Vision, Modeling, and Visualization (2015)
D. Bommes, T. Ritschel, and T. Schultz (Eds.)

# Light Field Imaging through Household Optics

A. Wender[1], J. Iseringhausen[2], B. Goldlücke[3], M. Fuchs[1], M. B. Hullin[2]

[1]University of Stuttgart, Germany [2]University of Bonn, Germany [3]University of Konstanz, Germany

**Abstract**

*Although light fields are well-established as a tool in image-based rendering and computer vision, their capture is still at a relatively early stage. In this article, we search for imaging situations similar to uncalibrated integral optics, noticing that they are common in everyday life. We investigate light field capturing scenarios which are provided by commonly available items like cutlery as optical building blocks. Using a generic calibration approach based on structured light, we reconstruct the light path providing an unorthodox light field capturing setup. As the resulting data is unstructured and poorly sampled and thus unsuited for standard image-based rendering pipelines, we propose techniques for the processing of such light fields. Additionally, we have implemented a novel depth estimation scheme to guide the rendering process. We demonstrate the potential of these techniques on different scenes, both static and dynamic, recorded by combining a DSLR camera with household items.*

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and scanning

## 1. Introduction

The light field [LH96, GGSC96] is a powerful concept that describes light emitted from a scene on a ray-by-ray basis. Light fields do not only enable new approaches like image-based rendering, they have also shown to facilitate the solution of long-standing computer vision problems. Yet, a widespread deployment of light field imagers is still hampered by practical challenges. The market shared among commercial solutions (Lytro, Pelican, Raytrix) is still a niche and by far outweighed by conventional 2D imagers.

At the heart of a light field imager is an optical system to multiplex the 4D space of rays onto the 2D sensor plane. Such systems are designed to provide a carefully chosen trade-off between spatial and angular resolution, and overall optimal imaging properties (light efficiency, cross-talk, sampling and filtering, etc.). Opening our eyes to the world around us, we discover many optical scenarios (Fig. 1) that could potentially convert a traditional 2D camera into a light field imager. In practice, all such unorthodox light field cameras have imperfect optical properties, and calibrating their effect on light rays is a challenging task. In this article, we propose a generic way of recovering the ray-space transform of refractive and reflective surfaces, and show that different everyday objects can in fact be used for single-shot light field capture of dynamic scenes.

The key contributions of this paper are as follows:

- We propose the combination of conventional 2D cameras with household items for the capture of light fields.
- We introduce a calibration scheme to recover the ray-space mapping caused by refractive or reflective objects to high precision, allowing to capture unstructured light field data through items like structured glass, cutlery or decorative objects.
- We introduce a rendering scheme that is guided by a novel depth estimation approach, and demonstrate its use for post-capture depth-of-field rendering.

## 2. Related work

Light fields have a rich research history, as evidenced by the recent survey by Ihrke et al. [IWLH11]. In this section, we briefly review publications that are the most relevant to our work. Since the seminal works introducing image-based rendering with light fields [GGSC96, LH96], new applications such as microscopy [LNA*06] or depth estimation [KZP*13, THMR13, WG14] have been demonstrated.

Much theoretical work has been done on light fields, most of it relating back to Adelson and Bergen's definition of the plenoptic function [AB91]. Milestones in light field analysis include the development of a sampling framework for image-based rendering by Chai et al. [CTCS00], Ng's Fourier slice theorem that identifies 2D images with 4D slices of the light field in Fourier domain [Ng05] and Wetzstein et al.'s general sampling theory that unifies the

© The Eurographics Association 2015.

DOI: 10.2312/vmv.20151271

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

**Figure 1:** *Examples of integral imagers as they appear in everyday situations. From left to right: spider web in morning dew, cathedral glass window, building façade with mirrored windows. In this paper, we explore the possibilities of using such data for typical light field applications, like depth estimation and refocusing.*

multiplexing of light fields with other plenoptic dimensions [WIH13]. All these works were done under the fundamental premise that either the spatial or angular domain is sampled densely, or at least in a structured grid that would allow for reasonably straightforward interpolation of missing data points [SYGM03]. In contrast, a central property of unorthodox light field cameras is that they produce extremely sparse and scattered samplings of both domains. The calibration, rendering and depth estimation schemes proposed in this work cater to this particular challenge.

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 [LH96], camera arrays [WJV*05], as well as multiplexing optics like lenslet arrays [Ng05, GZC*06], amplitude masks [VRA*07], elaborate mirror arrangements [MTK*11, FKR13] and even kaleidoscopes [HP03, MRK*13]. Since we use refractive or reflective objects as imaging optics, our work is an instance of the multiplexing principle. Our goal, however, is to move away from well-designed setups of specialized optical elements and towards using objects from daily life. A calibration step must make up for what our systems are lacking in geometric well-definedness and optical quality.

Calibrating the ray geometry of an unknown integral imager is closely related to the problem of capturing the geometry of reflectors and refractors [IKL*08]. The majority of literature deals with extensions to traditional laser scanning [HFI*08] and structured light [TLGS05, WORK13]. Kutulakos and Steger investigated the conditions and constraints in recovering reflective and refractive geometry [KS08]. Using angular-domain light field probes, Wetzstein et al. qualitatively visualized the refractive properties of transparent objects [WRH11]. Shan et al. [SCK10] explored optical and computational ways to undo the distorting effect

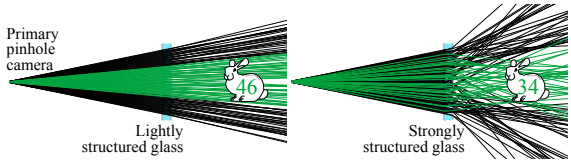



**Figure 2:** *An object observed by a pinhole camera through light field transformers of different directional spread. Out of 100 viewing rays from the camera, those that intersect with the object volume are marked in green. The more strongly the transformer scrambles the rays, the more diverse viewing directions on the object are obtained. However, this also means that the light field is sampled more sparsely.*

of structured glass, but did not go as far as to exploit its effects for the purpose of light field sampling.

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 [TF14]. 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 [GGSC96, DLD12]. In this tradition, we hope that our no-design, all-calibration approach will do its part to inspire upcoming work.

## 3. Light field imaging

Capturing a light field amounts to sampling the function of radiance $L(\mathbf{r})$ transported along any ray that passes through an interface surface. To identify such a ray, four parameters are required. Common choices of parameterization include a point and a direction, $\mathbf{r} = (\mathbf{x}, \mathbf{d})$, and the intersection coordinates with two planes, $\mathbf{r} = (u, v, s, t)$. In traditional single-shot light field photography, carefully designed optical elements are used to map incoming light rays to individual measurement devices, $M(\mathbf{r}) = \mathbf{r}'$.

Our attention in this work is on the exact opposite: a combination of a standard camera (the *primary camera*) with an optical system (the *light field transformer*) that is provided by the world around us, with sub-optimal optical properties and virtually no control over $M$. In this scenario, a light field transformer can take on many forms, including structured glass or metalized objects, even water drops (Fig. 1).

Before we introduce the calibration of different light field transformers, we need to understand the effects the transformer has on rays of light.

**Projection.** Assuming a pinhole projection for the primary camera, each image coordinate corresponds to exactly one viewing ray. After traversing the light field transformer (i.e., the reverse mapping $M^{-1}$), those viewing rays sample the volume of interest (Fig. 2).

© The Eurographics Association 2015.

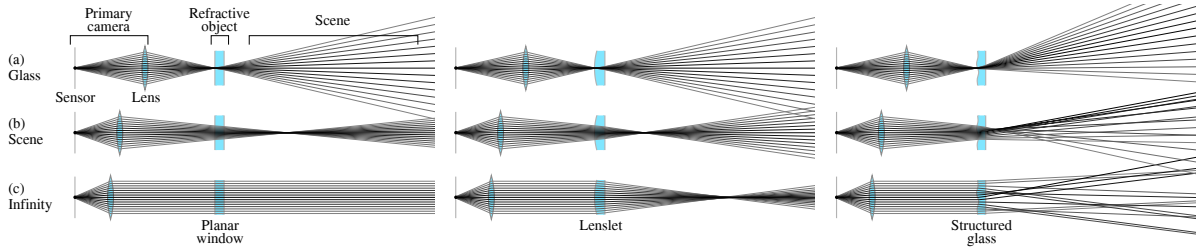



**Figure 3:** *Focusing in case of a scattered light field: Ray tracings of three refractive geometries as seen by a thin-lens camera at different focal settings: on the glass surface (a), into the scene (b), or at infinity (c). To calibrate the geometry of viewing rays (Sec. 3.1), the working volume needs to be sufficiently well focused. Consequently, the focal distance should be set inside the reconstruction volume (b) or at infinity (c). Simple refractive geometries like a plano-parallel slab (left) or a spherical lenslet (center) are capable of keeping the viewing rays bundled across those focal settings. For arbitrarily structured glass (right), however, only the worst choice of focusing on the first glass surface (a) produces even predictable defocus.*

The relative placement of camera, transformer and scene has an important influence not only on this sampling, but also on the net usable portion of the image. A major challenge lies in the divergence of the ray bundle. Viewing rays that already miss the working volume in the primary projection will likely also escape after transformation. In order to bias those far off-axis rays back onto the scene, specialized light field imagers employ additional prisms [GZC*06] or oblique projections [Ng05]. Our light field transformers, by and large, have no such biasing function and preserve existing ray divergence. Only through the spread of their effect on the ray direction, they can divert *some* viewing rays back toward the working volume. The ideal primary projection to minimize this problem would be orthographic with perfectly parallel rays. When using consumer cameras, the likeliest option is to zoom to telephoto mode.

**Focus.** Relaxing the pinhole model to a lens of finite aperture size, we find that focus will be challenging to attain. In his use of a lenslet array, Ng [Ng05] matches the f-number of the lenslets to that of main lens; a choice that maximizes light throughput while avoiding cross-talk and aliasing. Our improvised light field cameras do not consist of well-defined lenses but general, asymmetric, inhomogeneous optical elements. As Fig. 3 illustrates, the best option for a general light field transformer is to stop down the aperture and focus on the imaging object, placing the scene out of focus.

**Inventory.** Throughout this paper, we refer to various household objects that we use as light field transformers: two differently structured glass panes (cathedral glass and a cross-ribbed glass reminiscent of lenslet arrays), small glass ball ornaments and table spoons. Together, they cover a wide range of different scales and other properties:

| Name | Optics | Surface | Pattern | Views | # Rays |
|---|---|---|---|---|---|
| Cathedral | Refractive | Rough | Random | / | 16.3M |
| CrossRib | Refractive | Smooth | Regular | $28\times18$ | 2.6M |
| Spheres | Reflective | Smooth | Discrete | $(13\ldots17)\times6$ | 258K |
| Spoons | Reflective | Smooth | Discrete | $4\times3$ | 892K |

The portion of usable pixels, or rays, ranges from about

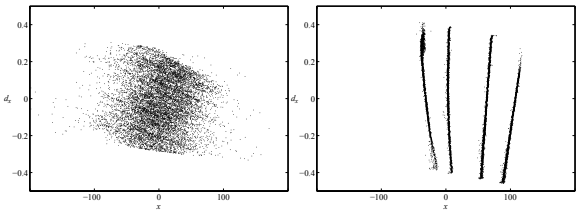


**Figure 4:** *Light field samplings of the rather random Cathedral glass (left) and the middle row of Spoons (right), visualized as scatter plot. On the vertical axis, the x-component of direction vector, $d_x$; on the horizontal axis, the x position. Shown is a subset of about 10000 rays.*

1% (Spheres) to more than half of the sensor's total resolution of 21 megapixels (Cathedral). The number of views relates to approximate centers of projection caused by bump-like structures. By using these transformers, we obtain a rather diverse set of light field samplings, illustrated in Fig. 4 for Cathedral and Spoons with position-vs.-angle scatter plots. This ray geometry is recovered through a versatile calibration procedure, which we will introduce in the following.

### 3.1. Calibration procedure

An integral imager augments a *primary camera* with a light field transformer, such as a free-form mirror surface or a structured glass pane. In order to perform light field imaging in this scenario, we need to establish the ray that is ultimately observed in each pixel position on the sensor of the primary camera. As we recover this map independently per-pixel, we will discuss the calibration of a single ray in the following.

We describe the ray geometry using the usual two-plane parameterization [LH96] as intersection coordinate pairs $(u,v)$ and $(s,t)$ of the ray against a near and far plane, respectively. As [KL08] have shown, intersection coordinates $(u,v)$ of rays with planes of known geometry can be determined by displaying structured light scanning patterns on a flatscreen monitor. Moving the monitor to a second position permits us to recover $(s,t)$ as well. In practice, we use

© The Eurographics Association 2015.

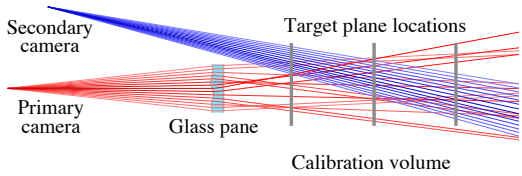
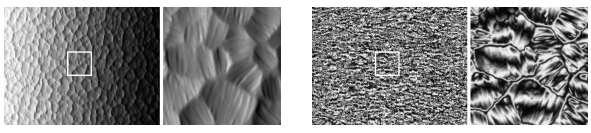



**Figure 5:** *Drawing of our calibration geometry. The calibration target, an LCD monitor, is placed in different target planes and displays structured light patterns of varying orientation and frequency to recover intersection coordinates between viewing rays (red) and the chosen plane. The number of target planes is arbitrary, but in our measurements 3 have proven sufficient. To register the different positions of the calibration target, a secondary camera is used (blue).*

three different plane positions and perform a least-squares fit for the ray coordinates. The quality of this fitting is estimated for each ray through the length of an error vector $\mathbf{e}(\mathbf{r})$ which is calculated between the measured and a interpolated point on the middle screen position. The interpolated point is given by a line which is defined by the measured points on the nearest and farthest planes and its intersection with the middle plane. For the results shown in this article, we have used horizontal and vertical sine patterns [SS03] on up to five different scales with five phase shifts each. Sinusoidal patterns have been chosen, as the reconstruction of the phase of a sinusoidal signal provides robust results, as demonstrated by [KL08] comparing different coding strategies. The gamma curve of the television screen has been recorded in a preparation step enabling a linearized output and recording of the emitted grayscale values. Examples of input images for Cathedral are shown in Fig. 6. As monitor, we used a 55-inch digital TV with 3840×2160 pixels placed between 100 and 215 cm from the primary camera. In absence of a light field transformer, we achieve a prediction accuracy of less than one monitor pixel distance, approximately 300$\mu$m, for more than 79% of the observed rays on a monitor position in between the calibration planes.

Note that this approach does not require to move the monitor to exact a-priori known locations. We use a secondary XIMEA MQ042CG-CM machine vision camera which is calibrated and observes the monitor directly, and compute the plane position with a robust RANSAC fit using the same sinusoids. Fig. 5 illustrates our setup.



(a) (b)

**Figure 6:** *Sinusoidal patterns observed through cathedral glass, after amplitude normalization: (a) horizontal gradient, 1 cycle / image; (b) vertical gradient, 8 cycles / image.*

## 4. Scattered light field analysis and rendering

### 4.1. Rendering

Rendering with a virtual camera in case of scattered light field data requires us to interpolate radiances $L(u,v,s,t)$ from all radiances $L(u_i,v_i,s_i,t_i)$ observed by the primary camera. We reparameterize the light field so that $(u,v)$ coordinates are relative to an aperture plane containing the optical center of the virtual view, and $(s,t)$ are relative to its focal plane. Either plane is perpendicular to the optical axis. We can estimate

$$L(u,v,s,t) = \frac{\sum_i w_i L(u_i,v_i,s_i,t_i)}{\sum_i w_i} \tag{1}$$

with bilateral Gaussian weights

$$w_i = \frac{1}{2\pi\sigma_a\sigma_f} \exp\left(-\frac{(u-u_i)^2+(v-v_i)^2}{2\sigma_a^2}\right) \cdot \exp\left(-\frac{(s-s_i)^2+(t-t_i)^2}{2\sigma_f^2}\right). \tag{2}$$

In this formulation, $\sigma_f$ controls the sampling kernel in the image plane of the virtual camera, $\sigma_a$ controls the aperture. For small values of $\sigma_a$, we obtain an all-in-focus picture, for larger values, geometry located in the focal plane stays sharp, while the remainder of the scene is rendered with a soft bokeh depth of field effect. This approach is closely related to the light field rendering described in [IMG00]. Because there are no assumptions regarding the light field data structure, we do not separate the aperture sampling from the interpolation of missing ray information. It would be inefficient to integrate all observed rays for all output pixels. As usually $\sigma_f \ll \sigma_a$, we found it sufficient to integrate only over the 500 rays which intersect the focal plane closest to its intersection point with $(u,v,s,t)$, located with a 2D $k$-d tree storing the radiance values indexed by the intersection coordinates.

### 4.2. Depth reconstruction

Reconstructing depth from scattered light field data is a very challenging task. Indeed, the light having been reflected at or refracted through complex surfaces, the ray calibration might be not exact, and ray density may vary wildly between 3D points. While numerous robust methods for 3D reconstruction are already available, they are often tailored to the scenario of several wide-baseline views [SCD*06]. The ones more specialized to light fields usually rely on having either densely sampled sets of views for epipolar plane image analysis [KZP*13, WG14], or a dense focal stack to estimate shape from focus [NN94]. To our knowledge, there is no method available yet which can be directly applied to our scenario of a unordered collection of rays.

We therefore carefully construct an appropriate cost volume related to the plane-sweeping idea [Col96] which is able to deal with sparse or missing data. A global optimum is obtained in the functional lifting framework [PCBC10] to ob-

© The Eurographics Association 2015.

**Figure 7:** *Using the bilaterally weighted rendering scheme from Sec. 4.1, we produce depth-of-field renderings for various scenes taken through the Spoons (top row) and Cathedral (bottom rows) cameras. From left to right, the input view and refocused rendering on near and far planes are shown. Please see Fig. 9 for results obtained using our depth-guided rendering scheme.*

tain a good geometry proxy in form of a depth map for an individual view, see Fig. 8.

First, we choose a camera and a stack of depth layers parallel to the image plane $\Omega$ of the camera with distances $\{z_1, \dots z_k\}$ from the center of projection, so that a large quantity of measured rays passes through all the planes. For a choice of pixel $x \in \Omega$ and depth layer $z$, we compute a cost $\rho(x, z_i)$ to quantify how well a surface at this location is in accordance with the set of measured rays.

Further, we construct a layer image $I_z$, which can be thought of as a rendering of the measured light field focused on the corresponding plane. For each pixel $x$, we compute the projection $A_{x,z}$ onto that particular $z$-plane, which will be a rectangle. We then collect the set of all rays $R_{x,z}$ of the light field passing through $A_{x,z}$, and compute the weighted average over all $\mathbf{r} \in R_{x,z}$ of the radiances $L(\mathbf{r})$, with weights equal to the confidence $w(\mathbf{r}) \sim \|\mathbf{e}(\mathbf{r})\|^{-1}$ in its measurement, see Sec. 3.1. The result will be the radiance $L_z(x)$ of the layer image at $x$. See Fig. 10 for a few examples.

The assignment cost $\rho(x, z)$ is then the weighted standard deviation

$$\rho(x,z) = \sqrt{\frac{\sum_{\mathbf{r}\in R_{x,z}} w(\mathbf{r}) \|L(\mathbf{r}) - L_z(x)\|^2}{\sum_{\mathbf{r}\in R_{x,z}} w(\mathbf{r})}}, \qquad (3)$$

of all radiances. If there is only one ray sample available, the above cost does not give a useful value. Thus, as a post-processing step, the cost function is completed by layer-wise application of total variation inpainting, which is available in the open source library `cocolib` for continuous convex optimization [GSC12]. We now have defined a cost volume $\rho(x, z)$, which gives an assignment cost for every point $x$ in the image plane and every depth layer $z$. This allows us to finally solve for a depth map $d$ on $\Omega$ by minimizing

$$E(d) = \int_\Omega \|\nabla d(x)\| + \lambda \rho(x, d(x)) \mathrm{d}x, \qquad (4)$$

thus minimizing the total assignment cost while regularizing with the total variation of $d$ to obtain a smooth depth map. Smoothness can be controlled with the parameter $\lambda > 0$. Using the method of calibration and the idea of functional lifting, the above energy can be minimized to global optimality, although the data term is non-convex [PCBC10]. The central idea is to reformulate the problem in a higher-dimensional space (hence the name "lifting"), where the equivalent energy turns out to be convex. Being able to find a true global optimum is an important property in our context, since we have to deal with input which is less than ideal for 3D correspondence search. Thus, the data term is highly irregular and noisy, with the functional having many local minima corresponding to bad reconstruction results. We use the CUDA implementation of [PCBC10] for energies of the form (4) also available in `cocolib`. A smoothness parameter between $\lambda = 0.5$ and $\lambda = 2$ usually achieves reasonable results, note that this is the only free parameter of the method.

From the reconstructed depth map $\hat{d}$ and the layer images, an all-in-focus rendering $I_f$ can be generated by selecting for each $x \in \Omega$ the radiance from the corresponding depth image, i.e. $I_f(x) := L_{\hat{d}(x)}(x)$. See Fig. 10 for results for depth and all-in-focus renderings, note that this approach is coarse and the image is quite noisy.

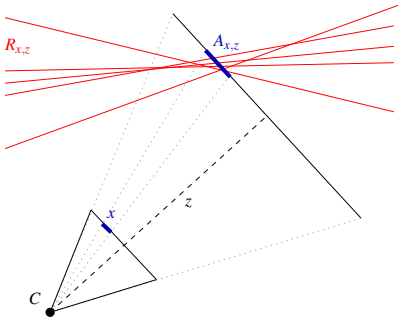



**Figure 8:** *In order to reconstruct a geometry proxy, we compute a depth map from a reference camera view C. For each pixel x in the reference view and each depth layer z, an assignment cost $\rho(x, z)$ is computed from the set of rays $R_{x,z}$ in the light field that intersect the pixel footprint $A_{x,z}$ on the depth layer, a lower value indicating higher agreement on the pixel color. The resulting cost function and a regularization term are globally optimized to obtain a smooth and accurate depth map, see Fig. 10.*

In order to obtain renderings of higher quality, one can also employ a slightly modified version of the bilateral rendering approach in Sec. 4.1. First, the sampled rays for the light field are sorted into individual 2D trees per depth

© The Eurographics Association 2015.

layer according to their intersection with the scene geometry. Then, we sample the aperture and generate viewing rays through the focal plane, accumulating the radiance in the first intersected layer according to Eq. 1. Results can be observed in Fig. 9.

**Figure 10:** *Depth reconstruction results for scenes captured through the Spoons and Cathedral imagers. Small images (left to right, top to bottom): input scene, reconstructed depth as seen from a single camera view (brighter = closer), depth layer images for the center and the back of the scene, respectively. Large image: all-in-focus rendering, using the color from the reconstructed depth layer.*

## 5. Results

In order to test our acquisition, analysis and rendering pipeline, we have captured a variety of static and dynamic scenes, evaluating the capabilities of the Cathedral, CrossRib, Spheres and Spoons cameras. We show representative examples rendered with the bilateral radiance estimation technique from a single shot with the primary camera (Canon EOS 5D Mark II). Please refer to the supplemental materials for additional results, including a reconstruction of the dynamic Mobile scene.

### 5.1. Reflective transformers

**Spheres.** The scattered light field in the top row of Fig. 12 is characterized by a large number of reflectors with simple geometry. While the ray geometry can be reconstructed faithfully, resolution is limited and each sphere captures only a very small number of rays. Thus, a rendering of the reflected scene with different focus planes is possible, but details can of course not be recovered.

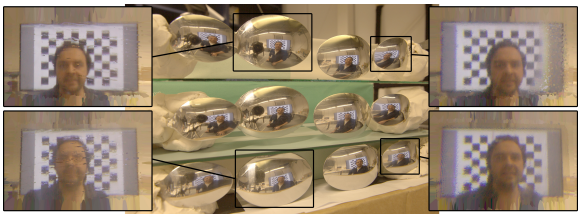


**Figure 11:** *From the input image shown in the center, we can extract multiple rectified views. Note the parallax between head and background.*

**Spoons.** The scene in the first row of Fig. 7 has a relatively small number of reflectors, each imaged with larger resolution than in the previous example. Each spoon can be considered as a camera capturing a distorted view of the scene, and indeed, we can compute undistorted views for the Spoons camera at a resolution comparable to the one at which the spoon appears in the scene image (Fig. 11). The wide baseline permits the reconstruction of a depth map (Fig. 10) which smoothly transitions from near to far.

### 5.2. Refractive transformers

**CrossRib.** The scattered light field in the bottom row of Fig. 12 is acquired through a sheet of glass with a periodic cross-rib pattern. Thus, on large scale, the ray distribution is comparable to the one from a plenoptic camera. Although good reconstruction results might be expected, they are limited by two factors: gaps in the ray coverage and high curvature structures on top of the basis shape which limits calibration resolution. Despite these challenges, we obtain a faithful reconstruction of the plant and can refocus on details of the branch and leaf structure.

**Cathedral.** The Cathedral camera creates a locally chaotic ray distribution. In total, though, it achieves a rather dense and even ray coverage (see Fig. 4). This permits surprisingly precise renderings (middle and bottom row in Fig. 7). Even depth reconstruction is possible, Fig. 10: in particular, we recover a reasonable set of depth labels even for the complex geometry of the plant, enabling all-in-focus rendering to globally increase sharpness.

## 6. Discussion and outlook

**What is possible?** Our setup is able to calibrate almost arbitrary reflective or refractive objects. By characterizing the irregular ray-space distortions of our optical system, we can reconstruct the ray geometry to sufficient precision to enable the extraction of multiple views, depth estimation, and synthetic aperture rendering.

**What are the limiting factors?** Our imaging optics, by definition, are not made for the purpose of light field imaging. As expected from the beginning, surface roughness, absorption and scattering degrade the image quality significantly.

© The Eurographics Association 2015.

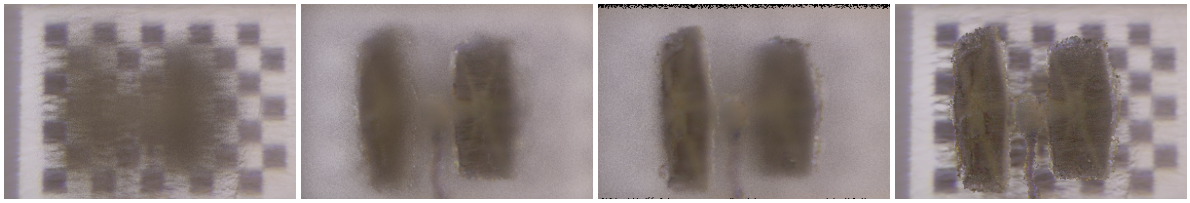


**Figure 9:** *Combination of the depth layer composition from Sec. 4.2 with aperture sampling based on the bilateral filtering scheme introduced in Section 4.1. From left to right, focus on background, on right wing, on left wing, and all-in-focus rendering. Compare to Fig. 10, which shows the unfiltered reconstructions used to estimate the depth map.*

**Figure 12:** *The Spheres transformer (top row) produces individual views that cover an almost spherical field of view, but view few pixels per sphere. The CrossRib glass (bottom row) bundles rays in a way that leaves large gaps in the acquired light field data. From the input image (left), renderings focusing on the near (center) and far plane (right) demonstrate nevertheless, that refocusing effects are still clearly visible.*

Our light field transformers have no radial arrangement of prisms that would counter the divergence of viewing rays, like Georgiev et al.'s setup does [GZC*06]. This drastically reduces the number of usable rays in the input images. In the case of the Spheres camera, the person's head only covers about $10\times15$ pixels per view.

Finally, the focus challenge illustrated in Fig. 3 and discussed in Sec. 3 limits our primary camera to a pinhole-like setting. While it would be possible to obtain reasonably good focus of an object through a single symmetric lens, no single focal setting could achieve this for lenses of different radius.

**What might become possible, and how?** Regarding focus, if the optical surface is "smooth enough", one will probably be able to focus more deeply into the scene. As we learn more about the imaging properties of improvised light field transformer, we expect to find a sweet spot for the focal setting that minimizes scene defocus and yields higher effective resolution than what is currently possible.

By identifying key properties of the imaging system, and making better use of them, we expect that significant progress can be made in calibration and use of other light field transformers. As of now, our calibration is ray-by-ray and does not exploit any high-level knowledge about the light field transformer. Glass sheets, for instance, can be described by two height fields, a fact that could not only be used to constrain the space of possible ray mappings, but also to allow for movement of the primary camera with respect to the transformer after calibration.

Finally, many of our problems could be tackled if suitable signal processing techniques were available. The most important recurring challenge is blur from various sources, i.e., the observation of light path mixtures. We expect that by generalizing deconvolution techniques to arbitrarily scattered ray-space data, a major leap in reconstruction quality could be obtained.

## 7. Conclusion

In this work, we have found a variety of household items suitable for integral imaging. Thanks to a ray-space calibration framework, we can now use light field transforming objects as diverse as glass sheets or kitchen spoons to capture light fields of scenes in a single shot. The light fields obtained from such light field cameras are degraded by blur and scattering effects and their sampling is highly irregular and sparse. By developing a layer-based depth estimation scheme, we were able to make the best of our data and use it for rendering purposes, producing effects such as post-capture parallax or refocusing.

We have drawn from a rich tradition of research that attempts to replace carefully designed and highly specialized capture setups with a combination of casually captured data, careful calibration and computational reconstruction. With our contribution of a lightweight capture paradigm for integral images, we hope that this paper will serve as a source of inspiration for future work.

**Acknowledgements:** *This research was financially supported by the Juniorprofessorenprogramm Baden-Württemberg and the X-Rite Chair and Graduate School for Digital Material Appearance.*

## References

[AB91] ADELSON E. H., BERGEN J. R.: The plenoptic function and the elements of early vision. *Computational models of visual processing 1*, 2 (1991). 1

[Col96] COLLINS R.: A space-sweep approach to true multi-image matching. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on* (Jun 1996), pp. 358–363. 4

© The Eurographics Association 2015.

[CTCS00] CHAI J.-X., TONG X., CHAN S.-C., SHUM H.-Y.: Plenoptic sampling. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (2000), SIGGRAPH '00, pp. 307–318. 1

[DLD12] DAVIS A., LEVOY M., DURAND F.: Unstructured light fields. *Comp. Graph. Forum 31*, 2 (May 2012), 305–314. 2

[FKR13] FUCHS M., KÄCHELE M., RUSINKIEWICZ S.: Design and fabrication of faceted mirror arrays for light field capture. *Computer Graphics Forum 32*, 8 (2013), 246–257. 2

[GGSC96] GORTLER S., GRZESZCZUK R., SZELISKI R., COHEN M.: The lumigraph. In *Proc. 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)* (1996), pp. 43–54. 1, 2

[GSC12] GOLDLUECKE B., STREKALOVSKIY E., CREMERS D.: The natural vectorial total variation which arises from geometric measure theory. *SIAM J. Imaging Sciences 5*, 2 (2012), 537–563. 5

[GZC*06] GEORGIEV T., ZHENG K. C., CURLESS B., SALESIN D., NAYAR S., INTWALA C.: Spatio-angular resolution tradeoffs in integral photography. In *17th Eurographics Conference on Rendering Techniques* (2006), EGSR '06, Eurographics Association, pp. 263–272. 2, 3, 7

[HFI*08] HULLIN M. B., FUCHS M., IHRKE I., SEIDEL H.-P., LENSCH H. P. A.: Fluorescent immersion range scanning. *ACM Trans. Graph. (Proc. SIGGRAPH) 27*, 3 (Aug. 2008), 87:1–87:10. 2

[HP03] HAN J. Y., PERLIN K.: Measuring bidirectional texture reflectance with a kaleidoscope. *ACM Trans. Graph. (Proc. SIGGRAPH)* (2003), 741–748. 2

[IKL*08] IHRKE I., KUTULAKOS K., LENSCH H., MAGNOR M., HEIDRICH W.: State of the art in transparent and specular object reconstruction. In *EUROGRAPHICS STAR* (2008). 2

[IMG00] ISAKSEN A., MCMILLAN L., GORTLER S. J.: Dynamically reparameterized light fields. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), pp. 297–306. 4

[IWLH11] IHRKE I., WETZSTEIN G., LANMAN D., HEIDRICH W.: State of the art in computational plenoptic imaging. In *EUROGRAPHICS STAR* (2011). 1

[KL08] KAMMEL S., LEÓN F. P.: Deflectometric measurement of specular surfaces. *IEEE T. Instrumentation and Measurement 57*, 4 (2008), 763–769. 3, 4

[KS08] KUTULAKOS K. N., STEGER E.: A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision 76*, 1 (2008), 13–29. 2

[KZP*13] KIM C., ZIMMER H., PRITCH Y., SORKINE-HORNUNG A., GROSS M.: Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph. (Proc. SIGGRAPH) 32*, 4 (2013), 73:1–73:12. 1, 4

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proc. 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)* (1996), pp. 31–42. 1, 2, 3

[LNA*06] LEVOY M., NG R., ADAMS A., FOOTER M., HOROWITZ M.: Light field microscopy. *ACM Trans. Graph. (Proc. SIGGRAPH)* (2006), 924–934. 1

[MRK*13] MANAKOV A., RESTREPO J. F., KLEHM O., HEGEDÜS R., EISEMANN E., SEIDEL H.-P., IHRKE I.: A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. *ACM Trans. Graph. (Proc. SIGGRAPH) 32*, 4 (2013), 47:1–47:14. 2

[MTK*11] MUKAIGAWA Y., TAGAWA S., KIM J., RASKAR R., MATSUSHITA Y., YAGI Y.: Hemispherical confocal imaging using turtleback reflector. In *Proc. ACCV 2010*, Lecture Notes in Computer Science. Springer, 2011, pp. 336–349. 2

[Ng05] NG R.: Fourier slice photography. *ACM Trans. Graph. (Proc. SIGGRAPH)* (2005), 735–744. 1, 2, 3

[NN94] NAYAR S., NAKAGAWA Y.: Shape from focus. *IEEE Trans. Pattern Analysis and Machine Intelligence 16*, 8 (1994), 824–831. 4

[PCBC10] POCK T., CREMERS D., BISCHOF H., CHAMBOLLE A.: Global Solutions of Variational Models with Convex Regularization. *SIAM J. Imaging Sciences* (2010). 4, 5

[SCD*06] SEITZ S., CURLESS B., DIEBEL J., SCHARSTEIN D., SZELISKI R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Int. Conf. Computer Vision and Pattern Recognition* (2006), pp. 519–526. 4

[SCK10] SHAN Q., CURLESS B., KOHNO T.: Seeing through obscure glass. In *Proceedings of the 11th European Conference on Computer Vision: Part VI* (2010), ECCV'10, pp. 364–378. 2

[SS03] SCHARSTEIN D., SZELISKI R.: High-accuracy stereo depth maps using structured light. In *Int. Conf. Computer Vision and Pattern Recognition* (2003), vol. 1, pp. I–195–I–202 vol.1. 4

[SYGM03] STEWART J., YU J., GORTLER S. J., MCMILLAN L.: A new reconstruction filter for undersampled light fields. In *Proceedings of the 14th Eurographics Workshop on Rendering* (2003), EGRW '03, pp. 150–156. 2

[TF14] TORRALBA A., FREEMAN W.: Accidental pinhole and pinspeck cameras. *International Journal of Computer Vision 110*, 2 (2014), 92–112. 2

[THMR13] TAO M. W., HADAP S., MALIK J., RAMAMOORTHI R.: Depth from combining defocus and correspondence using light-field cameras. In *Computer Vision (ICCV), 2013 IEEE International Conference on* (2013), IEEE, pp. 673–680. 1

[TLGS05] TARINI M., LENSCH H., GOESELE M., SEIDEL H.-P.: 3D acquisition of mirroring objects using striped patterns. *Graphical Models 67*, 4 (2005), 233–259. 2

[VRA*07] VEERARAGHAVAN A., RASKAR R., AGRAWAL A., MOHAN A., TUMBLIN J.: Dappled photography: mask-enhanced cameras for heterodyned lightfields and coded aperture refocusing. *ACM Trans. Graph. (Proc. SIGGRAPH) 26*, 3 (2007). 2

[WG14] WANNER S., GOLDLUECKE B.: Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Analysis and Machine Intelligence 36*, 3 (2014), 606–619. 1, 4

[WIH13] WETZSTEIN G., IHRKE I., HEIDRICH W.: On plenoptic multiplexing and reconstruction. *International Journal of Computer Vision 101*, 2 (2013), 384–400. 2

[WJV*05] WILBURN B., JOSHI N., VAISH V., TALVALA E.-V., ANTUNEZ E., BARTH A., ADAMS A., HOROWITZ M., LEVOY M.: High performance imaging using large camera arrays. *ACM Trans. Graph. (Proc. SIGGRAPH)* (2005), 765–776. 2

[WORK13] WEINMANN M., OSEP A., RUITERS R., KLEIN R.: Multi-view normal field integration for 3d reconstruction of mirroring objects. *Proceedings of the International Conference on Computer Vision* (Dec. 2013), 2504–2511. 2

[WRH11] WETZSTEIN G., RASKAR R., HEIDRICH W.: Hand-held schlieren photography with light field probes. In *Computational Photography (ICCP), International Conference on* (April 2011), pp. 1–8. 2

© The Eurographics Association 2015.