🔓 **Open Access**

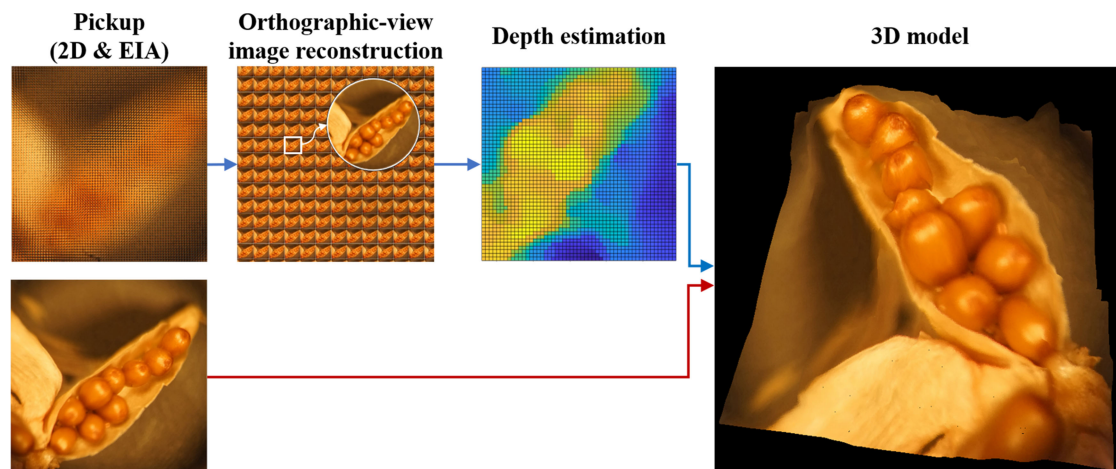# Advanced Three-Dimensional Visualization System for an Integral Imaging Microscope Using a Fully Convolutional Depth Estimation Network

Ki-Chul Kwon
Ki Hoon Kwon
Munkh-Uchral Erdenebat
Yan-Ling Piao
Young-Tae Lim
Yu Zhao
Min Young Kim
Nam Kim

Advanced Three-Dimensional Visualization System for an Integral Imaging Microscope Using a Fully Convolutional Depth Estimation Network

Ki-Chul Kwon [ID],[1] **Ki Hoon Kwon** [ID],[2] **Munkh-Uchral Erdenebat** [ID],[1]
Yan-Ling Piao [ID],[1] **Young-Tae Lim** [ID],[1] **Yu Zhao,**[3] **Min Young Kim** [ID],[2,4]
and Nam Kim [ID] [1]

[1]School of Information and Communication Engineering, Chungbuk National University, Cheongju, Chungbuk 28644, South Korea
[2]School of Electronics Engineering, Kyungpook National University, Daegu 41566, South Korea
[3]College of Information Engineering, Yangzhou University, Yangzhou 225009, China
[4]Research Center for Neurosurgical Robotic System, Kyungpook National University, Daegu 41566, South Korea

Abstract: In this paper, we propose an advanced three-dimensional visualization method for an integral imaging microscope system to simultaneously improve the resolution and quality of the reconstructed image. The main advance of the proposed method is that it generates a high-quality three-dimensional model without limitation of resolution by combining the high-resolution two-dimensional color image with depth data obtained through a fully convolutional neural network. First, the high-resolution two-dimensional image and an elemental image array for a specimen are captured, and the orthographic-view image is reconstructed from the elemental image array. Then, via a convolutional neural network-based depth estimation after the brightness of input images are uniformed, a more accurate and improved depth image is generated; and the noise of result depth image is filtered. Subsequently, the estimated depth data is combined with the high-resolution two-dimensional image and transformed into a high-quality three-dimensional model. In the experiment, it was confirmed that the displayed high-quality three-dimensional model could be visualized very similarly to the original image.

Index Terms: Integral imaging microscopy, resolution enhancement, high-quality reconstruction, fully convolutional depth estimation network.

1. Introduction

Integral imaging is an autostereoscopic three-dimensional (3D) imaging technique that captures and reproduces a full 3D information using a two-dimensional (2D) array of lenses [1]–[8]. It yields natural full-parallax, full-color, and continuous-viewing 3D images within a specific field-of-view and depth-of-field. By using the integral imaging technique, the integral imaging microscope (IIM) acquires the parallax and depth information of a specimen with a microlens array (MLA) in a single capture [9]–[11]. The elemental image array (EIA) contains the 3D information of specimen and is typically reconstructed from parallax-view images (i.e., orthographic-view images) or depth-slices, to provide 3D visualization [12]–[15]. However, the reconstructed 3D image has poor quality due to limited resolution of elemental images and others such as effects of illumination direction and small numerical aperture of the objective lens; therefore, the resolution of the IIM should be improved for the comfortable viewing. Also, even the depth-of-field of IIM is wider than optical 2D microscope according to the geometric or wave optics theories [9]–[11], [16], [17], it is quite insufficient for the comfortable viewing. Therefore, the resolution and depth-of-field of IIM should be improved. Note that the depth-of-field and resolution of IIM affect to each other [17], [18], [23].

For the improvement of depth-of-field, various methods can be applied [19]–[22]. For example, IIM systems using amplitude-modulated microlens arrays [19], spatial multiplexing method [20], bifocal holographic optical element-micro lens array [21] and switchable bifocal liquid-crystalline polymer micro lens array [22] have been suggested, and reported that shown very good results. However, due to the limited characteristics of IIM, it is quite difficult to apply various methods to improve the resolution. The resolution of the reconstructed image is enhanced by increasing the number of elemental images via mechanical movement of an optical device (e.g., an MLA) and/or software-based interpolation methods [23]–[25]. Also, the IIM with a configuration of Fourier plane recording that placing the MLA at the Fourier plane of the objective lens, not only improves the lateral resolution of the IIM but also improves the depth-of-field at the same time [26]. Although these methods enhance the resolution of the reconstructed images, they cannot provide satisfactory results because of poor image quality. Therefore, a method that can enhance the resolution of reconstructed images without degrading their quality is necessary for 3D visualization using an IIM.

A method for accurate depth estimation is required for an IIM system that can provide higher resolution and higher quality visualization. Although the depth information of a specimen is acquired by the EIA, the extraction of this information is quite difficult. The depth data can be acquired by active sensor-based methods using special equipment (e.g., depth cameras), passive sensor-based methods involving binocular or multi-view images, or hybrids of the two methods [27]–[30]. Active acquisition of depth data is accurate and rapid; however, the special equipment such as infrared or structured light, is required. In contrast, passive generation of a depth map from binocular or multi-view images requires only the image, but the subsequent calculations are complicated, and the quality is not consistent. In addition, after setting various constraints [31] (e.g., epipolarity, uniqueness, ordering, brightness changes, and smoothness), an optimal solution must be obtained for the accuracy of depth information.

In the case of general integral imaging and related light-field techniques except IIM, the depth information is extracted from the disparity between elemental images [32], [33]. Such methods do not require camera calibration; they also yield quite accurate depth information. Using the redundant information in IIM, the depth information can be derived for a specimen. However, if the above-mentioned active depth estimation methods are used for IIM, complex devices cannot be used. In addition, the extraction of depth information by existing methods does not provide good results because the reconstructed image from an IIM cannot yield images of uniform illumination using the passive depth estimation method. Note that more parallax information can be acquired through IIM than typical light field cameras and microscopes, but low depth-of-field is obtained [34].

Recently, deep learning-based image processing algorithms have been applied to integral imaging systems [35] and confirmed that such deep learning-based methods are suitable particularly for the resolution-enhancement of IIM [36]. However, regardless of the extent of improvement,
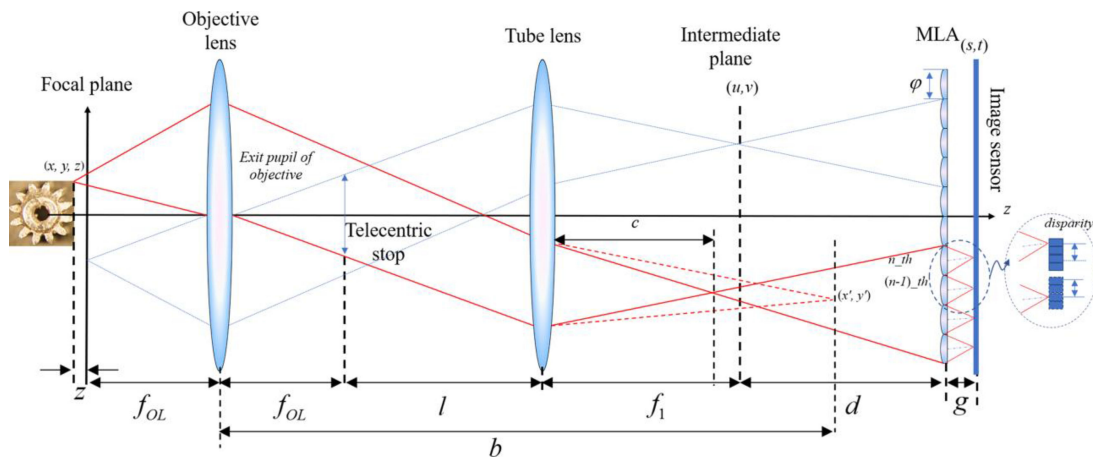
Fig. 1. Basic scheme of an IIM.

the resolution of reconstructed images for corresponding EIA captured by IIM, remains limited. Note that deep learning-based algorithms cannot recover the resolution degradation caused by IIM limitations. If the high-resolution and high-quality original 2D image and accurately estimated depth information of the specimen can be included in the visualization, reconstructed images could be obtained with better quality and improved resolution. As mentioned above, it is quite difficult to accurately estimate the depth information using only existing methods, due to various limitations of IIM systems. Here, we propose a 3D visualization method for IIM that provides improved resolution and quality for reconstruction, using a convolutional neural network (CNN)-based depth estimation algorithm. By using CNN-based depth estimation, the proposed algorithm generates high-quality 3D models (the resolution improvement/reduction does not affect to the quality) from the low-resolution depth image and high-resolution 2D color image of the specimen. To overcome the limitations of IIM systems, such as low resolution and low quality of the reconstruction caused by illumination effects and insufficient parallax disparity, the CNN-based model estimates depth using appropriate pre- and post-processing. The virtual 3D model is then regenerated by matching the high-resolution 2D red-green-blue (RGB) color image (captured after the MLA is removed) with the estimated depth image, instead of directly displaying the reconstructed images. When resolution is expanded through the proposed visualization method, the quality of reconstructed 3D model almost isn't deteriorated; therefore, the depth, color, and parallax data are represented simultaneously with much improved resolution and quality, which resolves the limitations of existing IIM systems.

2. Depth Extraction in IIM

2.1 Geometrical Analysis for the Disparity Between Elemental Images

Fig. 1 shows a schematic diagram of an IIM with an infinity-corrected optical system that consists of objective and tube lenses, an MLA, and an image sensor. A light beam emitted from a specimen passes through the objective and tube lenses, and an enlarged visualization of the specimen is imaged at the intermediate image plane. The MLA forms elemental images on the (s,t) plane from the intermediate image plane [6]–[12], [17]–[26]. Using the captured EIA, the orthographic-view image and depth slices can be reconstructed [13]–[15], [37], [38]. The number of elemental images defines the resolution of each reconstructed directional-view image of an orthographic-view image. In addition, each elemental image includes slightly different parallax information, and the calculation of the disparity between element images by geometric analysis is necessary for depth estimation.

When an arbitrary point on the specimen at position (x,y,z) is formed at position (x',y') and is separated from the objective lens by distance b, the distance b can be expressed as:

$$b = \frac{(z + f_{OL})f_{OL}}{z},$$ (1)

where z is the distance from the focal length of the objective lens to a corresponding arbitrary point on a specimen, and f_{OL} is the focal length of the objective lens. A new coordinate of the imaged point (x',y') can be expressed as:

$$(x', y') = \left(\frac{b}{z + f_{OL}}x, \ \frac{b}{z + f_{OL}}y \right).$$ (2)

When an arbitrary point is placed on the focal plane of an objective lens, the intermediate image plane is formed on the (u,v) plane, which is located at the focal length of the tube lens. If the point is not placed on the focal plane of the objective lens, the intermediate image plane will be formed at c, which is given by:

$$c = \frac{(f_{OL} + l - b)f_1}{(f_{OL} + l - b) - f_1} = \frac{(zl - f_{OL}^2)f_1}{z(l - f_1) - f_{OL}^2},$$ (3)

where f_1 is the focal length of the tube lens and l is the length of the infinity space of the infinity-corrected optical system. The coordinates of the intermediate image point (u,v) are given by:

$$(u, v) = \left(\frac{c}{f_{OL} - l - b}x', \ \frac{c}{f_{OL} - l - b}y' \right) = \left(\frac{f_{OL}f_1}{z(l - f_1) - f_{OL}^2}x, \ \frac{f_{OL}f_1}{z(l - f_1) - f_{OL}^2}y \right).$$ (4)

The coordinates (u,v) of the intermediate image plane correspond to the coordinates (s,t) of the image sensor through the MLA. When the arbitrary point (x, y, z) corresponds to the image sensor surface via each element lens of the MLA, the coordinates (s, t) after passing through the n-th elemental lens are given by:

$$(s, t)_n = \left(\frac{n_x\varphi g(z(l - f_1) - f_{OL}^2 - gf_{OL}f_1 x}{z(d(l - f_1) - f_1^2) - df_{OL}^2}, \ \frac{n_y\varphi g(z(l - f_1) - f_{OL}^2 - gf_{OL}f_1 y}{z(d(l - f_1) - f_1^2) - df_{OL}^2} \right),$$ (5)

where g is the gap between the MLA and the image sensor, d is the distance between the intermediate plane and the MLA, and φ is the pitch of the elemental lens. The disparity between the n-th and the $(n$-1)-th elemental images is expressed as:

$$disparity = \frac{\varphi g(z(l - f_1) - f_{OL}^2)}{z(d(l - f_1) - f_1^2 - df_{OL}^2}$$ (6)

However, the position of the image sensor corresponding to each element lens of the MLA (with respect to an arbitrary point $[x, y, z]$ of the specimen) is represented by an integer; therefore, the above geometrical analysis is limited in its ability to generate an accurate depth map.

2.2 CNN-Based Depth Estimation

Existing image processing methods, such as the previously reported cost-volume method-based multiview stereo method [39], can show good depth estimation performance for common multiview imaging systems. However, they are ineffective when applied directly to microscopes due to the limitations of the microscopes' imaging systems. On the other hand, several image-based depth estimation methods based on CNN which is one of the deep learning algorithms specialized in image processing, have been studied and have shown good results in terms of depth estimation accuracy and computation speed in recent years. [39], [40]. The epipolar geometry for depth from light field images (EPINET) proposed by Shin *et al.* is a recent algorithm with good performance, which enables more accurate depth estimation using CNN [42]. Originally, the EPINET model requires a dataset composed of directional-view images and disparity map for learning, However,
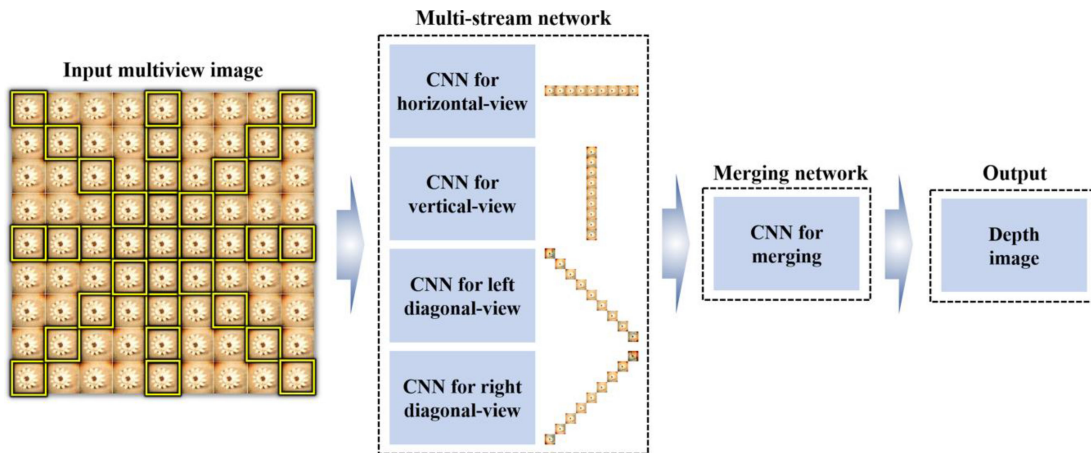
Fig. 2. Light field depth estimation architecture of conventional EPINET.

it is difficult to obtain the dataset that satisfies the conditions above; so, in the case of EPINET, the amount of data is inflated through data augmentation and utilized in the learning process. It considers the light field epipolar geometry and consists of a multi-stream network that operates only in four view directions (horizontal, vertical, and left- and right-diagonal views), as well as a merging network that matches the results of the four directions (Fig. 2). Unlike existing methods, which utilize all perspectives as input images, the EPINET is more rapid and more efficient. In addition, the depth estimation accuracy is higher because the convolution kernel for depth estimation considers the geometry in four view directions; it also utilizes only 2×2 kernels, and is therefore suitable for low-resolution imaging systems, especially for the proposed IIM system.

3. Proposed IIM System Using CNN-Based Depth Estimation Algorithm

Good-quality EIA cannot be obtained by IIM due to inferior uniformity between viewpoints caused by the low-resolution, light reflection, and high-magnification-related distortion of the acquired image. To provide high-resolution reconstruction for IIM, in this paper, a visualization method using a CNN-based depth estimation (consisting of the CNN model with appropriate pre- and post-processing for depth estimation), is proposed and implemented. After optical capture of the EIA by the IIM, an orthographic-view image is reconstructed. Notably, existing depth estimation methods cannot estimate the depth of the specimen due to limitations of elemental images acquired by the IIM, such as low-resolution and inferior image uniformity. In order to solve these limitations, in the proposed system, the approximate depth information is estimated by CNN model with appropriate pre- and post-processing, then scaled up to achieve a size equal to that of a high-resolution 2D image of the specimen that captured through high-resolution camera without contain a MLA. By mapping the high-resolution 2D image (RGB) and depth image (Depth) estimated through CNN-based depth estimation, a high-quality color 3D model (RGB-D) that can be directly edited in a 3D coordinate system, is generated. Note that the point cloud generation is an effective way to check the accuracy of the information of a 3D object, like depth map [30]. Finally, the point cloud interpolation process generates a greater number of points, enabling visualization of a smoother and higher quality 3D model without limitations regarding its resolution. Fig. 3 shows a schematic diagram of the proposed CNN-based depth estimation algorithm-based visualization system for IIM.

The proposed method consists of several stages for precise depth estimation and visualization: IIM-based EIA capture and orthographic-view image reconstruction; preprocessing that produces directional-view images with uniformed brightness and depth estimation through CNN model from the disparity between the sub-images; and post-processing for noise filtering, in order to generate
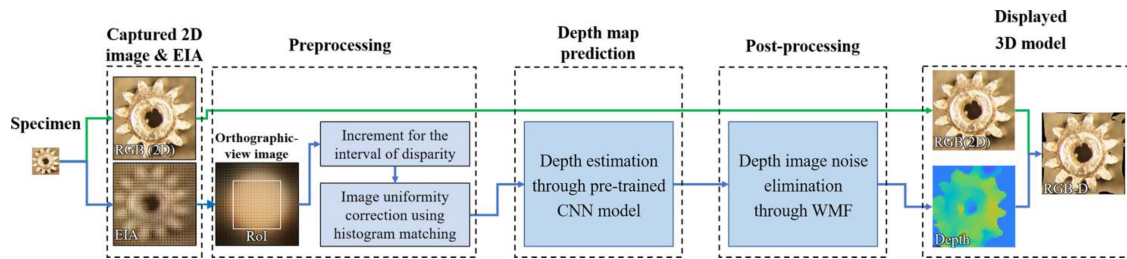
Fig. 3. Schematic diagram of the proposed IIM system with a 3D visualization method using the CNN-based depth estimation algorithm.
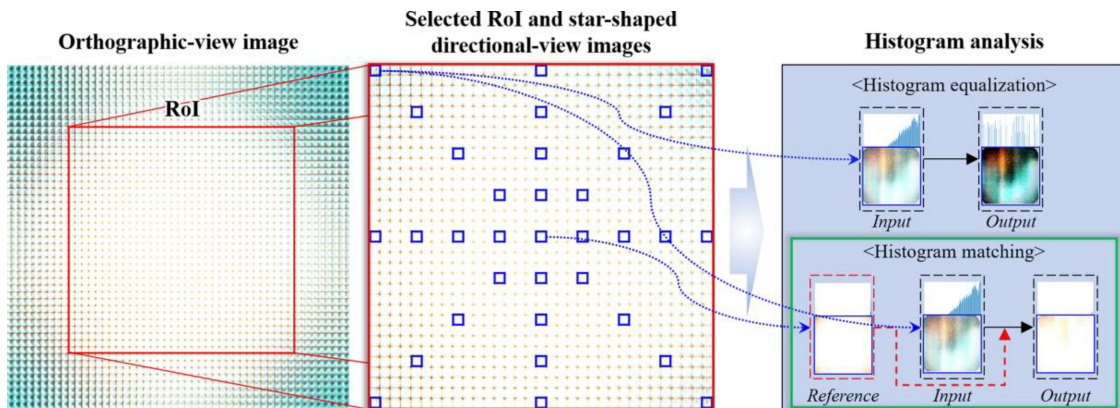

Fig. 4. Histogram matching for the selected directional-view images.

and display the high-quality 3D point cloud model. Preprocessing involves increasing the interval of disparity between the input directional-view images, which are selected in a star-shaped pattern, and correcting the imbalance in brightness between the images. When the orthographic-view image is reconstructed from the EIA captured through the IIM, as shown in Fig. 4, the images become progressively darker from the center outward due to the characteristics of the microscopic illumination and MLA. Therefore, the images at the edges cannot be inputted for depth estimation; accordingly, the region-of-interest (RoI) is selected first. When the neighboring directional-view images are input, the imbalance of brightness is smaller between input images, but the parallax information is insufficient for depth estimation. Therefore, the gap between the input directional-view images is then considered. To utilize the required number of directional-view images with minimum illumination influence, the input directional-view images are selected with intervals of a given number of images apart according to orthographic-view image properties, instead of using neighboring images. If the selection interval is large, the parallax between directional-view images that will be input to CNN model is increased; however, the brightness of the output images can be non-uniform due to the influences of the lens and illumination, or poor-quality outer-view images can be selected.

The brightness distribution of directional-view images gradually changes from the center to the outer portions of the orthographic-view image reconstruction of IIM for any object. The non-uniformity of brightness between images causes errors and noise in depth estimation. Histogram matching (i.e., histogram specification) is used to overcome this issue. Here, the central directional-view image is chosen as a reference, and the histograms of other selected images are matched to the histogram of the reference image. Unlike histogram equalization performed separately, as shown in Fig. 4, histogram matching using the central image as a reference can facilitate the
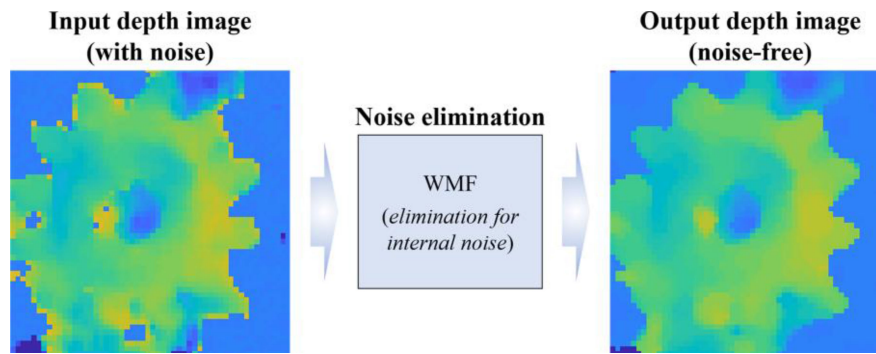
Fig. 5. Elimination of internal noise by the WMF method.

uniformity among all input directional-view images, including those in the outer portion. Therefore, the result images are much smoother and brighter like reference image than the original images. In order to present the actual shape of the orthographic-view image and enhance understanding of the histogram matching explained above, the EIA of white background, and reconstructed orthographic-view image with different brightnesses are presented in Fig. 4.

In the next stage, the histogram-matched directional-view images are input into a pre-trained CNN model to estimate their depth. Unlike the conventional EPINET algorithm, depth estimation is performed by CNN model after appropriate preprocessing stage. However, internal noise in the output images can occur due to the limitations of the IIM system. Therefore, appropriate post-processing is required to compensate for noise in the depth estimation process. Here, a weighted median filtering (WMF) method is utilized during post-processing for depth estimation. In particular, the WMF removes the internal noise while preserving boundary information, which is important for depth estimation. Thus, it is suitable for use in proposed method. Here, the central directional-view image is used as a reference image for the WMF. When this process is completed, the depth estimation image is acquired. Fig. 5 shows that a better-quality image is obtained by applying the WMF to depth estimation by proposed method.

Based on the estimated depth information and high-resolution 2D color image of a specimen, a 3D point cloud model is generated. Note that the high-resolution 2D image is captured by a digital single-lens reflex camera mounted on a microscope, by removing the MLA after EIA capture, from the same specimen. The high-resolution 2D color image and corresponding estimated depth image, which is resized for matching to the resolution of 2D color image, are stored as pixels in a matrix arrangement; each matrix contains red-green-blue values that are dependent on the intensity of each component. Image pixels storing color data are used for matrix operation/manipulation to generate high-level variations for a 3D point cloud. 3D mapping commences with feature extraction and matching, followed by geometric verification. Note that this point cloud model is generated based on the directional-view images; thus, when the viewing direction is slightly changed, empty holes can be observed. To create a smoother point cloud model, additional object points are interpolated by using the surface-interpolating corner vertices and boundary curves to fill the holes. Fig. 6 shows examples of initial and interpolation-based (hole-filled) 3D point cloud models for a given object.

The final regenerated high-quality 3D model includes accurate parallax and depth information of the specimen than the reconstructed directional-view images.

4. Experimental Results

The prototype IIM system consisted of the infinity-corrected optical system, an MLA, and a digital single-lens reflex camera (Fig. 7(a)). The device specifications are listed in Table 1.
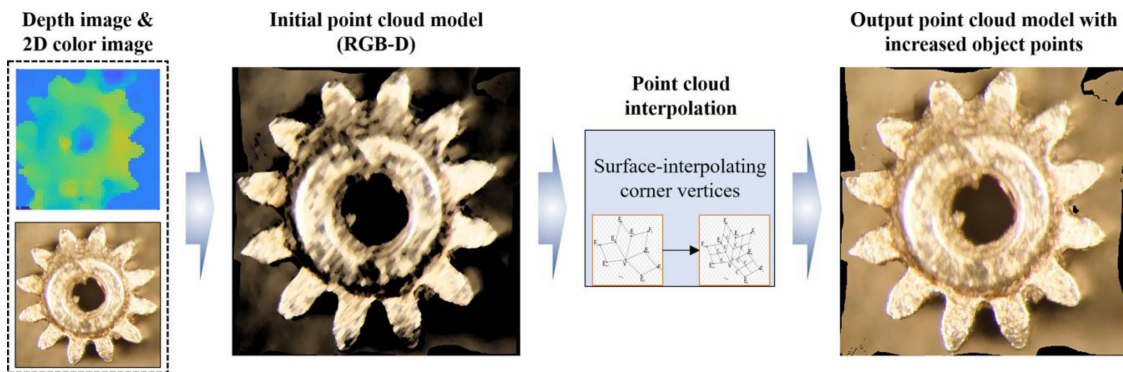
Fig. 6. The initial point cloud model generated from the RGB and depth images, and final point cloud model after the interpolation process.

TABLE 1
Devices' Specifications for the Implemented System

Devices	Indices	Specifications
IIM unit	Objective lens	×10
		NA=0.28
	Tube lens	×10
MLA	Number of lenses	100×100 lenses
	Elemental lens pitch	125 μm
	Focal length	2.4 mm
Camera	Model	Sony α6000
	Sensor resolution	6000×4000 pixels
	Pixel pitch	3.88 μm
PC	CPU	Intel i7-4790 3.6 GHz / 8 Cores
	Memory	16 GB
	Operating system	Windows 10 Pro (64-bit)

Three different objects were utilized in the experiment: a chip resistor (Fig. 7 (b)), a seedpod of a flower (Fig. 7 (c)), and a watch gear (Fig. 7 (d)). The EIAs of the three objects were captured with 76×76 elemental images, and the total resolution was 4028×4028 pixels, which was identical to that of the 2D color images. The orthographic-view images, consisting of 53×53 directional-view images, were reconstructed from the EIAs, where the resolution of each directional-view image was 76×76 pixels. Here, the RoIs were selected for each orthographic-view image with 36×36 directional-view images; selected 33 directional-view images (nine images in each of four directions in a star shape at intervals of three images) were chosen for depth estimation.

4.1 Depth Estimation
Before depth estimation, a histogram matching process was conducted to correct for imbalanced brightness in the selected directional-view images. Fig. 8 shows the histogram-matched
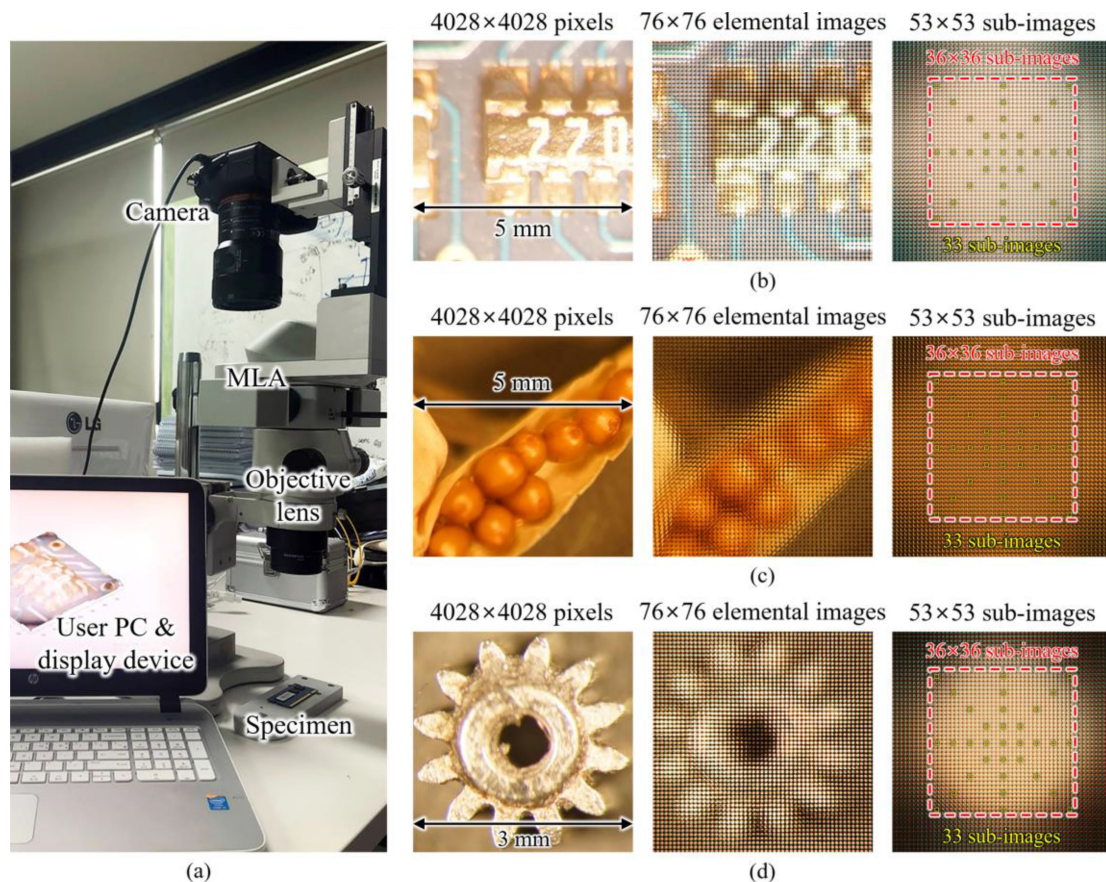
Fig. 7. (a) Prototype of the proposed IIM system; 2D images (left), EIAs (center), and orthographic-view images (right) of the (b) chip, (c) seedpod, and (d) gear. Red squares are the RoIs; yellow squares are 33 selected directional-view images for the proposed CNN-based depth estimation algorithm.

directional-view images of the three objects. The upper parts show the input images, and the lower parts represent the images with uniformed brightness through the histogram matching process (preprocessing). It has been confirmed that the selected directional-view images' brightnesses have been converted similarly to the reference image (central directional-view image).

Fig. 9 shows examples of directional-view images with uniform brightness for the three objects, as well as the estimated depth results (initial depth images and noise-filtered images using the WMF algorithm) from the input images. The proposed CNN-based depth estimation algorithm created a depth map for each object. Here, the depth images generated by the depth estimation algorithm were compared with images generated using existing methods (e.g., the conventional EPINET method and the cost volume-based stereo method) [39], [42]. Using the previous methods, the depth difference between the object and background was unclear and there was excessive noise (Fig. 9(c) and (d)). The results were imperfect due to the constraints of the input directional-view images; however, the depth between the object and background was distinguished more clearly, and the internal depth information of the object was successfully estimated (Fig. 9(b)). Importantly, the shape of the object was evident in the depth images.

4.2 3D Visualization

Fig. 10(a) and Videos 4–6 show the final 3D visualizations of three objects (Video 4: chip; Video 5: seedpod; Video 6, gear). Here, the 3D point cloud models were regenerated based on the
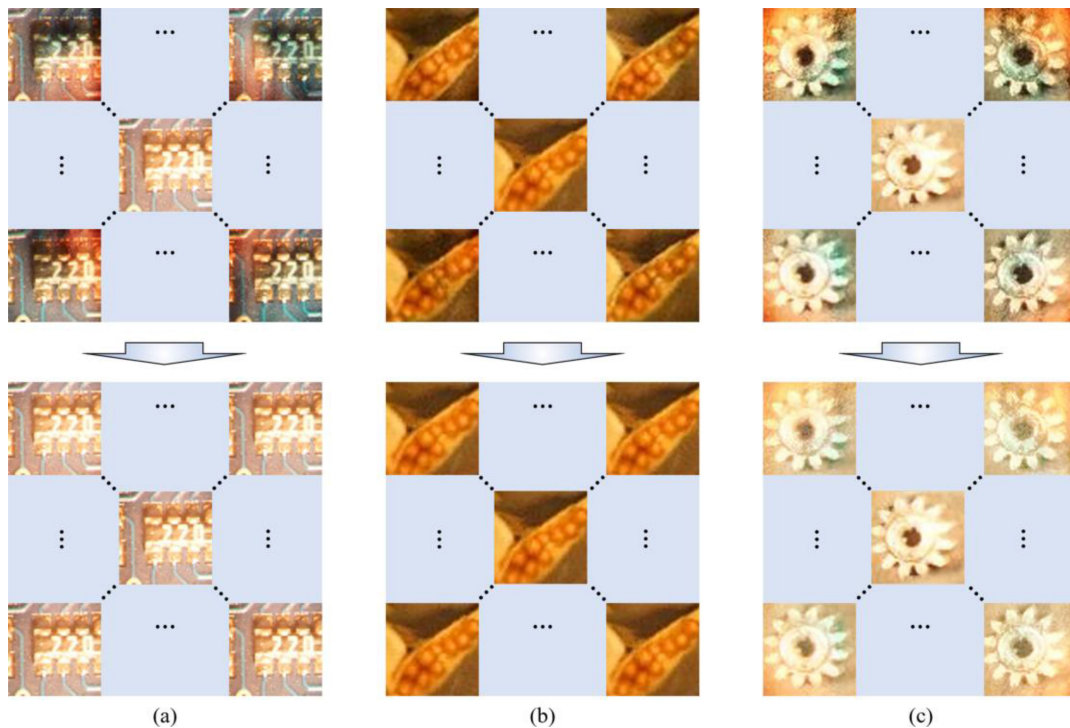
Fig. 8. Correction for non-uniform brightness in directional-view reconstructed images of the (a) chip, (b) seedpod, and (c) gear.

estimated depth images and point cloud interpolation and displayed on the display device. The 3D point cloud models were regenerated using 577,600 object points (2916 initially) for the chip, 291,600 object points (2916 initially) for the seedpod, and 290,393 object points (2916 initially) for the gear, following surface subdivision-based point interpolation that included nearly 10–20-fold more points than the initially generated point cloud models. From these regenerated 3D models, each user can select the desired images. Additionally, Fig. 10(b) shows directional-view images of the three objects. The comparison shows that the proposed CNN-based depth estimation method yields more accurate and clearer 3D visualizations than existing orthographic-view image-based IIM displays. The proposed method also enables visualization with better resolution and includes zoom-in/zoom-out functions.

4.3 Evaluations for Output Images

The proposed CNN-based depth estimation method yielded higher-quality visualizations (Figs. 11 and 12); we also performed quantitative evaluations of image quality. Specifically, image contrast was evaluated by the discrete entropy method [43] and the quality of the displayed images was evaluated by the power spectral density (PSD) method [20]–[22].

In addition, the depth images (Fig. 11) where exhibited better contrast than the existing methods, as determined by the discrete entropy formula. For the chip, the discrete entropy values were 4.71 for the proposed CNN-based depth estimation method, 3.16 for the conventional EPINET method, and 1.78 for the cost volume-based stereo method. For the seedpod, the discrete entropy values were 6.59 for the proposed method, 4.83 for the conventional EPINET method, and 2.99 for the cost volume-based method. For the gear, the discrete entropy values were 5.8 for the proposed method, 4.65 for the conventional EPINET method, and 1.25 for the cost volume-based method.
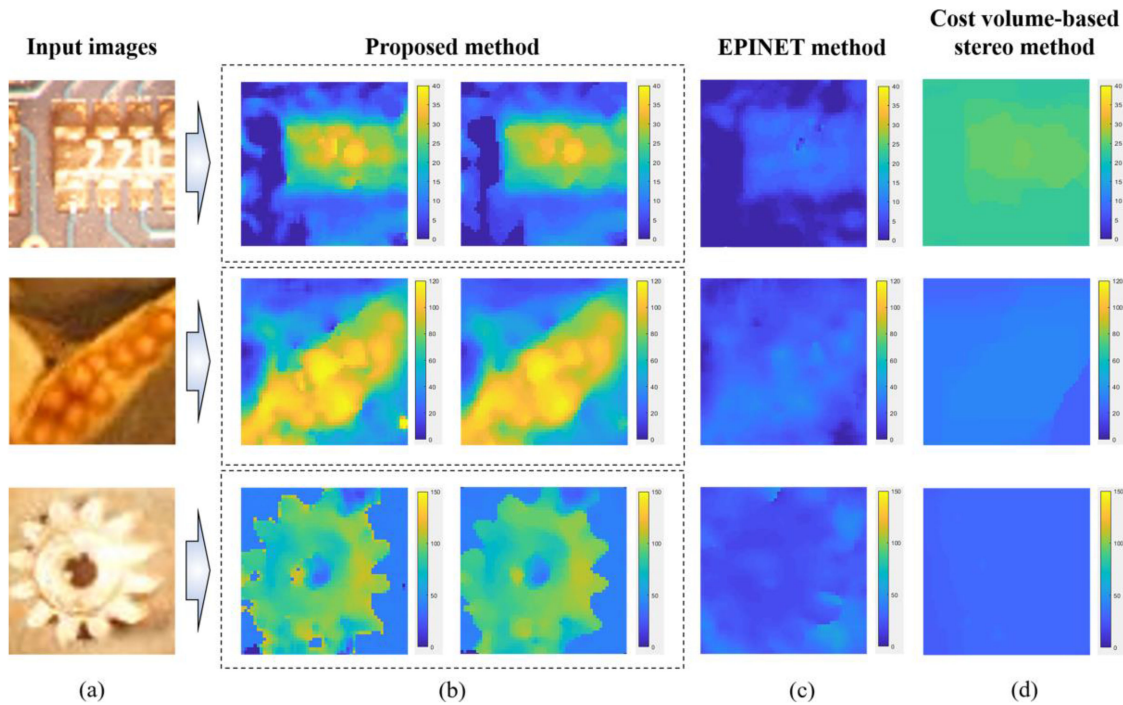
Fig. 9. (a) For input histogram-matched directional-view images, (b) the depth images of the three objects (left) were generated, and noise of each image was eliminated through post-processing (right, Video 1: chip; Video 2: seedpod; Video 3: gear). Depth images generated by (c) conventional EPINET [42] and (d) cost-volume-based stereo [39] methods, from the same input images.
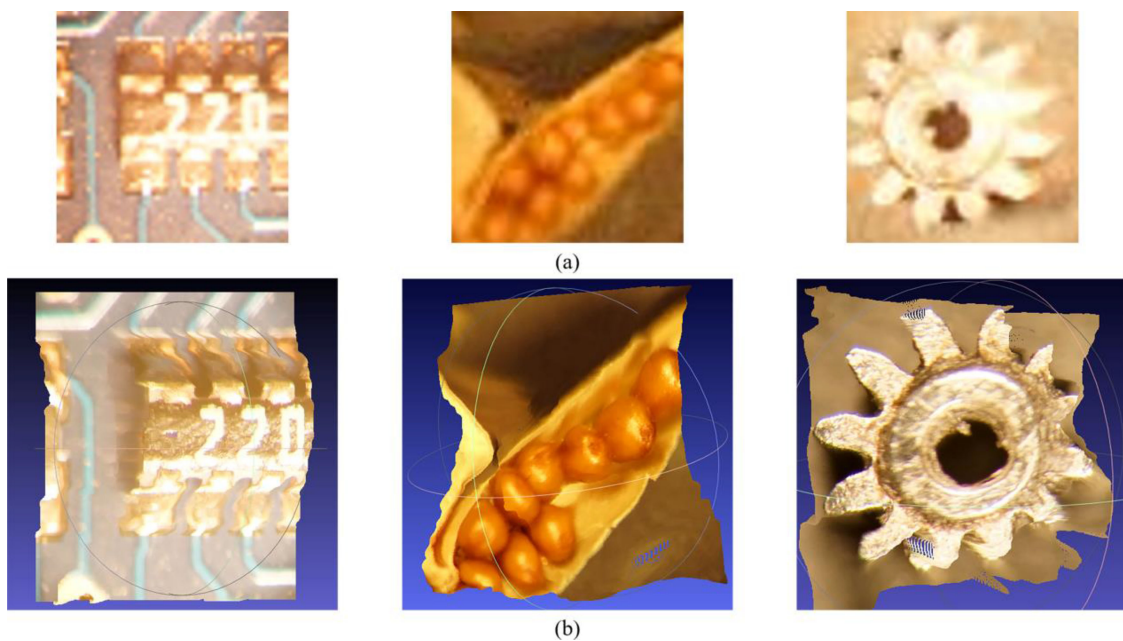

Fig. 10. 3D point cloud models for (a) chip (Video 4), seedpod (Video 5), and gear (Video 6), compared with the (b) corresponding directional-view images with different viewpoints.
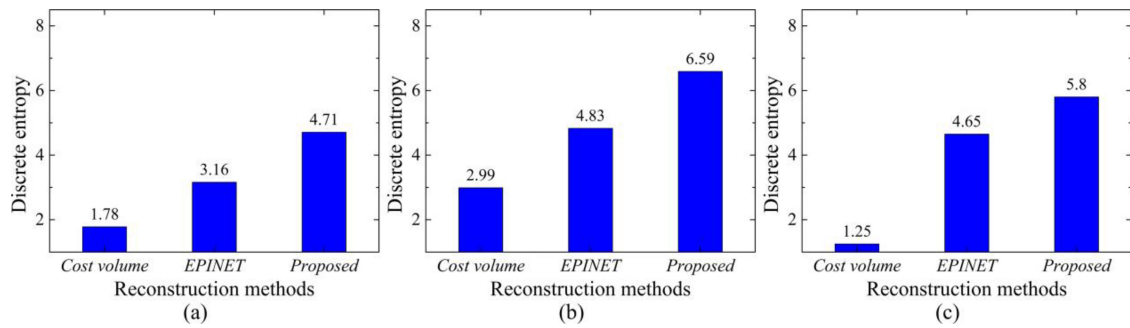
Fig. 11. Discrete entropy values for (a) chip, (b) seedpod, and (c) gear using the cost volume-based stereo [39], EPINET [42] and proposed CNN-based depth estimation methods.
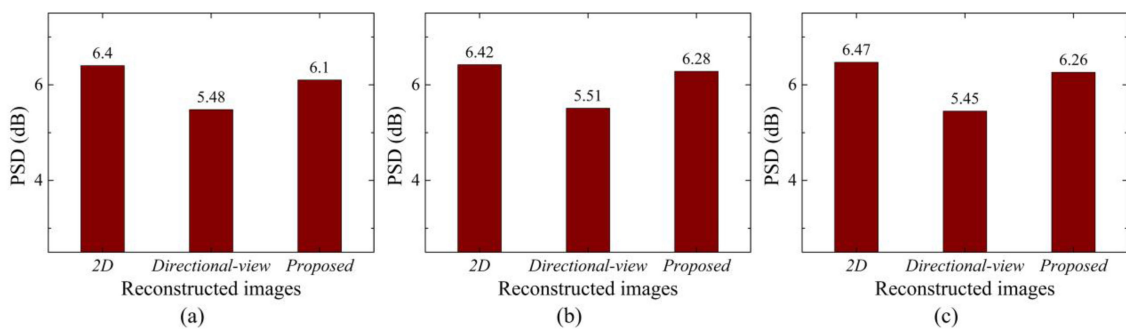

Fig. 12. Measured PSD values of 2D images, central directional-view images, and central viewpoints of final 3D point cloud models for (a) chip, (b) seedpod, and (c) gear.

Fig. 12 shows the measured PSD values for the proposed and existing orthographic-view image reconstructions. The center viewpoints of three 3D models, the central directional-view images for the three objects, shown in Fig. 10, were utilized in this comparison. Because, we measured the PSD values of corresponding high-resolution 2D images additionally, therefore, in order to determine whether the results from both reconstructions were like the corresponding original observations. The PSD values of images (reconstructed by proposed method) shown in Fig. 10(a) were 6.1 dB for the chip (Fig. 12(a)), 6.28 dB for the seedpod (Fig. 12(b)), and 6.26 dB for the gear (Fig. 12(c)); the PSD values of images shown in Fig. 10(b) were 5.48 dB, 5.51 dB, and 5.45 dB, respectively. The PSD values for 2D images were 6.4 dB (chip), 6.42 dB (seedpod), and 6.47 dB (gear). In Fig. 12, "2D" indicates the PSD values of 2D color images, "Initial" denotes the PSD values of reconstructed central directional-view images, and "Proposed" shows the PSD values of the central viewpoint of 3D models output by the proposed CNN-based depth estimation-based visualization method.

These PSD values verified that the proposed method enabled clearer 3D visualization of the three specimens. Furthermore, these values confirmed that the proposed method recapitulated the original appearance of the objects.

5. Conclusion

In this study, we investigated an advanced visualization method for IIM display using a high-quality 3D model generation method based on a CNN-based depth estimation algorithm. The proposed method avoids the poor-quality 3D reconstructions of IIM display, and simultaneously improves the

resolution and quality of the displayed image. In other words, the proposed method contains the advantages of both existing high-resolution 2D microscope and 3D IIM; and can be a distinguished solution for the disadvantages of both methods that 2D-only display of conventional 2D optical microscope and low image quality of IIM (the original image quality cannot be restored). Here, first, the EIA for the specimen is captured optically by the IIM and the characteristics of the elemental images are analyzed; the EIA is reconstructed into directional-view images suitable for depth estimation. Next, CNN-based depth estimation is performed using the proposed CNN-based depth estimation method from the disparity between the directional-view images. The estimated depth data are converted into the point cloud, while a high-resolution 2D image supports the color and texture, and a high-quality 3D model is visualized without any limitation of resolution.

Existing IIM displays present 2D directional-view images or depth-slices of poor quality and resolution by time- or spatial-division methods. Therefore, the existing display can be regarded as 2D visualization, including 3D information. The proposed approach involves a high-quality 3D model with various functions (e.g., rotation along three axes and zoom-in/zoom-out), which improves the resolution and quality of the image. The experimental results confirmed that the proposed system is highly efficient for metal objects, such as the watch gear used in this study. Especially, there was almost no deterioration in image quality compared to the original 2D high-resolution images, when the actual 3D model was generated. Further research should focus on 3D visualization and measurement by acquiring and analyzing more data to enable to create a deep learning model for accurate 3D information regarding micro objects. Also, further research considers the implementation of real-time IIM 3D display system including two optical axes for capturing the EIA and the original 2D image at once.

References

[1] G. Lippmann, "Reversible prints giving the sensation of relief," *J. Phys. Theor. Appl.*, vol. 7, pp. 821–825, 1908, doi: 10.1051/jphystap:019080070082100.
[2] F. Okano, H. Hoshino, J. Arai, and I. Yuyama, "Real-time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.*, vol. 36, pp. 1598–1603, 1997.
[3] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: Sensing, display, and applications," *Appl. Opt.*, vol. 52, pp. 546–560, 2013.
[4] N. Kim *et al.*, "3D display technology," *Display Imag.*, vol. 1, no. 1, pp. 73–95, 2014.
[5] N. Kim, M. A. Alam, L. T. Bang, A.-H. Phan, M.-L. Piao, and M.-U. Erdenebat, "Advances in the light field displays based on integral imaging and holographic techniques," *Chin. Opt. Lett.*, vol. 12, no. 6, 2014, Art. no. 06005.
[6] N. Kim and M.-U. Erdenebat, "3-D integral photography," SL18, SPIE Press, 2016.
[7] M. Martínez-Corral, A. Dorado, J. C. Barreiro, G. Saavedra, and B. Javidi, "Recent advances in the capture and display of macroscopic and microscopic 3-D scenes by integral imaging," in *Proc. IEEE*, vol. 105, pp. 825–836, 2017.
[8] M. Martínez-Corral and B. Javidi, "Fundamentals of 3D imaging and displays: A tutorial on integral imaging, light-field, and plenoptic systems," *Adv. Opt. Photon.*, vol. 10, pp. 512–566, 2018.
[9] J.-S. Jang and B. Javidi, "Three-dimensional integral imaging of micro-objects," *Opt. Lett.*, vol. 29, no. 11, pp. 1230–1232, 2004.
[10] B. Javidi, I. Moon, and S. Yeom, "Three-dimensional identification of biological microorganism using integral imaging," *Opt. Express*, vol. 14, pp. 12096–12108, 2006.
[11] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," in *Proc. SIGGRAPH*, 2006, pp. 924–934.
[12] A. Hassanfiroozi, Y.-P. Huang, B. Javidi, and H.-P. D. Shieh, "Hexagonal liquid crystal lens array for 3D endoscopy," *Opt. Express*, vol. 23, pp. 971–981, 2015.
[13] H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," *Opt. Lett.*, vol. 26, pp. 157–159, 2001.
[14] A. Stern and B. Javidi, "3D image sensing and reconstruction with time-division multiplexed computational integral imaging (CII)," in *Proc. SPIE ITCom*, 2003, pp. 131–139.
[15] S.-H. Hong, J.-S. Jang, and B. Javidi, "Three-dimensional volumetric object reconstruction using computational integral imaging," *Opt. Express*, vol. 12, pp. 483–491, 2004.
[16] C.-G. Luo, X. Xiao, M. Martínez-Corral, C.-W. Chen, B. Javidi, and Q.-H. Wang, "Analysis of the depth of field of integral imaging displays based on wave optics," *Opt. Express*, vol. 21, pp. 31263–31273, 2013.
[17] Y.-T. Lim, J.-H. Park, K.-C. Kwon, and N. Kim, "Analysis on enhanced depth of field for integral imaging microscope," *Opt. Express*, vol. 20, no. 21, pp. 23480–23488, 2012.
[18] F. Jin, J.-S. Jang, and B. Javidi, "Effects of device resolution on three-dimensional integral imaging," *Opt. Lett.*, vol. 29, pp. 1345–1347, 2004.
[19] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Integral imaging with improved depth of field by use of amplitude-modulated microlens arrays," *Appl. Opt.*, vol. 43, pp. 5806–5813, 2004.

[20] K.-C. Kwon, M.-U. Erdenebat, M. A. Alam, Y.-T. Lim, K. G. Kim, and N. Kim, "Integral imaging microscopy with enhanced depth-of-field using a spatial multiplexing," *Opt. Express*, vol. 24, no. 3, pp. 2072–2083, 2016.

[21] K.-C. Kwon, Y.-T. Lim, C.-W. Shin, M.-U. Erdenebat, J.-M. Hwang, and N. Kim, "Enhanced depth-of-field of an integral imaging microscope using a bifocal holographic optical element-micro lens array," *Opt. Lett.*, vol. 42, no. 16, pp. 3209–3212, 2017.

[22] K.-C. Kwon *et al.*, "Enhancement of the depth-of-field of integral imaging microscope by using switchable bifocal liquid-crystalline polymer micro lens array," *Opt. Express*, vol. 25, no. 24, pp. 30503–30512, 2017.

[23] Y.-T. Lim, J.-H. Park, K.-C. Kwon, and N. Kim, "Resolution-enhanced integral imaging microscopy that uses lens array shifting," *Opt. Express*, vol. 17, no. 21, pp. 19253–19263, 2009.

[24] K.-C. Kwon, J.-S. Jeong, M.-U. Erdenebat, Y.-T. Lim, K.-H. Yoo, and N. Kim, "Real-time interactive display for integral imaging microscopy," *Appl. Opt.*, vol. 53, no. 20, pp. 4450–4459, 2014.

[25] K.-C. Kwon, J.-S. Jeong, M.-U. Erdenebat, Y.-L. Piao, K.-H. Yoo, and N. Kim, "Resolution-enhancement for an orthographic-view image display in an integral imaging microscope system," *Biomed. Opt. Express*, vol. 6, no. 3, pp. 736–746, 2015.

[26] A. Llavador, J. Sola-Pikabea, G. Saavedra, B. Javidi, and M. Martínez-Corral, "Resolution improvements in integral microscopy with Fourier plane recording," *Opt. Express*, vol. 24, pp. 20792–20798, 2016.

[27] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation of early vision," *Int. J. Comput. Vision*, vol. 70, no. 1, pp. 41–54, 2006.

[28] Y.-S. Ho and W.-S. Jang, "Occlusion detection using warping and cross-checking constraints for stereo matching," in *Proc. Era Interactive Media*, 2013, pp. 363–372.

[29] Q. Yang, K.-H. Tan, B. Culbertson, and J. Apostolopoulos, "Fusion of active and passive sensors for fast 3D capture," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2010, pp. 69–74.

[30] A. S. Malik, T.-S. Choi, and H. Nisar, "Depth map and 3D imaging applications: Algorithms and technologies," IGI Global, 2012.

[31] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, no. 4262, pp. 283–287, 1976.

[32] J.-H. Park, S. Jung, H. Choi, Y. Kim, and B. Lee, "Depth extraction by use of a rectangular lens array and one-dimensional elemental image modification," *Appl. Opt.*, vol. 43, no. 25, pp. 4882–4895, 2004.

[33] J.-Y. Jang, J.-I. Ser, S. Cha, and S.-H. Shin, "Depth extraction by using the correlation of the periodic function with an elemental image in integral imaging," *Appl. Opt.*, vol. 51, no. 16, pp. 3279–3286, 2012.

[34] S. Zhu, A. Lai, K. Eaton, P. Jin, and L. Gao, "On the fundamental comparison between unfocused and focused light field cameras," *Appl. Opt.*, vol. 57, no. 1, pp. A1–A11, 2018.

[35] A. Markman and B. Javidi, "Learning in the dark: 3D integral imaging object recognition in very low illumination conditions using convolutional neural networks," *OSA Continuum*, vol. 1, pp. 373–383, 2018.

[36] K.-C. Kwon *et al.*, "Resolution-enhancement for an integral imaging microscopy using deep learning," *IEEE Photon. J.*, vol. 11, no. 1, Feb. 2019, Art. no. 6900512.

[37] J.-H. Park, G. Baasantseren, N. Kim, G. Park, J.-M. Kang, and B. Lee, "View image generation in perspective and orthographic projection geometry based on integral imaging," *Opt. Express*, vol. 16, no. 12, pp. 8800–8813, 2008.

[38] D.-C. Hwang, D.-H. Shin, S.-C. Kim, and E.-S. Kim, "Depth extraction of three-dimensional objects in space by the computational integral imaging reconstruction technique," *Appl. Opt.*, vol. 47, no. 19, pp. D128–D135, 2008.

[39] H.-G. Jeon *et al.*, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 1547–1555.

[40] A. Kendall *et al.*, "End-to-end learning of geometry and context for deep stereo regression," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 66–75.

[41] P. Knobelreiter, C. Reinbacher, A. Shekhovtsov, and T. Pock, "End-to-end training of hybrid CNN-CRF models for stereo," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 2339–2348.

[42] C. Shin, H.-G. Jeon, Y. Yoon, I. S. Kweon, and S. J. Kim, "EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 4748–4757.

[43] B. H. Kim, C. Bohak, K. H. Kwon, and M. Y. Kim, "Cross fusion based low dynamic and saturated image enhancement for infrared search and tracking systems," *IEEE Access*, vol. 8, pp. 15347–15359, 2020.