3D point cloud and depth map reconstruction with a monocular liquid lens optical system

Xavier Berthelon
Corning Incorporated
Lyon, France
berthelox@corning.com

Dr. Markus Jäger
SciCaTec GmbH
Holderäckerstraße 8
70499 Stuttgart, Germany
markus.jaeger@scicatec.com

Abstract

Liquid lenses are variable focus components actuated by specific external control with high repetitive accuracy. If the focusing distance is varied over a range large enough, each object is captured once in focus by the image sensor. Taking advantage of these circumstances and the known optical properties of the liquid lens, an all-in-focus image of the environment can be created, also called as image with Extended Depth of Field (EDoF). Furthermore, it is possible to create a 3D (more precisely a 2.5D) representation of the entire view by using only one camera resp. image sensor.

In this white paper we present a technological approach and implementation that demonstrates the described principle by using Corning® Varioptic® Lenses. The required image and video processing, which includes the algorithmic core, is realized by a GPU implementation within an Embedded System, the NVIDIA® Jetson Nano™. In addition, performance data like image processing rate, spatial resolution, and 3D representation acquisition time are given.

Keywords (order of all keywords can be changed)

Liquid Lenses, Extended Depth of Field, EDoF, All-in-Focus, 3D Reconstruction, 3D Model, Image Processing, GPU, SoC

Introduction to liquid lenses

Inspired by the functionality of a human eye, liquid lenses offer manufacturers and OEMs improved speed and reliability over mechanical solutions. The human eye can adjust its focus to the environment at incredibly fast speeds; similarly, Corning's liquid lenses emulate the eyes' fluid and adaptable characteristics to create a rapid response to variable circumstances. This process is made possible by a technology called electrowetting, which uses an electrical signal to manipulate a liquid solution into a workable lens.

Traditional mechanical solutions are limited in their ability to deliver sharp images continuously and reliably. Corning® Varioptic® Lenses (Figure 1) offer innovative solutions to complex optical challenges. Varioptic Lenses enable fast focus and micro-focus without moving parts. Traditional camera systems require moving parts which could begin to wear down and fail over the lifetime of the device. In comparison, liquid lenses function without the use of mechanically moving parts, eliminating much of the maintenance typically associated with vision systems.

Figure 1 : Liquid lens manufactured by Corning Varioptic

Methods for 3D Imaging

Over the past decade, there has been a growing interest for the capture, processing, and imaging of 3D information. The applications for this imaging modality are wide, ranging from information technology to life sciences or even more recently entertainment. 3D information refers to the additional depth cues that are added to a standard 2D image, captured by an imaging system. Many techniques have been developed to estimate depth using either dedicated software and/or hardware. The most commons are the following:

- Stereovision: The principle is to capture a 3D scene under different viewing angles to estimate the depth of objects using either multiple images acquisition or binocular systems.

- Time of Flight or Phase Shift: A dedicated sensor module sends and receives a laser impulse or a modulated signal, the analysis of which determines the depth of objects depending on the received reflection of the light signal.

- Structured lighting: A structured pattern is projected and deformed by objects in the scene depending on their depths, which allows a 3D reconstruction with a monocular system.

- Depth from focus/defocus: Image processing and contrast analysis gives information on the depth of objects using a variable focus system.

This last technique of depth from focus/defocus benefits from a simple hardware with a monocular system (one camera only), no specific lightning or external sensor and relies on the vari-focus properties of the imaging optical system. Liquid lenses have here a key role to play since their working principle results in an optical power variation depending on the voltage applied to their electrodes. The integration of liquid lenses into imaging devices, such as camera modules, together with some image processing (contrast analysis) create an optical system capable of performing 3D reconstruction using the depth from focus modality.

Theory behind the depth from focus /defocus

When the voltage applied to the liquid lens changes, its optical power is modified. During an optical power sweep, objects as seen by the imaging system appear successively out of focus, then in focus and out of focus again. The area where an object appears sharp is called the depth of field. By matching the optical power of the system to the sharpness of an object, we can determine the depth of this sharp object thanks to the laws of geometrical optics that link the focal length to the object distance (Figure 2):

$$m \times m' = f'^2$$

With **m** the object's distance to the object focal point, **m'** the image distance to the focal image point and **f'** the focal length of the optical system.

The depth of field, and therefore the accuracy to determine the object's depth, depends on the optical properties of the system. Any real optical system has a finite aperture size limited by the dimensions of its optical components. The spatial resolution is also limited by the pixel size of the sensor below which it is not possible to distinguish focus. As a result, a range of several points will appear in focus as long as the image spot size remains smaller than the pixel size. This range defines the depth of field as shown on Figure 2.

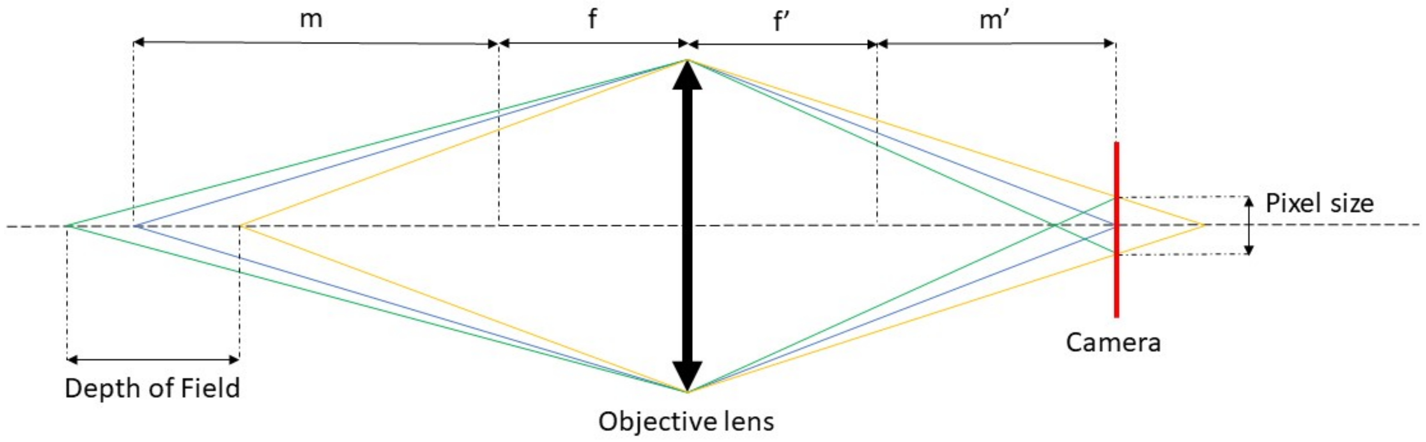

Figure 2 : Illustration of the depth of field of an optical system based on its optical parameters

Determining Sharpness and Depth Information

When sensor image information is available in a record of a pixel array with its width, height, and pixelwise color then there are various possibilities to extract sharpness information for each pixel considering a certain neighborhood. A common method to achieve is applying the well-known Laplace filter after pixel color values were converted to gray scale. Figure 3 shows the absolute value behavior of the Laplace filter result, for one specific pixel in the camera image during a lens voltage sweep from 47V to 27.3V. This result can be taken as a measure of sharpness. We can estimate that at 38V the sharpness reaches a maximum. The sharpness will develop a distribution around the maximum when the chosen voltage step is small enough. The width of this distribution can vary depending on the optical parameters of the system and on the algorithm used to analyze sharpness. Each combination leads to its specific sharpness detection efficiency. Typical indicators, which can behave in the same way to the detection efficiency, are the ratio of the peak sharpness amplitude and the blur region sharpness

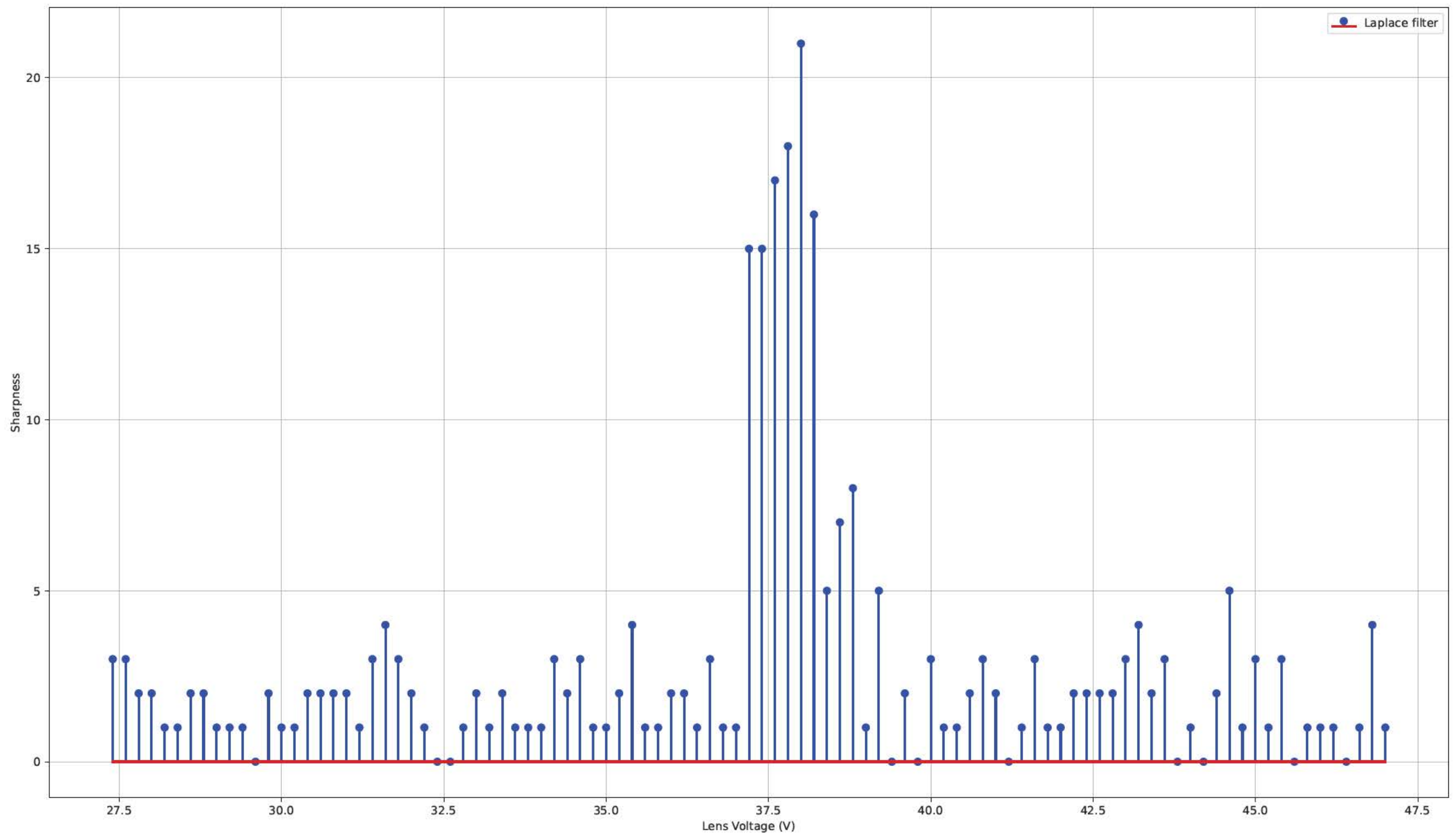

Figure 3 : Absolute value of Laplace filter result for a specific pixel during a lens voltage sweep from 47V to 27.3V

When the lens voltage matching the best sharpness is determined for every pixel in the image, then the corresponding pixel colors of the original camera image can be merged to the EDoF image. Additionally, each pixel can be translated into a point in 3D space taking into account the following circumstances:

1. The correct depth which appears sharp (sharpness plane) on the image sensor can vary over azimuth and elevation around the optical axis. In such cases the sharpness plane is not a sphere.

2. Depending on the chosen lens type and lens voltage this sharpness plane can be flat, concave, or convex distorted along the optical axis.

3. The relation mentioned in point 1 additionally does not need to be axisymmetric to the optical axis.

4. Image sensor and optical axis could not be center aligned which causes an offset of the relation mentioned in point 1.

5. The x and y coordinates of each point in 3D space are depending also on the determined depth (z coordinate) information.

Figure 4 illustrates the principle described above. The blue line shows the image sensor offset which is a transversal displacement to the optical axis. The sharpness plane (3) is for the constantly applied lens voltage concave distorted in z direction.

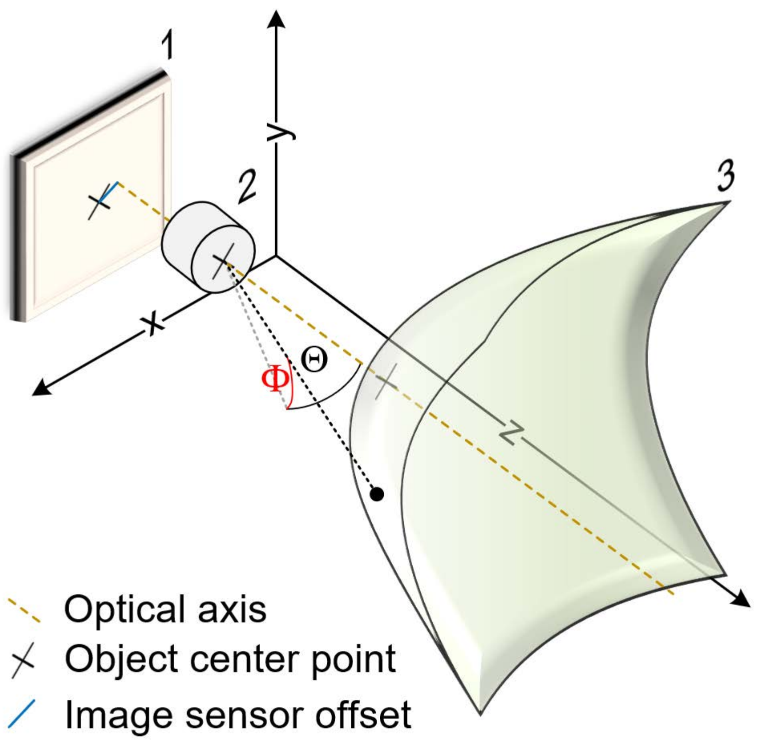

Figure 4 : Sharpness Plane (3) in relation to image sensor (1) and liquid lens (2) for a specific lens voltage

GPU Implementation

We developed an algorithm which sweeps the lens voltage over a given range, acquires all information and generates a new EDoF image and a 3D point cloud when the sweep is completed.

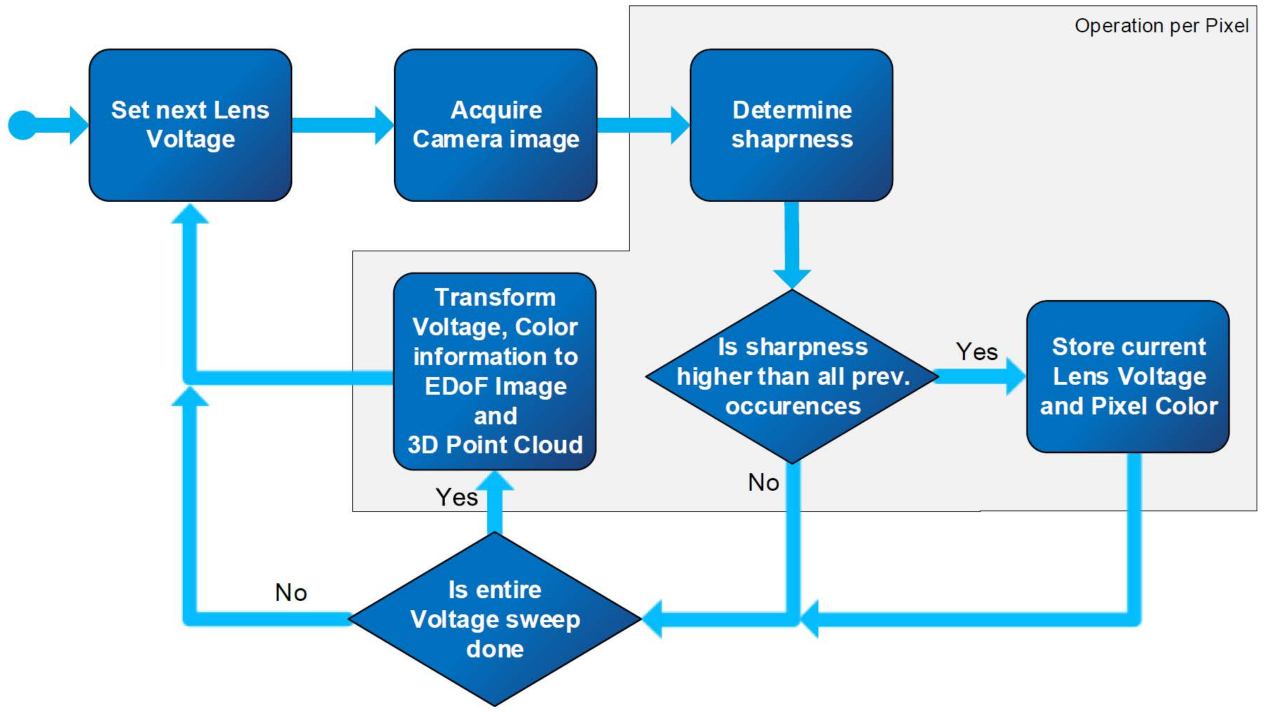

Figure 5 : Flowchart of image processing pipeline and algorithm to create EDoF image and 3D point cloud

Figure 5 illustrates the flowchart of the algorithm. A critical point is the synchronization of the image acquisition for a corresponding lens voltage. A slack in the alignment can influence the overall spatial resolution especially when operations are not executed at the desired times. Furthermore, it is important to point out that in Figure 5, the processing intensive operations to transform all pixel voltages into 3D points is only necessary to perform at the end of a complete voltage sweep. So, the most processing time-consuming parts of the algorithm are illustrated in Figure 5 by the shapes with the dashed border.

To encapsulate the entire working principle in a compact demonstrative setup we created a Development Kit able to implement the algorithm shown in Figure 6 in an Embedded System. It shows the schematic structure of the development kit's hardware and interfaces. As base we use the C-25H0-075 liquid lens which can bring objects between 70mm and 30mm in focus with the appropriate back focal adjustment. The Development Kit is also available with larger angle of view lenses to cover over 55° horizontally and vertically. As image sensor system we chose a color MIPI camera with HD (1280x720) image resolution and at least 100 frames per second (FPS) image acquisition rate to shorten the voltage sweep time. The higher the frame rate, the shorter the time to complete an entire voltage sweep of any range. When voltage sweep time is minimized then the achievable frame rate of the resulting EDoF image and 3D point cloud is maximized. To process as much as possible acquirable images by a software implementation we decided to use GPU Processing to execute the time-critical parts of the algorithm [1]. A suitable low-cost Embedded System type is the NVIDIA® Jetson Nano™ System on Module (SoM) [2] which is available including a carrier board as Development Kit [3].

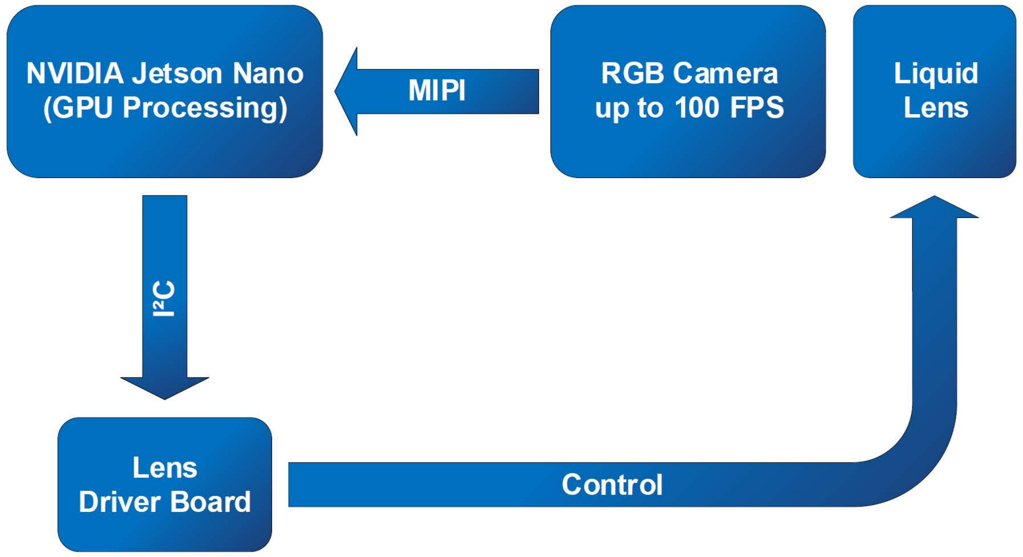

Figure 6 : Schematic structure of the CORNING Development Kit including interfaces

On the Jetson Nano SoM an Embedded Linux system is running and provides the camera image, the resulting EDoF Image and a 3D point cloud representation as Open GL ES graphics. By connecting a screen, keyboard, and mouse to the Jetson Nano carrier board it is possible to control and interact with the application software. It is possible to observe the reconstructed 3D point cloud from any point of view. The following figures show the actual setup of the CORNING Development Kit Figure 7 and a screen shot of the outputs from the Embedded Linux Desktop application on Figure 8.

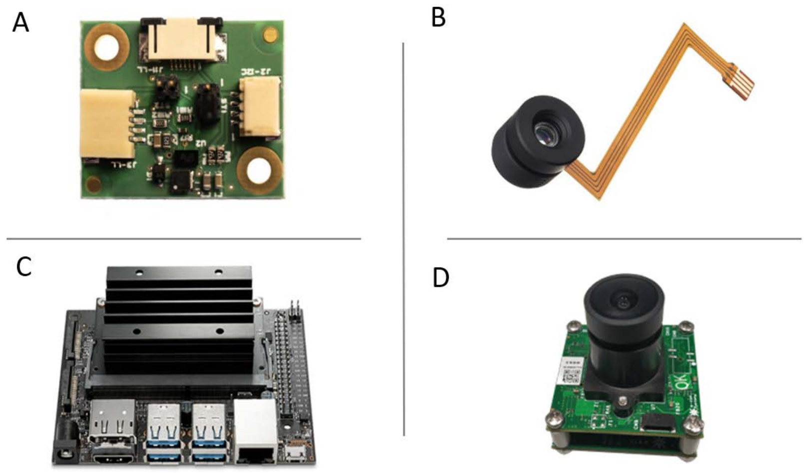

Figure 7 : CORNING Development Kit for 3D and edof reconstruction. A: Maxim Driveboard. B: Liquid lens module. C: NVIDIA® Jetson Nano™ D: Econ camera module.

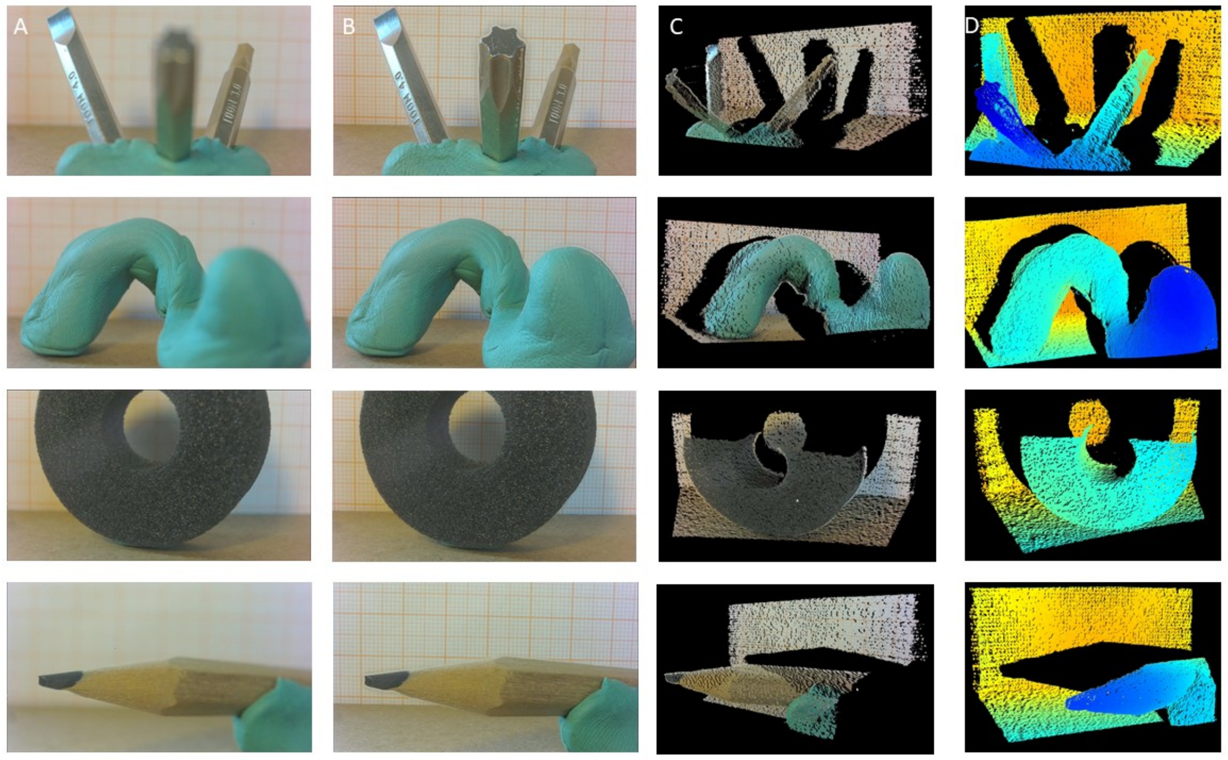

Figure 8 : Results of the EDOF and 3D point cloud estimation. A: original image captured by the camera. B: EDOF image. C: 3D point cloud. D: color depth map.

With this implementation and the Jetson Nano SoM as Processing Unit, we achieve an image processing rate of up to 44 FPS. At the used image resolution this leads to a pixel processing rate of up to 40.5 MPixels per second. This throughput counts for the algorithm shown on Figure 6 to process acquired images from the camera during the lens voltage sweep. After each completed voltage sweep the result images (EDoF image and 3D point cloud) are immediately presented on the screen without any time-consuming postprocessing. This implies that the resulting frame rate of the EDoF image and the refreshing rate of the 3D point cloud are depending on the voltage range of the sweep and the used voltage step size. Here, for typical scenarios with a nominal 3D point cloud resolution of below 1 mm one total lens voltage sweep takes 8 seconds, resulting frame or refreshing rate of 0.12 FPS.

Conclusion

Thanks to the vari-focus property of liquid lenses, we can get the estimation of depth in a 3D scene based on the sharpness analysis of a set of images. The synchronization between image acquisition and liquid lens voltage control makes it possible to match sharp pixels to an object distance. With the appropriate image processing pipeline, it is possible to reconstruct both an extended depth of field image of a scene and a 3D point cloud which can be viewed under different angles.

The imaging possibilities are dictated by the optical camera module and its parameters (focal length, aperture, pixel size) and can be adapted depending on the use case. The precision of the image point cloud will depend on the acquisition parameters, mainly the voltage steps, and the narrowness of the depth of field. A precision test under normal day light conditions showed that a spatial resolution with a standard deviation of 0.25 mm can be achieved by our implementation.

The main benefit of this 3D imaging method lies in the simplicity of liquid lens driving. The automatic voltage sweep of these variable focus lenses makes them top-quality candidates for the depth from defocus imaging technique with a monocular system.

References

(1)	CUDA by Example : an introduction to General-Purpose GPU Programing by Jason Sanders, Edward Kandrot
(2)	DATA SHEET NVIDIA® Jetson Nano™ System-on-Module
(3)	JETSON NANO DEVELOPER KIT User Guide

CORNING

SciCaTec
Dr. Markus Jäger

For additional product or technical information, please visit www.corning.com/varioptic
email varioptic@corning.com or call +33 (0) 4 37 65 35 31