Intrinsic Light Field Images

Elena Garces[1], Jose I. Echevarria[1,3], Wen Zhang[2], Hongzhi Wu[2], Kun Zhou[2] and Diego Gutierrez[1]

[1]Universidad de Zaragoza, I3A
[2]State Key Lab of CAD & CG, Zhejiang University
[3]Adobe Systems

Abstract

We present a method to automatically decompose a light field into its intrinsic shading and albedo components. Contrary to previous work targeted to 2D single images and videos, a light field is a 4D structure that captures non-integrated incoming radiance over a discrete angular domain. This higher dimensionality of the problem renders previous state-of-the-art algorithms impractical either due to their cost of processing a single 2D slice, or their inability to enforce proper coherence in additional dimensions. We propose a new decomposition algorithm that jointly optimizes the whole light field data for proper angular coherency. For efficiency, we extend Retinex theory, working on the gradient domain, where new albedo and occlusion terms are introduced. Results show our method provides 4D intrinsic decompositions difficult to achieve with previous state-of-the-art algorithms. We further provide a comprehensive analysis and comparisons with existing intrinsic image/video decomposition methods on light field images.

1. Introduction

Intrinsic scene decomposition is the problem of separating the integrated radiance from a captured scene, into a physically-based and more meaningful reflectance and shading components, so that $Scene = Albedo \times Shading$; enabling quick and intuitive edits of the materials or lighting of a scene.

However, this decomposition is a very challenging, ill-posed problem. Given the interplay between the illumination, geometry and materials of the scene, there are more unknowns than equations for each pixel of the captured scene. To mitigate this uncertainty, existing *intrinsic decomposition* works assume that some additional properties of the scene are known. However, the prevailing goal is always the same: the gradients of the depicted scene need to be classified as coming from a variation in albedo, shading, or both. In this work, we build on classical theories of Retinex to obtain better predictors of these variations leveraging information from the light field data.

At the same time, light field photography is becoming more popular, as multi-view capabilities are progressively introduced in commercial cameras [Lyt13, Ray13], including mobile devices [VLD*13]. Such captured light fields are 4D structures that store both spatial and angular information of the radiance that reaches the sensor of the camera. This means a correct intrinsic decomposition has to be coherent in the angular domain, which increases the complexity with respect to 2D single images and 3D videos (x,y,t). Not only because of the number of additional information to be processed, but also because of the kind of coherence required.

A naïve approach to intrinsic light field decomposition would be to apply any state-of-the-art single image algorithm to each view of the light field independently. However, apart from not taking advantage from the additional information provided by multiple views, angular coherence is not guaranteed. So, additional processing would be required to make all the partial solutions, typically around 9×9, converge into a single one. Another approach could be to extend intrinsic video decompositions to 4D light field volumes, as these techniques rely on providing an initial solution for a 2D frame (usually the first), which is then propagated along the temporal dimension. These algorithms are already designed to keep consistency between frames, but they do not respect the 4D structure in a light field as all images need to be arranged as a single sequence, where the optimal arrangement is unknown. Moreover, the 2D nature of the decomposition propagated back and forth does not fully exploit the information implicitly captured in 4D.

Therefore, we propose an approach that jointly optimizes for the whole light field data, leveraging its structure for better cues and constraints for solving the problem; and enforcing proper angular coherency by design. We test our algorithm with both synthetic light fields, and real world ones captured with Lytro cameras. Our results demonstrate the benefits of working in 4D in terms of coherence and quality of the decomposition itself.

2. Related Work

Intrinsic decomposition of the shading and albedo components of an image is a long-standing problem in computer vision and graphics since it was formulated by Barrow and Tenembaum in the 70s [BT72]. We review previous intrinsic decomposition algorithms based on their input, and then briefly cover related light field processing.

Single Image. Several works rely on the original Retinex theory [LM71] to estimate the *shading* component. By assuming that shading varies smoothly, either pixel-wise [TFA05, ZTD*12] or cluster-based [GMLMG12] optimization is performed. Clustering strategies have also been used to obtain the *reflectance* component, e.g. assuming a sparse number of reflectances [GRK*11, SY11], using a dictionary of learned reflectances from crowd-sourcing experiments [BBS14], or flattening the image to remove shading variations [BHY15]. Alternative methods require user interaction [BPD09], jointly optimize the shape, albedo and illumination [BM15], incorporate priors from data driven statistics [ZKE15], train a Convolutional Neural Network (CNN) with synthetic datasets [NMY15], or use depth maps acquired with a depth camera to help disambiguate shading from reflectance [BM13, CK13, LZT*12]. Although some of these algorithms can produce good quality results, they require additional processing for angular coherence, and they do not make use of the implicit information captured by a light field. Our work is based on the Retinex theory, with 2D and 4D scene-based heuristics to classify reflectance gradients.

Multiple Images and Video. Several works leverage information from multiple images of the same scene from a fixed viewpoint under varying illumination [Wei01, HWU*14, LB15, SMPR07]. Laffont et al. [LBP12] coarsely estimate a 3D point cloud of the scene from non-structured image collections. Pixels with similar chromaticity and orientation in the point cloud will be used as reflectance constraints within an optimization. Assuming outdoor environments, the work of Duchene et al. [DRC*15] estimates sunlight position and orientation and reconstructs a 3D model of the scene, taking as input several captures of the same scene under constant illumination. Although a light field can be seen as a structured collection of images, we do not make assumptions about the lighting nor the scale of the captured scene.

Video. A few methods dealing with intrinsic video have been recently presented. Ye et al. [YGL*14] propose a probabilistic solution based a casual-anticasual, coarse-to-fine iterative reflectance propagation. Bonneel et al. [BST*14] present an efficient gradient-based solver which allows interactive decompositions. Kong et al. [KGB14] rely on optical flow to estimate surface boundaries to guide the decomposition. Recently, Meka et al. [MZRT16] present a novel variational approach suitable for real-time processing, based on a hierarchical coarse-to-fine optimization. While this approach can provide coherent and stable results even applied straightforwardly to light fields, the actual decomposition is performed on a per-frame basis, so it shares the limitations with previous 2D methods.

Light fields. Related work on intrinsic decomposition of light field images and videos has been published concurrently. Bonneel et al. [BTS*17] present a general approach for stabilizing the results of *per-frame* image processing algorithms over an array of images and videos. Their approach can produce very stable results, but its generality does not exploit a 4D structure that can be used to handle complex non-lambertian materials [TSW*15, SAMG16]. On the other hand, Alperovich and Goldluecke [AG16] present an approach similar to ours posing the problem in ray space. By doing this, they ensure angular coherency and also handle non-lambertian materials. While we do not handle such materials explicitly, our algorithm produces sharper and more stable results, with comparable reconstructions of reflectances under specular highlights.

Light Field Editing. Our work is also related to papers that extend common tools and operations for 2D images to 4D light fields. This is not a trivial task, given again the higher dimensionality of light fields. Jarabo et al. [JMB*14] present a first study to evaluate different light field editing interfaces, tools and workflows, this study is further analyzed by Masia et al. [MJG14], providing a detailed description of subjects' performance and preferences for a number of different editing tasks. Global propagation of user strokes has also been proposed, using a a voxel-based representation [SK02], a multi-dimensional downsampling approach [JMG11], or preserving view coherence by reparameterizing the light field [AZJ*15], while other works focus on deformations and warping of the light field data [BSB16, COSL05, ZWGS02]. Cho et al. [CKT14] utilize the epipolar plane image to extract consistent alpha mattes of a light field. Guo et al. [GYK*15] stitch multiple light fields via multi-resolution, high dimensional graph cuts. There are also considerable interests in recovering depths from a light field. Existing techniques exploit defocus and correspondence depth cues [THMR13], carefully handle occlusions [WER15], or use variational methods [WG14]. As most of these works, we also rely on the epipolar plane for implicit multi-view correspondences and processing.

3. Formulation

To represent a light field, we use the two-plane parametrization on ray space $L(x,y,u,v)$, which captures a light ray passing through two parallel planes: the sensor plane Π_{uv}, and the virtual camera plane or image plane Ω_{xy}. Analogous to its 2D image counterpart, the problem of intrinsic light field decomposition can be formulated as follows: for each ray of the light field L, we aim to find its corresponding reflectance and shading components R and S, respectively.

$$L(x,y,u,v) = R(x,y,u,v) \times S(x,y,u,v) \tag{1}$$

Instead of solving for single rays directly, the problem can be formulated in the gradient domain for the image plane Ω_{xy}:

$$\nabla l(x,y,u^*,v^*) = \nabla r(x,y,u^*,v^*) + \nabla s(x,y,u^*,v^*) \tag{2}$$

more compactly $\nabla l = \nabla r + \nabla s$. Where l, r and s denote the single views for each $\{u^*,v^*\} \in \Pi_{uv}$ for each input view l, its reflectance r and shading s in log spaces. Note that we denote single views computed in log domain with lowercase, while uppercase letters denote the whole light field in the original domain.

The classic Retinex approach [LM71] proposes a solution to this formulation by classifying each gradient as either shading or albedo. As seen before, different heuristics have been proposed over the years, with the simplest one associating changes in albedo with changes in *chromaticity*. Although this provides compelling

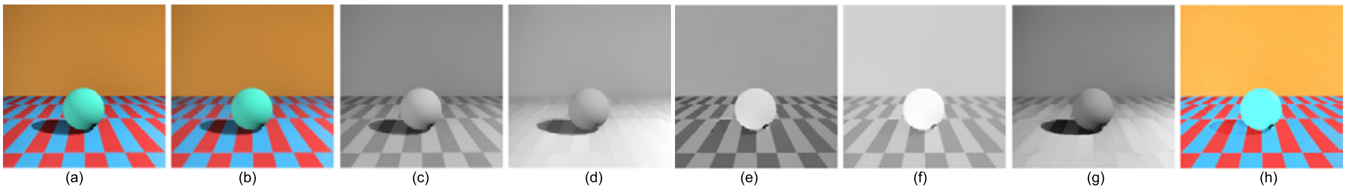


Figure 1: *Complete pipeline with a simple scene [AZJ*15]. The central view is shown here and the whole light field is shown in the Supplementary Material [Gar]. (a) Input light field L. (b) Filtered light field $\hat{L}$. (c) Normalized input $||\hat{L}||_2$. (d) Resulting shading S_1 from line 10 in 1 and Equation 8; note that although it looks consistent in one view, the global coherency is not guaranteed as shown in the Supplementary Material videos. (e) Resulting reflectance R_1 from from line 10 in 1 and Equation 8. (f) Filtered reflectance $\hat{R}_1$. (g) Final shading S_f. (h) Final reflectance R_f.*

results for some scenes, it still has the following limitations: chromatic changes do not always correspond to albedo changes; the solution is very sensitive to high frequency texture; and more importantly it does not take into account the effects of occlusion boundaries, where shading and albedo vary at the same time.

4. Our method

4.1. Overview

Our approach to the problem of intrinsic light field decomposition is based on a multi-level solution detailed in Algorithm 1: In a first step, we perform a global 4-dimensional l_1 filtering operation, which generates a new version of the light field with reduced high frequency textures and noise, to promote relevant gradients and edges, as well as improved angular coherency. The resulting light field, which we call $\hat{L}$, will serve to initialize a first estimation of the reflectance R_0 and shading components S_0 (Section 4.2). These initial estimations will then be used to compute the albedo and occlusion cues needed for the actual intrinsic decomposition, which is done locally per view (Sections 4.3.1 and 4.4), benefiting from the previous global processing of the whole light field volume. A final global 4D l_1 filtering operation (Section 4.5) performed over the reflectance finishes promoting angular coherency and stability, as can be seen in the results section and the Supplementary Material. The complete pipeline is shown in Figure 1.

4.2. Initialization

Inspired by the work of Bi et al. [BHY15], we noticed that better predictions of the albedo discontinuities can be done by performing an initial l_1 filtering of the light field volume, since it enhances edges and removes noise that could introduce errors in the estimation of gradients. In particular, we regularize the total variation (TV-l_1):

$$\min_{\hat{L}} \frac{1}{2}||\hat{L}-L||_2^2+\beta||\hat{L}||_1 \qquad (3)$$

As a result, from the original light field L, we obtain a filtered version $\hat{L}$, close to the original input but with sharper edges due to the use of l_1 norm on the second term. Additionally, the use of this norm effectively removes noise while prevents smoothing out other important features. The regularization factor β controls the degree of smoothing, where in our experiments $\beta = 0.05$.

Algorithm 1 Intrinsic Light Field Decomposition

```
1:  Input: Light field $L(x,y,u,v)$
2:                                        ▷ Initialization (Section 4.2)
3:  $\hat{L} \leftarrow$ TV-L$_1(L, \beta = 0.05)$
4:  $S_0 \leftarrow ||\hat{L}||_2$
5:  $R_0 \leftarrow \hat{L}/S_0$
6:                          ▷ Global Analysis (Sections 4.3.1 and 4.4)
7:  $\omega_a \leftarrow$ getAlbedoTh($\hat{L}$, $R_0$)
8:  $\omega_{occ} \leftarrow$ getOcclusionGradient($L_{depth}$)
9:                                    ▷ Local intrinsic decomposition
10: $R_1, S_1 \leftarrow \mathcal{G}(\hat{L}, \omega_a, \omega_{occ})$ ▷ Note that $R_1$ and $S_1$ are both single
    channel
11:                                  ▷ Global coherency (Section 4.5)
12: $\hat{R}_1 \leftarrow$ TV-L$_1(R_1, \beta = 0.05)$
13: $S_f \leftarrow ||\hat{L}||_2/\hat{R}_1$
14: $R_f \leftarrow L/S_f$
15: Result: $R = R_f(x,y,u,v)$, $S = S_f(x,y,u,v)$
```

Working with light fields means that we need to solve this multidimensional total variation problem in 4D. Since efficiency is key for our method to be practical, we use the ADMM solver proposed by Yang et al. [YWF*13]. ADMM combines the benefits of augmented Lagrangian and dual decomposition methods. It decomposes the original large global problem into a set of independent and small problems, which can be solved exactly and efficiently in parallel. Then it coordinates the local solutions to compute the globally optimal solution.

Figure 2, shows the difference in angular coherency and noise between the input L, a filtered version obtained from processing each single view independently, and our $\hat{L}$ obtained from the described global filtering. From $\hat{L}$, we compute the initial shading as, $S_0 = ||\hat{L}||_2$. This is a convenient step to obtain a single-channel version of the input image, with other common transformations like the RGB average or the luminance channel from CIELab [GMLMG12] providing similar performance. Taking S_0 as baseline, we compute the initial RGB reflectance R_0 simply from $\hat{L}/S_0$. It is important to note that S_0 and R_0 serve only as the basis over which our heuristics are applied to obtain the final cues to solve for the actual intrinsic decomposition (Equation 4). Figure 3 shows the impact of this l_1 regularization on the detection of albedo variations.

4.3. Intrinsic Estimation

As motivated before, for our efficiency requirements we follow a Retinex approach. We build on Zhao's closed-form formulation, extending it to take into account our albedo and occlusion cues obtained from the 4D light field volume. For each view l of the light field, the system computes the shading component s by minimizing the following equation:

$$\min_{s} \lambda_1 f_1(s) + \lambda_2 f_2(s) + \lambda_3 f_3(s) \tag{4}$$

where f_1 is the Retinex constraint, f_2 is an absolute scale constraint, and f_3 is a non-local texture cue; and λ_1, λ_2, and λ_3 are the weights which control the influence of each term, set to $\lambda_1 = 1$, $\lambda_2 = 1$ and $\lambda_3 = 1000$. In this work we extend f_1, so please refer to the original paper [ZTD*12] for the full details of f_2 and f_3.

4.3.1. Retinex-Based Constraint

The original Retinex formulation assumes that while shading varies smoothly, reflectance tends to cause sharp discontinuities, which can be expressed as:

$$f_1(s) = \sum_{i,j \in \mathcal{N}_{xy}} (\nabla s_{ij}^2 + \omega_{ij}^a \nabla r_{ij}^2) \tag{5}$$

where $\mathcal{N}_{xy}$ is the set of pairs of pixels that can be connected in a four-connected neighborhood defined in the image plane Ω_{xy}, and ω_{ij}^a is commonly defined as a threshold on the variations in the chromatic channels (Section 4.4). Following Equation 2, we define the following transformation, needed to solve Equation 4.

$$\nabla r = \nabla \hat{l} - \nabla s \tag{6}$$

However, we found that this equation ignores the particular case of occlusion boundaries, where shading and reflectance may vary at the same time. In order to handle such cases, we introduce a new additional term ω_{ij}^{occ}, which has a very low value when an occlusion is detected, so it does not penalize the corresponding gradients (more details in Section 4.4):

$$f_1(s) = \sum_{i,j \in \mathcal{N}_{xy}} \omega_{ij}^{occ} (\nabla s_{ij}^2 + \omega_{ij}^a \nabla r_{ij}^2) \tag{7}$$

We define as $\mathcal{G}$ the function that takes the whole light field and the global cues to obtain the corresponding shading and reflectance layers:

$$\mathcal{G}(\hat{L}, \omega^a, \omega^{occ}) = (S_1, R_1) \tag{8}$$

It is important to note that s has a single channel (an interesting future work would be to lift this restriction to allow colored illumination), so Equation 6 is also a single channel operation, where $\hat{l}$ is $||\hat{l}||_2$. Therefore, Equation 4 yields single channel shading s, and reflectance $r = ||l||_2 - s$ in log-spaces. Then, S_1 and R_1 are:

$$\forall u, v \in \Pi_{uv} \quad \begin{array}{rcl} S_1(x,y,u,v) & = & e^s \\ R_1(x,y,u,v) & = & e^r \end{array} \tag{9}$$

4.4. Gradient Labeling

In the following, we describe our extensions to the classic Retinex formulation: the albedo and occlusion terms in Equation 7. Note that this labeling is independent from solving the actual system (Equation 4), so each cue is computed in the most suitable color space, or additional available dimensions like depth.

4.4.1. Albedo Gradient (ω^a)

Albedo gradients are usually computed based on the chromatic information in CIELab color space. However, as we have shown, our initial RGB reflectance R_0 is better suited for this purpose, since it shows more relevant albedo variations. Staying in RGB space, we are inspired by the planar albedo assumption of Bousseau et al. [BPD09] and propose an edge-based analysis where if neighboring pixels $\{i, j\}$ are co-linear, their albedo is assumed to be constant. This is a heuristic which works reasonably well in practice except for black and white albedo, which are handled separately. We thus compute our weights as:

$$\omega_{ij}^a = \begin{cases} 0, & \text{if } \widehat{R_0^i, R_0^j} > 0.04 \\ 1, & \text{otherwise} \end{cases} \tag{10}$$

Setting $\omega_{ij}^a = 0$ in Equation 7, means that such gradient comes from albedo, so the gradient of the shading should be smooth. We found a difference of 0.04 radians works well in general, producing good results. We can see an example in Figure 3, where our measure is compared to the original Zhao's estimator, which only used Euclidean distances.

Our proposed heuristic works reasonably well when there is color information available, however it fails when colors are close to pure black or white. Thus, we choose to detect them independently and use them as similar cues as for regular albedo, so the final shading is not affected. We propose an approach based on the distance from a color to the black and white references in CIELab space (given its better perceptual uniformity than RGB), which gives a measure of the probability of a color being one of them.

From the light field $\hat{L}$, we compute the perceptual distance of each pixel to the white color as $\mathcal{D}_i^w = ||\hat{L}_i - w||_2^2$, and analogously the distance to black $\mathcal{D}_i^b$; where w and b may change depending on the implementation. With that, we compute the probability of a pixel of being white or black as $\mathcal{P}_i^w = \exp(-\mathcal{D}_i^w / \mathcal{D}_b^w)$, with $\mathcal{D}_b^w$ being the maximum distance in CIELab space (see Figure 4). Then, we label the gradients as:

$$g_{ij}^w = \begin{cases} 0, & \text{if } (\mathcal{P}_i^w \geq \tau || \mathcal{P}_j^w \geq \tau_1) \wedge (|\mathcal{D}_i^w - \mathcal{D}_j^w| > \tau_2) \\ 1, & \text{otherwise} \end{cases} \tag{11}$$

where $\tau_1 = 0.85$ and $\tau_2 = 0.05$. And we impose the additional condition that it must be a real gradient, so $|\mathcal{D}_i^w - \mathcal{D}_j^w| > \tau_2$ avoids marking pixels inside uniform areas. The black albedo labeling g_{ij}^b is analogously formulated. τ_1 and τ_2 were set empirically, but work well for all tested scenes. Then, we compute the final threshold for each gradient as $\omega_{ij}^a = \max(\omega_{ij}^a, g_{ij}^w, g_{ij}^b)$. The result of this step is a binary labeling, where each gradient is labeled as albedo or shading change (Figure 4).

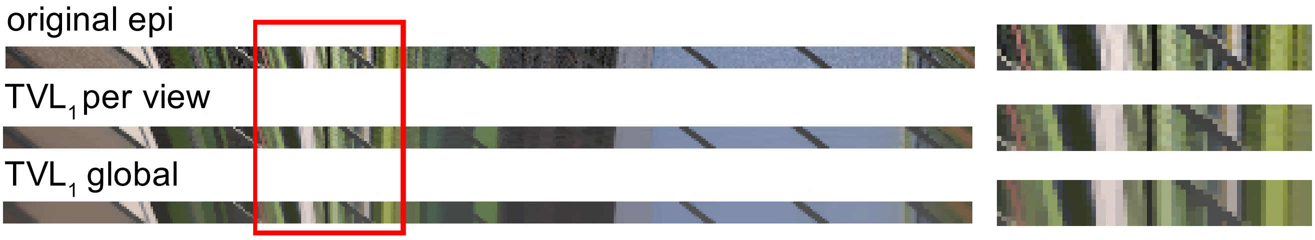


Figure 2: *Visualization of the horizonal epi view for the red scanline in Figure 3 (a). From top to bottom: the epi from the original light field; the epi after applying TVL_1 filter to each view separately; the same epi after applying a 4D TVL_1 filter to the whole light field volume using our approach. We can observe (by zooming in the digital version), areas with very similar colors are flattened, while sharp discontinuities are preserved, effectively removing noise and promoting angular coherence.*

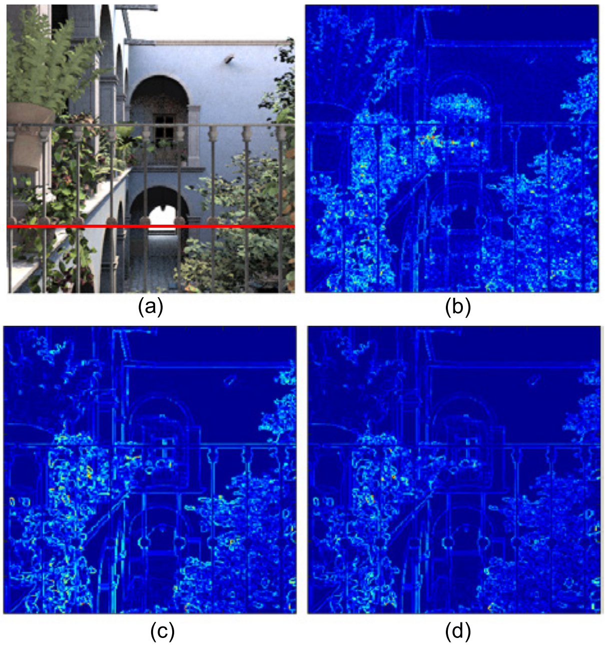


Figure 3: *(a) Central view of an input light field. (b) Albedo variations computed as the angle between RGB vectors for neighboring pixels $\widehat{L^i, L^j}$, from the original light field L. (c) Albedo variations obtained from our initial reflectance estimation, $\widehat{R_0^i, R_0^j}$. (d) Albedo variation from the chromaticity norm, $||\hat{L}^i - \hat{L}^j||$, used by Zhao et al [ZTD*12]. Our approach (c) yields cleaner gradients than (b), and captures more subtleties than (d). Note for example the green leaves at the right of the image. Every image is normalized to its maximum value.*

4.4.2. Occlusion Gradient (ω^{occ})

Previous work assume that discontinuities come from changes in albedo or changes in shading, but not both. However, we found they can actually occur simultaneously at occlusion boundaries, becoming an important factor in the intrinsic decomposition problem. Our key idea then is to detect the corresponding gradients and assign them a low weight ω_{ij}^{occ} in Equation 7, so larger changes are allowed in shading and albedo at the same time. Contrary to single 2D images, 4D light fields provide several ways to detect occlusions, like analyzing the epipolar planes [AF05, WG14] or using defocus cues [WER15]. In the following, we describe a simple heuristic assuming an available depth map [TSW*15], although it can be easily adjusted if only occlusion boundaries are available:

$$\omega_{ij}^{occ} = \begin{cases} 0.01, & \text{if } |D_i - D_j| > 0.02 \\ 1, & \text{otherwise} \end{cases} \tag{12}$$

where the depth map D is normalized between 0 and 1. Note that we cannot set $\omega_{ij}^{occ} = 0$ because it would cause instabilities in the optimization. Figure 5 (c), show the effect of including this new term.

4.5. Global Coherency

After solving Equation 8 we get S_1 and R_1. Given the way normalization of shading values is performed in Equation 4, we found some views may become a bit unstable, affecting the angular coherence of the results. A straightforward approach could be to apply another 4D l_1 filter (Equation 3) over S_1. But, this tends to remove details, wrongly transferring them to the reflectance producing an over-smoothed shading layer and a noisier reflectance one.

We found filtering R_1 provides better results. Because R_1 already features uniform regions of color, the 4D l_1 filter finishes flattening them for enhanced angular coherence, obtaining $\hat{R}_1$. Again, we use $\beta = 0.05$. From there, we compute our final smooth and coherent shading S_f as $||\hat{L}||_2 / \hat{R}_1$. And the final RGB reflectance as $R_f = L/S_f$.

5. Results and Evaluation

We show the whole pipeline in Figure 1. The central view is shown after each step of the Algorithm 1, plus the whole light field is shown in the Supplementary Material [Gar]. The input light field L, the filtered version $\hat{L}$ and the normalized version $||\hat{L}_2||$ are shown in Figures (a) to (c). We observe that the variation between the original light field L and the filtered one $\hat{L}$ is very subtle. In particular, in

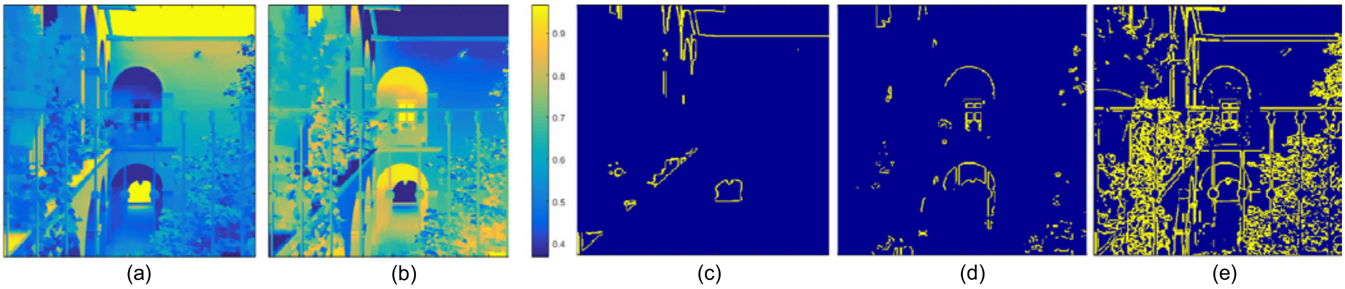
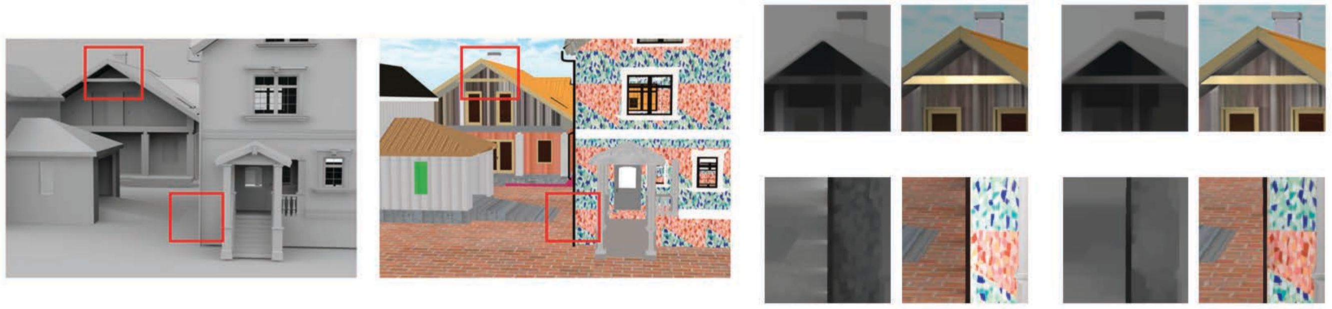


Figure 4: *(a) Probability of being white, $\mathcal{P}_i^w$ (b) Probability of being black, $\mathcal{P}_i^b$ (c) White pixels masked after g_{ij}^w (d) Black pixels masked after g_{ij}^b (e) Final albedo weights ω_{ij}^a taking into account color, white, and black information.*



(a) ground truth shading and reflectance (b) without depth cue (c) with depth cue

Figure 5: *(a) Ground truth shading. (b) Ground truth reflectance. (c) Without ω^{occ}, the algorithm classifies some prominent gradients as albedo, so it enforces continuous shading, causing artifacts. Taking occlusions into account fixes this limitation, producing results closer to the reference.*

this figure, it is more noticeable in very dark regions where black gradients become grayish. This is favorable to the gradient-based solver we use to solve Equation 4, which is very sensitive to very dark areas (with values close to zero). The output from Equation 8 is shown in Figures (d) and (e), and, although the shading looks pretty consistent in one view, it lacks of angular consistency when the whole volume is visualized (as shown in the Supplementary). Finally, from the filtered reflectance $\hat{R}_1$ (f) and the original light field L, we are able to recover the coherent shading S_f (g) and reflectance layers R_f (h). Note that the initial filtering operation also removes small details in shadows and texture, which are recovered in the reflectance layer. This is favorable if the details removed are high frequency texture, as we can see in Figure 8 (first row), but may also cause small remnants of shading in the reflectance, as we can see in Figure 1 (h).

In addition to the scenes shown for the comparisons, we also provide a different set of results with our method in a variety of real and synthetic scenes in our Supplementary Material. In Figure 6 we show the full result for *sanmiguel* scene without and with the occlusion cue. In this example, knowing the depth map improves the albedo decomposition as the left-most part of the image is more balanced. In the other two scenes (*plants* and *livingroom*) the difference between both scenarios is more subtle so we just show here the output with the cue. We can observe again that our filtering step favors high frequency albedo details. As has been noted in related work, there is a close relationship between intrinsic estimation and high frequency detail removal [BHY15].

Intrinsic light field decomposition extends the range of edits that can be performed to a light field with available tools [JMB*14, MJG14]. Figure 7 shows two examples, where simple albedo and shading edits allow to change the appearance coherently across the angular domain. Please note more advanced manipulations like texture replacement are still an open problem in 4D.

5.1. Discussion

In the following, we discuss and compare our approach with related work and some straightforward alternatives. Our results for the comparisons do not make use of the occlusion cues. For all of them we show the final decomposition for the central view of the light field. Angular coherence can be inspected in the animated sequences included in the Supplementary Material [Gar].

Single Image. Figure 8 shows a comparison with 2D state-of-the-art methods that use a single color image as input. The method of

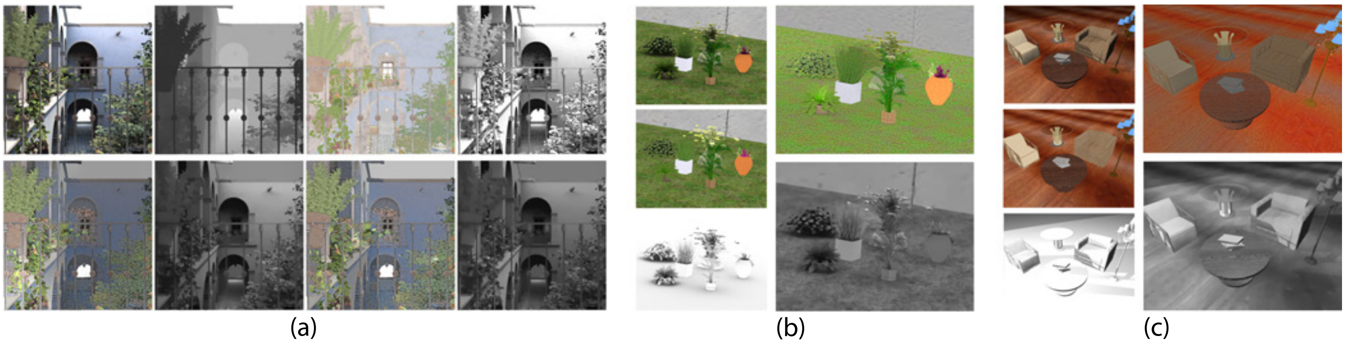


Figure 6: *(a)* sanmiguel. *First row: input, depth map and ground truth albedo and shading. Second row: left, our result without occlusion cue; right, our result with occlusion cue. (b)* living room. *Left column: input, ground truth albedo and shading. Right column: our result with occlusion cue. (c)* plants. *Left column: input, ground truth albedo and shading. Right column: our result with occlusion cue.*

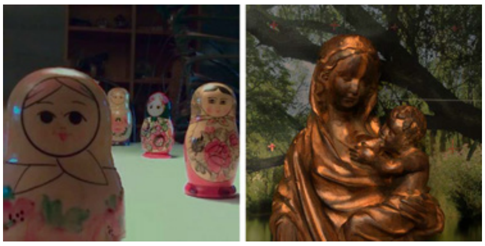


Figure 7: *Simple editing operations performed by modifying the albedo (left) and shading (right) layers independently. Check the accompanying videos to see the complete edited light field.*

Chen et al. [CK13] requires an additional depth map, which in comparable real scenarios could be reconstructed from the light field itself (we use Wang et al. [WER15] for this matter). In terms of overall accuracy of the decomposition, it could be argued that the RGB-D approach provides better results, specially in the shading component. However, results tend to be overly smooth and artifacts appear when the reconstructed depth map is not accurate enough. But more important, this approach requires non-trivial additional processing for solving the remaining views given depth maps are usually computed only for the central view. Compared to the other single image inputs, our method provides very similar results per view, while it keeps the angular coherency (see the Supplementary Material to observe the flickering artifacts that appear solving the decomposition per view). Straight processing of the whole array of views as a single image is obviously impractical given the huge number of equations to be solved.

Video. If the different views captured in a 4D light field are arranged as a single sequence, they can be interpreted as a video, and so previous intrinsic video solutions can be applied. While the optimal sequence is unknown, we chose the one in Figure 9 (left). Apart from specific intrinsic video algorithms, we also tested a more general approach based on blind temporal consistency [BTS*15], where the single image solutions from the previous paragraph were applied per frame, to be then processed for enhanced coherence (an approach that can be also found in concurrent work [BTS*17]). As can be seen in Figure 8, both methods, Bonneel et al. [BST*14] and Meka et al. [MZRT16], produce results that tend to be too smooth, with visible flickering and haloing artifacts when played in a different order from the original sequence (proper angular coherence needs to be independent of the order of visualization). Blind temporal consistency [BTS*15] applied over single frames from Zhao et al. [ZTD*12] and Bell et al. [BBS14] is able to produce stable results when the baseline between views is very little as the per view decompositions are very similar. However, while this seems to be an effective way of enforcing angular coherence, working independently over single frames has some limitations when it comes to extensions to handle non-lambertian surfaces. This is something out of the scope of our paper, but an interesting venue for future work as already demonstrated in related work [TSW*15, AG16, SAMG16].

Light Field Images. Finally, concurrent work has appeared also decomposing 4D light field images into their intrinsic components. In their paper, Alperovich and Goldluecke [AG16] also pose the problem in the 4D ray space, with the additional goal of separating specular reflections from the albedo and shading. Figure 10 shows comparisons between the processed central views, while the animated sequences in the Supplementary Material showcase angular coherence. From the static images, similar overall quality is achieved. It is interesting to see, however, that although we do not explicitly process specular highlights, our reflectance layers are able to recover better values in some of these regions (mirror ball in *Mona's room* and the blue owl figurine). From the animated sequences, our results show less flickering and so better angular coherence. It is worth mentioning that because of the computing requirements, we were not able to get the full decomposed light fields from Alperovich and Goldluecke [AG16]. Our method, however, has still room for optimization, given each 2D view can be solved in parallel, before and after the 4D operations.

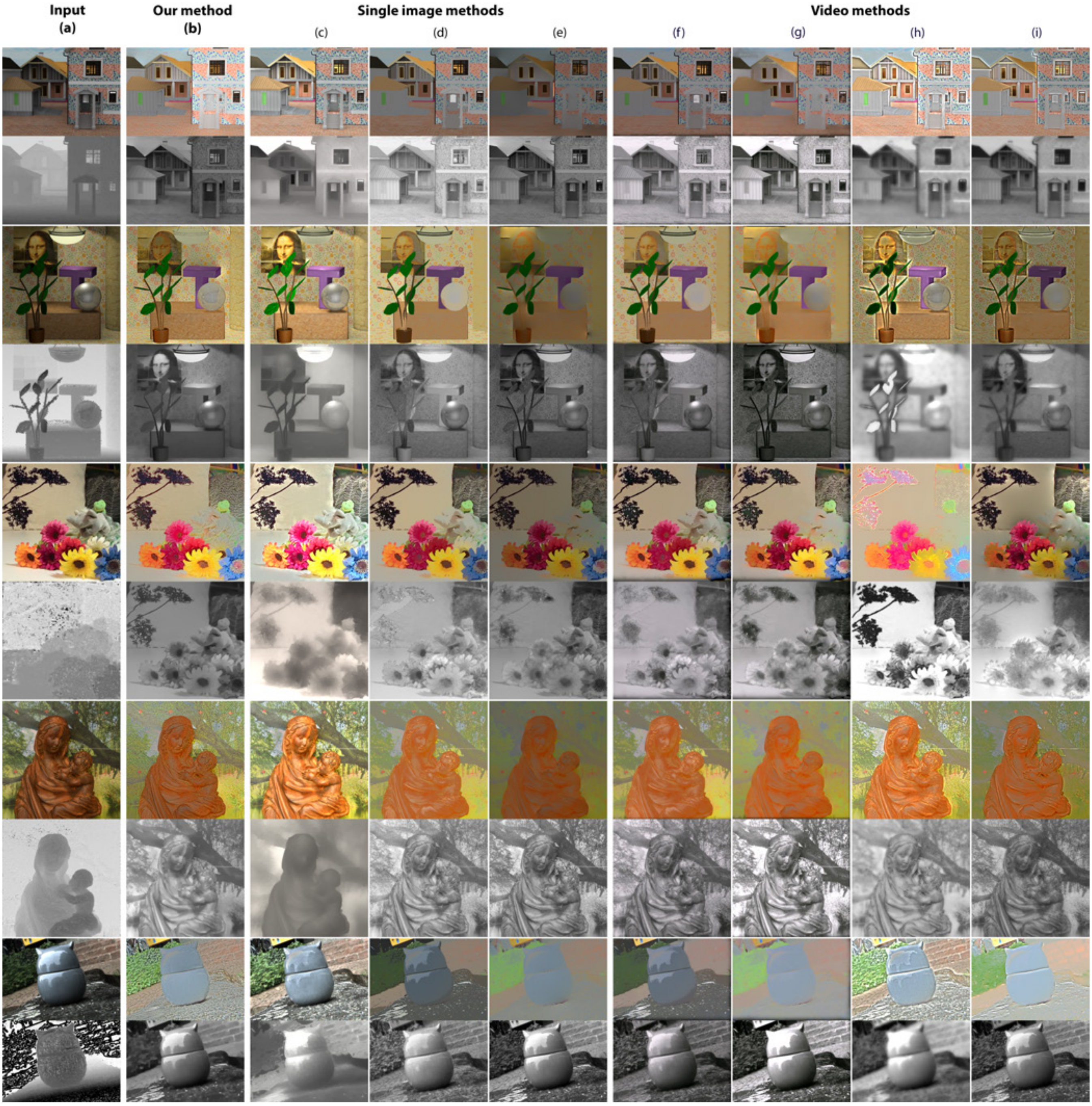


Figure 8: *(a) Input RGB and depth data (computed using Wang et al. [WER15]). (b) Our results. Single image approaches: (c) Chen and Koltun [CK13], (d) Bell et al. [BBS14], (e) Zhao et al. [ZTD* 12]. Video approaches: (f)-(g) single image methods ((d) and (e)) filtered using blind temporal consistency [BTS* 15], (h) Meka et al. [MZRT16], (i) Bonneel et al. [BST* 14]. The scenes are named, from top to bottom: outdoor, Monas' room, frog, Maria, and owlstr.*

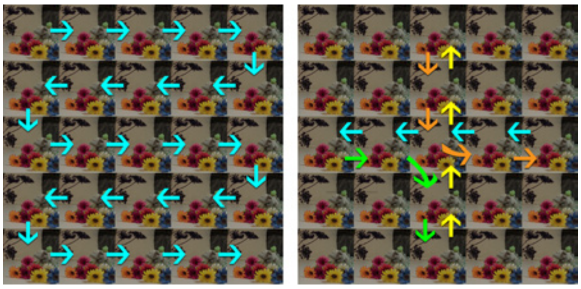


Figure 9: *Left: Sequence for video processing. Right: sequence for the animations in the Supplementary Material.*

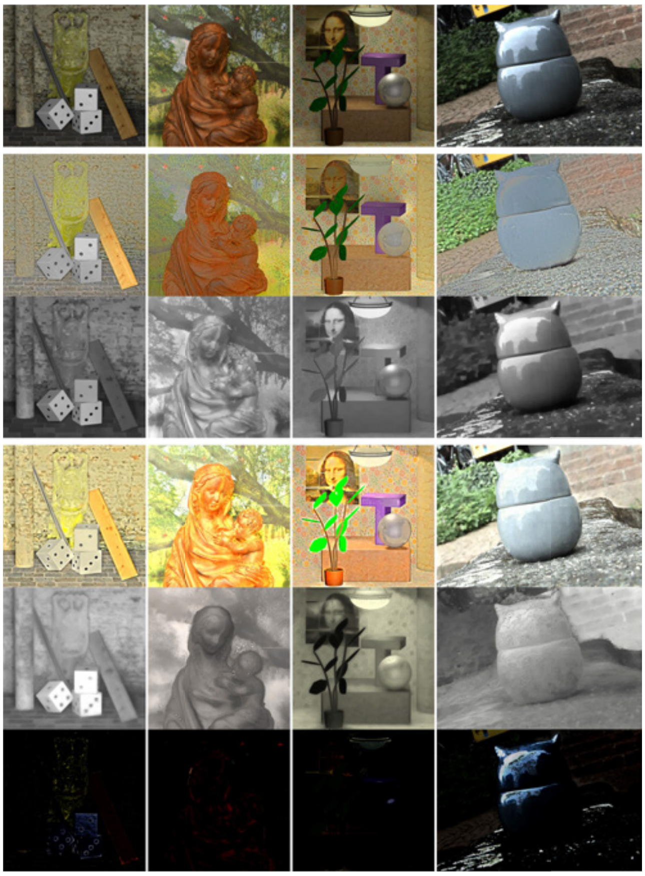


Figure 10: *Light field methods. From top to bottom: center view of input light field; our results; results from Alperovich and Goldluecke [AG16], including their additional specular layer. Given this extra layer, it is easier to compare results based on reflectance alone, where we are able to recover more plausible values in areas covered by strong specular highlights.*

6. Conclusions and Future Work

We have presented a new method for intrinsic *light field* decomposition, which adds to existing approaches for single images and video, enabling practical and intuitive edits in 4D. Our method is based on Retinex formulation, reviewed and extended to take into account the particularities and requirements of 4D light field data. We have shown results on both synthetic and real datasets, which compare favorably against existing state-of-the-art methods, as shown by the accompanying videos in the supplemental material.

For our albedo and occlusion cues, we currently rely on simple thresholds. A more sophisticated solution could make use of multi-dimensional Conditional Random Fields [JKG16]. Despite the flexibility of our formulation with respect to depth data, a current limitation is that its quality can directly affect the final results. More sophisticated occlusion heuristics could combine information from the epipolar planes to make this term more robust.

Finally, to reduce the complexity of the intrinsic decomposition problem, some simplifying assumptions are usually made, with the most relevant ones about the color of the lighting (white light) and the material properties of the objects in the scene (non-specular lambertian surfaces). As we have seen, although some approaches adapted from video processing can arguably match our method in terms of stability and quality of the decomposition, extensions to handle more complex materials and scenes can be posed more naturally and effectively in 4D space, paving the way for interesting future work.

7. Acknowledgements

We thank the reviewers for their insightful comments, Anna Alperovich for their datasets, Nicolas Bonneel and Abhimitra Meka for kindly providing the necessary comparisons, Adrian Jarabo and Belen Masia for fruitful discussions and synthetic scenes. This research has been funded by the European Research Council (Consolidator Grant, project Chameleon, ref. 682080), as well as the Spanish Ministry of Economy and Competitiveness (project LIGHTSLICE, ref. TIN2013-41857-P). The authors from Zhejiang University were partially supported by the National Key Research & Development Plan of China (2016YFB1001403), NSFC (No. U1609215) and the Fundamental Research Funds for the Central Universities.

References

[AF05] Apostoloff N., Fitzgibbon A.: Learning Spatiotemporal T-junctions for Occlusion Detection. In *Proc. Conference on Computer Vision and Pattern Recognition* (June 2005), IEEE. 5

[AG16] Alperovich A., Goldluecke B.: A Variational Model for Intrinsic Light Field Decomposition. In *Proc. Asian Conference on Computer Vision* (2016). 2, 7, 9

[AZJ*15] Ao H., Zhang Y., Jarabo A., Masia B., Liu Y., Gutierrez D., Dai Q.: Light Field Editing Based on Reparameterization. In *Proc. Pacific-Rim Conference on Multimedia* (2015), Springer. 2, 3

[BBS14] Bell S., Bala K., Snavely N.: Intrinsic Images in the Wild. *ACM Trans. Graphics (Proc. SIGGRAPH) 33*, 4 (2014). 2, 7, 8

[BHY15] Bi S., Han X., Yu Y.: An L 1 Image Transform for Edge-Preserving Smoothing and Scene-Level Intrinsic Decomposition. *ACM Trans. Graphics (Proc. SIGGRAPH) 34*, 4 (2015). 2, 3, 6

[BM13] Barron J. T., Malik J.: Intrinsic Scene Properties from a Single RGB-D Image. In *Proc. Computer Vision and Pattern Recognition* (2013), IEEE. 2

[BM15] Barron J., Malik J.: Shape, Illumination, and Reflectance from Shading. *IEEE Trans. Pattern Analysis and Machine Intelligence 37* (2015). 2

[BPD09] Bousseau A., Paris S., Durand F.: User-assisted Intrinsic Images. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 28*, 5 (2009). 2, 4

[BSB16] Birklbauer C., Schedl D. C., Bimber O.: Nonuniform Spatial Deformation of Light Fields by Locally Linear Transformations. *ACM Trans. Graphics 35*, 5 (2016). 2

[BST*14] Bonneel N., Sunkavalli K., Tompkin J., Sun D., Paris S., Pfister H.: Interactive Intrinsic Video Editing. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 33*, 6 (2014). 2, 7, 8

[BT72] Barrow H. G., Tenenbaum J. M.: Recovering Intrinsic Scene Characteristics from Images. In *Proc. Computer Vision Systems* (1972). 2

[BTS*15] Bonneel N., Tompkin J., Sunkavalli K., Sun D., Paris S., Pfister H.: Blind Video Temporal Consistency. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 34*, 6 (2015). 7, 8

[BTS*17] Bonneel N., Tompkin J., Sun D., Wang O., Sunkvalli K., Paris S., Pfister H.: Consistent Video Filtering for Camera Arrays. *Computer Graphics Forum (Proc. Eurographics) 36*, 2 (2017). 2, 7

[CK13] Chen Q., Koltun V.: A Simple Model for Intrinsic Image Decomposition with Depth Cues. In *Proc. International Conference on Computer Vision* (2013), IEEE. 2, 7, 8

[CKT14] Cho D., Kim S., Tai Y.-W.: Consistent Matting for Light Field Images. In *Proc. European Conference on Computer Vision* (2014), Springer. 2

[COSL05] Chen B., Ofek E., Shum H.-Y., Levoy M.: Interactive Deformation of Light Fields. In *Proc. Symposium on Interactive 3D Graphics and Games* (2005), ACM. 2

[DRC*15] Duchêne S., Riant C., Chaurasia G., Lopez-Moreno J., Laffont P.-Y., Popov S., Bousseau A., Drettakis G.: Multi-View Intrinsic Images of Outdoors Scenes with an Application to Relighting. *ACM Trans. Graphics 34*, 5 (2015). 2

[Gar] Intrinsic Light Fields - Supplementary Material. http://webdiis.unizar.es/~elenag/projects/intrinsicLF/supplementary/supplementary.html. Accessed: 2017-02-28. 3, 5, 6

[GMLMG12] Garces E., Munoz A., Lopez-Moreno J., Gutierrez D.: Intrinsic Images by Clustering. *Computer Graphics Forum (Proc. EGSR) 31*, 4 (2012). 2, 3

[GRK*11] Gehler P. V., Rother C., Kiefel M., Zhang L., Schölkopf B.: Recovering Intrinsic Images with a Global Sparsity Prior on Reflectance. In *Proc. Neural Information Processing Systems* (2011). 2

[GYK*15] Guo X., Yu Z., Kang S. B., Lin H., Yu J.: Enhancing Light Fields through Ray-Space Stitching. *IEEE Trans. Visualization and Computer Graphics*, 99 (2015). 2

[HWU*14] Hauagge D., Wehrwein S., Upchurch P., Bala K., Snavely N.: Reasoning about Photo Collections using Models of Outdoor Illumination. In *Proc. British Machine Vision Conference* (2014). 2

[JKG16] Jampani V., Kiefel M., Gehler P. V.: Learning Sparse High Dimensional Filters: Image Filtering, Dense CRFs and Bilateral Neural Networks. In *Proc. Computer Vision and Pattern Recognition* (2016), IEEE. 9

[JMB*14] Jarabo A., Masia B., Bousseau A., Pellacini F., Gutierrez D.: How Do People Edit Light Fields? *ACM Trans. Graphics (Proc. SIGGRAPH) 33*, 4 (2014). 2, 6

[JMG11] Jarabo A., Masia B., Gutierrez D.: Efficient Propagation of Light Field Edits. In *Proc. SIACG* (2011). 2

[KGB14] Kong N., Gehler P. V., Black M. J.: Intrinsic Video. In *Proc. European Conference on Computer Vision* (2014), Springer. 2

[LB15] Laffont P.-Y., Bazin J.-C.: Intrinsic Decomposition of Image Sequences from Local Temporal Variations. In *Proc. International Conference on Computer Vision* (2015), Springer. 2

[LBP12] Laffont P., Bousseau A., Paris S.: Coherent Intrinsic Images from Photo Collections. *ACM Trans. Graphics (Proc. SIGGRAPH) 31*, 6 (2012). 2

[LM71] Land E. H., McCann J. J.: Lightness and Retinex Theory. *Journal of the Optical Society of America 61*, 1 (1971). 2

[Lyt13] Lytro Inc.: The Lytro camera. http://www.lytro.com, 2013. 1

[LZT*12] 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), Springer. 2

[MJG14] Masia B., Jarabo A., Gutierrez D.: Favored Workflows in Light Field Editing. In *Proc. CGVCVIP* (2014). 2, 6

[MZRT16] Meka A., Zollhöfer M., Richardt C., Theobalt C.: Live Intrinsic Video. *ACM Trans. Graphics (Proc. SIGGRAPH) 35*, 4 (2016). 2, 7, 8

[NMY15] Narihira T., Maire M., Yu S. X.: Direct Intrinsics: Learning Albedo-Shading Decomposition by Convolutional Regression. In *Proc. International Conference on Computer Vision* (2015), Springer. 2

[Ray13] Raytrix GmbH: 3D Light Field Camera Technology. http://www.raytrix.de, 2013. 1

[SAMG16] Sulc A., Alperovich A., Marniok N., Goldluecke B.: Reflection Separation in Light Fields based on Sparse Coding and Specular Flow. In *Proc. Vision, Modeling & Visualization* (2016), Eurographics. 2, 7

[SK02] Seitz S. M., Kutulakos K. N.: Plenoptic Image Editing. *International Journal of Computer Vision 48*, 2 (2002). 2

[SMPR07] Sunkavalli K., Matusik W., Pfister H., Rusinkiewicz S.: Factored Time-lapse Video. *ACM Trans. Graphics (Proc. SIGGRAPH) 26*, 3 (2007). 2

[SY11] Shen L., Yeo C.: Intrinsic Images Decomposition using a Local and Global Sparse Representation of Reflectance. In *Proc. Computer Vision and Patter Recognition* (2011), IEEE. 2

[TFA05] Tappen M., Freeman W., Adelson E.: Recovering Intrinsic Images from a Single Image. *IEEE Trans. Pattern Analysis and Machine Intelligence 27*, 9 (2005). 2

[THMR13] Tao M. W., Hadap S., Malik J., Ramamoorthi R.: Depth from Combining Defocus and Correspondence Using Light-Field Cameras. In *Proc. International Conference on Computer Vision* (2013), IEEE. 2

[TSW*15] 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 Trans. Pattern Analysis and Machine Intelligence* (2015). 2, 5, 7

[VLD*13] Venkataraman K., Lelescu D., Duparré J., McMahon A., Molina G., Chatterjee P., Mullis R., Nayar S.: PiCam: An Ultra-thin High Performance Monolithic Camera Array. *ACM Trans. Graphics 32*, 6 (2013). 1

[Wei01] Weiss Y.: Deriving Intrinsic Images from Image Sequences. In *Proc. International Conference on Computer Vision* (2001), IEEE. 2

[WER15] Wang T.-c., Efros A. A., Ramamoorthi R.: Occlusion-aware Depth Estimation Using Light-field Cameras. In *Proc. International Conference on Computer Vision* (2015), Springer. 2, 5, 7, 8

[WG14] WANNER S., GOLDLUECKE B.: Variational Light Field Analysis for Disparity Estimation and Super-Resolution. *IEEE Trans. Pattern Analysis and Machine Intelligence 36*, 3 (2014). 2, 5

[YGL*14] YE G., GARCES E., LIU Y., DAI Q., GUTIERREZ D.: Intrinsic Video and Applications. *ACM Trans. Graphics (Proc. SIGGRAPH) 33*, 4 (2014). 2

[YWF*13] YANG S., WANG J., FAN W., ZHANG X., WONKA P., YE J.: An Efficient ADMM Algorithm for Multidimensional Anisotropic Total Variation Regularization Problems. In *Proc. International Conference on Knowledge Discovery and Data Mining* (2013), ACM. 3

[ZKE15] ZHOU T., KRÄHENBÜHL P., EFROS A. A.: Learning Data-driven Reflectance Priors for Intrinsic Image Decomposition. In *Proc. International Conference on Computer Vision* (2015), IEEE. 2

[ZTD*12] ZHAO Q., TAN P., DAI Q., SHEN L., WU E., LIN S.: A Closed-Form Solution to Retinex with Nonlocal Texture Constraints. *IEEE Trans. Pattern Analysis and Machine Intelligence 34*, 7 (2012). 2, 4, 5, 7, 8

[ZWGS02] ZHANG Z., WANG L., GUO B., SHUM H.-Y.: Feature-based Light Field Morphing. *ACM Trans. Graphics 21*, 3 (2002). 2