



# Master in Computer Vision *Barcelona*

Module M4: 3D Vision

Lecture 4.9b: **3D Sensors and 3D data**

Lecturer: Josep R. Casas

# Outlook

## Session 9b: 3D Sensors and 3D data

- Motivation & Principles
  - Image vs Range sensingList (29)
- 3D Sensors
- 3D data

## Session 10: 3D processing & applications

- Pointcloud processing:
  - PCL (2011) & Open3D (2018)
  - Organized/Unorganized processing
  - Applications

# Introduction to 3D Sensors, Data and Processing

3D vision has been introduced from the concepts of:

- **Projective transformations**  
perspective projection, *projectivities* (3D, plane-to-plane,  
 $n$ -dimensional spaces, homography, invariants...)
- **Multiple-View geometry**  
reconstruct real world scenes from several images (projections)  
or from a moving camera (SfM)

Ok. That's the natural way into the field of 3D vision as, for visual perception, humans come equipped with:

- Two ‘**projective**’ sensors
- Multi-view (**stereoscopic**) vision
- And ‘Free Viewpoint View’ (FVV or rather “Capture”)



# Motivation and fundamental questions

*Motto:*

*Computer Vision = “Teaching computers to see”*

***Antonio Torralba (MIT)***

Talk@CVC 20th Anniversary, Barcelona July 9th, 2015



*Exciting time for CV: new architectures, DBs, productivization, future  
Fundamental problems: reconstruct 3D world, recognize...*

*Questions around this:*

*Q1: Is “projective vision” a natural way to capture the 3D world?*

*Q2: Do we need photometry to get geometry?*

*Q3: Does 3D vision and 3D geometry mean the same?*

*Q4: Does 2D/3D matter for “Teaching computers to see”?*

# What if...?



Tar stain?

## Same problem in other contexts:



*Person seen?*

*Q1: Is “**projective vision**” a natural way to capture the 3D world?  
Projective vision ambiguities...*

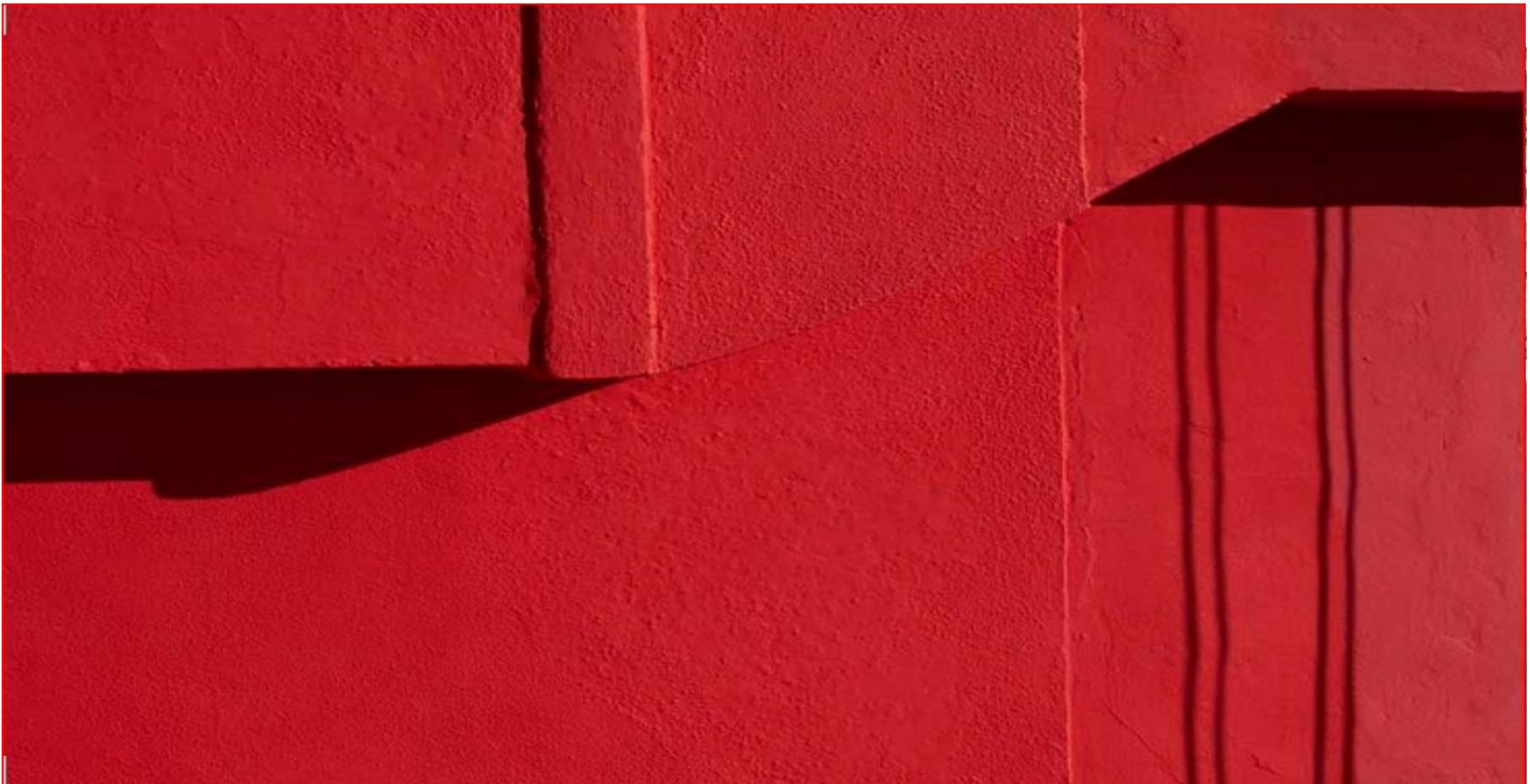


(c) 2018 jessicabackhaus.net



(c) 2011 Santiago Bañón

*Q1: Is “**projective vision**” a natural way to capture the 3D world?  
...used smartly  
[not keen in bullfighting... at all!]*



*A1: Is “**projective vision**” a natural way to capture the 3D world?*

## *Q2: Do we need to measure photometry to get geometry?*

### 2D vs 3D

- 2D vs 3D

“2D imaging

...projects 3D scenes onto a planar surface (retina, sensor)...  
...so that the depth (Z) dimension is lost”

- *Can we backproject systematically?*
- *Are there imaging sensors not projecting onto a plane?*
- *How can depth (range) be perceived directly?*
- *What would be the main advantage of capturing 3D directly?*

## Q2: Do we need to measure photometry to get geometry?

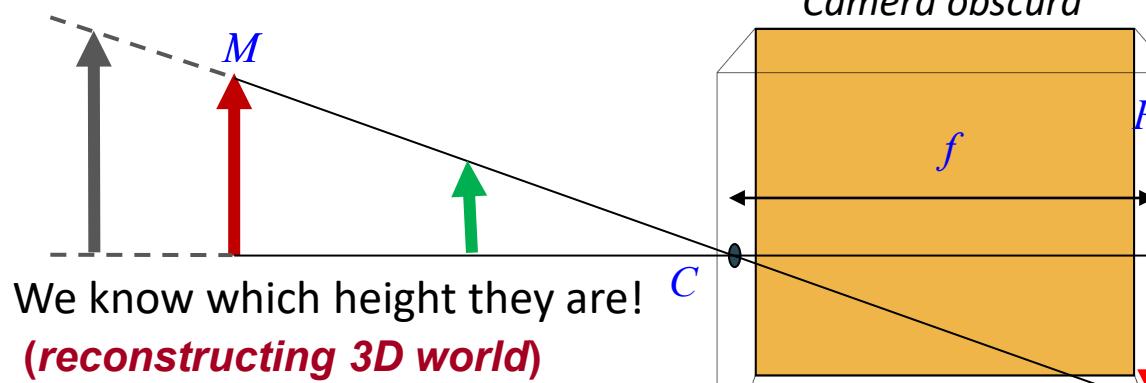
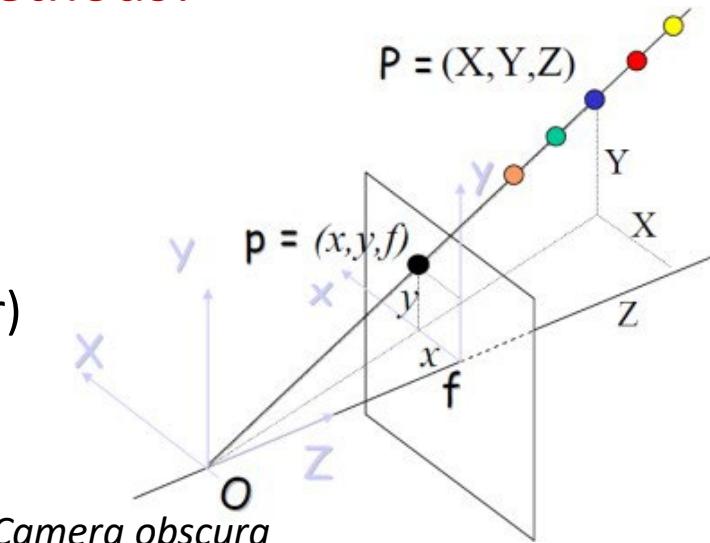


### 2D vs 3D

Yes! with distance measurement methods!

Range sensors (scanners!)

- 3D scanners and Lidar systems
  - Light coding systems (e.g. Kinect™)
  - TOF cameras (e.g. Kinect™ v2, Lidar)
- ... as if we could **backproject!**



We know which height they are!  
*(reconstructing 3D world)*

A2: Do we need to measure photometry to get geometry?

### *Q3: Does 3D vision mean the same than 3D geometry?*

Example of 3D capture



... and then...

*Will this result in a complete description of the scene?*

*...like a CAD design?*

*Will we have a whole view like Panoramic / Surround view?*

*...and Free viewpoint?*

**A3: Does 3D vision mean the same than 3D geometry?**

*Q4: Does 2D/3D matter for “Teaching computers to see”?*

What can we do with 3D vision?

Example I of 3D processing (AR)



AR magic mirror using Kinect (cf. T. Blum, N. Navab, TUM)

## *Q4: Does 2D/3D matter for “Teaching computers to see”?*

Example II of 3D processing (AR)



*A4: Does 2D/3D matter for “Teaching computers to see”?*

## What if...

... we could always backproject to 3D our 2D images?

- Geometry of objects and scenes in the 3D world can greatly help analysis and recognition
- 3D geometry has the potential to avoid the large variability in appearance and missing information due to projection, occlusions, motion, shadows...
- Yes, 3D geometry may definitely **help computers to see** and "understand" complex visual scenes, reasoning about events that evolve in space and time...



# 3D/range/scan sensors

vs

# Multiple View

## Advantages 3D/range/scan

- Direct measure of 3D geometry
- Independent of photometry (active)
  - does not hold for reflectivity
- From scan to CAD blueprints (illusion!)

## Disadvantages

- 2,5D!
  - Surface vs 3D (interface air-matter vs volumetric 3D scanner, i.e. PET)
- Single viewpoint
  - Neither panoramic, nor surround view  
→ still need MultiView!!!
- Accuracy?
  - vs MPix cams!

# Symmetric paradigm of 3D Data

Double nature of 3D data

- **Geometric information**

Pure data, measures.

Numerical representation of objects

→ calculation, simulation, measurements

... + **reflectivity**

Data

3D Rendering

Visual

- **Photometric information (images)**

Can be rendered, displayed and presented visually,  
by exploiting our perception capabilities

→ exploration, analysis and understanding  
(same than looking at the physical world)

3D data

3D capture

Images

# Reconstruct the world with range sensors?

## Illusion!

- Occlusions  
→ incomplete view
- Multiple viewpoints  
→ need multiple range sensors (or moving around on static scenes)
- Only air-opaque interfaces (not the interior of objects)  
→ Range is not X-Ray ....fortunately! ☺

...but better than stereo/multi-view!

- *Back-projection* with a single sensor (“externalized” processing)
- Independent on scene illumination and surface texturing

# Photogrammetry

**Photogrammetry** refers to the practice of deriving 3D measurements from photographs.

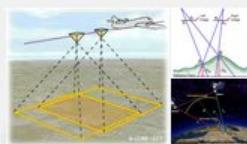
- Advances in digital cameras, processors, and computation, yield **extremely dense and accurate 3D surface data** from a limited number of photos with standard digital photography
- Structure from Motion (SfM) and sub-pixel image matching yield **3D-dimensional structure by analyzing projected 2D motion fields** created by a sequential change of position of the camera sensor relative to the object. Photographic sequences are captured to maximize information available from the change in viewpoint
- Resulting **data sets are software platform-independent** and can be reused

*Disadvantage: **computation time...***

# Photogrammetry examples



Aerial



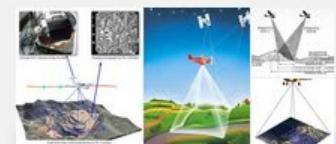
Stereoscopic



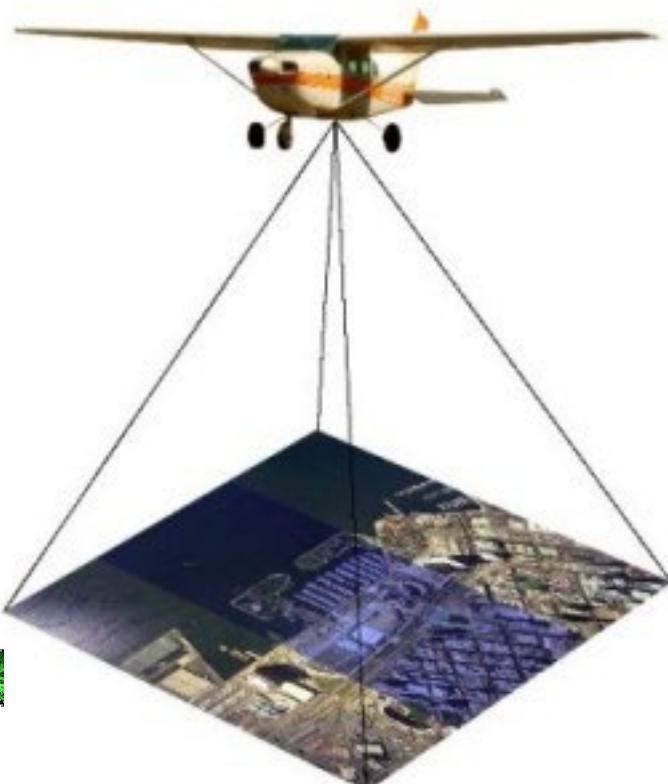
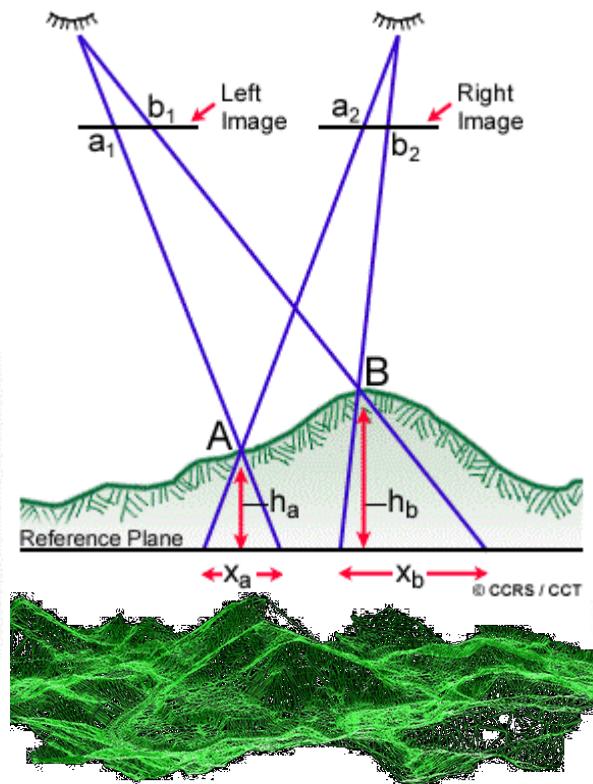
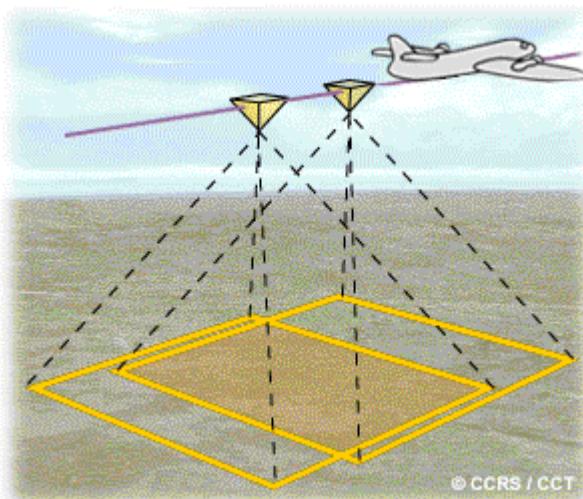
Architecture



Close Range



Surveying



# Recall & Outlook

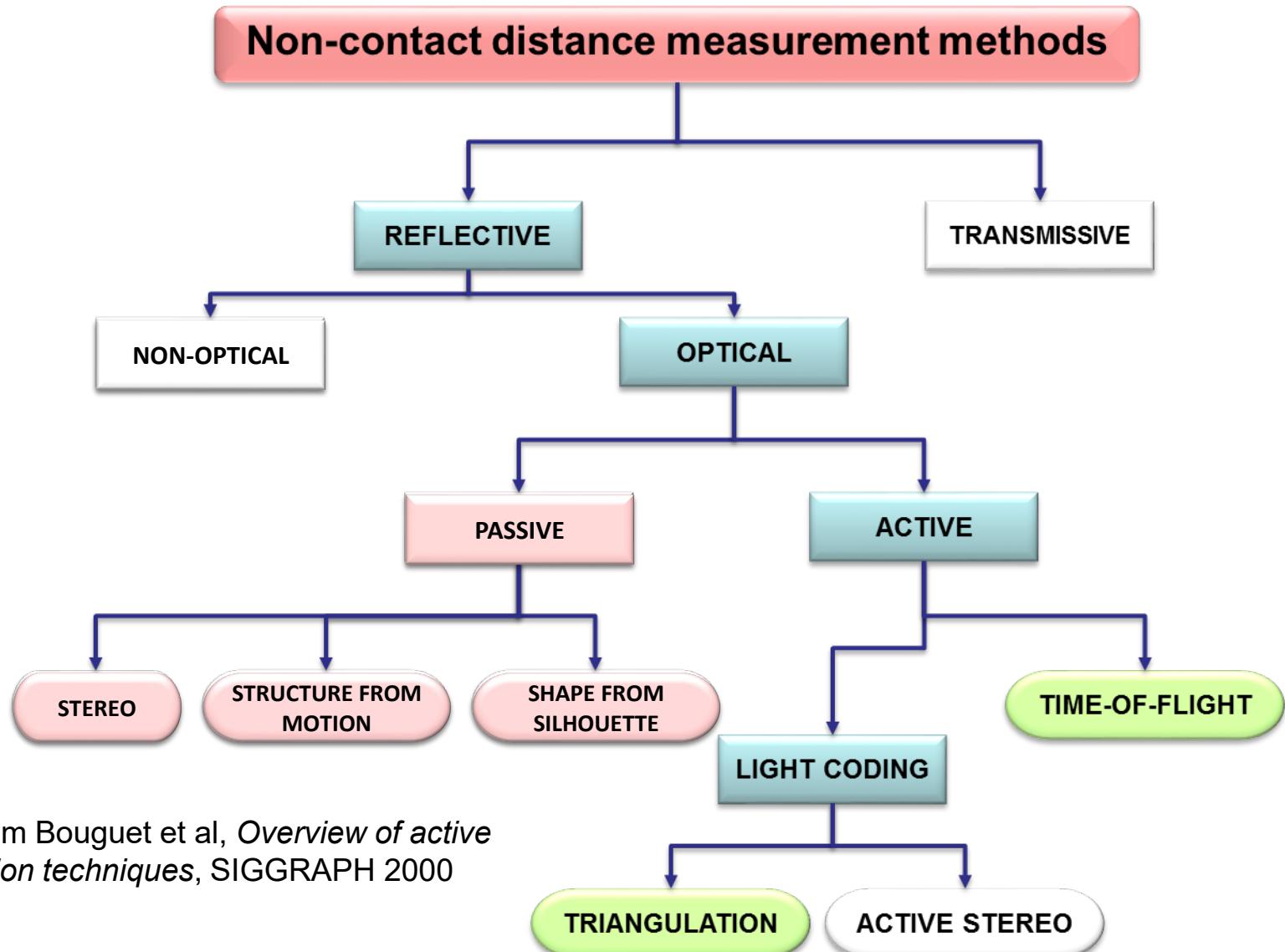
## 3D Sensors and 3D data

- Intro: images vs range data
- Sensors: ontology & examples
  - Structured Light (e.g. Kinect 1, Xtion, Orbbec)
  - TOF (e.g. Kinect 2, Lidar)
    - Kinect 1 SL strategy
    - Lidar examples
- 3D data: pointclouds
  - Projectability
- 3D data processing

# 3D sensors

Depth, range, scan...

# “Non-contact” – distance measurement methods



From Bouguet et al, *Overview of active vision techniques*, SIGGRAPH 2000

# Depth or Range sensing Methods

Reflective optical methods:

- Passive range sensing

3D distance measurement by the way of **radiation already present in the scene**. Not necessarily in the visible spectrum

- E.g. stereo, multi-camera (triangulation, SfS...)

*Off-the-shelf light field cams have demonstrated improved depth estimation using multiview stereo configs [Bishop 2012]*

- Active range sensing

3D distance measurement obtained by projecting in the scene **some form of radiation**

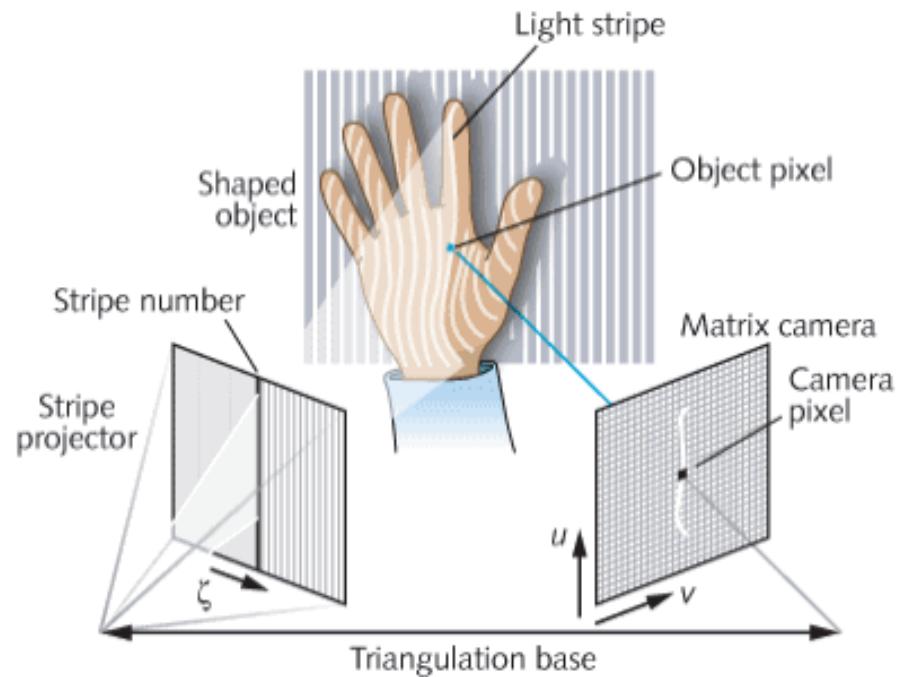
- ToF (IR/laser) → high frame rate, low res, noise [Foix 2011]
  - Lidar (UV/visible/IR) → sequential (low frame rate)
  - Structured Light (IR/visible) → artifacts, noise, indoor [Salvi 2010]



# Structured Light methods

# Structured light methods

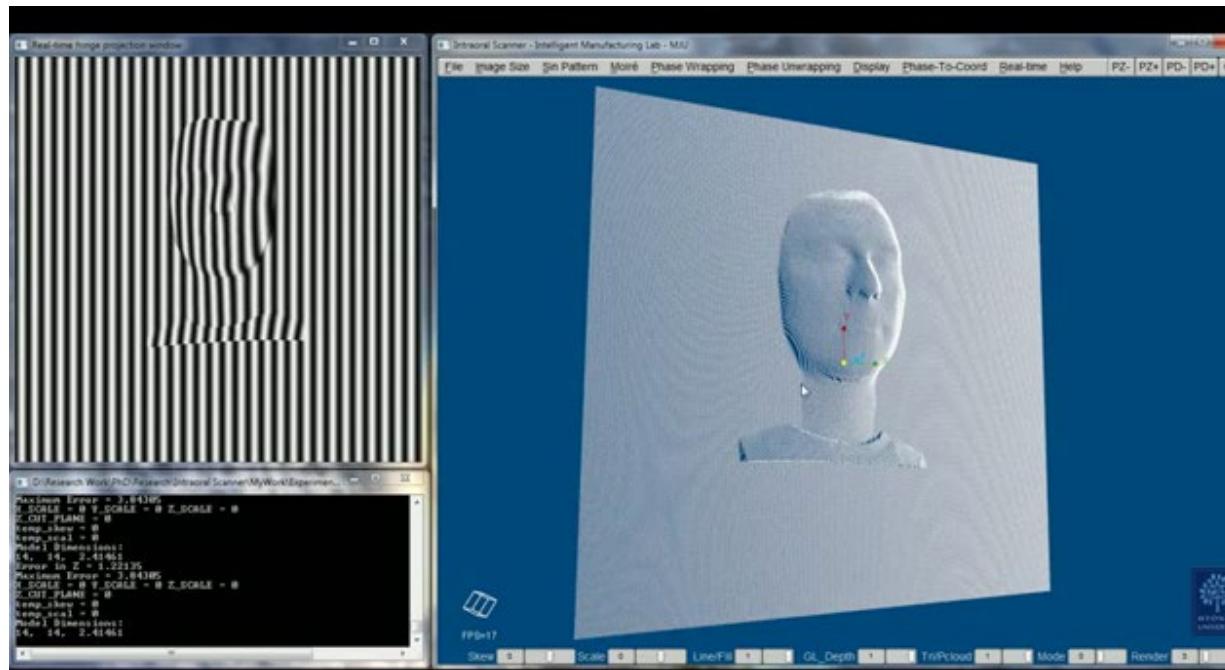
- Calculate the shape by how the strip is distorted.



<http://www.laserfocusworld.com/articles/2011/01/lasers-bring-gesture-recognition-to-the-home.html>

# Real time Virtual 3D Scanner - Structured Light Technology

- Demo



[http://www.youtube.com/watch?v=a6pgzNUjh\\_s](http://www.youtube.com/watch?v=a6pgzNUjh_s)

→ “Computational depth sensing”

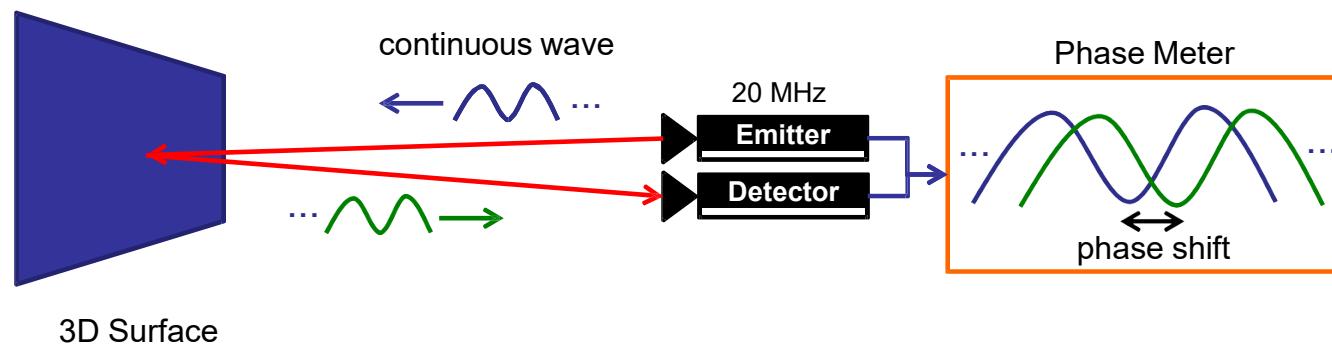
# TOF Imaging



# Principles of ToF sensors

## Continuous Wave Modulation

- Continuous light waves instead of short light pulses
- Modulation in terms of frequency of sinusoidal waves
- Detected wave after reflection has shifted phase
- Phase shift proportional to distance from reflecting surface



# TOF *Imaging*



Swissranger™ SR4000  
Mesa Imaging



[Kinect for Xbox One](#) by  
Microsoft

## Matricial Time-Of-Flight Cameras

- Active sensors
- Acquire 3D geometry at video rate (up to 50fps)
- E.g.      MESA Imaging  
                  PMD Technologies  
                  Optrima SoftKinectic  
                  CANESTA (*acquired by Microsoft in 2010*)

...



D-IMager  
Panasonic



pmd[vision] CamCube  
PMD Technologies



FOTONIC-B70  
Fotonic



3D MLI Sensor  
IEE S.A.

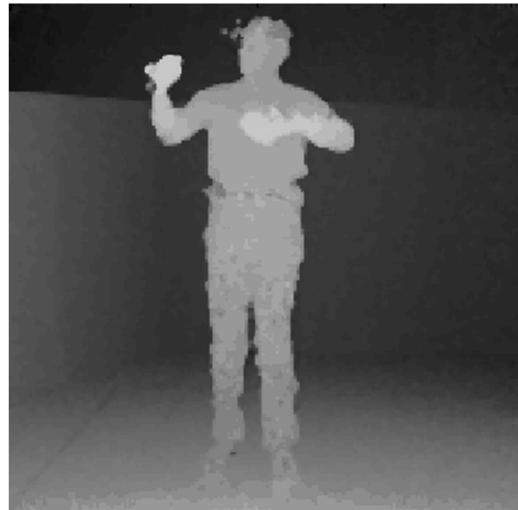


pmd[vision] CamBoard  
PMD Technologies

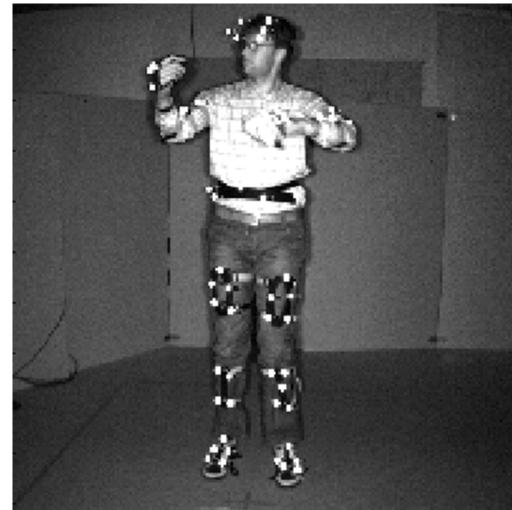
# Principles of ToF Imaging

## Continuous Wave Modulation

- Simultaneous availability of (co-registered) range and amplitude images



Depth Image



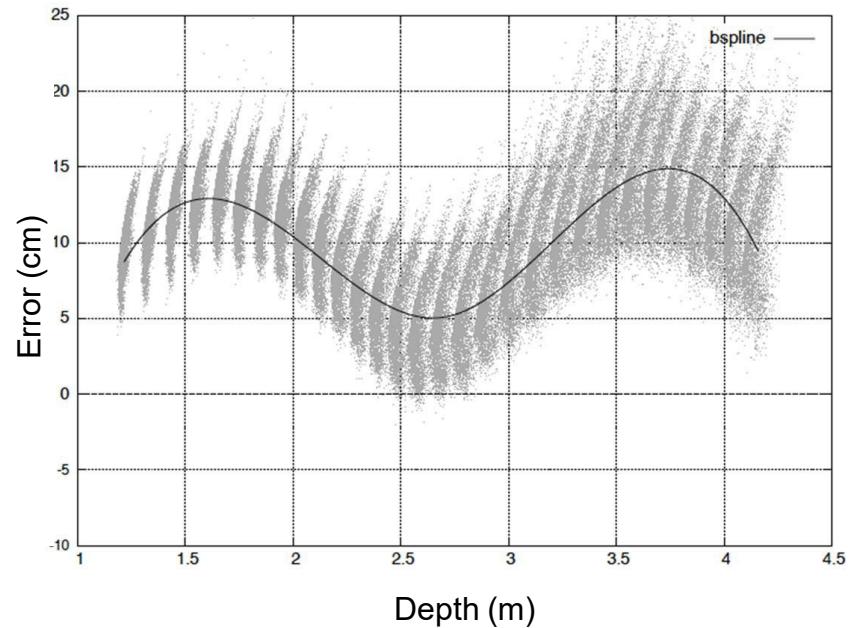
Amplitude Image

# Computer Vision with ToF Cameras

## Measurement Errors and Noise

### Systematic distance error

- Perfect sinusoidal signals hard to achieve in practice
- Depth reconstructed from imperfect signals is erroneous
- Solution 1: camera-specific calibration to know distance error
- Solution 2: alternative demodulation techniques not assuming perfect sinusoidal signals

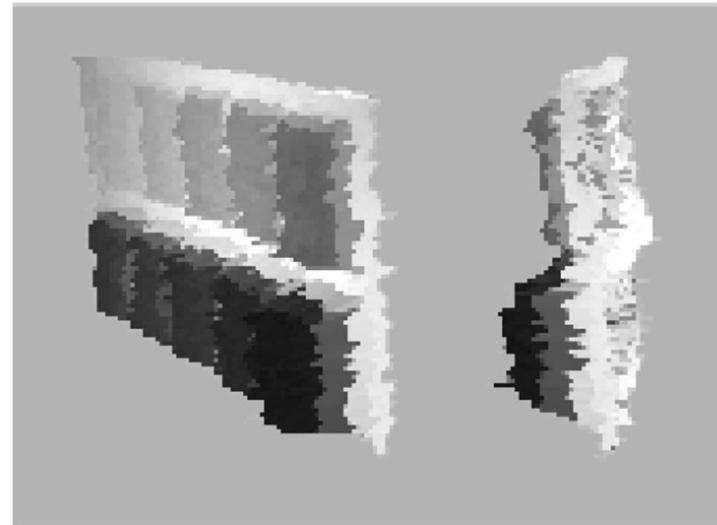


# Computer Vision with ToF Cameras

## Measurement Errors and Noise

### Intensity-related distance error

- Computed distance depending on amount of incident light
- Inconsistencies at surfaces with low infrared-light reflectivity
- Correction by means of corresponding amplitude image



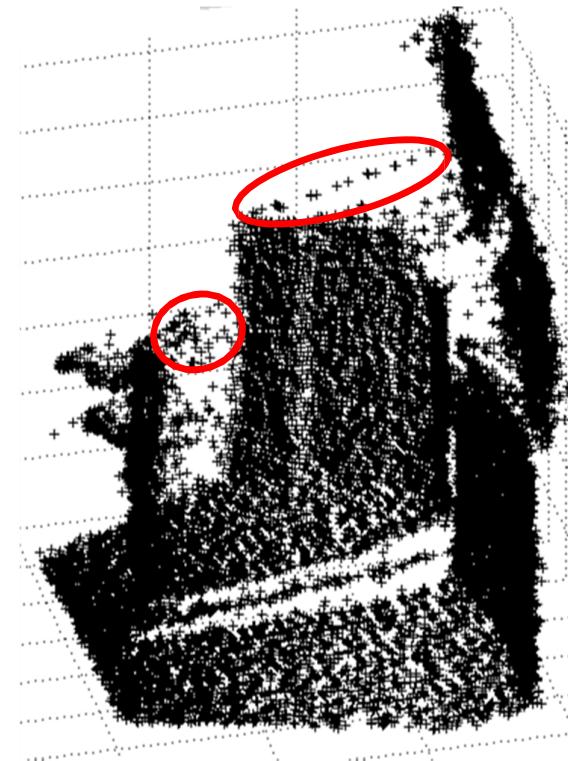
Depth images of planar object  
with patches of different reflectivity

# Computer Vision with ToF Cameras

## Measurement Errors and Noise

### Depth inhomogeneity

- Current ToF cameras have low pixel resolution
- Individual pixels get different depth measurements
- Inhomogeneous
- "**Flying pixels**", especially at object boundaries
- Correction: discard pixels along rays parallel to viewing direction



# Kinect



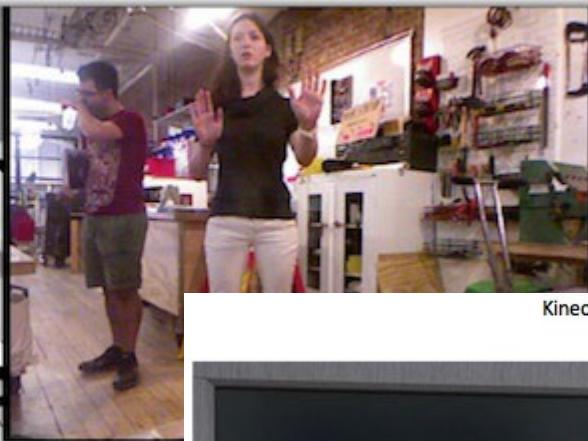
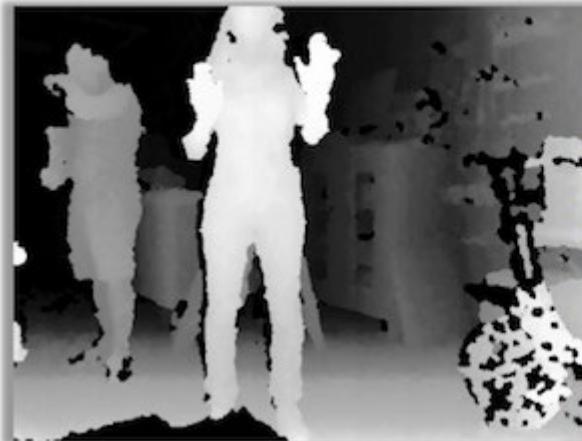
- RGB, Depth Sensor, and Multi-mic Array
- Works with structured light (dots rather than strips) to determine distance for each pixel

# History

- Original technology developed in 2005
- Announced in 2009 as codenamed Natal after the city and because of its relation to being “of or related to birth”
- Released Kinect for Windows Beta on June 16, 2011
- On February 1<sup>st</sup>, released commercial version
- Kinect for Xbox ONE (v2, 2013)
  - Time-of-flight sensor replacing the old technology from PrimeSense with CANESTA technology

# Kinect

Kinect For Windows 1



Kinect For Windows 2



Image via <http://blogs.msdn.com>

# Some Commercial Depth Sensors: Specs



	<u>MS Kinect</u>	<u>MS Kinect2</u>	<u>Asus Xtion Pro Live</u>	<u>Occipital Structure</u>
Color	640x480 (2x!) @30fps	1920x1080 @30fps	1280x1024 @30fps	(iPad cam)
Depth	320x240 16bit	512x424	640x480 (320x240@60)	640x480 (320x240@60)
Range (m)	(0.4)0.8-3,5	0.5-4,5	0.8-3,5	0.4-3,5
FoV (VxH)	43º x 57º	60º x 70º	58º x 45º	58º x 45º
Tilt motor Accelerom.	±27º (V) 2G(1º)/4G/8G	–	–	–
Audio	16bit PCM @16kHz 4 mics	16bit PCM @16kHz 4 mics	2 mics	(iPad mic)
USB	2.0	3.0	2.0	2.0 (hack)
Tracker	20 joints (2x)	26 joints (6x)	–	–

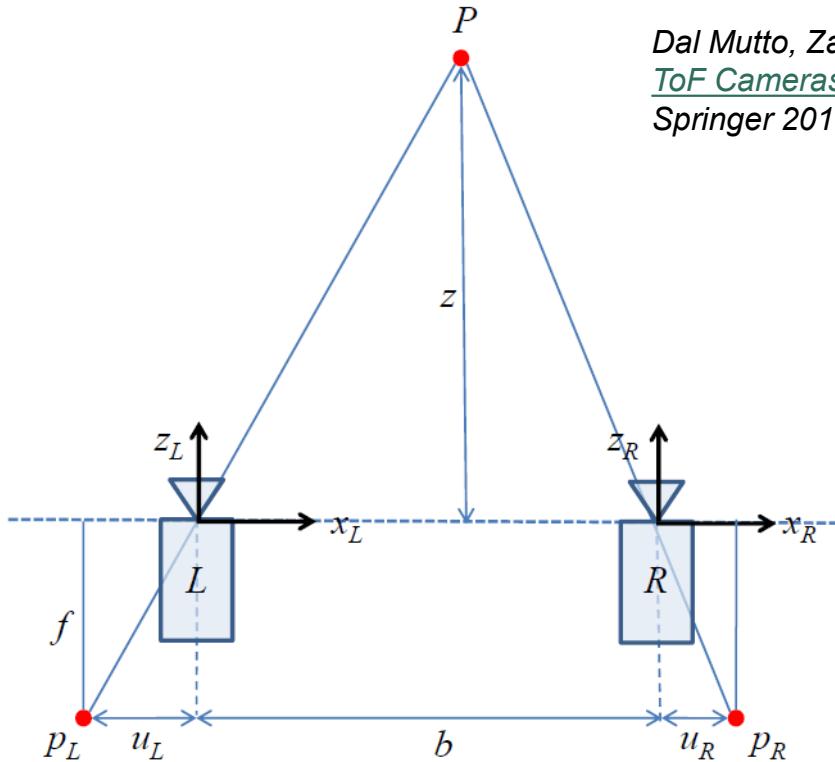
Others: e.g. Carmine from *PrimeSense*, Orbbec sensors, VicoVR or Stereolabs ZED

# Kinect: depth imaging principles

- Active triangulation
- Matricial arrangement
- Structured light coding
  - Correspondence solving: codewords
- Artifacts
- Light coding strategies
  - Kinect pattern
- Practical issues
  - Calibration, artifacts
  - Comparison to other 3D scanning methods
- Demo

# Active Triangulation

## Stereo Triangulation



Conjugate pixels:

$p_L, p_R$

Disparity:

$d = u_L - u_R$

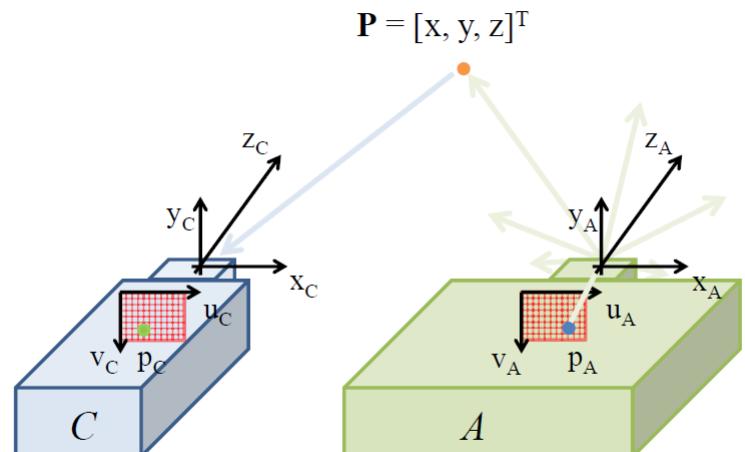
Depth:

$z = bf/d$

(calibrated and rectified)

Dal Mutto, Zanuttigh, Cortelazzo  
ToF Cameras and MS Kinect  
Springer 2012

## Active Triangulation



Same geometry:

camera ( $R$ ) replaced by projector ( $A$ )

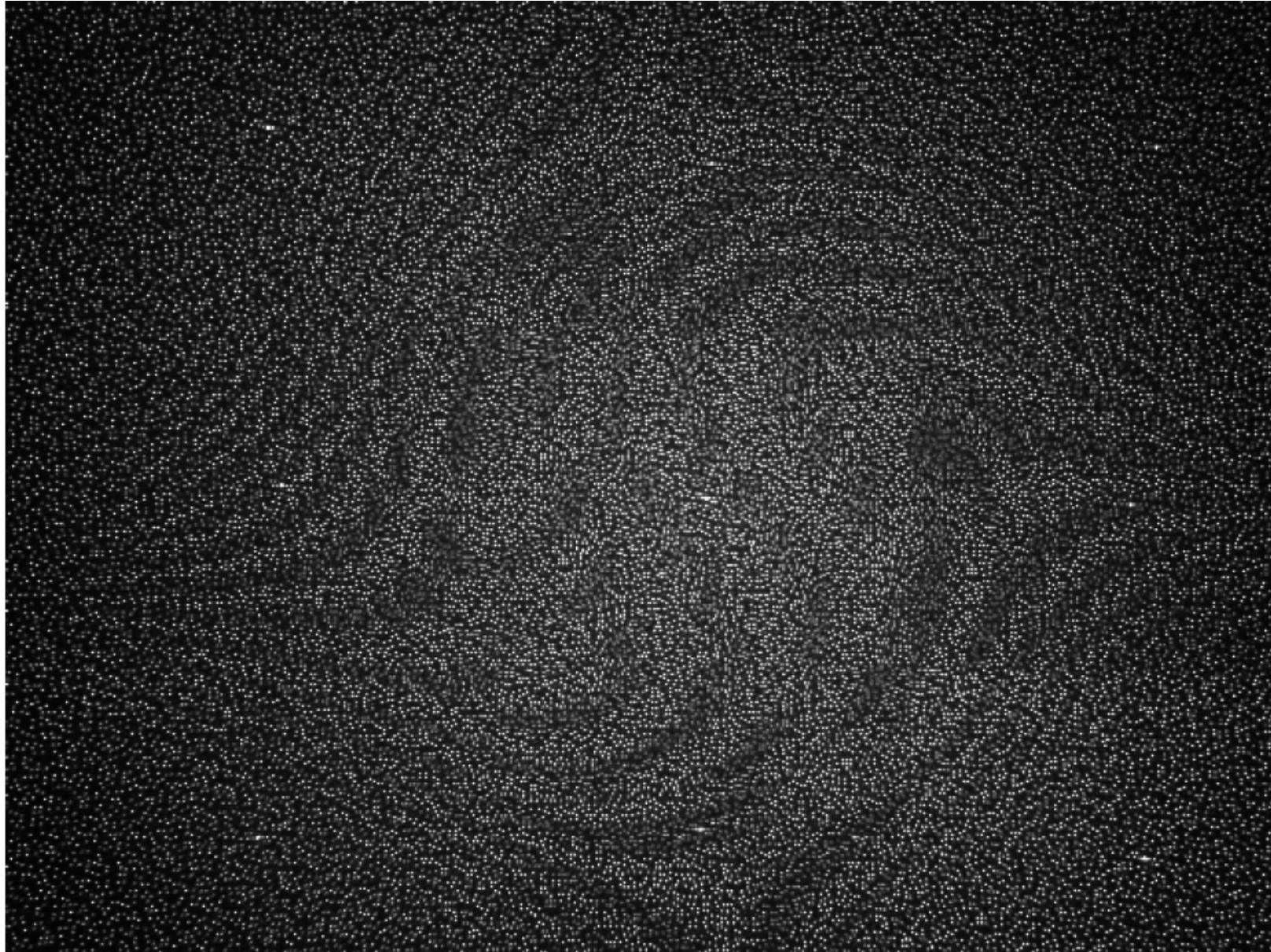
$p_A$  pattern pixel,  $p_C$  camera pixel

$p_A, p_C$  are conjugate  $\rightarrow$  disparity  $d$

Depth of scene point  $P$ :  $z = bf/d$

(calibrated and rectified)

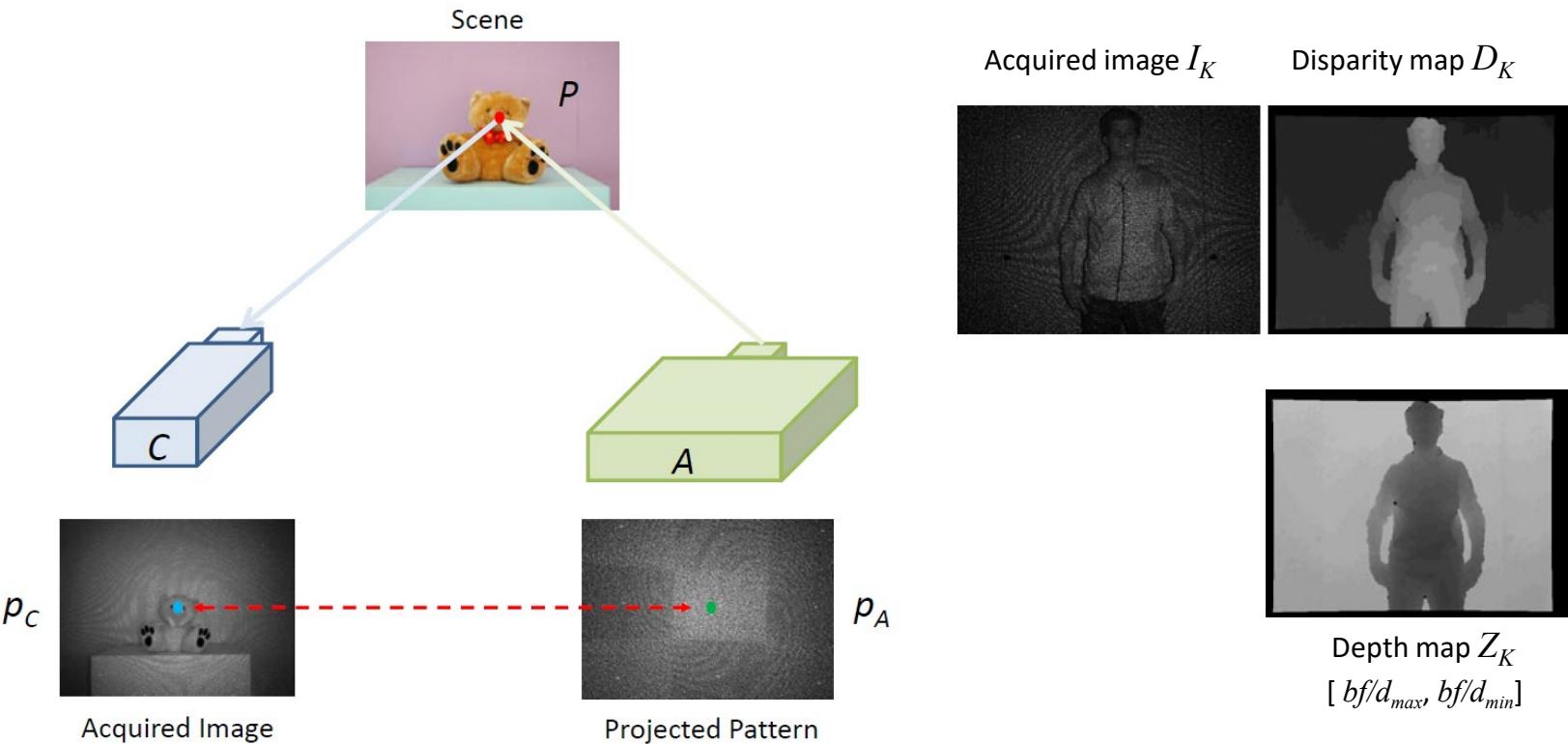
# Kinect pattern projected



(from Dal Mutto, et al 2012)

# Matricial Active Triangulation

IR Projector and camera (invisible)



(from Dal Mutto, et al 2012)

# Structured light coding



- Solves the correspondence problem
  - Each pixel of the pattern is associated to a *codeword* (specific local configuration of the projected pattern)
  - The pattern is projected ( $A$ ), reflected in scene and captured ( $R$ )
  - A *correspondence estimation* algorithm analyzes the received codewords in the acquired image ( $I_K$ ) in order to compute the conjugate of each projected pattern
  - Pattern design adopts *codewords effectively decodable* in presence of non-idealities of the projection/acquisition process

## Decodable codewords

The projection/acquisition process introduces an horizontal shift  $d$  (disparity) proportional to the inverse of the depth  $z$  of scene point  $P$  for each pixel  $p_A$  which is associated to a codeword (window centered at  $p_A$ ).

Codewords should be:

- **Different enough**: low cardinality!
- **Row-separable**

In calibrated & rectified setups conjugate points lie in horizontal lines

- The more the local pattern distribution of a single pixel differs from the local pattern distribution of other pixels of the same row, the more **robust** will be the coding
- Local distribution of the pattern for pixel  $p_A$  given by the illumination values of the pixels in a window around  $p_A$
- For windows  $n_W$  pixels, there are  $n_P^{n_W}$  possible pattern configurations  
 $n_P = 2^8$  for an 8-bit grayscale projector



# Codeword artifacts

Artifacts in the projection acquisition process:

- **Perspective** distortion, due to varying depths in scene
- **Color** (gray level) distortion, due to reflectivity properties of objects
- **External illumination**, due to other light sources in the scene (IR noise)
- **Occlusions**, not all pixels projected will be seen by the camera
- Projector and camera **non-idealities** (non-linear characteristics)
- Projector and camera **noise** (Gaussian, additive)

# Lidar

Light Detection And Ranging  
(also called LIDAR, LiDAR, and LADAR)

# Lidar

Surveying method/Measurement technique:

Pulsed laser light illuminates an object, the reflected pulses are recovered and measured with a solid-state sensor

Time elapsed provides an automated measurement of distance to target (concept also referred to as lidar or ToF imaging)

The Marching Bear Mound Group in High Definition LiDAR



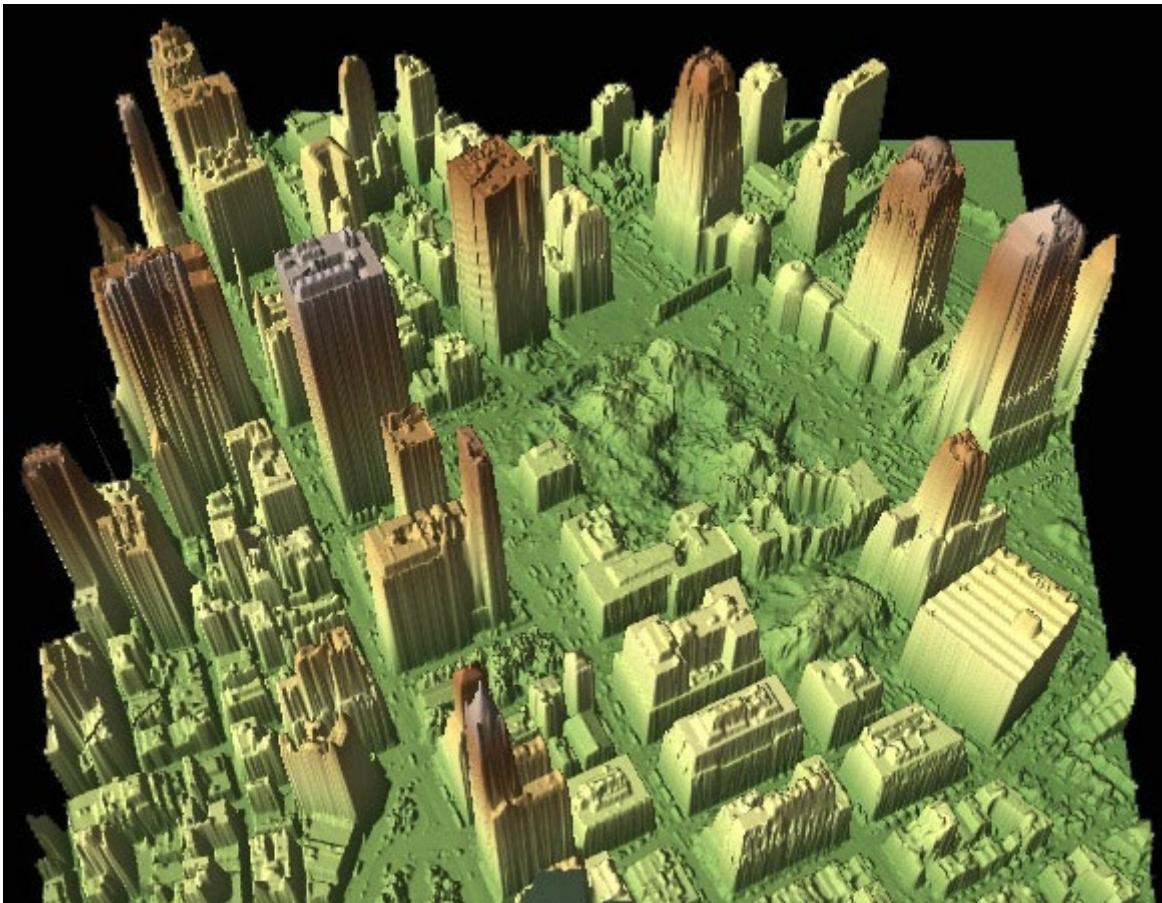
Popular applications:

Earth resources exploration, landing aids, object recognition, self-guided vehicles and safety and security applications in transport



# LIDAR light detection and ranging scanner

<http://www.youtube.com/watch?v=MuwQTc8KK44>



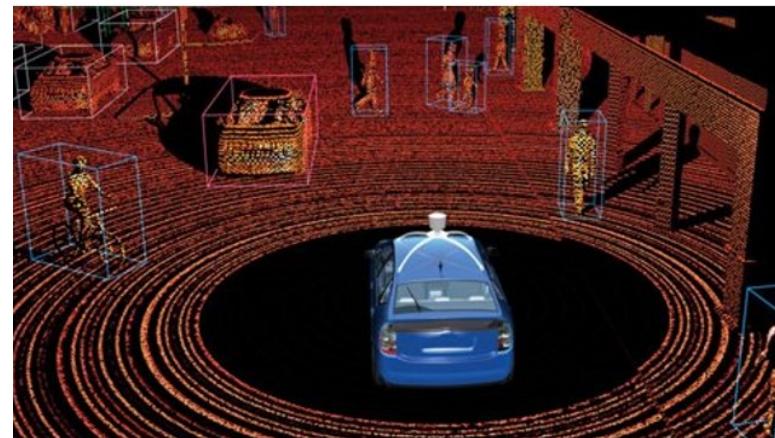
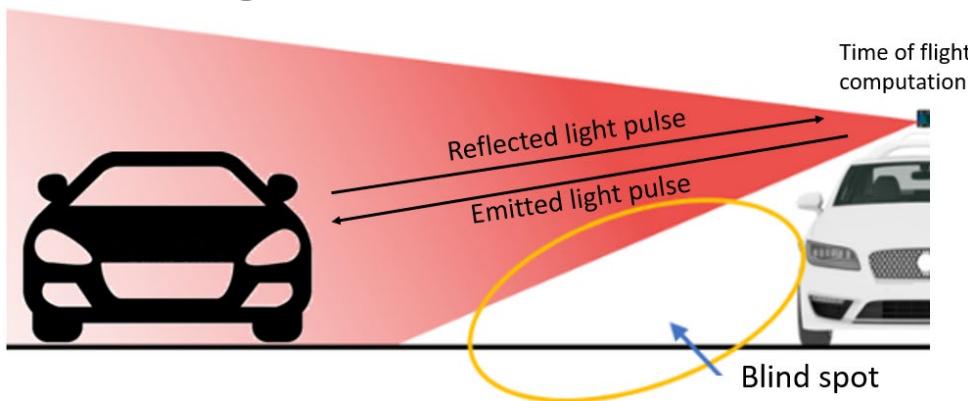
Leica  
lidar

<http://hodcivil.edublogs.org/2011/11/06/lidar-%E2%80%93-light-detection-and-ranging/>

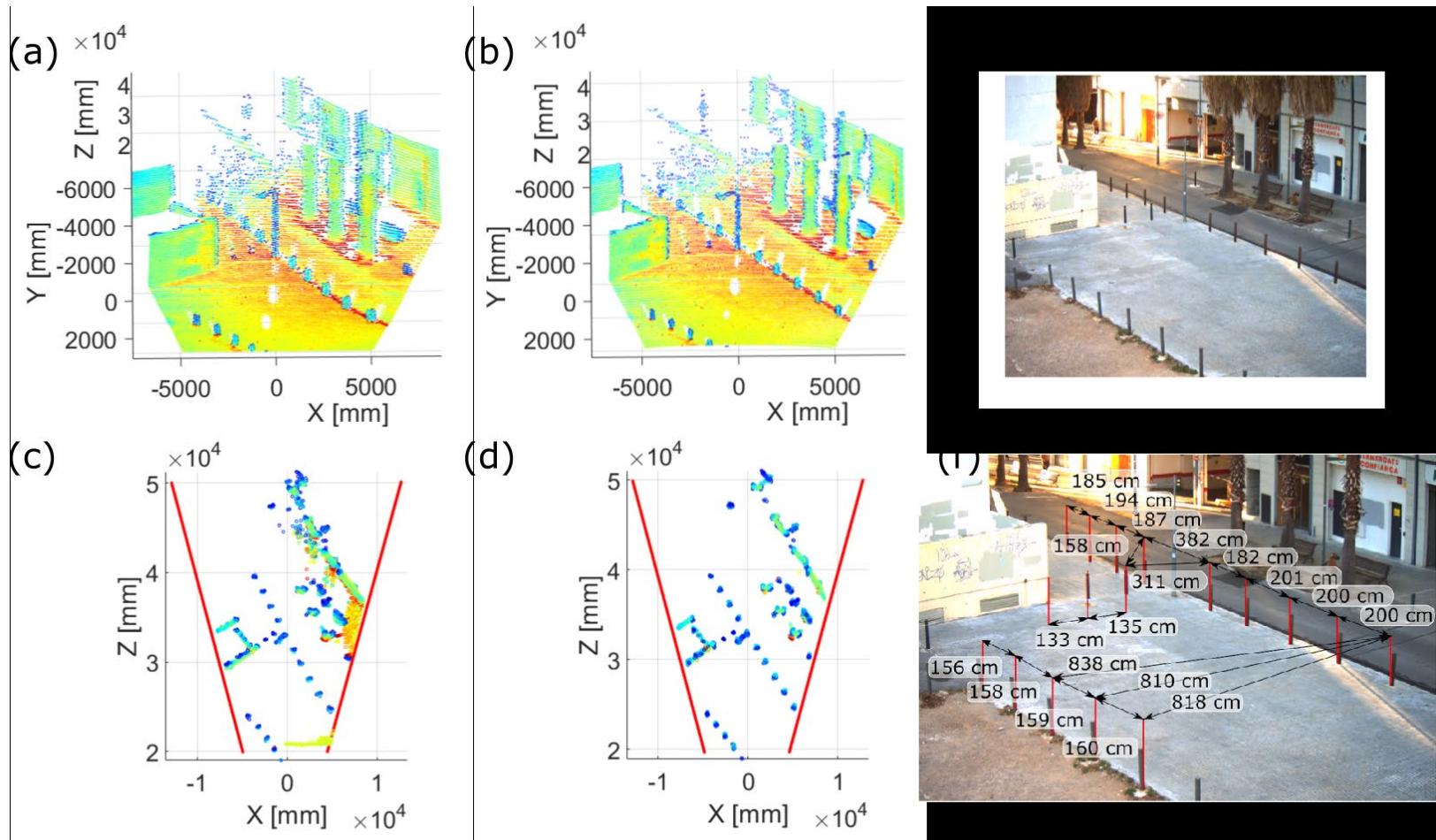
[http://commons.wikimedia.org/wiki/File:Lidar\\_P1270901.jpg](http://commons.wikimedia.org/wiki/File:Lidar_P1270901.jpg)

## LIDAR

- Uses a rotatory pulsated laser (360º field of view)
- Can estimate distances and map the surroundings with high precision
- Higher angular resolution than radar (narrow beam)
- Profile of detected objects
- Expensive technology
- Weak against adverse weather conditions

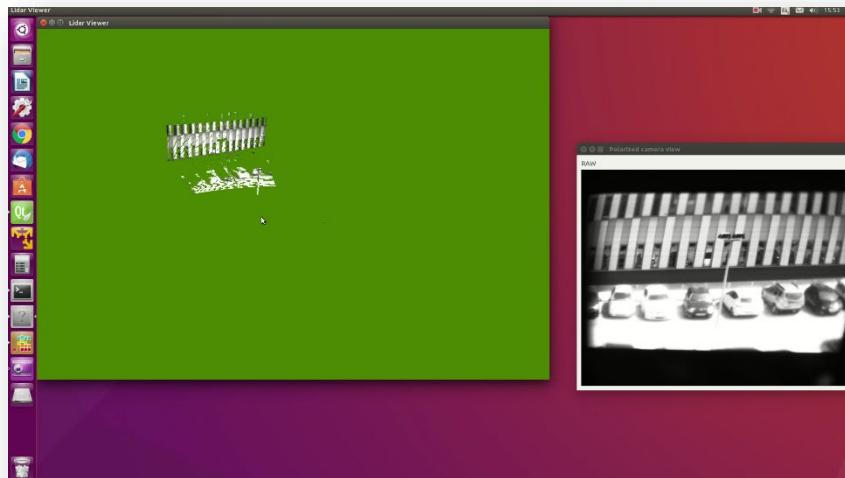


# Data fusion: RGB-Lidar

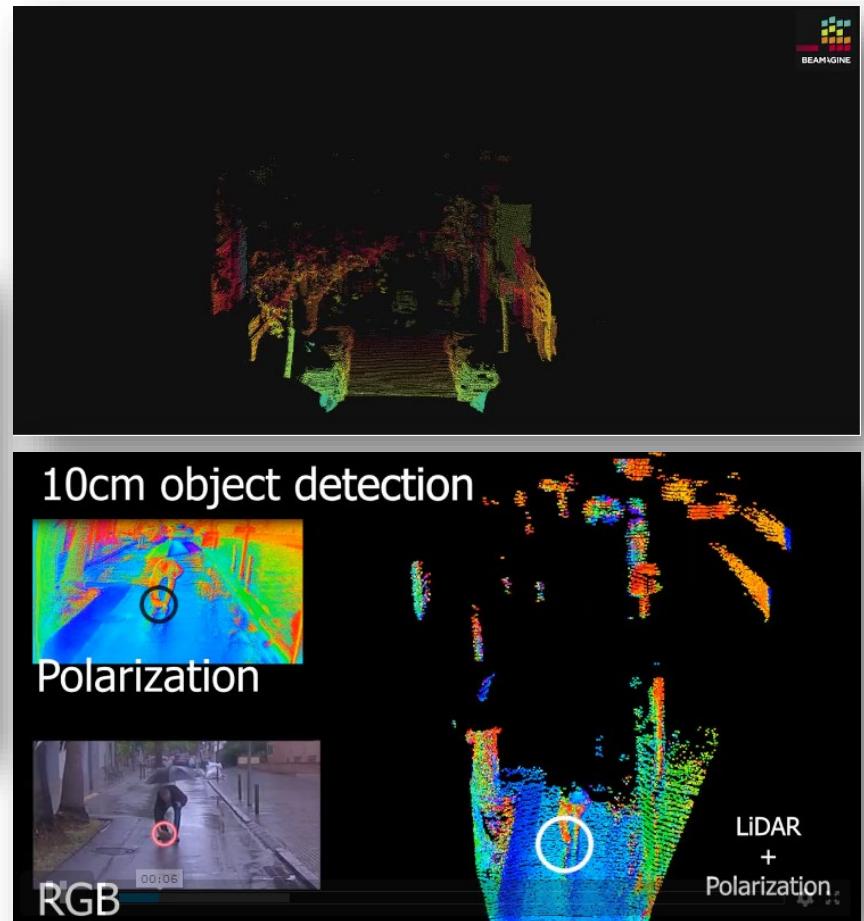


P. García-Gómez, et al, Geometric Model and Calibration Method for a Solid-State LiDAR,  
Sensors 20(10) 2020

# Examples



Dense Lidar (P. García-Gómez, et al)



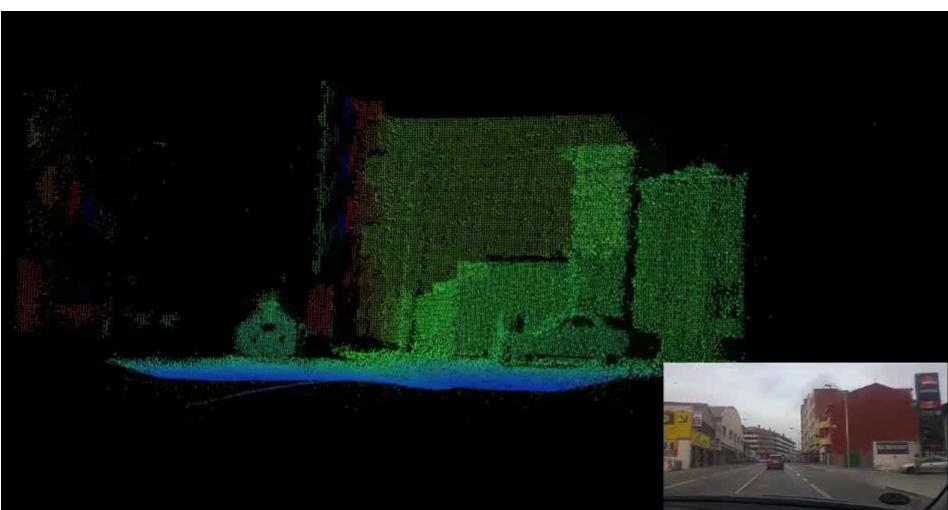
[beamagine.com](http://beamagine.com) (applications)

# Lidar Point Cloud

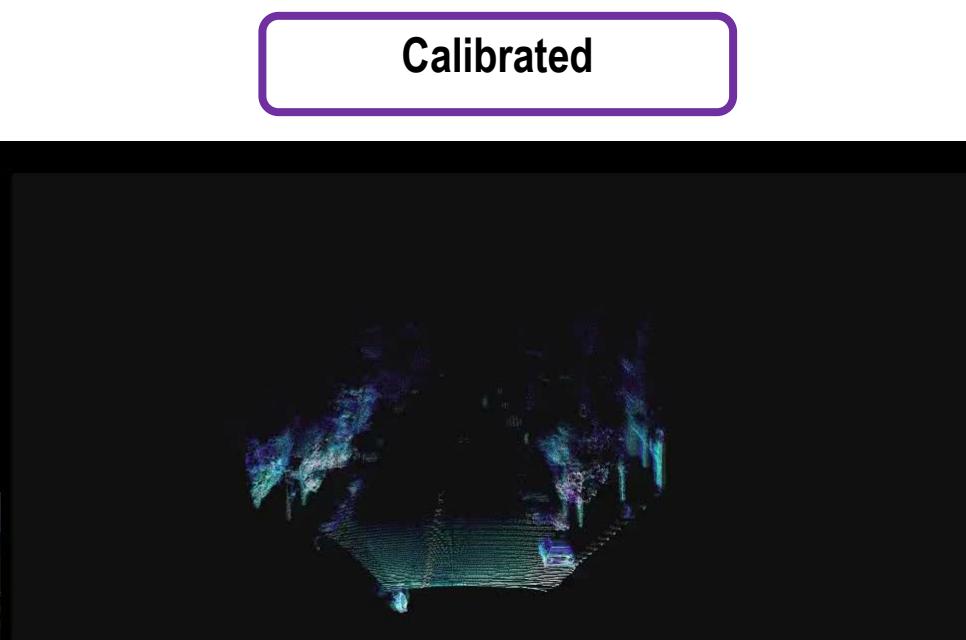


# Data fusion RGB-Lidar

Uncalibrated



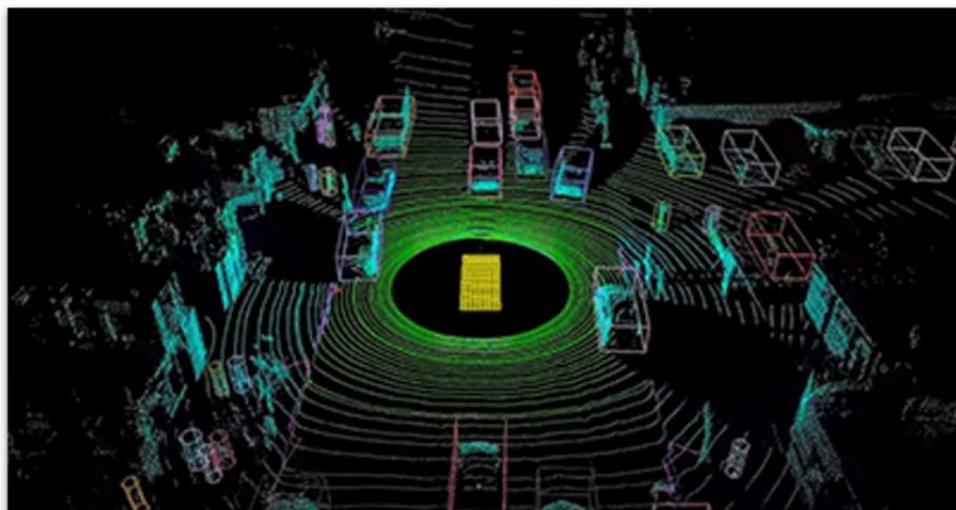
Calibrated



# Lidar 4D



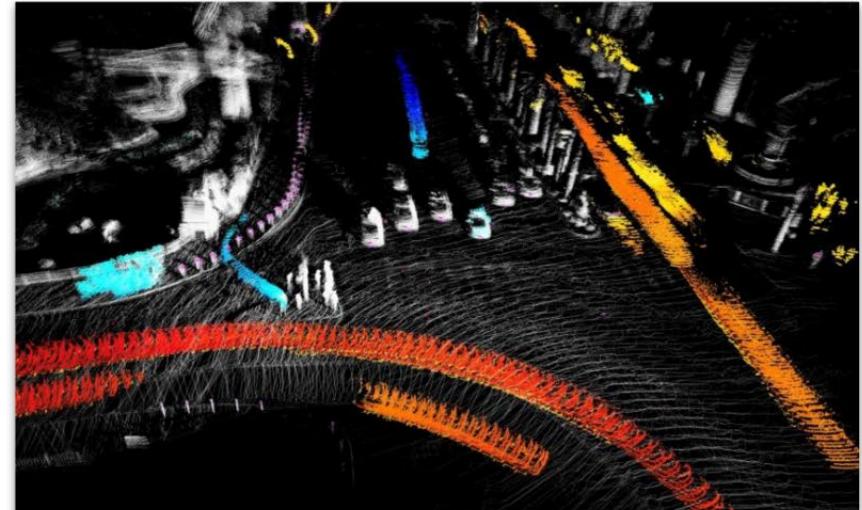
3D LiDAR



A 3D LiDAR that can estimate positions in 3D

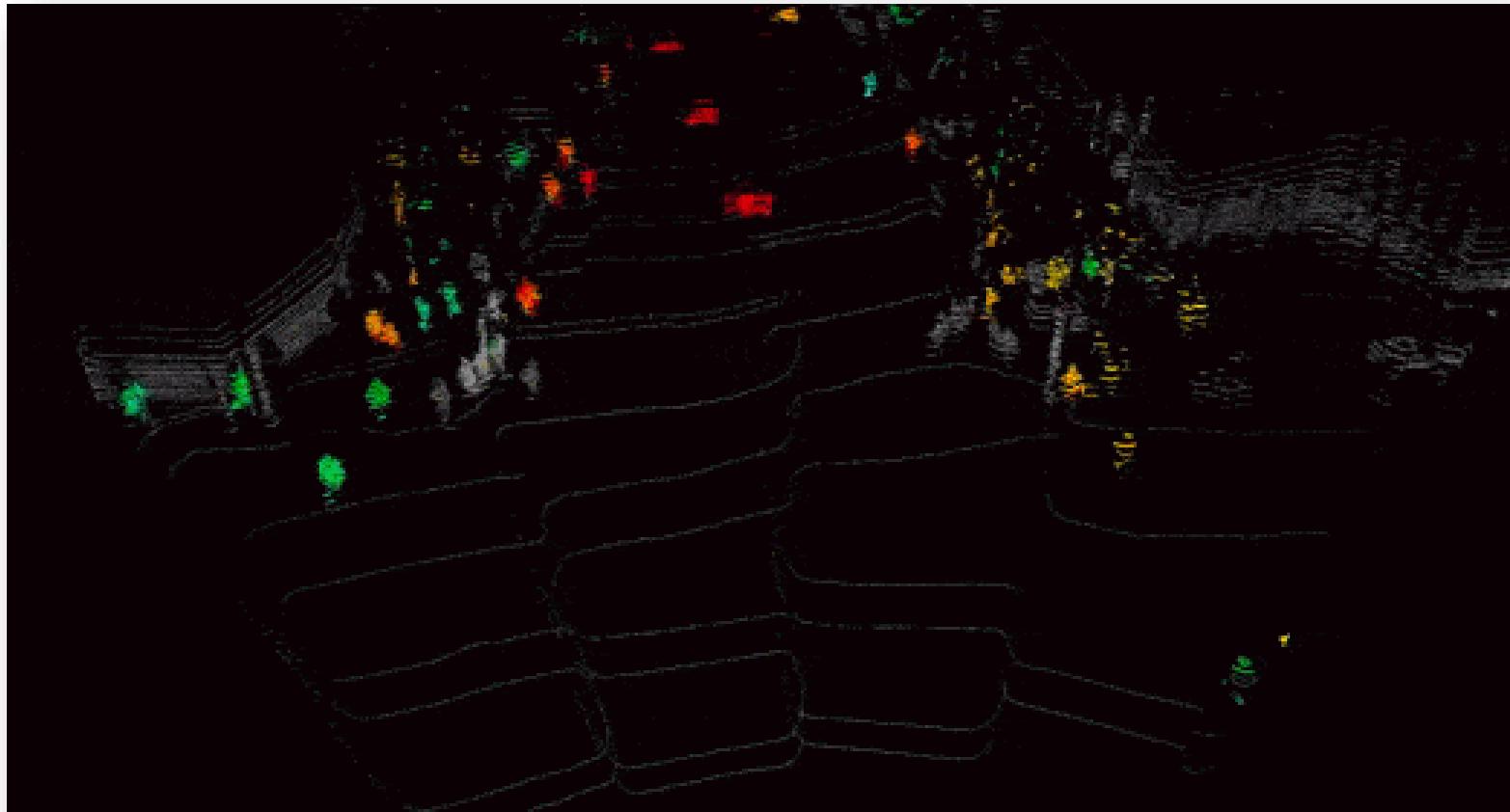


4D LiDAR

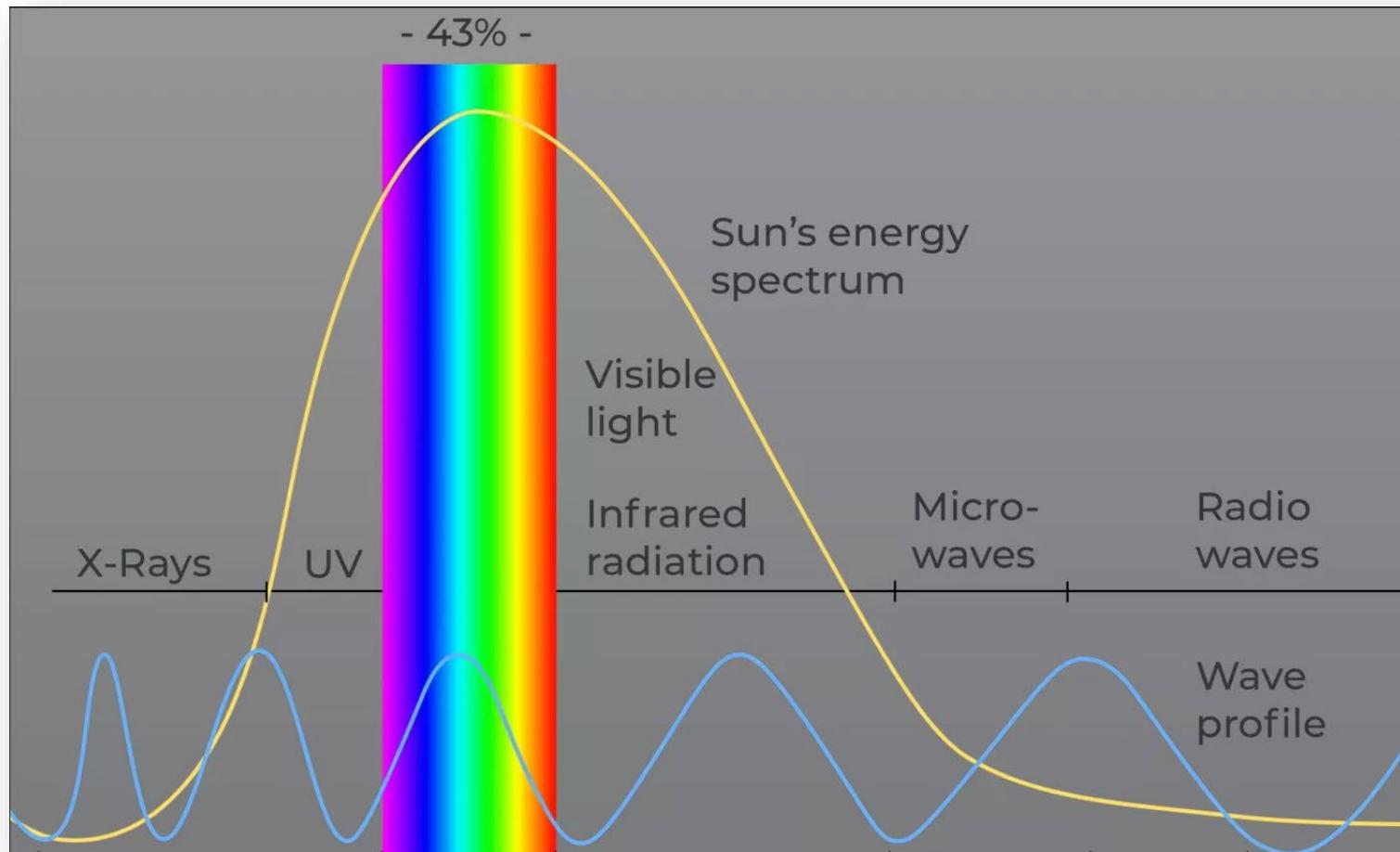


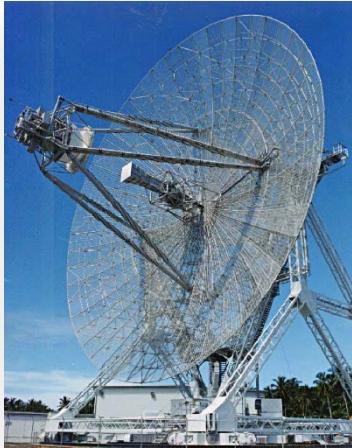
A 4D LiDAR that can estimate velocities and predict trajectories  
**blue:** approaching | **red:** receding

# Lidar 4D



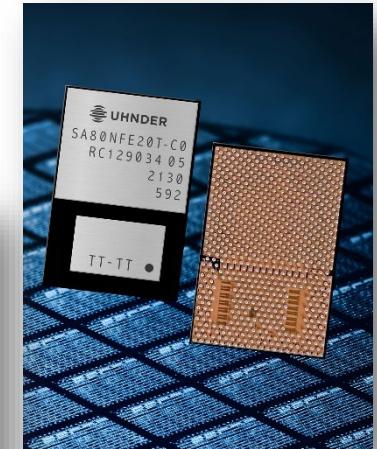
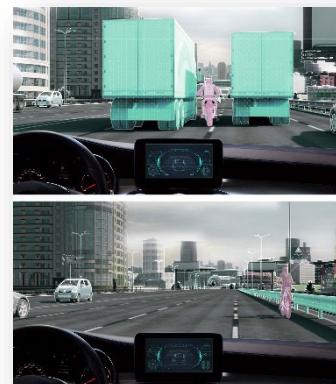
# Lidar 4D





## Radar

RAdio Detection And Ranging

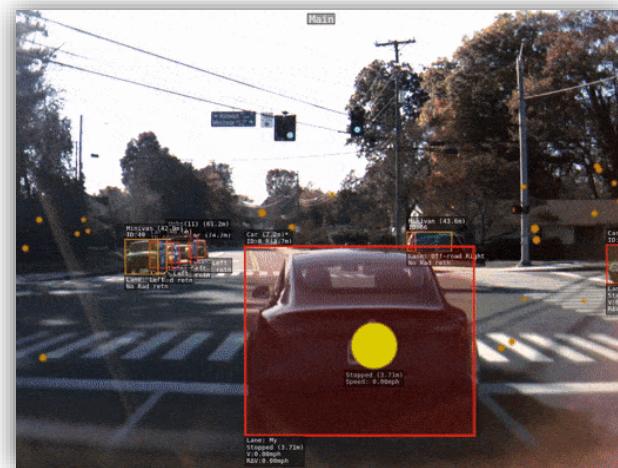


# Imaging Radar 4D

- Hype started CES2022



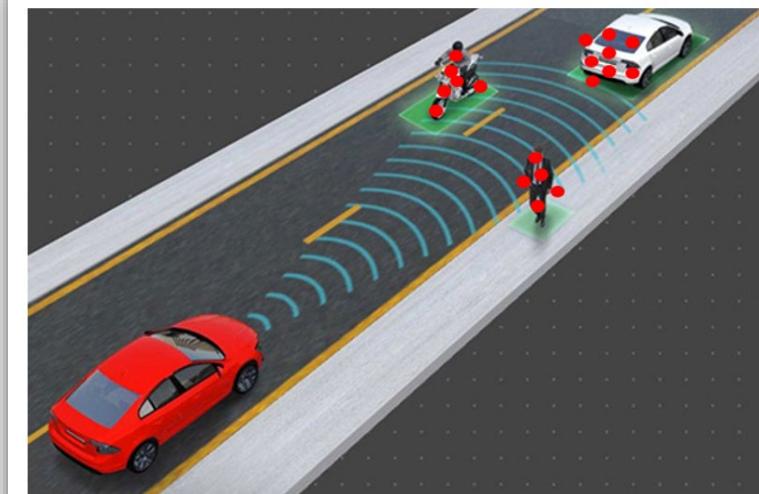
Mobileye Imaging Radar CES 2022



([source](#))

## FMCW RADAR

- Can estimate range, angle and velocity of several targets
- Strong to adverse weather conditions (rain, fog, etc...)
- Compact device. High operating frequency (77 GHz)
- TDM and antenna beamforming capabilities
- Several equipped in a single car
- Range up to 300 m



# References (3D sensors)

- [Foix 2011] S. Foix et al, **Lock-in Time-of-Flight (ToF) Cameras: A Survey**, IEEE Sensors 11(9), 2011
- [García-Gómez 2016] P. García-Gómez, et al, **Geometric Model and Calibration Method for a Solid-State LiDAR**, Sensors 20(10) 2020
- [McManamon 2012] P. McManamon, “Review of lidar: a historic, yet emerging, sensor technology with rich phenomenology,” OE 51(6), 2012
- [Mutto 2012] C. Dal Mutto, P. Zanuttigh, and G.M. Cortelazzo, **Time-of-flight Cameras and Microsoft Kinect**, Springerbriefs, Springer, 2012
- [Salvi 2010] J. Salvi et al, **A state of the art in structured light patterns for surface profilometry**, Pattern Recognition 43(8), 2010
- [Xiong 2017] Z. Xiong, et al, **Computational Depth Sensing : Toward high-performance commodity depth cameras**, IEEE SPM 34(3), 2017
- [Zhang 2012] Z. Zhang, **Microsoft Kinect Sensor and Its Effect**, IEEE Multim 19(2), 2012
- Websites: [structure.io](http://structure.io), [www.openni.org](http://www.openni.org) (now Apple!), [openkinect.org](http://openkinect.org)

# My answers

*Q1: Is “**projective vision**” a natural way to capture the 3D world?*

**Natural? Yes!**

Some living beings are equipped with a pair of projective (passive) sensors (eyes) performing stereoscopic vision to compute distance...

**Practical? Nope!**

Capturing 3D geometry can be better done with active sensors probing the actual distance to scene surfaces

*Q2: Do we need **photometry** to get **geometry**?*

**Yes!**

In 2D projective imaging we do need photometry, and then we compute correspondence disparities from stereoscopic vision or SfM to get scene geometry. But **Not!** for Lidar or TOF sensors, which can compute scene geometry by radar principles, and without resorting to disparity in photometric data...



# My answers

## *Q3: Does 3D vision mean the same than 3D geometry?*

Absolutely not!

3D vision usually assumes one (or several) points of view from which 3D geometry is computed, leaving part of the scene geometry unavailable due to occlusions!

On the contrary 3D geometry in Graphics or CAD design is a complete representation of the scene (even rendering occlusions and transparencies if needed)

## *Q4: Does 2D/3D matter for “Teaching computers to see”?*

Geometry of objects and scenes in 3D world helps analysis and recognition. It has the **potential to avoid the large variability in appearance** and missing information due to projection, occlusions, motion, shadows...

Yes, 3D geometry may definitely help computers to see and "understand" complex visual scenes, reasoning about events that evolve in space and time

