

Computer Vision

(in 5 minutes...)

Prof. Giuseppe Lisanti
giuseppe.lisanti@unibo.it

What is Computer Vision?

From pixels to meaning (and measurement)

In one sentence:

Computer Vision is about extracting information from images/video

Tasks

Feature extraction

Classification

Segmentation

Calibration

Detection

Vision+Language

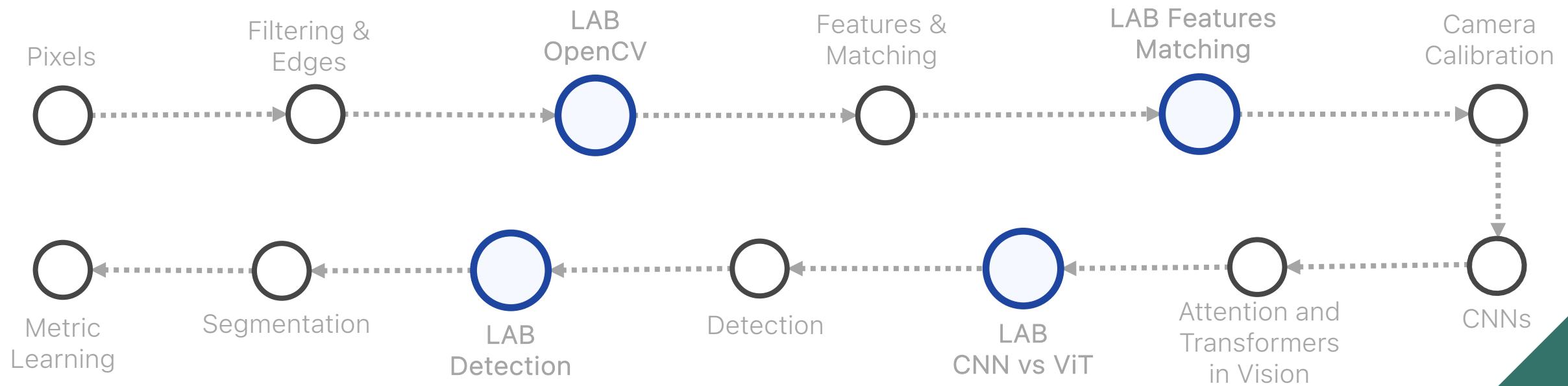
Why?

- Understanding: objects, people, scenes, actions
- Measurement: depth, 3D shape, geometry
- Search & interaction: "find similar" / "answer about the image"

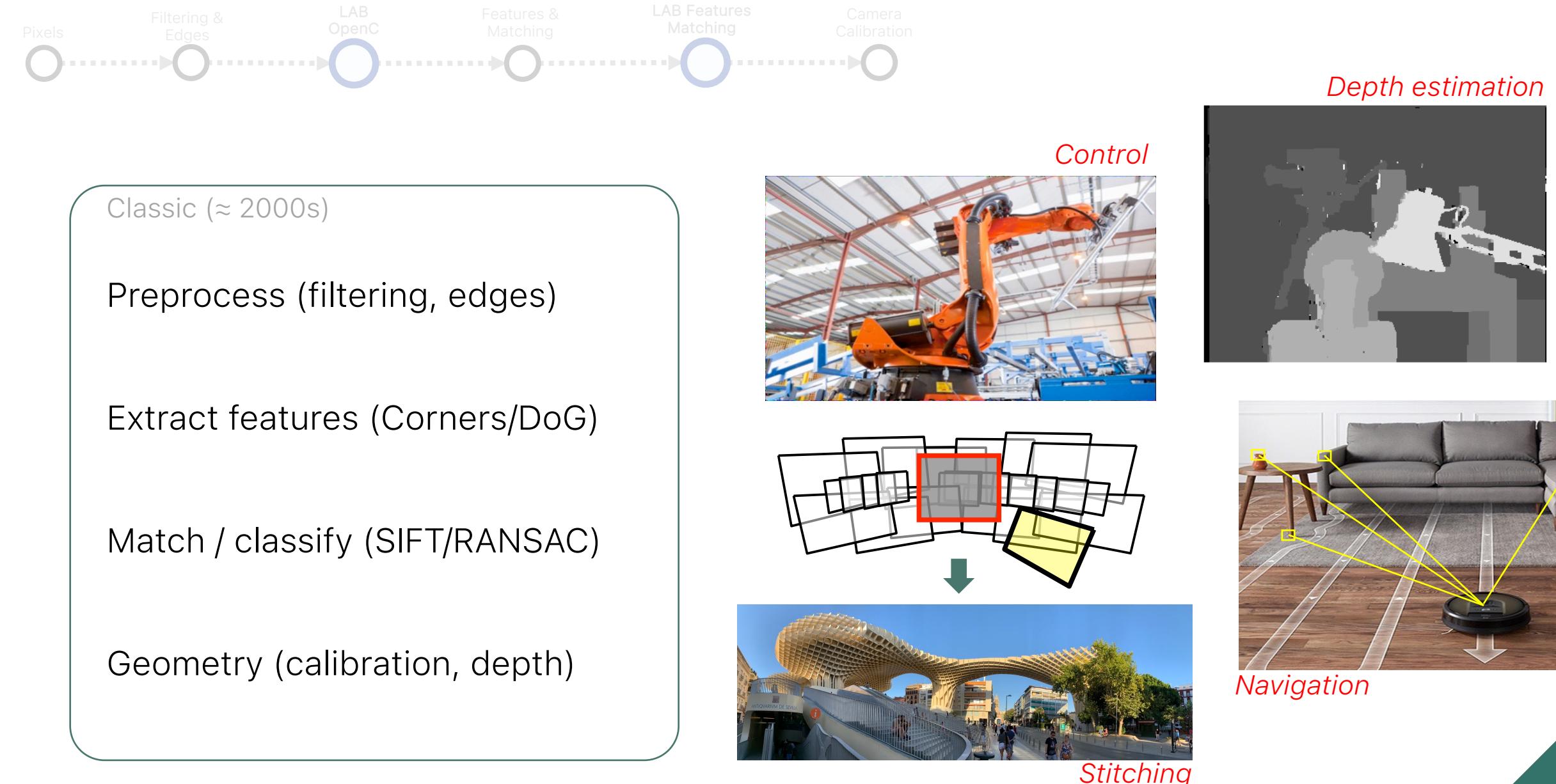
Course info and map

Our journey through lectures and laboratories

- The course is in English
- Prerequisites: Linear Algebra, Basic knowledge of Machine Learning, Programming and Python
- 2025/2026 Second semester => 2026/2027 First semester



From classic vision to modern vision



From classic vision to modern vision



Detection and Segmentation

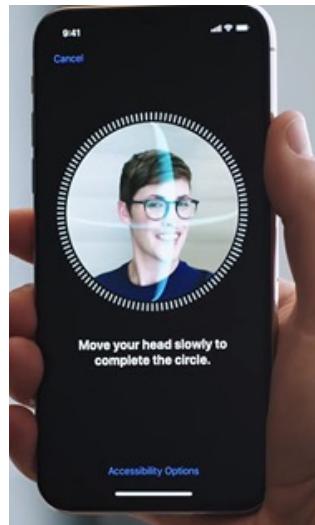


Neural Graphics (NERF)



Text-to-Image (Google)

Verification



Modern (2012 → today)

CNNs → large-scale ImageNet era

Transformers & attention (ViT)

Self-supervised learning
(DINO/MAE)

Multimodal & generative models
(vision+language, diffusion)

From classic vision to modern vision

- ✓ Interpretable steps
- ⚠ Lots of manual design

Classic (\approx 2000s)

Preprocess (filtering, edges)

Extract features (Corners/DoG)

Match / classify (SIFT/RANSAC)

Geometry (calibration, depth)



- ✓ End-to-end learning
- ⚠ Needs data + compute

Modern (2012 \rightarrow today)

CNNs \rightarrow large-scale ImageNet era

Transformers & attention (ViT)

Self-supervised learning
(DINO/MAE)

Multimodal & generative models
(vision+language, diffusion)

Thank you

Questions?