IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 5, 2018, accepted December 7, 2018, date of publication December 11, 2018,
date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2886235

Depth and All-in-Focus Image Estimation in Synthetic Aperture Integral Imaging Under Partial Occlusions

JOSÉ MARTÍNEZ SOTOCA[1], PEDRO LATORRE-CARMONA[1], FILIBERTO PLA[1], AND BAHRAM JAVIDI[2], (Fellow, IEEE)

[1]Institute of New Imaging Technologies, Universitat Jaume I, 12071 Castellon de la Plana, Spain
[2]Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-4157 USA

Corresponding author: Bahram Javidi (bahram.javidi@uconn.edu)

This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO) under Project SEOSAT ESP2013-48458-C4-3-P and Project MTM2013-48371-C2-2-PDGI, in part by the Generalitat Valenciana under Project PROMETEO-II-2014-062, and in part by the University Jaume I under Project UJIP11B2014-09. The work of B. Javidi was supported in part by the Office of Naval Research under Grant N000141712561 and Grant N000141712405, and in part by the Air Force Office of Scientific Research under Grant FA9550-18-1-0338.

ABSTRACT A common assumption in the integral imaging reconstruction is that a pixel will be photo-consistent if all viewpoints observed by the different cameras converge at a single point when focusing at the proper depth. However, the presence of occlusions between objects in the scene prevents this from being fulfilled. In this paper, a novel depth and all-in focus image estimation method is presented, based on a photo-consistency measure that uses the median criterion in relation to the elemental images. The interest of this approach is to find a solution to detect which camera correctly sees the partially occluded object at a certain depth and allows for a precise solution to the object depth. In addition, a robust solution is proposed to detect the boundary limits between partially occluded objects, which are subsequently used during the regularization depth estimation process. The experimental results show that the proposed method outperforms other state-of-the-art depth estimation methods in a synthetic aperture integral imaging framework.

INDEX TERMS Synthetic aperture integral imaging, depth map estimation, all-in-focus image, partial occlusions, 3D image processing.

I. INTRODUCTION

Three-dimensional (3D) optical image sensing and visualisation technologies are currently applied in areas like medical sciences, synthetic aperture radar in remote sensing, entertainment devices or robotics [1]–[5]. One of the most promising 3D approaches is based on Integral Imaging (II) [6]–[15]. II is an autostereoscopic imaging method that works under incoherent or ambient light. This is considerably helpful when compared to other sensing techniques (i.e holography, Ladar), which require an active illumination system [16]. II might be used to infer the three dimensional profile and the range of objects in a scene [17]. 3D sensing with an II architecture has specific benefits in some applications such as segmentation of objects from heavy background, and imaging through obscuration and scattering media (see e.g. [18], [19] for details).

In lenslet-based Integral Imaging systems, the achievable resolution is restricted 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 [20]. 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. This approach may be considered as Synthetic Aperture Integral Imaging (*SAII*) [21]. *SAII* enables larger fields of view (*FOV*) to be obtained with high resolution 2D images because each 2D image makes full use of the detector array and the optical aperture.

Several works [22]–[24] have tried to tackle two of the problems that affect most the quality in a 3D reconstruction scenario (and, in particular, in *SAII*). The first one is to define a robust photo-consistency measure that detects when the

surface of an object is *in focus*. The other problem comes from the existence of occlusions, which is a strong drawback to obtain accurate depth map estimations at objects' depth discontinuities. In particular, this problem comes from the fact that once an object is *in focus*, the blurring of the object can make it more difficult to estimate objects at higher depths.

In relation to the photo-consistency problem, several strategies have been applied. Some of them are based on the assumption that photo-consistency can be characterized by comparing pairs of images. For instance, the Normalised Cross-Correlation (NCC), the Sum of Squared Differences (SSD), Mutual Information based measures, etc. Other measures try to explicitly deal with, for instance, occlusions and highlights [23], [25].

The other main problem is associated with the presence of objects that partially occlude other objects in the scene. Several approaches exist that try to minimise the effect of occlusions on 3D reconstruction and depth estimation [26], [27]. Wang *et al.* [28] separate the occlusion edges in two view regions (occluded object versus occluder), where only one of regions obeys the photo-consistency criterion. Furthermore, the experiments are carried out under the framework of a light-field occlusion model with plenoptic images. One drawback of this technique is that it is used on a light-field camera with very low disparities between *EIs* and depth ranges on the reconstructed scene are smaller than in *SAII*.

Once the occluded region has been identified, there are different strategies to eliminate (or at least, mitigate) them. In other works the occluded pixels are substituted by pixels not belonging to occluding areas from other views. This is made by creating a variance map per elemental image and applying a clustering technique to classify pixels into two classes, *foreground* and *background* pixels, depending on the variance value each pixel has [29], [30]. In [24] a similar strategy has been used by means of a previous estimation of depths of the edges belonging to the occluding objects, in order to improve the accuracy in the depth map. Others apply methods to *fill in* the missing information using, for instance, *inpainting* techniques [31].

However, one should find a solution to the depth estimation problem that may not depend on the need to have a priori information about which camera correctly sees the whole object, without any occlusion. In this sense, [32] proposes two photo-consistency measures (shape from median and shape from entropy).

The present paper tries to solve the two problems outlined before. On the one hand, through the application of a photo-consistency measure, adapting a strategy proposed in [28] based on a defocusing strategy to deal with spatial information surrounding a pixel. On the other hand, by the application of a photo-consistency measure based on the use of the median distance [32] in order to deal with and mitigate information coming from occluding objects without any a priori information. Thus, the aim is to improve the photo-consistency criterion compared to previous proposals, in scenes with partial occlusions between objects. In addition, an algorithm aimed at detecting occlusions is included for the case of *SAII* in order to improve the results obtained during the depth map regularization process.

The main contributions proposed in this work are:

- A photo-consistency measure based on the median distance between the EIs is proposed, in order to solve the depth estimation errors due to the existence of partially occluded objects. This measure combines a defocus and a correspondence term for each RGB color channel.
- An improvement of the depth map regularization is presented using the minimum photo-consistency value for each pixel at the optimal depth.
- Based on this depth map regularization improvement, we also propose: (a) a confidence measure that estimates the pixel depth estimation accuracy and (b) a way to give approximate information on where the boundaries between objects are.
- In real scenes where the ground-truth depth map is not available, the estimation of the all-in-focus image error is proposed as an approximation to assess the importance of the artifacts generated by the depth map errors.

The present paper is organized as follows. Section II provides a brief overview of surface reconstruction, comparing the use of median and average criteria in the photo-consistency measure. A proposal to estimate occlusion boundaries to improve the regularization of depth map is presented in Section III. Section IV shows the different results applied to synthetic and real scenes. Finally, some conclusions are given in Section V.

II. PHOTO-CONSISTENCY BASED ON MEDIAN DISTANCES

Integral Imaging consists of a multi-view technique that allows a computational reconstruction of a 3D scene. To do this, the *EIs* obtained during the acquisition stage are projected onto an image plane at an arbitrary distance through a real pinhole or a lens. This is possible because the 3D objects of the scene can be viewed as a combination of multiple depth images. In this way, 3D information can be estimated and analyzed by generating a series of images at different depths.

This 3D reconstruction scheme with *SAII* can be carried out in the case where the cameras are located on a flat surface [25], but also in the case where the positions of those cameras are spread or in a free pose configuration in 3D space [33]–[35].

Integral Imaging offers us a series of advantages over other 3D imaging techniques that can be exploited to overcome the partial occlusion problem and extracting more accurate depth information: (a) the capability to obtain a stack of images *in focus* at different depths. (b) From a 3D reconstruction of the scene, we can build an *all-in-focus* image as a criterion to measure the visual quality of the depth map. (c) A photo-consistency measurement for each depth level can be estimated, and obtain a photo-consistency image measure for each depth level. With this photo-consistency image measure

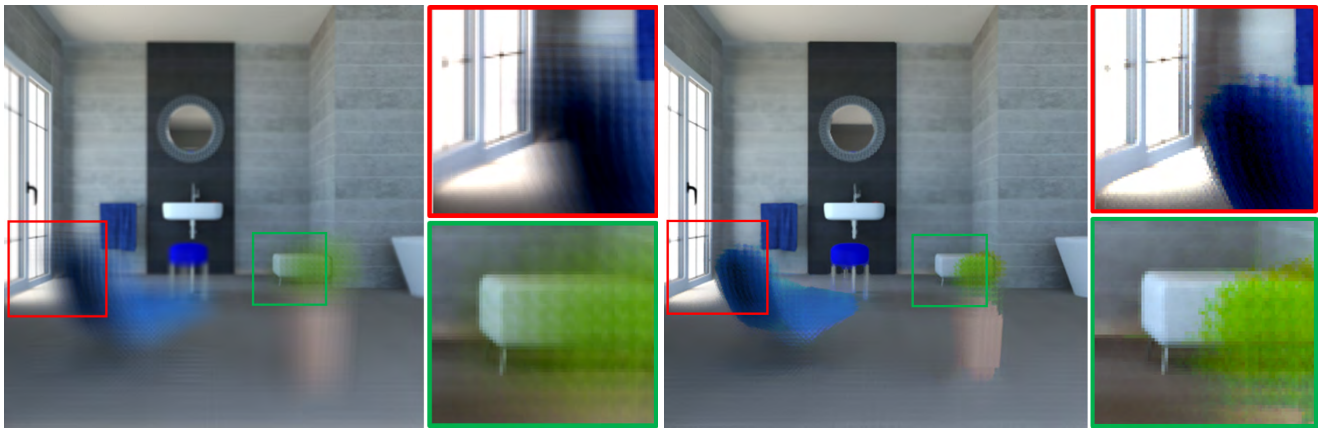


FIGURE 1. *Bathroom* **image focused at a depth of 710 cms from the camera array. In (a) we see the result obtained with the average criterion of the *EIs* contributing for each pixel position, while in (b) we can see the same result using the median criterion.**

we can get an assessment of the depth map's confidence and we can predict those areas with the highest uncertainty that are coincident with the boundaries between partially occluded objects.

Nevertheless, when estimating the depth map of a scene, the accuracy can be degraded by the texture of the objects, or by the fact that the occlusion between objects is seen differently in each camera. This makes the correspondence process between *EIs* more difficult to obtain, generating higher uncertainty in those regions. In this sense the photo-consistency criterion is key to handle with projective distortions and partial occlusions in the scene.

A. AVERAGE AND MEDIAN IMAGE PERFORMANCE IN SAII

Traditionally, the average criterion has been used as a method to obtain an image *in focus* for a specific depth. One problem that appears with the use of this criterion is that during the 3D reconstruction of the scene, objects close to the cameras, once they are *in focus*, have a tendency to expand their corresponding edges in the scene for images *in focus* reconstructed at higher depths (see Figure 1 (a)). This expansion effect initially varies depending on the FOV and the pitch value of the cameras. In other words, this expansion effect depends on the initial disparity that contains the *EIs* and varies depending on the reprojection of its pixels on 3D planes at different depths. In this work, we propose the use of the median criterion so that this effect is substantially reduced, being able to detect partially occluded regions more clearly (see Figure 1 (b)).

The median criterion represents the value of the variable in the central position of the dataset. In our case, when determining the RGB-values of a partially occluded object viewed from a set of cameras, the median criterion removes those cameras whose RGB-values are different from the majority of them. In these cases, the median criterion will be a better estimator for the surface intensity than the average criterion. Specifically, this assumes that to correctly detect the RGB values of a surface point occluded by a foreground object, this must be seen by more than half of the cameras. This is precisely the reason why median colors have been used in matte extraction [36].

B. DEPTH MAP AND ALL-IN-FOCUS RECONSTRUCTION FOR MEDIAN DISTANCES

In [28] and [37] the use of a photo-consistency measure is proposed by combining a *defocus* and a *correspondence* measure in a light-field camera. This measure, made by these two terms, was extended to the case of *SAII* in [24] and [33]. The defocus measure allows for an optimal contrast to be obtained in a certain patch of the image with the aim to improve stability over occluded regions. Nevertheless, out-of-focus regions, such as certain high frequency regions and bright lights, can produce a higher contrast that is not desired for an accurate depth estimation. In addition, the patch size also has an impact on the sensitivity measure because the defocus blur may exceed the patch size.

Correspondence measurements make depth estimation possible using photo-consistency measures and have been commonly applied in stereo problems. In these cases, a statistical measure is usually applied to resolve matching ambiguities in displacements between images. Furthermore, large displacements between *EIs* could cause correspondence measures at erroneous depths.

In order to combine the strategies of both criteria, correspondence and defocus, based on the proposed cost function defined by Vaish *et al.* [32], we have defined a photo-consistency measure combining a defocus and a correspondence term. Consider the case of a 3D point $P_j = (X_j, Y_j, Z_j)$ belonging to a plane surface, which is viewed by a set of m cameras. Consider that the distance of that plane in relation to the optical center of a reference camera with focal distance f is given by $Z = d$. This reference camera that we will call *central camera* will perform the reconstruction process of the scene. Denote by $p_j^i = (x_j^i, y_j^i)$ the pixel coordinates of the

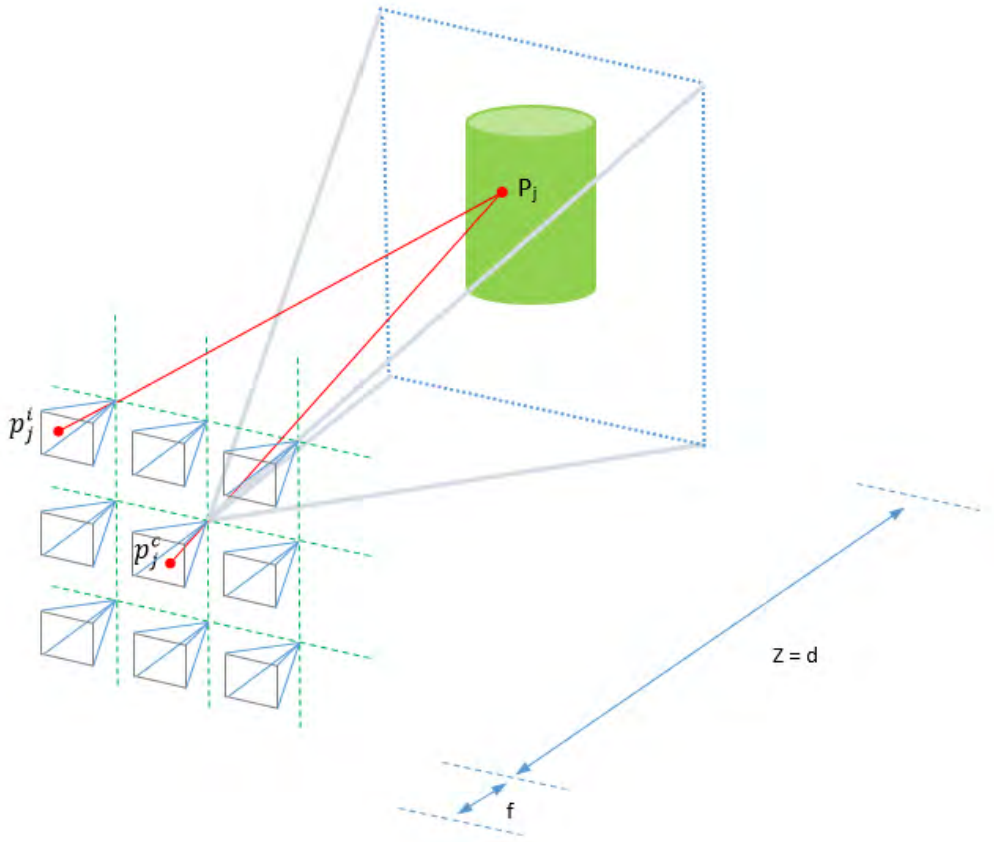


FIGURE 2. Camera setup acquisition of a 3D point of the scene.

point P_j projected into the *EI* of the i-th camera and $E_i^d(p_j^i)$ its RGB-value. Denote by $S_j^d = \{E_i^d(p_j^i) : i = 1, \ldots, m\}$ the set of the corresponding pixels projected from the point P_j to each one of the m input *EIs* for a depth $Z = d$. From the information stored in S_j^d, we can estimate a measure for the average for each RGB color $E_{avg}^d = \text{mean}(S_j^d)$ and the median for each RGB color $E_{med}^d = \text{median}(S_j^d)$ of the *EIs*.

In our approach, the first term (correspondence term) defines a cost function equal to the sum of median distances for each RGB color, of all pixels projected $E_i^d(p_j^i)$ in the m cameras with respect to E_{med}^d and with the pixel p_j^c corresponding to the *EI* of the central camera E_c with RGB-value $E_c(p_j^c)$. It is worthwhile to indicate that the projections of the 3D point vary at different depths on the *EIs* for all cameras, except in the case of the *EI* of the central camera. Thus, the correspondence term is defined by the following expression:

$$C_{RGB}^d(p_j^c) = \sum_{r=1}^{RGB} \Big(median\{|E_i^d(p_j^i) - E_{med}^d|_r\} + median\{|E_i^d(p_j^i) - E_c(p_j^c)|_r\} \Big), \quad \forall i = 1, \ldots, m \tag{1}$$

where r refers to the number of the color channel in the RGB image. Analogously, for the defocus term, we define the distance between $E(p_j^c)$ and E_{med}^d as follows

$$D_{RGB}^d(p_j^c) = \sum_{r=1}^{RGB} |E_c(p_j^c) - E_{med}^d|_r \tag{2}$$

Therefore, the proposed photo-consistency measure $P_{RGB}^d(p_j^c)$ would be given by the sum of the two terms: $P_{RGB}^d(p_j^c) = D_{RGB}^d(p_j^c) + C_{RGB}^d(p_j^c)$. These distances based on the median criterion are estimated for each pixel and for each color channel by taking normalized RGB-values between 0 and 1 in the *EIs*. In addition, the photo-consistency estimate is made from the sum over the three color channels.

When comparing the photo-consistency image for each depth, some noise can be observed in these images, specially in scenes with high texture surfaces or with many edges between objects. This noise can be reduced applying an accurate noise filtering technique based on Total Variation regularization [38]. This technique applies a denoising technique and depends on a parameter λ that affects the balance between removing noise and preserving the photo-consistency image content for each depth. In our experiments we have used a value of $\lambda = 60$.

Additionally, neighborhood information can be added by applying a bilateral filter technique with a spatial mean and a zero mean Gaussian kernel function G_s to the image intensity differences of E_c, centered at the current pixel around a window W, and with a standard deviation equal to 0.1. This is an empirical parameter that establishes the influence of the photo-consistency measurements of the neighbors in the filter. Thus, for each pixel p_j^c with coordinates (x_j^c, y_j^c) in $E_c(p_j^c)$, we can define the following cost function for each depth value d as:

$$L_{RGB}(p_j^c, d) = \frac{\sum_{p^c \in W} P^d_{RGB}(p_j^c) G_s(|(E_c(p_j^c) - E_c(p^c))|)}{\sum_{p^c \in W} G_s(|(E_c(p_j^c) - E_c(p^c))|)} \tag{3}$$

In relation to the window size, the larger the size of the window, the greater accuracy is reached in the depth map because it achieves a better coherence between neighboring pixels. However, in scenes with thin objects a large window can degrade the accuracy of the depth in those areas. One way to solve this problem is to discriminate the low texture areas with respect to the areas that contain texture or have edges of objects in the central image camera. Thus, we define a pixel-wise measure as a gray-level variance typically used as a focus measure [39].

$$F(x^c, y^c) = \sum_{(x^c, y^c) \in W} (E_c(x^c, y^c) - \mu)^2 \tag{4}$$

where W is the $r \times r$ neighborhood of a pixel at position (x^c, y^c) and μ is the mean gray-level of pixels within W. The selection of r is a trade-off between robustness to noise and spatial resolution. In this work we use a value of $r = 11$ such that when $F(x^c, y^c) < 0.0001$, the pixel is considered belonging to a low texture area, and considered as significantly textured, otherwise. From this threshold, we apply the bilateral filter with a window size of $W = 11 \times 11$ for the first case and a window size of $W = 3 \times 3$ for the second case.

Finally, the optimal depth is determined over all depth planes minimizing the cost function $L_{RGB}(p_j^c, d)$ as:

$$\widehat{L_{RGB}}(p_j^c, d^*) = \underset{d \in [Z_{min}, Z_{max}]}{\arg\min} \; L_{RGB}(p_j^c, d) \tag{5}$$

where $\widehat{L_{RGB}}(p_j^c, d^*)$ is the minimum photo-consistency value for each pixel p_j^c at the optimal depth d^*.

The method proposed in this section is referred as Photo-consistency measure based on Median distances (*Photo-Med*) and it will be used in Section IV.

The method proposed here satisfies the three specifications discussed at the beginning of Section II: (a) establish a stack of images E^d_{med} at different depths where the objects of the scene are *in focus* at that depth. (b) Since we have an estimated depth for each pixel, we can get the RGB value for that depth and create an *all-in-focus* image. This image has a strong dependence with depth estimation because depth errors appear as artifacts in the *all-in-focus* image, especially in textured surfaces. (c) We can obtain the photo-consistency measurement for each depth level. This allows us to represent an image that we call E_{photo} (value for each pixel in which its value is minimum) and that we will use to predict the limits between partially occluded objects.

III. THE ROLE OF OCCLUSIONS IN DEPTH REGULARIZATION

After an initial depth map estimation (see Eq. 5), we can refine the results with a global regularization term using a smoothness term similar to that applied in [28]. In that work, the authors define a predictor composed by three terms for assessing if a particular pixel is occluded, by combining depth, correspondence and refocus cues.

One drawback associated to the depth information is that it contains errors due to the fact that the photo-consistency measure is not able to focus at the proper depth generating false edges that mislead the regularization process.

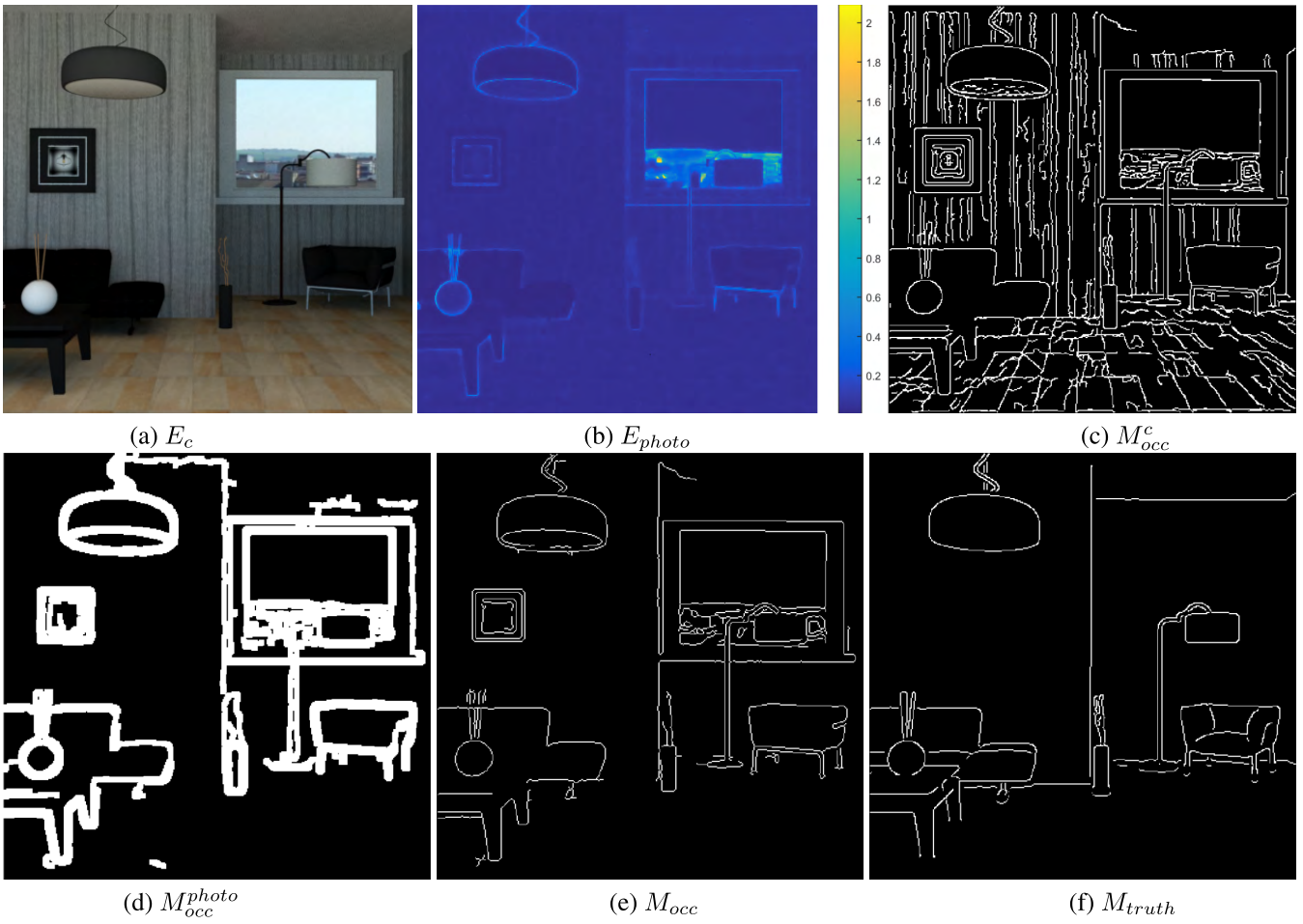
The second and third terms are related to the strategy used to obtain the depth map. Authors apply their method in a light-field camera, where the disparities between the *EIs* are very low. This allows the non-occluded zone to maintain the visibility to the corresponding object, at different depths. However, when the disparity between *EIs* increases (for instance, in *SAII*), the number of cameras that see the partially occluded object varies with depth, and that approach fails in *SAII* where there are larger disparities in EIs than in lenslet-based plenoptic images. Therefore, we propose a new approach based on the photo-consistency measure to solve this problem.

A. ESTIMATION OF PARTIAL OCCLUSIONS IN SAII

During the initial depth map reconstruction process, we have stored the photo-consistency measurement for each depth level. This allows us to represent an image called E_{photo} with the minimum photo-consistency value $\widehat{L_{RGB}}(p_j^c, d^*)$ for each pixel p_j^c (see Figure 3 (b)). When observing this figure we can see how the photo-consistency measure obtains low values in regions where there are no occlusions, while this value increases in occluded areas and around them. This indicates that there is a conflict in the distance measure for the different cameras in relation to E^d_{med} and $E_c(p_j^c)$. Note that the photo-consistency measure decreases in the occlusion areas around the object edges. Therefore, the occlusion zones are on the edges of the image where there is also high uncertainty in E_{photo}.

The analysis of the photo-consistency measure E_{photo} can be used to get information about two fundamental aspects in the regularization process: (a) to have a confidence measure that estimates the pixel depth estimation accuracy and (b) to provide us an approximate information on where the boundaries between objects are.

In order to detect the boundary limits of the depth discontinuities between objects, first we apply a *Canny* filter on the *EI* of the central camera E_c obtaining a binary mask



(a) E_c (b) E_{photo} (c) M^c_{occ}

(d) M^{photo}_{occ} (e) M_{occ} (f) M_{truth}

FIGURE 3. Occlusion predictor for a synthetic scene (a). The edges (e) are obtained through the intersection of the masks M^c_{occ} (b) and M^{photo}_{occ} (c). We show the *Canny* filter (f) applied on the ground-truth depth map of the scene.

M^c_{occ} (see Figure 3 (c)). This mask contains accurate edges between objects but it also contains edges due to changes in color. Next, we apply a *Canny* filter on E_{photo} and apply a morphology filter to dilate the edges M^{photo}_{occ} (see Figure 3 (d)). Finally, we estimate the intersection between these two binary masks and refine the output mask by joining small segments into only one (see Figure 3 (e)):

$$M_{occ} = M^c_{occ} \times M^{photo}_{occ} \tag{6}$$

In Figure 3 (f) we show the *Canny* filter applied to the ground-truth depth map of the scene (M_{truth}). The main differences are found in the picture hanging on the wall and in the window that has the same depth as that of the corresponding walls in the ground-truth depth map.

B. DEPTH REGULARIZATION

Given an initial depth map obtained by the proposed method (see Eq. 5), a regularization process with a Markov Random Field (MRF) is applied to generate the final depth map. Thus, we can define and minimize the energy with unary and pairwise terms as

$$F = \sum_p E_1(p, l_d(p)) + \sum_{p,q} E_2(p, q, l_d(p), l_d(q)) \tag{7}$$

where p and q are neighboring pixels, and $l_d(p)$ and $l_d(q)$ are the depth label values for the pixels p and q respectively. The first term $E_1(p, l_d(p))$ expresses that depth labels $l_d(p)$ should agree with the observed data while the second term measures the depth smoothness level between neighboring pixels.

Consider that $N(l_d)$ is the number of depth levels in the interval $[Z_{min}, Z_{max}]$ and $l^*_d \in [1, \ldots, N(l_d)]$ is set of all possible depth label values. If $l_d(p)$ is the initial depth map label value of pixel p, then the first data energy term for all possible depth labels l^*_d is given by

$$E_1(p, l_d(p)) = \min\{|l_d(p) - l^*_d|, N(l_d)/2\} \quad \forall l^*_d \in [1, \ldots, N(l_d)] \tag{8}$$

and the second data energy term relates the depth map level value of pixel p with its corresponding neighboring pairwise

FIGURE 4. Synthetic scenes used in the work. From left to right: ***Bathroom***, ***Living-room*** and ***Toysroom***.

pixels and it is defined as

$$E_2(p, q, l_d(p), l_d(q)) = \frac{E_{photo}(p)^{0.1} + E_{photo}(q)^{0.1} + 1}{|\nabla(E_c(p)) - \nabla(E_c(q))| + k|M_{occ}(p) - M_{occ}(q)|} \quad (9)$$

where $E_{photo}(p)$ is the minimum photo-consistency value or confidence function, $\nabla(E_c)$ is the gradient of the *EI* of the central camera E_c and M_{occ} is a binary mask with value equal 1 if it is an occlusion edge, and 0 otherwise. Parameter k with value equal to 10^5 is a weighting factor to penalize the propagation in the regularization process when detecting an occlusion, that is, a change in the mask M_{occ}.

The numerator measures the degree of confidence in the depth map, while the denominator penalizes the difference between the image gradients and the difference in the occlusion mask M_{occ} between neighboring pixels. The minimization is solved using a standard graph-cut algorithm [40]–[42] to obtain the final depth map estimation.

IV. EXPERIMENTAL RESULTS

In order to analyze the performance of the technique proposed in this paper, we compare it with two other 3D reconstruction techniques based on the average image criterion for each depth E^d_{avg}. The first one is the proposal by Wang *et al.* [28] where we have only changed the image reconstruction process of each depth used in a light-field camera by the one used in *SAII*. The rest of the method remains the same as the original work including estimation of occlusions using a metric on angular pixel patches, color consistency constraint, occlusions prediction and depth regularization.

The second one called Elemental Imaging from Edge removal (*EIEd-rem*) approach [24], removes the edges of objects in the *EIs* when they are *in focus*. This is a problem that appears when using the variance where the objects close to cameras, once they are *in focus*, have a tendency to expand their corresponding edges in the scene for images *in focus* at higher depths. This method does not treat the problem of partially occluded objects. Therefore, in order to include this method in the comparison, we have added to the method the same proposal of occlusion estimation and depth regularization to the one proposed in this work.

In both techniques, the photo-consistency measure has also two terms of the same nature as the one proposed in this paper: correspondence and defocus, but the definition of these terms is different. The correspondence term is equal to the variance and the assessment of distances between the *EIs* in relation to E^d_{avg} is based on the average criterion for each pixel and for each color channel. The defocus term is similar to our proposal but changing E^d_{med} by E^d_{avg}. In addition, we have analyzed the occlusion estimator performance, and its role in depth regularization from the initial photo-consistency image E_{photo} obtained.

A. DATASET DESCRIPTION

A series of images were used to test the performance of the proposed depth estimation with occlusion handling strategy. Some of them, of synthetic nature, were created using Autodesk 3DS Max. These synthetic images correspond to indoor scenes (and they will be called *Bathroom*, *Living-room* and *Toysroom*) (see Figure 4). A ground-truth depth map based on the *z-buffer* algorithm is available for all the synthetic images, in terms of graphical units. The equivalence $1GU \equiv 1cm$ is used in these scenes.

The other group of images corresponds to real images of three different scenes, acquired using a 3×3 array of Stingray $F080B$ cameras. These cameras were located in a *square grid* with the optical axes pointing in parallel directions. Image resolution for each camera was 1024×768 pixels. Images in Figure 5 show the elemental image corresponding to the central camera, for the three scenes. In the first scene, a USAF test was printed and pasted on a ceramic tile. This ceramic tile allowed us to have the test in an orthogonal position in relation to the optical axes of the cameras. The second scene corresponds to a series of toys placed at different depths in relation to the camera array. Finally, the third scene shows

FIGURE 5. Real scenes obtained in our laboratory. From left to right: *USAF test*, *Toys* and *OcclusionTree*.

TABLE 1. Experimental setup parameters. The first three images are the synthetic images created using Autodesk 3DS Max. The other three are real images acquired with the 3 × 3 camera array. In all case the units are in centimeters.

Image name	Cam	Z_{min}:Z_{step}:Z_{max}	(c_x, c_y)	$pitch$	f
Living-room	7×7	370:10:900	(3.6,3.6)	5	5
Bathroom	7×7	220:10:830	(3.6,3.6)	5	5
Toysroom	7×7	220:10:750	(3.6,3.6)	5	5
USAF test	3×3	80:variable:550	(4.76,3.57)	3.0,3.5	0.8
Toys	3×3	80:variable:550	(4.76,3.47)	3.0,3.5	0.8
OcclusionTree	3×3	80:variable:550	(4.76,3.47)	3.0,3.5	0.8

a person that is sat down behind an indoor plant, which is partially occluding a gesture he is doing.

Table 1 shows the acquisition configuration parameters used in these experimental results. Second and third columns show the camera rack configuration, and the depth range from Z_{min} to Z_{max} with a step size of Z_{step}. Fourth and fifth columns give the specification details of the II pickup process, where (c_x, c_y) is the physical size of the camera sensor and *pitch* is the pitch of the cameras. A focal length of $f = 50mm$ is used for the synthetic scenes and $f = 8mm$, for the real scenes.

In the case of the real scenes, we have used a variable step size Z_{step}, being this step length smaller when it is closer to the camera array, and larger as the distance increases. To establish a variable depth scale d in the scene, we define a range of values $L(d) \in [0, \ldots, N(l_d)-1]$, and we apply the following rule:

$$d = \frac{Z_{min}}{1 - L(d) * \left(\frac{1 - \frac{Z_{min}}{Z_{max}}}{(N(l_d)-1)} \right)} \tag{10}$$

In addition, the depth map ground-truth is available for the synthetic scenes. This allows us to assess the quality of the depth estimation results through the application of two figures of merit. The first one of them is the Root Mean Squared Error (*RMSE*), defined as:

$$RMSE = \sqrt{\frac{1}{r \times c} \sum_{(i,j)} [\hat{Z}(i,j) - Z^*(i,j)]^2} \tag{11}$$

where r and c are the number of rows and columns in each image, $\hat{Z}$ is the ground-truth depth map and Z^* is the estimated depth map for a pixel at position (i, j).

The second figure of merit is the Structural Similarity Index (*SSIM*) [43]. This index is used to assess the similarity between two images. Its value depends on a window size defined by the user. In this work, a window size of 11 × 11 pixels has been used to estimate this measure. Its SSIM index for two windows, W_x and W_y, both of size 11 × 11 would be given by:

$$SSIM(W_x, W_y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{12}$$

where μ_x and μ_y refer to the average values for windows W_x and W_y, respectively and σ_x and σ_y to its corresponding variance values. σ_{xy} refers to the covariance of W_x and W_y. $c_1 = (k_1 * N(l_d))^2$ and $c_2 = (k_2 * N(l_d))^2$ are two constants used to stabilize the ratio in case the denominator gets low values. These constants depend on two small values $k_1 = 0.01$ and $k_2 = 0.03$ and $N(l_d)$ that is the number of levels used to reconstruct the depth map. In this paper we have adopted the same values for k_1 and k_2 as those used in the original work [43] for the case of a Gaussian window. $SSIM \in [-1, 1]$, and the closer the *SSIM* value to 1, the most similar the two windows, W_x and W_y, are.

In practice, a single overall quality measure for the entire depth map is considered. In [43], the authors use a mean SSIM (*MSSIM*) index defined over all possible windows as:

$$MSSIM = \frac{1}{r \times c} \sum_{(i,j)} SSIM(W_x(i,j), W_y(i,j)) \tag{13}$$

TABLE 2. ***RMSE*, *MSSIM-Depth* and *MSSIM-Focus* values for the 3 synthetic scenes without and with regularization. Best values for each case are given in bold.**

	without Reg			*with Reg*		
	Wang et al.	*EIEd-rem*	*Photo-Med*	*Wang et al.*	*EIEd-rem*	*Photo-Med*
Bathroom						
RMSE	74.2836	52.2387	41.0338	52.7664	36.8538	**34.4419**
MSSIM-Depth	0.2646	0.4810	0.4048	0.5556	**0.7688**	0.7534
MSSIM-Focus	0.9762	0.9720	**0.9825**	0.9615	0.9637	0.9709
Living-room						
RMSE	69.9511	46.1948	42.2109	40.3469	**28.7221**	32.7828
MSSIM-Depth	0.3039	0.4481	0.5407	0.5707	**0.7491**	0.7394
MSSIM-Focus	0.9847	0.9804	**0.9885**	0.9718	0.9691	0.9826
Toysroom						
RMSE	107.2091	66.1476	72.8093	38.6366	**20.4748**	27.5309
MSSIM-Depth	0.1661	0.2189	0.2370	0.5133	0.7101	**0.7378**
MSSIM-Focus	0.9905	0.9871	**0.9925**	0.9795	0.9780	0.9881
Mean						
RMSE	83.8146	54.8604	52.0180	43.9166	**28.6836**	31.5852
MSSIM-Depth	0.2449	0.3827	0.3962	0.5465	**0.7427**	0.7435
MSSIM-Focus	0.9838	0.9798	**0.9888**	0.9709	0.9798	0.9805

This second figure of merit has been used to measure two types of errors in terms of visual quality. First, if the ground-truth depth map is available, we can estimate the similarity between the estimated depth map and the ground-truth depth map. We have called this error measure *MSSIM-Depth*. Furthermore, we can obtain the *all-in-focus* image and check the importance of the artifacts generated by the errors in the depth map. In this case we can compare the *all-in-focus* image with the *EI* of the central camera E_c. We have called this error measure *MSSIM-Focus*.

B. RESULTS AND DISCUSSION FOR SYNTHETIC SCENES

Figure 6 presents the ground-truth depth map images in the first row. In these images we can see how the boundaries of the objects are well defined, recognizing in detail the changes in the shape of the objects at different depths. In rows 2, 3, and 4 the results of the initial depth map for the method by *Wang et al.*, *EIEd-rem* and *Photo-Med* are shown. In rows 5, 6 and 7 the results of the depth map regularization process for the three methods are shown.

Table 2 shows the *RMSE*, *MSSIM-Depth* and *MSSIM-Focus* values associated to the depth estimation grouped per synthetic scene for the cases of Figure 6. The mean values show the average estimated value for the three scenes.

Analyzing the results in Table 2, we observe that *Photo-Med* gets better performance results in the initial depth map in terms of *RMSE* values than *Wang et al.* and *EIEd-rem* for *Bathroom* and *Living-room* scenes. In addition, rows 2 to 4 in Figure 6 show that in *Photo-Med* the shape of the objects' depth discontinuities is better defined within the map especially in the regions with thin shapes.

In the case of *Toysroom*, the results of our method in terms of *RMSE* are worse than *EIEd-rem*. However, when analyzing the depth map visual quality (see rows 3 and 4 of Figure 6), we see that *Photo-Med* defines the shape of the objects in a better way but includes more noise in the background wall which occupies more than 50% of the image. In this sense, the advantage of *EIEd-rem* is because the method removes the edges of the objects and reduces the background noise but increases the computational cost to generate the final depth map. However, none of the methods is able to accurately estimate the depth for objects whose surface has low texture.

In relation to the approach proposed by *Wang et al.*, the angular patch around de edges of the occluding objects divides the set of cameras into two regions, where only one of them obeys the photo-consistency criterion. This hypothesis is valid provided that the disparity between the *EIs* is low, so that the region of cameras that correspond to the non-occluded zone maintains the visibility of the object at different depths. Nevertheless, this is not fulfilled as the disparity and the range of depths increases, and therefore the amount of cameras that see the partially occluded object varies with the depth, which affects the method accuracy.

Regarding the measurements of *MSSIM-Depth* and *MSSIM-Focus*, we can say that, in general, higher values are obtained by *Photo-Med* and therefore the visual quality that can be observed is better.

When analyzing the results of the regularization process for the three methods, we should emphasize how the results significantly increase the depth accuracy. In *Bathroom* scene, the value of RMSE is more accurate in the regularized version of *Photo-Med*, while in *Living-room* and *Bathroom*, *EIEd-rem* obtains better results (see Table 2). For *Living-room* the differences between *EIEd-rem* and *Photo-Med* are in the lamp area around the window and the vase on the table (see rows 6 and 7 of Figure 6). In *Toysroom* the errors are concentrated

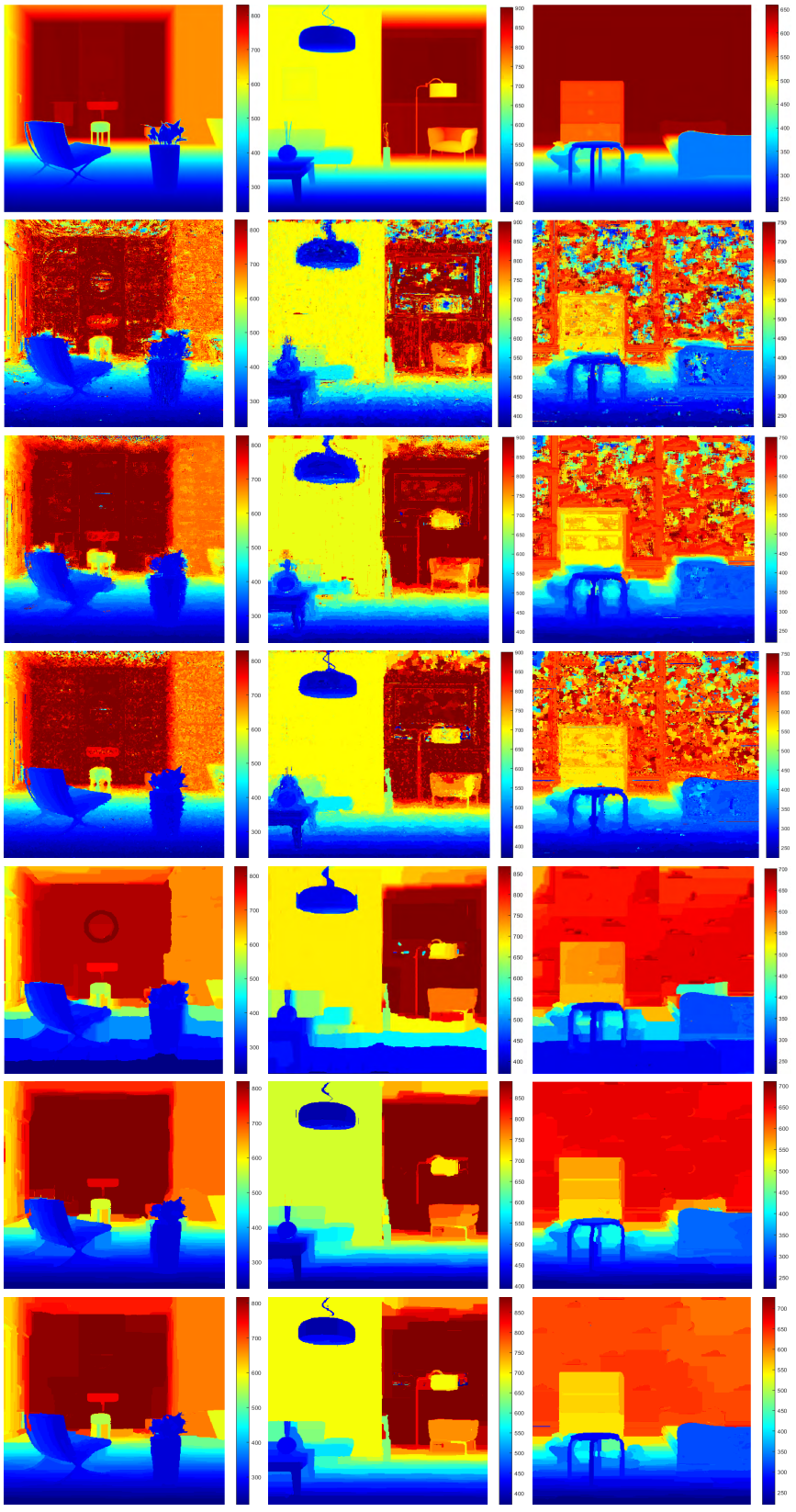


FIGURE 6. Methods are given in rows and synthetic images in columns. The first row shows ground-truth depth maps. The methods that were tested are: ***Wang et al.***, ***EIEd-rem*** and ***Photo-Med***, without regularization, and the corresponding versions after regularization.

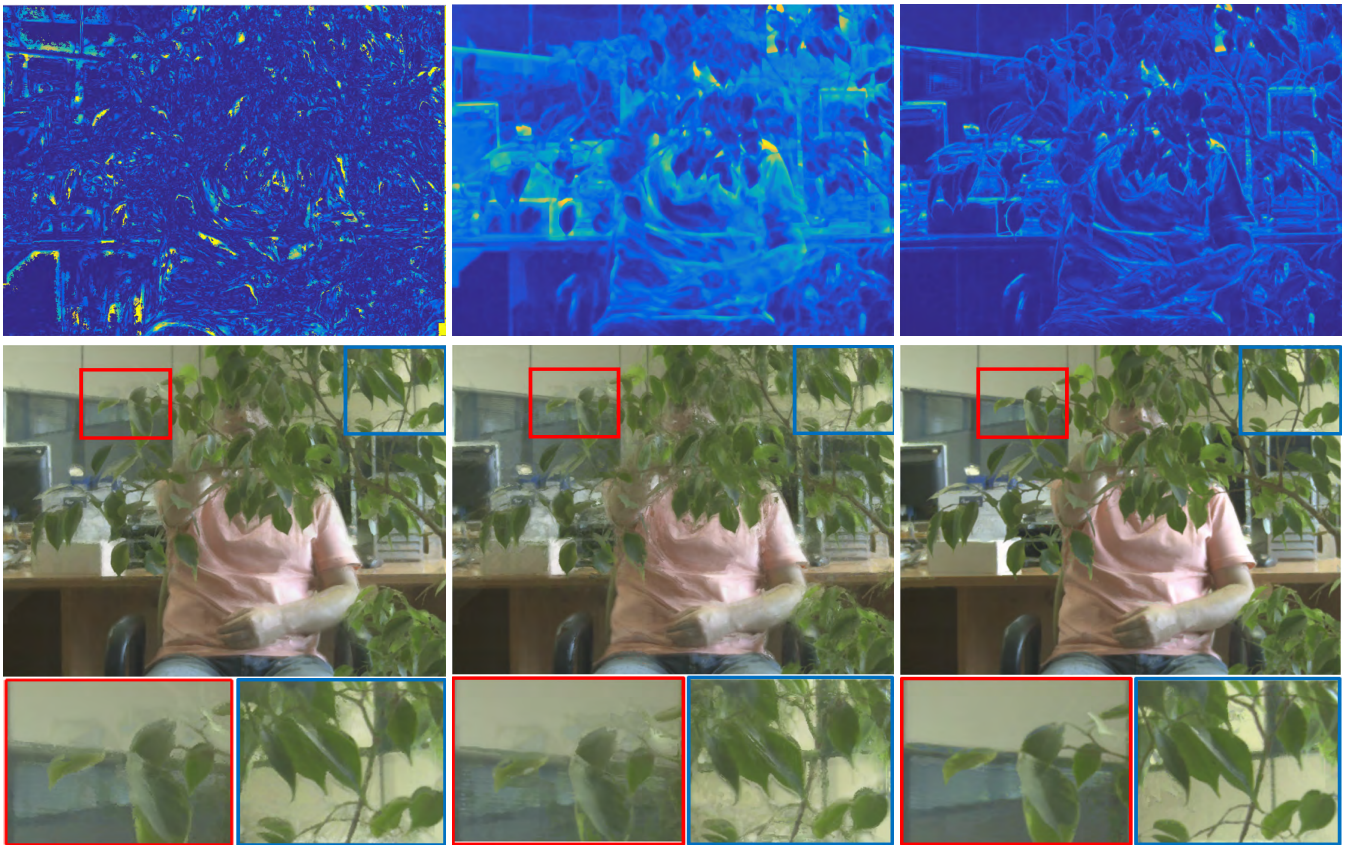


FIGURE 7. First row shows (from left to right) the photo-consistency image E_{photo} for the methods *Wang et al.*, *EIEd-rem* and *Photo-Med* respectively. Second row shows the corresponding *all-in-focus* image for these methods. Zoomed-in image areas have been included in these images to highlight the increase in the *all-in-focus* image quality.

in the back wall, where the initial depth map has many inaccuracies due to the existence of low texture regions. In this sense the regularization algorithm homogenizes the surface, but this does not mean that the correct depth is found.

Although the proposed algorithm to estimate occlusions is able to detect the borders between partially occluded objects and use this information in the regularizer, in the case of *EIEd-rem* the photo-consistency values of the E_{photo} images are not so precise, losing part of the depth of some objects, especially for the case of thin areas. This is due to the ability of the median criterion to deal with reconstruction values at depth discontinuities and occluding areas.

In the case of the *Wang et al.* approach, the depth map result improves after the regularization process. However in our experiments, the proposed Energy functional in the equations 8 and 9 for the depth regularization gets a more accurate 3D map in the SAII scenes than the expressions used in [28]. This may be due to the errors in the initial depth map labels used in its expression for the depth regularization and the occlusion prediction method they propose, which was designed for small *EIs* disparities in lenslet-based plenoptic images.

Finally, we would like to point out how the *all-in-focus* image for the case of the regularized depth map is usually worse than for the case without regularization (see the *MSSIM-Focus* values in Table 2). Thus, although homogenizing the depth values in a certain area improves the depth map, it may result in the loss of certain visual details on object surfaces in the *all-in-focus* image that the initial depth map contained. Thus, specific details of the scene such as curved surfaces or objects with thin areas may be impaired in their accuracy performance. In addition, initial errors in the depth map can be amplified during the regularization process.

C. RESULTS AND DISCUSSION FOR REAL SCENES

Figure 8 shows the results obtained by the different methods on the images acquired by the 3 × 3 camera array. Rows 1, 2 and 3 show the results of the initial depth map for *Wang et al.*, *EIEd-rem* and *Photo-Med* approaches. In rows 4, 5 and 6 the results of the depth map regularization process for the three methods are also shown.

Table 3 shows *MSSIM-Focus* values associated to the errors detected in the *all-in-focus* image when compared to the *EI* of the central camera. In this case, being a real scene, we do not have information about the correct depth values of the objects in the scenes, and therefore we can only analyze the visual information of the depth map and the errors that the estimated map makes when the *all-in-focus* image is created.

Table 3 clearly shows how *Photo-Med* method gets better accuracy results in terms of *MSSIM-Focus* value when

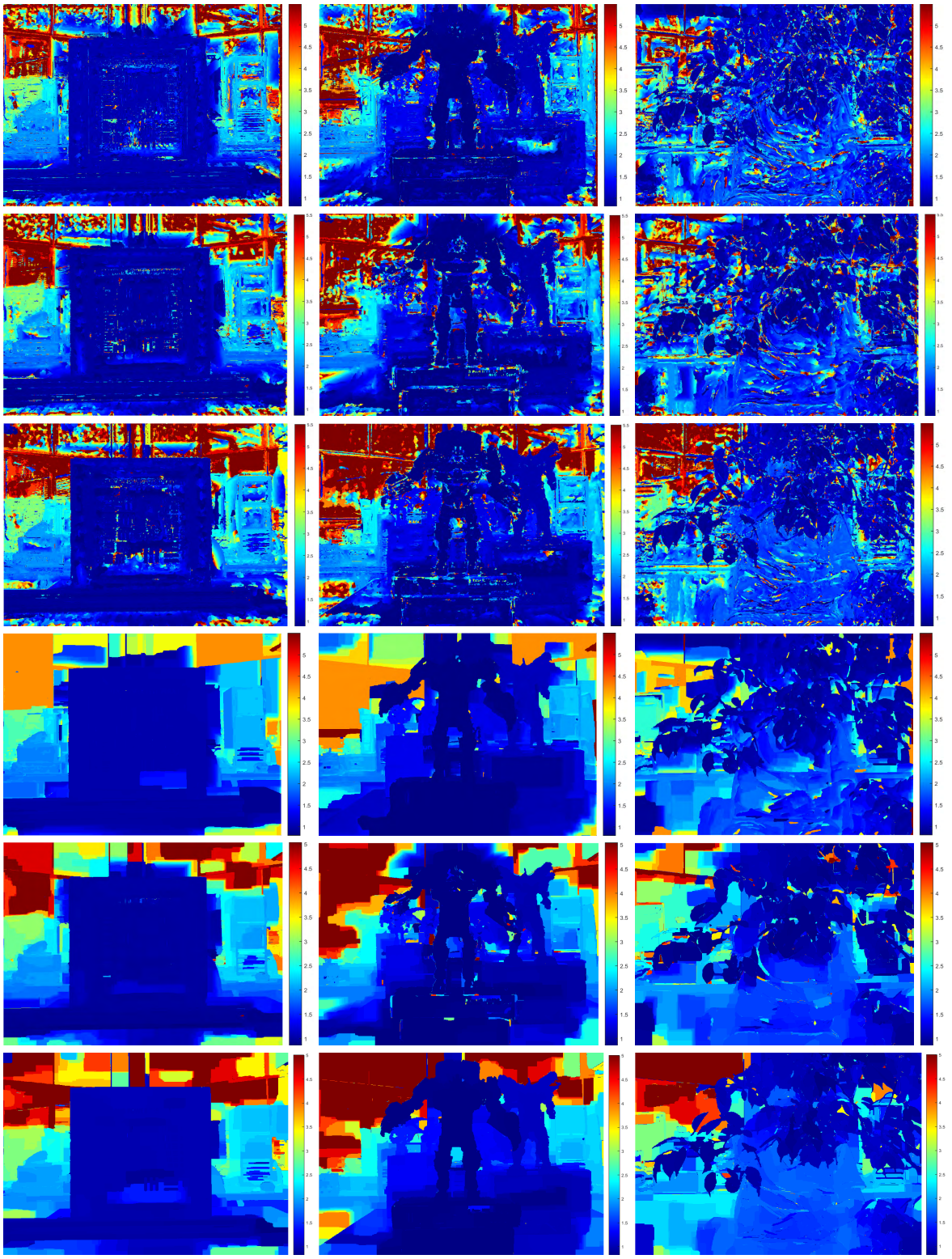


FIGURE 8. Depth estimation results for three real scenes. For each one of them, the following methods are given in rows (from left to right): ***Wang et al.***, ***EIEd-rem*** and ***Photo-Med*** approaches. The corresponding regularized versions are in rows 4, 5 and 6.

compared to *Wang et al.* and *EIEd-rem*. Note that *EIEd-rem* is the most imprecise and generates more artifacts in its corresponding *all-in-focus* image.

Figure 7 shows the result of the *all-in-focus* image generated by the three methods for the scene *OcclusionTree*. It can be seen how the quantity of artifacts or noise due to

TABLE 3. ***MSSIM-Focus*** **values for the three real scenes. Best values for each case are given in bold.**

	without Reg			*with Reg*		
	Wang et al.	*EIEd-rem*	*Photo-Med*	*Wang et al.*	*EIEd-rem*	*Photo-Med*
USAF test	0.9794	0.9519	**0.9869**	0.9325	0.9323	0.9668
Toys	0.9553	0.9010	**0.9763**	0.9015	0.8548	0.9371
OcclusionTree	0.8842	0.8034	**0.9425**	0.7657	0.6992	0.7805
Mean	0.9396	0.8854	**0.9686**	0.8666	0.8288	0.8948

inaccuracies in the depth map is much greater in the case of the *EIEd-rem* and *Wang et al.* methods than for *Photo-Med*.

Nevertheless, in complex scenes the *Photo-Med* method does not always achieve a good correspondence between the different cameras in order to give a satisfactory response, as it can be seen in the partially occluded crystal in the upper right part of the image or traces of leaves that appear on the wall in the background.

These errors also affect to the information that is obtained through the photo-consistency values of the E_{photo} image, and that are used to detect the occlusion boundaries between objects that are used to assign a confidence value to each pixel in the regularizer. This can be seen in the Figure 7 where the proposed measure in this work achieves a higher precision to detect occlusions. In the second row in Figure 7, images also show two zoomed areas where the *all-in-focus* image results are better for the *Photo-Med* method, in relation to the minimization of different types of artifacts that usually appear around occluding objects.

These errors also affect the information that is obtained through the photo-consistency values of the E_{photo} image, and that are used to detect the occlusion boundaries between objects and to assign a confidence value to each pixel in the regularizer. This can be seen in Figure 7, where the proposed measure in this work achieves a higher precision to detect occlusions. In the second row in Figure 7, images also show two zoomed areas where the *all-in-focus* image results are better for the *Photo-Med* method, in relation to the minimization in the appearance of different types of artifacts around occluding objects.

When comparing *Wang et al.* and *EIEd-rem* vs. *Photo-Med* for the real scenes in Figure 8 we can see that the *EIEd-rem* method introduces a larger amount of noise in the depth maps. Specifically, in the *EIEd-rem* method the area of the objects expands, and therefore their shapes are not well defined (at different depths). This appears clearly in the *USAF test* y *Toys* scenes. This also applies to the stapler and other objects that appear on the top of the laboratory tables. In the case of the *Toys* scene, the methods are able to find the approximate distance for most of toys although as we have indicated *Wang et al.* and *EIEd-rem* do not define the details of the objects with the same degree of precision as the proposed *Photo-Med* depth boundaries.

In relation to the regularized images, it is worth mentioning that the results obtained by *Photo-Med* are quite convincing in their final visual appearance. Thus, some problems related to surfaces with low texture or changes in illumination like the reflections that are produced by the glass in the laboratory windows are relatively well resolved. However, the photo-consistency measure can only ensure that it will work well on surfaces that contain enough texture, and therefore allowing to focus the object at the proper depth.

V. CONCLUSIONS

This paper has presented a depth estimation method for *SAII* based on a novel photo-consistency measure that uses the median distance in order to deal with and mitigate the errors coming from occluding objects. The main interesting feature in this approach is to find a solution in the 3D reconstruction process that does not take into account any a priori information about which camera correctly sees the object at a certain depth in the case of scenes with occlusions. The only assumption is related to the number of cameras that correctly see the object point surface to reconstruct. In addition, a robust solution is proposed to detect the boundary limits between partially occluded objects for the case of *SAII*. This is a critical aspect in order to improve the depth map during the regularization process.

For the comparison with other state-of-the-art methods, two photo-consistency-based techniques have been included that also contain the correspondence and defocus terms as the one proposed in this work. The methods have been tested in two groups of images, one of synthetic nature and other consisting of real scenes acquired with a 3×3 camera array. Furthermore, two measures were used to assess the quality of the depth estimation results.

Analyzing the results we can conclude that the proposed method obtains satisfactory results with a better performance than the other methods for most of the real and synthetic scenes. In addition, regarding visual appearance, we can indicate that the shape of the boundaries in the objects is better defined in the resulting depth map with the proposed method as compared to the other ones considered in our experiments.

It is also worthwhile noting that the observed depth map resulting from the proposed approach generates a minor error in the all-in-focus image, which serves as a practical indicator of the depth map quality, especially in the case of real scenes where, in general, it is not possible to have information about the ground-truth depth map.

This loss of accuracy in the *all-in-focus* image for the case of synthetic scenes opens the possibility of a future line of research that would use a regularization function that would

APPENDIX

See Table 4.

TABLE 4. Symbols Used in this Paper.

Symbol	Explanation
$P_j = (X_j, Y_j, Z_j)$	$3D$ point.
$\{m, f\}$	Number of cameras, and focal distance.
$Z = d$	Distance of the plane in relation to the optical center of the central camera.
$p_j^i = (x_j^i, y_j^i)$	Pixel coordinates of the point P_j projected in the Elemental Image of i-th camera.
$E_i^d(p_j^i)$	RGB colors of the pixel coordinates p_j^i.
S_j^d	Set of pixels projected from P_j to each m input EI.
$E_{avg}^d = \text{mean}(S_j^d)$	Average value for each RGB color in the set S_j^d for a distance d.
$E_{med}^d = \text{median}(S_j^d)$	Median value for each RGB color in the set S_j^d for a distance d.
$p_j^c = (x_j^c, y_j^c)$	Pixel coordinates of the point P_j projected in the EI of the central camera.
$E_c(p_j^c)$	RGB colors of the pixel coordinates p_j^c corresponding to the EI of the central camera.
$median\{\lvert E_i^d(p_j^i) - E_{med}^d\rvert_r\} + median\{\lvert E_i^d(p_j^i) - E_c(p_j^c)\rvert_r\}$	Sum of the median distances for each RGB color respect to the E_{med}^d and $E_c(p_j^c)$.
$C_{RGB}^d(p_j^c)$	Correspondence term of the pixels projected by P_j for a distance d.
$D_{RGB}^d(p_j^c)$	Defocus term for pixels projected by P_j for a distance d.
$P_{RGB}^d(p_j^c)$	Photo-consistency measure of the pixels projected by P_j for a distance d.
λ	Parameter used in the Total Variation regularization to remove noise in the photo-consistency image.
G_s	Zero mean Gaussian kernel function.
p^c	Neighbor pixels of p_j^c in the window W.
$L_{RGB}(p_j^c, d)$	Photo-consistency measure of $P_{RGB}^d(p_j^c)$ after applying a bilateral filter.
$\widehat{L_{RGB}}(p_j^c, d^*)$	Minimum photo-consistency at the optimal depth d^*.
E_{photo}	Image with the minimum photo-consistency value for each pixel.
M_{occ}^c	Canny filter on the EI of the central camera E_c.
M_{occ}^{photo}	Canny filter on E_{photo} and morphology filter to dilate the edges.
M_{occ}	Intersection of M_{occ}^c and M_{occ}^{photo}.
M_{truth}	Canny filter applied to the ground-truth depth map.
F	Energy function for the regularization process with Markov Random Fields (MRFs).
E_1, E_2	Unary and binary terms of Eq. 7.
p, q	Neighbouring pixels.
$l_d(p), l_d(q)$	Depth labels values of pixels p and q, respectively. A range of values $[1, \ldots, N(l_d)]$ of depth levels is used.
$\nabla(E_c)$	Gradient of E_c.
k	Weighting factor to penalize the propagation in the regularization process when detecting an occlusion.
c_x, c_y	Physical size of the camera sensor.
$\widehat{Z}$	Ground-truth depth map.
Z^*	Estimated depth map.
W_x, W_y	Windows used to estimate the Structural Similarity Index.
μ_x, μ_y	Average values for windows W_x and W_y.

not only take into account the depth map information but also the error that is committed when the *all-in-focus* image is reconstructed.

REFERENCES

[1] S. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys, "Interactive 3D architectural modeling from unordered photo collections," *ACM Trans. Graph.*, vol. 27, no. 5, p. 159, 2008.

[2] J.-Y. Son, W.-H. Son, S.-K. Kim, K.-H. Lee, and B. Javidi, "Three-dimensional imaging for creating real-world-like environments," *Proc. IEEE*, vol. 101, no. 1, pp. 190–205, Jan. 2013.

[3] B. Javidi *et al.*, "Multidimensional optical sensing and imaging system (MOSIS): From macroscales to microscales," *Proc. IEEE*, vol. 105, no. 5, pp. 850–875, May 2017.

[4] S.-H. Wang, J. Sun, P. Phillips, and G. Zhao, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *J. Real-Time Image Process.*, vol. 15, no. 3, pp. 631–642, 2018.

[5] Y. Zhang, L. Wu, and G. Wei, "A new classifier for polarimetric SAR images," *Prog. Electromagn. Res.*, vol. 94, pp. 83–104, 2009.

[6] H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," *Opt. Lett.*, vol. 26, no. 3, pp. 157–159, 2001.

[7] G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

[8] H. E. Ives, "Optical properties of a Lippmann Lenticulated sheet," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 21, no. 3, pp. 171–176, 1931.

[9] T. Okoshi, "Three-dimensional displays," *Proc. IEEE*, vol. 68, no. 5, pp. 548–564, May 1980.

[10] N. Davies, M. McCormick, and L. Yang, "Three-dimensional imaging systems: A new development," *Appl. Opt.*, vol. 27, no. 21, pp. 4520–4528, 1988.

[11] F. Okano, H. Hoshino, J. Arai, and I. Yuyama, "Real-time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.*, vol. 36, no. 7, pp. 1598–1603, 1997.

[12] J. Arai *et al.*, "Integral three-dimensional television using a 33-megapixel imaging system," *J. Display Technol.*, vol. 6, no. 10, pp. 422–430, 2010.

[13] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: Sensing, display, and applications," *Appl. Opt.*, vol. 52, no. 4, pp. 546–560, 2013.
[14] M. Martinez-Corral and B. Javidi, "Fundamentals of 3D imaging and displays: A tutorial on integral imaging, lightfield, and plenoptic systems," *Adv. Opt. Photon.*, vol. 10, no. 3, pp. 512–566, 2018.
[15] T. H. Jen, X. Shen, G. Yao, Y. P. Huang, H. P. Shieh, and B. Javidi, "Dynamic integral imaging display with electrically moving array lenslet technique using liquid crystal lens," *Opt. Express*, vol. 23, no. 14, pp. 18415–18421, 2015.
[16] S. A. Benton and V. M. Bove, *Holographic Imaging*. Hoboken, NJ, USA: Wiley, 2008.
[17] M. Daneshpanah and B. Javidi, "Profilometry and optical slicing by passive three-dimensional imaging," *Opt. Lett.*, vol. 34, no. 7, pp. 1105–1107, 2009.
[18] M. Cho and B. Javidi, "Three-dimensional visualization of objects in turbid water using integral imaging," *J. Display Technol.*, vol. 6, no. 10, pp. 544–547, 2010.
[19] M. Cho, A. Mahalanobis, and B. Javidi, "3D passive photon counting automatic target recognition using advanced correlation filters," *Opt. Lett.*, vol. 36, no. 6, pp. 861–863, 2011.
[20] F. Okano, J. Arai, K. Mitani, and M. Okui, "Real-time integral imaging based on extremely high resolution video system," *Proc. IEEE*, vol. 94, no. 3, pp. 490–501, Mar. 2006.
[21] J. S. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," *Opt. Lett.*, vol. 27, no. 13, pp. 1144–1146, 2002.
[22] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *Int. J. Comput. Vis.*, vol. 38, no. 3, pp. 199–218, Jul. 2000.
[23] G. Vogiatzis, C. H. Esteban, P. H. S. Torr, and R. Cipolla, "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2241–2246, Dec. 2007.
[24] J. M. Sotoca, P. Latorre-Carmona, H. Espinos-Morato, F. Pla, and B. Javidi, "Depth estimation improvement in 3D Integral Imaging using a edge removal approach," *Pattern Anal. Appl.*, pp. 1–13, 2018. [Online]. Available: https://link.springer.com/article/10.1007/s10044-018-0721-4, doi: 10.1007/s10044-018-0721-4.
[25] A. Martínez-Usó, 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.*, vol. 12, no. 12, pp. 1715–1723, Dec. 2016.
[26] S.-H. Hong and B. Javidi, "Three-dimensional visualization of partially occluded objects using integral imaging," *J. Display Technol.*, vol. 1, no. 2, pp. 354–359, Dec. 2005.
[27] M. Zhang, Z. Zhong, and Y. Piao, "Visual quality enhancement of three-dimensional imaging reconstruction of partially occluded objects using exemplar-based image restoration," *J. Inf. Commun. Converg. Eng.*, vol. 14, no. 1, pp. 57–63, 2016.
[28] T.-C. Wang, A. A. Efros, and R. Ramamoorthi, "Depth estimation with occlusion modeling using light-field cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2170–2181, Nov. 2016.
[29] X. Xiao, M. Daneshpanah, and B. Javidi, "Occlusion removal using depth mapping in three-dimensional integral imaging," *J. Display Technol.*, vol. 8, no. 8, pp. 483–490, Aug. 2012.
[30] X. Shen, A. Markman, and B. Javidi, "Three-dimensional profilometric reconstruction using flexible integral imaging and occlusion removal," *Appl. Opt.*, vol. 56, no. 9, pp. D151–D157, 2017.
[31] B.-G. Lee, B. Ko, S. Lee, and D. Shin, "Computational integral imaging reconstruction of a partially occluded three-dimensional object using an image inpainting technique," *J. Opt. Soc. Korea*, vol. 19, no. 3, pp. 248–254, 2015.
[32] V. Vaish, R. Szeliski, C. L. Zitnick, S. B. Kang, and M. Levoy, "Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust methods," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2006, pp. 2331–2338.
[33] J. M. Sotoca, P. Latorre-Carmona, F. Pla, X. Shen, S. Komatsu, and B. Javidi, "Integral imaging techniques for flexible sensing through image-based reprojection," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 34, no. 10, pp. 1776–1786, 2017.
[34] M. Daneshpanah and B. Javidi, "Three-dimensional imaging with detector arrays on arbitrary shaped surfaces," *Opt. Lett.*, vol. 36, no. 5, pp. 600–602, 2011.
[35] J. Wang, X. Xiao, and B. Javidi, "Three-dimensional integral imaging with flexible sensing," *Opt. Lett.*, vol. 39, no. 24, pp. 6855–6858, 2014.
[36] Y. Wexler, A. Fitzgibbon, and A. Zisserman, "Bayesian estimation of layers from multiple images," in *Proc. Eur. Conf. Comput. Vis.*, 2001, pp. 487–501.
[37] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 673–680.
[38] P. Getreuer, "Rudin-Osher-Fatemi total variation denoising using split Bregman," *Image Process. On Line*, vol. 2, pp. 74–95, May 2012, doi: 10.5201/ipol.2012.g-tvd.
[39] S. Pertuz, D. Puig, M. A. Garcia, and A. Fusiello, "Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1242–1251, Mar. 2013.
[40] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1222–1239, Nov. 2001.
[41] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.
[42] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

JOSÉ MARTÍNEZ SOTOCA received the B.Sc. degree in physics from the Universidad Nacional de Educación a Distancia, Madrid, Spain, in 1996, and the M.Sc. and Ph.D. degrees in physics from the University of Valencia, Valencia, Spain, in 1999 and 2001, respectively.

His Ph.D. work was on surface reconstructions with structured light. He is currently an Assistant Lecturer with the Departamento de Lenguajes y Sistemas Informáticos, Universitat Jaume I, Castellón de la Plana, Spain.

He has collaborated in different projects, most of which are in the medical application of computer science. He has published more than 45 scientific papers in national and international conference proceedings, books, and journals. His research interests include pattern recognition and biomedical applications, including image pattern recognition, hyperspectral data, structured light, and feature extraction and selection.

PEDRO LATORRE-CARMONA received the B.S. degree in physics from the University of Valencia, Spain, in 1999, and the Ph.D. degree in computer science from the Polytechnical University of Valencia, in 2005. He is a Postdoctoral Researcher with the Departamento de Lenguajes y Sistemas Informaticos, Universitad Jaume I, Castellón de la Plana, Spain.

His current research interests are 3-D image analysis, feature selection and extraction, pattern recognition, multispectral (including remote sensing) image processing, colorimetry, and vision physics.

FILIBERTO PLA received the B.Sc. and Ph.D. degrees in physics from the Universitat de València, Spain, in 1989 and 1993, respectively. He is currently a Full Professor with the Departament de Llenguatges i Sistemes Informàtics, University Jaume I, Castellón de la Plana, Spain. He has been a Visiting Scientist with the Silsoe Research Institute, the University of Surrey, the University of Bristol, U.K., CEMAGREF, France, the University of Genoa, Italy, the Instituto Superior Técnico, Lisbon, Portugal, the Swiss Federal Institute of Technology, ETH-Zurich, the Idiap Research Institute, Switzerland, and the Technical University of Delft, The Netherlands.

He is currently the Director of the Institute of New Imaging Technologies, University Jaume I. His current research interests are color and spectral image analysis, visual motion analysis, 3-D image visualization, and pattern recognition techniques applied to image processing. He is a member of the Spanish Association for Pattern Recognition and Image Analysis, which is a part of the International Association for Pattern Recognition.

BAHRAM JAVIDI (S'82–M'83–SM'96–F'98) received the B.S. degree from George Washington University, and the M.S. and Ph.D. degrees from Pennsylvania State University, all in electrical engineering. He is the Board of Trustees Distinguished Professor at the University of Connecticut. He has over 1000 publications, including nearly 450 peer-reviewed journal article, over 450 conference proceedings, over 120 plenary addresses, keynote addresses, and invited conference papers. His papers have been cited 33 000 times according to the Google Scholar Citations (h-index=85, i10-index=537). He is a co-author of nine best paper awards.

Dr. Javidi received the Quantum Electronics and Optics Prize for Applied Aspects from the European Physical Society, in 2015. He has been named as a Fellow of several scientific societies, including OSA and SPIE. In 2010, he was a recipient of The George Washington University's Distinguished Alumni Scholar Award, the university's highest honor for its alumni in all disciplines. In 2008, he received a Fellow award by the John Simon Guggenheim Foundation. He received the 2008 IEEE Donald G. Fink Prized Paper Award among all (over 150) IEEE transactions, journals, and magazines. In 2007, The Alexander von Humboldt Foundation awarded him with the Humboldt Prize for outstanding U.S. scientists. He received the Technology Achievement Award from SPIE, in 2008. In 2005, he received the Dennis Gabor Award in Diffractive Wave Technologies from SPIE. He was a recipient of the IEEE Photonics Distinguished Lecturer Award twice, in 2003–2004 and 2004–2005. He was awarded the IEEE Best Journal Paper Award from the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY twice, in 2002 and 2005. Early in his career, the National Science Foundation named him a Presidential Young Investigator and he received The Engineering Foundation and the IEEE Faculty Initiation Award. He was selected, in 2003, as one of the nation's top 160 engineers between the ages of 30–45 by the National Academy of Engineering to be an Invited Speaker at The Frontiers of Engineering Conference, which was co-sponsored by The Alexander von Humboldt Foundation. He has been an alumnus of the Frontiers of Engineering of The National Academy of Engineering, since 2003. He has served on the Editorial Board of the PROCEEDINGS OF THE IEEE (ranked #1 among all electrical engineering journals), the advisory board of the IEEE PHOTONICS JOURNAL, and he was on the Founding Board of Editors of the IEEE/OSA JOURNAL OF DISPLAY TECHNOLOGY.