Making Sense of Commercial Stereo 3D Sensors

Part 1: Geometric Optical Performance

Introduction: Over the past years, 3D Sensing and Perception have increased in popularity. In addition to the Autonomous Vehicle (AV) fervor, Robotics, another autonomy-driven market, is beginning to make greater use of 3D Perception. 3D (or Depth) sensors are being used to enhance the awareness of Robots (and vehicles) to their surroundings. 3D Stereo sensing will play a key role in technological advancement by increasing applications and improving performance and safety.

Similar to other 3D Sensing and Perception modalities, there are many companies now offering Stereo 3D sensors for a variety of applications. Stereo 3D is highly extensible because not only does it provide 3D Depth perception, but can provide color imagery for use with well-understood and widely available Computer Vision (CV) algorithms. Given the utility and accessibility of Stereo 3D, how might one go about selecting the best Stereo 3D Sensor for a given application?

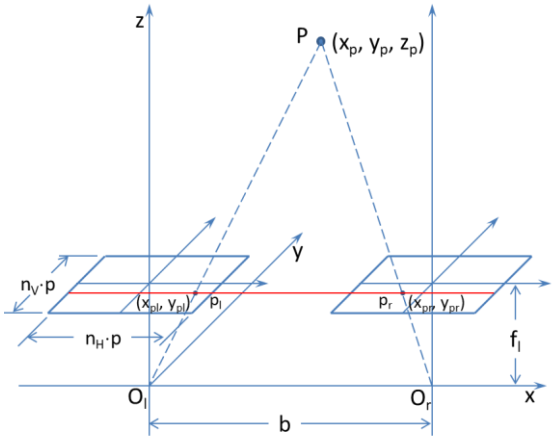

Figure 1: Stereo Geometry and Epipolar Plane

This document is the first in a series, written with the objective of assisting engineers with the selection of Stereo 3D Imagers for their applications. It provides a simple framework useful for characterizing and understanding the geometric optical performance of Stereo 3D imagers. After a review of the technical background, useful relationships will be presented to improve understanding of Stereo 3D performance given the basic parameters. In addition, the presented framework is applied to a set of commercially available Stereo 3D imagers from companies such as; Intel RealSense, Occipital, MyntAI, StereoLabs, e-Con Systems, etc.

Technical Background: In typical form, Figure 1 shows the geometry associated with Stereo 3D imaging. There are two image sensors, which are pixelated focal planes, separated by an intraocular distance, referred to as the baseline, b. The points O_l and O_r represent the lens centers. The effective focal length of the lenses, f_l, is also shown. Because the two image sensors are coplanar, they share an epipolar line along which images of the point P, p_l and p_r, lie. The difference in locations of points p_l and p_r on each focal plane is known as the disparity, d. As derived from figure 1, the three basic equations below show the 3D location of point P (i.e., z_p) is inversely proportional to the disparity.

$$z_p = \frac{b \cdot f_l}{(x_l - x_r)} = \frac{b \cdot f_l}{d} \tag{1}$$

$$y_p = \frac{b \cdot y_l}{(x_l - x_r)} = \frac{b \cdot y_r}{(x_l - x_r)} = \frac{b \cdot y_l}{d} = \frac{b \cdot y_r}{d} \tag{2}$$

$$x_p = \frac{b \cdot x_l}{(x_l - x_r)} = \frac{b \cdot x_r}{(x_l - x_r)} + b = \frac{b \cdot x_l}{d} = \frac{b \cdot x_r}{d} + b \tag{3}$$

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

After calibration, (1), (2), and (3) can be used to find the 3D location of any point P in the shared field of view of the Stereo 3D imager. (shown light blue in figure 2)

Figure 2 shows the horizontal fields of view (HFoV) of the stereo pair. At depth D_0 there are zero pixels of overlap between the imagers (i.e., $OL(z_p=D_0) = 0$)). As z_p increases above D_0, the (horizontal) depth field of view (DHFoV) also increases. Accordingly the size of the shared field, $OL(z_p)$, increases linearly with depth, z_p. Minimum useful depth, D_{min}, therefore depends on the useful size of the shared field, $OL(z_p)$.

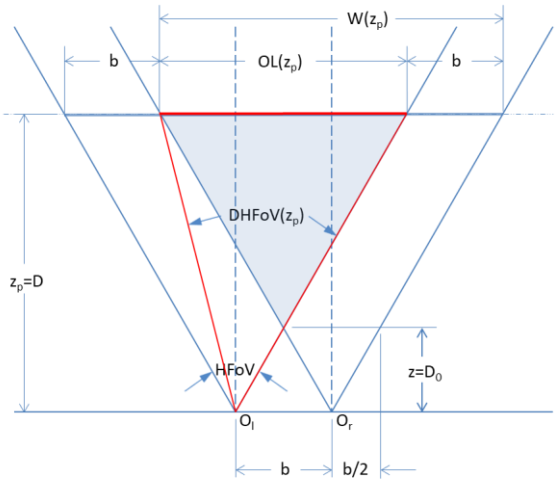

Figure 2: Shared Field (OL) and Depth Field of View (DHFoV)

Practical Stereo 3D Sensors: Figures 1 and 2 show a few of the parameters necessary to understand the geometric optical performance of Stereo 3D Sensors. These are the baseline, b, the imager focal lengths, f_l (assumed equal), the image sensor pixel pitch, p, and the imager format (n_H, n_V) (representing the number of pixels in the horizontal and vertical directions) and the horizontal fields of view, HFoV (assumed equal). While these parameters are important, it is not immediately apparent how these relate to a useful specification that is derived from or suited to a specific market application.

Generally, when specifying a Stereo 3D Sensor an engineer will consider; the maximum useful depth, $z_p = D_{max}$, the minimum useful depth, $z_p = D_{min}$, the horizontal and vertical fields of view, HFoV and VFoV, the lateral and depth resolution, the speed at which the measurements can be made (or frame rate), FR, in addition to environmental characteristics (e.g., indoor/outdoor, lighting, moisture and humidity, etc.). Available Stereo 3D sensors use imagers with different formats (n_H, n_V) and pixel pitches, p. It is easier to understand the suitability of a Stereo 3D sensor for use in specific applications if the Stereo 3D sensor parameters; p, f_l, (n_H, n_V), b, and FR can be related to application parameters, such as; D_{min}, D_{max}, lateral resolution, depth resolution, (HFoV, VFoV), and FR.

Stereo 3D Sensor Resolution: Although at this time, it is generally true that larger pixels, p, provide a potentially higher signal to noise ratio (SNR), larger pixels also increase the size of the lenses and therefore the overall size of the Stereo 3D sensor. To ease comparison of the geometric optical performance of various Stereo 3D sensors, it is useful to normalize the imaging arms of the left and right imagers, by considering a quantity, iFoV*, similar to the instantaneous field of view. This imager parameter also sets the lateral resolution of the sensor.

$$iFoV^* = \frac{p}{f_l} \tag{4}$$

Similarly, rather than considering the actual displacement (x_l, y_l) and (x_r, y_r) on each of the pixelated focal planes, normalized displacements can be taken as the number of pixels (n_{Hl}, n_{Vl}) and (n_{Hr}, n_{Vr}) to determine the normalized disparity, Δn (difference in the x-direction/horizontal pixel count). (1), (2), and (3) can be accordingly rewritten.

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

$$z_p = \frac{b \cdot f_l}{(x_l - x_r)} = \frac{b \cdot f_l}{d} = \frac{b}{iFoV^*} \cdot \frac{1}{\Delta n} = FoM \cdot \frac{1}{\Delta n} \quad (1a)$$

(1a) shows a parameter referred to as the figure of merit, FoM, for a Stereo 3D sensor. FoM scales the 'inverse disparity' to determine depth, z_p. Therefore, one would expect that a larger FoM indicates a greater depth determination capability. However, because z_p varies inversely with Δn, the depth resolution becomes poorer/more coarse as disparity decreases (range increases). Since maximum useful depth, D_{max}, is important for specifying Stereo 3D sensors, understanding how FoM might relate to D_{max} is useful. To explore the depth resolution, (1a) can be differentiated with respect to disparity.

$$\frac{dz_p}{d\Delta n} = -\frac{b}{iFoV^*} \cdot \frac{1}{(\Delta n)^2} = -FoM \cdot \frac{1}{(\Delta n)^2} = -\frac{iFoV^*}{b} \cdot {z_p}^2 = -\frac{1}{FoM} \cdot {z_p}^2 \quad (5)$$

(5) shows that a large FoM also improves/makes finer the depth resolution of the Stereo 3D sensor.

The lateral resolution is proportional to the depth, z_p, and is determined by multiplying z_p by iFoV*, see (4), whereas the depth resolution is proportional to the square of depth, $(z_p)^2$, and is determined by dividing $(z_p)^2$ by the FoM, see (5).

Stereo 3D Sensor Maximum (useful) Depth: Given the foregoing, it would be valuable to determine an expression for the maximum useful depth, D_{max}, of the Stereo 3D sensor. One way to define a useful depth measurement would be to find D_{max} subject to some constraint on the resolution at $z_p = D_{max}$. Specifically, one might want to know that, at D_{max}, a variation of one pixel in the disparity would limit the depth error to some fraction, k_D, of D_{max}. (D_{max} can be enhanced by sub-pixel interpolation).

$$D_{max}\left(\left|\frac{dz_p}{d\Delta n}\right| \le k_D \cdot z_p\right) = k_D \cdot \frac{b}{iFoV^*} = k_D \cdot FoM; \;\; 0 \le k_D < 1 \quad (6)$$

For example, if k_D = 10%, then a one pixel variation in disparity will cause a 10% error in depth, z_p at D_{max}. It should also be noted, by comparison with (1a), that for $z_p = D_{max}$, $\Delta n = (k_D)^{-1}$, or in this case, 10 pixels. This definition for D_{max} provides a useful result: D_{max} is proportional to the FoM, as set by k_D.[1].

Stereo 3D Sensor Minimum (useful) Depth and Depth (Horizontal) Field of View: As discussed above and shown in figure 2, the useful minimum depth, D_{min}, relates to the shared field (referred to as overlap, $OL(z_p)$), as subtended by the DHFoV. Figure 2 can be used to derive these important expressions.

$$W(z_p) = n_H \cdot iFoV^* \cdot z_p \quad (7)$$

$$OL(z_p) = n_H \cdot iFoV^* \cdot z_p - b \quad (8)$$

$$\%OL(z_p) = 1 - \left(\frac{b}{n_H \cdot iFoV^*} \cdot \frac{1}{z_p}\right) = 1 - \left(\frac{FoM}{n_H} \cdot \frac{1}{z_p}\right) \quad (9)$$

$$D_{min}(\%OL) = \frac{b}{(1-\%OL) \cdot n_H \cdot iFoV^*} = \frac{FoM}{(1-\%OL) \cdot n_H} \quad (10)$$

$$(n_H)_D(z_p) = \%OL(z_p) \cdot n_H \quad (11)$$

$$DHFoV(z_p) = \frac{HFoV}{2} + \tan^{-1}\left(\frac{n_H \cdot iFoV^*}{2} - \frac{b}{z_p}\right) \quad (12)$$

[1] A rule of thumb exists for Stereo 3D sensors used in parts measurement; D_{max} = 30·b. This means k_D = 3%.

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

where $(n_H)_D(z_p)$is depth (horizontal image) field at depth, z_p, (i.e., the disparity map size). Similar to the use of k_D to set the maximum useful depth, D_{max} given the sensor FoM, %OL can be used to set the minimum useful depth, D_{min} given the FoM and number of horizontal pixels, n_H. A large FoM also produces a larger D_{max} and a larger D_{min}. In this case, increasing the number of horizontal pixels, which has the effect of increasing the HFoV, can decrease D_{min}, subject to lens performance.

Framework for Comparison of Stereo 3D Imager Performance: Today, there are many companies offering Stereo 3D sensors. These sensors have different specifications and use different imagers and lenses. The foregoing allows a framework for comparing the geometric optical performance of Stereo 3D sensors. Engineers who would like to apply these sensors use application parameters such as (HFoV, VFoV), D_{max}, D_{min}, lateral and horizontal resolutions, and FR, while sensor designers select design parameters such as, (n_H, n_V), f_l and p (together iFoV*), b, and FR. HFoV can be written as a function of design parameters.

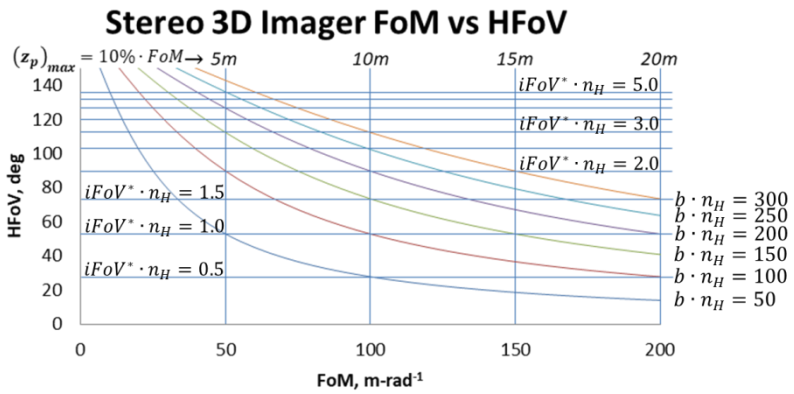

Figure 3: FoM-HFoV Plane (with b·n_H and iFoV*·n_H contours)

$$HFoV = 2 \cdot \tan^{-1}\left(\frac{n_H \cdot iFoV^*}{2}\right) = 2 \cdot \tan^{-1}\left(\frac{b \cdot n_H}{2 \cdot FoM}\right) \quad (13)$$

Based on (13), the FoM-HFoV plane framework allows various Stereo 3D sensors to be compared by plotting them as couples (FoM, HFoV), which scale to (D_{max}, HFoV) through selection of k_D appropriate for the intended application. Note that sensor depth resolution improves (becomes finer) with increasing FoM.

Considering (13), figure 3 shows a set of constant b·n_H and iFoV*·n_H contours. Therefore, given an image sensor, as defined by n_H and p, horizontal lines would suggest constant f_l contours. In any event, the intersection of the b·n_H and iFoV*·n_H contours define specific (HFoV, FoM) or (HFoV, D_{max}) designs. For example, for k_D=10%, one can immediately see that a design with (D_{max}, HFoV) = (5m, 90°) requires b·n_H=100 and iFoV*·n_H=2.0, while a design with (D_{max}, HFoV) = (10m, 90°) requires b·n_H=200 and iFoV*·n_H=2.0. It shows the well-known result; increasing only the baseline by a factor of 2 will double D_{max} (and D_{min}).

Accordingly, Stereo 3D imagers with larger b·n_H can attain higher FoM at a given HFoV and visa-versa. This means given an HFoV, higher b·n_H means higher D_{max}, D_{min} and better depth resolution. A lower HFoV translates to a smaller iFoV* if the image sensor format, n_H, remains constant.

Figure 4 shows a comparison of 16 Stereo 3D sensors from six different vendors. To illustrate the framework, the x-axis is labeled at the top in terms of D_{max}, for k_D=10%. This allows the geometric

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

optical performance of the Stereo 3D sensors to be compared and assessed for suitability for specific applications.

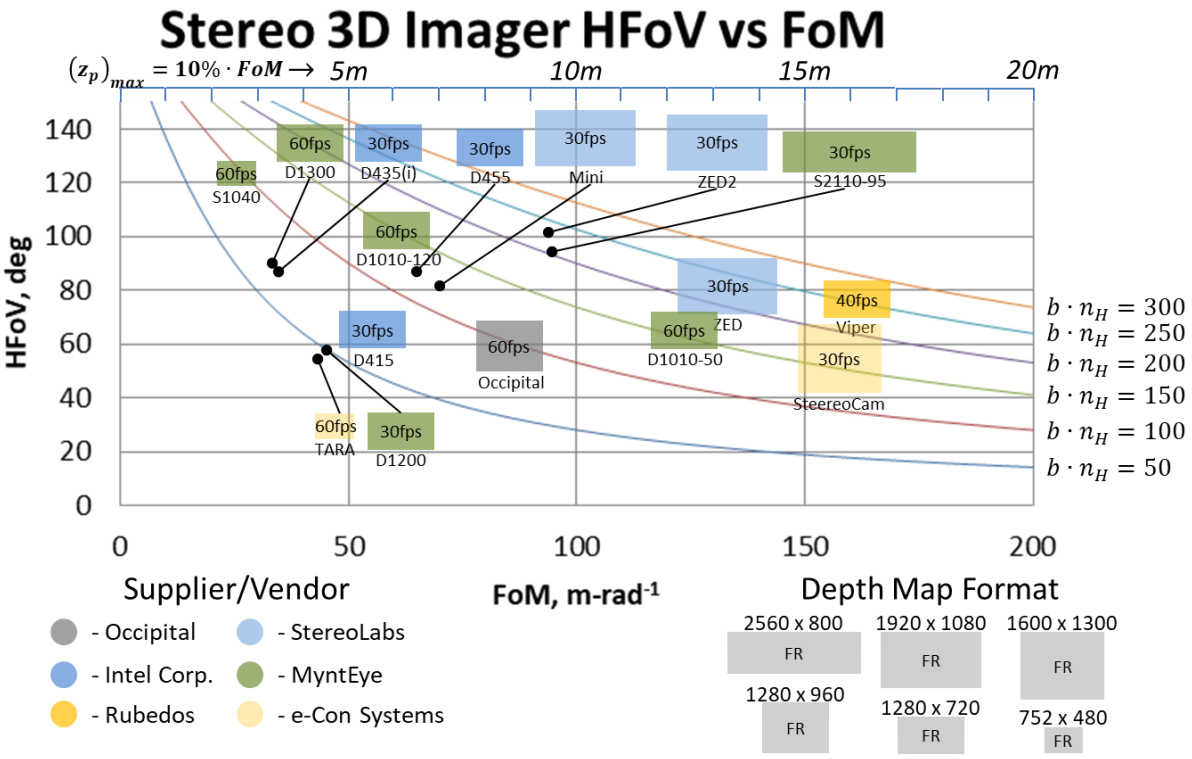

Figure 4: Stereo 3D Sensors plotted in the D_{max} vs. HFoV Plane

Commercial Offerings: Figure 4 shows 16 different Stereo 3D Sensor offerings, plotted in the FoM-HFoV plane to compare their geometric optical performance. Each sensor is represented by a rectangle showing its $z_p = \infty$ depth map format with frame rate indicated. An attempt has been made to show the highest pixel rate configurations (i.e., $FR \cdot n_H \cdot n_V$) for each supplier. Note: these sensors have other important differences: some include on-board depth processors, some utilize a pair of monochrome depth cameras, some supplement monochrome depth cameras with a color camera, others use a pair of color depth cameras, etc. In addition, some sensors include a near infrared (NIR) illuminator, an inertial measurement unit (IMU), etc. Although important, those topics will be covered in another document.

As mentioned previously, larger $b \cdot n_H$ typically means better geometric optical performance with lens focal length, f_l (or iFoV*) used to trade HFoV with FoM (or D_{max}). Therefore, before consideration of the other Stereo 3D Sensor attributes, choosing sensors with higher $b \cdot n_H$ is beneficial. However, larger n_H means higher data rates at a given frame rate, FR. So, the frame rate should also be considered, trading b for n_H to obtain the needed (D_{max}, HFoV) at data rates compatible with the application. (Note: for some targets, higher b can mean poorer correspondence performance and D_{min} will also increase.)

Competition: Figure 4 shows that StereoLabs, Intel RealSense, and MyntAI have offerings with similar performance. Recently, Intel RealSense and MyntAI have each introduced new Stereo 3D sensors:

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

RealSense D455 and MyntAI S2110-95 and MyntAI D1300. These compete directly with existing offerings. (Note: MyntAI offers S-series and D-series, the latter series offering depth processors.)

It can be seen that MyntAI's D1300 is in direct competition with the Intel RealSense D435i, both providing ($D_{max}(k_D$=10%), HFoV) ≅ (3.2m, ~90°) with similar D_{min}(%OL=90%)) ≅ 0.25m, although the D1300 operates at FR = 60fps vs 30fps.

MyntAI's S2110-95 challenges StereoLabs' ZED2 with ($D_{max}(k_D$=10%), HFoV, FR) ≅ (9.3m, ~100°, 30fps). MyntAI's D1010 series also challenges StereoLabs' ZED and ZED Mini, optically.

Intel RealSense has recently offered their highest $b{\cdot}n_H$ offering, the D455, with optical performance that competes squarely with StereoLabs ZED Mini – ($D_{max}(k_D$=10%), HFoV) = (~6.5m, ~85°) at 30fps. (MyntAI's D1010-120 is similar with ($D_{max}(k_D$=10%), HFoV) = (~6.1m, ~100°) at 60fps.) Interestingly, especially because MyntAI's and Intel RealSense's release dates are so close, the D455 falls short of the expected performance of MyntAI's S2110-95, which shows a higher FoM, resulting in a few meters of additional range, and also besting the D455's depth and lateral resolution. For the right applications, like AR/MR, Occipital's Structure seems in its own class. E-Con also offers a sensor with excellent D_{max}. The Rubedos Viper boasts high performance – it is a 'tweener, bridging performance of these commercial Stereo 3D sensors with higher end sensors from companies like; Carnegie Labs, Nerian, Roboception, etc. Various end-of-arm tooling (EOAT) companies are beginning to offer their own Stereo 3D Sensors, including OnRobot and Kinova, possibly licensed from Intel RealSense.

Applications: The application spaces for Stereo3D are growing rapidly. Table 1 attempts to improve understanding by summarizing which parameters are most important for each application. Therefore, engineering judgement should be used in specifying a Stereo 3D sensor for any application.

Table 1: Applications and Stereo 3D Parameters

	Application	*$b \cdot n_H$*	*FoM*	*HFoV*	*FR*	*IMU*
	Volume Meas.	low/mod.	mod./high	mod.	mod.	
	Facial Bio	mod.	mod.	small	mod.	
	Skeletal Bio	high	high	mod./high	mod./high	
	3D Scanning	low/mod.	mod./high	low/mod	mod./high	yes
	AR/MR	low/mod.	low/mod.	mod.	mod./high	yes
	Pick & Place	high	high	mod.	mod.	
Mobile Robots	**Vacuum/Lawn**	mod./high	mod./high	mod.	mod.	yes
	AMRs	mod./high	mod./high	mod.	mod./high	yes
	Retail Service	high	high	mod.	mod.	yes
	Security	high	high	mod./high	mod.	yes
	Ground Delivery	high	high	mod./high	high	yes
	Drone Inspection	high	high	low	mod./high	yes
	Drone Delivery	high	high	mod./high	high	yes

Although not covered in this document, the use of an IMU is valuable for Visual Odometry and simultaneous localization and mapping (SLAM), especially in mobile robotic applications. Also, an IMU makes it possible to do 3D scanning of rooms and objects, because it can be used to stitch together multiple frames, in addition to enhancing other processing capabilities useful for AR/MR.

Summary: This document has provided a framework useful for characterizing and understanding the geometric optical performance of Stereo 3D imagers. The main goal has been to help clarify how Stereo 3D sensor specifications relate to application specifications. In addition, equations were derived to allow engineers to estimate the useful maximum depth, D_{max}, given a constraint on the depth resolution

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

as a fraction of range at $z_p=D_{max}$. For this, we introduced the concept of the Stereo 3D sensor figure of merit, FoM. Because the measurement resolution of the Stereo 3D sensor is often important, mathematical expressions were presented – succinctly, lateral resolution is enhanced by decreasing iFoV* and depth resolution is enhanced by increasing FoM. Overarching are the $b \cdot n_H$ contours, which provide the capability of any Stereo 3D imager to simultaneously provide large D_{max} and HFoV, subject to data rate constraints.

Lastly, using the framework, 16 commercially available Stereo 3D Imagers were compared on their geometric optical performance and a brief discussion was provided on the competition between various players at different locations in the FoM-HFoV plane, which each correspond better to certain applications. The next documents will focus on other practical matters related to Stereo 3D Perception and Sensing.

Dave Dozor is the President of Vision Optronix, providing embedded vision and embedded motion solutions for industrial and advanced manufacturing, scientific and metrology, and security and commercial applications. For over 25years, Dave has led teams, programs, and companies to develop and commercialize advanced technologies incorporating vision and motion for a variety of applications. For a small sampling of work, please visit: http://www.vision-optronix.com/programs.

Companies mentioned that offer Stereo 3D Sensors:

Intel RealSense: https://www.intelrealsense.com

Occipital: https://structure.io/

MyntAI: https://www.mynteye.com

StereoLabs: https://www.stereolabs.com

e-Con Systems: https://www.e-consystems.com/

Rubedos: https://www.rubedos.com/

Carnegie Robotics: https://carnegierobotics.com

Nerian Systems: https://nerian.com

Roboception: https://roboception.com/en/

OnRobot: https://onrobot.com/en

Kinova: https://www.kinovarobotics.com/en

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

Appendix A: Sensitivities of the Stereo 3D Depth and Lateral Measurements

With regard to the lateral sensitivity, a pixel in object space has dimension:

$$p_{obj}(z_p) = \frac{z_p}{f_l} \cdot p = iFoV^* \cdot z_p \quad \text{(A-1)}$$

(Strictly speaking, the instantaneous field of view, iFoV, decreases as we move across the FoV. Other than for small angles, the iFoV is always lesser than the definition used in this document; iFoV* = p/f_l. It has no bearing on the results presented in this document.)

This geometric relationship can be applied to the lateral directions, similarly.

With respect to an error in disparity, it can be shown;

$$\frac{dz_p}{d\Delta n} = -\frac{b \cdot f_l}{p} \cdot \frac{1}{(\Delta n)^2} = -\frac{1}{FoM} \cdot {z_p}^2 \quad \text{(A-2)}$$

Similarly;

$$\frac{dy_p}{d\Delta n} = -\frac{b \cdot y_l}{p} \cdot \frac{1}{(\Delta n)^2} = -\frac{1}{FoM} \cdot y_p \cdot z_p \quad \text{(A-3)}$$

and;

$$\frac{dx_p}{d\Delta n} = -\frac{b \cdot x_l}{p} \cdot \frac{1}{(\Delta n)^2} = -\frac{1}{FoM} \cdot x_p \cdot z_p \quad \text{(A-4)}$$

Therefore, it can be seen that the sensitivities are inversely proportional to the FoM and proportional to the product of the depth the direction of the sensitivity sought (i.e., x_p, y_p, z_p).

In a practical sense, the 'voxels' formed in space (i.e., the shape in which a specific measurement might lie) are 'frustum' (or truncated pyramids, which are symmetric at the center of the field of view). The relationships show that these become elongated with the square of the depth and wider in proportion to depth. (In this case, the lateral directions are proportional to iFoV*. Note: iFoV* in this document is a definition not to be confused with the instantaneous field of view, iFoV, which varies across the field.)

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017

Appendix B: Instantaneous Field of View

In the foregoing document, a quantity referred to as iFoV* has been defined as the ratio of the pixel pitch p, to the effective focal length, f_l. For small angles, iFoV* is approximately equal to the instantaneous field of view, iFoV. However, the instantaneous field of view, iFoV is a function of the pixel number, m. As derived from figure B1 and plotted in figure B2, the relationship between the iFoV and pixel number is as follows.

$$iFoV(m) = \tan^{-1}\left[\frac{\left(\frac{p}{f_l}\right)}{1+m(m+1)\cdot\left(\frac{p}{f_l}\right)^2}\right]; \quad for\ \left(\frac{n}{2}\right) \le m \le \left(\frac{n}{2}-1\right) \tag{B-1}$$

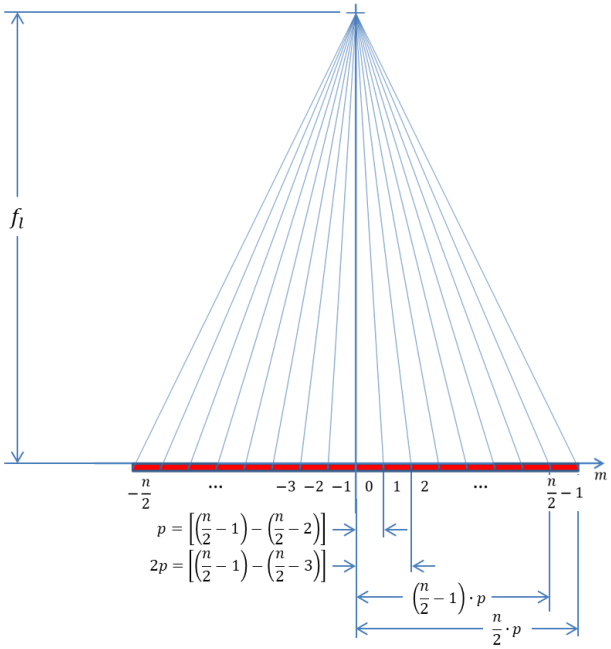

Figure B1: Diagram showing $iFoV(m)$

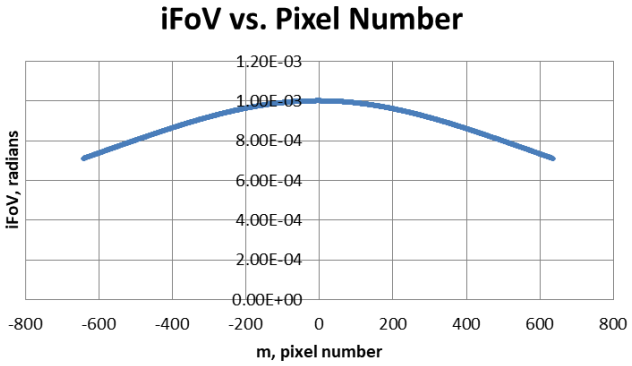

Figure B2: Plot of $iFoV(m)$ for n_H = 1280 and $iFoV^* = \frac{p}{f_l} = 1.0\ mrad$

P.O. BOX 29, BELMONT, MA 02478 USA | info@voptronix.com | PHONE: +1 617.489.1017