Camera Calibration and Optimization for Indoor Location

Soren Bjornstad Nathan Maveus TianYu Pang

February 9, 2016

Abstract

There are many applications in which it is useful to know the position of a device in space with reasonable accuracy – for instance, navigation, opening apps or displaying information on a mobile device when it is in a particular area, or displaying targeted advertising. GPS is a commonly used technology for location, but it does not work effectively indoors, and many of the most useful applications of location services are indoors. We propose an alternative to existing indoor positioning systems based on creating a 3D model from images of a building, then matching feature points in an image taken at the time the position fix is needed to known locations in that 3D model.

To begin, we created a 3D model of part of the St. Olaf College science building. We took stereoscopic pairs of images of the building and of calibration images using DSLR cameras and a tripod. Then we used camera calibration techniques to determine the parameters of the camera (focal length, translation of the principal point, distortion, etc.). We then estimated the positions at which the images were taken and their rotations. Finally, we located corners on the images and found corresponding coordinates for those points in the real 3D space using the building's blueprint. By projecting the points between the two spaces, we were able to optimize our model using numerical optimization techniques with some code written at St. Olaf for this purpose, based on the Google Ceres framework.

With the model created, it is possible to determine the camera location for an image taken in the space by finding corners present in the model in that image, then matching them with the corresponding point in the world coordinate system. This technique appears to have the potential to offer higher precision than most existing IPS systems without requiring the installation of complex and expensive equipment.

1 Introduction

There are many applications in which it is useful to know the position of a device in space with reasonable accuracy – for instance, navigation, opening apps or displaying information on a mobile device when it is in a particular area, or displaying targeted advertising. GPS is a commonly used technology for location, but it does not work effectively indoors, and many of the most useful applications of location services are indoors. We propose an alternative to existing indoor positioning systems based on creating a 3D model from images and a blueprint of a building, then matching feature points in an image taken at the time the position fix is needed to known locations in that 3D model.

2 Background

2.1 GPS and IPS

GPS (Global Positioning System) is a collection of satellites that orbit around the Earth, providing estimates of a users current location. These satellites broadcast information from space to the users GPS device, and using four of these satellites, the users position can be triangulated.

In total, thirty-one satellites have been launched by the United States and the Air Force to ensure that GPS is available at least 95% of the time [8]. However, the accuracy of GPS may not be what is required for some users. Various services called augmentation systems have also been developed to improve accuracy and availability with regard to positioning, navigation, and timing [1]. Together, these systems and GPS can give information about a user's position with remarkable accuracy.

Unfortunately, GPS has a difficult time tracking users that are indoors for several reasons, including that metal completely blocks the signal and that it is difficult to have a clear line of sight between a users position and the satellites [5]. Therefore, IPS (Indoor Positioning Systems) have been developed. Instead of relying on satellites, these systems use a variety of other inputs, including Wi-Fi fingerprints (looking up nearby Wi-Fi routers in a database of positions) [6], device accelerometers, altimeters, camera images, Bluetooth, and other sources of location and motion information. Often several are integrated into one system for greater precision.

2.2 Camera Calibration

Existing IPS systems based on radio waves, infrared light, or ultrasound either have relatively low accuracy or require special, expensive infrastructure. Camera calibration offers an alternative approach to finding a precise location inside a building.

Camera calibration is a mathematical technique used to determine the extent to which cameras distort and transform 3D space as they create 2D images, and to condense these distortions to a set of camera parameters. Given the camera parameters, and knowing the cameras position, it is possible to accurately project any point in 3D space onto a point on the 2D image and vice versa. The precision can often reach sub-pixel accuracy on the image.

Many techniques have been developed for camera calibration. Sturm and Maybank [10] as well as Zhang [11] take pictures of a special planar surface, or use grid-like planar surfaces. Others have used circles or spheres in images [3]. Generally, once known points have been found, the process of camera calibration goes like this: 1) using a camera or a pair of cameras, take multiple photographs (with different views) of a target; 2) distinguish different contours and locate feature points on the target in each image; 3) providing a parameterized model for the camera, estimate parameters to minimize discrepancy between the model and the observed locations of feature points.

This equation describes the mapping from 3D to 2D:

$$s\left[\begin{smallmatrix} u \\ v \\ 1 \end{smallmatrix}\right] = \left(\begin{smallmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{smallmatrix}\right)[R\ t]\left[\begin{smallmatrix} X \\ Y \\ Z \\ 1 \end{smallmatrix}\right].$$

In this mapping, there is a 3×3 upper triangular matrix called the *intrinsic parameter matrix*, a set of numbers that describe the camera itself. α and β represent the focal lengths, (u_0, v_0) represents the optical center (or principal point), and γ is the skewness of the image. (We set skew to zero for simplicity, as Zhang does – with most cameras the skew is so close to zero that it is not worth worrying about.) The are also the *extrinsic parameters* R and t, which represent the rotation and translation of the camera in the 3D world (that is, how the camera is placed and oriented in space). The two vectors $\left[\begin{smallmatrix} u \\ v \\ 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} X \\ Y \\ Z \\ 1 \end{smallmatrix}\right]$ are homogeneous coordinates of the point we wish to map in the image and 3D space respectively.

This model assumes that the lens produces no distortion. This is not a good assumption with real lenses, so it is necessary to correct for the distortion. There are two major kinds of distortion: tangential distortion and radial distortion. Varying the alignment of two optical elements produces tangential distortion. However, for most cameras today, tangential distortion is not a significant problem any more and can be ignored. Radial distortion occurs when light rays bend more near the edges of a lens than they do at its optical center. The *Brown-Conrady model*

[2] can be used to correct radial distortion. This model fits a power series to distortion. In practice, this infinite series is usually simplified to a short finite polynomial, as the amount of correction contributed by each term drops off rapidly:

$$\widetilde{x} = x + x[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$
$$\widetilde{y} = y + y[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$

In these equations, (x, y) and $(\widetilde{x}, \widetilde{y})$ are the ideal (distortion-free) and real (distorted) normalized image coordinates, and k_1, k_2 are the coefficients of the radial distortion.

By solving the homographies of the mapping, one can obtain the camera parameters and use these parameters to estimate camera positions. See *Zhang's paper* [11] for details about the computation of camera calibration.

We propose that projection of known points in 3D space through a camera with known parameters, such as a cell phone or other portable camera that is calibrated in advance, can be used to precisely locate the position of that camera for purposes of indoor location and navigation. Unlike existing IPS methods, this method does not require any infrastructure, can locate positions very accurately, and requires only a pair of cameras and a copy of the buildings blueprint.

3 Method

3.1 Obtaining Data

Our data set consists of parameters of two cameras that we use, parameters from image pairs taken at various angles throughout the second floor of the St. Olaf College science building, and coordinates of feature points on these images in both a world coordinate system and the image coordinate system.

The camera parameters of the left and right cameras include the translation of each camera from the tripod position, rotation of the camera from the tripod position, the focal length and principal point of the cameras, and the distortion factors of the camera lens. We got our camera parameters from a previous group of researchers who took the photos, including special calibration images. These cameras were mounted on a metal bar attached to a tripod to ensure that they remained at the same angle and distance with respect to each other throughout the data collection. For convenience, we considered the tripod position to be at the left camera and so fixed the left camera's translation and rotation at 0.

Parameters for each image pair are also necessary, including the translation of the tripod for each image pair and the rotation of the tripod from a fixed orientation in the world coordinate system. Rough estimates were made of the tripod location and the orientation of the cameras for each image pair. The positions can be estimated by measuring the rough position of the foreground of an image on the building's blueprint, or by using a tape measure and dead reckoning from the previous camera position as the pictures are being taken. On the other hand, the rotations can be estimated by looking at the image and its relation to the world coordinate system and taking an educated guess. We use Euler angles to represent the rotation, by estimating how much each camera rotates around the x, y, and z axes respectively.

Finally, we need to know the correspondences of some specific points between 2D coordinates and 3D coordinates in order to set up relations between them. Once the images were taken, we used *eriol*, a 3D modeling program developed at St. Olaf, to create contours, feature points, and corners on the image pairs, and to merge tiles and corners present in multiple images together. Working in eriol, we placed three feature points at the intersection of the floor and a wall on each image and had eriol record the 2D coordinates in pixels of these feature points. (We chose positions on the floor so that we would not have to measure their height, which is more difficult to find on a blueprint.) To find the matching world coordinates of the point in 3D, we measured the distance between the origin and that same point on the blueprint using Photoshop (this would be a useful thing to automate in the future). This gave us, for each image, three points in the world coordinate system mapped to their corresponding points in the image coordinate system.

3.2 Optimization

Once we had estimates of the camera parameters, we used the *Google Ceres* [9] framework for numerical optimization to minimize the error in our camera position and orientation estimates. We used code written by others at St. Olaf as a cost function to project the points. To increase the accuracy of the estimates, we chose to hold the positions of the corners both in 3D space and in the 2D images constant (the 3D positions were obtained from the blueprint, the 2D positions from user selections in eriol). We decided we did not have enough data to optimize the positions of the camera parameters further than the previous researchers had, so we held those constant as well. In contrast, we had the optimizer vary the position and rotation of the cameras on the world coordinate system. Also, we had to account for a small error in our data collection: we found that the rubber pad underneath the cameras would flex slightly when the shutter button was pushed, which meant that a small amount of unpredictable rotation and dilation was added to the images. To counteract this, we additionally created and optimized parameters for additional rotation of the right camera and additional scale factor of the right camera.

4 Results

4.1 Accuracy of GPS

According to data submitted to the Federal Aviation Association, 95% of measurements taken by a good GPS system outdoors will be within 3.4 meters horizontally and 4.7 meters vertically of the actual value [4]. However, we know that GPS is far less accurate indoors. To confirm the inaccuracy of indoor GPS for ourselves in our own space of interest, one of the authors obtained the *Easy GPS* app on an iPhone 5 and walked around the building with it, pacing off 10 steps (approximately 30 feet) at a time. In addition to giving a reading on one's current latitude and longitude, the app gives a margin of error estimate; unfortunately, we could not find specific details on how this number is calculated. Results in the entire science building ranged from a low of ± 21 ft to a high of over ± 300 ft; on the second floor, which we used exclusively for our tests, the margin of error was universally ± 123 ft.

4.2 Accuracy of IPS

We do not have any kind of existing IPS system in our building, so we were unable to test its accuracy here. In terms of existing systems, the Korean Institute of Communications Information Sciences studied different means of calculating distances with IPS systems, and found that accuracy was typically between 2.9 and 4.9 meters.[7] This is far better than GPS, but the error involved can still be larger than a small room, which is less than ideal.

4.3 Accuracy of Our Method

To test the accuracy of our method, we used 20 pairs of images (40 images total) taken at a number of tripod positions in a hallway on the second floor of the building. We found feature points and estimated other parameters for this data as described above.

There are several sources of error in this process. One preventable factor is that the cameras were set on auto-focus, which was intended to improve the quality of the images but likely changes the camera parameters and may introduce error; unfortunately, we have no quantitative information on the size of this effect. Another possible source of error is the original measurements of where the camera tripods were positioned. The authors were not involved in these measurements and do not know how they were made, although we know they were supposed to be measured to the nearest tenth of a meter, which does not seem to have happened looking at the average changes in position after optimization. If measurements were on the sloppy side of that tolerance and dead reckoning was used, it is possible that the accumulation of error could have exceeded the tolerance of the optimization process and caused incorrect or less accurate results.

A third source of error, which cannot be prevented entirely, is human precision in placing feature points. To see how much error might be created by poor placement of feature points, we conducted two tests in which we deliberately moved feature points 1, 10 and 50 pixels off the correct points in both the X and Y directions. The translation error in the final X, Y, and Z values created by these movements was then measured; measurements are shown in the table below.

Movement	**X (m)**	**Y (m)**	**Z (m)**
1 px – X	0.016	0.013	0.029
1 px – Y	0.005	0.01	0.01
10 px – X	0.15	0.13	0.30
10 px – Y	0.06	0.11	0.12
50 px – X	0.78	0.67	1.49
50 px – Y	0.28	0.53	0.57

This data shows that positioning errors can cause a significant drop in the accuracy of these positioning methods, so it is worth being careful about where feature points are placed. However, our experience doing 3D modeling at St. Olaf has shown that 1-pixel error in feature point placement is usually quite achievable, which means the error created by this factor should be no more than a couple of centimeters.

A final, related source of error is the precision with which the 3D positions of the feature points are measured. We obtained our positions by measuring a digital copy of the blueprints in Photoshop, so the accuracy is determined by the resolution of the image. In our version of the blueprint, there were 50 pixels per foot of the building, so we could expect measurement errors of ± 0.01 ft.

5 Conclusion

While we have not been able to fully account for all sources of error and we do not have a way to precisely measure the positions of the cameras in images that were previously taken to directly compare the accuracy of our technique, this technique looks able to offer high precision. Distortion of the camera parameters can be avoided by using a fixed focus and a larger number of images, and more careful data collection and estimates will likely eliminate some of the other issues we saw. The possibility of errors in finding coordinates remains. However, even with a 50-pixel error in feature point placement, which should be far larger than we ever see in actual 3D models, we saw errors considerably smaller than those of existing IPS systems – a small fraction of a meter compared to several meters. While this technique is not yet developed enough to be confidently used in an actual positioning system, we feel it looks promising and deserves further research.

References

[1] "Augmentation Systems." GPS.gov, National Coordination Office for Space-Based Positioning, Navigation, and Timing, n.d., Web., 28 Jan 2016.

[2] Brown, Duane C., "Decentering distortion of lenses", *Photogrammetric Engineering*, 32 (3): 444462, May, 1966

[3] Dinh, H., and Tang, H., *"Simple method for camera calibration of roundabout traffic scenes using a single circle"*, IET Intelligent Transport Systems, 8(3), 175., 2014.

[4] *Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Analysis Report* (n.d.): n. pag. *Federal Aviation Association.* Faa.gov, 31 July 2014. Web.

[5] "Introduction to GPS and Its Use in Geology Labs", n. pag. Colorado State University, Colorado State University. Web., n.d.

[6] Leedham, Graham, and Dan Patterson. *Expert Systems with Applications*. 23rd ed. Vol. 42. N.p.: Elsevier, 1997. 9263-278. Print.

[7] Park, Namjun, Sukhoon Jung, and Dongsoo Han. "An Adaptive Hybrid Filter for Practical WiFi-based Positioning Systems." *ICT Express (2015)*: n. pag. Web.

[8] "Space Segment." GPS.gov, National Coordination Office for Space-Based Positioning, Navigation, and Timing, n.d., Web., 28 Jan 2016.

[9] Sameer Agarwal and Keir Mierle and Others, "Ceres Solver", `http://ceres-solver.org`.

[10] Sturm, P.F. and Maybank, Stephen J, "On plane-based camera calibration: a general algorithm, singularities, application", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 432-437, 1999.

[11] Zhang, Zhengyou, "A flexible new technique for camera calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):13301334, 2000.