Depth Fields: Extending Light Field Techniques to Time-of-Flight Imaging

Suren Jayasuriya
Cornell University
sj498@cornell.edu

Adithya Pediredla
Rice University
adithya.k.pediredla@rice.edu

Sriram Sivaramakrishnan
Cornell University
ss2462@cornell.edu

Alyosha Molnar
Cornell University
molnar@ece.cornell.edu

Ashok Veeraraghavan
Rice University
vashok@rice.edu

Abstract

A variety of techniques such as light field, structured illumination, and time-of-flight (TOF) are commonly used for depth acquisition in consumer imaging, robotics and many other applications. Unfortunately, each technique suffers from its individual limitations preventing robust depth sensing. In this paper, we explore the strengths and weaknesses of combining light field and time-of-flight imaging, particularly the feasibility of an on-chip implementation as a single hybrid depth sensor. We refer to this combination as depth field imaging. Depth fields combine light field advantages such as synthetic aperture refocusing with TOF imaging advantages such as high depth resolution and coded signal processing to resolve multipath interference. We show applications including synthesizing virtual apertures for TOF imaging, improved depth mapping through partial and scattering occluders, and single frequency TOF phase unwrapping. Utilizing space, angle, and temporal coding, depth fields can improve depth sensing in the wild and generate new insights into the dimensions of light's plenoptic function.

1. Introduction

The introduction of depth sensing to capture 3D information has led to its ubiquitous use in imaging and camera systems, and has been a major focus of research in computer vision and graphics. Depth values enable easier scene understanding and modeling which in turn can realize new computer vision systems and human-computer interaction. Many methods have been proposed to capture depth information such as stereo, photometric stereo, structured illumination, light field, RGB-D, and TOF imaging.

However depth cameras typically support only one depth sensing technology at a time which limits their robustness and flexibility. Each imaging modality has its own advan-

Feature	Stereo	Photometric Stereo	Structured Illumination	Light Field	Time-of-Flight	Depth Fields (proposed)
On-chip pixel implementation	No	No	No	Yes	Yes	Yes
Illumination source	Passive	Active	Active	Passive	Active	Active
High resolution depth maps	No	Yes	Yes	No	Yes	Yes
Texture needed for depth	Yes	No	No	Yes	No	No
Ambiguity in depth measurement	No	Yes	No	No	Yes	No

Table 1. Table that summarizes the relative advantages and disadvantages of different depth sensing modalities including the proposed depth fields.

tages and disadvantages for attributes such as on-chip implementation, cost, depth resolution, etc that are summarized in Table 1. We argue that hybrid 3D imaging systems which utilize two or more depth sensing techniques can overcome these individual limitations. Furthermore, a system that combines modalities with an on-chip implementation would be cost effective and mass producible, allowing ubiquitous robust depth sensing.

We propose combining light field and TOF imaging into a hybrid 3D imaging system. This system inherits light field advantages such as post-capture digital refocusing with TOF advantages of high resolution depth information and the mitigated multipath interference using coded signals. Further, light field and TOF imaging both have been implemented on-chip [10, 28], and we can design hybrid pixel structures to combine both modalities on-chip as well. Each modality has its relative disadvantages: depth from light fields require textured surfaces and is dependent on object distance for disparity, and single frequency TOF imaging suffers from phase wrapping and is limited to small aperture cameras with low shutter speeds. However, we show that combining light field and TOF imaging can alleviate all of these limitations.

We call this extension of spatio-angular information captured traditionally by light fields to TOF depth maps as **depth field imaging**. Our main contributions include:

- Formulation of depth field imaging as an extension of the light field framework for TOF imaging

- Methods to capture depth fields using camera arrays and single-shot camera systems.

We show that capturing depth fields leads to many new applications that improve robust depth sensing in the wild including:

- Digital refocusing of depth images and extended depth of field.

- Phase unwrapping for single frequency TOF imaging.

- Depth imaging through partial occluders.

- Depth imaging and refocusing past scattering media.

A larger vision for introducing depth fields is to have a layer of post-capture control for depth sensing which can combine synergistically with higher level algorithms such as structure from motion (SfM) [5], Simultaneous Localization and Mapping (SLAM) [35], and reconstructions from collections of online images [32] used for 3D reconstruction and scene modeling/understanding.

2. Related Work

We survey related work in light field, TOF, and fusion algorithms for depth imaging to show the context of depth sensing technologies that depth field imaging relates to.

Light Field Imaging captures 4D representations of the plenoptic function parametrized by two spatial coordinates and two angular coordinates, or equivalently as the space of non-occluded rays in a scene [11, 27]. Light fields are used for image-based rendering and modeling, synthesizing new viewpoints from a scene, and estimating depth from epipolar geometry. In the context of cameras, light fields have been captured by using mechanical gantries [26] or large dense camera arrays [40], or by single-shot methods including microlenses [1, 28], coded apertures [25], transmission masks [38], or diffraction gratings [15]. Light fields can extend the depth of field and use digital refocusing to synthesize different apertures in post-processing [28], thus enabling a level of software control after the photograph has been taken. We will exploit this control in depth field imaging.

Time-of-Flight Imaging works by encoding optical path length traveled by amplitude modulated light which is recovered by various devices including photogates and photonic mixer devices [2, 10, 23, 30]. While yielding high

resolution depth maps, single frequency TOF suffers from limitations including phase wrapping ambiguity and multipath interference caused by translucent objects and scattering media. Proposed techniques to overcome these limitations include phase unwrapping with multifrequency methods [29], global/direct illumination separation [19, 41], deblurring and superresolution [42], and mitigating multipath interference with post-processing algorithms [3, 4]. Recently, new temporal coding patterns for these sensors help resolve multiple optical paths to enable seeing light in flight and looking through turbid media [12, 13, 21]. Similar to this paper, camera systems have been proposed to fuse together TOF + stereo [44], TOF + photometric stereo [36], and TOF + polarization [20].

Fusion of depth maps and intensity images has been used to enable 3D reconstruction by explicit feature detection [14, 16]. Real-time interaction for camera tracking and 3D reconstruction have been demonstrated via KinectFusion [17]. While conceptually similar to depth fields by acquiring per-pixel values of depth and intensity, these fusion methods do not systematically control the spatio-angular sampling or transcend the traditional capture tradeoffs between aperture and depth of field for depth imaging. In this way, we hope that depth field algorithms can serve as the foundation upon which fusion algorithms can improve their reconstruction quality, leading vertical integration from camera control all the way to high level scene modeling and understanding.

3. Depth Fields

In this section, we combine the mathematical formulations of light field and TOF imaging into the concept of a depth field. We show both how to capture these fields and how to invert the forward model to recover light albedo, defined as the reflectance value of an object with respect to the active illumination, and depth as a function of 2D spatial coordinates and 2D angular coordinates. This approach is similar to Kim et al. [22] who capture depth maps for different perspective views, but they do not use TOF imaging or show applications such as digital refocusing or depth mapping through partial/scattering occluders.

To describe the forward model for capturing depth fields, we first briefly discuss the forward models for light field and TOF imaging.

3.1. Light Fields

Light fields are commonly parameterized by the two plane model $l(u, v, x, y)$ where (u, v) is the angular coordinates at the lens plane, and (x, y) are the spatial coordinates of the sensor plane [27]. The output of this function represents the radiance of the ray parametrized by its intersection with the two planes. The forward model for light
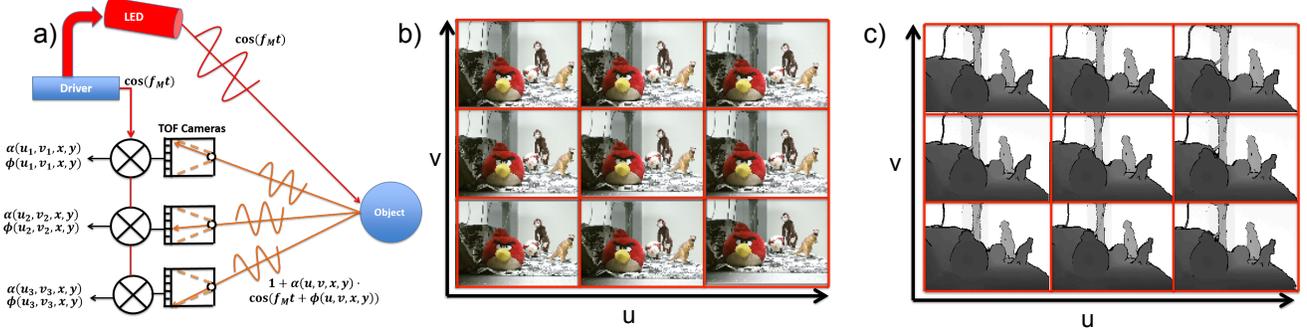
Figure 1. (a) Capturing a depth field conceptually using an array of TOF cameras, (b) albedo as 4D function of (u,v,x,y), (c) phase corresponding to TOF depth values as a 4D function of (u,v,x,y)

field capture has been modeled in [39] as follows:

$$i_{LF}(x,y) = \int_u \int_v m(u,v,x,y) \cdot l(u,v,x,y) du dv \quad (1)$$

where $i_{LF}(x,y)$ is the intensity measured by the detector and $m(u,v,x,y)$ is the modulation/multiplexing function that encodes the incoming light rays. The modulation function represents the different optical elements that could be used to sense the light field including pinholes ($m(u,v,x,y) = \delta(u,v,x,y)$), Fourier masks, random codes/masks, or diffraction gratings where the modulation functions are Gabor wavelets [15]. Discretizing the above equation, $\mathbf{i}_{LF} = \mathbf{M}\mathbf{l}$ where $\mathbf{i}_{LF}, \mathbf{l}$ are the vectorized images and light fields, and $\mathbf{M}$ is the modulation matrix, and both linear and nonlinear inversions can recover back the light field [39].

3.2. Time-of-Flight Imaging

In contrast, TOF is typically modeled using a cross-correlation between the incoming light signal and the reference code sent to the sensor. Given that incoming light is of the form: $1 + \alpha \cos(f_M t + \phi(x,y))$ where ϕ is the phase accumulated due to the optical path traveled from light source to object to camera and α is the albedo, the intensity at the sensor (normalized to integration time) is:

$$i_{TOF}(\tau,x,y) = (1 + \alpha(x,y)\cos(f_M t + \phi(x,y))) \otimes cos(f_M t)$$

$$\approx \frac{\alpha(x,y)}{2}\cos(f_M \tau + \phi(x,y)). \quad (2)$$

Here, τ is the cross-correlation parameter which controls the phase shift of the reference signal. By choosing different τ such that $f_M \tau = 0, \pi/2, \pi, 3\pi/2$, we can recover both the albedo α and the phase ϕ at each spatial location (x,y) using quadrature inversion:

$$\phi(x,y) = \tan^{-1}((i_{TOF}(\tfrac{3\pi}{2}) - i_{TOF}(\tfrac{\pi}{2}))/(i_{TOF}(\pi) - i_{TOF}(0))),$$

$$\alpha = \sqrt{(i_{TOF}(\tfrac{3\pi}{2}) - i_{TOF}(\tfrac{\pi}{2}))^2 + (i_{TOF}(\pi) - i_{TOF}(0))^2}. \quad (3)$$

Note that $d = \frac{c \cdot \phi}{4\pi f_M}$ can directly recover depth d from phase ϕ for TOF imaging.

3.3. Depth Fields

We now introduce the concept of the **depth field** as the ordered pair of albedo and depth (encoded in phase) (α, ϕ) that occurs at every (u,v,x,y) spatio-angular coordinate, i.e. $\alpha = \alpha(u,v,x,y), \phi = \phi(u,v,x,y)$. Note that depth fields are not recoverable from TOF measurements alone since TOF assumes a pinhole camera model, which sample ϕ and α at a particular fixed (u,v). We now describe the forward model of depth field imaging as follows:

$$i(\tau,x,y) = \int_u \int_v m(u,v,x,y) \cdot$$
$$(1 + \alpha(u,v,x,y)\cos(f_M t + \phi(u,v,x,y))) du dv$$
$$\otimes \cos(f_M t) \quad (4)$$

which is approximately

$$i(\tau,x,y) \approx \int_u \int_v m(u,v,x,y) \cdot$$
$$\frac{\alpha(u,v,x,y)}{2} \cdot \cos(f_M \tau + \phi(u,v,x,y)) du dv. \quad (5)$$

To invert this model, we take four measurements $f_M \tau = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ to get images $i(0), i(90), i(180), i(270)$ at each spatial location. Then we calculate $\mathbf{M}^{-1}i(\tau)$ to invert the light field matrix for each of these images (Note: this inverse can be either done at lower spatial resolution or using sparse priors or modeling assumptions to retain resolution). Thus we recover albedo and phase mixed together at every (u,v,x,y):

$$D' = \frac{\alpha(u,v,x,y)}{2} \cdot \cos(f_M \tau + \phi(u,v,x,y)). \quad (6)$$

To unmix the albedo and phase, we can perform quadrature inversion on D' for $f_M\tau = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ as before in TOF to recover the depth field.

4. Methods to Capture Depth Fields

We describe the potential for single-shot capture of depth fields (Note: single-shot is a misnomer since 4 phase measurements are performed per shot, however such functionality can be built into hardware to work in a single exposure). As in most light field sensors, we can align microlenses above CMOS TOF sensors such as photogates, photonic mixer devices, etc. Doing so allows sampling the angular plane by sacrificing spatial resolution at the sensor plane. The main lens can widen its aperture, allowing more light transmission while each of the sub-aperture views underneath the microlenses maintains a large depth of field [28]. This is advantageous since existing TOF cameras sacrifice exposure time to keep a small aperture and large depth of field. One limitation is the need for fine optical alignment of the microlenses at the conjugate image plane in the camera body.

Another depth field sensor can use amplitude masks between the main lens and the sensor plane of photogates to filter incoming angular rays [38]. While allowing less light transmission as microlenses, masks can be designed with different coding patterns for improved reconstruction of the depth field and can be flexibly interchanged within the camera body unlike fixed optical elements. We note a similar technique from [9] which uses a coded aperture in front of LIDAR system to extend the system's depth of field.

We also propose a fully integrated CMOS pixel design that does not require alignment of external optical elements: integrated diffraction gratings over interleaved photogates similar to [31]. This sensor works by diffracting the incoming light to form a Talbot pattern that is imaged by the photogates underneath similar to other diffractive sensors [15]. Note that this pixel can achieve better light efficiency with phase gratings and reduce its pixel size with interleaved photogates while maintaining the advantages of CMOS integration for cost and mass-production.

We outline the design of these concept pixels in Figure 2. The only image sensor fabricated to date capable of capturing depth fields in a single shot (that we know of) integrates metal diffraction gratings over single photon avalanche diodes [24], however this sensor is used for lensless fluorescence imaging. All these single-shot methods sacrifice spatial resolution to multiplex the incoming depth field.

Since fabricating a CMOS sensor takes significant time and resources, in this paper we motivate the need for depth field imaging using a custom acquisition setup. This acquisition setup captures depth fields at high spatial resolution by moving a TOF camera on a two axis stage sequentially
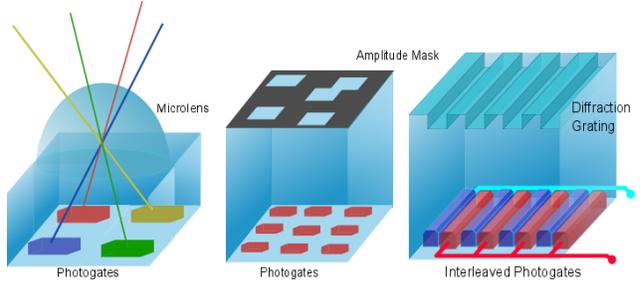

Figure 2. Pixel designs for single-shot camera systems for capturing depth fields. Microlenses, amplitude masks, or diffraction gratings are placed over top of photogates to capture light field and TOF information simultaneously.
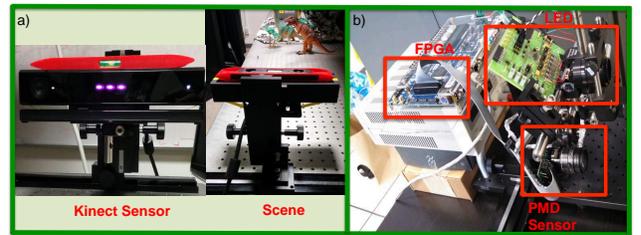

Figure 3. Setup to capture depth fields in practice. (a) A Kinect is placed on a XY translation stage on an optical bench, and a representative imaging scene, (b) PMD sensor with FPGA for code generation and LED setup as in [21]

in the (u, v) plane, as if having an array of TOF cameras, to scan a depth field. See Figure 1 for a schematic depiction of depth field capture.

5. Experimental Setup

In this section, we describe the depth field acquisition setup that scans a TOF sensor. We choose this approach since existing TOF sensors have limited resolution so single-shot methods would result in even smaller spatial resolution and the need of precise alignment of microlenses or masks. Our system has some limitations including a bulky setup and static scene acquisition, but we still demonstrate advantages of depth field imaging.

We move a TOF sensor on a two axis stage at different (u, v) positions. We utilize both the Microsoft Kinect One which has a 424 x 512 depth resolution (see Figure 3), and a custom PMD sensor of 160 x 120 resolution which enables us to send custom modulation codes directly to the silicon. The PMD sensor setup is the same as that described in [21]. All depth fields were captured with 1" spacing in the (u, v) plane at the coarse resolution of 5x5.

Figure 4. a) Captured scene, b-e) Digital refocusing on different focal planes for the depth map of the scene, showing how depth field imaging can break the tradeoff between aperture and depth of field for range imaging

6. Applications of Depth Fields

In this section, we highlight new applications of depth field imaging.

6.1. Synthetic Aperture Refocusing

One main disadvantage of TOF imaging is the necessity of a small aperture for large depth of field to yield accurate depth values. Having a shallow depth of field or wide aperture causes optical blur which corrupts TOF depth values. However, a small aperture limits the shutter speed and increases the acquisition time for these systems. In contrast, light field imaging breaks this tradeoff between depth of field and aperture size by using synthetic aperture refocusing. A plenoptic sensor with microlenses above its pixels can open its aperture and allow more light transmission while keeping the sub-aperture images beneath the microlenses in-focus, albeit at the loss of spatial resolution. After capture, one can digitally refocus the image, thus extending the depth of field by shearing the 4D light field and then summing over (u, v) to synthesize images with different focal planes [28].

Similarly, we show that the same techniques can be applied to depth fields. In Figure 4, we show digital refocusing of the 4D $\phi(u, v, x, y)$ information by applying the same shear and then average operation [28]. We are able to synthesize capture through a large virtual aperture for the scene which has not been shown in depth maps before, and may be combined with wide aperture light intensity images for enhanced artistic/photographic effect. In addition, this validates that single-shot depth field sensors such as TOF sensor with microlenses can allow more light through the aperture, thus increasing exposure while maintaining the same depth of field. This enables decreased acquisition time for TOF sensors at the expense of computationally recovering the lost spatial resolution and depth of field in post-processing algorithms. We note that [9] also showed extended depth of field for a LIDAR system using a coded aperture, but they don't extend their framework to show applications such as digital refocusing.

6.2. Phase wrapping ambiguities

One main limitation for single frequency TOF is that the phase has 2π periodicity, and thus depth estimates will wrap around the modulation wavelength. For modulation frequencies in the tens of MHz, this corresponds to a depth range of a few meters, which can be extended further by using multiple frequencies [6, 29] or phase unwrapping algorithms [7]. However, as modulation frequencies scale higher, phase wrapping becomes more severe.

We observe that capturing depth fields at a single modulation frequency also allows us to unwrap the phase periodicity by utilizing inherent epipolar geometry from different viewpoints. We use the depth from correspondence algorithm from [34] which is coarse and distance dependent, but does not suffer from phase wrapping, and thus can unwrap the depth measurements given by TOF.

In Figure 5, we simulate the Cornell Box scene and capture a depth field using the ray tracer Mitsuba [18]. We simulate phase wrapping and calculate depth from correspondence. In order to perform phase unwrapping, we select a continuous line in the image (the side wall in this scene) to determine the number of times the TOF image wraps upon itself in the scene. We use this mapping to match the wrapped TOF depth values to the depth values from correspondence, leading to unwrapped TOF depth values for the entire image as shown in Figure 5d. We also use a median filter to alleviate edge discontinuities in calculating depth from correspondence.

In the Microsoft Kinect we use for capturing depth fields, the maximum modulation frequency is 30MHz, which makes showing phase wrapping difficult on a standard optical bench. Thus we change the bit settings on the Kinect TOF sensors from N bits to N-1 bits to simulate phase wrapping for a real scene (identical to the wrapping caused by periodicity at a higher modulation frequency of 60MHz). We show the results of our phase unwrapping algorithm in Figure 6. Note that the reconstruction quality is limited by the lack of a good fiducial line in the scene that clearly corresponds light field depths to TOF wrapped depths. This is a limitation of our method, and it would be interesting to explore automatic calibration for phase unwrapping.
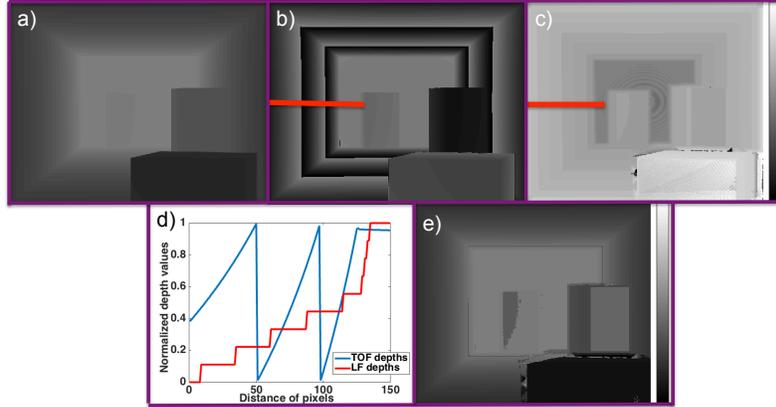
Figure 5. Phase unwrapping algorithm on synthetic data. a) Cornell box scene with ground truth depth values, b) a phase wrapped scene with red fiducial line for calibration marked, c) depth map given by light field correspondence algorithm. We identify the same calibration line in this scene for phase unwrapping, d) we map the TOF wrapped values to the depth values from correspondence for the given calibration line, e) unwrapped depth map.

6.3. Refocusing through partial occluders

The large synthetic aperture that can be synthesized by capturing 4D depth fields allows us to image past partial occluders in the foreground. This technique, which blurs out the foreground to reveal the background, has been shown in light fields [37] to look through bushes and plants. Note that in Figure 7, applying the same technique to the depth field works correctly for the albedo (one can see the object clearly while blurring out the foreground), but it does not work for the phase. This is because while visually we can perceptually tolerate some mixing of foreground and background color, this same mixing corrupts our phase measurements, leading to inaccurate depth values.

To solve this mixing problem when refocusing light fields, researchers have simply not added rays that are from the foreground when averaging over the sheared light field. A key assumption to their algorithm is that the foreground object rays are identified either by shooting continuous video [40] or by constructing an epipolar image, finding the corresponding depths, and then separating foreground relative to the background [37]. These algorithms are computationally expensive to identify the occluding objects pixels. We note that [43] use a combination of unstructured multiview stereo views and a depth sensor to refocus an intensity image through a partial occluder, and use the depth information to create a probabilistic model for occluders.

In contrast, we utilize the depths directly captured via TOF measurements to construct a histogram of depths observed in the scene as shown in Figure 7. We then can simply pick a foreground cluster using K-means or another computationally efficient clustering algorithm, which is faster than constructing an epipolar image, estimating line slopes, and then forming a histogram to do clustering. In
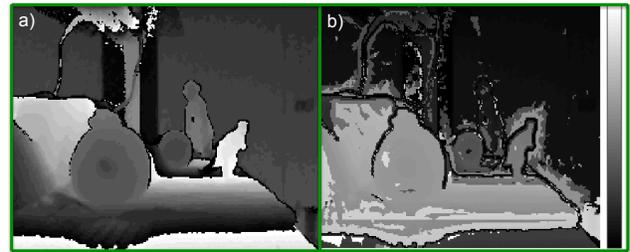

Figure 6. a) Phase unwrapping on real data with synthetic phase wrapping induced (due to prototype limitations). b) Recovered depth map. Notice that the monkey in back is not recovered because there does not exist a calibration marker line in the scene that extends all the way back in the TOF image.

Figure 7, you can see the results of our algorithm.

6.4. Refocusing past scattering media

While the previous subsection dealt with partial occluders that block the background for certain (u, v) viewpoints, other occluders such as scattering media or translucent objects are more difficult because they mix multiple phase measurements corresponding to different optical path lengths together at a single pixel. We approach the problem via coded TOF, specifically the depth selective codes by [33]. Mainly, we show how coded TOF extends the capabilities of our depth field camera systems by imaging past scattering media, and then use spatial information to perform digital refocusing. In Figure 8, we image through backscattering nets to get a depth field past the scattering media. We place nets in front of the camera to act as strong backscatterers, notice how the depth values are corrupted by the scattering. Using the depth selective codes, we can

image past the nets, and using multiple shots at different (u, v) viewpoints, we can capture the depth field beyond the nets and do digital refocusing. This demonstrates how depth field imaging can leverage the advantages of coded TOF techniques, and poses interesting questions of how to design the best possible codes for single-shot depth field imaging systems.

7. Discussion

Depth fields unify light field and TOF imaging as a single function of spatio-angular coordinates, and are useful for various applications. Besides the simple extensions of adding two imaging modalities, they can inform each other and make algorithms computationally more efficient and conceptually simpler, particularly in solving the problem of various occluders for light fields by using TOF information and breaking tradeoffs between aperture and depth of field for TOF cameras by adding light field capability. Improvements in light field depth estimation such as in [34] can also be applied for depth field cameras leading to improved depth resolution.

A key question that concerns depth field cameras is their pixel size which makes pixel multiplexing, including the sensor designs outlined in this paper, problematic. We note that TOF pixels have shrunk currently to 10um [2] which is only 10x larger than regular pixels (1um), and that technological advances such as stacked image sensors may help alleviate these multiplexing worries. However, the clear advantages for depth field cameras are applications where spatial resolution is not the limiting factor. This includes imaging systems that are limited by aperture (as argued in Section 6.1) and lensless imaging where spatial pixel layout is not a factor [8].

7.1. Limitations

Some limitations include long computational algorithms to recover lost spatial resolution for single-shot depth field cameras, or increased acquisition time for large TOF camera arrays or TOF cameras on mechanical gantries to scanline a depth field. Many applications provide partial robustness to depth sensing in the wild, but rely on modeling assumptions (foreground vs. background separation, scattering media is not immersing the object) that limits their deployment in real autonomous systems.

7.2. Future directions

One main future direction is to fabricate CMOS photogates with integrated diffraction gratings for an on-chip depth field sensor. This chip would be a lensless depth sensor for biomedical or 3D printing applications where a physical lens is bulky and prevents deployment of the image sensor in confined locations. Our analysis of a depth field shows that one can invert a farfield image using angular resolution similar to lensless lightfield cameras [8].

8. Conclusion

We presented the depth field, a representation of light albedo and phase from optical path length as function of space and angle. We developed the forward model to inform new ways to capture depth fields, and showed a myriad of applications that having this information possesess. We are inspired by the possibilitiy of image sensors that can perform hybrid depth imaging in general.

References

[1] E. H. Adelson and J. Y. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992. 2

[2] C. Bamji, P. O'Connor, T. Elkahtib, S. Mehta, B. Thompson, L. Prather, D. Snow, O. C. Akkaya, A. Daniel, A. D. Payne, T. Perry, M. Fenton, and V.-H. Chan. A 0.13um CMOS System-on-Chip for a 512x424 Time-of-Flight Image Sensor With Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC. *IEEE Journal of Solid-State Circuits*, 50:303–319, 2015. 2, 7

[3] A. Bhandari, A. Kadambi, and R. Raskar. Sparse linear operator identification without sparse regularization? applications to mixed pixel problem in time-of-flight/range imaging. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 365–369. IEEE, 2014. 2

[4] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar. Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization. *Optics Letters*, 39(6):1705–1708, 2014. 2

[5] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun. Structure from motion without correspondence. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000. 2

[6] D. Droeschel, D. Holz, and S. Behnke. Multi-frequency phase unwrapping for time-of-flight cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1463–1469. IEEE, 2010. 5

[7] D. Droeschel, D. Holz, and S. Behnke. Probabilistic phase unwrapping for time-of-flight cameras. In *Proc. of Joint 41st Int. Symposium on Robotics and 6th German Conference on Robotics (ISR/ROBOTIK)*, pages 1–7. VDE, 2010. 5
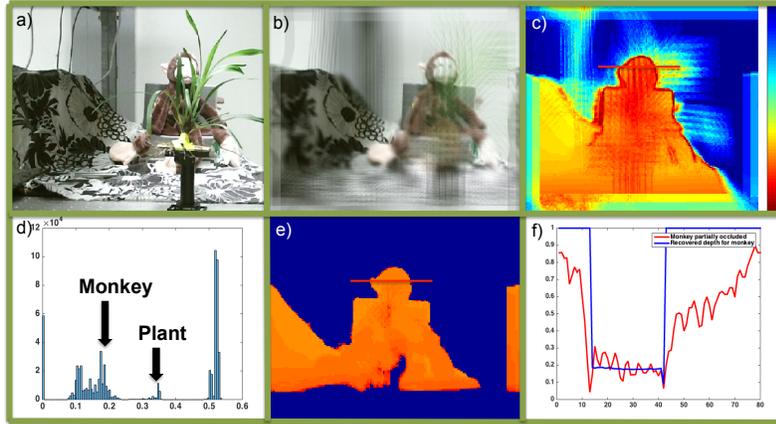
Figure 7. Refocusing in spite of foreground occlusions: (a) Scene containing a monkey toy being partially occluded by a plant in the foreground, (b) traditional synthetic aperture refocusing on light field is partially effective in removing the effect of foreground plants, (c) synthetic aperture refocusing of depth displays corruption due to occlusion, (d) histogram of depth clearly shows two clusters corresponding to plant and monkey, (e) virtual aperture refocusing after removal of plant pixels shows sharp depth image of monkey, (f) Quantitative comparison of indicated scan line of the monkey's head for (c) and (e)
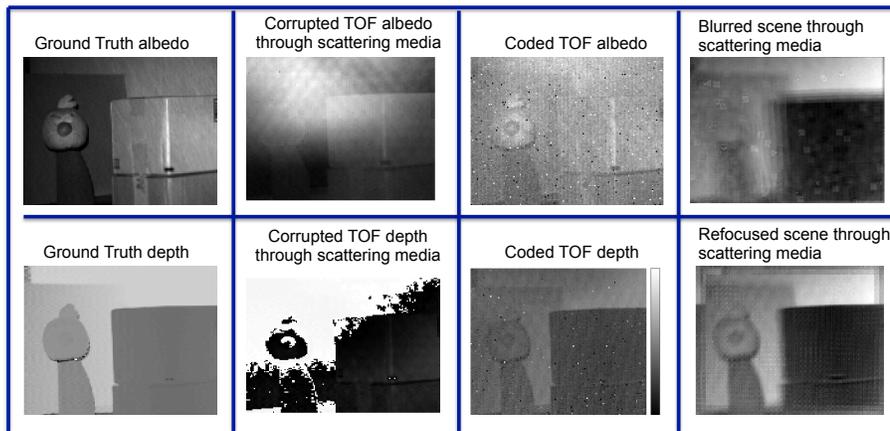

Figure 8. We use coding techniques from [33] to image beyond backscattering nets. Notice how the corrupted depth maps are improved using the codes. We show how digital refocusing can be performed on the images without the scattering occluders by combining depth fields with coded TOF.

[8] P. R. Gill, C. Lee, D.-G. Lee, A. Wang, and A. Molnar. A microscale camera using direct fourier-domain scene capture. *Optics Letters*, 36(15):2949–2951, 2011. 7

[9] J. Godbaz, M. Cree, and A. Dorrington. Extending amcw lidar depth-of-field using a coded aperture. *Asian Conference on Computer Vision*, pages 397–409, 2010. 4, 5

[10] S. B. Gokturk, H. Yalcin, and C. Bamji. A time-of-flight depth sensor - system description, issues and solutions. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 3:35, 2004. 1, 2

[11] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. SIGGRAPH*, pages 43–54, 1996. 2

[12] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (TOG)*, 32(4):45, 2013. 2

[13] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics Express*, 22(21):26338–26350, 2014. 2

[14] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *Int. J. Rob. Res.*, 31(5):647–663, Apr. 2012. 2

[15] M. Hirsch, S. Sivaramakrishnan, S. Jayasuriya, A. Wang, A. Molnar, R. Raskar, and G. Wetzstein. A Switchable Light Field Camera Architecture with Angle Sensitive Pixels and Dictionary-based Sparse Coding. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2014. 2, 3, 4

[16] B. Huhle, T. Schairer, P. Jenke, and W. Straíer. Fusion

of range and color images for denoising and resolution enhancement with a non-local filter. *Comput. Vis. Image Underst.*, 114(12):1336–1345, Dec. 2010. 2

[17] S. Izadi, A. Davison, A. Fitzgibbon, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, and D. Freeman. Kinect Fusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, page 559, 2011. 2

[18] W. Jakob. Mitsuba renderer, 2010. 5

[19] A. Kadambi, A. Bhandari, R. Whyte, A. Dorrington, and R. Raskar. Demultiplexing illumination via low cost sensing and nanosecond coding. *IEEE International Conference on Computational Photography (ICCP)*, 2014. 2

[20] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar. Polarized 3D: High-Quality Depth Sensing with Polarization Cues. In *IEEE International Conference on Computer Vision (ICCV)*, 2015. 2

[21] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph.*, 32(6):167, 2013. 2, 4

[22] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, 32(4):73:1–73:12, July 2013. 2

[23] R. Lange, P. Seitz, A. Biber, and S. Lauxtermann. Demodulation pixels in CCD and CMOS technologies for time-of-flight ranging. *Proceedings of SPIE*, 3965:177–188, 2000. 2

[24] C. Lee, B. Johnson, and A. Molnar. An on-chip 72x60 angle-sensitive single photon image sensor array for lens-less time-resolved 3-d fluorescence lifetime imaging. In *Symposium on VLSI Circuits*. IEEE, 2014. 4

[25] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26(3), July 2007. 2

[26] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas. Synthetic aperture confocal imaging. *ACM Trans. Graph.*, 23(3):825–834, Aug. 2004. 2

[27] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. SIGGRAPH*, pages 31–42, 1996. 2

[28] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11), 2005. 1, 2, 4, 5

[29] A. D. Payne, A. P. Jongenelen, A. A. Dorrington, M. J. Cree, and D. A. Carnegie. Multiple frequency range imaging to remove measurement ambiguity. *Proceedings of the 9th Conference on Optical 3-D Measurement Techniques*, 2009. 2, 5

[30] R. Schwarte. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). *Proceedings of SPIE*, 3100:245–253, 1997. 2

[31] S. Sivaramakrishnan, A. Wang, P. R. Gill, and A. Molnar. Enhanced angle sensitive pixels for light field imaging. In *Proc. IEEE IEDM*, pages 8–6, 2011. 4

[32] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM Trans. Graph.*, 25(3):835–846, July 2006. 2

[33] R. Tadano, A. Pediredla, and A. Veeraraghavan. Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography. In *IEEE International Conference on Computer Vision (ICCV)*, (Accepted) 2015. 6, 8

[34] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *IEEE International Conference on Computer Vision (ICCV)*, pages 673–680. IEEE, 2013. 5, 7

[35] S. Thrun. Exploring artificial intelligence in the new millennium. chapter Robotic Mapping: A Survey, pages 1–35. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003. 2

[36] C. Ti, R. Yang, J. Davis, and Z. Pan. Simultaneous time-of-flight sensing and photometric stereo with a single tof sensor. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4334–4342, 2015. 2

[37] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B. Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2331–2338. IEEE, 2006. 6

[38] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph. (SIGGRAPH)*, 26(3):69, 2007. 2, 4

[39] G. Wetzstein, I. Ihrke, and W. Heidrich. On Plenoptic Multiplexing and Reconstruction. *IJCV*, 101:384–400, 2013. 3

[40] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):765–776, 2005. 2, 6

[41] D. Wu, A. Velten, M. O'toole, B. Masia, A. Agrawal, Q. Dai, and R. Raskar. Decomposing global light transport using time of flight imaging. *Int. J. Comput. Vision*, 107(2):123–138, Apr. 2014. 2

[42] L. Xiao, F. Heide, M. O'Toole, A. Kolb, M. B. Hullin, K. Kutulakos, and W. Heidrich. Defocus deblurring and superresolution for time-of-flight depth cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2376–2384, 2015. 2

[43] T. Yang, W. Ma, S. Wang, J. Li, J. Yu, and Y. Zhang. Kinect based real-time synthetic aperture imaging through occlusion. *Multimedia Tools and Applications*, pages 1–19, 2015. 6

[44] J. Zhu, L. Wang, R. Yang, and J. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. 2