IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2021, accepted June 6, 2021, date of publication June 16, 2021, date of current version July 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089938

3D Image Generation From Single Image Using Color Filtered Aperture and 2.1D Sketch-A Computational 3D Imaging System and Qualitative Analysis

RASHMI R. DESHPANDE[1], CH. RENU MADHAVI[2], (Senior Member, IEEE), AND MAHABALESWARA RAM BHATT[3]

[1]Research Center of Electronics and Instrumentation Engineering, R V College of Engineering, Bengaluru 560059, India, affiliated to: Visvesvaraya Technological University, Belagavi 590018, India
[2]Department of Electronics and Instrumentation Engineering, R V College of Engineering, Bengaluru 560059, India
[3]Renalyx Health Systems Pvt., Ltd., and Rx Digi Health Platform Pvt., Ltd., Bengaluru 560004, India

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

ABSTRACT A new method for generating 3D images using both disparity parameters due to the color filtered aperture (CFA) and depth alignment of the multi-plane images for an acquired 2D image using a novel computational imaging system is presented. A CFA is created using add-on red, green, and blue color filters in front of the objective lens for a conventional 2D camera system. The multiple plane images are computed using various image regions identified by 2.1D sketch and the corresponding inter-region depth information with the aid of color disparities caused in the acquired 2D image. Our emphasis is to handle possible occlusions in single 2D image regions due to various object features for the acquired scene using ordered multiple plane images and to generate a 3D image by exploiting the order existing in the computed multiple plane images and compositing the region with aligned inter-image region depth information. The simulation and analysis demonstrate the elegance of the 3D image generation. Ultimately, the proposed computational imaging system is not only cost-effective but also an alternative to the stereo-based imaging system, which is not suited for ophthalmic or endoscopic imaging applications.

INDEX TERMS 2.1D sketch, color filter, color filtered aperture(CFA), depth, disparity, image alpha matte, multi-plane image, phase correlation, semantic segmentation.

I. INTRODUCTION

The necessity of 3D image generation and its analysis plays a major role in the field of computer vision and many imaging based applications such as quick medical diagnosis, robot navigation, perception of a scene, industrial automation, surveillance, object recognition, animations and 3D modeling, and virtual/mixed reality in health diagnostic and surgical studies. A large amount of literature exists for depth estimation and 3D image generation using more than two images and only some of this is based on a single image. Primarily, the methods are either based on stereo images captured with a stereo camera or multiple images captured at different positions, and these are costly in terms of optical components, storage, and computational complexities.

Evidently, any scene is 3D in nature, where the location of each point is represented by three spatial co-ordinates spread across the length, breadth, and depth(height) and also represents a sequence of images either as motion, as time instants, or hyper-spectral, as spectrally as wavelengths. However, visible light intensity cameras can only capture a 2D image where each pixel point is projected as intensity values that lie spatially across the length and breadth. Conventionally, depth estimation relies on multiple images, either binocular, where depth recovery is based on simultaneous two views, or monocular single view cues (shading, shape, texture, motion, colors, etc.) are known to be sufficient [1]. The various other methods suggested are structure from motion [2] to arrive at both 3D motion and depth, depth from

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

focus (DFF) [3] for getting all-in-focus image by choosing sharp image out of multiple images obtained by differing focus setting and depth from defocus (DFD) [4] based on two images with differed focus setting used to find amount of blur to obtain scene depth with appropriate camera parameters. Specifically, in some contexts, it is not at all possible to acquire stereo or multiple posed images, such as retinal images and endoscopic images that involve imaging of internal organs of the human body. 3D scene generation becomes a challenging task not only due to the depth loss but also due to missing information that is invisible to the camera. In this article, we focus on 3D image based on visible information for a single view camera. The latter has led to the study related to non-line of sight images, which is not the scope of our research. The depth being inevitable for 3D images and depth recovery from a single 2D image is not only computationally intensive but also sensitive to small errors. Depth estimation is obtained using color disparities that would occur between RG and BG monochromatic images. 3D images are computed by fusing the depth from disparities obtained by the color filtered aperture (CFA), and multiple layer information obtained from the 2.1D sketch. The main aim of the proposed approach is to handle both the occlusion and the foreshortened color regions caused in the acquired 2D image where the objects have an inclined orientation from the camera's view while capturing a 2D image.

The contributions of this research article are summarized as follows:

3D image generation from a single image is viewed as a composition of multi-plane images from multi-region boundaries, inter-region depth, and texture mapping from original image belonging to respective boundaries.

This is achieved by:

- Devising a deterministic DIRECT global optimization algorithm to decompose into multi-region boundaries from semantic image segmentation using a 2.1D sketch based on the number of in-focus and out-of-focus regions in accordance with the objects lying in the DoF region, while acquiring a 2D image from CFA,
- Computing inter-plane depths from the respective inter-region disparities caused due to CFA from newly deduced relationship explicitly from camera space parameters and
- Image texture of each plane is obtained by adding the respective inter-image plane depth to the original 2D image belonging to pixels inside the corresponding boundaries.

In the course of achieving the above, we deduced the inter-region depth using image space parameters of the camera and formulated energy function as object level inference for depth ordering instead of pixel level for handling occlusions. Then, we propose to obtain the 2.1D sketch as a global minimum solution using a deterministic global optimal method. Section 2 describes reviews on related research efforts on depth estimation using CFA. Section 3 lists the objective of the proposed occlusion handling while generating a 3D image from a 2D image. Section 4 presents an optical setup for CFA with an off-the-shelf camera. Section 5 describes how multi-plane images are generated from both newly derived inter-depths from disparities and decomposed multi-regions. Section 6 describes a new 3D image generation algorithm in detail. Experimental results and discussion are presented in Section 7. Lastly, the concluding remarks and future scope for research are explained in Section 8.

FIGURE 1. Prototype arrangement of filters inside off-the-shelf camera lens.

II. RELATED RESEARCH EFFORTS

In this section, we restrict our review discussion and our scope of research exploration only to a single view image with a color coded aperture (CCA) as a passive depth sensor keeping in mind their less associated optical and electronic components requirement, although there is enormous research progress in binocular views, multi-orientation views, and active depth sensor integrated based approaches. Also, in the current study, we omit learning based approaches involving deep neural network as a possibility anticipated future research extension to improve the performances further on similar to other classical approaches.

We consider several existing methods of 3D image generation from a single image acquired using CFA and some relevant CCA camera based on disparities arise in their monocular images. This section also paraphrases the decomposition of the acquired single 2D image into multi plane images (MPIs) using multi-region segmentation obtained from the 2.1D sketch. Both depth from disparity and MPIs lay a foundation for our algorithmic approach for addressing occlusions that occur due to the overlapping of more than one object.

A. DEPTH ESTIMATION USING VARIOUS CFA

This subsection focuses on reviewing relevant research articles pertaining to CFA as a passive depth estimator and its outcomes. Although a wide-aperture lens would provide more amount of light to improve the signal-to-noise ratio of the acquired image, it provides only a shallow depth for a focused image. Yosuke *et al.* [5] utilizes CFA with primary colors, viz., red, green, and blue filters in a spatial arrangement that enables to create a single image with varying amounts of defocus in accordance with distances that produces spatial disparities, as shown in Fig.2(a). Each filter creates a color shift in the opposite and inverted direction with reference to the optical axis onto the imaging plane, as shown in Fig.2(b).

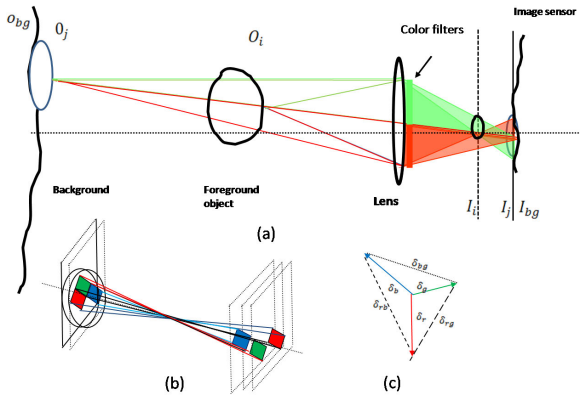

FIGURE 2. Optical Layout of CFA (a) the R band and G band light rays from CFA for the focused object reach the image sensor converging at the same point. For the non-focused objects the rays reach the sensor at different points which results in color shift. (b) 2D representation of the interactions between light rays from a CFA and its convergence to a point on the image sensor plane. (c) Disparities across R, G and B axis.

Although this approach exploits disparities between the red, green, and blue monochromatic images from the CFA RGB image, it would fail not only in identifying all the regions corresponding to the objects which are of the same color occluded by each other and are inclined to the optical axis. These failures may be due to the usage of tri-map oriented matting algorithm that segment the image into ''strictly foreground,'' ''strictly background,'' and ''unknown'' regions as instance image segmentation process, whose inferences are based on pixel-wise instead of object-wise. Moreover, the authors themselves remarked that their algorithm performs well when foreground and background objects are at 0.5 meters to 2.5 meters and about twice farther away from the camera. For farther scenes, the resolution of depth will decrease gradually, and the matte quality will also deteriorate as color misalignment will be small.

Maik *et al.* [6], adopted the color shift from multiple color filter aperture (MCA) to change the focal point of the given image at different locations to get the best focus and then used soft decision fusion and blending (SDFB) to obtain fully-focused images. Although SDFB smoothens the boundary transitions across region boundaries, it may not guarantee to handle occlusions similar to [5], since adaptive sum modified Laplacian (SML) based image segmentation method, is again dependent on pixel-wise inferences.

Lee *et al.* [7] formulated image enhancement to perform multi-focusing using MCA as a computational camera technique by estimating the depths of objects at different distances, image registration using phase correlation matching, and fusing apart from adaptive noise smoothing. Later, Sangjin Kim *et al.* [8] extended the same with real-time implementation for improving the quality of images acquired by the MCA system. In both the methods, in spite of adapting a configuration for handling objects at different distances by its color misalignment for both additional depth for objects at different distances and blur estimation, it would not assure to handle inclined objects and occlusions of the object with the same and different colors. This might be due to the cluster-based image segmentation and alpha map, which are instance image segmentation techniques.

Sangjin Kim, *et al.* [9] proposed an improved approach with MCA by considering not only spectrally varying misalignment due to colors but also spatially varying misalignment due to the distance of the objects from the plane of focus using the camera. The author uses cluster segmentation for color channels, which may not assure to distinguish different objects with the same color. Although this approach is shown to perform better for those objects with large inter-distances from the camera, say 5.8m and 21.4m again, this has not considered the image scene having occlusion with same color objects, inclined objects, and with lesser inter-objects distance.

Jaehyun *et al.* [10] proposed a real time single camera-based algorithm to detect both depth and track objects using color shift caused due to MCA and color probability distribution determined by the continuous adaptive mean shift algorithm. [11], [12] produces distortion at the boundary of the out-focused objects and may fail in tracking if multiple occluded objects have similar colors. Further, the accuracy of depth estimation is about 97.5 %, and the error range is $\pm$ 2m, which implies this would not assure objects overlapped with smaller distances.

Paramonov *et al.* [13] presented a CCA specially designed to provide high light efficiency, which could be integrated into a small camera of hand-held devices or a DSLR camera. Interestingly, it would provide depth in millimeters in real time for the whole frame, in contrast with all the above discussed approaches, demonstrated to valid for the center of the image frame and image restoration by applying color shifts using local disparity.

Panchenko *et al.* [14] presented an algorithm that yields continuous robust depth with respect to depth determination in a pixel, depth map sub-pixel estimation, depth propagation to low-textured areas, depth map edge restoration, depth quality enhancement, and raw data processing. This approach is computationally intensive and also the real-time performance has been achieved for FullHD frame for 50ms on GeForce GTX 780 and requires 15 seconds on Qualcomm Andreno 330 GPGPU. The algorithms described in [13], [14], and [15] focus on the performance of depth and disparity map for various specially designed CCA and possible light efficiency without any occlusion handling issues.

B. QUALITY OF DEPTH EXTRACTION AND 3D IMAGE SYNTHESIS

Even though the detection of depth from a single image is one of the essential necessary steps, the 3D image synthesis is dependent on the performance quality of the edge extraction process, non-missing of identification of objects in the scene while capturing a 2D image and its 3D scene interpretation. In this regard, we review the algorithm used in the above explained passive CFA-based depth sensing.

In [5], the scene depth map and alpha matte of the in-focus region corresponding to the foreground object are extracted automatically by dividing into three regions where either red, blue, or green rays can only pass to yield shifted views of scene distinctly. The main disadvantage in using alpha matte is that although it gives plausible results for in-focus regions alone, it fails to distinguish both overlapping and number of objects, mostly because matte is an instance image segmentation dependent only on pixels neighborhood. The synthetic refocus is performed after blurring the foreground color and in-focus background color differently.

A modified SML based algorithm used being again an instance image segmentation [16] may not ensure handling occlusions due to multiple objects with the same and different colors although it tackles objects at different heights in the scene. The all-in-focus scene synthesis [6] is done using a restoration technique similar to, would result in re-blurring and ringing artifacts in the focus region at a lower depth of field. In SDFB, alleviated the ringing artifact in focus region by optimal PSF estimation using phase correlation algorithm in image restoration in contrast to the method suggested in [16], but it may fail to handle occlusion problems. Mostly, this is due to the adaptive SML algorithm exploiting pixel-wise local inference only.

Note here that the method suggested by Lee *et al.* [7] has the same disadvantages that exist in [5] since it exploits alpha matte. Kim *et al.* [9] had a deficiency in handling occlusion due to same color objects because it exploits colors in cluster segmentation. The method used in Jaehyun *et al.* [10] may fail as objects overlapping with smaller depths and occlusion caused due to same color object by continuous adaptive mean shift algorithm is an instance image segmentation technique. The method proposed by Paramonov *et al.* [13] yields focused image synthesis only in the mid-region of the image. The approaches suggested in [13], [14], and [15] are not shown to handle any object occlusion problems.

On analyzing all the above, the algorithms described are based on instance image segmentation and some on image restoration seeming to fail to handle both occlusion due to objects at different height in the view directions, inclined objects, and objects with the same and different colors.

In this exploration, we adapt a semantic image segmentation method that is based on object-level inference instead of very local pixel-level segmentation.

C. GEOMETRIC 3D IMAGE REPRESENTATION

In this subsection, we review the existing geometry-based 3D image representation, which aids in our 3D image generation. Considering MPI representation for 3D images since it has attractive features because it not only handles multiple objects but also aids in boundary pixels on reflective-transparent objects with ''softness'', that enables the arrangement of image planes in terms of the layer depth image (LDI) as depth maps and associated color values [17], which facilitates to ''see around'' the foreground geometry to the occluded objects that lie behind. Zitnick *et al.* [18], represents scenes per-input-image depth maps, but also utilizes alpha matted layers around depth discontinuities to achieve high-quality interpolation. It is also similar to the classic layered representation for encoding moving image sequences by John *et al.* [19] and similar to the physical form described by [20].

III. OBJECTIVES OF OCCLUSION HANDLING WHILE GENERATING A 3D IMAGE FROM A 2D IMAGE

Usually, in many computer vision applications, the features existing in the acquired image need to be analyzed, especially in biomedical and robotic environments comprising of one or more scenarios, namely:

1) An image scene is a composition of several regions and objects, both of which exist in the scene at various depths;
2) It consists of several occluded image parts corresponding to the part of object or organ that is overlapped by one or more other objects;
3) The colors of the regions pertaining to the images corresponding to the two objects/organs in a scene could be the same;
4) There could be an object with a single color which is either occluded by other parts/organs or/and occludes some other parts/organs of different regions in the given scene; and
5) There could be regions in the image that would have been caused due to objects that are obliquely oriented or inclined in the camera view while capturing.

In view of the above difficulties, depth estimation definitely becomes a challenging task in terms of mathematical rigor, computational complexities, and hardware requirements such as processing speed and memory. Now, it can be recalled here that though the 3D image generation using only color disparity information suffices it may not ensure good results.

As mentioned in the preceding subsection, some of the proposed approaches distinguish between foreground and background using image matte computations, which is computationally intensive and require huge memory, and still would not ensure good occlusion handling capability. The emphasis in this article is not only to estimate the depth with good computation and quality performance but also to attempt to propose a more rigorous algorithmic framework to resolve the occlusion instead of a simple decision based on thresholding [10]. In this proposed approach, we formulate the 3D image generation problem as a sequence of the following salient processes given below:

1) Several objects are placed at known distances from the CFA camera and the RGB image is captured with one of the objects focused properly.
2) Disparity is estimated between RG and BG images for each of the salient pixels of the CFA image region boundaries, using phase correlation algorithm [21].
3) 2.1D sketch-based decomposition of the given image into multiple plane image layers using properties

related to regions/organs such as colors, texture, T-junctions, boundaries, and salient regions.

4) 3D image generation is carried by fusing depth estimated from color disparities and ordering arrangement of multiple plane images from 2.1D sketch to distinguish the overlapped regions.

IV. OPTICAL SETUP OF CFA FOR IMAGE CAPTURE

In this section, an optical setup required with additional modification for an off-the-shelf digital camera by considering CANON 50mm f1.8 II lens used to obtain the example results is described. A black opaque disc with three square-shaped holes is stuck with KODAK Wratten Red (No.47), Green (No.58), and Blue (No.25) optical filters and fixed just in front of the aperture diaphragms as shown in Fig.1. This modified aperture lens is attached to a CANON EOS 70D DSLR camera to capture the images. The effect of using these filters results in the color-shifting between red, green, and blue image planes. Typically, the image patch/region pertaining to the focused object has zero shift and those image patches/regions behind or in front of the focused object would undergo color shifts leading to disparities (depending on the distance from the focused object) in imaging planes. From Fig.2 we note that a right shift is observed in the red plane, an up shift in the green plane, and left shift in the blue plane [5] for the shown color filter arrangement. But, from Fig. 3, we note that the far point on the object plane is projected as a focused point near the lens in image space. Contrastingly, the near object plane is projected as far focused in the image plane.

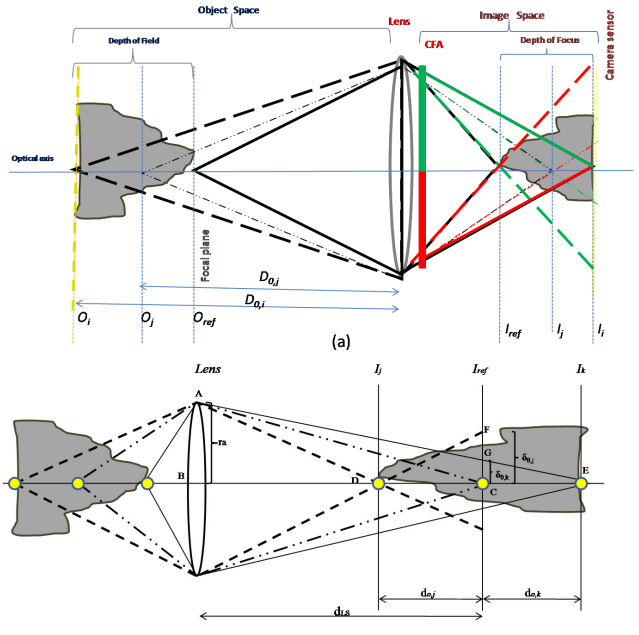

FIGURE 3. (a) Imaging geometry for DoF and DoFo for 2D CFA image formation depicting disparities on virtual image plane and real sensor plane for both near and far focus cases. (b) Ray construction for DoF and DoFo with virtual image planes and real sensor plane.

V. MULTI-PLANE IMAGES FROM MULTI-REGION IMAGE

A. DECOMPOSITION OF A SINGLE 2D IMAGE INTO MULTIPLE PLANE IMAGES USING 2.1D SKETCH BY DIRECT GLOBAL OPTIMIZATION METHOD

Essentially, any given single 2D image whose 3D image needs to be generated is decomposed into images of multiple planes on the basis of distinguishing the several distinct image regions/patches corresponding to various objects with either the same or differing colors having depths. Also, the decomposition process requires handling occlusions that would exist in the 2D image due to the overlapping of objects/organs. We employ 2.1D sketch extraction for the given 2D image, which is nothing but a process of cartooning the given 2D image with distinct regions into a set of multiple level sub-images (at either specific gray level and/or colors) in ordered fashion [22]. In other words, it is a process of slicing a given single 2D image into a set of multiple level image planes in an ordered fashion with each plane composed of only similarly clustered gray levels and/or colors apart from distinguishing the overlaps or occlusions.

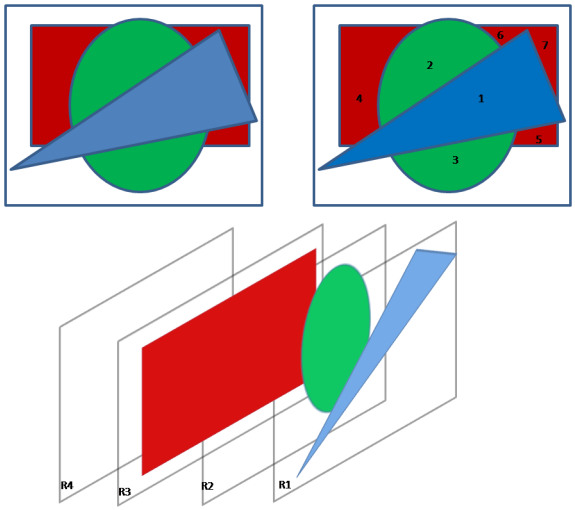

FIGURE 4. Typical 2.1D sketch. *top panel left*: Geometric shapes, *top plane right*: Fragmented boundary regions, *bottom panel*: Equivalent multi-layer image planes.

An example of cartoon image layers shown in top plane right of Fig. 4, where some of the regions are occluded by the others with partially visible image layers. We employ a 2.1D sketch as a method for determining the multiple plane images for any given scene image. A 2.1D sketch is nothing but cartooning the given natural image scene that decomposes into various regions corresponding to the various objects in the scene. Crucially, a 2.1D sketch gives not only a process of splitting the image into many regions of the given 2D scene image but also yields background and foreground regions ordering based on overlaps. For the purpose of explaining the 2.1D sketch, an example image comprising of only three objects, namely, square, circle, and triangle shown with three

different colors top plane left of Fig. 4 is considered. If one accounts only for the boundaries of the above three objects, this image consists of seven regions, as described top plane right of Fig. 4. But, the 2.1D sketch yields four sub-image planes as shown in bottom plane of Fig. 4 that correspond to three objects and a background. Conceptually, this method extracts the image regions with the help of T-junctions, edges, curves, cusps, crack tips, etc. [22]. Initially, among regional boundaries of 1, 2, and 3, the method identifies regions 2 and 3 as belonging to the same layer denoted as R_2, and region 1 as R_1. Consequently, a similar procedure is repeated among regional boundaries 2, 4, 5, 6, and 7, until all possible combinations are exhausted. Ultimately, these seven region boundaries are caused by four objects leading to four image planes. This could be generalized to L number of regions, R_1, $R_2, R_3, \cdots, R_L$, as sequences with layers on top of one another extracted with the help of T-junctions, edges, curves, cusps, crack tips, etc. [22]. Easily, we note the above procedure of L number of regions is combinatorial in nature. The energy minimization function involving T-junctions, boundaries, and saliency being the image cues to decide figure-ground (FG) regions and color and texture image cues to decide same-layer (SL) regions [23] is formulated and DIRECT optimization technique [24] is adapted for optimizing the same. The multiple plane images for the given 2D RGB image could be viewed as an optimal solution 2.1D sketch for an objective function $E_{2.1D}$ shown (1) with many adequate constraints [23], [24] as below:

$$\begin{aligned} E_{2.1D} &= E_{\lambda}\left(\alpha_i, \beta_i, \gamma_i, \zeta_i, \eta_i\right) \\ &= \left(\sum_{i=1}^{inum} \exp^{\lambda_t \alpha_i} + \exp^{\lambda_b \beta_i} + \exp^{\lambda_s \gamma_i}\right) \\ &\quad + \left(\sum_{i=1}^{inum} \exp^{\lambda_c \zeta_i} + \exp^{\lambda_{tex} \eta_i}\right). \end{aligned} \tag{1}$$

where the weighting parameter set, $\lambda_t, \lambda_b, \lambda_s, \lambda_c, \lambda_{tex}$, greater than 0, are the i^{th} features of α_i (T-junctions), β_i (boundaries), γ_i (saliency), ζ_i (colors) and η_i (textures) details, respectively. Further, *inum* denotes total number of features.

The above objective function shown in (1) is not only non-convex in nature but also involves a huge number of variables. Here, we advocate using DIRECT optimization algorithm [24] that is known to be deterministic in nature to get a global minimum (our preliminary results have been published in [25]) and here we extend the same for 3D image generation, which is the main intent of this article. Additionally, we recall from [23] that this function isolates the layers of objects in the image as FG that relates two adjacent over-segmented regions based on $T-$junctions, boundaries, and salient features, are not at the same depth. The decision on over-segmented regions in an image is taken based on color and texture.

The above obtained L number of segmented regions is re-organized as L number of multiple plane images by adding the respective inter-image depth computed as described at the beginning of this section. The details of 3D image generation are explained in the next section.

B. DISPARITIES ESTIMATION ALGORITHM USING 2D RGB IMAGE FROM CFA

A disparity determination using acquired 2D image from a CFA camera is proposed in the following steps:

1) Edge detection is carried out for red, green, and blue channels separately using a canny edge detector,
2) Because the strength of the green channel edge is less, scale the pixel values magnifying by some large value, say 10,
3) Form sub-images i_r, i_g, and i_b from acquired 2D image of block size 128×128 for all red, green, and blue planes,
4) Obtain color shifts/disparities, one between red and green planes, and the other between blue and green planes by performing phase correlation on each sub-image, as mentioned below:
   a) Apply appropriate Hanning window separately for all rows and all columns of the blocks (for all red, green, and blue channels) to get window operated outputs, Wi_r, Wi_g and Wi_b, respectively.
   b) Apply Fourier transformation on red, green, and blue images, for Hanning windowed output,

$$I_r = \mathscr{F}\left(Wi_r\right). \tag{2}$$
$$I_g = \mathscr{F}\left(Wi_g\right). \tag{3}$$
$$I_b = \mathscr{F}\left(Wi_b\right). \tag{4}$$

   c) Calculate the cross power spectrum $I_{cps}\,(rg)$ and $I_{cps}\,(bg)$ between RG planes and BG planes of the blocks as:

$$I_{cps}\,(rg) = \frac{I_r \circ I_g^*}{|I_r \circ I_g^*|}. \tag{5}$$
$$I_{cps}\,(bg) = \frac{I_b \circ I_g^*}{|I_b \circ I_g^*|}. \tag{6}$$

   $\circ$ and $*$ represents element-wise multiplications and complex conjugation, respectively.
   d) Calculate $i\,(rg)$ and $i\,(bg)$ using inverse Fourier Transform, as below:

$$i\,(rg) = \mathscr{F}^{-1}\left(I_{cps}\,(rg)\right). \tag{7}$$
$$i\,(bg) = \mathscr{F}^{-1}\left(I_{cps}\,(bg)\right). \tag{8}$$

   e) Determine shifts as below:
      i) The location of the maximum value from $i\,(rg)$ and $i\,(bg)$ for rows gives the horizontal shifts $S_H\,(rg)$ and $S_H\,(bg)$, respectively.
      ii) The location of the maximum value from $i\,(rg)$ and $i\,(bg)$ for columns gives vertical shifts $S_V\,(rg)$ and $S_V\,(bg)$, respectively.
   f) Calculate disparity $\delta\,(rg)$, $\delta\,(bg)$ and δ as

$$\delta_{(rg)} = \sqrt{S_H\,(rg)^2 + S_V\,(rg)^2}. \tag{9}$$

$$\delta_{(bg)} = \sqrt{S_H\,(bg)^2 + S_V\,(bg)^2}. \quad (10)$$

$$\delta = max\left(\delta_{(rg)}, \delta_{(bg)}\right). \quad (11)$$

C. INTER-PLANE IMAGES DEPTH FROM INTER-REGION IMAGE DISPARITY

We consider the geometric construction corresponding to the camera setup shown in Fig.3(a). The distances from the camera to any two specified objects O_i and O_j in a scene are at $D_{0,i}$ and $D_{0,j}$, respectively as the depth of field (DOF) in the object space side. The same are formed as regions I_i and I_j as the depth of focus (*DoFo*) in image space. These image regions are captured by the reference sensor plane at I_{ref} as a 2D image. Note here that I_i and I_j are extracted as two image segments belonging to two image planes as 2.1D sketch. It is also important to see these image segments that are not focused would have resulted in focused images if the image sensor plane is positioned at I_i and I_j.

The red, green, and blue filters produce shifts/disparities δ_r, δ_g and δ_b, respectively, in the captured 2D RGB image from CFA camera. This is depicted in the Fig.2(a) and Fig.2(b). The shift between red and green planes is denoted by δ_{rg} can be expressed as

$$\delta_{rg} = \left((\delta_r)^2 + \left(\delta_g\right)^2\right)^{\frac{1}{2}}. \quad (12)$$

Similarly, The shift between blue and green planes is denoted by δ_{bg} can be expressed as:

$$\delta_{bg} = \left((\delta_b)^2 + \left(\delta_g\right)^2\right)^{\frac{1}{2}}. \quad (13)$$

Finally, the overall shift $\delta_{0,i}$ is obtained using (11) as

$$\delta_{0,i} = \max\left(\delta_{rg}, \delta_{bg}\right). \quad (14)$$

For the sake of keeping the diagram simple, Fig.3(b) is displayed with one ray instead of three rays corresponding to red, green, and blue filters. The sensor captures the color filtered 2D RGB image and disparities in the regions are estimated using equations (9) through (11). Now, let r_a and d_{LS} denote radius of the aperture and inter-distance between aperture lens and image sensor plane I_{ref}, respectively. Let I_j and I_k be the image planes that would give focused images for the k^{th} and j^{th} regions. The sensor plane is represented as I_{ref}. Let the inter-distance between planes I_j and I_k with respect to sensor plane I_{ref} be $d_{0,j}$ and $d_{0,k}$, respectively. But, these region boundaries are projected on to the image sensor, referred here as I_{ref} creating the shifted but unfocused region boundaries. These non-focused region boundaries create disparities $\delta_{0,j}$ and $\delta_{0,k}$ for the j^{th} and k^{th} regions respectively. In Fig.3(b), the disparity distances CF and CG are shown as $\delta_{0,j}$ and $\delta_{0,k}$, respectively. Further, the depths CD and CE are represented as $d_{0,j}$ and $d_{0,k}$, respectively.

On noting similarity triangles ABE and GCE, and its ratio property, we express the following ratios

$$\frac{r_a}{\delta_{0,k}} = \frac{d_{LS} + d_{0,k}}{d_{0,k}}. \quad (15)$$

Similarly, on the basis of similarity triangles ABD and FDC we express another relation as:

$$\frac{r_a}{\delta_{0,j}} = \frac{d_{LS} - d_{0,j}}{d_{0,j}}. \quad (16)$$

In view of equations (15) and (16), the inter-depths $d_{0,j}$ and $d_{0,k}$ between image plane I_j and I_k with reference to the image sensor layer I_{ref} expressed as below:

$$d_{0,k} = \frac{\delta_{0,k} \times d_{LS}}{r_a - \delta_{0,k}}. \quad (17)$$

$$d_{0,j} = \frac{\delta_{0,j} \times d_{LS}}{r_a + \delta_{0,j}}. \quad (18)$$

D. GEOMETRICAL OPTICS PRINCIPLE BEHIND DEPTH ESTIMATION USING CFA

On noting $\delta_{0,i}$ and $\delta_{0,j}$ as inter-image region disparities either bottom layer or top layer of the object as focus point we make the following observations.

1) Inter-depths $d_{0,j}$ and $d_{0,k}$ between the image regions are related to aperture radius r_a and disparities $\delta_{0,j}$ and $\delta_{0,k}$ when d_{LS} is held constant.
2) Inter-depths are also proportional to d_{LS} when r_a is held constant, but this case is possible when the camera has adjustable sensor planes.
3) Inter-depths are almost proportional (as $\delta \ll r_a$) to disparities alone when both d_{LS} and r_a are held constant.

In this study, we make use of the latter observation in estimating the depth from a 2D image acquired using $d_{0,j}$ and $d_{0,k}$ for the optical setup comprising the camera and object shown in Fig.3(b). Although Fig.3(a) is shown for three regions, this is applicable for L number of regions if the captured 2D RGB image at the image sensor plane I_{ref}. It is the composition of a projected $(L-1)$ number of defocused image regions and one focused image region where each image region would have been focused one if they were captured at respective image sensor plane positions in the depth of focus region. The inter-image region's depth array could be viewed as follows.

For near focus:

$$d_{0,k} = \frac{\delta_{0,k} \times d_{LS}}{\left(r_a - \delta_{0,k}\right)}, \quad k = 1, 2, \cdots (L-1). \quad (19)$$

$$d_{L,k} = d_{LS} + d_{0,k}. \quad (20)$$

For far focus:

$$d_{0,j} = \frac{\delta_{0,j} \times d_{LS}}{\left(r_a + \delta_{0,j}\right)}, \quad j = 1, 2, \cdots (L-1). \quad (21)$$

$$d_{L,j} = d_{LS} - d_{0,j}. \quad (22)$$

where $d_{L,k}$ and $d_{L,j}$ represents distance from lens to k^{th} and j^{th} virtual planes, respectively. Now, 2D RGB image regions are seen as a 2.1D sketch having L number of segmented 2D images. It is important to note that inter-region depth is expressed only as image space parameters of the given camera. The principle behind the algorithm is described in section V(E). Equations (20) and (22) provide the distances of objects with reference to the lens used for scatter plots.

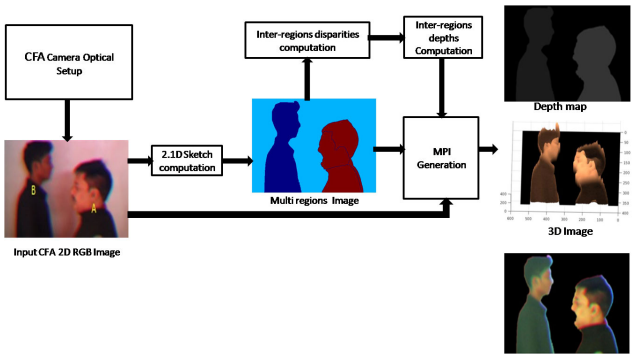

FIGURE 5. Overview of computational 3D imaging system.

E. PRINCIPLES OF 3D IMAGE GENERATION PROCESS FOR EXTRACTED MULTIPLE IMAGE PLANES

In this subsection, we explain method adopted to generate the 3D images as per any i^{th} and j^{th} image planes from the extracted 2.1D Sketch.

1) Consider i^{th} and j^{th} image plane from the extracted 2.1D sketch. In the current study i^{th} plane is set as focused plane.
2) Compute inter depth $d_{(i,j)}$ between above planes using equations (19) for near focus and (21) for far focus.
3) Extract corresponding pixel values from camera acquired 2D image patches corresponding to the segmented j^{th} image plane.
4) Add inter-depth value $d_{(i,j)}$ to every pixel in the above selected j^{th} image plane for each red, green, and blue channels of CFA 2D image.
5) Repeat steps 1-4 for all the non-focused regions.
6) Compose a 3D image $I_{3D}(r, g, b)$ as union of all the results generated for non-focused regions together with focused region.
7) Normalize all the pixel values in the above obtained $I_{3D}(r, g, b)$ such that its minimum value and maximum is same as original given single image.

VI. 3D IMAGE GENERATION ALGORITHM USING BOTH DISPARITIES AND 2.1D SKETCH FROM A 2D IMAGE

This section describes a new algorithm for 3D image generation for a given single image by making use of both disparities from CFA and multiple image planes from 2.1D sketch. The computational imaging system with a functional diagram is shown in Fig.5. Note here that the depth map I_{depth} obtained by the above CFA algorithm is of the same size as that of a 2D image. Also, the size of 2.1D sketch image $I_{2.1Dsketch}$ is the same as that of both original image and I_{depth}. Typically, these L layers are decided based on a number of sub-images corresponding to salient objects as depicted in top plane right of Fig. 4. Further, layers are designated as $0, 1, \cdots\cdots i, \cdots, j, \cdots, (L-2), (L-1)$, by noting the fact that the lesser gray level image value is relatively nearer to the camera when compared to the larger gray level image value, while designating the decomposed image planes as per 2.1D sketch. Moreover, the decomposed L number of layers could be viewed as either a number of different regions or objects. It is crucial to note that the inter-depth is estimated only with respect to boundaries with reference to images obtained by 2.1D sketch. To arrive at a 3D image for the corresponding acquired 2D image, we perform a fusion of (i) inter-depth estimated using the edge data from 2.1D sketch image (ii) the texture content inside the boundary of the region with the respective region's original acquired CFA image. We formulate a novel 3D image generation algorithm described below:

1) On noting L number of image planes pertaining to 2.1D sketch image $I_{2.1Dsketch}$ as

$$\begin{pmatrix} I_B = I_0(R_0), \\ I_{B+1}(R_1), \\ \vdots \\ I_{B+i}(R_i) \\ \vdots, \\ I_{B+j}, (R_j), \\ \vdots, \\ I_{(B-(L-2))}(R_{L-2},), \\ I_{(B-(L-1))}(R_{L-1}) = I_F \end{pmatrix}. \tag{23}$$

In the above array of image layers, B and F are designated as top and bottom-most layers, respectively.

2) Obtain 3D image patch for all designated Layers $l = 0, 1, 2, \cdots, (L-1)$ as per 2.1D sketch as below:
   - Using the 2.1D sketch layers, for each non-focused regions, obtain the disparities/shifts $\delta_{i,1}, \delta_{i,2}, \cdots, \delta_{i,l}$ as per section V(B). The disparities for non-focused regions vary according to the depth as depicted in Fig.3(b).
   - The inter-depth $d_{i,j}$ between each non-focused layer $I_{B+j}(R_j)$ and focused layer $I_{B+i}(R_i)$ is estimated using (19) for near focus objects and (21) for far focus objects. The inter-depths be denoted by: $d_{i,1}, d_{i,2}, \cdots, d_{i,j}, \cdots, d_{i,l}$ where l is the number of non-focused layers.
   - Extract the original R, G, B image patches for each of the non-focused layer $I_{2d_{B+l}}(R_l)$ and add the corresponding inter-distance depth value to each pixel. $I_{R_l}(c) = I_{2d_{B+l}}(R_l) + d_{i,l}$
3) Compose all the above R,G,B set of images to form $I_{3D}(r, g, b) := \bigcup_{l=1}^{L-1} I_{R_l}(r, g, b)$.
   Together with all the non-focused $I_{3D}(r, g, b)$, the original 2D R,G,B image patch for focused region is also included.
4) Normalize all the pixel values in the above obtained $I_{3D}(r, g, b)$ such that its minimum and the maximum value is the same as the original given single image.

It is important to recall the fact that though any human could visually observe the focused region or object, it is difficult for any computational machine-based algorithm to understand a specific focused region in the acquired 2D image. But,

in the next section, we consider a few examples of 3D image generation by considering a case where inter-depth between two regions with reference to a known focused region or image plane from the decomposed multiple plane images is realized, but in advance in some ways.

VII. EXPERIMENTAL EVALUATIONS

This section describes the experimental setup, qualitative results to validate the above proposed 3D image generation using a single image with both depth estimation from CFA and multiple planes segregation from 2.1D sketch.

A. EXPERIMENTAL SETUP CONFIGURATIONS FOR CFA

The experimental setup shown in Fig.3(b) is employed with the specifications as radius of aperture being $r_a = 10mm$, where in three RGB filters arranged as per Fig.1 and distance between lens and sensor being $d_{LS} = 52.6316mm$. for canon 50mm f1.8 II when focused at 100cm, We note here that for a fixed focal length and aperture size, DoF will increase as the object-to-camera distance increases and vice versa. On the other hand, for a fixed object distance to the camera and aperture size, DoF decreases as the focal length increases and vice versa. Apart from the above, the inter-objects depths in the scene, the DoF range depends on image sensor parameters used in the specific camera, such as size, types, etc. Further, diffraction also contributes to the varying depths of field. It makes the function relationship more complicated in nature. We also keep the aperture size constant. Hence, we determine 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 purpose of demonstration. Based on this, we have obtained inter-object distance in the scene to be 10cm. This indicates that the specific camera selection is an important requirement for lesser inter-object distances.

B. QUALITATIVE ANALYSIS

This section highlights the outcomes of the above newly suggested 3D image generation algorithm. To validate the algorithm's performance, nine images using CFA camera have been acquired, out of which the first five images obtained with the arrangement of different objects in Fig.6 (a-e) are synthetic and the other four are natural images Fig.6 (f-i). Images in column (I) of Fig.6 are mentioned below.

1) Image shown in row(a) of Fig.6 is obtained using CFA camera by keeping the objects A (in dark brown color), B (in light brown color), and C (in dark brown) at the distances 60cm, 100cm, and 130cm respectively without overlapping between them. Object B is focused.
2) Two objects X and Y are placed as displayed in the image shown in row(b) of Fig.6. Object X marked with BA is inclined in the direction of the optical axis at 81cm to 179cm and overlapped by object Y which is focused and placed at 100cm marked as C.
3) The arrangement of objects X and Y in the image shown in row(c) of Fig.6 obtained using CFA camera is similar to the arrangement of second. Object Y is focused at 80cm. Object X is placed in a more inclined manner to the direction of the optical axis such that it extends from 72cm to 80cm in front of the focused object whereas the region of object X in the back of the focused object Y is placed horizontally with one end at 80cm and the other end at a distance of 132cm.
4) Image shown in row(d) of Fig.6 is obtained using a CFA camera by placing one object X (in brown color) such that one end is at 80cm and the other end at a distance of 109cm. Object Y (in black color) is focused at 80cm overlapping over the former. Object Y and one end of object X are both placed at the same distance.
5) Image shown in row(e) of Fig.6 is obtained using CFA camera by placing object X (in green color texture) such that one end is placed at 50cm and the other end at 83cm. The other object Y(in black color) focused at 50cm overlaps over the former object X.
6) Image shown in row(f) of Fig.6 shows a boy standing behind the stone, where the stone is focused and the boy is not focused.
7) Image shown in row(g) of Fig.6 is captured with CFA focused at 80cm. The distance between two persons is 50cm. Both are not focused.
8) Image shown in row(h) of Fig.6 is obtained, where two stones placed on the ground, with one being focused at 100cm from the camera and the other is not focused.
9) Image shown in row(i) of Fig.6 is a natural image in which the bark of a tree at the extreme left is focused and the rest of the image regions are not focused.

It is important to note here that all the distances are calculated with respect to the camera lens in the above-mentioned object arrangement. While capturing the images, the camera image space configurations are maintained as described in the experimental setup section VII(A). The above generated images are used as inputs to extract 2.1D sketches comprising of multiple segmented images as shown in Fig.6 column(II) as the output of the algorithm described in section V(A). It is to be noted that here the DIRECT optimization algorithm is the central computational part to get 2.1D sketches very accurately. Also, the same 2D RGB images obtained through CFA are utilized to compute the inter-depth between various regions of the image patches using the disparities estimation algorithm described in section V(B). The computed inter-depth based on disparity is used to arrive at a depth map between the corresponding inter-regions shown in 2.1D sketches. Now, column(III) of Fig.6 shows the scatter map for depths which could be viewed as multiple decomposed image layers with varying depths across 2D RGB images. The resultant depth maps are used to derive the depth images shown in column(IV) of Fig.6. The depth map images can be viewed as images comprising of varied low to high gray values for each region, as seen from the camera lens on the image side. The inter-depths are calculated with reference to an image sensor which is at 52.6316mm from the lens when focused at 100cm. All possible inter-depths $d_{i,j}s$ are

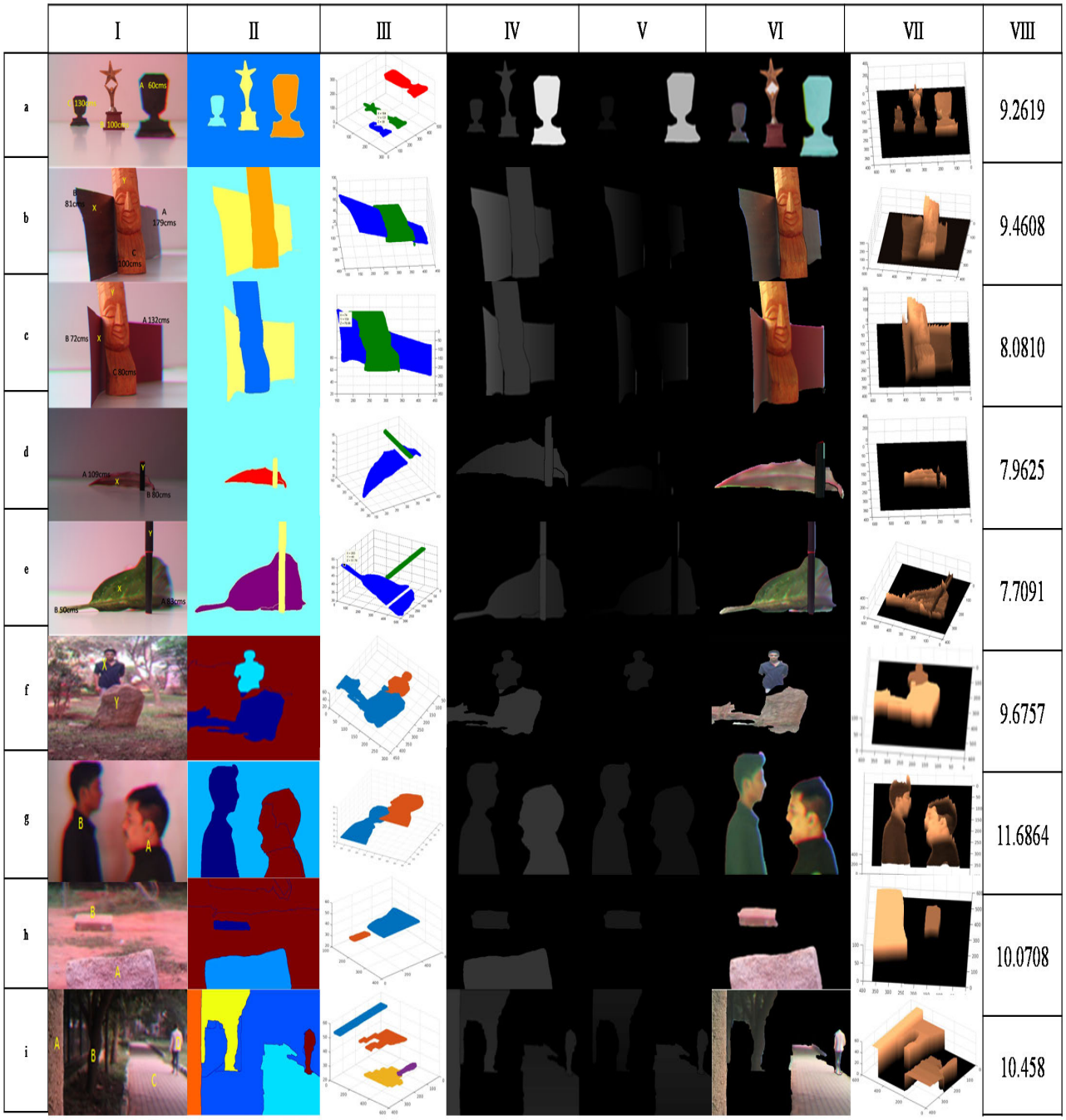

FIGURE 6. **Visual comparison of various salient scenarios are shown as rows and corresponding results obtained at different stages as columns: *col.(I)*: CFA captured 2D RGB images of salient scenarios, *col.(II)*: 2.1D sketch, *col.(III)*: Scatter map, *col.(IV)* Depth map, *col.(V)*: Depth map with respect to focused region, *col.(VI)*: Resolved images, *col.(VII)*: Generated 3D images from proposed approach, *col.(VIII)*: computational time elapsed in seconds for the proposed method, *row(a)*: image scene with three objects at different distances, *row(b)*: image scene with inclined objects with different colors with occlusions, *row(c)*: image scene with both inclined and horizontal objects with occlusion, *row(d)*: image scene with inclined leaf occluded by pen,*row(e)*: image scene with leaf occluded by pen,*row(f)*: image scene with person partially occluded by stone object, *row(g)*: image scene with two persons facing each other with different face color complexion, *row(h)*: image scene with two stones of different size without occlusion, *row(i)*: image scene with street, walking person and house without occlusion.**

computed as distances between image planes I_i and I_j, for all the j^{th} planes with respect to i^{th} plane as the reference plane (here the reference plane is the focused image plane). Column(V) of Fig.6 are depth maps of non-focused regions with respect to the focused region of the corresponding CFA images in column(I). The 3D resolved images are shown in column(VI) of Fig.6. Ultimately, 3D generated images are composed using both depth from disparity and multiple layer images as per the algorithm described in section VI, shown in column(VIII) of Fig.6. To demonstrate the performance of the

proposed algorithm, we consider the case where the top layer of the object is focused, which gives rise to a focus region in the acquired 2D image. It is crucial to compute inter-region depth while handling occlusion problems effectively. The estimated time in seconds for the proposed method is given in the last column(VIII). The decomposed multiple image plane to exhibit results are explained as below:

1) Column(II) of Fig.6 displays the output of the 2.1D sketch using the algorithm described in section V(A). It is crucial to note that 2.1D sketches with pseudo-colored painted regions are obtained accurately for all the images that correspond to different arrangements. This results in segmented image regions corresponding to the nature of objects in the scene under 2D image capture. In fact, 2.1D sketches help to extract the regions of each of the image patches that exist in the captured 2D image for 3D image generation. Further, the boundaries of 2.1D sketches aid in assigning the depth values between the inter-image patches.
2) Column(III) of Fig.6 depicts the 3D plots having different depths as scattered plots. Interestingly, these figures demonstrate the ability of the proposed algorithm to obtain proportional inter-depth in the respective image parts for all the images that are composed of overlapped and non-overlapped regions with inclined or horizontal position orientation.
3) Column(IV) of Fig.6 are depth maps for images of column(I). Here, higher gray level representation indicates more depth as compared to the lower gray levels.
4) Column(V) of Fig.6 are depth images represented as varying gray values with respect to the reference layer (focused layer).
5) 3D generated images obtained as per the algorithm proposed in section VI for all the cases shown in column(VI) of Fig.6.
6) Column(VII) of Fig.6 are 3D view representations corresponding to 3D generated images shown in column(VI) of Fig.6, display the quality of depth estimations.
7) On comparing 2D RGB images shown in column(I) with the corresponding 3D generated images column(VII), it is observed that the algorithm has displayed good 3D images with almost all regions resolved according to their depth. The comparisons between the approach similar to our approach is given in section VII(C).
8) The respective time in seconds taken to obtain the respective 3D images is displayed in column(VIII) of Fig.6. These results are based on the implementation using Intel® Core i5-4460 CPU @ 3.20GHz, Operating system: UBUNTU 16.04 LTS and 8.00GB memory, on MATLAB software platform.

On observation, the generated 3D image shown in column(VI) of Fig.6 corresponding to 2D RGB images rows(b-f) marked with sub-images as X occluded by region Y demonstrates good results with appropriate occlusion and overlap of regions.

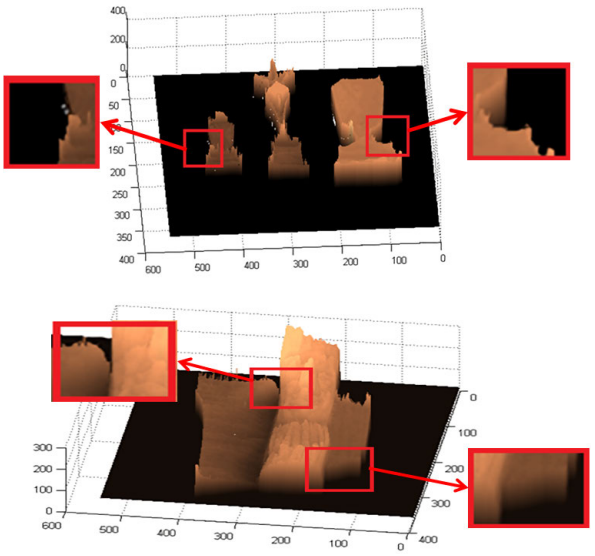

FIGURE 7. Generated 3D image from proposed approach with zoomed areas indicated by red markings to demonstrate the good of inter-depth extraction between the image regions, *top panel*: image without any occlusion, *bottom panel*: image with occluded image region.

C. 3D IMAGE GENERATIONS WITH OCCLUSION

In this subsection, we present a few points related to the occlusion handling capability of the proposed 3D image generation algorithm using the CFA 2D images. The images shown in Fig.6 for image rows(b-f) that have overlapping between the regions in various fashions are considered and observations are presented below:

1) On comparing the 2D RGB image shown in row(b) of Fig.6 and the corresponding 3D generated image shown in column(VI), it is observed that the proposed algorithm has been able to maintain both color differences and inclined objects for resolving all the occluded regions accurately according to their depths. Zoomed regions are shown in Fig.7.
2) On comparing 2D images shown in row(c) of Fig.6 with the corresponding 3D generated image shown in column(VI), it can be seen that the ability to generate 3D image yields good inclination for horizontal regions with respect to the optical axis with all regions resolved based on depth.
3) On comparing the 2D RGB image shown in row(d) of Fig.6 with the corresponding 3D generated image shown in column(VI), it is seen that the occluded leaf by pen is well resolved and seen bright in the generated 3D image.
4) On comparing the 2D CFA image shown in row(e) of Fig.6 with the corresponding 3D generated image shown in column(VI), it is noticed that the resultant 3D image especially, the blurred portion of the leaf (which is to the right of the pen region in row(e)) is well recognized.

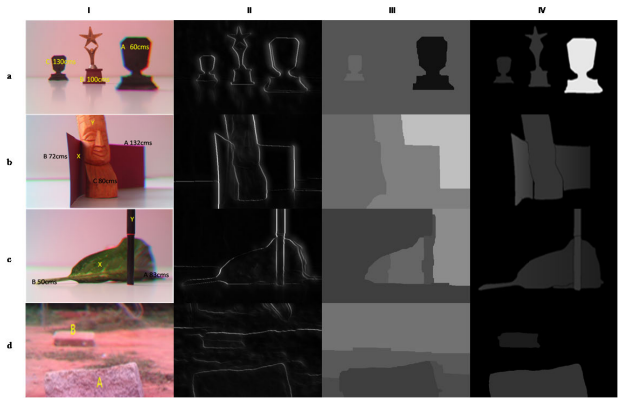

FIGURE 8. Comparison using ground truth. col. (I) CFA images col. (II) ground truth col. (III) Bando method col. (IV) our method.

5) On comparing the 2D image shown in row(f) of Fig.6 with the corresponding 3D generated image shown in column(VI), the boy occluded by stone is shown to be accurately resolved.
6) As expected, the proposed algorithm is able to provide a 3D image with inter-depths between regions by assuming 2D CFA RGB image regions, which is projected view along the given optical axis of the camera.

The comparative performances between Bando [5] method and our method for the examples with ground truth are listed below:

1) Results depicted in row(a) of Fig.8 show that the proposed method is able to segment all the three object scenes better than Yosuke *et al.* [5], where the latter fails to segment the focused object,
2) Results shown in rows(b) and (c) of Fig.8 show that our method is able to handle occlusion efficiently over Yosuke *et al.* [5],
3) The result shown in row(d) displays our method able to segment portion A and B, whereas it fails in the Yosuke *et al.* [5] method.

From the above observations, we note that multiple image region identification using 2.1D sketch is demonstrated to be better than Bando *et al.* which employs the image matting method [5]. We believe that our approach is able to handle occlusion because 2.1D sketch aids not only to segment the region but also enable depth ordering efficiently. This depth ordering in the 2.1D sketch is based on object level inferences in contrast with the depth estimation adopted in Bando *et al.* which is a pixel level inferences because image matting is pixel based operation.

In summary, we have presented and characterized a new computational imaging system that yields 3D images by utilizing the depth estimated based on the disparity that exists in the single CFA 2D RGB image and depth-ordering based on computed multiple image layers for the same CFA image. Furthermore, this would enable us to generate 3D images using a single image in applications where there is no better way to either obtain multi-view images or stereo images. Moreover, in endoscopes, ophthalmoscopes, etc, this computational imaging system would reduce not only the cost but also the space requirement where both are very crucial. Although the currently proposed 3D image generation is able to resolve occluding and occluded regions clearly, this is suitable for diagnostic studies of biomedical image. Additionally, we believe that it is not appropriate to compare any currently available deep learning based results mainly because our optical setup involves additional CFA instead of conventional image capture.

VIII. CONCLUSION AND FUTURE SCOPE

In this paper, we presented a new 3D image generation method using a single RGB image that exploits depth estimation using CFA and the decomposed multi-plane images from a CFA image using a 2.1D sketch. This involves depth estimation by computing disparities that exist between RG and BG planes using a phase correlation technique and the computation of the inter-depth between decomposed multiple images by employing the DIRECT optimization technique, which is known to be deterministic in nature. The results are shown to be good for cases, where occlusion regions in the acquired 2D image are caused not only due to overlaps of various regions in the scene but also have inclined orientation to the optical axis of the camera view occluded by other object in the scene. Further, a 2.1D sketch effectively aids in organizing various regions of the image in order, that helps to decide which regions are occluded or non-occluded during 3D image generation. Also, 3D image generation exploits inter-region depths with respect to any specific referenced layer. In the current study, the referenced layer is the focused region. It is easy to see that the computation of inter-region depth could be obtained with reference to either background or foreground image planes. The results shown in this paper are arrived at by considering inter-depth as per non-focused layers of a 2.1D sketch with respect to the focused layer. Ultimately, a 3D image is generated by modifying the interior of identified regions by adding the inter-depth value appropriately to the corresponding layer texture pixels for all R, G, and B planes of acquired 2D images separately and finally by combining all the three layers. A qualitative analysis has been demonstrated for various cases to show how it yields a good 3D image using a single RGB image acquired using an off-the-shelf camera and with less additional setup for a CFA. In the current study, although the qualitative analysis on the 3D image composition as a computational 3D imaging system has been discussed, we would be exploring the quantitative aspects as a future scope research study. Further, we believe that it is worth investigating the possible research outcome by implementing the suggested algorithm in terms of deep learning algorithms.

ACKNOWLEDGMENT

The authors would like to thank Dr. Ganesh Murthy C N S, Principal Engineer, Mercedes Benz Research and Development, Bengaluru, India, for his valuable and constructive suggestions during the development of this research work.

REFERENCES

[1] R. Szeliski, *Computer Vision Algorithms and Applications*. London, U.K.: Springer, 2011.
[2] T. Tan, K. Baker, and G. Sullivan, "3D structure and motion estimation from 2D image sequences," *Image Vis. Comput.*, vol. 11, no. 4, pp. 203–210, May 1993.
[3] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415–1432, May 2013.
[4] C. Zhou, S. Lin, and S. K. Nayar, "Coded aperture pairs for depth from defocus and defocus deblurring," *Int. J. Comput. Vis.*, vol. 93, no. 1, pp. 53–72, May 2011.
[5] 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.
[6] 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, Jul. 2007.
[7] E. Lee, W. Kang, S. Kim, and J. Paik, "Color shift model-based image enhancement for digital multifocusing based on a multiple color-filter aperture camera," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 317–323, May 2010.
[8] S. Kim, "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.
[9] 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. 4152–4166, Sep. 2012.
[10] I. Jaehyun, J. Jaehoon, and P. Joonki, "Single camera-based depth estimation and improved continuously adaptive mean shift algorithm for tracking occluded objects," *Pacific Rim Conf. Multimedia*, vol. 9315, pp. 246–252, Sep. 2015.
[11] Z. Li, J. Gao, Q. Tang, and N. Sang, "Improved mean shift algorithm for multiple occlusion target tracking," *Opt. Eng.*, vol. 47, no. 8, pp. 1–6, Aug. 2008.
[12] G. R. Bradski, "Real time face and object tracking as a component of a perceptual user interface," in *Proc. 4th IEEE Workshop Appl. Comput. Vis. (WACV)*, Oct. 1998, pp. 214–219, doi: 10.1109/ACV.1998.732882.
[13] 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.
[14] 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.
[15] V. P. Paramonov, *Depth Camera Based on Color-Coded Aperture*. Cham, Switzerland: Smart Algorithms for Multimedia and Imaging, May-2021.
[16] Stanciu, M. Dragulinescu, and G. Stanciu, "Sum-modified-Laplacian fusion methods experimented on image stacks of photonic quantum ring laser devices collected by confocal scanning laser microscopy," *UPB Sci. Bull. A, Appl. Math. Phys.*, vol. 73, no. 2, pp. 139–146, Jan. 2011.
[17] J. Shade, S. Gortler, L.-W. He, and R. Szeliski, "Layered depth images," in *Proc. 25th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, Jul. 1998, pp. 231–242.
[18] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600–608, Aug. 2004.
[19] 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.
[20] M. Holroyd, I. Baran, J. Lawrence, and W. Matusik, "Computing and fabricating multilayer models," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 1–8, Dec. 2011.
[21] M. Peeters, "Implementation of the phase correlation algorithm: Motion estimation in the frequency domain," Technische Universiteit Eindhoven, Eindhoven, The Netherlands, Tech. Rep. ICS-ES 820, 2003, vol. 7839.
[22] M. Nitzberg and D. Mumford, *The 2.1-D Sketch*. Piscataway, NJ, USA: IEEE Computer Society Press, Dec. 1990, pp. 138–144.
[23] C.-C. Yu, Y.-J. Liu, M. T. Wu, K.-Y. Li, and X. Fu, "A global energy optimization framework for 2.1D sketch extraction from monocular images," *Graph. Models*, vol. 76, no. 5, pp. 507–521, Sep. 2014.
[24] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the lipschitz constant," *J. Optim. Theory Appl.*, vol. 79, no. 1, pp. 157–181, Oct. 1993.
[25] R. R. Deshpande and M. R. Bhatt, "Extraction of 2.1 D sketch from monocular images using DIRECT optimization algorithm," in *Proc. IEEE 7th Int. Advance Comput. Conf.*, Hyderabad, India, Jan. 2017, pp. 645–650.

RASHMI R. DESHPANDE received the B.E. degree in electronics and communication engineering from Karnataka University, Dharwad, in 2001, and the M.Tech. degree in digital electronics and communication from Visveswaraya Technological University, Bengaluru, in 2009, where she is currently pursuing the Ph.D. degree in image processing with the R. V. College of Engineering. She worked as an Assistant Professor for four years with the T. John Institute of Technology, Bengaluru, and for five years with the PES School of Engineering, Bengaluru. Her research interests include color filters, computational cameras, and 3D imaging.

CH. RENU MADHAVI (Senior Member, IEEE) received the M.Tech. degree in electronic instrumentation from NIT Warangal and the Ph.D. degree in the area of 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 the Ph.D. scholars. She has over 30 years of teaching experience. She has around 25 research publications in various national and international journals. She was a reviewer at few international conferences and journal conferences. She has Springer book chapter publications. Her research interests include biomedical signals, image processing, and the IoT.

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 University of Mysore, in 1986, and the Ph.D. degree from the Electrical Engineering Department, Indian Institute of Technology Bombay (IIT Bombay), in 1992. He served with research institutions, as a Senior Research Engineer for IIT Bombay, from 1989 to 1994, and as a Scientist E-2 for the Centre for Development of Imaging Technology, Thiruvananthapuram, from 1994 to 1996. He also served with several multinational organizations as a Research and Development Personnel. From 1996 to 1997, he held the position of Project Manager at LG Soft Development Center, Bengaluru. From 2000 to 2004, he was the Senior Software Manager at Analog Devices, Bengaluru Center. From 2005 to 2006, he held the position of Senior Manager with C2silicon, Bengaluru (a design center for IPFlex, Tokyo, Japan). From 2006 to 2010, he served as the Principal Member of Technical Staff with AMD Hyderabad Design Center. He was also a Senior IP Development Consultant, from 2010 to 2018. Further, he served for various academic institutions at VIT, currently known as VIT University, Vellore, as a full-time Professor, from 1996 to 1997; the Department of Electronics and Communication Engineering, T. John Institute of Technology, Bengaluru, from 2010 to 2013; and the Department of Medical Electronics, BMS College of Engineering, Bengaluru, from 2013 to 2019. He is currently serving as a Senior Consultant for Renalyx Health Systems Pvt., Ltd., and Rx Digi Health Platform Pvt., Ltd., Bengaluru. He is the author of over 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.