Photometric Stereo for Outdoor Webcams

Jens Ackermann Fabian Langguth Simon Fuhrmann Michael Goesele
TU Darmstadt

Abstract

We present a photometric stereo technique that operates on time-lapse sequences captured by static outdoor webcams over the course of several months. Outdoor webcams produce a large set of uncontrolled images subject to varying lighting and weather conditions. We first automatically select a suitable subset of the captured frames for further processing, reducing the dataset size by several orders of magnitude. A camera calibration step is applied to recover the camera response function, the absolute camera orientation, and to compute the light directions for each image. Finally, we describe a new photometric stereo technique for non-Lambertian scenes and unknown light source intensities to recover normal maps and spatially varying materials of the scene.

1. Introduction

In the last years, researchers have discovered online image collections as a valuable data source for various applications. Examples include multi-view stereo [7, 8], segmentation [33], and reflectance recovery [10]. While image collections are certainly important, time-lapse sequences from webcams are another type of data suited for computer vision research, yet much less explored. Despite recent work in the area, we believe that this type of data is much less understood than online image collections. In particular, to our knowledge, no complete photometric stereo approach has been demonstrated on outdoor webcam data.

This is somewhat surprising since, conceptually, it should be easy to apply photometric stereo on outdoor webcam images, e.g., by combining a robust photometric stereo approach with an illumination estimation technique. In fact, Brooks and Horn [4] proposed already in 1985 an algorithm to recover surface orientation in a theoretical outdoor setting for a Lambertian object. Likewise, Sato and Ikeuchi [29] proposed a photometric stereo technique operating under solar illumination. Still, we are unaware of any reports that these methods or any other photometric stereo approach developed for lab settings have been successfully applied to the challenging case of outdoor webcam data. We believe that the reason for this is that naïvely applying existing techniques to such datasets will lead to bad results. In particular, any successful method needs to address the following challenges that are unique to outdoor webcams:

- Selection of suitable images from the time-lapse sequence. Especially bad weather images need to be rejected, and object illumination needs to be considered.
- Calibration of webcam and image data. In particular, subpixel image alignment, radiometric and photometric calibration, recovery of the lighting directions and intensities, and shadow detection need to be performed.
- Application of a suitable and robust photometric stereo technique handling spatially-varying, non-Lambertian surface reflectance.

In this paper, we present a complete system that addresses these challenges. We first select suitable images by analyzing the sky and object regions of the webcam images. Our main goal is to select daytime images where the object is directly illuminated by the sun. We then align the selected images with sub-pixel accuracy (compensating for any slight movements of the camera), estimate the orientation of the webcam in geo-referenced coordinates, and detect shadowed pixels in each of the input images. Although we perform a radiometric calibration step in order to recover the response curve of the camera, the relative light intensities in the images are unknown, which not only depend on the sun's observed intensity (influenced by weather conditions), but also on the auto-exposure system of the webcam. We recover the relative light intensities within our photometric stereo approach for non-Lambertian surfaces.

The remainder of the paper is organized as follows: We first discuss relevant related work (Section 2) before we state our assumptions and give an overview of our approach (Section 3). We then describe our selection and calibration techniques in Sections 4 and 5, respectively. Details about the photometric stereo approach are given in Section 6. We then present our results (Section 7) and show some directions for future work (Section 8).

In Proc. of the *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, 2012.

2. Related Work

Outdoor Webcam Data: The *Archive of Many Outdoor Scenes* [18] contains many time-series of webcam data and gives insights into long term analysis of static webcams using PCA. Given a time-series for a webcam, several authors proposed techniques to calibrate various aspects of the camera. Kim *et al*. [21] recover the exposure and radiometric response functions from static cameras using the intensity variations of groups of pixels over time. Jacobs *et al*. [19] geo-locate webcams by combining appearance variations and geo-registered satellite imagery. Extending the idea of webcam geo-location, Lalonde *et al*. [23] developed a technique to determine the orientation of a static camera using the sky region in webcam images as calibration target. We assume that the location of the webcam is known and recover its absolute orientation using this technique.

There has been little work on recovering shape from webcam sequences. Jacobs *et al*. [16] find that the intensities of two scene points due to cloud shadows behave similarly over time. They exploit this property to infer depthmaps for static outdoor cameras providing images every few seconds. Unfortunately, many webcams are updated in larger intervals limiting the applicability of this approach. Koppal and Narasimhan [22] cluster scene normals in image sequences with unstructured light paths but do not reconstruct normal maps. Sunkavalli *et al*. [34] describe a technique to factor a time-lapse sequence into shadow, illumination, and reflectance. In later work, Sunkavalli *et al*. [35] analyze the color changes over time caused by the varying spectrum of sunlight and separate the effects of sun and ambient sky illumination. Both papers also reconstruct some information about the orientation of surfaces in the scene but are restricted to projections of the normals on the solar plane. In addition, they are unable to handle varying exposure of the images. In contrast, our approach allows for orientation recovery from outdoor webcam time-series data alone without being restricted to the solar plane while also taking varying sun intensity and exposure into account.

Photometric Reconstruction with Non-Lambertian Reflectance: There is a large body of work on photometric reconstruction, including [14, 37, 11, 3]. We restrict our discussion to algorithms that consider aspects especially important for outdoor webcams such as the ability to cope with spatially-varying, non-Lambertian materials and unknown light source intensities and camera response. Those aspects are addressed independently in different approaches, none of which fits our purpose without modification.

Shi *et al*. [32] use color profiles to determine the camera response curve and object surface orientation in the diffuse setting. Still assuming Lambertian scenes, Shen and Tan [31] extend photometric stereo to Internet images with varying viewpoints. Other methods which can handle unknown light conditions due to matrix factorization, but ignore camera properties, are for example [11, 26, 20]. We use Hayakawa [11] to initially estimate relative light intensities under the assumption of a Lambertian scene.

While all these techniques assume Lambertion reflectance, Hertzmann and Seitz [12] use example objects in the scene with known geometry to reconstruct objects with arbitrary, varying BRDFs. Ackermann *et al*. [1] combine this approach with multi-view stereo to replace the example object. Both methods do not need to consider the response curve of the camera since they are based on orientation consistency. Higo *et al*. [13] also exploit reflectance cues that are preserved during capturing, but can only handle surfaces with either diffuse or specular reflectance. Sato and Ikeuchi [29] apply photometric stereo on outdoor scenes by separating specular and diffuse components as well as sun and skylight contributions. To handle more complex reflectance properties, Goldman *et al*. [9] solve for both, geometric information and BRDFs by alternating between optimizing the shape given the BRDF model, and optimizing the BRDF given the optimized shape. Recovering shape and even non-parametric BRDFs is possible when using approximations to the reflectance function as proposed by Alldrin *et al*. [2] in a purely data-driven approach. However, they assume a certain circular light setup which is infeasible for sunlight illumination. Given the initial light source intensities from [11], we extend the approach by Goldman *et al*. [9] by additionally solving for the light source intensities and skylight contributions in each iteration.

Image Selection: A key issue when working with Internet data is the ability to select a suitable subset of images with beneficial properties. This insight has been used in several works on multi-view stereo [8, 7, 6]. Hornung *et al*. [15] present and analyze a 3-step selection process for multi-view stereo. Related to this is also the research in *view-planning* which was originally concerned with active sensor placement to ensure a certain coverage and quality of measurements [30]. Lensch *et al*. [24] speed up the acquisition of spatially-varying BRDFs with an algorithm that proposes view and light positions for efficient sampling. Most of these techniques cannot be applied to photometric stereo. The closest work in terms of image selection is Lalonde *et al*. [23]. They fit a quadratic function to the intensities in the sky, and compare the region to a smooth vertical gradient. In contrast, we propose to also consider the illumination of the objects in the scene.

3. Overview and Assumptions

Our algorithm is designed to process a large time-lapse image sequence from a static web camera. We assume that the GPS coordinates of the camera as well as per-image time stamps are given. To guide image selection and estimate

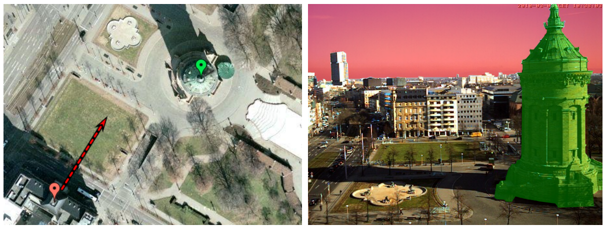

Figure 1. The geographic context of the scene with camera (red marker) and target object (green marker) on the left, and an example image from the camera with sky mask (red region) and object mask (green region) on the right. Sat image © by Google Inc.

the camera orientation, sky and object masks are provided to the system. This implies that the sky region is visible; if this is not the case, the camera orientation can also be specified manually. An example of the user input is given in Figure 1 (right). As in most photometric stereo reconstruction systems, we assume an orthographic camera model and treat the sun as distant point light source. Our algorithm can handle multiplicative lighting changes (e.g., changing exposure or varying sunlight intensity) as well as additive lighting phenomena (caused by ambient and indirect lighting).

The algorithm can be divided into two separate and independent steps (see Figure 2). The first step selects a small set of images suited for photometric stereo, removes pixel mis-alignments caused by camera movement, and computes the position of the sun (the light source) for each image. The camera response function is recovered and the pixel intensities are linearized. The selected images together with per-image sun positions are the input to the second stage: An initial Lambertian photometric stereo is performed and the result is iteratively optimized, yielding a normal map, a predefined number of basis materials, material maps that describe the per-pixel mixture of the fundamental materials, and an image that represents the indirect illumination.

4. Image Capture and Selection

Given a large set of webcam images taken over a long period of time we would like to select a small subset of 30 to 50 images well suited for reconstruction. (Although fewer images are sufficient for photometric stereo, more images increase robustness.) Our requirements for this subset are that the object of interest is directly illuminated by sunlight and that we observe sufficient variability in the incoming light directions. The latter can not be fullfilled by images taken in the span of a few days because the sun then moves on a single plane. We therefore capture images over the course of at least 6 months.

In order to recover the absolute camera orientation from the sky [23], we additionally need another subset of images with clear sky. Since manual selection is a tedious task, we present a novel algorithm to determine this subset automatically. This step requires a timestamp for each image and makes use of both the sky and the object mask to select suitable images.

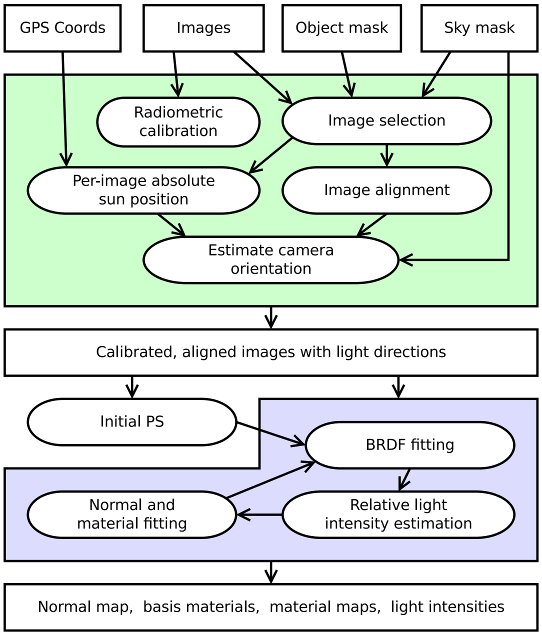

Figure 2. Algorithm overview. We first select suitable images (green) before we recover the final normal map, basis materials, corresponding material maps, and light intensities (blue).

We first compute a set of measures for each image which we will describe shortly. Based on these measures, the images are pre-filtered to remove unsuitable images. We then derive a score for the remaining images and greedily select the desired amount of images based on our measures.

Image Filtering: Images with more than 10% overexposed pixels in either the whole image or the object region are rejected immediately. We also reject images not taken during daytime where the sun zenith angle is larger than 85° as given by the Solar Position Algorithm [28]. To remove images with bad weather conditions, we exclude dark images. To this end, we compute the median intensity of the sky region I_{sky} and the 75th percentile intensity of the object region I_{obj} as defined by the masks. The latter rewards images where the object is well illuminated but also allows the object to be partially shadowed. We then discard 50% of the images with lowest score $S_I = I_{sky} + I_{obj}$.

Image Selection: From the remaining images, we now select two subsets, clear sky images for camera calibration and good-weather images with a well-illuminated object for photometric stereo, by combining several criteria.

We compute the average of the gradient magnitude in x- and y-direction G_{sky} for the sky region and G_{obj} for the object region. Low values of G_{sky} correspond to clear sky images without clouds. High values of G_{obj} indicate images with high-contrast and well exposed object region, as opposed to images where the object lies in fog. Evaluating V_{obj}, which is computed as the variance of all intensity values in the object region, has proven to be very useful to exclude images where the object appears diffuse (typically caused by clouds occluding the sun). These low-contrast images are particularly harmful for photometric stereo. A plausible good-weather indicator is the blue sky assumption. We calculate the average sky color $\bar{S}$ and measure the blueness of the sky as $B_{sky} = \bar{S}_B - \max(\bar{S}_R, \bar{S}_G)$, where $\bar{S}_{R,G,B}$ stands for the respective color channel.

Each of the quantities G^i_{sky}, B^i_{sky}, G^i_{obj}, and V^i_{obj} is normalized with respect to its minimum and maximum value over all images. We then compute two scores for image i,

$$\begin{aligned} S^i_{clearsky} &= (1 - G^i_{sky}) \cdot B^i_{sky} \cdot V^i_{obj} \cdot P^i && (1) \\ S^i_{PS} &= G^i_{obj} \cdot B^i_{sky} \cdot V^i_{obj} \cdot P^i, && (2) \end{aligned}$$

corresponding to the two image sets. P is a penalty value which is initially one. In a greedy search, we select the image with best score. After an image has been selected, we penalize images with similar sun position: We evaluate a 2D Gaussian function for each not yet selected image j

$$P^j := P^j - P^j \cdot e^{-\frac{(x_j - x_i)^2}{2\sigma_x^2} - \frac{(y_j - y_i)^2}{2\sigma_y^2}} \qquad (3)$$

where x_j and x_i correspond to the day of the year for image j and i, and y_j and y_i correspond to the minute of the day for image j and i respectively. We achieved good results with $\sigma_x = 10$ and $\sigma_y = 30$, which corresponds to a Gaussian spread of 10 days over the year and 30 minutes over the day. After penalizing, we select the next best image and iterate until the required number of images is selected (see Figure 3 for an exemplary selection on the Church dataset).

5. Image and Webcam Calibration

Image Alignment: Outdoor conditions such as strong winds can cause the webcam to shake, resulting in small camera motions. We apply a sub-pixel alignment step since even subtle mis-alignments have serious impact on the reconstruction quality. The dramatic variations in image appearance disqualifies most naïve methods. E.g., aligning gradient images directly using Lucas-Kanade [25] fails since the gradients are heavily influenced by varying lighting and shadowing. We instead align the gradient images to the average gradient image, calculated from all input images, similar to Jacobs *et al*. [17]. Figure 4 shows the alignment for about 50 images from the Church dataset.

Webcam Calibration: To linearize our input images we recover a response function for each webcam using the specialized approach of Kim *et al*. [21].

Afterwards, we obtain the sun position for each image using the Solar Position Algorithm [28]. This algorithm needs only a time stamp and a rough estimate of the camera's geo-location to recover the absolute zenith and azimuth angles of the sun. We then employ the method by Lalonde *et al*. [23] that uses the sky as calibration target to recover zenith and azimuth angles for the camera itself. We use these angles to transform the absolute position of the sun into the camera coordinate system for every image, which gives us the light direction L_i for photometric stereo.

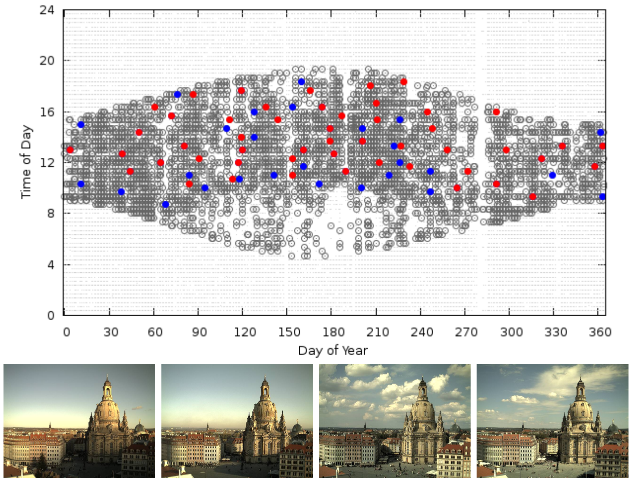

Figure 3. Image selection for the *church* dataset. The plot shows selected clear sky images (blue) and selected images for photometric stereo (red), all candidate images (dark gray) and all available non-overexposed images (light gray). Example images are shown on the bottom with two clear sky images (left) and two photometric stereo images (right).

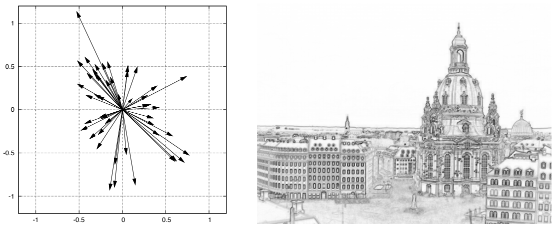

Figure 4. Motion vectors for image alignment (left): Each arrow corresponds to an image that has been aligned along the direction of its arrow. The axes show the distance of the alignment in pixels. One can see that alignment is very subtle, barely more than a single pixel. The average gradient image is shown on the right.

Shadow Detection: Shadows can provide a useful cue to infer scene structure [5]. In photometric stereo, they are, however, typically a source of additional errors. This is especially true for outdoor scenes as drop shadows dramatically change the appearance of the images.

To detect a shadowed pixel, we use the method originally proposed by Sunkavalli *et al*. [34]. The basic idea is that the pixel intensity $I_{i,p}$ of pixel p in the i-th image will be significantly lower when in shadow than when exposed to direct sunlight. If we do not observe an intensity difference of at least a factor of 1.4 between the highest and the lowest pixel value, we assume that the pixel was shadowed in all images. For each other pixel p we first calculate the median value $m_{\min}$ of the n smallest intensity values. The pixel is detected as shadowed in image i, if its intensity $I_{i,p}$ in that image is smaller than $K \cdot m_{\min}$. We use $n = 10\%$ of the total number of images and $K = 1.5$ in our examples.

6. Photometric Stereo

The image intensity at a pixel p and color channel c can be separated into a sun and sky contribution $I_{i,p,c} = I_{\text{Sun},i,p,c} + I_{\text{Sky},i,p,c}$. We model the reflectance function for a scene point observed at pixel p as a linear combination $\sum_{m=1}^{M} \gamma_{p,m} f_{m,c}$ of basis materials $f_{m,c}$. Furthermore, we assume the sun as a point light source $\delta(L - L_{i,\text{Sun}})$ from direction $L_{i,\text{Sun}}$ with intensity $\tilde{l}_{i,c}$. This yields:

$$I_{\text{Sun},i,p,c} = \tilde{l}_{i,c} \sum_{m=1}^{M} \gamma_{p,m} f_{m,c}(n_p, L_{i,\text{Sun}}) V_p(L_{i,\text{Sun}}), \quad (4)$$

where V_p encodes the portion of the sky visible from the scene point and n_p is the corresponding normal.

We approximate the sky as a spatially uniform light source of intensity $\widetilde{S}_{i,c}$ that is partially visible at a pixel p according to the visibility V_p:

$$\begin{aligned} I_{\text{Sky},i,p,c} &= \widetilde{S}_{i,c} \int \sum_{m=1}^{M} \gamma_{p,m} f_{m,c}(n_p, L) V_p(L)\, dL && (5) \\ &= \widetilde{S}_{i,c} \cdot \text{const}_p\,. && (6) \end{aligned}$$

If we further assume that the intensity of the sky scales linearly with the (observed) sun intensity, i.e. $\widetilde{S}_{i,c} \propto \tilde{l}_{i,c}$, we can factor this into per-image and per-pixel terms

$$I_{\text{Sky},i,p,c} = \tilde{l}_{i,c} \cdot S_{p,c}. \quad (7)$$

For webcams, camera parameters like exposure time and aperture size are usually unknown and may vary over time. This means the observed intensity in the image might differ from the true intensity in the scene by a scalar factor. We incorporate these effects by considering relative light intensities $l_{i,c}$ instead of absolute intensities $\tilde{l}_{i,c}$. This gives the final image creation model

$$\frac{I_{i,p,c}}{l_{i,c}} = \sum_{m=1}^{M} \gamma_{p,m} f_{m,c}(n_p, L_{i,\text{Sun}}) V_p(L_{i,\text{Sun}}) + S_{p,c}. \quad (8)$$

Our algorithm optimizes for the relative light intensities $l_{i,c}$ in each image, the basis materials $f_{1,c}, \ldots, f_{M,c}$ in the scene, the surface orientation n_p, and material mixing coefficients $\gamma_{p,m} \geq 0$ at each pixel. Sun visibility $V_p(L_{i,\text{Sun}})$ is handled by our shadow detection and we replace it by introducing the sets $\mathcal{I}_p$ of images such that p is not in shadow. We represent the materials using an isotropic Ward model [36] and encode the respective parameters in a vector α_m per basis material: $f_{m,c}(n_p, L_{i,\text{Sun}}) = f_c(n_p, L_{i,\text{Sun}}, \alpha_m)$.

6.1. Initialization

During initialization we ignore the sky term $S_{p,c}$ and set it to zero. In a first step, we assume constant light intensities and treat the whole scene as Lambertian, ignoring the errors that will arise if a more complex material is present. In our experiments, we found that all scenes contained enough points that conform to this simplification and result in decent initial estimates of surface normals and albedo. With this information we find an initialization of the relative light intensities and use them to re-estimate the normals, which considerably improves the first initialization. Finally, we cluster the resulting albedos and fit *pure* basis materials to get an initialization of the per-pixel reflectance properties. We now describe each of these in steps in more detail.

Classical Photometric Stereo: We solve for the normal $n_{p,c}$ and albedo $\rho_{p,c}$ in each color channel by minimizing

$$E(\rho_{p,c}, n_{p,c}) = \frac{1}{|\mathcal{I}_p|} \sum_{i \in \mathcal{I}_p} \left(I_{i,p,c} - \rho_{p,c}\, l_{i,c}\, n_{p,c} \cdot L_i\right)^2 \quad (9)$$

The relative intensities $l_{i,c}$ are set to 1.0 during initialization. If $\mathcal{I}_p$ contains too little variation in the light directions or if the least-squares error is too high, we ignore the pixel. For valid pixels we select the normal from the color channel with the lowest error.

Relative Light Intensities: For a first estimate of the relative light intensities we employ the approach by Hayakawa [11]. To overcome certain ambiguities this method needs at least six surface points with similar albedos and sufficient variation in normal direction. We determine these points automatically by clustering all pixels into four albedo clusters according to the albedo estimate from Equation (9). For the most frequent albedo we then cluster the normals of the set into 30 normal clusters. Finally, we select eight normals from different clusters and ensure that the corresponding pixels are almost never in shadow (see Figure 6 for an example). In both steps the clustering is done using Expectation Maximization for a Gaussian mixture model.

Initial Material Estimation: Following Goldman *et al.* [9], we use the albedos to compute an initial distribution of the fundamental materials in the scene. The number M of fundamental materials is chosen beforehand. (Typically, two or three materials are sufficient for accurate material reconstruction.)

While Goldman *et al.* suggest to cluster the albedos in the HSV color space, we found that clustering in sRGB color space gives better results for our datasets. Again, the clustering is based on a mixture of Gaussians. For each pixel, we assign its cluster weights as initial material combination $\gamma_{p,1}, \ldots, \gamma_{p,M}$ (normalized to $\gamma_{p,1}+\ldots+\gamma_{p,M} = 1$). Given these initial estimates for the mixing coefficients, we now find good initializations for the parameters of each material. For each $m \in \{1, \ldots, M\}$, we build a set of pixels $\mathcal{M}_m$ that represent the *pure* material. We do this by selecting pixels with the mixing coefficient for that material at least ten times greater than the others. Based on these pure pixels, we fit the BRDF parameters using non-linear Levenberg-Marquardt optimization [27], minimizing:

$$E(\alpha_m) = \sum_{c,p\in\mathcal{M}_m, i\in\mathcal{I}_p} \left(I_{i,p,c} - l_{i,c} f_c(n_p, L_{i,\text{Sun}}, \alpha_m)\right)^2 . \tag{10}$$

6.2. Iterative Refinement

We iterate the next steps to refine the normal map, light intensities, and materials. An overview of the entire algorithm is shown in Figure 2. As each step decreases the objective function given by Equation (12), the algorithm is guaranteed to converge. Depending on the scene 25-50 iterations were sufficient to reach convergence. After each of the following steps we update the current intensity estimate of pixel p in image i and color channel c which we denote

$$e_{i,p,c} = l_{i,c} \left(\sum_{m=1}^{M} \gamma_{p,m} f_c(n_p, L_i, \alpha_m) + S_{p,c} \right) . \tag{11}$$

Material Fitting: In Equation (10) we optimized parameters for each material separately. This already provides us with a good initial estimation. Now, we find the optimal parameters for all materials simultaneously, i.e., for the concatenation α of all parameter vectors α_m and not restricted to *pure* pixels $\mathcal{M}_m$. Given the current estimate for the normal map and per-pixel material weights, we minimize

$$E(\alpha) = \sum_{i\in\mathcal{I}_p,p,c} \left(I_{i,p,c} - e_{i,p,c}\right)^2 . \tag{12}$$

Light Intensity Optimization: To improve the relative intensities during our optimization we analytically solve for the best intensity update $U_{i,c} = l^{\text{new}}_{i,c} / l^{\text{current}}_{i,c}$ in every image, while keeping all other variables fixed. We want to minimize

$$E(U_{i,c}) = \sum_p \left(I_{i,p,c} - U_{i,c} e_{i,p,c}\right)^2 . \tag{13}$$

Setting the $\partial E / \partial U_{i,c} = 0$ yields the intensity update.

Material and Normal Map Optimization: The next step calculates the material weight maps, the per-pixel sky contribution $S_{p,c}$, and the normal for each pixel while material parameters and light intensities are fixed. For each pixel we minimize

$$\begin{aligned} E(n_p, S_p, \gamma_{p,1}, \ldots, \gamma_{p,M}) \\ = \sum_c \left(\sum_{i\in\mathcal{I}_p} \left(I_{i,p,c} - e_{i,p,c}\right)^2 + \lambda S_{p,c}^2 \right) . \end{aligned} \tag{14}$$

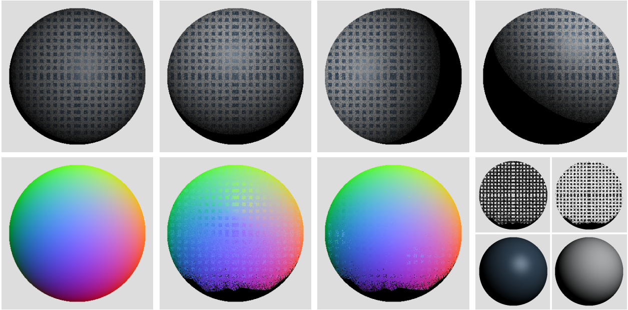

Figure 5. *Top row:* Input images with varying light intensities and randomly mixed materials. *Bottom row:* Ground truth normal map, initial normal map, normal map after optimization, and the final material maps with corresponding materials.

with a Levenberg-Marquardt optimization. Note that the material weights are no longer restricted to sum to one and that we include a penalty term on the skylight. In our experiments, we weight this term with $\lambda = 0.1$.

7. Results

Synthetic Data: We first evaluate our algorithm on a synthetic dataset with known ground truth. The dataset shows a sphere consisting of two materials, diffuse gray and specular blue. We explicitly used an orthogonal projection and parallel light rays from a distance point light source. Figure 5 shows the result of our algorithm for ten input images with varying light intensities and random indirect (additive) illumination per pixel. Additionally, around one third of the pixels are random mixtures between both materials.

The optimized normals show a significant improvement over the purely diffuse photometric stereo used for initialization. Some small errors remain if the optimization runs into wrong local optima, but we are able to properly recover the material parameters and maps.

Webcam Data: We now demonstrate our technique on three outdoor datasets, the *tower* dataset (≈20 k images), the *church* dataset (≈27 k images), and the *castle* dataset (≈20 k images). For each of the datasets we downloaded images every 20 to 60 minutes over the course of a year to get enough variation in the light directions. We then automatically selected 50 images for calibration and photometric stereo. Figure 9 shows some of these images for the *church* and *castle*. We confirmed the calibration results visually by projecting the viewing direction to the ground plane as, e.g., shown in Figure 1.

Figure 6 visualizes the shadow detection and some components of the initialization for the *tower* dataset. The tower consists of a diffuse base material and a more specular greenish material at the roof (corroded copper). The different materials can be clearly distinguished in our reconstruction, see Figure 7.

Figure 6. One input image, detected shadow regions, selected points for intensity estimation and the recovered object albedo.

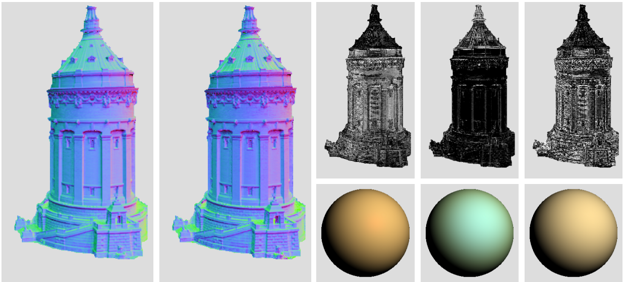

Figure 7. The initial normal map, the final normal map, and the four recovered BRDFs with corresponding material map.

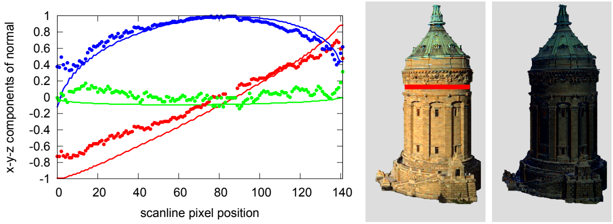

Figure 8. *Left:* Scanline through a cylindrical section of the *tower* showing the x (red), y (green), and z (blue) components of the normal vectors. Solid lines show normals from a reference cylinder and dots show the reconstruction of our algorithm. *Center:* The region used for the scanline marked in red. *Right:* Recovered contribution of the skylight.

The central part of the tower can be well approximated by a cylinder. Thus, to evaluate the performance of our technique, we rendered the normal map of a cylinder with corresponding radius and height as seen from a perspective camera . Figure 8 shows normals for one scanline from our reconstruction and the reference which we reproduce quite accurately. Most deviations occur at the far right and left where the surface is seen at gracing angles. In Figure 8 we also visualize the recovered sky term that shows the tower as seen without direct sun light.

The *church* dataset shows overall a similar behavior as demonstrated in Figure 10. While the initial normal map, reconstructed using the diffuse photometric stereo from Section 6.1, looks already promising, the final normal map exhibits more pronounced y directions of the normals in the dome region. The church is almost completely built from yellow sandstone which is also reflected in the three recovered material maps. Note that our method fits a fixed number of basis BRDFs that optimally explain the scene appearance. Since most scene points in this example have a similar albedo, the recovered materials essentially represent the specular variations.

Figure 9. Some input images of the *church* dataset, and one input image for the *castle* dataset (scanline region marked in red).

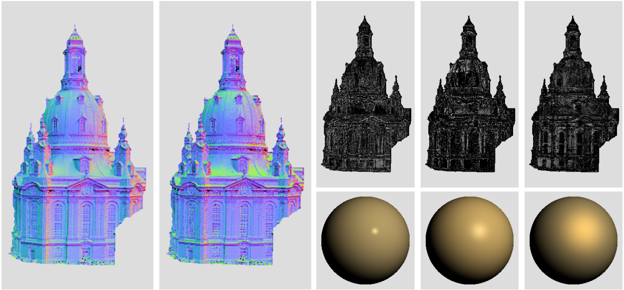

Figure 10. The initial normal map, the final normal map, and the three recovered BRDFs with corresponding material map.

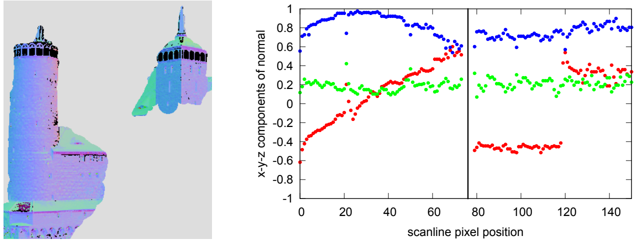

Figure 11. The final normal map, and a scanline showing the x (red), y (green), and z (blue) components of the normal vectors for the *castle* dataset. See Figure 9 for scanline position.

Figure 11 presents the results on the *castle* dataset. The scanline through the towers shows a slightly flattened cylindrical shape on the left. The tower on the right shows a sharp corner with the x-component jumping approximately 90 degrees. The base of the tower is occluded by a tree and was therefore excluded from the object mask.

8. Conclusion

We presented a method that recovers shape and reflectance information from outdoor webcams. Such reconstructions have so far only been studied in controlled settings. In practice, some unique challenges such as unknown lighting, uncontrolled cameras, and strongly varying scene appearance have to be addressed. To make processing tractable, we developed an image selection technique geared towards the requirements of photometric stereo, reducing the data by orders of magnitude. We presented an initialization and optimization scheme that jointly improves normals, materials, relative light intensities, and indirect lighting. The computed quantities result in a comprehensive scene representation which is relevant in areas such as scene understanding, scene modeling and computer graphics. The representation can, for example, be used for relighting tasks and to visualize the scene given a different set of BRDFs.

In future work we would like to explore the opportunities offered by this kind of data for 3D-modeling and rendering. It could also be interesting to extend our image selection to the classification of more complex weather conditions.

Acknowledgments: This work was supported in part by the DFG Emmy Noether fellowship GO 1752/3-1. The authors thank Seon Joo Kim for his help with the radiometric calibration.

References

[1] J. Ackermann, M. Ritz, A. Stork, and M. Goesele. Removing the example from example-based photometric stereo. *ECCV Workshop RMLE*, pages 1–14, 2010. 2

[2] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. *CVPR*, 2008. 2

[3] R. Basri and D. Jacobs. Photometric stereo with general, unknown lighting. *CVPR*, pages 374–381, 2001. 2

[4] M. Brooks and B. Horn. Shape and source from shading. In *Proc. IJCAI*, pages 932–936, 1985. 1

[5] M. Daum and G. Dudek. On 3-D surface reconstruction using shape from shadows. In *CVPR*, 1998. 4

[6] J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, and M. Pollefeys. Building Rome on a cloudless day. In *ECCV*, pages 368–381, 2010. 2

[7] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Towards Internet-scale multi-view stereo. In *CVPR*, 2010. 1, 2

[8] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *ICCV*, 2007. 1, 2

[9] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying BRDFs from photometric stereo. In *ICCV*, pages 341–348, 2005. 2, 5

[10] T. Haber, C. Fuchs, P. Bekaert, H.-P. Seidel, M. Goesele, and H. P. Lensch. Relighting objects from image collections. In *CVPR*, pages 627–634, 2009. 1

[11] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA*, pages 3079–3089, 1994. 2, 5

[12] A. Hertzmann and S. Seitz. Example-based photometric stereo: shape reconstruction with general, varying BRDFs. *PAMI*, 2005. 2

[13] T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo. In *CVPR*, 2010. 2

[14] B. K. P. Horn and K. Ikeuchi. An application of the photometric stereo method, 1979. 2

[15] A. Hornung, B. Zeng, and L. Kobbelt. Image selection for improved multi-view stereo. In *CVPR*, 2008. 2

[16] N. Jacobs, B. Bies, and R. Pless. Using cloud shadows to infer scene structure and camera calibration. In *CVPR*, 2010. 2

[17] N. Jacobs, W. Burgin, R. Speyer, D. Ross, and R. Pless. Adventures in archiving and using three years of webcam images. *CVPR Workshops*, pages 39–46, 2009. 4

[18] N. Jacobs, N. Roman, and R. Pless. Consistent temporal variations in many outdoor scenes. In *CVPR*, 2007. 2

[19] N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless. Geolocating static cameras. In *ICCV*, 2007. 2

[20] C. Julia, F. Lumbreras, and A. D. Sappa. A factorization-based approach to photometric stereo. *International Journal of Imaging Systems and Technology*, 21(1):115–119, 2011. 2

[21] S. Kim, J. Frahm, and M. Pollefeys. Radiometric calibration with illumination change for outdoor scene analysis. In *CVPR*, 2008. 2, 4

[22] S. J. Koppal and S. G. Narasimhan. Appearance derivatives for isonormal clustering of scenes. *PAMI*, 31, 2009. 2

[23] J.-F. Lalonde, S. G. Narasimhan, and A. Efros. What do the sun and the sky tell us about the camera? *IJCV*, 2010. 2, 3, 4

[24] H. P. A. Lensch, J. Lang, A. M. Sá, and H.-P. Seidel. Planned sampling of spatially varying BRDFs. *CGF*, 22(3), 2003. 2

[25] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Imaging Understanding Workshop*, pages 121–130, 1981. 4

[26] D. Miyazaki and K. Ikeuchi. Photometric stereo under unknown light sources using robust svd with missing data. In *ICIP*, pages 4057–4060, 2010. 2

[27] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. 2 edition, 1992. 6

[28] I. Reda and A. Andreas. Solar position algorithm for solar radiation applications. *Sol. Energy*, 2004. 3, 4

[29] Y. Sato and K. Ikeuchi. Reflectance analysis under solar illumination. In *Proc. Workshop on Physics-Based Modeling in Computer Vision*, pages 180–187, 1995. 1, 2

[30] W. R. Scott, G. Roth, and J.-F. Rivest. View planning for automated three-dimensional object reconstruction and inspection. *ACM Comput. Surv.*, 35(1):64–96, 2003. 2

[31] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. *CVPR*, 2009. 2

[32] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *CVPR*, 2010. 2

[33] I. Simon and S. M. Seitz. Scene segmentation using the wisdom of crowds. In *ECCV*, 2008. 1

[34] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz. Factored time-lapse video. *ACM TOG*, 2007. 2, 4

[35] K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? In *CVPR*, 2008. 2

[36] G. J. Ward. Measuring and modeling anisotropic reflection. *SIGGRAPH*, 1992. 5

[37] R. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 1980. 2