Understanding 3D Imaging Performance in Sensors with Angle-Sensitive Pixels

Pascal Grégoire*, Niloufar Faghihi, Alexandre Favron, Gil Summy

Airy3D, Montréal (Qc) Canada
*Email: pascal.gregoire@airy3d.com

***Abstract*- Image sensors equipped with angle-sensitive pixels (ASPs) can extract depth images at a fraction of the cost, resources, and power requirements of current solutions, but some technical challenges remain to obtain high accuracy 3D imaging. Off-the-shelf CMOS image sensors are upgraded into ASP cameras by the addition of a transmissive diffraction mask (TDM) patterned directly on top of the sensor. In this work, we model for the first time the depth sensitivity of ASP cameras. To demonstrate the validity of the model, the depth sensitivity is measured for diverse lenses and sensor architecture. This simple approach allows to maximise depth performances for a wide range of applications.**

Keywords—3D imaging, single-sensor, angle sensitive pixels, transmissive diffraction mask, PSF centroid

I. Introduction

Image sensors equipped with angle-sensitive pixels (ASPs) can extract depth information encoded into the defocus blur of typical 2D images. It is a simple approach compared to other types of 3D imaging, such as stereo vision or time-of-flight sensors, which rely upon several components (e.g., infrared emitters and receivers, multiple sensors). ASP cameras can be realized in several different ways. For example, Airy3D has developed a 3D imaging solution to generate near-field depth based on a transmissive diffraction mask (TDM) [1], [2]. The TDM is added above the microlenses of an image sensor (Fig. 1a) and modifies the angular response of each pixel to encode depth information. Recently, other approaches based on ASPs, such as dual-pixel (DP) cameras, have been applied to depth map estimation [3]–[6]. Initially used for auto-focus [7], [8], DP requires careful design of microlenses, photodiodes and signal readouts. On the other hand, the addition of a TDM can transform an existing image sensor into a 3D sensor with no change to the pixel architecture. In both cases, the pixel angular response and the optical characteristics of the lens system determine the sensitivity of depth measurements. However, there is no clear description in the literature on how those parameters interplay [9], [10] and we are not aware of any predictive tool for depth performance.

In this work, we develop a simple and physically accurate model, based on a modified pillbox PSF, to quantify and optimize depth sensitivity. Section II describes the pixel structure and the link between 3D performance and the depth sensitivity, while Section III details the modified pillbox PSF model. Section IV shows the sensitivity measured on various 3D imagers, including a DP camera, and discusses the validity of the model using ray-tracing simulations.

II. Pixel Structure And 3D Imaging

Fig. 1a shows a typical pixel structure for a TDM-based ASP camera. The TDM structure is added on top of an existing image sensor via standard mass-production techniques. It is composed of a few-microns thick spacer layer (pedestal) and a transmissive phase grating (TDM). Incoming light on the TDM experiences a phase modulation, which converts into an angle-dependent intensity modulation once the light propagates through the structure and is

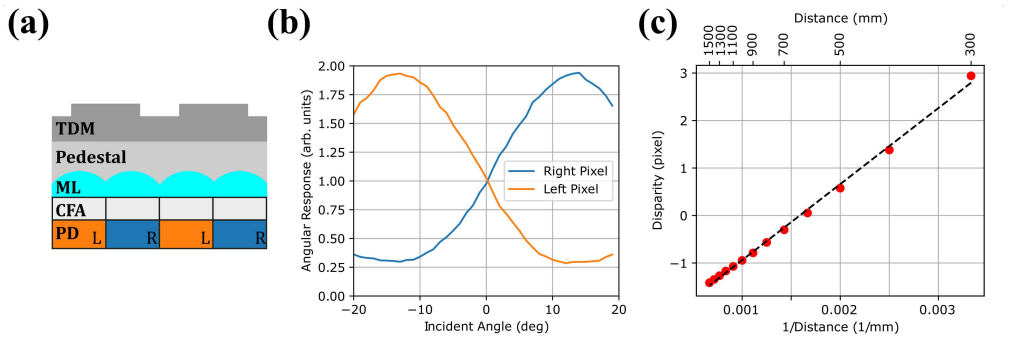

Fig. 1. (a) Schematic of an image sensor with a TDM. The photodiodes (PD), the color filter array (CFA) and the microlenses (ML) are unmodified components of an existing image sensor. A pedestal layer (light grey) and the transmissive diffraction mask (dark grey) are added above the microlenses. The TDM modifies the angular response of the photodiodes and leads to *left* (L) and *right* (R) pixels. **(b)** Measured angular response of left and right pixels for an image sensor with 1 µm pixel pitch. **(c)** Corresponding sensitivity curve using a 3.7mm F/2.0 lens. The measured disparity (red dots) is linear versus 1/distance. The slope of the linear fit (dashed line) is the depth sensitivity S = 1601 mm*pixel.

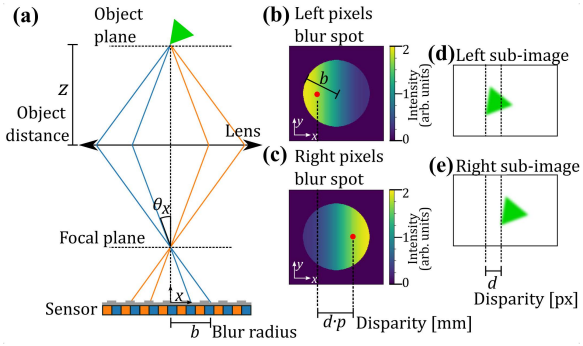

Fig. 2. (a) Thin lens model of an ASP camera. A point source at distance z from the lens forms a defocused spot of radius b at the sensor plane. The sensor is composed of alternating left (orange) and right (blue) pixels having angular responses as in Fig. 1b and a pixel pitch p. The blur spot intensity distribution of **(b)** the left pixels subset and **(c)** the right pixels subset are affected by the asymmetry of the angular response. A ray crossing the aperture on the right side of the lens reaches the sensor with a negative incident angle. The left blur spot is then more intense on the left side since the left pixels are more sensitive to negative angles. The red dots illustrate the PSF centroids, and their position difference is the disparity (Section III). Considering the image of an object placed at distance z, the left and right pixels form **(d)** a left and **(e)** a right sub-image, respectively, where the relative position of the object in the image is shifted by d pixels.

integrated by the photodiodes. The TDM effectively redistributes the light between adjacent pixels as a function of the incident angle, which leads to two subsets of pixels with asymmetric angular responses (Fig. 1b).

As illustrated in Fig. 2, each subset of pixels captures a different viewpoint of the scene, analogous to a stereo-camera system. It is then possible to measure the distance of an object based on its apparent displacement between the two sub-images (left versus right viewpoints, see Fig. 2d-e). This displacement, usually measured in pixels, is the disparity d and is related to the object distance z using

$$\mathrm{d} = \mathrm{S}(1/z - 1/\mathrm{z_F}) \qquad (1)$$

where S is the depth sensitivity and $\mathrm{z_F}$ is the focus distance. As also shown in Fig. 1c, the disparity is inversely proportional to the distance and is zero when an object is in focus ($\mathrm{z} = \mathrm{z_F}$). Once the parameters S and $\mathrm{z_F}$ are determined, the measured disparity can be transformed into a depth measurement as shown in Fig. 3.

An error in the disparity evaluation Δd leads to a depth error Δz of the form

$$\Delta z = \frac{z^2}{S}\Delta d \qquad (2)$$

which shows that maximising the depth sensitivity S reduces the depth error (increases the precision). It should be noted that Δd depends on the type of algorithm to extract the disparity, amongst other factors, while the sensitivity S is entirely dependent on the camera system parameters (lens, pixel size, angular response of pixels).

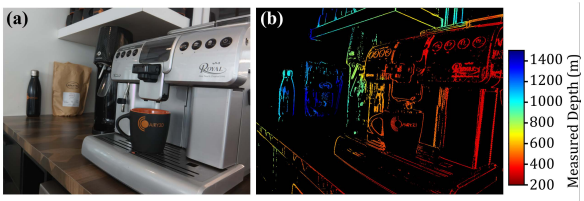

Fig. 3 Typical scene acquired with a TDM-enabled ASP camera using a 1.0 µm pixel Bayer sensor equipped with a f = 3.7mm F/2.0 lens. The color 2D image **(a)** and the depth map **(b)** are extracted from a single capture.

III. Depth Sensitivity Model

Having an accurate and simple model for the depth sensitivity S is then crucial to optimize a camera design and produce high quality depth maps.

The first step is to relate the point-spread function (PSF) of the camera to the disparity. Our starting point is the image formation model of a point source P, expressed as $\mathrm{I^{(i)}} = \mathrm{P} \otimes \mathrm{PSF^{(i)}}$, where $i = L, R$ for the left or right pixels, $\otimes$ denotes a convolution and $\mathrm{I^{(i)}}$ is the image of the point source for the left or right subset of pixels. One possible way to define the position of a defocused point source at the sensor plane is by its centroid. The disparity is then the displacement of the centroids between the left and right sub-images and expressed in pixel units p:

$$\mathrm{d} = \frac{1}{p}\left(\langle I^{(L)}\rangle_x - \langle I^{(R)}\rangle_x\right) = \frac{1}{p}\left(\langle \mathrm{PSF^{(L)}}\rangle_x - \langle \mathrm{PSF^{(R)}}\rangle_x\right), \qquad (3)$$

the centroid along the x axis of a function $\mathrm{g(x,y)}$ being defined as

$$\langle g\rangle_x = \frac{\iint x\, \mathrm{g(x,y)}\, dx\, dy}{\iint \mathrm{g(x,y)}\, dx\, dy} \quad . \qquad (4)$$

In (3), the second equality comes from the image formation model and the properties of convolutions [11]. This example can be generalized to any complex scene, not only point sources, meaning the centroid of the left and right PSFs are directly related to the disparity.

The next step is to define a model for the PSFs. We chose to use geometrical optics, for sake of simplicity, and adapted a pillbox PSF model [12]. First, assuming a thin lens approximation, there is a relationship between the angle of incidence θ_x of a ray and its position on the sensor plane x, which is related to the optical blur radius b:

$$\mathrm{b} = \frac{f^2}{2N}(1/z - 1/\mathrm{z_F}), \qquad \mathrm{f} \ll \mathrm{z} \qquad (5)$$

$$\mathrm{x} = f^2(1/z - 1/\mathrm{z_F})\theta_x, \qquad \theta_x \ll 1 \qquad (6)$$

with f the focal length and N the f-number of the lens.

The model is simple, the PSF is approximated by a circle of radius b and its intensity is modulated by the

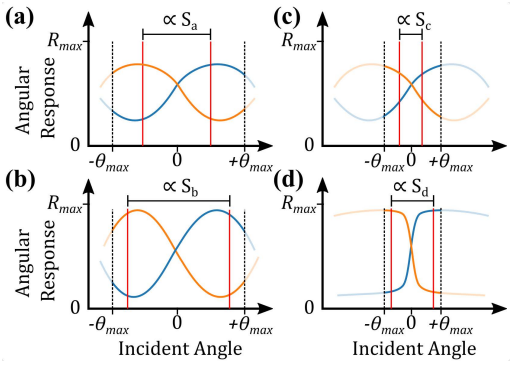

Fig. 4. Illustrating various typical PSF cross-sections in the angular space. The angular responses of the left pixels $R^{(L)}$ (orange) and the right pixels $R^{(R)}$ (blue) are shown. **(a)** The dashed lines are the maximal angles from the lens determined by the f-number and define the limits of the PSFs. The red lines illustrate the PSF centroid positions in the angle space $\langle \mathrm{PSF}^{(L)} \rangle_{\theta_x}$ and $\langle \mathrm{PSF}^{(R)} \rangle_{\theta_x}$, and their difference is proportional to the depth sensitivity S as in (9). **(b)** The amplitude of the angular response is bigger, giving a greater sensitivity. **(c)** A larger f-number reduces the sensitivity; the angular response is *clipped*. **(d)** The angular response shape maximizes the sensitivity for a larger f-number.

angular response of the pixel. The validity of this approximation is discussed in Section IV.

$$\mathrm{PSF}^{(i)} = \mathrm{A}(x, y, z) \cdot \mathrm{R}^{(i)}(\theta_x, \theta_y) \tag{7}$$

where $R^{(i)}(\theta_x, \theta_y)$ is the angular response of pixels $i = L, R$ (potentially a 2D angular response), and where A is a step function defined by the blur radius b,

$$\mathrm{A}(x, y, z) = \begin{cases} 1, & if\ x^2 + y^2 \leq b(z)^2 \\ 0, & elsewhere \end{cases} \tag{8}$$

with (x, y) the position on the sensor plane with respect to the optical axis. The left and right pillbox PSFs are illustrated in Fig. 2b-c, where (6) allows to express $R^{(i)}$ as a function of position instead of angle.

Combining (3) to (8), the disparity is

$$d = \underbrace{\frac{f^2}{p}\left(\langle \mathrm{PSF}^{(L)} \rangle_{\theta_x} - \langle \mathrm{PSF}^{(R)} \rangle_{\theta_x}\right)}_{S} (1/z - 1/z_F) \tag{9}$$

which has the same form as (1). Stated explicitly, the depth sensitivity is independent of the distance z

$$S = \frac{f^2}{p}\left[\langle a \cdot R^{(L)} \rangle_{\theta_x} - \langle a \cdot R^{(R)} \rangle_{\theta_x}\right] \tag{10}$$

with a the step function A now expressed with respect to angle (a is the numerical aperture):

$$a(\theta_x, \theta_y) = \begin{cases} 1, & if\ \theta_x^2 + \theta_y^2 \leq \theta_{max}^2, \\ 0, & elsewhere \end{cases} \tag{11}$$

with $\theta_{max} = 1/(2N)$. Equation 10 shows easily that a high focal length f and a small pixel pitch will maximise S. In practice, a long focal length is not always desirable since the focal length and the sensor size determine the angular field-of-view of the camera. Balancing the sensor size (the cost), the field-of-view and the depth sensitivity is then highly application-specific.

The last term of (10) is illustrated in Fig. 4 and can be understood as follow: the angular response is only contributing within the numerical aperture (determined by the f-number N in our model), the rest outside $[-\theta_{max}, +\theta_{max}]$ is clipped by the lens. The centroid of this *clipped* angular response needs to be as off-axis as possible to maximize the difference between the left and right PSFs. It can be accomplished with a strong asymmetry in the angular response (Fig. 4b, high ratio between the minimum and the maximum of the responses) and/or by shaping the profile of the angular response curve. For example, Fig. 4d shows a sharp transition followed by a plateau up to the maximal angle, which is optimal to increase the centroids difference.

Figure 4 shows that increasing the f-number (reducing the aperture) generally means a smaller sensitivity. Notably, two regimes can be distinguished. When the angular response is approximately linear around $\theta_x \approx \pm\theta_{max}$, such as in Fig. 4c, the sensitivity will scale as $S \propto 1/N^2$. Instead, if the angular response is constant around $\theta_x \approx \pm\theta_{max}$, like in Fig. 4d, then $S \propto 1/N$.

In the general case, the angular response needs to match with the targeted f-number. Although the response in Fig. 4d is optimal for a wide range of f-numbers, it is not possible to obtain easily with most image sensors. A sinus-like response, as in Fig. 4a-c is the norm. There, if the peak of the angular response occurs outside the numerical aperture of the lens, the contribution from the peak is lost and the sensitivity is further reduced ($1/N^2$ regime). It is possible to bring the peak of the angular response closer to $\theta_x = 0$ by adjusting the TDM design.

IV. Results And Discussion

To illustrate the validity of our model, we measured the depth sensitivity of various ASP cameras; multiple TDM-enabled image sensors and a digital full-frame camera with a DP architecture. The TDMs were applied using standard CMOS processes onto four types of image sensors with distinct pixel architectures, spanning from a 3.2 µm pixel industrial sensor to a 1.0 µm pixel Bayer sensor for mobile applications. The cameras were focused at a distance $z_F = 500$ mm, and the depth sensitivity was measured by capturing a series of planar scenes at known distances z between 300 and 1500 mm. The disparity was extracted from each image using our custom algorithm (DepthIQ™). A linear fit of the disparity vs. $1/z$ curve gives the sensitivity S (see Fig. 1c). The sensitivity in the central portion of the field-of-view is reported in Fig. 5.

The modeled sensitivity needs the angular responses, which have been measured using the sensors without a lens. Each sensor is illuminated with a distant point

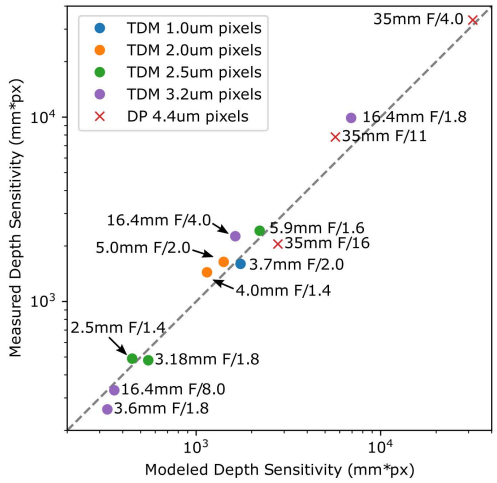

Fig. 5. Measured vs. modeled depth sensitivity for various sensor and lens combinations. The dashed line represents a perfect agreement between the measurement and the model. The measurement error is estimated to be ± 5%. Each TDM (dots) or DP (crosses) camera is labeled with the lens focal length and f-number. The depth sensitivity increases for higher focal lengths and lower f-numbers (larger apertures).

source, and the angle of the sensor plane is varied by 1 degree steps using a goniometer. The angular response is reconstructed from captures taken at each angle. As shown in Fig. 5, good agreement is found between the model and measured performances in all cases, over two orders of magnitude in depth sensitivity.

To confirm the validity of our adapted pillbox PSF model, we carried out a ray tracing simulation using Zemax OpticStudio. An F/1.8 lens with a 16.4 mm focal length (Edmund Optics 86-571) was combined with a custom layer that modulates the ray intensity at the sensor plane according to the angle of incidence. The PSF of the left and right pixels was simulated using the angular response measured from the 3.2 µm TDM-enabled sensor. Figure 6 compares the PSF of the left pixel to the case without ASPs (flat angular response). For the out-of-focus distance (Fig. 6b-c), the PSF shape is close to the geometrical optics regime. The impact of the TDM is almost exclusively a modulation of the intensity, which is similar to the pillbox model behavior. At the focus distance (Fig. 6d-e), geometrical optics is no longer valid, and the angular response is no longer visible; the PSFs with or without ASPs are identical. Even if the pillbox model is inexact in this regime, the disparity extracted by the PSF centroids as in (3) is still linear in 1/distance (Fig. 6a). In fact, the sensitivity predicted by the pillbox model (S=6920 mm*px) is very close to the sensitivity using the complete ray-tracing simulation (S=6770 mm*px). What matters is not the exact shape of the PSFs, only their centroid position, which is well captured by the pillbox model at all distances.

V. Conclusion

We have presented a complete model for the depth sensitivity of an ASP camera, based on the centroid of

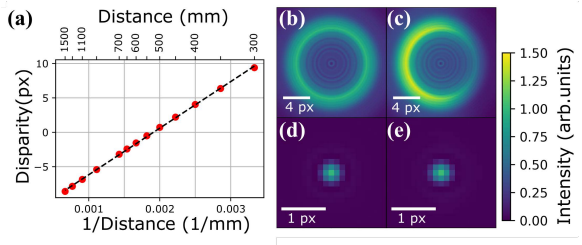

Fig. 6. Zemax OpticStudio simulation using an F/1.8 lens with a 16.4mm focal length and the 3.2 µm pixels TDM-enabled sensor. **(a)** Simulated disparity using the OpticStudio PSFs at various distances. The disparity is extracted using (3). The dashed line is a linear fit and its slope gives the depth sensitivity S=6770 mm*px. **(b)** and **(c)** are simulated out-of-focus PSFs (object distance z = 400mm, focus at z_F = 500mm) without and with ASPs, respectively. **(d)** and **(e)** are simulated PSFs at the focus distance $z = z_F = 500$mm without and with ASPs, respectively.

a modified pillbox PSF, and confirmed it accurately describes the 3D imaging capabilities of dual-pixel and TDM-based cameras. It highlights the trade-offs when optimizing the depth performance of an ASP camera for a specific application, notably the impact of the aperture size and the desired angular field-of-view onto depth sensitivity. Our approach is then a necessary tool for designing high performance angle-sensitive 3D imager.

References

[1] N. Kunnath, "Depth from Defocus using Angle Sensitive Pixels based on a Transmissive Diffraction Mask," M.S. thesis, McGill University, Montreal, 2018.

[2] G. Summy and J. Mihaychuk, "Diffraction mask design brings 3D imaging to standard CMOS image sensors," *Laser Focus World*, 2020. Accessed: Dec. 05, 2022.

[3] R. Garg, N. Wadhwa, S. Ansari, and J. T. Barron, "Learning Single Camera Depth Estimation using Dual-Pixels," in *ICCV*, Apr. 2019.

[4] A. Punnappurath, A. Abuolaim, M. Afifi, and M. S. Brown, "Modeling Defocus-Disparity in Dual-Pixel Sensors," in *ICCP*, 2020.

[5] N. Wadhwa *et al.*, "Synthetic depth-of-field with a single-camera mobile phone," *ACM Trans Graph*, vol. 37, no. 4, 2018.

[6] S. Xin *et al.*, "Defocus Map Estimation and Deblurring from a Single Dual-Pixel Image," in *ICCV*, 2021. Accessed: Dec. 04, 2022.

[7] E. S. Shim *et al.*, "All-Directional Dual Pixel Auto Focus Technology in CMOS Image Sensors," in *IEEE Symposium on VLSI Circuits, Digest of Technical Papers*, Jun. 2021, vol. 2021-June.

[8] M. Kobayashi *et al.*, "A Low Noise and High Sensitivity Image Sensor with Imaging and Phase-Difference Detection AF in All Pixels," in *IISW*, 2015. Accessed: Dec. 05, 2022.

[9] B. S. Choi *et al.*, "Analysis of disparity information for depth extraction using CMOS image sensor with offset pixel aperture technique," *Sensors (Switzerland)*, vol. 19, no. 3, Feb. 2019.

[10] K. Fukuda, "A Compressed N×N Multi-Pixel Imaging and Cross Phase-Detection AF with N×1RGrB + 1×NGb Hetero Multi-Pixel Image Sensors," in *IISW*, 2021.

[11] E. W. Weisstein "Convolution." From MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/Convolution.html

[12] S. W. Smith "The scientist and engineer's guide to digital signal processing," San Diego: California Technical Publishing, 1997, pp.400-402.