

<https://doi.org/10.31896/k.23.2>

Original scientific paper

Accepted 25. 11. 2019.

GEORG GLAESER

Focus Stacking from a Purely Geometrical Point of View

Focus Stacking from a Purely Geometrical Point of View

ABSTRACT

In order to create extremely sharp photographs, focus stacking has become a widely used method nowadays, mainly in macro or micro photography. Whereas the aspect of computationally detecting sharp regions of an image has been dealt with in many publications, there is only little published about the geometric background. This paper analyzes the process from a purely geometrical point of view, revealing some non-trivial aspects that may potentially also lead to improvements in a variety of applications such as the 3D scanning of small objects. It is shown that – under calibrated conditions and with certain restrictions – focus stacking of a scene leads to arbitrarily many geometrically correct perspective images of this scene, even including normal projections. The reason for this is that the process of photography never leads to purely two-dimensional images, but collinearly distorted spatial images.

Key words: ray optics, focus stacking, Depth of Field, computational photography, image registration, light field camera

MSC2010: 51P05, 65T60, 78A05

Podešavanje dubinske oštine s čisto geometrijskog gledišta

SAŽETAK

Da bi se stvorile izuzetno oštre fotografije u današnje se vrijeme često koristi metoda podešavanja dubinske oštine. Posebno je česta njezina upotreba u makro ili mikro fotografiji. Iako je tema računalnog određivanja oštih područja slike obrađena u mnogim publikacijama, o njegovoj je geometrijskoj pozadini objavljeno malo radova. Ovaj rad analizira spomenuti postupak s čisto geometrijskog gledišta otkrivajući neke netrivialne aspekte koji mogu dovesti do poboljšanja u mnogim njegovim primjenama kao što je 3D skeniranje malih predmeta. Pokazano je da podešavanje dubinske oštine scene – pod kalibriranim uvjetima i uz određena ograničenja – može proizvesti po volji mnogo geometrijski ispravnih perspektivnih slika te scene, pa čak i njezinu ortogonalnu projekciju. Razlog tome je činjenica da proces fotografiranja nikada ne rezultira čistim dvodimenzionalnim slikama, već kolinearno iskrivljenim prostornim slikama.

Ključne riječi: geometrijska optika, podešavanje dubinske oštine, dubinska oština, računski rekonstruirana fotografija, registracija slike, light field kamera

1 Are photographs central projections?

Photography is a source of fascination for many people – especially for mathematicians and geometry enthusiasts. In geometry classes, we sometimes say (in simplified terms): photographs correspond to central projections (perspectives) of space. This means a reduction of three-dimensional space into the two-dimensional plane. This can be “proved” by the fact that it is possible to reconstruct

the photographed three-dimensional scene quite accurately from a number of photographs. It is quintessential that high-quality lens systems are used that reproduce straight edges as exactly straight.

Even though this idea works quite well for scenes with larger technical objects, such as polyhedra (e.g., a furnished room or a building), it poses huge problems in the context of macro photography, meaning photographs of objects that measure just a few centimeters or even less



Figure 1: Two flies, depicted in different ways. Left: Focus-stacked with medium aperture and without flash, right: single photo, closed aperture, twin flash. Whereas the left image is completely sharp, the image on the right illustrates the limits of macro photography, even with professional equipment.

than that. There, the “Depth of Field” (DoF) is comparatively much smaller than in regular photography.

2 An impossible photograph

The flies in Figure 1, measuring about 1 cm, are fundamentally different in terms of the photographic techniques used to depict them. The picture on the left (with some disgorged digestive juice) is in sharp focus overall. An insect photographer will be puzzled: it seems impossible to take such a photograph – even if we used highly expensive equipment with special macro objectives, macro flashes and the highest aperture number possible (as in the picture on the right, where minuscule droplets of water can be seen on the complex eyes.)

This poses two questions: why can an object as small as a fly not be rendered in sharp focus overall, and how does it seem to be possible after all? This paper will explain the situation from a mathematical/geometrical point of view.

3 The lens formula

In physics, the method of operation of a lens (or a well-aligned lens system) is explained as follows: Let P be a point in the real world. It emits (reflects) light rays in all directions. Two of them will have easily predictable properties. The *principal ray* through the lens center C will not be refracted, while the ray parallel to the optical axis will

go through the focal point F^* after the refraction. Behind the lens (the lens system), the two rays – and all others – meet in the pixel P^* .

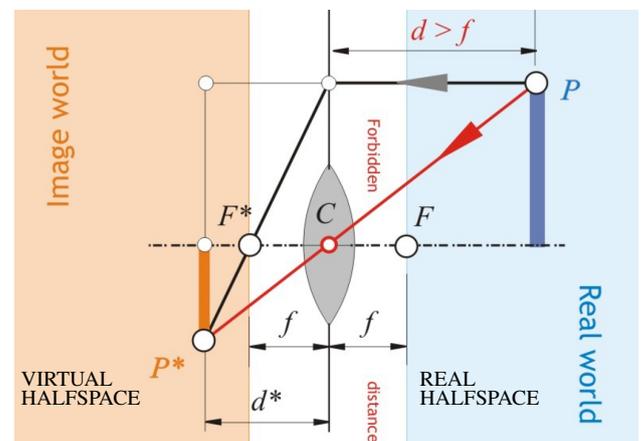


Figure 2: The principle of geometric ray optics (simplified): From the infinitely many light rays emitted from a space point, the principal ray through the lens center C is not refracted, and the ray parallel to the optical axis will go through the focal point F^* after the refraction.

From Figure 2, using similar triangles, we can derive the *lens formula*:

$$\frac{1}{f} = \frac{1}{d} + \frac{1}{d^*}. \quad (1)$$

Here, f denotes the focal distance and d and d^* the oriented distances of the point in space P and the pixel P^* from the symmetry plane through the lens center C , respectively.

Using this method, we can determine the corresponding pixel for each point in space.

4 The Gaussian Collineation

Let us define $d = kf$. Using the lens formula, we then get

$$d^* = \frac{k}{k-1} \cdot f = \frac{d}{k-1}. \quad (2)$$

According to the intercept theorems, we get the following simple relation between the distances of a point in space P and its pixel P^* from the center C :

$$\overline{CP^*} = \frac{1}{k-1} \cdot \overline{CP}. \quad (3)$$

Even though the geometrical mapping $P \mapsto P^*$ works in both directions, a camera will only be able to depict the half space whose points lie at a greater distance from the lens than the focal distance f ($k > 1$): points in the plane through the point F perpendicular to the optical axis are projected onto far points, because the denominator $k - 1$ vanishes in that case.

It is easy to show that the mapping $P \mapsto P^*$ preserves straight lines: Let g be an arbitrary straight line in space. It can always be defined as the intersection of two special planes ε and φ , with ε being the connecting plane of g with the center C and φ being the plane through g that is parallel to the optical axis. ε is transformed into itself ($\varepsilon^* = \varepsilon$) because we can think of the plane as a pencil of principal rays. The other plane, φ^* , can be conceived of as rays parallel to the optical axis that transition into a pencil through the focal point F^* and the intersection line of φ with the symmetry plane. The image g^* of g is the intersection of ε^* and φ^* and therefore a straight line ([1]). The mapping $P \mapsto P^*$ is thus a collineation – which is the technical term for images that preserve straight lines. This mapping is a very special form of *perspective collineation*: the center lies in the *collineation plane* (the symmetry plane). Such a collineation is called *elation*. This insight goes back to C. F. GAUSS ([2]).

¹Using our common sense, we may have assumed that in a photographic image, all those points would be rendered sharp that have a certain constant distance (thus lying on a sphere around C) from the lens center, which depends on the distance d^* from the sensor plane. However, according to the lens formula, this is not the case, as all of those points lie in a plane, the *plane of sharpness*, also called focal plane, at a distance d .

We briefly describe the Gaussian collineation $P(x, y, z) \mapsto P^*(x^*, y^*, z^*)$ analytically. The coordinate system shall have its origin in the camera center C , and the z -axis shall be the optical axis. Then we have

$$\begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix} = \frac{f}{f-z} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (4)$$

When we switch to homogenous coordinates

$$x = \frac{x_1}{x_0}, y = \frac{x_2}{x_0}, z = \frac{x_3}{x_0}, \text{ and } x^* = \frac{x_1^*}{x_0^*}, y^* = \frac{x_2^*}{x_0^*}, z^* = \frac{x_3^*}{x_0^*},$$

we obtain

$$\begin{pmatrix} x_0^* \\ x_1^* \\ x_2^* \\ x_3^* \end{pmatrix} = \begin{pmatrix} f x_0 - x_3 \\ f x_1 \\ f x_2 \\ f x_3 \end{pmatrix}. \quad (5)$$

From this, we can immediately read all important facts of the collineation, namely that it is an elation with center C . The fixpoints lie in $x_3 = 0$, i.e., the lens symmetry plane. The zero plane and the vanishing plane are parallel to the fixpoint plane at distance $\pm f$, respectively.

5 How does this relate to photography?

Using the simple formula (3), we can very easily transform spatial objects made of a number of points into equally *spatial* objects. How does this relate to photography, which after all produces a *two-dimensional* result?

Let us consider a point P at a distance d from the symmetry plane (collineation plane). If the plane π of the sensor of our camera happens to be located at a distance d^* from the collineation plane, the pixel P^* will lie in π . Thus, the intersection of the object that we aim to reproduce with the “plane of sharpness” (also called focal plane), sometimes also referred as through P at a distance d parallel to the collineation plane is in sharp focus.¹ All other points are rendered more or less sharp.

The extent of blurring is dependent on a number of different parameters, as we shall see. One crucial parameter is the size of the distance of the photographed object proportionally to the focal length.

6 Photographing elephants vs. flies

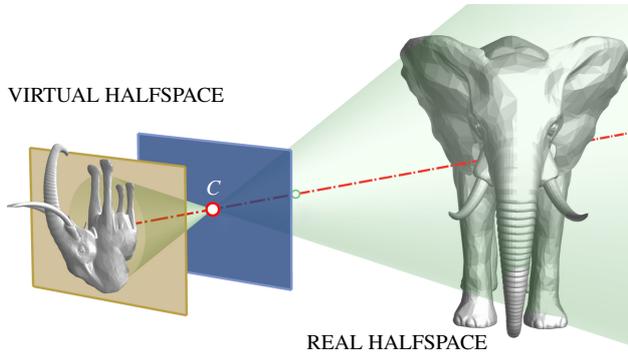


Figure 3: *Depicting large objects from a great distance. The corresponding virtual 3D image is “almost flat”, i.e., all image points are close to the sensor plane. The image in the sensor plane is more or less “sharp”.*

Let us first consider the image of a “large” object from a great distance (what we mean is the relation of the object size to the focal length f (Figure 3)).

For points at a great distance $d = k \cdot f$ ($k \gg 1$), the image distance $d/(k-1)$ (formula 2) does not vary a great deal.

The collinear virtual object behind the lens will thus be strongly oblate, which means that there will be minor blurring of those points that do not lie precisely in the plane of sharpness. Taking a sharp full-size photograph of an elephant thus poses no problem.

The smaller the object that we want to photograph, and the closer it is to the “forbidden” vanishing plane (Figure 4), the more the expression $f/(f-d) = 1/(1-k)$ in the transformation formula (3) will vary. This means that the points will become more blurry, which leads to a genuine problem.

A brief remark: Short focal lengths f apparently have a positive effect on the focus depth, as a fly or a snail becomes larger proportionally to f . Cameras with small sensors have a correspondingly shorter focal length.²

7 Geometry vs. physics

In geometry, the matter seems to be trivial: Let us intersect the light ray through the lens center with the sensor plane. Seen from the perspective of physics, this, of course, does not work: a single light ray is not sufficient to expose the sensor. We will thus have to install a circle-shaped opening

²In technical specifications, the focal length (e.g., 100 mm) of a lens is often provided in terms of a 35 mm equivalent, which means that for a “full format sensor size” of 24 mm × 36 mm, the lens has a focal length of 100 mm.

If the sensor, however, only has a size of, for example, 6 mm × 9 mm, the same visual impression can be achieved with a focal length of just 25 mm. In this special case, we have a *crop factor* of 4. Nowadays, it is possible to take amazingly sharp macro photographs with good smartphones, owing to their extremely short focal lengths and correspondingly tiny sensors (with much larger crop factors.) However, this leads to the problem of having a large number of pixels in the tightest of spaces, which inevitably leads to a loss of quality.

³Artistic photographers often encounter the opposite problem: they deliberately want to work with blurred areas. In such cases, using larger focal lengths and a *wide-open aperture* is recommendable.

in the collineation plane – the aperture. All light rays emanating from a point in space P will then lie inside a skewed circular cone through the opening, which is itself refracted in a skewed circular cone (Figure 5.)

The entirety of light rays in this refracted skewed circular cone exposes the sensor plane in a dot-shaped way only if P lies on the *plane of sharpness*. In all other cases, there is a so-called *circle of confusion* on the sensor (CoC for short.)

We could now assume that we only need sufficient lighting (flash) in order to keep the aperture as small as possible (we speak of a high aperture number in such cases.) However, that is only possible up to a certain limit (the aperture should be bigger than 1 mm at any rate.) If we reduce the size further, the wave properties of light further complicate the matter: this leads to diffraction on the edges of the aperture, which results in inconvenient *diffraction blurring*. Optimal results can be obtained by using the *optimum aperture* indicated by the manufacturer of the lens. Photographers know that exceeding the optimum aperture will reduce the image quality.

8 Focus stacking

In the picture of the elephant, we hardly ran into any problems – if a photographer targets a point that lies approximately at the end of the first third of the desired distance range, the picture will be sufficiently sharp.³

In macro photography, however, a lack of sharpness is a serious problem – especially if we are not dealing with artistic but scientific images.

Nowadays, a technique called *focus stacking* has been established; in essence, it works in the following way: the camera takes a number of images of a scene in as short an interval as possible, varying the distance of the focal plane. This way, we get an image series where different layers of the object are focused consecutively.

The theory of image processing is quite advanced by now, and software used for this purpose is able to distinguish sharp from blurry pixels. Here, we can only briefly mention two methods and do not go into details: One method is to use the shape of the edge gradient profile at each particular edge point to classify edge sharpness (see, e.g., [3]). The other method is to compute the fast Fourier Transform and analyze the result. The Fourier transform tells us which frequencies are present in the image. If there is a

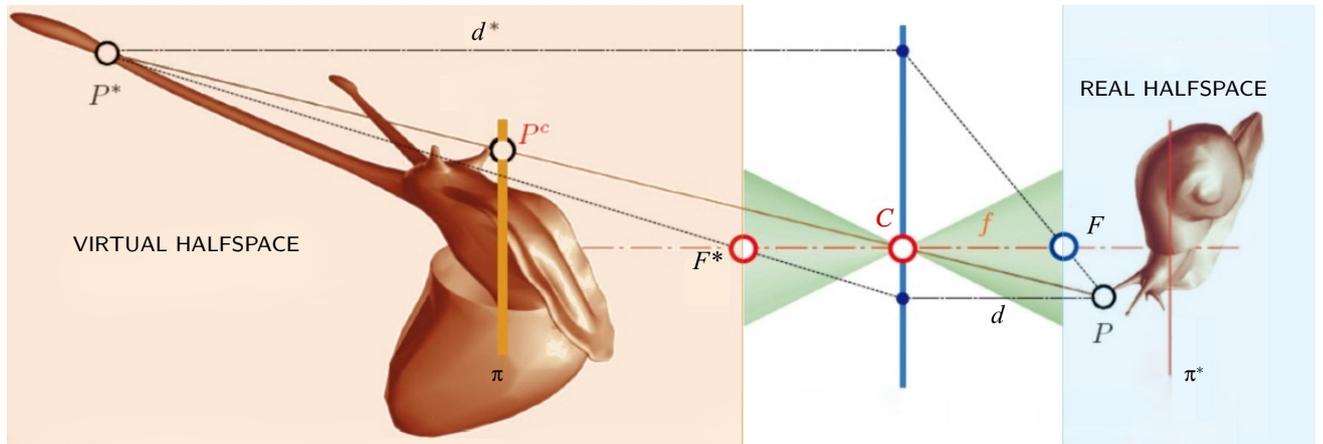


Figure 4: When a small object close to the lens is being photographed, the 3D distortion of the corresponding virtual image is considerable. Only points in the focal plane π^* will therefore be depicted in sharp focus.

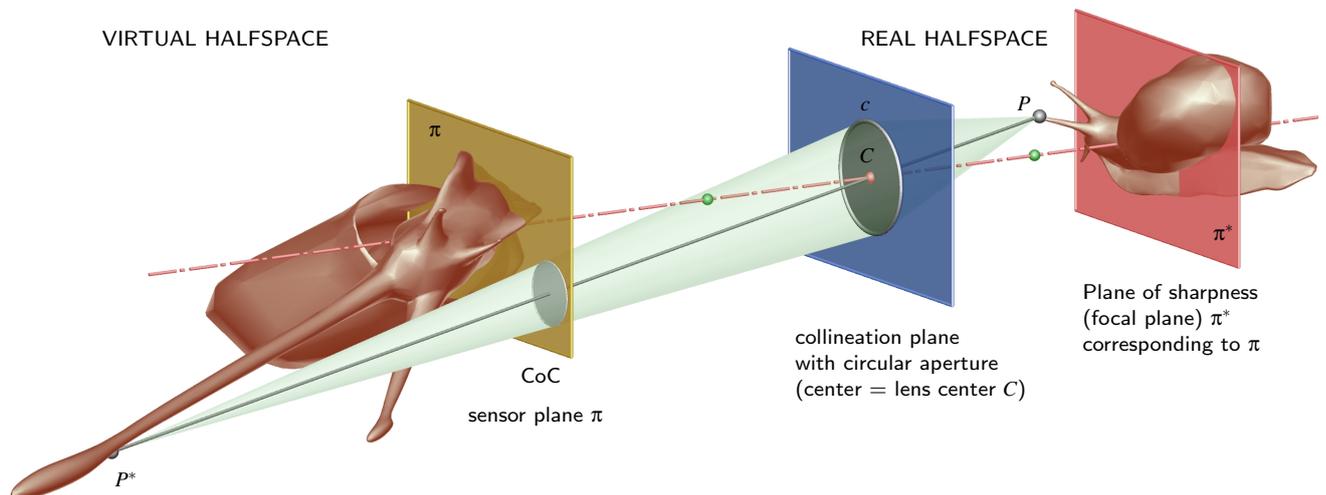


Figure 5: The Circle of Confusion (CoC) in the sensor plane π can be interpreted as the intersection of an oblique cone passing through the circular aperture opening c .



Figure 6: Two mating dragon flies, photographed several times with varying lens centers (and focal points). The first and the last photo of the series can be seen on the left. The focus-stacked image on the right can be considered sharp in the zone that is defined by the two animals. The background still remains blurred, which is – aesthetically speaking – a big advantage. Thus, focus stacking can also be considered as an artistic tool.

small amount of high frequencies, then the image is blurry (e.g., [4]).

In a second step, *one* sharp image is created from this entire image series. How well this works even without a tripod is shown in Figure 6: on the left, we see the first and the last picture of such a series by way of example; on the right, we see the final product.⁴

If the photo series is done without tripod, there is of course a slight movement of the camera to be expected between the single photos. There even might be a tilt and/or motion blur. In [5], solutions for these problems are proposed that are based on affine transformations of the images.

There is an interesting connection to “light field cameras” that capture information about the light field emanating from a scene (conventional cameras only record light intensity).

One type of light field camera uses an array of micro-lenses placed in front of an otherwise conventional image sensor to sense intensity, color, and directional information. Multi-camera arrays are another type of light field camera. Holograms are a type of film-based light field image.

⁴An additional advantage of this method is the fact that we usually do not get an “infinitely” sharp image, but that there is a certain layer rendered in sharp focus. Blurry backgrounds facilitate the isolation of objects from the background and prevent the viewer’s gaze from getting caught in unnecessary details.

9 The focal plane sweeps through the object

For any position of the sensor plane (distance d^*), there is thus exactly one plane of sharpness (distance d) in the Gaussian collineation; its position follows from the lens formula (1):

$$d = fd^*/(d^* - f). \quad (6)$$

9.1 Focus stacking with a microscope

Before we continue with “ordinary photography” (especially macro photography), let us take a quick look at “micro photography”, i.e., taking pictures by means of a microscope (Figure 7). Here, the Depth of Field (DoF) is extremely shallow. When we use focus stacking, we sweep the focal plane in tiny steps. In contrast to classic photography, however, the entire lens system – including the sensor plane – is just translated. Thus, the distance of the lens center to the sensor plane stays constant during the sweeping process.

As a consequence, we do not have to care about relative scaling of the corresponding images. In principle, we get section lines of our object and we even know

the accurate distance of these section lines: the translation of either the lens system or the object carrier respectively. Therefore, the scanned object is completely recognized three-dimensionally (see Figure 8; the microscopic photos show salt crystals and were taken by Johannes Weber, University of Applied Arts Vienna).



Figure 7: When the focal plane of a microscope sweeps through the object, this is equivalent to a perfectly controlled translation of the object along the optical axis. The single pictures do not have to be registered – they are all scaled in the same manner.

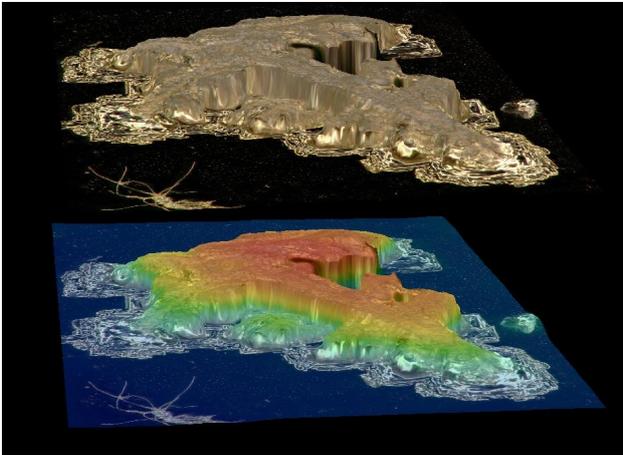


Figure 8: Having control of all coordinates, the object is well-determined and can be displayed and measured arbitrarily.

9.2 Focus stacking with a normal camera

In contrast to focus-stacked pictures taken by a microscope, focus stacking is a bit more complicated in ordinary photography: In commercially available cameras, the position of the sensor plane is fixed and the position of the lens center C moves forward and backward on the optical

axis. If we focus on points that are very far away, C lies in front of the sensor plane at a distance f (if $d^* = f$ then $d = \infty$). If we photograph a two-dimensional shape – such as, for instance, a drawing – that lies in a plane parallel to the sensor plane at a distance of $s = d + d^*$ and focus it, the center C will have the position at a distance d^* from the sensor plane that we want to calculate.

With the formula (6), the following applies:

$$s = d^{*2}/(d^* - f) \quad \text{or} \quad d^{*2} - sd^* + sf = 0 \quad (7)$$

The ambiguous solution of this quadratic equation is

$$d^* = \frac{s}{2} \pm \sqrt{\frac{s^2}{4} - sf}.$$

In order for the expression below the root not to be negative, $s \geq 4f$ must apply. This is, indeed, always the case because of the requirement $d > f$ (if $d = d^* = 2f$, the solutions coincide). Furthermore, both solutions are always valid, even though one would – for practical reasons – tend to stick to *one* algebraic sign when computing a series of camera positions.

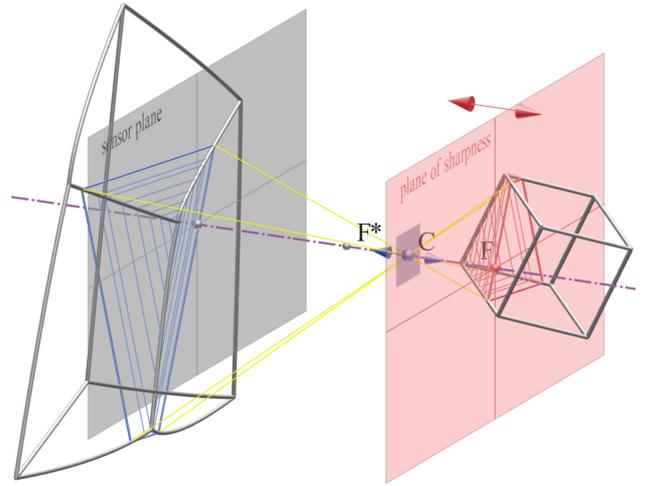


Figure 9: When the center C of the lens system (and therefore also the focal point F) is moving along the optical axis (with fixed sensor plane π), the plane of sharpness which corresponds to the sensor plane is moving in parallel. Its intersection lines with the spatial object are depicted in sharp focus, but in an absolute size that depends on the distance $\overline{C\pi}$.

If the lens system lies within the computed distance d^* in front of the fixed sensor plane π , our two-dimensional shape will firstly be rendered sharp overall on the sensor, and secondly, it will appear *similar*, and thus perspectively undistorted, even though it will not keep its original size.

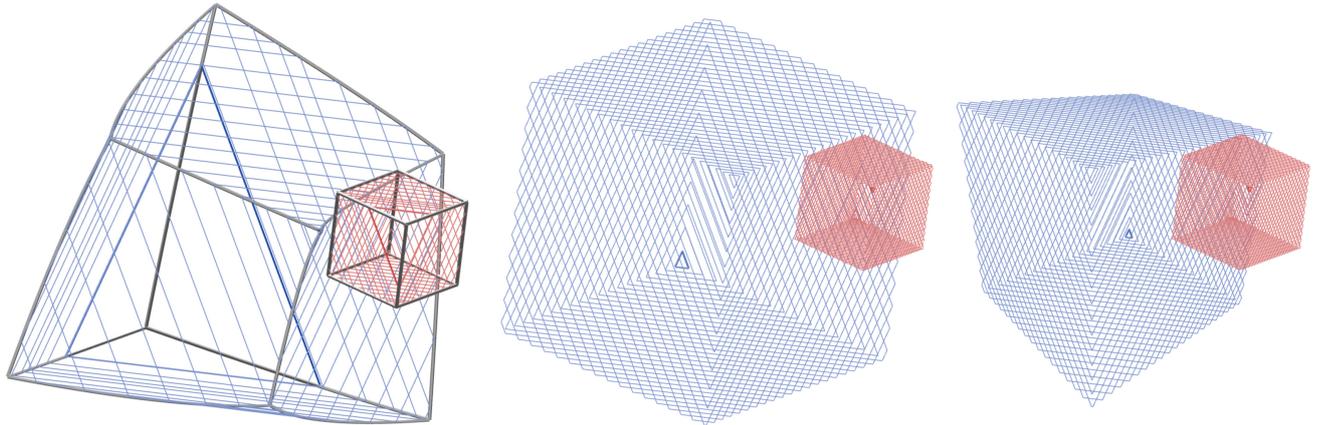


Figure 10: A spatial object (in this case a red cube) is “scanned” by a “sweeping focal plane.” Only the intersections with the plane are depicted in the sensor plane. If the corresponding images of the section are not scaled, the result is neither a central nor a normal projection (image on the left). Scaled with the factor in Equation 8, the image is a normal projection (middle image); scaled with the factor in Equation 9 as in the image to the right, the result is a perfect central projection (= “perspective”). In the latter case, the center of the perspective can vary almost arbitrarily.

Let us now turn back to three-dimensional objects. If we take a series of photographs while purposefully varying the distance of the lens center, we will get photographs in which one section line of the object is rendered sharp *and largely undistorted* (it is merely scaled.)

Figure 9 illustrates this matter: in a simulation of this process, a cube was mapped by way of layers, which can be triangles, quadrangles, pentagons or hexagons. However, if we now simply put these pictures on top of one another, we will end up with strangely distorted perspectives (see also Figure 10 on the left).

Straight lines – such as the edges of the cube – are depicted as curved (it is possible to show that they form parabolas.) The reason for this is of course that, owing to the different distance of the lens center from the sensor plane, the section lines are scaled proportionately to this distance.

10 Converting the scanning process into normal and central projections

Strictly speaking, we scanned our object three-dimensionally in this way – even though we must bear in mind that only those section lines were recorded that were visible from the respective center. Let us now conduct some skilful scaling, in order to compute the relations in the visible three-dimensional real space during the scanning process.

Let t be a line segment in the plane of sharpness, and thus in space, and t^* the line segment parallel to it in the sensor

plane π . According to the intercept theorems, the following applies: $t : d = t^* : d^*$. With the scaling factor

$$\lambda = d/d^*, \quad (8)$$

we can thus calculate the genuine length t from the image length. If we scale all pictures of the series, with the factor changing from one picture to another, we get a *normal projection* of the object (see Figure 10, middle). This is remarkable insofar as it is *impossible* to achieve a normal projection with a single photograph, except if we used an astronomical telescope with an almost infinite focal length.⁵

In order to achieve an exact central projection (perspective), which exclusively relies on unambiguously recorded points, we will choose a position as the center where the lens center used to be when the last still visible points of the object to be depicted appeared sharp ($d = d_m$ maximal $\Rightarrow d^* = d_m^*$ minimal). The i th picture will then not only be scaled according to the formula (8), with the factor $\lambda_i = d_i/d_i^*$, but additionally with the factor d_m^*/d_m : If we now feed the image series into a software designed to recognize sharp pixels, we will get a picture of graphical precision that is sharp and perspective-correct, as can be seen in Figure 10 on the right.

$$\mu = \frac{d_i}{d_i^*} \cdot \frac{d_m^*}{d_m} \quad (9)$$

⁵There is, however, one limitation: even from a number of positions on the optical axis, we cannot always see as much of a surface as we would in a genuine normal projection. Just think, for example, of a spaceship directly approaching the moon. From its position, we will never be able to see 50% of the moon’s surface, the way we do from Earth.



Figure 11: A square stamp of typical macro size is being focus-stacked. With the application of the corresponding scalings of the images, the result is a geometrically perfect perspective.

The square stamp in Figure 11 passes the relatively strict geometrical tests that can be used in order to test the correctness of a perspective, and the same is true of the “geometric still life” in Figure 12. This is important because we do not usually photograph miniature geometrical figures in practice, but living beings and natural objects. Creating a *full-size* sharp photograph of the stamp would be a classic task of macro photography, while the still life is already a medium-sized scene and thus easier to photograph with a sharp focus overall.



Figure 12: A typical geometric scene (cube, square, cylinder of revolution). From the geometric point of view, the stacked image fulfills all requirements of a single central projection.

11 Outlook

Speaking from a mathematical/geometrical point of view, it is important to note that much more can be done with a series of pictures of an object, owing to the fact that these pictures provide information of the object’s location in space – one example would be 3D models. In any case, focus stacking on the macro level should make it possible to achieve results that would rival those of laser scanners, if they were done under laboratory conditions – even though the technology behind it is simpler, quicker, and cheaper. Using the above described geometrical insights, methods like the ones proposed by [6] could be enhanced.



Figure 13: Here, only three photos were stacked, allowing major parts of the praying mantis to appear sharp. The blurred rest of the stacked photo is deliberate.

If the objects that we want to photograph happen to be small animals, we are faced with the additional problem that these seldom tend to freeze in place, and usually at least move their feelers or individual limbs (Figure 13). In such cases, the image series should be processed in the tenth of a second at the most, which will probably be possible in a few years’ time, bearing in mind the rapid technological advancement in recent years. Currently, commercially available cameras still need a full half second for a complete series of 8 to 10 pictures – the bottleneck does not occur when saving the pictures but is a result of the continuous re-adjustment of the focus. In order to depict as many parts of the object’s surface as possible, and also for additional accuracy, one could use several rigidly

connected cameras simultaneously. The registration of the surface parts should easily be possible since all camera positions are well known.

From an aesthetic point of view, we often only need two or three pictures of an insect for impressive photographs. In Figure 13, it was important to focus on the tongs and eyes of the praying mantis – it is secondary that the rest of its body appears blurred.

References

- [1] G. GLAESER, 3D-Images in Photography?, *J. Geom. Graph.* **13** (1) (2009), 113–120.
- [2] P. W. HAWKES, E. KASPER, *Principles of Electron Optics*, Volume 1: Basic Geometrical Optics (2nd edition), Academic Press (2018), 256–257.
- [3] N. NEVEROVA, N. H. KONIK, Edge-based method for sharp region extraction from low depth of field images, *Visual Communications and Image Processing* (2012), 1–6.
- [4] L. RENTING ET AL., Image partial blur detection and classification, *IEEE Conference on Computer Vision and Pattern Recognition*, (2008).
- [5] C. ZHANG ET AL., Extended depth-of-field via focus stacking and graph cuts, *IEEE International Conference on Image Processing* (2013).
- [6] G. KONTOGIANNI ET AL., Enhancing Close-up Image Based 3D Digitisation with Focus Stacking. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* **42** (2017).

Georg Glaeser

e-mail: georg.glaeser@uni-ak.ac.at

Department of Geometry

University of Applied Arts Vienna

Oskar Kokoschka-Platz 2, A-1010 Vienna, Austria