View-dependent Scene Appearance Synthesis using Inverse Rendering from Light Fields

Dahyun Kang, Daniel S. Jeon, Hakyeong Kim, Hyeonjoong Jang, and Min H. Kim

Abstract—In order to enable view-dependent appearance synthesis from the light fields of a scene, it is critical to evaluate the geometric relationships between light and view over surfaces in the scene with high accuracy. Perfect diffuse reflectance is commonly assumed to estimate geometry from light fields via multiview stereo. However, this diffuse surface assumption is invalid with real-world objects. Geometry estimated from light fields is severely degraded over specular surfaces. Additional scene-scale 3D scanning based on active illumination could provide reliable geometry, but it is sparse and thus still insufficient to calculate view-dependent appearance, such as specular reflection, in geometry-based view synthesis. In this work, we present a practical solution of inverse rendering to enable view-dependent appearance synthesis, particularly of scene scale. We enhance the scene geometry by eliminating the specular component, thus enforcing photometric consistency. We then estimate spatially-varying parameters of diffuse, specular, and normal components from wide-baseline light fields. To validate our method, we built a wide-baseline light field imaging prototype that consists of 32 machine vision cameras with fisheye lenses of 185 degrees that cover the forward hemispherical appearance of scenes. We captured various indoor scenes, and results validate that our method can estimate scene geometry and reflectance parameters with high accuracy, enabling view-dependent appearance synthesis at scene scale with high fidelity, i.e., specular reflection changes according to a virtual viewpoint.

Index Terms—Light field, view synthesis, inverse rendering

1 Introduction

Light fields have been used broadly to capture dense depth maps [1], create novel view images [2], refocus depth of field [3], capture 3D contents for holographic displays [4], etc. These applications are founded on the ground of *multiview geometry* from light fields. They commonly assume that object surfaces are perfectly *diffuse*, i.e., when an object is observed from a different view, only its geometric shape changes, not affecting its appearance. However, this assumption does not hold with real-world objects that have mixtures of diffuse and specular reflectance. When obtaining scene geometry from light fields, view-dependent appearance, such as *specular reflection*, has caused the regional failure of geometry estimation because stereo correspondence search fails in multiview geometry.

From the perspective of view synthesis, specular reflection is one of the most critical view-dependent appearance phenomena to achieve high-fidelity realism. It changes appearance depending on directions of light, view, and surface normals. To simulate view-dependent appearance changes, it is critical to evaluate the geometric relationships among these directions with high accuracy. Accurate 3D geometry of scenes is necessary for high-quality view-dependent appearance synthesis. For object-scale light fields, traditional view synthesis of the bidirectional appearance of light and view has been achieved by employing an additional process of 3D scanning [5], [6] that can capture polygonal surfaces of the object geometry. For scene-scale light fields, 3D scanning based on active illumination can provide sparse geometry of point clouds. However, it is too sparse to compute view-dependent appearance parameters, and also lacks the surface normal information in each light field.

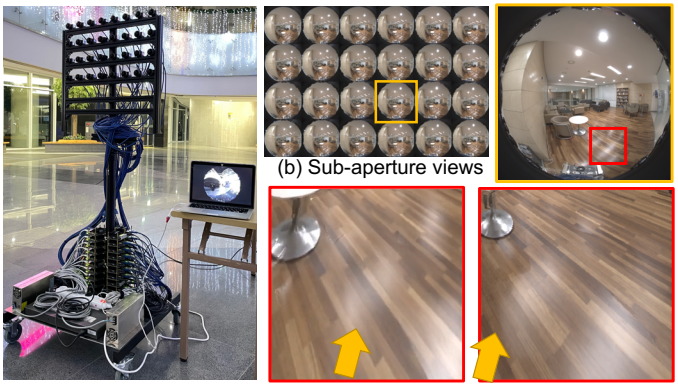


Fig. 1: (a) Our prototype of wide-baseline light field imaging for capturing scene-scale light fields. (b) Captured light-field images and a sub-aperture view image. (c) Results of our view-dependent appearance synthesis. Our method successfully simulates appearance changes of specular reflection across different views.

To achieve view-dependent appearance synthesis from light fields at scene scale without relying on additional 3D scanning, there are several technical challenges that need to be solved. First, when object surfaces are smooth and specular, geometry information from light fields is inaccurate. Thus, it is hard to estimate specular parameters appropriately by evaluating geometric relationships among lights, views, and normals in scenes. Second, without successful prediction of specular components, the traditional shape-from-shading (SfS) approach cannot refine scene geometry

• *D. Kang, D.S. Jeon, H. Kim, H. Jang, M.H. Kim are with the School of Computing, KAIST, South Korea, 34141.*
E-mail of the corresponding author: minhkim@kaist.ac.kr

due to insufficient photometric cues. Lastly, view synthesis on specular regions produces unrealistic artifacts due to inaccurate shape and appearance parameters.

To mitigate these challenges, we present a novel scene-scale appearance synthesis method based on inverse rendering from wide-baseline light fields. We first estimate the initial geometry from wide-baseline light fields captured by fisheye cameras using an optical flow method. The initial geometry is enhanced using diffuse-specular separation and enforcing photometric consistency. We then approximate illumination in the front hemisphere as a set of point lights by means of the captured light fields and the estimated scene geometry. Given the scene geometry and the approximated illumination, we formulate an inverse rendering problem that jointly optimizes spatially-varying bidirectional reflectance distribution function (SVBRDF) parameters including diffuse and specular reflectance.

To validate the proposed method, we built a prototype of a wide-baseline light field imaging system. Our system consists of 32 machine vision cameras, each equipped with a 185-degree fisheye lens and an embedded computer. Cameras are synchronized to capture light fields at the same time. We captured various indoor scenes that include diffuse and specular surfaces on various geometry. We quantitatively and qualitatively evaluate the accuracy of our view synthesis and depth estimation results on real-world and synthetic scenes. Results validate that our method can capture scene geometry and reflectance parameters with high accuracy from wide-baseline light fields. The simulated view-dependent appearance of specular reflection shows a good agreement with that of the real-world scene.

2 Related Work

2.1 View Synthesis from Light Fields

Geometry-based View Synthesis. Synthesizing a novel view from multiple views has been studied extensively in recent decades. Hedman et al. [2], [7], [8] reconstruct a mesh from the estimated camera poses and depth maps. They then stitch input multiple images and depth maps into a single seamless panorama and convert them into a triangle mesh. However, the mesh and the synthesized images' quality depend on the accuracy of the estimated depth maps. Cho et al. [9] synthesize novel view images by estimating camera poses and reconstructing a mesh from multiple 360° images. They render a novel view image by reprojecting each pixel point to a mesh and the closest 360° image and sample the color from it. They reconstruct the 3D mesh geometry using structure from motion (SfM), which also suffers from inaccurate geometry over specular surfaces, resulting in an incorrect shape of the mesh. Luo et al. [10] render free-viewpoint images from multiple views on a spherical grid, but they synthesize novel view images without reconstructing a 3D mesh. Instead, they precompute inter-view motion fields and use them to blend colors from neighboring input images. However, the accuracy of the estimated optical flow still depends on the diffuse assumption, i.e., photometric consistency of scenes in multiple views. In traditional object-scale view synthesis methods [5], [6], an extra 3D scanning process has been used to capture 3D models, which allows for accurate estimations of geometric relationships among light, view, and surface normal directions. However, for scene-scale view synthesis, 3D scanning only obtains partial scene geometry near surfaces, or overly sparse as point clouds. These geometry-based methods do not account for view-dependent appearance changes, such as specular reflection. In contrast, our method estimates view-dependent parameters of surface reflectance, in addition to surface normals at scene scale.

Image-based View Synthesis. Since the representation of layered depth images was proposed by Shade et al. [11], depth image-based rendering has been popularly used for view synthesis of light fields. Flynn et al. [12] employ a plane sweeping volume between neighboring stereo images as an input to the network and blend color values from different depth planes with corresponding weight values. Similarly, Zhou et al. [13] use plane sweeping volumes from neighbor images as inputs of the neural network to output multi-plane images (MPIs), which consist of color and alpha images for every depth plane. Then, they synthesize novel view images based on the MPI representation. Subsequent studies, such as [14], [15], have improved the network architectures. Also, there are other researches [16], [17] that encode depth probability in a multi-layer representation and utilize them to improve the quality of synthesized images. Mildenhall et al. [18] employ a set of multiple MPIs from each local input view and improve view synthesis results of non-Lambertian surfaces. These methods can roughly synthesize reflections of specular materials. Still, since they are not based on the reflection model, their view synthesis results cannot cover the complete angular resolution. They cannot add other light sources and render reflections that are not originally included in the input. Broxton et al. [19] employ multiple cameras with large field-of-view (FoV) of fisheye lenses to synthesize wide FoV novel view images. They extend MPI to multi-sphere image (MSI). They convert the MSI to a layered mesh to efficiently render novel views, but still, they require a lot of computational resources because of the large amount of data. Wu et al. [20] and Wang et al. [21] make use of EPI upsampling for view synthesis from light fields. These works are especially effective for narrow baseline light fields, in which disparity ranges up to about five pixels, while wide-baseline light fields inherently suffer from extreme discontinuity in EPI. Wu et al. [20] also propose a method for larger disparities by shearing EPI with given disparity; however, estimating disparity in non-diffuse regions is a long-lasting hard problem in the literature. Compared to previous works above, our approach estimates the appearance parameters of a reflectance model through inverse rendering and uses them to refine 3D geometry to be more robust on even extreme scenes with strong light sources. Also, acquiring the reflectance parameters leads us to render more realistic specular surfaces with even more light sources, which are not originally captured in the scene.

2.2 Shape and Reflectance from Light Fields

To capture shape and reflectance more robustly from light fields, many works analyze the characteristics of specular

reflection captured at different viewpoints of light fields. In addition to the traditional photo-consistency, Tao et al. [22] devise a line-consistency method in color space derived from dielectric material property. However, this approach is highly dependent on the acquired color vector of specular reflection, which can vary in large amounts for non-dielectric material, such as metal. Also, chromatic aberration could affect the line-consistency assumption due to the exponentially high intensity of pixels with specular peaks.

For short-baseline light fields, Wang et al. [23] employ a differential approach in optimization to reconstruct the shape and SVBRDF of objects with generalized reflectance. A BRDF-invariant equation is derived by jointly formulating impacts of the view change and the spatial change. Their BRDF-invariance approach eliminates the view-dependent property and relates depths and normals only. Following their work, Li et al. [24] proposed a robust energy minimization method achieving a lower error rate. However, their short-baseline light field is inherently limited to tiny objects and a narrow range of acquired viewing direction. Also, their differential approach is hard to be extended to a wide-baseline light field at scene scale due to its extreme disparity ranging to hundreds of pixels and drastic appearance change.

More recently, a probabilistic framework for joint estimation of shape and depth map under natural illumination is proposed by Ngo et al. [25]. They relax the previous works' laboratory environment restriction to a known natural illumination. They iteratively ease the Lambertian assumption by refining previously estimated depth, normal, and reflectance. Assuming homogeneous material objects, their algorithm exploits more combinations of incident lights. However, in the real world, the light sources usually are not so distant that the light rays emitted from the same light source incident at different angles at different object surfaces.

3 View-dependent Scene Appearance Synthesis

Overview. We first estimate depth from wide-baseline high-dynamic-range light fields at the reference viewpoint of the central camera. We then estimate scene illumination of the frontal scene captured by fisheye lenses. Given initial geometry and illumination, we estimate appearance parameters of diffuse albedos, specular albedos, surface smoothness parameters and also refine surface normals jointly through inverse rendering optimization. Using the appearance parameters, geometry, and illumination, we synthesize appearance changes from a novel viewpoint. See Figures 1(b) and (c) for an example. Figure 2 provides an overview of our algorithm workflow.

Acquisition Setup. We built a wide-baseline light-field imaging system with 32 machine vision cameras on a grid of 4×8, FLIR Blackfly (BFS-U3-120S4C-CS, 12 MP, 31 fps). Each camera is equipped with a fish-eye lens with 185-degrees FOV that covers the forward hemispherical environment, with a resolution of 3000×3000. In order to enable concurrent capture of these machine vision cameras, each camera is connected to an embedded system, Nvidia Jetson Nano.

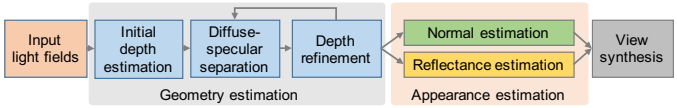


Fig. 2: Overview. We first estimate depth from input light fields and then decompose them into diffuse and specular reflection components. Then, we jointly optimize normals and reflectance parameters, enabling view-dependent appearance synthesis of scenes.

Each sub-system is connected to a local network switch with its local static IP address. We employ a controller machine that broadcasts the capture-instruction packet so that each camera system can capture a scene simultaneously with an average timing error below $\pm 1/30$ seconds measured with a 60 Hz counter. Also, the control machine manages captured image file transfer, controls the power of each client, and shares camera parameters including gain, shutter speed, and gamma-correction, to make sure consistent radiometric parameters of each camera. To estimate scene illumination, we capture high-dynamic-range (HDR) light-field images. Each camera captures five multiple exposures by controlling the shutter time with one-stop intervals to capture an HDR image.

The baseline between horizontal and vertical adjacent cameras are 8 cm and 10 cm respectively to ensure large enough disparity. Intrinsic and extrinsic parameters such as positions, rotations, focal lengths, and lens distortions are calibrated per camera and scene using the feature matching-based calibration method by Pozo et al. [2]. See Figure 1(a) for our prototype of light-field imaging.

3.1 Specularity-aware Geometry Estimation

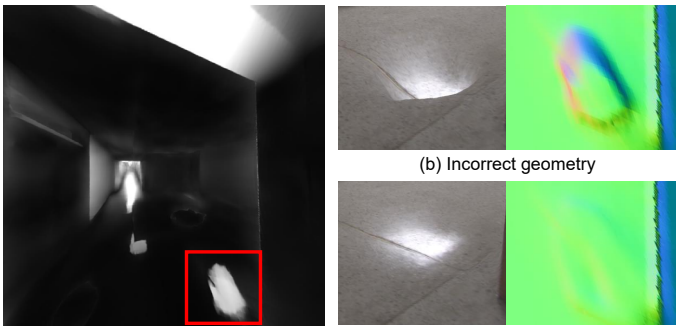
Our depth estimation pipeline consists of two main stages. We first estimate the initial depth under the diffuse assumption at the center camera view and then refine it by filling invalid depth values caused by strong specular reflection.

Initial Depth. We estimate the initial depth map Z_0 for a center camera under diffuse reflection assumption. Theoretically, a pair of cameras should be enough for disparity estimation. However, we utilize all 32 camera views with the purpose of estimating specular reflection and dense geometry with high accuracy.

We first employ a learning-based optical flow algorithm [26] to estimate pairwise optical flow maps $f_{0\to i}$ from the center camera (denoted by 0) to the i-th camera. Note that this initial depth may include inaccurate depth values in specular regions due to photometric inconsistency. 32 disparity maps are then integrated to a depth z_0 at a pixel (u_0, v_0) of the center camera, yielding a depth map Z_0. It is calculated as 3D points $\hat{\mathbf{q}}$ with respect to the center camera, using the least-squares that minimizes as follows:

$$\hat{\mathbf{q}} = \arg\min_{\mathbf{q}} \sum\nolimits_i ||\mathbf{q} - \Pi_i(u_i, v_i, z_i)||^2 , \qquad (1)$$

where pixel position $(u_i, v_i) = f_{0\to i}(u_0, v_0)$ is obtained by the optical flow map, and $\Pi_i(\cdot)$ is a backprojection function of the i-th camera that yields a 3D point in the world coordinates from a given depth z_i at a pixel (u_i, v_i).



(a) A depth disagreement map (b) Incorrect geometry (c) Corrected geometry

Fig. 3: Depth values of the pixels with the significant values in the depth disagreement map are corrected by our depth correction method.

By minimizing the loss, the i-th camera's depth value, at which the light ray starts from the origin of the i-th camera and passes through pixel (u_i, v_i), becomes closest to 3D point $\mathbf{q}$ in the world coordinates. We compute $\hat{\mathbf{q}}$ by minimizing the sum of the squared differences between two directions $\bar{\mathbf{r}}_{\mathbf{q}}$ and $\bar{\mathbf{r}}_i$: $||\bar{\mathbf{r}}_{\mathbf{q}} - (\bar{\mathbf{r}}_i \cdot \bar{\mathbf{r}}_{\mathbf{q}})\,\bar{\mathbf{r}}_i||$ in a matrix form. Here, $\bar{\mathbf{r}}_{\mathbf{q}} = \mathbf{q} - \mathbf{c}_i$ is a direction of $\mathbf{q}$ from i-th camera's origin, and $\bar{\mathbf{r}}_i = \frac{\Pi_i(u_i, v_i, 1) - \mathbf{c}_i}{||\Pi_i(u_i, v_i, 1) - \mathbf{c}_i||}$ is projected $\mathbf{q}$ on i-th camera's ray vector passing through a pixel (u_i, v_i).

Finally, we perform bilateral filtering to mitigate blocky artifacts in the initial depth map Z_0 caused by the large-scale upsampling at the end of Teed et al. [26]'s flow estimation pipeline.

Depth Refinement on Specular Reflection. We then refine the initial depth map Z_0 in a specular-aware manner. The main intuition of our depth refinement algorithm is that the areas, where specular reflection is dominant, present a lower agreement of the optimized depth map among 32 views due to miscalculated stereo correspondence. To this end, we define the measure of depth disagreement using the residual of the least-squares problem (Equation (1)). Figure 3(a) shows a depth disagreement map, in which bright pixels indicate the high probability of specular component existence. For each pixel (u_0, v_0) with a higher disagreement value than the threshold, denoted by a mask of 'depth-holes' H_0, the depth value is re-calculated by selectively adopting much more reliable observations only.

We propose a diffuse-reflection separation method to distinguish specular reflection against diffuse reflection using the depth agreements in wide-baseline light fields. As a specular reflection acts as a positive addition to diffuse reflection $I_i = D + S_i$, common diffuse reflection D at the center camera can be approximated by the darkest pixel value among sub-aperture views:

$$D(u_0, v_0) = \min_i I_i\left(\hat{u}_i, \hat{v}_i\right),$$

where $(\hat{u}_i, \hat{v}_i) \leftarrow \Pi_i^{-1}\left(\Pi_0(u_0, v_0, z_0)\right)$ is the pixel coordinates of a 3D world point $\Pi_0(u_0, v_0, z_0)$ reprojected to the i-th camera. Then specular reflection S_i observed at camera i and warped onto the center camera is defined as:

$$S_i\left(u_0, v_0\right) = I_i\left(\hat{u}_i, \hat{v}_i\right) - D(u_0, v_0). \tag{2}$$

Figure 5 shows an example of the separated D and S_i.

In order to estimate depth on specular reflection, we define the set of pixels as masks H_i, whose S_i values are higher than a threshold. Since pixels in H_i tend to produce an inaccurate depth value due to their specular reflection, they are excluded and marked as depth holes. After that, depth holes are filled by integrating local depth maps $\bar{Z}_i$ in a specular-aware manner. Local depth map $\bar{Z}_i$ is computed from four local optical flow maps $f_{i\to j}$ similarly to Equation (1), where $j \in N(i)$ and $N(i)$ is a set of the four nearest cameras to the i-th camera. Note that we use specular-free images as an input of the optical flow algorithm with an expectation of that diffuse objects will be less affected by the violation of photometric consistency.

For the purpose of reducing computational cost, $\bar{Z}_i$ are computed after subsampled by four and inside of the mask $H_0 \backslash H_i$. The values of Z_0 in the depth holes H_0 are then refined to the mean of $\bar{Z}_i$. Figure 3 compares the initial depth Z_0 and the iteratively refined depth Z_0. Algorithm 4 summarizes our depth estimation algorithm.

Algorithm 1 Specular-aware Depth Estimation

Input: light field images $\mathcal{I} = I_{1,\cdots,n}$
Output: specular-aware depth map at center camera Z

```
Z_0, H_0 ← InitialDepthEstimate(I)
for m iterations do
    D ← SeparateDiffuse(I)
    S_{1,...,n} ← I_{1,...,n} − D
    H_{1,...,n} ← SpecularHole(I, S_i)
    Z̄_{1,...,n} ← LocalDepthEstimate(I, H_0 \ H_i)
    Z_0 ← Mean(Z̄_{1,...,n})
end for
```

3.2 Inverse Rendering from Light Fields

Reflectance Model. We employ the Blinn-Phong model [27] to encode the view-dependent appearance of the scene. Given incident light direction $\omega_{\mathbf{i}}$, view direction $\omega_{\mathbf{o}}$, and surface normal $\mathbf{n}$, the ratio of reflected light follows:

$$R(\mathbf{n}, \omega_{\mathbf{i}}, \omega_{\mathbf{o}}) = \rho_{\mathrm{d}}\langle \mathbf{n}, \omega_{\mathbf{i}} \rangle + \rho_{\mathrm{s}}\langle \mathbf{n}, \mathbf{h} \rangle^{\alpha}, \tag{3}$$

where R is the reflected light, ρ_d and ρ_s are diffuse and specular albedo respectively, $\mathbf{h}$ is a half-vector of $\omega_{\mathbf{i}}$ and $\omega_{\mathbf{o}}$, α is specular smoothness, and $\langle , \rangle$ denotes positive dot product.

Given the geometry and the reflection ratio R, image formation from i-th camera can be modeled by the discretized rendering equation:

$$I_i(\hat{u}_i, \hat{v}_i) = \sum\nolimits_{l=1}^{\Lambda} R(\mathbf{n}_{\mathbf{p}}, \omega_{\mathbf{l}_l \to \mathbf{p}}, \omega_{\mathbf{p} \to \mathbf{c}_i}) \frac{L_l}{d^2},$$

where Λ is the number of point lights, $(\hat{u}_i, \hat{v}_i) = \Pi_i^{-1}\left(\Pi_0(u_0, v_0, z_0)\right)$ is a pixel coordinate in the i-th camera, $\mathbf{p} = \Pi_0(u_0, v_0, z_0)$ is an unprojected 3D point from (u_0, v_0) at the center camera, $d = ||\mathbf{l}_l - \mathbf{p}||^2$ is the attenuation factor by distance, and L_l is the emitting radiance from the l-th lights.

Lighting. Typical approaches in lighting using an environmental map assume the light sources are placed at infinite so that lighting applied to each scene point remains constant. However, in our case, the infinite-distant light source assumption is inappropriate. View-dependent appearance

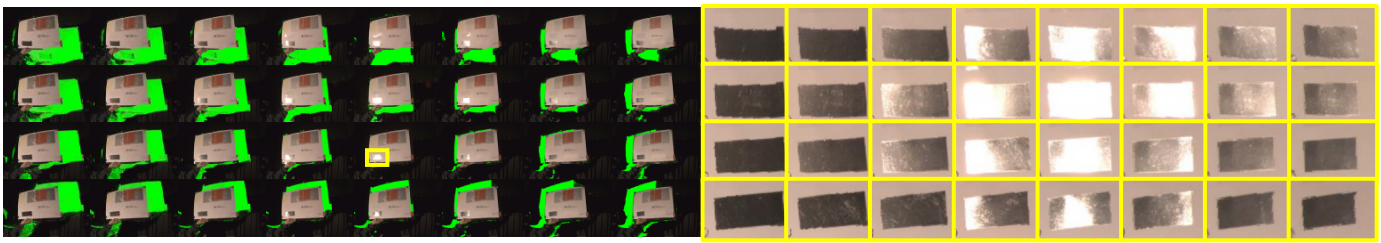


Fig. 4: An example of view-dependent appearance in light fields. A flat surface with different smoothness materials: diffuse color paper, tin-foil tape, and diffuse stickers. Sub-aperture images are warped onto the center view (green mask indicates occlusion). The silver-colored aluminum shows a strong specular reflection in the camera view at (2, 5) on the grid while it becomes significantly darker in other views. The right images show closeups of the aluminum sticker.

(a) Original image at the center camera (b) Diffuse image (c) Specular image at the center camera

Fig. 5: An example of diffuse reflection D and specular reflection at the center view.

does not depend only on view directions but also the incident light angle, which varies depending on scene geometry.

We, therefore, model our lights as a set of point lights described as 3D coordinate $\{\mathbf{l}_l\}$. We assume our light source has the following characteristics:

- Each light source is visible to the center camera,
- Each light source emits the same radiance and is not directional, and
- The light source that significantly influences specular reflection on other surfaces in the scene is clamped to a constant level.

We approximate illumination of area light as a group of the subsampled point light sources. An input 185-degree fisheye image with the fastest shutter speed (8 milliseconds) is downscaled to the 1/10 resolution of the original resolution. Then, the locations of pixels saturated at the lowest exposure image are counted as the positions of point light sources.

Specular Parameters Estimation. Given the center camera's depth map Z and the set of sampled point lights, we estimate ρ_d, ρ_s, α, and $\mathbf{n}$. Ideally, the view-dependent appearance of materials can be achieved by solving an inverse rendering problem, which finds all those parameters describing the reflectance model per pixel. However, it is a significantly ill-posed problem especially if exact incident lights are unknown. Instead, we restrict our problem to finding the parameters, which directly influence the appearance rendered at novel views. As the diffuse reflection term in Equation (3) is independent of the view direction $\omega_\mathbf{o}$, the value of $\rho_d\langle\mathbf{n}\cdot\omega_i\rangle$ remains the same in the observed views and any synthesized views. Thus, our interests are narrowed to specular parameters ρ_s and α. The geometric normals obtained from Z is inaccurate yet to solve the inverse problem. Materials with some degree of shininess such as plastics have a high α value ranging in a log scale and get closer to perfect mirror reflection for shinier materials. Thus calculating accurate $\mathbf{n}$ is critical for inverse rendering.

Our specular parameters estimation algorithm consists of two stages. First, ρ_s, α, and $\mathbf{n}$ are optimized per a screen-space cluster in order to make use of many observations. Those parameters are then optimized together, expecting appropriate propagation.

In the first stage, the diffuse image of the center view D is clustered into $\mathcal{K}_{1,\cdots,K}$ according to colors, pixel coordinates, and normals by using the K-means clustering algorithm. All pixels in the same cluster are assumed to have the same ρ_s and α, but variable $\mathbf{n}$. We find ρ_s, α, and $\mathbf{n}$ by minimizing the following loss function:

$$\mathcal{L}_\mathrm{total} = \lambda_\mathrm{render}\mathcal{L}_\mathrm{render} + \lambda_\mathrm{smooth}\mathcal{L}_\mathrm{smooth}.$$

We take advantage of our diffuse-specular separation method (Equation (2)) to suppress the interference between variables in the optimization process. Our rendering loss is sum of diffuse and specular loss:

$$\mathcal{L}_\mathrm{render} = \mathcal{L}_\mathrm{diffuse} + \mathcal{L}_\mathrm{specular},$$

where our diffuse loss is defined as:

$$\mathcal{L}_\mathrm{diffuse} = \sum_{(u_0,v_0)\in\mathcal{K}_k}\left(D\left(u_0,v_0\right) - \sum_l \rho_\mathrm{d}\langle\mathbf{n}(u_0,v_0),\hat{\mathbf{l}}_l\rangle\frac{L}{d^2}\right)^2$$

where $\hat{\mathbf{l}}_l = \frac{\mathbf{l}_l-\mathbf{p}}{||\mathbf{l}_l-\mathbf{p}||}$ is the directional vector to light $\mathbf{l}_l$ from point $\mathbf{p}$. Note that the diffuse reflectance term influences the rendered appearance more drastically and widely compared to the specular term. Thus, eliminating the diffuse variables from our rendering loss effectively reduces the search space.

Our specular loss is defined as:

$$\mathcal{L}_\mathrm{specular} = \sum_i\sum_{(u_0,v_0)\in\mathcal{K}_k}\left(S_i\left(u_0,v_0\right) - \sum_l \rho_\mathrm{s}\langle\mathbf{n}(u_0,v_0),\mathbf{h}\rangle^\alpha\frac{L}{d^2}\right)^2.$$

A challenge to optimize the specular loss is in the linear multiplication term of $\rho_s L$, which disturbs the other parameters $\mathbf{n}$ and α to take individual gradient movement. Rather than optimizing ρ_s together with other variables, approximating ρ_s by a linear fit helps the other parameters to be converged more robustly to the wrong initial ρ_s:

$$\rho_s = \frac{\sum_i\sum_{(u_0,v_0)\in\mathcal{K}_k}\left(S_i\left(u_0,v_0\right)\sum_l\left(\langle\mathbf{n}(u_0,v_0),\mathbf{h}\rangle^\alpha\frac{L}{d^2}\right)\right)}{\sum_i\sum_{(u_0,v_0)\in\mathcal{K}_k}\left(\sum_l\langle\mathbf{n}(u_0,v_0),\mathbf{h}\rangle^\alpha\frac{L}{d^2}\right)^2}. \tag{4}$$

We define normal smoothness loss that enforces local smoothness of normal directions:

$$\mathcal{L}_{\text{smooth}} = \sum\nolimits_{(u_0,v_0)\in\mathcal{K}_k} \left(\left(\frac{\partial \mathbf{n}}{\partial u}(u_0,v_0) \right)^2 + \left(\frac{\partial \mathbf{n}}{\partial v}(u_0,v_0) \right)^2 \right).$$

The optimization is performed over ∼500 iterations using the Adam optimizer and mesh rendering pipeline [28]. Once ρ_s and α per clusters and $\mathbf{n}$ per pixels of the clusters are estimated, boundaries are regularized by filtering to ensure a continuous transition.

3.3 View-dependent Scene Rendering

Rendering at a novel view is straightforward given depth map Z, normal map $\mathbf{n}$, specular albedo map ρ_s, specular smoothness map α, diffuse image D, and point light sources. The depth map Z is used to generate a simple quad-mesh whose vertices are the pixels visible in the center view. This quad-mesh enables interpolation of the vertex colors and the normals. Pixel values of a diffuse image D are assigned to each vertex as diffuse color. Vertex normals are perturbed with our optimized normal map $\mathbf{n}$. Plugging specular parameters ρ_s and α to each vertex, a novel view of the mesh can be rendered under known light sources.

As a drawback of the mesh-based rendering at the moved position, the stretched object boundary artifacts appear. Elongated triangles over depth edges need to be removed [2]. Those vertices at which the angle between the viewing direction and the geometric normal is over a threshold are filled by the color of neighboring background surfaces.

4 Results

We qualitatively and quantitatively evaluate our view synthesis and depth acquisition results captured by our real prototype (Figure 1(a)), also comparing them with results by other state-of-the-art methods. In addition, we validate the accuracy of the results using a synthetically rendered dataset as ground truth.

4.1 Geometry Estimation

In order to quantitatively evaluate the accuracy of our geometry estimation from light fields, we synthetically rendered a scene with our light-field camera's configuration to obtain the ground-truth depth information. Fisheye images of a 185-degree field of view are rendered at 32 positions with a displacement of 10 cm and 8 cm on a sub-aperture grid of 4×8 to capture wide-baseline light fields. We compare our depth estimation results with those of two other state-of-the-art methods [2], [26] as shown in Figure 6. Pozo et al. [2], [26]'s method relies on patch-wise similarity to search correspondences from light fields, assuming that only diffuse surfaces exist in the scene. However, when surfaces present strong specular reflection, their assumption fails, resulting in inaccurate depth estimation. In contrast, our specular-aware depth estimation algorithm estimates depth values over specular surfaces by integrating depth information of different sub-aperture views locally estimated from wide-baseline light fields. Also, Figure 7 compares qualitative results of the estimated depths and normals of two real

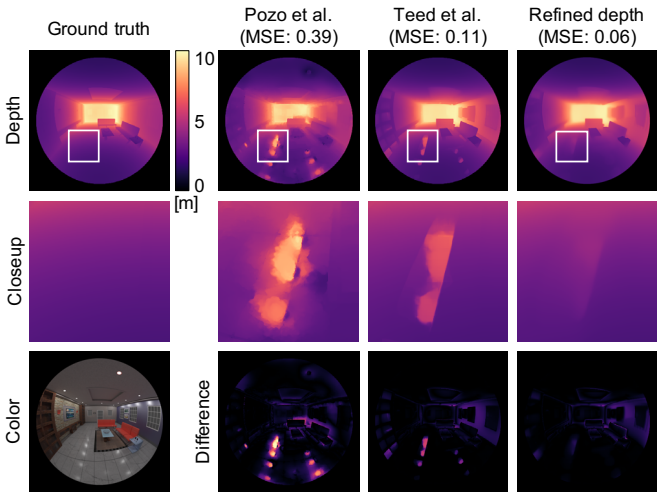


Fig. 6: We compare the accuracy of our depth estimation results on specular surfaces with that of state-of-the-art light field depth estimation methods [2], [26]. The left column shows the ground truth depth and colors of a synthetically rendered scene using the camera parameters same as our prototype system. The second column shows the depth map by Pozo et al. [2] that stands on diffuse assumption. The third presents depth estimation by a learning-based optical flow [26] that we use for initial depth. The fourth column shows our depth result. Over specular surfaces, our method improves the initial depth.

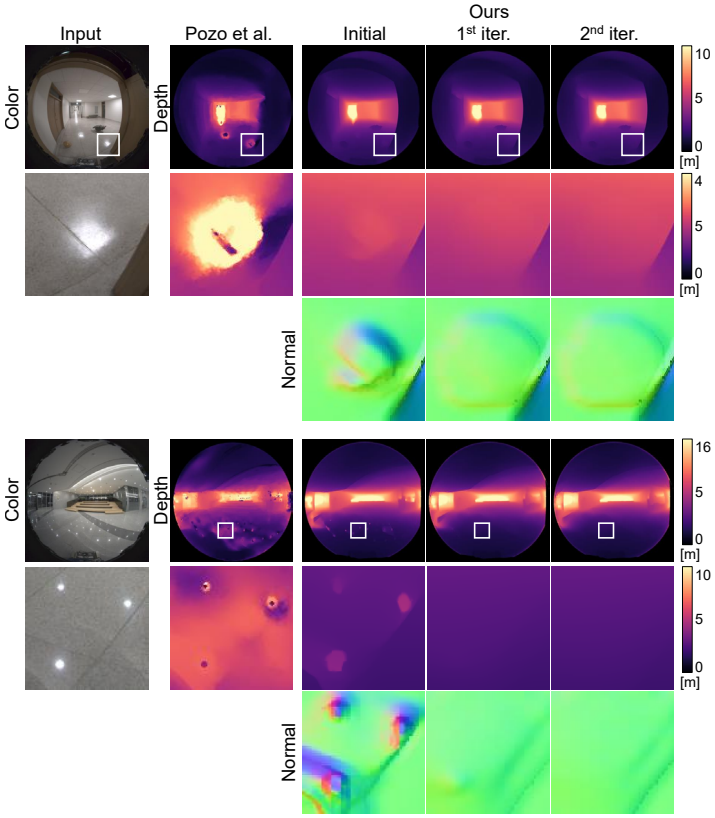


Fig. 7: The real-scene results of our iterative depth refinement process are compared with a state-of-the-art method [2]. Our initial depth and geometric normal estimated by an optical flow method [26] are gradually refined, correcting inaccurate initial depth by specular reflection.

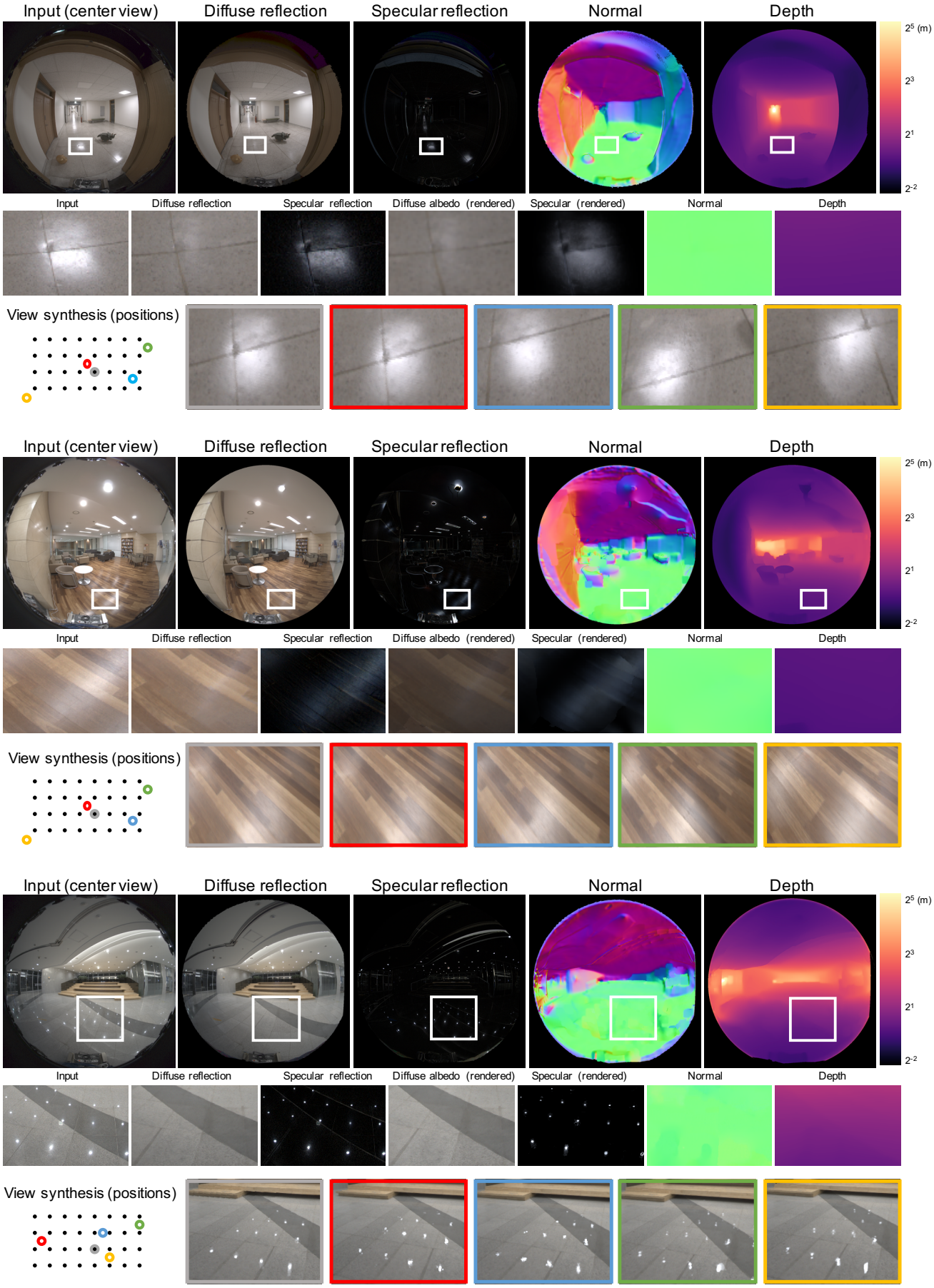


Fig. 8: View synthesis results of three real indoor scenes captured by our prototype. The first row shows an input image at the center view, the separated diffuse and specular reflection, the estimated normals and depths, respectively. The second row presents closeups of results. The last row shows our novel view synthesis with specular changes at different positions.

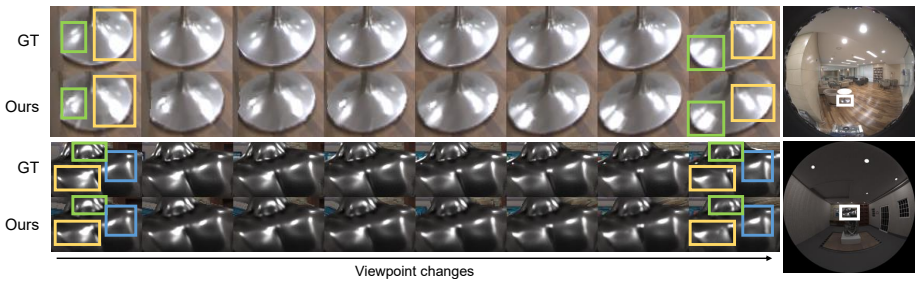


Fig. 9: View synthesis results of our method on non-planar surfaces. View synthesis results on rounded surfaces present a good agreement with ground-truth images with high accuracy. The average SSIM values of these two regions are 0.8858 and 0.8914, respectively.

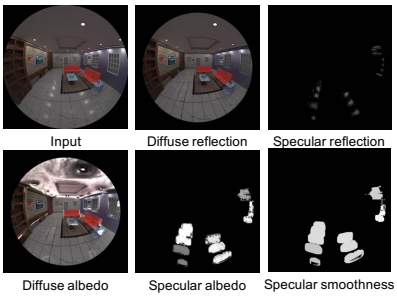


Fig. 10: Intermediate results of appearance parameters estimated by our inverse rendering optimization.

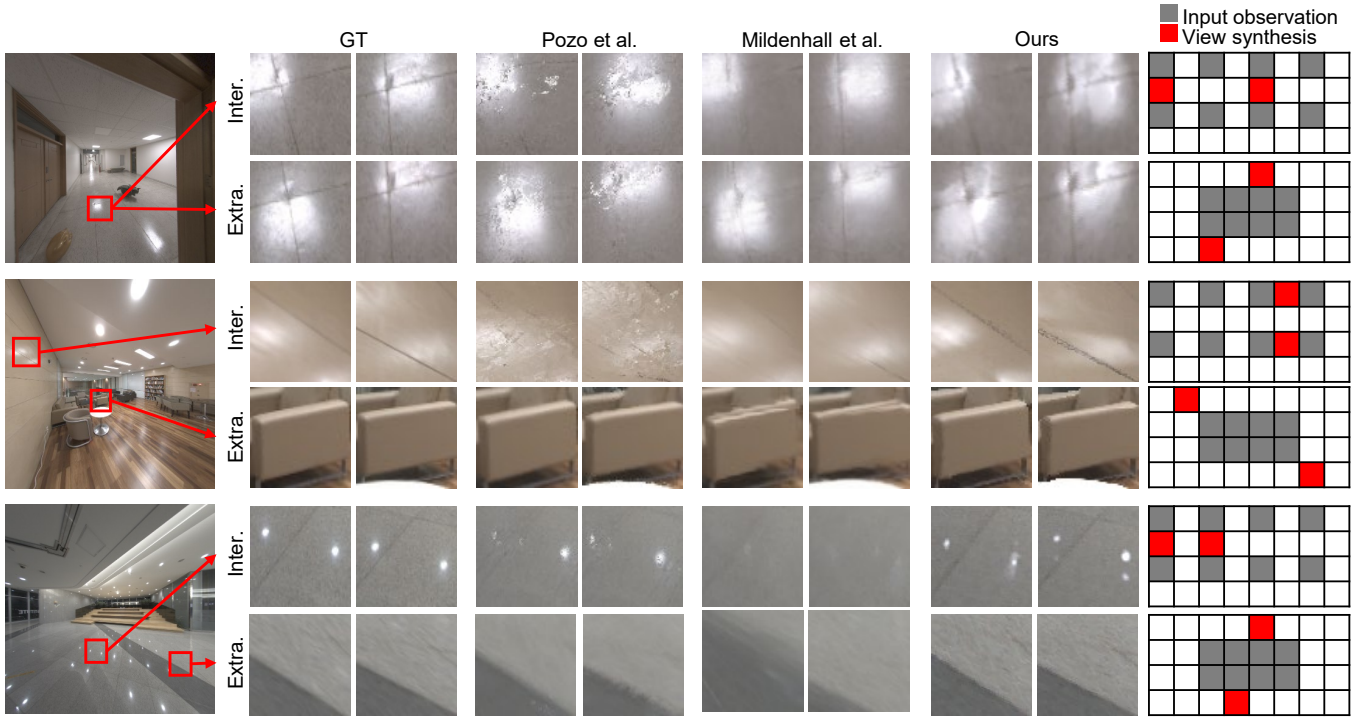


Fig. 11: Comparison of view synthesis results of interpolation and extrapolation with three real scenes captured by our prototype. Eight views are used for input observation for each method, and other views are used for evaluation. While overall view synthesis results of each method are highly comparative to each method, specular reflection results of our method appear more plausibly and similar to the real images. As overall scales from each method vary on an arbitrary scale, we calculated the structural similarity index (SSIM) rather than PSNR on an absolute scale. The averaged SSIM values of the cropped areas by three methods are 0.8584, 0.8902, and 0.9039, respectively.

scenes estimated by the state-of-the-art method [2] and our method with a different number of iterations.

4.2 View Synthesis

We captured three real scenes by using the prototype that we built (shown in Figure 1(a)). The real indoor scene results include a wide range of normal variation and depth range approximately from ∼0.25 m to ∼30 m. Unlike conventional short-baseline light-field cameras, such as Lytro and Raytrix, our wide-baseline light field imaging prototype can cover the entire range of large scene geometry successfully.

Given a set of 32 input images of a scene, our method first separates diffuse and specular reflection and then estimates normal, depth, diffuse albedo, and specular smoothness parameters. The estimated specular parameters (as described in Section 3.2) are shared within each material cluster in order to ensure enough observations in the optimization step. For view synthesis rendering, we use specular albedo calculated by solving the per-cluster least-squares problem among 32 views using Equation (4).

Figure 8 shows the results of three real indoor scenes captured by our prototype. Our inverse rendering method can separate specular reflection against diffuse reflection successfully. Our appearance parameters estimated by our inverse rendering algorithm allows us to interpolate and extrapolate view-dependent appearance at novel viewpoints faithfully. Actual daily objects in various shapes, such as rounded tables or statues, are demonstrated at a relatively small scale within the scenes. Figure 9 demonstrates more closeup results of view synthesis of our method. Along with the changed viewpoints, the specular reflection of view synthesis results over rounded surfaces shows a good

agreement with ground truth images with high SSIM values. Intermediate results of the estimated rendering parameters are demonstrated in Figure 10. Refer to our supplemental video for more results.

Comparison. We qualitatively evaluate our view synthesis results in two different ways: interpolation and extrapolation, compared with results of two state-of-the-art view synthesis methods [2], [18]. We first subsample our eight sub-aperture views from all the views to evaluate the performance of interpolation and extrapolation of the view synthesis methods, as shown in the rightmost column of Figure 11. The dark grey boxes indicate the locations of input sub-aperture views, and the red boxes show the locations of the synthesized views. Figure 11 qualitatively compares the captured real images with the synthesized views.

These two compared methods take more extensive computation across all the input views than ours. The image-based method [18] calculates a set of per-view MPIs, and the geometry-based method [2] makes use of a set of per-view depth maps and per-view mesh blending to calculate a novel view. Therefore, these two baseline methods show good performance overall in different views. In contrast, our method computes all the reflectance parameters and geometry with respect to the center view only, and thus our method performances better near the reference view. In terms of diffuse appearance synthesis, all three methods' performances are highly competitive.

However, in terms of specular appearance synthesis, the baseline methods present suboptimal performance near specular reflection as shown in Figure 11. For instance, the geometry-based method [2] shows noise around specular reflection because their depth estimation tends to convey depth errors near specular reflection. The image-based method [18] overly smooths out some regions while blending several plane images, losing sharp specular reflection. In contrast, our method presents a view-dependent specular appearance plausibly in both interpolation and extrapolation cases.

However, we also found that the lobby floor at the last row of Figure 11 is a hard case due to its extreme smoothness like the mirror. In naive optimization, the specular smoothness parameter diverges to extremes because the specular rendering loss becomes excessively sensitive to even small changes of surface normals. To mitigate the problem, we optimize the specular smoothness parameter in the logarithmic scale.

Impact of Depth Quality. We evaluate the impact of the depth accuracy with respect to the quality of view synthesis. Figure 12 shows the view synthesis results of our method with different depth values at different viewpoints. Each row shows ground-truth images, our view synthesis results given the GT depth maps, depth maps estimated by Teed et al., and depth maps estimated by our method, respectively. It is not surprising that GT depth produces the most accurate synthesis results. However, our depth estimation method allows for better results than results using the state-of-the-art depth estimation method.

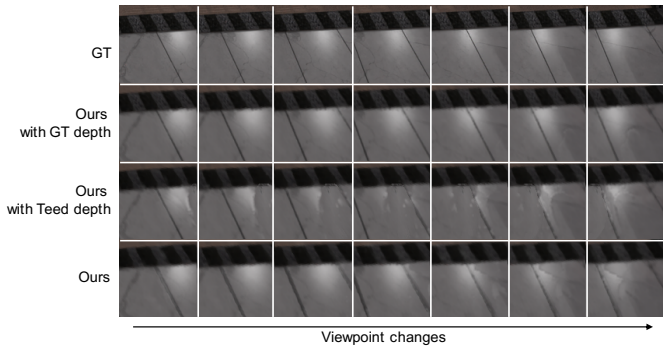


Fig. 12: View synthesis results with GT, Teed et al. [26] and our depth estimation. The average SSIM values of these images are 0.9689, 0.9174, and 0.9363, respectively.

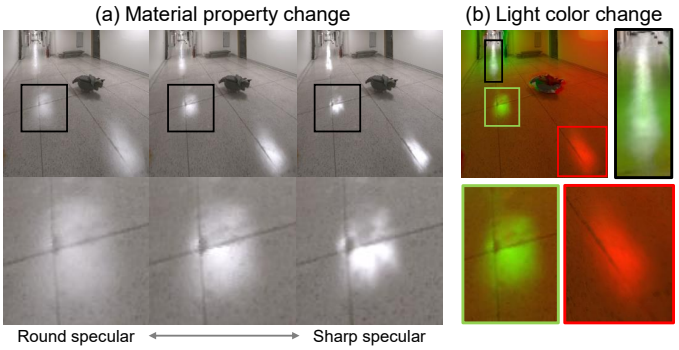


Fig. 13: (a) The center column shows a rendered image of specular reflection using the estimated specular albedo, smoothness, normal, and light by our method. The left column presents an edited appearance with a more rough surface by reducing the specular smoothness by 0.5 in the $\log_{10}$ scale and dividing the specular albedo by 3. The right column shows a smoothed material appearance to cause sharper specular reflection. (b) The colors of the light sources on the ceiling have been changed to different colors. The closest light and the middle light are changed to red and green light respectively and result in corresponding color changes of specular reflection on the floor.

4.3 Applications of Computational Photography

Compared to the methods that blend the observed images to synthesize a seamless view at an unobserved position, our algorithm explicitly estimates the scene geometry and reflectance parameters via inverse rendering. This enables us to go further than the faithful synthesis of the captured scenes and illumination. In our method, the material properties and the colors of scene illumination can be edited freely, allowing for various computational photography applications.

Material Editing. Figure 13 shows two scene edition results using the scene representation parameters estimated by our method. In Figure 13(a), by changing the specular albedo and smoothness parameters while keeping the same diffuse albedos, new images with the different shininess of the material can be created. Also, the colors of existing light sources can be changed. In Figure 13(b), the closest light on the ceiling is changed to red, and the middle light is changed to green, while the furthest light remains the same. While the whole room is filled with smooth red and green

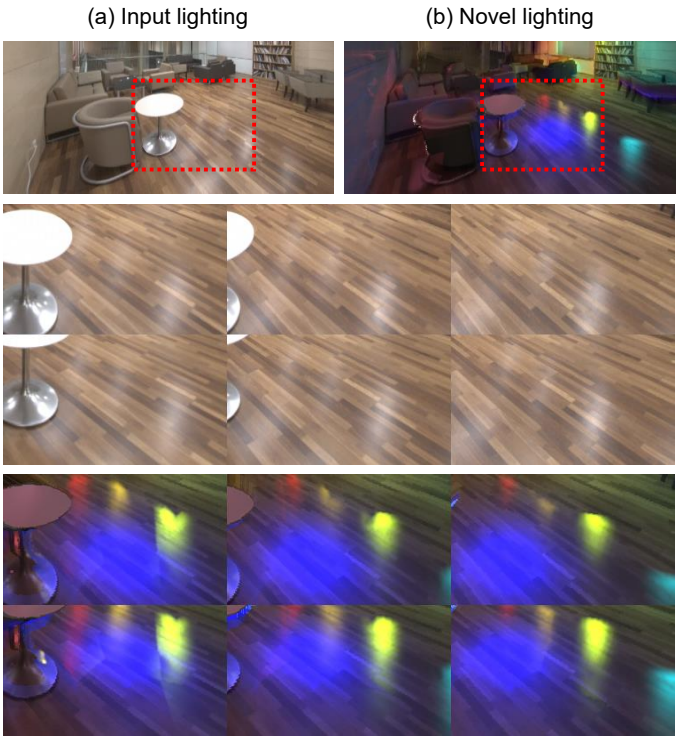


Fig. 14: (a) Input lighting can be eliminated and override by novel point light sources, whose position and color are free to set. (b) The view-dependent specular reflection shown in the input views are substituted by another view-dependent appearance under the novel lighting.

Fig. 15: Upsampled EPI results from wide-baseline light fields.

of the new light colors following their diffuse albedo, the specular reflection individually follows the color of the light source, causing the reflection.

Relighting. Having scene geometry together with the material property, the scene illumination can also be eliminated or added in 3D space, resulting in a dramatic change in scene appearance. In Figure 14(b), the input illumination is virtually eliminated by reducing the existing light source power by ten times so that the view-dependent specular highlights on the floor in the input image disappear. Then, various point light sources of different colors are added in the air so that new colored specular reflections appear according to the colors of the new light sources. See our supplemental video for more application results.

EPI Upsampling. Epipolar plane image (EPI) is the principal representation of light fields frequently used in many works, such as depth estimation and occlusion handling. Enabling sub-aperture view synthesis at any viewpoint, our algorithm can be used to obtain the high-resolution EPIs from wide-baseline light fields. Unlike micro-baseline light fields, such as Lytro, EPI of wide-baseline light fields are severely aliased and discontinuous as shown in the first row of Figure 15. Our algorithm enables us to simulate intermediate observations of the input light fields continuously, plausibly upsampling the EPI while accounting for material appearance.

5 Discussion

Impact of Material Clustering. The number of clusters needs to be determined by the number of materials in the scene. This is a long-lasting issue even in the existing inverse rendering algorithms, where the material numbers are empirically determined and a fixed number of base materials [29]. To tackle this problem, an exhaustive search approach, such as the elbow algorithm, could be used. However, considering computational cost, we determined to adjust the hyperparameter empirically. Figure 16 shows the effect of choice of the number of clusters.

Fig. 16: Impact of the number of clustering on specular reflection.

Memory Efficiency in Rendering. Once the intermediate scene information is estimated, our pipeline can synthesize a novel view with view-dependent appearance using only a few numbers of 2D screen space maps. Explicitly, they are camera parameters, lights, depth, normal, diffuse and specular albedo, and specular smoothness at the center view, which occupies a constant times of an input image resolution. Those are significantly smaller in size compared to the raw light field sub-aperture images, while still have the power of expressing the appearance of light rays along with various directions. Also, our rendering stage can be easily implemented by extending any mesh renderer, empowering its application. In contrast, the image-based method [18] uses multiple MPIs calculated at each input view to synthesize a novel view. Rendering a view requires hundreds of images to be stored and that storage is proportional to the number of input views multiplied by the number of MPI layers.

6 Limitations

Though our method estimates depth and diffuse albedo of each pixel, specular albedo and smoothness are estimated only for surface points that have shown specular reflection at least once during our observation. This is a fundamental limitation of inverse rendering since we cannot infer specular appearance parameters for pixels that we never observe specular reflection. This limitation can be eased under homogeneous material assumption as proposed by [25], or it would be possible to aggregate parameters on the area observed with a specular reflection for future work.

We formulate an inverse rendering optimization problem based on rasterization, in which global illumination is excluded for the sake of simplicity. While our approach

realizes computational optimization of inverse rendering, it introduces an inevitable artifact in factorizing illumination and diffuse albedo over surfaces of the same direction as the indoor light source. For instance, the brightness of the ceiling is dominated by global illumination, where the light on the ceiling is reflected back by the neighboring wall and floor. The diffuse albedo parameters of the ceiling, oriented towards the same direction as the light source, are often overestimated and causing severe artifacts in relighting.

Our method approximates the scene illumination by a set of diffuse point light sources; those should be visible in the hemispherical center view for modeling full 3D locations and incident directions of light rays. Another possible approach for modeling scene illumination is to use environmental map illumination. However, it is inappropriate for scene-scale inverse rendering because object 3D positions are so wide-ranging that the infinite-distance assumption is disabled. Also, many large-scale scenes contain light sources inside each scene, which need to be modeled for view-dependent appearance synthesis. Our lighting model fails under light conditions that cannot be approximated with diffuse point light sources, e.g., a surface light source with a wide area and directional light, such as a spotlight. Especially, we experimented on indoor scenes only, as our algorithm has a limitation on expressing wide, omnidirectional, and indirect illuminations of outdoor scenes.

Also, the reflectance property of many real-world materials, such as metals or brushed surfaces, is hard to be approximated with the isotropic Blinn-Phong model. Transparent or highly reflecting at all points also make our initial depth estimation method fail. These limitations remain as our future work.

7 Conclusion

We have presented an inverse rendering-based view synthesis algorithm that estimates geometric properties and represents view-dependent appearances from wide-baseline light fields of large-scale indoor scenes. We first estimate the scene depth robustly to the view-dependent specular reflection, exploiting plentiful change of observation positions in wide-baseline light fields. Separating diffuse and specular reflection and generating a specular-free image reinforces our base depth estimation process. Based on the estimated geometry and scene lighting, normal and surface reflection properties, i.e., diffuse albedo, specular albedo, and specular smoothness, are estimated through inverse rendering. These results enable re-rendered views at novel positions to obey the view-dependent appearance of specular reflection. For validating our method, we built a wide-baseline light field imaging prototype equipped with 32 fisheye cameras. In experiments, we achieved plausible results with various real scenes. Furthermore, compared to the baseline methods that properly blend the input images for synthesizing views, our inverse-rendering-based method takes advantage of a possible scene edition of material property and lighting.

Acknowledgements

Min H. Kim acknowledges MSIT/IITP of Korea (2017-0-00072), in addition to a partial support of Korea NRF grants (2019R1A2C3007229) and Samsung Research Funding Center of Samsung Electronics (SRFC-IT2001-04).

References

[1] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 673–680.

[2] A. P. Pozo, M. Toksvig, T. F. Schrager, J. Hsu, U. Mathur, A. Sorkine-Hornung, R. Szeliski, and B. Cabral, "An integrated 6DoF video camera and system design," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, pp. 1–16, 2019.

[3] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, p. 69, 2007.

[4] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an interactive 360 light field display," in *ACM SIGGRAPH 2007 papers*, 2007, pp. 40–es.

[5] D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle, "Surface light fields for 3D photography," in *SIGGRAPH*, 2000, pp. 287–296.

[6] H. P. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel, "Image-based reconstruction of spatial appearance and geometric detail," *ACM Transactions on Graphics (TOG)*, vol. 22, no. 2, pp. 234–257, 2003.

[7] P. Hedman, S. Alsisan, R. Szeliski, and J. Kopf, "Casual 3d photography," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–15, 2017.

[8] P. Hedman and J. Kopf, "Instant 3d photography," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–12, 2018.

[9] H. Cho, J. Kim, and W. Woo, "Novel view synthesis with multiple 360 images for large-scale 6-dof virtual reality system," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 880–881.

[10] B. Luo, F. Xu, C. Richardt, and J.-H. Yong, "Parallax360: Stereoscopic 360 scene representation for head-motion parallax," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1545–1553, 2018.

[11] J. Shade, S. Gortler, L.-w. He, and R. Szeliski, "Layered depth images," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 1998, pp. 231–242.

[12] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "Deepstereo: Learning to predict new views from the world's imagery," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5515–5524.

[13] T. Zhou, R. Tucker, J. Flynn, G. Fyffe, and N. Snavely, "Stereo magnification: Learning view synthesis using multiplane images," *arXiv preprint arXiv:1805.09817*, 2018.

[14] P. P. Srinivasan, R. Tucker, J. T. Barron, R. Ramamoorthi, R. Ng, and N. Snavely, "Pushing the boundaries of view extrapolation with multiplane images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 175–184.

[15] J. Flynn, M. Broxton, P. Debevec, M. DuVall, G. Fyffe, R. Overbeck, N. Snavely, and R. Tucker, "Deepview: View synthesis with learned gradient descent," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2367–2376.

[16] E. Penner and L. Zhang, "Soft 3D reconstruction for view synthesis," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–11, 2017.

[17] I. Choi, O. Gallo, A. Troccoli, M. H. Kim, and J. Kautz, "Extreme view synthesis," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7781–7790.

[18] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–14, 2019.

[19] M. Broxton, J. Flynn, R. Overbeck, D. Erickson, P. Hedman, M. DuVall, J. Dourgarian, J. Busch, M. Whalen, and P. Debevec, "Immersive light field video with a layered mesh representation," vol. 39, no. 4, pp. 86:1–86:15, 2020.

[20] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on EPI," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6319–6327.

[21] Y. Wang, F. Liu, Z. Wang, G. Hou, Z. Sun, and T. Tan, "End-to-end view synthesis for light field imaging with pseudo 4DCNN," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 333–348.
[22] M. W. 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*, vol. 38, no. 6, pp. 1155–1169, 2015.
[23] T.-C. Wang, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "SVBRDF-invariant shape and reflectance estimation from light-field cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5451–5459.
[24] Z. Li, Z. Xu, R. Ramamoorthi, and M. Chandraker, "Robust energy minimization for brdf-invariant shape from light fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5571–5579.
[25] T.-T. Ngo, H. Nagahara, K. Nishino, R.-i. Taniguchi, and Y. Yagi, "Reflectance and shape estimation with a light field camera under natural illumination," *International Journal of Computer Vision*, vol. 127, no. 11-12, pp. 1707–1722, 2019.
[26] Z. Teed and J. Deng, "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow," in *European Conference on Computer Vision*. Springer, 2020, pp. 402–419.
[27] J. F. Blinn, "Models of light reflection for computer synthesized pictures," in *Proceedings of the 4th annual conference on Computer graphics and interactive techniques*, 1977, pp. 192–198.
[28] N. Ravi, J. Reizenstein, D. Novotny, T. Gordon, W.-Y. Lo, J. Johnson, and G. Gkioxari, "Accelerating 3D Deep Learning with PyTorch3D," *arXiv:2007.08501*, 2020.
[29] G. Nam, J. H. Lee, D. Gutierrez, and M. H. Kim, "Practical SVBRDF acquisition of 3D objects with unstructured flash photography," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–12, 2018.

Dahyun Kang received her B.Sc. (2019) and M.Sc. (2021) degrees in Computer Science from Korea Advanced Institute of Science and Technology (KAIST). She is currently studying towards Ph.D. degree in KAIST. Her research interests include light field, 3D retrievals, and computational photography.

Daniel S. Jeon received his B.Sc. (2014) and M.Sc. (2016) degrees in Computer Science from Korea Advanced Institute of Science and Technology (KAIST). He is currently studying towards Ph.D. degree in KAIST. His research interests include computational imaging, optics, hyperspectral imaging, BRDF acquisition, and computer graphics.

Hakyeong Kim received her B.Sc. (2019) and M.Sc. (2021) degrees in Computer Science from Korea Advanced Institute of Science and Technology (KAIST). She is currently studying towards Ph.D. degree in KAIST. Her research interests include machine learning for computer vision, image reconstruction, and neural rendering.

Hyeonjoong Jang received his B.Sc. (2017) and M.Sc. (2019) degrees in Computer Science from Korea Advanced Institute of Science and Technology (KAIST). He is currently studying towards Ph.D. degree in KAIST. His research interests include 360 imaging, 3D reconstruction, and view synthesis.

Min H. Kim is an associate professor of computer science at Korea Advanced Institute of Science and Technology (KAIST), leading the Visual Computing Laboratory. Prior to KAIST, he worked as a postdoctoral researcher at Yale University. He received his Ph.D. in computer science from University College London (UCL) in 2010 with a focus on color reproduction in computer graphics. In addition to serving on many conference program committees, such as SIGGRAPH and CVPR, he has been working as an associate editor in various journals: ACM Transactions on Graphics and IEEE Transactions on Computational Imaging. His research interests include computational imaging, computational photography, 3D imaging, and hyperspectral imaging, in addition to color and visual perception.