INTERACTIVE SHADOW ANALYSIS FOR CAMERA HEADING IN OUTDOOR IMAGES

Matthew Clements and Avideh Zakhor

Signetron Inc. and U.C. Berkeley
{clements, avz}@eecs.berkeley.edu

ABSTRACT

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. This could be useful pruning the search space in geo-localization applications, for example by removing roads with incompatible orientations from a database such as Open Street Map. In this paper, we develop a novel interactive 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 approximate 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 sensors integrated with the imaging device such as GPS or compass in order to determine the location and 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. Compass data, which relies on detection of Earth's magnetic field, is unreliable in the presence of power transmission lines which generate their own magnetic fields and steel cabinets which disturb existing magnetic fields [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 geo-localization is to match the query against a pre-collected geo-tagged image database [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. It is also possible to geo-localize an image from the content of the image itself without matching it to any databases, be they overhead or terrestrial. For example, the sunlight and shadows in an outdoor image can be used to deduce its heading, if time of day and time of year are known. Even in the absence of the latter, a rough approximation of heading may be made: for instance, hunters, hikers, and other outdoorsmen wandering northern hemisphere wildernesses often use their analog watches as makeshift compasses by pointing the hour hand of the watch toward the sun, knowing that South lies halfway between that 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 nonetheless motivates an avenue of analysis based on shadows.

In this paper, we propose a novel interactive 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 approximate 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 the proposed method, and Section 4 shows results.

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.

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. The position of the sun as a function of time 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 [14].

In our problem set up, where the position of the query image is assumed known only to within 100 kilometers or about a degree and a half of latitude, such positional precision as the models in [14] provide is unnecessary. Further, evaluation of these models requires precise capture time information, which in our problem set up is not known. Thus, our approach is to iterate over the many times and dates in a year to arrive at a $range$ of possible headings, rather than to estimate one value for heading.

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 [14] further, neglecting higher-order effects such as the action of the Moon and Jupiter, the eccentricity of the Earth's orbit and the wobble of its axis, and human constructs such as time zones and daylight savings time. The resulting model provides the position of the sun as its angle of $altitude$ above the horizon, and its angle of $azimuth$ clockwise from North, as a function of the time of day, the time of year, and the coarse latitude of our query.

The position of the sun ($altitude$ and $azimuth$) provided by the model is related via one-to-one mapping to attributes of the shadow such as its $length\ ratio$ and $heading$, as shown in Equation 1 [14].

$$\begin{aligned} length\ ratio &= \cot(altitude) \\ heading &= azimuth + \pi \end{aligned} \tag{1}$$

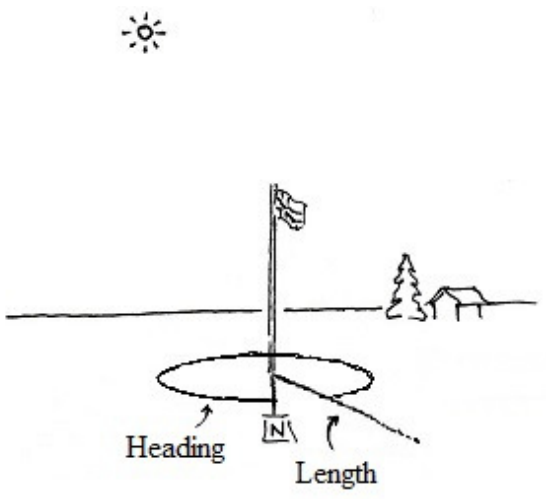

Fig. 1. Illustration of key shadow descriptors.

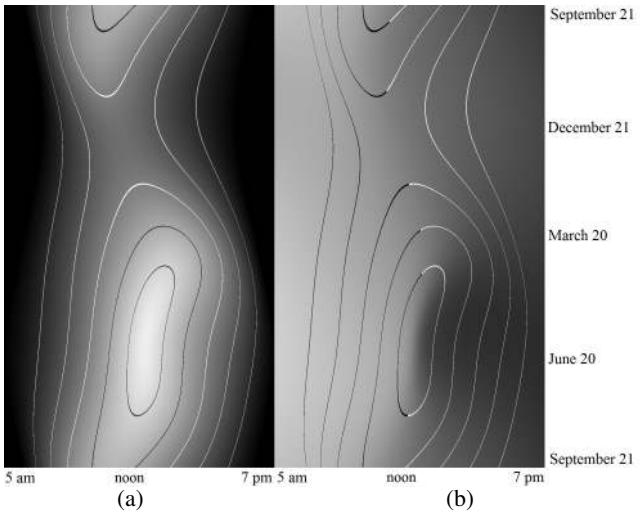

Fig. 2. Plots of shadow (a) $length\ ratio$ and (b) $heading$

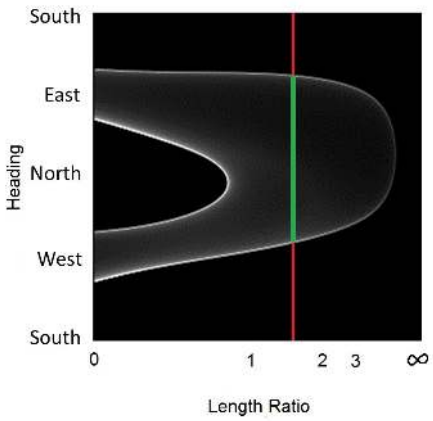

Fig. 3. A histogram of co-occurrences of shadow $heading$ and $length\ ratio$.

Shown in Figure 1, the shadow's $heading$ is defined to be the angle clockwise between North and the ground-plane direction the shadow falls in. Its $length\ ratio$ is the shadow length normalized by dividing out the height of the shadow's caster; this allows information contained in the length to be compared to other shadows possible under the model.

An example of the results of the simplified model which has been iterated over date and time for $length\ ratio$ and $heading$ at an approximate latitude of 30° North is shown in Figures 2(a) and 2(b) respectively. In this figure, the vertical axis indicates time of year with September 21 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 rightmost point indicating 7 pm. In Figure 2(a) darker regions correspond to longer shadows; in Figure 2(b) lighter/darker indicates westerly/easterly, with North being neutral gray. Isocontours of $length\ ratio$ 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 shadow $length\ ratio$ in the horizontal axis and $heading$ on the vertical. The intensity in Figure 3 corresponds to the relative frequency of a particular $length\ ratio$/$heading$ co-occurrence within the time window, 5 am to 7 pm of every day of the year. Thus, brighter points correspond to higher temporal "likelihood" of shadow $length\ ratio$ and $heading$, measured over a year. In the event that capture time is known, $length\ ratio$ and $heading$ can be pinpointed.

3. PROPOSED APPROACH

Our approach is to compare information extracted from the query image against that of the model, narrowing the possible camera headings from a range 2π to something more discriminative. Specifically, we extract the two quantities shown in Figure 4: shadow length relative to caster height or $length\ ratio$ and $offset\ angle$ from the shadow $heading$ clockwise to the camera axis. The main concept is to use the $length\ ratio$ from the query together with our model to arrive at a range of possible shadow $heading$s in the global coordinate frame, and then to add the $offset\ angle$ angle to these shadow $heading$s to derive the $range$ of possible camera headings.

The $length\ ratio$ of a query shadow, together with the information contained in the plot of 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 the tuple of (time of day, time of year), rather than the ranges of the two values independently. Intuitively, a man casting a shadow five or more meters is not doing so at noon on the summer solstice, nor is a tree whose shadow does not extend beyond its dripline photographed early in the morning, late in the evening, or in the depths of winter. With this range in hand we look up the shadow $heading$s possible at those times, either from a plot such as Figure 2(b), or more directly from a corresponding plot in the form of the one in Figure 3, where the shadow $length\ ratio$ selects a vertical line in the plot, whose non-zero values represent possible shadow $heading$s. We then need to devise a mechanism to derive camera heading from shadow $heading$. This amounts to estimating the $offset\ angle$ between the two headings, measured in the ground plane from the shadows in the query to the camera's optical axis.

Fig. 4. $offset\ angle$ is the angle between the shadow (blue) and the camera's optical axis (orange), measured in the ground plane.

Both $offset\ angle$ and $length\ ratio$ can be estimated by a human 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 calibration parameters such as $roll$, $pitch$, and $focal\ length$ to estimate world coordinates for a trio of points describing the shadow, and compute $length\ ratio$ and $offset\ angle$ from those. Even in the latter scenario of using camera parameters, a user is still required to click on the trio of points, namely the base of an object casting a shadow, the tip of that shadow, and the top of the object corresponding to that tip.

The estimation of camera calibration parameters from a single image is a well studied topic [15]. In this work, camera calibration parameters obtained via vanishing point/line analysis were supplied to us. As discussed later, these parameters are noisy and not necessarily error free. However, the use of vanishing point analysis allows us to both calibrate and take measurements from the same image. The use of $roll$, $pitch$, and $focal\ length$ in the analysis of the shadows boils down to providing the basis for a transformation from the pixel coordinates of the camera frame to world coordinates. However, camera parameters and pixel coordinates alone only specify a ray in the world frame, rather than a unique point; therefore, additional information is needed. This additional information is provided by making two simplifying assumptions that seem reasonable and generally hold: (a) we assume a level ground plane into which the shadow falls, locally planar and with a vertical normal; (b) we assume a vertical shadow caster, with its top directly above its base. Operating under these assumptions, along with camera parameters and pixel coordinates, 3D coordinates of the shadow tip, caster top, and shared base can be computed.

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 z_0, the x-axis extends below the camera's optical axis, and the y-axis extends to the camera's left, or into the plane of Figure 5. We take pixel coordinates $(u,v)_{pixel}$ to be the real-valued coordinates of a pixel with respect to the image center at $(0,0)_{pixel}$, after applying a 2D rotation in the pixel space to compensate for camera $roll$.

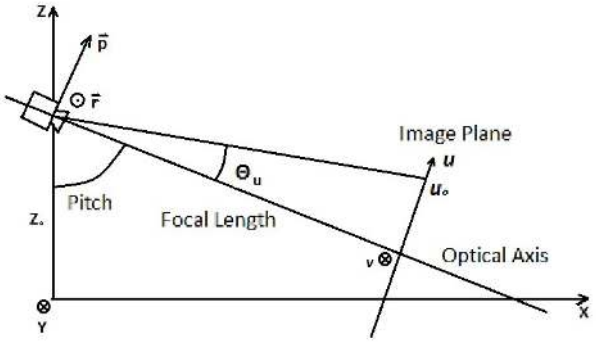

Fig. 5. Geometry of the camera with respect to world coordinates.

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 pixel coordinate $u = u_0$, as shown in Figure 5, is the image of the yz-plane rotated $pitch + \theta_u = pitch + \arctan(u_0/focal\ length)$ about the vector $\vec{r} =< 0, -1, 0 >$ centered on the camera and extending out of the plane of the figure; such a plane appears as a horizontal line in the camera image. Similarly, a plane of constant pixel coordinate $v = v_0$ (appearing as a vertical line in the camera image and not shown in the Figure) is the image of the zx-plane rotated $\theta_v = \arctan(v_0/focal\ length)$ around the camera and the vector $\vec{p} =< \cos(pitch), 0, \sin(pitch) >$, or directly up from the camera's point of view. For the two points on the ground, at the base and the tip of the shadow, the third plane is the xy-plane, which we defined to be coincident with 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 point of the shared base.

Given the coordinates of these points, $(x_{base},y_{base},0)_{world}$, $(x_{tip},y_{tip},0)_{world}$, and $(x_{base},y_{base},z_{top})_{world}$, $length\ ratio$ and $offset\ angle$ are given by:

$$\begin{aligned} length\ ratio &= \frac{\sqrt{(x_{tip}-x_{base})^2+(y_{tip}-y_{base})^2}}{z_{top}} \\ offset\ angle &= \arctan\left(\frac{y_{tip}-y_{base}}{x_{tip}-x_{base}}\right) \end{aligned} \tag{2}$$

Once we estimate the $length\ ratio$ and $offset\ angle$ of the query shadow, we can run our model at the latitude of our region of interest and for all times, recording the $heading$s of any model shadows that have $length\ ratio$s matching that of the query, adding the $offset\ angle$ 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 capture time is as likely as any other. Under such an interpretation, the $range$(s) of possible camera headings, mean(s), and standard deviation(s) can also be extracted.

In summary, the end-to-end processing of a query is an interactive process and consists of the following stages: (1) the user selection of the three points in an image that describe a shadow: the base of shadow and caster, the top of the caster, and the tip of the shadow, (2) the automatic extraction of vanishing points and estimation of camera parameters $roll$, $pitch$, and $focal\ length$, (3) the computation of world coordinates of the selected points and calculation of shadow $length\ ratio$ and $offset\ angle$ angle; this can be done in two ways, either through spatial reasoning of an interactive user as

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

described in Section 4, or automatically from Equation 2 and Figure 5, (4) selection of shadow $headings$ from model shadows that match the $length\ ratio$ from the query as shown in Figure 3, (5) addition of query $offset\ angle$ to determine the final $range$ of query camera headings.

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. Coarse latitude of each query is assumed to be known to within 1.5°. In all chosen images and/or video frames shadows and their casters are clearly visible.

Typical ranges of possible headings for the 54 queries varied from 0.8π to 1.2π, in either one contiguous arc as in the top two images in Figure 6, or two disjoint arcs that were independently contiguous, as in the bottom three. The former typically corresponds to midday queries and the latter to morning/evening queries. The proposed analysis was successful on 52 out of 54; "success" in this case is defined as "ground truth heading is within the returned range of possible headings"; the probability of flipping 52 heads on 54 coin tosses is less than 1.62×10^{-13}, and since the returned ranges tended toward π, it may be expected that achieving the same success rate for the proposed method through chance alone would have comparably minuscule probability. Five examples of "success", together with the heading ranges, are shown in Figure 6. The two failures are shown in Figure 7.

Fig. 7. The two miss cases.

Two major weaknesses of the proposed method are that (a) we presently have no way to automatically extract the pixel coordinates of points describing shadows of interest, and (b) the proposed method is highly vulnerable to noise, both in the estimation of camera parameters and in selection of pixel coordinates of shadow tip, caster top, and shared base. As such, user interaction was required on all queries, to select the points describing the shadows. Furthermore, user intervention was required on 32 of the 54 queries, to estimate $length\ ratio$ and $offset\ angle$ manually. In all such intervention cases, shadow $length\ ratio$s or $offset\ angle$s computed automatically were clearly erroneous, e.g. $length\ ratio$s differing by a factor of three or more from what the expected, or $offset\ angle$s differing by $\pi/3$ or more. The failures of the automated estimation fall into three classes: (a) queries in which the shadows are small, occupying no more than tens of pixels; (b) queries in which the shadows are distant, and the plane of constant u shown in Figure 5 and discussed in Section 3 is nearly parallel to the ground xy-plane; and (c) queries in which the shadows fell on the sides of buildings, such that the shared "base" of shadow and caster was a point in open air and difficult to select precisely. For the two failure cases shown in Figure 7, neither manual nor automated estimation of shadow descriptors resulted in success.

For the 22 queries in which no user intervention was needed, the manually estimated values for $altitude$s of the sun (computed from the shadow $length\ ratio$s as in Equation 1) and shadow $offset\ angle$s differed from those computed automatically 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 $length\ ratio$ and thus $altitude$ were on queries for which the $offset\ angle$ was near 0° or 180°; the bulk of the $offset\ angle$ discrepancies were near offsets of ±90°. In all cases where automated computation of shadow descriptors $length\ ratio$ and $heading$ was accepted, the ground truth lay within the obtained camera heading $range$. Had the user intervention been skipped for the 32 queries requiring it, 13 would have remained "successes", but that is likely due to chance.

5. REFERENCES

[1] A. Hallquist and A. Zakhor, "Single view pose estimation of mobile devices in urban environments," in *IEEE Workshop on the Applications of Computer Vision (WACV) 2003*. IEEE.

[2] E. Tzeng, A. Zhai, M. Clements, R. Townshend, and A. Zakhor, "User-driven geolocation of untagged desert imagery using digital elevation models," in *CVPR 2013 Workshop on Visual Analysis and Geo-Localization of Large-Scale Imagery*, June.

[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, and A. Zakhor, "Location-based image retrieval for urban environments," in *ICIP 2011*.

[5] J.Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based localization in indoor environments," in *International Conference on Computing for Geospatial Research and Applications*, July.

[6] J.Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Reduced-complexity data acquisition system for image based localization in indoor environments," in *IPIN 2013*.

[7] W. Zhang and J. Kosecka, "Image based localization in urban environments," in *Third International Symposium on 3D Data Processing, Visualization, and Transmission*, June 2006, pp. 33–40.

[8] J. Kosecka, L. Zhou, P. Barber, and Z.Duric, "Qualitative image based localization in indoors environments," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, June 2003.

[9] 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, pp. 208–216, April 2005.

[10] J. Wolf, W. Burgard, and H. Burkhardt, "Robust vision-based localization for mobile robots using an image retrieval system based on invariant featrues," in *ICRA '02*, 2002, vol. 1, pp. 359–365.

[11] I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," in *ICRA '00*, 2000, vol. 2, pp. 1023–1029.

[12] Torsten Sattler, Tobias Weyand, Bastian Leibe, and Leif Kobbelt, "Image retrieval for image-based localization revisited," in *BMVC*, 2012.

[13] Thomas Moulard, Pablo Fernandez Alcantarilla, Florent Lamiraux, Olivier Stasse, and Frank Dellaert, "Reliable indoor navigation on humanoid robots using vision-based localization," .

[14] Robert Walraven, "Calculating the position of the sun," *Solar Energy*, vol. 20, pp. 393–397, September 1977.

[15] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, March 2004.