

Shadow and Specularity Priors for Intrinsic Light Field Decomposition

Anna Alperovich, Ole Johannsen, Michael Strecke, and Bastian Goldluecke

University of Konstanz
anna.alperovich@uni-konstanz.de

Abstract. In this work, we focus on the problem of intrinsic scene decomposition in light fields. Our main contribution is a novel prior to cope with cast shadows and inter-reflections. In contrast to other approaches which model inter-reflection based only on geometry, we model indirect shading by combining geometric and color information. We compute a shadow confidence measure for the light field and use it in the regularization constraints. Another contribution is an improved specularity estimation by using color information from sub-aperture views. The new priors are embedded in a recent framework to decompose the input light field into albedo, shading, and specularity. We arrive at a variational model where we regularize albedo and the two shading components on epipolar plane images, encouraging them to be consistent across all sub-aperture views. Our method is evaluated on ground truth synthetic datasets and real world light fields. We outperform both state-of-the-art approaches for RGB+D images and recent methods proposed for light fields.

1 Introduction

Intrinsic image decomposition is one of the fundamental problems in computer vision, and has been studied extensively [23,3]. For Lambertian objects, where an input image is decomposed into albedo and shading components, numerous solutions have been presented in the literature. Depending on the input data, the approaches can be divided into those dealing with a single image [36,18,14], multiple images [39,24], and image + depth methods [9,20]. However, in the real world, there are few scenes with only Lambertian objects. Instead, they have specular surfaces, which makes the decomposition problem harder due to the complicated nature of specular reflection.

For these scenes, the recent work [1] proposes an intrinsic light field model for non-Lambertian scenes, where decomposition is performed with respect to albedo, shading, and specularity. The authors exploit the structure of the light field: while albedo and shading are view-independent and therefore constant along projections of the same scene point in the epipolar plane images (EPIs), specularity is view-dependent and shows a different behaviour.

While the previous model thus correctly identifies specular regions, there is still room for improvements in modeling the albedo and shading components. Shading is modeled by relying only on surface geometry, which might suffer from inaccurate disparity estimation. Furthermore, the previous model does not provide any information about inter-reflections and cast shadows in the scene.

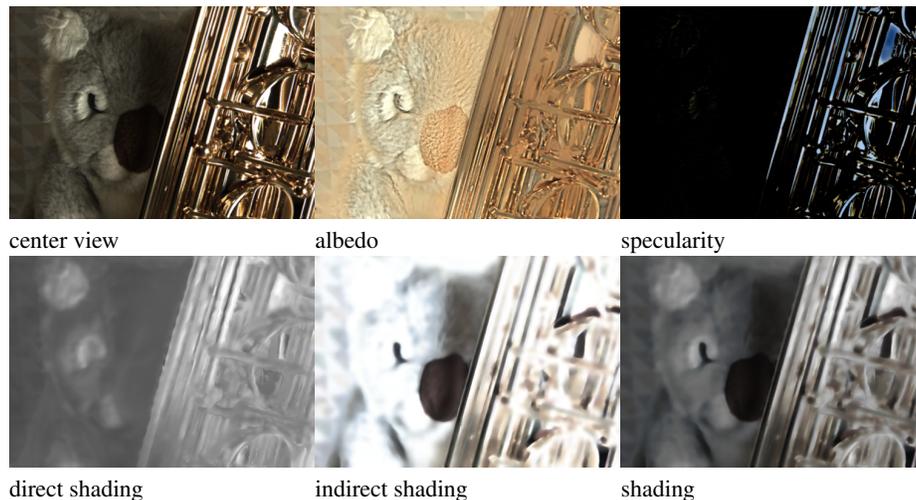


Fig. 1. We present a variational framework to decompose light field radiance into the intrinsic components albedo, shading and specularity. By further breaking down shading into direct and indirect contributions, we achieve superior performance compared to previous work. This light field is captured with Lytro Illum plenoptic camera, 9×9 views, resolution 434×625 .

Contributions. In this work, we propose a model for intrinsic light field decomposition which combines geometrical and color information to define a novel shading prior. We design a shadow detection model with light field data, which to our knowledge is the first time this problem has been addressed for light fields. We apply the estimated shadow score to model cast shadows and inter-reflections explicitly, which results in more consistent shading compared to previous approaches, with better identification of soft and hard shadows. Furthermore, we use pre-estimated specularity positions to make both albedo and shading priors less affected by highlights. For this, we compute a specular free representation of the input light field with [33], and use this information to model albedo and shading. A thorough evaluation of our algorithm on synthetic datasets with ground truth as well as the real world examples demonstrates that we outperform previous approaches based on RGB+D images or light fields.

2 Related Work

Intrinsic Images. Intrinsic image decomposition was first introduced by Barrow and Tenenbaum [3], who propose to decompose an input image into two components, albedo and shading. Since then, plenty of algorithms were proposed for intrinsic image decomposition. One of the first was Retinex, introduced by Land and McCann [23], where they assume that large discontinuities in image derivatives correspond to albedo or reflectance, and the remainder accounts for shading. Over the years, the basic Retinex algorithm was improved by many authors [18,15,36]. Besides the Retinex approach, there are many other interesting methods for single image decomposition [14,21,30].

$N_{3D}^{loc}(\mathbf{r})$	k -nearest neighbors of \mathbf{r} in 3D space.
$N_{3D}^{glob}(\mathbf{r})$	random subset of k among the M -nearest neighbors of \mathbf{r} in 3D space.
$N_{6D}^{loc}(\mathbf{r})$	k -nearest neighbors of \mathbf{r} in 6D space (spatial coordinate + normal).
$N_{6D}^{glob}(\mathbf{r})$	random subset of k among the M -nearest neighbors of \mathbf{r} in 6D space (spatial coordinate + normal).
$N_{obj}(\mathbf{r})$	subset of k among the M -nearest neighbors of \mathbf{r} in 3D space within the same object as \mathbf{r} .

Fig. 2. Different neighborhoods used to define the smoothness priors. They differ by (i) where the nearest neighbor search is performed, either in 3D spatial domain or 6D domain of spatial coordinates and normal vectors, (ii) whether the neighborhood is local, i.e. the k -nearest neighbors are chosen directly, or global, i.e. a random subset of k elements is chosen from the M -nearest neighbors, with a large number M , and (iii) whether we try to sample rays within the same object. To identify objects, we use a simple low-level approach based on Triantafyllidis et al. [37]. Throughout all experiments, we set $k = 20$ and $M = 900$. For global neighborhood in the energy for the indirect shading (17) $k = 80$.

Intrinsic image decomposition was improved a lot by using additional information, for example image sequences [39,11], video [42,6] and RGB+D sensors. Lee et al. [25] recover intrinsic components from a video sequence, Jeon et al. [20] and Chen and Koltun [9] use RGB+D images. Barron and Malik [2] propose the Scene-SIRFS model that outputs an improved depth map, a set of surface normals, a reflectance image, a shading image, and a spatially varying model of illumination. Garces et al. [13] extend the Retinex theory to light fields, where they perform intrinsic light field decomposition for 4D rays instead of 2D images. Alperovich and Goldluecke [1] propose intrinsic light field decomposition for non-Lambertian scenes, where they introduce an additional term to model specularity.

Shadow Detection and Removal. There are several approaches to tackle this problem. They can be divided into user-assistant methods, shape-based, and color-based techniques. Xiao et al. [40] use RGB+D images to construct cast shadow priors, which utilize geometrical information, particularly spatial locations and corresponding outer normals of scene points. Finlayson et al. [12] use an illumination invariant 1D representation of images to detect and remove shadows with Retinex. Lalonde et al. [22] use edge information to detect shadows on the ground for outdoor scenes. Panagopoulos et al. [28] use the concept of a bright channel cue to obtain shadow confidence from a single image.

Specularity Detection. The model for decomposition into specular and diffuse components or *dichromatic model* was introduced by Shafer [29]. This model was adopted for light fields by Tao et al. [35]. They propose a depth estimation and specular removal method that utilizes angular consistency of the diffuse component. Sulc et al. [32] proposed a specular removal approach for light fields based on sparse coding and specular flow. Wang et al. [38] remove specularity by clustering specular pixels into "unsaturated" and "saturated" categories with further color refinement.

3 Intrinsic Light Field Decomposition

We review the intrinsic light field model proposed by Alperovich and Goldluecke [1], and extend it to a model with an additional prior for cast shadows.

Light Field Structure. We start with briefly reviewing notation and basic definitions of the light field structure. A light field is defined on 4D ray space $\mathcal{R} = \Pi \times \Omega$, where a ray is identified by four coordinates $\mathbf{r} = (s, t, u, v)$, which describe the intersections with two parallel planes. Here (s, t) are view point coordinates on the focal plane Π , and (u, v) are coordinates on the image plane Ω . Epipolar plane images (EPIs) can be obtained by restricting 4D ray space to 2D slices. For more information and a thorough introduction on light field geometry, we refer to [17,26,5].

Input to The Model. Our model is based on reconstructed 3D geometry of the scene, in particular we use disparity, spatial locations of the 3D points and surface normals. Since we work with non-Lambertian scenes, disparity estimation might not be accurate in the specular regions, thus the resulting 3D geometry suffers from noise and artifacts. To obtain consistent 3D geometry, we estimate depth with an occlusion-aware algorithm that exploits focal stack symmetry and offers joint regularization of depth and surface normals [31]. We assume that objects have a piecewise smooth geometry, thus surface normals point in similar directions for spatially close points [43,27].

Light Field Decomposition. According to dichromatic reflection model [29], [10] the total radiance L of reflected light is the sum of two independent parts, the radiance of the reflected light at the surface body (diffuse reflection) and radiance of the reflected light at interface (specular reflection). By further breaking down diffuse component into albedo and shading, our intrinsic light field model assumes that the radiance L of a ray \mathbf{r} can be decomposed as

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

where A is an albedo component that represents the surface color independently of illumination and angular camera position, S is a shading component that represents intensity changes due to surface geometry and illumination conditions, and H is specularly or highlights. All intrinsic components are modeled as tri-chromatic vectors, corresponding to color channels R, G, B .

We consider the case that the scene is illuminated with a single white light source. In case of non-white illumination or multiple light sources the model should be multiplied with the illumination color, which can be estimated with color constancy algorithms [41,34].

Inspired by the approach by Chen and Koltun [9], we extend the model (1) by introducing an additional term for cast shadows and inter-reflections. Thus, the new decomposition model

$$L(\mathbf{r}) = A(\mathbf{r})s_d(\mathbf{r})S_i(\mathbf{r}) + H(\mathbf{r}) \quad (2)$$

has components s_d and S_i that describe direct and indirect shading, see figure 1. Direct shading can be understood as the shading that an object would have if it were alone in the scene, i.e. without shadows or reflected light. The second component S_i models inter-reflections and cast shadows. However, in contrast to Chen and Koltun [9], where S_i is modeled to be smooth in 3D space, we model it using a shadow confidence measure that is proportional to shadow intensities in the scene. This allows us to handle hard shadows, where the spatial smoothness assumption is violated. Also, since even under white light inter-reflections depend on the object colors, we model S_i as tri-chromatic, while s_d is mono-chromatic.

To obtain a linear decomposition equation from (2), we apply the logarithm,

$$L^{log}(\mathbf{r}) = A^{log}(\mathbf{r}) + \mathbf{1}s_d^{log}(\mathbf{r}) + S_i^{log}(\mathbf{r}) + H^{log}(\mathbf{r}, A, s_d, S_i, H) \quad (3)$$

$$\text{with } H^{log} = 1 + \frac{H(\mathbf{r})}{A(\mathbf{r})s_d(\mathbf{r})S_i(\mathbf{r})},$$

and ignore the dependence of the specular component H on albedo and shading by treating H^{log} as another independent variable.

4 Optimization

The decomposition (3) is ambiguous, and we thus need strong prior assumptions on all of the unknown intrinsic components. In addition, we need to leverage the light field structure, and enforce a consistent decomposition across all subaperture views. We thus reformulate the decomposition problem as minimization of the energy

$$\arg \min_{(A^{log}, s_d^{log}, s_c^{log}, H^{log})} \left\{ \|D(\mathbf{r})\|_2^2 + E_a(A^{log}) + E_d(s_d^{log}) + E_i(S_i^{log}) + \dots \right. \quad (4)$$

$$\left. + E_s(H^{log}) + J(A^{log}, s_d^{log}, S_i^{log}) \right\}.$$

The first data term enforces consistency of the decomposition, as the residual $D = L^{log} - A^{log} - \mathbf{1}s_d^{log} - S_i^{log} - H^{log}$ should be small. To reduce complexity of the problem (4), we substitute the specularity variable for the difference $L^{log} - A^{log} - \mathbf{1}s_d^{log} - S_i^{log}$ between the input light field and other intrinsic components. The energies E denote the aforementioned priors, to be explained later, while the term J enforces consistency with the light field structure. The idea is to encourage smoothness of albedo and the two shading components along the projections of scene points in the epipolar plane images, and spatial smoothness on each subaperture view by means of a total generalized variation regularizer [7]. We refer to [16] for details on definition and implementation.

The problem (4) can be rewritten as

$$\arg \min_{(A^{log}, s_d^{log}, s_c^{log}, H^{log})} \left\{ F(A^{log}, s_d^{log}, S_i^{log}, H^{log}) + J(A^{log}, s_d^{log}, S_i^{log}) \right\}. \quad (5)$$

where $F = \|D(\mathbf{r})\|_2^2 + E_a(A^{log}) + E_d(s_d^{log}) + E_i(S_i^{log}) + E_s(H^{log})$ is a convex and differentiable functional on the ray space \mathcal{R} . The regularizer

$$J = \mu J_{xs} + \mu J_{yt} + \lambda J_{st} \quad (6)$$

is convex and closed as the sum of convex and closed functionals, but not differentiable, while λ and μ are constants determining smoothness.

The problem (5) is now solved in three steps:

1. Pre-compute all data necessary for albedo, shading, and specularity terms: spatial neighborhoods, chromaticity, the specular and shadow masks that will be discussed in the following sections.

2. Initialize albedo A^{log} and direct shading s_d^{log} variables using values obtained from Color Retinex [18], initialize indirect shading S_i^{log} and specularity H^{log} with zeros.
3. Iteratively solve the problem within the framework proposed by Goldluecke and Wanner [16] for inverse problems on light fields:
 - Optimize F using gradient descent with respect to all intrinsic variables A^{log} , s_d^{log} , S_i^{log} and H^{log} for every $(s, t) \in \Pi$.
 - Enforce angular consistency of A^{log} , s_d^{log} and S_i^{log} by optimizing J_{yt} and J_{xs} with subgradient descent.
 - Regularize A^{log} with total variation (TV), and s_d^{log} , S_i^{log} with total generalized variation (TGV) for every subaperture view $(s, t) \in \Pi$.

To speed up convergence, we recompute the step size of the gradient descent in each iteration such that the cost is minimized as fast as possible, then we perform implicit subgradient descent for the regularizers with the same step size.

What is left to describe is how we model the prior energies for albedo, shading, and specularity. For this, we first start with specular confidence estimation and computation of the diffuse light field in section 5. This is followed up with estimation of a shadow confidence map in section 6. Finally, in section 7, we can assemble the energy terms for all intrinsic components.

5 Specularity Detection

The aim of this section is to first compute a specular mask h , which takes values in $[0, 1]$ and can be understood as the probability of a ray to be specular. Afterwards, we use this mask to obtain pseudo-diffuse radiance L_d for the complete light field.

Confidence Mask h . Our main modeling assumptions are that specularity either is view-dependent or "non-saturated", and thus has high variation in pixel intensities over different sub-aperture views, or it is "saturated", and thus appears bright through all sub-aperture views. We mask out occlusion boundaries while computing the specular mask, thus we assume that there is no specularity on object edges. Dealing with specularity on occlusion boundaries is left for future work.

Under the assumption of known disparity l , we construct angular patches $S(\mathbf{x}, l(\mathbf{x}))$ for every scene point \mathbf{x} . The angular patch represents the corresponding 3D point's color over all sub-aperture views, see figure 3. Contrary to [1] and [38], we not only divide points within $S(\mathbf{x}, l(\mathbf{x}))$ into specular and specular free clusters, which might be inaccurate due to the small baseline of light field cameras, which results in low separability between clusters. Instead, we apply a Fourier transform to every angular patch $S(\mathbf{x}, l(\mathbf{x}))$, and search for significant low-frequency oscillations in every color channel of the patch. If such oscillations exist, then we conclude that the angular patch represents a specular point. We assume that high frequency oscillations are caused by noise, and are not caused by specularity. See figure 3 for an overview of the proposed technique. Our algorithm proceeds in four steps:

1. Apply Fourier transform to logarithm of $S(\mathbf{x}, l(\mathbf{x}))$ for each color channel independently.

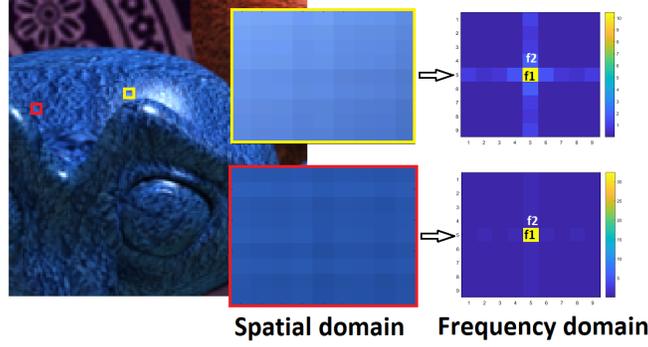


Fig. 3. Specularity detection mechanism. *Left:* part of the center view of the light field with a diffuse point in the red rectangle and a specular point in the yellow rectangle. *Middle:* angular patches for the diffuse and specular points. *Right:* Fourier transform of one color channel with non-zero frequencies for the specular point and almost zero frequencies for the diffuse point.

2. Compute the ratio $r(\mathbf{x}) = \frac{f_2(\mathbf{x})}{f_1(\mathbf{x})}$ between amplitude of the lowest frequency $f_2(\mathbf{x})$ and amplitude of the frequency $f_1(\mathbf{x})$ that corresponds to the mean value of the angular patch.
3. Choose the largest ratio $r(\mathbf{x})$ over the color channels. If $r(\mathbf{x})$ is close to 1, then there is a significant low-frequency component which we interpret as specularity. Usually, specularity exhibits high pixel intensity in at least one sub-aperture view. Thus, we multiply $r(\mathbf{x})$ with the maximum brightness of the patch $S(\mathbf{x}, l(\mathbf{x}))$ to filter out other low frequency color variations.
4. Compute a specular mask $h = \begin{cases} 1, & \text{if } r \geq \tau_f \\ 0, & \text{otherwise} \end{cases}$ by thresholding low values of $r(\mathbf{x})$.

To identify saturated points, we search for the brightest regions in every sub-aperture view that are connected to specular mask h via an eight-neighborhood, and mark them as specular. We define the p -th brightness percentile as the threshold for a point to be bright. Finally, we apply a Gaussian filter with standard deviation σ_h , see figure 4(a) for the result.

Pseudo-diffuse Radiance L_d . To model intrinsic components, we do not need actual diffuse color of the scene point. Instead, we need to know whenever chromaticities of the neighboring points are similar. For this purpose, we compute the pseudo-diffuse color $L_d(r)$ with the algorithm proposed in [33], where the authors suggest to compute a specular-free component that differs from the true diffuse component only in surface color. The authors assume that the scene contains only dielectric materials, and is illuminated under a white light source.

The specular-free representation L_d is computed according to

$$L_d = L - \frac{1}{3} \left(\sum_{i=R,G,B} L_i - \frac{\tilde{L}(3\tilde{c} - 1)}{\tilde{c}(3\tilde{A} - 1)} \right), \quad \tilde{L} = \max_{c=\{R,G,B\}} L_c, \quad (7)$$

with diffuse maximum chromaticity $\tilde{A} = 0.5$ and $\tilde{c} = \max_{c=\{R,G,B\}} \frac{L_c}{(L_R + L_G + L_B)}$. For saturated rays, we instead compute the average color over the k -nearest non-specular



Fig. 4. *Left:* specular confidence mask with brighter regions corresponding to the higher probability of surface to be specular. *Middle:* point-wise shadow confidence β_{pw} . *Right:* resulting final shadow score β after regularization and incorporating the brightness channel cue.

($h = 0$) spatial neighbors $N_{3D}^{loc}(\mathbf{r})$ of L_d . See the supplementary material for an example of the estimated pseudo-diffuse chromaticity.

Note that throughout the paper, we employ different neighbourhoods of rays for different prior terms. These are summarized in figure 2 for convenience.

6 Cast Shadow Detection

In order to identify regions with cast shadows and remove them from the direct shading component, we compute a shadow score $\beta \in [0, 1]$. For this, we first estimate the point-wise confidence β_{pw} of each ray to be shadowed, and then encourage shadowing to be consistent across all sub-aperture views. From these, we will later compute a shadow boundary score to set up the smoothness priors for the intrinsic shading components.

Point-wise Shadow Confidence. Inspired by the work of Xiao et al. [40], we use the spatial locations d , surface normals n , and chromaticity χ of the scene to decide which rays are likely to be shadowed. Since chromaticity images still may contain shadows, we follow previous work [28] and use several illumination-free representations based on RGB , HSV , and $c1c2c3$ color spaces to compute the chromaticity difference. In addition we compute RGB chromaticity from pseudo-diffuse representation L_d that was described in 5. The chromaticity is computed by dividing each channel of L_d by the sum of all channels $R+G+B$. For the different representations, we obtain different ray-wise chromaticity value candidates $\chi(\mathbf{r})$, which all enter the final score below.

Our main assumption is that if two points are spatially close to each other, share the same orientation, and have similar values in chromaticity space, but different intensities, the point with lower intensity is likely to be shadowed. To put this into formulas, we compute a weight $\theta_{\mathbf{r},\mathbf{q}}$ that measures likely similarity in shading between rays \mathbf{r} and \mathbf{q} , in the sense that $\theta_{\mathbf{r},\mathbf{q}}$ is close to one if the two points should have similar shading. We set

$$\theta_{\mathbf{r},\mathbf{q}} = w_{\mathbf{r},\mathbf{q}}^{\chi} w_{\mathbf{r},\mathbf{q}}^n w_{\mathbf{r},\mathbf{q}}^d, \text{ where } w_{\mathbf{r},\mathbf{q}}^{\chi} = \exp\left(-\frac{\min_{\chi} \|\chi(\mathbf{r}) - \chi(\mathbf{q})\|^2}{\sigma_{\chi}^2}\right),$$

$$w_{\mathbf{r},\mathbf{q}}^n = \exp\left(\frac{\cos(n_{\mathbf{r}}, n_{\mathbf{q}}) - 1}{\sigma_n^2}\right) \quad \text{and} \quad w_{\mathbf{r},\mathbf{q}}^d = 1 - \frac{\|d_{\mathbf{r}} - d_{\mathbf{q}}\|^2}{\max_{\tilde{\mathbf{q}} \in N_{obj}(\mathbf{r})} \|d_{\mathbf{r}} - d_{\tilde{\mathbf{q}}}\|^2}. \quad (8)$$

The minimum for w^χ is computed over the different ways to compute chromaticity values introduced above. Based on the similarity scores, we can then estimate color

$$m(\mathbf{r}) = \left(\sum_{\mathbf{q} \in N_{obj}(\mathbf{r})} \theta_{\mathbf{r},\mathbf{q}} L_d(\mathbf{q}) \right) / \left(\sum_{\mathbf{q} \in N_{obj}(\mathbf{r})} \theta_{\mathbf{r},\mathbf{q}} \right) \quad (9)$$

for each ray by computing a weighted average over its neighbors. See figure 2 for the definition of $N_{obj}(\mathbf{r})$. This neighborhood is designed to restrict comparison to suitable rays within the same object.

Finally, the point-wise shadow confidence score β_{pw} is based on the observation that the estimated and actual intensities $v(\cdot)$ should be the same for an unshadowed ray. Here by intensity we mean average over all color channels R , G and B . Thus, we set ray-wise

$$\beta_{pw} = 1 - \exp \left(-\frac{1}{\sigma_v^2} \max(v(m) - v(L_d), 0) \right), \quad (10)$$

with $\sigma_v^2 = 0.25$. See figure 4(b) for an example of a point-wise shadow confidence map.

Consistency Across The Light Field. While the locations of shadows are correct in most of the cases, point-wise estimation suffers from noise, and some shadow-free surfaces are classified as shadowed. To correct the initial shadow estimation, we minimize an energy $E(\beta)$, where we impose smoothness and consistency by embedding it in the optimization framework [16] similar to the main energy minimization problem (4).

In addition to light-field consistency, shadowed regions β are required to coincide with those regions where not only point-wise shadow confidence is high, but also intensity is low. For this, we combine the point-wise confidence with a shadow score derived from the bright channel concept for shadow detection introduced by Panagopoulos et al. [28]. We want our shadow score to be equal to the inverse bright channel cue

$$\bar{br}(\mathbf{r}) = 1 - \max_{c \in \{R,G,B\}} L_c(\mathbf{r}), \quad (11)$$

in regions where point-wise shadow confidence β_{pw} is high. We also include a sparsity prior to remove regions with low confidence altogether.

Thus, the final energy we minimize is

$$E(\beta) = \lambda_d \int_{\mathcal{R}} \beta_{pw}(\mathbf{r}) \|\beta(\mathbf{r}) - \bar{br}(\mathbf{r})\|^2 d\mathbf{r} + \lambda_\beta^s \|\beta(\mathbf{r})\|_1 + J(\beta), \quad (12)$$

defined over all sub-aperture views. In the regularizer J , we weight the spatial regularization with $1 - w_{\mathbf{r},\mathbf{q}}^\chi$, since shadow should be the same if chromaticity is changing, while discontinuities that are related to reflectance changes should be smoothed. Figure 4(c) shows final shadow confidence after optimization.

7 Prior Energies for Intrinsic Components

In this section, we can finally define the prior energies E for albedo, shading and specularity for our model (4). In particular, we show how to employ the previously defined shadow and specularity confidence measures to improve upon the regularization terms.

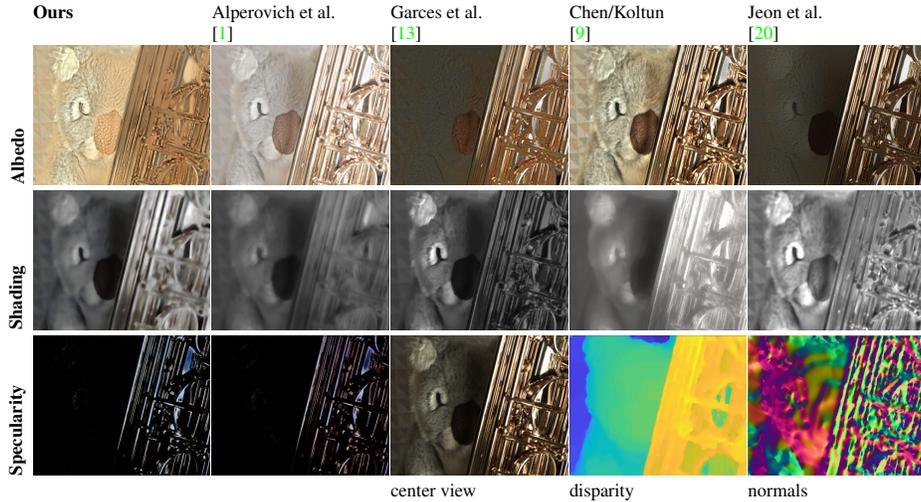


Fig. 5. Qualitative comparisons on the real world light field from figure 1. This scene has highly specular saxophone and almost Lambertian koala toy. Background and koala have similar pallets, which creates difficulties in albedo estimation. Our method successfully removes cast shadows, especially on the wall behind the koala and from saxophone. Our specular component contains all highlights that appear on saxophone.

Albedo. We model albedo energy based on the state-of-the-art approaches also used in [20,9] and many other works. We assume that points with close chromaticity values are likely to have similar albedo. Thus, we define the albedo term as

$$E_a(A^{log}) = \lambda_{ch} \int_{\mathcal{R}} \sum_{\mathbf{q} \in N_{3D}^{glob}(\mathbf{r})} \alpha_{\mathbf{r},\mathbf{q}} \|A^{log}(\mathbf{r}) - A^{log}(\mathbf{q})\|^2 d\mathbf{r}. \quad (13)$$

However, in contrast to previous work, to compute weights $\alpha_{\mathbf{r},\mathbf{q}}$, we use flexible thresholds for chromaticity differences, based on our novel specular confidence h . To remove artifacts caused by specularity from the albedo component, we give more weight to points that have high specular confidence. Thus, we model final regularization weights as

$$\alpha_{\mathbf{r},\mathbf{q}} = (1 + \gamma_a \max(h(\mathbf{r}), h(\mathbf{q}))) w_{\mathbf{r},\mathbf{q}}^x. \quad (14)$$

The weights for the pairs of rays above are defined in (8).

Direct Shading. To model direct shading, we assume that spatially close surface points that share the same orientation should have similar shading. We thus formulate the corresponding direct shading term

$$E_d(s_d^{log}) = \lambda_d^l \int_{\mathcal{R}} \sum_{\mathbf{q} \in N_{6D}^{loc}(\mathbf{r})} \|s_d^{log}(\mathbf{r}) - s_d^{log}(\mathbf{q})\|^2 d\mathbf{r} + \dots \\ + \lambda_d^g \int_{\mathcal{R}} \sum_{\mathbf{q} \in N_{6D}^{glob}(\mathbf{r})} w_{\mathbf{r},\mathbf{q}}^{glob} \|s_d^{log}(\mathbf{r}) - s_d^{log}(\mathbf{q})\|^2 d\mathbf{r} \quad (15)$$

Specularity detection	Shadow detection	$\sigma_n^2 = 0.1$	Weights	$\lambda_d^g = 0.1$	Iterations
$\tau_f = 0.1$	$\lambda_d = 1$	Priors	$\lambda_s = 10$	$\lambda_i^l = 1$	global iter = 50
$\sigma_h = 2$	$\lambda_\beta^s \in [0.1, 0.5]$	$\gamma_\alpha = 10$	$\lambda_{ch} = 0.1$	$\lambda_i^g = 5$	local iter = 10
$p = 85$	$\sigma_{X_i}^2 = 0.01$	$\gamma_d = 10$	$\lambda_d^l = 1$	$\lambda_i^s = 10$	$\lambda_h = 5$

Table 1. Main parameters for intrinsic light field decomposition.

as the combination of local and non-local direct shading priors. See figure 2 for a definition of the neighborhoods, which here live in the six-dimensional space of spatial locations d and normals n corresponding to the individual rays. For the global direct shading consistency, we also weight the similarity between two rays \mathbf{r} and \mathbf{q} with the angular difference of the outer normals and spatial locations,

$$w_{\mathbf{r},\mathbf{q}}^{glob} = (1 + \gamma_d \max(\beta(\mathbf{r}), \beta(\mathbf{q}))) w_{\mathbf{r},\mathbf{q}}^n w_{\mathbf{r},\mathbf{q}}^d. \quad (16)$$

Since direct shading should be free of cast shadows, we add additional weight for neighbors with high cast shadow score β .

Indirect Shading. We model indirect shading by means of the shadow confidence measure β , whose detailed computation is described in Section 6. The main modeling assumptions are:

- A1. Shading is spatially smooth except on shadow boundaries, i.e. near discontinuities of β .
- A2. Assume two points are spatially close to each other, none of them specular, they share the same orientation and their chromaticity is similar. Then $\theta_{\mathbf{r},\mathbf{q}} = 1$ in (8), and their shadow free representation should be the same.
- A3. The distribution of indirect shading is sparse except inside the areas within cast shadows.

These assumptions lead to the indirect shading term

$$E_i(S_i^{log}) = \lambda_i^l \int_{\mathcal{R}} \sum_{\mathbf{q} \in N_{3D}^{log}(\mathbf{r})} \delta_{\mathbf{r},\mathbf{q}} \|S_i^{log}(\mathbf{r}) - S_i^{log}(\mathbf{q})\|^2 d\mathbf{r} + \lambda_i^g \int_{\mathcal{R}} \sum_{\mathbf{q} \in N_{3D}^{glob}(\mathbf{r})} \theta_{\mathbf{r},\mathbf{q}} \beta(\mathbf{r}) \cdot \dots \cdot \|(L^{log}(\mathbf{r}) - S_i^{log}(\mathbf{r})) - (L^{log}(\mathbf{q}) - S_i^{log}(\mathbf{q}))\|^2 d\mathbf{r} + \lambda_i^s \int_{\mathcal{R}} (1 - \beta(\mathbf{r})) \|S_i^{log}(\mathbf{r})\|^2 d\mathbf{r}. \quad (17)$$

We define δ_β to be the norm of the Gaussian filtered gradient of β , and thus the weight $\delta_{\mathbf{r},\mathbf{q}} := (1 - \max(\delta_\beta(\mathbf{r}), \delta_\beta(\mathbf{q})))$ is small if and only if any of the two rays is close to a shadow boundary.

Specularity. The specularity prior we define as

$$E_s(H^{log}) = \lambda_s \int_{\mathcal{R}} (1 - h(\mathbf{r})) \|H^{log}(\mathbf{r})\|^2 d\mathbf{r} + \lambda_h \|H^{log}(\mathbf{r})\|_1. \quad (18)$$

Specularity is encouraged to have non-zero values only inside the specular mask h .

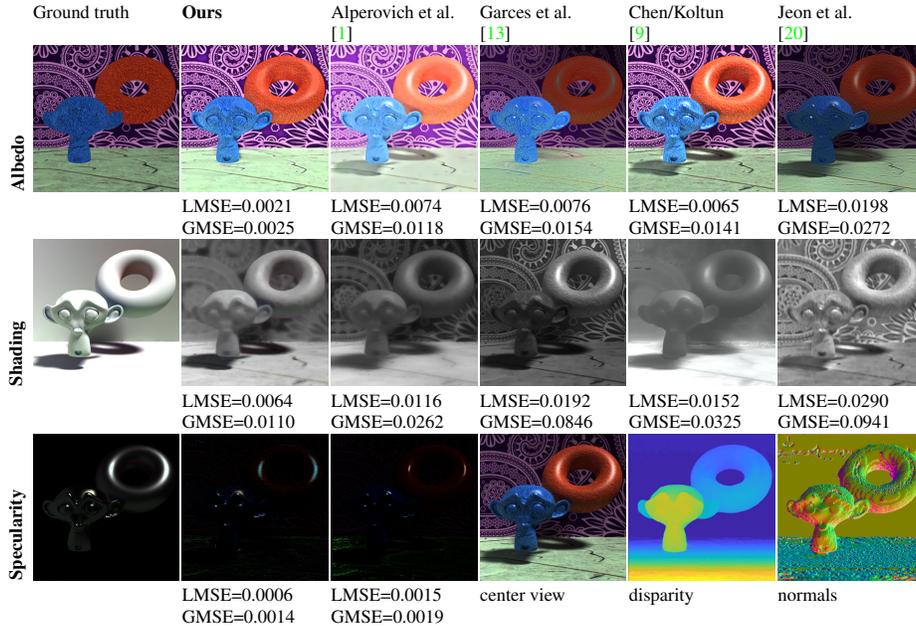


Fig. 6. Quantitative comparison on synthetic data rendered with *Blender*. The scene illustrates how the algorithms can cope with hard shadows, where chromaticity might vary in the shaded area. The disparity range is from -1.5 to 1.5 , light field size is $9 \times 9 \times 512 \times 512$. We find that our proposed method outperforms the other approaches in all measures. In the visual comparison, we see that our approach successfully removed shadow from albedo component.

8 Results

We validate our decomposition method on ground truth synthetic data sets as well as real world light fields. Our main goal is to achieve superior results for the center view of the light field, thus we present decomposition and evaluations only for the center view. Overall, our approach makes use of a light field with 9×9 views. Since estimated disparity is less accurate for the edge views we solve the final optimization (4) for a cross-hair shaped subset of 13 views from a light field with 7×7 views, where edge views are excluded. In the main paper we present a single synthetic data set (figure 6) and two real world light fields (figure 1, figure 5, figure 8). For more results, we refer to the supplementary material.

Note that since there is currently no ground truth available for intrinsic light field decomposition, and benchmark data sets for intrinsic image decomposition presented in [18,4,8] are not applicable to our setting, we created our own ground truth data. Synthetic datasets were generated with Blender using the Cycles rendering engine and a light field plugin [19]. Internally, Blender combines direct shading D , indirect shading I and object color C for diffuse and glossy reflection separately by evaluating $(D_{\text{diffuse}} + I_{\text{diffuse}})C_{\text{diffuse}} + (D_{\text{glossy}} + I_{\text{glossy}})C_{\text{glossy}}$. The separate components can be stored individually to use as ground truth for evaluation. Albedo corresponds to the diffuse color, shading is the sum $D_{\text{diffuse}} + I_{\text{diffuse}}$ of direct and indirect illumination

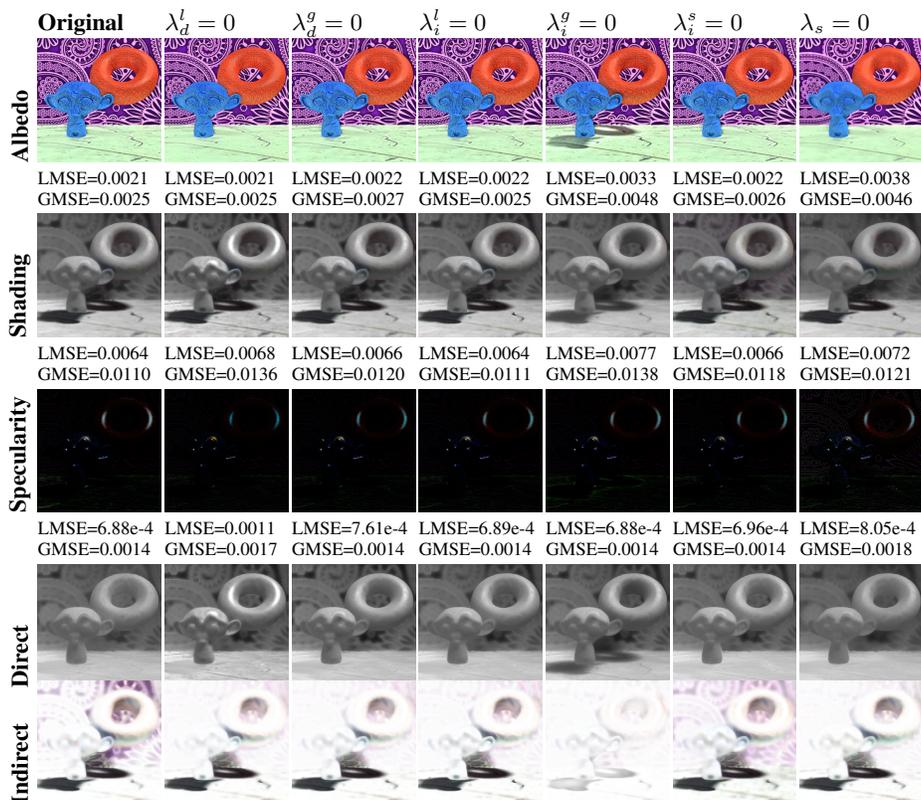


Fig. 7. The leftmost column shows the original result with experimentally determined parameter settings as specified in table 7. The other columns show the consequences of excluding certain priors from the model. We can see that local direct and indirect shading priors λ_d^l and λ_i^l , respectively, influence the smoothness of the shading component, while the global indirect shading prior λ_i^g controls the amount of cast shadows left in the albedo component. The specularity prior λ_s is required to prevent over-smoothing of albedo and shading.

components, and specularity is the difference $H = L - AS$ between input light field and product of albedo and shading.

Our definition of direct and indirect shading differs from one that used in Blender, thus for quantitative evaluations we compute resulting shading $S = s_d S_i$ and then compare to the ground truth shading $S_{ground\ truth} = D_{diffuse} + I_{diffuse}$.

We compare our method to the two RGB+D approaches proposed by Chen and Koltun [9] and Jeon et al. [20], as well as recent methods for light fields proposed by Alperovich and Goldluecke [1] and Garces et al. [13]. For quantitative evaluations, we selected two error measures. The first is local mean-squared error (LMSE) [18] computed patch-wise. The idea is to reduce scaling ambiguity by adjusting the brightness of the image patch such that it corresponds to that of the ground truth, and then to compute MSE. In our experiments, we use rectangular overlapping patches with a size of 20% of the total image size.

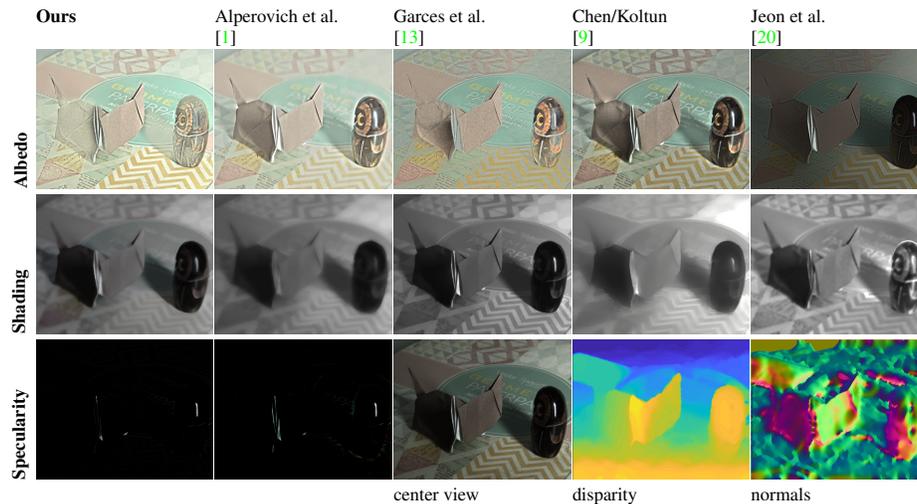


Fig. 8. Qualitative comparisons on the real world data set captured with Lytro Illum plenoptic camera, 9×9 views, resolution 434×625 . Proposed method outperforms the previous approaches, especially albedo component has much less cast shadows, and shading is more smooth and contains less texture.

Although the LMSE is a more reasonable measure than pure MSE, sometimes it can produce non-sensible results in shaded areas. Thus, we also compute global mean-squared error (GMSE) that adjusts the test image brightness value to the ground truth for the whole image. With GMSE, if shaded regions are present in the albedo image, the influence of their error will be more reasonably reflected in the error measure - MSE would otherwise be very small for dark regions. Since aggregated quantitative evaluation can only give a partial picture about the performance, we also visualize all decomposition results for qualitative evaluations. Resulting direct and indirect shading computed with our method for the above datasets can be observed in figures 1 and 7. In order to assess individual contributions of proposed priors we present results where some prior are excluded by setting corresponding weights to zero 7.

9 Conclusion

In this work, we propose a new model for intrinsic non-Lambertian scene decomposition with light field data, where we fully use information available in the light field to model priors for albedo, shading, and specularity. One of our main contributions is a novel shading term that describes cast shadows and inter-reflections. By means of this term, we recover consistent shading components, both in case of hard as well as soft shadows. Moreover, we improve albedo and specularity estimation by embedding specularity information in the albedo prior, which makes albedo less affected by highlights. We introduce ground truth evaluations on synthetic light fields, where we show qualitatively and quantitatively that our method significantly outperforms existing algorithms. In addition, we perform qualitative evaluations for the real world examples captured with Lytro Illum plenoptic camera.

References

1. Alperovich, A., Goldluecke, B.: A variational model for intrinsic light field decomposition. In: Proc. ACCV (2016) [1](#), [3](#), [6](#), [10](#), [12](#), [13](#), [14](#)
2. Barron, J.T., Malik, J.: Intrinsic scene properties from a single rgb-d image. IEEE TPAMI 38(4), 690–703 (2015) [3](#)
3. Barrow, H.G., Tenenbaum, J.M.: Recovering intrinsic scene characteristics from images. Computer Vision Systems, Academic Press 23(1), 3–26 (1978) [1](#), [2](#)
4. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. ACM Trans. on Graphics (SIGGRAPH) 33(4), 159:1–159:12 (2014) [12](#)
5. 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) [4](#)
6. Bonneel, N., Sunkavalli, K., Tompkin, J., Sun, D., Paris, S., Pfister, H.: Interactive intrinsic video editing. ACM Trans. Graph. 33(6) (Nov 2014) [3](#)
7. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences 3(3), 492–526 (2010) [5](#)
8. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: Proc. ECCV (2012) [12](#)
9. Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: Proc. ICCV (2013) [1](#), [3](#), [4](#), [10](#), [12](#), [13](#), [14](#)
10. Finlayson, G., Schaefer, G.: Solving for colour constancy using a constrained dichromatic reflection model. International Journal of Computer Vision 42(3), 127–144 (2001) [4](#)
11. Finlayson, G.D., Drew, M.S., Lu, C.: Intrinsic images by entropy minimization. In: Proc. ECCV (2004) [3](#)
12. 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(1), 59–68 (2006) [3](#)
13. Garces, E., Echevarria, J.I., Zhang, W., Wu, H., Zhou, K., Gutierrez, D.: Intrinsic light field images. Computer Graphics Forum (2017) [3](#), [10](#), [12](#), [13](#), [14](#)
14. Garces, E., Munoz, A., Lopez-Moreno, J., Gutierrez, D.: Intrinsic images by clustering. Computer Graphics Forum (Proc. EGSR 2012) 31(4) (2012) [1](#), [2](#)
15. Gehler, P.V., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. In: NIPS (2011) [2](#)
16. Goldluecke, B., Wanner, S.: The variational structure of disparity and regularization of 4D light fields. In: Proc. CVPR (2013) [5](#), [6](#), [9](#)
17. Gortler, S., Grzeszczuk, R., Szeliski, R., Cohen, M.: The Lumigraph. In: Proc. SIGGRAPH. pp. 43–54 (1996) [4](#)
18. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground truth dataset and baseline evaluations for intrinsic image algorithm. In: Proc. ICCV (2009) [1](#), [2](#), [6](#), [12](#), [13](#)
19. Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4d light fields. In: Proc. ACCV (2016) [12](#)
20. Jeon, J., Cho, S., Tong, X., Lee, S.: Intrinsic image decomposition using structure-texture separation and surface normals. In: Proc. ECCV (2014) [1](#), [3](#), [10](#), [12](#), [13](#), [14](#)
21. Kim, S., Park, K., Sohn, K., Lin, S.: Unified depth prediction and intrinsic image decomposition from a single image via joint convolutional neural fields. In: Proc. ECCV (2016) [2](#)
22. Lalonde, J.F., Efros, A.A., Narasimhan, S.G.: Detecting ground shadows in outdoor consumer photographs. In: Proc. ECCV (2010) [3](#)
23. Land, E.H., McCann, J.J.: Lightness and retinex theory. Journal of the Optical Society of America 61(1), 1–11 (1978) [1](#), [2](#)

24. 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. ECCV (2012) [1](#)
25. Lee, T.S.: Top-down influence in early visual processing: a Bayesian perspective. *Physiology and Behavior* *77*, 645–650 (2002) [3](#)
26. Levoy, M.: Light fields and computational imaging. *Computer* *39*(8), 46–55 (2006) [4](#)
27. Olsson, C., Boykov, Y.: Curvature-based regularization for surface approximation. In: Proc. CVPR (2012) [4](#)
28. Panagopoulos, A., Wang, C., Samaras, D., Paragios, N.: Estimating shadows with the bright channel cue. In: CRICV (2010) [3](#), [8](#), [9](#)
29. Shafer, S.: Using color to separate reflection components. *Color Research & Application* *10*(4), 210–218 (1985) [3](#), [4](#)
30. Shen, L., Tan, P., Lin, S.: Intrinsic image decomposition with non-local texture cues. In: Proc. CVPR (2008) [2](#)
31. Strecke, M., Alperovich, A., Goldluecke, B.: Accurate depth and normal maps from occlusion-aware focal stack symmetry. In: Proc. CVPR (2017) [4](#)
32. Sulc, A., Alperovich, A., Marniok, N., Goldluecke, B.: Reflection separation in light fields based on sparse coding and specular flow. In: Proc. VMV (2016) [3](#)
33. Tan, R.T., Ikeuchi, K.: Separating reflection components of textured surfaces using a single image. In: Proc. ICCV (2003) [2](#), [7](#)
34. Tan, R.T., Nishino, K., Ikeuchi, K.: Color constancy through inverse-intensity chromaticity space. *J. Opt. Soc. Am. A* *21*(3), 321–334 (Mar 2004) [4](#)
35. 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 TPAMI* *38*(6), 1155–1169 (2015) [3](#)
36. Tappen, M.F., Freeman, W.T., Adelson, E.H.: Recovering intrinsic images from a single image. *IEEE TPAMI* *27*(9), 1459–1472 (2005) [1](#), [2](#)
37. Triantafyllidis, G., Dimitriou, M., Kounalakis, T., Vidakis, N.: Detection and classification of multiple objects using an rgb-d sensor and linear spatial pyramid matching. *ELCVIA Electronic Letters on Computer Vision and Image Analysis* *12*(2) (2013) [3](#)
38. Wang, H., Xu, C., Wang, X., Zhang, Y., Peng, B.: Light field imaging based accurate image specular highlight removal. *PLoS ONE* *11*(6) (2016) [3](#), [6](#)
39. Weiss, Y.: Deriving intrinsic images from image sequences. In: Proc. ICCV (2001) [1](#), [3](#)
40. Xiao, Y., Tsougenis, E., Tang, C.K.: Shadow removal from single RGB-D images. In: Proc. CVPR (2014) [3](#), [8](#)
41. Yang, K., Gao, S., Li, Y.: Efficient illuminant estimation for color constancy using grey pixels. In: Proc. CVPR (2015) [4](#)
42. Ye, G., Garces, E., Liu, Y., Dai, Q., Gutierrez, D.: Intrinsic video and applications. *ACM Transactions on Graphics (SIGGRAPH)* *33*(4) (2014) [3](#)
43. Zeisl, B., Zach, C., M.Pollefeys: Variational regularization and fusion of surface normal maps. In: 3DV (2014) [4](#)