Focused plane segmentation in integral imaging scene reconstruction

Author: Pol Navarro Pérez.
Advisors: Artur Carnicer González.
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Abstract: Computational methods of Integral Imaging allow for the reconstruction of 3D scenes from information extracted from 2D arrays of elemental images. By combining the information from these images, the scene can be decomposed into different planes, enabling focus on the desired depth. The proposed method segments focused regions at each plane of the reconstructed scene, thereby identifying objects at specific depths. This approach allows for the extraction of information from each plane of interest and significantly reduces the effects of occlusions between planes. As a result, our method enhances the accuracy and efficiency of object segmentation by accurately distinguishing between focused and unfocused regions across different depth planes.

I. INTRODUCTION

Image segmentation based on blur level is a daily application in numerous fields. Whether for distinguishing focused objects, depth estimation, or tracking moving objects. Methods for this type of segmentation are constantly improving, implementing increasingly complex codes. This process allows for the extraction of information from a specific part of an image.

We have reconstructed a 3D scene, identified planes providing information, and observed the effect of occlusions through such segmentation based on blur. The blurriness of a part of the image implies the loss of information in it. Many efforts focus on improving the different phases of blur-based segmentation to minimize information loss. These phases include preprocessing, the application of one or several blur operators, pixel classification as blurry or focused, and post-processing of the obtained results. Numerous studies have explored many of these operators, examining in which situations each one is better suited.

The objective of this work is to achieve segmentation of different planes, focused at various distances. We will commence with the construction of a 3D scene from a set of 2D images. Through the reconstruction of the integral image, we will be able to extract information from different planes based on depth and observe the effect of occlusions. The scene will contain several planes of interest with various objects from which we aim to gather information. Once the images of the planes of interest are constructed, objects within each plane will appear focused while the rest will appear blurred. The detailed steps for blur-based segmentation, which will be discussed in the following section, will enable us to observe the effect of occlusions and determine the extent to which we can extract relevant information.

The structure of the paper is as follows: Section 2 will briefly detail the steps taken for the reconstruction of the 3D image. In Section 3, we will explain the process that has led us to segment the objects with relevant information for the different planes of interest in the scene. The final results will be presented in Section 4. Finally, Section 5 will outline the conclusions of the study.

II. BACKGROUND

The human visual system perceives the depth of objects through binocular disparity. This occurs because the two eyes observe the same scene from slightly different viewpoints due to the distance between them. The first approach to designing a system capable of capturing the plenoptic field of 3D objects was proposed by Lippmann, who suggested a multi-view camera system [1]. This system involved inserting an array of lenses in front of a light sensor, allowing for the capture of two-dimensional images from different perspectives, called elemental images (EIs). To avoid the superposition of these images, a set of physical barriers was required. The elemental images contained angular information corresponding to the rays passing through the vertex of the corresponding lens.

In the capture of plenoptic images, it is considered that the rays impacting the center of the camera pixels are the most significant. This implies that each pixel of a conventional photographic camera captures a plenoptic field confined to a segment of constant spatial coordinate but with a variable angular coordinate. The irradiance distribution in the image taken by the camera can be calculated by angular integration of the plenoptic function, using the Abel transform for a plane [2].

Additionally, in the literature, a method for calculating voxel size based on the Gaussian beam distribution model has been proposed [3], which could be relevant for understanding how precise measurements of irradiance distribution in plenoptic images can be obtained.

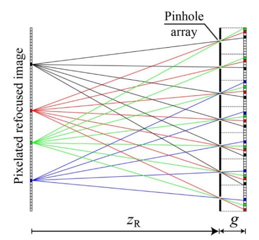

FIG. 1: Pinhole array approach to the back-projection algorithm [2]. © 2018 Optical Society of America.

To reconstruct the 3D image from the elemental images, we need to propagate the rays coming from the elemental images through a similar microlens array used for the recording. This process generates reverse propagation to form a 3D image where the object is originally located. In this work, we have used the elemental images from [4], extracted with the polarimetric integral imaging system they proposed.

III. METHODOLOGY

A. Computational Reconstruction

Starting from this point, we are interested in extracting the different angular information contained in each elemental image of the scene. Computationally, this will be done using the Shift and Sum method [2]. Defining $I_{k,l}(x,y)$ as the pixel value at (x,y) for the elemental image (k,l), we can find the images for a depth z according to:

$$I(x,y,z) = \sum_{k=0}^{EI_x-1} \sum_{l=0}^{EI_y-1} I_{k,l}\left(x - k\frac{N_x p f}{c_x z}, y - l\frac{N_y p f}{c_y z}\right) \tag{1}$$

To perform this operation, we need to use the parameters used in [4], as shown in the following table:

Elemental Image	$EI_x = 5,\ EI_y = 5$
Distance among cameras	$p = 5$mm
Camera focal length	$f = 50$mm
Sensor dimensions	$C_x = 36$mm, $C_y = 24$mm
Pickup distance	$z = [440, 540, 690]$mm
Number of pixels of elemental images	$N_x = 1248,\ N_y = 832$

TABLE I: Values of the variables in the experimental setup [4].

With the aim of centering the elemental image at the center of the overlap, we have introduced the variable change $k'\ (p') = k\ (p) + 2$. From this point on, the relative shift of the elemental images will vary according to the depth, variable z, that we want to focus on. Through computational reconstruction, each z value corresponds to all points on the plane with a consistent pixel displacement, known as disparity [2]. Consequently, the resulting computational Integral Imaging is a composite of all points sharing the same disparity, achieved through summation and averaging.

In the scene, objects have been placed at different depths, some in front of others. To study these occlusions, we have worked with the planes where these objects of interest are located; a tree, two cars, and some bushes in the background.

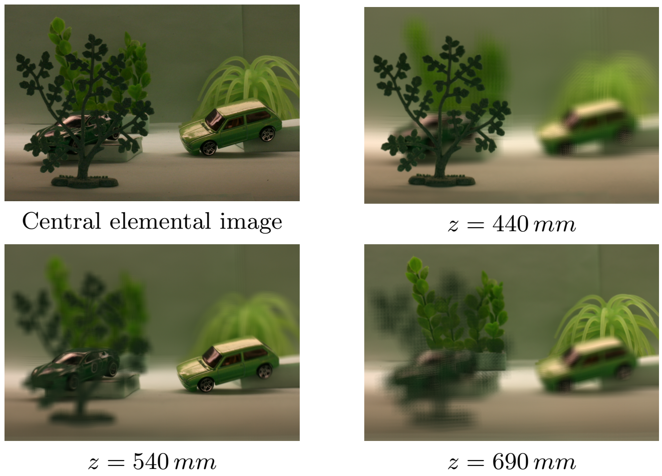

FIG. 2: Central elemental image and reconstructed images of planes at different depths using the Shift and Sum algorithm.

B. Blur segmentation

1. *Preprocessing*

Starting from the depth images Fig.4 a), we are interested in highlighting the finer objects before applying blurring operators. To achieve this goal, we have implemented a dilation operation using a structural element in the shape of an ellipse and a Top-Hat operation, which highlights the finer elements with higher intensity than their surroundings.

To conclude this phase, we have applied the Discrete Wavelet Transform (DWT). This technique decomposes signals and images into different scales and orientations, allowing the capture of fine details and global features. By applying DWT before blurring operators such as smoothing filters, undesired high-frequency components, like noise, can be effectively removed [5].

Equation 2 describes the convolution operation between an image I and a filter F, where $I_{\text{conv}}(i,j)$ is the resulting value of the convolution at position (i,j) of the convolved image:

$$I_{\text{conv}}(i,j) = \sum_{m=0}^{M-a} \sum_{n=0}^{N-b} I(m,n) \cdot F(i-m, j-n) \tag{2}$$

$I(m,n)$ represents the values of the original image at position (m,n), and $F(i-m, j-n)$ represents the values of the filter at position $(i-m, j-n)$. At the equation (M,N) and (a,b) are the image and filter dimnesions resepectively. Both a high-pass filter and a low-pass filter have been applied.

2. *Blur operators*

In this phase, we will apply two blur operators that will allow us to establish a value for each pixel based on whether it is focused or not. For the selection of the different operators, we have based ourselves on the work carried out in [2].

The first of these will be the Laplacian variation. Given a certain neighborhood of pixels $\Omega(x, y)$, we characterize the central pixel (x, y) as follows:

$$M_{VL}(x,y) = \sum_{(i,j)\in\Omega(x,y)} (\Delta I(i,j) - \bar{\Delta I})^2 \tag{3}$$

Here, $\Delta I(i, j)$ represents the Laplacian of pixel (i, j), which measures the second spatial derivative of the intensity at that pixel. $\bar{\Delta I}$ denotes the mean Laplacian value computed over the neighborhood $\Omega(x, y)$ surrounding pixel (x, y). As shown in [6], applying this magnitude gradient of the second derivative allows us to find the degree of focus. The size of the neighborhood used in Eq.3 has been 21 pixels for all images.

The other operator we have employed is the Local Binary Pattern. Focused regions exhibit a greater distinction in the obtained binary patterns than defocused areas [7]. Computationally, this is achieved by studying the blurriness measure of each pixel x_c, y_c as follows:

$$\begin{aligned} \mathrm{LBP_{P,R}}(x_c, y_c) &= \sum_{p=0}^{P-1} S(n_p - n_c)\cdot 2^p \\ \text{where } S(x) &= \begin{cases} 1 \text{ if } |x| \geq T_{\mathrm{LBP}} \\ 0 \text{ if } |x| < T_{\mathrm{LBP}} \end{cases} \end{aligned} \tag{4}$$

Where n_c is the central pixel, n_p is the set of pixels in the neighborhood P, R is the radius of the circle enclosing the center, and T_{LBP} is the threshold. Similar to Eq.3, the size of the neighborhood is 21 pixels.

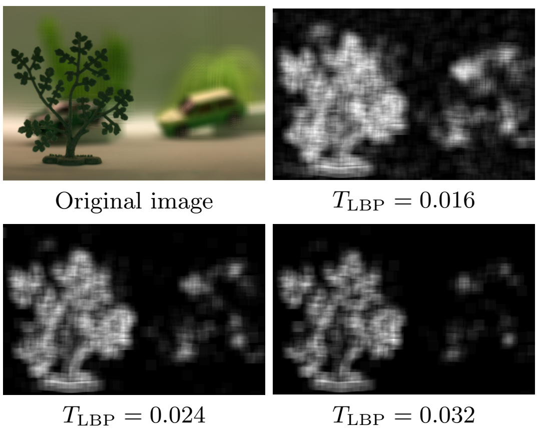

FIG. 3: Local Binary Pattern operator applied on plane 1 (z = 440mm) with different values for the threshold T_{LBP}.

As observed in Fig.3, the threshold determines how permissive we are when considering whether it is focused or not. In each image, different values for T_{LBP} have been tested, retaining those that yielded the best results, as shown in Table.II.

Depth (mm)	Threshold T_{LBP}
440	0.016
540	0.020
690	0.010

TABLE II: Threshold values based on the image depth.

At this point, we have two blur maps that show values in different ranges. In order to combine both , we have normalized each of the images $M(x, y) \in$ [0,1] according to

$$\hat{M}(x,y) = \frac{\mathrm{M(x,y) - min(M)}}{\mathrm{max(M) - min(M)}}, \tag{5}$$

where min(M) and max(M) represent the minimum and maximum values of M respectively.

3. *Clasification and post-processing*

At this point, we will need to binarize the image based on whether we consider the pixel to be focused or not. We have used a new threshold, different for each image, to determine if the pixel belongs to the focused region or not.

Other works mention methods involving a double threshold, and unclassified pixels are refined through a second step [7].

In the post-processing of the binarized image, the `remove_small_holes` and `remove_small_objects` functions from the `skimage` library were applied. These operations were crucial for improving the segmentation quality by removing small undesired regions. In both cases, a circular structural element was used.

IV. RESULTS

Using the binarized images as masks and overlaying them onto the corresponding section of the original image, we can obtain the results we are looking for.

To compare these results with the ideal objective, we manually extracted the focused region as ground truth. The Fig.4 illustrates the entire process mentioned in this work. As can be observed, we have worked with thresholds that minimize noise as much as possible. This has affected us in some cases, considering pixels as unfocused when they are indeed focused.

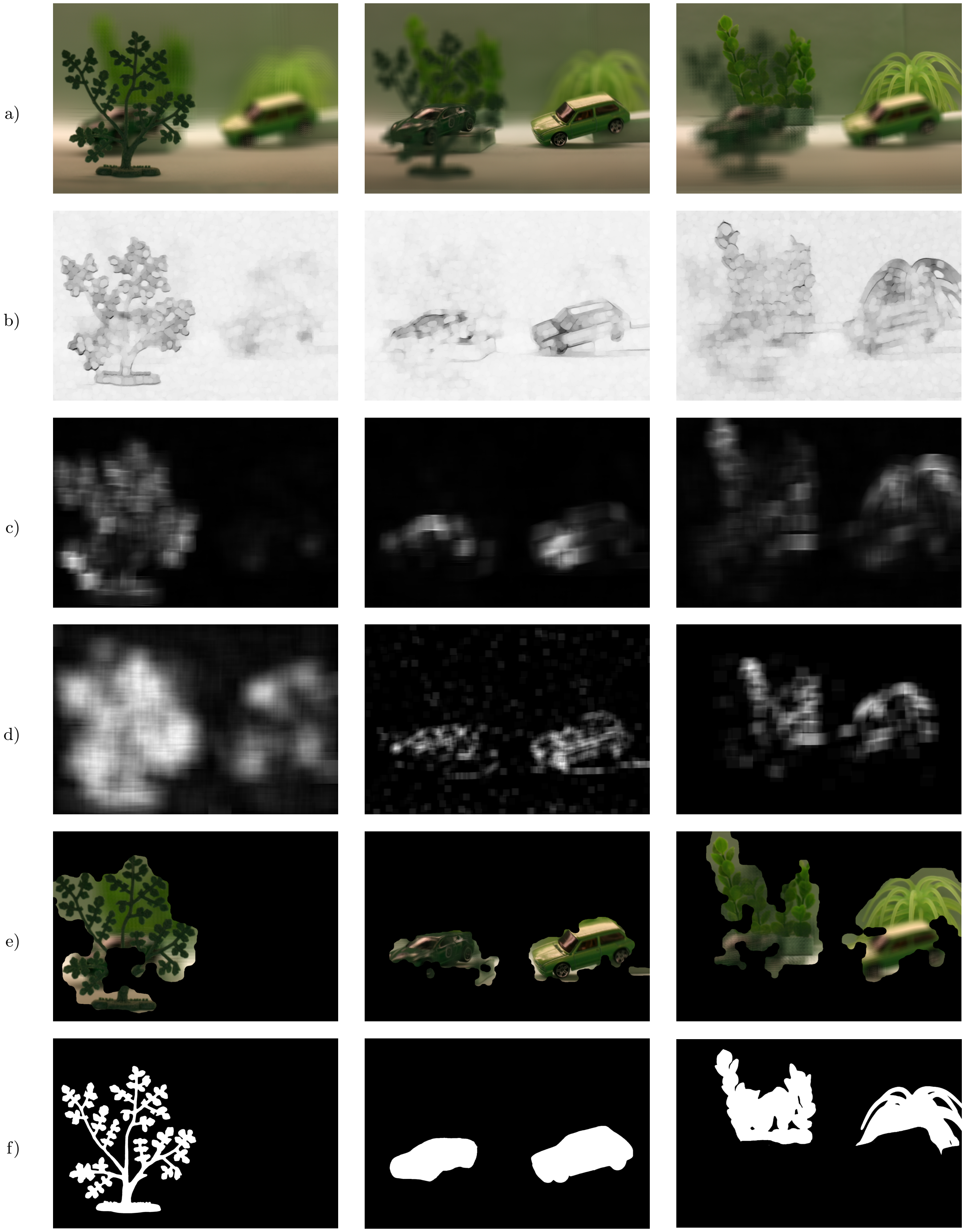

FIG. 4: Results of the entire process performed: a) Refocused images after the Shift and Sum algorithm. b) Images obtained from applying preprocessing, the Top-Hat and DWT algorithm. c) Variation of the Laplacian blur map. d) Local Binary Pattern blur map. e) Mask applied to the initial image. f) Ground truth.

Analyzing qualitatively the different planes, Fig.c) 4 depicts the tree located at $z = 440$ mm with a small occlusion, due to the presence of the car situated at a relative distance of 100 mm behind.

These occlusions remain evident in the following plane, $z = 540$ mm. Both cars are affected by occlusions from the tree as well as from the two bushes behind, resulting in a figure with poorer results.

In the plane of the background bushes, there is fairly good segmentation. In this case, the two cars situated $z = 150$ mm ahead are also visible. In this instance, the minimum threshold required to segment the corresponding plane has led us to include part of the previous plane as well.

Throughout the previous sections, we have adjusted various parameters such as thresholds, neighborhoods, or structural elements. We have aimed to segment each plane as much as possible, minimizing interference from others. Using less permissive values could have helped avoid occlusions, but we would not have extracted the entire plane.

V. CONCLUSIONS

A three-dimensional scene has been reconstructed from two-dimensional images. From the depth map, we have selected those planes that contain information of interest; z = 440, 540 and 690 mm.

Through a sequence of steps, we have segmented this information of interest and observed small differences with the ground truth due to the occlusions of other planes. Qualitatively, we can see how the occlusions correspond to the previous or subsequent plane. The occlusion becomes more pronounced as the adjacent plane gets closer.

In this work, we have carried out a process starting from flat images from which we have obtained a depth map. Subsequently, through an image processing process, we have segmented the focused part of the image, whose objects are located at the plane we are seeking. Using not very complex segmentation methods, we have managed to approximate the area that we consider focused.

Other blur operators could yield better results, being more sensitive to occlusions [5]. In other works, more complex systems are used for pixel classification once the blur map is obtained [8]. These improvements are more precise with the initial steps and allow avoiding the mentioned post-processing, thereby losing precision.

To access the source code, visit the following repository on GitHub:`https://github.com/polnape/Blur_segmentation.git`.

Acknowledgments

I would like to thank my advisor, Dr. Artur Carnicer Gonzalez. I am deeply grateful to my parents, sister, and grandparents, whose support, encouragement, and sacrifices have been essential along this journey.

The author is indebted to the MOSIS group, Electrical and Computer Engineering Department, University of Connecticut, for providing the integral image set. Additionally, thanks are extended to the authors of [7] for their contribution of part of the code used in the LBP operator.

[1] Lippmann, G. (1908). La photographie intégrale CR Séances Acad.

[2] Martínez-Corral, M., & Javidi, B. (2018). Fundamentals of 3D imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems. Advances in Optics and Photonics, 10(3), 512-566.

[3] Luo, C. G., Wang, Q. H., Deng, H., Gong, X. X., Li, L., & Wang, F. N. (2012). Depth calculation method of integral imaging based on Gaussian beam distribution model. Journal of Display Technology, 8(2), 112-116.

[4] Xiao, X., Javidi, B., Saavedra, G., Eismann, M., & Martinez-Corral, M. (2012). Three-dimensional polarimetric computational integral imaging. Optics express, 20(14), 15481-15488.

[5] Zhang, Y., & Hirakawa, K. (2013). Blur processing using double discrete wavelet transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1091-1098).

[6] Pech-Pacheco, J. L., Cristóbal, G., Chamorro-Martinez, J., & Fernández-Valdivia, J. (2000, September). Diatom autofocusing in brightfield microscopy: a comparative study. In Proceedings 15th International Conference on Pattern Recognition. ICPR-2000 (Vol. 3, pp. 314-317). IEEE.

[7] Yi, X., & Eramian, M. (2016). LBP-based segmentation of defocus blur. IEEE transactions on image processing, 25(4), 1626-1638.

[8] Su, B., Lu, S., & Tan, C. L. (2011, November). Blurred image region detection and classification. In Proceedings of the 19th ACM international conference on Multimedia (pp. 1397-1400).