Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Optical design and development of a snapshot light-field laryngoscope

Shuaishuai Zhu
Peng Jin
Rongguang Liang
Liang Gao

SPIE.

Shuaishuai Zhu, Peng Jin, Rongguang Liang, Liang Gao, "Optical design and development of a snapshot light-field laryngoscope," *Opt. Eng.* **57**(2), 023110 (2018), doi: 10.1117/1.OE.57.2.023110.

Optical design and development of a snapshot light-field laryngoscope

Shuaishuai Zhu,[a,b] Peng Jin,[b,*] Rongguang Liang,[c] and Liang Gao[a,d,*]
[a]University of Illinois at Urbana–Champaign, Department of Electrical and Computer Engineering, Urbana, Illinois, United States
[b]Harbin Institute of Technology, Center of Ultraprecision Optoelectronic Instrument, Harbin, China
[c]University of Arizona, College of Optical Sciences, Tucson, Arizona, United States
[d]University of Illinois at Urbana–Champaign, Beckman Institute for Advanced Science and Technology, Urbana, Illinois, United States

Abstract. The convergence of recent advances in optical fabrication and digital processing yields a generation of imaging technology—light-field (LF) cameras which bridge the realms of applied mathematics, optics, and high-performance computing. Herein for the first time, we introduce the paradigm of LF imaging into laryngoscopy. The resultant probe can image the three-dimensional shape of vocal folds within a single camera exposure. Furthermore, to improve the spatial resolution, we developed an image fusion algorithm, providing a simple solution to a long-standing problem in LF imaging. [DOI: 10.1117/1.OE.57.2.023110]

Keywords: three-dimensional image acquisition; computational imaging; medical optics instrumentation.

Paper 171934L received Dec. 4, 2017; accepted for publication Feb. 7, 2018; published online Feb. 27, 2018.

1 Introduction

Currently, ~7.5 million people in the United States suffer from voice disorders due to either trauma or diseases. Human vocal fold vibration is a complex 3-D movement. An unusual 3-D shape of the vocal fold is a hallmark of a variety of vocal diseases, such as polyps, nodules, recurrent nerve paralysis, and cancer.[1,2] The acquisition of 3-D data can facilitate the theoretical modeling of vocal fold dynamics, providing insights into vocal fold pathology[3,4] and fundamental phonation research.[5,6]

The standard in-office methods for diagnosing voice disorders include videostroboscopy[7] and high-speed videoendoscopy.[8] Both techniques image only the horizontal movement of vocal folds. They cannot measure the movement of vocal folds along the air flow direction. Despite its vital importance, 3-D laryngeal imaging is currently only available via a few methods—namely, computed tomography (CT),[3,9] magnetic resonant imaging (MRI),[10,11] laser triangulation,[12,13] optical coherence tomography (OCT),[14,15] and structured illumination.[16,17] Although CT, MRI, and structured illumination can measure the full 3-D profile of vocal cords, the prolonged acquisition time restricts their use in imaging transient dynamics. In addition, CT and MRI are costly, and they require special operating rooms. Alternatively, laser triangulation and OCT feature a high acquisition speed. Nonetheless, it measures depths at only selected points or lines, resulting in a limited field of view. The lack of an en face image jeopardizes the sensitivity and specificity of diagnosis.

To enable fast imaging of vocal folds in 3-D, for the first time we introduce the paradigm of light-field (LF) imaging[18] into laryngoscopy. The resultant system, which we term a LF laryngoscope (LFL), can capture a volumetric image of vocal folds within a single snapshot. Rather than acquiring only two-dimensional (2-D) (x, y) (x, y, spatial coordinates) images, LF cameras acquire both the spatial and angular information of remittance. The resultant four-dimensional (4-D) (x, y, θ, ϕ) (θ, ϕ, 2-D light-emitting angles) datacube can be mathematically converted into a 3-D (x, y, z) (z, depth) image through postprocessing.[19] Since no scanning is required, the 3-D frame rate is limited by only the camera's data readout bandwidth. Although the LF imaging was first proposed by Lippmann[20] in 1908, not until the last decade were breakthroughs achieved in demonstrating its biomedical applications. For example, Bedard et al.[21] developed an LF otoscope for 3-D imaging of the tympanic membrane in vivo. Hassanfiroozi et al.[22] constructed an LF endoscope using a hexagonal liquid crystal lens array. Turola and Gruppetta[23] showed the potential of LF cameras in retinal imaging. Lastly, using an LF microscope, Prevedel et al. recorded neuronal activity in 3-D with an unprecedented frame rate.[24]

2 Light-Field Laryngoscope Design

We show the optical schematic and a photograph of the distal end of LFL probe in Figs. 1(a) and 1(b), respectively. The illumination light is guided to the tip of the probe through a multimode glass fiber (Thorlabs M28L01) and reflected toward the object by a right-angle prism (Edmund 84-506). The back-reflected light is collected by an objective lens (Edmund 49-657, $f = 18$ mm), forming an intermediate image S_1 at the distal end of a gradient-index (GRIN) lens [Gradient Lens Corporation COAT14-45-219; length, 219 mm (one pitch)]. This intermediate image is then relayed by the GRIN lens to its proximal end, followed by being magnified by an optical system which consists of a microscope objective (Nikon CF Plan, $f = 40$ mm, NA = 0.13) and a tube lens (Thorlabs AC254-100-A, $f = 100$ mm). The magnified image is directed toward two-imaging channels by a beam splitter. Although the transmitted image is directly measured by a high-resolution (HR) detector

*Address all correspondence to: Peng Jin, E-mail: p.jin@hit.edu.cn; Liang Gao, E-mail: gaol@illinois.edu

0091-3286/2018/$25.00

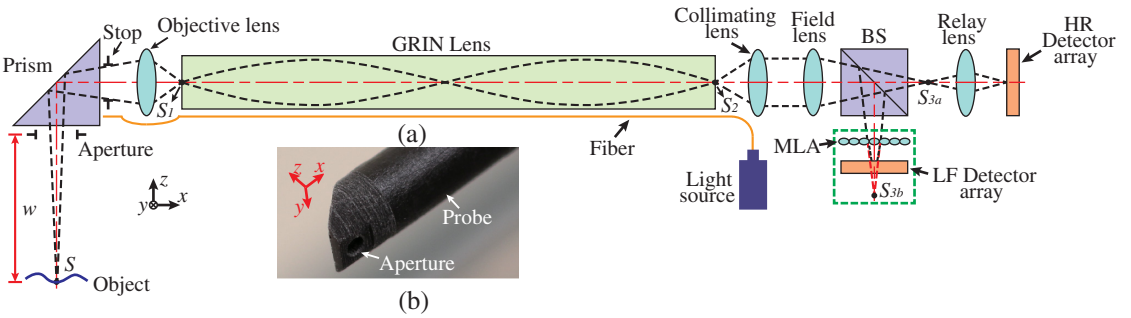


Fig. 1 Schematic of an LFL. (a) Optical setup. The detector arrays in the HR and LF channel are referred to as HR and LF detector array, respectively. (b) Photograph of the distal end of the probe. GRIN, graded-index; BS, beam splitter; MLA, microlens array; HR, high-resolution; LF, light-field.

Table 1 Specifications of two channels.

	Spatial resolution (pixels)	Depth precision	Depth range
HR channel	1920 × 1080	Not applicable	Not applicable
Light-field channel	640 × 360	0.37 mm	62.5 to 67.5 mm

array (Point Gray CR-POE-20S2C-CS), the reflected image is acquired by a custom LF camera which comprises a microlens array (MLA) (Advanced Microoptic Systems GmbH APO-Q-P148-R0.73, $f = 1.6$ mm) and a detector array (Point Gray BFLY-PGE-20E4M-CS). Herein, we adopt a 2.0 LF camera configuration—the distance between the MLA and the detector array is smaller than the focal length of the MLA.[18] We summarize the imaging parameters of two channels above (HR and LF) in Table 1. The geometrical dimensions of our prototype probe are similar to those of commercial laryngoscopes, with a nominal working distance of 65 mm and an outside diameter around 10 mm.

Figure 2 shows the image processing pipeline, which consists of four steps, namely (I) resolution enhancement, (II) disparity estimation, (III) depth reconstruction, and (IV) combination of depth map and HR image. In step I, we first derive the resolution ratio between the HR image and a single elemental image by imaging a calibration object (a chess board). Then we superresolve each elemental image in the LF image using the HR image as the reference by a patch-based image superresolution algorithm.[25] We

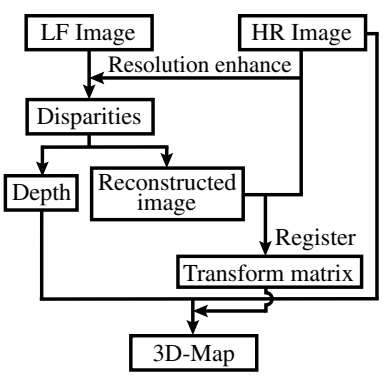


Fig. 2 Flowchart of the image processing pipeline.

downsample the HR image by a factor of the resolution ratio, followed by extracting a series of image patch pairs $\{h_i, l_i\}_{i=1}^{n}$. Here, h_i and l_i denote the image patches extracted from the original and the downsampled HR images, respectively, and i is the index enumerating the image patches. We save these image patches in dictionary D_{ref}. For each patch p_j (5×5 pixels) in an elemental image, we search in D_{ref} and identify nine patches $\{l'_k\}_{k=1}^{9}$, which have the smallest distances in the L_2 norm from p_j. We estimate the HR representation $\hat{h}_j$ of p_j by

$$\hat{h}_j = \frac{\sum_{k=1}^{9} w_k h'_k}{\sum_{k=1}^{9} w_k}, \tag{1}$$

where $w_k = \exp \frac{-\|p_j - l'_k\|^2}{2\sigma^2}$. Here, σ^2 is a hyperparameter, and we determine its value using the Stanford light-field database[26] as a cross-validation dataset.

In step II, we consider the MLA as an array of stereo cameras and derive disparities from the correspondent elemental images pairwise. We illustrate the underlying principle using a simplified one-dimensional example (Fig. 3). The purple dashed lines denote the chief rays associated with microlenses. The extensions of these light rays (red dashed lines) converge to a virtual image point S_{3b}. Figure 3(b) shows the elemental images M_1, M_2, and M_3 formed behind the correspondent microlenses. For each elemental image

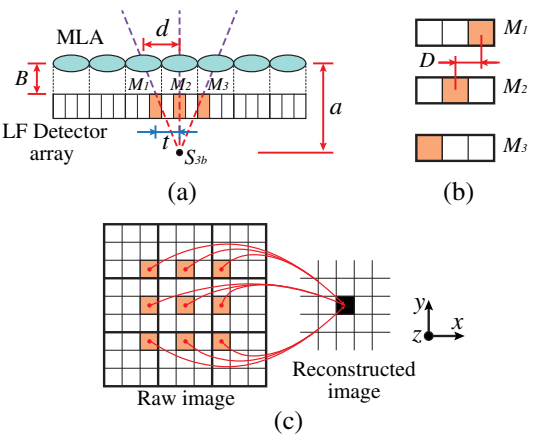


Fig. 3 Light-field reconstruction. (a) Image formation in one dimension. (b) Zoomed-in view of elemental images M_1, M_2, and M_3. The disparity D is calculated as the relative distance between two matched image pixels. (c) Two-dimensional image reconstruction.

pair, we identify the matched features by a searching algorithm based on correlation distance.[27] In brief, we first extract feature sets $\{f_{i,1}\}_{i=1}^{m_1}$ and $\{f_{i,2}\}_{i=1}^{m_2}$ from two elemental images, respectively. For each feature $f_{i,1}$, we search the correspondent neighborhood in $\{f_{i,2}\}_{i=1}^{m_2}$ and identify $f'_{j,2}$, which has the smallest correlation distance to $f_{i,1}$. We term $f'_{j,2}$ as the matched feature of $f_{i,1}$. Next, we calculate the disparity, D, as the relative distance between these two matched image features [Fig. 3(b)].

In step III, we derive depths through disparities. As shown in Fig. 3(a), in the global coordinate, we use t to denote the absolute distance between the two matched pixels in M_1 and M_2. Then the disparity can be calculated by $D = d - t$, where d is the MLA pitch. Using trigonometric relations, we get

$$(a - B)/a = t/d, \tag{2}$$

where B is the distance from the MLA to the LF detector array, and a is the distance from the MLA to the virtual image S_{3b}. Substituting d with $d = D + t$ yields

$$a = B \times d/D. \tag{3}$$

To calculate the object depth, w, we project the virtual image S_{3b} back to the object space. The relation between w and a can be experimentally determined through calibration.[28]

In Fig. 3(c), we further generalize the scheme above to the 2-D case. In each elemental image, the orange pixel collects the light rays converging to the same virtual intermediate image point, S_{3b}. We group these pixels and map their values to a single pixel in the intermediate image. Following this procedure pixelwise yields a reconstructed image.[27]

In light-field imaging, there is a trade-off between the spatial and angular resolution because the total number of 4-D light-field datacube voxels cannot exceed the total number of sensor pixels. To some extent, this trade-off can be mitigated by employing compressed sensing algorithms.[29–31] However, these techniques are computationally extensive, and they highly rely on the ill-posed assumption that the light-field is sparse in a given domain. Also, the requirement of multiple camera exposures[29,30] makes them unsuitable for imaging dynamic scenes. By contrast, in the proposed LFL, we alleviate this problem through fusing the depth map with an HR reference image in step IV. We warp the depth map to the actual size of HR image through a transform matrix derived by registering the reconstructed image from the LF channel and the HR image. Then, we mathematically combine the warped depth map and the HR image to generate an HR 3-D representation of the original scene.

3 Experiments

The depth precision in the LFL is determined by a myriad of factors, namely pitch size of MLA, the distance from the MLA to the sensor, the pixel size the sensor, NA, and the vignetting.[32] To evaluate the depth precision in our prototype, we scanned a point source along both the x- and z-axes. Twelve steps were taken along the x-axis with step size set as 0.8 mm, whereas nine steps were taken along the z-axis with step size set as 0.64 mm. The mean and standard deviation of the measured depths along the x-axis at each z-axis step are shown in Fig. 4. The black dashed line

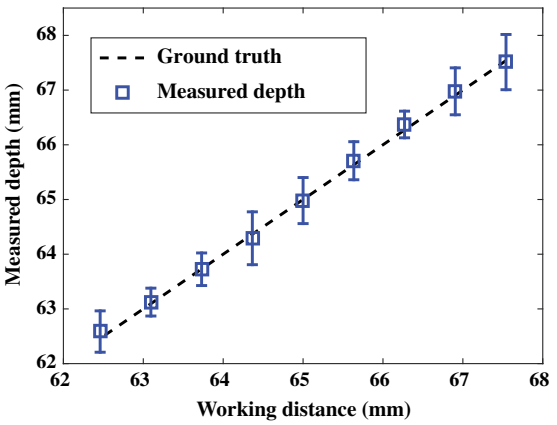
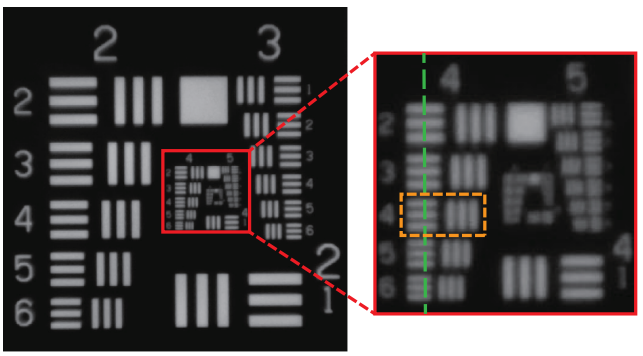


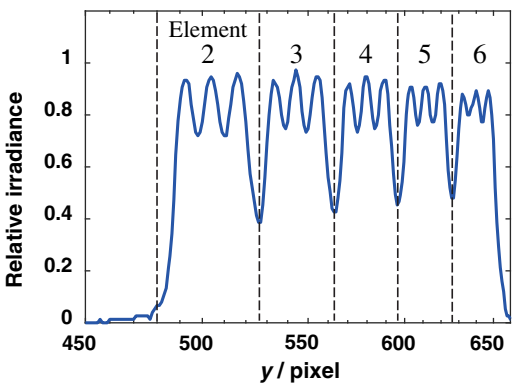
Fig. 4 Quantification of depth precision in LFL.

shows the ground truth. The root mean square error of the average measured depth along the x-axis is 0.07 mm. The depth precision is estimated as the average standard deviation along the z-axis. The result ~0.37 mm, providing an effective depth-to-resolution >10.

To assess the lateral resolution, we imaged a 1951 USAF resolution test target at the nominal working distance (65 mm) of the LFL. Figure 5(a) shows the reconstructed image of the test target with a zoomed-in inset view of bars in groups 4 and 5. In Fig. 5(b), we plot the intensity profile along a green dashed line in Fig. 5(a) (inset). The Lord Rayleigh's criterion states that two overlapping slit



(a)



(b)

Fig. 5 Spatial resolution of the LFL. (a) Reconstructed in-focus image of a 1951 USAF resolution target. (b) Intensity profile along the green dashed line in Fig. 5(a) inset.

images are resolvable when the irradiance of the saddle point between two fringes is lower than $8/\pi^2$ times of the maximum irradiance.[33] Based on this criterion, the bars of group 4 element 4, as shown in the orange dashed rectangle, are the finest resolvable features. Therefore, the lateral resolution of the LFL is 22.6 lp/mm.

Next, we compare this value to the diffraction limit, which is calculated as $R_d = 1/2r = \frac{\mathrm{NA}}{1.22}\lambda$, where r is the radius of Airy disk, λ is the wavelength of the incident light, and NA is the numerical aperture. Given $\lambda = 0.6$ μm and NA $= 0.02$, we have $R_d = 27.3$ lp/mm, which is greater than the experimental resolution of the LFL. We attribute this discrepancy to two reasons. First, we constructed the LFL prototype using only off-the-shelf lenses. The cumulative geometric aberrations blur the image. Second, the employment of a GRIN lens introduces a considerable level of chromatic aberration,[34] which also degrades the imaging performance. Although beyond the scope of this paper, we can potentially overcome these problems by replacing these off-the-shelf lenses and GRIN lens with custom ones.

Finally, to demonstrate the 3-D imaging capability of the LFL, we performed two phantom experiments. First, we used a tilted paper surface with letters as an object. A reference photograph is shown in Fig. 6(a). Captured by a single snapshot, the reconstructed 3-D image is shown in Fig. 6(b). The recovered surface tilt angle matches with the experimental setup. Next, we imaged a vocal fold phantom [Fig. 6(c)], using "vessels" and "vocal fold edges" as features for disparity estimation. Because LFL measures depths only at distinct feature points, we filled the blank areas using interpolation on the assumption that the phantom surface is naturally continuous in slope and curvature. Figure 6(d) shows the reconstructed 3-D image, agreeing well with the ground truth.

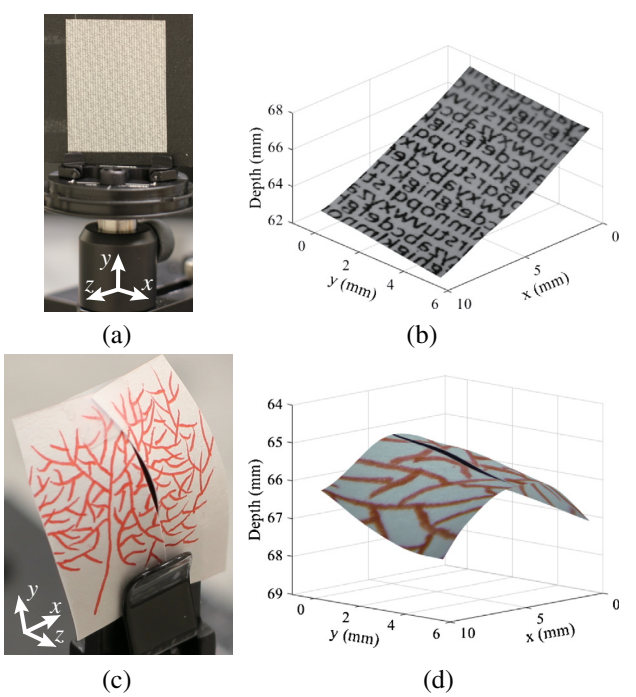


Fig. 6 3-D phantom imaging. (a) Reference photograph of a tilted paper surface with letters, (b) reconstructed 3-D image, (c) reference photograph of a vocal fold phantom, and (d) reconstructed 3-D image.

The overall layout of LFL is simple, requiring rudimentary training to operate. The projected manufacturing cost is comparable with the conventional medical laryngoscopes that are routinely used in the primary care clinics. However, due to the use of only off-the-shelf components, the current probe has an outside diameter ~10 mm and therefore is relatively hard to tolerate. Although not demonstrated, such a drawback could be overcome by replacing the commercial lenses with custom miniaturized optics in the future.

4 Conclusions

In summary, we constructed a 3-D imaging LFL. Rather than measuring only the spatial information, the LFL acquires the spatial and angular information of the incident light rays in parallel. Such a measurement leads to a recovery of a 3-D representation of the original scene with high fidelity. Due to a snapshot acquisition format, the 3-D imaging speed is limited by only the camera's readout speed, which can be potentially up to 1000 volumes/s when coupled to a high-speed camera.[35] In light of its unprecedented 3-D imaging performance, we anticipate that LFL will open an area of investigation in both the clinical diagnostics and fundamental phonation research.

Acknowledgments

This work was supported in part by NSF CAREER grant 1652150 and discretionary funds from UIUC. We thank Kuida Liu for his contribution to the superresolution algorithm and James Hutchinson for the close reading of the letter. We also gratefully acknowledge the financial support from the China Scholarship Council. A patent on this technology is currently pending.

References

1. J. W. Dankbaar and F. A. Pameijer, "Vocal cord paralysis: anatomy, imaging and pathology," *Insights Imaging* **5**(6), 743–751 (2014).
2. F. G. Dikkers and P. G. Nikkels, "Benign lesions of the vocal folds: histopathology and phonotrauma," *Ann. Otol. Rhinol. Larynol.* **104**(1), 698–703 (1995).
3. H. Bakhshaee et al., "Three-dimensional reconstruction of human vocal folds and standard laryngeal cartilages using computed tomography scan data," *J. Voice* **27**(6), 769–777 (2013).
4. T. Sanuki, "Endoscopic mode for three-dimensional CT display of normal and pathologic laryngeal structures," *Otolaryngol. Head Neck (Tokyo)* **69**(3), 211–216 (1997).
5. L. Cveticanin, "Review on mathematical and mechanical models of the vocal cord," *J. Appl. Math.* **2012**(5), 928591 (2012).
6. Q. Xue et al., "Subject-specific computational modeling of human phonation," *J. Acoust. Soc. Am.* **135**(3), 1445–1456 (2014).
7. D. D. Mehta and R. E. Hillman, "Current role of stroboscopy in laryngeal imaging," *Curr. Opin. Otolaryngol. Head Neck Surg.* **20**(6), 429–436 (2012).
8. D. D. Deliyski et al., "Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution," *Folia Phoniatr. Logop.* **60**(1), 33–44 (2007).
9. E. Yumoto et al., "Three-dimensional endoscopic mode for observation of laryngeal structures by helical computed tomography," *Laryngoscope* **107**(11), 1530–1537 (1997).
10. T. Frauenrath et al., "High spatial resolution 3D MRI of the Larynx using a dedicated TX/RX phased array coil at 7.0T," in *Proc. Int. Society for Magnetic Resonance in Medicine*, Stockholm, Sweden, D. K. Sodickson, Ed., p. 894 (2010).
11. T. Chen et al., "A new method of reconstructing the human laryngeal architecture using micro-MRI," *J. Voice* **26**(5), 555–562 (2012).
12. S. Schuberth et al., "High-precision measurement of the vocal fold length and vibratory amplitudes," *Laryngoscope* **112**(6), 1043–1049 (2002).
13. N. A. George et al., "Depth-kymography: high speed calibrated 3D imaging of human vocal fold vibration dynamics," *Phys. Med. Biol.* **53**(10), 2667–2675 (2008).
14. L. Yu et al., "Office-based dynamic imaging of vocal cords in awake patients with swept-source optical coherence tomography," *J. Biomed. Opt.* **14**(6), 064020 (2009).

15. C. A. Coughlan et al., "In vivo cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography," *Sci. Rep.-UK.* **6**, 22792 (2016).
16. J. Geng, "Structured-light 3D surface imaging: a tutorial," *Adv. Opt. Photonics* **3**(2), 128–160 (2011).
17. J. Zhu et al., "Accurate and fast 3D surface measurement with temporal-spatial binary encoding structured illumination," *Opt. Exp.* **24**(25), 28549–28560 (2016).
18. A. Lumsdaine and T. Georgiev, "Full resolution light field rendering," Technical Report, Indiana University and Adobe Systems (2008).
19. E. Y. Lam, "Computational photography with plenoptic camera and light field capture: tutorial," *J. Opt. Soc. Am. A* **32**(11), 2021–2032 (2015).
20. G. Lippmann, "Epreuves reversibles, photographies integrales," *J. Acad. Sci.* **146**(3), 446–451 (1908).
21. N. Bedard et al., "Light field otoscope design for 3D in vivo imaging of the middle ear," *Bio. Opt. Exp.* **8**(1), 260–272 (2017).
22. A. Hassanfiroozi et al., "Hexagonal liquid crystal lens array for 3D endoscopy," *Opt. Exp.* **23**(2), 971–981 (2015).
23. M. Turola and S. Gruppetta, "4D light field ophthalmoscope: a study of plenoptic imaging of the human retina," presented at *Frontiers in Optics, Orlando, United States*, Paper JW3A.36, Optical Society of America (2013).
24. R. Prevedel et al., "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nat. Methods* **11**(7), 727–730 (2014).
25. V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *IEEE Int. Conf. on Computational Photography*, Santa Clara, United States, 2–4 May, pp. 1–10, IEEE (2014).
26. A. Adams et al., "The (New) Stanford light field archive," http://lightfield.stanford.edu/lfs.html (28 November 2017).
27. C. Perwass and L. Wietzke, "Single lens 3D-camera with extended depth-of-field," *Proc. SPIE* **8291**, 829108 (2012).
28. L. Gao, N. Bedard, and I. Tosic, "Disparity-to-depth calibration in light field imaging," presented at *Computational Optical Sensing and Imaging*, Paper CW3D.2, Optical Society of America, Heidelberg, Germany (2016).
29. A. Ashok and M. A. Neifeld, "Compressive light field imaging," *Proc. SPIE* **7690**, 76900Q (2010).
30. S. D. Babacan et al., "Compressive light field sensing," *IEEE Trans. Image Process.* **21**(12), 4746–4757 (2012).
31. K. Marwah et al., "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.* **32**(4), 46 (2013).
32. S. Zhu et al., "On the fundamental comparison between unfocused and focused light field cameras," *Appl. Opt.* **57**(1), A1–A11 (2018).
33. E. Hecht, *Optics*, p. 424, Addison-Wesley, Boston, Massachusetts (1998).
34. K. Siva, R. Krishna, and A. Aharma, "Chromatic aberrations of radial gradient-index lenses. I. Theory," *Appl. Opt.* **35**(7), 1032–1036 (1996).
35. Photron, "Fastcam mini WX (Photron)," https://photron.com/mini-wx/ (28 November 2017).