Master in Computer Vision *Barcelona*

Module: M4 3D Vision

Lecture: **4.9b 3D sensors**

Lecturer: Josep R. Casas

Motto 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** mean the same than **3D geometry**?*

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

Introduction to 3D Sensors and Range Data

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 FVV: 'free viewpoint view')

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

Introduction to 3D Sensors and Range Data (cont)

- 2D vs 3D

  "2D imaging

  …projects 3D scenes onto a planar surface (retina, sensor)…

  …so that the depth (Z) dimension is lost"

  – *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**?*

Introduction to 3D Sensors and Range Data (cont)

- Distance measurement methods…

Range sensors (scanners!)

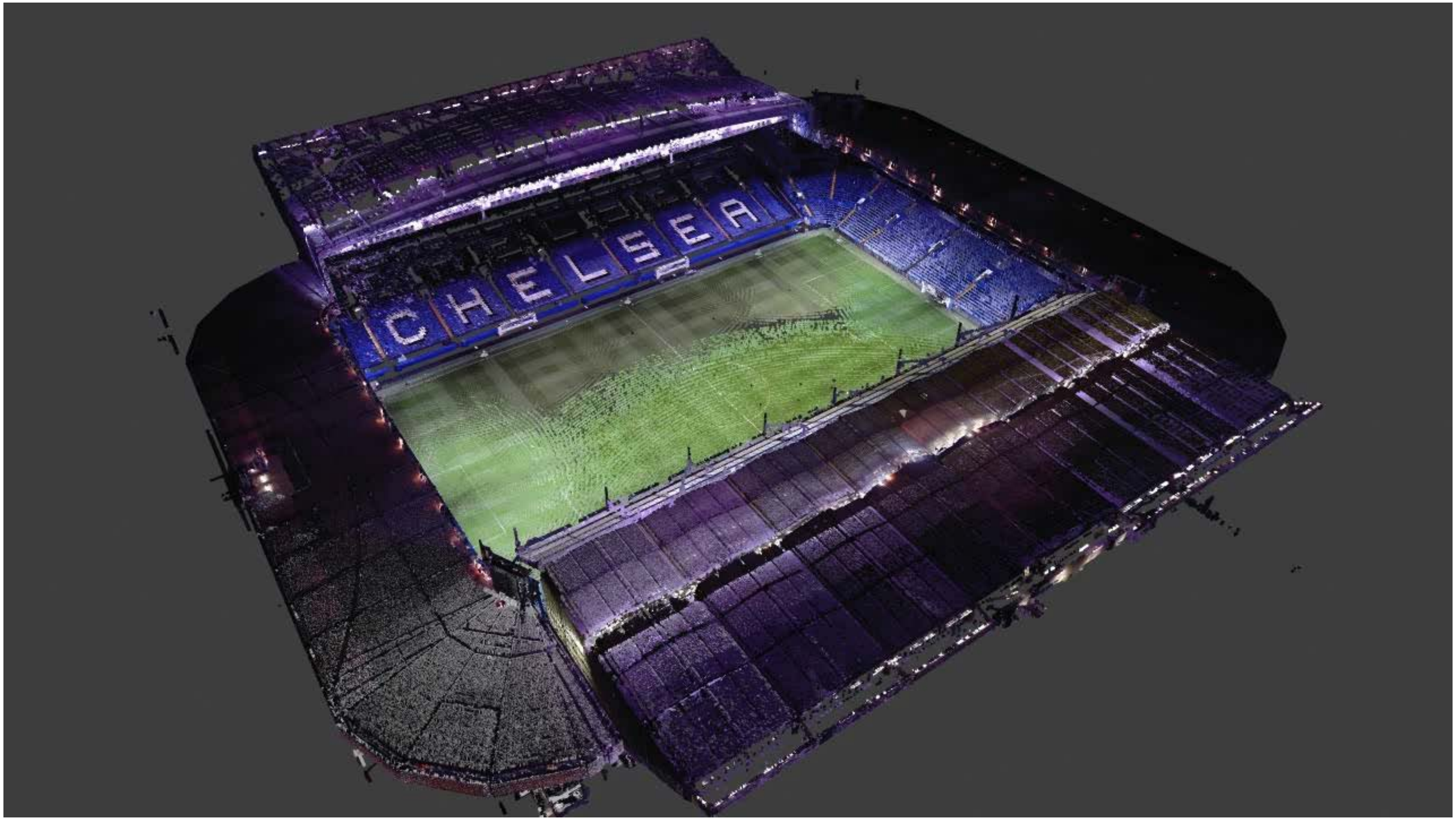
- 3D scanners and Lidar systems
- Light coding systems (e.g. Kinect™)
- TOF cameras (e.g. Kinect™ v2)

…and then

- *Will this result in a complete view… like a CAD design?*
- *Panoramic / Surround view?*
- *Free viewpoint?*

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

Example of 3D capture



Lidar capture of the Stamford Bridge (cf. FascinatE project)

Example of 3D processing

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

3D/range/scan sensors
vs
Multiple View

Advantages 3D/range/scan

- …
-

…

- Direct measure of 3D geometry
- Independent of photometry (active)
- From scan to CAD blueprints

Disadvantages

- …

…

- 2,5D!

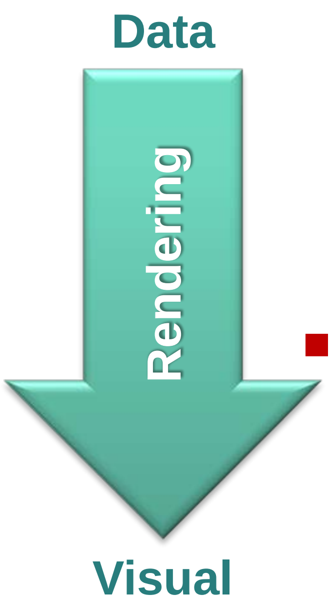
  Surface vs 3D (interface air-matter vs volumetric 3D scanner, i.e. PET)
- Single viewpoint

  Neither panoramic, nor surround view
- Accuracy?

  vs MPix cams!

3D Data

Double nature of 3D data



- Geometric information

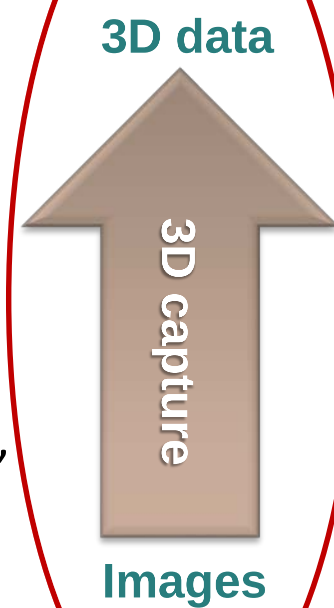
  Pure data, measures.
  Numerical representation of objects
  → calculation, simulation, measurements

  … + reflectivity

- Photometric information (images)

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

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



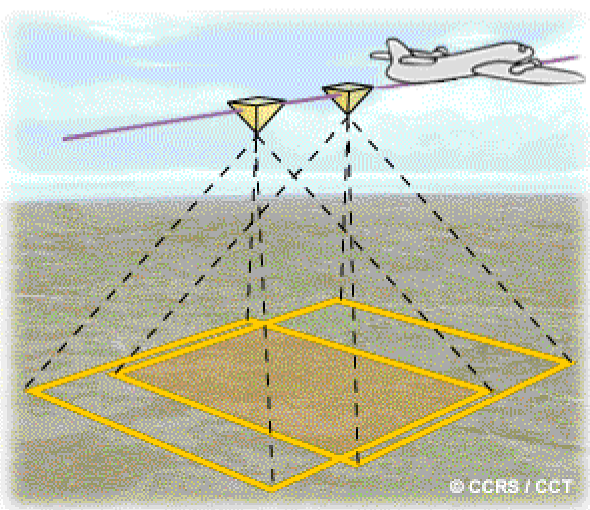
Photogrammetry

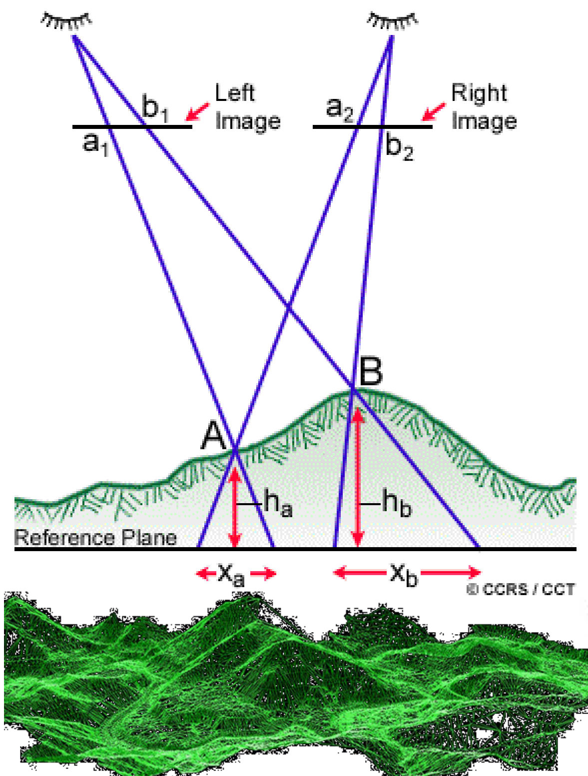
Photogrammetry refers to the practice of deriving 3D measurements from photographs.

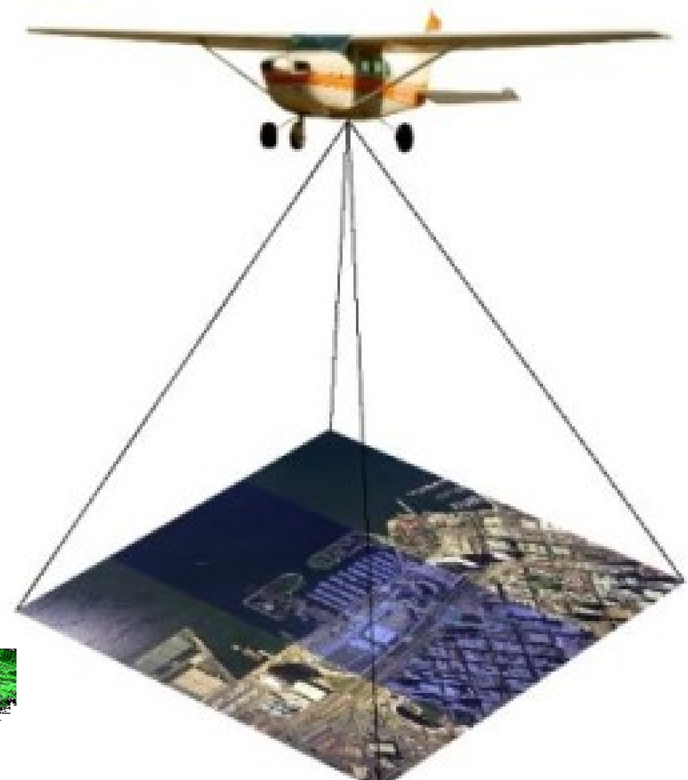
- Advances in digital cameras, processors, and computation, yield **extremely dense and accurate 3D surface data** from a limited number of photos with standard digital photography
- Structure from Motion (SfM) and sub-pixel image matching yield **3D-dimensional structure by analyzing projected 2D motion 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

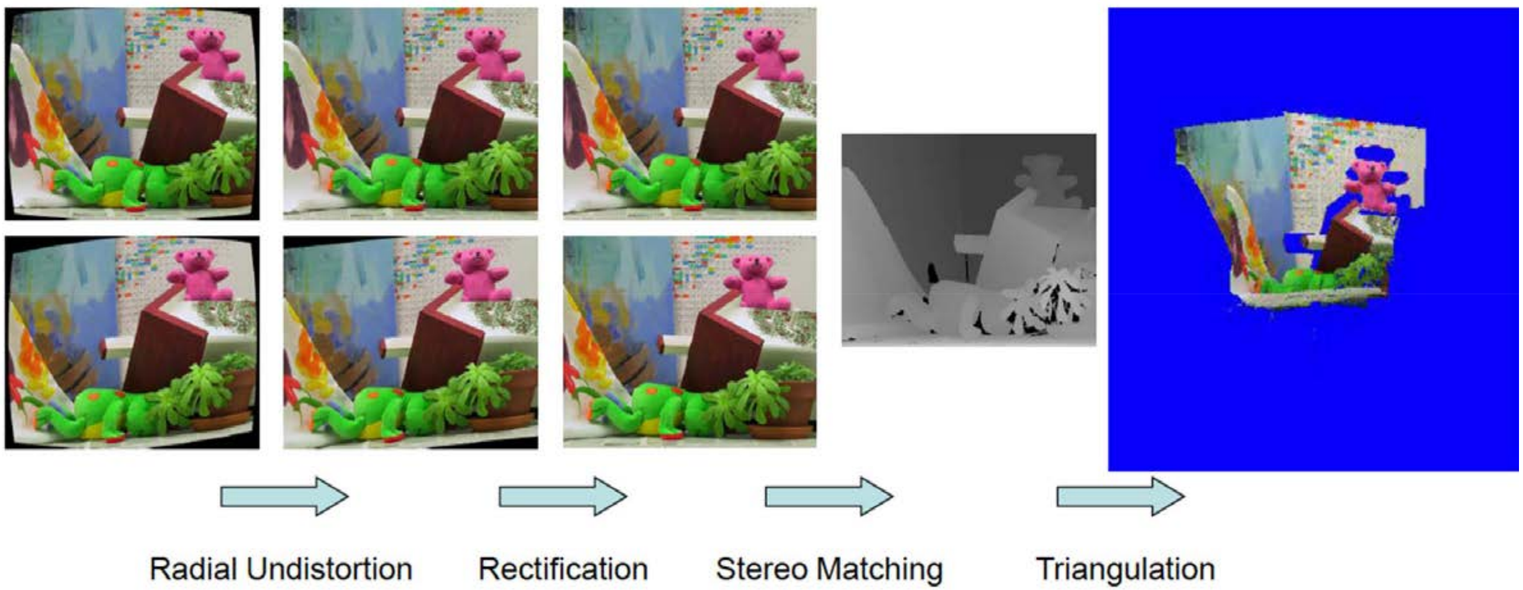






3D/range/scan sensors
vs
Multiple View

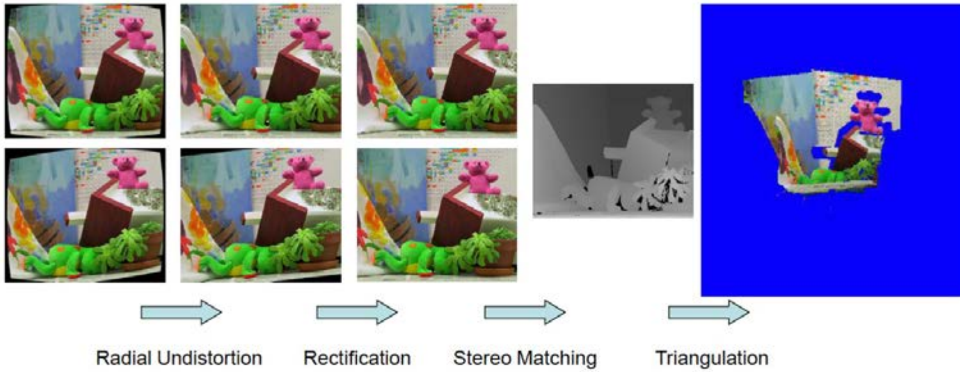
Depth Measurement Using Multiple Camera Views



From N. Navab et al, *TUM* 2011

3D/range/scan sensors
vs
Multiple View

Depth Measurement Using Multiple Camera Views



Disadvantages:
- At least two calibrated cameras required
- Multiple computationally expensive steps
- Dependence on scene illumination
- Dependence on surface texturing

From N. Navab et al, *TUM* 2011

3D data

Depth, cloud, mesh...

3D data representation

- Depth map
  - **2.5D** (concept: RGBD = 2D + depth)
- Point cloud
  - **organized**: keeps relationships in sensor neighborhood
  - **unorganized**: one can ***just*** compute nearest neighbors in 3D
- Mesh
  - nice scanned/reconstructed surfaces: watertight / convex...

Point cloud data

- Organized point-cloud

  Resemble an organized image (or matrix-like) structure, with data split into rows and columns (data from stereo, depth or TOF sensors)

  → **projectable** point cloud: has a correlation according to a pinhole camera model between the (u,v) index of a point in the organized point cloud and the actual 3D values.
  This correlation can be expressed as: u = f*x/z and v = f*y/z

  → knowing the relationship between adjacent points (e.g. pixels), nearest neighbor operations are much more efficient, thus speeding up the computation and lowering the costs of certain algorithms in PCL

- Unorganized point-cloud

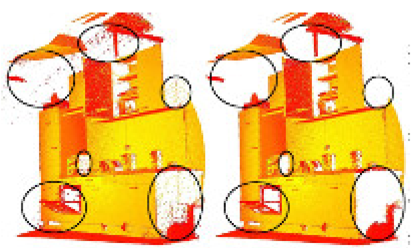
  Non-regular sampling of 3D space
  Neighborhood operations require KD tree search!

http://pointclouds.org/documentation
[Rusu 2009, Rusu 2011]

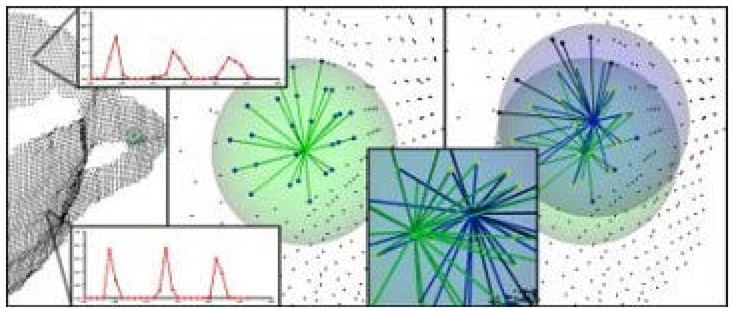
Point Clouds

- New trend to process raw data produced by scanners
  - Point Cloud → PCL library

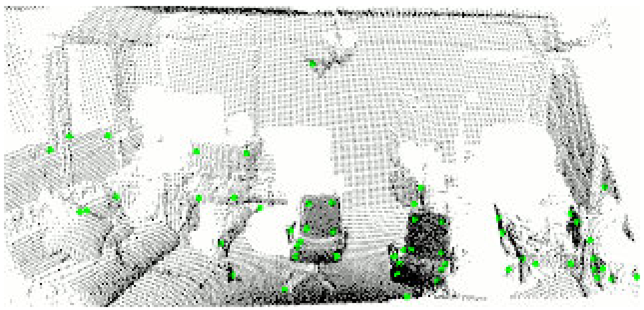
pointclouds.org



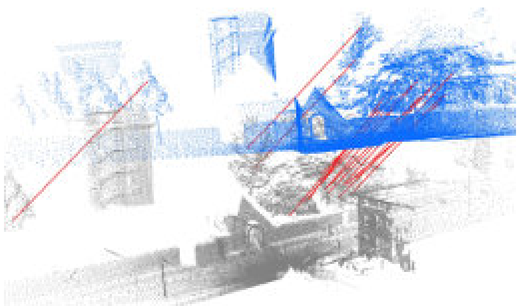
Filters



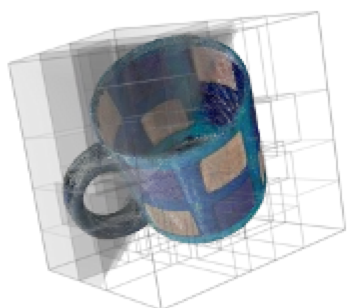
Features



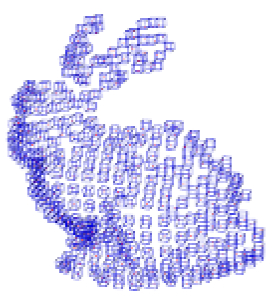
Key points



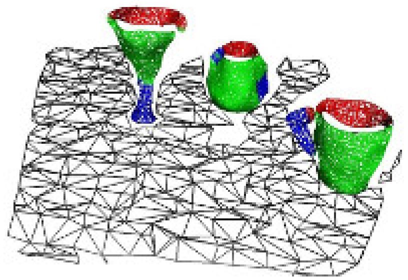
Segmentation



RANSAC



Octree

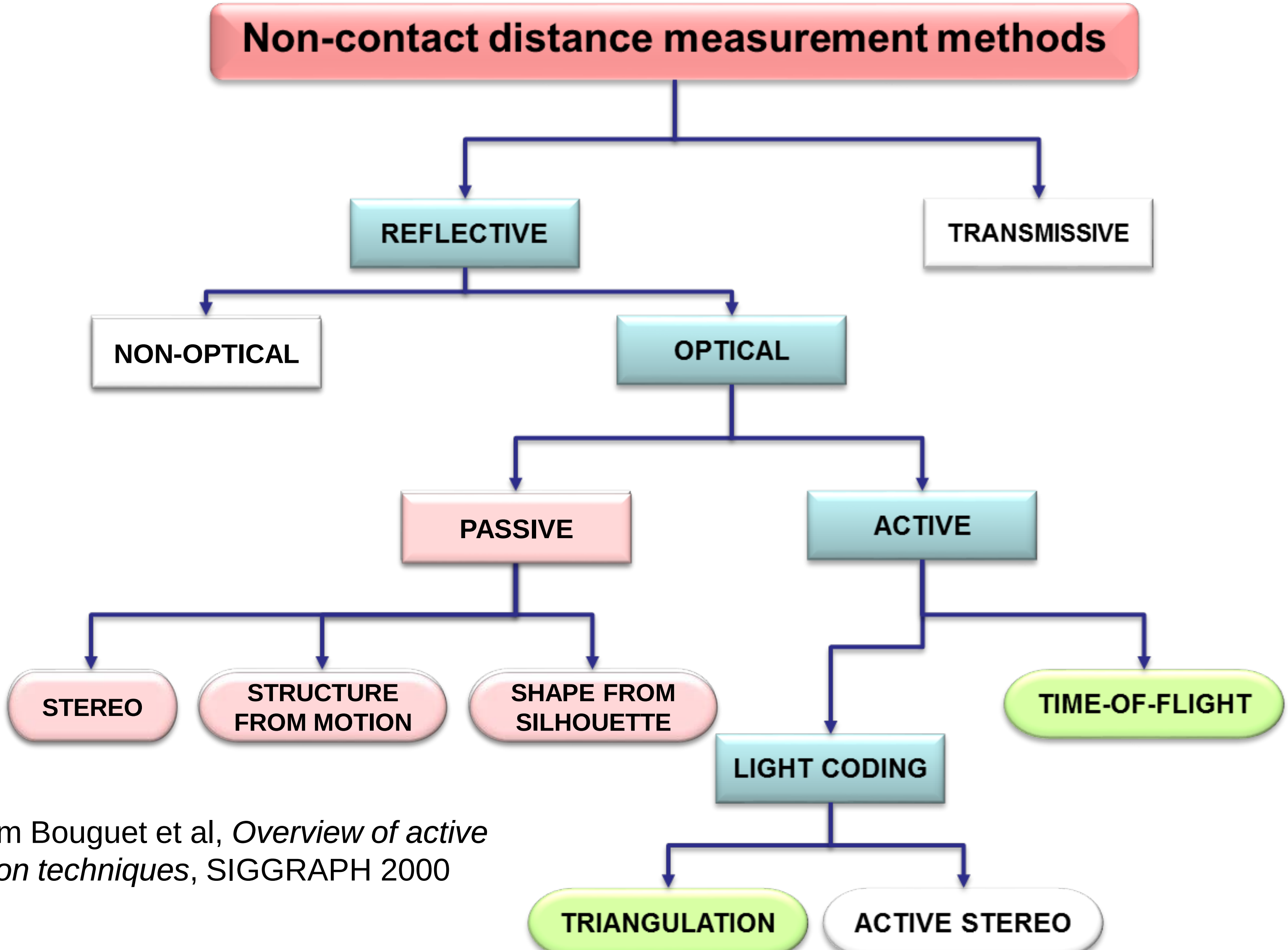


Visualization

3D sensors

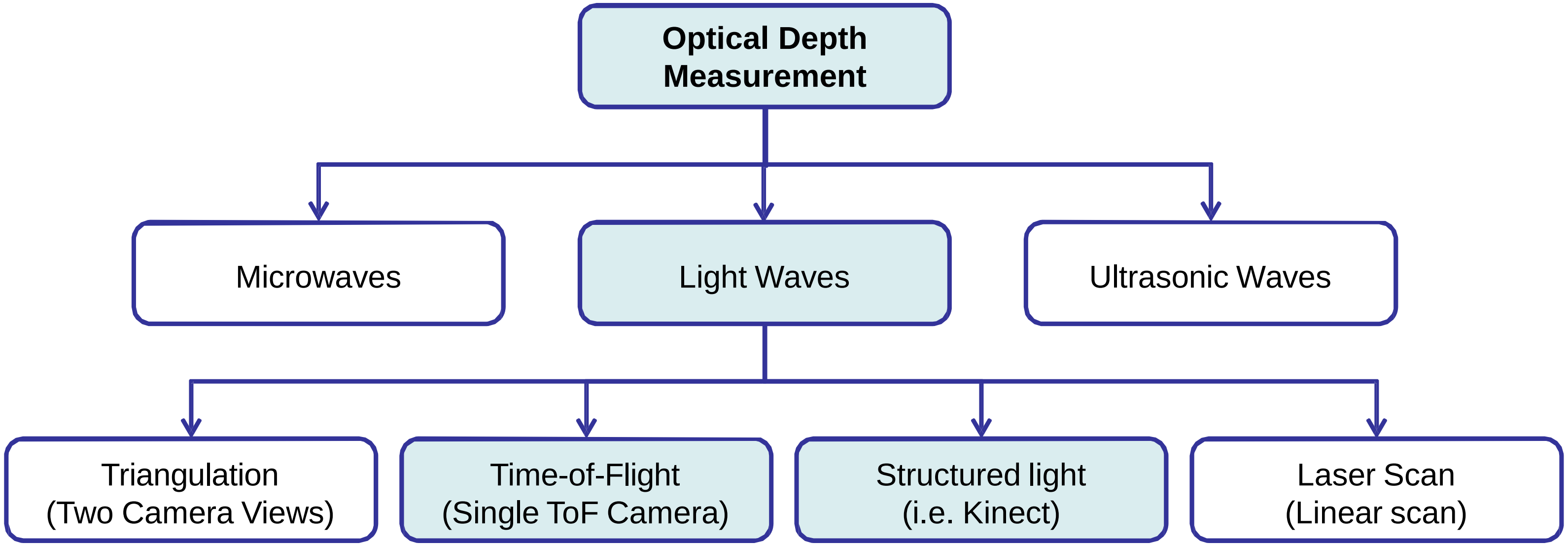
Depth, range, scan...

“Non-contact” – distance measurement methods



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

Optical Depth Measurement Techniques



From N. Navab et al, *TUM* 2011

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

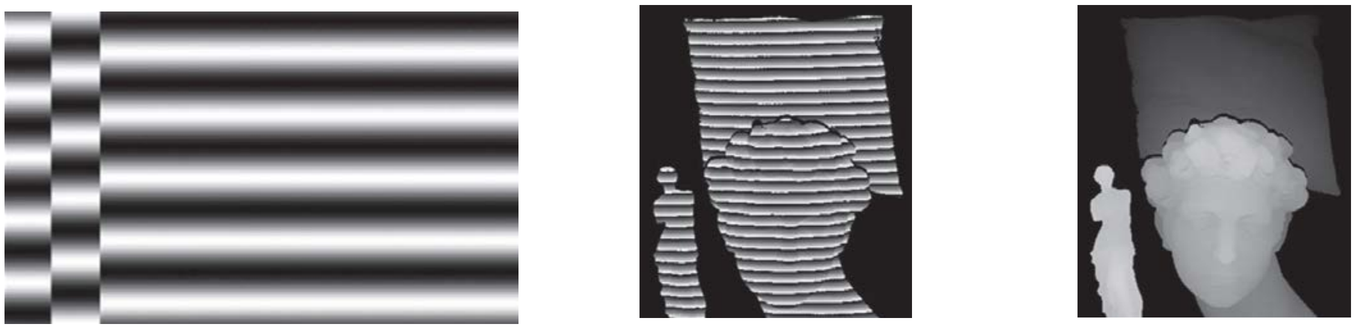
  3D distance measurement obtained by projecting in the scene some form of radiation

  – 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]

Computational Depth Sensing

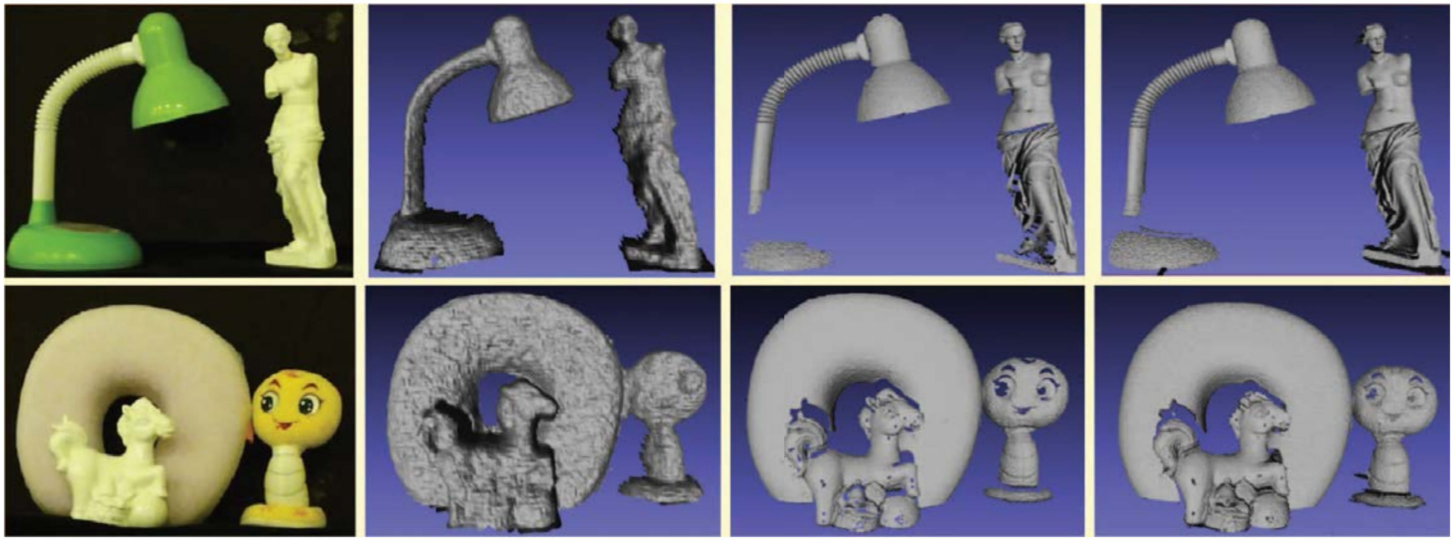
- Depth cameras do not sense depth information directly, rather, through either the space deformation or the time delay of light signals
- Computation needed!

Computational Depth Sensing: image sensors with onboard advanced signal processing algorithms (Redesign [Xiong 2017])



Example of phase shifting (Xiong et al, 2017

3D reconstruction results



3D reconstruction results for texture-less surfaces:

(a) color image

(b) ToF results

(c) three-frequency phase shifting (nine patterns)

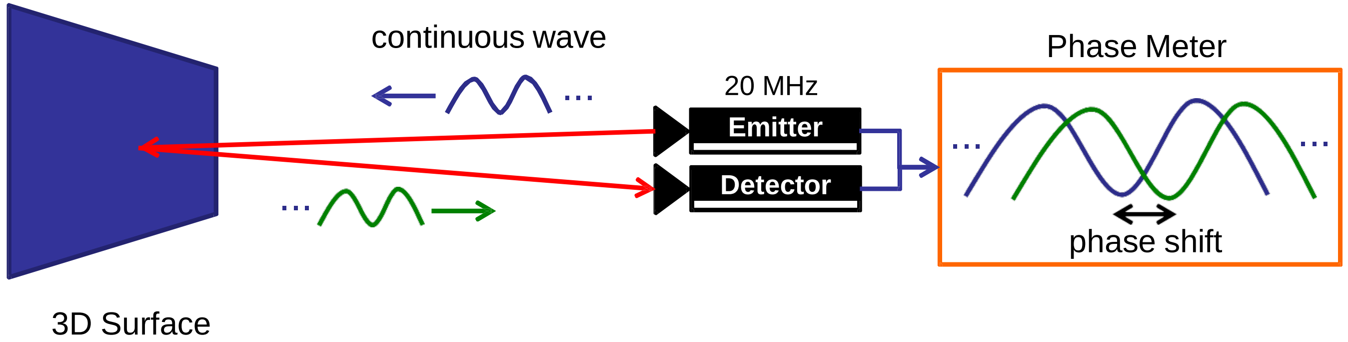
(d) ToF and phase-shifting fusion (three patterns)

(Xiong et al, 2017)

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



Principles of ToF sensors

Continuous Wave Modulation

Advantages:

- Variety of light sources available as no short/strong pulses required
- Applicable to different modulation techniques (other than frequency)
- Simultaneous range and amplitude images

Disadvantages:

- In practice, integration over time required to reduce noise
- Frame rates limited by integration time
- Motion blur caused by long integration time

TOF *Imaging*

Swissranger™ SR4000
Mesa Imaging

Matricial Time-Of-Flight Cameras

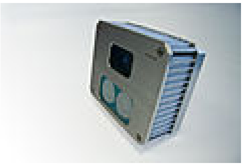
- Active sensors
- Acquire 3D geometry at video rate (up to 50fps)
- E.g. MESA Imaging
  PMD Technologies
  Optrima SoftKinectic
  CANESTA (*acquired by Microsoft in 2010*)
  ...

Kinect for Xbox One by Microsoft

D-IMager
Panasonic

pmd[vision] CamCube
PMD Technologies

FOTONIC-B70
Fotonic



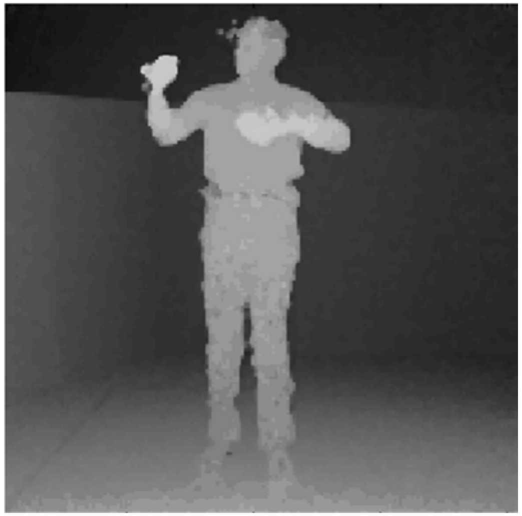
3D MLI Sensor
IEE S.A.

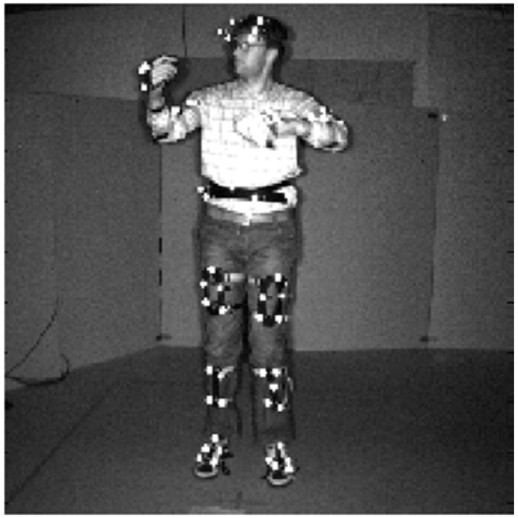
pmd[vision] CamBoard
PMD Technologies

Principles of ToF Imaging

Continuous Wave Modulation

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


Depth Image

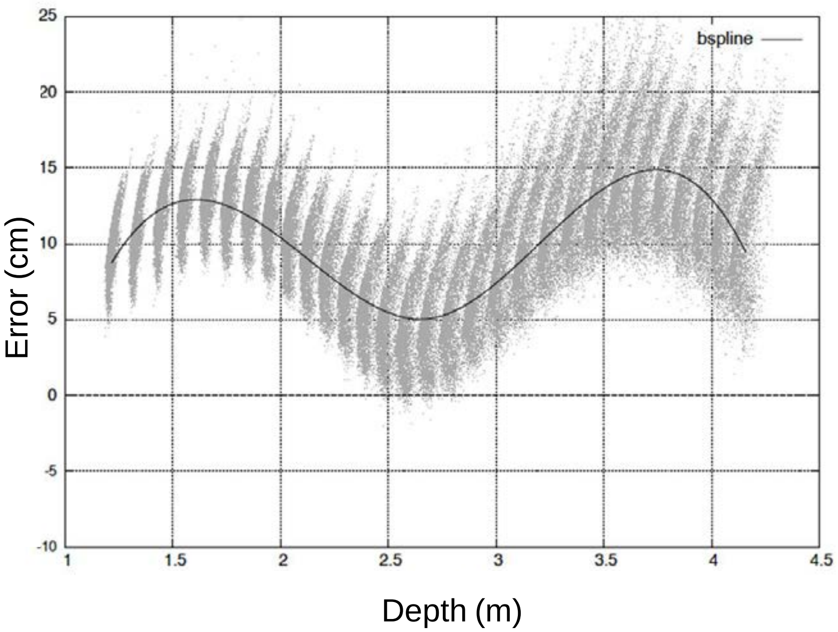

Amplitude Image

Computer Vision with ToF Cameras

Measurement Errors and Noise

Systematic distance error

- Perfect sinusoidal signals hard to achive 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

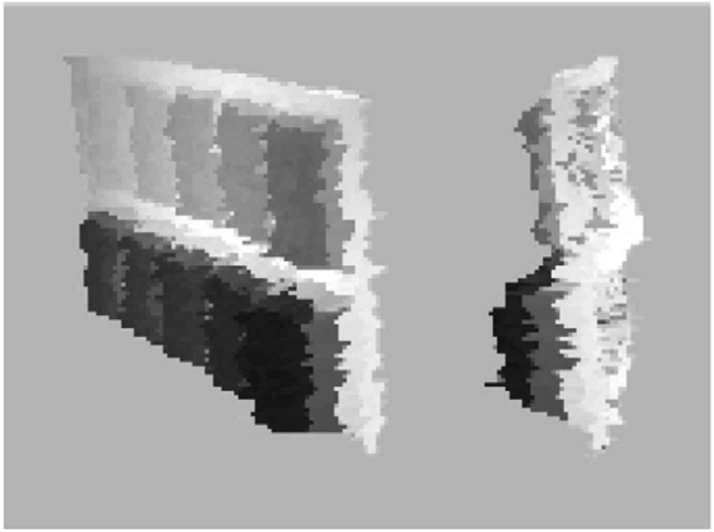


Computer Vision with ToF Cameras

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

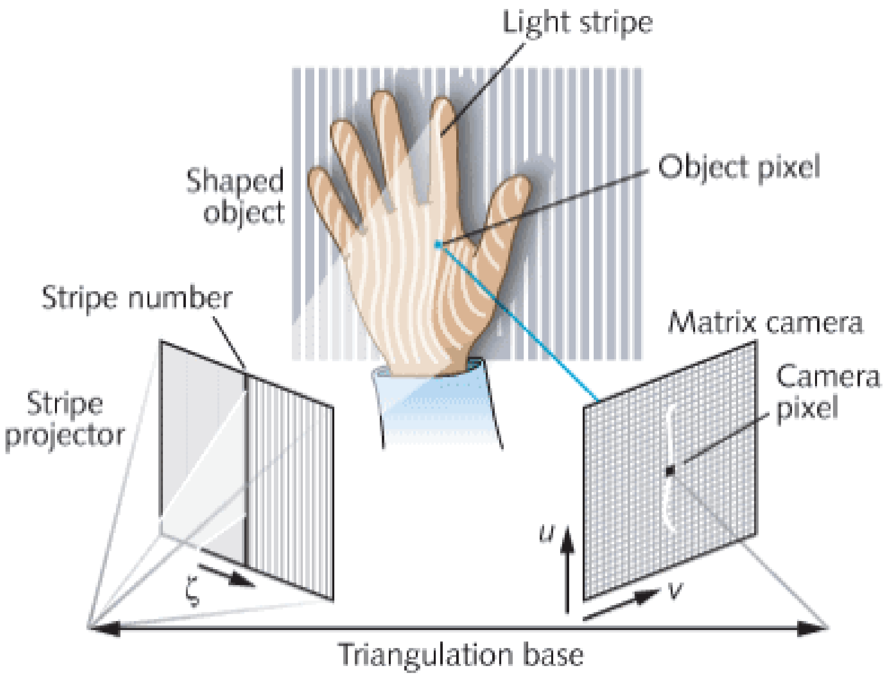
- Current ToF cameras have low pixel resolution
- Individual pixels get different depth measurements
- Inhomogeneous
- „Flying pixels“, especially at object bondaries
- Correction: discard pixels along rays parallel to viewing direction

Red circles: „flying pixels“

Structured Light methods

Structured light methods

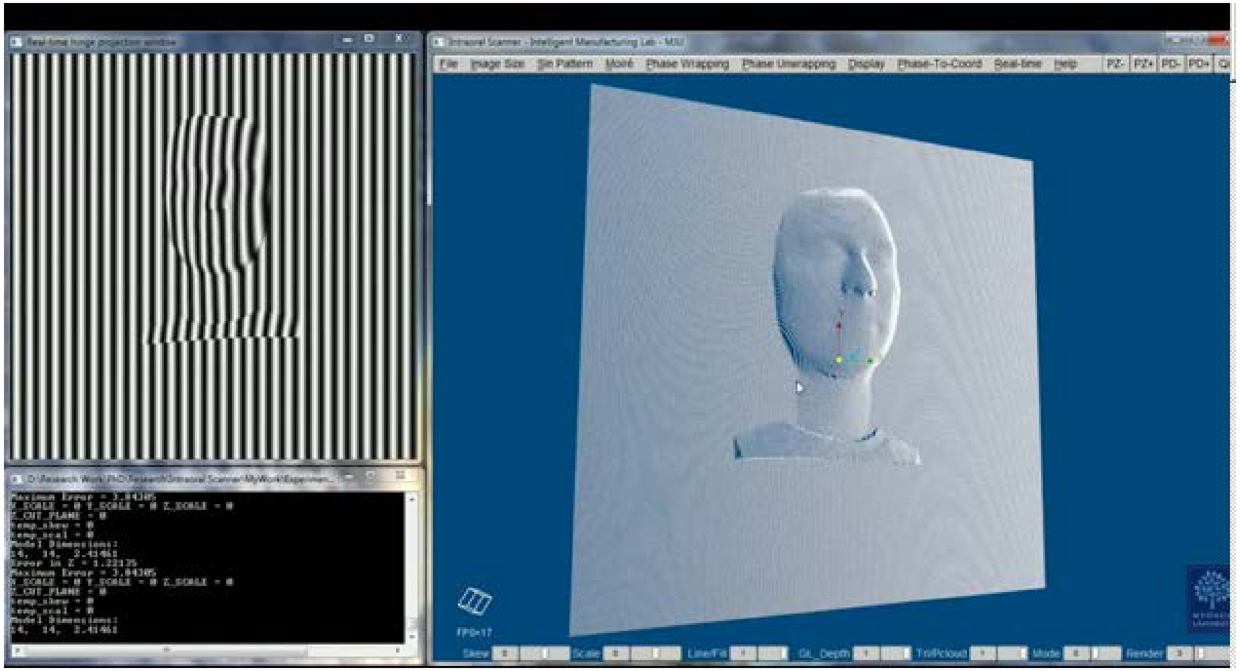
- Calculate the shape by how the strip is distorted.



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

What is 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 1st, 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

Commercial Depth Sensors: Specifications



	[Kinect]()	[Kinect-2]()	[Xtion Pro Live]()	[Structure]()
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)	–	–

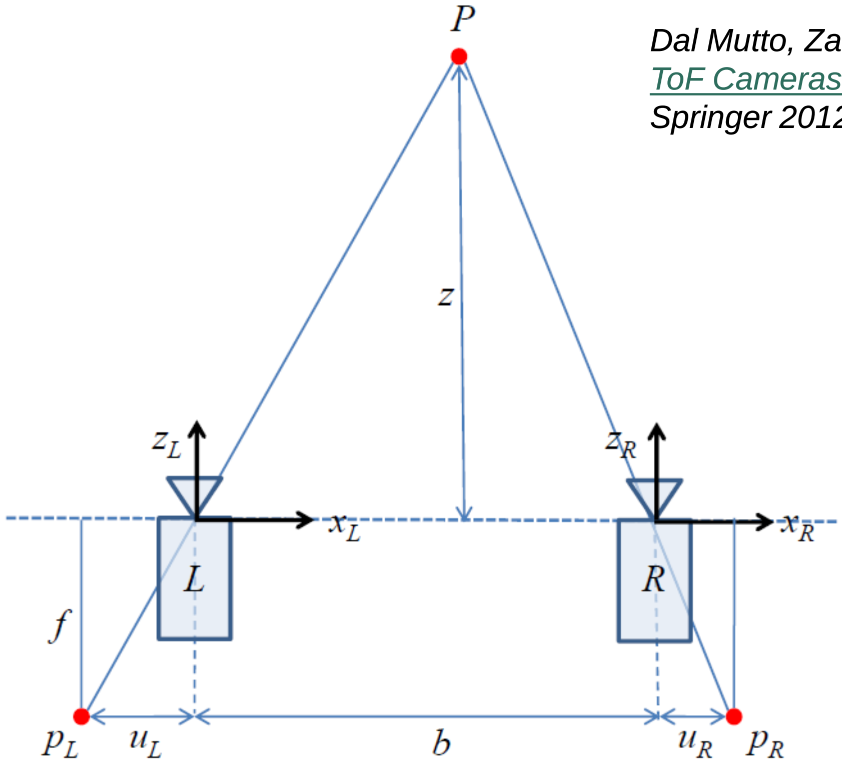
Others: e.g. *faceshift*

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



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

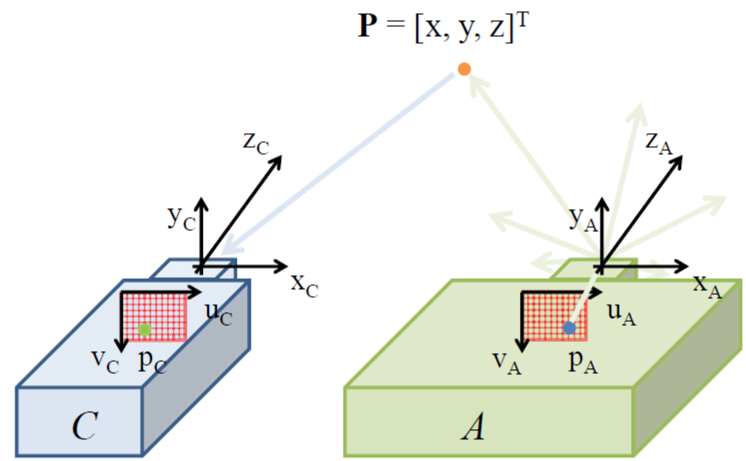
Conjugate pixels: p_L , p_R

Disparity: $d = u_L - u_R$

Depth: $z = bf/d$

(calibrated and rectified)

Active Triangulation



Same geometry:

camera (R) replaced by projector (A)

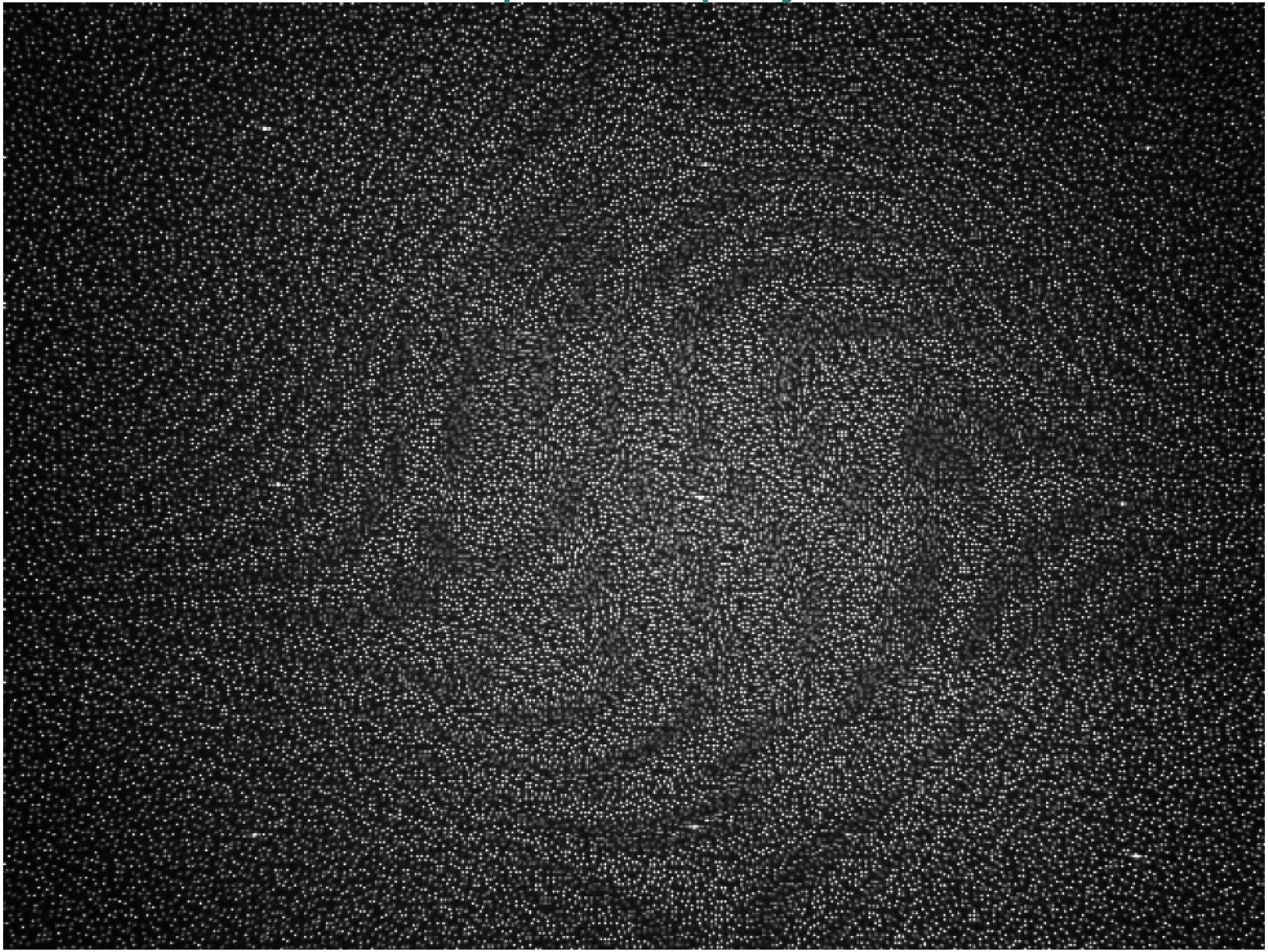
p_A pattern pixel, p_C camera pixel

p_A, p_C are conjugate → disparity d

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

(calibrated and rectified)

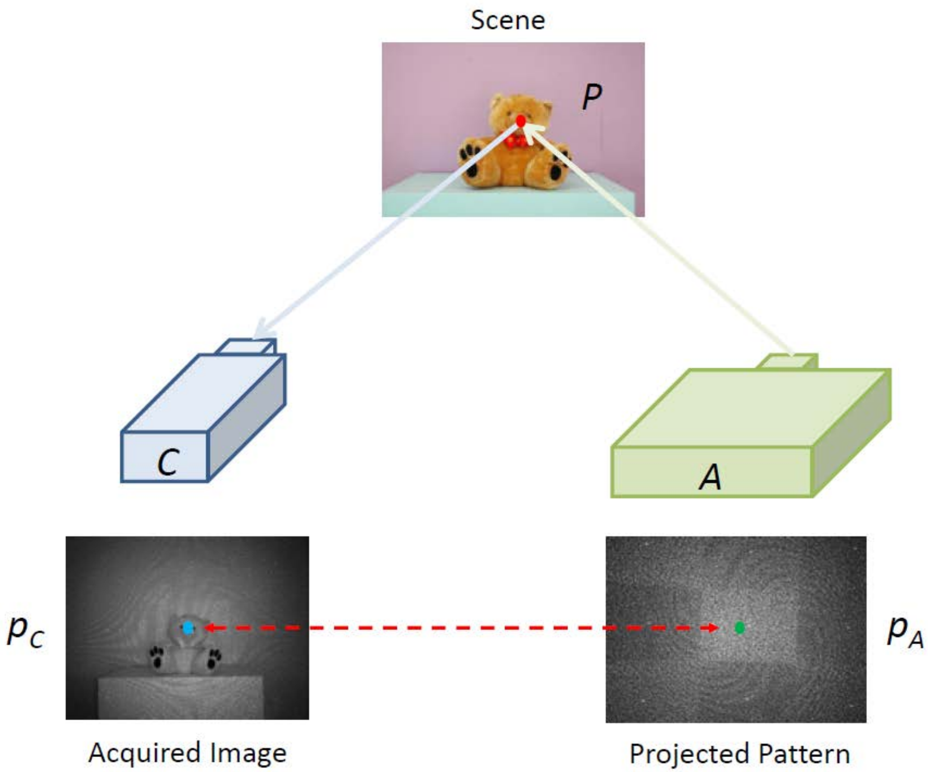
Kinect pattern projected

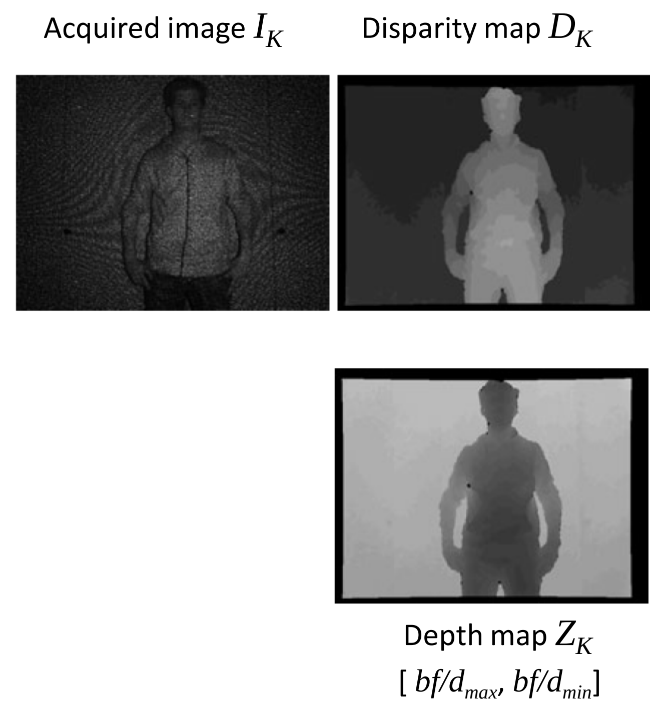


(from Dal Mutto, el al 2012)

Matricial Active Triangulation

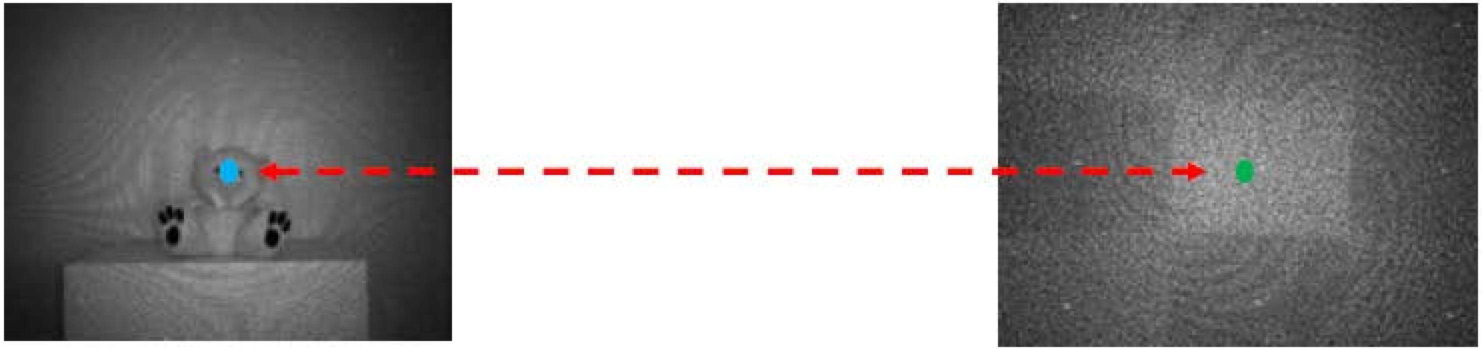
IR Projector and camera (invisible)





(from Dal Mutto, el al 2012)

Structured light coding

p_C  p_A

Acquired Image   Projected Pattern

- 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)

Light coding strategies

- Direct coding
  - unique pattern value at every pixel → n_p codewords

  A: able to deal with occlusions and perspective distortion

  D: sensitive to color/gray distortion, noise and camera non-idealities
- Time multiplexing
  - sequence of T values projected/measured T times → n_P^T codewords

  A: similar direct coding, but more robust against noise

  D: need some integration time: not suited to dynamic scenes
- Spatial multiplexing → Kinect sensor (PrimeSense)
  - spatial distribution in n_W pixels window around p_A → $n_P^{n_W}$ codewords

  A: can handle color/gray distortion, reflectivity, illumination, noise

  B: worse with occlusion and perspective distortion

  → smaller n_W, more robust to perspective, but less to noise, color/gray distortion

Lidar

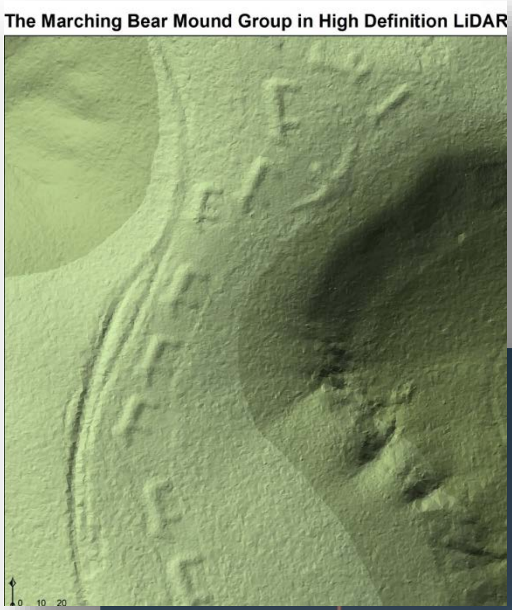
Surveying method/Measurement technique:

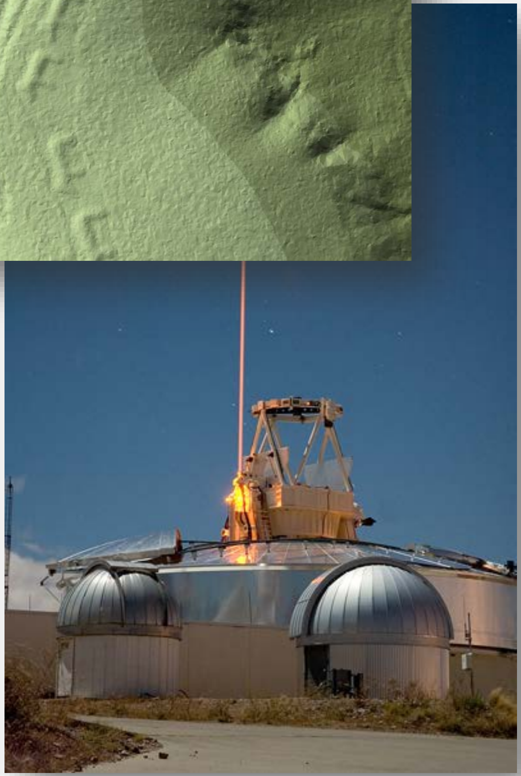
Pulser laser light illuminates an object, the reflected pulses are recovered and measured with a solid-state sensor

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

Popular applications:

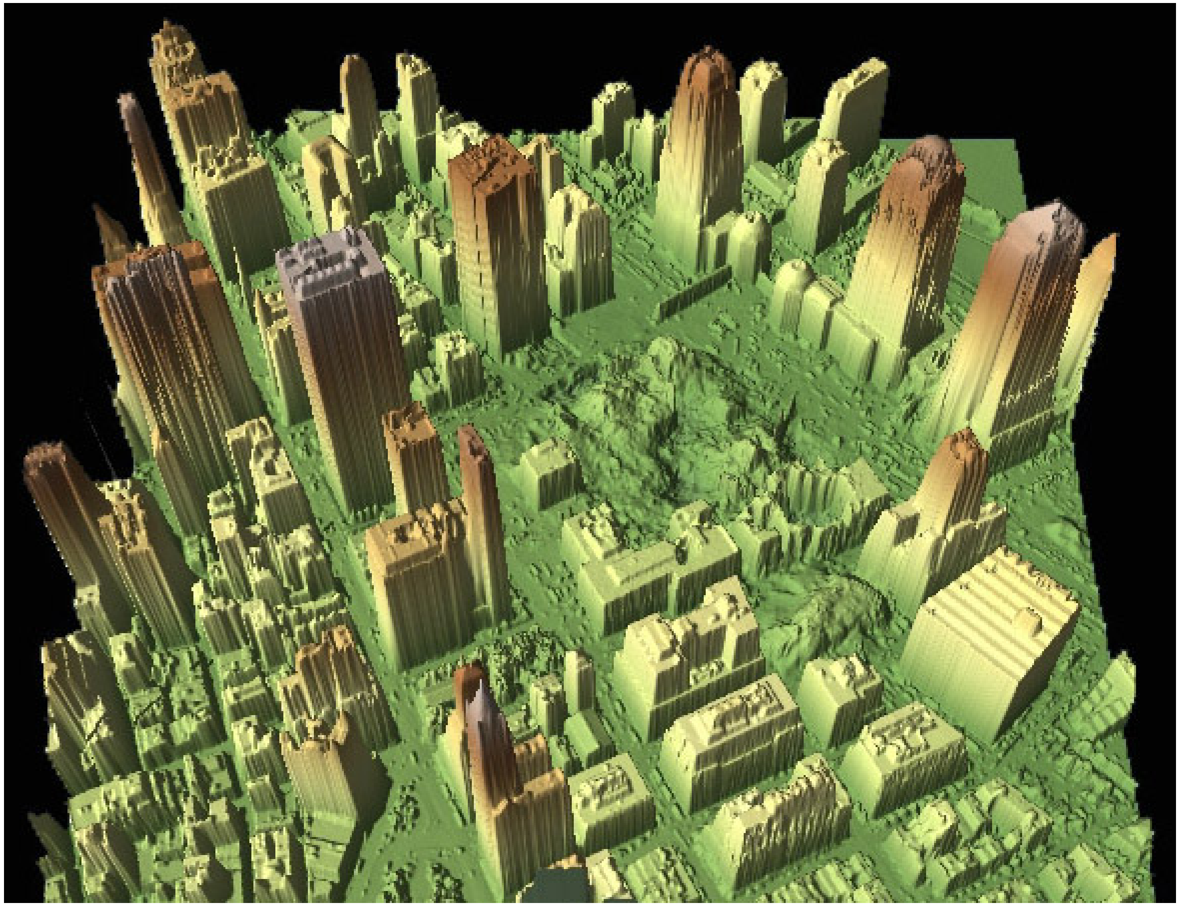
Earth resources exploration, landing aids, object recognition, self-guided vehicles and safety and security applications in transport





LIDAR light detection and ranging scanner

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

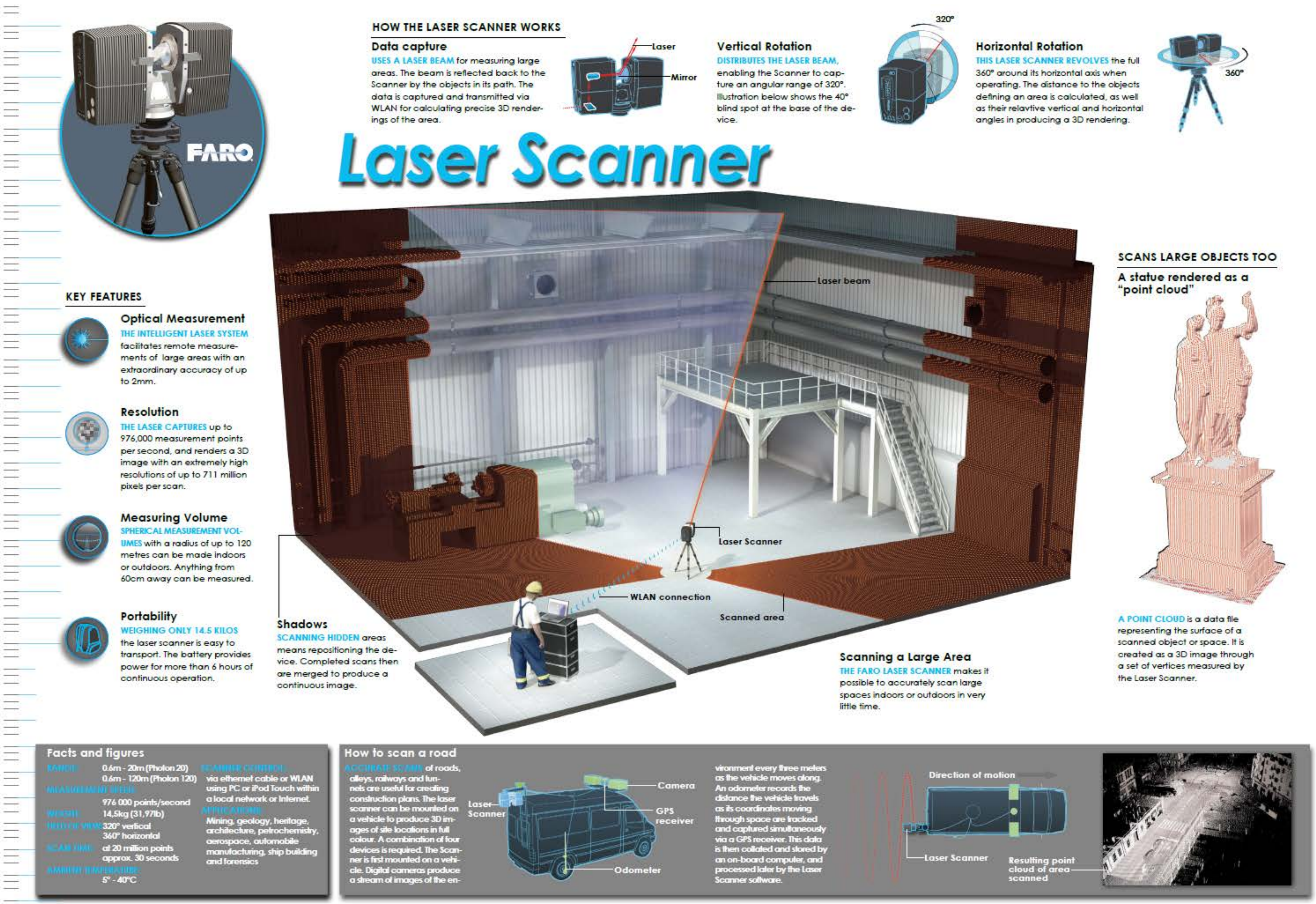


Leica lidar

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

http://commons.wikimedia.org/wiki/File:Lidar_P1270901.jpg

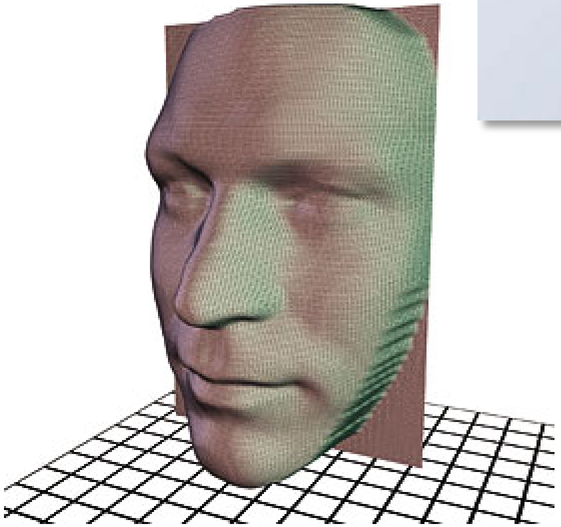
FARO scanners



Applications of 3D scanners

- Industrial design “CAD/CAM”.
- Orthotics and Prosthetics
- Entertainment industry: production of movies and video games
- Reverse engineering
- Inspection and documentation of cultural artifacts

E.g. Metrology (QC): nub3d.com, FARO
Face scanners



Comparison chart

Technology	Outdoor	Detection range [m]	Accuracy [cm]	FOV	Angular resolution	Obst Det. Terr Mapp.
HD-LIDAR	✓	120	4	27°V 360°H	0.4°V 0.09°H	Not included
Multi-Plane LIDAR	✓	120	10	3°V 85°H	0.8°V 0.125°H	OD, external unit
Kinect	✗	4-5	Up to 0.2	43°V 57°H	0.09°V 0.09°H	Not included
3DV	✓	**Up to 40**	**Variable**	**32-50°V 40-65°H**	**0.06°V 0.06°H**	**OD+TM, plugins**

3DV VisLab.it (video)

Motto 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** mean the same than **3D geometry**?*

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

References (3D sensors)

[Bhandari 2016] A. Bhandari, R. Raskar, **Signal processing for TOF imaging sensors: An introduction to inverse problems in computational 3-D imaging**, IEEE SPM 33(5), 2012

[Bishop 2012] E. Bishop, P. Favaro: **The Light Field Camera: Extended Depth of Field, Aliasing, and Superresolution**. IEEE PAMI 34(5), 2012

[Mutto 2012] C. Dal Mutto, P. Zanuttigh, and G.M. Cortelazzo, **Time-of-flight Cameras and Microsoft Kinect**, *Springerbriefs*, Springer, 2012

[Foix 2011] S. Foix et al, **Lock-in Time-of-Flight (ToF) Cameras: A Survey**, IEEE Sensors 11(9), 2011

[Salvi 2010] J. Salvi et al, **A state of the art in structured light patterns for surface profilometry**, Pattern Recognition 43(8), 2010

[Smisek 2011] J. Smisek, et al, **3D with Kinect**, *ICCV Workshops* 2011

[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: www.openni.org, openkinect.org, www.primesense.com, kinecthacks.net

References (pointclouds, processing and meshes)

[Izadi et al 2011] S. Izadi et al, KinectFusion: **Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera**, *ACM Symposium on User Interface Software and Technology*, October 2011

[Kazhdan 2006] M. Kazhdan 2006, **Poisson Surface Reconstruction,** *research.microsoft.com/en-us/um/people/hoppe/proj/poissonrecon*

[Meshlab] Visual Computing Lab CNR (it), **Meshlab,** meshlab.sourceforge.net, tutorials

[Newcombe et al 2011] **KinectFusion: Real-Time Dense Surface Mapping and Tracking**, *IEEE ISMAR*, October 2011

[PCL] PointClouds.org, **PCL and tutorials,** pointclouds.org/documentation

[OpenMesh] OpenMesh.org, **OpenMesh,** www.openmesh.org

[Rusu 2009] R. B. Rusu, **Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments**, Tecvhnische Universität München, **PhD 2009**

[Salvador et al 2011] J. Salvador and J. Casas, **Multi-View Video Representation Based on Fast Monte Carlo Surface Reconstruction,** *IEEE Trans. IP* 22 (9) 2013

Complete courses

- Poisson Surface Reconstruction

  research.microsoft.com/en-us/um/people/hoppe/proj/poissonrecon
- Digital Geometry Processing, Technion

  webcourse.cs.technion.ac.il/236329/Spring2013
- VCGlib and Meshlab

  vcg.isti.cnr.it/vcglib, Visual Computing Lab CNR (it) course

  meshlab.sourceforge.net, tutorials
- PCL and tutorials

  pointclouds.org/documentation
- OpenMesh

  www.openmesh.org

What to do with the results?

MESHING

Motivation for meshing

In many domains, scanners are used to obtain virtual representations of 3D shapes

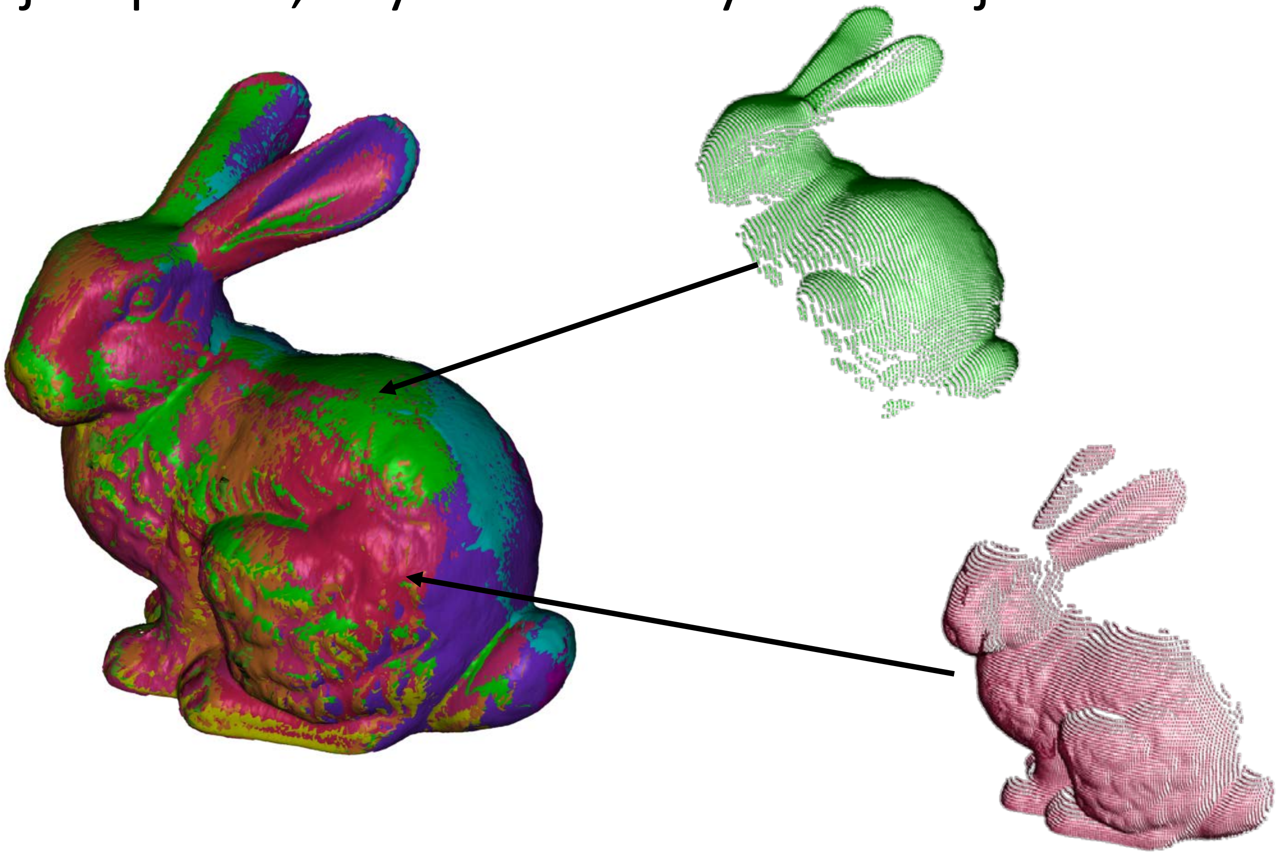
http://www.jhu.edu/digitalhammurabi/

http://graphics.stanford.edu/projects/mich/

Scanner results

Scanning often gives only local connectivity…

…or even just points, any connectivity at all → just a PointCloud!

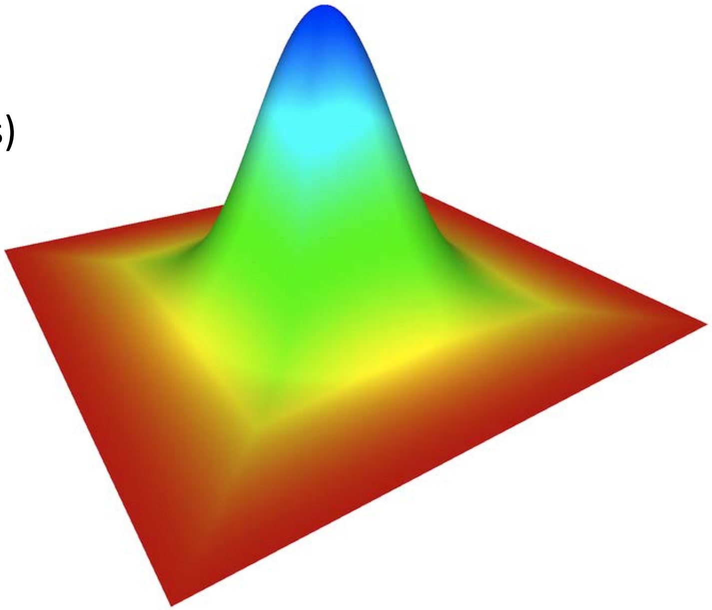


but yet… There is some motivation for meshing

We want a 3D Mesh for:

- Parameterization
- Computational Analysis
- Rapid Prototyping
- Rendering
- Collision Detection

Meshing



Mesh

Defines an ordering of the surface points (vertices) for interpolating a continuous surface in intermediate positions

Strategy

Greedy contour propagation under a set of rules for topological correctness

Implementation

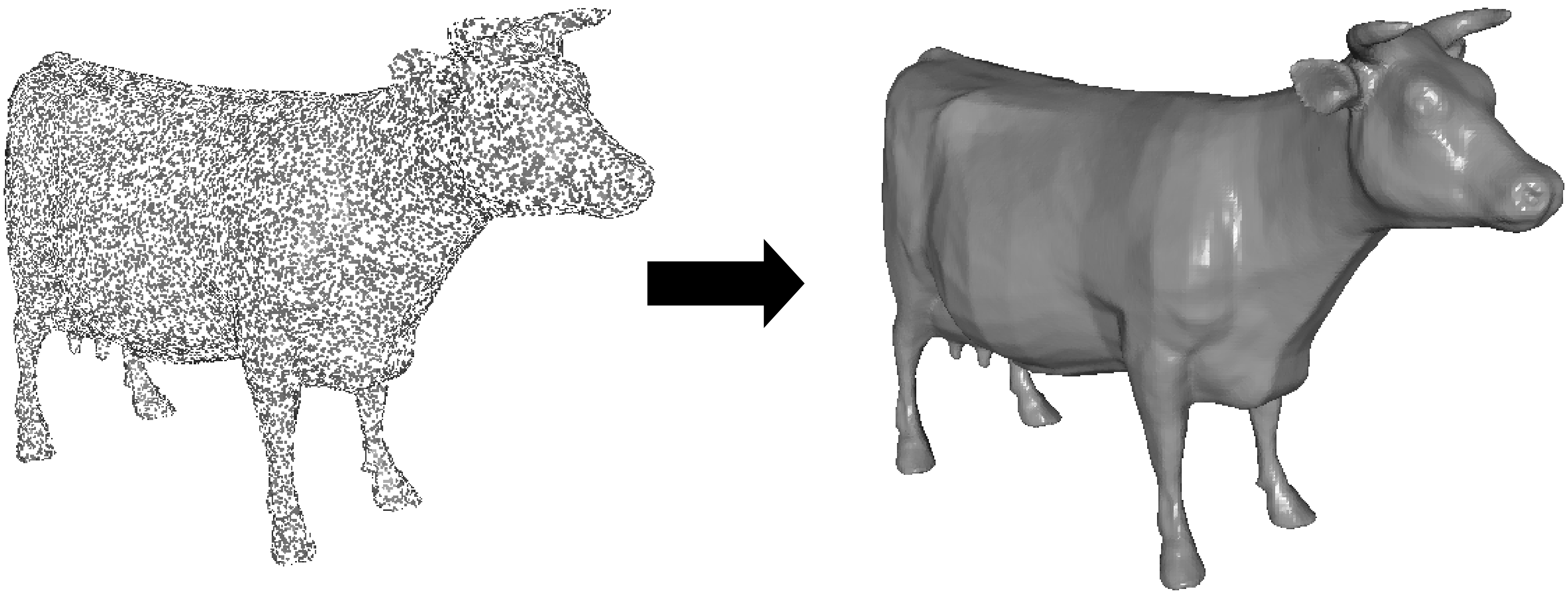
Efficient 3D spatial queries of neighbor points via *kd-tree*

applied iteratively for processing disconnected sets of points

Surface Reconstruction

M. Kazhdan *Johns Hopkins Univ. Poisson Reconstruction Eurographics 2006*

Generate a mesh from a set of surface samples



Surface Reconstruction

Generate a mesh from a set of surface samples

- Three general approaches:
  1. Computational Geometry
     - Boissonnat, 1984
     - Edelsbrunner, 1984
     - Amenta *et al.*, 1998
     - Dey *et al.*, 2003
  2. Surface Fitting
     - Terzopoulos *et al.*, 1991
     - Chen *et al.*, 1995
  3. Implicit Function Fitting
     - Hoppe *et al.*, 1992
     - Curless *et al.*, 1996
     - Whitaker, 1998
     - Carr *et al.*, 2001
     - Davis *et al.*, 2002
     - Ohtake *et al.*, 2004
     - Turk *et al.*, 2004
     - Shen *et al.*, 2004
     - Kazhdan, 2005

Meshing benchmark

- RMS *Hausdorff* distance between two non-empty subsets *X* and *Y* of a metric space (*M, d*), with *d* the Euclidean distance in 3D space:

$$d_H(X,Y) = \sqrt{\frac{1}{N}\sum_{x\in X} \| \inf_{y\in Y} d(x,y)\|^2}$$

- Computation time
- Memory footprint