Mixed Reality and Tangible interface
3D scene acquisition

Jeanne Vezien
Vezien@limsi.fr

Master Interaction– HCI Master - Université Paris-Saclay

Plan of the lecture

1. Intro : 3D acquisition for AR

2. Techniques for 3D data acquisition

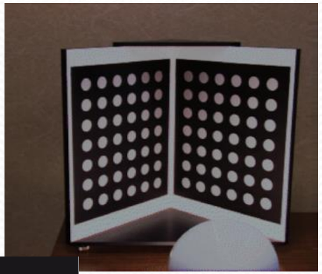
3. 3D acquisition and Tangible Interfaces

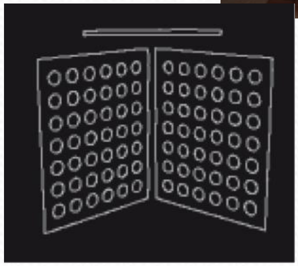
3D Acquisition for AR

http://nickzucc.blogspot.com/2012/03/3d-kinect-home-scanner.html

Augmentation needs 3D acquisition

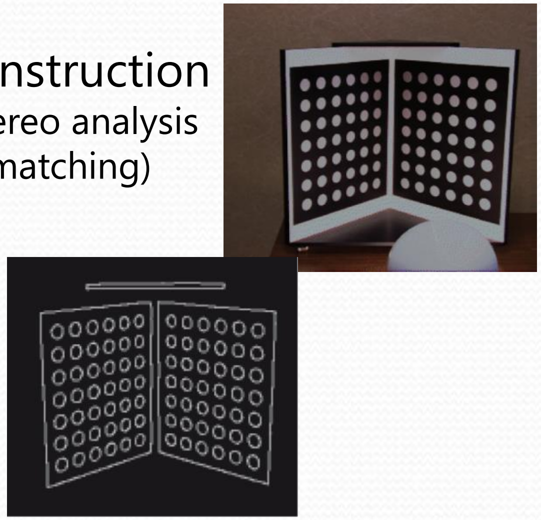
3D reconstruction
(here : stereo analysis
+ 3D matching)



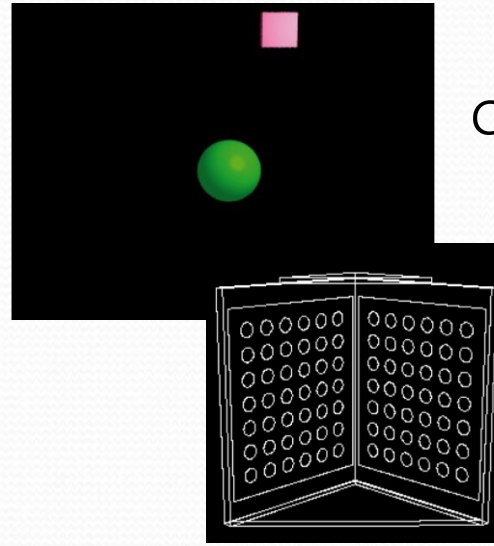


Augmentation needs 3D acquisition

3D reconstruction
(here : stereo analysis
+ 3D matching)



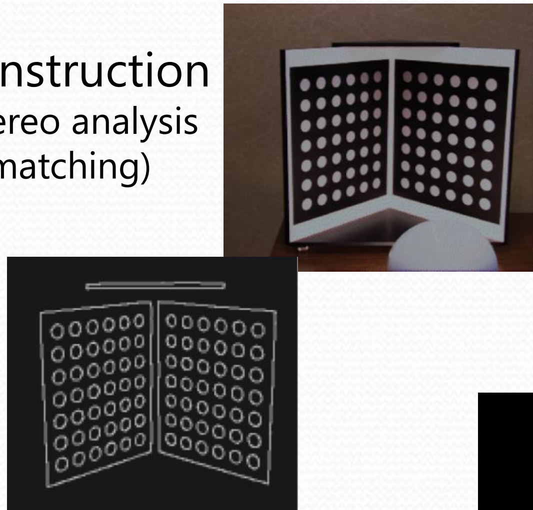
+



CG graphics

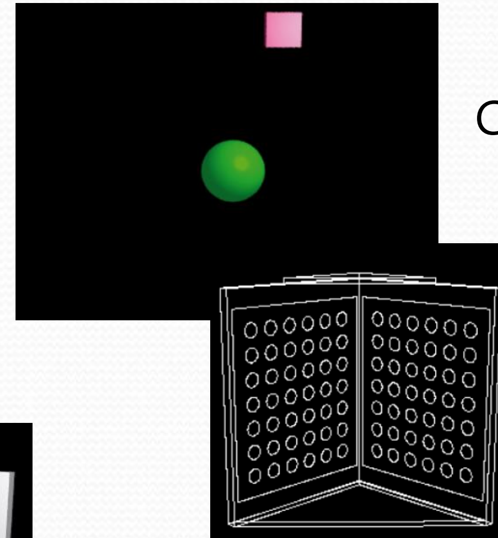
Augmentation needs 3D acquisition

3D reconstruction
(here : stereo analysis
+ 3D matching)

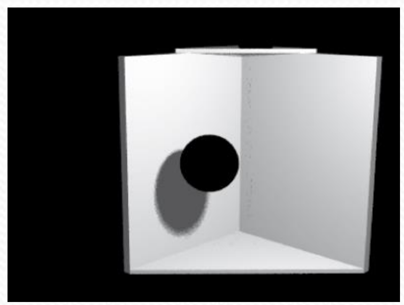


+

match move



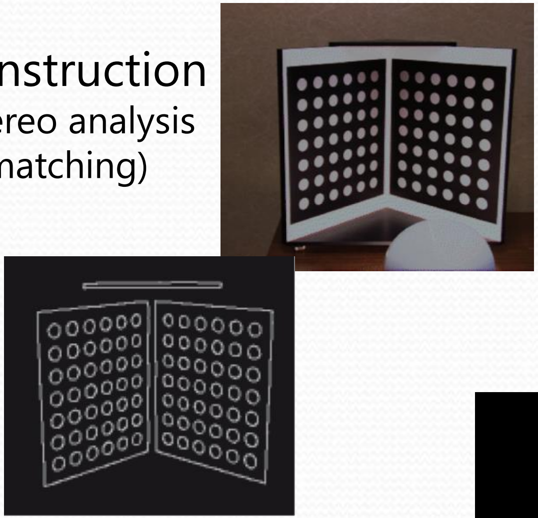
CG graphics



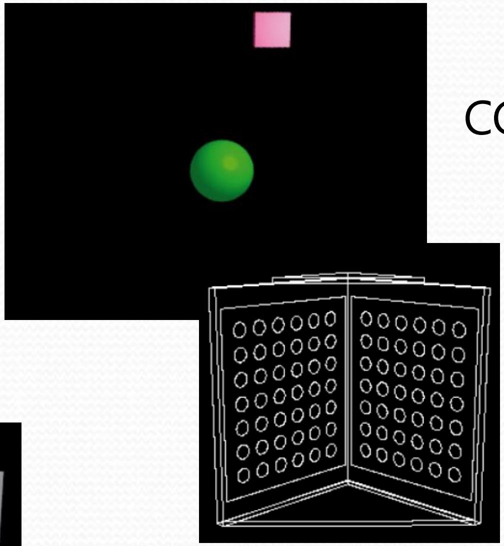
Compositing masks
(shadows, occlusions…)

Augmentation needs 3D acquisition

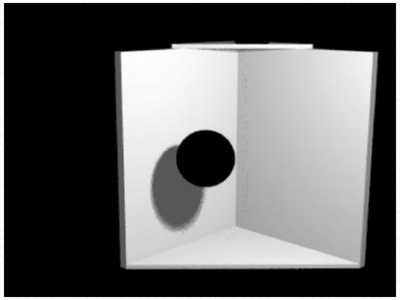
3D reconstruction
(here : stereo analysis
+ 3D matching)



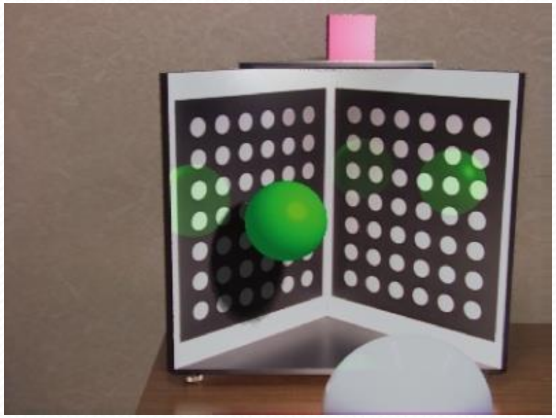
\+
match move



CG graphics



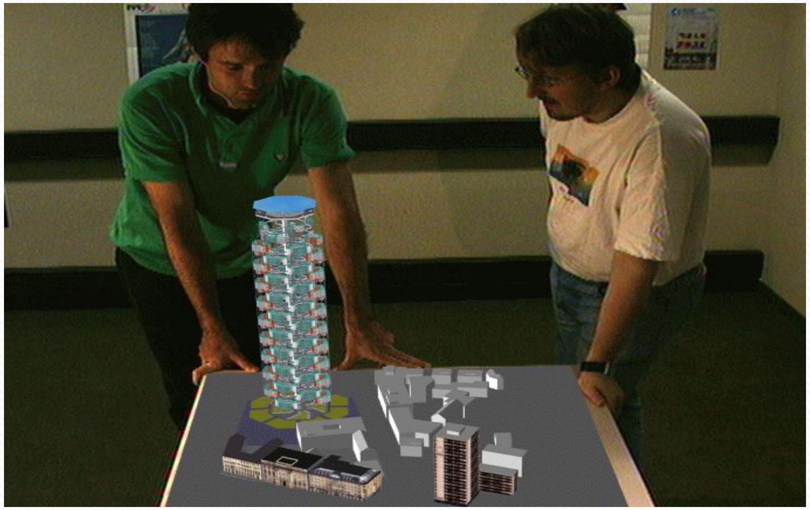
Compositing masks
(shadows, occlusions...)
(real-time *shaders*)



Augmented image

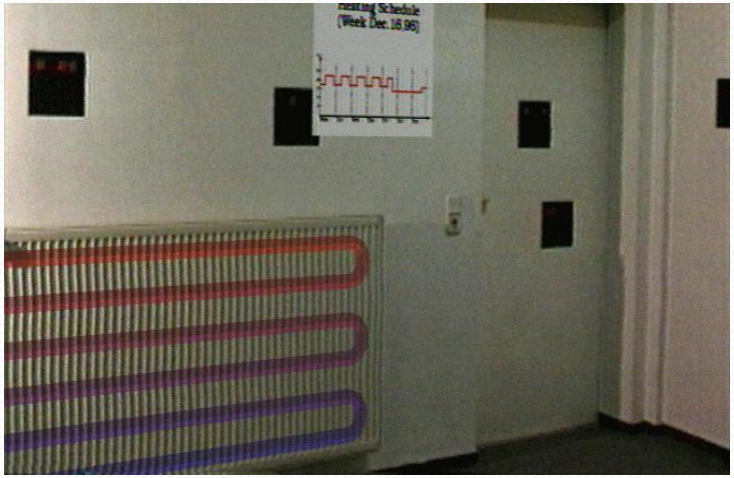
LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

Do not forget…



3D scene analysis can be simple yet effective

Augmented content can be 3D but not geometry-related



Augmented reality

Many applications:

- Architecture
- Design
- Medical
- Crisis simulation
- Marketing
- Multimedia/games
- Military …

3D Acquisition : how ?

3D vision: How ?

Sensor and techniques for 3D imaging:

- *Passive* sensors : observe « rays » from scene
  - ➢ Work everywhere, easier to transport
  - ➢ No need for fancy equipement
  - ➢ Safe for the humans
  - ➢ Historically first in use
- *Active* sensors : send « rays » on scene
  - ➢ Limited use (power supply, size …)
  - ➢ More expensive and sophisticate use
  - ➢ Safety issues

3D vision: How ?

Sensor and techniques for 3D imaging:

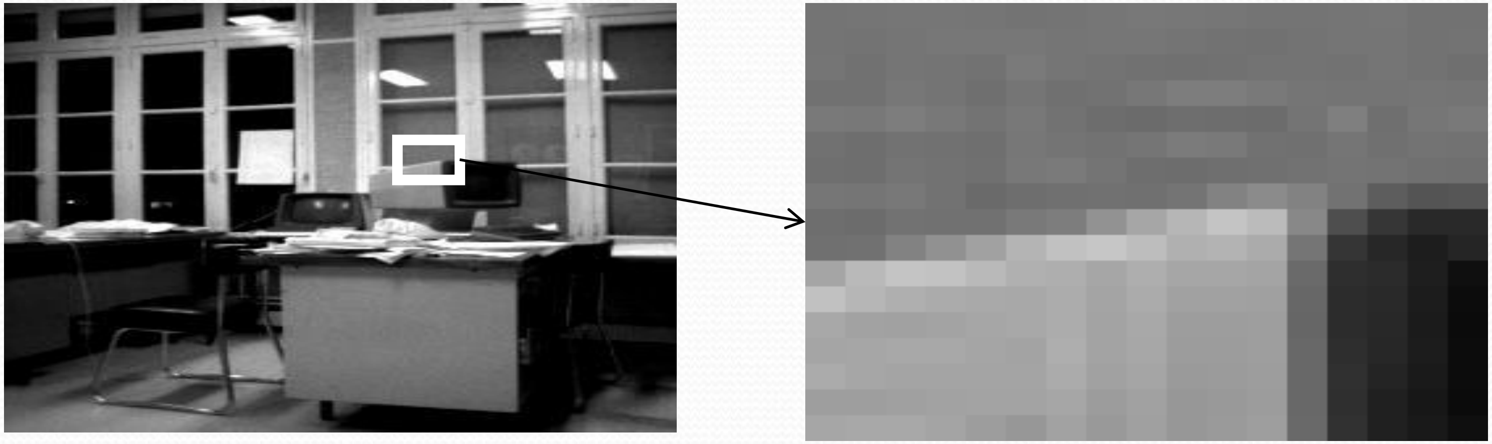
- Passive sensors
- Active sensors

Passive sensors

- Receive natural light, no active intervention

  Ex: CCD or CMOS arrays

- *Many* techniques can extract 3D information from images



Passive sensors : shape from X

The aspect of an object is a function of:

Intrinsic properties
- geometry (form, rugosity …)
- photometry (color, texture)

\+

Extrinsic properties
- Light sources
- Position of camera w.r.t. object

$\Longrightarrow$ Use extrinsic constraints to compute intrinsic properties

Passive sensors : shape from X

X = base for feature extraction in image

- X = motion
- X = shading (not shadows !)
- X = illumination
- X = texture
- X = …, combination of previous

X = motion

- matching features: *points*, *segments*, *regions*, between images.
- known camera motion → reconstruction of (X, Y, Z).

Stereovision (special case of motion)

Computing Differential Properties of 3-D Shapes from Stereoscopic Images without 3-D Models (Devernay, 1994)

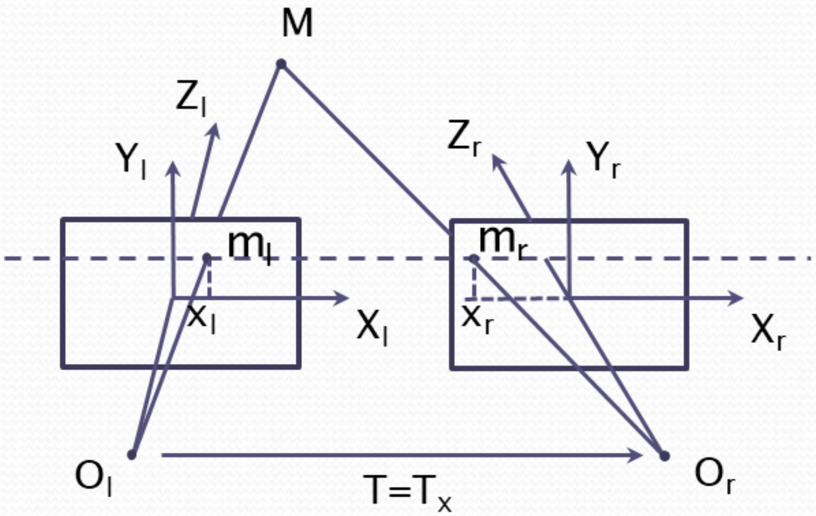
Generate dense 3D maps
(data redundancy)

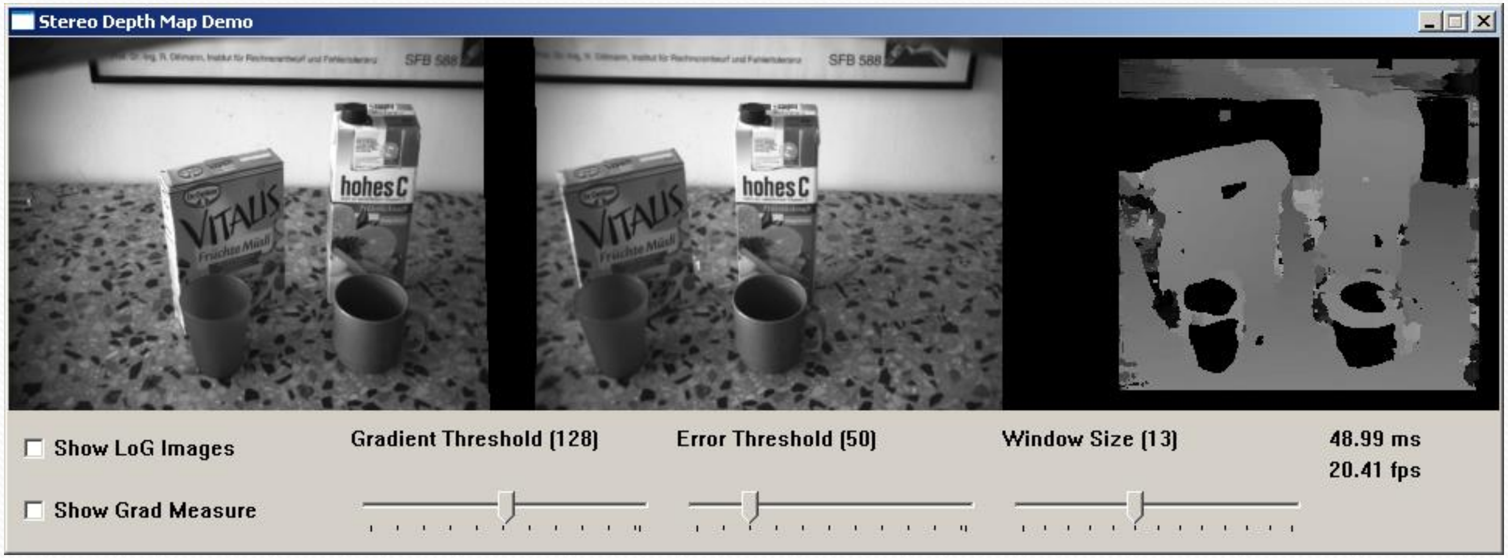
Rectified stereo (parallel cameras)

$Z.(X_l - X_r) = F\ T_x = Cte.$

Disparity : $\Delta X = .(X_l - X_r)$

Disparity map = depth map !





Infra red stereo: Leap motion

Gesture capture device

- IR lighting (3 LEDS)
- Fast camera (150 fps)
- volume is 600 x 600 x 600 mm
- stated 0.01mm depth resolution
- precision < 0.2 mm static
- 1.2 mm for dynamic setups
- Real-time stereo computation for fingers: all is in the software
- Patents pending !

1/10-inch High End VGA, CMOS Camera 640H x 480V. 16-bit wide data (2 x 8 bits interleaved images)

Tremor amplitude varies between 0.4 mm ± 0.2 mm

Stereovision

Many advantages:

- Rather simple, accessible, many existing software solutions

- Robust w.r.t. outside conditions

- Dense 3D maps possible

- Real-time computation possible (GPU, FPGA)

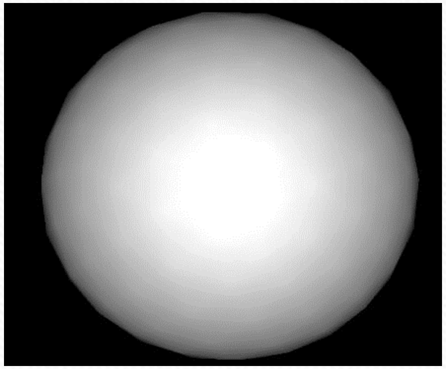
Stereovision

Limitations, constraints:

- need to match primitives (points, segments, regions)
- calibrate stereoscopic cameras… or buy specific hardware
- resulting 3D model is weakly structured (cloud of 3D points)

X = shading

- Compute surface orientation from *single* image, with *hypothesis* on:
  - Surface nature (homogeneous + type)
  - Illumination (<u>Ex:</u> point source at infinity)
- Approach introduced at the beginning of XXth century

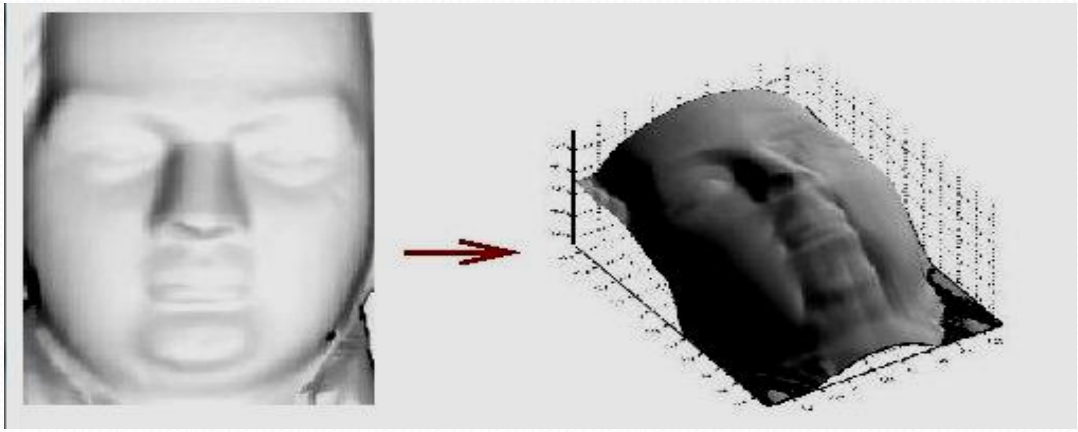


Opposition Surge: moon is made of Regolith

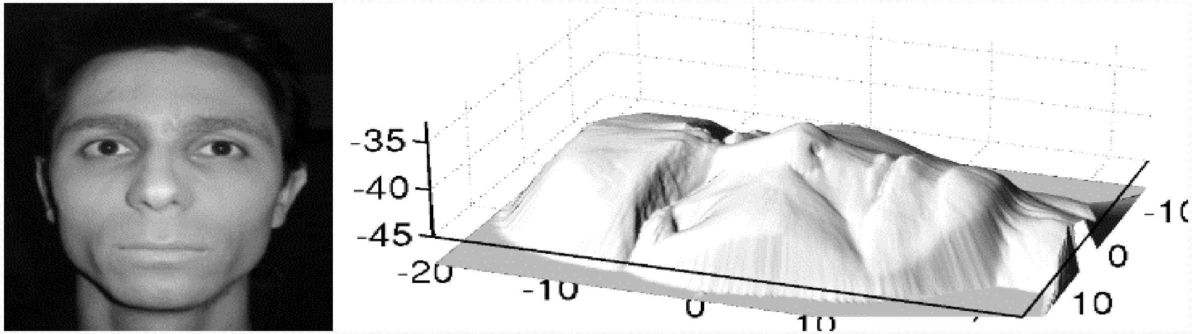
Full moon is 12x brighter than half crescent
(should be x 2)

→ *Hapke parameters*

Shape from shading : results



INRIAlpes, 2006



Shape from shading : results

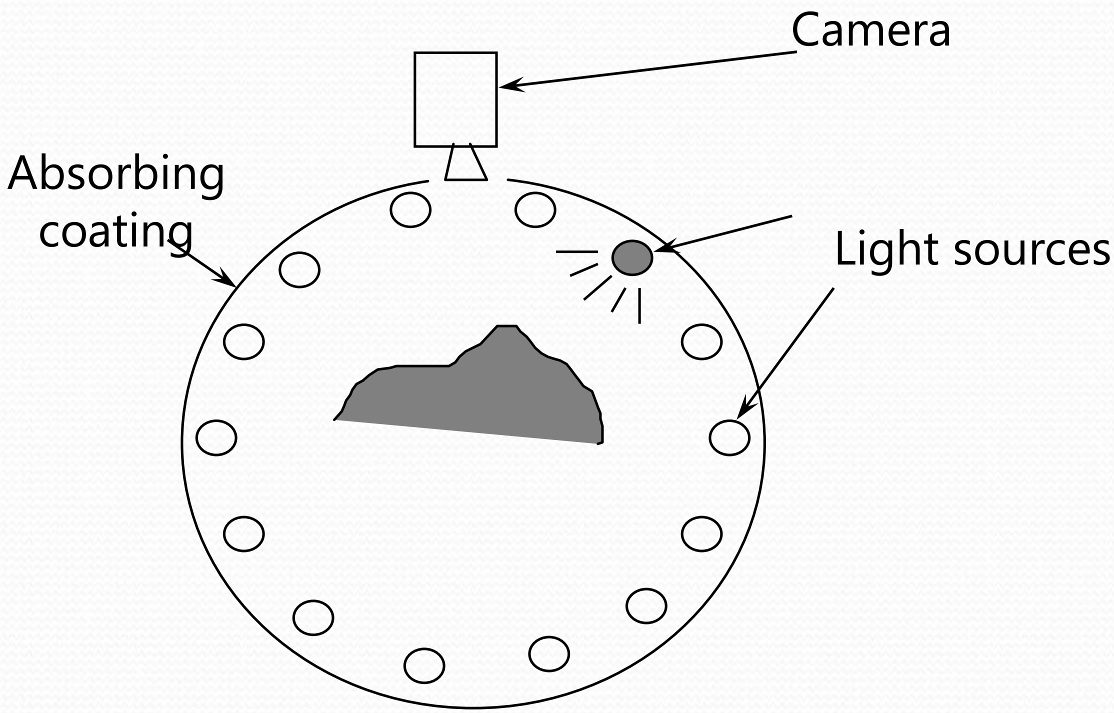
Many assumptions:

- One homogeneous surface
- Very sensitive to illumination
- Orientation only

Well adapted to platenary or aerial observation

Shape from lighting

- The aspect of a surface varies a lot as a function of illumination



Shape from lighting

- Hypotheses: surface is continuous (Z=f(x,y)), with *uniform material*.
- Each light produces one constraint I(x,y) per point of surface Z=f(x,y). Enough constraints → problem solved.

- Ex: *Lambertian* surface:

$$\text{minimize} \quad E = \sum_{i=1}^{M} \left[I_i - A\cos(\theta_i - \theta_n) \right]^2$$

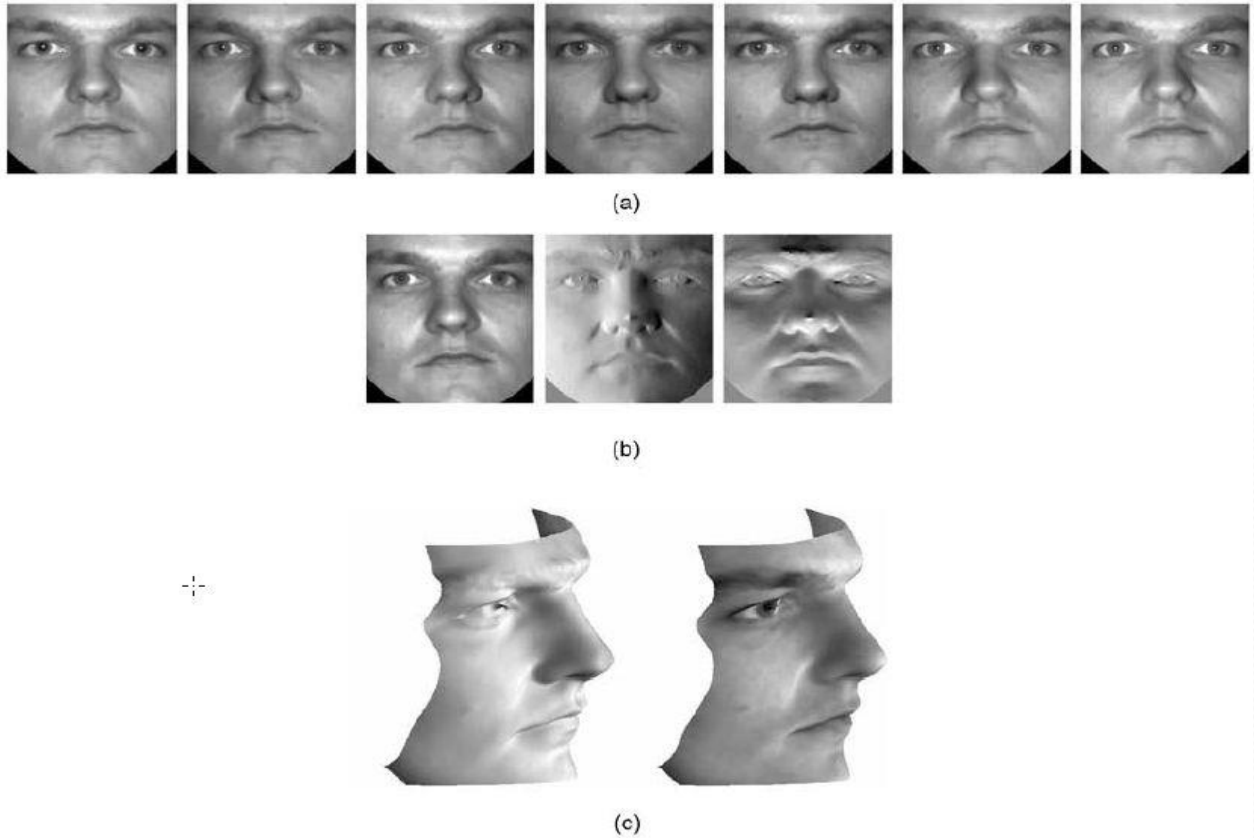
$$\theta_i = \text{position of light source i}$$

$$I_i = \text{image} \quad \text{i}$$

$$0 < \text{A} < 1 = \text{reflexion factor}$$

$$\theta_n = \text{surface orientation}$$

Shape from lighting: results



(a)

(b)

(c)

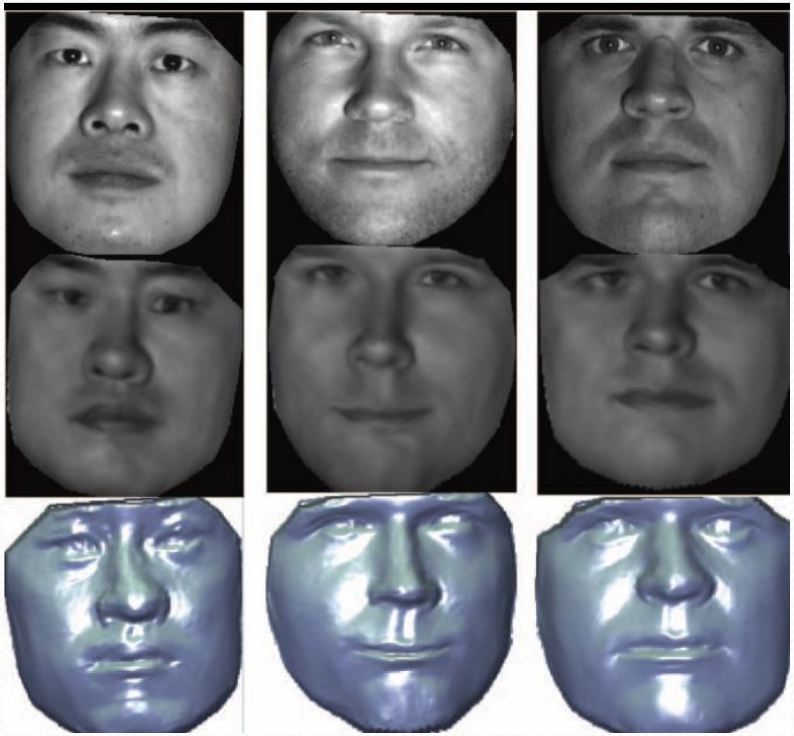
Georghiades (2001)

Shape from lighting

Problems:

- Too many assumptions
- Contrained environment
- Qualitative results

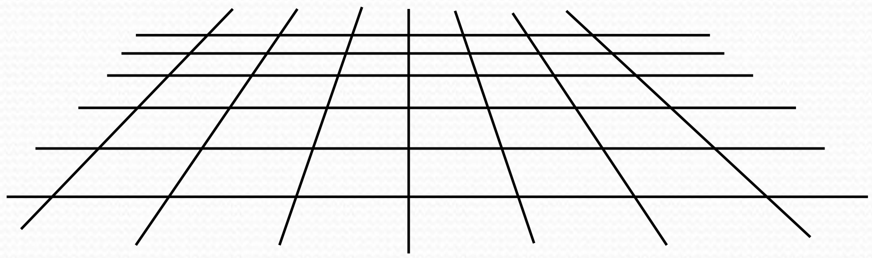
Statistical SFS

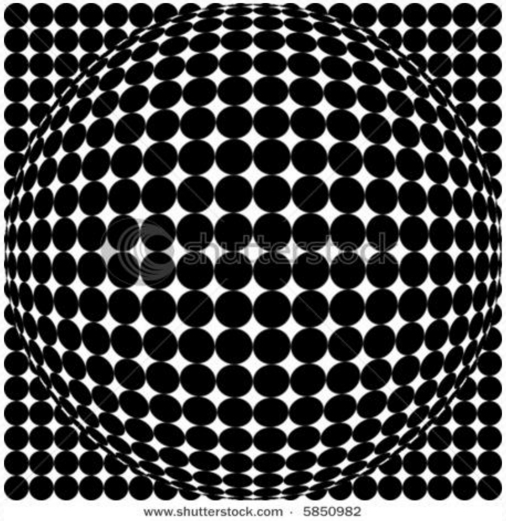


2010: “3D face recovery from intensities of general and unknown lighting using partial least squares”
Ham Rara, Shireen Elhabian, Thomas Starr, Aly Farag
CVIP Laboratory, University of Louisville (USA)

X = texture

- Surface with regular (repetitive) patterns





pattern deformation = f (depth)

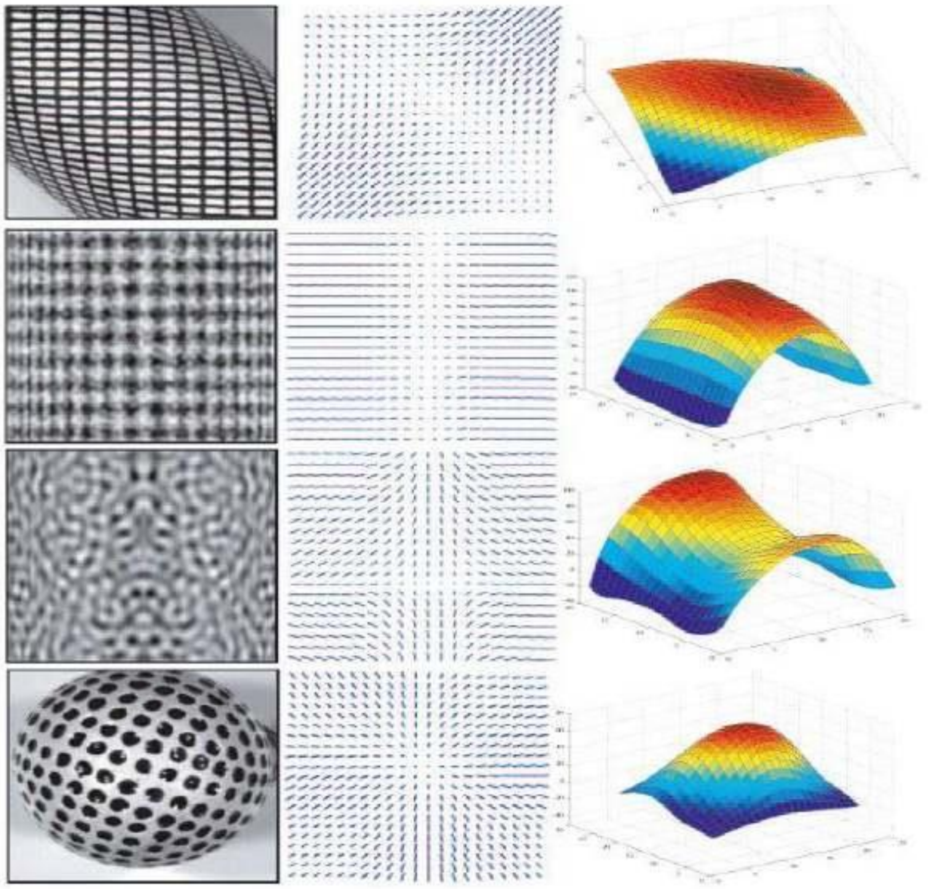
Shape from texture

Two steps:

- Image analysis: Compute a transform capturing deformation. Ex: Fourier transform, 2D momentums, etc.
- Convert measure into depth with regularity assumption

Combined with stereo: stereo from texture

Shape from texture: results



B. Super and A. Bovik. (1995)

Shape from texture

Limitations:

- Single surface, single homogeneous texture.
- Sensitive to illumination conditions.
- Use with caution

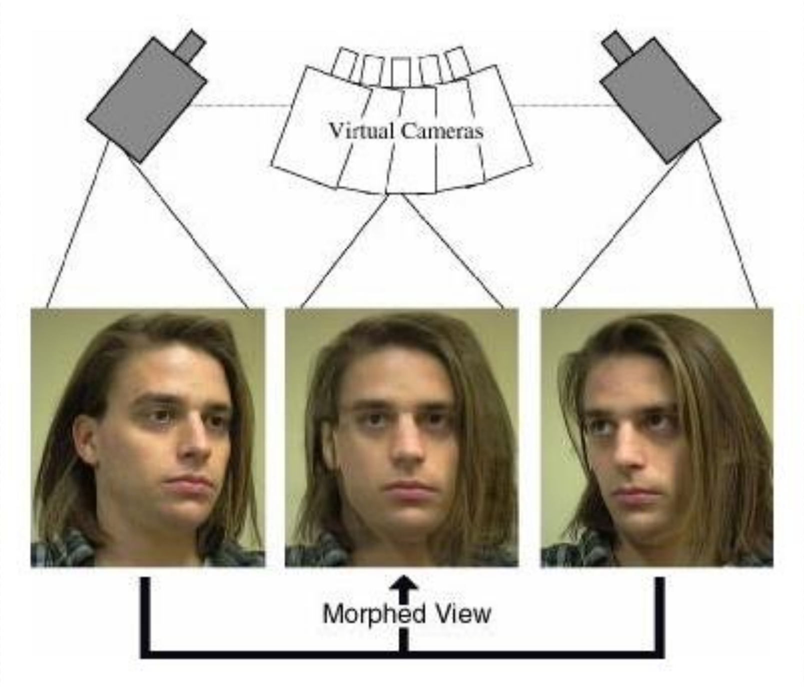
Implicit reconstruction

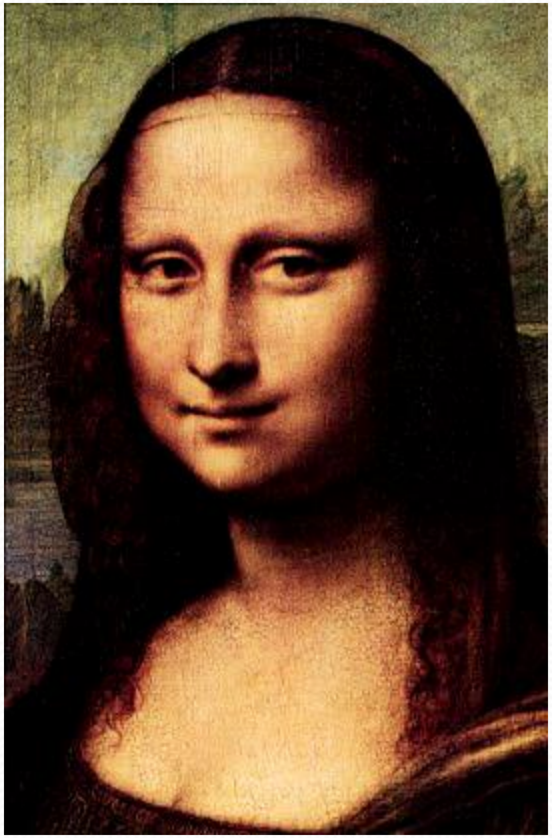
- It is not mandatory to compute 3D to see 3D objets (depend on application). Ex: movie industry.
- Is it possible to generate «intermediary» images between 2 reference frames ?

= Estimation of equivalent motion for all pixels in « virtual » view.

Knowing $P_1(x,y)$ et $P_2(x,y)$, can one predict $P_3(x,y)$, assuming the position Cam_3 relative to Cam_1 et Cam_2 is known ?

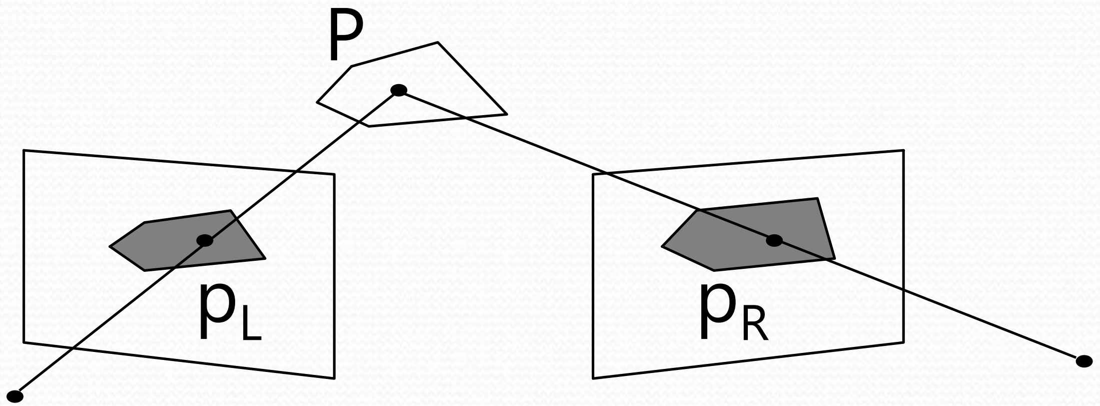
Implicit image interpolation





Implicit 3D surface reconstruction

- Ex: for a plane, projections are linked by *homography* :

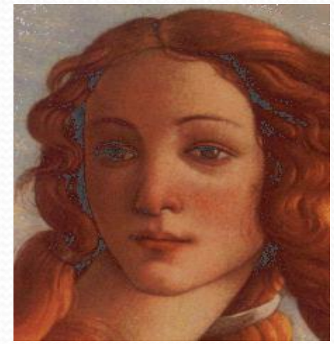


$$p_R = H\, p_L \;\; with \;\; \mathrm{H} = \begin{pmatrix} . & . & . \\ . & . & . \\ . & . & . \end{pmatrix} \;\; and \;\; \mathrm{p} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

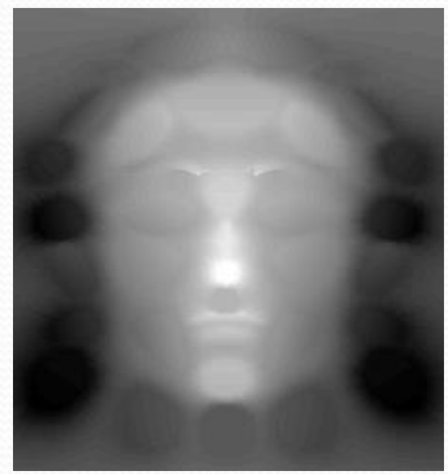
Estimation of H : matching features between images (points, regions…)

3D Interpolation : example

Mirror original painting



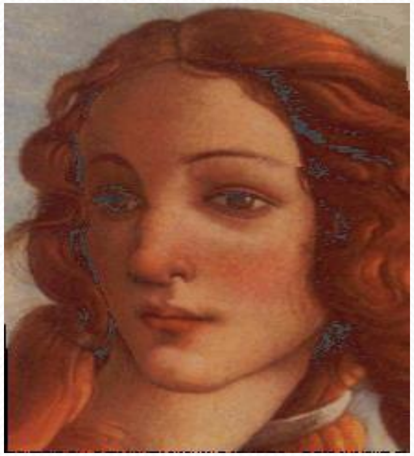
Region map (hand drawn)



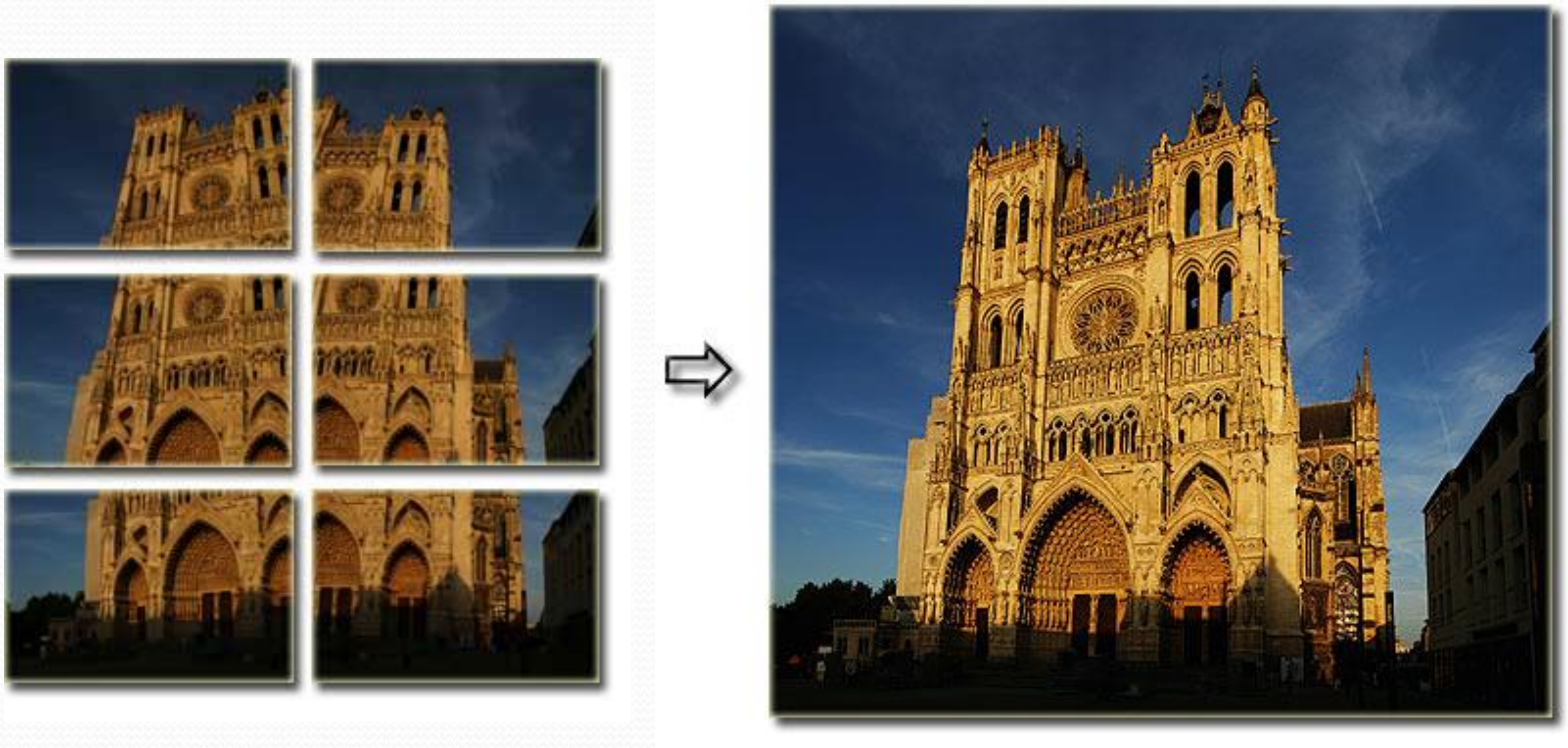
Depth map

Global 3D Planar Reconstruction with Uncalibrated Cameras and Rectified Stereo Geometry Jean-Philippe Tarel (1999).

Augmented image acquisition: result



Augmented image acquisition: stitching



Hugin : http://hugin.sourceforge.net/

3D vision: How ?

Sensor and techniques for 3D imaging:

- Passive sensors

- Active sensors

Active sensors

- Generate a signal to analyze 3D environment, usually light:
  - Structured in geometry
  - Using a coherent source (laser)

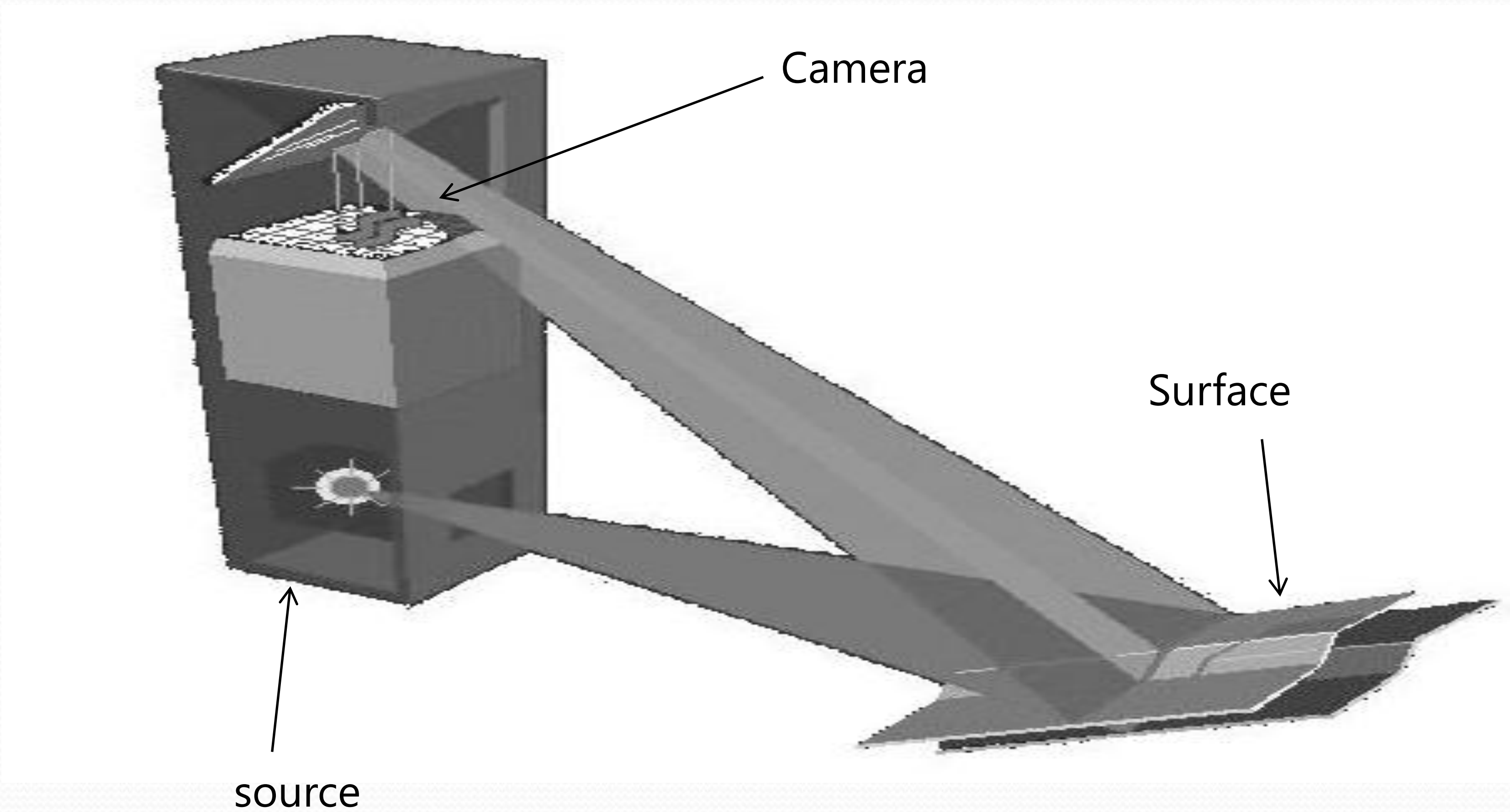
- Influence environment: application domain limited

Non optical active sensors

- Many sensors use waves to penetrate matter and analyze 3D content
  - ✓ Echography (sounds)
  - ✓ Radiography, tomography, scanner (X rays),
  - ✓ Nuclear Magnetic Resonance (alpha rays),
  - ✓ Etc.
- Very expensive, highly specialized (medical)

Triangulation-based sensor

Principle:

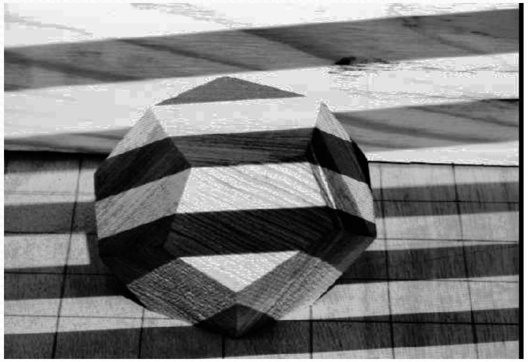


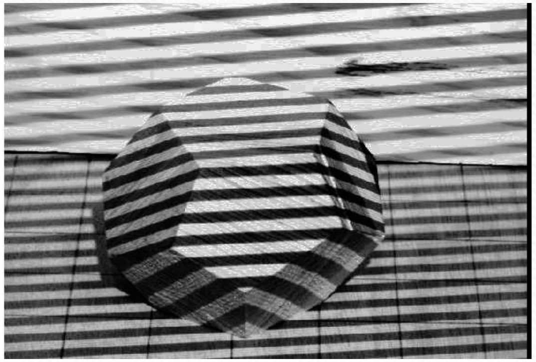
Triangulation-based sensor

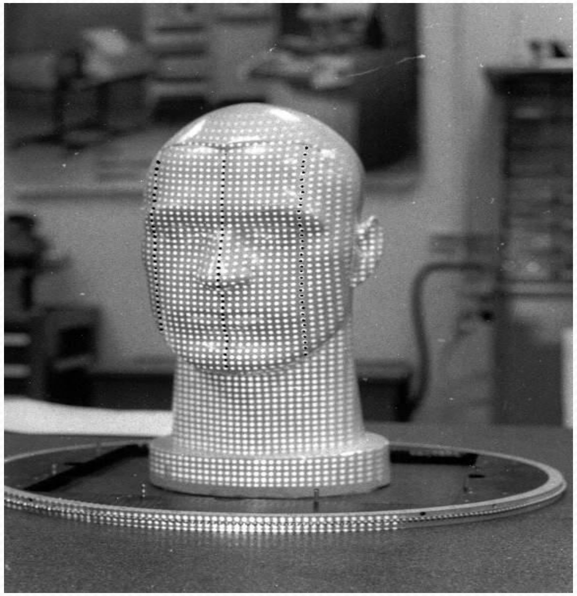
Several methods

- «natural » light , with several illuminating patterns
- coherent light: laser stripe created with cylindrical lens

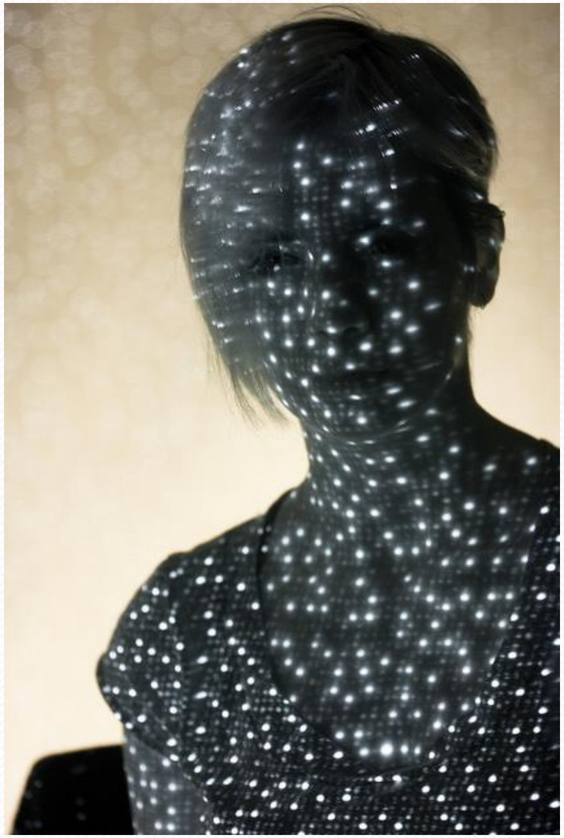
Structured lighting: results







Structured lighting: Kinect 1

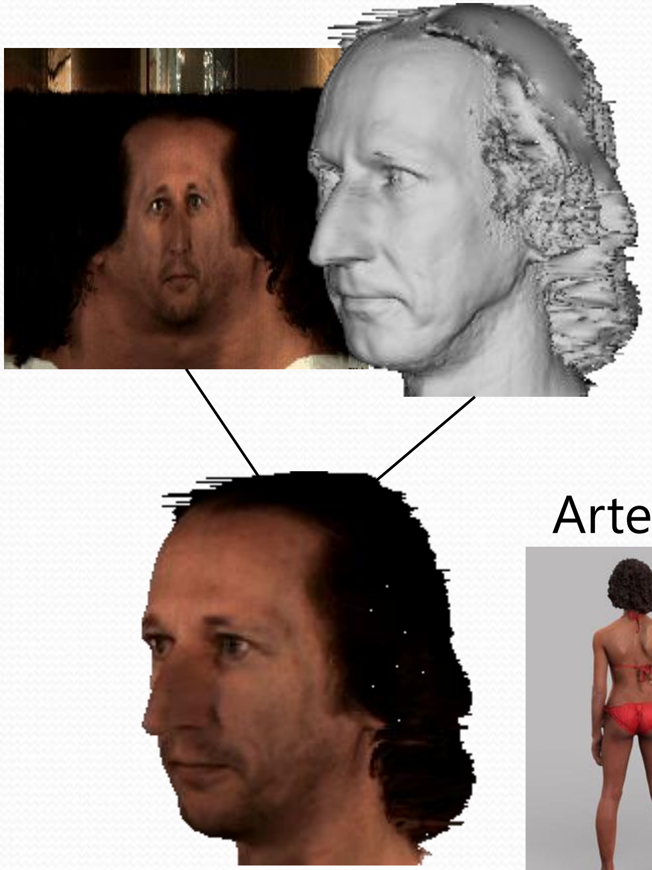


Kinect 1, 2010

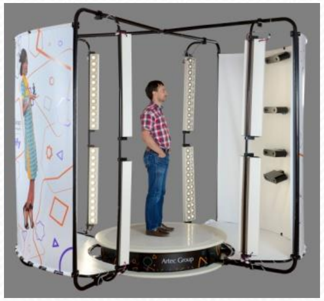
"Augmented Reality Magic Mirror using the Kinect "
Tobias Blum

Triangulation scanner : cyberware

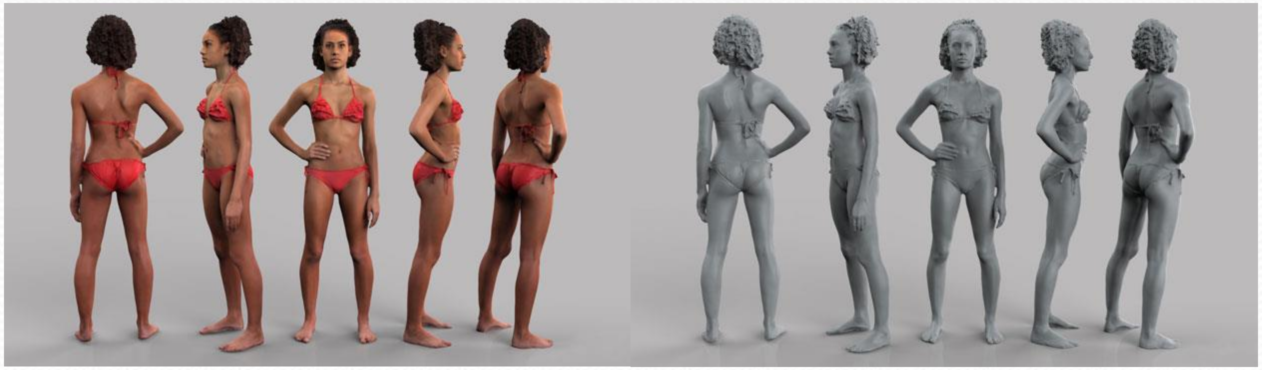
Cyberware (1998)



Laser sweep



Artec (2014): 12 sec. Scan.



LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

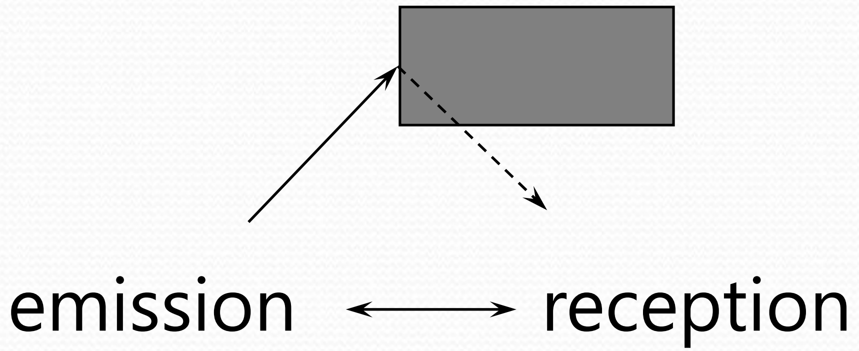
Triangulation sensors: problems

- Acquisition time: a few seconds (no real-time)
- Controlled environment: low ambient light, no scattering (metals), avoid blue colors.
- Distance: few centimeters to few meters
- Cost (100 k€)

Triangulation sensors: problems

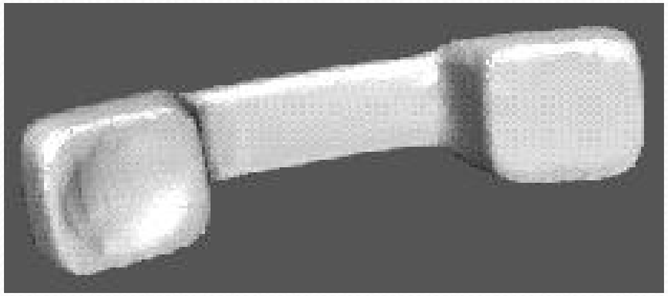
As the separation between emitter and receptor increases:

- Better precision
- Increasing chance of occlusions



Difficult compromise

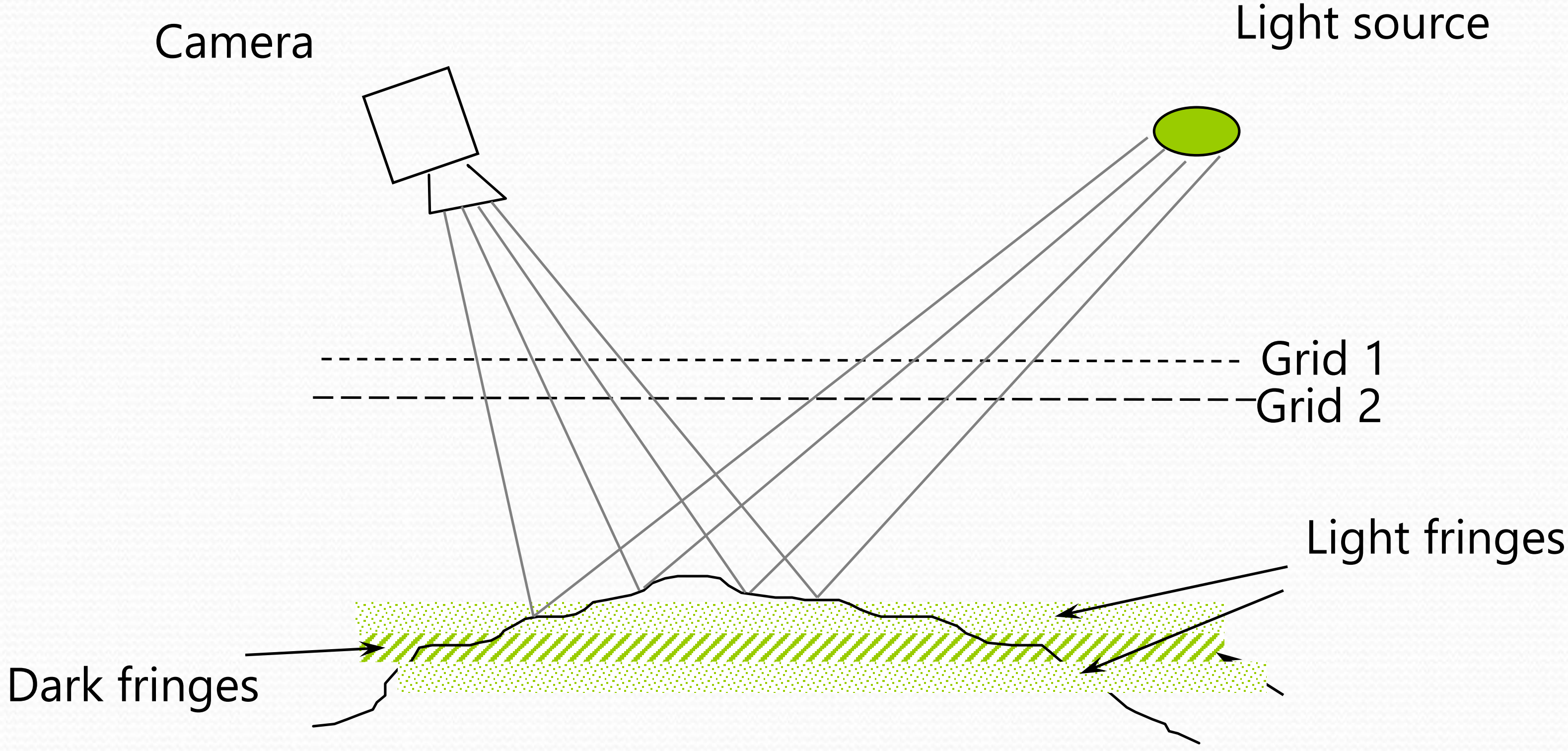
Triangulation: results



Triangulation: results

Integration of multiple scans: several million polygons

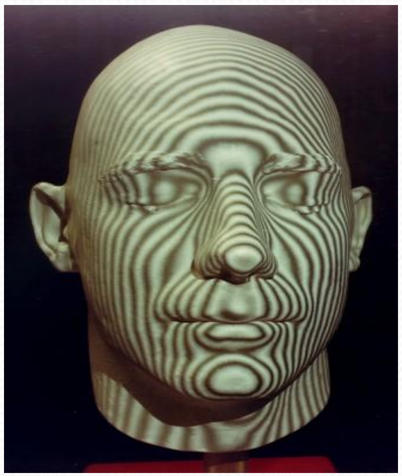
Moiré effect



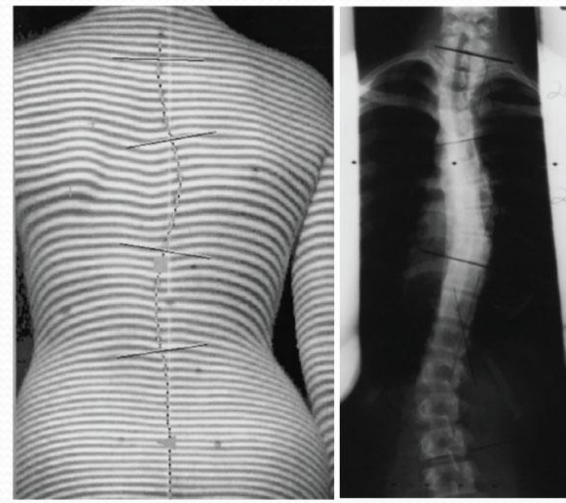
Moiré effect (Cont'd)

- Problems:
  - Small depth of field
  - Controlled environment
  - Sophisticated image analysis (not real-time)
  - No absolute Z: ambiguity.

Moiré : results



Face scan

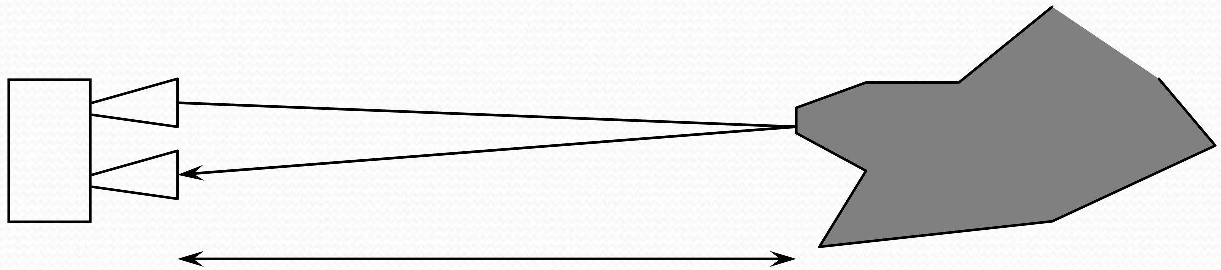


Posture analysis

Historical review and experience with the use of surface topographic systems in children with idiopathic scoliosis
XC Liu, JG Thometz, JC Tassone, LC Paulsen (2013)

LIDAR

- Principle (radar):



- Wave: infra-red laser with low amplitude modulated signal.
- Swipping with mirror

$$\lambda < 1m \Longrightarrow f > 300Mhz$$

LIDAR (Cont'd)

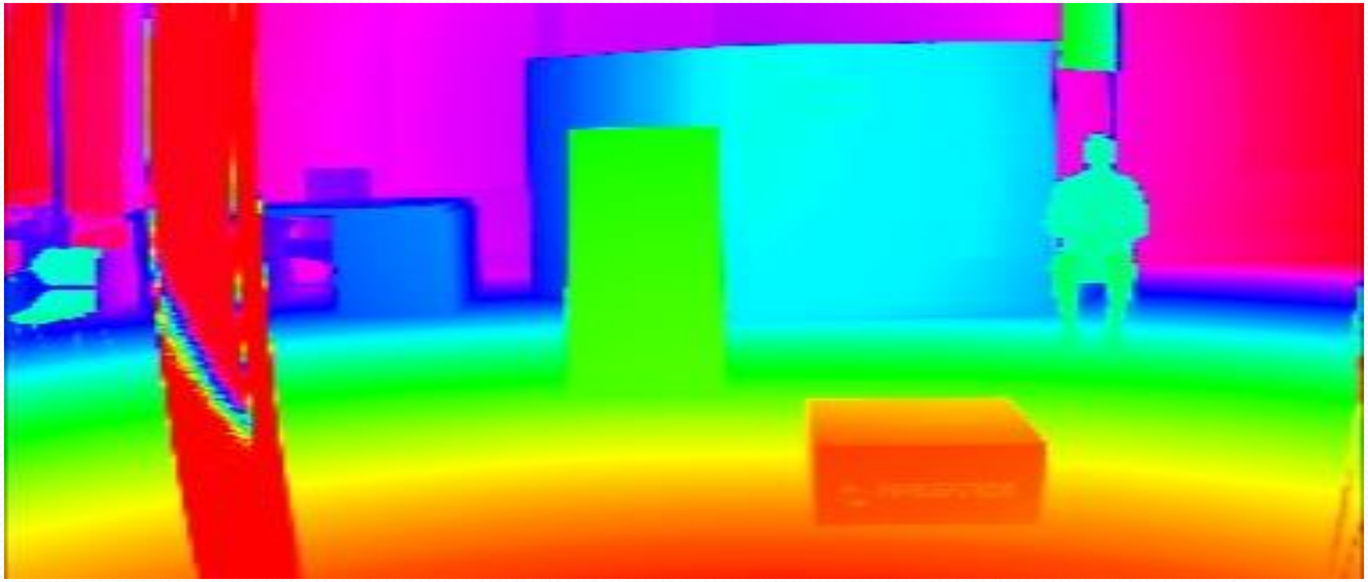
Ex.: **LADAR** from **Perceptron** Inc. (1993)

http://www.perceptron.com

- Laser diode @ 835 nm

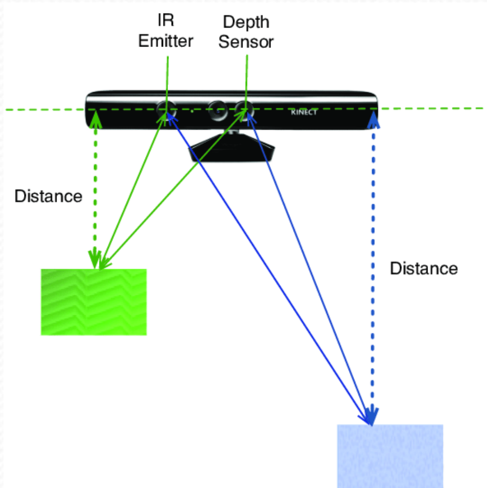
- 50 mW power
- Distance: 3 to 100 m.
- Viewing angle
  - horizontal: from 15 to 60 °
  - vertical: from 3 to 72 °
- Resolution: 1000 x 2000 pixels

LIDAR (Cont'd)

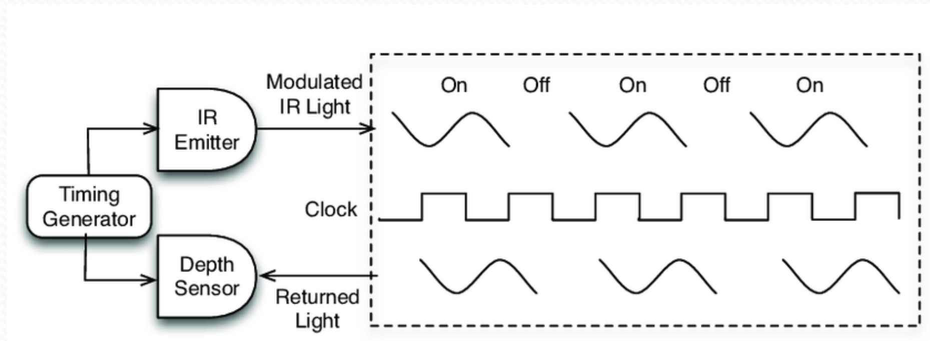


Price tag: 100k$!

Kinect2 (2014)



Modulated IR → measure phase shift in return signal



All this for 80 € !

Lun&AI: Survey of Applications and Human Motion Recognition with Microsoft Kinect International Journal of Pattern Recognition and Artificial Intelligence

Kinect Fusion

Kinect designed for 3D game interaction, but can do 3D reconstruction in real time !

Multiple Kinects

Goal: avoid sensor motion (best for room setup)

www.brekel.com

Smartphones !

Samsung S20

(but not after !)

Range : 3 meters
indirect time-of-flight (i-ToF) sensor
Modulated IR → measure phase shift
Single illuminator
Too costly and not user effective

iPhone 12 Pro

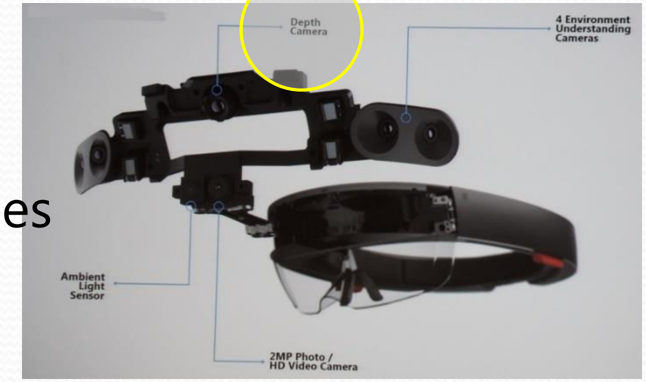
(and later)

Range : 6 meters
Direct ToF (DToF)
sends impulses on multiple points
Exclusive licence with manufacturer (Sony)

https://www.youtube.com/watch?v=FOxxqVzDaaA&feature=emb_rel_end

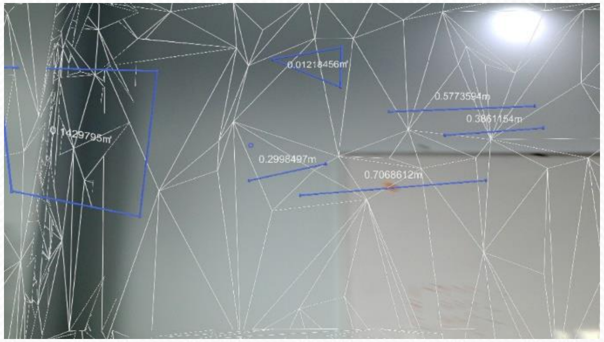
Lidar Lighting : Hololens

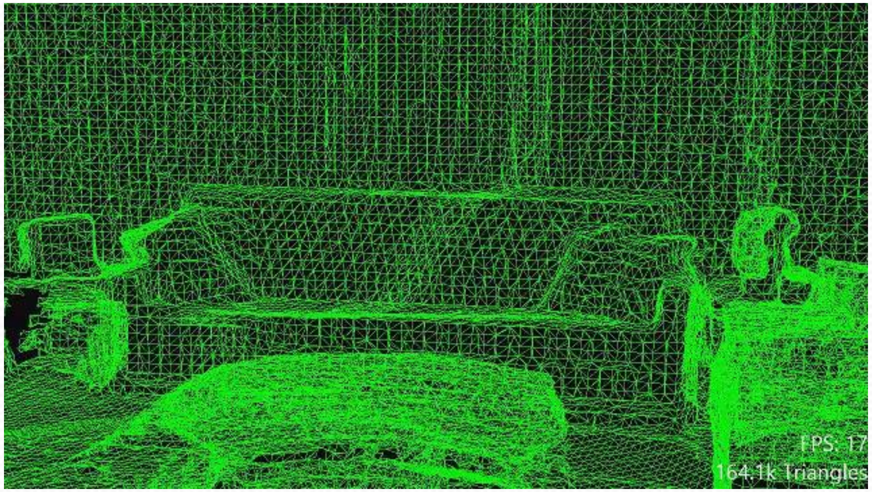
Hololens (2016, 3000$)



- Augmented reality « see-through » glasses
- Markerless 3D scanner with :
  - Depth camera (Kinect2-like) → Spatial Mapping

SLAM = Simultaneous Localization and Mapping = markerless tracking !





3D Acquisition and Tangible Interfaces

Tangible interfaces

Tangible user interface (TUI): an 3D object used as an interface with the digital world (i.e. your computer running a simulation)

SandScape device installed in the Children's Creativity Museum in San Francisco

Need **tracking** to locate **interactor** w.r.t. the **reconstructed** object

= Adaptive rendering from the interactor's point of view

https://www.youtube.com/watch?v=Q3elMIRCYSk

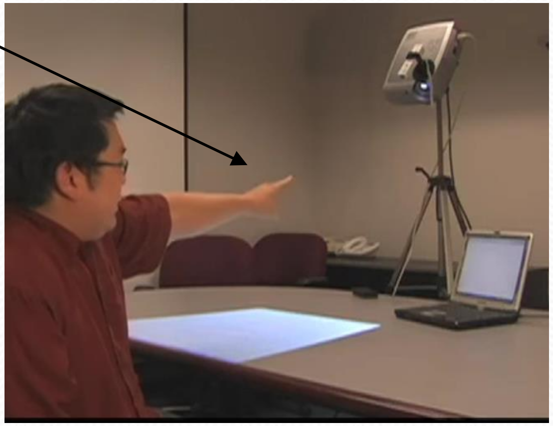
Tangible interfaces: Interactive whiteboard with the Wii

Johnny Lee's video (dec. 2007)

https://www.youtube.com/watch?v=5s5EvhHy7eQ

Reconstruction for TUI : Computation flow

1. *Tracking* = Get the interactor position w.r.t. the rendering surface



Reconstruction for TUI : Computation flow

1. *Tracking* = Get the interactor position w.r.t. the rendering surface

Reconstruction for TUI : Computation flow

1. *Tracking* = Get the interactor position w.r.t. the rendering surface

→ Use the projector / camera as the reference frame

Reconstruction for TUI : Computation flow

1. *Tracking* = Get the interactor position w.r.t. the rendering surface

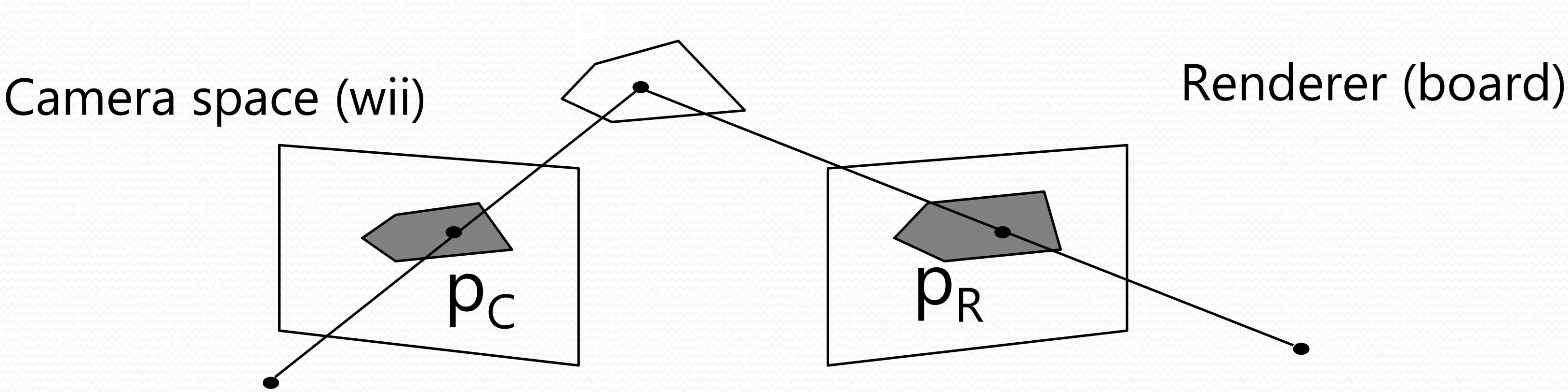
→ Use the projector / camera as the reference frame

2. Drawing on a surface: no actual 3D reconstruction needed = *implicit reconstruction*

i.e. 2D correspondence between image of the interactor (in camera space) and the rendering screen

Reconstruction for TUI : Computation flow

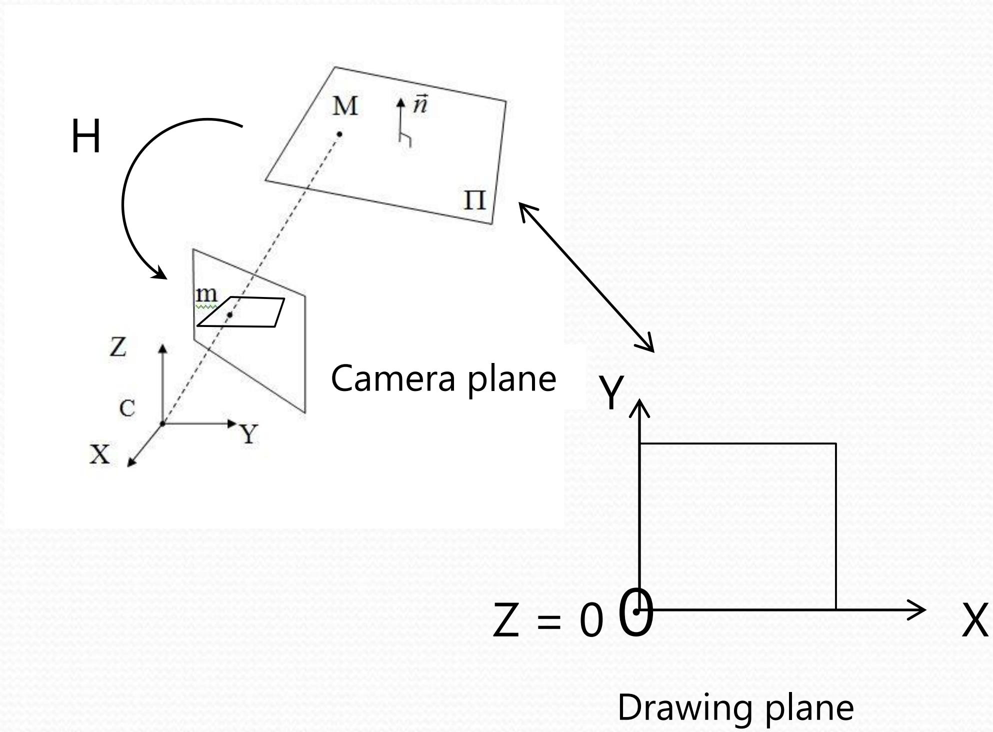
For a plane, images are linked by *homography* :



$$p_R = H\,p_C \quad with \quad \mathrm{H} = \begin{pmatrix} . & . & . \\ . & . & . \\ . & . & . \end{pmatrix} \quad and \quad \mathrm{p} = \begin{pmatrix} \mathrm{x} \\ \mathrm{y} \\ 1 \end{pmatrix}$$

Estimation of H **:** needs at least 4 points

Homography for TUI



$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

$$\tilde{m} = H.M$$

2D → 2D transform !

If we find H^{-1}, we can remap image into the original 2D space

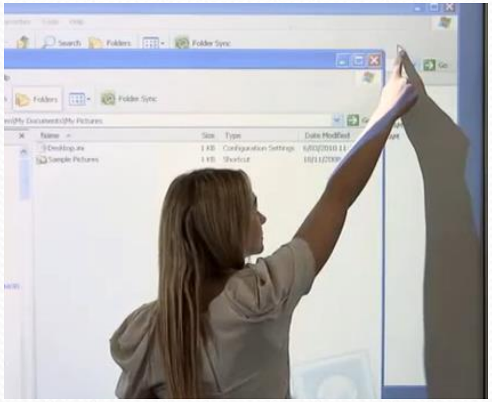
Homography for TUI

$$H = \left(R + \frac{Tn^t}{d} \right)$$

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = H \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

Points are in (X,Y) plane → Eq: $Z = Z_0$

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ 0 & h_{22} & h_{23} \\ 0 & 0 & 1 \end{pmatrix}$$



4 points calibration: drawing space corners

→ write on any surface

Reconstruction for TUI : Computation flow

2. *Render* the interaction (here writing): in the screen coordinate system

- Capture 2D position of pen in (wii) camera space – infrared image: with WiiMote SDK (C#)

```
using WiimoteLib;
Wiimote remote;
wiimotePoints[0].x = remote.WiimoteState.IRState.RawX1 / 768.0f;
wiimotePoints[0].y = remote.WiimoteState.IRState.RawY1 / 768.0f;
```

- Compute H (real-time, easy)
- H-Transform pen position in rendering space
- draw, interact

interaction for *projected reality*

3D Vision: applications

3D VISION: applications

- **Medical imaging**
- **Ergonomics / Biometry**
- **Building, CAD**
- **Inspection, control**
- **Exploration of 3D environments**
- **Assisted driving**
- **Military applications**
- **Satellite imagery**
- **Virtual Reality**
- **Multimedia content creation**

Medical imaging

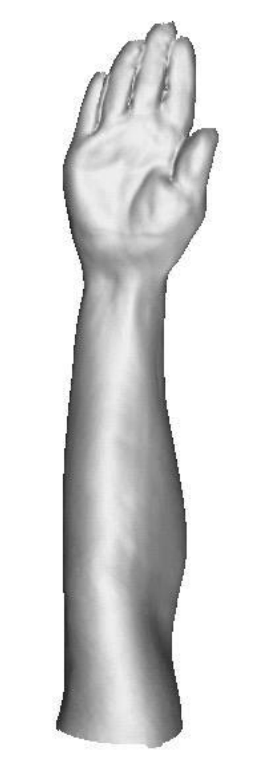
- Computer assisted diagnosis: indication of volumes, shapes, etc.

- Surgery planning
- Forensic medecine

Medical imaging

Techniques :

Moiré, stereovison, laser-based imaging (shape analysis)



Application:

- Detect deformations
- Prosthetic fitting

Design And Virtual Prototyping Of Rehabilitation Aids (1997)

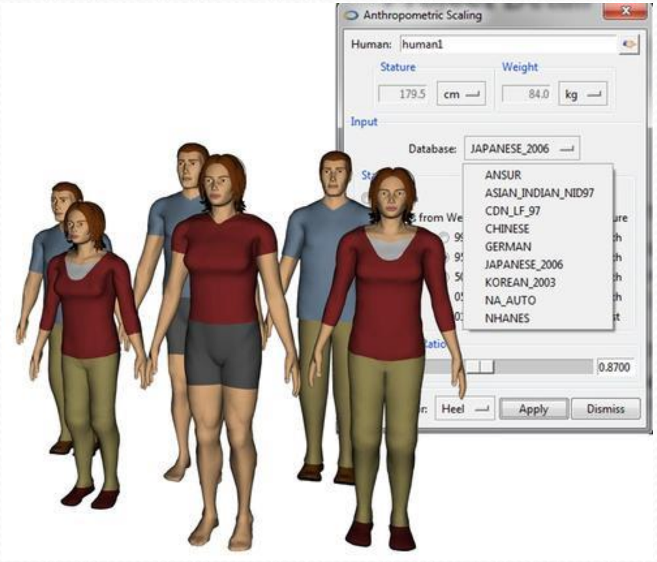
Venkat Krovi , G. K. Ananthasuresh , Jean-marc Vezien

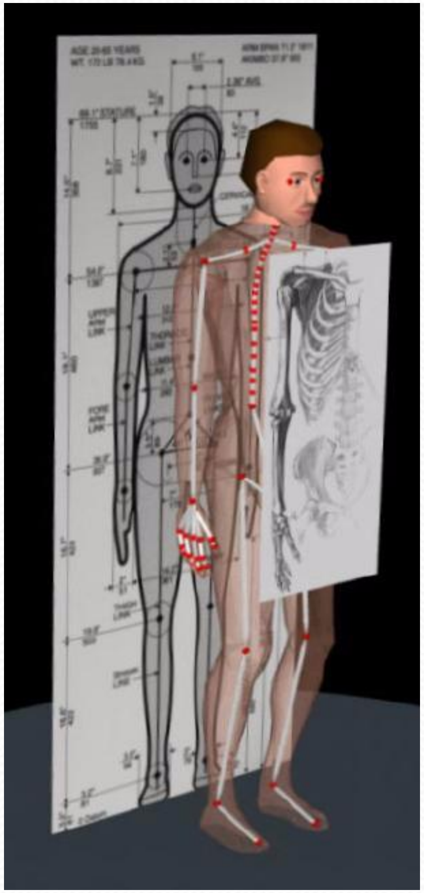
Ergonomics, anthropometric measures

- Goal: determine precise characteristics for each individual

- Objective: measure, adapt, recognize

Jack (U.penn): graphical system for human simulation





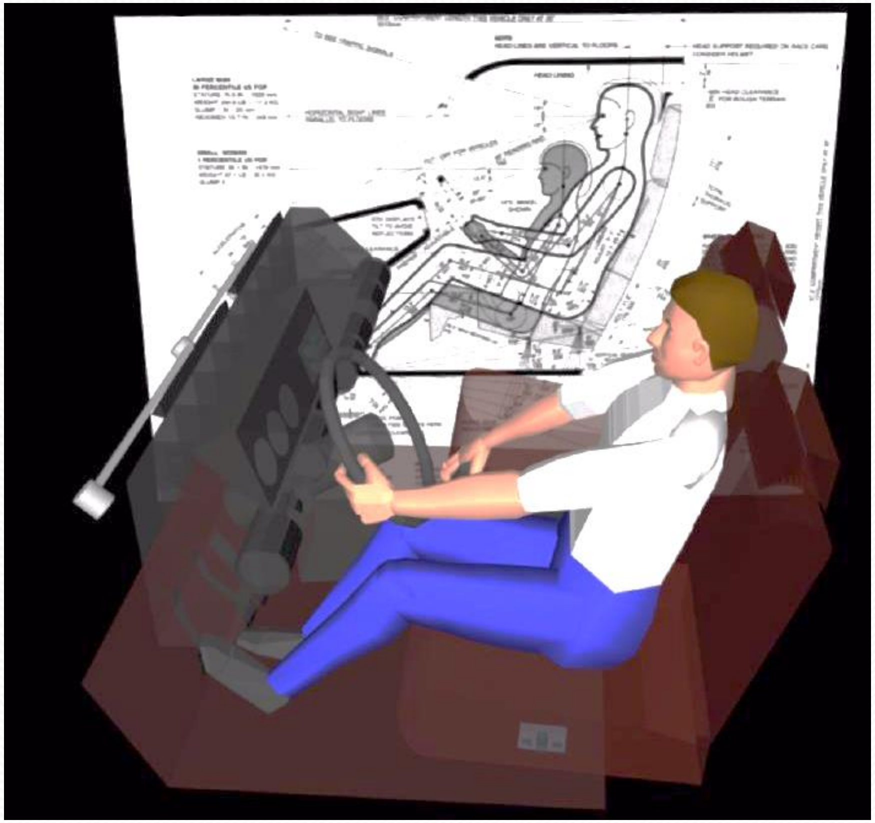
Ergonomics, anthropometric measures

Techniques:

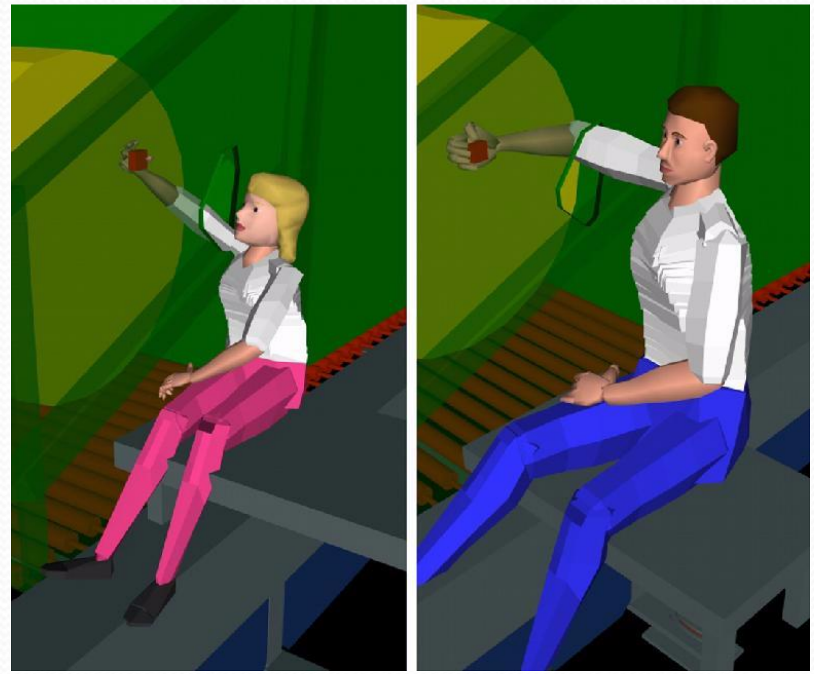
- Laser
- Stereo
- direct measurements

Application:

- Design/ergonomics

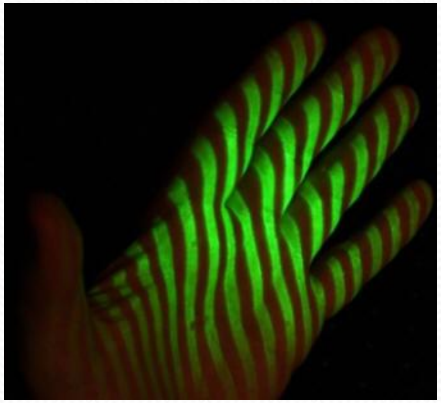


Ergonomics, anthropometric measures



Workplace design

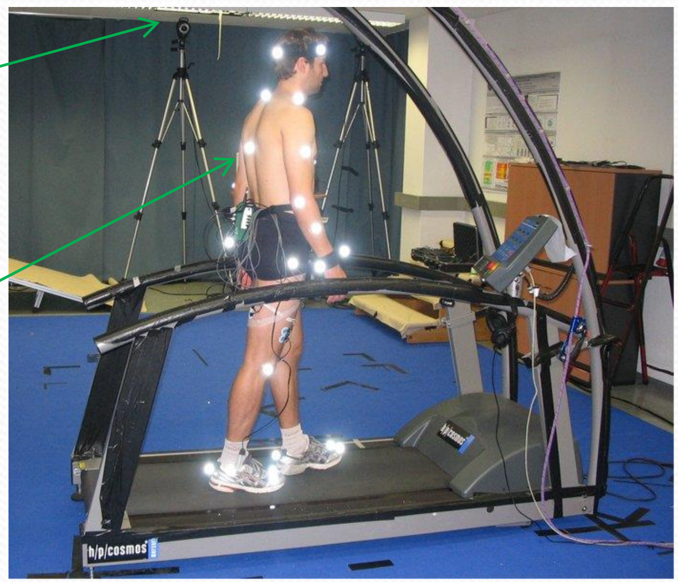
Ergonomics, anthropometric measures



3D Motion capture and analysis

Cameras

IR markers

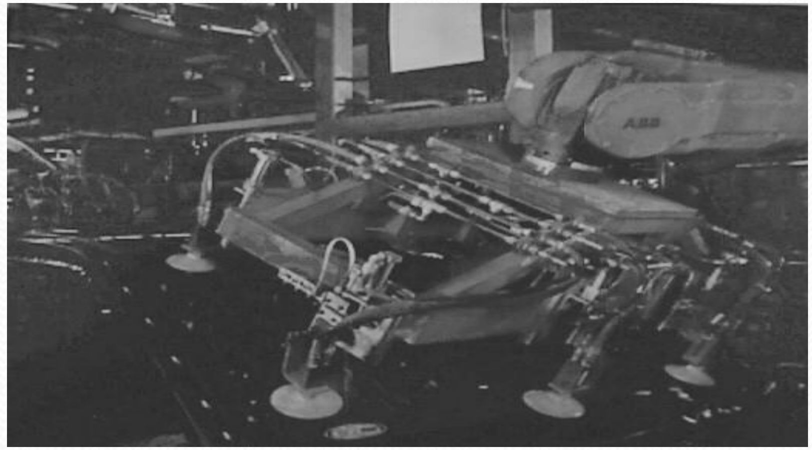


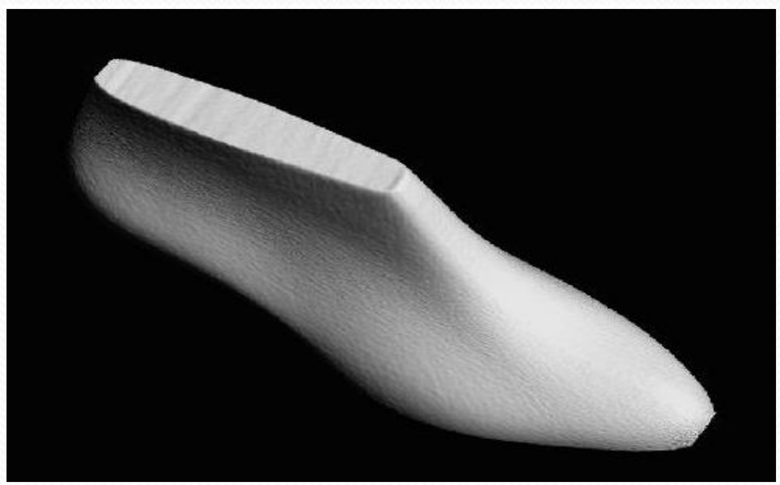
Construction, CAD

Help make computer-generated objects:

- Sensors to guide tools precisely
- Design a virtual model (design, simulations).
- Tools: laser, stereo.

Construction, CAD





Inspection, control

Check that a manufactured object is within specifications

• Micro-electronics, microchip production.
• cars (Ex: wheel alignement)
Structured light, excellent precision: 1/1000 mm !

http://www.cognitens.com/
http://www.cyberoptics.com/

Exploration

- Autonomous robots:

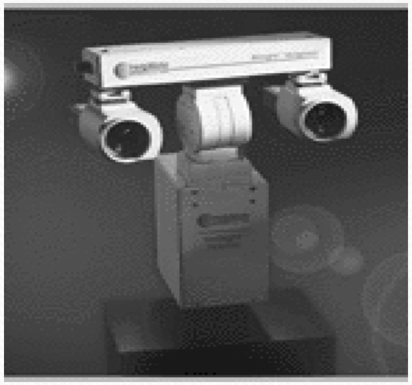
1988: K3A Navmaster
(Cybermotion, Inc.)

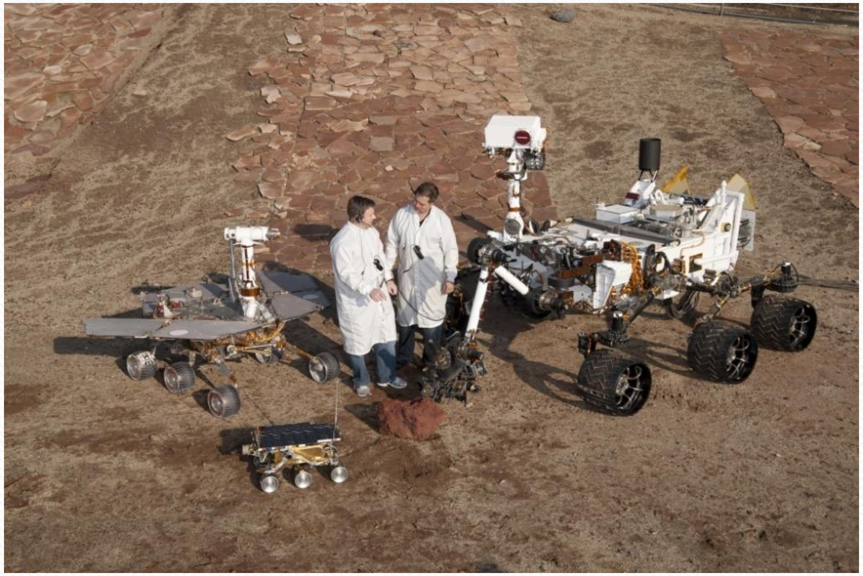
2010 : MDARS(Mobile Detection Assessment Response System) - US Army

http://www.public.navy.mil/spawar/Pacific/Robotics/Pages/Publications.aspx

Exploration

Binocular heads (stereovision maps)





Semi-autonomous robots

Exploration

Techniques: Stereo, Lidar, ultrasounds

- images: fast but not precise
- Laser (lidar): slower but more precise.

Problems:

- *fusion* of geometric data coming from multiple sources
- *calibration* of dynamic data

Exploration

Partial recontruction of the environment :

- (semi-) autonomous driving
- Limited interventions

Applications:

- Hostile environment (nuclear, chemical, explosive)
- Repetitive tasks (plants)

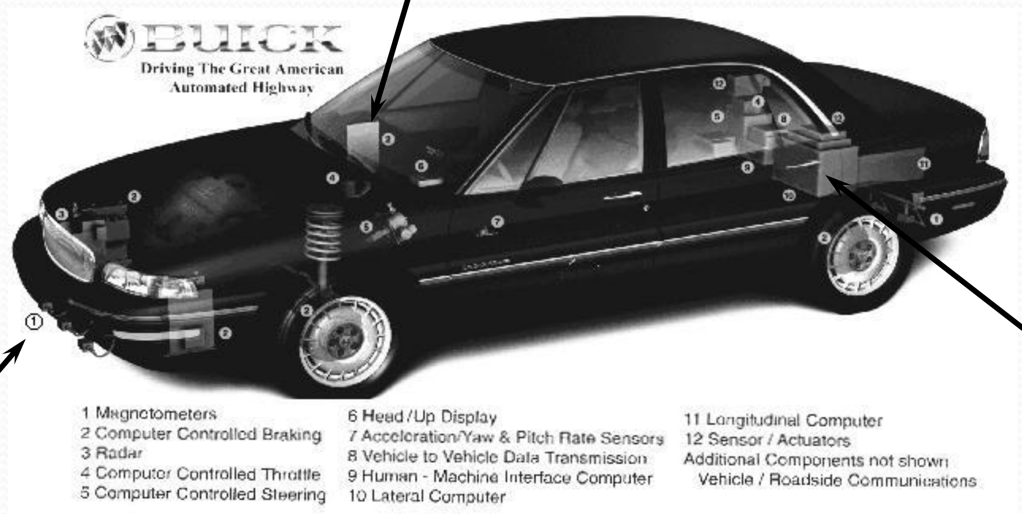
Still costly !

Assisted driving (smart car)

«intelligent» car :
environnement sensing → driver

Techniques:
- Laser, Stereo
- Ultrasound

Augmented windshield



Proximity sensor

On-board computing

Assisted driving (smart car)

Navlab project (univ. de Carnegie Mellon, USA):

- Several run-of-the-mill car tested.
- Highway testing in real conditions
- No vision (sensitivity) : lidar, magnetic beacons

DARPA Challenge : race between automated vehicules

- Tough competition
- High reward 2 M$ → high incentive
- Soon : urban driving

132-mile Mojave Desert course
7 Hours at speed = 19.1 mph

Assisted driving (smart car)

Tesla (2015)

- Model S assistive driving
- camera on top of the windshield
- GPS sensor
- 360° ultrasonic acoustic location sensors (close range)
- forward-looking Lidar (long range)
- 100 000 $
- Cloud-connected learning

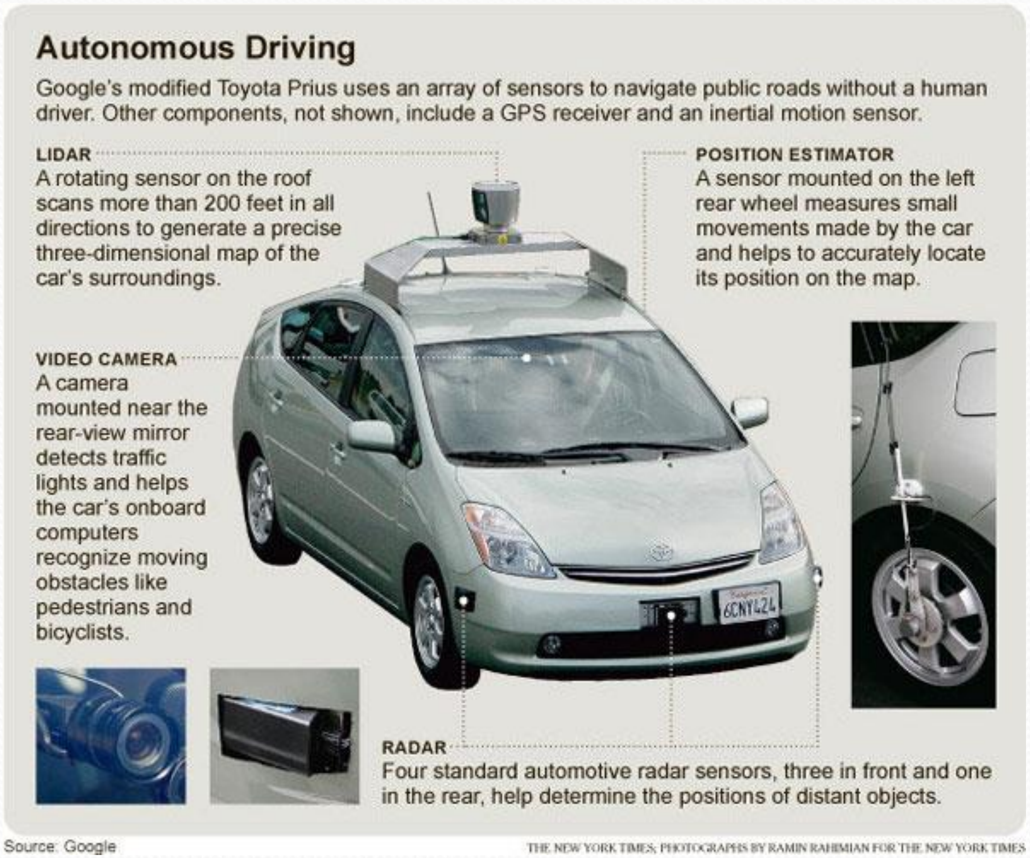
→ Hands-off possible (road following, soon traffic lights…)

July 2016: Tesla driver dies in first fatal crash while using autopilot mode

The autopilot sensors on the Model S failed to distinguish a white tractor-trailer crossing the highway against a bright sky.

Assisted driving (smart car)

Google car: a completely autonomous car !



Military applications

- Hostile/unknown environment

Objectives:

Analyze terrain to assist driving (wounded occupant, bad visibility, night)

Techniques:

Lidar: obstacle avoidance, route finding
Video + neural net = driving

NAVLAB 2

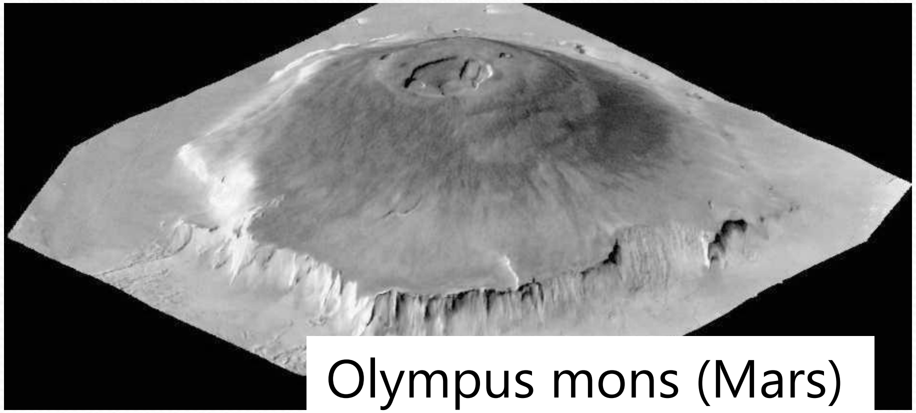
Aerial and satellite imagery

Terrain reconstruction from a collection of images from above.

Applications:

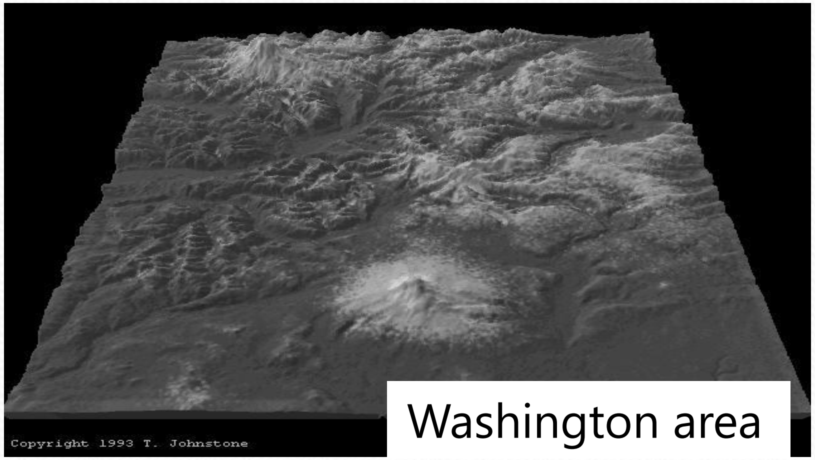
- climate (water set)
- ecological surveillance (rainforest, flood areas).
- military
- topography and mapping (google maps)

Aerial and satellite imagery


Olympus mons (Mars)

Techniques:

- Stereoscopy
- Shape from shading

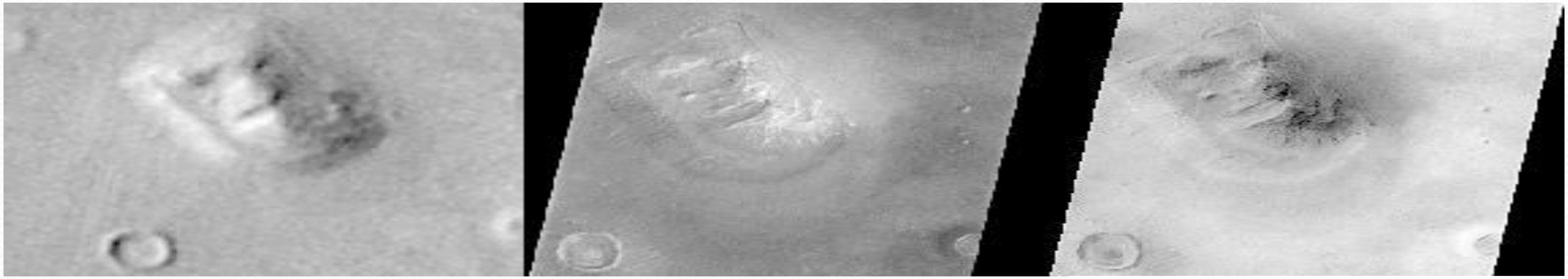


Washington area

Aerial and satellite imagery

Limitations:

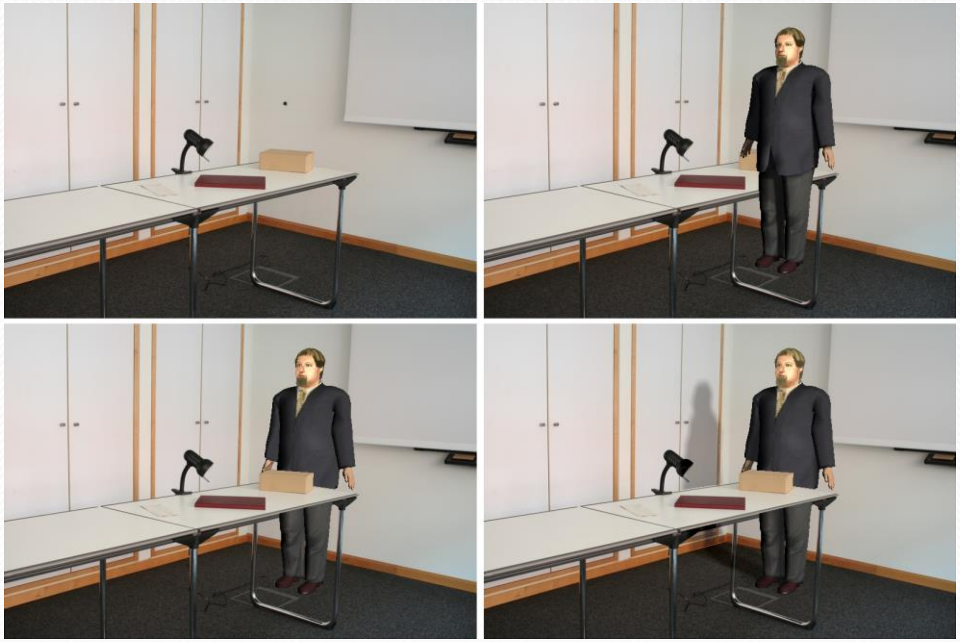
Beware of illumination assumptions !



Analyze → Modify reality

Techniques:

- Stereoscopy
- Marker tracking
- Shape from motion
- Laser scan
- 3D object registration



THE END !