sensors

MDPI

Article

Robust Depth Estimation for Light Field Microscopy

Luca Palmieri [1,*], Gabriele Scrofani [2], Nicolò Incardona [2], Genaro Saavedra [2], Manuel Martínez-Corral [2] and Reinhard Koch [1]

[1] Department of Computer Science, Christian-Albrecht-University, 24118 Kiel, Germany; rk@informatik.uni-kiel.de

[2] Department of Optics, University of Valencia, E-46100 Burjassot, Spain; gabriele.scrofani@uv.es (G.S.); nicolo.incardona@uv.es (N.I.); genaro.saavedra@uv.es (G.S.); manuel.martinez@uv.es (M.M.-C.)

* Correspondence: lpa@informatik.uni-kiel.de; Tel.:+49-431-880-4843

Received: 18 December 2018; Accepted: 22 January 2019; Published: 25 January 2019

check for updates

Abstract: Light field technologies have seen a rise in recent years and microscopy is a field where such technology has had a deep impact. The possibility to provide spatial and angular information at the same time and in a single shot brings several advantages and allows for new applications. A common goal in these applications is the calculation of a depth map to reconstruct the three-dimensional geometry of the scene. Many approaches are applicable, but most of them cannot achieve high accuracy because of the nature of such images: biological samples are usually poor in features and do not exhibit sharp colors like natural scene. Due to such conditions, standard approaches result in noisy depth maps. In this work, a robust approach is proposed where accurate depth maps can be produced exploiting the information recorded in the light field, in particular, images produced with Fourier integral Microscope. The proposed approach can be divided into three main parts. Initially, it creates two cost volumes using different focal cues, namely correspondences and defocus. Secondly, it applies filtering methods that exploit multi-scale and super-pixels cost aggregation to reduce noise and enhance the accuracy. Finally, it merges the two cost volumes and extracts a depth map through multi-label optimization.

Keywords: depth estimation; light field; microscope; stereo matching; defocus

1. Introduction

Light field microscopy was first introduced at Stanford in 2006 [1], and later improved in the same laboratory [2–4]. It consists of placing a microlens array (MLA) at the image plane of a conventional microscope, allowing for the capture of light field that records simultaneously both angular and spatial information of microscopic samples.

The main limitation of light field in microscopy is the spatial resolution [5]. To overcome this problem, a change of paradigm was necessary, so that the MLA is set, not at the image plane, but at the Fourier plane [6,7]. This realization of light field concept was named as Fourier integral Microscopy (FiMic). Light field microscopy has been used for several applications, such as brain imaging of neural activities in [8–10].

A common goal in microscopy is to estimate the three-dimensional structure of the observed sample. Industrial solutions reach a high accuracy of the reconstruction using different techniques as confocal microscopy [11], interferometry or variational focus [12], scanning electron microscopy (SEM) [13], optical profilometry [14,15], or stereo cameras [16].

However, these methods also present disadvantages in terms of real-time feasibility, sample preparation and costs. Light field concept provides a simpler and inexpensive approach that allows for many applications in real-time. On the other hand, because of the lack of texture and the presence of repetitive patterns that characterize microscopic samples, the task of extracting a depth map from such light fields presents many challenges.

To address these challenges, different methods have been proposed, whose main limitation still is the final resolution: in [17] optical flow and triangular meshes are used, and in [18] a Lytro consumer camera is used to build a light field microscope and a variational multi-scale optical flow algorithm is used to estimate the depth. An interesting approach to estimate depth of thin structure has been proposed in [19], obtaining high quality results at the price of targeting only a small subset of biological images.

Depth estimation from light field images has already been largely studied and different approaches were proposed, using epipolar plane images [20,21], angular or spatial information [22,23], focal stack [24] and correspondences cues combined [25], explicitly modeled occlusion-aware approaches [26], robust pseudo random field [27] and learning based costs [28].

In previous works we extended similar applications to the case of focused plenoptic cameras (plenoptic 2.0) using MLA with a large number of microlenses, where the raw image contains micro-images capturing a portion of the scene. The main targets were the lens selection process for such MLA-based cameras [29] and evaluating different methods for depth estimation [30].

However, FiMic images constitute another special case. Because of MLA position and structure, the light field is sampled differently. It samples the light field as a conventional plenoptic camera (plenoptic 1.0), while at the same time the perspective views are arranged on a hexagonal grid and exhibit large disparities between them, thus an interpolation to transform them into a rectangular grid would produce strong artifacts.

The main contribution of our work is the creation of a framework where different methods are unified and adapted to obtain a more robust and accurate approach. It takes the above-mentioned conditions into account and provides a method for recovering accurate depth information using a sparser light field, i.e., lower number of views with higher disparity shift. The method can be applied not only to FiMic, but also to images acquired with conventional light field cameras, by choosing only a subset of the perspective views, or to sparser light fields.

The structure of the rest of the paper is as follows: in Section 2, the Fourier Integral Microscope will be described; in Section 3, the depth estimation workflow is explained; in Section 4, a Comparative Performance Analysis to prove the quality of the proposed method is performed; then in Section 5, a potential application is shown to enhance the importance of the contributions; and finally in Section 6, a brief summary of the proposed work is given.

2. Fourier Integral Microscope

The presented work takes as input light field microscopy images. The light field microscope used here is the FiMic described in [7]. In the FiMic design shown in Figure 1 the MLA is conjugated with the aperture stop (AS) of the microscope objective (MO). In this way, the sensor, that is placed at the focal plane of the MLA, captures the perspective views directly.

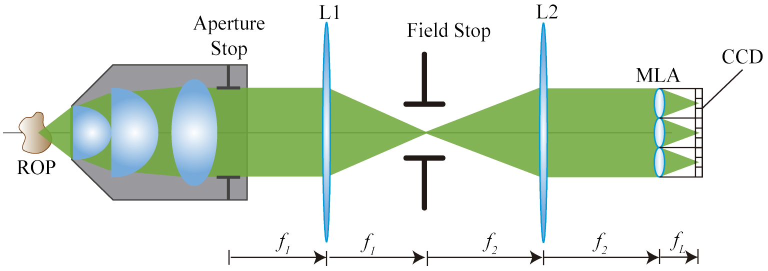


Figure 1. This is the schematic of the FiMic design. From left to right it is possible to distinguish the object, the microscope objective, two lenses (L1 and L2), one field stop, the MLA and the CCD sensor.

Changing the focal length of the two lenses L1 and L2, the designer can change the number of microlenses that fit in the diagonal of the AS. This affects the resolution limit (r) and the depth of field (DOF) as follows [7]:

$$r = M_n \frac{\lambda}{2NA} \tag{1}$$

$$DOF = \frac{5}{4} \frac{\lambda}{NA^2} M_n^2. \tag{2}$$

In these equations λ is the wavelength of the incoming light, NA is the numerical aperture of the MO and M_n is the number of microlenses that fit in the diameter of the AS. Note that usually the resolution capacity of an optical system is evaluated in terms of r^{-1}. It must be underlined that increasing M_n has two effects; a decrease of the resolution capacity and an increase of the DOF. An optimal setup for 3D microscopy aims to reach the highest resolution and the largest DOF, but as previously stated, these two factors are inversely affected by M_n. Therefore, depending on the sample to be reconstructed, M_n will be chosen according to the resolution and DOF required.

A typical example of an image acquired with this microscope can be seen in Figure 2, which includes seven elemental images, i.e., perspective views. These images can be seen as a particular case of a multi-view stereo imaging system, where different viewpoints are arranged on an hexagonal grid, having images aligned along at least three epipolar lines. This makes it suitable for correspondence matching.

Moreover, from such views a focal stack of refocused images can be extracted, by overlapping to the central one a shifted version of the elemental images, where the shift has to be oriented towards the center and will be the measure of the depth of the focal plane obtained. The depth of refocusing (z_R), with respect to the focal plane of the MO (considered as $z_R = 0\mu m$) can be calculated with the following formula [7]:

$$z_R = s \frac{f_{MO}^2}{f_L} \left(\frac{f_2}{f_1} \right)^2 \frac{\delta}{p}. \tag{3}$$

Here f_{MO}, f_L, f_2 and f_1 are respectively the focal lengths of the MO, MLA, L2 and L1. Besides, δ is the pixel's pitch, p is the MLA's pitch, and s is the integer number of shifted pixels applied to the focal stack. The use of defocus cue is then directly applicable to the generated focal stack. Due to these basic considerations, our approach aims to create a depth map by combining these two types of vision cues that are suitable for light field images.

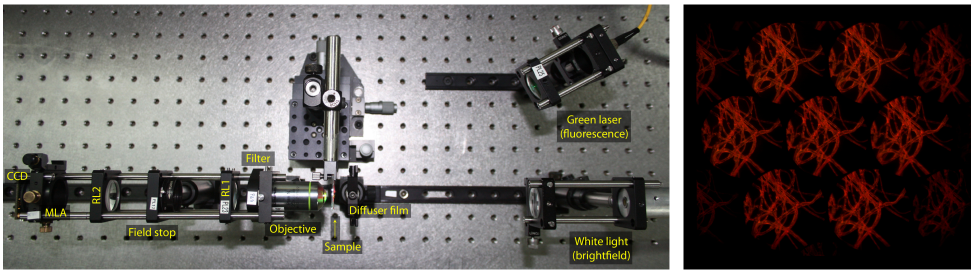


Figure 2. The setup used for the acquisition with the fluorescence laser used to illuminate the samples, and a sample output image acquired with such setup, from where the seven elemental images are visible.

3. Depth Map Calculation

The presented work builds on several successful ideas proposed for the depth estimation. The core consists in combining different visual cues, namely focus and correspondences, as in [25], to create a more versatile method, but it differs in the depth estimation.

The main contribution of this section is the combination of existing ideas and novel implementations. As it is possible to see in Figure 3, the depth estimation process can be divided into intermediate steps that contributes to the outcome of the algorithm. To achieve high quality results without losing in robustness and flexibility, we designed the estimation process to allow full controls over the parameters and fine-tuning for each single step.

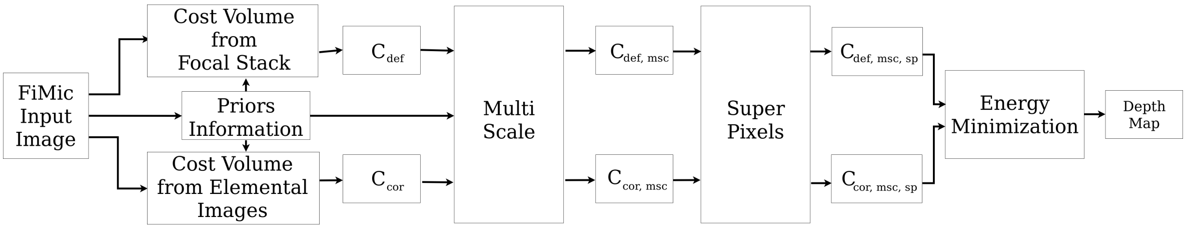


Figure 3. Pipeline of the depth estimation process. The name of cost volume (e.g., C_{def}, C_{cor}) are consistent with the ones used in the paper.

The pipeline consists in several steps: first, two cost volume cubes are calculated using respectively the elemental images, in Section 3.2, and the focal stack, in Section 3.3. These two cubes are then refined using a multi-scale approach similar to [31] in Section 3.4 and a contribution from superpixels inspired from [32] in Section 3.5. In Appendix A an overview of the parameters is given in Table A1.

Moreover, we introduce additional steps to increase the robustness of the algorithm and to adapt to a vast variety of input images. In Section 3.1 priors are incorporated in the fusion of the data under three different forms. A matting mask is applied to address the dark regions of the scenes, appearing mostly in transparent biological samples, where the empty areas do not capture light. Areas with high and low frequencies are weighted differently on high or low resolution depth estimation, and a failure prediction map is used to weight the contributions of the stereo matching along different epipolar lines.

The next step consists in fusing the two cost volumes and extracting a depth map. This is approached in Section 3.6 as an energy minimization problem, building an energy function that combines both cost volumes and finding the minimum using a multi-label optimization framework.

Finally, we used a post-processing filter to improve the estimation and obtain a smoother depth map. We applied a weighted median filter with integer weights followed by a guided filter, using in both cases the central image as a reference image.

3.1. Priors Information

Priors information have been incorporated into the depth estimation pipeline to deal with different kinds of input images, increasing the robustness and the reliability of the proposed approach.

3.1.1. Frequency Mapping

As explored in the literature in [33,34], the best improvements for the multi-scale approach are visible across areas with different frequencies: in high-frequencies areas high-resolution images can obtain the best results, while in low-textured and low-frequency areas a lower resolution leads towards more robust estimation.

Based on this consideration, we built a frequency map image using the difference of gaussians. The algorithm, calculating the difference in the amount of blur between two different blurred versions of the same image, is able to compute a frequency value for each pixel of the reference image.

$$F_{map} = I * (g(\sigma_1) - g(\sigma_2)) \tag{4}$$

where the reference image is denoted as I and $g(\sigma_1)$, $g(\sigma_1)$ are gaussian kernels of variance σ_1 and σ_2. To obtain the desired result, we must ensure $\sigma_1 \neq \sigma 2$. The pixel values will then be use to weight the sum of the depth estimation at different scales, by quantizing the images into F_s levels, where each level corresponds to one scaled version.

3.1.2. Failure Prediction Map

In the stereo matching case, the correspondences search is conducted along the epipolar line: this inevitably leads to error in estimation of structures along that line, as pointed out in [35].

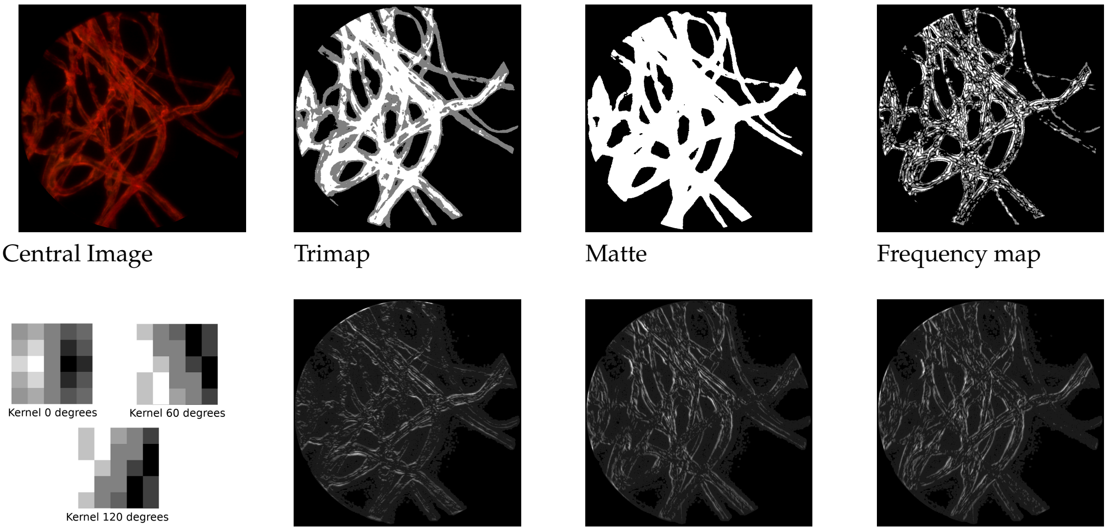
We built three different sobel-like kernel filters to detect edges at respectively 0, 60 and 120 degrees along the three epipolar lines of the hexagonal grid structures. By convolving the image with such a filter we obtain another weight map that favours the contribution of the estimation coming from the most appropriate direction. In Figure 4 it is possible to see the three kernels and an example of failure maps.

3.1.3. Matting

Some of the assumptions for natural images are not valid when dealing with microscope images. Because of the nature of the object and fluorescence illumination, some scenes exhibit a uniformly dark colored background.

In such dark and uniformly colored areas, almost every possible approach for depth estimation is doomed to fail. Based on these considerations, we adapt a matting approach to mask the areas we do not want to analyze. The proposed method has two steps: the first is creating a so-called trimap, that consists in heavily quantizing the image to three levels representing respectively foreground, unknown areas and background. This can be done by applying a multi-level thresholding on the focused image.

After this trimap is built, the next step consists in determining if the unknown areas belong to the foreground or background. We do this by applying one of the top performer methods in this areas, the three-layer graph approach in [36]. It introduces a new measure for discriminating pixels by using non-local constraint in the form of a three layer graph model to supplement local constraint, consisting of color line model, forming a quadratic programming that can be solved to obtain the alpha matte. The results are shown in Figure 4.



Central Image
Trimap
Matte
Frequency map

Sobel-like Kernels
Failure Map for 0 degrees
Failure Map for 60 degrees
Failure Map for 120 degrees

Figure 4. Priors under different form. For coherence, all maps are represented with brighter areas describing higher values. For the failure maps, higher values indicate higher likelihood to fail. In the Sobel-like kernels, bright pixels indicate positive values and dark pixels negative values.

3.2. Cost Volume from Correspondences

A stereo matching is performed to compute a three-dimensional cost volume from the elemental images. The cost function used for this scope is

$$C_{cor}(p) = \alpha_c TAD(p) + (1 - \alpha_c)\xi(p) \tag{5}$$

where we use the notation $C(p)$ to indicate the cost of a pixel p. In this formula the combination of the two terms is controlled by the value $\alpha_c \in [0, 1]$. The first term TAD is the truncated sum of absolute difference and is calculated between two images:

$$TAD(x, y) = \sum_{r=-hws}^{+hws} \sum_{q=-hws}^{+hws} min(\tau, |I_1(x + r, y + q) - I_2(x + d_x + r, y + d_y + q)|)w_{(x,y,r,q)} \tag{6}$$

In this notation we substitute the pixel p with its x and y coordinates within the image. The sum is then computed on a window, indices r, q are used to reach the pixel within the window and hws indicates half of the size of the window. To deal with three different epipolar lines, the disparity can be seen as a vector $\mathbf{d} = (d, \theta)$ where d is the disparity value and θ the direction ($\theta = 0, 60, 120°$) and is divided into its two horizontal and vertical components, respectively $d_x = d\cos(\theta)$ and $d_y = d\sin(\theta)$. τ is the threshold for truncating the difference.

The second term $\xi(p)$ indicates the census transform and is calculated as:

$$\xi(x, y) = \sum_{r=-hws}^{+hws} \sum_{q=-hws}^{+hws} HD(\xi(I_1(x + r, y + q), \xi(I_2(x + d_x + r, y + d_y + q))w_{(x,y,r,q)} \tag{7}$$

where $HD(\cdot)$ indicates the Hamming difference between the census transform of the window around the pixel and the weights are calculated as described in [37]:

$$w_{(x,y,r,q)} = e^{-\left(\frac{|I(x,y) - I(x+r,y+q)|}{\sigma_C} + \frac{dist(x,y,r,q)}{\sigma_d}\right)} \tag{8}$$

where $dist(x, y, r, q) = \sqrt{r^2 + q^2}$ is the euclidean distance between the central pixel and the considered one and σ_c, σ_d are the corresponding parameters regulating the contribution of color and distance.

The cost volume is generated as follows: the correspondences search is completed along the three epipolar lines obtaining $(N_{EI} - 1)$ different cost volumes, where N_{EI} is the number of elemental images ($N_{EI} = 7$ in our case). To merge the cost volumes, the same slice corresponding to a possible disparity value is taken from each cost volume and combined using a weighted average. The failure map calculated in Section 3.1 contains the weights used.

As described in [38], the cost volume is filtered using a guided image filter that takes the colored image as a reference.

3.3. Cost Volume from Defocus

To calculate an accurate defocus map the central image is used as a reference and a difference image for each focal plane is calculated. This allows for a more precise calculation with respect to the defocus response case as defined in [25]: a measure of how much a pixel is in focus at a certain distance.

The difference images are calculated with respect to each focal plane in a similar manner as in the correspondence case:

$$C_{def}(p) = \alpha_d TAD(p) + (1 - \alpha_d)NCC(p) \tag{9}$$

where TAD is defined in Equation (6) and the linear combination of the two terms is controlled by $\alpha_d \in [0, 1]$. The term NCC indicates the normalized cross correlation, calculated as:

$$NCC(x, y) = \sum_{r=-hws}^{+hws} \sum_{q=-hws}^{+hws} \left(\frac{\sigma^2_{I_1 I_2}(x + r, y + q)}{\sigma_{I_1}(x + r, y + q)\sigma_{I_2}(x + r, y + q)} \right) w_{(x,y,r,q)} \tag{10}$$

where the weights $w_{(x,y,r,q)}$ are defined in Equation (8) and σ_{I1I2}, σ_{I1}, σ_{I2} indicates respectively the joint variance, the variance of the first and of the second image. The cost volume is built by stacking the difference images calculated for each focal plane. The cost volume is also filtered using a guided filter as in [38].

3.4. Multi-Scale Approach

Cost volume filtering using a multi-scale approach has shown promising results in refining the cost volume for a higher accuracy of the final depth map [31]. To add consistency and robustness to the proposed cost function, we then adopt a multi-scale approach. Taking inspiration from [31], we build three different layers with a scaling factor $s = 2$, to ensure coherence among images. It has been verified in [31] that building more levels does not significantly improve the final estimation. The cost volume calculated on the smaller scale is then upscaled back and propagated to the initial cost volume.

Differently from [31], however, we do not propagate only the best results of the down-scaled estimation. This idea ensures faster computations, but can lead to larger errors. Instead, we sum the whole cost volume using a weighted average based on the priors information, in this case the frequency map.

$$C_{msc}(p) = w_{s0}(p)C(p) + w_{s1}(p)C_{s1}(p) + w_{s2}(p)C_{s2}(p) \tag{11}$$

As shown in Equation (11), the multi-scale cost (C_{msc}) is a weighted contribution of the cost computed at different scales, being C_{s1} and C_{s2} respectively the costs computed at scale $s1 = 2$ and $s2 = 4$. In Equation (11) $C(p)$ indicates a general cost volume: in our case this is applied to both cost volumes, calculated in Equations (5) and (9). From now on they will be denoted as $C_{def,msc}$ and $C_{cor,msc}$.

The weights $w_{si}, i = 0, 1, 2$ come from the frequency map, that is quantized into a number of levels matching with the number of scales used, and calculated as:

$$w_{si}(p) = \begin{cases} \gamma_1 & \text{if } p \in f_i \\ \gamma_2 & \text{if } p \in f_{i\pm1} \\ \gamma_3 & \text{otherwise} \end{cases} \tag{12}$$

where we ensure $1 \geq \gamma_1 \geq \gamma_2 \geq \gamma_3 \geq 0$, and typical values are $\gamma_1 = 0.6$, $\gamma_2 = 0.3$ and $\gamma_3 = 0.1$. Here we denote with w_{si} the weight relative to the i-th scale, with γ_{si} the parameter controlling the weight of i-th scale contribution and with f_i and $f_{i\pm1}$ the i-th and (i ± 1)-th level of the frequency map.

This allows to shape the cost volume based on the characteristics of the pixels, e.g., pixels that belong to high frequency areas will have a cost based on higher resolution and viceversa, pixel from texture-less regions will have a cost built using the lower resolution.

Tuning the parameters allows us to control the impact of the down-scaled version, and by changing them we can obtain more detail-preserving or smoother depth maps. Note that by setting both $\gamma_1 = \gamma_2 = \gamma_3 = 0.33$ we obtain a standard multi-scale approach that does not make use of the priors information.

3.5. Superpixels

Another technique that reported significant improvement in the cost volume filtering is using superpixels. Superpixels were introduced in [39] and exploited in the depth estimation task [19,32]. The idea behind this consists in grouping pixels with similar characteristics to obtain a more consistent depth estimation. The superpixels are built using two parameters, controlling respectively the approximate size and the similarity between the pixels.

For the depth estimation task there are two main ways of using them: by choosing a small size it can be assumed that the portion of the image corresponding to this superpixel belong to a plane. This allows to compute a single depth value for each superpixel, as done in [19]. This works particularly well for structures that do not exhibit abrupt changes in depth. A different way is shown in [32], where larger size is chosen to allow different depths within a single superpixel. A histogram is built and the best depth estimations are selected and used to filter the cost volume.

Based on these observations we build a flexible approach that takes inspiration from the latter one, but can be restricted to the first. We take the cost volume of the pixels within the superpixels and extract a tentative depth map by extracting the minimum values. From this, a histogram is built, where the highest N_p peaks are selected. Then a penalizing function for the labels not being a peak is built:

$$z(t) = \max\left(0, 1 - \sum_{i=1}^{N_p} \frac{t - ind(N_i)}{\sigma}\right) \tag{13}$$

where t loops through the labels (different focal planes or disparity values), the maximum function is used to avoid negative functions, $ind(N_i)$ indicates the index of the i-th peak and σ controls the strictness of the index, i.e., how much a peak is widened to its neighbours. This way of building $z(t)$ ensures $z(t) \in [0, 1]$, that provides an easy handling of the superpixels contribution. In fact, we can sum this back to our cost volume using a penalizing factor. We can then write:

$$C_{sp}(p) = C(p) + \rho c_{sp}(p) \tag{14}$$

where we call C_{sp} the cost volume updated with the superpixel contribution shown as $c_{sp}(p)$, while ρ is the factor that controls the impact of the contribution. In Equation (14) $C(p)$ indicates a general cost volume: in our case this is applied to both $C_{def,msc}$ and $C_{cor,msc}$ to obtain respectively $C_{def,msc,sp}$ and $C_{cor,msc,sp}$. Typical values are $\rho = 0.2$, $N_p = 3$ and $\sigma = 3$.

Note that by changing the parameters and ensuring a smaller size of the superpixels, a maximum number of peak $N_p = 1$ and a small σ, we could obtain a single depth for superpixels, as in [19]. The ρ parameter controls the contribution to the cost volume, with a large value of ρ leading to strongly shape the cost volume for pixels belonging to the same superpixel, and a small value of ρ reducing the cost for having different values within the same superpixel.

3.6. Depth Map Extraction

Many approaches are applicable to extract the depth map from the cost volume. Local and semi-global approaches as winner-takes-all (WTA), semi-global matching (SGM) and more global matching (MGM)[40] are shown to be outperformed from global approaches, where the depth map refinement is posed as an energy minimization problem. The general solution is obtained through minimization of an energy function that depends on two terms, one data term accounting for the cost volume and a smoothness to ensure consistency between neighbouring pixels.

The energy minimization problem can be addressed using several approaches: the most used approaches consist in using Markov random field, as in [25] where the two cost volumes are refined based on their confidence, or in [41] where different energy functions are analyzed, using graph cuts as in [23] or belief propagation as in [42], or a combination of the above methods [43].

We have chosen to use the graph cuts method to minimize an energy function defined as:

$$E(p) = E_{data}(p) + E_{smooth}(p) \tag{15}$$

Intuitively, the data term should include both cost volumes. We used a strategy that allows to cleverly merge the two cubes. By extracting two tentative depth maps using a winner-takes-all approach, some initial considerations can be made. We mainly encounter two cases: one where both guesses agree on a depth value and one where the cost curves have different shapes. The idea here is to apply different weights in these two situations: in the case where there is a large difference, we assume the pixel is unreliable, being either in a texture-less area or part of a repetitive pattern, thus we choose to stick with the defocus estimation, that is most likely to have a guess similar to the real one. In the second case, we choose to give more weight to the correspondence matching, because it is most likely to be more accurate.

We model this by creating an absolute difference map $M_{ad} = \frac{1}{K}|d_{def,wta} - d_{cor,wta}|$, where $d_{def,wta}$ and $d_{cor,wta}$ are respectively the tentative depth map from the defocus and correspondence cost volumes and K is just a normalization factor, that will be used to weight the two contributions of each slice of the cost volume.

Moreover, we want to exploit the points where the two estimations agree. We thus select the reliable pixel where the difference map has its minimum ($M_{ad} = 0$) and for these pixels we penalize the curve with proportion to the distance from the minimum, creating ground control points that are less likely to change their value during the optimization and will serve to distribute the correct value to unreliable neighbours.

The final sum of these contributions can be expressed as

$$E_{data}(p) = (1 - M_{ad}(p))C_{def,msc,sp}(p) + M_{ad}(p)C_{cor,msc,sp}(p) + \beta P_{gcp}(p) \tag{16}$$

where C_{def} and C_{cor} are respectively the cost volume for defocus and correspondences and P_{gcp} is the penalty function used on the ground control points cost curve to enhance the minimum and β is just a scaling factor. The smoothness term ensures consistency across the four adjacent neighbours and can be expressed as $E_{smooth}(p) = \sum_{q \in N} |l(p) - l(q)|$ where N contains the four neighbours of p and $l(p)$ indicates the pixel's label, i.e., its depth value.

The energy is minimized using a multi-linear optimization as explained in [44–46]. The final depth map is filtered using a weighted median filter with integer weights and a guided filter to remove last outliers while preserving structures.

4. Comparative Performance Analysis

The algorithm has been planned for images acquired using the FiMic light field microscope [7], where only visual comparison is possible. Therefore, to prove the worth of the proposed work, additional comparisons were made. We created a set of synthetically generated images with its corresponding ground truth, where numerical analysis can be conducted.

We compared state-of-the-art work on light field microscopy on their available images and state-of-the art in depth estimation for both real and synthetic images. Methods for light fields are not directly applicable to our images, therefore a fair comparison could not be made.

4.1. State-of-the-Art in Light Field Microscopy

Due to its recent introduction in the microscopy field, there are not many approaches available.

To the best of our knowledge, the best approach in literature belongs to [18], that built a light field microscope using a Lytro first generation camera. The light field in this case consists in 9 × 9 views of 379 × 380 pixels, with lateral views exhibiting strong presence of noise. Moreover, the scale of such images, in the order of millimeters, are quite different from the images acquired with FiMic, that can discriminate up to some micrometers.

Despite the image conditions, the extracted depth maps achieve a high accuracy in the reconstruction of the captured objects and maintains consistency in the structure, as shown in Figure 5.

The first image is the head of a daisy, where both approaches reach a satisfying solution, but ours shows a higher level of detail and robustness, visible in the three-dimensional representation.

The second object is an interesting case, where noise, defocus and very low-light condition increases the challenges. Nevertheless, our algorithm is able to reconstruct the structure of the tip of a pencil, that is symmetrical on the y-axis, where the previous approach could not find a solution. This is particularly visible on the lower part of the image.

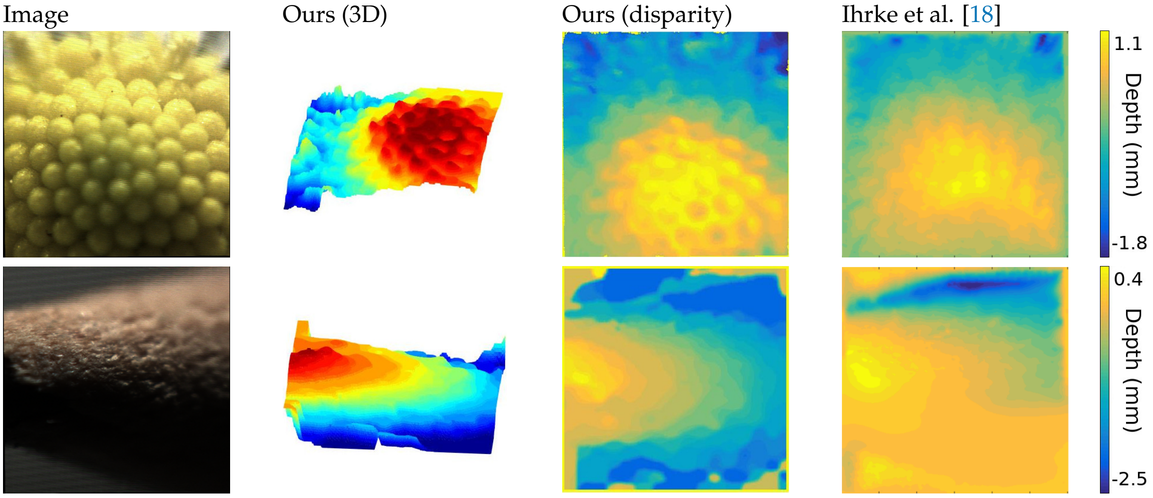


Figure 5. Comparison of depth images from the INRIA dataset. The color coded 3D representation shows the structure of the objects, and the disparity is scaled to try to match the colormap used in [18].

4.2. State-of-the-Art in Depth Map Estimation

To ensure that the proposed approach achieves high quality and accurate results, we also compare against the top performing method for stereo and depth from defocus.

We have chosen to compare with the shape from the focus method, originally published in [47] by taking the best focus measure as analyzed in [48] and a stereo method using neural network to estimate disparity through patches of an image, described in [49]. Unfortunately, we cannot train the network because of lack of dataset, so the pre-trained model is used.

The first set of images, shown in Figure 6, consists of biological samples: the first two images show cotton fibers stained with a fluorescent dye. The hird one shows the head of a zebrafish. The target is excited with a single wavelength laser and is emitting light with longer waveleght (i.e., lower energy). With a bandstop filter the laser's light is filtered out and the light emitted from the sample is captured by the sensor. The wavelength and the color depends on the fluorescent die of the sample.

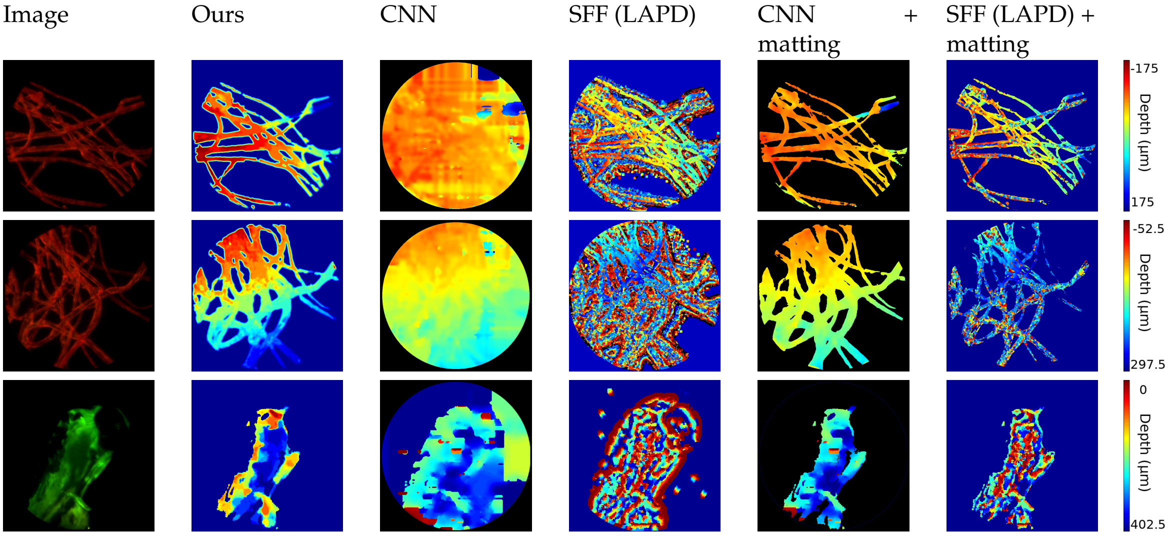


Figure 6. Comparison with Neural Networks (CNN) [49] and Shape from Focus (SFF) [48] on dataset of biological samples: first two rows consist of cotton fibers, last row is the head of a zebrafish.

Therefore, the image exhibits a dark background, where the matting technique described in Section 3.1.3 can be applied. In this case the FiMic was set in order to have: r = 2.2 µm, DOF = 46.4 µm and Z_R = 17.5 µm.

As expected, approaches from depth map estimation in natural images fail to follow the thin structure because of the lack of texture in the dark region and the presence of repetitive patterns in the fibers.

To ensure a fair comparison, we also applied our matting results to the estimated depth map, obtaining more consistent results. This confirms the high quality of our approach and the robustness of our algorithm that can tackle different kinds of input images.

This being a special case, we also evaluate the results on a different dataset. The second set of images, shown in Figure 7, consists of images of small opaque electrical components. By using a luminance ring it is possible to illuminate the object avoiding most of the shades. For this experiment the FiMic was built differently so it led to: r = 14.5 µm, DOF=1812 µm and Z_R = 17.5 µm. Such targets do not need a matting pre-processing step and exhibit structure and texture, being therefore more suitable for standard approaches.

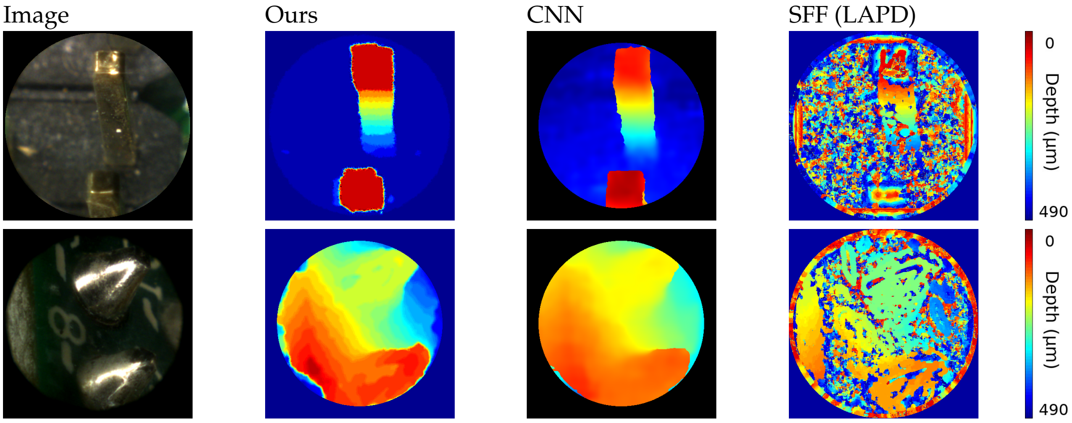


Figure 7. Comparison with CNN [49] and SFF [48] on dataset of opaque electrical components.

Results on these images lead to some considerations: because of the size of the image, the luminance condition and the reflectance of the metal, the depth map calculations are highly challenging, as proven from results of the shape from focus method [48], that recovers only a very noisy reconstruction. Nevertheless, more sophisticated approaches that incorporate filtering steps, as [49], are able to reach satisfying results, obtaining comparable outcomes.

We acknowledge the lack of ground truth images in this field, therefore we propose a numerical evaluation via synthetic images. In recent works Blender has shown to be suitable for emulating light field behaviour and complex scenes [30,50]. We have chosen to use the Blender engine to simulate realistic images of the fibers as if they were captured using the FiMic, thus obtaining the respective ground truth data. To recreate realistic conditions, a special material was generated to emulate the behaviour of semi-transparent fibers and laser illumination. Such images are shown in Figure 8. By doing so, we also obtain the ground truth that allows for a more detailed analysis. Here we show an example of a synthetically generated image and we recap our results in Table 1.

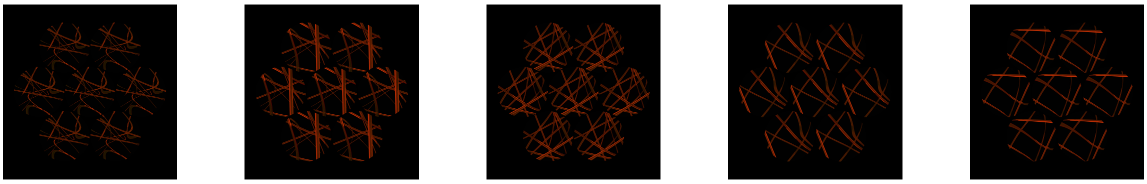


Figure 8. Synthetic images generated with Blender. They simulate the behaviour of the FiMic, shown in Figure 2.

The algorithm was evaluated computing the error as the absolute value of the difference between the estimated disparity and the real one, using the matting mask computed in our approach to reduce the evaluation to the interesting pixels. Results were averaged over a set of five images with different difficulties and shapes to ensure a fair and consistent comparison. The range of the disparities varies among different images, but is consistent within different approaches.

Table 1. Table with results from synthetic images.

Approach	Error	Standard Deviation
Ours	2.32555	1.8154478
CNN [49]	2.4275436	2.4392762
SFF [48]	9.379839	4.1694072

The error is lower in the proposed approach, as well as the standard deviation, confirming the accuracy and the robustness of the method. The algorithm based on neural network [49] obtains similar performances but still exhibits larger errors, as suggested from the higher variance value, while the shape from focus approach [48] fails in achieving high accuracy as expected, being the most naive approach without post-processing. Nevertheless, it shows that the generated images constitute a challenge for off-the-shelf methods for the disparity estimation and therefore enhance the importance of tackling such challenges.

5. Applications

Depth information can also be used for displaying the 3D information of the microscopic objects. With the technique of [51], one of the views and its corresponding depth map can be used to generate an integral image. This image, projected in an integral imaging (InI) monitor, provides a 3D display of the sample.

To generate the integral image, the view and its corresponding depth map are merged into a 3D point cloud. From this point cloud, a set of synthetic views are computationally generated, and finally processed to obtain the final integral image, which will be projected in the InI monitor. An InI monitor is implemented placing a microlens array in front of a pixelated screen: the lenslets of the MLA integrate the rays proceeding from the pixels to reconstruct the 3D scene. This kind of 3D display is autostereoscopic (glasses-free), it allows multiple observers to experience full parallax and overcomes the accommodation-convergence conflict [52].

We implemented the InI monitor through a Samsung SM-T700 and a MLA composed of lenlets having pitch $p = 1.0$ mm and focal length $f = 3.3$ mm (MLA from Fresneltech, model 630). We used as input for the technique of [51] the RGB and depth images of Figure 7. The integral image obtained and its projection in the 3D InI monitor are shown in Figure 9.

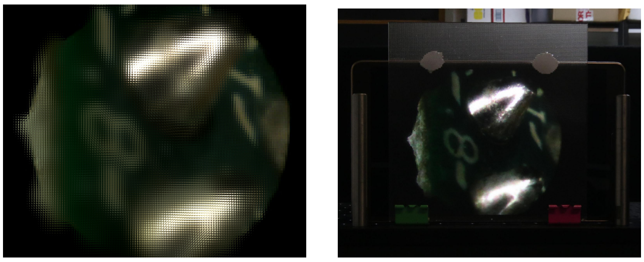


Figure 9. Integral image produced with the method described in [51] and its projection in the 3D monitor.

In this example the image reference plane is set at the background, so the MLA reconstructs the 3D object as a real image, floating in front of the InI monitor. The observer, through his binocular vision system, perceives the depth of the reconstructed objects. A video recording the InI display is visible at the address https://youtu.be/0fPQtckzc_8, in which the parallax and the depth sensation of the reconstructed 3D object is apparent.

In this video, the object shows increasing parallax as we move from far to closer objects. So the left part of the chip and the solder edges move their position with respect to the point of view, while the right part of the chip is almost fixed. This is because during the capture, the chip was tilted with respect to the microscope, as reflected in the depth map of Figure 7.

This technique is very effective for the visualization of the 3D structure of the microscopic samples and proves the usefulness of our work on depth estimation.

6. Summary

The work presented in this paper addresses an important challenge, namely estimating the depth, i.e., the three-dimensional structure of microscopic images. Because of the nature of these images, usually different from natural images, standard approaches may fail. Moreover, due to recent development in the light field technology, light field microscope became a valid alternative for its several applications.

We then use the light field images captured from the recently introduced FiMic [7] that exhibit higher resolution and we showed that the proposed algorithm is capable of accurate reconstruction of challenging scenes, even where previous approaches were failing. These improvements can be helpful for several applications, as presented in the last chapter for the case of lenticular stereoscopic displays, therefore constituting an important contribution for the community.

Appendix A. Implementation Details

Here we provide a table (Table A1) for the parameters relative to the implementation.

Table A1. Parameters Table. The names are consistent with the above formulas which they are referred to. The proposed values are a recommendation and should be adjusted for specific applications. The knowledge of the meaning and the range allow an easier manipulation of such values.

Name	Value	Range	Equation	Meaning
σ_1, σ_2	11, 20	$\sigma_1 > 0, \sigma_2 > 0,$ $\sigma_1 \neq \sigma_2,$	(4)	The variance of the two gaussians kernel used in the difference of gaussians, to build the frequency map.
α_c	0.7	$\alpha_c \in [0, 1]$	(5)	Regulates contributions of TAD and Census in the correspondences cost volume: a higher α leads to higher weights for Census values.
hws	3	$hws > 0,$ $\text{mod}(hws, 2) = 1$	(6), (7), (10)	Half of the size of the windows around the pixel in the cost calculations. Window size can be calculated as $ws = 2 \times hws + 1$.
α_d	0.9	$\alpha_d \in [0, 1]$	(9)	Regulates contributions of TAD and NCC in the defocus cost volume: a higher α leads to higher weights for NCC values.
$\gamma_1, \gamma_2, \gamma_3$	0.6, 0.2, 0.1	$\gamma_1, \gamma_2, \gamma_3 \in [0, 1],$ $\gamma_1 \geq \gamma_2 \geq \gamma_3$	(12)	Used in the multi-scale part. Weights for pixels at different scales with different labels in the frequency map.
σ	3	$\sigma > 0$	(13)	Used in the superpixels section. The width of the peaks in the function penalizing the labels not being a peak.
N_p	3	$N_p > 0$	(13)	Used in the superpixels section. Number of selected peaks. Such peaks will not be penalized. Larger numbers allow more values within the same superpixels.
ρ	0.2	$\rho > 0$	(14)	Controls the weight of the superpixels contribution with respect to the cost volume. $\rho = 0$ cancels superpixels contribution.

The parameters are application dependent, thus the optimal value may not be the same across different executions. We then propose for each parameter a recommended value and the range within which it should vary, followed by a brief explanation of its meaning.

Supplementary Materials: Open-source code is available at: https://github.com/PlenopticToolbox/RobustDepthLFMicroscopy. Images acquired with the Light Field Microscope (FiMic) are also provided.

Author Contributions: Conceptualization, L.P., G.S. and M.M.-C.; instrumentations, G.S.; methodology, L.P.; applications: N.I.; software, L.P.; resources: L.P., G.S. and N.I.; investigation, L.P.; writing–original draft preparation, L.P. (Sections 1, 3, 4, 6), G.S. (Sections 2) and N.I. (Section 5); writing–review and editing, G.Sa., M.M.-C. and R.K.; funding acquisition, G.Sa., M.M.-C and R.K.

Funding: This project has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Skłodowska-Curie Actions Grant Agreement No. 676401. On the other hand, M. Martínez-Corral, G. Saavedra and N. Incardona aknowledge funding from Spanish Ministry of Economy and Competitiveness (Grant DPI 2015-66458-C2-1-R).

Acknowledgments: The authors would like to thank Tim Michels (University of Kiel) for the help in preparing the Blender material used to simulate the fibers in the synthetic images.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Levoy, M.; Ng, R.; Adams, A.; Footer, M.; Horowitz, M. Light field microscopy. *ACM Trans. Graph.* **2006**, *25*, 924–934.
2. Levoy, M.; Zhang, Z.; McDowall, I. Recording and controlling the 4D light field in a microscope using microlens arrays. *J. Microsc.* **2009**, *235*, 144–162.
3. Broxton, M.; Grosenick, L.; Yang, S.; Cohen, N.; Andalman, A.; Deisseroth, K.; Levoy, M. Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express* **2013**, *21*, 25418–25439.
4. Cohen, N.; Yang, S.; Andalman, A.; Broxton, M.; Grosenick, L.; Deisseroth, K.; Horowitz, M.; Levoy, M. Enhancing the performance of the light field microscope using wavefront coding. *Opt. Express* **2014**, *22*, 24817–24839.
5. Hong, J.Y.; Yeom, J.; Kim, J.; Park, S.G.; Jeong, Y.; Lee, B. Analysis of the pickup and display property of integral floating microscopy. *J. Inf. Disp.* **2015**, *16*, 143–153.
6. Llavador, A.; Sola-Pikabea, J.; Saavedra, G.; Javidi, B.; Martínez-Corral, M. Resolution improvements in integral microscopy with Fourier plane recording. *Opt. Express* **2016**, *24*, 20792–20798.
7. Scrofani, G.; Sola-Pikabea, J.; Llavador, A.; Sanchez-Ortiga, E.; Barreiro, J.; Saavedra, G.; Garcia-Sucerquia, J.; Martínez-Corral, M. FIMic: Design for ultimate 3D-integral microscopy of in-vivo biological samples. *Biomed. Opt. Express* **2018**, *9*, 335–346.
8. Prevedel, R.; Yoon, Y.G.; Hoffmann, M.; Pak, N.; Wetzstein, G.; Kato, S.; Schrödel, T.; Raskar, R.; Zimmer, M.; Boyden, E.S.; et al. Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* **2014**, *11*, 727–730.
9. Cong, L.; Wang, Z.; Chai, Y.; Hang, W.; Shang, C.; Yang, W.; Bai, L.; Du, J.; Wang, K.; Wen, Q. Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio). *eLife* **2017**, *6*, e28158.
10. Pégard, N.C.; Liu, H.Y.; Antipa, N.; Gerlock, M.; Adesnik, H.; Waller, L. Compressive light-field microscopy for 3D neural activity recording. *Optica* **2016**, *3*, 517–524.
11. Olympus Stream. Industrial Microscope. Available online: https://www.olympus-ims.com/en/microscope/stream2/#! (accessed on 19 December 2018).
12. Sensofar. 3D Surface Metrology Applications. Available online: https://www.sensofar.com/metrology/technology/ (accessed on 19 December 2018).
13. Lucideon. Three Dimensional Scanning Electron Microscopy (3D SEM). Available online: https://www.lucideon.com/testing-characterization/techniques/three-dimensional-scanning-electron-microscopy-3dsem (accessed on 17 December 2018).
14. Zygo. 3D Optical Surface Profilers . Available online: https://www.zygo.com/?/met/profilers/ (accessed on 17 December 2018).
15. Filmetrics. The World's First $19k 3D Profilometer: The Profilm3D. Available online: https://www.filmetrics.com/profilometers/profilm3d (accessed on 19 December 2018).
16. FLIR—Machine Vision. Rolling Down the Cost of 3D Confocal Microscopy. Available online: https://www.ptgrey.com/case-study/id/10878 (accessed on 18 December 2018).

17. Jung, J.H.; Hong, K.; Park, G.; Chung, I.; Park, J.H.; Lee, B. Reconstruction of three-dimensional occluded object using optical flow and triangular mesh reconstruction in integral imaging. *Opt. Express* **2010**, *18*, 26373–26387. doi:10.1364/OE.18.026373.
18. Mignard-Debise, L.; Ihrke, I. Light-field microscopy with a consumer light-field camera. In Proceedings of the 2015 International Conference on 3D Vision (3DV), Lyon, France, 19–22 October 2015; pp. 335–343.
19. Liu, C.; Narasimhan, S.G.; Dubrawski, A.W. Matting and Depth Recovery of Thin Structures using a Focal Stack. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6970–6978.
20. Wanner, S.; Goldluecke, B. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 606–619.
21. Zhang, Y.; Lv, H.; Liu, Y.; Wang, H.; Wang, X.; Huang, Q.; Xiang, X.; Dai, Q. Light-field depth estimation via epipolar plane image analysis and locally linear embedding. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 739–747.
22. Chen, C.; Lin, H.; Yu, Z.; Bing Kang, S.; Yu, J. Light field stereo matching using bilateral statistics of surface cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1518–1525.
23. Jeon, H.G.; Park, J.; Choe, G.; Park, J.; Bok, Y.; Tai, Y.W.; So Kweon, I. Accurate depth map estimation from a lenslet light field camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1547–1555.
24. Lin, H.; Chen, C.; Kang, S.B.; Yu, J. Depth recovery from light field using focal stack symmetry. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3451–3459.
25. Tao, M.W.; Hadap, S.; Malik, J.; Ramamoorthi, R. Depth from combining defocus and correspondence using light-field cameras. In Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 673–680.
26. Wang, T.C.; Efros, A.A.; Ramamoorthi, R. Occlusion-aware depth estimation using light-field cameras. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chil, 7–13 December 2015; pp. 3487–3495.
27. Huang, C.T. Robust Pseudo Random Fields for Light-Field Stereo Matching. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 11–19.
28. Jeon, H.G.; Park, J.; Choe, G.; Park, J.; Bok, Y.; Tai, Y.W.; Kweon, I.S. Depth from a Light Field Image with Learning-based Matching Costs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 297–310.
29. Palmieri, L.; Koch, R. Optimizing the Lens Selection Process for Multi-focus Plenoptic Cameras and Numerical Evaluation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1763–1774.
30. Palmieri, L.; Koch, R.; Veld, R.O.H. The Plenoptic 2.0 Toolbox: Benchmarking of Depth Estimation Methods for MLA-Based Focused Plenoptic Cameras. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 649–653.
31. Zhang, K.; Fang, Y.; Min, D.; Sun, L.; Yang, S.; Yan, S.; Tian, Q. Cross-scale cost aggregation for stereo matching. *IEEE Trans. Circ. Syst. Video Technol.* **2017**, *27*, 965–976.
32. Furuta, R.; Ikehata, S.; Yamaskai, T.; Aizawa, K. Efficiency-enhanced cost-volume filtering featuring coarse-to-fine strategy. *Multimedia Tools Appl.* **2017**, *77*, 12469–12491.
33. Lee, Z.; Nguyen, T.Q. Multi-resolution disparity processing and fusion for large high-resolution stereo image. *IEEE Trans. Multimedia* **2015**, *17*, 792–803.
34. Choi, E.; Lee, S.; Hong, H. Hierarchical Stereo Matching in Two-Scale Space for Cyber-Physical System. *Sensors* **2017**, *17*, 1680.
35. Meier, L.; Honegger, D.; Vilhjalmsson, V.; Pollefeys, M. Real-time stereo matching failure prediction and resolution using orthogonal stereo setups. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 5638–5643.
36. Li, C.; Wang, P.; Zhu, X.; Pi, H. Three-layer graph framework with the sumD feature for alpha matting. *Comput. Vis. Image Underst.* **2017**, *162*, 34–45.
37. Yoon, K.J.; Kweon, I.S. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 650–656.

38. Hosni, A.; Rhemann, C.; Bleyer, M.; Rother, C.; Gelautz, M. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 504–511.
39. Ren, X.; Malik, J. Learning a classification model for segmentation. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; p. 10.
40. Facciolo, G.; De Franchis, C.; Meinhardt, E. MGM: A Significantly More Global Matching for Stereovision. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; pp. 90.1–90.12.
41. Szeliski, R.; Zabih, R.; Scharstein, D.; Veksler, O.; Kolmogorov, V.; Agarwala, A.; Tappen, M.; Rother, C. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1068–1080.
42. Meltzer, T.; Yanover, C.; Weiss, Y. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 1, pp. 428–435.
43. Tappen, M.F.; Freeman, W.T. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In Proceedings of the Eigth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; p. 900.
44. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239.
45. Kolmogorov, V.; Zabih, R. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 147–159.
46. Boykov, Y.; Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1124–1137.
47. Nayar, S.K.; Nakagawa Y. Shape from Focus; *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 824–831.
48. Pertuz, S.; Puig, D.; Garcia, M.A. Analysis of focus measure operators for shape-from-focus. *Pattern Recognit.* **2013**, *46*, 1415–1432.
49. Zbontar, J.; LeCun, Y. Stereo matching by training a convolutional neural network to compare image patches. *J. Mach. Learn. Res.* **2016**, *17*, 2287–2318.
50. Michels, T.; Petersen, A.; Palmieri, L.; Koch, R. Simulation of plenoptic cameras. In Proceedings of the 2018—3DTV-Conference: The True Vision—Capture, Transmission and Display of 3D Video (3DTV-CON), Helsinki, Finland, 3–5 June 2018; pp. 1–4. doi:10.1109/3DTV.2018.8478432.
51. Incardona, N.; Hong, S.; Martínez-Corral, M.; Saavedra, G. New Method of Microimages Generation for 3D Display. *Sensors* **2018**, *18*, 2805.
52. Son, J.Y.; Javidi, B. Three-dimensional imaging methods based on multiview images. *J. Disp. Technol.* **2005**, *1*, 125–140.