3D Point Cloud Reconstruction from Single Plenoptic Image

Francesca Murgia, Cristian Perra, *Senior Member, IEEE*, and Daniele Giusto, *Senior Member, IEEE*

Abstract — **Novel plenoptic cameras sample the light field crossing the main camera lens. The information available in a plenoptic image must be processed, in order to create the depth map of the scene from a single camera shot. In this paper a novel algorithm, for the reconstruction of 3D point cloud of the scene from a single plenoptic image, taken with a consumer plenoptic camera, is proposed. Experimental analysis is conducted on several test images, and results are compared with state of the art methodologies. The results are very promising, as the quality of the 3D point cloud from plenoptic image, is comparable with the quality obtained with current non-plenoptic methodologies, that necessitate more than one image.**

Keywords — **Light field reconstruction, Plenoptic reconstruction, 3D reconstruction.**

I. INTRODUCTION

REPRESENTING the real world on a digital screen is a very important topic in the area of computer graphics. An accurate 3D reconstruction of a real scene is needed for several applications such as engineering, architecture, design, videogames, virtual reality, just to mention some.

A good 3D model of the scene is the main objective for successive processing such as rendering, 2D/3D view extraction, distance measurements.

A 3D model of a real object can be obtained in several ways. The method currently producing very accurate results is based on capturing a point cloud representation of the object.

Point clouds are usually obtained by means of high cost technologies such as the total station, the laser scanner, professional digital cameras.

Moreover, the quality of the data captured so far depends on complicated setup of the capturing environment, high complexity processing with user intervention.

Francesca Murgia is with the Department of Electrical and Electronic Engineering, University of Cagliari, Via Marengo, 2, 09123 Cagliari, Italy (e mail: francesca.murgia@diee.unica.it).

Cristian Perra is with the Department of Electrical and Electronic Engineering, University of Cagliari, Via Marengo, 2, 09123 Cagliari, Italy (e mail: c.perra@unica.it).

Daniele Giusto is with the Department of Electrical and Electronic Engineering, University of Cagliari, Via Marengo, 2, 09123 Cagliari, Italy

An open research issue is to define novel methodologies to be used for the reconstruction of the 3D point cloud of a real object with low cost devices and low complexity algorithms.

Data captured by a plenoptic camera, thanks to the singular capturing technique, could bring to interesting applications in the field of 3D point cloud reconstruction.

Digital images captured by a plenoptic camera provide unique properties. A microlens matrix behind the camera digital sensor splits the light ray from the different directions. The sensor is able to sample this information that is called the light field of the scene.

The digital light field of the scene obtained so far can be subsequently processed by a software application. For example, it is possible to reconstruct 2D images at different focus points and it is possible to reconstruct the depth map of the scene.

The intrinsic properties of plenoptic images, such as the high number of light rays captured by the sensor allows the reconstruction of 2D images at different focus planes. Such focus planes are distributed from the far background to the near foreground. From these planes, it is possible to reconstruct the depth map using, for example, a Layer Depth Image (LDI) [1] of the scene to be used as a starting point for the 3D point cloud reconstruction of the captured scene.

Fig. 1 shows two images obtained by post processing a single plenoptic image. In Fig. 1(a) the image has been reconstructed setting the focus in the foreground of the scene, while in Fig. 1(b) the focus is in the background of the scene.

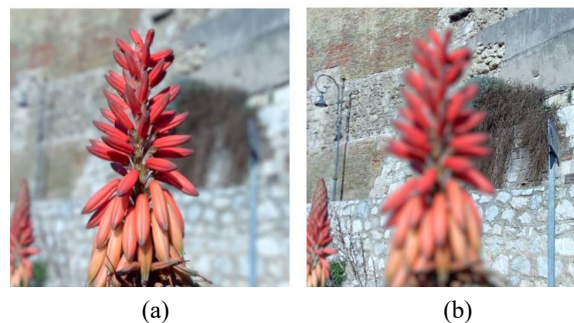(a)   (b)

Fig. 1 Different focus from a single shot. (a) Focus in the foreground. (b) Focus in the background.

One of the main contributions of this work is the proposal of an algorithm for the reconstruction of a 3D point cloud from a single plenoptic image.

The paper is organized as follows. The related work is briefly summarized in Section II. The description of the

plenoptic camera architecture is provided in Section III. The proposed algorithm is presented in Section IV. Section V discusses the obtained results. Section VI concludes the paper and suggests possible future developments of the work.

II. RELATED WORK

The reconstruction of a 3D point cloud is an intermediate step of several application areas. There are several approaches available in literature addressing the problem with different methodologies. In this paper, the problem addressed is the exploration of plenoptic images as source information for the reconstruction of 3D point cloud, from a single image. Continuous development and advances in 3D computer graphics, makes possible a detailed reconstruction of 3D models of a real object.

In the framework of image processing, the reconstruction of 3D models from images is a very challenging research area. Structure-from-motion (SFM) and dense multi-view 3D reconstruction (DMVR) are some of the main promising methodologies for 3D reconstruction.

SFM makes use of a set of unordered images, depicting a scene from different viewpoints. The 3D point cloud of the scene is reconstructed combining the image information with the photo camera parameters (focal length, image size, exposure time, ISO sensitivity, etc.).

The quality of a 3D point cloud can be increased by using professional cameras [2].

In the following sections, an open source software named VisualSFM [3] is used as a measurement tool in order to compare the proposed methodology with the image based rendering (IBR) [4] chosen as reference.

The methods discussed so far require a large number of images in order to reconstruct the 3D models of the objects.

Another approach is based on the reconstruction of the 3D model from a single image. Some of the proposed methods for the object reconstruction are [5]-[15].

Some of the methods require user interaction during all the process of 3D reconstruction [5], [12], [14]. In these cases, better results are obtained against a high number of user interventions.

Other methods impose constraints on the image type such as, for example, planarity and symmetry of the object to be modeled [7]-[9], [11], [12].

In some cases, user intervention becomes fundamental for the generation of the object's primitives [5], [14]. In other cases, it becomes necessary to manually draw the object silhouette before starting the 3D reconstruction process [6], [10], [13]. Differently from cited works, the method presented in this paper does not require object planarity or symmetry, and obtains appreciable results in non planar regions.

Another approach is based on the Epipolar Plane Images spatial direction. This method involves the use of two stereoscopic images of the same scene, so as to exploit the geometric differences for the reconstruction. This method is also used in [4].

The reconstruction of the 3D object requires an accurate estimation of the scene depth. Methods exploiting 2D gradients of subsets of light field data [16] can provide good results. A recent evolution of scene depth estimation exploits a stereo multi-vision approach [17] where the scene depth is computed from the calibration of different images of the same scene.
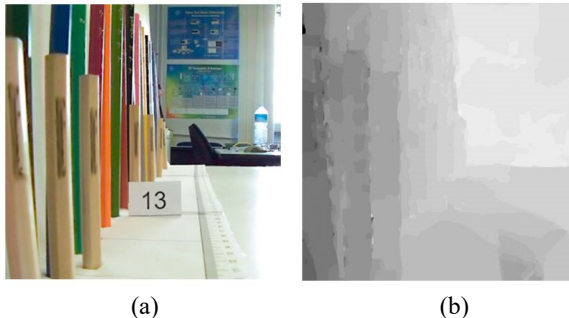

(a) (b)
Fig. 2 Files exported from the Lytro camera:
(a) all-in focus TIFF image; (b) PNG depth map.

III. PLENOPTIC CAMERA

A plenoptic camera is a photo camera that makes use of an array of microlenses, for capturing the 4D information of the light field of a scene.

If we consider the light affecting a pixel in a conventional digital image, the information gathered from the pixel will not tell anything about how the light has arrived at the pixel position. The lack of information related to the direction of the light entering the camera, does not allow further image processing. Instead, a plenoptic system, not only captures the image projection onto the sensor in terms of light intensity, but even the light ray directions are stored.

The plenoptic camera has been theorized by Lipmann [18]. Nevertheless, adequate technologies and sufficient computational capacity, for developing a plenoptic camera, have become available only in recent years.

In 2005 Ng [19], [20] developed a first consumer plenoptic camera, named Lytro [21] and commercialized it in 2006. Since then other companies started the development and selling of a different technology of plenoptic cameras. For example, the Raytrix [22] developed a camera defined 2.0 based on a microlens array capturing the information on three focal planes.

Other types of plenoptic cameras are available as consumer products or as testing devices in research centers. Nevertheless, the Lytro and the Raytrix are the main off-the-shelves products. Lytro and Raytrix cameras have different target applications. The first one is dedicated to consumer photography applications, while the second one is intended for research and industrial application. The applications that have given major prominence to plenoptic cameras are several. The main interests are related to the possibility of post processing digital refocusing [23] and the possibility to compute all-in-focus images from the plenoptic image [24].

Refocusing a scene allows for changing the image plane to a different position with respect to the main image plane. This process implies to possibility to determine the scene depths and to generate the depth maps of the scene [25], [26], [27], [28]. Fig. 2 shows an example of image captured with a Lytro camera (Fig. 2(a)) and the

corresponding depth map (Fig. 2(b)). Considering that raw plenoptic images are usually very large, another open issue is to explore data compression algorithms applied to plenoptic data [29], [32], [33].

IV. PROPOSED ALGORITHM

The input data for the proposed algorithm is originated by the Lytro cameras. The reconstruction of the 3D point cloud of the scene follows a sequential approach as the output of each processing block is the input for the following processing block.

Plenoptic data acquired from the plenoptic cameras are embedded with metadata providing additional information such as: shot date and time, exposition time, lens aperture, focal length and more.

Plenoptic cameras taken into consideration realize a preprocessing consisting in the computation of the depth map of the image. In particular the main inputs of the algorithm are an all-in focus image in TIFF format and a depth map of the scene in PNG format.

The proposed algorithm adopts a preprocessing step consisting in masking the subject of the image to be rendered. This step is accomplished by means of the image segmenter tool available from Matlab R2014b.

Mask creation is the only iteration required to the user. The steps of the proposed algorithm are:
A. Processing the depth map;
B. Edge detection;
C. Image fusion;
D. Object extraction;
E. Point cloud generation.

Fig. 3 shows an example of each plenoptic data processing step.

A. Processing the depth map

In the first step, the PNG depth map image (Fig. 3(b)) is processed. The resolution of the depth map is 16 bit. The higher the number of depth levels, the better the quality of the final 3D reconstruction. Nevertheless, experimental tests have shown that only very few depth planes, six on average, are usually available. This step is composed by two sub-steps. At first, in order to better separate the depth planes, a histogram stretching process is applied to the image

$$g_{ij} = \frac{f_{ij} - f_{min}}{f_{max} - f_{min}} \cdot 255 \qquad (1)$$

where f_{ij} is input image, f_{min} and f_{max} are respectively the minimum and maximum of the gray intensity in the histogram values.

Then, a 15×15 average filter is applied. The main purpose of this step is the reduction of the distance between adjacent depth layers

$$a_{ij} = \frac{1}{N} \sum g_{ij} w_{ij} \qquad (2)$$

where $N = 255$, g_{ij} is the image previously obtained and w_{ij} is the average filter.

B. Edge detection

The second step consists in the contour extraction by the application of a Sobel filter in the TIFF image.
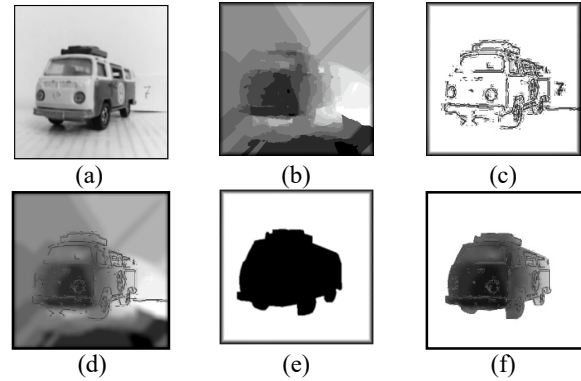

(a) (b) (c)

(d) (e) (f)

Fig. 3. Example of plenoptic data processing. (a) All-in focus input image in TIFF format; (b) Depth map input image in PNG format; (c) Edge detection; (d) Image fusion; (e) Mask generated with image segmenter (Matlab); (f) Object extraction.

Convolution takes place between the TIFF image (Fig. 3(a)) and the two Sobel masks, M_X and M_Y, defined as follows:

$$M_X = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, M_Y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad (3)$$

The convolution is related to the following equation:

$$G_X = a_{ij} M_X \qquad (4)$$

$$G_Y = a_{ij} M_Y \qquad (5)$$

The final image will be then obtained as follows:

$$b_{ij} = K(G_X^2 + G_Y^2) \qquad (6)$$

where K is a normalized parameter for gradient levels, respect depth map levels.

Final result is shown in Fig. 3(c).

C. Image fusion

In the third step, the depth map, processed in the first step (Fig. 3(b)), and the edge image (Fig. 3(c)) are fused together, combined by means of absolute difference operation, in the following way

$$c_{ij} = \left| a_{ij} - b_{ij} \right| \qquad (7)$$

where a_{ij} and b_{ij} are obtained from Equations (2) and (6).

Thanks to the extraction of the contours, it is possible to insert details of the depth map, as shown in (Fig. 3(d)).

D. Object extraction

In the fourth step, image segmentation is obtained using the Matlab Image Segmenter. It should be observed that, in the depth map (Fig. 3(b)), the object is not completely isolated with respect to the background. This problem can be overcome to some extent by manually masking the object with respect to the background. The user will manually define the object contour to be reconstructed. The resulting mask is shown in Fig. 3(e).

Then, the output of the third step (Fig. 3(d)) and the output of the fourth step (Fig. 3(e)) are summed together. The object is extracted by adding the image obtained from the third step and the object mask as follows:

$$d_{ij} = \left| k_{ij} + c_{ij} \right| \qquad (8)$$

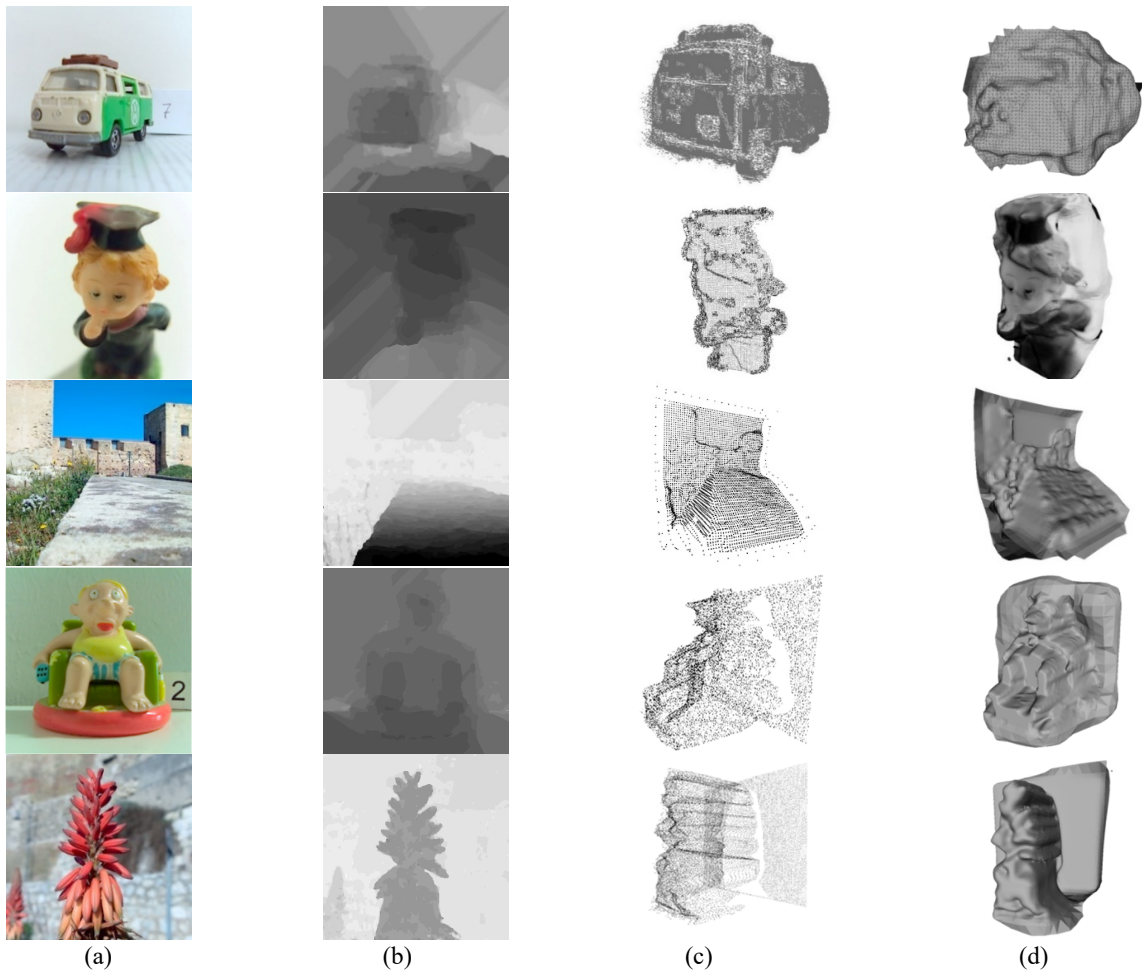|			
(a)	(b)	(c)	(d)

Fig. 4. Results obtained with images extracted from Lytro camera: columns (a) input TIFF image; column (b) input depth map; column (c) output point cloud; column (d) point cloud elaborated in MeshLab.

where k_{ij} is the mask obtained using the Matlab image segmenter. The extraction result is shown in (Fig. 3(f)).

E. Point cloud generation

In the fifth step the depth of the points in the object to be reconstructed is computed. For each pixel at coordinates (x, y) a third information representing the depth of the pixel (z) is computed from the depth map of the scene. The 3D point cloud obtained so far is characterized by a high level of detail, more or less 800 points/cm^2. The reason of this high level of detail is because in this method for each pixel a corresponding 3D point is generated in the model space.

Nevertheless, considering that the information is obtained from a single shot, the object is affected by perspective deformation. In order to reduce the perspective distortion error, a stereoscopic model system using two or more images taken from different views is needed. Consequently the problem of perspective deformation is currently under analysis and will be addressed in future research. The proposed method produces a 3D point cloud that can be further processed by modeling software.

V. RESULTS

This section presents experimental analysis and obtained results. A data set has been realized using a Lytro camera. The data set is composed by images with the following characteristics: TIFF image: 328×328 pixel, 24 bit; PNG image: 328×328 pixel, 16 bit.

Fig. 4 shows some of the obtained results. Starting from the input images, the all-in focus image (Fig. 4, column (a)) and the depth map (Fig. 4, column (b)), this method is able to produce a cloud point in PLY format, shown in (Fig. 4, column (c)).

Then the cloud point is imported into modelling software. Open source software, named Meshlab [30], has been used in order to process the model and produce the mesh and the texture. Later, in order to compare the proposed methodology with the image based rendering (IBR) methodology, open source software, named VisualSFM, has been chosen as a measurement tool. The same object has been modelled first with the plenoptic method from a single image, then with the IBR method using images taken from different points of view with a conventional camera, in order to evaluate the quality of the obtained results.
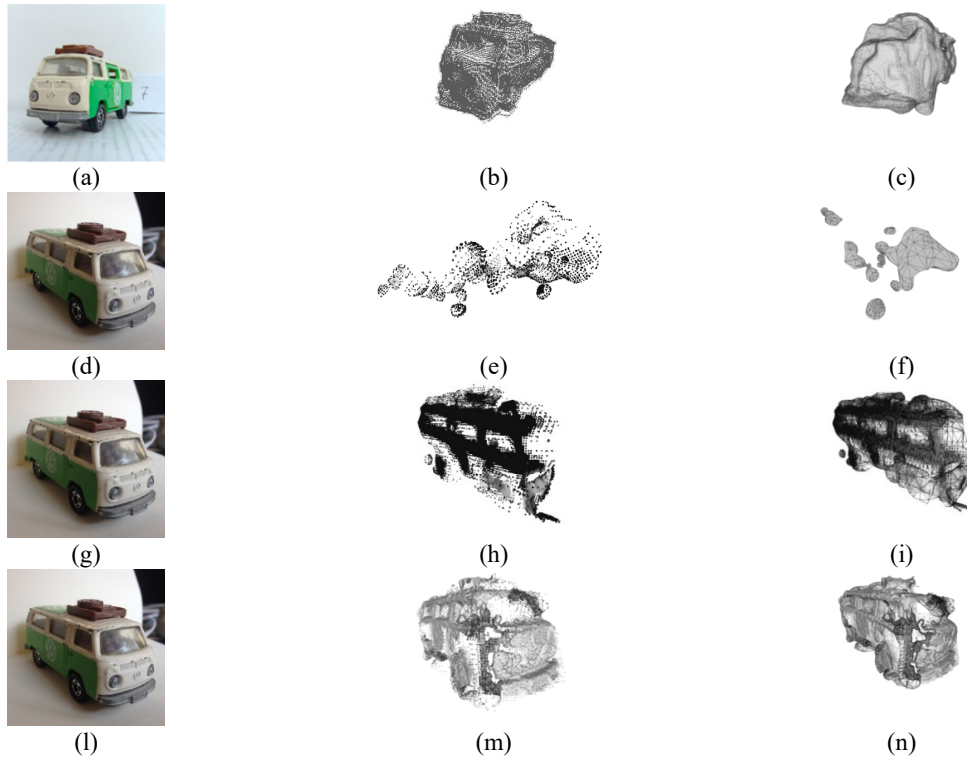
Fig. 5. Comparison between the proposed method and the IBR method:
(a) Plenoptic input image; (b) Output cloud point from plenoptic image; (c) Meshlab model from plenoptic image;
(d) One of the two stereoscopic conventional input images; (e) Output point cloud from two stereoscopic images; (f)
Meshlab model from the two stereoscopic images; (g) One of the three stereoscopic conventional input images; (h)
Output point cloud from three stereoscopic images; (i) Meshlab model from the three stereoscopic images, (l) One of the
four stereoscopic input images, (m) Output point cloud from the four stereoscopic images, (n) Meshlab model from the
four stereoscopic images.

TABLE 1: LOD COMPARISON BETWEEN RECONSTRUCTION FROM PLENOPTIC IMAGE (LYTRO)
AND 5 SHOT IMAGE BASED RENDERING.

3D METHOD	number of vertices	number of faces	dimension (kb)	processing time (min)	total time (min)
Proposed	30415	60745	1247	3.11(●)+1.50(♦)	4.60
IBR (1 shot)	-	-	-	-	-
IBR (2 shots)	3015	5990	124	1.55(□)+0.47(♦)	2.02
IBR (3 shots)	24635	49089	913	3.11(□)+0.43(♦)	3.54
IBR (4 shots)	43988	87646	1629	6.33(□)+3.16(♦)	9.49

(●) = MATLAB, (♦) = MESHLAB, (□) = VISUALSFM

This comparison is shown in (Fig. 5). It can be seen that to obtain a model of the object with acceptable visual quality, at least four different views of the object are necessary.

On the contrary, with the presented method it is possible to have a clear and detailed representation of the object taken with the plenoptic camera from a single image, even if it is limited to the visible part. The same comparison is also presented in Table 1, where the same models of Fig. 5 are represented in a numerical format.

In computer graphics, the level of detail (LOD) [31] is a measure based on the number of vertexes or surfaces composing an object, because curved surfaces are better defined if composed by a high number of vertexes and surfaces. In general, the LOD is adopted in order to define the number of polygons providing a given quality of the 3D object model, depending on the distance between the object and the point of view.

The average reconstruction time is 5 minutes processing the data with Matlab and Meshlab, and using a 1.50 GHz, 4GB RAM personal computer.

Table 1 shows that for a number of vertices that approximates those of the proposed method, the processing time is nearly doubled. Therefore, from the obtained results it can be seen that the proposed point cloud method generates a large number of vertices by processing a single image. It should be observed that for a conventional IBR method it is impossible to process the reconstruction from a single image and hence the IBR (1 shot) row is empty.

This is possible thanks to the direct relation that exists between pixels and number of points/cm^2. This relation is shown in Table 2. The first column shows the resolution of the plenoptic image, the second column shows the size of

the plenoptic image, the third column shows the number of vertex points extracted from the 3D model. It should be noted that the result in the third column shows that the number of vertex points is the same as the image resolution ($328 \times 328 = 107584$). The number of vertex divided by the image surface area (reported in column four) results in the point density reported in column five ($107584 / 133.86 \approx 804$). This is the reason why the number of points in the 3D model is directly proportional to the image resolution and shows that the level of detail of the 3D model, obtained with the plenoptic method, is nearly 800 points/cm^2.

TABLE 2: PLENOPTIC IMAGE CHARACTERISTICS
AND NUMBERS CLOUD POINT EVALUATION

Resolution (pixels)	Size (cm)	N° model Points	Surface Area (cm^2)	Point Density ($1/cm^2$)
328×328	11.57×11.57	107584	133.86	804

VI. CONCLUSIONS

An algorithm for 3D point cloud reconstruction from a single plenoptic image has been proposed in this paper. The process of 3D reconstruction of an object, from a single photo camera shot, has been developed exploiting the intrinsic characteristics of the plenoptic images. The algorithm requires little user intervention. The proposed method is able to model a portion of the entire object. The accuracy of the proposed method has been measured in terms of level of detail information. Future research will explore how to reconstruct entire objects, capturing the scene information with more plenoptic images while minimizing the number of images with respect to stereoscopic based methods.

REFERENCES

[1] J. Shade, S. Gortler, L.-W. He, R. Szeliski, "Layered depth images" in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM, 1998, pp. 231–242.

[2] K. Anestis, B. Vidmar, and F. Arnaoutoglou. "Performance evaluation of a multi-image 3D reconstruction software on a low-feature artefact," *Journal of Archaeological Science* 40.12 (2013): 4450-4456.

[3] C. Wu, "VisualSFM: A visual structure from motion system," vol. 9, 2011. Available: http://edu/ccwu/vsfm

[4] K. Rzazewska and M. Luckner, "3D model reconstruction and evaluation using a collection of points extracted from the series of photographs" in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on. IEEE*, 2014, pp. 669–677.

[5] T. Chen, Z. Zhu, A. Shamir, S.-M. Hu, and D. Cohen-Or, "3-sweep: extracting editable objects from a single photo," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, pp. 195, 2013.

[6] D. Jelinek and C. J. Taylor, "Reconstruction of linearly parameterized models from single images with a camera of unknown focal length," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 7, pp. 767–773, 2001.

[7] Z. Li, J. Liu, and X. Tang, "A closed-form solution to 3D reconstruction of piecewise planar objects from single images," in *Computer Vision and Pattern Recognition*, 2007. CVPR'07. IEEE Conference on. IEEE, 2007, pp. 1–6.

[8] P. Sturm and S. Maybank, "A method for interactive 3D reconstruction of piecewise planar objects from single images," in *The 10th British Machine Vision Conference* (BMVC'99). The British Machine Vision Association (BMVA), 1999, pp. 265–274.

[9] T. Xue, J. Liu, and X. Tang, "3-d modeling from a single view of a symmetric object," *Image Processing, IEEE Transactions on*, vol. 21, no. 9, pp. 4180–4189, 2012.

[10] L. Zhang, G. Dugas-Phocion, J.-S. Samson, and S. M Seitz, "Single-view modelling of free-form scenes," *The Journal of Visualization and Computer Animation*, vol. 13, no. 4, pp. 225–235, 2002.

[11] Z. Changqing, et al. "Sketch-based 3-D modeling for piecewise planar objects in single images." *Computers & Graphics* 46 (2015): 130-137.

[12] N. Jiang, P. Tan, and L.-F. Cheong, "Symmetric architecture modeling with a single image," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5, pp. 113, 2009.

[13] E. Toppe, C. Nieuwenhuis, and D. Cremers, "Relative volume constraints for single view 3D reconstruction," in *Computer Vision and Pattern Recognition* (CVPR), 2013 IEEE Conference on. IEEE, 2013, pp. 177–184.

[14] F. Yan, M. Gong, D. Cohen-Or, O. Deussen, and B. Chen, "Flower reconstruction from a single photo," in *Computer Graphics Forum*. Wiley Online Library, 2014, vol. 33, pp. 439–447.

[15] F. Murgia, D. Giusto, C. Perra. "3D reconstruction from plenoptic image," *Telecommunications Forum Telfor (TELFOR), 2015 23rd*. IEEE, 2015.

[16] D. Dansereau and L. Bruton, "Gradient-based depth estimation from 4D light fields," in *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*. IEEE, 2004, vol. 3, pp. III–549.

[17] Y. Zhan, et al. "Line assisted light field triangulation and stereo matching." *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013.

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

[19] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, 2005.

[20] R. Ng, Digital light field photography, Ph.D. thesis, Stanford University, 2006.

[21] Ren Ng, "Lytro web site". Available: https://www.lytro.com

[22] C. Perwass and L. Wietzke, "Single lens 3D-camera with extended depth-of-field," in *Proc. SPIE*, 2012, vol. 8291, p. 829108.

[23] F Murgia, SB Fernandez, DD Giusto, and C Perra, "Unfocused plenoptic camera calibration," in *Telecommunications Forum Telfor (TELFOR), 2014 22nd. IEEE, 2014*, pp. 517–520.

[24] T. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 5, pp. 972–986, 2012.

[25] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "Plenoptic depth map in the case of occlusions," in *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics*, 2013, pp. 86671S–86671S.

[26] J. P. B. L. Custodio, "Depth estimation using light-field cameras," 2014.

[27] D. Reddy, J. Bai, and R. Ramamoorthi, "External mask based depth and light field camera," in *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*. IEEE, 2013, pp. 37–44.

[28] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu, "Line assisted light field triangulation and stereo matching," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2792–2799.

[29] C. Perra, "Lossless plenoptic image compression using adaptive block differential prediction," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, April 2015, pp. 1231–1234.

[30] P. Cignoni, M. Corsini, and G. Ranzuglia, "Meshlab: an open-source 3D mesh processing system," *Ercim news*, vol. 73, pp. 45–46, 2008.

[31] A. Koutsoudis, B. Vidmar, G. Ioannakis, F. Arnaoutoglou, G. Pavlidis, and C. Chamzas, "Multi-image 3D reconstruction data evaluation," *Journal of Cultural Heritage*, vol. 15, no. 1, pp. 73–79, 2014.

[32] C. Perra, "On the coding of plenoptic raw images," *Telecommunications Forum Telfor (TELFOR), 2014 22nd*. IEEE, 2014.

[33] A. Vieira, H. Duarte, C. Perra, L. Tavora, P. Assuncao, "Data formats for high efficiency coding of Lytro-Illum light fields," in *International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2015.