4D Imaging through Spray-On Optics

JULIAN ISERINGHAUSEN, University of Bonn
BASTIAN GOLDLÜCKE, University of Konstanz
NINA PESHEVA and STANIMIR ILIEV, Bulgarian Academy of Sciences
ALEXANDER WENDER, University of Stuttgart
MARTIN FUCHS, Stuttgart Media University
MATTHIAS B. HULLIN, University of Bonn

Light fields are a powerful concept in computational imaging and a mainstay in image-based rendering; however, so far their acquisition required either carefully designed and calibrated optical systems (micro-lens arrays), or multi-camera/multi-shot settings. Here, we show that fully calibrated light field data can be obtained from a single ordinary photograph taken through a partially wetted window. Each drop of water produces a distorted view on the scene, and the challenge of recovering the unknown mapping from pixel coordinates to refracted rays in space is a severely underconstrained problem. The key idea behind our solution is to combine ray tracing and low-level image analysis techniques (extraction of 2D drop contours and locations of scene features seen through drops) with state-of-the-art drop shape simulation and an iterative refinement scheme to enforce photo-consistency across features that are seen in multiple views. This novel approach not only recovers a dense pixel-to-ray mapping, but also the refractive geometry through which the scene is observed, to high accuracy. We therefore anticipate that our inherently self-calibrating scheme might also find applications in other fields, for instance in materials science where the wetting properties of liquids on surfaces are investigated.

CCS Concepts: • **Computing methodologies** → **Computational photography**; **3D imaging**; **Reconstruction**; **Physical simulation**;

Additional Key Words and Phrases: Plenoptic imaging, inverse rendering, analysis by synthesis

ACM Reference format:
Julian Iseringhausen, Bastian Goldlücke, Nina Pesheva, Stanimir Iliev, Alexander Wender, Martin Fuchs, and Matthias B. Hullin. 2017. 4D Imaging through Spray-On Optics. *ACM Trans. Graph.* 36, 4, Article 35 (July 2017), 11 pages.
DOI: http://dx.doi.org/10.1145/3072959.3073589

1 INTRODUCTION

Light fields [Gortler et al. 1996; Levoy and Hanrahan 1996] describe light leaving a scene on a ray-by-ray basis. They do not only form the foundation of image-based rendering, but have also been shown to facilitate the solution of long-standing vision problems such as depth estimation. For the capture of light fields, few commercial solutions are available; to this day, 2D imagers by far dominate the market. The defining component of a light field imager is an optical and/or mechanical system that maps the 4D space of rays onto the 2D sensor plane. Most such systems are carefully designed to trade between spatial and angular resolution, and to achieve optimal overall imaging performance by maximizing light efficiency and sharpness while avoiding cross-talk and aliasing, all under the given design constraints. On the other end of the scale are "casual" or "random" light field cameras that use every-day reflective or refractive objects [Wender et al. 2015] or randomized optical elements [Antipa et al. 2016; Fergus et al. 2006]. They replace careful optical design by exhaustive calibration of the pixel-to-ray mapping. Here, we take this idea of exploiting low-end optical devices for integral imaging a significant step further. By focusing on a particular, but very common, optical scenario (a window wetted by water drops), we can make extensive use of domain knowledge and physical simulation to greatly facilitate the calibration process. The result is a heterogeneous pipeline that comprises low-level image analysis steps for drop segmentation and feature detection, drop shape simulation to recover the refractive geometry, and a custom bundle adjustment scheme to refine the estimated geometry. With that, our work for the first time enables both the calibration of a dense pixel-to-ray mapping and the acquisition of a light field from a single input image taken through a wetted window.

We consider the following to be our key contributions:

- We propose the use of physical simulation to facilitate the calibration of a-priori unknown imaging systems; in particular, liquid drops as optics for light field imaging.
- We introduce a pipeline for ray-space calibration and the extraction of light field data from a single input image. It combines simple image analysis steps with drop shape simulation, an algorithm for matching and refinement of 2D features, and a custom bundle adjustment scheme to jointly estimate a cloud of sparse 3D features and refine the estimated drop geometry.
- We experimentally validate our pipeline on a selection of static and dynamic scenes.
- Finally, for lack of experimental ground truth data, we evaluate the accuracy of our ray-space calibration and the recovered 3D water drop geometries using synthetic experiments.

2 RELATED WORK

Before we explain our method in detail, we will start by discussing existing works that served as a source of inspiration for our work.

Liquid mirrors and lenses. Liquids have been used for optical purposes throughout history, but it was not until the late 19th century that a rapid technical developments and deeper physical understanding enabled astronomers to construct mirror telescopes from liquid mercury, a technology that is still in use today [Hickson

This work is supported by the German Research Foundation (HU-2273/2-1), the X-Rite Chair for Digital Material Appearance, the ERC Starting Grant "Light Field Imaging and Analysis" and the SFB Transregio 161 "Quantitative Methods in Visual Computing".
Authors' addresses: {iseringhausen|hullin}@cs.uni-bonn.de; bastian.goldluecke@uni-konstanz.de; {nina|stani}@imbm.bas.bg; alexander.wender@visus.uni-stuttgart.de; fuchsm@hdm-stuttgart.de.

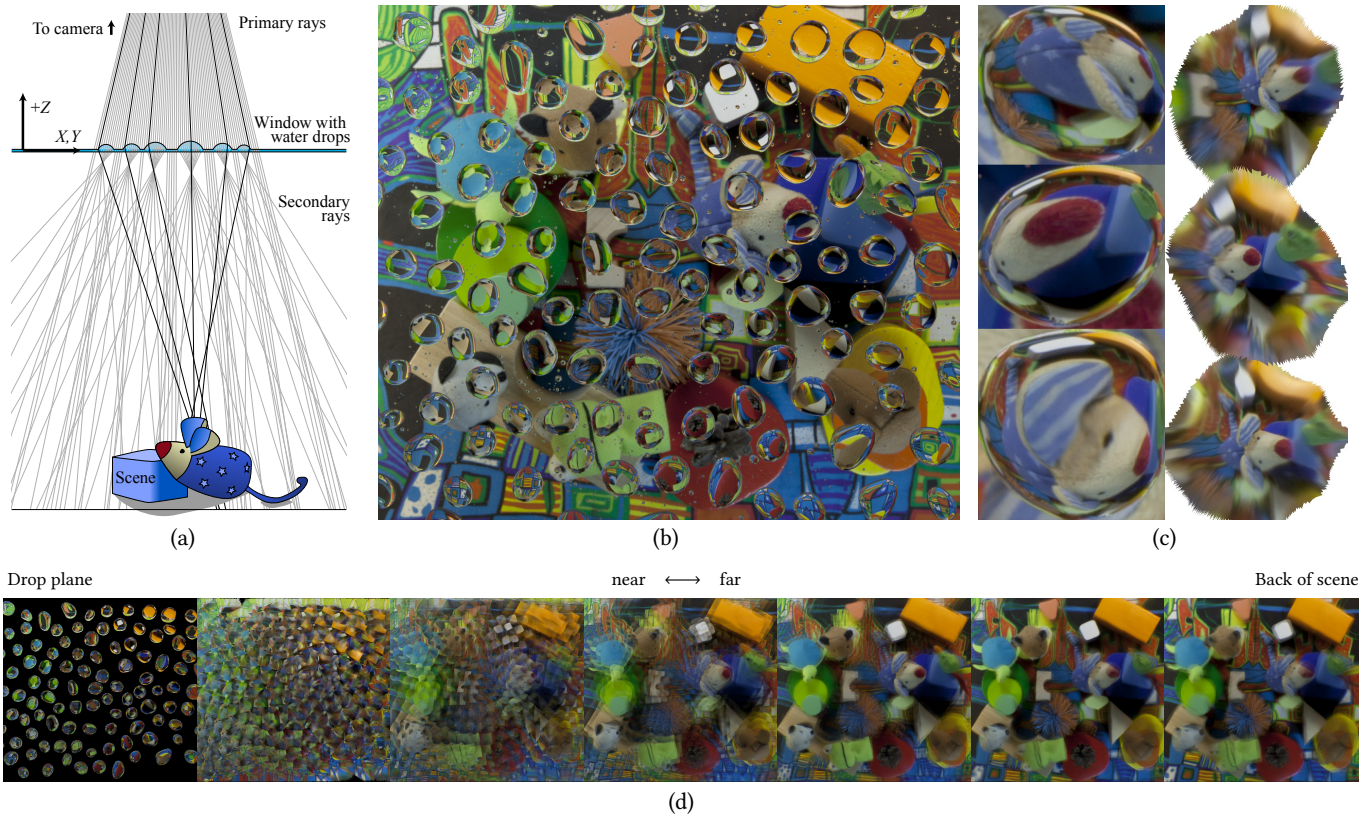


Fig. 1. Using liquids to image light fields ("Animals" example). (a), Our capture setting: the scene is observed by a 2D camera (not in illustration) through a wetted window. Light rays falling through water drops are refracted and sample the scene's light field. (b), Our input is a single image of the scene, as seen by the primary camera. Using drop shape simulation, we establish tentative pixel-to-ray mappings that allow to undistort the individual drop views (c) and, after further refinement, to render a weighted focal stack (d).

et al. 1998]. In technical optics, today's possibilities include variable lenses controlled e.g. by microfluidic channels [Chronis et al. 2003] or electrowetting [Kuiper and Hendriks 2004], and the fabrication of microlens arrays from photoresist through reflow processes [O'Neill and Sheridan 2002]. The computer graphics community has discovered water not only as a natural phenomenon worthy of digital simulation, but also as a display medium [Barnum et al. 2010; Hullin et al. 2011]. Just as we propose in this paper, in these works liquids were exposed to weakly controlled conditions, letting them assume a-priori unknown free-form shapes. Only very recently have researchers succeeded in using such settings for multi-view reconstruction [You et al. 2016]; to our knowledge, our work is the first to perform a full ray-space calibration from a single image taken through water drops.

Light fields. The research history on light fields, while significantly shorter, is nevertheless very rich and diverse [Ihrke et al. 2011]. In this section, we briefly review publications that are the most relevant to our work. They can serve as a starting point for a deeper exploration of the field.

The idea of capturing ray-space radiance measurements can be traced back to Lippmann [1908]. Yet, it was not until the computer age that light field data could be used to synthesize novel images [Gortler et al. 1996; Levoy and Hanrahan 1996], paving the way for a widespread adoption in the graphics and vision communities. Light fields are not only a mainstay of image-based rendering, but have also proven a valuable tool in a wide range of applications, including post-capture refocusing and parallax [Levoy et al. 2006; Ng 2005], depth estimation [Kim et al. 2013; Tao et al. 2013; Wang et al. 2016; Wanner and Goldlücke 2014], as well as for advanced filtering purposes like glare removal [Raskar et al. 2008].

Much theoretical work has been done on light fields, most of it relating back to Adelson and Bergen's definition of the plenoptic function [1991]. Milestones in light field analysis include the development of a sampling framework for image-based rendering by Chai et al. [2000], Ng's Fourier slice theorem [2005] that identifies 2D images with 4D slices of the light field in Fourier domain and Wetzstein et al.'s theory [2013] that unifies the multiplexing of light fields with other plenoptic dimensions. Motivated by practical challenges in the construction of light field imagers, Wei et al. [2015] proposed a unified sampling framework that takes into account lens aberrations and misalignment.

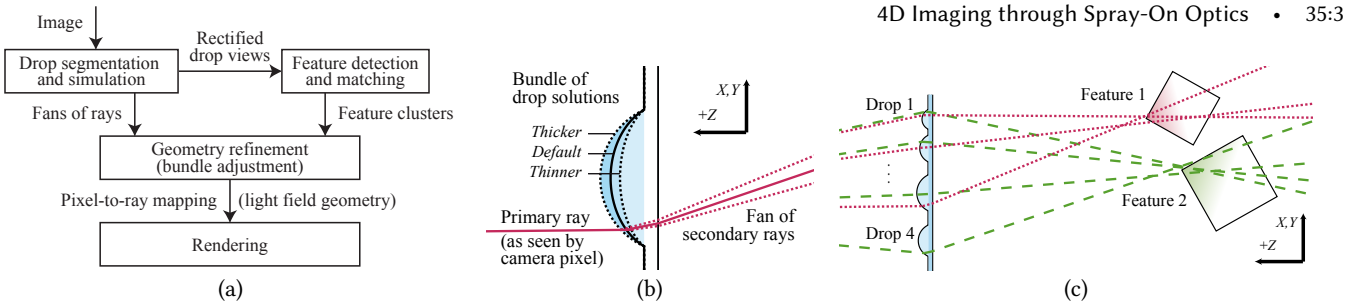


Fig. 2. Illustrations of the imaging pipeline and the underlying ray geometry at different stages. (a), Flow diagram of the reconstruction scheme, which combines a strong physical model (drop shape simulation) with computer vision elements such as image segmentation, feature detection and matching, and bundle adjustment. (b), Until the drop parameter is uniquely determined, each image location (primary ray) corresponds not to a single secondary ray but a fan of rays. (c), Secondary rays from different drops that have been identified to belong to the same scene-space feature (here illustrated by the red and green ray bundles) should intersect as closely as possible. We express this constraint in a cost function (Eq. 2) that sums up, for each feature f, the mutual line-line distances over all pairs of secondary rays belonging to that feature under the given drop volume parameters.

Since light fields in their most common definition are a four-dimensional representation of ray space, their capture poses numerous practical challenges as well. Among the setups proposed are robotic gantries [Levoy and Hanrahan 1996], camera arrays [Wilburn et al. 2005], as well as multiplexing optics like lenslet arrays [Georgiev et al. 2006; Ng 2005], amplitude masks [Veeraraghavan et al. 2007], elaborate mirror arrangements [Fuchs et al. 2013; Mukaigawa et al. 2011; Taguchi et al. 2010], kaleidoscopes [Han and Perlin 2003; Manakov et al. 2013], random elements [Antipa et al. 2016; Fergus et al. 2006] and even household items [Wender et al. 2015]. We note that calibrating an unknown integral imager's ray geometry is closely linked to capturing the geometry of reflective and transparent objects [Ihrke et al. 2008]. Here, most of literature deals with extensions to structured light scanning [Hullin et al. 2008; Tarini et al. 2005; Weinmann et al. 2013]. Kutulakos and Steger investigated the conditions and constraints under which reflective and refractive geometry can be recovered [2008]. In our approach, we constrain ourselves to optical surfaces that follow well-explored physical laws. We integrate this knowledge to estimate the shape of our refractive surface, and hence the geometry of viewing rays, using physical simulation.

Finally, on a higher level, we draw a great deal of inspiration from works on lightweight or free-hand capture techniques, recently culminating in Torralba and Freeman's explorative paper on accidental cameras [2014]. From the first days of light field acquisition, researchers have aimed to avoid high-precision robotic and opto-mechanical designs, instead augmenting the available hardware by appropriate calibration steps [Davis et al. 2012; Gortler et al. 1996]. By replacing optical design with calibration, and calibration with simulation, our work continues in this tradition.

3 EXPERIMENTAL SETUP AND PROCEDURE

In this section, we describe the experimental setup used to capture light fields through water drops.

Parts. Our camera was a Canon EOS 5D Mark II with the 24-105 mm f/4 kit lens set to a fixed 105 mm focal length and f/22 aperture. As substrate for our drops, we used 2 mm thick acrylic (PMMA) sheets. The liquid was tap water. Our model can account for slight changes in refractive index or surface energies by adjusting the drop volume parameter (see Section 4.1). Four diffused 50 W LED area lights served as the light source.

Setup. An illustration of our setup can be found in Fig. 1a. Using a checkerboard target at various distances and the Camera Calibration Toolbox for MATLAB [Bouguet 2004], we calibrated the intrinsic camera parameters to obtain a pixel-to-ray mapping. The camera was then mounted on a tripod and faced down approximately vertically, which we confirmed by placing a small spirit level on the camera's rear display. The tripod mounting point was located approximately 100 cm above the floor. To obtain stationary drops (a requirement for simulation), we mounted the acrylic sheet horizontally at an approximate distance of 50 cm (measured with tape) below the camera's tripod mounting point, and focused the lens to its surface. The LED lights were mounted immediately underneath the window, facing downward onto the scene. Although our method works in ambient light, reflections in the drop surfaces had to be avoided since they interfere with the drop segmentation and distort the measured light field. Our coordinate system is oriented such that the X and Y axes lie in the plane of the window, with the Z axis pointing toward the camera. The pixel-per-millimeter scale in the drop plane was obtained by combining the intrinsic camera calibration and the known distance of the substrate.

Capturing procedure. To capture a light field, we first arranged the scene and ensured that it was well lit. We then used a spray bottle to apply water drops to the acrylic surface. The drops typically take a few seconds to assume their final shape, a process that can be accelerated by gently tapping on the substrate. We triggered image exposure using a remote control. For the CarStunt scene, we used a microcontroller to simultaneously release four toy cars using a solenoid mechanism, and to time the camera exposure. The resulting raw images were converted to 16-bit PNG format using the Camera Raw importer in Adobe Photoshop CS5. Example input images can be seen in Fig. 1b and 3a.

4 RECONSTRUCTION PIPELINE

The input to our reconstruction pipeline consists of a single image, like the one shown in Fig. 1b, as well as a small number of additional parameters like camera projection, the distance of the window and the physical properties of the materials involved (density, refractive

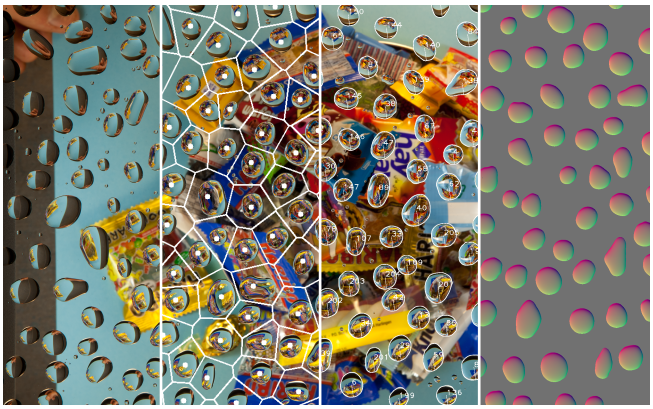


Fig. 3. Segmentation of drops in an image, and simulation of their shape. (a), Input image. (b), Result of semi-automatic circle detection visualized as Voronoi diagram. (c), Final drop contours. Both drop segmentation steps were corrected by additional manual input where needed. (d), Visualization of drop surfaces after simulation. Shown is the solution for the default drop volume parameter.

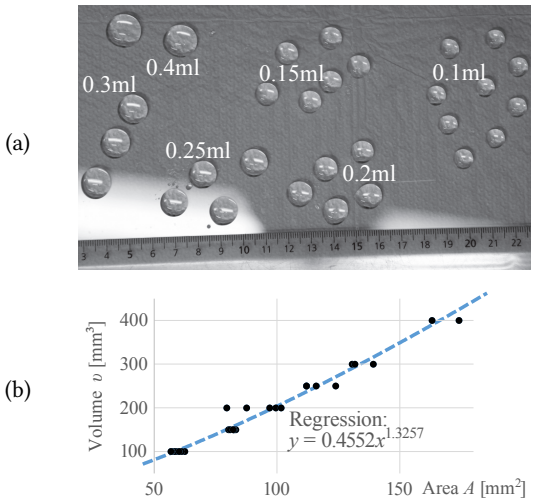


Fig. 4. Without the influence of gravity, area A and volume v of a spherical drop would relate as $v \sim A^{3/2}$. (a), In a pilot experiment, we placed drops of known volume on the substrate and measured their contact area. (b), Regression of a power function reveals that the actual exponent is slightly lower ($v \sim A^{1.33}$). We use this result to initialize the default drop volume.

index, surface energy). The desired output is a dense mapping from pixels in the input image to light field rays, 3D drop surface reconstructions, as well as depth estimates and renderings of the scene from new virtual camera positions. To achieve this goal, we propose a reconstruction pipeline (Fig. 2a) that consists of four major analysis and processing stages:

- extraction and simulation of water drops and ray geometries,
- extraction of scene features that serve as stereo constraints,
- a refinement step (bundle adjustment) to determine the volume parameter for each drop and establish the final pixel-to-ray mapping, and
- post-processing of the resulting light field (depth estimation and rendering).

Here, we motivate and explain these stages.

4.1 Drop extraction and simulation

Since the surface of a sessile drop is energy minimizing, for known physical parameters, the geometry is determined up to a single scalar parameter by the contact line (where drop surface and substrate meet) [Adamson and Gast 1997]. So the first step is to find this contour in the input image. Fully automatic segmentation of drops in images is an unsolved computer vision problem; existing approaches to image restoration [Eigen et al. 2013; Shan et al. 2010] only produce drop contours as a by-product and are not accurate enough to serve as input for drop shape simulation. We approach this problem in a semi-automatic fashion. Since all drops are more or less round, we initialize a map of coarse drop locations with a circle detector (Fig. 3b), drop centers serve as foreground constraints and their Voronoi diagram as background constraints. A state-of-the-art image segmentation algorithm [Gulshan et al. 2010] is then used to determine accurate drop contours. To aid the automatic segmentation in ambiguous or otherwise challenging regions, the user can provide additional constraints by annotating additional drop and background regions. The result is a contour line for each drop (Fig. 3c) which serves as input for a physical simulation that computes the drop geometries [Iliev and Pesheva 2006]. A detailed description of this simulation step can be found in Appendix A. Although we experimentally established a rough relation between a drop's contact area and its expected *default volume* by using a small syringe to place drops of roughly known volume on an acrylic window and fitting a power function to the observations (Fig. 4), the exact volume parameter is not yet known at this stage. For each drop, we simulate a bundle of surfaces that sample a range of values around the default parameter value. One such default solution is visualized in Fig. 3d.

Under a geometric optics model, each pixel samples a *primary ray* entering the camera, which in turn corresponds to a *secondary ray* of light leaving the scene. Given the refractive geometry of glass pane and water drops, the relation between primary and secondary rays can now be computed via ray tracing. For each primary ray, we thus obtain a *fan* of secondary rays, one ray for each value of the (yet unknown) drop volume parameter (Fig. 2b).

4.2 Feature extraction and matching

To further constrain the solution, we use SIFT [Lowe 1999] to extract keypoints from the image and identify scene features that are visible in multiple neighboring drops. The main challenge in this stage is that the drop views in the input image are strongly distorted, making scene features appear quite differently in different views (Fig. 1c). Prior to keypoint extraction, we therefore undistort the drop views using the default pixel-to-ray mapping from the previous stage. In particular, we perform a simple projection of each drop view to a plane located roughly at the distance of the scene. This effectively rectifies the view (Fig. 1d), allowing SIFT to perform well despite the fact that the default drop volume estimate (used for undistortion)

Fig. 5. Two examples of feature clusters found in different scenes, projected back into the original images. The total number of such clusters and the number of keypoints in each cluster depend on the visual complexity of the scene, as well as the drop arrangement.

may not be the final one. The next step is to match keypoints found in neighboring views that correspond to the same scene feature. Using the algorithm from Appendix B, we obtain a set of scene features that are visible in more than one drop, and for each of the features a set of keypoints in the input image that show the feature (the *feature cluster*, Fig. 5). We define the matching matrix G to reflect the relation between scene-space features and image-space keypoints,

$$G(f,k) = \begin{cases} 1, & \text{if keypoint } k \text{ belongs to feature } f, \\ 0, & \text{else.} \end{cases} \tag{1}$$

In combination with the results from the previous stage, we further know the fan of secondary rays that belongs to each keypoint as a function of the drop volume parameter.

4.3 Geometry refinement

The features found in the previous stage now become the stereo constraints in our reconstruction: all secondary rays belonging to the *same feature* should intersect in the same point in space (Fig. 2c). At the same time, the secondary rays belonging to features in the *same drop* are all controlled jointly by that drop's volume parameter. The purpose of this stage is to determine the vector of volume parameters $\mathbf{v} = (v_1, \ldots, v_m)$ (one parameter per drop) that produces the best global agreement between secondary rays. To this end, we define a cost function $F(\mathbf{v})$ that sums up, across all features f and all pairs of image keypoints (k_i, k_j) that represent a given feature in drops i and j, the line-line distance $\mathrm{dist}_{\mathrm{ray}}$ between the corresponding secondary rays,

$$F(\mathbf{v}) = \sum_f \sum_{k_i \neq k_j} G(f,k_i)\, G(f,k_j)\, \mathrm{dist}_{\mathrm{ray}}^{(v_i,v_j)}(k_i,k_j). \tag{2}$$

This formulation is closely related to bundle adjustment, or the joint estimation of viewing parameters and scene geometry from multi-view stereo images [Hartley and Zisserman 2004]. Rather than the usual reprojection error of features in image space, our cost function measures the distance between rays in scene space. To approach the high-dimensional non-linear problem of minimizing $F(\mathbf{v})$, we use an iterative coordinate descent scheme. We simultaneously perform line searches along all coordinate axes (volume parameters) and choose the solution with the lowest cost. This updating step is iterated until a local minimum of F is reached. To increase the chance of obtaining a good solution close to the global optimum, we restart the optimization process $n_{\mathrm{iterations}} = 3$ times with perturbed solution vectors.

The outcome of the refinement stage is a vector of drop volumes $\mathbf{v}$ that is locally optimal under Eq. 2. This results in a dense and uniquely defined mapping from input pixels to secondary rays, which concludes the geometric calibration of the light field. To validate the outcome, we also compute the root mean square (RMS) scene feature localization error. We obtain it from the pairwise line-line distances across all pairs of matched keypoints, a value that will increase when either drop or scene geometries are inconsistent.

4.4 Rendering

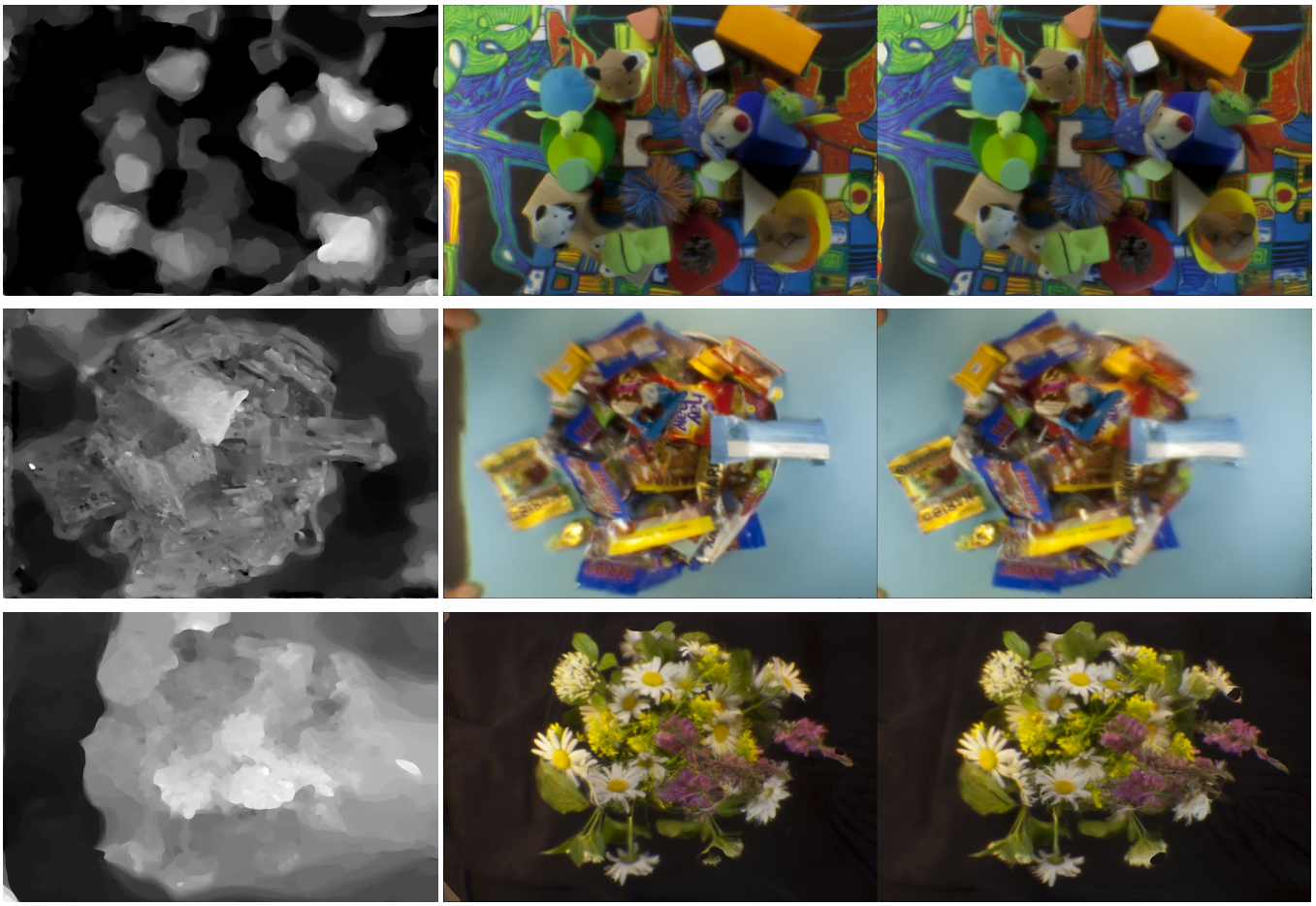
For the further assessment of the resulting light fields, we implemented a specialized renderer. Unlike light fields captured using properly designed optical systems, the ones reconstructed from liquid drops using the described method are irregularly and sparsely sampled. In addition, the estimated ray geometry is affected by residual inaccuracies.

To obtain high-quality 2D images from these liquid light fields, we use a rendering scheme that is guided by a per-pixel depth estimate. First, we set the parameters of a synthetic camera. For the desired viewpoint, we define a stack of planes of sufficient extent and resolution to fully contain the scene. By propagating all rays to the plane and integrating them there, a focused image is obtained from a light field [Levoy and Hanrahan 1996]; all focused images together form a focal stack (Fig. 1d). The sparsity of views necessitates careful selection of rays and a specific weighting scheme. At any given location in a given plane, we retrieve a set of rays that intersect in this location. From these rays and the corresponding pixel values in the input image, we compute a weighted average color value, and the uncertainty as the weighted standard deviation of radiance samples. The underlying assumption is that if all samples have the same color, they probably originate from the same point in the scene. Hence, a low standard deviation indicates a likely depth value. We use this relation to extract a per-pixel depth assignment from the focal stack (Fig. 6a).

As the final step, we follow the standard practice [Wanner and Goldlücke 2014] of using the depth map to extract an all-in-focus image from the focal stack (Fig. 6b). To render the scene under a different synthetic view, all steps including the focal stack computation are repeated. We provide implementation details and parameters in Appendix C.

5 RESULTS

To demonstrate our method, we acquired liquid light fields of six scenes, three of which are shown in Fig. 6. All input images as well as the recovered ray mappings are available as supplemental datasets to this paper. We further provide a collection of animated results in the supplemental video. All reconstructions rely exclusively on "wet" rays that passed through drops, except Fig. 7 where some of the artifacts introduced by "dry" rays can be seen.



(a) (b)

Fig. 6. Depth estimation and rendering for the "Animals" (top), "Candy" (middle) and "Flowers" (bottom) light fields. From a weighted focal stack (Fig. 1d), we estimate a depth map (a) and use it to render all-in-focus images (b). The cross-eye stereogram shown here was obtained by performing all rendering steps twice under different camera settings. Animated versions of these results are available in the supplemental video.

Scene	n_{drops}	# Clusters	RMS error	ET	$\bar{A}_{\text{drops}}$	$\bar{f}_{\text{drops}}$	n_{sec}	$\bar{\delta}_{\text{s}}$	$\bar{\alpha}$
Animals	126	1924	4.46 mm	250 ms	113.72 mm^2	89.49 mm	6 457 957	0.10 mm	3.03°
Candy	210	5454	0.79 mm	40 ms	85.01 mm^2	109.04 mm	5 064 711	0.10 mm	2.68°
Flowers	123	1868	1.31 mm	500 ms	112.50 mm^2	89.59 mm	6 236 003	0.10 mm	2.98°
CarStunts*	226	3424	2.66 mm	5 ms	84.09 mm^2	103.44 mm	5 389 975	0.10 mm	2.65°
Dwarfs*	143	2188	4.02 mm	250 ms	93.01 mm^2	84.17 mm	6 214 855	0.09 mm	2.72°
Firework*	205	489	1.33 mm	125 ms	85.39 mm^2	106.02 mm	5 036 900	0.10 mm	2.72°

Table 1. Our example scenes in numbers: count of drops n_{drops} used for reconstruction, number of feature clusters, RMS localization error of 3D features, exposure time, average drop footprint $\bar{A}_{\text{drops}}$, average drop focal length $\bar{f}_{\text{drops}}$, number of secondary rays n_{sec} in final light field, average spacing $\bar{\delta}_{\text{s}}$between secondary rays at a typical scene distance, average angle $\bar{\alpha}$ between drop views at scene depth (view separation). Results for the scenes marked with * are presented and discussed in the supplemental document.

The colorful "Animals" scene consists of plush animals and wooden building blocks in front of a richly textured Hundertwasser pattern. All surfaces are of mostly Lambertian (diffuse) reflectance. After undistorting the drop views using the initial drop estimate, the algorithm produces a large number of plausible clusters that reach even into the peripheral parts of some drops (Fig. 5), proving the good quality of the rectification step. After the light field calibration,

Fig. 7. Rendering of the "Animals" data set using both "wet" and "dry" rays. The usage of "dry" rays increases the resolution (see e.g. the furry texture at the mouse's nose) but also introduces artifacts due to unsegmented drops and incomplete coverage.

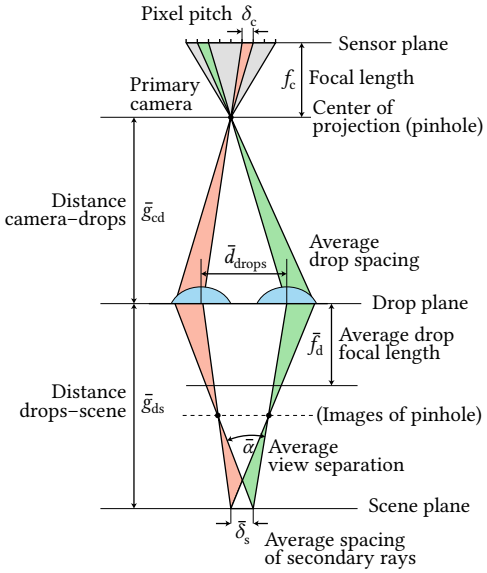


Fig. 8. Description of geometric parameters used in Section 6.1.

the alignment of the drop views and the depth estimates are of sufficient quality (Fig. 6b) to produce all-in-focus renderings that are rich in detail (Fig. 6c) and convey a good depth impression. In the drop estimation step, the 3D localization errors for the sparse feature clusters are on the order of 4.5 mm and hence relatively high compared to the other datasets. We notice that features located around depth discontinuities tend to produce the highest errors. A possible explanation is that in regions with prominent occlusion effects, detected features may not correspond to real points in space and can therefore be stereo-inconsistent.

Using the same scene, we also experimented with the usage of "dry" rays for rendering (Fig. 7). We observed a noticeable increase in detail for projections close to the primary camera projection, but also heavy artifacts caused by the numerous unsegmented small drops and the "Swiss cheese" topology of the direct view. To our knowledge, there is no fully automatic, pixel-precise and robust segmentation method that would enable the use of "dry" rays in the geometry refinement step as well. Here, mislabeled pixels would not only produce visual artifacts but also add an uncontrollable error source to the drop volume estimation.

The "Candy" scene is an arrangement of different kinds of candy (chocolate bars, gummy bears, etc.) in small plastic packages. It exhibits strongly non-Lambertian reflectance, since many of the packages are made of of high-gloss material or even partly transparent. The scene has a relatively shallow depth range (7 cm) which, despite the challenging materials, allows the feature optimization to achieve sub-millimeter localization errors. As expected from the view-dependent nature of glossy and transparent materials, the reconstructed depth maps are not as smooth as in the other scenes. Still, the recovered depth estimates coarsely reflect the overall scene structure and are sufficient to produce output renderings of relatively high resolution (Fig. 6). In fact, the stereo pair conveys a decent stereo impression of the scene, including view-dependent specular highlights. We note that in regions of constant color, small errors in the depth estimate may have little or no effect on the rendered outcome.

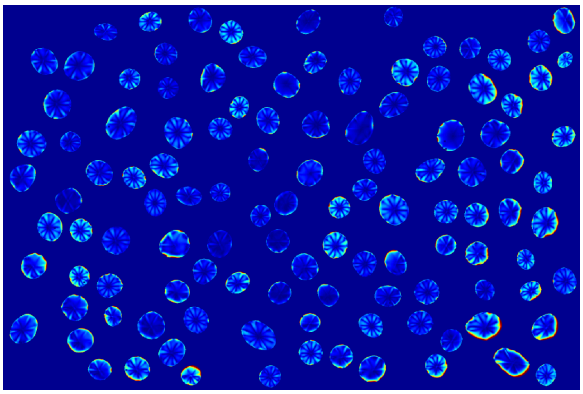
The "Flowers" scene consists of an arrangement of meadow flowers that are of mostly diffuse reflectance. The recovery of ray geometry works robustly, as evidenced by a small feature reconstruction error. Nevertheless, this light field proves to be extremely challenging to render: the recovered depth maps and, consequently, the renderings, contain numerous artifacts (Fig. 6). We identify several factors that may contribute to this problem. They include the total scene depth (measured with a ruler at 25 cm), the presence of repetitive features (daisy petals and small yellow flowers), and overall high spatial and angular frequencies which are not adequately sampled by the sparse and low-resolution drop views.

6 SYSTEM PERFORMANCE AND QUANTITATIVE EVALUATION

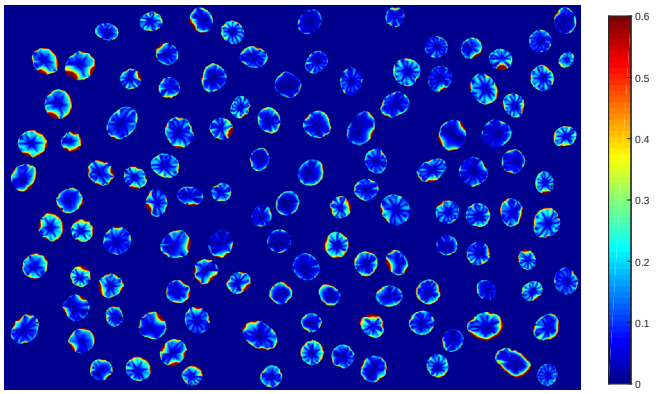
Spray-on optical systems are highly volatile and therefore hard to impossible to fully characterize "in the wild". Here, we list basic geometric relations for scattered arrangements of lens-like elements, and discuss the factors that affect the ray-optical system resolution under a pinhole model for the primary camera. We further use a synthetic replica of our experimental setup to measure the reconstruction accuracy of our pipeline under realistic conditions.

6.1 Resolution

Since light field imagers commonly trade spatial resolution against angular resolution, we used the following three measures to characterize our system: The average spacing between secondary rays when intersecting a plane at a typical scene depth ($\bar{g}_{ds}$ = 300 mm), the average angular separation $\bar{\alpha}$ between different drop views at that depth, and the total number n_{sec} of secondary rays. Assuming the drops to behave like thin lenses and taking into account the geometric parameters introduced in Fig. 8, we can estimate the spatial

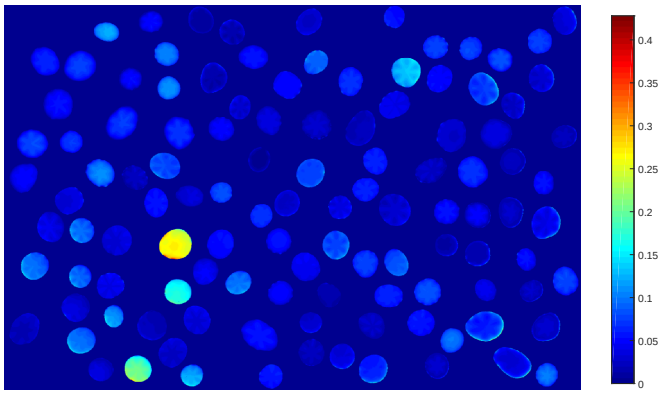


(a) Error of secondary rays [°], ground truth segmentation



(b) Secondary ray error [°], semiautomatic segmentation (Section 4.1)

(c) Synthetic input image, f/22



(d) Drop surface error [mm], ground truth segmentation

Fig. 9. False color error plots for our light field calibration on a synthetic scene.

resolution $\bar{\delta}_s$ of a setup in the paraxial limit as

$$\bar{\delta}_s = \frac{(\bar{g}_{ds} - \bar{f}_d) \cdot \bar{g}_{cd}}{\bar{f}_d \cdot f_c} \delta_c \tag{3}$$

and its average view separation $\bar{\alpha}$ as

$$\bar{\alpha} = 2 \tan^{-1}(\bar{d}_{drops}/2\bar{g}_{ds}). \tag{4}$$

Example values from our experimental datasets can be found in Table 1. Supposing uniformly distributed drops, the total number of secondary rays n_{sec} can be estimated as

$$n_{sec} = n_{pr} \frac{\bar{A}_{drops} \cdot n_{drops}}{A_{sensor} \cdot (g_{cd}/f_c)^2}, \tag{5}$$

where, in addition to the symbols introduced in Fig. 8, $\bar{A}_{drops}$ is the average drop footprint (area), n_{drops} the total number of segmented drops, and A_{sensor} the sensor area.

6.2 Synthetic experiment

Since we are not aware of any solutions for 3D scanning water drop surfaces, we assessed the accuracy of our algorithm using a synthetic experiment. Using the Mitsuba renderer [Jakob 2010], we modeled our imaging setup, procedurally generated and rendered a scene with random clutter under different aperture settings ($f/2$, $f/4$, $f/8$, $f/22$, pinhole), and extracted ground-truth primary and secondary ray geometries. An example rendering can be found in Fig. 9c. The textures were randomly sampled from the Describable Textures Dataset [Cimpoi et al. 2014], and for the 116 virtual water drops we re-used meshes from previous simulations, which fulfill the Young-Laplace equation and can therefore be assumed to be physically plausible under the given constraints.

We then performed a full ray-space calibration (starting with drop simulation) using our pipeline, and computed the RMS angular error in secondary rays and the RMS error in the intersection point between primary ray and drop. For both measures, the perfectly known "dry" rays were of course excluded. By randomly removing drops from the set, we varied the density of views fed into the bundle adjustment step. As the error plot in Fig. 10 shows, the typical ray-space calibration error thus obtained was 0.1° to 0.2° with a typical RMS drop surface error of 0.06 mm. Notably, up to $f/8$ the calibration quality was mostly independent of the aperture and even across a wide range of drop numbers. The pipeline only started to break down when neighboring views stopped to share the same scene features due to the increased distance between them. Example error maps for the full set of drops are shown in Fig. 9a,b. We observe that a few drops show significantly higher errors than the rest, which we attribute to mismatched keypoints.

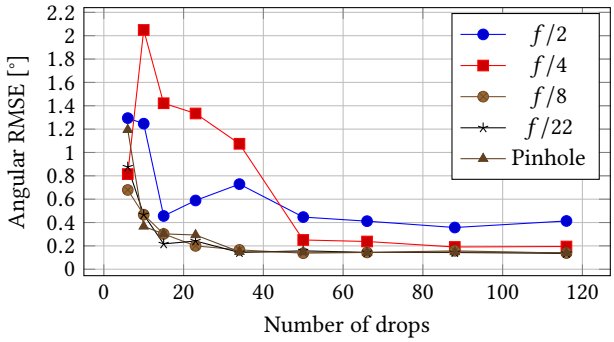


Fig. 10. The RMS angular error for secondary rays, plotted as function of the number of drops used for the reconstruction. Only as about 80% of the drops are removed from the set, the error starts increasing significantly. For large apertures ($f/2$, $f/4$) this effect can be observed earlier.

These results were obtained using ground-truth segmentation of drop contours, also obtained from the renderer. To evaluate the influence of errors in the segmentation, we also performed the semi-automatic segmentation step as described in Section 4.1. For the full set of drops at f/22, this change increased the RMS angular error from 0.136° to 0.234°.

7 DISCUSSION AND OUTLOOK

What is possible? We were able to show, to our knowledge for the first time, that capturing a light field through weakly controlled liquid optics, like water drops on a window, is an ambitious but realistic goal. From a single input image, our pipeline successfully recovers drop geometries, pixel-to-ray mappings and depth maps. Since water drops are minimal surfaces and hence smooth, the resulting image quality is at least comparable to what has been achieved using randomly structured reflective or refractive materials [Antipa et al. 2016; Fergus et al. 2006; Wender et al. 2015], even though our approach does not rely on exhaustive calibration. The recovered drop geometries, while technically a by-product, are of high quality, so depending on one's viewpoint one might also interpret our method as a 3D scanner for water drops that exploits stereo cues from the surrounding light field.

What are the limiting factors? The main limitations of our method are the restriction to a horizontal plane and the need for manual interaction during the drop segmentation step. For non-trivial scenes, like CarStunts, using colored water can reduce the amount of manual intervention required. We captured our experimental data in a conservative, near-pinhole setting ($f/22$) to achieve good focus in the plane and in the scene. This limitation is not exclusive to our method; in fact, a large part of light field research relies on synthetic or experimental reference data obtained under pinhole [Honauer et al. 2016] or near-pinhole [Kim et al. 2013; Vaish et al. 2008] settings. On the other hand, our evaluation on synthetic data suggests that the ray-space recovery and the drop surface estimation work reliably for much wider apertures as well. Finally, we note that rendering new views from sparsely and irregularly sampled light fields (especially with some residual ray-space uncertainty) remains a major challenge that even state-of-the-art techniques are still not quite up to. In fact, most image-based techniques do not generalize to our setting, so significant work will have to be done on depth estimation and filtering techniques to obtain the highest possible output quality under the given constraints.

What might become possible, and how? Water drops on slanted substrates constitute a dynamic phenomenon that is currently not covered by our model. The simulation of such scenarios is of great interest in various application fields (like architectural and automotive design) and the subject of ongoing research. It is therefore our hope that a solution could become possible in the not-too-distant future. While we demonstrate our high-level approach on the recovery of light fields, the quality obtained in a setting as uncontrolled as ours will obviously never rival that from a properly designed optical system. However, we can imagine many imaging situations under unfavorable conditions that could benefit from restoration techniques based on similar ideas. The recovered drop geometries can be of interest in materials science where the wetting behavior of liquids on surfaces is an important area of investigation. Estimating a handful additional material parameters (like the surface tensions, currently assumed to be known) seems like a plausible leap regarding the hundreds of degrees of freedom we are already recovering.

8 CONCLUSION

In this work, we set out to explore the challenge of capturing light fields through drops sitting on a clear window. To this end, we introduced a novel approach for establishing the ray geometries in this scenario, and crafted a reconstruction pipeline from it. Starting from a 2D input image, our algorithm segments drop outlines, simulates drop shapes, traces rays through the drops to undistort the image, and uses image features to refine the parameters. A key feature of our pipeline is its transparency, modularity and robustness regarding the choice of the individual components. The resulting light fields typically contain 100 to 250 scattered views (one per drop), which can then be combined to render the scene from novel viewpoints.

Our research is motivated by a line of work that aims to replace carefully designed and highly specialized capture setups with a combination of casually captured data, careful calibration and computational reconstruction. By contributing a novel take on integral imaging, and by showcasing the use of physical simulation to regularize severely underconstrained imaging tasks, we hope that this paper will serve as a source of inspiration for future work.

REFERENCES

Adamson, A. and A. Gast. 1997. *Physical Chemistry of Surfaces.* Wiley.
Adelson, E. H. and J. R. Bergen. 1991. The plenoptic function and the elements of early vision. *Computational models of visual processing* 1, 2 (1991).
Antipa, N., S. Necula, R. Ng, and L. Waller. 2016. Single-Shot Diffuser-Encoded Light Field Imaging. In *IEEE International Conference on Computational Photography (ICCP).*
Barnum, P. C., S. G. Narasimhan, and T. Kanade. 2010. A Multi-layered Display with Water Drops. *ACM Trans. Graph.* 29, 4, Article 76 (2010), 7 pages. DOI: https://doi.org/10.1145/1778765.1778813
Bouguet, J.-Y. 2004. Camera calibration toolbox for MATLAB. (2004).
Brakke, K. A. 1992. The surface evolver. *Experimental mathematics* 1, 2 (1992), 141–165.
Brakke, K. A. 2013. Surface Evolver 2.70. (2013). http://facstaff.susqu.edu/brakke/evolver/evolver.html.

Chai, J.-X., X. Tong, S.-C. Chan, and H.-Y. Shum. 2000. Plenoptic Sampling. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*. 307–318.

Chronis, N., G. L. Liu, K.-H. Jeong, and L. P. Lee. 2003. Tunable liquid-filled microlens array integrated with microfluidic network. *Opt. Express* 11, 19 (Sep 2003), 2370–2378. DOI: https://doi.org/10.1364/OE.11.002370

Cimpoi, M., S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi. 2014. Describing Textures in the Wild. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Davis, A., M. Levoy, and F. Durand. 2012. Unstructured Light Fields. *Comp. Graph. Forum* 31, 2 (May 2012), 305–314. DOI: https://doi.org/10.1111/j.1467-8659.2012.03009.x

De Gennes, P.-G., F. Brochard-Wyart, and D. Quéré. 2004. *Capillarity and wetting phenomena: drops, bubbles, pearls, waves*. Springer Science & Business Media.

Eigen, D., D. Krishnan, and R. Fergus. 2013. Restoring an Image Taken through a Window Covered with Dirt or Rain. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 633–640.

Fergus, R., A. Torralba, and W. T. Freeman. 2006. *Random lens imaging*. Technical Report MIT-CSAIL-TR-2006-058.

Fuchs, M., M. Kächele, and S. Rusinkiewicz. 2013. Design and Fabrication of Faceted Mirror Arrays for Light Field Capture. *Computer Graphics Forum* 32, 8 (2013), 246–257. DOI: https://doi.org/10.1111/cgf.12201

Georgiev, T., K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala. 2006. Spatio-angular Resolution Tradeoffs in Integral Photography. In *Proceedings of the 17th Eurographics Conference on Rendering Techniques (EGSR '06)*. Eurographics Association, 263–272. DOI: https://doi.org/10.2312/EGWR/EGSR06/263-272

Goldlücke, B., E. Strekalovskiy, and D. Cremers. 2012. The natural vectorial total variation which arises from geometric measure theory. *SIAM Journal on Imaging Sciences* 5, 2 (2012), 537–563.

Gortler, S. J., R. Grzeszczuk, R. Szeliski, and M. F. Cohen. 1996. The Lumigraph. In *Proc. 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*. ACM, New York, NY, USA, 43–54. DOI: https://doi.org/10.1145/237170.237200

Gulshan, V., C. Rother, A. Criminisi, A. Blake, and A. Zisserman. 2010. Geodesic star convexity for interactive image segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 3129–3136.

Han, J. Y. and K. Perlin. 2003. Measuring Bidirectional Texture Reflectance with a Kaleidoscope. *ACM Trans. Graph. (Proc. SIGGRAPH 2003)* (2003), 741–748. DOI: https://doi.org/10.1145/1201775.882341

Hartley, R. I. and A. Zisserman. 2004. *Multiple View Geometry in Computer Vision* (2nd ed.). Cambridge University Press.

Hickson, P., E. F. Borra, R. Cabanac, S. C. Chapman, V. De Lapparent, M. Mulrooney, and G. A. Walker. 1998. Large Zenith Telescope project: a 6-m mercury-mirror telescope. In *Astronomical Telescopes & Instrumentation*. International Society for Optics and Photonics, 226–232.

Honauer, K., O. Johannsen, D. Kondermann, and B. Goldluecke. 2016. A dataset and evaluation methodology for depth estimation on 4D light fields. In *Asian Conference on Computer Vision*. Springer.

Hullin, M. B., M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. A. Lensch. 2008. Fluorescent Immersion Range Scanning. *ACM Trans. Graph. (Proc. SIGGRAPH 2008)* 27, 3 (Aug. 2008), 87:1–87:10.

Hullin, M. B., H. P. A. Lensch, R. Raskar, H.-P. Seidel, and I. Ihrke. 2011. Dynamic Display of BRDFs. In *Computer Graphics Forum (Proc. EUROGRAPHICS)*, Oliver Deussen and Min Chen (Eds.). Eurographics, Blackwell, Llandudno, UK, 475–483.

Ihrke, I., K. Kutulakos, H. Lensch, M. Magnor, and W. Heidrich. 2008. State of the art in transparent and specular object reconstruction. In *EUROGRAPHICS 2008 STAR*.

Ihrke, I., G. Wetzstein, D. Lanman, and W. Heidrich. 2011. State of the art in computational plenoptic imaging. In *EUROGRAPHICS 2011 STAR*.

Iliev, S. 1995. Iterative method for the shape of static drops. *Computer Methods in Applied Mechanics and Engineering* 126, 3 (1995), 251–265.

Iliev, S. 1997. Static drops on an inclined plane: equilibrium modeling and numerical analysis. *Journal of colloid and interface science* 194, 2 (1997), 287–300.

Iliev, S. and N. Pesheva. 2003. Wetting properties of well-structured heterogeneous substrates. *Langmuir* 19, 23 (2003), 9923–9931.

Iliev, S. and N. Pesheva. 2006. Nonaxisymmetric drop shape analysis and its application for determination of the local contact angles. *Journal of colloid and interface science* 301, 2 (2006), 677–684.

Jakob, W. 2010. Mitsuba renderer. (2010). http://www.mitsuba-renderer.org.

Kim, C., H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross. 2013. Scene Reconstruction from High Spatio-Angular Resolution Light Fields. *ACM Trans. Graph. (Proc. SIGGRAPH 2013)* 32, 4 (2013), 73:1–73:12.

Kuiper, S. and B. Hendriks. 2004. Variable-focus liquid lens for miniature cameras. *Applied Physics Petters* 85, 7 (2004), 1128–1130.

Kutulakos, K. N. and E. Steger. 2008. A Theory of Refractive and Specular 3D Shape by Light-Path Triangulation. *International Journal of Computer Vision* 76, 1 (2008), 13–29. DOI: https://doi.org/10.1007/s11263-007-0049-9

Levoy, M. and P. Hanrahan. 1996. Light Field Rendering. In *Proc. 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*. ACM, New York, NY, USA, 31–42. DOI: https://doi.org/10.1145/237170.237199

Levoy, M., R. Ng, A. Adams, M. Footer, and M. Horowitz. 2006. Light Field Microscopy. *ACM Trans. Graph. (Proc. SIGGRAPH 2013)* (2006), 924–934. DOI: https://doi.org/10.1145/1179352.1141976

Lippmann, G. 1908. La photographie intégrale. *CR Acad. Sci.* 146 (1908), 446–451.

Lowe, D. G. 1999. Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision (ICCV)*. 1150–1157.

Manakov, A., J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H.-P. Seidel, and I. Ihrke. 2013. A Reconfigurable Camera Add-on for High Dynamic Range, Multi-Spectral, Polarization, and Light-Field Imaging. *ACM Trans. Graph. (Proc. SIGGRAPH 2013)* 32, 4, Article 47 (July 2013), 14 pages. DOI: https://doi.org/10.1145/2461912.2461937

Mukaigawa, Y., S. Tagawa, J. Kim, R. Raskar, Y. Matsushita, and Y. Yagi. 2011. Hemispherical Confocal Imaging Using Turtleback Reflector. In *Computer Vision – ACCV 2010*. Springer, 336–349. DOI: https://doi.org/10.1007/978-3-642-19315-6_26

Ng, R. 2005. Fourier Slice Photography. *ACM Trans. Graph. (Proc. SIGGRAPH 2005)* (2005), 735–744. DOI: https://doi.org/10.1145/1186822.1073256

O'Neill, F. T. and J. T. Sheridan. 2002. Photoresist reflow method of microlens production Part I: Background and experiments. *Optik-International Journal for Light and Electron Optics* 113, 9 (2002), 391–404.

Pock, T., D. Cremers, H. Bischof, and A. Chambolle. 2010. Global Solutions of Variational Models with Convex Regularization. *SIAM Journal on Imaging Sciences* (2010).

Raskar, R., A. Agrawal, C. A. Wilson, and A. Veeraraghavan. 2008. Glare Aware Photography: 4D Ray Sampling for Reducing Glare Effects of Camera Lenses. *ACM Trans. Graph. (Proc. SIGGRAPH 2008)* 27, 3, Article 56 (Aug. 2008), 10 pages. DOI: https://doi.org/10.1145/1360612.1360655

Shan, Q., B. Curless, and T. Kohno. 2010. Seeing Through Obscure Glass. In *Proceedings of the 11th European Conference on Computer Vision: Part VI (ECCV'10)*. Springer-Verlag, Berlin, Heidelberg, 364–378. http://dl.acm.org/citation.cfm?id=1888212.1888241

Taguchi, Y., A. Agrawal, A. Veeraraghavan, S. Ramalingam, and R. Raskar. 2010. Axial-Cones: Modeling Spherical Catadioptric Cameras for Wide-Angle Light Field Rendering. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2010)* 29, 6 (Dec 2010), 172:1–172:8.

Tao, M. W., S. Hadap, J. Malik, and R. Ramamoorthi. 2013. Depth from Combining Defocus and Correspondence Using Light-Field Cameras. In *IEEE International Conference on Computer Vision (ICCV)*. 673–680.

Tarini, M., H. P. A. Lensch, M. Goesele, and H.-P. Seidel. 2005. 3D acquisition of mirroring objects using striped patterns. *Graphical Models* 67, 4 (2005), 233–259.

Torralba, A. and W. Freeman. 2014. Accidental Pinhole and Pinspeck Cameras. *International Journal of Computer Vision* 110, 2 (2014), 92–112. DOI: https://doi.org/10.1007/s11263-014-0697-5

Vaish, V. and others. 2008. The (New) Stanford Light Field Archive. (2008). http://lightfield.stanford.edu/lfs.html.

Veeraraghavan, A., R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. 2007. Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing. *ACM Trans. Graph. (Proc. SIGGRAPH 2007)* 26, 3, Article 69 (2007). DOI: https://doi.org/10.1145/1276377.1276463

Wang, T.-C., A. Efros, and R. Ramamoorthi. 2016. Depth estimation with occlusion modeling using light-field cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2016).

Wanner, S. and B. Goldlücke. 2014. Variational Light Field Analysis for Disparity Estimation and Super-Resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 3 (2014), 606–619.

Wei, L.-Y., C.-K. Liang, G. Myhre, C. Pitts, and K. Akeley. 2015. Improving Light Field Camera Sample Design with Irregularity and Aberration. *ACM Trans. Graph.* 34, 4, Article 152 (2015), 11 pages. DOI: https://doi.org/10.1145/2766885

Weinmann, M., A. Osep, R. Ruiters, and R. Klein. 2013. Multi-View Normal Field Integration for 3D Reconstruction of Mirroring Objects. *Proceedings of the International Conference on Computer Vision* (Dec. 2013), 2504–2511.

Wender, A., J. Iseringhausen, B. Goldlücke, M. Fuchs, and M. B. Hullin. 2015. Light Field Imaging through Household Optics. In *Vision, Modeling & Visualization*, David Bommes, Tobias Ritschel, and Thomas Schultz (Eds.). Eurographics Association, 159–166. DOI: https://doi.org/10.2312/vmv.20151271

Wetzstein, G., I. Ihrke, and W. Heidrich. 2013. On Plenoptic Multiplexing and Reconstruction. *International Journal of Computer Vision* 101, 2 (2013), 384–400. DOI: https://doi.org/10.1007/s11263-012-0585-9

Wilburn, B., N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. 2005. High Performance Imaging Using Large Camera Arrays. *ACM Trans. Graph. (Proc. SIGGRAPH 2005)* (2005), 765–776. DOI: https://doi.org/10.1145/1186822.1073259

You, S., R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi. 2016. Waterdrop Stereo. *CoRR* (2016). arXiv:1604.00730v1

A DROP SHAPE ANALYSIS

The drop shape is approximated by a triangle mesh (we use $n_{\text{vertices}} = 12781$ vertices), which we initialize as a spherical cap that fulfills the

given volume and the contact angle that follows from the material-specific wetting parameters under Young's Law [De Gennes et al. 2004]. An iterative procedure then gradually transforms the initial circular contact line until the desired contact line L is obtained [Iliev and Pesheva 2006]. During this transition, the drop surface is gradually updated to fulfill the Young-Laplace equation while preserving the drop volume. The core numerical method employed is an iterative minimization procedure, first developed for homogeneous surfaces [Iliev 1995] and then extended to treating heterogeneous surfaces and line tension effects [Iliev 1997; Iliev and Pesheva 2003]. Equivalent tools are available in the public domain, for example *Surface Evolver* [Brakke 1992, 2013]. Physical constants used in the simulation are: $g = 9.81\,\mathrm{m/s^2}$ for the gravity acceleration, and the respective material values to model the wetting behaviour of water on acrylic glass (the mass density $\rho_{\mathrm{water}} = 1000\,\mathrm{kg/m^3}$ of the liquid and the surface tensions $\gamma_{\mathrm{water}} = 72.8\,\mathrm{mN/m}$, $\gamma_{\mathrm{PMMA}} = 41.0\,\mathrm{mN/m}$).

B FEATURE CLUSTERING

Keypoints k_1 and k_2 that form a correspondence match should not only be visually similar but also geometrically plausible. We therefore define the distance measure

$$\begin{aligned}\mathrm{dist}^{(v_1,v_2)}(k_1,k_2) &= \alpha\,\mathrm{dist}_{\mathrm{ray}}^{(v_1,v_2)}(k_1,k_2)\\ &\quad+(1-\alpha)\,\mathrm{dist}_{\mathrm{SIFT}}(k_1,k_2),\end{aligned}\tag{6}$$

where $\mathrm{dist}_{\mathrm{ray}}{}^{(v_1,v_2)}(k_1,k_2)$ is the line-line distance between the two corresponding secondary rays predicted under the drop volume parameters v_1 and v_2, and $\mathrm{dist}_{\mathrm{SIFT}}(k_1,k_2)$ the Euclidean distance between SIFT feature vectors. To achieve compatibility, both distance functions are normalized to the interval $[0,1]$ by dividing by the maximum respective distance across all pairs of keypoints. The parameter $\alpha \in [0,1]$ controls the relative weighting of the two terms. We keep it constant at $\alpha = 0.2$.

Using this distance measure, we construct a sparse graph of feature correspondences by adding clusters of scene-space features. We start with the pair of keypoints that are closest to each other with a distance $d_{\min}$, and proceed by adding keypoints from adjacent drops with a distance no greater than $\beta \cdot d_{\min}$ to the existing ones. This procedure is iterated until every drop belongs to at least $n_{\mathrm{clusters}} = 15$ clusters. In all our experiments, we set $\beta = 2$; keypoints that already belong to a cluster will no longer be considered in following iterations.

C RENDERING

For depth estimation, we use a variant of a plane sweep algorithm in order to deal with the irregular set of rays. The depth map is viewpoint-dependent. For a given camera and target resolution, we initialize a range of 75–100 depth layers at discrete distances $z \in \{z_1,\ldots,z_N\}$ from the camera. For each depth layer z and pixel x, we compute the color vector $I_z^{\{r,g,b\}}(x)$ as a weighted average of the radiances L over the set of rays $R_{x,z}$, a subset of all rays intersecting the plane within the footprint of the pixel,

$$I_z^{\{r,g,b\}}(x) = \frac{1}{\sum w_{\mathbf{r}}}\sum_{\mathbf{r}\in R_{x,z}} w_{\mathbf{r}} L^{\{r,g,b\}}(\mathbf{r}).\tag{7}$$

For the set $R_{x,z}$ we choose the five intersecting rays that have the smallest angular distance $\alpha_{\mathbf{r}}$ to the query ray that belongs to pixel x in the virtual camera, i.e., that are most representative for the desired synthetic view. The weights $w_{\mathbf{r}}$ are given by

$$w_{\mathbf{r}} = 1 - \frac{\alpha_{\mathbf{r}}}{\max_{\mathbf{q}\in R_{x,z}}(\alpha_{\mathbf{q}})}.\tag{8}$$

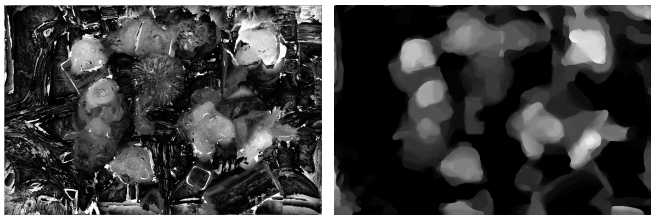
An example of such a weighted focal stack $I_z(x)$ is shown in Fig. 1d. We use it to compute the cost $\rho(x,z)$ for assigning depth z to x using the root-mean-square-deviation

$$\rho(x,z) = \frac{1}{3}\sum_{c\in\{r,g,b\}}\left(\sum_{\mathbf{r}\in R_{x,z}}\frac{\left(L^c(\mathbf{r}) - I_z^c(x)\right)^2}{|R_{x,z}|}\right)^{1/2}\tag{9}$$

over the radiances $L(\mathbf{r})$. Minimizing this cost for each pixel independently results in a noisy depth estimate with significant errors around depth discontinuities (Fig. 11). Therefore, we formulate the cost of the full depth map d on the image plane Ω as

$$E(d) = \int_\Omega \|\nabla d(x)\| + \lambda\rho(x, d(x))\,\mathrm{d}x.\tag{10}$$

The total-variation (TV) penalty of the gradient of the depth map encourages piecewise smooth solutions and can be optimized using the technique of functional lifting [Pock et al. 2010]. We use the implementation provided by `cocolib` [Goldlücke et al. 2012]. Given the depth map d, we obtain the all-in-focus image I_{all} from the chosen view point by extracting the color from the layer corresponding to the correct depth label, i.e. setting $I_{\mathrm{all}}(x) = I_{d(x)}(x)$.



(a) Naïve (unregularized) depth map (b) TV-regularized depth map

Fig. 11. Effect of regularization on the depth estimate.