Intrinsic Light Fields

Elena Garces
Universidad de Zaragoza
egarces@unizar.es

Jose I. Echevarria
Universidad de Zaragoza
jecheva@unizar.es

Wen Zhang
Zhejiang University
zwcszju@zju.edu.cn

Hongzhi Wu
Zhejiang University
hwu@acm.org

Kun Zhou
Zhejiang University
kunzhou@acm.org

Diego Gutierrez
Universidad de Zaragoza
diegog@unizar.es

Abstract

We present the first 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.

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.

On the other hand, light field photography is becoming more popular, as multi-view capabilities are progressively introduced in commercial cameras [27, 32], including mobile devices [40]. Such captured light fields are 4D structures that store both spatial and angular information of the radiance that reach 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. To the best of our knowledge, no intrinsic decomposition method exists suited to handle these particularities.

A naïve solution 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, due to the extensive computational time that these algorithms require (Bell et al. [6] takes around 10 minutes for a 400×400 image), this is not an option given a typical light field contains around 9×9 views. And even then, angular coherence across views would not be guaranteed. 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. However, the propagation mechanism is designed to ensure smooth temporal transitions in a single dimension, assuming a normal playback of the video. While the 4D structure of a light field would allow its different views to be "stacked" and processed as a single video sequence, its angular dimensions cannot be assumed to be explored in a single way. Thus, all possible sequences should be taken into account, making this approach highly inefficient.

Therefore, we propose a new decomposition algorithm that jointly optimizes for the whole light field data effi-

ciently, is scalable, and maintains proper angular coherency. With this work, our goal is not to obtain the best decomposition for each single view, but the most coherent one, in a practical way. This not only keeps adding to the limited set of tools for light field editing [20, 46], but paves the way for other applications that require robust segmentation or selection of the same object and areas across all the different views. Our algorithm has the following characteristics:

- It is devised with 4D light field data in mind, enforcing angular coherence in the results.
- It is computationally efficient, while working with the whole 4D data volume.
- It builds on existing Retinex approaches, extending them with additional terms on the formulation.

We show results both with synthetic light fields (with ground truth references for reflectance, shading and depth), and real world light fields captured with a Lytro camera. We show how our method outperforms previous approaches when used with light field data, and is robust even in the presence of imperfect depth information.

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 [5]. 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 [25] to estimate the *shading* component. By assuming that shading varies smoothly, either a pixel-wise [39, 48] or cluster-based [15] optimization is performed. Clustering strategies have also been used to obtain the *reflectance* component, e.g. assuming a sparse number of reflectances [16, 34], using a dictionary of learned reflectances from crowd-sourcing experiments [6], or flattening the image to remove shading variations [7]. Alternative methods require user interaction [10], jointly optimize the shape, albedo and illumination [3], incorporate priors from data driven statistics [49], train a Convolutional Neural Network (CNN) with synthetic datasets [30], or use depth maps acquired with a depth camera to help disambiguate shading from reflectance [4, 12, 26]. For efficiency, our work is also 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 [43, 18, 24, 35]. Laffont et al. [23] 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. [14] 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 avoid the additional work required to build and process such proxy 3D models, by directly leveraging the 4D structure for a more effective and efficient approach.

Video. A few methods dealing with intrinsic video have been recently presented. Ye et al. [45] propose a probabilistic solution based a casual-anticasual, coarse-to-fine iterative reflectance propagation. Bonneel et al. [9] present an efficient gradient-based solver which allows interactive decompositions. Kong et al. [22] rely on optical flow to estimate surface boundaries to guide the decomposition. Recently, Meka et al. [29] present a novel variational approach suitable for real-time processing, based on a hierarchical coarse-to-fine optimization. While these methods work in a 3D domain, they are devised to keep a smooth coherency in the temporal dimension, which is assumed to be navigated in a continuous way. 4D Light fields, however, need to be consistent in all angular dimensions, without any assumption about the way they are going to be explored.

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. [20] present a first study to evaluate different light field editing interfaces, tools and workflows, this study is further analyzed by Masia et al. [28], 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 [33], a multi-dimensional downsampling approach [21], or preserving view coherence by reparameterizing the light field [1], while other works focus on deformations and warping of the light field data [8, 11, 47]. Cho et al. [13] utilize the epipolar plane image to extract consistent alpha mattes of a light field. Guo et al. [17] 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 [38], carefully handle occlusions [41], or use variational methods [42]. As most of these works, we also rely on the epipolar plane image for implicit and efficient multi-view correspondences and processing.

3. Formulation

To represent a light field, we use the two-plane parametrization on ray space $L(x, y, q, t)$, which captures a light ray passing through two parallel planes: the sensor plane Π_{qt}, 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.

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

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

$$\nabla L(x, y, q^*, t^*) = \nabla R(x, y, q^*, t^*) + \nabla S(x, y, q^*, t^*) \tag{2}$$

more compactly $\nabla l = \nabla r + \nabla s$. Where l, r and s denote the single views for each $\{q^*, t^*\} \in \Pi_{qt}$ 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 [25] 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 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.

Algorithm 1 Intrinsic Light Field Decomposition

Input: Light field $L(x, y, q, t)$
▷ Initialization (Section 4.2)
$\hat{L} \leftarrow$ TV-L$_1$(L, $\beta = 0.05$)
$S_0 \leftarrow ||\hat{L}||_2$
$R_0 \leftarrow \hat{L}/S_0$
▷ Global Analysis (Sections 4.3.1 and 4.4)
$\omega_a \leftarrow$ getAlbedoTh($\hat{L}$, R_0)
$\omega_{occ} \leftarrow$ getOcclusionGradient(L_{depth})
▷ Local intrinsic decomposition
$R_{g1}, S_1 \leftarrow \mathcal{G}(\hat{L}, \omega_a, \omega_{occ})$ ▷ Note that R_{g1} and S_1 are both single channel
▷ Global coherency (Section 4.5)
$\hat{R}_1 \leftarrow$ TV-L$_1$(R_{g1}, $\beta = 0.05$)
$S_f \leftarrow ||\hat{L}||_2/\hat{R}_1$
$R_f \leftarrow L/S_f$
Result: $R = R_f(x, y, q, t)$, $S = S_f(x, y, q, t)$

4.2. Initialization

Inspired by the work of Bi et al. [7], 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):

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

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 $\beta = 0.05$.

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. [44]. 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 1, 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 [15] 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 2 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 [48] 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:

$$\mathcal{G}(\hat{L}, \omega^a, \omega^{occ}) = (S_1, R_{g1}) \tag{8}$$

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 $r_g = ||l||_2 - s$ in log-spaces. Then, S_1 and R_{g1} are:

$$\forall q, t \in \Pi_{qt} \quad \begin{array}{lcl} S_1(x,y,q,t) & = & e^s \\ R_{g1}(x,y,q,t) & = & e^{r_g} \end{array} \tag{9}$$

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 choose to analyze albedo based on Omer and Werman's color lines model [31], which states that if the RGB vectors of two neighboring pixels $\{i, j\}$ are co-linear, their albedo is assumed to be constant. 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 2, 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

Figure 1. Visualization of the horizonal epi view for the red scanline in Figure 2 (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.

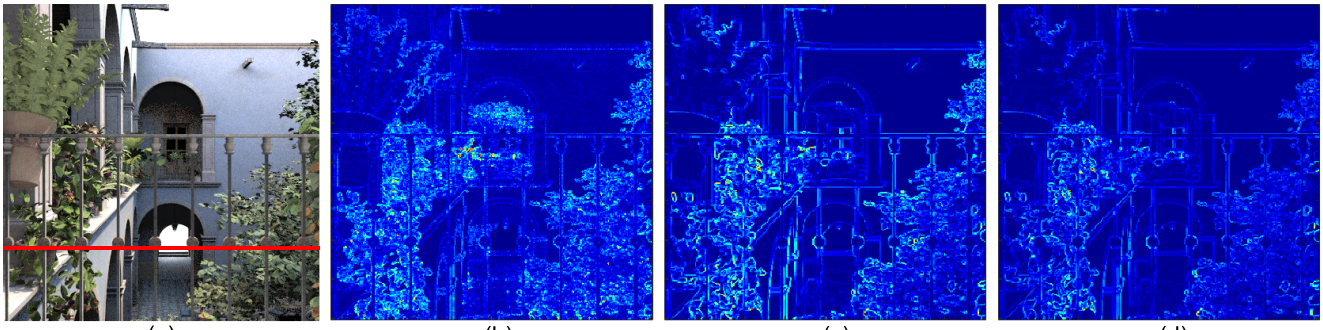


Figure 2. (a) Central view of an input light field. (b) Albedo variations computed as the angle between RGB vectors for neighboring pixels $\widehat{L^i, L^j}$, from the original light field L. (c) Albedo variations obtained from our initial reflectance estimation, $\widehat{R_0^i, R_0^j}$. (d) Albedo variation from the chromaticity norm, $||\hat{L}^i - \hat{L}^j||$, used by Zhao et al [48]. 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.

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 3). 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 albedo 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 3).

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 [2, 42] or using defocus cues [41]. In the following, we describe a simple heuristic assuming an available depth map [37], although it can be easily adjusted if only occlusion boundaries are available:

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

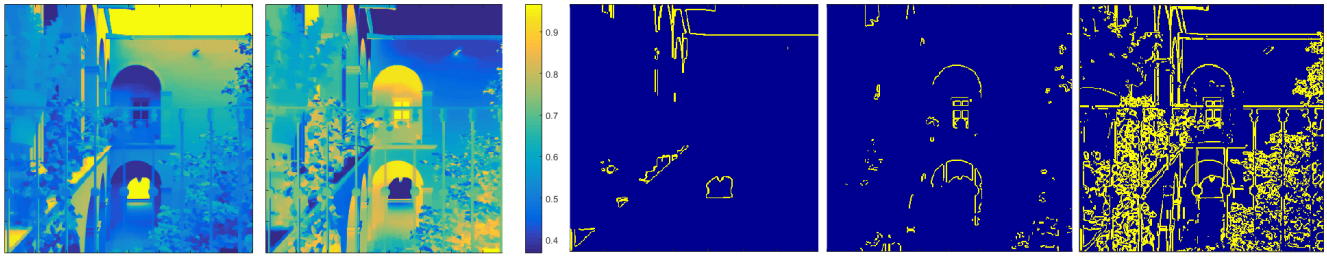


Figure 3. (a) Probability of being white, $\mathcal{P}_i^w$ (b) Probability of being black, $\mathcal{P}_i^b$ (c) White pixels masked after g_{ij}^w (d) Black pixels masked after g_{ij}^b (e) Final albedo weights ω_{ij}^a taking into account color, white, and black information.

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 4 (c), show the effect of including this new term.

4.5. Global Coherency

After solving Equation 8 we get S_1 and R_{g1}. 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_{g1} provides better results. Because R_{g1} 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 5. The central view is shown after each step of the Algorithm 1, plus the whole light field is shown in the Supplementary Material [36]. 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 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. 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 $\hat{R}_{g1}$ (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 6 (top), but may also cause small remnants of shading in the reflectance, as we can see in Figure 5 (h).

Figures 6 and 7 show several results using our method. We refer the reader to the Supplementary Material [36] to visualize the whole set of light fields with ground truth data and comparisons, where the critical improvement in angular coherence over state-of-art methods can be fully appreciated. All our results have been generated automatically with the fixed parameters given in the text.

Figure 6 shows two synthetic scenes of our dataset along with comparisons with the works of Zhao et al. [48] and Bell et al. [6]. In the case of Bell's method, it is based on a fully connected conditional random field, which produces a dense graph connecting all pixels in the image. This strategy does not converge to a coherent solution even for adjacent views of the light field, which translates into obvious flickering artifacts. Moreover, the reported computational cost of the method is of the order of 10 minutes for a moderate-size image, whereas our method, which is based on a much simpler linear system, performs in about 10 seconds. Since all the light fields shown here have 9×9 views, the efficiency of our method becomes crucial to make the method practical. Zhao's method, as can be observed in the accompanying videos, is not free of angular flickering, which is eliminated almost completely with our new albedo and occlusion gradients, plus 4D filtering.

Intrinsic light field decomposition extends the range of edits that can be performed to a light field with available tools [20, 28]. Figure 8 shows two examples, where simple albedo and shading edits allow to change the appearance coherently across the angular domain, something very difficult to achieve in RGB space.

6. Conclusions and Future Work

We have presented the first method for intrinsic *light field* decomposition, which follows up existing approaches for single images and video, enabling practical and intuitive ed-

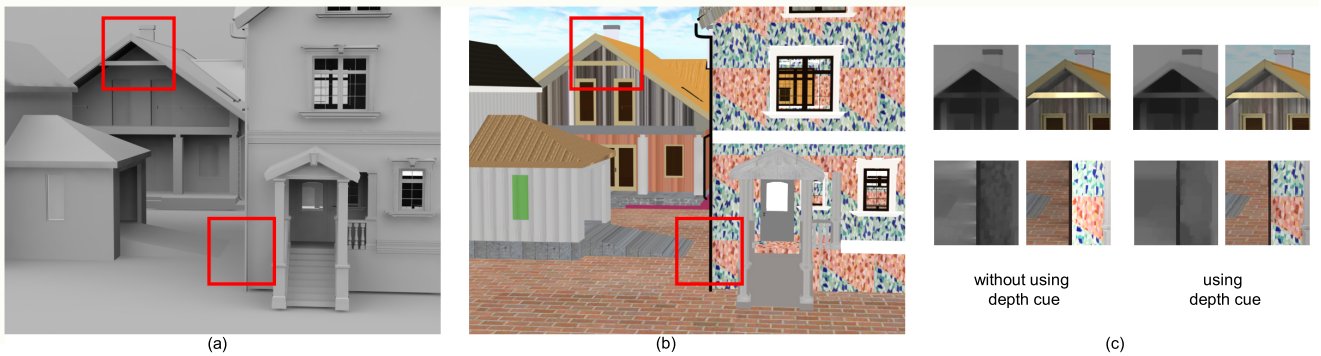


Figure 4. (a) Ground truth shading. (b) Ground truth reflectance. (c) Without ω^{occ}, the algorithm classifies some prominent gradients as albedo, so it enforces continuous shading, causing artifacts. Taking occlusions into account fixes this limitation, producing results closer to the reference.

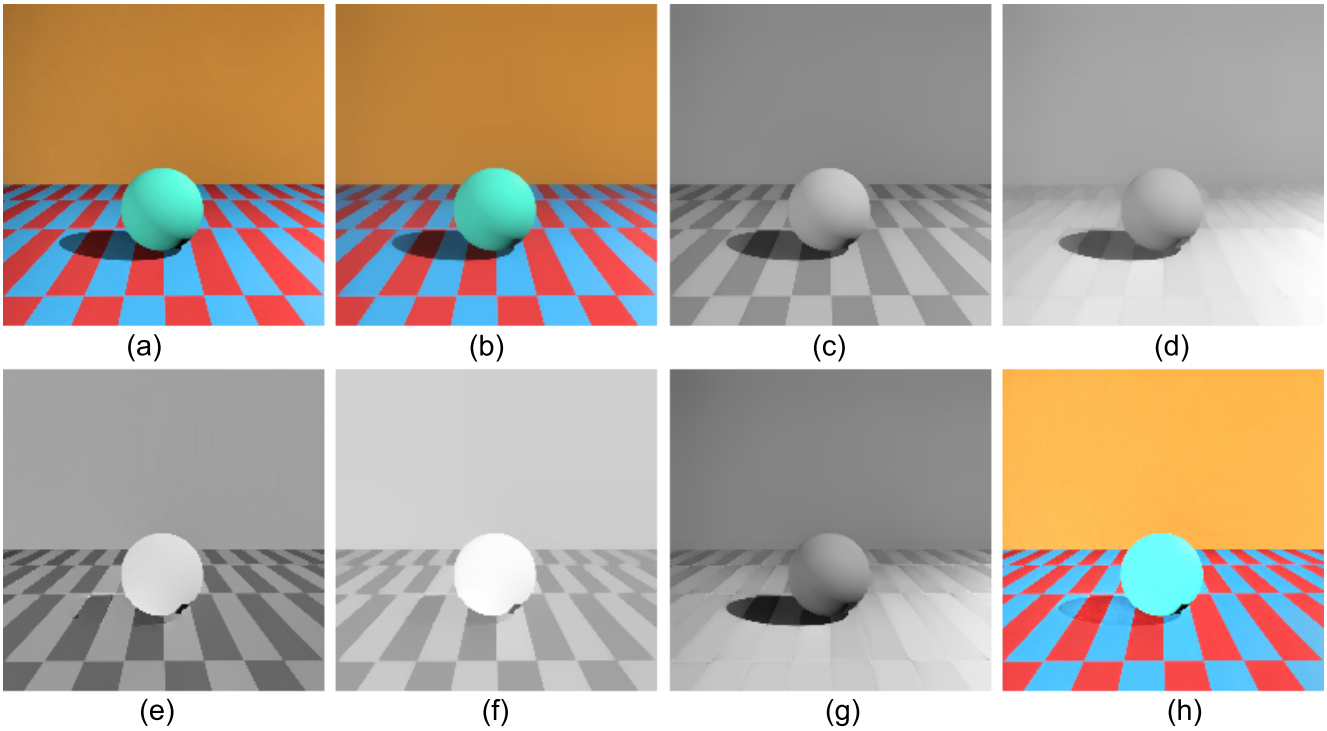


Figure 5. Complete pipeline with a simple scene [1]. The central view is shown here and the whole light field is shown in the Supplementary Material [36]. (a) Input light field L. (b) Filtered light field $\hat{L}$. (c) Normalized input $||\hat{L}||_2$. (d) Resulting shading S_1 from Equation 8; note 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_{g1} from Equation 8. (f) Filtered reflectance $\hat{R}_{g1}$. (g) Final shading S_f. (h) Final reflectance R_f.

its 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. Our method is efficient (a crucial aspect given the higher dimensionality of light fields), and our formulation still produces high quality results even in the absence of accurate depth information.

We have shown a straightforward editing example, but it would be interesting to see other applications enabled by our enhanced angular coherency, which will help in quick selection of objects or propagation of edits across the whole set of views.

For our albedo and occlusion cues, we currently rely on simple thresholds. A more sophisticated solution could make use of multidimensional Conditional Random Fields [19]. Despite the flexibility of our formulation with respect to depth data, a current limitation is that its qual-

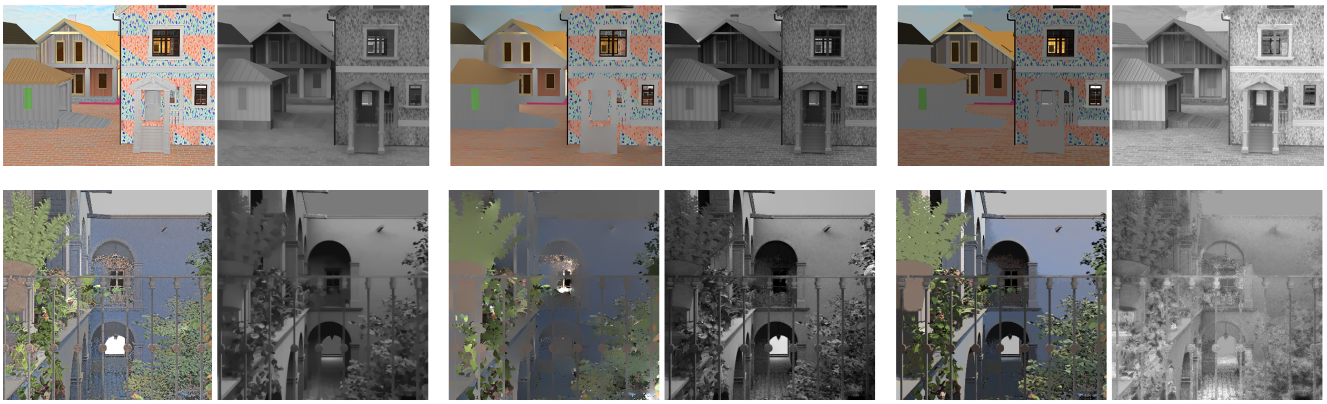


Figure 6. Comparison with previous work, using synthetic datasets [41]. (a) Our proposed method. (b) Zhao et al. [48]. (c) Bell et al. [6]. Single view differences can be appreciated in these representative frames, where all the results look similar. However, the critical aspect to evaluate in light fields is the angular coherency and the total processing time. Please refer to the videos in the Supplemental Material [36] for a visual comparison of the angular coherence in the solutions.

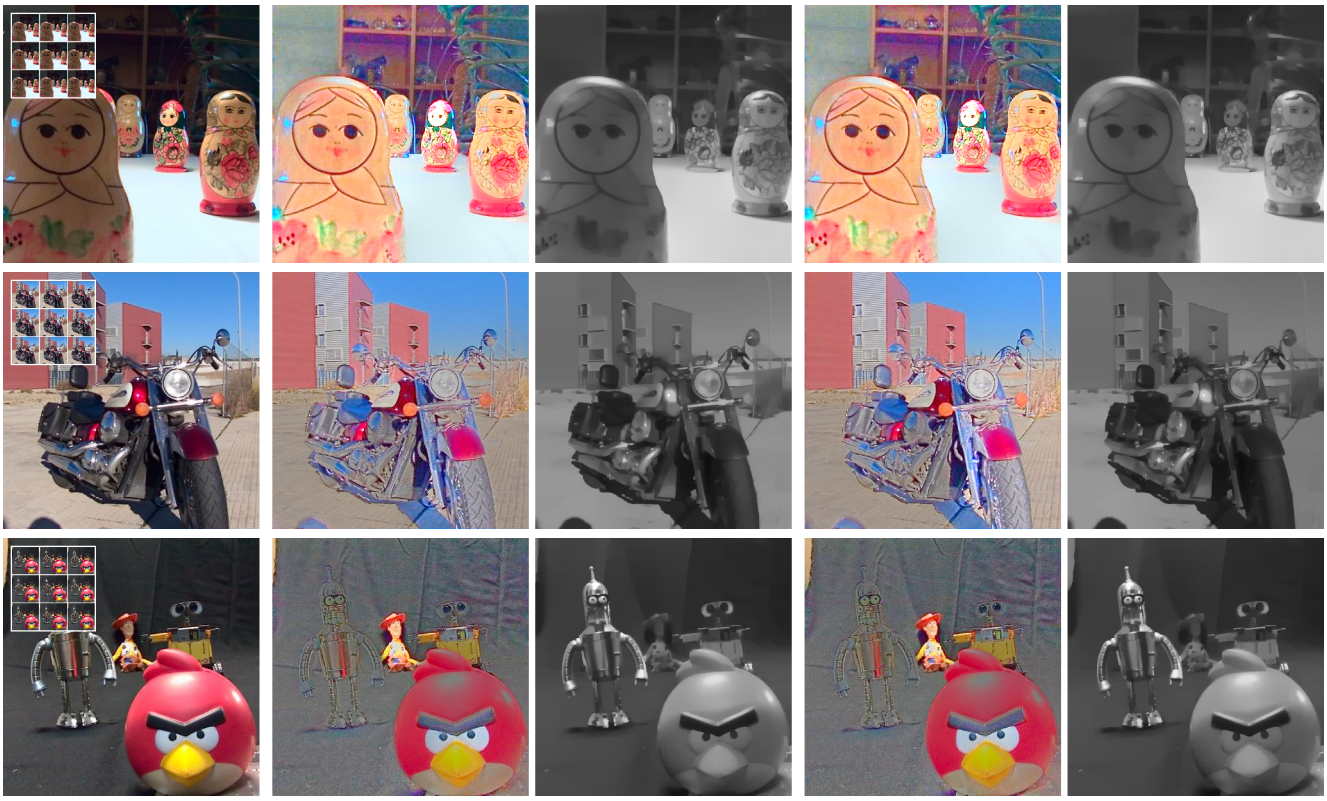
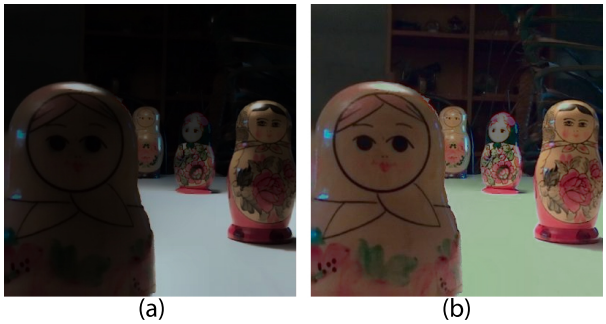


Figure 7. Results of our method on three real light fields taken with the Lytro^{TM} camera [20]. We show the decomposition for two opposite views. Please refer to the videos in the Supplementary Material [36].

ity 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). In this work we focused on coherency over subtleties in the single-view decompositions. However, we believe light field data



(a) (b)

Figure 8. Editing operations performed by modifying the shading (a) and the albedo (b) layers independently. Check the accompanying videos to see the complete edited light field.

captures rich radiometric scene information that will help lifting such limiting assumptions in the future.

References

[1] H. Ao, Y. Zhang, A. Jarabo, B. Masia, Y. Liu, D. Gutierrez, and Q. Dai. *Light Field Editing Based on Reparameterization*, pages 601–610. Springer International Publishing, Cham, 2015.

[2] N. Apostoloff and A. Fitzgibbon. Learning spatiotemporal t-junctions for occlusion detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 553–559 vol. 2, June 2005.

[3] J. Barron and J. Malik. Shape , Illumination , and Reflectance from Shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

[4] J. T. Barron and J. Malik. Intrinsic Scene Properties from a Single RGB-D Image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 17–24. Ieee, jun 2013.

[5] H. G. Barrow and J. M. Tenenbaum. Recovering intrinsic scene characteristics from images. In A. Hanson and E. Riseman, editors, *Computer Vision Systems*, New York, 1972.

[6] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics*, 33(4):1–12, 2014.

[7] S. Bi, X. Han, and Y. Yu. An L 1 Image Transform for Edge-Preserving Smoothing and Scene-Level Intrinsic Decomposition. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 34(4):1–12, 2015.

[8] C. Birklbauer, D. C. Schedl, and O. Bimber. Nonuniform spatial deformation of light fields by locally linear transformations. *ACM Trans. Graph.*, 35(5):156:1–156:12, June 2016.

[9] N. Bonneel, K. Sunkavalli, J. Tompkin, D. Sun, S. Paris, and H. Pfister. Interactive Intrinsic Video Editing. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 2014.

[10] A. Bousseau, S. Paris, and F. Durand. User-assisted intrinsic images. *ACM Trans. Graph. (SIGGRAPH Asia)*, 28(5):130, 2009.

[11] B. Chen, E. Ofek, H.-Y. Shum, and M. Levoy. Interactive deformation of light fields. In *Proceedings of the 2005 Symposium on Interactive 3D Graphics and Games*, I3D '05, pages 139–146, New York, NY, USA, 2005. ACM.

[12] Q. Chen and V. Koltun. A Simple Model for Intrinsic Image Decomposition with Depth Cues. In *IEEE International Conference on Computer Vision (ICCV)*, pages 241–248, 2013.

[13] D. Cho, S. Kim, and Y.-W. Tai. *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV*, chapter Consistent Matting for Light Field Images, pages 90–104. Springer International Publishing, Cham, 2014.

[14] S. Duchêne, C. Riant, G. Chaurasia, J. Lopez-Moreno, P.-y. Laffont, S. Popov, A. Bousseau, and G. Drettakis. Multi-View Intrinsic Images of Outdoors Scenes with an Application to Relighting. *ACM Transactions on Graphics*, 2015.

[15] E. Garces, A. Munoz, J. Lopez-Moreno, and D. Gutierrez. Intrinsic images by clustering. *Computer Graphics Forum (EGSR)*, 31(4):1415–1424, 2012.

[16] P. V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf. Recovering intrinsic images with a global sparsity prior on reflectance. In *Proc. NIPS*, page 765, 2011.

[17] X. Guo, Z. Yu, S. B. Kang, H. Lin, and J. Yu. Enhancing light fields through ray-space stitching. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2015.

[18] D. Hauagge, S. Wehrwein, P. Upchurch, K. Bala, and N. Snavely. Reasoning about Photo Collections using Models of Outdoor Illumination. In *British Machine Vision Conference*, pages 1–12, 2014.

[19] V. Jampani, M. Kiefel, and P. V. Gehler. Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[20] A. Jarabo, B. Masia, A. Bousseau, F. Pellacini, and D. Gutierrez. How do people edit light fields? *ACM Transactions on Graphics (SIGGRAPH 2014)*, 33(4), 2014.

[21] A. Jarabo, B. Masia, and D. Gutierrez. Efficient propagation of light field edits. In *Proceedings of SIACG '11*, pages 75–80, 2011.

[22] N. Kong, P. V. Gehler, and M. J. Black. Intrinsic Video. In *European Conference in Computer Vision (ECCV)*, 2014.

[23] P. Laffont, A. Bousseau, and S. Paris. Coherent intrinsic images from photo collections. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2012.

[24] P.-y. Laffont and J.-C. Bazin. Intrinsic Decomposition of Image Sequences from Local Temporal Variations. *International Journal of Computer Vision*, 2015.

[25] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1), 1971.

[26] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin. Estimation of intrinsic image sequences from image + depth video. In *Proc. ECCV*, pages 327–340. Springer, 2012.

[27] Lytro Inc. The Lytro camera. `http://www.lytro.com`, 2013.

[28] B. Masia, A. Jarabo, and D. Gutierrez. Favored workflows in light field editing. In *Proceedings of CGVCVIP '14*, 2014.

[29] A. Meka, M. Zollhöfer, C. Richardt, and C. Theobalt. Live intrinsic video. *ACM Transactions on Graphics (Proceedings SIGGRAPH)*, 35(4), 2016.

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

[31] I. Omer and M. Werman. Color lines: Image specific color representation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2:946–953, 2004.

[32] Raytrix GmbH. 3D light field camera technology. `http://www.raytrix.de`, 2013.

[33] S. M. Seitz and K. N. Kutulakos. Plenoptic image editing. *International Journal of Computer Vision*, 48(2):115–129, 2002.

[34] L. Shen, C. Yeo, and B.-s. Hua. Intrinsic Image Decomposition Using a Sparse Representation of Reflectance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2904–2915, 2013.

[35] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz. Factored time-lapse video. *ACM Trans. Graph. (SIGGRAPH)*, 26(3):101, 2007.

[36] Supplementary. Intrinsic light fields - supplementary material. `http://webdiis.unizar.es/~elenag/projects/intrinsicLF/supplementary/supplementary.html`. Accessed: 2016-08-08.

[37] M. Tao, J.-C. Su, T.-c. Wang, J. Malik, and R. Ramamoorthi. Depth Estimation and Specular Removal for Glossy Surfaces Using Point and Line Consistency with Light-Field Cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (August):1–14, 2015.

[38] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ICCV '13, pages 673–680, Washington, DC, USA, 2013. IEEE Computer Society.

[39] M. Tappen, W. Freeman, and E. Adelson. Recovering intrinsic images from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(9):1459–1472, 2005.

[40] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar. Picam: An ultra-thin high performance monolithic camera array. *ACM Trans. Graph.*, 32(6):166:1–166:13, 2013.

[41] T.-c. Wang, A. A. Efros, and R. Ramamoorthi. Occlusion-aware Depth Estimation Using Light-field Cameras. *International Conference on Computer Vision*, 2015.

[42] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(3):606–619, Mar. 2014.

[43] Y. Weiss. Deriving intrinsic images from image sequences. In *Proc. ICCV*, volume 2, pages 68–75. IEEE, 2001.

[44] S. Yang, J. Wang, W. Fan, X. Zhang, P. Wonka, and J. Ye. An efficient admm algorithm for multidimensional anisotropic total variation regularization problems. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 641–649, New York, NY, USA, 2013. ACM.

[45] G. Ye, E. Garces, Y. Liu, Q. Dai, and D. Gutierrez. Intrinsic video and applications. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 33(4):1–11, 2014.

[46] F. L. Zhang, J. Wang, E. Shechtman, Z. Y. Zhou, J. X. Shi, and S. M. Hu. Plenopatch: Patch-based plenoptic image manipulation. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2016.

[47] Z. Zhang, L. Wang, B. Guo, and H.-Y. Shum. Feature-based light field morphing. *ACM Trans. Graph.*, 21(3):457–464, July 2002.

[48] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin. A closed-form solution to retinex with nonlocal texture constraints. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(7):1437–1444, 2012.

[49] T. Zhou, P. Krähenbühl, and A. A. Efros. Learning Data-driven Reflectance Priors for Intrinsic Image Decomposition. *International Conference on Computer Vision (ICCV)*, 2015.