Three-Dimensional Sensors
Lecture 1: Introduction to 3D Sensors

Radu Horaud
INRIA Grenoble Rhone-Alpes, France
Radu.Horaud@inria.fr
http://perception.inrialpes.fr/

What Is a 2D Image?

- We recall that a *color* camera provides an RGB (red-green-blue) **two-dimensional** image of the **three-dimensional** world.
- The RGB values at each pixel location depend on many factors, such as:
  - Light source: position, wave length, number of sources.
  - Object: color, physical properties (opaque, translucide, reflective), orientation.
  - Scene composition: mutual illumination (shading).
  - Projection: loss of information, projective distorsion.
  - Sensor capabilities.
- In spite of this, we are very good at inferring 3D information from a single image – This remains a difficult computational problem!

What About Two (or Several) 2D Images?

- This is called *stereo vision*: binocular, multiple-camera, etc.
- It is also referred as *passive range*.
- All the animals have two eyes, but only a few species have stereo: predators (tigers, cats), birds (owls), primates (monkeys, humans).
- Hint: Observe eye placement, eye movements, and nose size!
- Other animals have developed *active range* strategies, e.g., sonars (bats, whales, etc.).

3D Sensor Technologies

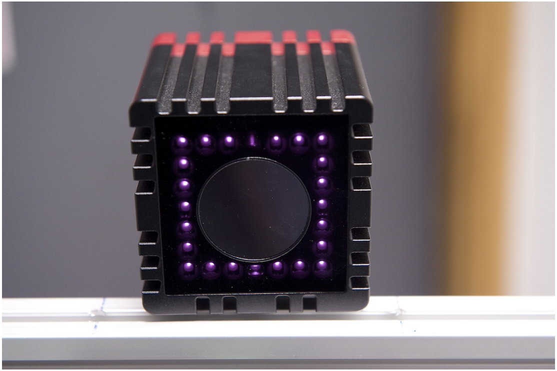
- Sensors that can measure depth (or distance, or range) to an object based on illuminating the object with a *laser* (or other kinds of light source), and measuring the backscattered light. There are two types of such sensors:
- Projected-light sensors that combine the projection of a light pattern with a standard 2D camera and that measure depth via **triangulation**, and
- Time-of-flight sensors that measure depth by estimating the **time delay** from light emission to light detection. There are two type of TOF cameras:
  - A point-wise TOF sensor is mounted on a two dimensional (pan-tilt) scanning mechanism, also referred to as a LiDAR, or *light detection and ranging*.
  - Matricial TOF cameras that estimate depth in a single shot using a matrix of TOF sensors (in practice they use CMOS or CCD image sensors coupled with a lens).

The Kinect Projected-Light Depth Camera

The ENSENSO *Stereo 3D Camera*

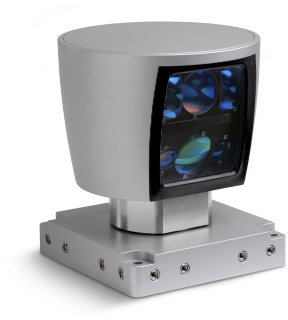
`http://fr.ids-imaging.com/ensenso.html`

The Time of Flight SR4000 Depth Camera



http://www.mesa-imaging.ch/

The VELODYNE LiDAR Scanners



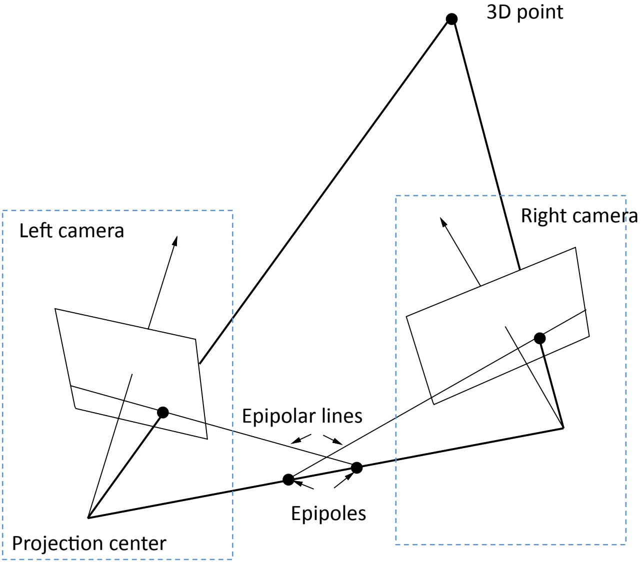
HDL-64E HDL-32E

http://velodynelidar.com/lidar/lidar.aspx

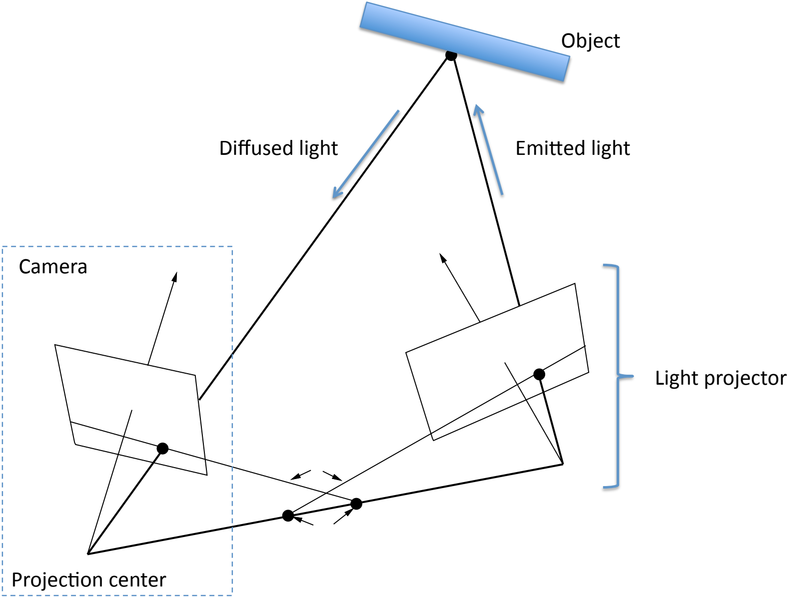
TigerEye 3D Flash LIDAR Camera

http://www.advancedscientificconcepts.com/index.html

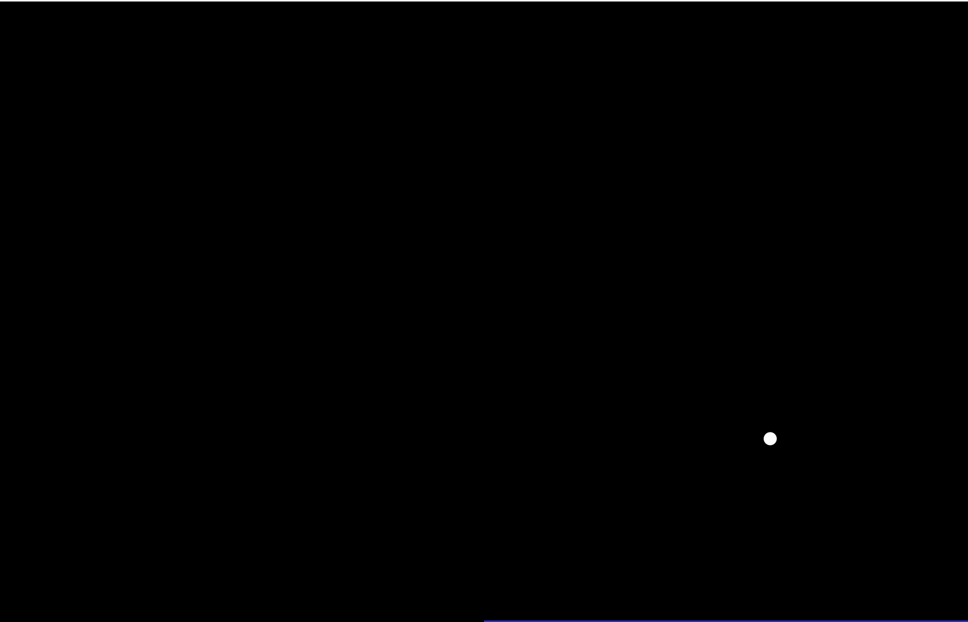
The Geometry of Passive Stereo



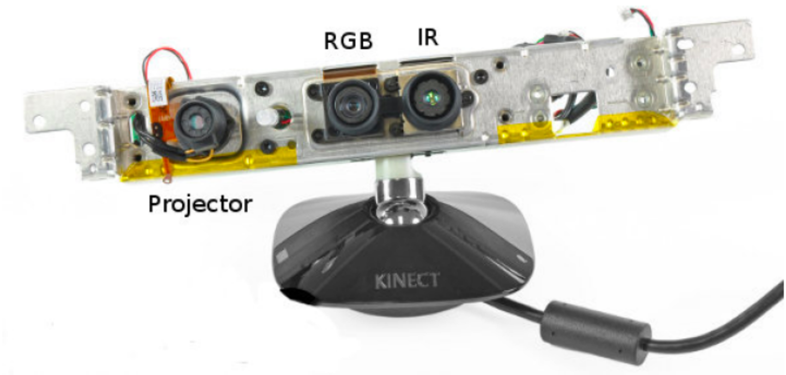
The Geometry of Active Stereo



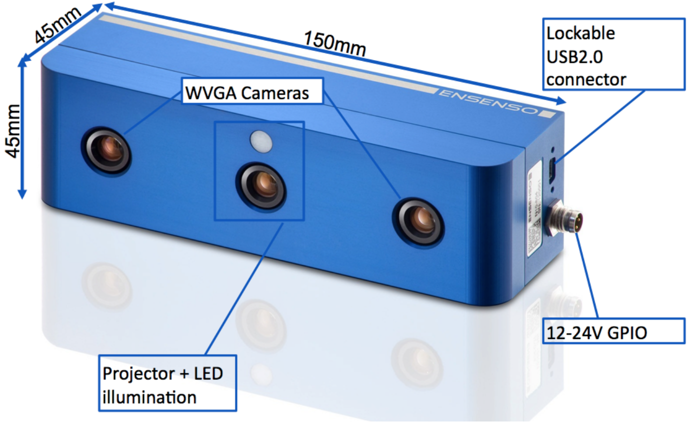
Projecting a Light Beam!



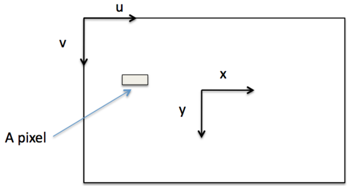
Inside the Kinect Depth Camera



ENSENSO: Stereo and Projected Light



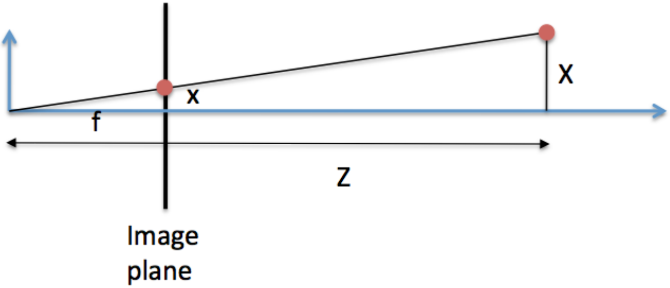
Camera Calibration Made Simple (I)



$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \approx \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

- horizontal and vertical scale factors (in pixels/mm): k_u and k_v
- the image center (in number of pixels): u_0 and v_0
- the transformation is written in *homogeneous* or *projective* coordinates defined up to a scale factor, i.e., $\approx$.

Camera Calibration Made Simple (II)



- $x = f\frac{X}{Z}$ and $y = f\frac{Y}{Z}$
- This can be written as a projective transformation:

$$\begin{pmatrix} sx \\ sy \\ s \end{pmatrix} \approx \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Camera Calibration Made Simple (III)

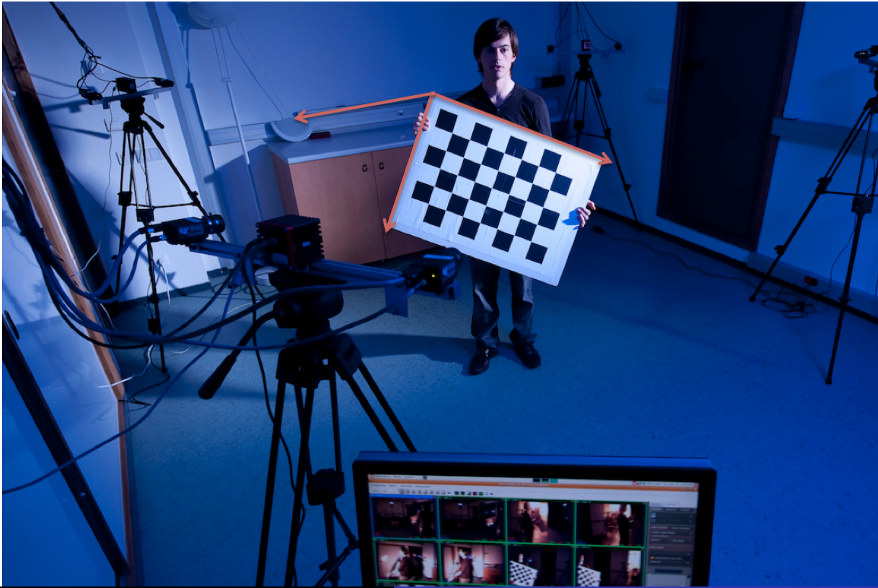
- A change of 3D coordinates from the *calibration frame* to the camera frame:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}}_{\mathbf{R} \text{ (rotation)}} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} + \underbrace{\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}}_{\boldsymbol{t} \text{ (transl.)}}$$

- In projective coordinates:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \approx \begin{bmatrix} \mathbf{R} & \boldsymbol{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}$$

Practical Calibration Setup



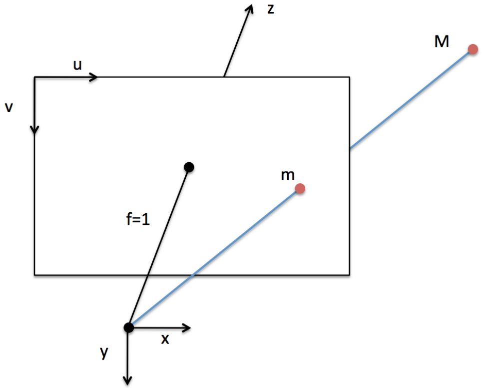
Camera Parameters

- Putting everything together, we obtain:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \approx \underbrace{\begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A} \text{ (intrinsic)}} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} sx \\ sy \\ s \end{pmatrix} \approx \underbrace{\begin{bmatrix} \mathbf{R} & \boldsymbol{t} \end{bmatrix}}_{\mathbf{D} \text{ (extrinsic)}} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}$$

The *Virtual* Camera Coordinates



- The coordinates of the image point m:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \approx \mathbf{A}^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

- The coordinates of the world point M:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = Z \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

- Depth along a line of sight through m: $\frac{\|\boldsymbol{M}\|}{\|\boldsymbol{m}\|} = \frac{Z}{1}$ (1 is the focal length)

Linear Triangulation

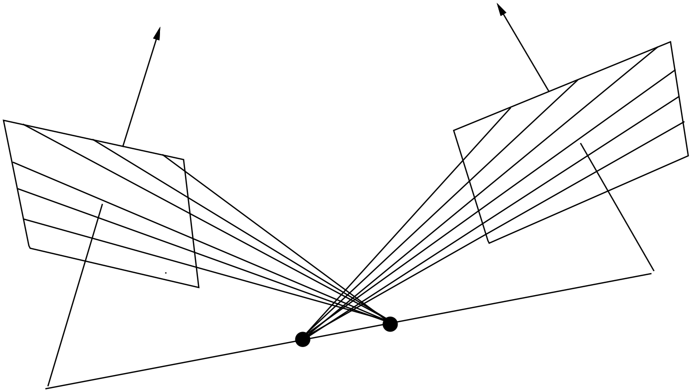
The left and right cameras are calibrated and by a simple change of reference frame, we express everything in the reference frame of the left camera:

- The left camera: $s\boldsymbol{m}_l \approx \left[\begin{array}{cc} \mathbf{I} & \mathbf{0} \end{array} \right] \boldsymbol{M}$
- The right camera: $s'\boldsymbol{m}_r \approx \left[\begin{array}{cccc} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{r}_3 & \boldsymbol{t} \end{array} \right] \boldsymbol{M}$

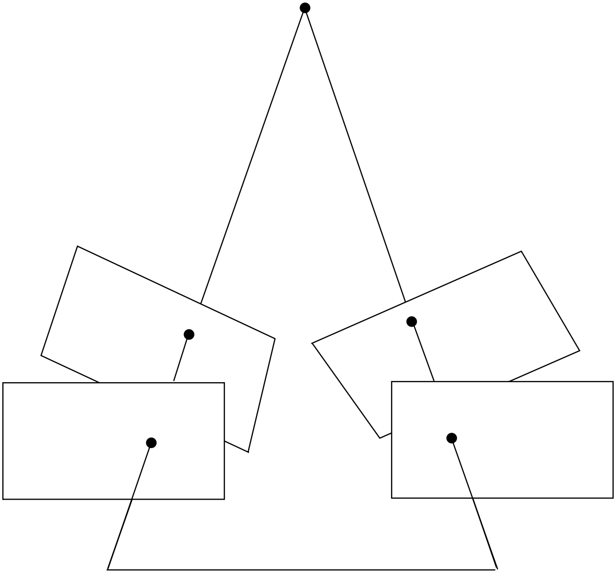
This yields a set of four linear equations in the unknowns X, Y, and Z:

$$\begin{cases} X = x_l Z \\ Y = y_l Z \\ \boldsymbol{r}_1^\top \boldsymbol{M} + t_1 = x_r(\boldsymbol{r}_3^\top \boldsymbol{M} + t_3) \\ \boldsymbol{r}_2^\top \boldsymbol{M} + t_2 = y_r(\boldsymbol{r}_3^\top \boldsymbol{M} + t_3) \end{cases}$$

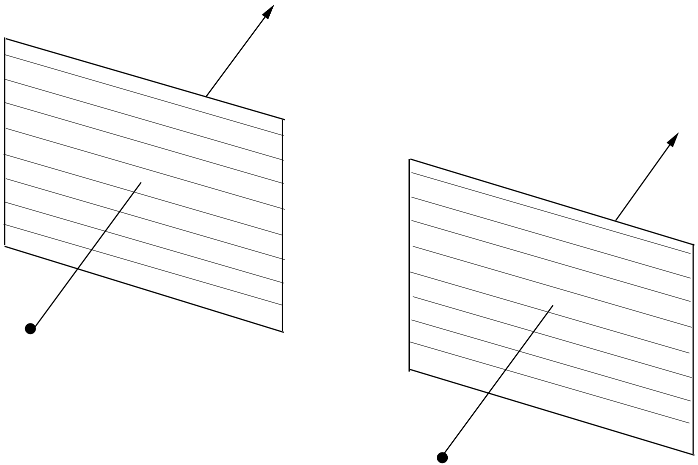
General Stereo Configuration



Rectification



Rectified Stereo Pair



Disparity

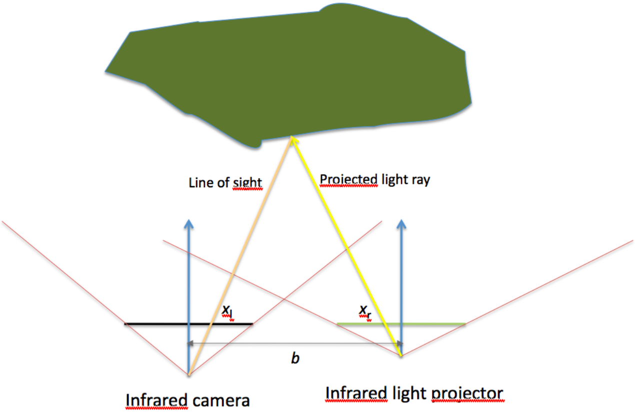
A rectified image pair:

- The left camera: $s\boldsymbol{m}_l \approx \left[\begin{array}{cc} \mathsf{I} & \mathbf{0} \end{array}\right] \boldsymbol{M}$
- The right camera: $s'\boldsymbol{m}_r \approx \left[\begin{array}{cc} \mathsf{I} & \boldsymbol{t} \end{array}\right] \boldsymbol{M}$ with $\boldsymbol{t} = (b\ 0\ 0)$
- This yields a set of simplified equations:

$$\begin{cases} X = x_l Z \\ Y = y_l Z \\ X + b = x_r Z \\ Y = y_r Z \end{cases}$$

- Disparity: $d = x_r - x_l$
- Depth is inversely proportional to disparity: $Z = \frac{b}{d}$

Geometry of Kinect (Rectified)



Time of Flight Sensors

- Basic principle: measure the absolute time that a light pulse needs to travel from a targeted object to a detector.
- The speed of light is of 0.3 meters/nano-second.
- TOF systems use either *pulsed-modulation* or *continuous wave modulation*:
  - **Pulsed-modulation** measures the time-of-flight directly. Allows long-distance measurements. The arrival time must be detected very precisely. Needs very short light pulses with fast rise and fall times and with high optical power – lasers or laser diodes.
  - **CW-modulation** measures the phase difference between the sent and received signals. Different shapes of signals are possible, e.g., sinusoidal, square waves. Cross-correlation between the received and sent signals allows phase estimation which is directly related to distance if the modulation frequency is known.

Depth Estimation by Cross-Correlation (I)

- The distance is computed with: $d = \frac{1}{2}c\tau$.
- Relationship between phase and travel time: $\phi = 2\pi f \tau$ where f is the modulation frequency.
- Emitted signal: $s(t) = a\cos(2\pi f t)$
- Received signal: $r(t) = A\cos(2\pi f(t-\tau)) + B$
- Demodulation: Cross-correlation between received and sent signals:

$$C(x) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} r(t)s(t+x)dt$$

- The solution is:

$$C(x) = \frac{aA}{2}\cos(\underbrace{2\pi f(\tau}_{\phi} + x)) + B$$

Phase, Amplitude, and Offset

- The cross-correlation function is evaluated at four selected phases, namely $2\pi f x_0 = 0^0$, $2\pi f x_1 = 90^0$, $2\pi f x_2 = 180^0$, $2\pi f x_3 = 270^0$.
- This yields the following solutions for the phase, amplitude and offset images:

$$\begin{aligned}
\phi &= 2\pi f \tau = \arctan\left(\frac{C(x_3) - C(x_1)}{C(x_0) - C(x_1)}\right) \\
A &= \frac{1}{2a}\sqrt{(C(x_3) - C(x_1))^2 + (C(x_0) - C(x_1))^2} \\
B &= \frac{1}{4}\left(C(x_0) + C(x_1) + (C(x_2) + C(x_3)\right)
\end{aligned}$$

Depth from Phase and Modulation Frequency

- Depth: $d = \frac{1}{2}c\tau = \frac{c}{2f}\frac{\phi}{2\pi}$
- Minimum depth: $d_{\min} = 0 \ \ (\phi = 0)$
- Maximum depth: $d_{\max} = \frac{c}{2f} \ \ (\phi = 2\pi)$.
- Phase wrapping: there is a 2π ambiguity in the computation of the phase. Hence the distance is computed up to a **wrapping** ambiguity:

$$d = \left(\frac{\phi}{2\pi} + n\right) d_{\max}$$

  where $n = 0, 1, 2, \ldots$ is the number of wrappings.
- For $f = 30$MHz, the unambiguous range is from 0 to 5 meters.
- Solving for this ambiguity is called **phase unwrapping**.

Geometry of a TOF camera