Towards calibration-free geo-localization of stationary outdoor webcams

Frode Eika Sandnes

Faculty of Engineering
Oslo University College

Abstract

This study proposes two strategies for determining the approximate geographical location of outdoor webcams based on time-series comprising regularly sampled images. The strategies require an accurate account of universal time and the date to be known, then the intensity of the images are used to estimate the sunrise and sunset times. Given accurate sunrise and sunset times a celestial model can be used to extract the latitude and longitude of the webcam. Webcams vary in quality, dynamic pixel depth and sensitivity to light. Two strategies are therefore proposed for avoiding the need to perform calibration. The first technique involves normalizing and noise-damping the image intensity measurements. The second technique employs a self-normalizing brightness ratio. The brightness ratio is computed from the overall brightness of the upper part of the image in relation to the bottom part. During day the sky is much brighter than the ground, while at night the sky is much darker than the ground if the ground is lit up. Experiments demonstrate that the intensity normalization strategy is the most robust and it is capable of determining the geographical location of webcams with an accuracy of approximately 2 degrees.

1 Introduction

Geographical information is important in many domains including image collection management. Due to the low cost of storage and wide availability of digital cameras multimedia collections are growing at an explosive rate. The organization and effective retrieval of images in such large collections is an active research area. Image browsers, such as Picasa, are commonly organizing images into folders with names given by the authors, or chronologically according to date [7] or other EXIF-coded information [8]. Often, image collections are too large to be manually labeled. One avenue of research is looking at developing games that makes it fun to label image collections collaboratively [3, 4]. Images can also be organized according to geographical information [1], but still very few cameras are equipped with GPS hardware capable of capturing geographical information. Consequently, few images are geo-coded. Several researchers have therefore looked at other ways of extracting geographical information from images. In one

This paper was presented at the NIK-2010 conference; see http://www.nik.no/.

ambitious attempt landmark recognition is used to deduce the location of images [16]. Other approaches are inspired by the sextant where direct sun height observations are used to deduce the geographical information [2, 15] but these strategies require information about the lens optics and either a very wide angle lens or low sun elevations. Because of the problems associated with obtaining direct sun observations it has also been proposed to indirectly measure the sun elevation using the shadows cast by objects to obtain geographical information [9, 10]. However, the automatic detection of shadows and their respective objects is an open research problem. It has also been shown that longitudinal information with an accuracy of less than 30 degrees can be obtained simply by looking at the mean timestamp for a 24-hour window of images [11]. Next, geographical information has also been obtained for image collections using overall brightness signatures as an indirect measure of sun elevation [12]. Webcams is another image source. There are thousands of webcams available worldwide continuously sampling images that are immediately available for download over the Internet. Often, the locations of these webcams are provided, but in some instances the geographical location is not provided. Moreover, one may wish to corroborate the location information. Strategies have been proposed to deduce the relative geographical location of a network of webcams [5, 6]. This study takes this one step further by deducing the absolute location of a single webcam by employing a simple celestial model.

2 Method

The following sections outline the proposed strategies including the celestial model used to determine the geographical locations and the two sunrise/sunset detection strategies for estimating the sunrise/sunset parameters.

Celestial model

The strategies presented herein are based on a celestial model where the latitude and longitude of a stationary webcam are deduced from sunrise and sunset estimates. First the latitude of the webcam is found by solving the following equation for latitude φ:

$$\cos(a_{sunset}) = \frac{\sin(-0.83) - \sin(\varphi)\sin(\delta)}{\cos(\varphi)\cos(\delta)} \tag{1}$$

where δ is the declination of the sun. The declination of the sun is related to the Earth's tilt which is 23.5 degrees and is the reason for the seasons. During the winter solstice the declination is equal to the Earth's tilt such that the southern hemisphere is more exposed to the sun. During the summer solstice the declination also equal to the Earth's tilt, but with maximum solar exposure on the northern hemisphere. During the spring and autumn equinoxes the tilt angle is perpendicular to the earth-solar vector and the declination is therefore zero. A good approximation of the cyclic sun declination is given by

$$\delta = -0.4092797 \cos\left(\frac{2\pi}{365}(d+10)\right) \tag{2}$$

where d is the day of the year with the first day being 1. Next, the angular sunset a_{sunset} is defined as

$$a_{sunset} = \frac{\pi}{12}(t_{sunset} - t_{midday}) \tag{3}$$

where t_{sunset} is the time of the sunset in universal time and t_{midday} is the time of the solar noon, given by

$$t_{midday} = \frac{t_{sunset} - t_{sunrise}}{2} \tag{4}$$

and $t_{sunrise}$ is the time of the sunset in universal time. Next, the longitude λ of the webcam is simply found using

$$\lambda = 2\pi \frac{12 - t_{midday}}{24} \tag{5}$$

The parameters used in the celestial model are summarized in Figures 1 and 2.

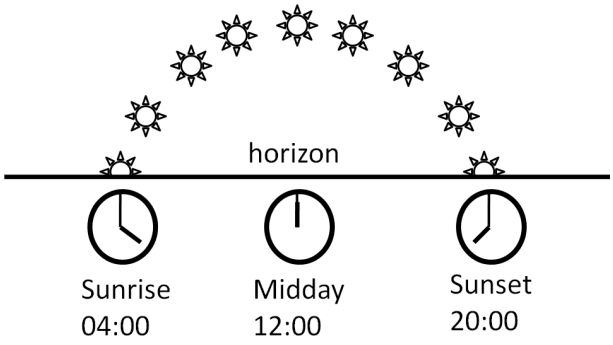

Figure 1: Sunrise, midday and sunset parameters.

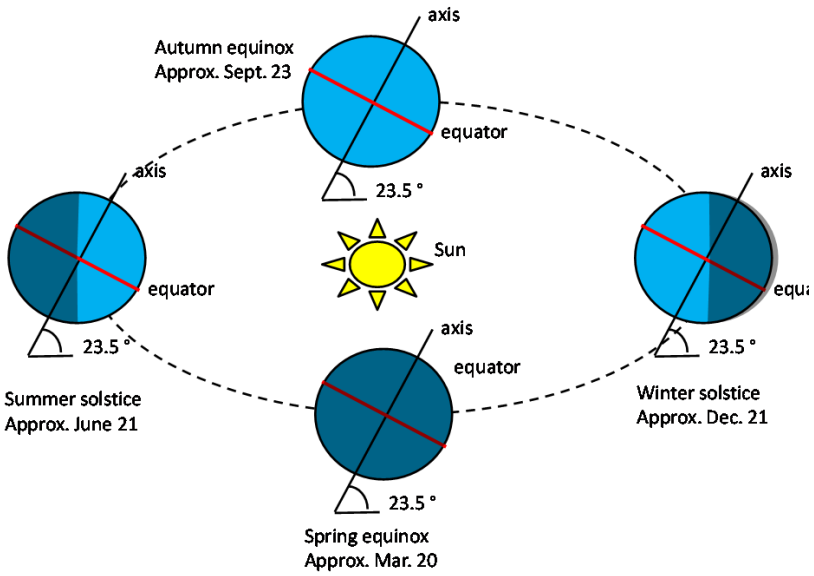

Figure 2: Declination of the sun.

Sunrise/sunset detection - normalized intensity

The problem in this study is to accurately determine $t_{sunrise}$ and t_{sunset} based on a series of time-stamped image samples. It is slightly related to the indoor-outdoor image classification problem [13, 14]. A sunrise occurs when the scene changes from dark to bright and a sunset occurs when the image scene changes from bright to dark. Usually, daybreak is preceded by a gradual increase in brightness up until the point when the sun is visible and the brightness has reached a stable level. Similarly, a sunset is often proceeded by a gradual decrease in brightness from the maximum brightness to darkness. This transition typically takes around an hour depending on the location and the season. Clearly, at night it is dark and at day it is bright, and the brightness of an image can be computed as

$$s = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} I_{x,y} \tag{6}$$

Were $I_{x,y}$ is the intensity of pixel x, y which is directly computed from the rgb components as follows

$$I_{x,y} = \frac{I_{x,y,r} + I_{x,y,g} + I_{x,y,b}}{3} \tag{7}$$

A low intensity value indicates night and a bright value indicates day. However, what is bright and what is dark depends on the scene, the lighting conditions, camera orientation in relation to light sources, the lens optical characteristics and the camera image sensor sensitivity. To reduce the impact of such factors it is proposed that the image values are linearly normalized. Given an image from a 24-hour collection with intensity s_i then the normalized intensity s'_i is

$$s'_i = \frac{s_i - s_{min}}{s_{max} - s_{min}} \tag{8}$$

Where s_{min} and s_{max} are the images in the 24-hour sample with the smallest and largest intensities, respectively. Moreover, to further enhance the extremes of night and day the following noise-damping transformation is used

$$f(x) = \left[\frac{1}{2} sin\left(\pi x - \frac{\pi}{2}\right) + \frac{1}{2} \right] \tag{9}$$

This transformation moves values less than 1/2 closer to 0 and values above 1/2 closer to 1. The noise-damping transformation function is plotted in Figure 3.

Sunrise/sunset detection - brightness ratio

The second hypothesis of this study is that the spatial intensity balance in an image can be used as a measure to classify night and day images. This measure is based on the assumption that the ground and close objects are located at the bottom part of the image and that the far away sky is located in the upper part of the image. During the day the sky will be much brighter than most objects and the ground, while during night the sky is dark and the ground may be lit. The top-bottom intensity balance measure is therefore introduced, namely

$$b(I) = \frac{p_{top}(I)}{p_{bottom}(I)} \tag{10}$$

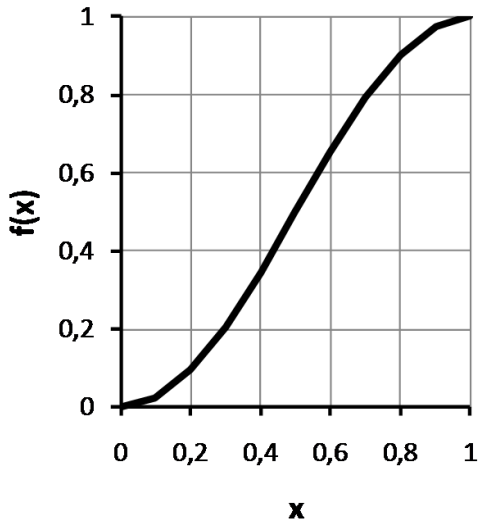

Figure 3: The noise-damping transformation function.

where

$$p_{top}(I) = \frac{3}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y/3} I_{x,y} \tag{11}$$

and

$$p_{bottom}(I) = \frac{3}{XY} \sum_{x=1}^{X} \sum_{y=\frac{2}{3}Y}^{Y} I_{x,y} \tag{12}$$

In simple terms the brightness ratio is based on computing the brightness of the top third of the image in relation to the bottom third of the image, where the brightness is simply summing the intensity values of the pixels in the given region. The above principle is illustrated in Figure 4 where the first illustration (a) is a daytime scene. On a brightness scale from 0 to 10 where 0 is total darkness and 10 is the brightest, the brightness ratio is 10/5 which equals 2. Illustration (b) shows the scene just before sunset and the brightness ratio is 5/4. Illustration (c) shows the scene during sunset with a brightness ratio of 4/3. Illustration (d) shows the scene after sunset with a brightness ratio of 2/2 which equals 1. Illustration (e) shows the scene approximately 30 minutes after sunset and the brightness ratio is 1/2. Finally, illustration (f) shows a nighttime scene with a brightness ratio of 0/2 which equals 0. This strategy is independent of the optical sensor characteristics such as sensitivity to light, etc, as such specific characteristics are cancelled by the ratio. During day the sky will always be brighter than the ground. During night it is assumed that the scene is somewhat lit as otherwise there would be no point having the webcam installed. However, at night the sky will be dark and the ground brighter than the sky since it is closer to the webcam and lit up. This assumption ensures that the intensity of the lower part of the image is never zero which will make the ratio cause a division by zero. Should the intensity of the lower part be zero then one is guaranteed that the captured image represents night.

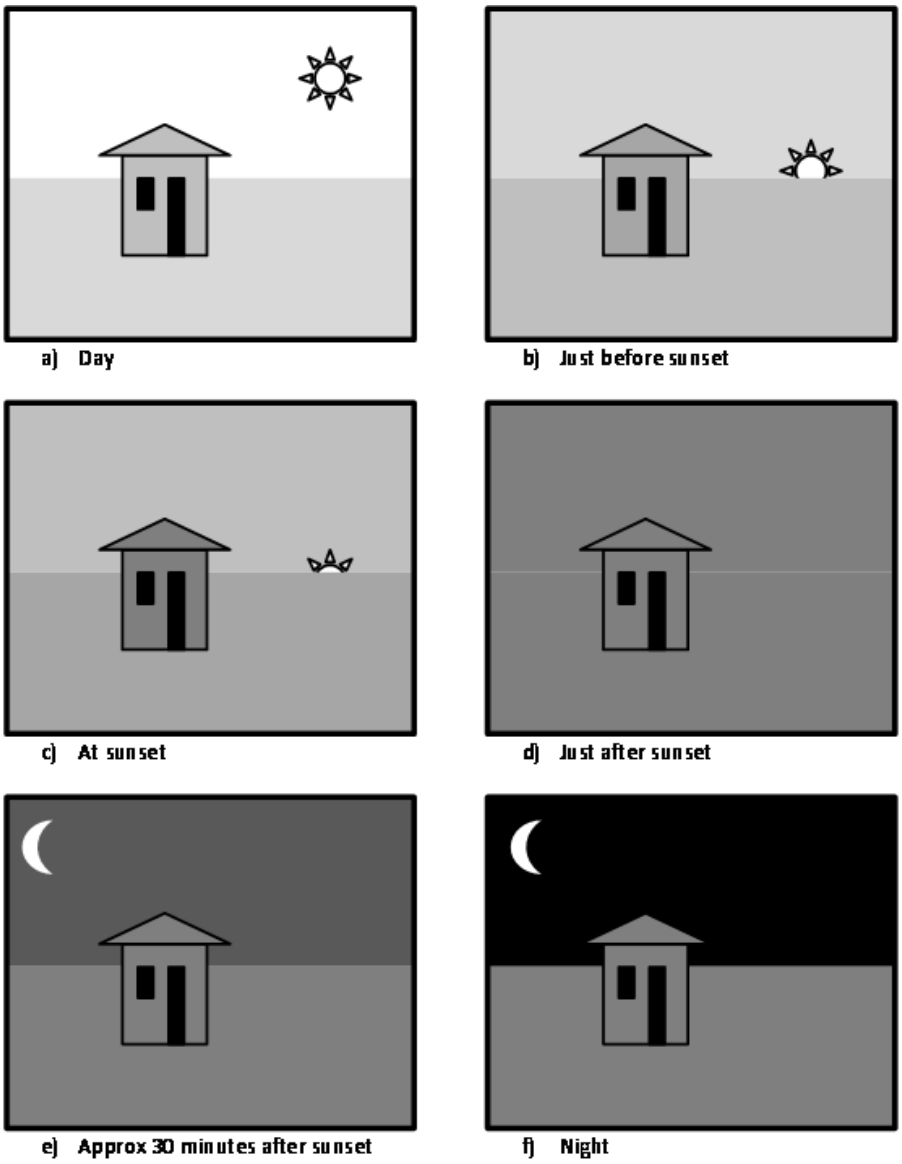

Figure 4: Characteristics of outdoor webcam images during a sunset.

3 Experimental evaluation

Testsuite

The webcam images used in this study are taken from the AMOS webcam image repository [5]. The collection comprises regularly sampled sequences of images taken by four webcams over the duration of five days starting June 2, 2006. The images are recorded in Central Time which equals UTC-5 or UTC-6. The four webcams are located in the United States but only the exact location of one webcam is known, namely New York, USA, overlooking the statue of liberty. An example day and night image for the four webcams are shown in Figure 5.

Sunrise/sunset detection

Figure 7 shows the normalized (blue) and noise-damped transformed intensity curves (red) for the four webcams during June 2, 2006. The curves in Figure 6 reveal that the

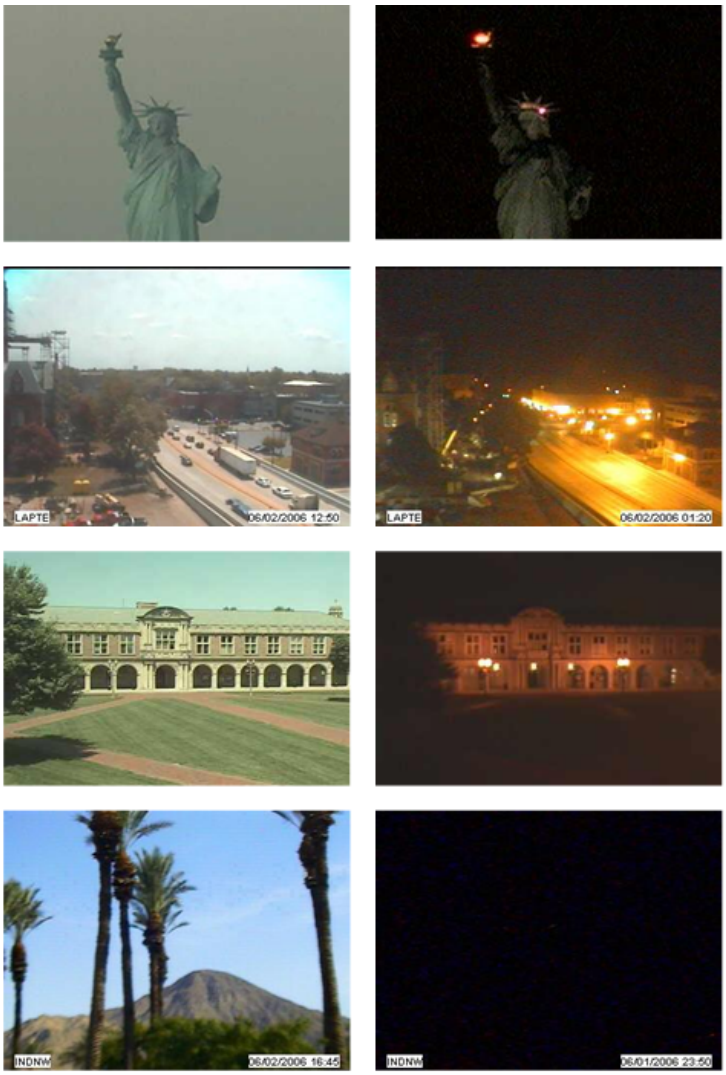

Figure 5: Daytime and nighttime scenes captured with the three webcams labeled 15, 190, 4 and 82 in the AMOS repository.

normalization procedure works well as the intensity of the four webcams becomes easily comparable. It is quite evident from the plot where the sunrise and sunset points are and where the night and day segments are located. The transformed plots reveal that the transformation greatly reduces high frequency noise in the data and makes the distinction between night and day more distinct. In this study a 80

Figure 8 shows the brightness ratio plots for the same data set as just described. All the plots besides the first plot has a distinct low-high-low shape where the sunrise is identified as occurring when the ratio is increasing rapidly and the sunset is revealed by a rapid drop in brightness ratio. All of these three plots seem to stabilize with a brightness ratio of about 2 during the day, that is, the top part or sky, is approximately twice as bright as the ground represented by the bottom part. The night is slightly more varied as both the second and forth plots seem to have a nightly brightness ratio of approximately 0.5 signaling that the ground is twice as bright as the sky. However, the third plot has a nightly brightness value of approximately 1 signaling that both the ground and the sky have similar brightness. Figure 4 reveals that this is because the ground in the lower part

Figure 6: A sunrise caught by a webcam (webcam 4).

of the image is not lit up at night, just the building in the middle of the image. The first plot in Figure 8 has a different pattern than the three others. It starts with a chaotic pattern with large oscillations, followed by a more gentle sequence with smaller oscillations, and the plot ends with large oscillations. The entire plot is centered just above a brightness ratio of 1. This indicates that the sky and the ground have similar brightness during both the day and night. Figure 4 reveals that this is because there is no ground. Although one may get a rough idea about the sunrise and sunset locations from the plot due to the large oscillations that occurs during the night this graph suggests that the brightness ratio is not suitable for this type of image and that one instead need to resort to the overall brightness value.

The chaotic nature of the brightness ratios for this webcam is revealed in Figure 9 which shows three consecutive snapshots taken at 2:00, 2:14 and 2:25. The images reveal that the webcam in the first image is zoomed in on the statue of liberty with the brightest parts towards the bottom of the image, the second image is zoomed out with the brightest part towards the top of the image and the third image zoomed far out revealing the base of the statue and with the brightest parts towards the center of the image. Clearly, the brightness ratio curves are more chaotic and noisy than the intensity plots and it is much harder to clearly identify the sunrise and sunset points from these plots. In comparison it is much easier to identify the sunrise and sunset points in the normalized transformed intensity plots. For webcam 15 overlooking the statue of liberty it is nearly impossible to identify the sunrise and sunset points while with the normalized intensity measure this plot was probably the most easy to decipher. The conclusion is therefore that the intensity measure should be employed and the brightness ratio rejected.

Geo-information extraction

Table 1 lists the results achieved using the normalized noise-damped intensity strategy across five days, namely mean estimated sunrise and sunset times (in decimal format),

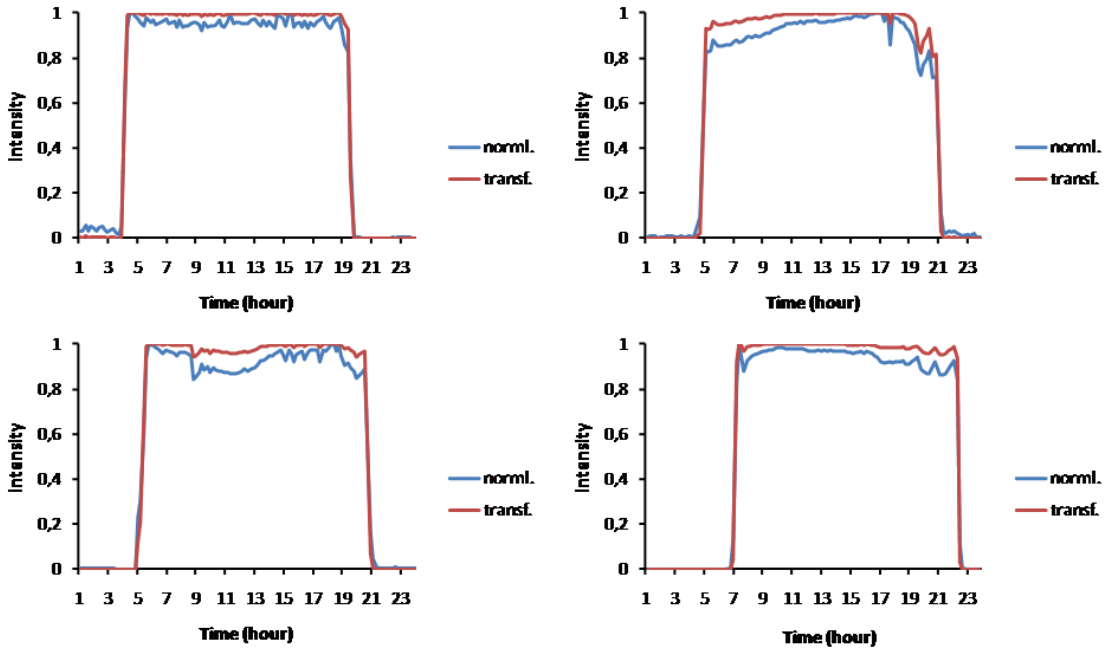

Figure 7: Normalized and transformed intensity plots during a 24 hour cycle for the four webcams 15, 190, 4 and 82 during June 2, 2006.

Table 1: Results obtained with the four webcams.

webcam	sunrise		sunset		latitude		longitude	
	mean	SD	mean	SD	mean	SD	mean	SD
15	4.29	0.15	19.45	0.18	42.6	2.3	73.1	1.2
190	5.00	0.06	20.75	0.06	47.4	0.5	88.1	0.9
4	5.54	0.11	20.49	0.07	40.8	0.9	90.2	1.2
82	7.24	0.07	22.26	0.10	41.4	1.4	116.3	0.8

and mean estimated latitude and longitude. Webcam 15 is located in New York City, USA which geographical coordinates are 40.7 degrees north and 74.0 degrees west. At June 2nd, 2006 the local sunrise and sunset times were 4.43 and 19.35, respectively. Two online solar calculators were used to verify these values. During this season the local time in New York City corresponds to UTC+5. The mean sunrise is 8 minutes early and the sunset is 6 minutes late yielding a day length error of 14 minutes. Next, the estimated latitude and longitudes for the New York City webcam is 42.6 degrees north and 73.1 degrees west, respectively. The latitudinal error is thus 1.9 degrees, and the longitudinal error is 0.9 degrees. The results for the three cameras positioned at unknown locations all have small standard deviations suggesting that the results are consistent. Google maps was used for plotting the extracted coordinates. These plots revealed that webcam 190 with coordinates (47.4, -88.1) was at Eagle Hauk, Michigan, United States, webcam 4 with coordinates (40.8, -90.2) was at Maquon, Illinois, United States and webcam 82 with coordinates (41.4, -116.3) was at Humboldt National Forrest, Nevada, United States. The correctness of these estimations has not yet been confirmed.

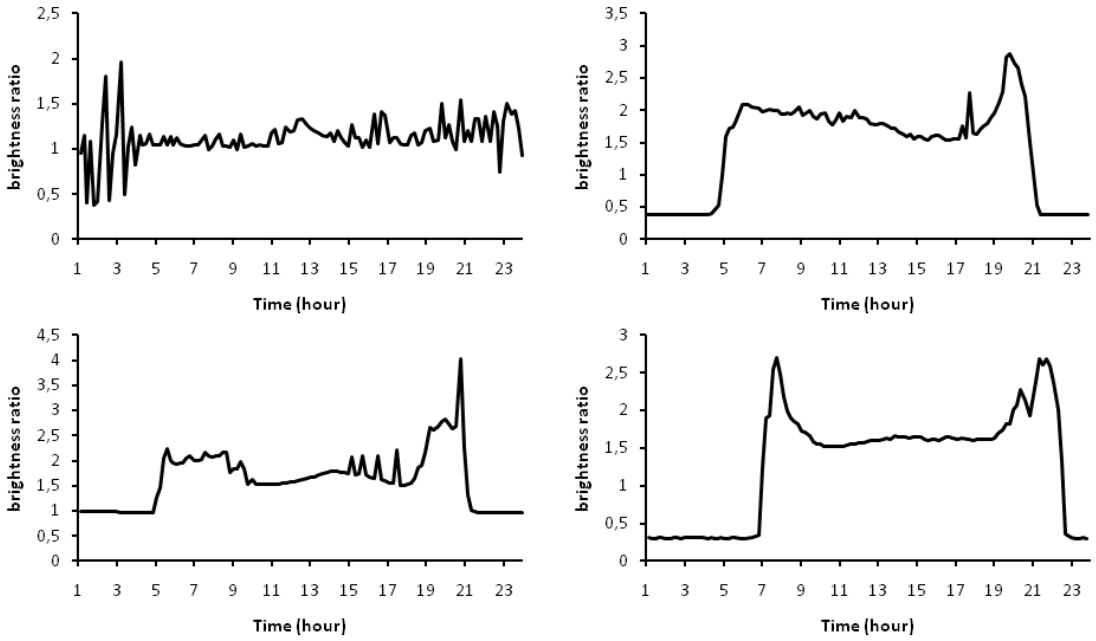

Figure 8: Brightness ratios during a 24 hour cycle for the four webcams 15, 190, 4 and 82 during June 2, 2006.

Figure 9: The Webcam 15 zooms in and out on an object resulting in measurement noise.

4 Conclusions

Two novel strategies for estimating the approximate geographical location of webcams were investigated. One strategy is based on a normalized and noise-damped image intensity contour of regular sequences of webcam images. The second method uses a brightness ratio which is computed as the ratio between the intensity of the upper third of the image compared to the intensity of the bottom third of the image. Both strategies are computationally effective and can be made even more effective by sampling fewer data points randomly dispersed across the same image areas. Experimental evaluations revealed that the normalized and noise-damped intensity contour is robust and allows the sunrise and sunsets to be easily identifiable, while the brightness ratio contours contained much more noise and it was much harder to identify sunrises and sunsets. Using the normalized noise-damped intensity contours a latitudinal and longitudinal accuracy of about 2 degrees was achieved on the dataset studied. It is likely that the accuracy can be further increased with image sequences sampled at higher rates.

References

[1] S. B. Barnes. Douglas Carl Engelbart: Developing the Underlying Concepts for Contemporary Computing. *IEEE Annals of the History of Computing 19:3*, 16-26. 16-26, 1997.

[2] Carboni, D., Sanna, S., and Zanarini, P. GeoPix: image retrieval on the geo web, from camera click to mouse click. *in the proceedings of Proceedings of the 8th conference on Human-computer interaction with mobile devices and services, 2006, pp. 169-172.*

[3] Cozman, F. and Krotkov, E. Robot localization using a computer vision sextant, *in the proceedings of IEEE International Conference on Robotics and Automation, 1995, pp. 106-111.*

[4] Diakopoulos, N. and Chiu, P. Photoplay: A collocated collaborative photo tagging game on a horizontal display. *in the proceedings of UIST '07, 2007, pp. 53-54.*

[5] Golder, S. and Huberman, B. A. The structure of collaborative tagging systems. *Journal of Information Sciences, 32, 2* (2006), pp. 198-208.

[6] Jacobs, N., Roman, N., and Pless, R. Toward Fully Automatic Geo-Location and Geo-Orientation of Static Outdoor Cameras. *in the proceedings of IEEE Workshop on Applications of Computer Vision, 2008, pp. 1-6.*

[7] Jacobs, N., Satkin, S., Roman, N., Speyer, R., and Pless, R. Geolocating Static Cameras. *in the proceedings of IEEE 11th International Conference on Computer Vision (ICCV 2007), 2007, pp. 1-6.*

[8] Jang, C.-J., Lee, J.-Y., Lee, J.-W., and Cho, H.-G. Smart Management System for Digital Photographs using Temporal and Spatial Features with EXIF metadata. *in the proceedings of 2nd International Conference on Digital Information Management, 2007, pp. 110-115.*

[9] Romero, N. L., Chornet, V. V. G. C. G., Cobos, J. S., Carot, A. A. S. C., Centellas, F. C., and Mendez, M. C. Recovery of descriptive information in images from digital libraries by means of EXIF metadata. *Library Hi Tech, 26, 2* (2008), pp. 302-315.

[10] Sandnes, F. E. Determining the Geographical Location of Image Scenes based on Object Shadow Lengths. *Journal of Signal Processing Systems for Signal, Image, and Video Technology,* (2010), in press.

[11] Sandnes, F. E. Sorting holiday photos without a GPS: What can we expect from contents-based geo-spatial image tagging? *Lecture Notes on Computer Science, 5879*, (2009), pp. 256-267.

[12] Sandnes, F. E. Unsupervised and Fast Continent Classification of Digital Image Collections using Time. *in the proceedings of ICSSE 2010, 2010, IEEE Press, pp. 516-520.*

[13] Sandnes, F. E. Where was that photo taken? Deriving geographical information from image collections based on temporal exposure attributes. *Multimedia Systems, 16, 4-5* (2010), pp. 309-318.

[14] Serrano, N., Savakis, A., and Luo, A. A computationally efficient approach to indoor/outdoor scene classification. *in the proceedings of 16th International Conference on Pattern Recognition, 2002, pp. 146-149.*

[15] Szummer, M. and Picard, R. W. Indoor-outdoor image classification *in the proceedings of IEEE International Workshop on Content-Based Access of Image and Video Database, 1998,* pp. 42-51.

[16] Trebi-Ollennu, A., Huntsberger, T., Cheng, Y., and Baumgartner, E. T. Design and analysis of a sun sensor for planetary rover absolute heading detection. *IEEE Transactions on Robotics and Automation, 17, 6* (2001), pp. 939 - 947.

[17] Zheng, Y.-T., Ming, Z., Yang, S., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.-S., and Neven, H. Tour the world: Building a web-scale landmark recognition engine. *in the proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), 2009, pp. 1085 - 1092.*