between the left and right cameras. Stereo self-calibration is possible, but, without additional information, the absolute scale of the stereo baseline cannot be determined. In this paper, we formulate stereo extrinsic parameter calibration as a batch maximum likelihood estimation problem, and use GPS measurements to establish the scale of both the scene and the stereo baseline. Our approach is similar to photogrammetric bundle adjustment, and closely related to many structure from motion algorithms. We present results from simulation experiments using a range of GPS accuracy levels; these accuracies are achievable by varying grades of commercially-available receivers. We then validate the algorithm using stereo and GPS data acquired from a moving vehicle. Our results indicate that the approach is promising.**

I. Introduction

Stereo vision is a rich sensing modality that is able provide dense bearing, range and appearance information. However, recovery of accurate range data from stereo, which is required for metric mapping and in some cases for path planning and obstacle avoidance, depends on careful calibration of the *extrinsic parameters* that define the transform between the stereo cameras. Calibration is typically carried out offline, using a specialized calibration target with known geometry.

The need for precision calibration limits our ability to build power-on-and-go robotic systems in which stereo is the primary sensor. Further, although stereo self-calibration (auto-calibration) is possible, it is well known that information about the absolute scale of the translation between the cameras cannot be obtained without external measurements, i.e., the length of the stereo baseline is a free parameter [1].

In contrast to stereo, which is primarily useful for short-range navigation, wide-area navigation systems such as GPS can supply positioning information over the entire globe. Commodity GPS receivers have become cheap and ubiquitous, and with the removal of selective availability, the accuracy of these receivers has improved dramatically. It should be possible to leverage GPS for a variety of tasks beyond positioning. In this paper we ask the following question: can we use GPS and image feature measurements alone to calibrate the extrinsic parameters of a robot-mounted stereo camera rig, while the robot is operating? We seek a metric calibration of the parameters, with a known scale factor.

To answer this question, we present a theoretical approach and simulation studies and experiments which characterize the feasibility of using GPS for calibration. We formulate the calibration problem using a batch maximum likelihood framework, in which point landmarks are viewed from multiple camera poses. Information about the absolute scale of the scene and the stereo baseline is derived entirely from GPS measurements. The calibration algorithm also produces a map of the landmarks in the environment — this allows calibration to be performed as part of a larger mapping task. Our simulation results are based on trajectory data acquired from a Pioneer 2-AT robot with an on-board GPS receiver. Although we focus on the use of GPS here, the algorithm we describe can be adapted for use with any sensor that is able to provide coarse, wide-area three-dimensional position measurements.

The remainder of the paper is organized as follows. We review related work in Section II below. In Section III, we formally define the calibration problem and motivate our approach. Section IV discusses our methods for landmark position and robot pose initialization, while Section V details the maximum likelihood calibration algorithm. We describe our simulation studies and vehicle experiments in Sections VI and VII, respectively, and present results in Section VIII. Finally, we offer some conclusions and directions for future work in Section IX.

II. Related Work

The problem of camera calibration has been studied extensively in the photogrammetry community, with work dating to the 1940s [2]. Much of the early research focused on calibration for, e.g., aerial mapping, where the necessary level of precision demands the use of sophisticated calibration equipment. The algorithmic techniques, such as bundle adjustment [3], developed for these applications have now been adopted by computer vision researchers.

Close-range self-calibration of both the intrinsic and extrinsic parameters of a stereo rig is demonstrated by Zhang, Luong

This work was funded in part by the US NSF (grants IIS-1017134 and CCF-0120778) and by a gift from the Okawa Foundation. Jonathan Kelly was supported by an Annenberg Fellowship from the University of Southern California and by an NSERC doctoral postgraduate scholarship from the Government of Canada.

[†]Jonathan Kelly and Gaurav S. Sukhatme are with the Department of Computer Science, University of Southern California, Los Angeles, California 90089, {jonathsk, gaurav}@usc.edu.

[*]Larry H. Matthies is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, lhm@jpl.nasa.gov.

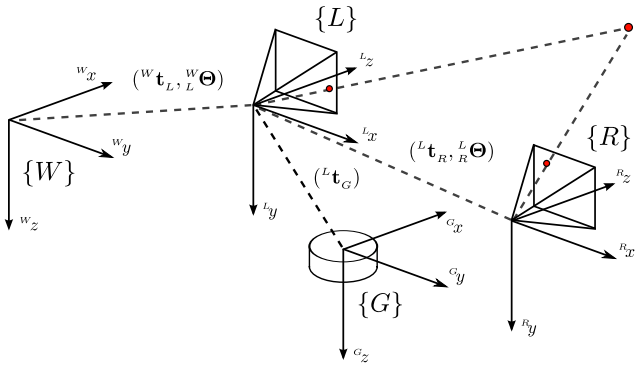

Fig. 1. Relationships between the world $\{W\}$, GPS $\{G\}$, left camera $\{L\}$, and right camera $\{R\}$ reference frames and a landmark point (red circle). The transform from the left camera frame to the world frame is $({}^{W}\mathbf{t}_{L}, {}^{W}_{L}\mathbf{\Theta})$. The (unknown) transform from the right camera frame to the left camera frame is $({}^{L}\mathbf{t}_{R}, {}^{L}_{R}\mathbf{\Theta})$. The translation of the GPS frame relative to the left camera frame is ${}^{L}\mathbf{t}_{G}$.

and Faugeras in [4], [5]. Their method involves nonlinear optimization over the parameter space, where the cost function is based on epipolar constraints between point landmarks viewed in multiple camera frames. Using the calibration result, metric information about the landmark positions can be recovered, but distances in the scene and the length of the stereo baseline are only specified up to scale. Our work is primarily concerned with determining the scale factor; we also use landmarks that lie at greater distances from the cameras.

Basic 2D stereo self-calibration for robot navigation is described in [6], where distances to obstacles are reported in a velocity-dependent coordinate system. This is appropriate for the task of obstacle avoidance only; the technique also depends on continuous motion of the robot platform.

In [7], Martinelli, Scaramuzza and Siegwart present online calibration of the extrinsic parameters of an omnidirectional camera, using an extended Kalman filter (EKF) to fuse observations of a stationary LED light. Instead of an artificial light source, we use salient point landmarks in the environment, and depend on an image feature descriptor, such as SIFT [8] or SURF [9], that is invariant to changes in feature scale and at least partially invariant to changes in feature orientation.

Our calibration algorithm borrows from recent work on structure from motion (SFM, e.g., [10]) and simultaneous localization and mapping (SLAM, e.g., [11]). The same estimation machinery used to solve many SFM and SLAM problems can be used for calibration. Although we have GPS data, we must still determine the camera orientation at each camera position (for both the left and right cameras), and estimate the positions of the landmarks. In [12], Solà suggests that accurate stereo self-calibration can be performed by fusing the output from two independent monocular SLAM estimators running in parallel, although, again, absolute scale information is missing.

III. Problem Formulation

The standard method for calibrating the extrinsic parameters of a stereo rig involves measuring the projections of known 3D landmarks (e.g., points on a calibration target) in both the left and right camera images. An optimization procedure then minimizes the image reprojection error with respect to the translation and rotation between the camera reference frames.

In our case, the problem is more difficult because the positions of the landmarks, and of the GPS antenna relative to the cameras, are initially unknown and therefore must also be estimated. We formulate this task in a batch maximum likelihood framework, where m point landmarks are observed (in both the left and right cameras) from a set of n left camera poses. This procedure is essentially an augmented form of bundle adjustment [3], in which there are constraints between the poses of the left and the right cameras, and the position of the GPS antenna, due to the rigid body transforms between the sensors. To begin, we define four reference frames:

1. the *world frame* $\{W\}$, which is an Earth-centered, Earth-fixed northing-easting-down (NED) frame,
2. the *GPS frame* $\{G\}$, with its origin at the center of the GPS receiver antenna and with the same orientation as the world frame,
3. the *left camera frame* $\{L\}$, with its origin at the optical center of the left camera, and
4. the *right camera frame* $\{R\}$, with its origin at the optical center of the right camera.

The relationship between the reference frames is shown in Figure 1. We use the computer vision convention for the orientation of the camera frames, in which the z axis is aligned with the optical axis of the lens. The origin of the world frame is selected arbitrarily, based on, e.g., the UTM zone in which the robot or vehicle is operating.

In the sections below, we denote vectors and matrices in boldface. We indicate that a vector is expressed in a particular reference frame by prefixing the vector with a left superscript that identifies the frame, e.g., ${}^{W}\mathbf{p}$ for the vector $\mathbf{p}$ expressed in the world frame.

A. System Parameterization

Our task is to jointly determine the positions of m landmarks in the world frame, the n poses of the left camera in the world frame, the transform from the right camera frame to the left camera frame, and the translation of the GPS receiver antenna relative to the left camera frame. The position of the i^{th} landmark in the world frame is represented by a 3×1 vector ${}^{W}\mathbf{p}_{l_i}, i = \{1, \ldots, m\}$. Similarly, the j^{th} pose of the left camera in the world frame is represented by a 6×1 vector

$$\mathbf{u}_j = \begin{bmatrix} {}^{W}\mathbf{t}_{L_j}^{T} & {}^{W}_{L_j}\mathbf{\Theta}^{T} \end{bmatrix}^{T}, \; j = \{1, \ldots, n\}, \tag{1}$$

where the first term, ${}^{W}\mathbf{t}_{L_j}$, is a 3×1 vector that defines the translation of the camera optical center relative to the origin of the world frame, and the second term, ${}^{W}_{L_j}\mathbf{\Theta}$, is a 3×1 vector of roll, pitch and yaw Euler angles that defines the orientation of

the camera frame relative to the world frame. The transform from the right camera to the left camera is represented by the 6×1 vector

$$\mathbf{v}_R = \begin{bmatrix} {}^L\mathbf{t}_R^T & {}^L_R\mathbf{\Theta}^T \end{bmatrix}^T, \quad (2)$$

where the 3×1 vector ${}^L\mathbf{t}_R$ defines the translation of the right camera optical center relative to the left camera optical center, and the vector ${}^L_R\mathbf{\Theta}$ defines the orientation of the right camera frame relative to the left camera frame.

We concatenate the landmark positions and left camera poses together with the right-to-left camera transform and the GPS-to-left camera translation to build the complete parameter vector

$$\mathbf{X} = \begin{bmatrix} {}^L\mathbf{t}_G^T & \mathbf{v}_R^T & \mathbf{u}_1^T & \dots & \mathbf{u}_n^T & {}^W\mathbf{p}_{l_1}^T & \dots & {}^W\mathbf{p}_{l_m}^T \end{bmatrix}^T, \quad (3)$$

where ${}^L\mathbf{t}_G$ is the 3×1 vector that defines the translation of the GPS antenna in the left camera frame. The size of the complete parameter vector is $9+6n+3m$. Note that all of the entries in the vector are static quantities which do not depend on time.

We parameterize orientations using a minimal set of three Euler angles. Although there are singularities in this representation, constraints on the motion of the platform prevent us from reaching any of the singular configurations (e.g., the pitch and roll of a land vehicle are typically limited to ± 15 degrees).

B. Camera Sensor Model

We use an ideal projective (pinhole) model for both the left and right cameras, and assume that the intrinsic and lens distortion parameters are known.[1] To compute the predicted image measurements, we begin by expressing the position of the i^{th} landmark, at position ${}^W\mathbf{p}_{l_i}$ in the world frame, in the left and right camera frames

$${}^{L_j}\mathbf{p}_{l_i} = \mathbf{C}^T({}^W_{L_j}\mathbf{\Theta})({}^W\mathbf{p}_{l_i} - {}^W\mathbf{t}_{L_j}), \quad (4)$$

$${}^{R_j}\mathbf{p}_{l_i} = \mathbf{C}^T({}^L_R\mathbf{\Theta})\left(\mathbf{C}^T({}^W_{L_j}\mathbf{\Theta})({}^W\mathbf{p}_{l_i} - {}^W\mathbf{t}_{L_j}) - {}^L\mathbf{t}_R\right). \quad (5)$$

Here, $\mathbf{C}(\mathbf{\Theta})$ is a direction cosine (rotation) matrix, parameterized by the vector $\mathbf{\Theta}$ of Euler angles.

Measurements ${}^{L_j}\mathbf{z}_{l_i}$ and ${}^{R_j}\mathbf{z}_{l_i}$ are the projections of the i^{th} landmark onto the left and right camera image planes, respectively, from the j^{th} left camera pose:

$${}^{L_j}\mathbf{z}_{l_i} = \begin{bmatrix} {}^Lu_{ij} \\ {}^Lv_{ij} \end{bmatrix} = \begin{bmatrix} {}^Lx_{ij}/{}^Lz_{ij} \\ {}^Ly_{ij}/{}^Lz_{ij} \end{bmatrix} + \boldsymbol{\eta}_{ij}, \begin{bmatrix} {}^Lx_{ij} \\ {}^Ly_{ij} \\ {}^Lz_{ij} \end{bmatrix} = (\mathbf{K}_L)\,{}^{L_j}\mathbf{p}_{l_i}, \quad (6)$$

$${}^{R_j}\mathbf{z}_{l_i} = \begin{bmatrix} {}^Ru_{l_{ij}} \\ {}^Rv_{l_{ij}} \end{bmatrix} = \begin{bmatrix} {}^Rx_{ij}/{}^Rz_{ij} \\ {}^Ry_{ij}/{}^Rz_{ij} \end{bmatrix} + \boldsymbol{\eta}_{ij}, \begin{bmatrix} {}^Rx_{ij} \\ {}^Ry_{ij} \\ {}^Rz_{ij} \end{bmatrix} = (\mathbf{K}_R)\,{}^{R_j}\mathbf{p}_{l_i}, \quad (7)$$

[1] We plan to explore full calibration of the both the intrinsic and extrinsic stereo parameters in future work.

where $\begin{bmatrix} u_{ij}, v_{ij} \end{bmatrix}^T$ is the vector of observed left (resp. right) horizontal and vertical image coordinates, $\mathbf{K}$ is the 3×3 camera intrinsic parameter matrix, and $\boldsymbol{\eta}_{ij}$ is a 2×1 white Gaussian measurement noise vector with covariance matrix $\mathbf{W}_{ij}$.

C. GPS Sensor Model

Each GPS measurement gives the position of the receiver in the world frame. Accounting for the moment arm of the GPS antenna relative to the left camera optical center, we have

$${}^{G_j}\mathbf{z} = {}^W\mathbf{t}_{L_j} + \mathbf{C}({}^W_{L_j}\mathbf{\Theta})\,{}^L\mathbf{t}_G + \mathbf{n}_j \quad (8)$$

where $\mathbf{n}_j$ is a 3×1 white Gaussian noise vector with covariance matrix $\mathbf{S}_j$.

This model neglects gross systematic errors due to multipath interference. The occurrence of these types of errors is largely dependent on the operating environment. To avoid incorporating pose measurements that include systematic errors, a chi-squared distribution test can be used to reject GPS fixes that lie outside of a specific confidence ellipsoid, based on the measurement covariance [13]. Also, for non-holonomic vehicles such as the Pioneer 2-AT robot, constraints on plausible motions may be used as an additional validation gate for the GPS data — e.g., we typically drive slowly and we know *a priori* that the robot cannot move significantly in a lateral direction over a short time interval.

IV. Landmark Position and Robot Pose Initialization

The batch calibration approach described in Section V is only valid for small-residual problems, in which the initial parameter values are reasonably close to their true values. In particular, if there are large errors in the estimates of one or more landmark positions, the calibration algorithm can converge to the wrong solution, or diverge and fail to provide an answer. The success of the algorithm therefore depends on acquiring good initial landmark and camera pose estimates. We use triangulation in combination with a *maximum disparity* heuristic to determine the initial landmark positions; camera pose estimates are derived from GPS data. The initialization techniques are described below.

A. Maximum Disparity Initialization

Estimating the camera-relative depth of a landmark in the environment using stereo normally involves some form of triangulation. However, distance values derived from triangulation are significantly affected by small errors in the estimated orientation of either the left or the right camera; these small orientation errors can produce very large errors in an estimated landmark position. The problem is most severe for landmark that lie far from the stereo rig.

We attempt to reduce the effects of triangulation errors using a *maximum disparity* heuristic. *Disparity* is a measure of the difference in the projected positions of the landmark on the left

and right camera image planes. For a fronto-parallel camera configuration, the horizontal disparity of landmark point i is

$$d_{ij} = {}^{R}u_{ij} - {}^{L}u_{ij} \tag{9}$$

where ${}^{R}u_{l_{ij}}$ and ${}^{L}u_{l_{ij}}$ are the projected horizontal image coordinates for the landmark in the right and left cameras, respectively; the disparity value is a negative quantity. For a given, fixed left horizontal image coordinate, landmarks with larger absolute disparity will be located *nearer* to the cameras.

Most stereo cameras will not, in general, be aligned in a perfectly fronto-parallel configuration, and our initial estimate of the camera pose will have some amount of error (otherwise there would be no need for calibration). However, we can still use our knowledge of the approximate relative pose of the cameras, and of the horizontal disparity, to produce a rough estimate of the landmark position. This approximation is poor for small disparities, but reasonably good for large disparities.

Our approach is to delay initializing the position of landmark i until we find the left camera pose for which the left-right horizontal disparity is the largest possible, relative to all poses where the landmark is visible. We further constrain the image plane points to lie within a fixed horizontal distance of the principal point, which is less than the full size of the image plane. This prevents initialization using points which lie at the edges of the left or right image plane (in our current implementation, points must lie within 250 pixels of the principal point, on either side). We then triangulate the left camera-relative the position of the landmark. The result is that, in the majority of cases, the position of the landmark is initialized when the robot and the landmark are in close proximity, and the initial estimate of the landmark position is reasonably close to the true position. This technique can still fail, however, in cases where one or more landmarks lie far from all of the camera poses (and the maximum disparity value is small); we discuss this issue further in Section VI.

B. Initial Pose Estimation and Landmark Triangulation

The calibration algorithm requires an initial estimate of the left camera pose at the time each GPS measurement is acquired. We initialize the left camera position using the available GPS fix and an approximate GPS antenna translation vector (from, e.g., hand measurements or CAD data etc.). This gives the translation of the left camera relative to the origin of the world frame, but in general GPS does not provide reliable information about the heading of the robot. For a non-holonomic platform (such as the Pioneer 2-AT), and assuming that the left camera optical axis is approximately aligned with the longitudinal axis of the robot, we can estimate heading using a line segment joining the positions defined by two GPS measurements spaced closely in time.[2] Because GPS altitude data is usually less accurate than the horizontal positioning information, we assume that the optical axis of camera is initially horizontal.

[2]Here we also assume that the robot moves slowly, so that its trajectory is approximately linear between GPS updates.

Fig. 2. Pioneer AT-2 with μBlox LEA-5H GPS unit, configured for data logging experiments on the USC campus. The GPS antenna is visible at the center of the aluminum crossbeam.

For camera pose j, the initial 3D positions of the visible landmarks (which have maximum disparity at pose j) are then found by stereo triangulation, using the technique described in [14]. Given a pair of corresponding left and right image point measurements, ${}^{L_j}\mathbf{z}_{l_i}$ and ${}^{R_j}\mathbf{z}_{l_i}$, we back-project rays from the left and right camera optical centers through the image plane points. If the image plane measurements were error-free, these rays would intersect at a 3D single point, however noise and matching errors inevitably cause the rays to diverge. Instead, we find the midpoint of the shortest perpendicular segment connecting the rays. This midpoint is selected as the initial landmark position, after transforming from the left camera frame to the world frame. More recently, we have also explored the use of an inverse depth-based parameterization for landmark positions, to better represent the landmark position uncertainty [15].

V. Calibration Algorithm

We use a batch iterated maximum likelihood formulation for the complete calibration problem, in which we simultaneously solve for the landmark positions, left camera poses, translation of the GPS antenna, and the extrinsic calibration parameters. First, we stack the image plane and GPS measurements to form the complete observation vector

$$\mathbf{Z} = \begin{bmatrix} {}^{L_1}\mathbf{z}_{l_1}^T & {}^{R_1}\mathbf{z}_{l_1}^T & \dots & {}^{L_n}\mathbf{z}_{l_k}^T & {}^{R_n}\mathbf{z}_{l_k}^T & {}^{G_1}\mathbf{z}^T & \dots & {}^{G_n}\mathbf{z}^T \end{bmatrix}^T. \tag{10}$$

The value k is the index of the last landmark visible from pose n. The observation covariance matrices for the image plane and GPS measurements are, respectively,

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{0}_{2\times 2} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{2\times 2} & \cdots & \mathbf{W}_{kn} \end{bmatrix}, \; \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & \cdots & \mathbf{0}_{3\times 3} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{3\times 3} & \cdots & \mathbf{S}_n \end{bmatrix}. \tag{11}$$

The complete observation covariance matrix is then

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix}. \tag{12}$$

Note that, because the image and GPS measurements are independent and uncorrelated, the covariance matrix is block diagonal and can be inverted quickly.

We define the parameter and observation error vectors as, respectively,

$$\delta\hat{\mathbf{X}} = \mathbf{X} - \hat{\mathbf{X}}, \quad \delta\mathbf{Z} = \mathbf{Z} - \hat{\mathbf{Z}}. \tag{13}$$

where $\hat{\mathbf{X}}$ is the current estimated parameter vector and $\hat{\mathbf{Z}}$ is the predicted observation vector based on $\hat{\mathbf{X}}$.[3] The update step of the iterated maximum likelihood algorithm involves linearizing about the current parameter estimate, and we therefore require the Jacobians of the image plane and GPS measurements with respect to the pose and calibration parameters. The Jacobians of an image plane point with respect to the i^{th} landmark position and the j^{th} left camera pose are computed as

$$\mathbf{H}_{\mathbf{z}_{l_i},\mathbf{p}_{l_i}} = \begin{bmatrix} \dfrac{\partial {}^{L_j}\mathbf{z}_{l_i}}{\partial \mathbf{p}_{l_i}} \\ \dfrac{\partial {}^{R_j}\mathbf{z}_{l_i}}{\partial \mathbf{p}_{l_i}} \end{bmatrix}, \quad \mathbf{H}_{\mathbf{z}_{l_i},\mathbf{u}_j} = \begin{bmatrix} \dfrac{\partial {}^{L_j}\mathbf{z}_{l_i}}{\partial \mathbf{u}_j} \\ \dfrac{\partial {}^{R_j}\mathbf{z}_{l_i}}{\partial \mathbf{u}_j} \end{bmatrix}. \tag{14}$$

The Jacobian of an image plane point with respect to the right camera extrinsic parameters, the Jacobian of a GPS measurement with respect to the j^{th} left camera pose, and the Jacobian of a GPS measurement with respect to the GPS translation parameters are

$$\mathbf{H}_{\mathbf{z}_{l_i},\mathbf{v}_R} = \frac{\partial {}^{R_j}\mathbf{z}_{l_i}}{\partial \mathbf{v}_R}, \quad \mathbf{H}_{\mathbf{u}_j} = \frac{\partial {}^{G_j}\mathbf{z}}{\partial \mathbf{u}_j}, \quad \mathbf{H}_{\mathbf{t}_G} = \frac{\partial {}^{G_j}\mathbf{z}}{\partial {}^{L}\mathbf{t}_G}. \tag{15}$$

The complete Jacobian matrix $\mathbf{H}$ is formed by inserting the partial derivative matrices above at the appropriate row and column positions.[4] Matrix $\mathbf{H}$ is sparse and block diagonal except for the first nine columns – as such, operations involving $\mathbf{H}$ are amenable to optimization using sparse matrix multiplication techniques.

The maximum likelihood estimate for the parameters is obtained by iteratively performing a Levenberg-Marquardt update, solving the system

$$\left(\mathbf{H}^T\mathbf{\Sigma}^{-1}\mathbf{H} + \lambda \cdot \text{diag}(\mathbf{H}^T\mathbf{\Sigma}^{-1}\mathbf{H})\right)\delta\mathbf{X} = \mathbf{H}^T\mathbf{\Sigma}^{-1}\delta\mathbf{Z}. \tag{16}$$

Here, λ is a damping factor which controls the direction of motion along the parameter error surface; larger values of λ force the update more towards gradient descent [3]. The updated estimate for the parameter vector at iteration i is

$$\hat{\mathbf{X}}_{i+1} = \hat{\mathbf{X}}_i + \delta\hat{\mathbf{X}}_i. \tag{17}$$

This process is iterated until convergence. We determine that the estimate has converged when the two-norm of the difference between the six extrinsic parameters over consecutive iterations is less than a small positive constant, ϵ (in our implementation, $\epsilon = 10^{-6}$). If, on iteration $i+1$, the squared observation error increases relative to iteration i, we update the damping factor as $\lambda_{new} = 10\lambda_{old}$ and repeat the iteration using the previous parameter estimate. Otherwise, we decrease λ by a factor of 10 and continue; this is a standard heuristic used in Levenberg-Marquardt optimization. As a last step, we determine the maximum likelihood parameter vector by performing an update with $\lambda = 0$.

[3] Estimated quantities are denoted with the ˆ (hat) symbol.

[4] For brevity, we omit the complete Jacobians. The Jacobians with respect to the camera poses are particularly complex because the measurement involves division by the camera-relative landmark depth.

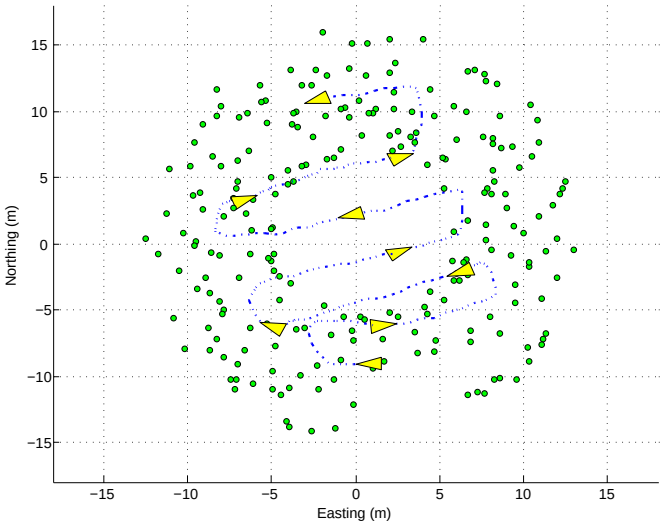

Fig. 3. Robot trajectory (dashed blue line) with overlaid synthetic point landmarks (green dots). The complete trajectory is approximately 79 meters in length. A total of 225 landmarks are visible from 229 camera poses. The yellow triangles indicate the estimated orientation of the robot and left camera at various positions along the trajectory.

VI. Simulation Studies

To evaluate the performance of the calibration algorithm, we initially performed a series of simulation experiments using a combination of real and synthetic data. We drove a Pioneer 2-AT robot equipped with an on-board μBlox LEA-5H GPS receiver [16] in an open area on the USC campus, while logging GPS and wheel odometry data in real time. The update rates for GPS and odometry were 1 Hz and approximately 10 Hz, respectively. Total length of the trajectory was 79.3 m as measured by wheel odometry. The odometry data was used only to verify that there were no gross errors in the GPS measurements.

Based on the area covered by the (real) trajectory of the robot, we then generated a set of 240 landmark points at random positions on an annulus with an inner radius of 5 meters and an outer radius of 13 meters, and with random heights between -0.5 and 0.5 meters. We used this trajectory and set of synthetic landmark points, shown in Figure 3, as ground truth for our simulations.

For each simulation trial, we captured a pair of simulated images from the stereo cameras at every position along the trajectory for which a GPS fix was available. After projecting all visible landmarks into both the left and right image planes, we added independent, zero-mean Gaussian noise with a standard deviation of 1.0 pixels to the image coordinates, to simulate errors in feature localization. Each (simulated)

camera had a resolution of 640×480 pixels. We used the same intrinsic parameter matrix for both cameras, with a focal length of 500 pixels and with the principal point located at the center of the image plane.

We chose to run the simulation trials using the error characteristics of a selection of three different, commercially-available GPS receivers: the NovAtel OEMV-1 and OEMV-2 [17], [18] and the μBlox LEA-5H [16]. Receiver accuracy is usually quoted in terms of the Circular Error Probable (CEP) value, i.e., the radius of a (local tangent plane) circle in which 50% of the observed measurements will lie [19]. Depending on the type of aiding employed, the receivers have accuracies, listed in Table I, that vary from 0.02 m CEP to 2.0 m CEP. We assume a Gaussian error distribution on the GPS measurements, and convert the CEP values to the 1σ values shown in the third column of the table. Note that the accuracy values range across two order of magnitude.

The initial positions of the robot were generated by perturbing the GPS measurements with zero-mean Gaussian noise, according to the selected level of accuracy in Table I. Our early experiments indicated that the calibration algorithm is very sensitive to situations in which landmarks and/or camera poses are only *weakly observable*, that is, when the available measurements weakly constrain the relevant landmark or pose parameters. To ensure that the problem is fully observable, we keep only landmarks which have been observed (by both the left and right cameras) from at least six poses, and require that at least six landmarks are visible in both cameras at every pose.

VII. Experiments

Based on our simulation studies, we then analyzed data from a calibration test run with a ground vehicle; the vehicle is shown in Figure 4. A stereo beam was mounted on the roof, with two black and white Flea FireWire cameras from Point Grey Research (640×480 pixel resolution), mated to 4 mm Navitar lenses (58° horizontal FOV, 45° vertical FOV). The stereo baseline was 30 cm. GPS measurements were recorded from a μBlox LEA-5H receiver, with the antenna placed just above center of the vehicle's front windshield.

We gathered experimental data during a test run near the Santa Monica airport in Los Angeles, California. The car was driven along a semi-circular trajectory, over a distance of approximately 140 m. We logged a total of 532 stereo pairs and 296 GPS measurements. Prior to the start of data logging, we gathered calibration data for the cameras using a standard planar camera calibration target. These values were then compared with those produced by the GPS-based algorithm (see Section VIII).

VIII. Results and Discussion

Simulation results for receiver configurations 1 and 2, with 0.02 meter CEP and 0.20 meter CEP accuracy, respectively, are listed in Table II. We were not able to obtain reliable calibration results using GPS measurements with a 2.0 meter CEP (Table I, line 3); this amount of noise resulted in initial camera pose and landmark position estimates that were too far from the true values for the algorithm to converge consistently. As such, we restrict the discussion below to configurations 1 and 2. We also focus primarily on our simulation results, as the results from our vehicle experiments are preliminary.

Fig. 4. Ground vehicle equipped for data collection experiments. The stereo cameras and GPS receiver are mounted on the roof of the vehicle, as shown in the inset image.

The complete trajectory shown in Figure 3 consists of 229 left camera poses, from which a total of 225 landmarks were visible (from six or more poses each) out of the 240 candidate landmarks. Calibration values in columns three and five of Table II are averages over 10 trials, with different randomly-generated Gaussian noise for each trial.

At the start of each trial, we set the true extrinsic parameters to the values shown in the second column of Table II. These values correspond to a fronto-parallel stereo geometry with a 30 centimeter baseline. We then initialized the estimated (erroneous) parameter values as follows: a 35 centimeter baseline in x, 2 centimeters in y and -2 centimeters in z,

TABLE I
GPS receiver noise characteristics used for simulation studies.

Configuration	CEP (m)	1σ (m)	Example
1	0.020	0.017	NovAtel OEMV–2 RT-2
2	0.200	0.170	NovAtel OEMV–1 RT-20
3	2.000	1.700	μBlox LEA–5H SBAS

TABLE II
RIGHT CAMERA EXTRINSIC PARAMETER CALIBRATION RESULTS FOR GPS RECEIVER CONFIGURATIONS 1 AND 2.

		Configuration 1		Configuration 2	
Parameter	Truth	Average	σ	Average	σ
x (mm)	300.0	299.3	0.6	297.2	4.7
y (mm)	0.0	1.4	0.2	-0.4	0.5
z (mm)	0.0	3.6	0.3	21.1	3.4
Roll α (mdeg)	0.0	0.4	0.1	0.1	0.2
Pitch β (mdeg)	0.0	-1.9	0.1	-1.3	0.4
Yaw γ (mdeg)	0.0	1.5	0.1	0.2	0.1

with 2 degrees of positive roll and yaw error, and 4 degrees of negative pitch error, i.e. with the cameras verged by 4 degrees.

The results show that, for both configurations, the average residual camera orientation error after calibration is on the order of one millidegree in roll, pitch and yaw. This is to be expected, as camera rotation errors have a large effect on the estimated landmark positions, and the batch estimator must therefore drive the orientation errors close to zero to obtain a low overall residual error. Indeed, we observed exactly this behavior over sequential iterations of batch algorithm: the rotation parameters typically converged first, followed by the translation parameters.

The average residual errors for the translation parameters are somewhat larger. For configuration 1, the average error is less than 4 millimeters along all axes. This result, however, depends on a level of GPS accuracy that can only be achieved by real-time carrier-phase differential receivers – at present, these units are very expensive and their deployment is limited.

For configuration 2, the average residual error is less than 3 millimeters in x and y – along the x direction, the error is less than 6% of the original error value at the start of the simulation. The average residual error along the z axis is larger than along the other axes, however, and slightly larger on average than the initial error introduced at the start of the simulation. Achieving better calibration results for the z translation parameter may simply be a matter of collecting more data. We are exploring this issue.

In analyzing our results, we noted that accurate stereo calibration can sometimes be obtained even when there are relatively large errors in the 3D positions of several landmarks. This is because the calibration algorithm minimizes image reprojection error — for landmarks that are visible from a small number of clustered poses only, and that lie at a significant distance from the cameras in all cases, the reprojection error is relatively insensitive to landmark depth.

Our simulation results are based on incremental GPS measurements that are acquired as the robot or vehicle moves, navigating or performing some other activity. It is possible to obtain more accurate positioning information simply by remaining stationary and filtering a large amount of GPS data. This increased accuracy comes at the expense of the additional time, however, which may be unacceptable in some situations.

TABLE III
RIGHT CAMERA EXTRINSIC PARAMETER CALIBRATION RESULTS FOR VEHICLE EXPERIMENT.

Parameter	Target-Based Value	GPS-Based Value
x (mm)	298.2	293.1
y (mm)	2.6	4.6
z (mm)	2.9	13.6
Roll α (deg)	-0.12	-0.14
Pitch β (deg)	0.14	0.15
Yaw γ (deg)	-0.74	-0.72

Results for our experiment with the test vehicle are also in agreement with the values determined using the standard target-based calibration procedure, as shown in Table III. We note that, in practice, when a large number of GPS satellites are in view, the μBlox LEA-5H receiver is able to obtain position fixes with an accuracy significantly better than 2.0 m CEP. We are presently conducting additional experiments to determine the performance of the algorithm under a wider variety of conditions.

IX. CONCLUSIONS AND FUTURE WORK

This paper presented an approach for calibrating a robot- or vehicle-mounted stereo rig, using GPS measurements to determine the absolute scale of the scene and of the stereo baseline. This work is a step towards developing robots that can operate for long periods of time without requiring manual sensor re-calibration.

Our results are promising: we obtained reasonable calibration accuracy using GPS measurements with a CEP of up to 0.2 meters. A CEP of 0.2 meters approaches the accuracy available with standard differential GPS, which is readily available in many locations. The batch algorithm establishes a benchmark for other approaches — we believe

Fig. 5. Example left stereo camera image acquired during one vehicle calibration experiment.

that incremental solutions, which incorporate larger numbers of observations over time, should be able to improve upon these results. Further, as the global satellite navigation network grows to incorporate the Russian GLONASS and European Galileo constellations, we can expect even better positioning accuracy from commodity receivers. Also, the algorithm we have described is not limited to GPS — it can be adapted for use with other sensors that provide positioning or ranging information.

There are several directions for future work. We are currently exploring the use of a combination of wheel odometry and GPS measurements to perform full calibration of both the intrinsic and extrinsic parameters of the stereo rig. Based on the batch solution, we are also developing an alternative, sequential estimator formulation. Lastly, we would like to define optimal or near-optimal robot trajectories that enable rapid calibration in the field, based on the uncertainty associated with each calibration parameter.

REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, United Kingdom: Cambridge University Press, Nov. 2003.

[2] J. C. McGlone, E. M. Mikhail, and J. S. Bethel, Eds., *Manual of Photogrammetry*, 5th ed. American Society for Photogrammetry and Remote Sensing, July 2004.

[3] B. Triggs, P. F. McLauchlan, R. Hartley, and A. W. Fitzgibbon, "Bundle Adjustment — A Modern Synthesis," in *Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science, B. Triggs, A. Zisserman, and R. Szeliski, Eds. Berlin: Springer, Jan. 2000, vol. 1883/2000, ch. 21, pp. 298–372.

[4] Z. Zhang, Q.-T. Luong, and O. D. Faugeras, "Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction," INRIA, Sophia Antipolis Cedex, France, Tech. Rep. 2079, Oct. 1993.

[5] ——, "Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction," in *Proc. 12th IAPR Int'l Conf. Pattern Recognition (ICPR'94)*, vol. 1, Jerusalem, Israel, Oct. 1994, pp. 695–697.

[6] R. A. Brooks, A. M. Flynn, and T. Marill, "Self Calibration of Motion and Stereo Vision for Mobile Robot Navigation," Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, USA, Tech. Rep. AIM-984, Aug. 1987.

[7] A. Martinelli, D. Scaramuzza, and R. Siegwart, "Automatic Self-Calibration of a Vision System during Robot Motion," in *Proc. IEEE Int'l Conf. Robotics and Automation (ICRA'06)*, Orlando, USA, May 2006, pp. 43–48.

[8] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 2, no. 60, pp. 91–110, Nov. 2004.

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," in *Computer Vision — ECCV 2006: 9th European Conf. Computer Vision, Graz, Austria, May 7–13, 2006, Proc. , Part I*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin: Springer, 2006, vol. 3951/2006, pp. 404–417.

[10] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from Motion Causally Integrated Over Time," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 523–535, Apr. 2002.

[11] S. Thrun, "Robotic Mapping: A Survey," Robotics Institute, Carnegie Mellon University, Pittsburgh, USA, Tech. Rep. CMU-CS-02-111, Feb. 2002.

[12] J. Solà, "Multi-camera VSLAM: from former information losses to self-calibration," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems Workshop on Visual SLAM: An Emerging Technology*, San Diego, USA, Oct./Nov. 2007.

[13] S. Sukkarieh, E. M. Nebot, and H. F. Durrant-Whyte, "A High Integrity IMU/GPS Navigation Loop for Autonomous Land Vehicle Applications," *IEEE Trans. Robotics and Automation*, vol. 15, no. 3, pp. 572–578, June 1999.

[14] Y. Cheng, M. W. Maimone, and L. H. Matthies, "Visual Odometry on the Mars Exploration Rovers," in *Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics*, vol. 1, Big Island, USA, Oct. 2005, pp. 903–910.

[15] J. M. M. Montiel, J. Civera, and A. J. Davison, "Unified Inverse Depth Parametrization for Monocular SLAM," in *Proc. Robotics: Science and Systems (RSS'06)*, Philadelphia, USA, Aug. 2006.

[16] "LEA-5 u-blox 5 Modules for GPS and GALILEO Data Sheet," uBlox AG, Thalwil, Switzerland, Jan. 2008. [Online]. Available: http://www.u-blox.com/en/lea-5a.html

[17] "NovAtel OEMV-1 GPS Receiver Data Sheet," NovAtel Inc., Calgary, Canada, Jan. 2008. [Online]. Available: http://www.novatel.com/products/gnss-receivers/oem-receiver-boards/oemv-receivers/

[18] "NovAtel OEMV-2 GPS+GLONASS Receiver Data Sheet," NovAtel Inc., Calgary, Canada, Jan. 2008. [Online]. Available: http://www.novatel.com/products/gnss-receivers/oem-receiver-boards/oemv-receivers/

[19] J. A. Farrell and M. Barth, *The Global Positioning System & Inertial Navigation*, 1st ed. New York: McGraw-Hill, Dec. 1998.