Generic camera model and its calibration for computational integral imaging and 3D reconstruction

Weiming Li and Youfu Li*

Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong
**Corresponding author: meyfli@cityu.edu.hk*

Received May 7, 2010; revised October 11, 2010; accepted December 13, 2010; posted December 23, 2010 (Doc. ID 128165); published February 10, 2011

Integral imaging (II) is an important 3D imaging technology. To reconstruct 3D information of the viewed objects, modeling and calibrating the optical pickup process of II are necessary. This work focuses on the modeling and calibration of an II system consisting of a lenslet array, an imaging lens, and a charge-coupled device camera. Most existing work on such systems assumes a pinhole array model (PAM). In this work, we explore a generic camera model that accommodates more generality. This model is an empirical model based on measurements, and we constructed a setup for its calibration. Experimental results show a significant difference between the generic camera model and the PAM. Images of planar patterns and 3D objects were computationally reconstructed with the generic camera model. Compared with the images reconstructed using the PAM, the images present higher fidelity and preserve more high spatial frequency components. To the best of our knowledge, this is the first attempt in applying a generic camera model to an II system.

OCIS codes: 100.3010, 110.6880, 100.6890.

1. INTRODUCTION

Integral imaging (II) [1–3] is important for implementing 3D autostereoscopic display systems, which can operate in natural light and provide full parallax, continuous viewing points to users without any special glasses. In general, II consists of two processes: pickup and reconstruction. In the pickup process, light rays from objects are captured through a lenslet array by an image sensor such as a charge-coupled device (CCD). The light rays passing through each lenslet are from a particular view and are recorded in an elemental image (EI). All the elemental images form an elemental image array (EIA) that simultaneously captures multiple views of the 3D scene. The reconstruction process is a reverse process of the pickup, and it can be approached by either optical integral image reconstruction (OIIR) [4–7] or computational integral imaging reconstruction (CIIR) [8–11]. In OIIR techniques, the EIA is displayed to an optical setup similar to the one used for pickup to generate optical 3D image to observers with the naked eye. On the other hand, CIIR explores appropriate digital processing to computationally synthesize images of the viewed 3D scene or, further, to reconstruct a 3D digital model of the scene. This work focuses on the model and calibration in CIIR for reconstructing 3D information from the EIA.

Because CIIR aims at digitally reversing the optical pickup process, using an appropriate model to describe the optical pickup process is of crucial importance. The most often used model in existing CIIR methods is a pinhole array model (PAM) [10,11]. The PAM works well for many II systems in existing work [11–14]. However, there are still II systems whose optical configurations are known only within a degree of certainty. Theoretical analysis [15,16] shows that the misalignment between optical pickup components would degrade the image reconstruction quality. In particular, a loss of the high spatial frequency component is observed in CIIR images [15]. Furthermore, there might be nonlinear optical distortion existing in the lenses, which cannot be accounted for by pinhole models. In some specific systems with limited fabrication precision, when optical distortions and installation misalignments amongst different optical components are coupled, the real optical process becomes more complicated.

In this work, instead of assuming a PAM, we explore a generic camera model [17–19] that accommodates more generality to describe the real optical pickup process. The generic camera model was first proposed in modeling and calibrating omnidirectional vision sensors whose optical geometry cannot be described by traditional pinhole camera models. The highly generic nature of generic camera models makes them suitable for describing imaging systems with arbitrarily complex imaging geometry, such as the II system studied in this work. The generic camera model is an empirical model based on measurements. In this work, a calibration setup was constructed for its calibration. Using an LCD panel as an active calibration object, the correspondences between pixels in the EIA and light rays in 3D space can be automatically established in high density and precision. The integral imaging system (IIS) used in this work is depicted in Fig. 1. The setup is similar to those in the existing work [8,11].

Our experiments show that there is a significant difference between the generic camera model of this IIS and a PAM. Using the generic camera model, images of objects can be computationally reconstructed. Compared with the images reconstructed based on a PAM, the visual quality is enhanced in terms of the reconstruction fidelity and the preservation of high spatial frequency components. Experiments were

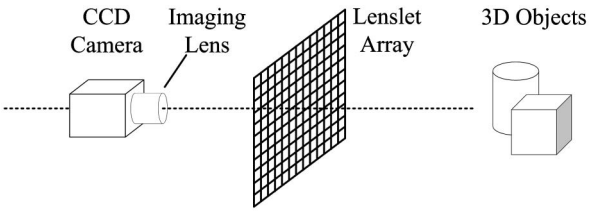

Fig. 1. Configuration of the IIS used in this work.

conducted to reconstruct the positions of an object in a 3D Euclidean coordinate frame.

Although this work focuses on a particular type of IIS, its application can be extended to other IISs with different optical configurations. The generic calibration can accommodate nonstandard distortions and thus allows an IIS to adopt more diversified nontraditional optical designs to achieve the desired properties such as larger field of view or larger depth of field.

2. PINHOLE ARRAY MODEL

CIIR can be performed using a PAM, as depicted in Fig. 2. In a PAM, each EI is formed through a pinhole projection. Then a virtual pinhole array is used to simulate the reverse process of the optical pickup of EIA. Several assumptions are made: (i) the EIs are aligned to form a square array, (ii) the pinhole of each EI is aligned with the EI's image center in the lateral direction, and (iii) all the pinholes are located in a virtual plane parallel to the elemental image plane. Following the PAM, a CIIR method [10,11] can reconstruct an image by projecting each EI through this virtual pinhole array and magnifying it by a factor of A/a, where a is a focal length of the virtual pinhole and A is a distance between the pinhole array and the desired reconstruction plane (see Fig. 2). Then an image is obtained on the reconstruction plane by linear superposition of each inversely mapped and magnified EI. The reconstructed image in this case presents a low depth of field (DOF), and only the 3D objects that are located at the depth of this virtual plane are clearly focused. Another method for CIIR extracts one pixel or a set of pixels from each corresponding EI and combines them into an image [12]. Depending on the selected pixel position, the reconstructed images have different viewing points. The PAM is also implicitly assumed in such a method.

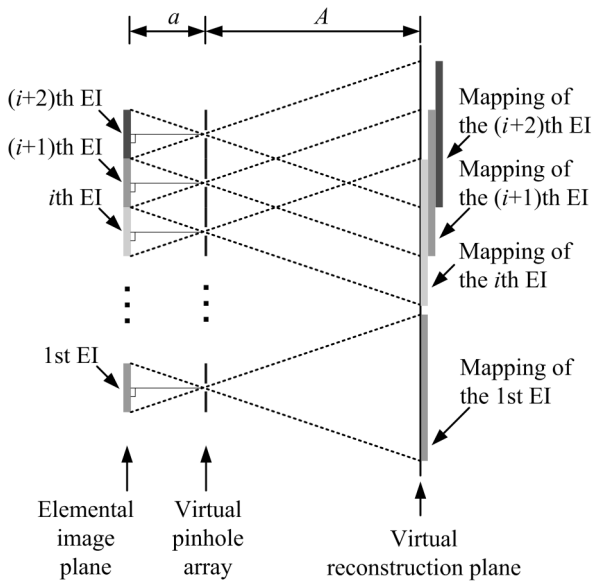

Fig. 2. CIIR by using a PAM.

3. GENERIC CAMERA MODEL

A. Integral Imaging System

Figure 3(a) shows the apparatus of our IIS, which consists of a lenslet array, a 25 mm imaging lens, and a color CCD camera. A standard LCD panel is used as the calibration object. The lenslet array is fabricated by attaching 625 lenslets on the surface of a flat glass panel, as shown in Fig. 3(b). The lenslets are arranged in a 25 by 25 matrix. Each lenslet has a 4 mm by 4 mm square shape with a spherical surface, whose equivalent optical focal distance is 8.95 mm. All the optical axes of the lenslets are parallel, and they are perpendicular to the flat glass panel. The lenslet array and the camera are mounted on a precision translation stage, which allows fixing and adjusting their relative positions. As limited by the fabrication precision, the alignment of the lenslets presents some deficiencies, as shown in the close-up view in Fig. 3(c).

B. Construction of a Generic Camera Model

In the IIS of this work, the light rays captured in each EI are refracted through both the corresponding lenslet and the imaging lens. The misalignment of the lenslets also introduces extra distortions. Instead of constructing a mathematical model

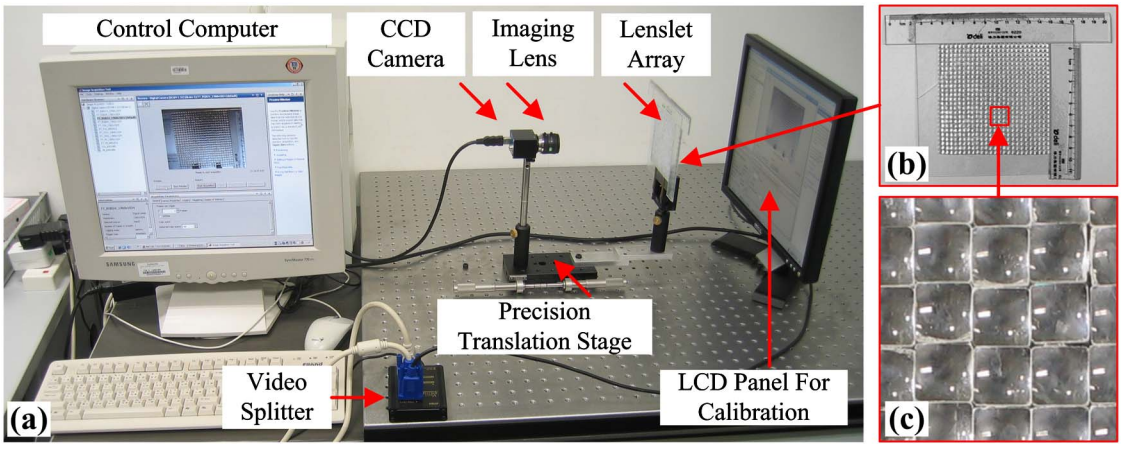

Fig. 3. (Color online) Integral imaging system and the calibration setup.

to explicitly describe all the optical process during image formation, this work focuses on an empirical approach that explores the generic camera calibration concept proposed in the work [17–19]. Compared with the PAM, the generic camera model releases all three assumptions of PAM mentioned in Section 2. This makes it provide more generality to describe the real optical pickup process.

Following the generic camera model concept, the studied IIS is considered as a black box. The input to the IIS is the intensity of light rays in 3D space. This information is commonly modeled as A light field and is encoded in the plenoptic function [20]. The plenoptic function $\Phi(\boldsymbol{q},\boldsymbol{p},t,\lambda)$ gives the intensity of light at each point $\boldsymbol{q}\in\mathbf{R}^3$ in space, from direction $\boldsymbol{p}\in\mathbf{R}^3$, at an instant of time t, and at wavelength λ. By considering still images and CCD sensors with A determined spectrum response, the time and wavelength can be set apart and the plenoptic function becomes a function on $\mathbf{R}^6$, depending only on the position and direction: $\Phi(\boldsymbol{q},\boldsymbol{p})$. In this work, a "two plane parameterization" method is used to represent the plenoptic function. This method parameterizes a light ray by its intersection points with two reference planes. Let φ_1 and φ_2 be the two reference planes and $\boldsymbol{q}_{\varphi1}=(X_{\varphi1},Y_{\varphi1},Z_{\varphi1})$ and $\boldsymbol{q}_{\varphi2}=(X_{\varphi2},Y_{\varphi2},Z_{\varphi2})$ be the two intersection points defined in a 3D Euclidean coordinate X–Y–Z. Then the plenoptic function becomes $\Phi(\boldsymbol{q}_{\varphi1},\boldsymbol{q}_{\varphi2})$. Let the two planes be parallel to the X–Y plane and pass $(0,0,\tilde{Z}_{\varphi1})$ and $(0,0,\tilde{Z}_{\varphi2})$, respectively. Then $\Phi(\boldsymbol{q}_{\varphi1},\boldsymbol{q}_{\varphi2})$ is essentially dependent only on four variables: $X_{\varphi1}$, $Y_{\varphi1}$, $X_{\varphi2}$, and $Y_{\varphi2}$. The cost of this simplification is the inability of describing light rays parallel to φ_1 and φ_2. However, because the field of view of IIS is far less than 180°, $\Phi(\boldsymbol{q}_{\varphi1},\boldsymbol{q}_{\varphi2})$ is sufficient to characterize all the light rays directed into the IIS.

The output of the IIS is the EIA, which is denoted as $E(x,y)$. Here (x,y) is defined in an image plane. As the EIA is a digital image, the pixel location (x,y) takes discrete values. For each pixel location, the intensity value $E(x,y)$ reflects the light energy from a bundle of light rays in space. For computational purposes, this bundle of rays is represented by a chief light ray. Following this, a correspondence is established between each pixel (x,y) and a light ray $(\hat{\boldsymbol{q}}_{\varphi1},\hat{\boldsymbol{q}}_{\varphi2})$, as shown in Fig. 4. Then $E(x,y)$ can be considered as a sampling of the plenoptic function and it can be written as

$$E(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c\bullet\delta(\boldsymbol{q}_{\varphi1}-\hat{\boldsymbol{q}}_{\varphi1},\boldsymbol{q}_{\varphi2}-\hat{\boldsymbol{q}}_{\varphi2})\bullet\Phi(\boldsymbol{q}_{\varphi1},\boldsymbol{q}_{\varphi2})\mathrm{d}\boldsymbol{q}_{\varphi1}\mathrm{d}\boldsymbol{q}_{\varphi2}. \tag{1}$$

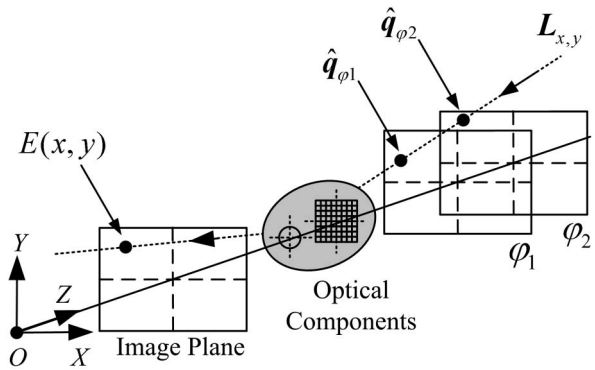

Fig. 4. Illustration of the generic camera model.

Here c is a constant that represents a linear intensity response of the CCD. The Dirac delta function $\delta(\hat{\boldsymbol{q}}_{\varphi1},\hat{\boldsymbol{q}}_{\varphi2})$ equals to one when $\boldsymbol{q}_{\varphi1}=\hat{\boldsymbol{q}}_{\varphi1}$ and $\boldsymbol{q}_{\varphi2}=\hat{\boldsymbol{q}}_{\varphi2}$ and equals to zero elsewhere. It can be seen that in Eq. (1), the correspondence between image pixels and light rays in 3D space is defined by $\delta(\hat{\boldsymbol{q}}_{\varphi1},\hat{\boldsymbol{q}}_{\varphi2})$. As $\hat{\boldsymbol{q}}_{\varphi1}$ and $\hat{\boldsymbol{q}}_{\varphi2}$ are both dependent on (x,y), they can be written as $\hat{\boldsymbol{q}}_{\varphi1}=(\hat{X}_{\varphi1}(x,y),\hat{Y}_{\varphi1}(x,y),\tilde{Z}_{\varphi1})$ and $\hat{\boldsymbol{q}}_{\varphi2}=(\hat{X}_{\varphi2}(x,y),\hat{Y}_{\varphi2}(x,y),\tilde{Z}_{\varphi2})$. Note that $\tilde{Z}_{\varphi1}$ and $\tilde{Z}_{\varphi2}$ are constants that are determined in the world coordinate. Therefore, the geometric model of the IIS is specified by the four mapping functions: $\hat{X}_{\varphi1}(x,y)$, $\hat{Y}_{\varphi1}(x,y)$, $\hat{X}_{\varphi2}(x,y)$, and $\hat{Y}_{\varphi2}(x,y)$.

For the studied IIS, the lenslets are in a square array of M rows and N columns. As the configuration is determined, the lenslet boundaries can be outlined. Besides these boundary pixels, denote $\boldsymbol{e}_{m,n}$ to be the image region of the EI corresponding to the lenslet in row m and column n. In this work, the mapping from image to light field within each lenslet is assumed to be smooth. Following this, the geometric model of the IIS consists of four piecewise smooth mappings $\hat{X}_{\varphi1}(x,y)$, $\hat{Y}_{\varphi1}(x,y)$, $\hat{X}_{\varphi2}(x,y)$, and $\hat{Y}_{\varphi2}(x,y)$, which are defined on the EI regions $(x,y)\in\boldsymbol{e}_{m,n}$, for $m=1,2,\ldots,M$ and $n=1,2,\ldots,N$. In this work, for an EIA image with $R\times C$ pixels, the parameters of the generic camera model are represented by four $R\times C$ matrices. These matrices are stored as lookup tables and accessed at high speed. This work makes no further assumptions for the four mapping functions except for the piecewise smooth property. Therefore, the four mapping functions must be determined via calibration, which will be described in Subsection 4.A.

By using this model, the light ray associated with pixel (x,y) can be expressed as

$$\boldsymbol{L}_{x,y}(r)=\hat{\boldsymbol{q}}_{\varphi1}(x,y)+r(\hat{\boldsymbol{q}}_{\varphi2}(x,y)-\hat{\boldsymbol{q}}_{\varphi1}(x,y)). \tag{2}$$

Here r is a variable that makes $\boldsymbol{L}_{x,y}(r)$ traverse all positions in the light ray.

4. METHODS AND ALGORITHMS

A. Method to Calibrate the Generic Camera Model

In this work, the generic camera model is experimentally determined through a generic calibration method. A calibration setup is used as shown in Fig. 3(a) and illustrated in Fig. 5. A flat LCD panel is used as an active calibration plane. As stated in [21], an LCD panel is a precise planar object whose planarity deviation is no more than 0.05 μm. The LCD panel is placed on the optical table, which is equipped with threaded holes at 25.0 mm square grids. Using these holes, the translation of LCD on the optical table can be precisely controlled and measured. By displaying a bright circular spot on the LCD, a set of points can be highlighted with known positions in 3D space. This provides an accurate source for calibrating the GCM. The proposed method is similar with that used in [18]. In their work, the LCD was translated to two positions in the calibration. To make the estimation more accurate, more positions have been used here. Also, a bright spot pattern is used to encode the plane positions instead of the Gray code [18] to make the method more robust to the nonlinear intensity variations in the EIA.

The calibration method can be summarized in the following steps:

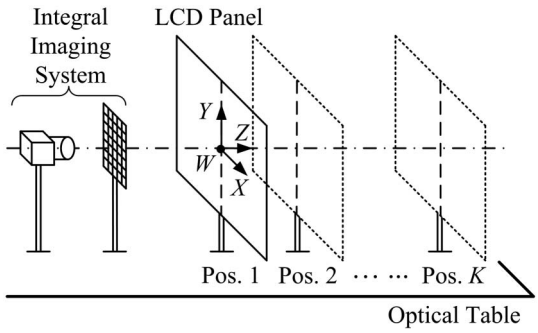

Fig. 5. Illustration of the calibration setup.

1. Place the LCD in front of the IIS at an initial position denoted as position 1 (see Fig. 5). The LCD plane is adjusted to be perpendicular with the optical table and parallel to the lenslet array plane. Define a world coordinate frame X–Y–Z with its X–Y plane parallel to the LCD plane. With the optical table, the LCD's position on the X–Z plane can be precisely controlled. Then, the LCD is translated along the Z axis to K positions, where the Z coordinates are $Z_1, Z_2, \ldots, Z_K$. At each position, steps (2) and (3) are performed.
2. Establish a correspondence between points on the LCD and image pixels in the EIA. Assume that the LCD is at position k with depth Z_k. A point on the LCD is denoted by (X_k, Y_k), and an EIA pixel position is denoted by (x, y). Then the correspondence can be represented by two 2D mappings $X_k(x, y)$ and $Y_k(x, y)$. They are then estimated through a computer controlled procedure as follows:

a. Display a circular bright spot centered at (C_X, C_Y) on the LCD against an all-black screen.
b. Capture an EIA. Because the EIs share overlapping views, the spot will appear in multiple EIs. We apply an image processing algorithm to extract these spots and fit each one to an ellipse [22]. The centers of the ellipses are recorded as a set of pixel locations $\{(x_i, y_i)\}$. Then for each of them, let $X_k(x_i, y_i) = C_X$ and $Y_k(x_i, y_i) = C_Y$.
c. Move (C_X, C_Y) to a new position $(C_X + \Delta X, C_Y + \Delta Y)$ and repeat steps (a) and (b).

Repeat steps (a), (b) and (c) until the LCD screen is traversed by (C_X, C_Y) at intervals ΔX and ΔY. The values of ΔX and ΔY determine the sampling rate on the LCD, and they should be small enough so that a sufficient number of pixels would observe some points on the LCD for at least one time.

3. Interpolate the correspondence in step (2) to be dense. The correspondence established in step (2) is represented as $K \times 2$ mapping functions: $X_k(x, y)$, $Y_k(x, y)$, where $k = 1, 2, \ldots, K$. However, only a subset of pixels in these functions has been assigned values. As the optical property within each EI is considered to be smooth in the GCM, a smooth surface is fit to each function and the mapping functions are interpolated to be dense.
4. For each pixel (x, y) in the EIA, given the mapping functions obtained in step (3), the light ray associated with (x, y) passes a set of K points $\{X_k(x, y), Y_k(x, y), Z_k\}$. Then a straight line in 3D space is fit to the K points with a least-squares minimization (LSM) estimation [23]. This associates a straight line (light ray) to each pixel (x, y) in the EIA.
5. For each pixel (x, y) in the EIA, let the straight line obtained in step (4) intersect with two reference planes φ_1 and φ_2 and assign the coordinates to $\hat{X}_{\varphi 1}(x, y)$, $\hat{Y}_{\varphi 1}(x, y)$, $\hat{X}_{\varphi 2}(x, y)$, and $\hat{Y}_{\varphi 2}(x, y)$ respectively. This provides all the parameters of the GCM.

B. Integral Imaging Using the Generic Camera Model

With Eq. (2), given a pixel (x, y) in the EIA, a light ray $\boldsymbol{L}_{x,y}$ can be determined based on the GCM. Denote all the light rays captured by the IIS as a set $\boldsymbol{S} = \{\boldsymbol{L}_{x,y}\}$. It is known that these light rays intersect within a volume of 3D space. By choosing a viewpoint within this space, a digital image can be computationally reconstructed. For this, a GCM-based CIIR method is introduced, which is referred to as the GCM-CIIR method. In comparison, we refer to the method in [10] as the PAM-based CIIR method (PAM-CIIR). The GCM-CIIR method consists of the following steps:

1. Select a viewpoint $\boldsymbol{v} = (X_v, Y_v, Z_v)$ and determine a subset of light rays that pass a small neighborhood of $\boldsymbol{v}$. Denote this subset of light rays as $\boldsymbol{S}_v \subset \boldsymbol{S}$.
2. For a light ray that satisfies $\boldsymbol{L}_{x,y} \in \boldsymbol{S}_v$, a pixel location in the EIA can be determined as (x, y). Assume that $\boldsymbol{S}_v$ contains a number of W light rays and the associated set of image pixel locations are indicated by $\boldsymbol{R}_v = \{(x^{(m)}, y^{(m)})\}$, where $m = 1, 2, \ldots, W$.
3. Let Ψ be a plane that is parallel to the lenslet array plane and is located at a depth f_Ψ from viewpoint v along the Z axis. By projecting local EI image patches that are centered at positions $\boldsymbol{R}_v = \{(x^{(m)}, y^{(m)})\}$ onto Ψ, an output image I_{output} can be computationally reconstructed. This is performed by the following W iterations. Let $I^{(0)}_{\text{output}}$ be a null image with all white pixels, then from $m = 1$ to $m = W$, repeat the following operation:

$$\begin{aligned} I^{(m)}_{\text{output}}(x^{(m)}_\Psi + \Delta x, y^{(m)}_\Psi + \Delta y) = {} & I^{(m-1)}_{\text{output}}(x^{(m)}_\Psi + \Delta x, y^{(m)}_\Psi + \Delta y) \\ & * (1 - G(\Delta x, \Delta y)) + E(x^{(m)} \\ & + \Delta x, y^{(m)} + \Delta y) \\ & * G(\Delta x, \Delta y). \end{aligned} \tag{3}$$

Here $(x^{(m)}_\Psi, y^{(m)}_\Psi)$ is the image coordinate of the intersection point of the light ray $\boldsymbol{L}_{x^{(m)},y^{(m)}}$ with Ψ. $E(x, y)$ is the EIA image. Pixel coordinate $(\Delta x, \Delta y)$ is defined on a circular region with radius w so that $\sqrt{\Delta x^2 + \Delta y^2} < w$. $G(x, y)$ is a weighting function that is constructed by normalizing the function values of a 2D Gaussian to the interval $[0, 1]$. The final obtained image is $I_{\text{output}} = I^{(W)}_{\text{output}}$.

The above process projects the captured light rays that pass near the selected viewpoint onto a virtual image plane. Therefore, the reconstructed image is an approximation of a perspective image. The resolution of the reconstructed image is determined by a number of factors, such as the light field sampling rate of the lenslet array, the image resolution of the EIs, and the configuration of the image reprojection process. Configuration of the reprojection process includes the selection of the viewpoint, the size of the neighborhood of the viewpoint, and the depth of the virtual image plane. Dependent on the selected viewpoint, a different subset of the captured light

rays are used in image reconstruction. This would lead to different sizes and shapes of a reconstructed image, which is dependent on the specific optical path of the involved light rays. For a fixed viewpoint, the farther the virtual image plane is, the larger the image size is. However, the largest virtual image plane distance is limited by the sampling rate of the light field.

With the image reconstruction method, increasing the object distance would lead to degradation in the reconstructed image quality because some assumptions become less satisfied. In the proposed algorithm, the neighborhood of a selected viewpoint defined in step 1 works as an equivalent aperture. Ideally, if this neighborhood shrinks to a single point and the light field is infinitely dense, the reconstructed image is a perspective image with an infinite depth of field. However, because the EIA is only a discrete sample of the light field, two compromises have been made in the algorithm. First, the neighborhood has a volume to contain a sufficient number of light rays. Second, an image patch centered at each selected pixel is used in the projection to fill the gaps in the virtual image plane. In the optical process, an image patch corresponds to a cone in 3D space. Thus, this cone is used to approximate a light ray. For an object, as its distance increases, a larger part of the object surface would be included in this cone. This makes the approximation to be less appropriate. The consequent image quality degradation appears as the blur of image details. This effect is shown with an experiment in Subsection 5.A.2.

C. 3D Reconstruction Using the Generic Camera Model

Because an object point can be simultaneously observed in a set of EIs from different viewpoints, the position of the point in 3D space can be reconstructed by multiview triangulation. For point $\boldsymbol{q} \in \mathbb{R}^3$, denote the set of its corresponding pixels in the EIA as $\boldsymbol{B}_{\boldsymbol{q}} = \{(x^{(n)}, y^{(n)})\}$, where $n = 1, 2, \ldots, Q$. Then a set of light rays can be determined as $\boldsymbol{S}_{\boldsymbol{q}} = \{\boldsymbol{L}_{x^{(n)},y^{(n)}}\}$, where $n = 1, 2, \ldots, Q$. As the calibrated GCM has already provided the light ray functions for each pixel in the EIA, $\boldsymbol{S}_{\boldsymbol{q}} = \{\boldsymbol{L}_{x^{(n)},y^{(n)}}\}$ is known. Then the 3D position of point $\boldsymbol{q}$ is determined by finding an optimal point $\tilde{\boldsymbol{q}}$ that satisfies the following equation:

$$\tilde{\boldsymbol{q}} = \arg\min_{\boldsymbol{q}} \left(\sum_{(x^{(n)},y^{(n)}) \in \boldsymbol{B}_{\boldsymbol{q}}} D(\boldsymbol{q}, \boldsymbol{L}_{x^{(n)},y^{(n)}}) \right). \tag{4}$$

Here $D(\boldsymbol{q}, \boldsymbol{L}_{x^{(n)},y^{(n)}})$ indicates the Euclidean distance between a point $\boldsymbol{q}$ and a straight line $\boldsymbol{L}_{x^{(n)},y^{(n)}}$. In this work, the optimization problem is approached by a LSM [23]. The proposed 3D reconstruction method is essentially a variant of the multiview stereo method. As the stereo algorithms, the precision of the estimated depth would decrease as the object distance becomes large.

5. EXPERIMENT

In the experiment, the EIA image used has a resolution of 815 pixels by 700 pixels and contains 14 (columns) by 12 (rows) EIs. The size of each EI is 56 pixels by 56 pixels. For the calibration procedure in Subsection 4.A, the LCD was translated to four positions with an interval of 25 mm. The smallest distance between the lenslet array and the LCD panel is 163 mm. The obtained GCM consists of four mapping functions, and they are stored as four 815 by 700 matrices. The algorithms are programmed in MATLAB and executed on a PC with a 2.2 GHz CPU and 2 G RAM. The computational time is about 7 min.

A. Evaluation of the Generic Calibration by Computational Integral Imaging Reconstruction

The accuracy of the calibration is closely related with the quality of the CIIR. CIIR was performed on both 2D and 3D objects to examine the accuracy of the calibration. An image reconstructed by the GCM-CIIR method in Subsection 4.B is referred to as a GCM-CIIR image, and an image reconstructed by the PAM-CIIR method [10] is referred to as a PAM-CIIR image. For comparison, all the reconstructed CIIR images are cropped to contain the same object of interest, and they are normalized to 100 pixels by 75 pixels.

1. Computational Integral Imaging Reconstruction for 2D Planar Patterns

The GCM-CIIR and PAM-CIIR methods were applied to a checkerboard pattern. The pattern is displayed on an LCD panel parallel to the lenslet array, as shown in Fig. 6(a). The EIA image is shown in Fig. 6(b).

Figure 6(c) shows the GCM-CIIR image and Fig. 6(e) shows the PAM-CIIR image. The distortion and misalignment of the lenslets are not accounted by the PAM, which induces artifacts near the edges in the PAM-CIIR image. In Figs. 6(d) and 6(f), the dotted lines represent the edges of the check board and are generated by aligning the checkerboard with that in the reconstructed image. These dotted lines represent the ground truth. The Harris corner detector [24] is often used to automatically locate corners in an image. It is known that the localization precision of Harris corners decreases when the image is blurred. Therefore, the magnitude of deviation of the Harris corners from their true positions reflects the lost of image information. The detected Harris corners in Figs. 6(c) and 6(e) are shown in Figs. 6(d) and 6(f) as crosses. The average deviation is two (pixels) in the GCM-CIIR image and six (pixels) in the PAM-CIIR image. This shows that the GCM-CIIR image provides more spatial accuracy in representing the geometry of a viewed object. Notice that in Fig. 6(d), there are still some alignment errors left between the detected Harris corners and the real corners. These errors can be further reduced by improving the optical component fabrication and the calibration method. Because this experiment focuses on the comparison between the GCM-CIIR image and the PAM-CIIR image, such accuracy improvement is left for a future work.

A resolution test pattern in Fig. 7(a) was used to examine the loss of details in the high spatial frequency component. The GCM-CIIR and PAM-CIIR images are shown in Figs. 7(c) and 7(e). The discrete Fourier transform (DFT) was performed for Figs. 7(a), 7(c), and 7(e). For visualization, the logarithm of the DFT magnitude was scaled to the same gray range as shown in Figs. 7(b), 7(d), and 7(f). The zero-frequency coefficient is displayed in the center. Figure 7(b) shows that the ideal image contains a rich spectrum of high spatial frequency components along different directions. Compared with Fig. 7(f), the spatial frequency components in Fig. 7(d) appear brighter in the high-frequency region (farther from the DFT image center). This shows that the GCM-CIIR image recovers more high spatial frequency components than the PAM-CIIR image.

Furthermore, the visual quality of the reconstructed images was compared quantitatively in terms of the

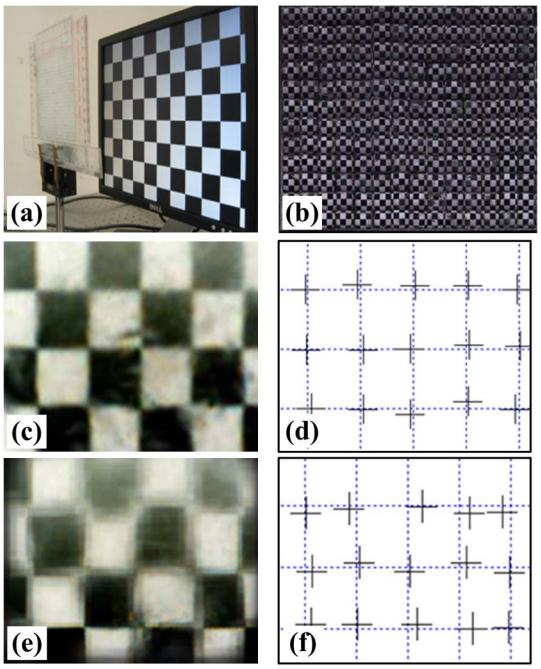

Fig. 6. (Color online) Experiment of CIIR for 2D planar pattern displayed on an LCD: (a) experiment setup, (b) EIA image, (c) GCM-CIIR image, (d) corners in (c), (e) PAM-CIIR image, and (f) corners in (e).

peak-signal-to-noise ratio (PSNR). The PSNR is often used in engineering to compare the visual quality of image reconstruction. A higher PSNR indicates a better approximation of the original image. The original image was aligned with the reconstructed image by image registration. Then the PSNR between them were calculated. The PSNR for the GCM-CIIR image in Fig. 6(c) is 18.60 dB, and the PSNR for the PAM-CIIR image in Fig. 6(e) is 14.43 dB. Also, the PSNR for the GCM-CIIR image in Fig. 7(c) is 16.42 dB, and the PSNR for the PAM-CIIR image in Fig. 7(e) is 13.77 dB. These comparisons show that the GCM-CIIR images are closer to the original images in visual quality.

2. Computational Integral Imaging Reconstruction for 3D Objects

The GCM-CIIR method was tested with a 3D object (a space shuttle model) as shown in Fig. 8(a). The EIA is shown in Fig. 8(b). GCM-CIIR images reconstructed from the same EIA with different viewpoints are shown in Figs. 8(c)–8(h). The viewpoint varies only along the horizontal direction in Figs. 8(c)–8(e). It can be seen that the relative position between the head and the vertical tail of the shuttle varies along the horizontal direction. The viewpoint varies only along the vertical direction in Figs. 8(f)–8(h). It can be seen that the relative position between the front window (black part) and the wings of the shuttle varies along the vertical direction.

Figure 8(i) shows a GCM-CIIR image, and Fig. 8(j) shows a PAM-CIIR image, which are both reconstructed from the same EIA in Fig. 8(b). It can be seen that all features on the space shuttle at different depth are focused in the GCM-CIIR image. In comparison, the PAM-CIIR image presents a narrow depth of field (DOF). Though the head is in focus, the wings and vertical tail are blurred. This demonstrates that the GCM-CIIR method achieves a larger DOF.

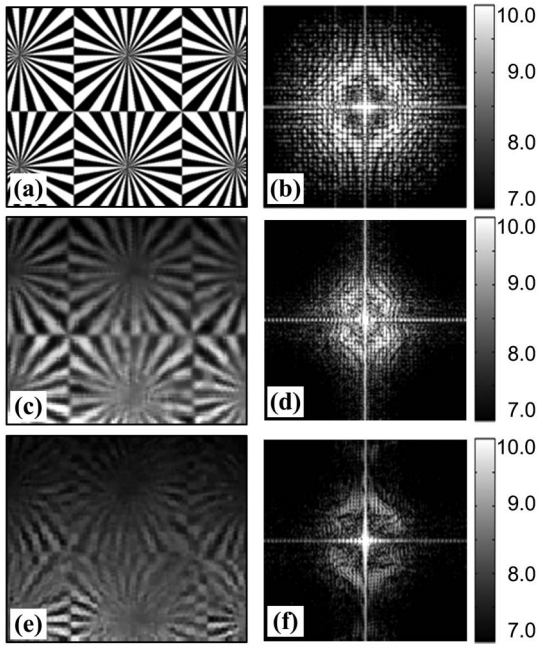

Fig. 7. Comparison of the DFT magnitudes of the GCM-CIIR image and the PAM-CIIR image.

Figure 9 shows the effect of object distance on the GCM-CIIR image. A ping-pong ball was placed at four depths and an image was reconstructed at each position. The diameter

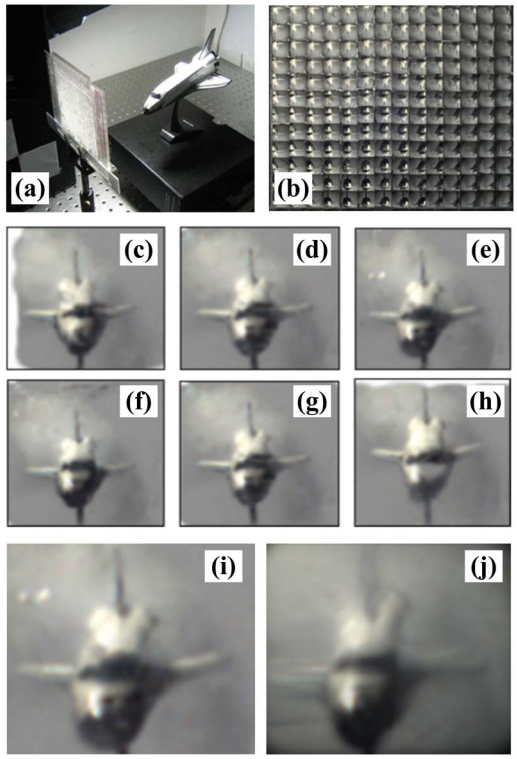

Fig. 8. Experiment of CIIR for a 3D object—a space shuttle model. (a) Experiment setup. (b) EIA image. (c)–(h) GCM-CIIR images from different viewpoints. A GCM-CIIR image (i) is compared with a PAM-CIIR image (j).

of the ball was 40 mm. The distance from the ball to the lenslet array was 275, 325, 375, and 425 mm for Figs. 9(a)–9(d), respectively. All the images are reconstructed with the same parameters. The figures show that the object contour appears more blurred when the distance increases. The blur is more visible at the edge of the black bar that supports the ping-pong ball.

B. Analysis of the Generic Camera Model

The GCM consists of four mapping functions stored in four matrices. For visualization, we retrieve the light ray associated with each pixel in EIA via Eq. (2) and calculate its angle to the Z axis. These angle values are encoded with color and shown in Fig. 10.

Figure 10 shows that the optical property of an EI is different when its location in the EIA varies. The angle distribution of light rays within an EI tends to be symmetric to the EI's center when the EI is near the EIA's center. As the EI moves away from the EIA's center, the angle distribution becomes more asymmetric. This is different from the situation in a PAM, where all the EIs have the same optical properties. For the studied IIS, the angle of view (calculated as the maximum angle spanned by the light rays) of an EI varies from 34.6° (for a lenslet at the center of the EIA) to 32.1° (for a lenslet at the fringe of the EIA). The light rays in an EI can be fit by a viewing cone. The axis of this cone can represent a viewing angle of the EI. The angle between the cone axis and the normal of the glass panel varies from 0° (for a lenslet at the center of the EIA) to 4.6° (for a lenslet at the fringe of the EIA).

Figure 11 quantitatively presents the above difference between the GCM and a PAM. Figure 11(a) shows an EIA taken for a white screen. Against the white screen, the lenslet boundaries appear as the dark pixels between the EIs. In the PAM, the pinholes are the centers of an array of squares on a virtual plane parallel to the image plane. Then the pinhole array can be depicted as an array of squares (outlined by the dotted lines) aligned with the EIA, as shown in Fig. 11(a). The grids are consistent with the lenslet boundaries, though some random misalignments exist due to the fabrication. Figure 11(b) shows the same grids as those in Fig. 11(a). A "+" icon at the center of each square indicates the pinhole. In the PAM, the pixel at this center position is assumed to be associated with a light ray parallel to the Z axis. On the other hand, using the GCM, as shown in Fig. 10, the pixel in each EI whose associated light ray is parallel to the Z axis can be identified and displayed as a "o" icon. An error vector in each EI is shown as a line segment connecting the icons "+" and "o." This error vector reveals the difference between the two models in each EI. In general, all the error vectors are oriented toward the center area of the EIA. As an EI is located farther away from the EIA center, the magnitude of the error vector becomes larger. The maximum error magnitude is 15 pixels. Noticing that the size of each EI is 56 pixels by 56 pixels, this error is significant.

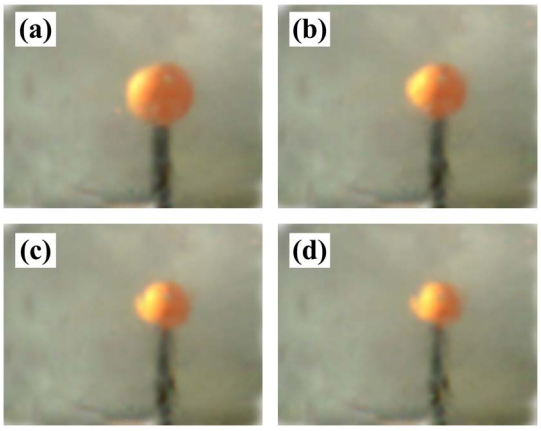

Fig. 9. (Color online) GCM-CIIR with different object distance.

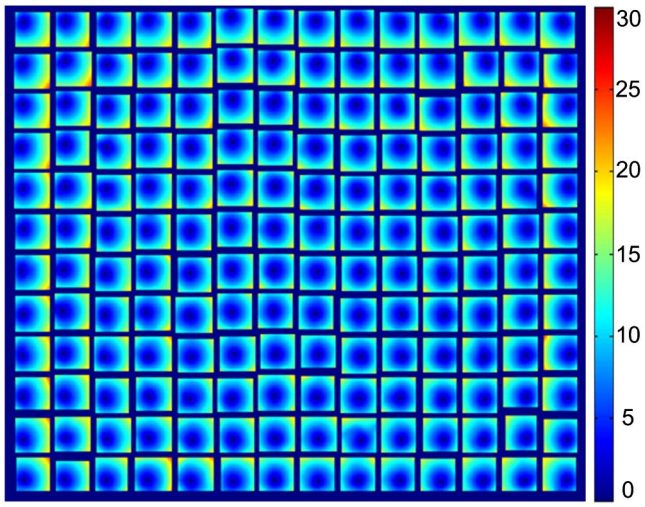

Fig. 10. (Color online) Visualization of the generic camera model.

Figure 3(c) shows some misalignments in the lenslets. These misalignments are one of the sources that make the imaging process different from that described by a PAM. Still, it can be seen that the misalignments alone cannot explain the structural error pattern in Fig. 11(b). This indicates that with a precision fabrication of this lenslet array to eliminate the misalignments, calibration is still necessary.

Compared with traditional camera models, the GCM has a larger number of parameters. However, the GCM provides a complete description of the imaging system's geometric property and a solid basis for building models of more appropriate parameterizations.

C. 3D Reconstruction Experiment

The method in Subsection 4.C was performed to reconstruct the position of a ping-pong ball. The experiment setup is shown in Fig. 12(a) and a top-down view is depicted in Fig. 12(b). A Euclidian world coordinate X_w–Y_w–Z_w is defined in Fig. 12(b). The optical table provides threaded holes arranged in 25 mm grids. By moving the optical holder to these threaded holes, the X_w and Z_w positions of the ball can be measured. By using the pillar holder, the Y_w position of the ball can be measured. This way, the ball can be set to known positions in 3D space. This experiment tested the positions on the two diagonals of a square, which are labeled a1–a9 and b1–b9 in Fig. 12(b). The ball was placed at each position, and the 3D reconstruction was performed. Note that there is a translational motion between the coordinate X_w–Y_w–Z_w and the coordinate X–Y–Z used in calibration. This motion was measured as the LCD's relative position to the optical table. The reconstructed 3D positions were then transformed to the coordinate X_w–Y_w–Z_w to be compared with the ground truth data.

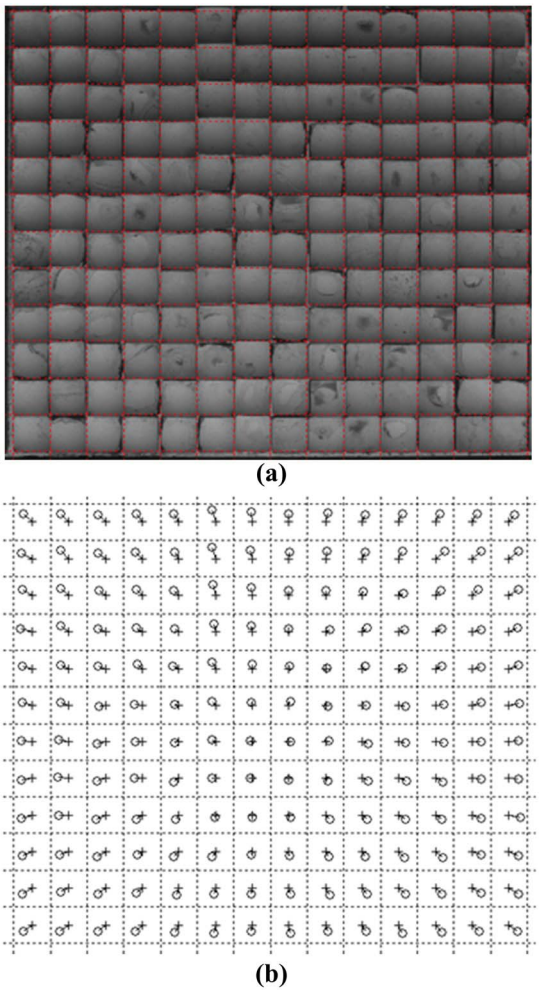

Fig. 11. (Color online) Difference between the GCM and PAM.

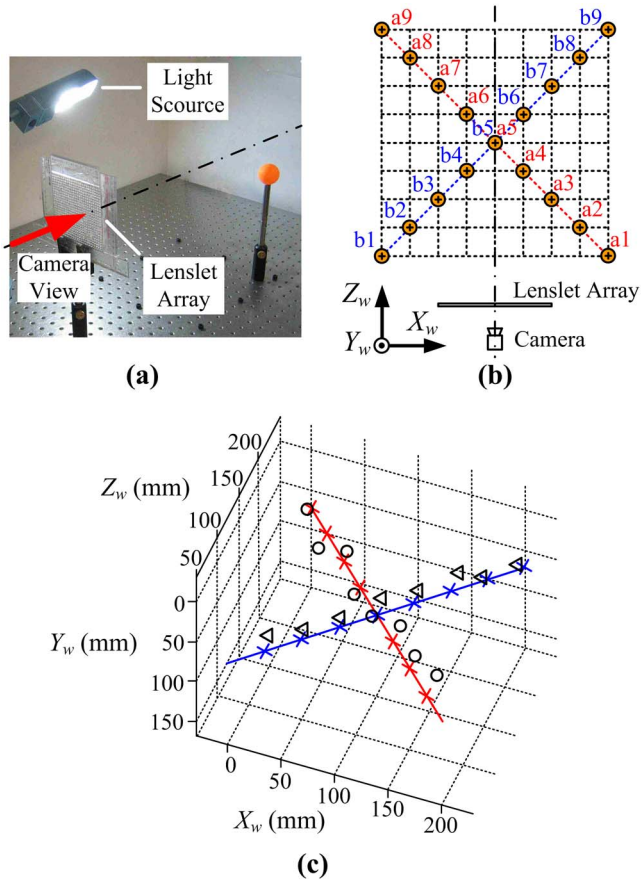

Fig. 12. (Color online) (a) Experiment setup for 3D reconstruction. (b) A top-down view of the setup in (a). A ping-pong ball is placed at the points marked a1–a9 and b1–b9. (c) 3D reconstruction result compared with ground truth.

The image regions of the ping-pong ball were extracted by using a color filter. Some image regions were degraded due to the lenslet defects, and they were removed by image processing. An ellipse fitting algorithm [22] was then used to fit the center of each region. The corresponding pixels of these centers are mapped to their corresponding light rays given in the GCM. Then, using the method in Subsection 4.C, the ball's 3D position was reconstructed.

The positions of the ball in Fig. 12(b) were reconstructed, and the results are shown in Fig. 12(c). In Fig. 12(c), the two diagonals are represented by two lines, and the ground truth data are represented by the crosses on them. The circular icons indicate the reconstructed positions for a2–a9, and the triangular icons indicate the reconstructed positions for b2–b9. Positions a1 and b1 are beyond the effective field of view of the IIS and cannot be reconstructed. In Fig. 12(c), we can see that the reconstructed 3D positions are distributed near the two reference diagonals and are consistent with the ground truth. The reconstructed position (X'_w, Y'_w, Z'_w) for each test position is compared with the ground truth position (X_w, Y_w, Z_w). The majority of the reconstructed X'_w and Y'_w are near their true values. The largest errors in X'_w and Y'_w are within 6 and 2 mm, respectively. Compared with X'_w and Y'_w, the accuracy of depth reconstruction Z'_w is not as high. The largest error in Z'_w reaches 27.1 mm. This error comes from the uncertainty of the image processing algorithms and the noise in the GCM. Further improvement on the reconstruction accuracy may rely on employing more advanced 3D reconstruction algorithms with capabilities to reject noise points by optimization.

6. CONCLUSIONS

In this work, a generic camera model has been investigated to describe and calibrate the optical pickup process of an IIS. Significant differences are found between the generic camera model and a PAM. Based on the generic camera model, CIIR for planar and 3D objects achieves better image quality in terms of reconstruction fidelity and the preservation of high spatial frequency components than the previous method. This shows that the nonstandard configuration in the IIS can be effectively compensated. Furthermore, based on this generic camera model, 3D positions of an object can be reconstructed in a Euclidian coordinate frame.

Future work will include improving the quality of image reconstruction and the precision of 3D reconstruction. The optical constraints enabled by the multiview nature of the integral imaging process can be further explored. By incorporating additional prior information such as the interrelationship between lenslets, parameterization and optimization can be introduced to further improve the calibration in terms of efficiency and accuracy.

ACKNOWLEDGMENTS

This work was supported by a grant from the Research Grants Council of Hong Kong under project CityU117507.

REFERENCES

1. F. Okano, H. Hoshino, J. Arai, M. Yamada, and I. Yuyama, "Three-dimensional television system based on integral photography," in *Three Dimensional Television, Video, and Display Technologies*, B. Javidi and F. Okano, eds. (Springer, 2002), pp. 101–122.
2. A. Stern and B. Javidi, "Three-dimensional image sensing, visualization, and processing using integral imaging," Proc. IEEE **94**, 591–607 (2006).
3. S. Min, Y. Kim, B. Lee, and B. Javidi, "Integral imaging using multiple display devices," in *Three Dimensional Imaging, Visualization, and Display*, B. Javidi, F. Okano, and J. Sun, eds. (Springer, 2009), pp. 41–54.
4. J. Arai, F. Okano, H. Hoshino, and I. Yuyama, "Gradient-index lens-array method based on real time integral photography for three-dimensional images," Appl. Opt. **37**, 2034–2045 (1998).
5. J.-S. Jang and B. Javidi, "Improved viewing resolution of three-dimensional integral imaging by use of nonstationary micro-optics," Opt. Lett. **27**, 324–326 (2002).
6. J.-S. Jang, Y. S. Oh, and B. Javidi, "Spatiotemporally multiplexed integral imaging projector for large-scale high-resolution three-dimensinoal display," Opt. Express **12**, 557–563 (2004).
7. M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Multifacet structure of observed reconstructed integral images," J. Opt. Soc. Am. A **22**, 597–603 (2005).
8. Y. Frauel and B. Javidi, "Digital three-dimensional image correlation by use of computer-reconstructed integral imaging," Appl. Opt. **41**, 5488–5496 (2002).
9. A. Stern and B. Javidi, "Three-dimensional image sensing and reconstruction with time-division multiplexed computational integral imaging," Appl. Opt. **42**, 7036–7042 (2003).
10. S. H. Hong, J.-S. Jang, and B. Javidi, "Three-dimensional volumetric object reconstruction using computational integral imaging," Opt. Express **12**, 483–491 (2004).
11. D. H. Shin and H. Yoo, "Image quality enhancement in 3D computational integral imaging by use of interpolation methods," Opt. Express **15**, 12039–12049 (2007).
12. H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," Opt. Lett. **26**, 157–159 (2001).
13. V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, "Using plane + parallax for calibrating dense camera arrays," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2004), pp. 2–9.
14. R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford technical report CTSR 2005-02 (2005).
15. B. Tavakoli, M. Daneshpanah, B. Javidi, and E. Watson, "Performance of 3D integral imaging with position uncertainty," Opt. Express **15**, 11889–11902 (2007).
16. J. Arai, M. Okui, M. Kobayashi, and F. Okano, "Geometrical effects of positional errors in integral photography," J. Opt. Soc. Am. A **21**, 951–958 (2004).
17. M. D. Grossberg and S. K. Nayar, "A general imaging model and a method for finding its parameters," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2001), pp. 108–115.
18. M. D. Grossberg and S. K. Nayar, "The raxel imaging model and ray-based calibration," Int. J. Comput. Vis. **61**, 119–137 (2005).
19. P. Sturm and S. Ramalingam, "A generic concept for camera calibration," in *Proceedings of the Eighth European Conference on Computer Vision*, T. Pajdla and J. Matas, eds. (Springer, 2004), pp. 1–13.
20. E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds. (MIT, 1991), pp. 3–20.
21. Z. Song and R. Chung, "Use of LCD panel for calibrating structured-light-based range sensing system," IEEE Trans. Instrumen. Meas. **57**, 2623–2630 (2008).
22. J.-N. Ouellet and P. Hebert, "A simple operator for very precise estimation of ellipses," in *Proceedings of the Fourth Canadian Conference on Computer and Robot Vision* (IEEE, 2007), pp. 21–28.
23. R. I. Hartley and A. Zisserman, "Least-squares minimization," in *Multiple View Geometry in Computer Vision* (2nd ed.), R. I. Hartley and A. Zisserman, eds. (Cambridge University, 2003), pp. 588–596.
24. C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the Fourth Alvey Vision Conference*, C. J. Taylor, eds. (AVC, 1988), pp. 147–151.