Heliometric Stereo: Shape from Sun Position

Austin Abrams, Christopher Hawley, and Robert Pless

Washington University in St. Louis
St. Louis, USA

Abstract. In this work, we present a method to uncover shape from webcams "in the wild." We present a variant of photometric stereo which uses the sun as a distant light source, so that lighting direction can be computed from known GPS and timestamps. We propose an iterative, non-linear optimization process that optimizes the error in reproducing all images from an extended time-lapse with an image formation model that accounts for ambient lighting, shadows, changing light color, dense surface normal maps, radiometric calibration, and exposure. Unlike many approaches to uncalibrated outdoor image analysis, this procedure is automatic, and we report quantitative results by comparing extracted surface normals to Google Earth 3D models. We evaluate this procedure on data from a varied set of scenes and emphasize the advantages of including imagery from many months.

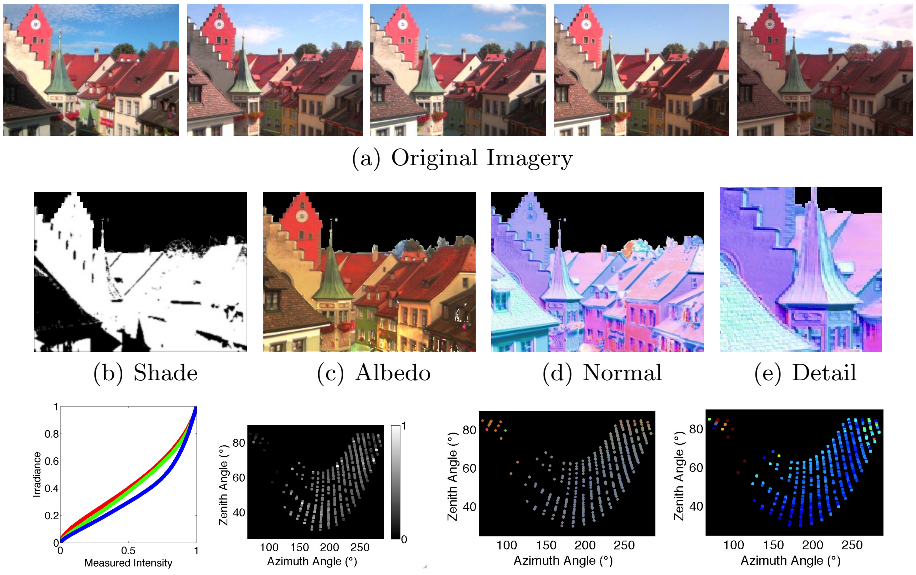
Key words: Photometric stereo, webcams, camera response

1 Introduction

This paper presents an approach to heliometric stereo — using the sun as a moving light source to recover surface normals of objects in an outdoor scene. This is a classic application of photometric stereo because the position of the sun is known very accurately, but made challenging because of variations in lighting and weather. Additionally, most long term imagery is captured by webcams that may not share geometric or radiometric calibration information. Thus, we explore what it would take to fully automate the solution to the photometric stereo problem for uncalibrated, outdoor cameras.

Our approach is to optimize the similarity between long term time-lapse imagery and an image formation model that includes the surface normal and albedo for each pixel in the scene, color and intensity of the ambient and direct lighting terms in each frame, and a shadow mask in each frame. This optimization results in estimates of all these parameters of the scene structure, albedo, lighting and camera properties. Although the WILD database [1] has sparked a rich recent literature in algorithms to understand outdoor image time-lapses, this paper is unique in tractably computing explicit, geo-referenced surface normals without any user interaction.

Alternative approaches without user interaction parameterize pixels with respect to time-series basis functions which have an unknown, non-linear relationship to surface normals [2], or cluster pixels groups that have approximately the



(a) Original Imagery

(b) Shade (c) Albedo (d) Normal (e) Detail

(f) Response (g) Ambient (h) Color (i) Exposure

Fig. 1. Given a sequence of geo-located, timestamped imagery taken over the span of several months (a), we recover a shadow mask for each image (b) (here shown as the mask for the leftmost original image), albedo (c), a dense normal map represented in East-North-Up Coordinates (d) (detail image shown in (e), and colormap described in Figure 2), nonlinear camera response (f), and for each image, measures of ambient lighting (g), light color (h), and exposure (i). We show the temporal variables (g)-(i) as a function of the image's solar azimuth and zenith angle. Best viewed in color.

same, but unknown surface normals [3]. Approaches that compute metric surface normals require user interaction to identify three pixel locations on orthogonal surfaces [4], and all these approaches result in surface normals in the coordinate system of the camera, not a geo-referenced coordinate system.

Thus, our approach is appropriate to deploy on very large webcam archives [2, 5] in order to extract quantitative remote sensing measurements without human input. Despite the allure of using already emplaced webcams for environmental monitoring, the few published uses of uncalibrated webcams are limited to simple color changes [6, 7]. We hope that our approach to automatically estimate scene shape, scene lighting and surface reflectance will support a larger range of environmental measurement uses of these webcam archives.

There are three major contributions of this work. First, we adapt the photometric stereo algorithm [8] to work for outdoor scenes by integrating a richer image formation model. We present a gradient descent approach and methods for initialization and regularization of this optimization. Second, we test this across a variety of types and scales of natural scenes and highlight the ability to capture

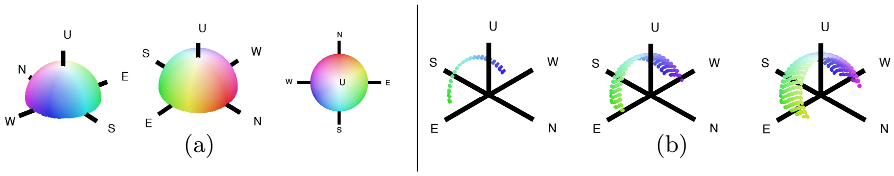


Fig. 2. (a) The colormap used for all normals in this paper. Notice that the normals are represented in an absolute, East-North-Up reference frame, rather than with respect to the camera's optical axis. (b) Input solar lighting directions for one camera, using progressively longer spans of time (1 month, 3 months, and 6 months, from left to right). The longer the input sequence, the richer the set of lighting directions.

very small scale surface structure We report surface normals geo-referenced to an "East-North-Up" coordinate system, and we are the first to offer quantitative comparisons between our results and 3D geometry from Google Earth. These highlight both the accuracy of our results and limitations in the completeness and resolution of the ground truth.

Third, we characterize performance under which this approach gives good results. This emphasizes the importance of using imagery from many months, because over the course of one day or one week, the solar path does not give sufficient constraints to recover surface normals (see Figure 2(b)).

2 Background and Related Work

The original constraints for photometric stereo [8] assumed multiple images of scenes with known lighting directions captured with a calibrated camera. While there is immense literature on variations of this problem, in this review we consider only those most relevant to our problem domain of long term, outdoor, webcam imagery. Specifically, our imagery is captured with known lighting directions because the primary light source is the sun whose direction can be computed given a timestamp and geolocation of the camera. However, lighting may vary due to weather conditions, and the camera may perform some unknown nonlinear radiometric distortion before publishing the image.

Papers that address uncalibrated photometric stereo have primarily concentrated on indoor scenes lit by unknown lights. One approach clusters pixels into groups with the same albedo and surface normal to provide constraints to solve for the unknown radiometric response and the generalized bas-relief ambiguity [9]. Another works with constant albedo objects and uses non-linear optimization to solve for the radiometric response, lighting directions, and surface normals [10]. Because these focus on indoor images (in a dark room), they have relatively noise-free images and no ambient light term.

There is less work on outdoor, uncalibrated photometric stereo. Shen and Tan [11] explored the solution for photometric stereo from unstructured internet imagery, combining multi-view stereo to register images from different viewpoints. Assuming the surface albedo and surface normals are constant at a scene location between images, an optimization was used in order to estimate the lighting parameters and therefore infer the weather of each image.

A related approach [12] focused on a photometric stereo variant for recovering surface orientation from webcam images. This method also exploited multi-view stereo to solve for scene geometry, which was then used in a photometric stereo by example approach, which transfers normals from locations with known surface normals to the rest of the scene. Our approach does not require or use such an example, and is applicable to time-lapse captured from a single location.

In [13], Sunkavalli et al. factor a time-lapse sequence of images taken from a single day into several meaningful components, including shadows, albedo, and a one-dimensional surface normal projection. We follow this work towards a more flexible model that accounts for solar variations over the span of a year, which uncovers the full three-dimensional normal field. In addition, we incorporate the effects of nonlinear camera response.

Kim et al. [14] present an approach to take a series of images taken through the span of a day and use changing light to infer radiometric calibration and exposure values for each image. Using a PCA basis for camera response functions introduced by Grossberg and Nayar [15], Kim et al. recover the response function that is most consistent with a Lambertian assumption. We take a similar approach toward solving the exposure and radiometric curves, but we simultaneously solve for surface normals in the process.

Recently, Ackermann et al. [16] describe an approach to perform photometric stereo on outdoor webcams. While they use a richer reflectance model to explicitly handle specularities, we offer a model which allows a trackable optimization, giving our algorithm an 8x speedup. We also perform a more rigorous evaluation, by comparing surface normals to Google Earth models and producing 3D shapes, and use a shadow estimation procedure appropriate for months of imagery.

Given a sequence of images taken under partly-cloudy conditions and an estimate of wind direction and speed, Jacobs et al. [17] provide an algorithm to extract shape based on cues derived from observing cloud shadows through time. While Jacobs et al. focus on time-lapse sequences on the scale of minutes or hours, we take advantage of the changing position of the sun over the span of many months. This increased timespan relaxes the assumption that the image was taken under specific meteorological conditions, and solves for surface normals rather than depth.

3 Method

This section presents our outdoor image formation model, and our methods for solving for the parameters of that model. The inputs to this model are timestamped images, and the direction to the sun computed from the timestamp and geolocation. Our algorithm extracts several components of our image formation model, including surface normals and photometric properties of the camera.

3.1 Model

The input consists of n images $I_1, \ldots, I_i, \ldots, I_n$, each represented as a p-pixel vector in $[0, 255]^p$. Given the latitude and longitude of the camera, as well as the timestamps from each image, we use [18] to determine the sun direction for each image $L_1, \ldots, L_n$, represented as a 3-vector in the East-North-Up coordinate frame. To keep consistent notation, temporally-indexed variables are annotated with the subscript i from 1 to n, and spatially-indexed variables with the subscript $\mathbf{x}$ from 1 to p.

The Lambertian lighting model assumes that the intensity of a pixel $\mathbf{x}$ in image i is a function of the surface normal $N_\mathbf{x} \in \mathbb{R}^3$, the albedo $\rho_\mathbf{x} \in \mathbb{R}$, and lighting direction $L_i \in \mathbb{R}^3$. More formally, the intensity of image i at pixel $\mathbf{x}$ is given by the linear model $\rho_\mathbf{x} L_i^\top N_\mathbf{x}$.

Webcam imagery from real scenes is also affected by ambient lighting conditions, which we model by including $a_i \in [0, 1]$ for each image i in sequence. Natural scenes also include both cast shadows, such as a tree shadow on a road, and attached shadows such as the side of a building that is not illuminated. This is modeled with a per-pixel-per-image shadow volume $S_{i,\mathbf{x}} \in [0, 1]$, that models how much light is received from the sun at each time at each pixel.

$$\rho_\mathbf{x}(L_i^\top N_\mathbf{x} S_{i,\mathbf{x}} + a_i) \tag{1}$$

The reported pixel intensity of a camera also depends on the radiometric camera calibration. Webcams rarely publish results in RAW format, rarely include meta-data describing their radiometric response, and may not be accessible to allow radiometric calibration using images of known calibration objects. Thus, we must solve for, and include in our model, an unknown, monotonic, nonlinear photometric response function $f : [0, 255] \rightarrow [0, 255]$, which we assume to be fixed through all the images in sequence, and we modulate each image by an exposure value $e_i \in \mathbb{R}$:

$$I_{i,\mathbf{x}} = f(e_i \rho_\mathbf{x}(L_i^\top N_\mathbf{x} S_{i,\mathbf{x}} + a_i)) \tag{2}$$

Here, we take advantage of the invertibility of f and rewrite the above as

$$f^{-1}(I_{i,\mathbf{x}}) = e_i \rho_\mathbf{x}(L_i^\top N_\mathbf{x} S_{i,\mathbf{x}} + a_i). \tag{3}$$

To make the estimation of f tractable, we use the Empirical Model of Response introduced by Grossberg and Nayar [15]. This is a PCA basis for typical camera response functions, which models the nonlinear response function f^{-1} as a linear composition of nonlinear bases:

$$f_v^{-1}(x) = f_0^{-1}(x) + \sum_{j=1}^{b} f_j^{-1}(x) v_j \tag{4}$$

where f_0^{-1} is the mean inverse response curve, $f_1^{-1}, \ldots, f_b^{-1}$ are the set of basis curves, v_j are the unknown, camera-specific coefficients that implicitly describe the shape of the nonlinear function, and b is the number of bases used (in our experiments, we use $b = 5$). We use f_v^{-1} as notation to define the radiometric response defined by a vector of coefficients v.

Color To incorporate color, we largely use the above model independently on each color channel. Allowing different exposures for each color channel can be interpreted as modeling the color and intensity of the sunlight for each image. For each image we have a single term for ambient intensity a_i, which models the ambient light as the same color as the direct light, but allows variation in its intensity. We also have a single normal $N_{\mathbf{x}}$ for each pixel, and a single shadow volume $S_{i,\mathbf{x}}$. Unless otherwise specified, we do not denote each color channel individually, and in a slight abuse of notation, treat Equation 3 as an equality over a three-element vector.

3.2 Optimization

Given a set of images and their lighting directions, we wish to extract each component of the lighting model. Our algorithm is a simple gradient-descent procedure that minimizes the following loss function:

$$\operatorname*{argmin}_{v,e,\rho,N,a} \frac{1}{pn} \sum_{\mathbf{x}=1}^{p} \sum_{i=1}^{n} ||f_v^{-1}(I_{i,\mathbf{x}}) - e_i \rho_{\mathbf{x}} (L_i^\top N_{\mathbf{x}} S_{i,\mathbf{x}} + a_i)||^2. \tag{5}$$

Notice that we do not solve for the shadow volume $S_{i,\mathbf{x}}$. Similar to Sunkavalli et al. [13], we first estimate a shadow volume and leave it fixed for the remainder of the optimization. See Section 3.3 for our shadow estimation approach.

To enforce a physically-based lighting model, we impose additional constraints on the variables in Equation 5. We constrain that the albedo and ambient lighting are bounded between 0 and 1, that the normals are unit length, that the exposures are all non-negative, and that the response function is monotonic (which can be expressed as a linear inequality constraint over v; see [15]).

To prevent overfitting the response curve, we employ a smoothness regularization term on the response function that penalizes large changes across intensity bins:

$$R_v = \sum_{i=1}^{255} \left|\left|f_v^{-1}(i) - f_v^{-1}(i-1)\right|\right|^2, \tag{6}$$

which is expressed as a system of linear equations over the curve coefficients v.

The overall optimization is therefore

$$\operatorname*{argmin}_{v,e,\rho,N,a} \frac{1-\lambda}{pn} \sum_{\mathbf{x}=1}^{p} \sum_{i=1}^{n} ||f_v^{-1}(I_{i,\mathbf{x}}) - e_i \rho_{\mathbf{x}} (L_i^\top N_{\mathbf{x}} S_{i,\mathbf{x}} + a_i)||^2 + \frac{\lambda}{255} R_v \tag{7}$$

subject to the constraints listed above. Here, λ is a regularization constant that defines the weight of satisfying the data term versus the smoothness term.

To make this optimization tractable, we take an alternating minimization strategy, which minimizes the normals and albedos in one step, then all other variables in another step, and repeats until convergence:

$$\underset{\rho,N}{\operatorname{argmin}} \frac{1-\lambda}{pn} \sum_{\mathbf{x}=1}^{p} \sum_{i=1}^{n} ||f_v^{-1}(I_{i,\mathbf{x}}) - e_i \rho_{\mathbf{x}} (L_i^\top N_{\mathbf{x}} S_{i,\mathbf{x}} + a_i)||^2 \tag{8}$$

$$\underset{v,e,a}{\operatorname{argmin}} \frac{1}{pn} \sum_{\mathbf{x}=1}^{p} \sum_{i=1}^{n} ||f_v^{-1}(I_{i,\mathbf{x}}) - e_i \rho_{\mathbf{x}} (L_i^\top N_{\mathbf{x}} S_{i,\mathbf{x}} + a_i)||^2 \quad + \frac{\lambda}{255} R_v \tag{9}$$

This formulation has substantially smaller computational overhead, because Equation 8 can be broken into p independent subproblems, one for each pixel. Although neither of these subproblems are convex, there are immediate linear approximations to each of these problems which can be solved by least squares.

Linear Approximation to Equation 8 We cannot simultaneously solve for ρ and N and achieve a convex solution, because the surface normal N must be unit length. We can rewrite the lighting model from Equation 3 as:

$$f^{-1}(I_{i,\mathbf{x}}) - e_i \rho_{\mathbf{x}} a_i = e_i L_i^\top (\rho_{\mathbf{x}} N_{\mathbf{x}}) S_{i,\mathbf{x}}. \tag{10}$$

Therefore, we approximate a solution by fixing the $\rho_{\mathbf{x}}$ on the left-hand side of the above equation as the last iteration's approximation for the albedo, then solve for the unconstrained 3-element vector $\rho_{\mathbf{x}} N_{\mathbf{x}}$ on the right hand side via least squares for each color channel, where the magnitude of the resulting variable is the albedo. Since this allows a surface normal for each color channel, we repeat this optimization by fixing the albedo and solving for the single best normal over all color channels via least squares, and then scaling the albedos so that each normal is unit length. Since Equation 8 is independent for each pixel, this optimization can be done for each pixel in parallel.

The above method makes the assumption that the albedo does not change dramatically from iteration to iteration. Therefore, when we minimize Equation 8, we make use of a learning rate that encourages a slow change in the albedo and normal. If ρ' and N' are the solution to the linear approximation, then we update the estimates of the albedo and normal for the next iteration as:

$$\rho \leftarrow \rho' \tau + \rho(1-\tau) \tag{11}$$

$$N \leftarrow \frac{N'\tau + N(1-\tau)}{||N'\tau + N(1-\tau)||}. \tag{12}$$

Each time we minimize Equation 8, we repeat the solve-and-update procedure 20 times, using a learning rate of $\tau = 0.1$.

Linear Approximation to Equation 9 Because the ambient lighting term is single-channel, Equation 9 is nonconvex. However, if we separate the ambient lighting into three channels, then we can solve for the radiometric parameters v, the ambient light in each channel a, and the exposure e. An auxiliary variable

$\hat{e}_i = e_i a_i$ transforms Equation 9 into a linear system with respect to unknowns $v, e_i, \hat{e}_i$:

$$f_0^{-1}(I_{i,\mathbf{x}}) + \sum_{j=1}^{b} f_j^{-1}(I_{i,\mathbf{x}})v_j = e_i \rho_{\mathbf{x}} L_i^\top N_{\mathbf{x}} S_{i,\mathbf{x}} + \hat{e}_i \rho_{\mathbf{x}}. \tag{13}$$

Notice that we can keep the constraint $a_i \leq 1$ by use of the linear constraint over the auxiliary variable as $\hat{e}_i \leq e_i$.

This solves for an independent ambient lighting term for each color channel. However, after fixing all other terms, the optimal single-channel a can be found through a simple closed-form solution.

The full algorithm therefore initializes its albedo, normals, ambient light, exposures, and nonlinear response and iterates between the two linear approximations until convergence.

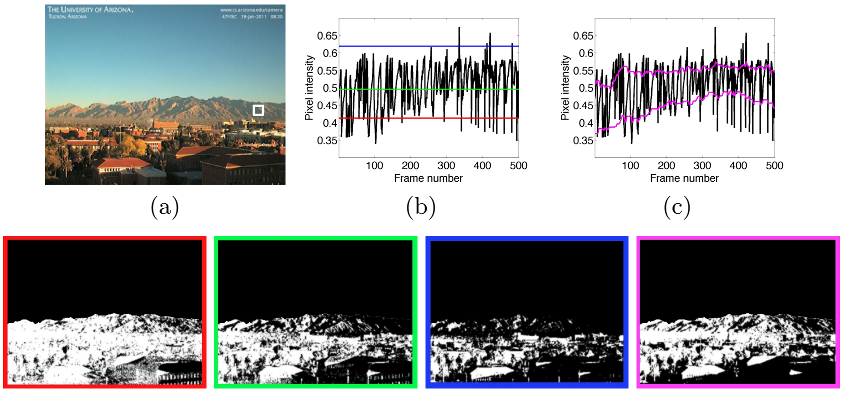
Ambiguity As presented, there is an ambiguity between the exposures and albedo, in that we can double the exposure and halve the albedo image, and keep the same reconstruction. We fix this ambiguity by scaling each term so that the mean exposure is 255.

3.3 Shadow Estimation

To initialize the shadow volume, we initially tried the shadow estimation algorithm from Sunkavalli et al. [13], which uses a per-pixel threshold to define at which times that pixel is in shadow. This threshold is defined as the median of the bottom 20% of intensities ever seen by that pixel, multiplied by some constant (Sunkavalli et al. use 1.5).

While this approach works well for one day's worth of video, over the span of a year, pixel intensities vary drastically as the position and intensity of the sun change; a shadowed pixel when the sun is highest in the sky may have the same intensity as a directly-lit pixel at the opposite time of year, and so there usually is not a single correct threshold.

We modify this thresholding technique by introducing an adaptive, per-pixel threshold which changes over time. For each pixel $\mathbf{x}$, we use two centroids $s_{\mathbf{x}}$ and $l_{\mathbf{x}}$ that model the intensity of that pixel in shade and direct sunlight, respectively. We initialize $s_{\mathbf{x}}$ and $l_{\mathbf{x}}$ similarly to Sunkavalli et al., where $l_{\mathbf{x}}$ is the median of the top 5% of pixels, and $s_{\mathbf{x}}$ is the median of the bottom 5% of pixels. Then, for each image i in chronological order, we compute whether the original image pixel $I_{i,\mathbf{x}}$ is closer to $s_{\mathbf{x}}$ or $l_{\mathbf{x}}$, set its corresponding pixel in the shadow volume, and update that threshold as $s_{\mathbf{x}} \leftarrow 0.8 s_{\mathbf{x}} + 0.2 I_{i,\mathbf{x}}$ or $l_{\mathbf{x}} \leftarrow 0.8 l_{\mathbf{x}} + 0.2 I_{i,\mathbf{x}}$. In this way, as we loop over all times i, from 1 to n, the values of $l_{\mathbf{x}}$ and $s_{\mathbf{x}}$ rise and fall as needed. Figure 3 shows that this method produces better shadow volumes than Sunkavalli et al. when working over the span of many months. As in [13], we perform bilateral filtering [19] to smooth the shadow images.



(a) (b) (c)

(d) (e) (f) (g)

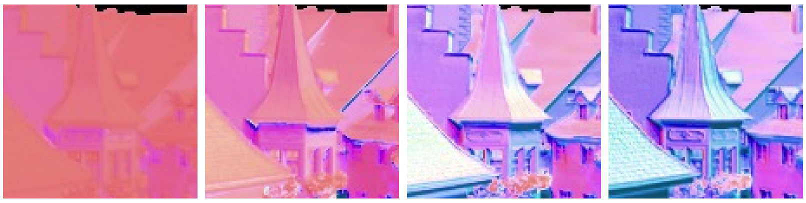
Fig. 3. (a) An example image. The pixel trajectory for the pixel centered in the white box is shown in (b), along with three scalar multiples of the threshold generated by the shadow estimation method of Sunkavalli et al. [13]. The blue line is the threshold suggested in [13]. (c) The centroids used in our adaptive approach. (d)-(g) The resulting shadow images from each approach for the image in (a), where the color of the border indicates which thresholding technique was used.

3.4 Implementation

Although we can decompose our optimization into a sequence of convex steps, the overall optimization is nonconvex and is subject to initialization. We initialize our variables rather simply: the normals begin as the all-up vector, the ambient intensity is 1 for all images, the exposure is 255 for each color channel, the response function is linear (i.e., we choose the v so that $f_v^{-1}(x) = x$), and we initialize the albedo as the mean in-shade image.

In all of the experiments used in this paper, we use $n = 500$ images. When selecting images, it's important to select from a wide range of lighting angles, but not to include any times of day when the sun is in view, producing lens flare effects in the image. Furthermore, we aim to select images that are the least overcast or hazy. Therefore, we begin by sampling 1000 lighting directions uniformly from the set of sun illumination directions the camera observes.

Because many cameras experience large numbers of cloudy days, and direct sunlight is important to constrain surface normals, we use a heuristic to encourage selection of sunny images. For each of the 1000 target lighting directions, we consider all images whose lighting direction is within 3 degrees of that target, and select the image with the largest saturation. From these 1000 candidate images that span the observed lighting directions, we select the 500 most-saturated images, in order to remove times of day which are consistently hazy, cloudy, or when the sun is in view.



(a) 1 month (b) $1\frac{1}{2}$ months (c) 2 months (d) 4 months

Fig. 4. The normal maps recovered from our algorithm as the duration of original imagery ranges from 1 to 4 months.

For all experiments, we manually mask off the sky, timestamps, streets, bodies of water, pathways, or other areas prone to transient objects. Optimization is performed only on the remaining pixels. We use $\lambda = 0.9$ in Equation 9.

Runtime depends on the size of the image and number of pixels, but we report timing for the results in Figure 1. On a machine with dual 2.66GHz processors with 6 cores each and 12GB of RAM, the entire inference process from loading imagery to completion takes 62 minutes on a sequence with 500 images and 224,052 non-sky pixels using our MATLAB implementation. We perform 10 iterations of alternating descent for each experiment. The main bottleneck is in solving for albedo and normals, which takes about four and a half minutes.

We emphasize that this data comprises 500 images captured over many months, so our algorithm is substantially faster than real-time.

4 Results

As discussed briefly in the introduction, photometric stereo requires lighting to come from a diverse set of directions. To evaluate its importance, we ran our algorithm on one camera and progressively increased the duration of the input sequence from one week to a few months, starting in August 2011. Figure 4 shows that only once we have about six weeks of imagery can we begin to uncover three-dimensional normals, and the quality of the normals increases as more imagery is included. Depending on the times of year used, even longer lengths of time may be required to reliably extract usable normals (i.e., a week near an equinox produces a larger solar cover than a week at a solstice).

Figure 5 shows the results of our algorithm on real-world cameras. We reliably recover shadows, response functions, and surface normals for a variety of scenes from the AMOS dataset [2].

4.1 Evaluation

To evaluate our approach, we compare our surface normals to the normals from Google Earth models. Using the interface from [20], we geo-calibrated two webcams and generated normals from the surrounding Google Earth geometry.

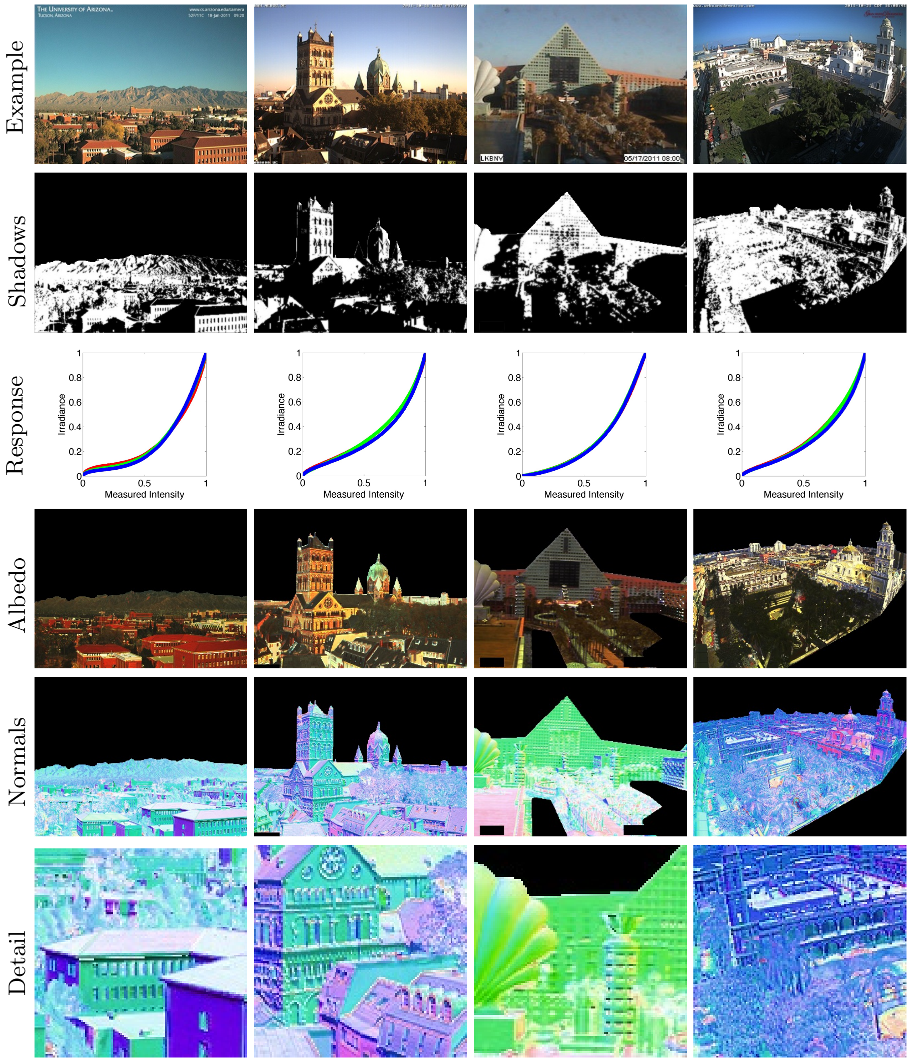


Fig. 5. Results on a variety of cameras. From top to bottom, we show an example image, the recovered shadow mask for that image, nonlinear camera response curves, albedo, surface normals, and a crop of the surface normal image to show detail. Notice that the shadow mask accurately captures large-scale shadows as well as small-scale details. We reliably recover physically-meaningful, nonlinear response curves, yet our model is flexible enough to allow variety of real-world responses. Because we allow each pixel a unique normal, we produce high-fidelity normals that capture the tiny changes in surface orientation due to windows and the detailed geometry of trees.

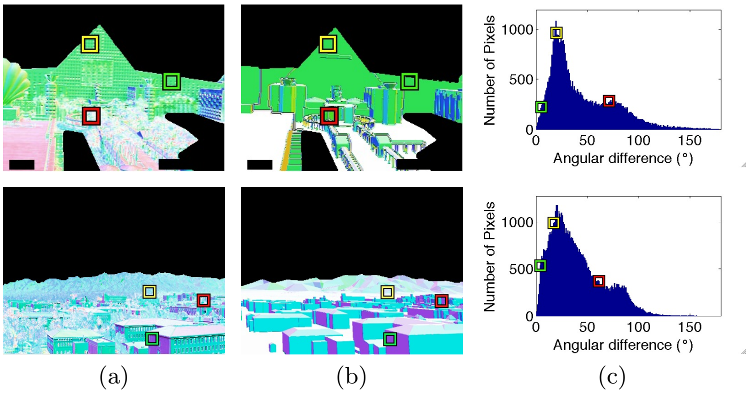


Fig. 6. Comparing our results to the models from Google Earth. (a) Our recovered surface normals. (b) The surface normals from Google Earth. (c) A histogram of the angular difference, in degrees, between our model and the Google model. The green, yellow, and red squares show locations on the images and histograms with 5, 20, and 70 degrees of error.

Figure 6 shows a quantitative evaluation of our approach, by measuring the angular difference between our normals and the normals from Google Earth. These histograms show that many locations in the image have substantial angular error. However, the colored squares help parse this difference, and show that in locations where the Google Earth geometry is an accurate reflection of the scene geometry, the angular error is on the order of 5 degrees, and that larger angular errors occur on objects not modeled by Google Earth (e.g. trees), or modeled as low resolution polygons (e.g. mountains in the distance).

4.2 Depth from Normal Field Integration

The normals from photometric stereo are often computed as an intermediate step toward inferring a depth map. However, the standard normal field integration equations make the assumption that the normals are represented in a coordinate frame relative to the camera (i.e., the optical axis is the Z-axis).

Since our algorithm returns surface normals in a geo-referenced coordinate system, we perform a search over pan and tilt angles of possible camera rotations, choosing the rotation that yields the best resulting integration error; see [21] for a more detailed discussion on normal field integration. Example results for image regions that consist of objects with interesting shapes are shown in Figure 7.

5 Conclusions and Future Work

We have presented an image formation model rich enough to capture variations in outdoor webcam imagery over long time periods. We acknowledge that there

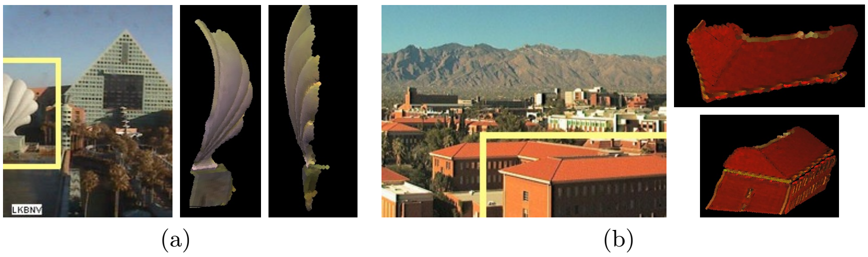


Fig. 7. 3D reconstructions of objects from webcams in the wild. In each example, we show an original image, and two novel views of the reconstructed object. The shape of the seashell in (a) is captured nicely (and shown from a very different viewpoint), and the rooftop of the building in the other scene is accurately reconstructed with a right angle when viewed from overhead in the top right of (b).

are many additional components that affect outdoor imagery, many of which we could add within our formation model and optimization. Over long time periods, both the surface normal and the albedos can change (buildings may undergo construction, snow may fall, trees may change color) and this could be captured by solving for normal maps or albedos with constraints on how they vary through time. The lighting model could be extended to include haze or non-uniform ambient light. The surface reflectance model can be modified to explicitly include a specular component [16], or we could use estimators for the surface normal that are robust to non-Lambertian effects [22].

Understanding how to incorporate additional terms within a tractable optimization scheme is something that we look forward to pursuing. We believe that this current work offers one step in that direction, by being the first to tractably return real-world surface reflectance properties from uncalibrated images, which could be used toward more reasoned environmental monitoring.

To facilitate future comparative studies, our data, code, and ground truth are available at `research.engineering.wustl.edu/~abramsa/heliometric` .

References

1. Narasimhan, S., Wang, C., Nayar, S.: All the Images of an Outdoor Scene. In: Proc. European Conference on Computer Vision. Volume III. (2002) 148–162
2. Jacobs, N., Roman, N., Pless, R.: Consistent temporal variations in many outdoor scenes. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2007)
3. Koppal, S.J., Narasimhan, S.G.: Clustering appearance for scene analysis. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2006)
4. Sunkavalli, K., Romeiro, F., Matusik, W., Zickler, T., Pfister, H.: What do color changes reveal about an outdoor scene? In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2008)

5. Lalonde, J.F., Efros, A.A., Narasimhan, S.G.: Webcam clip art: Appearance and illuminant transfer from time-lapse sequences. ACM Transactions on Graphics (SIGGRAPH Asia 2009) **28** (2009)
6. Jacobs, N., Burgin, W., Fridrich, N., Abrams, A., Miskell, K., Braswell, B.H., Richardson, A.D., Pless, R.: The global network of outdoor webcams: Properties and applications. In: Proc. ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL GIS). (2009)
7. Riordan, E., Graham, E., Yuen, E., Estrin, D., Rundel, P.: Utilizing public internet-connected cameras for a cross- continental plant phenology monitoring system. In: Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International. (2010) 1501 –1504
8. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineerings **I** (1980) 139–144
9. Shi, B., Matsushita, Y., Wei, Y., Xu, C., Tan, P.: Self-calibrating photometric stereo. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2010) 1118–1125
10. Mongkulmann, W., Okabe, T., Sato, Y.: Photometric stereo with auto-radiometric calibration. In: ICCV Workshops. (2011) 753–758
11. Shen, L., Tan, P.: Photometric stereo and weather estimation using internet images. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2009)
12. Ackermann, J., Ritz, M., Stork, A., Goesele, M.: Removing the example from example-based photometric stereo. In: ECCV Workshop on Reconstruction and Modeling of LargeScale 3D Virtual Environments. (2010)
13. Sunkavalli, K., Matusik, W., Pfister, H., Rusinkiewicz, S.: Factored time-lapse video. Proc. ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH) **26** (2007) 101–111
14. Kim, S.J., Frahm, J.M., Pollefeys, M.: Radiometric calibration with illumination change for outdoor scene analysis. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2008)
15. Grossberg, M.D., Nayar, S.K.: What is the space of camera response functions? In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Volume II. (2003) 602–609
16. Ackermann, J., Langguth, F., Fuhrmann, S., , Goesele, M.: Photometric stereo for outdoor webcams. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2012)
17. Jacobs, N., Bies, B., Pless, R.: Using cloud shadows to infer scene structure and camera calibration. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2010)
18. Reda, I., Andreas, A.: Solar position algorithm for solar radiation applications. In: NREL Report No. TP-560-34302. (2003)
19. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. IEEE International Conference on Computer Vision. (1998)
20. Abrams, A., Pless, R.: Webcams in context: Web interfaces to create live 3D environments. In: Proc. ACM SIGMM International Conference on Multimedia (ACMMM). (2010) 331–340
21. Zhang, L., Curless, B., Hertzmann, A., Seitz, S.M.: Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multiview stereo. In: Proc. IEEE International Conference on Computer Vision. (2003)
22. Wu, L., Ganesh, A., Shi, B., Matsushita, Y., Wang, Y., Ma, Y.: Robust photometric stereo via low-rank matrix completion and recovery. In: Proc. Asian Conference on Computer Vision. (2010)