Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Integral imaging techniques for flexible sensing through image-based reprojection

JOSE M. SOTOCA,[1,*] PEDRO LATORRE-CARMONA,[1] FILIBERTO PLA,[1] XIN SHEN,[2] SATORU KOMATSU,[2] AND BAHRAM JAVIDI[2]

[1]Institute of New Imaging Technologies, Universitat Jaume I, 12071 Castellón de la Plana, Spain
[2]Electrical and Computer Engineering Department, University of Connecticut, Storrs, Connecticut 06269-4157, USA
*Corresponding author: sotoca@uji.es

Received 9 May 2017; revised 7 August 2017; accepted 15 August 2017; posted 16 August 2017 (Doc. ID 295585); published 0 MONTH 0000

In this work, a 3D reconstruction approach for flexible sensing inspired by integral imaging techniques is proposed. This method allows the application of different integral imaging techniques, such as generating a depth map or the reconstruction of images on a certain 3D plane of the scene that were taken with a set of cameras located at unknown and arbitrary positions and orientations. By means of a photo-consistency measure proposed in this work, *all-in-focus* images can also be generated by projecting the points of the 3D plane into the sensor planes of the cameras and thereby capturing the associated RGB values. The proposed method obtains consistent results in real scenes with different surfaces of objects as well as changes in texture and lighting.

OCIS codes: (100.6890) Three-dimensional image processing; (110.3010) Image reconstruction techniques; (110.6880) Three-dimensional image acquisition.

https://doi.org/10.1364/JOSAA.99.099999

1. INTRODUCTION

As opposed to traditional two-dimensional (2D) imaging techniques, three-dimensional (3D) imaging technologies can potentially capture the 3D structure, range, and texture information of the different objects in a scene. Additionally, 3D imaging technologies are more robust to partial scene occlusion. There are many 3D imaging technologies, such as holography and related interferometry techniques [1], stereoscopy [2], pattern illumination techniques [3], LADAR [4], and time-of-flight techniques [5].

Multi-perspective imaging obtains 3D scene information by recording conventional 2D incoherent images from multiple views. Because standard 2D images are used, multi-perspective 3D imaging systems can be built using a single inexpensive camera with a lenslet array or an array of inexpensive sensors. However, thanks to the advances in optoelectronic sensors such as CMOS and CCDs, display devices such as LCDs, and commercially available digital computers, integral imaging is a very active area of research nowadays.

Integral imaging can be considered a class of multi-view imaging acquisition and display technology [6]. It has been applied in fields like visualization [7], target recognition and ranging [8], 3D photon-counting imaging [9,10], 3D imaging for objects under occlusions or in a scattering medium [11], 3D underwater imaging [12], biological or medical imaging [13], integral microscopy [14], and others [15], to cite just a few examples.

Integral imaging performs well under ambient or incoherent light, which compares favorably in relation to other sensing techniques, such as holography, LADAR, or structured light, that make use of an active illumination system. It also has specific benefits over 2D imaging as well as stereo imaging. For 3D visualization purposes in integral imaging, the microlenses produce differences in the light density within the space in front of the observer. Thus, there is a real reconstruction of the light structure produced by the original 3D scene. In lenslet-based integral imaging systems, the achievable resolution is limited by the size of the lenslet and the number of pixels allocated to each lenslet. In essence, the resolution of each elemental image (EI) is limited by three parameters: the pixel size, the lenslet point spread function, and the lenslet depth of focus [7,16]. In addition, aberrations and diffraction are significant because the size of the lenslet is relatively small. In contrast to the lenslet-based systems, integral imaging can be performed either in a synthetic aperture mode or with an array of high-resolution imaging sensors. Each perspective image can be recorded by a full-size CCD or CMOS sensor of several megapixels. This approach may be considered *synthetic aperture integral imaging* (SAII) [17]. SAII enables larger fields of view (FOVs) to be obtained with high resolution 2D images because each 2D image makes full use of the detector array and the optical aperture.

Traditionally, SAII consists of a setup formed by a camera array located on a planar surface. This configuration greatly limits the application of SAII in other situations where an arbitrary arrangement of the cameras is necessary. In Ref. [18], the authors propose the use of a lenslet-based integral imaging system that has an array of lenslets embedded in an elastic scaffold, integrating it into a flexible optoelectronic detector array in an arbitrary non-planar configuration. Recent advances in mechanics and material properties of conventional rigid wafer-based technologies, but with the ability to be stretched and deformed into arbitrary shapes, allow active components to be connected to create new engineering options in imaging devices, where the geometry of the detector array can be optimized together with the lens configuration [19]. The most promising initial possibilities for application are in surveillance, night vision, endoscopy, and retinal implants, or as active components on the eye to enhance vision.

Moreover, authors in Ref. [20] present a 3D integral image acquisition and reconstruction technique with unknown sensor positions and orientations placed on a flexible surface that increases the field of view of the 3D imaging system. In addition, the proposed estimation algorithm assumes that the relative pose of the first two cameras is known. This may not be very convenient if we want to carry out experiments in real-world scenarios without any constraints. Another problem that arises when seeking to solve 3D reconstruction with sensors on a flexible surface is how to obtain a criterion that is robust in this type of problem with an arbitrary camera arrangement. In Ref. [21], a methodology is developed to build a depth map of the scene using a minimum variance approach. Depth estimation accuracy will degrade when object surfaces do not satisfy the Lambertian assumption and requires a precise photometric calibration of the cameras, such as in the presence of partial occlusions or when concave surfaces exist. To address this problem, several proposals for multi-view photo-consistency measures have been developed, such as voxel coloring [22], space carving [23], standard deviation based on an adaptive threshold [24], and voting strategies [25], and in some deformable surface methods [26]. Similarly, the variational formulation relies on square intensity differences [27] or modeling the intensity deviations from brightness constancy by a multivariate Gaussian [28]. Other photo-consistency measures are based on the assumption that a comparison can be made between pairs of images used in stereo, such as normalized cross-correlation (NCC), the sum of squared differences (SSD), mutual information-based measures, and others [29]. Nevertheless, this does not remove any of the severe limitations of the Lambertian assumption.

Integral imaging offers a series of advantages in relation to other 3D imaging techniques. Three of them are: (a) its capability to reconstruct a scene on planes at a constant depth, where only the objects that are at that distance from the camera array are *in focus*; (b) the creation of an *all-in-focus* image from the stack of depth planes; and (c) the ability to infer a depth map of the scene.

The main contribution of this work is oriented toward providing a technique to adapt the reconstruction methodology applied in integral imaging for a *flexible sensing* configuration and show that these same features (i.e., focus on a given depth, creation of an *all-in-focus* image, estimation of the scene depth, etc.) can be obtained in a *flexible sensing* configuration. To that end, a precise calibration of the system is used based on [30], which does not need knowledge of any intrinsic or extrinsic parameter of the cameras setup.

To show the feasibility and accuracy of the proposed 3D plane reconstruction by reprojection, we analyze the problem to obtain a photo-consistency criterion introduced in Section 3 for flexible sensing setups. Thus, we apply the approach proposed in Ref. [31] for light field displays, consisting in a defocusing strategy to deal with spatial information surrounding a pixel. Furthermore, a comparison of this photo-consistency measure will be made with the method based on minimum variance that has been widely used in integral imaging.

Although in Ref. [31], an occlusion method for light fields is also proposed, this is not applicable to the case of an arbitrary flexible sensing setup due to the amount of disparity among the elemental images, which makes the occlusion problem worse than in usual integral imaging setups. In this sense, we have chosen an alternative occlusion method proposed in Ref. [32] and explained in Section 4.

The rest of the paper is organized as follows. Section 2 provides a brief explanation of how the calibration process has been solved in an arbitrary cameras setup. Section 3 explains the methodology proposed in this paper for robust depth estimation. Section 4 describes the creation of the depth map and the *all-in-focus* image estimation of the scene. Section 5 offers the results obtained by applying the techniques proposed here in real scenes and also discusses several aspects. Finally, several conclusions are given in Section 6.

2. MULTI-CAMERA SELF-CALIBRATION

Important advances have been recently made in the reconstruction of 3D scenes from multiple views. In this sense, the review by Ref. [33] and the associated Middlebury evaluation framework represented a milestone after which a lot of research has been conducted focusing on the multi-view reconstruction of objects taken under strictly controlled imaging conditions. However, most of these algorithms are not directly suited to large-scale outdoor scenes.

In a multi-view camera acquisition system, we also need a calibration algorithm that is sufficiently precise to be used for integral imaging techniques that may be able to perform well in outdoor scenes. In this section, we describe a calibration method and camera location for the case where the cameras have an arbitrary pose. The method is based on the work proposed in Ref. [30] for *m*-views using bundle adjustment for a projective reconstruction.

Consider the case where a set of n 3D points $\mathbf{X}_j = [X_j, Y_j, Z_j, 1]^T, j = 1, \ldots, n$ are viewed by a set of m cameras with projection matrices P^i. Denote by $\mathbf{x}_j^i = [x_j^i, y_j^i, 1]^T$ the coordinates of the j-th point as seen in the i-th camera. Our goal is to solve the reconstruction problem where, given a set of image coordinates $\mathbf{x}_j^i$, we aim to find the set of camera matrices P^i and their correspondence points $\mathbf{X}_j$ in the scene such that

$$\mathbf{x}_j^i = P^i \mathbf{X}_j. \tag{1}$$

If the image correspondence measurements has a high number of uncertainties, then Eq. (1) will not be satisfied exactly. Thus, we wish to estimate the projection matrices $\hat{P}^i$ and the 3D points $\hat{\mathbf{X}}_j$ that project exactly onto image points $\hat{\mathbf{x}}_j^i$ as $\hat{\mathbf{x}}_j^i = \hat{P}^i\hat{\mathbf{X}}_j$. Likewise, we also seek to minimize the image distance between the reprojected points and detected (measured) image points $\mathbf{x}_j^i$ for every view in which the 3D point is seen. This approximation (which minimizes the reprojection error) is defined as *bundle adjustment* [34].

Equation (1), representing the projective mapping, can be interpreted as true only up to a constant factor. Writing this constant factor explicitly, we have

$$\lambda_j^i \mathbf{x}_j^i = P^i \mathbf{X}_j. \quad (2)$$

Thus, the goal of the calibration is to estimate the scales λ_j^i and the camera projection matrices P^i. The weighting factors λ_j^i are called the *projective depths* of the points.

For the estimation of λ_j^i we have used Sturm and Triggs' method, exploiting epipolar geometry to obtain these projective depths [35]. In relation to the two alternative methodologies proposed by the authors, the solution based on a central image is more appropriate for wide baseline stereo, and it is the one we used in this section (see Martinec and Pajdla [36] for more details).

Provided that each point is visible in every view, we can put all the points and camera projections into the W_s matrix:

$$W_s = \begin{bmatrix} \lambda_1^1\mathbf{x}_1^1 & \ldots & \lambda_n^1\mathbf{x}_n^1 \\ \vdots & \vdots & \vdots \\ \lambda_1^m\mathbf{x}_1^m & \ldots & \lambda_n^m\mathbf{x}_n^m \end{bmatrix} = \begin{bmatrix} P^1 \\ \vdots \\ P^m \end{bmatrix} [\mathbf{X}_1 \ldots \mathbf{X}_n], \quad (3)$$

where W_s is called the *scaled measurement matrix*, $P = [P^1 \ldots P^m]^T$ and $X = [\mathbf{X}_1 \ldots \mathbf{X}_n]$. P and X are referred to as the *projective motion* and the *projective shape*, respectively [30].

If we collect enough noiseless points (x_j^i, y_j^i) and the scales λ_i^j are known, then W_s can be factored into P and X [35]. The factorization of Eq. (3) retrieves the motion and shape through a 4×4 projective transformation H:

$$W_s = PX = PHH^{-1}X = \hat{P}\hat{X}, \quad (4)$$

where $\hat{P} = PH$ and $\hat{X} = H^{-1}X$. The self-calibration process computes a matrix H, such that $\hat{P}$ and $\hat{X}$ become Euclidean. This process is sometimes called *Euclidean stratification* [30,37]. The matrix H can be solved by imposing certain geometrical constraints. The most general constraint is the assumption that some internal parameters of the cameras are the same.

In Ref. [36], projective reconstruction by factorization is applied, handling perspective views and occlusions jointly. The factorization algorithm also provides an optimal method for computing the new image points when they are not visible from all the cameras. In addition, the method proposed in Ref. [30] fills the missing points (those with unknown depths). This is implemented in two steps: first, triangulation to find the pre-image X, and then the reprojection as PX to generate its image in all views. In practice, triangulation and reprojection provide a method of "filling in" points that are missed during multiple view matching.

Another aspect to be considered is that lenses with short focal lengths are often used in immersive environments to guarantee sufficient field of view. However, such lenses have significant nonlinear distortion, which has to be corrected for precise 3D computation. Therefore, a distortion model is applied to assess the radial and tangential distortion, aiming at eliminating these distortion effects of the lenses in the elemental images obtained by the cameras during the calibration process.

3. PHOTO-CONSISTENCY RECONSTRUCTION BY IMAGE-BASED REPROJECTION

In integral imaging, an optical display or computational reconstruction method can be used to visualize a 3D scene. In the computational reconstruction approach, the elemental images obtained during the acquisition stage are projected onto the image plane at an arbitrary distance through a real pinhole or lens. Because a 3D object can be viewed as the combination of multiple depth images, 3D information can be observed and analyzed by generating a series of depth images.

For a flexible sensing setup, as is our case, we adapt the computational reconstruction used in integral imaging for a regular array of sensors to the case of a non-uniformly distributed flexible sensing integral imaging system, where the camera setup is not placed on a flat surface with known positions in a regular grid. Thus, an alternative strategy is to sweep a set of planes through the scene with respect to a reference camera [see Fig. 1(a)]. This is known as the *plane sweep* algorithm in the computer vision literature [38,39]. Sweeping to a depth D through a series of disparity hypotheses corresponds to mapping each input image into the reference camera defining the disparity space through a series of homographic transformations [38].

We have the projection matrices of the different cameras obtained by the calibration method explained in the previous section. Therefore, the approach presented here is aimed at achieving a depth reconstruction of the objects that are observed from this reference camera c, which we call the *central camera*, and whose projection matrix is defined by $P^c{:}\mathbb{R}^3 \to \mathbb{R}^2$. We denote $I^c{:}\Omega_c \subset \mathbb{R}^2 \to \mathbb{R}^d$ as the intensity of the image acquired by camera c in the set of pixels Ω_c. In practice, the parameter d defines the information stored in the pixels by taking the value $d = 1$ for grayscale images and $d = 3$ for RGB images. Thus,

$$\mathbf{X}_j^c = P_{\pi_Z^c}^{-1} \lambda_Z^c \mathbf{x}_j^c. \quad (5)$$

Let us consider that from the camera c we want to reproject the set of pixels Ω^c onto a 3D plane called *Plane* π_Z^c, which is located at a distance Z with respect to its optical center. Therefore, let us define a reprojection from the camera c onto the plane by $P_{\pi_Z^c}^{-1}{:}P^c(\Omega_c) \to \pi_Z^c$.

During the calibration process, each detected 3D point in the scene [see Eq. (2)] has a scaling factor λ_j^i that is different, and it depends on the depth in relation to the camera. To generate a 3D plane at a certain depth, we must only use a constant scale factor $\lambda_Z^c = Z$. Depending on the applied factor in $[Z_{\min}, Z_{\max}]$ on the image pixels, we can generate 3D points $\mathbf{X}_j^c$ on planes π_Z^c located at different depth ranges, as shown in Fig. 1(a).

Every time we reproject the pixels of the image I^c to a depth level Z, these 3D points can be seen by the rest of the cameras, and, therefore, their positions on their respective images can be

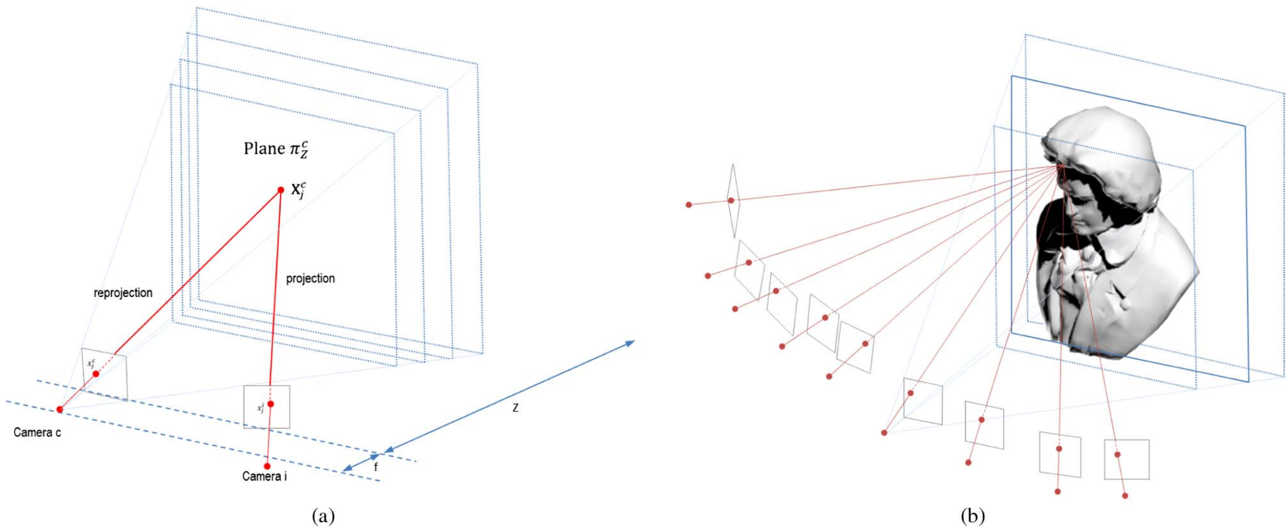


Fig. 1. (a) Reprojection and projection operations with respect to a point $\mathbf{X}_j^c$ on the plane π_Z^c. (b) Camera setup, with arbitrary distribution observing a point in the scene.

estimated. Thus, the value of the image observed by camera i via reprojection of the camera c will be expressed by

$$I^i \circ P^i \circ P_{\pi_Z^c}^{-1} : P^c(\Omega_c) \to \mathbb{R}^d. \quad \textbf{(6)}$$

When the plane is at the depth corresponding to the distance that the object is situated with respect to the optical center of the *center camera* and in the absence of occlusions, we can consider that all cameras will observe the same image value [see Fig. 1(b)], which is the principle of photo-consistency.

A. Photo-Consistency Measure

As indicated in the introduction, an accurate scene depth estimation may be degraded by the shape of the objects or by the intersection of objects with others seen by the different cameras. Hence, the matching process between the different views must handle projective distortion and partial occlusions. The use of local as well as global image intensity information can be exploited to improve the robustness to changes in appearance, without taking into account any approximation of shape, motion, or visibility.

An example of occlusion due to a convex shape can be seen in Fig. 2. We show a scheme with four cameras C_1 to C_4, producing four images of an object with intensities I^1 to I^4. Each 3D point $\mathbf{X}$ projects on the positions $\mathbf{x}_1$ to $\mathbf{x}_4$ of the images in the cameras. Model (object) point $\mathbf{X}$ projects to $\mathbf{x}_2$ and $\mathbf{x}_3$ with intensities I^2 and I^3 but not in $\mathbf{x}_1$ and $\mathbf{x}_4$. Thus, the intensities of I^1 and I^4 are not equal to I^2 and I^3. Needless to say, this is just one of the ways in which occlusion occurs, and other combinations can be produced.

The previous example does not satisfy the conditions of a Lambertian lighting model, where image intensity or color per pixel would be independent of the camera viewpoint. Therefore, the image intensities at pixels $\mathbf{x}_1$ to $\mathbf{x}_4$ should be identical apart from image noise and differences in the camera responses. Let a set of optical images with intensity values be $(I^1, \ldots, I^m)$. Thus, we can project each 3D point of the object $\mathbf{X}_j$ onto the corresponding pixel $\mathbf{x}_j^i$ for each camera i. Then, the arithmetic mean associated to the pixel values of the images corresponding to an object point would be given by:

$$\bar{I}_j = \frac{1}{m}\sum_{i=1}^{m} I^i(\mathbf{x}_j^i). \quad \textbf{(7)}$$

The variance between image intensities and the mean is defined as

$$V_j^2 = \frac{1}{m}\sum_{i=1}^{m} (I^i(\mathbf{x}_j^i) - \bar{I}_j)^2. \quad \textbf{(8)}$$

The variance criterion was one of the first photo-consistency criteria proposed and is also one of the most widely accepted.

Because of the errors in the photo-consistency estimation introduced by occlusions, and taking into account that this fact worsens in an arbitrary flexible sensing setup, the previous arithmetic mean [see Eq. (7)] is computed by applying a visibility criterion for each pixel $O_{occ}^i(\mathbf{x}_j^i, \pi_Z^c)$ that takes values 0 or 1, considering two conditions. The first condition establishes that for those 3D points in the plane π_Z^c that are not visible for the other cameras, the visibility criterion of pixels takes the

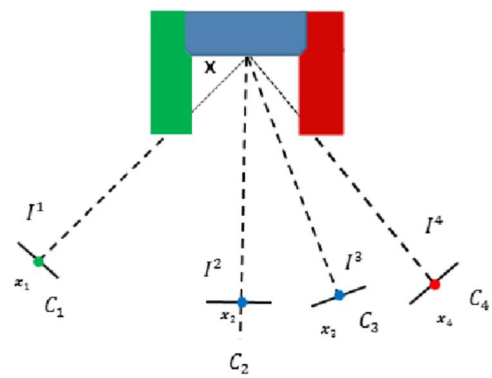


Fig. 2. Projection with occlusions. All the cameras do not see the same point in the scene.

value 0, otherwise it takes 1. As a second condition, we use the asymmetrical occlusion model of Wei and Quan [32] to evaluate the visibility of pixels $O^i_{occ}(\mathbf{x}^i_j, \pi^c_Z)$ for each level of Z. It is defined as being 0 if there exists another pixel $\mathbf{p}^i_j$ which projects onto the same point in camera i as pixel $\mathbf{x}^i_j$ and for which the projected depth is less than that of $\mathbf{x}^i_j$, otherwise it is 1. Therefore, the arithmetic mean associated to the pixel of the *central camera* and its reprojection on the plane π^c_Z is given by:

$$\bar{I}_{\pi^c_Z(j)} = \frac{\sum_{i=1}^m I^i(\mathbf{x}^i_j).O^i_{occ}(\mathbf{x}^i_j, \pi^c_Z)}{\sum_{i=1}^m O^i_{occ}(\mathbf{x}^i_j, \pi^c_Z)}. \tag{9}$$

In addition, the variance between image intensities would be:

$$V^2_{\pi^c_Z(j)} = \frac{\sum_{i=1}^m (I^i(\mathbf{x}^i_j) - \bar{I}_{\pi^c_Z(j)})^2.O^i_{occ}(\mathbf{x}^i_j, \pi^c_Z)}{\sum_{i=1}^m O^i_{occ}(\mathbf{x}^i_j, \pi^c_Z)}. \tag{10}$$

Algorithm 1: Depth map scene and *all-in-focus* image

Procedure Flexible Sensing Integral Imaging by reprojection of 3D planes
  Input:
  Ω^c: set of pixels of *central camera*
  P^i: set of camera projection matrices
  I^i: set of images captured by the cameras
  zstep: distance in depth between two planes π^c_Z
  Output:
  $Z_a = Z_b = Z_{min}$
  while $Z_b <= Z_{max}$ **do** ▷ work with two depths Z_a and Z_b
    $Z_b = Z_b + zstep$
    $\forall$ pixel $\mathbf{x}^c_j$ in Ω^c estimate 3D points $\mathbf{X}^c_{j,Z_a}$ in depth scene by reprojection
    $\forall$ pixel $\mathbf{x}^c_j$ in Ω^c estimate 3D points $\mathbf{X}^c_{j,Z_b}$ in plane $\pi^c_{Z_b}$ by reprojection
    $i = 1$
    while $i <= m$ **do** ▷ m is the number of cameras
      $\forall$ 3D point calculated, project $\mathbf{x}^i_{j,Z_a} = P^i\mathbf{X}^c_{j,Z_a}$
      $\forall$ 3D point in plane $\pi^c_{Z_b}$, project $\mathbf{x}^i_{j,Z_b} = P^i\mathbf{X}^c_{j,Z_b}$
      $\forall$ pixel $\mathbf{x}^i_j$ store the intensities $I^i(\mathbf{x}^i_{j,Z_a})$, $I^i(\mathbf{x}^i_{j,Z_b})$.
      Thus, store the visibility $O^i_{occ}(\mathbf{x}^i_{j,Z_a})$, $O^i_{occ}(\mathbf{x}^i_{j,Z_b})$
      $i = i + 1$
    return
    $\forall$ pixel $\mathbf{x}^c_j$ estimate $\bar{I}_{\pi^c_{Z_a(j)}}$, $\bar{I}_{\pi^c_{Z_b(j)}}$, $V^2_{\pi^c_{Z_a(j)}}$, $V^2_{\pi^c_{Z_b(j)}}$
    If $Z_b = Z_{min} + zstep$
      $\forall$ pixel $\mathbf{x}^c_j$ estimate $\widehat{Photo}_{\pi_{Z_a}}(\mathbf{x}^c_j)$
    EndIf
    $\forall$ pixel $\mathbf{x}^c_j$ estimate $\widehat{Photo}_{\pi_{Z_b}}(\mathbf{x}^c_j)$
    If $\widehat{Photo}_{\pi_{Z_b}}(\mathbf{x}^c_j) < \widehat{Photo}_{\pi_{Z_a}}(\mathbf{x}^c_j)$
      $\widehat{Photo}_{\pi_{Z_a}}(\mathbf{x}^c_j) = \widehat{Photo}_{\pi_{Z_b}}(\mathbf{x}^c_j)$ and $Z_{step}(\mathbf{x}^c_j) = Z_b$
    Else
      $Z_{step}(\mathbf{x}^c_j) = Z_a$
    EndIf
    If $Z_b <= Z_{max}$
      $Z_a \leftarrow Z_{step}$
    EndIf
  return $\bar{I}_{\pi^c_{Z_a}}, Z_a$

A drawback in this strategy is that it is usually applied as a per-pixel photo-consistency measure. To give more robustness to noise and to be able to deal with realistic imaging conditions, pixel neighborhood imaging information should be incorporated.

Authors in Ref. [31] estimate the depth of a scene by combining a *defocus* and a *correspondence* measure. However, they apply it to light fields where the object disparity in the elemental images is small. That is not our case. On the other hand, the defocus measure allows an optimal contrast in a certain region of the image to be obtained, but occlusions and lighting changes may easily affect the measurement accuracy. The patch size may also affect the measure sensitivity because the defocus measure may exceed the patch size. Correspondence measurement allows depth to be estimated using photo-consistency, and it has been widely used in stereo problems. In this case, a statistical measure is usually applied to resolve matching ambiguities.

In our approach, we propose a photo-consistency measure where the first term (correspondence term) defines an initial cost function equal to the square root of the variance $V^2_{\pi^c_Z(j)}$ [see Eq. (10)] in the plane π^c_Z for each point [see Fig. 1(a)]. The second term (defocus term) acts locally and involves the reconstructed mean image intensities defined as $\bar{I}_{\pi^c_Z(j)}$ in relation to the intensities I^c of the *central camera*. Thus, given a 3D point $\mathbf{X}^c_j$ reprojected from pixel $\mathbf{x}^c_j$ of the *central camera* in the plane π^c_Z, therefore,

$$\text{Photo}_{\pi_Z}(\mathbf{x}^c_j) = \left\{\sqrt{V^2_{\pi^c_Z(j)}} + |\bar{I}_{\pi^c_Z(j)} - I^c(\mathbf{x}^c_j)|\right\}. \tag{11}$$

Moreover, we add neighborhood imaging information by applying a bifiltering technique with a spatial mean and a zero mean Gaussian kernel function G_s on the image intensity differences, centered on the current pixel around a window W defined as

$$\widehat{\text{Photo}}_{\pi_Z}(\mathbf{x}^c_j) = \frac{\sum_{\mathbf{p}^c_i \in W} \text{Photo}_{\pi_Z}(\mathbf{p}^c_j).G_s(|I^c(\mathbf{p}^c_i) - I^c(\mathbf{x}^c_i)|)}{\sum_{\mathbf{p}^c_i \in W} G_s(|I^c(\mathbf{p}^c_i) - I^c(\mathbf{x}^c_i)|)}. \tag{12}$$

The idea of using color differences as a range filter to estimate the photo-consistency value is based on the observation that whenever a change of depth edge appears, a color change usually occurs between background objects with respect to foreground objects. This can be useful for comparing neighborhoods that are photo-consistent with others that are not.

Finally, the optimal depth is determined over all planes as

$$\widehat{\text{Photo}}(\mathbf{x}^c_j) = \arg \min_{Z \in [Z_{\min}, Z_{\max}]} \widehat{\text{Photo}}_{\pi_Z}(\mathbf{x}^c_j). \tag{13}$$

4. DEPTH MAP AND ALL-IN-FOCUS RECONSTRUCTION FOR FLEXIBLE SENSING

Algorithm 1 is presented as an example of the application of the image reprojection and photo-consistency criterion on the reconstructed 3D planes proposed in the previous section. In this algorithm, the depth maps and *all-in-focus* images on a certain 3D plane can be estimated. It also shows how a reprojection is performed with two depths Z_a and Z_b. In the case of Z_a, the range of values changes as the algorithm steps forward in depth over the scene with respect to the *central camera*, assigning for each pixel x_j^c the level of depth corresponding to the lowest photo-consistency value $\widehat{\text{Photo}}_{\pi_{Z_a}}(\mathbf{x}_j^c)$ assessed so far. In the case of Z_b, it acts like classical photo-consistency, generating planes at different levels of depth.

Each of the 3D points generated by reprojections $\mathbf{X}_{j,Z_a}^c$ and $\mathbf{X}_{j,Z_b}^c$ in the two proposed depth types is projected for each of the cameras i in the pixels $\mathbf{x}_{j,Z_a}^i$ and $\mathbf{x}_{j,Z_b}^i$ storing the intensities $I^i(\mathbf{x}_{j,Z_a}^i)$ and $I^i(\mathbf{x}_{j,Z_b}^i)$, respectively. Furthermore, the visibility of the projected pixels is assessed.

The main reason for the use of two levels is given by the occlusion algorithm of Wei and Quan [32], which suggests that if there is another pixel p_j^i with depth Z_a that projects to the same point in camera i as pixel $\mathbf{x}_j^i$ and for which the projected depth is less than that of $\mathbf{x}_j^i$ projected in that step with depth Z_b, then the pixel can be occluded. For multiple pixels warped into the same location, only the one with the smallest depth is visible, and it occludes all other projections.

Once the internal loop has finished, we can estimate the arithmetic mean of the intensities associated to the two types of depth estimation, $\overline{I}_{\pi_{Z_a}^c}$ and $\overline{I}_{\pi_{Z_b}^c}$, and their corresponding variances between the image intensities $V_{\pi_{Z_a}^c}^2$ and $V_{\pi_{Z_b}^c}^2$. With this information, we can assess (for each pixel belonging to the *central camera*) the proposed photo-consistency criterion and make a comparison, updating $\widehat{\text{Photo}}_{\pi_{Z_a}}(\mathbf{x}_j^c)$ and its depth Z_a if the photo-consistency criterion obtains a smaller value.

The algorithm satisfies the three specifications that were considered in the Introduction section: (a) establish depth planes where the objects that are at a specific scene depth are *in focus*. This is obtained by means of the arithmetic mean of the images $\overline{I}_{\pi_{Z_b}^c}$. (b) We can create an *all-in-focus* image to form the stack of depth images. This is a final by-product of the algorithm obtained when we have the final depths of the scene and project them over all the cameras. Observe that in Z_a we have stored the depth values with the lowest photo-consistency value reached until that depth plane. At the end of the loop in $Z_b = Z_{\max}$, we have the depth of scene in Z_a, and we can obtain the arithmetic mean of the images $\overline{I}_{\pi_{Z_a}^c}$. (c) A depth map of the scene can be built with the depth stored in Z_a when the algorithm finishes.

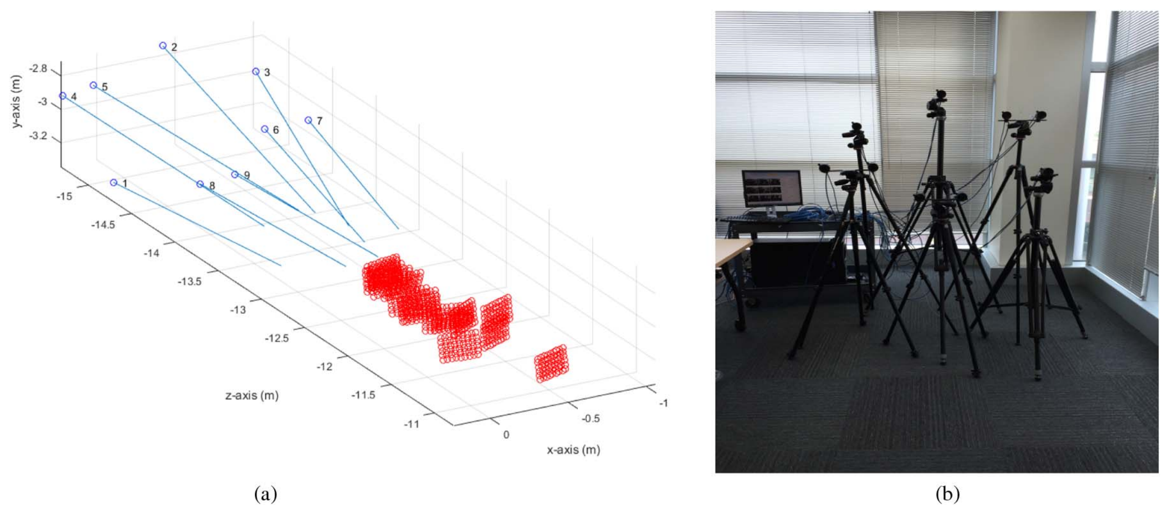
5. RESULTS

In this section, we will show some results obtained from using a flexible sensing setup. To show the capabilities of the proposed method, an experimental arrangement of the cameras, the *in focus* images obtained at different depths, and the depth map of the scene for some examples are described in the following sections.

A. Experimental Setup

Regarding the image acquisition setup, a Norpix camera array consisting of 9 AVT Mako G–192C PoE CMOS cameras (1/1.8″) was used. Camera resolution was 1600 × 1200 pixels. The focal length of the optics used in the experiment was 12.5 mm. The lenses were Ricoh 12.5–75 mm F1.8, manual focus.iris/zoom lens, C-mount, 2/3″ format, w/lock screws. The diagonal FOV was 39.3°. The software used for synchronized capturing was StreamPix6, for multiple camera use. The computer used to manage the entire system had a CPU Intel(R) Core(TM) i7 – 6700 K CPU at 4.0 GHz, and a speed of 2.5 GHz.

Figure 3(b) shows a picture of the experimental setup, including the array of cameras (nine cameras) and the computer used to control them. The spatial arrangement of the different cameras was located at different heights and depths, which produces a variation in the location of the objects and their size as seen by the different cameras. In addition, the cameras were positioned with arbitrary rotation to observe the scene from different points of view. As an example of what the cameras observe in the scene, we show four images (see Fig. 4). Camera number 6 [Fig. 4(c)] acted as the *central camera*, and the depth reconstruction of the scene was performed with respect to this camera. It can be observed that when choosing arbitrary posi-



(a) (b)

Fig. 3. (a) Centers and optical axes for the nine cameras of the setup are illustrated by solid lines in blue. Points in red represent the estimation of 3D points that belong to the movable checkerboard calibration pattern. (b) Image of camera array setup used in the experiments.

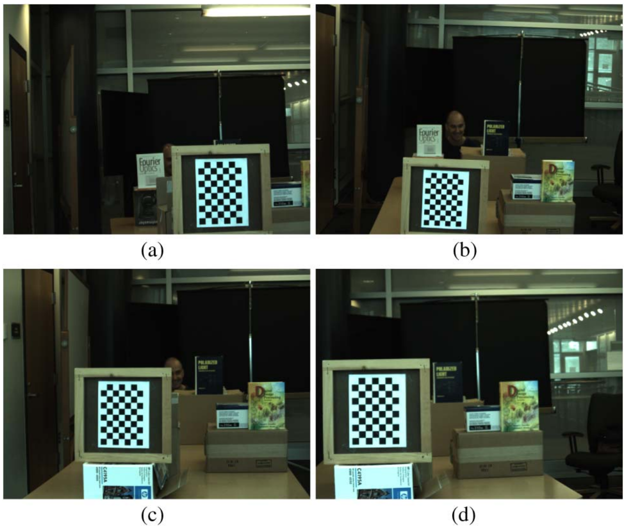


Fig. 4. Elemental images: (a) camera 1, (b) camera 2, (c) camera 6, and (d) camera 9.

tions of objects at different depths in a complex scene, some objects may be seen by some of the cameras and not by others. This camera setup enlarges the common FOV observed by the set of cameras, but it makes depth estimation of the objects in the scene a difficult task.

B. Camera Calibration Process

To solve the calibration problem and therefore estimate the matrix W_s, it is necessary to have a set of image points. Nevertheless, it is possible that the matrix may contain some missing points. Thus, the more complete the matrix is, the more accurate and stable the results achieved from the calibration process will be. Providing a correspondence between a set of scene points that can be directly visible for all cameras is a difficult task. Problems arise because the inaccuracies that are generated in the correspondence between the points seen by the different cameras may affect the calibration accuracy. Only when the objects that appear in the scene have a high level of texture detail is a robust correspondence likely to occur. An example of this type of technique is the use of the scale-invariant feature transform (SIFT), in combination with epipolar geometry and maximum likelihood estimation in the presence of outliers [40].

In other scenes where these objects do not appear, it is possible to apply a moving calibration pattern in order to obtain accurate information about this correspondence [41]. However, the moving calibration pattern poses the same problem as the direct acquisition of scene points in situations where the cameras are far away, i.e., the correspondence points cannot be visible in all the views, and the partially calibrated structures have to be chained together; this procedure is highly prone to errors. In our case, we chose a movable checkerboard calibration pattern, taking into account that the calibration method is able to solve the existence of points that are not observed by all the cameras. In Fig. 3(a), we show the camera centers and optical axes for all cameras estimated during the calibration process. Points in red represent the estimation of 3D points that belong to the movable checkerboard calibration pattern.

C. Focusing Images for Different Depths

Computational reconstruction techniques in integral imaging allow calculation of the image of the scene in a certain plane so that the objects that are at that depth are *in focus*. Therefore, to demonstrate the application of the proposed 3D image plane reprojection and photo-consistency measure calculation for flexible sensing, let us show some examples of depth maps and the *all-in-focus* images obtained from real scenes.

In Figs. 5(a)–5(e), we show five *in-focus* images of the scene estimated from different depth planes. To validate these images, an estimation of the distance of the object from the camera array was obtained using a Laser Distance Measure Model 40-6001 to measure the depth in meters of a set of objects in the scene [see Fig. 5(f)]. When observing the different images, it can be seen how the object *in focus* corresponds to a part of the scene with a sharp image while the rest of the scene is blurred. This is because when the plane is at the depth corresponding to that object, the cameras that see that object have the same distribution of intensities, and the object is photo-consistent at that depth.

In these demanding real experimental conditions, there is no ground-truth available. Besides, the depth values that are given in Fig. 5(f) are depth values taken with a laser measure system that points only to a part of the object, from a position close to the reference camera. Therefore, we cannot assign a depth to a complete object, and then these depth measures cannot be considered as measures for the object as a whole. Nevertheless, we confirmed that the reconstruction where the objects were in focus was the depth given by the laser measure system.

The application of a flexible sensing setup with cameras at arbitrary positions and different relative rotations, and therefore with different directions for their optical axes, has both positive and negative effects. On the one hand, the positive effect is that by amplifying the common FOV observed by the set of cameras, the size of the 3D plane that all the cameras observe can be expanded in a larger region of the scene forming a common mosaic. However, in this work, given that we have used a small number of cameras for the experiments, we have chosen a conservative configuration and limited the number of reprojected points in the reconstructed image planes to the number of pixels the *central camera* has.

On the other hand, the downside effect is related to an effect that appears in the 3D reconstruction process and that consists in the tendency of the objects that are close to the cameras, once they are *in focus*, to expand their corresponding boundaries in the scene for images *in focus* at higher depths. This expansion effect has been observed in experiments performed in classical integral imaging from an array of cameras with parallel optical axes and varies depending on the value of the FOV and the distance between the cameras. This effect can be amplified when a flexible arrangement of cameras is used, since the variation of where a close object is located by the different cameras is bigger. As an example, the position of the checkerboard shown in Figs. 5(c)–5(e) occupies an increasingly larger area, thus

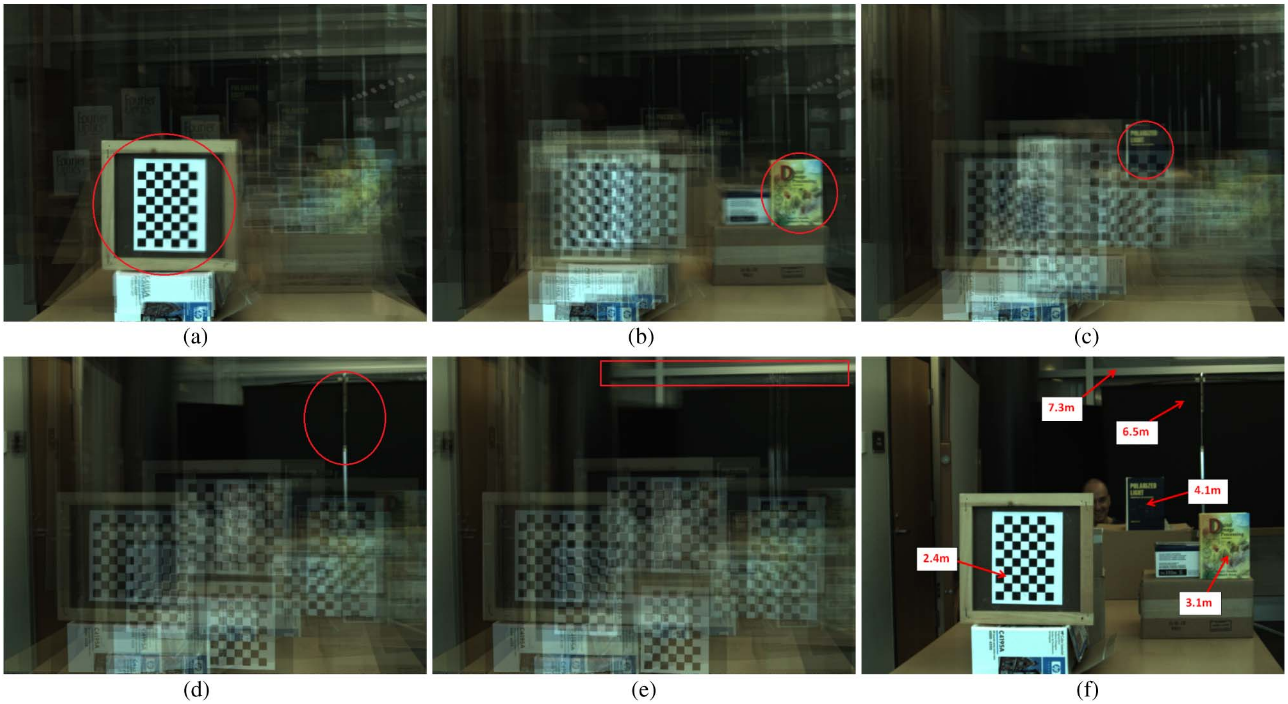


Fig. 5. (a)–(e) Reconstruction at the depth where the following objects are in focus: (a) the checkerboard; (b) first book; (c) second book; (d) the rear side of a projection screen; (e) the wall at the end of the room; (f) measured depths with the laser rangefinder.

interfering with objects that, for greater depths, should be seen sharply.

D. Depth Map and All-in-Focus Image

The two other aspects addressed in this work to show how integral imaging techniques can be extended to the flexible sensing approach are: the generation of an *all-in-focus* image and its corresponding depth map.

To analyze the visual quality of the proposed photo-consistency criterion, the results have been compared with those obtained by the *Min-Var* method [21]. To do so, three scenes have been analyzed, where the first one corresponds to the previous example of the scene that shows different objects on a table, and the other two scenes basically consist of a person making a gesture. Thus, two people are shown, focusing particularly on the reconstruction of the hand gesture and body of the two subjects.

Figure 6 shows the results of these three scenes. The first row shows the elemental images for the *central camera*. The second row shows the results of the *all-in-focus* images as a function of the depth estimation methodology proposed in this work. The third row shows the depth map results obtained by the *Min-Var* method, and the fourth row shows the depth map obtained by the photo-consistency criterion used in this work.

The generation of an *all-in-focus* image has a strong dependence on the photo-consistency criterion used because the imprecisions that are generated during the reconstruction process affect the degree of accuracy reached in the depth map. In the same way, because the calculation of *all-in-focus* images is a by-product of the depth map calculation, depth errors appear as artifacts in the *all-in-focus* images. When comparing the elemental images (first row) with their corresponding *all-in-focus* images (second row), we may observe how some artifacts and noise appear in certain regions of the scene that do not allow the reconstruction to be clearly visualized. This noise occurs mainly in regions that are further away from the camera setup and close to objects that were in focus at a certain depth and have become defocused at greater depths.

When we analyze the visual accuracy of the depth maps obtained by the *Min-Var* method (see depth maps in row three of Fig. 6), we must take into account that this method is based on a pixel-by-pixel variance of the RGB values obtained for the elemental images of each camera. We can see how this method is influenced by two circumstances: the first is given by the variance of intensities (called the "correspondence term" in this paper) between the cameras, and it is strongly influenced by the expansion effect we have previously mentioned. In addition, the occlusion between objects may make it difficult to find a precise correspondence of what the different cameras can observe. The second circumstance is given by the pixel-by-pixel measurement that impedes the analysis of neighboring pixels. This produces a very noisy depth map on the object surfaces. Furthermore, the photo-consistency measure obtained is more sensitive in the case of objects containing textured surfaces that generate visual irregularity in the objects.

In our work, we have also added a defocus term in order to measure the optimum contrast of one region of the scene that is focused at a specific depth. As with the correspondence term, occlusions or differences in pixel color distribution are related to the point of view of the scene of each camera. This fact may mean that object focusing can only be partially obtained, thereby degrading the performance of this measure.

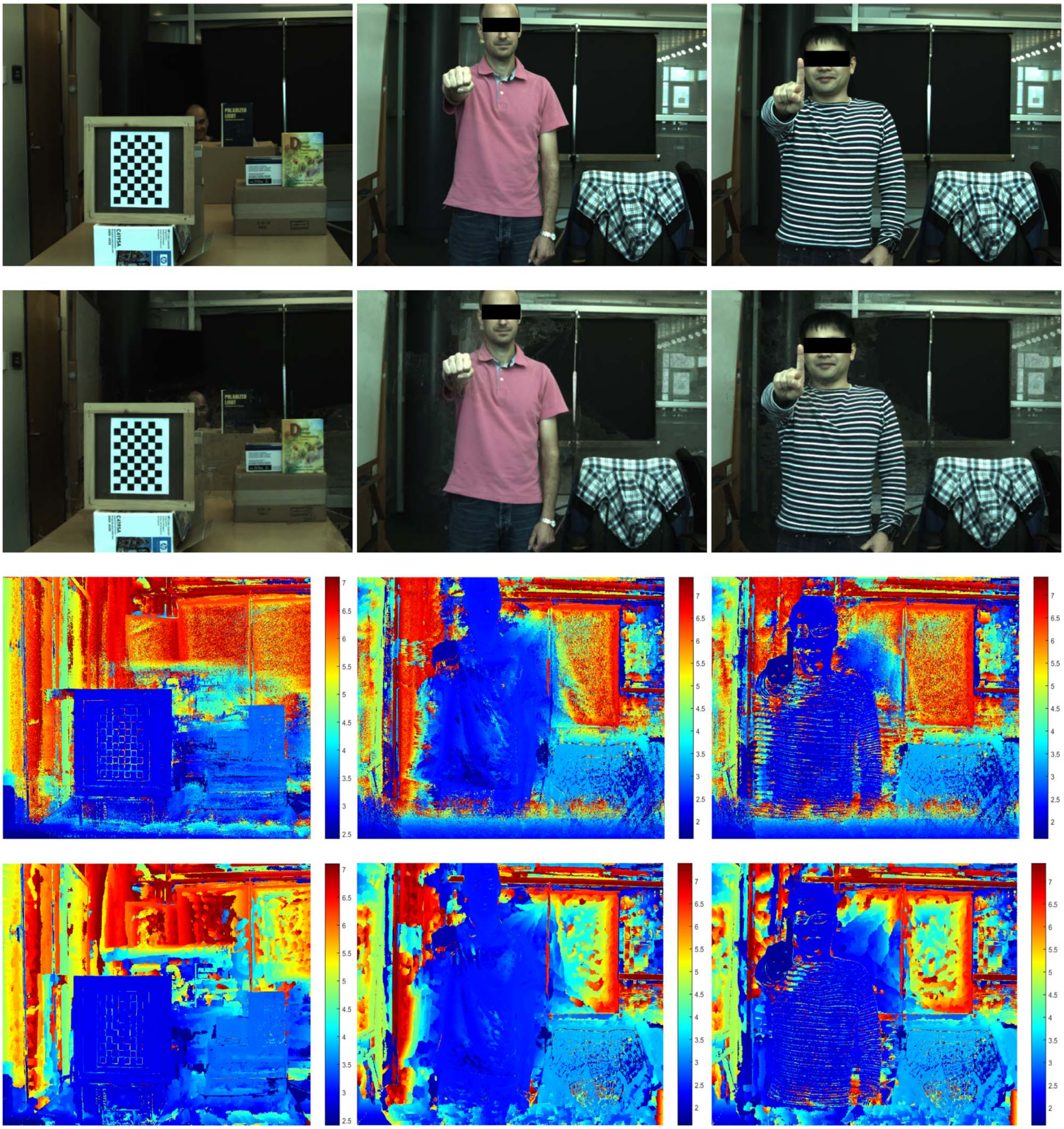


Fig. 6. *All-in-focus* images and depth maps results. From top to bottom rows, the elemental images of the *central camera*, results of the *all-in-focus* images, depth map results obtained by the *Min-Var* method, and depth map results obtained by the photo-consistency measure used in this work.

An example of this type of situation is Fig. 5(c), where objects near the checkerboard are severely affected (look at the book inside the circle). Another problem that this measure presents is the accuracy of the depth because the same object can be in focus in an interval range of depths, especially if the object has little texture.

When analyzing the visual results of the depth maps obtained with the photo-consistency measure used in this work (see depth maps in row four of Fig. 6), we can observe that the use of a bifiltering strategy based on a mean spatial filter and a Gaussian kernel function for the intensity differences allows us to obtain results with a more homogeneous estimation of the object surface depth. In this case, the spatial kernel is centered at each individual pixel around a window W. This allows the cameras to be matched, while also adding neighborhood imaging information. Notice from the results that when applying the bifiltering process, the depth map contains lower noise, with smoother depth areas as a final result.

In general terms, our method is more stable in fixing the correct depth of the objects since it takes into account information of the neighboring pixels. Nevertheless, if depth estimation is not correct, it not only affects a particular pixel, but it also affects the pixels in its neighborhood. For instance, we can see in rows 3 and 4 in Fig. 6 that the depth estimation for the checkerboard and the books is more robust in our case than for the *Min-Var* method. However, the depth estimation in the black projection wall is worse in our case.

6. CONCLUSIONS

The present work has proposed a 3D reconstruction approach based on integral imaging for a flexible sensing configuration of the cameras. It considers that the scene is observed from noticeably different points of view in such a way that the regions of the scene perceived by the cameras are difficult to match. The method is based on the reprojection into 3D planes at different depths that are orthogonal to the optical axis of a reference camera called the *central camera*.

To carry out the reconstruction of the scene, a photo-consistency measure combining a *defocus* and *correspondence* measure has been proposed. In addition, to add information from neighboring pixels, a bifiltering approach based on a mean spatial filter and a Gaussian kernel function for the intensity differences is applied. Based on the applied 3D plane reconstruction and photo-consistency criterion, it has been shown how some properties from integral imaging can be adapted to this scenario. In particular, a depth estimation and an *all-in-focus* image estimation algorithm are described to show how they can be performed in a free sensing setup. Experimental results show the feasibility of the proposed method and the level of accuracy obtained despite the fact that the errors produced by occlusions worsen in a free sensing setup. To tackle this problem, an accurate multi-camera calibration method and the 3D image plane reprojection approach are essential to obtain satisfactory results.

The results obtained are generally consistent in real scenes with different types of surfaces, although objects with a smooth texture or changes due to brightness can affect the result. A downside effect for objects close to the cameras is that, once they are in focus, they tend to expand when reconstructing planes through the scene at larger depths. This effect is amplified in the case of a flexible sensing configuration where the optical axes are not parallel. Future work will be aimed at improving the precision and visual quality of the generated depth map, and also at incorporating other aspects such as depth map regularization strategies, in an attempt to obtain smoother depth maps inside homogeneous surface objects and sharp estimations at depth discontinuities.

Funding. Generalitat Valenciana (PROMETEO-II/2014/062); Spanish Ministry of Economy and Competitiveness (MINECO) (ESP2013-48458-C4-3-P, MTM2013-48371-C2-2-P); University Jaume I (UJIP11B2014-09); National Science Foundation (NSF) (NSF/IIS-1422179); Office of Naval Research (ONR) (N00014-17-1-2561).

Acknowledgment. B. Javidi acknowledges support from NSF and ONR.

REFERENCES

1. S. Benton and M. Bove, *Holographic Imaging* (Wiley Interscience, 2008).
2. A. P. Sokolov, "Autostereoscopy and integral photography by professor Lippmann's method," in *Izd. MGU* (Moscow State University, 1911).
3. D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2003), Vol. **1**, pp. 195–202.
4. M. Hebert and E. Krotkov, "3d measurements from imaging laser radars: how good are they?" Image Vis. Comput. **10**, 170–178 (1992).
5. A. Bhandari and R. Raskar, "Signal processing for time-of-flight imaging sensors: an introduction to inverse problems in computational 3-d imaging," IEEE Signal Process. Mag. **33**(5), 45–58 (2016).
6. 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," Proc. IEEE **105**, 825–836 (2017).
7. F. Okano, J. Akai, K. Mitani, and M. Okui, "Real-time integral imaging based on extremely high resolution video system," Proc. IEEE **94**, 490–501 (2006).
8. J. H. Park and K. M. Jeong, "Frequency domain depth filtering of integral imaging," Opt. Express **19**, 18729–18741 (2011).
9. M. DaneshPanah, B. Javidi, and E. A. Watson, "Three dimensional object recognition with photon counting imagery in the presence of noise," Opt. Express **18**, 26450–26460 (2010).
10. D. Aloni, A. Stern, and B. Javidi, "Three-dimensional photon counting integral imaging reconstruction using penalized maximum likelihood expectation maximization," Opt. Express **19**, 19681–19687 (2011).
11. I. Moon and B. Javidi, "Three-dimensional visualization of objects in scattering medium by use of computational integral imaging," Opt. Express **16**, 13080–13089 (2008).
12. R. Schulein, C. M. Do, and B. Javidi, "Distortion-tolerant 3d recognition of underwater objects using neural networks," J. Opt. Soc. Am. A **27**, 461–468 (2010).
13. B. Javidi, I. Moon, and S. Yeom, "Three-dimensional identification of biological microorganism using integral imaging," Opt. Express **14**, 12096–12108 (2006).
14. A. Llavador, J. Sola-Pikabea, G. Saavedra, B. Javidi, and M. Martínez-Corral, "Resolution improvements in integral microscopy with Fourier plane recording," Opt. Express **24**, 20792–20798 (2016).
15. Y. Zhao, X. Xiao, M. Cho, and B. Javidi, "Tracking of multiple objects in unknown background using Bayesian estimation in 3d space," J. Opt. Soc. Am. A **28**, 1935–1940 (2011).
16. A. Stern and B. Javidi, "Three-dimensional synthetic aperture integral imaging," Proc. IEEE **94**, 591–607 (2006).
17. J. S. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," Opt. Lett. **27**, 1144–1146 (2002).
18. M. Daneshpanah and B. Javidi, "Three-dimensional imaging with detector arrays on arbitrary shaped surfaces," Opt. Lett. **36**, 600–602 (2011).
19. J. A. Rogers, T. Someya, and Y. Huang, "Materials and mechanics for stretchable electronics," Science **327**, 1603–1607 (2010).
20. J. Wang, X. Xiao, and B. Javidi, "Three-dimensional integral imaging with flexible sensing," Opt. Lett. **39**, 6855–6858 (2014).
21. M. Daneshpanah and B. Javidi, "Profilometry and optical slicing by passive three-dimensional imaging," Opt. Lett. **34**, 1105–1107 (2009).
22. S. Seitz and C. Dyer, "Photorealistic scene reconstruction by voxel coloring," Int. J. Comput. Vis. **35**, 151–173 (1999).
23. K. Kurulakos and S. Seitz, "A theory of shape by space carving," Int. J. Comput. Vis. **38**, 199–218 (2000).
24. G. Slabaugh, W. Culbertson, T. Malzbender, M. Stevens, and R. Schafer, "Methods for volumetric reconstruction of visual scenes," Int. J. Comput. Vis. **57**, 179–199 (2004).
25. A. Martínez-Uso, P. Latorre-Carmona, J. M. Sotoca, F. Pla, and B. Javidi, "Depth estimation in integral imaging based on a maximum voting strategy," IEEE J. Display Technol. (to be published).
26. M. Lhuillier and L. Quan, "Surface reconstruction by integrating 3d and 2d data of multiple views," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2003), Vol. **2**, 1313–1320.

27. C. Strecha, T. Tuytelaars, and L. V. Gool, “Dense matching of multiple wide-baseline views,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2003), Vol. **2**, 1194–1201.
28. C. Strecha, R. Fransens, and L. V. Gool, “Wide-baseline stereo from multiple views: a probabilistic account,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2004), Vol. **2**, 552–559.
29. Y. Furukawa and C. Hernandez, “Multi-view stereo: a tutorial,” Found. Trends Comput. Graph. Vis. **9**, 1–148 (2015).
30. T. Svoboda, D. Martinec, and T. Pajdla, “A convenient multi-camera self-calibration for virtual environments,” Presence Teleop. Virt. Environ. **14**, 407–422 (2005).
31. T.-C. Wang, A. A. Efros, and R. Ramamoorthi, “Depth estimation with occlusion modeling using light-field cameras,” IEEE Trans. Pattern Anal. Mach. Intell. **38**, 2170–2181 (2016).
32. Y. Wei and L. Quan, “Asymmetrical occlusion handling using graph cut for multi-view stereo,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2005), Vol. **2**, pp. 902–909.
33. S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2006), Vol. **1**, pp. 519–526.
34. R. Harley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University, 2000).
35. P. Sturm and B. Triggs, “A factorization based algorithm for multi-image projective structure and motion,” in *European Conference on Computer Vision* (Springer-Verlag, 1996), pp. 709–720.
36. D. Martinec and T. Pajdla, “Structure from many perspective images with occlusions,” in *European Conference on Computer Vision*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, eds. (Springer-Verlag, 2002).
37. G. Wang and Q. M. J. Wu, *Guide to Three Dimensional Structure and Motion Factorization* (Springer-Verlag, 2011).
38. R. Szeliski and P. Golland, “Stereo matching with transparency and matting,” Int. J. Comput. Vis. **32**, 45–61 (1999).
39. H. Saito and T. Kanade, “Shape reconstruction in projective grid space from large number of images,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1999), pp. 49–54.
40. P. H. S. Torr and A. Zisserman, “MLESAC: a new robust estimator with application to estimating image geometry,” Comput. Vis. Image Underst. **78**, 138–156 (2000).
41. Z. Zhang, “A flexible new technique for camera calibration,” IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).

Queries

1. AU: Please check my edits here: “(c) the ability to infer a depth map of the scene.” The original was “(c) to infer a depth map of the scene,” which didn’t quite fit with the grammatical structure of (a) and (b).

2. AU: Please check my edit in the sentence beginning, “Thus, we wish to estimate the projection matrices…” I changed “which” to “that.”

3. AU: Please check my edits in the sentence beginning, “Each of the 3D points generated…”

4. AU: Please provide value (8") in SI unit instead of “inches.”

5. AU: Please check my edit in the sentence beginning, “ The computer used to manage the entire system…”

6. AU: Please check my edits in the sentence beginning, “Only when the objects that appear in…”

7. AU: Please check my edits in the sentence beginning, “Therefore, to demonstrate the application…”

8. AU: Does *Min-Var* have to be capped and italicized, or could it be min-var? (I didn’t change it.)

9. AU: The funding information for this article has been generated using the information you provided to OSA at the time of article submission. Please check it carefully. If any information needs to be corrected or added, please provide the full name of the funding organization/institution as provided in the CrossRef Open Funder Registry (http://www.crossref.org/fundingdata/registry.html).

10. AU: Is updated information avaiable for Ref. [25]?