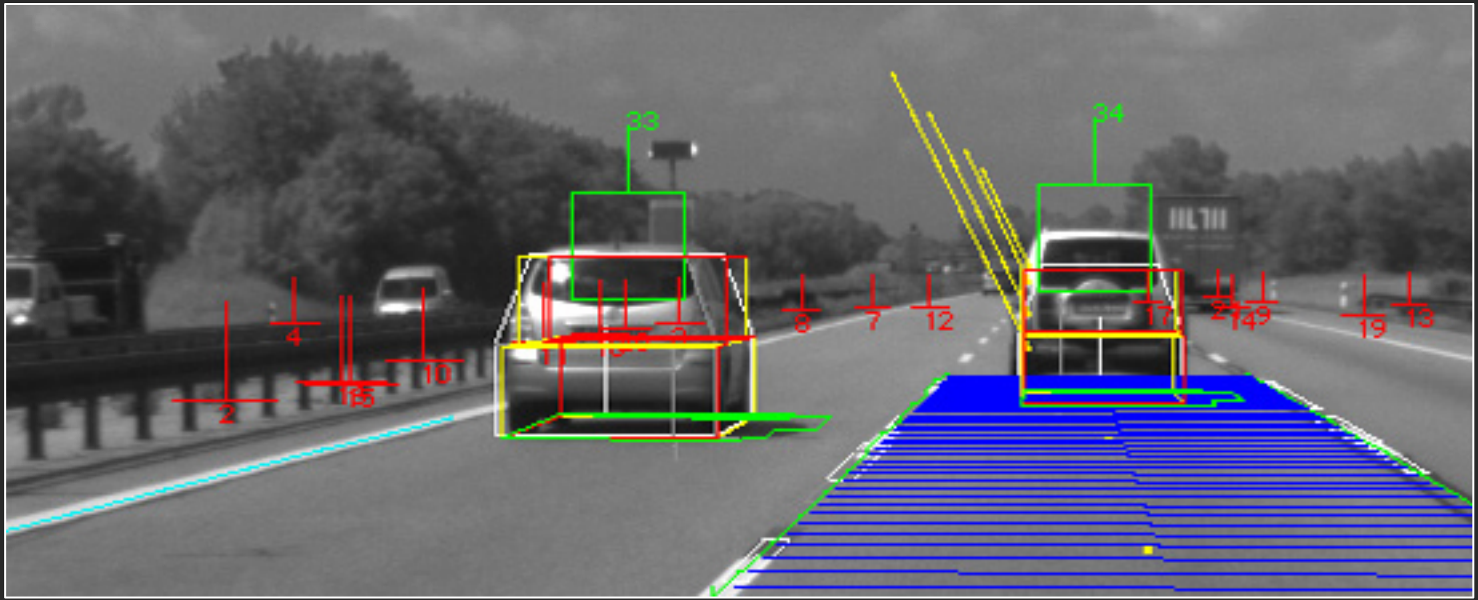
IMTR- A08A : Visión Artificial

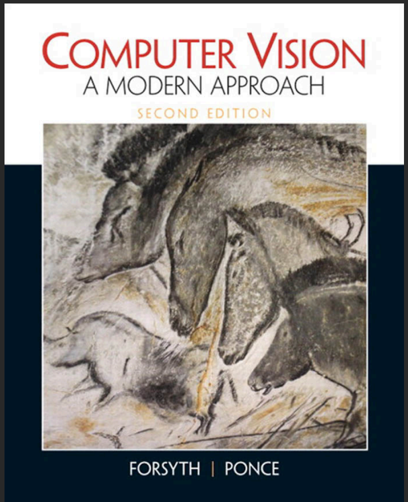


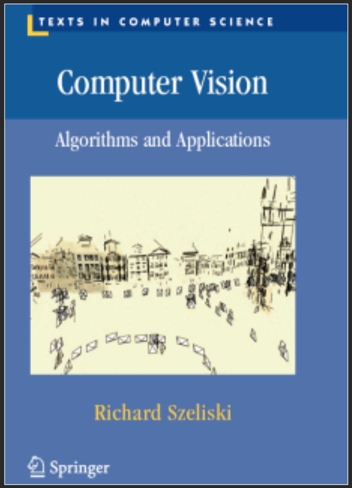
Andrés Marrugo, PhD

2021

Información general

- Profesor: Andrés Marrugo, PhD agmarrugo@utb.edu.co
- Horario de atención a estudiantes: Por anunciar, o con cita previa.
- El curso está apoyado en SAVIO.
- Textos (sugeridos):
  Forsyth & Ponce, *Computer Vision: A Modern Approach*
  Richard Szeliski, *Computer Vision: Algorithms and Applications* (available online)





The goal of computer vision

- To extract "meaning" from pixels

What we see

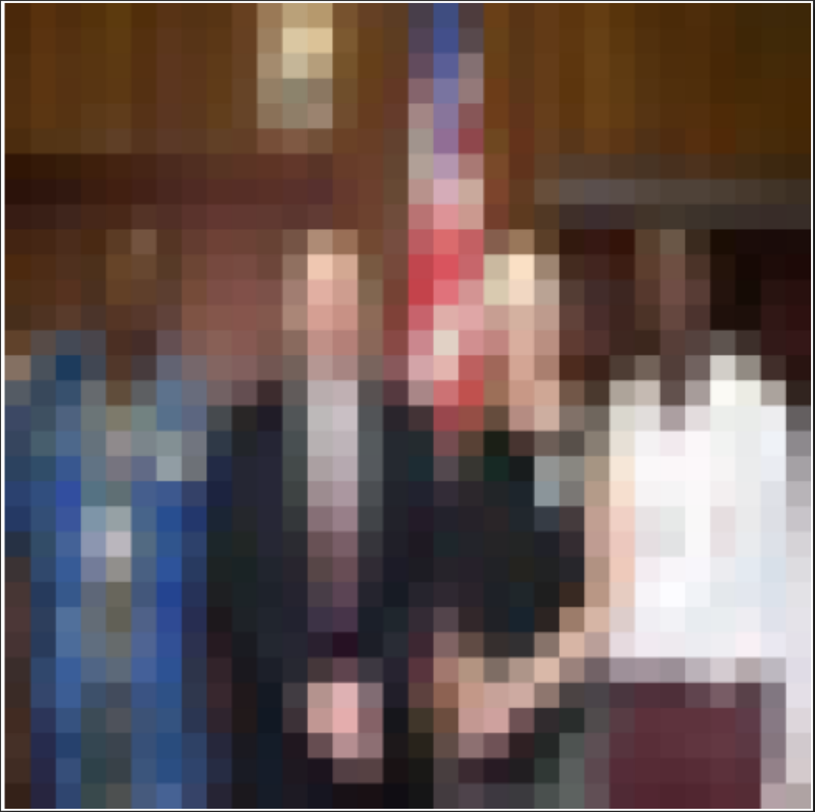
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: S. Narasimhan

The goal of computer vision

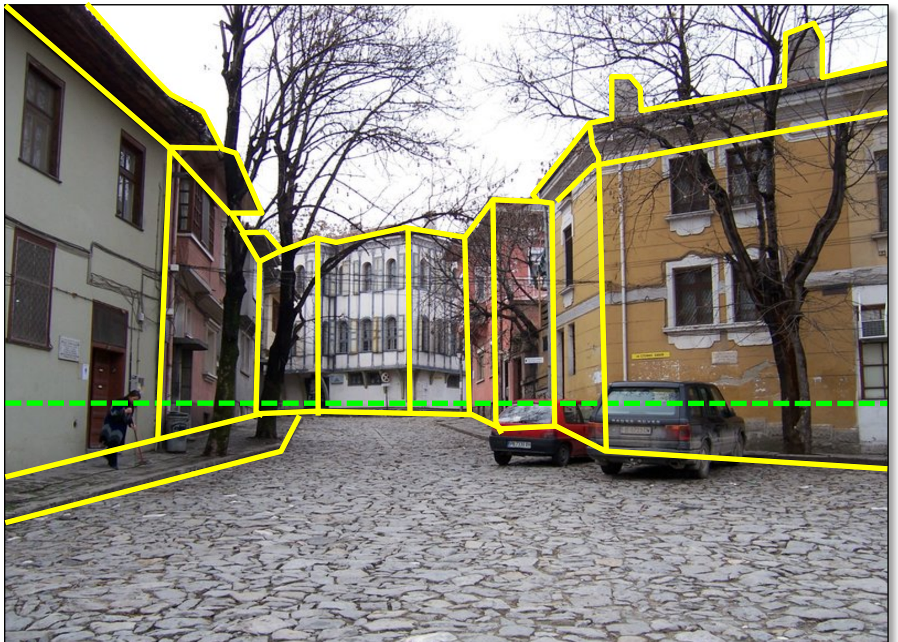
- To extract "meaning" from pixels



Humans are remarkably good at this…

Source: "80 million tiny images" by Torralba et al.

What kind of information can be extracted from an image?



Semantic information

Geometric information

Source: S. Lazebnik

Why study computer vision?

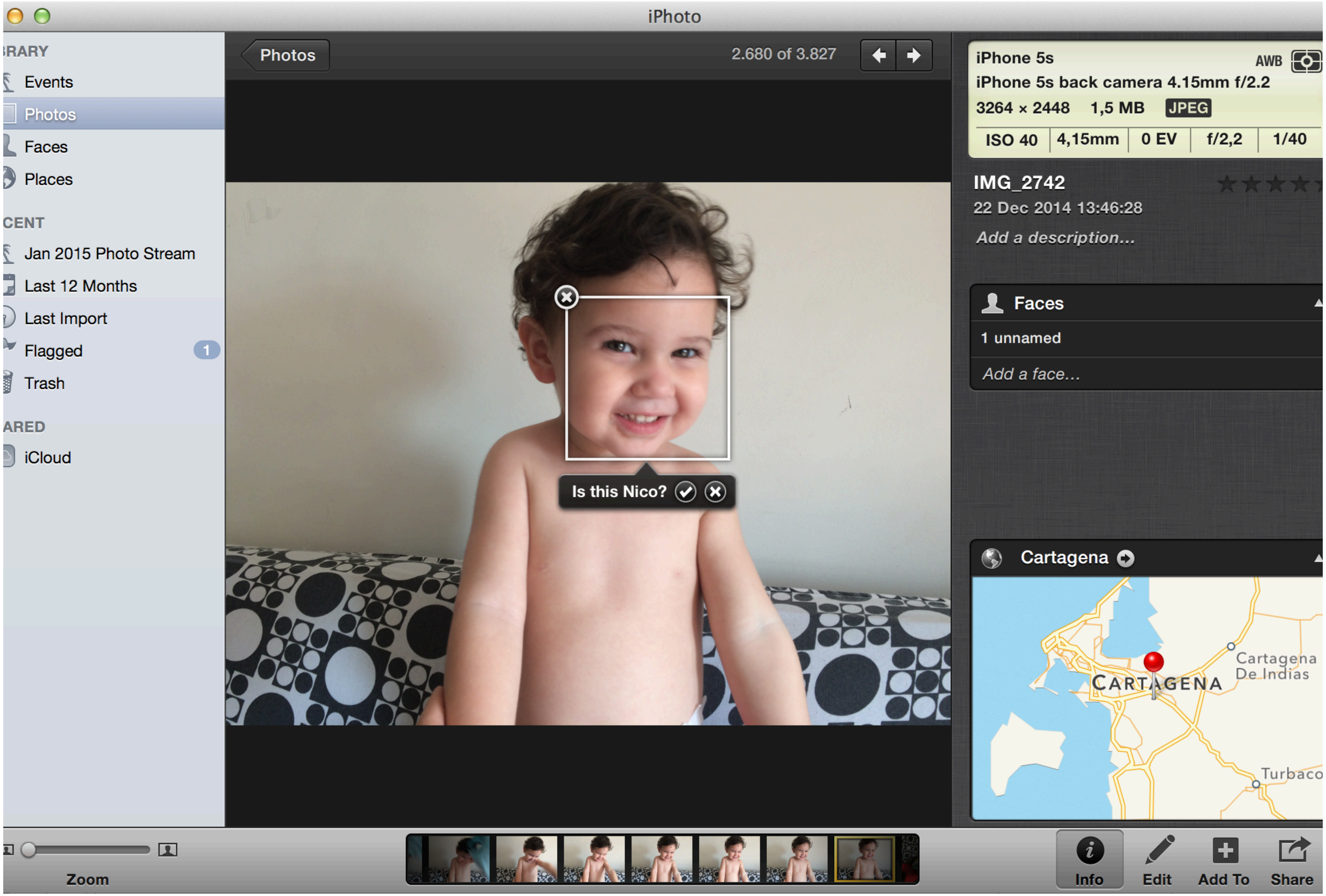
- Vision is useful
- Vision is interesting
- Vision is difficult
  - Half of primate cerebral cortex is devoted to visual processing
  - Achieving human-level image understanding is probably “AI-complete”

Successes of computer vision to date

“Simple” patterns

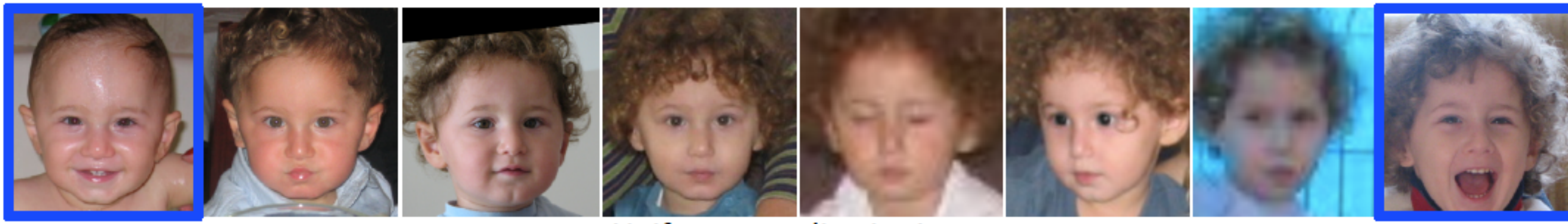
Source: S. Lazebnik

Face recognition: Apple iPhoto software

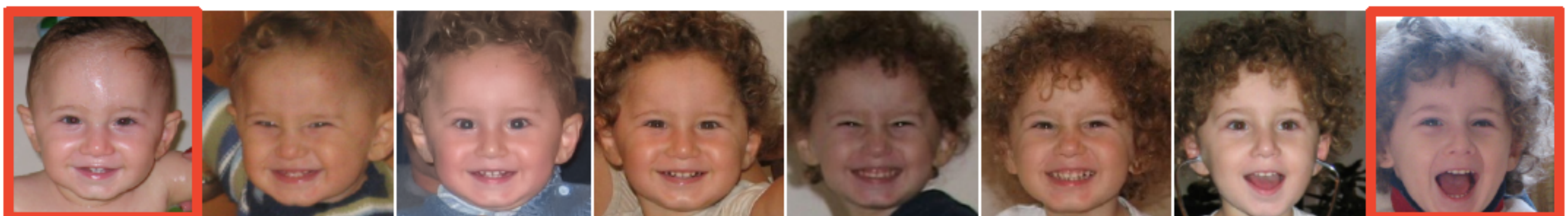


http://www.apple.com/ilife/iphoto/

Face movies



Uniform sampling in time

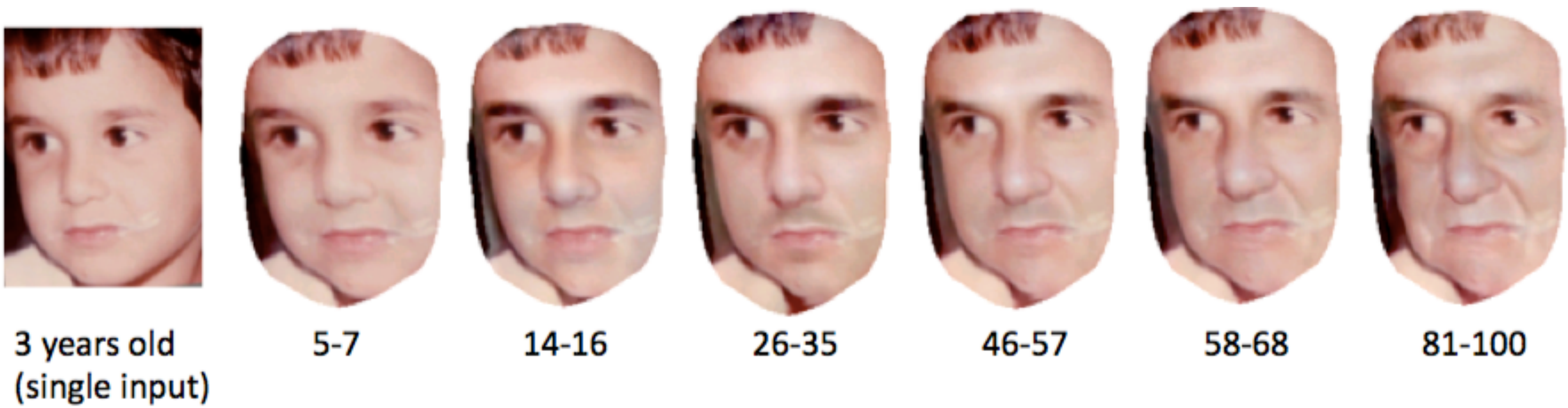


Our result

I. Kemelmacher-Shlizerman, E. Shechtman, R. Garg and S. Seitz, [Exploring Photobios](), SIGGRAPH 2011

[YouTube Video]()

Automatic age progression

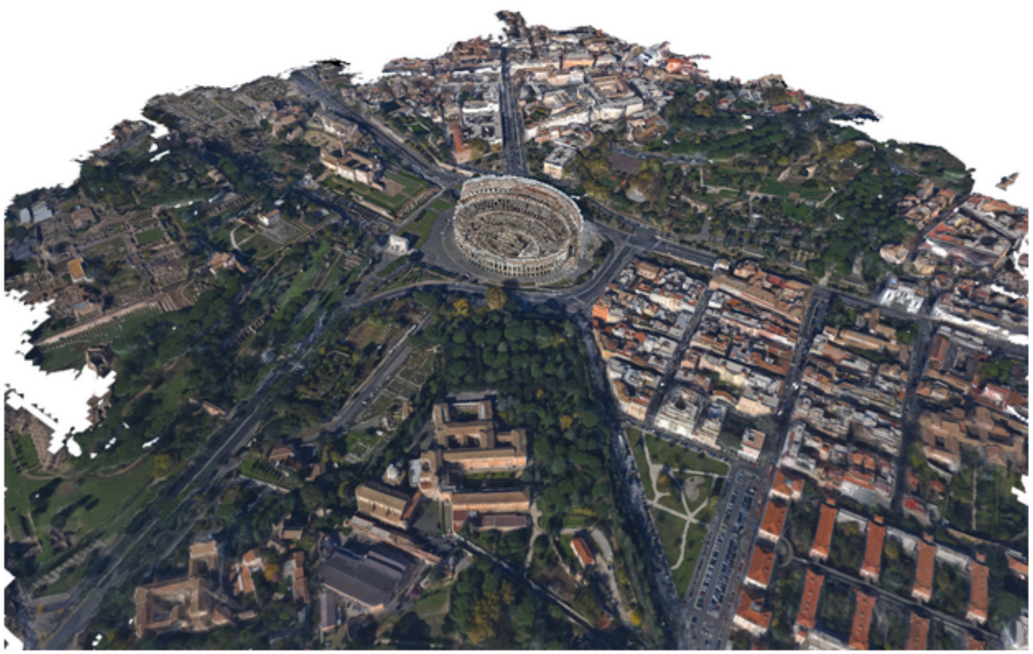


I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. Seitz, Illumination-Aware Age Progression, CVPR 2014

YouTube Video

Reconstruction: 3D from photo collections

Colosseum, Rome, Italy

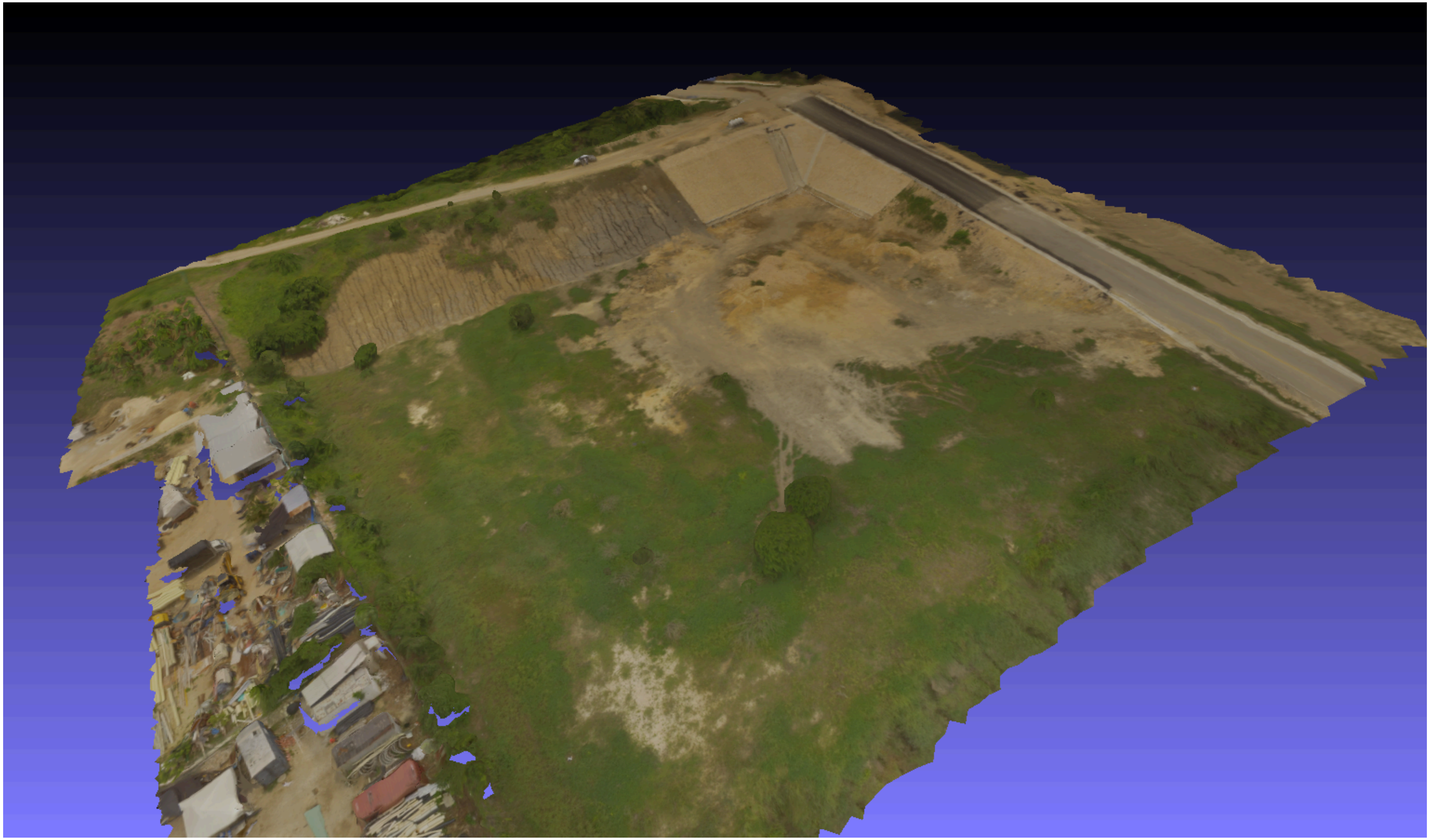


San Marco Square, Venice, Italy

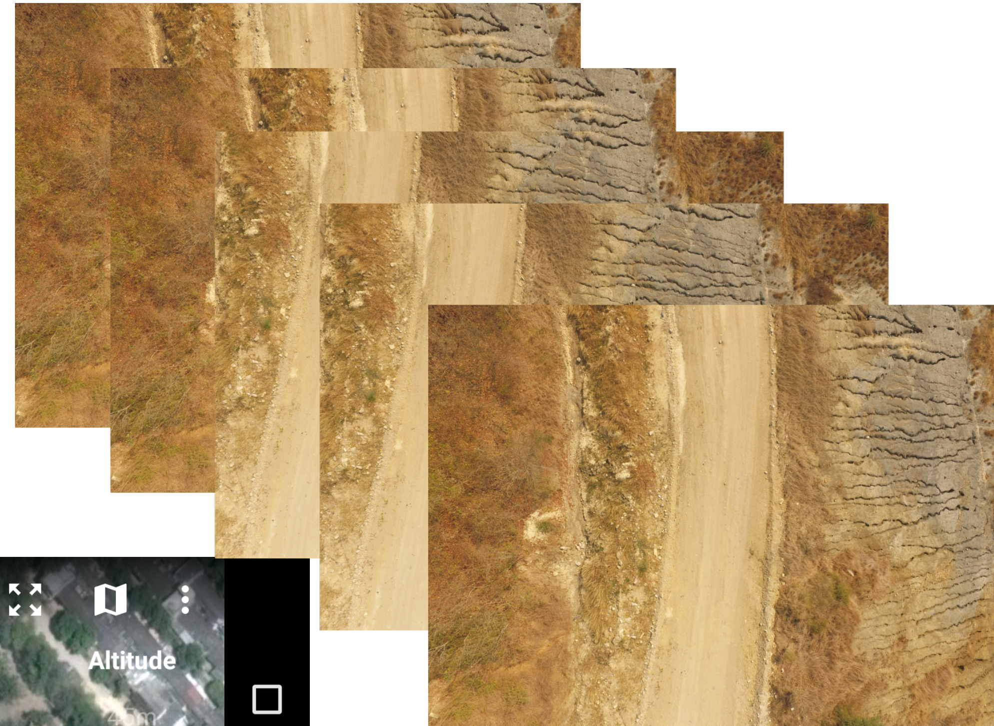
Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. Seitz, The Visual Turing Test for Scene Reconstruction, 3DV 2013

YouTube Video

Reconstruction: 3D from drone photos



J Meza et al., Comm. in Computer and Information Science, 2018



J Meza et al., Comm. in Computer and Information Science, 2018

J Meza et al., Comm. in Computer and Information Science, 2018

Reconstruction: 4D from photo collections

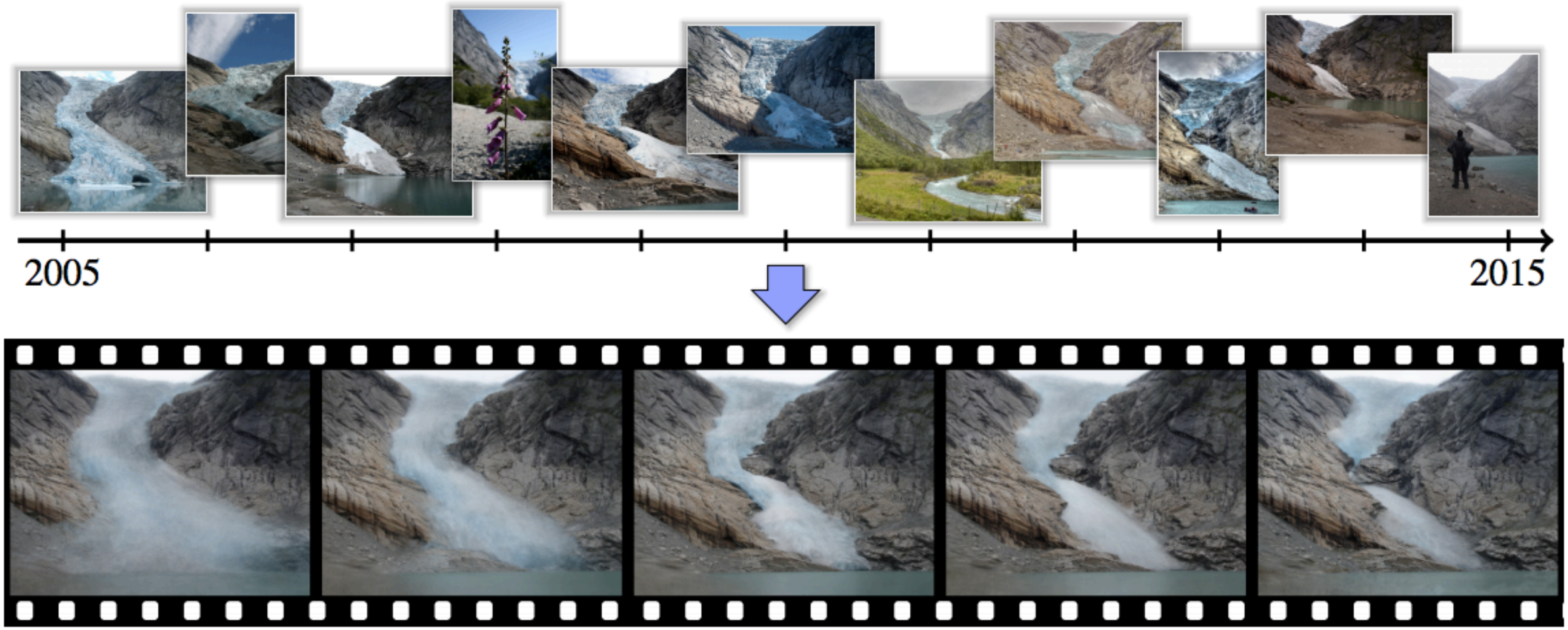


Figure 1: *We mine Internet photo collections to generate time-lapse videos of locations all over the world. Our time-lapses visualize a multitude of changes, like the retreat of the Briksdalsbreen Glacier in Norway shown above. The continuous time-lapse (bottom) is computed from hundreds of Internet photos (samples on top).* *Photo credits:* `Aliento Más Allá`, `jirihnidek`, `mcxurxo`, `elka_cz`, *Juan Jesús Orío, Klaus Wißkirchen,* `Daikrieg`, `Free the image`, `dration` *and Nadav Tobias.*

R. Martin-Brualla, D. Gallup, and S. Seitz, Time-Lapse Mining from Internet Photos, SIGGRAPH 2015

YouTube Video

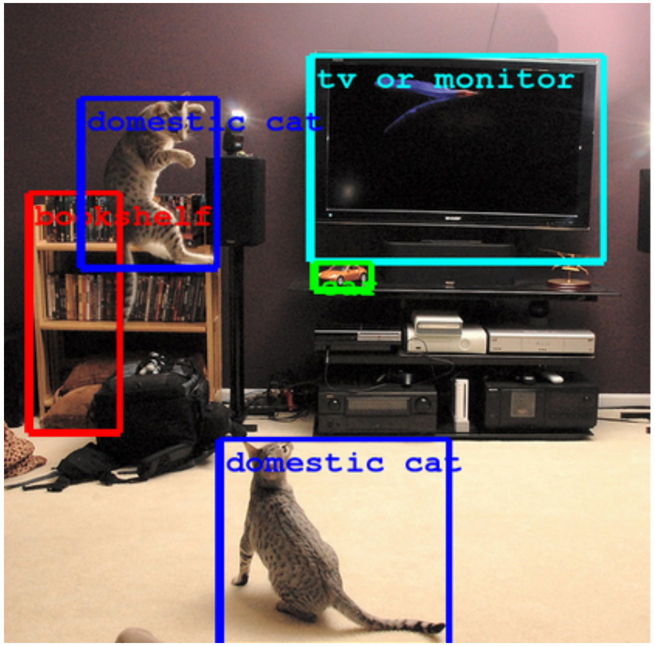
Reconstruction: 4D from depth cameras

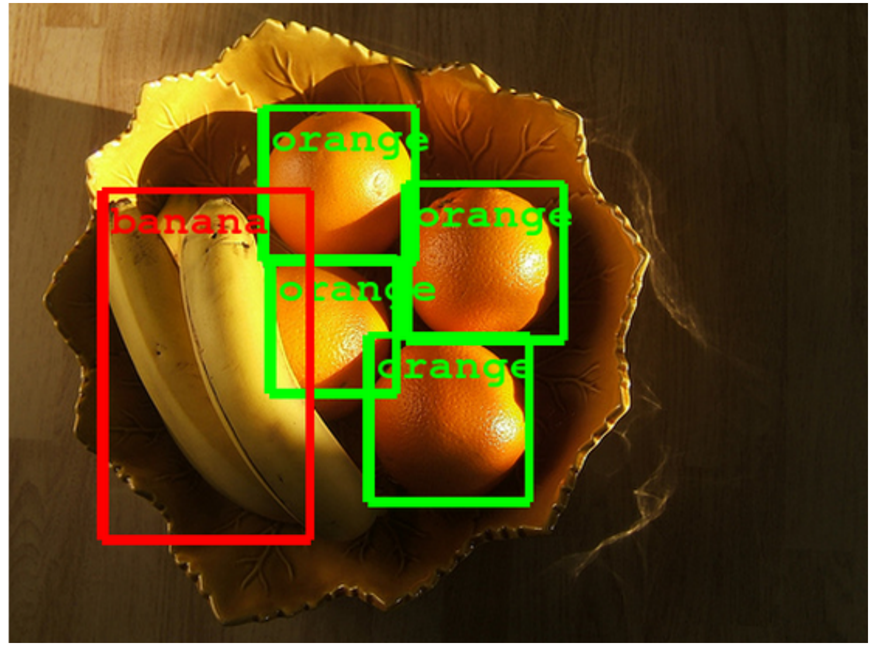
Figure 1: Real-time reconstructions of a moving scene with DynamicFusion; both the person and the camera are moving. The initially noisy and incomplete model is progressively denoised and completed over time (left to right).

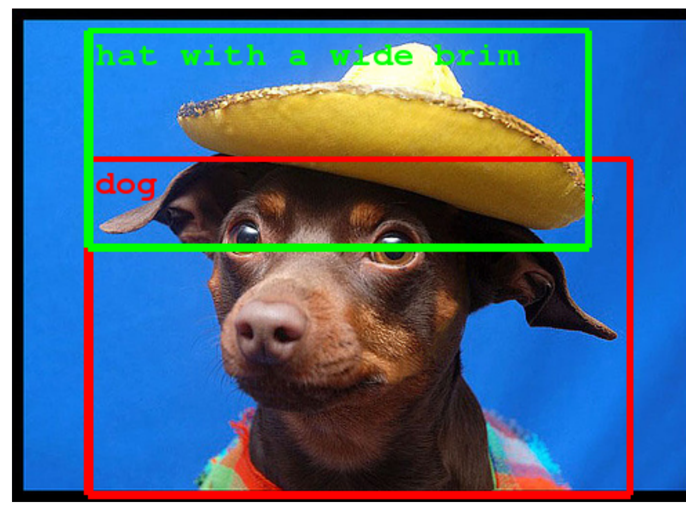
R. Newcombe, D. Fox, and S. Seitz, [DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time], CVPR 2015

[YouTube Video]

Recognition







- Computer Eyesight Gets a Lot More Accurate, NY Times Bits blog, August 18, 2014
- Building A Deeper Understanding of Images, Google Research Blog, September 5, 2014
- Baidu caught gaming recent supercomputer performance test, Engadget, June 3, 2015

Self-driving cars

TECHNOLOGY

The New York Times

For Now, Self-Driving Cars Still Need Humans

By JOHN MARKOFF JAN. 17, 2016

http://www.nytimes.com/2016/01/18/technology/driverless-cars-limits-include-human-nature.html

Why is computer vision difficult?

Challenges: viewpoint variation

Challenges: illumination

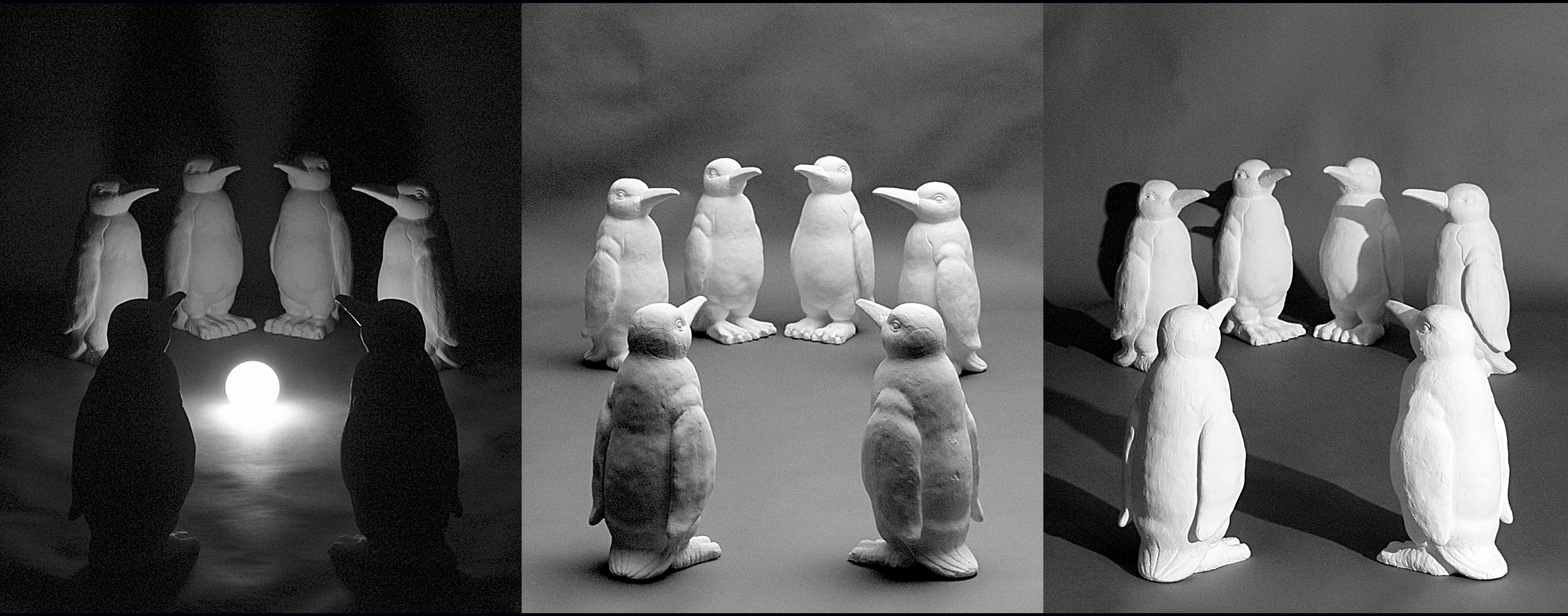
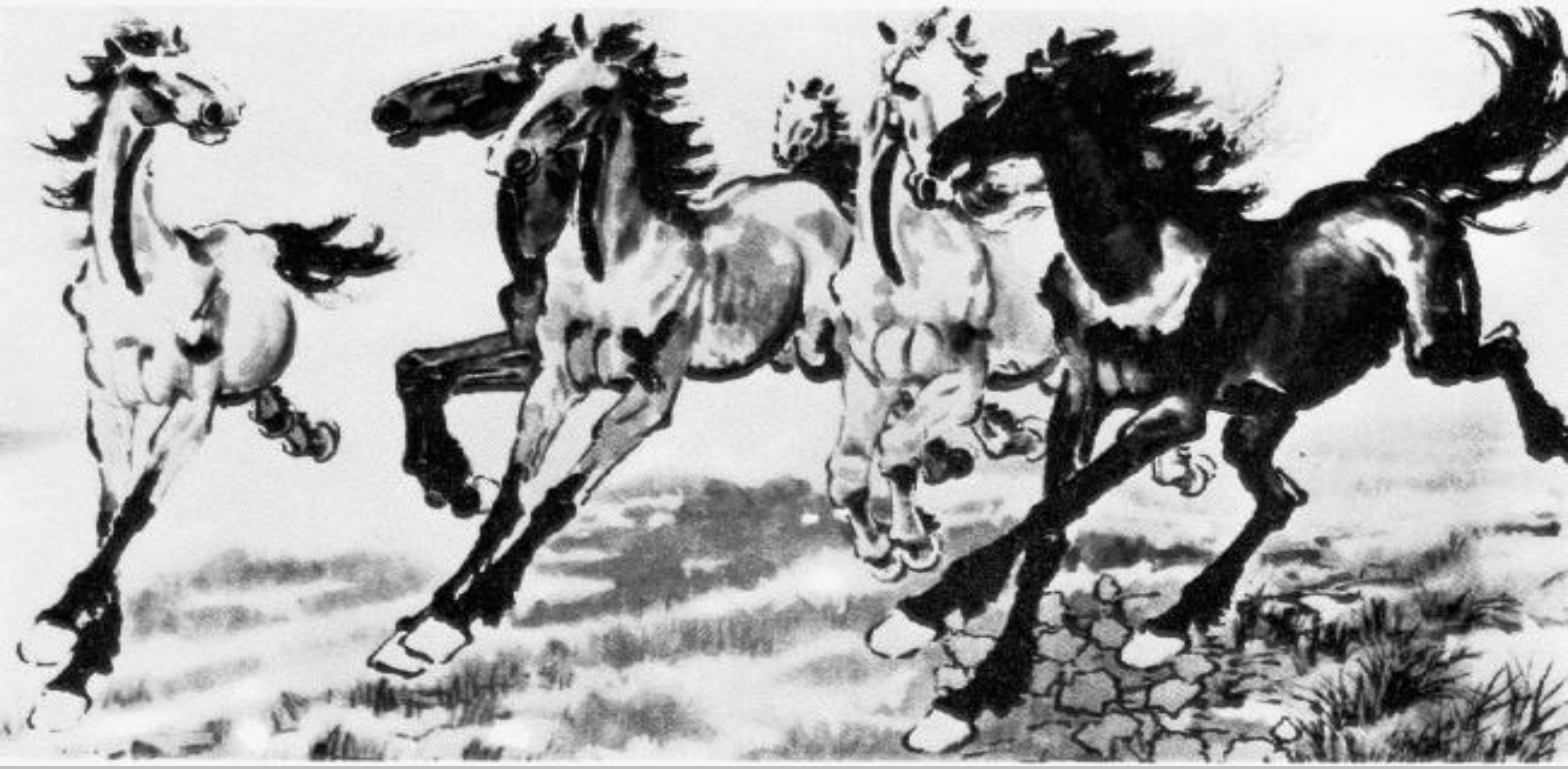


image credit: J. Koenderink

Challenges: scale

slide credit: Fei-Fei, Fergus & Torralba

Challenges: deformation



Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba

Challenges: object intra-class variation

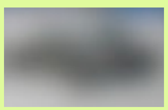
slide credit: Fei-Fei, Fergus & Torralba

Challenges: occlusion, clutter

Image source: National Geographic

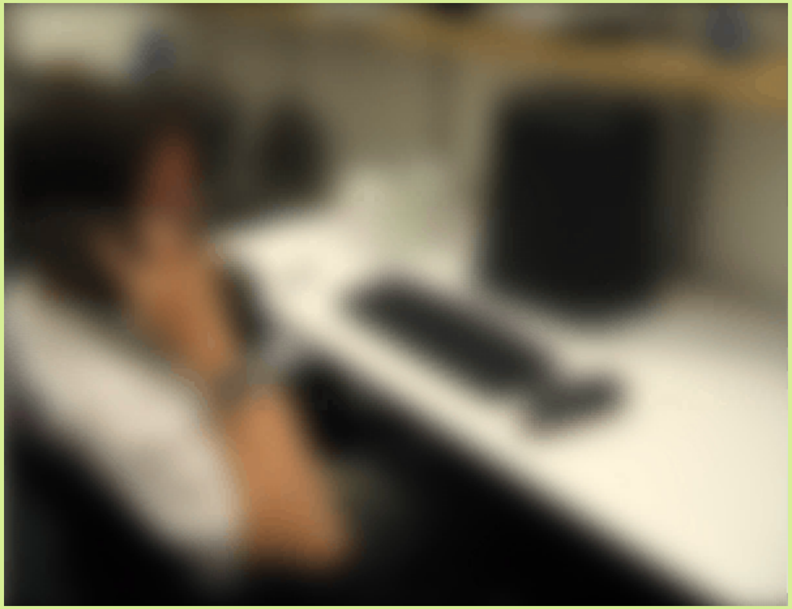
Challenges: Motion

Challenges: ambiguity



slide credit: Fei-Fei, Fergus & Torralba

Challenges: local ambiguity



Source: Rob Fergus and Antonio Torralba

Challenges: local ambiguity

Source: Rob Fergus and Antonio Torralba

Challenges: ambiguity

- Many different 3D scenes could have given rise to a particular 2D picture

Challenges or opportunities?

- Images are confusing, but they also reveal the structure of the world through numerous cues
- Our job is to interpret the cues!

Depth cues: Linear perspective

Depth cues: Parallax

Shape cues: Texture gradient

Shape and lighting cues: Shading

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Shape and lighting cues: Shading

Source: http://forum.bodybuilding.com/showthread.php?t=159440991

Shape and lighting cues: Shading

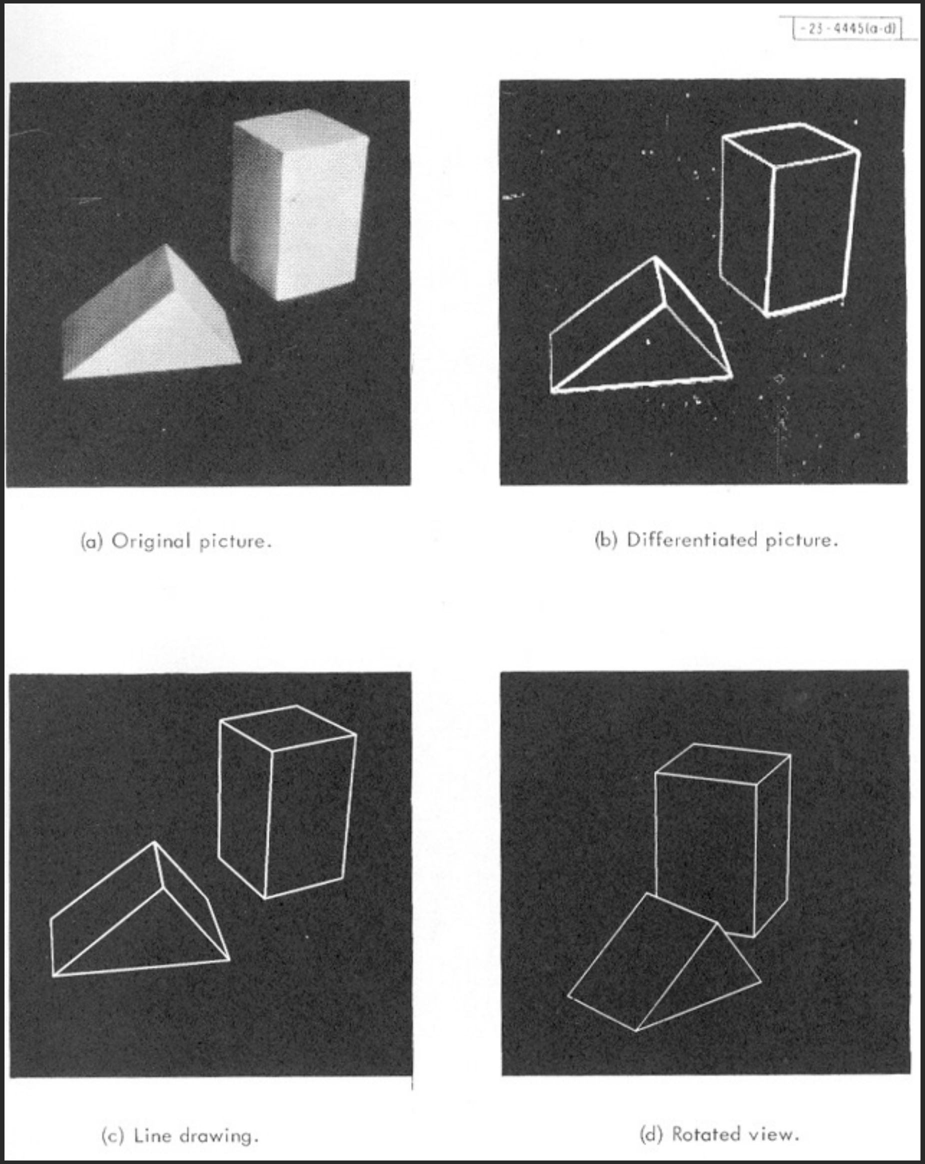
Source: http://galleryhip.com/famous-women-without-makeup.html

Grouping cues: Similarity (color, texture, proximity)

Grouping cues: “Common fate”

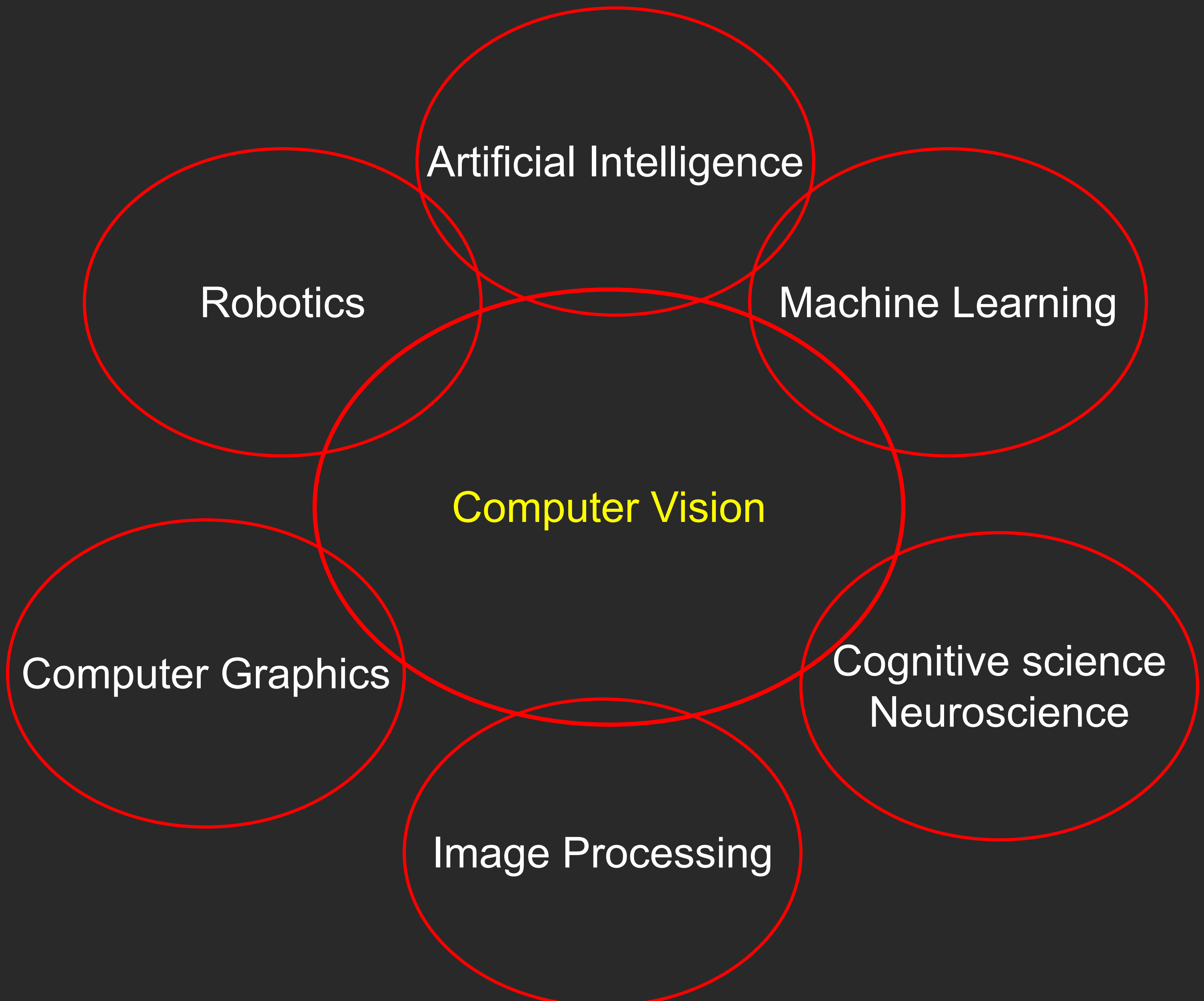
Image credit: Arthus-Bertrand (via F. Durand)

Origins of computer vision



L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Connections to other disciplines



The computer vision industry

- Corporate sponsors of CVPR:

Evaluación

- Filosofía: la visión artificial se experimenta mejor con la práctica.

- Tres cortes: 30%, 35%, 35%.
- Examenes parciales : 40%, 40%, 20%
  - Examen práctico-teorico.

- Tareas de programación y laboratorios: 40%, 40%, 20%

- Proyecto final: 10%, 10%, 60%
  - Una competencia?

- Participación y quices: 10%, 10%, 0%
  - Quices
  - Preguntas y respuestas en clase, foros, etc.

	1er Corte (30%)	2o Corte (35%)	3er Corte (35%)
Examen parcial	40%	40%	20%
Tareas y Labs	40%	40%	20%
Quices	10%	10%	
Proyecto	10%	10%	60%

Políticas sobre colaboración

- Pueden discutir las tareas entre ustedes, pero los programas se hacen individualmente (o en grupo cuando sea el caso).
- Puedes incorporar tips o código de internet siempre y cuando no vuelva la tarea trivial y debes citar las fuentes.
- Recuerda: Yo también sé usar Google.

Course overview

I. Early vision: Image formation and processing

II. Mid-level vision: Grouping and fitting

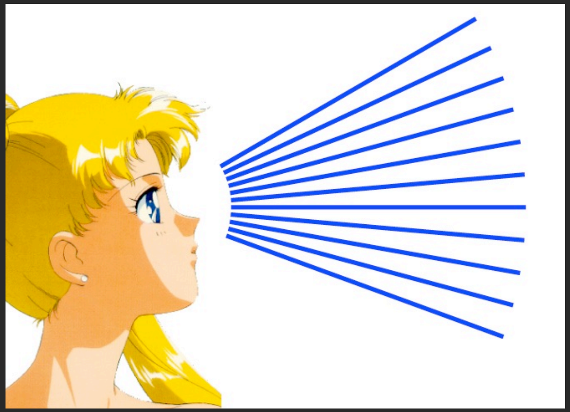
III. Multi-view geometry

IV. Recognition

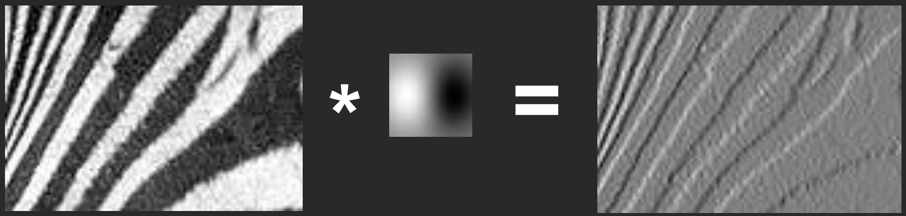
V. Additional topics

I. Early vision

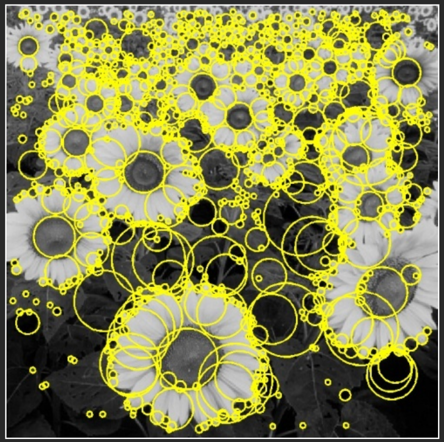
- Basic image formation and processing



Cameras and sensors
Light and color



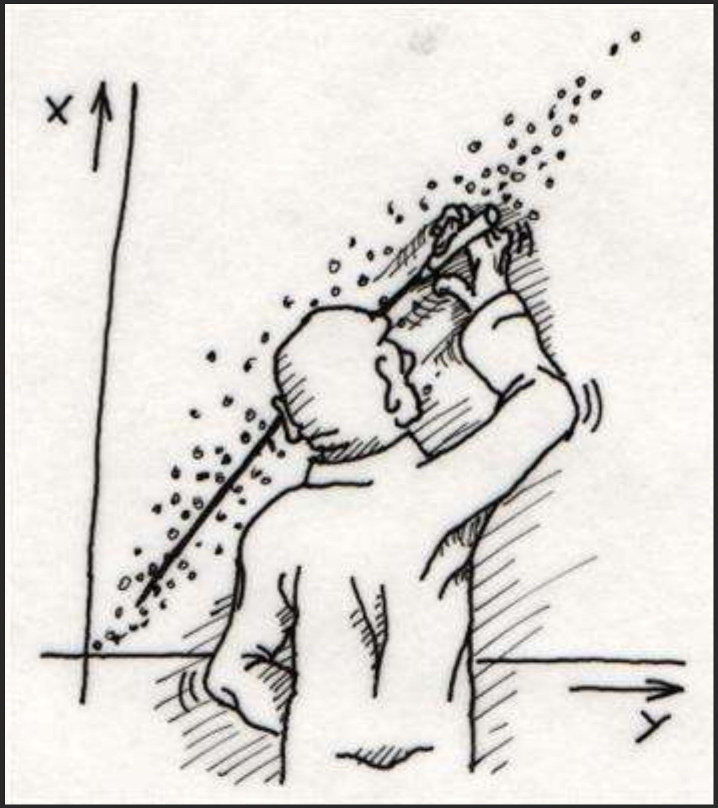
Linear filtering
Edge detection



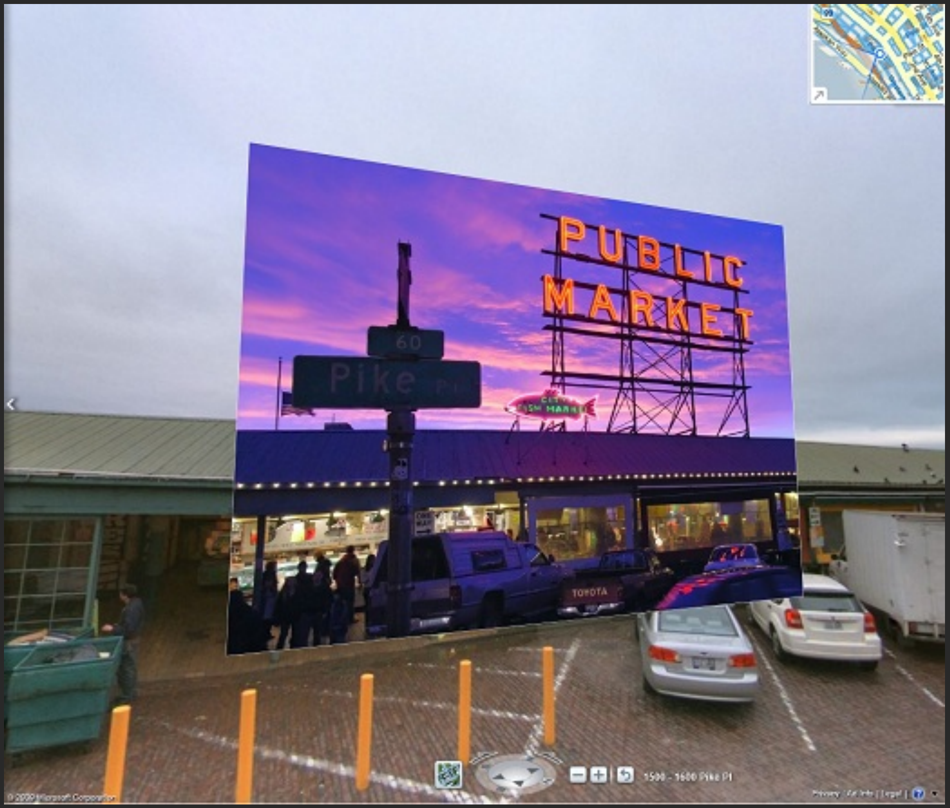
Feature extraction, feature tracking

II. “Mid-level vision”

- Fitting and grouping

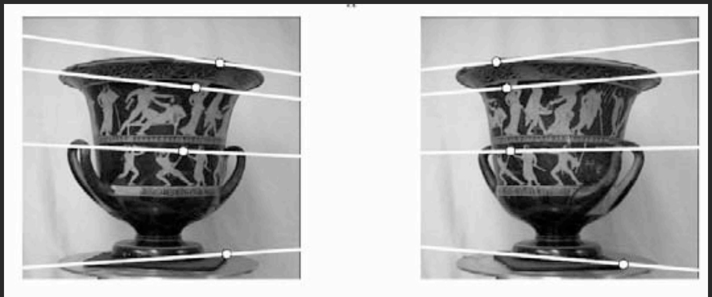


Fitting: Least squares
Hough transform
RANSAC



Alignment

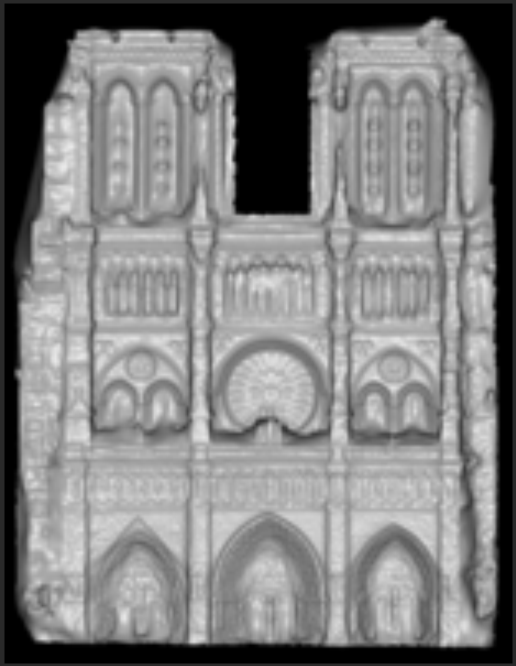
III. Multi-view geometry



Epipolar geometry

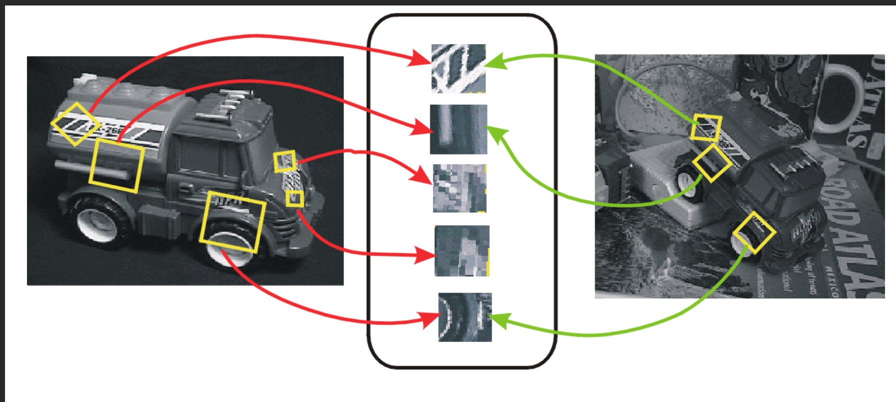
Stereo

Structure from motion



3D Photography

IV. Recognition



Instance recognition, large-scale alignment

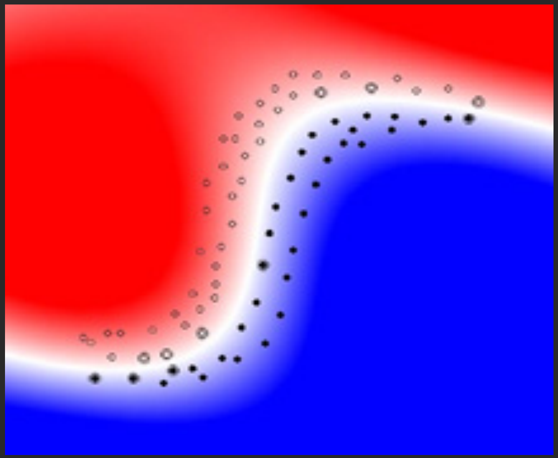
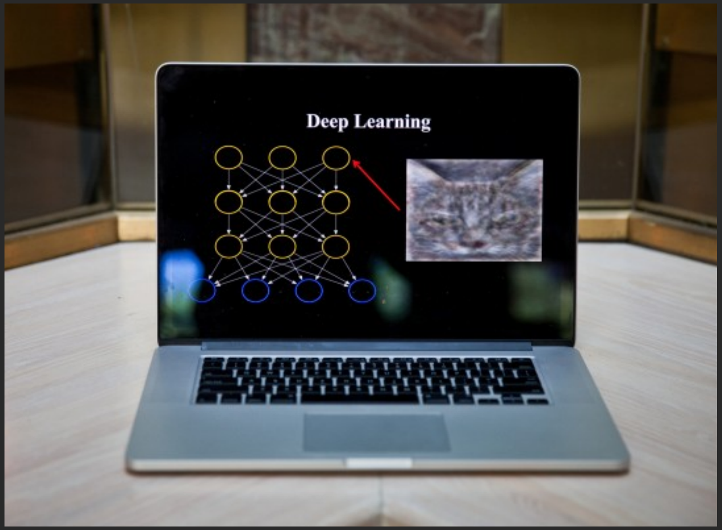


Image classification

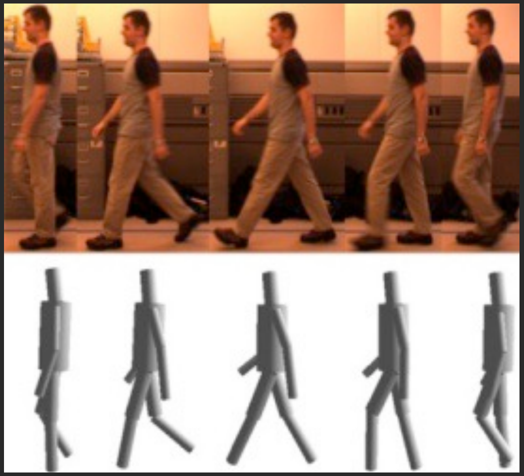
Object detection



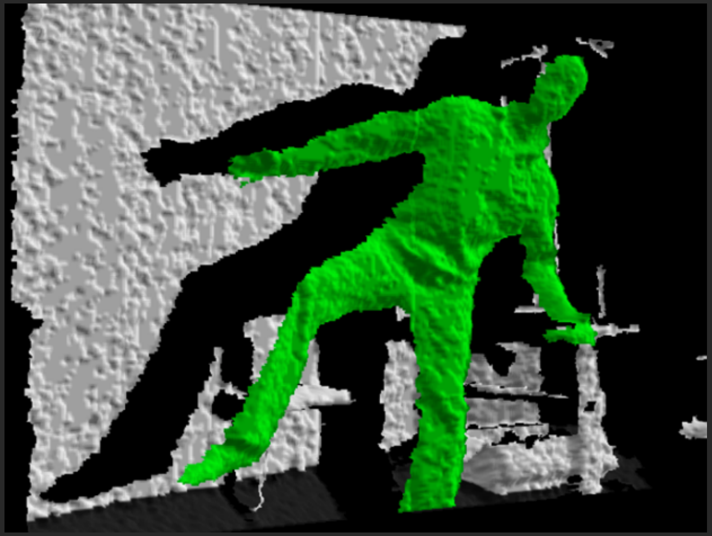
Deep learning

V. Additional Topics (time permitting)

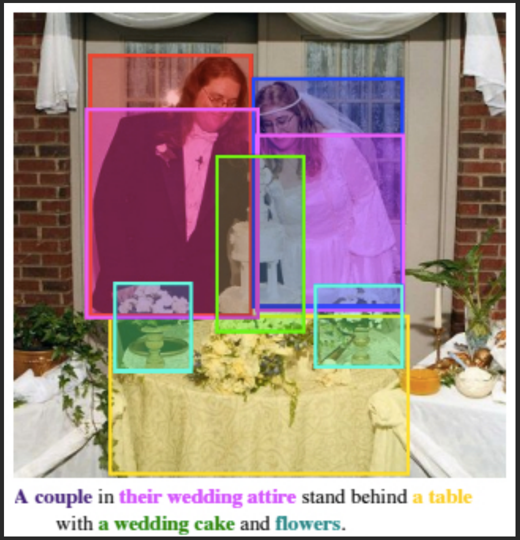
Segmentation



Video



RGBD images



Images and text