A Short Tutorial on Three-Dimensional Cameras

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

Binocular Vision

- This is also called *stereo vision*: depth can be inferred by computing a disparity field at each pixel location.
- 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 Sensors

- They measure depth based on illuminating the scene with a a controlled light source, and and on 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 TOF principles: (i) modulated-light and (ii) pulsed-light.
- There are two types 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 system).

Examples

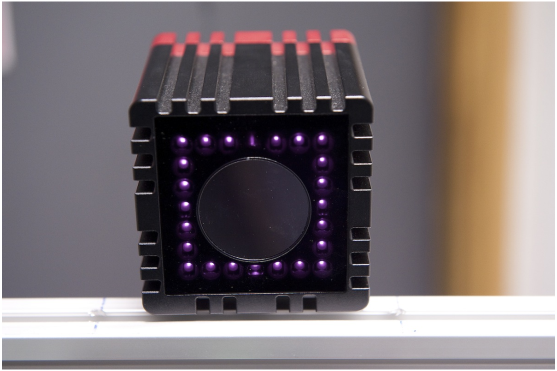
- Projected-light cameras: Kinect, Asus Xtion Pro Live, Ensenso, etc.
- Modulated-light time of flight: SR4000 (Swiss Ranger)
- Pulsed-light time of flight: Tiger Eye (Advanced Scientific Concepts, Inc.)
- Pulsed-light rotating scanner: Velodyne.

The Kinect Projected-Light Depth Camera

The ENSENSO *Stereo 3D Camera*

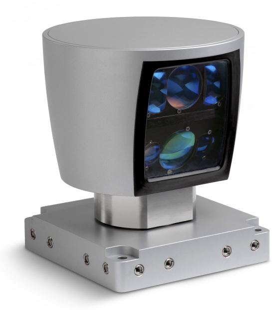
http://fr.ids-imaging.com/ensenso.html

The Time of Flight SR4000 Depth Camera



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

The VELODYNE LiDAR Scanners



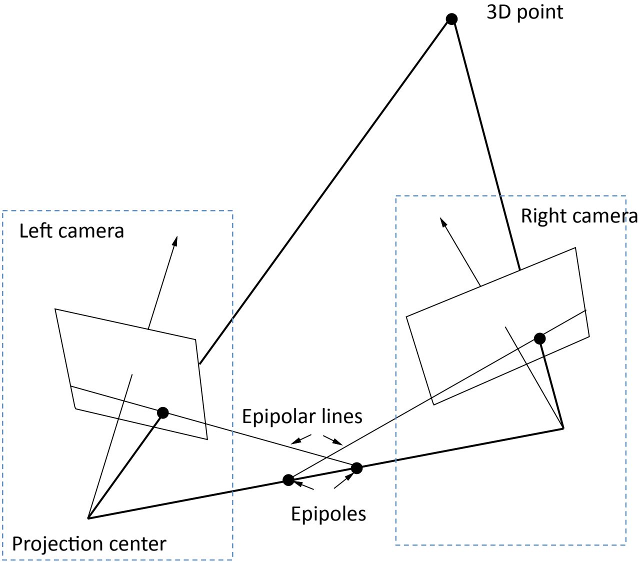
HDL-64E HDL-32E

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

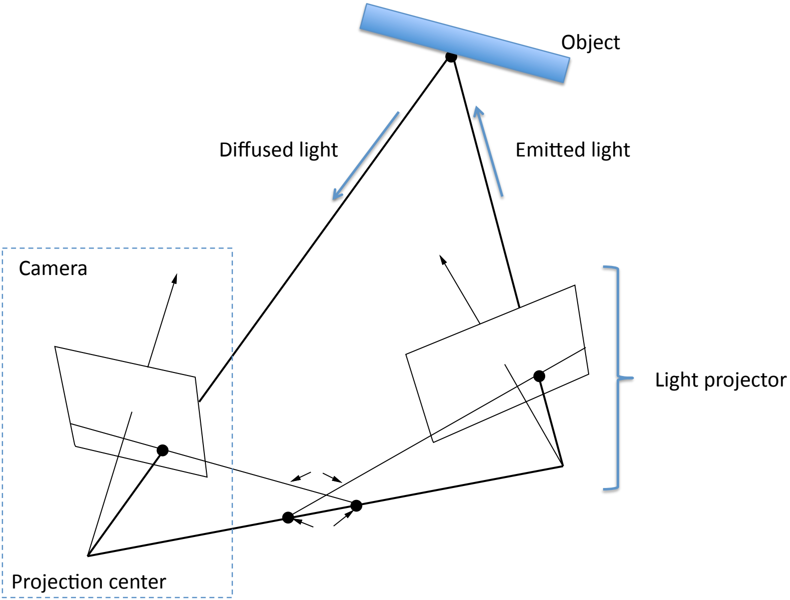
TigerEye 3D Flash LIDAR Camera

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

The Geometry of Passive Stereo



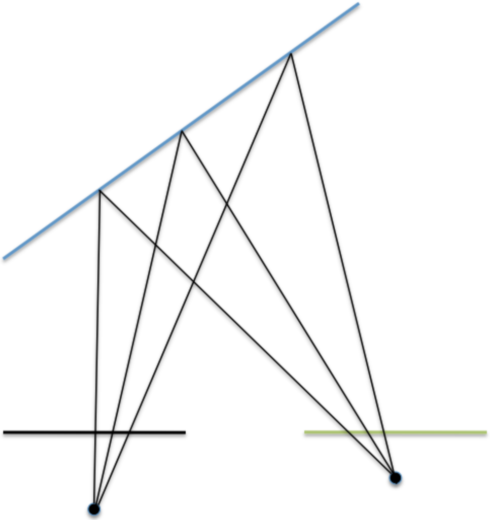
The Geometry of Active Stereo



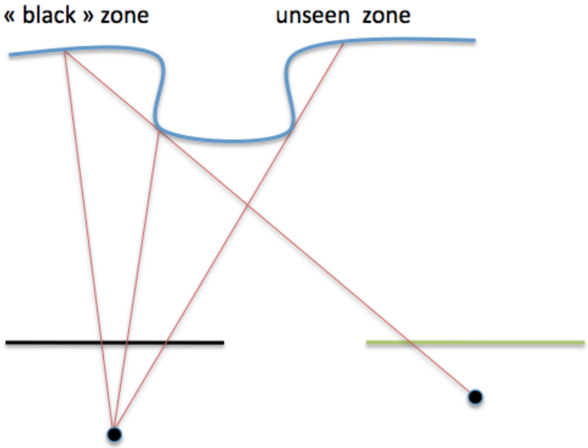
The Geometry of Active Stereo

- Perspective effects (foreshortening).
- Surface discontinuities.
- Cluttered scenes.

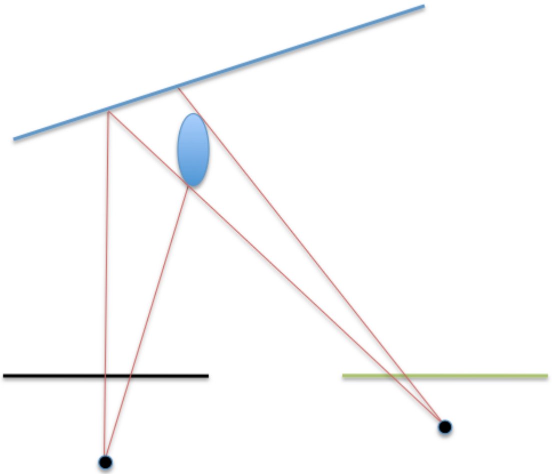
Foreshortening



Discontinuities



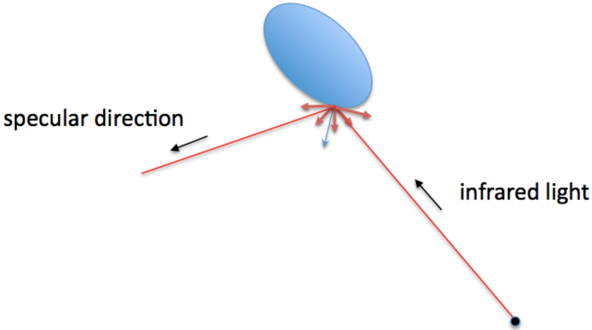
Clutter



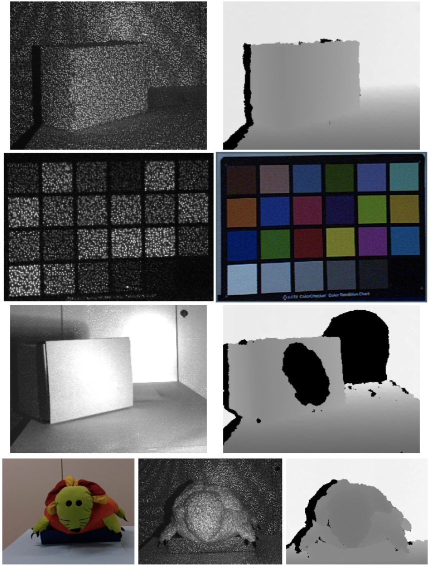
Surface Reflectivity and External Illumination

- This is a scientific topic in its own right.
- Reflectance (or absorption) depends on the material properties and of the wavelength of the incident light.
- A rough way to think of surface reflectivity is to divide surfaces into **diffuse** (matte) and **specular** (mirror).
- In general a surface is a combination of these effects plus other ones: metal, glossy, rough, transparent, translucent, etc. cause problems in practice.
- External illumination (daylight, artificial light, etc.) acts as *noise* added to the infra-red light emitted by the projector.

Reflectivity



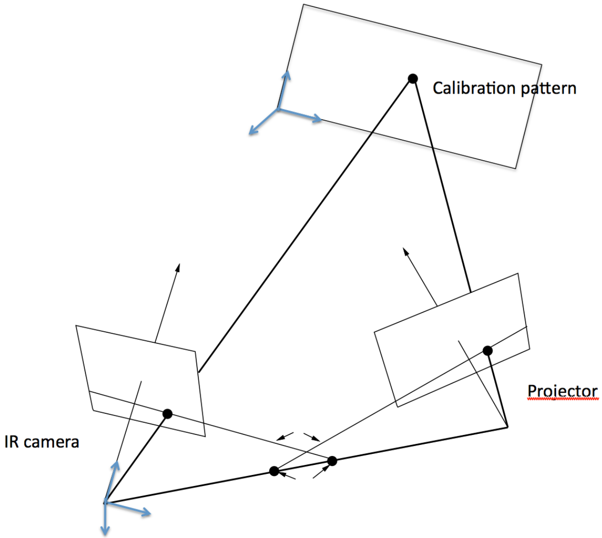
Artifacts with the Kinect Camera



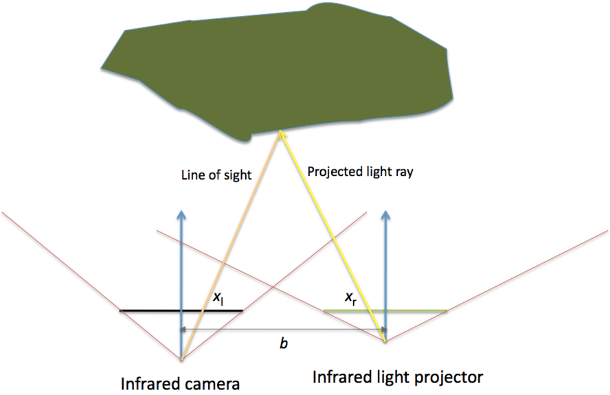
Back to Geometry

- Consider an IR camera and a projector in general positions (before rectification).
- Parameters to be estimated:
  - The internal parameters of the IR camera
  - The external parameters of the IR camera
  - The projection matrix of the light projector

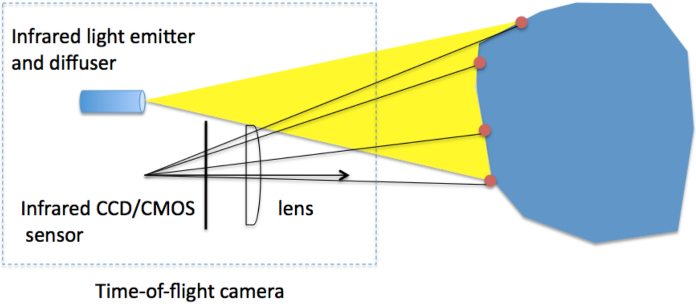
IR-Camera and Projector Geometry and Calibration



Geometry of Kinect (Rectified)



Geometry of a TOF camera



Depth Estimation

- The distance is computed with:

$$d = \frac{1}{2} c\tau$$

- In practice τ cannot be measured directly.
- Continuous wave modulation: The phase difference between the sent and received signals is measured.
- The modulation frequency is in the range 10 to 100 MHz.
- Relationship between phase shift and time of flight (f is the modulation frequency):

$$\phi = 2\pi f \tau$$

Cross-Correlation Between Sent and Received Signals

- Emitted signal: $s(t) = a\cos(2\pi f t)$
- Received signal: $r(t) = A\cos(2\pi f(t-\tau)) + B$
- Cross-correlation between emitted and received signals:

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

- The solution is:

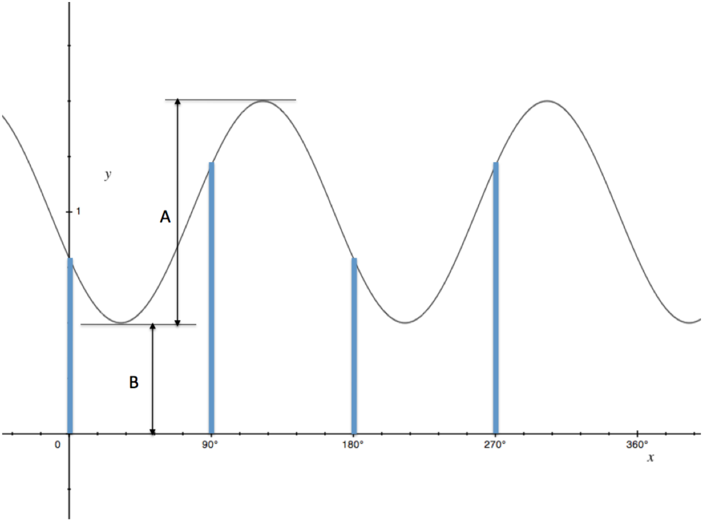
$$C(x) = \frac{aA}{2}\cos(\underbrace{2\pi f\tau}_{\phi} + \underbrace{2\pi f x}_{\psi}) + B$$

Demodulation Parameters

- ϕ is the phase and is defined up to 2π, this is called phase wrapping.
- A is the amplitude of the received signal and it depends on the object's reflectivity and of the sensor's sensitivity. The amplitude decreases with $1/d^2$ mainly due to light spreading.
- B is an offset coefficient due to the ambient illumination.

The 4-Bucket Method

- Estimate four values of $C(x)$ at $\psi_0 = 0^0$, $\psi_1 = 90^0$, $\psi_2 = 180^0$, $\psi_3 = 270^0$



Phase, Amplitude, and Offset

- From these four values of the cross-correlation signal we obtain the following solutions for the phase, amplitude and offset:

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

Implementation Based on CCD/CMOS Technologies

- The CCD (charge-coupled device) plays several roles:
  - Data acquisition or readout operation: the incoming photons are converted into electron charges.
  - Clocking.
  - Signal Processing (demodulation).
- After the demodulation, the signal $C(\psi)$ is integrated at four equally space intervals, over an equal-length Δt, within one modulation period T.
- These four signal values are stored independently.
- The cycle of integration and storage can be repeated over many periods.
- Example: for $f = 30$MHz and at 30 frames per second (FPS), 10^6 integration periods are possible.

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: an inherent 2π phase ambiguity.
- Hence the distance is computed up to a **wrapping** ambiguity:

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

where $n = 0, 1, 2, \ldots$ is the number of wrappings. This can also be written as:

$$d(i,j) = d_{\mathsf{tof}}(i,j) + n(i,j) d_{\max}$$

where d is the *real* depth value and d_{tof} is the *measured* depth value.

More on Phase Wrapping

- For $f = 30$MHz, the unambiguous range is from $d_{\min} = 0$ to $d_{\max} = 5$ meters.
- Solving for this ambiguity is called **phase unwrapping**.
- The modulation frequency of the SR4000 camera can be selected by the user. The camera can be operated at:
  - $29/30/31$ MHz corresponding to a maximum depth of 5.17m, 5m and 4.84m.
  - $14.5/15/15.5$ MHz corresponding to a maximum depth of 10.34m, 10m and 9.67m.
- The accuracy increases with the modulation frequency.

Observations

- A TOF camera works at a very precise modulation frequency. Consequently, it is possible to simultaneously and synchronously use several TOF cameras by using a different modulation frequency for each one of the cameras, e.g., six cameras in the case of the SR4000.
- A TOF camera needs a long integration time (IT), over several time periods, to increase SNR, hence accuracy. In turn, this introduces "motion blur" in the presence of moving objects. Because of the need of long IT, fast shutter speeds (as done with standard cameras) cannot be envisaged.
- Sources of errors:
  - Demodulation error;
  - Integration error;
  - Temperature error.

From Depth Values to Euclidean Coordinates

- The TOF camera measures the depth d from the 3D point M to the optical center, hence:

$$d = \|\boldsymbol{M}\|$$

.

- Relationship between the *image coordinates* $\boldsymbol{p} = (u\ v\ 1)^\top$ and the *virtual camera coordinates* $\boldsymbol{m} = (x\ y\ 1)^\top$ of the point m:

$$\boldsymbol{m} = \mathbf{A}^{-1}\boldsymbol{p}$$

- Hence:

$$Z = \frac{\|\boldsymbol{M}\|}{\|\boldsymbol{m}\|} = \frac{d}{\|\mathbf{A}^{-1}\boldsymbol{p}\|}$$

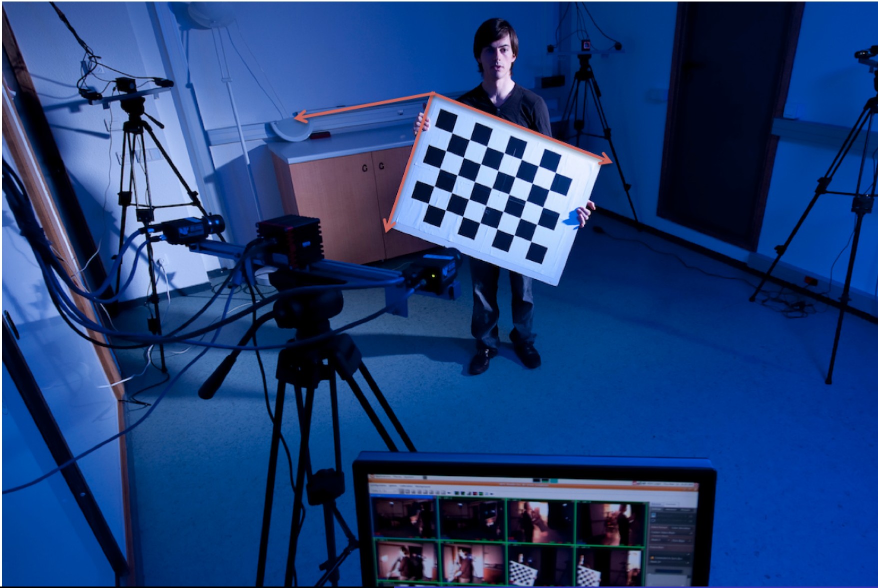
- We obtain for the Euclidean coordinates of the measured point M:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \frac{d}{\|\mathbf{A}^{-1}\boldsymbol{p}\|}\mathbf{A}^{-1}\begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

TOF Camera Calibration

- The CCD image sensor of a TOF camera provides both a depth image and an amplitude+offset image.
- The amplitude+offset image can be used by the OpenCV packages to calibrate the camera, namely its intrinsic and lens-distorsion parameters.
- However, the low-resolution (approximately 170×150 pixels) of the TOF images implies some practical considerations.

Practical Calibration of a TOF-Stereo Setup



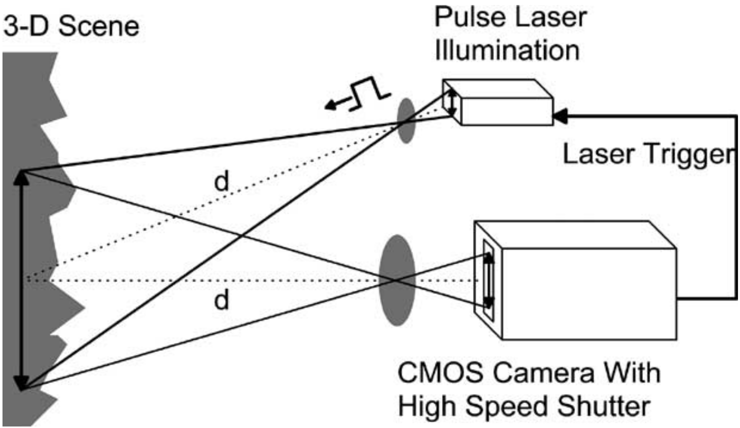
Time of Flight with Pulsed Light

- A light pulse of a few *nanoseconds* is generated by a laser;
- Distance is determined directly from the *time delay* between emitted light pulse and its reflection;
- It can use very short pulses with high optical power: The pulse irradiance is much higher than the background irradiance;
- The emitted laser energy remains low (class 1);
- Does not suffer from the phase ambiguity problem;
- It is the technology of choice in a number of outdoor applications under adverse conditions: surveying (static and mobile), autonomous driving, cultural heritage, planetary missions.

Sensor Types

- Laser scanners use a rotating mirror;
- 2D (line) scanners (horizontal or vertical);
- 3D scanners (horizontal and vertical);
- Multiple-laser 2D line scanners (horizontal and vertical);
- 3D flash Lidar cameras (provide a depth image without any rotating mechanism).

The Principle of 3D Flash Cameras

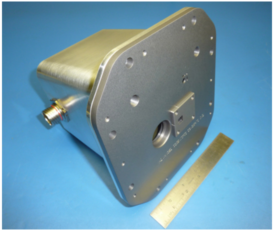


Elkhalili et al. (2004). IEEE Transactions on Solid-State Circuits. Vol. 39. No 7, pages 1208–1212

3D Flash Lidar Cameras from Advanced Scientific Concepts Inc.

Tiger Eye

Portable 3D



Dragon Eye

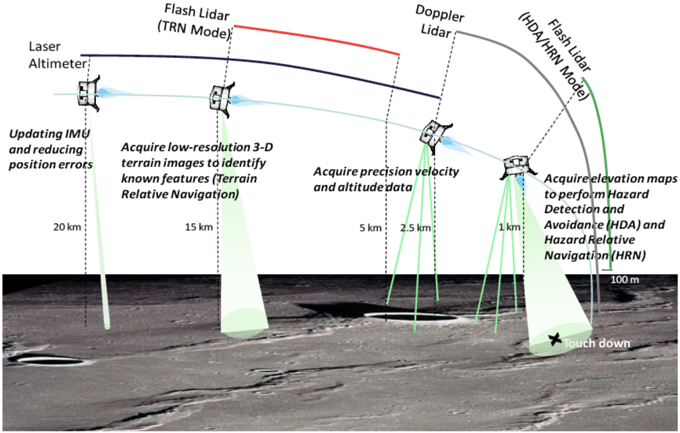
TigerEye 3D Video Camera

- 128 x 128 pixels APD (avalanche photo diode); 30Hz
- 1570 nm eye-safe laser
- 3^0 field of view (actual full FOV $= 3^0$ x 3^0); Range up to 1100 meters
- 9^0 field of view (actual full FOV $= 8.6^0$ x 8.6^0); Range up to 450 meters
- 45^0 x 22^0 field of view; Range up to 150 meters
- 45^0 field of view; Range up to 60 meters

DragonEye 3D Flash LIDAR Space Camera

- 128 $\times$ 128 pixels; 10FPS; Range and intensity video Camera
- 45^0 x 45^0 field of view (17mm)
- Range up to 1.5km inclusive (greater depending on diffuser/lens choice)
- Tested and used by NASA

Use of Lidar Technology for Planetary Exploration



Landing on Moon and Mars

- NASA's Autonomous Landing and Hazard Avoidance (ALHAT)
- Lidar sensors: 3-D Imaging Flash Lidar, Doppler Lidar, and Laser Altimeter
- Five sensor functions: Altimetry, Velocimetry, Terrain Relative Navigation (TRN), Hazard Detection and Avoidance (HDA) and Hazard Relative Navigation (HRN)
- The Inertial Measurement Units (IMU) suffer from drastic drift over the travel time from the Earth. The IMU drift error can be over 1 km for a Moon-bound vehicle and over 10 km for Mars

Conclusions and Discussion

- Depth sensors are based on structured light combined with CCD/CMOS technology.
- Depth is measured either by triangulation or by travel time.
- Projected-light sensors (Kinect) are more accurate in the range 1-4 meters than time-of-flight sensors.
- Several modulated-light time-of-flight cameras can work together and/or in combination with other color cameras – this opens the door to a wide range of devices that combine (low-resolution) depth cameras with (high-resolution) color cameras.
- Pulsed-light time-of-flight (3D flash Lidar) cameras can work outdoor, on Moon, Mars, etc. They are expensive!
- At INRIA in Montbonnot we investigate mixed camera systems (contact me for a short visit).

A Course on 3D Sensors (3DS)

http://perception.inrialpes.fr/people/Horaud/Courses/3DS_2013.html