

Received July 16, 2021, accepted August 13, 2021, date of publication September 3, 2021, date of current version September 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3109865

Accuracy in Depth Recovery and 3D Image Synthesis From Single Image Using Multi-Color Filter Aperture and Shallow Depth of Field

RASHMI R. DESHPANDE¹, MAHABALESWARA RAM BHATT^{2,3},
AND C. H. RENU MADHAVI⁴, (Senior Member, IEEE)

¹Research Center of Electronics and Instrumentation Engineering, R. V. College of Engineering (affiliated to Visvesvaraya Technological University-Belagavi), Bengaluru 560059, India

²Renalyx Health Systems Private Ltd., Bengaluru 560004, India

³Rx Digi Health Platform Private Ltd., Bengaluru 560004, India

⁴Department of Electronics and Instrumentation Engineering, R. V. College of Engineering, Bengaluru 560059, India

Corresponding author: Rashmi R. Deshpande (rashmideshpande01@gmail.com)

ABSTRACT A computational 3D image generation using a single view with multi-color filter aperture (MCA) and multi-plane representation is a cost-effective approach and most useful when there is no option to acquire either stereo or multi-views with orientation at all. Although this approach generates 3D perception image that includes multiple objects with both similar and dissimilar colors having occluded by each other, it may be insufficient for virtual/augmented reality applications due to inaccurate depth. In this article, we obtain a more accurate geometric depth estimation by formulating a suitable relationship between inter-objects depth of the 3D scene in the depth-of-field (DoF) zone and its corresponding inter-image plane depths of a 3D perception image in depth-of-focus (DoFo) zone of a given camera under shallow DoF zone constraint. But, this shallow depth zone is configured to be dependent only on the focal distance between the lens and object while the remaining parameters such as aperture diameter, focal length, and sensor sensitivity are held at constant values. All-in-focus 3D perception image is synthesized from multi-plane images (MPIs) by utilizing the inter-image plane depths computed from the disparities caused across the boundaries and its smooth surface from image textures inside the respective boundaries of the 2D MCA image. The 2.1D sketch is used as a semantic segmentation technique to determine the number of objects in the 3D scene as one in-focus region and the rest as out-of-focus regions due to the circle of confusions (CoCs) on the fixed image sensor plane. The same enables both ordering of the image regions and identifying occlusion wherever applicable. An accurate depth 3D image is synthesized, replacing accurate inter-depths in place of inter-depth between MPIs used for 3D perception image. In the end, the paper summarizes few experimental validations for the proposed approach with some salient examples having depth gaps between 0.5cm to 10.5cm.

INDEX TERMS 2D object plane, 2D image sensor plane, 2.1D Sketch, 3D perception image, 3D virtual object, accurate depth, computational 3D imaging, depth of field, depth of focus, disparity, multi-color filter aperture (MCA), multi-region image (MRI), multi-plane image (MPI).

I. INTRODUCTION

All in-focus 3D image generation having accurate depths is an essential need in many computer vision-based applications for getting crucial details regarding the 3D scene. In this article, we mean all-in-focus image as approximately computable such that every pixel in this image is in focus. Creating a virtual 3D object or 3D image synthesis is essential for

any virtual/mixed reality applications, especially for medical diagnosis and surgical training appliances. Any conventional optical imaging camera system that is used for visible light intensity imaging, a real-world 3D scene or 3D object typically comprises of three intrinsic components, namely, (i) 3D object space, (ii) image space, and (iii) optical imaging system [1]. Among the various known optical imaging systems that use various ranges of electromagnetic rays, we consider only visible spectral light rays based on the photographic system. Moreover, the optical image sensors used in the

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar¹.

image space in most digital imaging systems would be either CMOS or CCD sensors, which are typically 2D in nature for capturing the image of any 3D object, which may be due to a low-cost compulsion. The method of recovering the depth information to represent a 3D image needs some tricky ways to project the irradiant rays emitted from the scene/object surface onto the 2D sensor plane. To date, most of the approaches suggested are based on either single view, stereo-view or multi-views [1], [2] for generating 3D perception qualitatively for the real 3D world scene from the irradiant optical stimulations (mathematically it would be of one to many mapping), [3], which we term as “3D perception image”. Among those approaches that produce 3D perception image, single view image-based 3D image generation is a most challenging one though it is advantageous because of two reasons, namely, (i) it involves less optics with an extensive computational complexity and (ii) it is helpful in some scenarios where stereo or multi-view imaging is not possible. The generation of a 3D image from a single-view 2D image is an active area of research due to the following reasons, namely, (i) suitable imaging can be readily obtained by a commercially available and affordable camera with either minor modifications or plug-in attachments in contrast to the devices that involve time-of-flight [4] structured illumination [5] or stereo camera [5], etc., (ii) unlike the conventional camera, computational camera composes of both geometrical imaging optics having built-in image processing/analysis functionality with computational resources [6], (iii) especially, in medical imaging applications, all focused images would give good clarity in identifying pathological symptoms. In any imaging system, exposure parameters, namely shutter speed and aperture size, either one or both of them may be used to get the desired amount of light radiation onto the sensor. Further, the sensitivity of the sensor also plays a crucial role. However, the camera is designed and



FIGURE 1. Red, green and blue filters arrangement as aperture.

developed based on the desired sensor and lens, resulting in fixed sensor and lens characteristics. Moreover, in conventional cameras, flexibility lies only in aperture and shutter speed. The flexibility in the camera system is achieved in many ways by the aperture used in the camera, viz, (i) radius of the aperture or size of the aperture [1], (ii) multiple apertures (e.g. color filter aperture) [7]–[9] (iii) sparse aperture (e.g. color-coded apertures) [10], (iv) offset aperture [11] and (v) differential aperture photometry [12]. In recent days, computational photography systems are gaining more importance over traditional photography because it comprises not

only optics and electronics to capture the image but also software-based computational manipulation to improve the imaging capability [13]. These manipulations may involve to compute attributes, such as shape, position, resolution, structure, unfocused pixel value or pixel regions, variation of aperture, and control the amount of incoming light, etc. It is important to distinguish between 3D perception scene image and the more precise 3D virtual scene or image. In a 3D perception image, the depths between the various portions of the image pertaining to their corresponding features are proportional to give a feel for the 3D environment. However, in the 3D precise image, the depths mentioned above are of a more accurate estimate, which depicts the real measurement of the scenes as required for auto maneuvering of robots or vehicles. This article will extend the computational 3D perception imaging system comprising a camera, multi-color filter aperture (MCA), and computation of depth by disparities among various boundaries identified using a 2.1D sketch for the single image [14]. In this article, we adapt computational photography techniques to generate a 3D image from a single RGB image acquired from a conventional digital camera with add-on MCA using red, green, and blue filters. We propose to obtain multi-plane images (MPIs) that attributes to all-in-focus image regions based on geometric optics principle as shown in Fig. 1 with associated inter-depths between them by suitable computational processes. Ultimately, we aim to deduce the quantitative model utilizing associated camera parameters to arrive at the depth parameter using the acquired 2D RGB image. In this research effort, we develop a quantitative model for the computational imaging system suggested in [14] that employs both depth cues from MCA [8] and composition of a 3D image using MPIs by using multi-region image (MRI) decomposition of the acquired image using a 2.1D sketch [15] as semantic segmentation. The primary goal here is to arrive at better accurate 3D images or 3D virtual objects in terms of associated geometrical and intrinsic camera parameters. In brief, the method relates accurate inter-image depth in multiple object layers in DoF and its corresponding inter-image depths in MPIs in the DoFo region. However, inter-depths for the decomposed MPIs are associated with disparity parameters using MCA 2D image and 2.1D sketch. Ultimately, it leads to an affordable device due to minor optics modification as MCA with red, green, and blue filters arrangement shown in Fig. 1 embedded in CANON 50mm f1.8 II lens to the readily available CANON 70D DSLR camera.

In summary, our contributions are:

- 1) Deduction of explicit relation between inter-depth associated with MPIs of 3D perception image with inter-objects depth of real-world 3D scene with a shallow DoF constraint.
- 2) Generation of more accurate geometric depth 3D image from a single 2D RGB image as three steps, namely, (i) Generation of 3D perception image as a set of MPIs, using inter-image planes depth and image surface, (ii) The set of MPIs are obtained by multi-regions

due to varying CoCs as a consequence of shallow DoF determined using 2.1D sketch as semantic image segmentation, inter-image plane depths from the disparities across the boundaries caused due to MCA and smooth surface by adjusting the image textures from respective regions of acquired 2D MCA image, and iii) Final more accurate depth 3D image is obtained from 3D perception image by realigning with their inter-objects depth.

Section 2 reviews the relevant depth estimation using MCA and associated image representation for view synthesis by several researchers along with limitations. Motivation for this project is highlighted in section 3. Section 4 explains the geometric 3D perception of the image from a multi-plane image based on ordered multi-region image and inter-depth estimation. Section 5 describes a 3D accurate depth image or virtual object from MPIs based on a newly deduced relationship between the inter-depth of the perception image and its 3D accurate depth image or virtual scene. Section 6 demonstrates the experimental simulation to validate the above developed 3D image, Section 7 gives the concluding remarks with the future scope of research.

II. RELATED RESEARCH WORKS

Although substantial research has been carried out on the depth estimation and reconstruction of 3D perception image (typically termed as all-in-focus image) using MCA, we notice that hardly any discussion is carried out on precise depth. As mentioned earlier, our main aim involves 3D imaging using a single image with a conventional camera with more accurate depth based on cues associated with MCA and DoF. Some crucial computer vision applications such as robotic, medical diagnosis, and surgery planning need more exact shape, size, and volume-based object recognition. Particularly in the now being considered technique, this boils down to establish an exclusive relationship between perception depth image and actual depth of the object in the 3D scene. Hence, our review is centered around depth estimation using both DoF and MCA to arrive at all-in-focus image synthesis.

A. REVIEW ON DEPTH ESTIMATION FROM DISPARITY USING MCA

Jaehyun Im *et al.* [15] have suggested an optically modified color-filter aperture (CFA) for tracking objects with both depth and position expressed in color shifts by assuming there would be color discontinuities when there are two different objects in the scene that are located at different depths. The phase correlation is employed in the adaptive mean shift algorithm for real-time implementation of the CFA approach. However, in this method, occlusion handling entirely depends on empirical threshold selections based on pixels, and the method does not assure to take care of the multiple objects with the same color at different positions along the optical axis. Yosuke Bando *et al.* [8] proposed the improved depth estimation and usage of matte to distinguish

between foreground and background using color misalignment utilizing several image-editing techniques. More crucially, the disparities increase when the lens is focused nearer, wherein inter-depths between any two objects in the scene may not be consistently the same. It is noted that the better inter-depth results, when objects are at 50cm to 250cm, and if the distance between object and camera is about twice as far away from the background. Further, performance degrades for smaller color misalignment and the objects with the same color. On the other hand, the restriction of a distinct colored object limits the practical utility of this approach.

The color-coded aperture (CCA) based method by Ivan Panchenko *et al.* [16] needs calibration to arrive at better results for different coded aperture designs and is based on the calibration and the effective focal length. The demerits of CCA are its limited field of view (FOV) [17], and are claimed to perform better than pinhole collimators only with a very localized point source.

The incorporation of light efficient CCA was designed and calibrated in skillful ways by Vladimir Paramonov *et al.* [18] to arrive at depth resolution in millimeters for a given image frame. Moreover, their contribution demonstrates DSLR, smart-phone, and compact camera to real-time 3D scene generation and depth-based image effects. But, this resolution is applicable at the center of the acquired image only. Unfortunately, both of these methods exploit GeForce GTX 780 hardware GPGPU for handling computation intensiveness. There is no assurance to guarantee inter-depth for the objects with the same color and textures in the image. But, the accuracy of CCA method depends on many factors such as working range, stronger texture information for depth extraction, lower accuracy in strongly defocused areas, and image restoration to get a sharp image. A single MCA camera-based methodology suggested by Seungwon Lee *et al.* [19] and modified multi-focusing image misalignment by Sangjin Kim *et al.* [19] to track the object upon exploiting sparsely extracting edges due to red, green, and blue color shifting vectors at each channel assures good estimate of depth only in the central region of the image. However, both of these methods assure good depth only at the center of the image without any mention of occlusion handling issues. Recently, we have proposed a novel method of generating a 3D image from a single 2D RGB image for a scene utilizing the depth from disparity based on MCA that handles multiple objects in the scene both with occlusions and with the same colors at different depths along the optical axis. It utilizes a 2.1D sketch instead of image matte on the MCA images to determine the L number of image regions corresponding to L objects in the given scene. The MRIs obtained using a 2.1D sketch are based on: one in-focus and $(L - 1)$ out-of-focus regions with varying blurriness as a natural consequence of differing CoCs. This blurriness is due to objects lying at different distances in DoF regions. Synthesis of 3D perception image as MPIs corresponding to MRIs at different inter-depths are obtained using inter-image region disparities. [14].

Although the previously discussed algorithms deal with various aspects of how to arrive at 3D perception image having all-in-focus image surface with proportional depth information, acceptable in some computer vision applications, may not be adequate in some applications such as robotic environment, auto-driving, and some mixed reality medical diagnostic studies. Again, we note that they exploit only image-side information to arrive at 3D perception image that does not use many camera parameters, lays the foundation for our motivation discussed in the following subsection.

B. REVIEW ON LAYERED REPRESENTATION FOR 3D PERCEPTION IMAGE

The requirement of more accurate inter-depths between the objects is essential for geometry-aware manipulation of 3D virtual objects in most virtual or augmented reality applications when the 3D scene comprises more than one object. It is important to note that any 3D object in the 3D scene possesses non-shape changing attributes such as color, overall size, orientation, and shape-changing attributes such as the number of sides (parallel or non-parallel, straight or curved), vertices, edges, boundaries, etc. Many of the above attributes may exist in the 3D perception image in general except for the accurate depth information.

In the literature, we notice that layered depth images (LDI) representation is used as one of the compact coding representations for multi-view synthesis to address occlusion and hidden information associated with a 3D scene [20]. Vincent Jantet, *et al.*, [21] suggested an improved virtual synthesis based on object levels distinction between background and foreground, using region growing segmentation technique. But this algorithm is developed for multi-image views.

The multi-plane image (MPI) representation is another way to synthesize a 3D image that enables to render each pixel to get scene independent new views with consistent non-occlusion when multiple objects are involved in the scene. Each image plane is considered as a RGBA image belongs to the part of frustum with apex at lens and positioned at fixed equally spaced depth obtained as inverse of disparity [22]–[25], [26], [27]. Recently, we have utilized MPIs as L number of fronto-parallel planes for synthesizing 3D perception image corresponding to L number of MRIs, and each MPIs are positioned in the DoFo zone belonging to image space with non-uniform interval using respective inter-depths [14]. But, it lacks the assurance of true depth in the image. As mentioned earlier, 3D geometric modeling is required for virtual reality applications for generating required views from the reference. It implies that inter-depths involved are to be established based on camera parameters, as indicated in the previous subsection.

III. MOTIVATION

In this article, before bringing up salient points for quantification of depth in 3D imaging, we distinguish 3D visual luminance image as: 3D perception image realized with an illusion of depth to the 2D luminance image, whereas 3D

virtual object or 3D precise synthesized image is realized with a true depth estimated from realistic parameters to the 2D luminance image. Further, we point out the 3D virtual object or 3D precisely synthesized image is advantageous in various real-world applications. We note few relevant points, which are relevant for our quantified 3D imaging systems using shallow DoF are listed below:

- 1) The insufficiency of 3D perception image obtained intuitively using a central theme for generating an all-in-focus 3D image from a single image needs to decompose the given image data as appropriate ordered multiple layered images. On adapting only image data as a set of varying blurred or defocus image regions, we could generate only 3D perception images from the depth from disparities and decomposed multiple layer images.
- 2) Inability in acquiring all-in-focus 2D images or accurate depth-based 3D image corresponding to entire 3D objects of the scene using a 2D image sensor as there is no simplistic and cheaper known 3D image sensor other than costly 3D sensor like time of flight [2], Microsoft Kinect device [28], etc.,
- 3) Limitations associated with DoF, especially under limited illumination conditions [1],
- 4) Technically, deep to shallow DoF zones are parametrically controllable by aperture, focal length, and distance of the object from the objective lens of the camera [1],
- 5) The change in aperture radius will not only vary DoF zone but also FoV [1].
- 6) Interestingly, for a fixed aperture size or FoV, on one hand, DoF is directly proportional to object distance with focal length held constant, and on the other hand, DoF is inversely proportional to focal length when the object distance is held constant [1].
- 7) Very shallow DoF results very small all-in-focus 2D image [1],
- 8) DoF is sensitive on digital format size or CoC criterion for a camera [1],
- 9) Professional photographers use DoF photography, which involves controlling deep to shallow DoF for getting creative effects. Especially shallow DoF isolates the required portion of the image in the larger scene or FoV, since a shallow DoF zone in the object space leads to a small all-in-focus region along the optical axis, but the regions before or after the all-in-focus region along the optical axis are out-of-focus [1],
- 10) Changing the DoF by varying focal length is impossible once the lens is decided for the camera unless it is a liquid lens, which amounts to the extra cost.
- 11) The DoFo zone interval in image space is related to the DoF zone interval in object space and is dependent on both intrinsic and geometric parameters of the given camera [1].
- 12) The concept of exploiting DoF is used in the field of computer graphics or visual synthesis or virtual

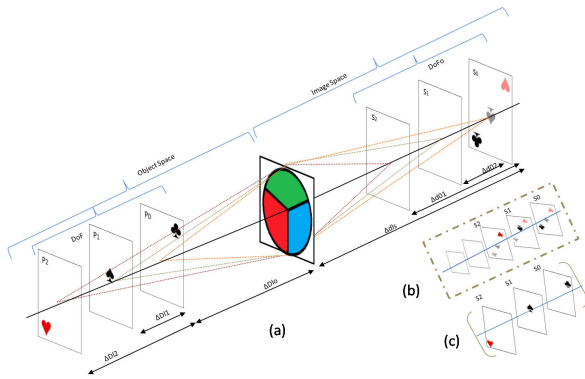


FIGURE 2. (a) MCA camera set up for varying defocused captured 2D image and computed 3D perception image as MPis and transforming to 3D virtual object/ estimated 3D image. (b) 3D perception image as MPis (c) 3D Virtual object or estimated 3D image.

reality [20]. But, it is used as a software simulation for generating 3D perception image to get the feel of photorealism only.

The above-enlisted issues and properties have motivated us to arrive at an affordable 3D image generation system. That would find applications for 3D imaging for biological surface tissues, collision-free autonomous driving, and maneuvering industrial robotics, where the objects that need to be imaged are always very close to the camera lens. This small focal distance (object to lens distance) is an obvious consequence of wide-angle aperture, and FoV [1]. A shallow DoF is a direct consequence of focal distance, which corresponds to a small DoFo leading to a smaller in-focus image region, and many image regions are out-of-focus, causing blurriness due to CoCs [1]. These factors prompted us to derive virtual positions for multiple image sensor planes as consequences of disparities caused due to blurriness on the fixed image sensor in the pre-configured camera setup to the acquired small in-focus region and remaining out-of-focus regions during single image acquisition of the scene. In this article, among the many interpretations for the 3D image corresponding to a 3D scene in nature, our focus is on geometrical aspects restricted either as a 3D perception image, and actual 3D image from a single image. The perception 3D image referred to the image of a scene that represents the geometrical environment displayable as an all-in-focus image with qualitative depth information, whereas a precise 3D image as a virtual object would be able to project with more realistic depth information.

IV. GEOMETRIC 3D PERCEPTION IMAGE FROM MULTI-PLANE IMAGE

In our previous work, we had considered single 3D scene image representation as a composition of multi-plane images from 2D multi-region images by exploiting the objects with varying amounts of blurriness, lying at a varying depth, that enables us to arrive at perception depth [19].

A. DECOMPOSITION OF MULTIPLE IMAGE REGIONS AND COMPUTATION OF INTER-DEPTHS

Typically, any given scene comprises multiple objects with the same color or different colors at varying depths, either with and without overlaps. These objects would result in varying blurriness in a captured 2D image due to DoF configurations. The first step is to decompose the given 2D image into various image regions corresponding to the L number of objects in the scene. The decomposition into multiple image regions with boundaries is obtained based on one in-focus region and $(L - 1)$ out-of-focus regions. For any camera setup shown in Fig. 2, in the captured 2D RGB image, focussed portion or object in the scene results in a region with no blurriness, and the remaining $(L - 1)$ regions corresponding to those objects yields varying blurriness that depends on other positions in 3D scene project on 2D sensor plane due to varied CoCs as consequences of DoF [1]. The second step is to obtain the disparities along the boundaries, which are of the above spaces are dependent on four parameters, namely (1) focal length, (2) sensor sensitivity, (3) aperture type with size, and (4) camera focus distance. In any camera system, with a decided sensor type and lens, once manufactured, makes those parameters constant. It is flexibly easy to select either shallow or deep focus region-range from adjusting d_a compared with focus distance. Further, for a specific object in a 3D scene, we suggest configuring d_a with a suitable value such that we obtain an appropriate FoV with shallow DoF that results in a small focussed image on the sensor plane and remaining objects located at different distances on the optical axis behind it generates differing defocus image regions due to CoCs. Indirectly, we presume that the acquired single image consists of L number of MRIs for the objects in the 3D scene. Since our intention is to tackle the occlusion aware decomposition of single image into MRIs, associated with varying amount of defocus, 2.1D sketch formulated by either Nitzburg and Mumford [15] or a 2.5D sketch David Marr [29] can be used, which enables to generate the 3D scene representation.

Especially in this article, we consider a 2.1D sketch since it segments the image regions that handle occlusion explicitly to decide the partial ordering of regions by indicating their ordering. This method exploits depth cues from edges, curves, cusps, crack tips, and T -junctions. Further, we solve energy minimization problem [30] using Divided RECTangle (DIRECT) search-based deterministic global optimization method [31], [32] that slices the given single 2D image into a set of MRIs with ordered top to bottom layers. This is based on similarly clustered gray levels and/or colors apart from distinguishing their overlaps or occlusions. Also, it gives information about the total number of different regions/patches present in a given 2D image.

Any given 2D image is described with domain \mathcal{D} , where $\mathcal{D} = ((i, j), 0 \leq i \leq (M - 1), 0 \leq j \leq (N - 1))$ and on adopting principle of 2.1D sketch as described by Nitzburg

and Mumford [15], we arrive at L number of MRIs as

$$\mathcal{D} = \bigcup_{l=0}^{(L-1)} \mathcal{R}'_l, \text{ where } \mathcal{R}'_l = \bigcup_{l < k, \mathcal{R}_l < \mathcal{R}_k} \mathcal{R}_l \quad (1)$$

where $<$ denotes partial ordering includes occlusion.

B. REVISITING THE PERCEPTION 3D IMAGE FROM MPI

We recall from our earlier research work the all-in-focus image as 3D perception image represented as L number of fronto-parallel MPIS, and its inter-plane depths is given by:

$$\Delta d_{0l} = \frac{\delta_l \times d_{ls}}{(d_a + \delta_l)}, \quad l = 1, 2, 3 \dots, (L - 1) \quad (2)$$

In natural scenario, these partial or full occlusions between the objects lying in a 3D scene may not be very well ordered but these occlusions are caused due to the objects at different heights and locations other than focus distance in the 3D scene. Further, as per geometric optics, shallow DoF would result in shift varying blurred regions across the captured single image. In our previous research, we are inspired to use 2.1D sketch since it enables occlusion aware image segmentation for decomposing the single image into MRIs, and allowing to take into account the partial occlusion of the farther object by those with nearer ones [1]. Briefly, the salient steps for perception 3D image generation algorithm is enumerated as below:

- 1) Identify various image regions using salient boundaries based on the change in disparities and obtain inter-depths between the regions using (2).
- 2) Order the MRIs from foreground to background layers using depth cues as per 2.1D sketch.
- 3) Generate 3D image as a composition of L fronto-parallel MPIS as a part of a frustum with a reference to the image sensor planes as shown in Fig. 2(b) (shown in dotted box as MPI frustum), each at the distance Δd_{0l} , $l = 0, 1, \dots, (L - 1)$ and its smooth surface of l^{th} MPI as $g(i, j) \in \mathcal{R}_l + \Delta d_l$.

The 3D perception image obtained above can be expressed as:

$$g(i, j) \in \mathcal{D} = \bigoplus_{i=0}^{(L-1)} (g(i, j) \in \mathcal{R}'_i) + (\Delta d_{0l})$$

$$\mathcal{D} = \bigcup_{l=0}^{(L-1)} \mathcal{R}'_l \ \& \ \bigcap_{i=0}^{(L-1)} \mathcal{R}'_i = \emptyset \quad (3)$$

where \mathcal{D} and \bigoplus denotes domain of the given image and compositing operation, respectively.

C. LIMITATION IN THE 3D PERCEPTION IMAGE

The above obtained inter-image depths from the boundaries of decomposed multiple image region and synthesizing 3D perception image may not assure the accurate depth, which is one of the vital requirements for the all-in-focus 3D image or virtual 3D object. Essentially, a virtual 3D object is nothing

but a 3D model of the realistic scene object. Indeed 3D image generation with accurate inter-image region depth is crucial for applications like biomedical, industrial robotics, and auto-driving vehicle maneuvering. It is easy to infer that the perception depth is not scale-invariant. For example, in robotic navigation or AR/VR application, there is a need to generate a 3D image scene that should be photo-realistic and deformable. We address this in the next section by relating the perception depth to the actual depth.

V. IMPROVED DEPTH 3D IMAGE OR VIRTUAL 3D OBJECT USING GEOMETRIC OPTICS

In the previous section, although the 3D perception image uses blur caused due to the DoF phenomenon, the sharpness in the image space is not a specific point, but the acceptable area caused due to CoC with no noticeable distortions. This fact implies that there must exist a relationship between the object space and image space for any camera in terms of specific parameters due to the following characteristics:

- 1) For a given FoV, the distant DoF (termed as far DoF) zone maps to the nearest DoFo (termed as near DoFo) zone for the camera and vice versa. The other points in the DoF zone starting from far point to near, will map in reverse order in DoFo zone. These DoF and DoFo zones are also subsets of object and image spaces, respectively. In this space, there would be negligibly noticeable degradation or blurriness in the image, if an image is focused well.
- 2) For a specific FoV of the camera and a 3D object placed in the DoF space, the irradiant rays originated from the points on various virtual planes in DoF zone form the least CoC or most focused image points at specific DoFo space points if the sensor plane is positioned at those locations along optical axis. But, it forms non-focused image points or larger CoC in any other location beyond or before this location (refer Fig. 3(a) and 3(b)).

Briefly, the far-field and near-field 3D scene points in object space forms a near focus and far focus image points in the image space, as shown in Fig. 3(a) and 3(b), respectively. The actual positions of the above spaces are dependent on four parameters, namely (1) focal length, f (2) sensor sensitivity, (3) aperture type with diameter, d_a , and (4) focus distance from the camera. In any camera system, rigid lens and sensor parameters would get frozen to constant value while manufacturing. Among the latter two parameters, we can consider the choice for shallow or deep DoF using camera to object distance, whereas aperture type and diameter d_a are pre-configured since we are using MCA.

On noting the above pre-configurations of many parameters, for shallow DoF range, the relationship is deduced for inter-depths between objects in the 3D scene as computed inter-depths between corresponding two MPIS for non-transparent and non-speculative 3D objects at the stationary position while imaging for optical setup shown in Fig. 3(b). These configurations in the imaging optics yield

the following characteristics when we look at both object space to image space together.

- 1) There exists a specific point on the DoFo zone in image space along the optical axis where the image sensor could be positioned to get a focused image points corresponding to a specific object surface belonging to the 3D scene in the DoF zone of the object space of the camera.
- 2) In any conventional camera, the 2D sensor is fixed at a distance d_{fs} from lens to capture the focus image region for a specific shallow surface of the object in the 3D scene and many out-of-focus image regions due to varying size CoCs proportional to the points lying on the optical axis in DoF zone with reference to the focused surface for remaining objects in the 3D scene.
- 3) Though the width of the DoF space and DoFo space are not equal, and apart from being distributed unequally, for the above specific conditions in the camera setup, it is possible to know the corresponding far and near field points in DoF space to near and far focus points in DoFo space.

A. ALL-IN-FOCUS IMAGE AS MULTI-PLANE IMAGES

In Fig. 3(a) and (b), we have shown only $l = 3$ number of planes for the sake of clarity, but we could have L number of virtual sensor planes which correspond to L number of MRIs. For the sake of generality, let us assume $\mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_{(L-2)}, \mathcal{S}_{(L-1)}$ be the virtual image sensor planes belong to the image space pertaining to a specific camera setup. Furthermore, we interpret these planes are specific planes of frustum whose base is fixed sensor plane \mathcal{S}_0 and the apex is at the lens. From the geometric optics principle, we know that the focussed image region on a fixed plane \mathcal{S}_0 is due to the object in the 3D scene at the focal distance from the lens that belongs to DoF zone in the object space. The objects before and after focal distance form defocused image regions on \mathcal{S}_0 . Obviously, the $(L - 1)$ number of objects which are at different distances from the camera lens would create a corresponding $(L - 1)$ number of defocused image regions on \mathcal{S}_0 depending on the area of respective CoCs. At the same time, every l^{th} MRIs would have resulted in the in-focus image if there is a virtual image sensor plane \mathcal{S}_l in the image space of the camera whose inter-depth as Δd_{0l} with respect to the \mathcal{S}_0 .

B. MULTI-PLANE OBJECTS AS 3D SCENE

In this subsection, we denote any 3D scene comprising of many objects that need to be imaged could be represented as slices of multi-plane objects as layered DoF as described by David C Schedl and Michael Wimmer [20], [33]. This is also in a similar line to the 3D image as multi-plane images discussed in the previous subsection IV(B). However, this could be viewed in two ways, as one being multi-plane object layers as a part of the frustum in the object space that is

intended to yield in-focus image region in DoFo space and arranging planes towards far focus point of DoF zone, and the other being multi-plane object layers as part of a frustum with referencing to the object plane which is intended to yield in-focus image in DoFo and arranging planes towards near focus point in DoF region. In both cases, the 3D scene is viewed as L number of slices as multi-plane scenes belonging to part of frustum coinciding with DoF with inter-object depths as $\Delta D_{0l}, l = 1, 2, \dots, (L - 1)$ as shown in Fig. 2(c). Essentially, we use this fact to deduce the accurate relationship between inter-depths in the scene and 3D images represented as MPis.

C. RELATIONSHIP BETWEEN INTER-IMAGE PLANES DEPTH IN DoFo AND INTER-OBJECTS DEPTH IN DoF

Due to the geometric optics principle in any camera system, for a specific object position in FoV and camera configurations, back and front DoF regions in object space form the corresponding front and back DoFo regions in the image space, respectively. To arrive at L objects corresponding to L MPis, we consider DoF virtual planes as $\mathcal{P}_l, l = 0, 1, 2, \dots, (L - 2), (L - 1)$ along the optical axis, respectively, belonging to the back DoF zone of the object space. When a camera is focused on the object, the selected virtual object planes form corresponding virtual image planes $\mathcal{S}_l, l = (L - 1), (L - 2), \dots, 2, 1, 0$, which are not only reverse ordered but also located at non-equidistant intervals.

As per [32], 3D image is generated using fusion of focused multiple layer image planes at points $\mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_{(L-1)}$ on DoFo space in correspondence to the DoF planes $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_{(L-1)}$ in DoF space along optical axis, respectively. 3D perception image seems to be blind for parameters related to camera, object-image geometric parameters, and lighting parameters except for the fact that there exist L number of objects in the 3D scene while imaging is carried out. It is important to consider the imaging process parameters for generating a 3D image, which plays a crucial role in computational photography.

In order to compute a 3D image with more realistic depth, we consider the object, image, and camera as shown in Fig. 3(b). For the sake of simplicity, we consider the case where we position the given 3D scene, in which the top portion of the 3D scene as C' (at object layer \mathcal{P}_0) yields an in-focus image region on the fixed image sensor \mathcal{S}_0 (as shown in Fig. 3(b)) and bottom portion of the 3D scene is assumed to be at A' (at object layer \mathcal{P}_2), which would form an in-focus image at virtual plane \mathcal{S}_2 in front DoFo zone of image space but captured as defocused image at the image sensor plane. It amounts to place the 3D object in the back DoF zone instead of the front DoF region. The irradiant light ray coming from the point A' forms a conical bundle at object plane \mathcal{P}_0 with sectional diameter δ_0 . Similarly, the irradiant light rays reflected from the object top position at C' on object plane \mathcal{P}_0 would create a sharp image at sensor plane \mathcal{S}_0 in image space at point C . Considering similar triangles $EA'D$

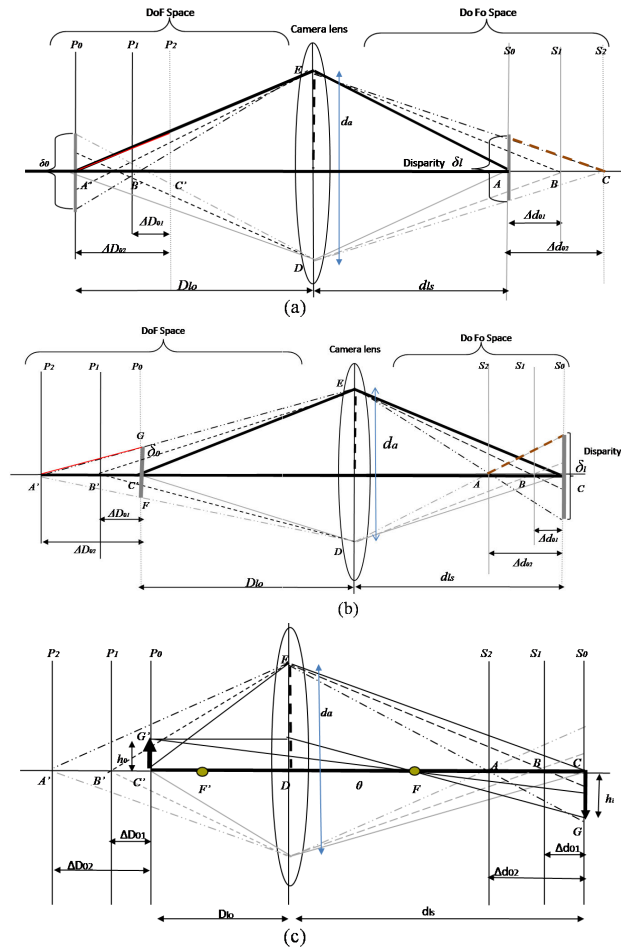


FIGURE 3. (a) Thin lens model for near focus planes (b) Thin lens model for far focus planes, (c) Thin lens model for lateral magnification.

and $GA'F$ in Fig. 3(b), we arrive at

$$\frac{D_{lo} + \Delta D_{0l}}{d_a} = \frac{\Delta D_{0l}}{\delta_0}, \quad \text{where } l = 1, 2, 3. \quad (4)$$

where D_{lo} is focal distance and ΔD_{0l} is inter-object depth between focused point C' and non-focused point A' . Noting magnification $M = \delta_l / \delta_0$, the equation for D_{lo} from (4), can be expressed as:

$$D_{lo} = \frac{\Delta D_{0l} (Md_a - \delta_l)}{\delta_l} \quad (5)$$

The depth estimation using geometric optics with Gaussian lens equation mentioned as:

$$\frac{1}{(D_{lo} + \Delta D_{0l})} + \frac{1}{d_{ls} - \Delta d_{0l}} = \frac{1}{f}, \quad (6)$$

where f represents focal length, distance between far focus point in DoF space to the lens is $(D_{lo} + \Delta D_{0l})$ (point A' in Fig. 3(b)) and distance between far focus point in DoFo space to the sensor as $(d_{ls} - \Delta d_{0l})$ (point A in Fig. 3(b)). Now, we recall the simplified relationship for far point A' using the intercept theorem by J.N.P. Martel *et al.* [34] in-terms of

intrinsic and extrinsic parameters as,

$$\frac{1}{(D_{lo} + \Delta D_{0l})} = \frac{d_a (d_{ls} - f) - f \delta_l}{d_a f d_{ls}}. \quad (7)$$

Further, we would like to mention that in J.N.P. Martel *et al.* [34] the real-time depth estimation is carried out by using a focal plane processor array with a tunable focus lens. Naturally, this method is not directly applicable to use the conventional camera because it involves changing both the image sensor and lens. Moreover, the tunable lenses are unstable in performance. But, our intention here is to estimate the depth using a rigid lens and fixed sensor with simple camera optics attachment with a computational approach from disparities. Magnification M can be represented as:

$$M = \frac{d_{ls} - f}{f}, \quad \text{where } f \text{ represents the focal distance.} \quad (8)$$

Due to above equations, (7) can be re-written as

$$\Delta D_{0l} = \frac{(d_{ls} d_a)}{(M d_a - \delta_l)} - D_{lo}, \quad (9)$$

where $l = 1, 2, 3 \dots L$ object layers.

On substituting D_{lo} from (5) in (9) we arrive at:

$$\Delta D_{0l} = \frac{(d_{ls} \delta_l)}{M (M d_a - \delta_l)} \quad (10)$$

Re-writing δ_l from (2) in terms of an inter-image layers depth between image sensor plane to the l^{th} image layer plane, δ_l as

$$\delta_l = \frac{d_a \Delta d_{0l}}{(d_{ls} - \Delta d_{0l})} \quad l = 1, 2, 3 \dots, (L - 1) \quad (11)$$

and substituting the same in (10), we arrive at a relationship between inter-objects layer depth between focus layer to non-focused l^{th} layer as below:

$$\Delta D_{0l} = \frac{\Delta d_{0l}}{M \left(M - \frac{(M+1)}{d_{ls}} \Delta d_{0l} \right)}. \quad (12)$$

Alternatively, (12) can be expressed in the form below:

$$\Delta D_{0l} = \frac{\Delta d_{0l}}{M^2 \left(1 - \frac{(M+1)}{M} \frac{\Delta d_{0l}}{d_{ls}} \right)}. \quad (13)$$

In the similar lines, we can obtain ΔD_{0l} for back DoF zone as:

$$\Delta D_{0l} = \frac{\Delta d_{0l}}{M^2 \left(1 + \frac{(M+1)}{M} \frac{\Delta d_{0l}}{d_{ls}} \right)}. \quad (14)$$

The interesting point regarding the above relationship is that the inter-depth of the virtual 3D object is expressed only as image space parameters and magnification of the camera, where the latter is measured without direct object space parameters. In Fig. 3(c), based on the geometric optics principle, we infer that transverse and axial magnifications of the camera setup will alter not only the size of the image but also the width of the object layers and corresponding image layers. Further, we also note that the equidistant points on the

optical axis in the object space will distribute unequally on image space along the optical axis. The relationship shown in (13) is used for obtaining true depth at various points on the 3D object surface with its corresponding point on the generated 3D image surface. Note here that the relationship given in (10) can also be used to identify the inter-depth from disparity δ_l itself without generating a 3D image, that helps in generating virtual objects required for mixed reality applications. Further, this way of identifying inter-object layer depth information is helpful in object detection, especially for industrial robotic and auto vehicle anti-collision applications. For the sake of simplicity, we have considered geometric optics with a thin lens camera setup. However, similar relations could be deduced for thick lens or more than one lens also though it is not discussed in the current study.

D. 3D VIRTUAL OBJECT FROM MULTIPLE PLANE IMAGES AND IMPROVED INTER-DEPTHS

A 3D topographic surface of a non-transparent and non-speculative 3D object visible under FoV of any given camera along its optical axis is defined as a discrete luminance function represented in the form of a 2D array as below:

$$G_{(i,j)} = Z_{i,j} \quad (i, j) \in \mathcal{D} \quad (15)$$

where $Z_{i,j}$ denotes the maximum topographic value of the 3D object surface with reference to the virtual plane \mathcal{P}_0 in the DoF zone of the camera’s object space. The above-said object surface points on various virtual object layers have corresponding positions \mathcal{S}_l , $l = 0, 1, \dots, (L - 1)$ of the image sensor planes that would generate in-focus MPI layers that are formed as out-of-focus image regions, if acquired at \mathcal{S}_0 . As pointed out earlier, out-of-focus image regions on an image sensor plane are caused due to the object points lying on any \mathcal{P}_l , the l^{th} object layer that is farther away from the in-focus object layer \mathcal{P}_0 and give rise to disparity as δ_l on the image sensor plane \mathcal{S}_0 .

E. SYNTHESIS OF VIRTUAL 3D OBJECT FROM 3D PERCEPTION IMAGE

In this subsection, we utilize multiple planes for the object representation similar to MPIs as shown in Fig. 2(a) that is a one-to-one correspondence with the layers of MPI, except for the fact that inter-object layer as ΔD_{0l} instead of Δd_{0l} . Any 3D virtual object is synthesized in similar lines like the 3D perception image described in section IV(B) using the relationship between inter-object layers depths given in equation (3) as below:

$$G(i, j) \in \mathcal{D} = \bigoplus_{i=0}^{(L-1)} (g(i, j) \in \mathcal{R}'_l) + \Delta D_{0l}$$

$$\mathcal{D} = \bigcap_{l=0}^{(L-1)} \mathcal{R}'_l \quad \& \quad \bigcup_{i=0}^{(L-1)} \mathcal{R}'_l = \emptyset \quad (16)$$

Indirectly, adding ΔD_{0l} to each pixel in the region enables to get in-focus image regions from the defocused pixels due

to the effect of shallow DoF. We interpret the above synthesized virtual 3D object or image as nothing but a 3D object representation expressed with actual depth in similar lines to 3D perception image. Again, this may find advantageous to surgical devices for probing and dissecting or LASER device pointing more accurately. The virtual 3D object is helpful to view using virtual or augmented or mixed reality visual systems. We believe that it aids in medical imaging appliances to analyze the medical practitioner for diagnosis of disease symptoms.

VI. DISCUSSION ON EXPERIMENTAL EVALUATIONS FOR 3D IMAGING SYSTEM

In this section we discuss some simulated experimental results to validate the quantified 3D image generation formulated in previous sections.

A. MCA CAMERA SETUP FOR IMAGE ACQUISITION

The experimental setup is formed using the scheme shown in Fig. 3(a) and 3(b) with the specifications as the diameter of the aperture being $d_a = 12mm$, three RGB filters arranged as per Fig. 1, and distance between lens and sensor being $d_{ls} = 52.63mm$ for canon 50mm f1.8 II when focused at 100cm. Hence, we determine the minimum inter-object in DoF range object space and its shifts in terms of pixels for the camera (CANON EOS 70D DSLR Camera) for our demonstration purpose.

B. DATA SET CREATIONS FOR EVALUATIONS

To demonstrate and evaluate the practical performance of the proposed approach, we need to create relevant image data set because there seems to be no standard color filter aperture image database. We have created the following sample MCA 2D RGB images by various arrangements with known inter-depths between the objects.

- Scene consisting of non-overlapping similar color objects arranged one behind the other but focal distance is on one of the intermediate object,
- Scene similar to above objects with different colors,
- Scene with one object overlaps the other with varying sizes and colors,
- Scene with an object in slant position, overlapped with its middle portion and visible at both sides,
- Scene with different color and sized objects one behind the other such that mid portions are occluded.
- Scene with both overlapped and non-overlapped objects.
- Most of the above scene with inter-object depth is maintained between 0.5cm to 10.5cm gap.

C. EXPERIMENTAL EVALUATIONS

This subsection describes both qualitative and quantitative results on images with non-trivial scenarios captured using MCA camera with color filter arrangement shown in given Fig. 1.

The qualitative evaluations on the depth map for all the generated data sets are discussed and the results are displayed in Fig. 4.

| | I | II | III | IV | V | VI |
|---|---|----|-----|----|---|-------|
| a | | | | | | 20.16 |
| b | | | | | | 18.23 |
| c | | | | | | 22.1 |
| d | | | | | | 19.45 |
| e | | | | | | 28.7 |
| f | | | | | | 16.8 |
| g | | | | | | 19.15 |

FIGURE 4. Illustration scene images of salient scenarios with overlaps, inclined positioning with respect to the optical axis of the MCA camera, and various combinations are displayed as rows with its intermediate and final stages results as columns, col (I): MCA images, col (II): 2.1D sketch, col (III): Depth map, col (IV): scatter map col (V): Accurate inter-depth 3D image plot, col (VI): Runtime in seconds, row (a): MCA image of pots placed one behind the other with a gap 3cm with no overlap, row (b): MCA image of non-overlapped blue and green blocks with a gap of 3.5cm and orange block overlapped by blue block at one end and green block at the other end at a gap of 4cm, row (c): Six color blocks one behind other at a gap of 3cm, row (d): MCA image of four similar color bottles kept at one behind other with a gap 7cm, 9cm, and 8cm, row (e): MCA image of the wooden spatula is overlapped by a red tomato, orange carrot and, green tomato, which are placed at the gap of 0.5cm, 4cm, 4cm, and 6.2cm, row (f): MCA images of three different color and size cups overlapping each other placed one behind other at a gap of 6.5cm and 10.5cm row (g): MCA image of a knife in slant position along the optical axis with one end at 3cm front and the other end at 10cm behind overlapping orange carrot.

- 1) In col. (II) of all rows, we note that the 2.1D sketch yields all the regions in the scene image with comparatively good accuracy in the boundaries irrespective of their occlusions, which play a vital role in arriving at good 3D perception and actual 3D image.
- 2) Depth maps for MCA images shown in col. (III) of Fig. 4 is represented in terms of gray levels, where higher gray value for a region, implies that it is nearer to

- the lens. Also, it has a very narrow range of gray levels since the inter-object depth gap is minimal.
- 3) In col. (IV) of all rows indicates the scatter map. Rows (b), (e), (f), and (g) indicates the detection of the overlaps that exist in the images very accurately, and the same is utilized to arrive at MPI's, resulting in good 3D images in the respective rows shown in col. (V).

TABLE 1. Quantitative comparison results reported with actual values between estimated inter-depths of 3D image and their respective ground truth inter-depth for salient examples for designated objects, A, B, C, D, E, and F present in the scene and its resultant 2D image for respective regions.

| Image details | | MCA images shown in Fig. 4 | | | | | | |
|---|-------------|----------------------------|--------|---------|---------|---------|--------|---------|
| Depth details | Image label | row(a) | row(b) | row(c) | row(d) | row(e) | row(f) | row(g) |
| Ground truth inter-object depth in mm when focussed object at 100cm from camera | A | 90 | 35 | -60 | 160 | -35 | 170 | -30 |
| | B | 60 | 35 | -30 | 90 | -40 | 65 | 100 |
| | C | 30 | 75 | 0 | 0 | 0 | 0 | 0 |
| | D | 0 | 0 | 30 | -80 | 40 | - | - |
| | E | -30 | - | 60 | - | 102 | - | - |
| | F | - | - | 90 | - | - | - | - |
| Estimated inter-image planes depths in mm for 3D perception image | A | 0.1796 | 0.1438 | -0.2172 | 0.323 | -0.2172 | 0.2868 | -0.0722 |
| | B | 0.1079 | 0.1438 | -0.1084 | 0.2868 | -0.2172 | 0.217 | 0.2868 |
| | C | 0.0720 | 0.2154 | 0 | 0 | 0 | 0 | 0 |
| | D | 0 | 0 | -0.0720 | -0.3630 | 0.1438 | - | - |
| | E | -0.1084 | - | 0.2154 | - | 0.2868 | - | - |
| | F | - | - | 0.2868 | - | - | - | - |
| Estimated inter-object depth in mm for 3D virtual object using (13) | A | 69 | 55 | -71 | 126 | -44 | 129 | -25 |
| | B | 40 | 55 | -37 | 116 | -44 | 84 | 116 |
| | C | 26 | 84 | 0 | 0 | 0 | 0 | 0 |
| | D | 0 | 0 | 26 | -67 | 54 | - | - |
| | E | -37 | - | 84 | - | 116 | - | - |
| | F | - | - | 116 | - | - | - | - |

TABLE 2. Performance errors between estimated inter-depth in the generated 3D image and true inter-depths in the scene.

| Image details | | MCA images shown in Fig. 4 | | | | | | |
|---|-------------|----------------------------|--------|--------|--------|--------|--------|--------|
| Error details | Image label | row(a) | row(b) | row(c) | row(d) | row(e) | row(f) | row(g) |
| Absolute error at each label in mm | A | 21 | 20 | 11 | 34 | 9 | 41 | 5 |
| | B | 20 | 20 | 7 | 26 | 4 | 19 | 16 |
| | C | 4 | 9 | 0 | 0 | 0 | 0 | 0 |
| | D | 0 | 0 | 4 | 13 | 14 | - | - |
| | E | 7 | - | 24 | - | 14 | - | - |
| | F | - | - | 26 | - | - | - | - |
| Relative error at each label of MCA image | A | 0.233 | 0.571 | 0.183 | 0.2125 | 0.257 | 0.241 | 0.166 |
| | B | 0.333 | 0.571 | 0.233 | 0.288 | 0.1 | 0.292 | 0.16 |
| | C | 0.133 | 0.12 | 0 | 0 | 0 | 0 | 0 |
| | D | 0 | 0 | 0.133 | 0.1625 | 0.35 | - | - |
| | E | 0.233 | - | 0.4 | - | 0.137 | - | - |
| | F | - | - | 0.288 | - | - | - | - |

- 4) From col. (IV), we see that the scatter map display depicts that the proposed algorithm resolves all regions very well.
- 5) Col. (VI) displays the running time in seconds to execute the algorithm for arriving at 3D virtual objects corresponding to the acquired 2D MCA image. The quoted runtime for algorithm is applicable for MATLAB based implementation that runs on specific hardware and software platform, having the Intel®core i5-4460 CPU @ 3.20GHz × 4 with pre-loaded UBUNTU 16.04 LTS operating system and 8.00GB memory without considering any effort on many possible optimizations. There are many scopes for improving the execution time by adopting optimization of algorithm not only with respect to hardware and software on any intended application scenarios but also with the elimination of redundant computations and memory requirements.

Table 1 presents the quantitative comparisons between ground truth and estimated values for all the images shown in Fig. 4. However Table 2. gives absolute and relative depth errors. The absolute error is expressed as the absolute difference between the ground truth and estimated depth at image labels to signify the amount of deviation for obtained depth

at labeled locations. The relative errors expressed as a ratio between the absolute error and the ground truth depth at labeled points signify the percentage error with respect to the ground truth depth. From Table 2, we note that the estimated inter-depth values are very near to the true values.

We present the detailed performance analysis as below:

- 1) On referring to row (c), row (e), and row (f), we see the error in estimating the respective inter-region depth is significantly less.
- 2) From row (a), row (c), and row (d), we notice that the error in estimated inter-depth is lesser when the actual gap between is less.
- 3) The estimated inter-depth value seeming to be less erroneous when the object is near to the focussed object (refer rows (a), (b), (d), (e), and (f)).
- 4) From Table 1, we note that the estimated inter-region depth in perception 3D image is not near-to the ground truth inter-objects depth. On the other hand, the computed inter-image depth in the 3D image from the newly derived relationship yields better results.
- 5) From the results shown in row (a), we could infer that the error increases as the actual gap between the objects increases.

- 6) As per row (e), the error is less when the gap between the objects is less, but on the other hand, it also exhibits the error that is comparatively more for the object that lies on the near focus side, though the object gap is same. Similar results are exhibited in rows (d) through (g).
- 7) On observing all rows (a) through (g), we infer that errors are minimal when the object depth is nearer to the focal distance.
- 8) The performance accuracy is similar in both rows (c) and (e).
- 9) From the results depicted in a row (b), we infer that the estimated depth is the same for objects A and B with less estimated error than the object C, which is behind object D. But, we note that it exhibits the slant orientation as expected.
- 10) From row (g), the estimated depth between B and C is comparatively lesser than that of inter-object depth between A and B since their inter-object is based on near-focus conditions. A similar characteristic is exhibited in row (c).
- 11) From Table 2, we observe that errors are less for near focus compared to far focus depth for all figures in rows (a) through (g).
- 12) We notice that in Table 2, for rows (b), (c), and (d) the errors are less near to the focus region than the errors away from focus regions.
- 13) Especially, in rows (a), (c), (d), and (e) with the focus region from the camera on one of the intermediate objects in the scene instead of the nearest object to the lens, error estimation for inter-depth are less for those objects behind focus object than the objects those in front of the focussed object. The same trend is true in the rows (b), (f), and (g), where focal distance is nearer to lens.
- 14) In general, the accuracy of the inter-object depth results are better when the object is focused, and the remaining objects are placed between the focus distance and distant endpoint in DoF zone of object space.
- 15) Table 2 shows that the relative depth measured for a near focus object has an error of 10%, and the percentage error increases gradually up to 57% as the objects are placed laterally away from the focused object.

VII. CONCLUSION AND FUTURE SCOPE

We have presented a novel method of synthesizing a more accurate geometric depth 3D image from the 3D perception depth image using the newly formulated inter-depth relationship. We have derived a specific non-linear dependency between inter-depths of two objects lying in 3D scene and corresponding inter-depths of two MPIs in image space parameters under shallow DoF zone constraints. On the basis of this, a more accurate geometric depth 3D image synthesis from a single image has arrived as three steps: (i) Generation of 3D perception image from MPIs using its inter-depths computed from inter-image region boundaries disparities.

(ii) The single 2D image is decomposed into one in-focus region and many out-of-focused regions caused due to varying CoCs using 2.1D sketch as a semantic image segmentation, and (iii) 3D image is composed by re-aligning the image surface and corresponding MPIs with accurate respective inter-depths.

The partial ordering of image regions using the 2.1D sketch enables us to determine the occlusions among the multiple objects in the scene better than the alpha-matting. Experiments show that the proposed method gives better results not only when the scene comprises multiple objects lying at different depths with dissimilar colors but also objects lying with the same color. Further, the proposed method has been demonstrated to yield smaller inter-depths of the order of few millimeters (range 5mm to 105mm gap between the objects in the 3D scene), making it quite qualified for real-time applications.

Few future scopes for research directions are enlisted as below:

- 1) Improving the accuracy of depth estimation for the objects lying on the off-axis with respect to the camera's optical axis.
- 2) Definitely worth making comparative studies between the suggested approach with plenoptic imaging to understand the importance of optical lens arrays.
- 3) The extension of the above-discussed methods for microscopic 3D image analysis could be interesting for many machine inspection and bio-imaging applications.
- 4) Exploration on deep layer network for robust and more precise 3D image generation using a single image to arrive at improved performance over the suggested approach in this article.
- 5) Exploration on the computation of 3D virtual and 3D perception images using front DoF and back DoFo zone is a worthwhile exercise to see the performance regarding the accuracy in inter-depths.

REFERENCES

- [1] U. Teubner and H. J. Bruckner, *Optical Imaging and Photography*. Berlin, Germany: De Gruyter, 2019.
- [2] F. Daniel, S. Yoni, K. Eyal, L. Ido, and S. Mirko, "SRA: Fast removal of general multipath for ToF sensors," in *Computer Vision—ECCV 2014*. Ithaca, NY, USA: Cornell Univ., 2014, pp. 234–249.
- [3] J. T. Todd, "The visual perception of 3D shape," *Trends Cogn. Sci.*, vol. 8, no. 3, pp. 115–121, Mar. 2004.
- [4] O. Wasenmüller and D. Stricker, "Comparison of kinect V1 and V2 depth images in terms of accuracy and precision," in *Proc. Asian Conf. Comput. Vis. Workshop (ACCV Workshop)*, 2017, pp. 35–45.
- [5] G. K. M. Cheung, S. Baker, and T. Kanade, "Visual hull alignment and refinement across time: A 3D reconstruction algorithm combining shape-from-silhouette with stereo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, pp. 375–382.
- [6] R. Lukac, *Computational Photography: Methods and Applications*, vol. 20. Boca Raton, FL, USA: CRC Press, 2010.
- [7] S. Kim, E. Lee, V. Maik, and J. Paik, "Real-time image restoration for digital multifocusing in a multiple color-filter aperture camera," *Opt. Eng.*, vol. 49, no. 4, Apr. 2010, Art. no. 040502.
- [8] Y. Bando, B.-Y. Chen, and T. Nishita, "Extracting depth and matte using a color-filtered aperture," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1–9, Dec. 2008.

- [9] V. Maik, D. Cho, J. Shin, and J. Paik, "Color shift model-based segmentation and fusion for digital autofocusing," *J. Imag. Sci. Technol.*, vol. 51, no. 4, pp. 368–379, 2007.
- [10] L. Galvis, E. Mojica, H. Arguello, and G. R. Arce, "Shifting colored coded aperture design for spectral imaging," *Appl. Opt.*, vol. 58, no. 7, p. B28, 2019.
- [11] M. U. K. Khan, A. Khan, J. Lim, S. Hamidov, W.-S. Choi, W. Yun, Y. Lee, Y.-G. Kim, H.-S. Park, and C.-M. Kyung, "Offset aperture: A passive single-lens camera for depth sensing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 5, pp. 1380–1393, May 2019.
- [12] H. J. Deeg and L. R. Doyle, "VAPHOT—A package for precision differential aperture photometry," in *Proc. 3rd Workshop Photometry*, Feb. 2001, p. 85.
- [13] R. Szeliski, *Computer Vision: Algorithms and Applications*. London, U.K.: Springer-Verlag, 2021.
- [14] R. R. Deshpande, C. R. Madhavi, and M. R. Bhatt, "3D image generation from single image using color filtered aperture and 2.1D sketch—A computational 3D imaging system and qualitative analysis," *IEEE Access*, vol. 9, pp. 93580–93592, 2021.
- [15] M. Nitzberg and D. Mumford, "The 2.1-D sketch," in *Proc. 3rd Int. Conf. Comput. Vis.* Washington, DC, USA: IEEE Computer Society, Dec. 1990, pp. 138–144.
- [16] I. Panchenko, V. Paramonov, and V. Bucha, "Depth estimation algorithm for color coded aperture camera," *Electron. Imag.*, vol. 2016, no. 21, pp. 1–6, Feb. 2016.
- [17] R. Accorsi, "Design of a near-field coded aperture cameras for high-resolution medical and industrial gamma-ray imaging," M.S. thesis, Dept. Nucl. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2001.
- [18] V. Paramonov, I. Panchenko, V. Bucha, A. Drogolyub, and S. Zagoruyko, "Depth camera based on color-coded aperture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 910–918.
- [19] S. Kim, E. Lee, M. H. Hayes, and J. Paik, "Multifocusing and depth estimation using a color shift model-based computational camera," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4152–4168, Sep. 2012.
- [20] J. Shade, S. Gortler, L.-W. He, and R. Szeliski, "Layered depth images," in *Proc. 25th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 1998, pp. 231–242.
- [21] V. Jantet, C. Guillemot, and L. Morin, "Object-based layered depth images for improved virtual view synthesis in rate-constrained context," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 125–128.
- [22] R. Tucker and N. Snavely, "Single-view view synthesis with multiplane images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 548–557.
- [23] P. P. Srinivasan, R. Tucker, J. T. Barron, R. Ramamoorthi, R. Ng, and N. Snavely, "Pushing the boundaries of view extrapolation with multiplane images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 175–184.
- [24] Y. Yang, S. Hallman, D. Ramanan, and C. C. Fowlkes, "Layered object models for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1731–1743, Sep. 2012.
- [25] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 625–638, Sep. 1994.
- [26] T. Zhou, R. Tucker, J. Flynn, G. Fyffe, and N. Snavely, "Stereo magnification: Learning view synthesis using multiplane images," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–12, Aug. 2018.
- [27] M.-L. Shih, S.-Y. Su, J. Kopf, and J.-B. Huang, "3D photography using context-aware layered depth inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8025–8035.
- [28] Y. Wan, J. Wang, J. Hu, T. Song, Y. Bai, and Z. Ji, "A study in 3D-reconstruction using kinect sensor," in *Proc. 8th Int. Conf. Wireless Commun., Netw. Mobile Comput. (WiCOM)*, Sep. 2012, pp. 1–7.
- [29] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. New York, NY, USA: Henry Holt, 1982.
- [30] R. R. Deshpande and M. R. Bhatt, "Extraction of 2.1D sketch from monocular images using DIRECT optimization algorithm," in *Proc. IEEE 7th Int. Adv. Comput. Conf. (IACC)*, Jan. 2017, pp. 645–650.
- [31] D. R. Jones, C. D. Pertunnen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *J. Optim. Theory Appl.*, vol. 79, no. 1, pp. 157–181, 1993.
- [32] M. Katz, *Introduction to Geometrical Optics*. Singapore: World Scientific, 2002.

[33] S. David and W. Michael, "A layered depth-of-field method for solving partial occlusion," *J. WSCG*, vol. 20, no. 3, pp. 1–8, 2012.

[34] J. N. P. Martel, L. K. Muller, S. J. Carey, J. Muller, Y. Sandamirskaya, and P. Dudek, "Real-time depth from focus on a programmable focal plane processor," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 3, pp. 925–934, Mar. 2018.



for four years with T. John Institute of Technology, Bengaluru, and five years with the PES School of Engineering, Bengaluru. Her research interests include color filters, computational cameras, and 3D imaging.



MAHABALESWARA RAM BHATT received the B.E. degree in electronics and communications engineering, in 1982, the M.Tech. degree in industrial electronics engineering from the Mysore University, in 1986, and the Ph.D. degree from the Department of Electrical Engineering, Indian Institute of Technology, Bombay, in 1992. From 1989 to 1994, he was as a Senior Research Engineer with IIT Bombay. From 1994 to 1996, he was a Scientist E-2 with the Centre for Development of Imaging Technology, Thiruvananthapuram. He served for several multi-national organizations as a research and development personnel. From 1996 to 1997, he was a Project Manager at LG Soft Development Center, Bengaluru; from 2000 to 2004, he was a Senior Software Manager at Analog Devices, Bengaluru; from 2005 to 2006, he was a Senior Manager at C2silicon, Bengaluru (a design center for IPFLEX, Tokyo, Japan); from 2006 to 2010, he was a Principal Member Technical Staff at AMD Hyderabad Design Center; and from 2010 to 2018, he was a Senior IP Development Consultant. He also served in various academic institutions as a full-time Professor at VIT, currently known as VIT University, Vellore, from 1996 to 1997, then at the Department of Electronics and Communication Engineering, T. John Institute of Technology, Bengaluru, from 2010 to 2013 and from 2013 to January 2019 as a full-time Professor with the Department of Medical Electronics, BMS College of Engineering, Bengaluru. He is currently serving as a Senior Consultant for Renalyx Health Systems Private Ltd., and Rx Digi Health Platform Private Ltd., Bengaluru. He is the author of more than 30 research publications and holds five U.S. patents and 12 patents are in various stages of patent publications and grants in India, WIPO, and USA.



C. H. RENU MADHAVI (Senior Member, IEEE) received the M.Tech. degree in electronic instrumentation from NIT Warangal and the Ph.D. degree in heart rate variability from Avinashilingam University for Women. She is currently working as an Associate Professor and the Head of the Department of Electronics and Instrumentation Engineering, R. V. College of Engineering, Bengaluru. She is also guiding Ph.D. scholars. She has 30 years of teaching experience. She has around 25 research publications in various national and international journals. She has Springer book chapter publications. Her research interests include biomedical signals, image processing, and the IoT. She was a reviewer of few international conferences and journal conferences.

• • •