

Single Lens 3D-Camera with Extended Depth-of-Field

Christian Perwaß and Lennart Wietzke

Raytrix GmbH, Schauenburgerstr. 116, 24116 Kiel, Germany

ABSTRACT

Placing a micro lens array in front of an image sensor transforms a normal camera into a single lens 3D camera, which also allows the user to change the focus and the point of view after a picture has been taken. While the concept of such *plenoptic cameras* is known since 1908, only recently the increased computing power of low-cost hardware and the advances in micro lens array production, have made the application of plenoptic cameras feasible. This text presents a detailed analysis of plenoptic cameras as well as introducing a new type of plenoptic camera with an extended depth of field and a maximal effective resolution of up to a quarter of the sensor resolution.

Keywords: Plenoptic Camera, Extended Depth-of-Field, 3D Estimation

1. INTRODUCTION

A camera design that has regained much popularity in the last couple of years is that of the plenoptic camera where a micro lens array is placed in front of the image sensor. Taking pictures with a plenoptic camera offers a whole new way to perceive images. Such images are not static, but allow the user to modify the image after it has been recorded. The user can interactively change the focus, the point of view and the perceived depth of field. As this is done computationally, these properties need not be modified in the same way for the whole image, but can be adjusted separately for each pixel. In addition the raw image of a plenoptic camera also contains all the information needed to calculate the 3D form of the recorded scene. That is, a plenoptic camera is in fact a 3D camera.

The plenoptic camera was, in fact, first described by M. G. Lippmann¹ in 1908 and before that using pinholes instead of lenses by Ives² in 1903. The idea was developed further in the past 100 years by Sokolov (1911), Ives³ (1930), Coffey (1935), Ivanov (1948), Chutjian (1968) with the first digital light field recording device, Dudnikov (1970) and Adelson⁴ (1991) who first called it plenoptic camera. More recently the technology was developed further by Ng⁵ (2005) who built a hand-held plenoptic camera, Levoy⁶ (2006) who applied the light field technology to microscopy, Georgiev⁷ & Lumsdaine⁸ (2006) who found a way to increase the spatial resolution of plenoptic cameras and Fife⁹ (2008) who developed a plenoptic CCD sensor. For a more detailed history see Roberts.¹⁰

The processing of the raw images generated by plenoptic cameras was first developed in a slightly different context by the computer graphics community, where the goal was to generate images from "light-field" data generated by many camera views of the same scene from different angles. Levoy & Hanrahan¹¹ (1996) coined the term "light-field", where the position and the direction of each light ray are stored. Isaksen¹² (2000) later presented a refocusing algorithm from such light-field data. In 2010 the company Raytrix¹³ produced the first commercially available plenoptic camera with an increased effective spatial resolution, an extended depth of field and algorithms for refocusing, view manipulation and scene depth calculation from a single shot.

Some of the reasons why it took over 100 years from the initial idea to the first commercial product are that high quality micro lens arrays, high resolution sensors and high computing power were not available or too costly for a long time to make a cost-effective product possible. This only changed over the last decade as high resolution sensors and high processing power on graphics cards became standard products for a large customer base. Another reason that it took some time for the plenoptic technology to take hold is that the initial plenoptic camera design, as for example implemented by Ng⁵ in 2005, generates refocused images whose number of pixels equals the number of micro lenses. That is, even when using a high resolution sensor with 16 mega pixels, say,

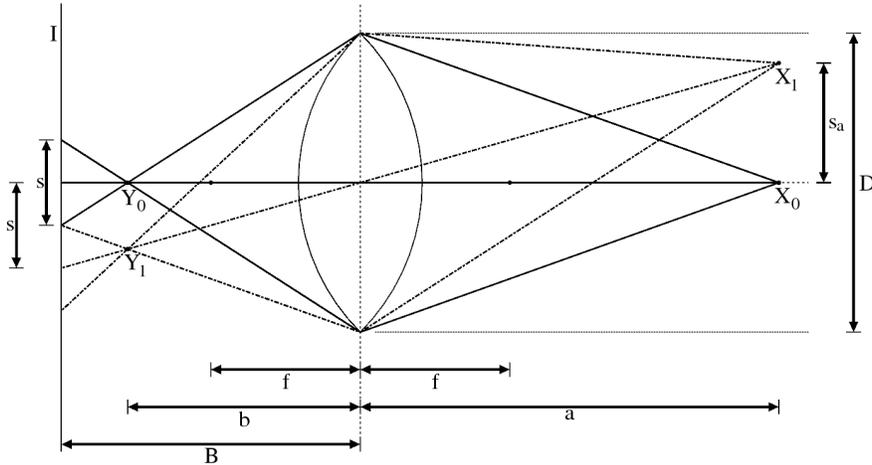


Figure 1. Simplified imaging process of a standard camera.

and a micro lens array with 40000 lenses, the refocused images have a resolution of 40000 pixels. For many applications this is not a practicable resolution. Georgiev & Lumsdaine^{7,8,14} presented a method to increase the effective refocus resolution at the cost of directional resolution by simply placing the micro lens array at a slightly different distance from the image sensor. In this way a 1 mega pixel effective resolution could be achieved from a 16 mega pixel sensor. Another approach to increase the effective resolution of a plenoptic camera is to apply super-resolution techniques,^{15,16} which depend on an exact knowledge of the point spread function of the optic system.

The main contributions of this paper are a detailed analysis of the optical properties of a plenoptic camera and the introduction of a multi-focus plenoptic camera. Other authors have investigated the plenoptic camera as general optical system as well.^{14,16,17} The analysis presented here is more detailed and results in a set of formulas for the design of plenoptic cameras with maximal effective lateral resolution. Furthermore, formulas for the depth of field in relation to the depth of field of a standard camera at the same effective resolution are given. It is also shown that the depth of field of a plenoptic camera can be smaller than that of a standard camera at the same effective resolution if the main lens position is not adjusted. The achievable depth resolution of a plenoptic camera using a triangulation method is also derived. It is also shown that it is important to consider an actual 2D micro lens array to derive the limits on the obtainable effective lateral resolutions.

In addition, a new type of plenoptic camera is introduced here, called the *multi-focus plenoptic camera* (MFPC), whose micro lens array consists of micro lenses with different focal lengths. A MFPC can therefore be regarded as a number of interlaced single focus plenoptic cameras. It is shown that a MFPC extends the depth of field compared to a single focus plenoptic camera and still allows effective lateral resolutions at up to one quarter of the sensor resolution.

The plan for this paper is to first re-derive some basic results for standard cameras, most importantly the effective resolution ratio (ERR) and the depth of field (DOF) in terms of the effective resolution. Then the properties of a plenoptic camera are derived and compared with those of a standard camera. This is followed by the description of algorithms for the synthesization of images and the calculation of the scene depth from a single image taken with a plenoptic camera. Finally we present results of the commercially available Raytrix R11 camera. Note that throughout this paper lens properties are approximated by the thin lens equation.

2. STANDARD CAMERA

Before discussing the plenoptic camera, it is helpful to look at a simplified standard camera setup as shown in figure 1. The main lens generates an image of a point on an object to its right, by focussing the various rays that emanate from that point to a single point on its left. In reality, the rays are not exactly focused to a point, but have an intensity distribution which depends on the whole imaging system. This intensity distribution is

usually called the point spread function (PSF). A fairly good approximation of the spacial extent of the PSF is given by $s_\lambda := 1.22 \lambda N$, where λ is the wavelength of the light and N is the working f -number of the imaging system, which is defined as $N := B/D$. The position b along the optical axis to which an object at distance a from the lens plane is projected, can be calculated by the thin lens equation

$$\frac{1}{f} = \frac{1}{a} + \frac{1}{b}. \quad (1)$$

While the thin lens equation is only an approximation for light rays near the optical axis that enter the lens at a small angle, it gives a good basic understanding of the optical system. For the design of a real system more complex optical properties have to be taken into account.

If, as shown in Figure 1, the point of best focus Y_0 does not coincide with the image plane I , the image of X_0 appears blurred on the image plane, as soon as the diverging projection rays are separated by more than one pixel at the image plane. This means that a standard camera only optimally focusses a particular depth range in object space onto its image plane. This depth range is called the depth-of-field (DOF). One method to increase the range in object space that generates a focussed image, is to decrease the size of the aperture, since then the projection rays diverge more slowly away from the optimal focal plane. The limit is to reduce the aperture to a single point, in which case the whole object space is in focus, but also the amount of light entering the camera is infinitesimal. In standard photography an appropriate compromise between the DOF and the camera speed, i.e. the size of the aperture, has to be found.

A more general concept than determining when something is in or out of focus, is that of an *effective resolution* of the imaging system. As shown in Figure 1, light emanating from point X_0 maps to a range s on the image plane about the optical axis, so that $|s|$ can be regarded as the *effective pixel size* for the object plane at separation a from the lens. Let D_I denote the extent of the image sensor, then the effective resolution R_e and the total resolution R_t are defined as

$$R_e := \frac{D_I}{\max[|s|, s_0]}, \quad R_t := \frac{D_I}{p}, \quad (2)$$

where $s_0 := \max[p, s_\lambda]$ is the minimal size of a projected point that can be resolved with the image sensor and p denotes the side length of a pixel. A more telling quantity is the *effective resolution ratio* (ERR) which is defined as

$$\epsilon_L := \frac{R_e}{R_t} = \frac{p}{\max[|s|, s_0]}. \quad (3)$$

Using the method of similar triangles, the relation between s and a can be found from figure 1 as

$$s = D \left(B \left(\frac{1}{f} - \frac{1}{a} \right) - 1 \right). \quad (4)$$

The ERR for a single lens ϵ_L can therefore be written as a function of the object distance a as

$$\epsilon_L : a \mapsto \frac{p}{\max \left[\left| D \left(B \left(\frac{1}{f} - \frac{1}{a} \right) - 1 \right) \right|, s_0 \right]}. \quad (5)$$

Figure 2 shows an exemplary plot of the ERR, where B , D and f are fixed and the pixel size is assumed to be $D/20$. Point X_0 is projected to point Y_0 , i.e. it gives the position of optimal focus. DoF denotes the depth of field, where the blur is smaller than or equal to the size of a pixel. The left and right borders of the DOF are functions of the ERR denoted by a^- and a^+ , respectively, given by

$$a^- : \epsilon \mapsto B \left[\frac{B}{f} - 1 + \frac{1}{\epsilon} P \right]^{-1}, \quad a^+ : \epsilon \mapsto B \left[\frac{B}{f} - 1 - \frac{1}{\epsilon} P \right]^{-1}, \quad \epsilon \in (0, p/s_0], \quad (6)$$

with $1/a^+ < 1/f - 1/B$ and $1/a^- > 1/f - 1/B$. $P := p/D$ is the pixel to lens aperture ratio. This can also be expressed as the ratio of the working f -numbers of the main lens $N_L := B/D$ and of a pixel $N_p := B/p$, such that $P = N_L/N_p$. Note that in photography the f -number N is defined as $N := f/D$ and the working

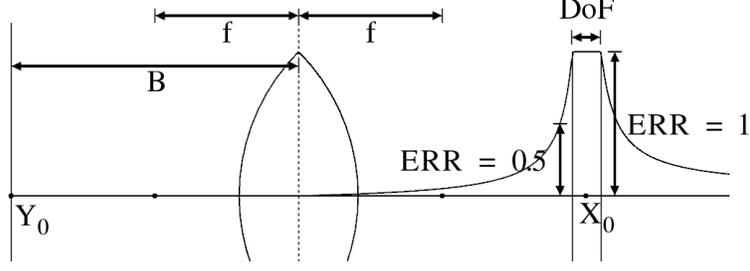


Figure 2. Exemplary plot of the effective resolution ratio (ERR).

f -number as $N_w := (1 + |m|)N$, where m is the magnification of the current focus setting. Let A be given such that $1/A + 1/B = 1/f$, then $m = B/A$ and thus

$$N_w = \left(1 + \frac{B}{A}\right) \frac{f}{D} = B \left(\frac{1}{B} + \frac{1}{A}\right) \frac{f}{D} = \frac{B}{D}. \quad (7)$$

Note that in this text all f -numbers used are *working* f -numbers. The DOF range $\delta(\epsilon)$ is now given by

$$\delta : \epsilon \mapsto \frac{1}{\epsilon} \frac{2BP}{\left(\frac{B}{f} - 1\right)^2 - \frac{1}{\epsilon^2} P^2}, \quad B > f, \epsilon \in (0, p/s_0]. \quad (8)$$

The DOF ratio is defined as $\rho : \epsilon \mapsto \delta(\epsilon)/\delta(1)$ and gives the relation how the DOF scales with a reduced ERR. For a typical camera the main lens aperture D will be much larger than the pixel aperture p , such that $P = (p/D) \ll 1$. Approximating $P^2 \approx 0$ in Equation (8) it follows that $\rho(\epsilon) = \frac{1}{\epsilon}$. That is, for a standard camera, when scaling the image down to half its size, the effective pixel size is increased by factor 2 and thus the ERR is $\epsilon = 1/2$. Hence, the DOF is doubled since $\rho(1/2) = 2$. This shows that simply by scaling the image of a standard camera down, the camera's DOF can be extended at the cost of its lateral resolution.

Next to the object side DOF as shown in Figure 2, it will also be useful to evaluate the image side DOF. This is derived by expressing ϵ_L in terms of b , with $1/b = 1/f - 1/a$. That is,

$$\epsilon_L : b \mapsto \frac{p}{\max\left[\left|D \left(\frac{B}{b} - 1\right)\right|, s_0\right]}. \quad (9)$$

If $B > f$ the borders of the image side DOF as functions of the ERR are given by

$$b^- : \epsilon \mapsto \frac{B}{1 + \frac{1}{\epsilon} P}, \quad b^+ : \epsilon \mapsto \frac{B}{1 - \frac{1}{\epsilon} P}, \quad \epsilon \in (0, p/s_0], \quad (10)$$

with $P := p/D$. The image side DOF γ as function of the ERR is thus:

$$\gamma : \epsilon \mapsto \frac{1}{\epsilon} \frac{2BP}{1 - \frac{1}{\epsilon^2} P^2}. \quad (11)$$

If $p \ll D$, then $\gamma(\epsilon)$ can be approximated by

$$\gamma(\epsilon) \approx \frac{1}{\epsilon} 2pN, \quad (12)$$

where $N = B/D$ is the working f -number of the lens.

3. PLENOPTIC CAMERA

The general plenoptic camera setup that will be discussed in the following is shown in figure 3. The main lens generates an image that lies somewhere to the left of it. Note that the image may also be virtual in the sense that it is formed behind the image plane. The image of the main lens is then projected by the micro lenses onto the image plane.

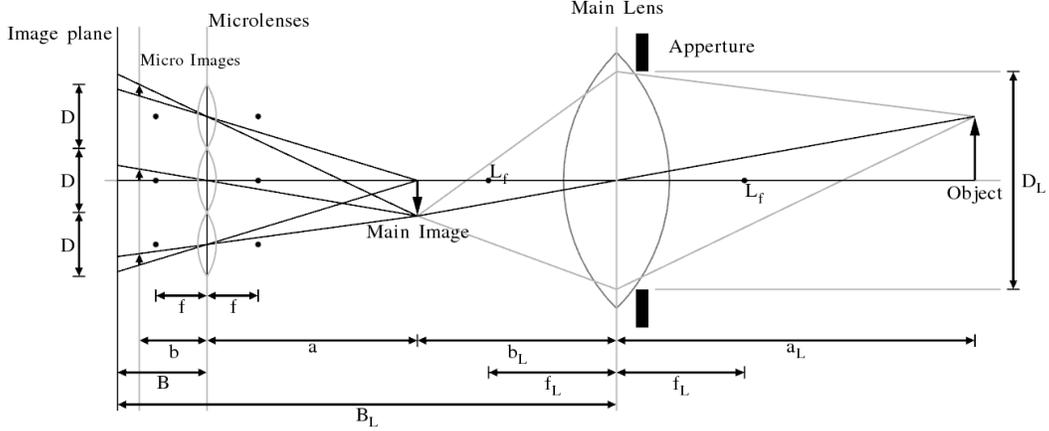


Figure 3. Basic imaging process of plenoptic camera.

3.1 F-Number Matching

The micro images generated by the micro lenses in a plenoptic camera should just touch to make the best use of the image sensor. With respect to Figure 3 this means that

$$\frac{B}{D} = \frac{B_L - B}{D_L} \iff N = N_L - \frac{B}{D_L}, \quad (13)$$

where N is the f -number of the micro lenses and N_L the f -number of the main lens. Since typically $B \ll D_L$, it follows that $N \approx N_L$. That is, the f -numbers of the main imaging system and the micro lens imaging system should match. This also implies that the design of the micro lenses fixes the f -number of the main lens that is used with the plenoptic camera.

3.2 Virtual Depth

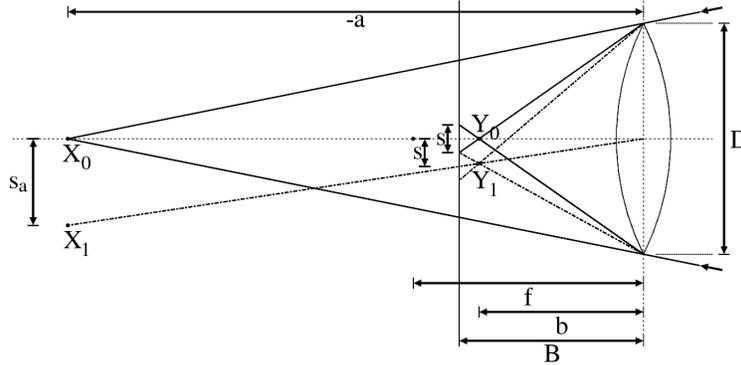


Figure 4. Projection of virtual main lens images by a single micro lens.

Another important concept is that of the *virtual depth*. Figure 1 can also be regarded as the projection of a real main lens image point X_0 onto the camera's image plane by a micro lens centered on the main lens's optical axis. Similarly, Figure 4 shows the projection of a virtual main lens image by a micro lens. The virtual depth of a main lens image point X_0 is defined as $v := a/B$, whereby a is positive if X_0 is a real main lens image as in Figure 1 and is negative if X_0 is a virtual main lens image as in Figure 4. In the following, the main lens will be disregarded and only the projection of its image onto the image plane by the micro lenses is considered. The properties of a plenoptic camera will be developed with respect to the virtual depth. How these properties translate into object space depends on the main lens used.

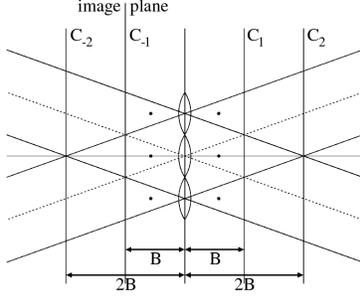


Figure 5. Multiple imaging of a point by micro lens array depending on the point's position.

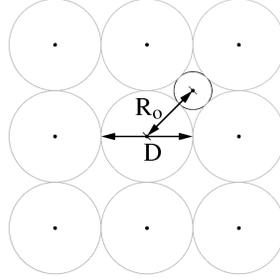


Figure 6. Micro lens projection cones at $|v| = 1$ in orthogonal grid.

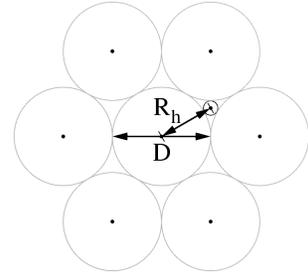


Figure 7. Micro lens projection cones at $|v| = 1$ in hexagonal grid.

3.3 Total Covering Plane

The number of micro lenses that project a point onto the image plane depends on the point's position with respect to the micro lens plane, i.e. its virtual depth. Figure 5 shows the central projection rays of three neighboring micro lenses in a 1D setup. The planes C_1 and C_{-1} are the *total covering planes* (TCP) of this setup as they are the planes closest to the micro lens plane where each point on the plane is projected by at least one micro lens. The virtual depths of the total covering planes are 1 and -1 , respectively. The planes C_2 and C_{-2} are called the *double covering planes* accordingly, as they are the planes closest to the micro lens plane, where each point is projected by at least two micro lenses.

For a 2D micro lens array the situation is somewhat different. Since the micro lenses and the main lens are assumed to be f -number matched, the border of the micro images generated by the micro lenses has the form of the main lens's diaphragm. That is, if the micro lenses are placed in an orthographic grid the main lens diaphragm should have a square form and be oriented along the axes of the MLA to ensure that there are no gaps between neighboring micro images. Similarly, if the micro lenses lie in an hexagonal grid, the diaphragm should be hexagonal. In that case the (TCP) of the camera is still at virtual depth $v = 1$. However, typically lenses have nearly round diaphragms.

Figures 6 and 7 show a top view of the projection cones of the micro lenses at virtual depth $|v| = 1$ for an orthogonal and hexagonal grid, respectively. Moving further away from the micro lens plane, i.e. increasing $|v|$, increases the micro lens projection cone radii R linearly with the virtual depth magnitude as $R = |v| D/2$. To achieve a total covering of the 2D projection plane, the projection cone radii have to be R_o for the orthogonal and R_h for the hexagonal grid, with

$$R_o = \frac{D}{\sqrt{2}}, \quad R_h = \frac{D}{2} \sqrt{1 + \tan^2(\pi/6)}. \quad (14)$$

Defining a *spacing factor* as $\kappa := 2R/D$, the spacing factors κ_o and κ_h for the orthogonal and hexagonal case are

$$\kappa_o = \sqrt{2} \approx 1.41, \quad \kappa_h = \sqrt{1 + \tan^2(\pi/6)} \approx 1.15. \quad (15)$$

The virtual depths of the i -times covering plane for the orthogonal and hexagonal case are therefore $v_i^o = \pm i \kappa_o$ and $v_i^h = \pm i \kappa_h$, respectively. Therefore a complete resultant image can be reconstructed at a lower virtual depth in the hexagonal case than in the orthogonal case. As depth estimation is only possible starting at the double covering plane, a hexagonal micro lens grid also allows for depth estimation at a lower virtual depth. As will be shown later, the effective lateral resolution of a plenoptic camera drops with increasing virtual depth. It is thus advantageous to use a hexagonal micro lens grid as compared to an orthogonal grid.

3.4 Effective Lateral Resolution

Consider again the plenoptic camera in the 1D case. The virtual depth gives in this case also the number of images generated by the micro lenses of a point. For example, a point at virtual depth $|v| = 3$ is imaged by three

micro lenses. This means that a point that is imaged to a single pixel by a standard camera, is imaged to three pixels in a plenoptic camera, in this case. Hence, the effective resolution ratio is 1/3. If the micro lenses were replaced by ideal pin holes with an infinite depth of field, the effective resolution ratio of a plenoptic camera would be $\epsilon = 1/|v|$. This is the maximal effective resolution ratio with respect to the virtual depth that any plenoptic system can achieve.

In a real plenoptic camera with micro lenses, the micro lenses have a finite depth of field. Their effective resolution ratio is thus the same as that for the standard camera given by equation (5). The ERR of a plenoptic camera ϵ_P is therefore the product of the ERR of an ideal plenoptic camera and the ERR of a microlens,

$$\epsilon_P : v \mapsto \frac{1}{|v|} \epsilon_L(v) = \frac{1}{|v|} \frac{p}{\max \left[\left| D \left(\frac{B}{f} - \frac{1}{v} - 1 \right) \right|, s_0 \right]}, \quad |v| \geq 1. \quad (16)$$

The ERR for the 2D case can be approximated by the square of the ERR in the 1D case.

Note that the same result is obtained when calculating the ERR of the plenoptic camera as $\epsilon_P = p/|s_a|$, where s_a is the effective pixel size in the main lens image plane. The relation between s_a and s is shown in Figures 1 and 4. Hence, $s_a = s a/B = s v$ and thus $\epsilon_P(v) = p/(|v| |s|) = \epsilon_L(v)/|v|$.

3.5 Depth of Field

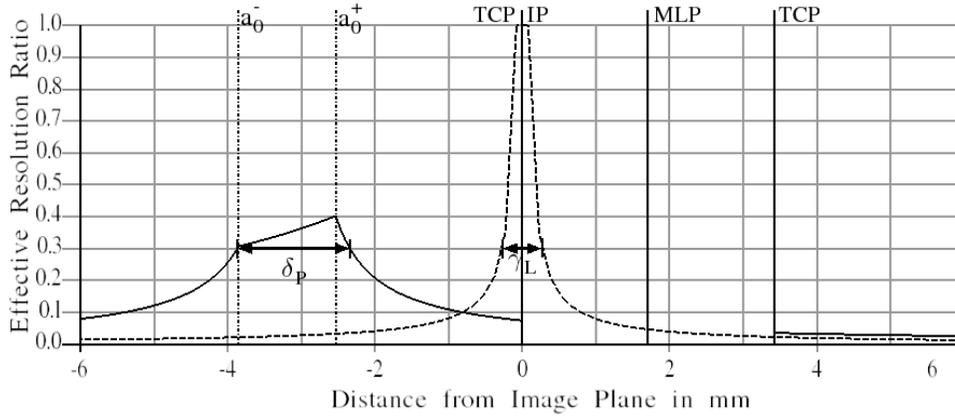


Figure 8. ERR and DOF of main lens and plenoptic camera. IP, MLP and TCP indicate positions of image plane, micro lens plane and total covering plane, respectively. Dotted graph is image side ERR of main lens and γ_L is image side DOF of main lens at $\epsilon_L = 0.3$. Solid line graph is ERR of plenoptic camera and δ_P is DOF of plenoptic camera at $\epsilon_P = 0.3$.

Figure 8 shows the ERR of a standard camera as dotted line and the ERR of a plenoptic camera using the same main lens as the standard camera as solid line. The camera parameters used here are: image sensor pixel size $p = 0.01\text{mm}$, working f -number of main lens $N_L = B_L/D_L = 8$, main lens focal length $f_L = 100\text{mm}$, total distance between image sensor and object the main lens is focused to $T_L = 5000\text{mm}$. The distance between the image sensor plane and the main lens plane B_L and the distance between the main lens plane and the focused object plane A_L are calculated as,

$$B_L = \frac{T_L}{2} \left(1 - \sqrt{1 - 4 \frac{f_L}{T_L}} \right); A_L = T - B_L. \quad (17)$$

The main lens aperture is then given as $D_L = B_L/N_L$. The micro lens diameter D is equal to the micro image diameter, which is set to $D = 21 * p = 0.21\text{mm}$. The distance between the micro image plane and the image plane is fixed by the f -number matching constraint given in equation (13), hence $B \approx 1.7\text{mm}$. The micro lens focal length is set to $f = 1.55 * B$.

As micro lenses project the image of the main lens, the DOF that have to be compared are the image side DOF of the main lens and the object side DOF of the micro lenses. That is, the main lens generates an image at position a relative to the micro lens plane, such that a is negative for main lens images that lie behind the image plane, i.e. to the left of the image plane in Figure 8. The planes denoted by a_0^- and a_0^+ in Figure 8 give the DOF range of the micro lenses. Between these planes the ERR of the plenoptic camera is proportional to $1/|a|$, where $|a|$ is the distance to the micro lens plane. The values of a_0^- and a_0^+ can be derived from equation (16),

$$a_0^- = \left[\frac{1}{f} - \frac{1}{B} \left(1 - \frac{s_0}{D} \right) \right]^{-1}, \quad a_0^+ = \left[\frac{1}{f} - \frac{1}{B} \left(1 + \frac{s_0}{D} \right) \right]^{-1}. \quad (18)$$

The following equations and their definition spaces depend on the sign of a . Recall that if $a < 0$ the main lens image lies behind the image plane and if $a > 0$ it lies in front of the image plane. For brevity, only the case $a < 0$ is considered. The case $a > 0$ can be derived analogously. A concise formula for the DOF ratio of a plenoptic camera compared to a standard camera with same main lens can be derived for the ERR range $\epsilon_P(-B) \leq \epsilon \leq \epsilon_P(a_0^-)$. That is, if the ϵ lies between the ERR at the image plane ($a = -B$) and the ERR at $a = a_0^-$. In this case, the DOF of the plenoptic camera $\delta_P(\epsilon)$ is given by

$$\delta_P : \epsilon \mapsto \frac{1}{\epsilon} \frac{2pN}{|B/f - 1|}, \quad B \neq f, \quad (19)$$

where $N = B/D$ is the working f -number of the micro lenses. Using equation (12), the image side DOF of the main lens can be approximated by $\gamma_L(\epsilon) \approx 2pN_L/\epsilon$, where N_L is the working f -number of the main lens. Making the additional approximation $N_L = N$, the image side DOF ratio of the plenoptic camera to the standard camera is given by

$$\rho_P(\epsilon) = \frac{\delta_P(\epsilon)}{\gamma_L(\epsilon)} \approx \frac{1}{|B/f - 1|}, \quad B \neq f. \quad (20)$$

That is, for $\epsilon_P(-B) \leq \epsilon \leq \epsilon_P(a_0^-)$ the DOF ratio is constant. In the example used for Figure 8, the DOF ratio evaluates to $\rho_P \approx 2.8$.

Note that this is only the image side DOF ratio. The actual DOF in object space depends on the main lens. When projecting the respective DOF ranges of the main lens and the plenoptic camera back into object space, it turns out that the main lens DOF is $\approx 1249\text{mm}$ and the plenoptic camera DOF is $\approx 580\text{mm}$ at $\epsilon = 0.3$. To actually utilize the DOF of the plenoptic camera in object space, the main lens has to be moved closer to the image sensor plane, so that the optimal main lens focus plane lies at the point of maximum ERR of the plenoptic camera DOF range. Placing the optimal main lens focus plane at the plane a_0^+ in the example shown in Figure 8 gives an object side DOF of the plenoptic camera of $\approx 2366\text{mm}$, which is nearly twice the DOF of the standard camera.

3.6 Standard Plenoptic Camera

The two limits of the general plenoptic camera are at $f \rightarrow \infty$ and $f = B$. The case $f \rightarrow \infty$ is the standard camera with an additional flat sheet of glass instead of a micro lens array. The other limit $f = B$ is, what is usually described as the standard plenoptic camera in the literature. One example is the camera described by Ng.⁵ The case $f \neq B$ is called *plenoptic 2.0* or *focused plenoptic* camera.^{14,15} The theory developed for the processing of light-field data can be readily applied to images generated with a standard plenoptic camera. As the micro lenses are focused to infinity, each pixel in a micro image relates to a particular angle of an incoming parallel light bundle.¹⁴ The whole micro image array therefore stores information about the incoming light direction at different positions, that is it stores the whole *light field*. The effective lateral resolution is found to be equal to the number of micro images. However, such an image exists for each direction stored. The image synthesization algorithm for a standard plenoptic camera is therefore very simple, as just one pixel from each micro image at the same relative position to the respective micro image center is taken and combined into the resultant image.

Setting $f = B$ in equation (16) it follows that

$$\epsilon_P(v) = \frac{1}{|v|} \frac{p}{\max[|D/v|, s_0]} = \min \left[\frac{p}{D}, \frac{p}{|v|s_0} \right], \quad |v| \geq 1. \quad (21)$$

Therefore, if $1 \leq |v| \leq D/s_0$ the effective resolution ratio is constant and equal to p/D . Since the effective resolution ratio is defined as $\epsilon = R_e/R_t$, where R_e is the effective resolution and R_t is the total image sensor resolution, the effective resolution of a standard plenoptic camera is

$$R_e = \epsilon R_t = \frac{p}{D} \frac{D_I}{p} = \frac{D_I}{D}, \quad 1 \leq |v| \leq D/s_0, \quad (22)$$

where D_I denotes the size of the image sensor. Therefore, the effective resolution of a standard plenoptic camera is shown to be constant and equal to the number of micro lenses (D_I/D) for the virtual depth ranges $[-D/s_0, -1]$ and $[1, D/s_0]$.

From equation (5) it follows that the micro lenses generate an optimally focused image if $|v| \geq D/s_0$, that is outside the constant range of $\epsilon_P(v)$. As the virtual depth $|v|$ is changed from D/s_0 to 1, the micro images go out of focus but each point is seen in less and less micro images. These two effects just cancel each other to give a constant ERR in the range $|v| \in [1, D/s_0]$. Since in a 1D array the virtual depth magnitude also gives the number of micro images a point at that virtual depth is projected to, a point at $|v| = D/p$ is seen in as many micro images as there are pixel in a micro image.

From the above analysis it follows that it is actually not advantageous to choose the distance between the micro lens array plane and the image plane to be exactly equal to the micro lenses' focal length. The micro images will appear blurred over a large range in this setting, which gives bad results for depth estimation and image synthesization. Instead, the distance B should differ slightly from f . This gives a larger depth range in which the micro images appear focused while approximating the theoretical image generation quite well.

3.7 Multi-Focus Plenoptic Camera

As the optimal effective resolution of a plenoptic camera is proportional to $1/|a|$, main image points close to the TCPs will have the highest effective resolution. For the image space in front of the image plane, objects that are closer are resolved with a higher effective resolution than objects that are far away. For the image space behind the image plane that means that objects far away from the plenoptic camera will be resolved with a higher effective resolution than objects close by. This is advantageous, because objects that are close also appear larger, which compensates the loss in effective resolution. With a particular choice of the main lens focal length, the effective resolution on the object can even be kept constant. Therefore, only the case $a \leq -B$ is considered in the following.

An optimal plenoptic camera should offer a high lateral resolution with a large DOF. Unfortunately, these two demands counteract each other. The highest lateral resolution is obtained if a_0^+ of equation (18) is $a_0^+ = -B$. Expressing the micro lenses' focal length as $f = \phi B$, where ϕ is a scalar factor, then the condition $a_0^+ \leq -B$ is equivalent to $\phi \leq D/s_0$. For $\phi = D/s_0$ it then follows that $\epsilon_P(-B) = p/s_0$, where $p/s_0 = 1$ if $s_\lambda \leq p$ and the DOF ratio between plenoptic and standard camera is approximately $\rho_P = 1/|s_0/D - 1| \approx 1$. That is, for a high effective resolution the DOF is only that of the standard camera. If a larger DOF is desired, then $\phi < D/s_0$ as $\rho_P = 1/|1/\phi - 1|$, which also reduces the maximal effective resolution $\epsilon_P(a_0^+)$.

A good compromise between a high maximal effective lateral resolution and the size of the DOF can be achieved by using not just micro lenses of one focal length but of a number of different focal lengths. Suppose two different types of regular 1D micro lens arrays are interlaced, then the spacing between two micro lenses of the same type is $2D$. Therefore, the virtual depth of the total covering planes is $|v| = 2$ (cf. Figure 5).

A 2D micro lens array consisting of three interlaced hexagonal micro lens arrays with different lens types is shown in Figure 9. It can be seen that the projection cones of the micro lenses of one type cover the whole plane when their radius doubles. Therefore, the TCPs of such an array are also at $|v| = 2$. It is thus possible to realize a plenoptic camera with three different types of micro lens sub arrays, while only reducing the maximum effective resolution ratio to $1/2$ in each dimension, if the micro lenses' focal lengths are chosen appropriately.

Figure 11 shows an example design of a multi-focus plenoptic camera. The TCP lies at $v = -2B$ which reduced the maximal effective resolution ratio to $1/2$. The focal lengths f_1 , f_2 and f_3 of the three different micro lens types are chosen such that the micro lenses' DOF just touch. In this way, the theoretically optimal effective resolution of a plenoptic camera can be obtained over a larger depth range. Note that the main lens has to be

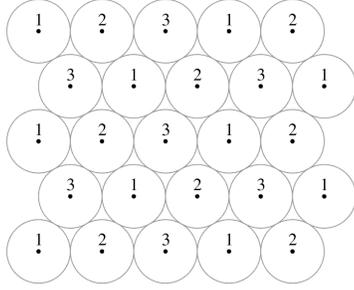


Figure 9. Regular ordering of three different micro lens types in an hexagonal array.

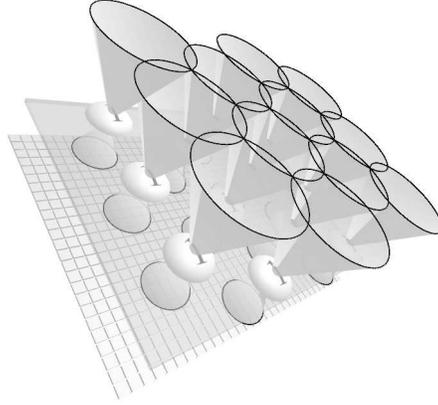


Figure 10. Projection cones of one micro lens type at TCP. The distance between the TCP and the micro lens plane is twice the distance between the image sensor plane and the micro lens plane. Note that the cones are drawn in front of the image plane for better visibility.

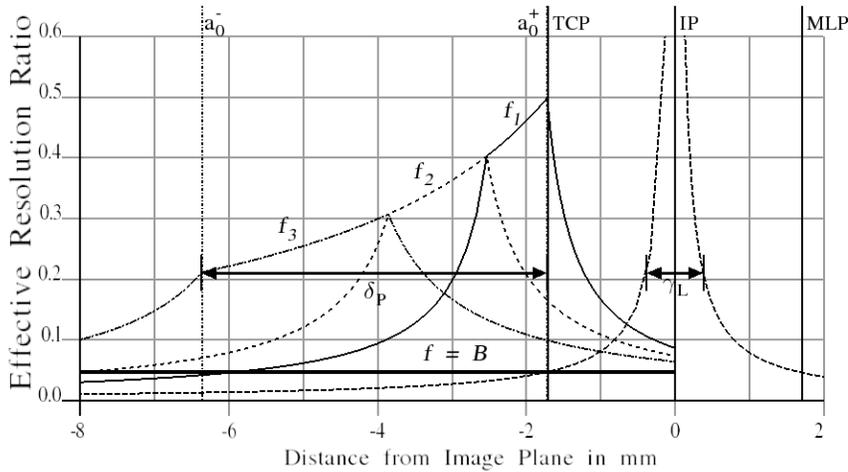


Figure 11. ERR and DOF of main lens and multi-focus plenoptic camera. IP, MLP and TCP indicate positions of image plane, micro lens plane and total covering plane, respectively. The graphs denoted by f_1 , f_2 and f_3 are the ERR graphs for the different micro lens types.

moved closer to the image sensor by the distance between image sensor and TCP to obtain an increased object side DOF. The ERR of a standard plenoptic camera is shown as thick solid horizontal line, denoted by $f = B$.

The object side DOFs of the main lens alone and the multi-focus plenoptic camera depend on the main lens focal length and the focus distance. Using the same camera parameters as before, then for a total focus distance of $T = 5000\text{mm}$ and at $\epsilon = 0.21$, the object side standard camera DOF ranges from 4159mm to 5976mm , which are 1817mm . The multi-focus plenoptic camera DOF ranges from 1584mm to 4898mm , which are 3314mm , 1.8 times more than the standard camera.

Keeping the main lens focal length constant but changing the focus distance to $T = 500\text{mm}$ gives a standard camera DOF range from 359.2mm to 364.4mm , which are 5.2mm . For the same main lens setting the multi-focus plenoptic camera DOF ranges from 333.5mm to 361.8mm , which are 28.3mm , nearly a 6 times improvement over the standard camera. This shows that the multi-focus plenoptic camera works best for long focal lengths and

short focus distances. This makes it particularly well suited for microscopy, for example.



Figure 12. Effect of different micro lens focal lengths on the micro images formed for different object depths. 1, 2 and 3 indicate the different micro lens types.

Figure 12 shows the effect of the different micro lens focal lengths for different object depths. Lens type 1 is in focus for objects near to the main lens focal plane, while lens type 3 is in focus for close up objects.

4. PLENOPTIC IMAGE PROCESSING

4.1 Image Synthesis

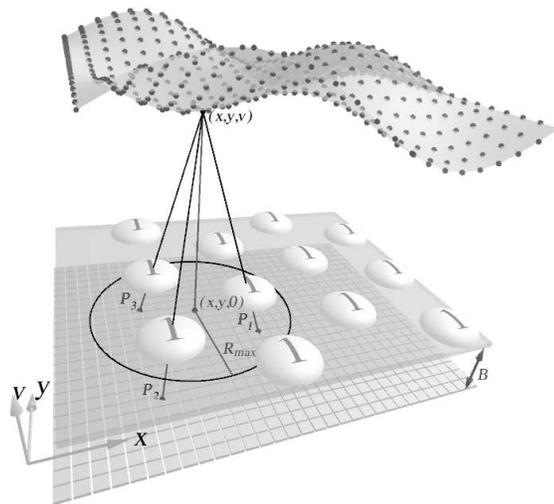


Figure 13. Synthesizing an image from the raw image of a plenoptic camera.

To synthesize a final image from the raw image generated by a plenoptic camera, the correct virtual depth of the pixel in the synthesized image has to be known. How this may be done will be discussed later. For now, it will be assumed that the correct depth is known. Figure 13 shows the geometric setup of the synthesizing algorithm. The wavy surface drawn at the top represents a synthesizing surface where the distance of a point on the surface from the micro lens plane is the virtual depth of that point denoted by v here. Note that the synthesization is shown here for a virtual depth surface in front of the image plane for clarity. The same algorithm also works for negative virtual depths.

To calculate the color or gray value at a position (x, y) on the synthesized image assuming a virtual depth v , the following procedure is applied:

1. Select the subset of micro images of that micro lens type that has the highest ERR at the given virtual depth v . In Figure 13 it is assumed that this is micro lens type 1.

- Find the set of micro lenses which do actually "see" the point at (x, y) . These are all those micro lenses whose centers fall inside a circle of radius R_{max} about the orthographic projection of (x, y) onto the micro lens plane. The radius R_{max} is given by

$$R_{max} = \frac{|v| D}{2 B}.$$

- For each of these micro lenses, calculate the pixel position on the image plane onto which the point (x, y) is projected by the respective micro lens.
- The final color value is a weighted average of the color values on the image plane at the pixel positions calculated in the previous step. The weighting can be taken from calibration data that gives the amount of intensity attenuation at each of the pixel positions for the projection of point (x, y, v) .

Note that if the virtual depth is chosen incorrectly at a point, the synthesized image will show artifacts. A defocussing of the synthesized image can therefore not be achieved by varying the virtual depths in the above algorithm away from their correct values. To obtain a final image with an apparently reduced depth of field, first an image has to be synthesized with the correct virtual depth values and then a blurring is applied to this image depending on the local distance between the correct and the desired virtual depths.

4.2 Depth Estimation

From a single raw image of a plenoptic camera the depth of the recorded scene can be estimated, at least at positions with sufficient local contrast. Why this works can be best understood by regarding the micro lens array as a micro camera array that looks at the virtual image generated by the main lens. As the main lens typically shrinks the object space, the micro lenses "look" at a much smaller version of the original scene. The distance between neighboring micro lenses with respect to this shrunken object space is large enough to offer sufficient parallax for calculating the depth via triangulation. As the distance and orientation of the micro lenses is fixed and assumed to be known, a depth estimation in terms of the virtual depth can be performed without the need of a 3D calibration of the system.

Before a triangulation can be calculated, it has to be known which pixel in neighboring micro images are images of the same object point. This can be done by calculating, for example, the correlation or sum of absolute differences (SAD) of small pixel patches along the epipolar lines of micro image pairs, just as in standard stereo camera approaches. Triangulation is therefore only possible in places with local image contrast, i.e. at edges, corners or other structures. Once the depth for the pixels in the micro images is known, the final depth image can be synthesized in much the same way as described in section 4.1. An alternative iterative approach to depth estimation is, for example, given by Bishop & Favaro.¹⁸

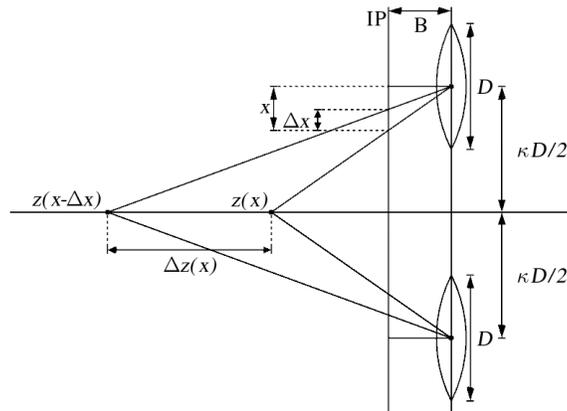


Figure 14. Relation between pixel resolution in micro images and depth resolution. IP denotes the image plane.

For the analysis of the achievable depth resolution by a plenoptic camera, only the depth resolution along a line parallel to the optical axes and lying half way between two micro lenses is considered. In general, the depth resolution will vary depending on the position of this line. Figure 14 shows the geometric construction used. An image point at distance x from the micro image center intersects the central bisecting line at point $z(x)$. If Δx is the size of a pixel, then an image point at the neighboring pixel at distance $x - \Delta x$ from the micro image center intersects the central line at point $z(x - \Delta x)$. The distance between $z(x)$ and $z(x - \Delta x)$ is regarded as the depth resolution for a pixel at distance x from the micro image center. If the match of corresponding image features in neighboring micro lenses can be calculated with sub-pixel accuracy, then Δx can be chosen accordingly.

The relation between z and x is given by

$$\frac{z}{\kappa D/2} = \frac{B}{x} \iff z = \frac{1}{\kappa} \kappa B \frac{D}{2}. \quad (23)$$

The point at distance $z(x)$ from the micro image plane can now be projected back through the main lens into the corresponding object space position $a(x)$ using the thin lens equation. Let z_0 denote the distance of the TCP from the micro lens plane and assume that the optimal focus plane of the main lens for a given image plane to object distance T_L , lies at the plane z_0 . Denote the distance between the main lens plane and the total covering plane at z_0 by b_0 . The depth positions in object space are then given by

$$a(x) = \left[\frac{1}{f_L} - \frac{1}{b_0 + z(x) - z_0} \right]^{-1}. \quad (24)$$

The depth resolution in object space is given as $|a(x) - a(x - \Delta x)|$.

From Figure 14 it becomes clear that just as for the lateral resolution, the image side depth resolution of the micro images is higher for far away objects than for close objects. How this depth resolution maps to object space depends on the focal length f_L of the main lens as well as the focus distance T_L between image plane and object plane. For a given main lens focal length f_L there exists a total focus distance T_L such that the depth resolution is constant over the whole depth range. Let $b_0 = f_L + z_0$, $a_0 = 1/(1/f_L - 1/b_0)$ and $T_L = a_0 + b_0$, then it may be shown that

$$a(x) = x \frac{2 f_L^2}{\kappa D B} + f_L. \quad (25)$$

The derivative of $a(x)$ is therefore constant and given by $\partial_x a(x) = 2 f_L^2 / (\kappa D B)$, such that the depth resolution is

$$\partial_x a(x) \Delta x = \frac{2 f_L^2}{\kappa D B} \Delta x.$$

One special property of a plenoptic camera is that a point is seen in more and more micro images as $|v|$ increases. Therefore, the depth of a scene does not have to be evaluated by only utilizing directly neighboring micro images. Instead, the micro images with the largest separation that still have an image of the same point, can be used. That is, a plenoptic camera has a varying baseline for triangulation over the depth range.

The parameter κ introduced in Figure 14 is a measure of the distance between the micro lens pair centers that are used for triangulation. In fact, z is minimal in Figure 14 if $x = D/2$, i.e. the pixel in question lies at the border of the micro image. In that case, $\kappa = z/B = v$, where v is the virtual depth. In other words, κ gives the minimal virtual depth for which a triangulation can be performed. For a hexagonal micro lens array the first ten consecutive κ values for micro lens pairs are: 2.0, 3.5, 4.0, 5.3, 6.0, 6.9, 7.2, 8.0, 8.7, 9.2, 10.

Using the same main lens settings as before, Figures 15, 17 and 16 show the depth resolution in object space for the ten κ values given above with main lens focus distances $T_L = 500\text{mm}$, $T_L \approx 3131\text{mm}$ and $T_L = 5000\text{mm}$, respectively. Figure 16 shows the constant depth resolution case at $b_0 = f_L + z_0$. Note that full pixel steps were assumed here to calculate the depth resolution. For scenes with a lot of fine structure and good contrast, the disparity of corresponding pixels patches can be calculated with up to a tenth of a pixel accuracy. The depth resolutions shown here then roughly scale down by the same factor.

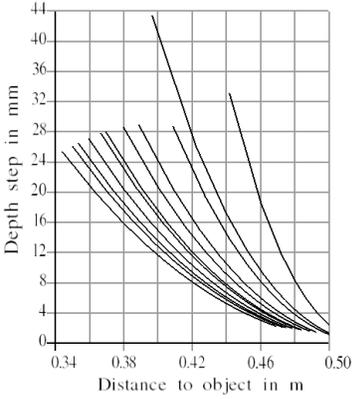


Figure 15. Depth resolution for focus distance $T_L = 500\text{mm}$.

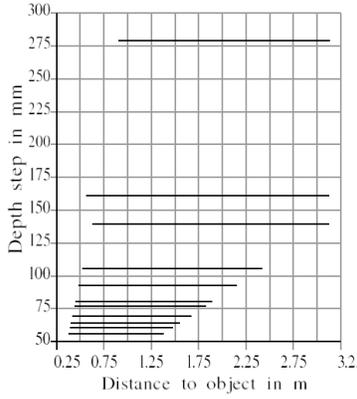


Figure 16. Depth resolution for $b_0 = f_L + z_0$, with $T_L \approx 3131\text{mm}$.

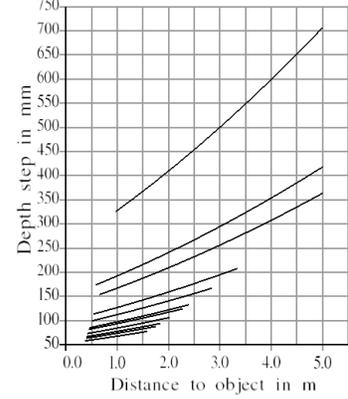


Figure 17. Depth resolution for focus distance $T_L = 5000\text{mm}$.

Note that the larger the DOF of a plenoptic camera, the larger also the range in which scene depth can be calculated, as sharp micro images are needed. With the multi-focus plenoptic camera introduced here, depth can be calculated over a larger depth range than with a standard stereo camera system or a single focus plenoptic camera.

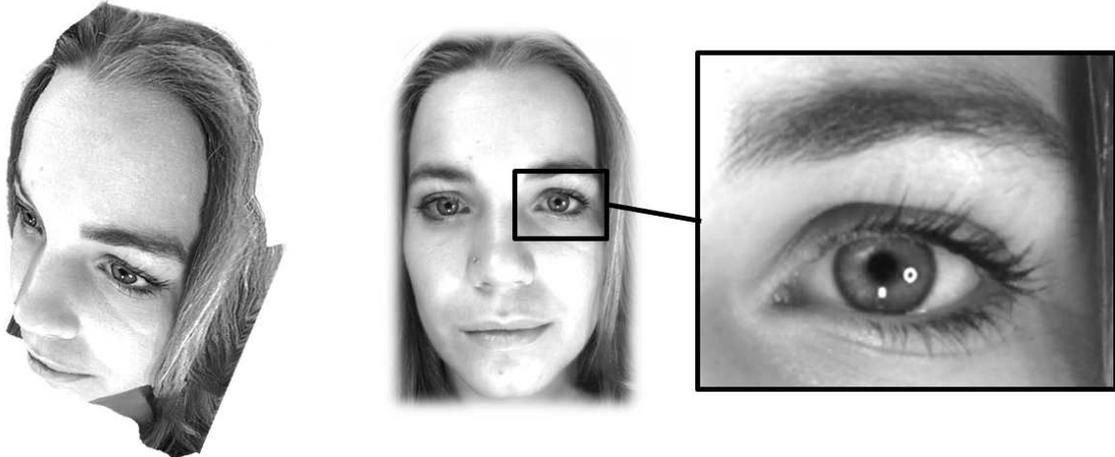


Figure 18. Example results of a portrait taken with a Raytrix R11 camera with a 100mm lens. From a single shot taken a high resolution image can be synthesized as well as calculating the 3D-form of the face.

Figure 18 shows example results of a portrait taken with a Raytrix R11 camera using a 100mm lens at aperture $f/8$. The synthesized image offers a high lateral resolution with up to 2.7 megapixels. In addition the 3D-form of the face can be reconstructed. The depth algorithm used employs a correlation technique to find pixel matches between micro images in the raw image. As depth information is only available at places where there is sufficient contrast, an additional algorithm is used to fill the sparse depth map. The depth algorithm has been implemented to run on an NVIDIA CUDA graphics card. For the present depth data the calculation time was approximately 200ms on a GTX580 graphics card. More example images can be found on www.raytrix.de.

5. CONCLUSION

Depending on the application area of a plenoptic camera, there are different constraints on the f -number, the focus distance and the field of view, while a particular minimal depth of field, lateral resolution and depth resolution are expected. The plenoptic camera analysis presented here is instrumental to design the best plenoptic

camera for the given constraints. The multi-focus plenoptic camera introduced here with its large effective resolution and its extended depth of field, extends the range of applications for plenoptic cameras. For example, automated optical inspection tasks, 3D fluid flow measurements and 3D face capture, are now possible with a single multi-focus plenoptic camera. This makes applications more robust and easier to set up. Raytrix (www.raytrix.de) makes such plenoptic cameras available commercially together with fast synthesization and depth estimation software since 2010.

REFERENCES

- [1] Lippmann, G., “Epreuves reversibles, photographies integrales,” *Academie des sciences*, 446451 (1908).
- [2] Ives, F., “US patent 725,567,” (1903).
- [3] Ives, H. E., “A camera for making parallax panoramagrams,” *Journal of the Optical Society of America* **17**, 435–439 (1928).
- [4] Adelson, E. H. and Wang, J., “Single lens stereo with a plenoptic camera,” *PAMI* **14**(2), 99–106 (1992).
- [5] Ng, R., Levoy, M., Bredif, M., Duval, G., Horowitz, M., and Hanrahan, P., “Light field photography with a hand-held plenoptic camera,” Tech. Rep. CTSR 2005-02, Stanford University (2005).
- [6] Levoy, M., Ng, R., Adams, A., Footer, M., and Horowitz, M., “Light field microscopy,” *ACM Trans. Graph.* **25**(3), 924934 (2006).
- [7] Georgiev, T. and Intwala, C., “Light-field camera design for integral view photography,” tech. rep., Adobe Systems, Inc. (2006).
- [8] Lumsdaine, A. and Georgiev, T., “Full resolution lightfield rendering,” tech. rep., Adobe Systems, Inc. (January 2008).
- [9] Fife, K., Gamal, A. E., and Wong, H.-S. P., “A 3mpixel multi-aperture image sensor with 0.7um pixels in 0.11um cmos,” *IEEE ISSCC Digest of Technical Papers*, 48–49 (2008).
- [10] Roberts, D. E., “History of lenticular and related autostereoscopic methods,” *Parallax* (2003).
- [11] Levoy, M. and Hanrahan, P., “Light field rendering,” in [*SIGGRAPH 96*], 31–42 (1996).
- [12] Isaksen, A., McMillan, L., and Gortler, S. J., “Dynamically reparameterized light fields,” in [*SIGGRAPH 2000*], 297306 (2000).
- [13] Raytrix, “Digital imaging system for synthesizing an image using data recorded with a plenoptic camera,” *European Patent EP09005628.4* (April 2009).
- [14] Lumsdaine, A. and Georgiev, T., “The focused plenoptic camera,” in [*International Conference on Computational Photography*], (April 2009).
- [15] Georgiev, T. and Lumsdaine, A., “Superresolution with plenoptic camera 2.0,” in [*Eurographics*], **28** (April 2009).
- [16] Bishop, T. E., Zanetti, S., and Favaro, P., “Light field superresolution,” in [*IEEE International Conference on Computational Photography*], (2009).
- [17] Georgiev, T. and Lumsdaine, A., “Depth of field in plenoptic cameras,” in [*Eurographics*], (April 2009).
- [18] Bishop, T. and Favaro, P., “Plenoptic depth estimation from multiple aliased views,” in [*3DIM 09*], (2009).