SHADOW ANALYSIS FOR CAMERA HEADING IN IMAGE GEO-LOCALIZATION

Matthew Clements and Avideh Zakhor[1]
Signetron Inc. and U.C. Berkeley
{clements@eecs., avz@}berkeley.edu

Image geo-localization is an important problem with many applications such as augmented reality and navigation. The most common ways to geo-localize an image are to use its meta-data such as GPS or to match it against a geo-tagged database. When neither of those is available, it is still possible to apply shadow analysis to determine the camera heading for outdoor images. In this paper, we develop a novel method for deducing the global heading of a query image using the shadows in it. We start by constructing a model of the sun-earth system to determine all shadows possible at a given latitude, and compare shadows within the query to those possible under the model to determine the range of possible headings. We demonstrate this on 54 query images with known ground truth, and show that in 52 cases the ground truth lies in the computed range.

1 Introduction

Image geo-localization is an important problem with many applications such as augmented reality and navigation. The easiest way to geo-localize an image is to use the meta data from the sensors integrated with the imaging device such as GPS or compass in order to determine the location and the heading of the captured imagery. However, many images lack such meta data, or have erroneous meta data due to various physical conditions. For example GPS is known to be erroneous in urban canyon areas, can at times be jammed by adversarial forces, and is unavailable indoors. Furthermore, compass data which relies on detecting Earth's magnetic field is unreliable in the presence of power lines which produce their own magnetic field, or steel cabinets which could distort the magnetic field indoors [1].

One approach to image localization is to match the query image to overhead databases such as Digital Elevation Maps (DEM) or satellite imagery. An example of this approach involves matching skylines from query images to synthetic skylines from DEM data [2,3]. Another approach to geolocation is to match the query image against a pre-collected geo-tagged image database [4-6,9-15]. It is also possible to geolocate an image from the content of the image itself without matching it to any databases, be it overhead or terrestrial. For example, the shadows in an image can be used to deduce its heading, if time of day and time of year are known. In the absence of the latter, it is possible to derive a range of possible headings for the captured image based on shadow analysis. This is reminiscent of the fact that hunters, hikers, and others who spend substantial periods of time in northern hemisphere wildernesses use an analog watch as a makeshift compass; they do so by simply pointing the hour hand of the watch toward the sun, resulting in south lying about halfway between the hour hand and 12:00 noon. While this simple example is not directly applicable to the problem of determining the heading of a camera taking a picture at an unknown time it motivates an avenue of analysis based on shadows.

In this paper, we propose a novel method for deducing the global heading of a query image using the shadows in it. We start by constructing a model of the sun-earth system to determine all shadows possible at a given latitude, and compare shadows within the query to those possible under the model to determine heading. In Section 2, we review prior work, Section 3 includes proposed method, and Section 4 shows results.

2 Related Work

Similar to the analog watch example, our proposed method exploits the position of the sun in order to prune the space of possible headings. This relationship is well studied, with empirical models on solar altitude and azimuth predicting heading to within a tenth of a second of angle, or rapidly computing them to within a minute of angle [7].

In our problem set up though, since the position of the query image is assumed to be known only to within a degree in latitude and longitude, such positional precision by models in [7] is unnecessary. Further, these models require precise capture time information, which in our problem set up is not

[1]Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, Air Force Research Laboratory (AFRL), or the U.S. Government.

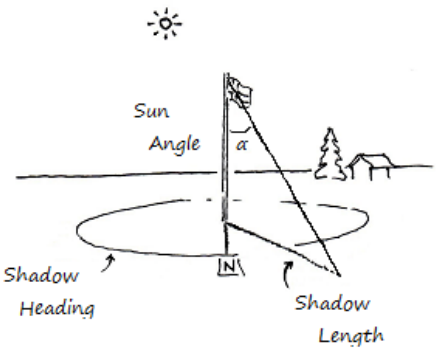


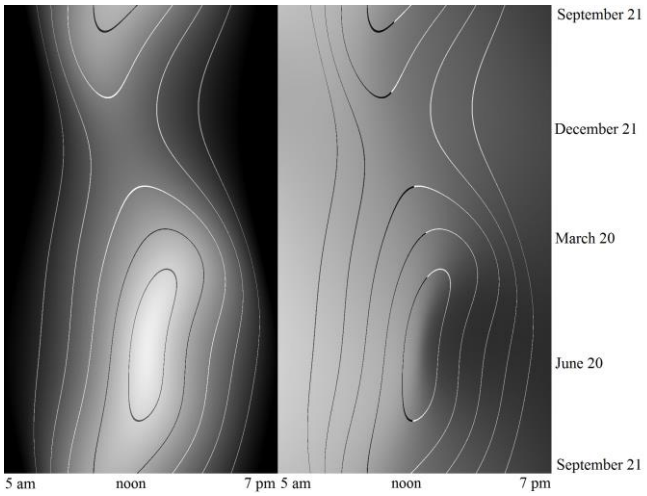
Figure 1: Illustration of sun angle, shadow heading, and shadow length for a vertical object such as a flagpole.

known. Thus, our approach is to iterate over the many times and dates in a year in order to arrive at a *range* of possible headings, rather than to estimate one value for it.

Given the less rigid precision requirements, and the need to evaluate the model many times in short order, we choose to simplify the model in [7] further, neglecting higher-order effects such as the action of the Moon and Jupiter, and the eccentricity of the Earth's orbit and wobble of its axis. This simplified model allows us to relate the position of the sun in the sky to shadow attributes such as "sun angle" and "shadow heading" as shown in Figure 1. As seen, shadow heading is defined as the counterclockwise angle from the North direction to the direction of the shadow. "Sun Angle" α is defined to be the angle between the shadow caster which is assumed to be vertical, and the line connecting the end of the shadow to the top of the caster. From simple geometry, we conclude that

$$\alpha = \arctan\left(\frac{shadow\_length}{caster\_height}\right)$$

An example of the results of the simplified model in [7] which has been iterated over date and time for "sun angle" and "shadow heading" at an approximate latitude of 30° North is shown in Figures 2(a) and 2(b) respectively. In this Figure, vertical axis indicates time of year with September 21st, the autumnal equinox, at the top and bottom of the axis. The horizontal axis denotes time of day, with the left most point indicating 5 am and the right most point indicating 7 pm. In Figure 2(a) darker regions correspond to larger sun angles, i.e. closer to 90° than zero, and consequently longer shadows; in Figure 2(b) lighter (darker) indicates westerly (easterly), with North being neutral gray. Isocontours of length from Figure 2(a) are reproduced in Figure 2(b) for visualization purposes only. Combining Figures 2(a) and 2(b), we obtain Figure 3, which shows a two dimensional histogram of the sun angle in the horizontal axis and shadow heading on the vertical axis. The intensity in Figure 3 corresponds to the relative frequency of a particular sun angle and shadow heading occurring over the time period 5 am to 7 pm of all days of a year. Thus,



(a) **(b)**

Figure 2: (a) sun angle; (b) shadow heading.

brighter points correspond to higher temporal "likelihood" of sun angle and shadow headings measured over a year.

3 Proposed Approach

Our approach is to compare extracted information from the query image against that of the model whose precision matches the 1° ambiguity in our latitude and longitude information. Specifically, we extract two quantities shown in Figure 4 from the query image: *sun angle α* and angle of *offset* from shadow to camera heading.

As shown in Figure 1, the *length* of the query shadows, relative to the heights of the objects casting them, provides an estimate of *sun angle α*. This together with the 2D plot in Figure 2(a) allows us to narrow the range of times at which the query could have been taken; by range of times, we mean both time of day, and time of year. Intuitively, a man casting a shadow five or more meters, for example, is not doing so at noon on the summer solstice, nor is a tree whose shadow does not extend from below its leaves photographed early in the morning, late in the evening, or in the depths of winter. Furthermore, by only recording information for sun angles from the model that match the *sun angle* of the shadows in the query, we can narrow the range of possible shadow *headings* for the query. This can be done by utilizing the 2D density plots such as the one shown in Figure 3. Specifically, the estimated sun angle specifies a vertical line in the plot which, when intersected with the non-zero portion of the density plot results in possible range of values for shadow headings.

We now need to devise a mechanism to derive camera heading from shadow heading. This amounts to estimating the angle of *offset* between the two headings, measured in the ground plane from the camera's optical axis to the shadows in the query. Both offset angle and sun angle can ideally be estimated by a user with a good eye and a strong grasp of spatial reasoning, but a more principled approach that relies less on the skill of the user would consist of using camera

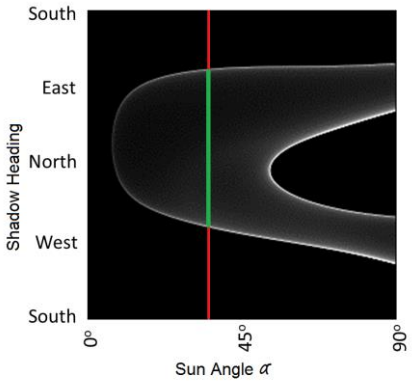


Figure 3: A histogram of co-occurrences of shadow heading and sun angle.

Figure 4: Offset is the angle between the shadow (blue) and the camera's optical axis (orange), measured in the ground plane.

parameters *roll*, *pitch*, and *focal length* to extract them.

The estimation of camera parameters from a single image is a well-studied topic [8]. Their use in the analysis of shadows boils down to building three-dimensional models of the shadows present in the query. This is done by making two simplifying assumptions that generally hold in practice: (a) level ground plane: we assume that the ground that the shadows fall onto is locally planar with a vertical normal; (b) vertical shadow caster: we assume that the tops of objects casting shadows stand directly above their bases. Operating under these assumptions, and using camera parameters to translate pixel coordinates into rays in the world frame, 3D coordinates of the following three points describing the shadow can be estimated: the common base of both shadow and caster, the tip of the shadow, and the top of the shadow caster.

To do so, we define the world coordinate frame such that the ground plane coincides with the xy-plane as shown in Figure 5; the camera lies on the z-axis at a fixed, arbitrary distance from the origin, the x-axis extends below the camera's optical axis, and the y-axis extends to the camera's left, or into the

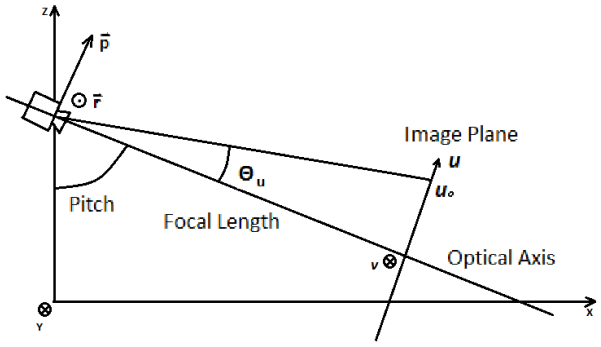


Figure 5: Geometry of the camera with respect to world coordinates.

plane of Figure 5. We take (u, v) to be the real-valued coordinate of a pixel with respect to the image center at (0, 0), after a rotation of the image to compensate for camera *roll*.

A point in space can be uniquely identified by the intersection of three non-parallel planes; our pixel coordinates can provide two such planes, with the third being supplied by our assumptions. Specifically, a plane of constant $u = u_0$, as shown in Figure 5, is the image of the yz-plane rotated *pitch* + θ_u = *pitch* + arctan(u_0 / *focal length*) about the vector $\vec{\boldsymbol{r}}$ = <0, –1, 0> centered on the camera and extending out of the plane of the figure. Similarly, a plane of constant $v = v_0$ (not shown in the figure), is the image of the zx-plane rotated θ_v = arctan(v_0 / *focal length*) around the camera and the vector $\vec{\boldsymbol{p}}$ = <cos(*pitch*), 0, sin(*pitch*)>, or directly up from the camera's point of view. For the two points at the base and tip of the shadow, the third plane is the ground plane; for the point at the top of the caster, the third plane is one parallel to the yz-plane and containing the common point at the base of the shadow and caster.

Given the 3D coordinates of these points, (x_{base},y_{base},0), (x_{tip},y_{tip},0), and (x_{base},y_{top},z_{top}), *length* and *offset* are given by:

$$length = \frac{\sqrt{\left(x_{tip} - x_{base}\right)^2 + \left(y_{tip} - y_{base}\right)^2}}{z_{top}}$$

$$offset = \arctan\left(\frac{y_{tip} - y_{base}}{x_{tip} - x_{base}}\right)$$

Regardless of how they are acquired, once we estimate *length* and *offset*, we can run our model for all times and at the latitude of our region of interest, recording the *headings* of any shadows in the model that match the *length* of the query shadows, adding the offset to each one to produce a candidate camera heading. As shown in the right column of Figure 6, these camera headings are collected into a list that can be interpreted as a probability density function, if one assumes that any time is equally likely as any other time for the picture to have been taken. Under such an interpretation, the range(s) of possible headings, mean, and standard deviation can also be extracted.

4 Results

The proposed analysis was applied to 54 images or video frames with ground truth, from 5 regions of interest with latitudes from 32° North to 33° South and terrains from desert to jungle and development from urban through rural to none. In all chosen images and/or video frames shadows and their casters are clearly visible. 25 of the 54 queries come from a region in Jordan, around Ammon, with fewer numbers coming from the other 4 regions, in Taiwan, India, the Philippines, and Chile; the desert terrain and climate provided consistently good visibility with little cloud cover.

Typical ranges of possible headings for the 54 queries varied from 0.8π to 1.2π, in either one contiguous arc as shown for the top two images in Figure 6, or two disjoint arcs that were independently contiguous as shown in the bottom three images in Figure 6. The former typically corresponds to midday queries and the latter to morning/evening queries. The proposed analysis was successful on 52; "success" in this case is defined as "ground truth heading is within the returned range of possible headings", and the probability of achieving such a success rate by chance alone is less than $1.62*10^{-13}$. Five examples of "success", together with the heading ranges, are shown in Figure 6. The two failures are shown in Figure 7.

Figure 6: A sampling of success cases. The blue marks indicate means of distributions; the red, ground truth headings.

Figure 7: The two miss cases.

Use of camera calibration in the analysis was applied to 22 of the 54 queries; for 32 queries, *length* and *offset* were extracted manually by a user, as the values produced by computation from camera parameters produced implausible results which did not match the approximate results generated by 3D spatial reasoning of a skilled human operator: estimated offsets nearly perpendicular to what human eyes see, or computed lengths with exceedingly large values, or negative ones. The reasons for these calibration failures varied from query to query, but fell into three categories: (a) Queries in which the shadows being examined did not fall on the ground, but rather on the sides of buildings or other objects; these were typically urban queries; (b) Queries for which estimating camera parameters without metadata is difficult due to a lack of prominent vanishing points; these were typically in rural or undeveloped areas; (c) Queries in which the shadows were distant or the camera was not elevated significantly above the ground plane, leading to increased response in x-coordinates to noise in *pitch*, *focal length*, or pixel coordinate u. For the two failure query images shown in the left column of Figure 7, neither camera calibration nor manual eye-balling resulted in a heading range in which the ground truth lay, as shown in the right column of Figure 7.

For the 22 queries in which calibration was successful, the *sun angles* and *offsets* estimated manually differed from those extracted by calibration by an RMS average of 7.9° and 16.0° and a maximum of 14.8° and 32.7°, respectively. The bulk of the discrepancies in *sun angle* were on queries for which the *offset* was near either 0° or 180°; the bulk of the offset discrepancies were near offsets or ±90°. Arithmetic mean discrepancies of –5.3° and –0.19° for *sun angle* and *offset* respectively indicate that the user in question either had an underestimation bias for shadow *lengths* and *sun angles*, or more strictly rejected calibration results reporting shorter shadows, but was not significantly biased in estimates of *offset*. In all cases where calibration was successful, the ground truth lay in the obtained heading range, regardless of whether manual processing or camera calibration was used to extract *length* and *offset*.

5 References

[1] A. Hallquist and A. Zakhor, "Single View Pose Estimation of Mobile Devices in Urban Environments," IEEE Workshop on the Applications of Computer Vision (WACV) 2013. Clearwater Beach, Florida, January 2013.

[2] E. Tzeng, A. Zhai, M. Clements, R. Townshend, and A. Zakhor, "User-Driven Geolocation of Untagged Desert Imagery Using Digital Elevation Models," CVPR 2013 Workshop on Visual Analysis and Geo-Localization of Large-Scale Imagery. Portland, Oregon, June 2013.

[3] G. Baatz, O. Saurer, K. Koeser, and M. Pollefeys. Large scale visual geo-localization of images in mountainous terrain. In Proc. European Conference on Computer Vision, 2012.

[4] J. Zhang, A. Hallquist, E. Liang, A. Zakhor, "Location-Based Image Retrieval for Urban Environments," ICIP 2011, Brussels, Belgium, September 11-14, 2011.

[5] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image Based Localization in Indoor Environments," International Conference on Computing for Geospatial Research and Applications, San Francisco, CA, July 2013.

[6] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Reduced-Complexity Data Acquisition System for Image Based Localization in Indoor Environments," IPIN 2013, Montbeliard, France, October 2013.

[7] Robert Walraven, "Calculating the Position of the Sun," *Solar Energy*, vol. 20, pp. 393-397, Sept. 1977.

[8] R. I. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision, Second Edition," Cambridge University Press, March 2004.

[9] Wei Zhang and J Kosecka, "Image Based Localization in Urban Environments," Third International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 33-40 Chapel Hill, NC, June 2006.

[10] J. Kosecka, Liang Zhou, P. Barber, and Z. Duric, "Qualitative Image Based Localization in Indoors Environments," 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2003.

[11] J. Wolf, W. Burgard, and H. Burkhardt, "Robust Vision-Based Localization by Combining an Image Retrieval System with Monte Carlo Localization," IEEE Transactions on Robotics, vol. 21, issue 2, pp.208-216, April 2005.

[12] J. Wolf, W. Burgard, and H. Burkhardt, "Robust Vision-Based Localization for Mobile Robots Using an Image Retrieval System Based on Invariant Features," ICRA '02, vol. 1, pp. 359-365, 2002.

[13] I. Ulrich and I. Nourbakhsh, "Appearance-Based Place Recognition for Topological Localization," ICRA '00, vol. 2, pp. 1023-1029, 2000.

[14] Torsten Sattler, Tobias Weyand, Bastian Leibe, and Leif Kobbelt, "Image Retrieval for Image-Based Localization Revisited", BMVC, 2012.

[15] Thomas Moulard, Pablo Fernandez Alcantarilla, Florent Lamiraux, Olivier Stasse, and Frank Dellaert, "Reliable Indoor Navigation on Humanoid Robots Using Vision-Based Localization"