Photometric Stereo with Coherent Outlier Handling and Confidence Estimation

Frank Verbiest[1]

Luc Van Gool[1,2]

[1]KU Leuven
Leuven, Belgium
`firstname.lastname@esat.kuleuven.be`

[2]ETH Zürich
Zürich, Switzerland
`vangool@vision.ee.ethz.ch`

Abstract

In photometric stereo a robust method is required to deal with outliers, such as shadows and non-Lambertian reflections. In this paper we rely on a probabilistic imaging model that distinguishes between inliers and outliers, and formulate the problem as a Maximum-Likelihood estimation problem. To signal which imaging model to use a hidden binary inlier map is introduced, which, to account for the fact that inlier/outlier pixels typically group together, is modelled as a Markov Random Field. To make inference of model parameters and hidden variables tractable a mean field Expectation-Maximization (EM) algorithm is used. If for each pixel we add the scaled normal, i.e. albedo and normal combined, to the model parameters, it would not be possible to obtain a confidence estimate in the result. Instead, each scaled normal is added as a hidden variable, the distribution of which, approximated by a Gaussian, is also estimated in the EM algorithm. The covariance matrix of the recovered approximate Gaussian distribution serves as a confidence estimate of the scaled normal. We demonstrate experimentally the effectiveness or our approach.

1. Introduction

Photometric stereo, first introduced by Woodham [16], allows the estimation of reflectance (albedo) and local surface orientation (normal) at each pixel by using several images captured from a fixed viewpoint but for different illumination conditions. Ideally, for each captured image, the surface is lit by a single distant light source. For a Lambertian surface, the amount of light reflected depends on the angle the normal makes with the illumination direction, the illumination intensity and the albedo, which can be expressed by the following equation,

$$z = \lambda \rho \mathbf{l}^{\mathrm{T}} \mathbf{n} = (\lambda \mathbf{l})^{\mathrm{T}} (\rho \mathbf{n}) = \mathbf{s}^{\mathrm{T}} \mathbf{b} \qquad (1)$$

where z is the pixel intensity observed at a surface patch, $\mathbf{b}$ is the unit normal $\mathbf{n}$ scaled by the albedo ρ and $\mathbf{s}$ is the unit
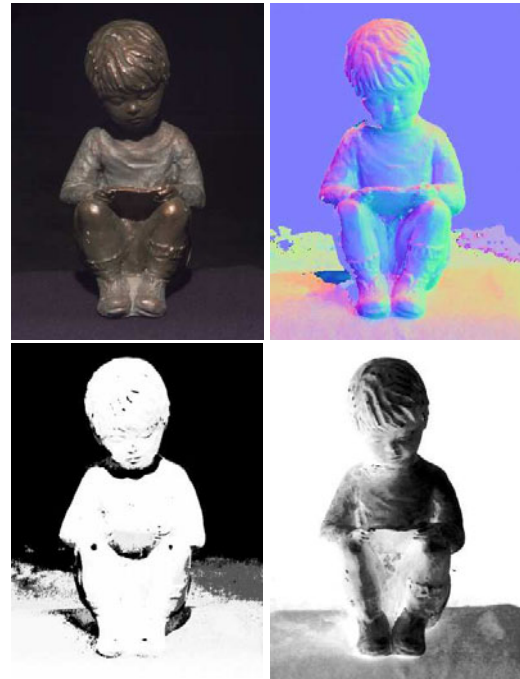

Figure 1. Top left: One of 169 captured input images. Top right: The mean of the posterior scaled normal distribution normalized, also known as the normal map (color coded in the usual way). Bottom left: Posterior inlier probability. Bottom right: The covariance of the posterior normal distribution, visualized by its trace. Dark (small trace, small covariance) indicates higher confidence.

illumination direction $\mathbf{l}$ scaled by the illumination intensity λ. Both scaled normals $\mathbf{b}$ and scaled lights $\mathbf{s}$ are basic entities in photometric stereo. With known scaled lights (calibrated photometric stereo), each pixel intensity measurement yields a linear equation in the scaled normal. Three such measurements, corresponding with linearly independent scaled lights, suffice to determine the scaled normal. Extraction of the albedo and normal is trivial. This way the surface shape and the pattern on the surface produced by varying albedo can be successfully separated. We will

1

not consider the case where both scaled normals and scaled lights are unknown (uncalibrated photometric stereo).

In reality, however, the presence of shadows and the complex material behavior exhibited by some surfaces render this Lambertian assumption invalid and force us to resort to methods which are robust against such complicating effects. Despite the extensive research on such robust methods (see subsection 1.1), the presence of severe shadows, highlights, complex geometry and materials, still pose serious problems to photometric stereo.

Recently [15, 17, 18], with the availability of cheap memory and computing power, the use of a dense set of images (> 100) in photometric stereo emerged as a viable alternative. Such methods start from the observation that for a wide range of illumination directions most pixel intensities measured exhibit a purely Lambertian behavior. Given enough such measurements this fact can be exploited to determine the normal and albedo, by iteratively trying to find a consensus around the Lambertian model, as expressed by Eq. 1. Also, using a larger number of illumination directions can lead to increased normal-noise averaging and thus higher quality normals. The approach presented in this paper falls into this category.

The successful application of a proper probabilistic model combined with Expectation-Maximization in domains such as multi-view stereo [12, 5] and face recognition [2], suggest similar techniques for solving the robust photometric stereo problem from a dense set of images. By fully exploiting the data redundancy inherent in the dense and noisy data, we are able to recover accurate normals and albedos despite the presence of severe shadows, highlights, complex geometry, and inaccurate illumination intensity calibration. The probabilistic imaging model used in this paper distinguishes between an inlier and outlier process for explaining observed pixel intensities. Prior information can be used to speed up convergence and guide the solution to more likely configurations. For this reason, the coherence present in inlier/outlier pixels is expressed by modelling the inlier map as a Markov Random Field. Indeed shadow and highlight pixels do not usually find themselves isolated. Everything is formulated in a Maximum-Likelihood setting. Model parameters and hidden variables are inferred through mean field Expectation-Maximization algorithm, of which the convergence properties are well known. Part of a solution is also to know how good your solution is, at least qualitatively. To get a confidence measure on the result, we add the scaled normals to the hidden variables instead of the model parameters. This way approximate posterior distributions of the requested scaled normals follow from the EM procedure, which then can be used for interpreting the (un)certainty of the result. If there a is need for smoothing the normals, these distributions can be used to guide the smoothing process, by not allowing normal configurations with low probability.

1.1. Related Work

Since Woodham [16], there has been much research on robust methods. One of the earliest extensions, due to Coleman and Jain [4], was to use 4 images to detect highlights in the absence of shadows. Each combination of 3 images yields an estimate for the albedo. If the recovered albedos differ significantly, a highlight should be present. The combination producing the smallest albedo contains only the Lambertian component and is used. The case of shadows in the absence of highlights follows by similar reasoning. Barsky and Petrou [1] can handle both cases by using color images to distinguish between highlights and shadows.

A large body of work is devoted to the recovery of more complex reflectance models along with the local surface orientation. Non-Lambertian effects are considered inliers of these reflectance models. Provided enough images are available, the reflectance parameters can be estimated locally, and such methods are therefore suitable for surfaces with spatially varying reflectance properties. Nayar *et al*. [10] applied photometric stereo to a so-called hybrid reflectance model which is a linear combination of Lambertian and specular components. Tagare and deFigueiredo [14] developed the theory of photometric stereo for the class of m-lobe reflectance maps. Goldman *et al*. [6] exploit the observation that for many objects, pixel behaviour can be well represented by a mixture of a small number of reflectance models. Surface orientation, reflectance parameters and mixture weights are estimated in an alternating fashion.

Other methods do not estimate parameters of reflectance models but rather exploit some of their properties. By computing a data-dependent rotation of RGB color space, Mallick *et al*. [8] show that the specular component in reflectance can be separated from the simpler, diffuse component for surfaces that can be modelled with a dichromatic reflection model. Images in this transformed color space are used to obtain photometric reconstructions. Hertzmann and Seitz [7] make use of orientation consistency to establish correspondence between normals of an unknown object and one or more reference objects imaged under the same illumination. The reference objects have known geometry and material properties similar to the unknown object. Clustering is used to obtain a full segmentation into different material types. However, both these methods [7, 8] do not take shadows into account.

Recently, several robust photometric stereo algorithms have been developed that use a dense set of images [15, 17, 18]. To improve results, Tang *et al*. [15] use a MRF formulation, inspired by Sun *et al*. [13], to impose discontinuity preserving smoothness on the normals. Efficient inference is accomplished by tensorial belief propagation. In

a similar MRF approach, Wu and Tang [17] use graph-cuts for inference. However, neighborhood constraints on normals have been reported to erase fine surface details [18]. The same authors, Wu and Tang [18], use Expectation-Maximization to recover the albedo while simultaneously clustering a set of initial normals obtained through ratio images. The matching cost defined in measurement space is augmented with an ad hoc clustering term.

2. Objective Function

Let $\boldsymbol{\theta}$ be the model parameters, and let $\mathbf{z}$ and $\mathbf{h}$ denote all the input images and hidden variables, respectively. The maximum likelihood (ML) solution of the model parameters is given by:

$$
\begin{aligned}
\boldsymbol{\theta}_{ML} &= \arg\max_{\boldsymbol{\theta}} p(\mathbf{z}\,|\,\boldsymbol{\theta}) \\
&= \arg\max_{\boldsymbol{\theta}} \Big\{ \sum_{\mathbf{h}} p(\mathbf{z},\mathbf{h}\,|\,\boldsymbol{\theta}) \Big\} \\
&= \arg\max_{\boldsymbol{\theta}} \Big\{ \log \sum_{\mathbf{h}} p(\mathbf{z},\mathbf{h}\,|\,\boldsymbol{\theta}) \Big\} \quad (2)
\end{aligned}
$$

The hidden variables $\mathbf{h} = \{\mathbf{b},\mathbf{v}\}$ contain a scaled normal map $\mathbf{b}$ and inlier maps $\mathbf{v}$. Included in the model parameters $\boldsymbol{\theta}$ are the inlier noise variance σ^2, the prior inlier fraction α_i for image i, and a parameter vector $\boldsymbol{\beta}_i$ used in the histogram to describe the outlier process for image i.

Given that we are solving the photometric stereo problem, the probabilistic distribution of the hidden scaled normal map is of particular interest. An approximation to this unknown distribution will be estimated during the Expectation-Maximization (EM) procedure in section 4. So by including the scaled normals into the set of hidden variables, instead of into the model parameters, a confidence measure becomes available.

3. Description of Probabilistic Model

In this section the complete data likelihood term $p(\mathbf{z},\mathbf{h}\,|\,\boldsymbol{\theta})$, used in Eq. (2), will be explained in detail. The assumption of independent $\mathbf{b}$ and $\mathbf{v}$ allows for the following decomposition,

$$
p(\mathbf{z},\mathbf{h}\,|\,\boldsymbol{\theta}) = p(\mathbf{z}\,|\,\mathbf{b},\mathbf{v},\boldsymbol{\theta})p(\mathbf{b}\,|\,\boldsymbol{\theta})p(\mathbf{v}\,|\,\boldsymbol{\theta}) \quad (3)
$$

where $p(\mathbf{z}\,|\,\mathbf{b},\mathbf{v},\boldsymbol{\theta})$ is the probability of observing the image data given the model parameters and the state of hidden variables, $p(\mathbf{b}\,|\,\boldsymbol{\theta})$ and $p(\mathbf{v}\,|\,\boldsymbol{\theta})$ are priors on the normal map and inlier maps, respectively.

The image likelihood $p(\mathbf{z}\,|\,\mathbf{b},\mathbf{v},\boldsymbol{\theta})$ can be factorized further over all individual pixel likelihoods as follows,

$$
\begin{aligned}
p(\mathbf{z}\,|\,\mathbf{b},\mathbf{v},\boldsymbol{\theta}) &= \prod_i p(\mathbf{z}_i\,|\,\mathbf{b},\mathbf{v}_i,\boldsymbol{\theta}) \\
&= \prod_i \prod_x p(z_{i,x}\,|\,\mathbf{b}_x, v_{i,x},\boldsymbol{\theta}) \quad (4)
\end{aligned}
$$

where $\mathbf{z}_i$ and $\mathbf{v}_i$ are image i and its corresponding inlier map, respectively. For the pixel location x in image i, $z_{i,x}$ and $v_{i,x}$ are pixel intensity and inlier status (either $+1$ or -1), respectively. The continuous 3-vector $\mathbf{b}_x$ corresponds to the scaled normal at pixel location x. We assume each of the observations $z_{i,x}$ independent of one another, and conditioned only on the hidden variables $v_{i,x}$ and $\mathbf{b}_x$ corresponding to the same image and location. The individual pixel likelihood used in this paper makes a distinction between an inlier process f(.) and an outlier process g(.). The inlier process generates the pixels for image regions not affected by shadows or non-Lambertian effects. Other regions are assigned to the outlier process. In the inlier process the observed pixel intensities are generated as follows,

$$
z_{i,x} = \mathbf{s}_i{}^{\mathrm{T}}\mathbf{b}_x + \epsilon \quad (5)
$$

in which we recognize Eq. 1, the basic equation of Lambertian photometric stereo perturbed by iid additive noise ϵ sampled from a zero mean univariate normal distribution with variance σ^2. We can characterize the inlier process by the probability density function (PDF) f(.),

$$
f(z_{i,x};\mathbf{s}_i{}^{\mathrm{T}}\mathbf{b}_x,\sigma^2) = \mathcal{N}(z_{i,x};\mathbf{s}_i{}^{\mathrm{T}}\mathbf{b}_x,\sigma^2) \quad (6)
$$

The outlier process is generally unknown, but the unknown PDF can be characterized by a normalized histogram g(.), of which the parameters $\boldsymbol{\beta}_i$ are estimated during the EM procedure. Each histogram bin (equally spaced in our case) corresponds to a range of pixel intensities of which the probability is given by an entry in the parameter vector $\boldsymbol{\beta}_i$. Using $v_{i,x}$ to signal the inlier status of measurement $z_{i,x}$, the individual pixel likelihood can now be summarized as follows,

$$
p(z_{i,x}\,|\,\mathbf{b}_x, v_{i,x},\boldsymbol{\theta}) = \begin{cases} f(z_{i,x};\mathbf{s}_i{}^{\mathrm{T}}\mathbf{b}_x,\sigma^2) & \text{if } v_{i,x} = +1 \\ g(z_{i,x};\boldsymbol{\beta}_i) & \text{if } v_{i,x} = -1 \end{cases}
$$

$$(7)$$

where σ^2 and $\boldsymbol{\beta}_i$ are model parameters. At this point we would like to highlight some differences with earlier approaches [15, 17, 18]. In [15, 17] the matching cost used in the inlier process is some kind of algebraic distance, necessitated by the albedo cancellation (ratio images). In [18], the inlier matching cost was properly defined in the observation space, similar to our own approach, but augmented with a clustering term. These alternative matching costs do not allow for an easy interpretation. Also in [18], by treating pixels individually each has its own noise variance σ^2, which seems unnecessary, and an outlier process with uniform distribution is used, instead of the more expressive histogram proposed here.

The prior $p(\mathbf{b} \mid \boldsymbol{\theta})$ expresses any prior knowledge we have regarding the scaled normals. If nothing is known beforehand, any distribution with very large covariance can be used to express that the scaled normal can be located almost anywhere in scaled normal space. For the mean we

then typically choose a small scaled normal oriented towards the camera. By adding a prior numerical problems are also avoided in case not enough inlier observations are available. The prior can be factorized as follows,

$$p(\mathbf{b}\,|\,\boldsymbol{\theta}) = \prod_x p(\mathbf{b}_x) \qquad (8)$$

where for reasons of simplicity all scaled normals are assumed independent of one another, and of $\boldsymbol{\theta}$. If ever a smoothness on the normals needed to be enforced, this prior would be the place to do it. In our work we assume the individual prior $p(\mathbf{b}_x)$ is modelled by a trivariate normal distribution with mean $\mathbf{m}_{x,prior}$ and covariance $\mathbf{C}_{x,prior}$, as specified here,

$$p(\mathbf{b}_x) = \mathcal{N}(\mathbf{b}_x; \mathbf{m}_{x,prior}, \mathbf{C}_{x,prior}) \qquad (9)$$

In the following we assume no correlation between shadows and highlights of different images. For some cases, such as cavities and images with nearby activated light sources, this assumption is almost certainly violated, however, the experiments validate its use. The prior $p(\mathbf{v}\,|\,\boldsymbol{\theta})$ can be then factorized as follows,

$$p(\mathbf{v}\,|\,\boldsymbol{\theta}) = \prod_i p(\mathbf{v}_i\,|\,\alpha_i) \qquad (10)$$

where α_i is a model parameter describing the fraction of pixels in image i generated by the inlier process. Shadows and highlights will not usually give rise to isolated outlier pixels, but rather form spatially coherent regions in an image. To take this coherence into account, similar to [2], we model each inlier map $\mathbf{v}_i$ as a Markov Random Field (MRF) with an associated Gibbs-prior distribution as follows,

$$p(\mathbf{v}_i\,|\,\alpha_i) = \frac{1}{Z(T)} \exp\left(\frac{-U(\mathbf{v}_i)}{T}\right) \cdot$$
$$\prod_x \alpha_i^{\frac{1+v_{i,x}}{2}} (1-\alpha_i)^{\frac{1-v_{i,x}}{2}} \qquad (11)$$

where

$$U(\mathbf{v}_i) = -\sum_x \sum_{y \in N(x)} v_{i,x} v_{i,y} \qquad (12)$$

is a measure of coherence, lower values mean higher coherence, T is a temperature parameter, Z is a normalization constant and $N(x)$ denotes the 4-neighborhood of x. Instead of defining the fraction of inlier pixels over all the pixel locations in a specific image, the approach in [18] defines it over all measurements in images at specific pixel location. This is again a consequence of their individual treatment of pixels. In one of our earlier approaches we followed a similar path, but found that convergence with EM was unstable. By defining it over all pixel locations within

the same image, the impact of individual pixels on the inlier fraction becomes negligible and therefore has a stabilizing effect. It also means less parameters have to be estimated. A similar argument can be made for σ^2.

The probabilistic model is now fully specified. The next section will give details on inference of the various model parameters and distributions of hidden variables. An EM algorithm is used.

4. Update by Expectation-Maximization

Even for modest sized images, the total number of configurations of $\mathbf{h}$ is huge, hence direct optimization of the log-likelihood in Eq. 2 is infeasible. The Expectation-Maximisation (EM) [3, 11] algorithm offers a solution to this problem. The algorithm is derived using a variational approach. We will briefly mention the steps involved. By Jensen's inequality, which follows from the fact that the log function is concave, a lower bound for the log-likelihood can be written down as follows,

$$\begin{aligned}
\log \sum_{\mathbf{h}} p(\mathbf{z}, \mathbf{h}\,|\,\boldsymbol{\theta}) &= \log \sum_{\mathbf{h}} q_h(\mathbf{h}) \frac{p(\mathbf{z}, \mathbf{h}\,|\,\boldsymbol{\theta})}{q_h(\mathbf{h})} \\
&\geq \sum_{\mathbf{h}} q_h(\mathbf{h}) \log \frac{p(\mathbf{z}, \mathbf{h}\,|\,\boldsymbol{\theta})}{q_h(\mathbf{h})} \\
&= \mathcal{F}(q_h(\mathbf{h}), \boldsymbol{\theta}) \qquad (13)
\end{aligned}$$

Any probability distribution $q_h(\mathbf{h})$ over the hidden variables gives rise to a lower bound $\mathcal{F}(q_h(\mathbf{h}), \boldsymbol{\theta})$, which is the negative of a quantity known in statistical physics as the free energy. Optimizing the functional $\mathcal{F}(q_h(\mathbf{h}), \boldsymbol{\theta})$ with respect to $q_h(\mathbf{h})$ and $\boldsymbol{\theta}$ results in a tightening of the bound. Unfortunately, in many interesting models the data are explained by multiple interacting hidden variables, such as the MRFs in our case, which can result in intractable posterior distributions $q_h(\mathbf{h})$. In the variational approach we can constrain the posterior distributions to be of a particular tractable form. This will not generally result in the bound becoming an equality, unless of course the exact posterior lies in the family of constrained posteriors. We use the mean field approximation, where $q_h(\mathbf{h})$ is fully factorized over the hidden variables,

$$\begin{aligned}
q_h(\mathbf{h}) &= q_b(\mathbf{b}) q_v(\mathbf{v}) \\
&= \left\{ \prod_x q_{b_x}(\mathbf{b_x}) \right\} \left\{ \prod_i \prod_x q_{v_{i,x}}(v_{i,x}) \right\} \quad (14)
\end{aligned}$$

We further constrain $q_{b_x}(\mathbf{b_x})$ and $q_{v_{i,x}}(v_{i,x})$ to have a particular analytical form, as follows,

$$q_{b_x}(\mathbf{b_x}) = \mathcal{N}(\mathbf{b_x}; \mathbf{m_x}, \mathbf{C_x}) \qquad (15)$$

$$q_{v_{i,x}}(v_{i,x}) = \begin{cases} w_{i,x} & \text{if } v_{i,x} = +1 \\ 1 - w_{i,x} & \text{if } v_{i,x} = -1 \end{cases}$$
$$(16)$$

With this choice of constrained posterior and the probabilistic model explained in the previous section, an analytical expression for the lower bound can be derived. The derivation is tedious but relatively straightforward, and will be made available as a technical report if the paper is accepted. Setting to zero the derivatives of this expression with respect to the posterior distribution parameters $\mathbf{m_x}$, $\mathbf{C_x}$, $w_{i,x}$ and the model parameters $\boldsymbol{\theta}$, results in what is classically called the E-step and M-step of the algorithm, respectively. The update equations are given by,

E-step

$$\mathbf{m}_x = (\mathbf{C}_{x,prior}^{-1} + \frac{1}{\sigma^2}\sum_i w_{i,x}\mathbf{s}_i\mathbf{s}_i^{\mathrm{T}})^{-1}.$$
$$(\mathbf{C}_{x,prior}^{-1}\mathbf{m}_{x,prior} + \frac{1}{\sigma^2}\sum_i w_{i,x}\mathbf{s}_i z_{i,x}) \quad (17)$$

$$\mathbf{C}_x = (\mathbf{C}_{x,prior}^{-1} + \frac{1}{\sigma^2}\sum_i w_{i,x}\mathbf{s}_i\mathbf{s}_i^{\mathrm{T}})^{-1} \quad (18)$$

$$w_{i,x} = \frac{1}{1 + \exp(-Q_{i,x})} \quad (19)$$

M-step

$$\alpha_i = \frac{1}{N}\sum_x w_{i,x} \quad (20)$$

$$\sigma^2 = \frac{\sum_i \sum_x w_{i,x} P_{i,x}}{\sum_i \sum_x w_{i,x}} \quad (21)$$

$$\beta_i^k \propto \sum_i \sum_x (1 - w_{i,x})\delta_k(z_{i,x}) \quad (22)$$

and the following definitions,

$$Q_{i,x} \triangleq \log \frac{\frac{\alpha_i}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{P_{i,x}}{2\sigma^2}\right)}{(1-\alpha_i)g(z_{i,x};\boldsymbol{\beta}_i)}$$
$$+ \frac{2}{T}\sum_{y \in N(x)}(2w_{i,y} - 1) \quad (23)$$

$$P_{i,x} \triangleq (z_{i,x} - \mathbf{s}_i^{\mathrm{T}}\mathbf{m_x})^2 + \mathbf{s}_i^{\mathrm{T}}\mathbf{C_x}\mathbf{s}_i \quad (24)$$

where $\delta_k(z_{i,x})$ is an indicator function which evaluates to 1 if the pixel value falls in the kth histogram bin and evaluates to 0 otherwise. The update equations are applied in alternating fashion until convergence, upon which the parameters $\mathbf{m_x}$ and $\mathbf{C_x}$ serve as an estimate of a scaled normal distribution at pixel location x. We found that the order in which these updates are applied is not so important.

Initialisation The algorithm needs some bootstrapping. Initially, for each pixel location the 50% brightest intensities are considered inlier measurements, since outliers will be predominantly dark shadowed pixels. Ignoring the prior terms in Eq. 17, a solution for $\mathbf{m_x}$ independent of σ^2 is obtained. Plugging Eq. 18 (also ignoring the prior term) into Eq. 21 and isolating σ^2, result in an initial solution for σ^2. From this σ^2 and Eq. 18 an initial solution for $\mathbf{C_x}$ follows. Histograms are initialised with each bin having equal probability, and prior inlier fractions are set to 0.5. After this the above update equations are applied in alternating fashion.

5. Results

As acquisition device, a dome structure containing 264 light sources, similar to [9], has been chosen. A digital camera is mounted on top of the dome, facing downwards at the object to be digitized. By construction the directions of illumination are known, but a calibration procedure to determine their intensities is still necessary. In one of the experiments (see Fig. 1) a larger version of this dome was used, where only a quarter of the sphere is covered with 169 light sources and the object of interest is sitting on a turntable viewed by a camera mounted on a gantry moving in the vertical plane. However, this does not mean that the method proposed in this paper is restricted to this type of acquisition devices. In all our experiments 64 histogram bins are used. The number of images used in our experiments always corresponds to the number of light sources.

Bob the Builder Photometric stereo reconstruction results of Bob the Builder are shown in Fig. 5. We show the recovered albedo, a 3D model obtained through normal integration and a synthetically generated image for a virtual light source (relighting). Also, the distinction between normal confidence and scaled normal confidence is illustrated. For surface patches with low albedo (zero in worst case), the normal cannot be estimated with great confidence, indeed large variation in normal orientation will only results in negligible pixel intensity variation. So, it is possible to have high confidence in the scaled normal estimate but low confidence in the normal estimate, which is most noticable on the dark feet of Bob. Given the scaled normal covariance $\mathbf{C}_b$, the normal covariance $\mathbf{C}_n$ can be shown to be $\frac{\partial \mathbf{n}}{\partial \mathbf{b}}\mathbf{C}_b\frac{\partial \mathbf{n}}{\partial \mathbf{b}}^T$.

Ground truth As part of our ground truth experiments we used rendered images of a sphere, examples of which are shown in Fig. 2. A Phong reflectance model was used and after rendering noise was added. The error between estimate and groundtruth is measured in average angular error of the normals (in degrees). In Fig. 3(a) comparisons are shown of the algorithm proposed here (without outlier coherence) and our implementation of [18]. We believe the clearly better results of our approach are due to a more nat-
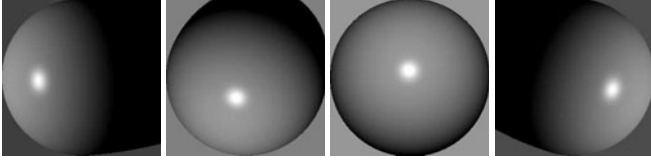
Figure 2. 4 example images used in the ground truth experiments.

Figure 4. Influence of the temperature parameter on posterior inlier probability coherence. From left to right the temperature is set to 50, 5 and 0.5, respectively.

ural formulation of the problem, and by not assigning a different prior inlier fraction and noise variance to each individual pixel location. In addition, we are also faster because we do not have to solve for an initial set of normals equal to the number of images, however, no running times are available. In Fig. 3(a) the noise averaging becomes evident as the number of images increases. In Fig. 3(b) and 3(c), we investigate the effect of the coherence temperature on the error. Both plots show a decreasing error for decreasing temperature, however, in Fig. 3(c) the error increases again for very low temperatures. At these temperatures, details in the inlier maps are elimated, the effect of which seems to be more pronounced in the error when more images are used as in Fig. 3(c).

Statue Photometric stereo reconstruction results of a small statue are shown in Fig. 1. For the acquisition the camera was placed such that only the right side of the statue was visible to the quarter sphere illuminating it. Taking a closer look, one notices that for surface patches on the left of the statue, which usually find themselves in the shadow, the reconstructed normals are noisier. These surface patches, having only a few or no inlier measurements, are clearly of lower confidence. The confidence estimation of our procedure clearly identifies this situation, and apropriately assigns a lower confidence to these surface patches, in Fig. 1 the confidence visualization is clearly lighter on the left (lower confidence) than on the right (higher confidence). Also shown is the posterior inlier probability, in which the cast shadows have been correctly detected as outliers. The specular reflections on the knees are detected as well.

By lowering the temperature parameter more coherence is enforced on the hidden inlier variables, the effect of which is clearly visible in Fig. 4. Lowering it too much causes small details to be eliminated.

Relighting There are applications where a fixed viewpoint suffices, but where one would like to interactively move around a virtual light source. One of these is cuneiform tablet reading, where e.g. assyriologists get a better idea of the indentations on the tablet by casting shadows in that manner. The most important property in conveying these indentations to a user are the surface normals. The surface normals and albedos recovered with our method can be used in combination with a virtual light source to generate corresponding virtual images by evaluating the ren-

dering equation Eq. 1, see Fig. 6. Shading can be done in real-time through the use of pixel shaders in graphics cards.

Non-Lambertian Finally, we show successful reconstructions for two challenging cases, see Fig. 7. In the first case, the surface is very specular, and for the second case, a piece of cloth with anisotropic material properties is reconstructed.

6. Conclusion

In this paper we proposed a robust method for dense photometric stereo reconstruction that can handle shadows and non-Lambertian effects. The redundancy present in a dense set of noisy images is fully exploited. Measurements not consistent with the underying Lambertian assumption are properly identified by a probabilistic model through the use of hidden inlier variables. The coherence of the hidden inlier variabels is expressed by a Markow Random Field prior. Confidence estimation of the desired scaled normals, is incorporated in our approach by considering them hidden variables. Inference of the Maximum-Likelihood solution, is made possible by a mean field Expectation-Maximization approach. Very good results have been obtained on a variety of challenging problems, such as severe shadows, specular surfaces and even surfaces with anisotropic material properties.

Acknowledgments

References

[1] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Anal-*

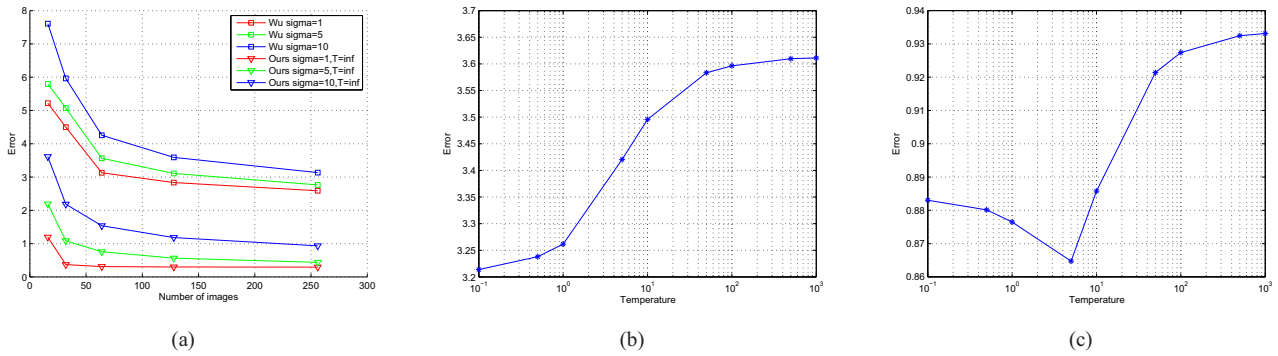(a)   (b)   (c)

Figure 3. Top row: Comparison of [18] with our algorithm for various noise levels and number of images N used. Bottom row: Influence of the coherence temperature on the error for $N = 16$ (left image) and $N = 256$ (right image) and with $\sigma = 10$.
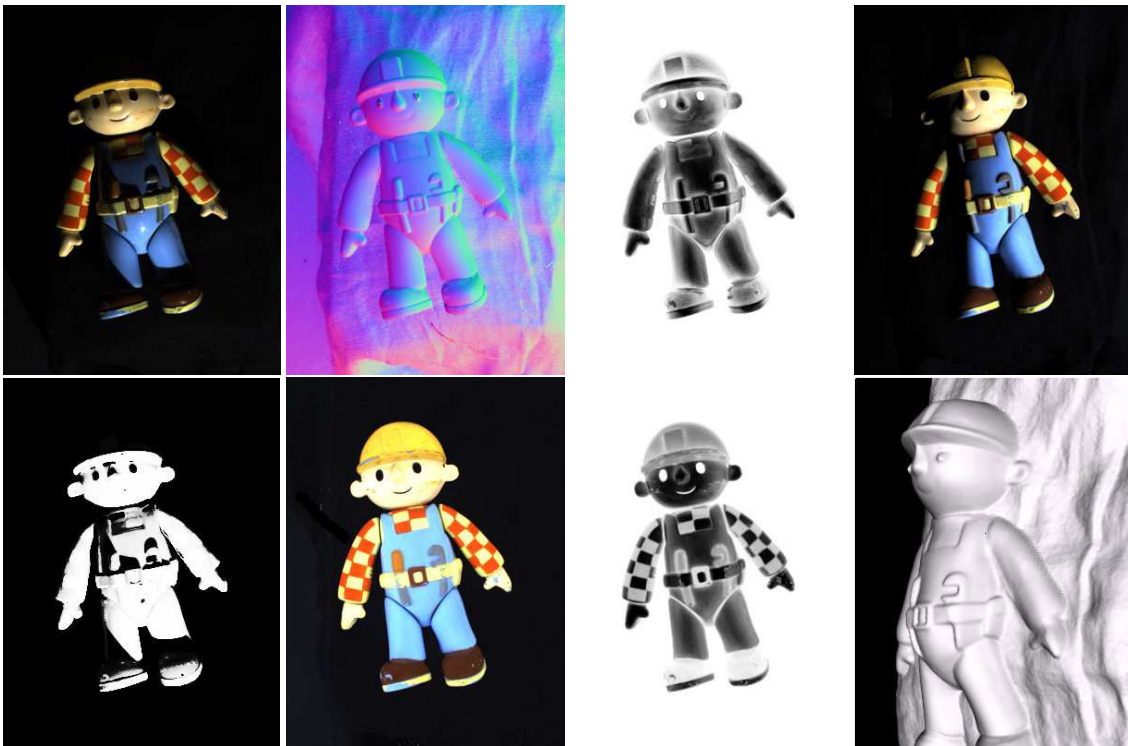

Figure 5. Top row from left to right: Captured image of Bob the Builder, recovered normal map, scaled normal confidence and a relighting of Bob using the recovered scaled normals. Bottom row from left to right: Posterior inlier probability, recovered albedo map, normal confidence and a rendering of the 3D model obtained through normal integration.

ysis and Machine Intelligence, 25(10):1239–1252, October 2003. 2

[2] M. De Smet, R. Fransens, and L. Van Gool. A generalized em approach for 3d model based face recognition under occlusions. In *Proc. of CVPR*, pages 1423–1430, 2006. 2, 4

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the *em* algorithm. *Journal of the Royal Statistical Society B*, 39(1):1–38, 1977. 4

[4] J. E. North Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18:309–328, April 1982. 2

[5] P. Gargallo and P. Sturm. Bayesian 3d modeling from images using multiple depth maps. In *Proc. of CVPR*, pages 885–891, 2005. 2

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

[7] A. Hertzmann and S. Seitz. Shape and materials by example: A photometric stereo approach. In *Proc. of CVPR*, pages

Figure 6. Example of an application in cuneiform reading. From left to right: Recovered normal map, albedo map and a relighting for a virtual light source using the recovered surface properties.
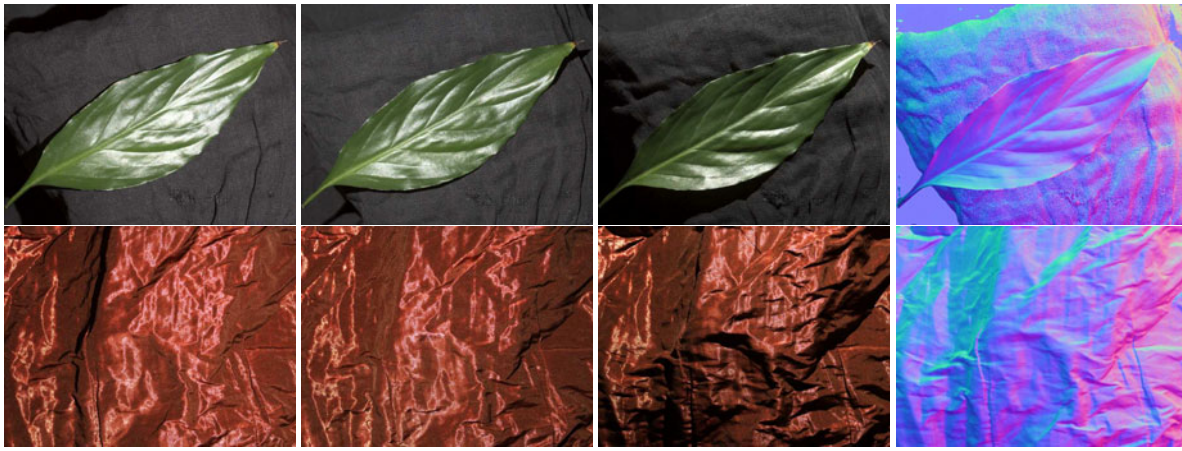

Figure 7. Top row: Three images of a highly specular leaf, and its recovered normal map. Bottom row: Three images of a piece of cloth with anisotropic material properties, and its recovered normal map.

533–540, 2003. 2

[8] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *Proc. of CVPR*, pages 619–626, 2005. 2

[9] T. Malzbender, D. Gelb, and H. Wolters. Polynomial texture maps. In *Proc. of SIGGRAPH 2001*, pages 519–528, 2001. 5

[10] S. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(4):418–431, August 1990. 2

[11] R. Neal and G. Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*. Kluwer, 1998. 4

[12] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: A probabilistic account. In *Proc. of CVPR*, pages 552–559, 2004. 2

[13] J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003. 2

[14] H. Tagare and R. deFigueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):133–152, 1991. 2

[15] K.-L. Tang, C.-K. Tang, and T.-T. Wong. Dense photometric stereo using tensorial belief propagation. In *Proc. of CVPR*, pages 132–139, 2005. 2, 3

[16] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 1, 2

[17] T.-P. Wu and C.-K. Tang. Dense photometric stereo using a mirror sphere and graph cut. In *Proc. of CVPR*, pages 140–147, 2005. 2, 3

[18] T.-P. Wu and C.-K. Tang. Dense photometric stereo by expectation maximization. In *Proc. of ECCV*, volume 3954, pages 159–172, 2006. 2, 3, 4, 5, 7