

# Inverse Lightfield Rendering for Shape, Reflection and Natural Illumination

Antonin Sulc, Ole Johannsen, and Bastian Goldluecke

University of Konstanz,  
antonin.sulc@uni-konstanz.de

**Abstract.** We propose an inverse rendering model for light fields to recover surface normals, depth, reflectance and natural illumination. Our setting is fully uncalibrated, with the reflectance modeled with a spatially-constant Blinn-Phong model and illumination as an environment map. While previous work makes strong assumptions in this difficult scenario, focusing solely on specific types of objects like faces or imposing very strong priors, our approach leverages only the light field structure, where a solution consistent across all subaperture views is sought. The optimization is based primarily on shading, which is sensitive to fine geometric details which are propagated to the initial coarse depth map. Despite the problem being inherently ill-posed, we achieve encouraging results on synthetic as well as real-world data.

**Keywords:** lightfield, inverse rendering, BRDF, natural illumination

## 1 Introduction

The irradiance of a ray received by a sensor depends on many factors, in particular surface properties like geometry and reflectance, but also the illumination from the environment. The difficult ill-posed problem of inverse rendering aims at recovering all such variables which lead to the formation of an image of a scene. While there has been tremendous progress in the recovery of each of these variables separately [13,15] or under very restrictive conditions [1,2], the general question of reconstructing a geometry, natural illumination and non-Lambertian reflectance from an image has not been addressed yet. The main problem is that an image can be explained in numerous ways, therefore strong priors or additional information such as multiple views are required. Thus, most of previous works based on traditional images focus on recovery of an object either with priors which restrict the set of possible solutions [3], restrict the type of analyzed object [2] or require additional information either about illumination [15] or shape [13].

In contrast, we rely only on the structure of the lightfield. Lightfield imaging is an efficient way to acquire a structured array of multiple views with a small baseline which attracted lots of attention in the past decade [23,26,8,11,24]. Despite the fact that the idea of lightfield is relatively old, the recent availability of hand-held cameras like the Lytro Illum or Raytrix gives a chance to answer many questions which for 2D imaging were ill-posed in practice using comparably simple off-the-shelf devices. Significant progress has, for example, been made in Lambertian depth estimation [12], where the

appearance of a surface point does not change for different viewpoints. However, it has also been shown that due to the rigid structure of a 4D lightfield, recovery of surface normals and single-lobe BRDF is robustly possible for non-Lambertian objects in case of a calibrated single distant light source [25].

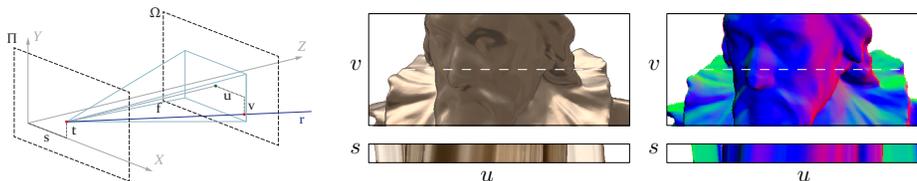
The problem of inverse rendering is severely ill-posed but the specific structure of the 4D lightfield as a regular grid of views gives us an opportunity to enforce a prior which encourages consistent solutions between views and does not suffer from the bias towards a certain class of solutions. Furthermore, although most algorithms for depth estimation fail when the Lambertian assumption is violated, a loss in accuracy is usually visible mostly on geometric details on fine scales, while the depth map is on a coarse level accurate enough. This is sufficient as an initial guess of geometry, which is necessary to initially estimate other missing variables as a starting point for further refinement.

**Contributions.** In this work, we present the first complete inverse lightfield rendering pipeline which can recover shape, reflectance, and natural illumination. The model is inspired by [13], where they formulated estimation of natural illumination and reflectance given surface shape. However, by leveraging the lightfield structure, we can in addition impose priors on geometry, i.e. depth and normals. We do this by employing regularization based on [21,26], which encourages the normal field to be consistent with depth and in particular respect the epipolar geometry across subaperture views. Overall, we obtain a unified energy-based formulation, where we can optimize for all variables in turn. In contrast to previous work based on individual images [3], our model does not overly restrict the space of solutions.

## 2 Related work

There have been tremendous efforts in lightfield depth estimation, nevertheless most approaches allow only very small deviation from the Lambertian assumption, if any. For an in-depth review of recent Lambertian methods, we refer to [12]. A recent lightfield based method which does not assume Lambertian reflectance is [25], where they use differential motion to estimate normals under the assumptions of Blinn-Phong reflectance and a single calibrated point light source placed at infinity. They can also recover BRDF parameters, but in contrast to their work, we estimate unknown natural illumination. In [15] they proposed an approach for shape estimation under the assumption of known natural illumination. They used strong geometric priors in order to enforce smoothness while maintaining similarity from exhaustive search in a discrete set of normals. Since the normals are estimated from a look-up table, the accuracy is highly dependent on the quality of input illumination. The work of [15] is further extended to multiple views [16] and can recover full 3D geometry of the object.

Shape-from-shading based methods underwent promising development under challenging conditions like natural illumination [10]. Here, light is modeled with spherical harmonics and the objects are assumed to be Lambertian. In [3], they search for the most likely parameters for geometry, reflectance, and illumination that would explain a single input image. In contrast to our work, they employ very strong priors to prune the set of possible solutions.



**Fig. 1.** *Light field and epipolar plane images.* Left: An incident ray  $\mathbf{r}$  is parametrized by its intersections with the *focal plane*  $\Pi$  and the *image plane*  $\Omega$  (red dots). The planes are parallel with distance equal to the focal length  $f$ . The intersection coordinates  $(s, t)$  are given in relation to the origin of the world coordinate system. The coordinates  $(u, v)$  are given relative to the intersection of the optical axis of a virtual camera placed at  $(s, t, 0)$  in  $Z$  direction with the second plane (green dot). Each of these virtual cameras gives a subaperture view of the light field. Middle: an example center view of a non-Lambertian object. One can observe in the EPI below that not all lines exhibit constant values, which is caused by viewpoint dependency of the material. Right: illustration of the normal map  $\mathbf{n}$  for the center view. The restriction  $\mathbf{n}_{tv}$  for fixed  $(t, v)$ -coordinates below shows that while for the input image the appearance of some points may change with viewpoint, normals stay constant and exhibit the line pattern.

Our work is partially based on [13]. They formulated the problem of estimation of natural illumination and reflectance as a maximum a-posteriori problem. In contrast to approaches where they use spherical harmonics [10] to represent illumination, they discretize distant illumination on the space of illumination directions. However, in contrast to their work, we do not assume known geometry, which makes the problem much harder. In [19] they proposed a method for BRDF estimation from a single image of a known shape in an unknown natural illumination. They leveraged a statistic about real-world illumination and estimated the reflectance that is most likely under a distribution of probable illumination environments.

### 3 Inverse lightfield rendering model

#### 3.1 Lightfield parametrization and rendering equation

In this paper, we adopt the absolute two-plane parametrization for a lightfield. The rays are parameterized as an intersection of two parallel planes, a *focal plane*  $\Pi$  and *image plane*  $\Omega$ , where the focal plane  $\Omega$  is closer to the scene at distance  $f$  from  $\Pi$ . Each ray is described by a four-dimensional vector  $\mathbf{r} = [u, v, s, t] \in \mathbf{R}^4$ . The intuition is that the coordinates  $(s, t) \in \Pi$  select the focal point of an ideal pinhole camera with a corresponding *sub-aperture view*, while  $(u, v) \in \Omega$  corresponds to image coordinates within this view, see figure 1 for an illustration.

The way how each image is formed is described by the *rendering equation*. The radiance  $R$  observed for a ray  $\mathbf{r}$  is a function of several variables. It depends (i) on the surface normal  $\mathbf{n}(\mathbf{r})$  at the point in 3D space where the ray intersects the scene geometry, (ii) on the reflectance function (BRDF) at this location, given by a global parametric model depending on parameters  $\Psi$ , and (iii), on global illumination  $\mathbf{L}$ . In particular, all rays which intersect in the same 3D point have the same normal associated

to them, a fact which we employ later to impose consistency. We have

$$R(\mathbf{r}, \mathbf{L}, \Psi) = \int_{H_{\mathbf{n}(\mathbf{r})}} \text{BRDF}_{\Psi}(\mathbf{v}(\mathbf{r}), \mathbf{n}(\mathbf{r}), \mathbf{l}) \mathbf{L}(\mathbf{l}) d\mathbf{l}, \quad (1)$$

where integration is performed over the hemisphere  $H_{\mathbf{n}(\mathbf{r})}$  of illumination directions  $\mathbf{l}$  which are oriented into the direction of the surface normal  $\mathbf{n}(\mathbf{r})$ , i.e.  $\mathbf{l} \cdot \mathbf{n}(\mathbf{r}) \geq 0$ , and  $\mathbf{v}(\mathbf{r})$  is the viewing direction for the ray. Global illumination  $\mathbf{L}$  is assumed to be distant and thus just a function of direction  $\mathbf{l}$ . Note that we can easily specialize to a setting with just one point light source by setting illumination to a delta distribution over the set of directions, in which case the integral reduces to a single evaluation of the reflectance function. In the following, we discuss the specific parametric models we use for BRDF and illumination, as well as our priors on geometry imposed by the lightfield structure.

### Reflectance and illumination

Reflectance is the ratio of the radiance of light which was reflected by material and the irradiance of incident light  $\mathbf{L}(\omega_i)$ . In order to keep the model simple, we model the reflectance with a spatially uniform Blinn-Phong parametric model, but model can be easily replaced with a more complex reflectance. The specular lobe is parametrized with respect to the half-angle vector  $\mathbf{h}$  and the normal  $\mathbf{n}$ . The constant  $\alpha > 0$  is the *shininess* of the material, while  $i^d, i^s \in \mathbb{R}^3$  are diffuse and specular color, respectively. The resulting BRDF is

$$\text{BRDF}_{\Psi}(\mathbf{v}, \mathbf{n}, \mathbf{l}) = i^d (\mathbf{n}^T \mathbf{l}) + i^s (\mathbf{n}^T \mathbf{h})^{\alpha} \quad (2)$$

where the  $\mathbf{h} = \mathbf{n}^T \mathbf{v}$  is half vector between normal and viewing direction [20].

The general assumptions about illumination are that it is natural, infinitely far from an observer. We adopt a similar approach to [15,13], where the illumination is discretized on the domain of  $\mathbf{l}$ . A linear combination of spherical harmonics [10] can be considered as an alternative for the illumination. See section 4 for further details.

### Depth and normal maps

We recover depth  $z = z(\mathbf{r})$  for each individual ray as the distance of the focal plane to the 3D point on the scene surface intersected by the ray. Knowing depth immediately gives us 3D coordinates of the point, and we can in turn recover disparity, i.e. the ratio of shift in the  $(u, v)$  plane vs. shift in the  $(s, t)$  plane when changing viewpoint [26]. In addition, we require normal vectors  $\mathbf{n}(\mathbf{r})$  for each ray to evaluate the rendering equation 1. While normals could be computed directly from the depth map  $z$  via suitable derivatives, in our approach, we decided to introduce two separate variables and link them via soft constraints. One of the important motivations is that normals are more sensitive on fine geometric details, which we are lacking on the smoothed input depth map. Therefore, the results of an optimized normal field provide a valuable clue about fine geometric details.

Our priors for geometry follow directly from the structure of the lightfield. We further impose area minimization and piecewise smoothness of the normal map to regularize the surface geometry. For this, we need to leverage and combine several insights from previous work. In [7], it was shown that if depth  $z$  is reparametrized in a new variable  $\zeta := \frac{1}{2}z^2$ , then the linear operator  $N$  given by

$$N(\zeta) = \begin{bmatrix} -\zeta_u/f \\ -\zeta_v/f \\ \hat{u}\zeta_u/f + \hat{v}\zeta_v/f + 2\zeta/f^2 \end{bmatrix} \quad (3)$$

maps a depth map  $\zeta$  to the map of corresponding normals scaled with the local area element of the parametrized surface. Above,  $(\hat{u}, \hat{v})$  are the homogenous image coordinates of the ray where the normal is computed, in particular,  $N$  varies across rays, while  $\zeta_u$  and  $\zeta_v$  denote partial derivatives of reparametrized depth with respect to the spatial view coordinates.

We leverage this map in several ways. First, we follow [7] and introduce a minimal surface regularizer by encouraging small total  $\|N\zeta\|$ , which corresponds to the surface area. Second, we follow [21], and encourage similarity of the normal map  $\mathbf{n}$  scaled by local area element  $\alpha$  and  $N\zeta$ . They enforce consistency of normals and depth as a soft constraint by penalizing  $\|N\zeta - \alpha\mathbf{n}\|_2^2$ . In particular, they propose an optimization strategy to jointly optimize for depth and  $\alpha$ , which we make use of in our framework, see below for more details.

### Consistency with the lightfield structure

The rigid structure of the lightfield becomes visible when considering *epipolar plane image space*, which is the restriction of ray space to 2D horizontal  $(s, u)$  or vertical  $(t, v)$  coordinates, respectively. In particular, a non-occluded point of a perfectly Lambertian surface has the same radiance regardless of viewpoint, thus all its observations in the lightfield should be the same. As it is well known [4,26], projections of a single point lie on a line in horizontal and vertical epipolar plane image space, whose orientation  $\mathbf{q} = [d, 1]^T$  can be computed from disparity  $d$ . This becomes immediately visible when considering the horizontal and vertical epipolar plane images (EPIs)  $I_{su}$  and  $I_{tv}$  of a Lambertian lightfield  $I$ , see figure 1, which consists of patterns of lines. Determining their slope is the same as reconstructing disparity [26].

In [26], this constancy was encouraged by means of an anisotropic regularizer  $J_{\mathbf{q}}$ , which for a vector-valued function  $\mathbf{u} : \Omega \rightarrow \mathbb{R}^n$  and a point-wise orientation field can be defined as

$$J_{\mathbf{q}}(\mathbf{u}) = \sum_{i=1}^n \int_{\Omega} \|\nabla u_i^T (\mathbf{q}\mathbf{q}^T) \nabla u_i\|_2 \, dx. \quad (4)$$

This regularizer prefers functions which are constant along  $\mathbf{q}$ . In our case, it is not the radiance which is constant for a surface point, as we are considering non-Lambertian scenes. However, an invariant property is the surface normal. Thus, a well-defined reconstruction of normals  $\mathbf{n}$  should be consistent with the disparity map in the sense that

it is constant along the corresponding directions on epipolar plane image space. We thus define a regularization term to obtain a consistent normal field as

$$J_d(\mathbf{n}) = \int J_{[d,1]}(\mathbf{n}_{su}) d(s, u) + \int J_{[d,1]}(\mathbf{n}_{tv}) d(t, v), \quad (5)$$

where we sum up the contributions of all epipolar plane images. See [26] for details.

### Final inverse rendering model

Putting together the terms from the previous subsections, we arrive at the energy

$$\begin{aligned} E(\zeta, \mathbf{n}, \Psi, \mathbf{L}) &= E_{\text{data}}(\mathbf{n}, \Psi, \mathbf{L}) + E_{\text{depth}}(\zeta, \mathbf{n}) + E_{\text{normals}}(\mathbf{n}), \\ \text{with } E_{\text{data}}(\mathbf{n}, \Psi, \mathbf{L}) &= \int_{\mathcal{R}} (R(\mathbf{n}(\mathbf{r}), \mathbf{L}, \Psi) - I(\mathbf{r}))^2 d\mathbf{r}, \\ E_{\text{depth}}(\zeta, \mathbf{n}) &= \|N\zeta - \alpha\mathbf{n}\|_2^2 + \|N\zeta\|_{2,1}, \\ E_{\text{normals}}(\mathbf{n}) &= J_{d(\zeta)}(\mathbf{n}), \end{aligned} \quad (6)$$

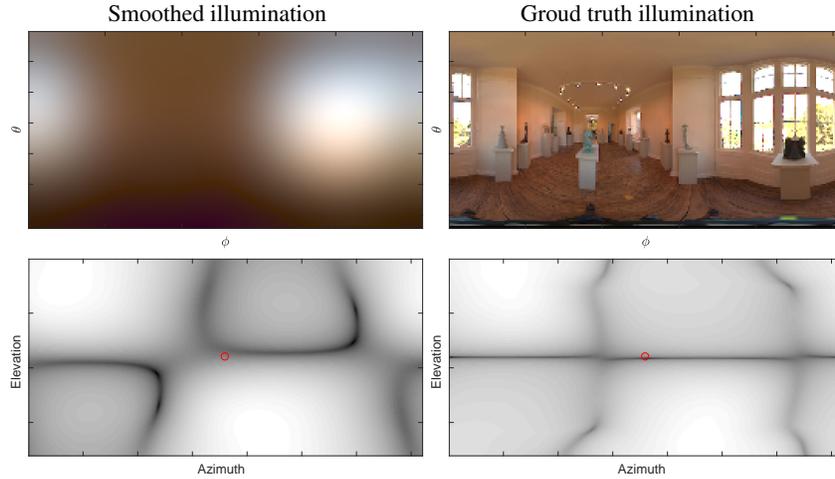
to be minimized for ray-wise depth  $\zeta$ , surface normal  $\mathbf{n}$ , global BRDF parameters  $\Psi$ , and global illumination  $\mathbf{L}$ .

Above, the data term  $E_{\text{data}}$  enforces similarity of the rendering with the input lightfield  $I$ . It is defined as an integral of the squared error over all rays. However, as we currently have only one global BRDF, we have to restrict to the rays which actually hit the object, and perform a semi-automatic pre-segmentation of the lightfield using [27]. The depth term  $E_{\text{depth}}$  links the depth map to the estimated normals, and is used for depth refinement. For depth regularization, we minimize the area of the scene surface as suggested in [7]. The scaling factor  $\alpha$ , equal to the local area element, is recovered during optimization [21]. Finally, the normal term  $E_{\text{normals}}$  regularizes the normal map and makes it consistent with the epipolar plane image structure encoded in the disparity map, as detailed above. Note that disparity can be computed directly from  $\zeta$  using the standard lightfield projection equations [26].

In the following section, we describe the strategy for optimization, which has to be very carefully performed as the energy is highly non-convex.

## 4 Optimization

In this section we describe how we minimize energy (6) to recover depth reparametrized as  $\zeta$ , normals  $\mathbf{n}$ , BRDF parameters  $\Psi$  and illumination  $\mathbf{L}$ . Overall, it follows three main iterative steps. The first is optimization with respect to variables  $\mathbf{n}, \mathbf{L}, \Psi$  which appear in the rendering equation (1). For BRDF parameters and illumination, we compute a global solution given all of the other variables. For the normals, we compute a descent step. The second is refinement of depth  $\zeta$  given the normal field, i.e. descent with respect to  $E_{\text{depth}}$  with fixed  $\mathbf{n}$ . In the third step, we finally adapt the normals  $\mathbf{n}$  to make them consistent with the lightfield structure. This is a descent step with respect to  $E_{\text{normals}}$ . All three steps are iterated until convergence to a local minimum. For initialization,



**Fig. 2.** *Influence of illumination on energy landscape for normals.* The top images show two different input illumination maps parametrized with respect to a normal with the direction  $[0; 0; 1]$ . Right is the ground truth, left a smoothed version of the ground truth. The figures at the bottom show how the log of the data term for different orientations of a normal for these two illumination maps. The correct normal is indicated by a red circle. One can see that the correct normal orientation can be coarsely identified in the case of ground truth illumination, but optimization is still quite prone to get stuck in a local minimum. For the incorrect smoothed illumination, the global minimum actually gives the wrong result.

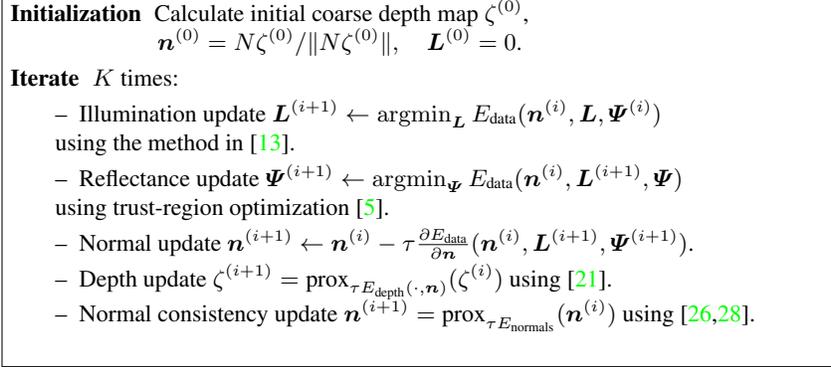
we require an initial geometry. For this, we compute a coarse depth map under the assumption that the scene is approximately Lambertian. In practice, we use [21], and apply strong regularization. The algorithm is summarized in figure 3. We now describe the descent steps in more detail.

### Optimization for $E_{\text{data}}$

In our framework, the data term  $E_{\text{data}}$  in the unknowns  $\mathbf{n}$ ,  $\Psi$  and  $\mathbf{L}$  is optimized in a relatively straight-forward manner by computing the gradients of the rendering equation finding a maximum-likelihood solution.

**Reflectance.** The Blinn-Phong reflectance is defined via the shininess parameter  $\alpha$ , the color of specular lobe  $\mathbf{i}^s = [i_r^s, i_g^s, i_b^s]^T$  and color of diffuse albedo  $\mathbf{i}^d = [i_r^d, i_g^d, i_b^d]^T$ , which make up the seven unknowns summarized in the vector  $\Psi$ . Because reflectance parameters occur only in the data-term, optimization is simple. The rendering equation is differentiable in all variables, and we can easily calculate explicit derivatives with respect to  $\Psi$ . In each iteration of reflectance estimation, the parameters  $\Psi$  are optimized with trust-region optimization [5] with constraints  $\alpha > 0$ ,  $\mathbf{i}^d \geq 0$  and  $\mathbf{i}^s \geq 0$ .

**Illumination.** The illumination  $\mathbf{L}$  is distant and thus a function of illumination direction  $\mathbf{l}$ . In our implementation, it is discretized as a 2D high-dynamic-range image, where coordinates represent azimuth and height of  $\mathbf{l}$ . Same as the reflectance, the illumination occurs only in data-term and explicit derivatives of  $\mathbf{L}$  are easy to calculate.



**Fig. 3.** Overview of the optimization algorithm for energy (6).

It is important to highlight that if the object’s surface is not a perfect mirror, material plays the role of a low-pass filter [18]. Therefore, there is a loss in high frequency components of illumination, which makes it impossible to recover fine details in incoming illumination. Each iteration of illumination estimation entails a few iterations of a limited memory BFGS-B [6].

**Normals.** Neither reflectance nor illumination is likely yet accurate enough to render a correct normal, as the energy landscape for normals is drastically influenced by both illumination and reflectance, see figure 2 for an illustration. For this reason, we do not look for a global minimum, but perform only a few steps of ray-wise gradient descent for the normals, with explicit exact derivatives calculated for the rendering equation in a straight-forward way. Unit length of the normals is enforced with a Lagrange multiplier applied during these gradient descent steps. Thus, we give the regularizer a better chance to propagate information across the lightfield in the next step, and avoid local minima.

### Refinement of depth by optimizing $E_{\text{depth}}$

We want to update our perspective depth map  $\zeta$  to account for the currently estimated normals  $\mathbf{n}$ . The framework for this was established in [22], which makes use of the linearized relationship of depth to normals from [7]. The optimization problem we solve in this step is an implicit subgradient descent for  $E_{\text{data}}$  in  $\zeta$  given the current solution  $\zeta^{(i)}$ , i.e.

$$\zeta^{(i+1)} = \operatorname{prox}_{\tau E_{\text{depth}}(\cdot, \mathbf{n})}(\zeta^{(i)}) = \operatorname{argmin}_{\zeta} \left\{ \frac{\|\zeta - \zeta^{(i)}\|_2^2}{2\tau} + E_{\text{depth}}(\zeta, \mathbf{n}) \right\}. \quad (7)$$

Once the operator  $N$  in  $E_{\text{depth}}$  has been computed, minimization is essentially a straight-forward implementation of [17]. Details can be found in [7,22]. Note that  $E_{\text{depth}}$  is convex in  $\zeta$ , so a solution can be uniquely determined.

### Consistent normals via optimization of $E_{\text{normals}}$

Finally, we compute an implicit subgradient descent step for the normal map regularizer  $J_d$  in  $E_{\text{normals}}$  to enforce consistent normals across the lightfield. This means we have to solve essentially a  $L^2$ -denoising problem for the current normal map  $\mathbf{n}^{(i)}$ ,

$$\mathbf{n}^{(i+1)} = \text{prox}_{\tau E_{\text{normals}}}(\mathbf{n}^{(i)}) = \underset{\mathbf{n}}{\text{argmin}} \left\{ \frac{\|\mathbf{n} - \mathbf{n}^{(i)}\|_2^2}{2\tau} + J_d(\mathbf{n}) \right\}. \quad (8)$$

This is done by using the idea in [28] of successively linearizing and updating the normals around the current solution to deal with the unit length constraint, and embed it in the framework of [26] for consistent regularization with respect to the lightfield structure imposed by current disparity  $d$ . For details, we refer to the respective papers and [22]. While the energy  $E_{\text{normals}}$  is formally convex in  $\mathbf{n}$ , the constraint  $\|\mathbf{n}\|_2 = 1$  is not, adding to the overall non-convexity and difficulty of the problem.

## 5 Experiments

For our experiments, we render several lightfields using a high-resolution HDR environment map for illumination<sup>1</sup>. In addition, we recorded a real-world scene using a Baumer LXG-200C camera designed for industrial inspection, mounted on a high-accuracy gantry. The camera was calibrated to minimize distortion and the disparity lies in  $[-1.5; 1.5]$ .

For the synthetic scenes first normals for all subaperture views are generated using the light field Blender Addon [9] and afterwards each view rendered with the engine provided in [13]. The calculations are performed with CUDA custom kernels on a Geforce GTX 1080 Ti GPU. The code will be publicly available upon publication of the paper. For the initial estimation of geometry, we use the method in [26], which is simple but achieves an accurate enough initial estimate of depth map  $\zeta^{(0)}$  after applying strong TV- $L^2$  denoising [17].

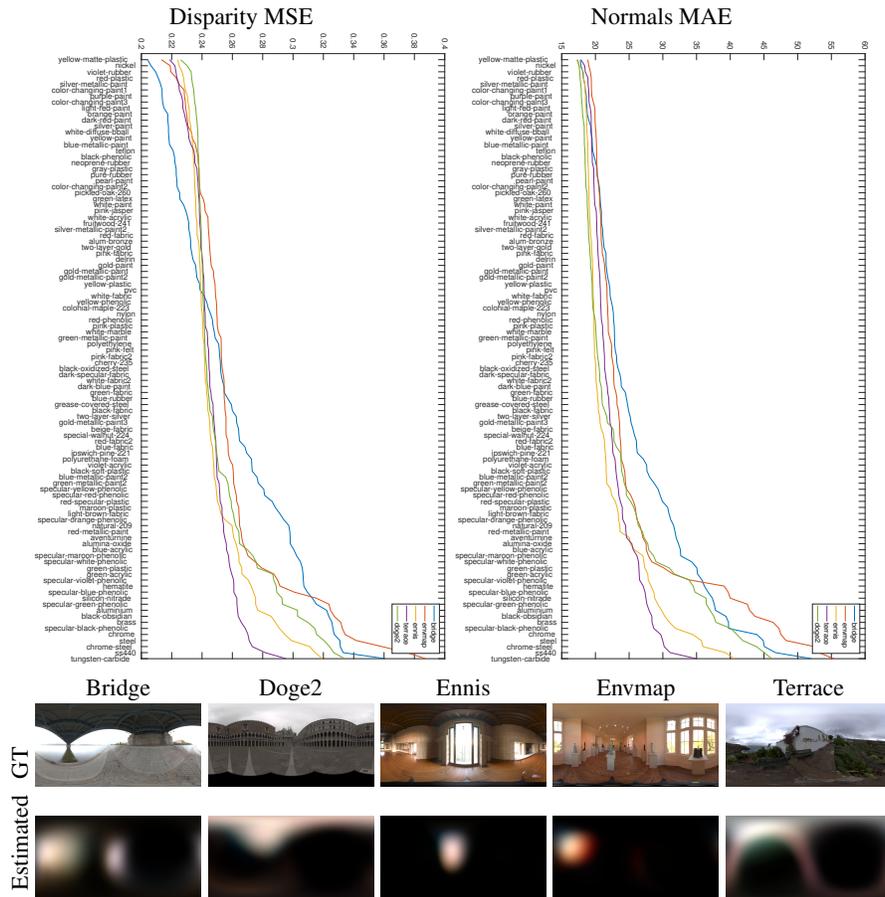
The inverse rendering problem is of course highly ill-posed, which is underlined by two observations. The first is that it is very complicated to separate color of illumination from the color of reflectance, and there is inherent ambiguity in how strong a material reflects vs. brightness of illumination. While possible in theory, this works only for very accurate geometry [13], which is difficult to achieve from our rough initial guess. The second problem arises from the nature of the BRDF. The BRDF is a low-pass filter on the illumination [18] and limits possible higher details of illumination.

We evaluate our method on three different synthetic models, a simple smooth sphere and two objects with complex geometry. The first lightfield shows the sphere, see figure 5. We evaluated mean angular error (MAE) of the estimated normal field and mean squared error (MSE) of the estimated disparity on all materials from MERL database [14] and five different environment maps, see figure 4. The second lightfield *lion* shows a statue with more complex geometry, rendered with the *light-red-paint* MERL BRDF and environment map *Envmap*, see figure 6. The third lightfield *cotton1* is rendered with

<sup>1</sup> Available online <http://gl.ict.usc.edu/Data/HighResProbes/>

the *delrin* MERL BRDF and environment map *Terrace*, see figure 7. Reflectance and illumination have very similar colours, therefore reflectance suffers from the above-mentioned illumination-reflectance ambiguity. The relatively small value of MAE is attributed to the good identification of strong light sources in the scene, however, some persistent inaccuracies in geometry cause the shininesses  $\alpha$  to be too small.

Finally, figure 8 shows results for a real-world lightfield. The dot in the left part of illumination shows a correctly reconstructed point light source, while the smooth right part is from light reflected from a slightly reflective table.



**Fig. 4.** Evaluation of a *ball* dataset MERL BRDF database [14]. The first row shows disparity MSE and normals MAE of our method on all MERL BRDFs with various environment maps. The second row shows ground truth environment maps used for reddening. The last row shows respective estimated environment maps averaged over all MERL materials

## 6 Conclusion

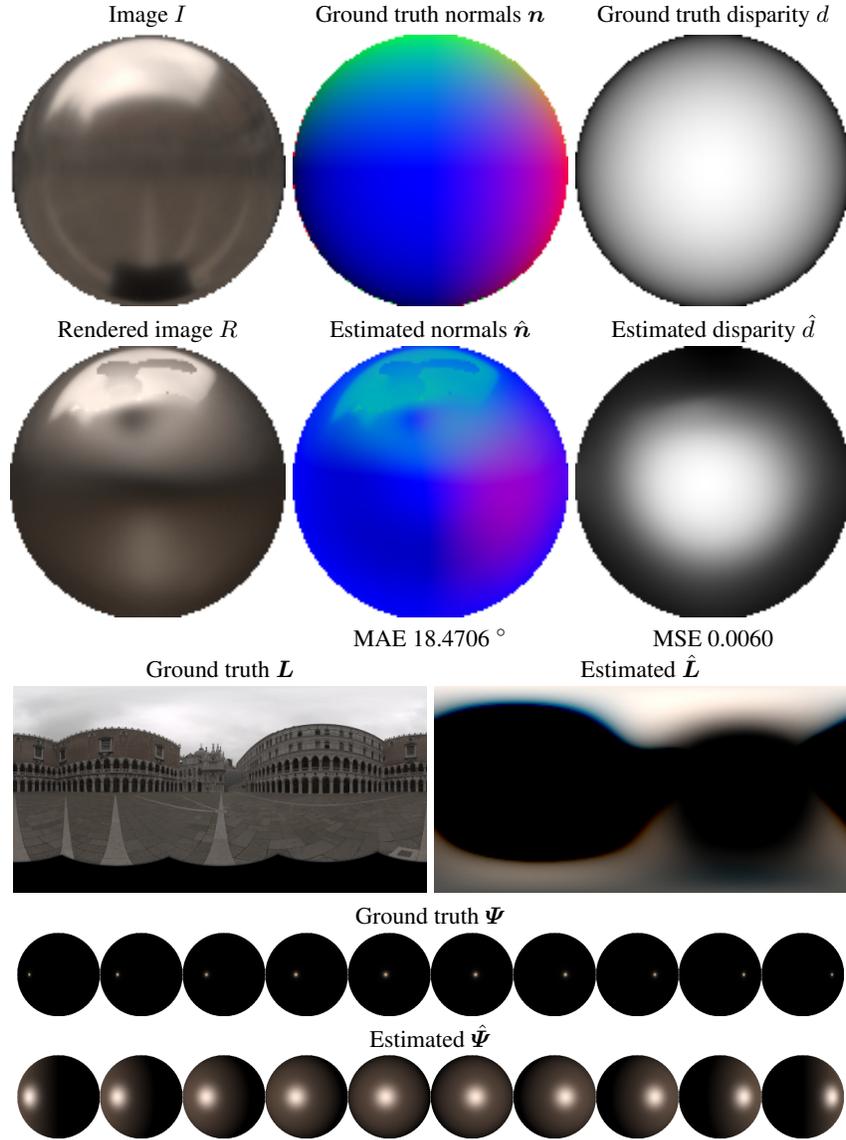
We propose the first approach to recover geometry, reflectance and illumination from a single lightfield. Our approach is based on energy-minimization and starts with a coarse initial depth map, computed under the Lambertian assumption, to provide an initial clue to roughly initialize all variables. Then we iteratively optimize for each of the variables in turn, imposing the inverse rendering dataterm, enforcing depth to be consistent with normals, and enforcing normals to be consistent with the epipolar plane image structure. We tested our approach on synthetic as well as real-world data, and achieve encouraging results despite having a substantially ill-posed problem with a non-convex energy.

*Acknowledgement* This work was supported by the ERC Starting Grant "Light Field Imaging and Analysis" (LIA 336978, FP7-2014).

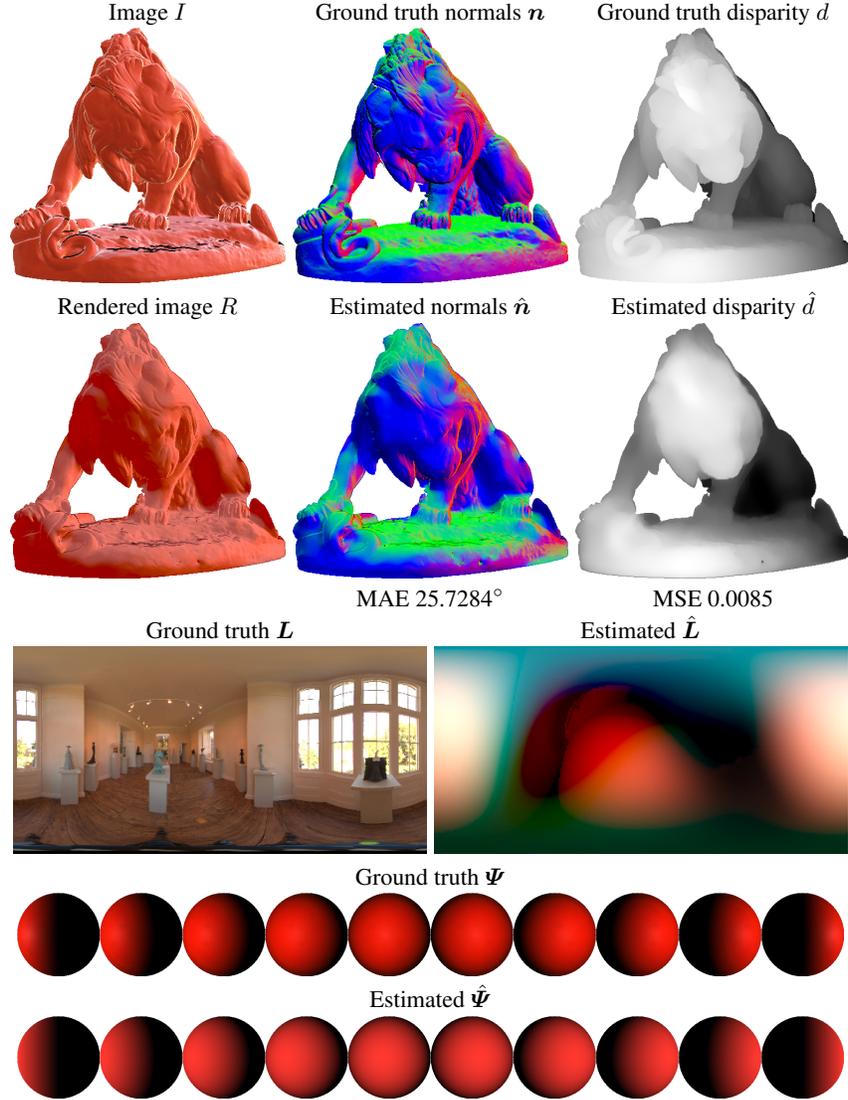
## References

1. Aldrian, O., Smith, W.A.: Inverse rendering of faces on a cloudy day. In: European Conference on Computer Vision. pp. 201–214. Springer (2012) [1](#)
2. Aldrian, O., Smith, W.A.: Inverse rendering of faces with a 3d morphable model. IEEE transactions on pattern analysis and machine intelligence 35(5), 1080–1093 (2013) [1](#)
3. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. IEEE transactions on pattern analysis and machine intelligence 37(8), 1670–1687 (2015) [1](#), [2](#)
4. Bolles, R., Baker, H., Marimont, D.: Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision 1(1), 7–55 (1987) [5](#)
5. Branch, M.A., Coleman, T.F., Li, Y.: A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. SIAM Journal on Scientific Computing 21(1), 1–23 (1999) [7](#), [8](#)
6. Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16(5), 1190–1208 (1995) [8](#)
7. Graber, G., Balzer, J., Soatto, S., Pock, T.: Efficient minimal-surface regularization of perspective depth maps in variational stereo. In: Proc. International Conference on Computer Vision and Pattern Recognition (2015) [5](#), [6](#), [8](#)
8. Heber, S., Pock, T.: Shape from light field meets robust PCA. In: Proc. European Conference on Computer Vision (2014) [1](#)
9. Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4d light fields. In: Asian Conference on Computer Vision. pp. 19–34. Springer (2016) [9](#)
10. Huang, R., Smith, W.A.: Shape-from-shading under complex natural illumination. In: Image Processing (ICIP), 2011 18th IEEE International Conference on. pp. 13–16. IEEE (2011) [2](#), [3](#), [4](#)
11. Jeon, H., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y., Kweon, I.: Accurate depth map estimation from a lenslet light field camera. In: Proc. International Conference on Computer Vision and Pattern Recognition (2015) [1](#)
12. Johannsen, O., Honauer, K., Goldluecke, B., Alperovich, A., Battisti, F., Bok, Y., Brizzi, M., Carli, M., Choe, G., Diebold, M., Gutsche, M., Jeon, H.G., Kweon, I.S., Park, J., Schilling, H., Sheng, H., Si, L., Strecke, M., Sulc, A., Tai, Y.W., Wang, Q., Wang, T.C., Wanner, S., Xiong, Z., Yu, J., Zhang, S., Zhu, H.: A taxonomy and evaluation of dense light field depth

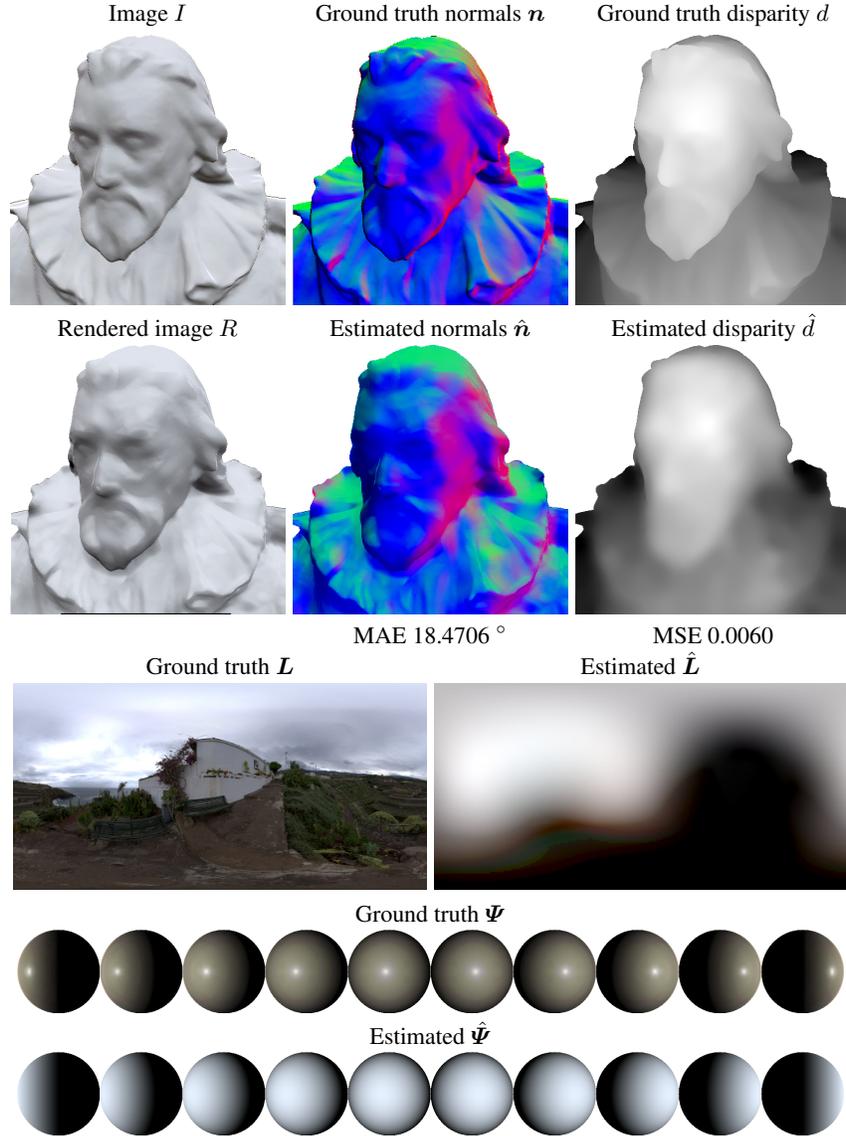
- estimation algorithms. In: 2nd Workshop on Light Fields for Computer Vision at CVPR (2017) [1](#), [2](#)
13. Lombardi, S., Nishino, K.: Reflectance and natural illumination from a single image. In: European Conference on Computer Vision. pp. 582–595. Springer (2012) [1](#), [2](#), [3](#), [4](#), [8](#), [9](#)
  14. Matusik, W., Pfister, H., Brand, M., McMillan, L.: A data-driven reflectance model. ACM Transactions on Graphics 22(3), 759–769 (July 2003) [9](#), [10](#), [13](#), [14](#), [15](#)
  15. Oxholm, G., Nishino, K.: Shape and reflectance from natural illumination. In: European Conference on Computer Vision. pp. 528–541. Springer (2012) [1](#), [2](#), [4](#)
  16. Oxholm, G., Nishino, K.: Multiview shape and reflectance from natural illumination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2155–2162 (2014) [2](#)
  17. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: International Conference on Computer Vision (ICCV 2011) (2011) [8](#), [9](#)
  18. Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for inverse rendering. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. pp. 117–128. ACM (2001) [8](#), [9](#)
  19. Romeiro, F., Zickler, T.: Blind reflectometry. In: European conference on computer vision. pp. 45–58. Springer (2010) [3](#)
  20. Rusinkiewicz, S.M.: A new change of variables for efficient brdf representation. In: Rendering techniques’ 98, pp. 11–22. Springer (1998) [4](#)
  21. Strecke, M., Alperovich, A., Goldluecke, B.: Accurate depth and normal maps from occlusion-aware focal stack symmetry. In: Proc. International Conference on Computer Vision and Pattern Recognition (2017) [2](#), [5](#), [6](#), [7](#), [8](#)
  22. Strecke, M., Alperovich, A., Goldluecke, B.: Accurate depth and normal maps from occlusion-aware focal stack symmetry. In: Proc. International Conference on Computer Vision and Pattern Recognition (2017) [8](#), [9](#)
  23. Tao, M., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from combining defocus and correspondence using light-field cameras. In: Proc. International Conference on Computer Vision (2013) [1](#)
  24. Wang, T., Efros, A., Ramamoorthi, R.: Occlusion-aware depth estimation using light-field cameras. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3487–3495 (2015) [1](#)
  25. Wang, T.C., Chandraker, M., Efros, A.A., Ramamoorthi, R.: Svbrdf-invariant shape and reflectance estimation from light-field cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5451–5459 (2016) [2](#)
  26. Wanner, S., Goldluecke, B.: Variational light field analysis for disparity estimation and super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(3), 606–619 (2014) [1](#), [2](#), [4](#), [5](#), [6](#), [8](#), [9](#)
  27. Wanner, S., Straehle, C., Goldluecke, B.: Globally consistent multi-label assignment on the ray space of 4D light fields. In: Proc. International Conference on Computer Vision and Pattern Recognition (2013) [6](#)
  28. Zeisl, B., Zach, C., Pollefeys, M.: Variational regularization and fusion of surface normal maps. In: Proc. International Conference on 3D Vision (3DV) (2014) [8](#), [9](#)



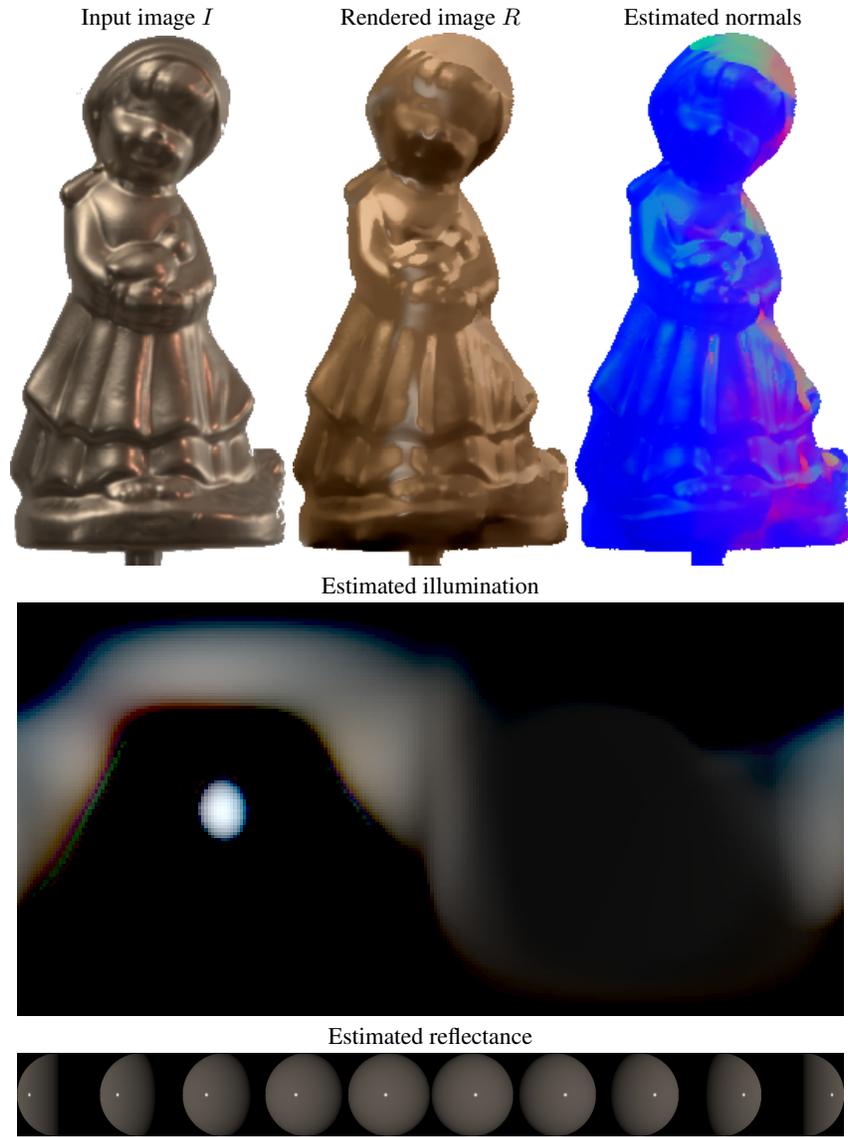
**Fig. 5.** Synthetic dataset ball with nickel reflectance from the MERL BRDF database [14] and *Doge2* illumination. The first row shows the center views of the input image  $I$ , ground truth normals  $\mathbf{n}$  and disparity  $d$ . The second row shows center views of re-rendered image for the estimated parameters  $R$ , estimated normals  $\hat{\mathbf{n}}$  and disparity  $\hat{d}$ . The third row shows ground truth illumination  $\mathbf{L}$  and the estimated illumination  $\hat{\mathbf{L}}$ . and last two rows show comparison of ground truth  $\Psi$  versus estimated reflectance  $\hat{\Psi}$  where the spheres are rendered with a point light which moves from left to right.



**Fig. 6.** Synthetic dataset lion with light-red-paint reflectance from the MERL BRDF database [14] and Envmap illumination. The first row shows the center views of the input image  $I$ , ground truth normals  $\mathbf{n}$  and disparity  $d$ . The second row shows center views of re-rendered image for the estimated parameters  $R$ , estimated normals  $\hat{\mathbf{n}}$  and disparity  $\hat{d}$ . The third row shows ground truth illumination  $L$  and the estimated illumination  $\hat{L}$ , and last two rows show comparison of ground truth  $\Psi$  versus estimated reflectance  $\hat{\Psi}$  where the spheres are rendered with a point light which moves from left to right.



**Fig. 7.** Synthetic dataset *cotton1* with *delrin* reflectance from the MERL BRDF database [14] and *Terrace* illumination. The first row shows the center views of the input image  $I$ , ground truth normals  $\mathbf{n}$  and disparity  $d$ . The second row shows center views of re-rendered image for the estimated parameters  $R$ , estimated normals  $\hat{\mathbf{n}}$  and disparity  $\hat{d}$ . The third row shows ground truth illumination  $\mathbf{L}$  and the estimated illumination  $\hat{\mathbf{L}}$ , and last two rows show comparison of ground truth  $\Psi$  versus estimated reflectance  $\hat{\Psi}$  where the spheres are rendered with a point light which moves from left to right.



**Fig. 8.** Results on a real world dataset recorded with our gantry. The left image shows the input image, the one in the middle is re-rendered image with parameters estimated using our pipeline, and the one on the right shows estimated normals. The second row shows estimated illumination and last row shows renderings of the unit sphere using estimated BRDF parameters with a point light moving from left to right.