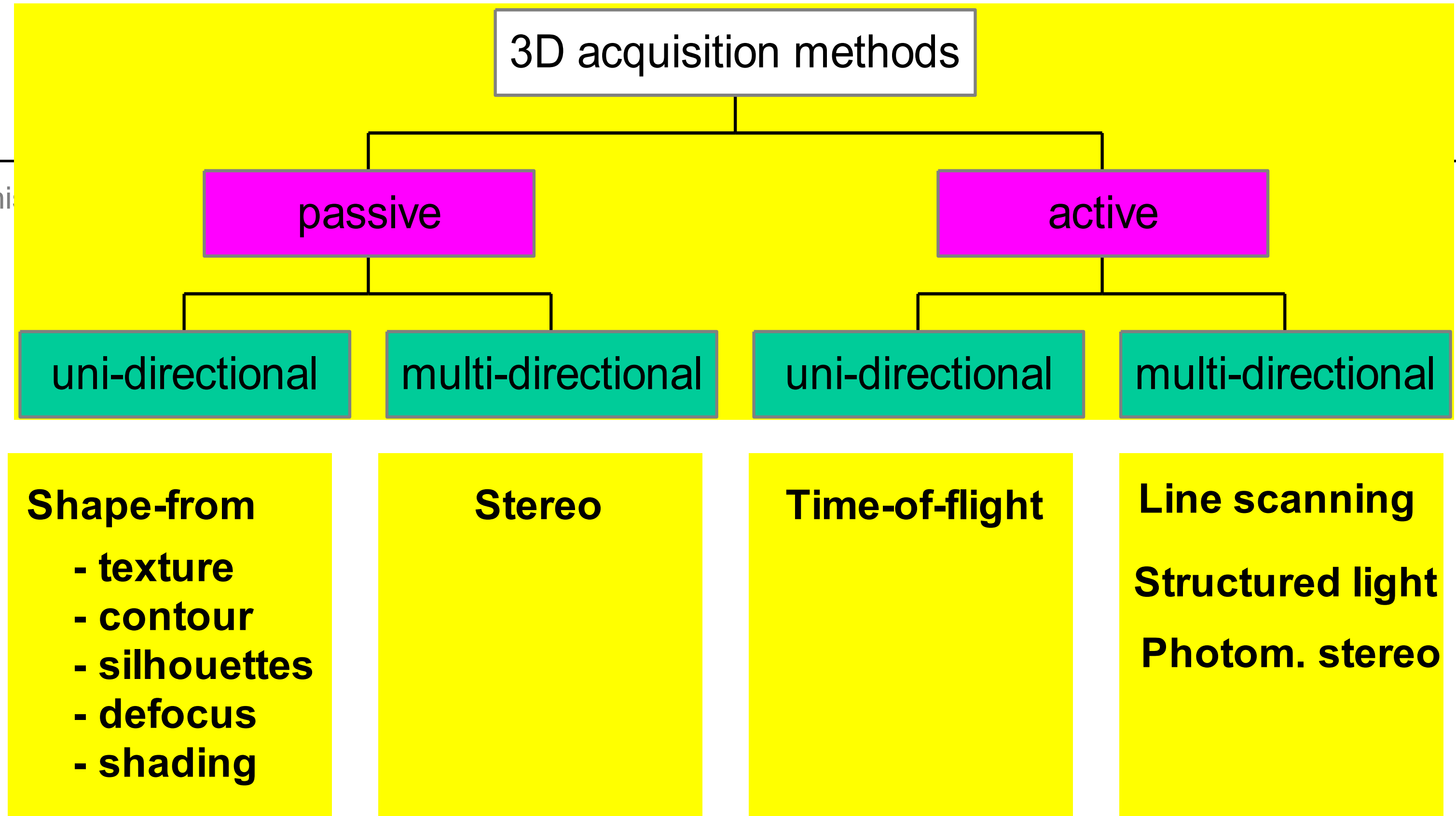
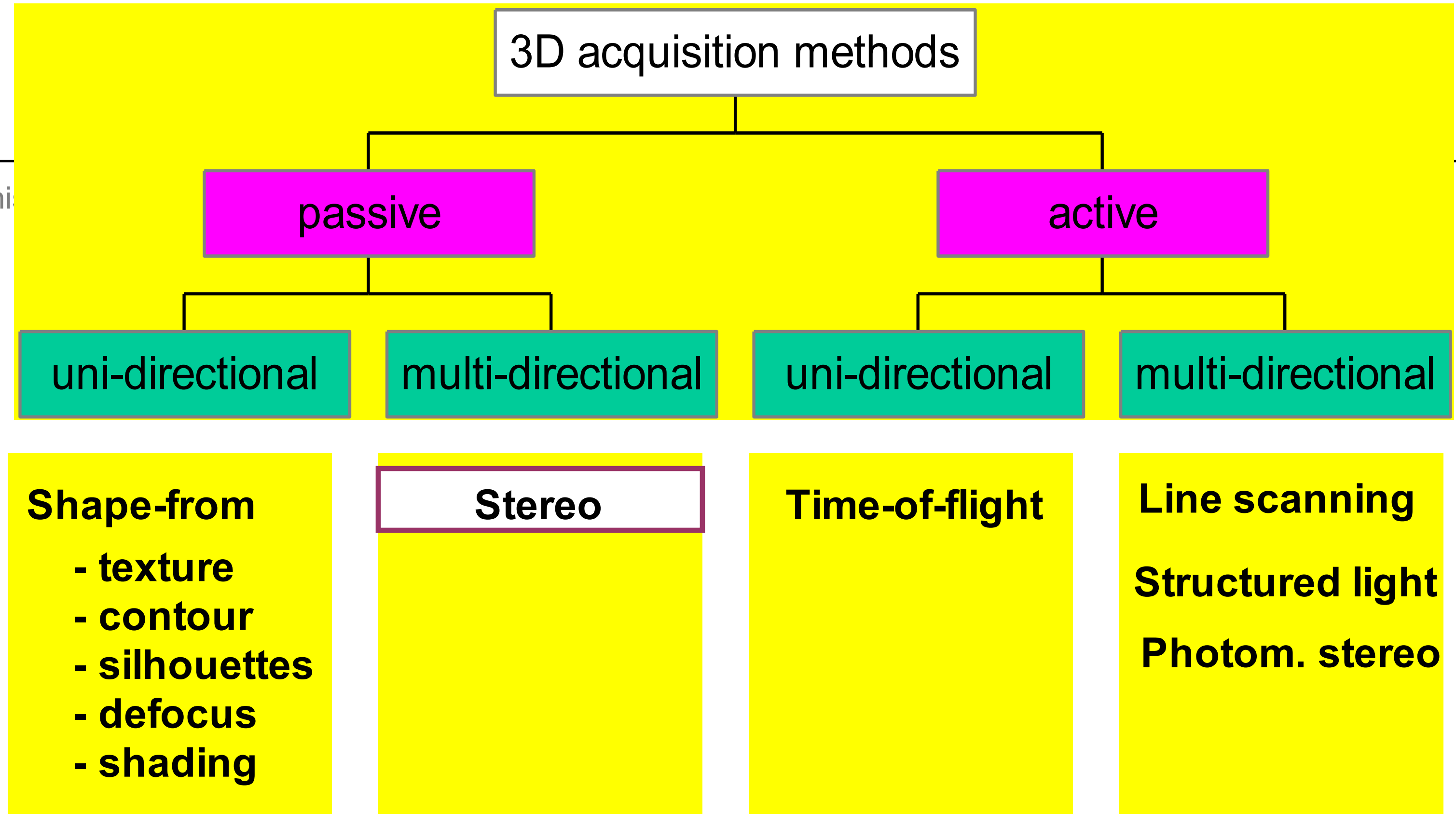
3D acquisition

3D acquisition taxonomy



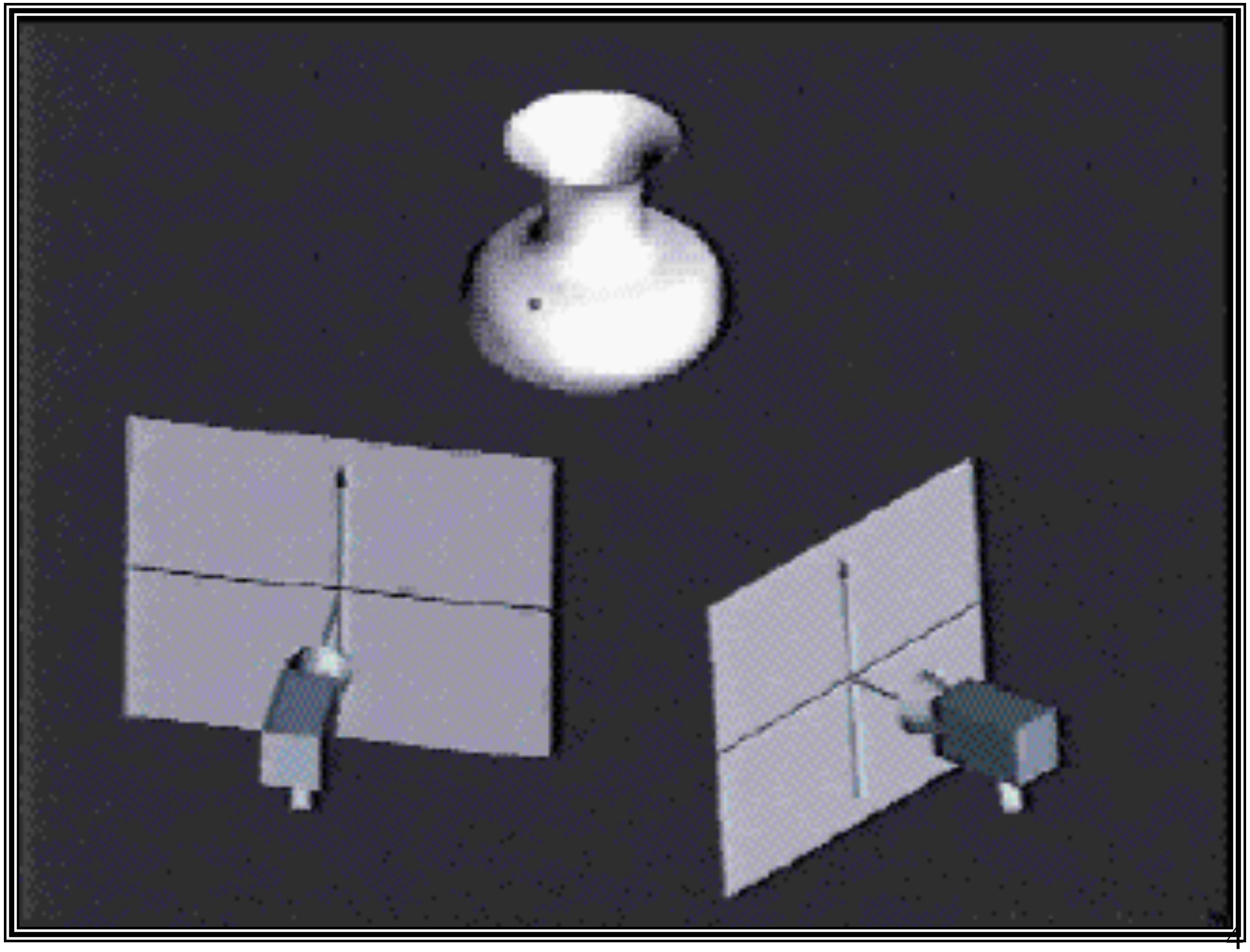
3D acquisition taxonomy



Computer Vision

Stereo

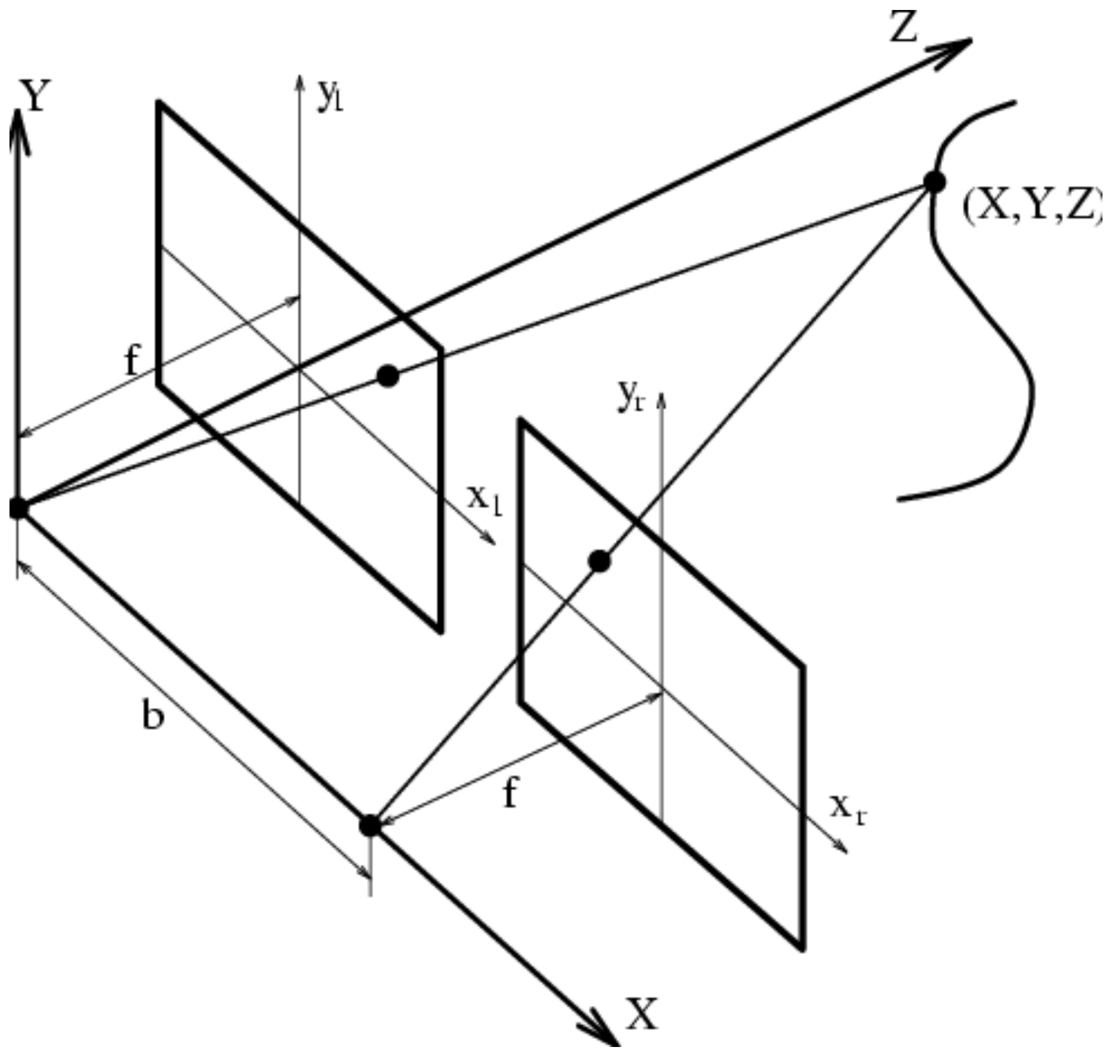
The underlying principle is "triangulation" :



Computer Vision

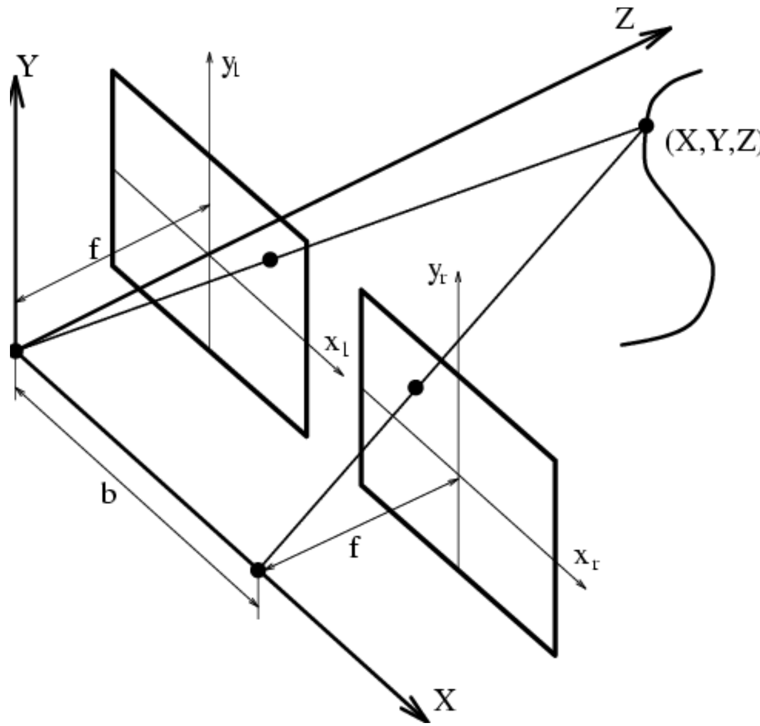
(Passive) stereo

Simple configuration :



Computer Vision

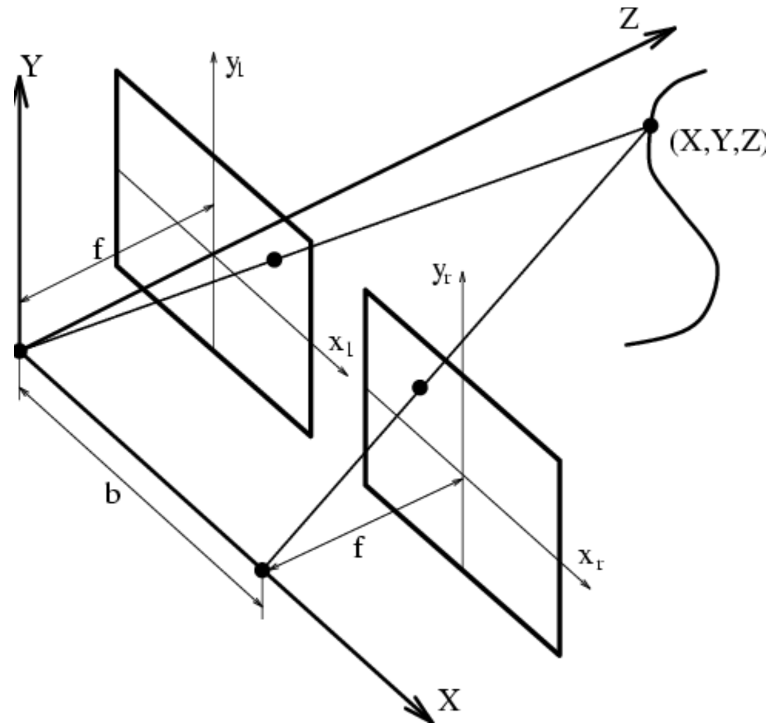
A simple stereo setup



- identical cameras
- coplanar image planes
- aligned *x*-axes

Computer Vision

A simple stereo setup



Reminder :

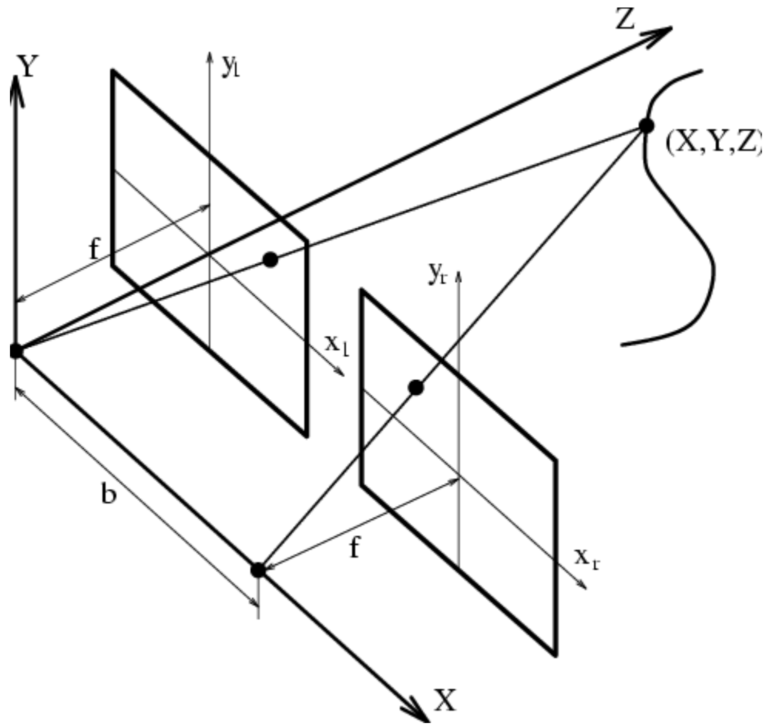
the camera projection can be formulated as

$$\rho p = KR^t (P - C)$$ for some non-zero $\rho \in \mathbb{R}$

Here R is the identity...

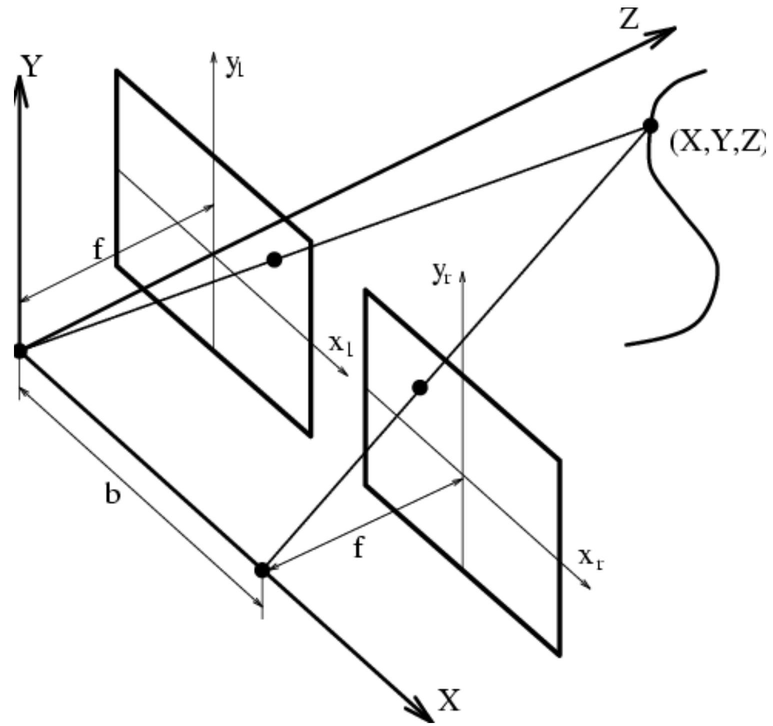
Computer Vision

A simple stereo setup



$$\rho \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = K \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad \rho' \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = K \begin{pmatrix} X - b \\ Y \\ Z \end{pmatrix} \quad K = \begin{pmatrix} fk_x & 0 & 0 \\ 0 & fk_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

A simple stereo setup



$$\begin{cases} x = \dfrac{fk_x X}{Z}, \\ y = \dfrac{fk_y Y}{Z}, \end{cases} \text{ and } \begin{cases} x' = \dfrac{fk_x (X-b)}{Z}, \\ y' = \dfrac{fk_y Y}{Z}, \end{cases}$$

Note that $y = y'$

Computer Vision

A simple stereo setup

The 3D coordinates of the point are

$$X = b\frac{x}{(x - x')},$$

$$Y = b\frac{k_x}{k_y}\frac{y}{(x - x')},$$

$$Z = bk_x\frac{f}{(x - x')}.$$

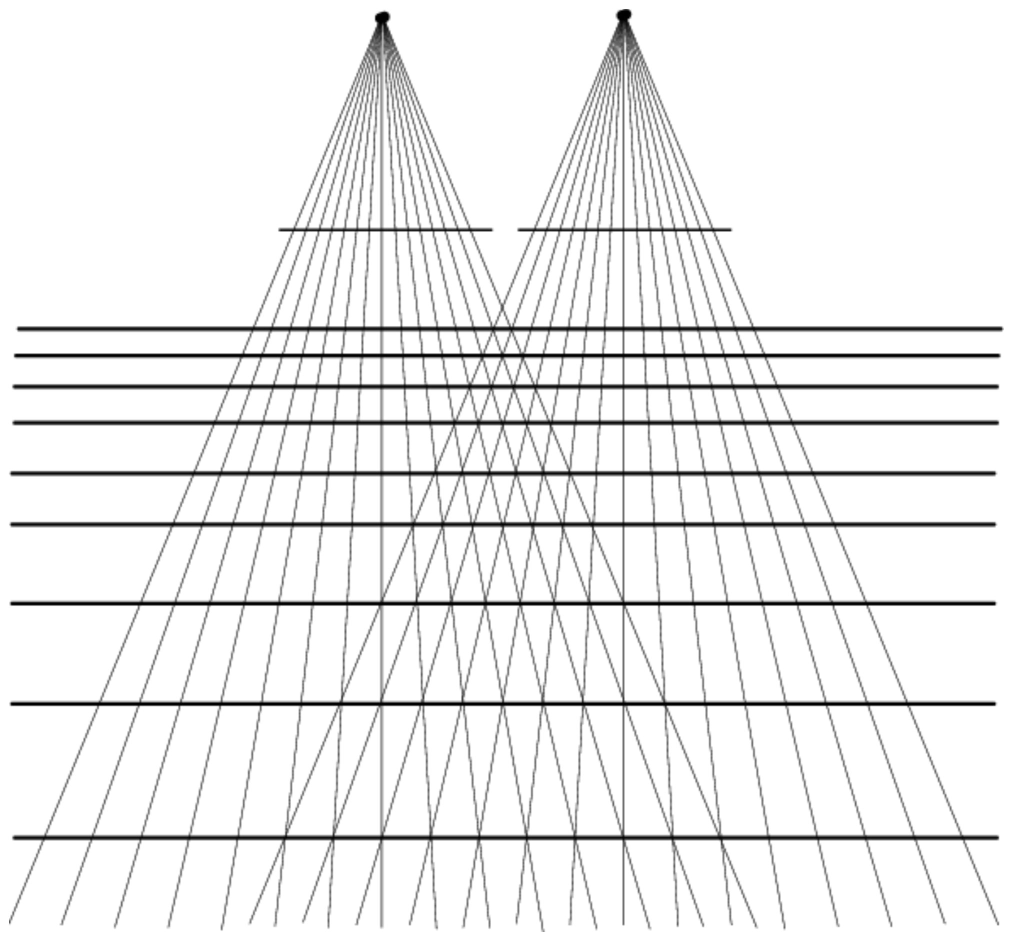
$(x - x')$ is the so-called *disparity*

Stereo is imprecise for far away objects,but increasing b and/or f can increase depth resolution

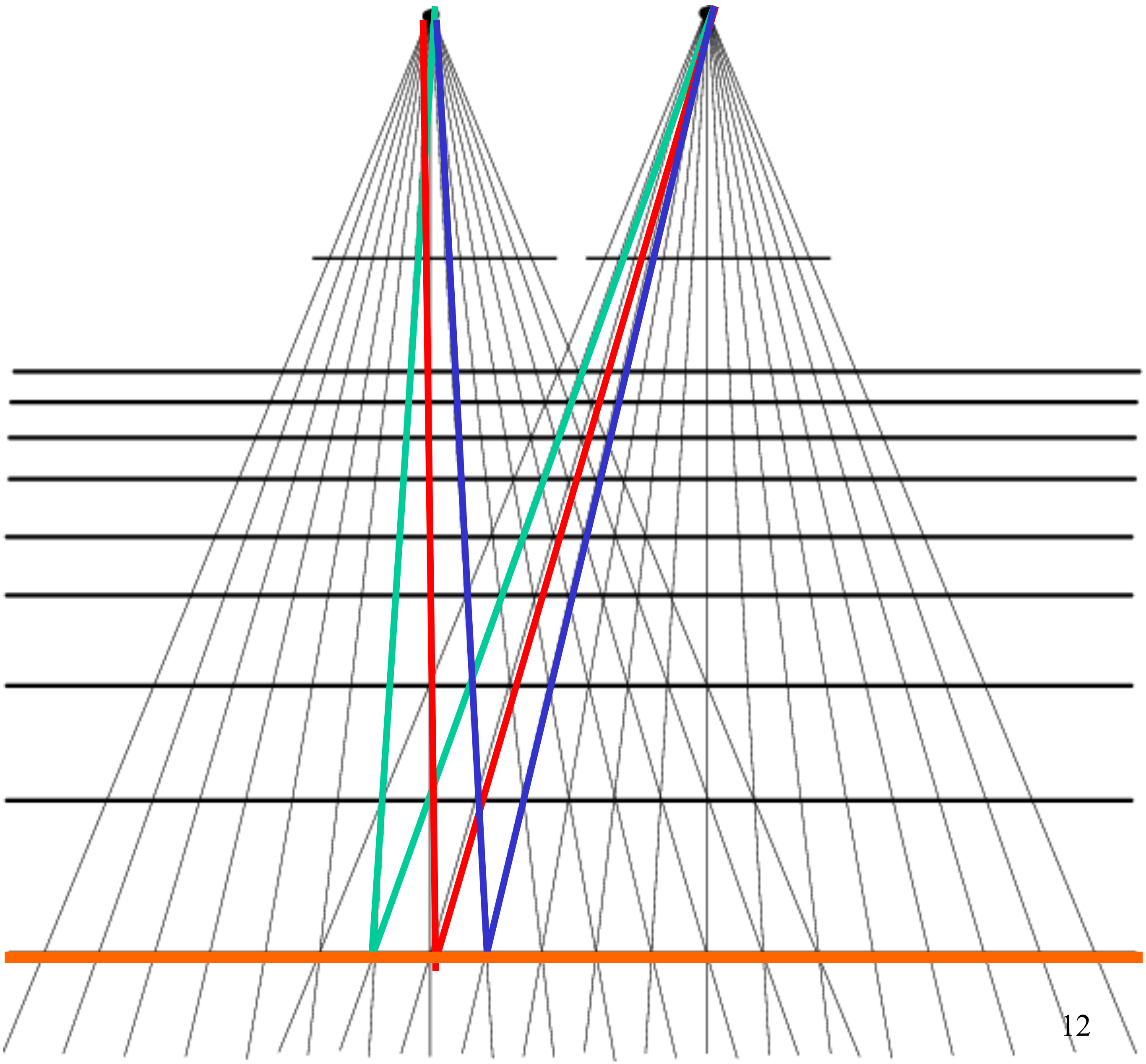
Computer Vision

A simple stereo setup

Notice: for this simple setup, same disparity means same depth

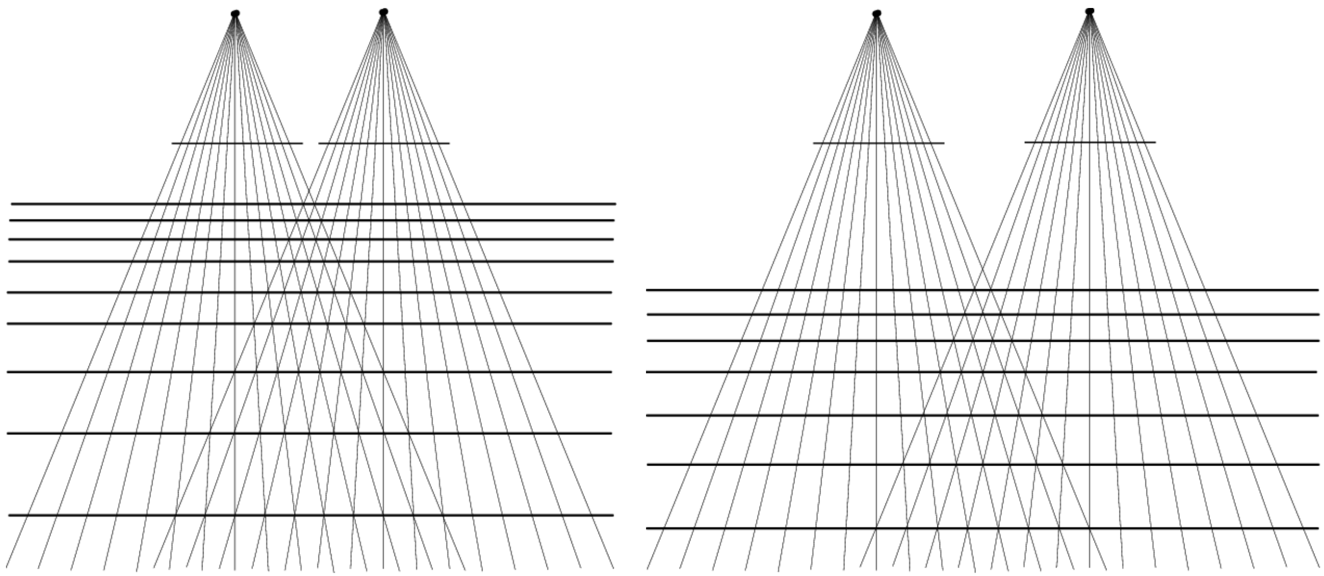


same disparity means same depth



A simple stereo setup

Increasing b increases depth resolution

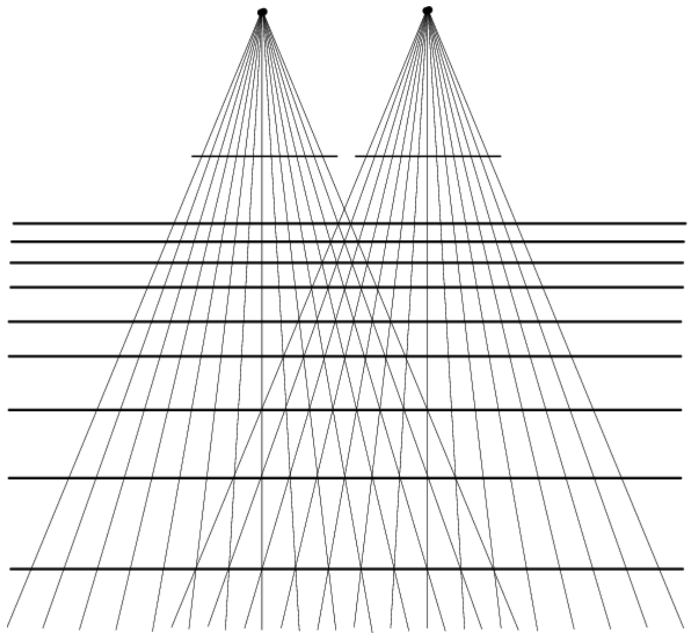


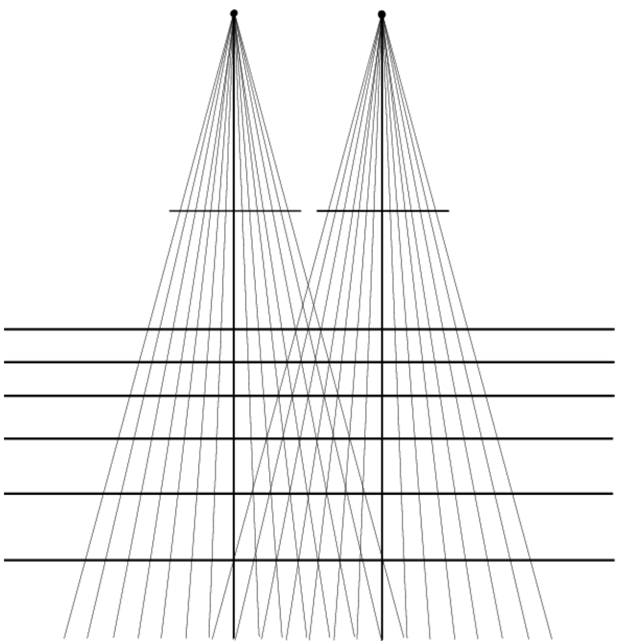
one has to strike a balance with visibility...

Computer Vision

A simple stereo setup

Increasing f increases depth resolution





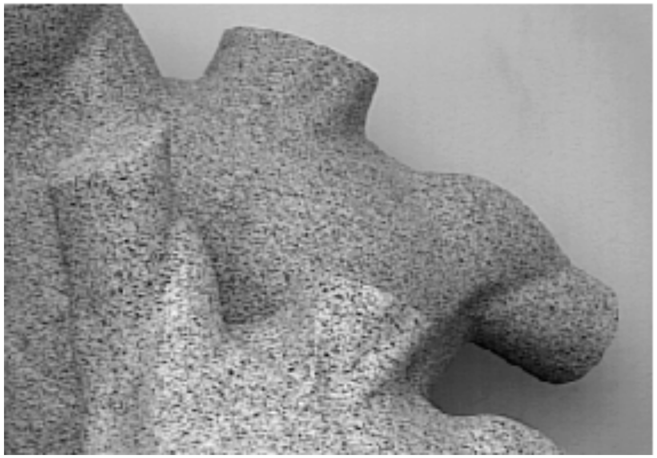
one has to strike a balance with visibility...

Remarks

- 1. increasing b and/or f increases depth resolution but reduces simultaneous visibility
- 2. iso-disparity loci are depth planes, not so for other configurations
- 3. as soon as the disparity gets too small, depth difference can no longer be seen;
  hence human stereo only works up to ± 10 m
- 4. the real problem is finding correspondences

Computer Vision

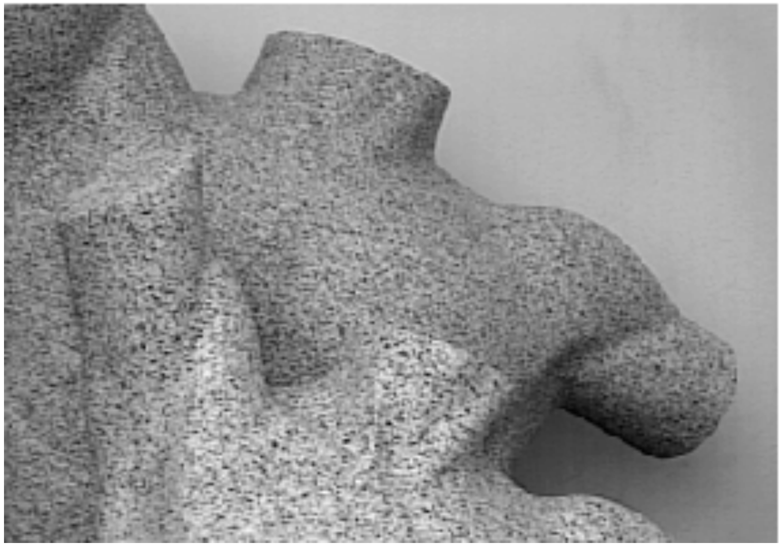
A simple stereo setup

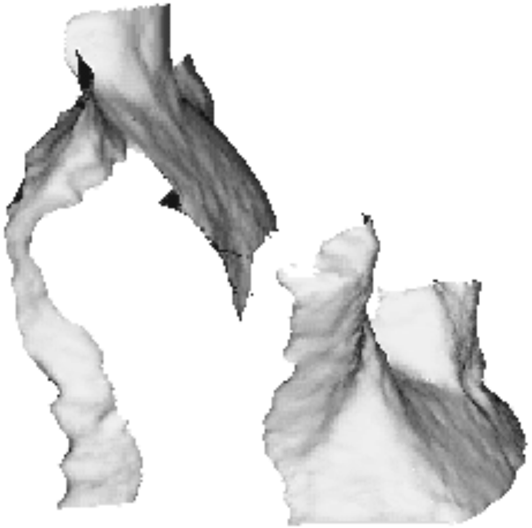


The HARD problem is finding the *correspondences*

Notice : no reconstruction for the untextured back wall...

A simple stereo setup





The HARD problem is finding the *correspondences*

Notice : no reconstruction for the untextured back wall...


film &
video
Sidmar

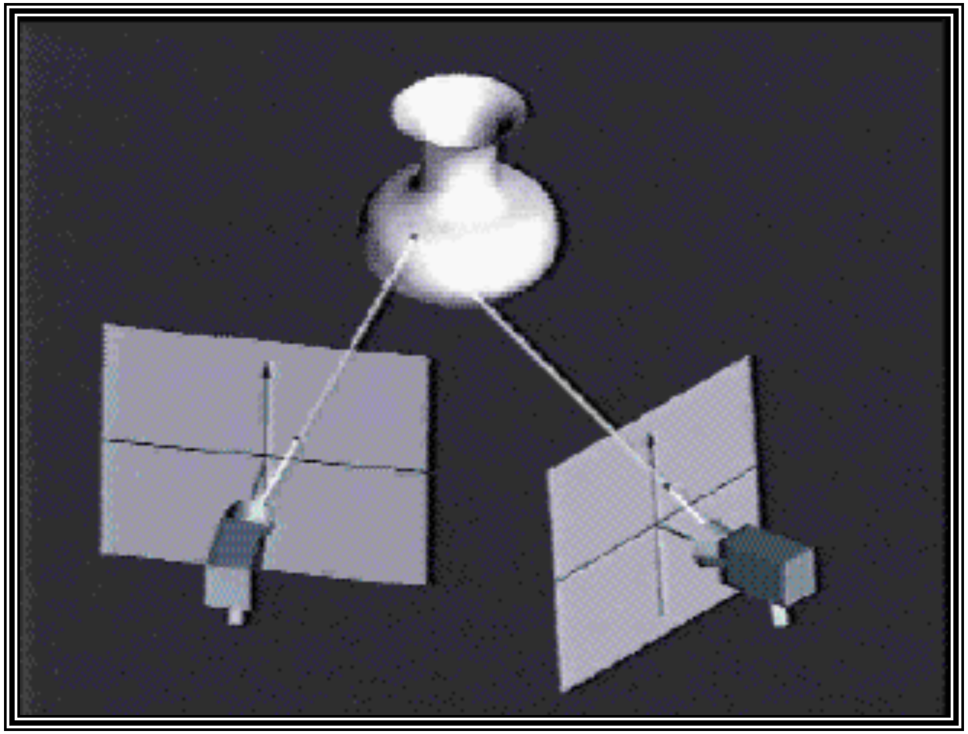
Computer Vision

Computer Vision

Stereo, the general setup

we start by the relation between the 2 projections of a point to ease correspondence search

in the second image the point must be along the projection of the viewing ray for the first camera :



Computer Vision

Stereo, the general setup

We cast this constraint in mathematical expressions :

p and p' are the two images of P

$$\mu\, p = K\, R^t (P - C)$$

$$\rho' p' = K' R'^t (P - C')$$

w.r.t. world frame P is on the ray with equation

$$P = C + \mu R K^{-1} p$$ for some $\mu \in \mathbb{R}$

Stereo, the general setup

so, the ray is given by

$$P = C + \mu R K^{-1} p \quad \text{for some} \quad \mu \in \mathbb{R}$$

Now we project it onto the second image
In general, points project there as follows :

$$\rho' p' = K' R'^{t} (P - C')$$

and thus, filling in the ray's equation

$$\rho' p' = \mu K' R'^{t} R K^{-1} p + K' R'^{t} (C - C')$$

Computer Vision

Stereo, the general setup

the projected ray was found to be

$$\rho' p' = \mu K' R'^{t} R K^{-1} p + K' R'^{t} (C - C')$$

the second term is the projection of the 1st camera's center, the so-called *epipole*

$$\rho'_e e' = K' R'^{t} (C - C')$$

the first term is the projection of the ray's point at infinity, the so-called *vanishing point*

finally, adopting the simplifying notation

$$A = \frac{1}{\rho'_e} K' R'^{t} R K^{-1}$$

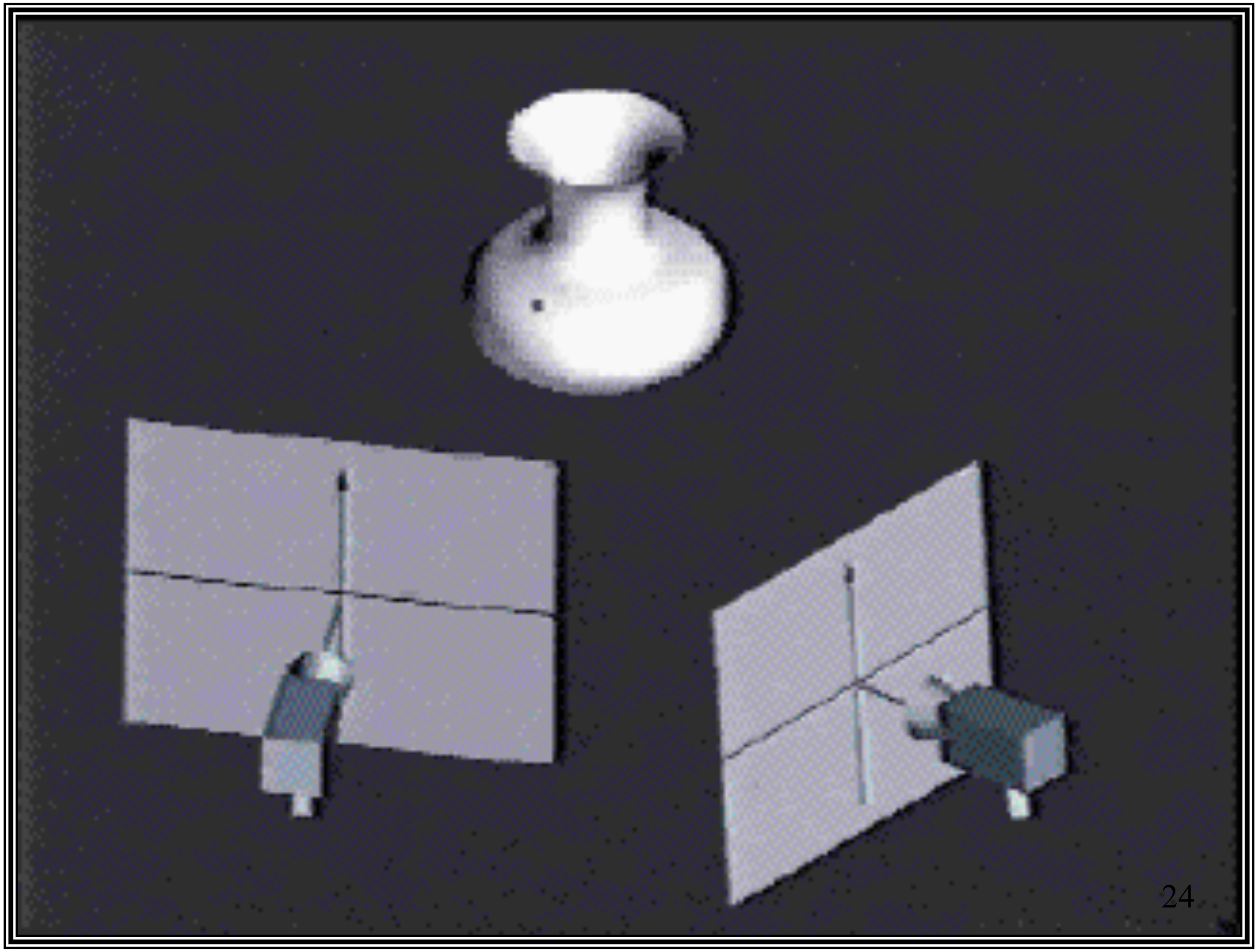
$$\rho' p' = \mu A p + \rho'_e e'$$

A is the *infinity homography*

Computer Vision

Stereo, the general setup

note that the epipole lies on all the epipolar lines



Stereo, the general setup

$$\rho' p' = \mu A p + \rho'_e e'$$

Computer Vision

Stereo, the general setup

$$\rho' p' = \mu A p + \rho'_e e'$$

expresses that p' lies on the line l' through the epipole e' and the vanishing point Ap of the ray of sight of p (in the 2nd image)

Computer Vision

Stereo, the general setup

$$\rho' p' = \mu A p + \rho'_e e'$$

the epipolar constraint (epipolar line)

we can rewrite this constraint as

$$\left| p' e' A p \right| = p'^t \left(e' \times A p \right) = 0$$

Stereo, the general setup

$$\left| p' e' Ap \right| = p'^t (e' \times Ap) = 0$$

can be written, given

$$[e']_\times = \begin{pmatrix} 0 & -e'_3 & e'_2 \\ e'_3 & 0 & -e'_1 \\ -e'_2 & e'_1 & 0 \end{pmatrix}$$

as

$$\left| p' e' Ap \right| = p'^t [e']_\times Ap$$

$F = [e']_\times A$ is the *fundamental matrix*

F is a 3x3 matrix, but has rank 2

Stereo, the general setup

$$p'^t [e']_\times A p = 0 \quad \rightarrow \quad \boxed{p'^t \; F \; p = 0}$$

The 3-vector $p'^t F$ contains the line coordinates of the epipolar line of p' (i.e. a line in the 1st image that contains its corresponding point p)

The 3-vector $F\,p$ contains the line coordinates of the epipolar line of p (i.e. a line in the 2nd image that contains its corresponding point p')

Hence, the epipolar matrix works in both directions

Computer Vision

Stereo, the general setup

Andrea Fusiello, CVonline

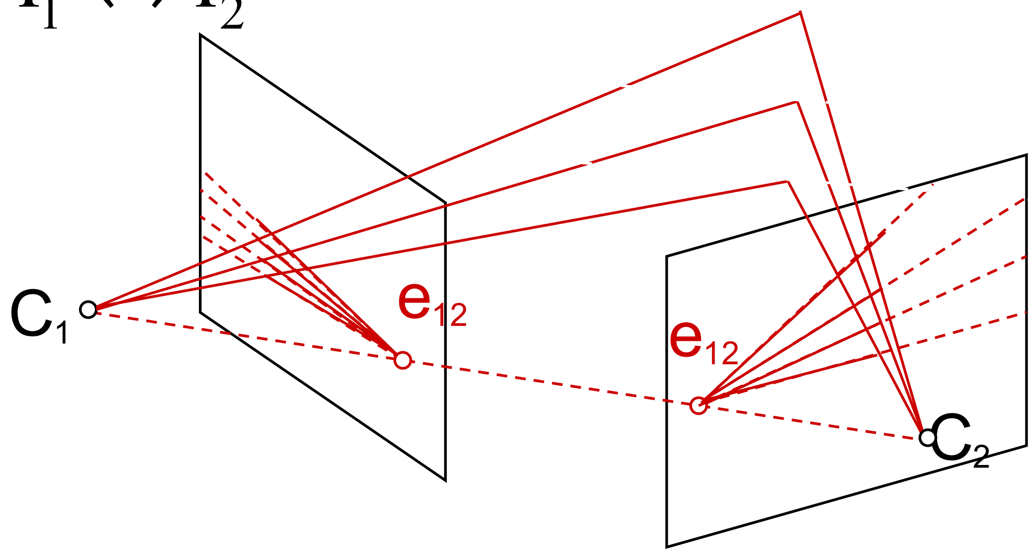
Computer Vision

Epipolar geometry cont'd

Epipolar geometry cont'd

- Epipolar lines are in mutual correspondence

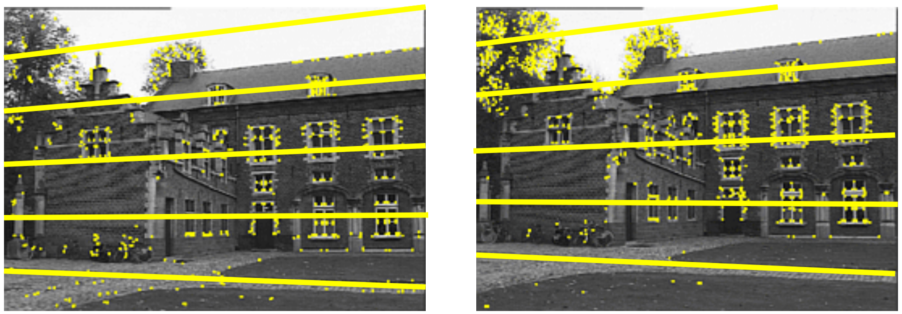
$$l_1 \leftrightarrow l_2$$

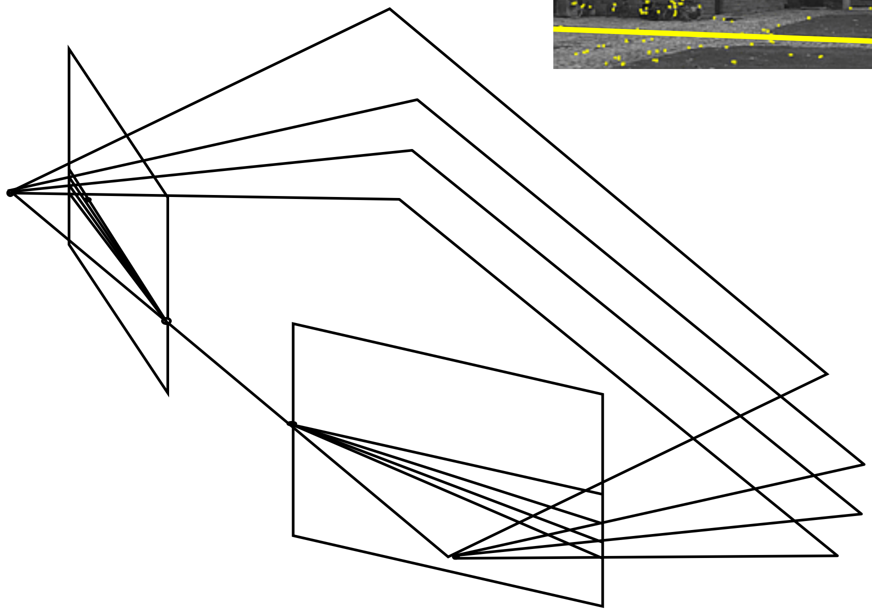


- allows to separate matching problem:
  matching pts on an epipolar line to pts on the corresponding epipolar line

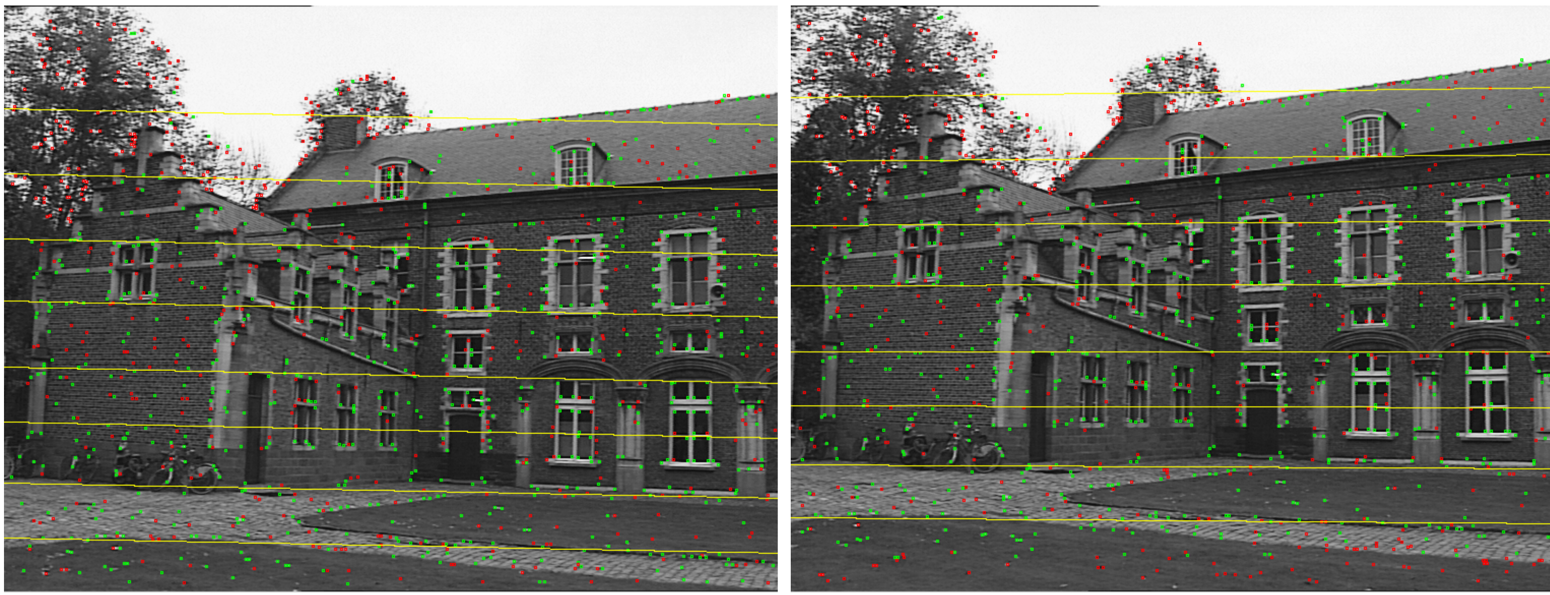
Exploiting epipolar geometry

Separate 2D correspondence search problem to 1D search problem by using two view geometry





Epipolar geometry cont'd



Computer Vision

Stereo, the general setup

- one point yields one equation $p'^t F\, p = 0$ that is linear in the entries of the fundamental matrix F
  so, we can actually obtain F without any prior knowledge about camera settings if we have sufficient pairs of corresponding points !!

- F can be computed *linearly* from 8 pairs of corresponding points,
  i.e. already from 8 `correspondences'
  (not 9, as this is a homogeneous system and one coefficient can be fixed to value 1 to fix the scale !)

- F being rank 2 yields an additional, but non-linear constraint. Thus, 7 correspondences suffice to *non-linearly* solve for F

Stereo, the general setup

Remarks :

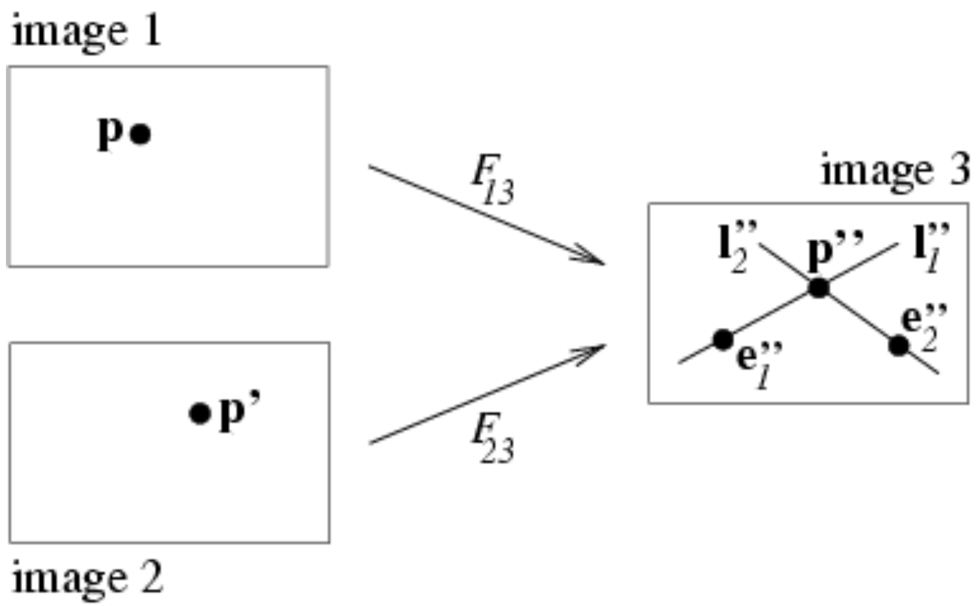
- Of course, in practice one wants to use as many Correspondences as available, e.g. for obtaining a least-squares solution, based on the linear system, followed by a step to impose rank 2.

- Often, F is found through a procedure called RANSAC (RANdom Sample Consensus). It starts from a randomly drawn subset of correspondences of minimal size (e.g. 8), and then keeps on drawing until a subset is found that yields an F so that many correspondences are seen to obey the epipolar constraint. RANSAC is good to fend off against correspondences that are wrong (`outliers')

Computer Vision

Relations between 3 views

one could use more than 2 images, e.g. 3
suppose P projects to p,p', and p''
p'' is found at the intersection of epipolar lines :



fails when the epipolar lines coincide

⇒ *trifocal constraints*

Computer Vision

Computer Vision

Correspondence problem : constraints

Reducing the search space :

- 1. Points on the epipolar line
- 2. Min. and max. depth ⇒ line segment
- 3. Preservation of order
- 4. Smoothness of the disparity field

Correspondence problem : methods

1. correlation
   - deformations…
   - small window ⇒ noise!
   - large window ⇒ bad localisation
2. feature-based
   - mainly edges and corners
   - sparse depth image
3. regularisation methods

Computer Vision

Stereo, the general setup

3D reconstruction

$$P = C + \mu R K^{-1} p$$

$$P = C' + \mu' R' K'^{-1} p'$$

Yields 6 equations in 5 unknowns X,Y, Z and μ, μ'

However, due to noise and errors, the rays may not intersect!

⇒ e.g. use the middle where the rays come closest

Computer Vision

3D city models

Computer Vision

3D city models – ground level

Mobile mapping example – for measuring

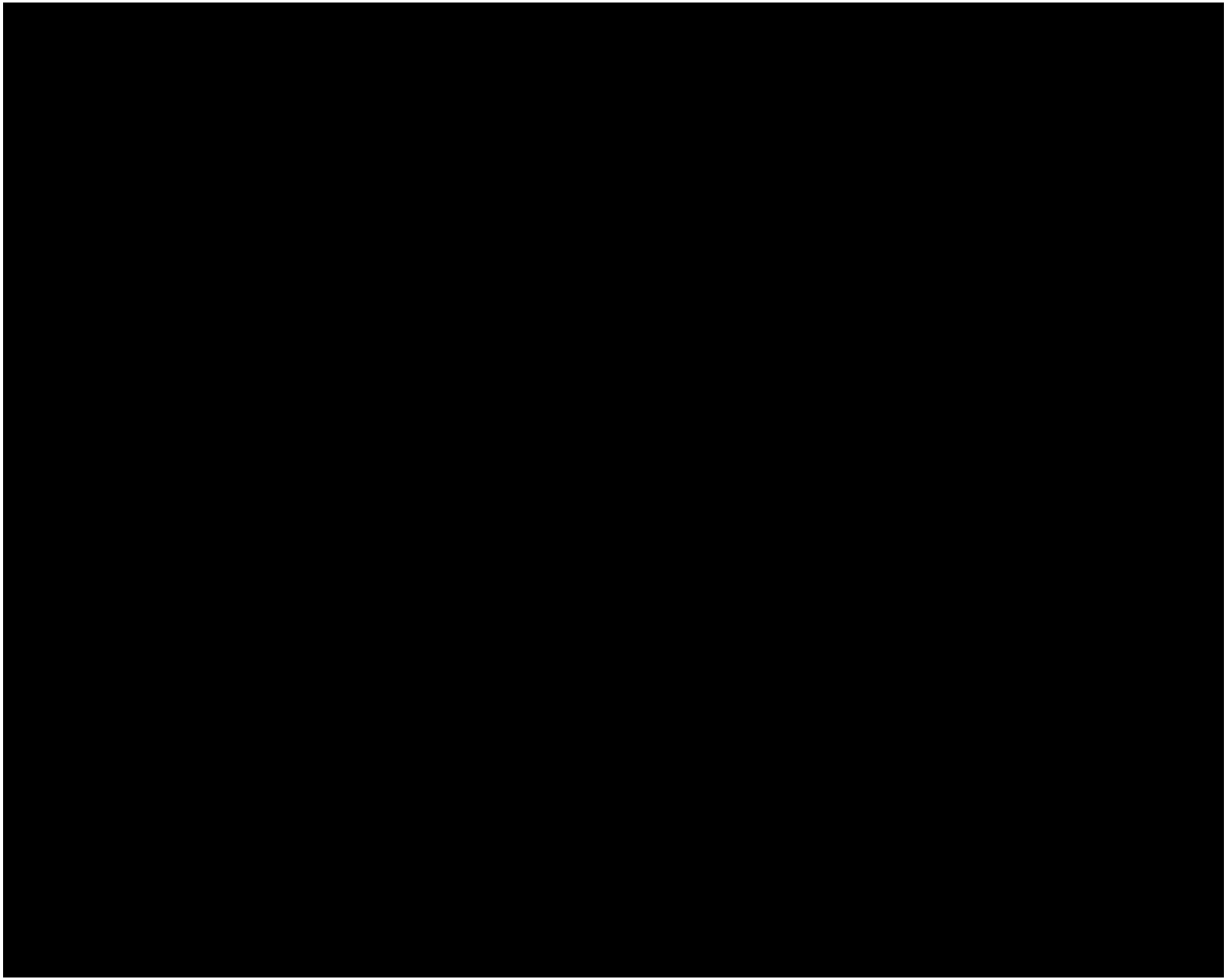
3D city models – ground level

Can also be turned into 3D for visualisation, but one needs to stay close to the camera viewpoints.

The example shown is of Quebec

Computer Vision

3D city models – ground level



Uncalibrated reconstruction

From 2 views...

If the camera translates...

An affine reconstruction can be made

A projective reconstruction is always possible

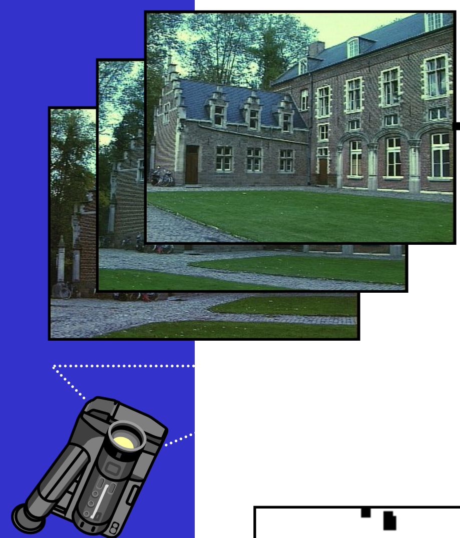
(if no pure rot.)

Uncalibrated reconstruction

From 3 general views
taken with the same camera parameters...

A metric reconstruction is possible

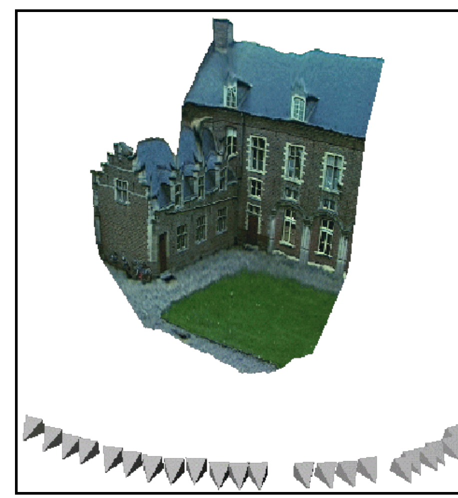
Uncalibrated reconstruction



Tracking and Calibration → **Dense depth estimation** → **3D surface modeling**

Points and cameras

Depth map



3D models

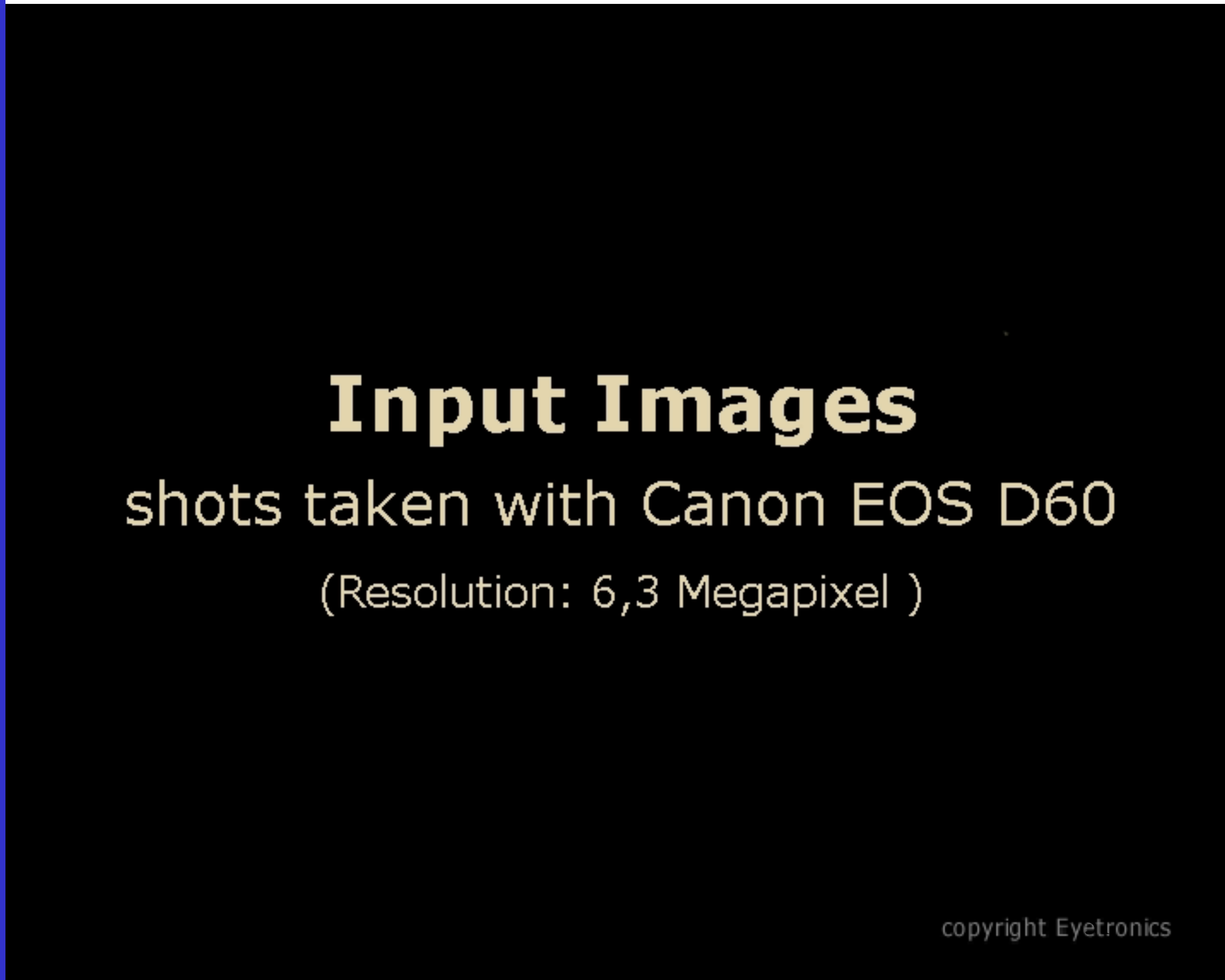
Uncalibrated reconstruction

Computer Vision

Uncalibrated reconstruction - example

Univ. of Leuven

Shape-from-stills

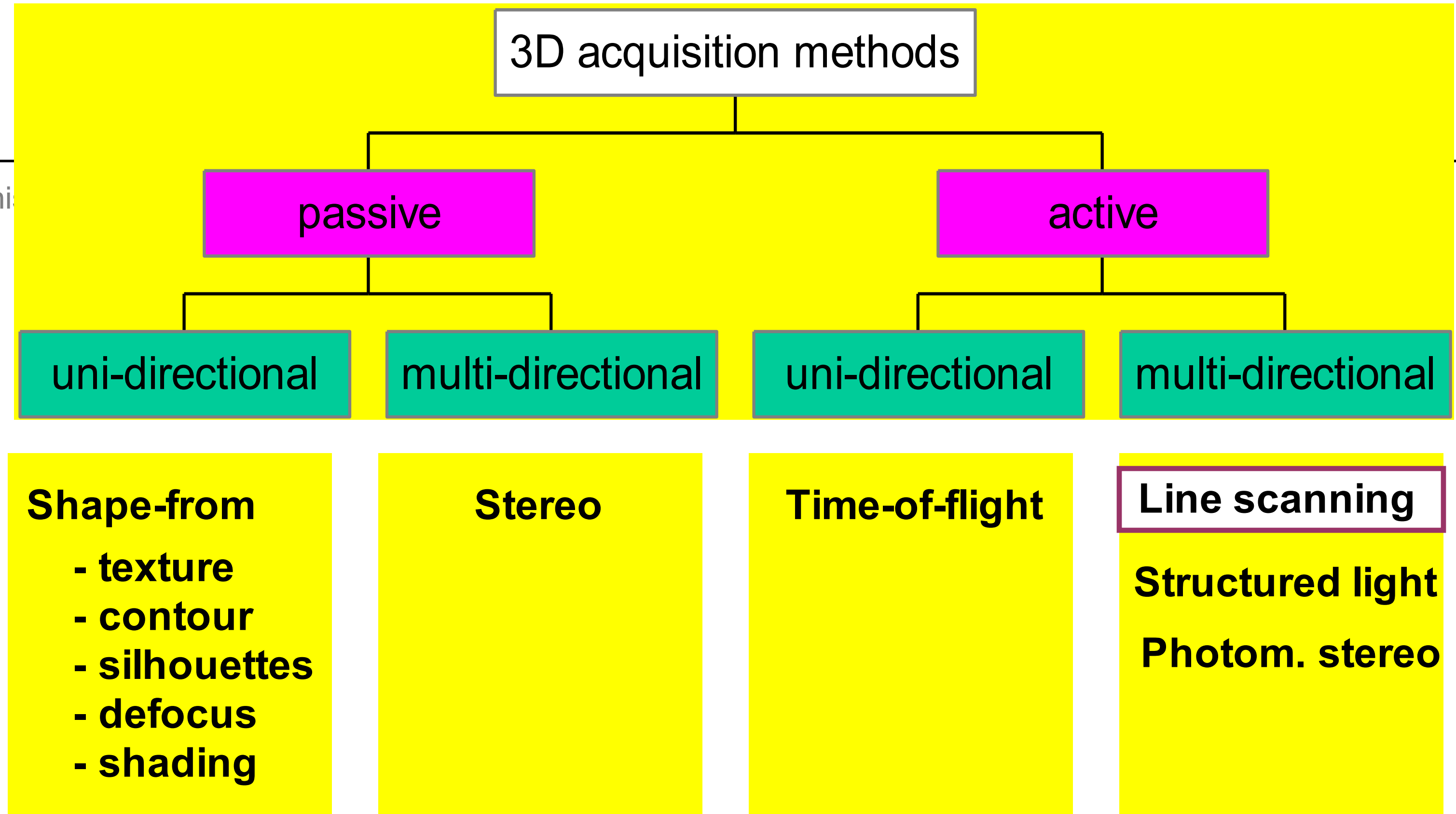


Shape-from-stills

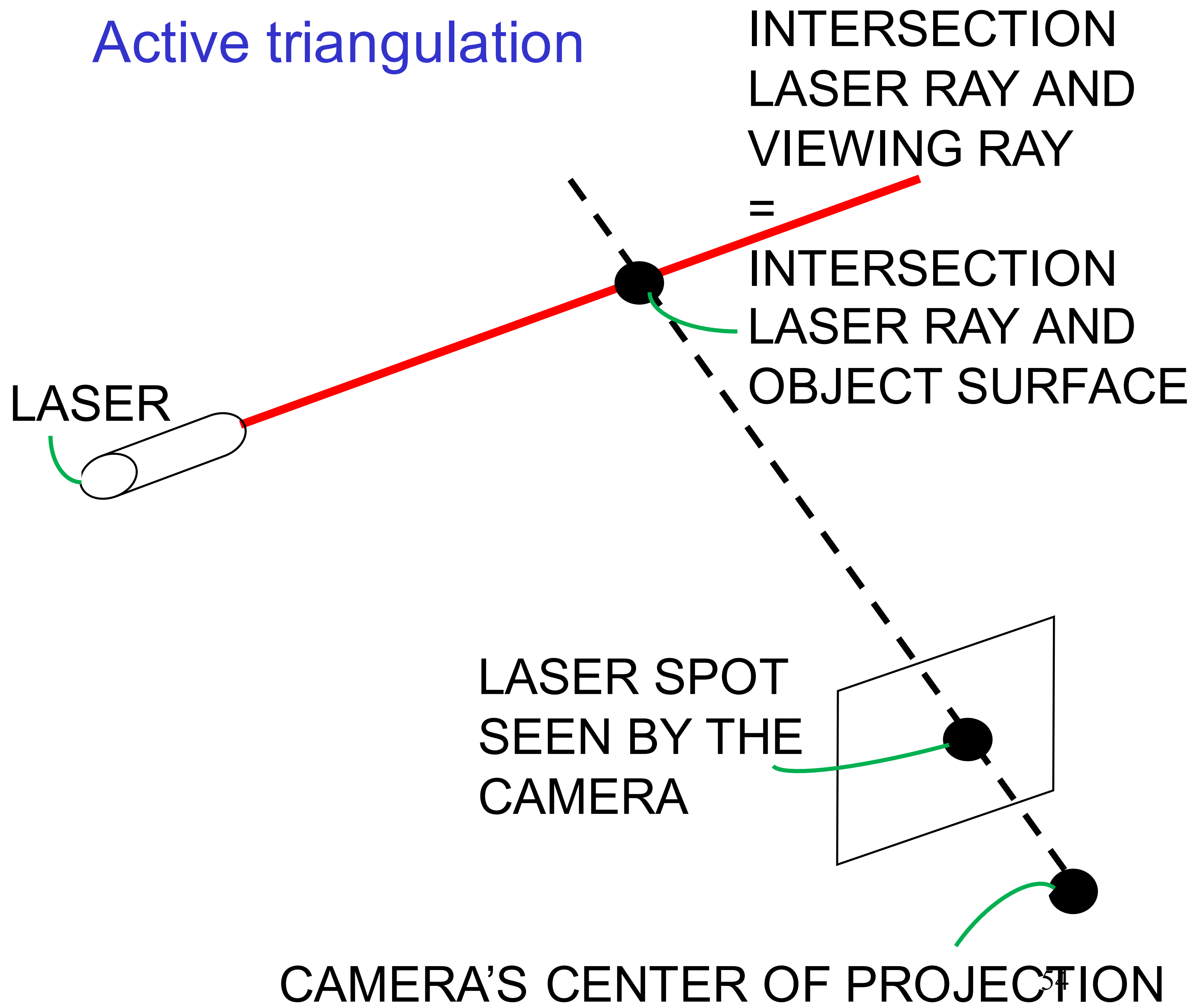
www.arc3d.be

Webservice,
free for non-commercial use

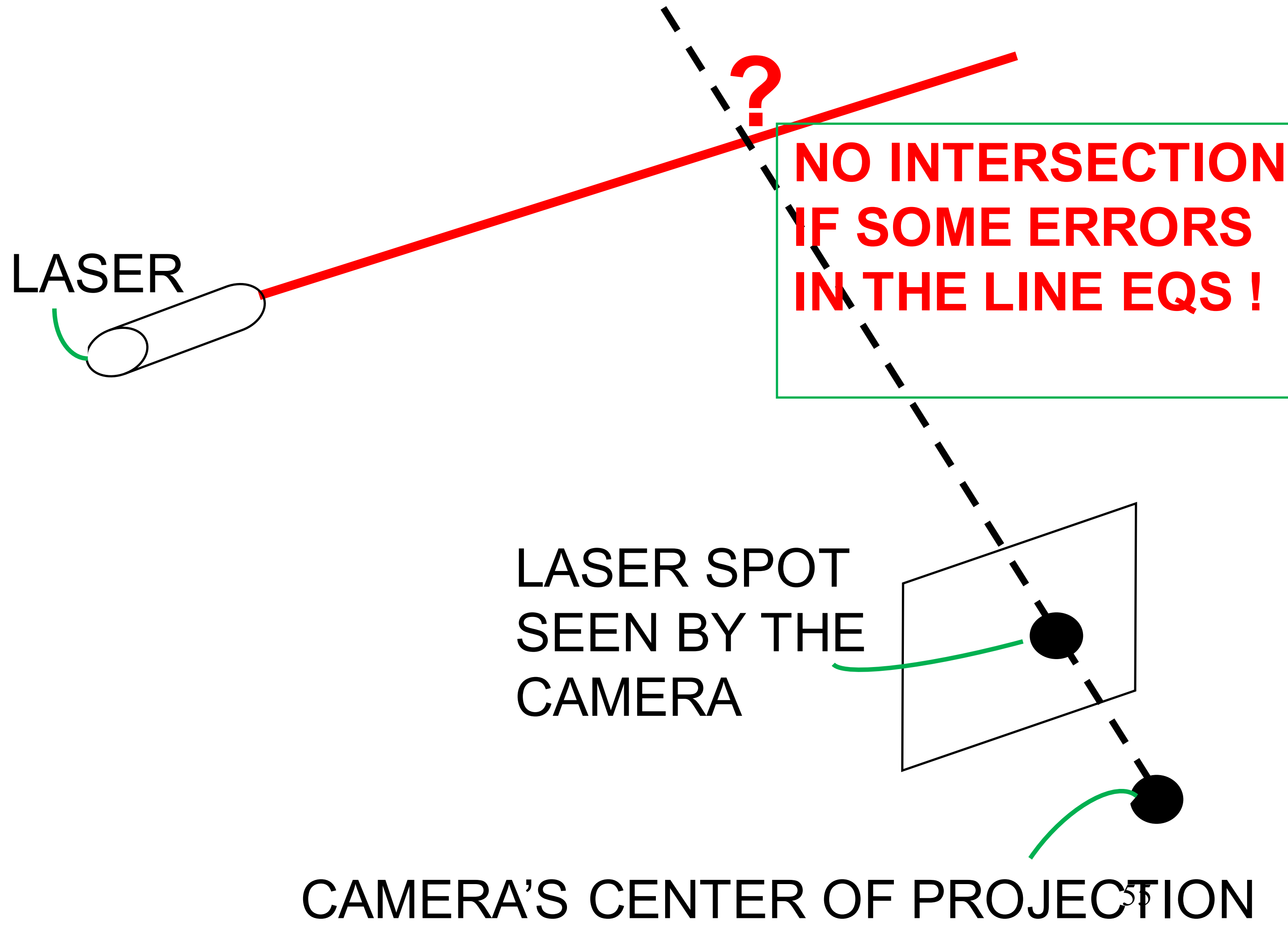
3D acquisition taxonomy



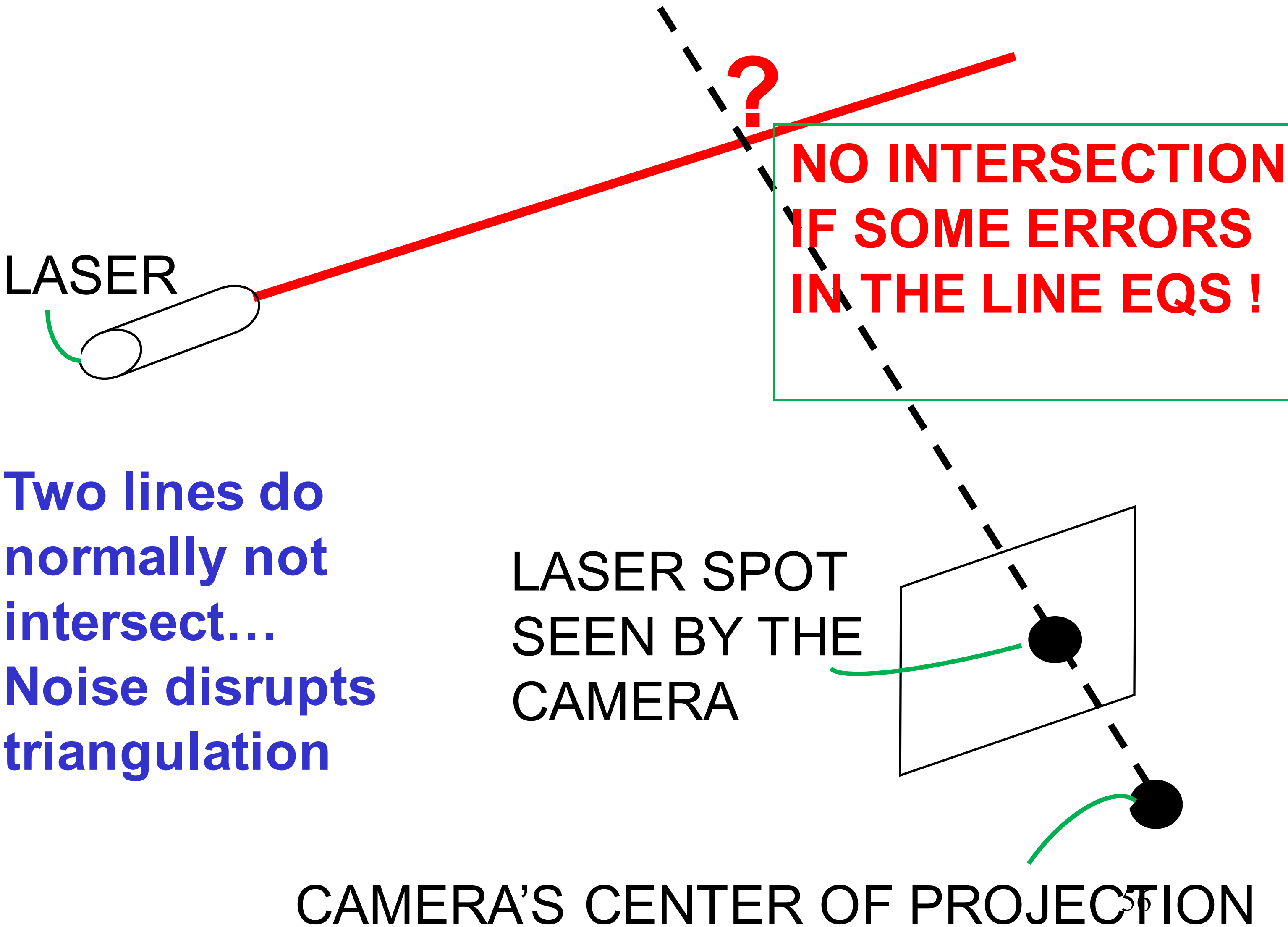
Active triangulation



Active triangulation



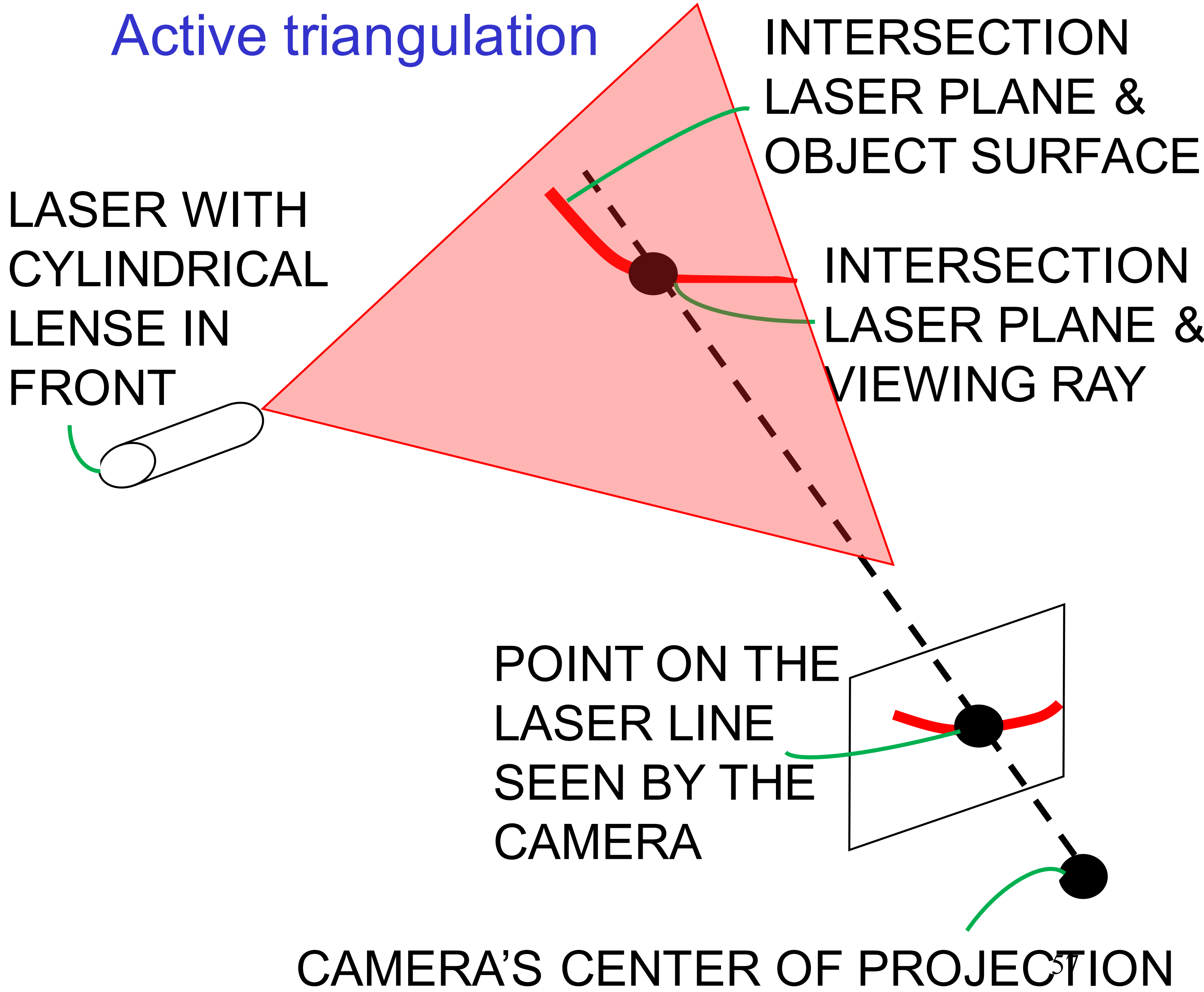
Active triangulation



Two lines do normally not intersect…
Noise disrupts triangulation

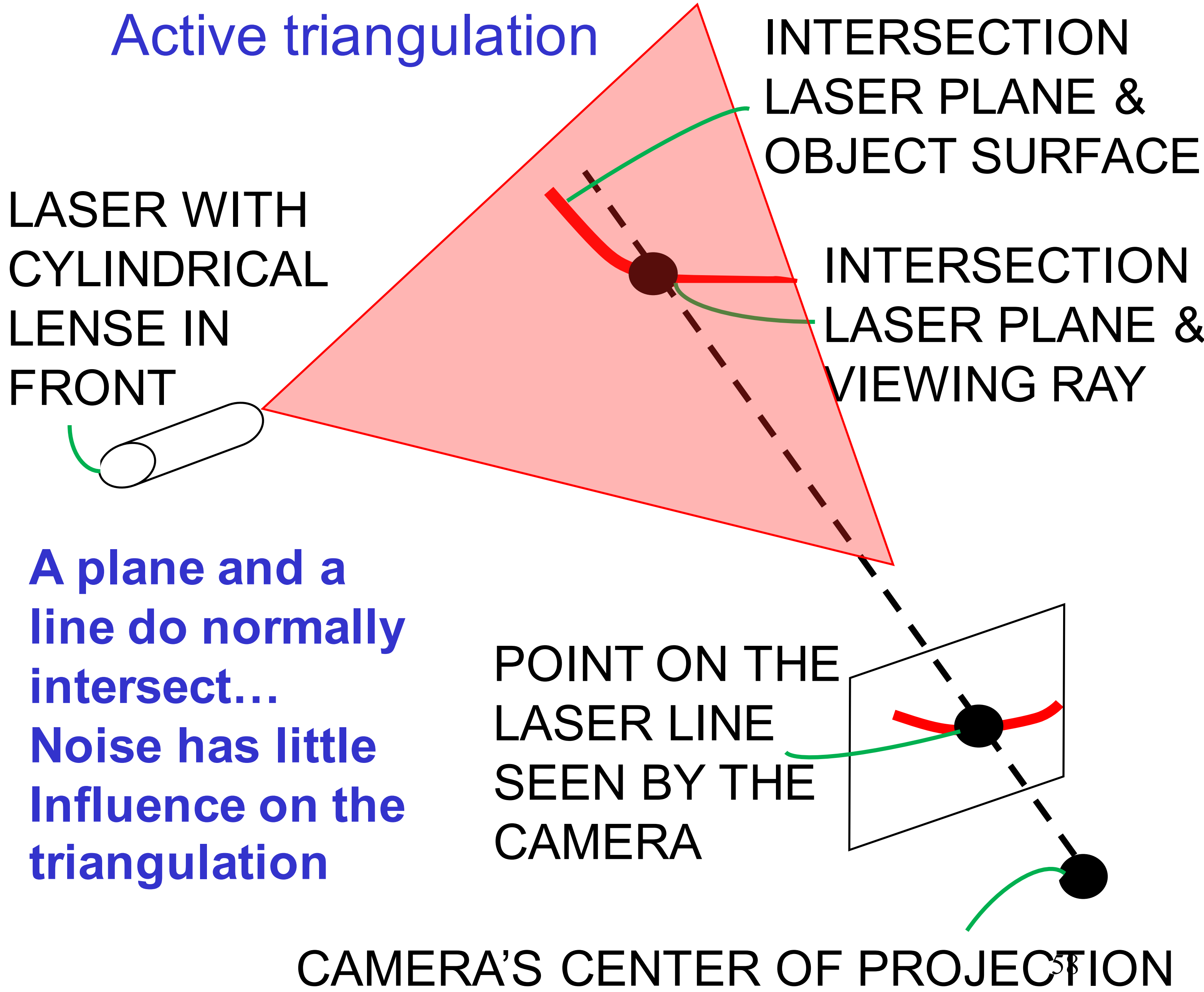
Computer Vision

Active triangulation



Computer Vision

Active triangulation



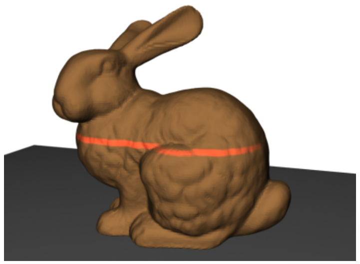
A plane and a line do normally intersect…
Noise has little Influence on the triangulation

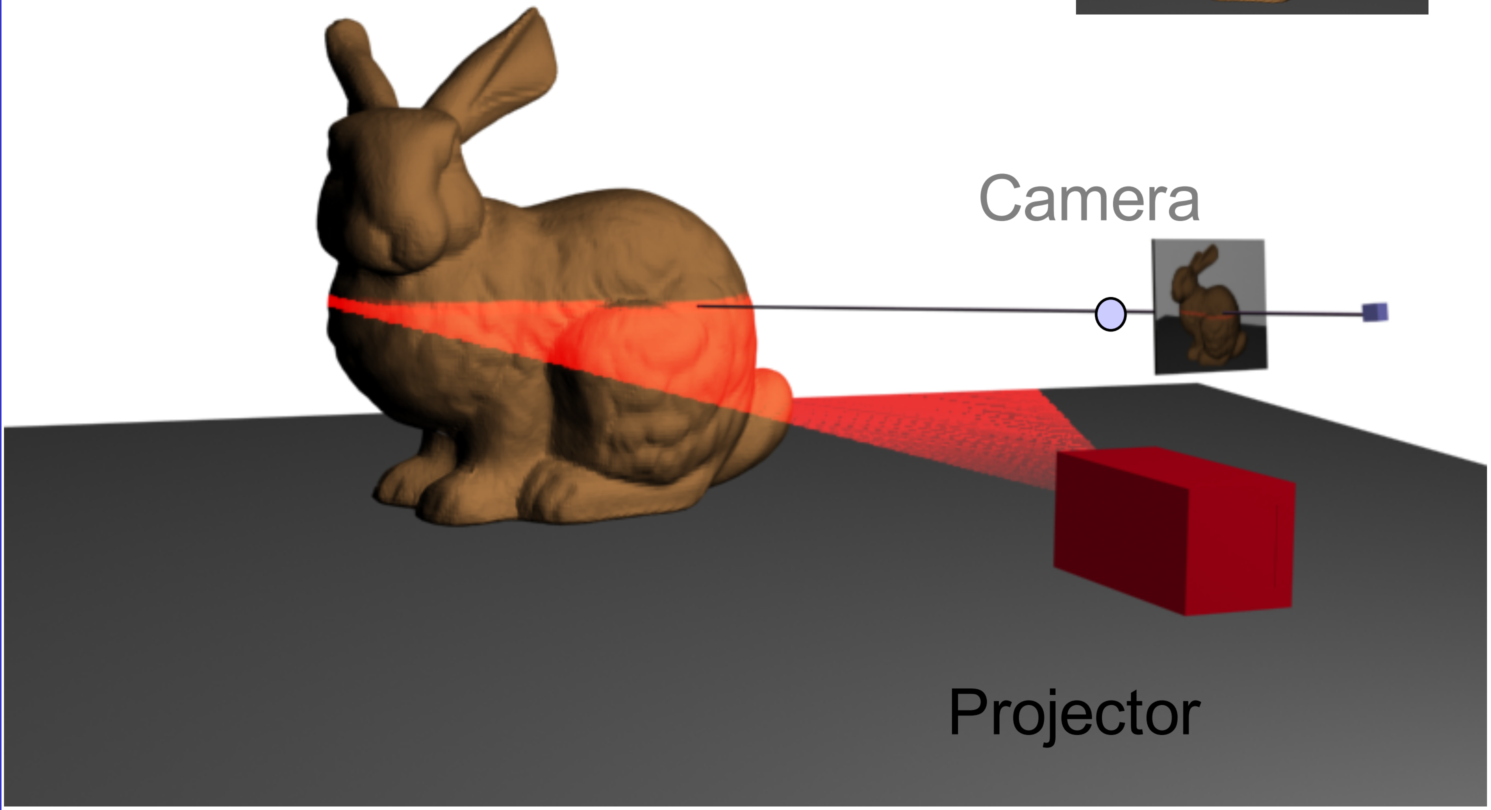
Active triangulation

Computer Vision

Active triangulation

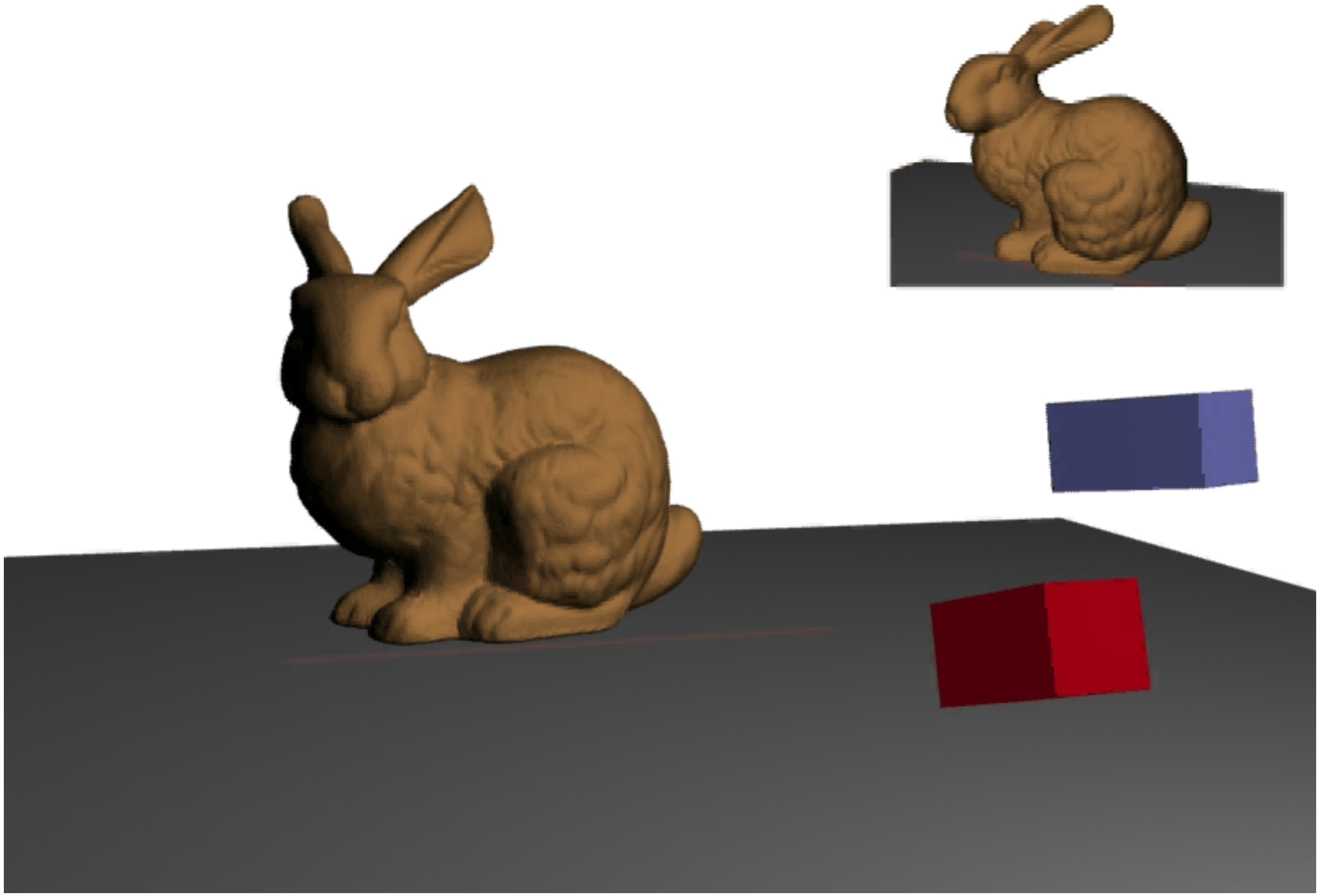
Triangulation → 3D measurements





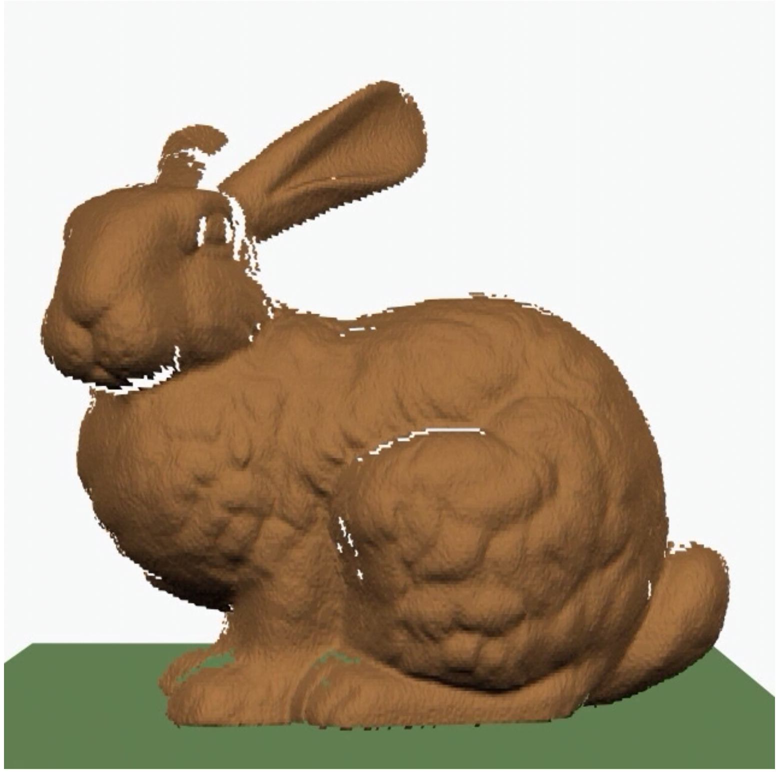
Active triangulation

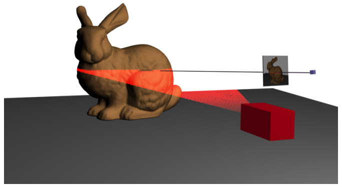
Camera image



Computer Vision

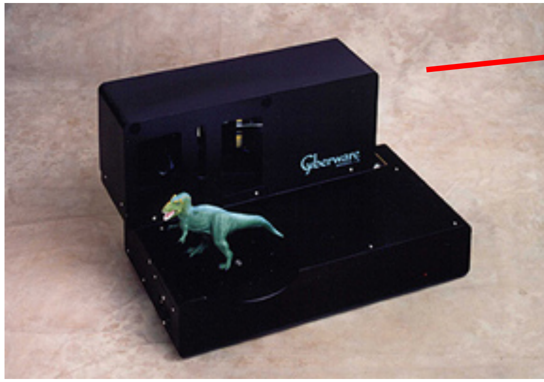
Active triangulation





Active triangulation

Example 1 Cyberware laser scanners



Desktop model for small objects

Medium-sized objects

Body scanner

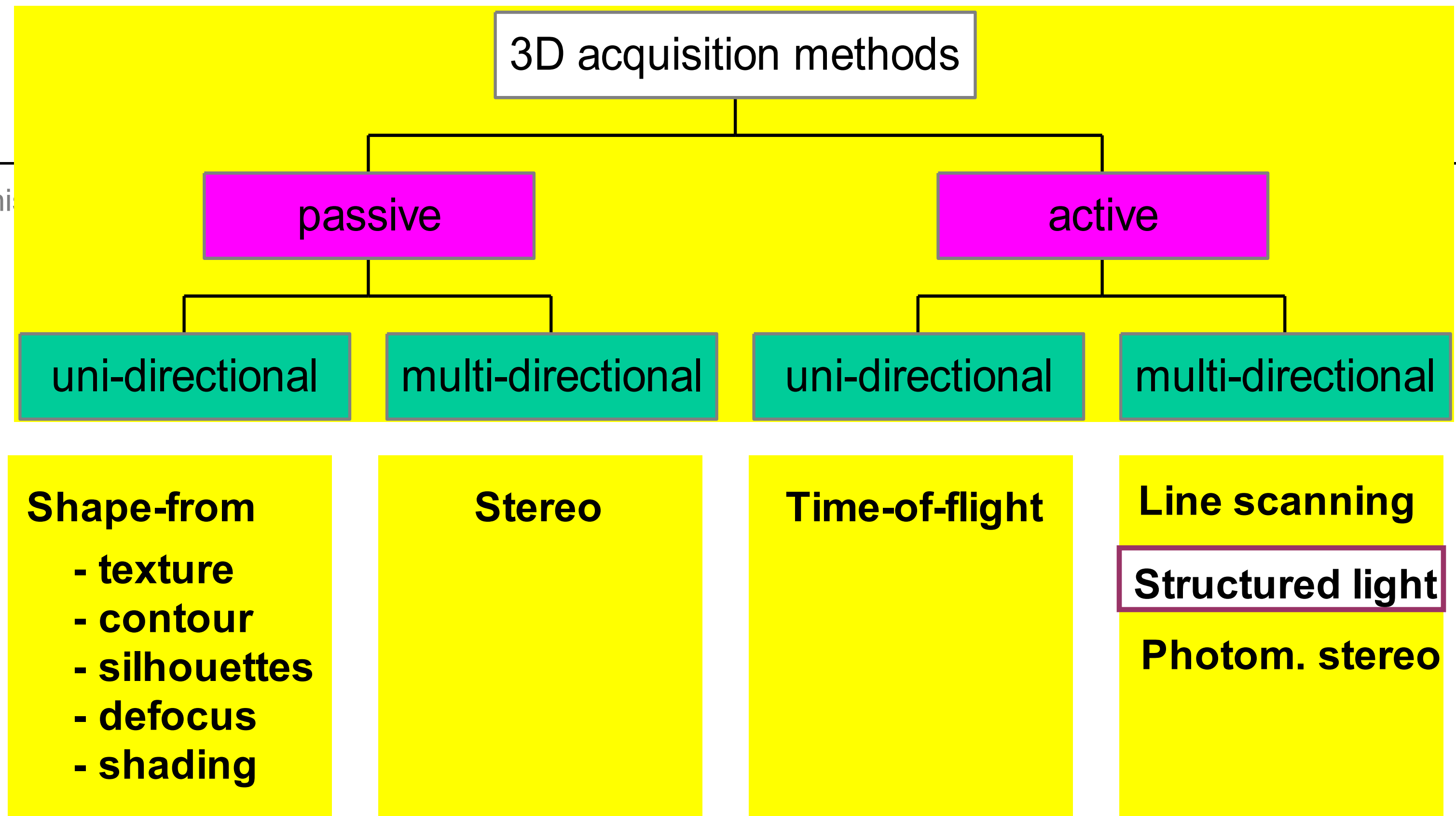
Head scanner

Active triangulation

Example 2 Minolta

Portable desktop model

3D acquisition taxonomy



Structured light

patterns of a special shape are projected onto the scene

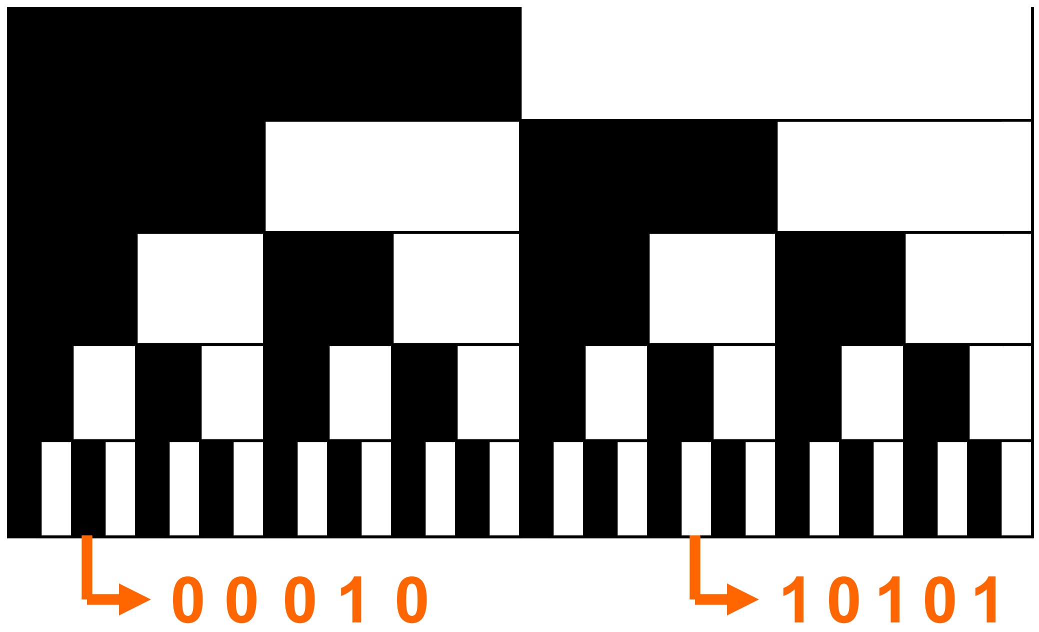
deformations of the patterns yield information on the shape

Focus is on combining a good resolution with a minimum number of pattern projections

Serial binary patterns

A sequence of patterns with increasingly fine subdivisions

Yields 2^n identifiable lines for only n patterns



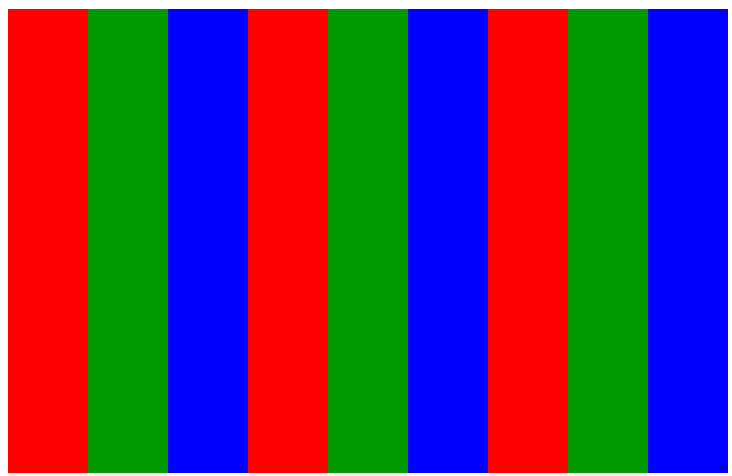
Reducing the nmb of projections: colour

Binary patterns

Yields 2^n identifiable lines for only n patterns

Using colours, e.g. 3,

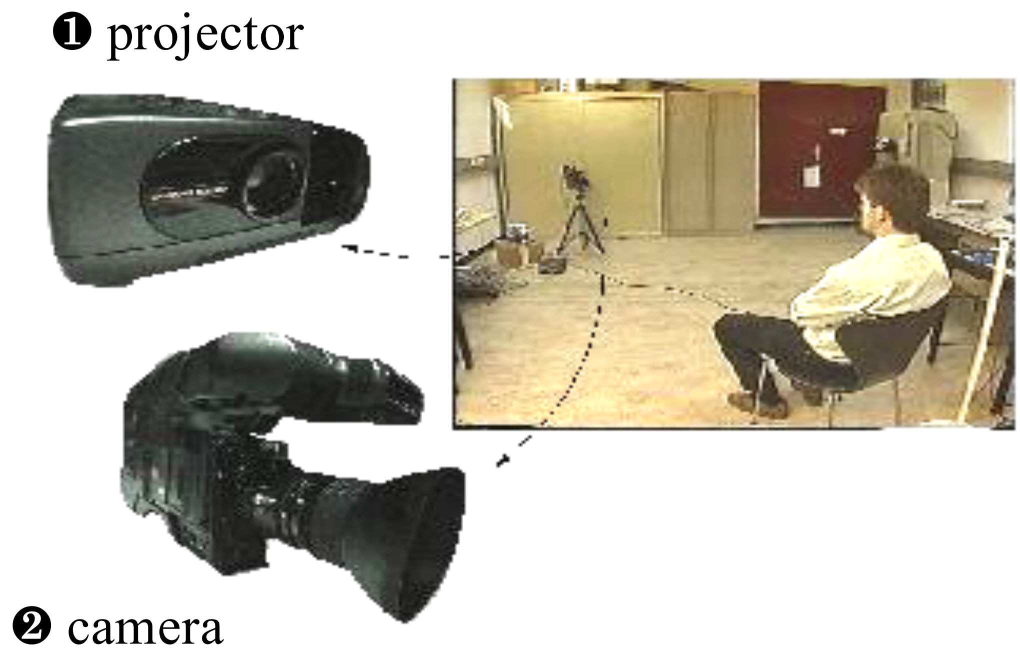
Yields 3^n identifiable lines for only n patterns



Interference from object colours...

One-shot implementation

3D from a single frame – KULeuven ‘96:



Computer Vision

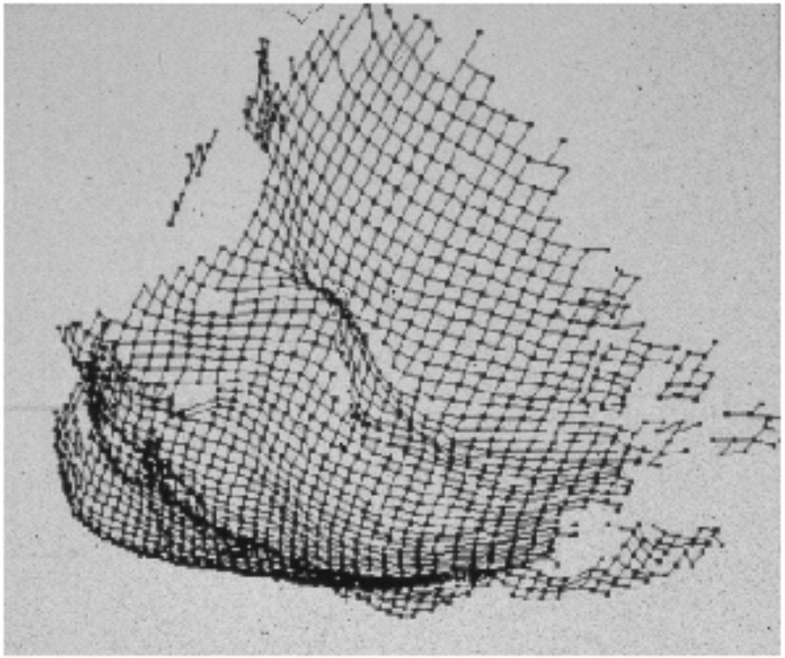
One-shot implementation

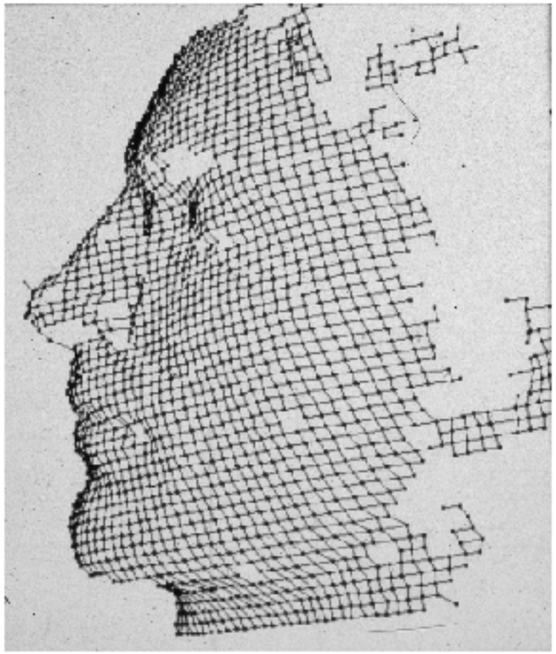
KULeuven '81: checkerboard pattern with column code

example :

Computer Vision

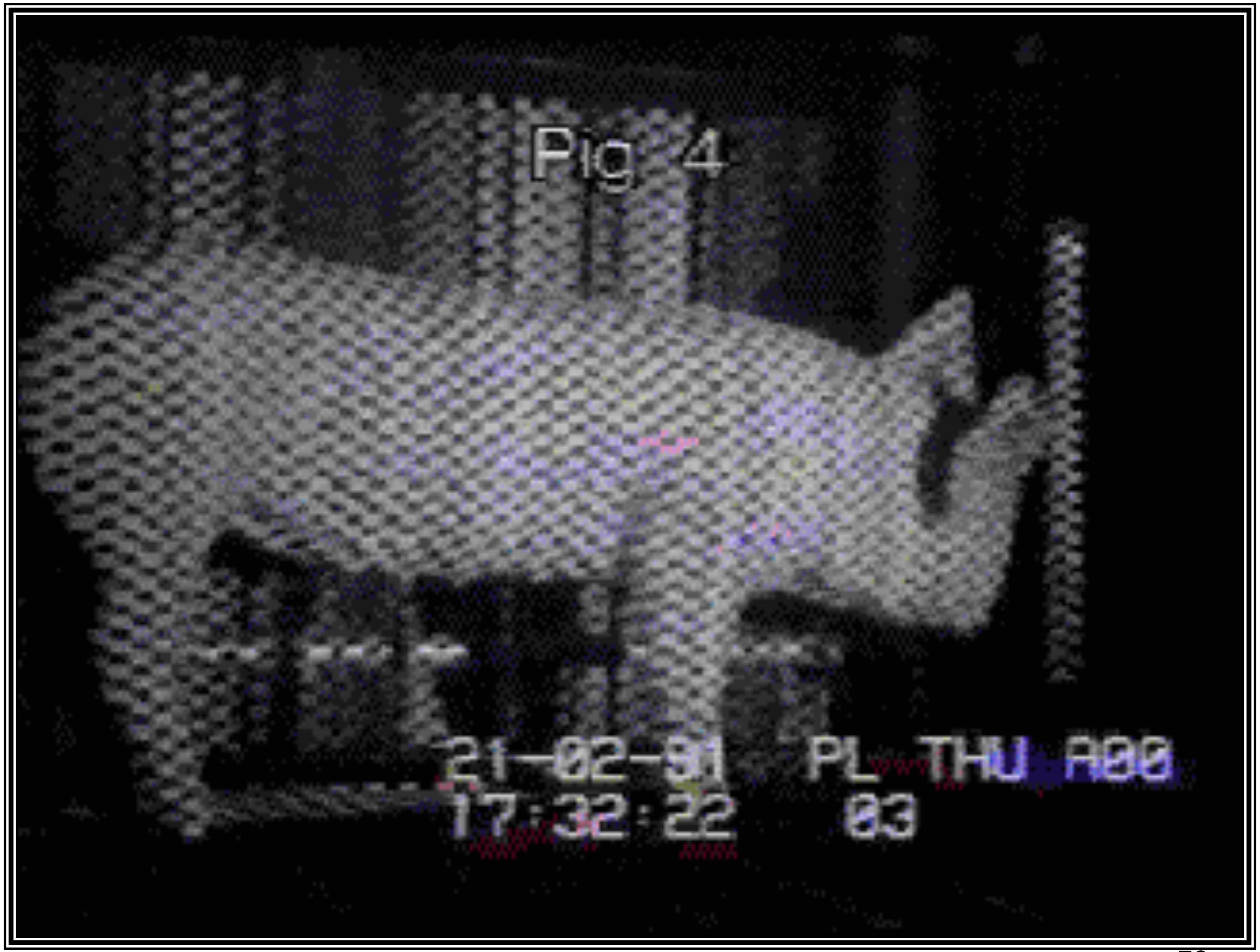
3D reconstruction for the example





Computer Vision

An application in agriculture



Computer Vision

One-shot 3D acquisition

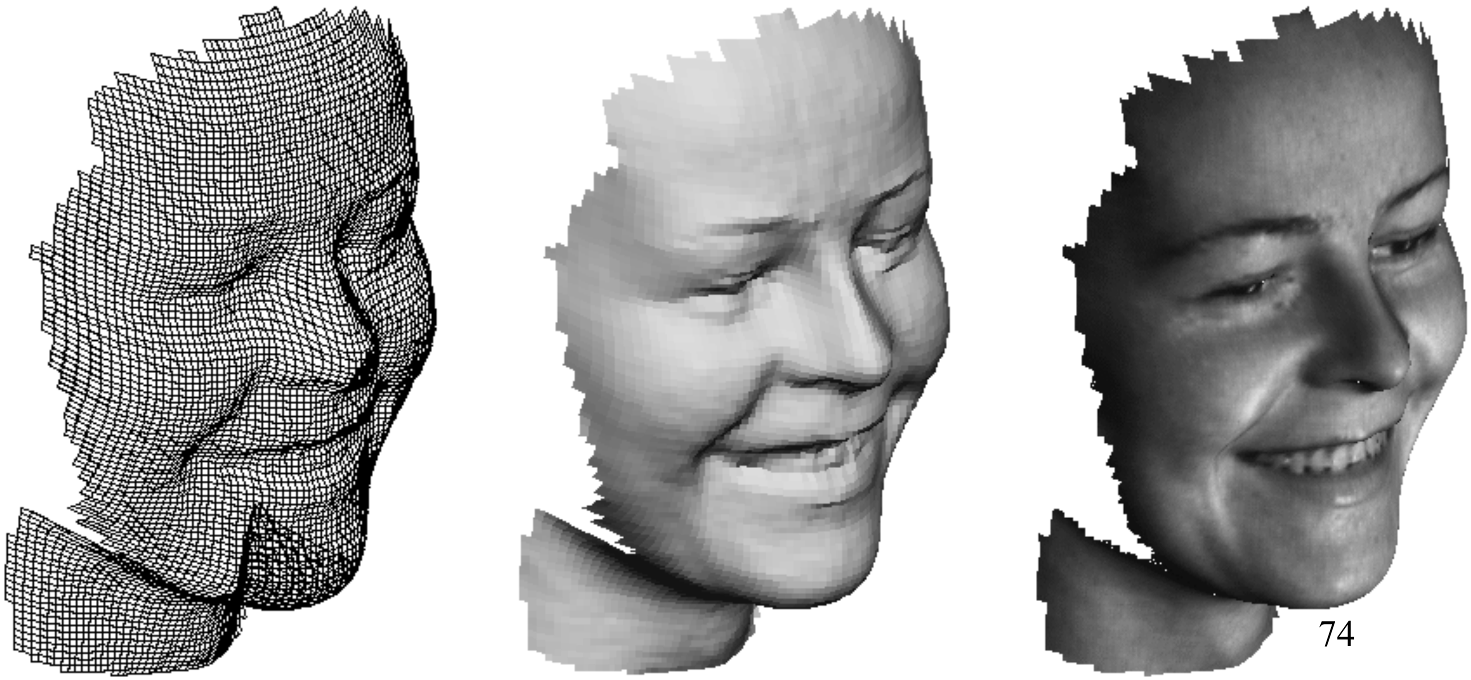
Leuven ShapeCam

Computer Vision

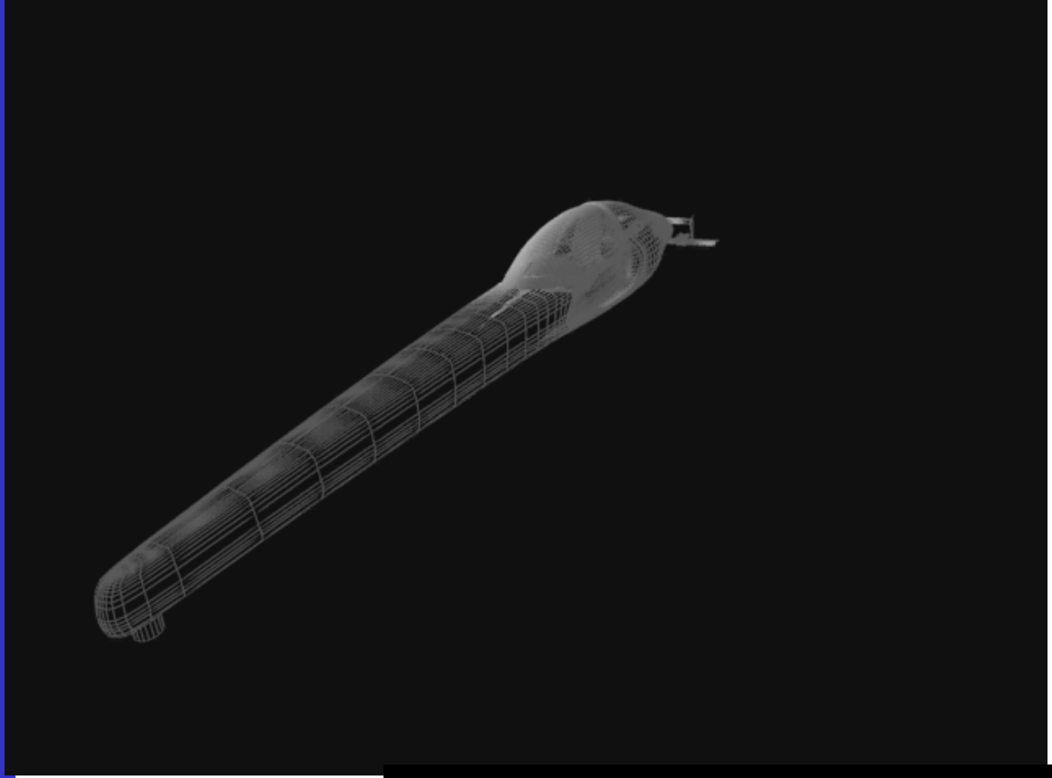
Shape + texture often needed

Higher resolution

Texture is also extracted



Computer Vision

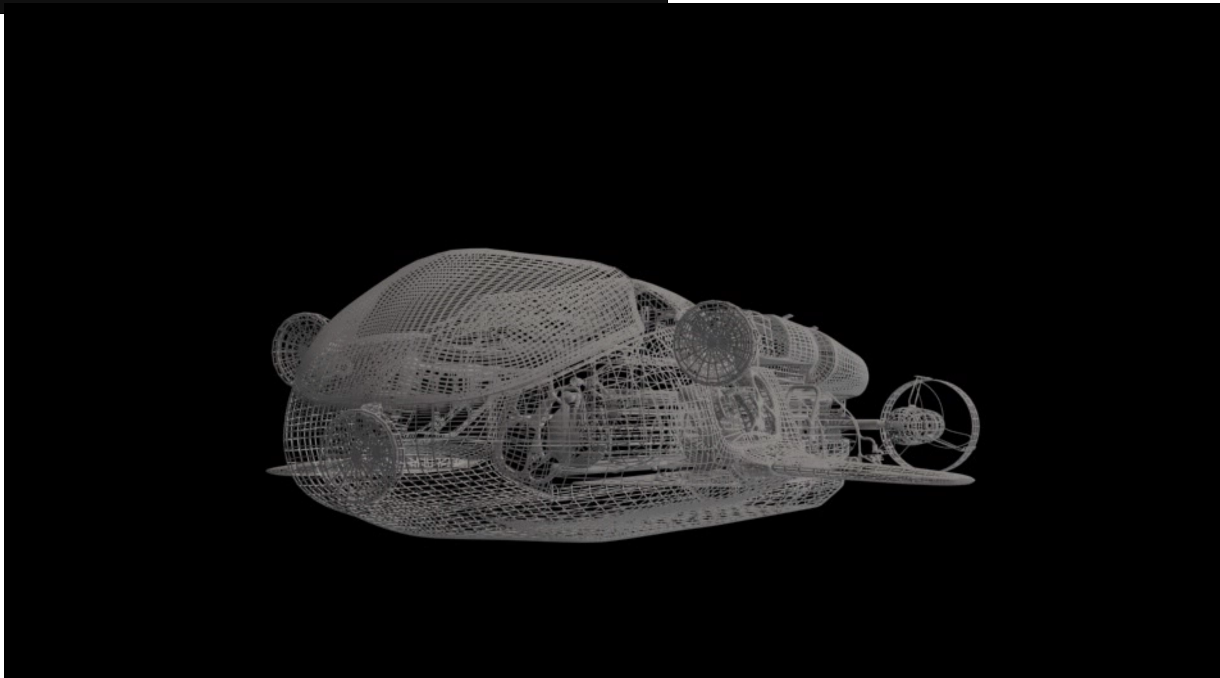


James Bond

Die another day

Lara Croft

Thomb Raider



Computer Vision

Active triangulation

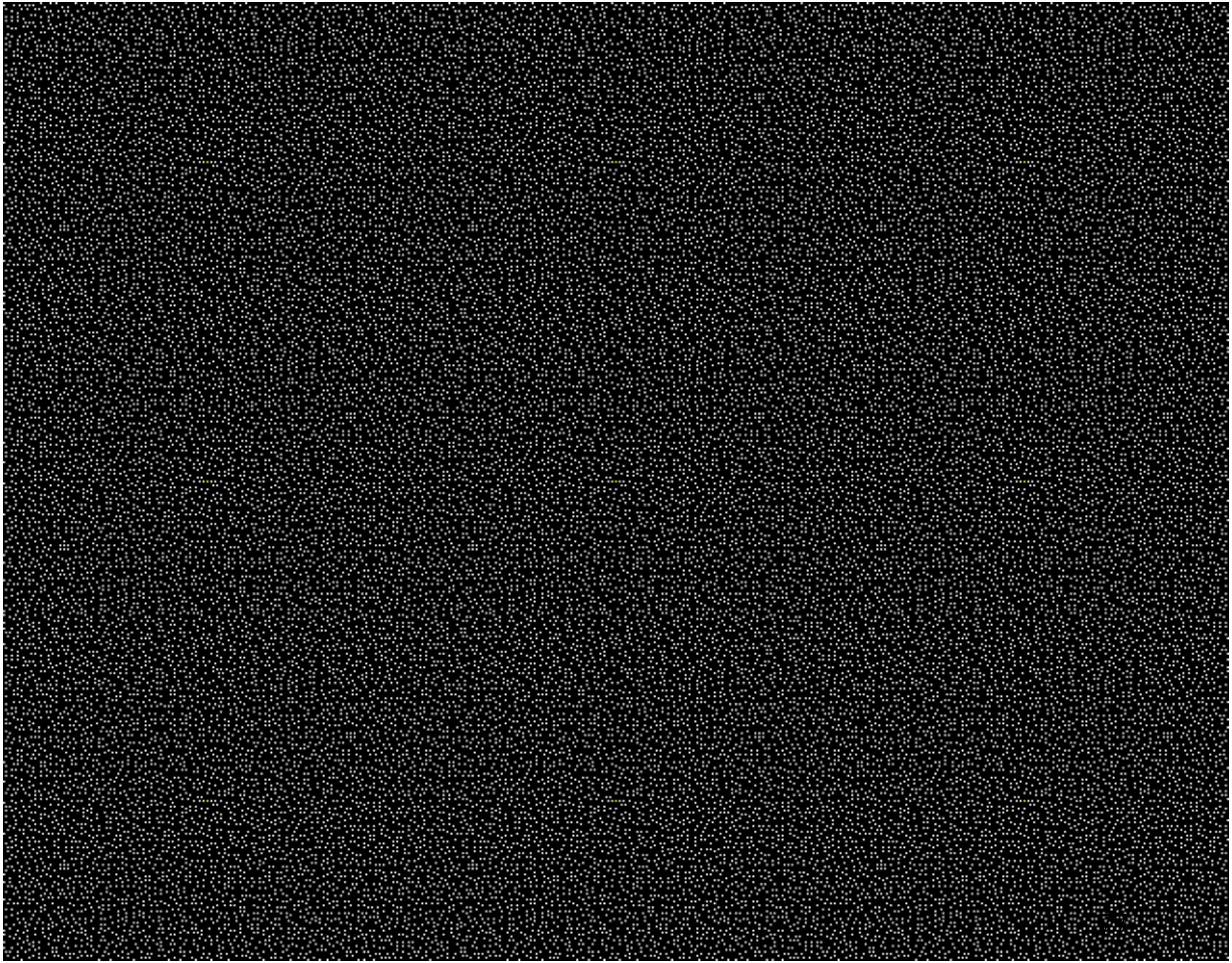
Recent, commercial example

Kinect 3D camera, affordable
and compact solution by Microsoft.

Projects a 2D point pattern in the NIR,
to make it invisible to the human eye

Computer Vision

Kinect: 9x9 patches with locally unique code



Computer Vision

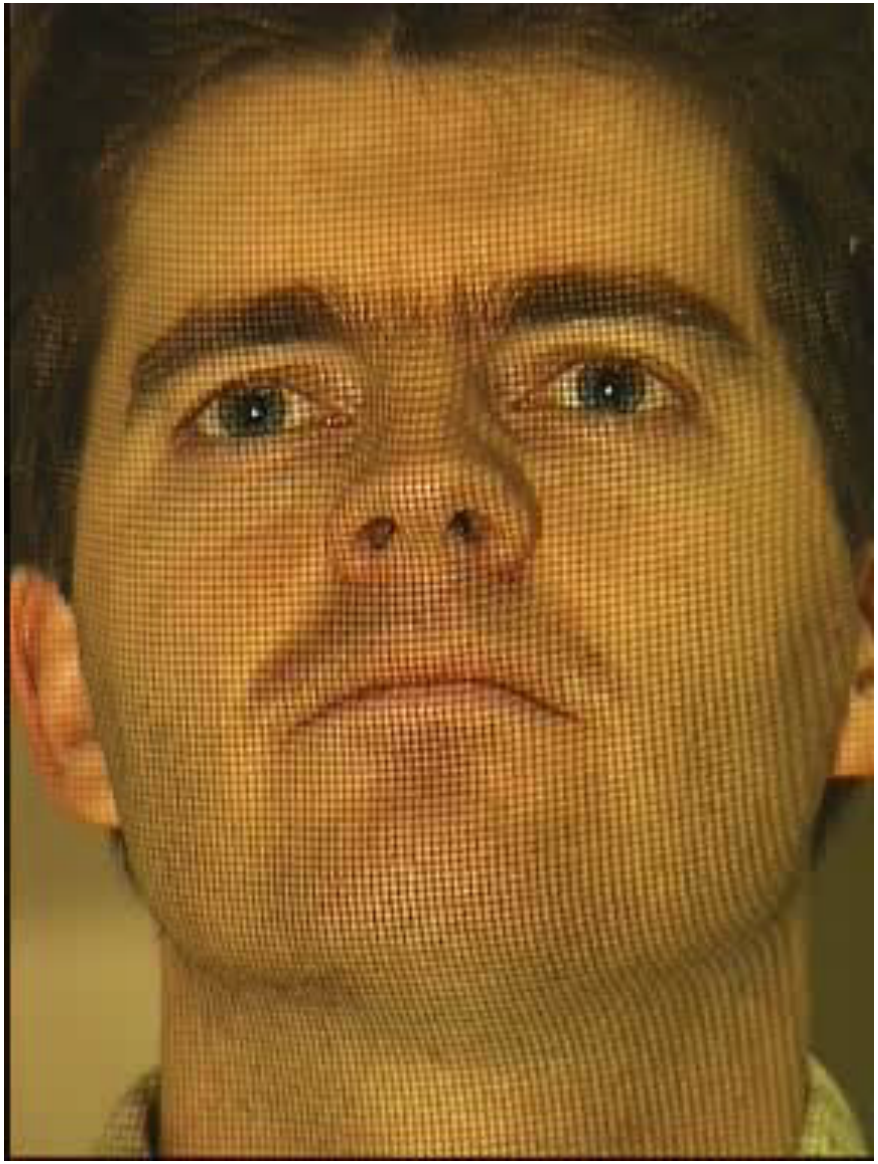
Kinect as one-shot, low-cost scanner

Excerpt from the dense NIR dot pattern:

http://research.microsoft.com/apps/video/default.aspx?i
15

Computer Vision

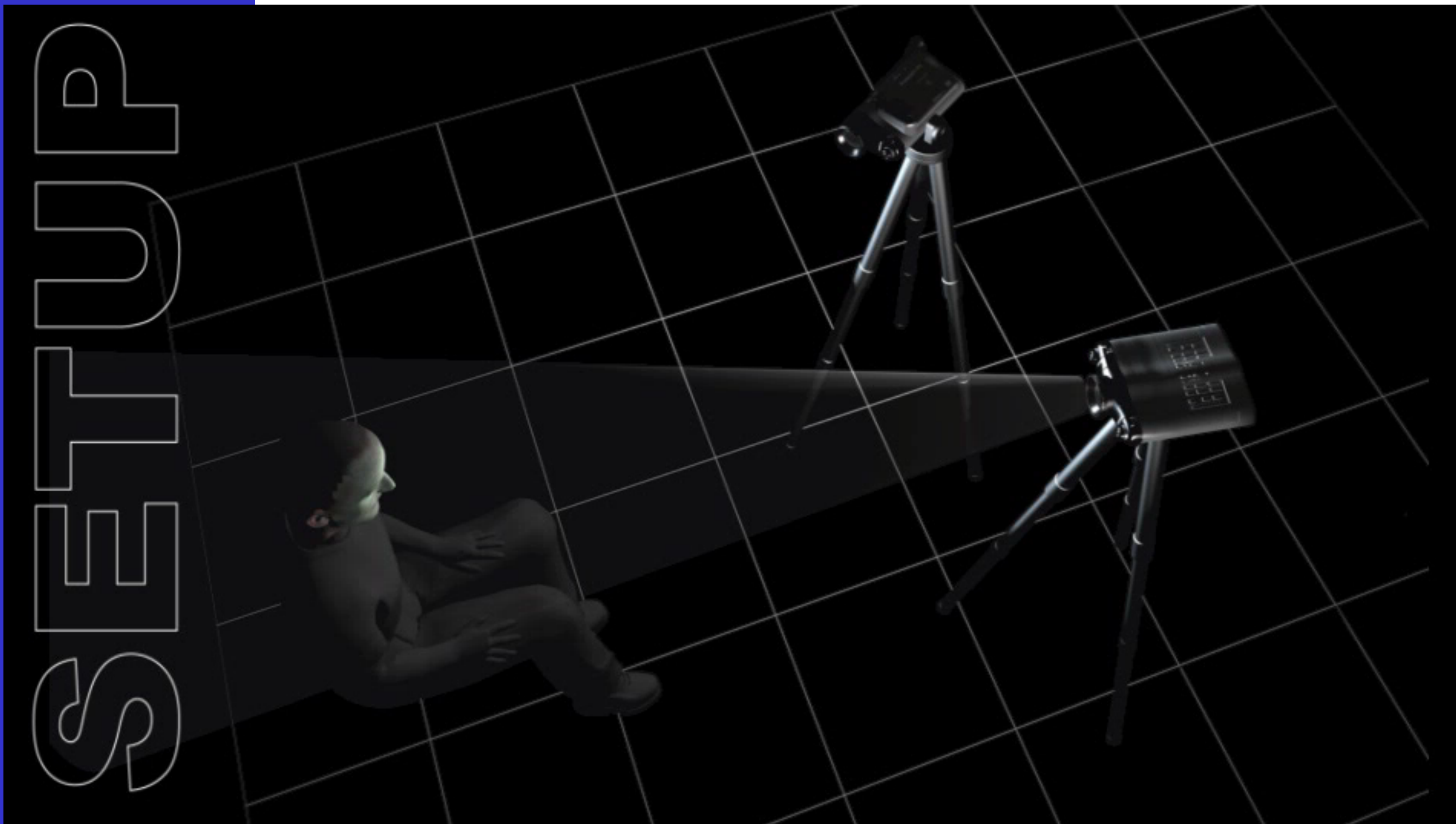
Face animation - input



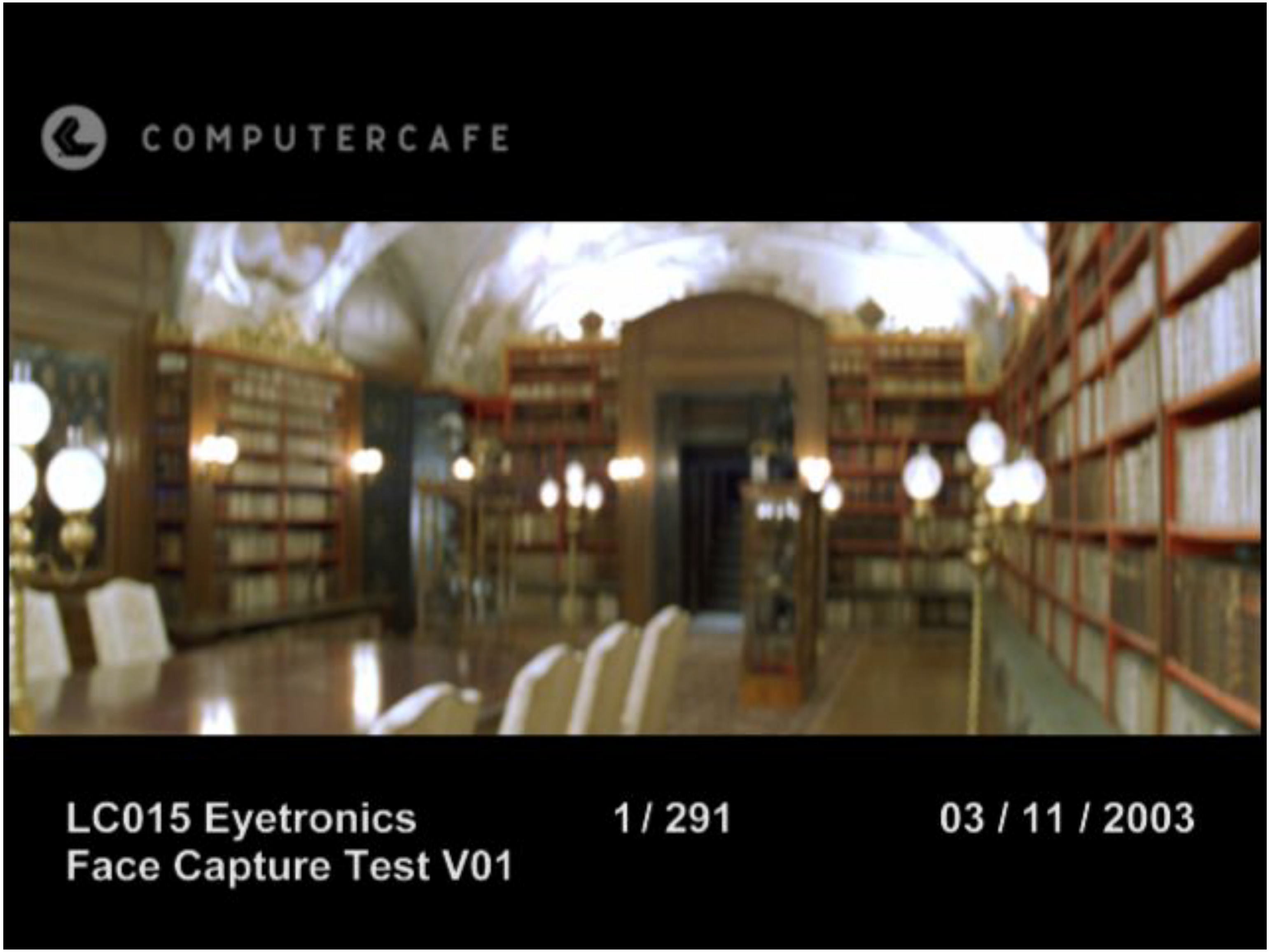
Face animation – replay + effects

Facial motion capture

motion capture for *League of Extraordinary Gentlemen*

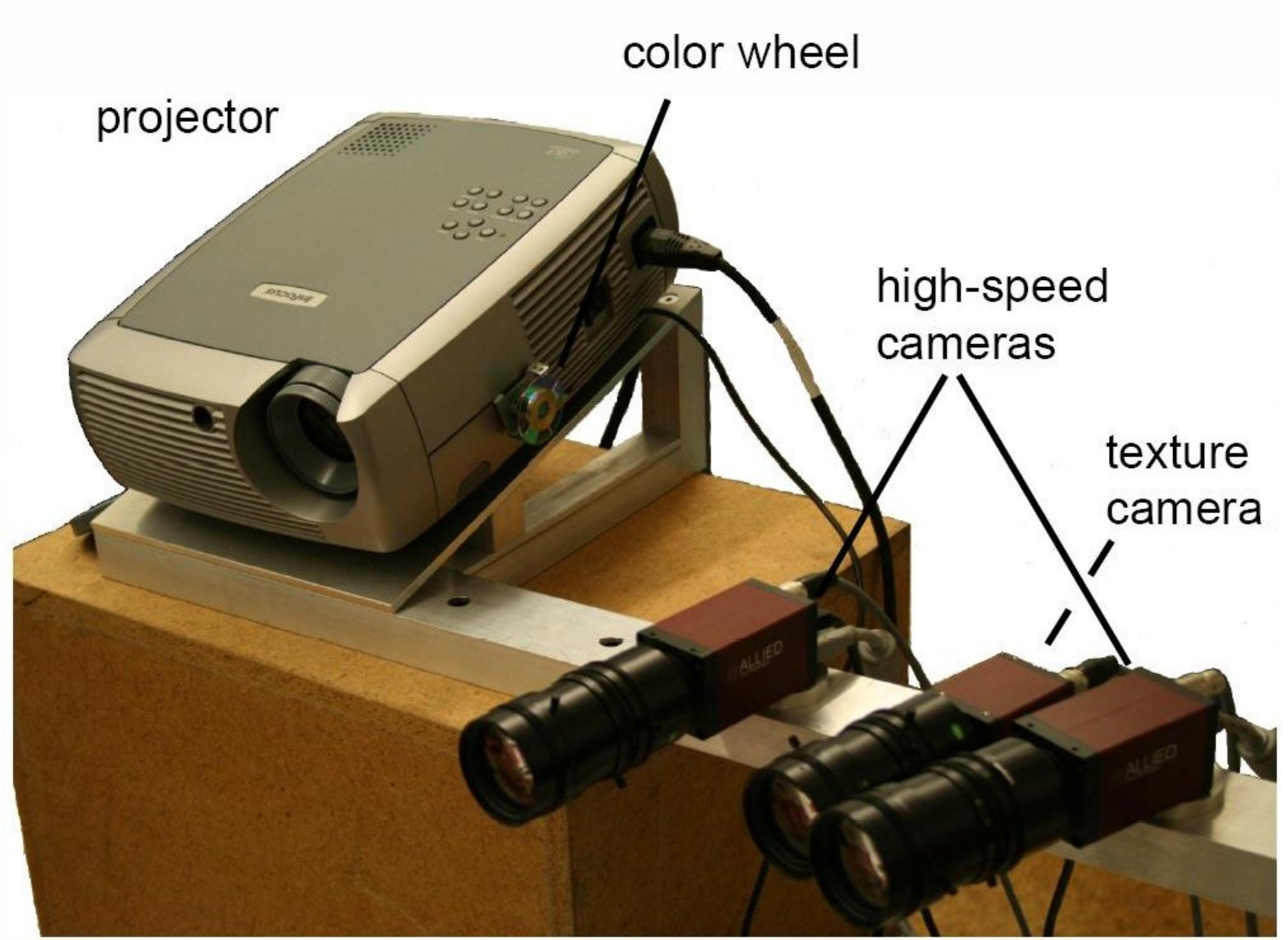


Facial motion capture



Computer Vision

Phase shift



Phase shift

$$I_r = A + R\cos(\phi - \theta)$$

$$I_g = A + R\cos(\phi)$$

$$I_b = A + R\cos(\phi + \theta)$$

1. detect phase from 3 subsequently projected cosine patterns, shifted over 120 degrees
2. unwrap the phases / additional stereo
3. texture is obtained by summing the 3 images / color camera w. slower integration

Phase shift

$$A = \frac{I_r + I_g + I_b}{3}$$

$$\phi = \arctan\left(\tan\left(\frac{\theta}{2}\right) \frac{I_r - I_b}{2I_g - I_r - I_b} \right)$$

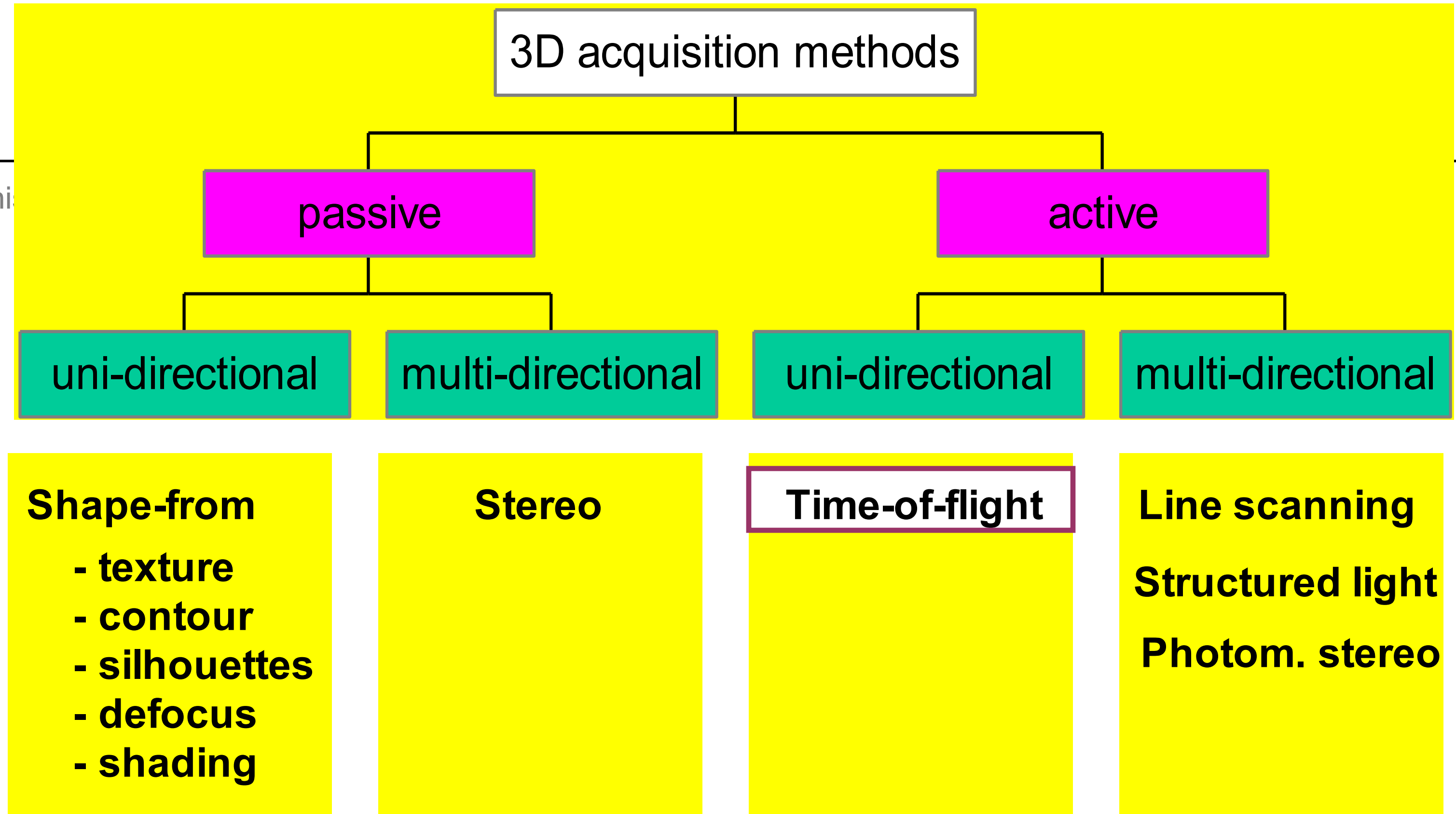
4D acquisition

Motion retargetting, from 3D phase shift scans

Face/Off: Live Facial Puppetry

PaperID 102

3D acquisition taxonomy



Time-of-flight

measurement of the time a modulated light signal needs to travel before returning to the sensor

this time is proportional to the distance

waves :

1. *radar*	low freq. electromagnetic
2. *sonar*	acoustic waves
3. *optical radar*	optical waves

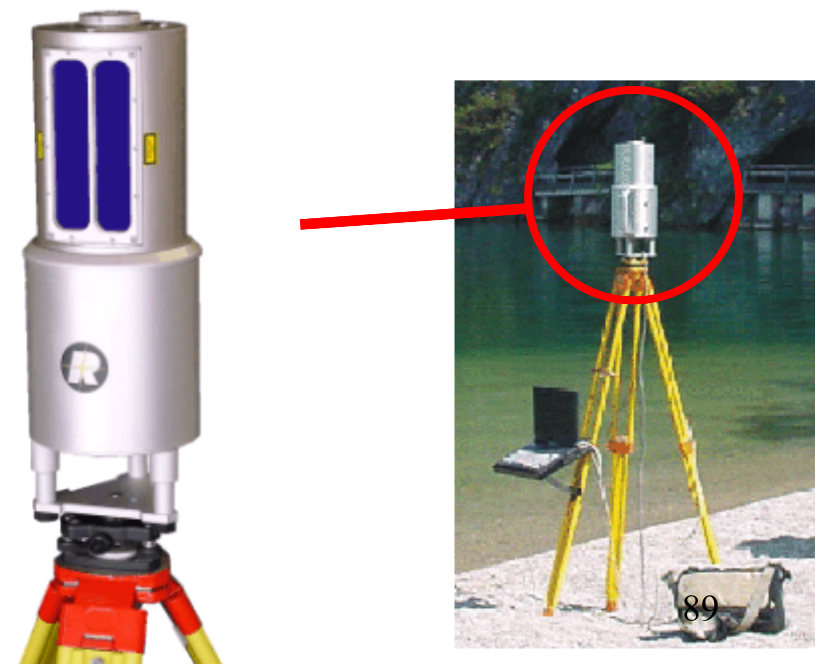
working principles :

1. pulsed
2. phase shifts

Time-of-flight

Example 1: Cyrax

Example 2: Riegl



Time-of-flight: example

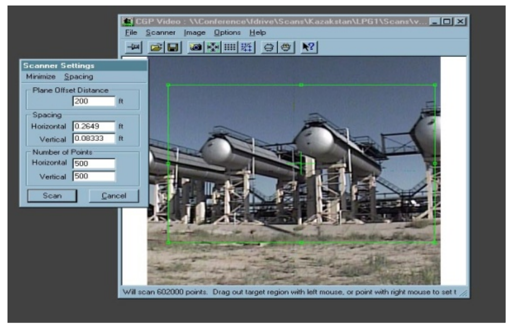
Cyrax ™

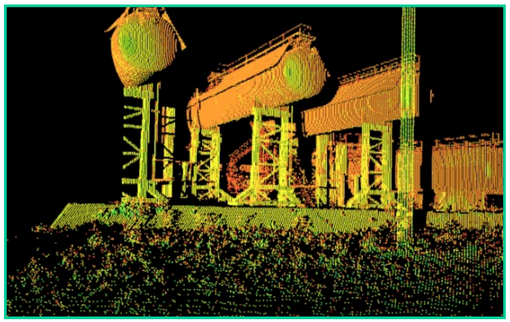
3D Laser Mapping

System

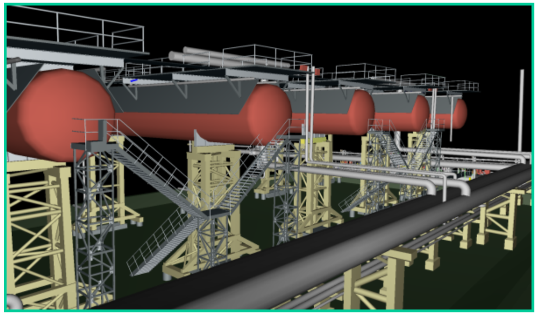
Cyrax

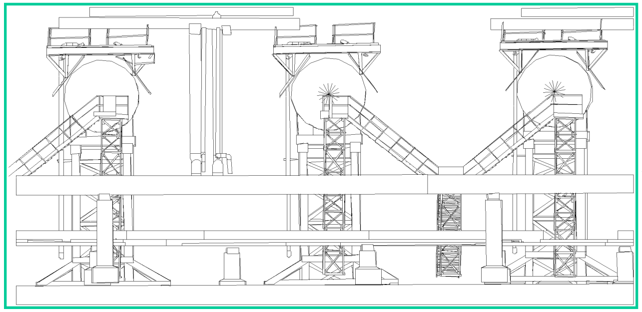
Accurate, detailed, fast measuring





2D / 3D CAD

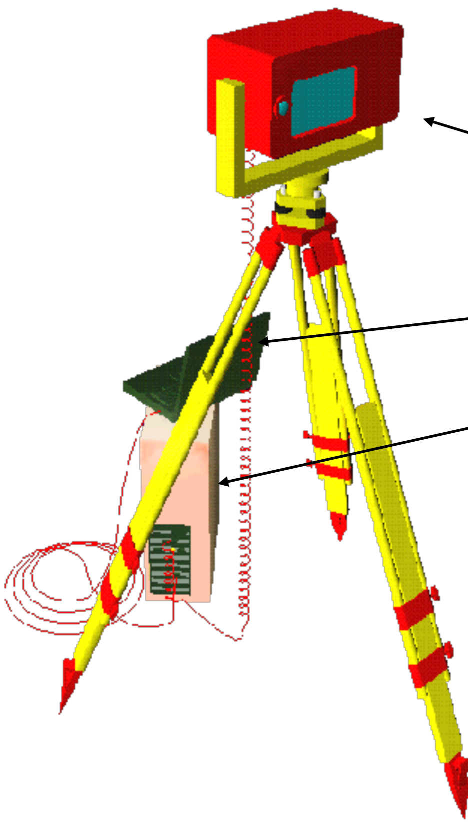




Integrated modeling

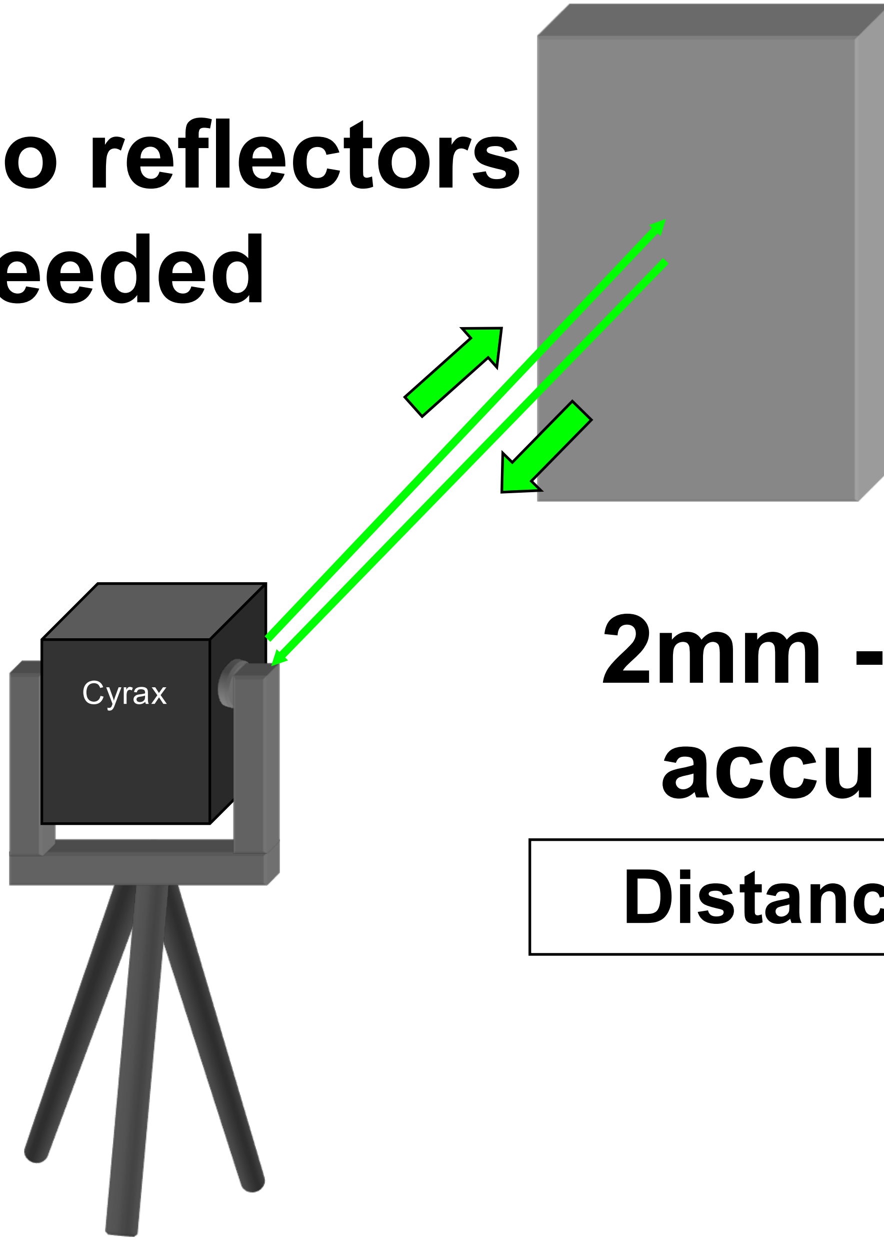
Computer Vision

Cyrax



- 3 Components
  - Laser Scanner
  - Laptop Computer & Software
  - Electronics & Power Supply
- Field Portable

Pulsed laser (time-of-flight)

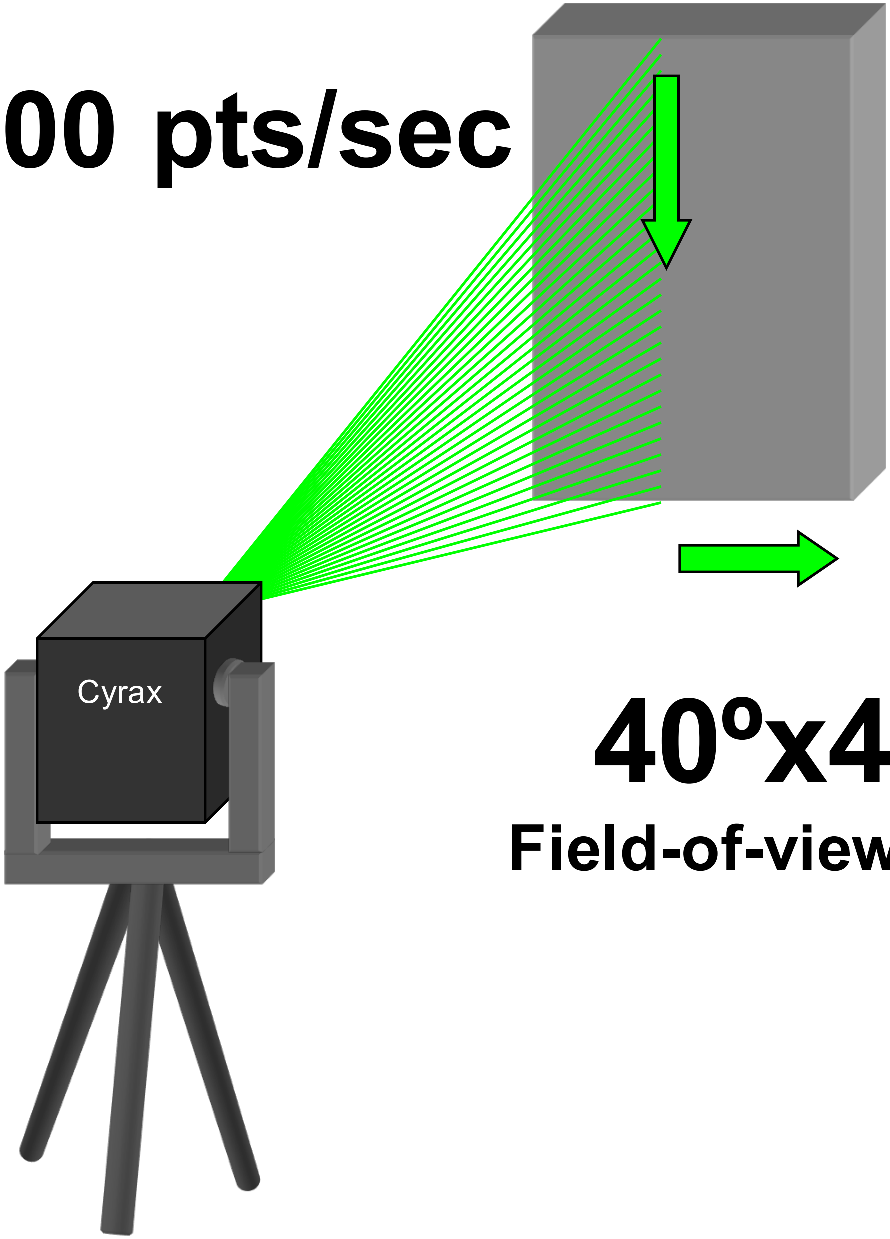


No reflectors needed

2mm - 6mm accuracy

Distance = C x ΔT÷2

Laser sweeps over surface



800 pts/sec

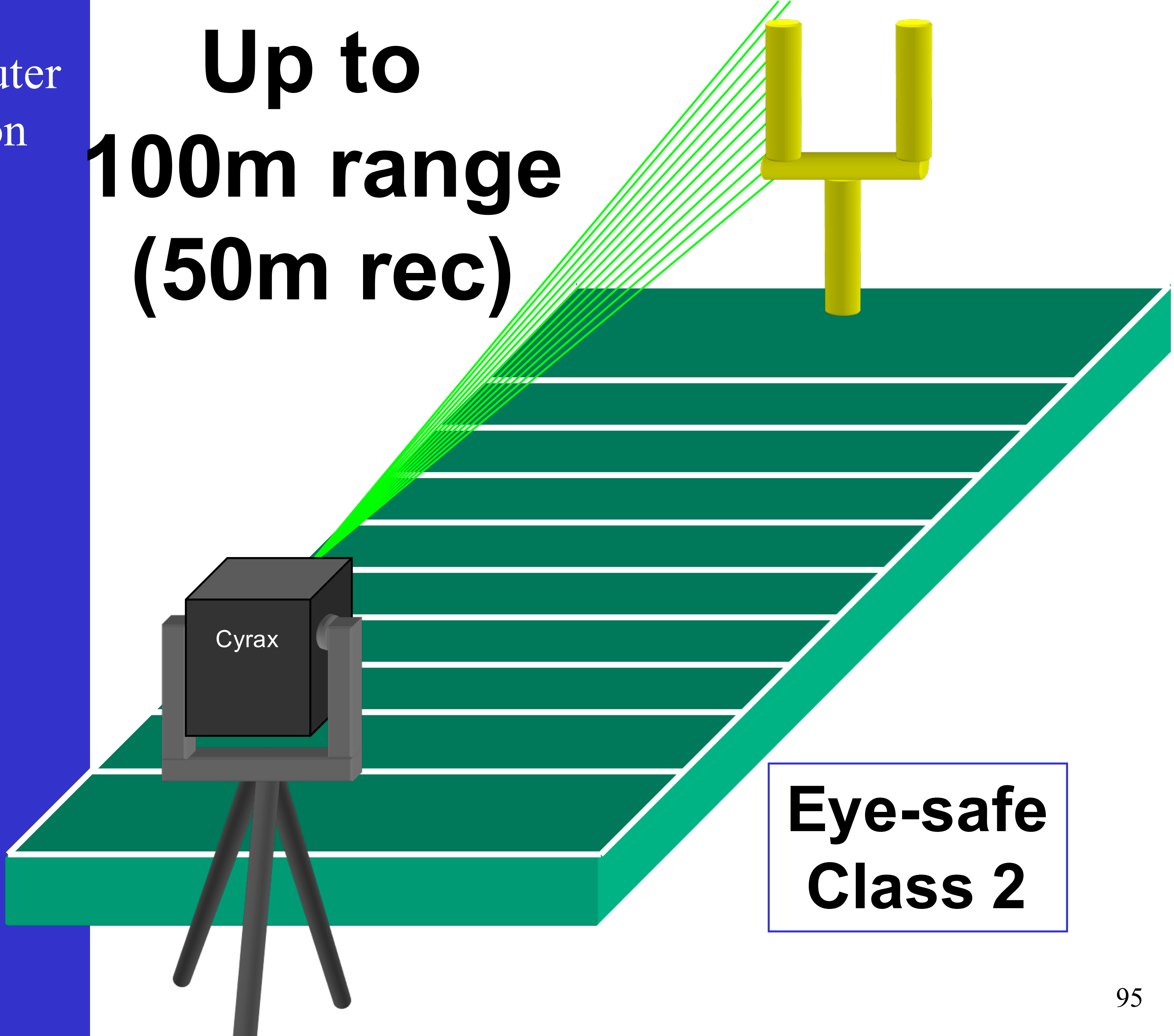
2mm min pt-to-pt spacing

40ºx40º

Field-of-view (max)

Computer Vision

Up to 100m range (50m rec)

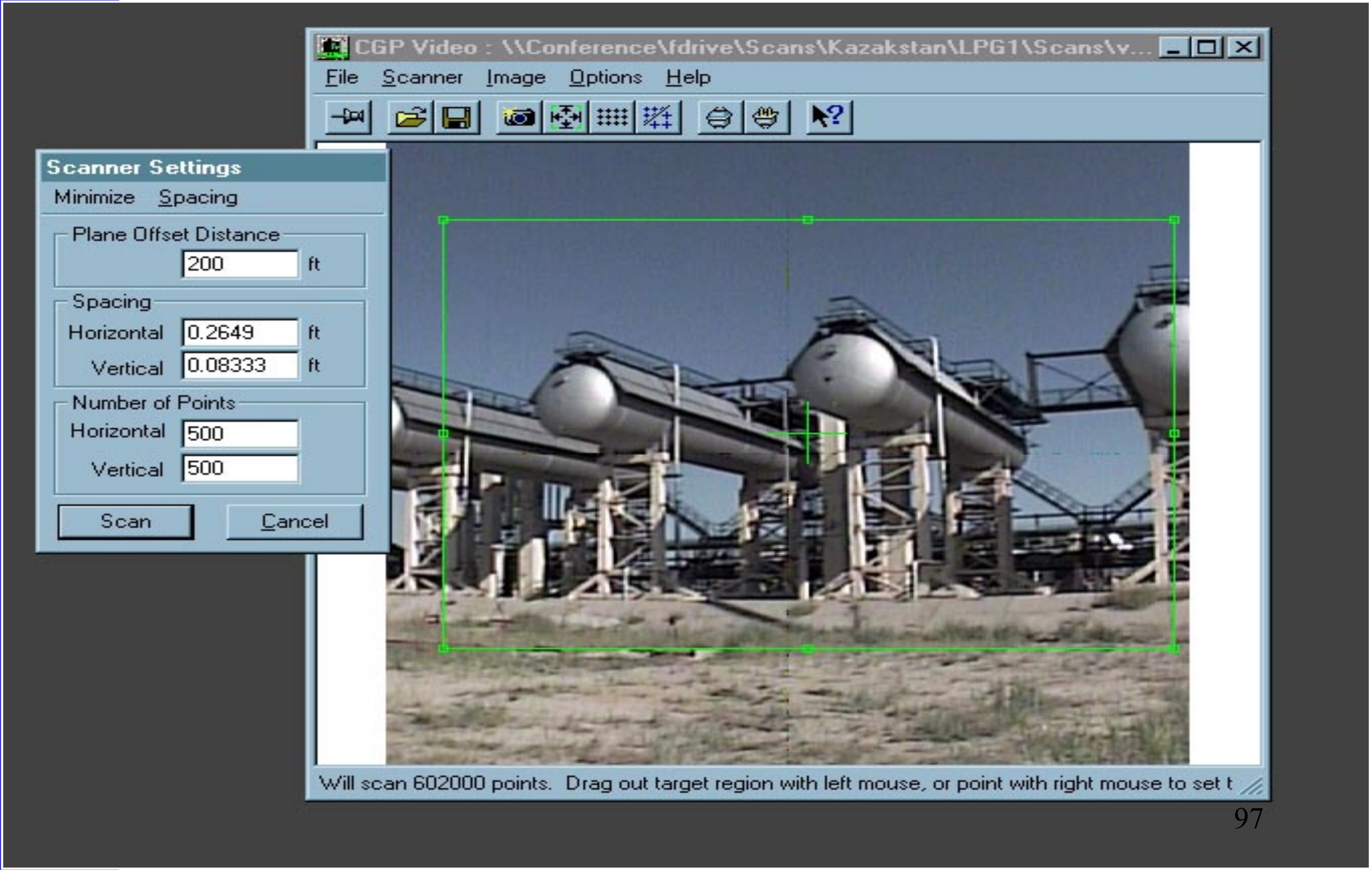


Eye-safe Class 2

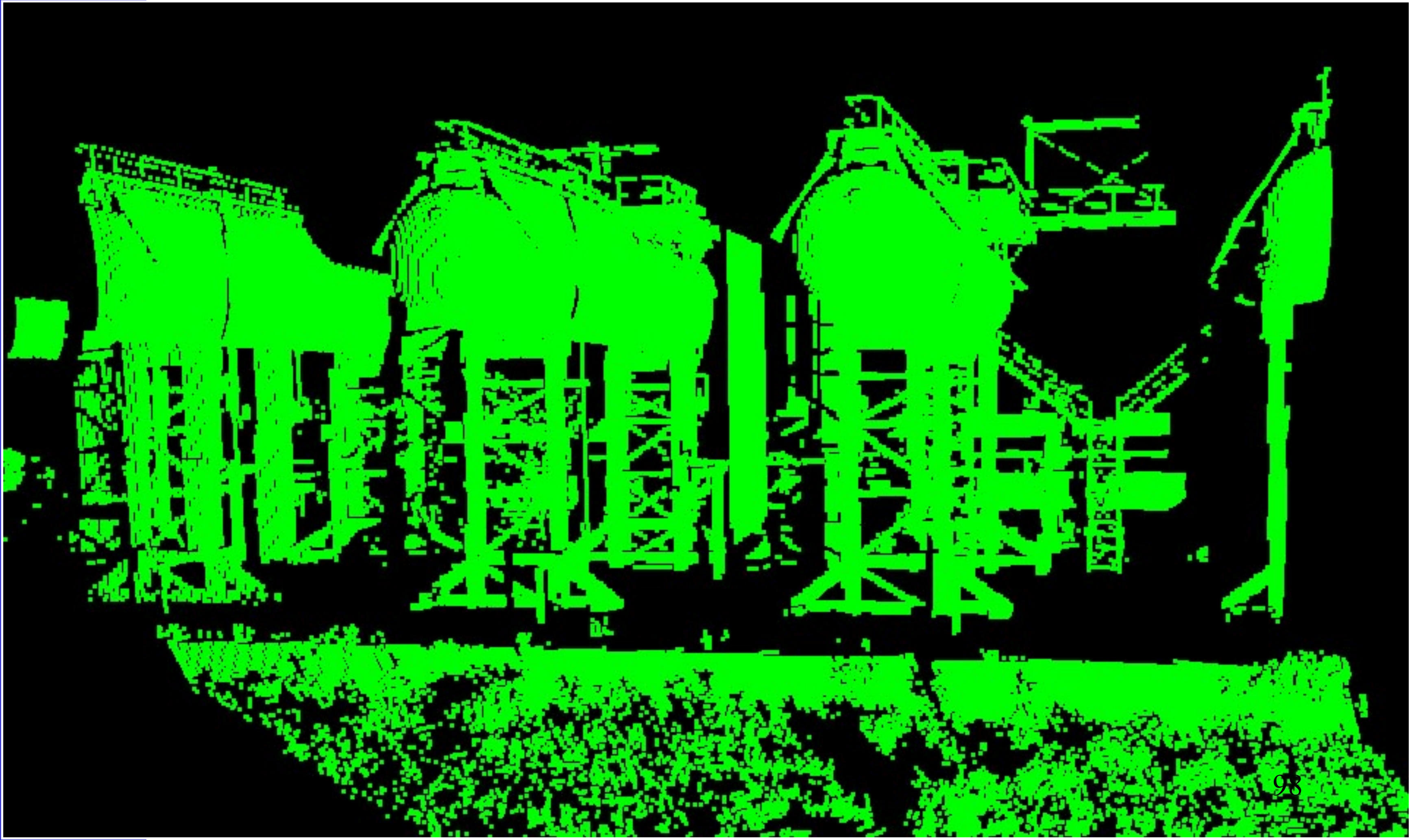
Cyrax is *also* a visualization tool

Cyrax detects the intensity of each reflected laser pulse and colors it

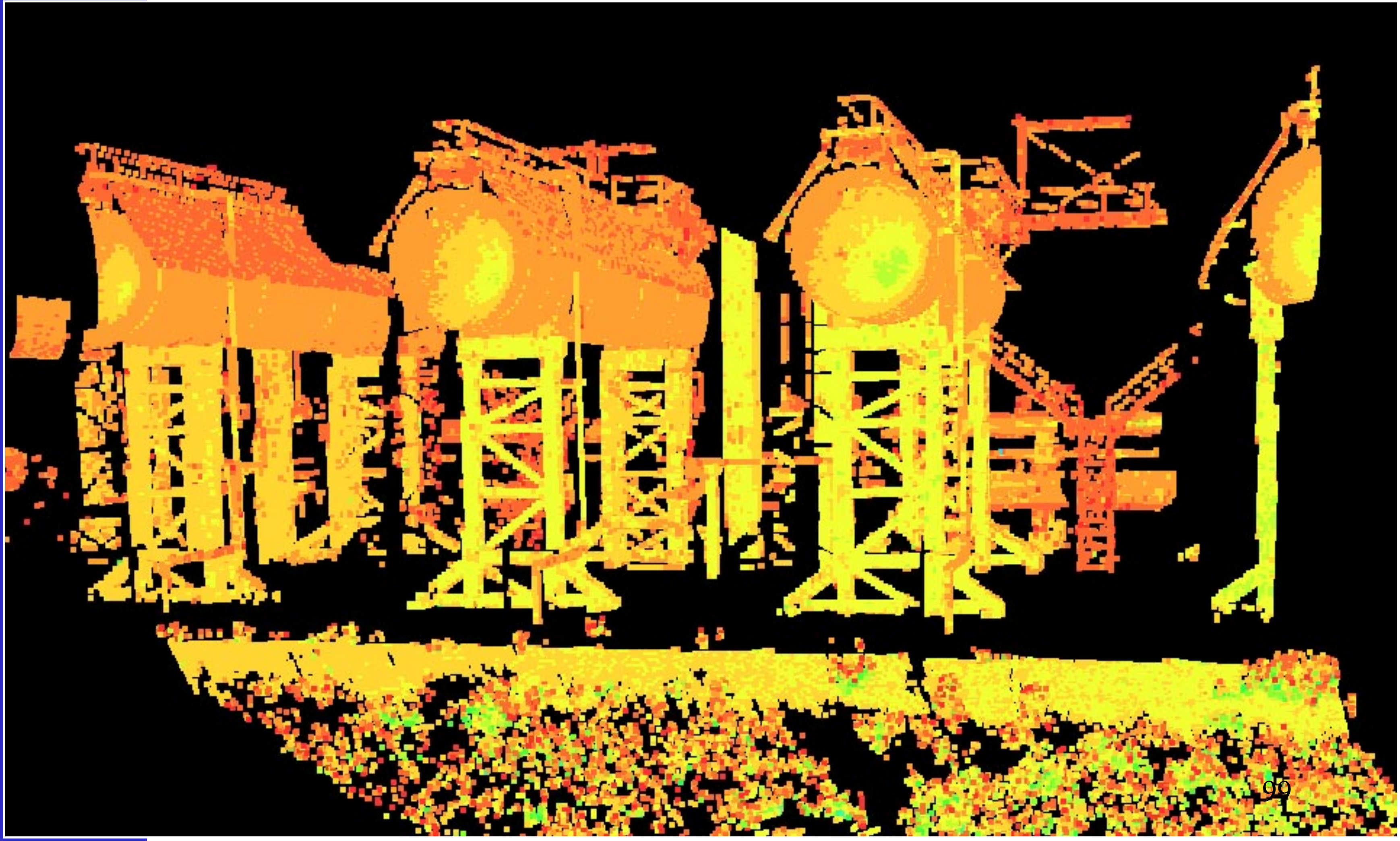
Step 1:
Target the structure



Step 2:
Scan the structure

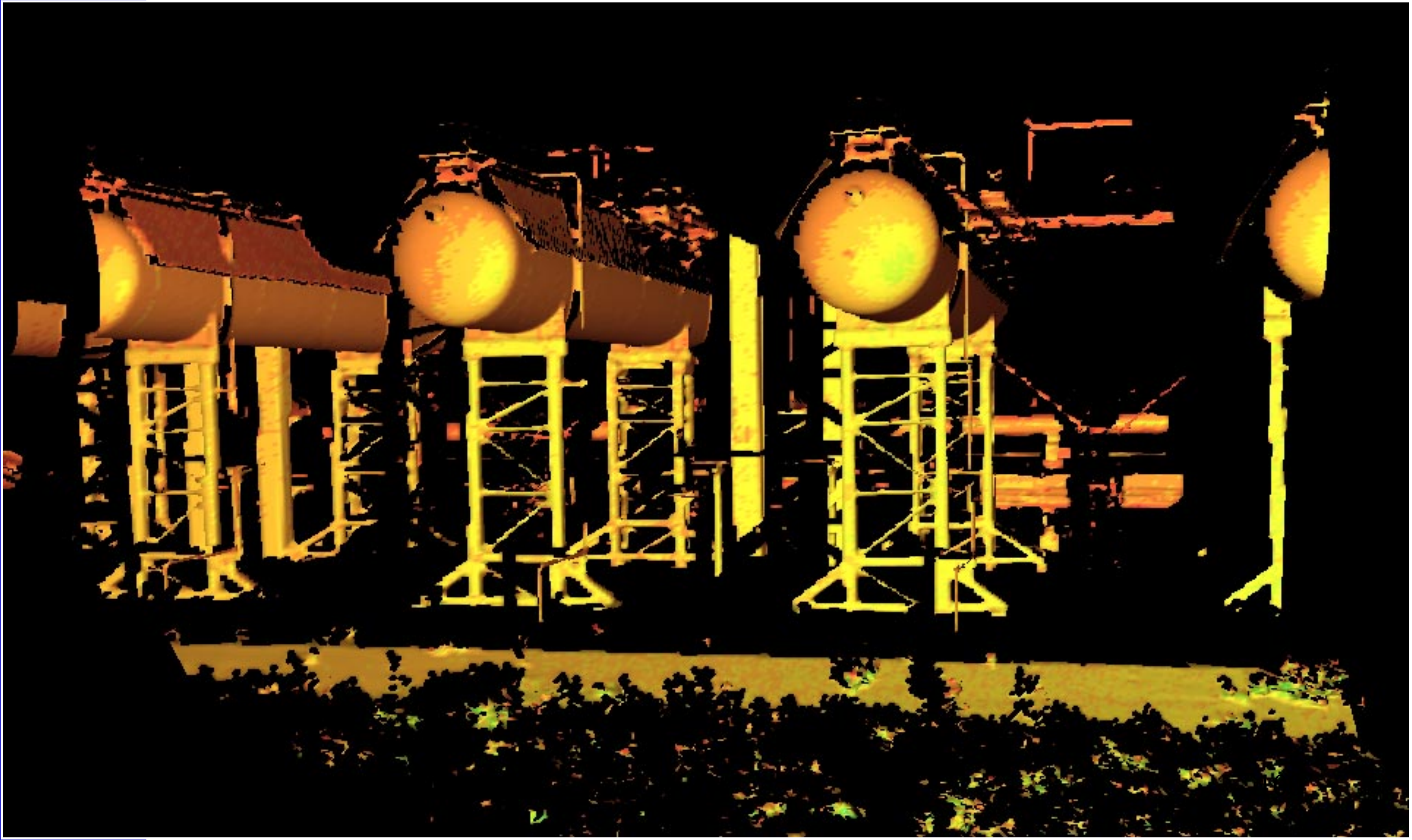


Step 3:
Color the points



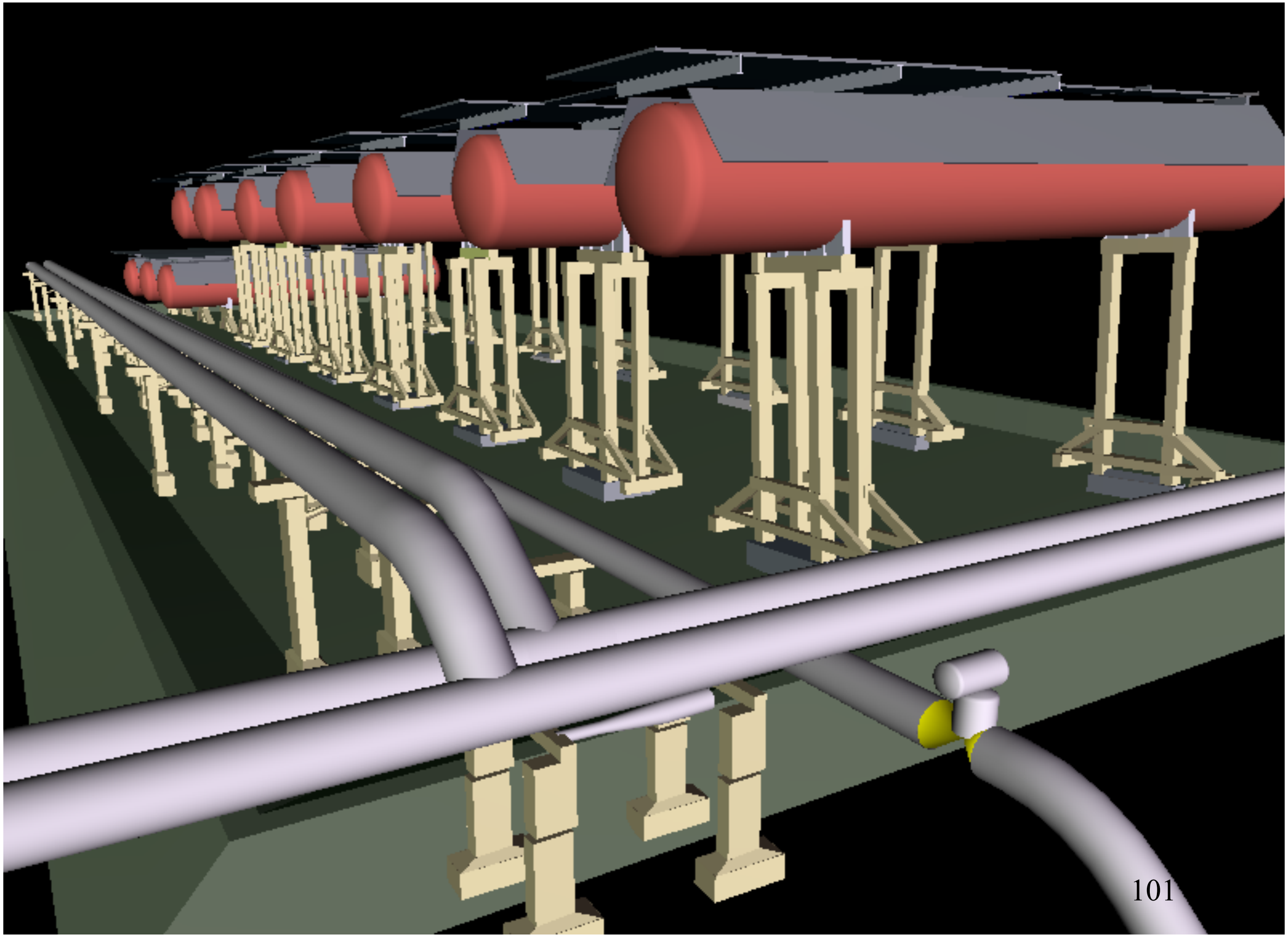
Computer Vision

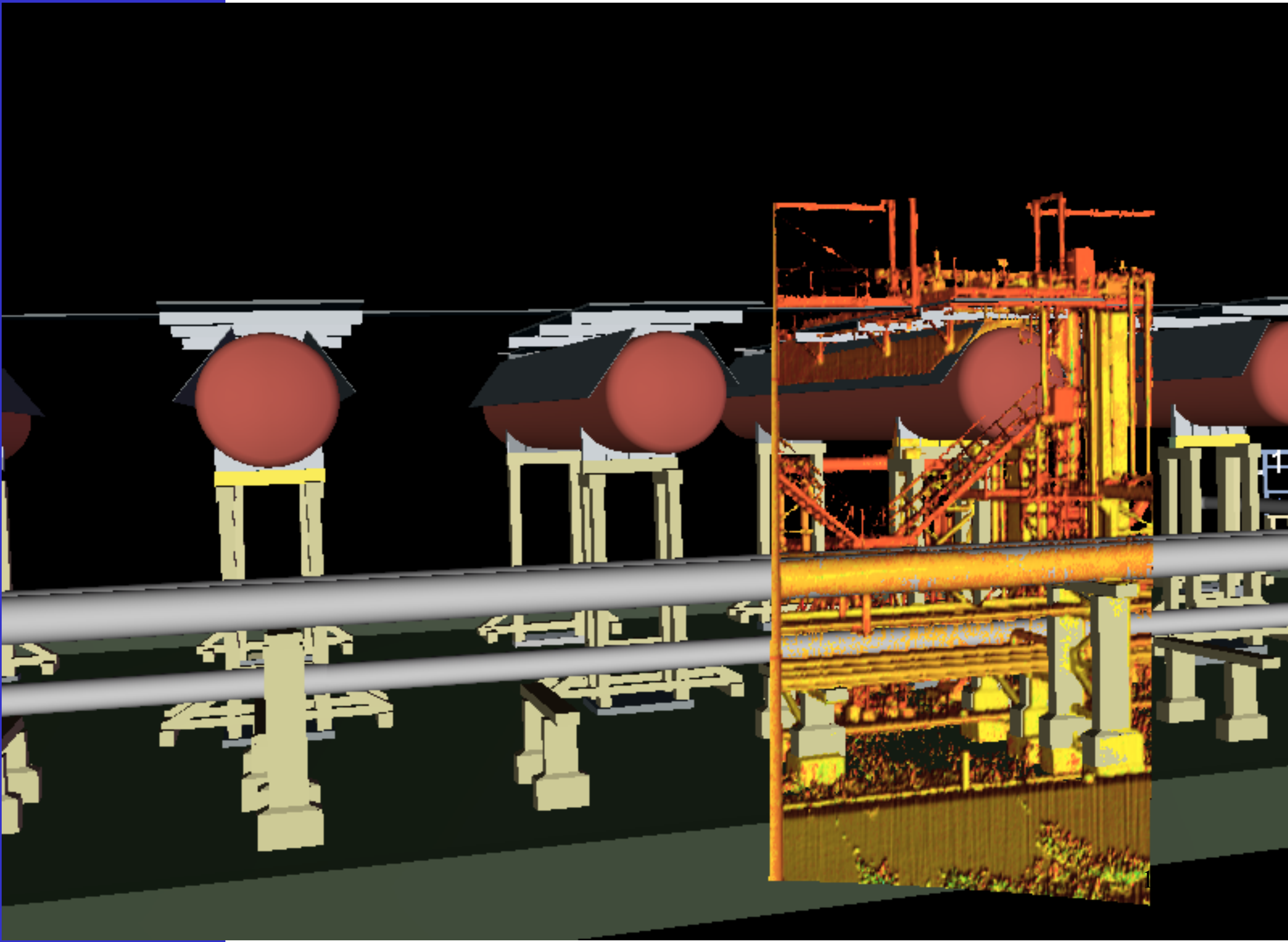
Step 4:
Model fitting in-the-field



Computer Vision

Result





Computer Vision

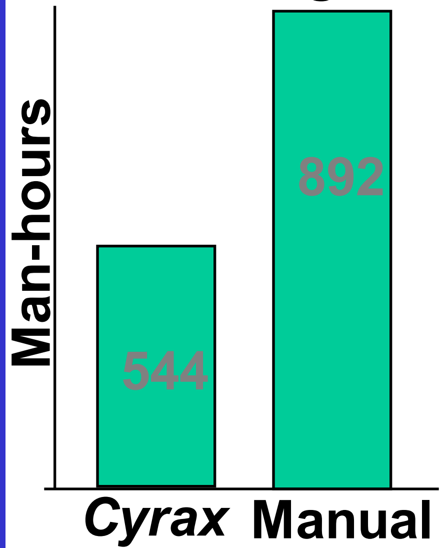
Project: As-built of Chevron hydrocarbon plant

- 400’x500’ area
- 10 vessels; 5 pumps
- 6,000 objects
- 81 scans from 30 tripod locations
- *Cyrax* field time = 50 hrs

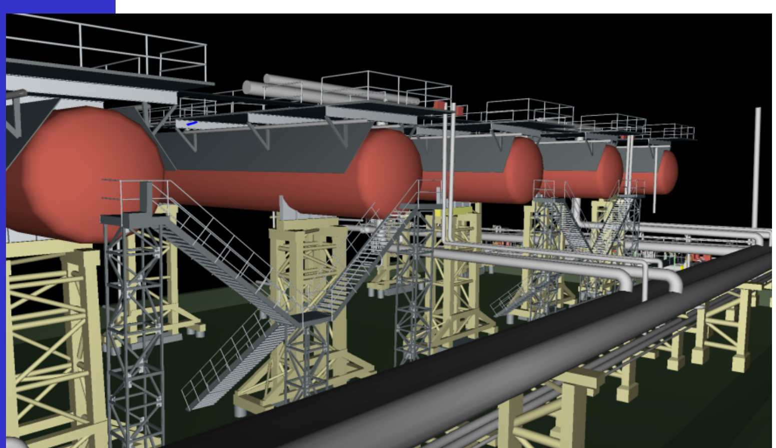
Cost Benefits

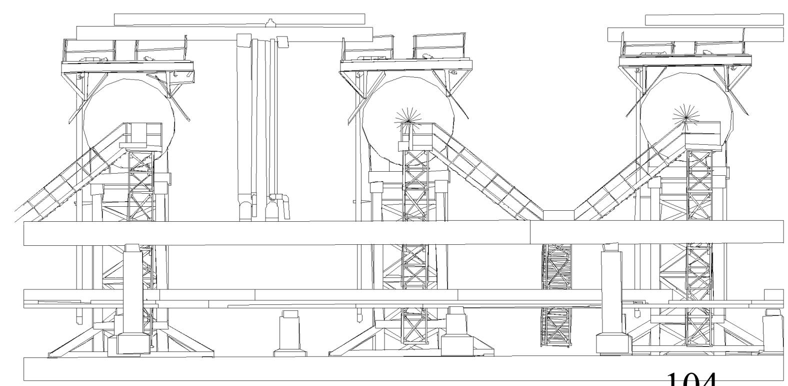
Added Value Benefits

Measuring & modeling



- Greater detail & no errors
- Higher accuracy
- Fewer construction errors
- 6 week schedule savings



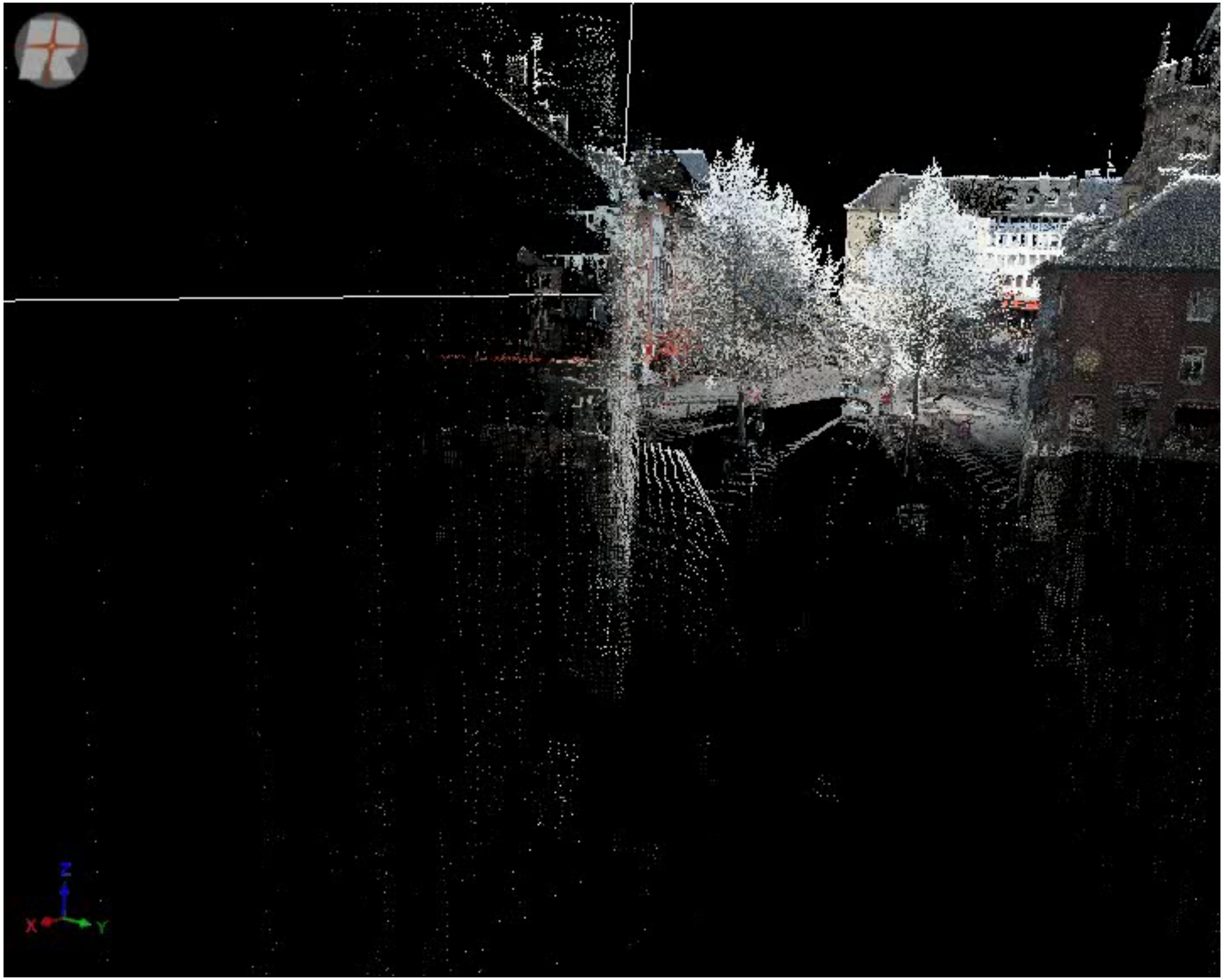


Application
Modeling movie sets

Image courtesy of Tippett Studio

Computer Vision

Lidar data with Riegl LMS-Z390i



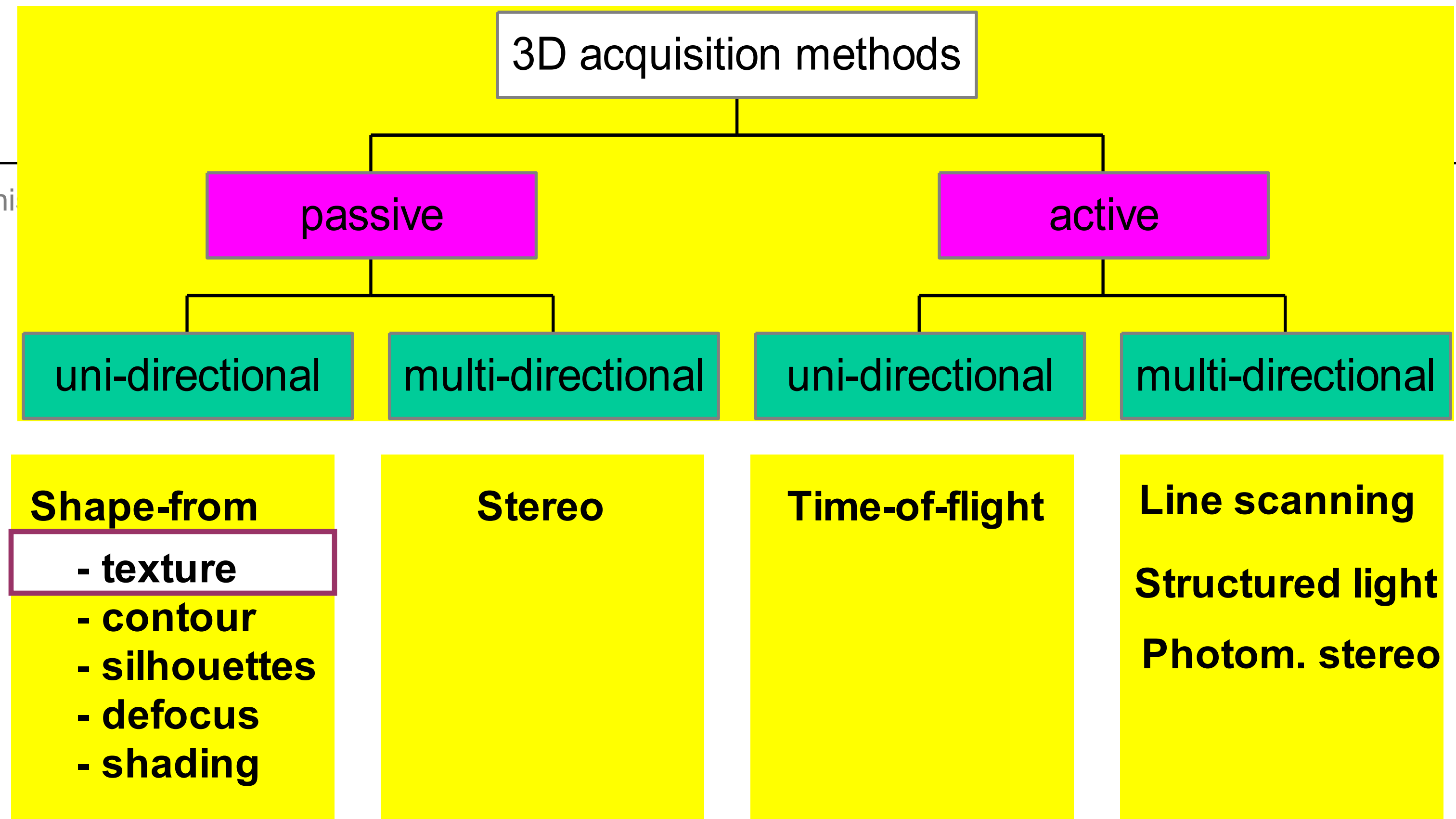
courtesy of RWTH Aachen, L. Kobbelt et al.

Computer Vision

Comparison Lidar - passive

Image courtesy of Tippett Studio

3D acquisition taxonomy



Shape-from-texture

assumes a slanted and tilted surface to have a homogeneous texture

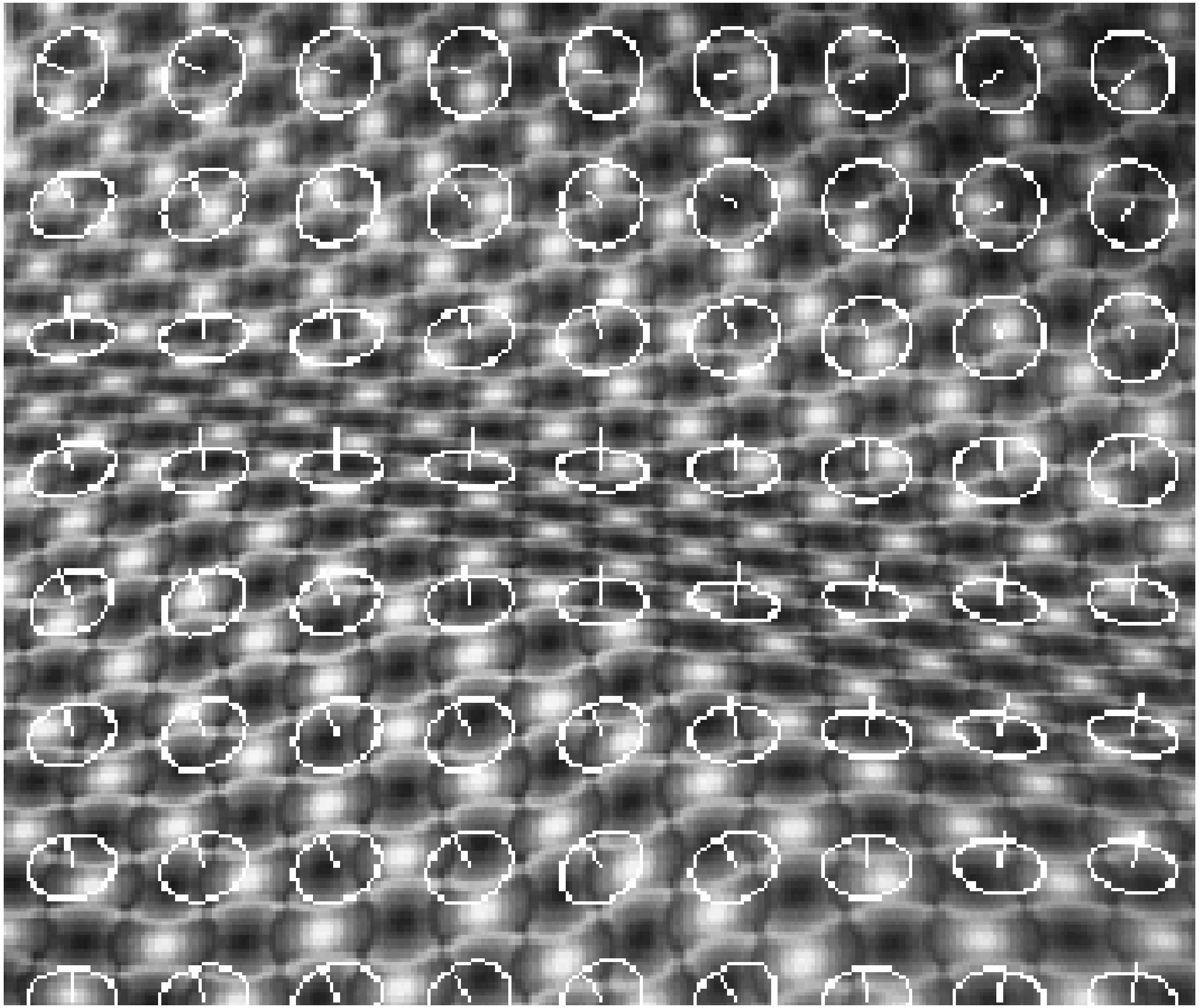
inhomogeneity is regarded as the result of projection

e.g. anisotropy in the statistics of edge orientations

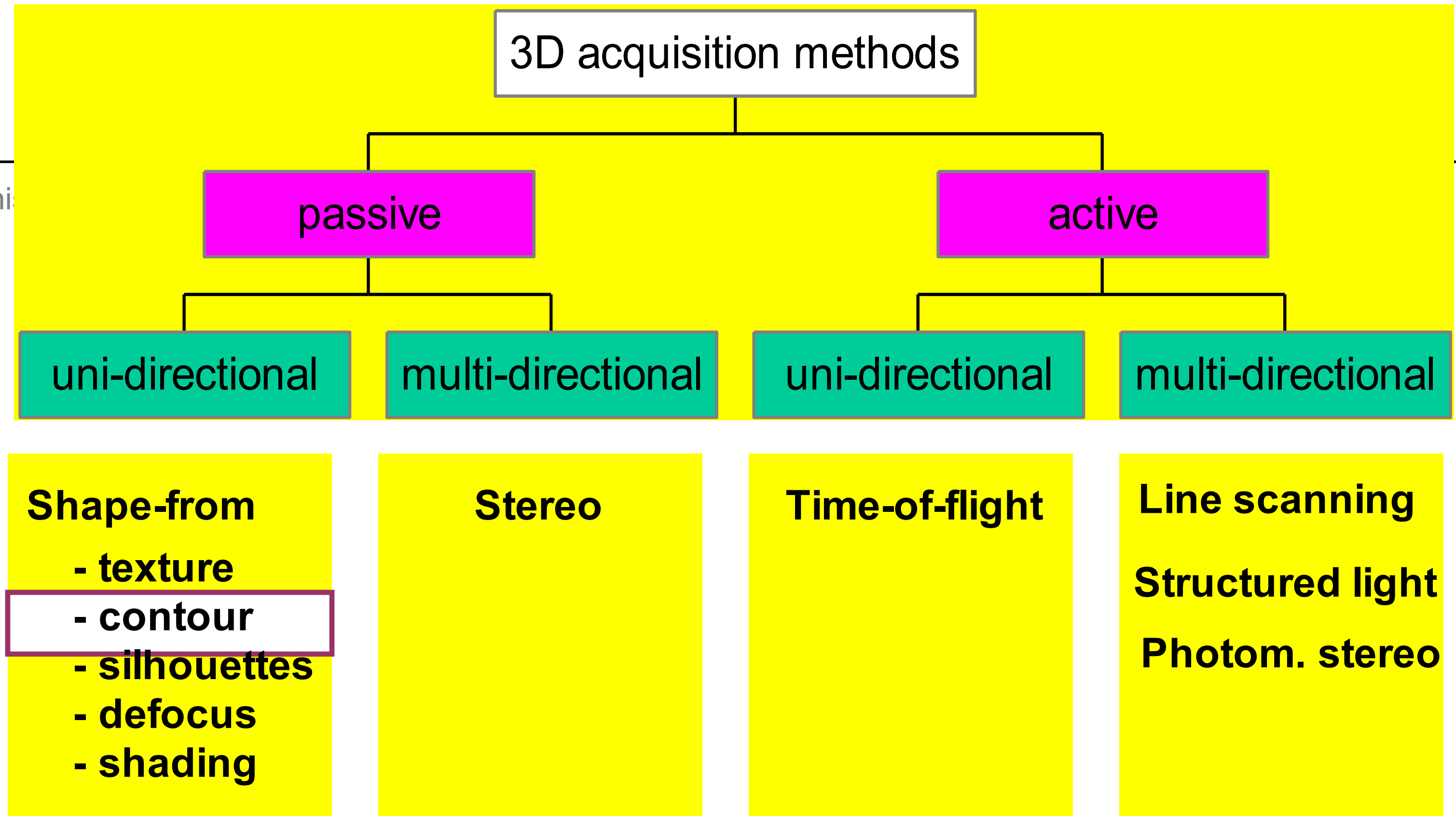
⇓

orientations deprojecting to maximally isotropic texture

Computer Vision



3D acquisition taxonomy



Shape-from-contour

makes assumptions about contour shape

E.g. the maximization of area over perimeter squared (compactness)

$\Downarrow$

ellipse ➔ circle

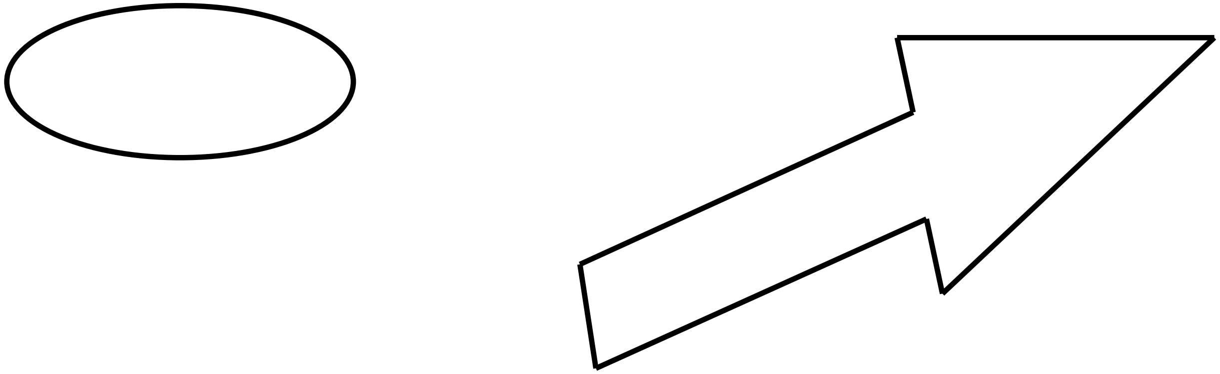
E.g. assumption of symmetry

$\Downarrow$

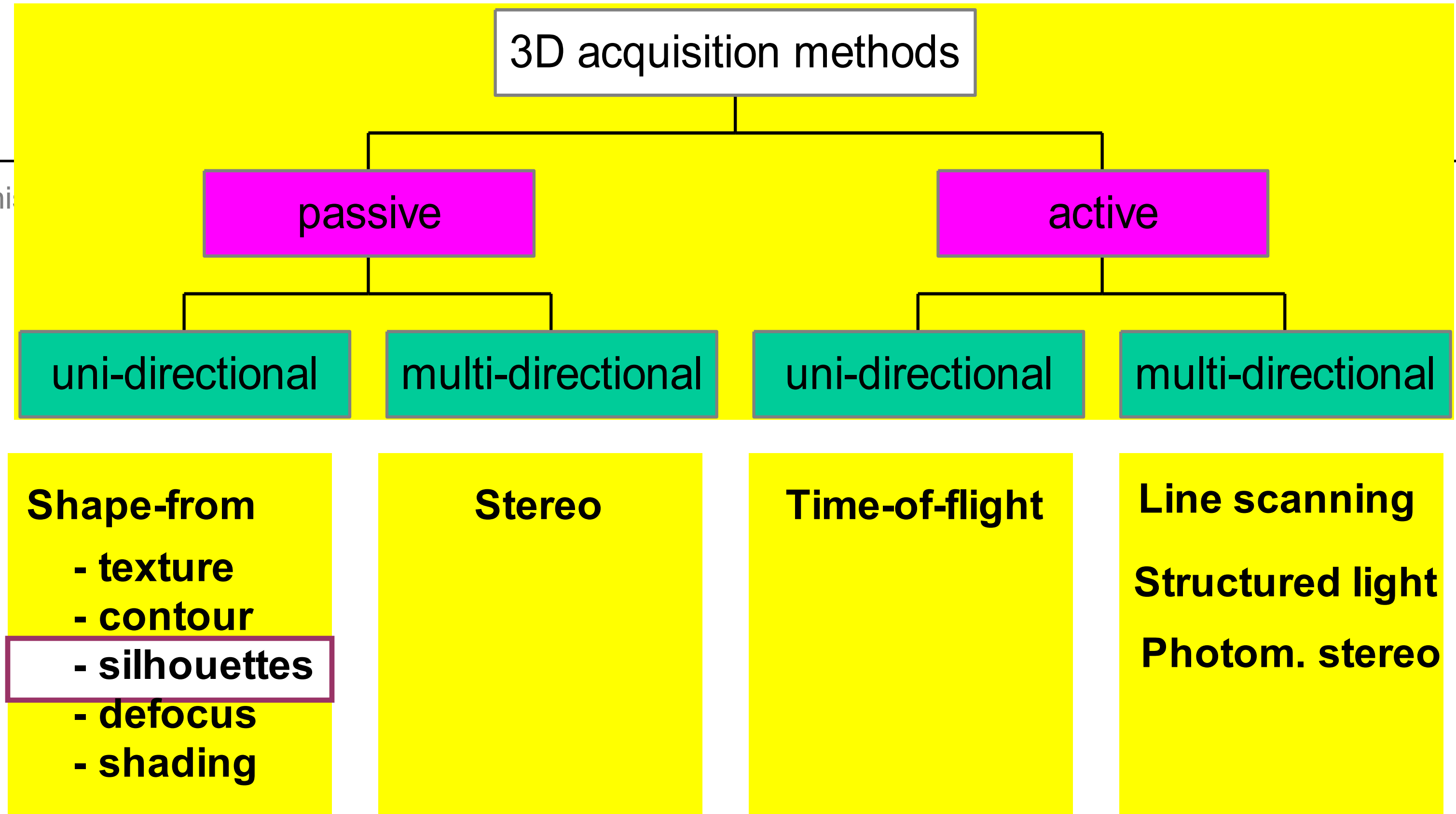
Symmetric contours ➔ surface of revolution

Computer Vision

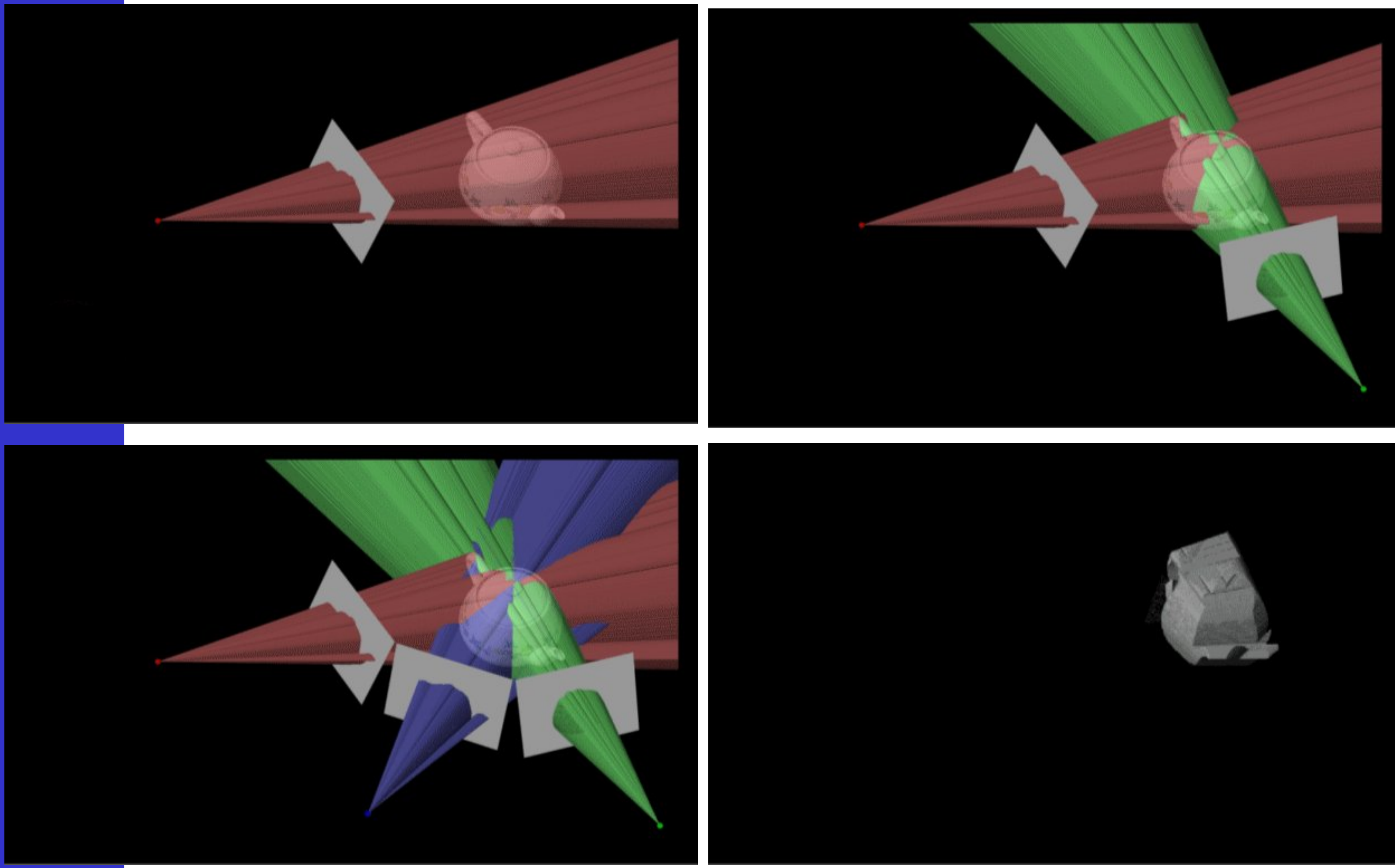
Shape-from-contour



3D acquisition taxonomy



Shape-from-silhouettes



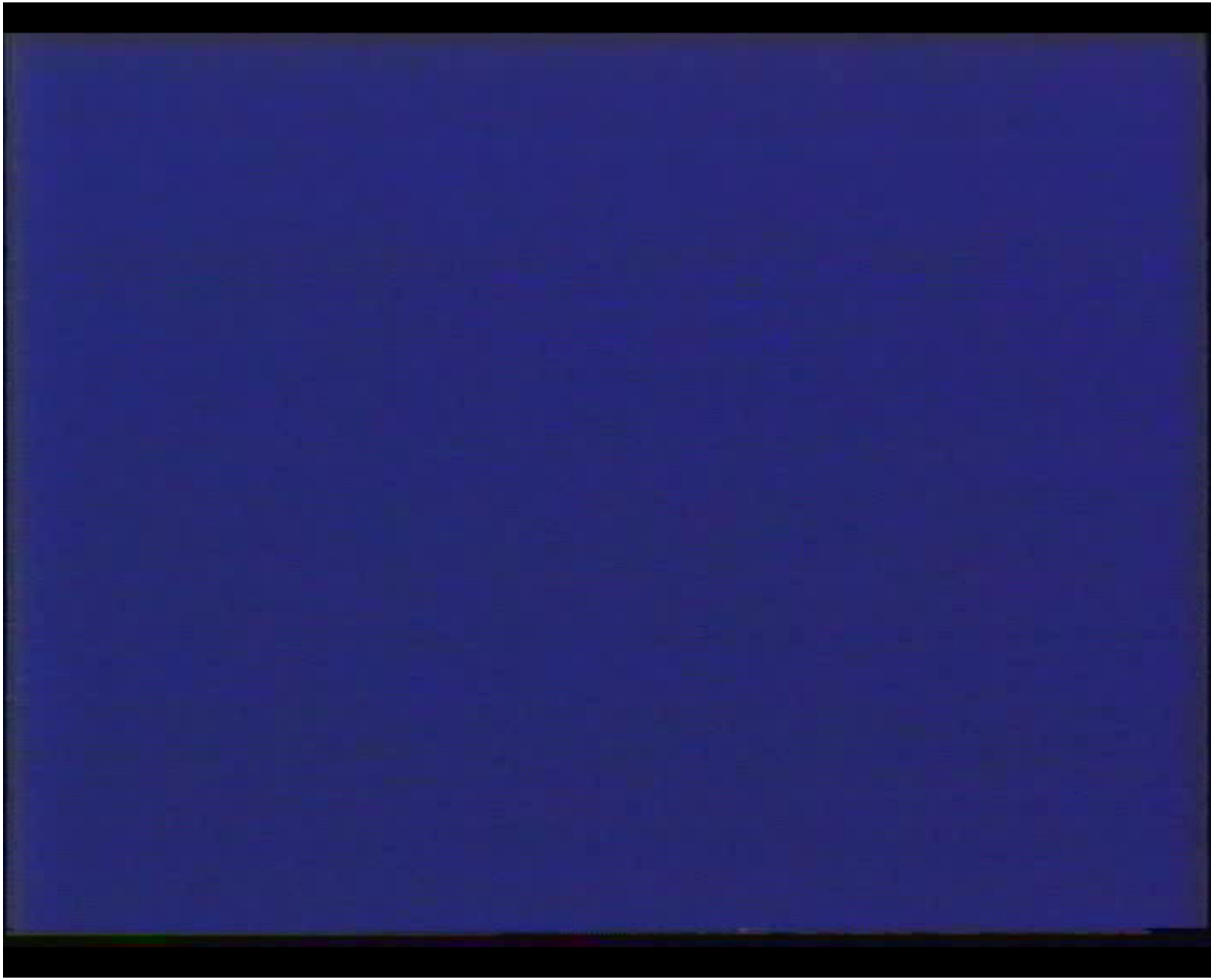
Shape from silhouettes - uncalibrated

- tracking of turntable rotation
- volumetric modeling from silhouettes
- triangular textured surface mesh

Turntable sequence

Camera tracking

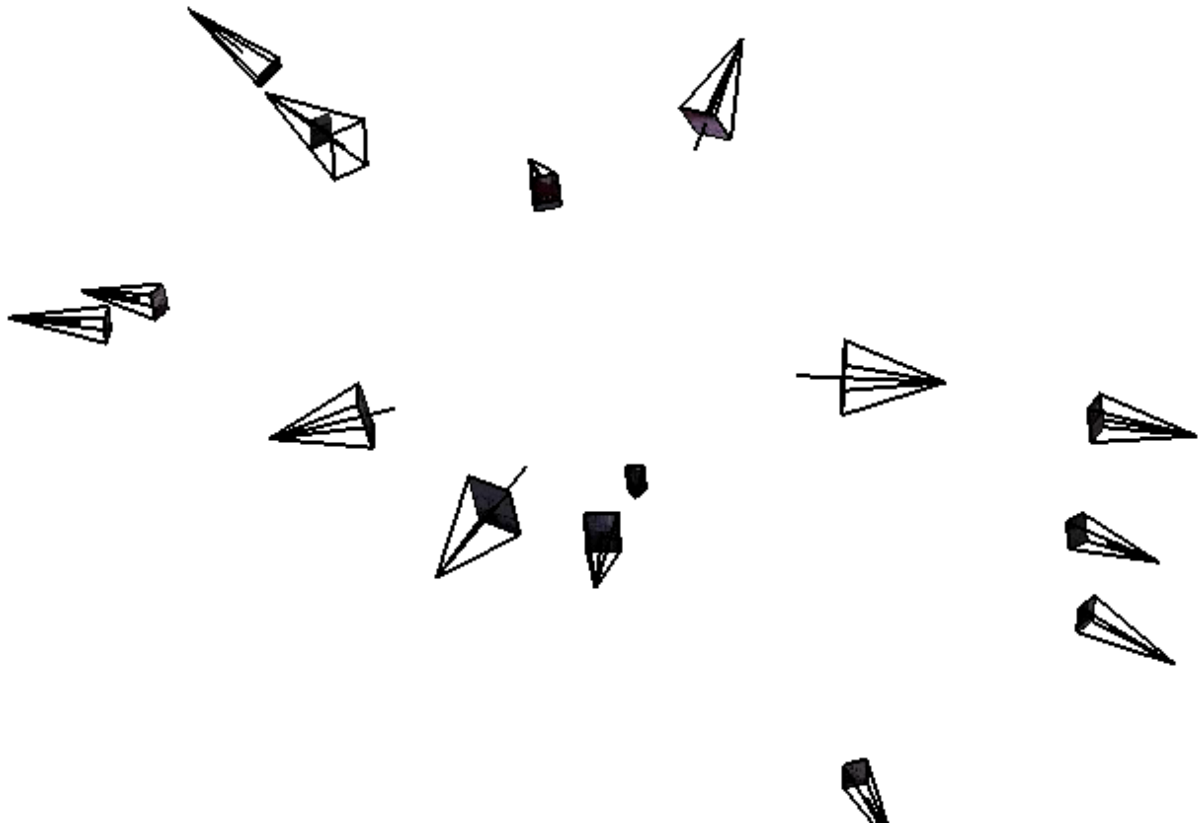
VRML model



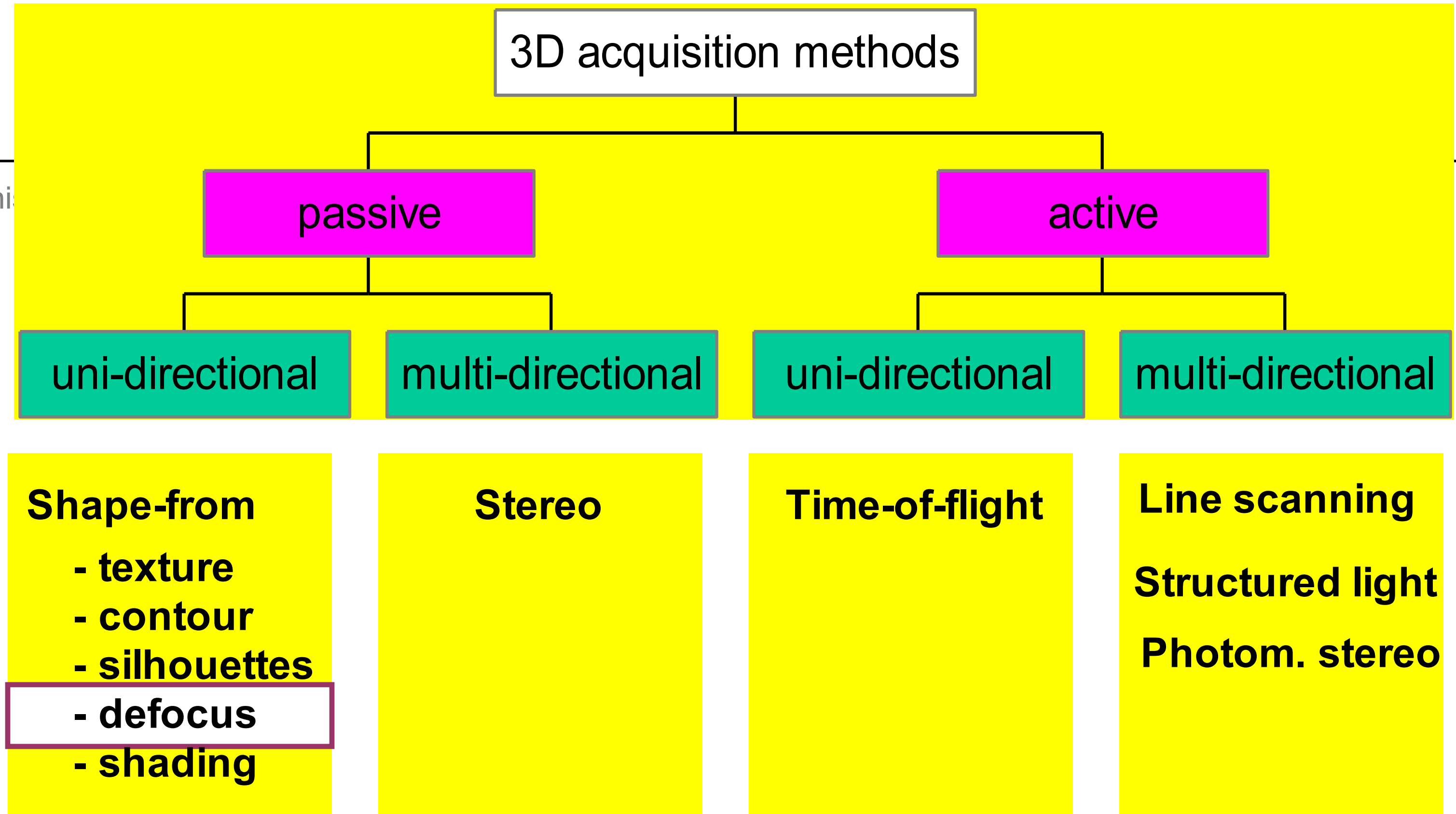
Computer Vision

Outdoor visual hulls

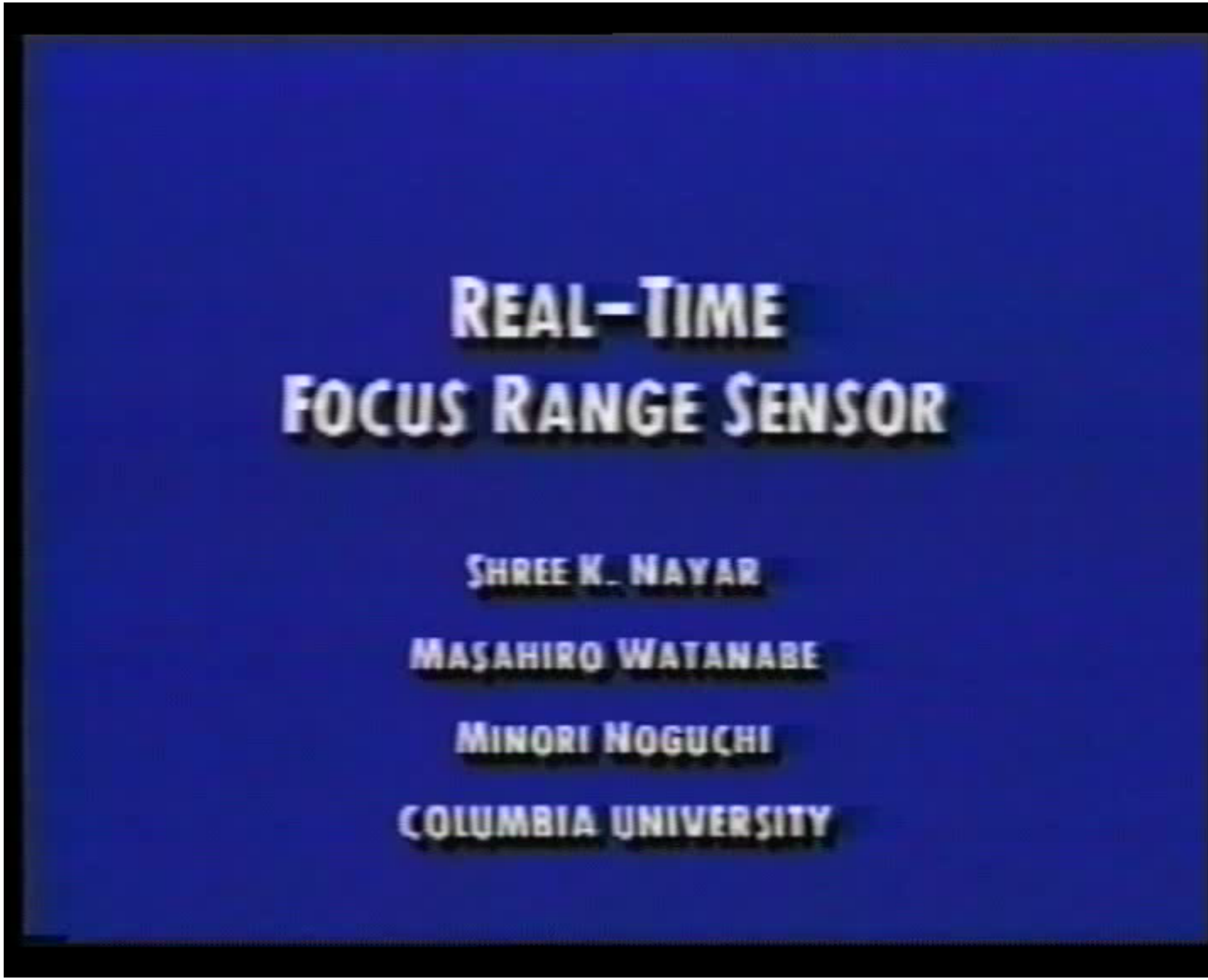
Outdoor visual hulls



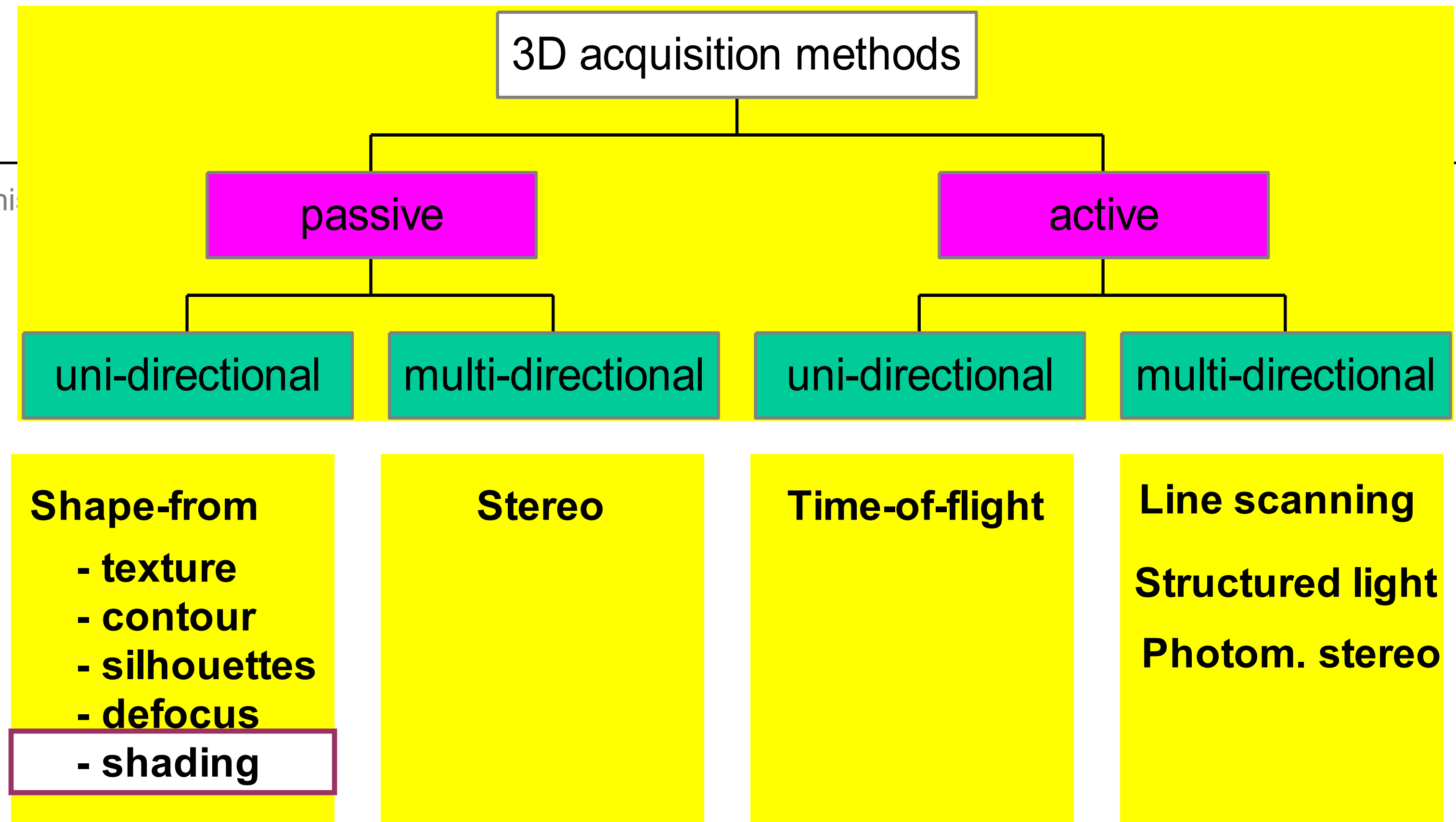
3D acquisition taxonomy



Computer Vision



3D acquisition taxonomy



Shape-from-shading

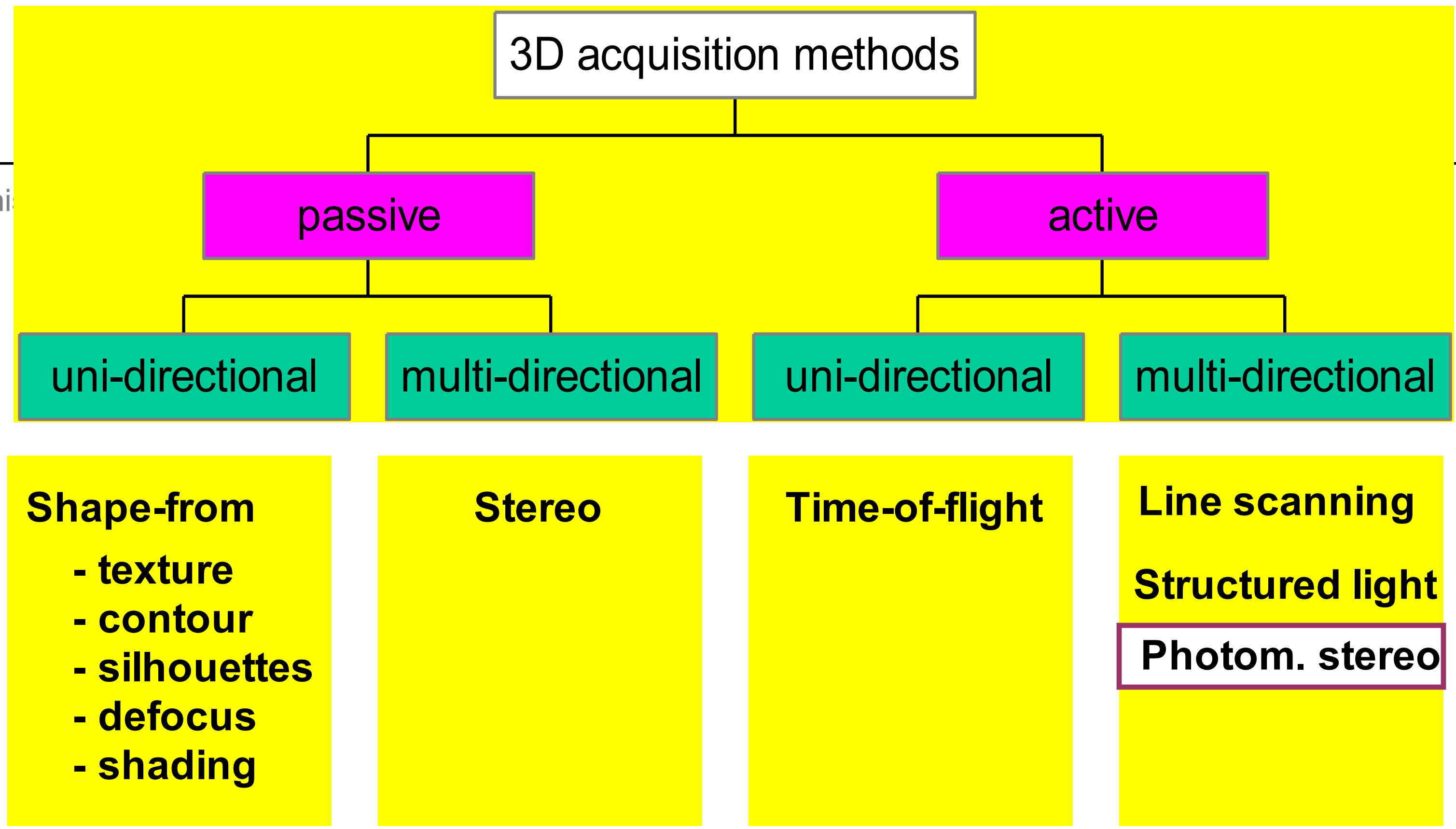
Uses directional lighting, often with known direction

local intensity is brought into correspondence with orientation via *reflectance maps*

orientation of an isolated patch cannot be derived uniquely

extra assumptions on surface smoothness and known normals at the rim

3D acquisition taxonomy



Photometric stereo

constraint propagation eliminated by using light from different directions

simultaneously when the light sources are given different colours

Mini-dome for photometric stereo

Instead of working with multi-directional light applied simultaneously with the colour trick, one can also project from many directions in sequence…

Mini-dome for photometric stereo

KATHOLIEKE UNIVERSITEIT
LEUVEN

Mini-dome

Computer

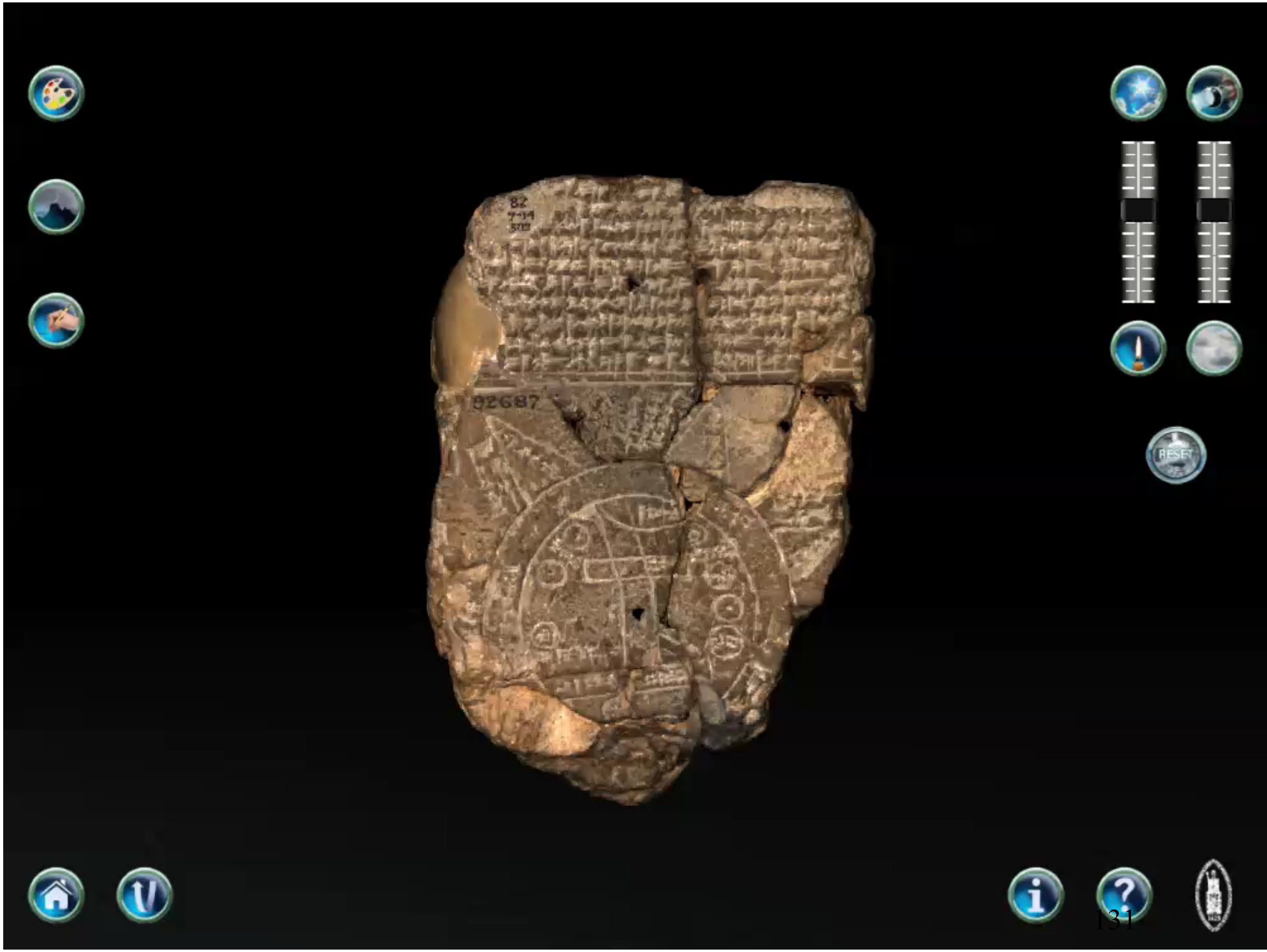
Mini-dome

Mini-dome for photometric stereo

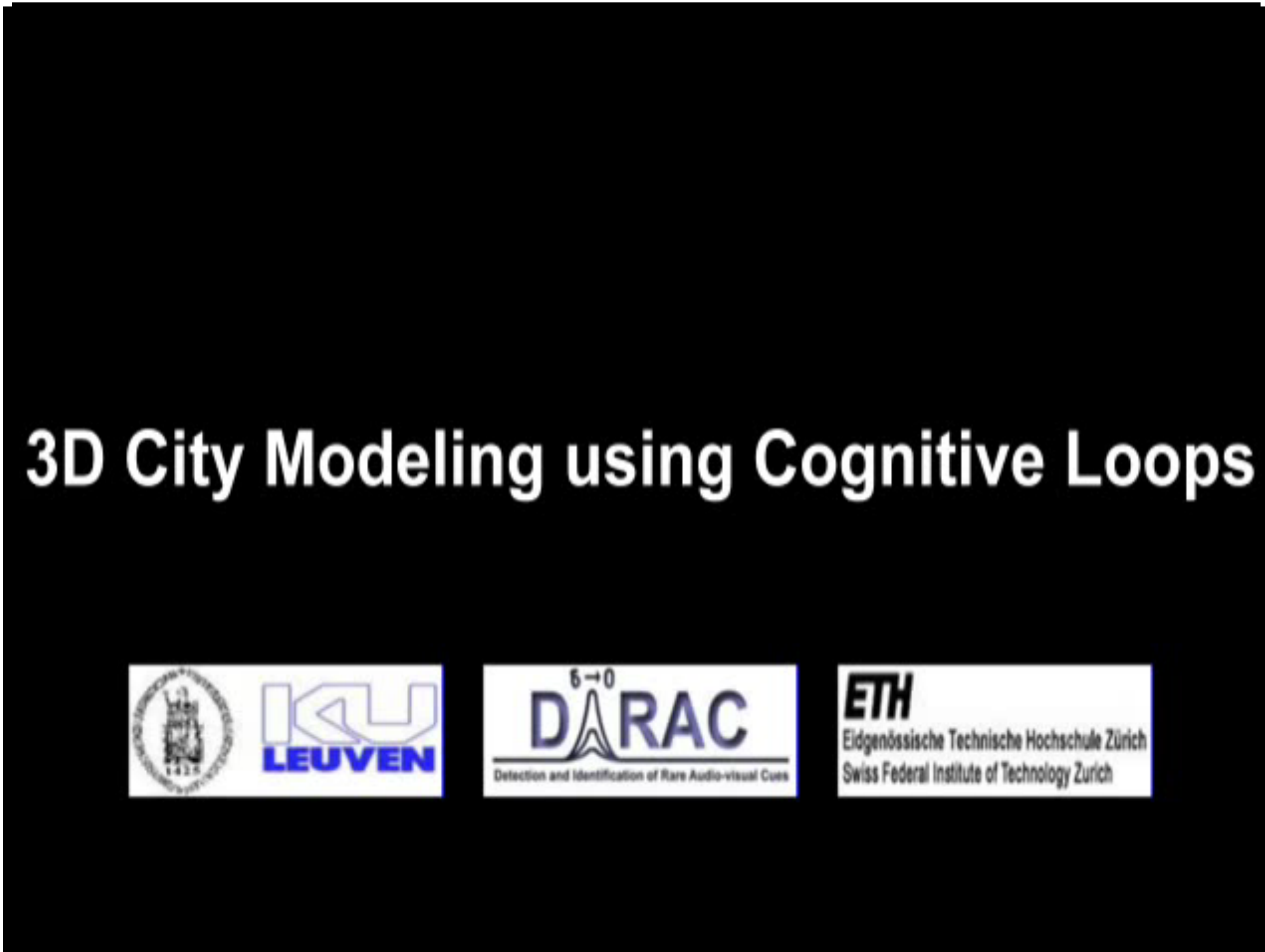
Example for tablet with first world map known, an exhibit at the British Museum:

http://homes.esat.kuleuven.be/~mproesma/mptmp/cuneiform

Mini-dome for photometric stereo



3D and recognition integrated



Computer Vision

Multi-walker tracker

Strongest 3D cues for us are 2D...