Calibration of an Outdoor Active Camera System *

Robert T. Collins and Yanghai Tsin
The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. 15213
Email: {rcollins,ytsin}cs.cmu.edu

Abstract

A parametric camera model and calibration procedures are developed for an outdoor active camera system with pan, tilt and zoom control. Unlike traditional methods, active camera motion plays a key role in the calibration process, and no special laboratory setups are required. Intrinsic parameters are estimated automatically by fitting parametric models to the optic flow induced by rotating and zooming. No knowledge of 3D scene structure is needed. Extrinsic parameters are calculated by actively rotating the camera to sight a sparse set of surveyed landmarks over a virtual hemispherical field of view, yielding a well-conditioned pose estimation problem.

1. Introduction

This paper develops a parametric projection model for the intrinsic (lens) and extrinsic (pose) parameters of a camera with active pan, tilt and zoom control. Calibration procedures are presented for estimating intrinsic parameters by fitting parametric models to the optic flow induced by rotating and zooming the camera. These calibration procedures are fully automatic and require no precise knowledge of 3D scene structure. We do not assume any special distribution of features in the world (e.g. a well-distributed set of distinctive corners or straight lines). Extrinsic parameters are calculated by sighting a sparse set of measured landmarks in the scene. Actively rotating the camera to measure landmarks over a virtual hemispherical field of view leads to a well-conditioned pose estimation problem.

The calibration procedures are specifically designed for *in-situ* (meaning "in place") camera calibration, as opposed to pre-calibrating the camera in a laboratory and then carrying it elsewhere. We believe that all cameras should be calibrated in an environment that resembles their actual operating conditions. This philosophy is particularly relevant for outdoor camera systems. Cameras get jostled during transport and installation, and changes in temperature and humidity can affect a camera's intrinsic parameters. Furthermore, it is impossible to recreate the full range of zoom and focus settings that are useful to an outdoor camera system within the confines of an indoor lab.

Unfortunately, outdoors is not an ideal environment for careful camera calibration. It can be cold, rainy, or otherwise unpleasant. Simple calibration methods are needed that can be performed with minimal human intervention. The active calibration procedures presented here fit this description.

2. Active Camera Model

In this section we develop a camera projection model for a camera with active pan, tilt and zoom. The model is a generalization of the well-known Tsai camera model [12]. We choose the Tsai model as a basis since it is representative of the vast majority of camera models used in computer vision and robotics research. Development of the model has relied heavily on the work of Willson [13, 14].

Although our active camera model is meant to apply to a broad class of pan-tilt-zoom cameras, the target camera platform is a Sony EVI-370 camera mounted on a Directed Perception (DP) PTU-46-70 pan-tilt unit. The EVI-370 spec sheet reports a 12X zoom (see Figure 1) divided into 1024 discrete zoom settings with a horizontal field of view of approximately 48.8 degrees at zoom setting 0 and 4.3 degrees at zoom setting 1023. The DP pan-tilt head has a resolution of 0.771 arc minutes over a pan angle range of 318 degrees and tilt range of 78 degrees.

Figure 1. Example of Sony EVI-370 12X zoom.

*Funded by DARPA VSAM contract DAAB07-97-C-J031.

Extrinsic Parameters:

The extrinsic camera equation is a kinematic chain representing the transformation of a scene point (X_w, Y_w, Z_w) into the same point (X_c, Y_c, Z_c) specified in a camera-centered coordinate system.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R_m \, R_\theta \, R_\phi \, R \, (\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} - \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}) \tag{1}$$

where $T = (T_x, T_y, T_z)$ specifies the scene location of the camera focal point, R is the scene orientation of the pan-tilt unit when pan = tilt = 0, R_ϕ is a rotation by pan angle ϕ, R_θ is a rotation by tilt angle θ, and R_m specifies the orientation of the camera as physically mounted on the pan-tilt head.

Intrinsic Parameters:

The intrinsic camera equations relate point (X_c, Y_c, Z_c) with its projected pixel location (X_f, Y_f) in the image frame.

$$\frac{d_x/s_x}{M(z)} \, [X_f - C_x(z)] \, (1 + \kappa r^2) \quad = \quad f \, \frac{X_c}{Z_c} \tag{2}$$

$$\frac{d_y}{M(z)} \, [Y_f - C_y(z)] \, (1 + \kappa r^2) \quad = \quad f \, \frac{Y_c}{Z_c} \tag{3}$$

with

$$r^2 = \left[\frac{d_x/s_x}{M(z)} \, (X_f - C_x(z)) \right]^2 + \left[\frac{d_y}{M(z)} \, (Y_f - C_y(z)) \right]^2 .$$

In these equations f is the focal length at zoom setting 0 (widest-angle), $M(z)$ is image magnification indexed by zoom setting, $C_x(z)$ and $C_y(z)$ are the pixel coordinates of the image center (see discussion below) indexed by zoom, s_x is a scale factor that compensates for non-square aspect ratio, and κ is the first-order coefficient of radial lens distortion (refer to [12] for details). Two predetermined constants in the Tsai model, d_x and d_y, specify the dimensions of a pixel in millimeters on the focal plane. We are content to measure camera parameters in pixel units rather than millimeters, and set $d_x = d_y = 1$.

Discussion

Some of the issues involved in designing the above camera model are discussed here. In many cases, a tradeoff has been made for simplicity over strict geometric accuracy. Some known factors have been intentionally left out of the model, since they have only second-order affects on image appearance. The overriding goal has been to devise a set of parameters that can be measured stably from image data without requiring precisely measured calibration targets.

1. Representing pan, tilt and the physical camera mount as rotation matrices assumes a pure rotation model where all rotation axes pass precisely through the camera focal point. Unless the pan-tilt mechanism is specially manufactured, this abstraction is unlikely to hold in practice. However, the amount of induced parallax is negligable for an outdoor camera viewing distant scene structure. This is one case where *in-situ* outdoor calibration has a benefit over calibration in the confines of an indoor laboratory.

2. Adding a magnification term $M(z)$ to represent lens zoom is non-standard. Typically both focal length f and radial distortion coefficient κ are written as functions of the zoom setting [13]. Writing $f(z)$ alone would not suffice because the effects of radial distortion decrease as the field of view narrows. In our formulation, the image pixel radius r^2 is implicitly a function of zoom (it has an $M(z)$ term in it), and therefore even when κ is constant the pixel displacements due to distortion will decrease as the zoom increases. In this respect, our model correctly reflects the qualitative behavior of radial distortion with respect to zoom, while using fewer parameters. To precisely capture the quantitative relationship would require computing values for κ at several different zoom settings, as in [13]. This is time consuming and hard to perform accurately outside of a calibration lab. A trade-off has been made here for simplicity (fewer parameters) over strict geometric accuracy.

3. Image center also varies with zoom [13], and thus pixel coordinates $C_x(z)$ and $C_y(z)$ are written as functions of the zoom setting. For our cameras, the image center can vary by as much as 40 pixels from low-zoom to high-zoom. It is also known that the location of the focal point T, an extrinsic parameter, is displaced minutely along the optic axis with changing zoom [8, 13]. For the distances between camera and scene structure that we are interested in, this tiny displacement of the focal point can be ignored.

4. There are many potential definitions of image center [14]. At least three different definitions potentially describe the meaning of C_x and C_y in Equations (2) and (3): principal point, center of zoom expansion, and center of radial distortion. It will be clear from the calibration procedure outlined in Section 3.2 that we compute C_x and C_y as the center of zoom expansion (as does [8]). The principal point is notoriously hard to compute accurately, particularly when the objects viewed are distant [7]. In contrast, the zoom center is easy to calculate correctly from image data. Alternatively, two different image centers, one for zoom and one for principal point, could be incorporated into the model, but at the cost of introducing two more parameters and the risk of overfitting. A similar argument applies to the center of radial distortion for narrow to moderate field-of-view lenses.

5. The equations include only one coefficient of radial distortion, and no tangential distortion terms. For narrow to moderate field-of-view lenses, the first coefficient of radial distortion dominates the effects of the other distortion terms on image appearance.

3. Intrinsic Calibration

Zooming and pure rotation of a camera induce image pixel displacements that do not depend on 3D scene structure. We use this fact to develop calibration methods that do not require knowledge of the scene geometry.

Previous calibration methods using active camera zoom and rotation have been reported. For zoom calibration, Willson [13] is the most comprehensive work to date. He methodically steps through the zoom and focus settings of the camera, performing a full camera calibration at each step. Li and Lavest [8] studied different feature configurations for zoom calibration, and reinforced the notion that a good set of features covers as much of the image as possible while being as densely spaced as possible.

It is well-known that intrinsic parameters can be calibrated using a set of images related by pure rotation. The process is known as *self-calibration* in the projective geometry literature [4]. Stein [10] provides an accessible description, and shows how to explicitly calibrate for intrinsic parameters including lens distortion. Basu and Ravi [1] develop simple methods for determining focal length, aspect ratio and image center using camera pan, tilt and roll.

Stevensen and Fleck [11] present an interesting active calibration approach using rotation and translation of a camera mounted on a robot arm. Feature extraction is simplified by using a single point light source in a dark room. The authors do not use a standard parameteric camera model, but instead directly tabulate a lookup table relating radial angle from the principal point to distance in the image.

All of these existing active camera calibration approaches use a sparse set of simple scene features such as corners or lines. The assumption is that a good distribution of such features across the entire field-of-view can be obtained. This is possible in an indoor environment, particularly if one is willing to paste calibration grids on the walls of the room. In an outdoor environment, a good distribution of corner or line features is not always possible.

3.1. Calibration by Image Warping

Our basic approach to intrinsic calibration is to perform a known camera zoom or rotation, and then compare the optic flow predicted by the camera projection equations (Eq. 2,3) with the actual observed pixel displacements. Relevant subsets of the camera parameters are adjusted until the sum of squared difference (SSD) between predicted and actual pixel positions achieves a minimum. Initial outdoor experiments using automatic detection and tracking of corner-like features through an image sequence soon exhibited serious limitations. Independently moving objects such as vehicles gave rise to outlier displacement vectors that caused problems due to the sparseness of the entire feature set. Furthermore, the natural distribution of "interesting" features in the scene was never as uniform across the field of view as one would like.

These observations led us to abandon sparse feature tracking methods in outdoor environments, and to focus instead on a dense optic flow approach based on image warping. The approach is similar to the work of Bergen et.al. [2] where a search through the space of affine or projective parametric warps is performed to align an incoming image with a reference frame. The major difference is that our warping transformations are written in terms of physically meaningful intrinsic camera parameters, and thus the process of discovering the best image alignment yields a direct estimate of the camera parameters.

Consider a reference image $I_1[x_1]$ indexed by pixel coordinates x_1. A change in the values of any of the camera parameters p will result in a new image $I_2[x_2]$ being observed. The displacement field $x_2 - x_1$ represents the *optic flow* induced by the change in camera parameters. The flow for active zoom and rotation of the camera does not depend on 3D scene structure, and we can write an invertible nonlinear transformation G that maps each pixel x_1 to its new location $x_2 = G(x_1; p)$, and vice versa $x_1 = H(x_2; p)$, where H is the inverse of G. We can thus predict how the new image will appear:

$$I_w[x] = I_1[H(x; p)] \,.$$

How well the predicted image I_w matches the actual observed image I_2 depends on how accurately we know the camera parameters p. We can improve our estimates of the parameters by adjusting them to minimize an SSD error function

$$E(p) = \sum_{x \in V} (I_2[x] - I_1[H(x; p)])^2 \,/ \sum_{x \in V} 1 \qquad (4)$$

where V is the set of "valid" pixels such that $H(x; p)$ is a proper index into image I_1. Bilinear interpolation is used to compute intensity values of noninteger pixel coordinates. The denominator of Eq. (4) serves to compute the average squared error over all valid pixels.

Care must be taken when computing E with raw intensity values. For a camera with 12X zoom and automatic gain control, the view at high-zoom is likely to have a significantly different brightness than the corresponding portion of the image seen at low-zoom. Furthermore, changes in outdoor scene illumination during a zoom or rotation sequence are possible. Motivated by [6], we perform a preprocessing step consisting of histogram equalization followed by reduction of the 8-bit intensity range to just 4 bits (16 distinct intensity values). This stretching and quantization normalizes intensity gain and offset, and removes intensity fluctuations due to noise.

Searching for parameter values that minimize the SSD function is performed using Powell's method [9]. This variant of coordinate descent optimization minimizes each parameter in turn using line search minimization. The method cycles repeatedly through all parameters until the function cannot be minimized further. Although it is slower than gradient-descent approaches such as Levenburg-Marquardt, it has the distinct benefit that no derivatives need to be computed for the function being minimized.

3.2. Calibration via Zooming

Our first calibration step is to compute image magnification $M(z)$ and the center of zoom expansion $(C_x(z), C_y(z))$ as lookup tables where z runs from 0 to 1023 for the Sony camera. For this step we simplify the projection equations (Eq. 2,3) by setting the radial distortion coefficient k to 0.

The calibration procedure is as follows. An initial image I_0 is taken at the widest-angle zoom setting $z = 0$. Subsequent images are taken at incrementally increasing zoom values with a step size of 5, yielding 205 images total. For each zoom level n, the best values for magnification and zoom center are found by minimizing the SSD of the predicted transformation between I_0 and the current image I_n. This transformation takes the form of an isotropic scaling:

$$X'_f = M(n)\,[X_f - C_x(n)] + C_x(n) \quad (5)$$
$$Y'_f = M(n)\,[Y_f - C_y(n)] + C_y(n) \quad (6)$$

All camera parameters are held fixed except for $M(n)$, $C_x(n)$ and $C_y(n)$, which are adjusted by Powell's method until the sum of squared differences between the predicted zoom image and the observed zoom image is at a minimum. Initial estimates for the zoom parameters are $M(n) = M(n-1)$, $C_x(n) = C_x(n-1)$ and $C_y(n) = C_y(n-1)$ with a base case of $M(0) = 1$, $C_x(0) = 320$ and $C_y(0) = 240$.

Each zoom image in the sequence of 205 images is processed separately, and the resulting sets of estimates for magnification and zoom center are linearly interpolated to yield lookup tables indexed from 0 to 1023. Figure 2a shows the resulting lookup tables for magnification wrt zoom for five different Sony EVI-370 cameras, superimposed on the same graph. Each estimated curve is very smooth, and all are in good agreement over the whole zoom range. Figure 2b shows the image center lookup tables computed for the five cameras. Each zoom center "travels" nearly in a straight line, starting from upper right for low zoom, to lower left for high zoom. For each camera, the estimates for image center are tightly clustered along a 10 pixel long curve over most of the zoom range, as shown in detail for one of the cameras in Figure 2c. Therefore, a fairly accurate single estimate of zoom center for each camera could be computed as the median of the C_x and C_y components.

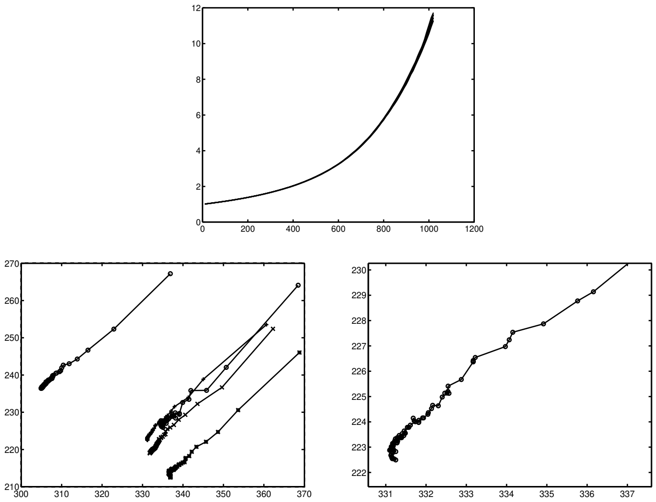

Figure 2. Calibration by zooming. Top: magnification vs. zoom for five cameras. Bottom left: image center vs. zoom. Bottom right: Detail of image center for one camera.

3.3. Calibration via Rotation

Calibration by zooming determines the values of parameters that vary with respect to zoom setting. In this section, calibration by rotating the camera is used to determine the values of the remaining intrinsic parameters (f, s_x, κ) and the camera mount orientation R_m. The most observable effect of the camera mount matrix is to cause a slight roll (rotation about the optic axis) in the image. To simplify the optimization procedure, we therefore reduce the three degrees of freedom of the camera mount orientation matrix R_m to a single roll angle ρ represented by the rotation matrix R_ρ.

Referring to the intrinsic camera equations, we define a nonlinear transformation P that projects camera-centered coordinates into pixel coordinates, and an inverse transformation Q that maps pixel coordinates into camera-centered coordinates:

$$P(X_c, Y_c, Z_c) \mapsto (X_f, Y_f)$$
$$Q(X_f, Y_f) \mapsto (X_c, Y_c, Z_c)$$

where in order for Q to be a well-defined transformation we impose a constraint like $(X_c^2 + Y_c^2 + Z_c^2) = 1$. Now consider the relationship between an image I_1 taken at pan angle ϕ_1 and tilt angle θ_1, and a second image I_2 taken after rotating the camera to pan angle ϕ_2 and tilt angle θ_2. Referring to the extrinsic camera equation (1), we can now write the transformation G that maps a pixel in I_1 into its predicted

Figure 3. Calibration by rotation. Left: Mosaic created from a pair of images using the intrinsic parameters found from active rotation calibration. Middle: Magnified portion of the mosaic where pixels from the two images were averaged. Right: Pixel bias vs. radial distance of active pan,tilt towards a user-selected feature.

location in I_2, and the inverse H that maps from I_2 to I_1:

$$\begin{aligned} G &\equiv P(R_\rho R_{\theta_2} R_{\phi_2} R_{\phi_1}^T R_{\theta_1}^T R_\rho^T Q(X_{f_1}, Y_{f_1})) \\ H &\equiv P(R_\rho R_{\theta_1} R_{\phi_1} R_{\phi_2}^T R_{\theta_2}^T R_\rho^T Q(X_{f_2}, Y_{f_2})) \,. \end{aligned}$$

In the absence of radial distortion, transformations G and H would be simple 2D projective transformations or *homographies*.

The calibration procedure consists of taking several pairs of images related by known rotations, and performing a nonlinear search over the space of parameters (f, s_x, κ, ρ) in order to minimize the sum of the SSD errors from Eq. 4 over all pairs simultaneously. Figure 3 shows sample results for one of the cameras. Nine pairs of images were used, composed of all combinations of images with pan angles of $\{-30, 0, 30\}$ and tilt angles of $\{-24, -20, -16\}$, and all at zoom setting 0. The best set of parameters found were used to create the two-image mosaic in Figure 3a. Pixels in the overlap between the two unwarped images were "blended" by taking the average of their intensity values, therefore any misalignments will show up as a blurring of structures in the image. Figure 3b shows a magnified subimage taken from the area of overlap. There is no apparent scene blurring – thin image structures such as the painted parking lines and sign post still appear sharp.

Another way to test the results of intrinsic calibration is to measure camera pointing accuracy, using a cross-hair drawn in the center of the image. The user selects the pixel coordinates of distinctive image features with a mouse, and a pan and tilt angle are computed that ideally will align the crosshair with that image feature (this involves mapping pixel coordinates to scene coordinates and back, using the intrinsic projection equations). After performing that camera rotation, the misalignment between the image feature and the crosshair is measured. Figure 3c shows three curves fit to data collected on pixel errors vs. distance of the target feature from the image center. These curves also illustrate the effects of removing various intrinsic parameters from the model. The solid curve is based on using only f and s_x model terms – pointing errors of up to 9 pixels occur near the edge of the image. Adding roll angle ρ improves the performance, as shown by the dashed curve, to roughly 3 pixels at the image edge. Finally, correcting for radial distortion by adding parameter κ results in a 2 pixel bias at the edges of the image (dot-dash curve).

4. Active Extrinsic Calibration

Extrinsic calibration involves solving for the location T and orientation R of the camera with respect to some Euclidean scene coordinate system, a process also known as *pose determination* [3, 7]. Camera pose is typically determined from a monocular view by finding an R and T that bring a set of projected 3D scene features into the best alignment with extracted 2D image features. Fully automated landmark-based pose determination is nearly impossible unless a good initial pose estimate is already known, due to the difficulty in determining the correspondence between 3D scene landmarks and extracted image features [7].

We sidestep such difficulties by manually determining the correspondence between a sparse set of 3D landmark points and viewing rays through the camera focal point. Rather than infer viewing rays from image pixel coordinates, our approach measures viewing orientations directly by actively panning and tilting the camera towards each landmark until its image projection precisely aligns with a crosshair at the median image center (C_x, C_y) computed during intrinsic calibration. The pan angle ϕ_i and tilt angle θ_i are noted for each visible landmark, yielding a set of viewing ray unit vectors $u_i = (\sin\phi_i \cos\theta_i, \sin\theta_i, \cos\phi_i \cos\theta_i)$ in the pan-tilt head coordinate system

Since viewing rays are determined by active camera rotation, measurements can be recorded over an extended hemispherical field of view. We develop an error metric based on comparing the angle between pan-tilt viewing rays and direction vectors from the camera to 3D landmark points, and search for the pose that brings these two sets of unit vectors into best alignment.

First consider the case where we already know the camera location T (say by prior GPS measurement), and we only need to estimate its orientation R. Each 3D landmark sighting P_i yields two unit vectors, the viewing ray u_i as above, and a corresponding scene direction vector $n_i = (P_i - T)/||P_i - T||$ directed from the camera center T to the landmark point. If there were no noise in the measurements, these two vectors whould be related by camera orientation R as $n_i = Ru_i$. In actuality, the 3D landmark coordinates and the pan and tilt angles all contain measurement errors. Following Horn [5], we solve for the rotation R that best aligns these two sets of unit vectors (u_i, n_i) in a least squares sense by maximizing

$$E = \sum(n_i \cdot Ru_i) = \sum(\frac{(P_i - T)}{||P_i - T||} \cdot R\, u_i) \qquad (7)$$

Using an intermediate unit quaternion representation, the rotation R^* that maximizes E can be computed in closed-form [5].

Now consider solving for full pose by maximizing (7) with respect to both T and R. As in earlier sections, we employ Powell's method. More specifically, we embed Horn's closed-form solution for R inside a Powell coordinate descent search on the three coordinates (T_x, T_y, T_z). For each tested value of T the closed-form solution for R is computed, and error function (7) is evaluated for that T and R. Experiments show that this hybrid Powell-Horn approach is remarkably insensitive to the initial estimate of T.

Experiments

A GPS survey was performed on five fixed-mount camera locations and two dozen landmark points located around the camera stations (see Figure 4). Measurements were taken with a Novatel RT-2 GPS receiver in communication with a similar Novatel base station located on site. These units provide dual carrier phase differential readings with roughly $2cm$ level accuracy.

For each camera, pan-tilt measurements were made of the landmark points visible to it, and the pose was computed using the hybrid Powell-Horn optimization procedure. Accuracy of the resulting pose estimate for each camera is summarized in the following table:

camera	# landmarks	dist err (m)	ang err (deg)
A	7	2.0	1.2
B	6	0.4	0.4
C	12	1.0	0.5
U	7	0.5	0.1
V	9	1.4	0.3

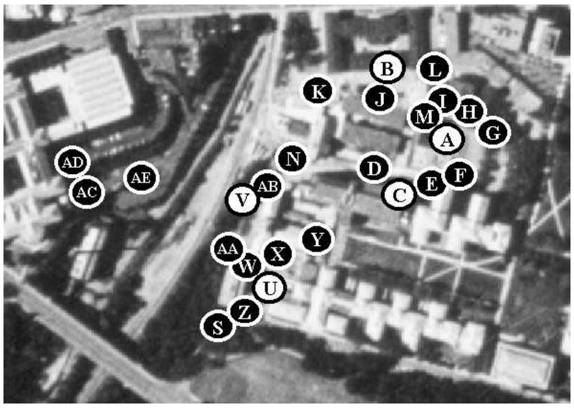

Figure 4. Camera locations (white dots) and landmark locations (black) surveyed by GPS for pose estimation.

where distance error is the difference between the computed location and the location as measured by GPS, and angular error is the mean angle between corresponding unit vectors n_i and Ru_i.

The hybrid Powell-Horn pose estimation process is very stable when used with landmarks spanning a virtual hemisphere. T does not need to be initialized near the actual camera location, within the convex hull of the landmark points, or even close to the site, in order to converge to an accurate pose estimate. Figure 5 illustrates this behavior for one of the cameras. Seven landmark points were sighted, yielding a set of pan angles spanning a total range of 266 degrees, and a set of tilt angles spanning a range of 37 degrees. Initial estimates of T were generated every 200 meters on a $60X60$ kilometer grid centered at the "ground-truth" camera location measured by GPS. For each initial position, camera pose was recovered using the Powell-Horn method, and the location component was compared to the ground-truth camera location. Each black grid cell represents an initial location from which the pose algorithm converged to within 2 meters of the ground-truth camera position. The entire set of landmark points is contained within a single grid cell in the center of the image (i.e. within a 200 X 200 meter area). The average radius of the convergence region is roughly 21 kilometers. Similar convergence properties were observed for the other cameras.

5. Applications and Future Work

We have developed a parameteric model for an active camera system with pan, tilt and zoom control. We have also incorporated active camera control into novel calibration methods for determining the intrinsic and extrinsic parameters of the model. This work was motivated by the

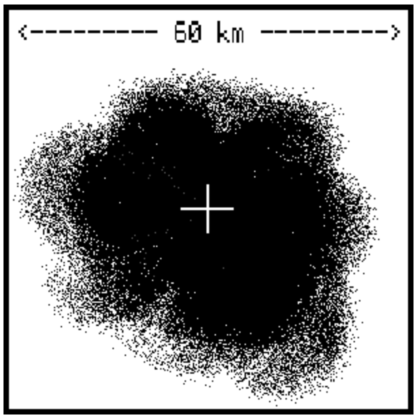

Figure 5. Black pixels mark initial location estimates from which pose determination converged to within 2 meters of the ground-truth camera location measured by GPS.

need to calibrate *in-situ* a network of outdoor cameras for video surveillance applications. Accurate camera calibration was crucial to performing several image understanding tasks involved in video surveillance: actively tracking moving objects while rotating and zooming the camera, building image mosaics for operator visualization, pointing the cameras at known scene landmarks such as doorways, and estimating 3D object locations by intersecting viewing rays with a terrain model.

Further experiments are needed to compare the accuracy of flow-based intrinsic calibration methods to traditional methods using precise calibration grids in a controlled environment. The level of accuracy achieved is clearly good enough for performing outdoor surveillance tasks, but the limits on accuracy need to be established. In the current extrinsic calibration framework, each camera is calibrated separately, even though many cameras can see overlapping sets of scene landmarks. In future work we will perform simultaneous calibration of all sensors using a bundle adjustment procedure to perform least-squares refinement of all sensor poses and landmark locations. Derivation of uncertainty bounds on computed pose is also future work.

Acknowledgements

We wish to thank our fellow VSAM team members: Alan Lipton, Dave Duggins, Hironobu Fujiyoshi, David Tolliver, Raju Patil and Yun-Ching Lee, for motivating this work and providing a cool working environment. Additional thanks go to Dave Duggins for installing the cameras and performing the GPS survey of landmark positions, and to Takeo Kanade for his mentorship.

References

[1] A.Basu and K.Ravi, "Active Camera Calibration Using Pan, Tilt And Roll," *IEEE Trans SMC,* Vol.B-27(3), June 1997, pp. 559-566.

[2] J.Bergen et.al., "Hierarchical Model-Based Motion Estimation," *ECCV,* 1992, pp. 237-252.

[3] R.M.Haralick *et.al.*, "Pose Estimation from Corresponding Point Data," *IEEE Trans SMC,* Vol. 19(6), Nov 1989, pp. 1426-1446.

[4] R.I.Hartley, "Self-Calibration from Multiple Views with a Rotating Camera," *ECCV*, 1994, pp.471-478.

[5] B.K.P.Horn, "Closed Form Solutions of Absolute Orientation Using Unit Quaternions," *JOSA-A,* Vol. 4(4), April 1987, pp. 629-642.

[6] T.Kanade et.al., "A Stereo Machine for Video-Rate Dense Depth Mapping and Its New Applications," *CVPR*, 1996, pp.196-202.

[7] R.Kumar and A.R.Hanson, "Robust Methods for Estimating Pose and a Sensitivity Analysis," *CVGIP,* Vol. 60(3), Nov. 1994, pp. 313-342.

[8] M.X.Li and J.M.Lavest, "Some Aspects of Zoom Lens Camera Calibration," *IEEE Trans. PAMI,* Vol.18(11), November 1996, pp. 1110-1114

[9] W.H.Press et.al., *Numerical Recipes in C,* Cambridge Univ Press, New York, 2nd edition, 1992.

[10] G.P.Stein, "Accurate Internal Camera Calibration Using Rotation, with Analysis of Sources of Error," *ICCV*, 1995, pp.230-236.

[11] D.Stevenson and M.M.Fleck, "Robot Aerobics: Four Easy Steps to a More Flexible Calibration," *ICCV*, 1995, pp.34-39.

[12] R.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,* Vol. RA-3, No. 4, August 1987, pp. 323-344.

[13] R.G.Willson, *Modeling and Calibration of Automated Zoom Lenses,* Ph.D. Thesis, Carnegie Mellon University, CMU-RI-TR-94-03, 1994.

[14] R.G.Willson and S.A.Shafer, "What is the Center of the Image?," *JOSA-A,* Vol.11(11), November 1994, pp. 2946-2955.