

# Depth estimation and image recovery using broadband, incoherent illumination with engineered point spread functions [Invited]

Sean Quirin and Rafael Piestun\*

Department of Electrical, Computer and Energy Engineering, University of Colorado, Boulder, Colorado 80309, USA

\*Corresponding author: piestun@colorado.edu

Received 20 August 2012; revised 19 October 2012; accepted 21 October 2012;  
posted 22 October 2012 (Doc. ID 174682); published 7 December 2012

The use of complementary engineered point spread functions is proposed for the joint tasks of depth estimation and image recovery over an extended depth of field. A digital imaging system with a dynamically adjustable pupil is demonstrated experimentally. The implementation of a broadband, passive camera is demonstrated with a fractional ranging error of  $4/10^4$  at a working distance of 1 m. Once the depth and brightness information of a scene are obtained, a synthetic camera is defined and images rendered computationally to emphasize particular features such as image focusing at different depths. © 2012 Optical Society of America

OCIS codes: 070.0070, 110.0110, 150.0150.

## 1. Introduction

The use of images for depth estimation has applications in fields as diverse as biology, entertainment devices, and robotics for industrial monitoring. General requirements include accuracy, precision, resolution, dynamic range, and ultimately the production of images corrected for the traditional information loss associated with defocus. Several approaches to this problem have emerged over time, but unfortunately no single solution fits all the applications. One common classification divides systems into active and passive, according to the use or lack of controlled illumination, respectively. With the advent of multimegapixel sensor arrays there has been a renewed interest in multiaperture systems to extract three-dimensional (3D) information [1–4]. Geometric optical solutions have already been demonstrated to be useful for computer vision applications. These techniques can exploit human vision depth cues such as stereo, which uses the disparity of object position in two images with different known

viewpoints [5], and defocus, which uses the projected size of the optical pupil in the image plane [6] or the spatial frequency content of the image to estimate depth [7,8]. As opposed to 3D imaging and tomography, all these techniques return data in the form of a transverse map of axial positions and object brightness. Light field cameras also operate on a geometric model to reconstruct the light field, i.e., the field of light rays described by the direction of a ray at each position in space. However, one of the disadvantages of light field cameras is the inherent loss of image resolution [9].

In contrast, diffraction-based engineered point spread function (PSF) solutions such as the double-helix PSF (DH-PSF) have recently been demonstrated for precise 3D position information transfer at the microscale and nanoscale [10,11]. Furthermore, early studies have shown that ranging from diffraction effects using coherent monochromatic light can improve precision relative to classical depth from defocus techniques [12]. It is therefore worthwhile to examine whether similar approaches can be applied to improve imaging and ranging of macroscale problems using passive and incoherent broadband light.

In this paper, a system is proposed for imaging a scene with broadband spatially and temporally incoherent light. There is no prior knowledge assumed about the objects except that the emission, reflection, or scattering falls within the spectral bandwidth of the system and that spatial frequency content exists within the system cutoff frequency. This is a significant departure from the application of engineered PSFs to 3D microscopy, where the object is well approximated as a pointlike source [10,13,14].

In this paper we assume the following image model:

$$i(u, v) = \frac{\mathbb{C}}{zd_i} \iiint_{-\infty}^{+\infty} h(u-x, v-y, z) o(x, y, z) dx dy dz + n, \quad (1)$$

where  $i(u, v)$  is the image,  $h(x, y, z)$  is the transverse shift-invariant and axially shift-variant PSF,  $o(x, y, z)$  is the surface brightness,  $d_i$  is the distance from the lens to the image detector,  $\mathbb{C}$  is a constant, and  $n$  is additive random noise. With a pointlike source, the object  $o$  can be approximated as a delta function. Therefore, the image closely resembles the PSF, i.e.,  $i \approx h(z)$  (note that noise must still be accounted for), and the estimation problem is reduced to an extreme case where direct calculation of axial position is feasible with a single image.

The passive ranging problem, i.e., depth estimation without active illumination, becomes ill-posed if this (or similar) prior knowledge about the object is removed, therefore admitting a larger class of objects. Such scenarios are typical, for instance, in the imaging of natural scenes or fluorescent biological samples. A solution is therefore necessary in order to extend the existing work of diffraction-based engineered PSFs to the application of passive ranging problems with arbitrary objects and broadband incoherent illumination. Additionally, this solution must be evaluated to determine whether engineered PSFs provide improved axial localization performance for macroscale applications, and if so, how the

localization precision depends on the specific object under investigation.

For the purpose of this study, the active use of illumination control (for example, time-of-flight and structured illumination imaging [15,16]) is excluded from consideration in favor of passive solutions, which will only require modification of the optical design. Passive systems are relevant for solutions where data must be collected without leaving any trace of the measurement in the form of emitted radiation—of paramount importance for surveillance applications. Passive solutions also provide an advantage in terms of power demands, since adequate lighting of a scene can be provided from ambient illumination instead of being actively controlled.

In this paper, a solution for creating depth and image brightness maps is proposed using a pair of complementary engineered PSFs. Background information is provided to the reader in Section 2. In Section 3, the diffraction-based engineered PSFs are introduced, including a pair of complementary engineered PSFs. The proposed system is then analyzed in terms of theoretical performance limitations in Section 4. This optical solution is experimentally demonstrated in Section 5. An example application of the acquired 3D information is proposed in Section 6, where digital refocusing is demonstrated on a data set collected from the joint-engineered PSF optical system. The results are discussed in Section 7.

## 2. Background

For passive ranging systems, a sense of depth can be provided from the defocus modulation of the PSF [17–19]. This axial dependence is demonstrated in Fig. 1A, where the PSF of a typical camera is shown as the image of a point is defocused. The result of defocus is an enlarged, rotationally symmetric, transverse pattern. From examination of Eq. (1), it is clear that this is consistent with the observation that out-of-focus objects become blurry and carry limited high-spatial-frequency information. Solutions exploiting this axial variation are attractive for

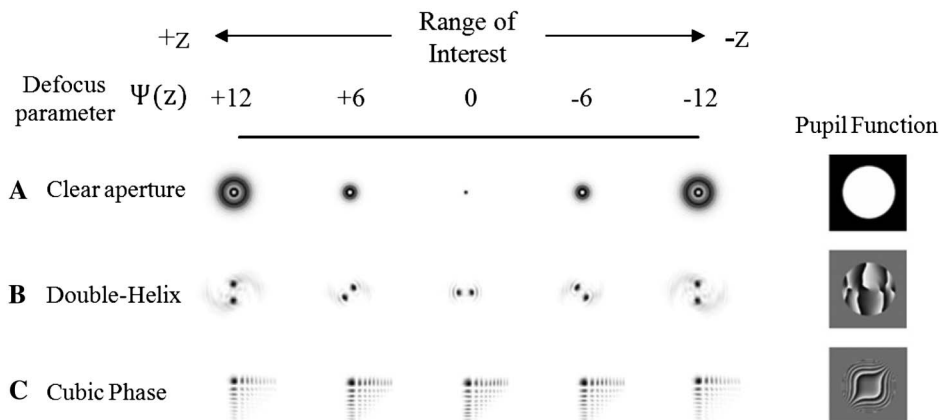


Fig. 1. Axial dependence of three PSFs. A, The typical camera PSF using a clear, circular aperture; B, the DH-PSF; and C, the cubic-phase PSF (CP-PSF) are compared through an equivalent range of defocus. The standard camera (clear aperture) PSF is symmetric about focus, the DH-PSF has a unique rotation angle at each position, and the CP-PSF has an axially invariant pattern.

ranging applications because they can provide range over a wide field of view. This means that the data input is parallelized along the transverse dimension of the image in contrast to a point-by-point scanning-based technique. The axial information of the object is then contained in a depth-dependent pattern encoded into the image. Note that in some cases this pattern can be ambiguous, with more than one axial position producing the same image. Such is the case for the standard optical pupil found in most cameras and demonstrated in Fig. 1A. This is one of the reasons depth-from-defocus systems require the use of more than one image via either the active adjustment of focus across an image series [18] or the use of multiple fixed focal planes [20]. Either case will provide unique depth discrimination as given by the rate of change of the radius of blur [17,19,20].

Coded pupil functions can be utilized to engineer the PSF for modifying depth sensitivity/information [12,21–24]. With suitable coding, it has been demonstrated that these methods can operate under passive and polychromatic illumination [22,25,26], thereby making them useful for a variety of imaging applications. Of these, it is expected that those based on diffraction will yield more precise axial location solutions as compared to geometrical optics approaches.

In particular, the DH-PSF has been demonstrated to be a superior solution for 3D localization tasks of point sources [27]. A generic DH-PSF can be found from the phase-only superposition of Laguerre–Gauss modes [12,28]. An implementation with improved optical efficiency can be generated from numerical optimization of the pupil-plane phase modulation to create the phase mask shown in Fig. 1B [26]. This phase mask operates on a limited depth range (see Section 3 for further details) by localizing more energy into each of the two DH-PSF peaks instead of the side lobes or background.

The DH-PSF advantage relative to a clear aperture is in its ability to range with enhanced precision when given *a priori* knowledge of the object [12,27]. Rather than encoding with blur, the DH-PSF encodes the axial position of the object with the rotational orientation from two copies of the object in the image. For point sources, these images will trace a double-helix pattern as the object is moved through focus, as shown in Fig. 1B. Prior work has studied the spectral response of the DH-PSF and demonstrated that only a slight variation of rotation rate exists ( $\pm 5^\circ$ ) for phase masks with a shift away from the design wavelength of up to 50 nm [26].

By way of contrast, the utility of the engineered PSF approach is highlighted by a PSF with limited axial dependence: the cubic phase mask [24] is one such example that generates a PSF that maintains its transverse profile with depth—hereafter referred to as the CP-PSF (Fig. 1C). Taken together, these two engineered PSF designs exemplify the operational gamut of the axial dependence solution space.

In this work, we investigate the functionality of a DH-PSF for passively ranging macroscopic objects with spatial frequency content within the spectral passband of the system while using broadband illumination. The main differences from prior work in depth estimation using diffraction [12] are (a) the use of two high-efficiency engineered phase masks rather than a low-efficiency hologram and a reduced aperture, (b) the use of broadband incoherent light rather than quasi-monochromatic illumination, and (c) the reconstruction of depth maps for relatively complicated scenes with multiple depths.

### 3. Complementary Selection of Engineered Point Spread Function Solutions

As discussed above, to satisfy the most general imaging applications, a system should operate with few assumptions about the types of objects present in the scene. Because the PSF applied to encode ranging information is necessarily axially shift variant, the problem of estimating position from a single image becomes underdetermined without additional information about the object. This can be seen from Eq. (1), where the measurement  $i$  contains no unique information for disambiguation of either the PSF,  $h$ , or the object,  $o$ . To address this problem, a second optical channel is proposed to provide unique object information that can be used to recover the PSF.

As the DH-PSF channel is well suited for ranging, it may be considered a dedicated channel for completing depth estimation tasks. In a similar way, the second optical channel could be chosen to be dedicated to the estimation and restoration of object brightness. The solution must provide axially uniform imaging performance/resolution that is complementary to the axial ranging channel (DH-PSF). Because the DH-PSF has an operational range well outside the traditional depth of field (DOF) of a clear circular aperture with equivalent  $f/\#$  (about a 12 $\times$  increase as defined by Hopkins' allowable defocus criteria in [24]), an extended depth of field (EDOF) channel is desirable to complement the DH-PSF channel. Prior work has utilized a standard optical aperture with higher  $f/\#$  to accommodate the depth-of-field requirement of an invariant PSF [12]. The increased  $f/\#$  reduces the aperture size and consequently decreases the optical efficiency. To avoid this loss of optical efficiency, an alternative is investigated using the cubic-phase (CP) pupil function [24], defined as

$$p_{\text{CP}}(m, n) = e^{i\alpha(m^3+n^3)}, \quad (2)$$

where  $m, n$  are the transverse coordinates of the pupil plane,  $\alpha$  is the amplitude coefficient of the CP variation of the pupil, and the axial dependence of the PSF is found from

$$h_{\text{CP}}(x, y; z) = \left| \mathcal{F} \left\{ p_{\text{CP}}(m, n) e^{i\frac{\alpha}{2}(m^2+n^2)\left(\frac{1}{f} - \frac{1}{z} - \frac{1}{d_i}\right)} \right\} \right|^2, \quad (3)$$

with  $\mathcal{F}$  denoting a Fourier transform. The axial invariance of  $h_{CP}$  is shown in Fig. 1C. Note that the quadratic phase function in Eq. (3) is the optical aberration associated with defocus. Defining this explicitly as the defocus parameter,

$$\Psi = \frac{\pi}{\lambda} (m^2 + n^2) \left( \frac{1}{f} - \frac{1}{z} - \frac{1}{d_i} \right), \quad (4)$$

the operational range of the DH-PSF mask can be found for a given optical system by noting that the 180° rotation coincides with  $|\Psi| \leq 12$ . Modulation of the coefficient  $\alpha$  is desired, then, to find an optimum match to the DOF of the DH-PSF channel [29], but as a minimum the original derivation had suggested that  $\alpha \gg 20$  (in normalized coordinates) in order to have a CP aberration that is significant relative to the defocus aberration, hence providing defocus invariance [24]. The use of a CP pupil function allows the  $f/\#$  of the two optical channels to be matched, thereby optimizing optical efficiency. Note that as  $\alpha$  is increased, the attenuation of spatial frequencies is correspondingly increased, resulting in a lower spectral SNR. Therefore, the correct value of  $\alpha$  should be carefully chosen [29].

From the CP-PSF channel, a restoration of the spatial frequency content of the object can be provided via Wiener deconvolution for image recovery. The Wiener deconvolution filter is a linear, least-squares solution to estimate the object from the image  $i_{CP}$  and the known axially invariant PSF  $h_{CP}$ ,

$$\hat{o}(x, y) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{i_{CP}\}}{\mathcal{F}\{h_{CP}\}} \left[ \frac{|\mathcal{F}\{h_{CP}\}|^2}{|\mathcal{F}\{h_{CP}\}|^2 + \text{SNR}(f)^{-1}} \right] \right\}, \quad (5)$$

where  $\mathcal{F}^{-1}$  denotes an inverse Fourier transform and the caret accent indicates that the object variable is an estimate. Note that when no noise is present [i.e.,  $\text{SNR}(f) \rightarrow \infty$ ], this reduces to solving for the object using the convolution theorem—which is a valid assumption for linear, transverse shift-invariant optical systems.

Using the complementary, double-helix-engineered PSF channel, the axially dependent DH-PSF is found from

$$\hat{h}_{DH}(x, y; z) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{i_{CP}\}^* \mathcal{F}\{i_{DH}\}}{|\mathcal{F}\{i_{CP}\}|^2 + \text{SNR}(f)^{-1}} \mathcal{F}\{h_{CP}\} \right\}, \quad (6)$$

where the relationship to the convolution theorem can again be seen for the DH-PSF channel when no noise is present and the object estimate from Eq. (5) is inserted. The estimate of  $\hat{h}_{DH}$  is then used to calculate a  $z$  position through calibration of the relationship between the recovered, relative rotation angle of each lobe from the DH-PSF and the axial position, as demonstrated in Fig. 2. This relationship can be characterized experimentally for each system in order to account for the presence of unknown aberrations. Note that inclusion of the spectral SNR [30] in Eqs. (5) and (6) is known as regularization, used to accommodate for the presence of noise. The result is that spatial frequencies with little noise will be restored with higher emphasis than spatial frequencies with higher noise. In practice, this spectral SNR—defined as the ratio of the spatial frequency power spectrum of the image relative to noise—is assumed to be white (i.e., constant) and found empirically. To estimate the  $z$  position of the object, the rotation angle, as measured from the centroid of each DH-PSF lobe, is found from  $\hat{h}_{DH}$  and related to  $z$  by a calibration curve. An axial position estimation technique based on the generic geometrical features of the recovered  $\hat{h}_{DH}$  is chosen in lieu of more elaborate estimation methods such as the maximum likelihood estimation, because this centroid/rotation angle algorithm (shown in Fig. 2) is expected to be more robust to the inherent noise from the deconvolution results. Figure 3 summarizes in a block diagram the different components of the system and how they relate the object being imaged to the data outputs.

#### 4. Theoretical Performance Limitations

To justify the use of the DH-PSF for this application, the Cramer–Rao lower bound (CRLB) is analyzed for

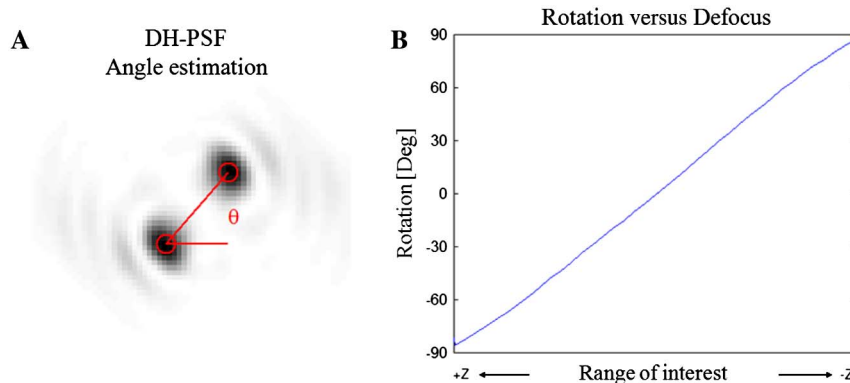


Fig. 2. (Color online) A, Rotation angle is associated with the estimated  $\hat{h}_{DH}$  by calculating the angle subtended by the centroids of each lobe and a frame-of-reference on the detector—here the horizontal axis. B, The DH-PSF rotation angle varies accordingly as a function of axial position and can be found experimentally to account for the presence of possible aberrations.



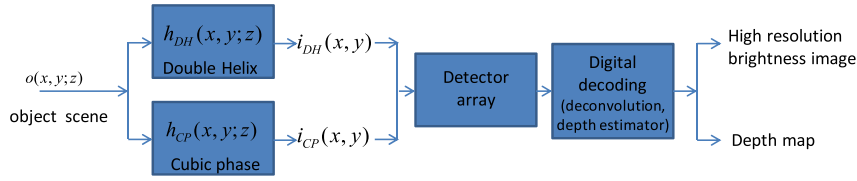


Fig. 3. (Color online) Block diagram of the dual-channel complementary PSF engineering digital optical system.

quantifying the theoretical limit of precision in axial ranging and to aid in the comparison with a classical lens [31]. In contrast to prior analysis [11,12,27], the CRLB will also be used here to develop intuition on how the precision depends on object type. Using the imaging model described in Eq. (1), the CRLB is calculated using two input objects: an ideal point source as well as a sample image taken of a die-cast model car (right-hand portion of the experimental image shown later in Fig. 7A). The point source represents a limiting performance for object types, as this is in principle an idealized object with infinite spatial frequency content.

The CRLB is calculated using

$$\sigma_{\text{CRLB}}^{-2}(z) = \sum_{m=1}^M \sum_{n=1}^N \frac{1}{\sigma_N} \left( \frac{\partial}{\partial z} i'[m, n] \right)^2, \quad (7)$$

where  $i'[m, n] = \int_{\lambda_1}^{\lambda_2} \eta(\lambda) i[m, n; z] d\lambda$  is the broadband, discretely sampled image,  $\sigma_N$  is the standard deviation of per-pixel Gaussian noise; and  $\sigma_{\text{CRLB}}$  is the lower limit to measurement precision (standard deviation) associated with axial estimation [32]. Values for the broadband spectrum  $\eta(\lambda)$  are selected to match with an experimentally characterized filter (shown in Fig. 5).

The axial CRLB performance of a single DH-PSF image is compared with that of a single, clear circular aperture image, which is representative of a typical camera, in Fig. 4. The CRLB is seen to depend on the

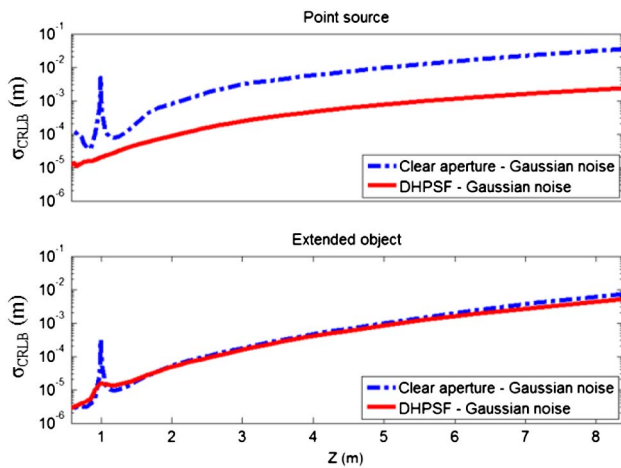


Fig. 4. (Color online) CRLB analysis of the axial estimation precision for a point source (top panel) and a die-cast car (bottom panel). The performance of the DH-PSF and standard camera is reported as the object is translated through focus at  $z = 1$  m. The relative performance gains from using DH-PSF optics are seen to depend on the spatial frequency content of the object.

axial position of each object. In the CRLB calculation, a Gaussian-dominated noise process with an SNR of 16 is assumed, where SNR is defined as the ratio of maximum detector pixel counts to the standard deviation from noise.

These results confirm that the depth estimation precision depends on the object under investigation. This dependence is understood by the fundamental observation that with engineered PSF methods only the spatial frequency content present in an image can be coded. For the point source, the full advantage of the engineered PSF design can thus be exploited. However, for a given object, the spatial frequency content can be limited, and therefore the relative advantage associated with the engineered PSF is decreased. Despite this, the performance of the DH-PSF channel is seen to be more uniform over the region of interest and can generally improve upon the performance of the standard camera. Furthermore, the CRLB analysis provides insight into the situations for which the DH-PSF excels, namely in the ranging of small objects and objects with fine texture or high level of detail.

## 5. Experimental Depth Estimation and Image Recovery

An experimental implementation of an engineered PSF camera capable of the joint depth estimation and image recovery tasks must be able to accommodate the two complementary optical channel designs introduced earlier. This requirement can be satisfied with the use of a serial engineered PSF imaging system, shown in Fig. 5. Here a phase-only spatial light modulator (SLM) is proposed to cycle through the appropriate pupil functions necessary for the encoding of the engineered PSFs.

This optical design incorporates an imaging lens paired with a chromatic filter ( $1/e^2 = 80$  nm) and linear polarizer, which is then followed with a relay imaging system. The SLM is placed in the Fourier plane to allow for the manipulation of multiple engineered PSF channels as a dynamic, serialized collection of task specific images. The reflective phase-only SLM (Boulder Nonlinear Systems XY-P512) has 8 bit dynamic range to control the phase modulation within  $0-2\pi$  on a  $512 \times 512$  pixel array. This SLM necessitates the use of a polarization analyzer to block the light with polarization orthogonal to the active modulation axis of the SLM. Because the SLM is chromatic, the color filter is chosen to coincide within the design range of a calibrated SLM. This filter also takes into account the chromaticity of the DH-PSF mask design that operates over a bandwidth in

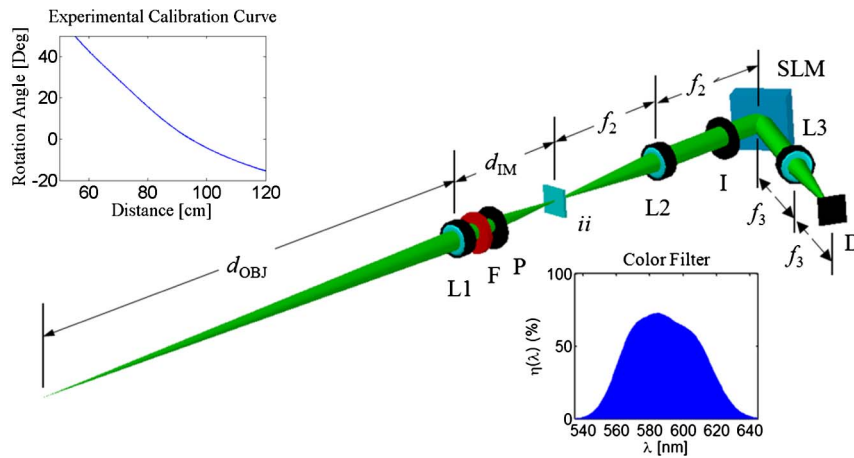


Fig. 5. (Color online) Experimental configuration of serialized engineered point spread function optical imager. The imaging lens (L1) is paired with a color filter (F) and polarization analyzer (P). The intermediate image (ii) is reimaged through a Fourier transform lens (L2) such that the phase of the spatial frequency components can be modulated with the SLM. An iris (I) is placed before the SLM to confine the incident light to the modulated region of the liquid-crystal device. The final Fourier transform lens (L3) is used to complete the optical encoding and form a signal on the detector (D) that is related to the object and encoded with the desired engineered PSF.

excess of 100 nm [26]. The angle of incidence to the reflective SLM is  $\theta = 18^\circ$ . However, the SLM refractive index control is optimized for normal incidence, and to compensate for the effect of incidence angle the DH-PSF mask was linearly scaled by a factor of 0.9 to minimize the DC spot observed in the image (Fig. 6). The final image is formed a distance  $f_3$  from the last lens. A monochrome detector (Point Grey GRAS-20S4M) with 8 bit dynamic range per pixel and a  $4.4 \times 4.4 \mu\text{m}$  pitch pixel is used for image capture of  $480 \times 640$  pixel images. The lenses have focal lengths  $f_1 = 60$  mm,  $f_2 = 80$  mm, and  $f_3 = 100$  mm. The focal lengths  $f_2 \neq f_3$  were chosen to minimize the incidence angle on the SLM due to the lens mounting hardware. The aperture size of the phase mask was limited to  $\phi = 4.19$  mm by an iris, and the object space focal plane was located at  $d_{\text{OBJ}} = 1$  m. A calibration curve for relating the rotation angle of the DH-PSF to the axial position of an object is created by using a pinhole that is translated axially. Using this technique, the calibration curve through the axial region of interest is presented in Fig. 5. The calibration curve was limited to distances of less than 120 cm due to both laboratory space constraints, which prevented moving a source further away, and an inherent loss in signal, as the source was moved further from the collection optics. To complement the DH-PSF channel, a CP mask was implemented for EDOF imaging with a cubic coefficient

of  $\alpha = 166$  in normalized coordinates [24]. In-focus experimental PSF images from both optical channels are shown in Fig. 6.

To properly operate, some assumptions are placed upon the optical system and the scene under investigation. These include that (a) a signal from the object exists within the temporal bandwidth of the system, (b) the signal is from a surface rather than from a volume, (c) the source is within the operational DOF of the DH-PSF, (d) the objects have sufficient detail to allow for encoding with engineered PSFs (as discussed earlier in Section 4), (e) the lenses used are sufficiently aplanatic across the field of view (although this can be remedied with thorough calibration techniques), and (f) the transverse translation of the CP-PSF does not shift the image of the object more than a few pixels. With these in mind, a demonstration scene is created consisting of a pair of die-cast cars located at two different axial distances and illuminated by a halogen bulb. A representative scene is shown in Fig. 7A.

For depth estimation tasks the axial position estimates are required as a function of transverse position. Therefore, to localize individual regions, a signal processing window is applied to the image to provide  $M \times M$  pixel subimages and uses these to estimate the axial position within the scene by the use of Eq. (6). In practice a Kaiser window, defined as  $w(x) = I_0(\beta \sqrt{1 - ((2n/M) - 1)^2}) / I_0(\beta)$ , where  $I_0$  is the zeroth-order modified Bessel function of the first kind,  $M$  is the size of the window (here,  $M = 128$ ),  $0 \leq n \leq M$ , and a damping coefficient  $\beta = 5$ , is applied to each subimage to smoothly attenuate the contribution of neighboring regions. The  $M \times M$  pixel Kaiser window is taken at every 8 pixel interval in the image, deconvolved with the CP image, and a rotation angle found from the DH-PSF to transform into an axial position and create a dense array of estimates. The resulting passive-ranging depth map therefore subsamples the



Fig. 6. A, Experimental DH-PSF image; B, experimental CP-PSF. Both are shown as measured after L3 in the engineered PSF imaging system.

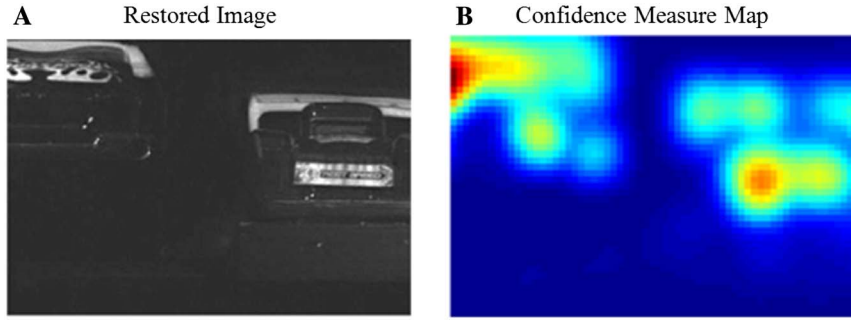


Fig. 7. (Color online) A, Restored image contains spatial frequency information that is used to provide local confidence measures with regard to the axial estimation and B, the confidence measures for the restored CP image.

original wide-field image by a factor of  $8\times$ . A reduced resolution of the depth map relative to the original images is common to depth from defocus methods [6]. However, by proper selection of the sampling, it is possible to target a required depth map resolution.

As discussed earlier, engineered PSF methods rely on the spatial frequency content of the scene; therefore, a confidence measure for each axial estimate can be made based on the Fourier content within each transverse window. The availability of a restored image from the CP channel allows for an accurate confidence measure regardless of the object position. The chosen confidence measure provides a normalized indication of the high-spatial-frequency content in the image as

$$CM = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\delta(x, y)|^2 dx dy}{\int_0^{\infty} \int_0^{\infty} |\delta(x, y)|^2 dx dy}. \quad (8)$$

In practice, the selection of  $\epsilon_x$  and  $\epsilon_y$  are found through an empirical search for the best results. The result for a representative experimental image is shown in Fig. 7B. This indicates the detailed portions from either car provide axial estimates with the highest confidence.

Histograms of the estimated axial positions in Fig. 8 suggest that the front and back details for each car can be resolved. The mean distance to the front of the cars was  $84.15 \pm 0.52$  cm and  $98.16 \pm 0.26$  cm for the left and right cars of Fig. 8A, respectively. The depth estimation measurement used 169 and 112 independent estimates for the left car and right car, respectively, resulting in an error of the mean estimate of 0.04 cm and 0.03 cm. The fractional ranging error is therefore  $4/10^4$ , defined as the average of the ratios of the error of the mean to the mean distance taken for each object. An image segmentation algorithm is applied to the restored image in Fig. 9A

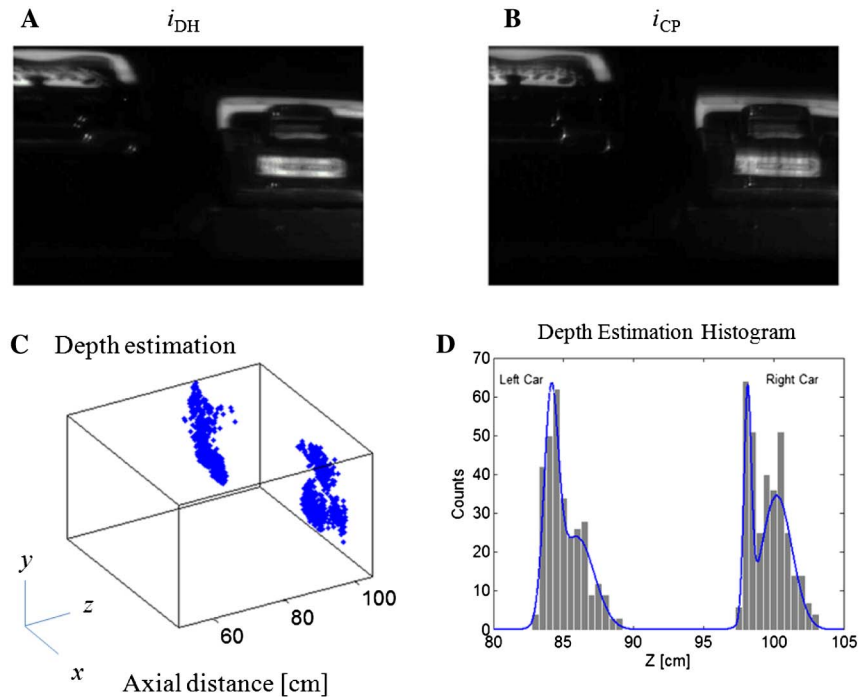


Fig. 8. (Color online) Experimental images from both the DH-PSF channel (A) and the CP-PSF channel (B) are used to calculate depth estimation results (C). Histograms for the dense collection of axial position estimates for each of the two cars are shown in (D). Two broad distributions, indicative of the locations of the two cars, are readily apparent (Media 1).

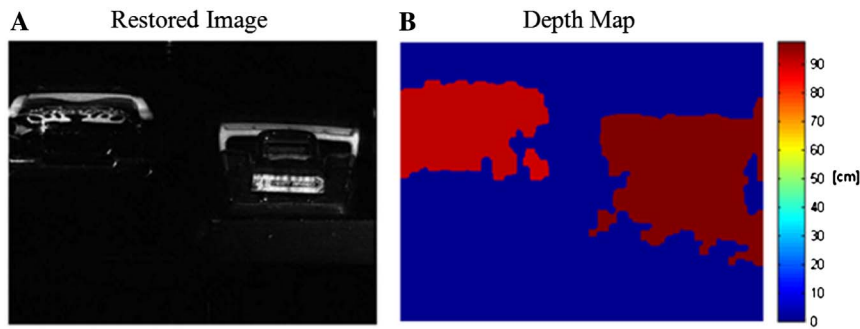


Fig. 9. (Color online) Restored object estimate of the scene from the cubic phase channel (A) is used for image segmentation. After segmenting the objects within the scene, an average axial distance taken from the depth estimation channel is assigned to each car (B).



Fig. 10. Digital refocusing of the scene as a postprocessing step. The proposed system returns a grid of axial estimates associated with the diffraction-limited information content (A) to define the surface and brightness of the scene. Images can then be generated synthetically, emphasizing focus to convey visual information. (B) and (C) alternatively focus on either object in the scene as a demonstration.

to identify the individual cars and associate a mean position as shown in Fig. 9B.

Deviation from CRLB performance in this experiment is expected due to three factors:

- The car has a true 3D profile and not a planar one as assumed in the CRLB calculation.
- The angle estimator is a trade-off between calculation speed and estimation efficiency—in this experiment, calculation speed was emphasized to achieve faster throughput. Work on more efficient DH-PSF estimators is also an active area of research [33].
- A depth estimate is given for  $8 \times 8$  neighborhoods of pixels rather than an estimate grid as dense as the discretely sampled image. This is partially due to the practical difficulty of addressing the influence of the axial-position-dependent transverse bias introduced in the CP reference image [34]. A second factor is related to the computation time required for estimating such dense arrays.

## 6. Digital Refocusing of Experimental Data

Given the estimated  $z$  positions and a digital representation of the restored object,  $\hat{o}$ , the 3D profile of the scene can now be utilized for synthetically generating images of functional interest using Eq. (1). Focus is used as one example, because it can be used to emphasize objects with visual cues. The use of this visual aid has partially motivated renewed research in the light field or plenoptic cameras [3]. One drawback of the plenoptic camera design is a sacrifice of

transverse resolution necessary to collect the information for refocusing. For example, the transverse resolution of the plenoptic image is determined by the number of microlenses in the image plane instead of the number of pixels on the detector. Additionally, in order to collect the necessary light field information for refocusing, each microlens must be subsampled by a number of pixels ( $\sim 14$  pixels per microlens is used in [3]), thereby inherently limiting transverse resolution. As an alternative, this system can render refocused images by defining a synthetic camera with alternative DOFs or PSFs. In addition, the object space focal plane can be alternatively shifted to either car to generate images that emphasize 3D features of the scene, as demonstrated in Fig. 10. The system can therefore realistically emphasize details of the scene while requiring minimal user input (only specifying the focal plane is necessary).

## 7. Conclusions

In summary, a serialized engineered PSF camera was proposed to recover the depth information and object brightness that describe the scene under investigation. The system performance limits were analyzed theoretically and compared with a classical method, the standard, clear, circular aperture used in depth from defocus. This development was used to illustrate how estimation precision can depend on the object's axial location within the scene. The experimental implementation was shown to have a fractional ranging error of  $4/10^4$ . The recovered scene function was then used to synthesize



functional images that convey particular information to the user like digital refocusing.

The main contribution of this work is in the use of diverse and complementary PSFs engineered to extract information from the scene that is normally lost with classical imaging systems. Indeed, the DH-PSF provides high depth discrimination over an extended depth range, while the CP provides in-focus information for a similar DOF. The two capabilities are thus complementary, and by joint digital postprocessing the scene brightness and depth map are retrieved. The use of diffraction-based depth retrieval is critical for providing enhanced depth discrimination in the focal region relative to competing methods. The basic property of interest in the DH-PSF is its rapid change through defocus that eases the estimation of depth. The use of efficient phase masks provides high light throughput, which is critical for low-light passive illumination scenarios.

As opposed to multiaperture systems such as light field cameras, the system proposed here provides brightness maps with full resolution as expected from the  $f/\#$  of the system. Also, in contrast to stereo imaging, the dual-channel PSF engineering solution does not suffer from the problems of correspondence and occlusion. The system is thus attractive for applications where a scene presents objects with high spatial frequencies such as that provided by an inherent texture or detailed structure. The time sequential imaging performed with complementary PSF can also be implemented in parallel with a dual aperture system or via a beam splitter device with either one or two cameras.

The current goal of incorporating engineered PSFs into arbitrary object and macroscale imaging problems should not prevent future solutions that also include an active illumination component designed to exploit the prior passive system designs. In fact, this is one potential avenue for future growth in the engineered PSF domain. In either case, the passive system design must first be understood and demonstrated, two goals accomplished here.

The authors acknowledge support from a Honda Initiation Grant, the National Science Foundation, and the University of Colorado Technology Transfer Office.

## References

1. G. Lippman, "Épreuves réversibles donnant la sensation du relief," *J. Phys. Théor. Appl.* **7**, 821–825 (1908).
2. E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Machine Intell.* **14**, 99–106 (1992).
3. R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford University Computer Science Tech Report CSTR 2005-02 (Stanford University, 2005).
4. H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," *Opt. Lett.* **26**, 157–159 (2001).
5. Y. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: how different really are they?" *Int. J. Comput. Vis.* **39**, 141–162 (2000).
6. S. Chaudhuri and A. N. Rajagopalan, *Depth from Defocus: A Real Aperture Imaging Approach* (Springer-Verlag, 1999).
7. Y. Xiong and S. A. Shafer, "Depth from focusing and defocusing," Technical report CMU-RI-TR-93-07 (Robotics Institute, Carnegie Mellon University, 1993).
8. T. Darrell and K. Worn, "Pyramid based depth from focus," in *Proceedings of Computer Vision and Pattern Recognition* (IEEE, 1988), pp. 504–509.
9. T. Georgev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala, "Spatio-angular resolution tradeoff in integral photography," in *Proceedings of the 17th Eurographics Conference on Rendering Techniques* (Eurographics Association, 2006), pp. 263–272.
10. S. R. P. Pavani, J. G. DeLuca, and R. Piestun, "Polarization sensitive, three-dimensional, single-molecule imaging of cells with a double-helix system," *Opt. Express* **17**, 19644–19655 (2009).
11. S. R. P. Pavani, A. Greengard, and R. Piestun, "Three-dimensional localization with nanometer accuracy using a detector-limited double-helix point spread function system," *Appl. Phys. Lett.* **95**, 021103 (2009).
12. A. Greengard, Y. Y. Schechner, and R. Piestun, "Depth from diffracted rotation," *Opt. Lett.* **31**, 181–183 (2006).
13. S. R. P. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. E. Moerner, "Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function," *Proc. Natl. Acad. Sci. USA* **106**, 2995–2999 (2009).
14. B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," *Science* **319**, 810–813 (2008).
15. A. Medina, F. Gaya, and F. del Pozo, "Compact laser radar and three-dimensional camera," *J. Opt. Soc. Am. A* **23**, 800–805 (2006).
16. M. A. A. Neil, R. Juskaitis, and T. Wilson, "Method of obtaining optical sectioning by using structured light in a conventional microscope," *Opt. Lett.* **22**, 1905–1907 (1997).
17. A. P. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal.* **PAMI-9**, 523–531 (1987).
18. S. K. Nayar and Y. Nakagawa, "Shape from Focus," *IEEE Trans. Pattern Anal.* **16**, 824–831 (1994).
19. P. Favaro and S. Soatto, *3-D Shape Estimation and Image Restoration—Exploiting Defocus and Motion Blur* (Springer-Verlag, 2007).
20. S. Ram, J. Chao, P. Prabhat, R. J. Ober, and E. S. Ward, "A novel approach to determining the three-dimensional location of microscopic objects with applications to 3D particle tracking," *Proc. SPIE* **6443**, 64430D (2007).
21. G. E. Johnson, E. R. Dowski, and W. T. Cathey, "Passive ranging through wave-front coding: information and application," *Appl. Opt.* **39**, 1700–1710 (2000).
22. A. Levin, R. Fergus, F. Durand, and B. Freeman, "Image and depth from a conventional camera with a coded aperture," *SIGGRAPH 2007* (ACM, 2007), article 70.
23. C. Zhou, S. Lin, and S. Nayar, "Coded aperture pairs for depth from defocus," *IEEE International Conference on Computer Vision* (IEEE, 2009), pp. 325–332.
24. E. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Appl. Opt.* **34**, 1859–1866 (1995).
25. J. van der Gracht, E. R. Dowski, M. G. Taylor, and D. M. Deaver, "Broadband behavior of an optical-digital focus-invariant system," *Opt. Lett.* **21**, 919–921 (1996).
26. S. R. P. Pavani and R. Piestun, "High-efficiency rotating point spread functions," *Opt. Express* **16**, 3484–3489 (2008).
27. G. Grover, S. R. P. Pavani, and R. Piestun, "Performance limits on three-dimensional particle localization in photon-limited microscopy," *Opt. Lett.* **35**, 3306–3308 (2010).
28. R. Piestun, Y. Y. Schechner, and J. Shamir, "Propagation-invariant wave fields with finite energy," *J. Opt. Soc. Am. A* **17**, 294–303 (2000).
29. S. Bagheri, P. E. X. Silveira, R. Narayanswamy, and D. Pucci de Farias, "Analytical optical solution of the extension of the depth of field using cubic-phase wavefront coding. Part II. Design and optimization of the cubic phase," *J. Opt. Soc. Am. A* **25**, 1064–1074 (2008).

30. S. Bagheri, P. E. X. Silveira, and G. Barbastathis, "Signal-to-noise-ratio limit to the depth-of-field extension for imaging systems with an arbitrary pupil function," *J. Opt. Soc. Am. A* **26**, 895–908 (2009).
31. S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory* (Prentice Hall, 1993).
32. H. Barrett, C. Dainty, and D. Lara, "Maximum-likelihood methods in wavefront sensing: stochastic models and likelihood functions," *J. Opt. Soc. Am. A* **24**, 391–414 (2007).
33. S. Quirin, S. R. P. Pavani, and R. Piestun, "Optimal 3D single-molecule localization for super-resolution microscopy with engineered point spread functions," *Proc. Natl. Acad. Sci. USA* **109**, 675–679 (2011).
34. W. T. Cathey and E. R. Dowski, "New paradigm for imaging systems," *Appl. Opt.* **41**, 6080–6092 (2002).