

A Variational Model for Intrinsic Light Field Decomposition

Anna Alperovich and Bastian Goldluecke

University of Konstanz

Abstract. We present a novel variational model for intrinsic light field decomposition, which is performed on four-dimensional ray space instead of a traditional 2D image. As most existing intrinsic image algorithms are designed for Lambertian objects, their performance suffers when considering scenes which exhibit glossy surfaces. In contrast, the rich structure of the light field with many densely sampled views allows us to cope with non-Lambertian objects by introducing an additional decomposition term that models specularity. Regularization along the epipolar plane images further encourages albedo and shading consistency across views. In evaluations of our method on real-world data sets captured with a Lytro Illum plenoptic camera, we demonstrate the advantages of our approach with respect to intrinsic image decomposition and specular removal.

1 Introduction

Intrinsic image decomposition aims at separating an illumination invariant reflectance image from an input color image. Such a decomposition has numerous applications in color enhancement, image segmentation, pattern recognition, and object tracking [1–3]. The separation of the shading component is used in BRDF estimation and shadow removal methods [4–6]. However, while intrinsic images have many applications, recovering them remains a substantial challenge for researchers. Estimation of intrinsic components is an ill-conditioned problem: a single image can be decomposed into infinitely many different combinations of reflectance and illumination. Thus, additional constraints or priors are needed to select an appropriate solution. Priors on reflectance (albedo) and shading are usually based on physical principles of light and object interaction, scene geometry, and material properties, as well as on expert knowledge of how intrinsic images should look like. Finally, decomposition into reflectance and illumination components is suitable only for diffuse (Lambertian) objects. According to the dichromatic model introduced by Shafer [7], if glossy (non-Lambertian) objects are present in a scene, a specular term should be taken into account. Many classical approaches fail when the target scene has non-Lambertian objects; as specularity depends on view point, it is hardly possible to estimate it from a single image.

To improve accuracy of intrinsic images, researchers use additional information, for instance, a video sequence instead of a single image, RGB-D imaging

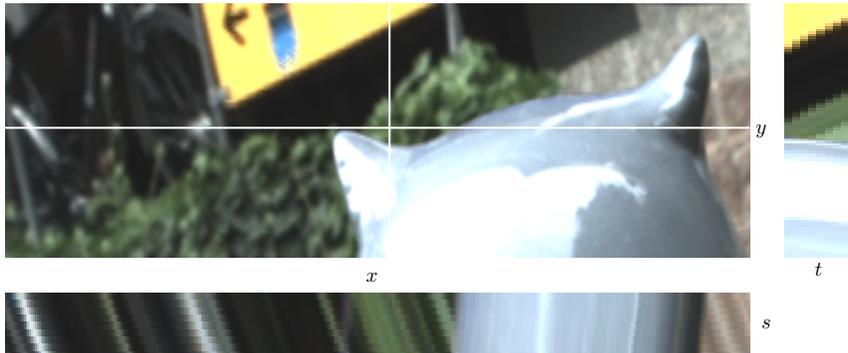


Fig. 1. The top-left image shows the center view of a light field parametrized by image coordinates x and y . On the bottom and right, the epipolar plane images (EPIs) for the white lines in the center view are shown, where s and t describe view point coordinates. As the camera moves, 3D scene points trace straight lines on the EPIs, whose slope corresponds to disparity. Any assignment of a property of a scene point to rays should be constant along these lines, which can be leveraged for consistent regularization [12].

sensors, or manual labeling. This information may be incomplete, suffer from sensor noise, calibration errors, and be dependent on a human factor. Computing this information may be time consuming, require complex experiments, and special equipment. Thus, it is hardly possible to use it in for example industrial applications.

In this work, we leverage light fields for intrinsic image decomposition. 4D light fields are widely used in image analysis and computer graphics. The key idea of light field is to represent a scene not as a traditional 2D image, which contains information about accumulated intensity at each image point, but as a collection of images of the same scene from slightly different view points, see figure 1. The specific structure of the light field allows a wide range of applications. It is used for efficient depth estimation, virtual refocusing, automatic glare reduction as well as object insertion and removal [8–10]. Recently, the inherent structure of the light field was leveraged for shape and BRDF estimation [11, 4].

Contributions. In this paper, we formulate and solve intrinsic light field decomposition by means of an optimization problem for albedo, shading, and specularity. As far as we are aware, this is the first time this problem is addressed for 4D light fields. Based on a detailed review of the state-of-the-art in intrinsic image decomposition, we propose priors for modeling all unknowns based on additional data available in the light field. Epipolar plane image constraints encourage albedo and shading to be constant for projections of the same scene point. By means of a novel term which is specific to light fields, we can also estimate specularity and highlights, and separate them from shading and albedo components. In experiments, we demonstrate that we outperform state-of-the-art intrinsic image decomposition based on RGB plus depth data [13], as well as an alternative approach to detect and remove light field specularity [14, 10].

2 Related work

Intrinsic images have been a challenging research topic for many years. First introduced by Barrow and Tenenbaum [15], they divide an observed image into the product of a reflectance and illumination image. According to Land and McCann [1], large discontinuities in pixel intensities correspond to changes in reflectance, and the remaining variation corresponds to shading. They proposed a Retinex theory that was successfully extended and implemented for intrinsic image decomposition by Tappen et al. [16], Chung et al. [17], Grosse et al. [18], Finalyson et al. [5, 6], and many others.

Besides the Retinex approach, it is common to include additional regularization terms that describe certain physical properties of intrinsic components. Barron and Malik [19–21] introduce priors on reflectance, shape, and illumination to recover intrinsic images. Shen et al. [22] employ texture information. Finalyson et al. [23] search for an invariant image which is independent of lighting and shading. Gehler and Rother [24] model reflectance values drawn from a sparse set of basis colors. Bell et al. [25] also assume that reflectance values come from a predefined set which is unique for every image, then they iteratively adjust reflectance values in this set.

Recently, a significant improvement in intrinsic image decomposition was achieved by using richer types of input data. Having a sequence of images with depth information available allows to penalize albedo and shading consistency between different views, Lee et al. [26]. Depth or disparity information allows to incorporate spatial dependencies between pixels to construct shading prior, Jeon et al. [27]. Chen and Koltun [13] develop a model based on RGB-D information. They separate shading into two components: direct and indirect irradiance that significantly improved decomposition results. Barron and Malik [21] use depth to extend their SIRFS model [20] such that it is applicable for natural scenes.

Although decomposition algorithms nowadays achieve spectacular results for Lambertian scenes, their performance is suffering in the non-Lambertian case in the presence of highlights or specularities. In our paper, we will make use of the rich structure in the light field to estimate specularities for non-Lambertian objects. According to the dichromatic model introduced by Shafer [7], diffuse and specular reflections behave differently. Diffuse objects reflect incident light in multiple directions equally, thus, their color is independent of viewpoint. Specular objects reflect light in a certain direction that depends on orientation, and thus their color depends on viewpoint, light source color, and physical material properties. Blake and Bühlhoff [28] made an extensive analysis of specular reflections, and propose a strategy for recovering 3D structure using specularities. Swaminathan et al. [29] study photometric properties of specular pixels, and model their motion depending on the surface geometry. Adato et al. [30] model specular flow with non-linear partial differential equations. Tao et al. [14, 10] introduced depth estimation for glossy surfaces. They leverage the light field structure to cluster pixels in specular and specular-free groups, then they remove specular components from the input light field.

3 Intrinsic light field model

Light Field structure We briefly describe the light field structure and review notation. For more detailed information, we refer to [31, 12]. A light field is defined on 4D ray space $\mathcal{R} = \Pi \times \Omega$, which parametrizes rays $\mathbf{r} = (x, y, s, t)$ by their intersection coordinates with two planes Π and Ω . Intersection with the focal plane Π gives view point coordinates (s, t) , while the image plane Ω denotes image coordinates (x, y) , see figure 1. A 4D light field is now a map $L : \mathcal{R} \rightarrow \mathbb{R}^n$ on ray space. It can be scalar or vector-valued for grey scale or color images, respectively.

Light field decomposition. We model an intrinsic light field as a function

$$L(\mathbf{r}) = A(\mathbf{r})S(\mathbf{r}) + H(\mathbf{r}), \quad (1)$$

where the radiance L of every ray \mathbf{r} is decomposed into albedo A , shading S , and specular component H . The functions $L, A, S, H : \mathcal{R} \rightarrow \mathbb{R}^3$ map ray space to RGB values. Albedo represents the color of an object independent of illumination and camera position. Shading describes intensity changes due to illumination, inter-reflections, and object geometry. Finally, specularity represents highlights that occur in case of non-Lambertian objects. They depend on illumination, object geometry, and camera position.

The common assumption in the literature related to intrinsic image decomposition is to model the shading component as mono-chromatic [18, 24, 16]. However, in case of multiple light sources or non-Planckian light, this modeling assumption is not sufficient. Thus, we further decompose shading into mono-chromatic shading s and trichromatic light source color C ,

$$S(\mathbf{r}) = s(\mathbf{r})C(\mathbf{r}). \quad (2)$$

We directly compute the illumination component C in a pre-processing step with the illuminant estimation algorithm developed by Yang et al. [32] applied to the center view, assuming that it will be similar across views. After illumination color is computed, we exclude it from the original light field by switching to the new decomposition model

$$\frac{L(\mathbf{r})}{C(\mathbf{r})} = A(\mathbf{r})s(\mathbf{r}) + \frac{H(\mathbf{r})}{C(\mathbf{r})} \quad (3)$$

which is illumination color free. Vector division is to be understood component-wise. As a further simplification, we obtain System (3) in linear form

$$L^{log}(\mathbf{r}) = A^{log}(\mathbf{r}) + \mathbf{1}s^{log}(\mathbf{r}) + H^{log}(\mathbf{r}, A, s, H) \quad (4)$$

by applying the logarithm. We now want to solve (4) with respect to albedo, shading, and specularity.

System (4) is ill-posed, since its number of variables is three times larger than the number of equations. To select a solution that agrees with physical

meaning of intrinsic components, we pose it as an inverse problem and introduce a number of constraints or regularization terms for albedo, shading, and specularity. As usual, dependence of H^{log} on all arguments except \mathbf{r} is ignored during optimization, and it is estimated as another independent component. We thus solve a global energy minimization problem where we weight the residual of (4) with different priors and regularization terms,

$$\begin{aligned} \arg \min_{(A^{log}, s^{log}, H^{log})} & \left\{ \|L^{log}(\mathbf{r}) - A^{log}(\mathbf{r}) - \mathbf{1}s^{log}(\mathbf{r}) - H^{log}(\mathbf{r})\|_2^2 + \dots \right. \\ & \left. \dots + P_{\text{albedo}}(A^{log}) + P_{\text{shading}}(s^{log}) + P_{\text{spec}}(H^{log}) + J(A^{log}, s^{log}) \right\}. \end{aligned} \quad (5)$$

The priors P_{albedo} and P_{shading} for albedo and shading essentially apply the key ideas in intrinsic image decomposition to every subaperture image. They are defined in section 4. The specularity prior P_{spec} is specific to light fields, and a main contribution of our work. It is described in detail in section 5. Finally, the smoothing prior J across ray space encourages spatial smoothness and in particular consistency across different subaperture images. It relies on disparity, and is described together with the optimization framework in section 6.

4 Albedo and shading priors

We start with describing the priors, which are the key for obtaining an accurate solution for intrinsic light field decomposition from the variational model (5). In this section, we introduce the priors P_{albedo} and P_{shading} for albedo and shading, respectively.

Albedo. To model albedo, we combine ideas of Retinex theory, which is widely used to decompose an image into shading and reflectance components [16, 33, 17], with the idea that pixels with equal chromaticity are likely to have similar albedo [34, 26, 13]. Thus, the prior for albedo is the sum of two energies, $P_{\text{albedo}}(A^{log}) = E_{\text{retinex}}(A^{log}) + E_{\text{chroma}}(A^{log})$, corresponding to these two models.

Under the simplifying assumption that image derivatives in the log-domain are caused either by shading or reflectance, we classify the derivative at every ray as caused by shading or albedo. The idea is to compute a modified gradient field \hat{g} which assigns a zero value to all derivatives that are caused by shading. The derivative classification is done with approach similar to Color Retinex used in [18, 17]. A partial spatial derivative L_x of the light field is classified as albedo if neighbouring RGB vectors into the direction of differentiation are not parallel, or if it is above a certain magnitude. Thus, the modified derivative is

$$\hat{g}_x = \begin{cases} L_x & \text{if } \mathbf{c}_{x+1,y} \cdot \mathbf{c}_{x,y} < \tau_{col} \text{ or } |L_x| > \tau_{grad}, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Above, $\mathbf{c} = (r, g, b)^T$, the constant $\tau_{col} > 0$ is a threshold above which two vectors are assumed to be parallel, and $\tau_{grad} > 0$ is another user-defined constant. In a

similar way, we estimate the modified partial derivative \hat{g}_y in the second spatial direction.

The gradient of the albedo should be equal to the gradient field modified by retinex, thus we finally obtain the retinex energy

$$E_{retinex}(A^{log}) = \lambda_{retinex} \int_{\mathcal{R}} \|\partial_x A^{log}(\mathbf{r}) - \hat{g}_x(\mathbf{r})\|^2 + \|\partial_y A^{log}(\mathbf{r}) - \hat{g}_y(\mathbf{r})\|^2 d\mathbf{r}. \quad (7)$$

The second regularization term is based on chromaticity similarities between adjacent rays. The basic idea is that if two neighboring rays of the same view have close chromaticity values, they have the same albedo. We use the chromaticity measure described by Chen and Koltun [13], which gives a weight $\alpha_{\mathbf{r},\mathbf{q}}$ for how likely it is that two rays \mathbf{r} and \mathbf{q} have the same albedo,

$$\alpha_{\mathbf{r},\mathbf{q}} = \left(1 - \frac{\|L^{ch}(\mathbf{r}) - L^{ch}(\mathbf{q})\|}{\max_{\mathbf{r}' \in \Omega, \mathbf{q}' \in N_A(\mathbf{r}')} \|L^{ch}(\mathbf{r}') - L^{ch}(\mathbf{q}')\|}\right) \sqrt{L^{lum}(\mathbf{r})L^{lum}(\mathbf{q})}, \quad (8)$$

where $N_A(\mathbf{r})$ is a neighborhood of \mathbf{r} , and L^{ch} and L^{lum} are chromaticity and luminance. The chromaticity energy

$$E_{chroma}(A^{log}) = \lambda_{chroma} \int_{\mathcal{R}} \sum_{N_A(\mathbf{r})} \alpha_{\mathbf{r},\mathbf{q}} \|A^{log}(\mathbf{r}) - A^{log}(\mathbf{q})\|^2 d\mathbf{r} \quad (9)$$

now penalizes dissimilarity of albedos that have chromaticity measure $\alpha_{\mathbf{r},\mathbf{q}}$ close to one. Note that we use a mixed continuous/discrete notation for \mathbf{r} and \mathbf{q} , as our choice of neighbourhood is inherently discrete, while we require a variational rayspace model in the optimization framework, see section 6.

To construct the neighborhoods $N_A(\mathbf{r})$ for every ray $\mathbf{r} \in \mathcal{R}$, we impose the assumption that spatially close points in \mathbb{R}^3 probably have similar albedo. We select k_A nearest neighbors in \mathbb{R}^3 for the point P on the scene surface intersected by \mathbf{r} , and choose m_A out of k_A neighbors randomly. We believe that this connectivity strategy has several advantages over fully random connectivity: by defining neighbors we increase the chance to meet points with similar chromaticity, and by random connectivity within neighboring points we avoid disconnected chromaticity clusters.

Shading. The shading prior is also the sum of two components, $P_{\text{shading}} = E_{\text{normal}} + E_{\text{spatial}}$. To model the first component, we adopt the well-known assumption [35, 26, 13] that scene points which are spatially close to each other and share the same orientation are likely to have similar shading. To facilitate this, we construct the six-dimensional set

$$\Gamma := \{(P(\mathbf{r}), \mathbf{n}(P(\mathbf{r}))) : \mathbf{r} \in \mathcal{R}\},$$

where $P(\mathbf{r})$ is again the point of the scene surface intersected by \mathbf{r} , and $\mathbf{n}(P(\mathbf{r}))$ the corresponding outer normal. The set of neighbours $N_S(\mathbf{r})$ now consists of



Fig. 2. The left image shows the center view of a light field captured with a Lytro Illum camera. The right image shows the specular mask obtained by our method.

the k_N -nearest neighbours of $\mathbf{r} \in \mathcal{R}$ in the six-dimensional space Γ . The regularization term

$$E_{normal}(s^{log}) = \lambda_{normal} \int_{\mathcal{R}} \sum_{\mathbf{q} \in N_S(\mathbf{r})} (s^{log}(\mathbf{r}) - s^{log}(\mathbf{q}))^2 d\mathbf{r} \quad (10)$$

thus penalizes shading components to be the same if corresponding 3D points are spatially close to each other and their outer normals have similar orientations.

To account for indirect shading, which is caused by inter-reflections between objects in a scene, we also include a purely spatial regularization term

$$E_{spatial}(s^{log}) = \lambda_{space} \int_{\mathcal{R}} \sum_{\mathbf{q} \in N_D(\mathbf{r})} (s^{log}(\mathbf{r}) - s^{log}(\mathbf{q}))^2 d\mathbf{r}, \quad (11)$$

where the neighborhood $N_D(\mathbf{r})$ denotes the k_D nearest neighbors of the 3D scene point first intersected by \mathbf{r} .

5 Prior for the specular component

In this section, we describe the specular prior in the variational energy (5). We first discuss the modeling assumptions, then show how to compute a mask for candidate specular pixels based on these assumptions, and finally construct the prior P_{spec} .

Modeling assumptions. We combine several approaches to model specularity [2, 14, 10, 28–30, 36]. According to the specular motion model [28, 29], specularity changes depend on surface geometry. For instance, regions of low curvature on a specular object create color intensity changes within different views. Specular regions of high curvature result in high pixel intensities in all subaperture views. Thus, curvature information can be useful while estimating specularity. In practice, however, it turns out that curvature estimation is very sensitive to inaccuracies of the 3D model of a scene. Imperfect disparity maps lead to a certain amount of noise in the estimated spatial coordinates, thus curvature information

becomes highly unreliable. Instead of using curvature information directly, we therefore propose a heuristic approach that estimates candidate regions where specularity or highlights can occur. Our main modeling assumptions are thus:

- S1. Specularity is view dependent.
- S2. If a projected 3D point has high pixel intensities and its color is constant across all subaperture views, then the point may be part of a specular surface.
- S3. If a projected 3D point has high variation in pixel intensities, and the color of the corresponding rays changes across subaperture views, then the point may belong to a specular surface.
- S4. If a point is classified as specular, then it is a part of specular surface, and its local neighborhood in \mathbb{R}^3 may result in specular pixels from a certain viewing angle.
- S5. The distribution of specularity is sparse.

Potential specular objects are identified based on magnitude and variation of pixel values over different views. We compute a specular mask for the center view, and propagate it to the remaining views according to disparity.

Computing the specular mask. Our proposed algorithm proceeds in 4 steps:

1. Let Ω_c be the image plane for the center view, and $V = \{(s_1, t_1), \dots, (s_N, t_N)\}$ the set of remaining N view points. For every $\mathbf{p} \in \Omega_c$, we compute the vector $\omega_{\mathbf{p}}$ of color intensity changes with respect to V according to

$$w_p^i = L_i(\mathbf{p} + \mathbf{v}_i d(\mathbf{p})), i = 1, \dots, N. \quad (12)$$

where $\mathbf{v}_i = (s_c - s_i, t_c - t_i)$ is the view point displacement and $d(\mathbf{p})$ the estimated scalar disparity of \mathbf{p} .

2. Identify pixels where color and intensity vary within subaperture views in three steps according to assumptions (S1) and (S3):
 - Filter out a percentage $\%n_{var}$ of pixels that have low luminance variation $\sigma(\omega_{\mathbf{p}})$, where by Ω_c^* we define a set of remaining pixels.
 - Exclude occlusion boundaries from Ω_c^* . To find occlusion boundaries, we compute the k -nearest neighbors in the image domain, and corresponding spatial coordinates in \mathbb{R}^3 . If neighboring pixels in Ω_c^* are far away in \mathbb{R}^3 , with distances larger than d_{occ} , then we classify those pixels as occlusion boundaries.
 - From the remaining pixels, finally exclude the percentage $\%n_{conf}$ with the lowest confidence scores similar to the approach proposed by Tao et al. [14]. To compute confidence, we cluster the corresponding values of $\omega_{\mathbf{p}}$ in two groups using K-means. Let $\mathbf{m}(\mathbf{p})$ be the cluster centroid with the larger mean $\mu(\mathbf{m})$. The confidence is computed as

$$c(\mathbf{p}) = \exp\left(-\frac{1}{\sigma_{spec}^2} \left(\frac{\beta_0}{\mu(\mathbf{m})} + \frac{\beta_1}{\xi(\mathbf{m})}\right)\right), \quad (13)$$

where $\xi(\mathbf{m})$ denotes the sum of all distances within the cluster. The confidence score grows with mean intensity and variation within the brightest

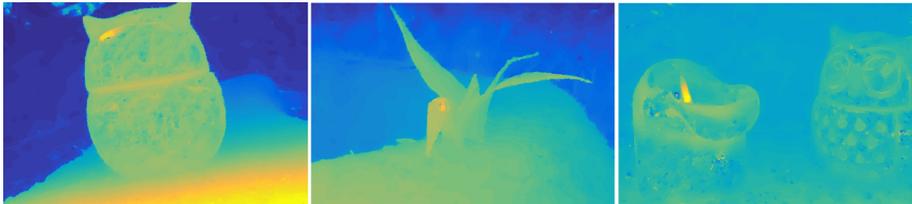


Fig. 3. Estimated disparity maps for scenes captured with a Lytro Illum camera. From left to right: an outdoor scene with the ceramic owl, a tinfoil swan, an indoor scene with the same owl and a candle. Disparities range between -1.5 and 1.5 .

cluster. Thus, we obtain pixels with varying values within subaperture views. Above, β_0 and β_1 are user-defined parameters that control exponential decay of brightness and distance terms, σ_{spec} scales the confidence function. We fix $\beta_0 = 0.5$, $\beta_1 = 10^{-3}$, $\sigma_{spec} = 2$.

3. Identify pixels where intensity is high and color not changing within all subaperture views according to assumption (S2). According to Tian and Clark [36], regions with high unnormalized Wiener entropy, which is defined as the product of RGB values over all pixels, are likely to be specular. We adopt their approach and also identify those regions.
4. Combine pixels found in steps 2 and 3 into the specular mask

$$h_{mask} = \begin{cases} 1, & \text{specular} \\ 0, & \text{non-specular,} \end{cases} \quad (14)$$

which is then grown according to assumption (S4) to include all k_{spec} -nearest neighbors for each specular pixel in the initial mask.

An example specular mask for a Lytro dataset is shown in figure 2.

Final prior on specularity. The specular component should be non-zero only within the candidate specular region given by the mask h_{mask} defined above. We therefore strongly penalize non-zero values outside this region by defining the final sparsity prior as

$$P_{spec}(H^{log}) = \lambda_{spec} \int_{\mathcal{R}} \gamma_w (1 - h_{mask}) \|H^{log}(\mathbf{r})\|^2 d\mathbf{r} + \lambda_{sparse} \|H^{log}\|_1. \quad (15)$$

where $\gamma_w \gg 0$ is a constant. We include an additional sparsity norm on H^{log} to account for assumption (S5).

6 Ray space regularization and optimization

We summarize the previously defined terms in the variational energy (5) as a functional F , so that to obtain the light field decomposition we have to solve

$$\arg \min_{(A^{log}, s^{log}, H^{log})} \left\{ F(A^{log}, s^{log}, H^{log}) + J(A^{log}, s^{log}) \right\}. \quad (16)$$

Table 1. Main parameters for intrinsic image decomposition used in implementation.

Priors weights	Optimization	Specularity
$\lambda_{chroma} = 0.25$	global iterations = 10	$\%n_{var} = 75\%$
$\lambda_{retinex} = 0.5$	local iterations = 10	$d_{occ} = 0.1$
$\lambda_{normal} = 0.5$	$\mu = 0.01$	$\%n_{conf} = 85\%$
$\lambda_{space} = 0.5$	$\lambda = 0.25$	Wiener entropy = 99%
$\lambda_{spec} = 0.5$	Neighborhood	$k_{spec} = 60$
$\lambda_{spec.sparse} = 0.1$	$k_A = 100$	$\gamma_w = 10$
Retinex	$m_A = 10$	
$\tau_{col} = 0.99$	$k_N = 20$	
$\tau_{grad} = 0.2$	$k_D = 20$	

As typical in intrinsic image decomposition, the overall optimization problem is rather complex. However, taking a detailed look at the individual terms, it turns out that we have a convex objective F . Furthermore, our intention is to define the global smoothness term J on ray space in a way that it enforces spatial smoothness within the views, as well as consistency with the disparity-induced structure on the epipolar plane images. Thus, the complete objective function exactly fits the light field optimization framework for inverse problems on ray space proposed by Goldluecke and Wanner [12]. The key advantage of this framework is that it is computationally efficient since it allows to solve subproblems for each epipolar plane image and view independently. Also, it is generic in the sense that we just need to provide a way to compute F and related proximity operators. We thus adopt their method to solve our problem.

In [12], the light field regularizer J in (16) is a sum of several contributions. First, there are individual regularizers J_{xs} and J_{yt} for each epipolar plane image, which depend on the disparity map and employ an anisotropic total variation to enforce consistency of the fields in the arguments with the linear patterns on the epipolar plane images, see figure 1. Second, for each view, there is a regularizer J_{st} , and as in the basic framework in [12], we use a simple total variation term for efficiency. In future work, we intend to move to something more sophisticated here.

Albedo and shading are independent of view point, thus their values should not vary between views. We want A^{log} and s^{log} to be constant in the direction of \mathbf{d} , except at disparity discontinuities. We also regularize both components within each individual view as noted above. The complete regularizer can thus be written as

$$J(A^{log}, s^{log}) = \mu J_{xs}(A^{log}, s^{log}) + \mu J_{yt}(A^{log}, s^{log}) + \lambda J_{st}(A^{log}, s^{log}), \quad (17)$$

where $\lambda, \mu > 0$ are user-defined constants which correspond to the amount of smoothing on the separate views and EPIs, respectively. The objective is convex, so that we achieve global optimality. For details and the actual optimization algorithm we refer to [12].

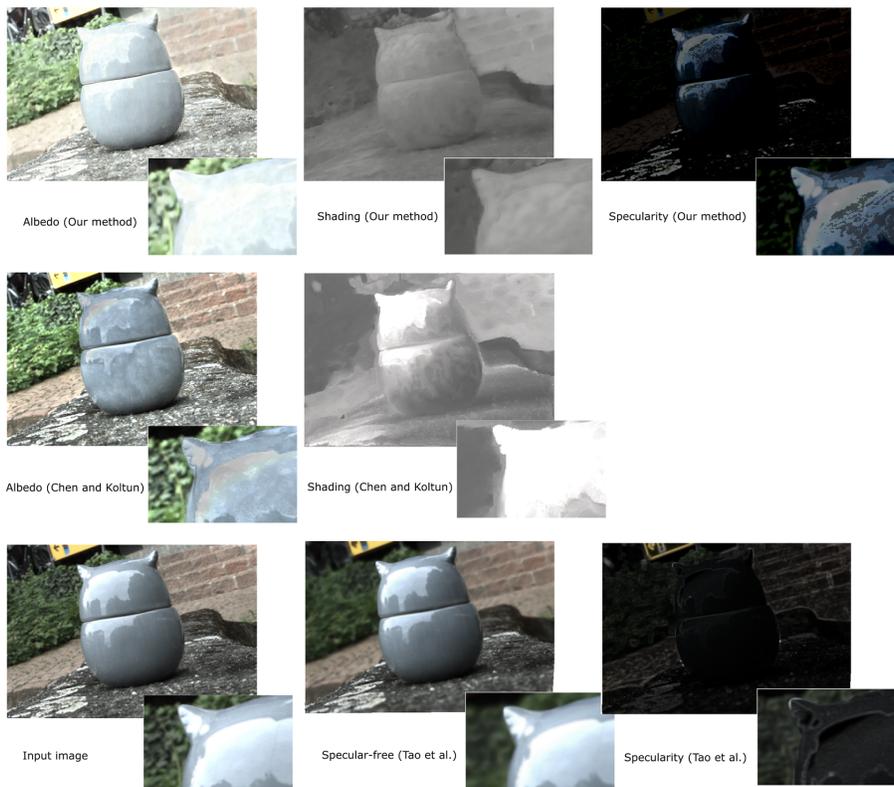


Fig. 4. Center view images showing the light field decomposition. The first row depicts a decomposition with our approach into: albedo, shading, and specularity. The second row illustrates Chen and Koltun’s algorithm [13], where the center view image is decomposed into albedo and shading with the additional input of our generated depth map. The third row illustrates original image, diffuse, and specular images obtained by Tao et al.’s method [10]. Due to EPI constraints and the specularity term, our shading component does not include specular highlights as Chen and Koltun’s result. Also, we removed most cast shadows from albedo image, while smoothness priors prevent albedo and shading discontinuities. Tao et al. detects less of the specular regions compared to ours. Their algorithm identifies mostly boundaries of specular regions, and removes those boundaries from the diffuse image. Disparity errors create variation of intensity values in subaperture views, which erroneously are classified as specularity on occlusion boundaries. Our algorithm detects the complete specular regions, since it is more robust to inaccurate disparity estimation.

7 Results

We validate our decomposition method on light fields captured with a Lytro Illum plenoptic camera, as well as on synthetic and gantry data sets provided by Wanner et al. [37]. In the paper, we present selected results for real world indoor and outdoor scenes, the rest we show in the supplementary material. While

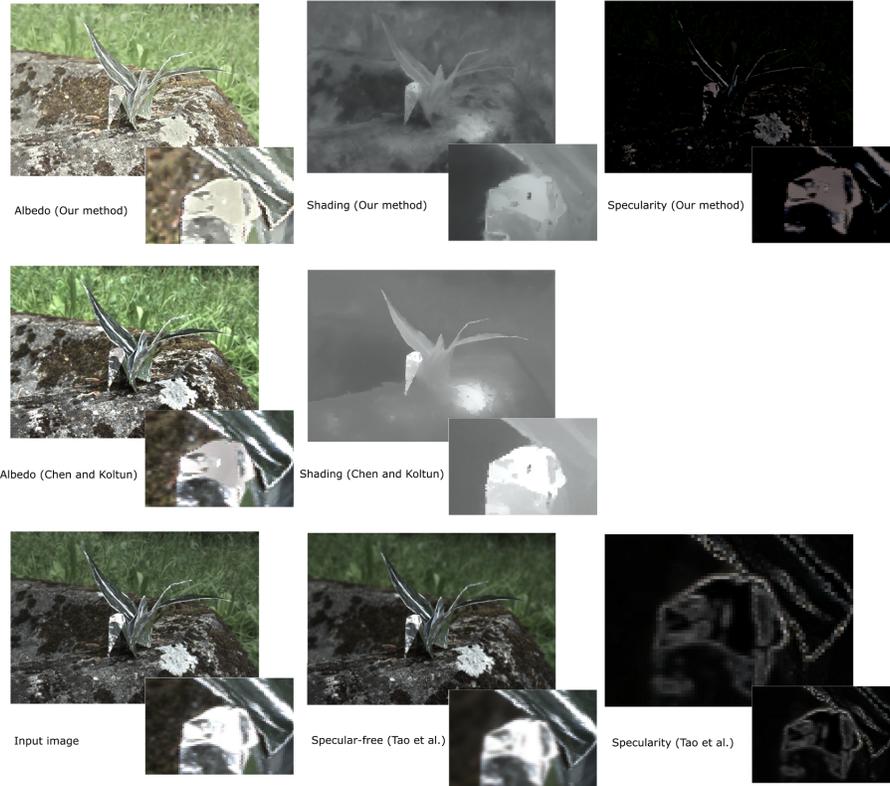


Fig. 5. Center view of the outdoor scene with an origami swan made from aluminum foil. Comparing shading and albedo images, we conclude that our algorithm detects more cast shadows than Chen and Koltun’s algorithm. Our specular component contains more correctly classified glossy regions than the one produced by algorithm of Tao et al. We observe that their approach predominantly detects boundaries of specular regions, thus only these are removed in the generated specular free image.

benchmark datasets for evaluating intrinsic image decomposition are presented in [18, 38, 25], those data sets are designed for algorithm evaluation on single RGB, RGB+D images, or on optic flow; they are not applicable to our light field based method. Since there are thus no ground truth intrinsic light fields available so far, we evaluate our method visually and with qualitative comparisons, deferring rendering of a novel benchmark to future work.

To recover a 3D model and estimate normals, we perform disparity estimation with the multi-view stereo method described in [9], with an improved more occlusion-aware data term and refined with further smoothing using a generalized total variation regularizer, see estimated disparity labels in figure 3. The main algorithm parameters and their values are presented in Table 1. Our method is implemented in Matlab, version R2015b, with run-times on a PC with Intel(R) Core i7-4790 CPU 3,60GHz and an NVIDIA GeForce GTX 980.

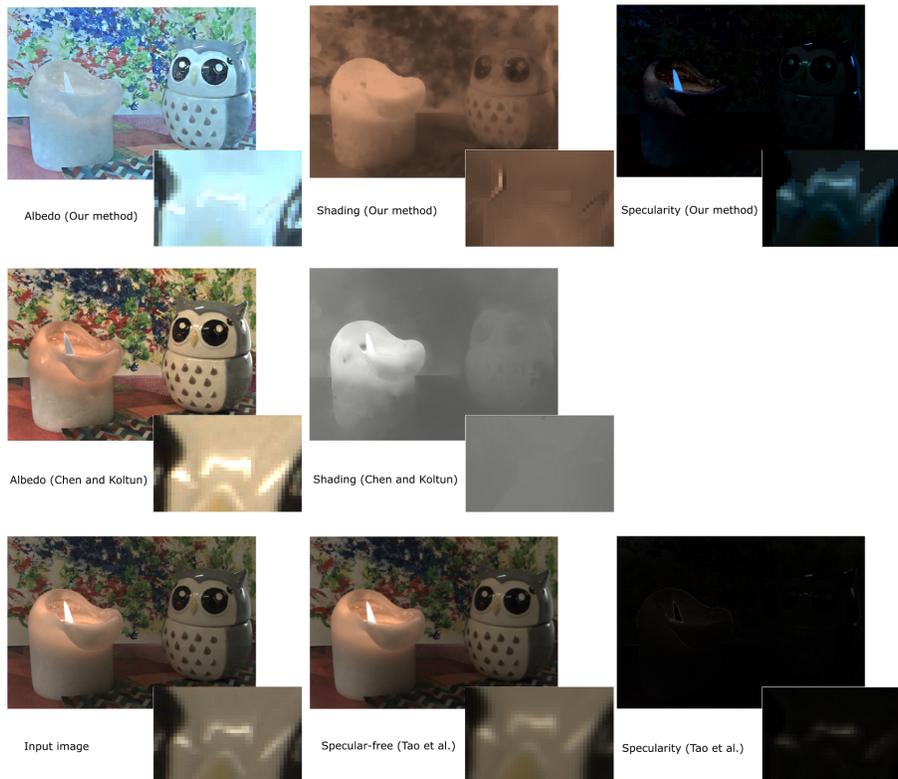


Fig. 6. Center view of an indoor scene with several light sources and non trivial chromaticity. We observe a difference in lighting of albedo images. In our approach, the albedo image is illumination free, compared to albedo images produced with Chen and Koltun’s algorithm. A reason can be that we first compute illumination color and exclude it from the optimization model, while in Chen and Koltun’s approach illumination color is included in the optimization. Both shading components are specular-free, while the albedo component of Chen and Koltun algorithm contains specularity. There is a near to zero specular component detected with Tao et al.’s algorithm. This can be explained by the bad initial disparity estimation which causes erroneous pixel classification. Our approach outperforms Tao et al., since our specular detection algorithm is occlusion aware and also analyses regions where pixels have high intensities.

Evaluation results. Since there are no intrinsic image decomposition algorithms that consider specularity, to compare results of our specular term we select a recent algorithm for depth estimation and specular removal developed for light field cameras by Tao et al. [10]. To compare albedo and shading terms, we investigated recently published algorithms that employ 3D information. There are several papers where depth information is used for intrinsic image decomposition [26, 21, 13, 27]. We selected the algorithm developed by Chen and Koltun [13] to compare against, since it outperforms other algorithms that use 3D information. For both comparisons, we use the authors implementations with default parame-

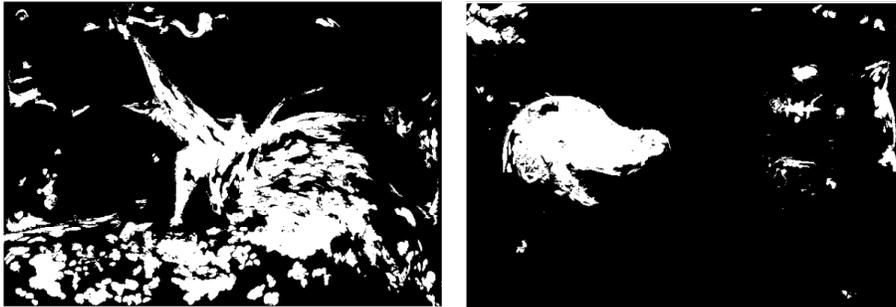


Fig. 7. Potential locations for specularity that were detected with our algorithm. The left image shows candidate specular regions for the origami swan, the right image depicts the specular mask for the owl and candle scene.

ter setting. Figures 4, 5, 6 illustrate original image, our proposed decomposition method, Chen and Koltun [13], and Tao et al. [10]. For all images, contrast was enhanced using the Matlab function `imadjust` for better visualization. Figure 7 illustrates specular masks for origami swan and owl with candle light fields.

We also compared runtime of the algorithms. The Chen and Koltun algorithm converges in 20 - 30 minutes for a single image, the method by Tao et al. (including depth estimation) takes 60 minutes. Our approach evaluated on a light field with a cross-hair shaped subset of 17 views from a light field with 9×9 views in total converges in 30-40 minutes, which amounts to 1.7 – 2.4 minutes per frame.

8 Conclusions

In this work, we propose the first approach towards solving the intrinsic 4D light field decomposition problem while leveraging the disparity-induced structure on the epipolar plane images. In contrast to existing intrinsic image algorithms, the dense collection of views in a light field allows us to define an additional specular term in the decomposition model, so that we can optimize over the specular component as well as albedo and shading by minimizing a single variational functional. As the inverse decomposition problem is embedded in a recent framework for light field labeling [12], we can ensure that albedo and shading estimates are consistent across and use information from all views. Experiments demonstrate that we outperform both a state-of-the-art intrinsic image decomposition method employing additional depth information [13], as well as a light field based method for specular removal [10, 14] on challenging non-Lambertian scenes.

Acknowledgements

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

References

1. Land, E.H., McCann, J.J.: Lightness and retinex theory. *Journal of the Optical Society of America* **61** (1978) 1–11
2. Shroff, N., Taguchi, Y., Tuzel, O., Veeraraghavan, A., Ramalingam, S., Okuda, H.: Finding a needle in a specular haystack. In: *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. (2011) 5963–5970
3. Beigpour, S., van de Weijer, J.: Object recoloring based on intrinsic image estimation. In: *IEEE International Conference on Computer Vision*. (2016)
4. Wang, T.C., Chandraker, M., Efros, A., Ramamoorthi, R.: Svbrdf-invariant shape and reflectance estimation from light-field cameras. In: *Proc. International Conference on Computer Vision and Pattern Recognition*. (2016)
5. Finlayson, G.D., Hordley, S.D., Drew, M.S.: Removing shadows from images using retinex. In: *In Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*. (2002)
6. Finlayson, G.D., Hordley, S.D., Lu, C., Drew, M.S.: On the removal of shadows from images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **28** (2006) 59–68
7. Shafer, S.: Using color to separate reflection components. *Color Research & Application* **10** (1985) 210–218
8. Levoy, M.: Light fields and computational imaging. *Computer* **39** (2006) 46–55
9. Wanner, S., Goldluecke, B.: Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36** (2014) 606–619
10. Tao, M., Su, J.C., Wang, T.C., Malik, J., Ramamoorthi, R.: Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **38** (2015) 1155–1169
11. Tao, M., Srinivasan, P., Hadap, S., Rusinkiewicz, S., Malik, J., Ramamoorthi, R.: Shape estimation from shading, defocus, and correspondence using light-field angular coherence. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2016) 1
12. Goldluecke, B., Wanner, S.: The variational structure of disparity and regularization of 4D light fields. In: *Proc. International Conference on Computer Vision and Pattern Recognition*. (2013)
13. Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: *Proc. International Conference on Computer Vision*. (2013)
14. Tao, M., Wang, T.C., Malik, J., Ramamoorthi, R.: Depth estimation for glossy surfaces with light-field cameras. In: *Proc. European Conference on Computer Vision Workshops*. (2014)
15. Barrow, H.G., Tenenbaum, J.M.: Recovering intrinsic scene characteristics from images. *Computer Vision Systems, Academic Press* **23** (1978) 3–26
16. Tappen, M.F., Freeman, W.T., Adelson, E.H.: Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27** (2005) 1459–1472
17. Chung, Y., Cherng, S., Bailey, R.R., Chen, S.W.: Intrinsic image extraction from a single image. *Journal of Information Science and Engineering* **25** (2009) 1939–1953
18. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground truth dataset and baseline evaluations for intrinsic image algorithm. In: *Proc. International Conference on Computer Vision*. (2009)

19. Barron, J.T., Malik, J.: High-frequency shape and albedo from shading using natural image statistics. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2011)
20. Barron, J.T., Malik, J.: Color constancy, intrinsic images, and shape estimation. In: Proc. European Conference on Computer Vision. (2012)
21. Barron, J.T., Malik, J.: Intrinsic scene properties from a single rgb-d image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38** (2015) 690–703
22. Shen, L., Tan, P., Lin, S.: Intrinsic image decomposition with non-local texture cues. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2008)
23. Finlayson, G.D., Drew, M.S., Lu, C.: Intrinsic images by entropy minimization. In: Proc. European Conference on Computer Vision. (2004)
24. Gehler, P.V., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. In: NIPS. (2011)
25. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. *ACM Trans. on Graphics (SIGGRAPH)* **33** (2014) 159:1–159:12
26. Lee, K.J., Zhao, Q., Tong, X., Gong, M., Izadi, S., Lee, S.U., Tan, P., Lin, S.: Estimation of intrinsic image sequences from image+depth video. In: Proc. European Conference on Computer Vision. (2012)
27. Jeon, J., Cho, S., Tong, X., Lee, S.: Intrinsic image decomposition using structure-texture separation and surface normals. In: Proc. European Conference on Computer Vision. (2014)
28. Blake, A., Bühlhoff, H.: Shape from specularities: computation and psychophysics. *Phil. Trans. R. Soc. Lond. B* **331** (1991) 237–252
29. Swaminathan, R., Kang, S.B., Szeliski, R., Criminisi, A., Nayar, S.K.: On the motion and appearance of specularities in image sequences. In: Proc. European Conference on Computer Vision. Volume I. (2002) 508–523
30. Adato, Y., Vasilyev, Y., Ben-Shahar, O., Zickler, T.: Toward a theory of shape from specular flow. In: Proc. International Conference on Computer Vision. (2007)
31. Bolles, R.C., H.H.B., Marimont, D.H.: Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision* **1** (1987) 7–55
32. Yang, K., Gao, S., Li, Y.: Efficient illuminant estimation for color constancy using grey pixels. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2015)
33. Weiss, Y.: Deriving intrinsic images from image sequences. In: Proc. International Conference on Computer Vision. (2001)
34. Finlayson, G.D., Drew, M.S., Lu, C.: Entropy minimization for shadow removal. In: *IJCV*. (2009)
35. Tao, M., Srinivasan, P., Malik, J., Rusinkiewicz, S., Ramamoorthi, R.: Depth from shading, defocus, and correspondence using light-field angular coherence. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2015)
36. Tian, Q., Clark, J.J.: Real-time specularity detection using unnormalized wiener entropy. In: *Computer and Robot Vision (CRV)*. (2013) 356–363
37. Wanner, S., Meister, S., Goldluecke, B.: Datasets and benchmarks for densely sampled 4D light fields. In: *Vision, Modelling and Visualization (VMV)*. (2013)
38. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: Proc. European Conference on Computer Vision. (2012)