

Finding Jane Doe: A Forensic Application of 2D Image Calibration

Abby Stylianou, Austin Abrams and Robert Pless

Washington University in St. Louis, USA
{astylianou,abramsa,pless}@wustl.edu

Keywords: camera calibration, photogrammetry, geolocation, forensic, case study

Abstract

Thirty years ago, a young girl was found decapitated. Her identity remains unknown, and neither her head nor her killer have been found. Until recently, the location of her grave was lost, preventing any efforts to identify her using modern forensic techniques. This paper presents a case study on the use of burial photos to accurately and precisely determine the location of the lost grave. We highlight challenges in finding good correspondences between points in thirty year old imagery and the actual scene today, characterize the sensitivity of a camera calibration's geometric constraints to inaccuracies in these correspondences, and discuss an interactive tool that allows an analyst to quickly identify problems with them. Using the geolocation pipeline and tools discussed in this paper, we localized the lost grave to an approximately 1.6 meter long uncertainty region. On June 16, 2013, the Saint Louis Police Department performed an exhumation according to this localization, quickly finding the body.

1 Introduction

Solving for the location from which a camera captured a photograph is a classic problem in photogrammetry. It has proven a useful task for anthropologists and historians, who use websites like WhatWasThere.com to understand how man-made scenes appeared in the past, and for climate scientists seeking to create repeat photography records of glacial change [10]. This problem has even recently been turned into games on the internet, including blogger Andrew Sullivan's View from Your Window contest [13] and the popular website GeoGuessr [1].

This paper offers a case study on a specific, successful effort to geolocate a lost grave from thirty year old images. Figure 1 shows a photograph from the burial and a modern image from approximately the same place. The largest challenge in this particular case was the dramatic change of the scene over time. Finding reliable matching features between the original images and the scene today was difficult, and yet geo-location accuracy needed to be sufficient to direct digging to find the grave.

Through this case study, we offer three specific contributions:



Figure 1. An image from the 1983 burial of a Jane Doe, and a comparable view from 2013. Note the amount of dense growth in the background of the scene, as well as the large tree in the foreground that was not present in 1983. Significant changes over the last 30 years made selecting correspondences between the original images and the current scene difficult.

- the description of an interactive system that provides an analyst with feedback on the geometric consistency of a set of correspondences between an image and an aerial view of the scene,
- a characterization of the sensitivity of geolocation results as a function of these correspondences,
- a generalization of lessons learned that can be applicable to other image localization applications.

1.1 Case Details

In 1983, an approximately eight-year-old girl was found decapitated in Saint Louis, Missouri. After no family came for-

ward to identify the girl and neither a suspect nor the girl’s head was found, the Saint Louis Police Department (STLPD) paid for an indigent burial. A photographer for the Saint Louis Globe Democrat captured several images of the burial for a story about the case [2]. A year after the burial, a group of college students raised funds to place a headstone at the gravesite.

In 2010, the STLPD attempted to exhume her body, hoping that new forensic techniques might help identify the girl or her killer. They dug at and around the headstone, exhuming three bodies in the process — all with heads. Based on this, and controversy about the poor record keeping at the cemetery, the Saint Louis Medical Examiner asserted that the headstone location had little relation to the actual grave, and banned any further attempts to exhume the body until further proof of the grave location was provided.

In 2013, for the thirty year anniversary of the Jane Doe’s death, the Saint Louis Post-Dispatch printed a story on the case and the lost grave, including two of the images taken by the Globe Democrat photographer [6]. The following sections will detail the geolocation pipeline we used to determine the precise location of the lost grave from these images.

1.2 Overview

We solve for the position of the grave by finding the exact position of the camera that captured one of the pictures during the burial. This is a photogrammetry (or camera calibration) problem based on matching locations in the image to notable points in the world. These matches allow one to solve for the position, orientation and zoom of a hypothetical camera. Our optimization, described at length in Section (2) minimizes the difference between the the points in the original image and real world points projected into image space by this hypothetical camera.

We then share the results of our initial attempts to locate the camera, which suffered from problems including bad matches and GPS inaccuracy, and our final attempt which was used to successfully exhume the body of the young Jane Doe.

1.3 Related Work

The problem of localizing a camera relative to a collection of points in a scene is not a new one in photogrammetry [14, 5, 4]. While significant progress has been made recently in the Computer Vision community to automatically recognize scenes and estimate their location, methods that determine accurate and precise camera locations, as opposed to more general localizations [15, 16], require either large networks of images where camera locations can be determined through feature matching between the non-geolocated images and images with known geolocation [12] or a user-in-the-loop referencing publicly available 3D geometry of a scene [3].

2 Image Calibration

Camera calibration is a method to optimize for the position of a camera given some information about the projective geometry

in the scene. Given a set of real-world coordinates (X, Y, Z) and the corresponding locations (x, y) on the image, camera calibration recovers the 3D position, orientation, and zoom level of the camera most consistent with that set of correspondences. Here, we discuss camera calibration with known 3D world-coordinates, and then discuss a special case of camera calibration that only requires 2D world-coordinates (i.e. latitude and longitude).

2.1 Notation

We assume that we have many correspondences $(X_i, Y_i, Z_i) \rightarrow (x_i, y_i)$. The goal is to recover the 3D location of the camera (L_X, L_Y, L_Z) most consistent with those correspondences. In the process, we will also optimize over the 3D orientation—expressed as a (pan, tilt, roll) = (p, t, r) triple in radians—and the focal length, f , of the camera¹.

2.2 3D calibration

Given the location, orientation, and focal length of the camera, a 3D point (X, Y, Z) projects into the image at (x, y) as determined by the following linear equation (see [8] for more details):

$$\begin{bmatrix} x\omega \\ y\omega \\ \omega \end{bmatrix} = \begin{bmatrix} f & 0 & \frac{w}{2} \\ 0 & f & \frac{h}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R(p, t, r) & \begin{bmatrix} L_X \\ L_Y \\ L_Z \end{bmatrix} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

where $R(p, t, r)$ is a 3×3 rotation matrix equivalent to panning, tilting, and rolling by the specified amounts, w and h are the width and height of the image (in pixels), and ω is a scaling factor. To recover the final (x, y) coordinate, we divide the left hand side by the third element:

$$\text{proj} \left(\begin{bmatrix} x\omega \\ y\omega \\ \omega \end{bmatrix} \right) = \begin{bmatrix} x \\ y \end{bmatrix} \quad (2)$$

To simplify notation, we rewrite the right hand side of Equation (1) as a 3×4 matrix M_{3D} that depends on the unknown parameters L, p, t, r , and f . Camera calibration performs an optimization over these unknowns so that this *reprojection error* over all n correspondences is as small as possible:

$$\underset{L, p, t, r, f}{\text{argmin}} \sum_{i=1}^n \left\| \text{proj} \left(M_{3D}(L, p, t, r, f) \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \right) - \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|^2 \quad (3)$$

Determining the world-coordinates (X, Y, Z) for a given (x, y) point in the image remains challenging. Other work in

¹The focal length, expressed in pixels, is a measure of the zoom level of the camera. A small focal length (say 50 pixels in a 640×480 image) corresponds to a very wide-angle field of view, while a larger focal length (say 2000 pixels) would be used in cameras more zoomed in, like a telephoto lens.

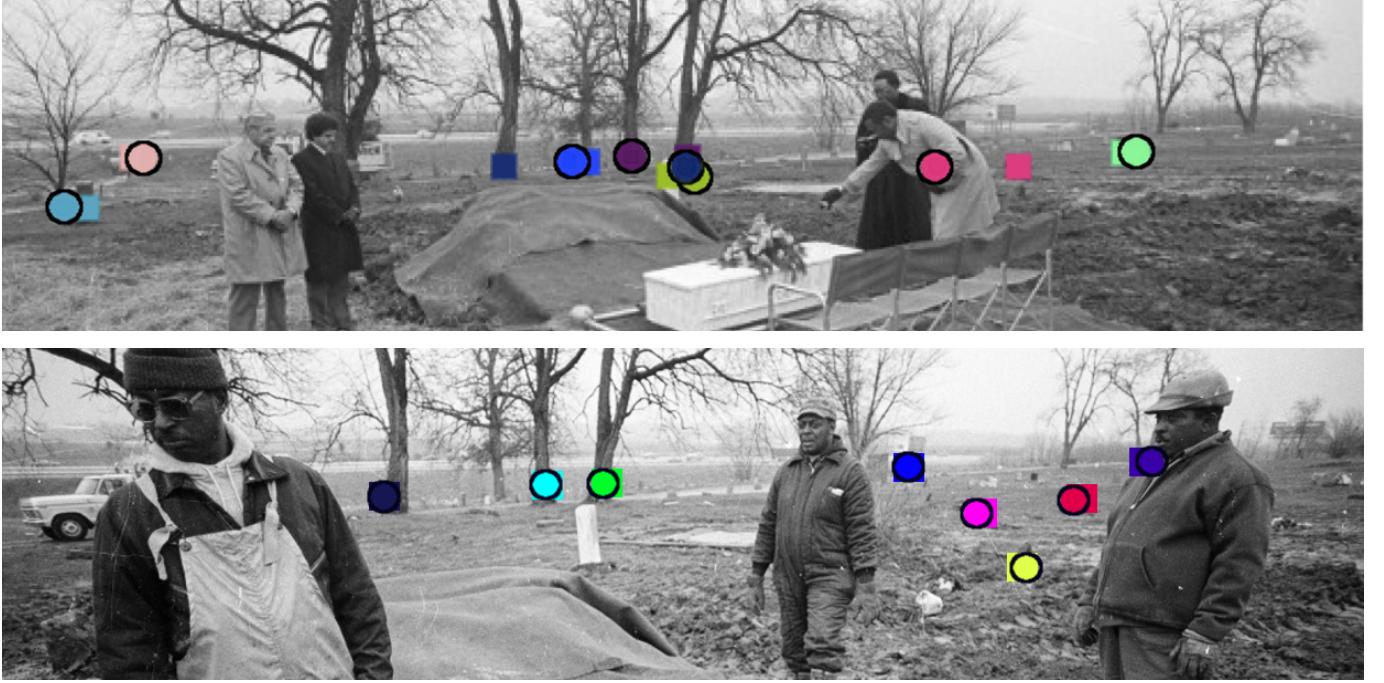


Figure 2. A visualization of the difference between the image point, (x, y) , and the real world GPS coordinate, (X, Y) . The GPS coordinate projected back into the image given a set of camera parameters is displayed with a colored circle, outlined in black. The corresponding image point is a square of the same color. The optimization process finds camera parameters that minimize the error defined in Equation (5). The top image was the visualization shown to an analyst using the interactive tool described in Section (4), after calibrating one photo with GPS data collected from cell phone images. This calibration attempt still included bad correspondences, such as the bright pink correspondences at a headstone that had been moved during efforts to restore the cemetery, and the dark blue correspondences at a tree with particularly inaccurate geolocation. The bottom image was the visualization for the image closest to the grave using the most accurate geolocations available, and after an analyst used the interactive tool to carefully prune bad correspondences.

camera calibration uses checkerboards with known 3D coordinates [14] or 3D geometry from Google Earth [3] to get this information. In our case, with imagery dating back thirty years and no Google Earth geometry for the scene, determining precise 3D locations for each correspondence is difficult.

2.3 2D calibration

However, if we know the latitude and longitude of each correspondence, this provides the coordinates in a 2D space (i.e. the X and Y , but not Z). Assuming the camera has no roll, a similar derivation shows the projective equations still depend on the 2D location of the camera:

$$\begin{bmatrix} x\omega \\ \omega \end{bmatrix} = \begin{bmatrix} f & \frac{w}{2} \\ 0 & 1 \end{bmatrix} \left[R(p) \mid \begin{matrix} L_X \\ L_Y \end{matrix} \right] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (4)$$

Notice that now, the projection only depends on the x coordinate in the image, and that only the rotation up to a pan angle p can be recovered. This is equivalent to calibrating a camera from a 1-dimensional image.

Still, this simplified projection offers an optimization that returns L_X and L_Y , the geo-location of the camera on a ground

plane:

$$\operatorname{argmin}_{L, p, f} \sum_{i=1}^n \left\| \operatorname{proj} \left(M_{2D}(L, p, f) \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} \right) - x_i \right\|^2 \quad (5)$$

where M_{2D} is the 2×3 matrix from Equation (4), and here $\operatorname{proj}([x]) = \frac{x}{y}$.

Therefore, given a set of image location (x, y) and their corresponding geolocations (X, Y) , we can optimize the above objective to find the most consistent camera location L , camera pan angle p , and focal length f .

A visualization of the difference between (x, y) and (X, Y) reprojected back into the image for a single set of parameters can be seen in Figure (2).

2.4 Optimization

Instead of directly using degrees latitude and longitude as (X, Y) , we convert our coordinate space into a Cartesian coordinate frame with axes corresponding to meters East and North of some arbitrary origin point. Although this does not model the curvature of the Earth, we are only working with points in an approximately kilometer square region, so these errors

are insignificant. After calibrating in this coordinate frame, we convert L back into a geographic coordinate frame.

Given a pan angle and focal length, the optimization becomes a system of linear equations over L . To speed up the optimization, we only optimize directly over p and f . At each step, we solve for the best L with linear least squares and report the reprojection error with respect to that L .

The nonlinear optimization will always reach a local minimum, but depending on initialization, may or may not reach the global minimum. Therefore, we evaluate the objective function from Equation (5) at a grid of (p, f) guesses, and optimize from the single parameter choice with the smallest objective. We sample p at 20 degree intervals from 0 to 360 degrees, and sample f every 50 pixels from 50 to 3000 pixels. This grid search gives us a good initial point, and from here, we use the Nelder-Mead simplex method [11] to find the best answer close to the initialization.

2.5 Accounting for Error

After performing this calibration, and projecting (X, Y) back into the image given the hypothetical camera parameters, these points may still not line up perfectly, even if the parameters were based on a ground truth camera. This is both because the GPS coordinates collected in the field have some error associated with the measurement device and also because of uncertainty about precisely which point in the real world corresponds to an exact pixel in an image. Some amount of error is unavoidable given that the scene today does not perfectly match the scene in 1983 (i.e., the trees have grown over thirty years). The calibration can also be corrupted due to the inclusion of completely erroneous correspondences, as discussed in Section (3).

To evaluate sensitivity to noise, we distorted each point by up to ϵ meters in a random direction. We performed 1,000 trials, each time generating a unique set of distorted GPS coordinates for all of the correspondence pairs. These correspondences were then passed to the 2D calibration described in Section (2.3), producing 1,000 sets of possible camera parameters. We describe our setting of ϵ in Sections (3) and (4).

3 Initial Attempt

In order to identify good correspondences for our calibration, we selected headstones with recognizable shapes, trees that seemed large enough to remain recognizable, and man-made features including a billboard and a highway underpass, from our original pictures. We then found GPS locations for these points based on a combination of Google Maps aerial views, and using smartphones and a low-cost, handheld Garmin at the cemetery to collect GPS measurements.

Using these point locations, we solve for the best fitting camera parameters using the optimization in Section (2.4). In the top image in Figure (2), we overlay circles showing the image location of a correspondence, and a square with the matching color the 3D GPS coordinates projected based on the optimized camera parameters, on top of the original image.

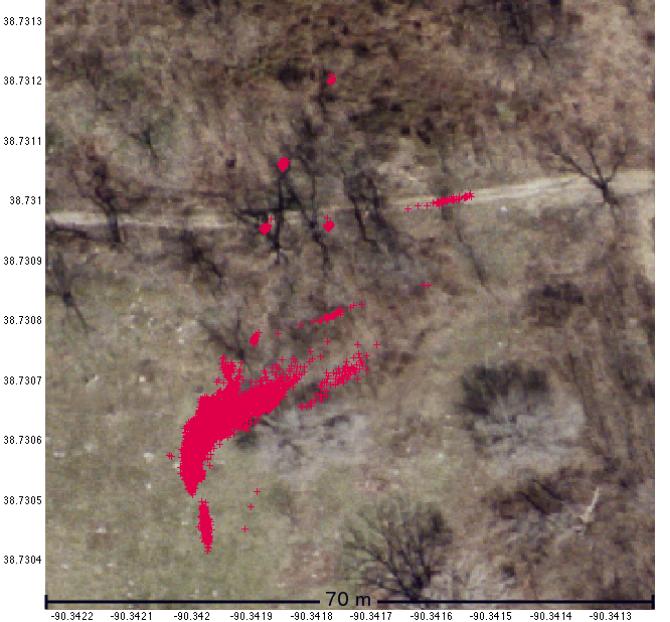


Figure 3. Our initial optimization produced a very large uncertainty region. The pink locations in this figure are the 1,000 possible locations from the optimization and error modeling described in Sections (2.4) and (2.5). The lack of precision in this optimization was the result of bad correspondences and poor GPS measurements.

This initial attempt at geolocating the lost grave using the methods described in Section (2.5) produced 1,000 possible camera locations covering approximately twelve meters across, as seen in Figure (3). While this uncertainty region ultimately did include the correct site of the grave, we knew that it was not sufficiently precise to direct digging by the STLPD.

3.1 Bad Correspondences

The large amounts of change over the last thirty years increase the possibility of selecting bad correspondences. For instance, the billboard that is present today at approximately the same location as the billboard circled in green in Figure (4) is actually newly constructed, approximately 2 meters west of the original. The two billboards have notably different base shapes. The original location was determined because the five steel beams that supported the original billboard seen in the burial images are still in the ground at the cemetery.

The trees circled in red on in Figure (4) were also a problem. These trees were important in estimating a reasonable field of view for the camera, as they were located on the right side of the image where we lacked other good correspondences. Unfortunately, they were unidentifiable on two separate visits to Washington Park Cemetery, as well as in publicly available aerial imagery (from Google Maps, Bing Maps and the USGS Earth Explorer). We were finally able to identify them using aerial imagery from 1981, provided by the St. Louis County GIS department, which showed two billboards, and had only two trees that could appear between these two billboards from

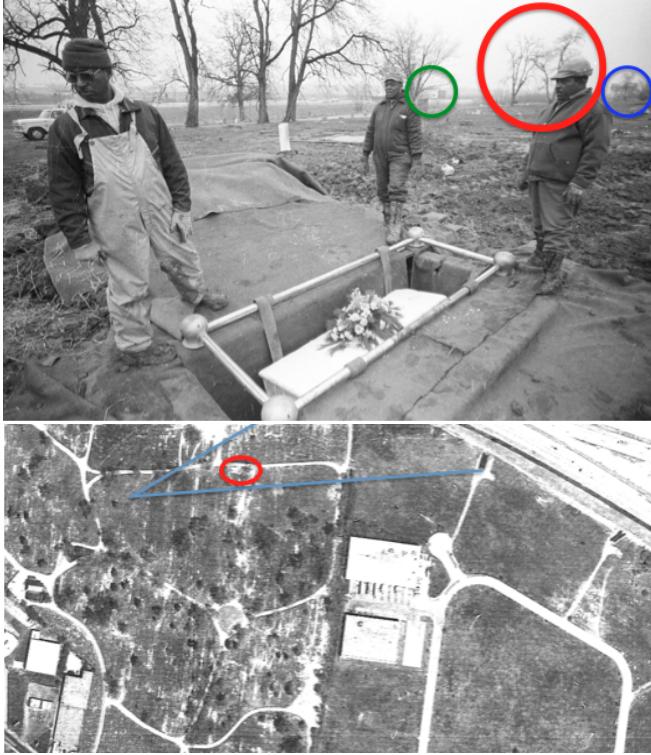


Figure 4. Some of the difficulties in working with thirty year old images are seen here. The billboard circled in green in the top image was rebuilt in the 1990s, while the one circled in blue has been torn down. The trees circled in red were impossible to locate in person and could only be found with the 1980s aerial imagery, shown in the bottom image.

the camera viewpoint.

Even a headstone in the original image that is still present today was moved from its original location by volunteers during attempts to restore the cemetery. We determined this when removing that correspondence from the calibration produced optimization results with significantly lower reprojection error. When we discussed this with volunteers at the cemetery, they consulted maps of the grave plots and realized that they had placed the headstone in the middle of what was once a pathway, rather than on top of a grave.

3.2 Data Collection

Initial attempts at determining the grave location were also hindered by a lack of precision with the smart phone and hand-held GPS devices that we used for geolocation.

To account for the difficulties in bad GPS data, we attempted to select reasonable ϵ values. To do this, we took GPS measurements at known locations using the same devices. The average error across these points was approximately a meter, with a maximum error of just over two meters. We allowed ϵ in this case to vary between 0 and 2. This large amount of error, combined with the inclusion of some bad correspondences, described above, resulted in the large uncertainty region seen in Figure (3).

4 Successful Attempt

In our final, successful attempt at locating the lost grave, we acquired a Trimble R8 VRS Rover real time kinematic GPS device with sub-meter accuracy. Unlike either the smart phones or the handheld GPS devices that we initially used, the Trimble reports both its location and the possible error associated with that reported location. This accuracy was verified on-site by confirming the reported location at sites with known geolocation (i.e., road intersections that are identifiable in modern day aerial imagery). Across all correspondences recorded at Washington Park Cemetery, the Trimble reported an average horizontal error of 0.066 meters. Even this small amount of error, however, is enough that the optimization in Section (2.4) may produce sub-optimal camera parameters. Therefore, we use the Trimble's reported error as ϵ in the attempts to account for geolocation error described in Section (2.5).

Another source of error in our optimization is the inclusion of bad correspondences. In order to prevent this, we built a web tool that provided instantaneous feedback on the possible veracity of correspondences. The interface showed an analyst the visualization of the reprojection error seen in Figure (2), given the results of the calibration in Section (2). The analyst could then “turn off” specific correspondences to see the effect on the error. This utility was used to identify the bad correspondences discussed in Section (3.1) and either correct them or exclude them in the final calibration.

Using this tool to identify confident matches allowed us to geo-locate the camera taking the image in Figure (4). The corner of the grave is clearly near the camera, but to give a specific estimate of the grave location, we used the computed focal length the camera, the position of the horizon line, and an estimated height of the camera of 1.57 meters to estimate that the near corner of the grave was 0.4 meters in front of the 2D camera location.

5 Results

The result of our final attempt to locate the grave was a narrow uncertainty region 1.628 meters long. On June 16, 2013, Saint Louis Police performed an exhumation at this location, unearthing the casket of the young Jane Doe within only a few hours [7, 9]. The mean point of our uncertainty region was 0.2 meters from the center of the found casket, and our uncertainty region encompassed the grave.

6 Conclusions

This case study led us to several conclusions that we think may generalize to other situations. First, by assuming a simple camera model with no roll, we were able to limit the geometric constraints in the camera calibration, and therefore reduce the number of correspondences needed to solve for the camera location. Second, for the purpose of very accurate geo-location (i.e., sufficient to direct digging efforts), measuring world-point locations with a smart phone or hand-held GPS device is not sufficient to yield high-accuracy results. Third, over the course of many years, there can be dramatic changes in a scene, even

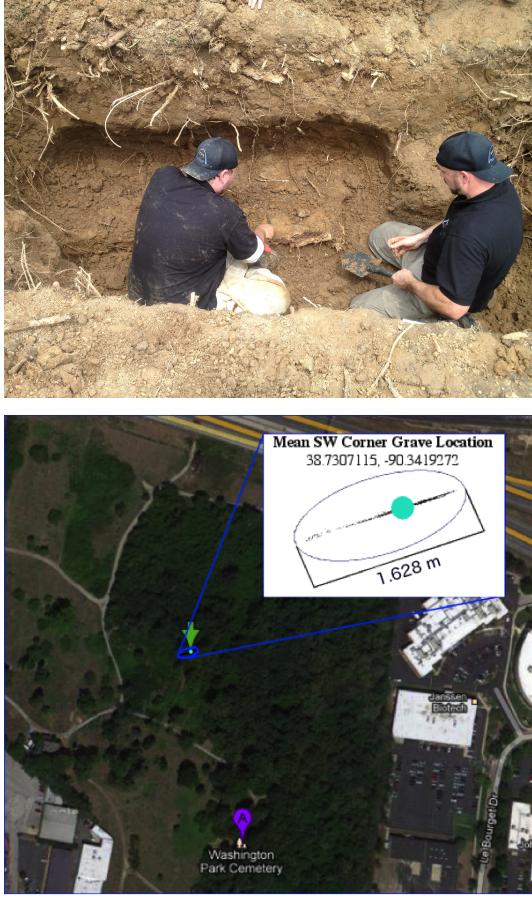


Figure 5. Saint Louis Medical Examiner employees carefully unearthing the casket of the young Jane Doe at the location determined in our successful camera calibration. The correct location is shown with present day aerial imagery of Washington Park Cemetery. Compared to the sparse trees in Figure (4), the cemetery is now heavily forested. Imagery ©2013 DigitalGlobe, U.S. Geological Survey, USDA Farm Service Agency, Map data ©2013 Google.

in a location as undeveloped as a cemetery, and a single bad correspondence can dramatically change an estimated camera location. Some of the difficulty posed by the change in time can be surmounted with good sources of similarly aged aerial imagery.

Although it did not affect this case, in related situations it may also be important to understand the provenance of the image. If the only image that is available was cropped then the $\frac{w}{2}$ and $\frac{h}{2}$ terms in Equation (1) may be different. If a negative has been scanned, that matrix may have a different form, and if there is non-linear camera distortion (e.g. fish-eye effects or barrel distortion), then a more complicated camera calibration model overall is necessary. Even in this more complicated case, we think an interactive, analyst-in-the-loop tool would be the most efficient way to find good correspondences between an image and the real world.

References

- [1] Geoguessr. <http://geoguessr.com/>.
- [2] Decapitated body of girl found on CLEMENS AVE. *Saint Louis Globe-Democrat*, February 28 1983.
- [3] Austin Abrams and Robert Pless. Webcams in context: Web interfaces to create live 3D environments. In *ACMMM*, pages 331–340, Jun 2010.
- [4] Abdel Y. I. Aziz and H. M. Karara. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Proc. of the Symposium on Close-Range Photogrammetry*, pages 1–18, Urbana, Illinois, 1971.
- [5] Duane C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.
- [6] Christine Byers. New hope in cold case of decapitated girl found in St. Louis, but now where's the body? *Saint Louis Post-Dispatch*, March 17 2013.
- [7] Christine Byers. Slain girl's remains found as part of 30-year-old St. Louis murder investigation. *Saint Louis Post-Dispatch*, June 17 2013.
- [8] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2000.
- [9] Sam Levin. St. Louis cops, researchers find remains of slain "Jane Doe," girl decapitated 30 years ago. *Daily Riverfront Times*, June 17 2013.
- [10] Betsy Mason. Then and now: Repeat photography captures changing landscapes. *Wired Science*, December 9 2010.
- [11] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.
- [12] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3D. In *SIGGRAPH Conference Proceedings*, pages 835–846, 2006.
- [13] Andrew Sullivan. The dish. <http://dish.andrewsullivan.com/>.
- [14] R. Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 364–374, Miami, June 1986.
- [15] Ling Xie and Shawn Newsam. IM2MAP: deriving maps from georeferenced community contributed photo collections. In *Proceedings of the 3rd ACM SIGMM international workshop on Social media*, pages 29–34, 2011.
- [16] W. Zhang and J. Kosecka. Image based localization in urban environments. In *International Symposium on 3D Data Processing, Visualization and Transmission*, 2006.