

# Photometric Stereo and Weather Estimation Using Internet Images

Li Shen      Ping Tan

Department of Electrical & Computer Engineering, National University of Singapore

## Abstract

*We extend photometric stereo to make it work with internet images, which are typically associated with different viewpoints and significant noise. For popular tourism sites, thousands of images can be obtained from internet search engines. With these images, our method computes the global illumination for each image and the surface orientation at some scene points. The illumination information can then be used to estimate the weather conditions (such as sunny or cloudy) for each image, since there is a strong correlation between weather and scene illumination. We demonstrate our method on several challenging examples.*

## 1. Introduction

The appearance of an outdoor scene depends on the scene geometry, surface reflectance, weather and lighting conditions. Scene appearance gives strong cues to these various factors, and humans can easily perceive them from an image. For example, in Fig. 1 we can tell that the picture of St. Basil Cathedral in column (a) is captured in a sunny day, as there is significant shading variation between the building facades. In contrast, the picture in column (c) is captured in a cloudy day with uniform illumination, and the shadings of the different building facades are more consistent.

To obtain these scene properties from images is also a fundamental problem in computer vision. Once recovered, these properties can provide useful information for various applications. For example, sunlight information can be used to localize webcams [4, 18], and scene depth can be recovered from appearance change caused by fog or haze [10]. Many different algorithms have been designed to retrieve these properties from images. Photometric stereo [21] computes scene geometry in the form of local surface orientation. Surface reflectance properties can be derived as a 4D bidirectional reflectance distribution function [9] from images. Appearance variations of a textured surface [20] can be used to recover the illumination direction. However, these algorithms are mostly designed for indoor lab environments with carefully controlled lighting.

Recently, there has been active interest in radiometric

techniques to handle general outdoor images. The seminal work by Basri and Jacobs [1] extended photometric stereo to work with images under unknown general distant illumination. Romeiro et al. [13] measured isotropic reflectance under general lighting. Kim et al. [5] calibrated the radiometric response function and the exposure of images for an outdoor scene. Lalonde et al. [6] calibrated camera parameters from a clear-sky image sequence. Narasimhan and Nayar [11] inferred weather conditions by analyzing the scattering of light sources in the atmosphere. Our work belongs to this category of radiometric image analysis under general outdoor illumination. Our method takes multiple outdoor images as input to infer weather conditions such as sunny or cloudy according to the estimated environment illumination.

Most of these techniques, [1, 5, 6], require a video sequence captured by a fixed camera as input. In contrast, we work with images and do not require the camera to be fixed. As a result, our method can take advantage of millions of pictures available online. Image search engines can retrieve tens of thousands of pictures from a popular tourism site. This number keeps growing with the prevalence of digital cameras and internet photo-sharing. Such images provide abundant information and, at the same time, pose as a significant challenge to vision algorithms. Several classic vision problems have been examined with internet images. Snavely et al. [15, 16] used internet images to compute a sparse 3D scene geometry and utilize it for image browsing. Hays and Efros [3] developed a method for image completion from millions of internet images. Stone et al. [17] retrieved the social network from internet images to enhance face recognition. Liu et al. [8] used internet images for colorization. Our method also makes use of internet images to tackle difficult vision problems. It is known [12] to be difficult to recover the global illumination from a single image even if the scene geometry is precisely given. Estimating the global illumination from a single image as shown in Fig. 1 (a) is very difficult. The internet provides millions of images of the same site. We use these images to constrain the problem to make it more tractable.

We follow the formulation originally developed in [12, 1] to model the appearance of an outdoor scene. We consider

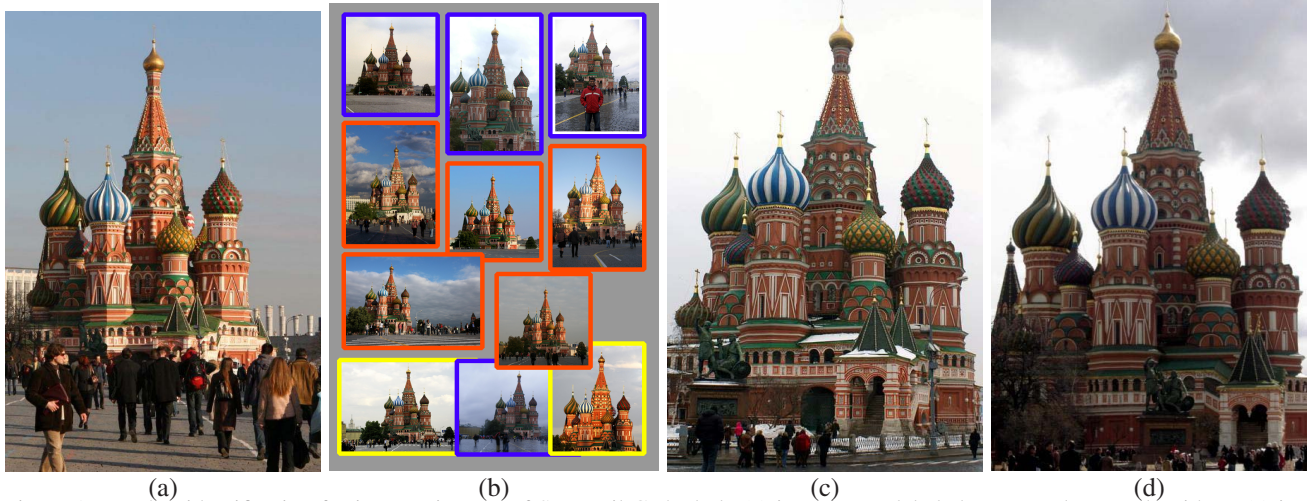


Figure 1. Weather identification for internet images of St. Basil Cathedral. (a) is an image labeled as sunny by our algorithm. (c) is an image labeled as cloudy. (b) shows more images, where blue frame indicates cloudy, red frame indicates sunny, and yellow frame indicates failed examples. These images have variant viewpoints and levels of image noise. (d) is a typical failed example. The algorithm considers this picture as cloudy because sunlight comes from backward and most of the scene points are in attached shadow.

a Lambertian scene illuminated by a distant global environment lighting. The scene radiance can be regarded as a convolution of the global lighting with a ‘half-cosine kernel’. With multiple images under different illuminations, surface normal directions can be recovered. One of the difficulties of applying this formulation to analyze an outdoor scene is that it is hard to collect sufficient images under different lighting conditions. Although sunlight direction varies in a day and provides a natural varying illumination, its movement is confined in a plane and cannot provide sufficient constraint for surface orientation. We avoid this problem by working with images searched from the internet. These images are captured across different seasons and years and often have sufficient illumination variations. However, they also have very different resolutions, contain significant noise and are captured at different viewpoints, which are challenges to vision algorithms. We extend the previous work on photometric stereo [1] to handle these very diversified internet images and recover a sparse set of normal directions as well as the low frequency illumination information for each image. The weather conditions are then identified from the recovered illumination information. As sunny day is associated with strong directional lighting and cloudy day is associated with uniform lighting, we can simply classify weather as sunny or cloudy according to the frequency domain characteristics of the illumination information.

## 2. Photometric Stereo under General Lighting

The theoretic framework for photometric stereo under general lighting is developed in [1, 12]. Here, we give a brief description of this framework to make this paper self-contained. A general lighting is considered as on a distant

sphere  $l(\theta_i, \phi_i)$ , where  $(\theta, \phi) \in \Omega = [0, \pi] \times [0, 2\pi]$  is a spherical coordinate. Its value at  $(\theta_i, \phi_i)$  indicates the incident flux from the direction  $\omega_i = (\theta_i, \phi_i)$ . We neglect the effects of cast shadow and near-field illumination. For a surface with a bidirectional reflectance distribution function (BRDF)  $\rho(\omega_i, \omega_o)$ , the reflected radiance at a point with a surface orientation  $n$  toward the direction of  $\omega_o = (\theta_o, \phi_o)$  should be

$$I(n, \omega_o) = \int_{\Omega} \rho(\omega_i, \omega_o; n) l(\omega_i) \max(\cos(n, \omega_i), 0) d\omega_i.$$

This integration can be regarded as a convolution of the spherical lighting with a kernel determined by the surface BRDF  $\rho(\omega_i, \omega_o)$ . If the scene reflectance is Lambertian, the kernel acts as a low pass filter and the reflected radiance can be approximated as

$$I(n) \approx \sum_{i=1}^K l_i s_i \quad (1)$$

where  $l_i$  and  $s_i$  are the coefficients of expanding  $l(\theta_i, \phi_i)$  and the ‘half cosine’ kernel in spherical harmonics, respectively.  $s_i$  depends on the surface normal direction  $n$  and the albedo  $\rho$ . Specifically, the first 9 coefficients are given by

$$\begin{aligned} s_1 &= \rho, s_2 = \rho x, s_3 = \rho y, s_4 = \rho z, s_5 = \rho(3z^2 - 1), \\ s_6 &= \rho x^2, s_7 = \rho xz, s_8 = \rho yz, s_9 = \rho(x^2 - y^2) \end{aligned} \quad (2)$$

where  $x, y, z$  are the components of the normal direction  $n$ .  $K = 4$  or  $9$  indicates the first or second order of the spherical harmonics approximation, respectively. Using a second order approximation, the accuracy of the approximation for any light function exceeds 98%. With a first order approximation, the accuracy exceeds 75%.

From multiple images captured by a fixed camera under varying illumination, the photometric stereo using a second order approximation is formulated as

$$\mathbf{I}_{N \times M} \approx \mathbf{L}_{N \times 9} \mathbf{S}_{9 \times M} \quad (3)$$

where  $N$  is the number of input images, and  $M$  is the number of pixels in each image.  $\mathbf{I}$  is the matrix formed by the reflected radiances recorded in the images. Each row of  $\mathbf{I}$  is the concatenated radiances from an input image. Each row of  $\mathbf{L}$  is the spherical harmonics coefficients of the lighting of the corresponding image. Each column of  $\mathbf{S}$  encodes the surface reflectance and orientation of a scene point. Each row of  $\mathbf{S}$  is a harmonic image, which is the image of the scene under a harmonic lighting. These harmonic images are grouped according to their order. The zeroth order harmonic image is the first row of  $\mathbf{S}$ ; the first order harmonic images include the second, third and fourth row of  $\mathbf{S}$ ; the second order harmonic images include the remaining five rows. The 9D linear space formed by these harmonic images is called the 9D *harmonics space*.

Matrix factorization techniques can be applied to retrieve both  $\mathbf{L}$  and  $\mathbf{S}$ . Typically, Singular Value Decomposition (SVD) is applied to extract a 9D subspace space  $\bar{\mathbf{S}}_{9 \times M}$  with the minimum Frobenius distance to  $\mathbf{I}$  since more than 98% of the image radiances are concentrated in the 9D *harmonics space*. Here,

$$\mathbf{I} = \bar{\mathbf{L}} \bar{\mathbf{S}},$$

and the factorized result  $\bar{\mathbf{S}}$  is different from the real configuration by an unknown linear transformation  $T$ , which can later be reduced to a rotation by identifying points of the same surface albedo [1]. If the surface normals in two points are known, the rotational ambiguity can be resolved further.

### 3. Photometric Stereo with Internet Images

To obtain images with sufficient illumination changes, we download multiple images from the same site using a popular search engine. Then, we extend a photometric stereo algorithm to work under general lighting for multi-view images. While photometric stereo is generally considered as a solved problem, it is nontrivial to extend it to work with internet images of variant viewpoints. We first let an image selection algorithm select a set of images that are captured at similar viewpoints. This ensures that they share a large number of common visible points. Photometric stereo is first run on these views via feature matching to detect the common visible points. The result of this procedure is a sparse set of normal directions and the global lighting conditions estimated from these images. Then other views are added one by one into the system to recover the environment lighting of each of them.

#### 3.1. View and Feature Selection

Conventional photometric stereo algorithms are designed for images captured at a fixed viewpoint. All input images are naturally registered. The input of our algorithm is obtained by searching the internet. Their viewpoints vary

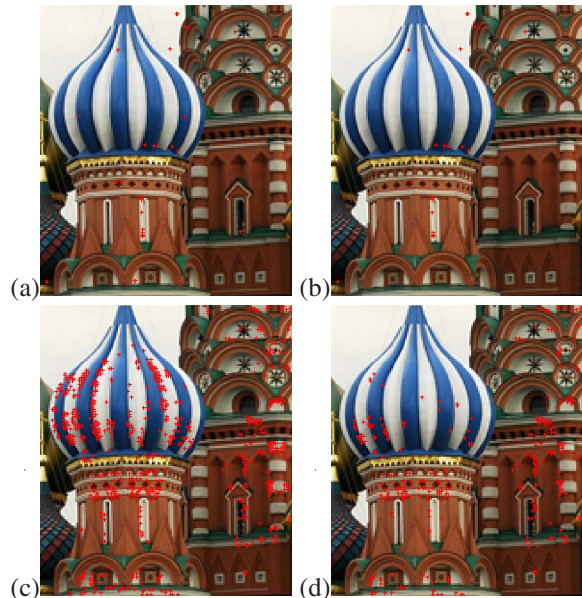


Figure 2. Feature selection. Red cross indicate the selected feature points at each stage. (a) is the initial SIFT matched feature. (b) is the filtered result by homography fitting. (c) is newly propagated features. (d) are the final set of matches used for photometric stereo.

significantly. A natural approach to handle this is to apply SIFT feature matching to extract corresponding points from these images. The factorization based algorithm requires that each surface point is visible in all involved images. Otherwise, there will be missing data in the observation matrix  $\mathbf{I}$ . Although there are existing techniques [14] to handle this problem, we avoid this problem by selecting close views sharing a large number of common visible surface points. We fit a homography for each two views from their matched SIFT features and choose images with small homography fitting errors as the initial set. Photometric stereo is performed on these images and later extended to other views.

Once the initial set of images is selected, we extract their common visible points by SIFT matching. These matches are filtered by the estimated homography. Matches with large discrepancy from the homography are discarded. SIFT feature points are generally at corners and edges where albedo changes quickly. A small inaccuracy in matching will cause a large difference of surface albedo, which makes the algorithm formulation invalid. To alleviate this problem, we interpolate more matches at smooth regions. Starting from the most reliable matches, we apply the method described in [7] to propagate matching to other surface points. Many of these propagated points have less albedo changes in their neighborhood, hence, less sensitive to matching inaccuracy. On the other hand, though these propagated matches have more consistent albedo, their matching quality is less reliable compared to those obtained through SIFT matching. As a result, they might

have inconsistent surface normal directions. Moreover, the matched features may include the points contaminated by cast shadows or interreflections. We further filter all these features to select an optimal subset.

We form a data matrix  $\mathbf{I}_{N \times M}$  by collecting radiance values at matched feature points in the images.  $M$  is the number of features visible in all  $N$  selected views. Assuming there are no effects of cast shadows and interreflections, Eq. (3) should be satisfied. We apply RANSAC to the SVD based matrix factorization to filter incorrect matches. If the matching is incorrect, or if there are cast shadows/interreflections, the radiance of the feature point will not lie close to the *harmonic space*. We select a set of inliers from all the features including SIFT matched features and propagated features. In practice, we find it is often helpful to normalize the image intensity before matrix factorization to make each input image have similar average intensity.

The feature selection is exemplified in Fig. 2. (a) is the initial SIFT matching, (b) is the matching filtered by homography estimation. We obtain 310 matches from SIFT, among them 230 are remained after homography fitting. Fig. 2 (c) shows the propagated matches, (d) shows the final set of features we used for photometric stereo. We obtain about 2000 matches after propagation. RANSAC choose 1200 from all these matches as inlier for the 9D *harmonics space* fitting.

### 3.2. Subspace Analysis

Starting from the subspace  $\bar{\mathbf{S}}$  obtained from SVD, we recover a surface normal direction for each matched feature point and the environment lighting for each image. As  $\bar{\mathbf{S}}$  is the 9D subspace with the minimum Frobenius distance to the row space of  $\mathbf{I}$ , and the 9D *harmonic space*  $\mathbf{S}$  should contain more than 98% of the radiance, it is often safe to treat these two linear spaces as identical, and  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  differ by a linear transformation  $\mathbf{S} = T\bar{\mathbf{S}}$ . The surface normal directions and albedos are encoded in the row vectors of  $\mathbf{S}$ , i.e. the harmonic images. It is not trivial, however, to extract those harmonic images from  $\bar{\mathbf{S}}$ . Basri et al. [1] propose to estimate a matrix  $A_{3 \times 9}$  to extract the first order harmonic images, i.e. the scaled surface normals, as

$$\mathbf{b}_{3 \times M} = A_{3 \times 9} \bar{\mathbf{S}}_{9 \times M}.$$

Here, each column of  $\mathbf{b}$  is a surface normal direction scaled by its associated albedo as  $[\rho x, \rho y, \rho z]^T$ . Given an  $A$ , the complete *harmonic space*  $\mathbf{S}_A$  can be reconstructed analytically from the scaled normal directions  $\mathbf{b}$  according to Eq. (2). Hence, the matrix  $A$  is searched by minimizing the difference between the row space of original image data and the reconstructed *harmonics space*, i.e.

$$A = \arg \min \|\mathbf{I} - \mathbf{L}_A \mathbf{S}_A\|. \quad (4)$$

Here,  $\mathbf{L}_A$  is computed by minimizing  $\|\mathbf{I} - \mathbf{L}_A \mathbf{S}_A\|$  for each  $\mathbf{S}_A$ . The optimization for  $A$  is computationally simple, as  $\mathbf{L}_A$  is linearly computed. However, Basri et al. [1] notice the estimated scaled normals from this optimization could differ from the true scaled normals by an unknown linear transformation because linearly transformed scaled normals lead to a similar *harmonic space*.

In practice, we find this ambiguity is even larger than an unknown linear transformation. The 3D space formed by true first order harmonic images can be different from the estimated 3D row space of  $A\bar{\mathbf{S}}$ . In other words, there is no linear transformation that can relate these two 3D spaces. We use a synthetic example to exemplify the problem. We render a uniform sphere at a fixed viewpoint under 60 different illumination conditions. A small Gaussian noise with standard deviation 0.001 is added. Its 9D *harmonic space* and nine harmonic images are constructed analytically. We find multiple different 3D spaces can be used as ‘the scaled normals’ and reconstruct the same *harmonic space* quite accurately. We exhaustively choose all combinations of three among the nine harmonic images to initial  $A$  in the minimization Eq. (4). Among the total 84 combinations, 51 combinations can lead to a result with average pixel error of 0.001. However, only one among these 51 initials generates the correct scaled normals that are related to the true configuration by a linear transformation. All the other 50 initials lead to a 3D row space of  $A\bar{\mathbf{S}}$  that is significantly different from the space of true first order harmonic images.

Fig. 3 shows some minimization result with different initials. The first row shows the surface albedo map. The second row is a map of the normal directions, where the  $x, y, z$  components of a normal are linearly encoded in the R,G,B channels. Column (a) is the ground truth. Our input images are synthesized with this albedo and normal map under 60 different environment lightings. Column (b) are the results computed by choosing the correct harmonic images, i.e. the second, third and fourth harmonic images, to initial  $A$ . The recovered normal directions differ from the ground truth by an unknown rotation. The recovered albedo map is very close to the ground truth. Column (c) is the results from initializing  $A$  with the first, second and third harmonic images. With these results, the original images can be reconstructed precisely with a normalized perpixel error 0.001. It means the reconstructed 9D harmonics space from this result matches the original space very well. However, the recovered scaled normals from this initial cannot be linearly transformed to the true scaled normals. This can also be verified from the recovered surface albedos. The real surface has a uniform albedo, yet the recovered results contain significant albedo variation although the cost function in Eq. (4) can be minimized to a very small numeric value.

We show that there are multiple distinct 3D spaces of

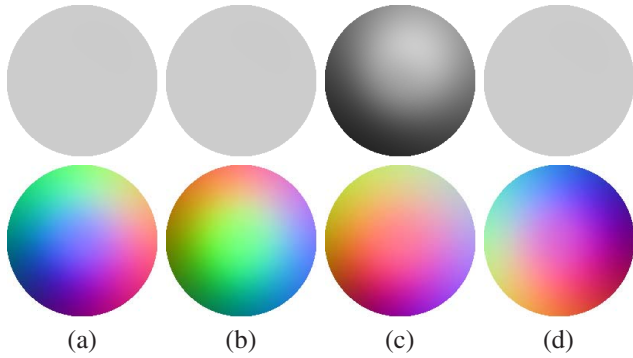


Figure 3. Comparison of different initial to the minimization of Eq. (4). The top row is the surface albedo map. The second row is the linearly coded normal directions. Column (a) is the ground truth. Column (b) and (c) are both computed from Eq. (4) by different initialization of  $A$ s. Both results achieve normalized per-pixel reconstruction error 0.001. The result in (b) is initialized by the second, third and fourth harmonic images. This result differs from the ground truth only by an unknown rotation. The result in (c) is initialized by the first, second and third harmonic images. This result cannot be transformed to the ground truth by any linear transformation. Among the all the 84 initializations (choosing any three harmonic images for initialization), 51 of them can yield highly accurate reconstructions. However, for 50 of these initials, the computed scaled normals cannot be converted to ground truth by any linear transformation. Column (d) is the result computed from Eq. (6) with the albedo consistency constraint. This result is different from ground truth by an unknown rotation.

the scaled normals that can faithfully reconstruct the same 9D harmonics space according to Eq. (1). There is no simple linear transformation between these different 3D spaces. Therefore a good initial guess is very critical to achieve correct results. Basri et al. [1] sort the row vectors of  $\bar{\mathbf{S}}$  according to their associated singular value and take the rows with second, third and fourth largest singular value to initialize  $A$ . However, the sorting based on singular values are less reliable to internet images which are contaminated by significant levels of noise. The singular values of high order harmonics could become large due to image noise, non-lambertian reflectance, etc. For example, Hallinan [2] observes, for human faces images, that the rows of  $\bar{\mathbf{S}}$  form two groups and the rows could exchange places within a group.

### 3.3. Albedo Consistency Constraint

Instead of relying on initialization, we seek additional constraints to solve this problem. We utilize the constraint from distinct points with the same surface albedo. This constraint was previously used in [1] after the matrix  $A$  was obtained to estimate an unknown linear transformation between the recovered scaled normals and the true scaled normals. However, as discussed, an incorrect initial could lead to a result that differs from the true configuration by more

than a linear transformation. Here, we advocate this constraint for the computation of  $A$  as well. We first consider the case of a uniform sphere to demonstrate this constraint.

According to the definition of  $A$ ,

$$\rho_{(i)}n_{(i)} = \mathbf{b}_{(i)} = A\bar{\mathbf{S}}_{(i)}.$$

Here,  $\bar{\mathbf{S}}_{(i)}$  is the  $i$ -th column of  $\bar{\mathbf{S}}$ .  $\rho_{(i)}n_{(i)}$  is the scaled normal at the  $i$ -th feature point. Hence,

$$\bar{\mathbf{S}}_{(i)}^T A^T A \bar{\mathbf{S}}_{(i)} = \rho_{(i)}^2. \quad (5)$$

For a uniform surface, we can consider  $\rho_{(i)} = 1$  as the absolute scale of the albedo that can be absorbed into illumination intensity. Each column of  $\bar{\mathbf{S}}$  gives an equation for  $A$ . If we define a symmetric matrix  $B_{9 \times 9} = A^T A$ . Eq. (5) is a linear constraint about the entries of  $B$ . Accumulating constraints from all the surface points could lead to a linear solution of  $B$ . However, these linear equations for different points are not all independent. We build such a linear system for all the points on a uniform sphere, the resulted solution space for  $B$  is 19D. Since matrix  $B$  has rank 3, we might use this to further constrain its solution. Here, we directly solve Eq. (5) for  $A$  with a general non-linear optimization algorithm. We find that the 3D space of the estimated scaled normals from this optimization could differ from that of the true scaled normals. For example, by solving Eq. (5), we might obtain a matrix  $A$  which satisfies  $A\bar{\mathbf{S}}_{(i)} = [\rho_{(i)}, 0, 0]^T$ . It means that the albedo consistency constraint along is not sufficient to determine the matrix  $A$ . Here, we revise the objective function Eq. (4) by including the constraint Eq. (5) and compute  $A$  as:

$$\arg \min \|\mathbf{I} - \mathbf{L}_A \mathbf{S}_A\| + \sum_{\rho_{(i)}=1} \|\bar{\mathbf{S}}_{(i)}^T A^T A \bar{\mathbf{S}}_{(i)} - 1\|. \quad (6)$$

In Fig. 3 (d), we show the recovered normals and albedos from this non-linear optimization, which are quite consistent to the ground truth. This exemplifies that this hybrid constraint in Eq. (6) is sufficient to extract the harmonic images from an initial 9D space with uniform albedo.

### 3.4. Complete System

For real images containing objects with different albedo, we group feature points with similar albedo based on similarity of their chromaticity. Chromaticity is defined as the normalized RGB value as  $R/(R+G+B)$ ,  $G/(R+G+B)$ . The chromaticity of a Lambertian surface is invariant to shading.  $A$  is computed by minimizing Eq. (6) with the albedo constraint for the feature points with similar albedo. Once  $A$  is computed by this iterative scheme, the scaled normals are determined at each feature point. The surface albedos and normal directions are then obtained from the scaled normals. The complete *harmonic space*  $\mathbf{S}_A$  is also constructed from the scaled normals analytically. We then compute the transformation  $T$  that maps  $\bar{\mathbf{S}}$  to  $\mathbf{S}_A$ . The illumination is recovered by  $\bar{\mathbf{L}}T$ .

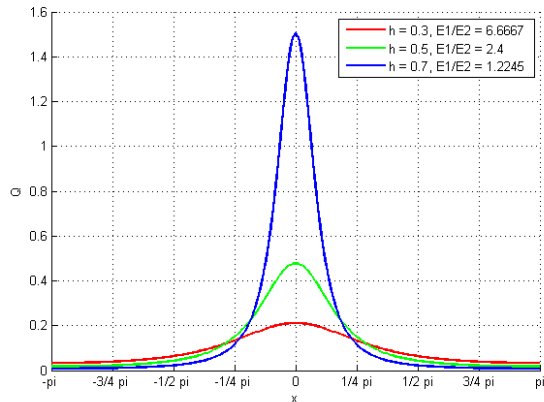


Figure 4. The Poisson kernel with different parameter. Larger parameter  $h$  means stronger peak.

Next, we extend the result on the initial set of images to other views with larger viewpoint changes. These images are processed one by one. For each of the images, we compute SIFT matches between it and the images in the initial set. The normal directions are already estimated for the matched feature points in the initial set of images. Hence, they can be used to compute the illuminations for other views according to Eq. (1).

#### 4. Weather Identification

The derived photometric stereo algorithm might be applied for many applications, e.g. 3D reconstruction. In this paper, we focus on the application of weather estimation for internet images. Outdoor scenes are illuminated by the sky and sunlight. Sunlight is strongly directional and skylight is relatively uniform. Hence, the illumination of a sunny day is directional, while a cloudy day has relatively uniform illumination. We identify weather condition from the directionality of an outdoor illumination.

Once the sparse set of normal directions and the global illumination have been computed with photometric stereo, we estimate the weather condition according to the recovered illumination. As there is an unknown rotation in the reconstructed result, we analyze the energy distribution of the global illumination at different order of harmonics. The energy at each order is defined as  $E_0 = l_0^2$ ,  $E_1 = l_1^2 + l_2^2 + l_3^2$  and  $E_3 = \sum_{i=5}^9 l_i^2$ .  $E_0$ ,  $E_1$  and  $E_2$  are invariant to the unknown rotation.

Next, we show how directional lighting and uniform lighting can be isolated from the analysis of these energy terms. Following [19], we model the directional light with a spherical Poisson function, and represent the uniform light as a spherical constant function. Thus, an environment light can be expressed

$$l(\omega) = c + Q_h(\cos(\omega, \omega_0)).$$

Here,  $c$  is a constant for the uniform light and  $\omega_0$  is the center of the directional light.  $h$  is a parameter of the Poisson



Figure 5. Three different environment lightings and their associated energy ratios. From top to bottom, the lightings become less directional, and their energy ratios  $E_1/E_2$  are 0.72, 2.0 and 15.2, respectively.

function  $Q_h(x)$ , which is defined as:

$$Q_h(x) = \frac{1}{4\pi} \frac{1 - h^2}{(1 - 2hx + h^2)^{3/2}}.$$

Larger  $h$  indicates stronger directionality. Some examples of the Poisson representation are shown in Fig. 4. We choose the Poisson function because its spherical harmonics expansion has simple analytical form, i.e.  $l_i \propto h^k$  for  $1 \leq i \leq 9$ , where  $k = 0, 1, 2$  is the order of  $l_i$ . The corresponding energy terms for the global light can be analytically derived:

$$E_0 = 0.0800 + c, E_1 = 0.2387h^2, E_2 = 0.3978h^4.$$

Both  $E_1$  and  $E_2$  are independent of the uniform light. The directionality property of the Poisson light, i.e. the associated parameter  $h$ , can be estimated from the energy ratio of  $E_1/E_2 = 0.6h^{-2}$ . Smaller ratio means the light is more directional.

Though a more sophisticated algorithm can be designed to decide if an image is of a sunny or cloudy scene according to the vector  $(E_0, E_1, E_2)$ , we simply threshold the energy ratio  $E_1/E_2$  to demonstrate the idea of weather identification. Here, we manually choose a cutting threshold at 2.4 for all the examples in this paper. A few real environment lightings and their associated energy ratios are shown in Fig. 5 to justify this threshold. The top image with ratio  $E_1/E_2 = 0.72$  is a scene of a sunny day and the bottom image with  $E_1/E_2 = 15.2$  is a scene of a cloudy day. The middle is sunny day with some clouds. Its energy ratio is 2.0.

#### 5. Results

We applied the method described in the paper to internet images and label images as sunny or cloudy. Fig. 1 shows the example of St. Basil Cathedral. We downloaded 100 pictures of St. Basil Cathedral from flickr. These images are captured under diversified times, sunlight directions,

weathers and viewpoints, and are from different cameras, affected by different amounts of noise. It presents a serious challenge for traditional photometric stereo. Our method, however, can correctly recover the normal directions at the matched points and the environment lighting for each image up to an unknown rotation. All images are then marked as sunny or cloudy according to the computed illumination conditions. Because no ground truth of the weather conditions is available, we also manually labeled each image as sunny or cloudy for verification. Column (a) of Fig. 1 shows an image labeled as sunny. Column (c) shows an image labeled as cloudy. Column (b) shows more examples, where blue frame indicates cloudy, red frame indicates sunny, and yellow frame highlights failed example. Column (d) is a failure example. The algorithm considers this image as cloudy, but it is actually sunny. This image is incorrectly classified because the sunlight comes from behind and most of the scene points are in shadows. Although there is very subtle highlight on the roof showing this is a sunny day, our algorithm fail on this challenging example. The estimated weather is consistent with the manually marked result for 74% of all the images.

Fig. 6 shows the example of the Taj Mahal. We downloaded about 80 images of the Taj Mahal from flickr and applied the photometric stereo algorithm to recover the surface normal directions and global lighting. Then, the images are labeled as sunny or cloudy according to the recovered illuminations. We also manually labeled each image for comparison. Our algorithm can correctly label 80% of these images. An image of a sunny day is shown in Fig. 6 (a). Fig. 6 (c) shows a cloudy day for comparison. More results are included in Fig. 6 (b). Similarly, red, blue and yellow frames are added to indicate sunny, cloudy and failed examples.

To verify the reconstruction accuracy of our photometric stereo algorithm, Fig. 8 visualize the recovered surface normals and global lightings. We manually assigned surface normal directions for two feature points to fix the unknown rotation to assist the visual comparison. Fig. 8 shows a closeup of the Taj Mahal example. A normal direction is recovered for each feature point in the image. Fig. 8 (b) is a visualization of the estimated normal directions in the framed roof region in (a). These normal directions are consistent with the spherical roof geometry. The estimated global lighting for the image shown in (a) is validated in (c), where a sphere is rendered under this estimated lighting. The illumination of this sphere is consistent with the input image in (a). Some of the difference is caused by the inaccuracy of the manually corrected rotation.

Another example shown in Fig. 7 is the Mount Rushmore National Memorial. We used 90 images of the spot downloaded from flickr in the example. An image labeled as sunny by our algorithm is shown in (a). An image labeled

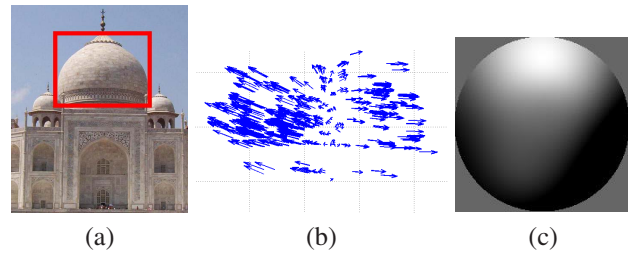


Figure 8. Validation of photometric stereo result. (a) shows a cropped region from one of the input image for the Taj Mahal example. The recovered normal directions for the framed region in (a) are visualized in (b). These normal directions on the roof are quite consistent with the scene geometry. (c) is a sphere rendered under the estimated global lighting for the image shown in (a). The illumination on this sphere is quite consistent with the input image.

as cloudy is shown in (c). More identifying results are shown in (b). With the high quality of the image set, our method achieved success rate of 90% on this example.

## 6. Conclusion

We have presented a method to perform photometric stereo and weather estimation using internet images. Our photometric stereo method can handle variant viewpoints and tolerate noisy internet images. The recovered global illumination is applied for weather estimation. Although we currently only identify two kinds of weather conditions (i.e. sunny and cloudy), similar approaches can be used with a more sophisticated classifier to identify other weather patterns such as fog or rain. Furthermore, our method recovers additional physical scene properties for internet images, which can be used for other applications or general image annotations.

## 7. Acknowledgement

We would like to thank Michael Brown for the fruitful discussion. The authors were supported by the Singapore FRC Grant R-263-000-477-112.

## References

- [1] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *Int'l Journal of Computer Vision*, 72(3), 2007.
- [2] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proc. of CVPR*, 1994.
- [3] J. Hays and A. A. Efros. Scene completion using millions of photographs. *ACM Trans. on Graphics (SIGGRAPH 2007)*, 2007.
- [4] N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless. Geolocating static cameras. In *Proc. of ICCV*, 2007.

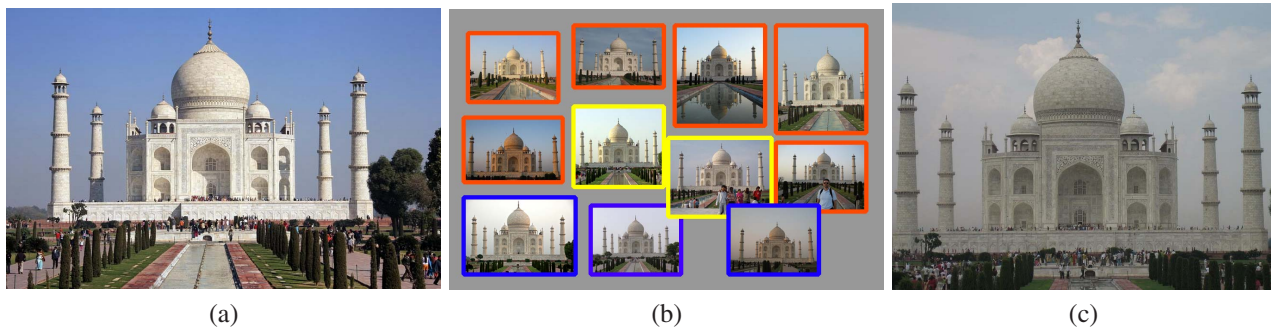


Figure 6. Weather identification for internet images of the Taj Mahal. (a) is an image labeled as sunny by our algorithm. (c) is an image labeled as cloudy. (b) shows more images, where blue frame indicates cloudy, red frame indicates sunny, and yellow frame indicates failed examples.

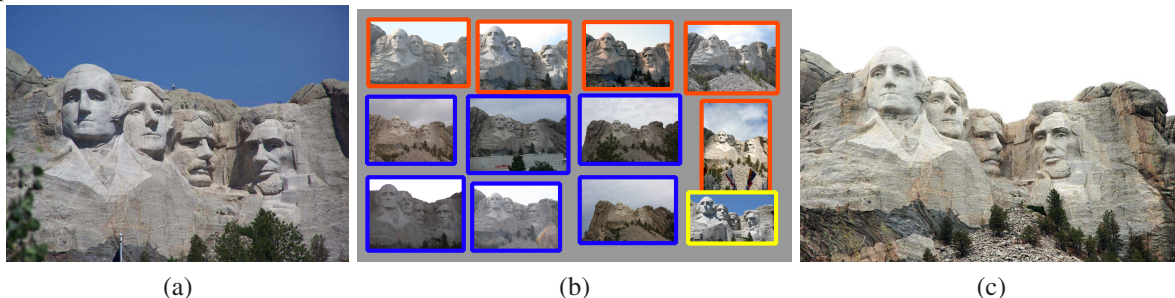


Figure 7. Weather identification for internet images of the Mount Rushmore National Memorial. (a) is an image labeled as sunny by our algorithm. (c) is an image labeled as cloudy. (b) shows more images, where blue frame indicates cloudy, red frame indicates sunny, and yellow frame indicates failed examples, respectively.

- [5] S. J. Kim, J. M. Frahm, and M. Pollefe. Radiometric calibration with illumination change for outdoor scene analysis. In *Proc. of CVPR*, 2008.
- [6] J. Lalonde, S. Narasimhan, and A. Efros. What does the sky tell us about the camera? In *Proc. of ECCV*, 2008.
- [7] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 24(8), 2002.
- [8] X. Liu, L. Wan, Y. Qu, T.-T. Wong, S. Lin, C.-S. Leung, and P.-A. Heng. Intrinsic colorization. 2008.
- [9] S. Marschner, S. Westin, E. Lafortune, K. Torrance, and D. Greenberg. Image-based brdf measurement including human skin. In *Proc. 10th Eurographics Workshop on Rendering*, pages 139 – 152, 1999.
- [10] S. NaraSimhan and S. Nayar. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 25(6), 2003.
- [11] S. Narasimhan and S. Nayar. Shedding light on the weather. In *Proc. of CVPR*, 2003.
- [12] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. 2001.
- [13] F. Romeiro, Y. Vasilyev, and T. Zickler. Passive reflectometry. In *Proc. of ECCV*, 2008.
- [14] H.-Y. Shum, K. Ikeuchi, and R. Reddy. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 17(9), 1995.
- [15] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring image collections in 3d. 2006.
- [16] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *Int'l Journal of Computer Vision*, 80(2), 2008.
- [17] Z. Stone, T. Zickler, and T. Darrell. Autotagging facebook: Social network context improves photo annotation. In *Proc. of First Workshop on Internet Vision*, 2008.
- [18] K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? In *Proc. of CVPR*.
- [19] Y.-T. Tsai and Z.-C. Shih. All-frequency precomputed radiance transfer using spherical radial basis functions and clustered tensor approximation. *IEEE Trans. on Graphics (SIGGRAPH 2006)*, 2006.
- [20] M. Varma and A. Zisserman. Estimating illumination direction from textured images. In *Proc. of CVPR*. IEEE, 2004.
- [21] R. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensities. In *Proc. SPIE 22nd Annual Technical Symposium*, 1978.