MAI4CAREU | Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 1: Introduction to Computer Vision

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

Co-financed by the European Union

Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Who we are
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications
- What we do

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Who we are

Visual Computing Group at CYENS Centre of Excellence

Melinos Averkiou
MRG Leader

Yiangos Georgiou
Research Associate (DTP)

Marios Loizou
Research Associate

Yeshwanth Kumar Adimoolam
Research Associate (DTP)

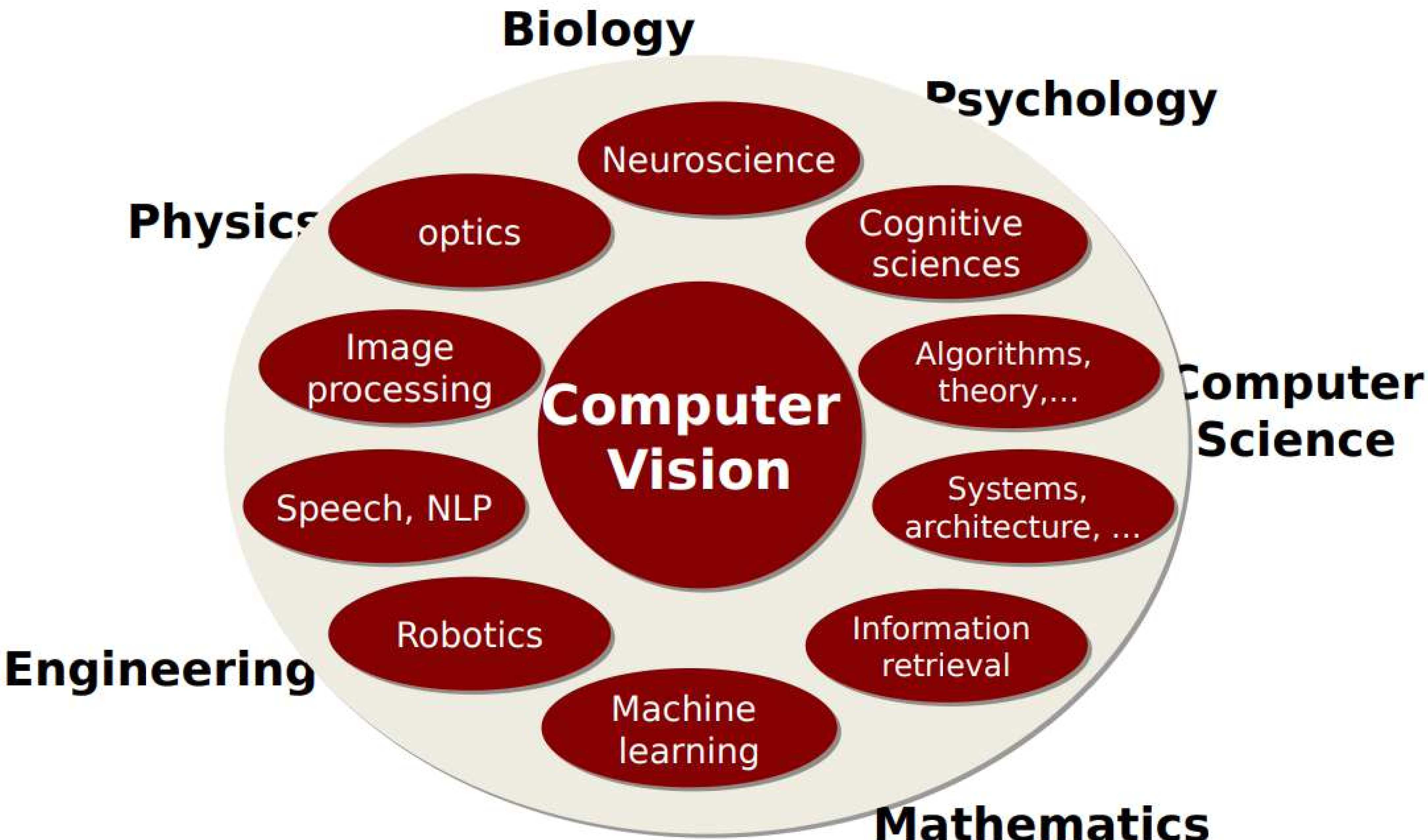
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

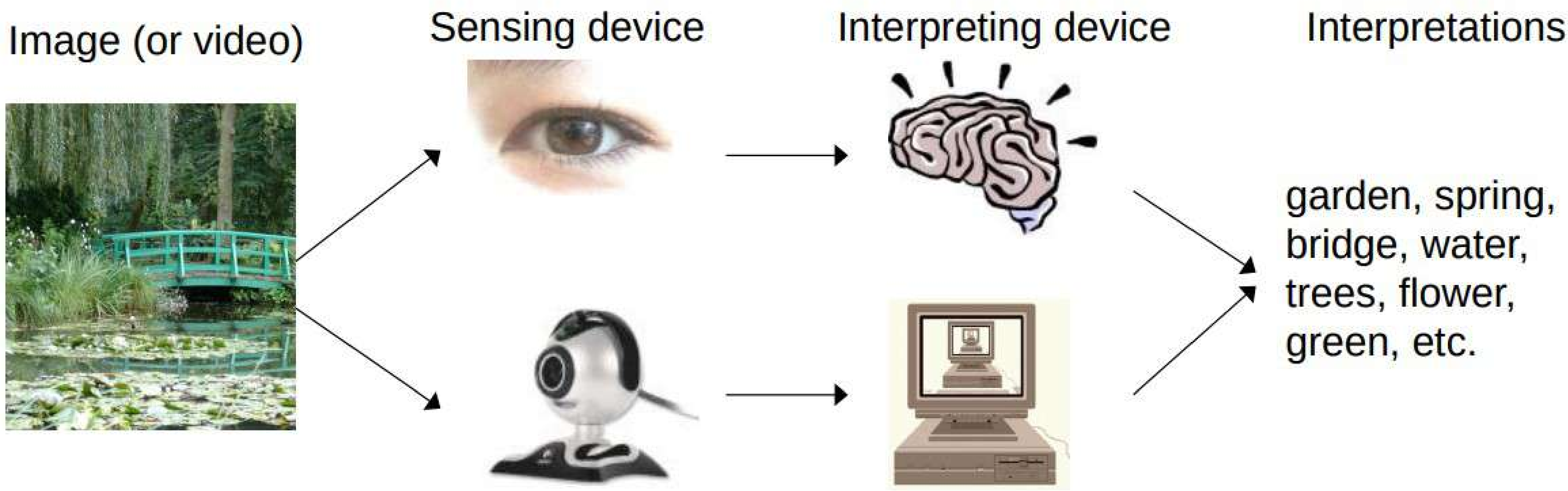
- Course Overview
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications
- What we do

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423


Biology
Psychology
Physics
Computer Science
Engineering
Mathematics
Neuroscience
optics
Cognitive sciences
Image processing
Computer Vision
Algorithms, theory,...
Speech, NLP
Systems, architecture, ...
Robotics
Information retrieval
Machine learning

[Niebles]

What is (Computer) Vision ?



[Niebles, Fergus]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What is (Computer) Vision?

- *Vision* is about discovering from images
  - **what** is present in the scene, and
  - **where** it is
- In *Computer Vision* a ***camera*** (or several ***cameras***) is linked to a computer
- The computer interprets ***images*** of a scene to obtain information
- Useful for tasks such as navigation, manipulation and recognition

What is Computer Vision?

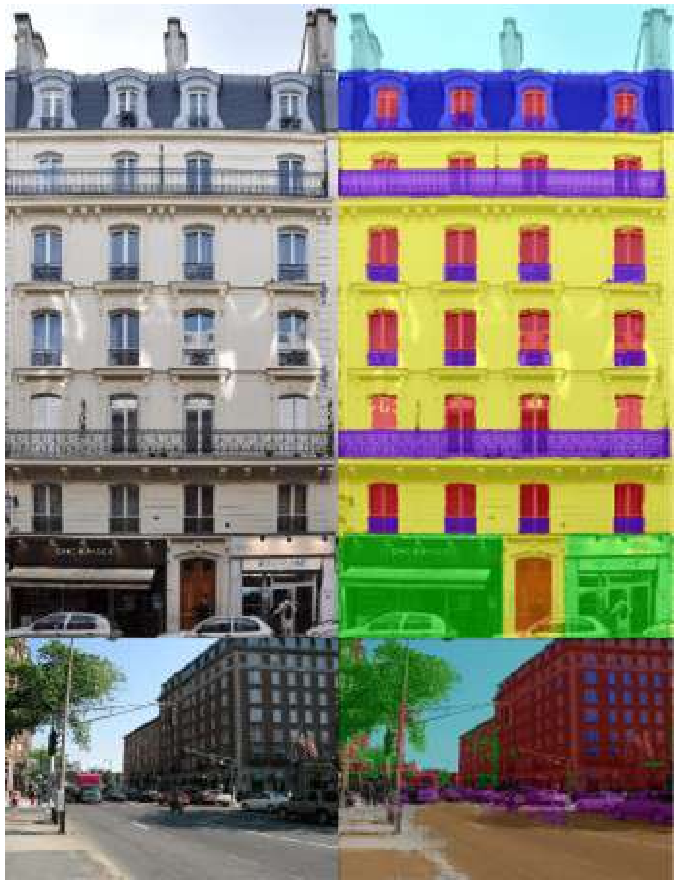
Computer Vision's **goal** is to obtain a high-level **understanding** of the world using images as input

Co-financed by the European Union
Connecting Europe Facility

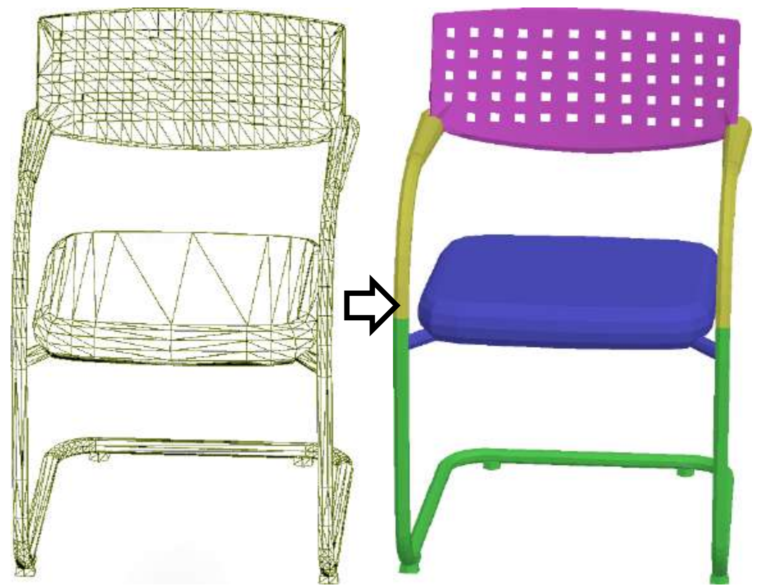
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Understand = Obtain **Semantics** & **Geometry**

3D Object layout
Input: RGBD Image



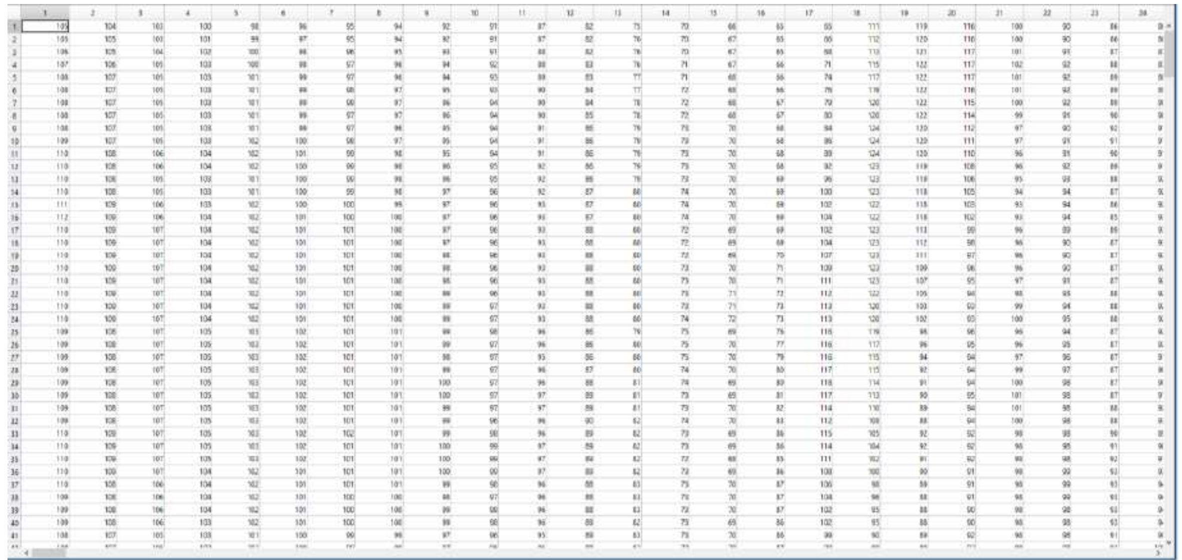
Building facade segmentation
Input: RGB Image



Object semantic segmentation
Input: 3D Mesh

Understanding the world is hard for machines

What we see



What the machine sees

Today's Agenda

- Course Overview
- Introduction to Computer Vision
    - What is Computer Vision
    - How hard is Computer Vision
    - Why is Computer Vision so hard
    - How to organize Computer Vision
    - Why study Computer Vision
    - Applications
- What we do

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

How hard is Computer Vision ?

- The Summer Vision Project – MIT AI Memo 100, 1966
  - ‘Solve vision in a summer project’ - almost an urban legend
    - Basic foreground/background segmentation,
    - Analyse scenes with simple non-overlapping objects,
    - Extend the system to more complex objects.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

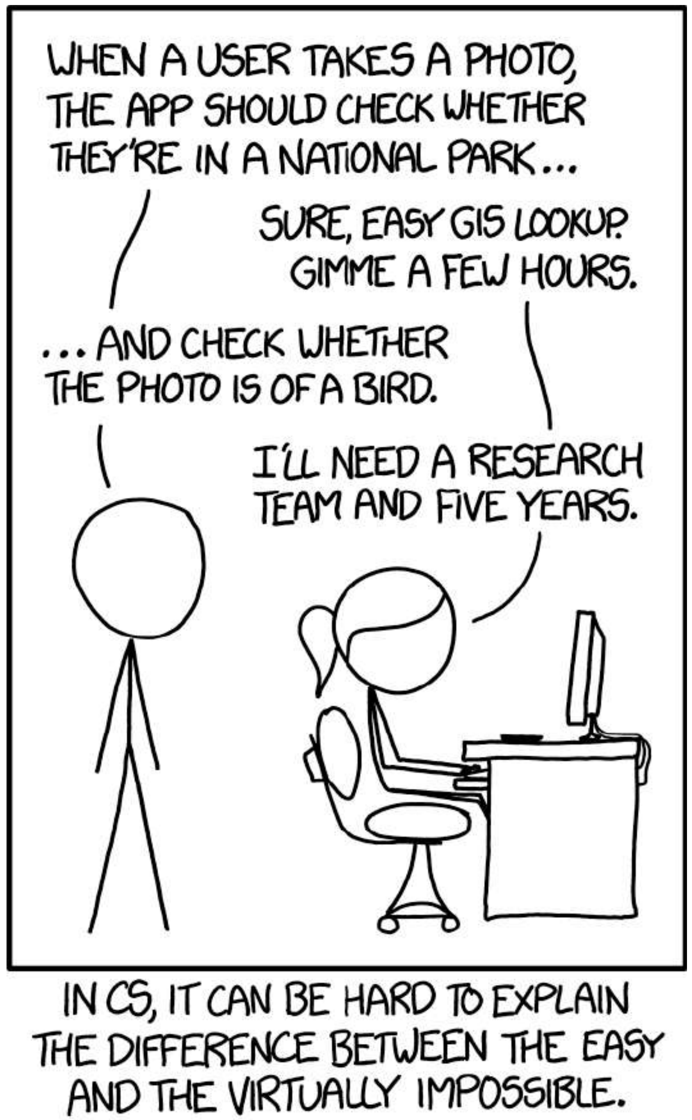
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

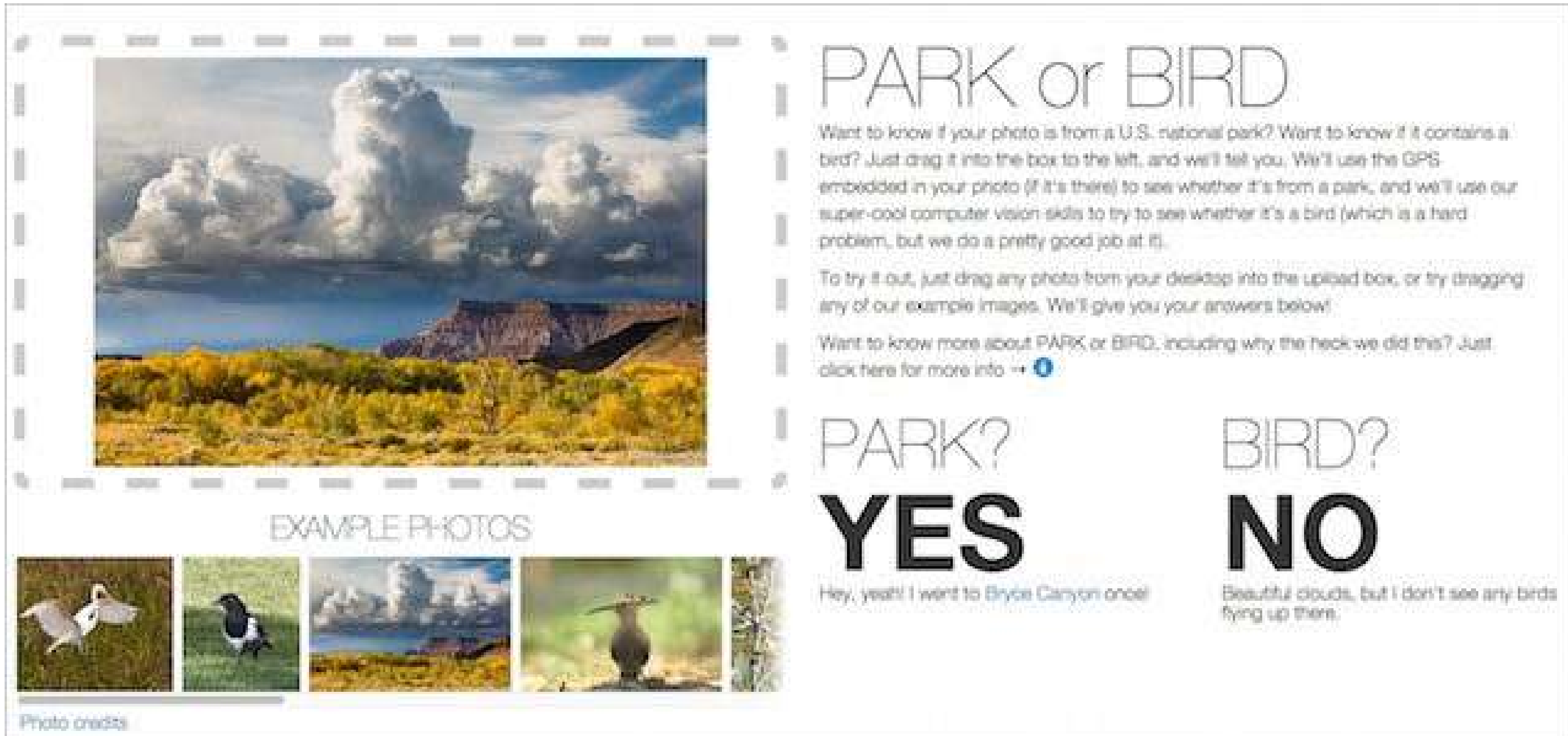
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How hard is Computer Vision ?



[XKCD]

How hard is Computer Vision ?



[Flickr 'solved' it]

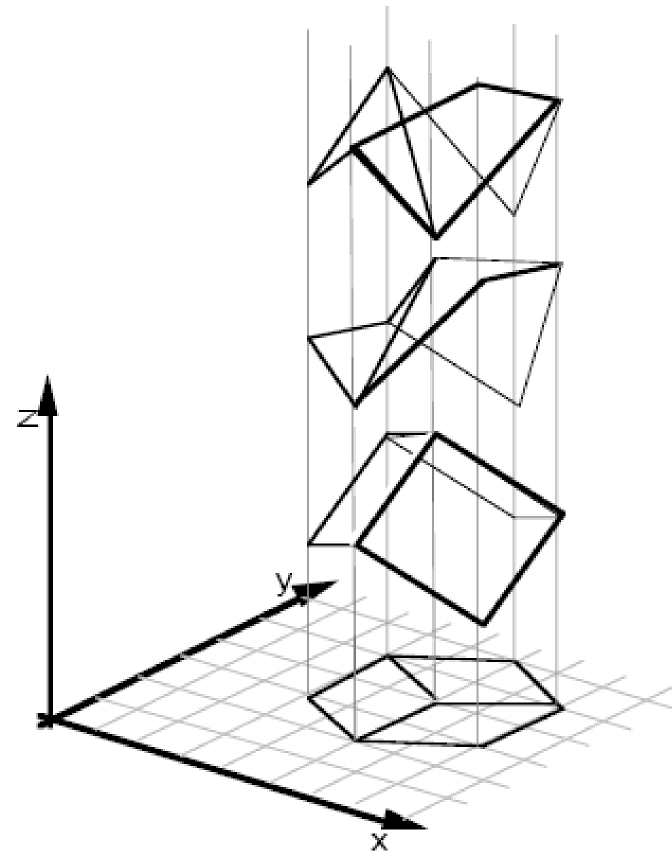
Today's Agenda

- Course Overview
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications
- What we do

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

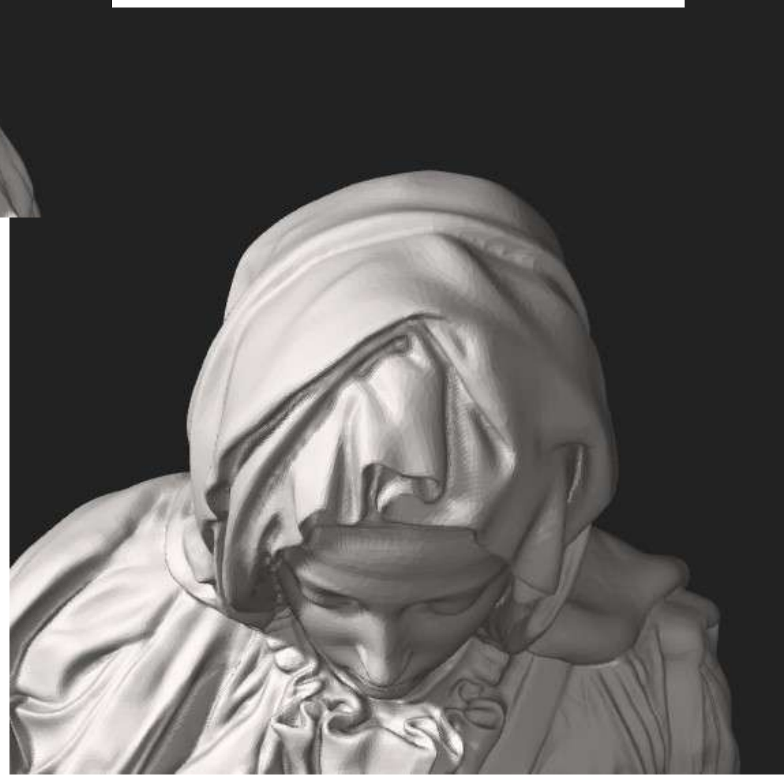
Why is Computer Vision so hard?

Because it is an ill-posed problem



[Sinha and Adelson 1993]

Challenge 1: viewpoint variation



Madonna della Pietà,
Michelangelo Buonarroti, 1498-99

[Fei Fei, Fergus & Torralba]

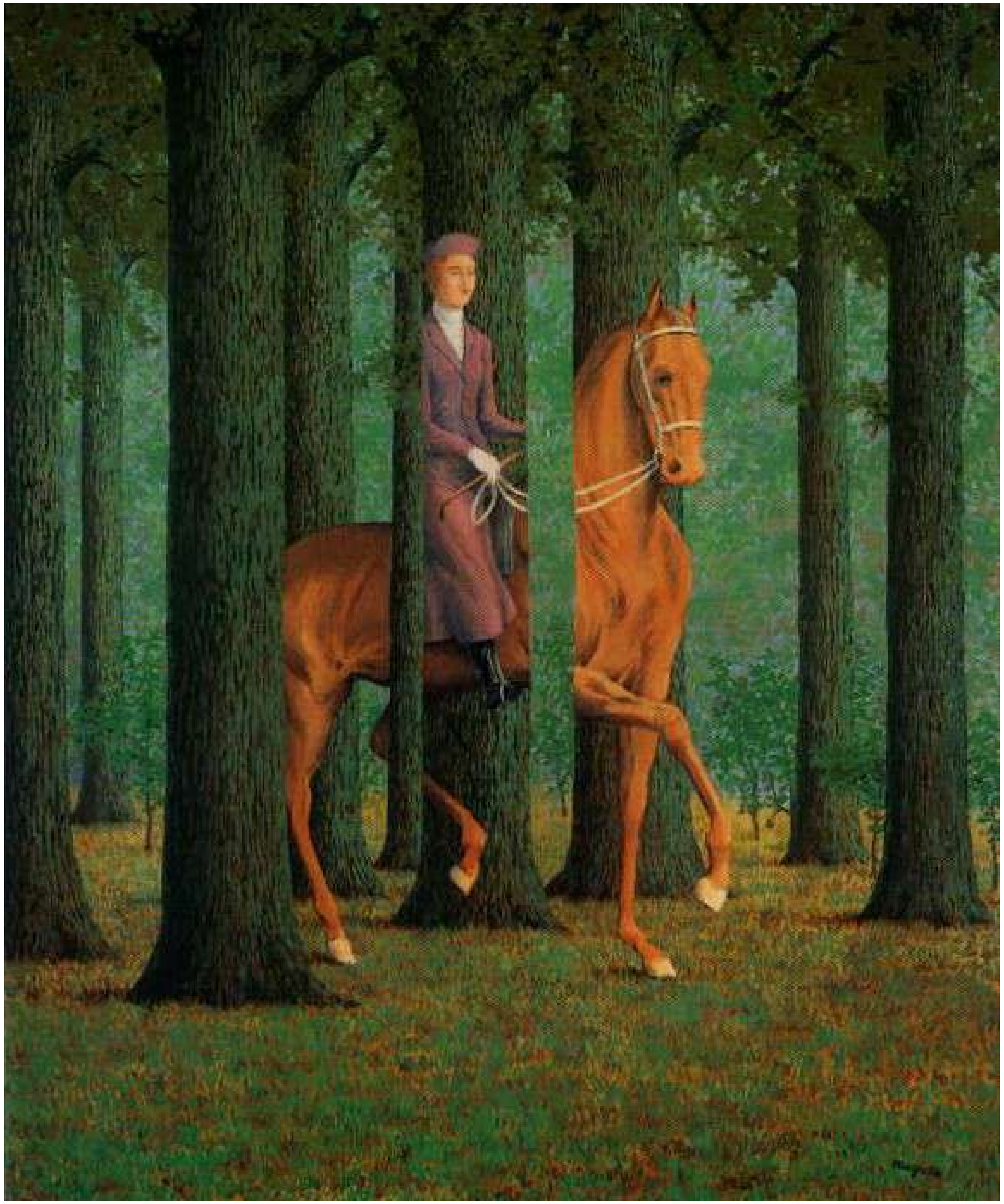
Challenge 2: illumination

[S. Ullman]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Challenge 3: occlusion



The Blank Signature, Rene Magritte, 1965

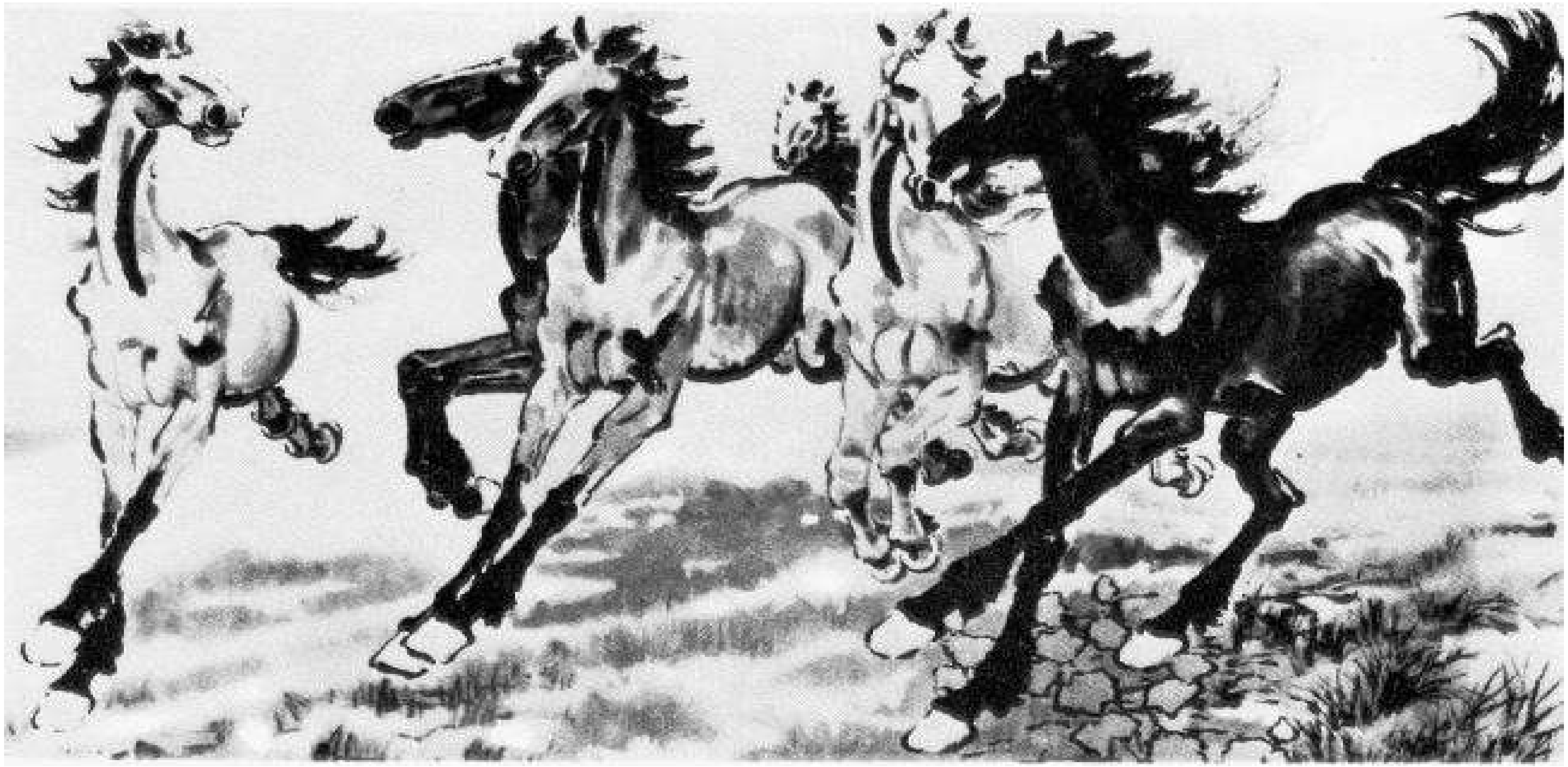
[Fei Fei, Fergus & Torralba]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Challenge 4: scale

[Fei Fei, Fergus & Torralba]

Challenge 5: deformation



Six Galloping Horses,
Xu Beihong, 1942

[Fei Fei, Fergus & Torralba]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Challenge 6: background clutter

The Maiden,
Gustav Klimt, 1913

[Fei Fei, Fergus & Torralba]

Co-financed by the European Union
Connecting Europe Facility

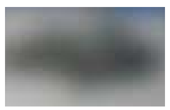
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Challenge 7: object intra-class variation

[Fei Fei, Fergus & Torralba]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Challenge 8: local ambiguity



[Fei Fei, Fergus & Torralba]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Course Overview
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications
- What we do

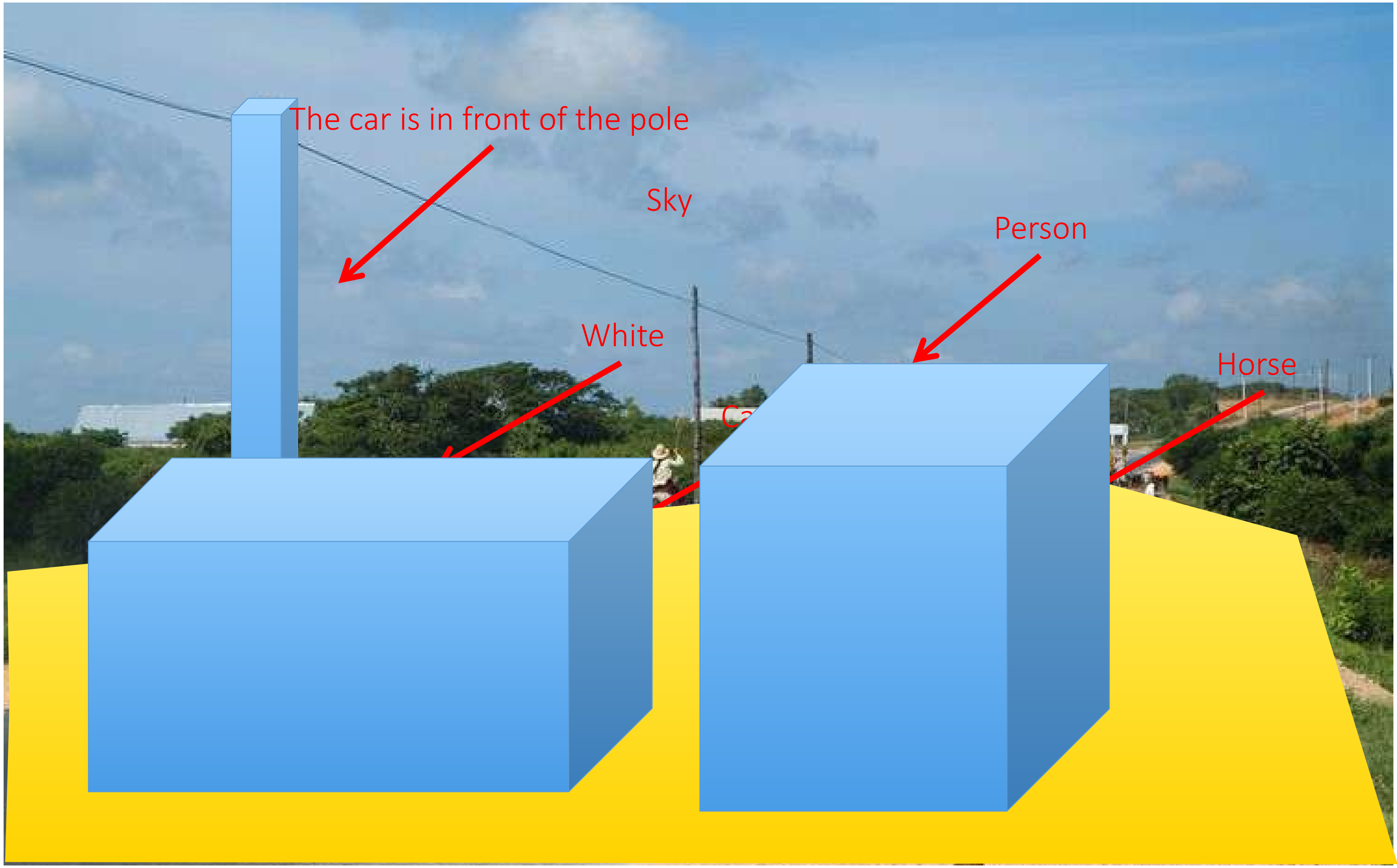
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How to organize Computer Vision ?

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

[Shapiro]



[Shapiro]

How to organize Computer Vision ?

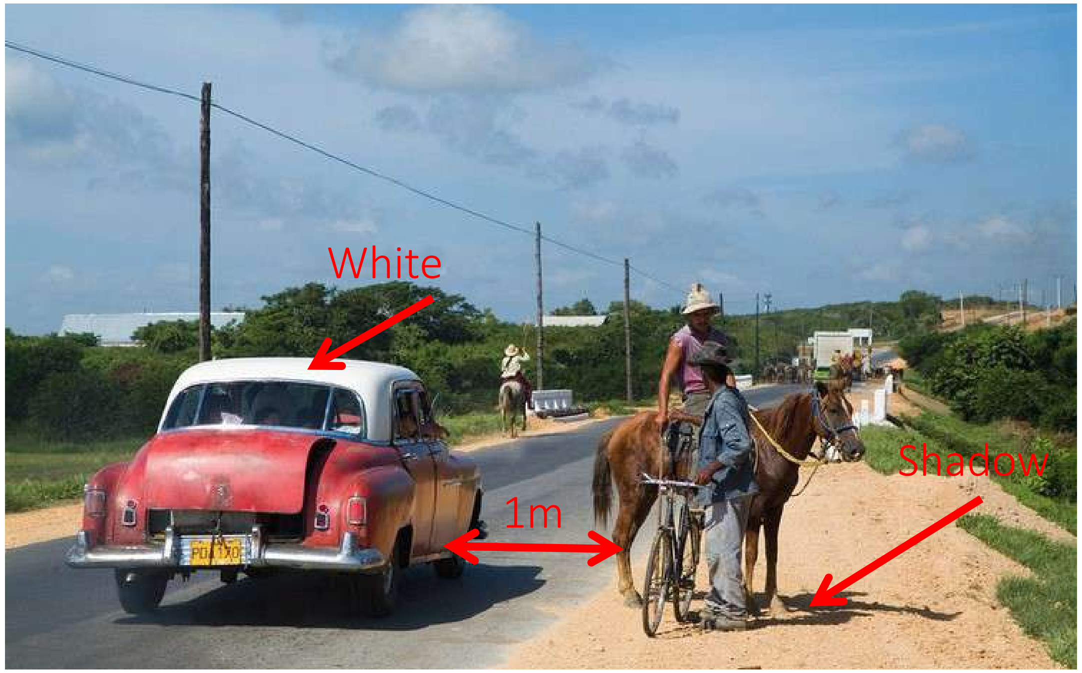
- Low Level Vision
  - Measurements
  - Enhancements
  - Region segmentation
  - Features
- Mid Level Vision
  - Reconstruction
  - Depth
  - Motion Estimation
- High Level Vision
  - Category detection
  - Activity recognition
  - Deep understanding

[Shapiro]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How to organize Computer Vision ?

- Low Level Vision
  - Measurements
  - Enhancements
  - Region segmentation
  - Features
- Mid Level Vision
  - Reconstruction
  - Depth
  - Motion Estimation
- High Level Vision
  - Category detection
  - Activity recognition
  - Deep understanding



[Shapiro]

Low-Level: Exposure

[Redmon]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Low-Level: Edges

[Redmon]

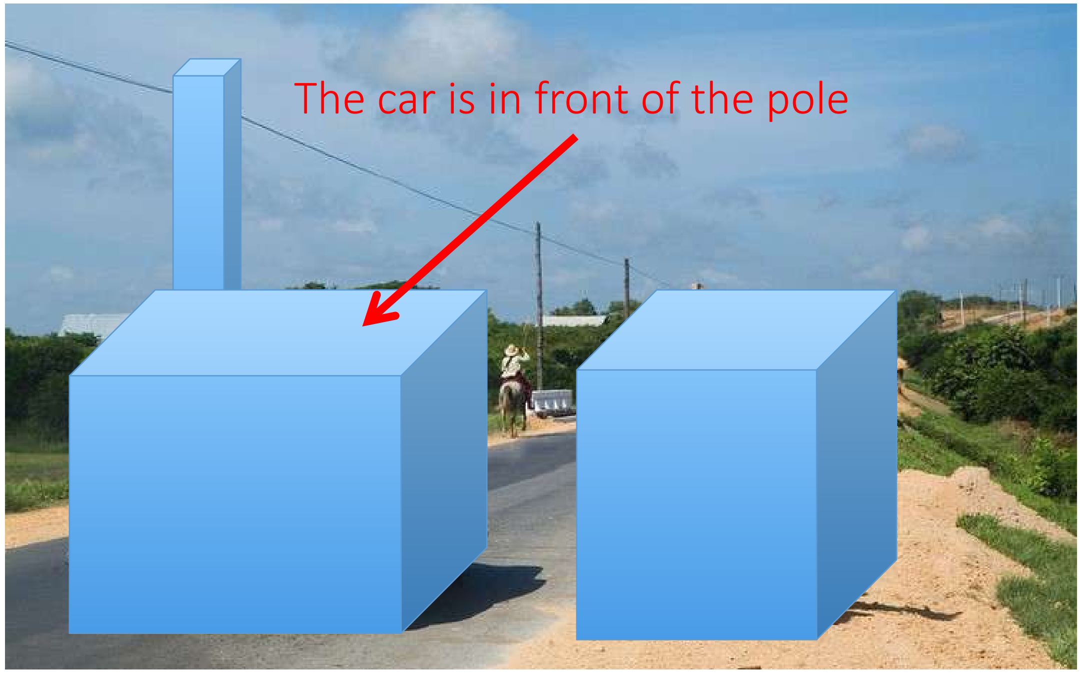
Low-Level: Segmentation (color)

[Redmon]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How to organize Computer Vision ?

- Low Level Vision
  - Measurements
  - Enhancements
  - Region segmentation
  - Features
- Mid Level Vision
  - Reconstruction
  - Depth
  - Motion Estimation
- High Level Vision
  - Category detection
  - Activity recognition
  - Deep understanding



[Shapiro]

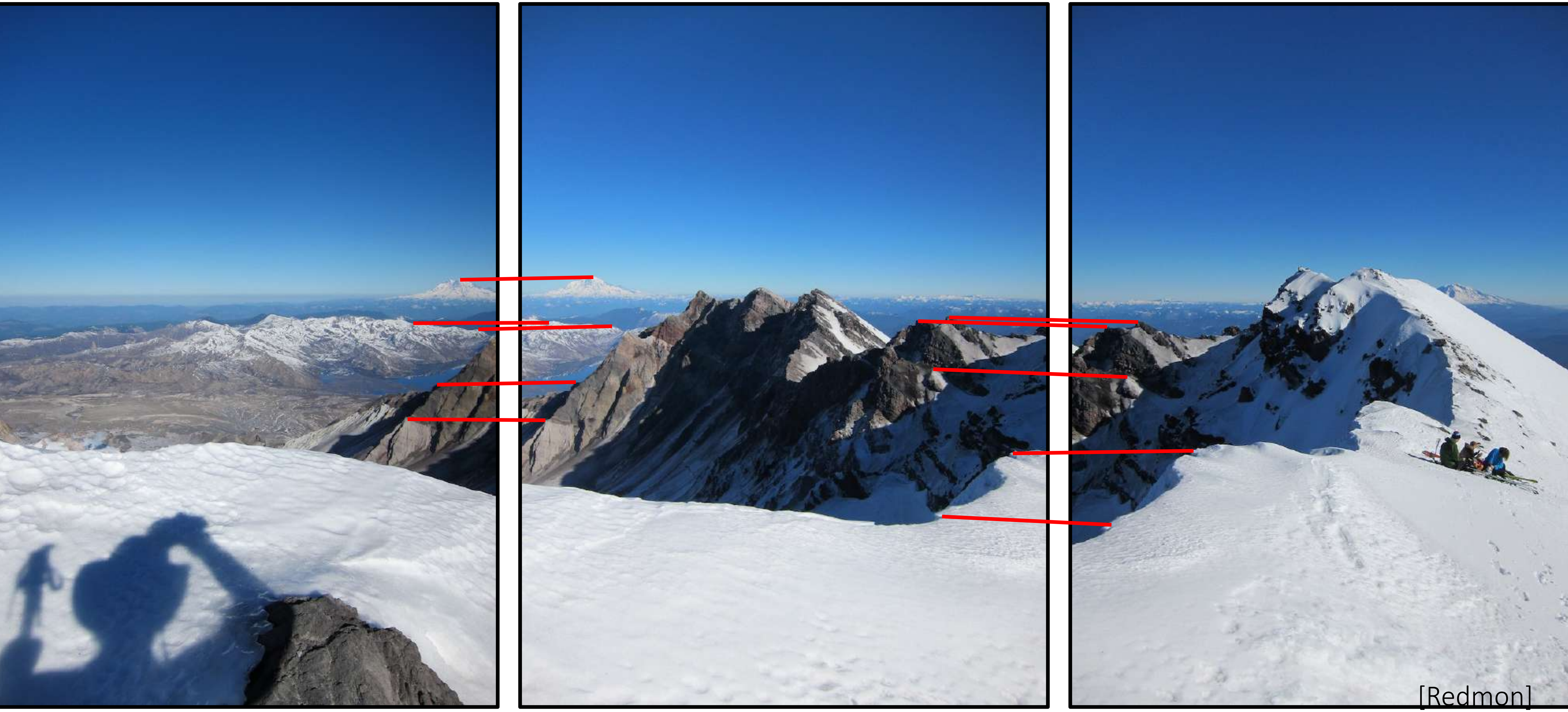
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mid-Level: Panorama Stitching

[Redmon]

Mid-Level: Panorama Stitching



Mid-Level: Panorama Stitching

[Redmon]

Mid-Level: Multi-View Stereo

[Redmon]

Mid-Level: Multi-View Stereo

[Redmon]

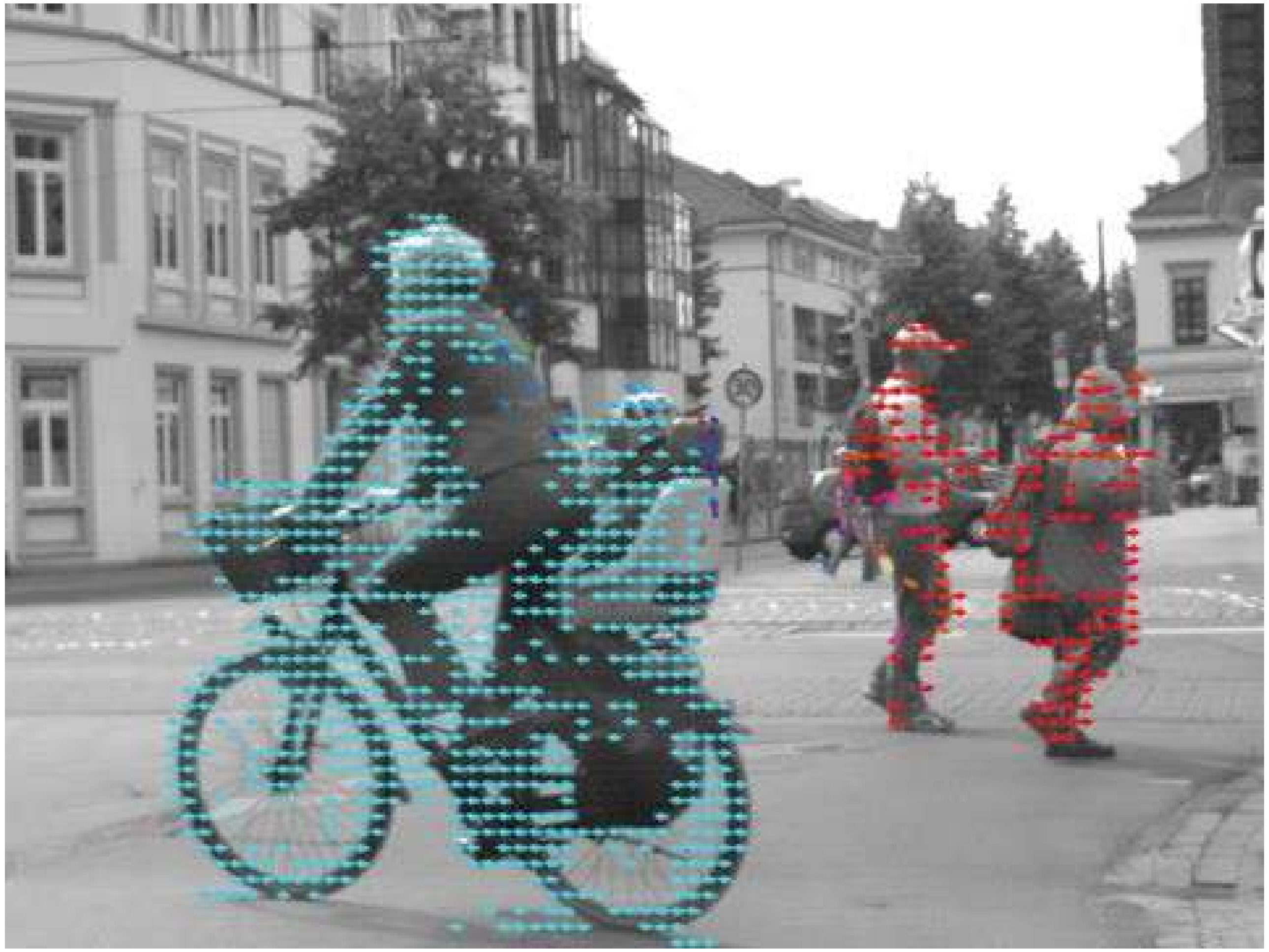
Mid-Level: Multi-View Stereo

[Building Rome in a Day, Agarwal et al., ICCV 2009]

The Colosseum, 2,106 images, 819,242 points

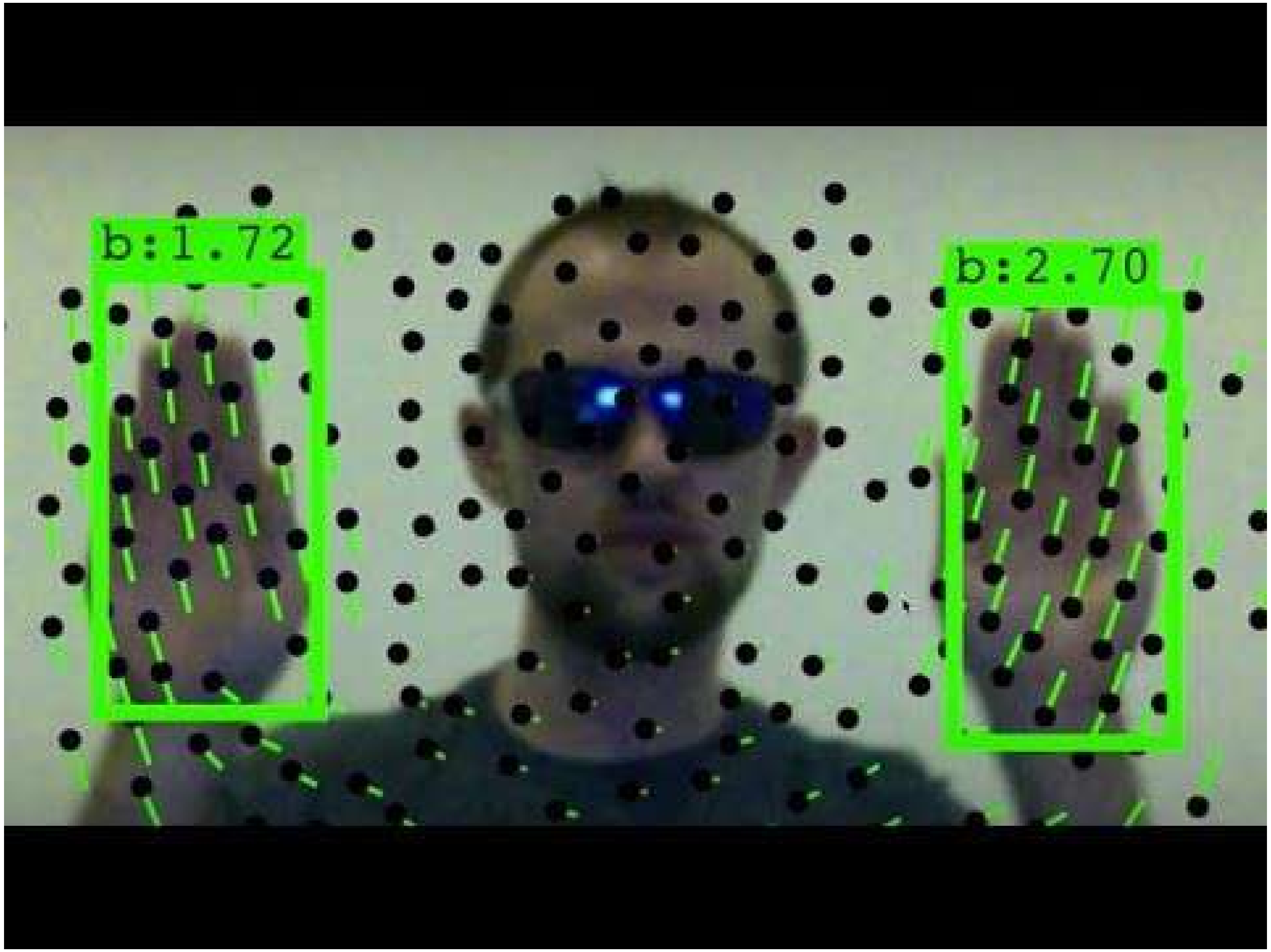
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mid-Level: Optical Flow



[Redmon]

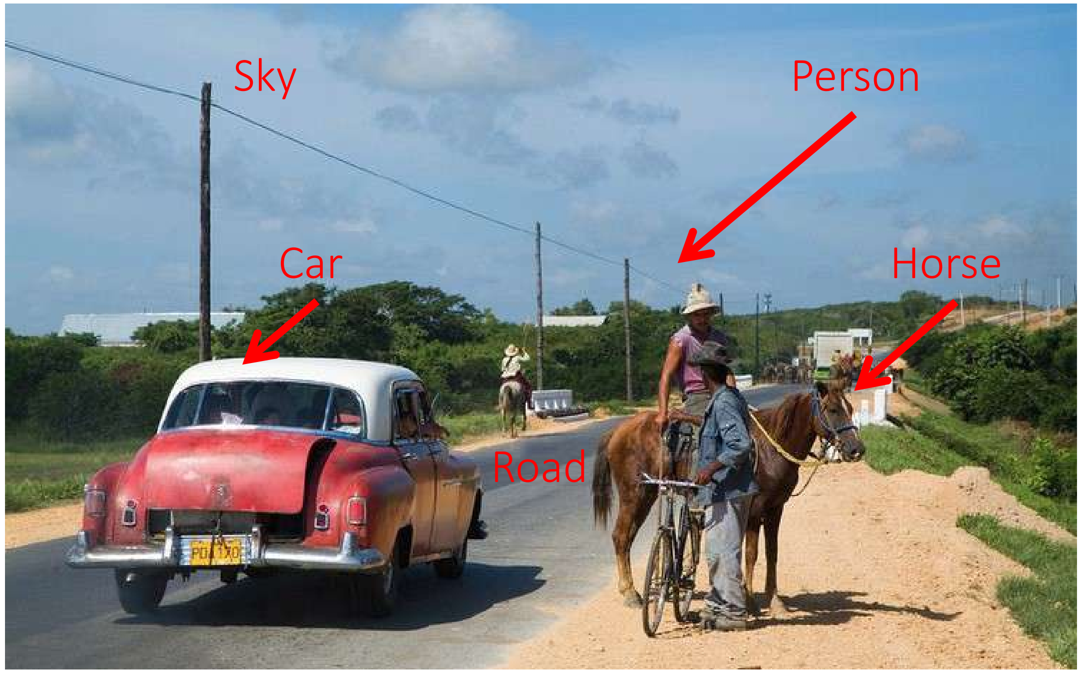
Mid-Level: Optical Flow



[Redmon]

How to organize Computer Vision ?

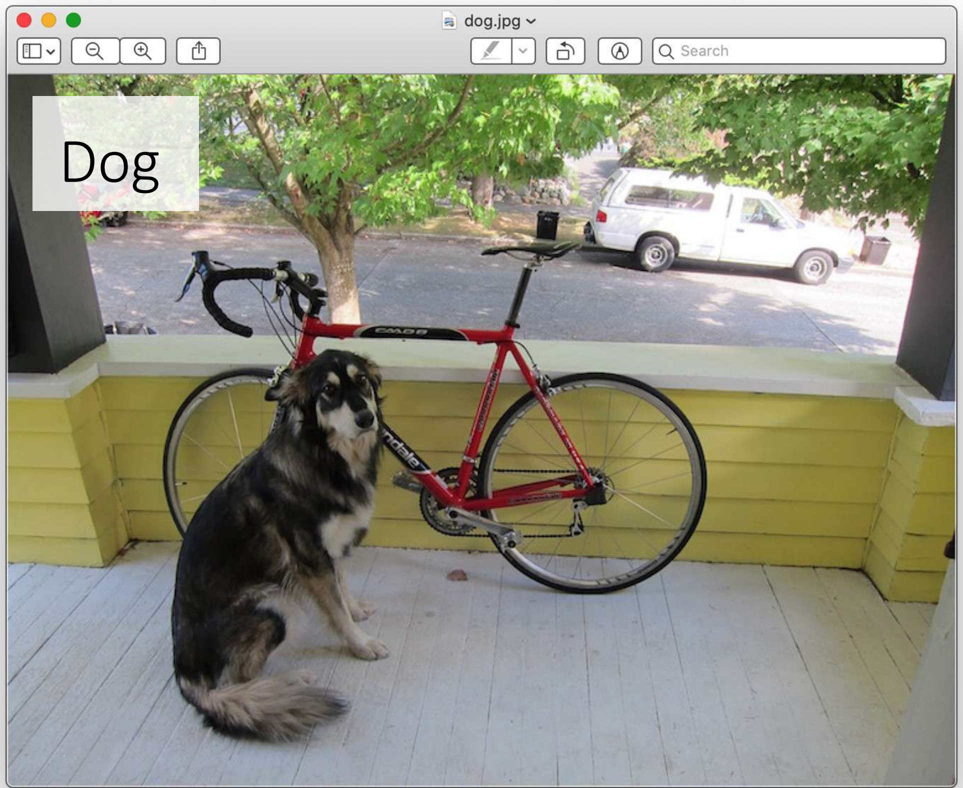
- Low Level Vision
  - Measurements
  - Enhancements
  - Region segmentation
  - Features
- Mid Level Vision
  - Reconstruction
  - Depth
  - Motion Estimation
- High Level Vision
  - Category detection
  - Activity recognition
  - Deep understanding



[Shapiro]

High-Level: Classification

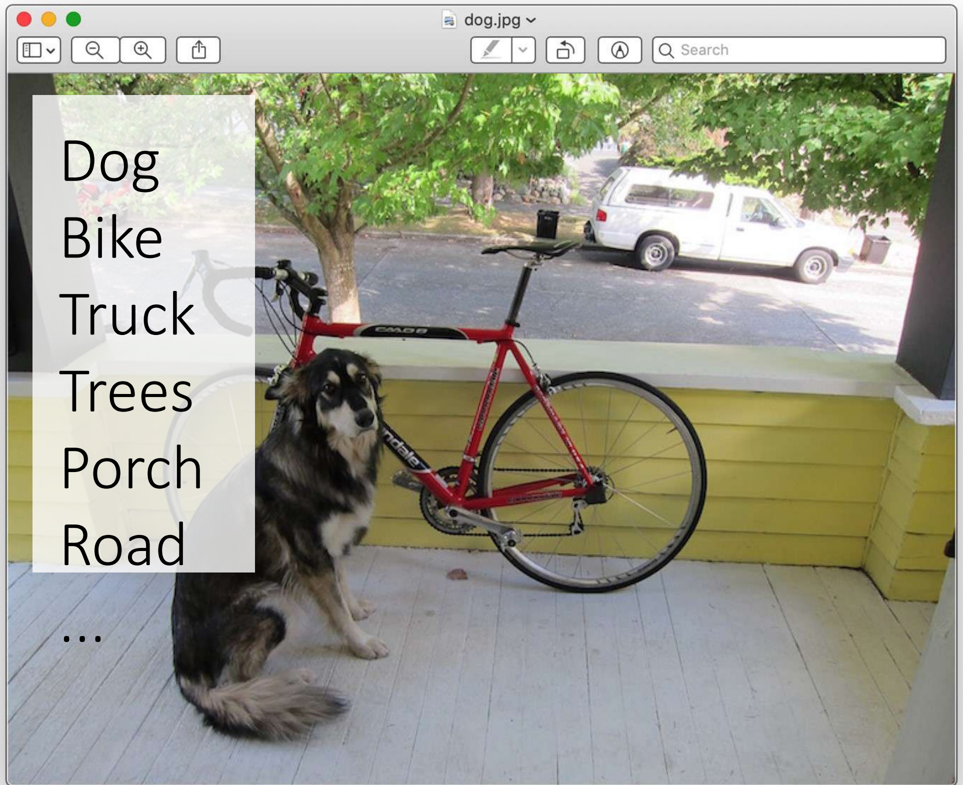
- What is in the image?



[Redmon]

High-Level: Tagging

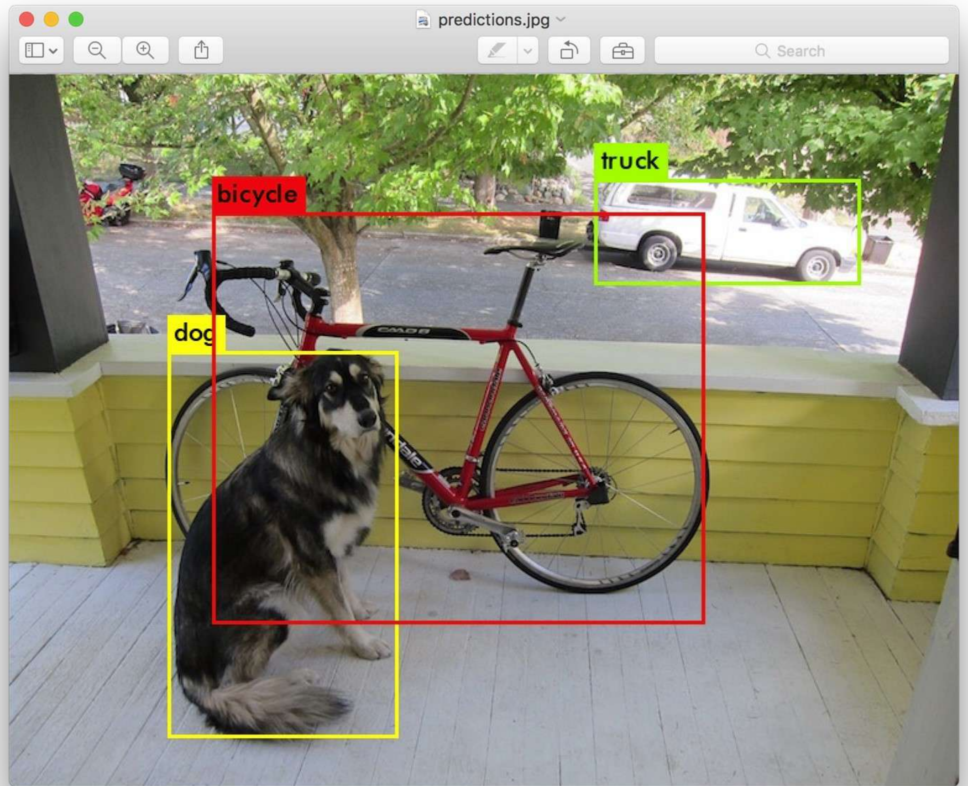
- What are ALL the things in the image?



[Redmon]

High-Level: Detection

- What are ALL the things in the image?
- Where are they?



[Redmon]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

High-Level: Semantic Segmentation

[Redmon]

High-Level: Instance Segmentation

http://www.youtube.com/watch?v=OOT3UIXZztE [Redmon]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Course Overview
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications
- What we do

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Why study Computer Vision?

- *Match* (or *beat*) human vision

  → central to Artificial Intelligence, countless applications

- Understand human vision → neuroscience
- Do research with huge impact
- Get a job in the industry
- Timing is perfect: AI revolution – big data, faster hardware, deep learning

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Do research with huge impact

Google Scholar

Top publications

Categories

	Publication
1.	Nature
2.	The New England Journal of Medicine
3.	Science
4.	The Lancet
5.	IEEE/CVF Conference on Computer Vision and Pattern Recognition
6.	Advanced Materials
7.	Nature Communications
8.	Cell
9.	Chemical Reviews
10.	Chemical Society reviews
11.	Journal of the American Chemical Society
12.	Angewandte Chemie
13.	Proceedings of the National Academy of Sciences
14.	JAMA
15.	Nucleic Acids Research

2021

Google Scholar

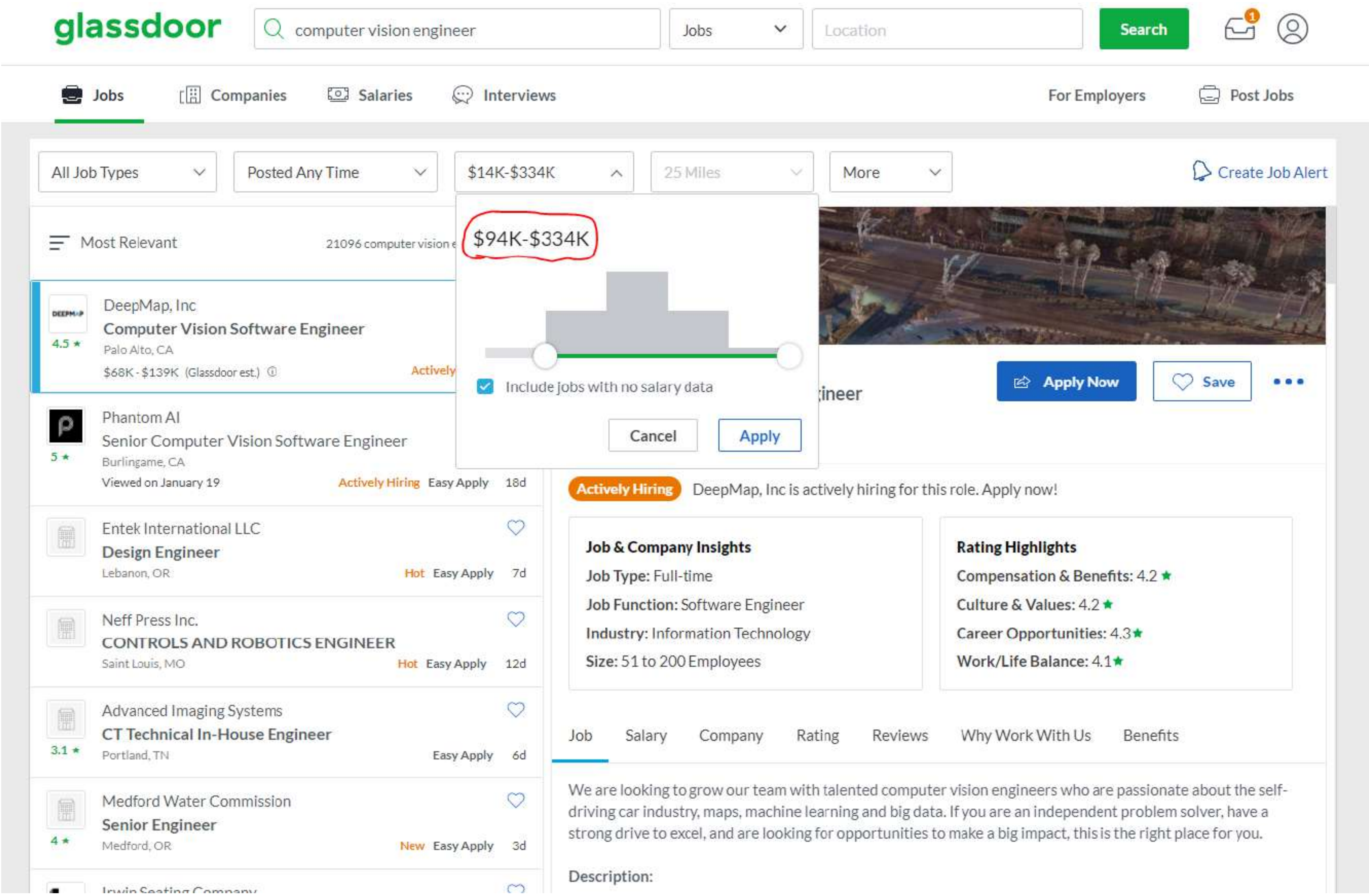
Top publications

Categories English

	Publication	h5-index	h5-median
1.	Nature	414	607
2.	The New England Journal of Medicine	410	704
3.	Science	391	564
4.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	356	583
5.	The Lancet	345	600
6.	Advanced Materials	294	406
7.	Cell	288	459
8.	Nature Communications	287	389
9.	Chemical Reviews	270	434
10.	International Conference on Learning Representations	253	470
11.	JAMA	253	446
12.	Neural Information Processing Systems	245	422
13.	Proceedings of the National Academy of Sciences	245	337
14.	Journal of the American Chemical Society	245	330
15.	Angewandte Chemie	235	314

2022

Get a job in the industry



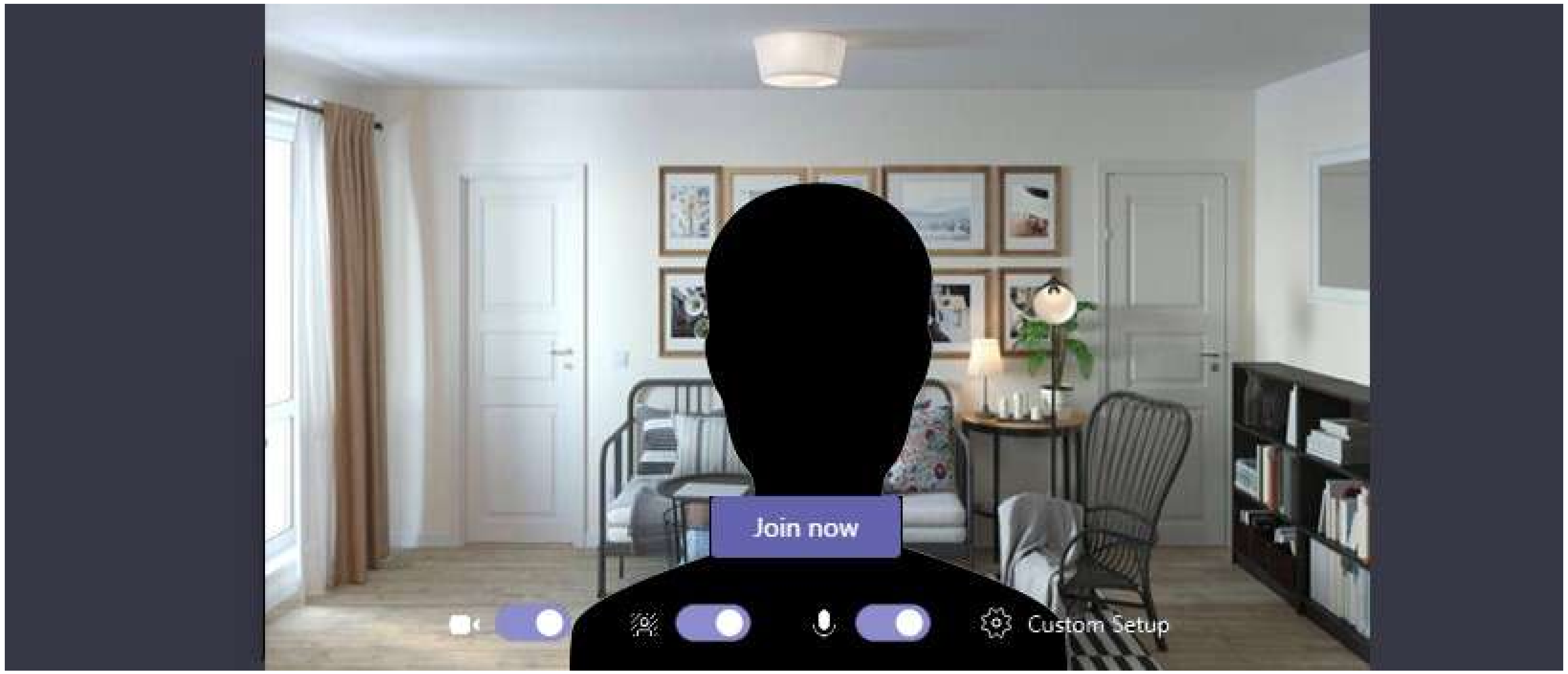
Today's Agenda

- Course Overview
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications
- What we do

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Segmentation and Matting



3D Maps

Apple Maps

[Seitz, Szeliski]

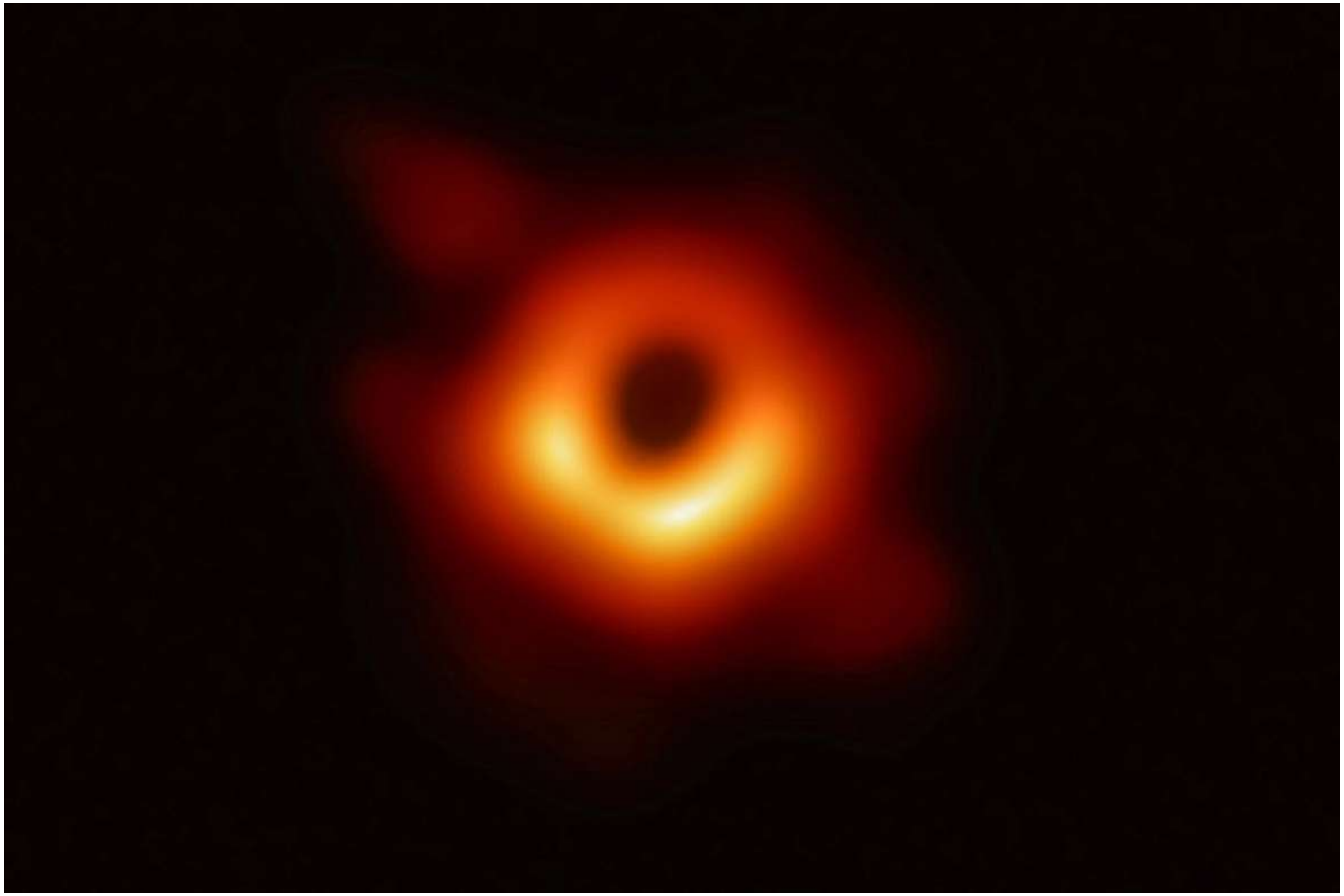
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Computational photography

Portrait mode
simulating wider aperture

[Seitz, Szeliski]

Even wider aperture...



How scientists captured the first image of a black hole, 2019

[Seitz, Szeliski]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

3D photos

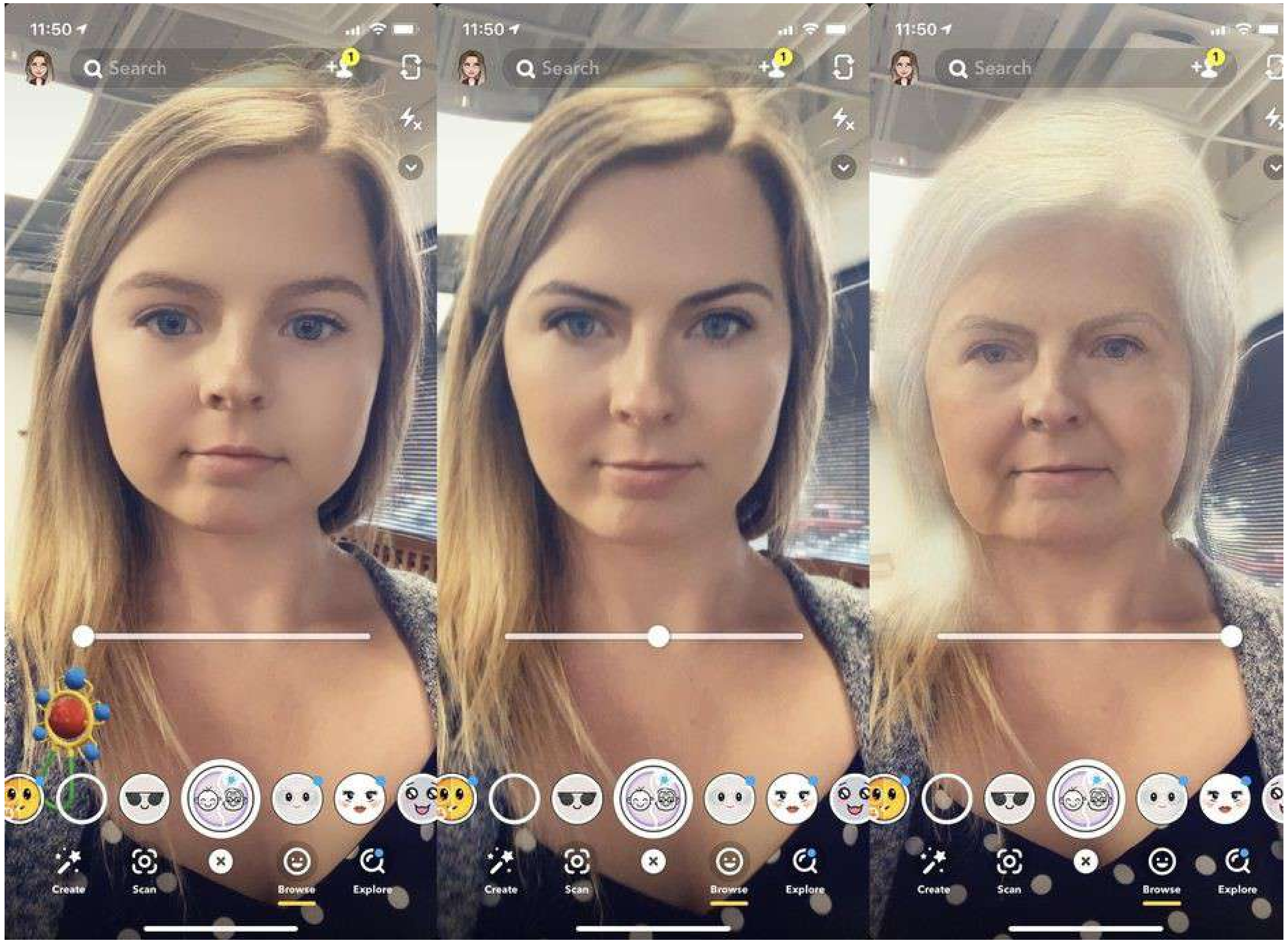
3D Photos on Facebook

Estimate depth from photo to create animation

[Seitz, Szeliski]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Age Simulation

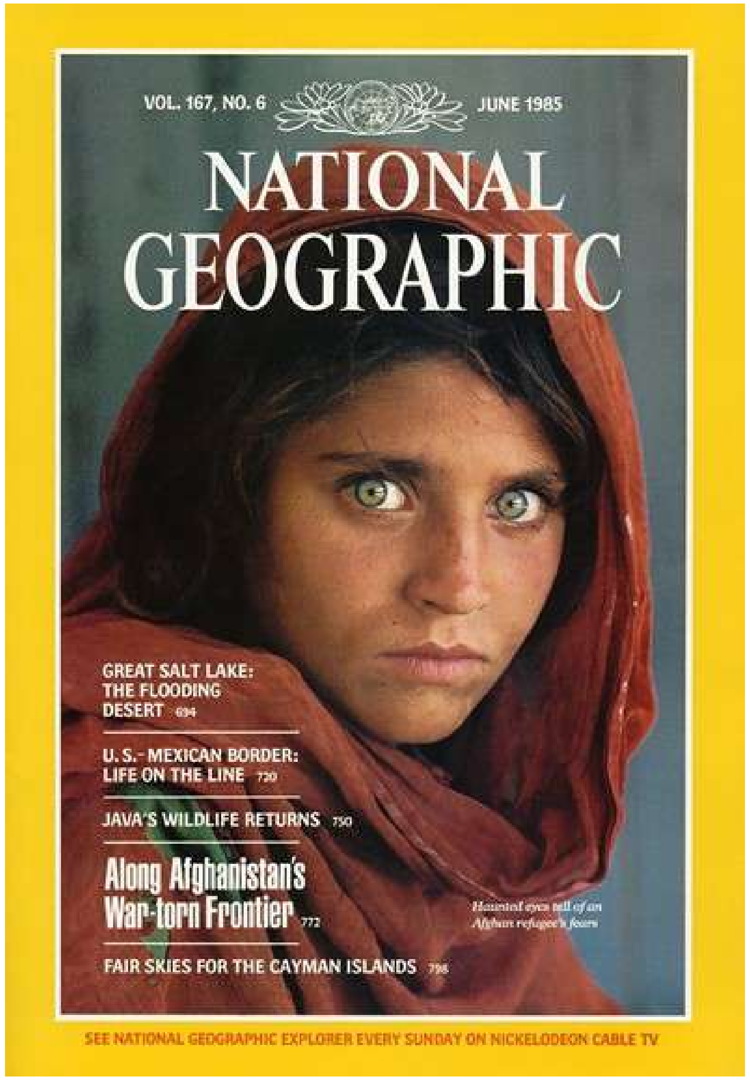


From CNET

Snapchat Time Machine

[Seitz, Szeliski]

Face recognition

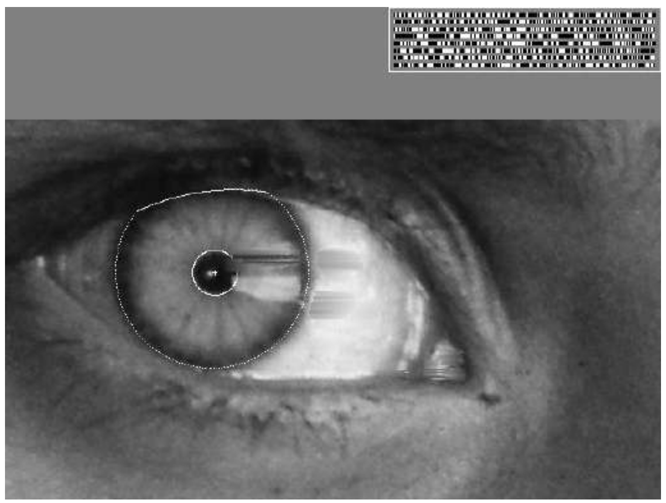


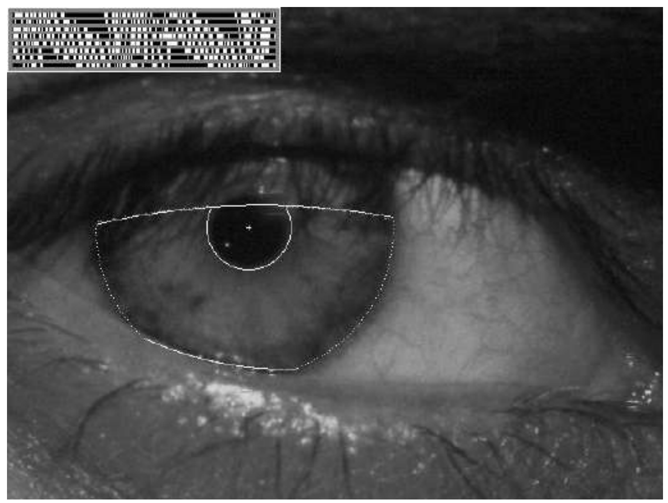
Who is she?

[Seitz, Szeliski]

Vision-based biometrics

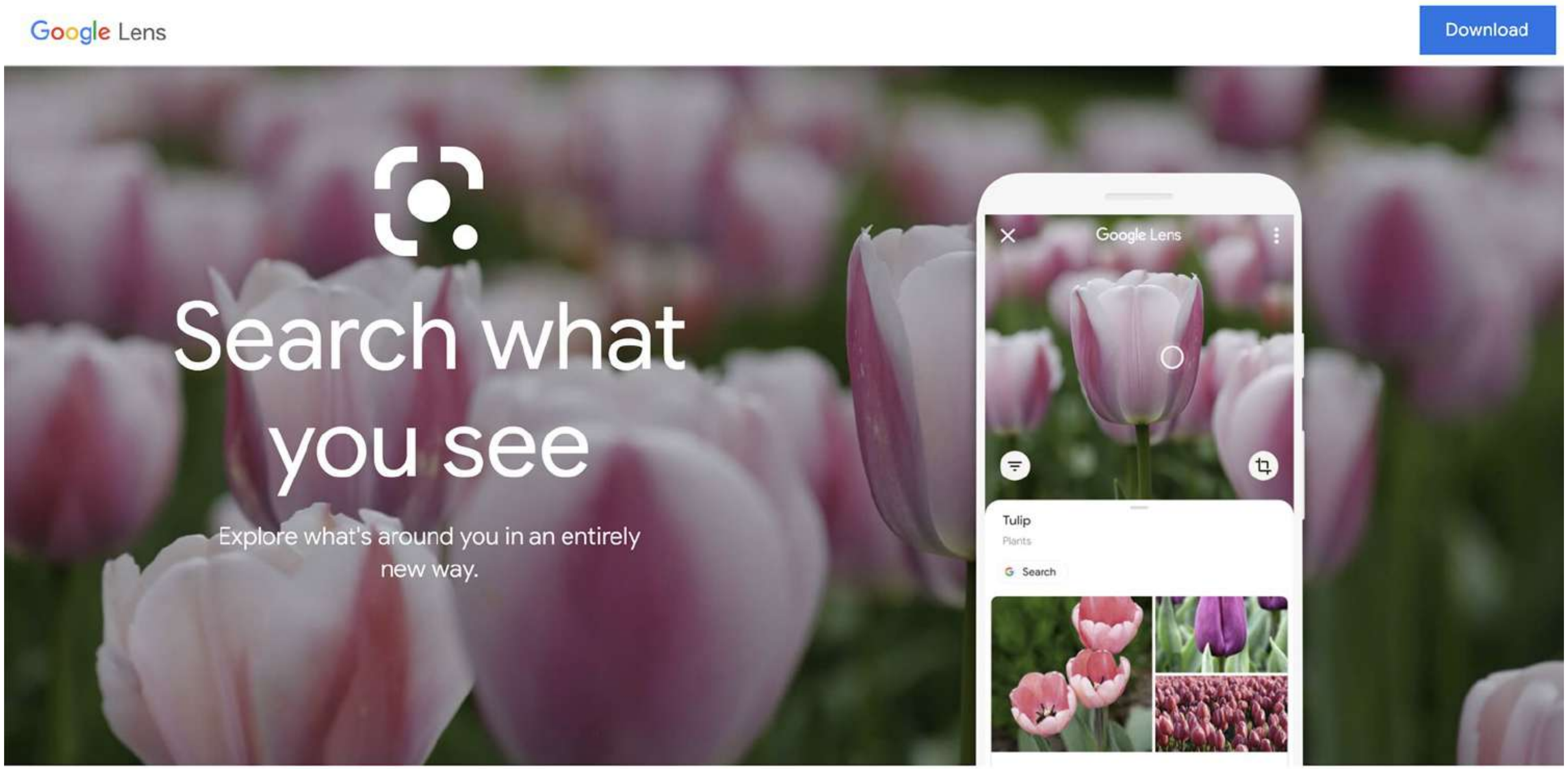
"How the Afghan Girl was Identified by Her Iris Patterns" Read the story





[Seitz, Szeliski]

Object recognition



[Seitz, Szeliski]

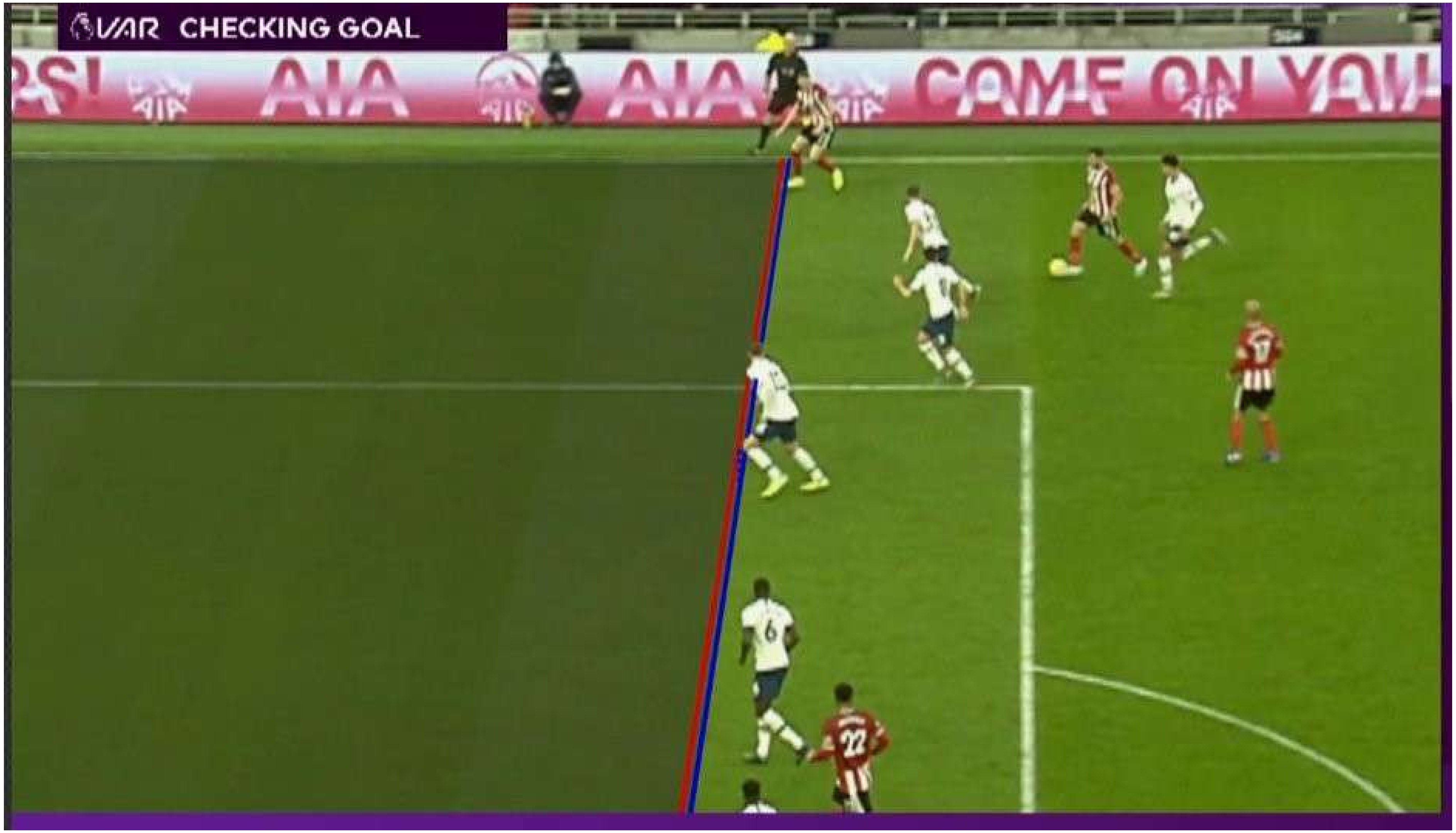
Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

[Seitz, Szeliski]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Sports - VAR



Games

Microsoft's XBox Kinect

[Seitz, Szeliski]

Virtual Reality - Metaverse

Oculus Quest, Beat Saber

[Seitz, Szeliski]

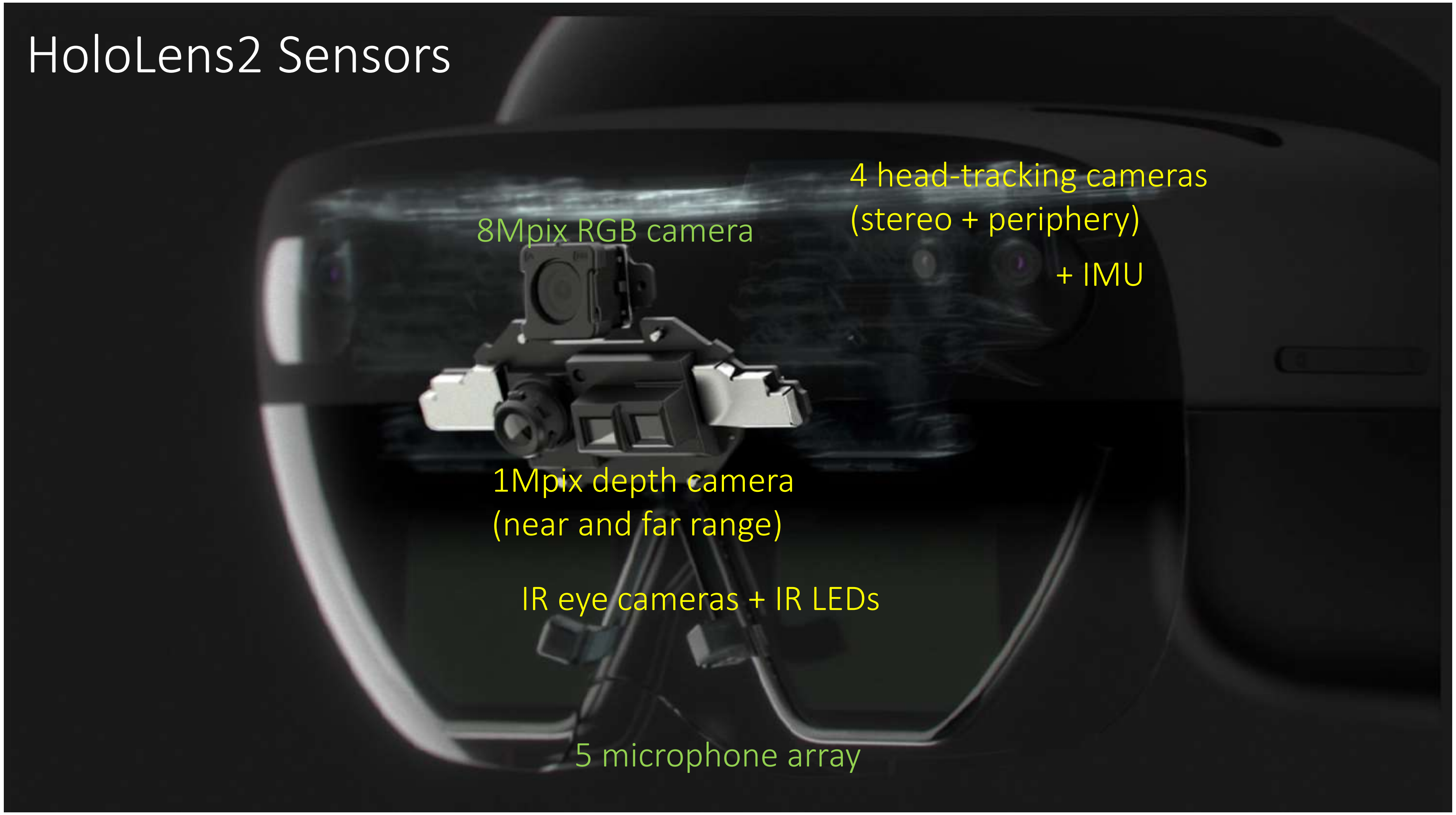
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Augmented Reality - Metaverse

Microsoft Hololens 2

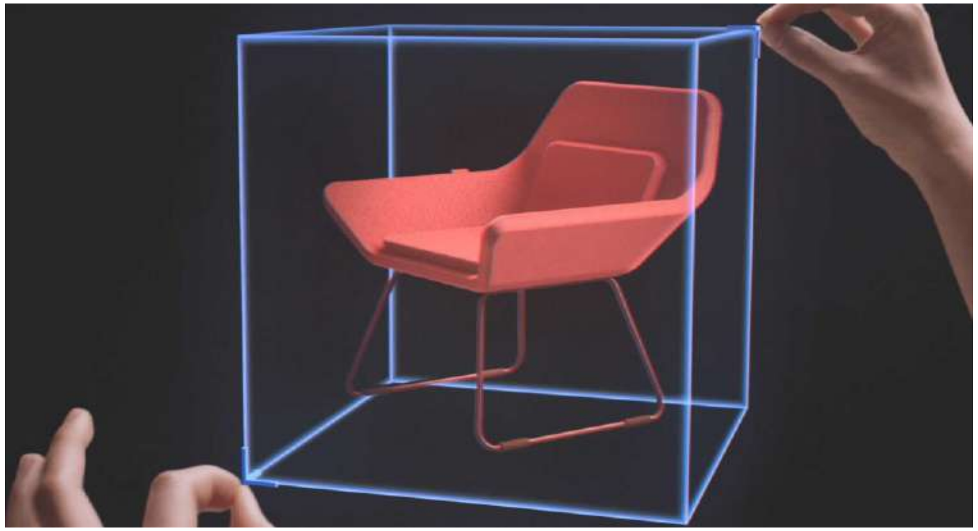
[Seitz, Szeliski]

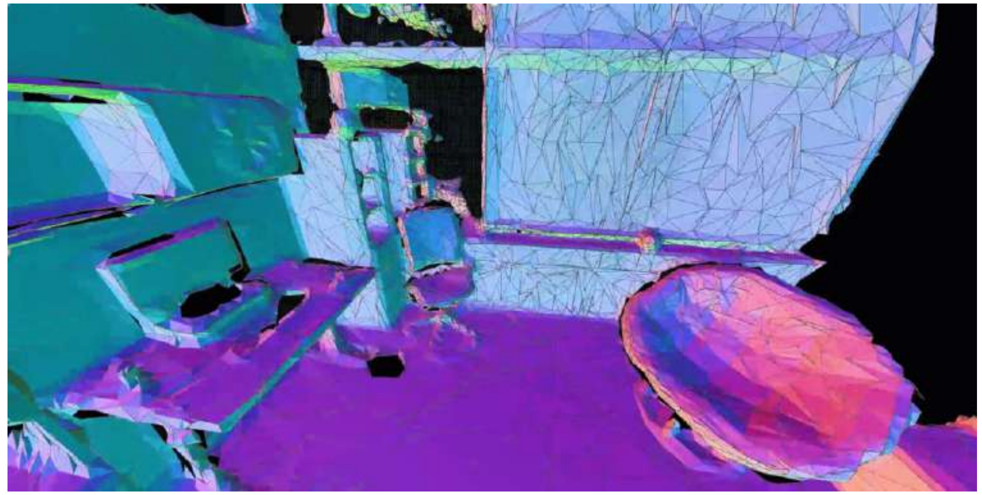
Augmented Reality - Metaverse

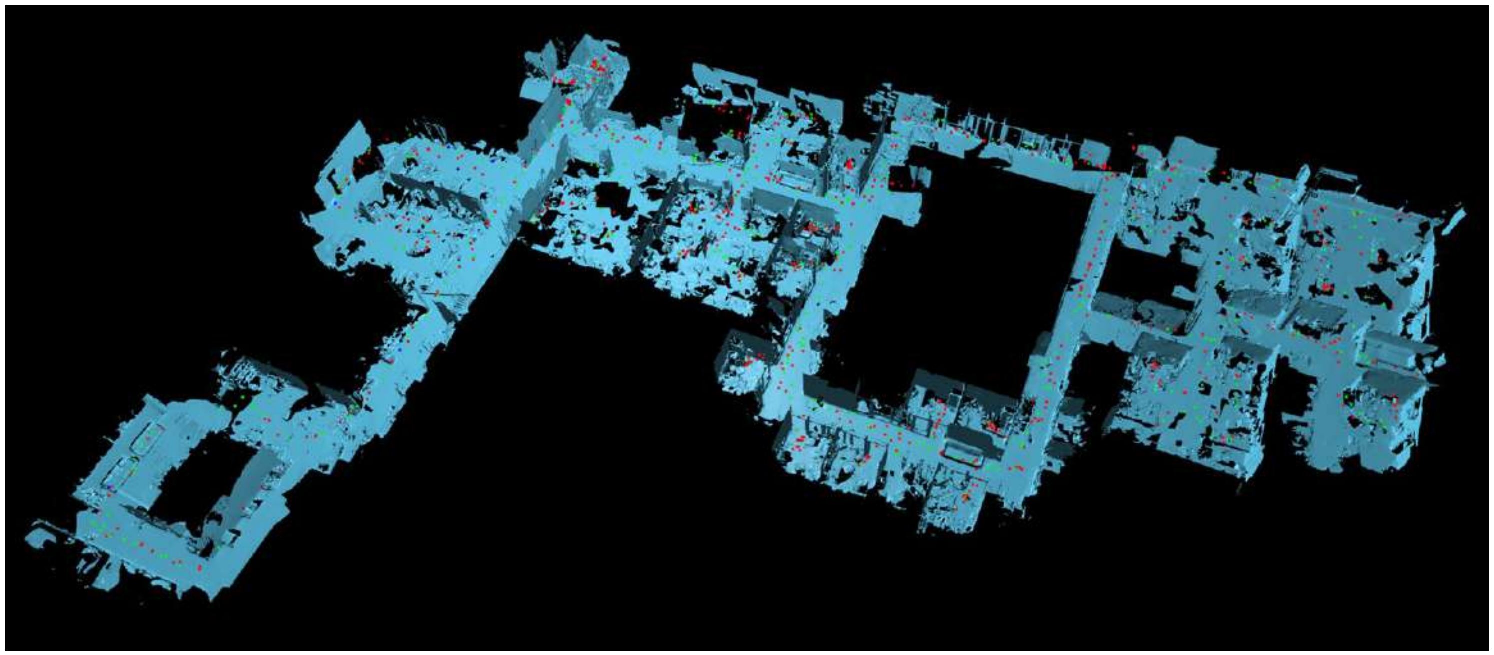


[Seitz, Szeliski]

Augmented Reality







[Seitz, Szeliski]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

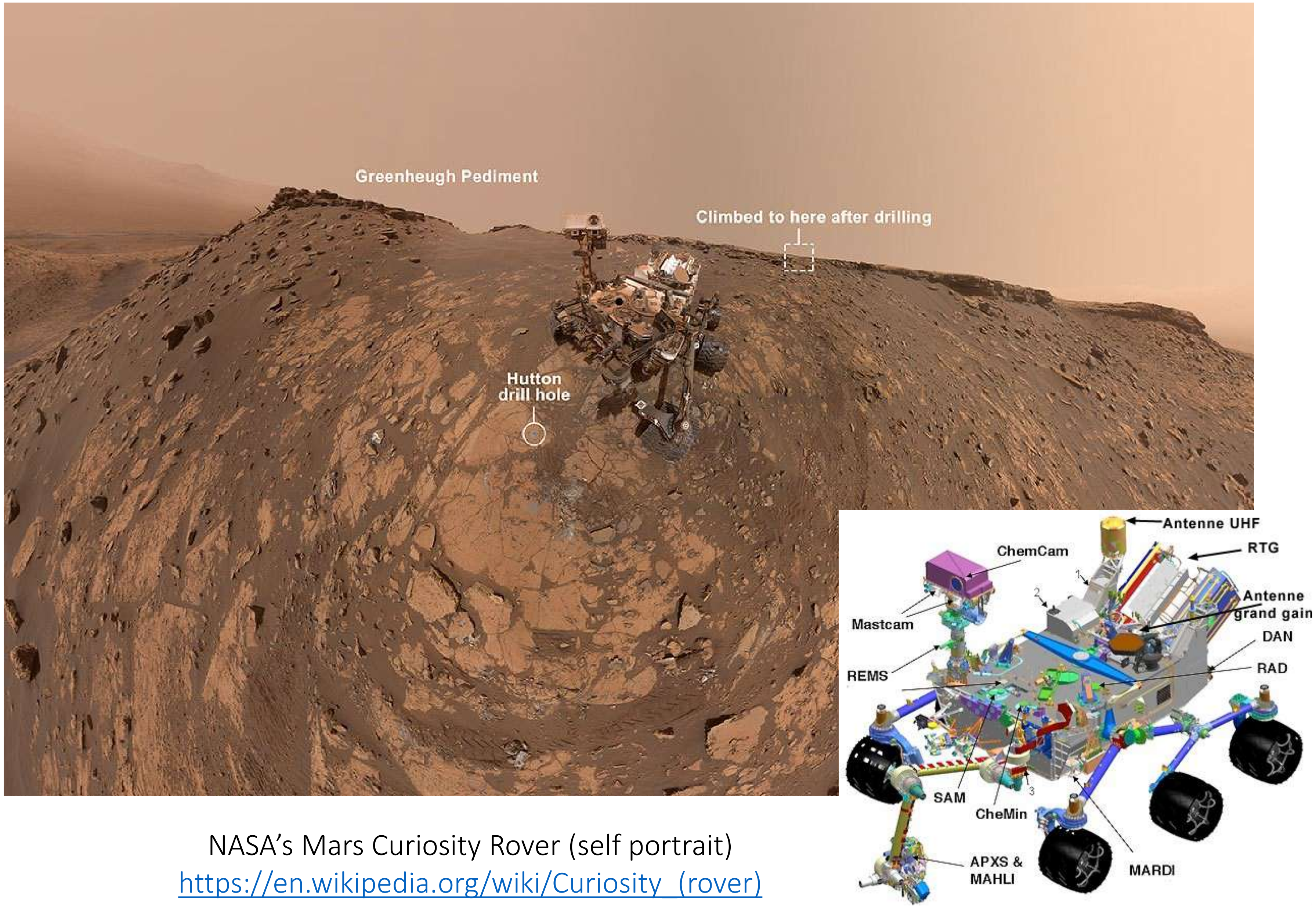
Phone-based AR

http://www.youtube.com/watch?v=0Pj-jzy6ESE

[Seitz, Szeliski]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Robotics



NASA's Mars Curiosity Rover (self portrait)
https://en.wikipedia.org/wiki/Curiosity_(rover)

[Seitz, Szeliski]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Smart cars

[Mobileye](#)

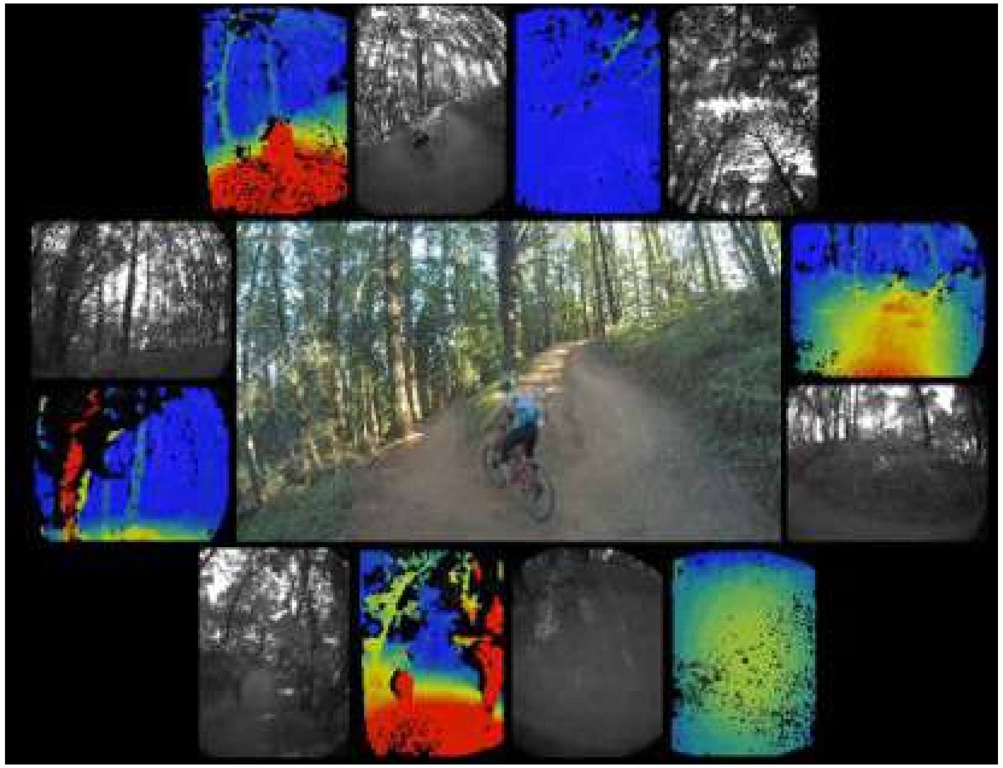
- Vision systems currently in high-end BMW, GM, Volvo models

[Seitz, Szeliski]

Self-driving cars

https://waymo.com/tech/

Drones



https://www.skydio.com/

[Seitz, Szeliski]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Research: Yolo

http://www.youtube.com/watch?v=MPU2Histivl

[Seitz, Szeliski]

Research: StyleGan

http://www.youtube.com/watch?v=BIZg_PPuj_0

[Seitz, Szeliski]

Today's Agenda

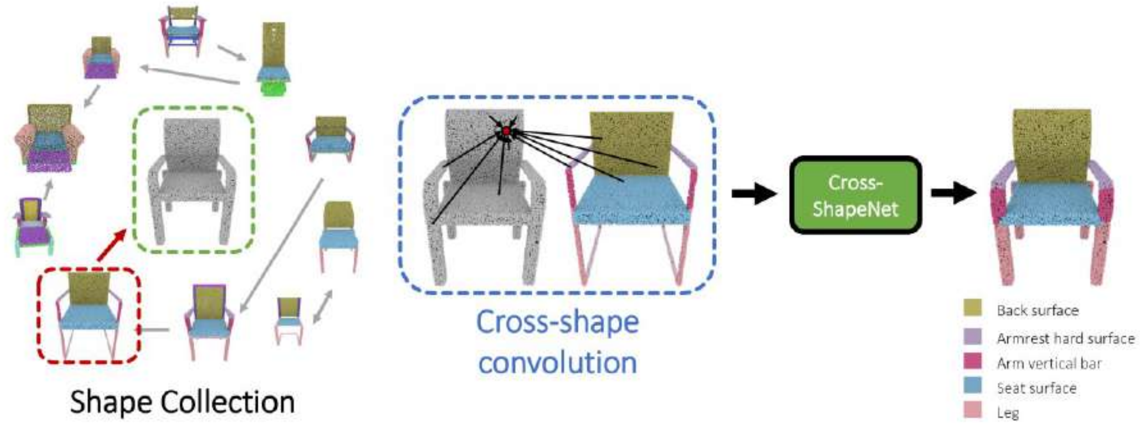
- Who are we?
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications
- What we do

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

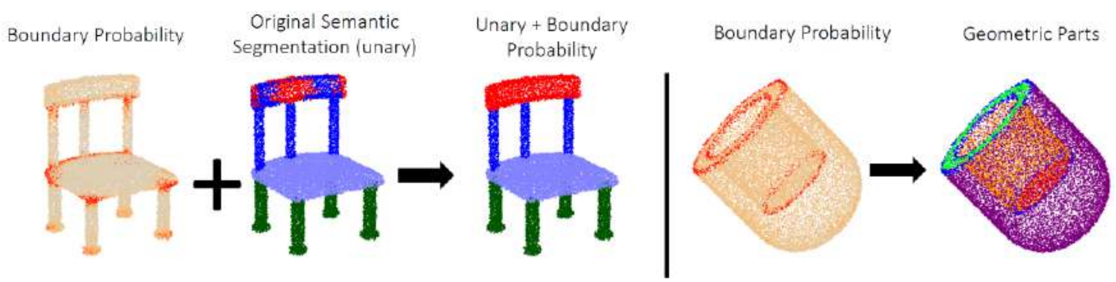
3D Shape Understanding

3D Building Semantic Understanding

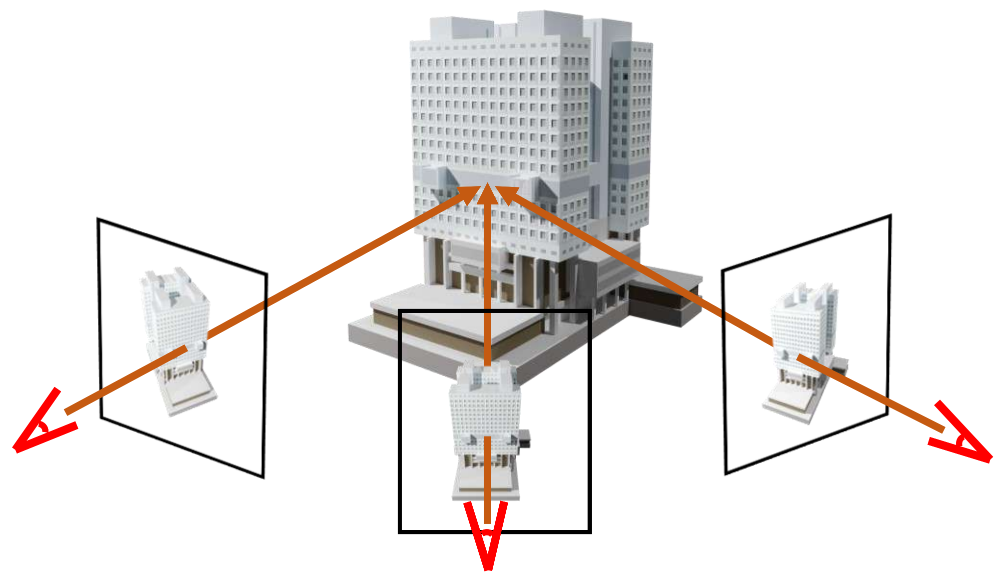
Cross-shape semantic segmentation



Geometric/Semantic Decomposition



Neural 3D Reconstruction

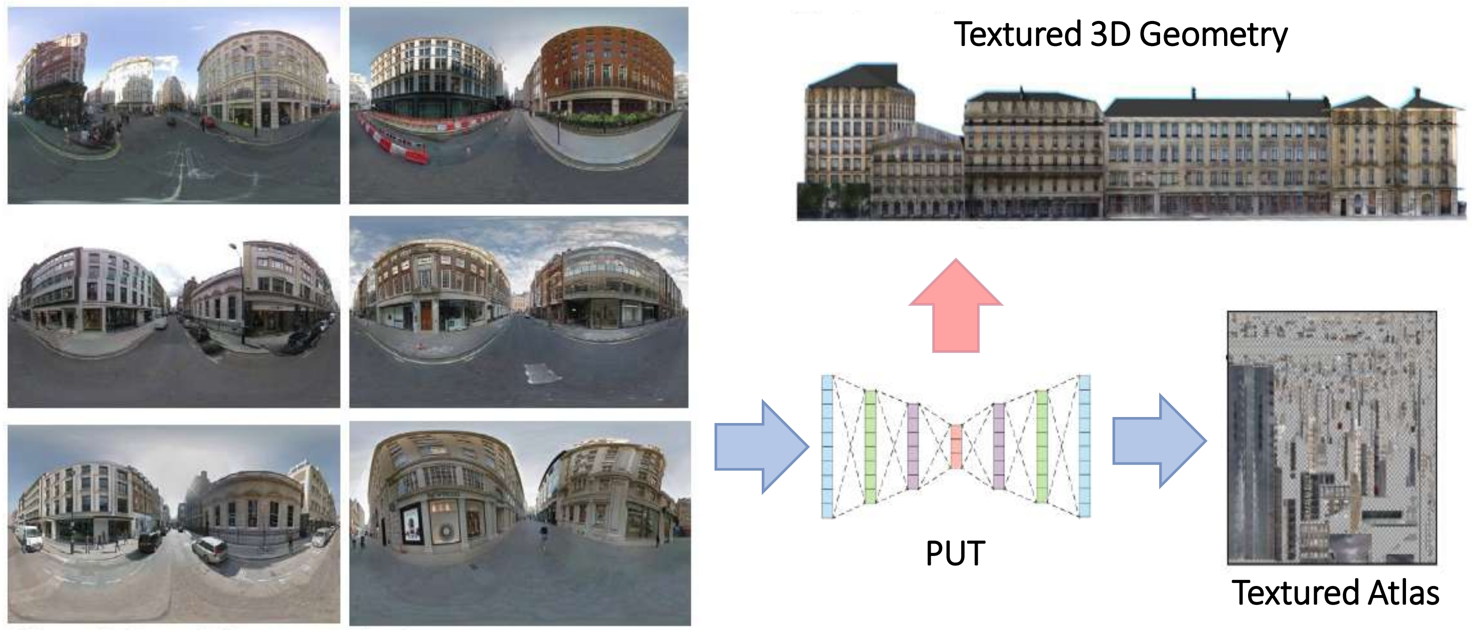


Texture Generation for 3D Data

Single-View Guided Façade Synthesis

Interactive Texturing

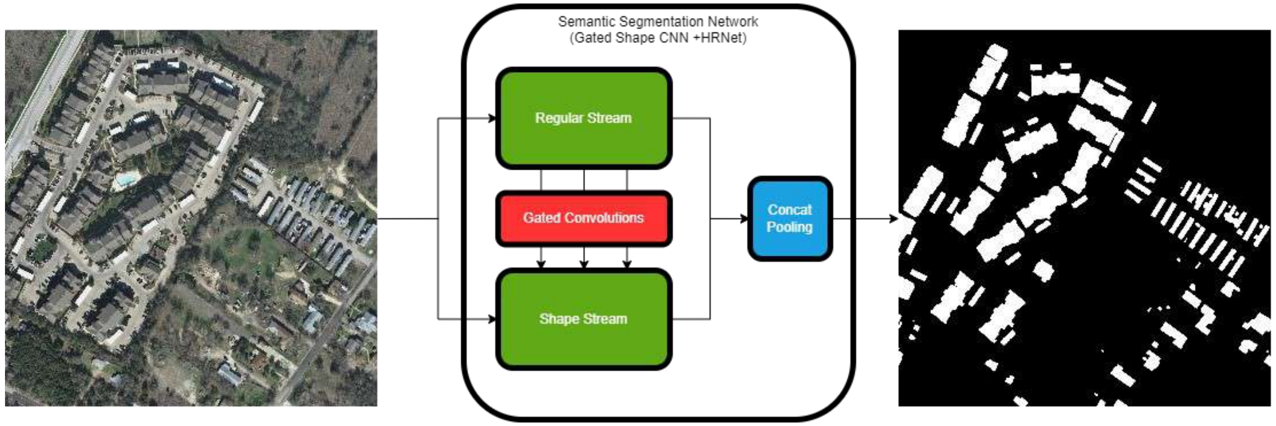
Projective Urban Texturing



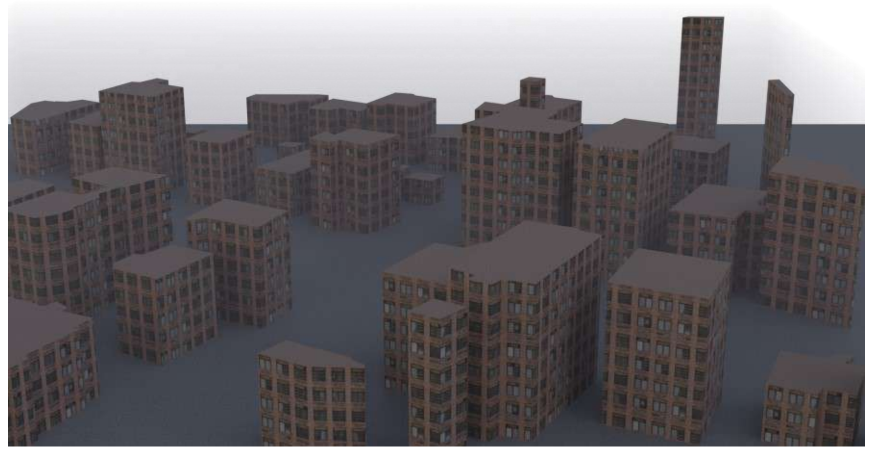
Street Level Panoramic Images

Urban Semantic Understanding from Remote Sensing Data Sources

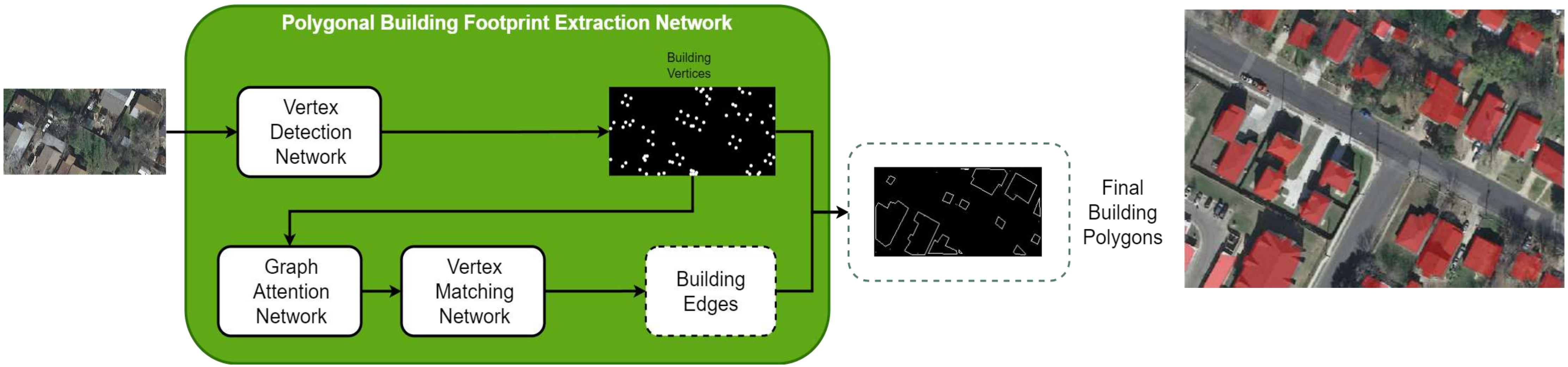
Semantic Segmentation of Buildings



Urban 3D Reconstruction



Building Footprint Extraction



Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 2: Fundamentals - Color

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Course Overview
- Introduction to Computer Vision
  - What is Computer Vision
  - How hard is Computer Vision
  - Why is Computer Vision so hard
  - How to organize Computer Vision
  - Why study Computer Vision
  - Applications

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda - Overview of Color

- Physics of color
- Human encoding of color
- Color spaces

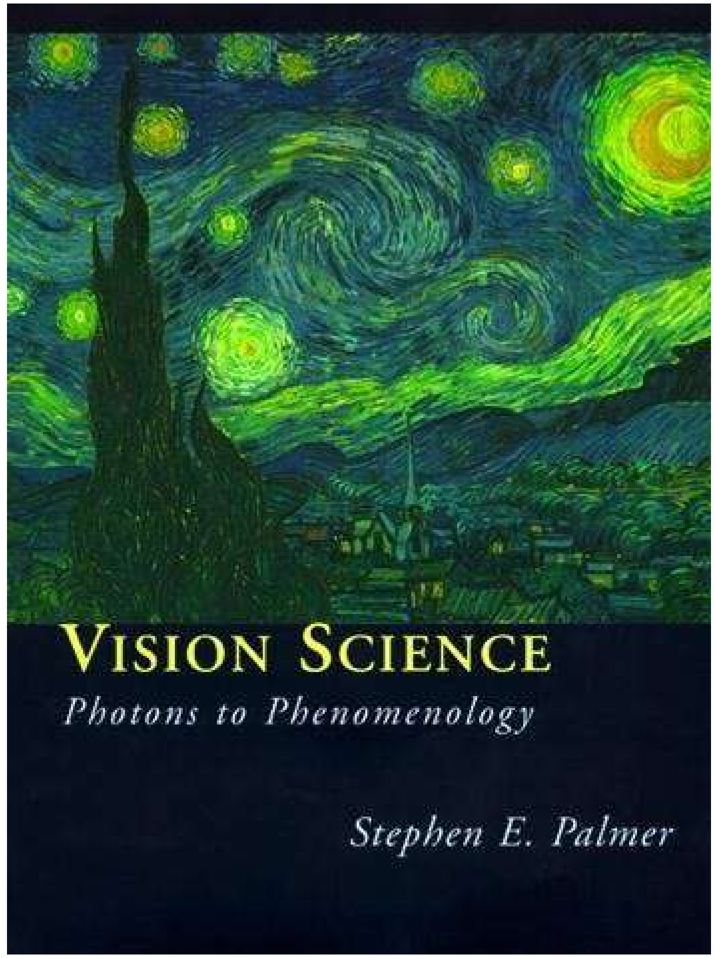
[Slides based on material by Niebles 2017]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What is color?

- The result of interaction between physical **light** in the environment and our **visual system**.
- A **psychological property** of our visual experiences when we look at objects and lights, **not a physical property** of those objects or lights.

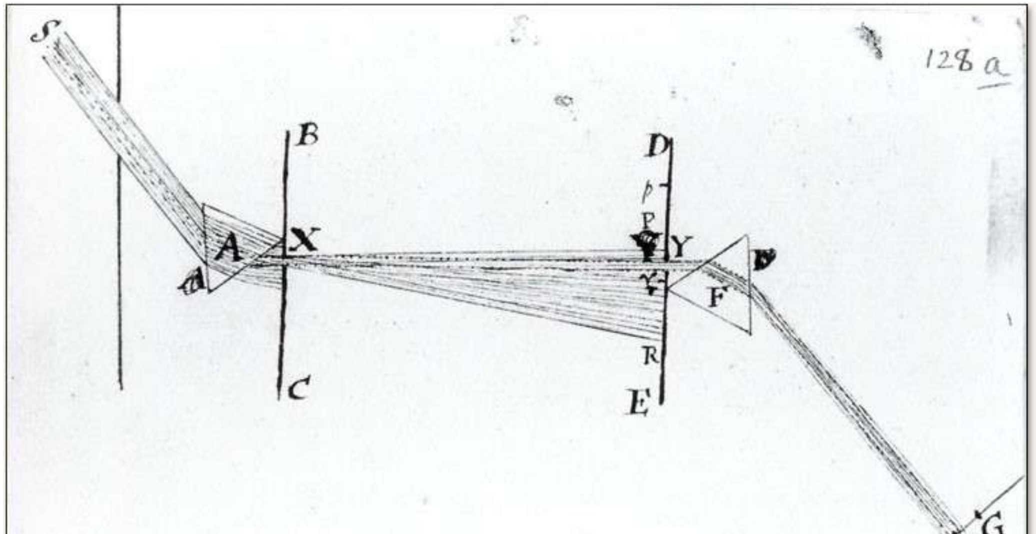


[Slide credit: Lana Lazebnik]

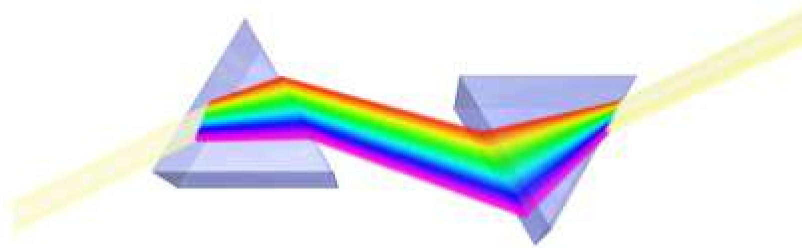
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Color and light

White light: composed of almost equal energy in all wavelengths of the **visible spectrum**

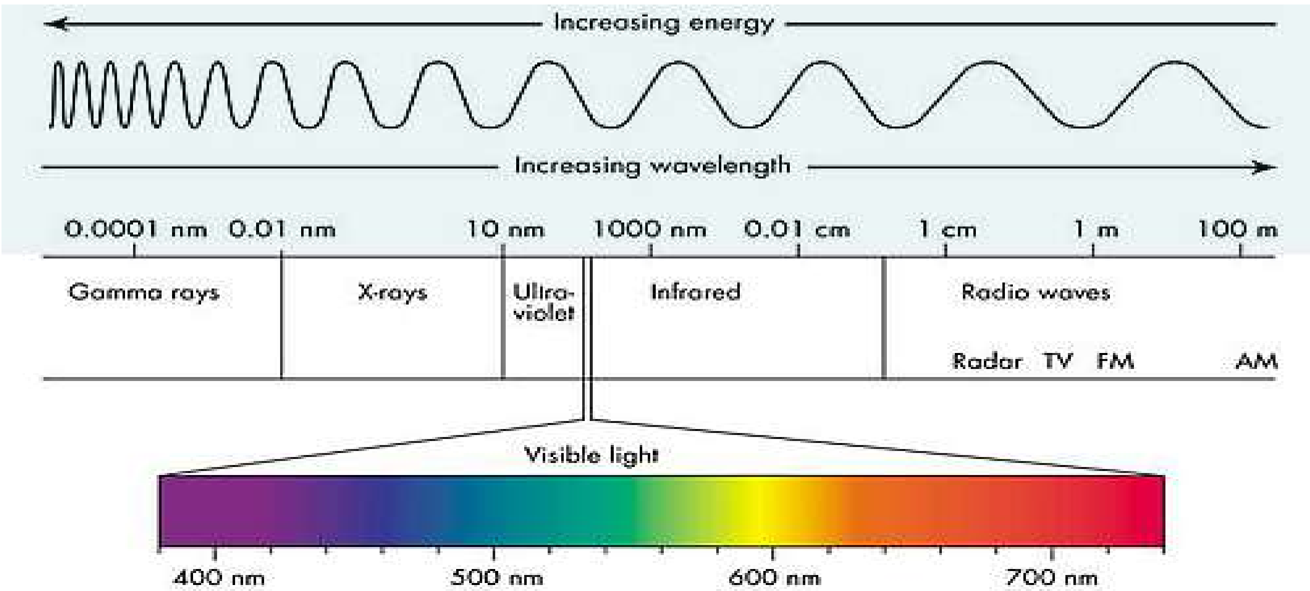


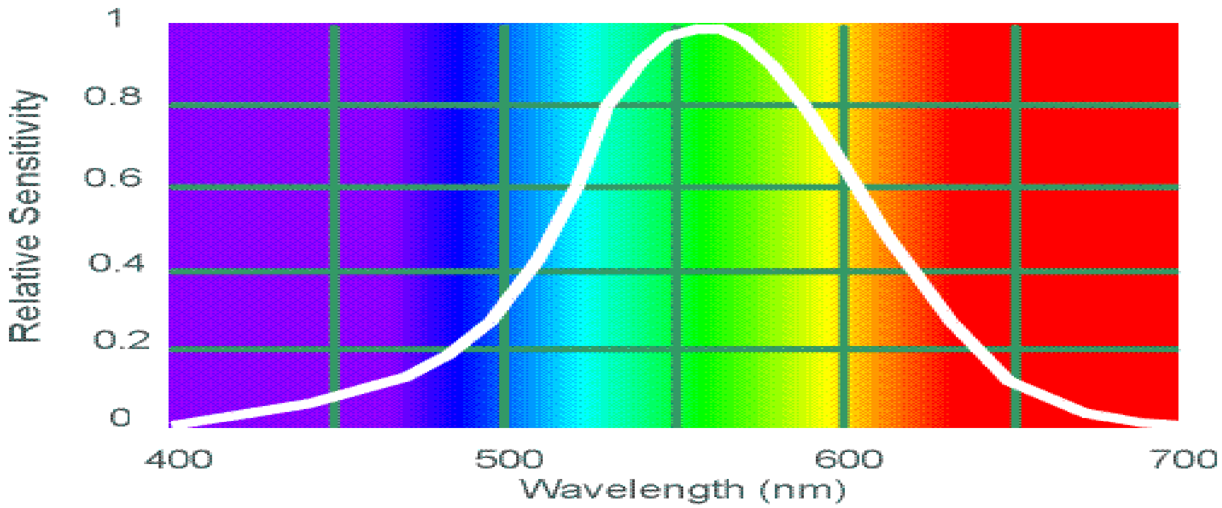
Newton 1665



Try it:
https://micro.magnet.fsu.edu/primer/java/scienceopticsu/newton/

Electromagnetic Spectrum





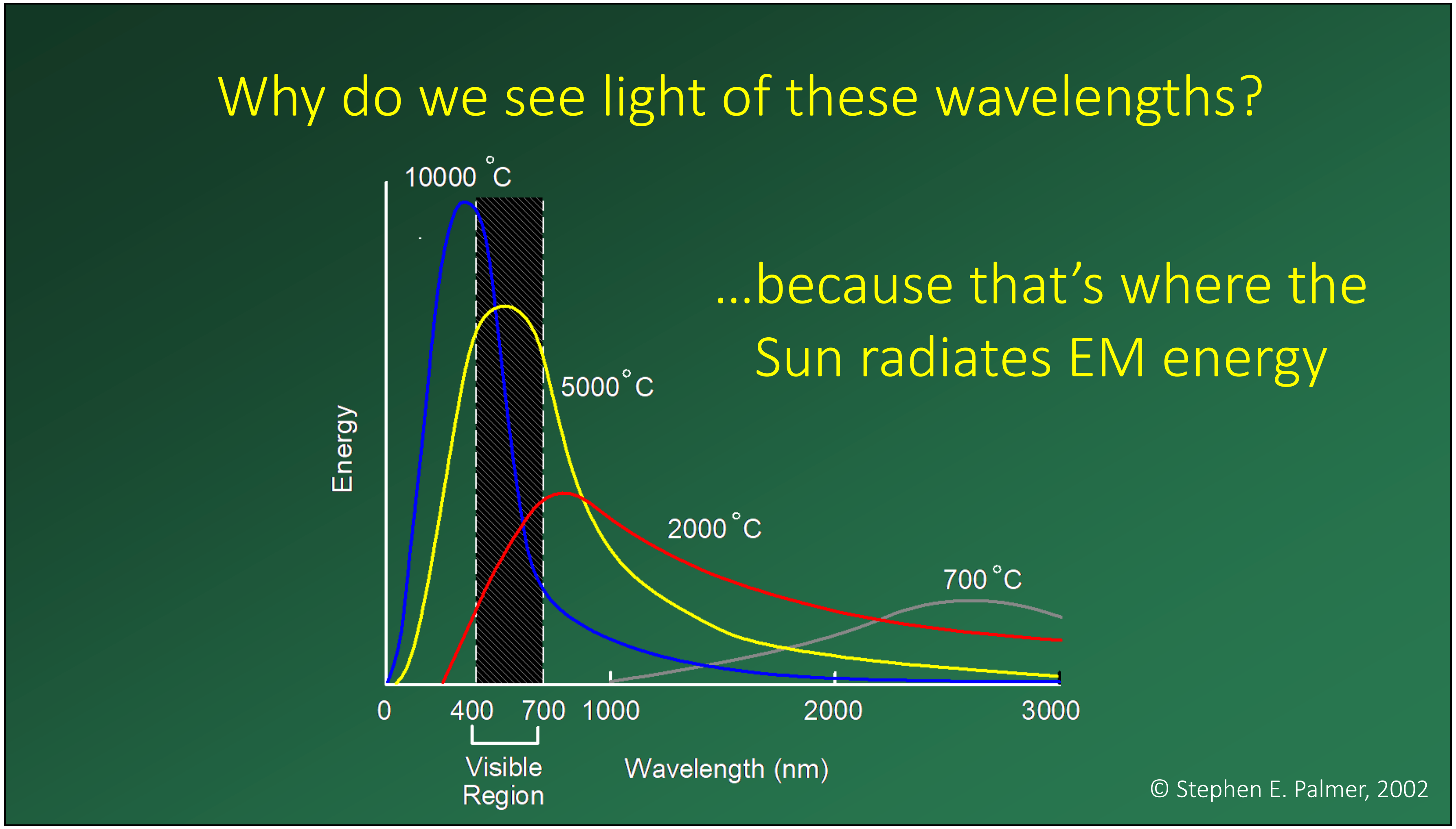
Human Luminance Sensitivity Function

http://www.yorku.ca/eye/photopik.htm

Co-financed by the European Union
Connecting Europe Facility

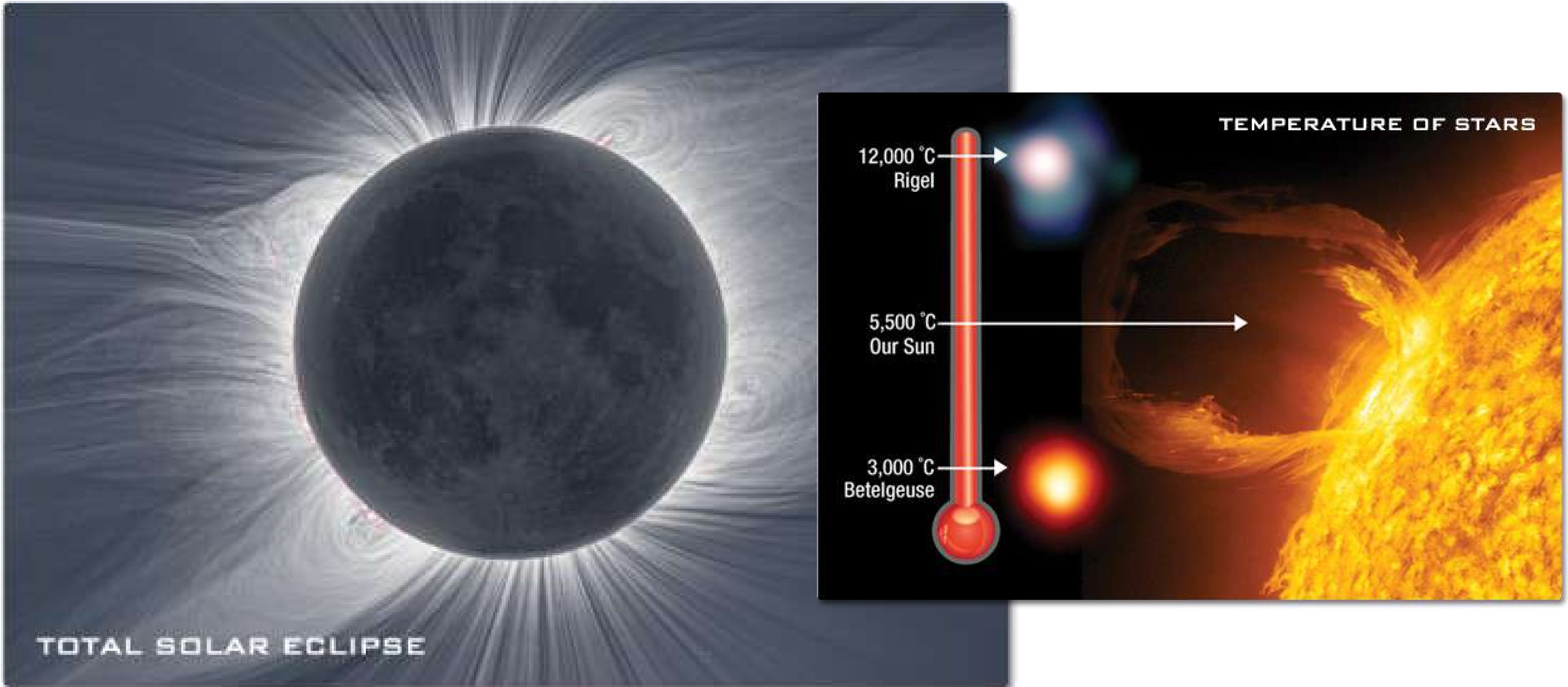
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Visible Light



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Sun’s temperature makes it emit yellow light more than any other color

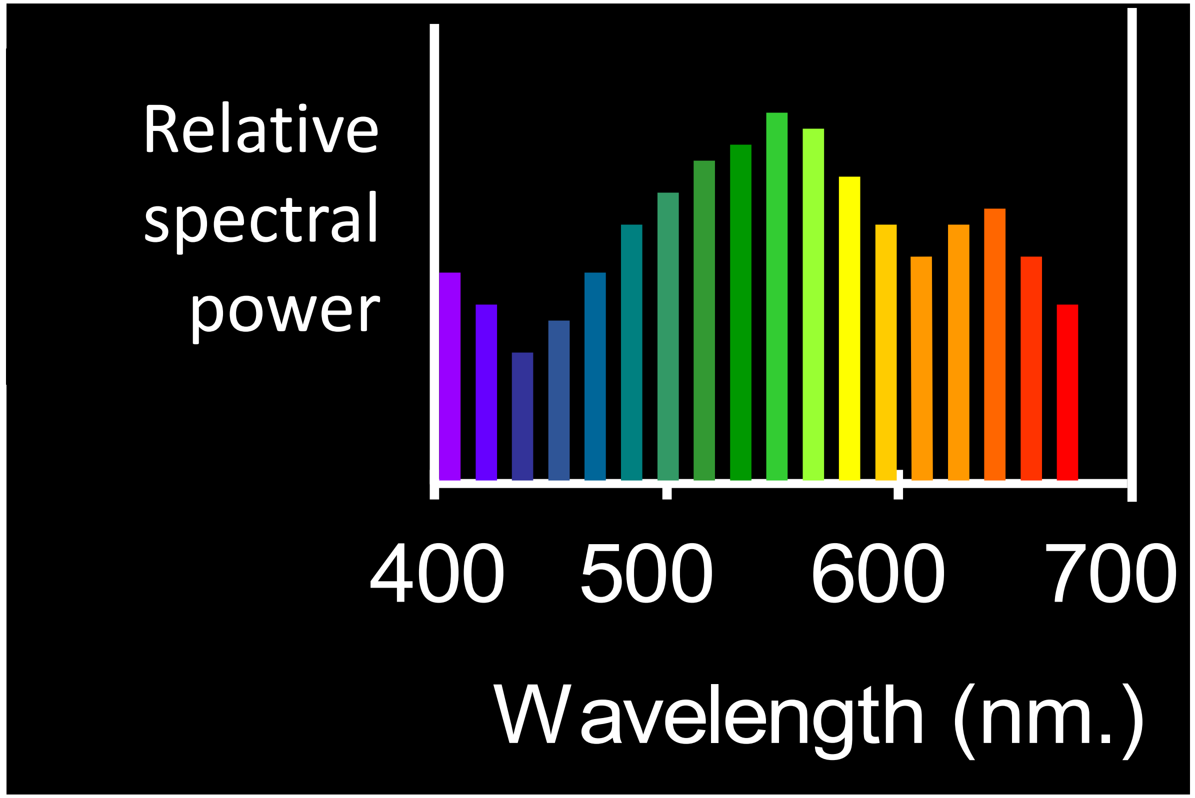


https://science.nasa.gov/ems/09_visiblelight

The Physics of Light

Any source of visible light can be completely described physically by its **spectrum**:

the amount of energy emitted (per time unit) at each wavelength 400 - 700 nm.

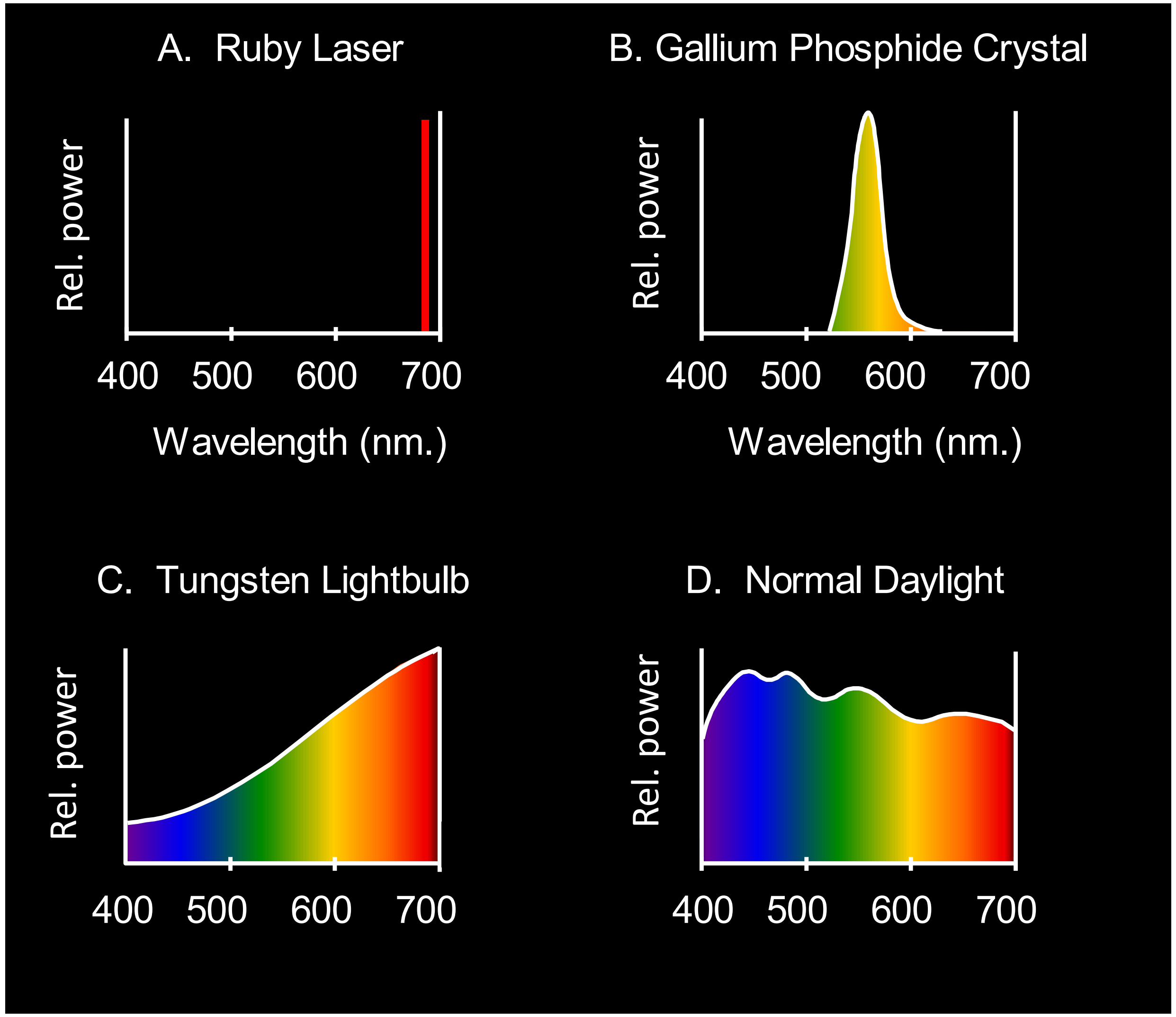


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The Physics of Light

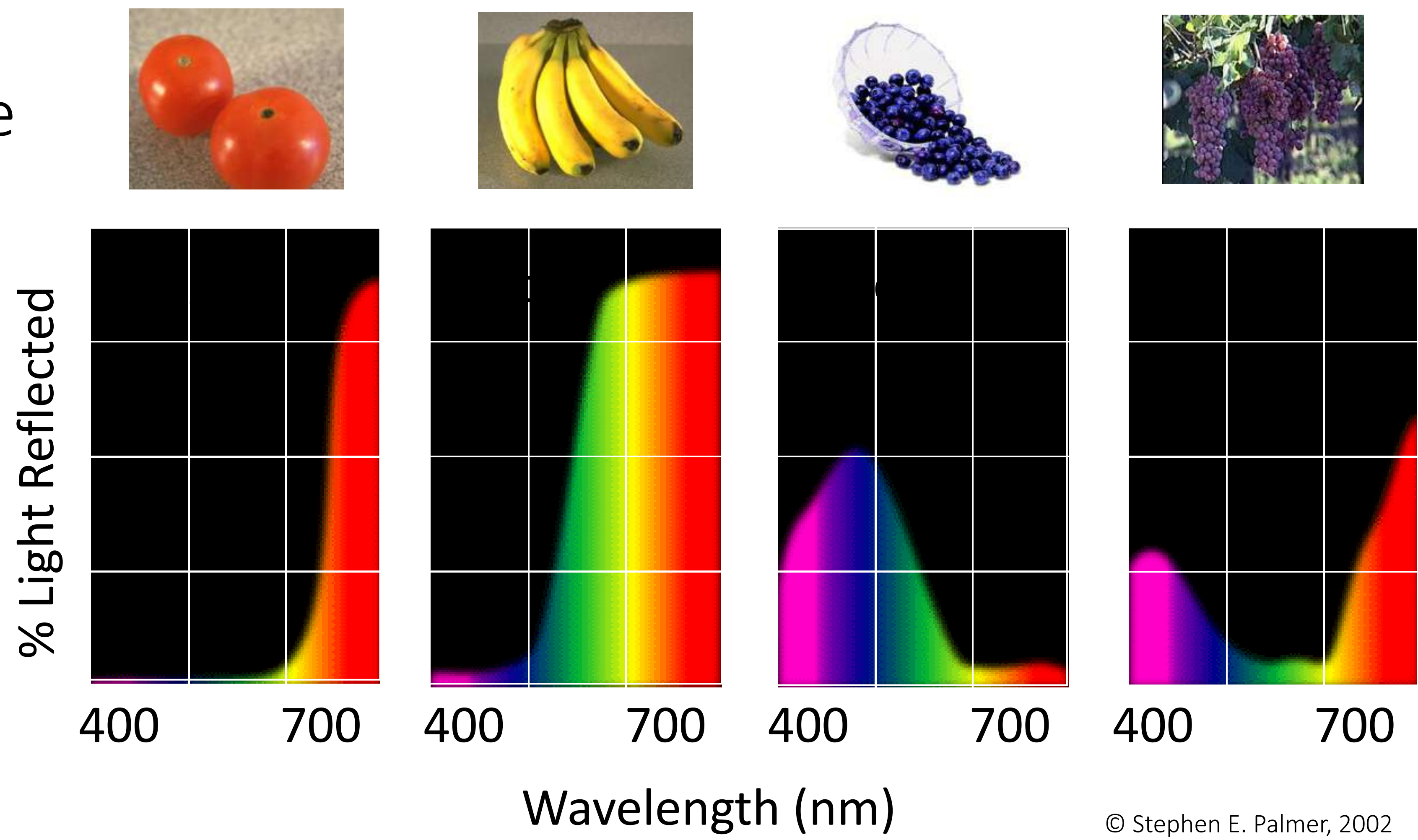
Some examples of the **spectra** of light sources



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The Physics of Light

Some examples of the **reflectance** spectra of surfaces

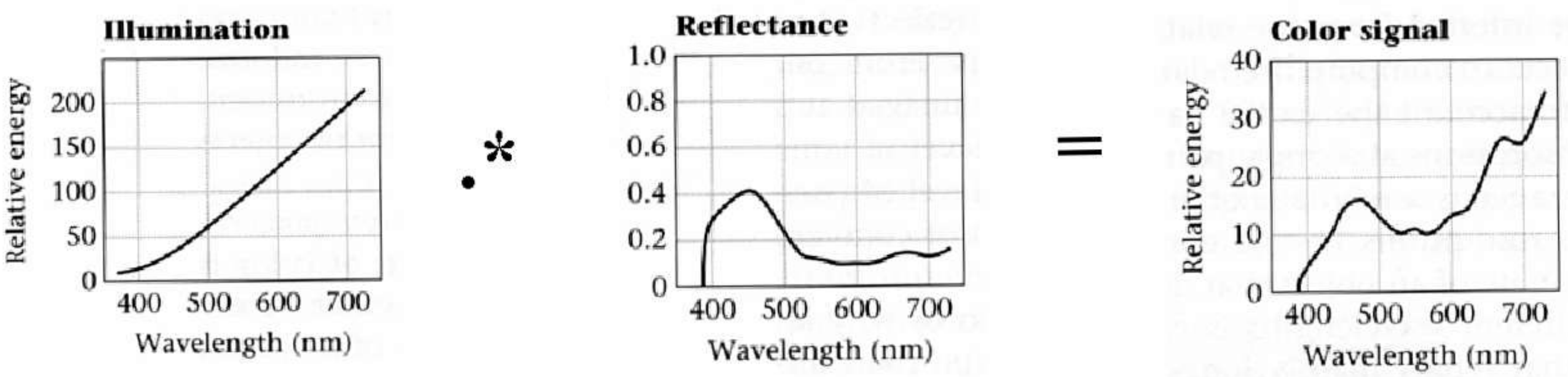


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Interaction of light and surfaces

Reflected color is the result of interaction of **light** source spectrum with **surface** reflectance



From Foundation of Vision by Brian Wandell, Sinauer Associates, 1995

Interaction of light and surfaces

What is the observed color of any surface under monochromatic light?

Olafur Eliasson, *Room for one color* [Slide by S. Lazebnik]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

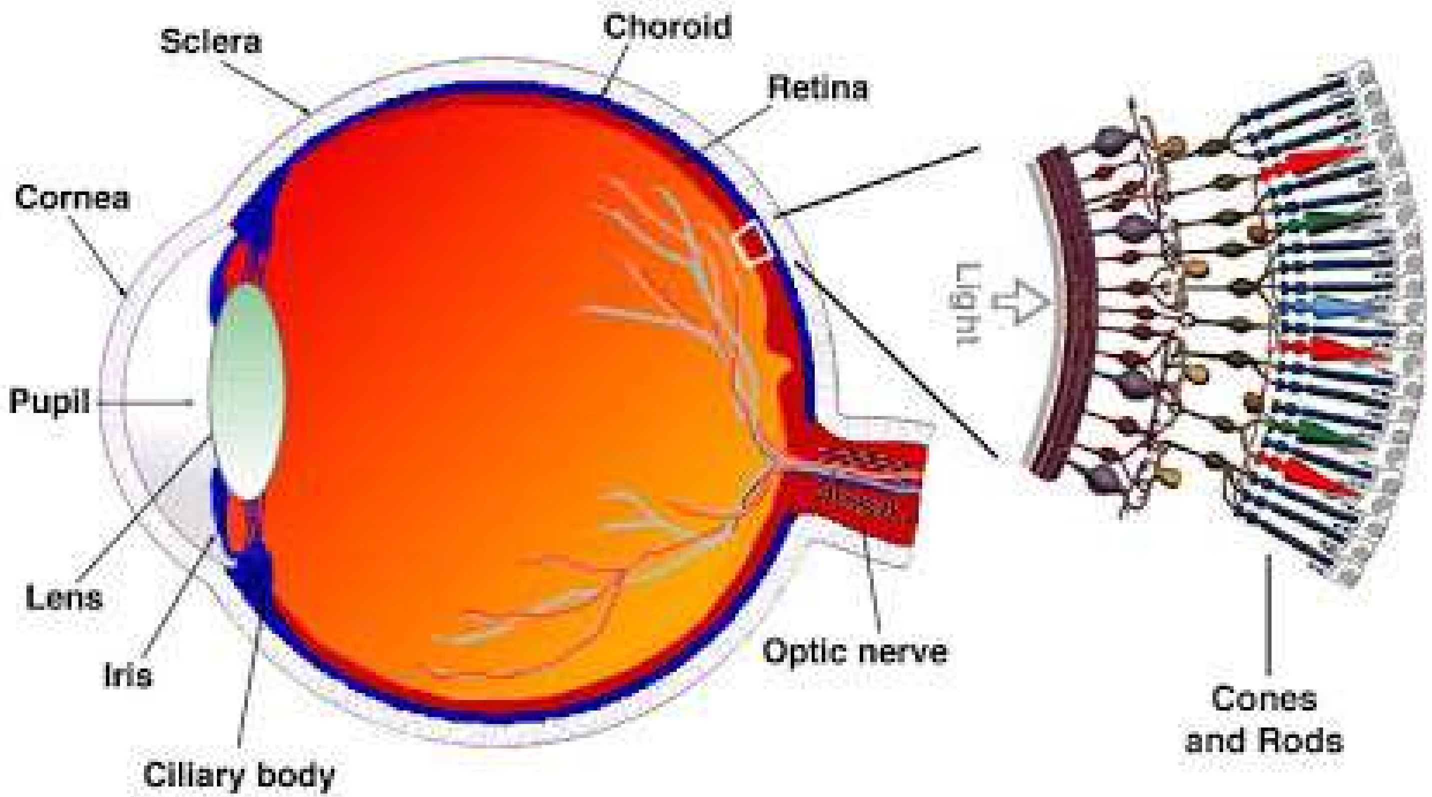
Interaction of light and surfaces

See here

Overview of Color

- Physics of color
- Human encoding of color
- Color spaces

Two types of light-sensitive receptors

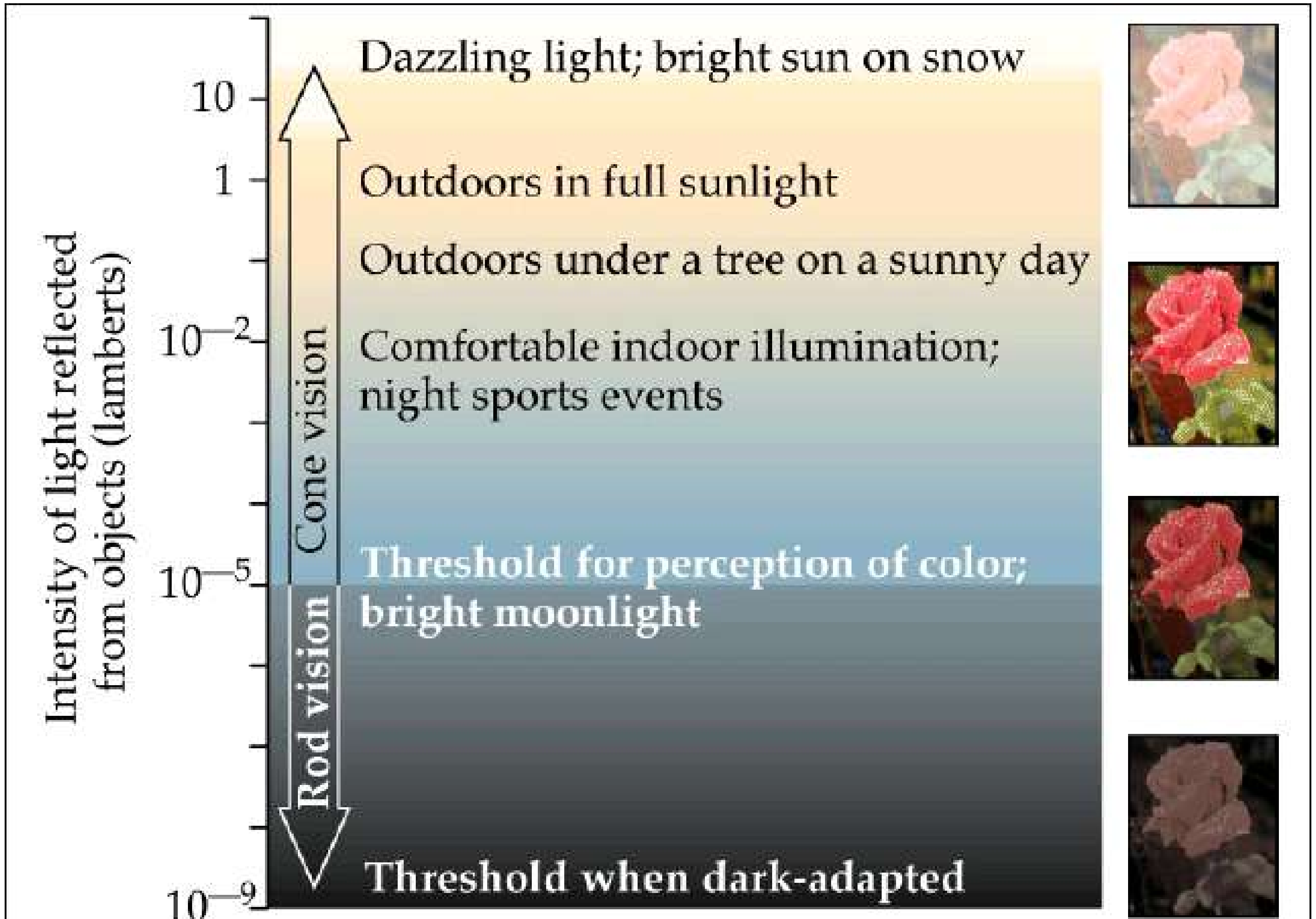


Cones
- cone-shaped
- less sensitive
- operate in high light
- color vision

Rods
- rod-shaped
- highly sensitive
- operate at night
- gray-scale vision

http://www.blueconemonochromacy.org/how-the-eye-functions/

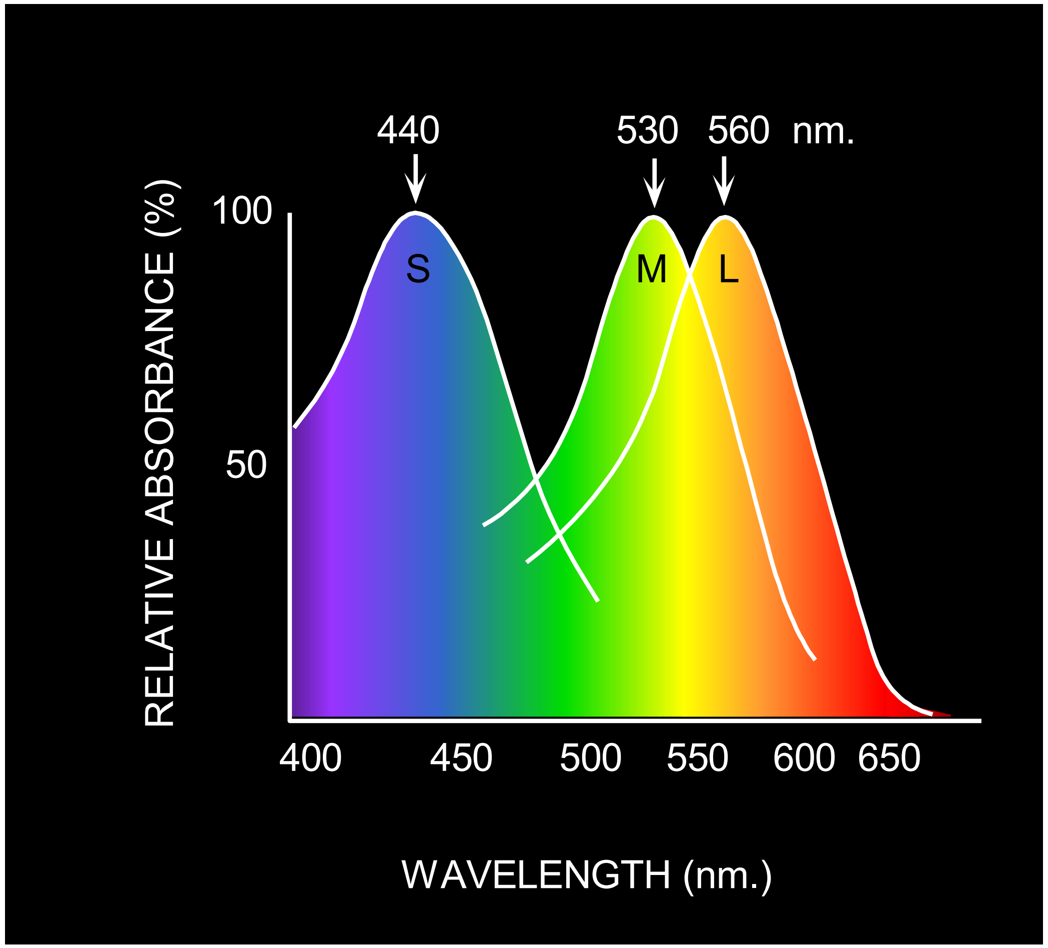
Rod / Cone sensitivity

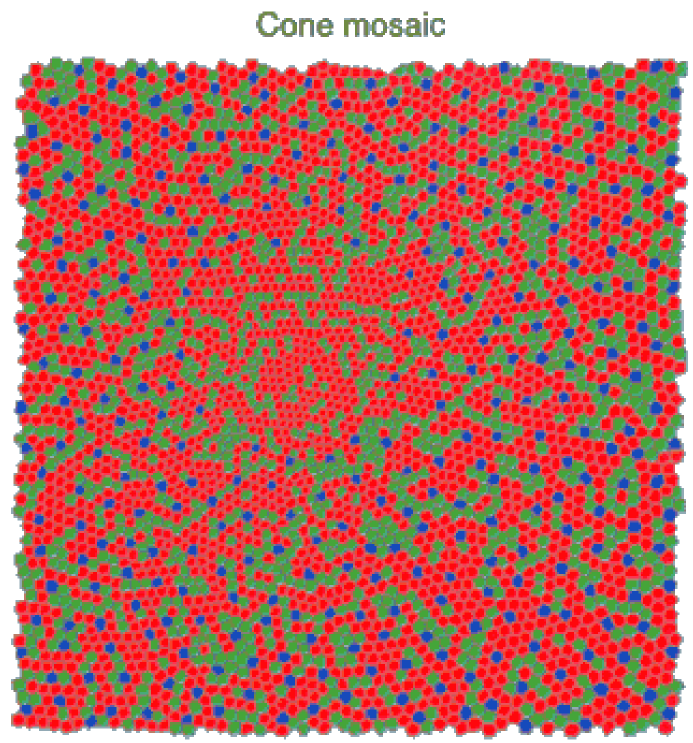


The famous sock-matching problem…

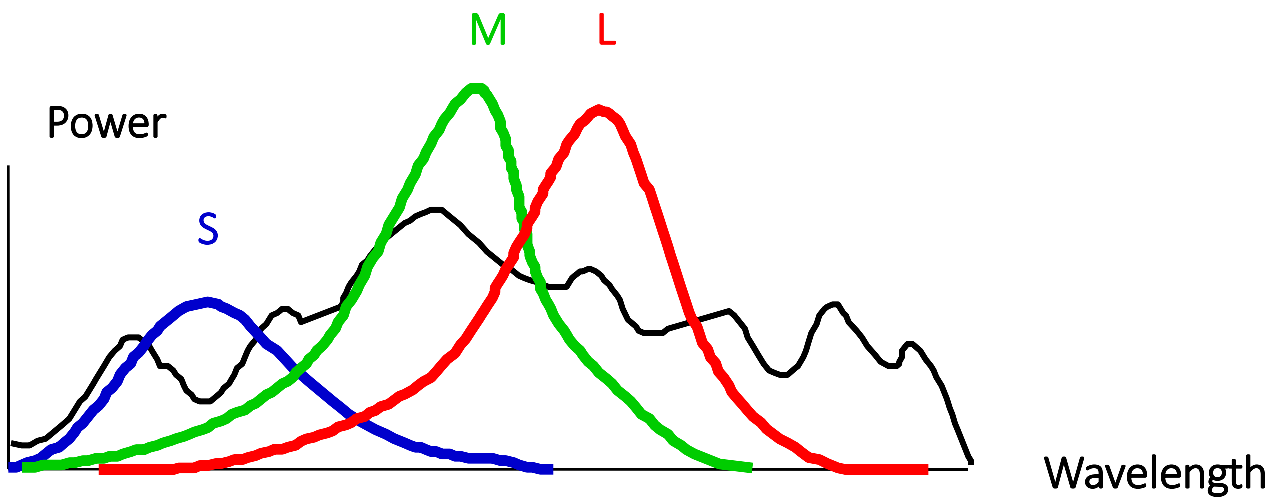
Physiology of Color Vision

Three kinds of cones: Short (S), Medium (M), Long (L)





Color perception



Rods and cones act as filters on the spectrum

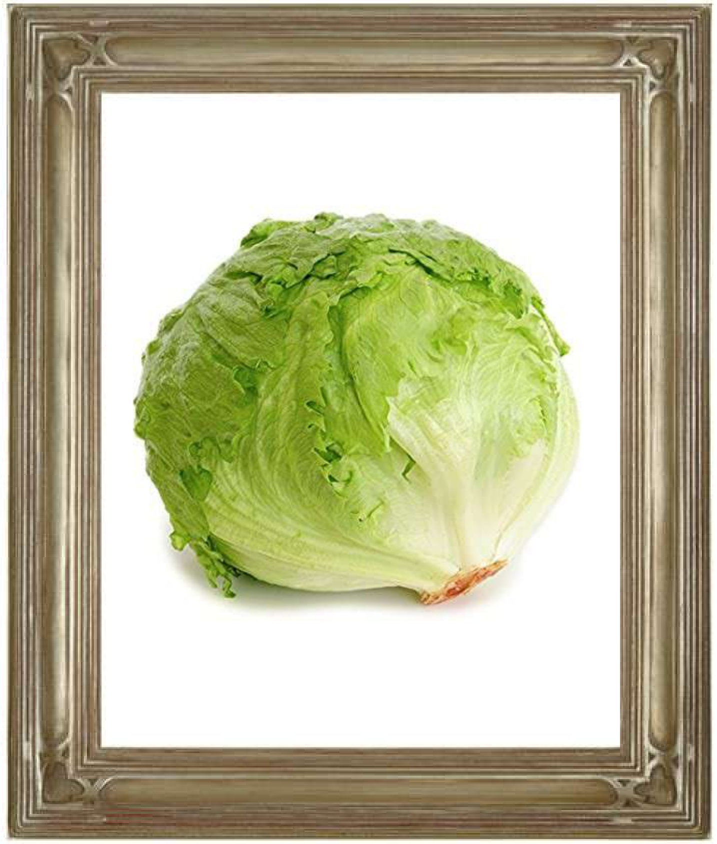
- To get the output of a filter, multiply its response curve by the spectrum, integrate over all wavelengths
  - Each cone yields **one** number

- Q: How can we represent an entire spectrum with 3 numbers?
- A: We can't! Most of the information is lost.
  - As a result, two different spectra may appear indistinguishable - such spectra are known as **metamers** (see demo)

[Slide by Steve Seitz]

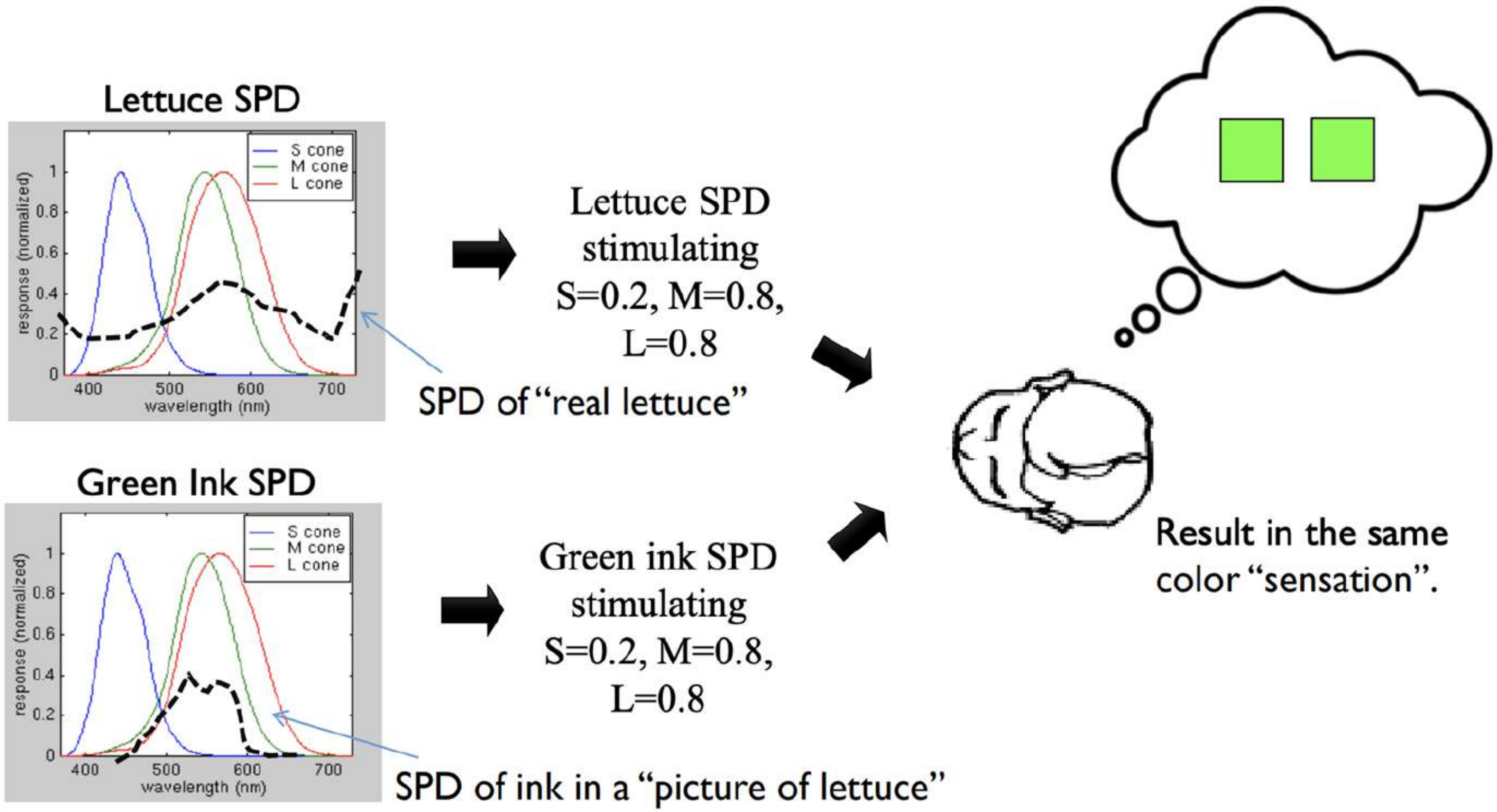
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We need metamers!



[Seitz]

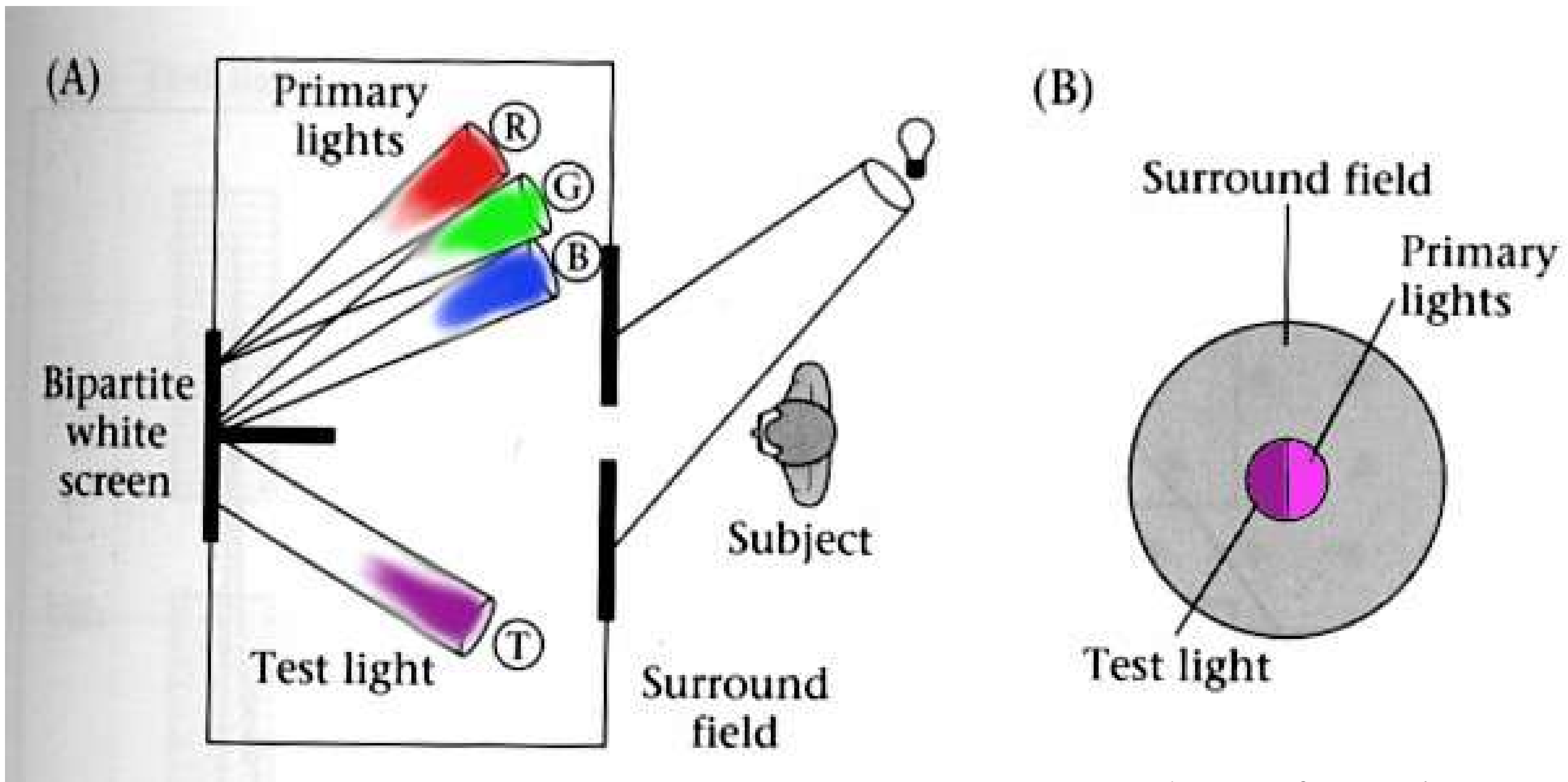
We need metamers!



[Seitz]

Standardizing color experience

- We would like to understand which spectra produce the same color sensation in people under similar viewing conditions
- Color matching experiments



Foundations of Vision, by Brian Wandell, Sinauer Assoc., 1995

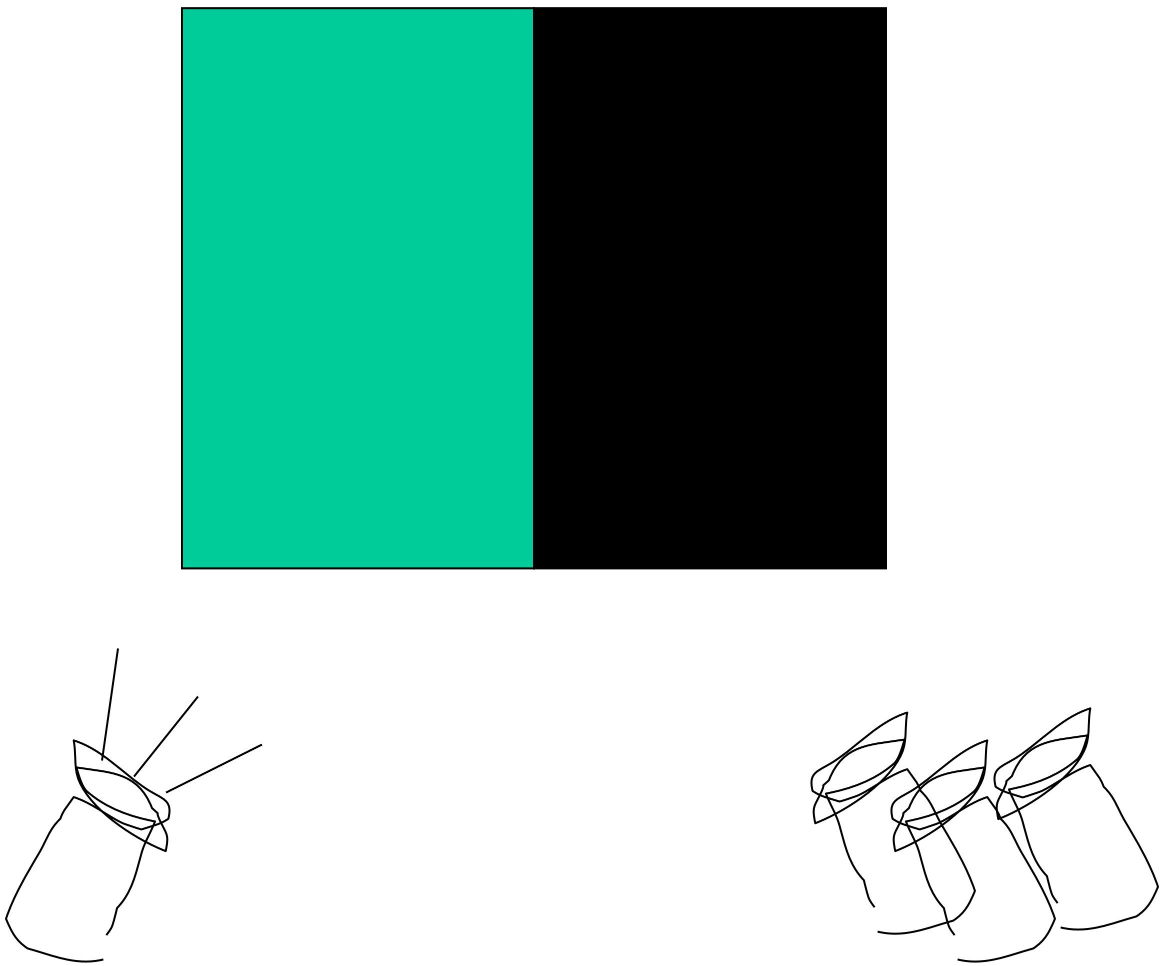
Trichromacy

- In color matching experiments, most people can match any given light with three primaries
  - Primaries must be *independent*
- For the same light and same primaries, most people select the same weights
  - Exception: color blindness
- Trichromatic color theory
  - Three numbers seem to be sufficient for encoding color
  - Dates back to 18th century (Thomas Young)

Co-financed by the European Union
Connecting Europe Facility

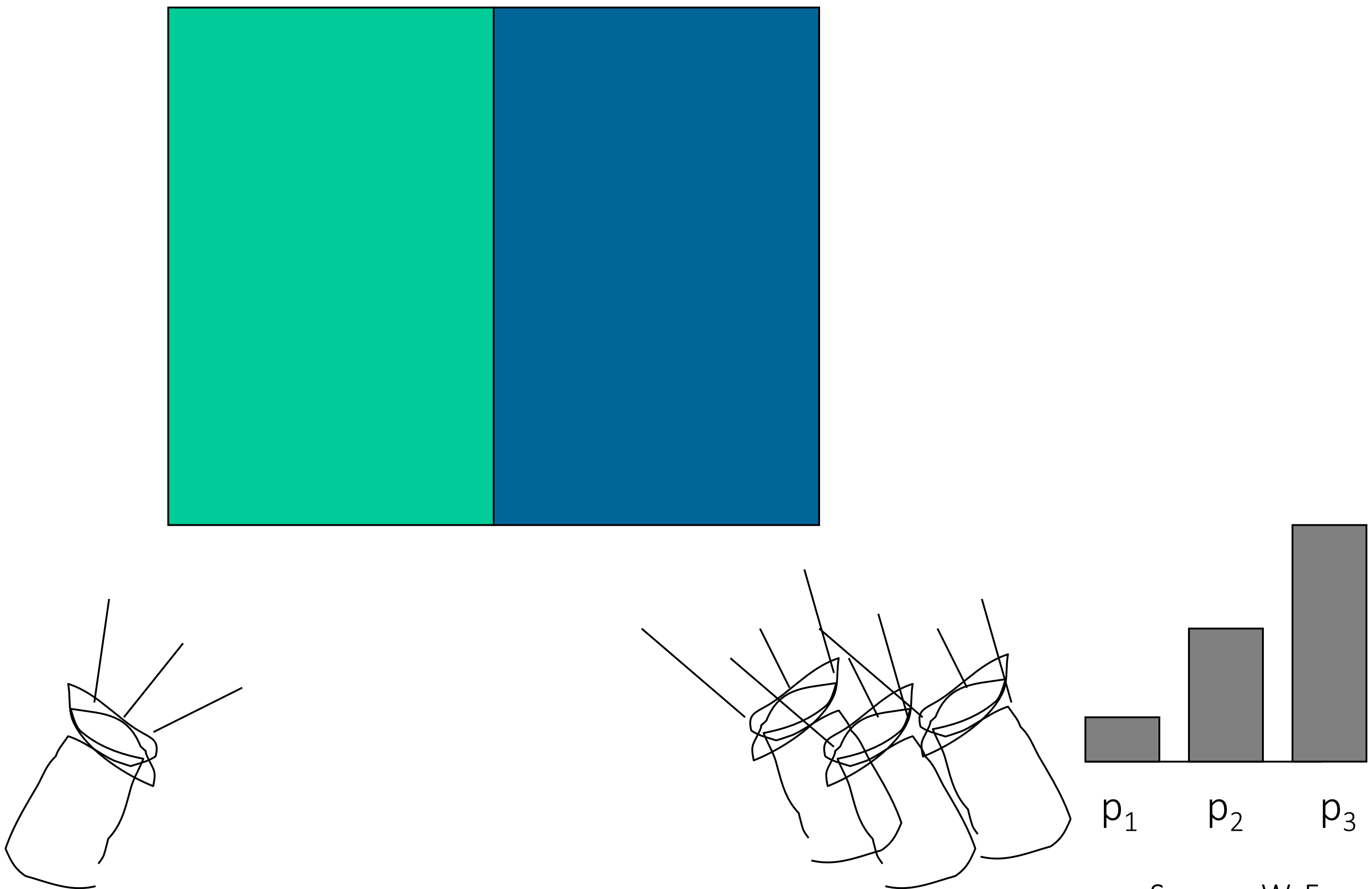
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Color matching experiment 1



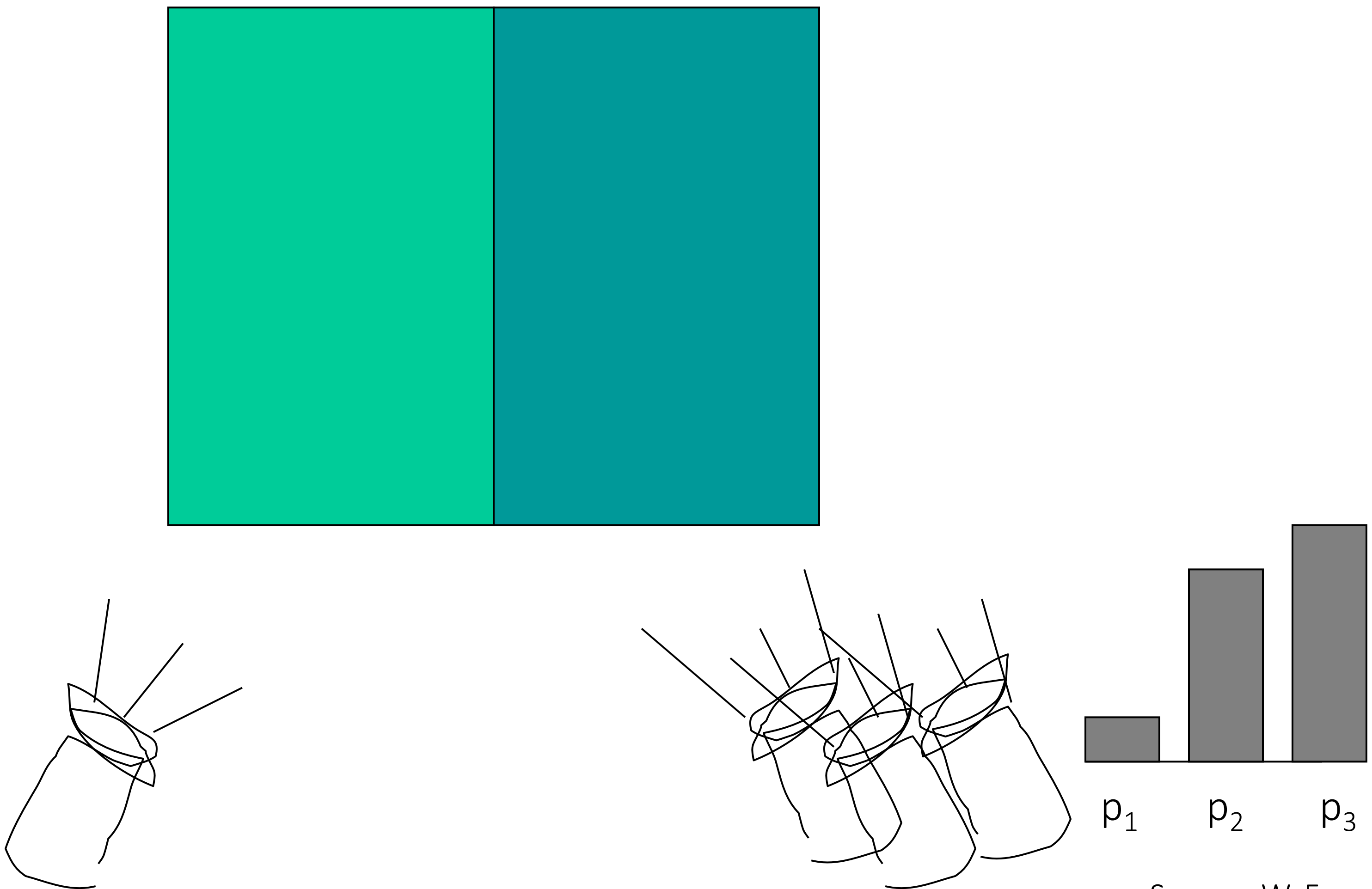
Source: W. Freeman

Color matching experiment 1



Source: W. Freeman

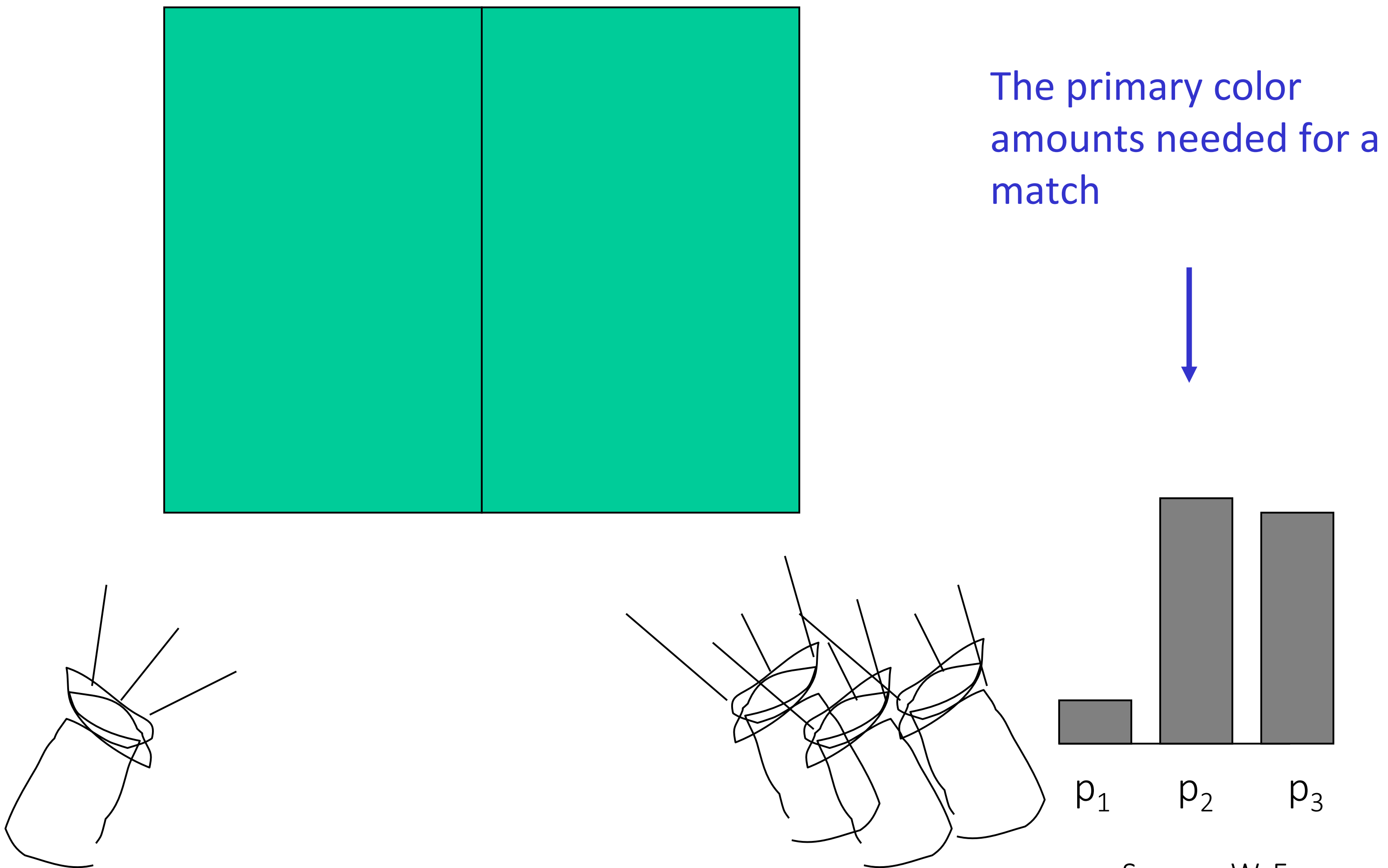
Color matching experiment 1



Source: W. Freeman

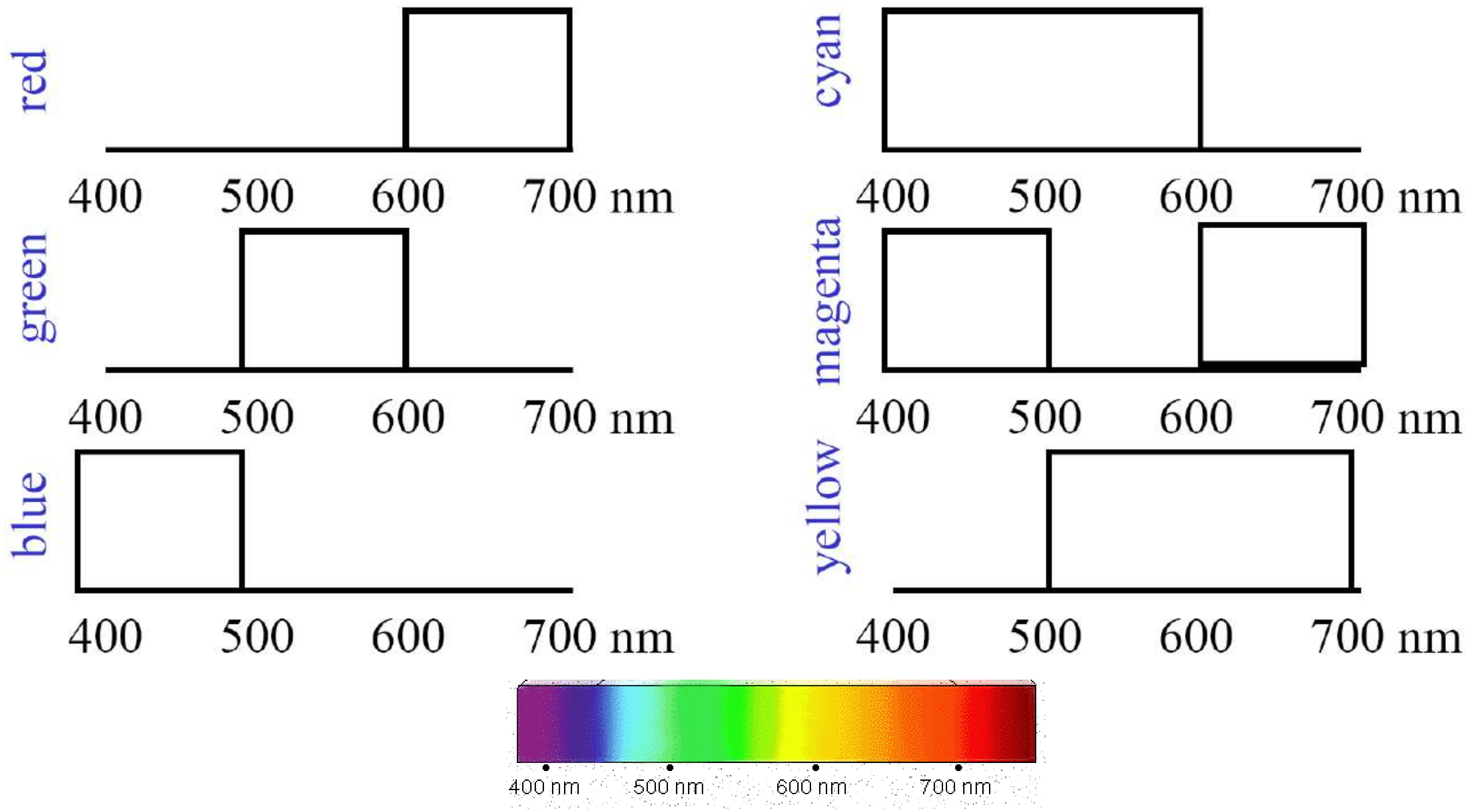
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Color matching experiment 1



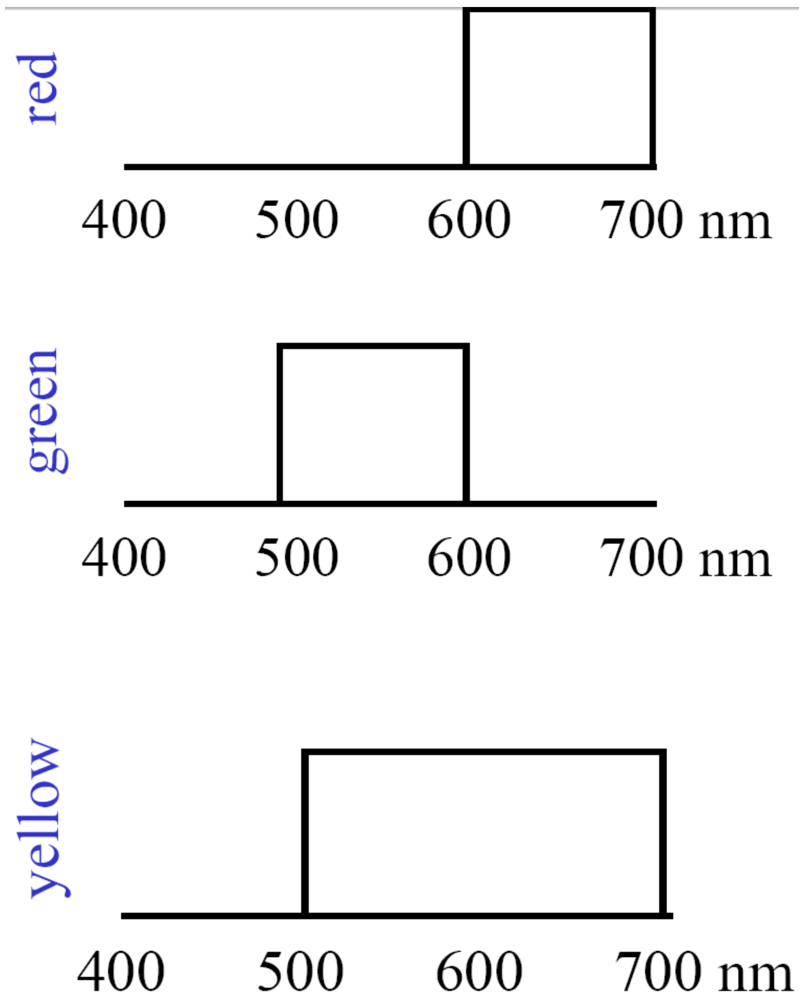
Source: W. Freeman

Color mixing

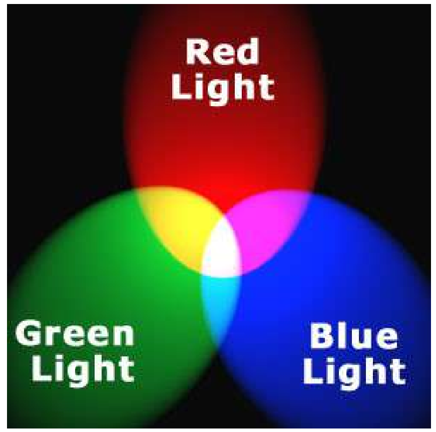


Source: W. Freeman

Additive color mixing



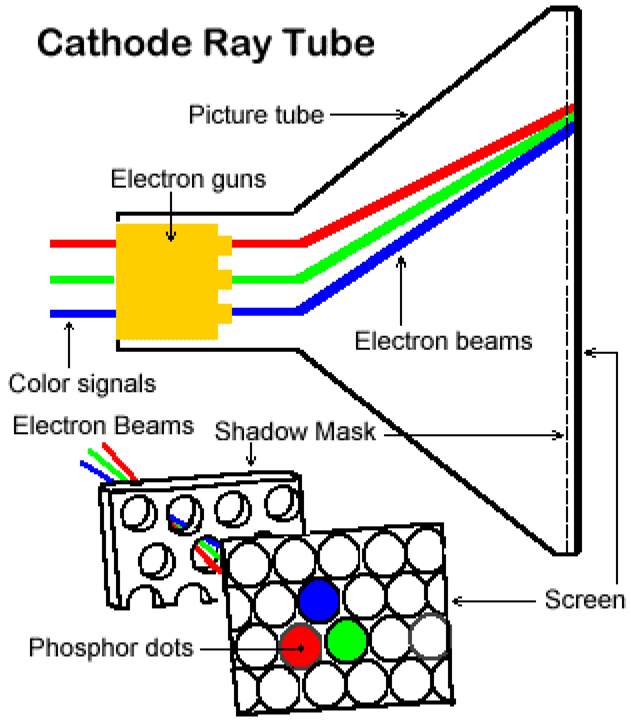
Colors combine by *adding* color spectra



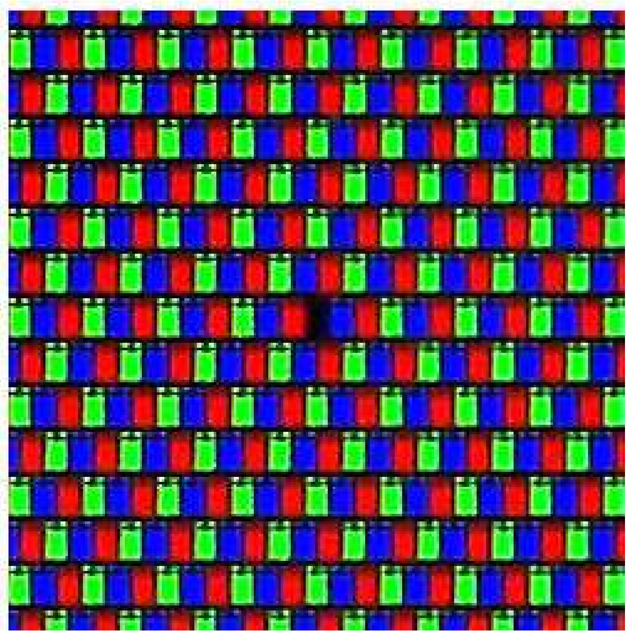
Light *adds* to existing black.

Source: W. Freeman

Examples of additive color systems

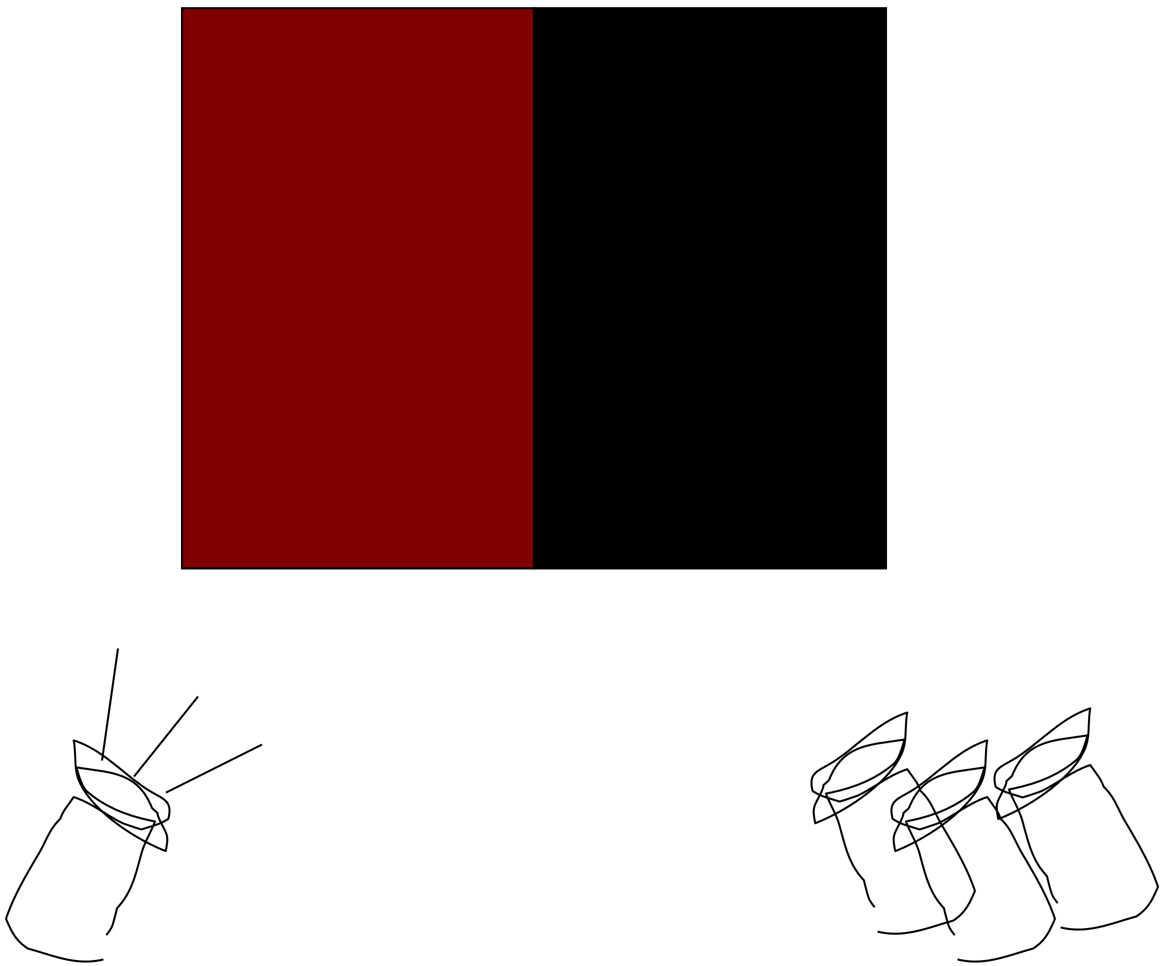


CRT phosphors



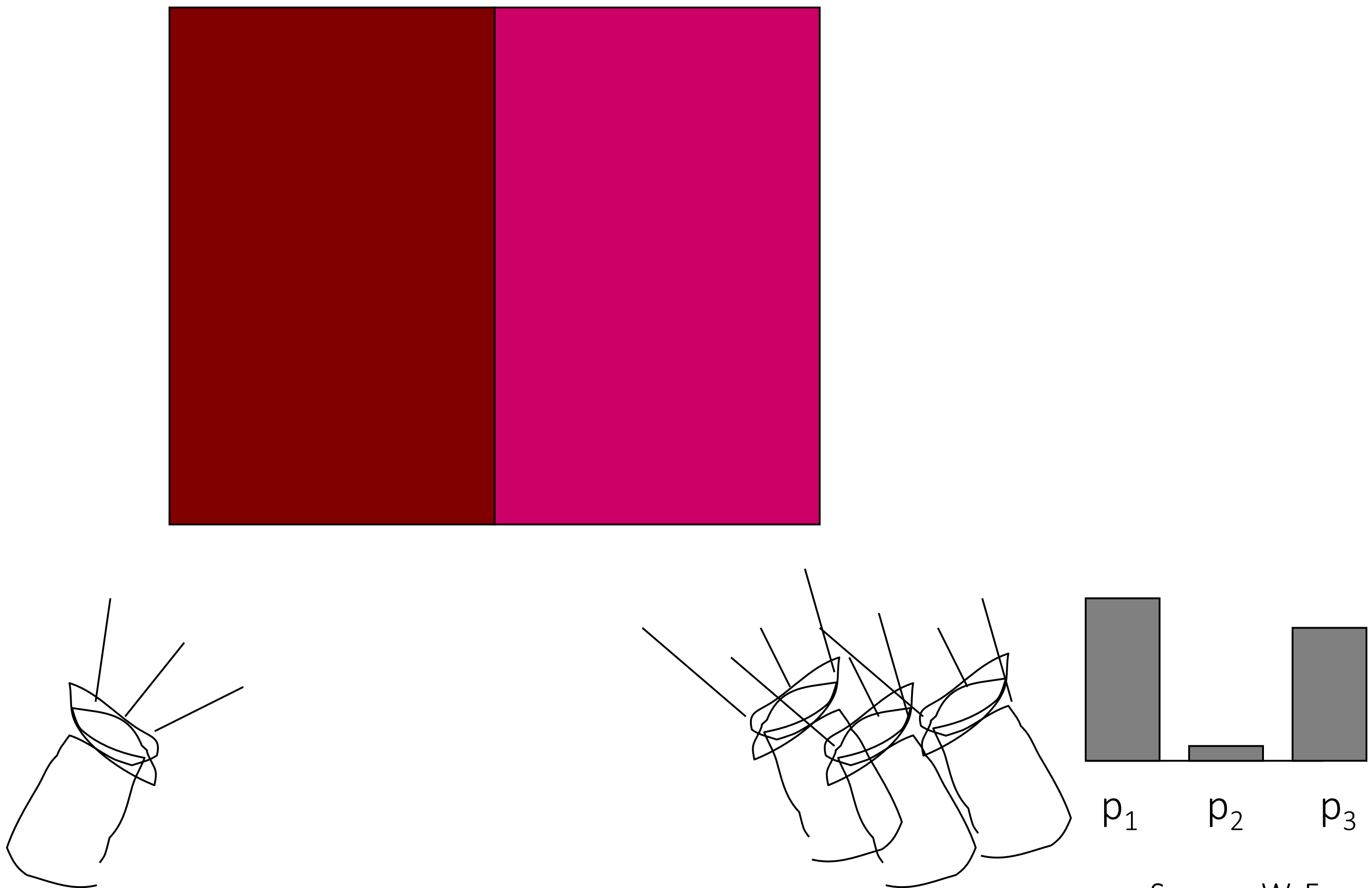
Multiple projectors

Color matching experiment 2



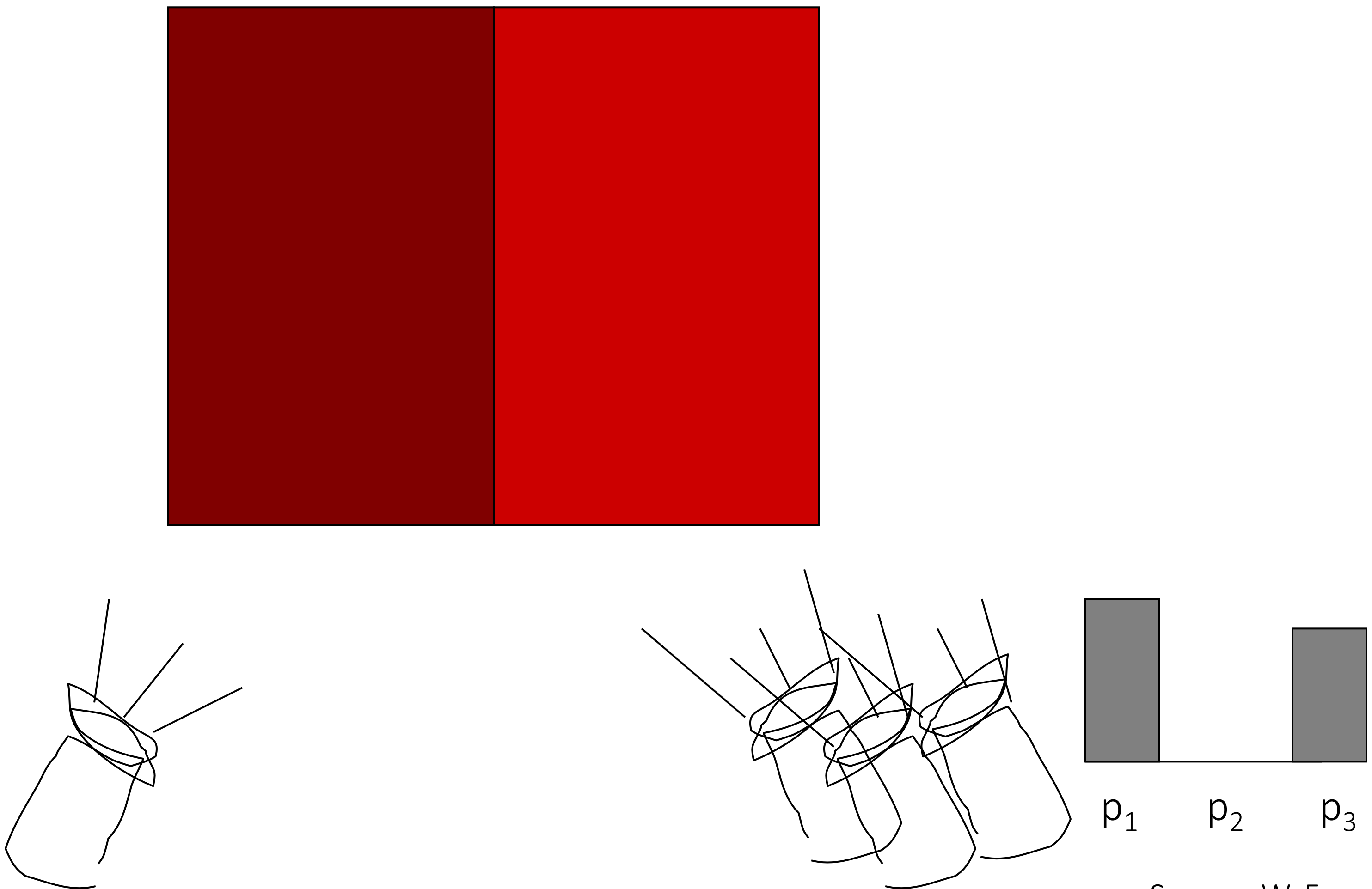
Source: W. Freeman

Color matching experiment 2



Source: W. Freeman

Color matching experiment 2

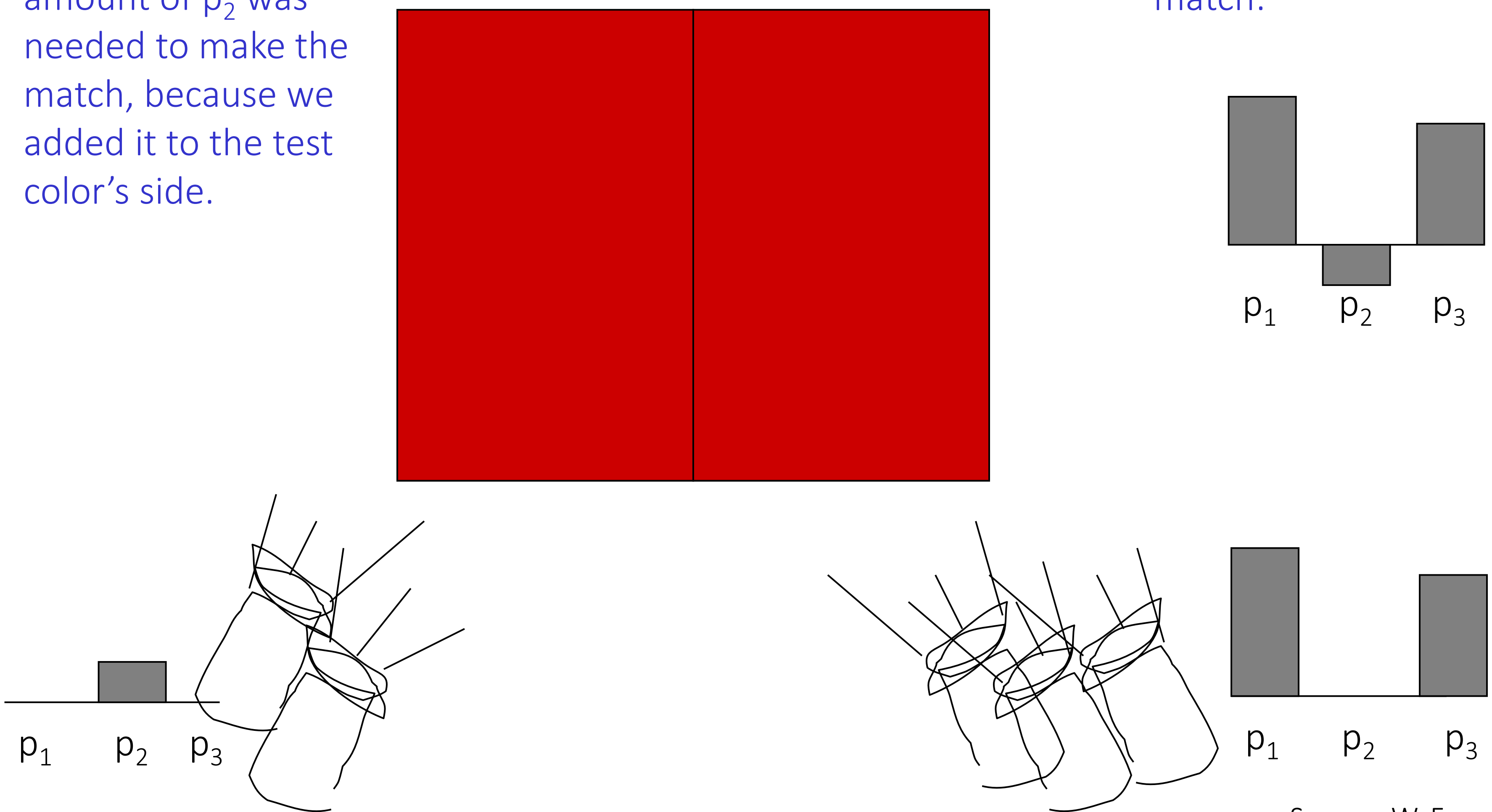


Source: W. Freeman

Color matching experiment 2

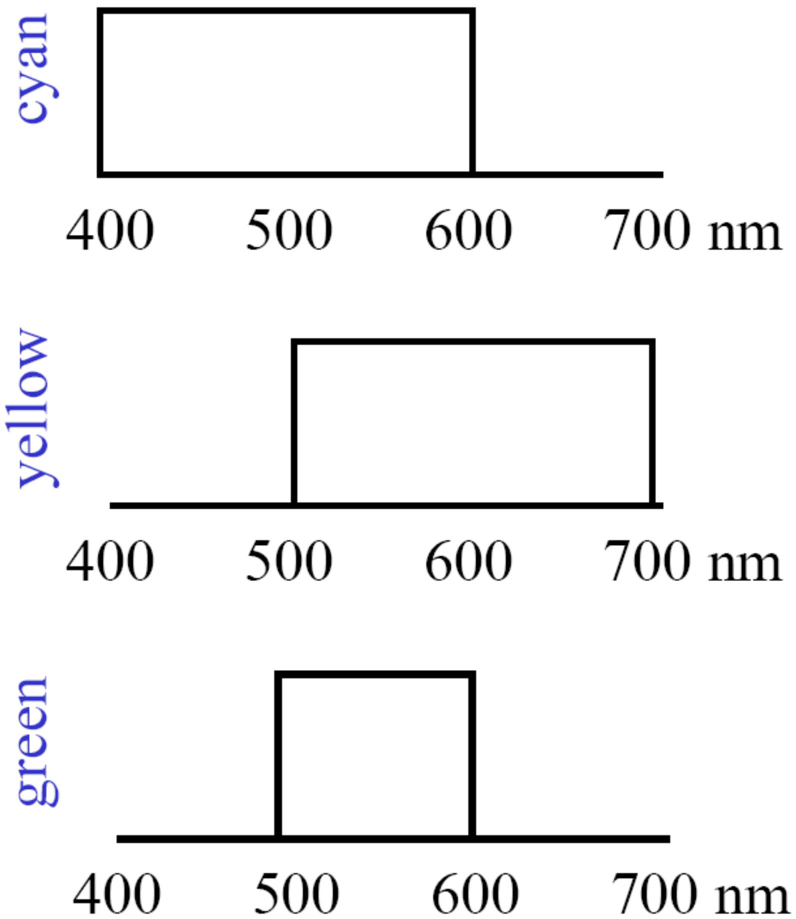
We say a “negative” amount of p_2 was needed to make the match, because we added it to the test color’s side.

The primary color amounts needed for a match:

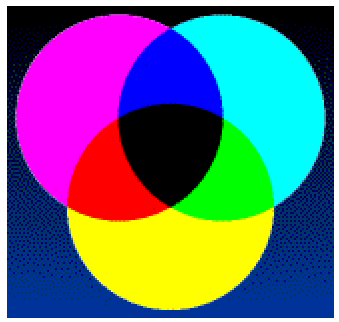


Source: W. Freeman

Subtractive color mixing



Colors combine by *multiplying* color spectra.



Pigments *remove* color from incident light (white).

Source: W. Freeman

Examples of subtractive color systems

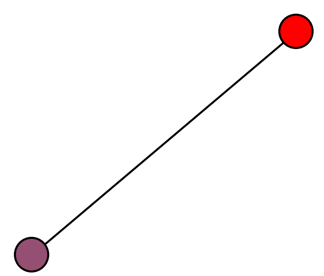
- Printing on paper
- Crayons
- Photographic film

Overview of Color

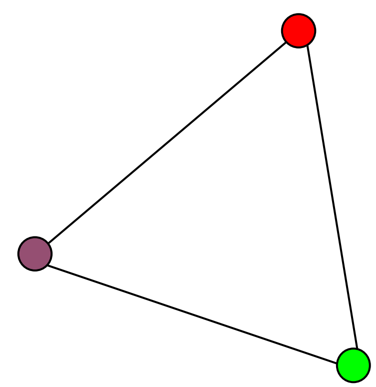
- Physics of color
- Human encoding of color
- Color spaces

Linear color spaces

- Defined by a choice of three *primaries*
- The coordinates of a color are given by the weights of the primaries used to match it



mixing two lights produces colors that lie along a straight line in color space



mixing three lights produces colors that lie within the triangle they define in color space

Co-financed by the European Union
Connecting Europe Facility

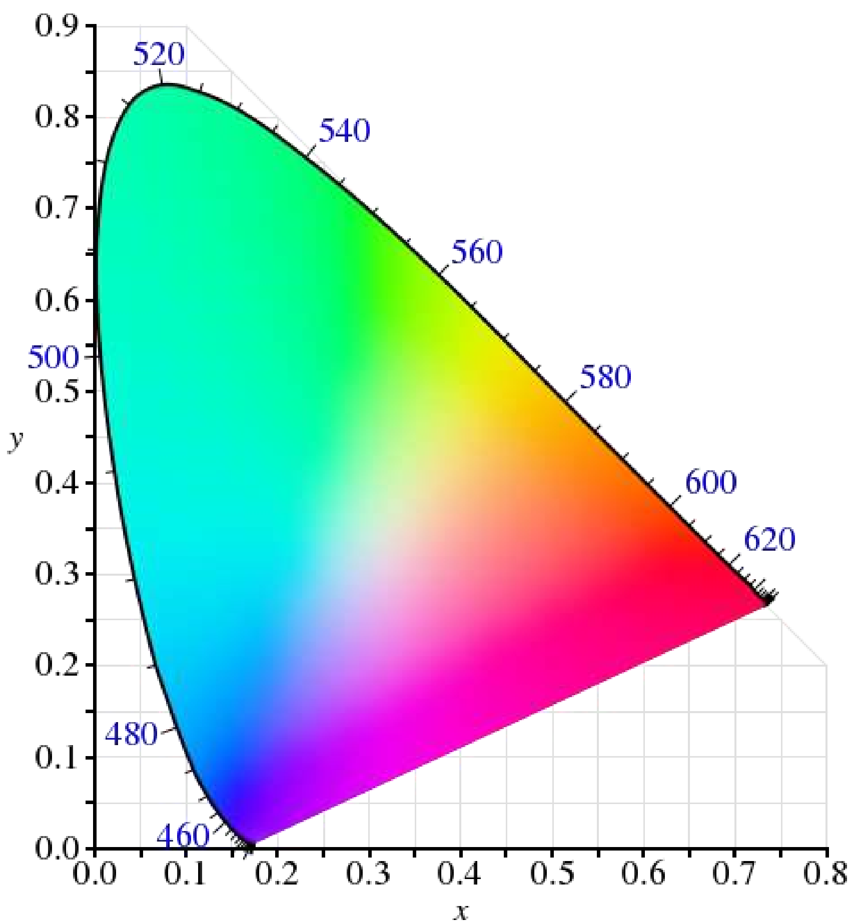
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Linear color spaces

- Pick some primaries
- Can mix those primaries to match any color inside the triangle
- There is a commission that studies color!
  - Commission internationale de l'éclairage (CIE) is a 100-year-old organization that creates international standards related to light and color.

[Shapiro]

CIE 1931 XYZ color space



- First attempt to produce a color space based on measurements of human color perception

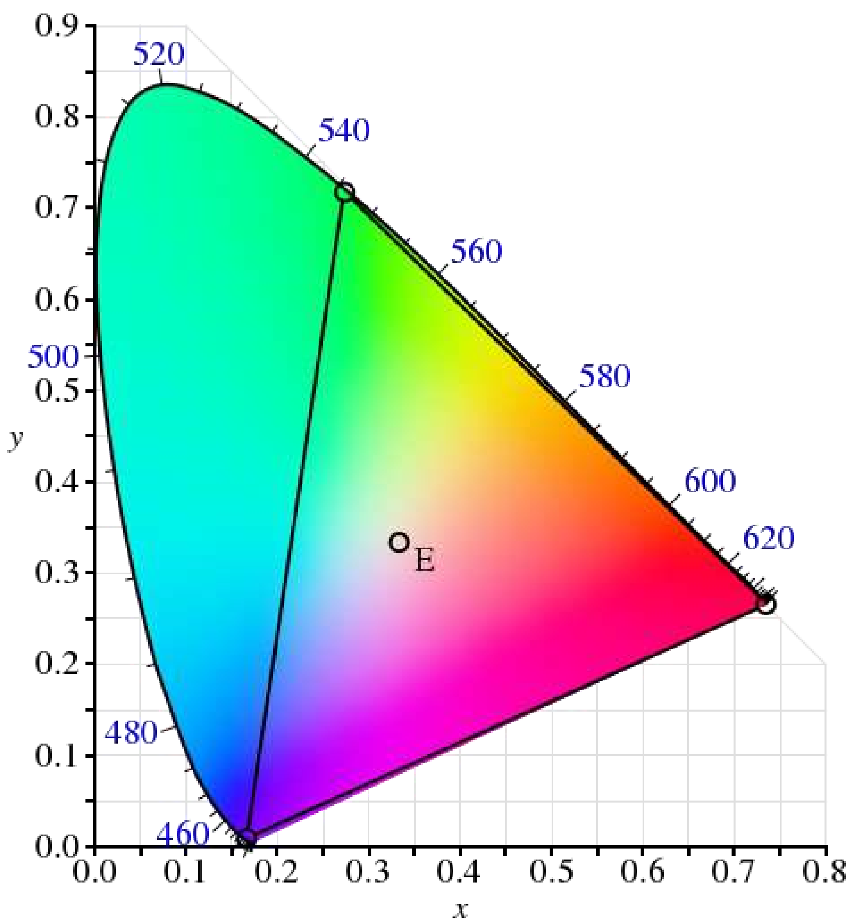
- Y corresponds to brightness or *luminance* of a color. Z is like blue light stimulation, and X is a mix (a linear combination) of cone response curves chosen to be nonnegative.

- 2D visualization - draw (x,y), where: $x = X/(X+Y+Z)$, $y = Y/(X+Y+Z)$

http://en.wikipedia.org/wiki/CIE_1931_color_space

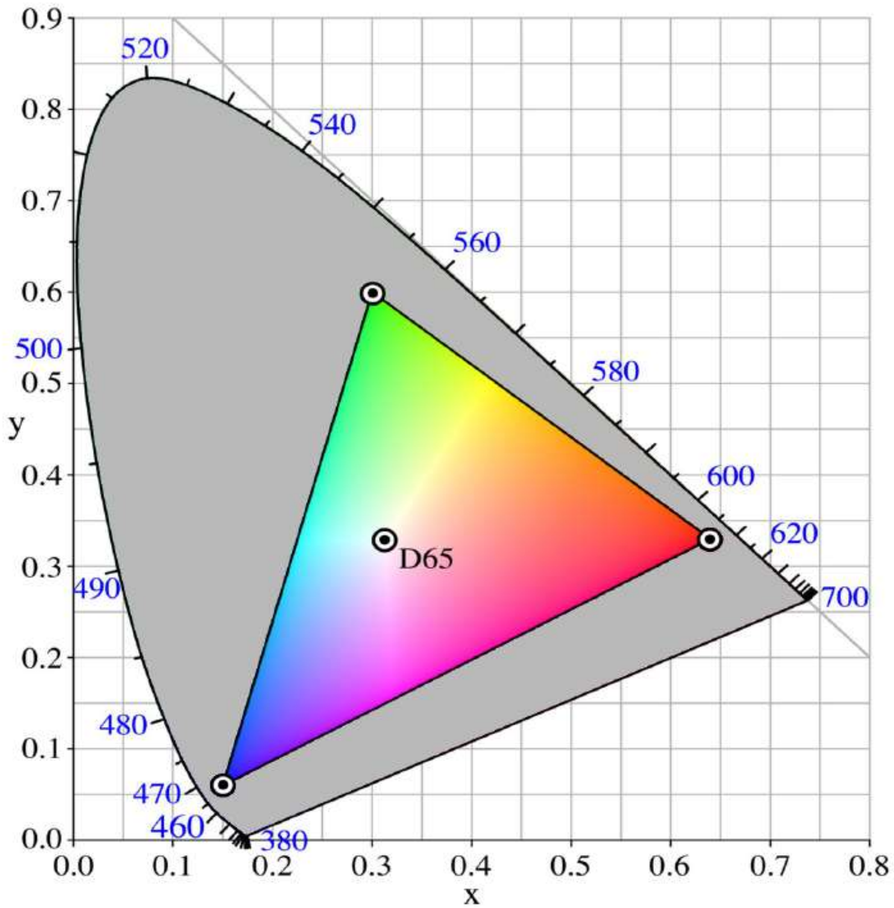
[Shapiro]

'Theoretical' CIE RGB primaries



[Shapiro]

Practical sRGB primaries, MSFT 1996



- sRGB (standard Red Green Blue) is an RGB color space that HP and Microsoft created cooperatively in 1996 to use on monitors, printers, and the Internet
- It was subsequently standardized by the IEC (International Electrotechnical Commission) as IEC 61966-2-1:1999

[Shapiro]

What does this mean for computers?

- We represent images as grids of pixels
- Each pixel has a color, 3 components: RGB
- Not every color can be represented in RGB!
  - Have to go out in the real world sometimes
- RGB is not fully accurate
- We can represent a color with 3 numbers
  - #ff00ff; (1.0, 0.0, 1.0); 255,0,255
  - What color is this?

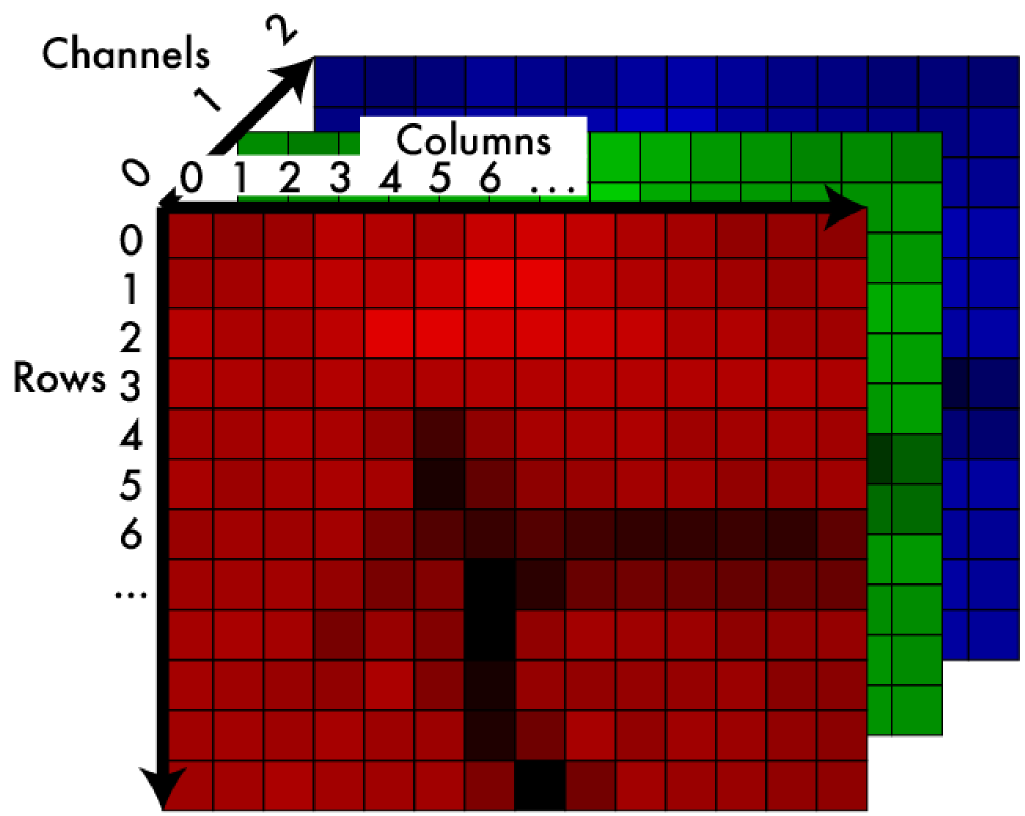
[Shapiro]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image: 2d array of color

- Some range
  - [0,255]
  - [0.0,1.0]
- We'll talk more about this later.

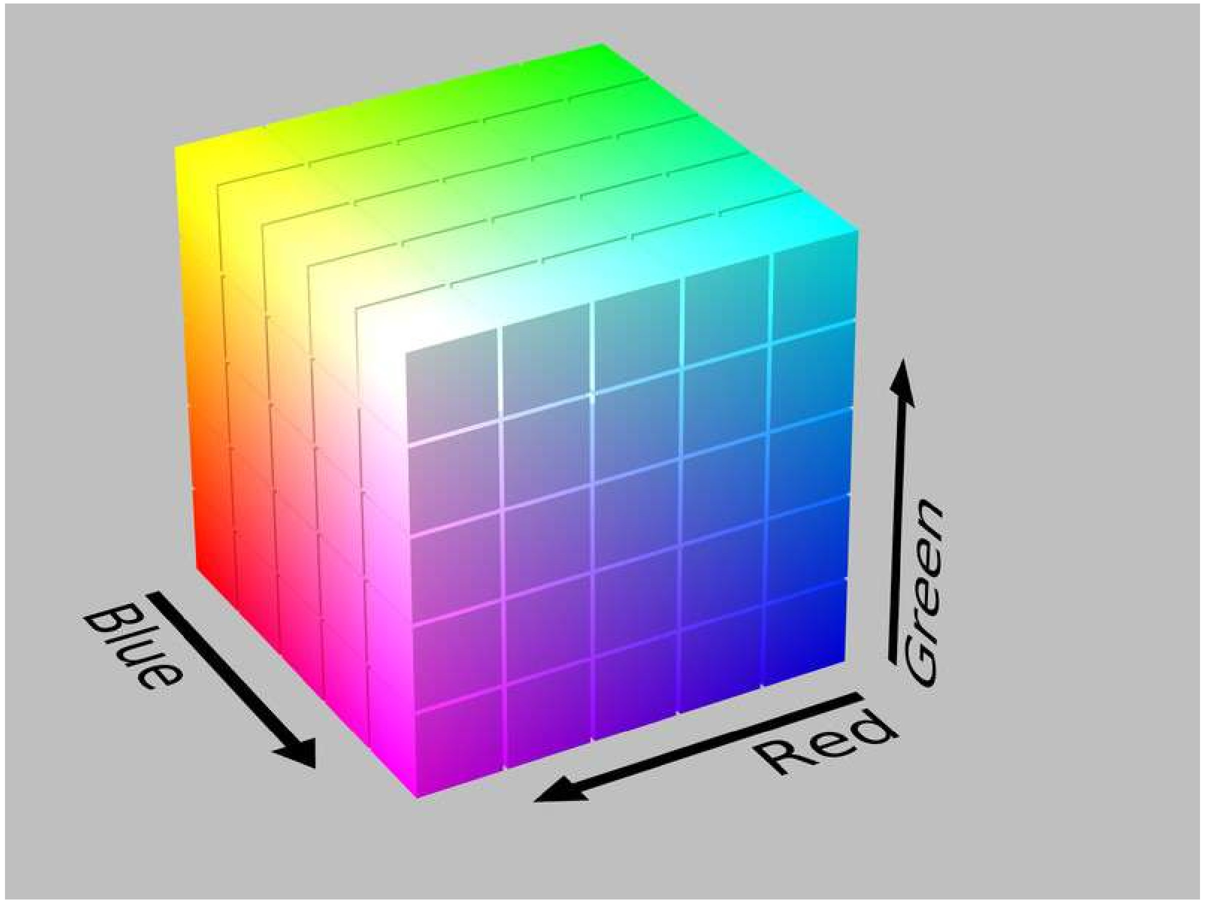


[Shapiro]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

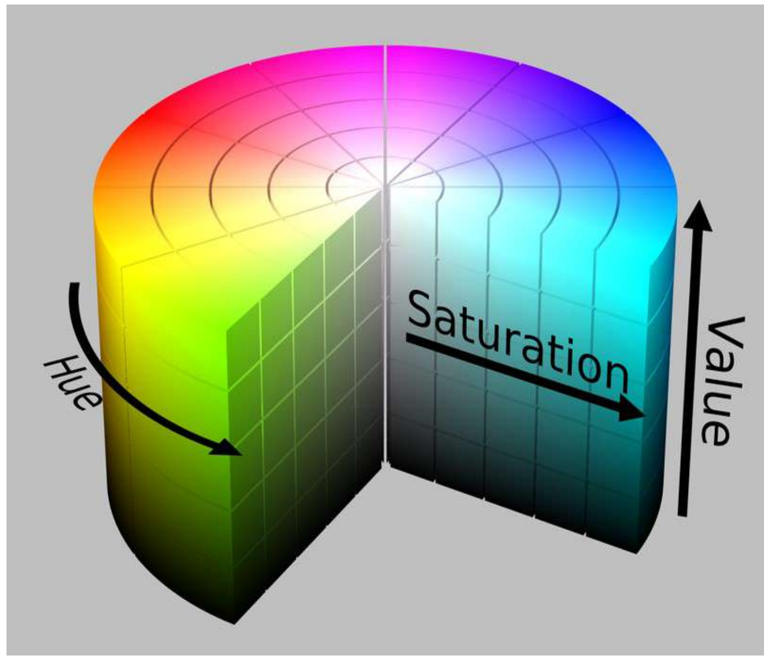
RGB is a cube...



[Shapiro]

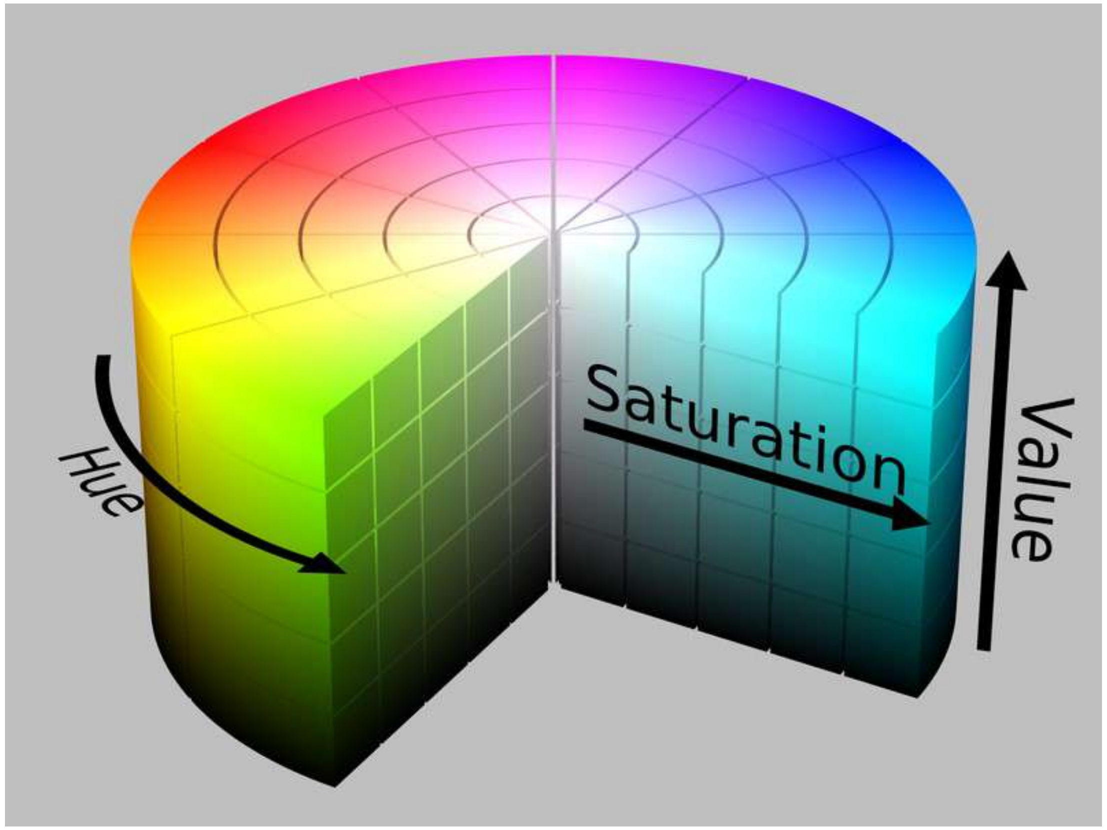
Non-linear color spaces: Hue, Saturation, Value

- Different model based on perception of light
- Hue: what color
- Saturation: how much color
- **Value**: how bright
- Allows easy image transforms
  - Shift the hue
  - Increase saturation



[Shapiro]

Hue, Saturation, Value: cylinder!



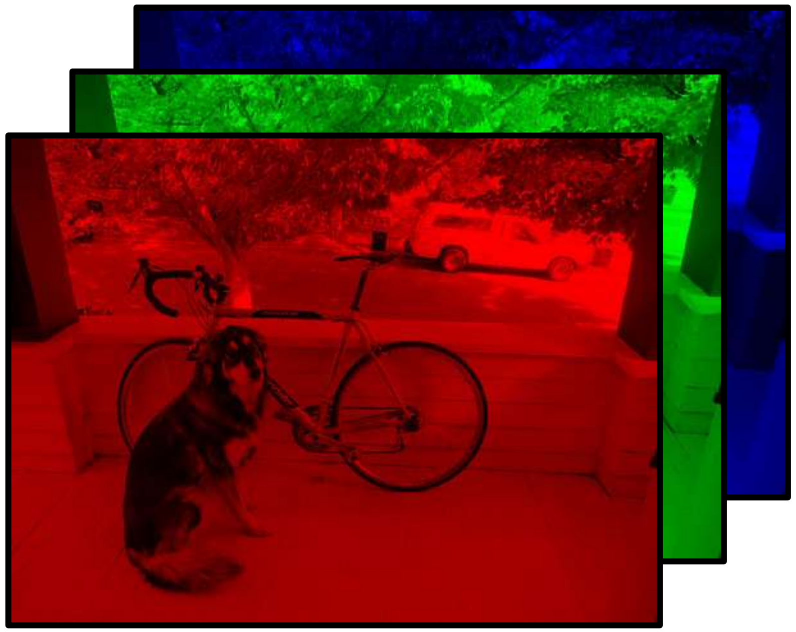
[Shapiro]

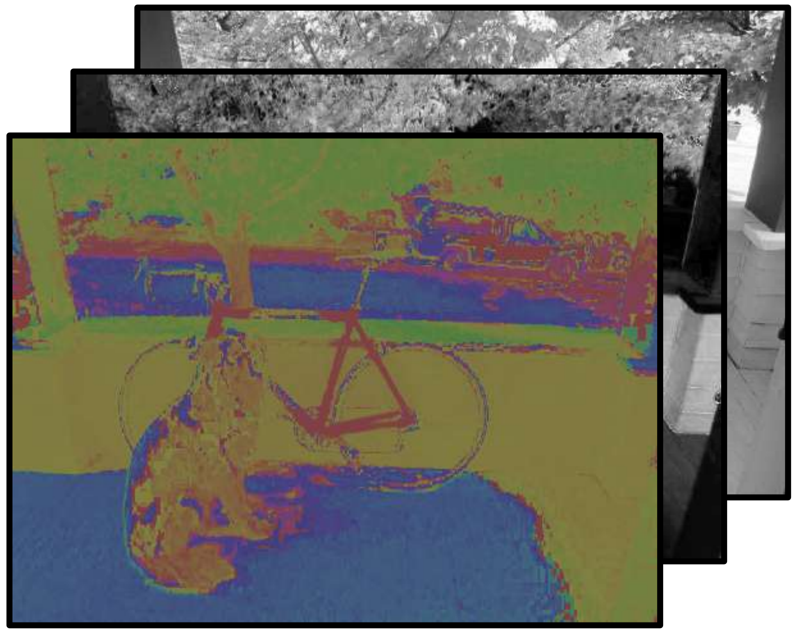
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

An RGB Image

[Shapiro]

Still 3d tensor, different info



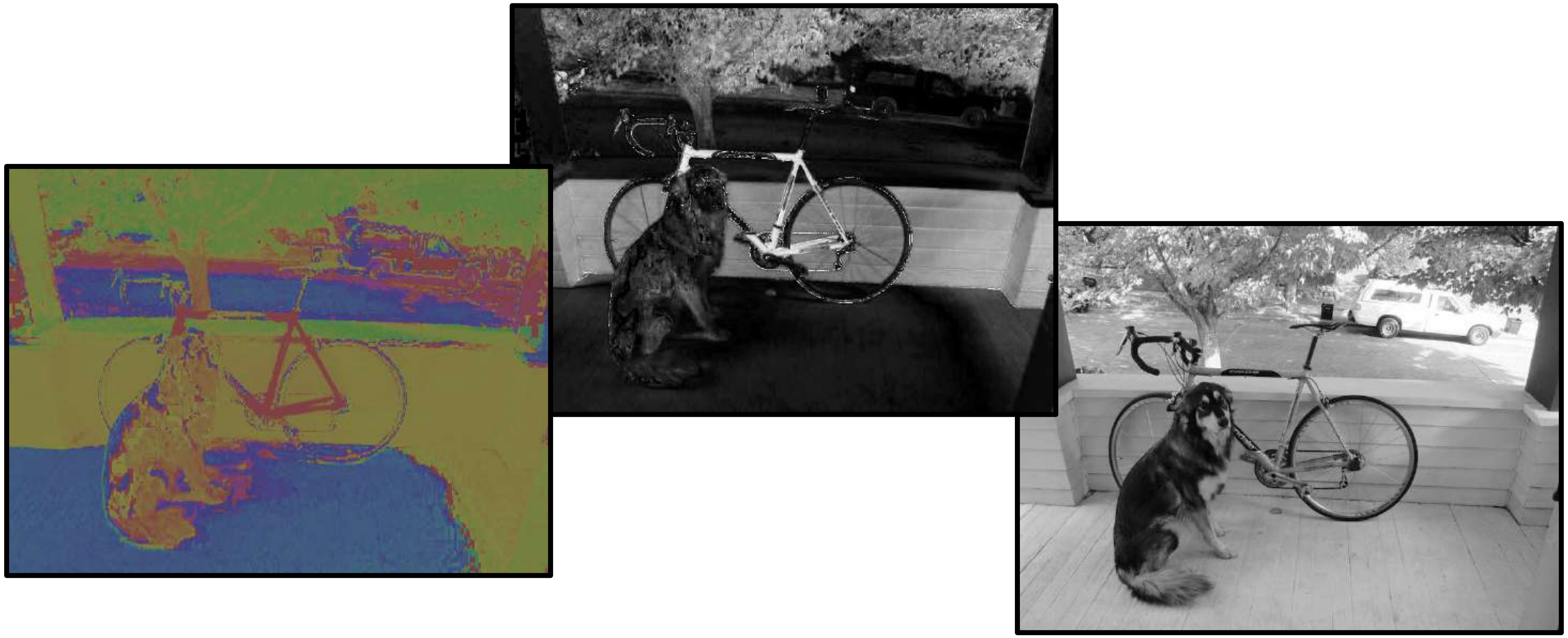


[Shapiro]

Co-financed by the European Union
Connecting Europe Facility

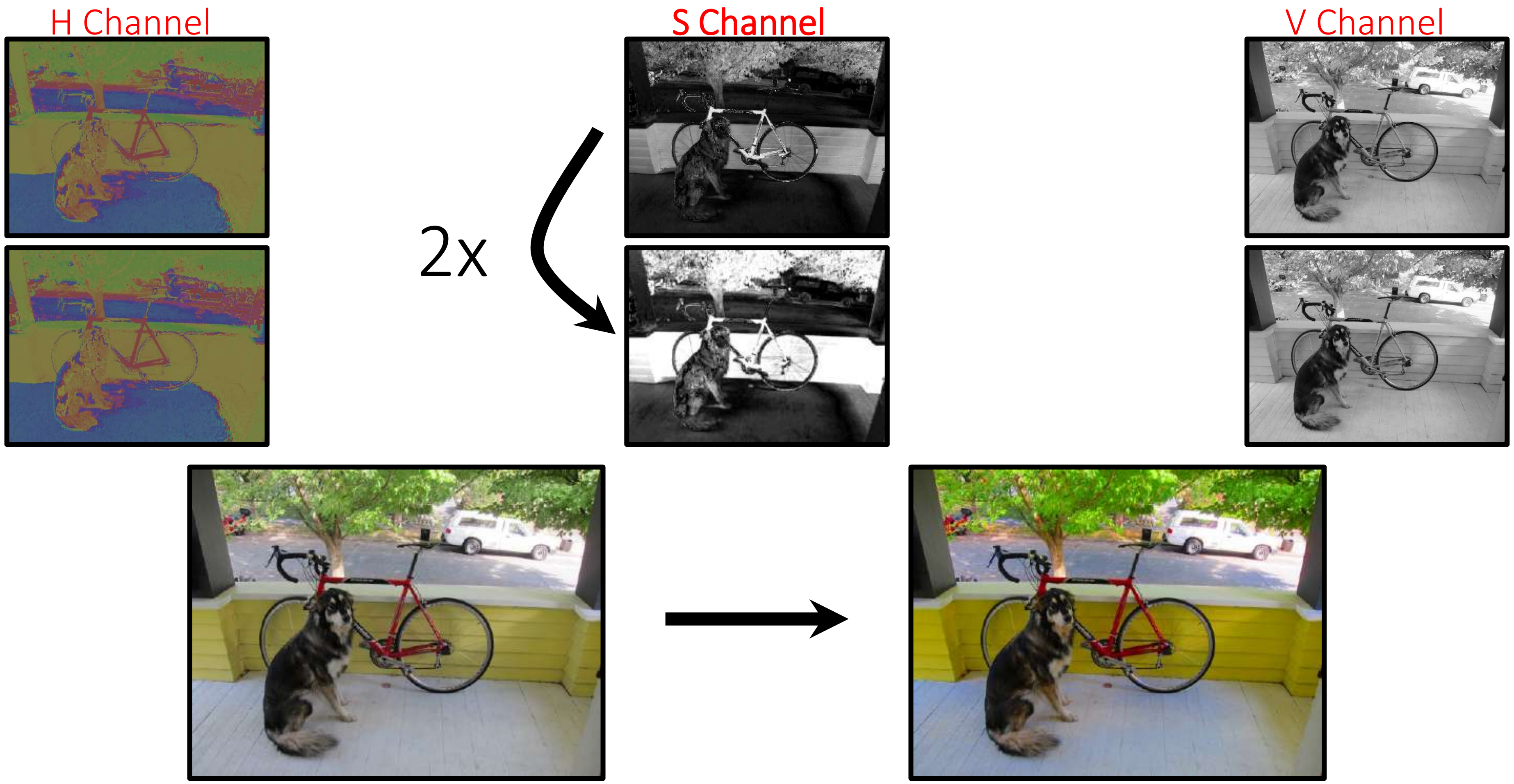
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Hue Saturation Value



[Shapiro]

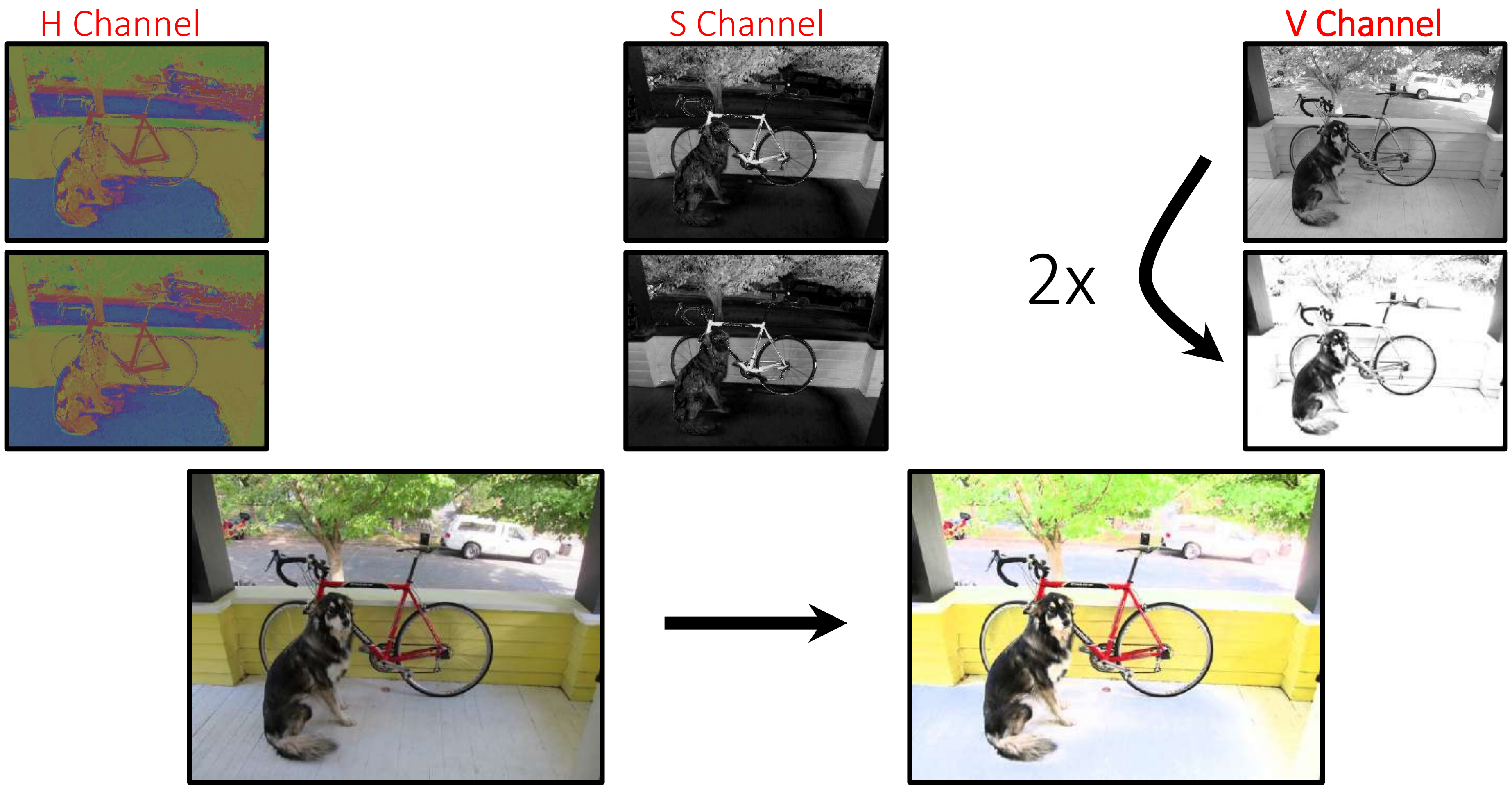
More saturation = intense colors



[Shapiro]

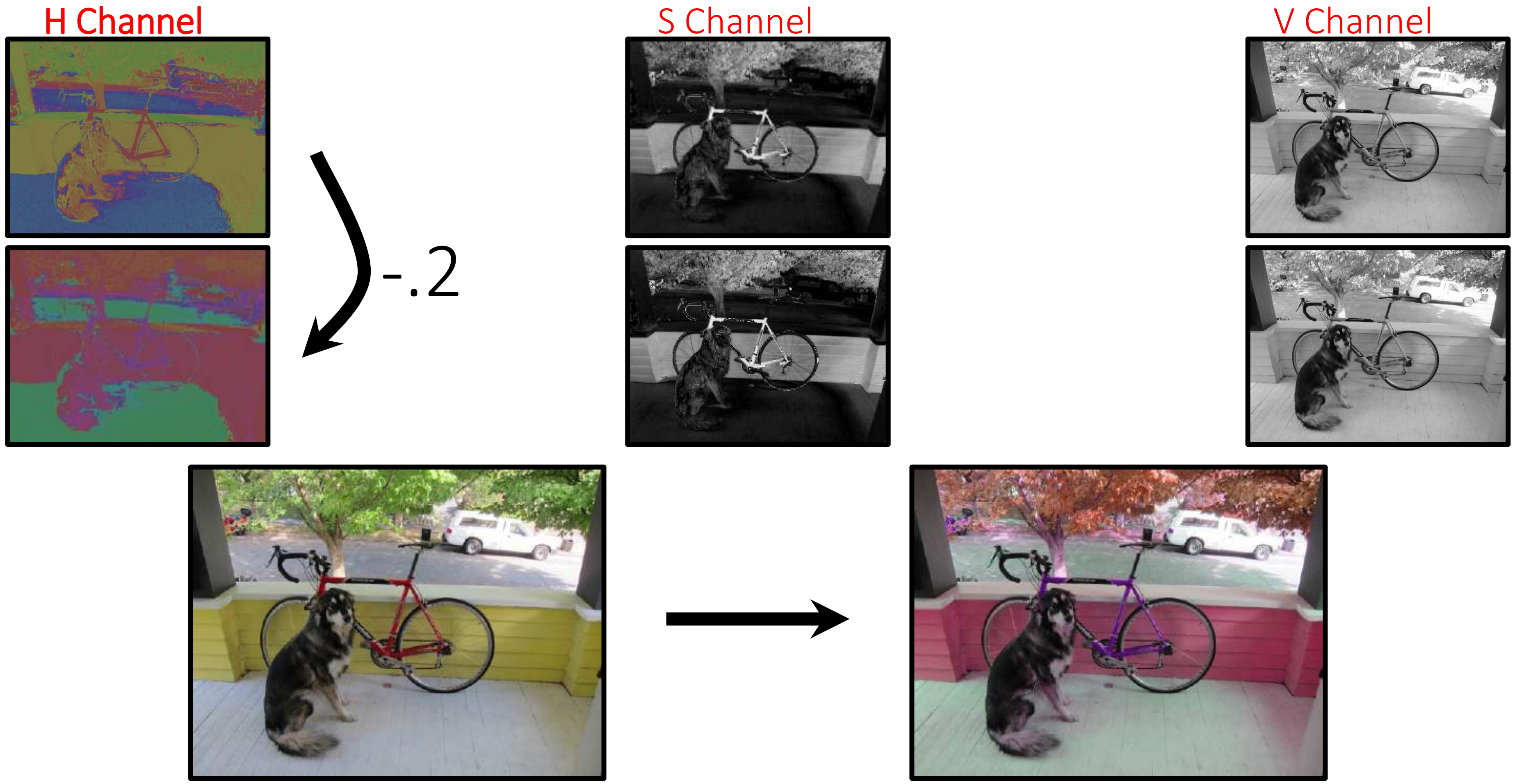
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

More value = lighter image



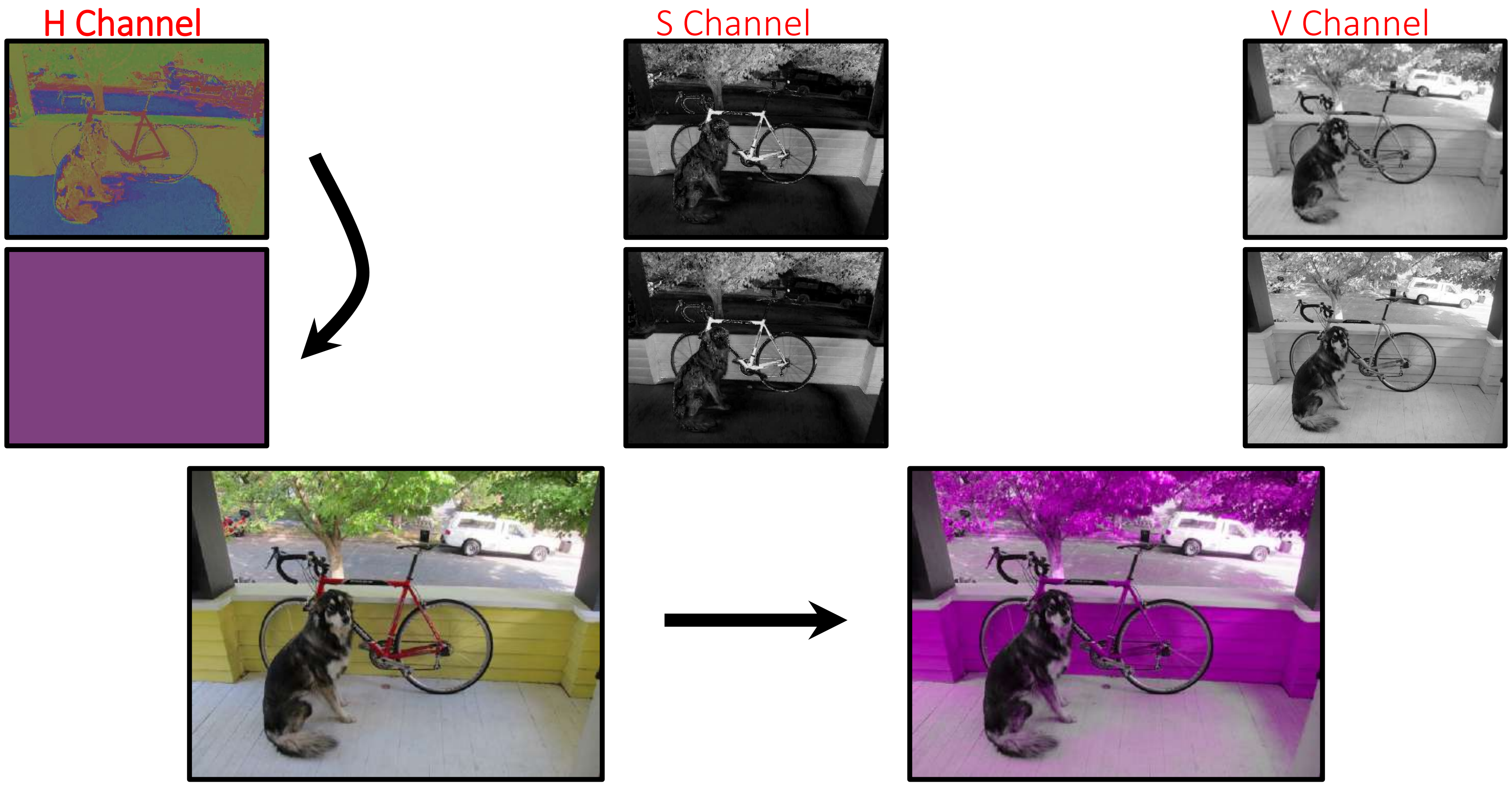
[Shapiro]

Shift hue = shift colors



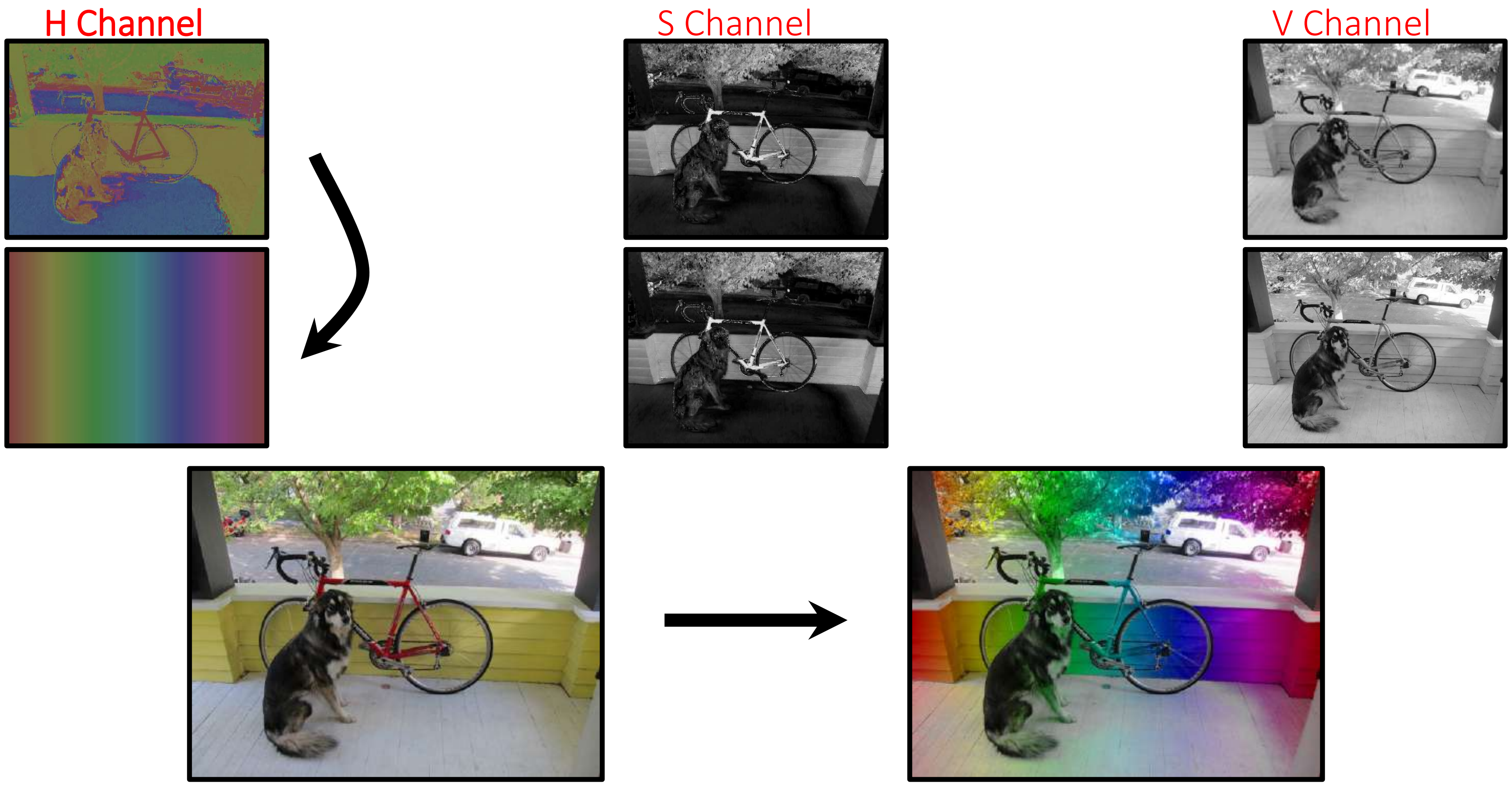
[Shapiro]

Set hue to your favorite color!



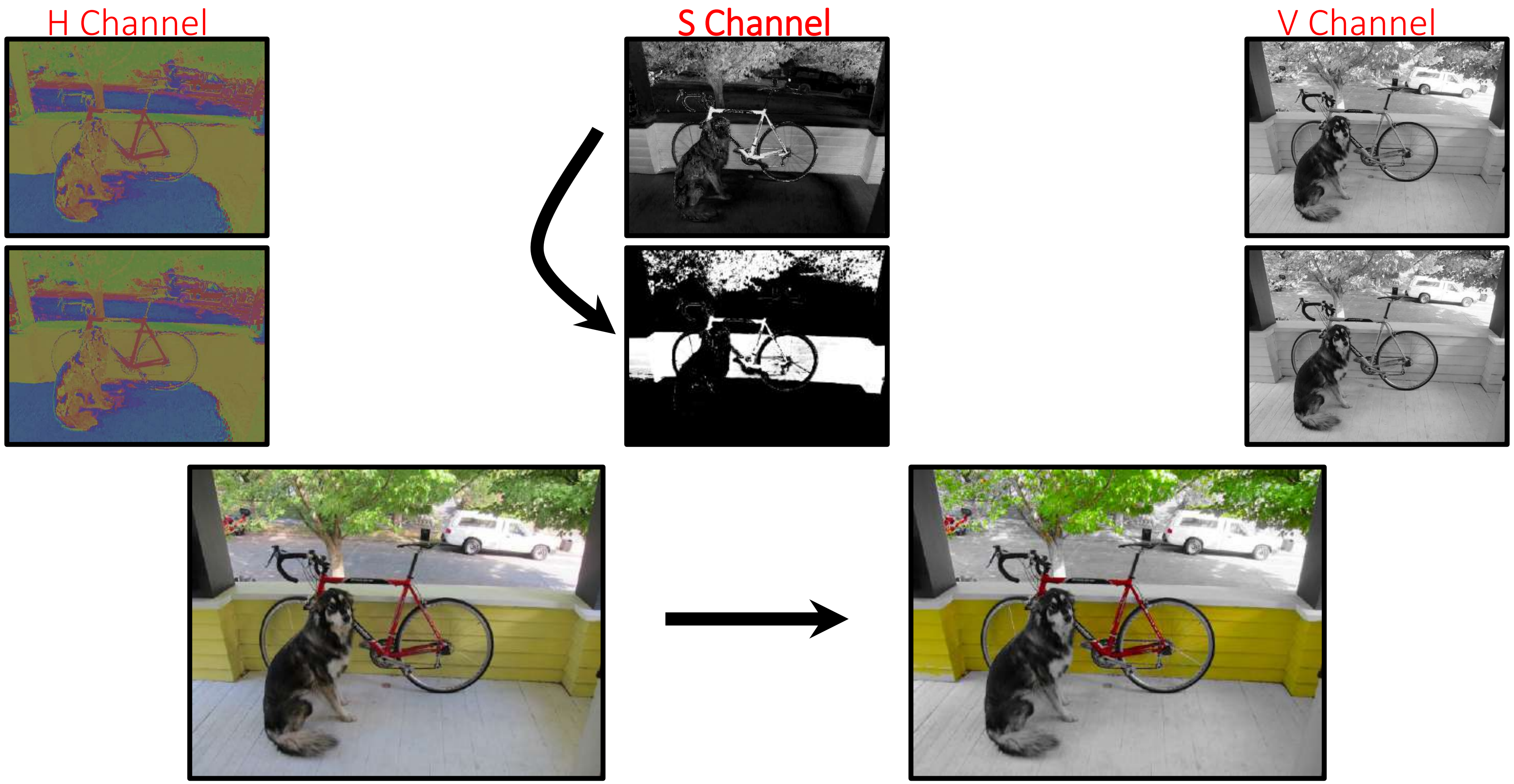
[Shapiro]

Or pattern...



[Shapiro]

Increase and threshold saturation



[Shapiro]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

[Shapiro]

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 3: Fundamentals - Cameras

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Physics of color
- Human encoding of color
- Color spaces

Today’s Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

Co-financed by the European Union
Connecting Europe Facility

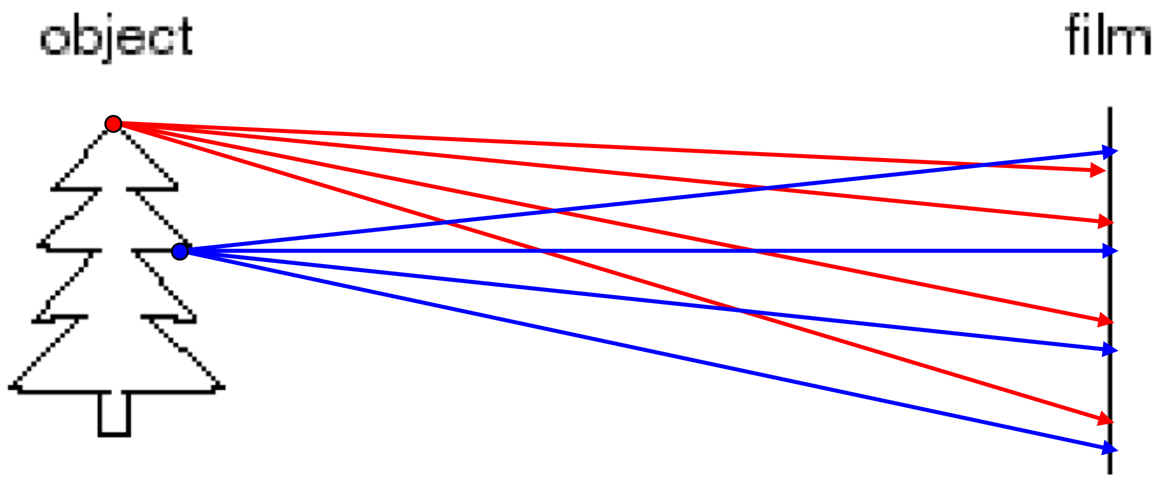
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Let's design a camera

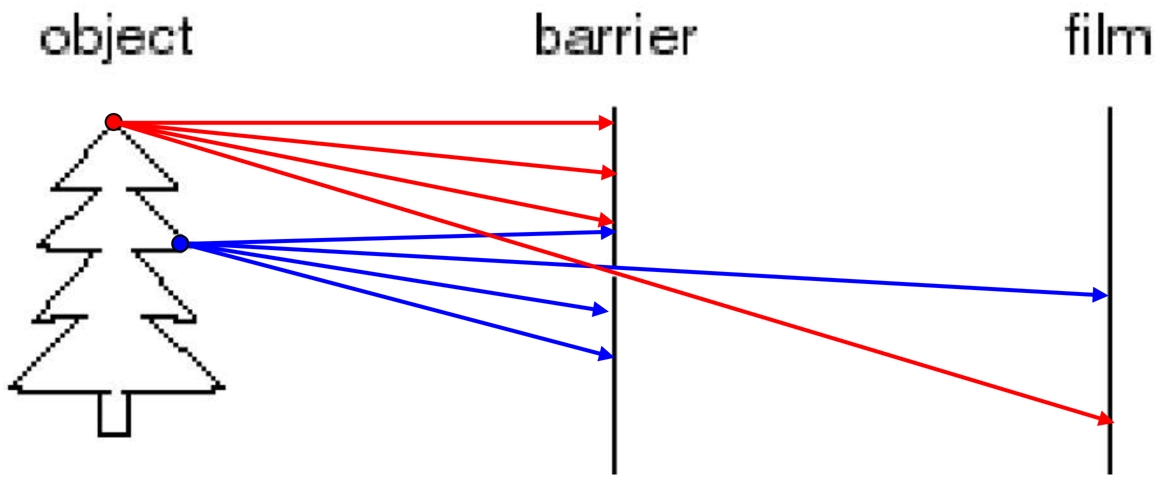


- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

[Slide by Steve Seitz]

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Pinhole camera



Add a barrier to block off most of the rays

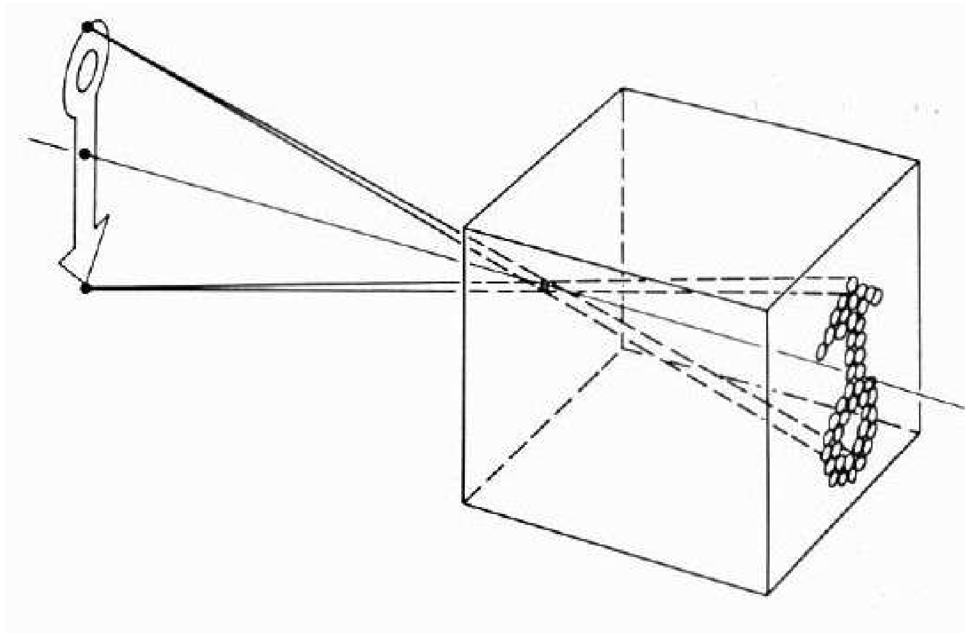
- This reduces blurring
- The opening known as the **aperture**
- How does this transform the image?

[Slide by Steve Seitz]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Pinhole Camera model

- Captures **pencil of rays** – all rays through a single point
- The point is called **Center of Projection**
- The image is formed on the **Image Plane**



[Slide by Steve Seitz]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Pinhole cameras everywhere

Sun “shadows” during a solar eclipse
by Henrik von Wendt http://www.flickr.com/photos/hvw/2724969199/

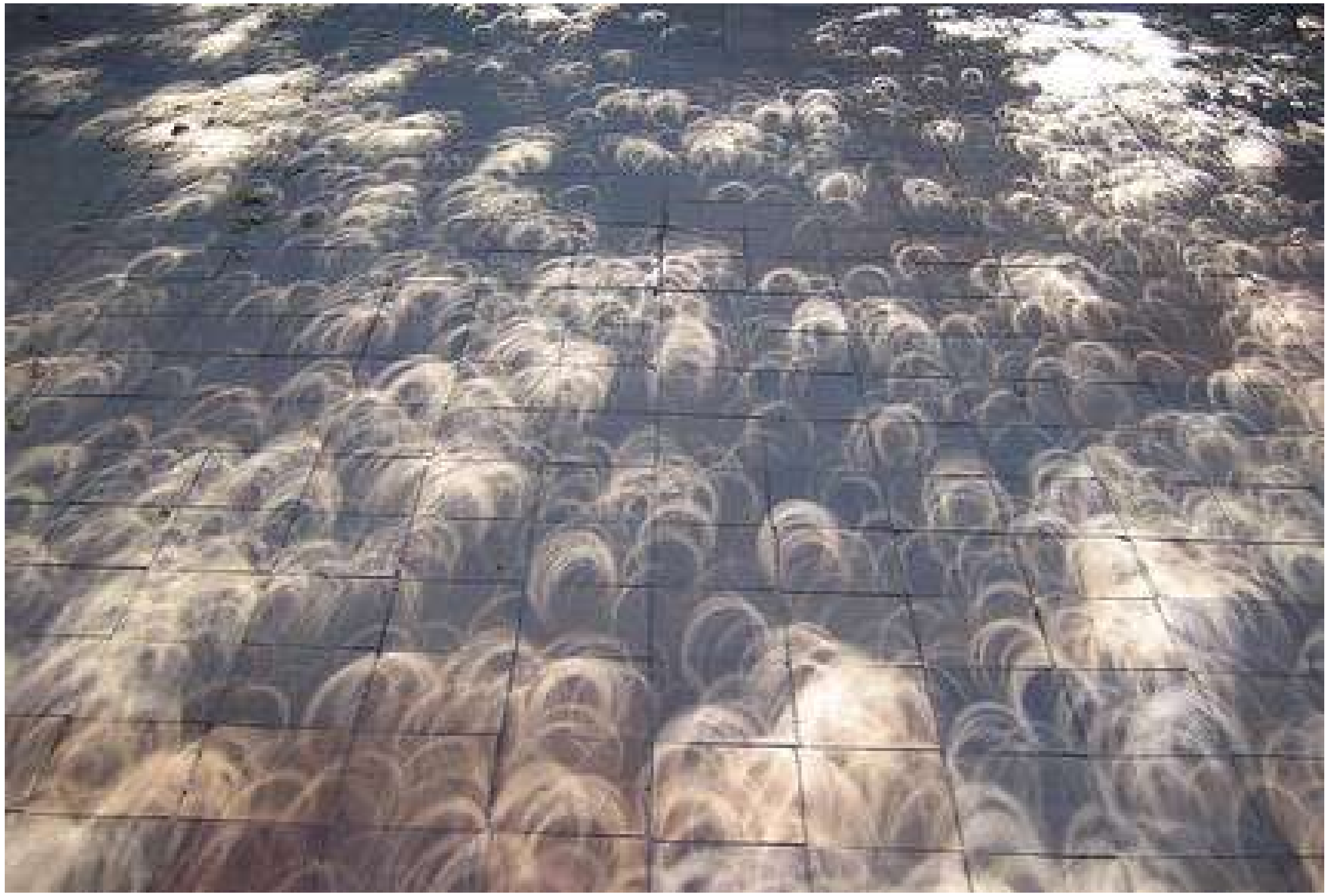
[Slide by Steve Seitz]

Pinhole cameras everywhere

Sun "shadows" during a solar eclipse
http://www.flickr.com/photos/73860948@N08/6678331997/

[Slide by Steve Seitz]

Pinhole cameras everywhere



Tree shadow during a solar eclipse

photo credit: Nils van der Burg http://www.physicstogo.org/index.cfm [Slide by Steve Seitz]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

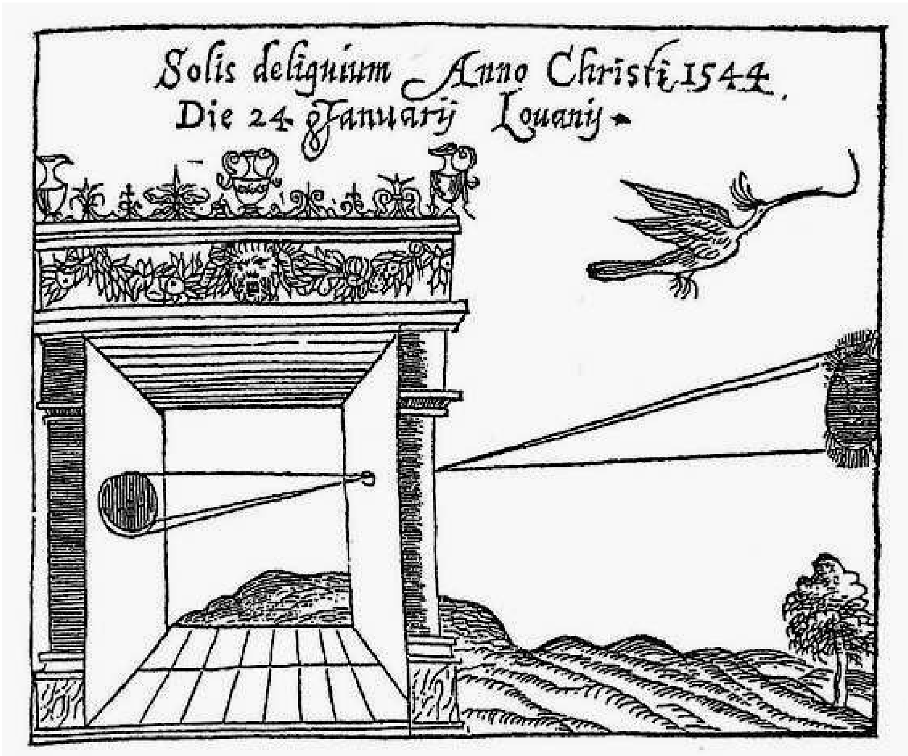
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Pinhole Camera - Camera Obscura

The first camera

- Known to Aristotle



Gemma Frisius, 1558

[Slide by Steve Seitz]

How to Turn a Room into a Camera Obscura

https://www.youtube.com/watch?v=hsXo4gD7iWI&ab_channel=GeorgeEastmanMuseum

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Home-made pinhole camera

Why so blurry?

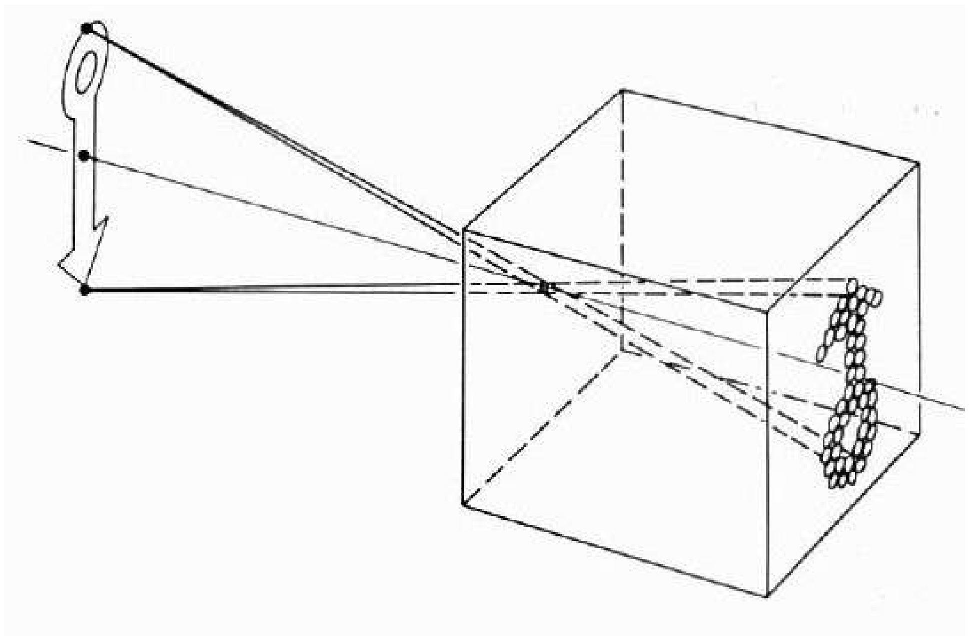
http://www.debevec.org/Pinhole/

[Slide by A. Efros]

Pinhole Camera - Camera Obscura

The first camera

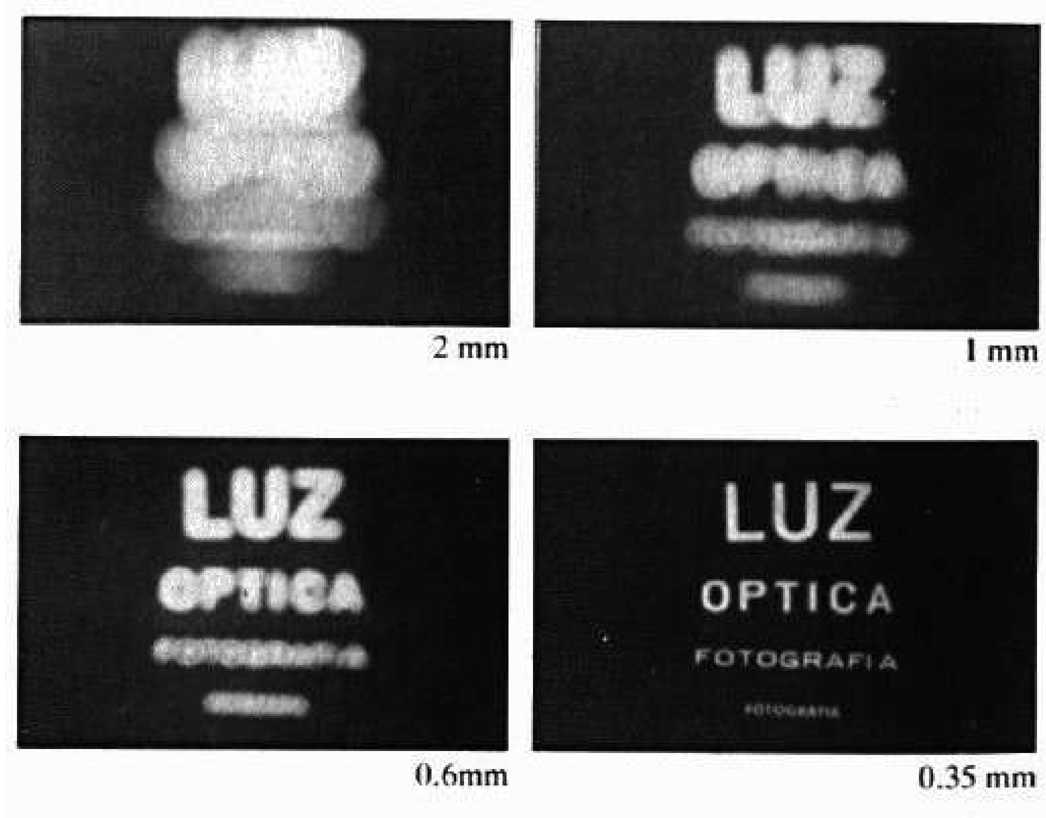
- Known to Aristotle
- How does the aperture size affect the image?



[Slide by Steve Seitz]

Shrinking the aperture

Why not make the aperture as small as possible?

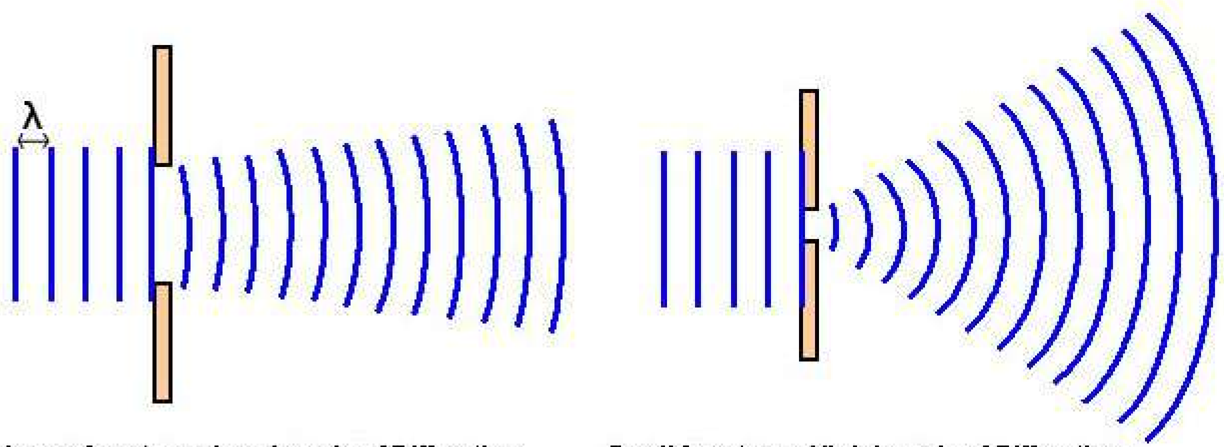


[Slide by Steve Seitz]

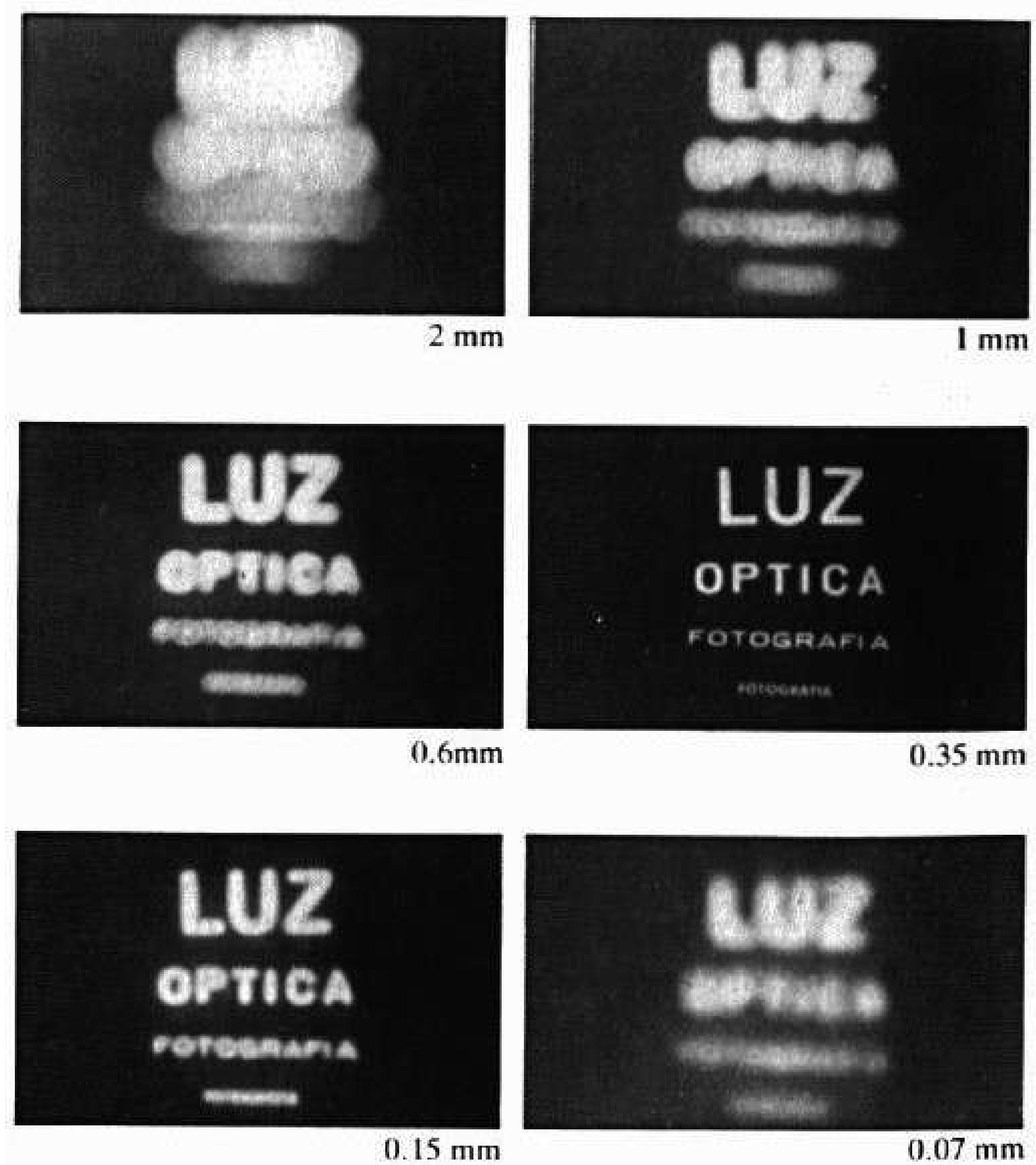
Shrinking the aperture

Why not make the aperture as small as possible?

- Less light gets through
- Diffraction effects



Large Aperture - Low Levels of Diffraction   Small Aperture - High Levels of Diffraction



[Slide by Steve Seitz]

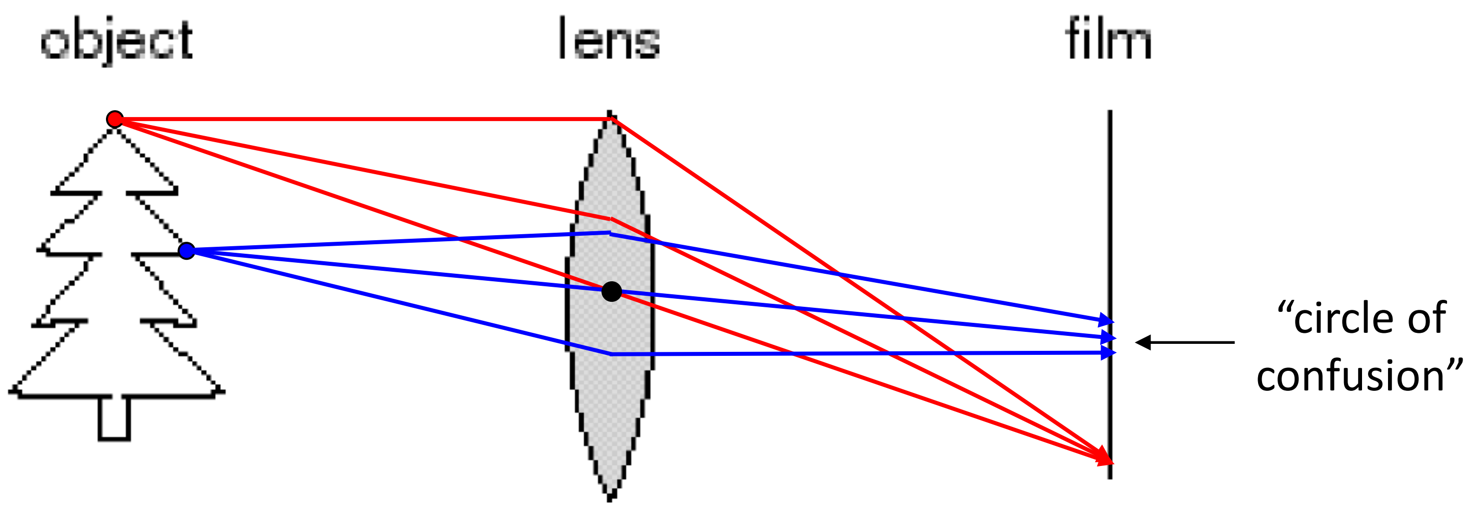
Today's Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Adding a lens



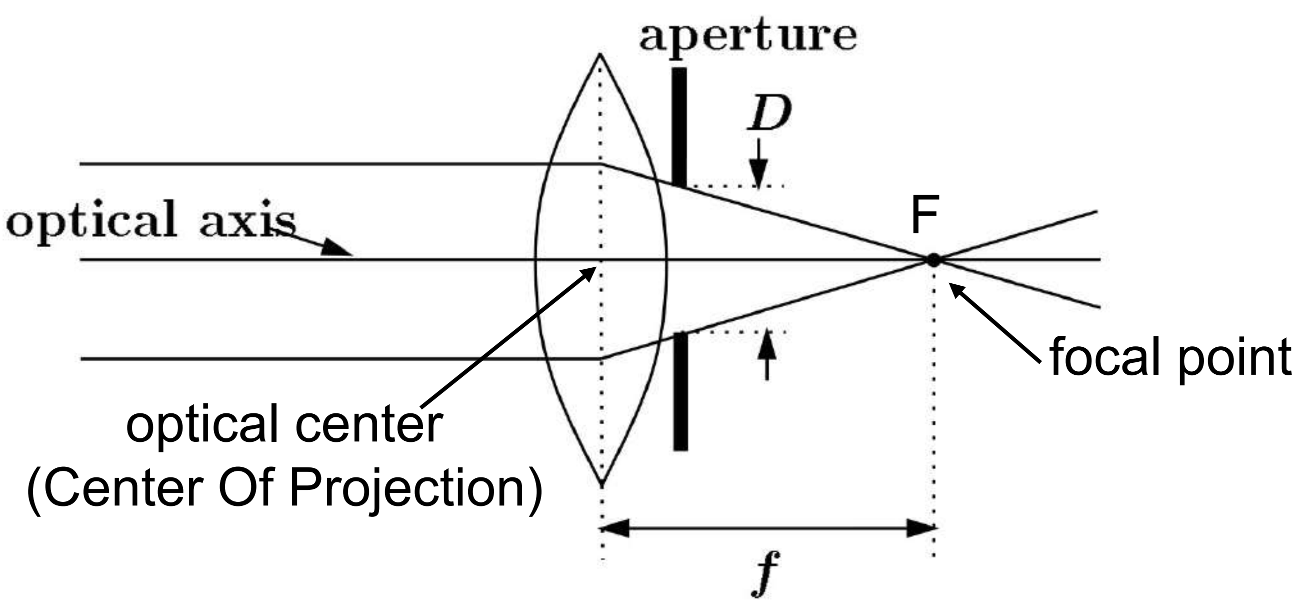
A lens focuses light onto the film

- Rays passing through the center are not deviated
- There is a specific distance at which objects are “in focus”
  - other points project to a “circle of confusion” in the image

[Slide by Steve Seitz]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Lenses



A lens focuses rays parallel to its axis onto a single **focal point**

- Focal point is on a plane located at a distance **f** (**focal length**) beyond the plane of the lens
  - f is a function of the shape and index of refraction of the lens
- Aperture of diameter **D** restricts the range of rays
  - aperture may be on either side of the lens
- Lenses are typically spherical (easier to produce)

[Slide by Steve Seitz]

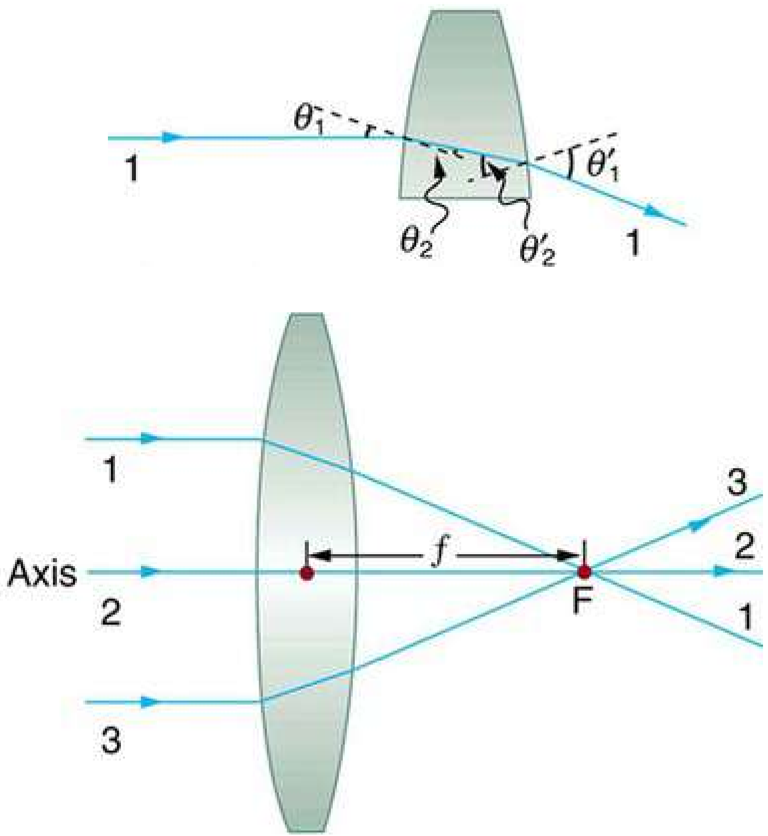
Today's Agenda - Overview of Cameras

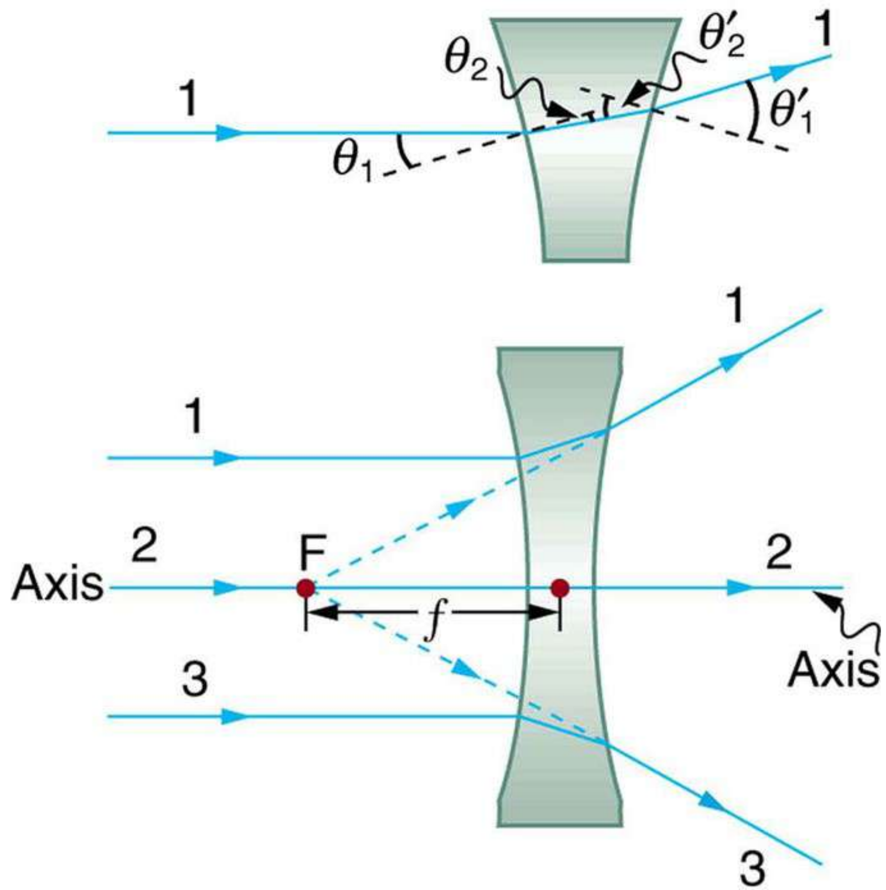
- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thin lenses



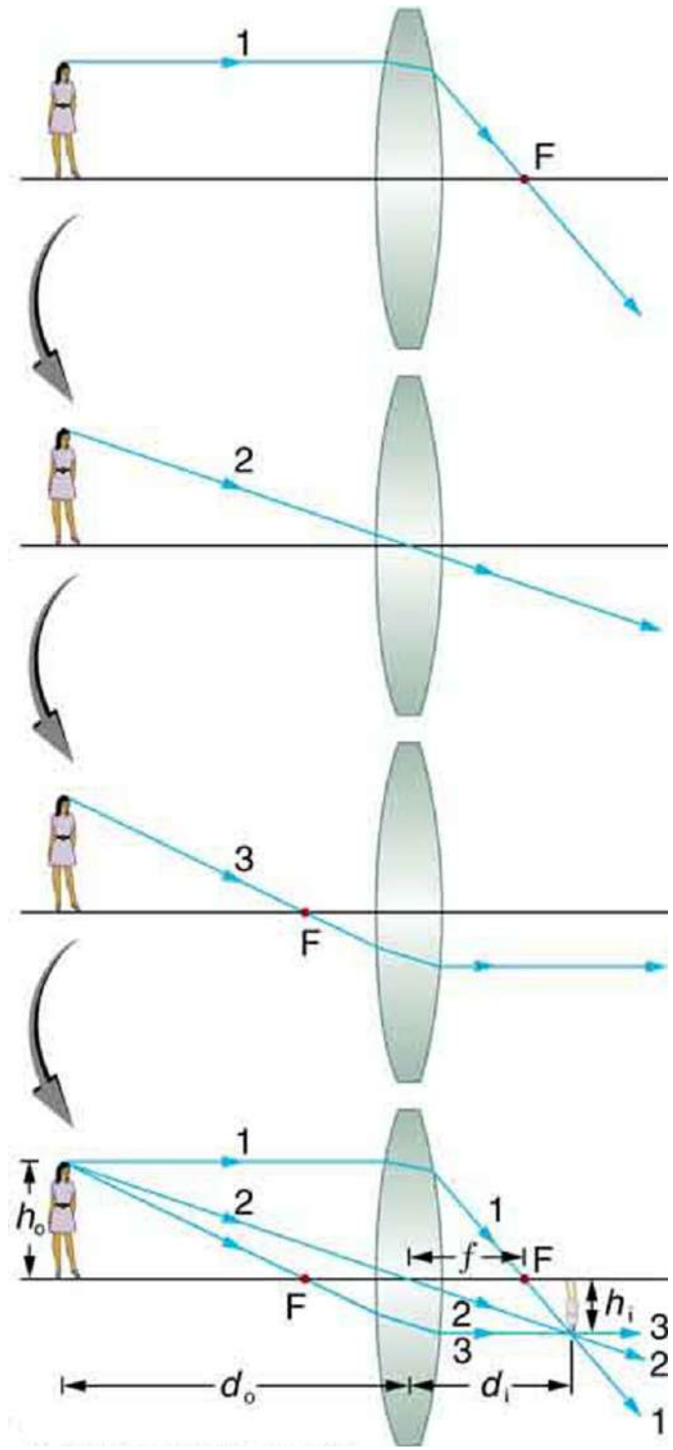


[See here](#)   Converging – Convex lens   Diverging – Concave lens

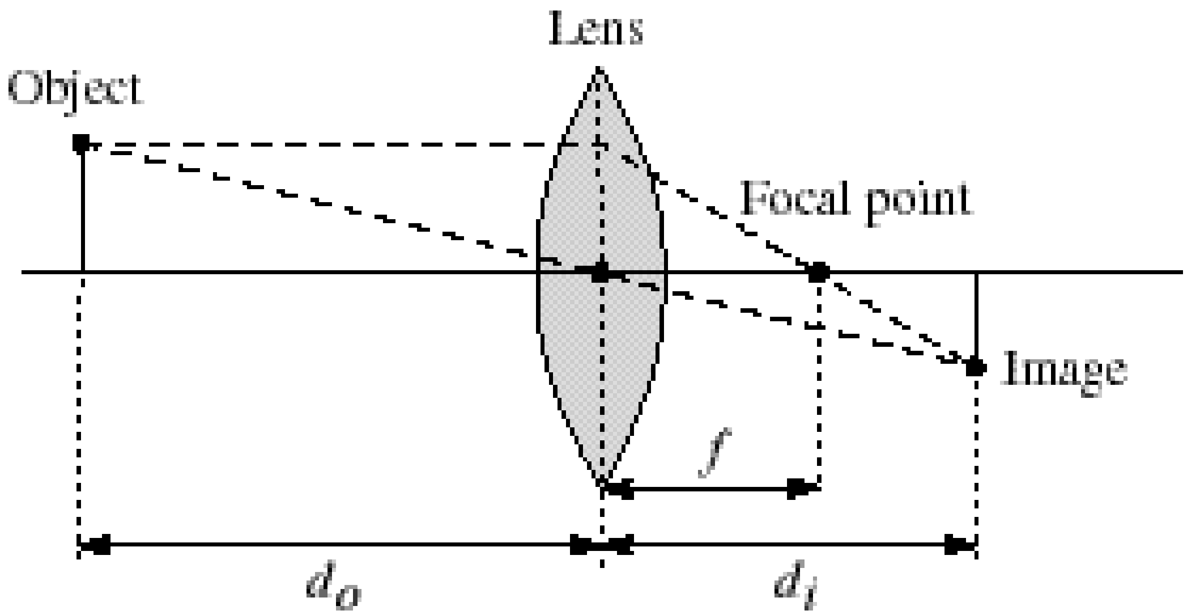
Thin lenses – ray tracing

- Five rules for image formation with thin lenses
  1. A ray entering a converging lens parallel to its axis passes through the focal point F of the lens on the other side.
  2. A ray entering a diverging lens parallel to its axis seems to come from the focal point F.
  3. A ray passing through the center of either a converging or a diverging lens does not change direction.
  4. A ray entering a converging lens through its focal point exits parallel to its axis.
  5. A ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis.

See here



Thin lens equation

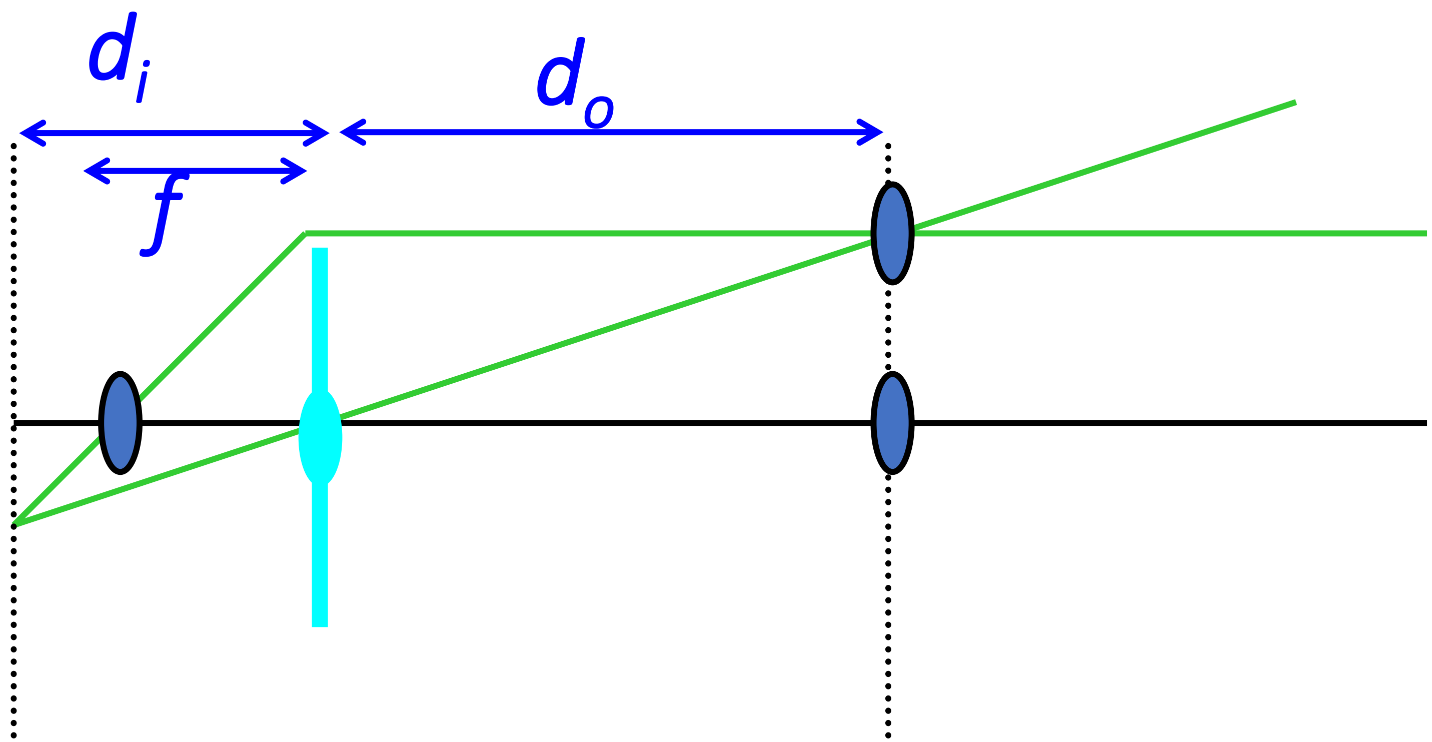


Thin lens equation: $$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

- Any object point satisfying this equation is in focus
- Thin lens applet (needs java player): http://www.phy.ntnu.edu.tw/java/Lens/lens_e.html (by Fu-Kwun Hwang)

[Slide by Steve Seitz]

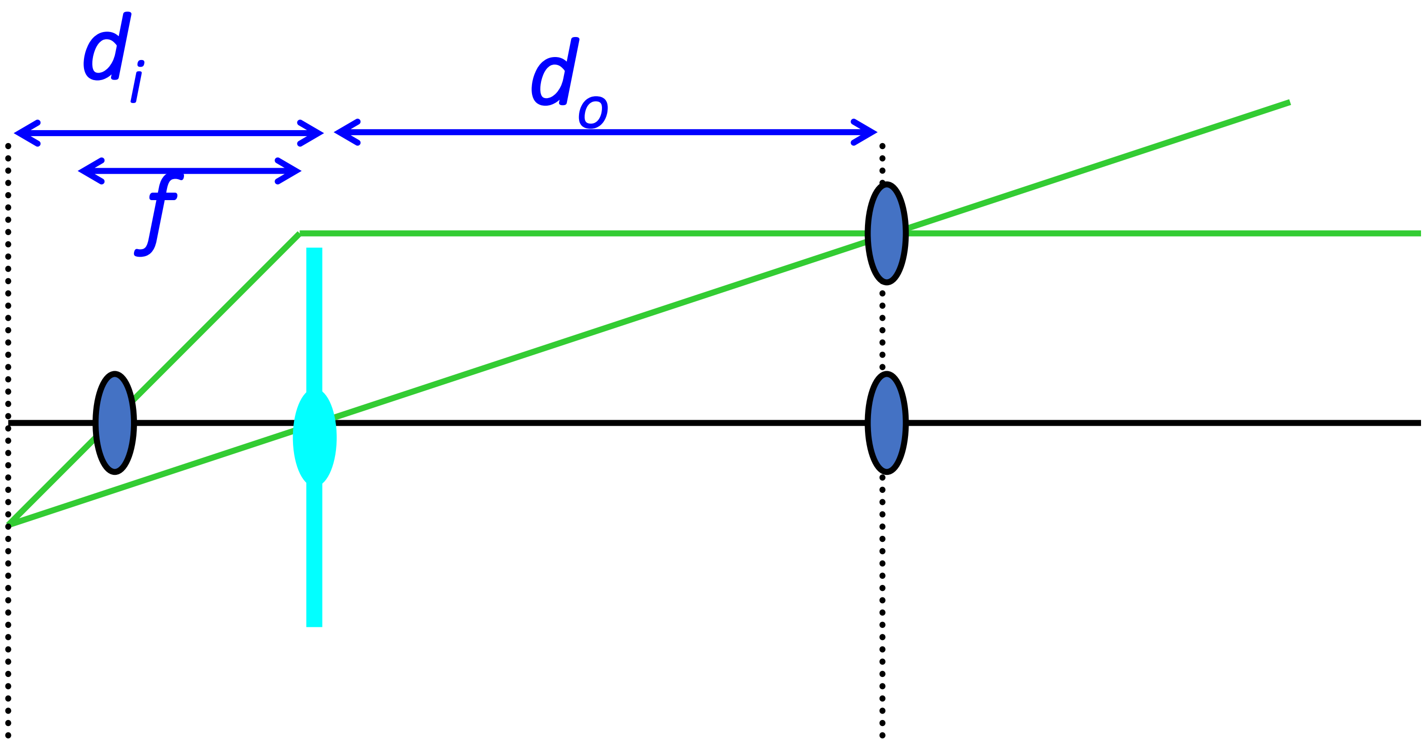
Thin lens formula



[Slide by Fredo Durand]

Thin lens formula

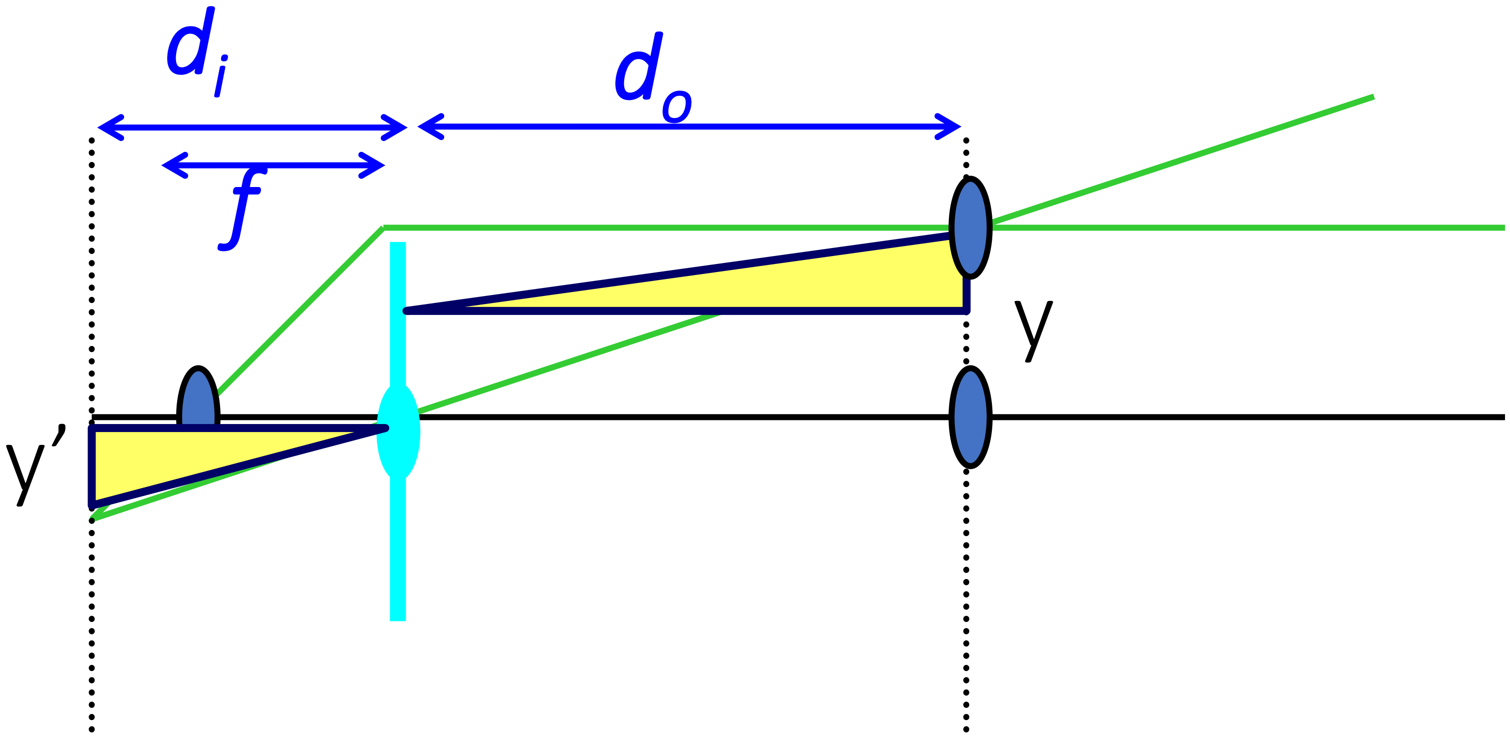
Use similar triangles!



[Slide by Fredo Durand]

Thin lens formula

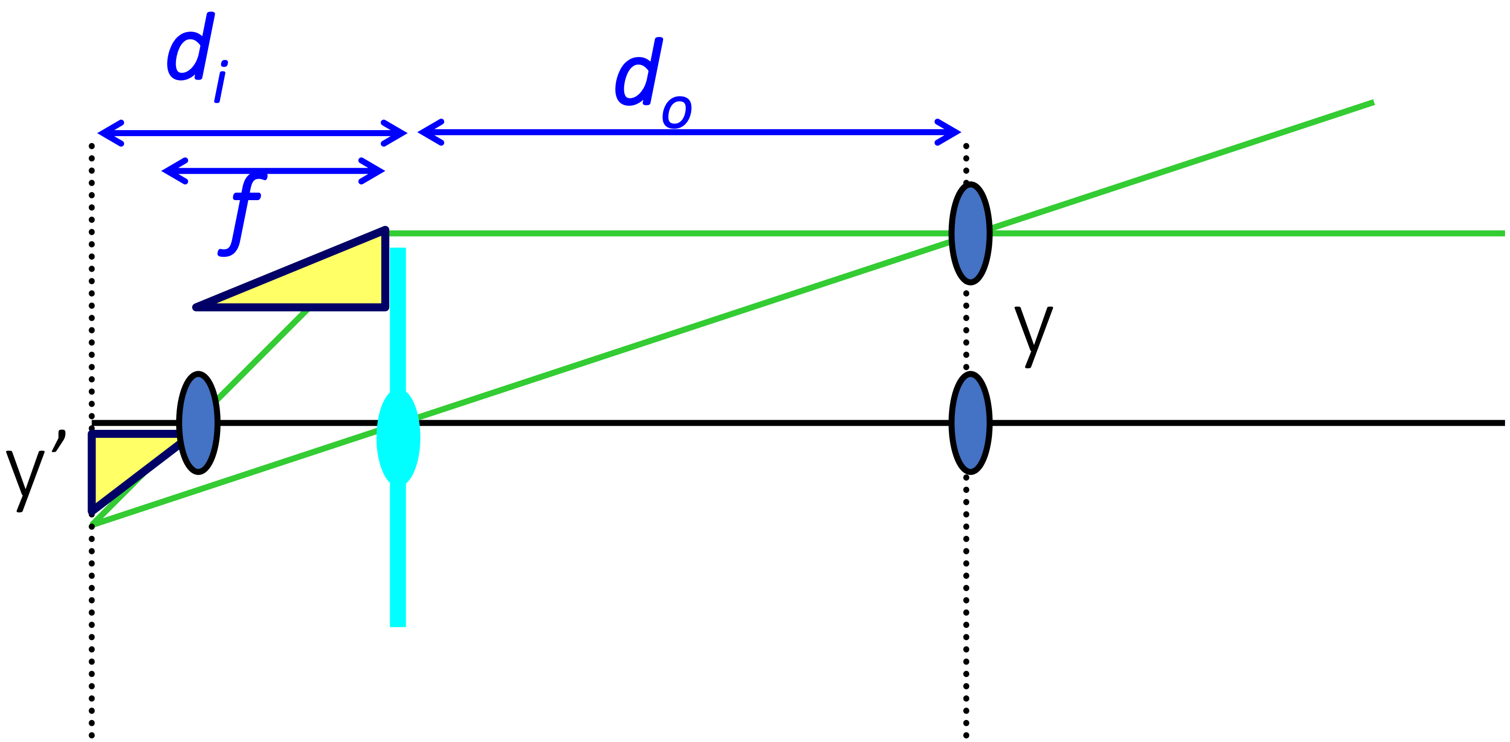
$y'/y = d_i/d_o$



[Slide by Fredo Durand]

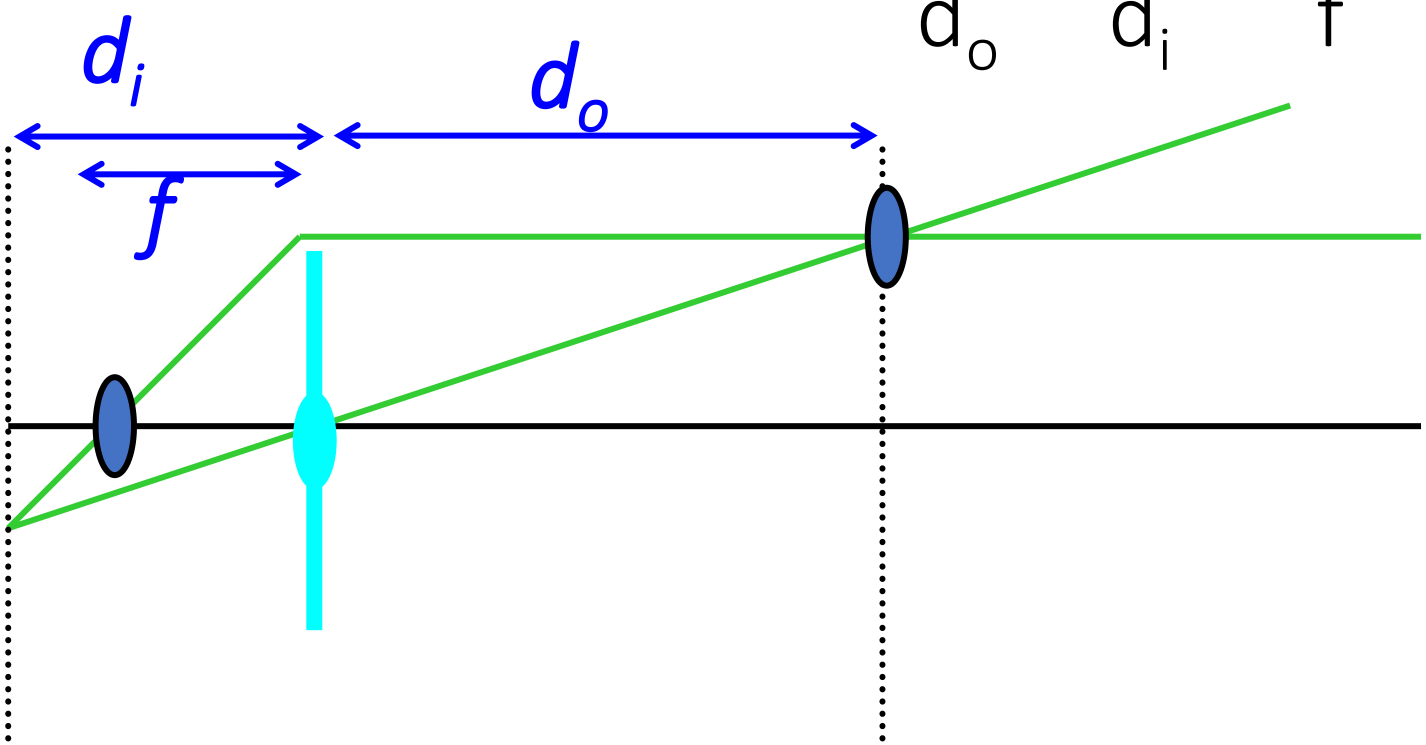
Thin lens formula

$$y'/y = d_i/d_o$$
$$y'/y = (d_i\text{-}f)/f$$



[Slide by Fredo Durand]

Thin lens formula



$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

Any point satisfying the thin lens equation is in focus.

[Slide by Fredo Durand]

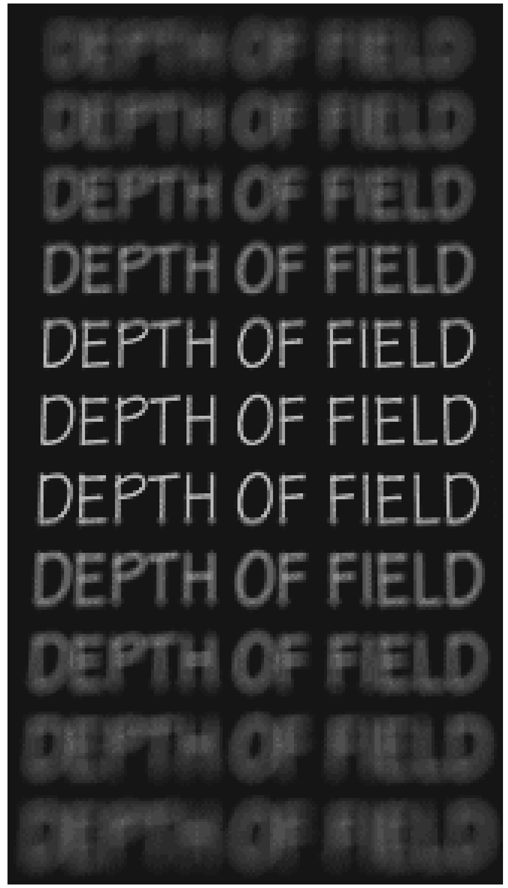
Today's Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Depth of field



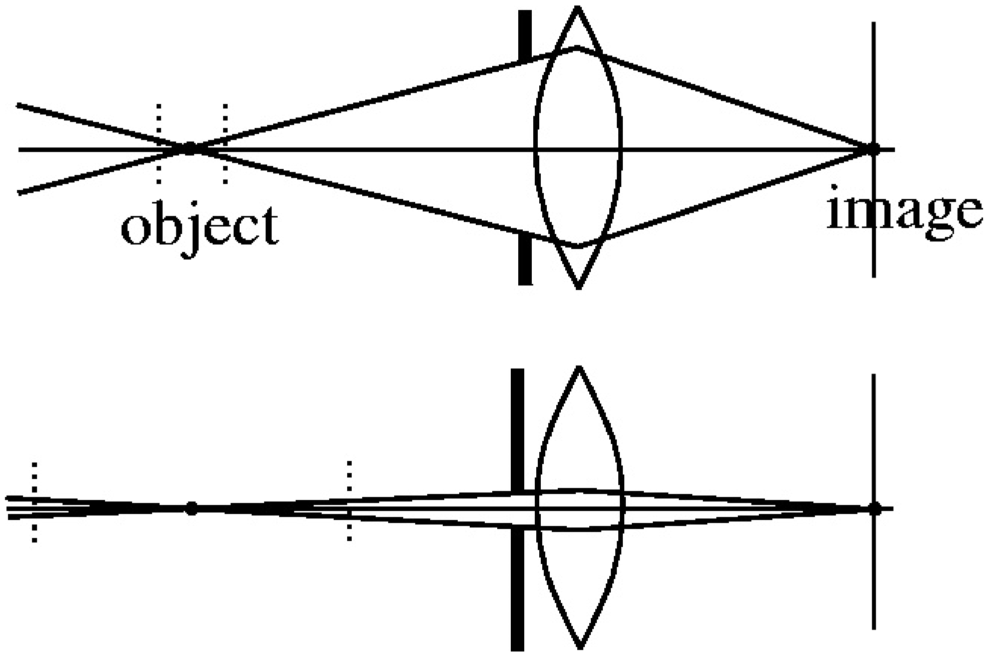
http://www.cambridgeincolour.com/tutorials/depth-of-field.htm

[Slide by A. Efros]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Depth of field




f / 5.6


f / 32

Changing the aperture size affects depth of field

- A smaller aperture increases the range in which the object is approximately in focus
- But small aperture reduces amount of light – need to increase exposure

Flower images from Wikipedia http://en.wikipedia.org/wiki/Depth_of_field

Depth of field

Large aperture = small DOF

Small aperture = large DOF

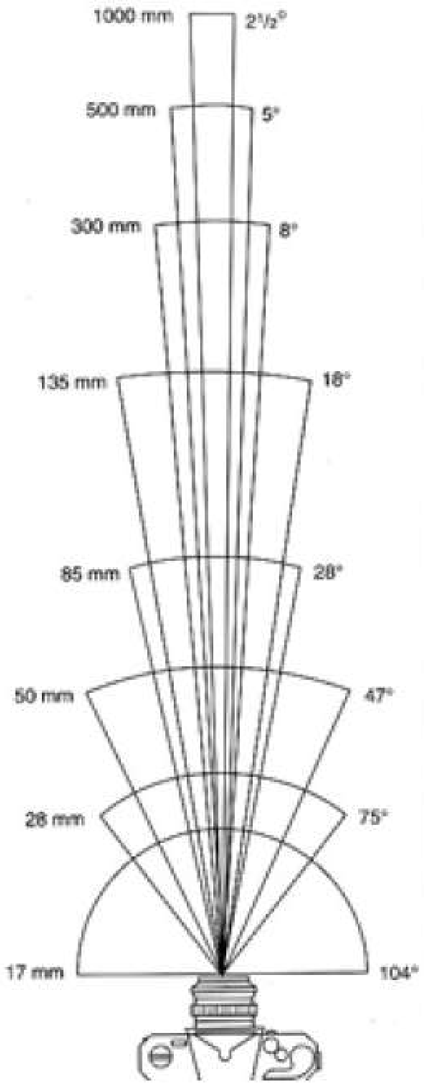
[Slide by A. Efros]

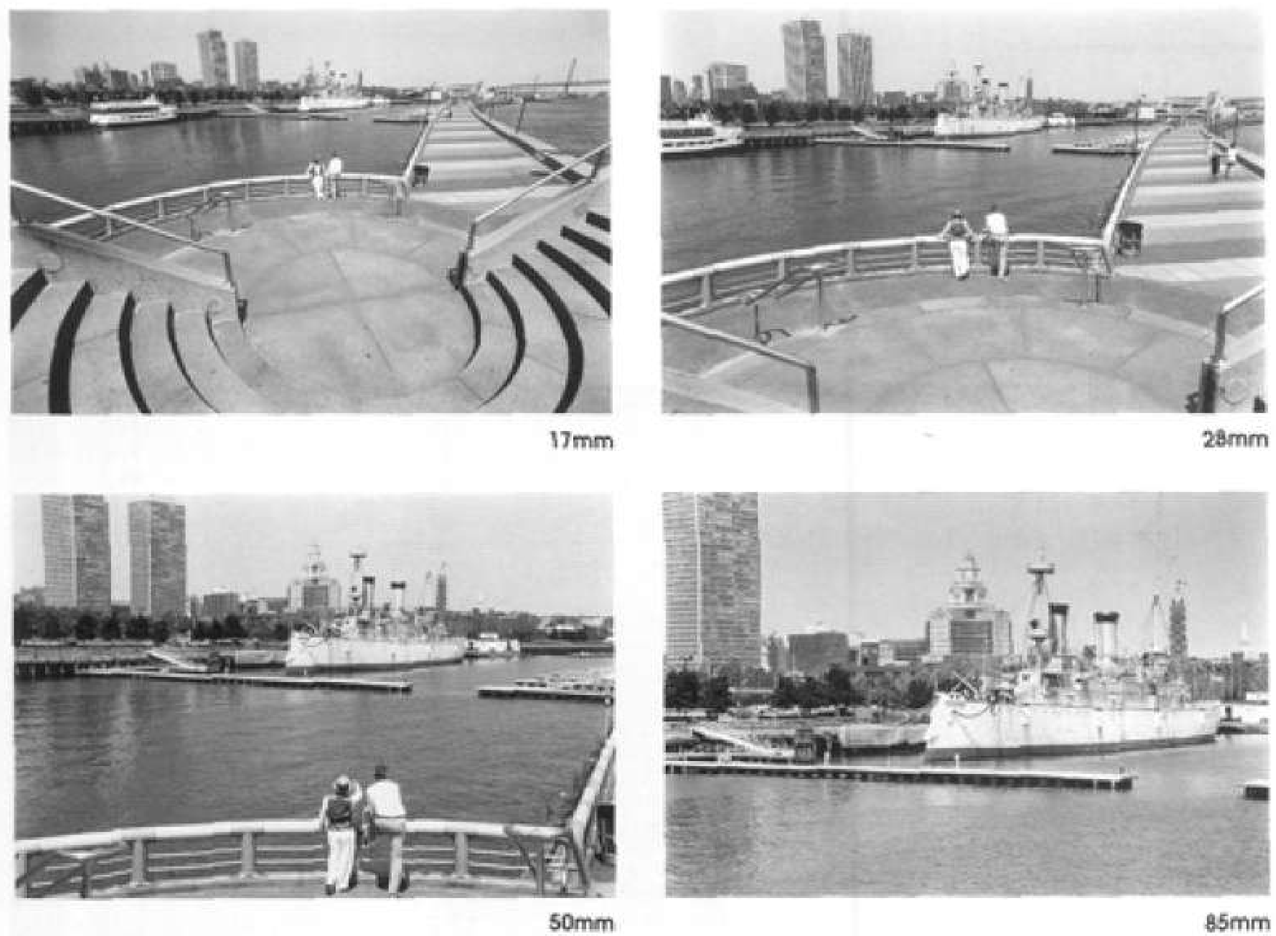
Today's Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Field of view

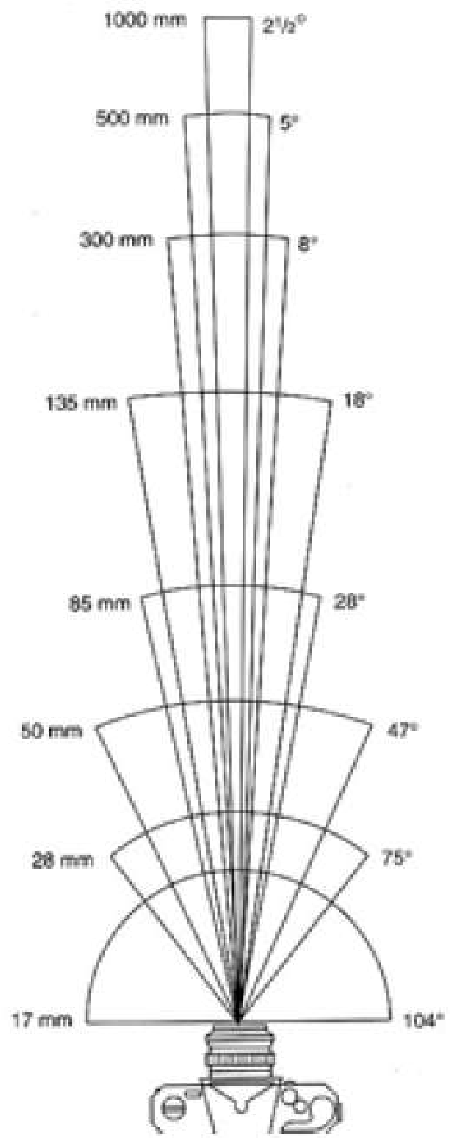


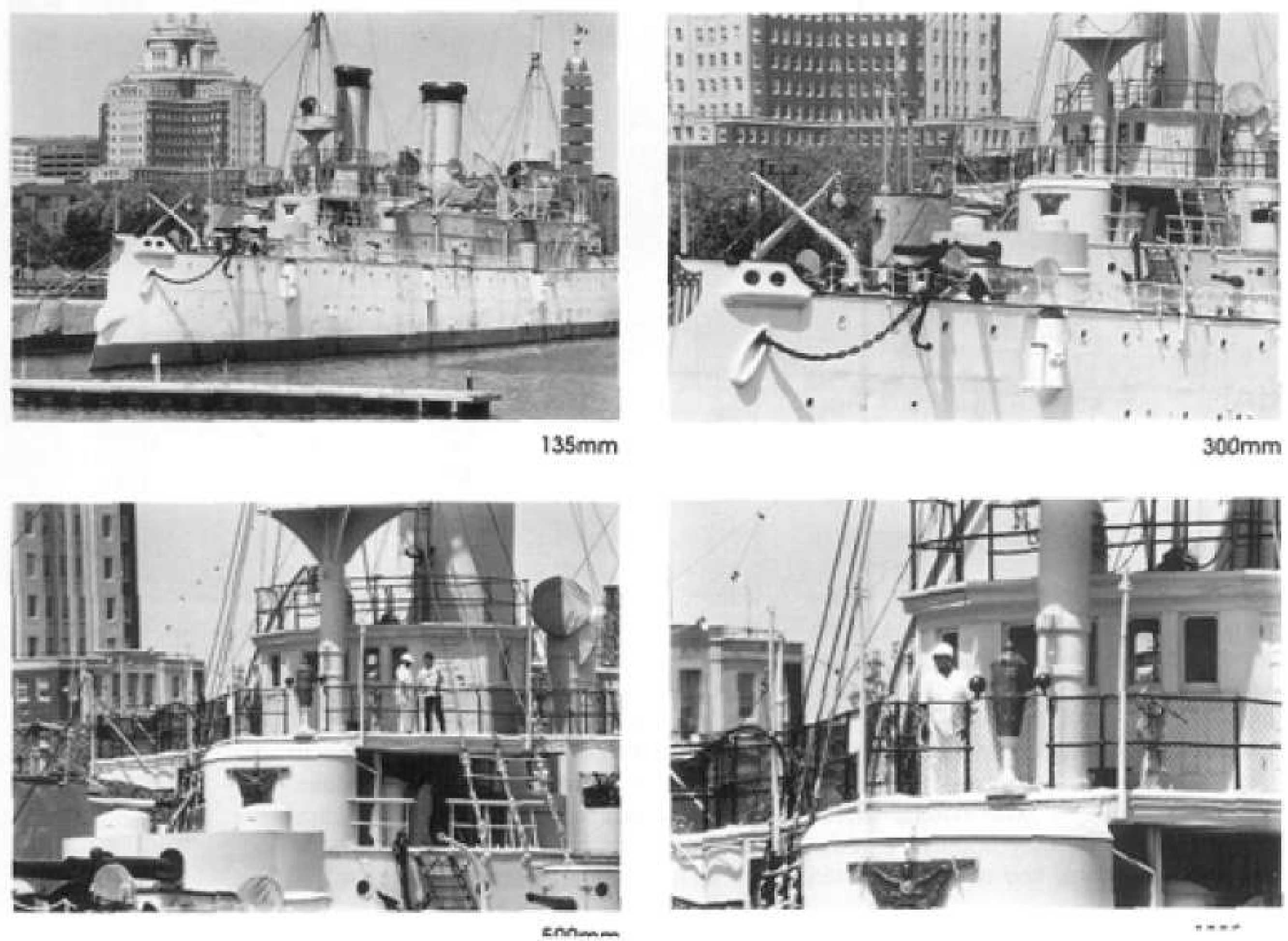


From London and Upton

[Slide by A. Efros]

Field of view



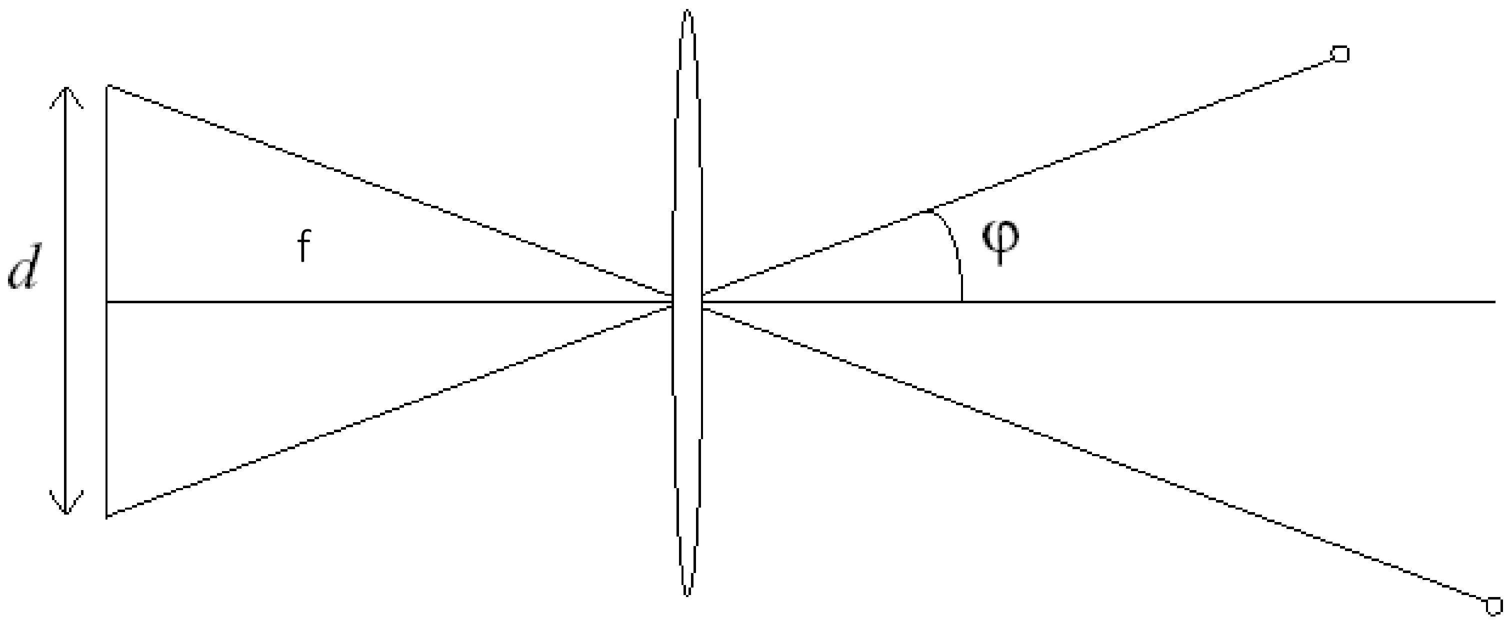


From London and Upton

[Slide by A. Efros]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Field of view



FOV depends on focal length f and size of the camera sensor size d

$$\varphi = \tan^{-1}\left(\frac{d}{2f}\right)$$

Smaller FOV = larger Focal Length

[Slide by A. Efros]

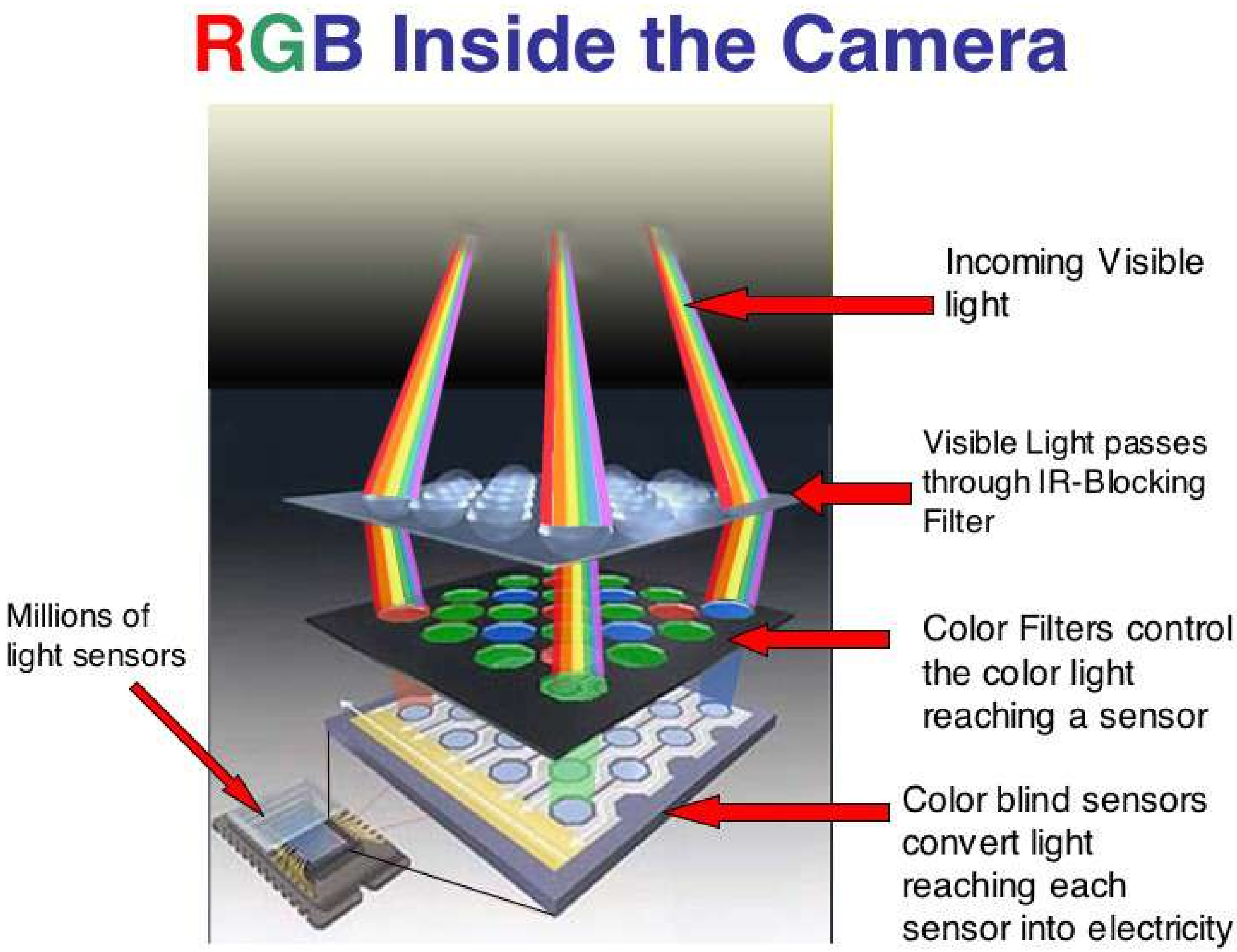
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

From light to pixels

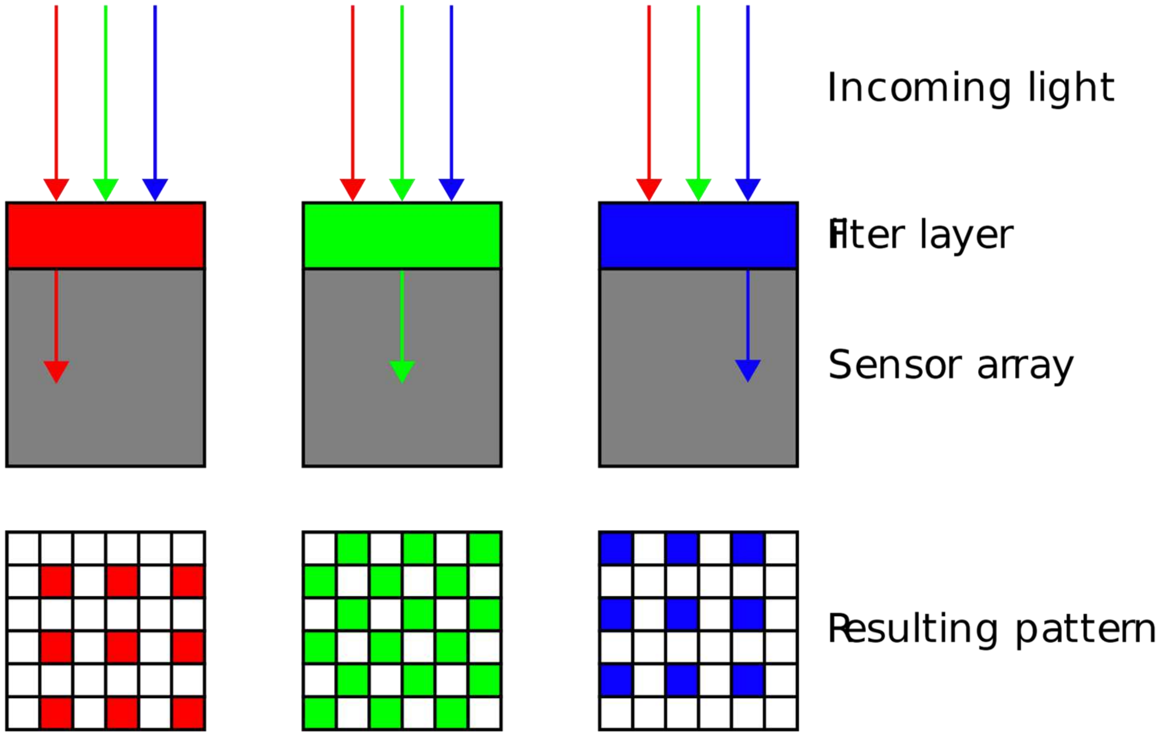


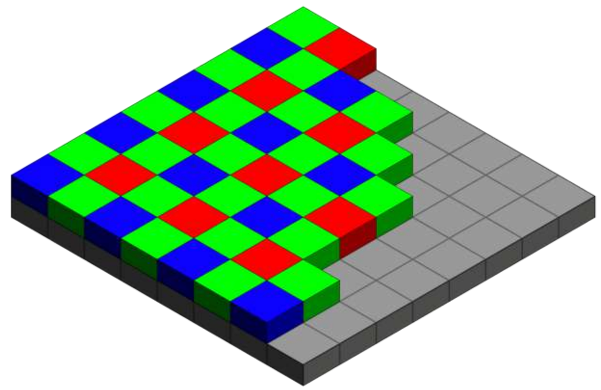
https://sites.google.com/a/globalsystemsscience.org/digital-earth-watch/tools/digital-cameras/light-entering-a-camera

[Slide by Steve Seitz]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bayer filters





https://en.wikipedia.org/wiki/Bayer_filter

¼ of pixels see red light (e.g.)

- Q: how do you get red at every pixel?
- A: Need to interpolate -- called *debayering*

[Slide by Steve Seitz]

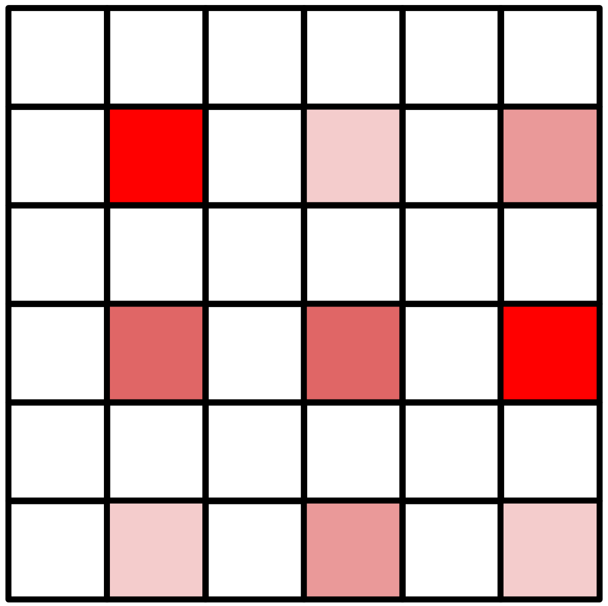
Today's Agenda - Overview of Cameras

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Debayering



	100		10		30
	50		50		100
	10		30		10

¼ of pixels see red light (e.g.)

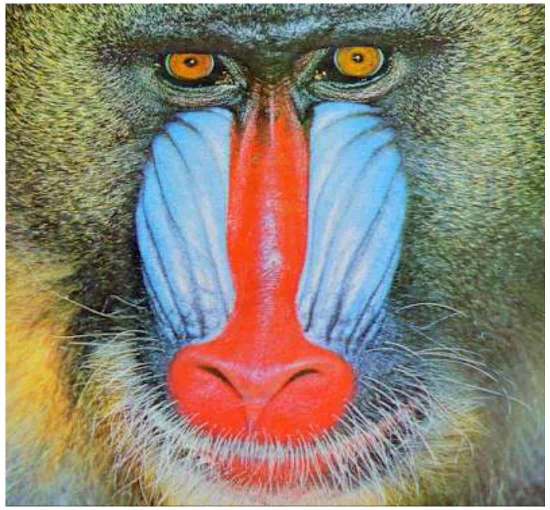
- Q: how do you get red at every pixel?
- A: Need to interpolate -- called *debayering*

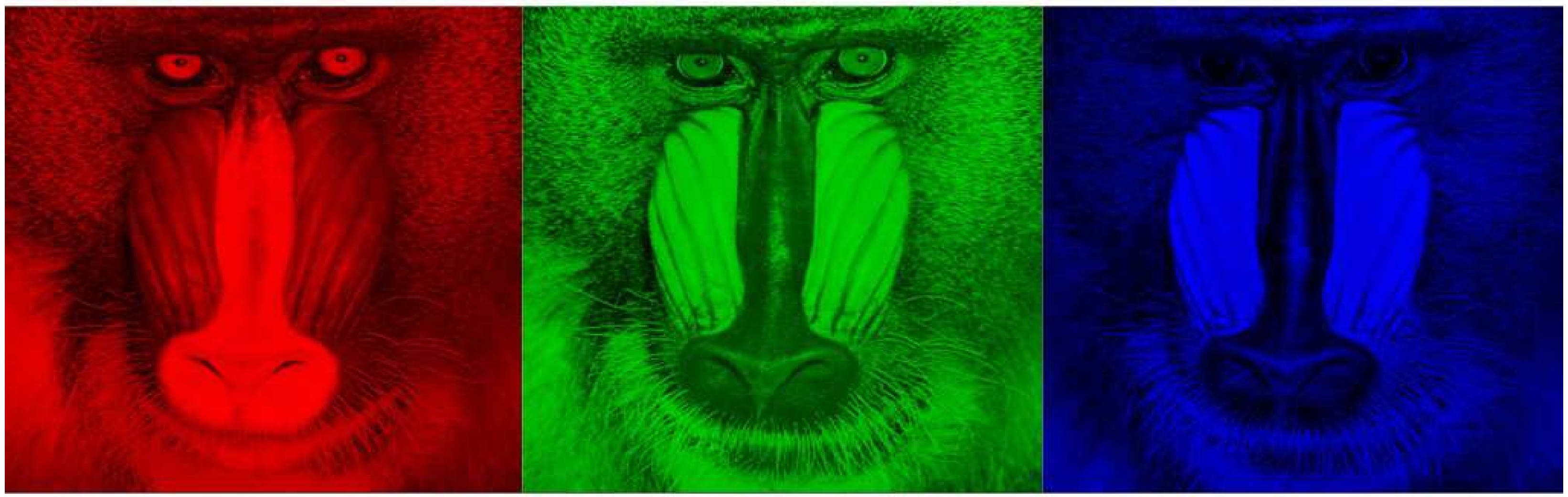
[Slide by Steve Seitz]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

RGB images (three channel)

What we see



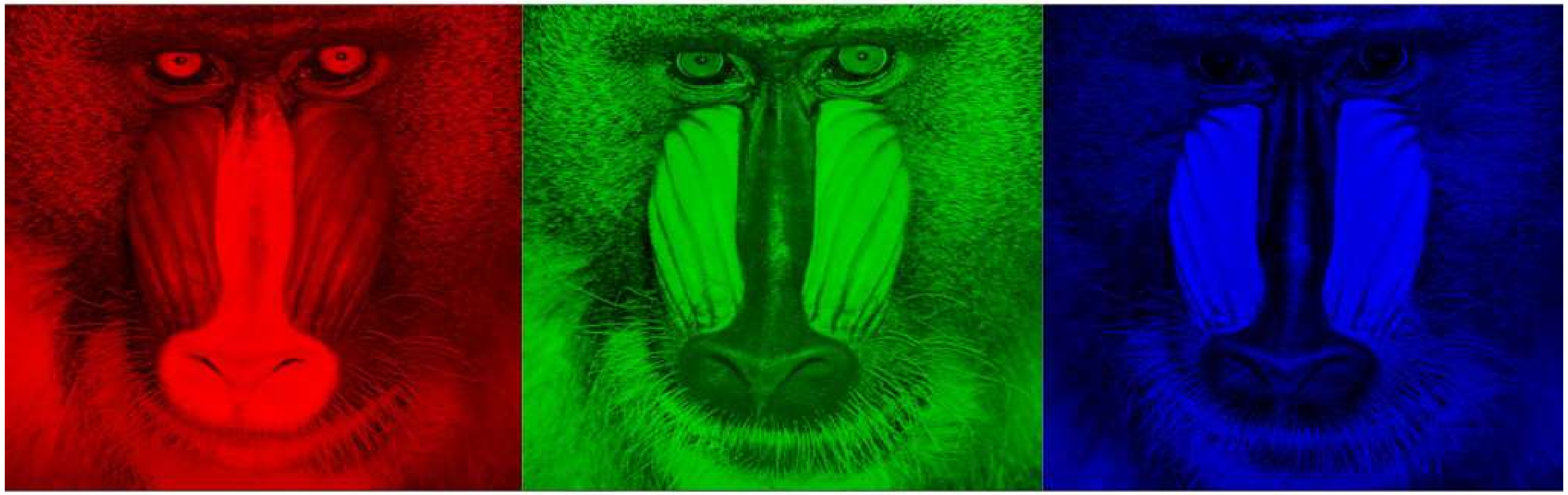


What we get out of the camera

[Slide by Steve Seitz]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

From now on: what to do with these RGB images



[Slide by Steve Seitz]

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 4: Interpolation – Resizing

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union

Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Pinhole Camera model
  - Aperture
  - Camera Obscura
- Cameras with lenses
  - Thin lens equation
  - Depth of field
  - Field of view
- Digital cameras
  - Bayer filters
  - Debayering

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

[material based on Joseph Redmon’s course]

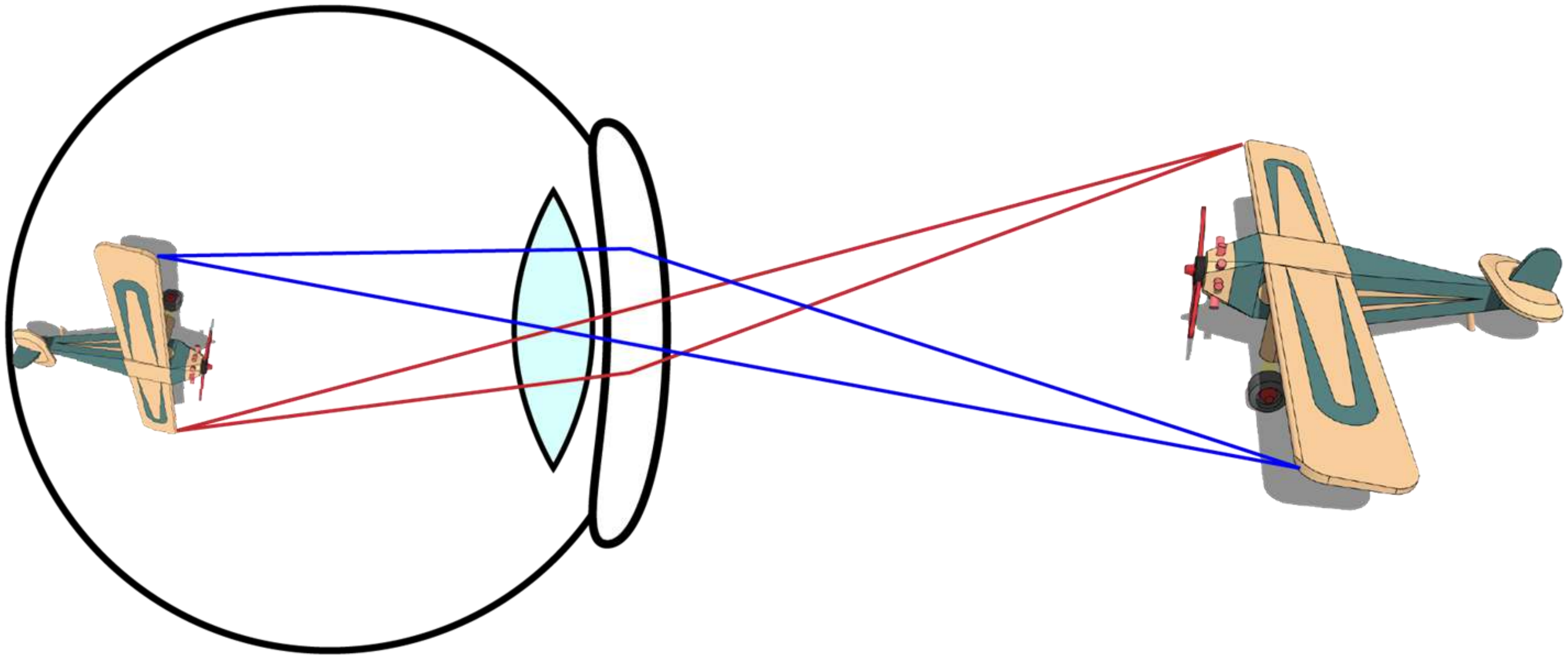
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

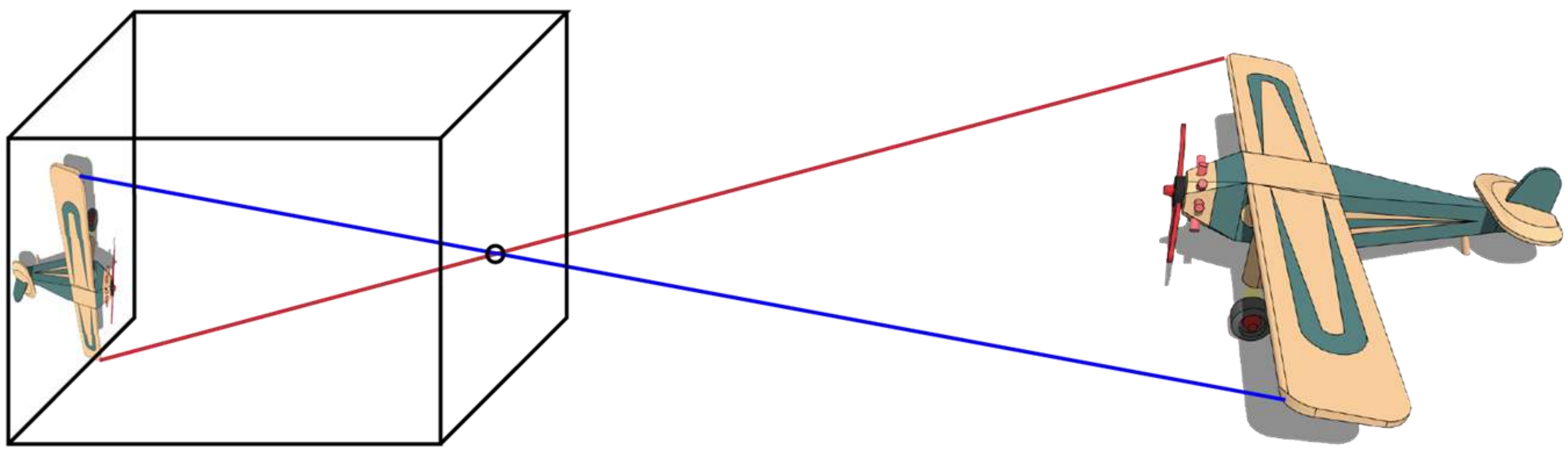
Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Eyes: projection onto retina



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

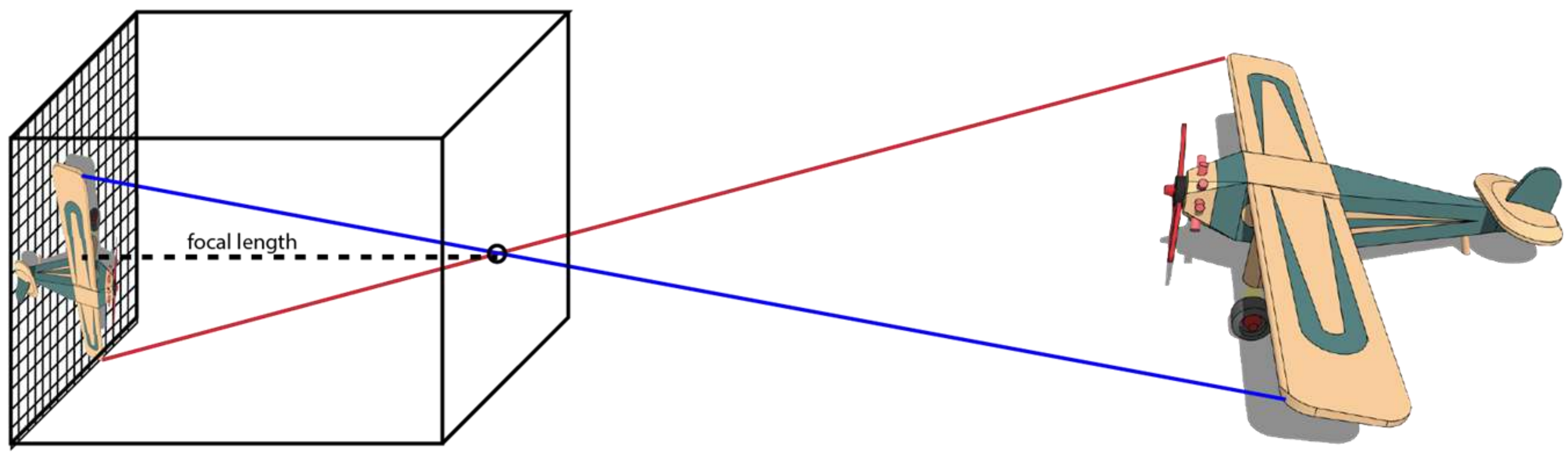
Model: pinhole camera



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

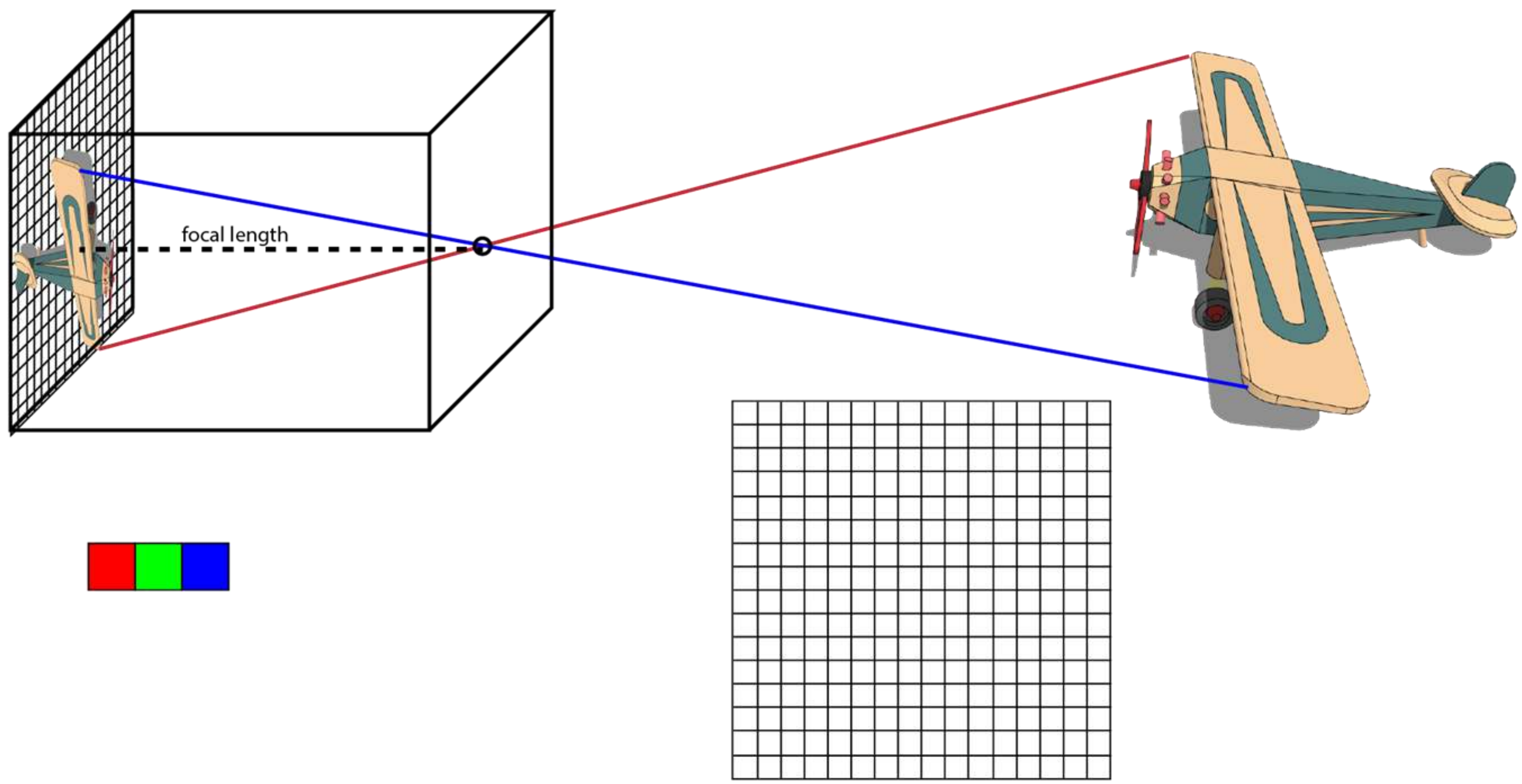
At each point we record incident light



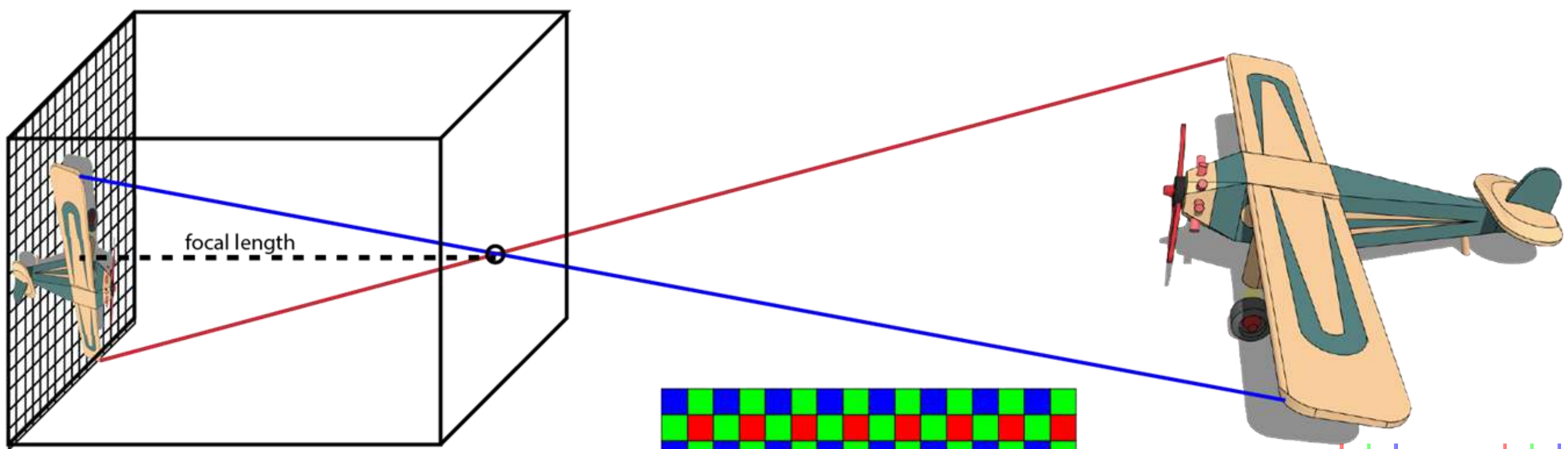
Co-financed by the European Union
Connecting Europe Facility

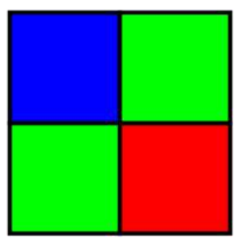
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

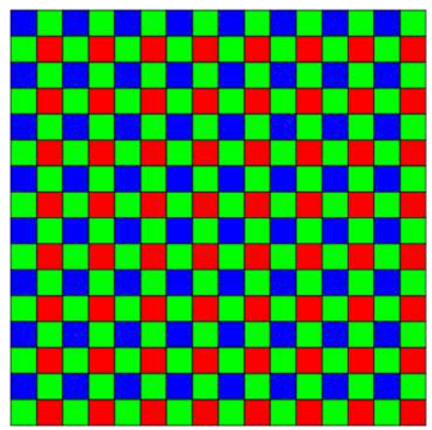
How do we record color?

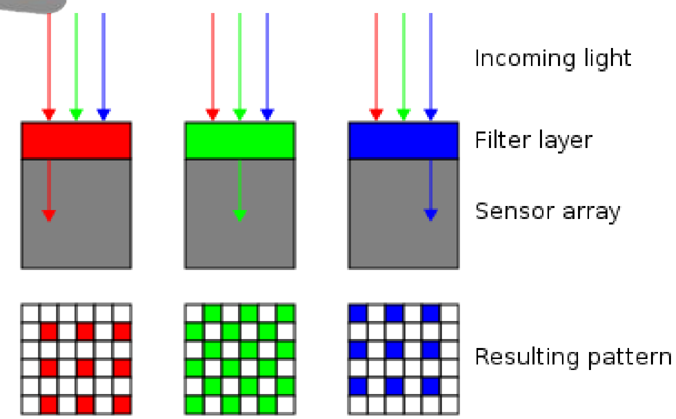


Bayer pattern for CMOS sensors





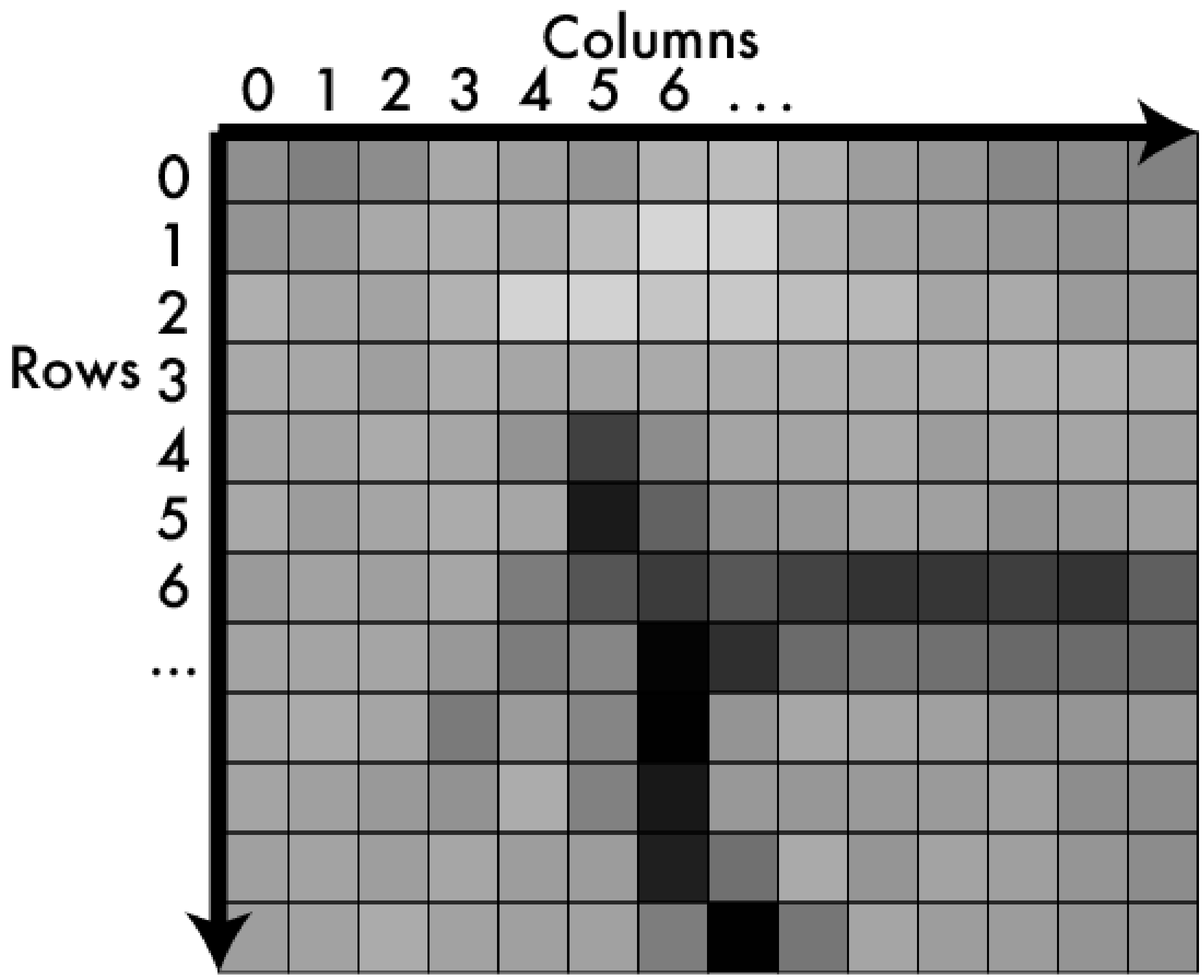




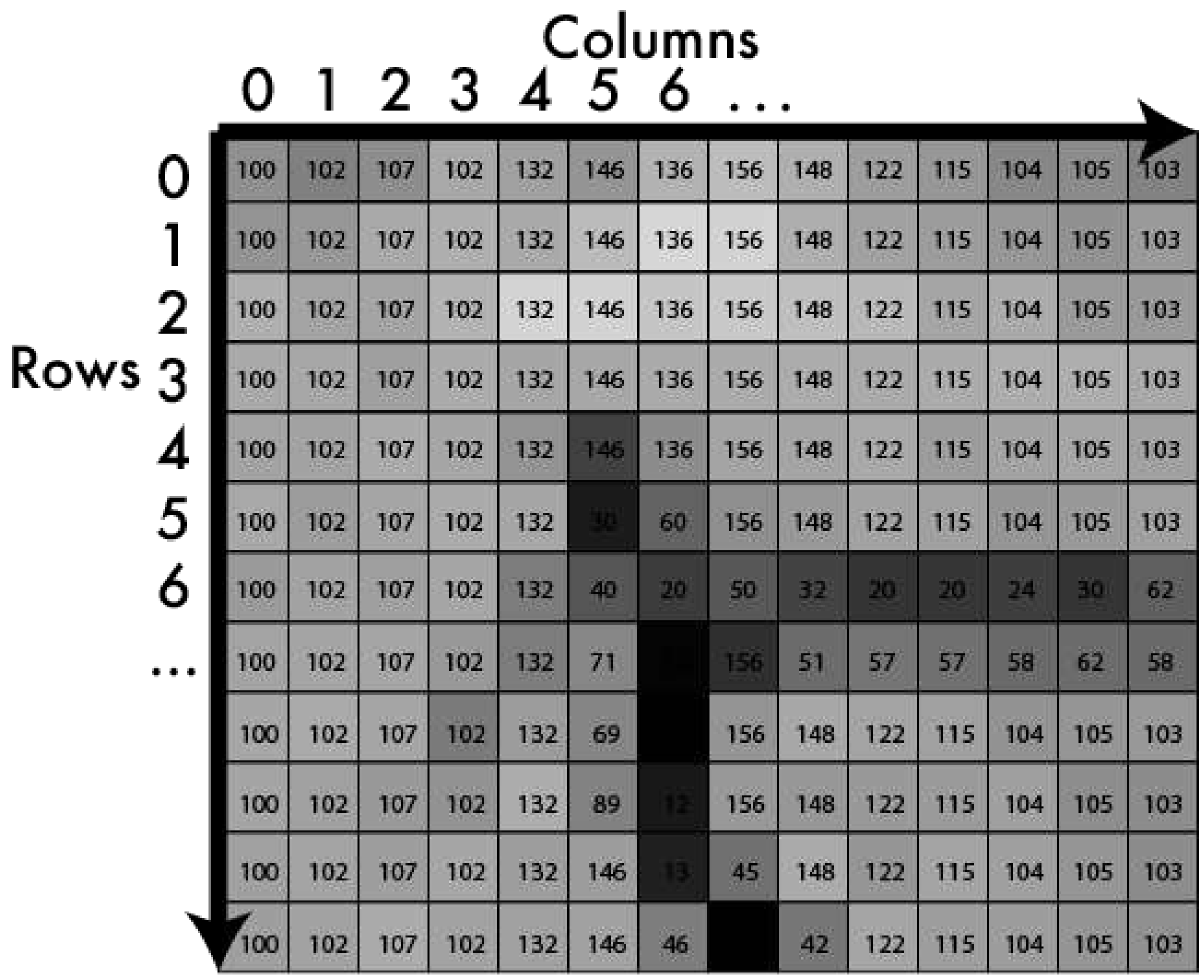
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

An image is a matrix of light



Values in matrix = how much light

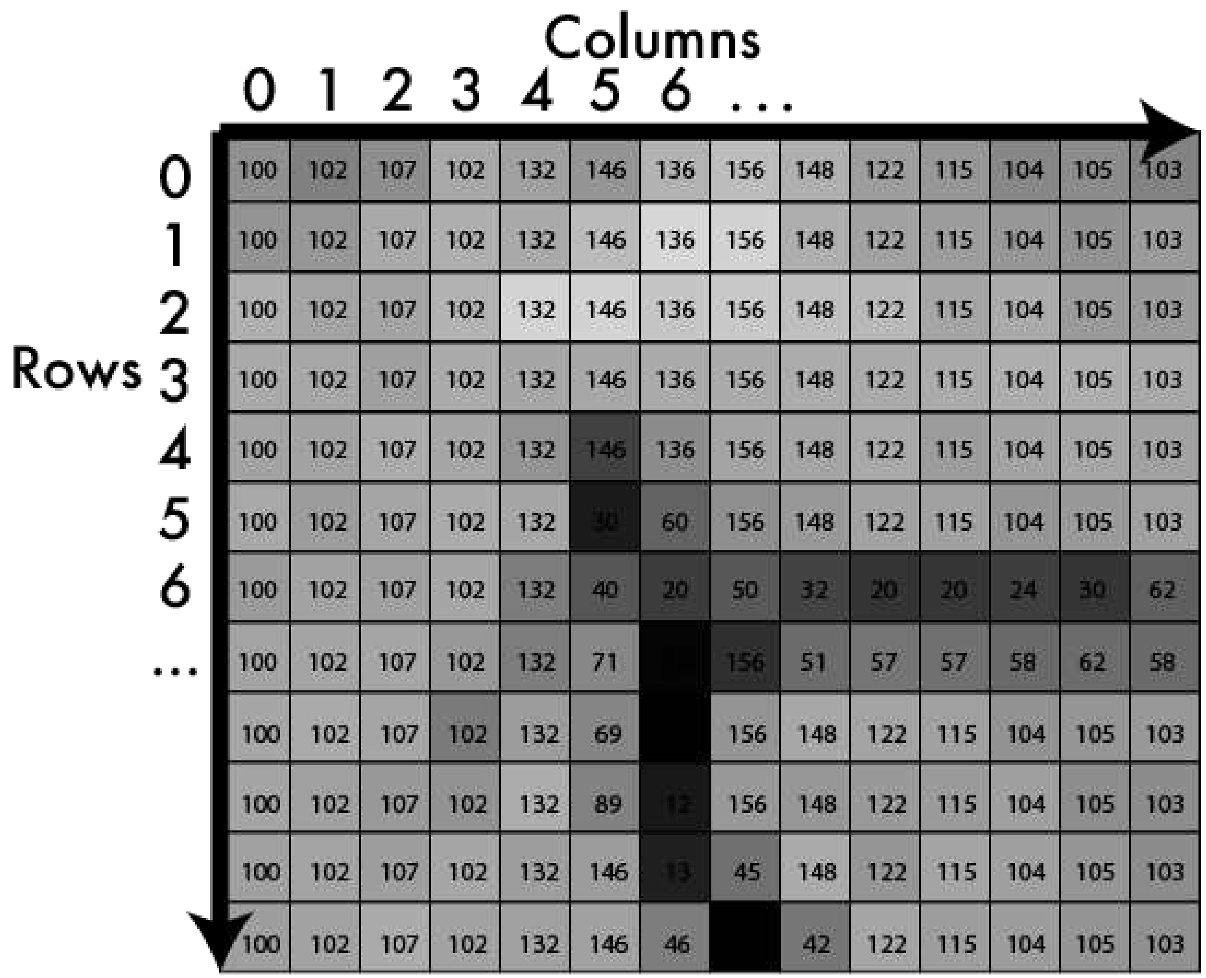


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Values in matrix = how much light

- Higher = more light
- Lower = less light
- Bounded
    - No light = 0
    - Sensor/device limit = max
    - Typical ranges:
        - [0-255], fit into byte
        - [0-1], floating point
- Called pixels

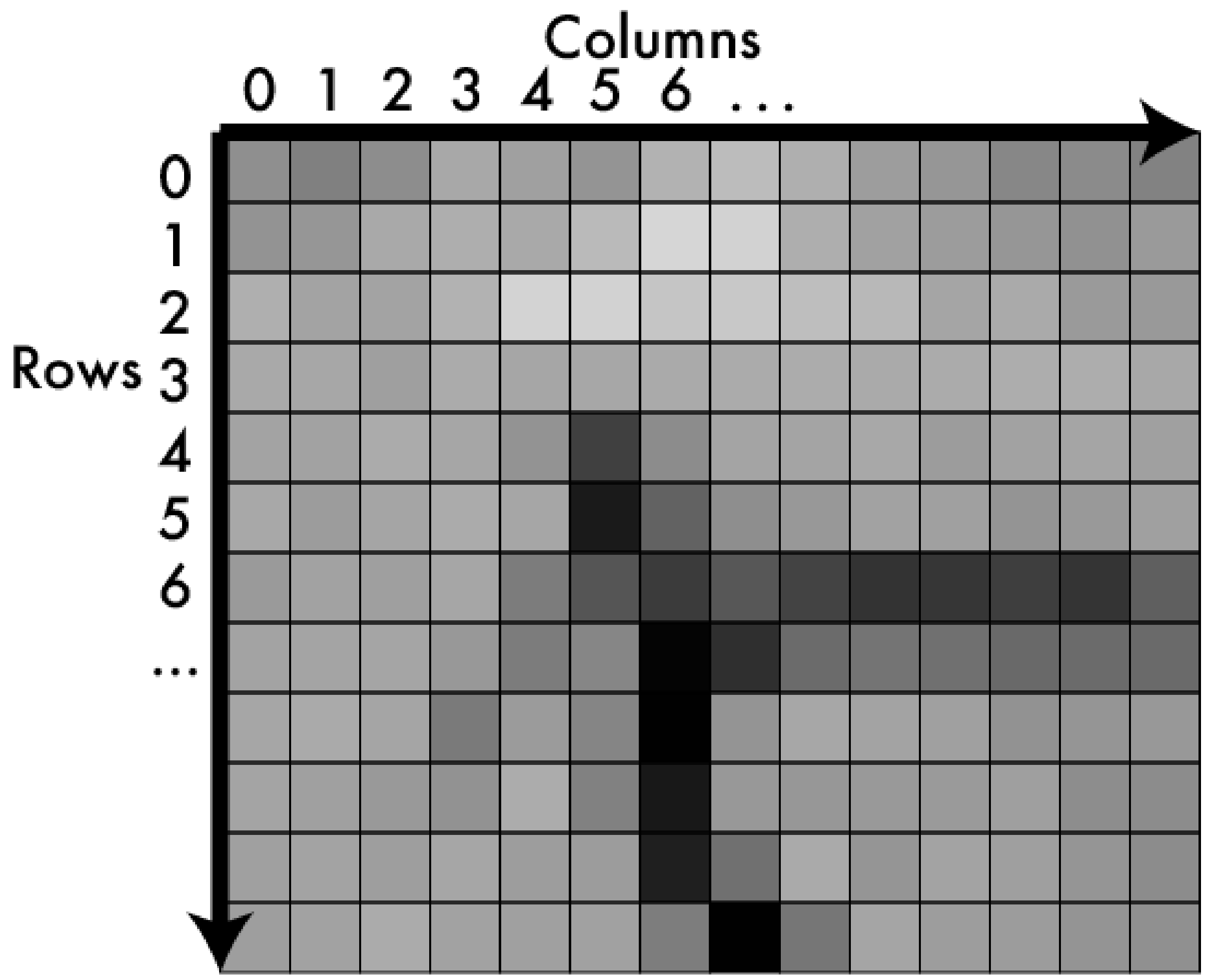


Co-financed by the European Union
Connecting Europe Facility

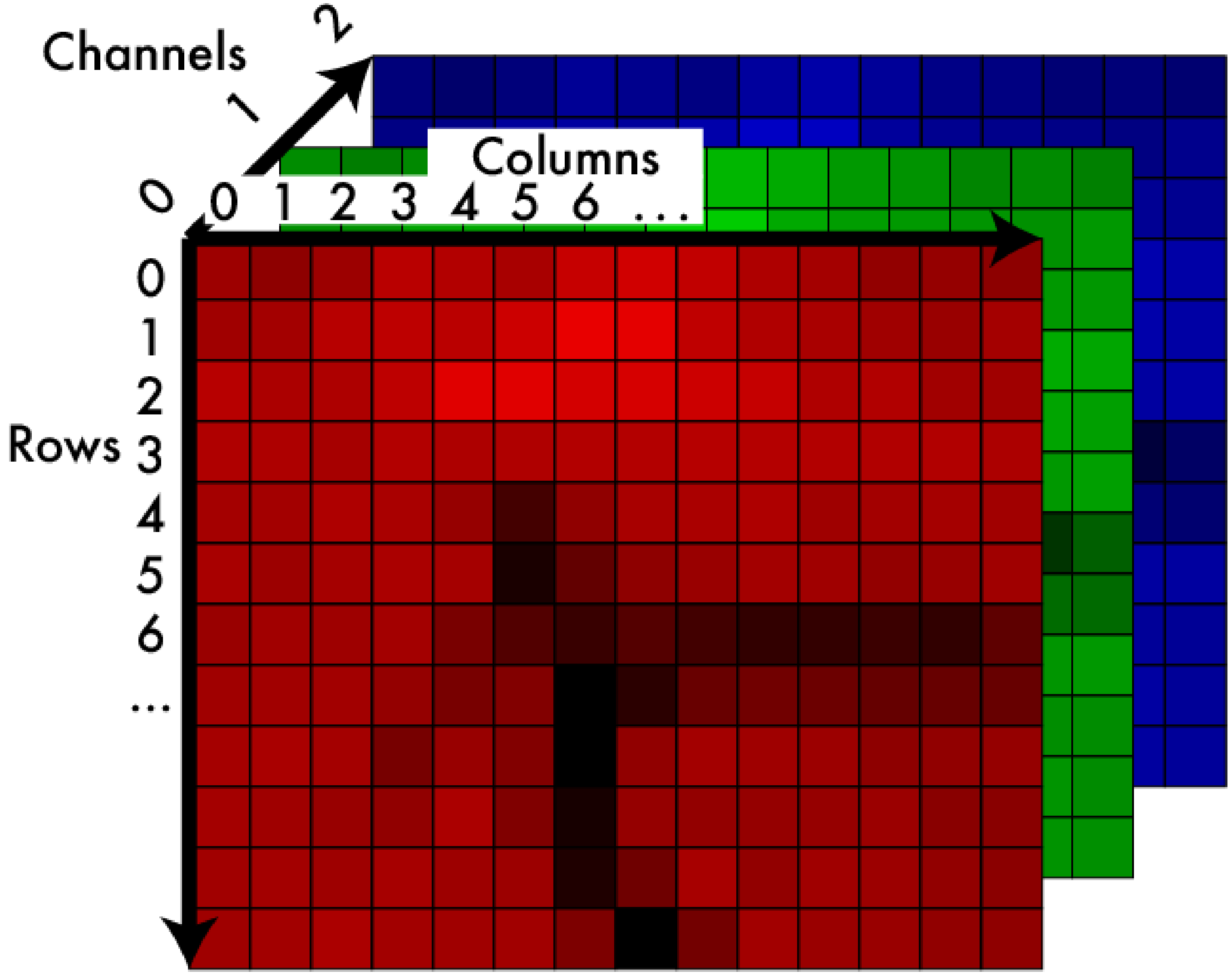
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Addressing pixels

- Ways to index:
  - (x,y)
    - Like cartesian coordinates
    - (3,6) is column 3 row 6
  - (r,c)
    - Like matrix notation
    - (3,6) is row 3 column 6
- We use (x,y)
  - Arbitrary
  - Only thing that matters is consistency



Color image: 3d tensor in colorspace

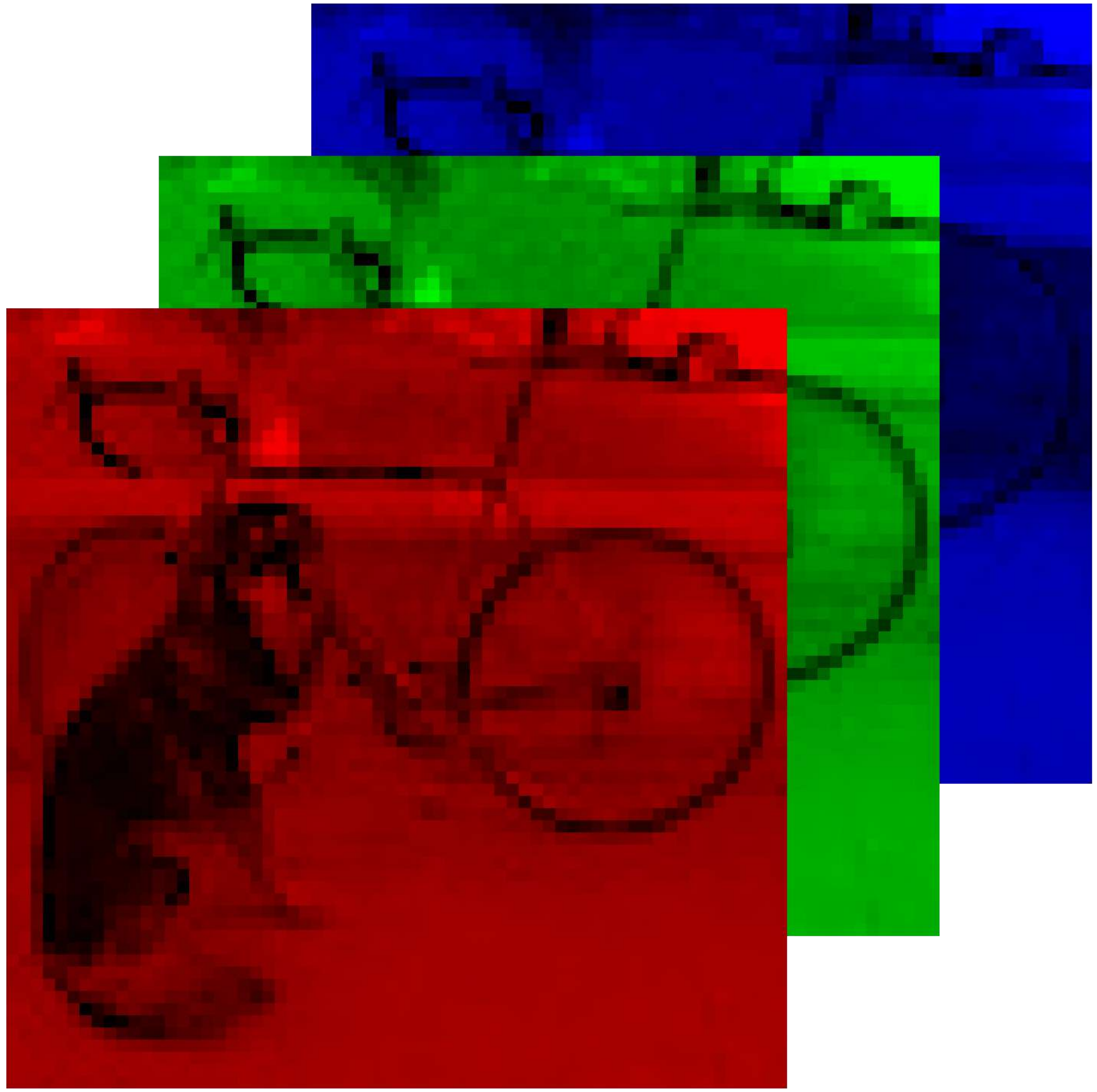


Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

RGB information in separate “channels”

Remember: we can match “real” colors using a mix of primaries.

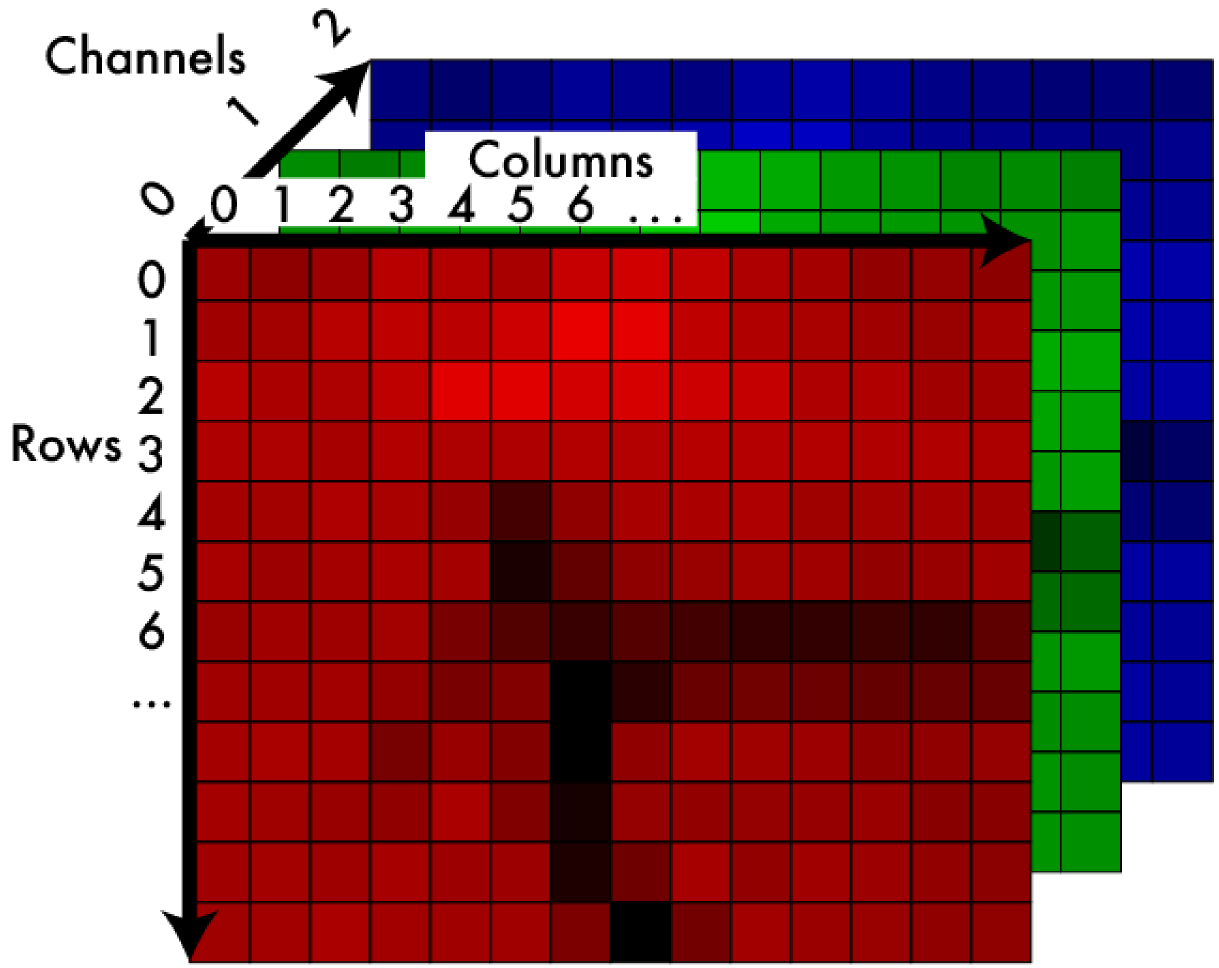
Each channel encodes one primary. Adding the light produced from each primary mimics the original color.



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Addressing pixels

- We use (x,y,c)
  - (1,2,0):
    - column 1, row 2, channel 0
- Still doesn't matter, just be consistent
- Also for size:
  - 1920 x 1080 x 3 image:
    - 1920 px wide
    - 1080 px tall
    - 3 channels



Today's Agenda

- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

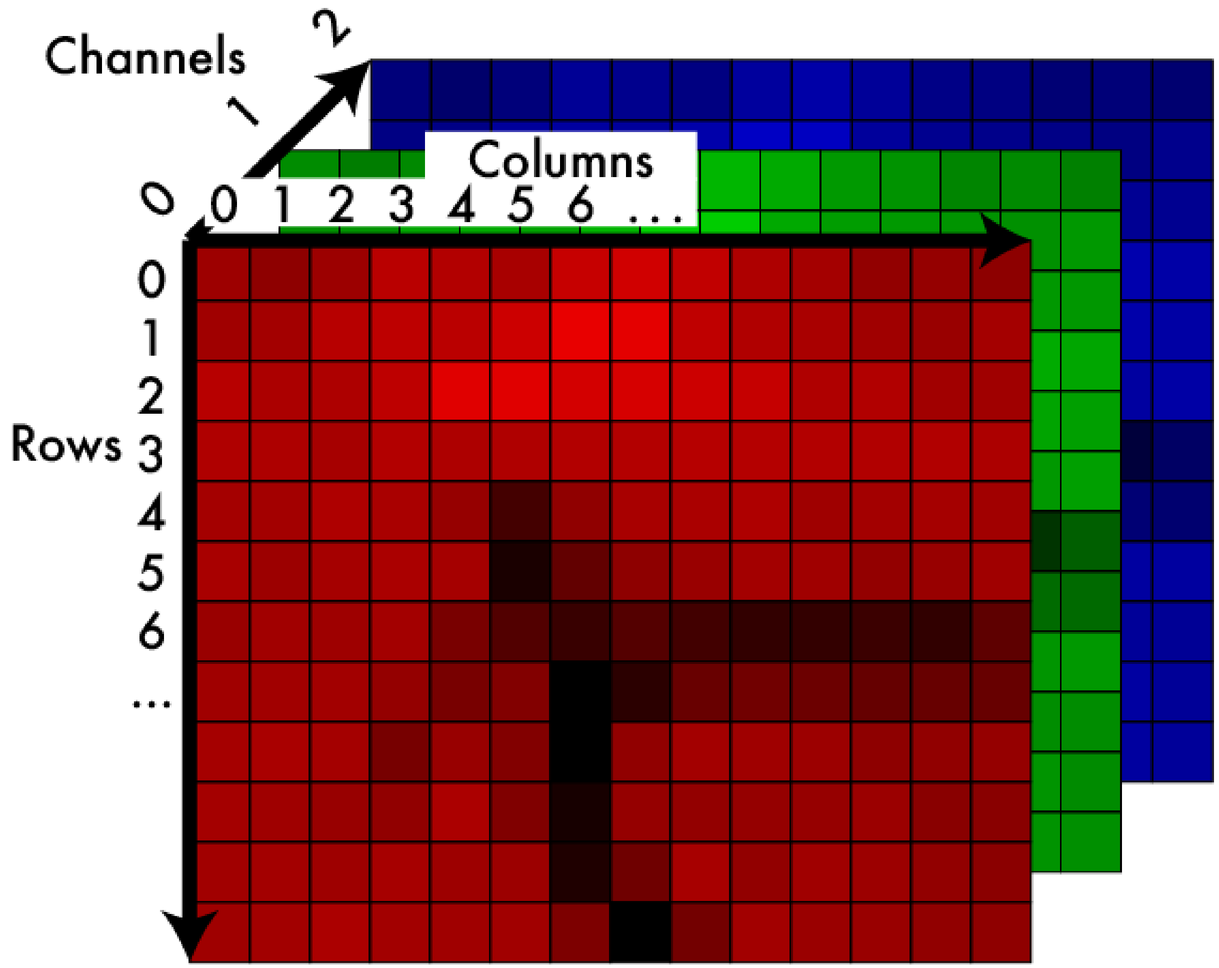
An image is like a function

An image is a mapping from indices to pixel value:

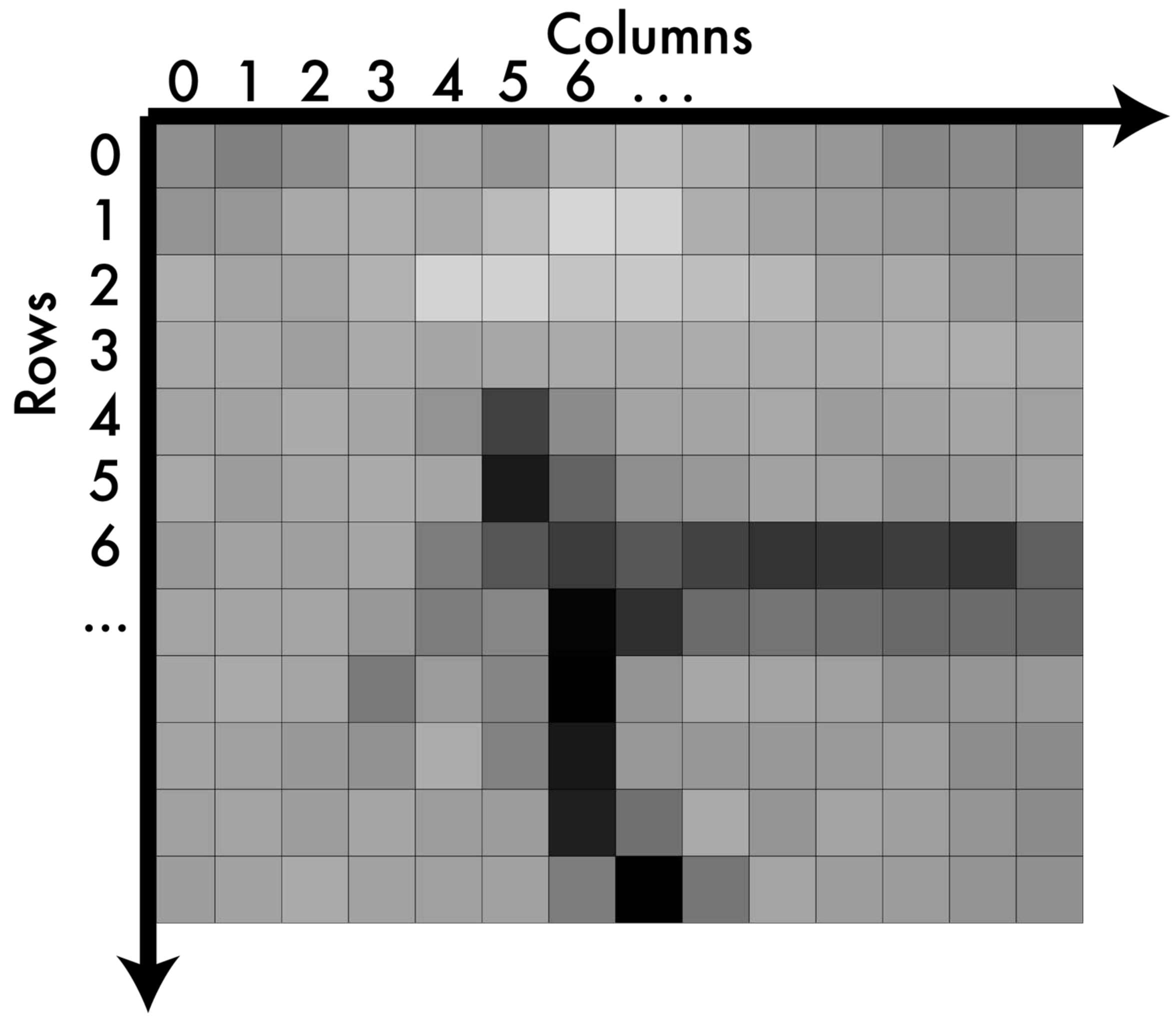
- Im: I x I x I -> R

We may want to pass in non-integers:

- Im’: R x R x I -> R

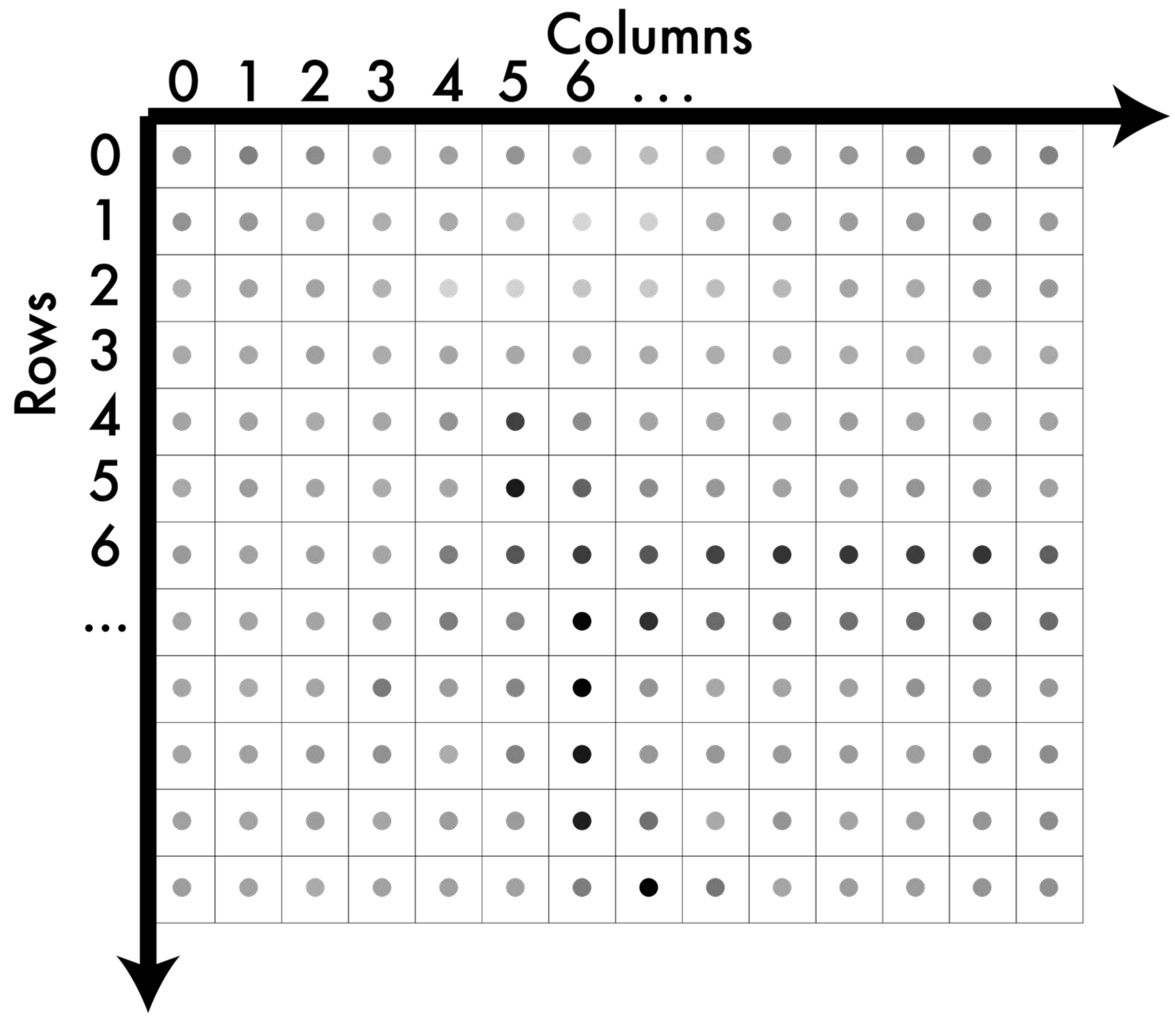


A note on coordinates in images



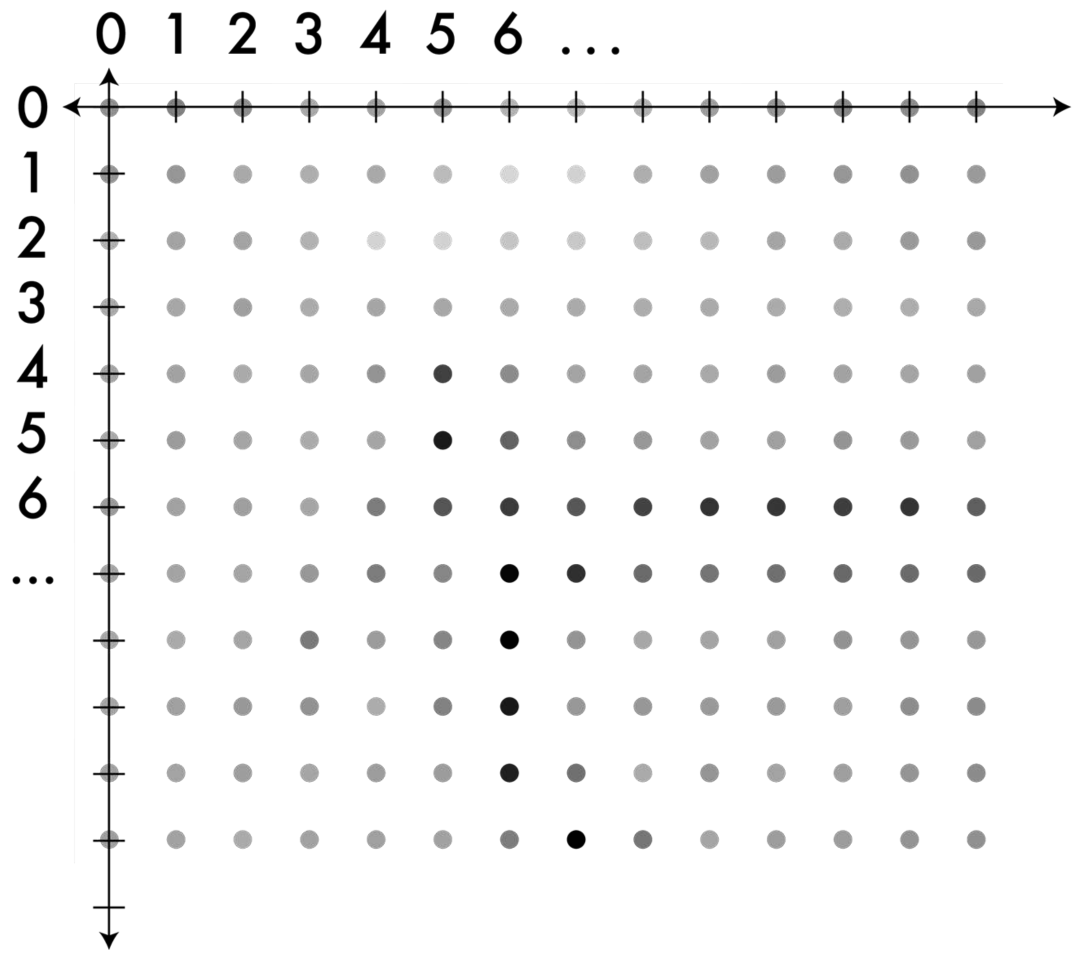
integer pixels

A note on coordinates in images



We can think of their values as being at the centers.

A note on coordinates in images

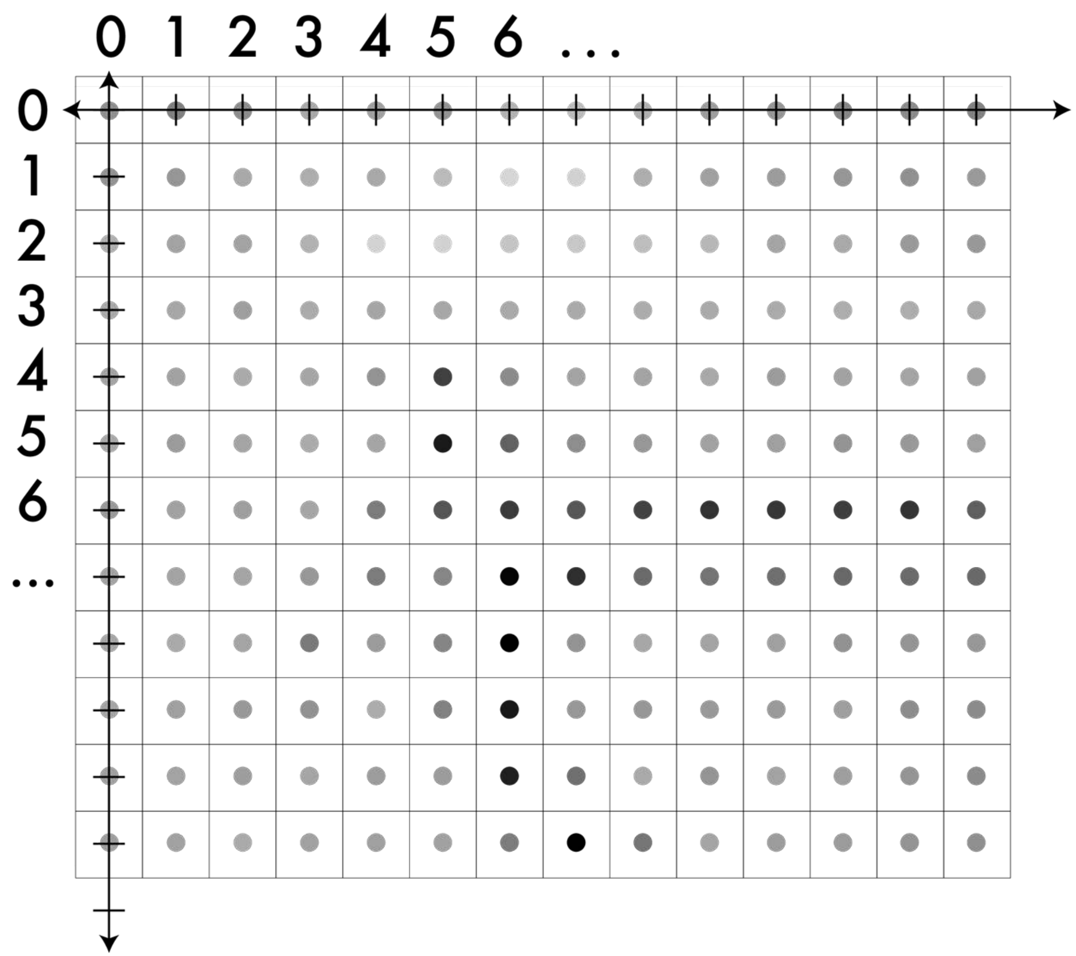


Now we can move to a real coordinate system.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A note on coordinates in images



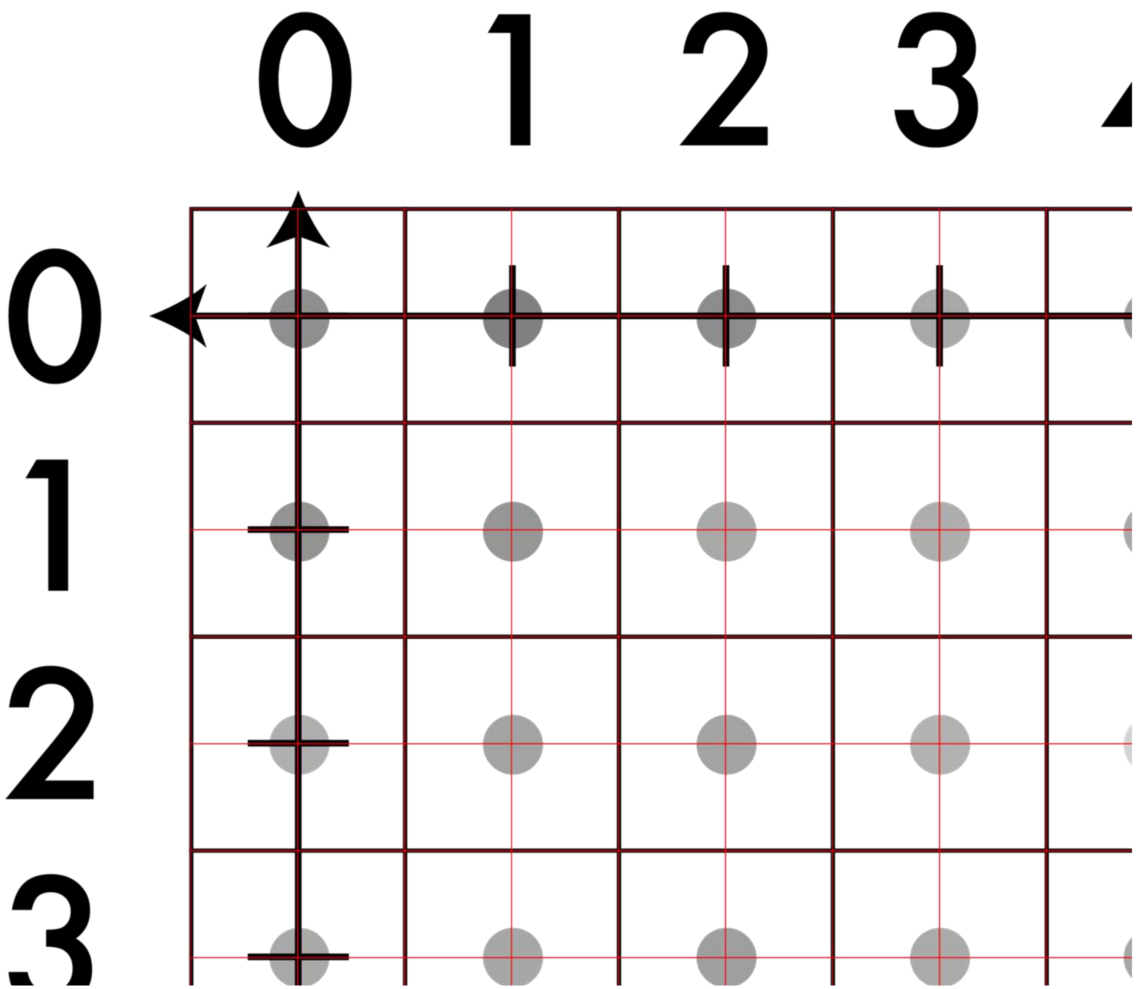
On the image

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

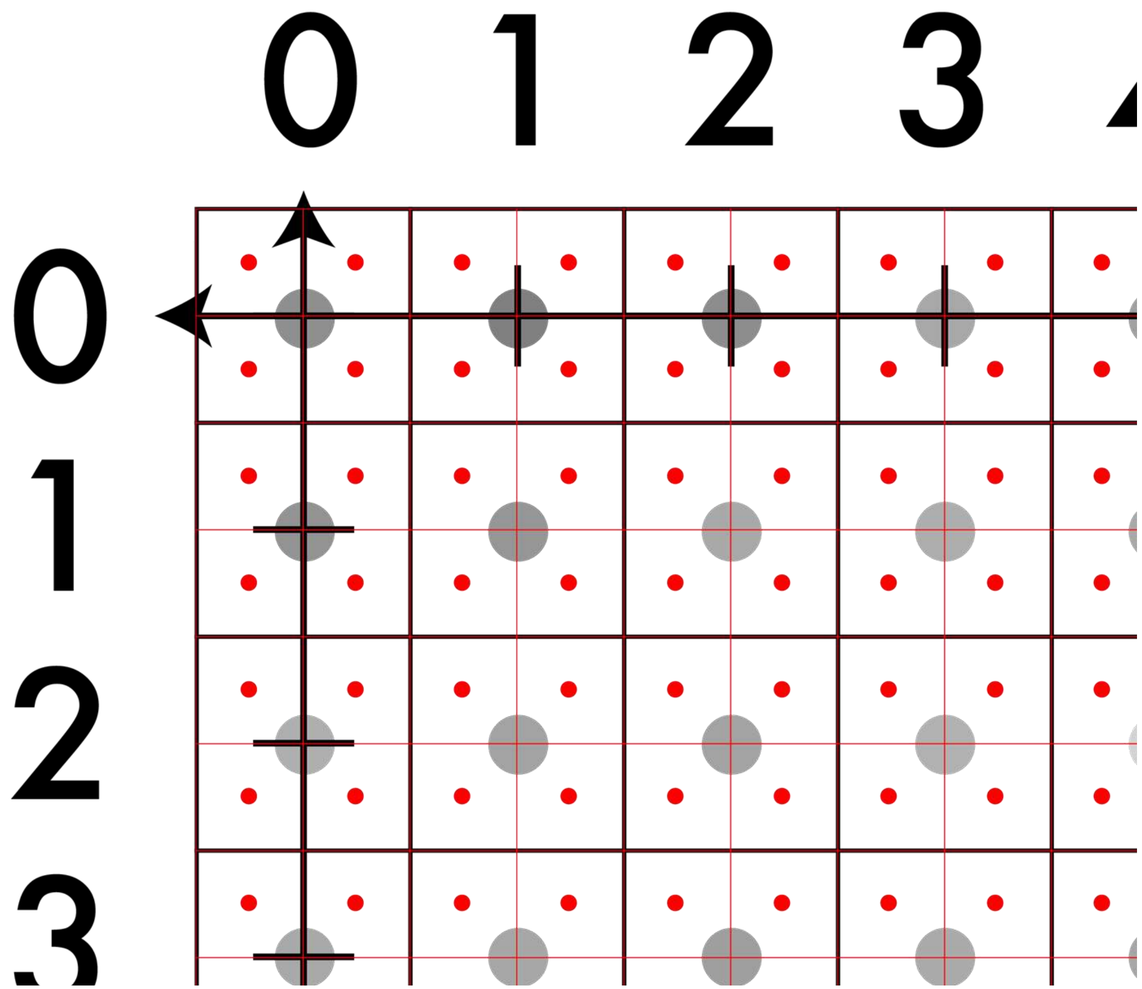
A note on coordinates in images

So, the value of the pixel (x,y) is now centered at (x,y).

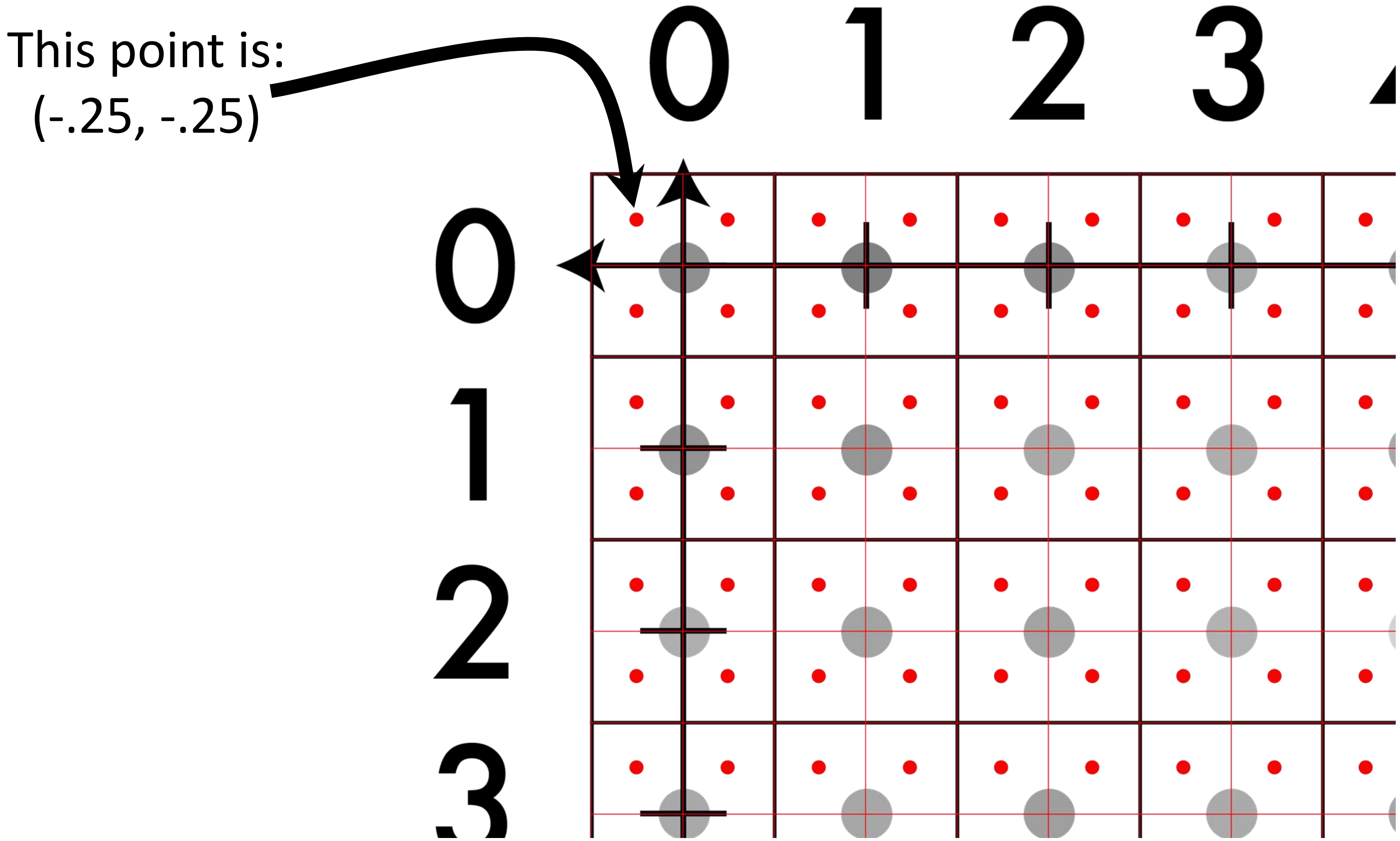


A note on coordinates in images

But there are other real-valued points.



A note on coordinates in images



Today's Agenda

- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Interpolation

How do we find out the VALUE of a non-integer point, when the image only comes with integer points, i.e. (25,45,3).

Two simple ideas:

1. Nearest-Neighbor Interpolation
2. Bilinear Interpolation

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Nearest neighbor: what it sounds like

f(x,y,z) = Im(round(x), round(y), z)

- Looks blocky
- Note: z is still int

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

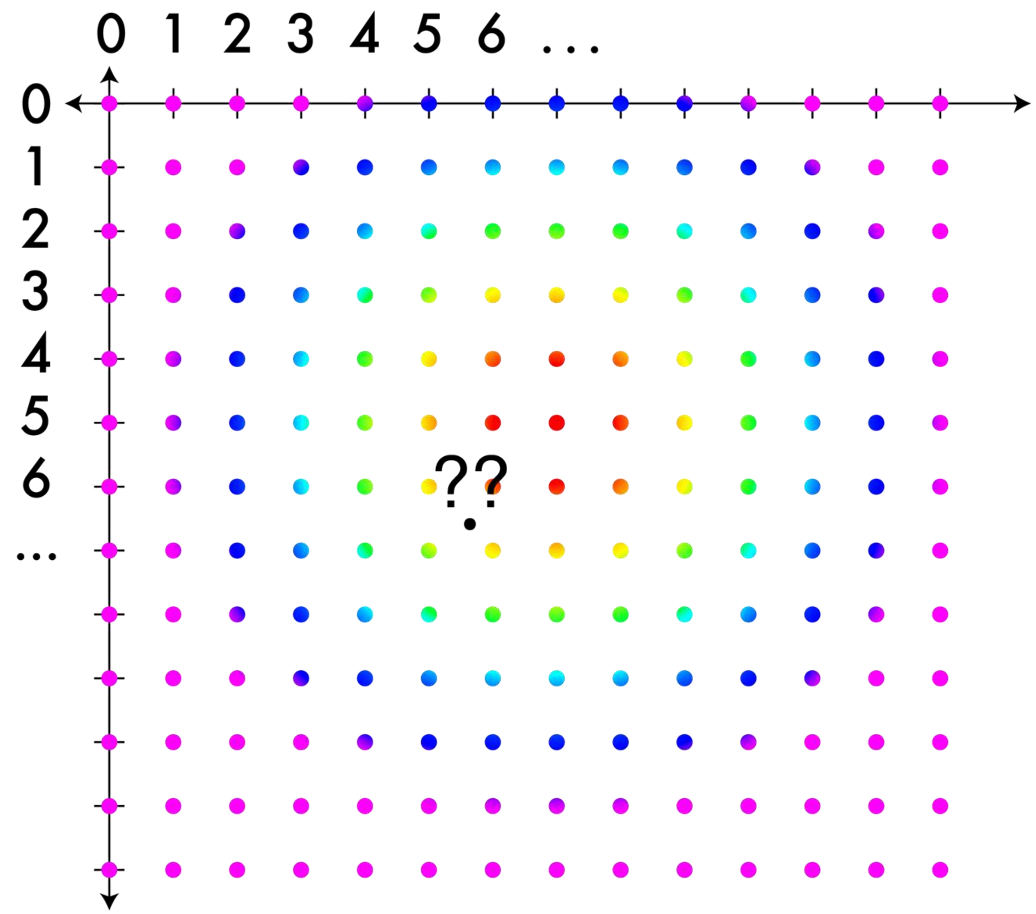
- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Bilinear interpolation: for grids, pretty good

This time find the closest pixels in a box

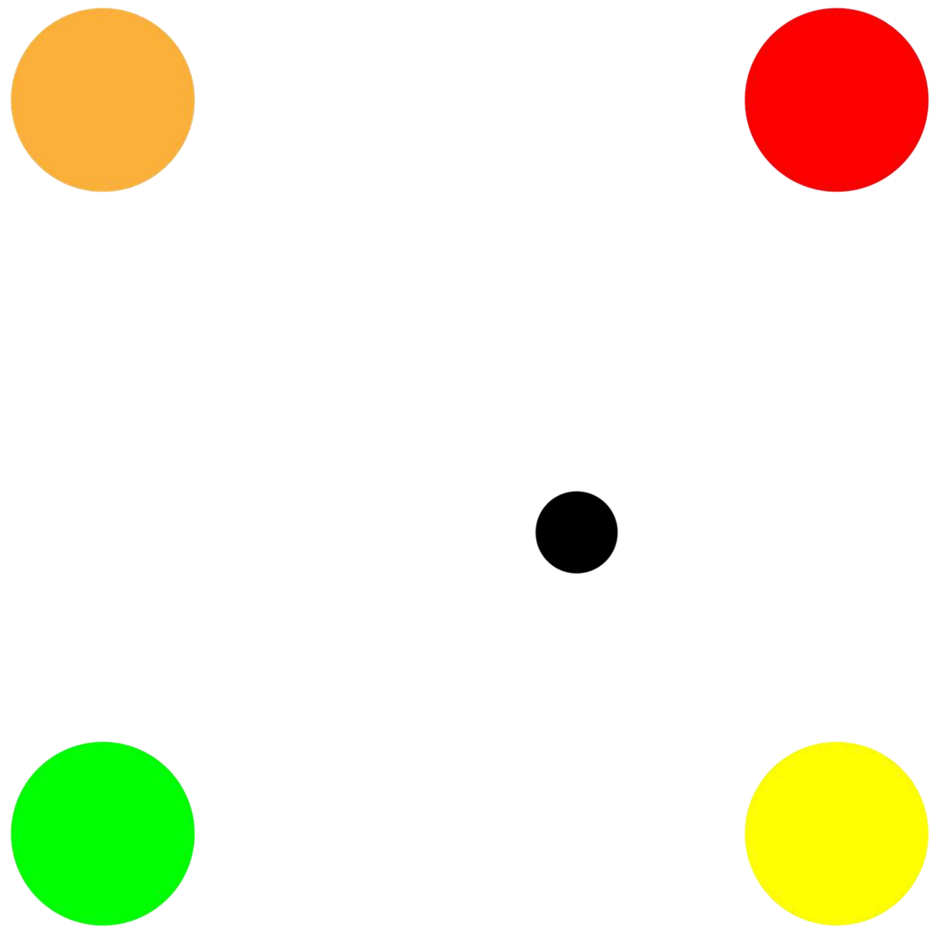


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bilinear interpolation: for grids, pretty good

This time find the closest pixels in a box

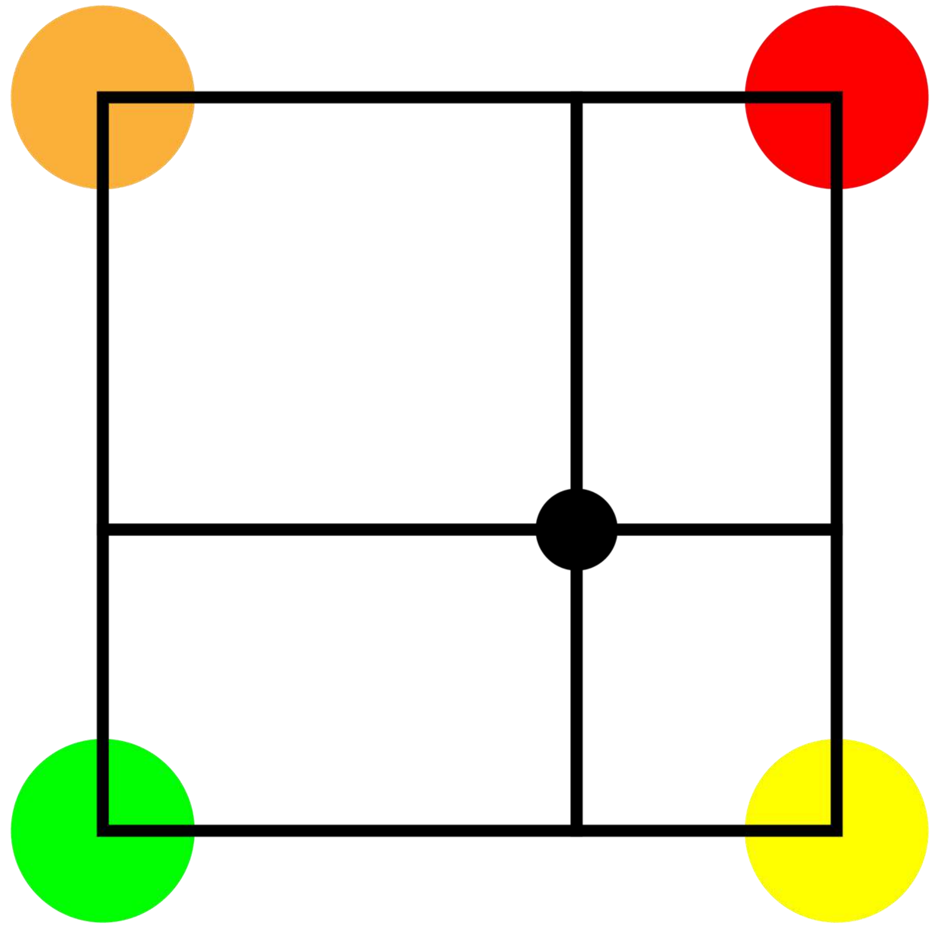


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bilinear interpolation: for grids, pretty good

This time find the closest pixels in a box



Bilinear interpolation: for grids, pretty good

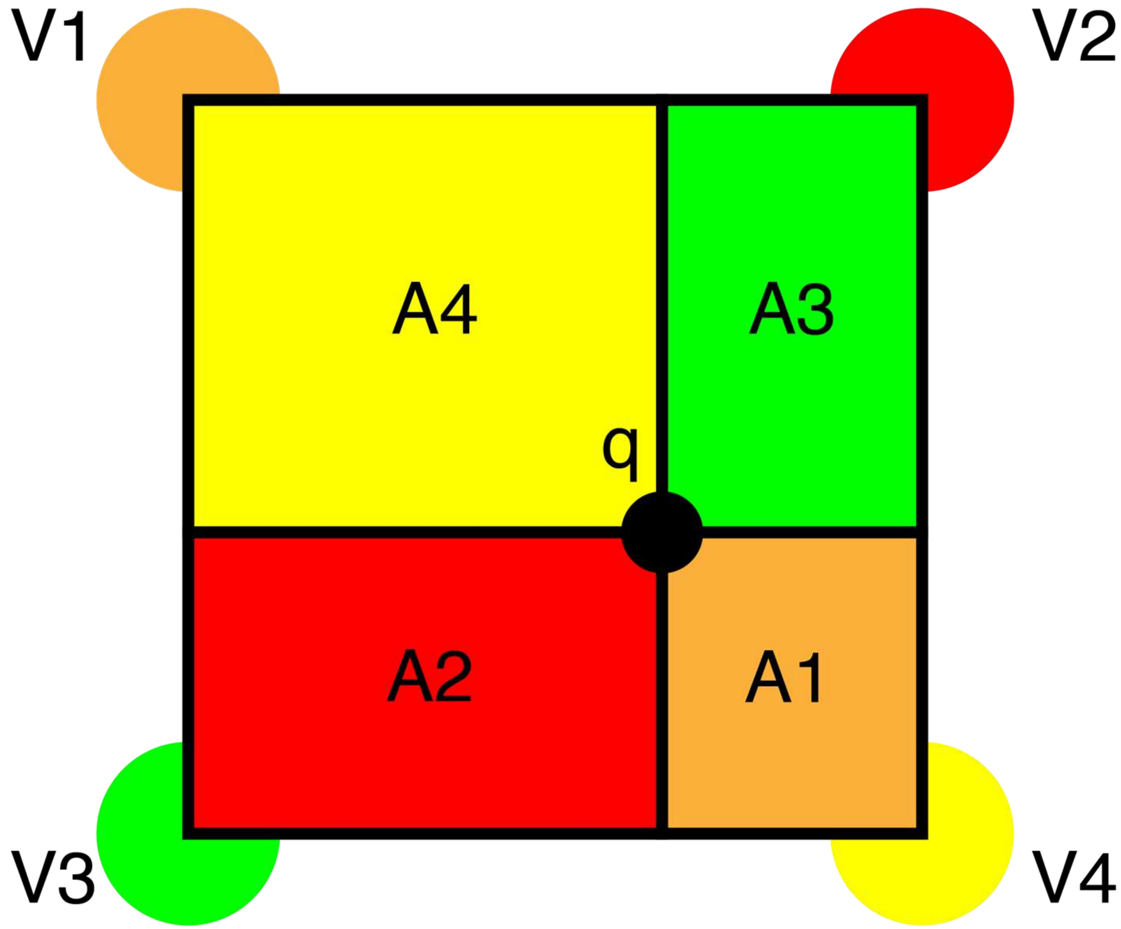
This time find the closest pixels in a box

Weighted sum based on area of opposite rectangle

q = V1*A1 + V2*A2 + V3*A3 + V4*A4

Need to normalize!

Or do we?

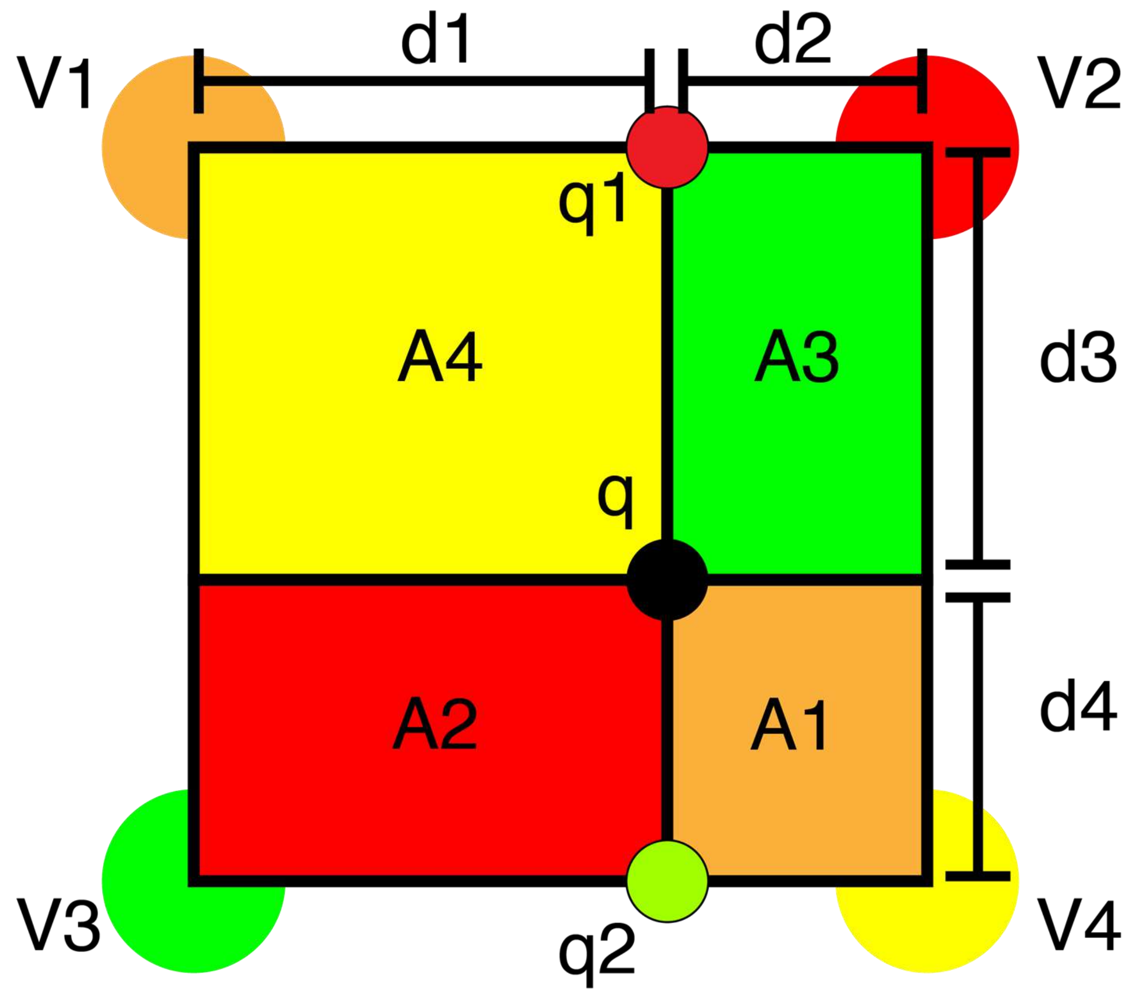


Bilinear interpolation: for grids, pretty good

q = V1*A1 + V2*A2 + V3*A3 + V4*A4

A1 = d2*d4
A2 = d1*d4
A3 = d2*d3
A4 = d1*d3

=> q = V1*d2*d4 + V2*d1*d4 + V3*d2*d3 + V4*d1*d3



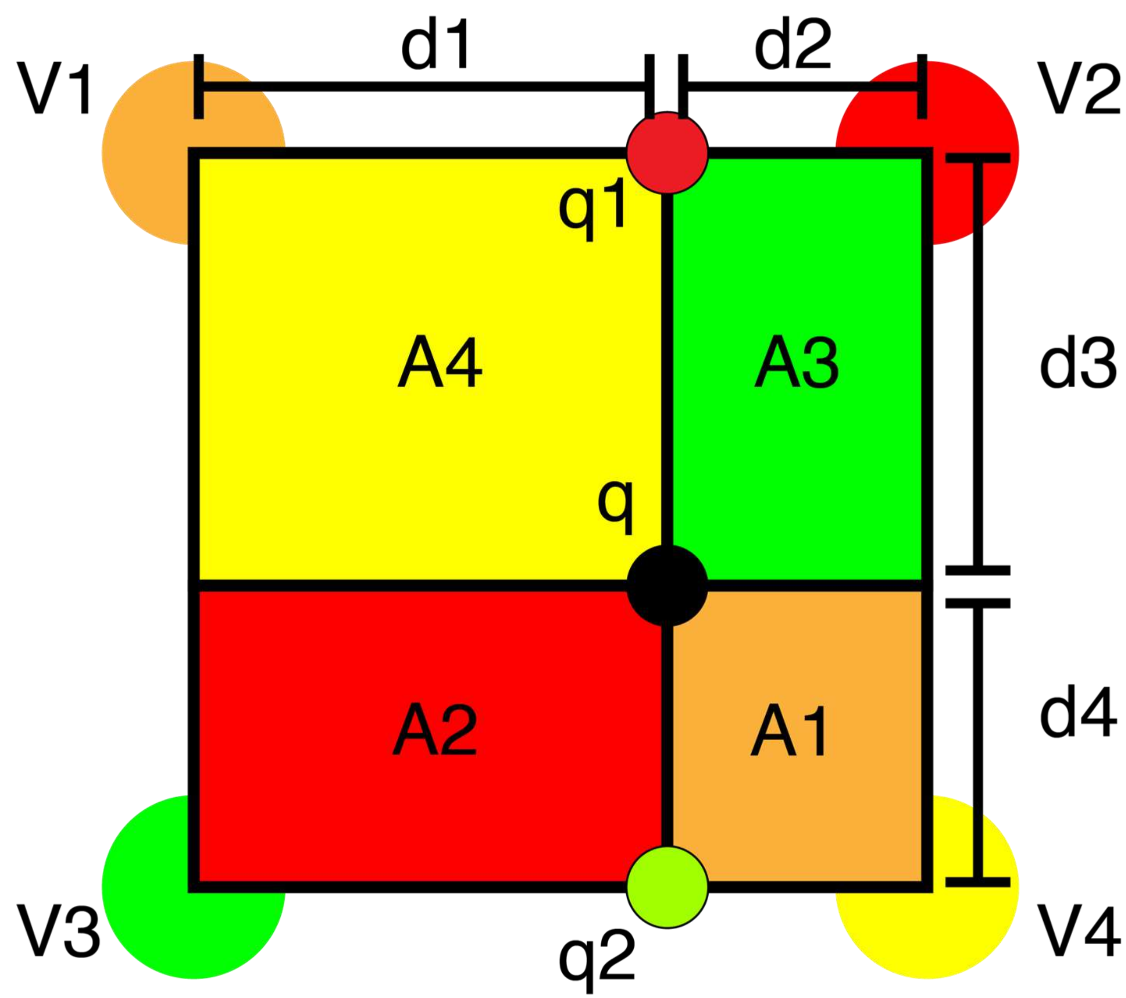
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bilinear interpolation: for grids, pretty good

Alternatively, linear interpolation of linear interpolates

q1 = V1*d2 + V2*d1
q2 = V3*d2 + V4*d1
q = q1*d4 + q2*d3



Bilinear interpolation: for grids, pretty good

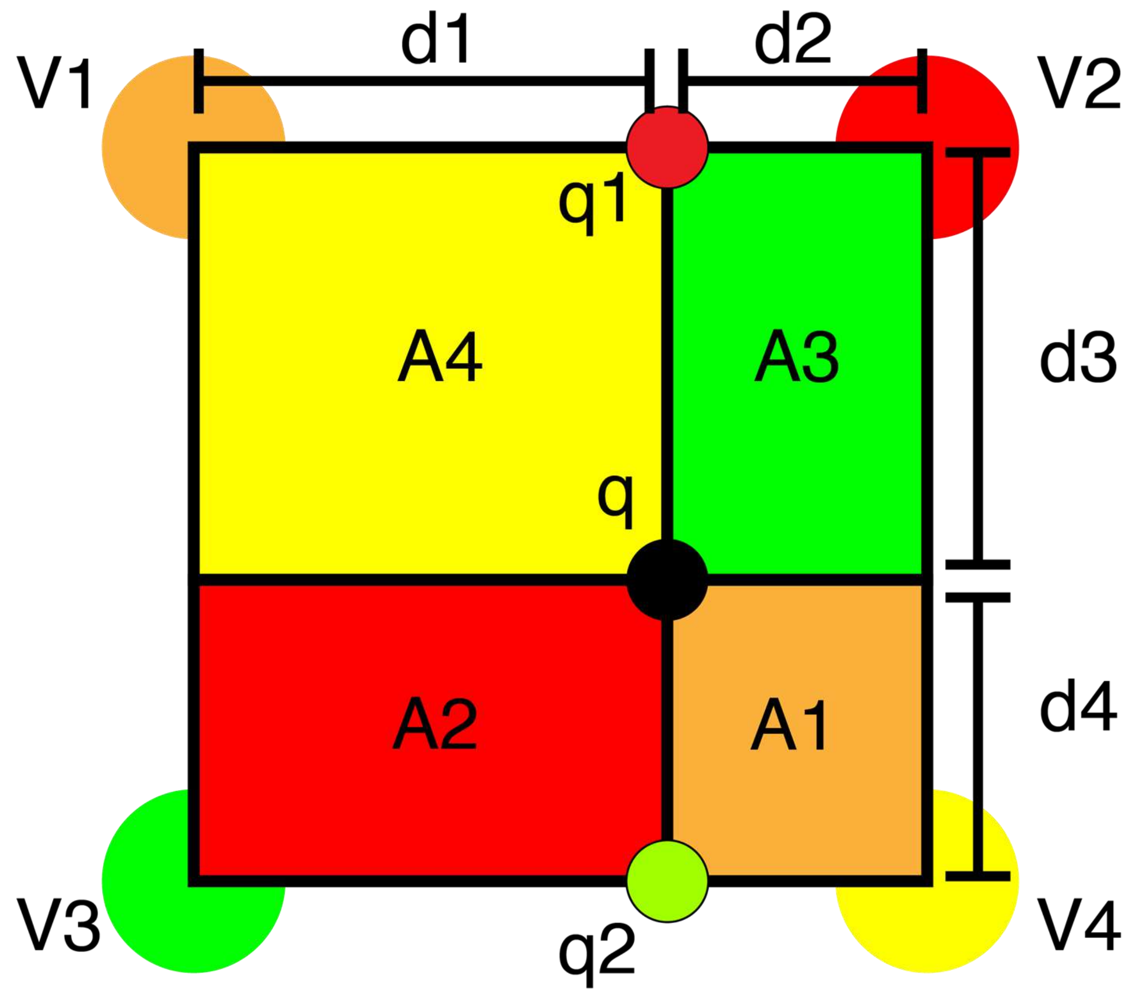
q1 = V1*d2 + V2*d1

q2 = V3*d2 + V4*d1

q = q1*d4 + q2*d3

Equivalent:

q = q1*d4 + q2*d3



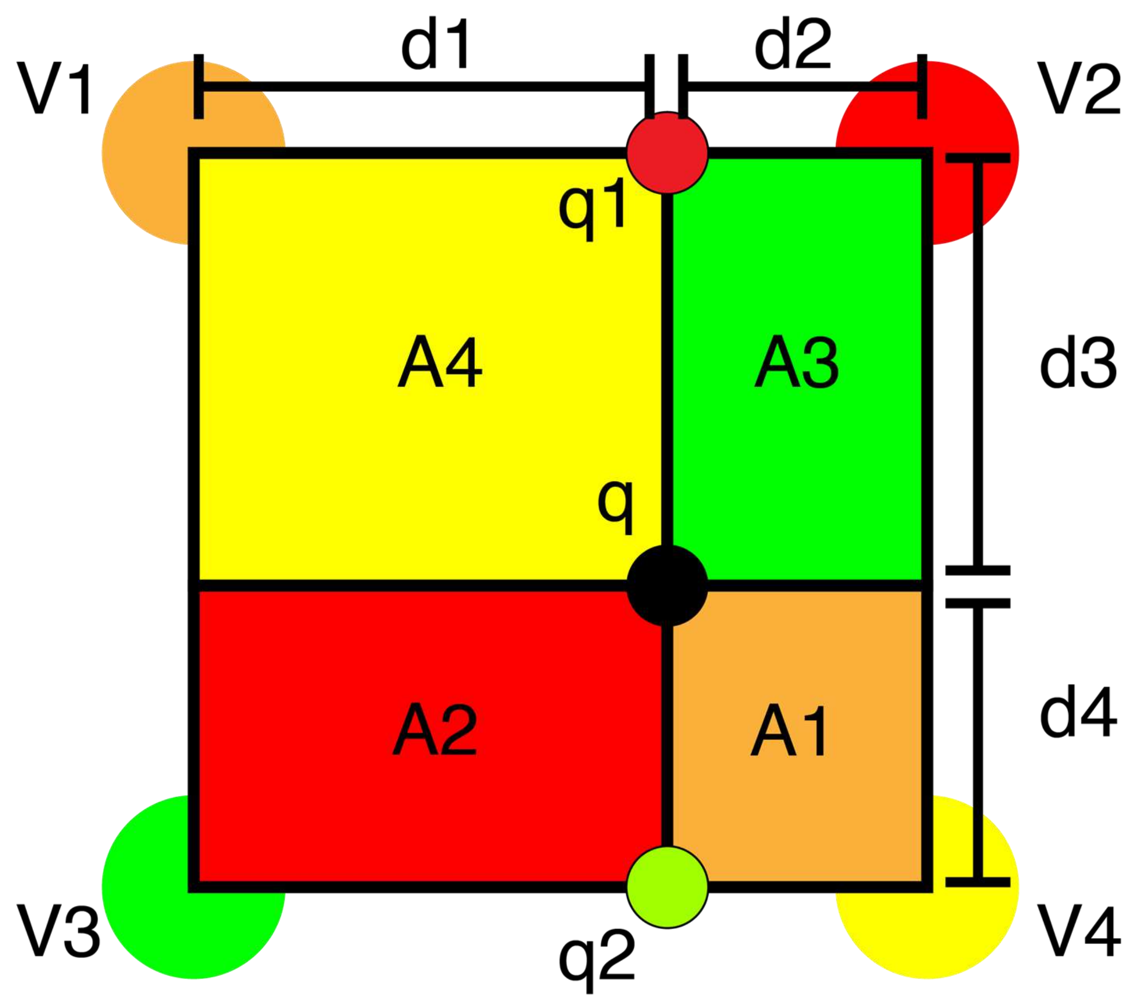
Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bilinear interpolation: for grids, pretty good

q1 = V1*d2 + V2*d1
q2 = V3*d2 + V4*d1
q = q1*d4 + q2*d3

Equivalent:

q = q1*d4 + q2*d3
q = (V1*d2 + V2*d1)*d4 + (V3*d2 + V4*d1)*d3 (subst)



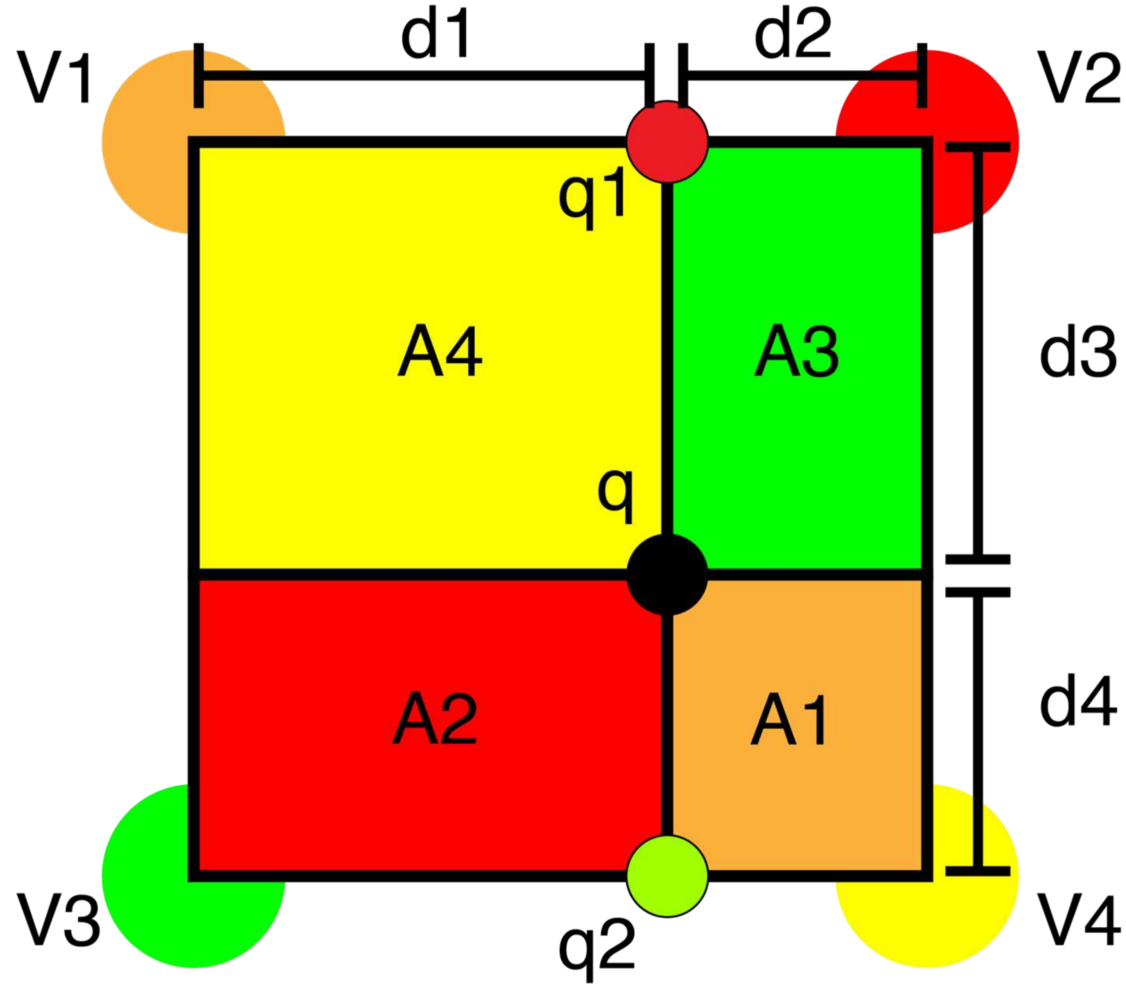
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bilinear interpolation: for grids, pretty good

q1 = V1*d2 + V2*d1
q2 = V3*d2 + V4*d1
q = q1*d4 + q2*d3

Equivalent:

q = q1*d4 + q2*d3
q = (V1*d2 + V2*d1)*d4 + (V3*d2 + V4*d1)*d3 (subst)
q = V1*d2*d4 + V2*d1*d4 + V3*d2*d3 + V4*d1*d3 (distribution)



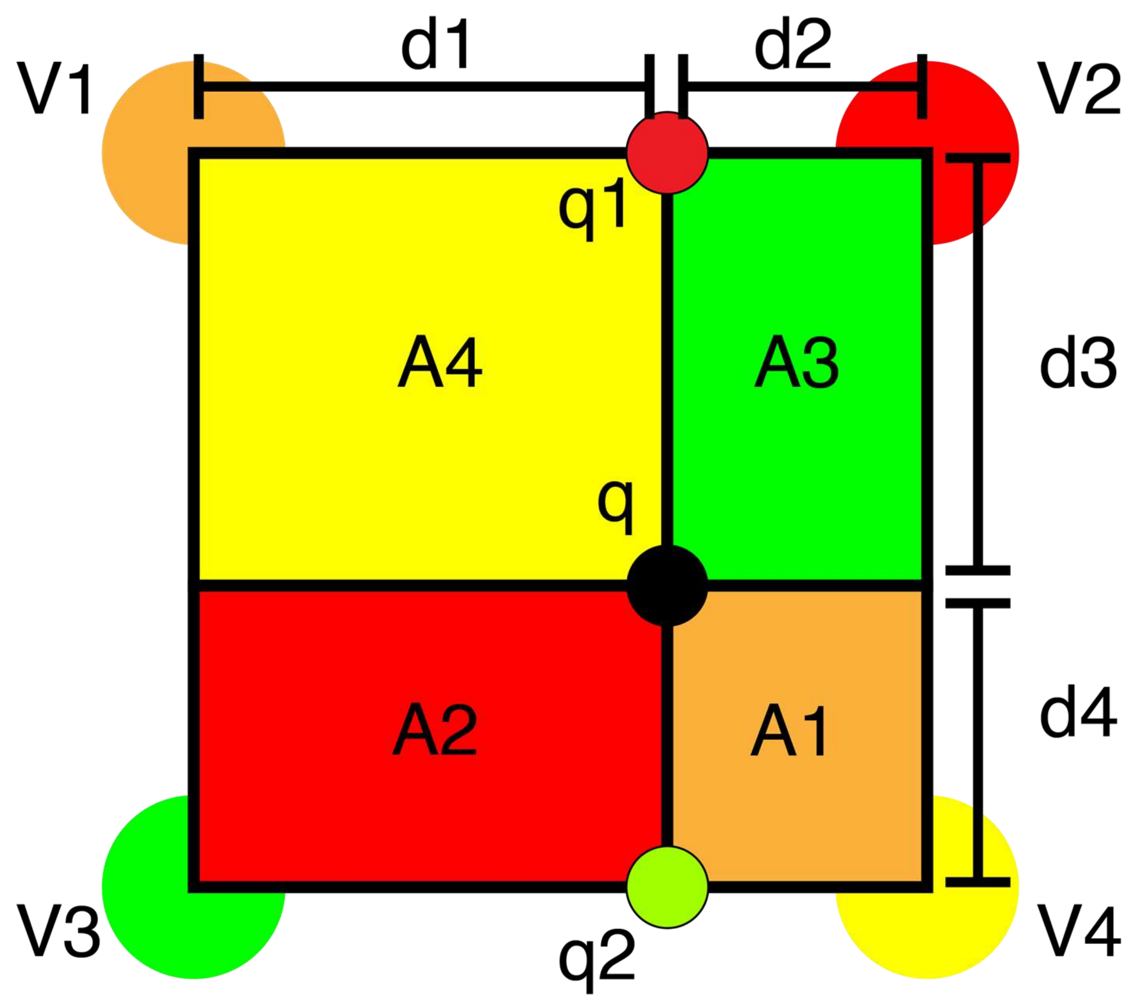
Bilinear interpolation: for grids, pretty good

q1 = V1*d2 + V2*d1
q2 = V3*d2 + V4*d1
q = q1*d4 + q2*d3

Equivalent:

q = q1*d4 + q2*d3
q = (V1*d2 + V2*d1)*d4 + (V3*d2 + V4*d1)*d3 (subst)
q = V1*d2*d4 + V2*d1*d4 + V3*d2*d3 + V4*d1*d3 (distribution)

Recall:
A1 = d2*d4
A2 = d1*d4
A3 = d2*d3
A4 = d1*d3



Bilinear interpolation: for grids, pretty good

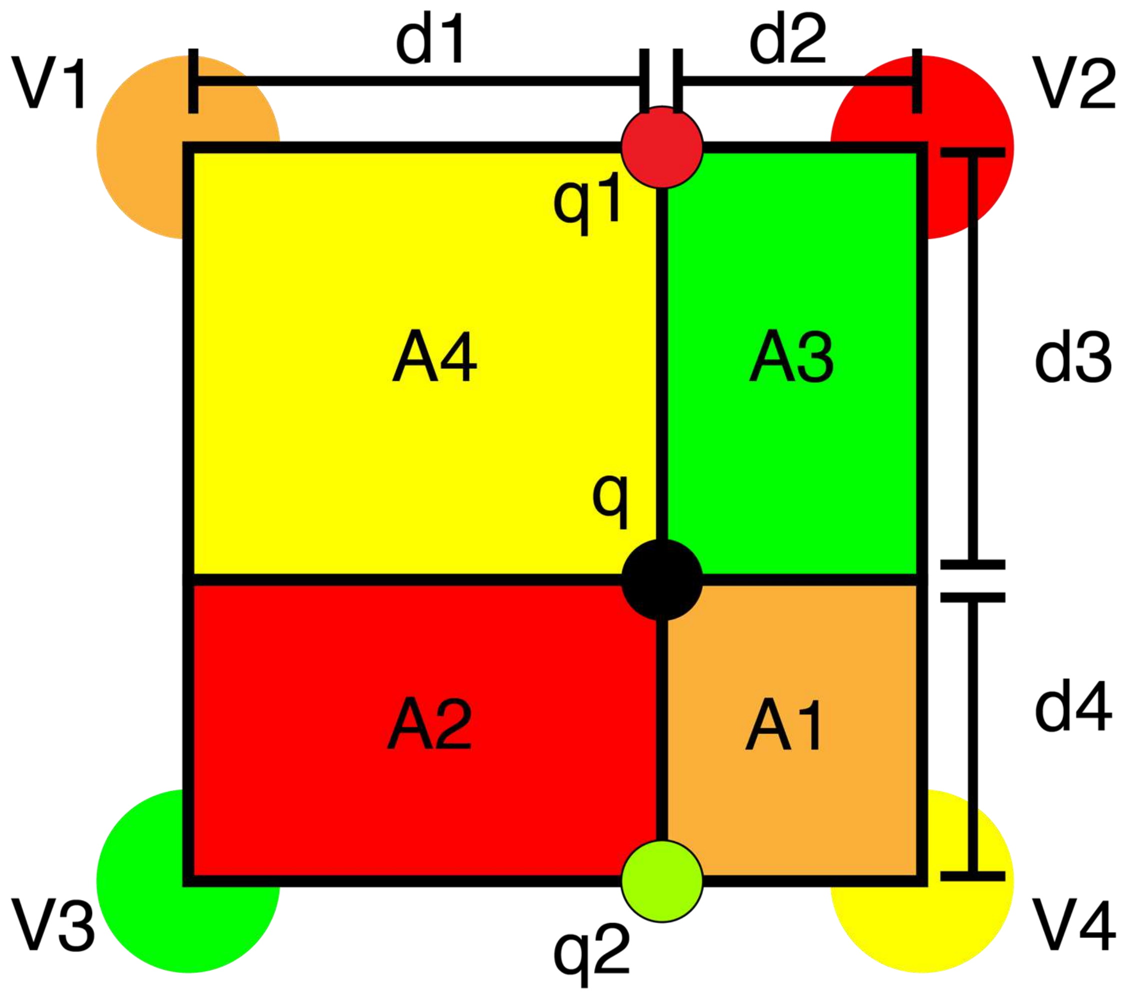
q1 = V1*d2 + V2*d1
q2 = V3*d2 + V4*d1
q = q1*d4 + q2*d3

Equivalent:

q = q1*d4 + q2*d3
q = (V1*d2 + V2*d1)*d4 + (V3*d2 + V4*d1)*d3 (subst)
q = V1*d2*d4 + V2*d1*d4 + V3*d2*d3 + V4*d1*d3 (distribution)

Recall:
A1 = d2*d4
A2 = d1*d4
A3 = d2*d3
A4 = d1*d3

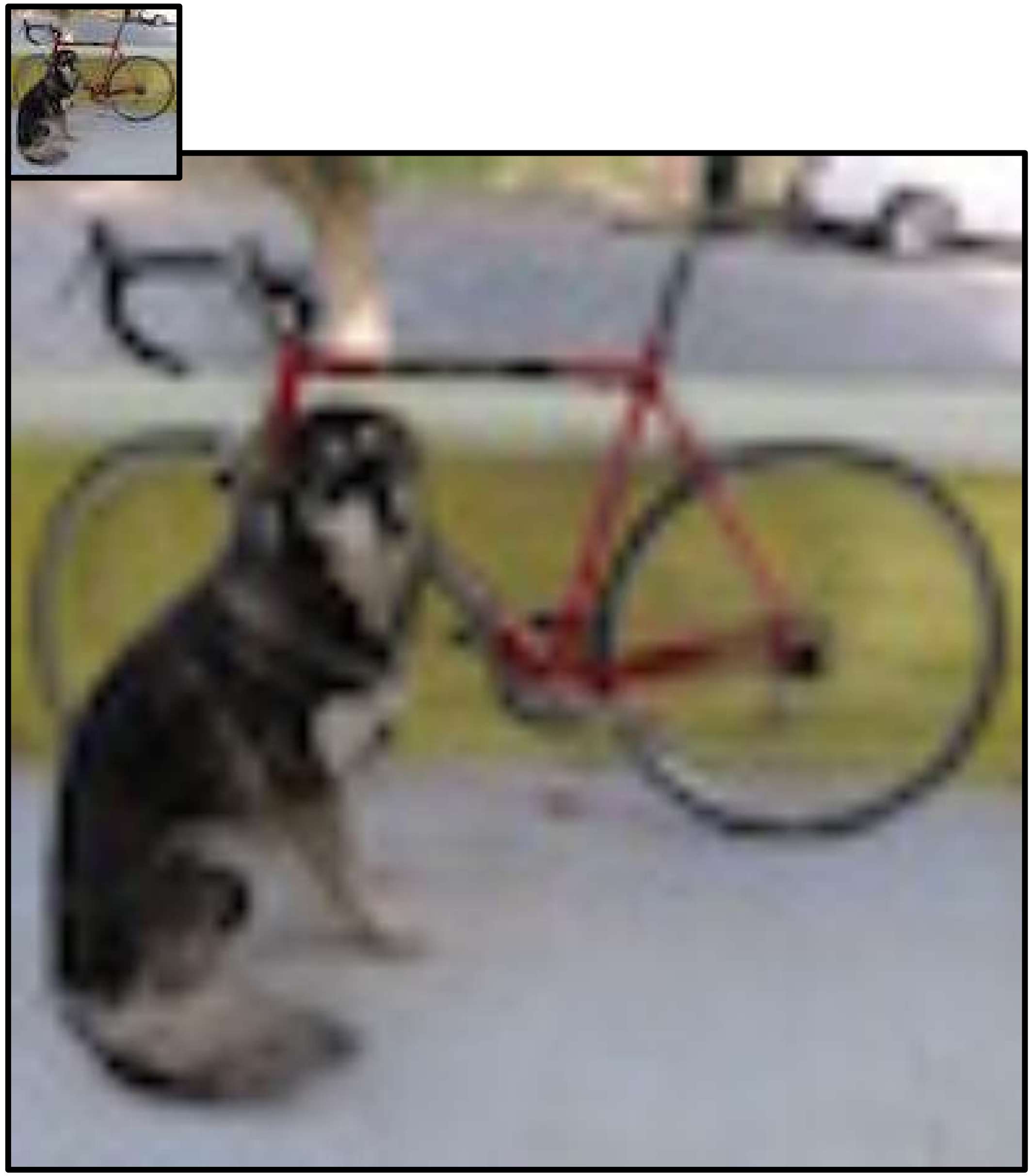
q = V1*A1 + V2*A2 + V3*A3 + V4*A4



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bilinear interpolation: for grids, pretty good

- Smoother than NN
- More complex
  - 4 lookups
  - Some math
- Often the right tradeoff of speed vs final result



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

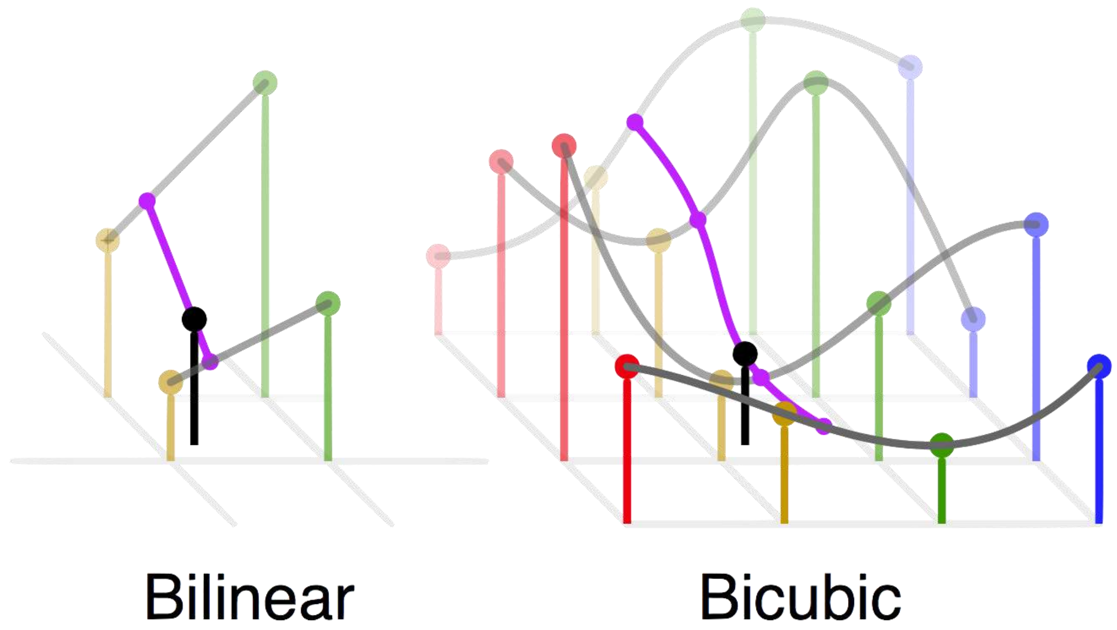
Today’s Agenda

- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

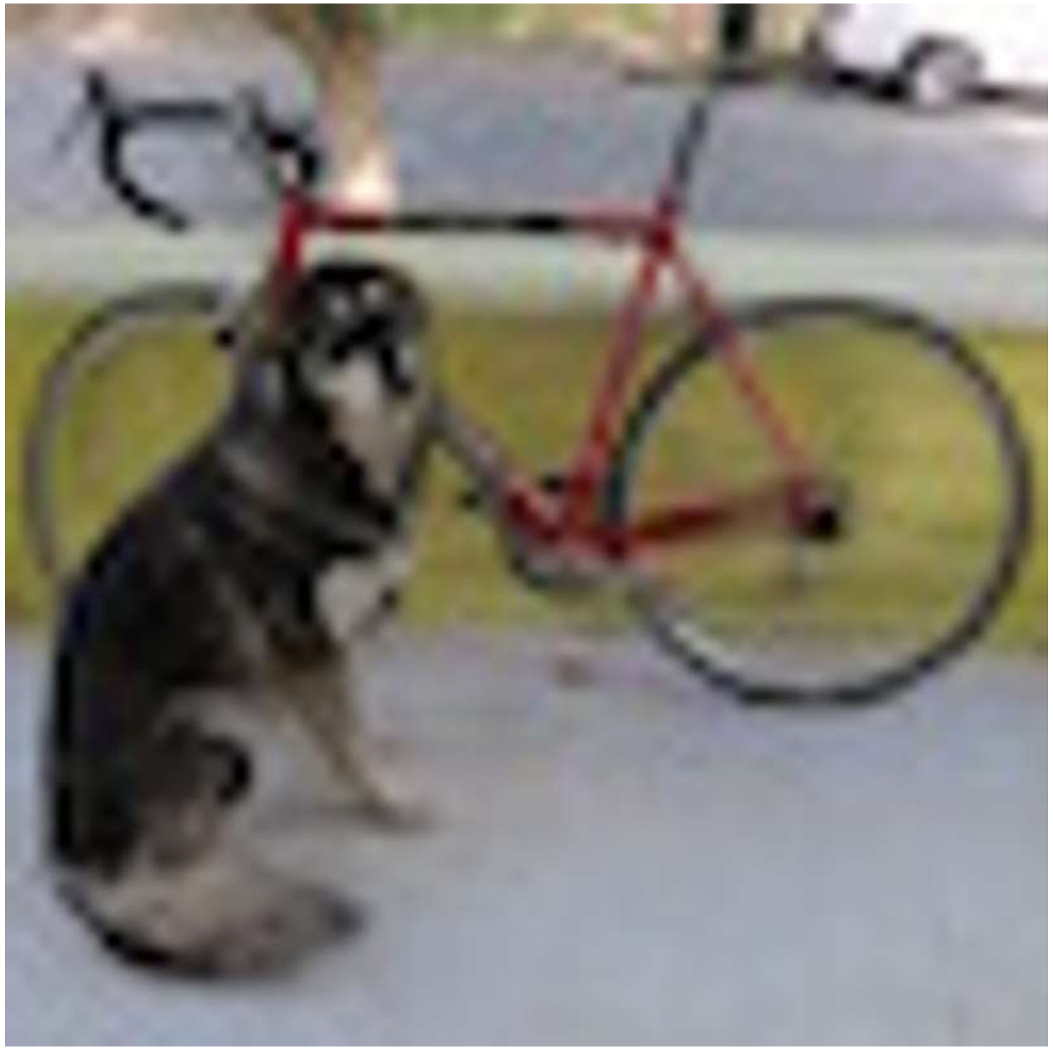
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bicubic sampling: more complex, maybe better?

- A cubic interpolation of 4 cubic interpolations
- Smoother than bilinear, no "star"
- 16 nearest neighbors
- Fit 3rd order poly:
  - f(x) = a + bx + cx^2 + dx^3
- Interpolate along axis
- Fit another poly to interpolated values



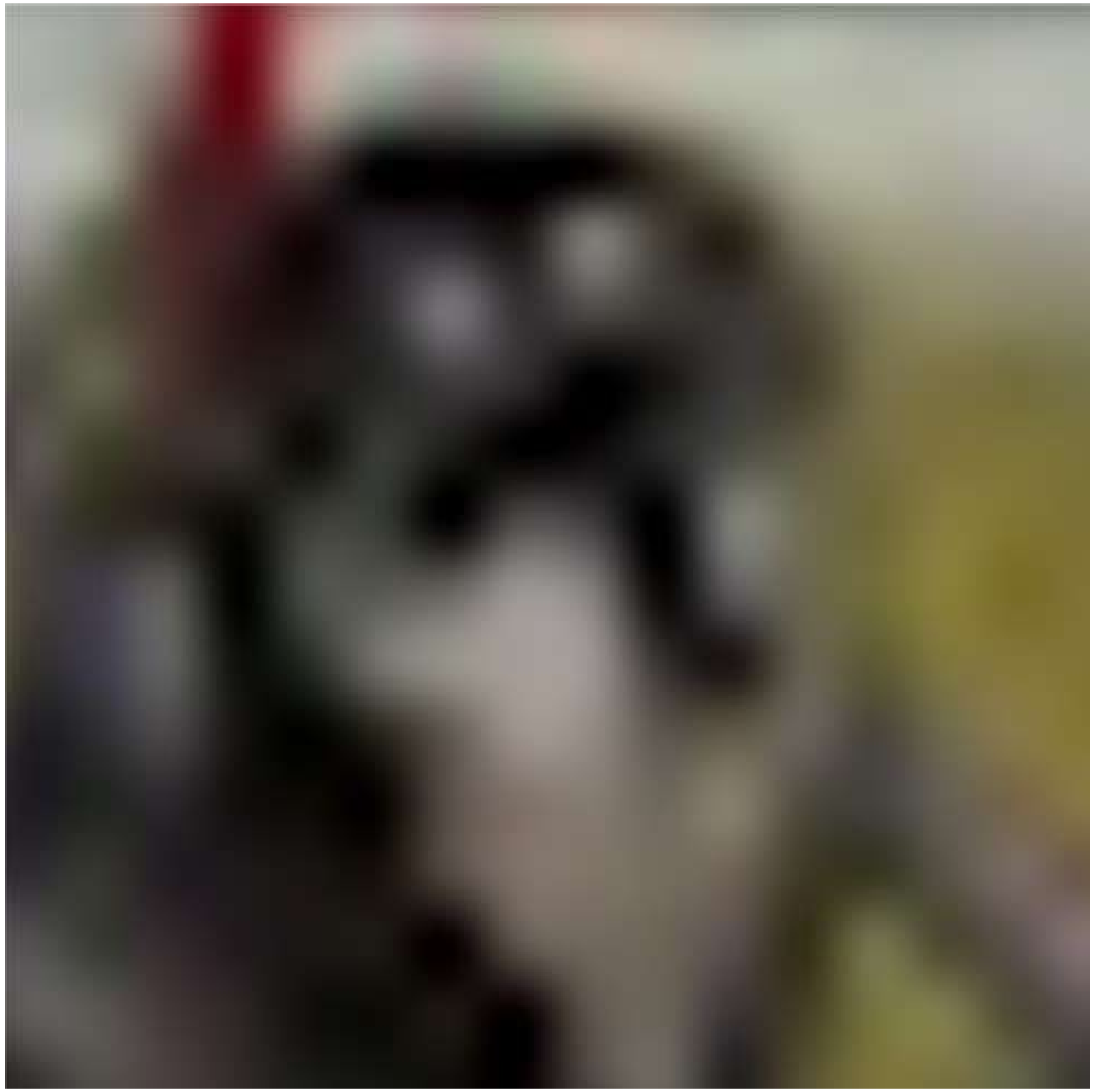
Bicubic vs bilinear

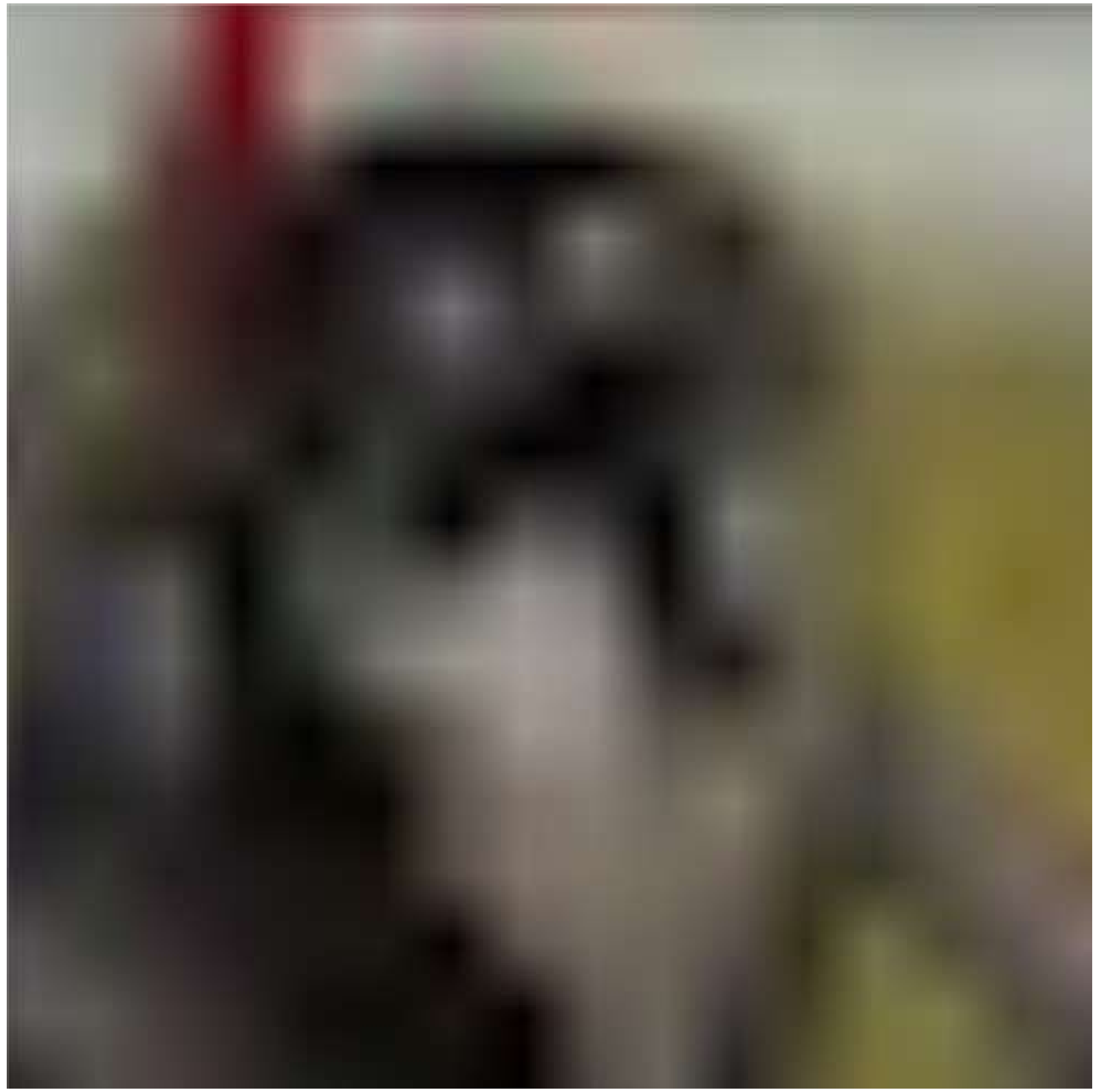


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bicubic vs bilinear



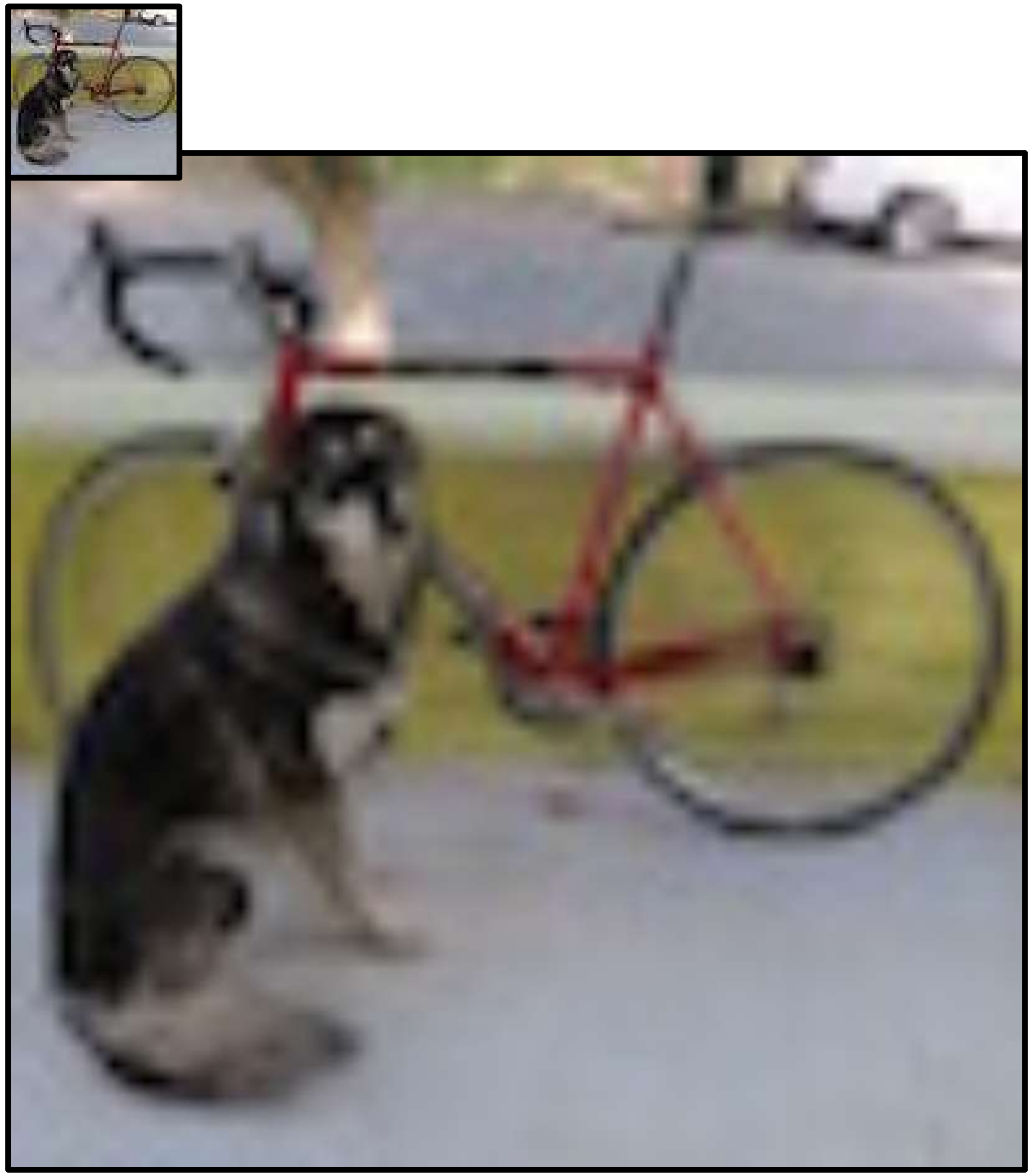


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Resize algorithm:

- For each pixel in new image:
  - Map to old im coordinates
  - Interpolate value
  - Set new value in image



So what is this interpolation useful for?

Today's Agenda

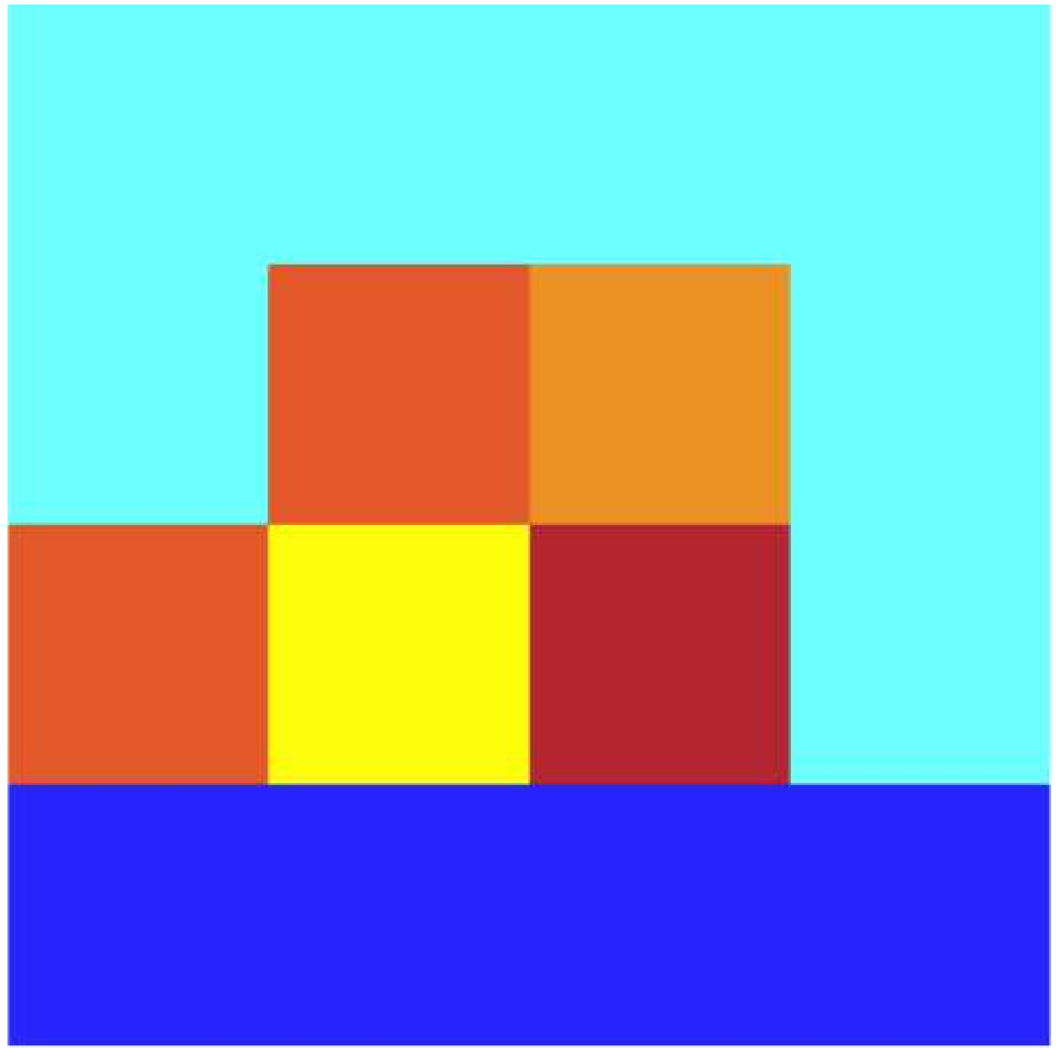
- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image resizing!

Say we want to increase the size of an image…

This is a beautiful image of a sunset… it's just very small…



Co-financed by the European Union
Connecting Europe Facility

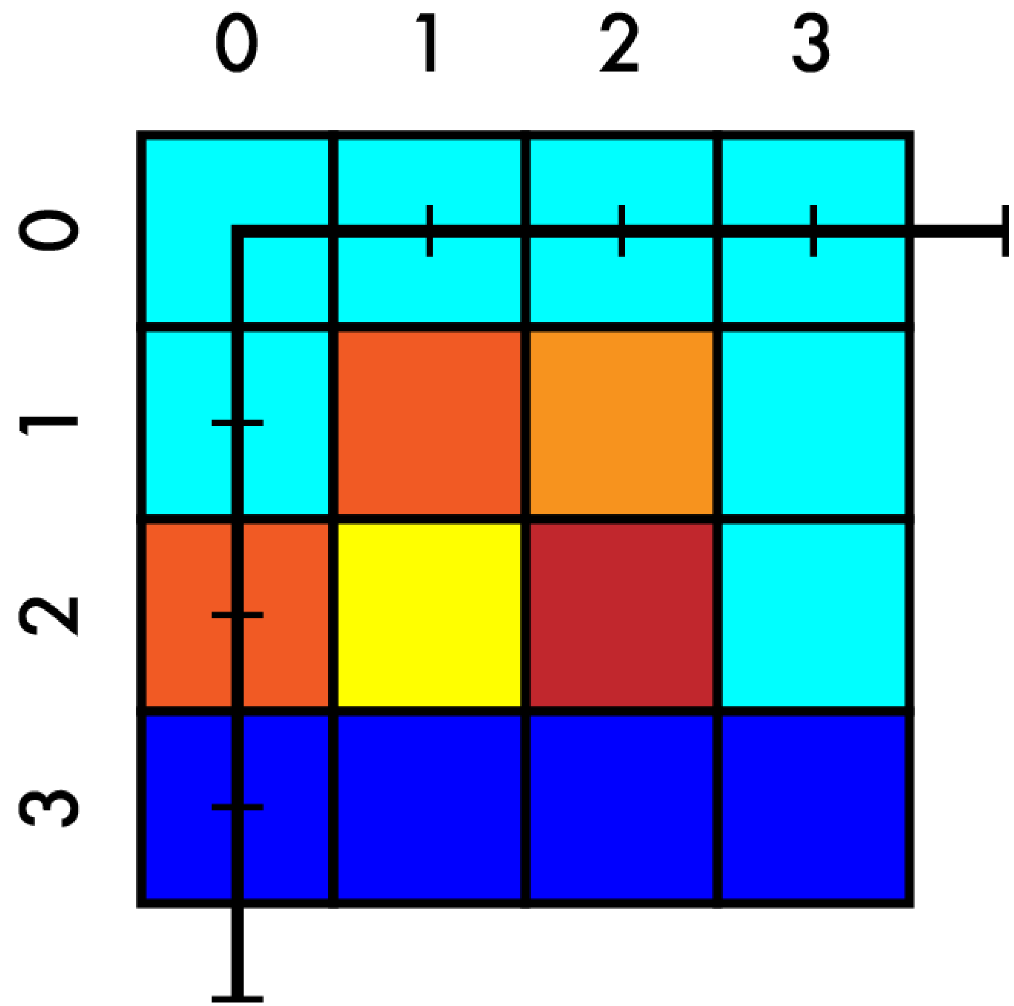
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image resizing!

Say we want to increase the size of an image…

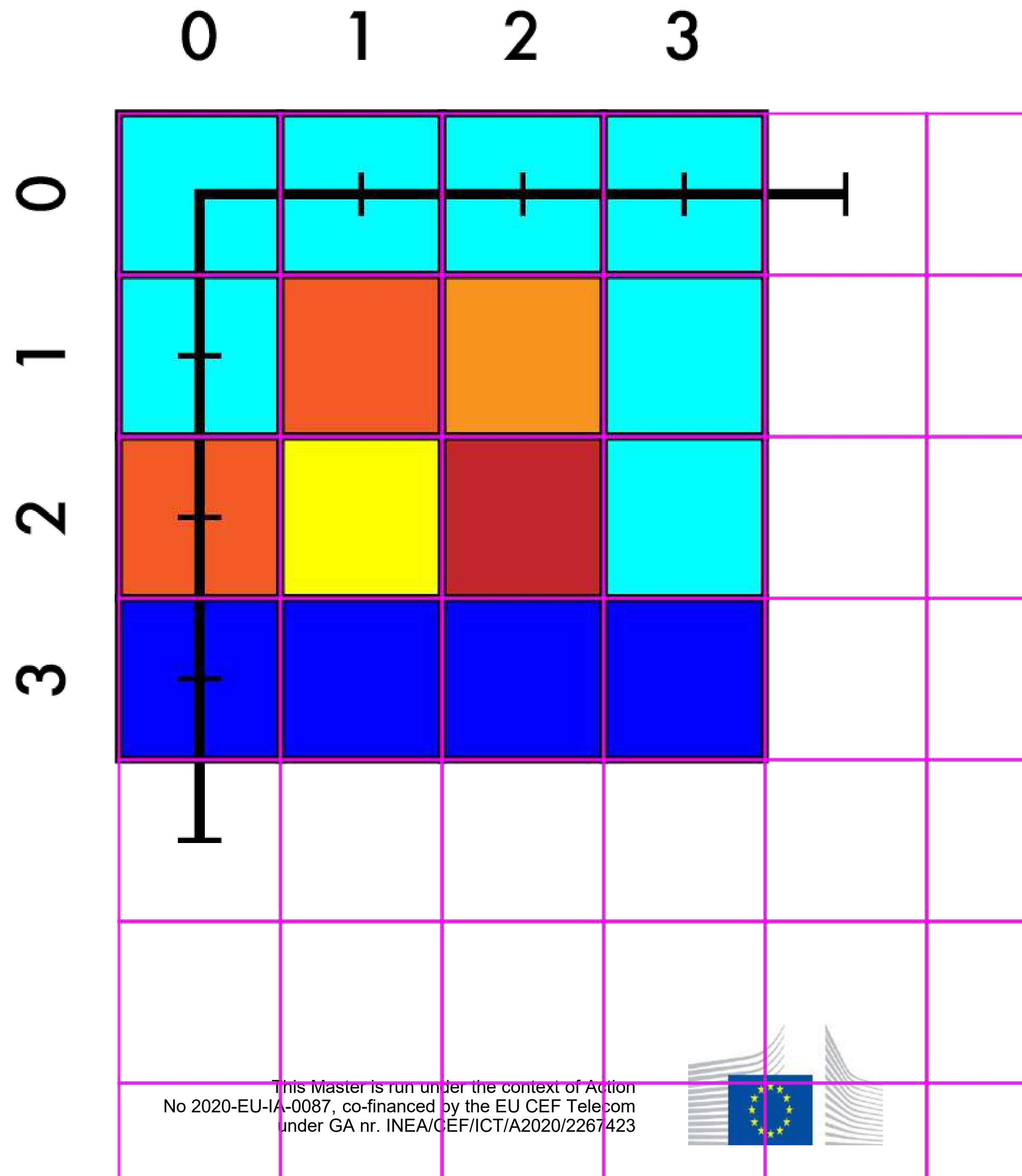
This is a beautiful image of a sunset… it’s just very small…

Say we want to increase size 4x4 -> 7x7



Resize 4x4 -> 7x7

- Create our new image

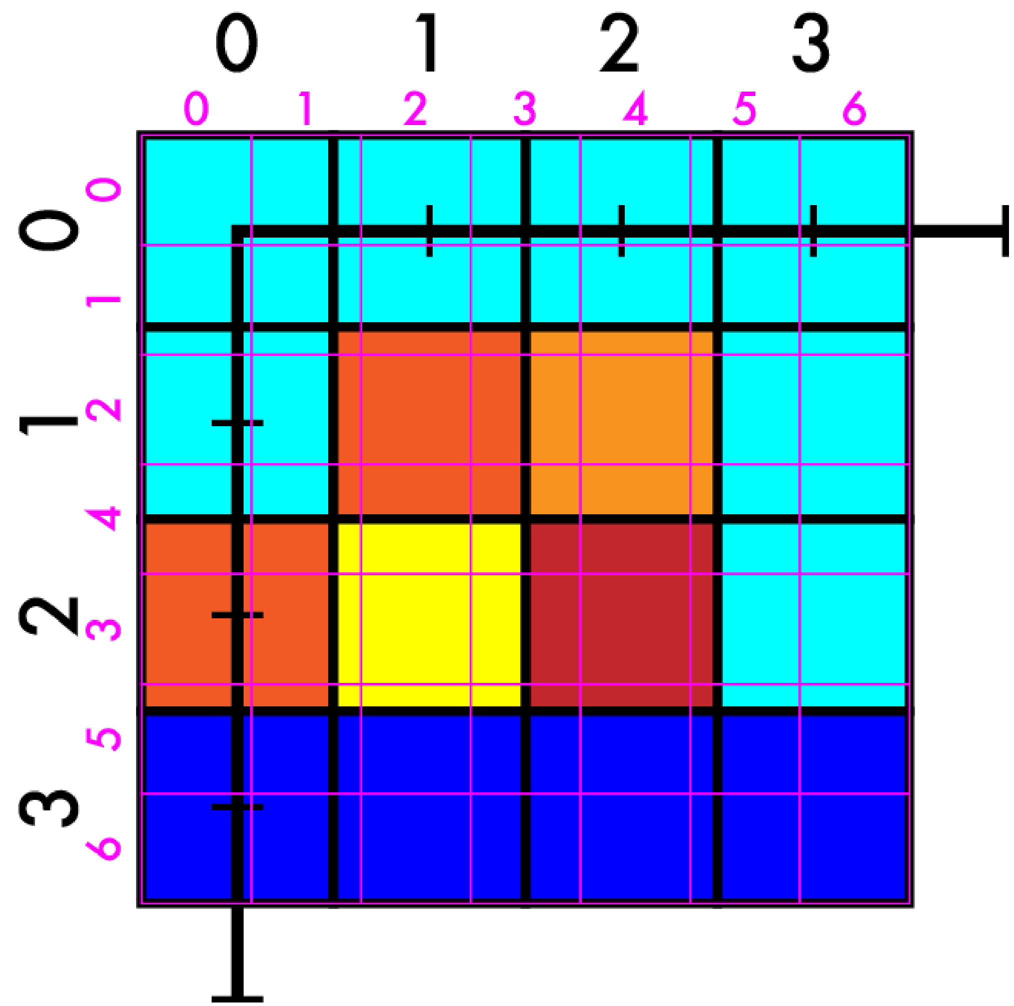


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

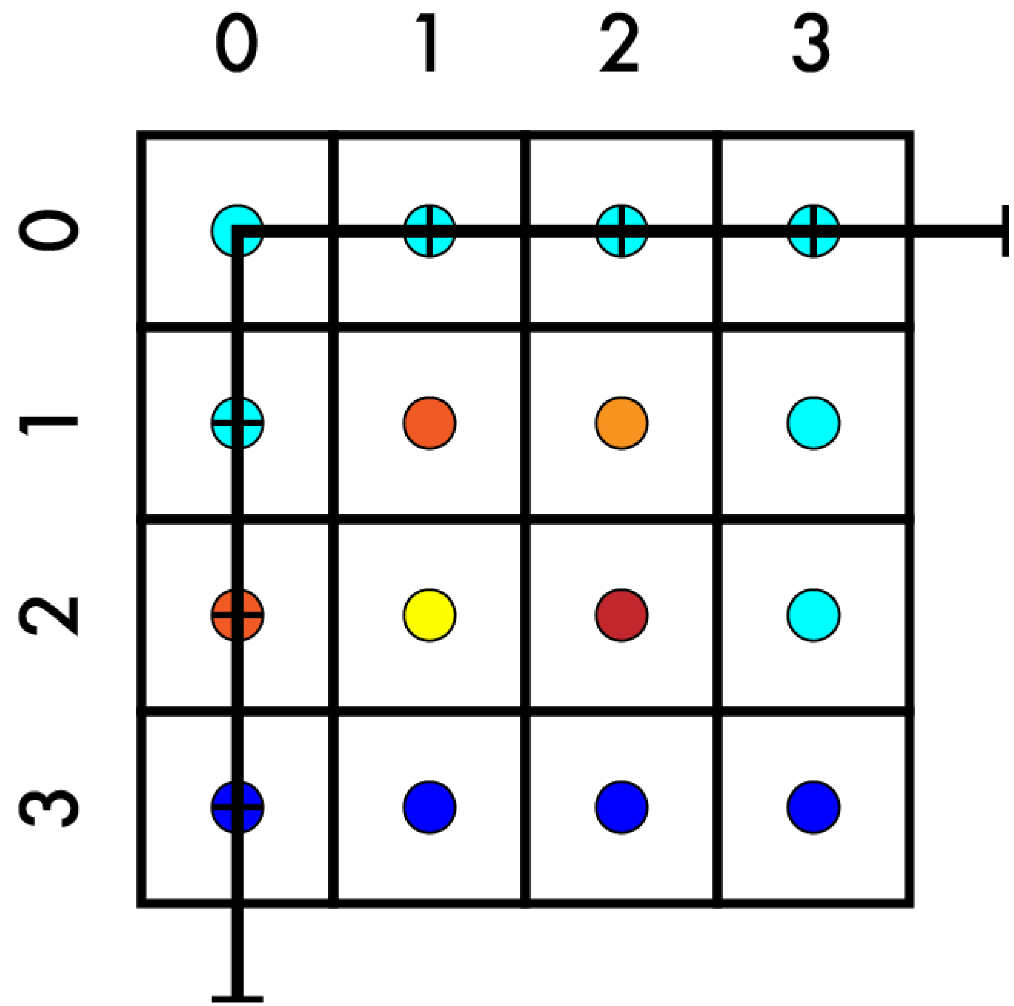
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates



Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates

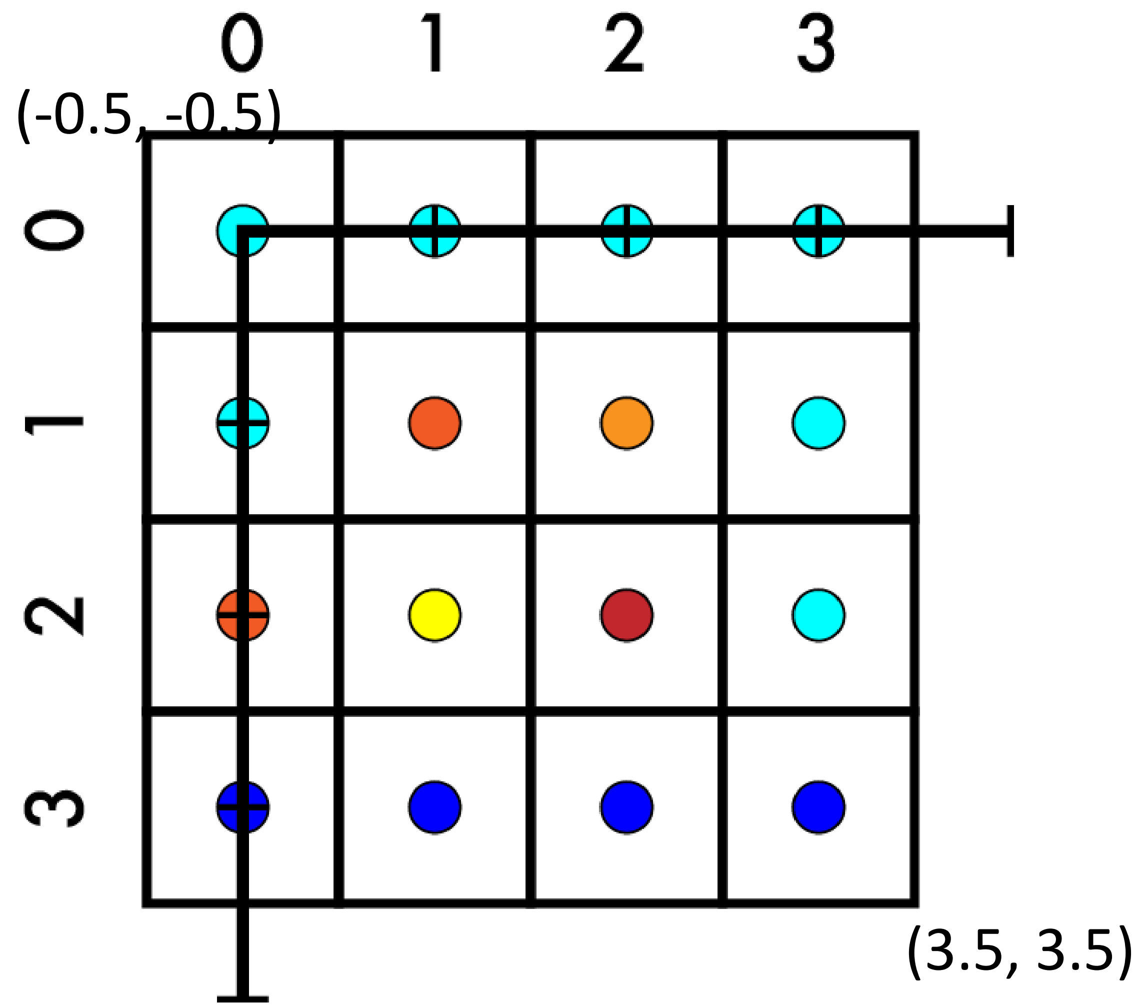


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates

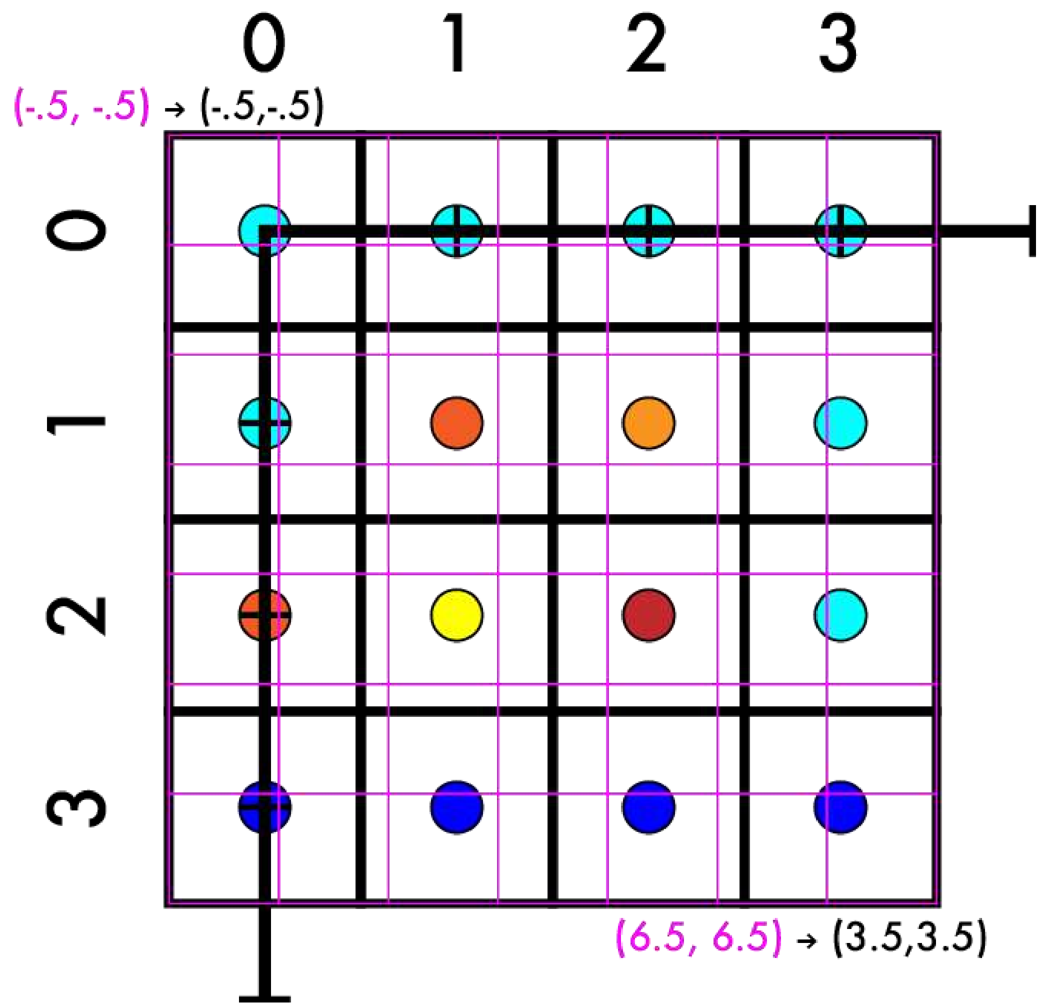


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

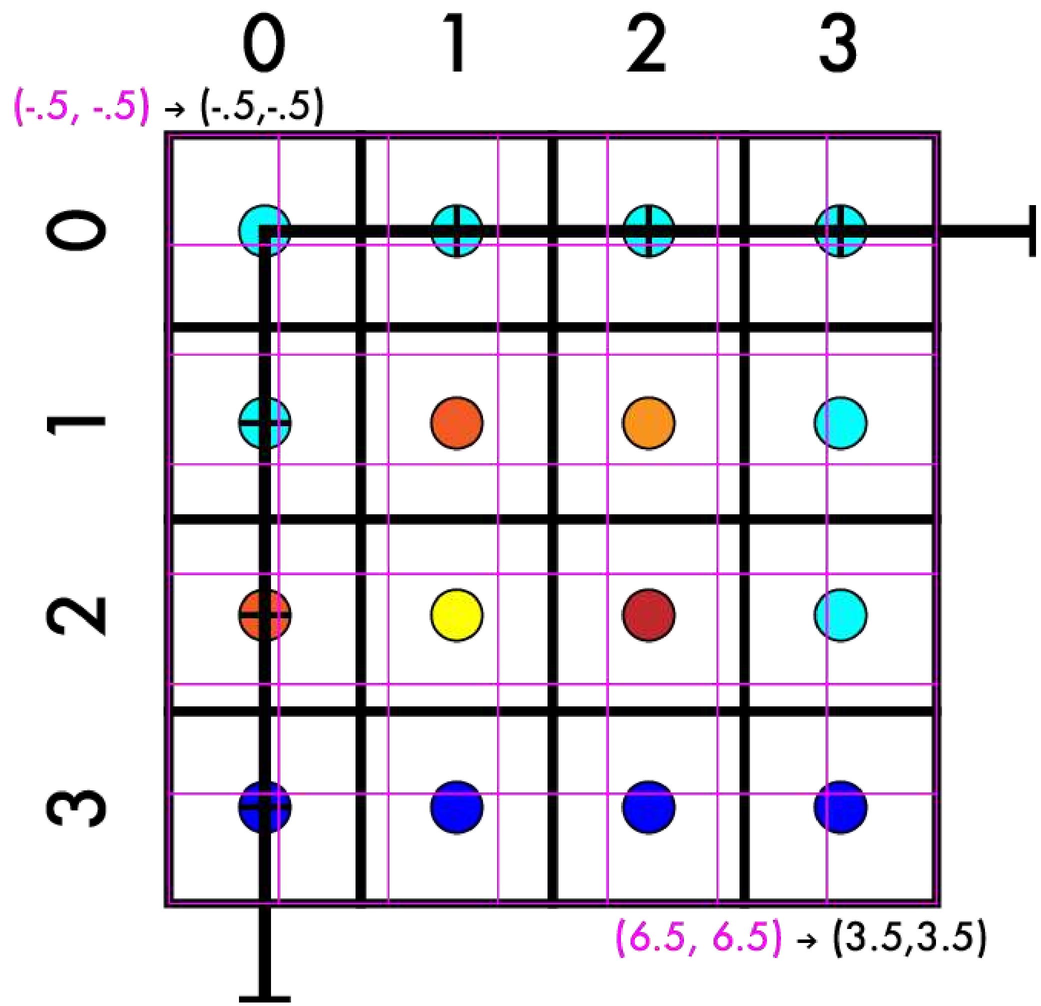
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5



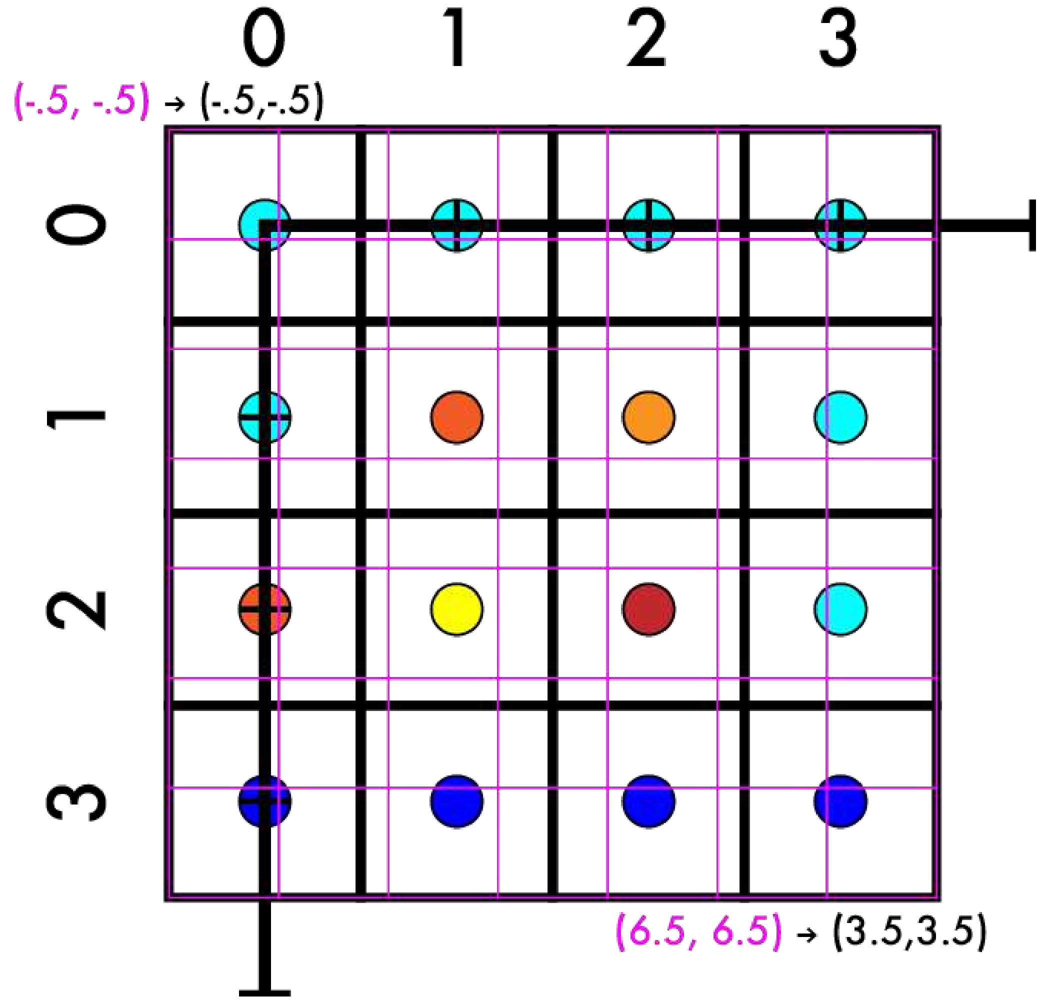
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
    - System of equations
    - aX + b = Y
    - a*-.5 + b = -.5
    - a*6.5 + b = 3.5
        - a*7 = 4



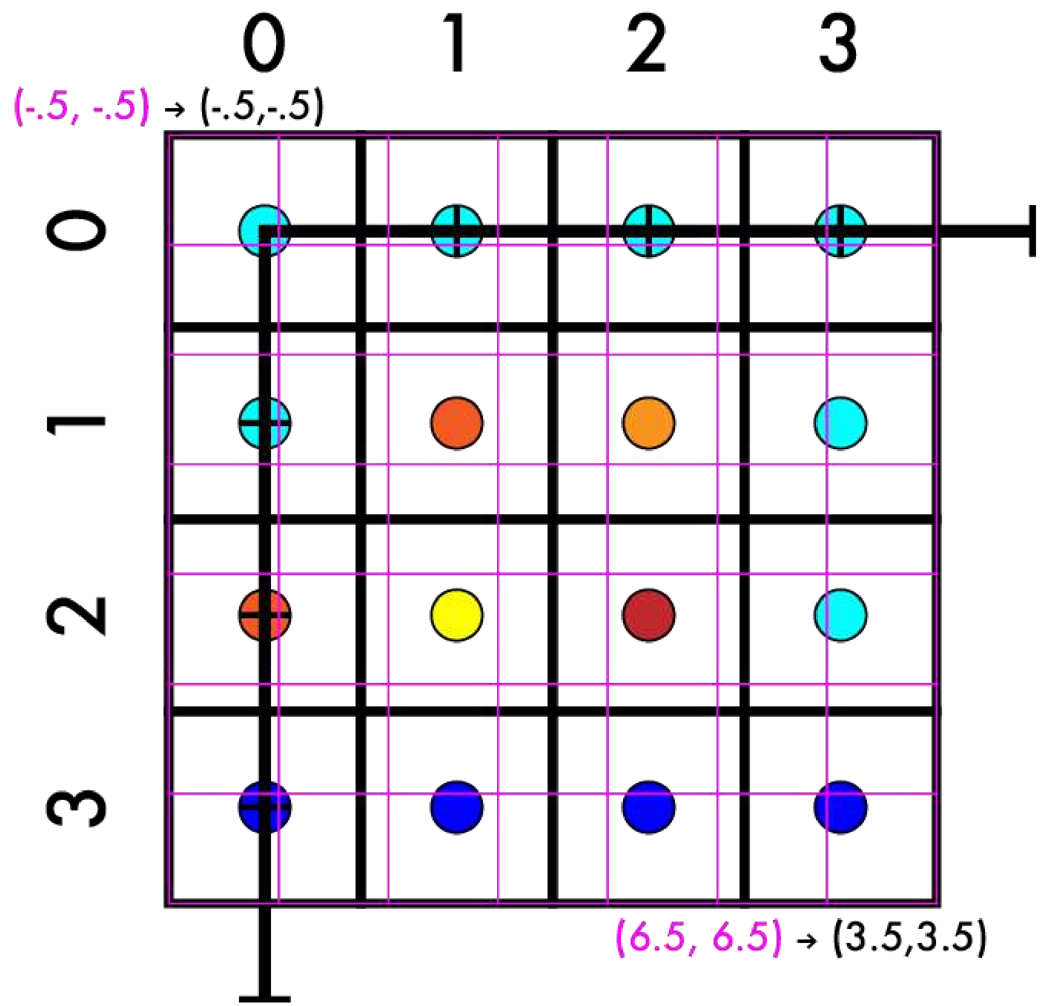
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5
    - a*7 = 4
    - a = 4/7



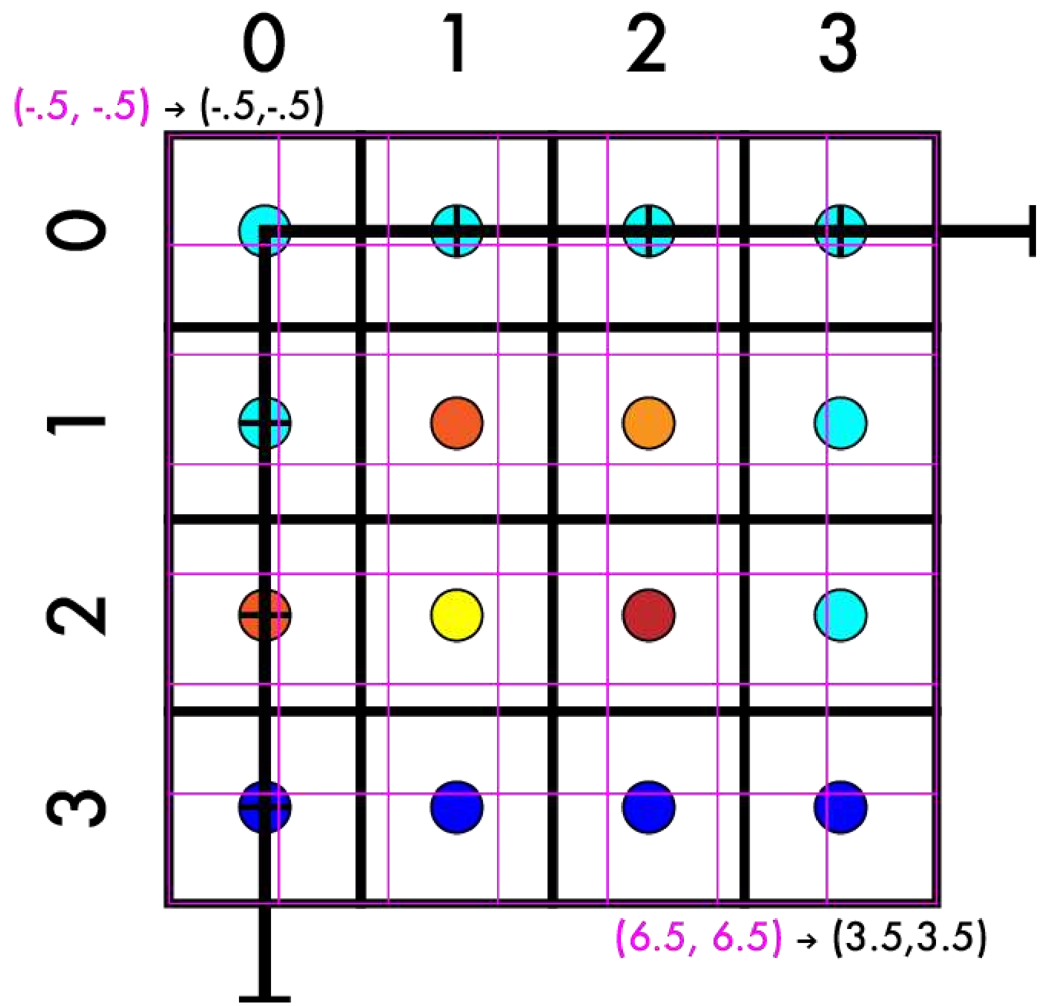
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5
  - a = 4/7



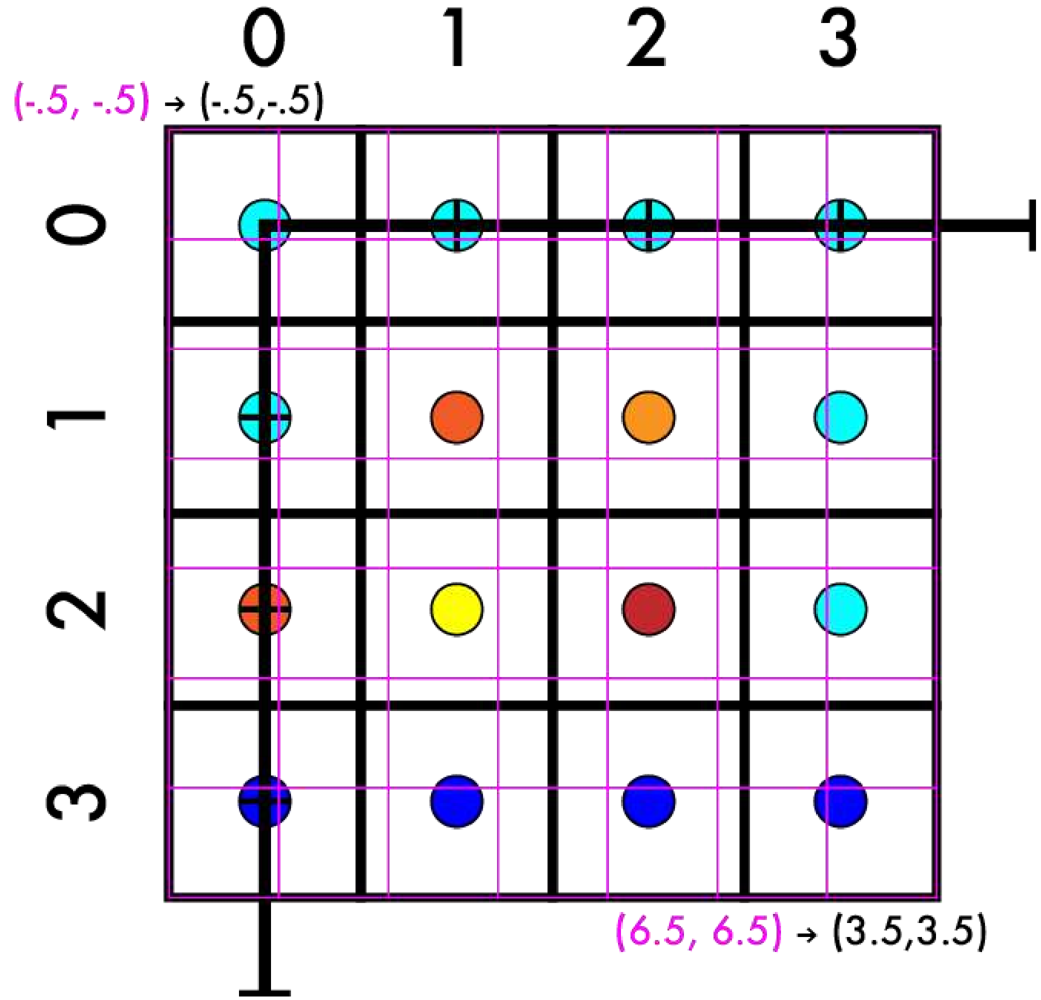
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5
  - a = 4/7
    - a*-.5 + b = -.5



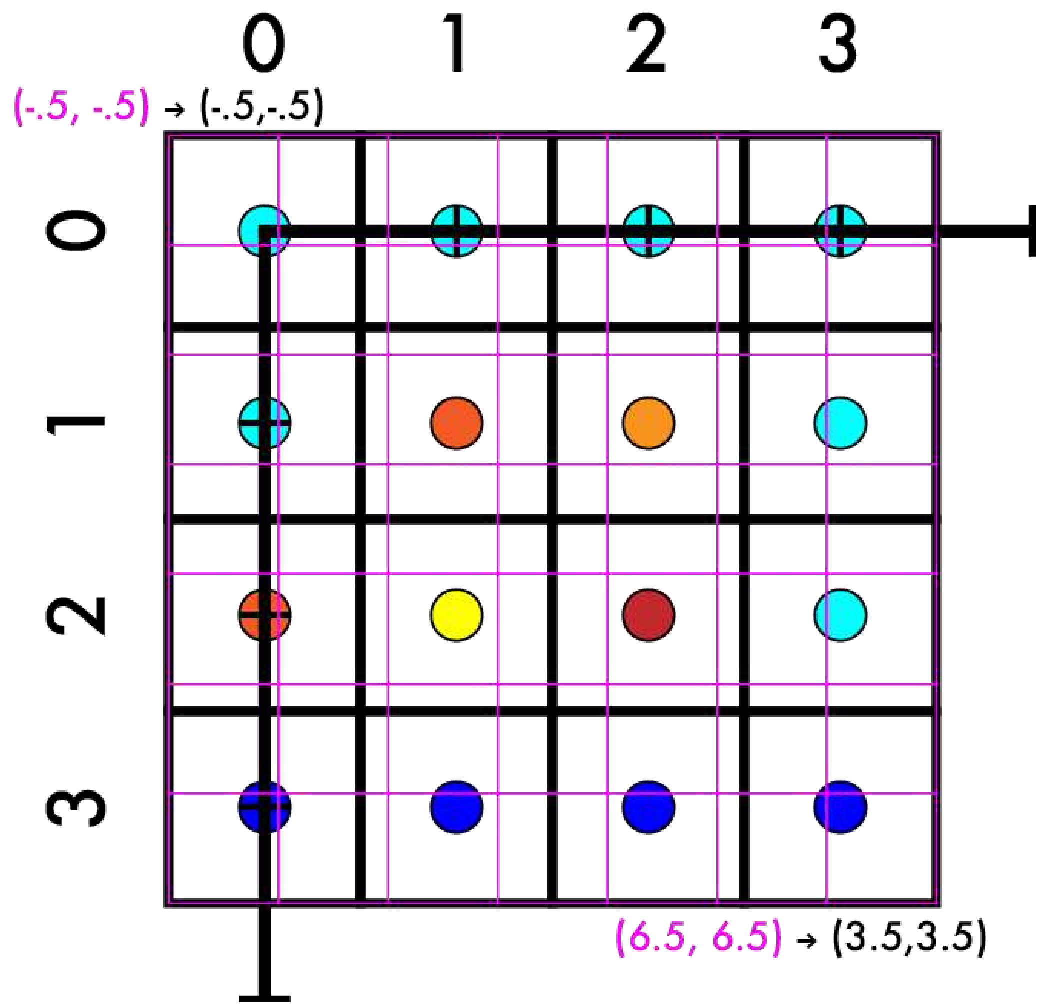
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5
  - a = 4/7
    - a*-.5 + b = -.5
    - 4/7*-1/2 + b = -1/2



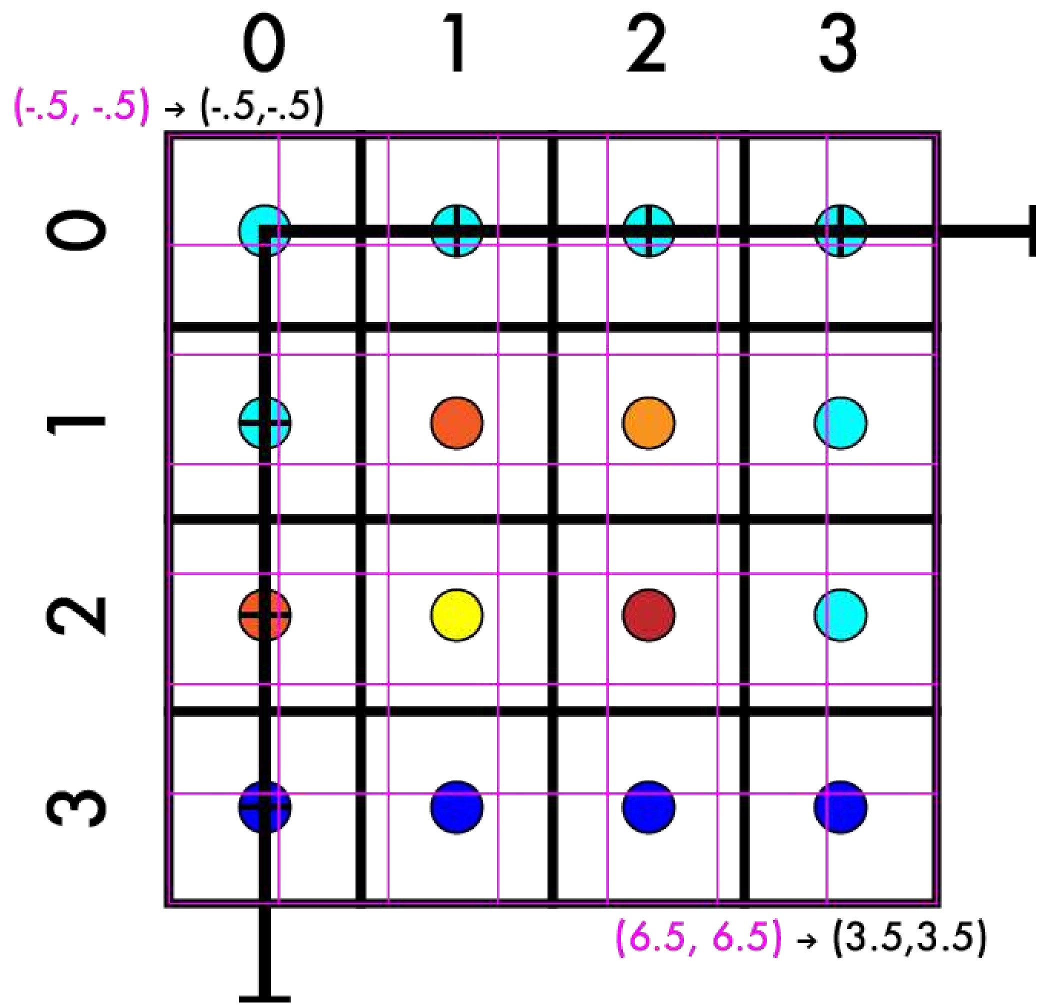
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5
  - a = 4/7
    - a*-.5 + b = -.5
    - 4/7*-1/2 + b = -1/2
    - -4/14 + b = -7/14



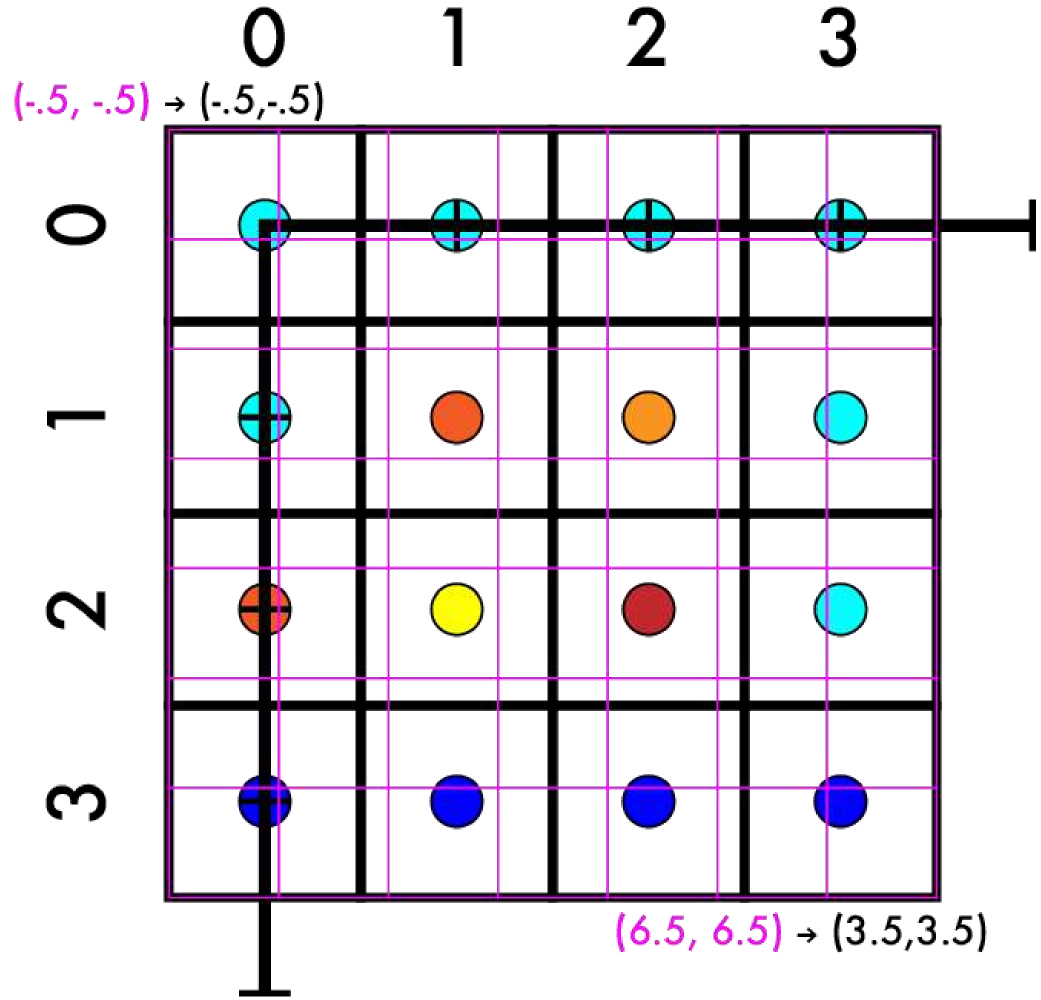
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5
  - a = 4/7
    - a*-.5 + b = -.5
    - 4/7*-1/2 + b = -1/2
    - -4/14 + b = -7/14
    - b = -3/14



Resize 4x4 -> 7x7

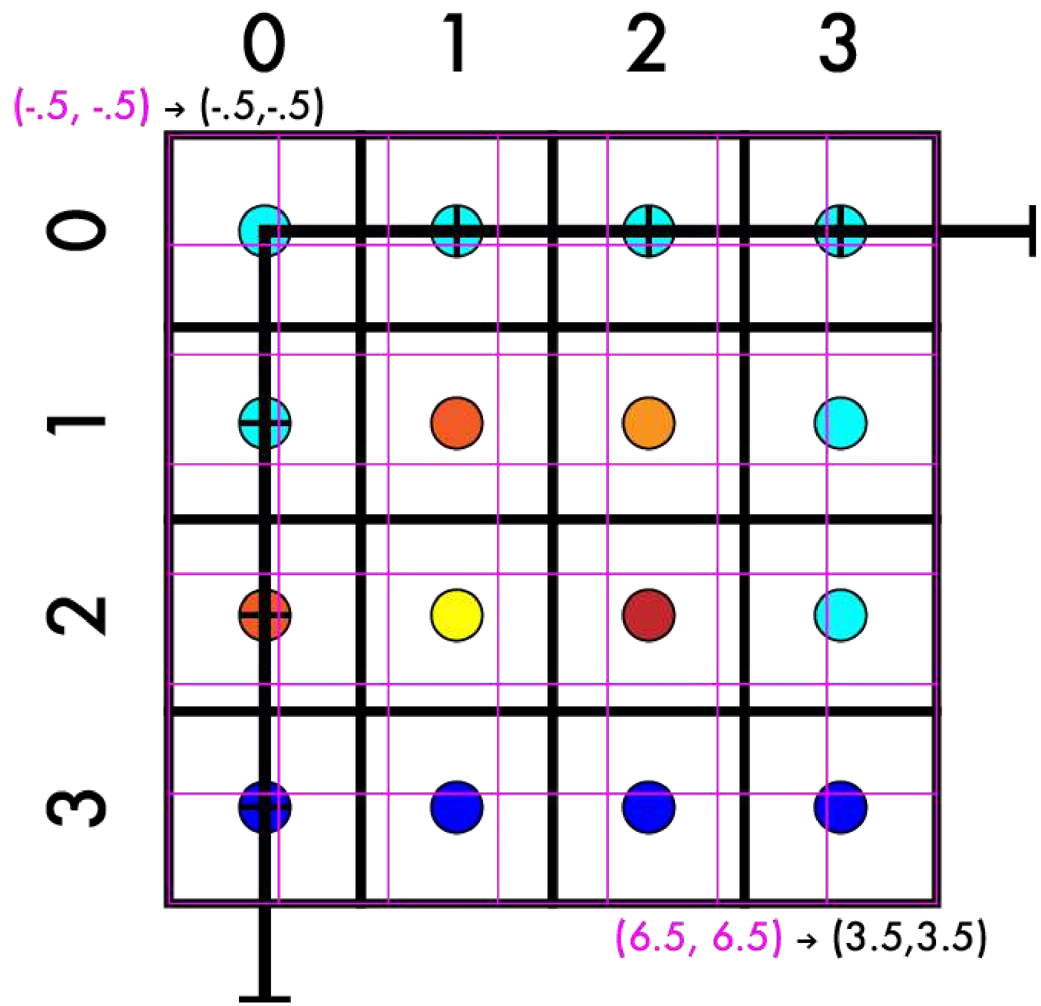
- Create our new image
- Match up coordinates
  - System of equations
  - aX + b = Y
  - a*-.5 + b = -.5
  - a*6.5 + b = 3.5
  - a = 4/7
  - b = -3/14



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

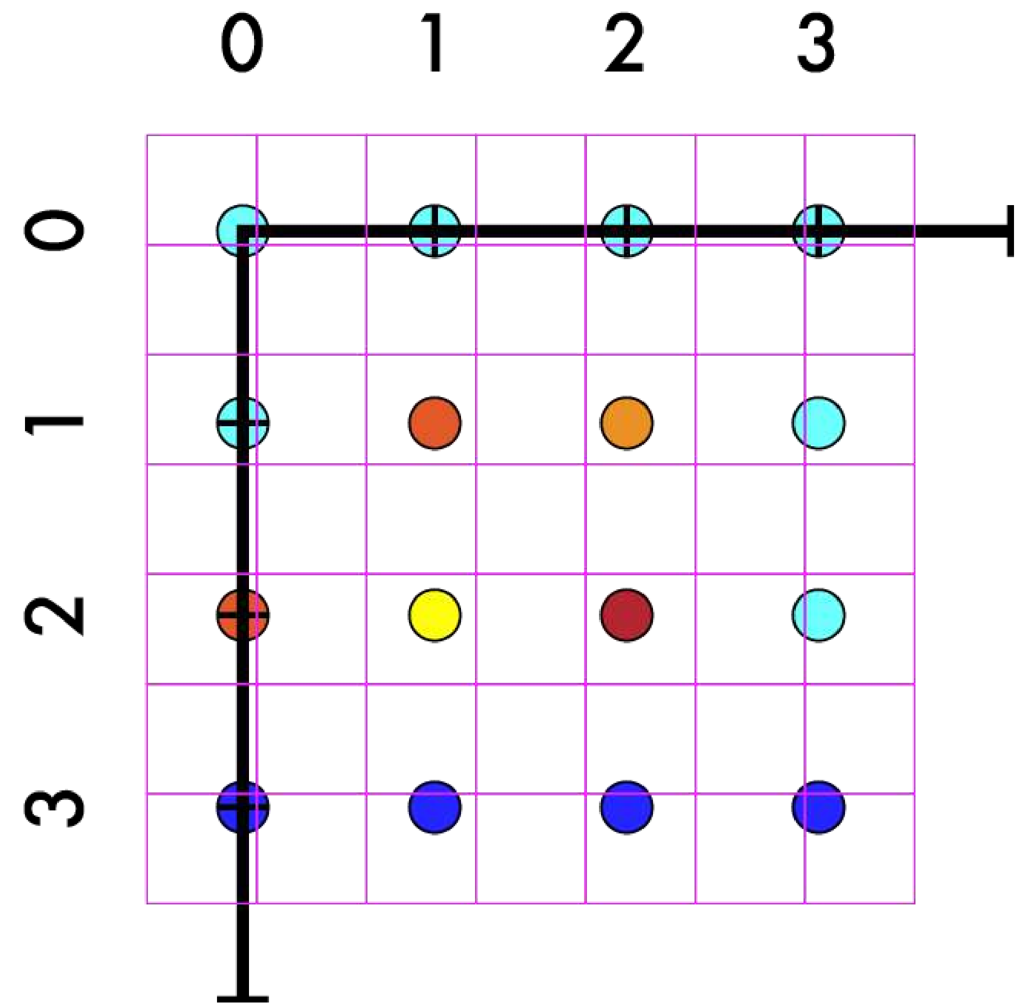
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - $4/7\ X - 3/14 = Y$



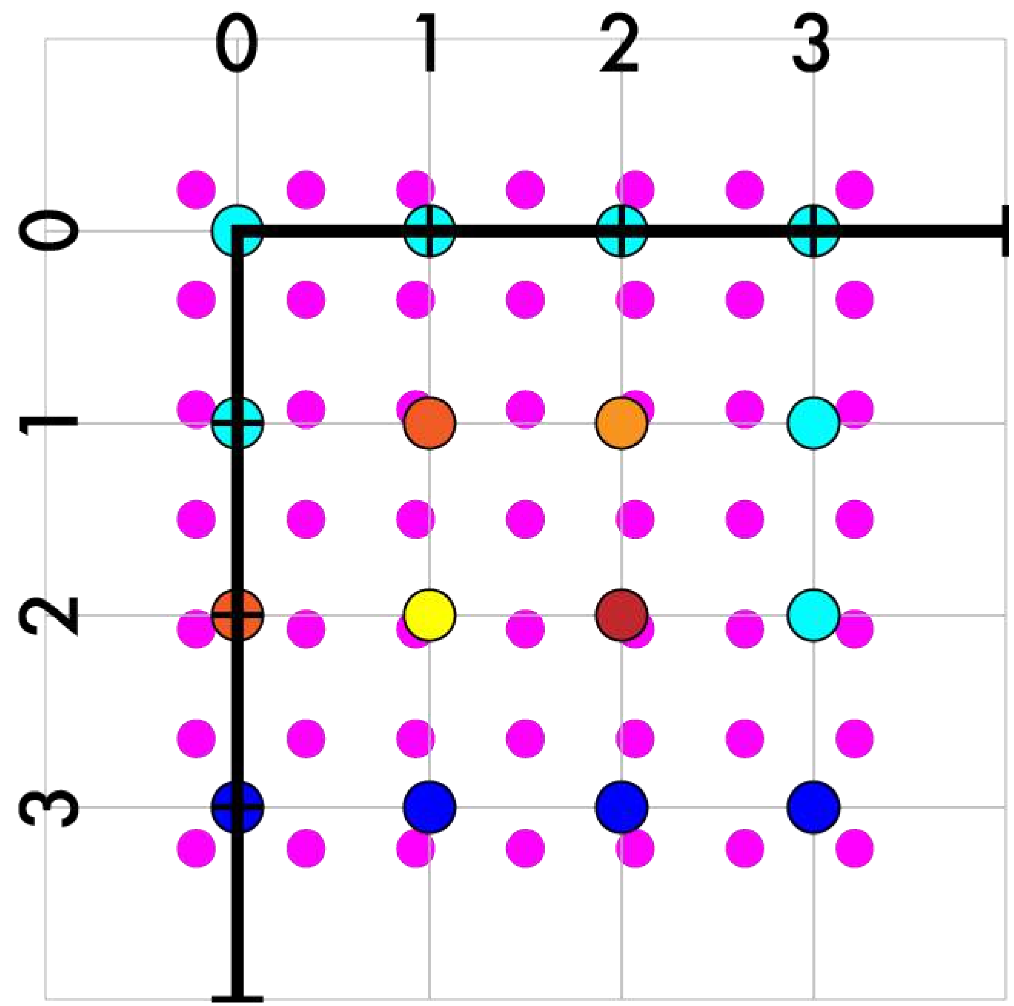
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y



Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts

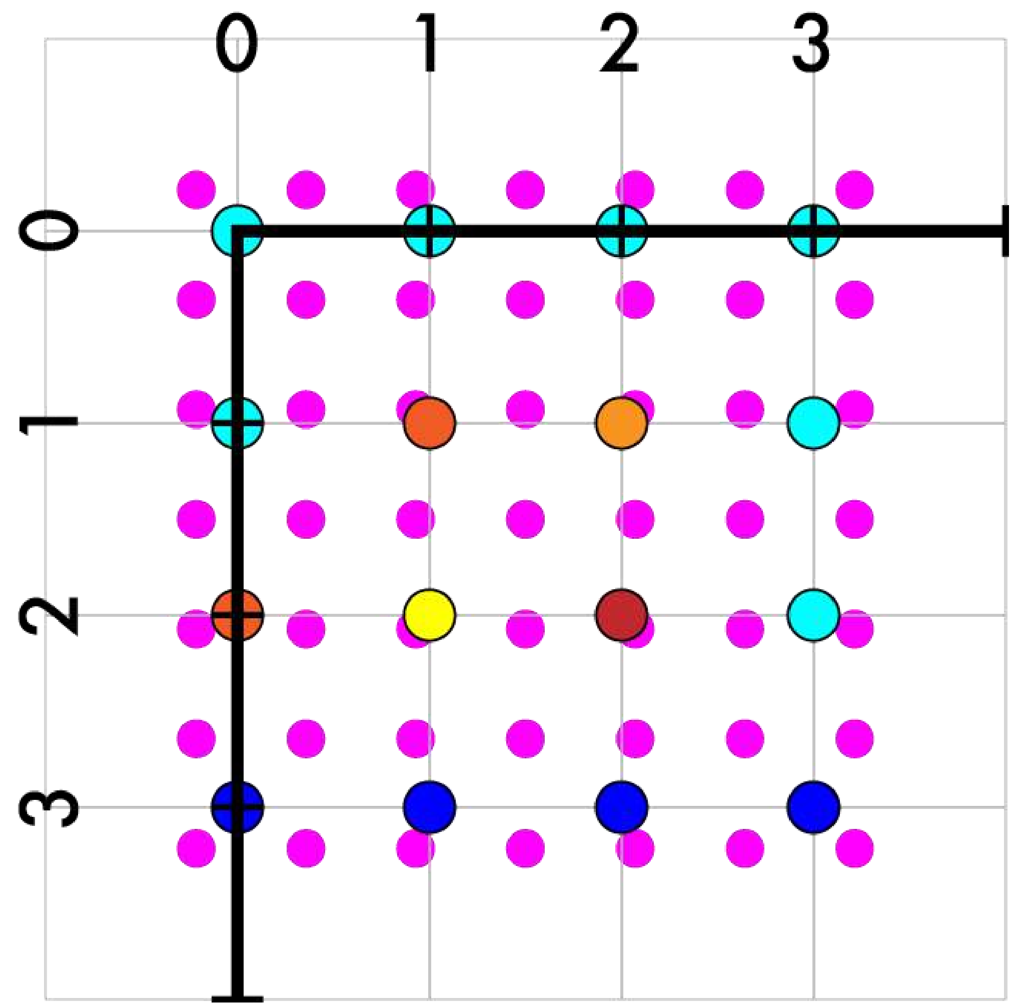


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

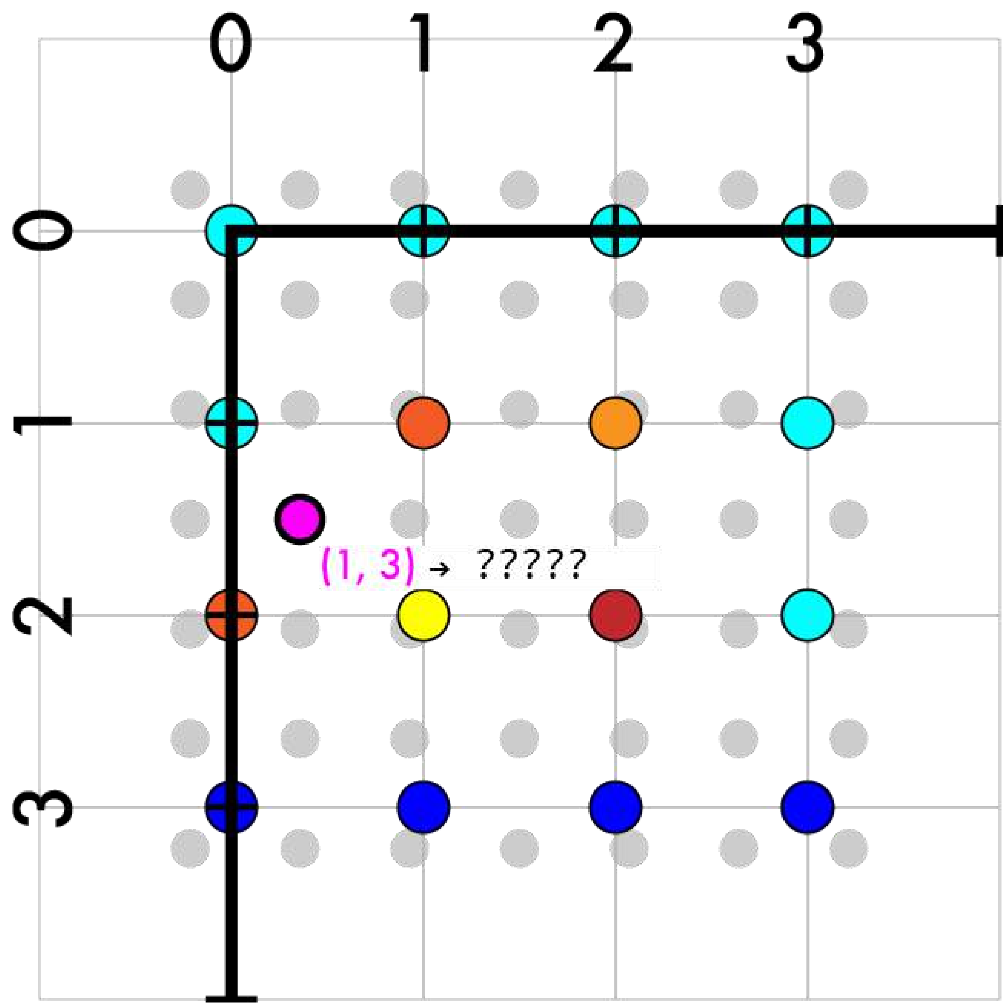
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords



Resize 4x4 -> 7x7

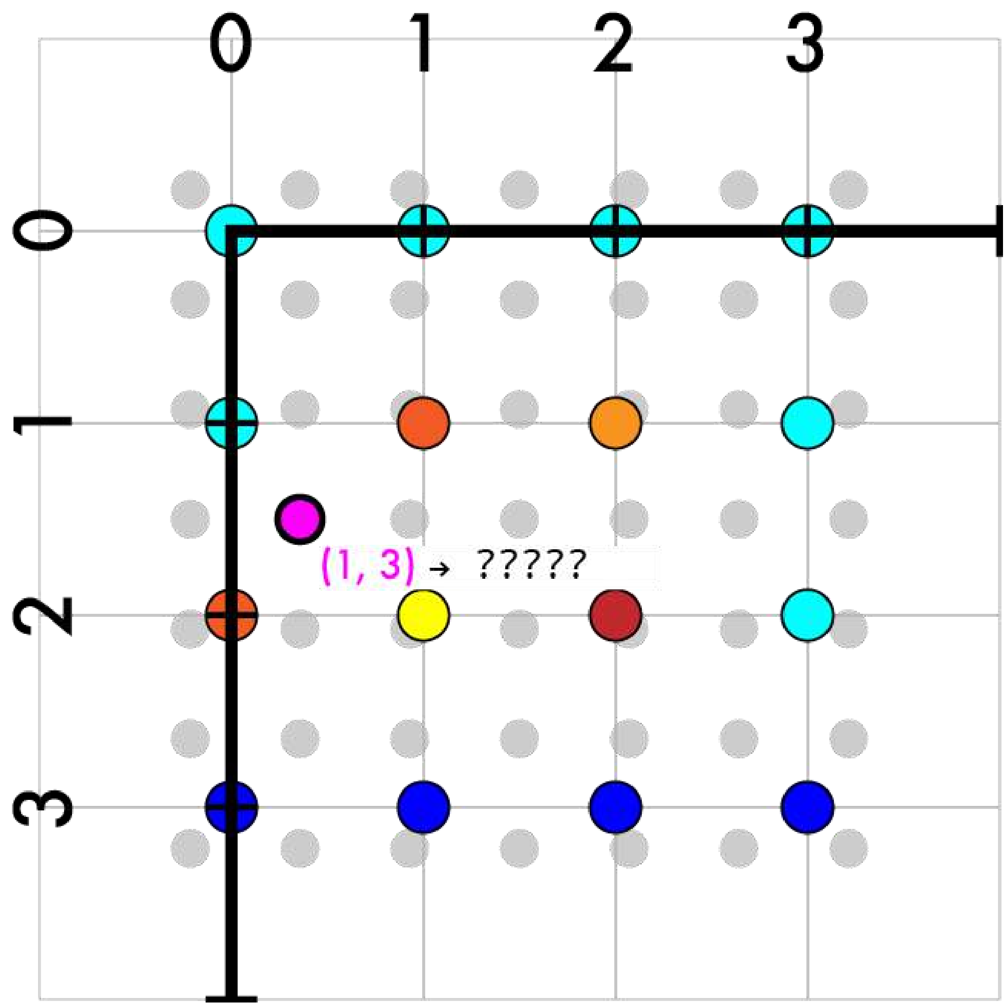
- Create our new image
- Match up coordinates
    - 4/7 X - 3/14 = Y
- Iterate over new pts
    - Map to old coords
    - (1, 3)



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

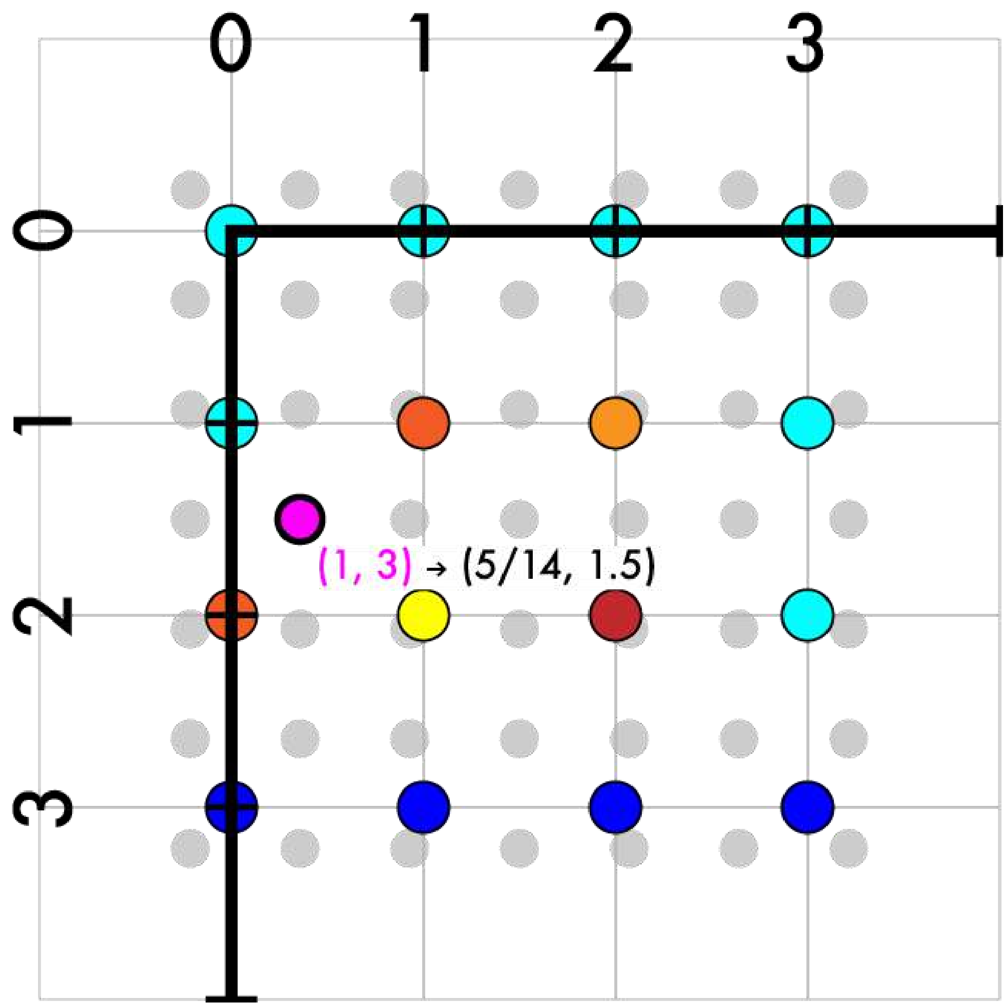
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3)
  - 4/7*1 - 3/14
  - 4/7*3 - 3/14



Resize 4x4 -> 7x7

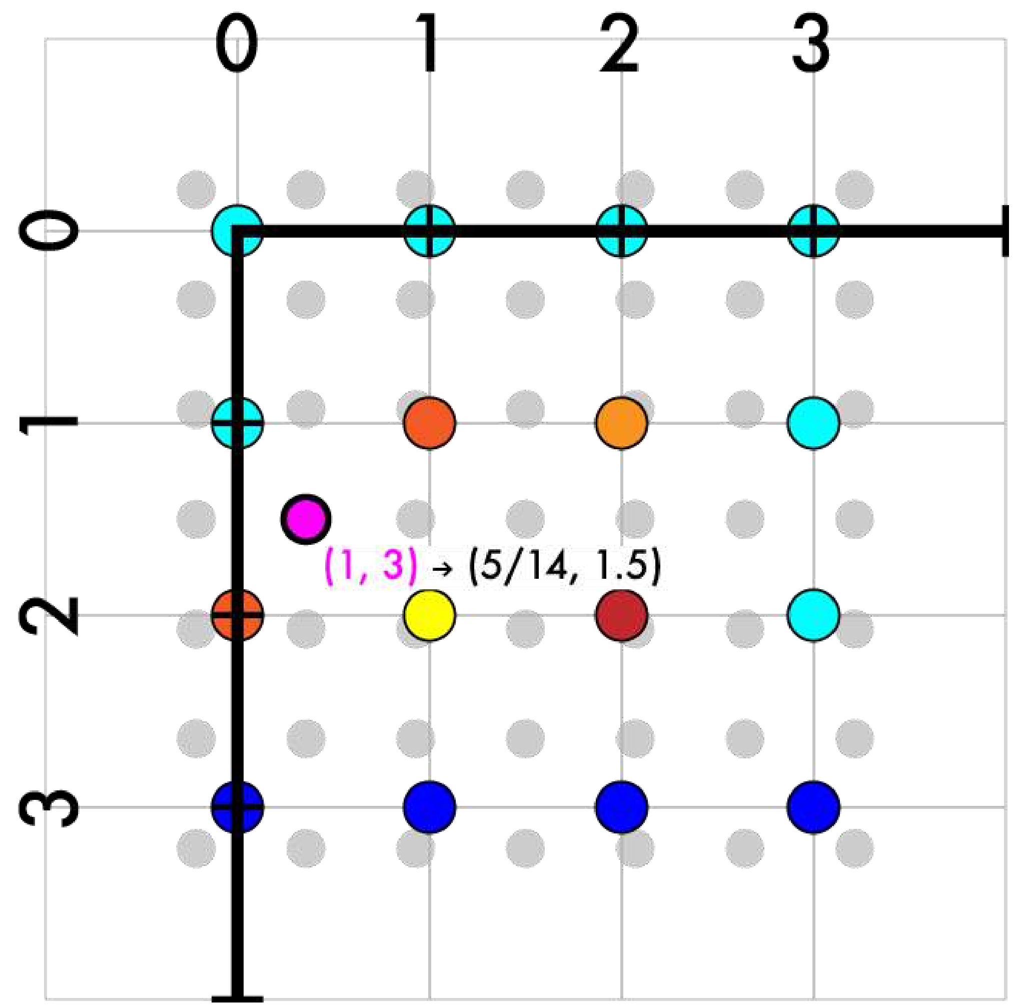
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3)
  - 4/7*1 - 3/14
  - 4/7*3 - 3/14
  - (5/14, 21/14)



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

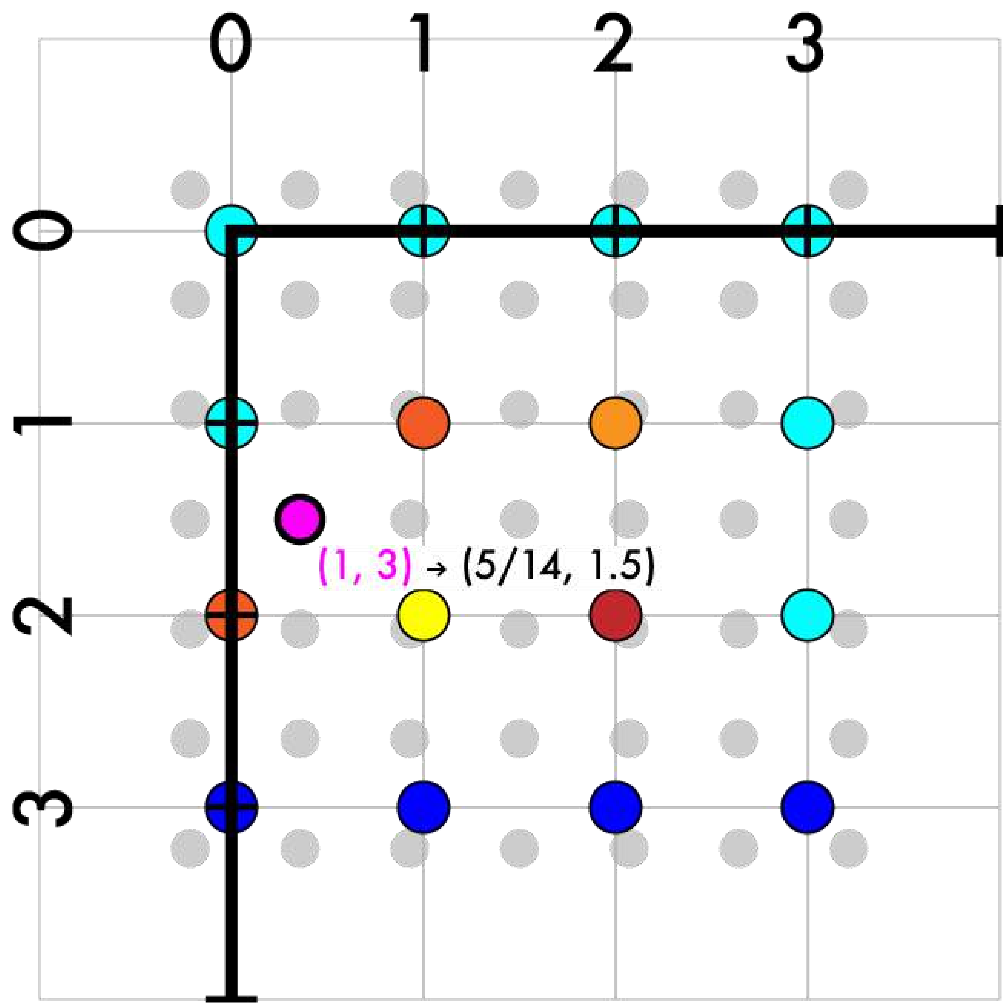
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)



Resize 4x4 -> 7x7

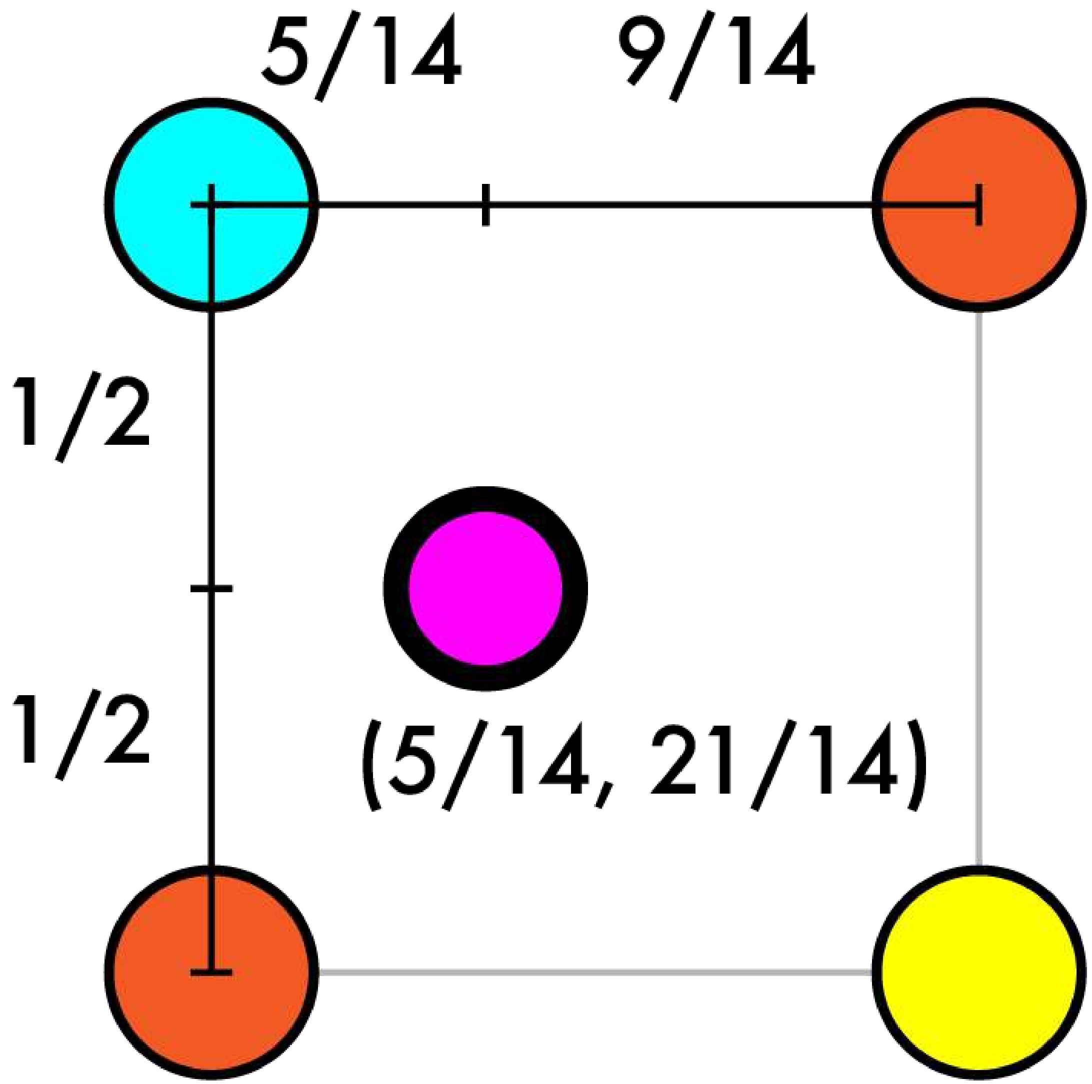
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Resize 4x4 -> 7x7

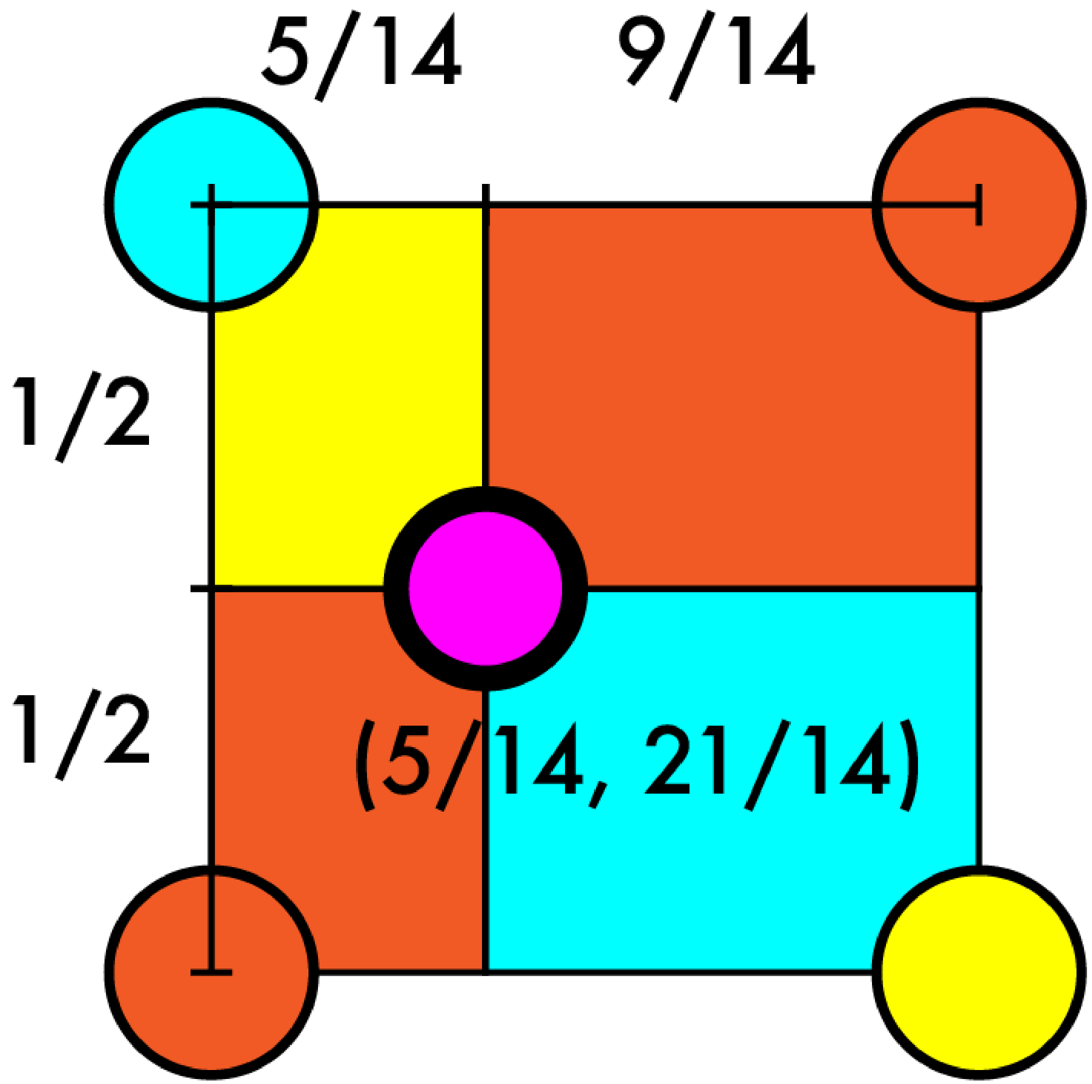
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

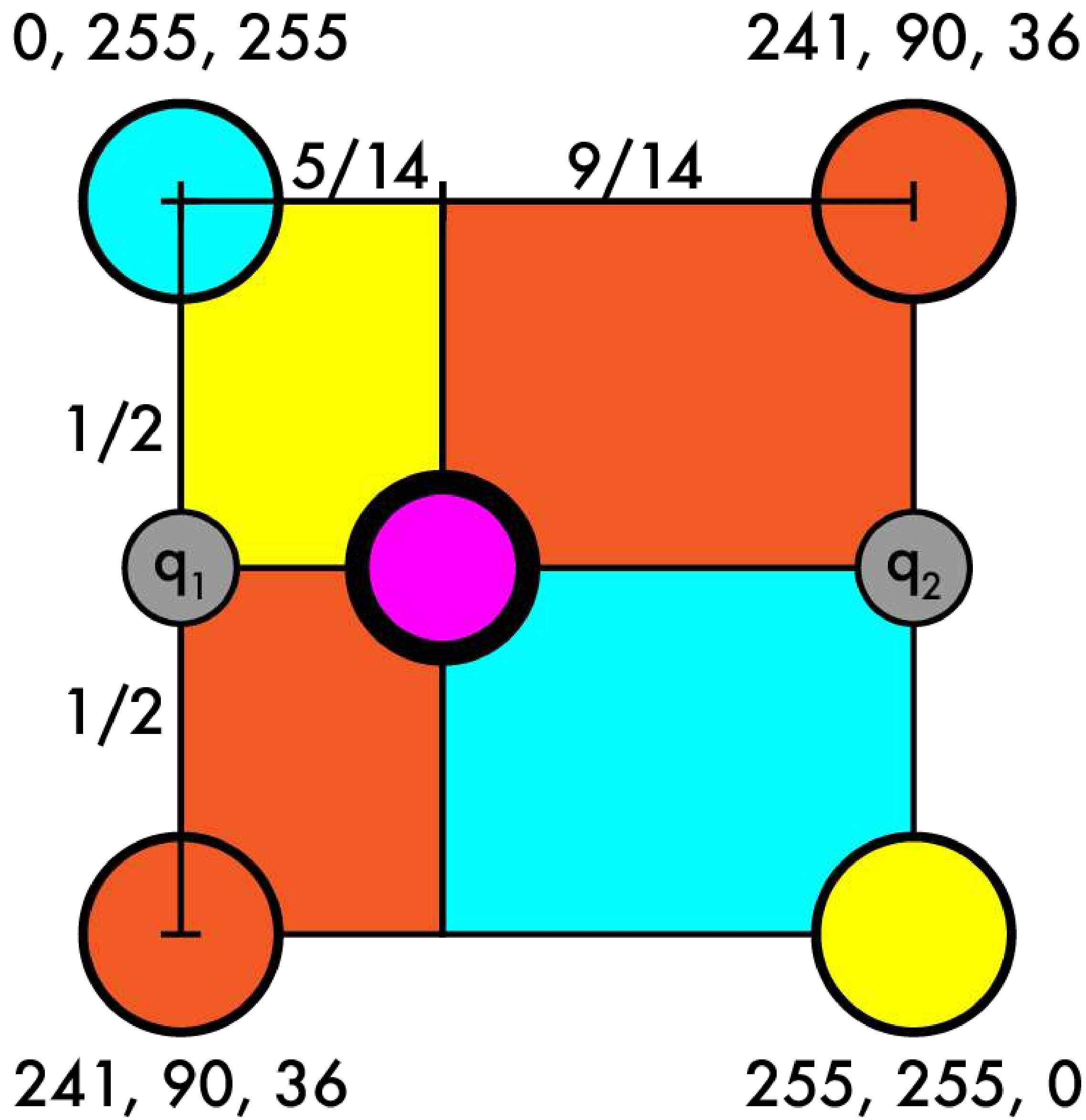
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - Size of opposite rects



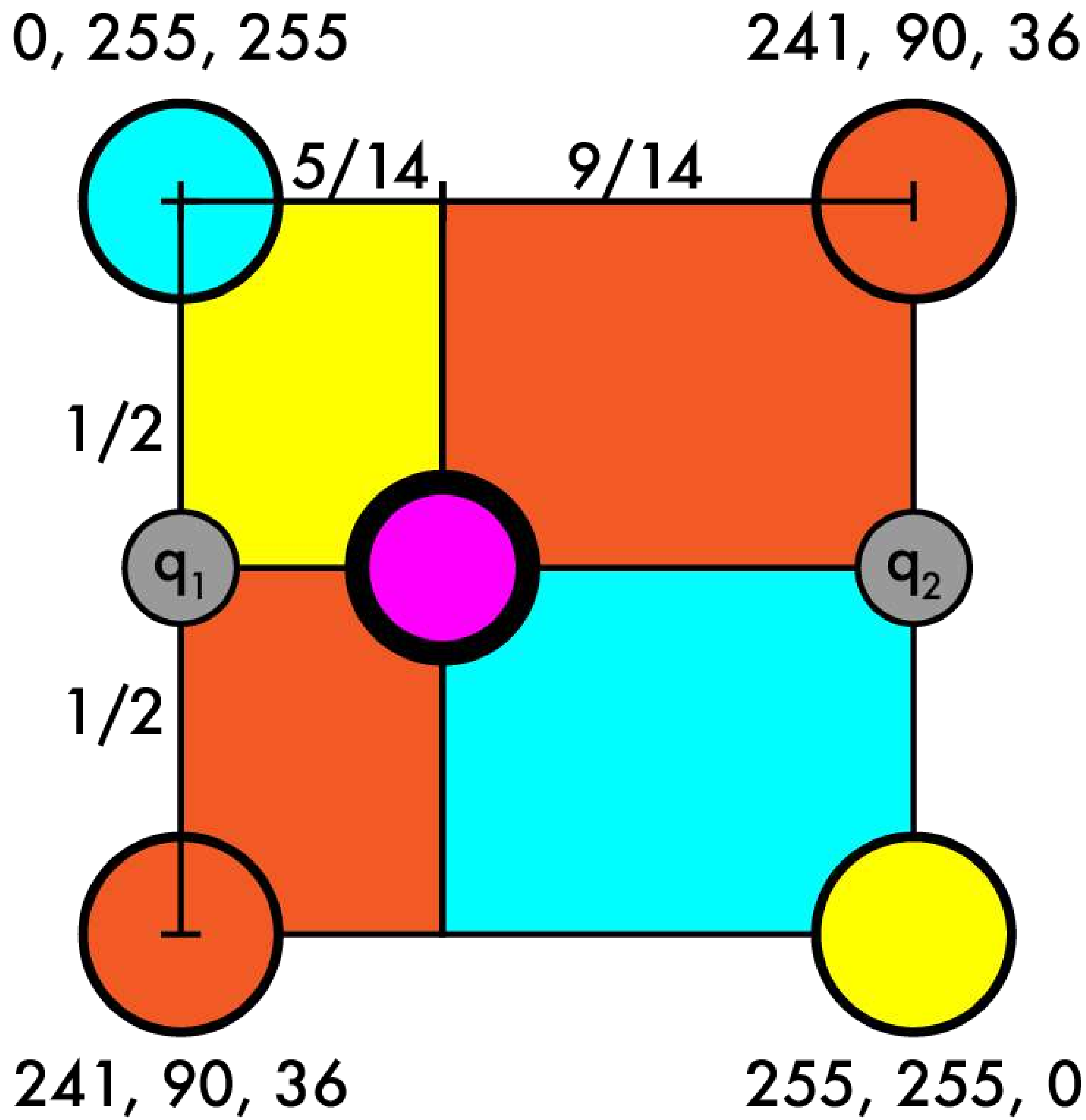
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - Size of opposite rects
    - OR find q1 and q2, then interpolate between them



Resize 4x4 -> 7x7

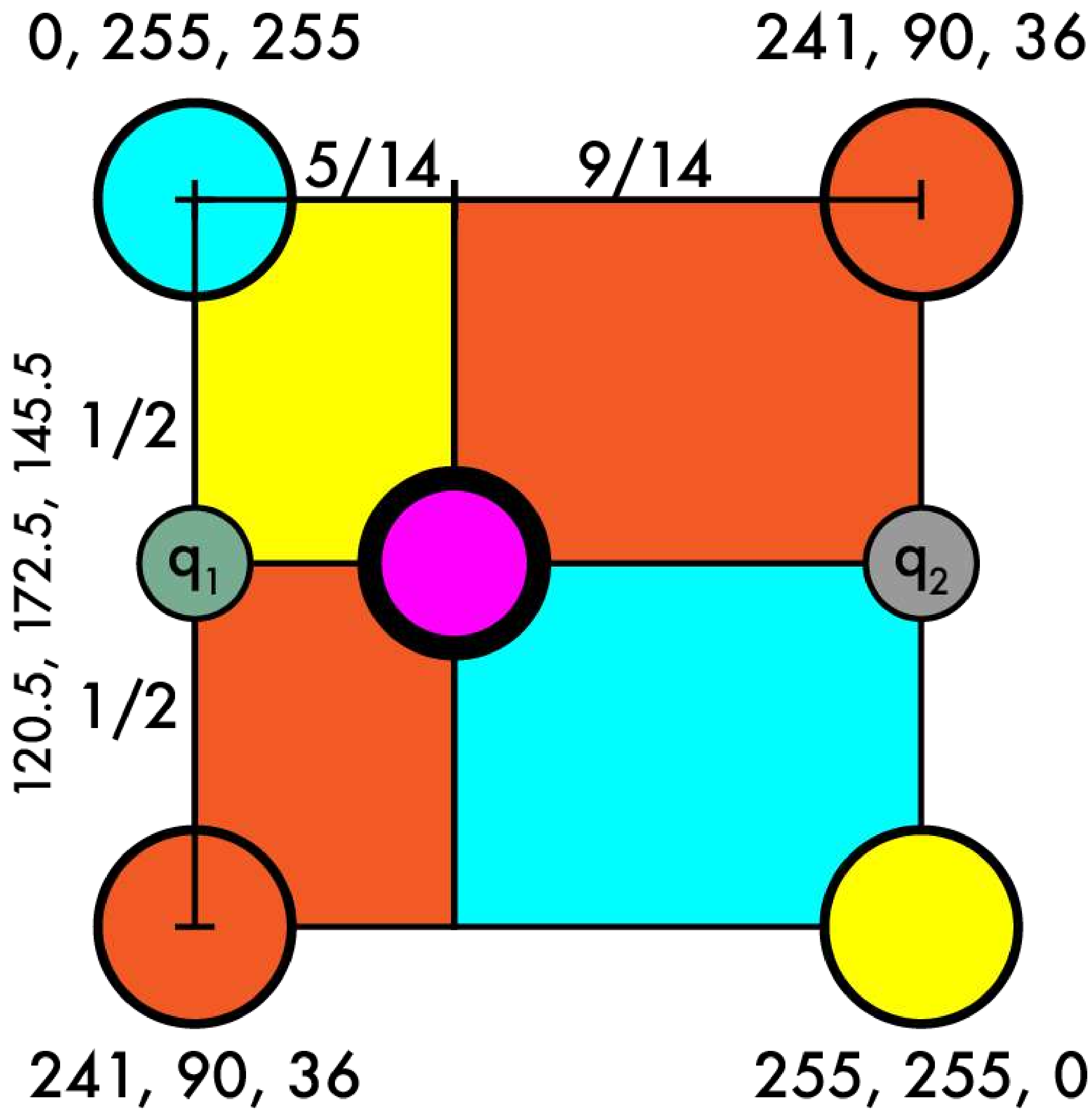
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q1 = r1, g1, b1
    - r1 = .5*0 + .5*241
    - g1 = .5*255 + .5*90
    - b1 = .5*255 + .5*36



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

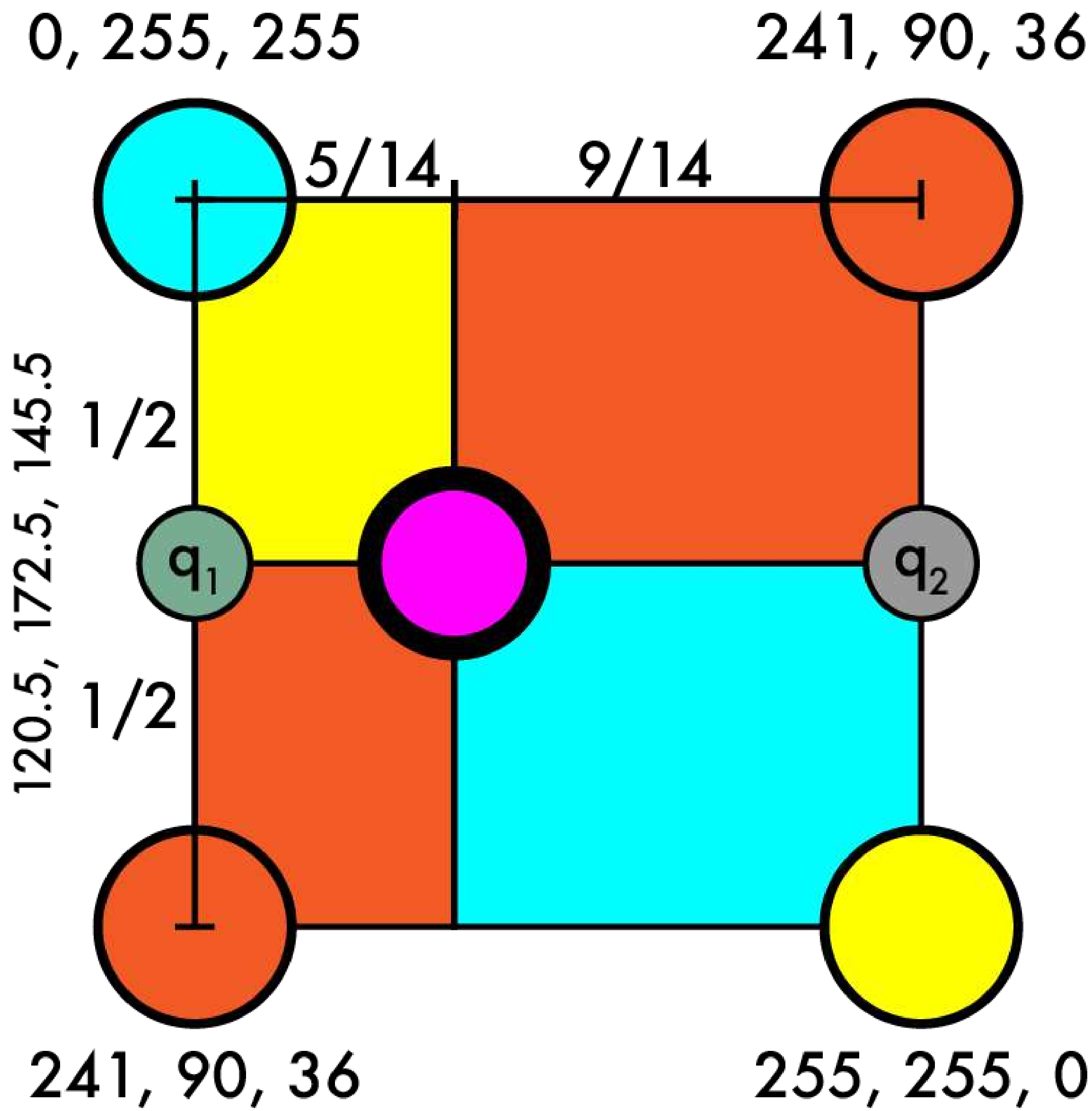
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q1 = (120.5, 172.5, 145.5)
    - q2 = r2, g2, b2



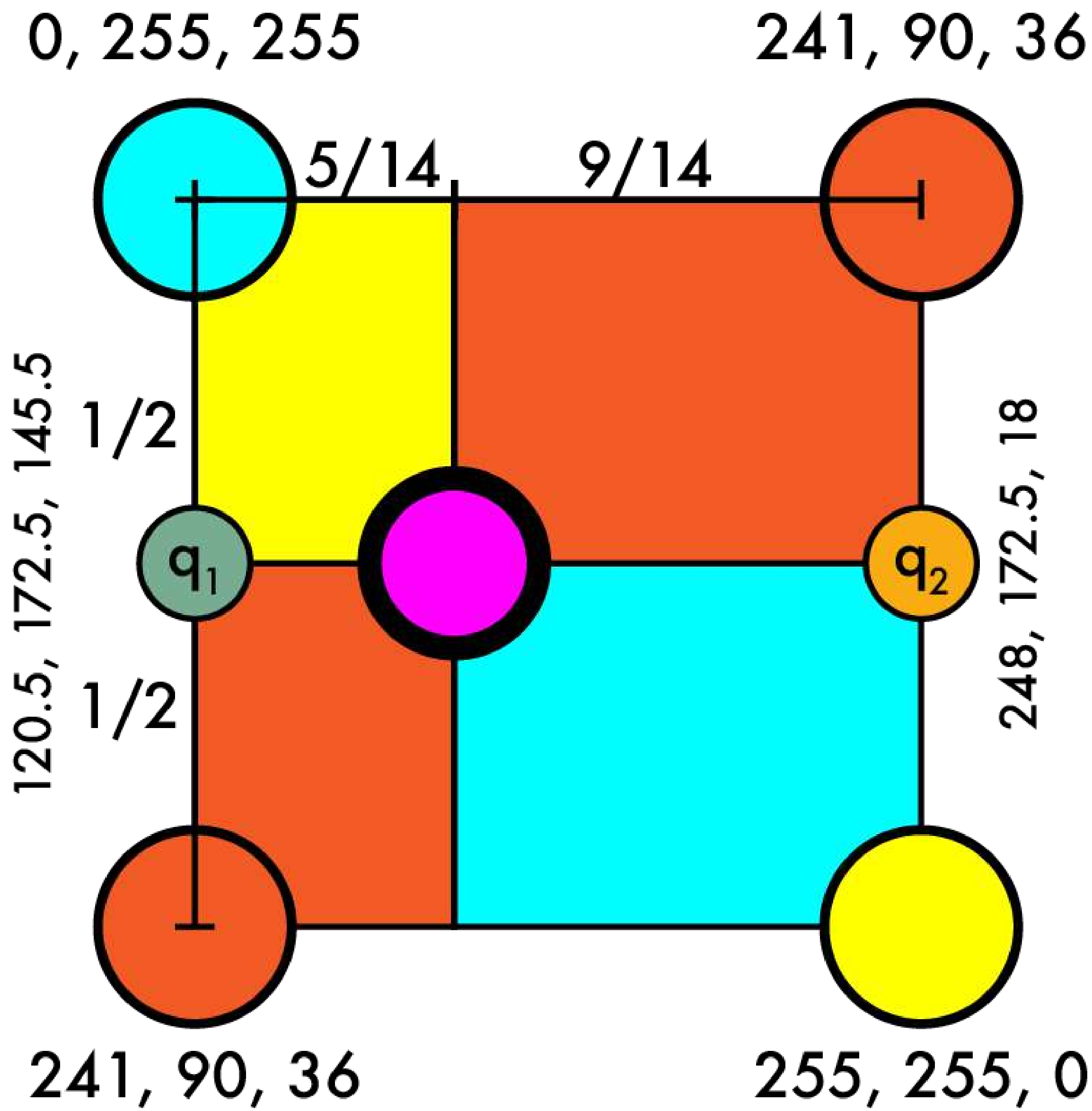
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q1 = (120.5, 172.5, 145.5)
    - q2 = r2, g2, b2
    - r2 = .5*241 + .5*255
    - g2 = .5*90 + .5*255
    - b2 = .5*36 + .5*0



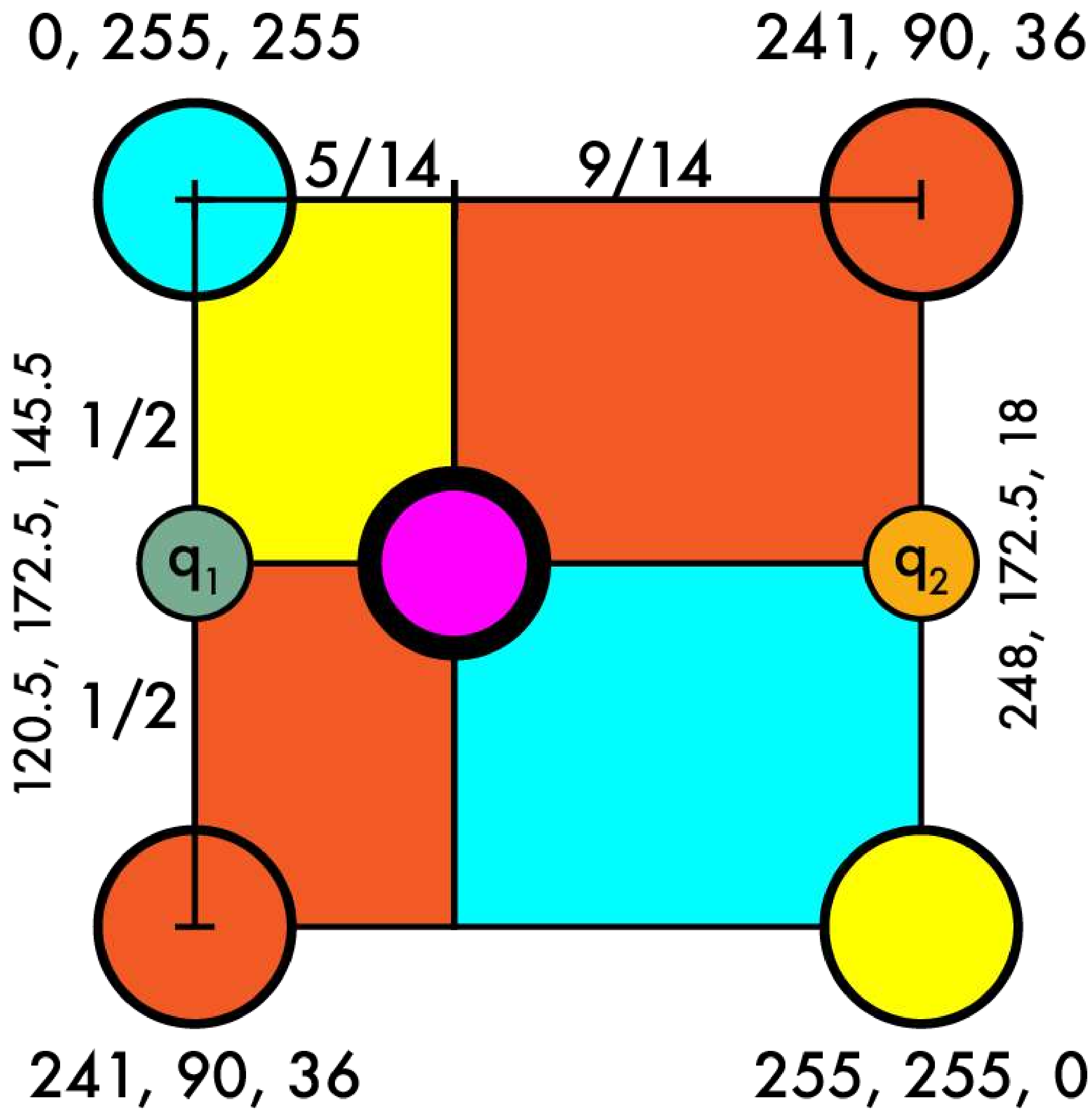
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q1 = (120.5, 172.5, 145.5)
    - q2 = (248, 172.5, 18)



Resize 4x4 -> 7x7

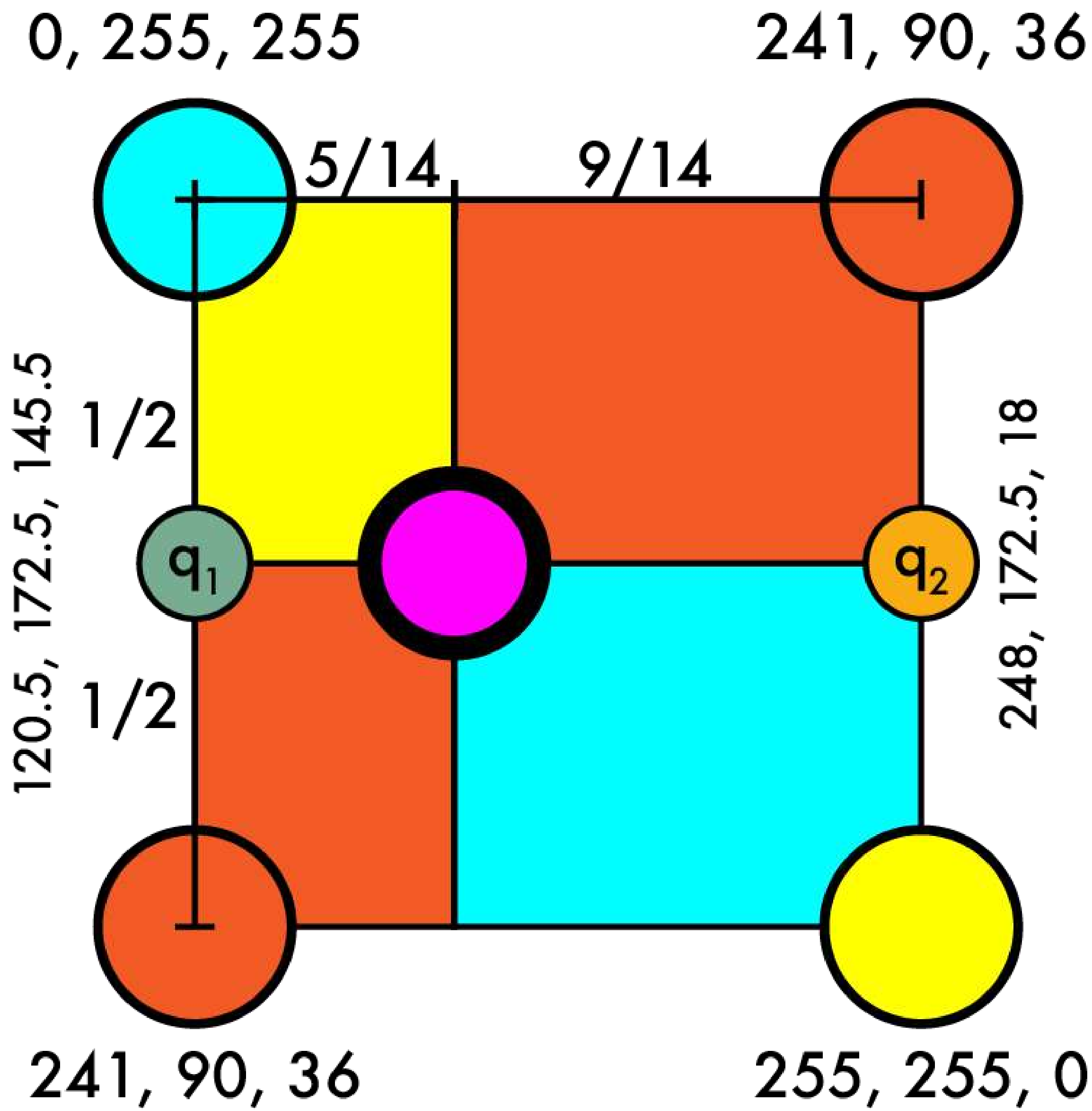
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q1 = (120.5, 172.5, 145.5)
    - q2 = (248, 172.5, 18)
    - q = r, g, b
    - q = 9/14*q1 + 5/14*q2



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

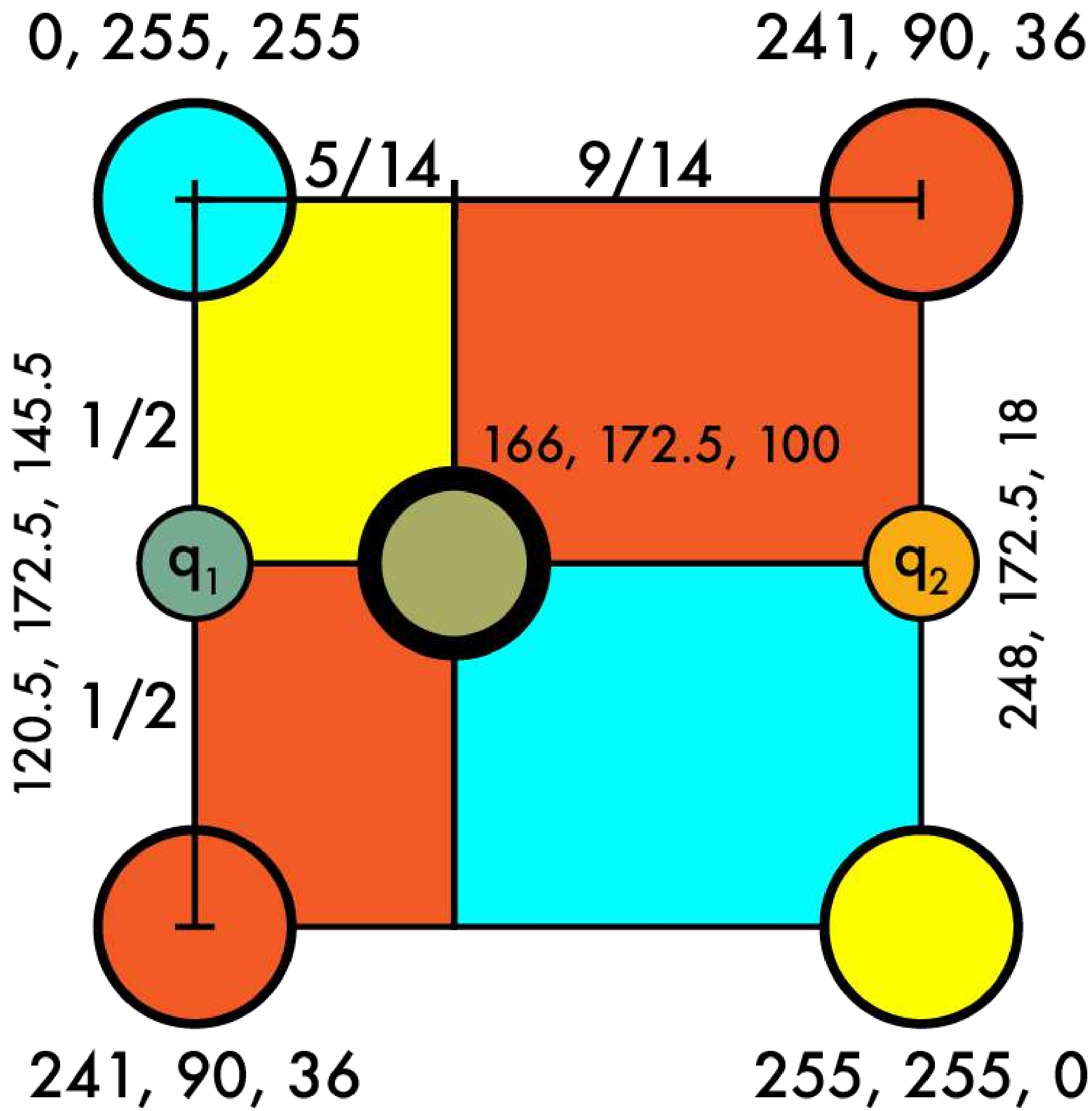
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q1 = (120.5, 172.5, 145.5)
    - q2 = (248, 172.5, 18)
    - q = r, g, b
    - q = 9/14*q1 + 5/14*q2
    - r = 9/14*120.5 + 5/14*248
    - g = 9/14*172.5 + 5/14*172.5
    - b = 9/14*145.5 + 5/14*18



Resize 4x4 -> 7x7

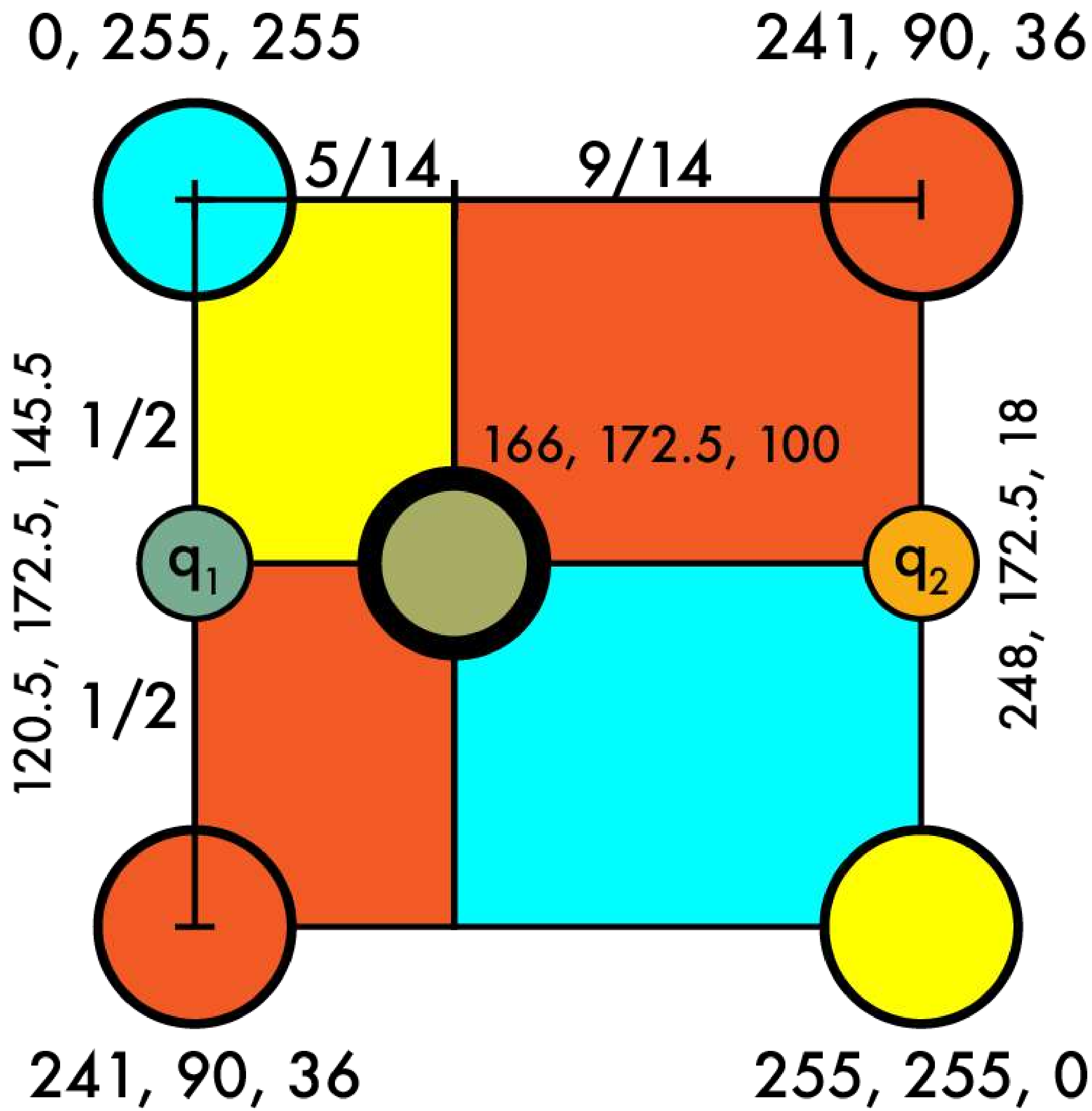
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q1 = (120.5, 172.5, 145.5)
    - q2 = (248, 172.5, 18)
    - q = (166, 172.5, 100)



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

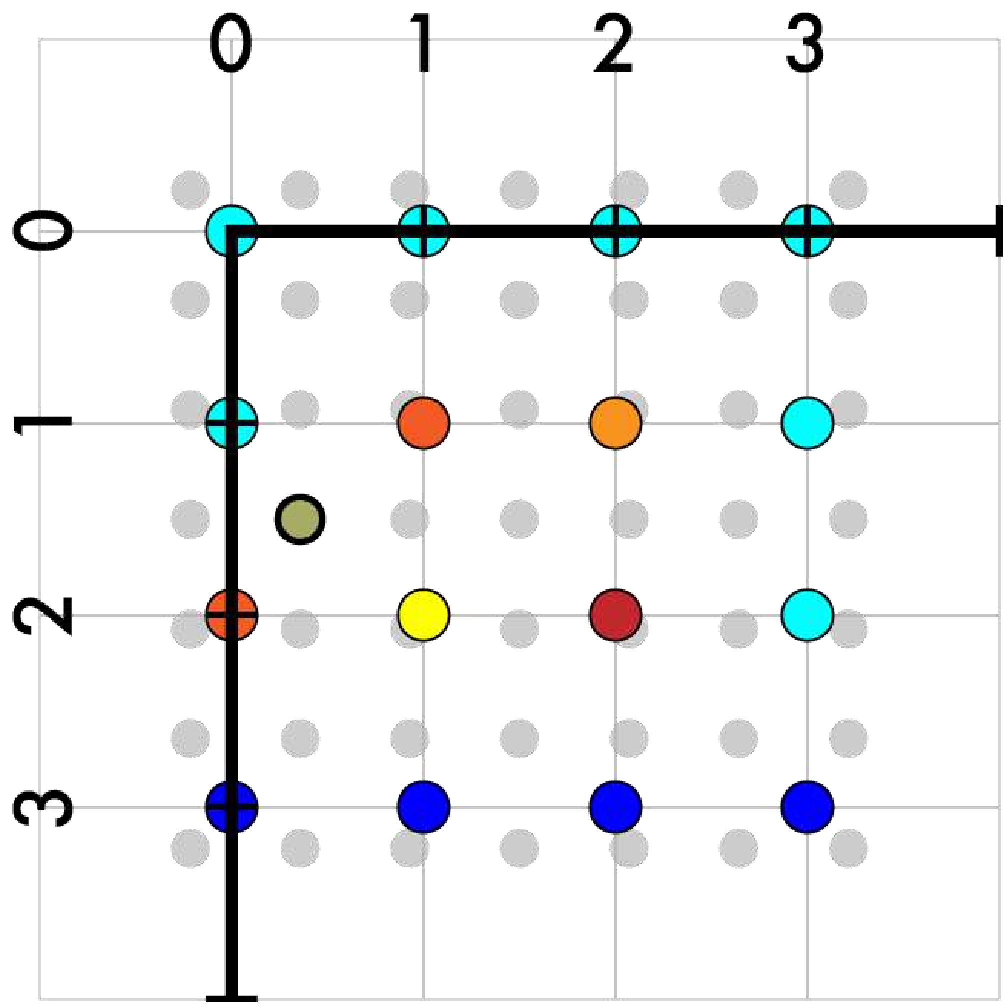
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q = (166, 172.5, 100)



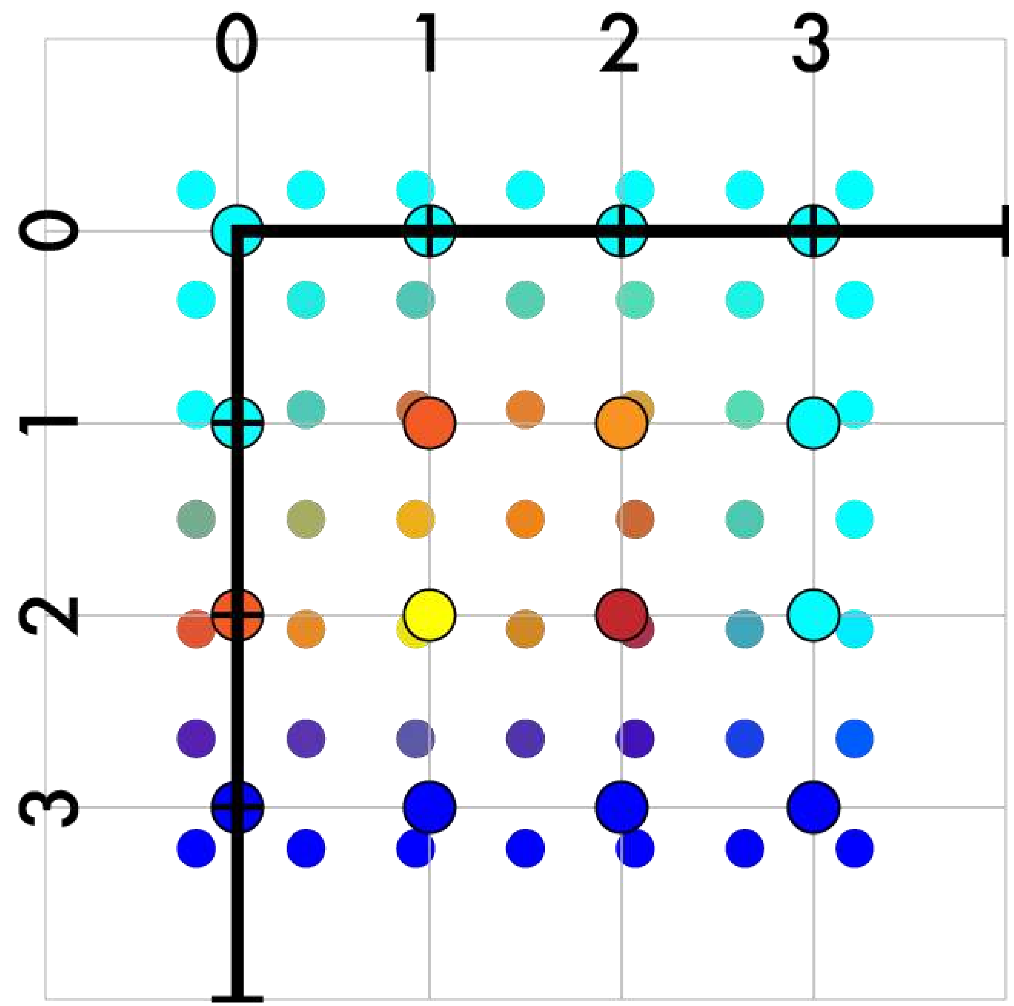
Resize 4x4 -> 7x7

- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q = (166, 172.5, 100)



Resize 4x4 -> 7x7

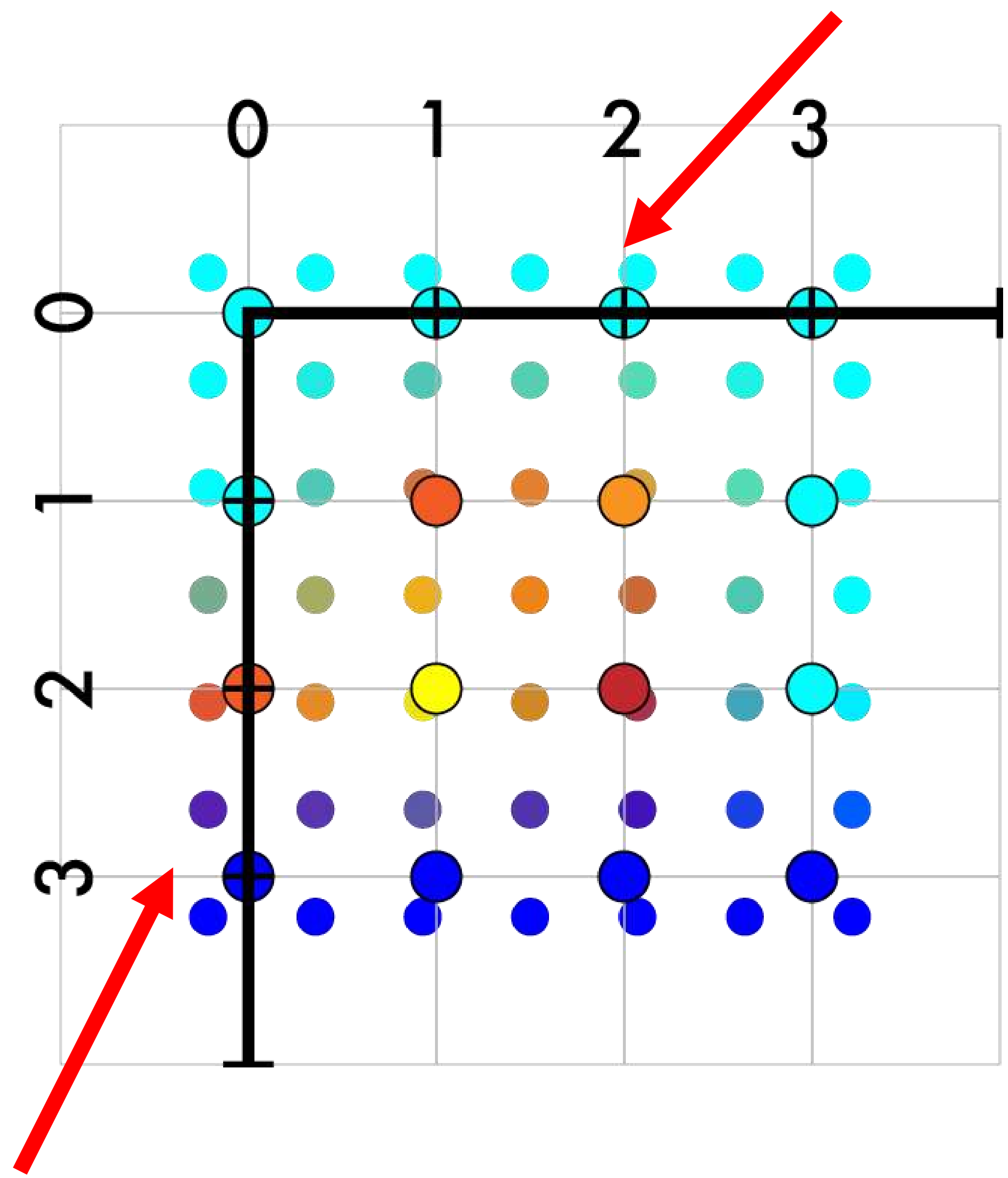
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q = (166, 172.5, 100)
- Fill in the rest



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Resize 4x4 -> 7x7

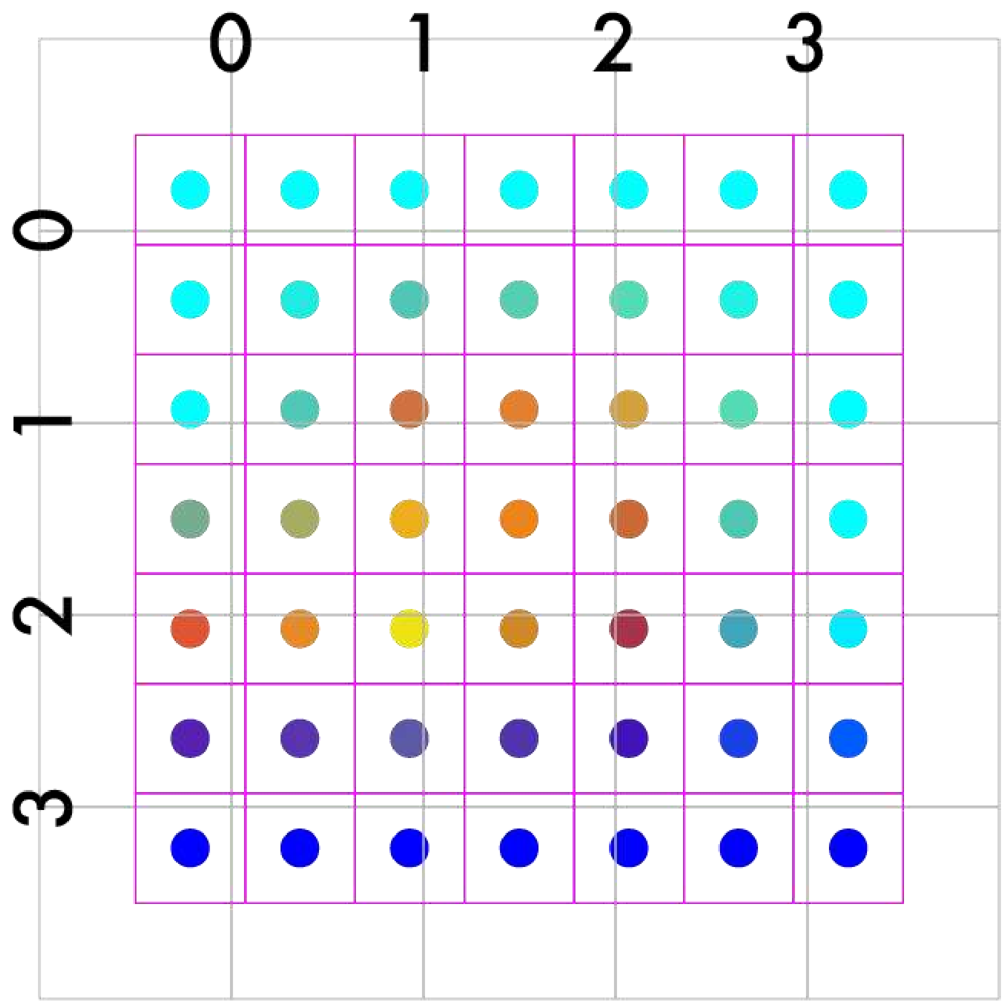
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q = (166, 172.5, 100)
- Fill in the rest
  - On outer edges use padding!



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Resize 4x4 -> 7x7

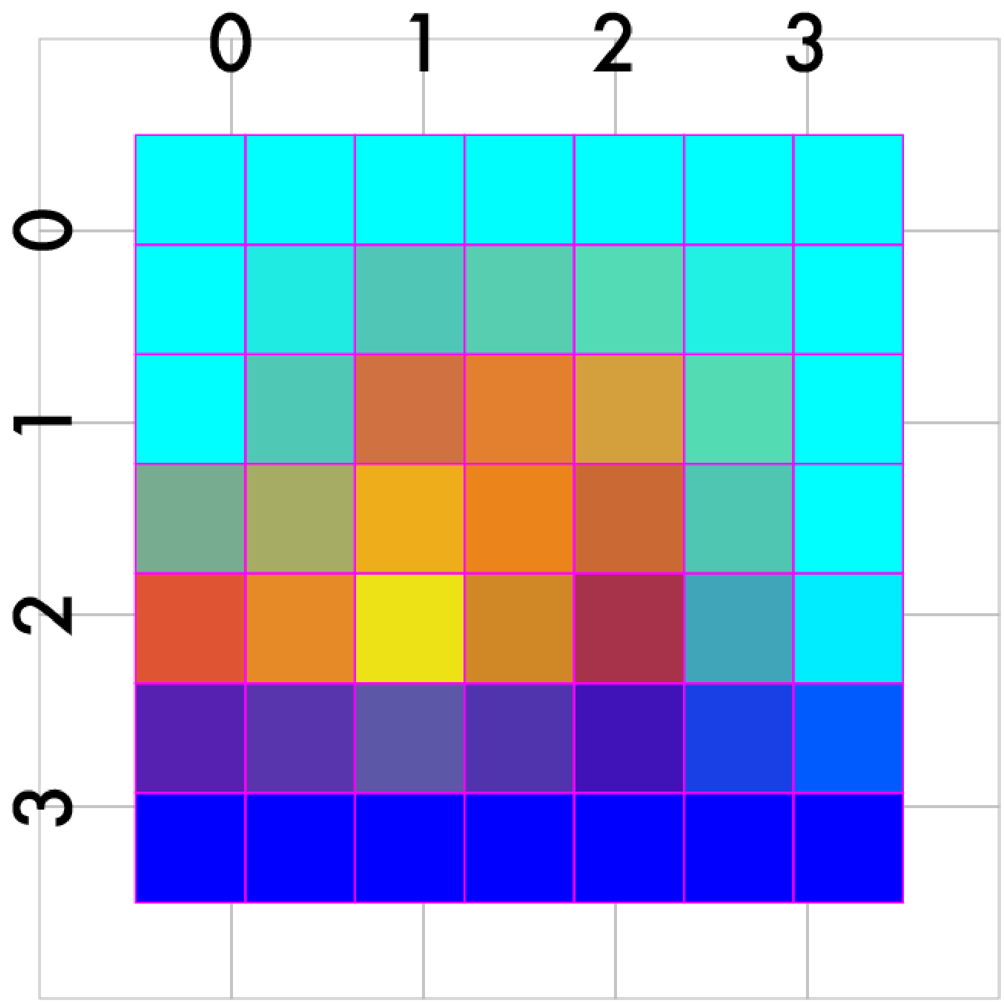
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q = (166, 172.5, 100)
- Fill in the rest



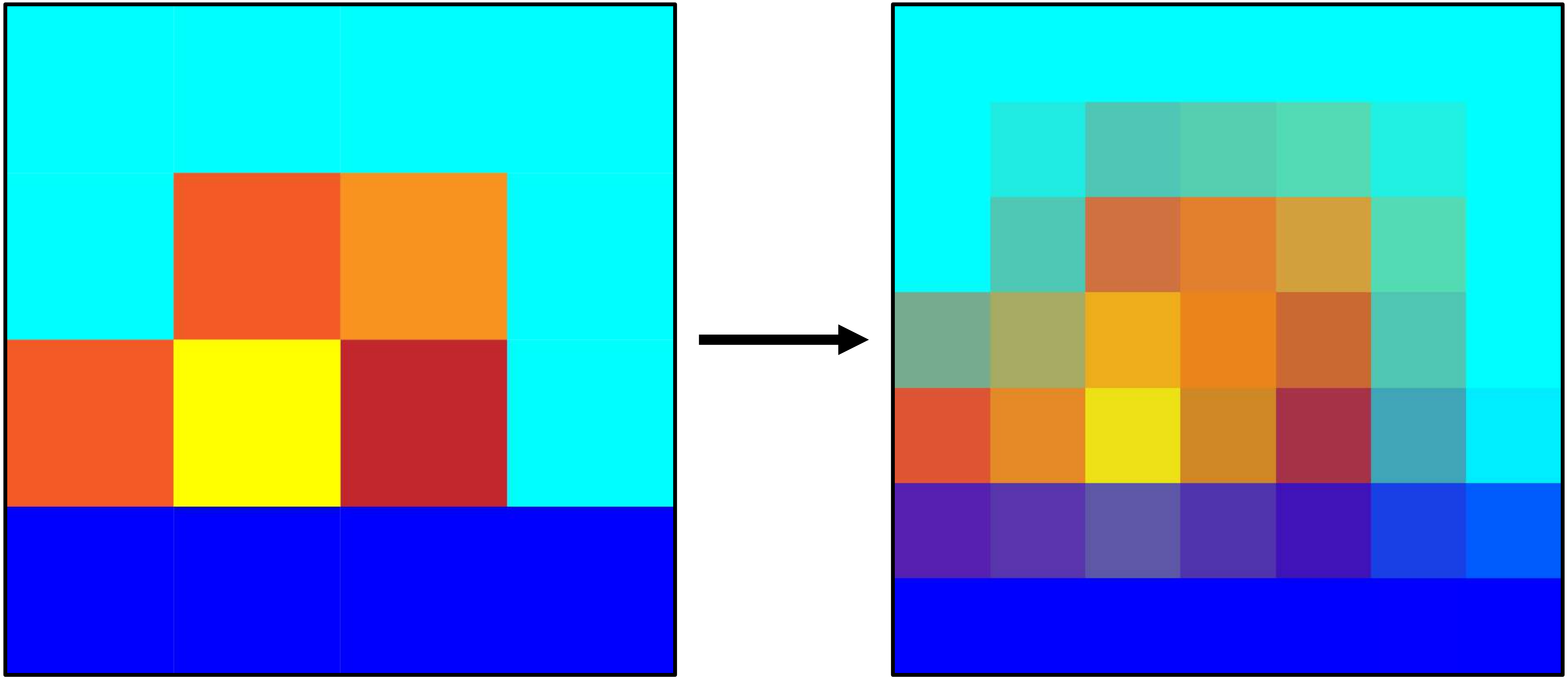
Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Resize 4x4 -> 7x7

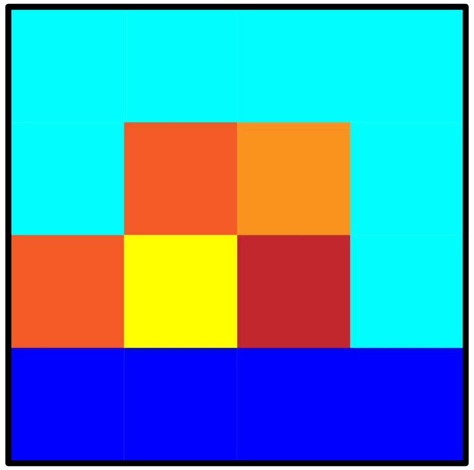
- Create our new image
- Match up coordinates
  - 4/7 X - 3/14 = Y
- Iterate over new pts
  - Map to old coords
  - (1, 3) -> (5/14, 21/14)
  - Interpolate old values
    - q = (166, 172.5, 100)
- Fill in the rest

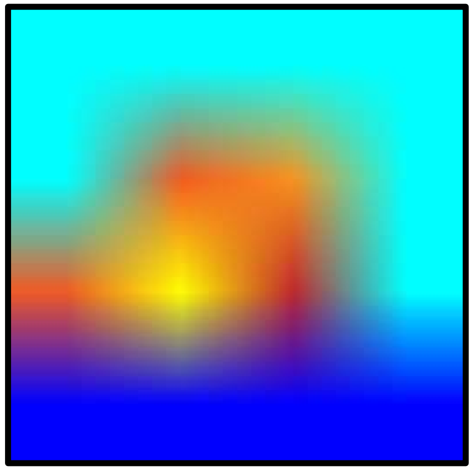


We did it!

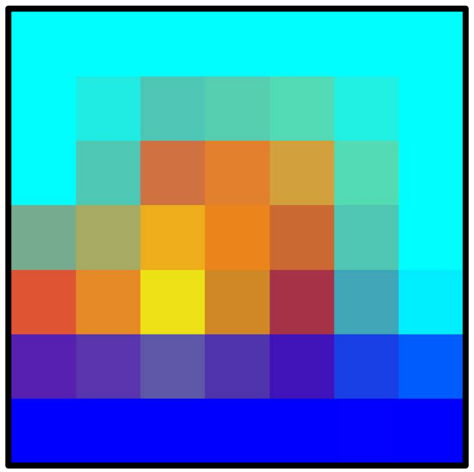


Different scales

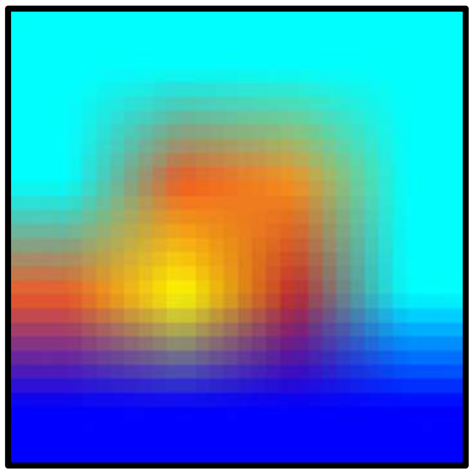




256x256



7x7

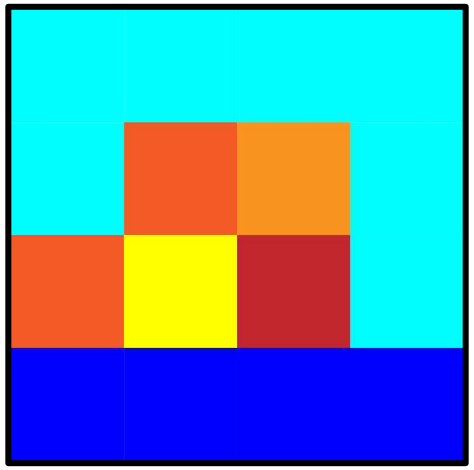


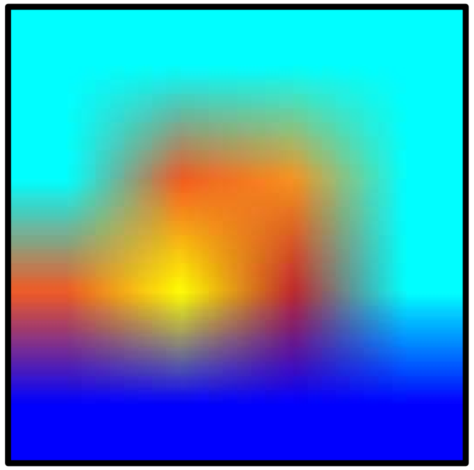
32x32

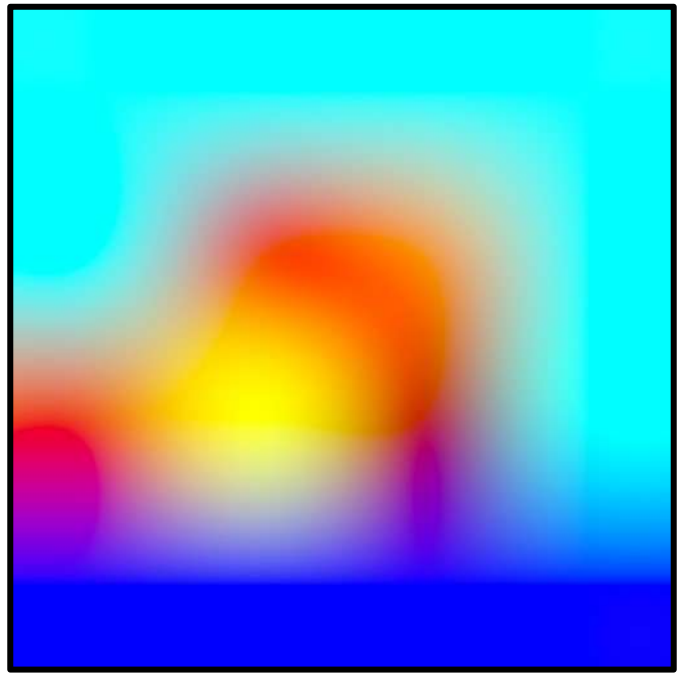
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

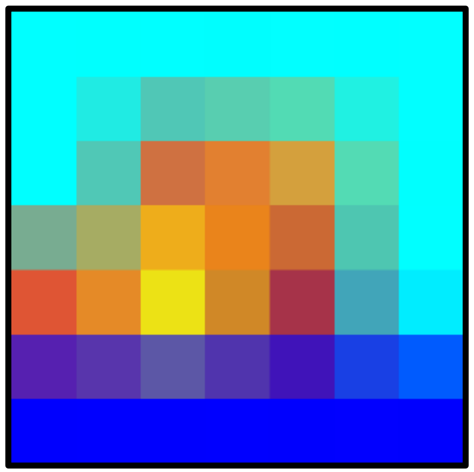
Different methods

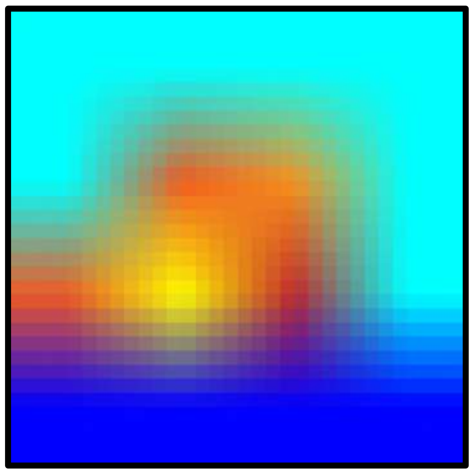


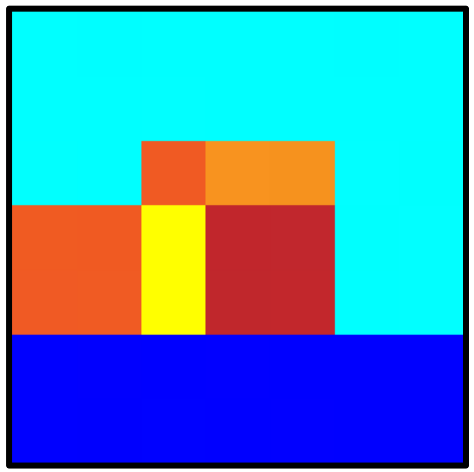




Bicubic







7x7
NN

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Want to make image smaller

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

448x448 -> 64x64

448x448 -> 64x64

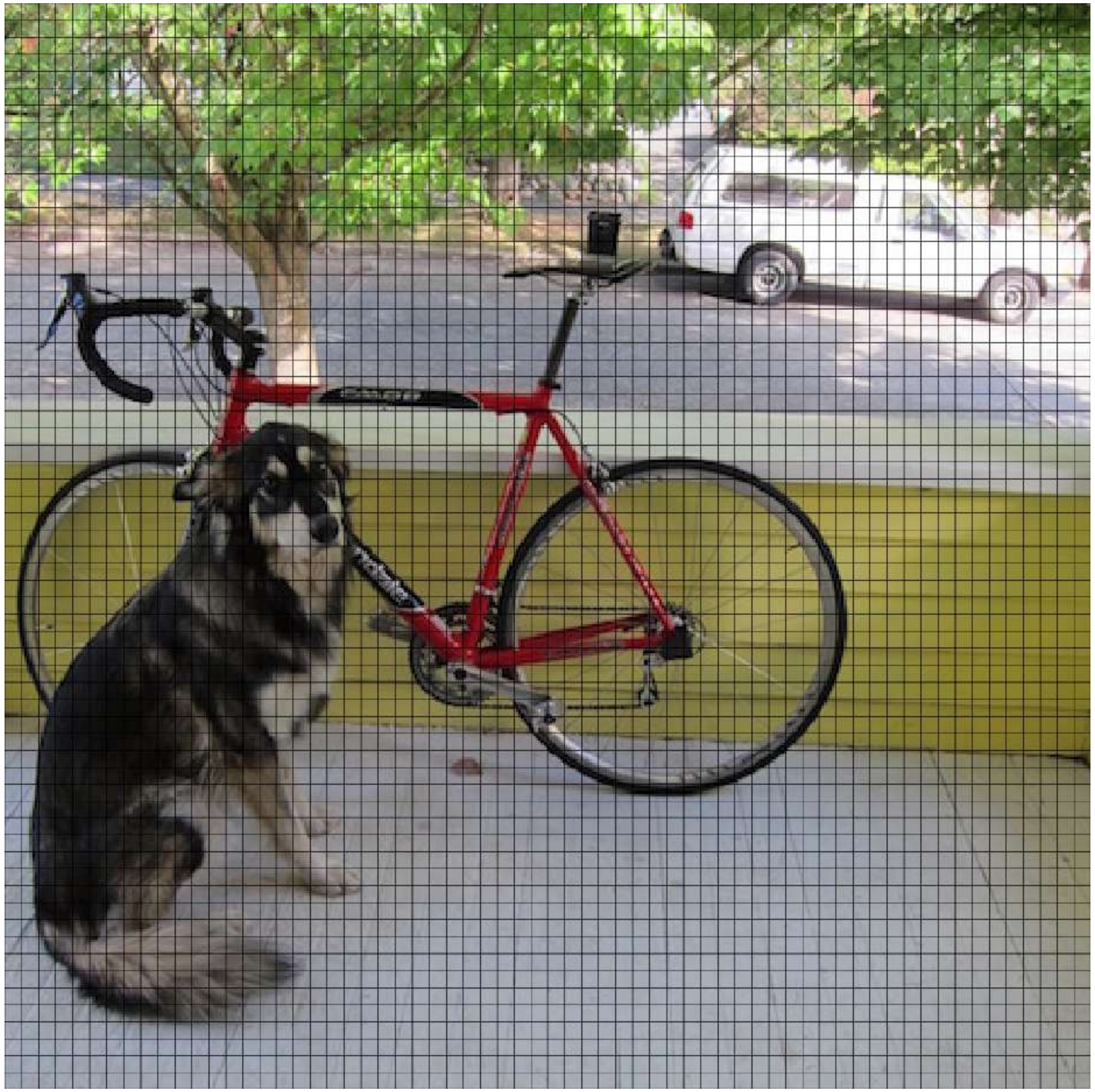
NN

Bilinear

448x448 -> 64x64

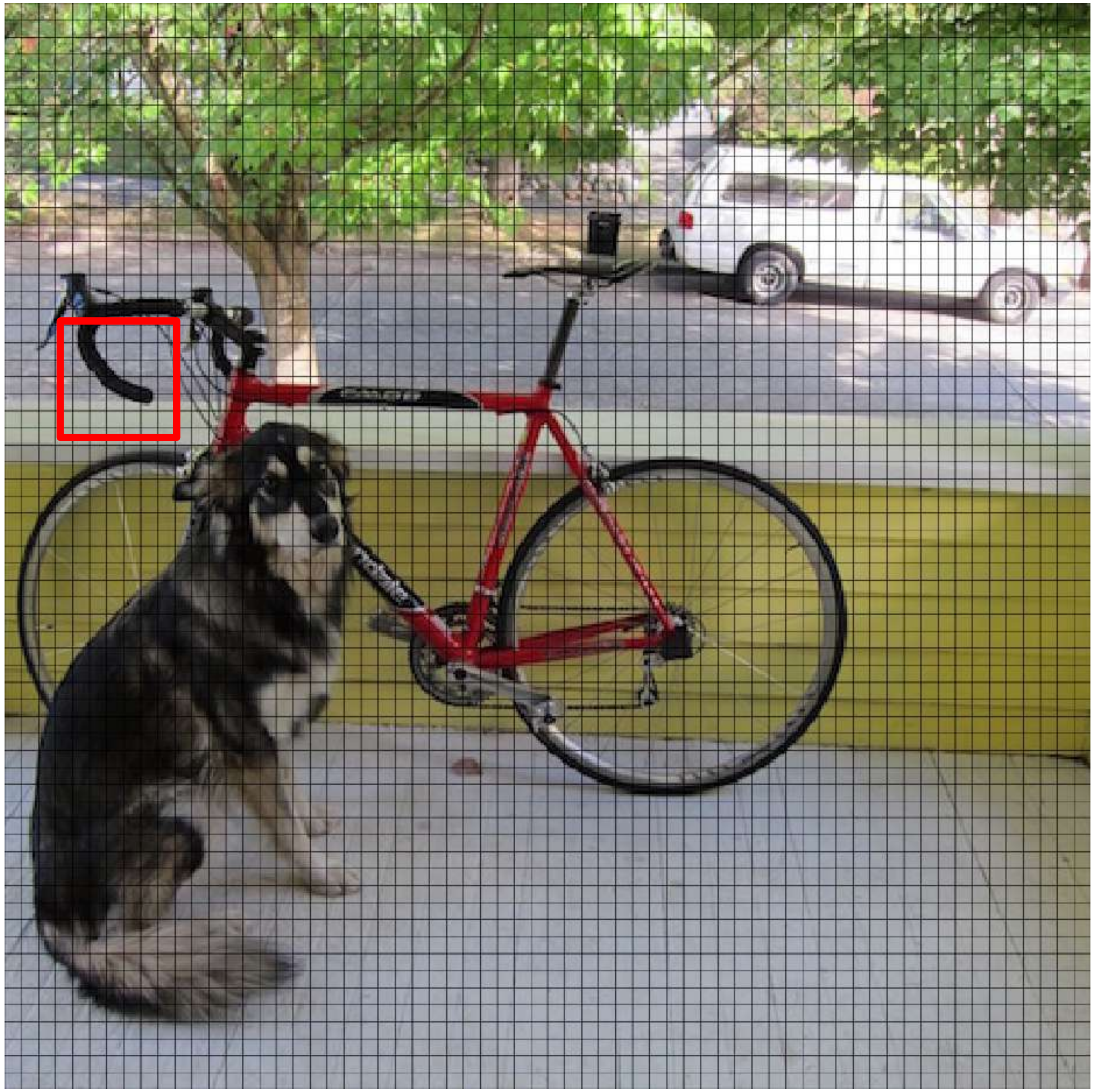
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

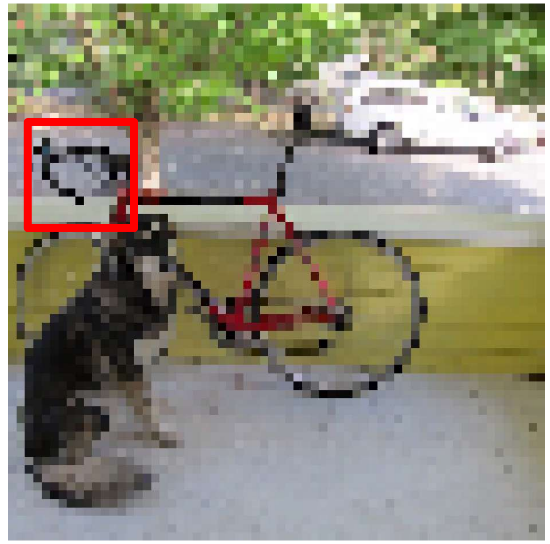
448x448 -> 64x64



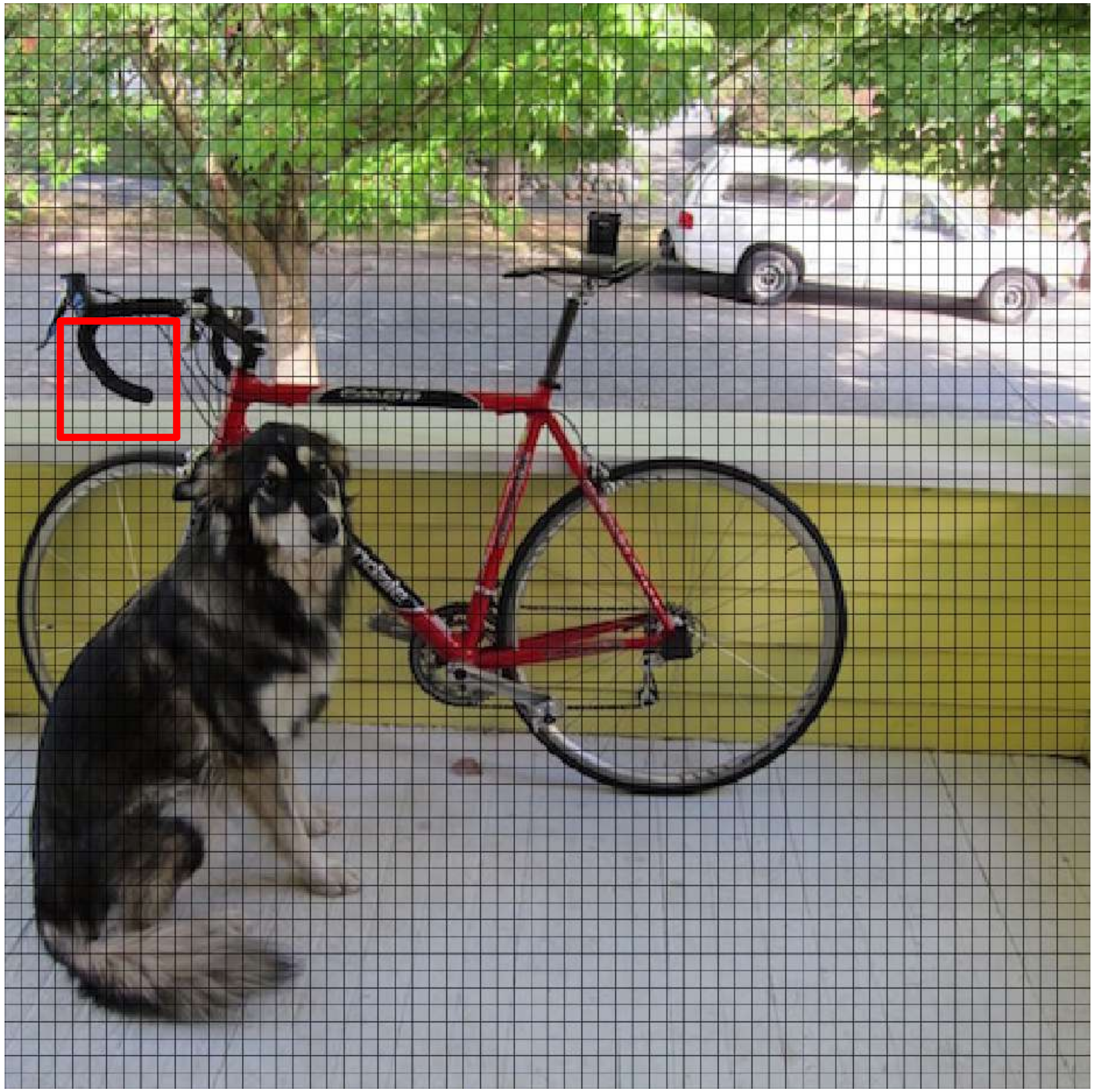
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

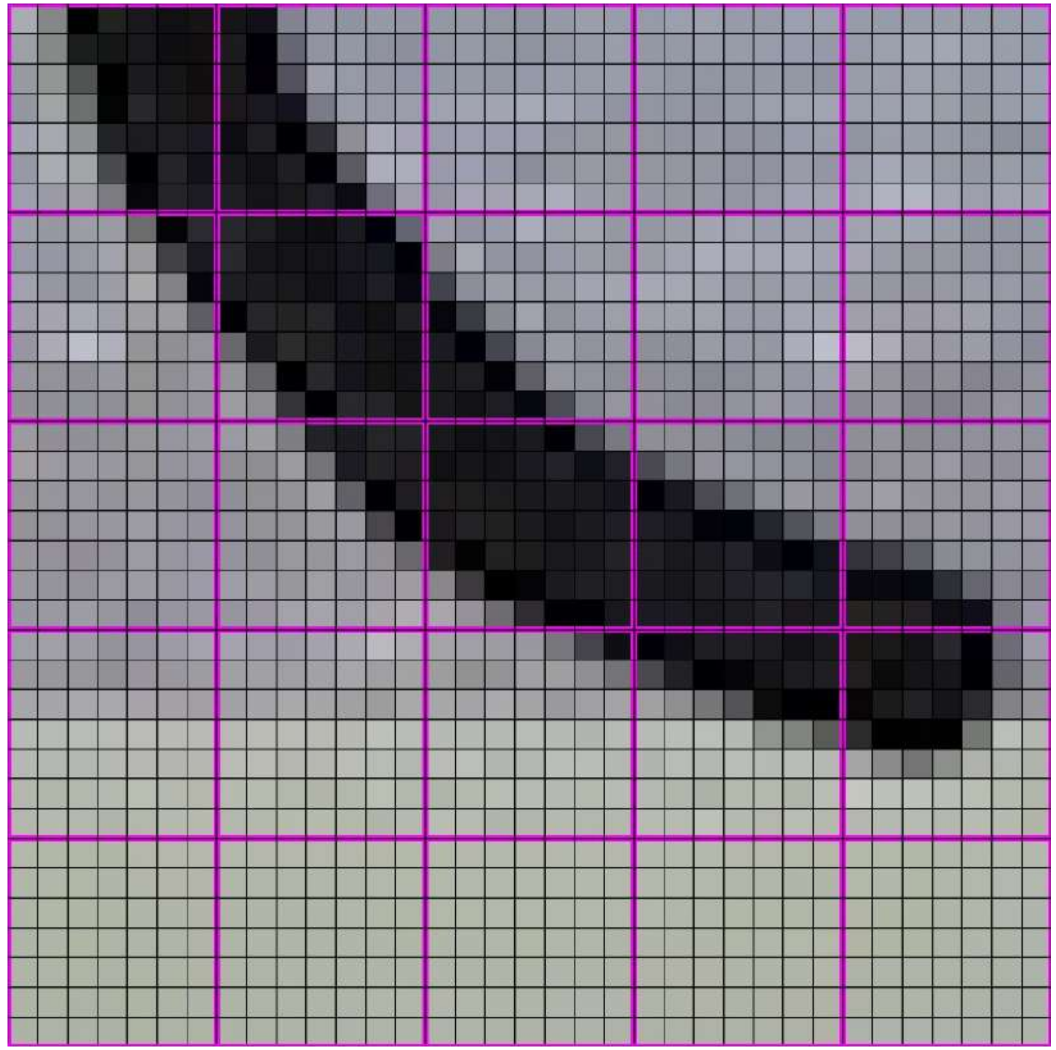
448x448 -> 64x64





448x448 -> 64x64

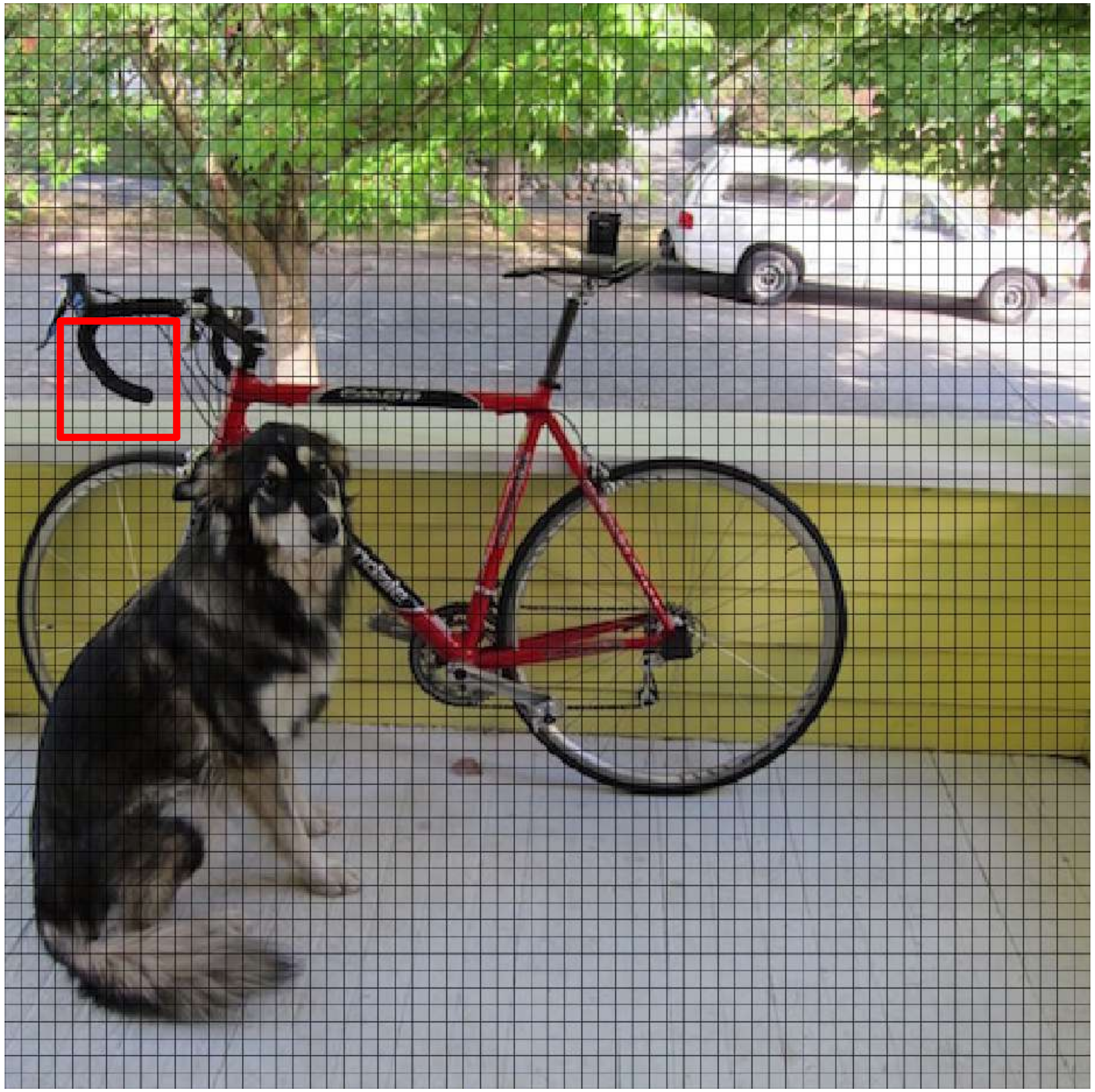


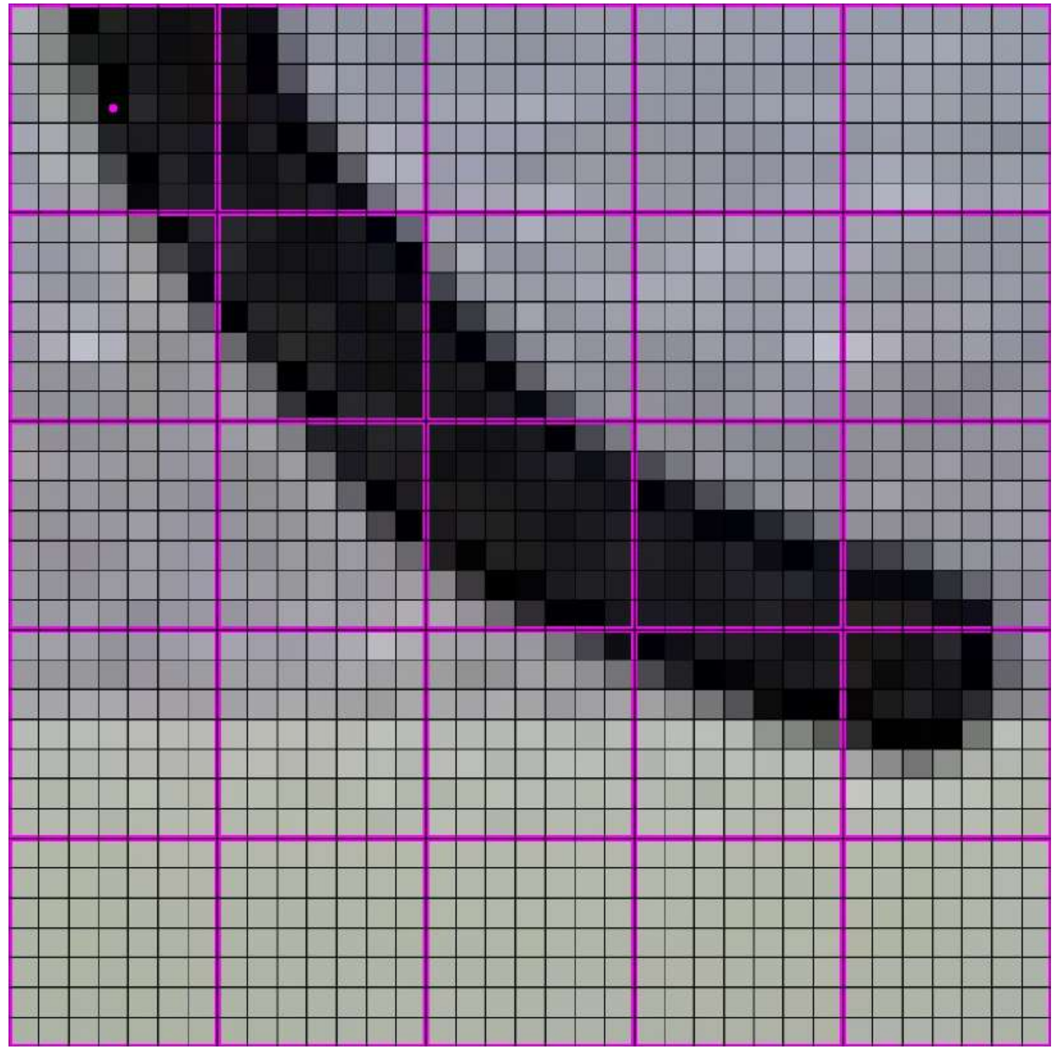


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

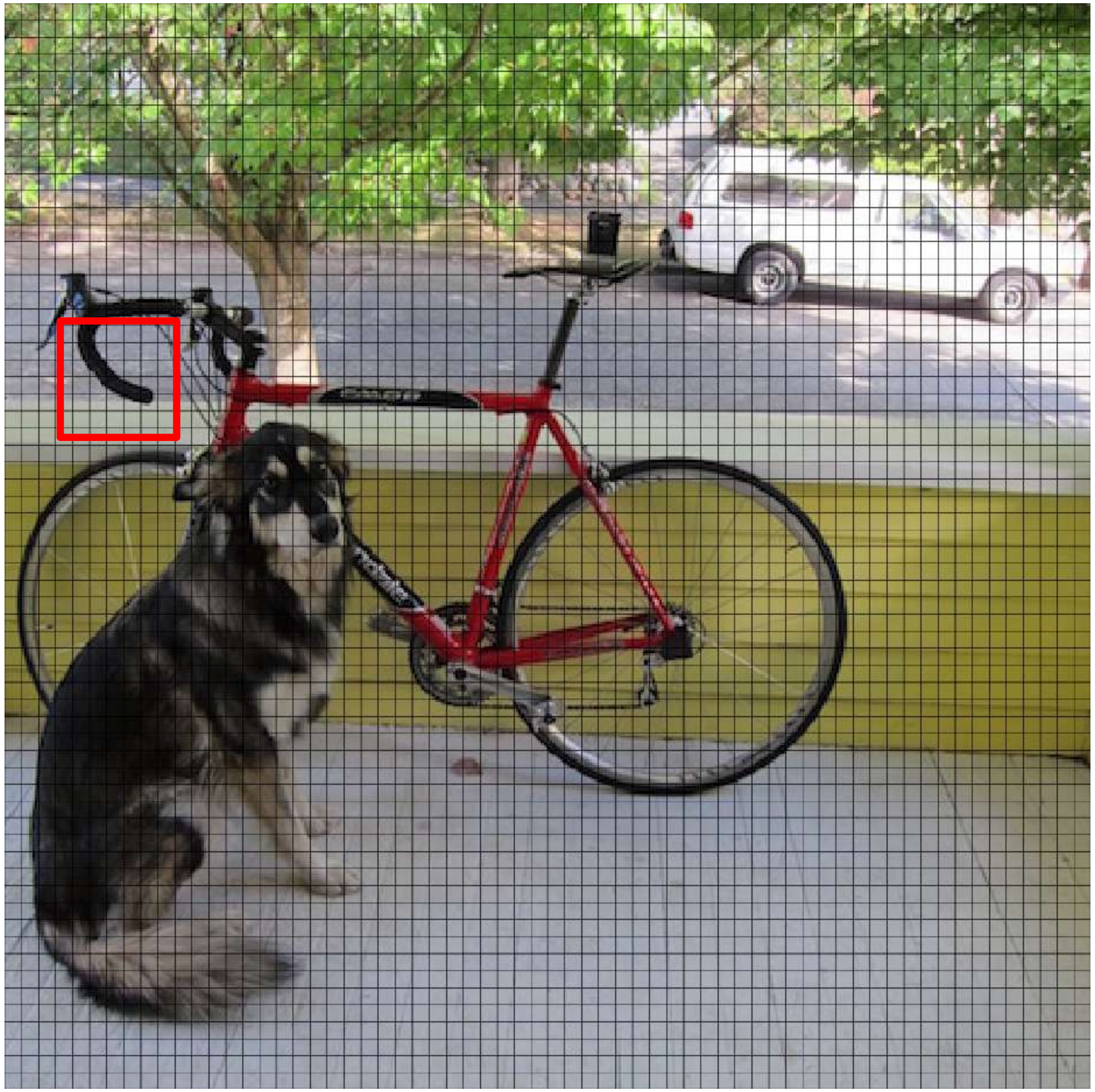
448x448 -> 64x64

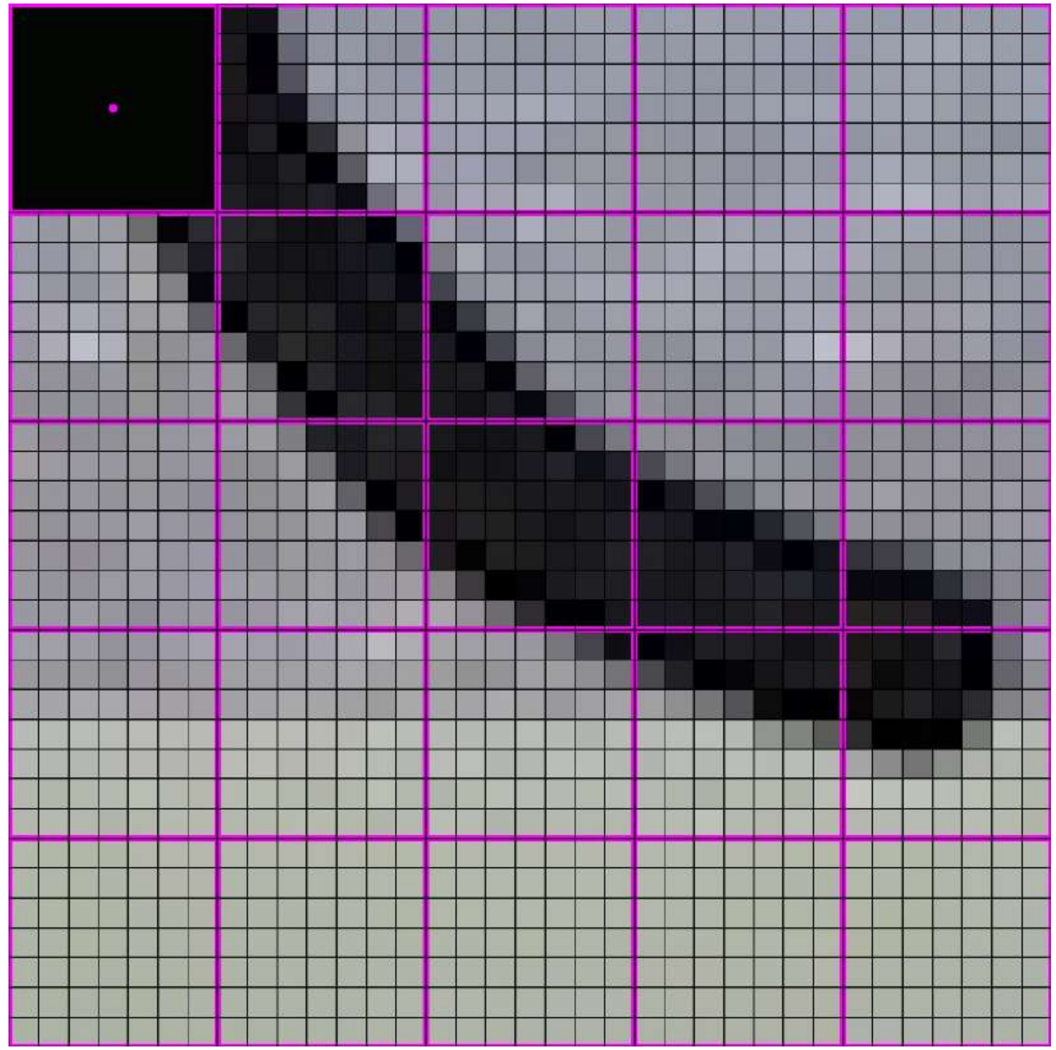




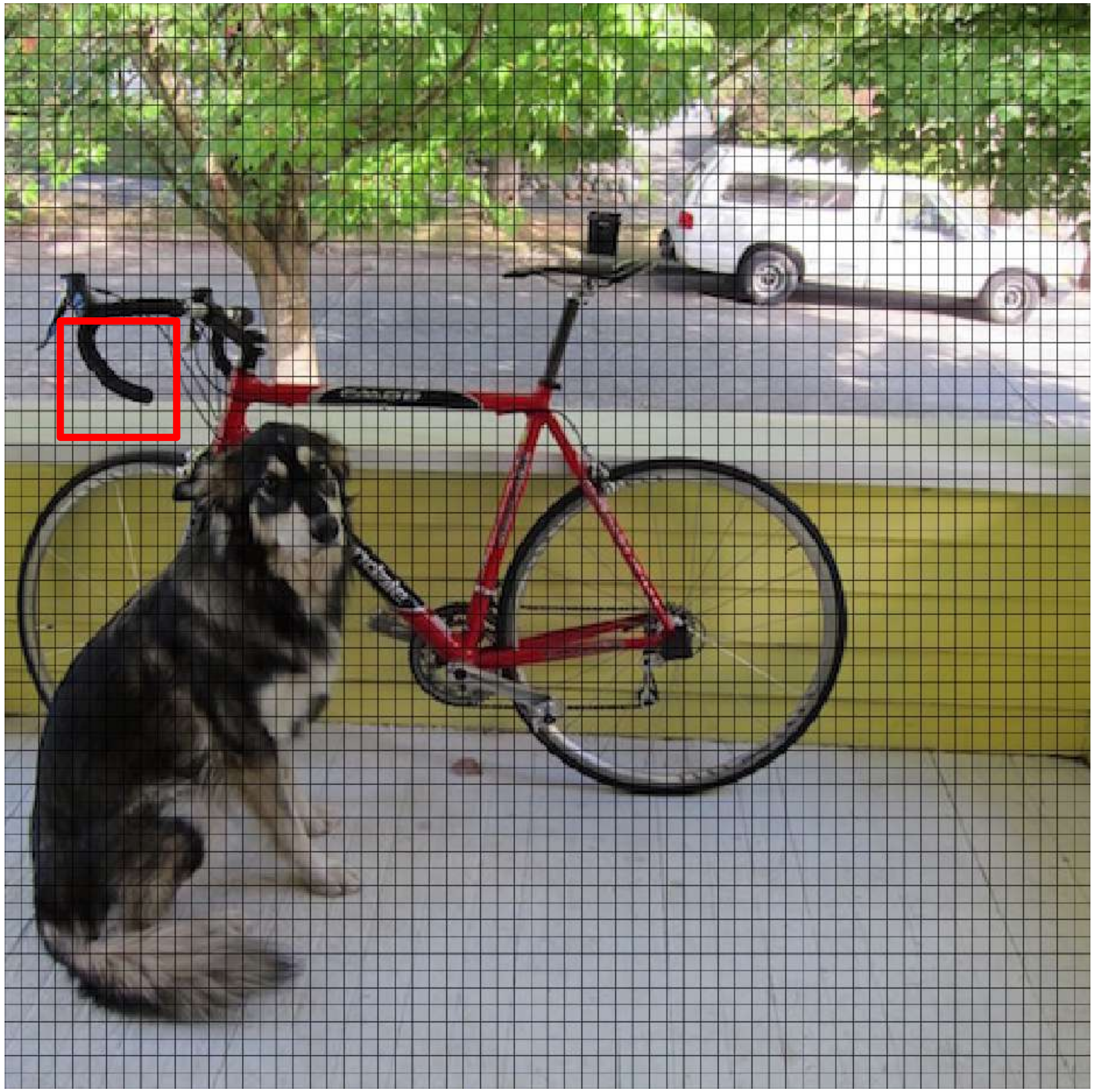
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

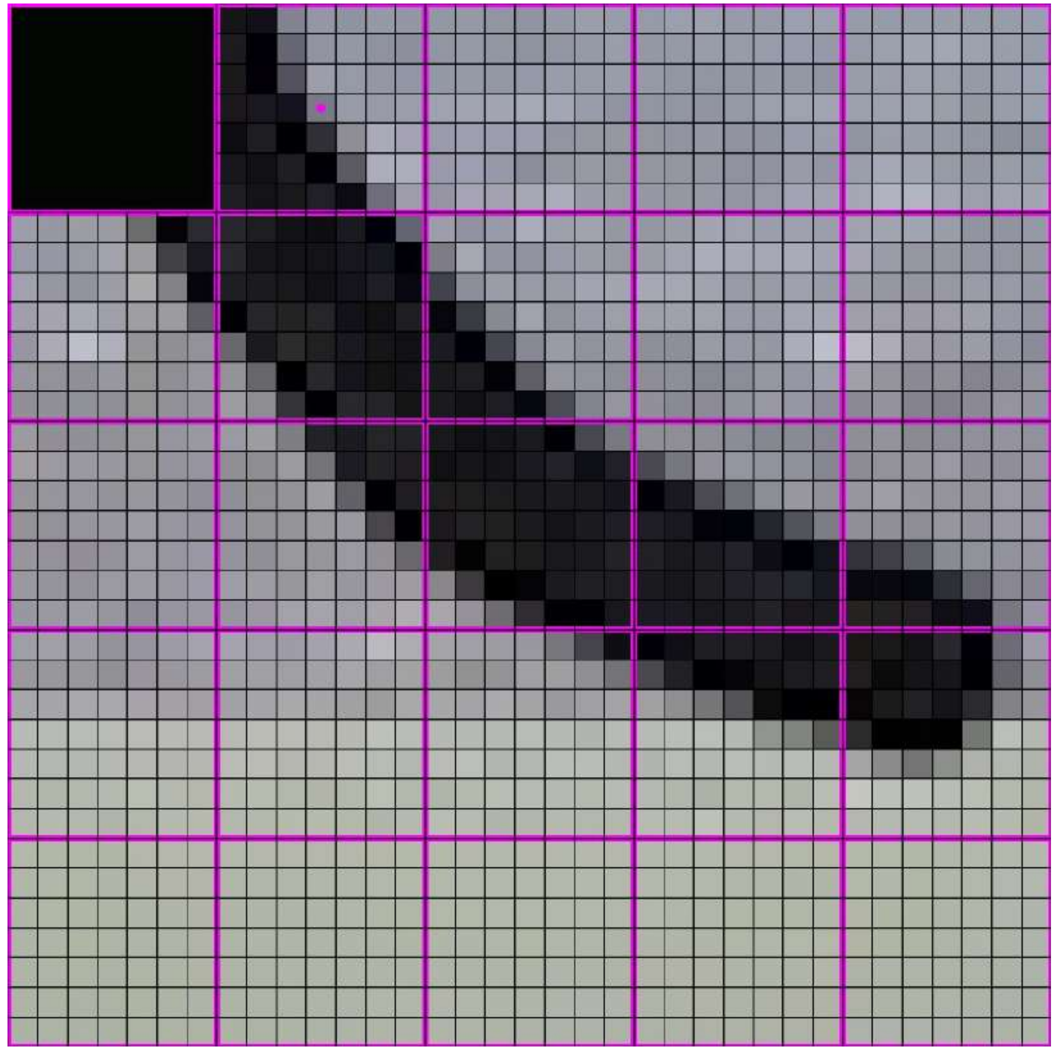
448x448 -> 64x64



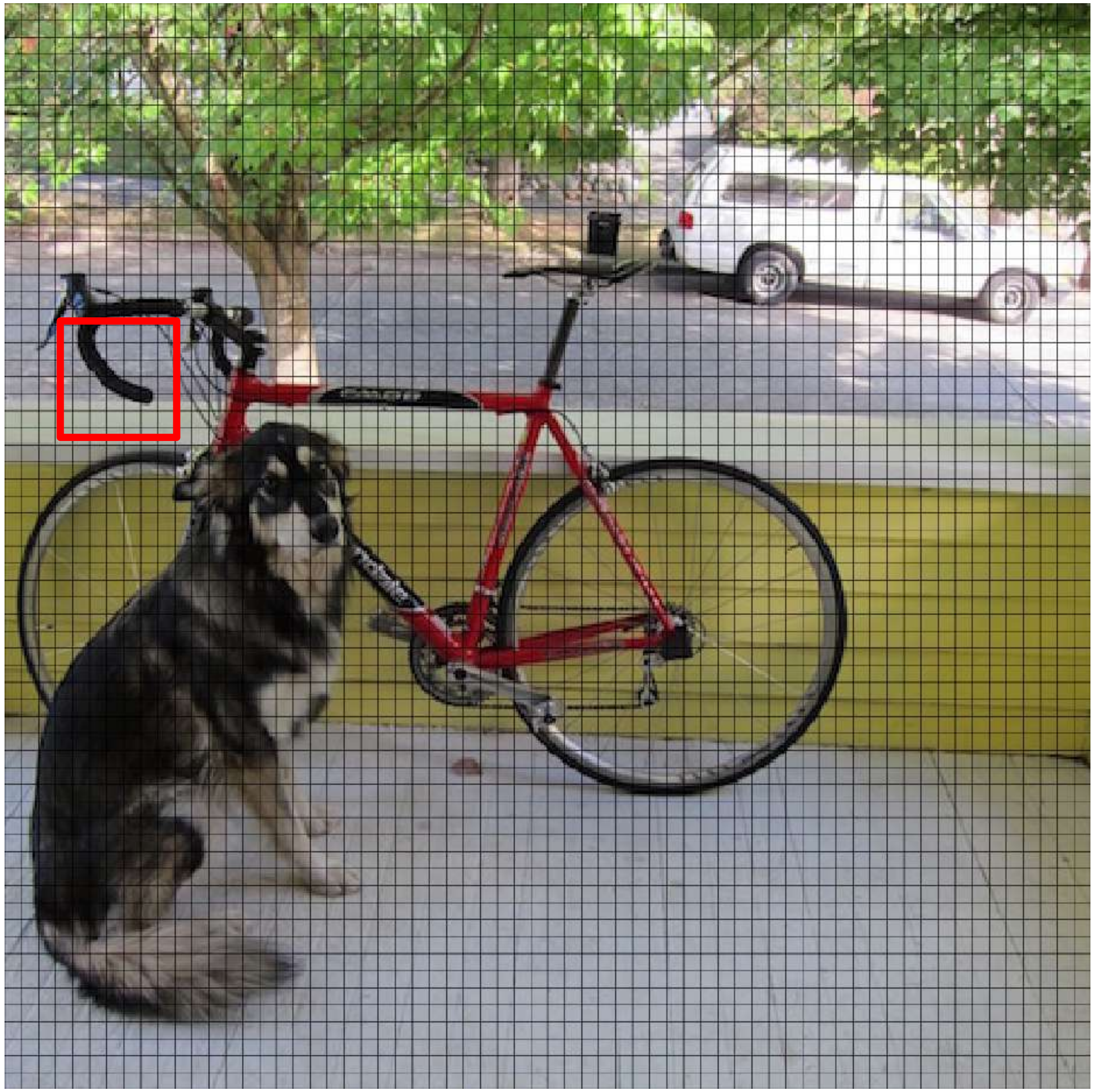


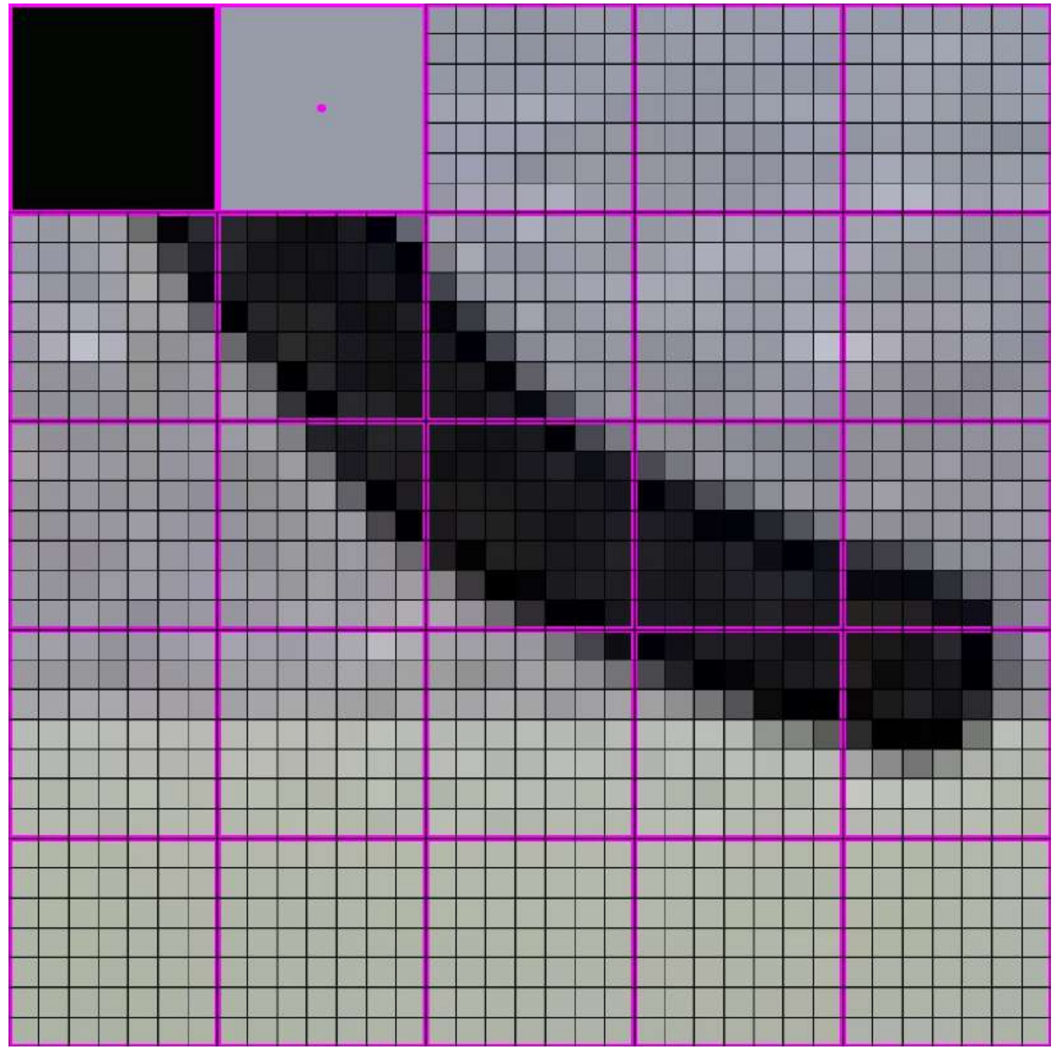
448x448 -> 64x64



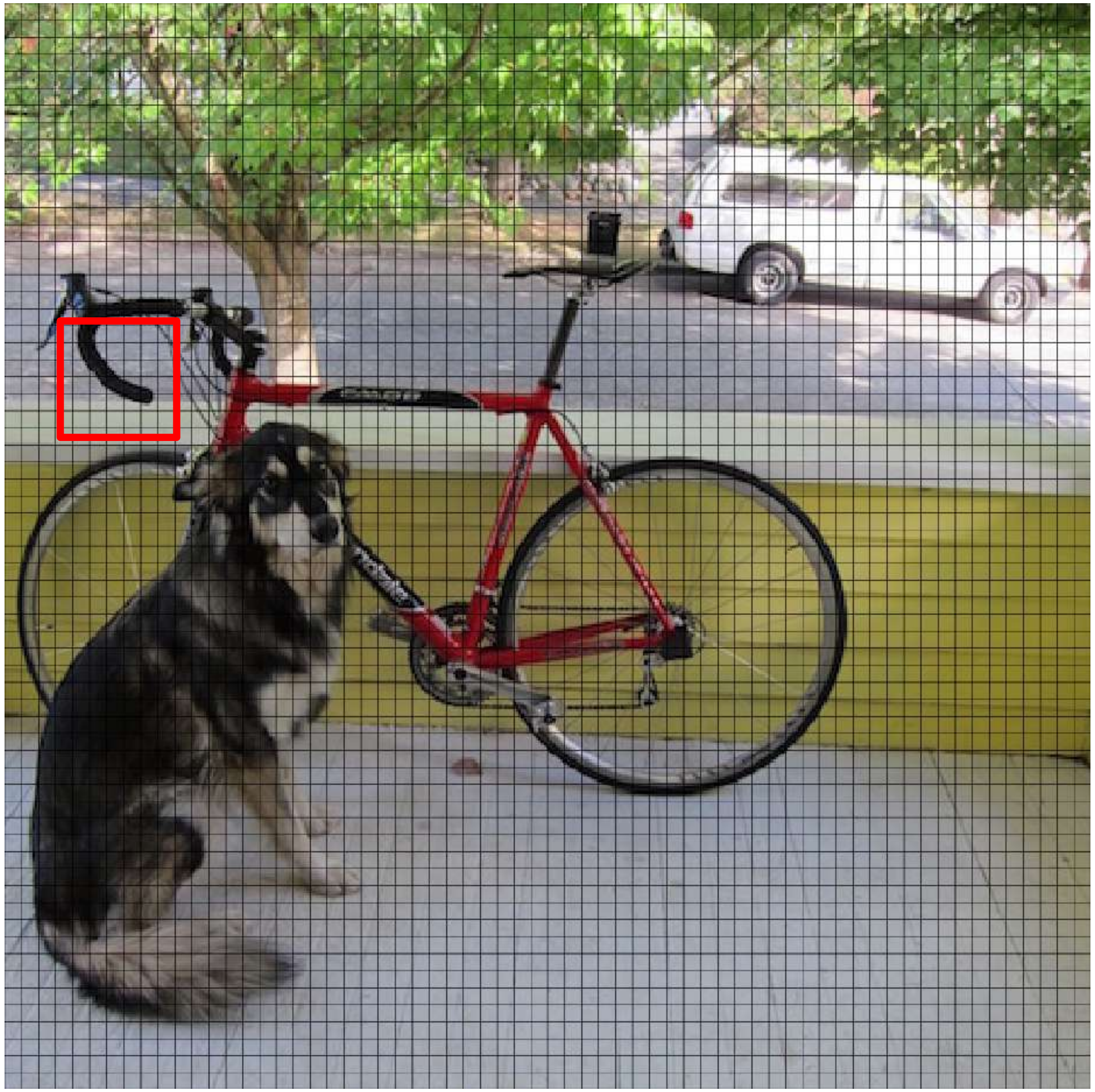


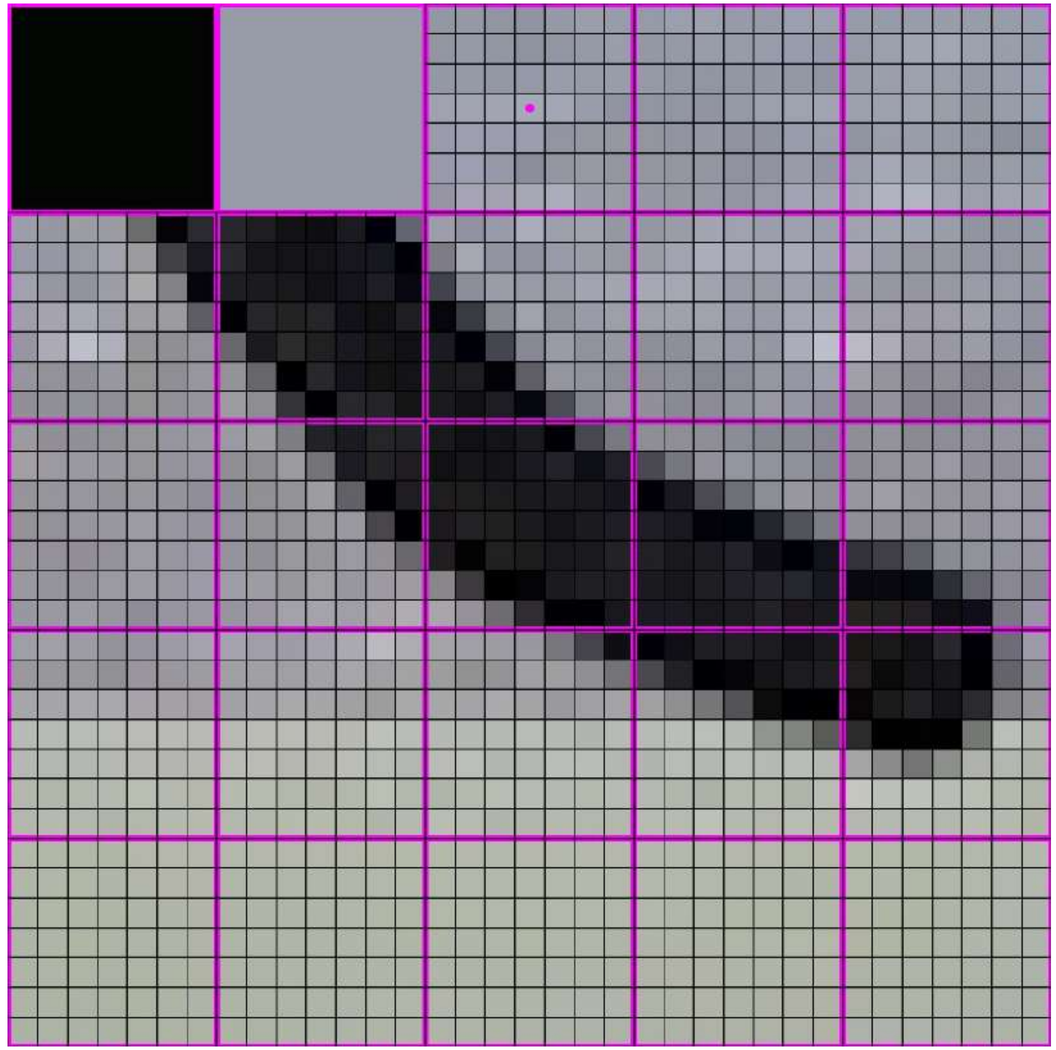
448x448 -> 64x64





448x448 -> 64x64

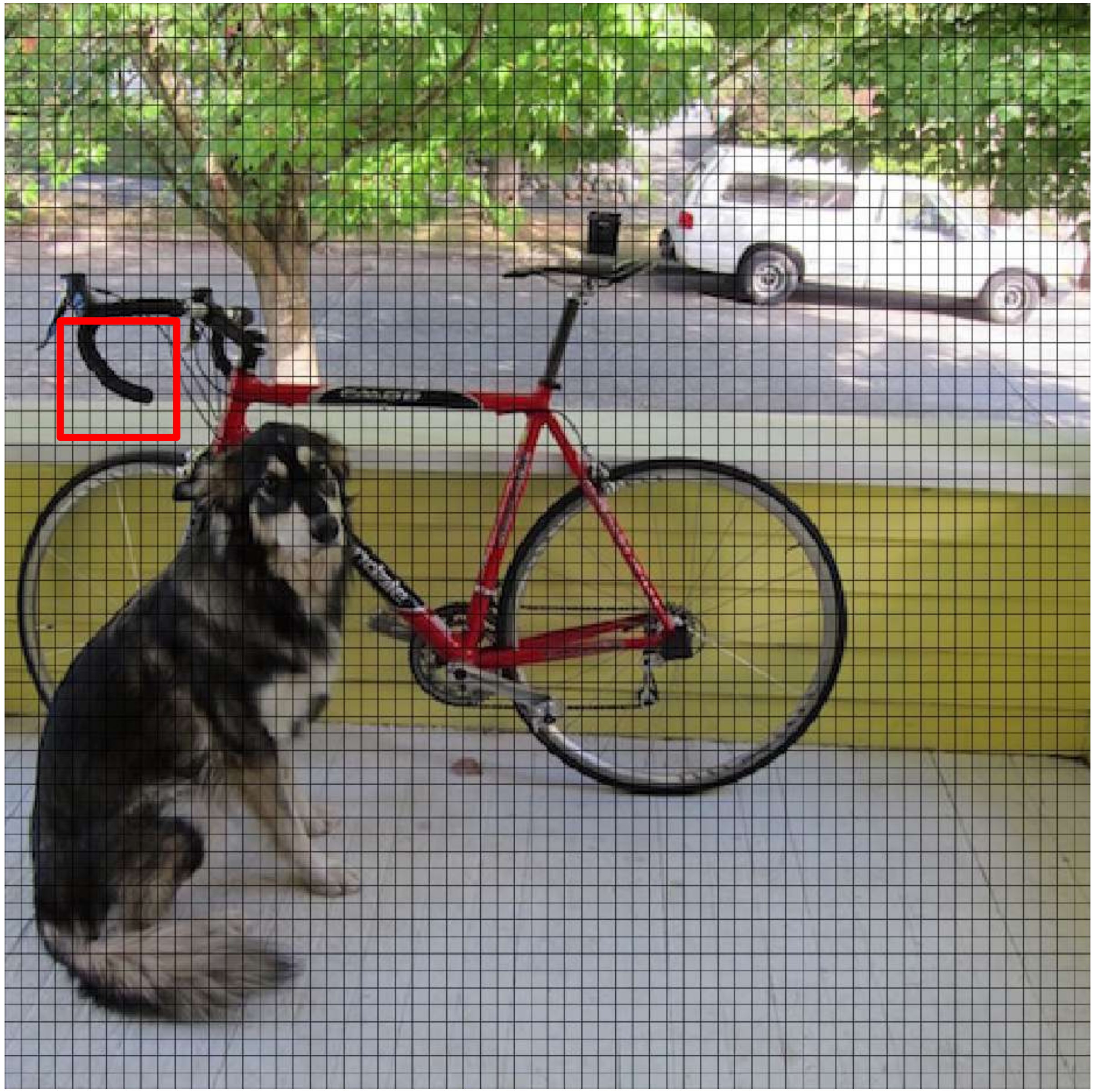


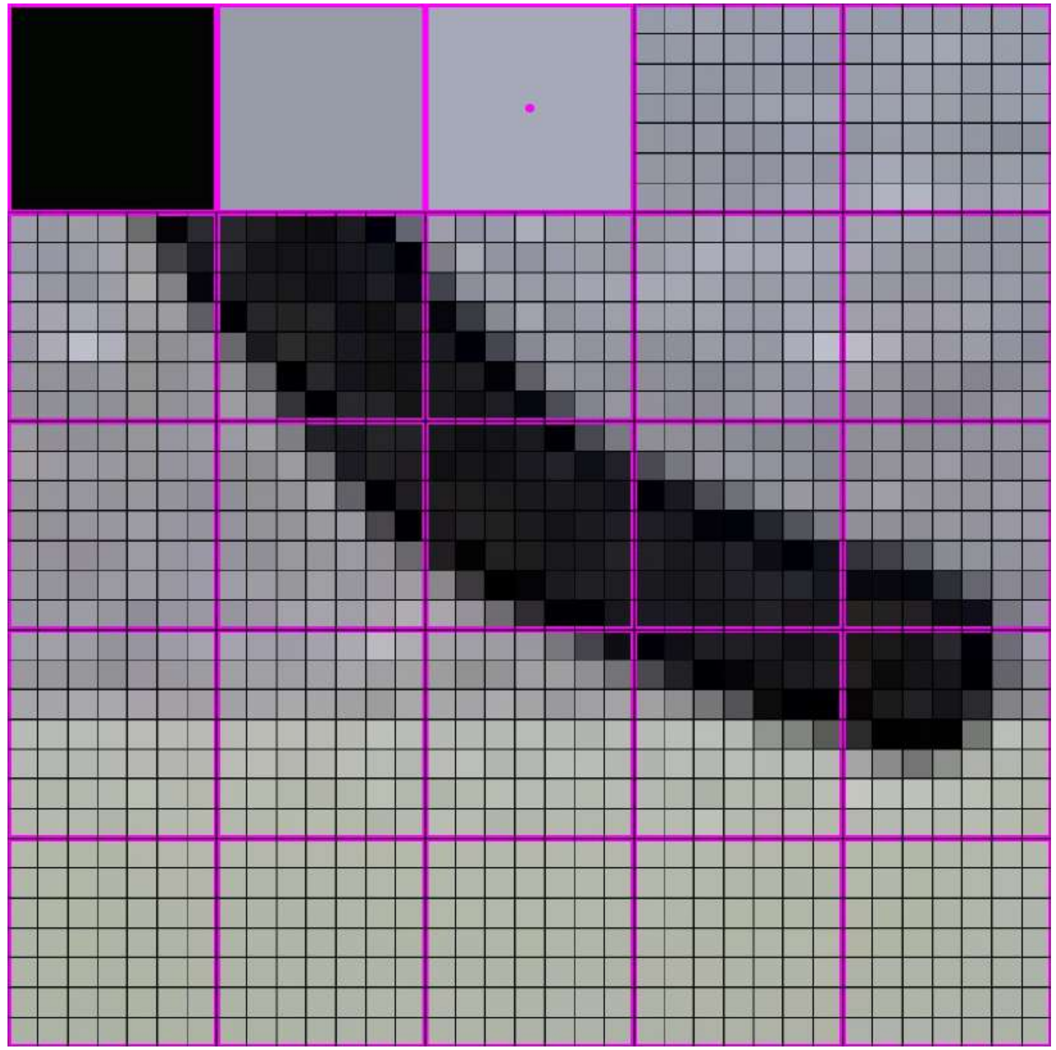


Co-financed by the European Union
Connecting Europe Facility

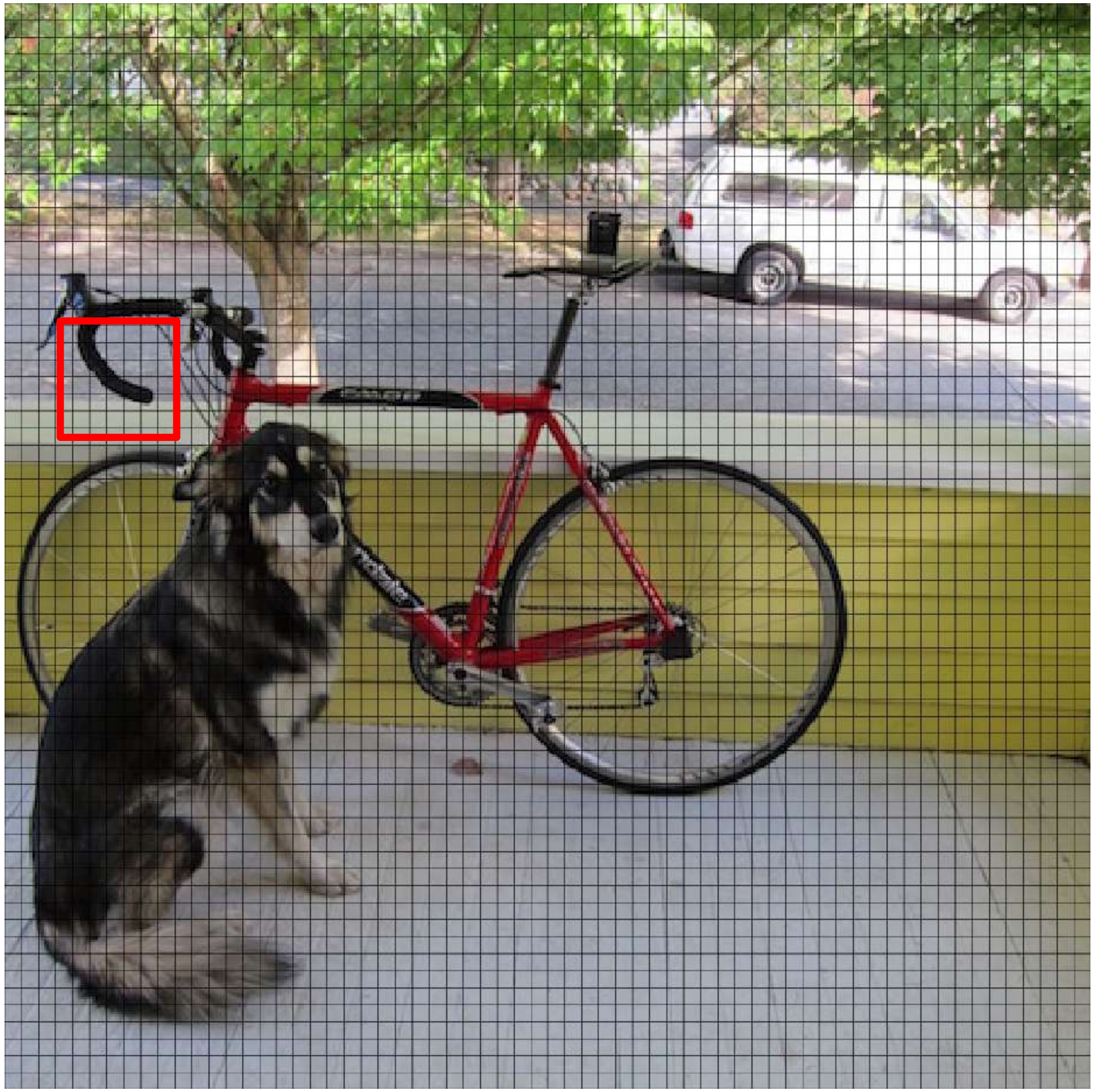
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

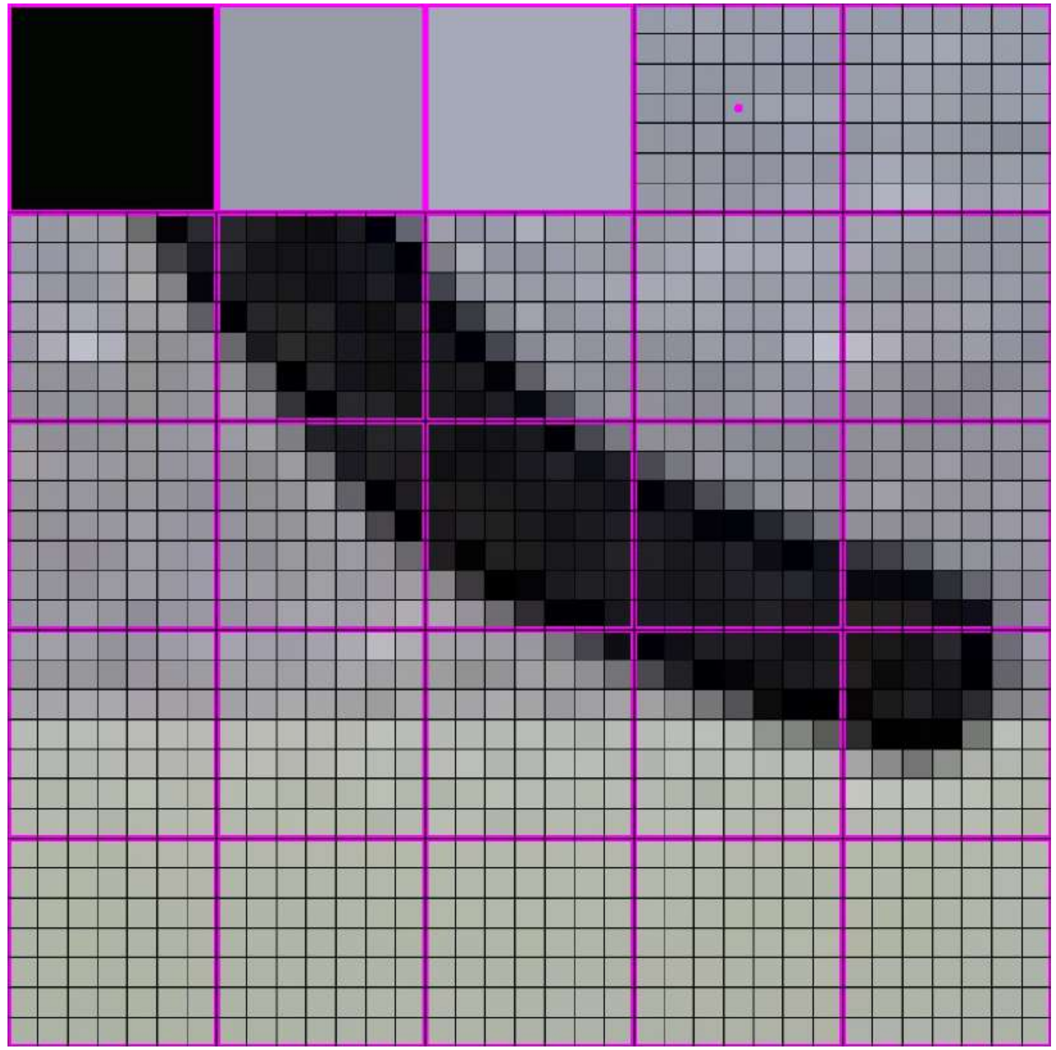
448x448 -> 64x64



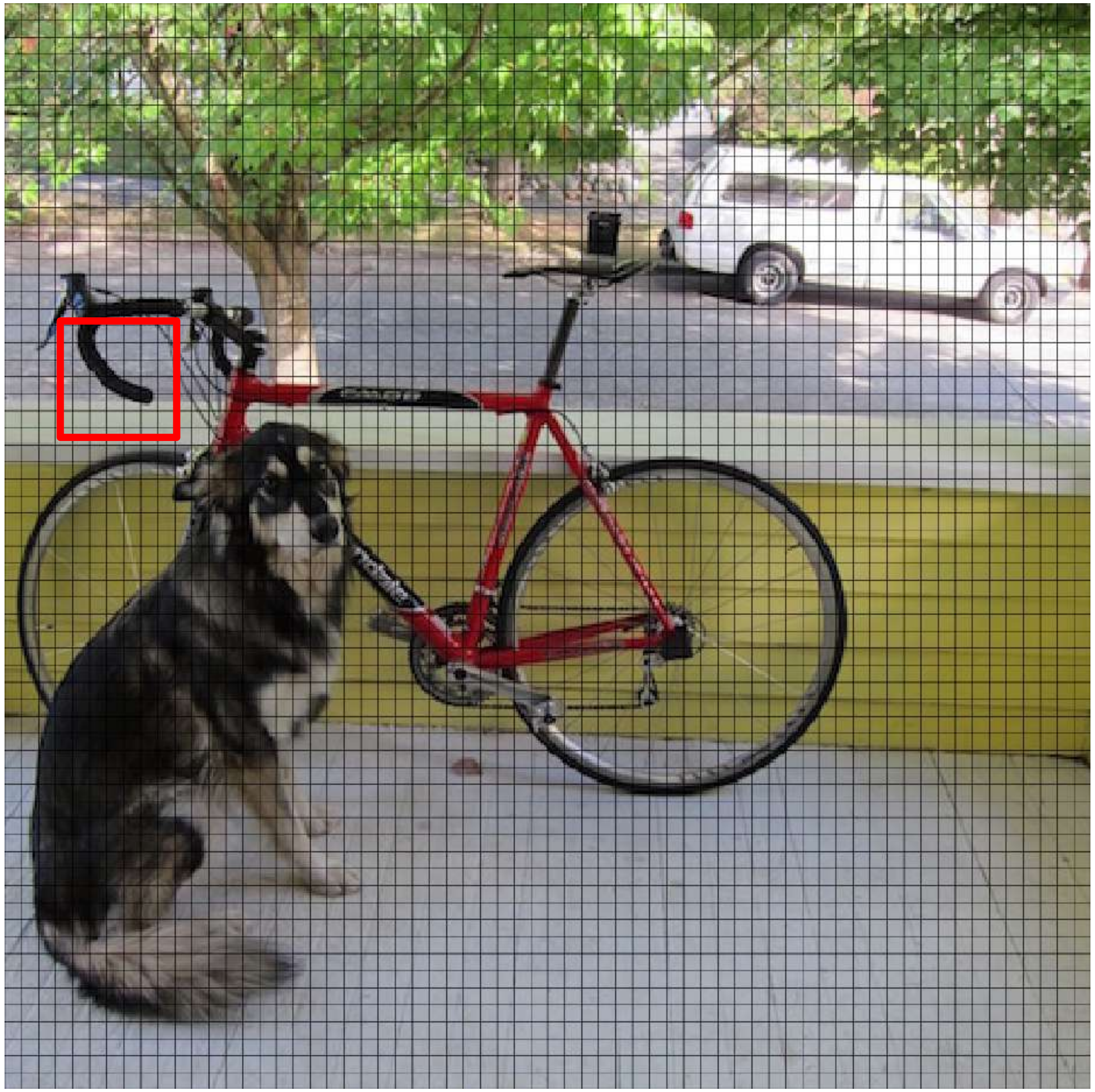


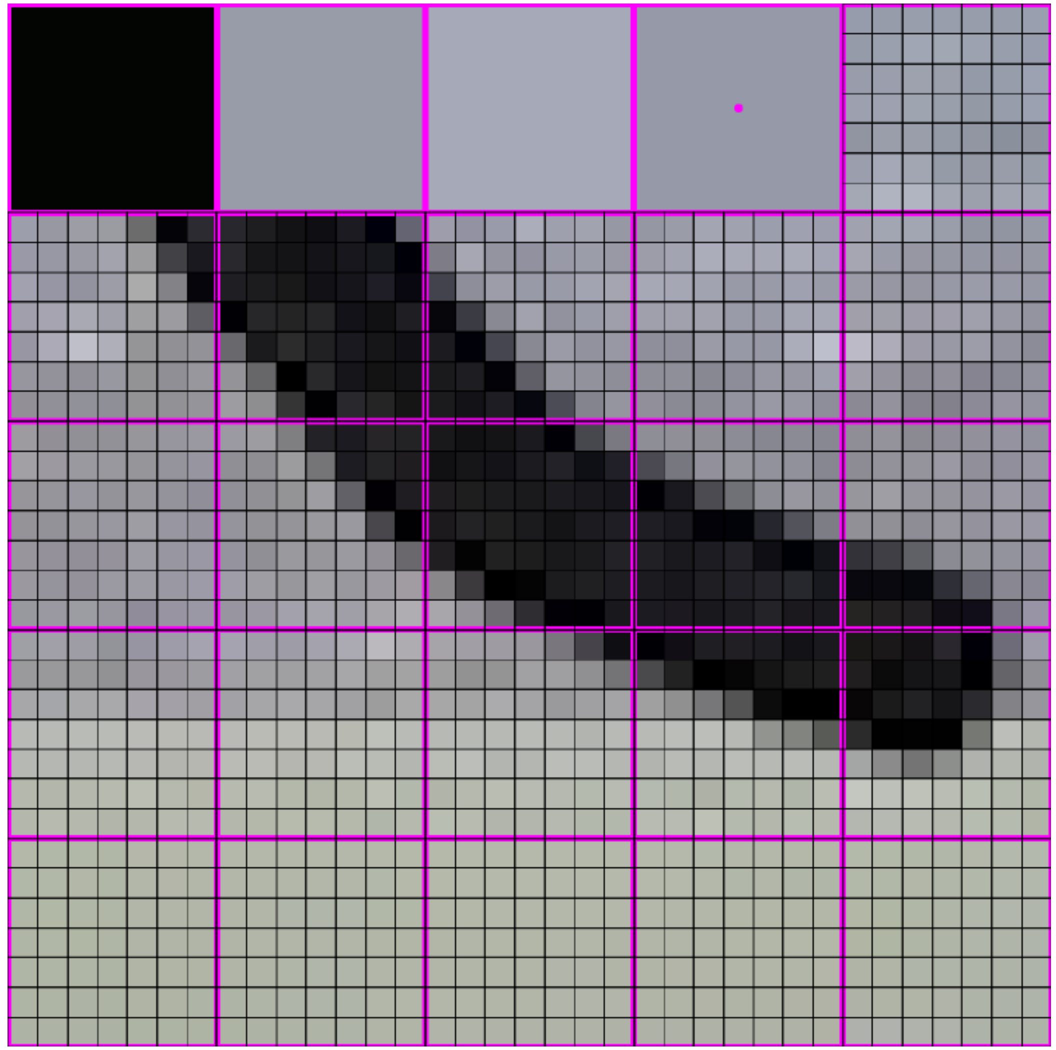
448x448 -> 64x64



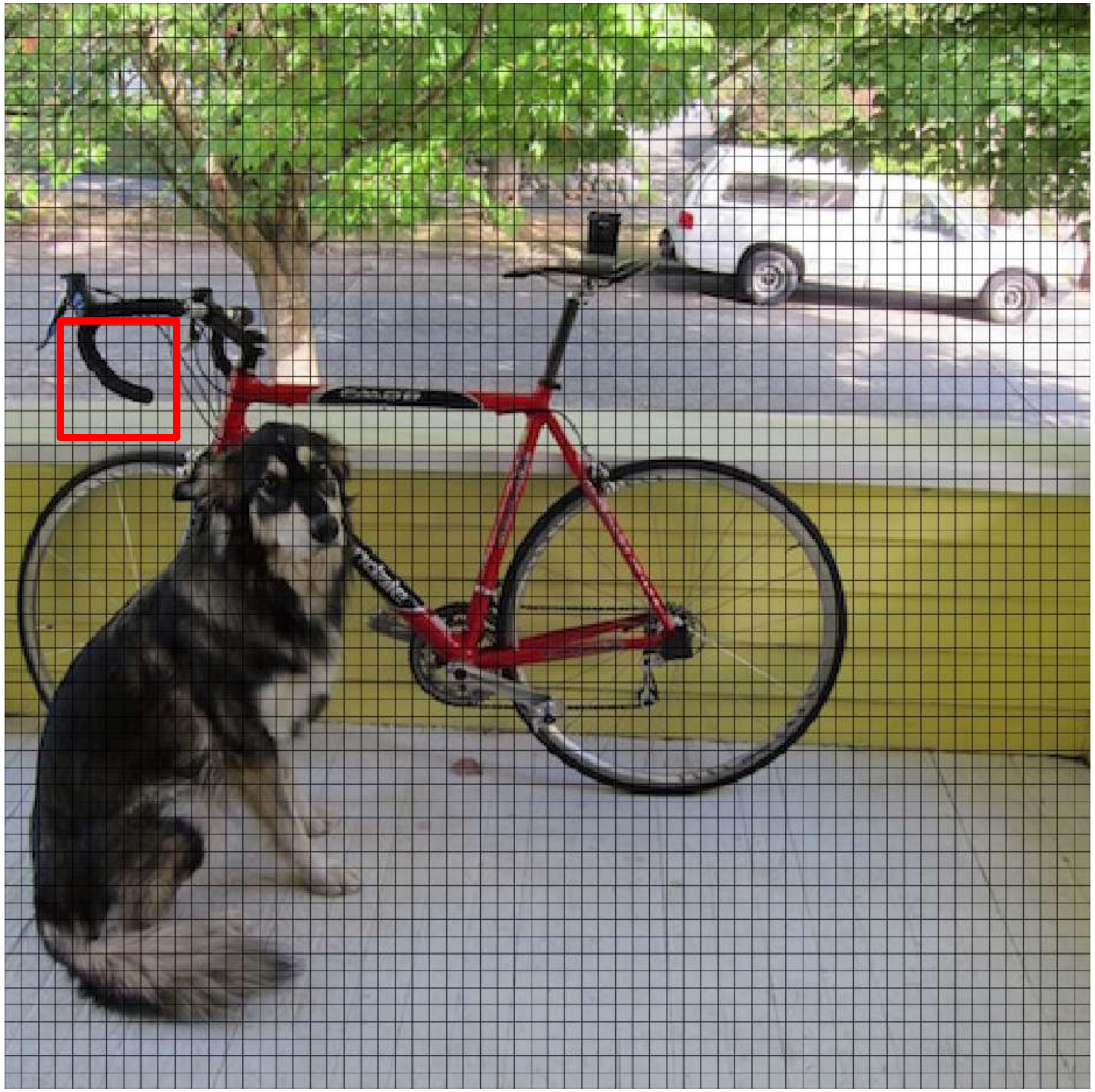


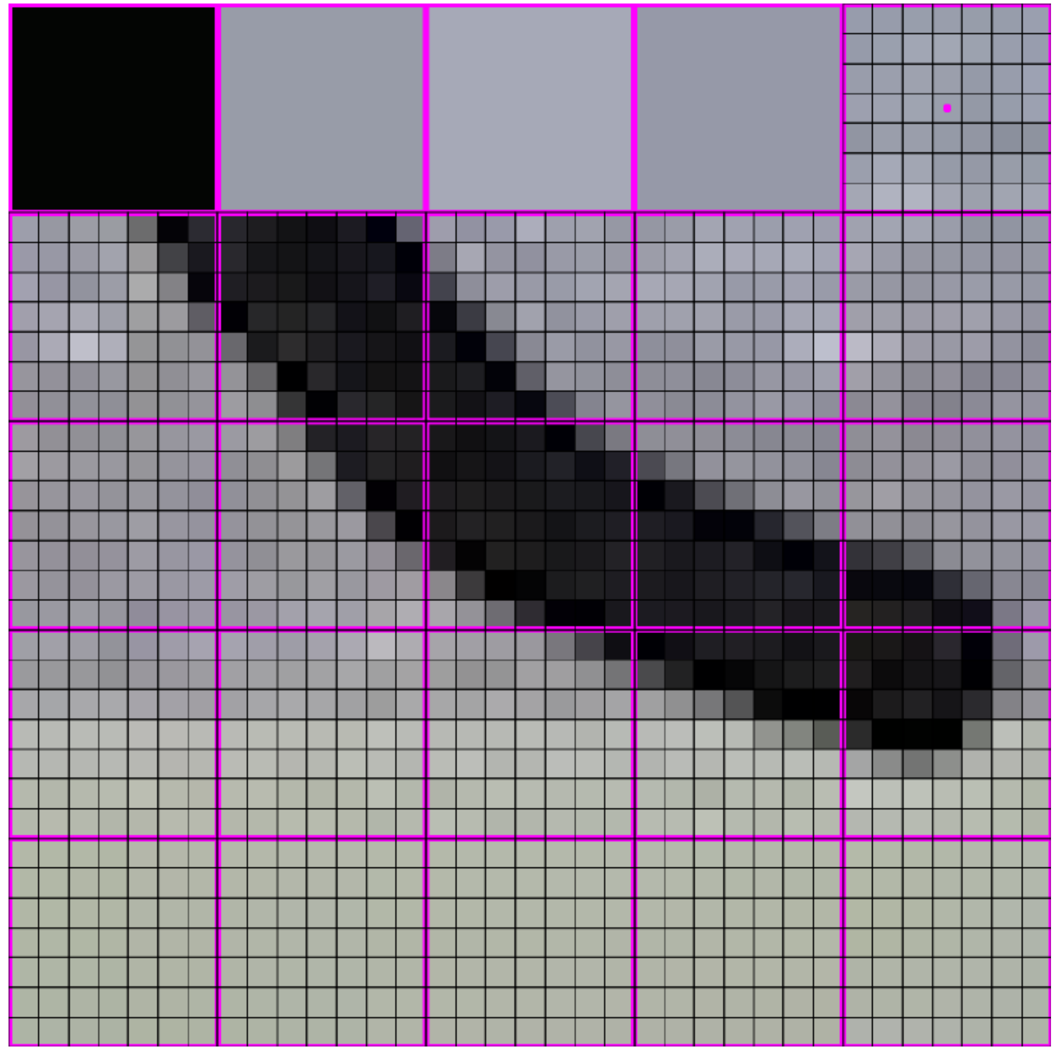
448x448 -> 64x64





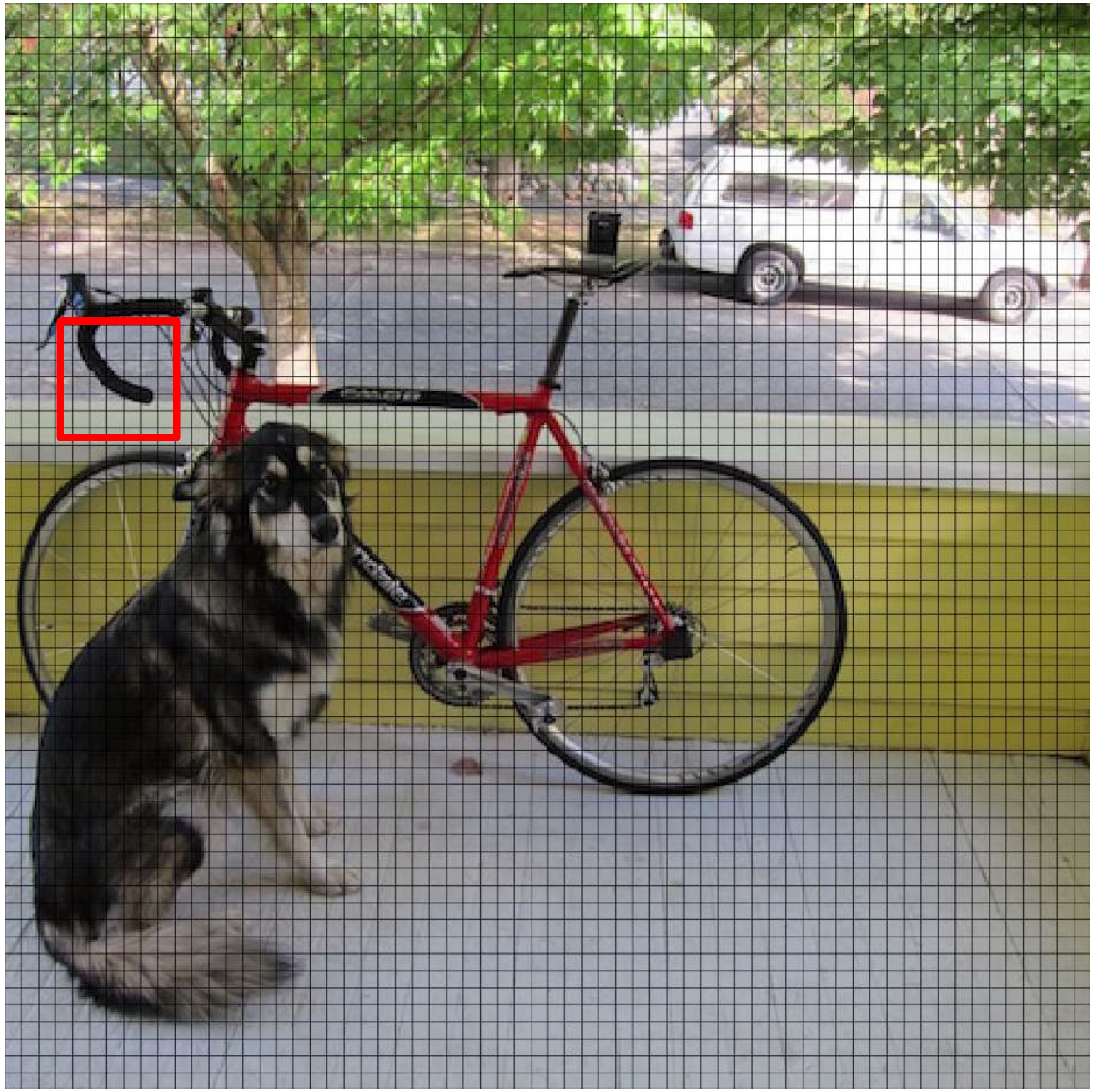
448x448 -> 64x64

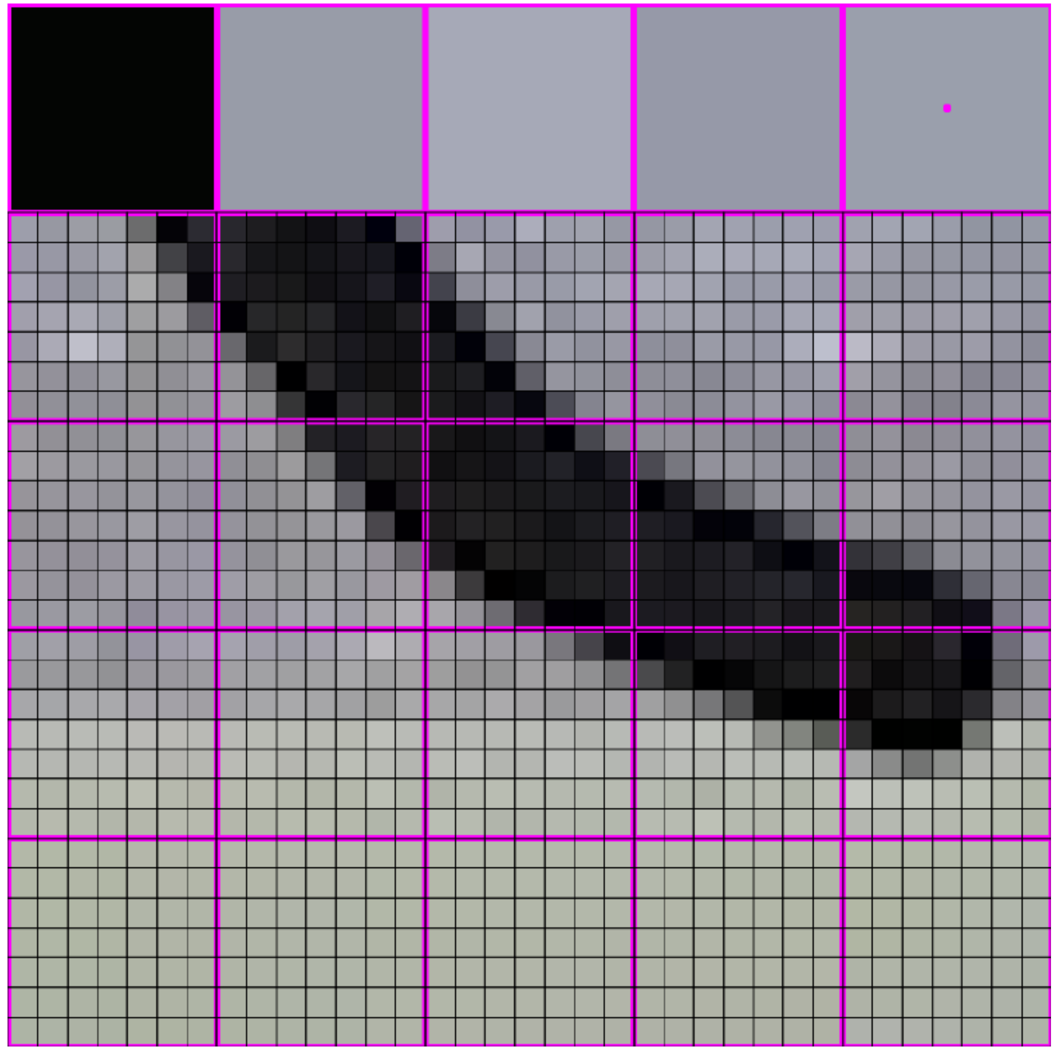




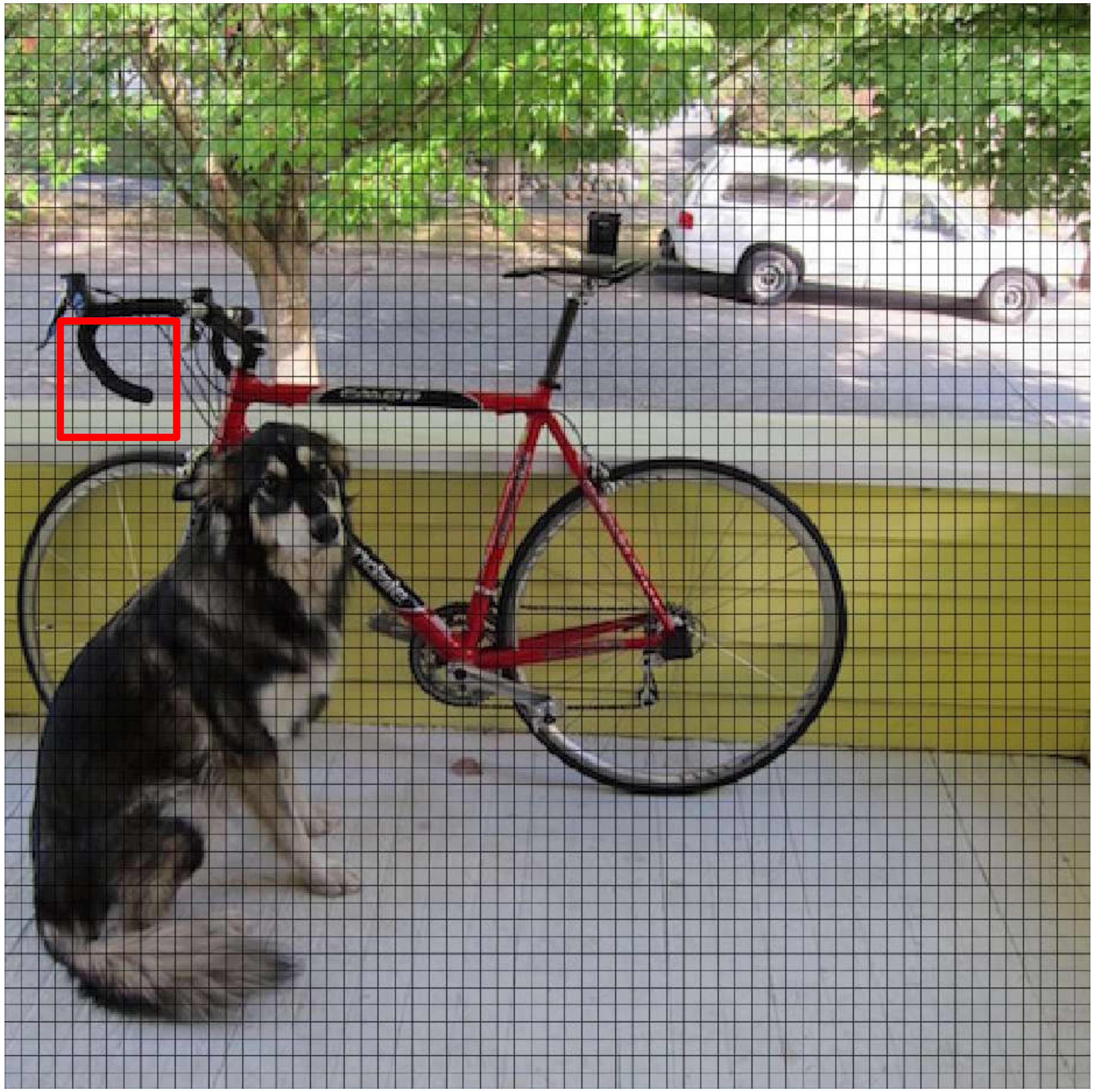
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

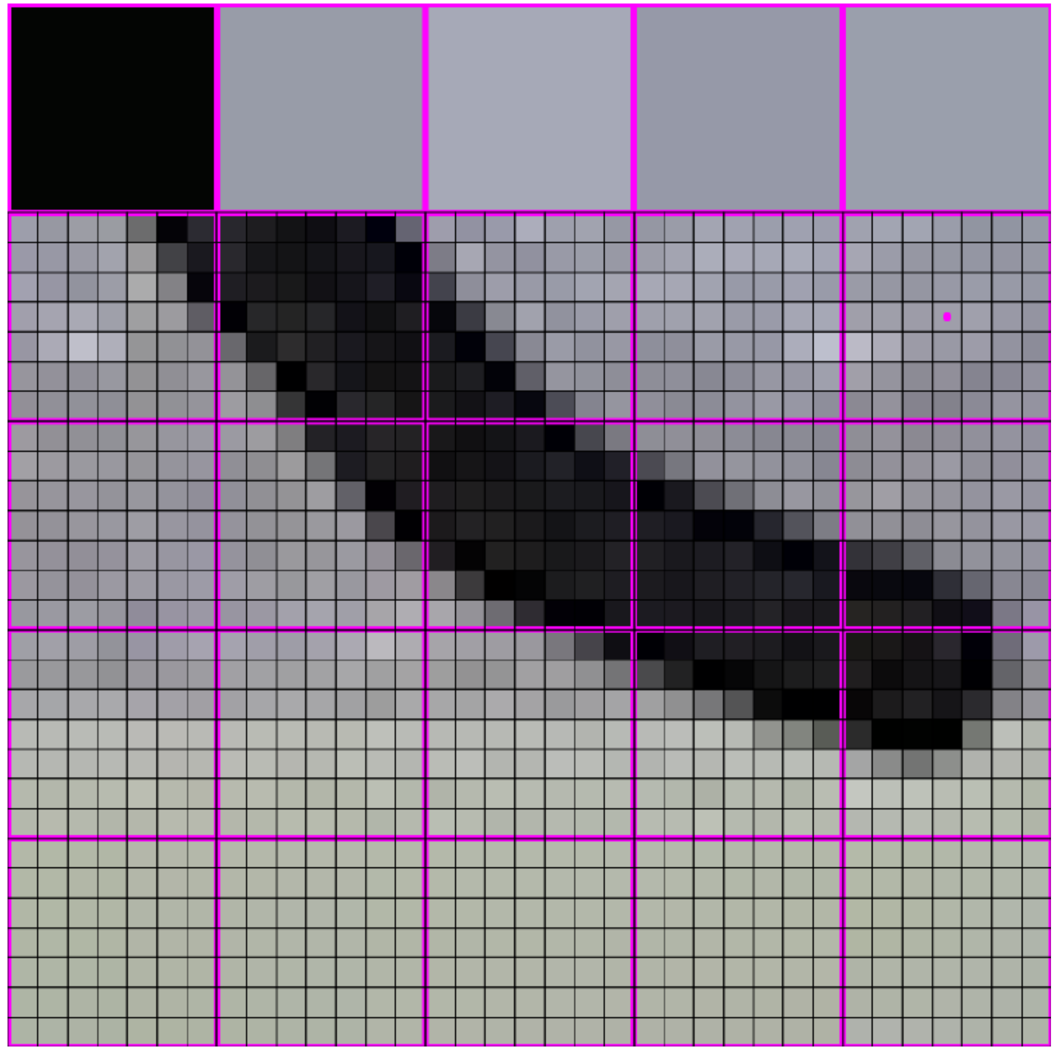
448x448 -> 64x64



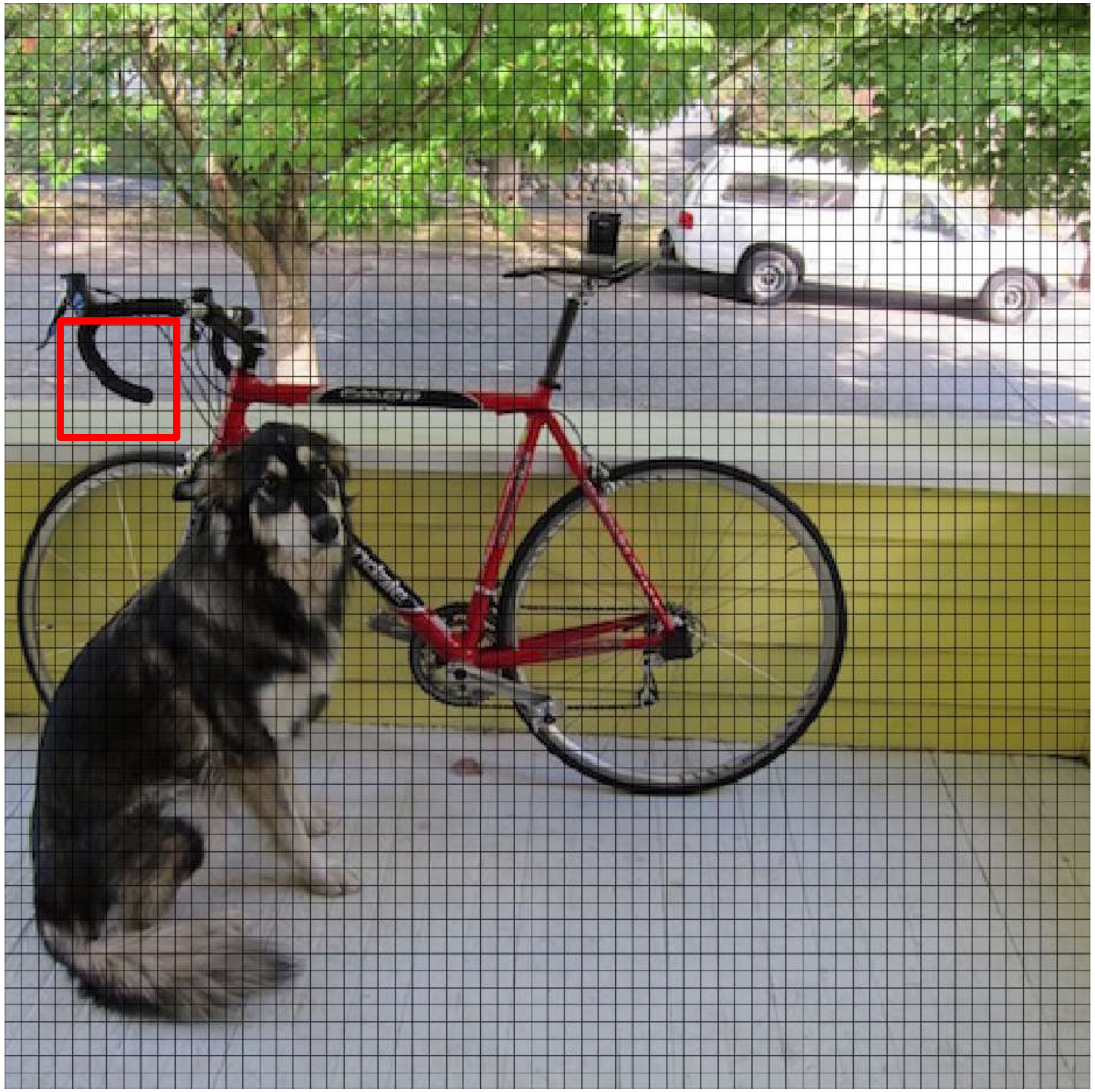


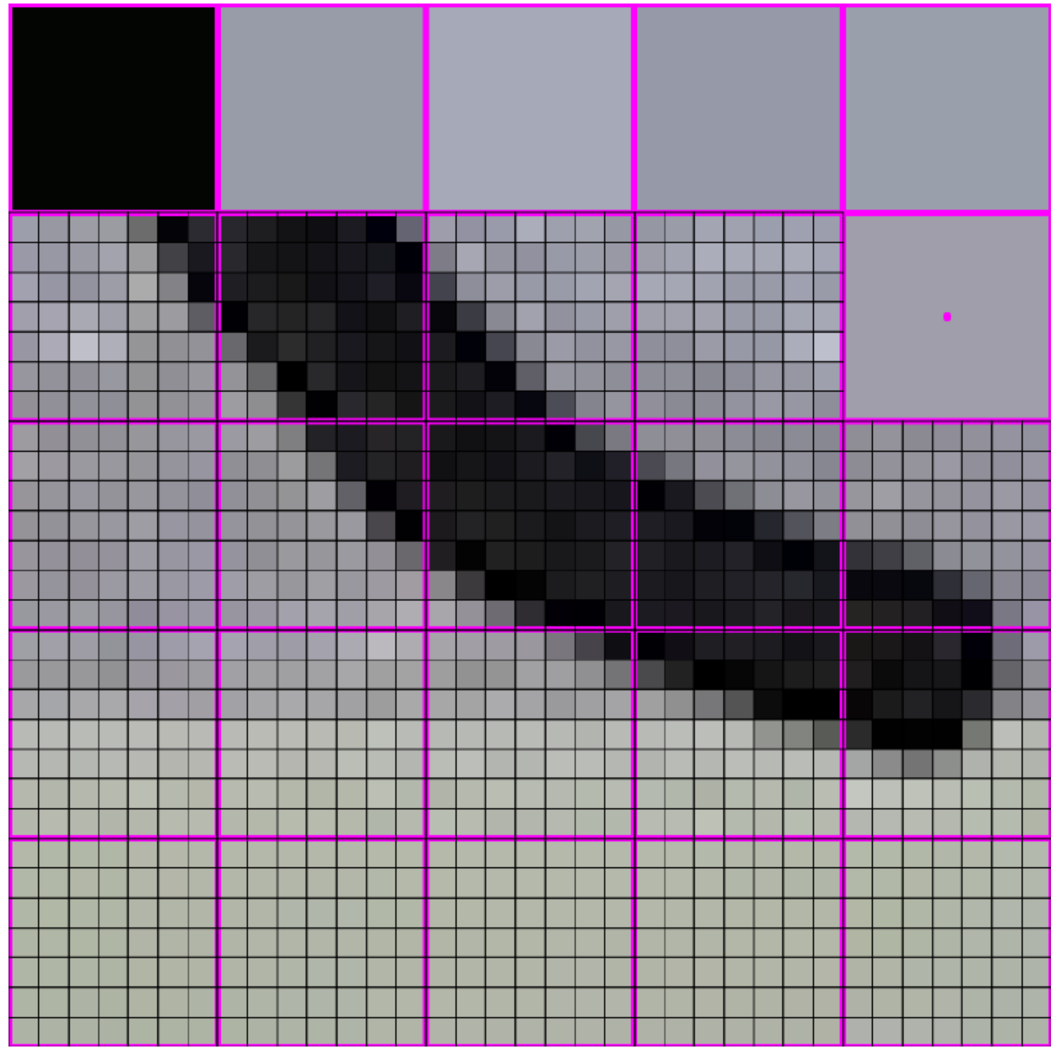
448x448 -> 64x64



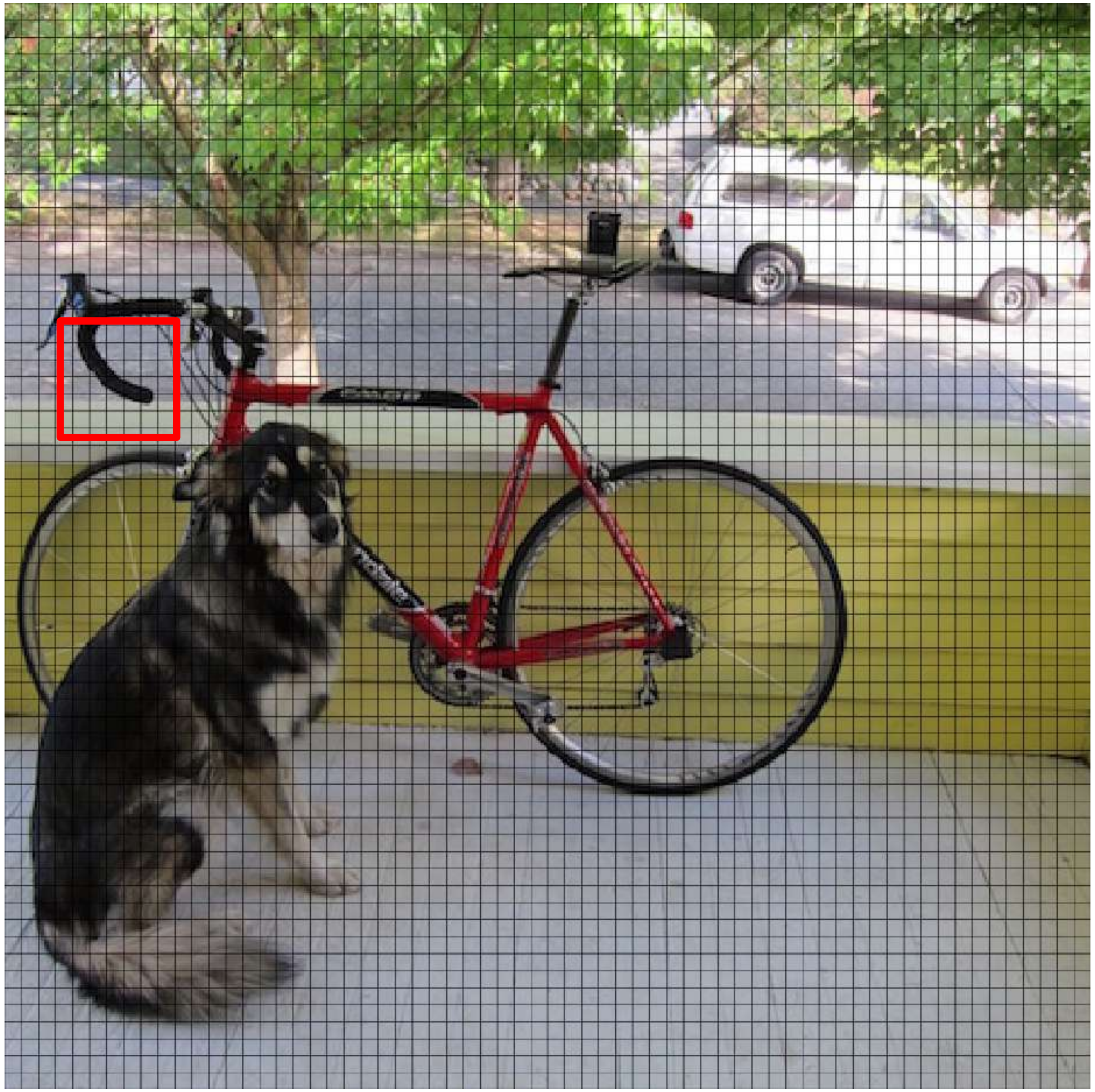


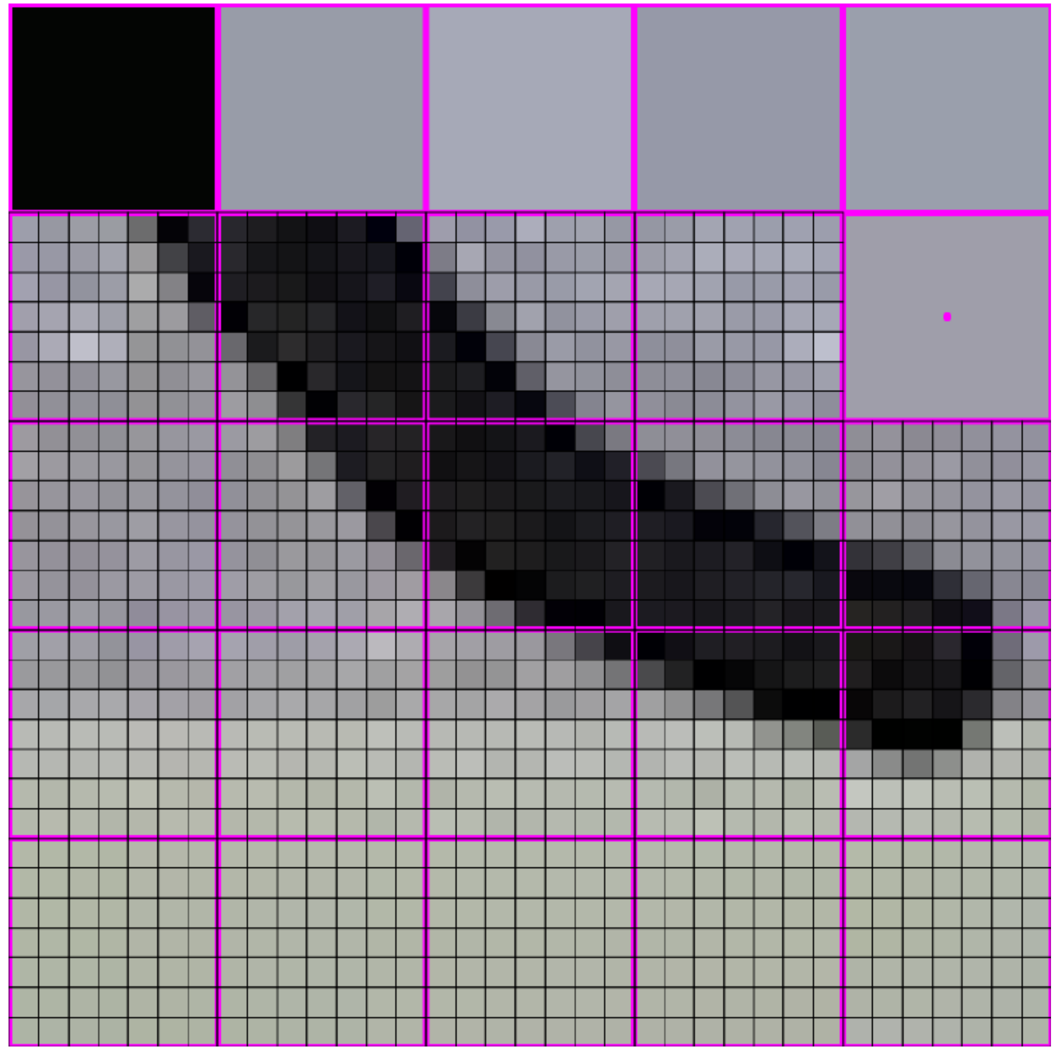
448x448 -> 64x64



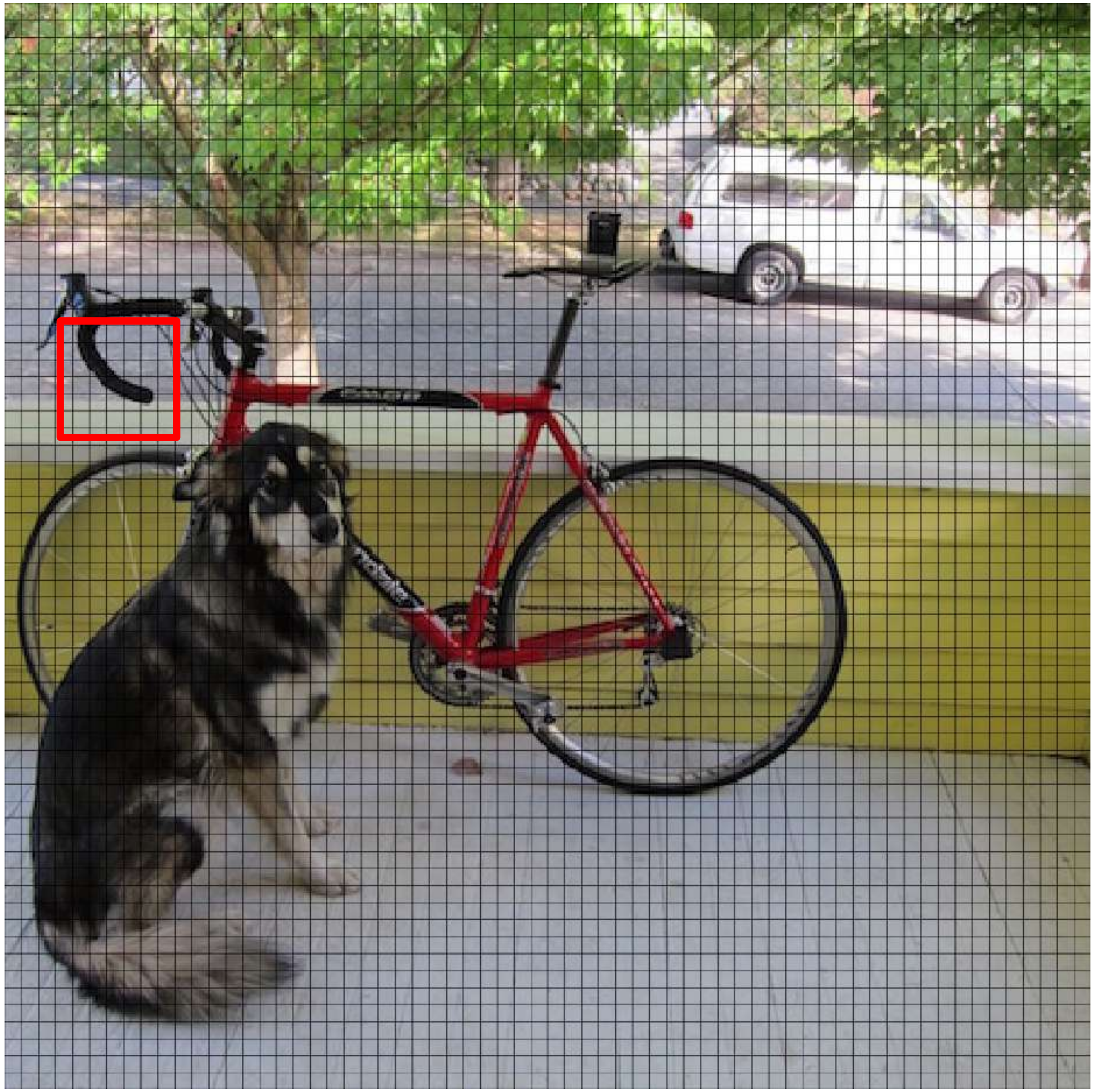


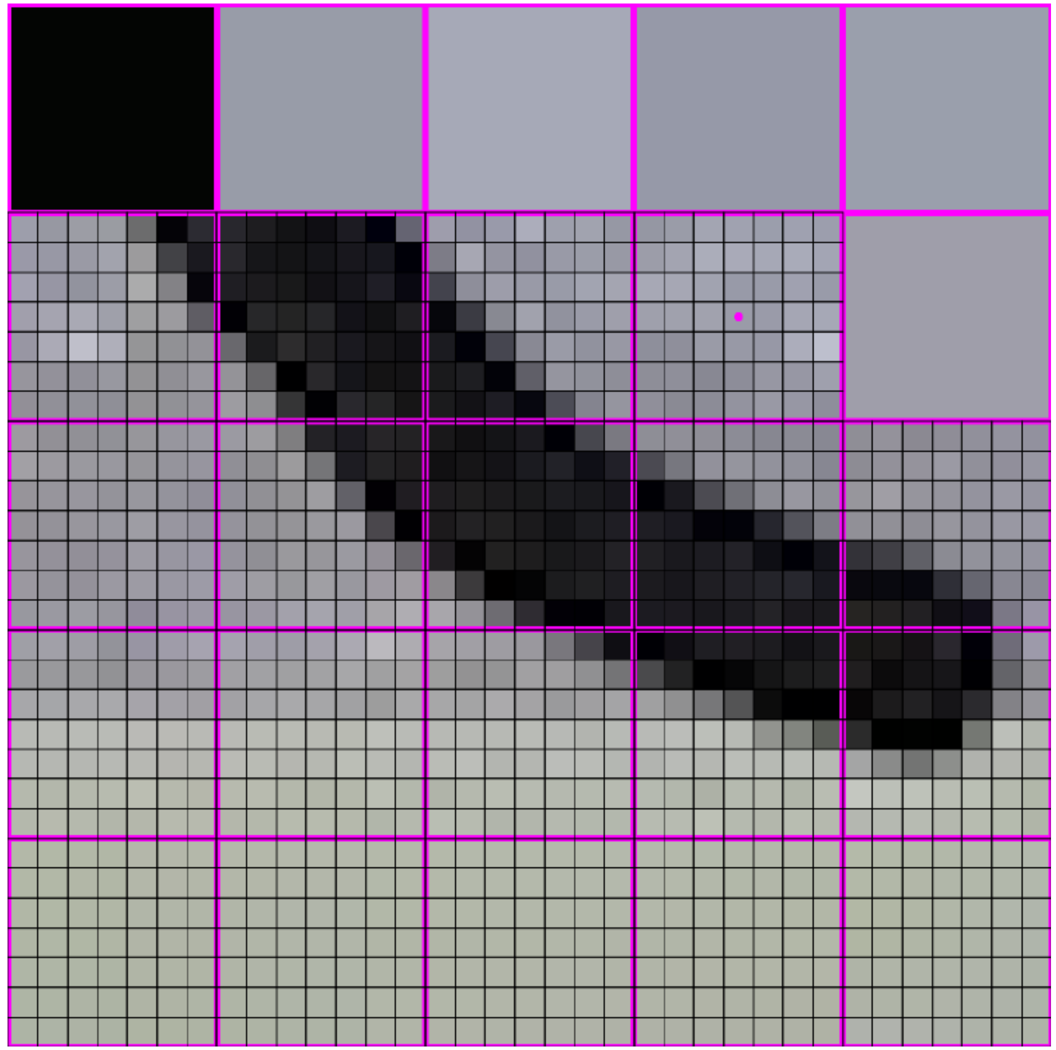
448x448 -> 64x64



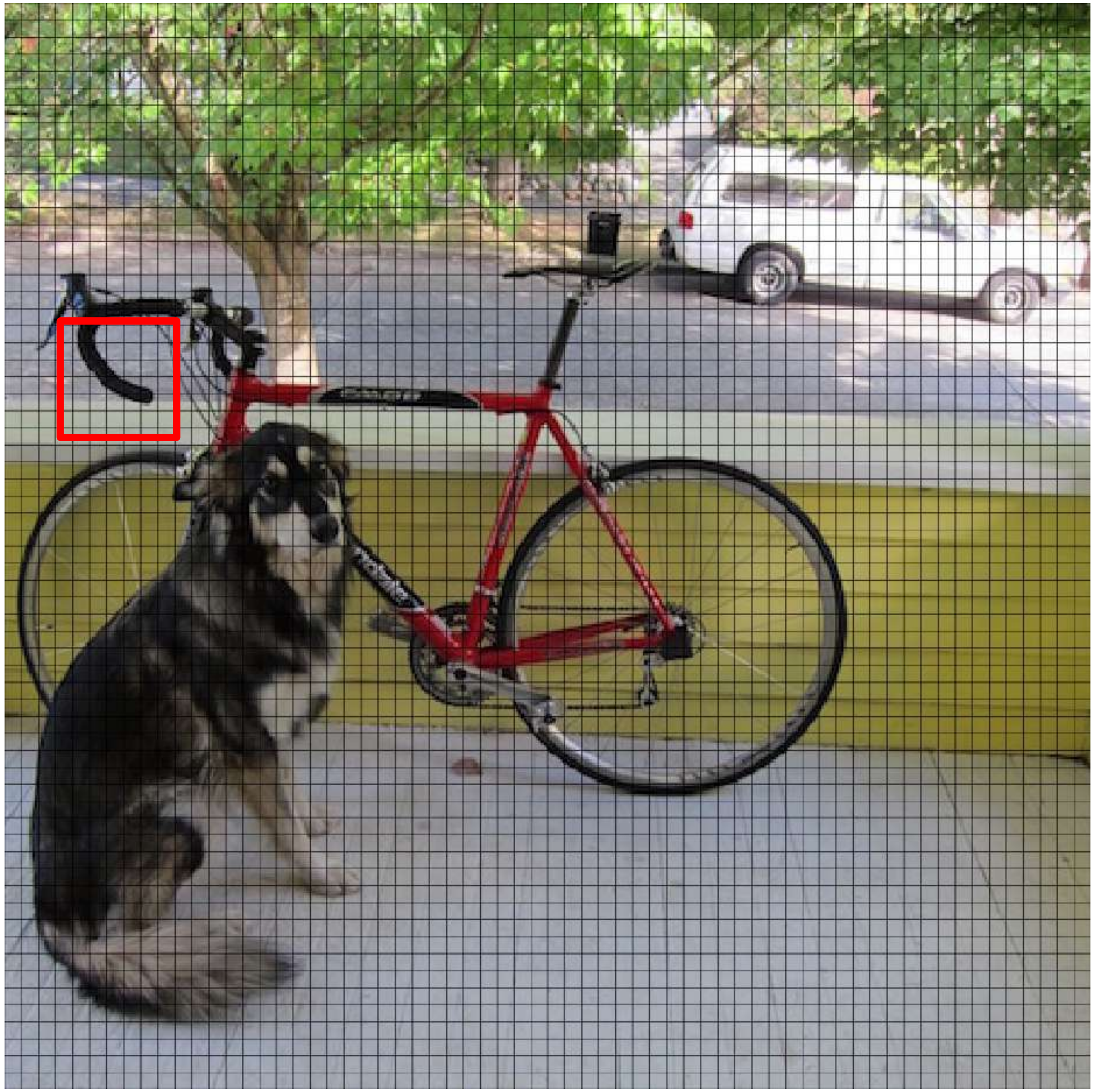


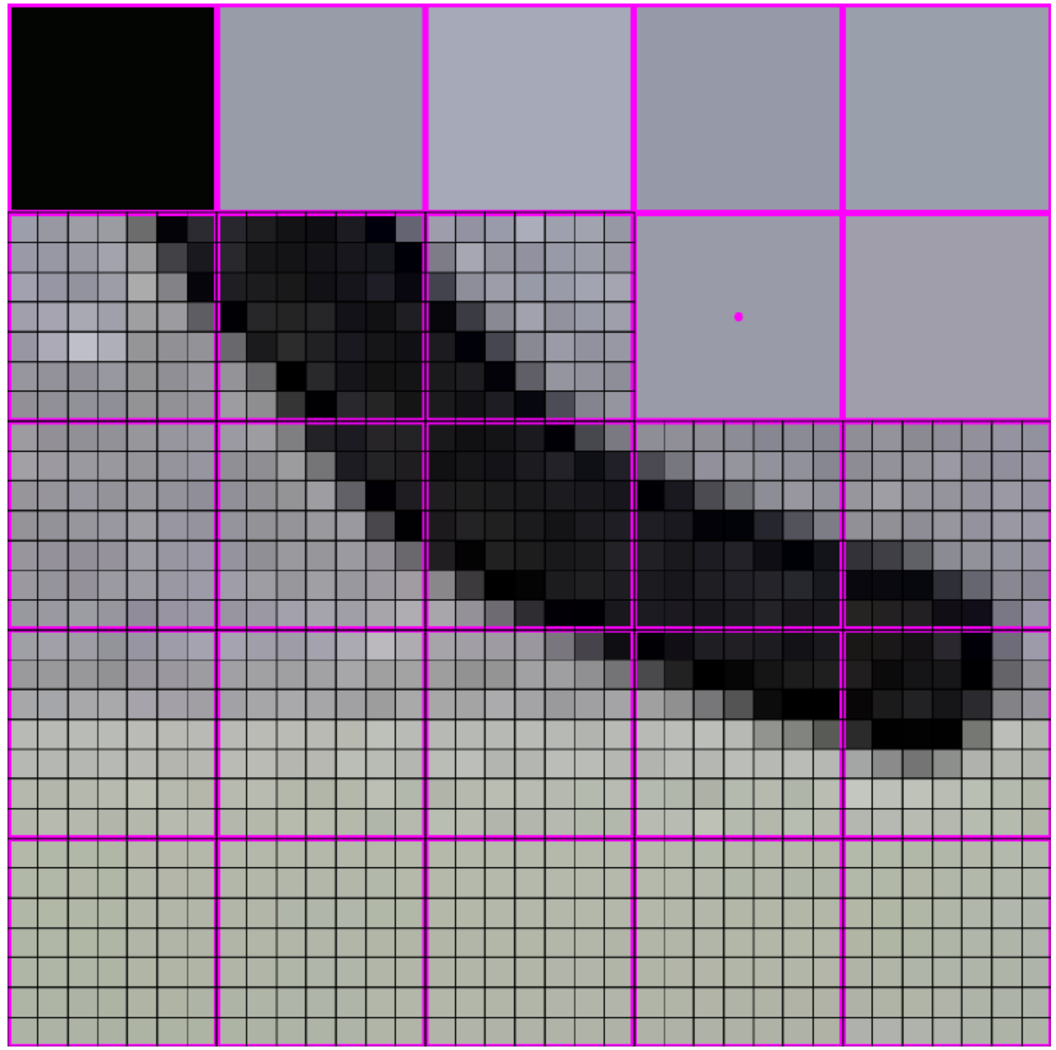
448x448 -> 64x64



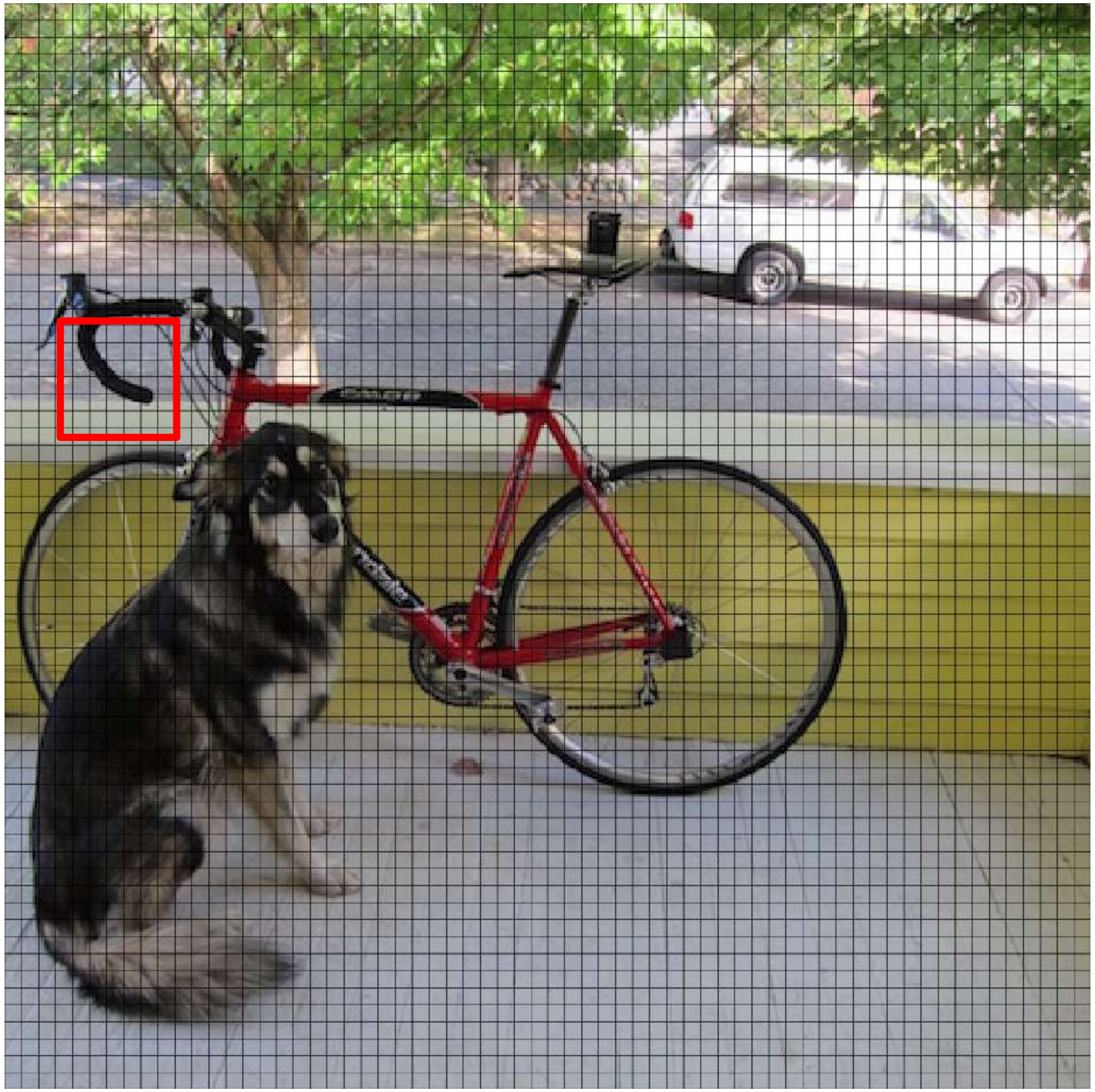


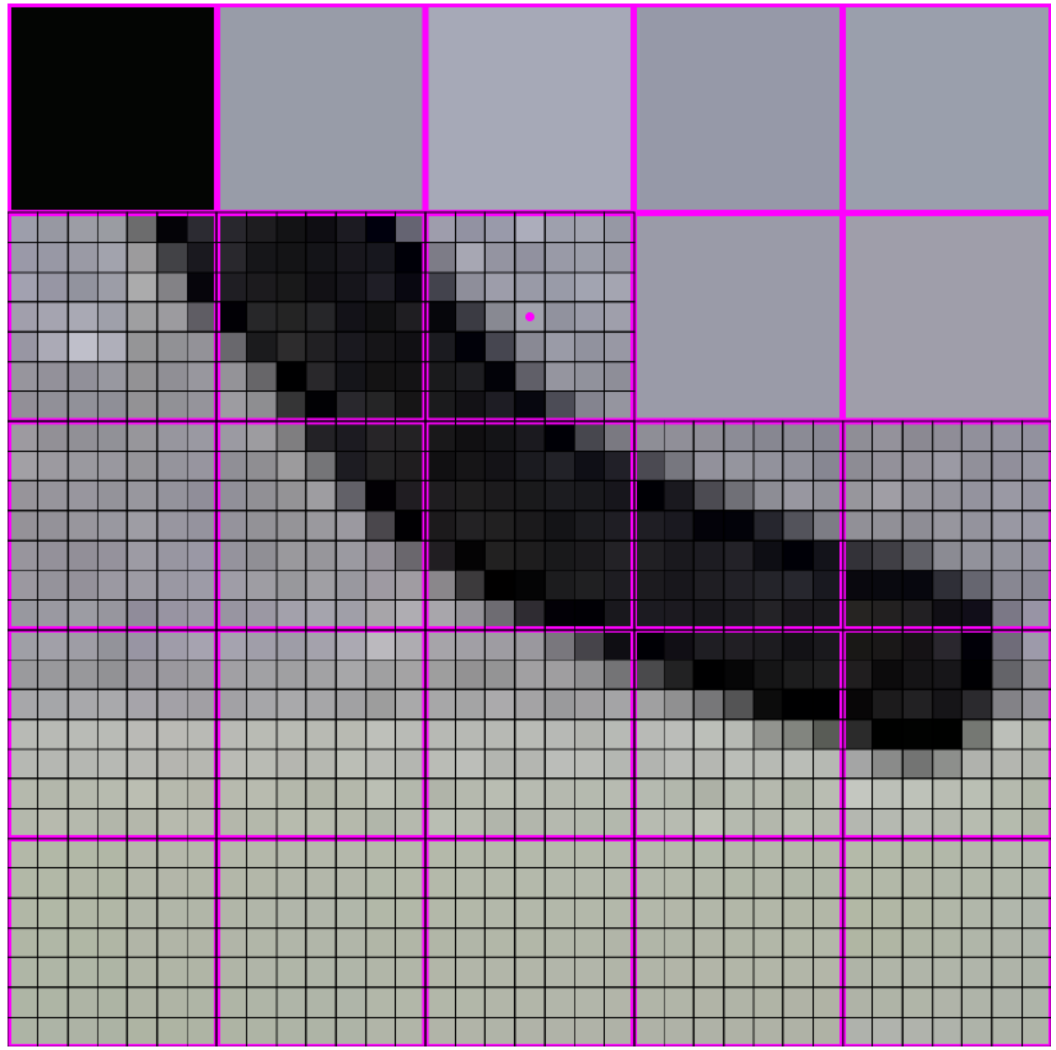
448x448 -> 64x64



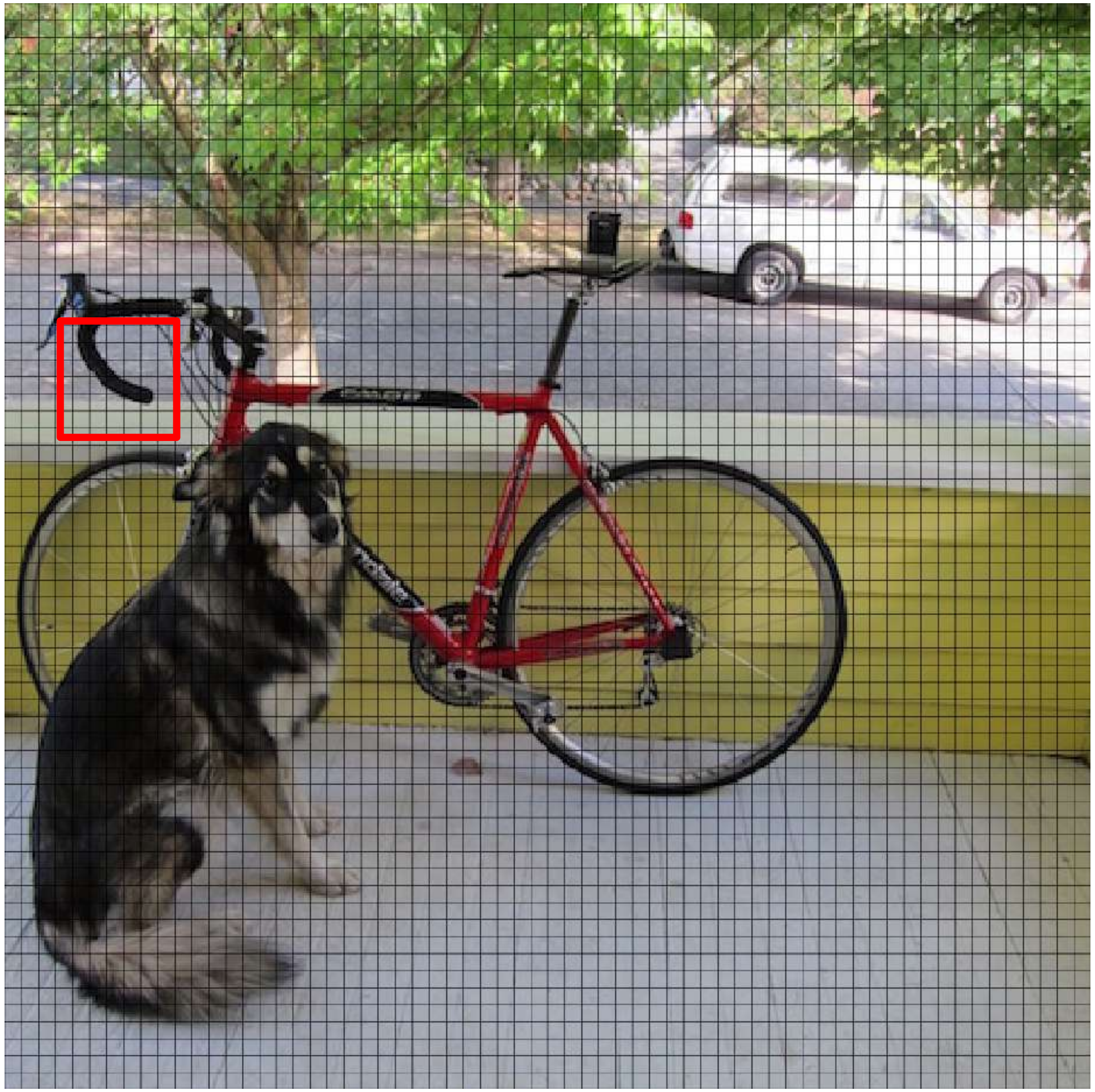


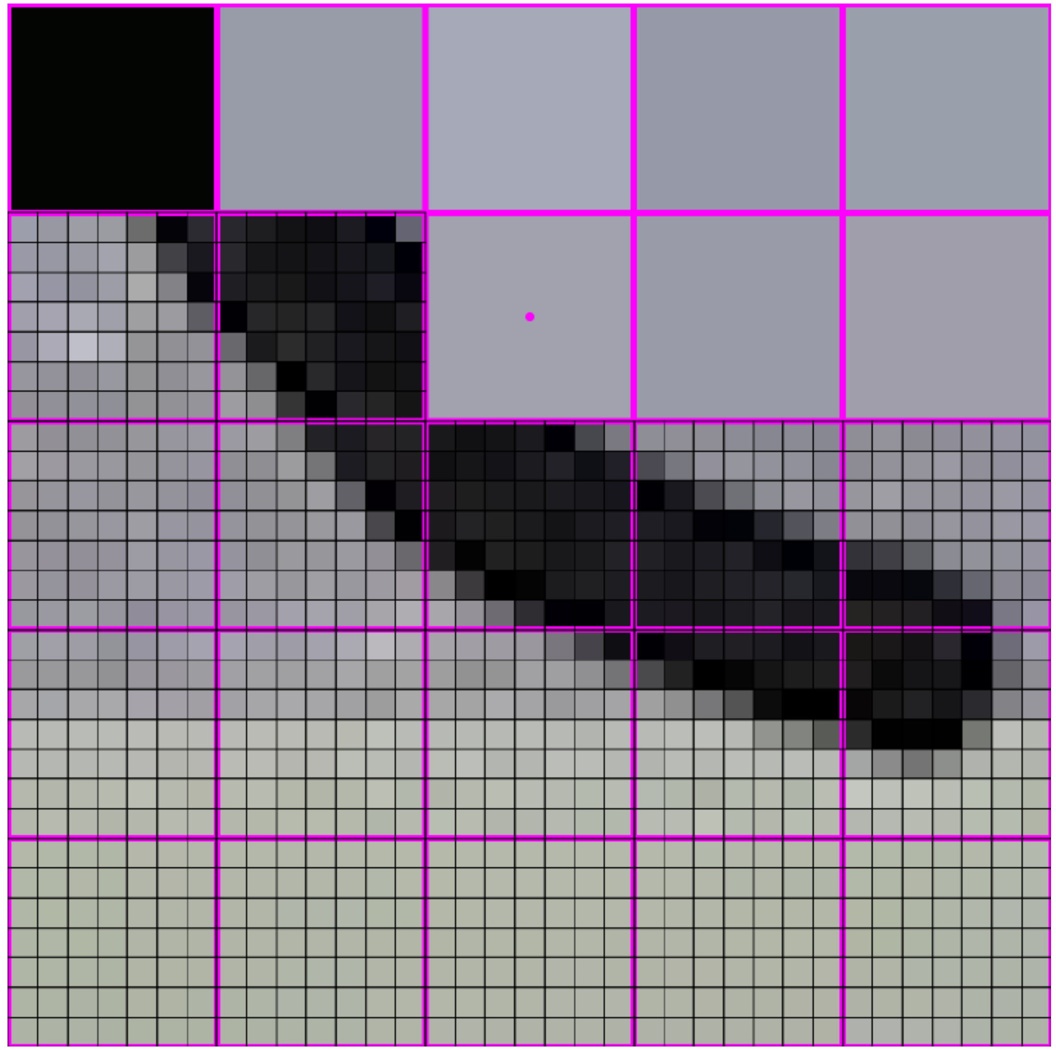
448x448 -> 64x64



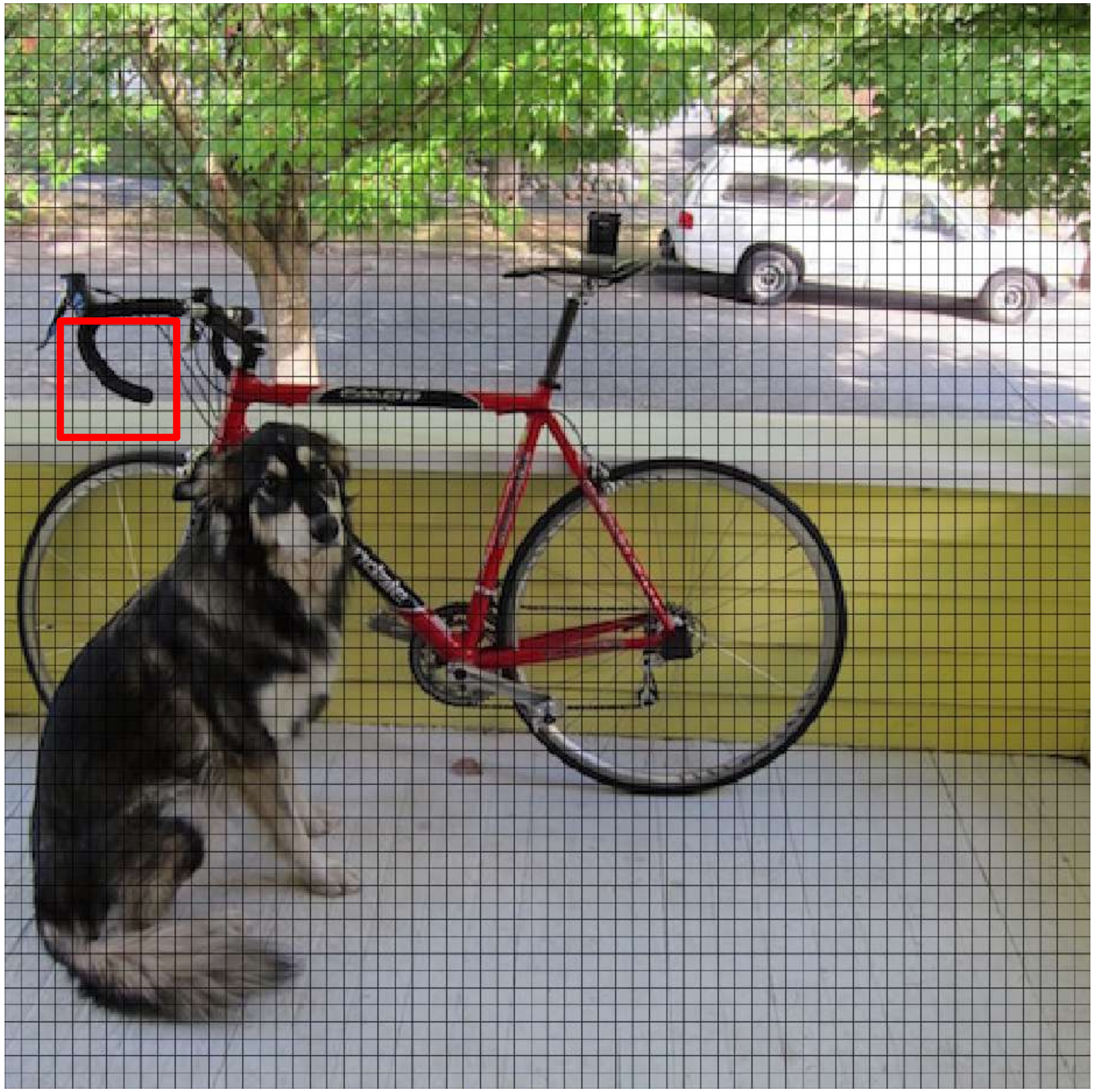


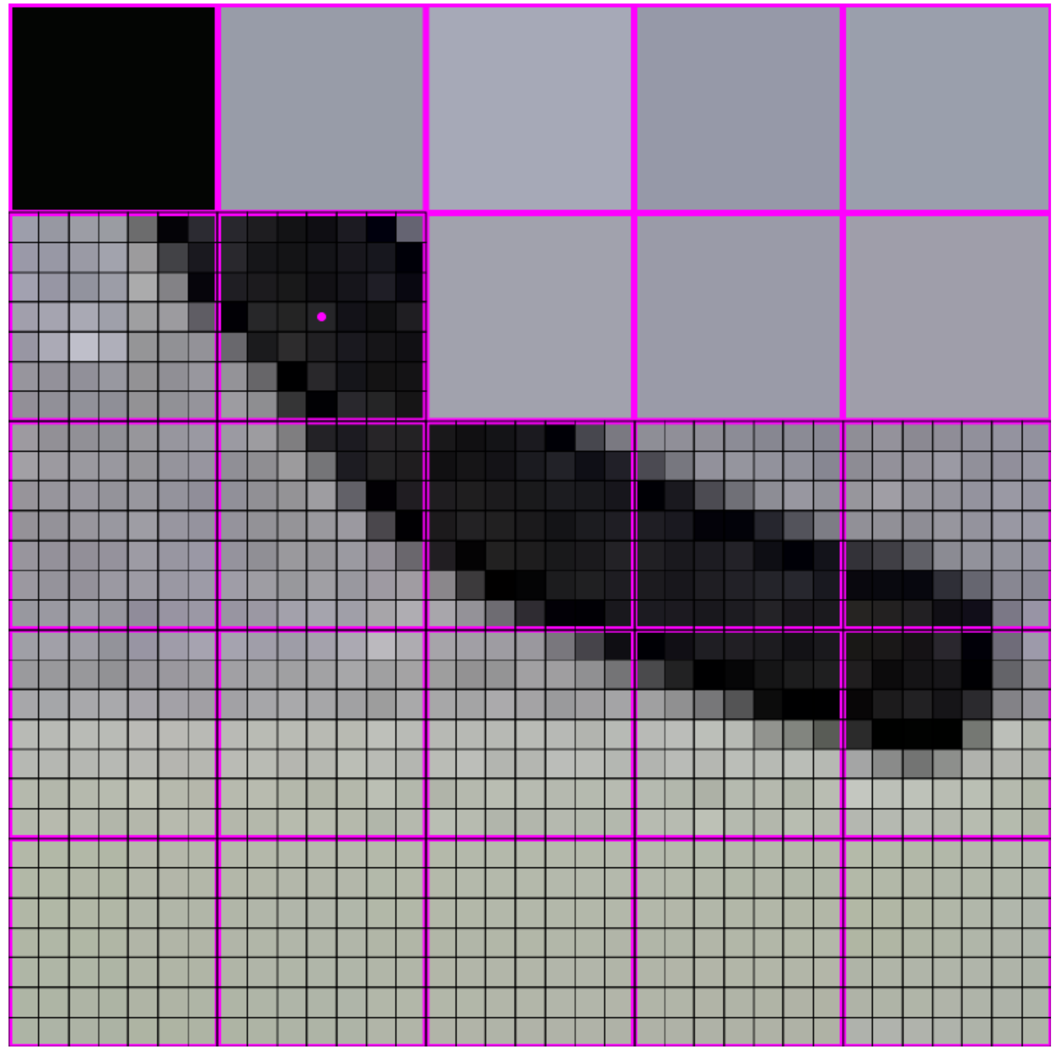
448x448 -> 64x64



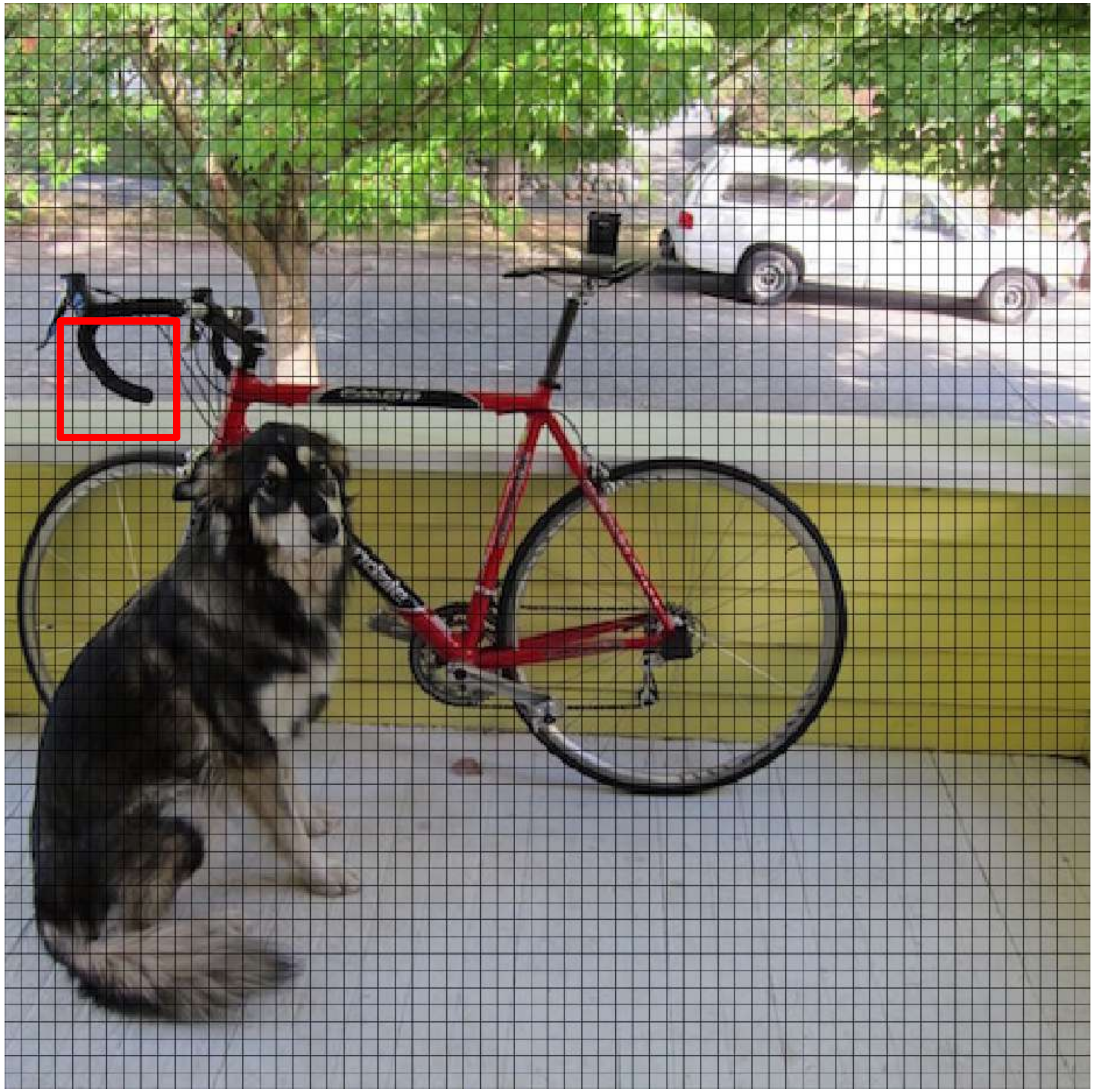


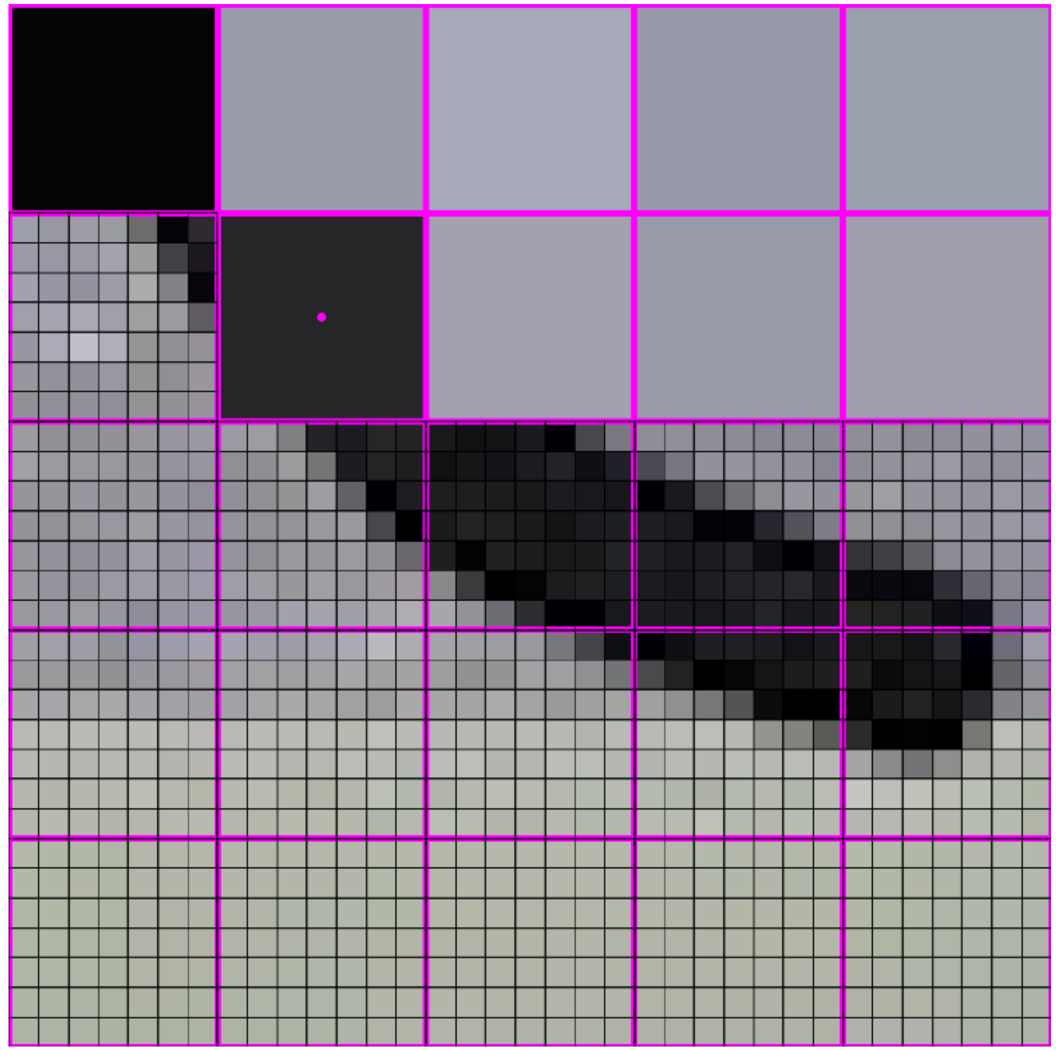
448x448 -> 64x64



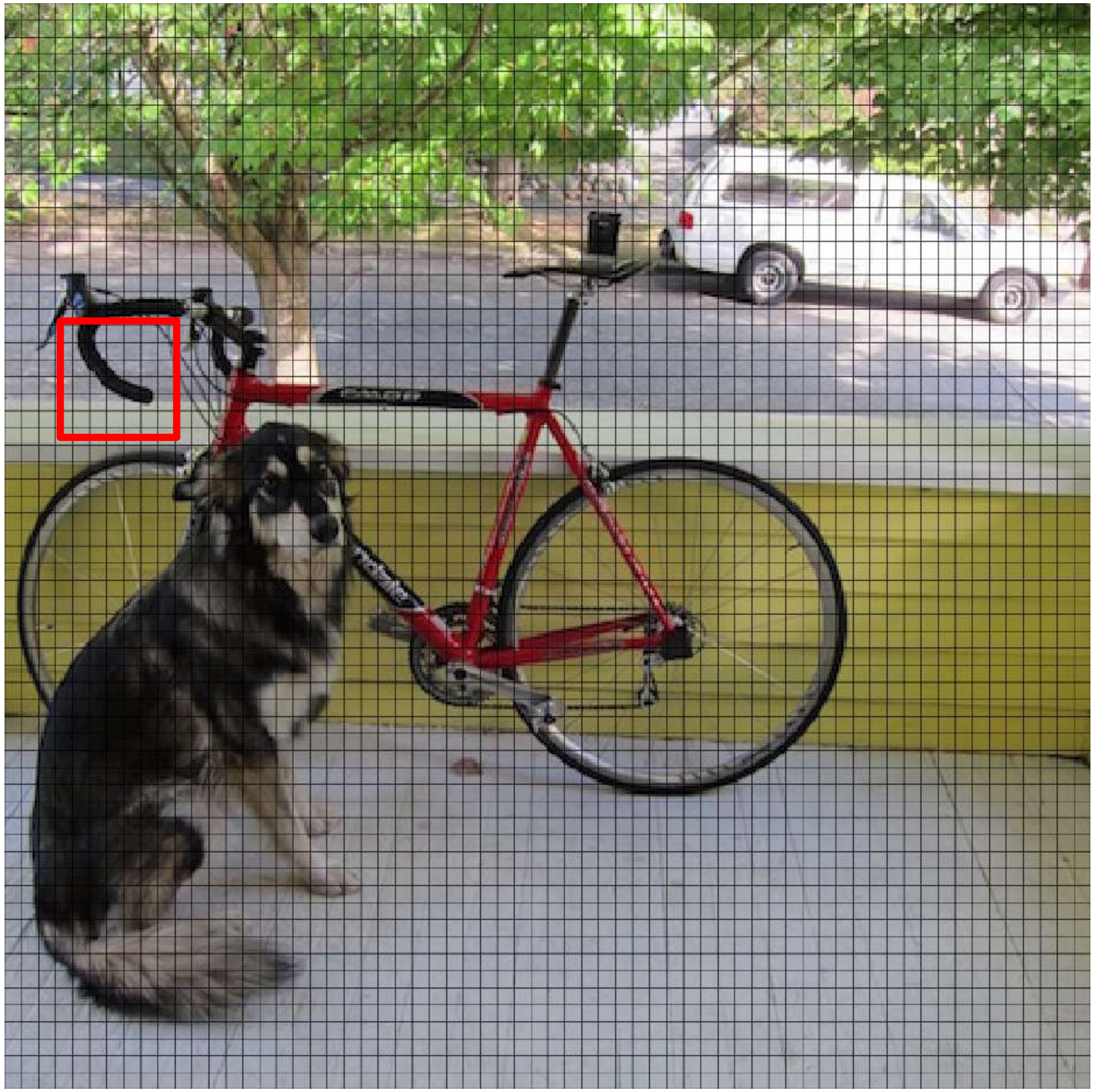


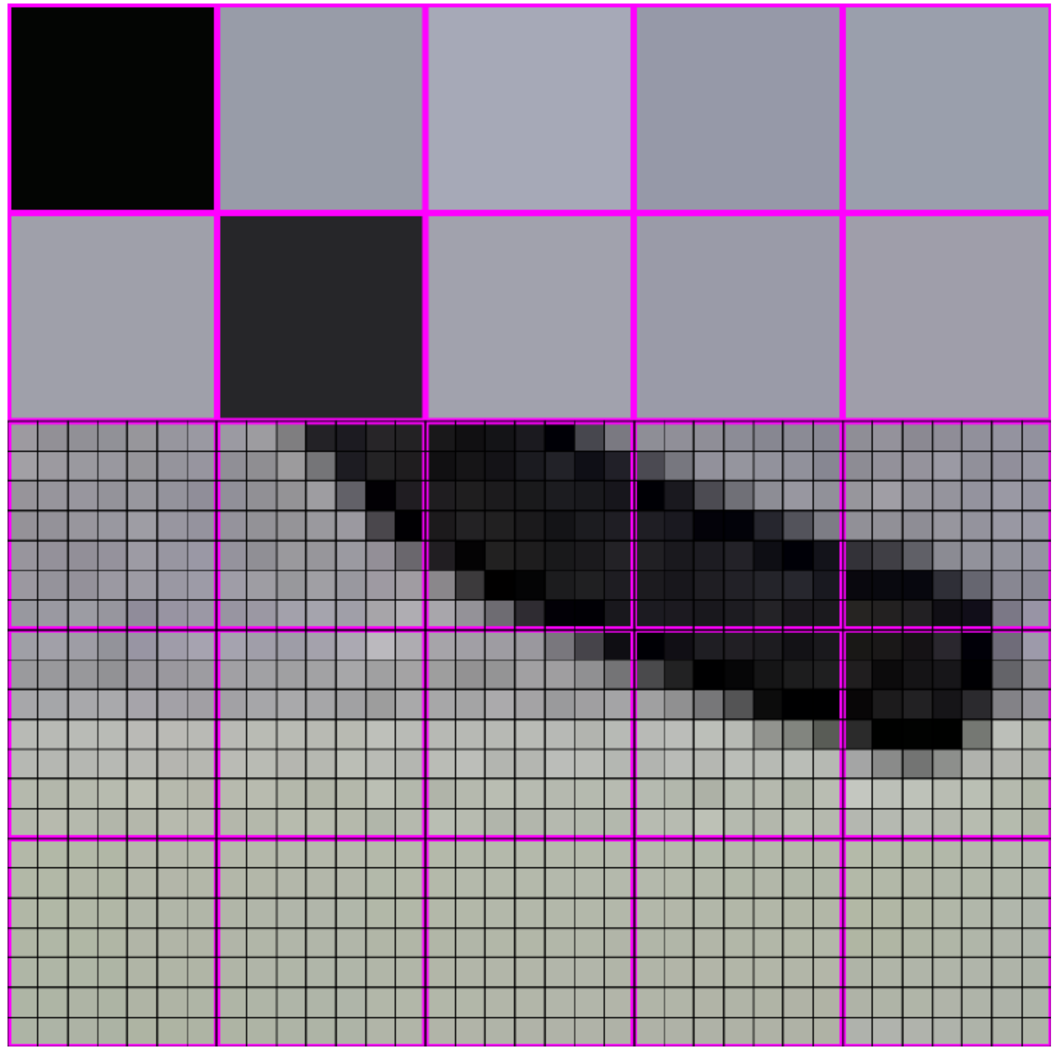
448x448 -> 64x64



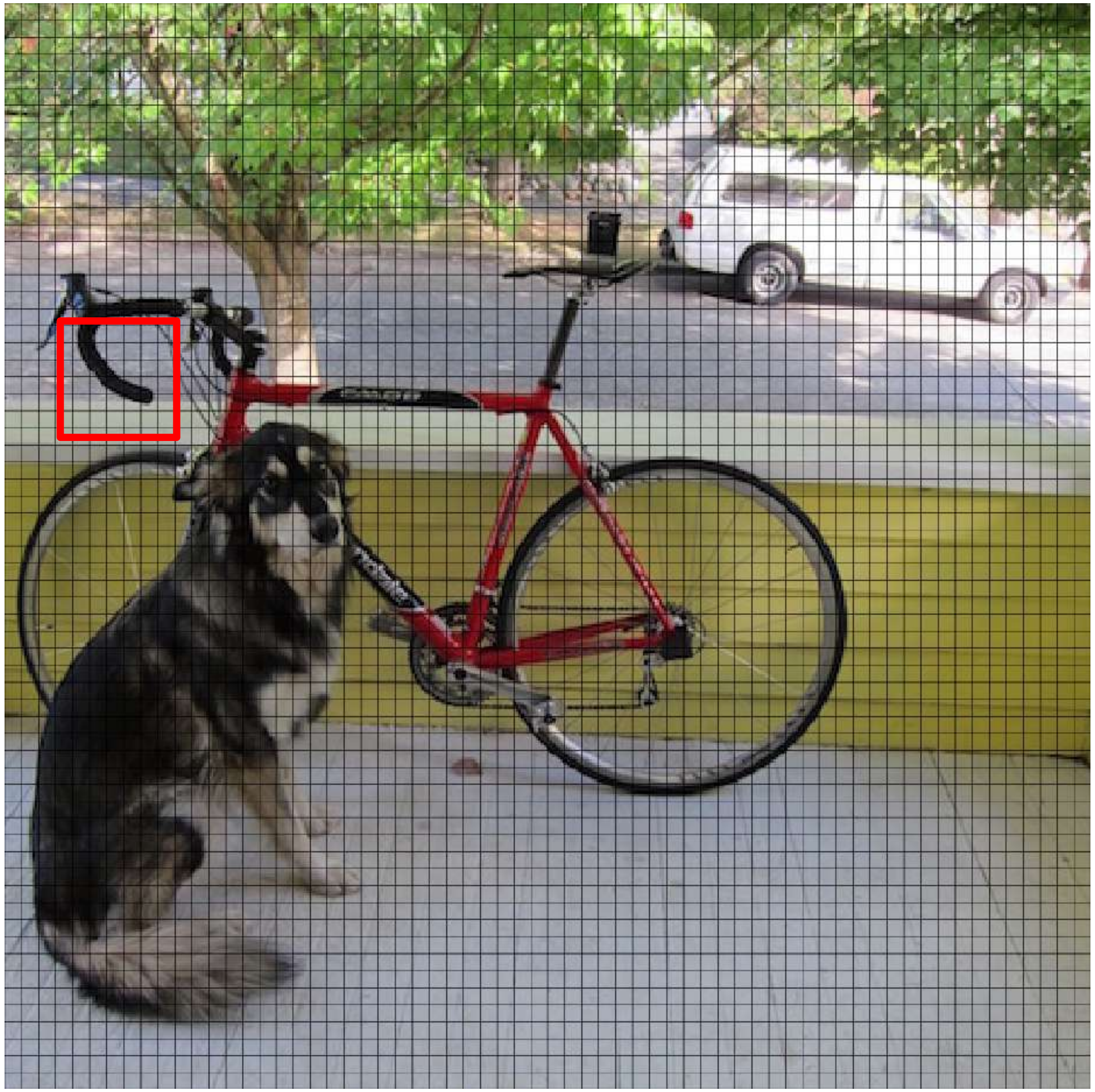


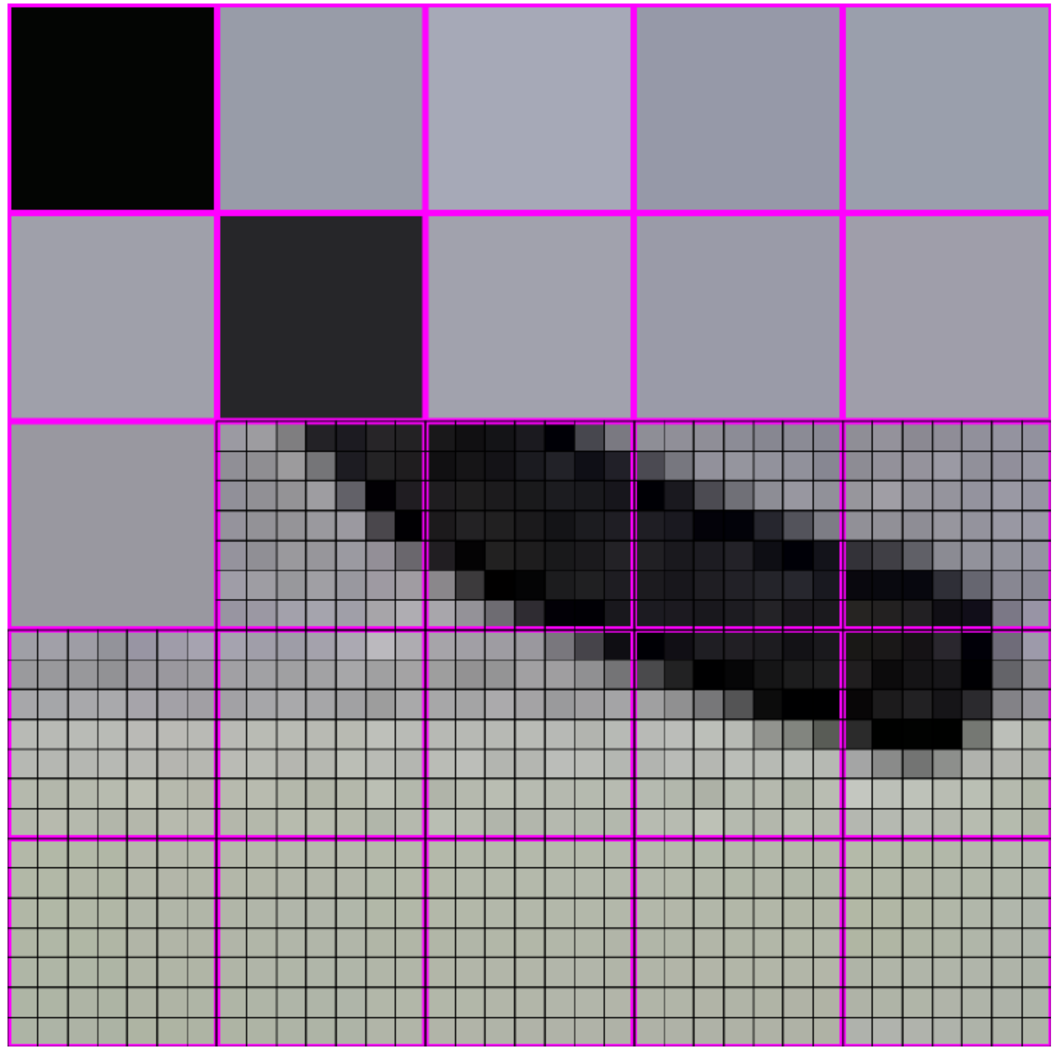
448x448 -> 64x64



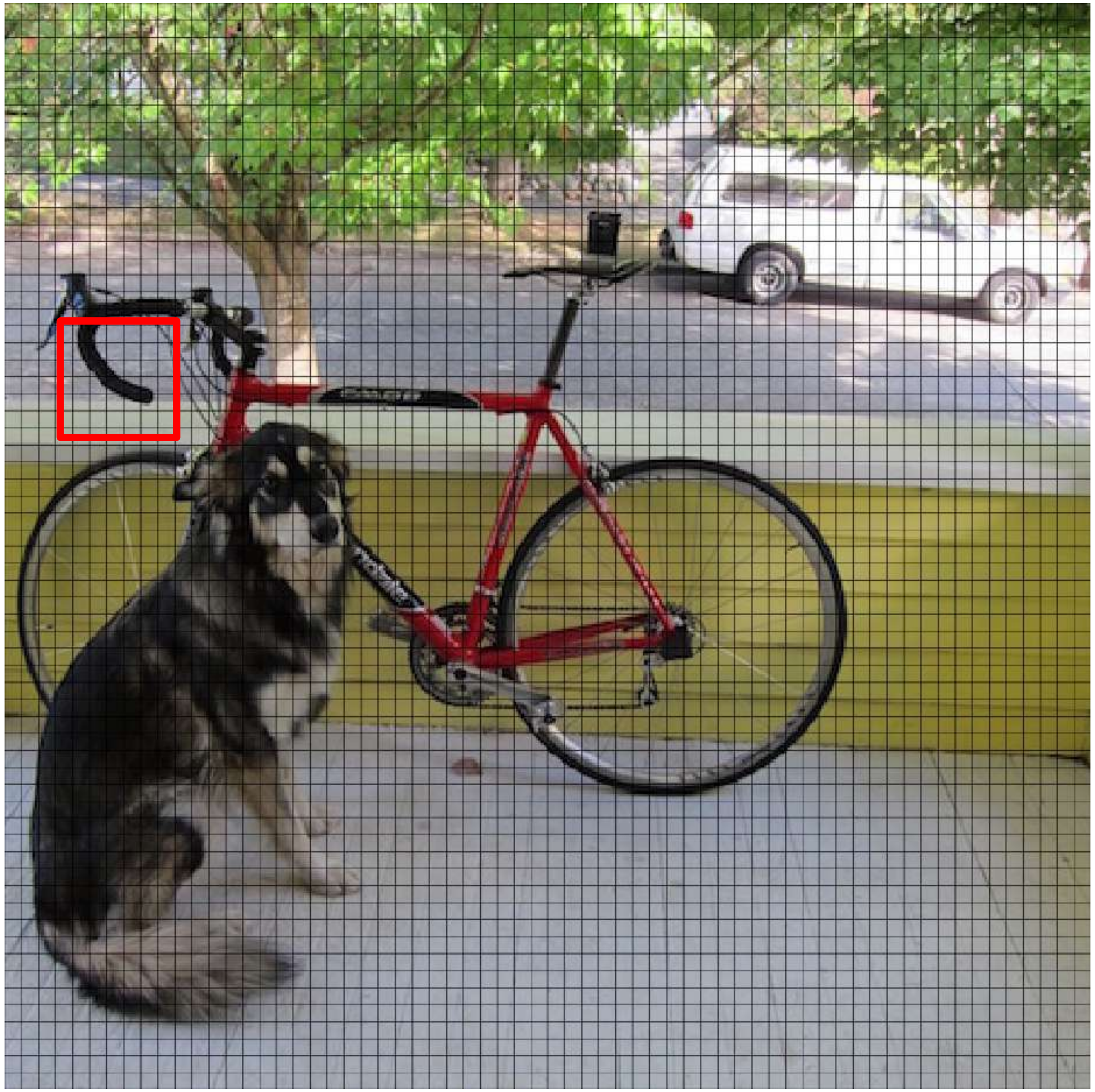


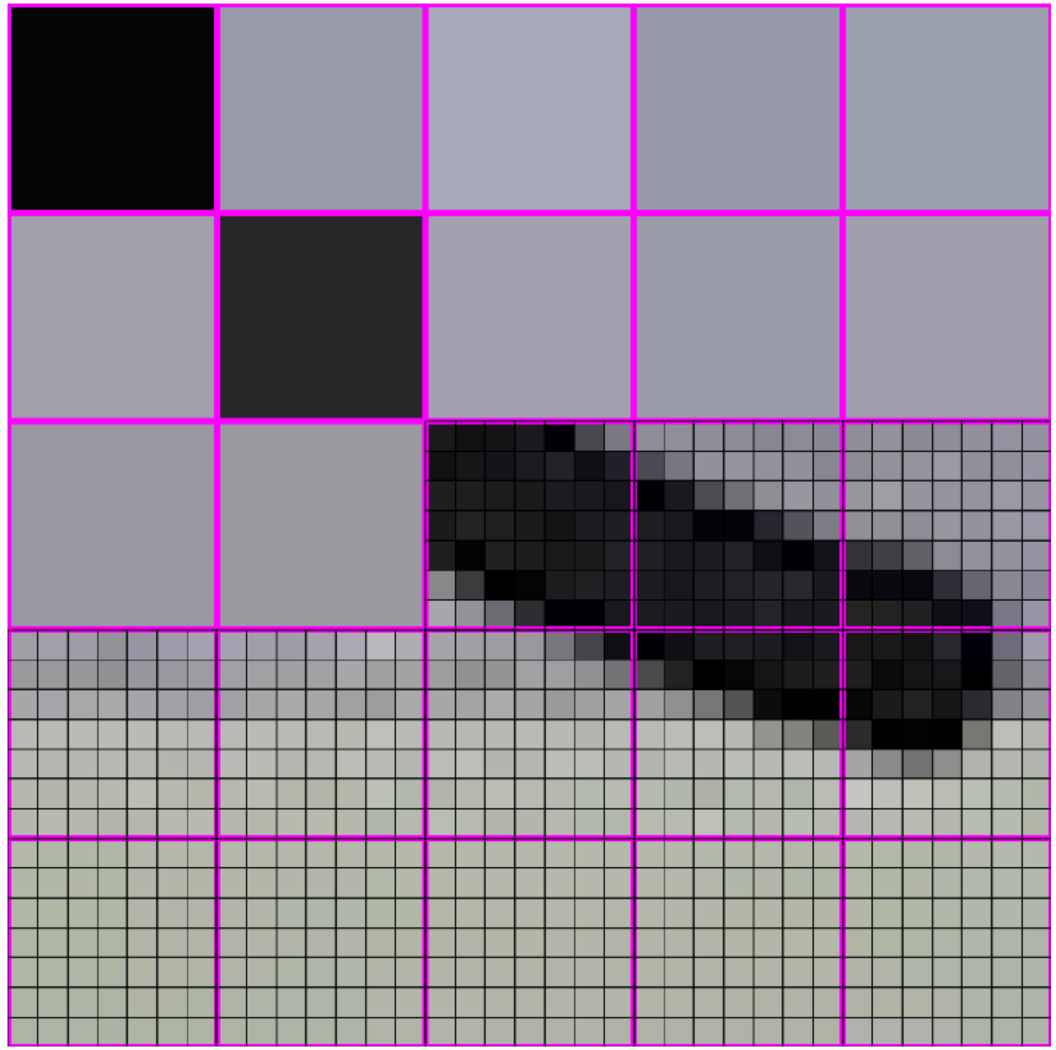
448x448 -> 64x64



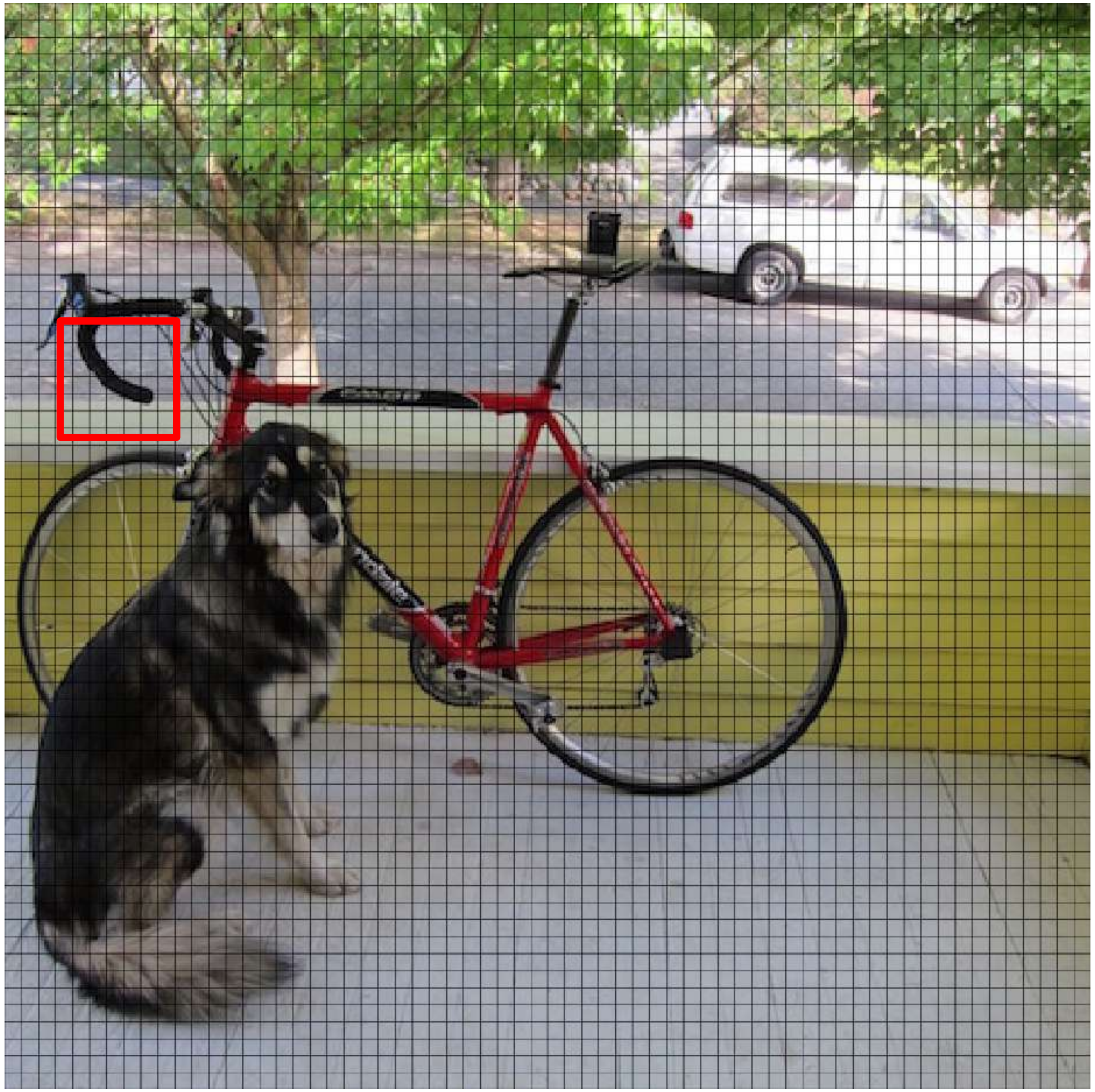


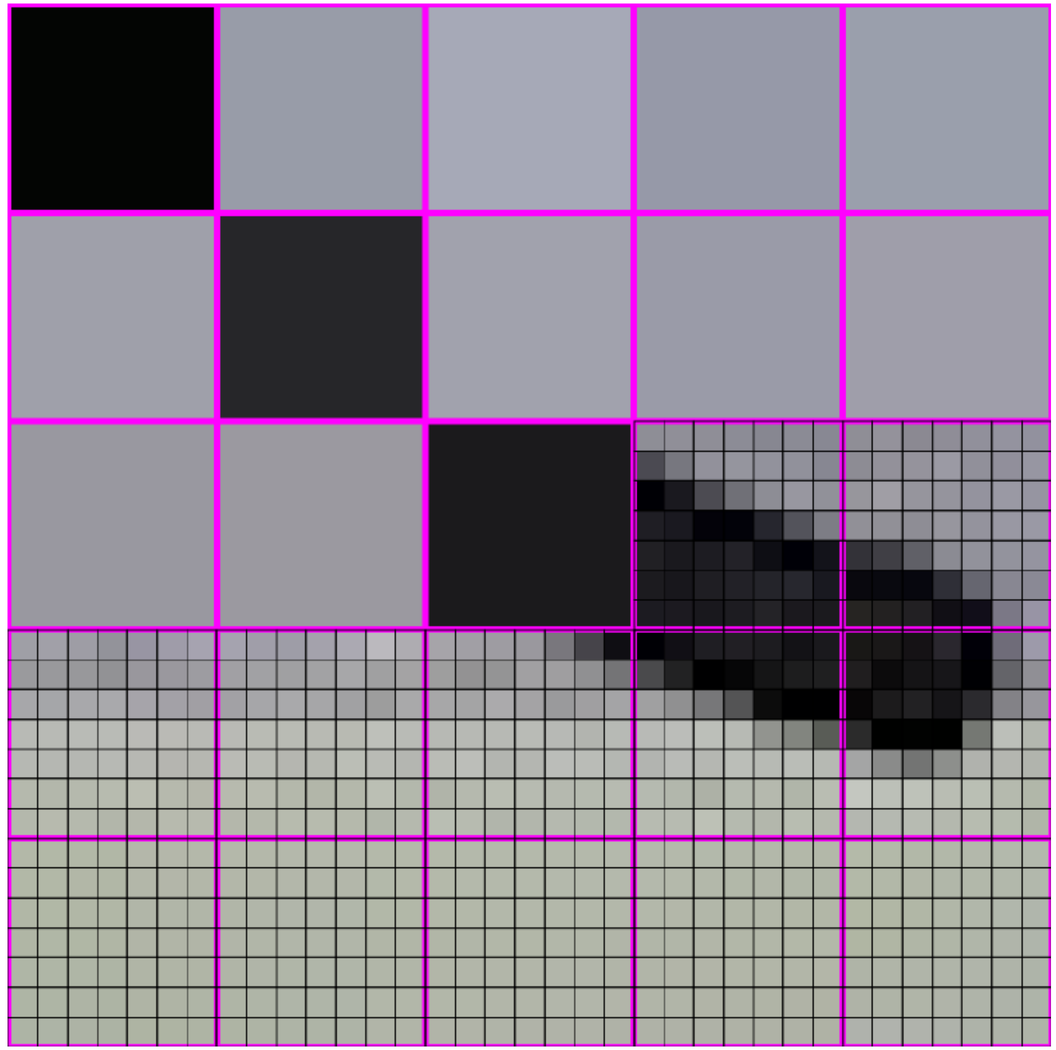
448x448 -> 64x64





448x448 -> 64x64

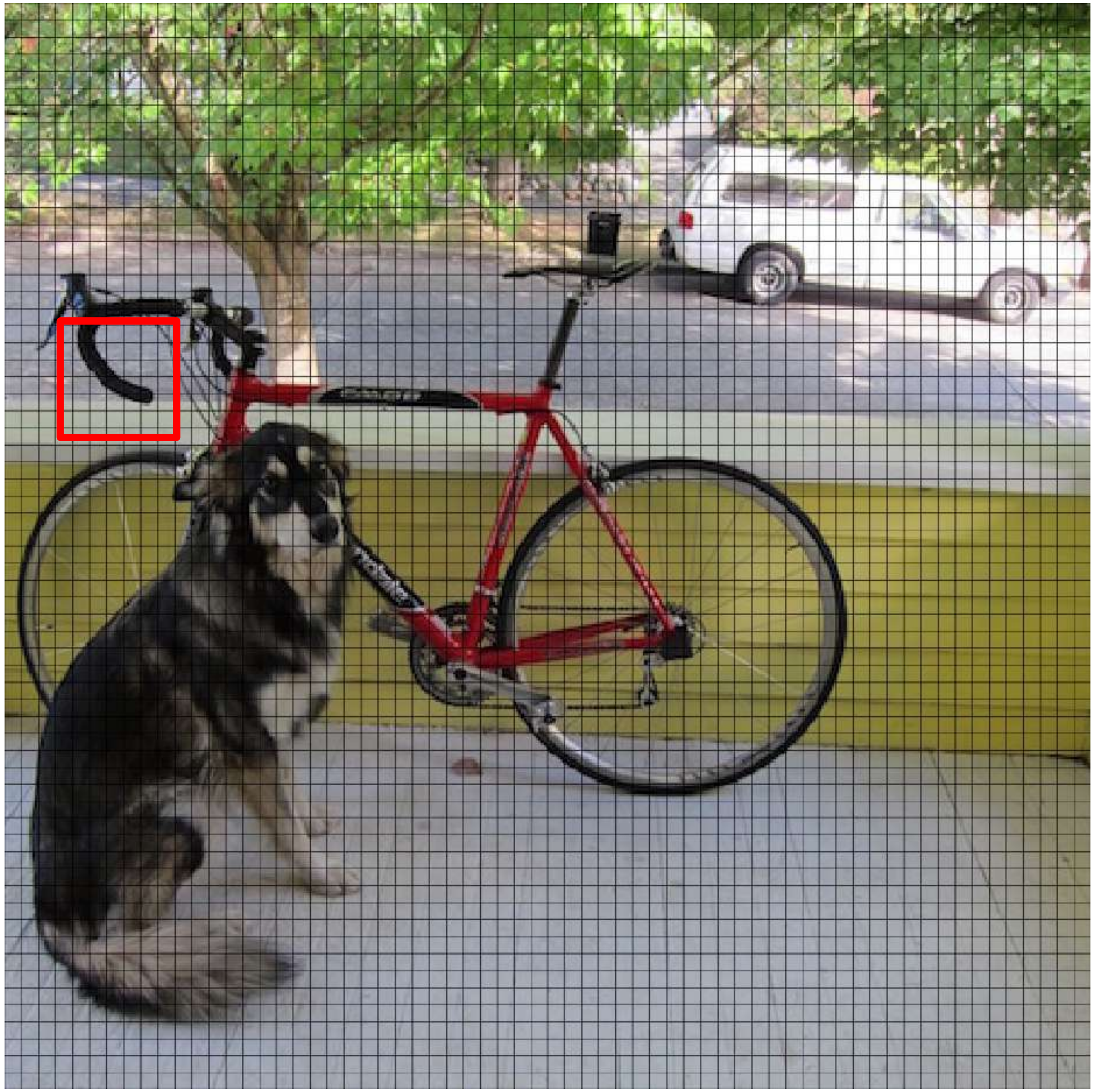


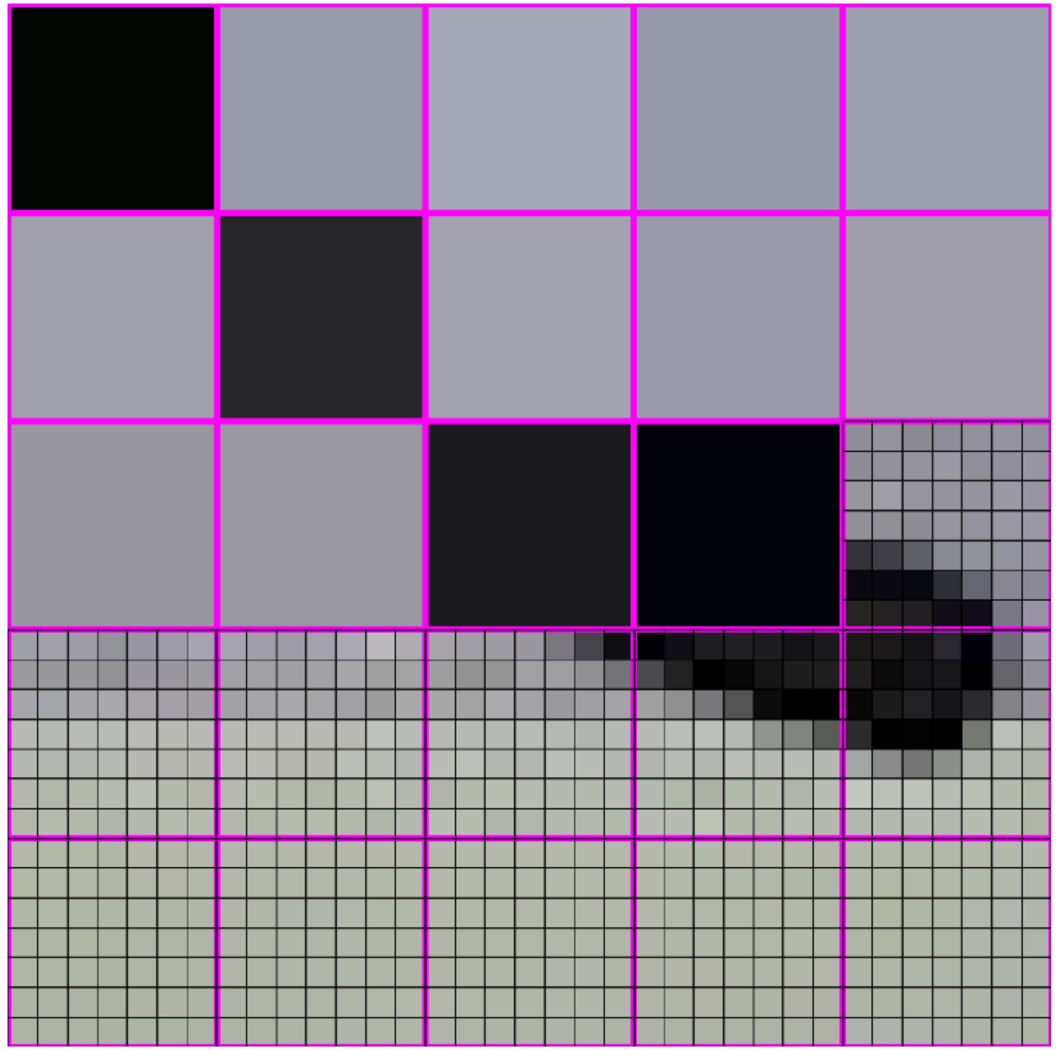


Co-financed by the European Union
Connecting Europe Facility

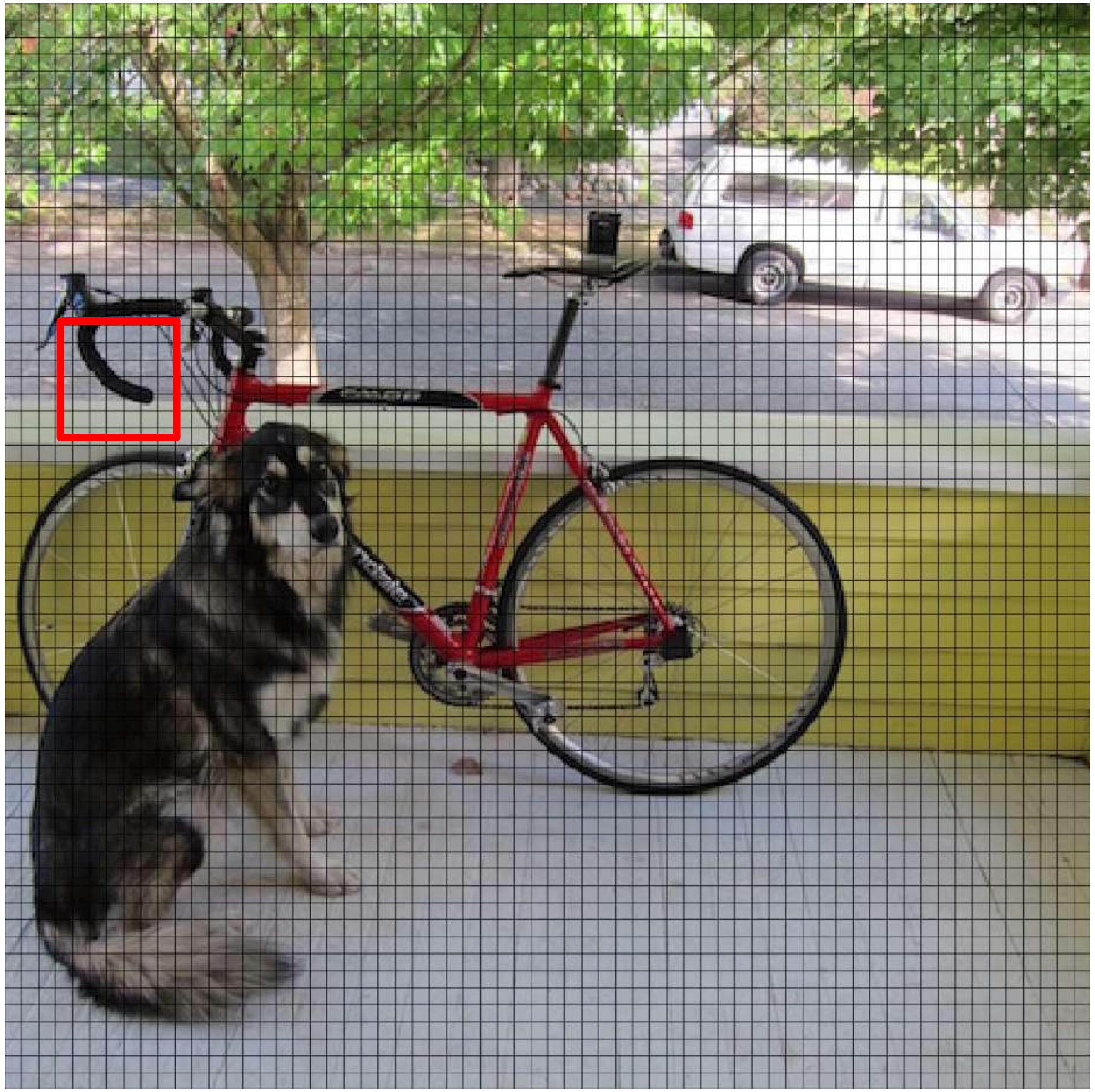
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

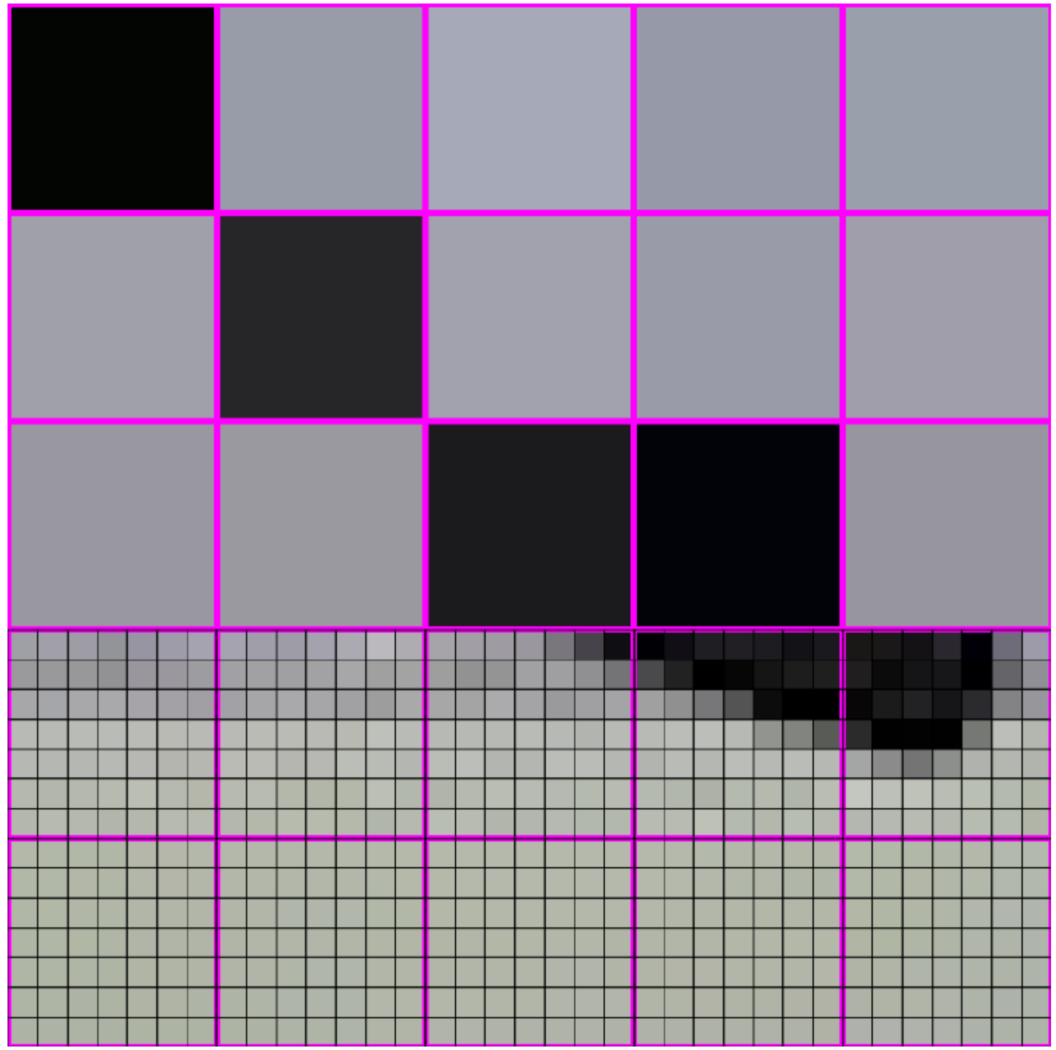
448x448 -> 64x64



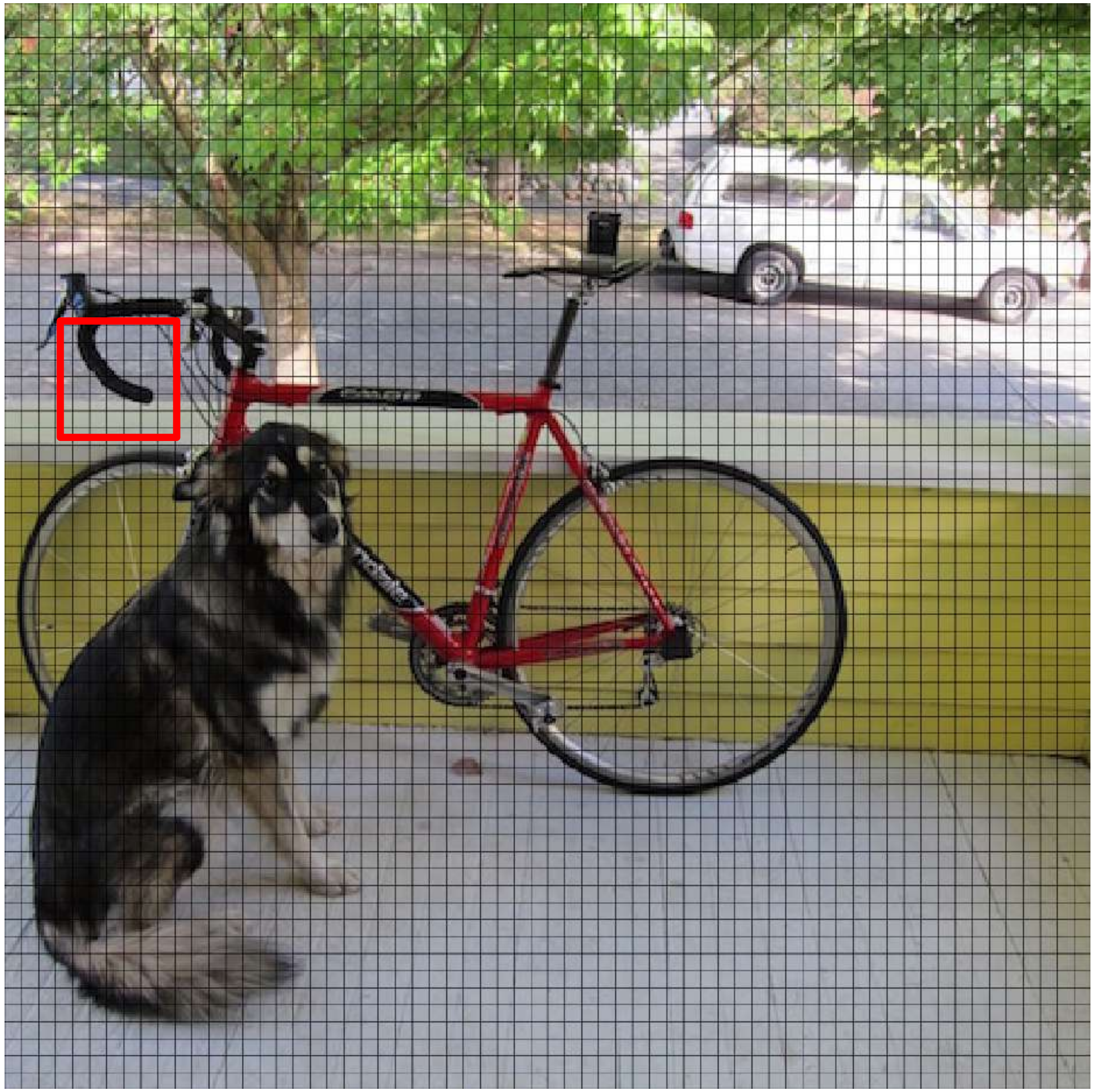


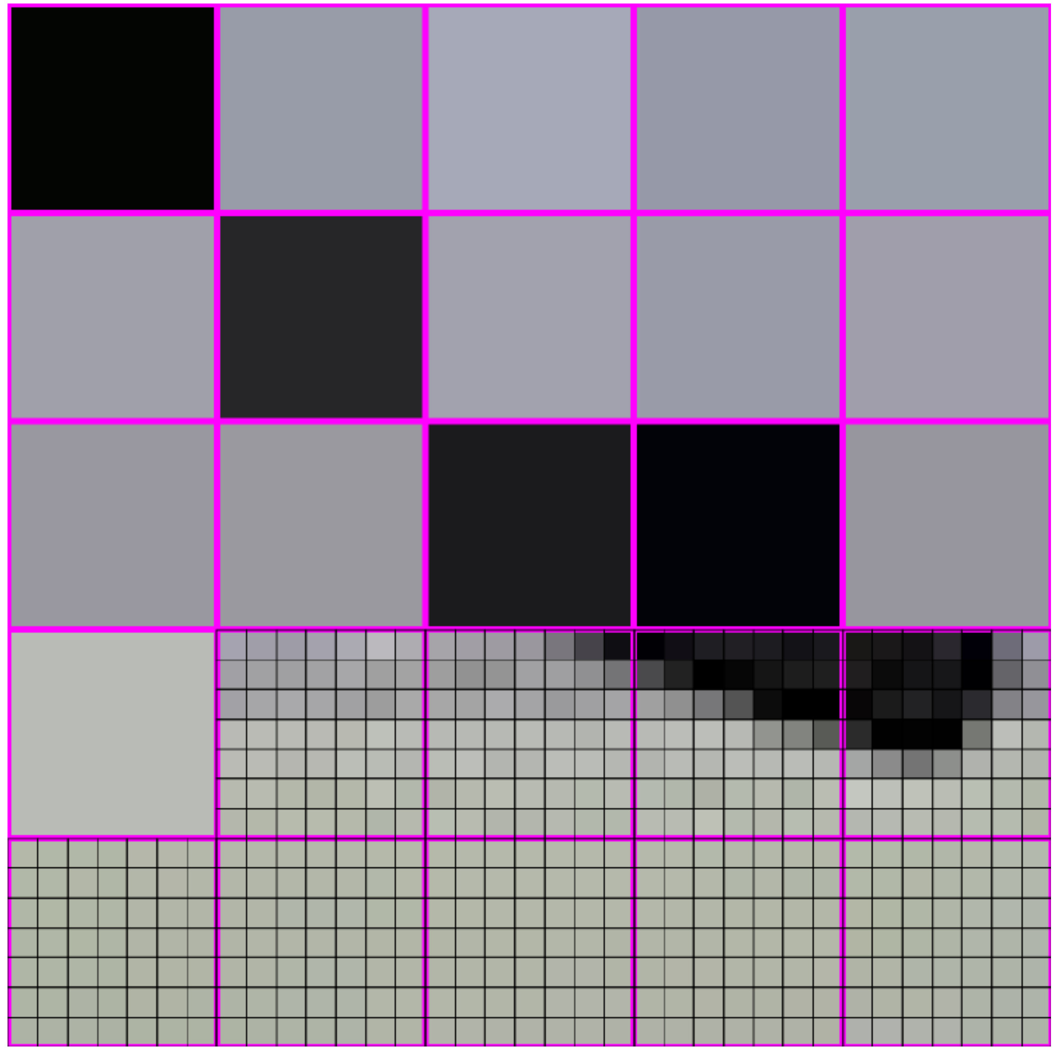
448x448 -> 64x64



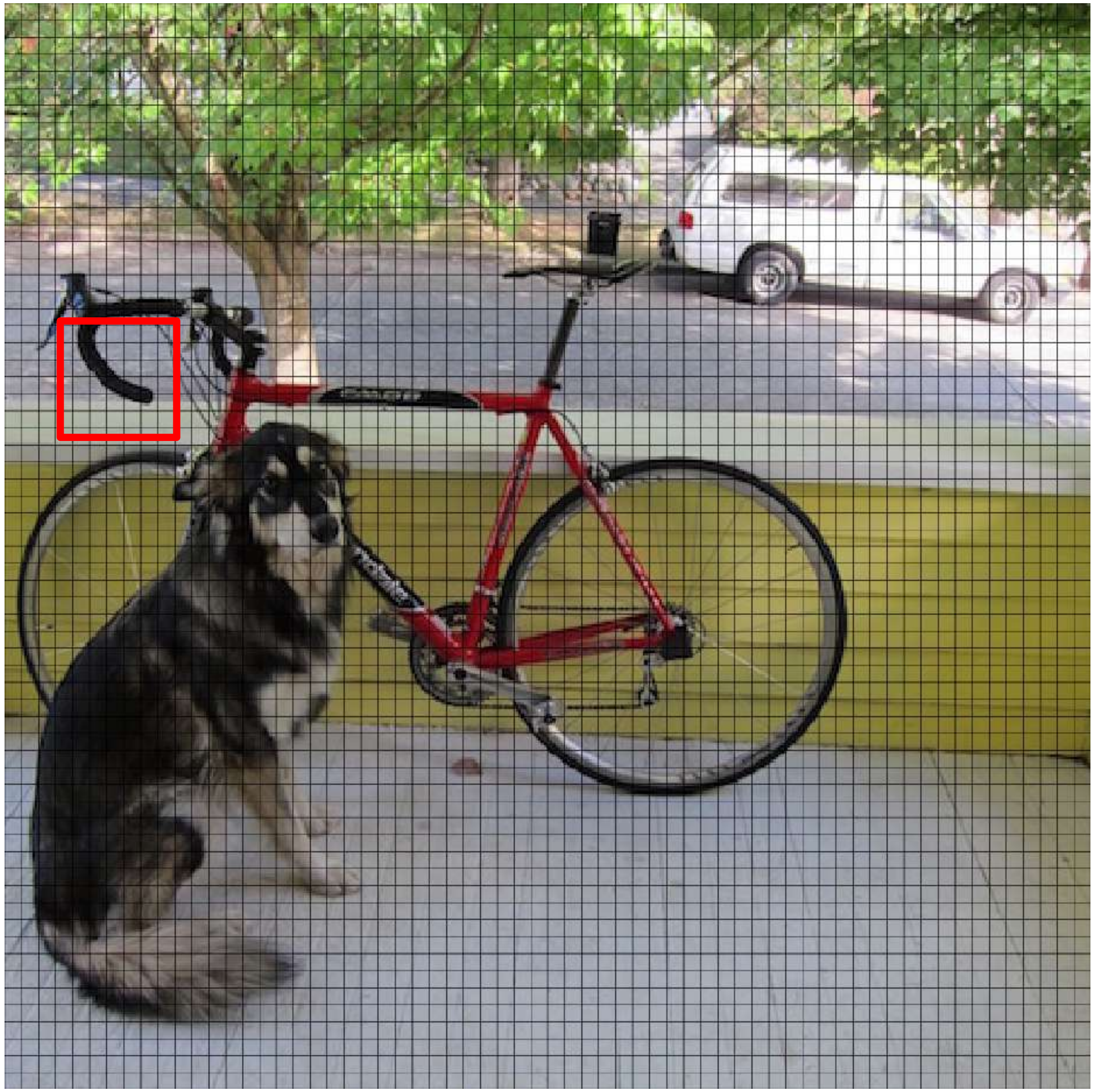


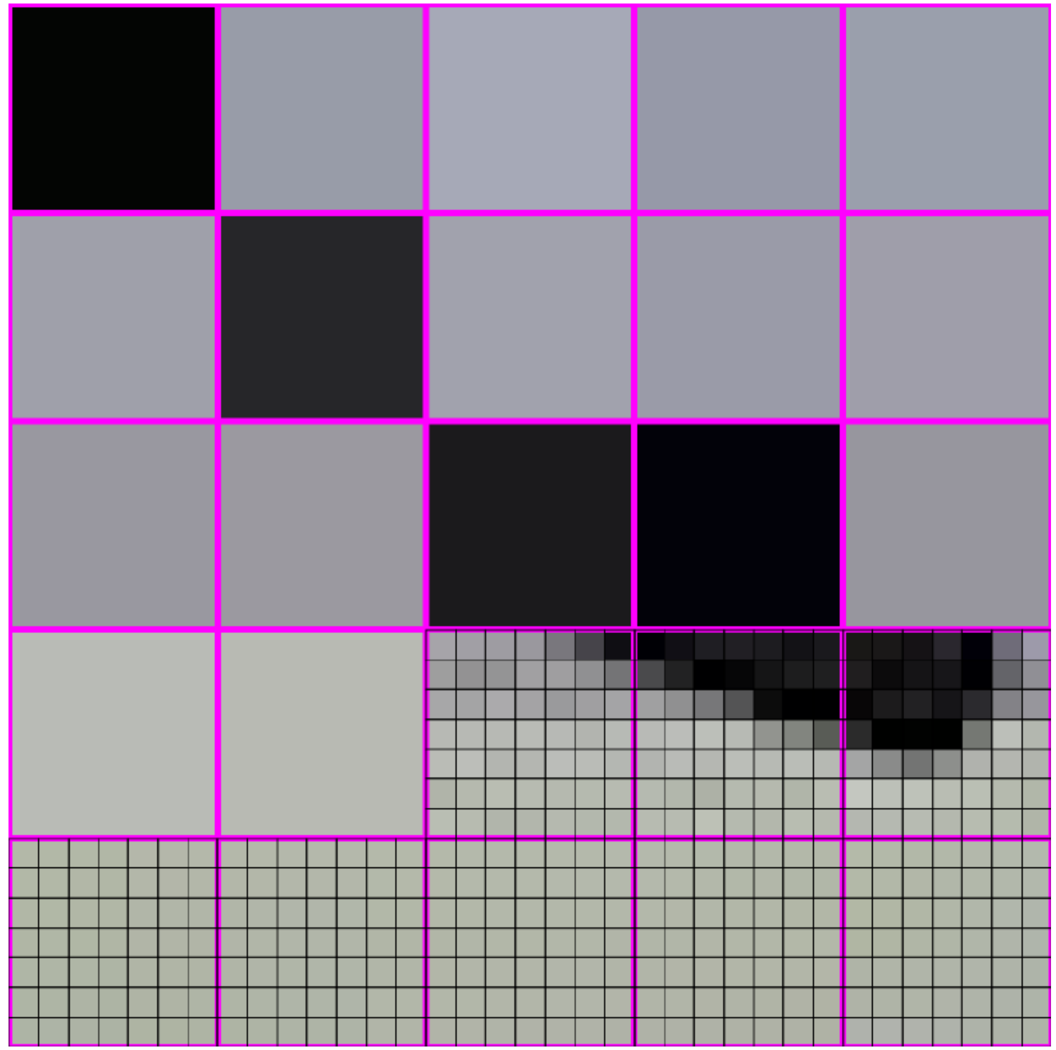
448x448 -> 64x64





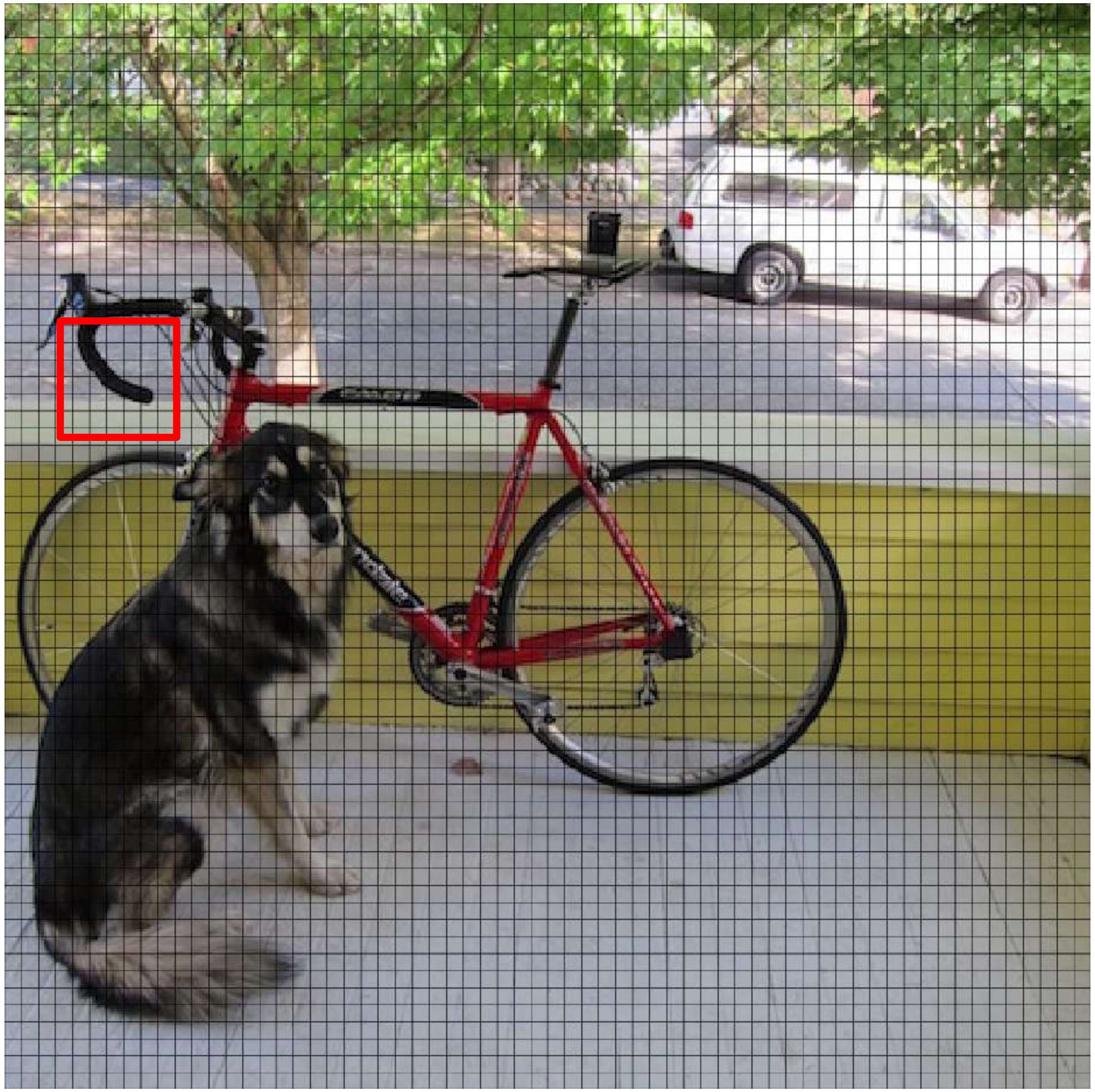
448x448 -> 64x64

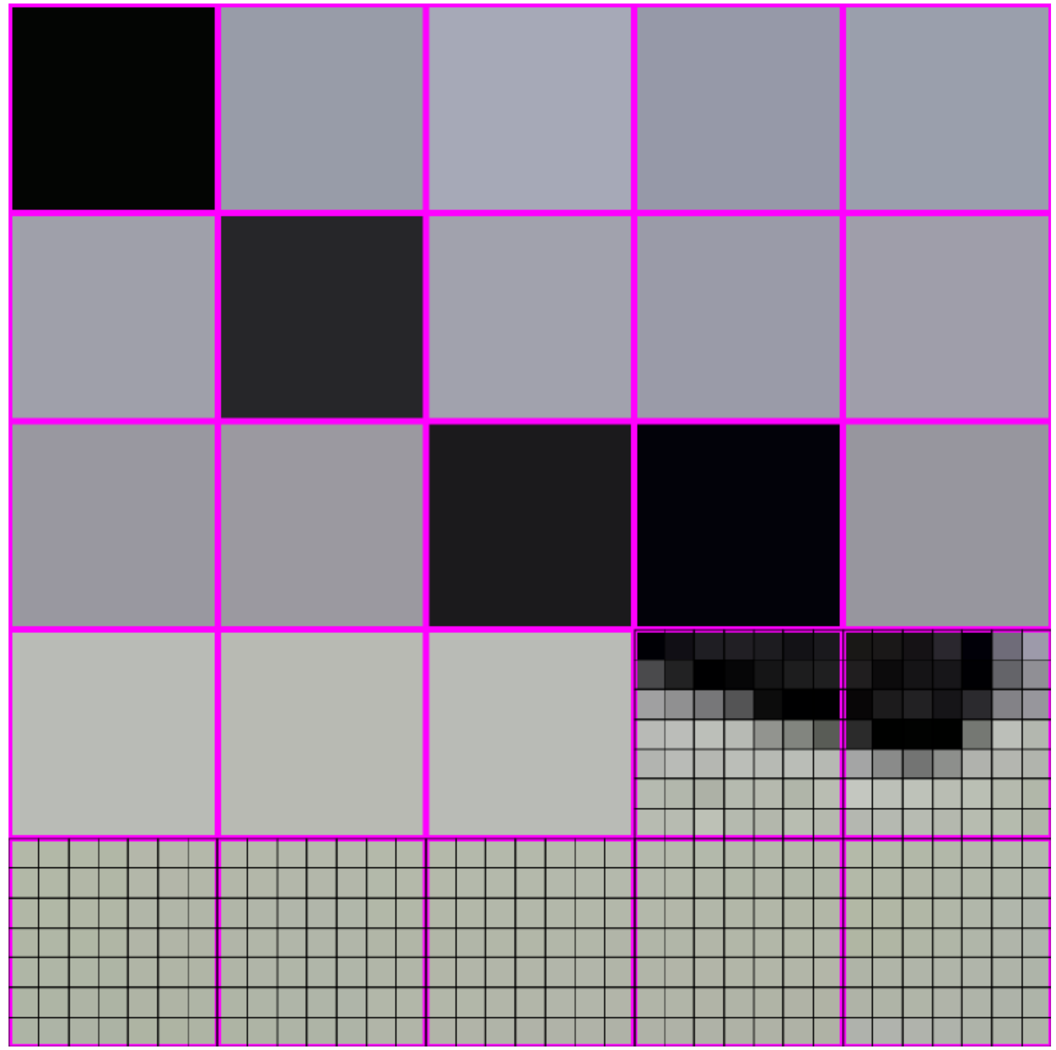




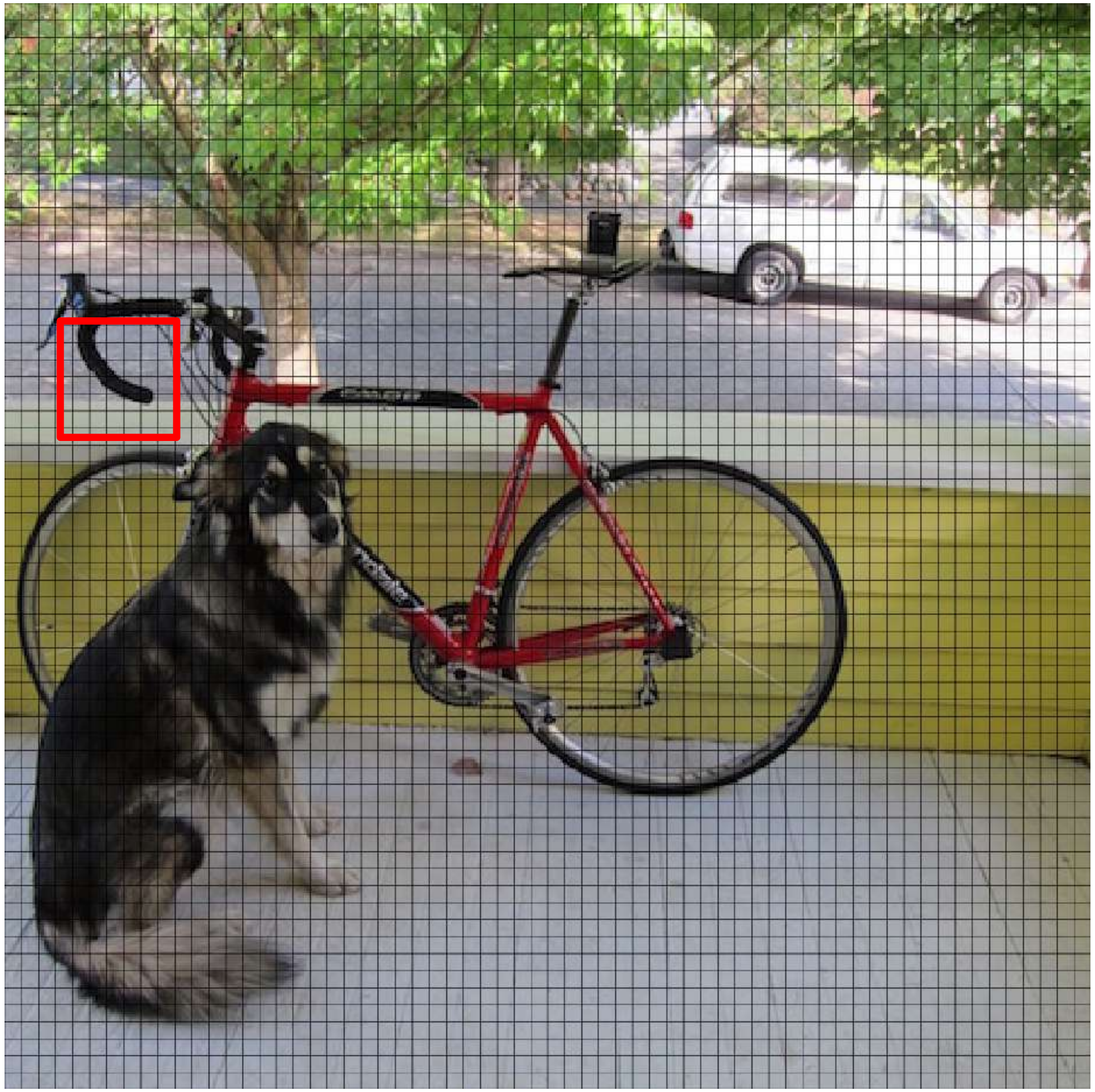
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

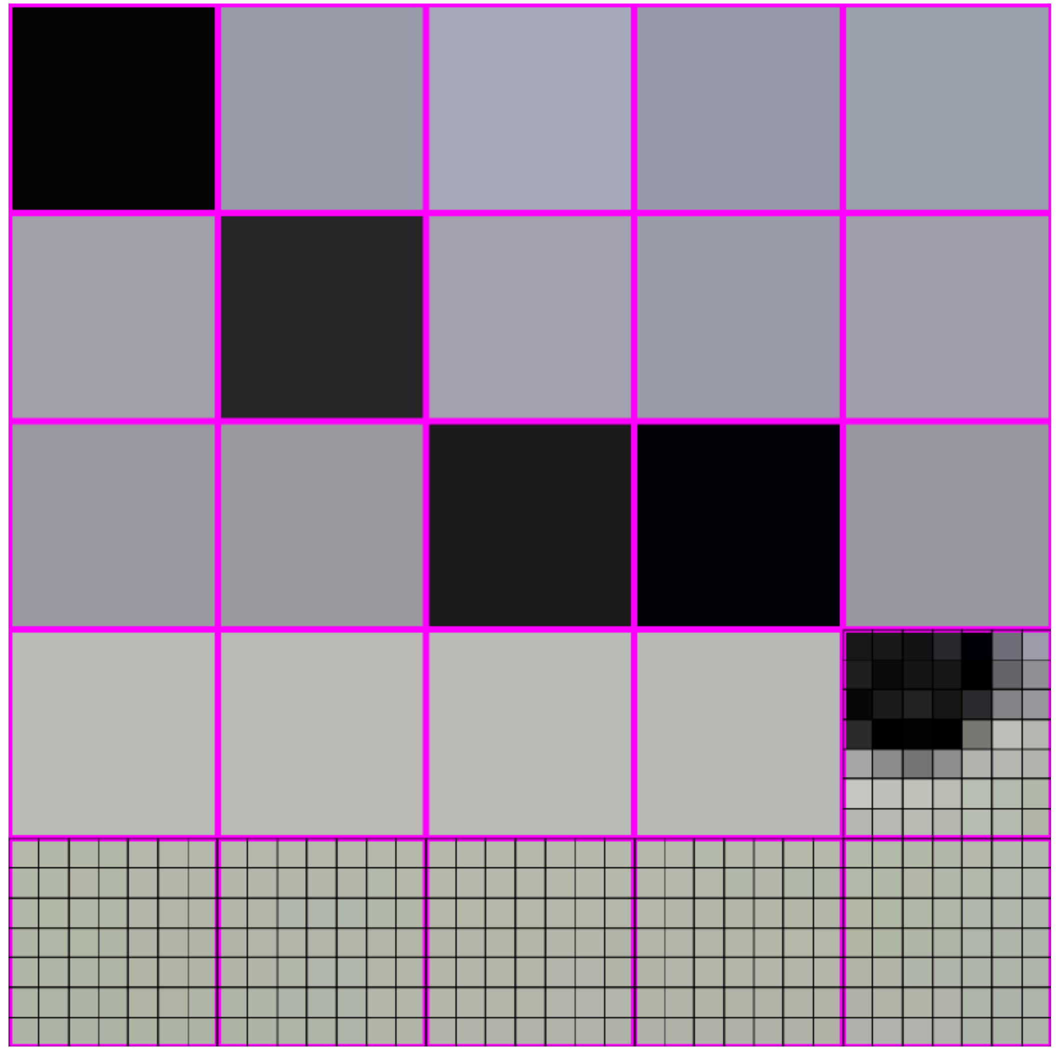
448x448 -> 64x64





448x448 -> 64x64

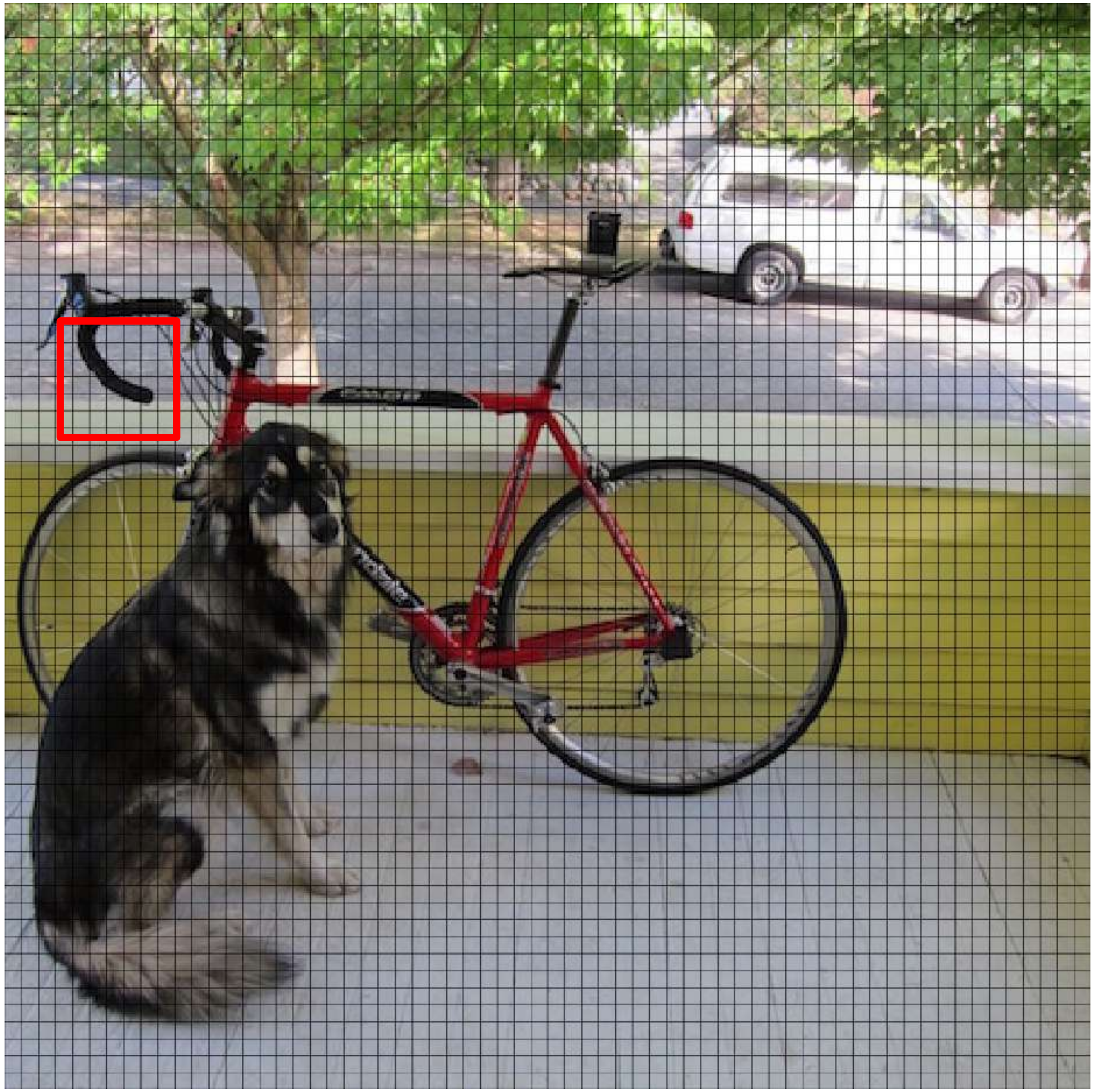


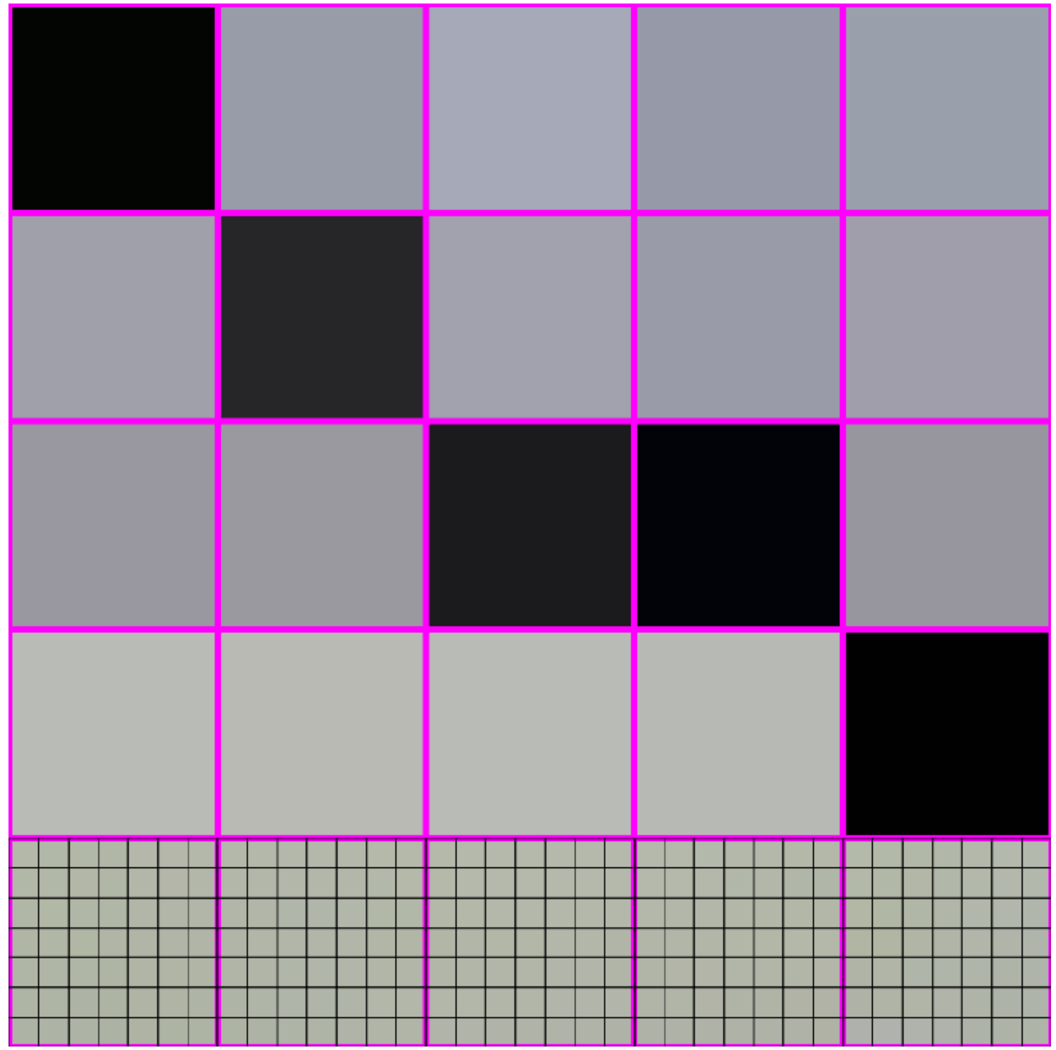


Co-financed by the European Union
Connecting Europe Facility

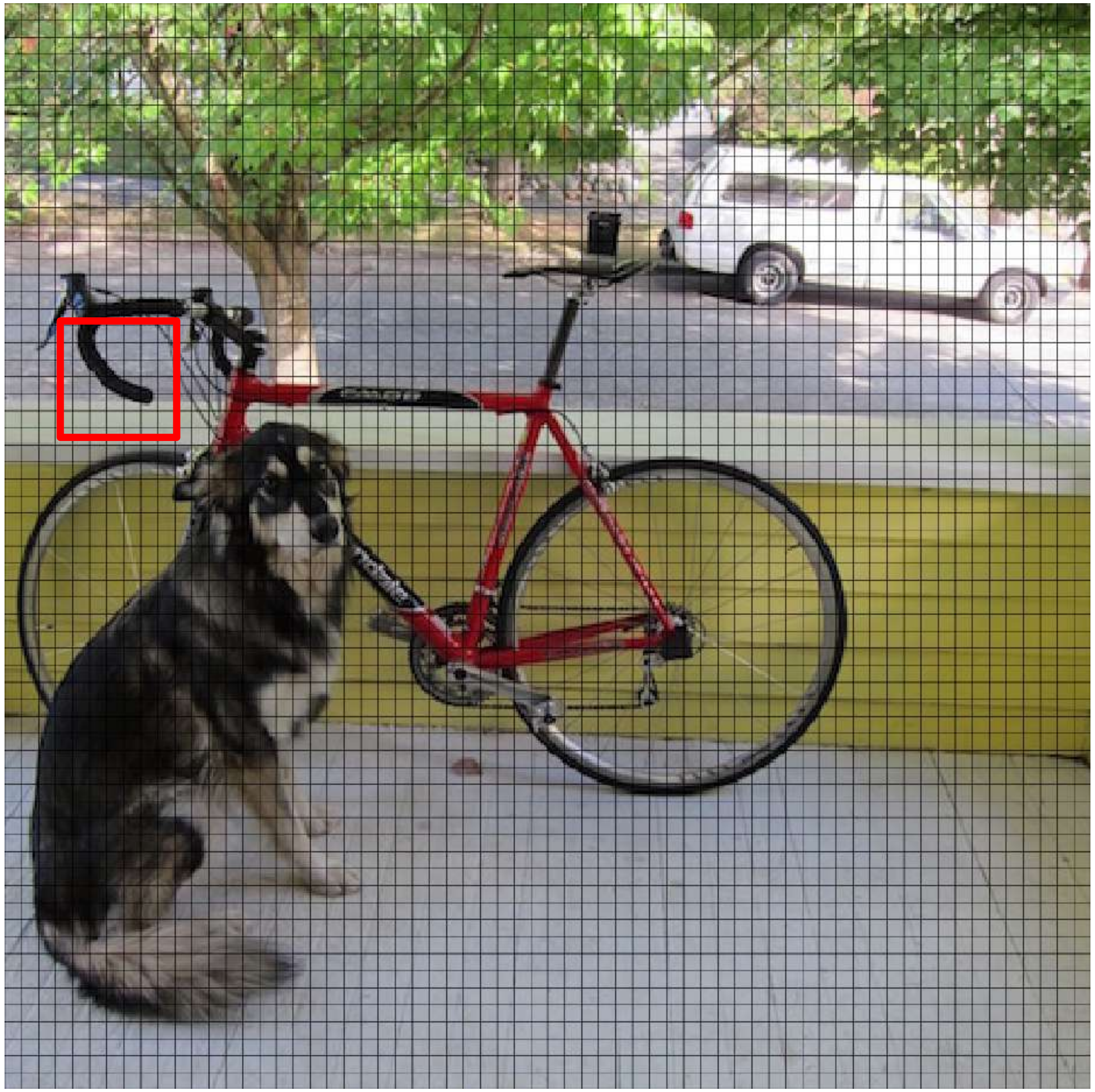
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

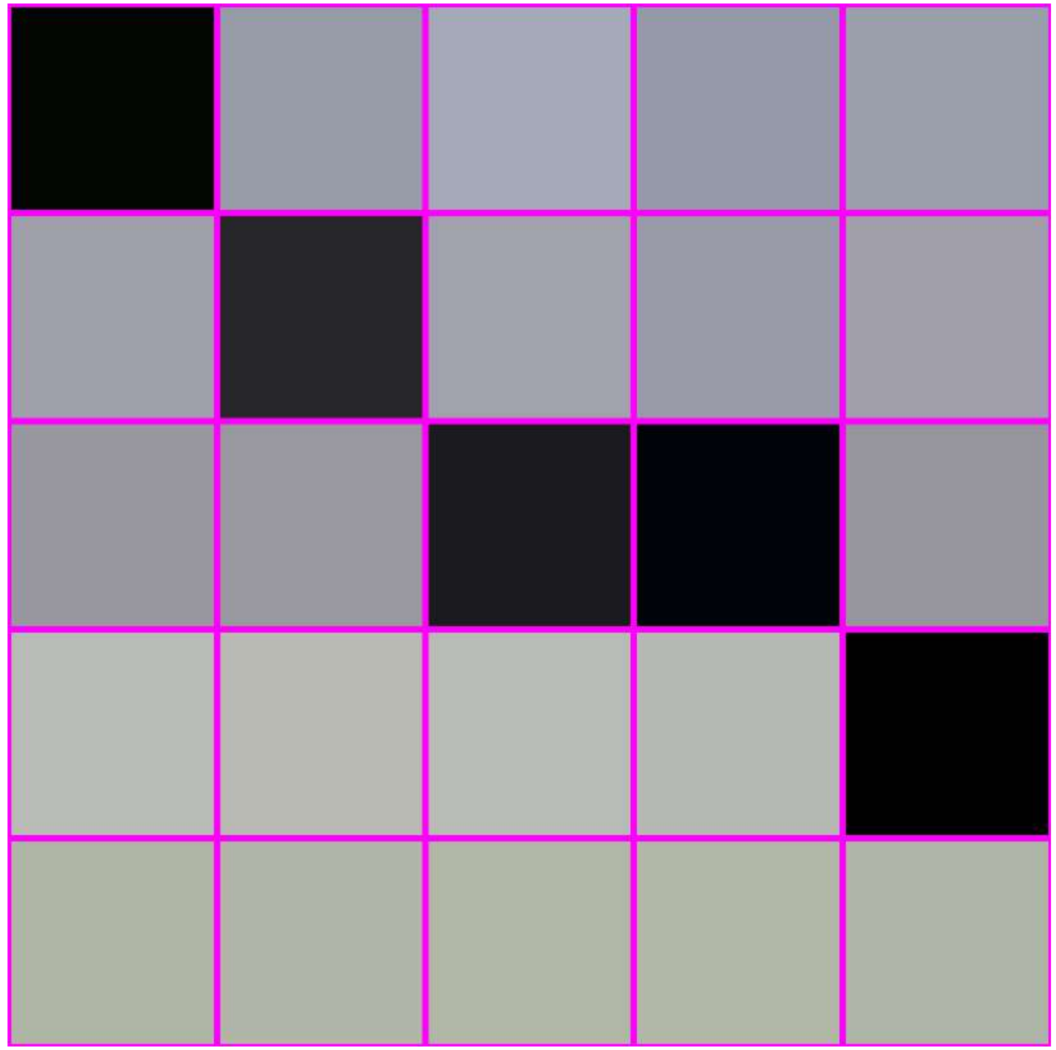
448x448 -> 64x64





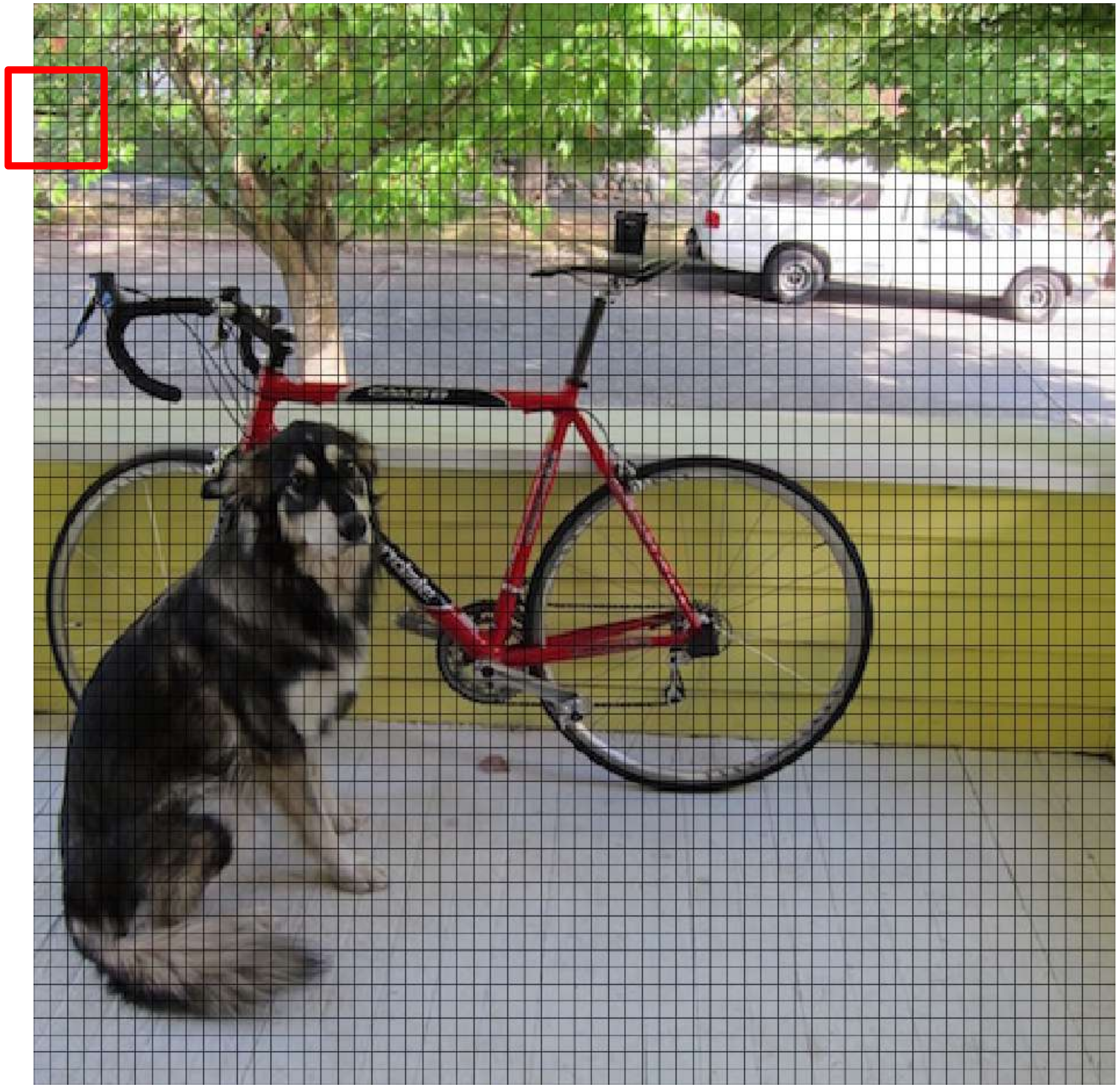
448x448 -> 64x64

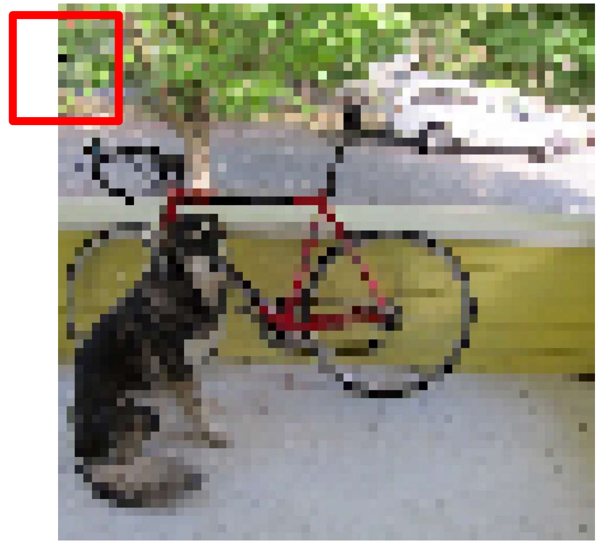




This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

448x448 -> 64x64

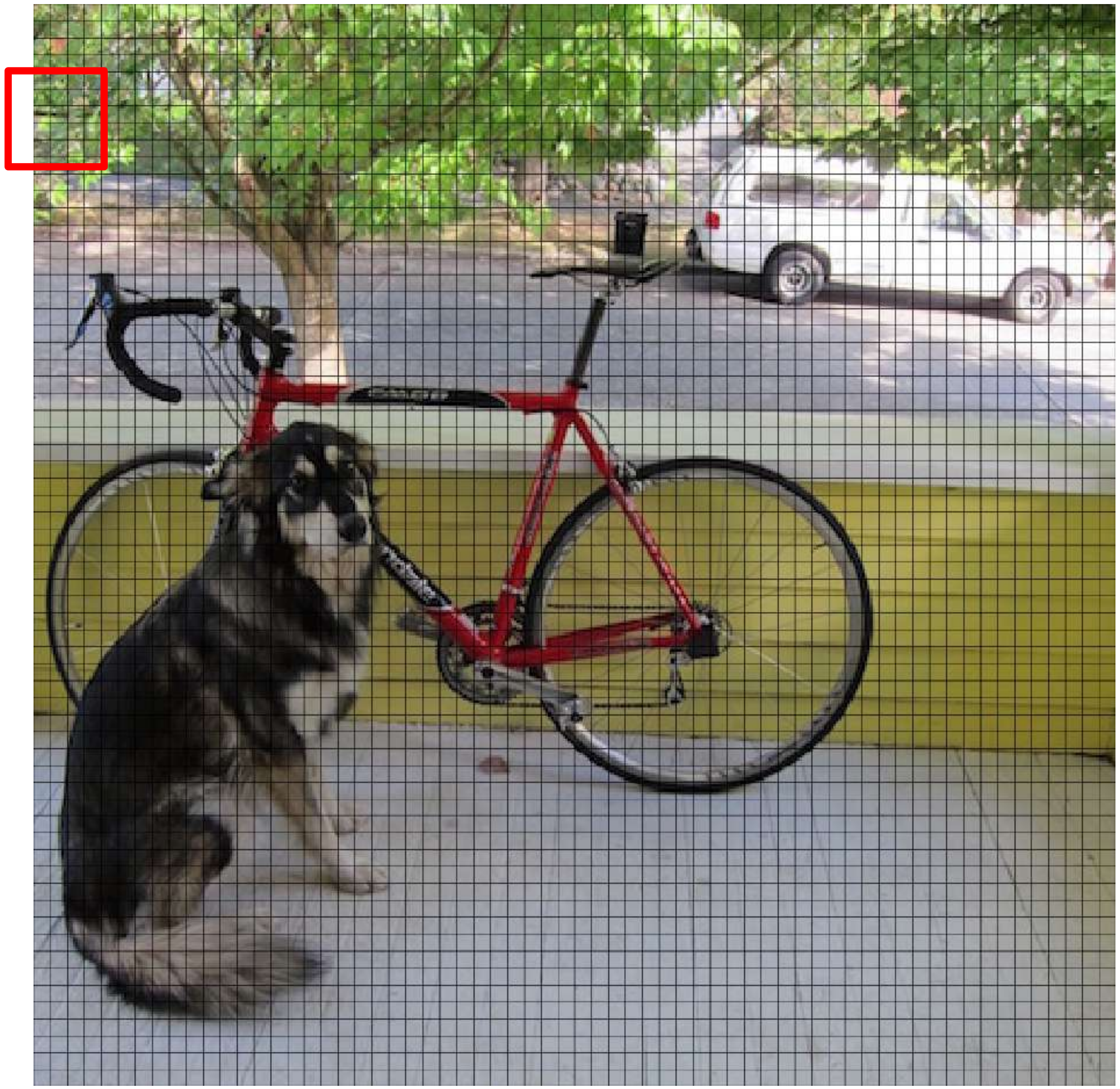


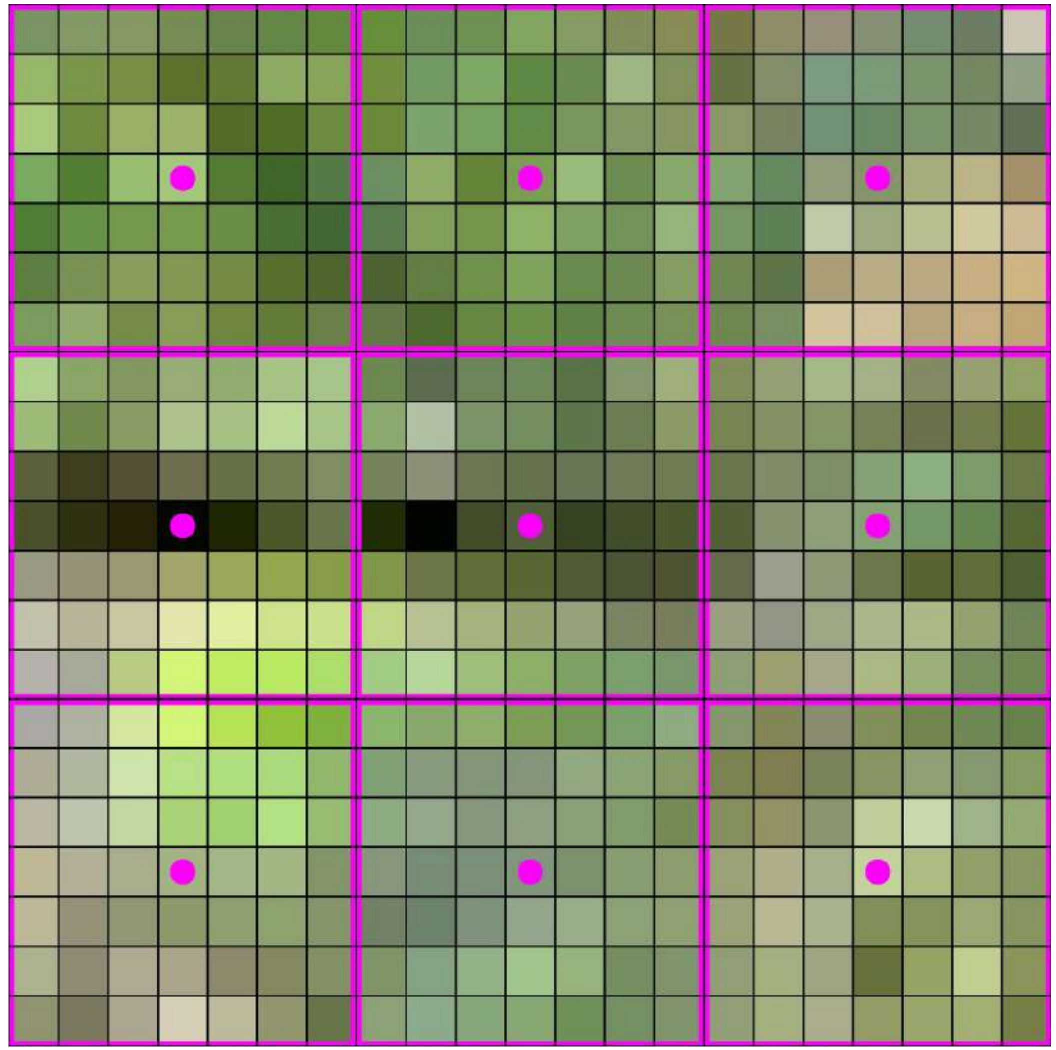


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

448x448 -> 64x64

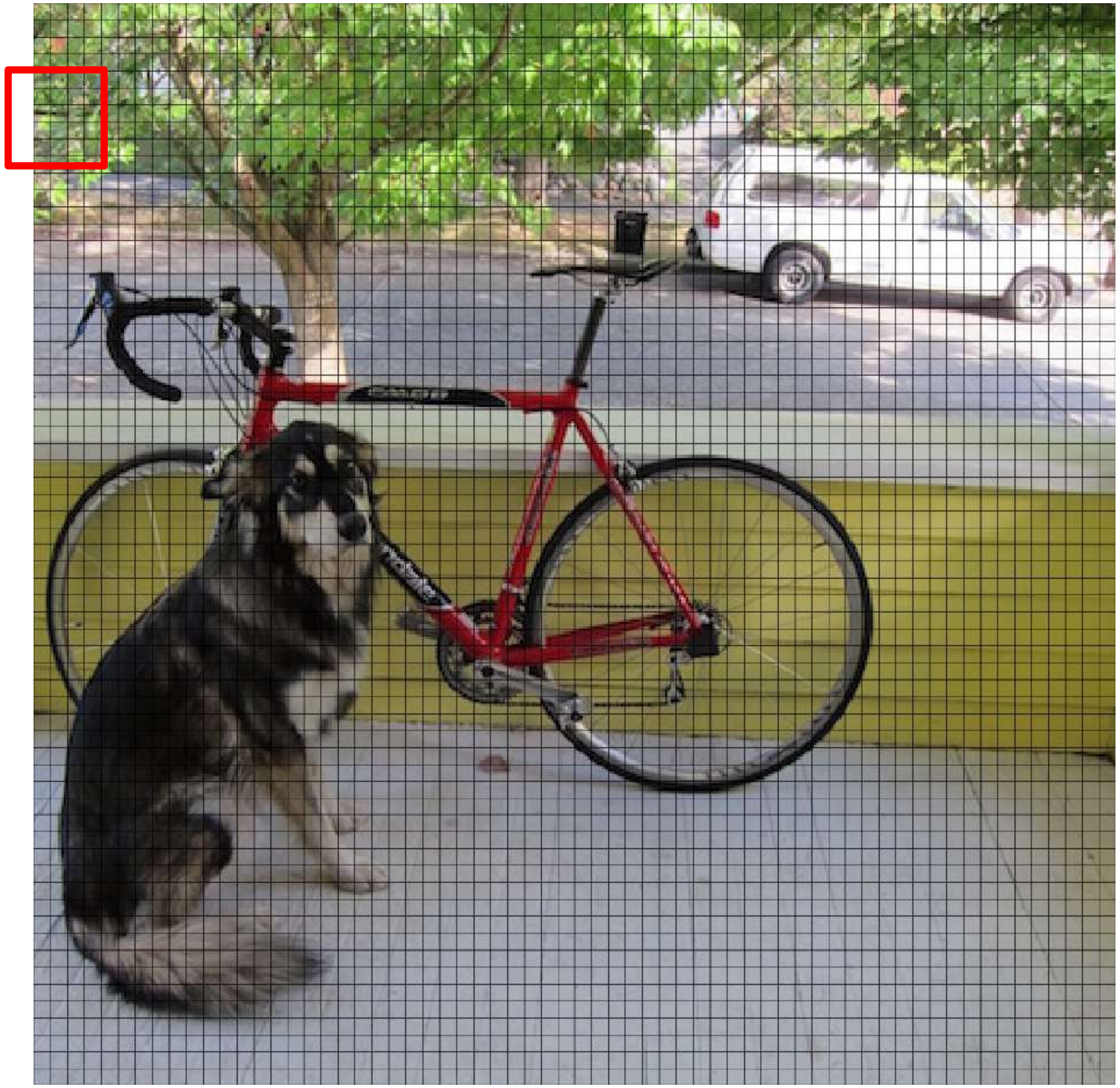


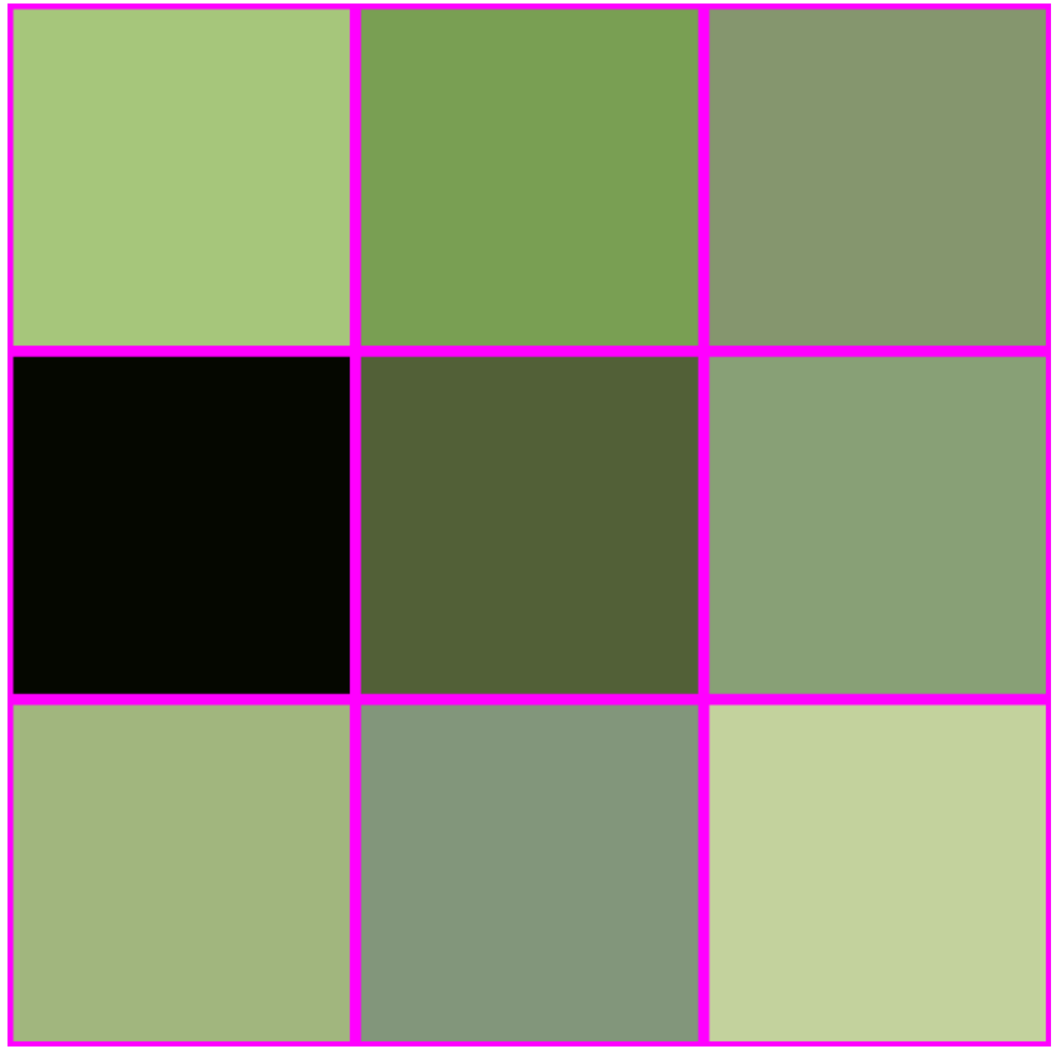


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

448x448 -> 64x64

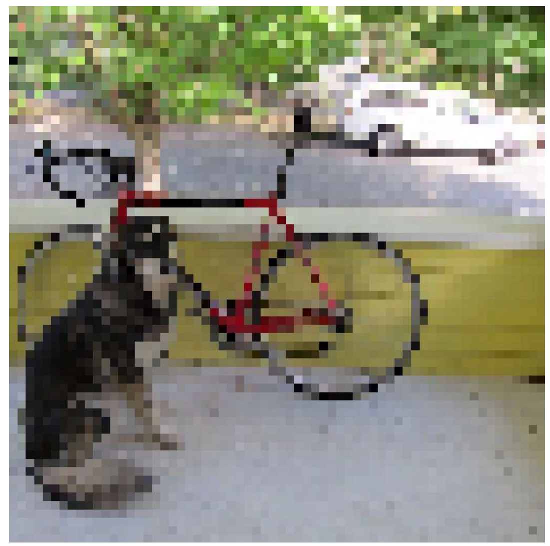




Lots of issues

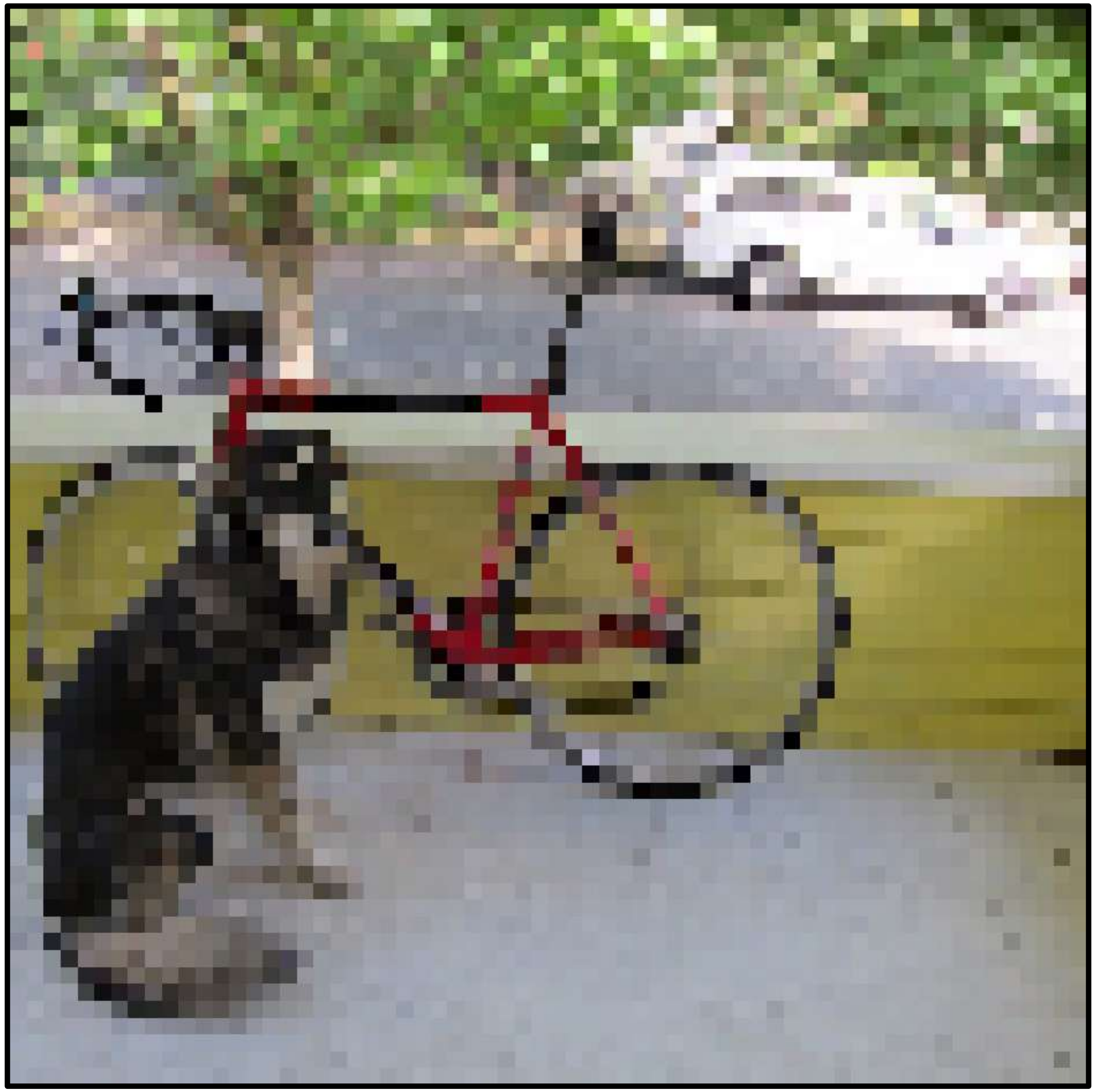
- NN and Bilinear only look at small area
- Lots of artifacting
- Staircase pattern on diagonal lines
- We'll fix this with filters!

NN



Bilinear

IS THIS ALL THERE IS??



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

THERE IS A BETTER WAY!

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 5: Filters – Convolution

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union

Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Image basics
  - What is an image – addressing pixels
  - Image as a function – image coordinates
- Image interpolation
  - Nearest neighbor
  - Bilinear
  - Bicubic
- Image resizing
  - Enlarge
  - Shrink

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Averaging vs Interpolation
- Systems - filters
- Convolution
  - Box Filter
  - Gaussian
  - Cross correlation vs Convolution
- Examples of filters

[material based on Joseph Redmon's course]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

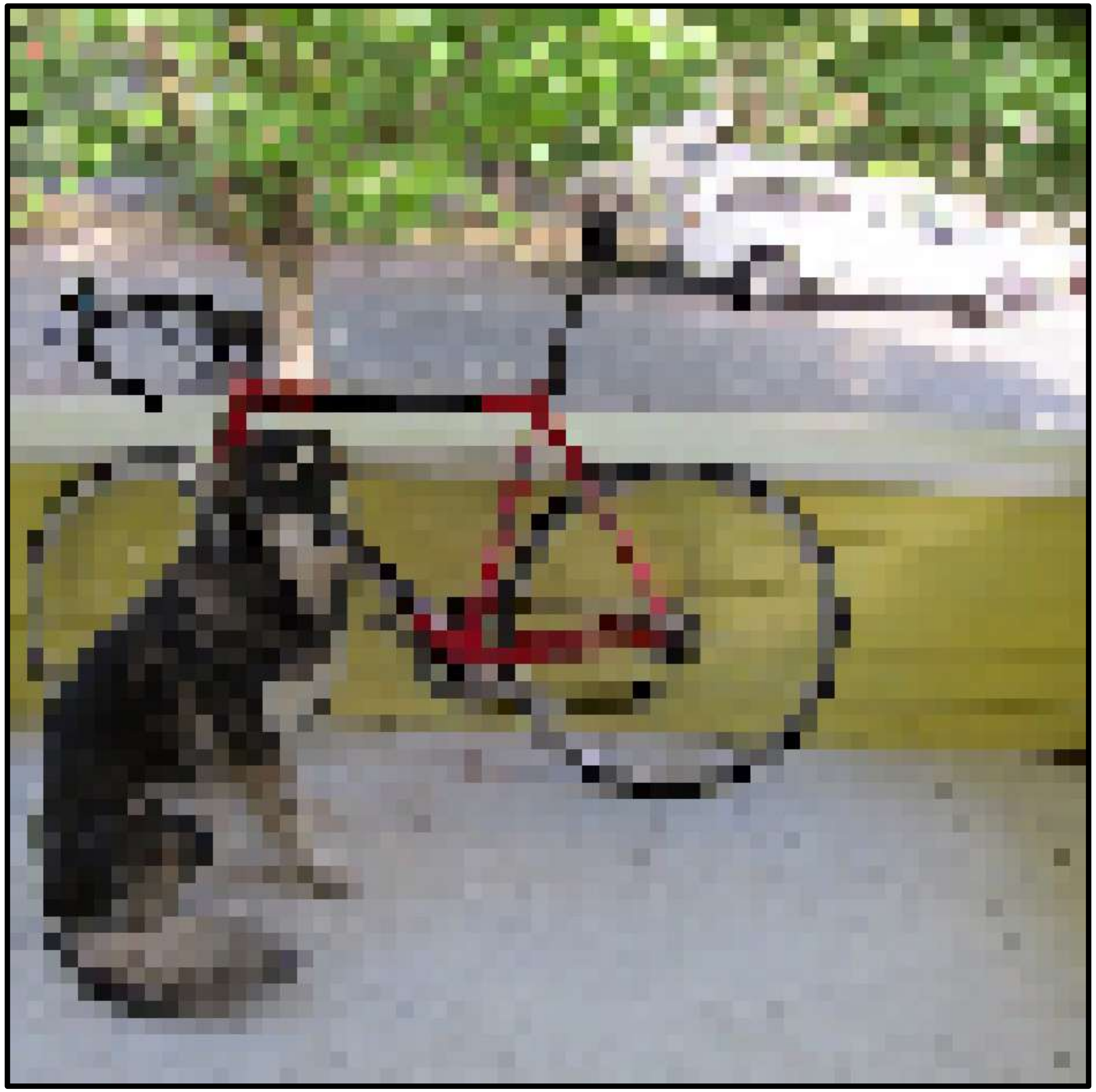
Today's Agenda

- Averaging vs Interpolation
- Systems - filters
- Convolution
    - Box Filter
    - Gaussian
    - Cross correlation vs Convolution
- Examples of filters

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Is this all there is ??



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Lots of issues

- NN and Bilinear only look at small area
- Lots of artifacting
- Staircase pattern on diagonal lines
- We'll fix this with filters!

NN

Bilinear

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

There is a better way!

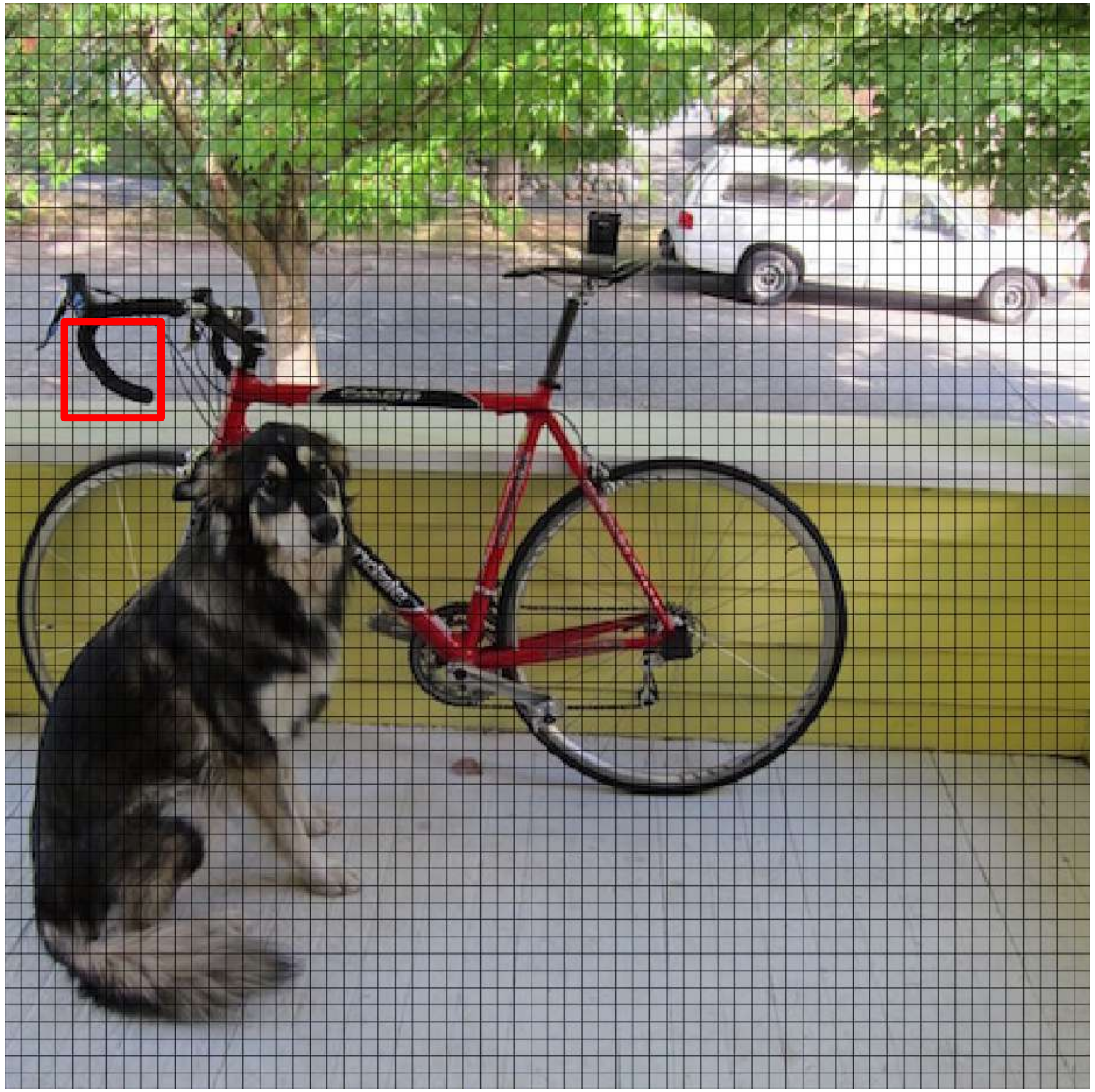
Look at how much better

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

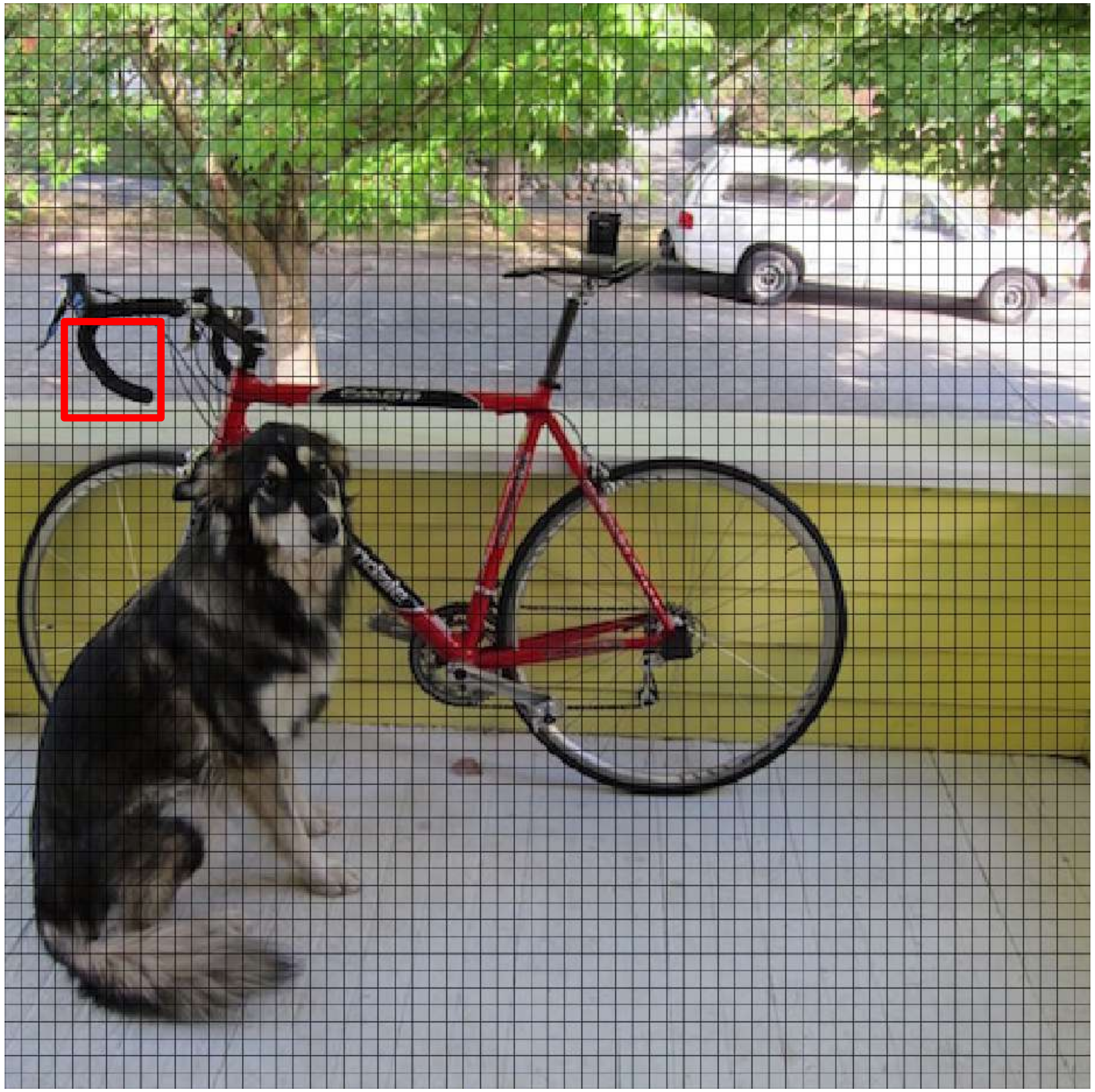
How?

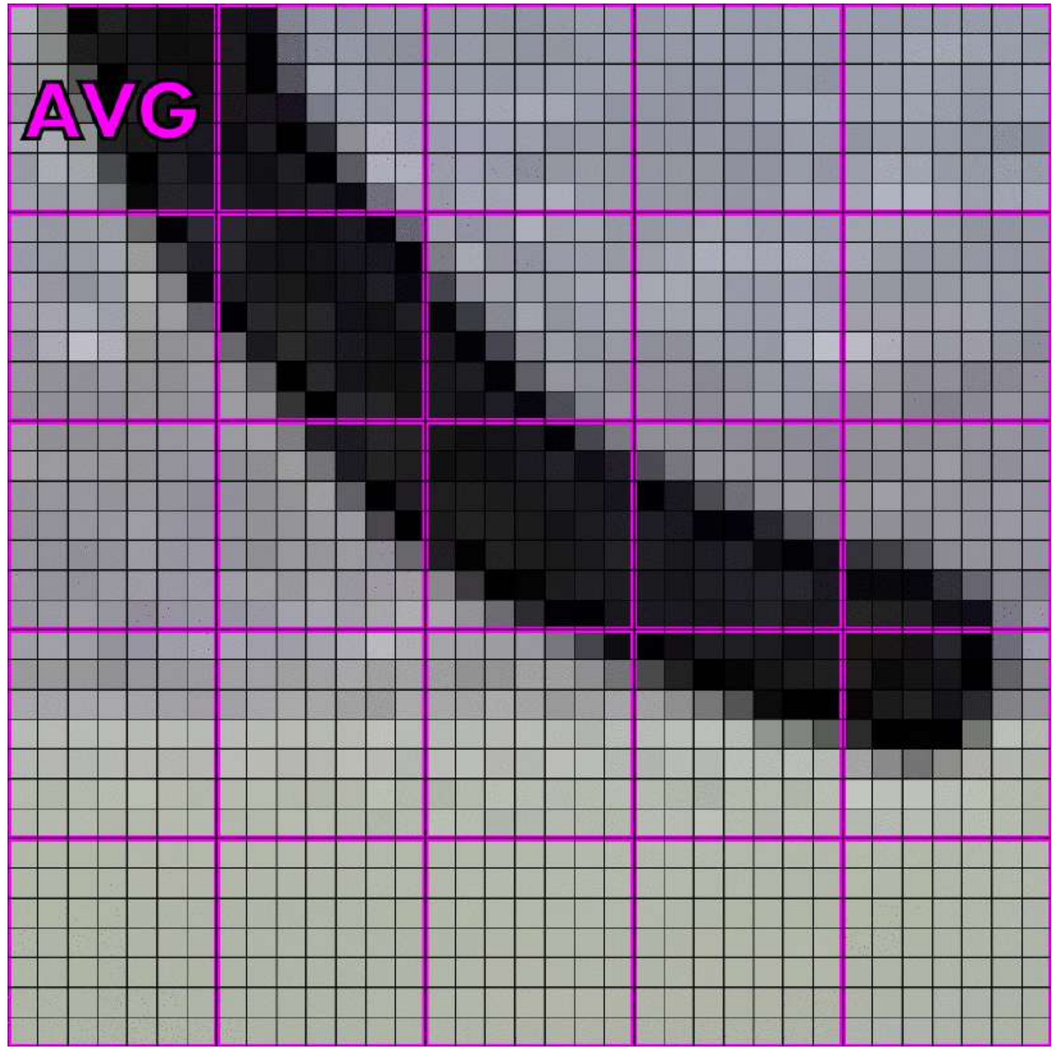


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

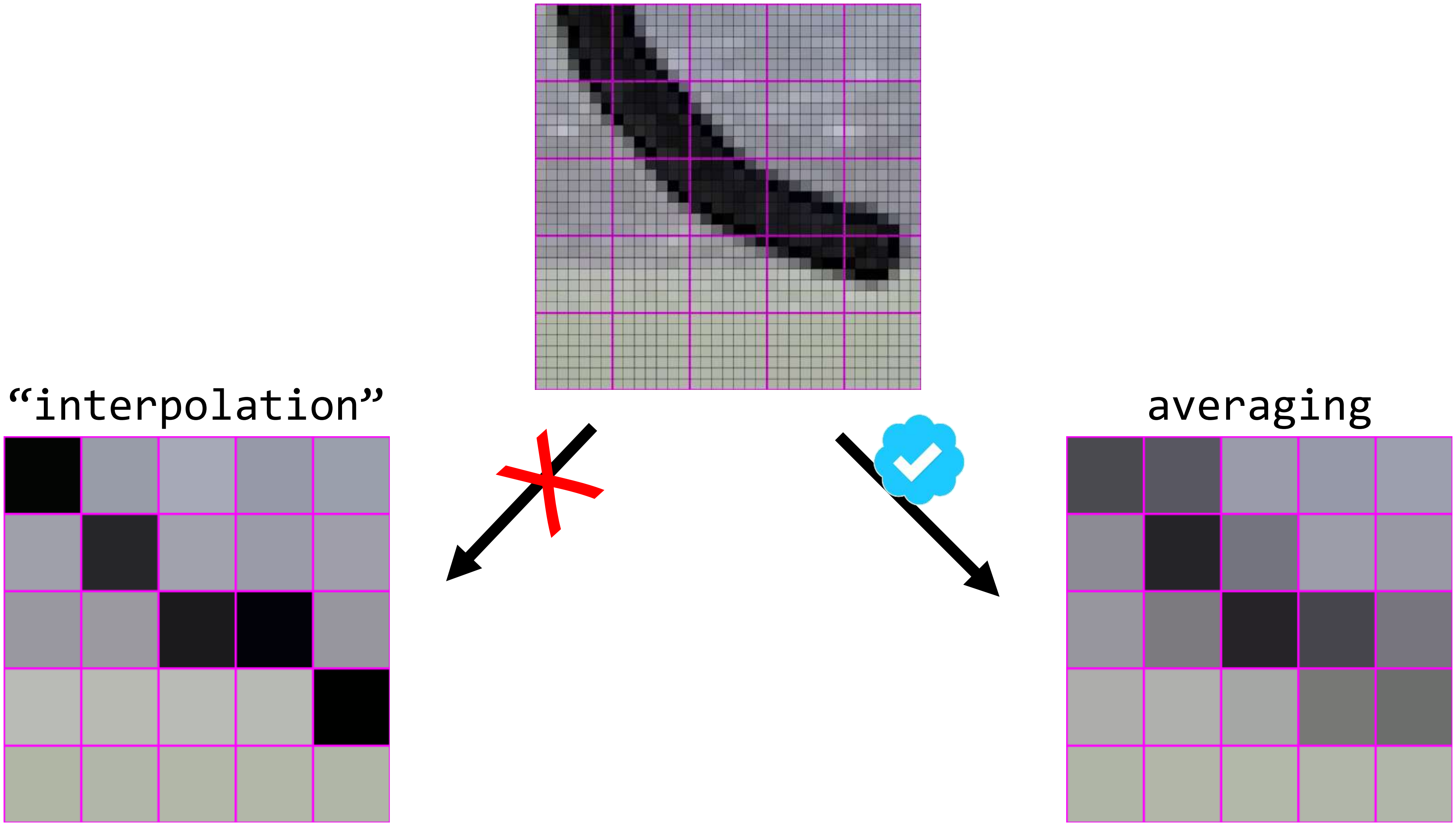
How? Averaging!



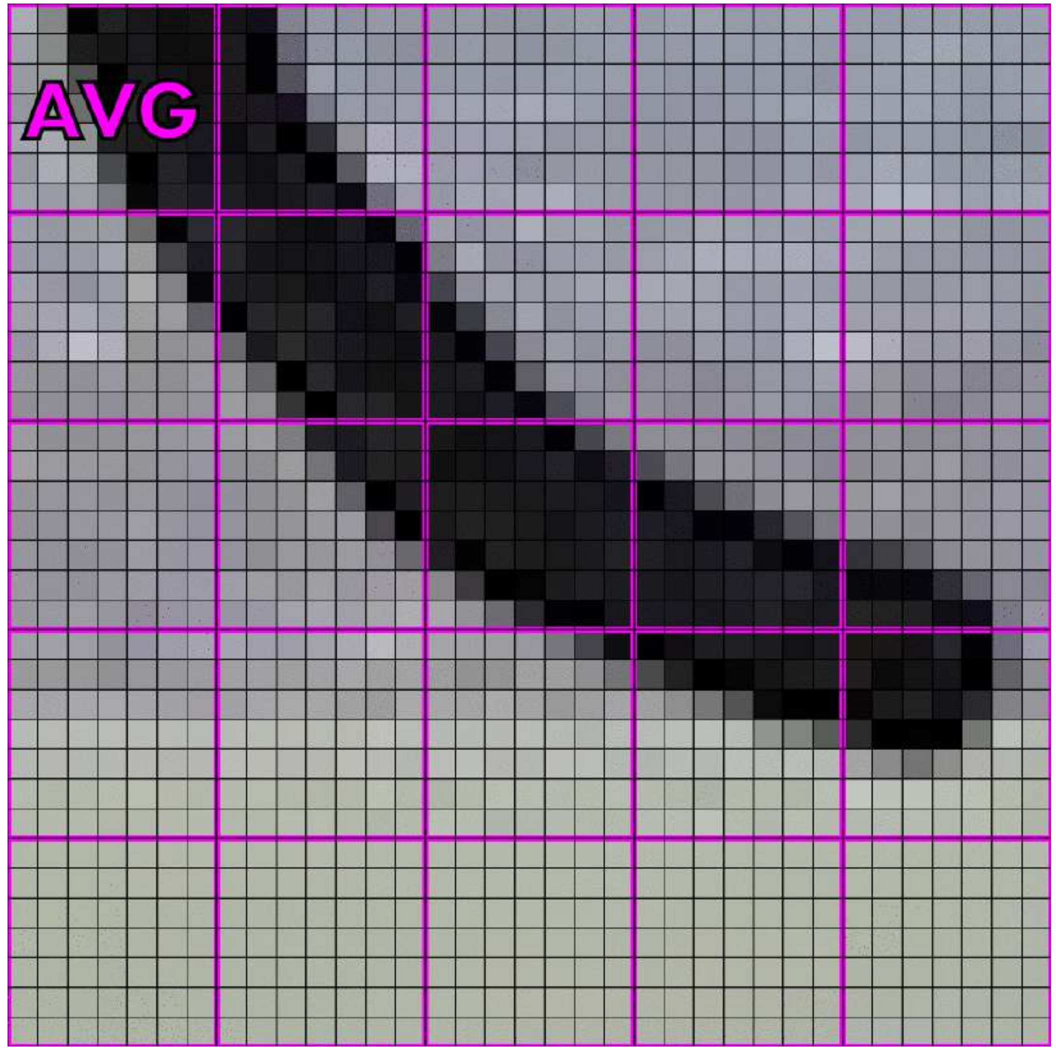


This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

How? Averaging!



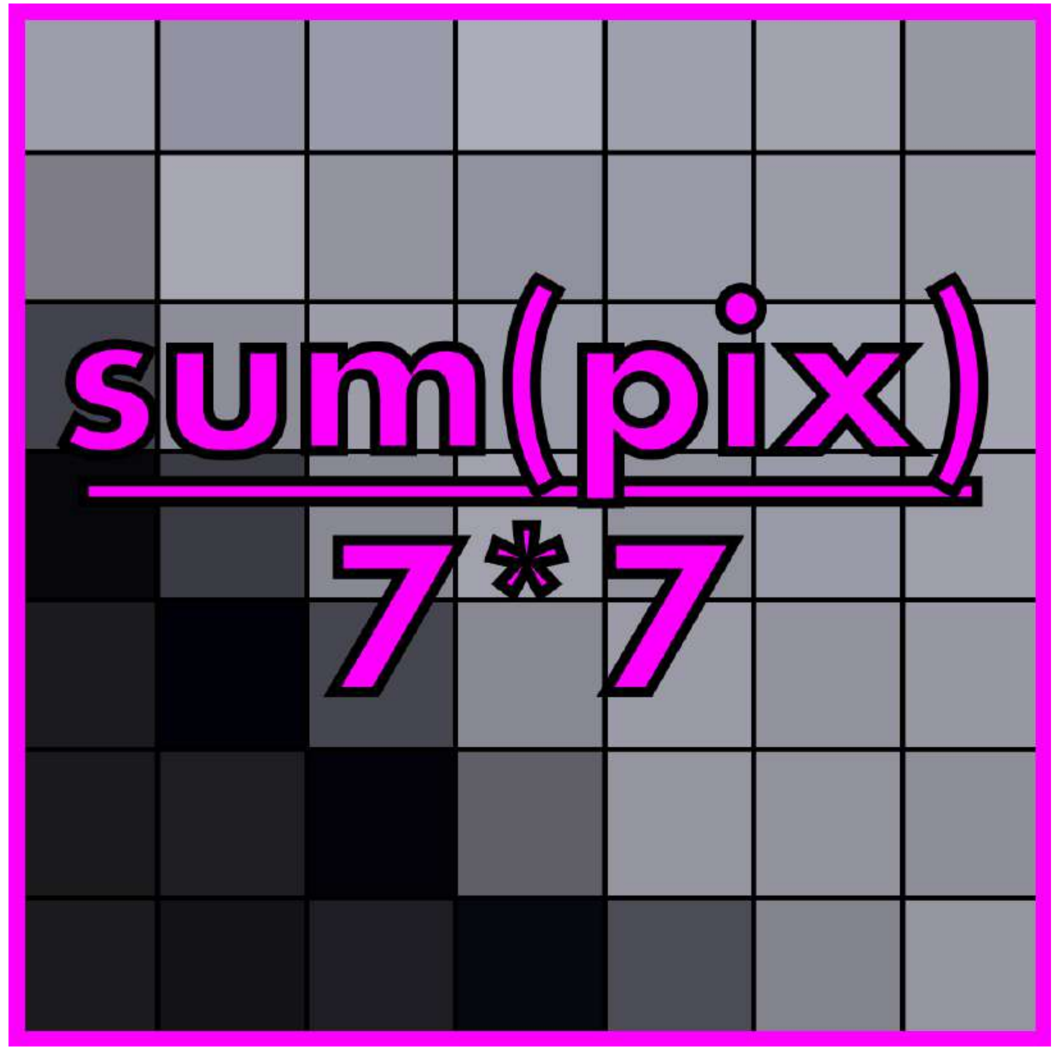
What is averaging?



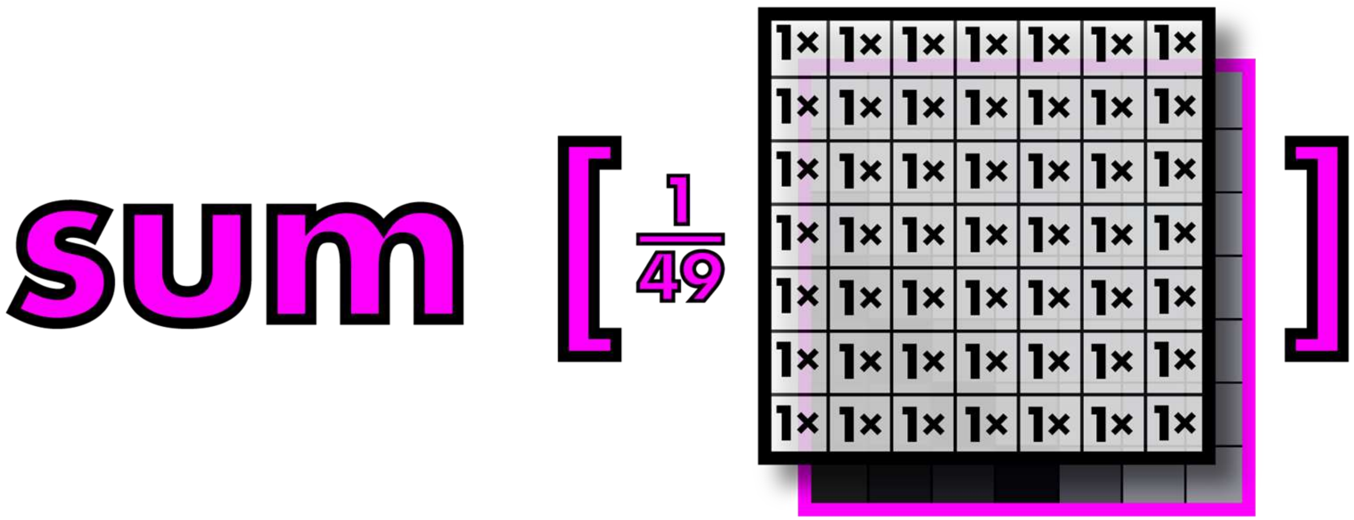
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What is averaging? A weighted sum



What is averaging? A weighted sum



What are the weights here ?

Today's Agenda

- Averaging vs Interpolation
- Systems - filters
- Convolution
  - Box Filter
  - Gaussian
  - Cross correlation vs Convolution
- Examples of filters

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Moving average is a filter

Filter or kernel

$$\frac{1}{49}\begin{bmatrix} 1\times & 1\times & 1\times & 1\times & 1\times & 1\times & 1\times \\ 1\times & 1\times & 1\times & 1\times & 1\times & 1\times & 1\times \\ 1\times & 1\times & 1\times & 1\times & 1\times & 1\times & 1\times \\ 1\times & 1\times & 1\times & 1\times & 1\times & 1\times & 1\times \\ 1\times & 1\times & 1\times & 1\times & 1\times & 1\times & 1\times \\ 1\times & 1\times & 1\times & 1\times & 1\times & 1\times & 1\times \\ 1\times & 1\times & 1\times & 1\times & 1\times & 1\times & 1\times \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Filtering

- Filtering
  - Forming a new image whose pixel values are transformed from original pixel values
- Goal is to extract useful information from images, or transform images into another domain where we can modify/enhance image properties
  - Features (edges, corners, blobs…)
  - Applications: super-resolution (resizing); in-painting; de-noising;

[Slide by Niebles]

Applications

De-noising

Salt and pepper noise

Super-resolution

In-painting

Bertamio et al

[Slide by Niebles]

Systems

- We define a **system** as a unit that converts an input function f[x,y] into an output (or response) function g[x,y], where (x,y) are the independent variables.
  - In the case of images, (x,y) represents the **spatial position in the image.**

[Slide by Niebles]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Moving average - example

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

[Slide by Seitz]

Moving average - example

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10							

[Slide by Seitz]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Moving average - example

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10	20						

[Slide by Seitz]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Moving average - example

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10	20	30					

[Slide by Seitz]

Moving average - example

$F[x, y]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$G[x, y]$

	0	10	20	30	30				

[Slide by Seitz]

Moving average - example

$$F[x, y]$$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$$G[x, y]$$

	0	10	20	30	30	30	20	10	
	0	20	40	60	60	60	40	20	
	0	30	60	90	90	90	60	30	
	0	30	50	80	80	90	60	30	
	0	30	50	80	80	90	60	30	
	0	20	30	50	50	60	40	20	
	10	20	30	30	30	30	20	10	
	10	10	10	0	0	0	0	0	

[Slide by Seitz]

Properties of systems

- Amplitude properties
  - Additivity $$S[f_i[n,m] + f_j[n,m]] = S[f_i[n,m]] + S[f_j[n,m]]$$
  - Homogeneity $$S[\alpha f_i[n,m]] = \alpha S[f_i[n,m]]]$$
  - Superposition $$S[\alpha f_i[n,m] + \beta f_j[n,m]] = \alpha S[f_i[n,m]] + \beta S[f_j[n,m]]$$
  - Stability $$|f[n,m]| \leq k \implies |g[n,m]| \leq ck$$
  - Invertibility $$S^{-1}[S[f_i[n,m]]] = f[n,m]$$

[Slide by Niebles]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Properties of systems

- Spatial properties
  - Causality $\quad$ for $n < n_0, m < m_0$, if $f[n,m] = 0 \implies g[n,m] = 0$
  - Shift invariance $\quad f[n-n_0, m-m_0] \xrightarrow{\mathcal{S}} g[n-n_0, m-m_0]$

[Slide by Niebles]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Linear Systems - filters

- Linear filtering
  - Form a new image whose pixels are a weighted sum of original pixel values
  - Use the same set of weights at each point
- S is a linear system (function) iff *S satisfies*

$$S[\alpha f_i[n,m] + \beta f_j[h,m]] = \alpha S[f_i[n,m]] + \beta S[f_j[h,m]]$$

superposition property

[Slide by Niebles]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Linear Shift Invariant Systems

- We call systems which satisfy the superposition and shift-invariant property *Linear Shift Invariant Systems* (LSI)

- Not all filters are LSI
  - Is thresholding linear?

$$g[n,m] = \begin{cases} 1, & f[n,m] > 100 \\ 0, & \text{otherwise.} \end{cases}$$

  - Consider:

    f1[n,m] + f2[n,m] > T
    f1[n,m] < T
    f2[n,m]<T

- LSI systems can be described by the **convolution** operation

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

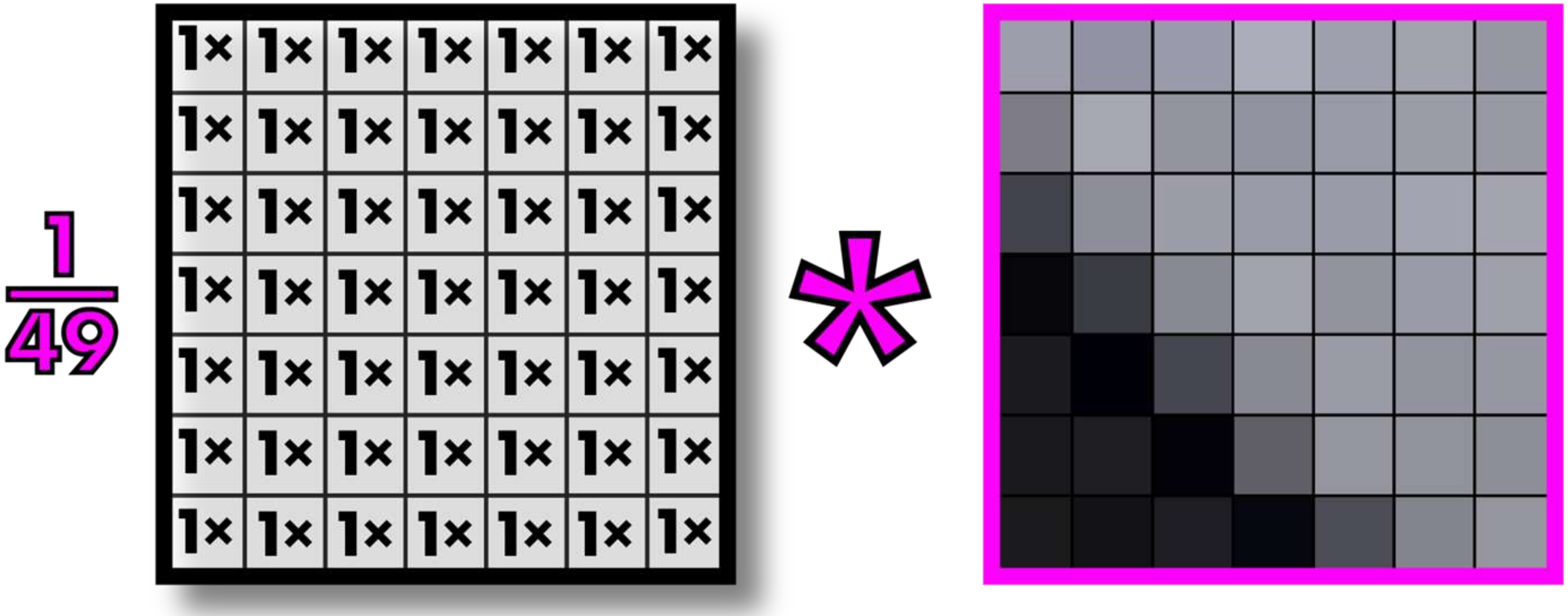
- Averaging vs Interpolation
- Systems - filters
- Convolution
  - Box Filter
  - Gaussian
  - Cross correlation vs Convolution
- Examples of filters

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Call this operation *"convolution"*

Filter or kernel

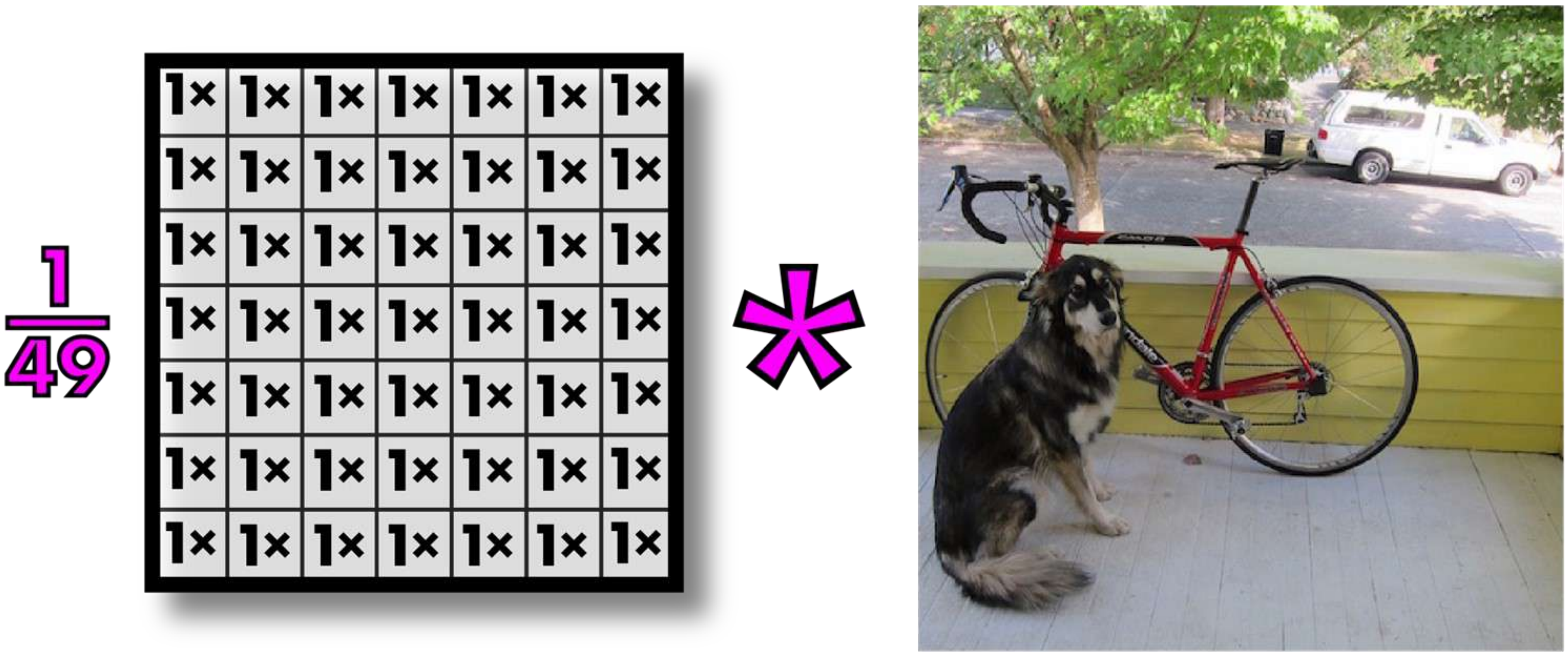


Note: multiplying an image section by a filter is actually called "correlation" and convolution involves inverting the filter first

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

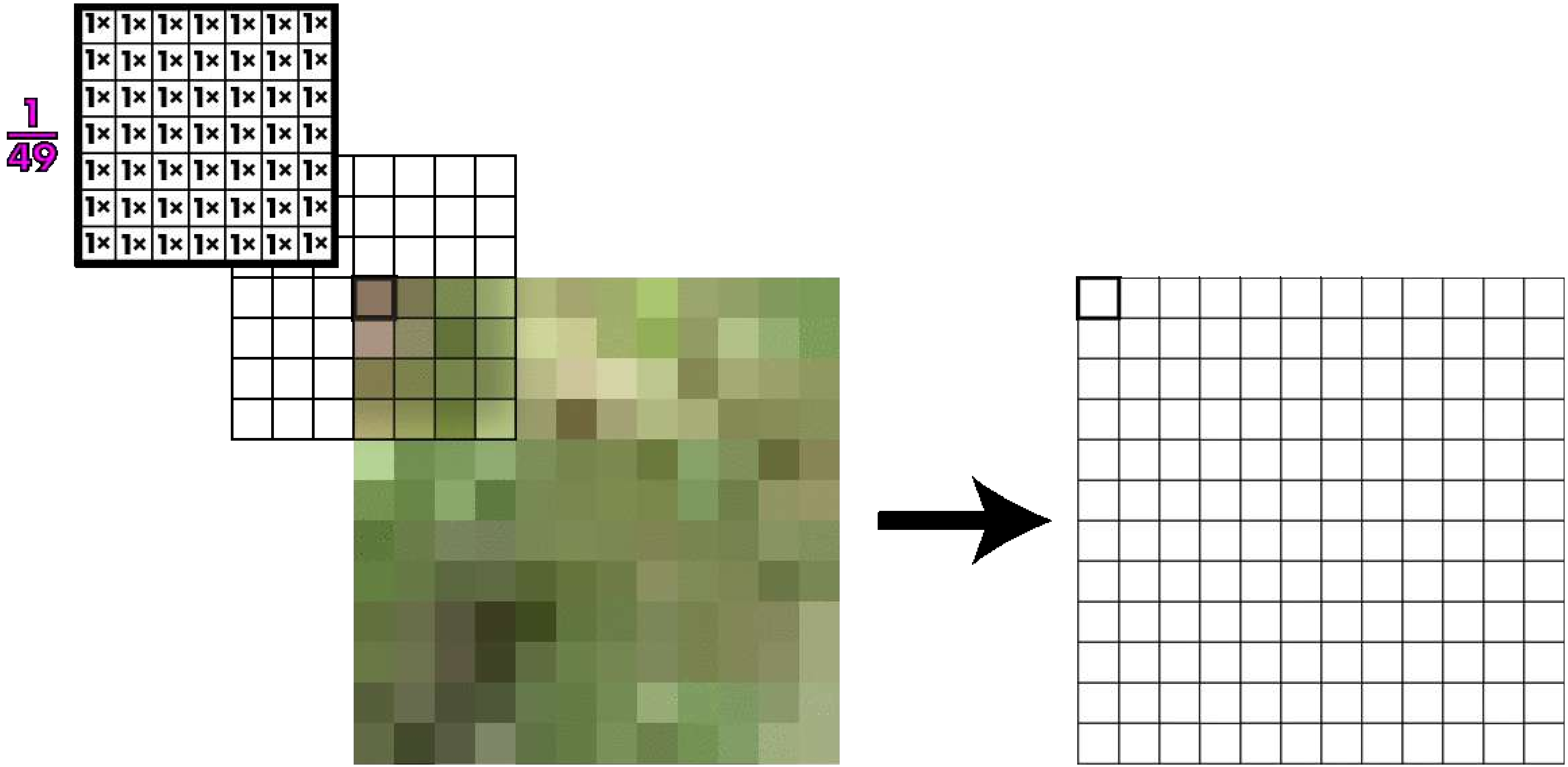
Convolutions on larger images



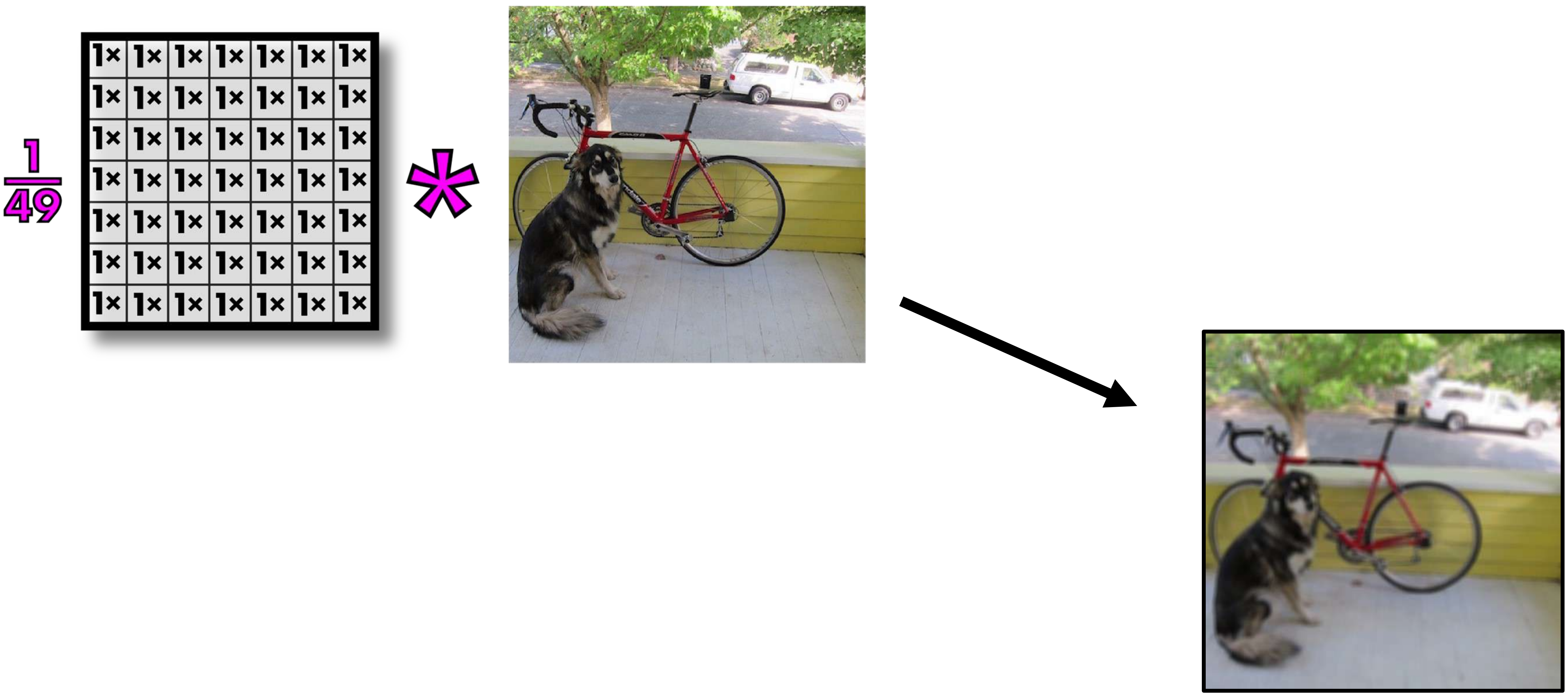
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Kernel slides across image

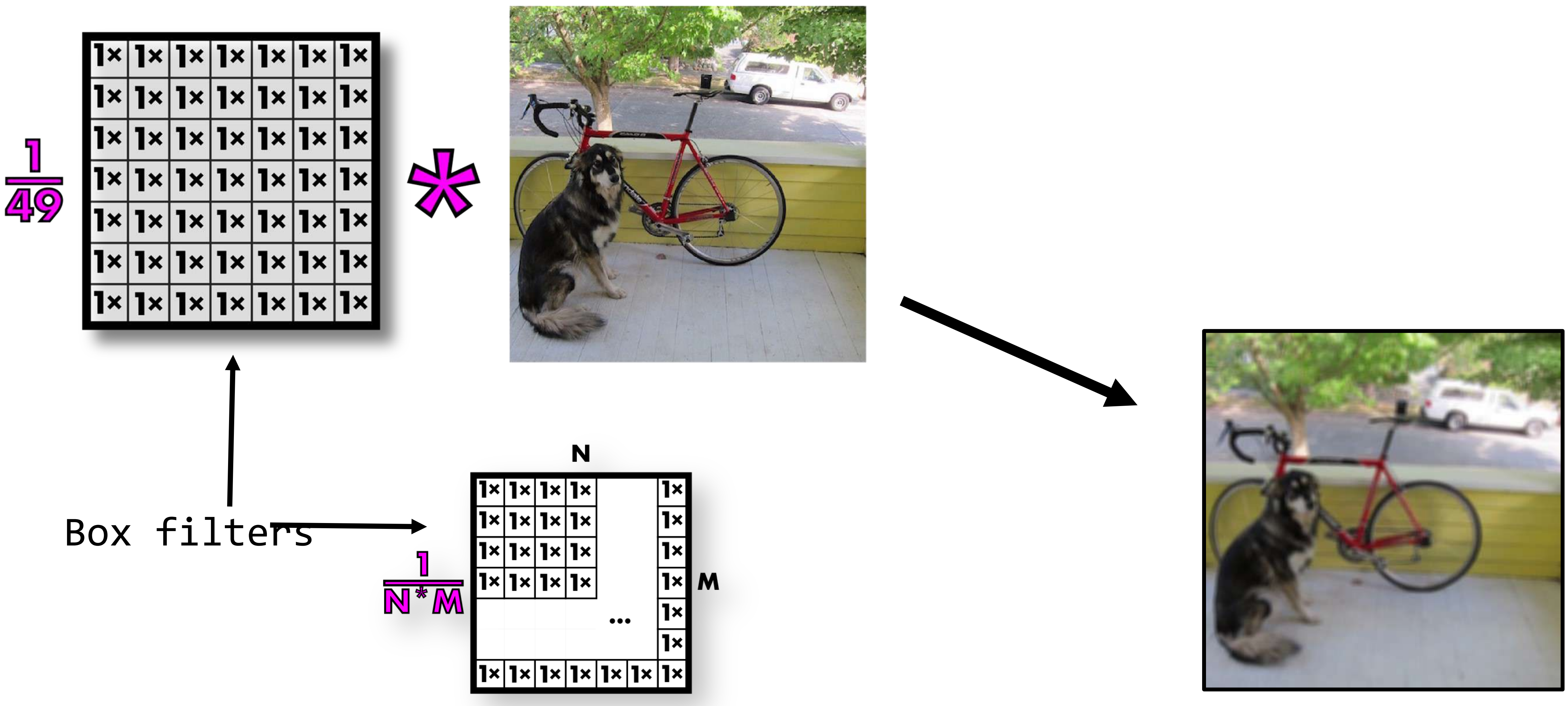


Convolutions on larger images

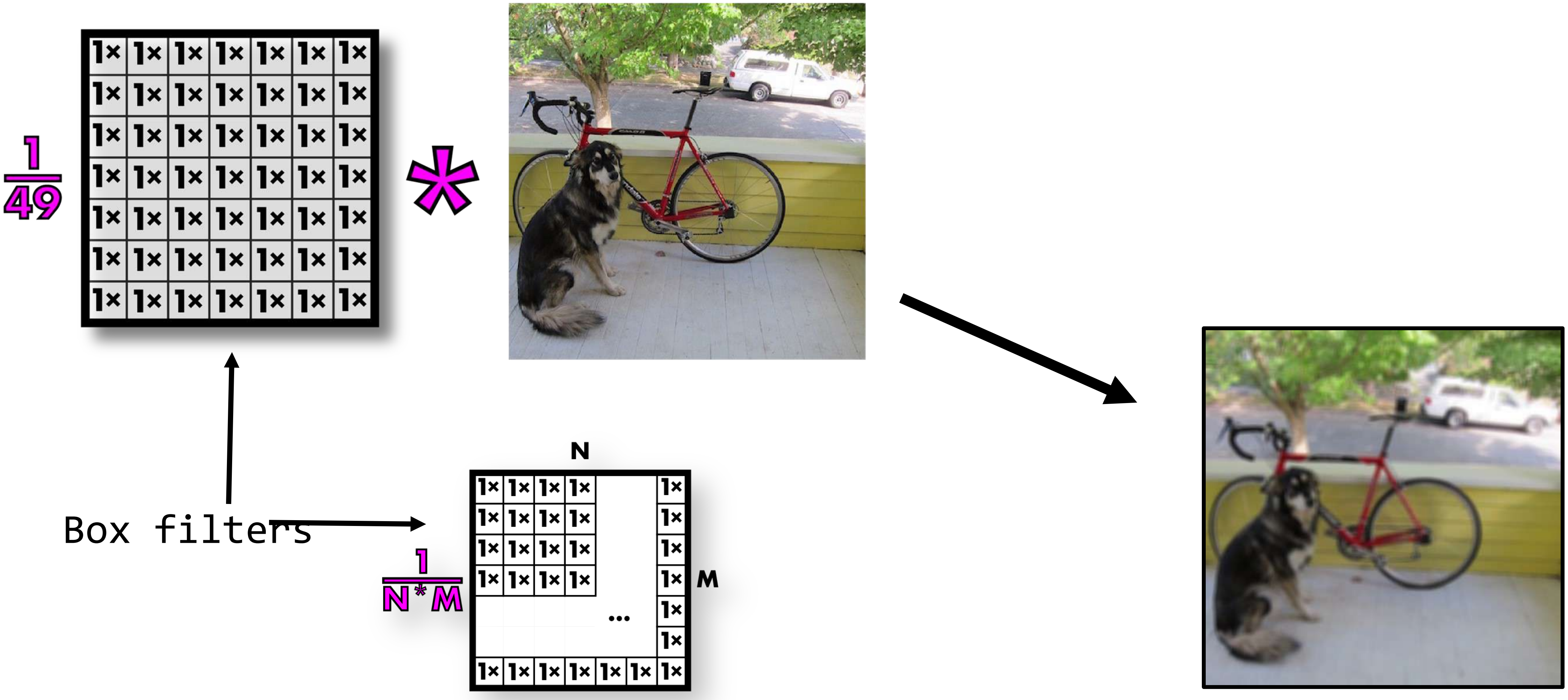


This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

This is called box filter

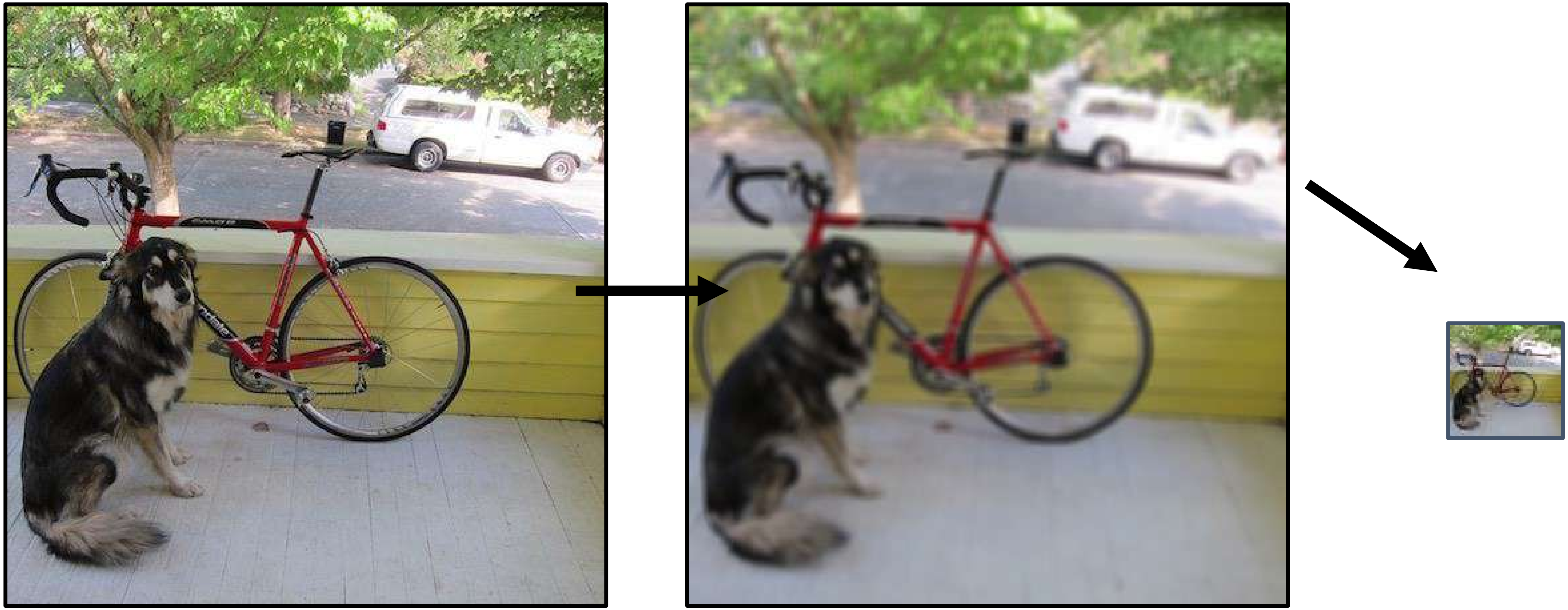


Box filters smooth image



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Now we resize our smoothed image

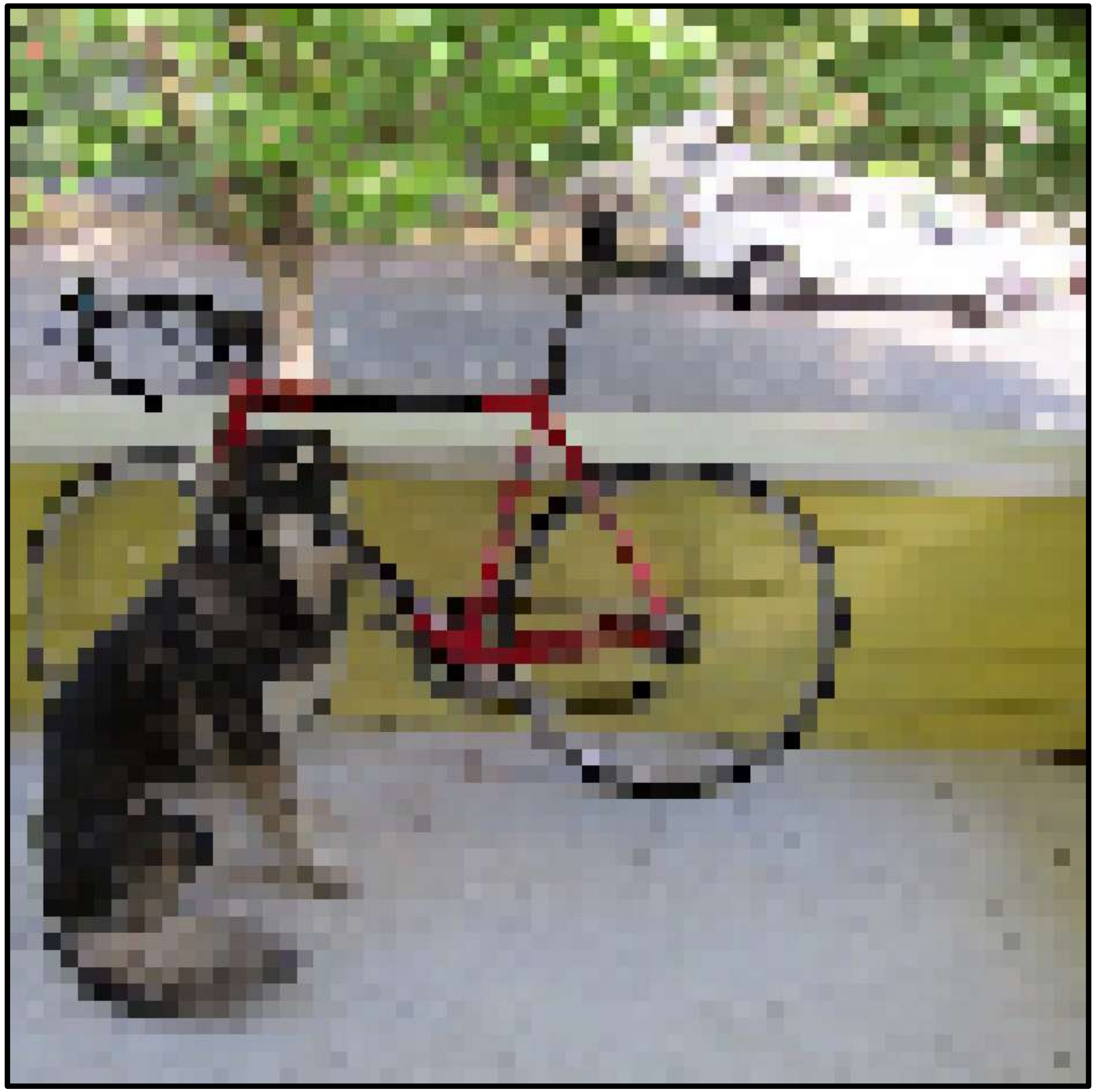


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

So much better!

Compare to interpolation



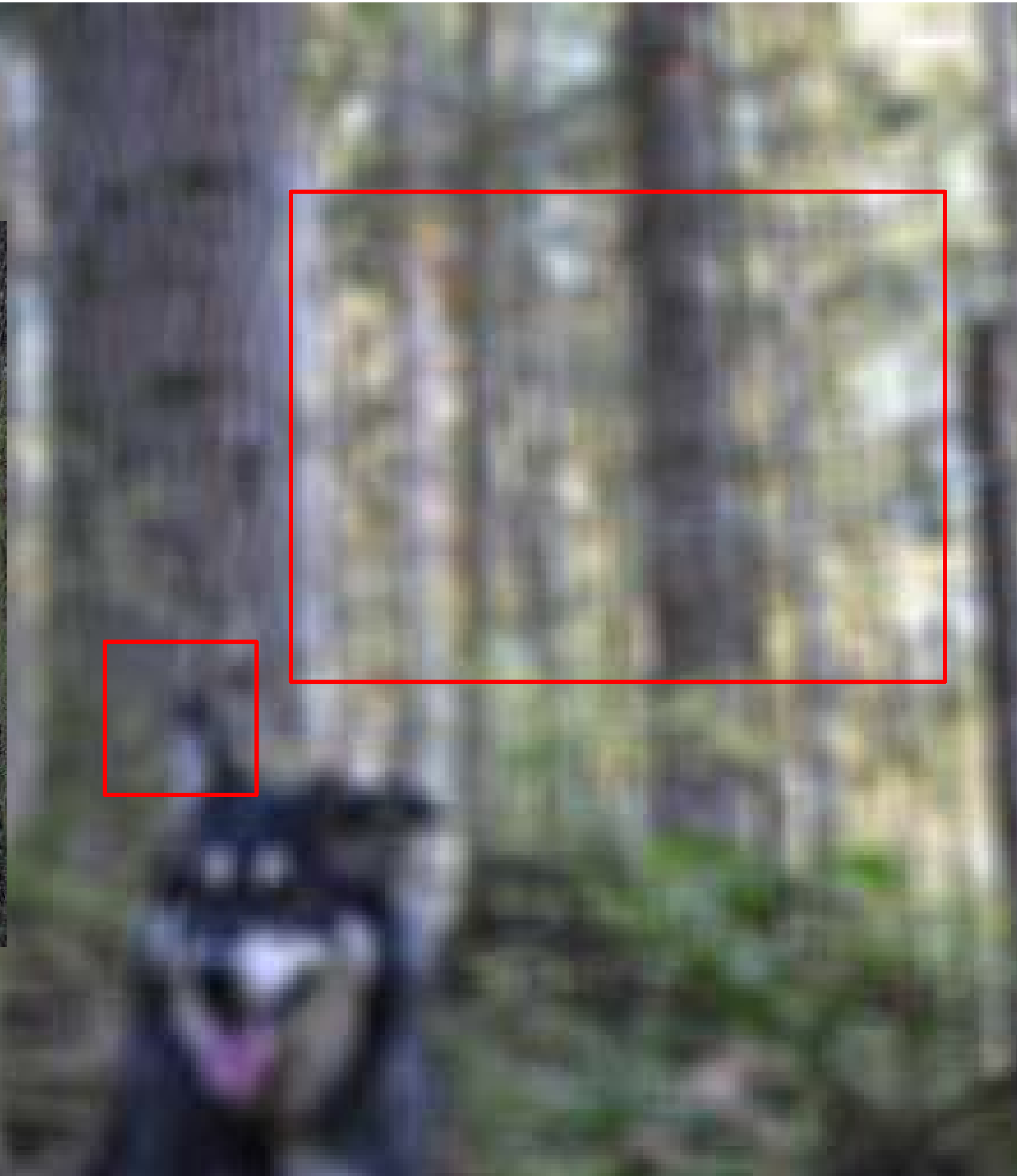
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Box filters have artifacts

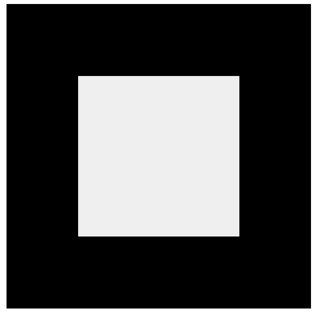
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Box filters have artifacts



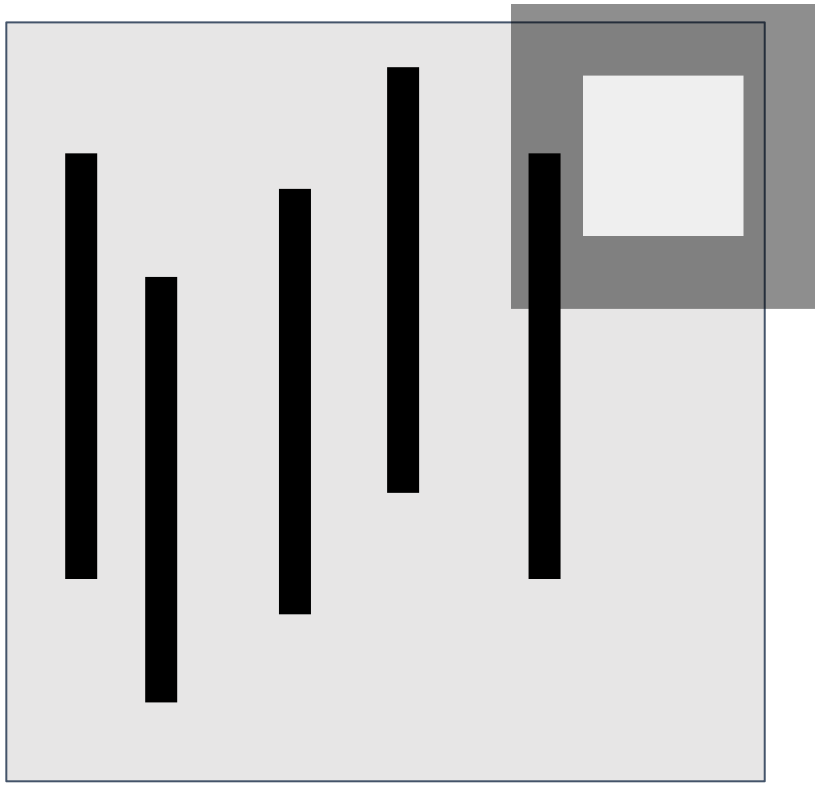
Box filters: vertical + horizontal streaking



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

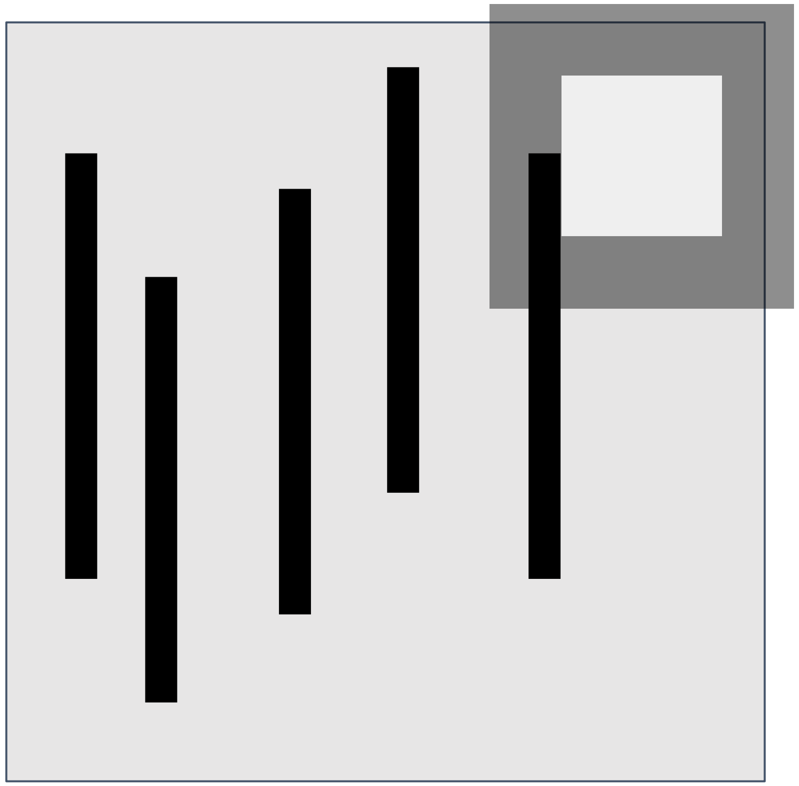
Box filters: vertical + horizontal streaking



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

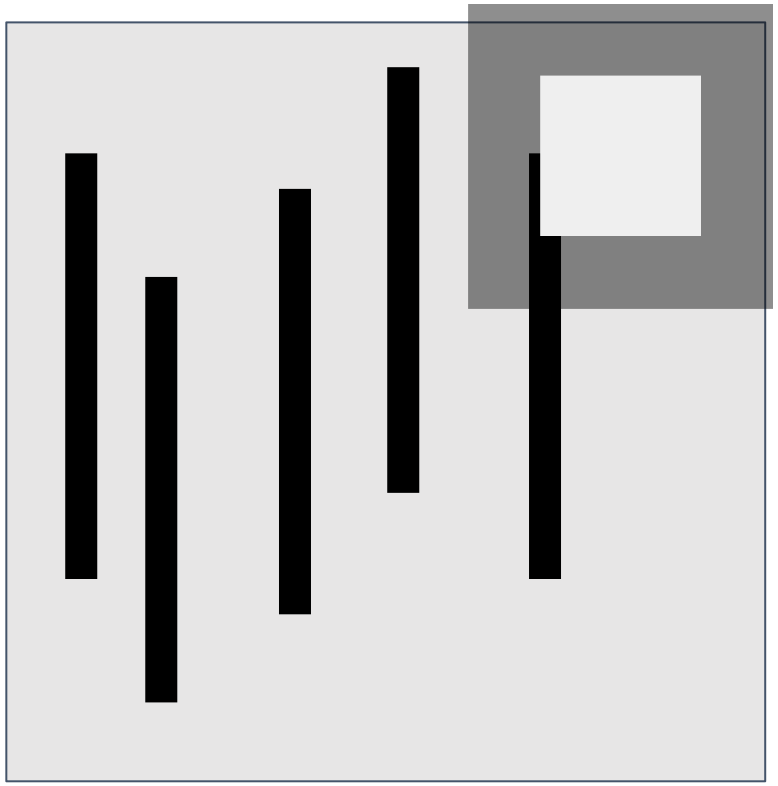
Box filters: vertical + horizontal streaking



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

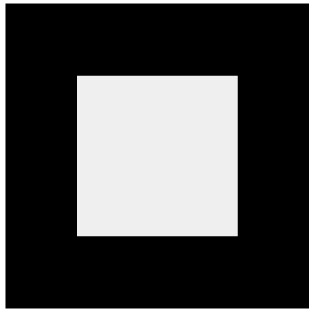
Box filters: vertical + horizontal streaking



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We want a smoothly weighted kernel



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

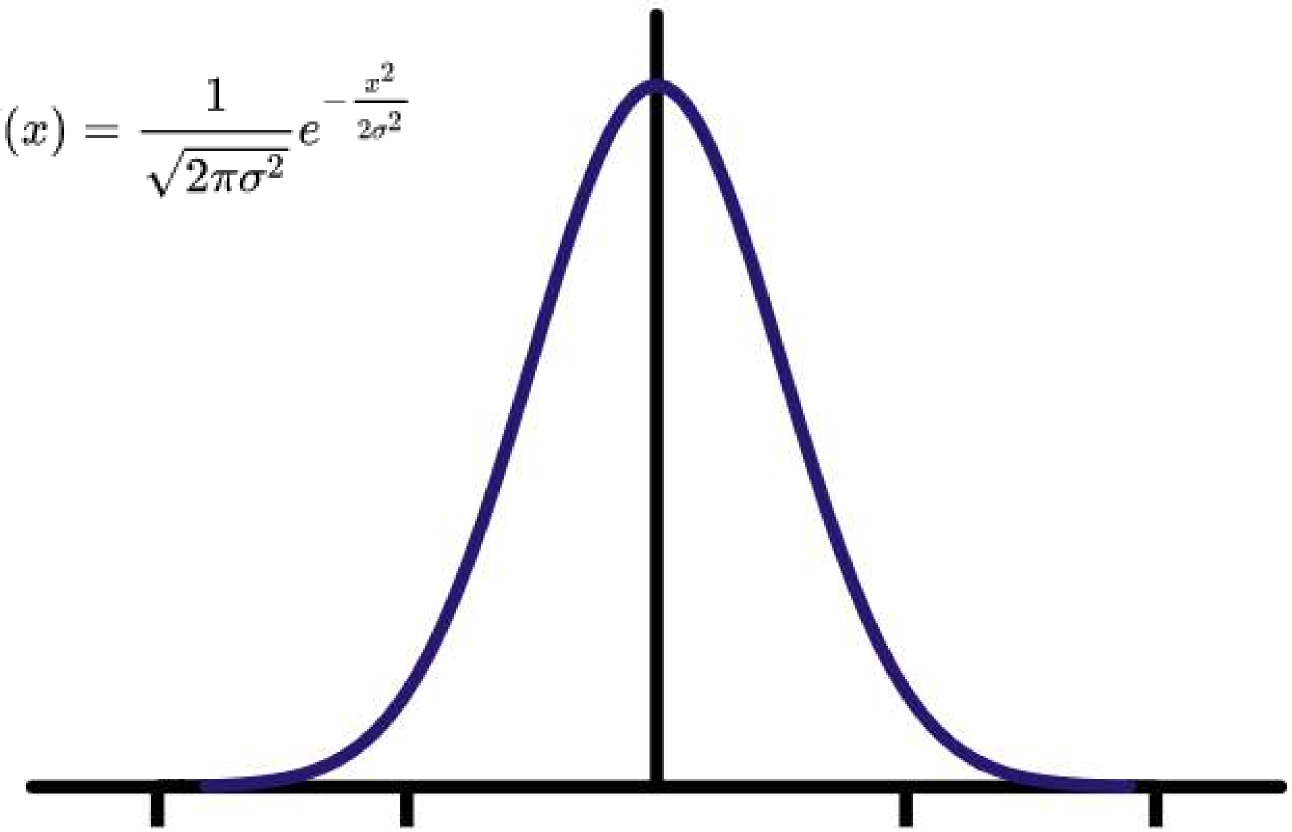
- Averaging vs Interpolation
- Systems - filters
- Convolution
  - Box Filter
  - Gaussian
  - Cross correlation vs Convolution
- Examples of filters

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

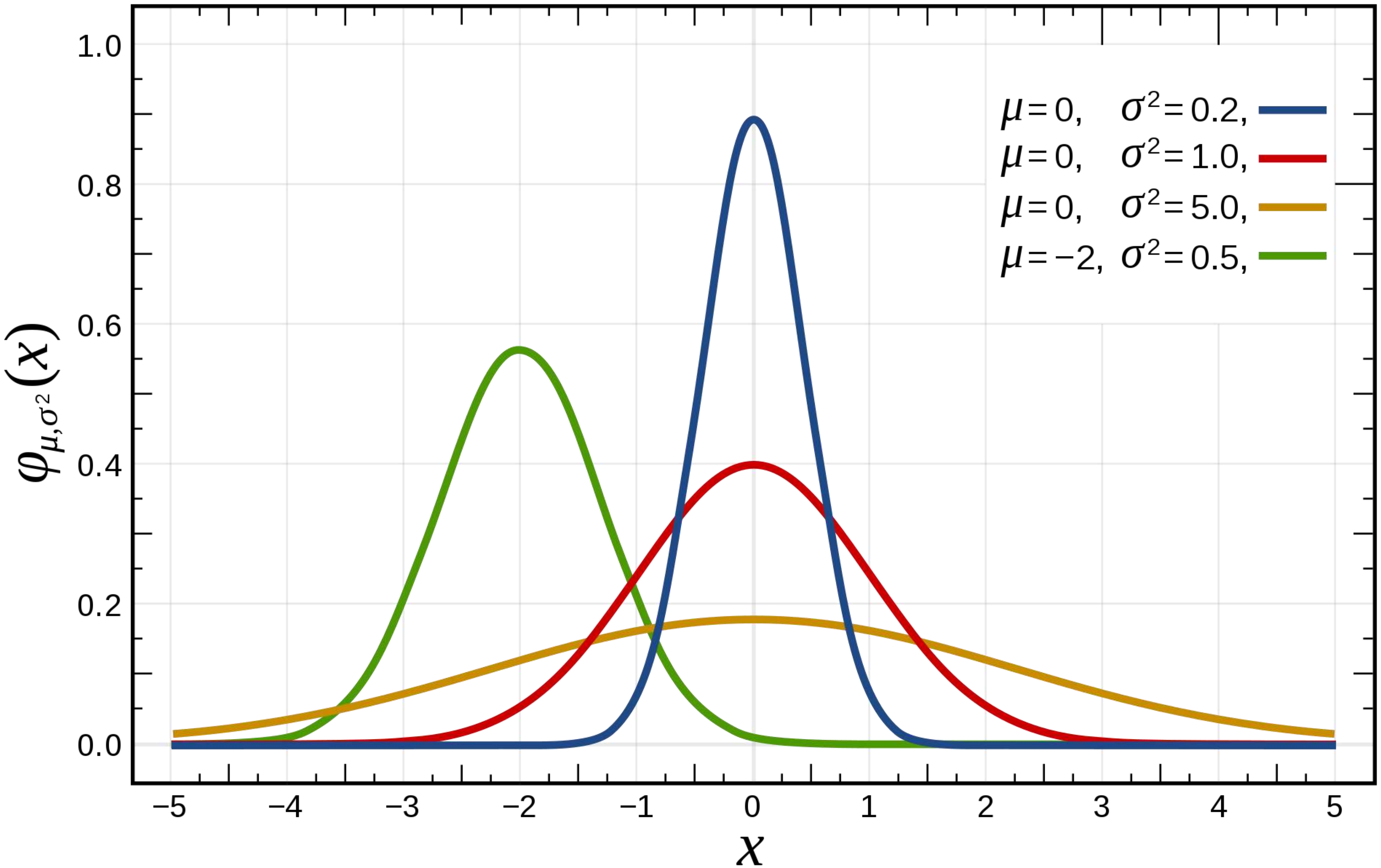
Gaussians

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$



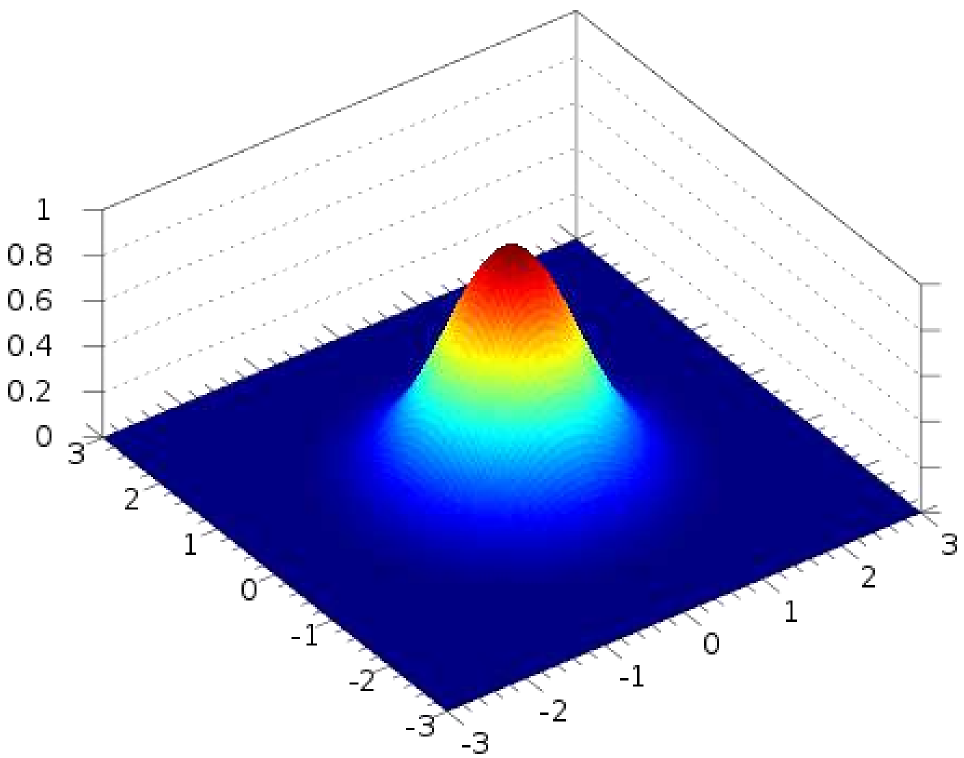
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

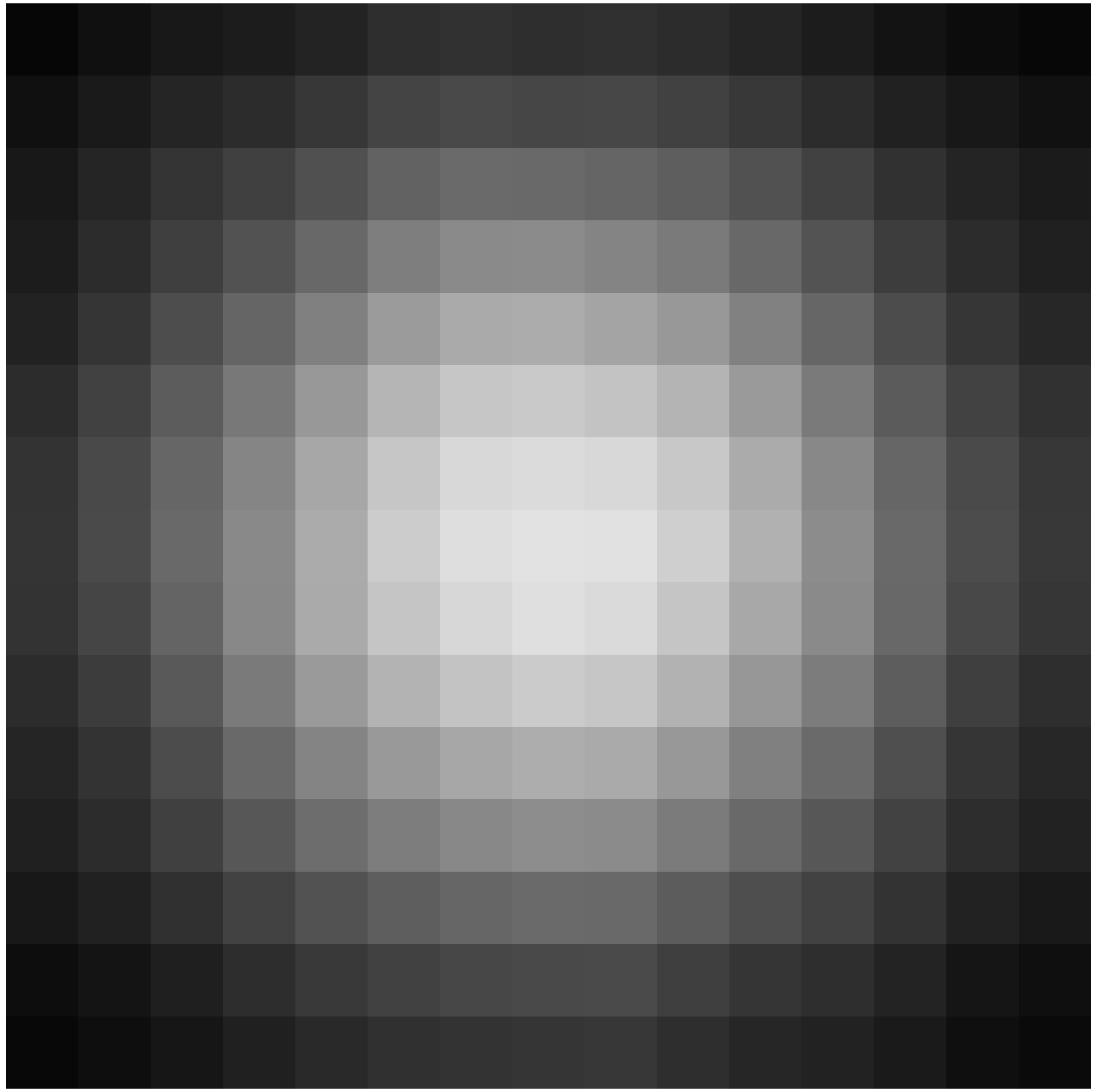
Gaussians – how σ affects the shape



2D Gaussian

$$g(x,y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}}$$



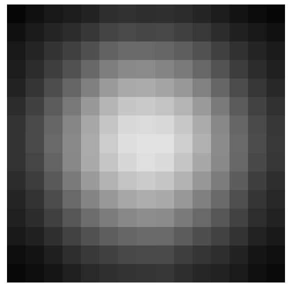


Example 7x7 Gaussian

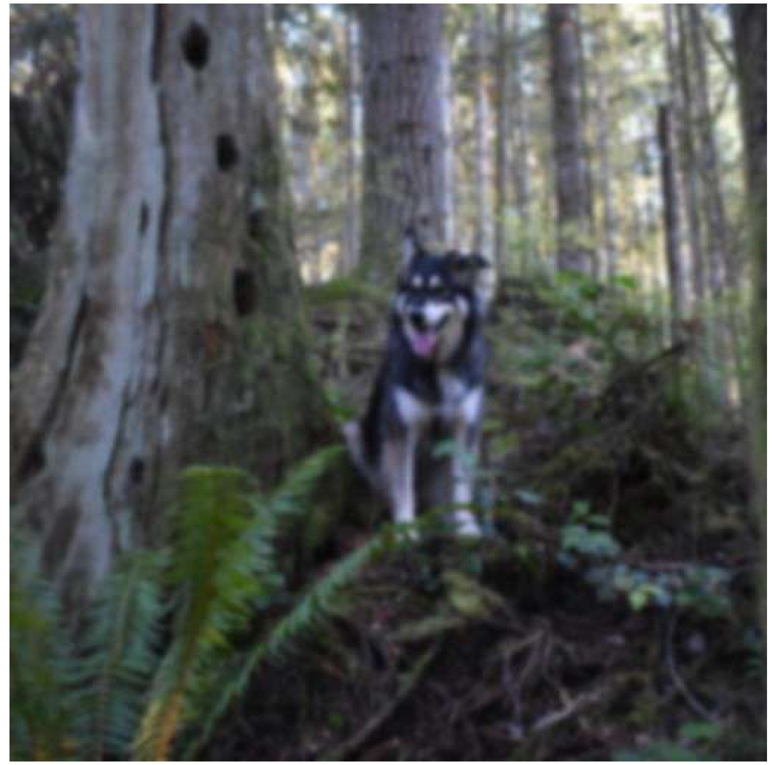
0.000	0.000	0.001	0.001	0.001	0.000	0.000
0.000	0.002	0.012	0.020	0.012	0.002	0.000
0.001	0.012	0.068	0.109	0.068	0.012	0.001
0.001	0.020	0.109	0.172	0.109	0.020	0.001
0.001	0.012	0.068	0.109	0.068	0.012	0.001
0.000	0.002	0.012	0.020	0.012	0.002	0.000
0.000	0.000	0.001	0.001	0.001	0.000	0.000

Better smoothing with Gaussians

$*$



$=$



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Better smoothing with Gaussians

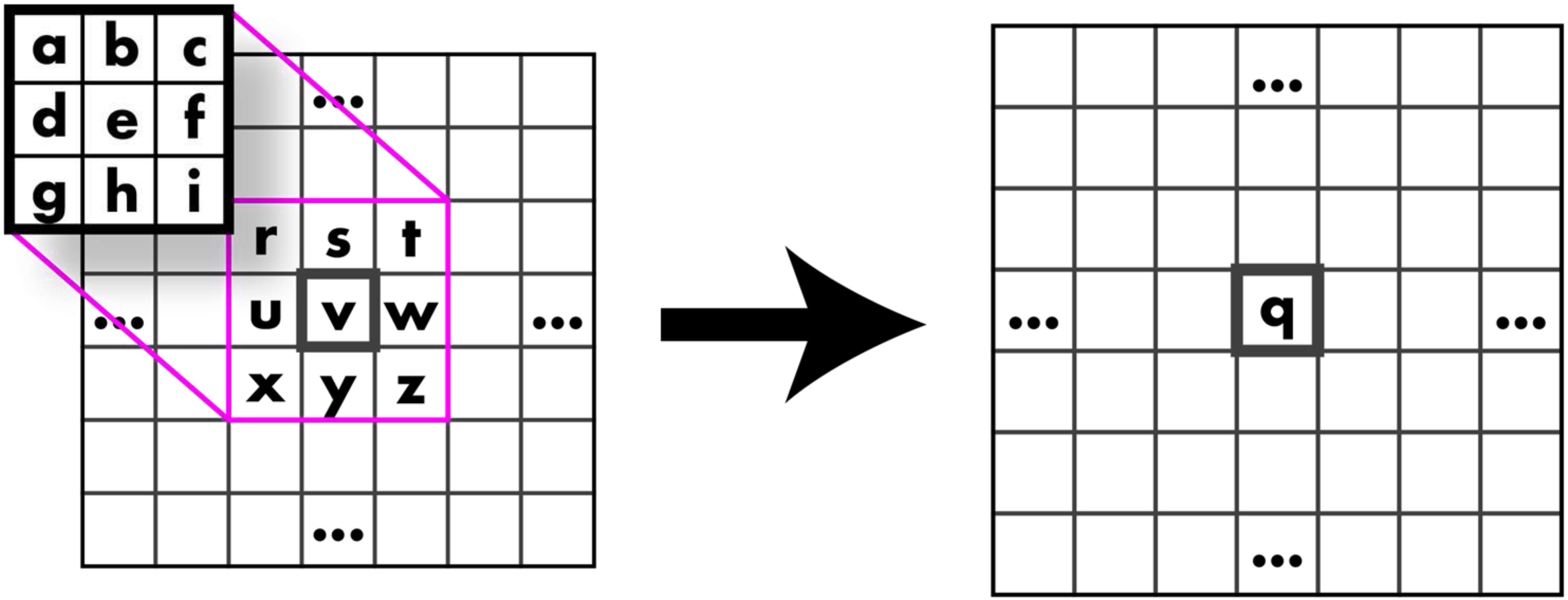
Better smoothing with Gaussians

Today’s Agenda

- Averaging vs Interpolation
- Systems - filters
- Convolution
  - Box Filter
  - Gaussian
  - Cross correlation vs Convolution
- Examples of filters

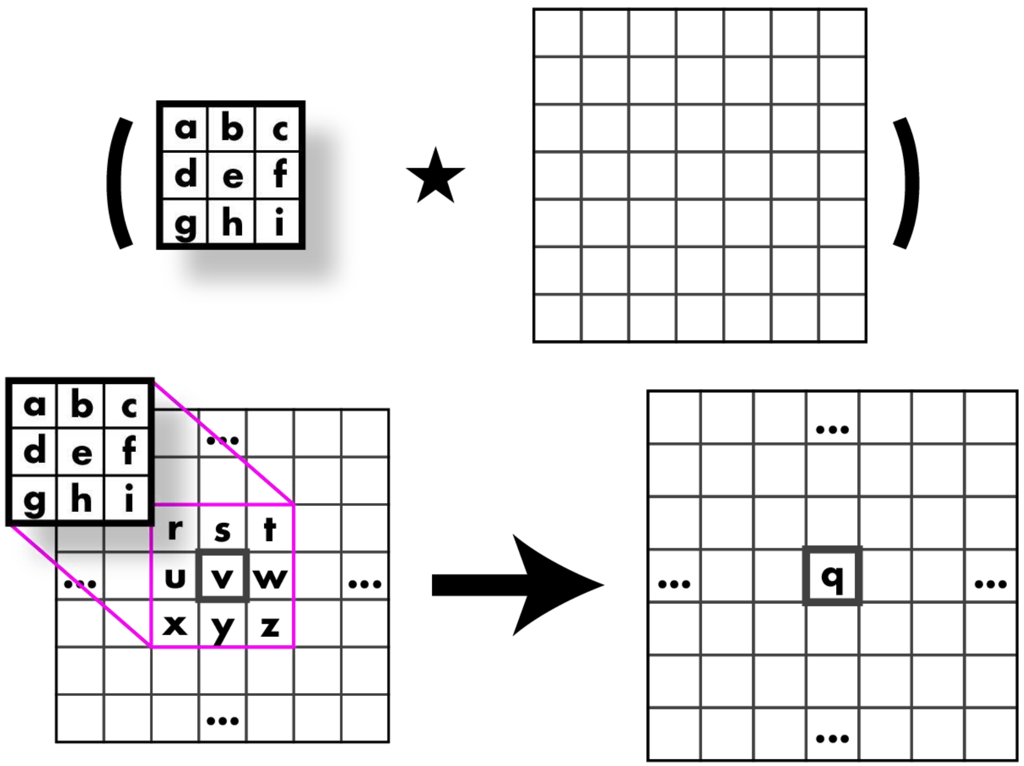
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

So what is convolution??



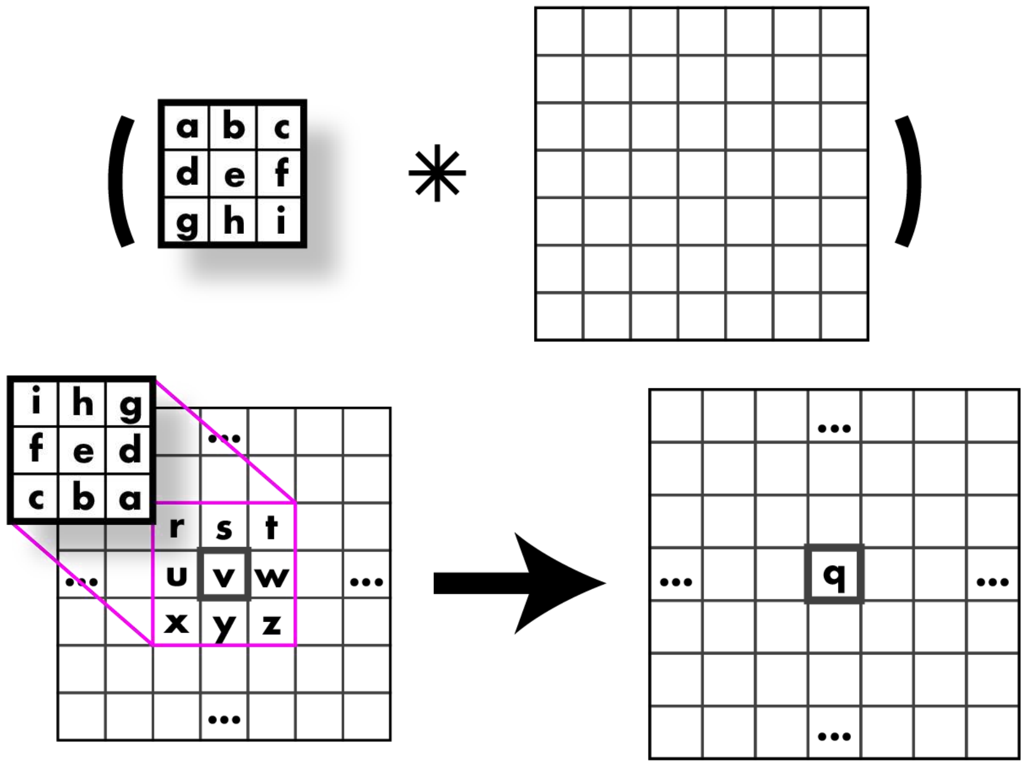
Cross-Correlation vs Convolution

Cross-Correlation



q = a×r + b×s + c×t + d×u + e×v + f×w + g×x + h×y + i×z

Convolution



q = i×r + h×s + g×t + f×u + e×v + d×w + c×x + b×y + a×z

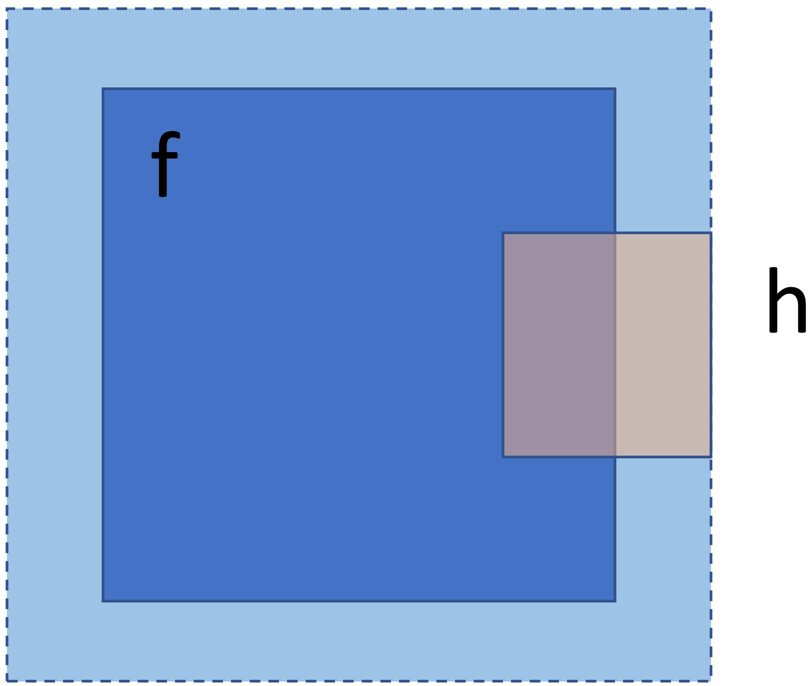
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Image support and edge effect

- A computer will only convolve **finite support signals**

- What happens at the edge?



- zero "padding"
- edge value replication
- mirror extension
- …

[Slide by Niebles]

2D convolution example

1	2	3
4	5	6
7	8	9

Input

n \ m	-1	0	1
-1	-1	-2	-1
0	0	0	0
1	1	2	1

Kernel

-13	-20	-17
-18	-24	-18
13	20	17

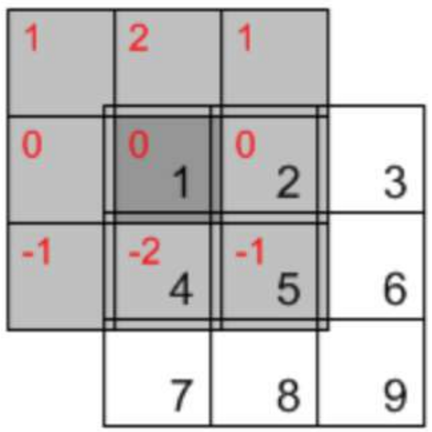
Output

[Slide by Song Ho Ahn]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

2D convolution example



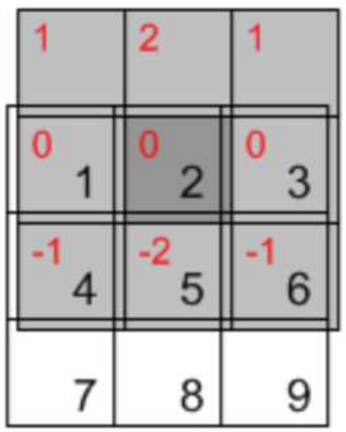
$$
\begin{aligned}
&= x[-1,-1]\cdot h[1,1] + x[0,-1]\cdot h[0,1] + x[1,-1]\cdot h[-1,1] \\
&\quad + x[-1,0]\cdot h[1,0] + x[0,0]\cdot h[0,0] + x[1,0]\cdot h[-1,0] \\
&\quad + x[-1,1]\cdot h[1,-1] + x[0,1]\cdot h[0,-1] + x[1,1]\cdot h[-1,-1] \\
&= 0\cdot 1 + 0\cdot 2 + 0\cdot 1 + 0\cdot 0 + 1\cdot 0 + 2\cdot 0 + 0\cdot(-1) + 4\cdot(-2) + 5\cdot(-1) = -13
\end{aligned}
$$

-13	-20	-17
-18	-24	-18
13	20	17

Output

[Slide by Song Ho Ahn]

2D convolution example



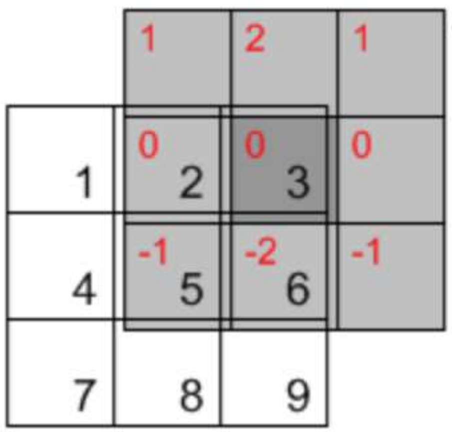
$$
\begin{aligned}
&= x[0,-1]\cdot h[1,1] + x[1,-1]\cdot h[0,1] + x[2,-1]\cdot h[-1,1] \\
&\quad + x[0,0]\cdot h[1,0] + x[1,0]\cdot h[0,0] + x[2,0]\cdot h[-1,0] \\
&\quad + x[0,1]\cdot h[1,-1] + x[1,1]\cdot h[0,-1] + x[2,1]\cdot h[-1,-1] \\
&= 0\cdot 1 + 0\cdot 2 + 0\cdot 1 + 1\cdot 0 + 2\cdot 0 + 3\cdot 0 + 4\cdot(-1) + 5\cdot(-2) + 6\cdot(-1) = -20
\end{aligned}
$$

-13	$\boxed{-20}$	-17
-18	-24	-18
13	20	17

Output

[Slide by Song Ho Ahn]

2D convolution example



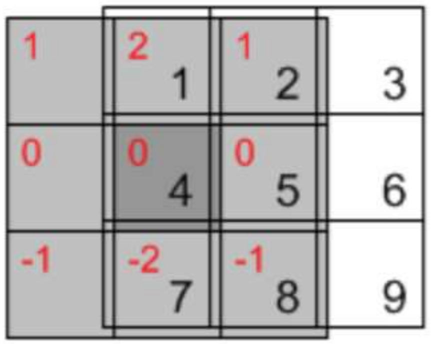
$$
\begin{aligned}
&= x[1,-1]\cdot h[1,1] + x[2,-1]\cdot h[0,1] + x[3,-1]\cdot h[-1,1] \\
&\quad + x[1,0]\cdot h[1,0] + x[2,0]\cdot h[0,0] + x[3,0]\cdot h[-1,0] \\
&\quad + x[1,1]\cdot h[1,-1] + x[2,1]\cdot h[0,-1] + x[3,1]\cdot h[-1,-1] \\
&= 0\cdot 1 + 0\cdot 2 + 0\cdot 1 + 2\cdot 0 + 3\cdot 0 + 0\cdot 0 + 5\cdot(-1) + 6\cdot(-2) + 0\cdot(-1) = -17
\end{aligned}
$$

-13	-20	$\boxed{-17}$
-18	-24	-18
13	20	17

Output

[Slide by Song Ho Ahn]

2D convolution example



$$
\begin{aligned}
&= x[-1,0]\cdot h[1,1] + x[0,0]\cdot h[0,1] + x[1,0]\cdot h[-1,1] \\
&\quad + x[-1,1]\cdot h[1,0] + x[0,1]\cdot h[0,0] + x[1,1]\cdot h[-1,0] \\
&\quad + x[-1,2]\cdot h[1,-1] + x[0,2]\cdot h[0,-1] + x[1,2]\cdot h[-1,-1] \\
&= 0\cdot 1 + 1\cdot 2 + 2\cdot 1 + 0\cdot 0 + 4\cdot 0 + 5\cdot 0 + 0\cdot(-1) + 7\cdot(-2) + 8\cdot(-1) = -18
\end{aligned}
$$

-13	-20	-17
$\boxed{-18}$	-24	-18
13	20	17

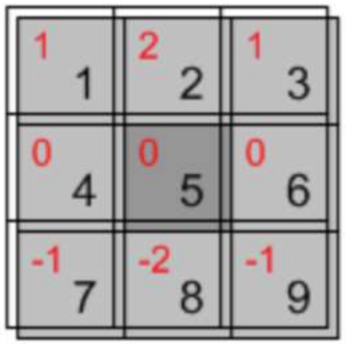
Output

[Slide by Song Ho Ahn]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

2D convolution example



$$
\begin{aligned}
&= x[0,0]\cdot h[1,1] + x[1,0]\cdot h[0,1] + x[2,0]\cdot h[-1,1] \\
&\quad + x[0,1]\cdot h[1,0] + x[1,1]\cdot h[0,0] + x[2,1]\cdot h[-1,0] \\
&\quad + x[0,2]\cdot h[1,-1] + x[1,2]\cdot h[0,-1] + x[2,2]\cdot h[-1,-1] \\
&= 1\cdot 1 + 2\cdot 2 + 3\cdot 1 + 4\cdot 0 + 5\cdot 0 + 6\cdot 0 + 7\cdot(-1) + 8\cdot(-2) + 9\cdot(-1) = -24
\end{aligned}
$$

-13	-20	-17
-18	$\boxed{-24}$	-18
13	20	17

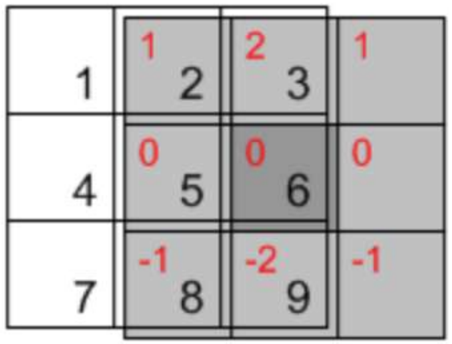
Output

[Slide by Song Ho Ahn]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

2D convolution example



$$
\begin{aligned}
&= x[1,0]\cdot h[1,1] + x[2,0]\cdot h[0,1] + x[3,0]\cdot h[-1,1] \\
&\quad + x[1,1]\cdot h[1,0] + x[2,1]\cdot h[0,0] + x[3,1]\cdot h[-1,0] \\
&\quad + x[1,2]\cdot h[1,-1] + x[2,2]\cdot h[0,-1] + x[3,2]\cdot h[-1,-1] \\
&= 2\cdot 1 + 3\cdot 2 + 0\cdot 1 + 5\cdot 0 + 6\cdot 0 + 0\cdot 0 + 8\cdot(-1) + 9\cdot(-2) + 0\cdot(-1) = -18
\end{aligned}
$$

-13	-20	-17
-18	-24	$\boxed{-18}$
13	20	17

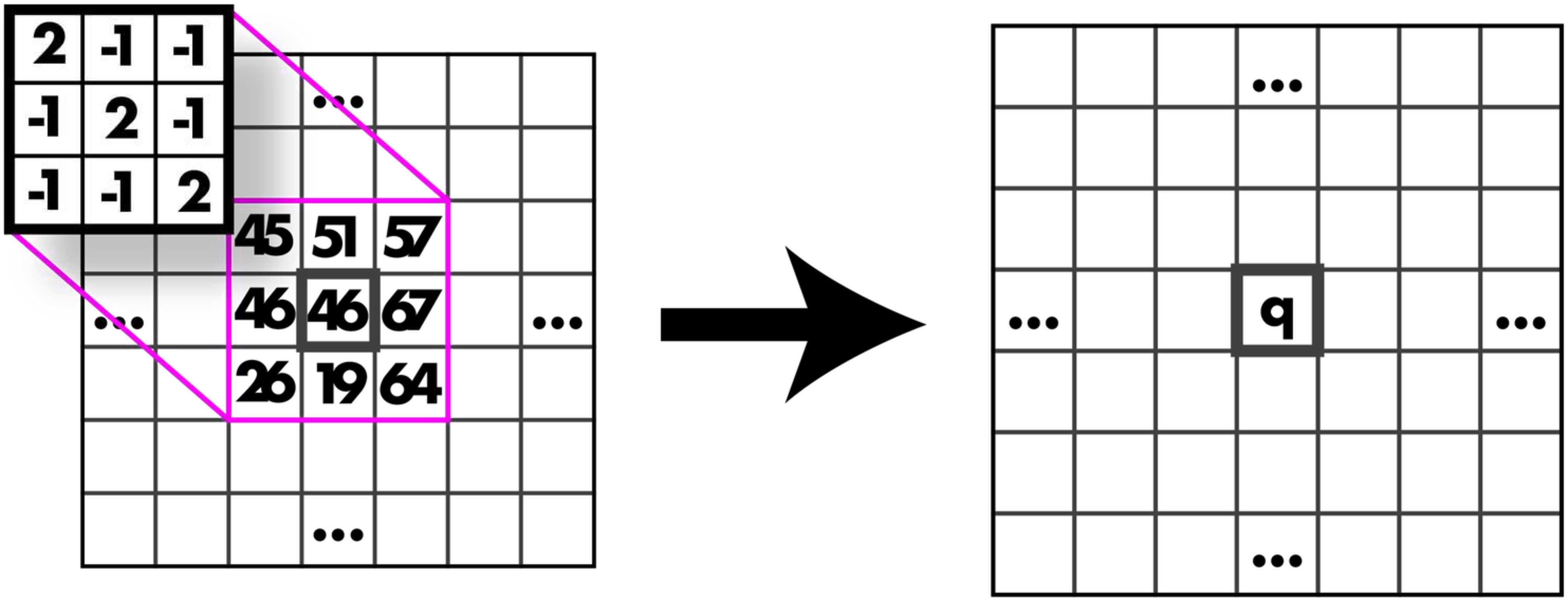
Output

[Slide by Song Ho Ahn]

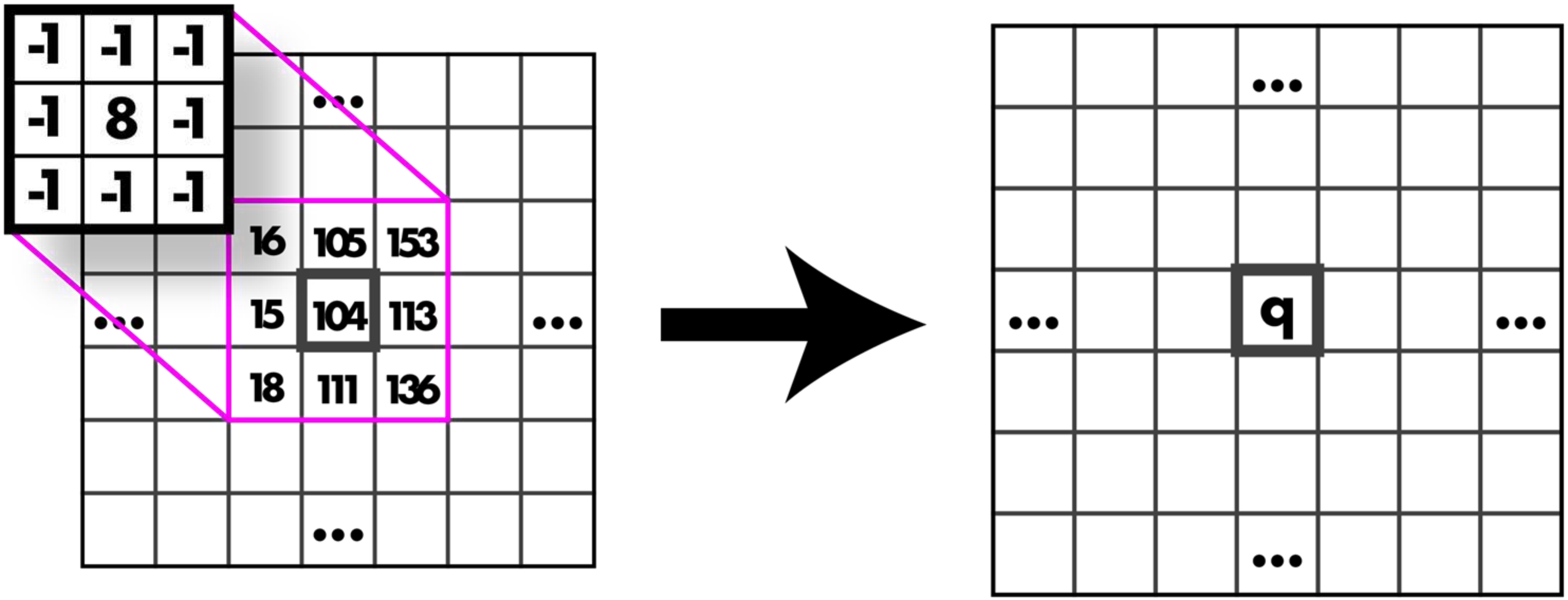
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Calculate it!



Calculate it!



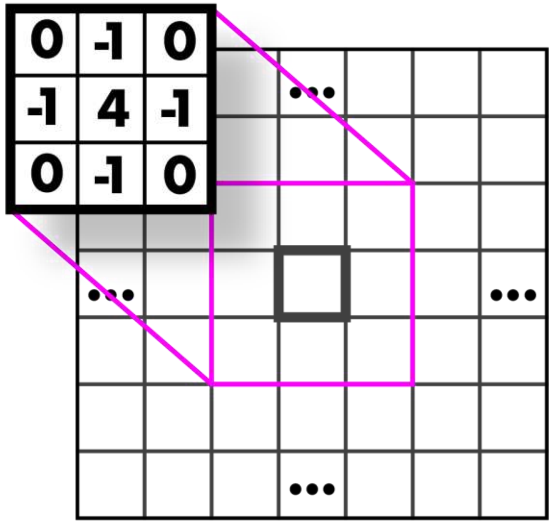
Today’s Agenda

- Averaging vs Interpolation
- Systems - filters
- Convolution
  - Box Filter
  - Gaussian
  - Cross correlation vs Convolution
- Examples of filters

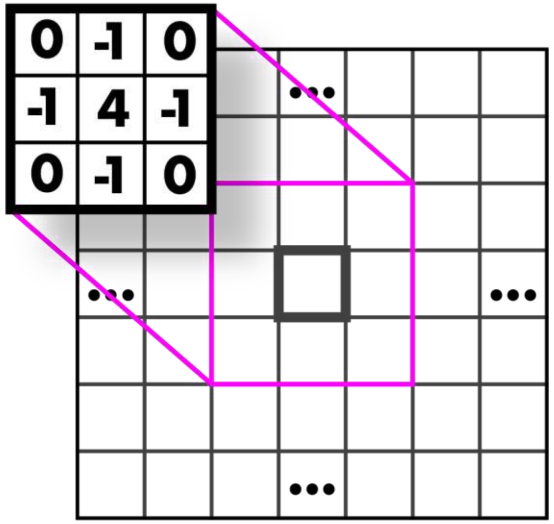
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Guess that kernel!

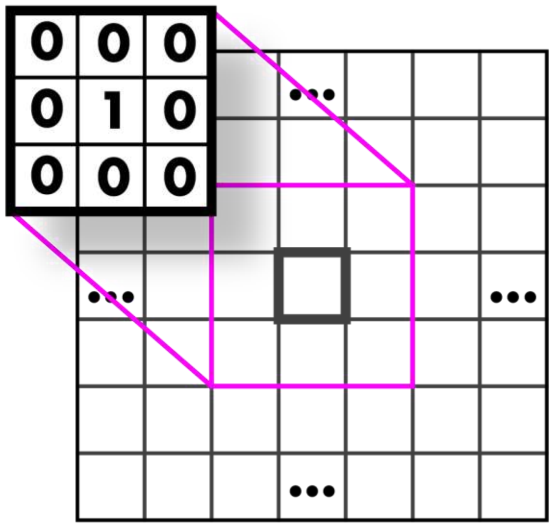


Highpass Kernel: finds edges

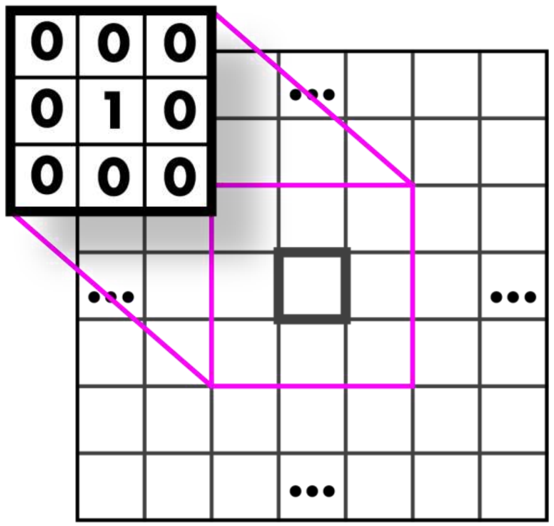


applied to grayscale

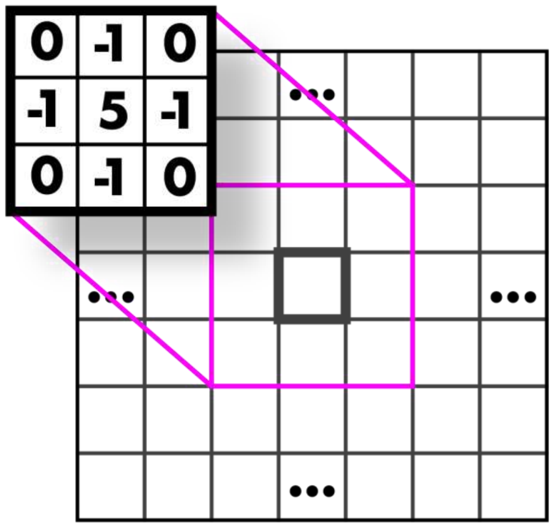
Guess that kernel!



Identity Kernel: Does nothing!

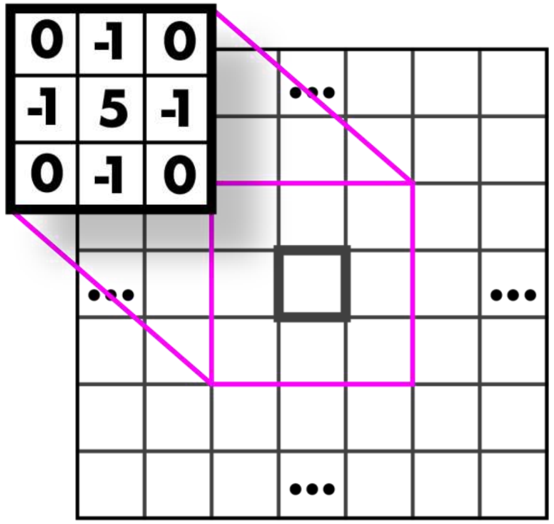


Guess that kernel!



Sharpen Kernel: sharpens!

applied to all three channels



Note: sharpen = highpass + identity! Why ?

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Sharpen Kernel: sharpens!

What does blurring take away?

Highpass

–

=

Let’s add it back:

Identity +
Highpass

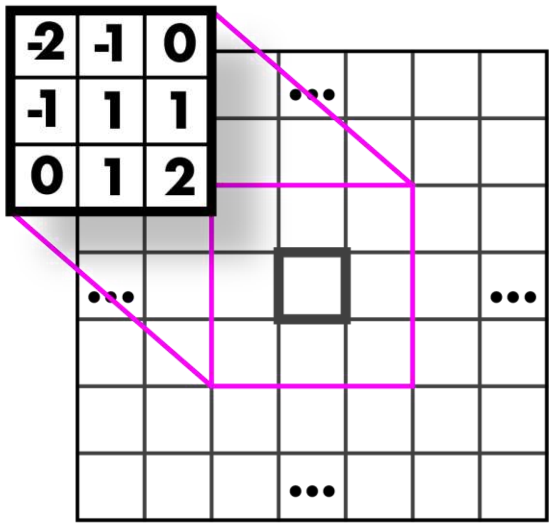
+

=

[Slide by D. Lowe]

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Guess that kernel!

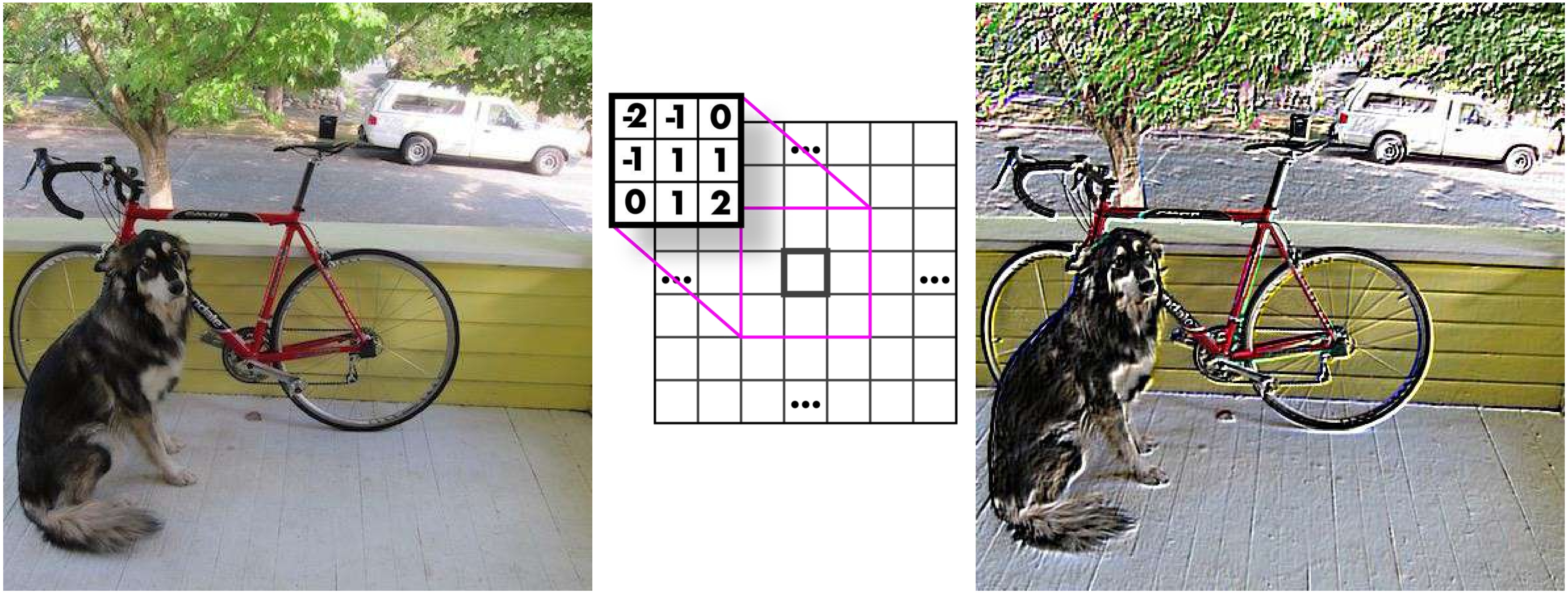


Co-financed by the European Union
Connecting Europe Facility

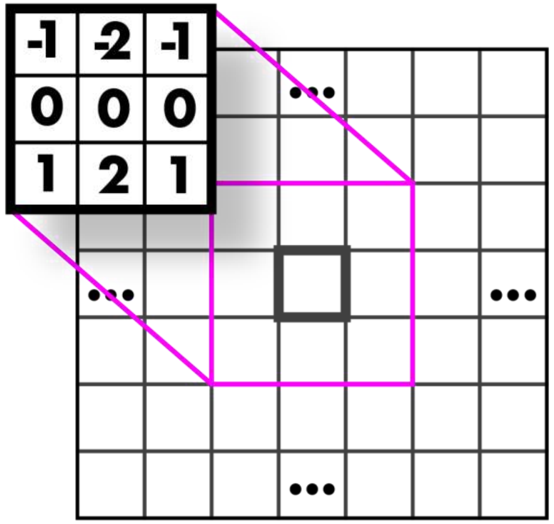
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

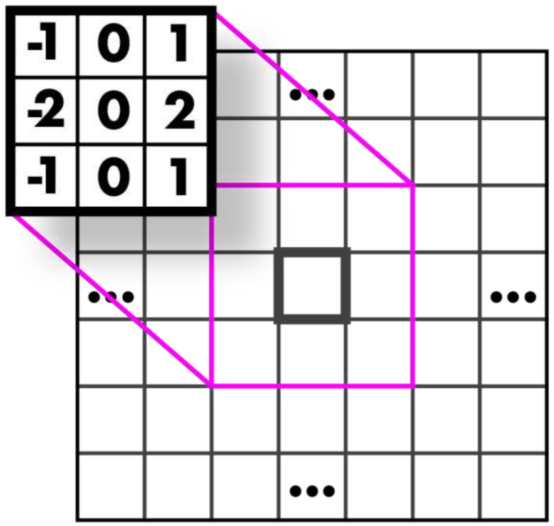
Emboss Kernel: styling

applied to all three channels



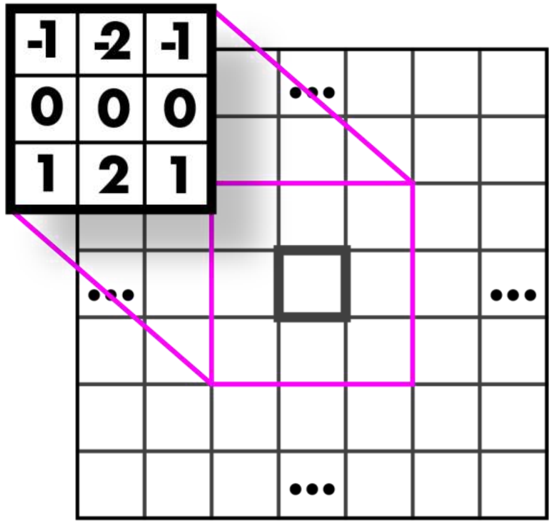
Guess those kernels!

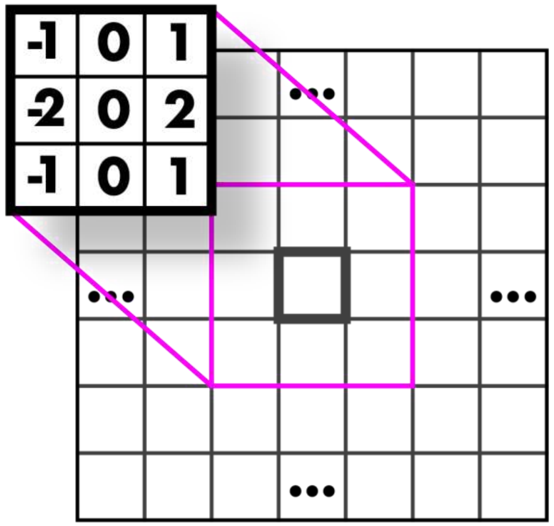




Sobel Kernels: edges and...

applied to grayscale and thresholded

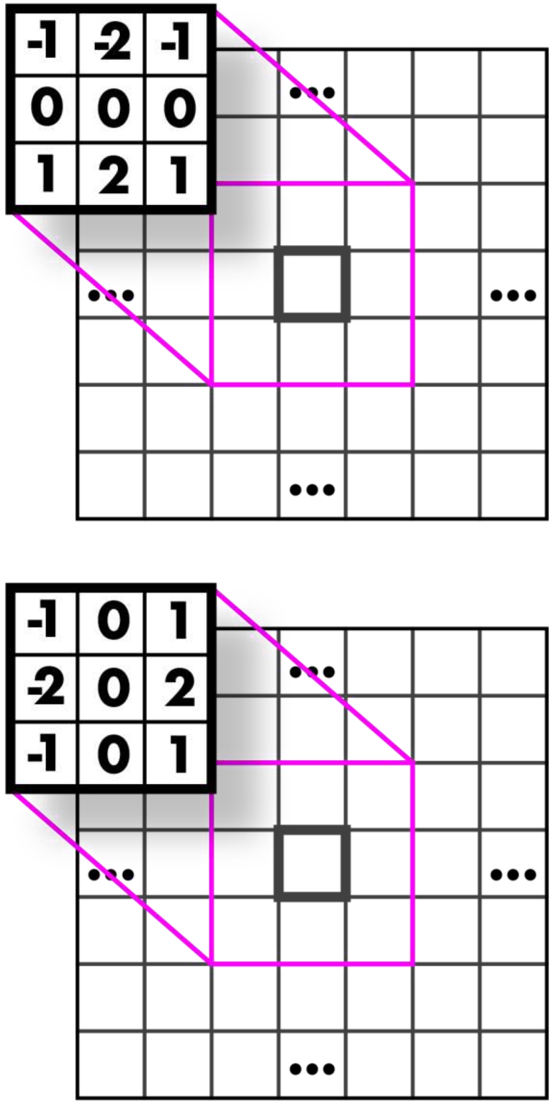


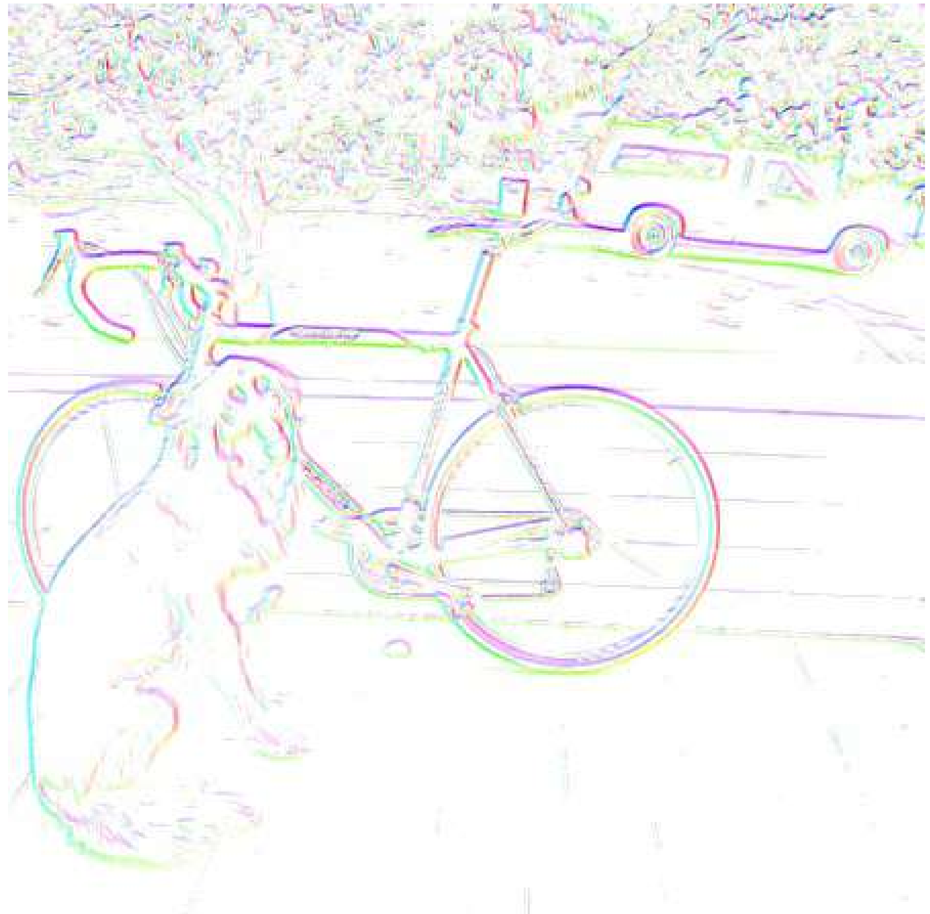


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Sobel Kernels: edges and gradient!





Sobel Kernels: edges and gradient!

This visualization is showing the magnitude and direction of the gradient

And so much more!!

- Next time
  - Edges
  - Features

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 6: Edges

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Averaging vs Interpolation
- Systems - filters
- Convolution
  - Box Filter
  - Gaussian
  - Cross correlation vs Convolution
- Examples of filters

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

[material based on Joseph Redmon's course]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What can we do with convolutions

Mathematically: nice linear properties

- Commutative
  - A*B = B*A
- Associative
  - A*(B*C) = (A*B)*C
- Distributes over addition
  - A*(B+C) = A*B + A*C
- Plays well with scalars
  - x(A*B) = (xA)*B = A*(xB)

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

What can we do with convolutions

This means some convolutions decompose:

- 2d gaussian is just composition of 1d gaussians
  - Faster to run 2 1d convolutions

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What can we do with convolutions

- Blurring
- Sharpening
- Edges
- Features
- Derivatives
- Super-resolution
- Classification
- Detection
- Image captioning
- ...

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What can we do with convolutions

- Blurring
- Sharpening
- Edges
- Features
- Derivatives
- Super-resolution
- Classification
- Detection
- Image captioning
- ...

A big part of computer vision is **convolutions**

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

So what can we do with these convolutions anyway?

- Blurring
- Sharpening
- Edges
- Features
- Derivatives
- Super-resolution
- Classification
- Detection
- Image captioning
- ...

A big part of computer vision is **convolutions**

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

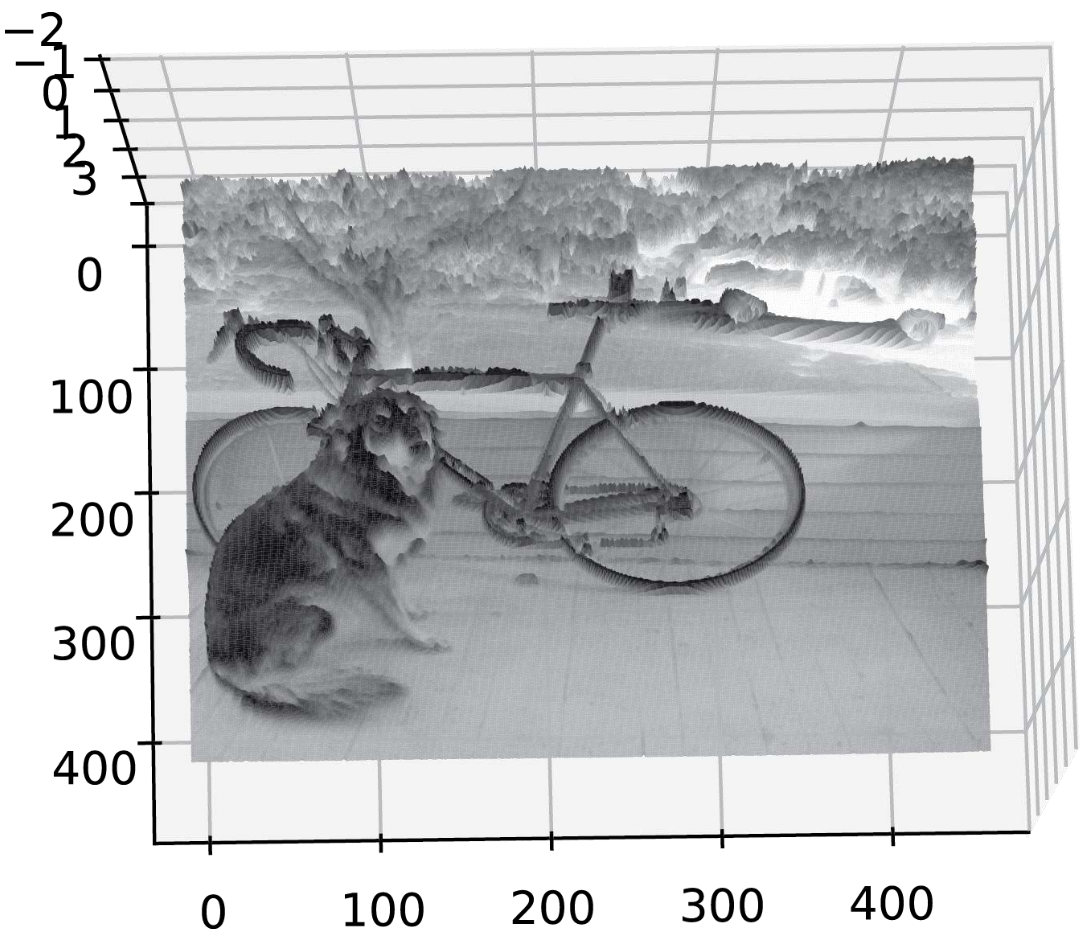
- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

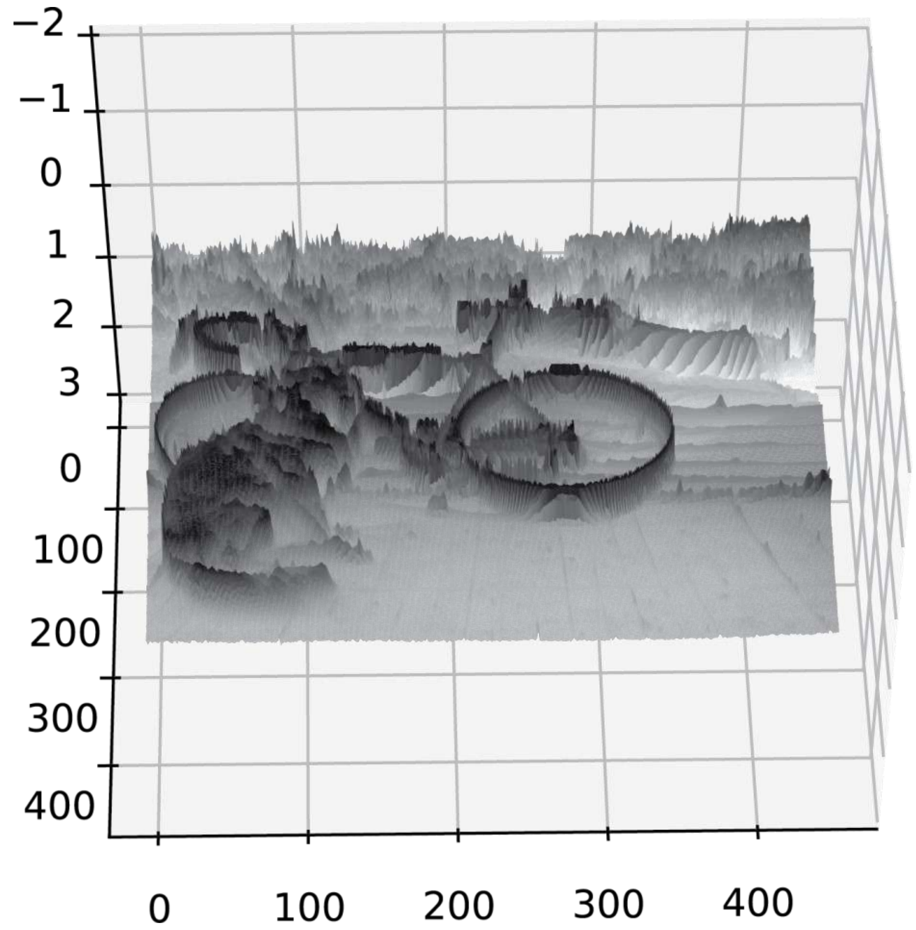
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What’s an edge?

- Image is a function
- Edges are rapid changes in this function



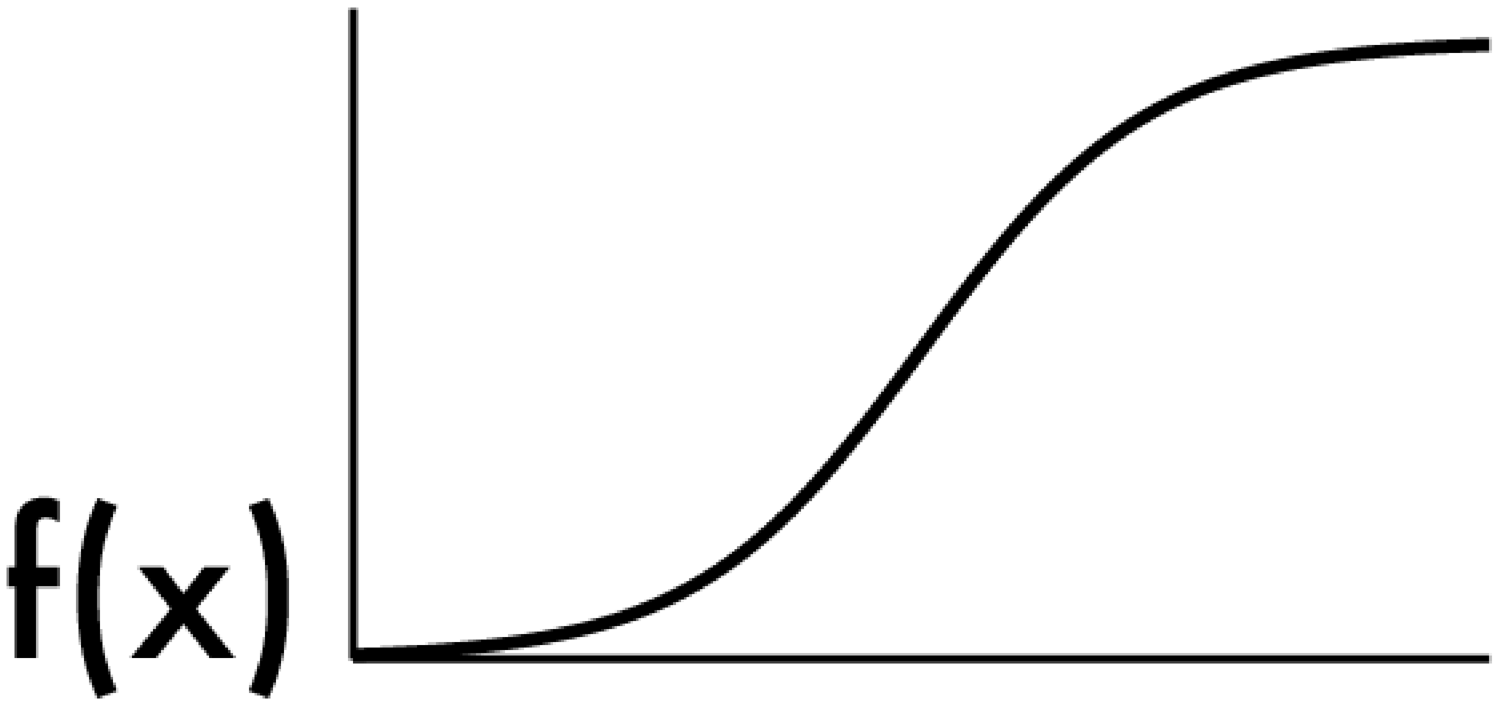


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What's an edge?

- Image is a function
- Edges are rapid changes in this function

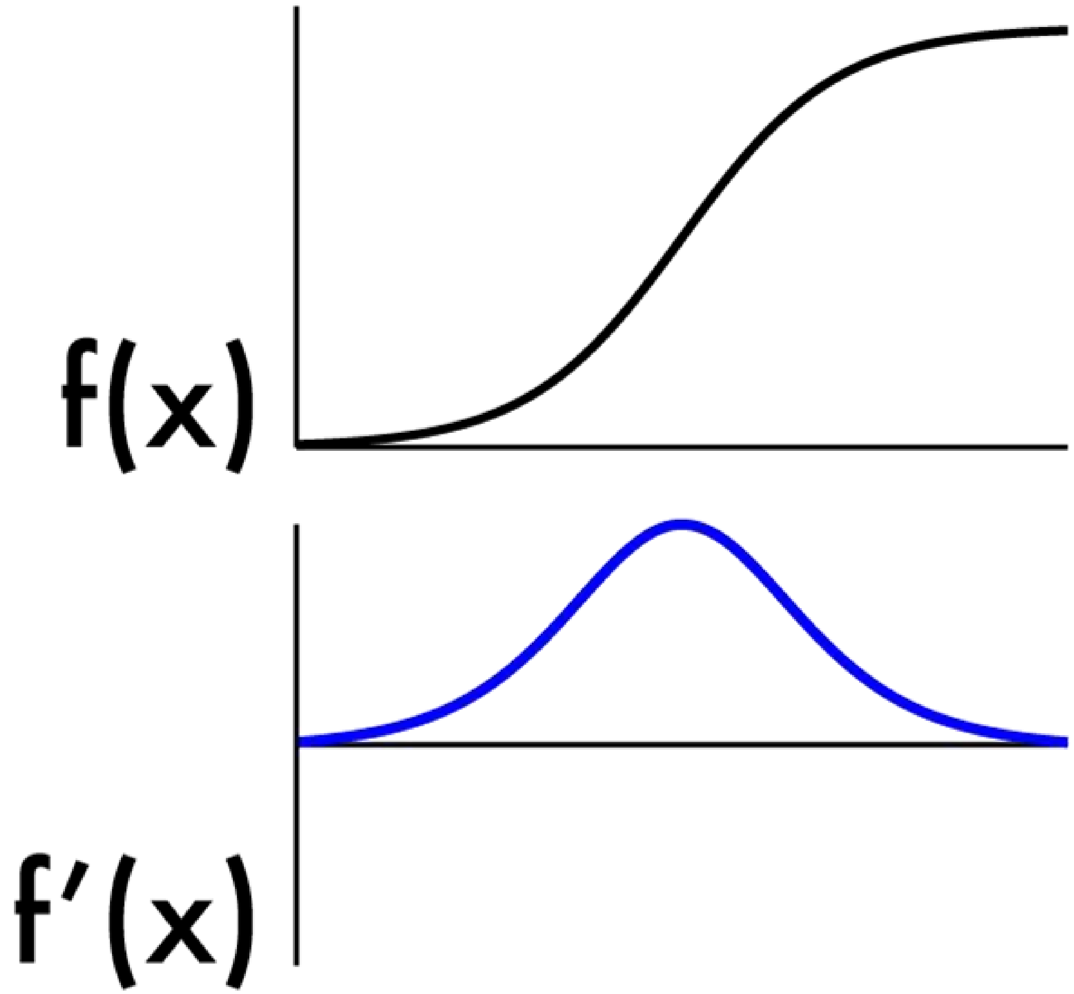


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Finding edges

- Could take derivative
- Edges = high response



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image derivatives

- Recall:
  - $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$
- We don't have an "actual" function, must estimate
- Possibility: set h = 1
- What will that look like?

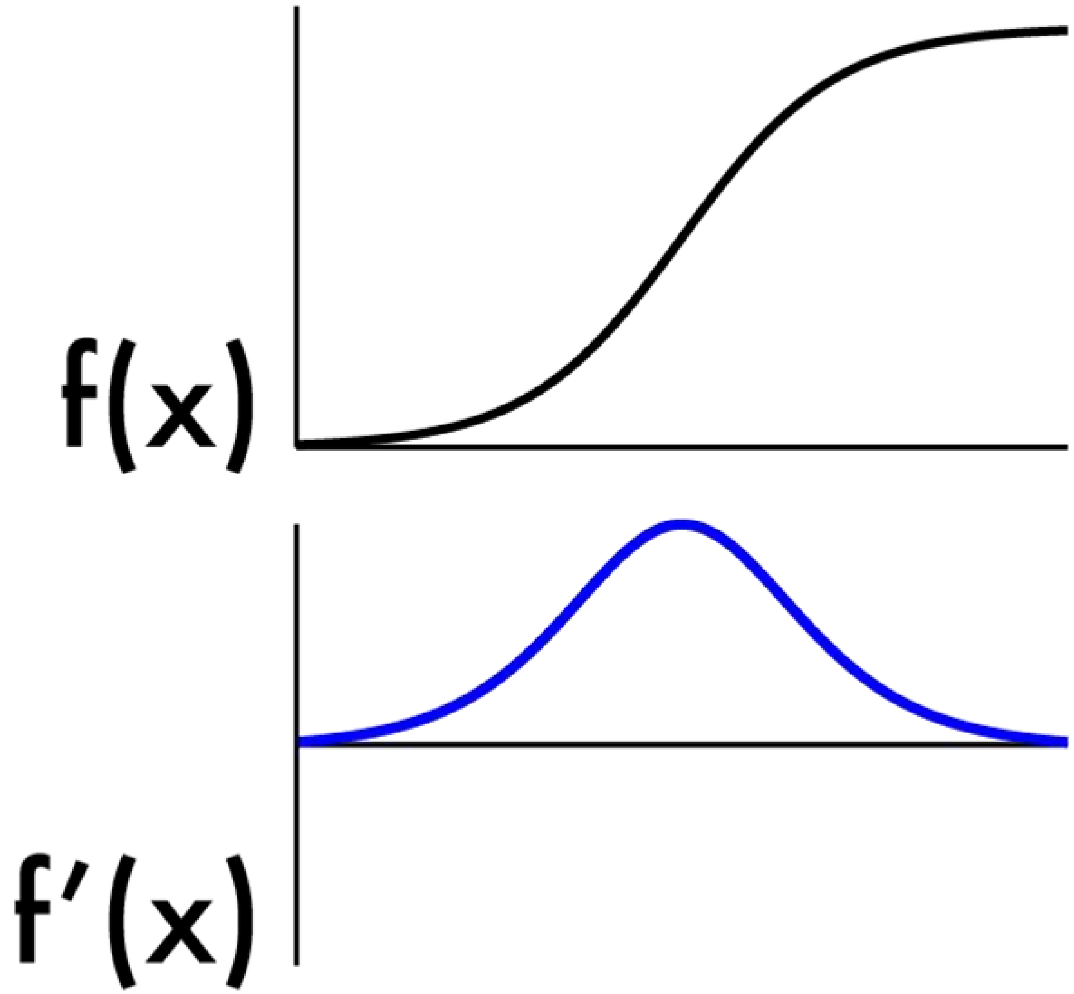
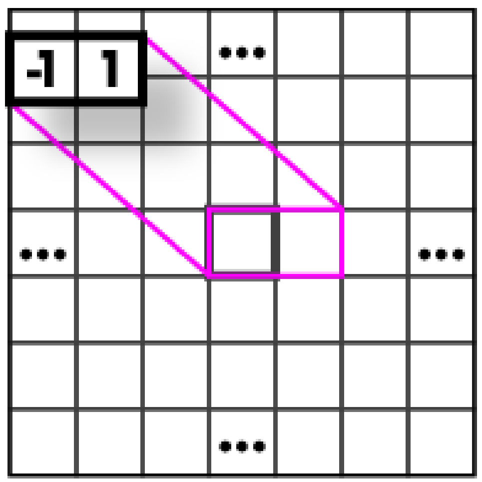


Image derivatives

- Recall:
  - $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$
- We don't have an "actual" function, must estimate
- Possibility: set h = 1
- What will that look like?



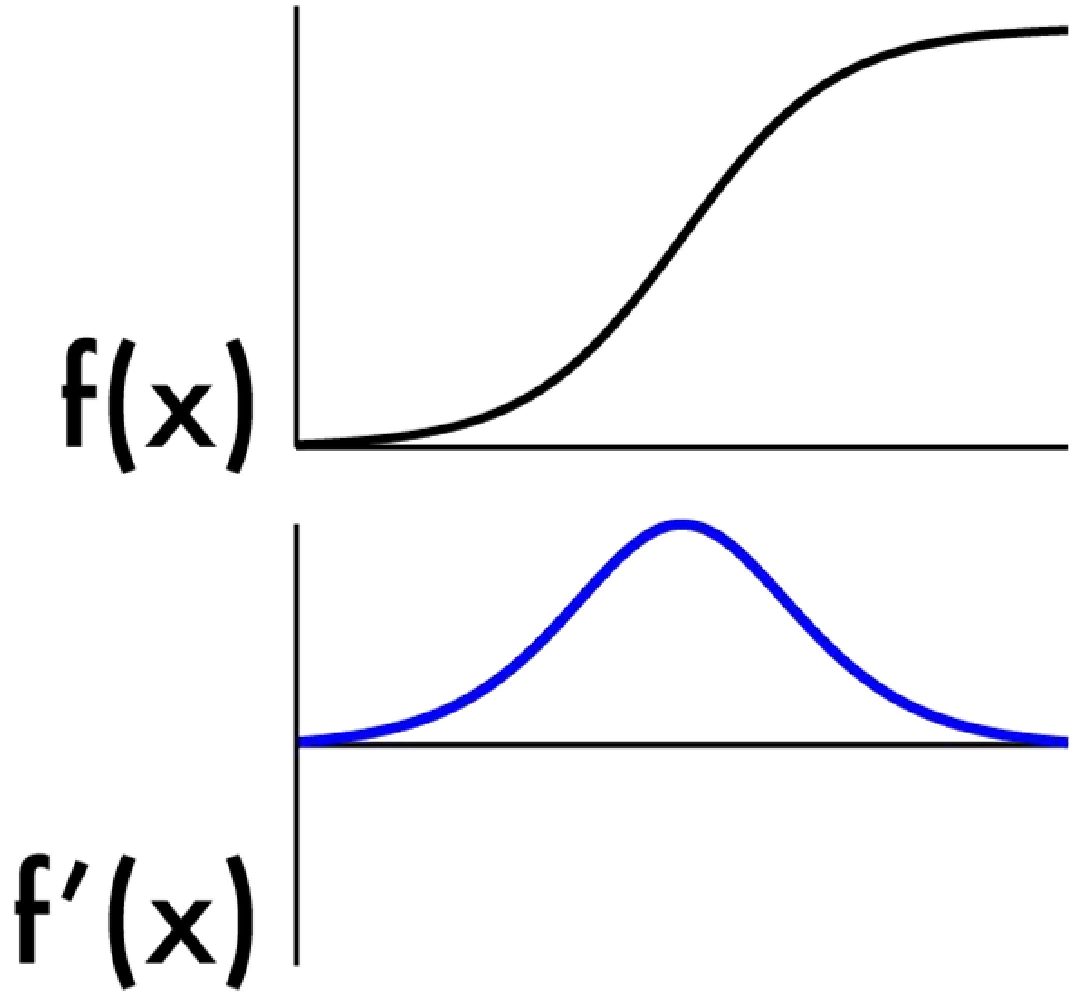
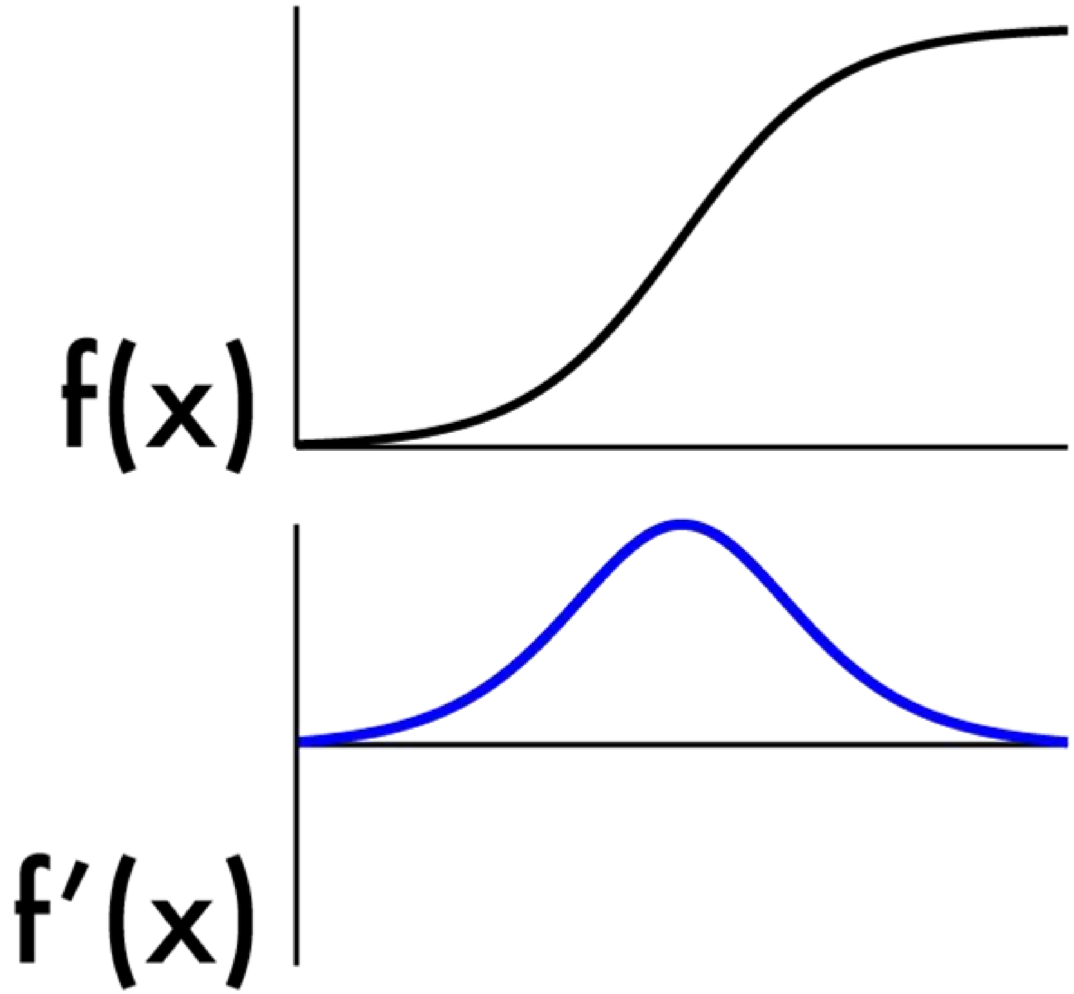


Image derivatives

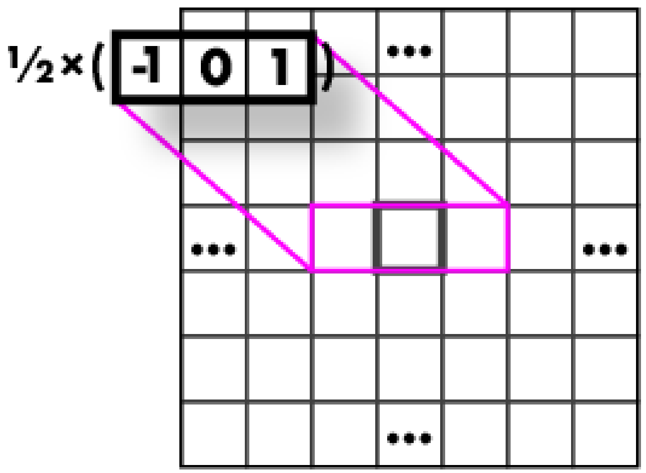
- Recall:
  - $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$
- We don't have an "actual" function, must estimate
- Possibility: set h = 2
- What will that look like?

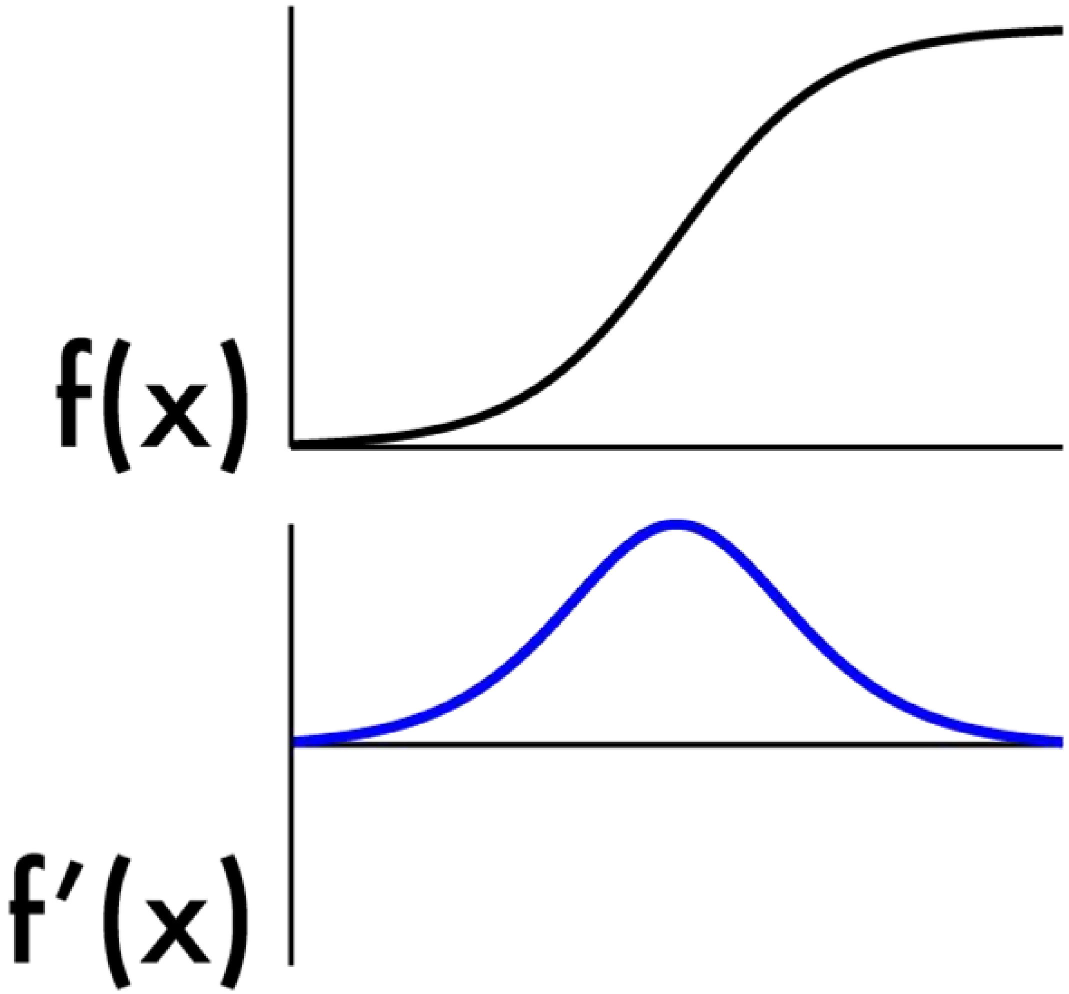


Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image derivatives

- Recall:
  - $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$
- We don't have an "actual" function, must estimate
- Possibility: set h = 2
- What will that look like?





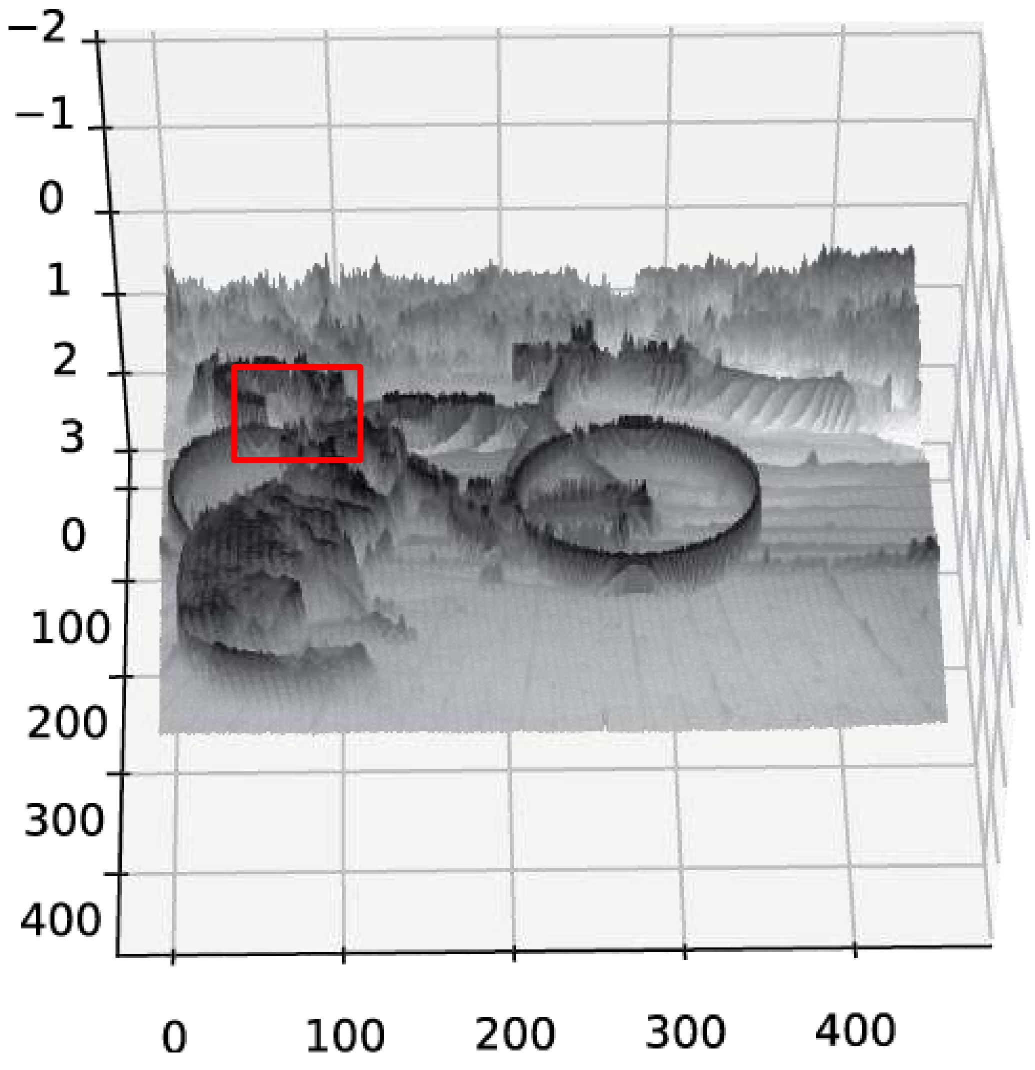
Today's Agenda

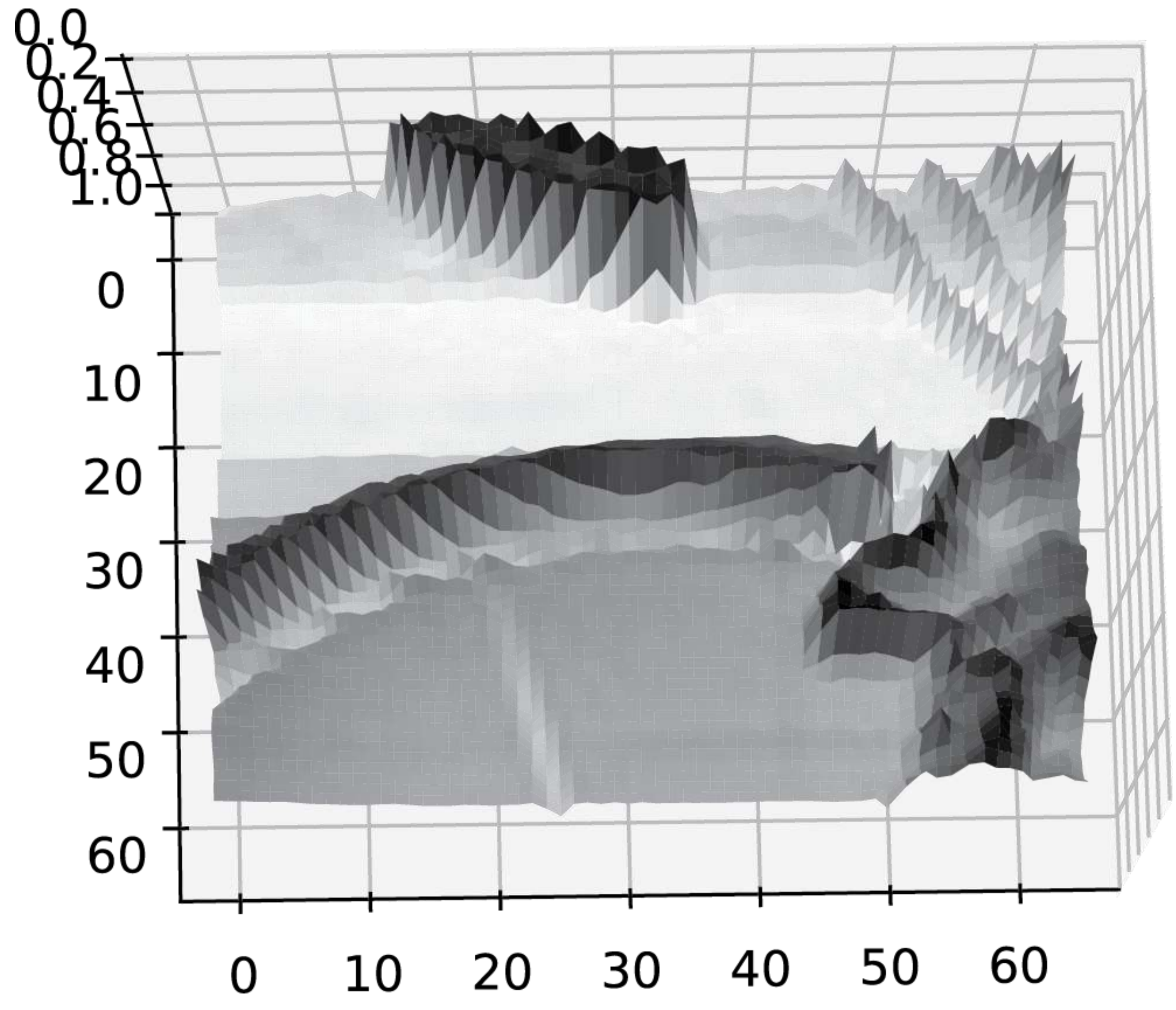
- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Images are noisy!

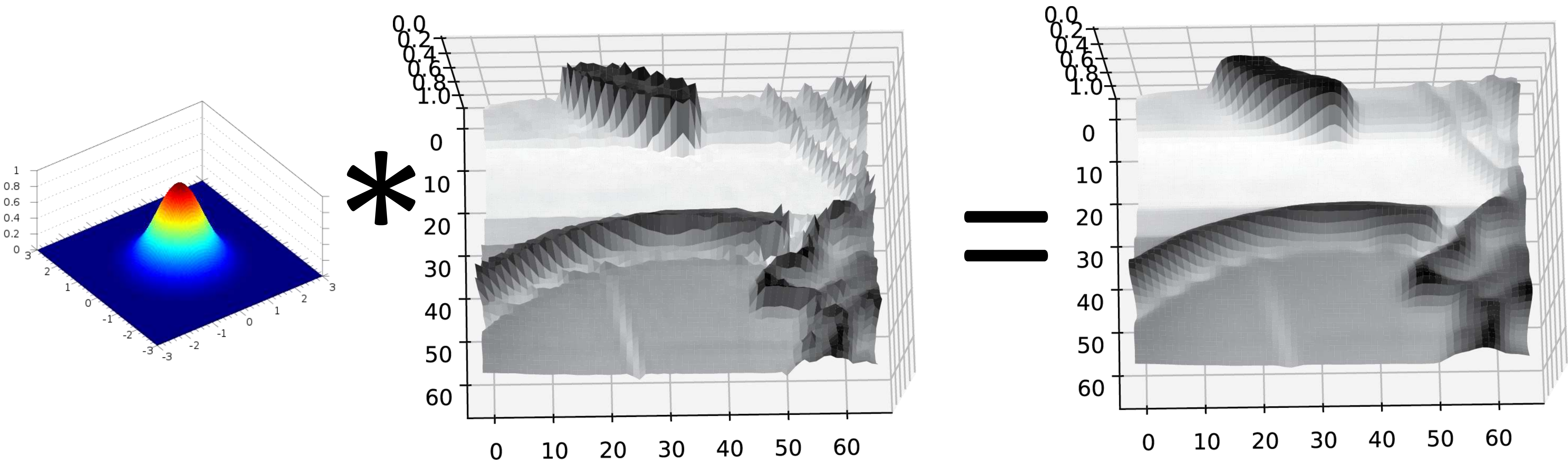




Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

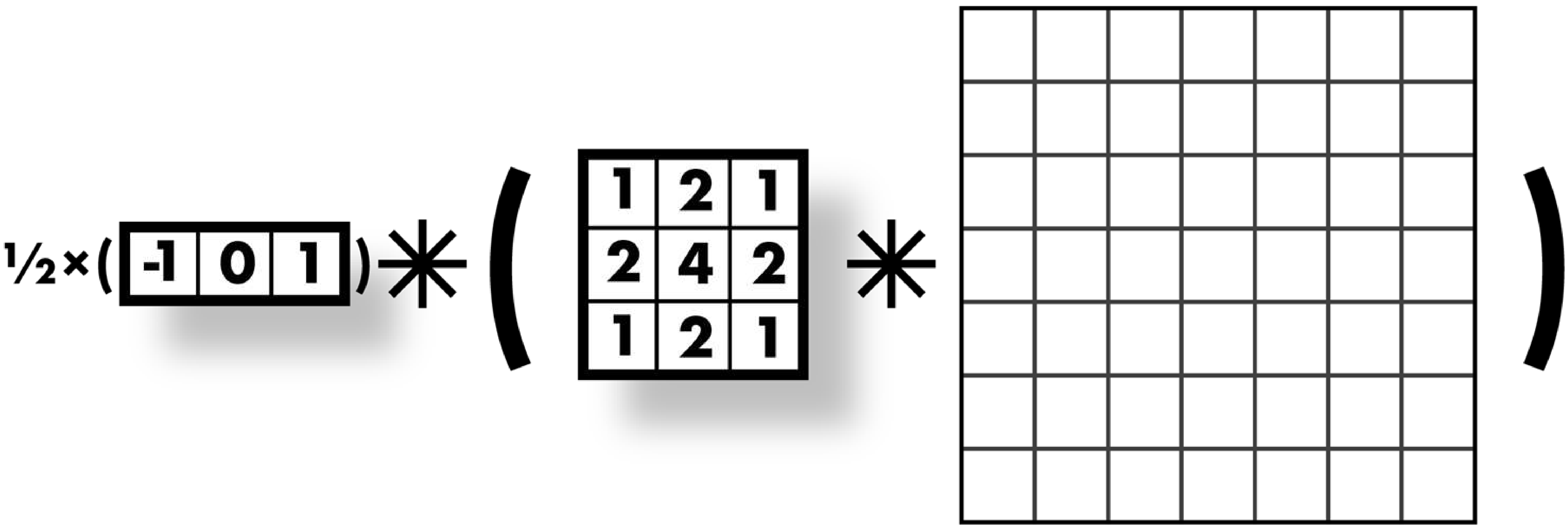
But we already know how to smooth



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

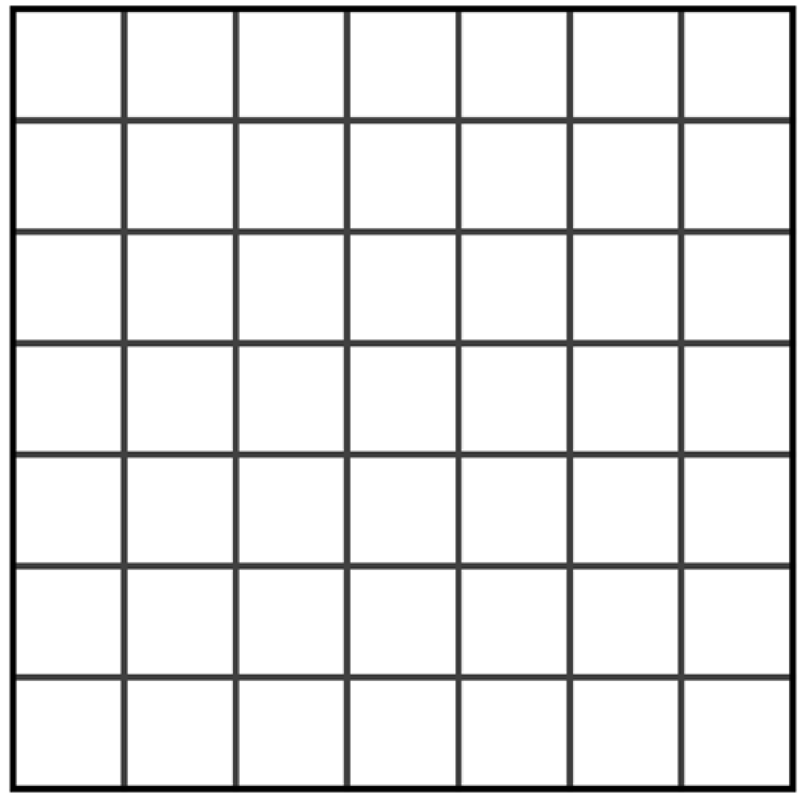
Smooth first, then derivative



This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Smooth first, then derivative

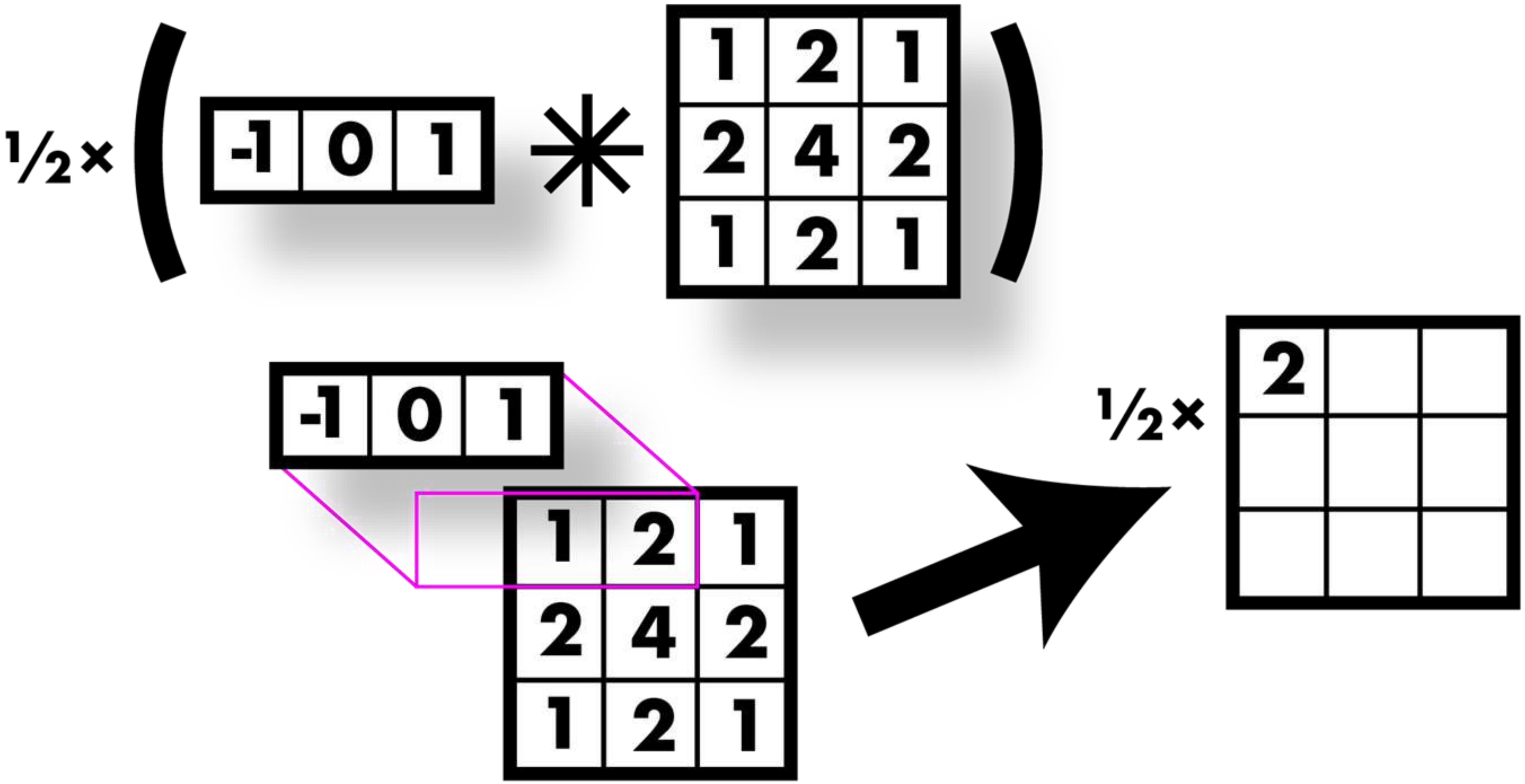
$$\frac{1}{2}\times\left(\begin{bmatrix}-1 & 0 & 1\end{bmatrix} * \begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix}\right) *$$



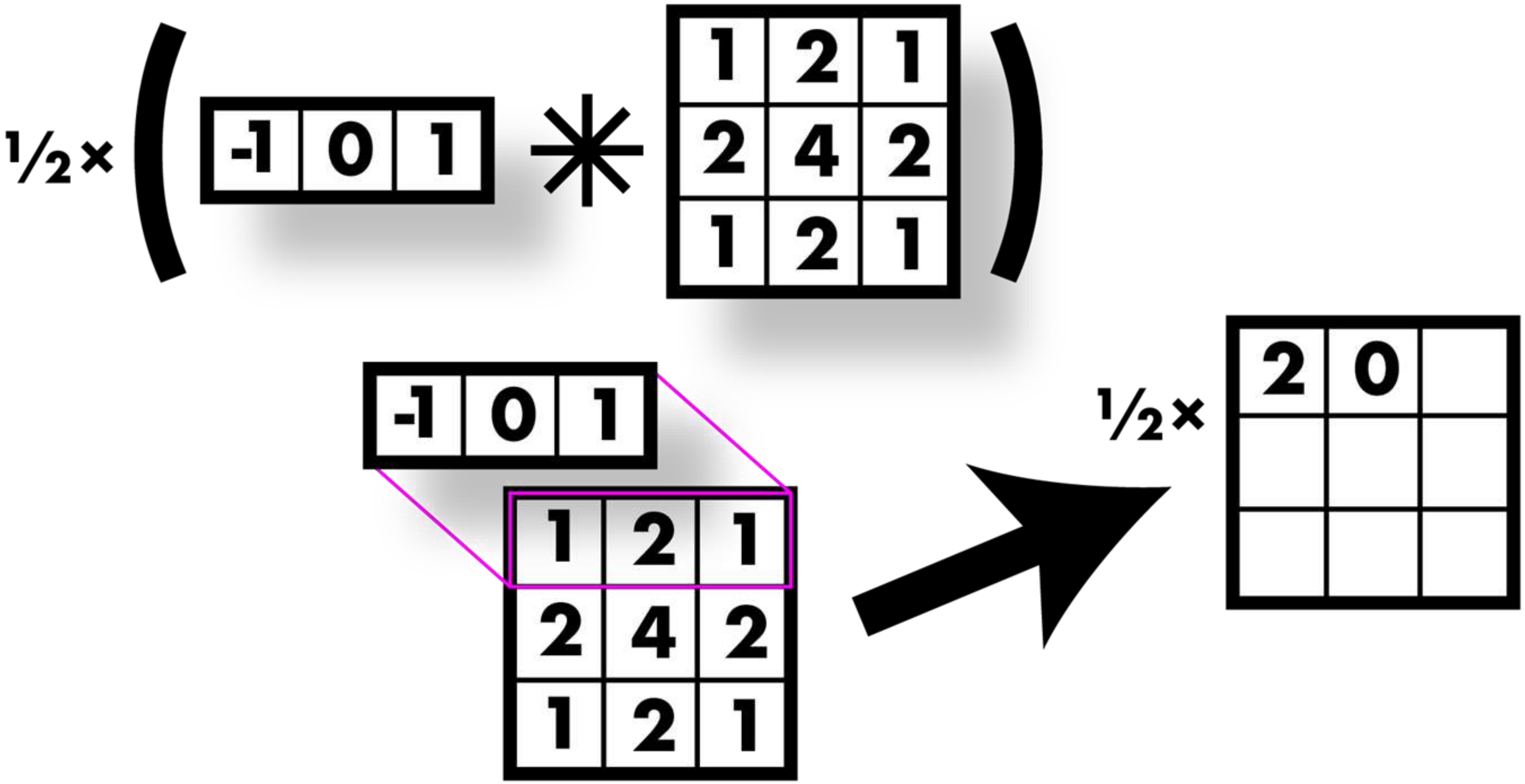
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Smooth first, then derivative

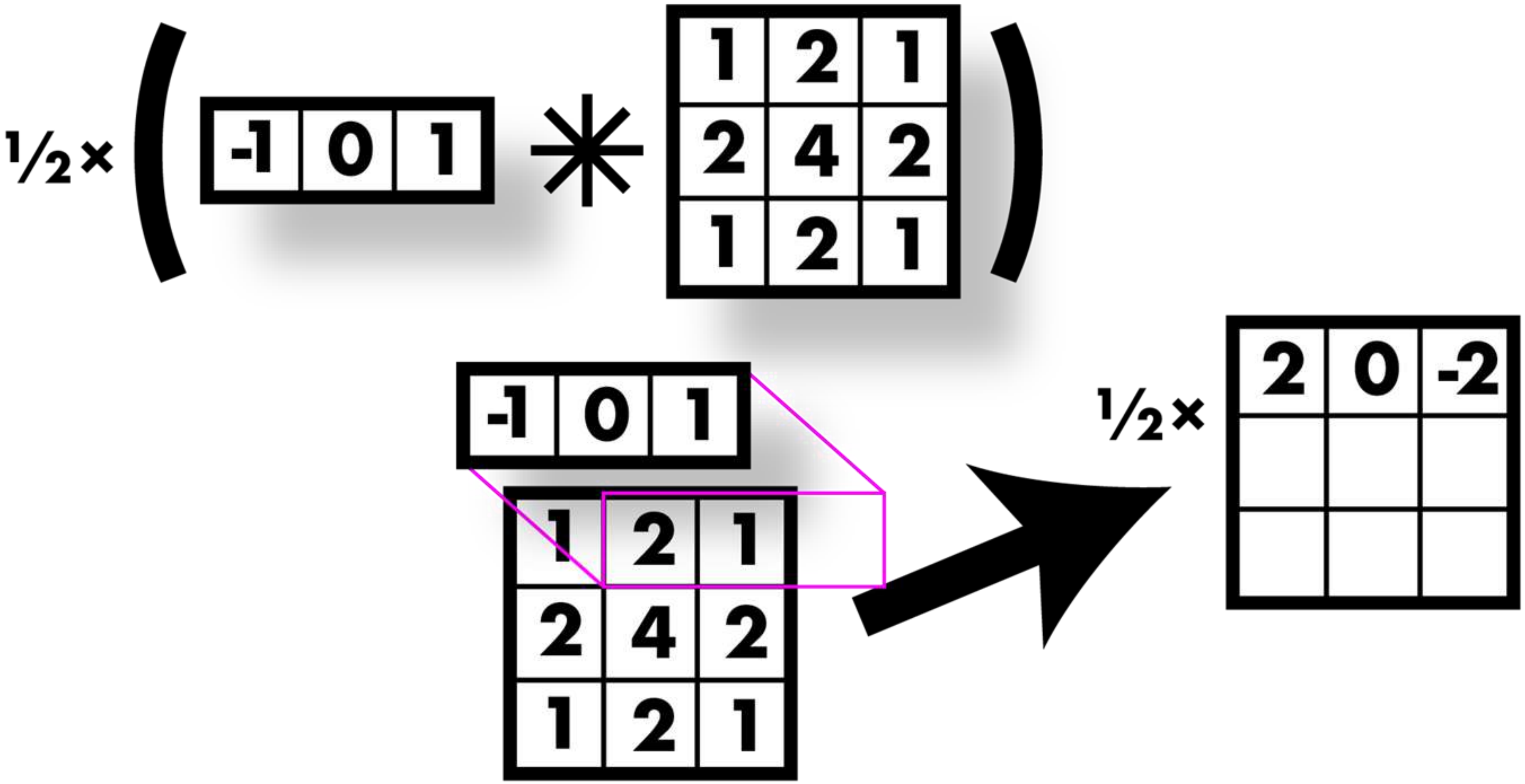


Smooth first, then derivative

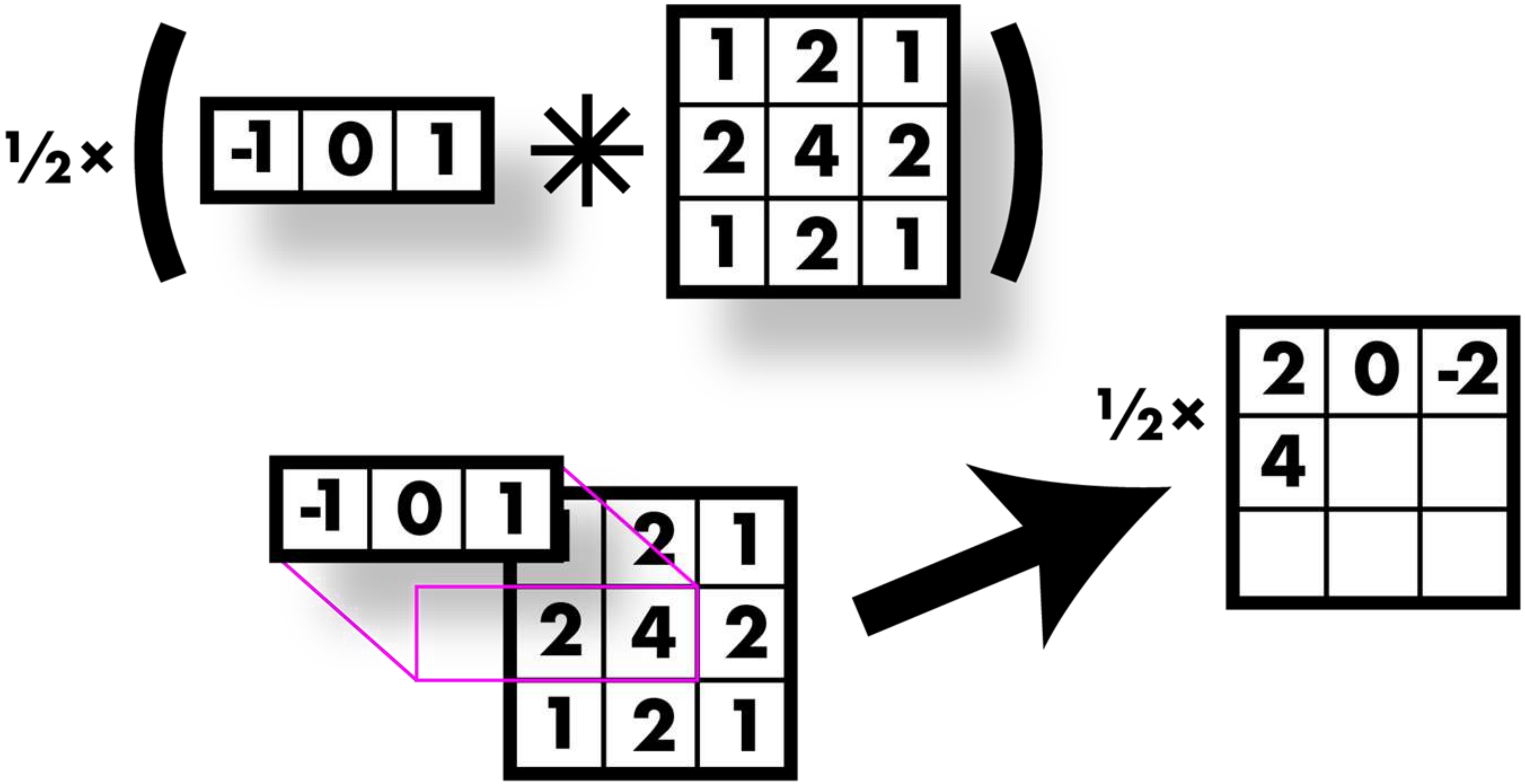


This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

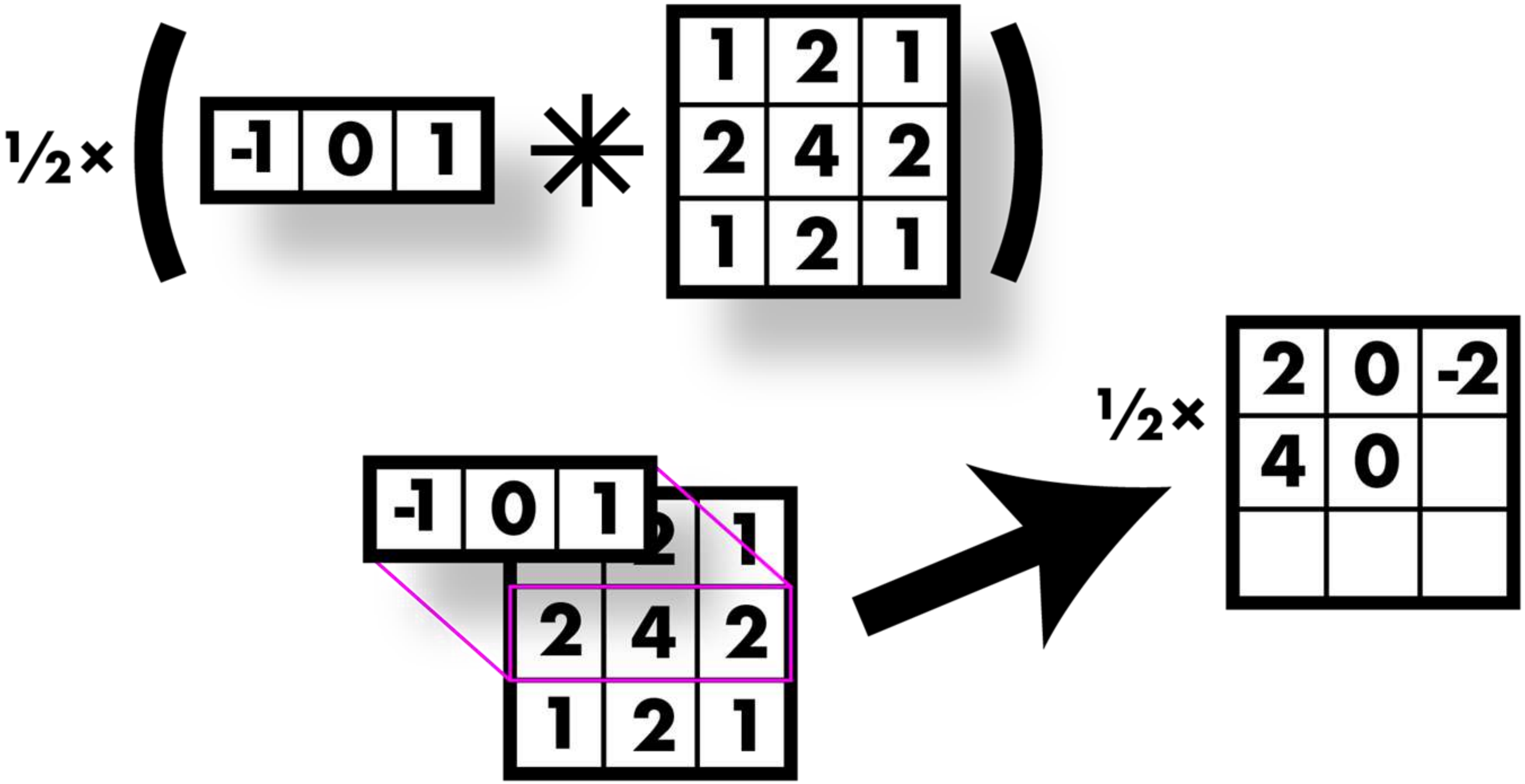
Smooth first, then derivative



Smooth first, then derivative

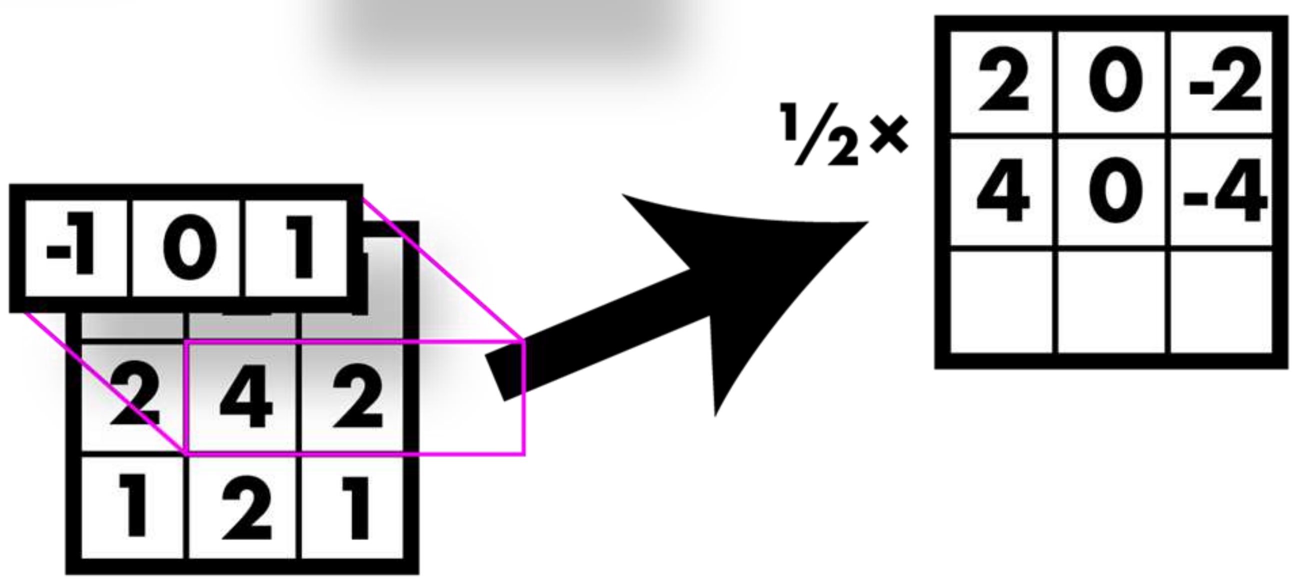


Smooth first, then derivative

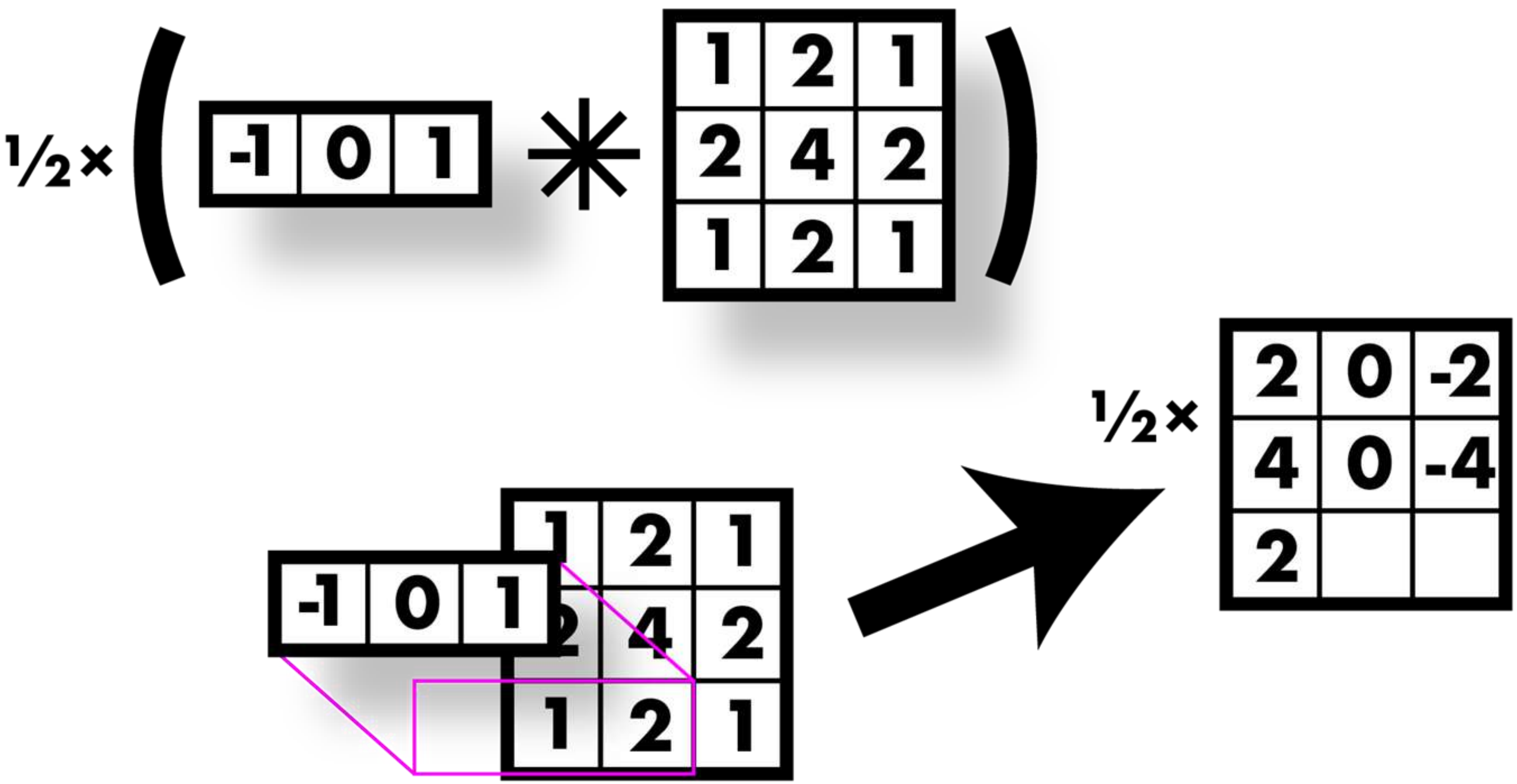


Smooth first, then derivative

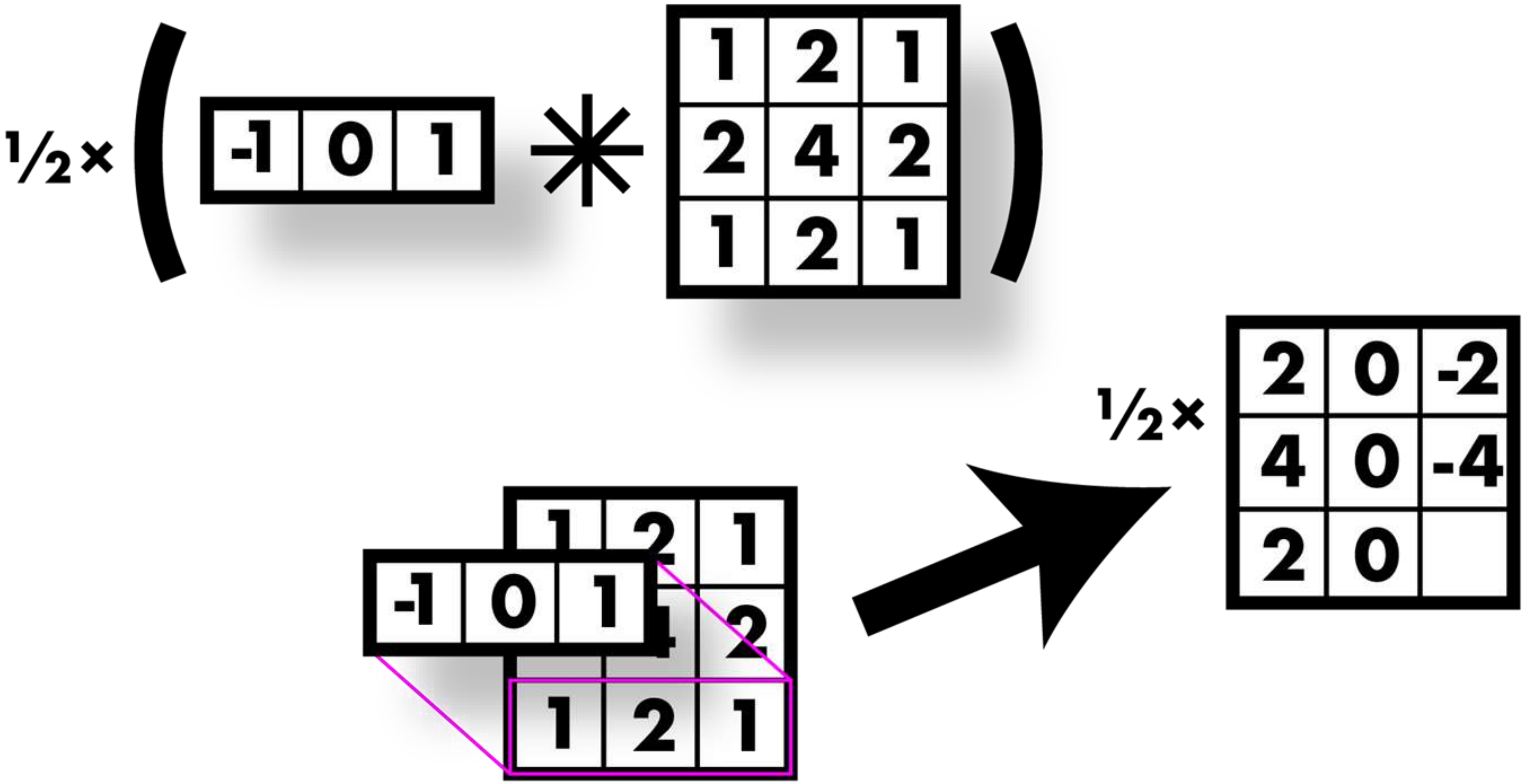
$$\frac{1}{2}\times\left(\begin{bmatrix}-1 & 0 & 1\end{bmatrix} * \begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix}\right)$$



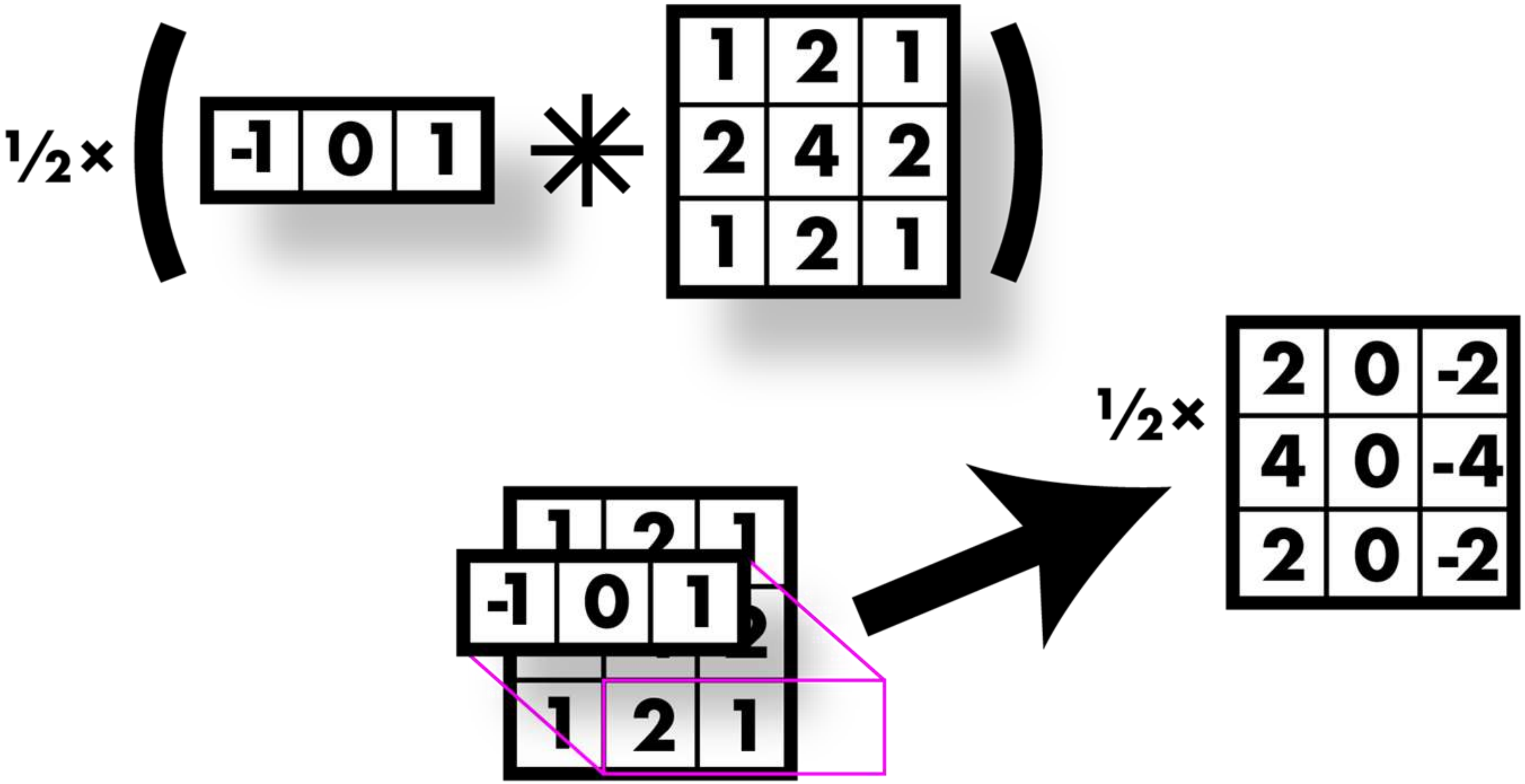
Smooth first, then derivative



Smooth first, then derivative

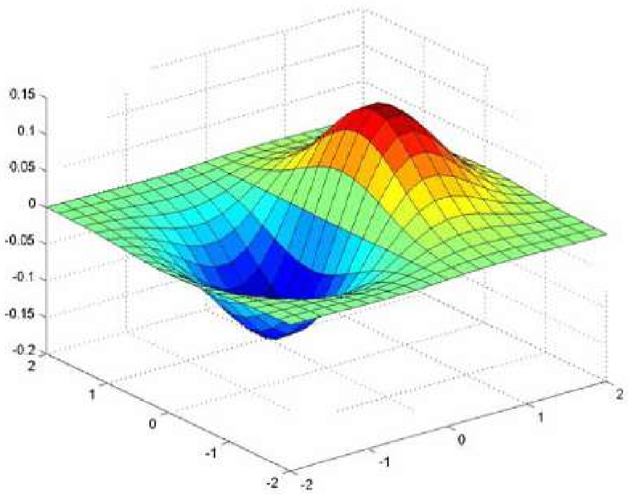


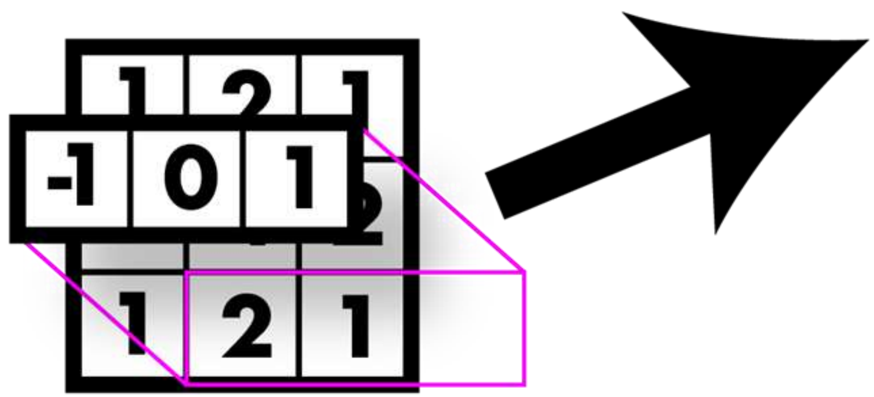
Smooth first, then derivative



Sobel filter! Smooth & derivative

$$\frac{1}{2}\times\left(\begin{bmatrix}-1 & 0 & 1\end{bmatrix} * \begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix}\right)$$

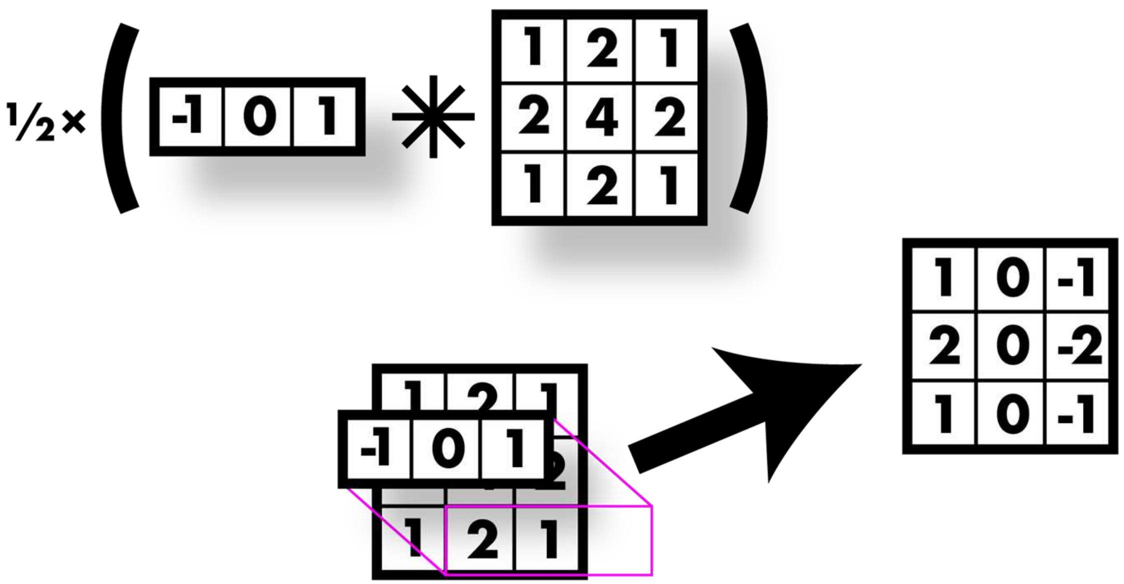


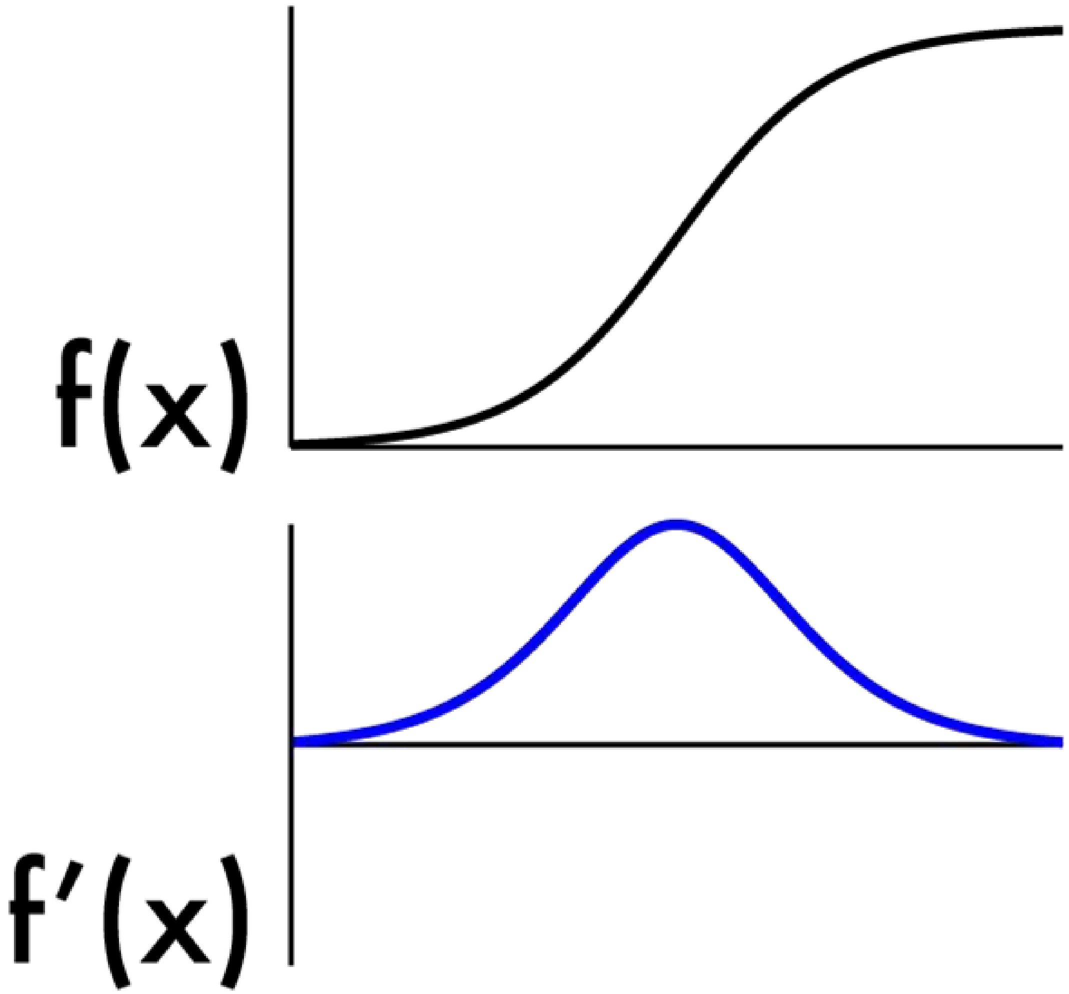


$$\begin{bmatrix}1 & 0 & -1\\ 2 & 0 & -2\\ 1 & 0 & -1\end{bmatrix}$$

Image derivatives

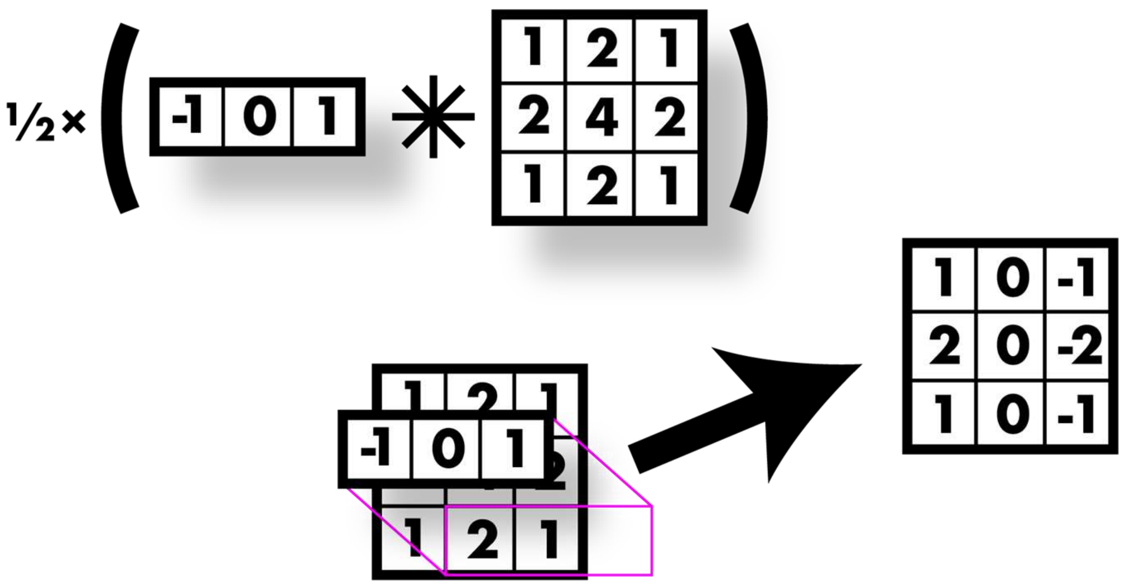
- Recall:
  - $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$
- Want smoothing too!

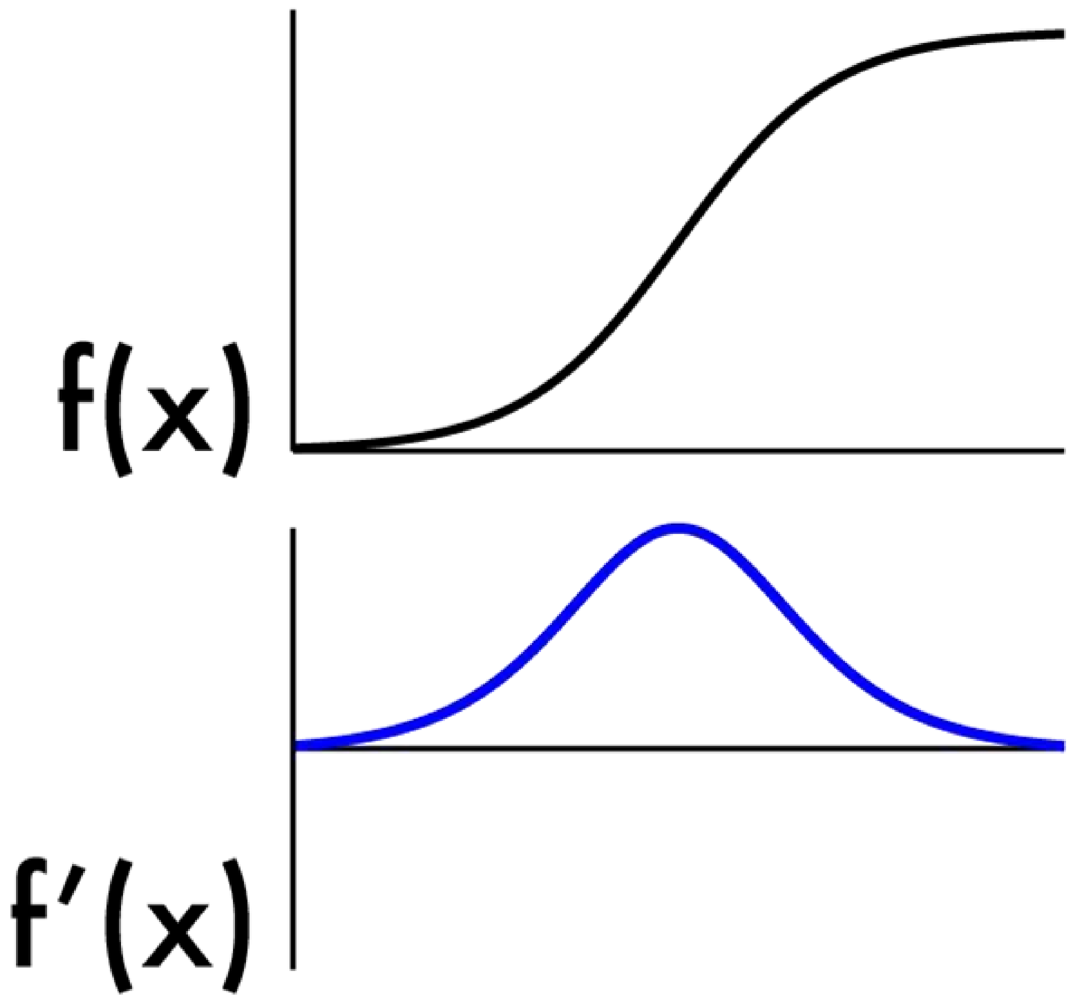




Finding edges

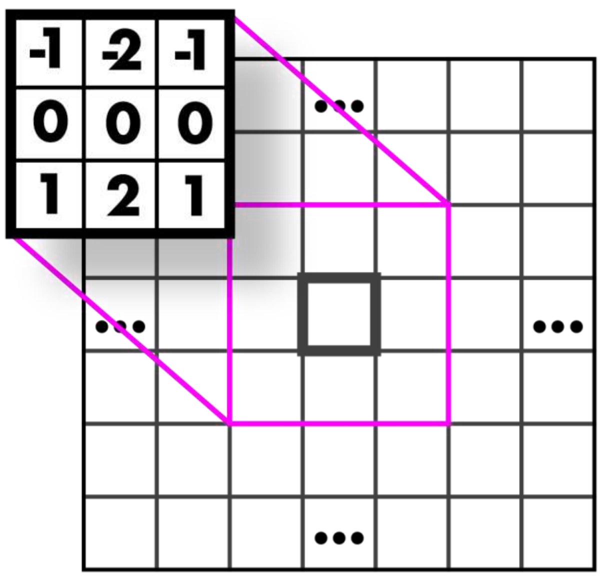
- Could take derivative
- Find high responses
- Sobel filters!
- What about y direction ?

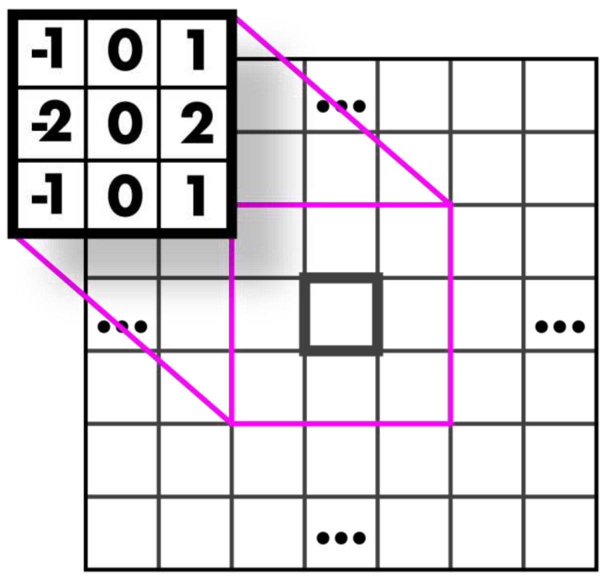


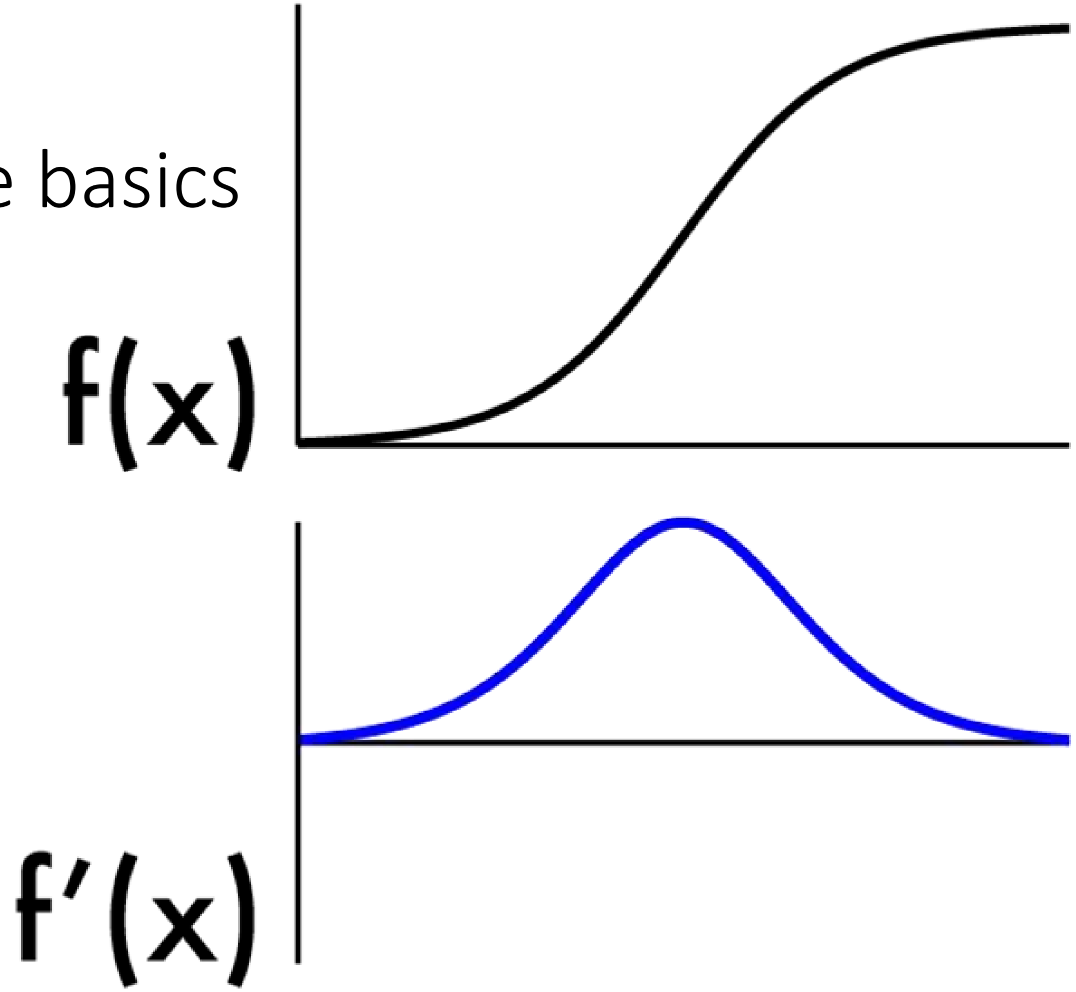


Finding edges

- Could take derivative
- Find high responses
- Sobel filters!
- Let's stop a moment and get some basics



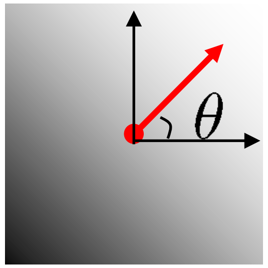




Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Simplest image gradient



$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$$

$$\frac{\partial f}{\partial x} = f(x+1, y) - f(x, y)$$

Likewise for df/dy

The **gradient direction** is $\theta = \tan^{-1}\left(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}\right)$

How does this relate to the direction of the edge? -Perpendicular

The *edge strength* is given by the **gradient magnitude**

$$\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

Sobel filters

$$g_x = \begin{array}{|c|c|c|} \hline 1 & 0 & -1 \\ \hline 2 & 0 & -2 \\ \hline 1 & 0 & -1 \\ \hline \end{array} \qquad g_y = \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -2 & -1 \\ \hline \end{array}$$

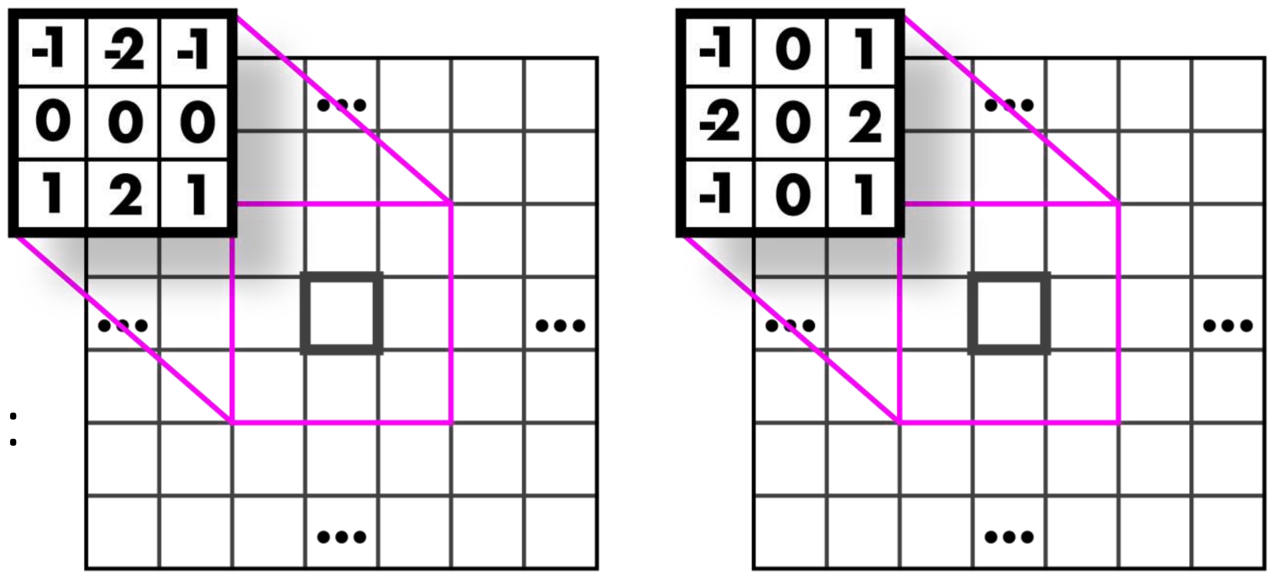
Magnitude:

$$g = \sqrt{g_x^2 + g_y^2}$$

Orientation:

$$\Theta = \tan^{-1}\left(\frac{g_y}{g_x}\right)$$

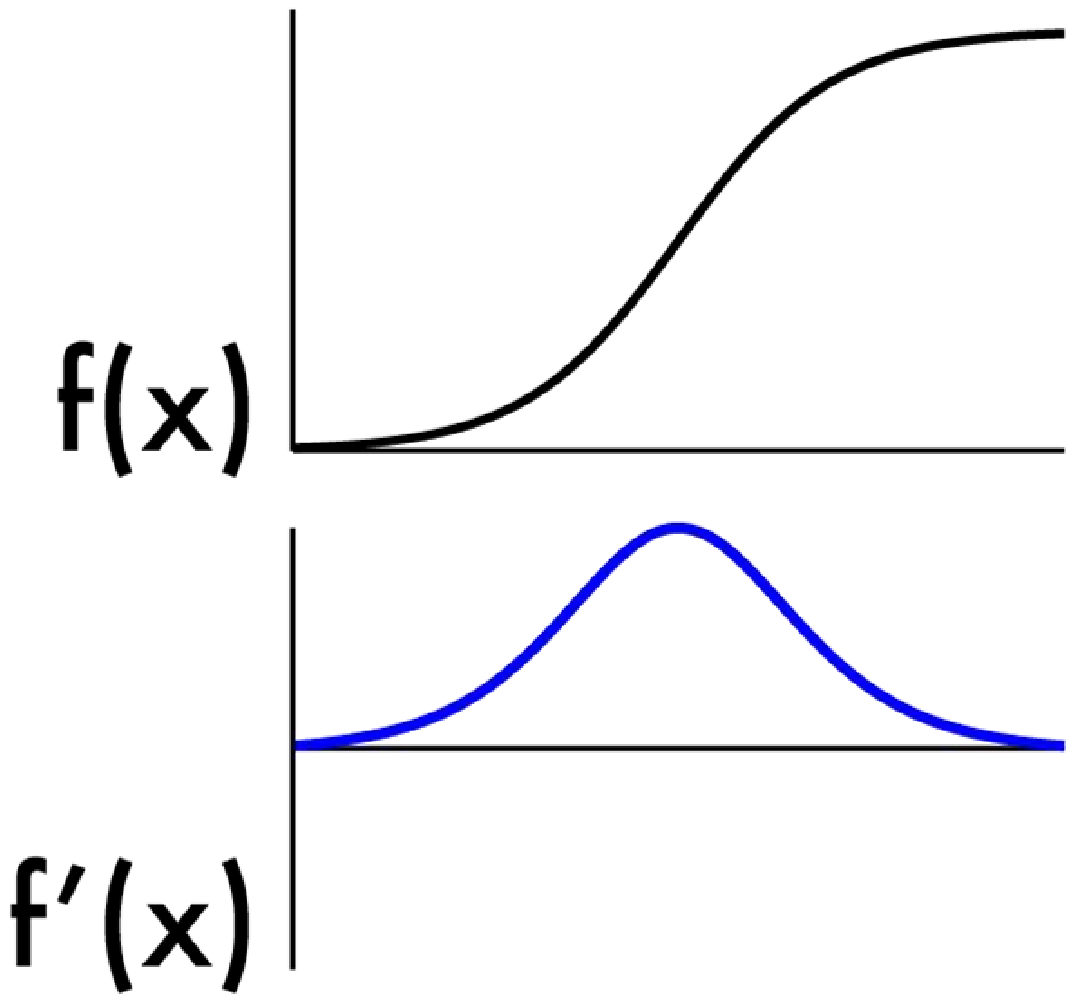
We can change the sign:



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

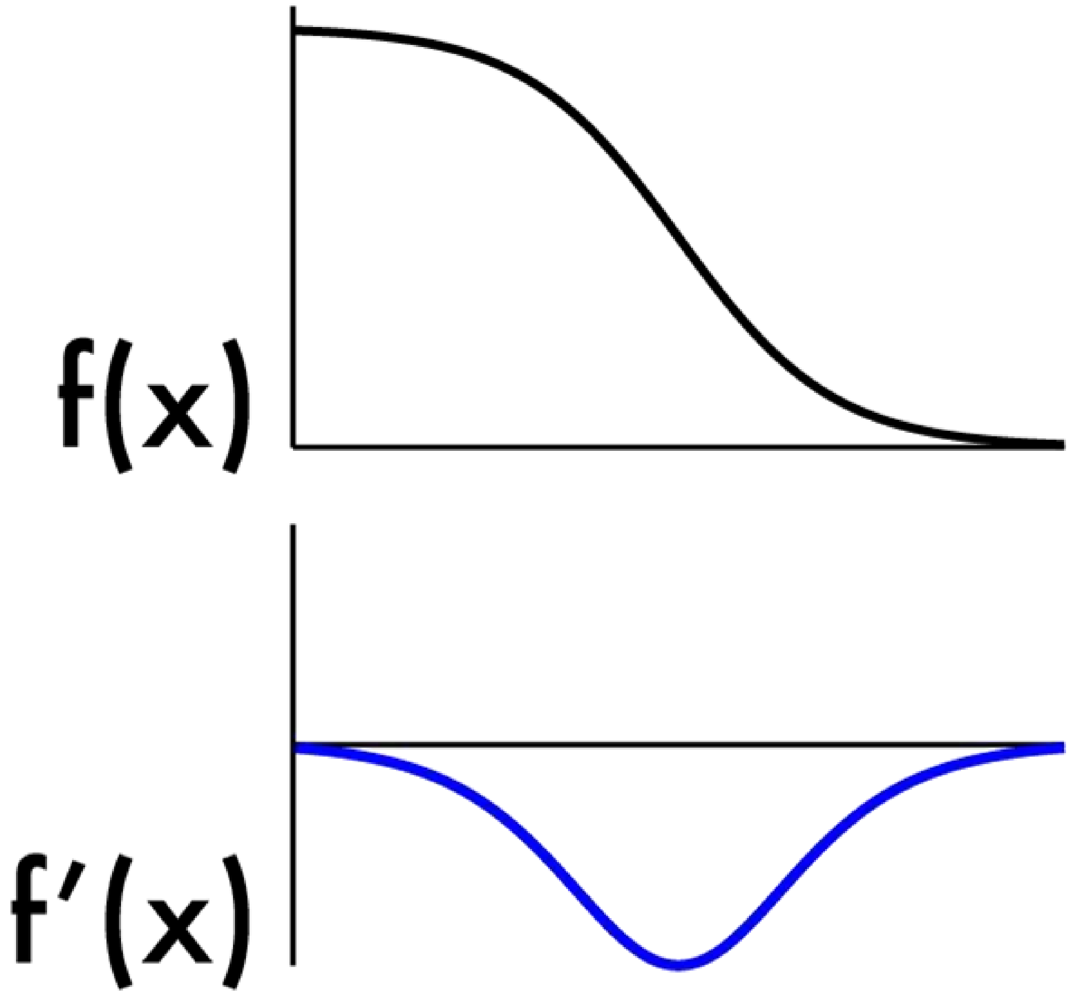
Finding edges

- Could take derivative
- Find high responses
- Sobel filters!
- But...



Finding edges

- Could take derivative
- Find high responses
- Sobel filters!
- But...
- Edges go both ways
- Want to find extrema



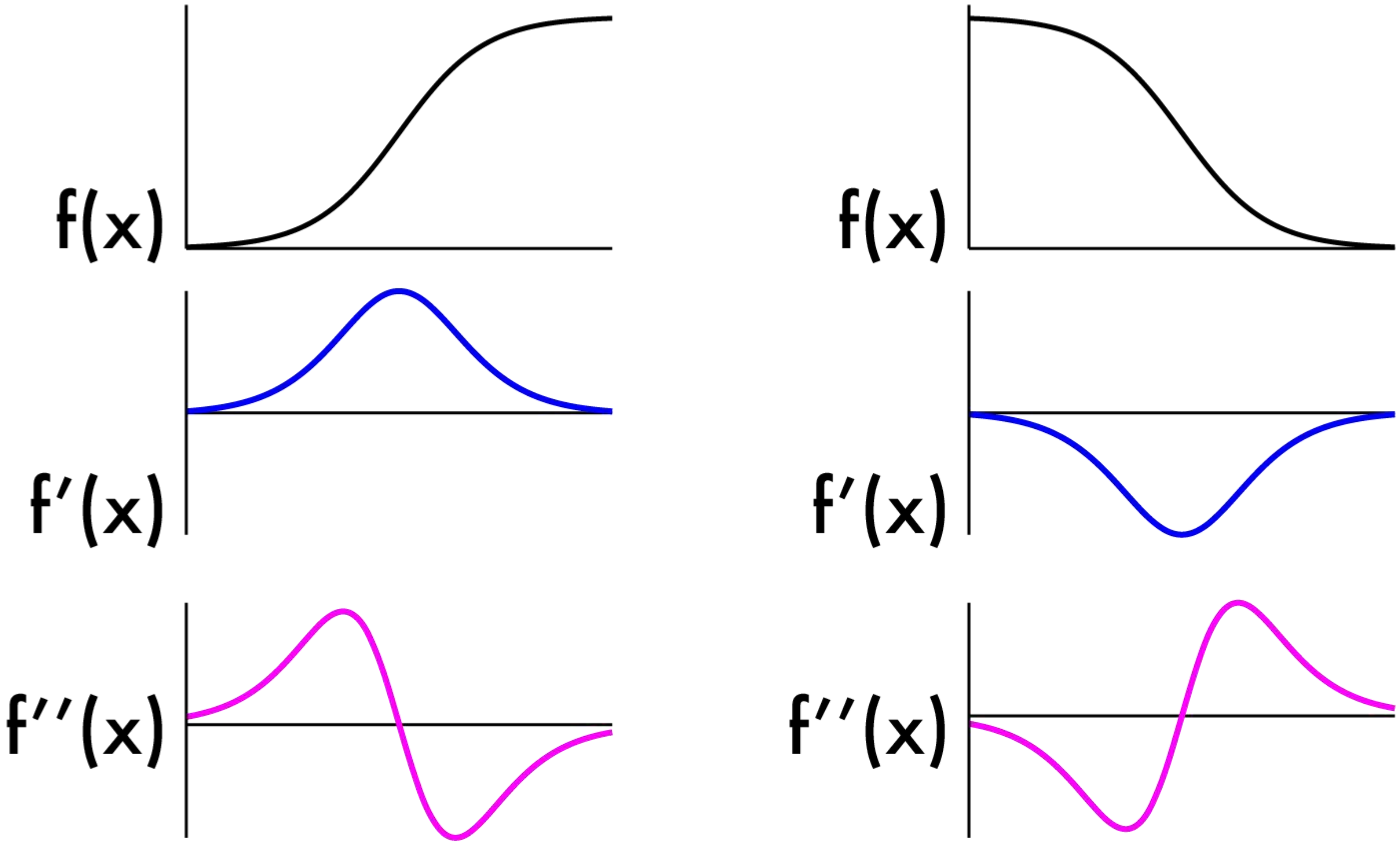
Today’s Agenda

- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

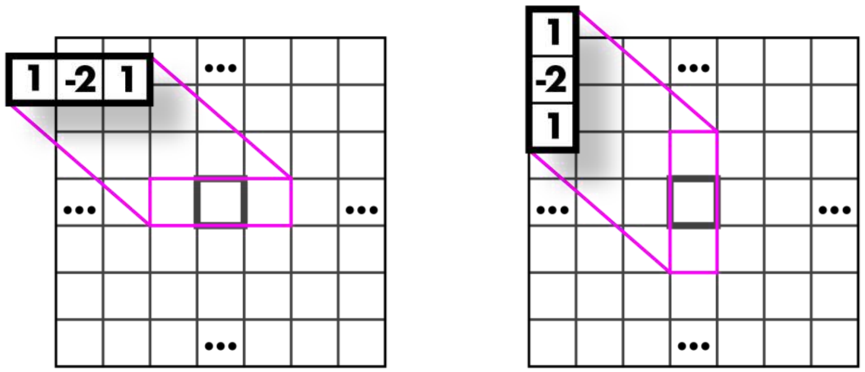
2nd derivative!

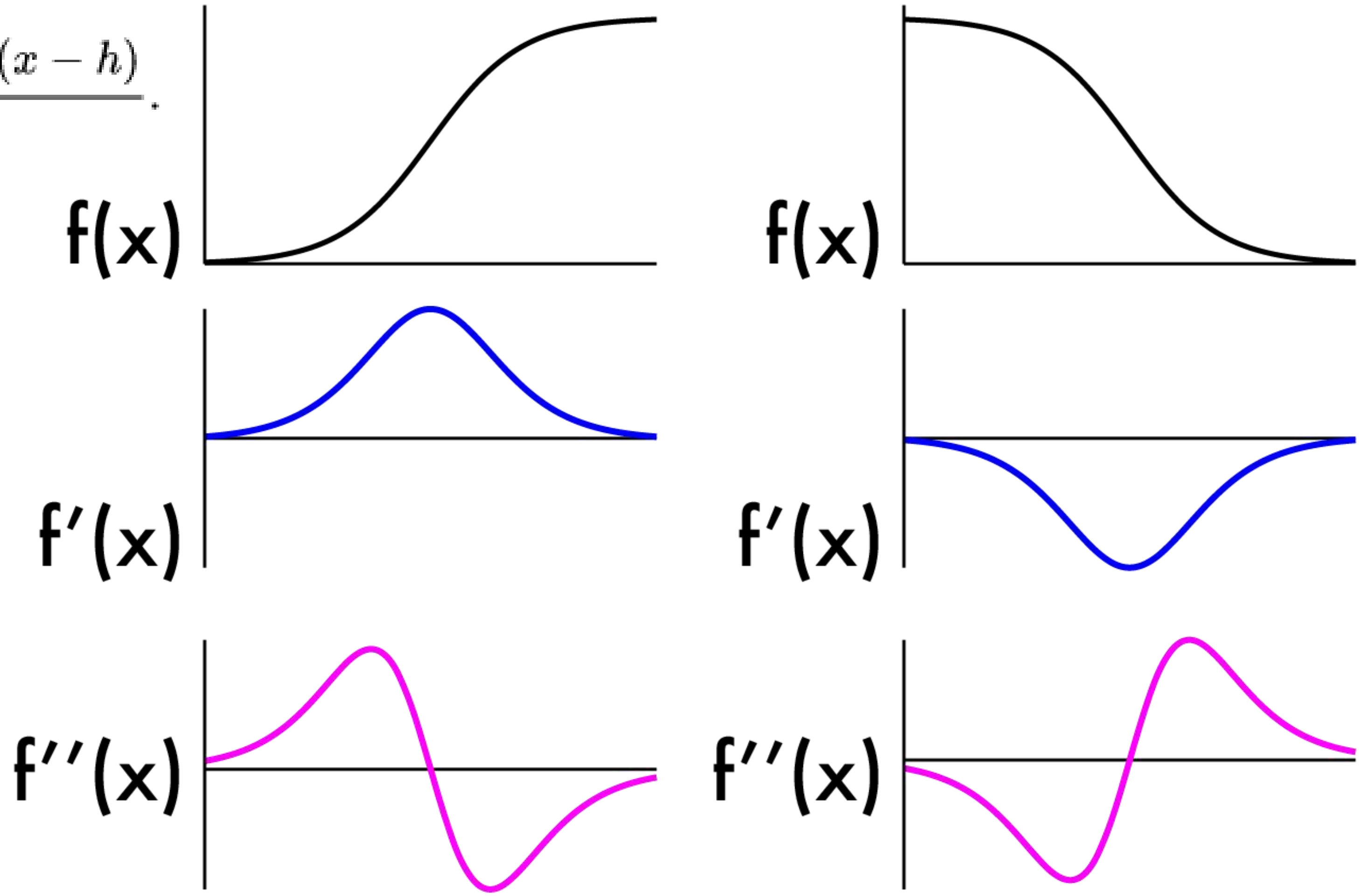
- Crosses zero at extrema



Laplacian (2nd derivative)!

- Crosses zero at extrema
- Recall:
  - $$f''(x) = \lim_{h \to 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}.$$
- Laplacian:
  - $$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$
- Again, have to estimate f''(x):





Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Laplacians

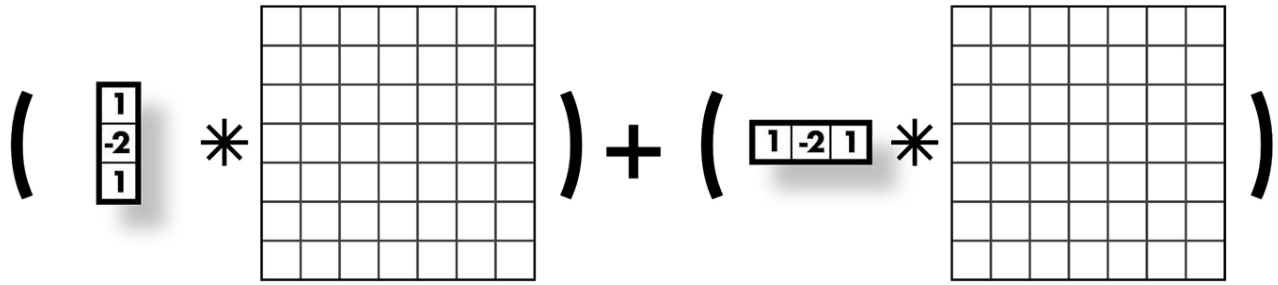
- Laplacian:
  - $$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

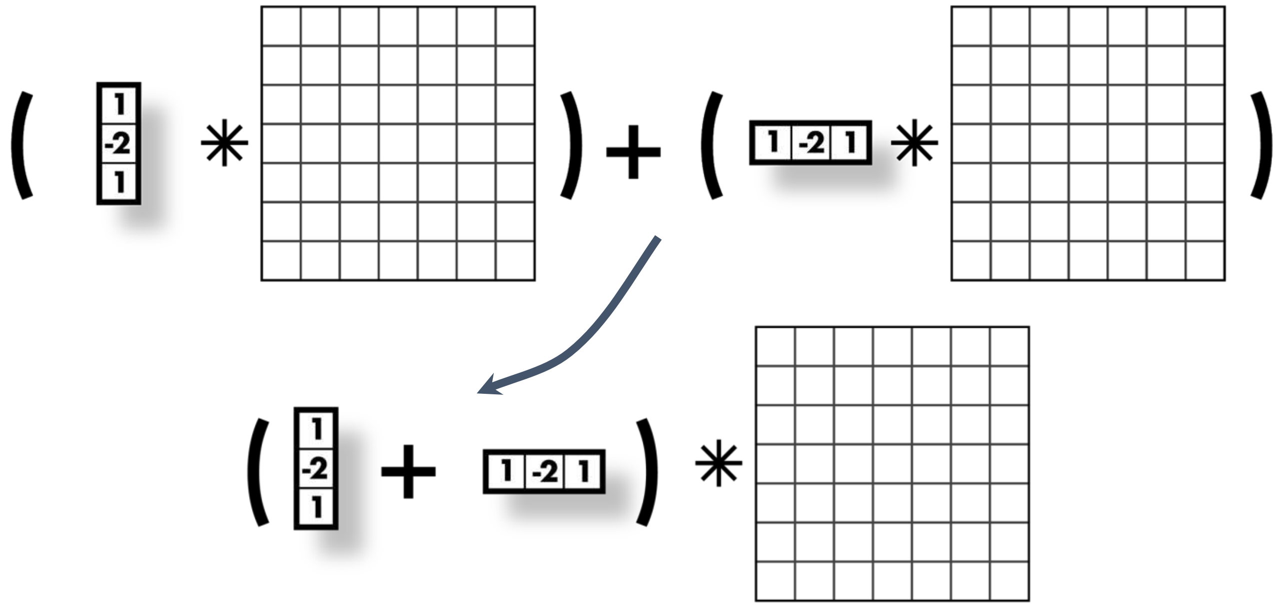
Laplacians

- Laplacian:
  - $$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$



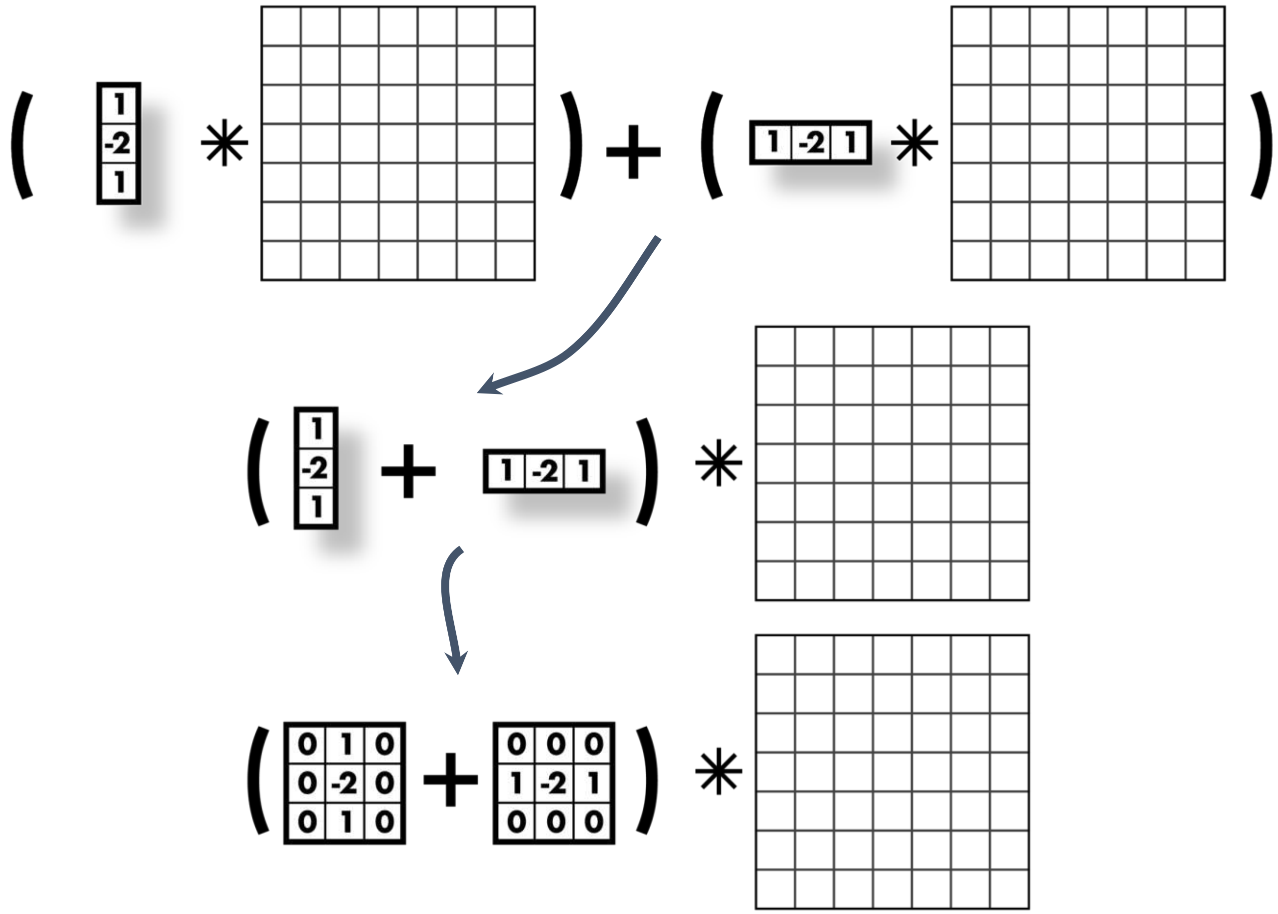
Laplacians

- Laplacian:
  - $$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$



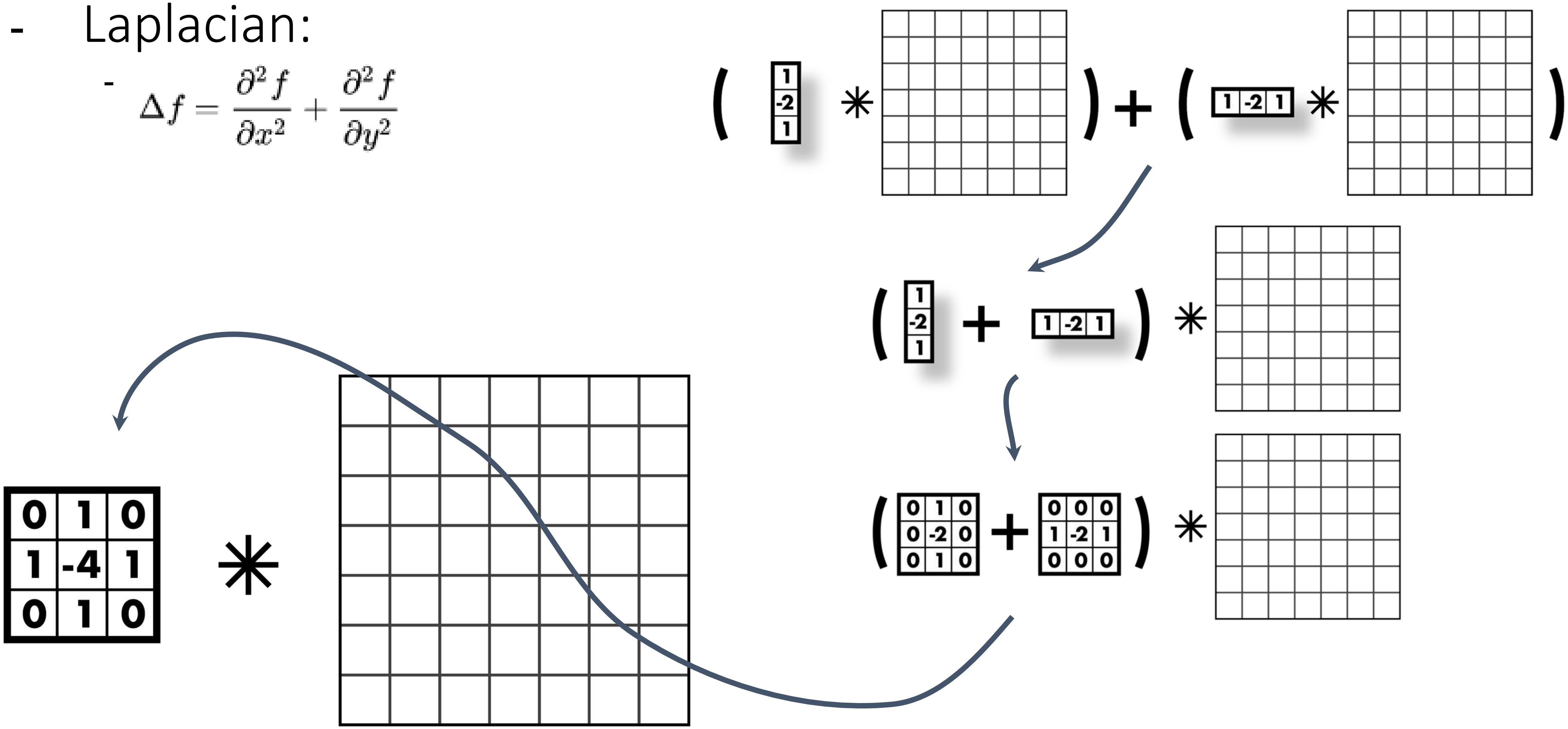
Laplacians

- Laplacian:
  - $$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$



Laplacians

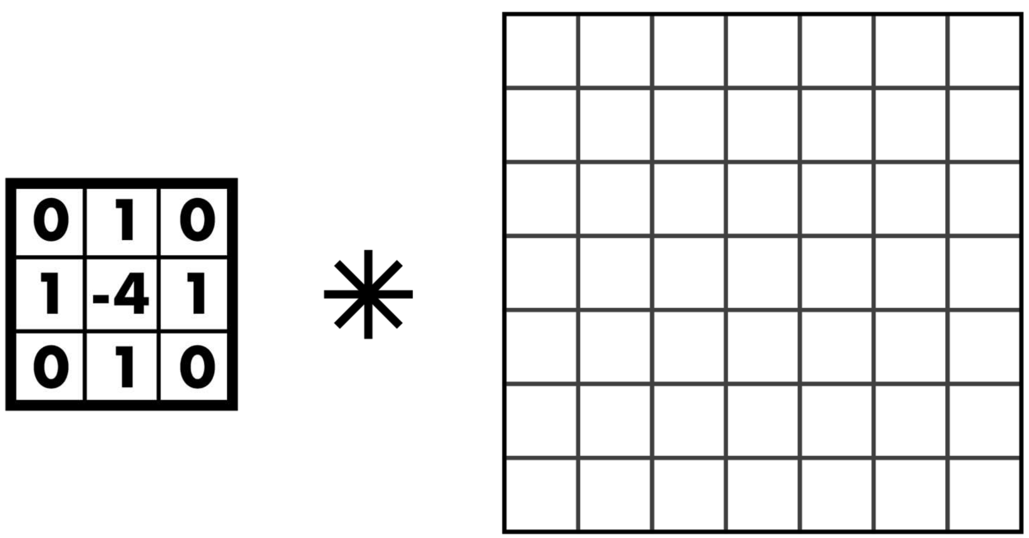
- Laplacian:
  - $$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

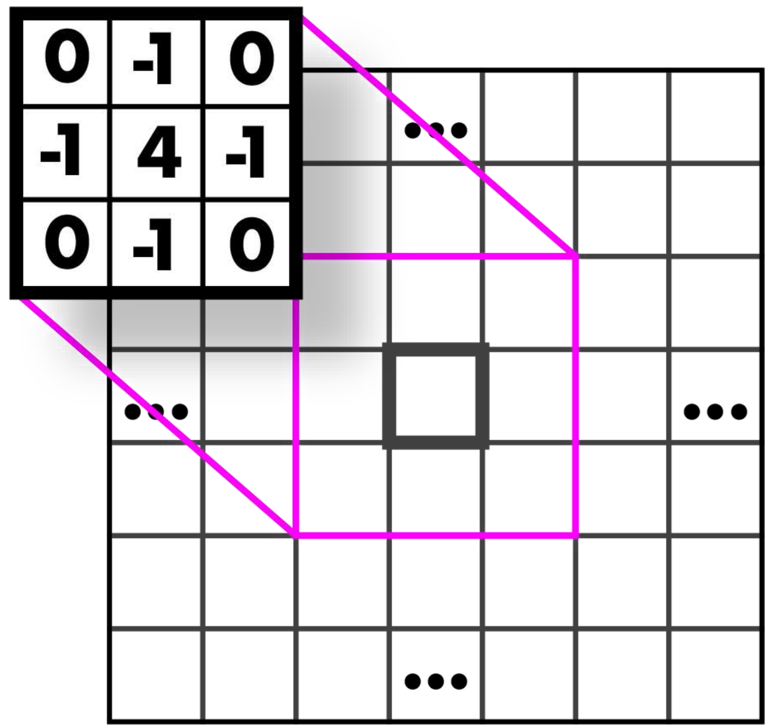


This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Laplacians

- Laplacian:
  - $$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$
- Negative Laplacian, -4 in middle
- Positive Laplacian --->

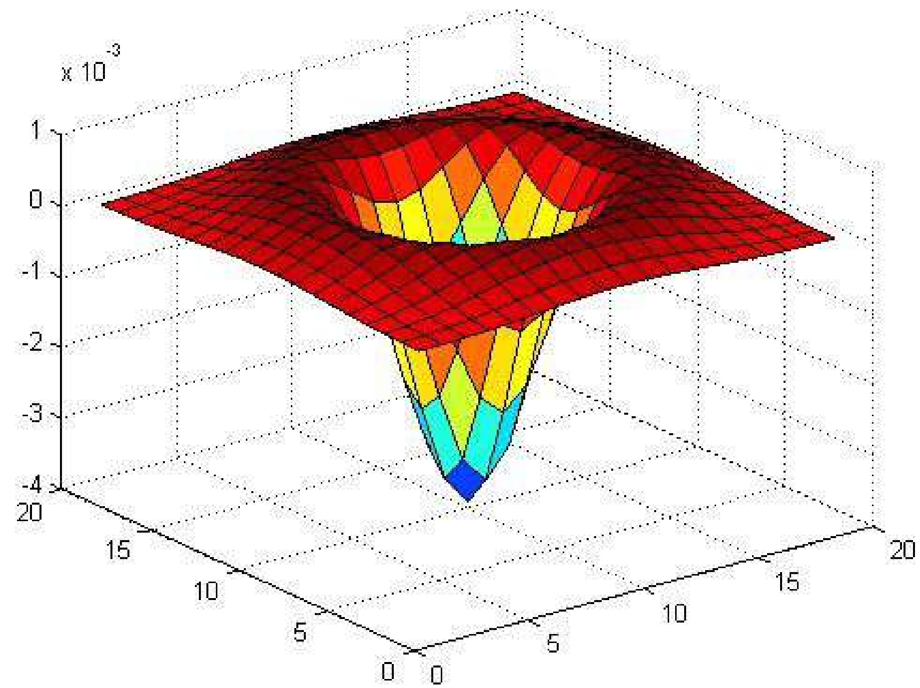




This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Laplacians also sensitive to noise

- Again, use gaussian smoothing
- Can just use one kernel since convs commute
- Laplacian of Gaussian, LoG
- Can get good approx. with
  5x5 - 9x9 kernels



Today's Agenda

- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Another edge detector

- Image is a function
  - Has high frequency and low frequency components
  - Think in terms of fourier transform
- Edges are high frequency changes
- Maybe we want to find edges of a specific size (i.e. specific frequency)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Difference of Gaussian (DoG)

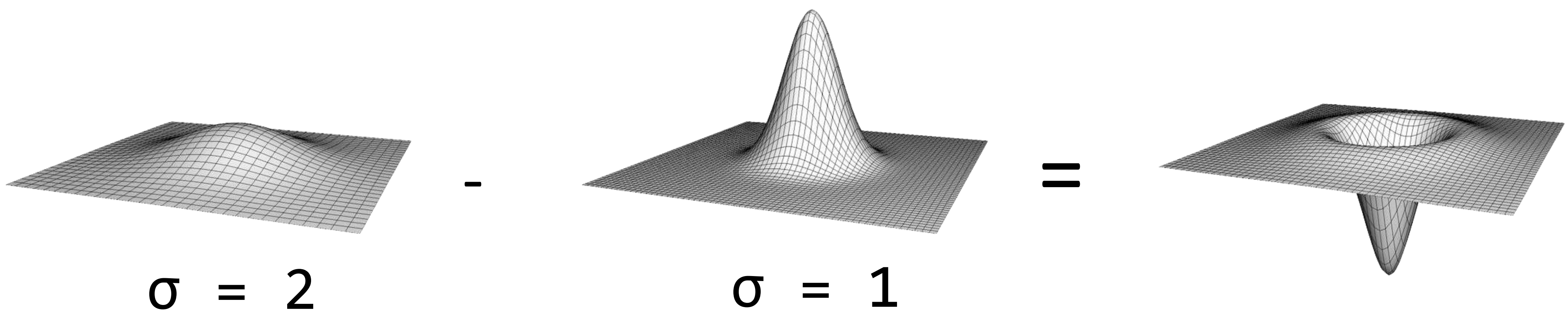
- Gaussian is a low pass filter
- Strongly reduce components with frequency f **>** σ
- (g*I) low frequency components
- I - (g*I) high frequency components
- g(σ1)*I - g(σ2)*I
    - Components in between these frequencies
- g(σ1)*I - g(σ2)*I = [g(σ1) - g(σ2)]*I

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Difference of Gaussian (DoG)

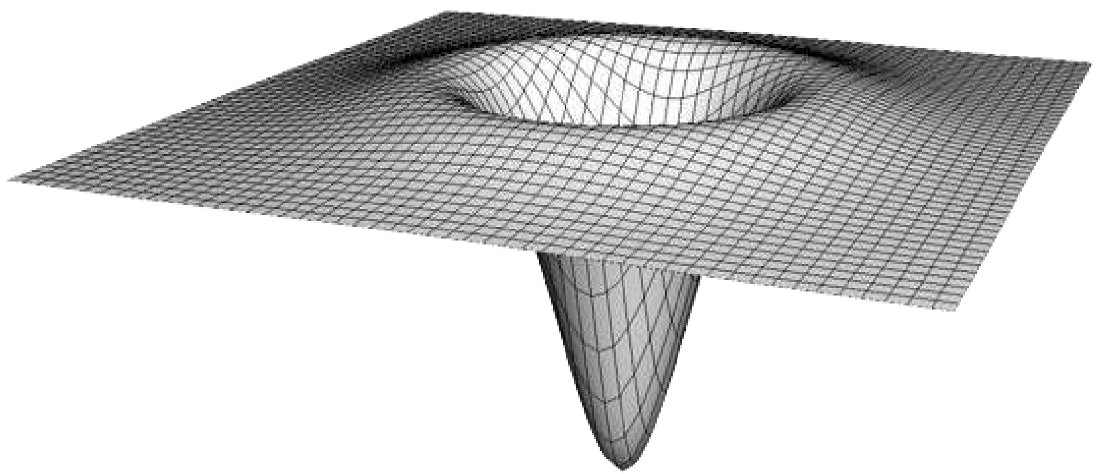
- $g(\sigma 1)*I - g(\sigma 2)*I = [g(\sigma 1) - g(\sigma 2)]*I$

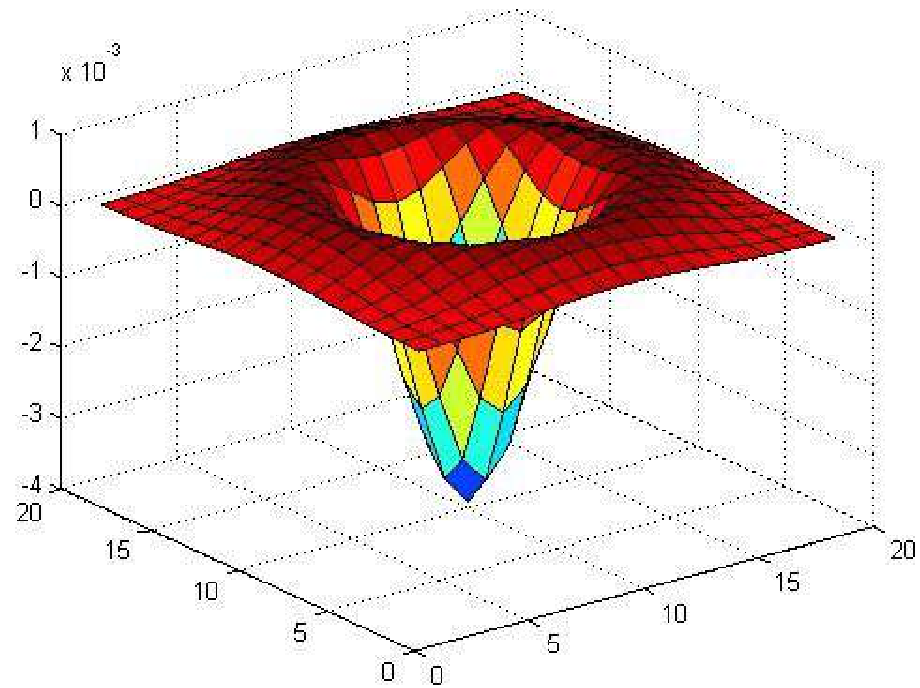


Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Difference of Gaussian (DoG)

- g(σ1)*I - g(σ2)*I = [g(σ1) - g(σ2)]*I
- This looks a lot like our LoG!
- (not actually the same but similar)





DoG (1 - 0)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

DoG (2 - 1)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

DoG (3 - 2)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

DoG (4 - 3)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

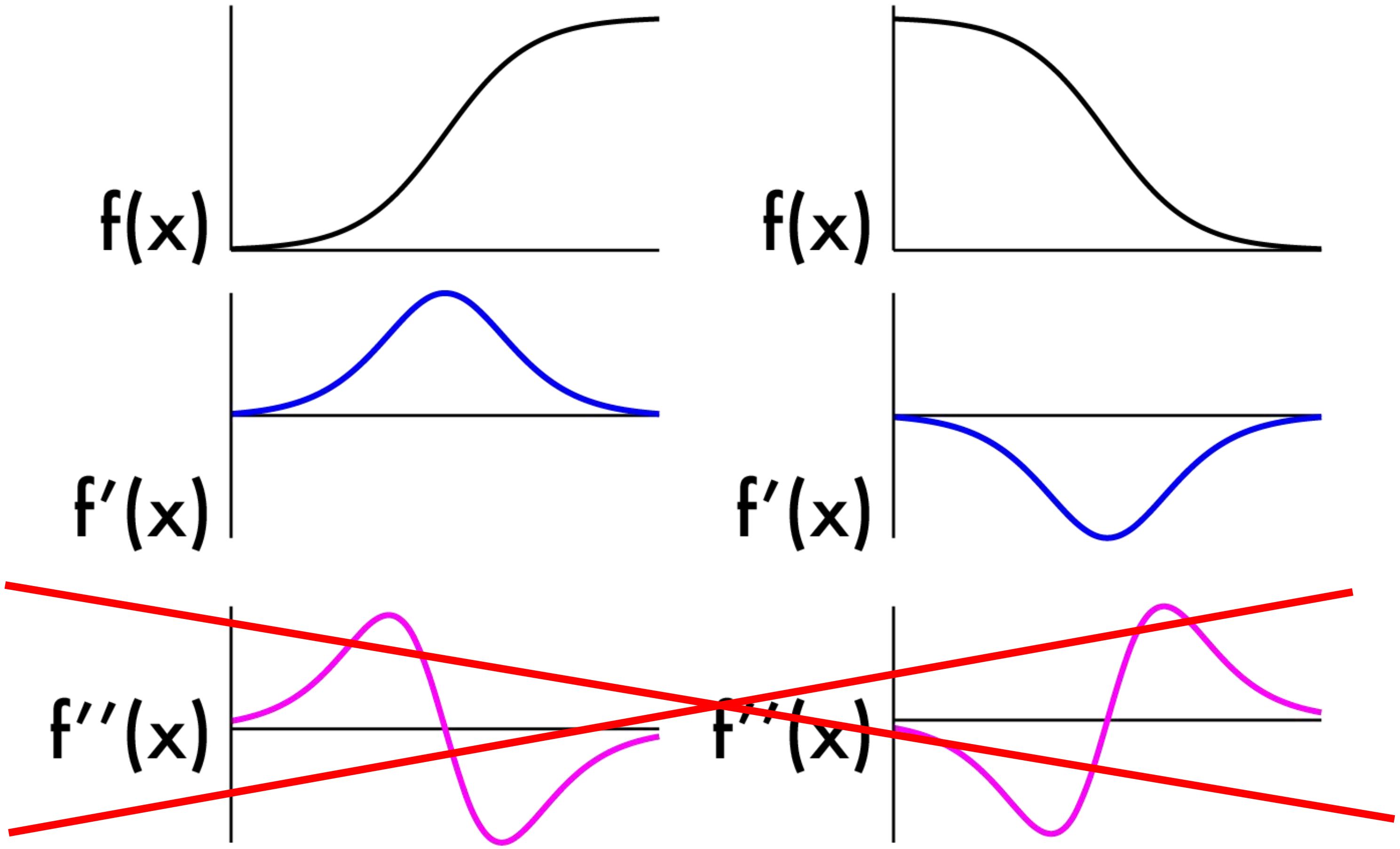
- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

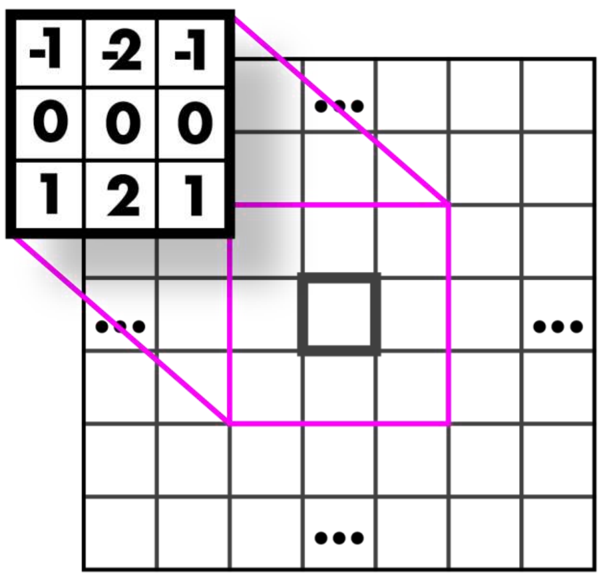
Another approach: gradient magnitude

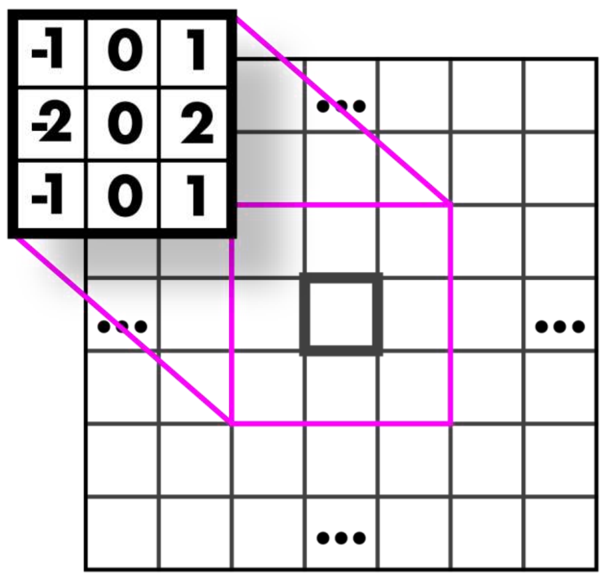
- Don’t need 2nd derivatives
- Just use magnitude of gradient



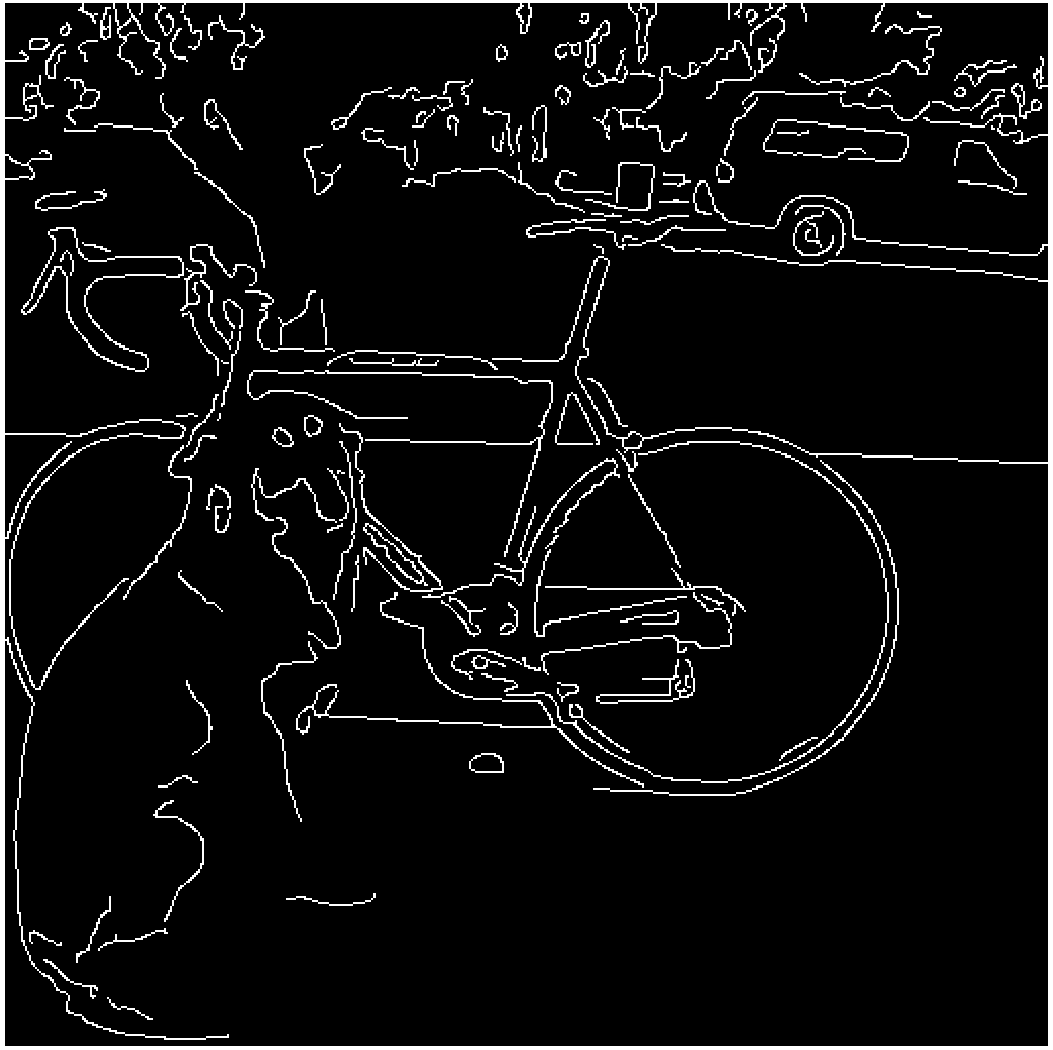
Another approach: gradient magnitude

- Don’t need 2nd derivatives
- Just use magnitude of gradient
- Are we done? No!





What we really want: line drawing



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Canny Edge Detection

Algorithm:

- Smooth image (only want "real" edges, not noise)
- Calculate gradient direction and magnitude
- Non-maximum suppression perpendicular to edge
- Threshold into strong, weak, no edge
- Connect together components

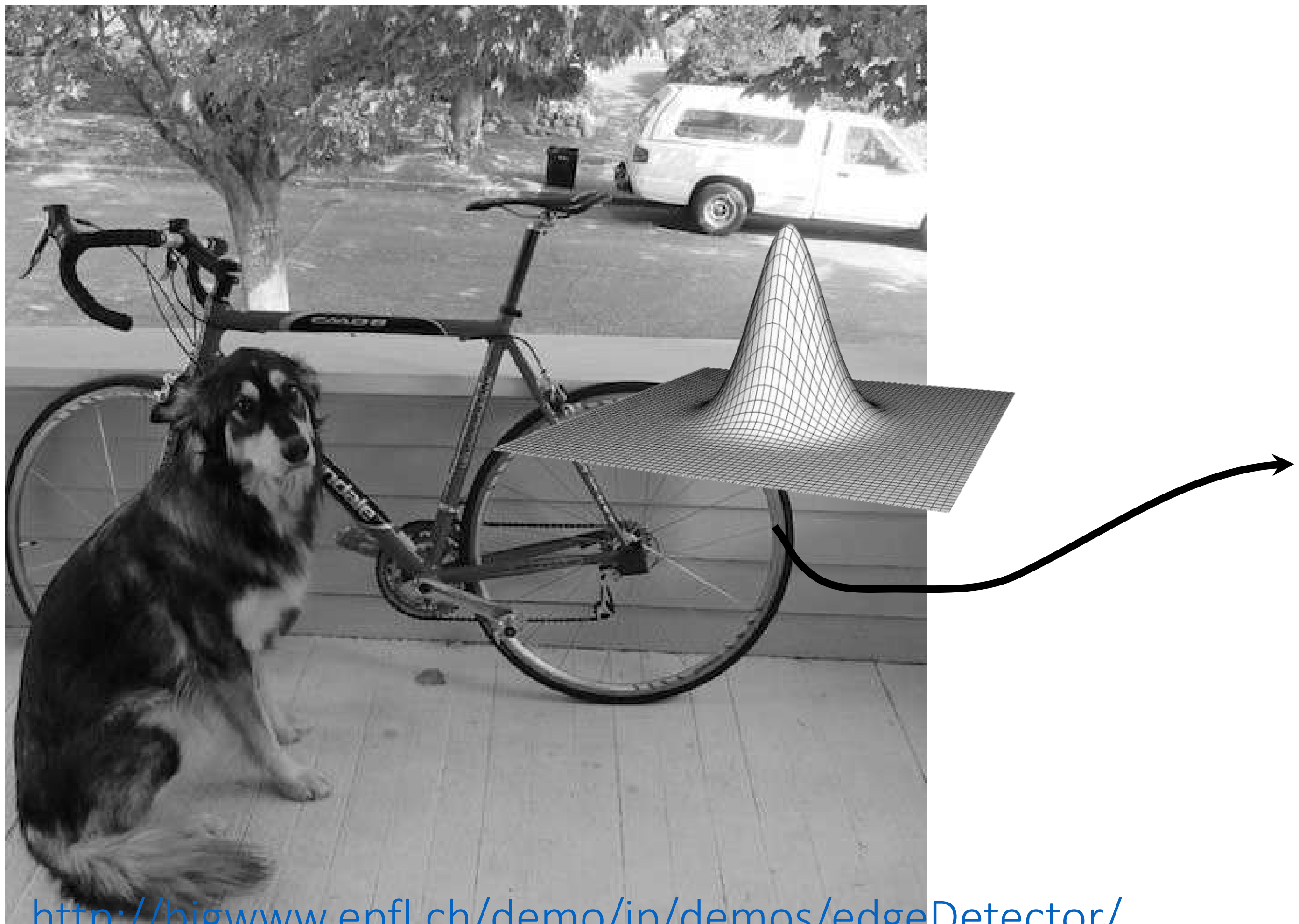
http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Smooth image

- You know how to do this, gaussians!



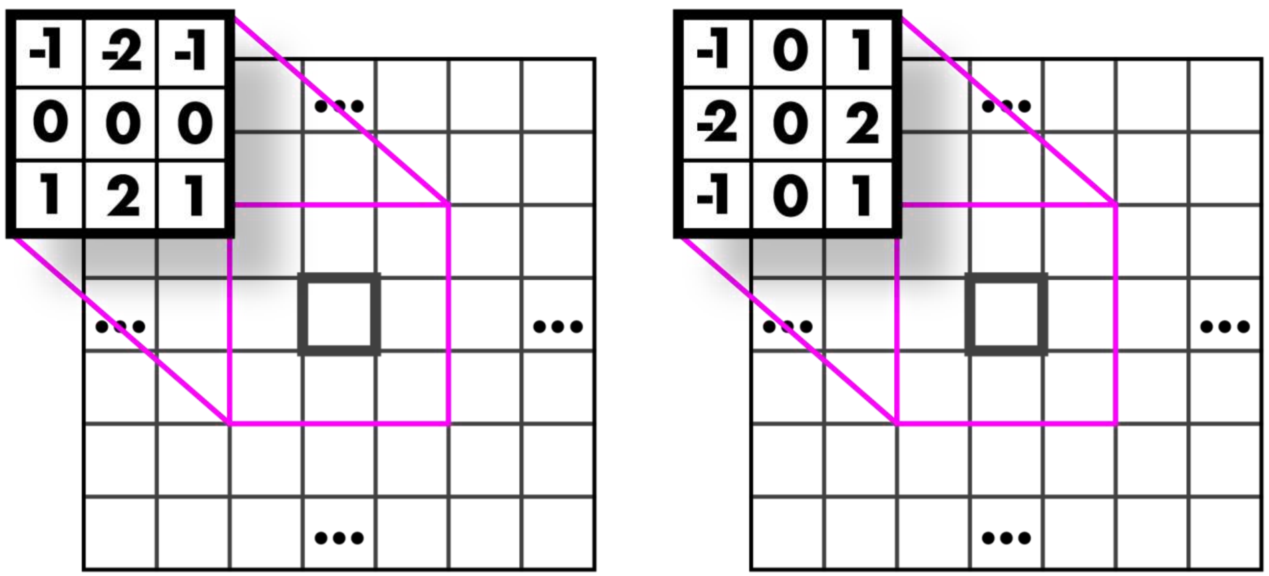
http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Gradient magnitude and direction

- Sobel filter



http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

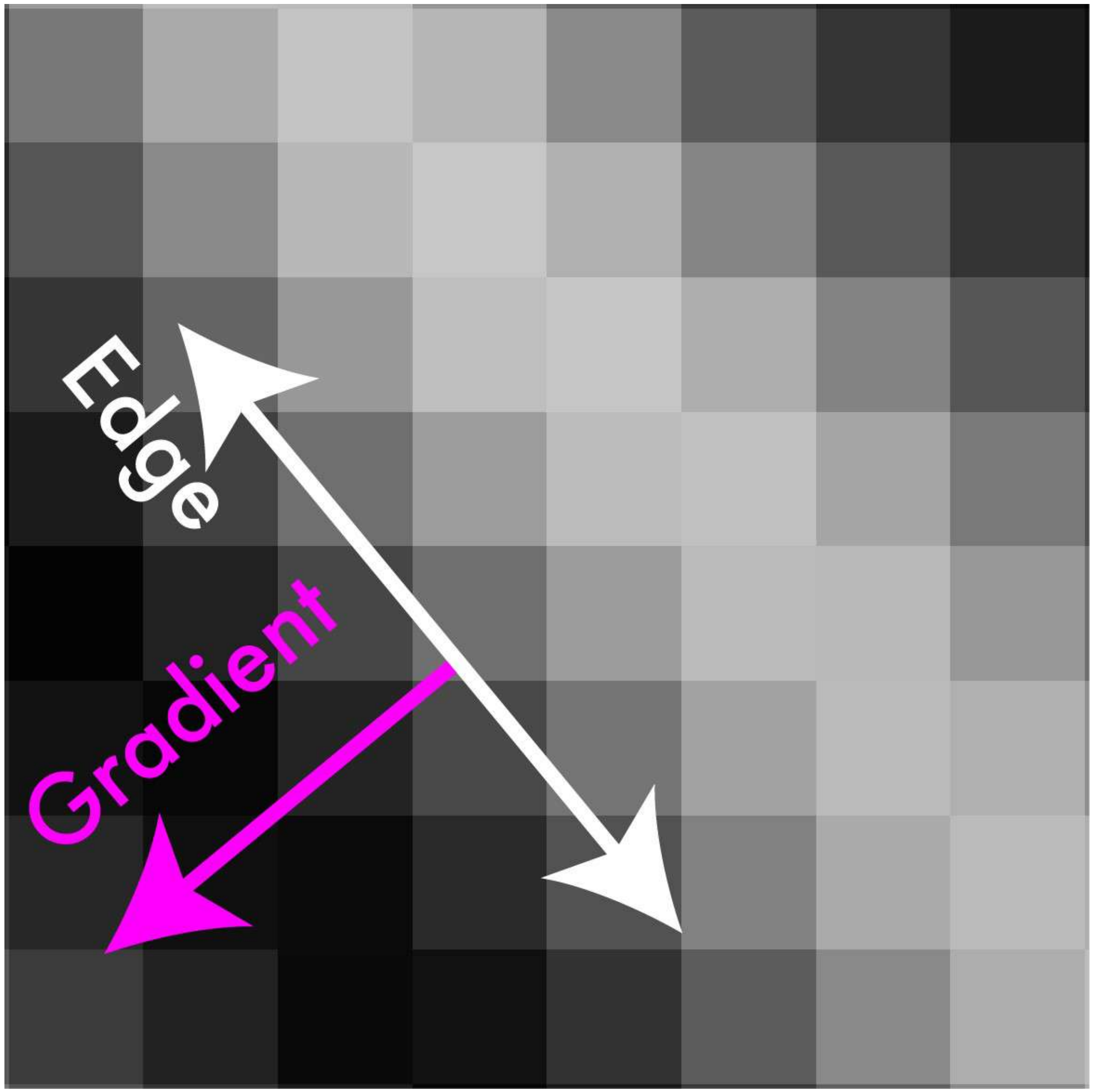
Non-maximum suppression

- Want single pixel edges, not thick blurry lines
- Need to check nearby pixels
- See if response is highest

http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

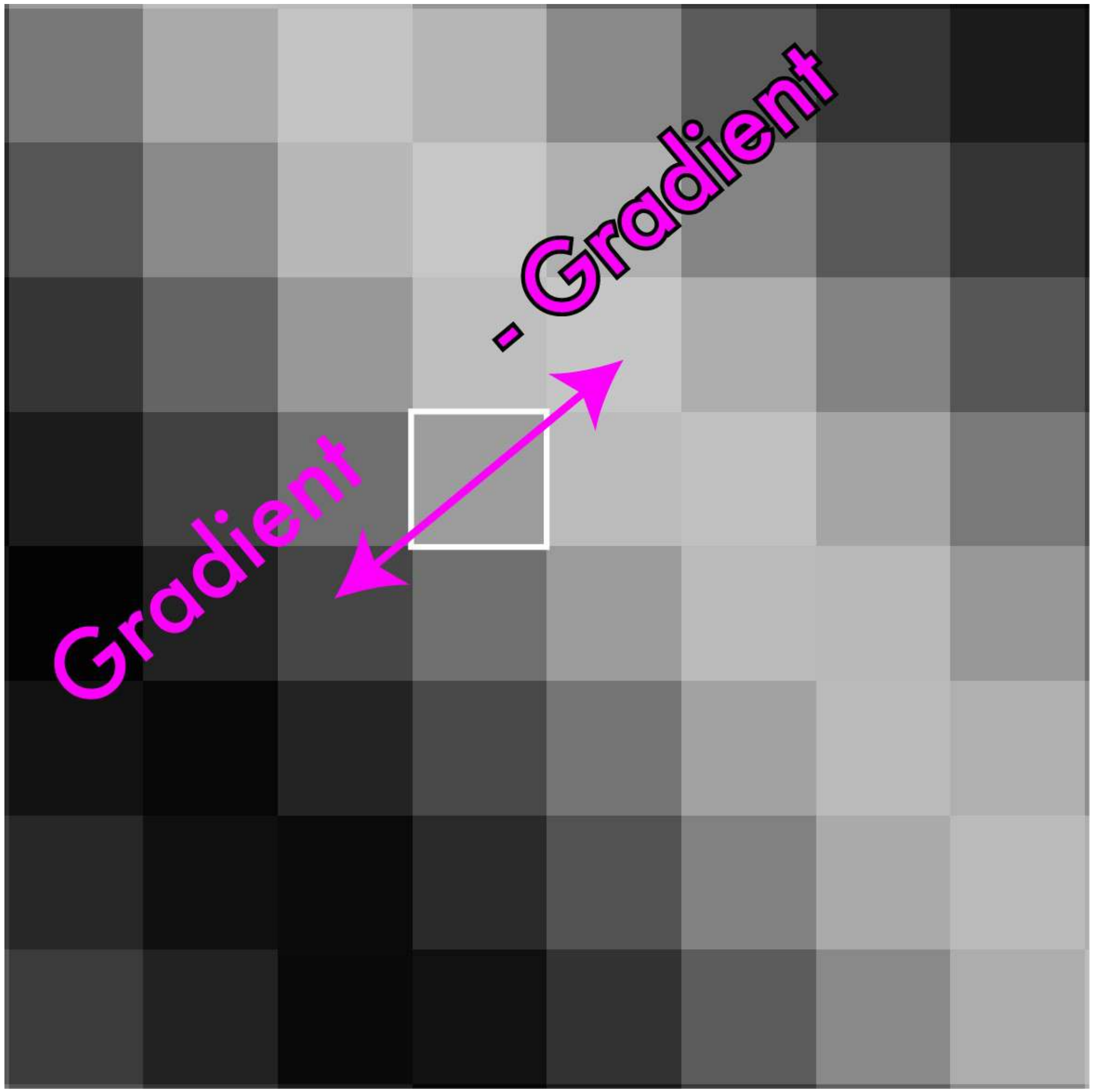
Non-maximum suppression



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

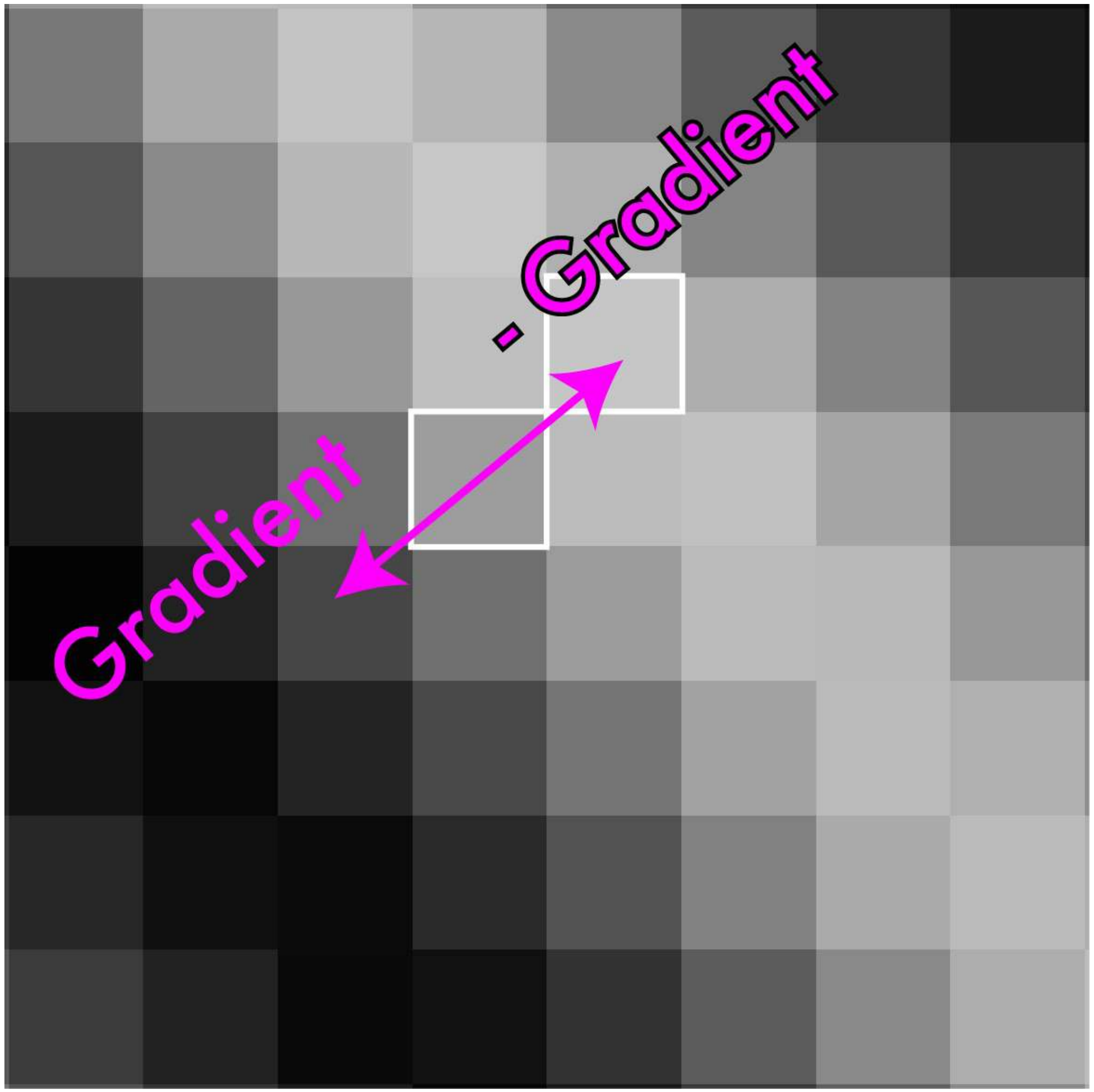
Non-maximum suppression



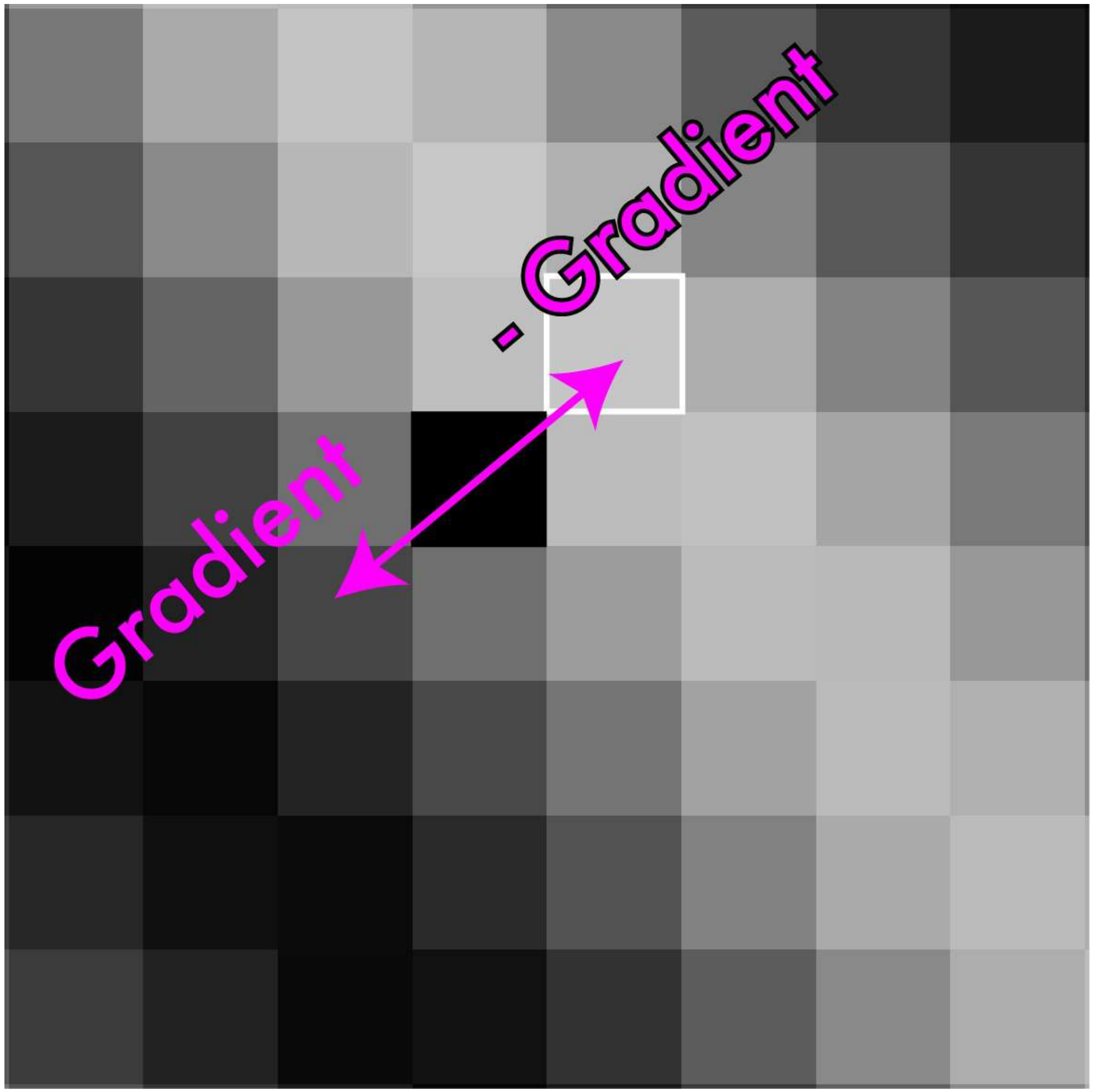
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Non-maximum suppression

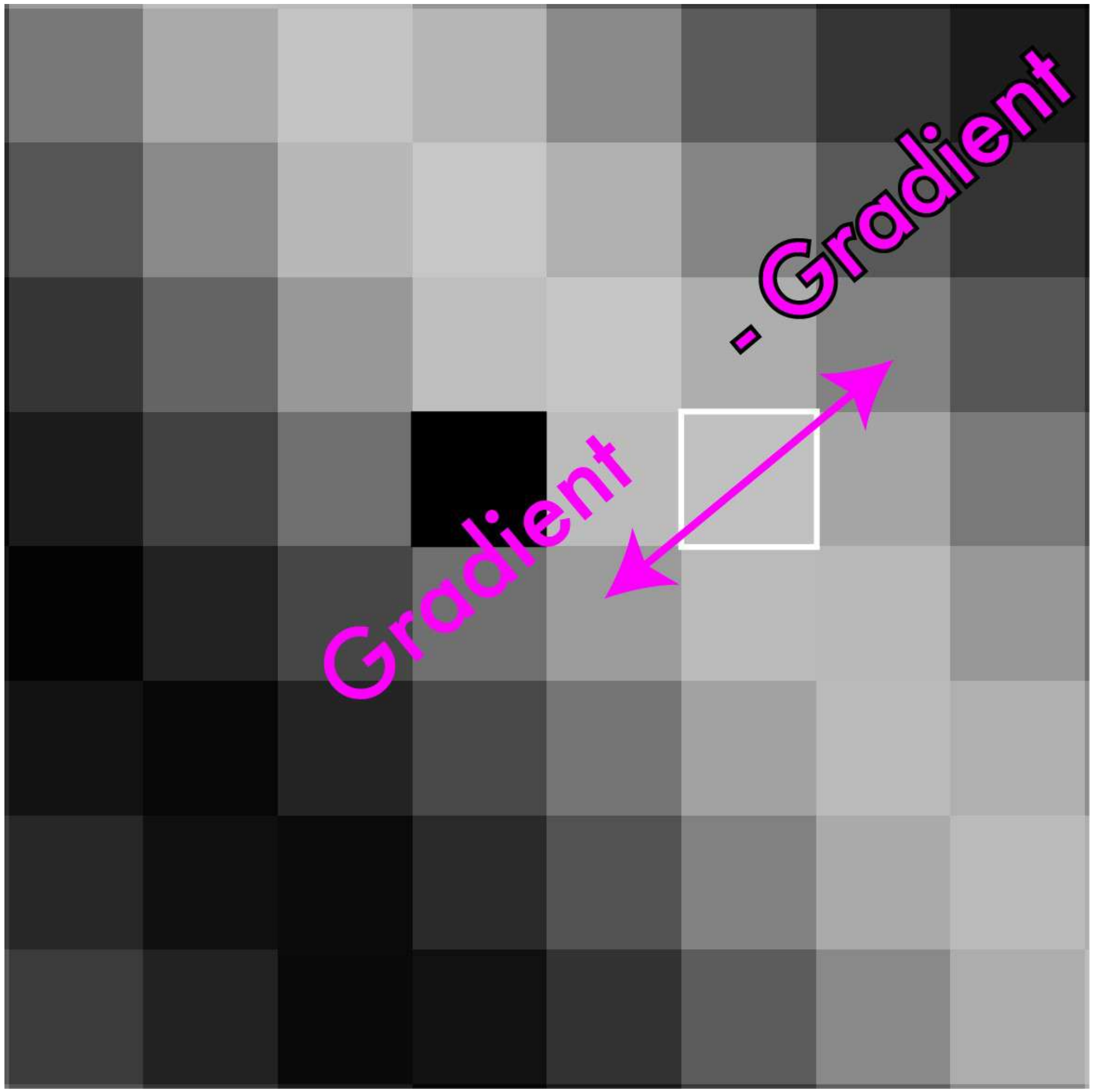


Non-maximum suppression

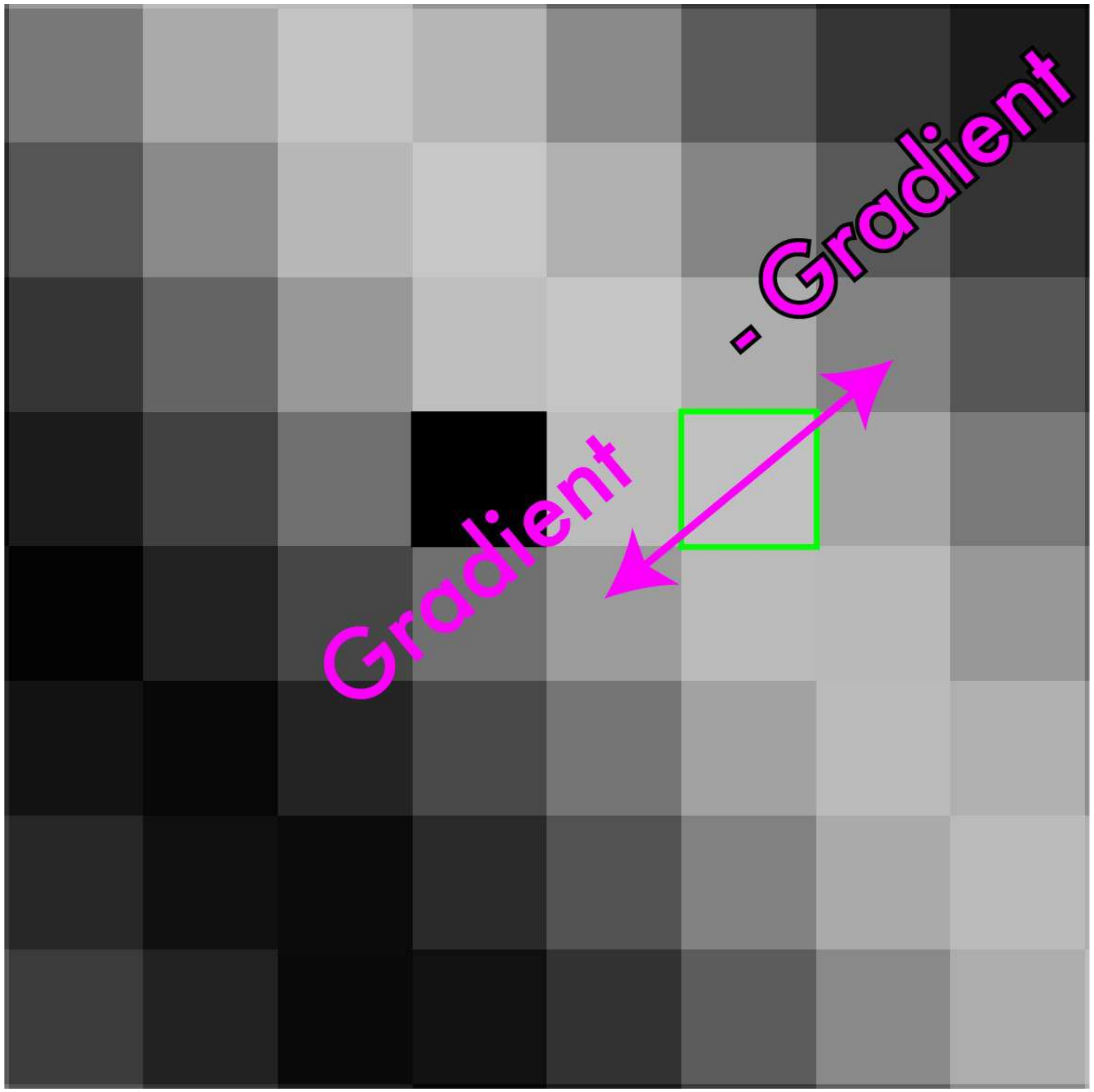


This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

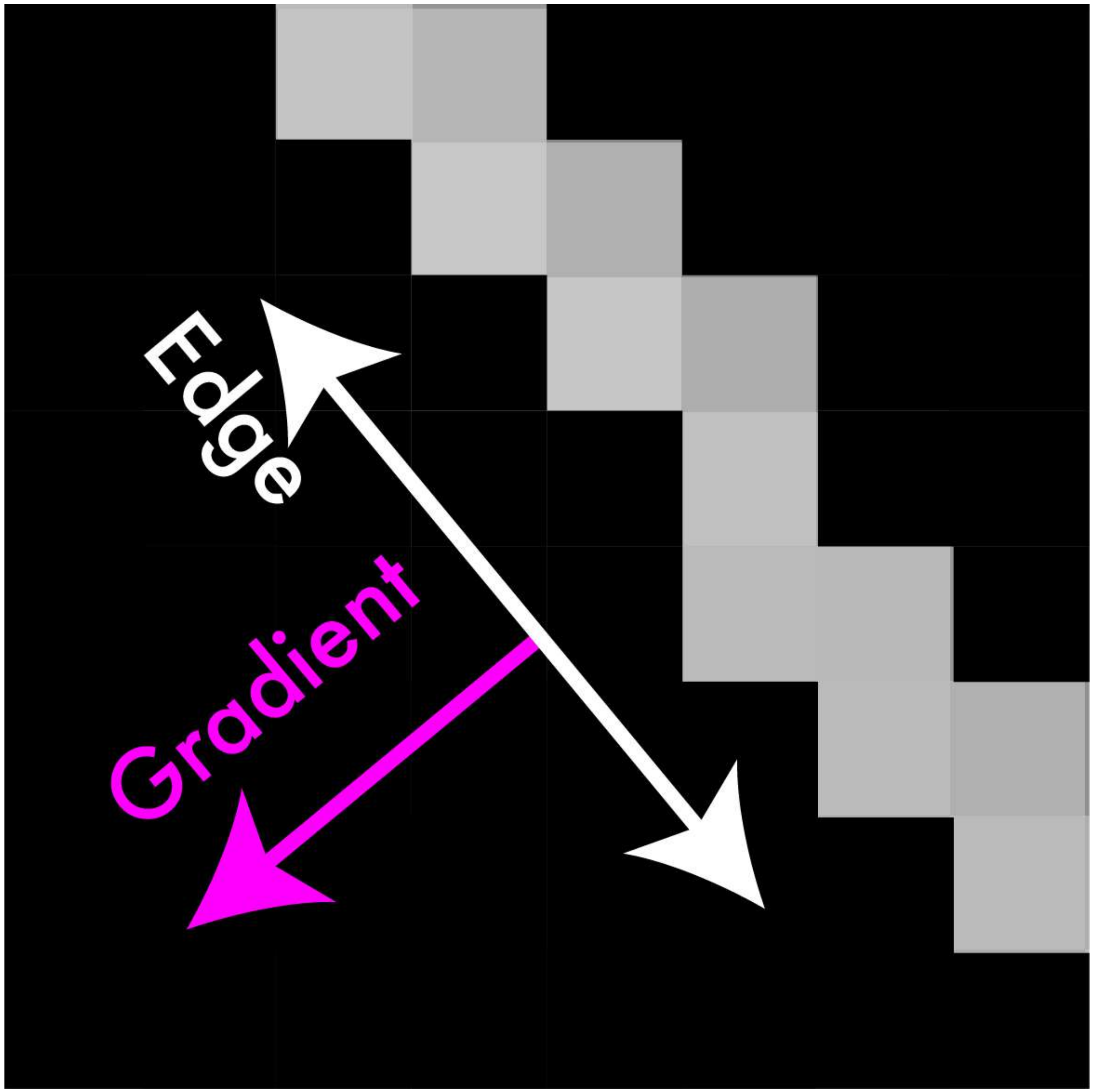
Non-maximum suppression



Non-maximum suppression

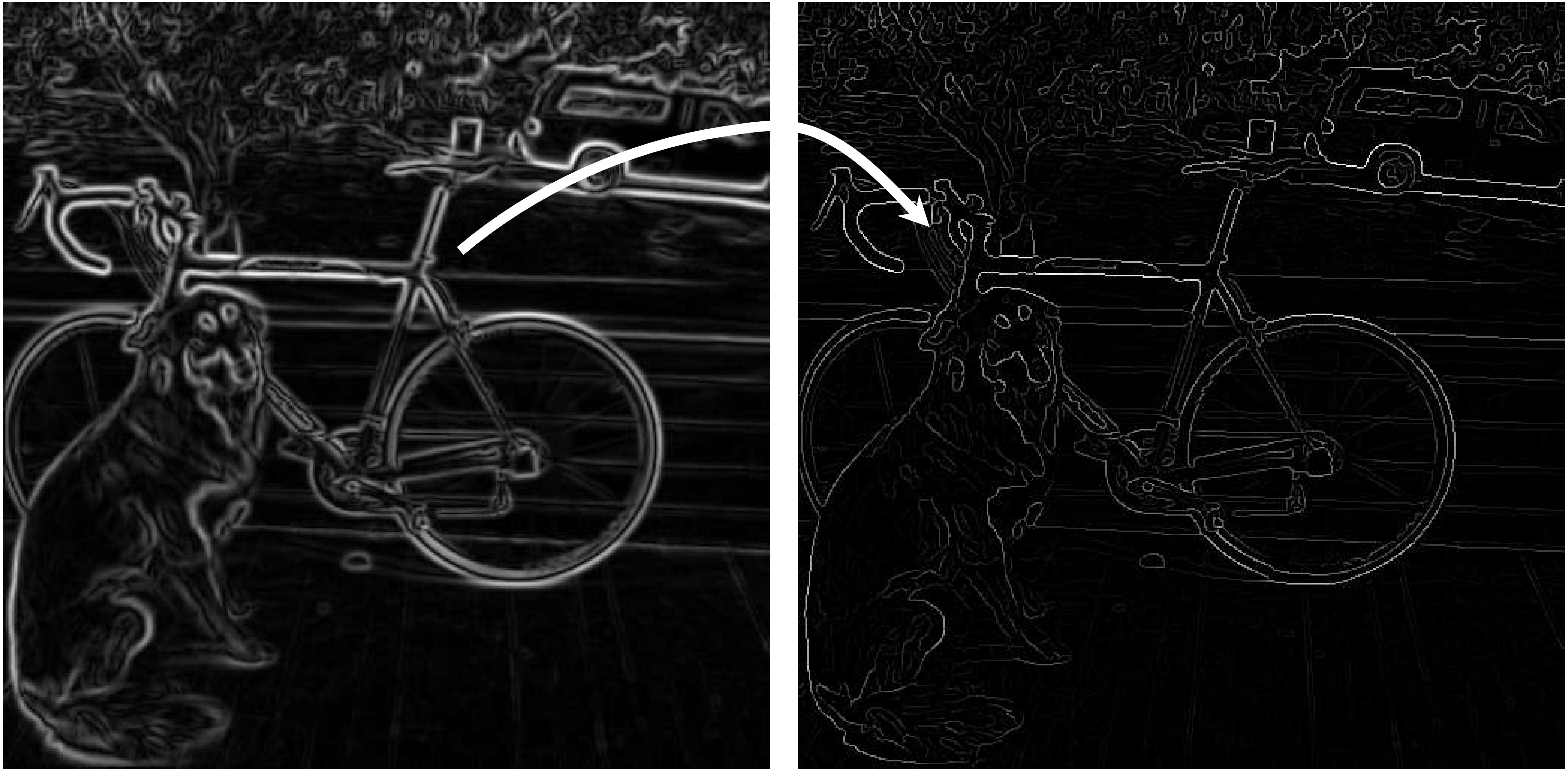


Non-maximum suppression



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Non-maximum suppression



http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Threshold edges

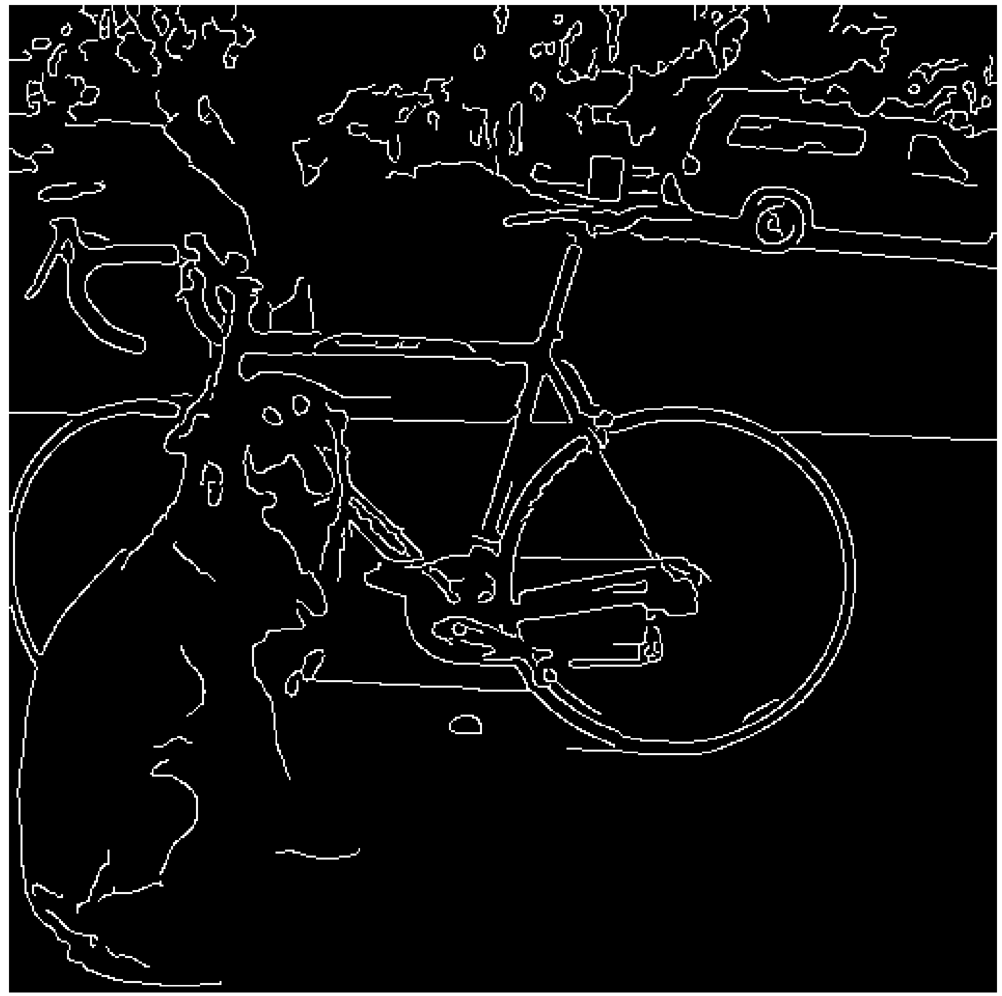
- Still some noise
- Only want strong edges
- 2 thresholds, 3 cases
  - R > T: strong edge
  - R < T but R > t: weak edge
  - R < t: no edge
- Why two thresholds?

http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Connect them up!

- Strong edges are edges!
- Weak edges are edges
  iff they connect to strong
- Look in some neighborhood
  (usually 8 closest)



http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

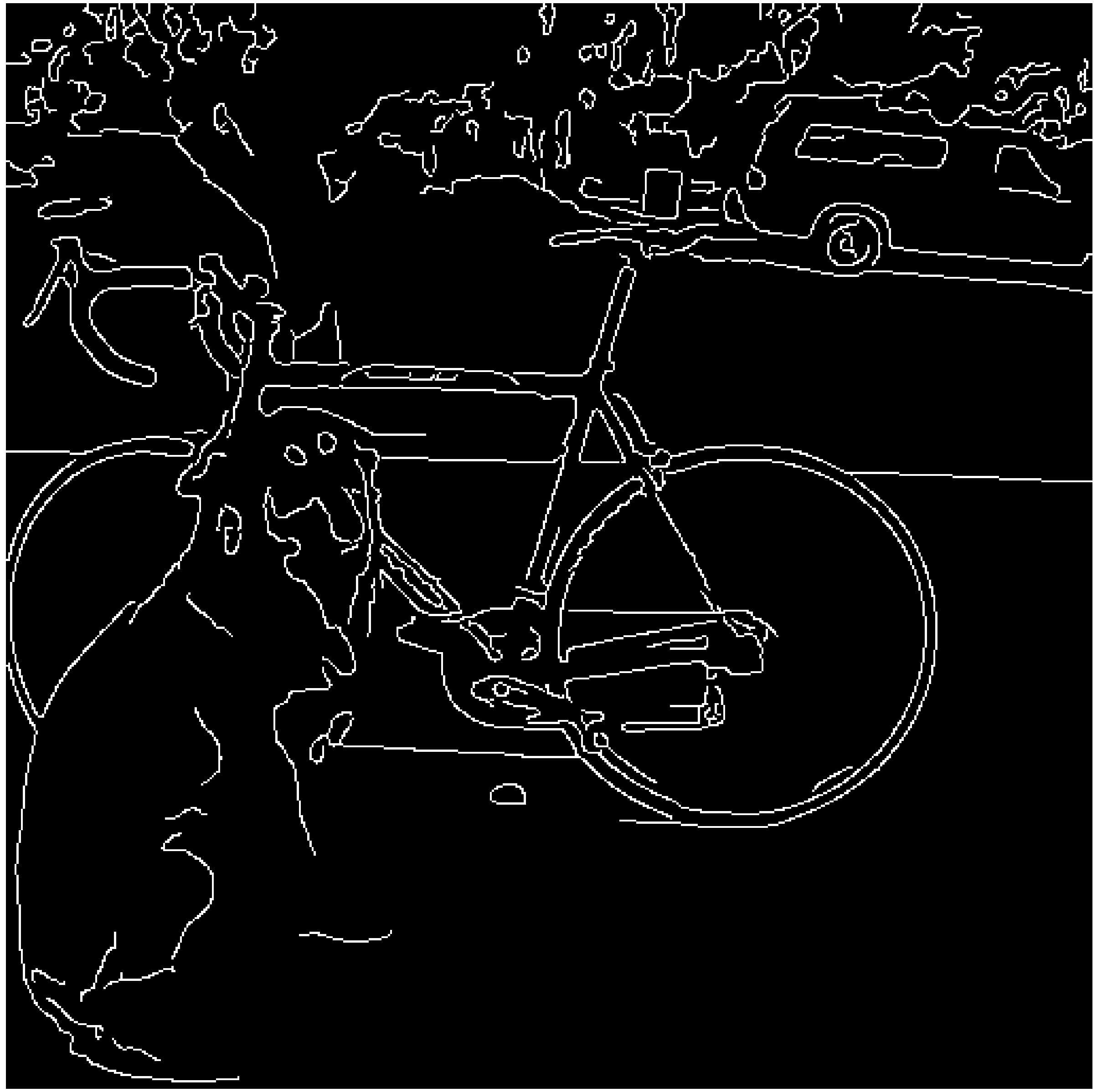
Canny Edge Detection

Algorithm:

- Smooth image (only want “real” edges, not noise)
- Calculate gradient direction and magnitude
- Non-maximum suppression perpendicular to edge
- Threshold into strong, weak, no edge
- Connect together components
- Tunable: Sigma, thresholds

http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

Canny Edge Detection



http://bigwww.epfl.ch/demo/ip/demos/edgeDetector/

Master programmes in Artificial Intelligence 4 Careers in Europe

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Master programmes in Artificial Intelligence 4 Careers in Europe

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 7: Features – Corners

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- What can we do with convolutions
- What is an edge – image derivatives
- Sobel filters
- Laplacian filters
- Difference of Gaussian filters
- Canny edge detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Features
- Self-difference
- Harris corner detection

[material based on Joseph Redmon’s course]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Features
- Self-difference
- Harris corner detection

What else is there?

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

So what can we do with these convolutions anyway?

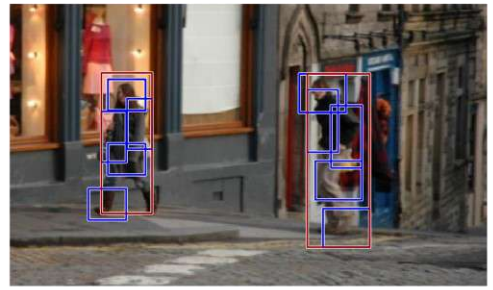
- Blurring
- Sharpening
- Edges
- Features
- Derivatives
- Super-resolution
- Classification
- Detection
- Image captioning
- ...

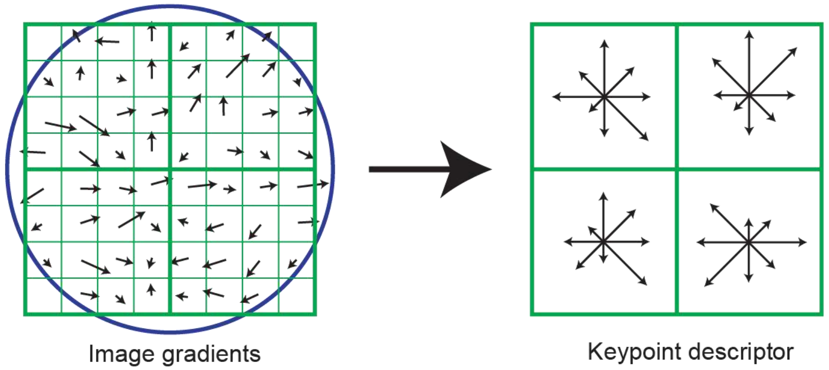
Co-financed by the European Union
Connecting Europe Facility

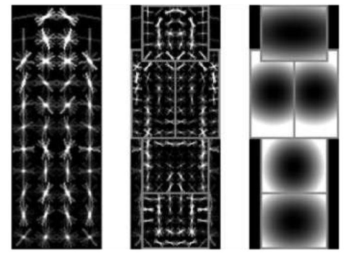
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

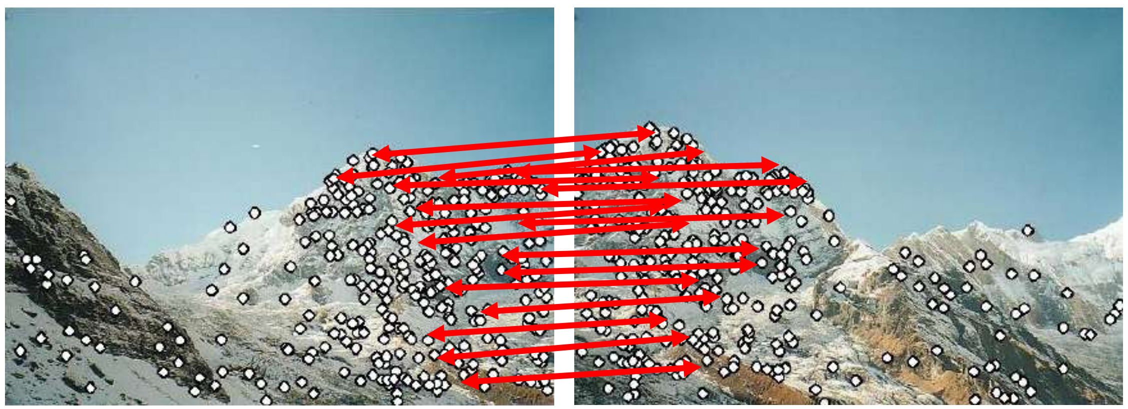
Features!

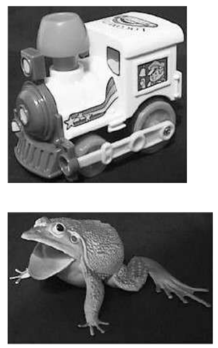
- Highly descriptive local regions
- Ways to describe those regions
- Useful for:
  - Matching
  - Recognition
  - Detection











What makes a good feature?

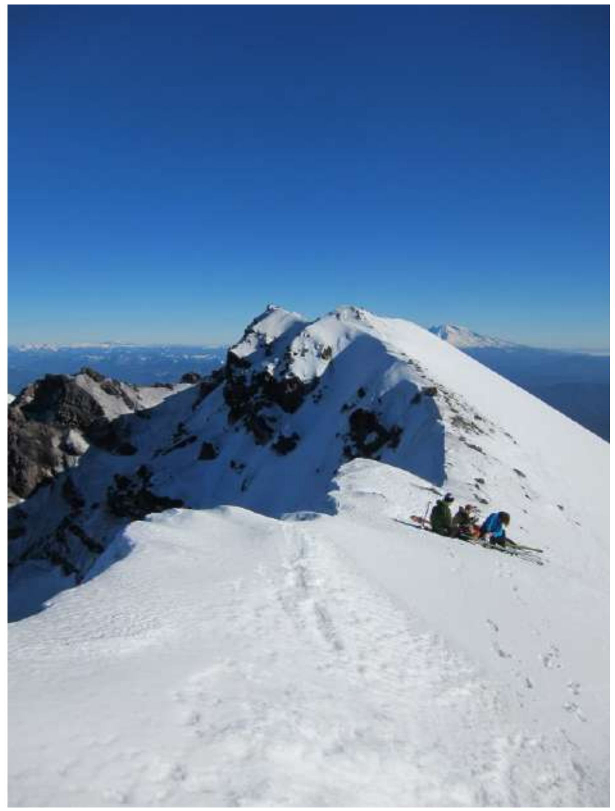
- Want to find patches in image that are useful or have some meaning
- For objects, want a patch that is common to that object but not in general
- For panorama stitching, want patches that we can find easily in another image of same place
- Good features are <u>unique</u>!
    - Can find the “same” feature easily
    - Not mistaken for “different” features

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Application - How to create a panorama

- Say we are stitching images to create a panorama
- Want patches in one image to match to patches in another image
- Hopefully only match one spot



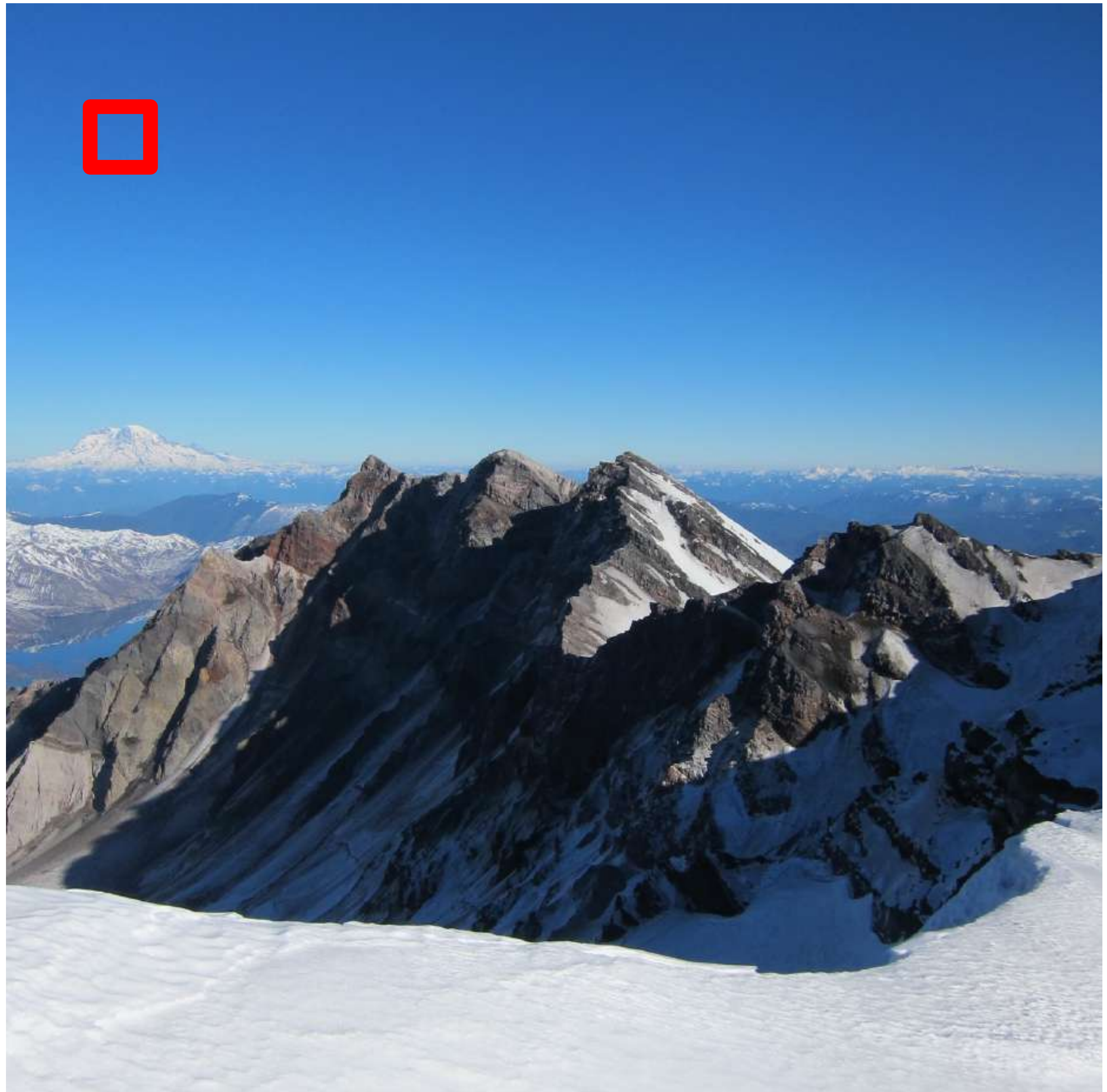
How close are two patches?

- (weighted) Summed squared difference
- Images I, J
- $\Sigma_{x,y}\ w(x,y)\ (I(x,y) - J(x,y))^2$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How can we find unique patches?

- Sky: bad
  - Very little variation
  - Could match any other sky



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

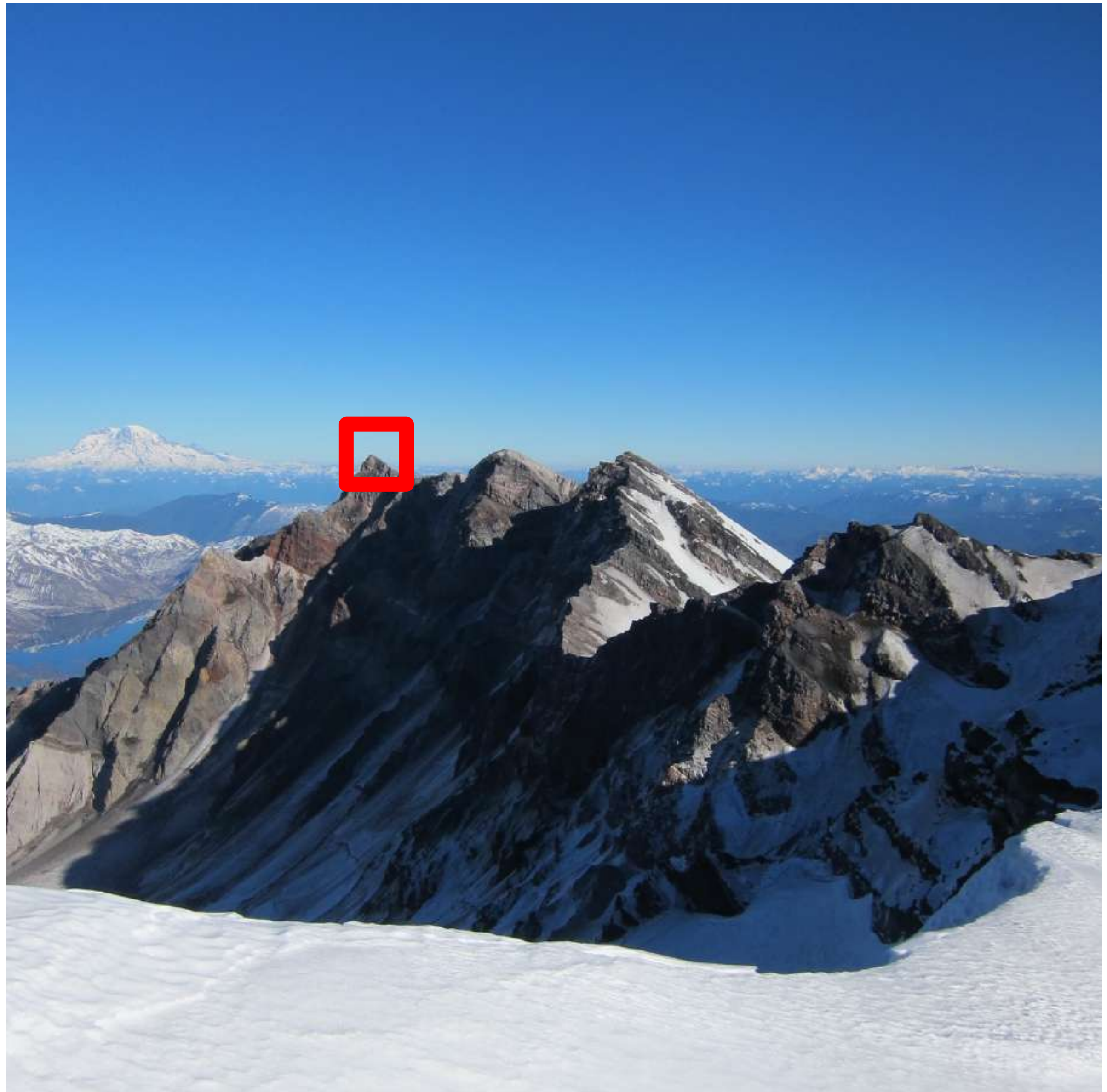
How can we find unique patches?

- Sky: bad
  - Very little variation
  - Could match any other sky
- Edge: ok
  - Variation in one direction
  - Could match other patches along same edge

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How can we find unique patches?

- Sky: bad
    - Very little variation
    - Could match any other sky
- Edge: ok
    - Variation in one direction
    - Could match other patches along same edge
- Corners: good!
    - Only one alignment matches



Today's Agenda

- Features
- Self-difference
- Harris corner detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How can we find unique patches?

- Want a patch that is unique in the image

- Can calculate distance between patch and every other patch, a lot of computation

- Instead, we could think about auto-correlation:
  - How well does image match shifted version of itself?

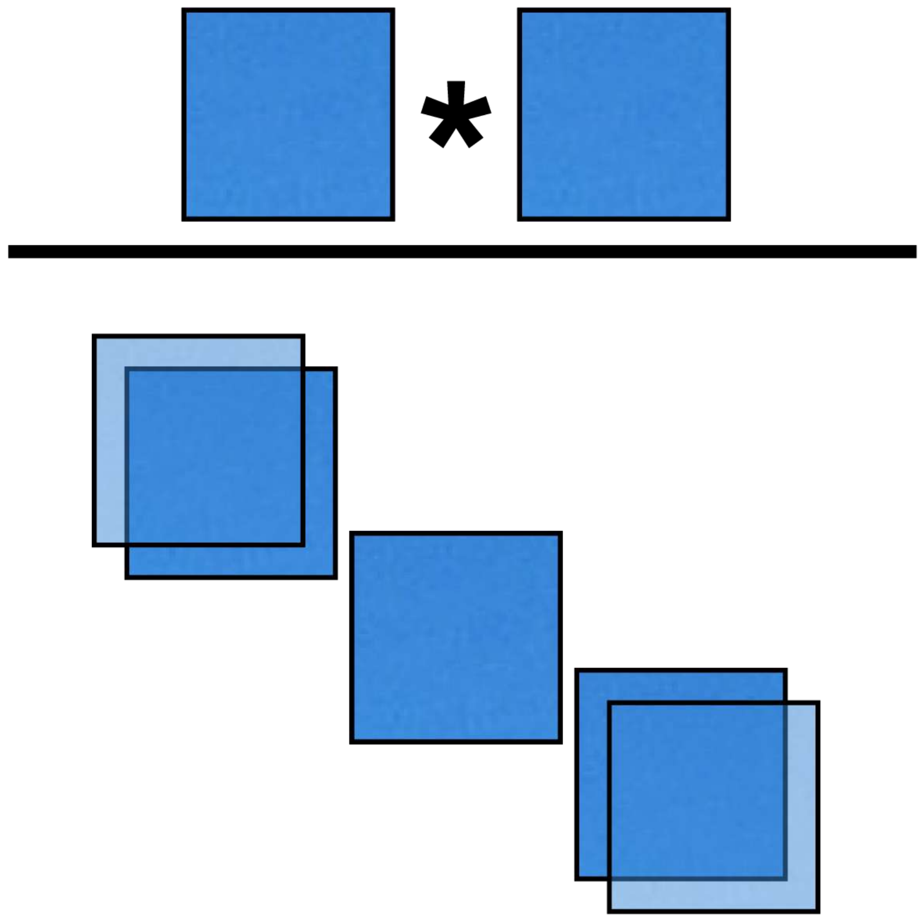
- Autocorrelation: $\Sigma_{\mathbf{d}}\Sigma_{x,y}\ w(x,y)\ (\ I(x+\mathbf{d}_x, y+\mathbf{d}_y) - I(x,y)\ )^2$

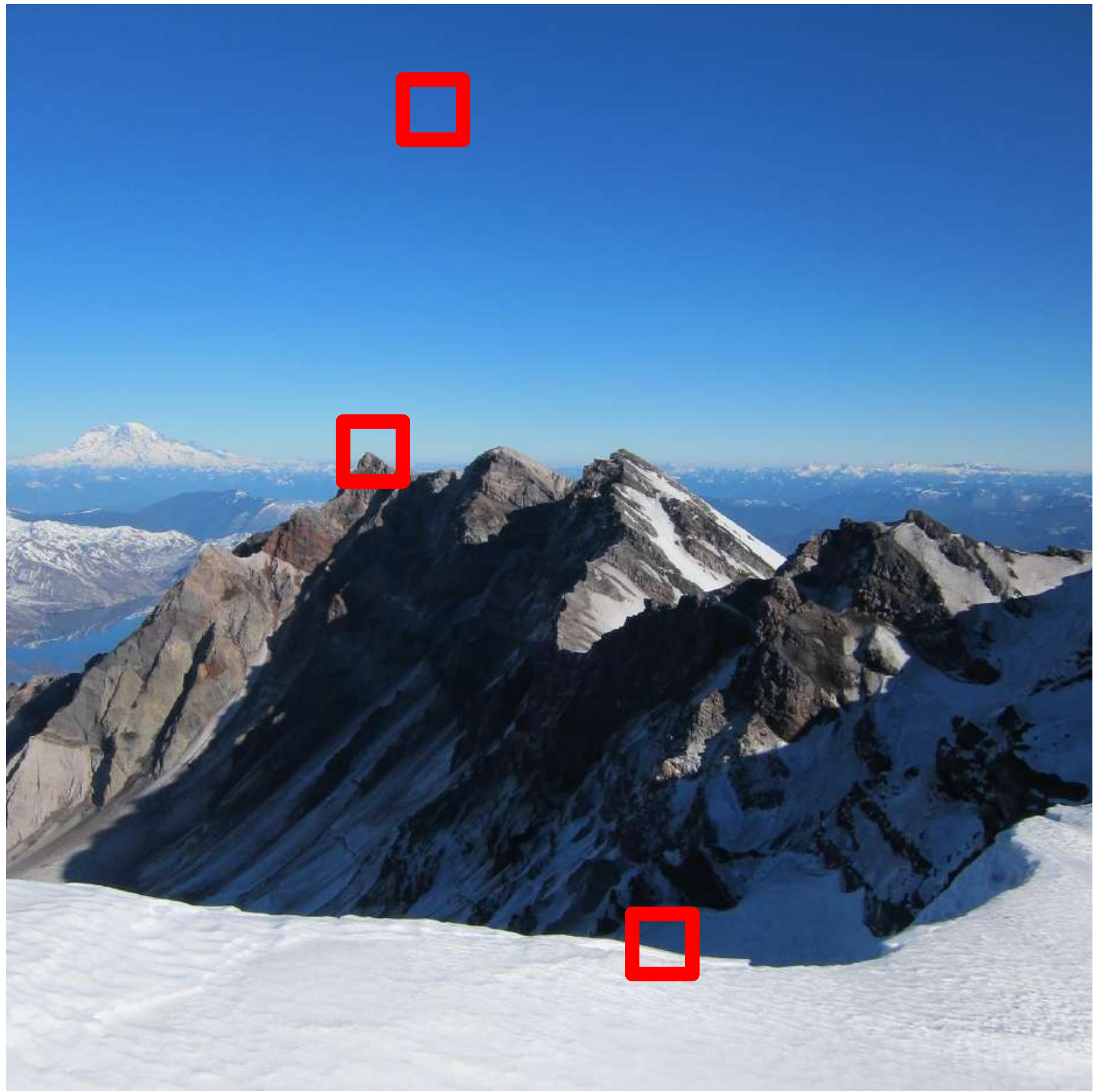
- Measure of self-difference (how much am I not myself?)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Self-difference

Sky: low everywhere

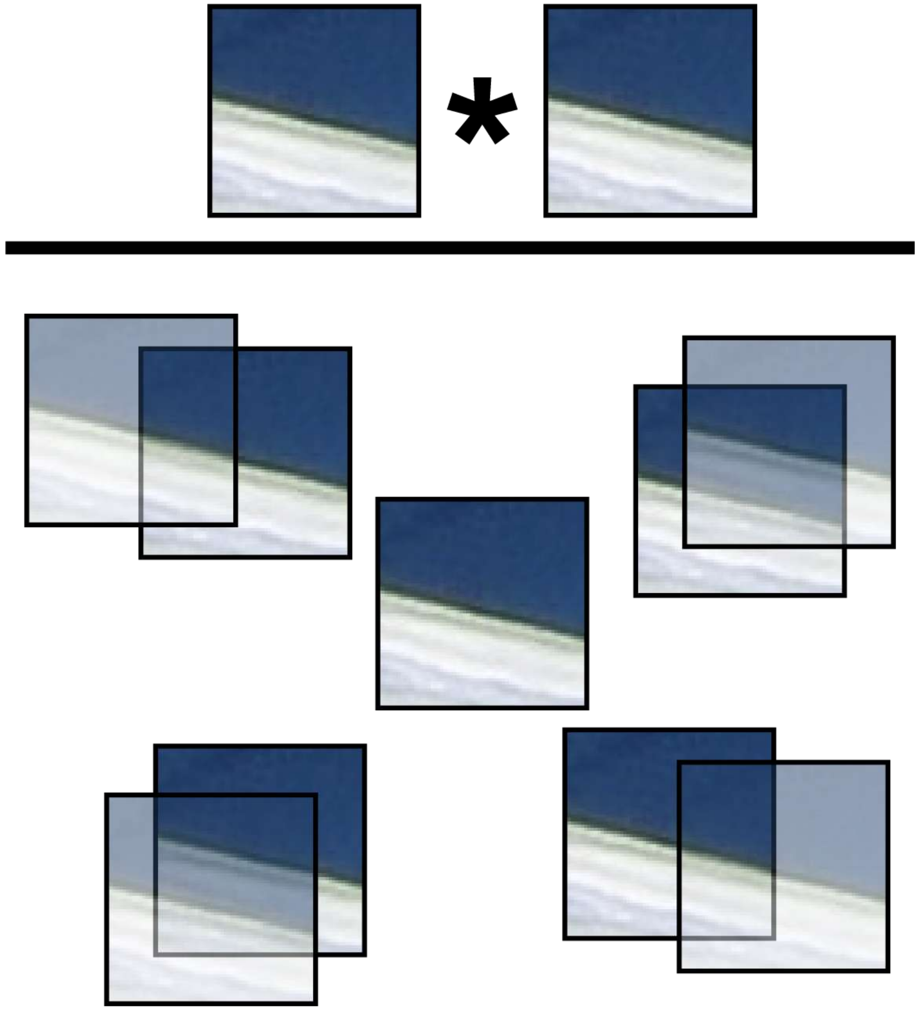


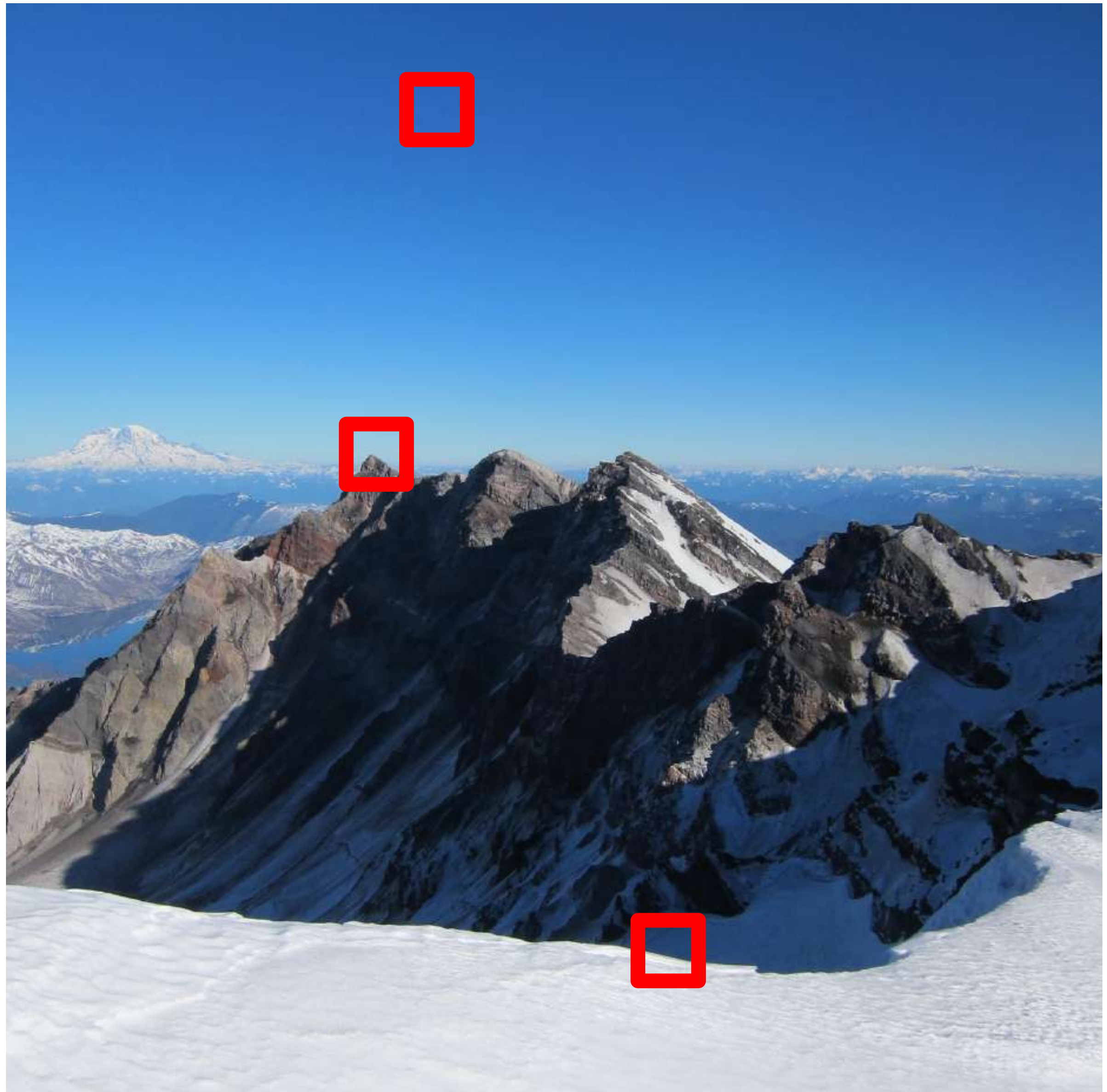


This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Self-difference

Edge: low along edge



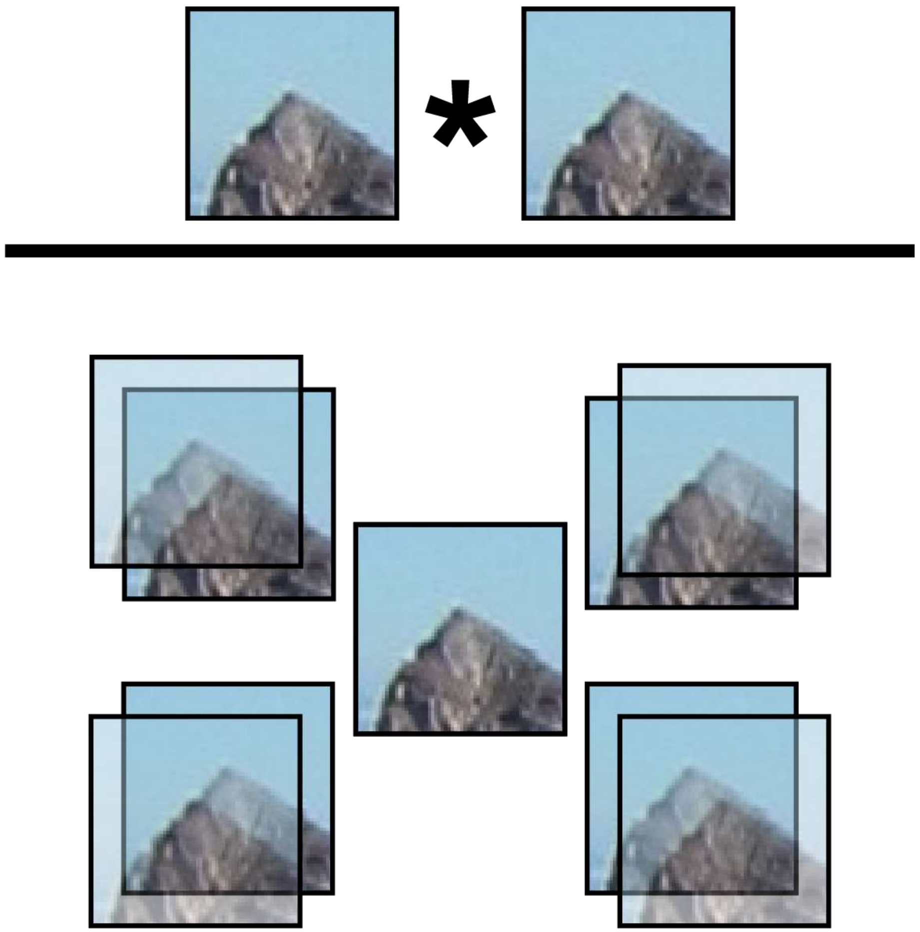


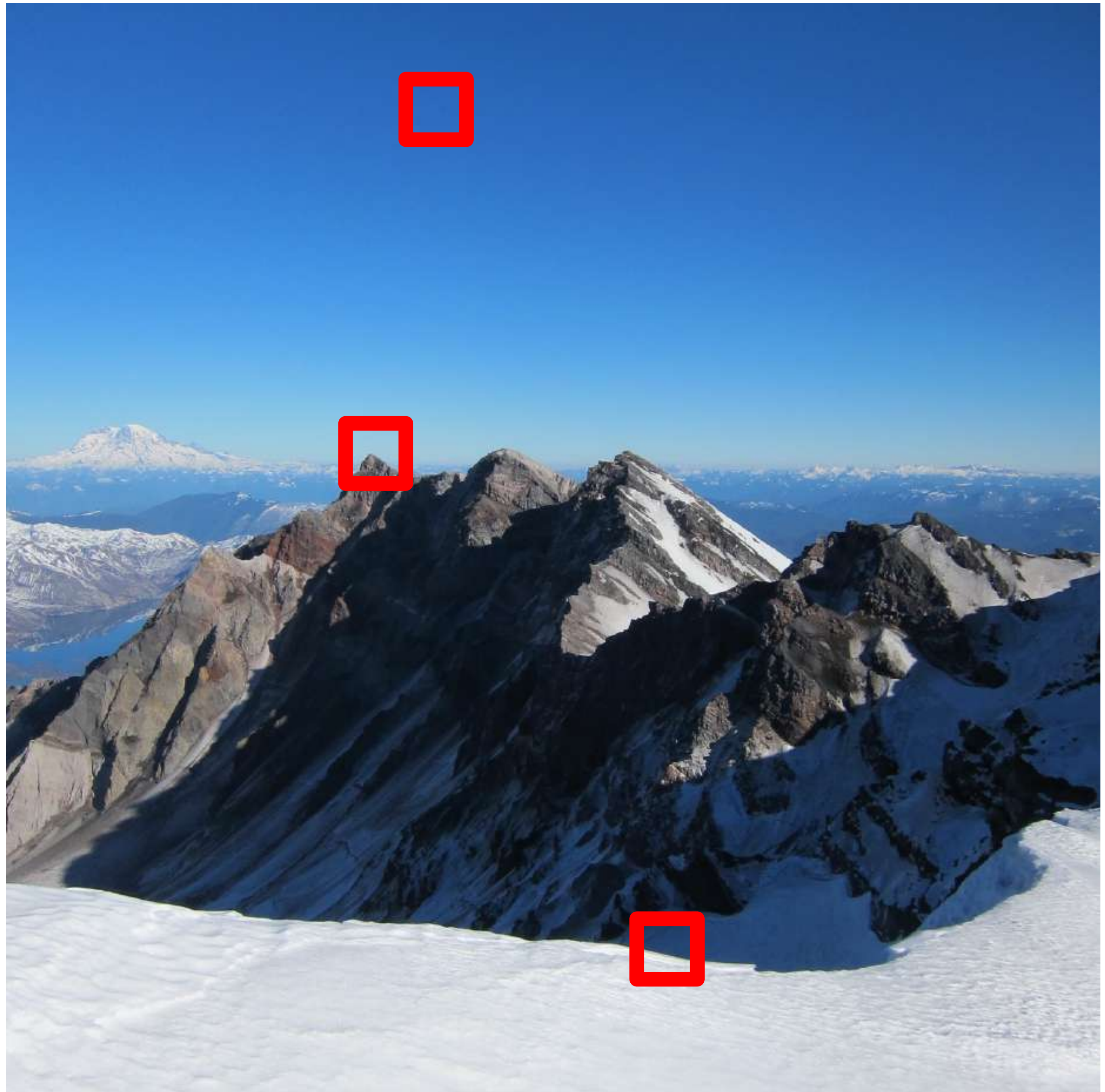
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Self-difference

Corner: mostly high





Co-financed by the European Union
Connecting Europe Facility

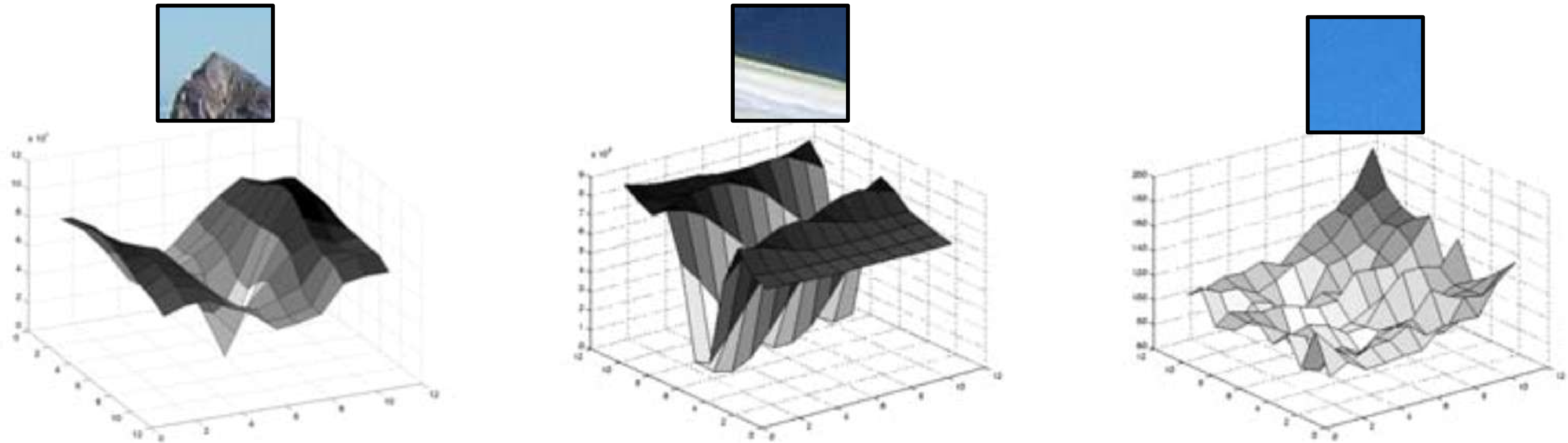
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Self-difference

Sky: low everywhere

Edge: low along edge

Corner: mostly high



Today's Agenda

- Features
- Self-difference
- Harris corner detection

Self-difference is still expensive

- $\Sigma_{\mathbf{d}}\Sigma_{x,y}\, w(x,y)\ (\ I(x+\mathbf{d}_x, y+\mathbf{d}_y) - I(x,y)\ )^2$
- Lots of summing
- Need an approximation
  - Use Taylor Series expansion of the image function (see Szeliski book Section 7.1.1)
  - $I(x+\mathbf{d}_x, y+\mathbf{d}_y) \sim= I(x,y) + \mathbf{d}_x I_x(x,y) + \mathbf{d}_y I_y(x,y)$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Approximate self-difference

- Look at nearby gradients I_x and I_y
- If gradients are mostly zero, not a lot going on
  - Low self-difference
- If gradients are mostly in one direction, edge
  - Still low self-difference
- If gradients are in two(ish) directions, corner!
  - High self-difference, good patch!

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Approximate self-difference

- How do we tell what's going on with gradients?
- Eigenvectors/values!
- Need structure matrix for patch, just a weighted sum of nearby gradient information

$$\begin{vmatrix} \Sigma_i w_i I_x(i) I_x(i) & \Sigma_i w_i I_x(i) I_y(i) \\ \Sigma_i w_i I_x(i) I_y(i) & \Sigma_i w_i I_y(i) I_y(i) \end{vmatrix}$$

- Not as complex as it looks, weighted sum of gradients near pixel

https://en.wikipedia.org/wiki/Structure_tensor

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Structure matrix

- Weighted sum of gradient information

$$\begin{vmatrix} \Sigma_i w_i I_x(i) I_x(i) & \Sigma_i w_i I_x(i) I_y(i) \\ \Sigma_i w_i I_x(i) I_y(i) & \Sigma_i w_i I_y(i) I_y(i) \end{vmatrix}$$

- Can use Gaussian weighting
- Eigenvectors/values of this matrix summarize the distribution of the gradients nearby
- λ_1 and λ_2 are eigenvalues
  - λ_1 and λ_2 both small: no gradient
  - $\lambda_1 >> \lambda_2$ or $\lambda_1 << \lambda_2$: gradient in one direction
  - λ_1 and λ_2 both large: multiple gradient directions, corner

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Estimating eigenvalues

- Use determinant and trace:
  - $\det(S) = \lambda_1 * \lambda_2$
  - $\mathrm{trace}(S) = \lambda_1 + \lambda_2$
- Response function (R score): $\det(S) - \kappa\, \mathrm{trace}(S)^2 = \lambda_1\lambda_2 - \kappa(\lambda_1+\lambda_2)^2$
- If this score is large, both λ_1 and λ_2 are large

Co-financed by the European Union
Connecting Europe Facility

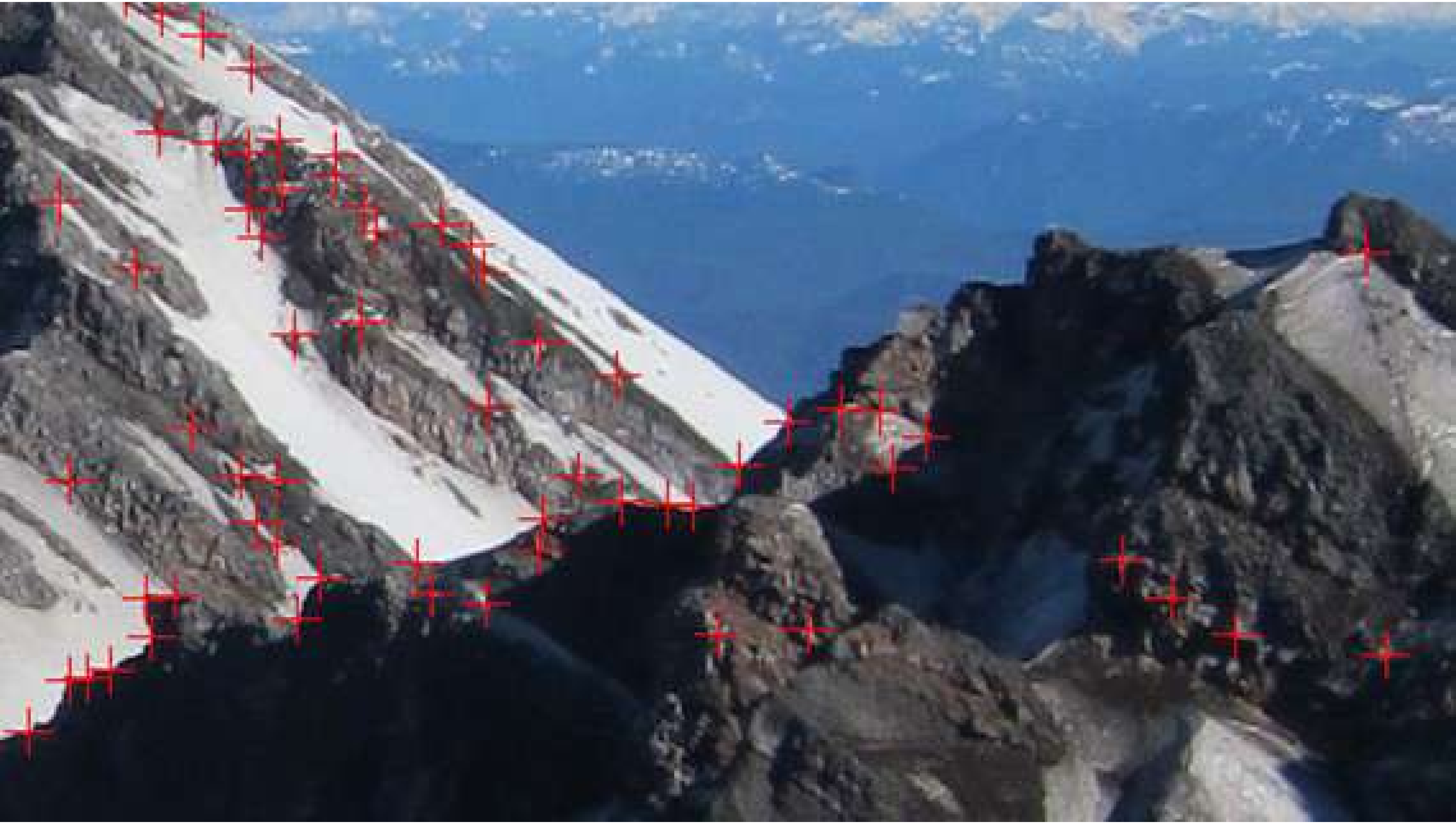
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Harris Corner Detector

- Calculate derivatives I_x and I_y
- Calculate 3 measures I_xI_x, I_yI_y, I_xI_y
- Calculate weighted sums
  - Want a (weighted) sum of nearby pixels, guess what this is?
  - Gaussian – or weights could just be ones
- Estimate response based on R score
- Non-max suppression in a neighborhood

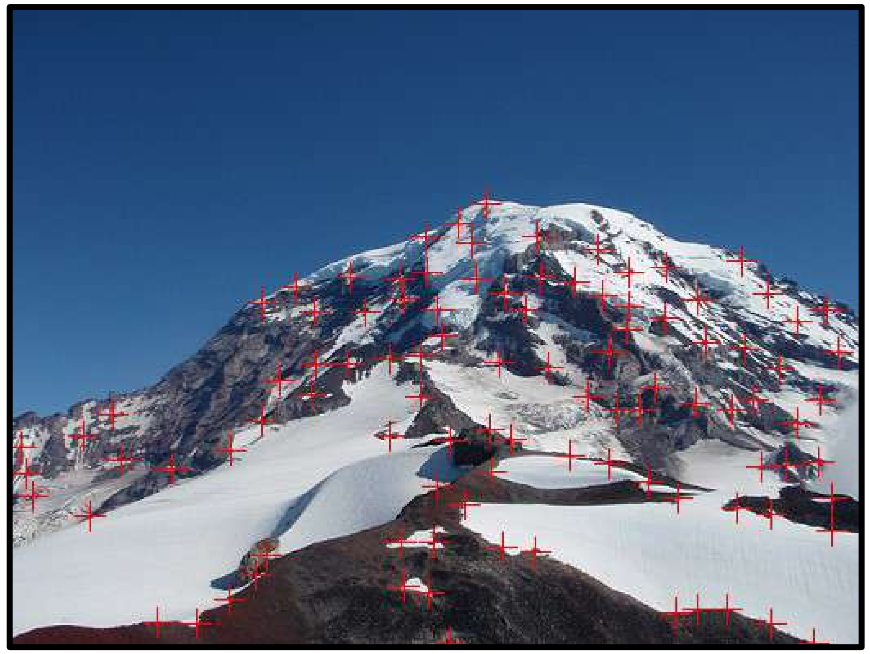
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423



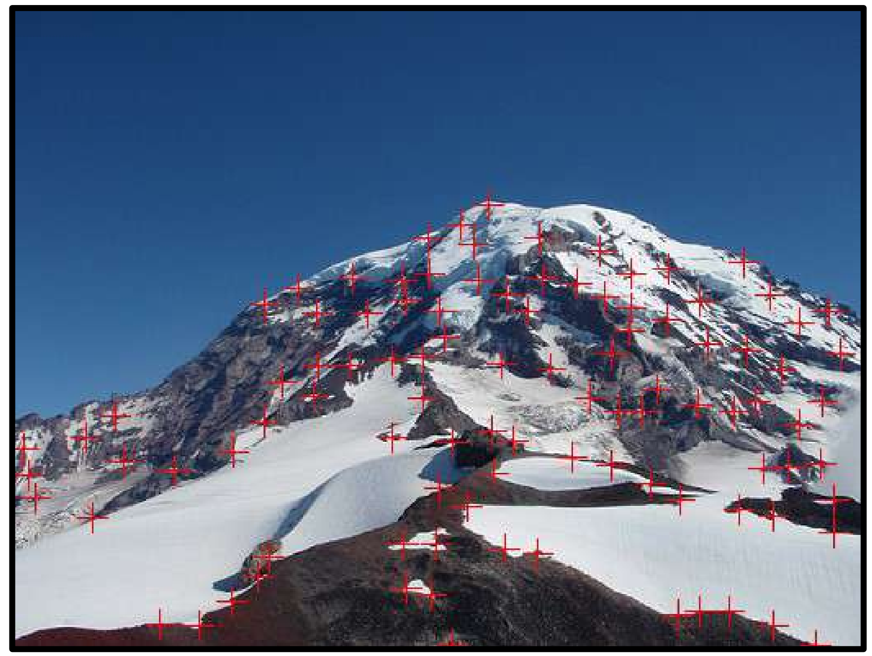
Ok, we found corners, now what?

- Need to match image patches to each other
- Need to figure out transform between images



Ok, we found corners, now what?

- Need to match image patches to each other
  - What is a patch? How do we look for matches? Pixels?
- Need to figure out transform between images
  - How can we transform images?
  - How do we solve for this transformation given matches?



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 8: Feature Descriptors and Image Transforms

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Features
- Self-difference
- Harris corner detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Basic feature descriptor and matching
- Histogram of Oriented Gradients
- SIFT
- Image transformations
- Estimate transformations

[material based on Joseph Redmon’s course]

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

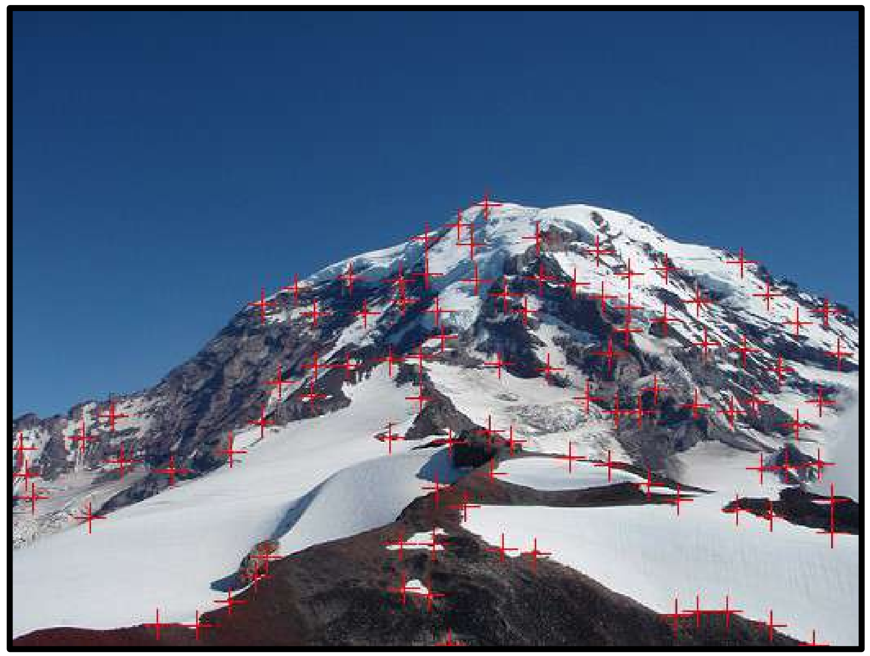
Today’s Agenda

- Basic feature descriptor and matching
- Histogram of Oriented Gradients
- SIFT
- Image transformations
- Estimate transformations

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ok, we found corners, now what?

- Need to match image patches to each other
- Need to figure out transform between images

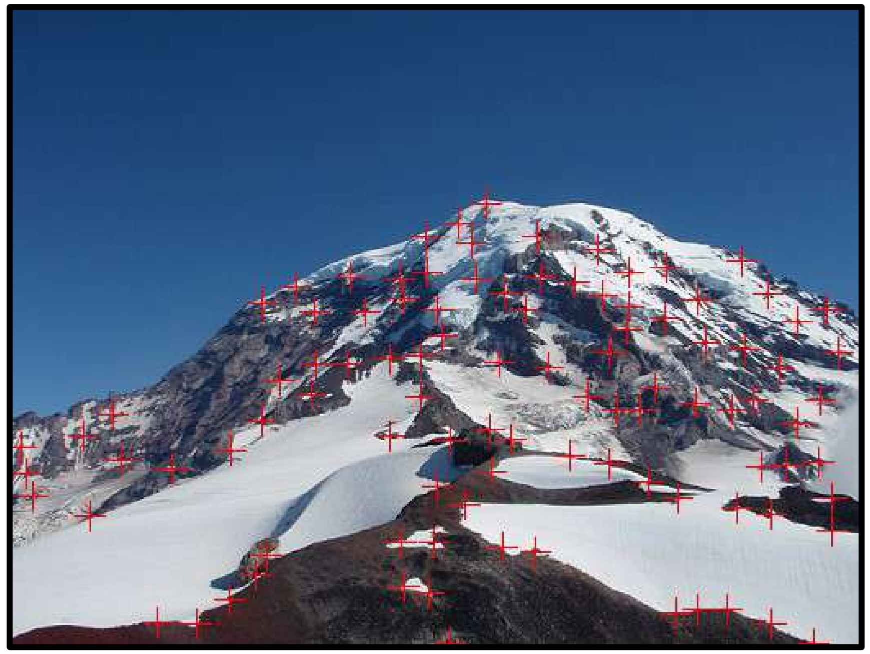


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ok, we found corners, now what?

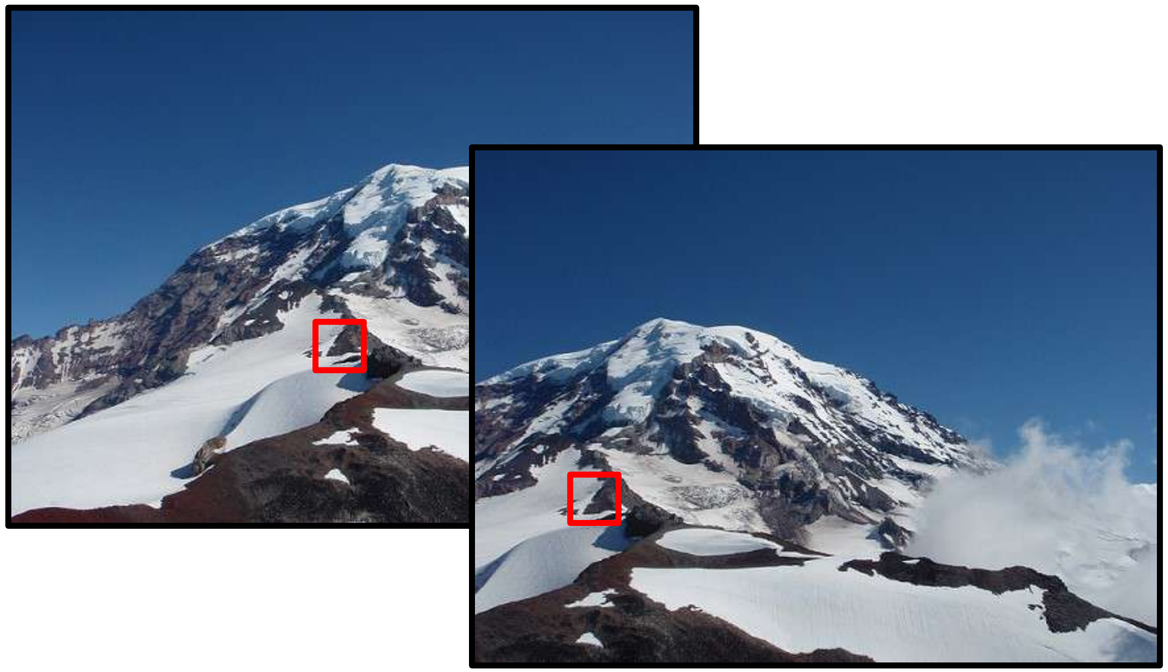
- Need to match image patches to each other
  - What is a match? How do we look for matches? Pixels?
- Need to figure out transform between images



This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

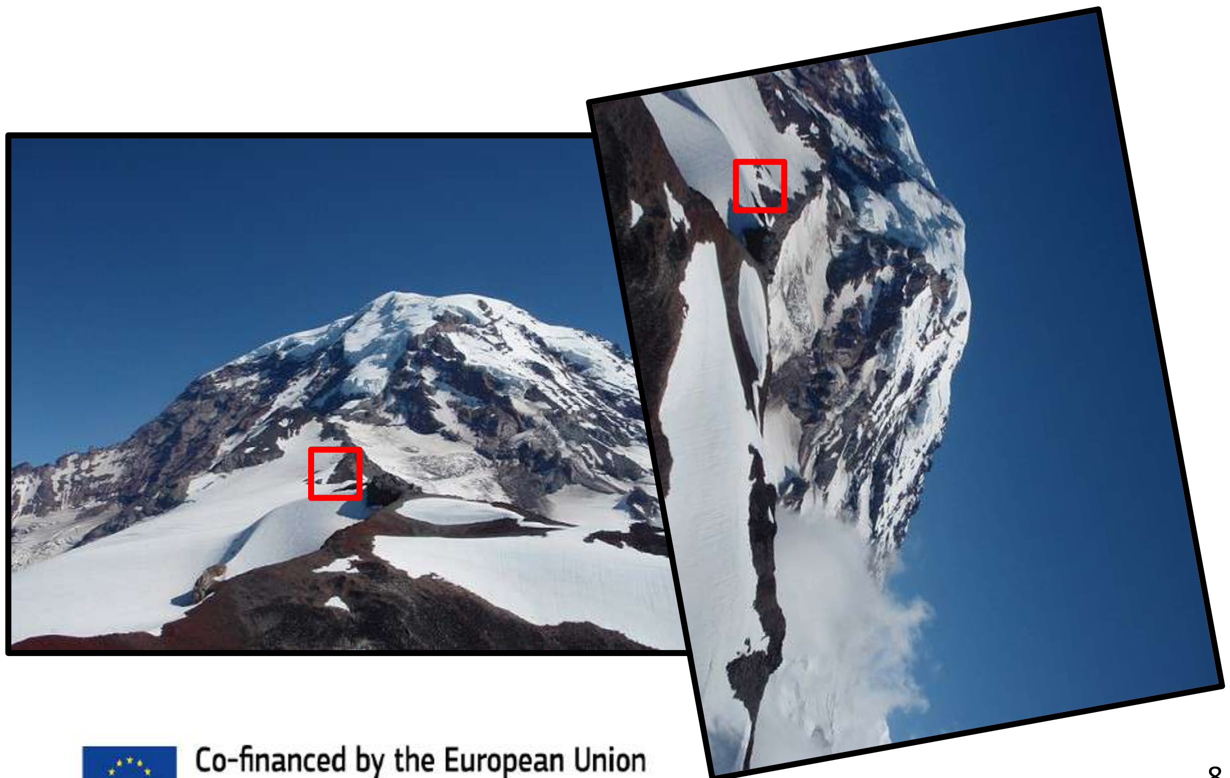
Matching patches: descriptors!

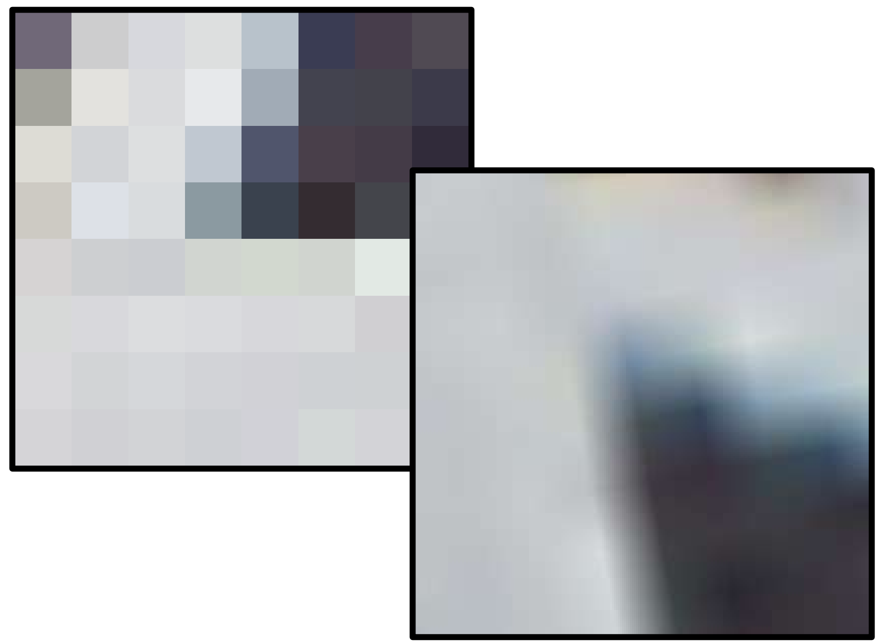
- We want a way to represent an image patch
- Can be very simple, just pixels!
- Finding matching patch is easy, distance metric:
  - $\Sigma_{x,y} (I(x,y) - J(x,y))^2$
  - What problems are there with just using pixels?



Matching patches: descriptors!

- We want a way to represent an image patch
- Can be very simple, just pixels!
- Finding matching patch is easy, distance metric:
  - $\Sigma_{x,y} (I(x,y) - J(x,y))^2$
  - Not invariant to some image transformations (e.g. rotation, scaling) !



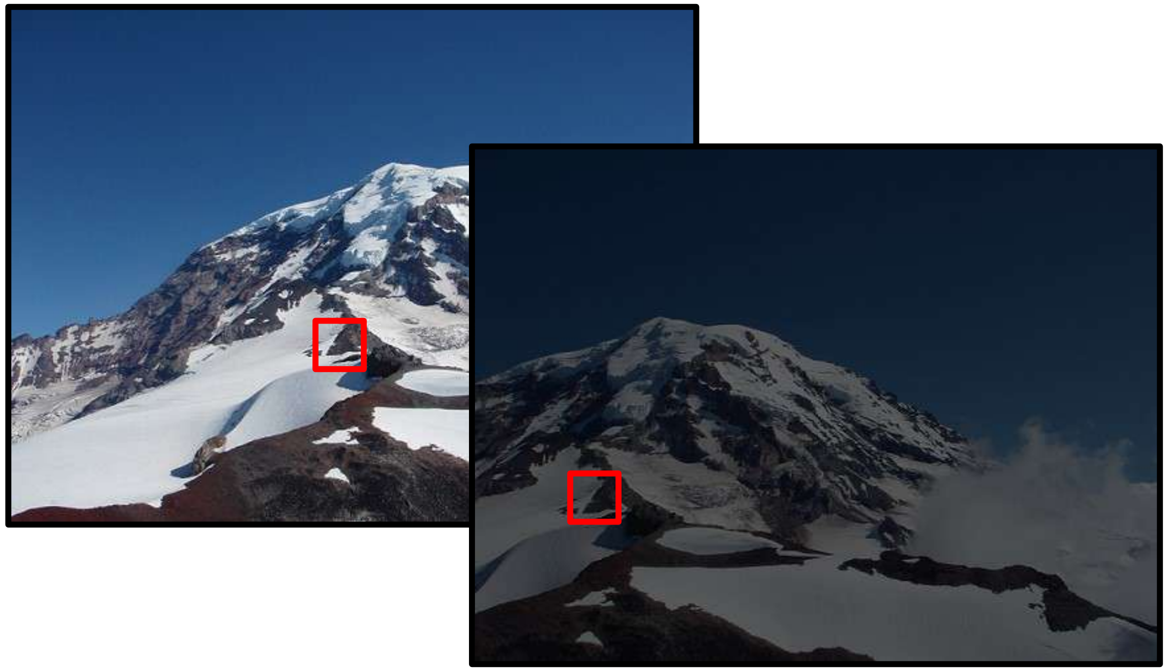


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Matching patches: descriptors!

- We want a way to represent an image patch
- Can be very simple, just pixels!
- Finding matching patch is easy, distance metric:
  - $\Sigma_{x,y} (I(x,y) - J(x,y))^2$
  - Not invariant to lighting changes !



Matching patches: descriptors!

- We want feature descriptors invariant to lighting and image transforms !
- Descriptors can be more complex
  - Gradient information
  - How much context?

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Basic feature descriptor and matching
- Histogram of Oriented Gradients
- SIFT
- Image transformations
- Estimate transformations

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Histogram of Oriented Gradients (HOG)

- By Dalal and Triggs 2005
- Better image descriptor
- Not reliant on magnitude, just direction
  - Invariant to some lighting changes
- They used it to train an SVM to recognize people

Binning from http://aishack.in/tutorials/sift-scale-invariant-feature-transform-features/

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Histogram of Oriented Gradients (HOG)

Steps to calculate HOG Feature Descriptor

1. Compute gradients

2. Bin gradient directions to create histogram

3. Normalize histograms of gradients

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Histogram of Oriented Gradients (HOG)

Steps to calculate HOG Feature Descriptor

1. Compute gradients

Gaussian smoothing (experimented with various σ), followed by a derivative filter

- σ =0 , i.e., no smoothing gave best results
- 1D filter [-1, 0, 1] gave best results

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

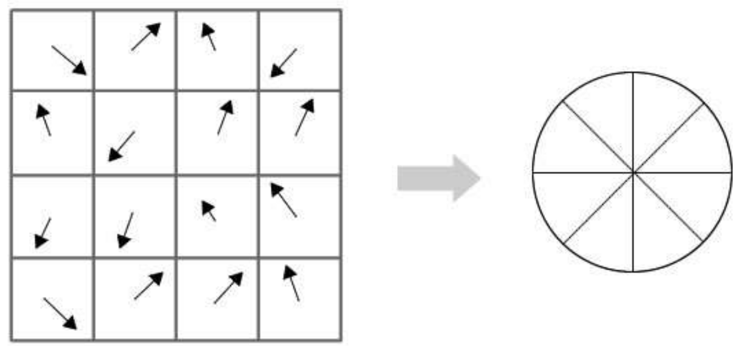
Histogram of Oriented Gradients (HOG)

Steps to calculate HOG Feature Descriptor

2. Bin gradient directions to create histogram

Split image into 8x8 'cells' and compute histogram for each cell

- Unsigned gradients, i.e., θ=0-180 degrees gave best results
- 9 bins gave best results



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Histogram of Oriented Gradients (HOG)

Steps to calculate HOG Feature Descriptor

3. Normalize histograms of gradients

Gather overlapping ‘cells’ into ‘blocks’, concatenate histograms and normalize

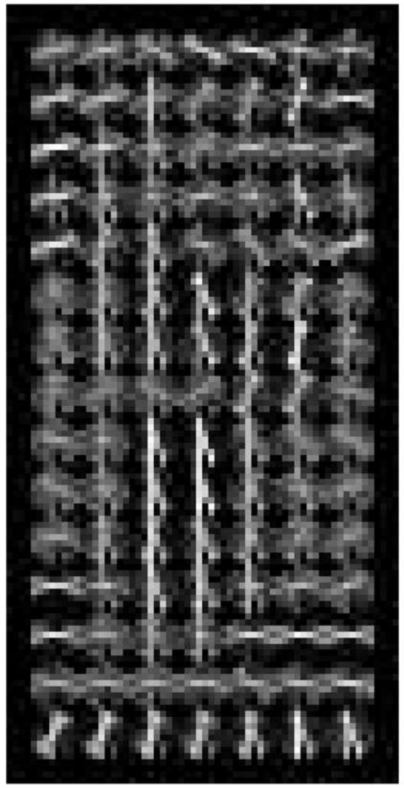
- 16x16 blocks of 4 (2x2) cells gave best results
- L2 Normalization gave best results

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

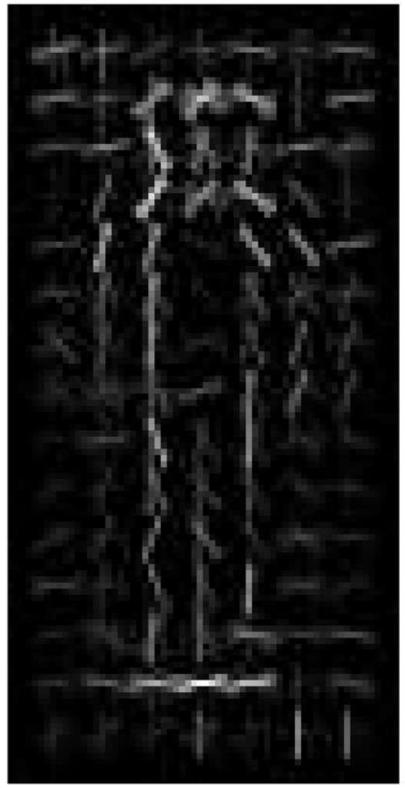
Histogram of Oriented Gradients (HOG)

For each training image of 64x128 there are 7x15 blocks, so the overall descriptor is 7x15x36 = 3780 dimensions

Training image



HOG descriptor of the image visualized for each 16x16 block

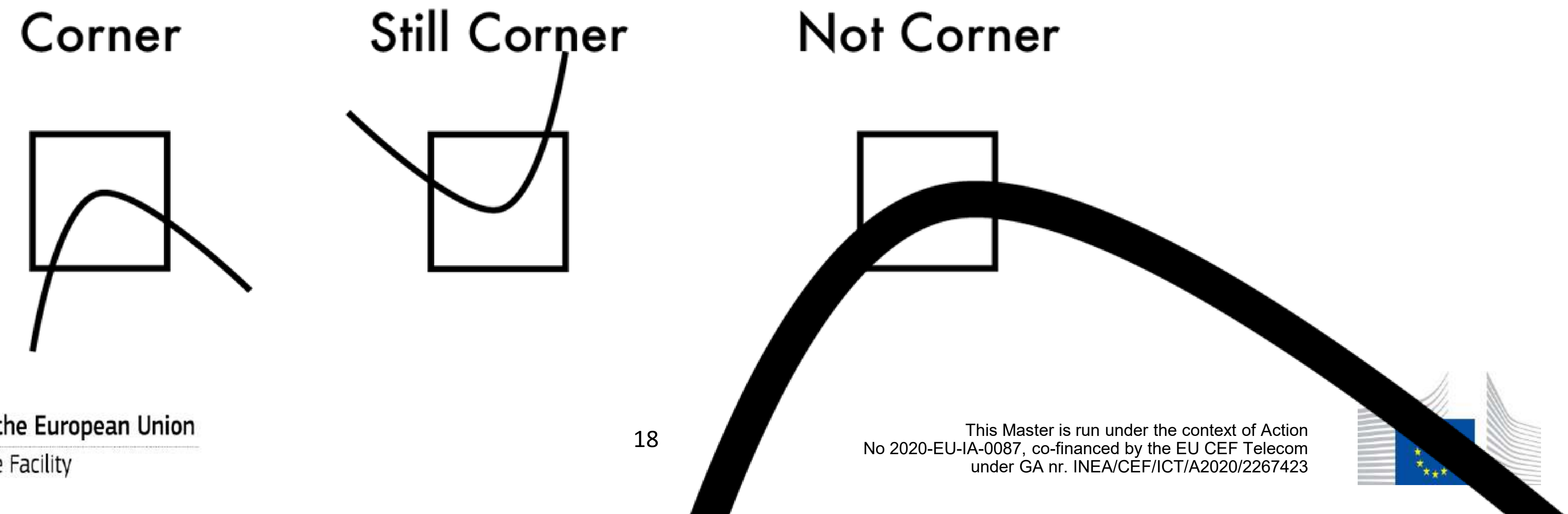


Descriptor weighted by the SVM weights

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

This is as good as it gets ?

- Not so fast…
- Harris has some issues:
  - Corner detection is rotation invariant
  - Harris not invariant to scale
- Descriptors are also hard
  - Just looking at pixels is not rotation invariant!
  - HOG also not rotation invariant



Today’s Agenda

- Basic feature descriptor and matching
- Histogram of Oriented Gradients
- SIFT
- Image transformations
- Estimate transformations

Co-financed by the European Union
Connecting Europe Facility

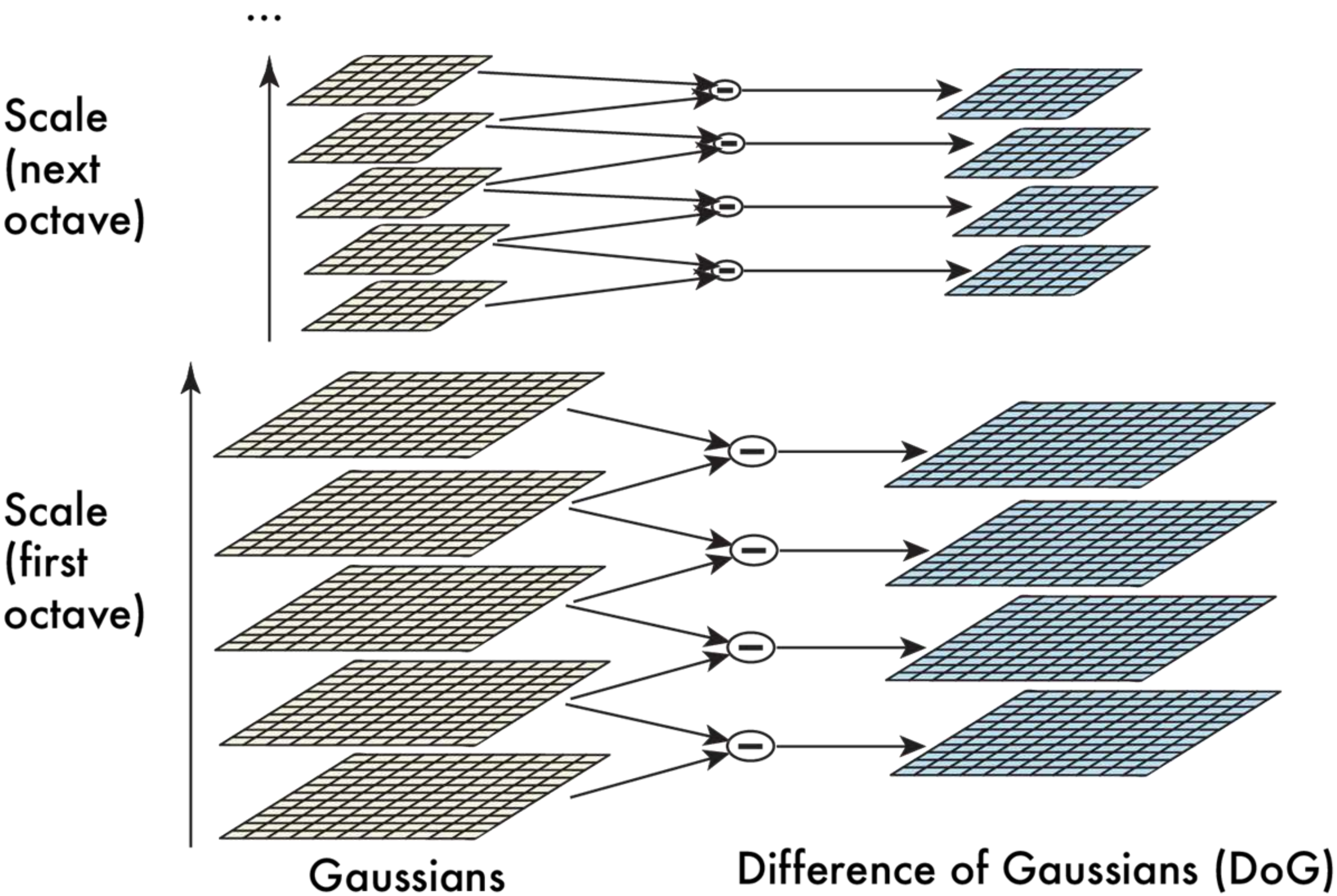
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Want features invariant to scaling, rotation, etc.

- Scale Invariant Feature Transform (SIFT)
    - Lowe et al. 2004, many images from that paper
- Get scale-invariant response map
- Find keypoints
- Extract rotation-invariant descriptors
    - Normalize based on orientation
    - Normalize based on lighting

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Extract DoG features at multiple scales

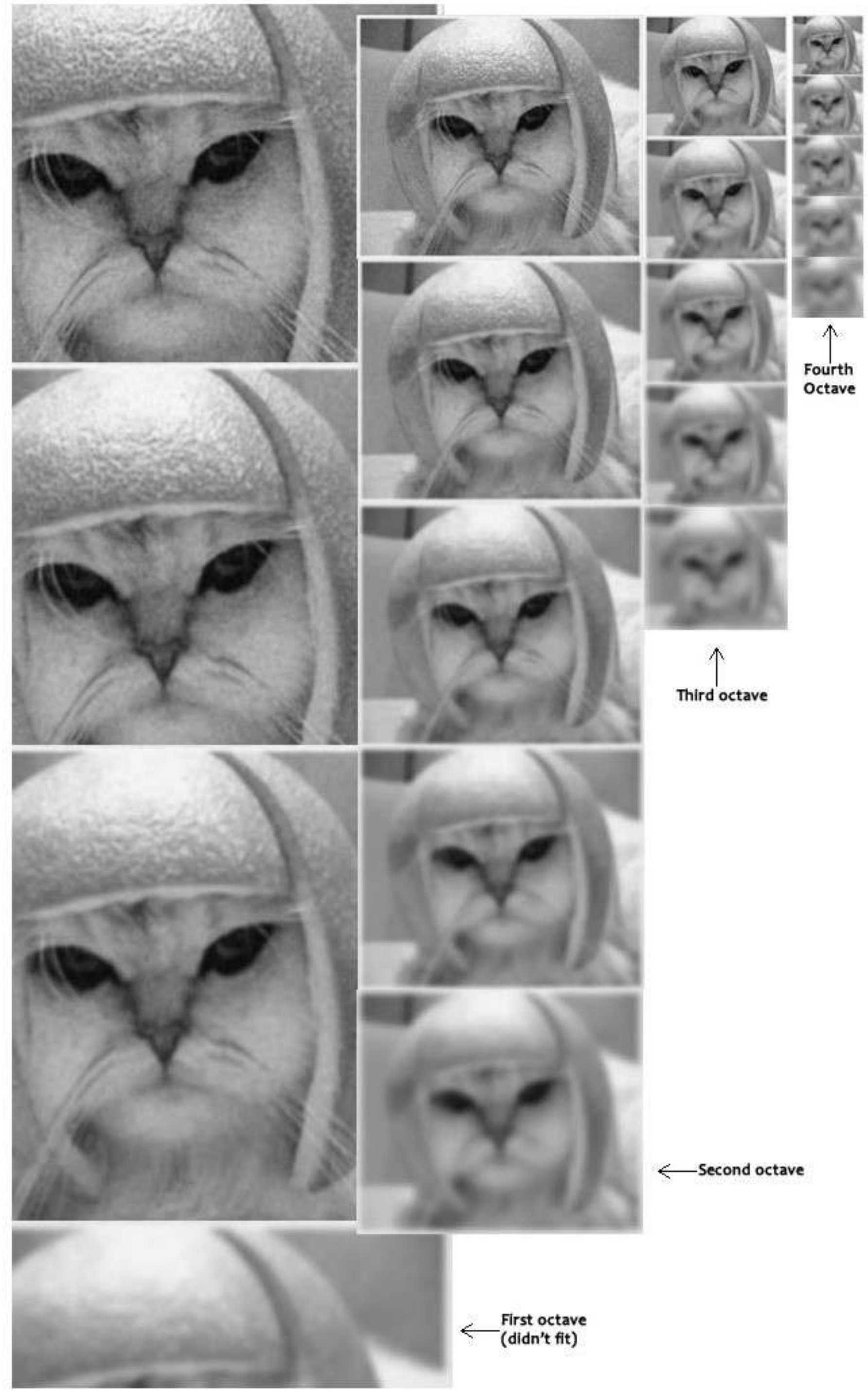


Scale space

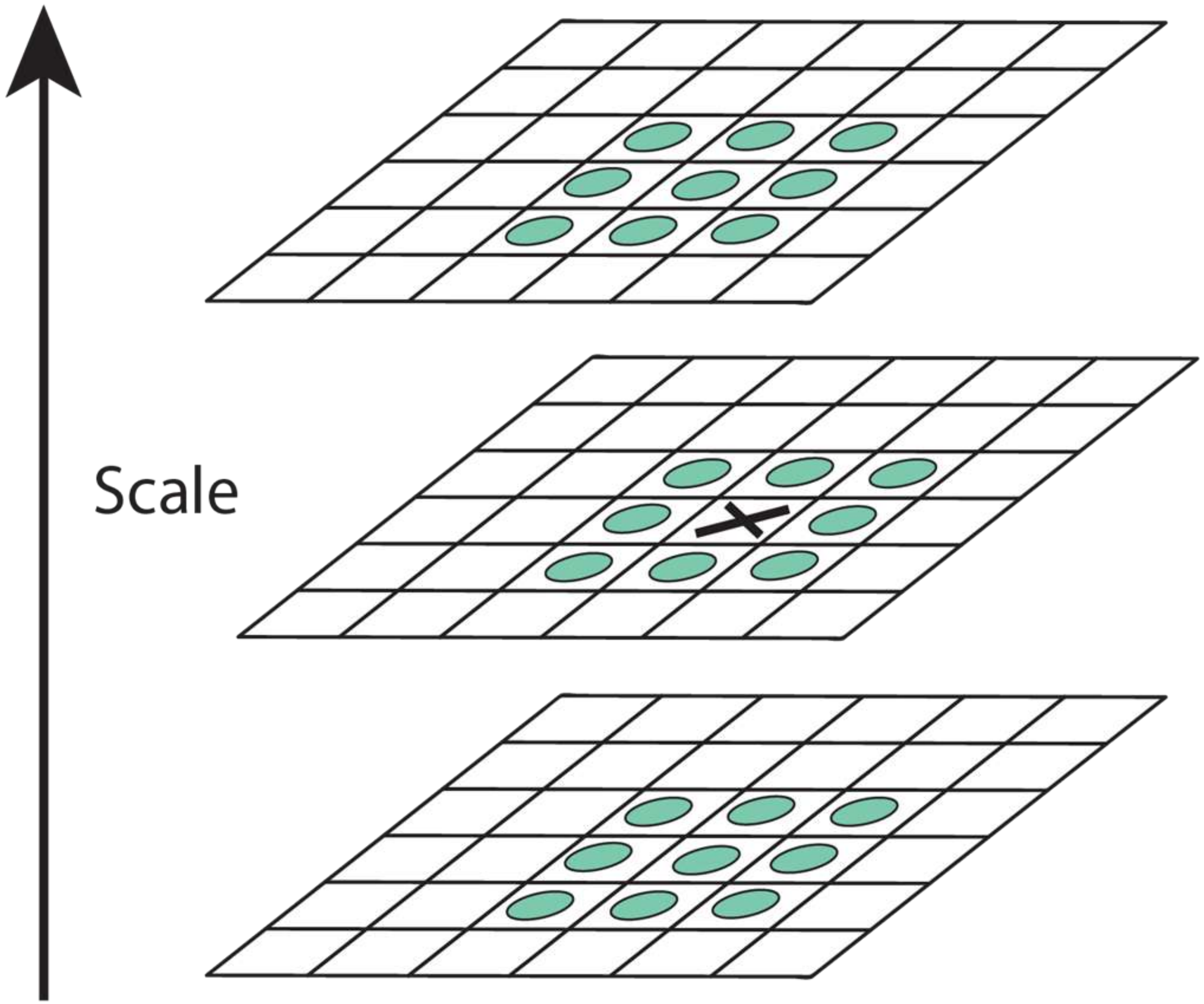
scale

octave

0.707107	1.000000	1.414214	2.000000	2.828427
1.414214	2.000000	2.828427	4.000000	5.656854
2.828427	4.000000	5.656854	8.000000	11.313708
5.656854	8.000000	11.313708	16.000000	22.627417



Find local-maxima in location and scale



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Throw out weak responses and edges

- Estimate gradients
  - Similar to before, look at nearby responses
  - Not whole image, only a few points! Faster!
  - Throw out weak responses
- Find cornery things
  - Same deal, structure matrix, use det and trace information
  - r : ratio of larger to smaller eigenvalue

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r},$$

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r+1)^2}{r}$$

Find main orientation of patches

- Look at weighted histogram of nearby gradients
  - Any gradient within 80% of peak gets its own descriptor
    - Multiple keypoints per pixel
  - Descriptors are normalized based on main orientation

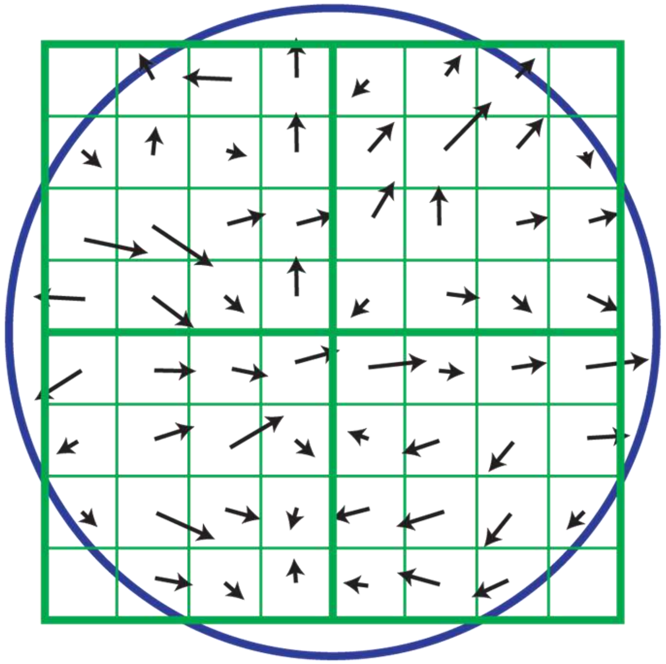
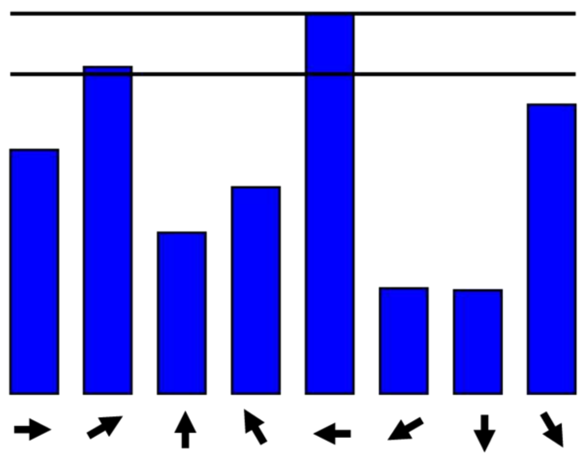


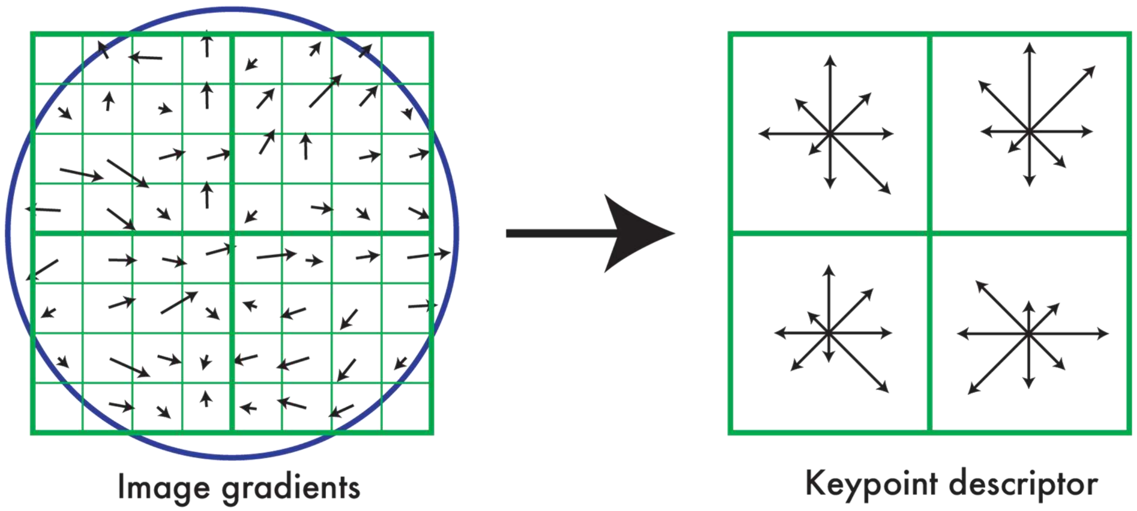
Image gradients



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Keypoints are normalized gradient histograms

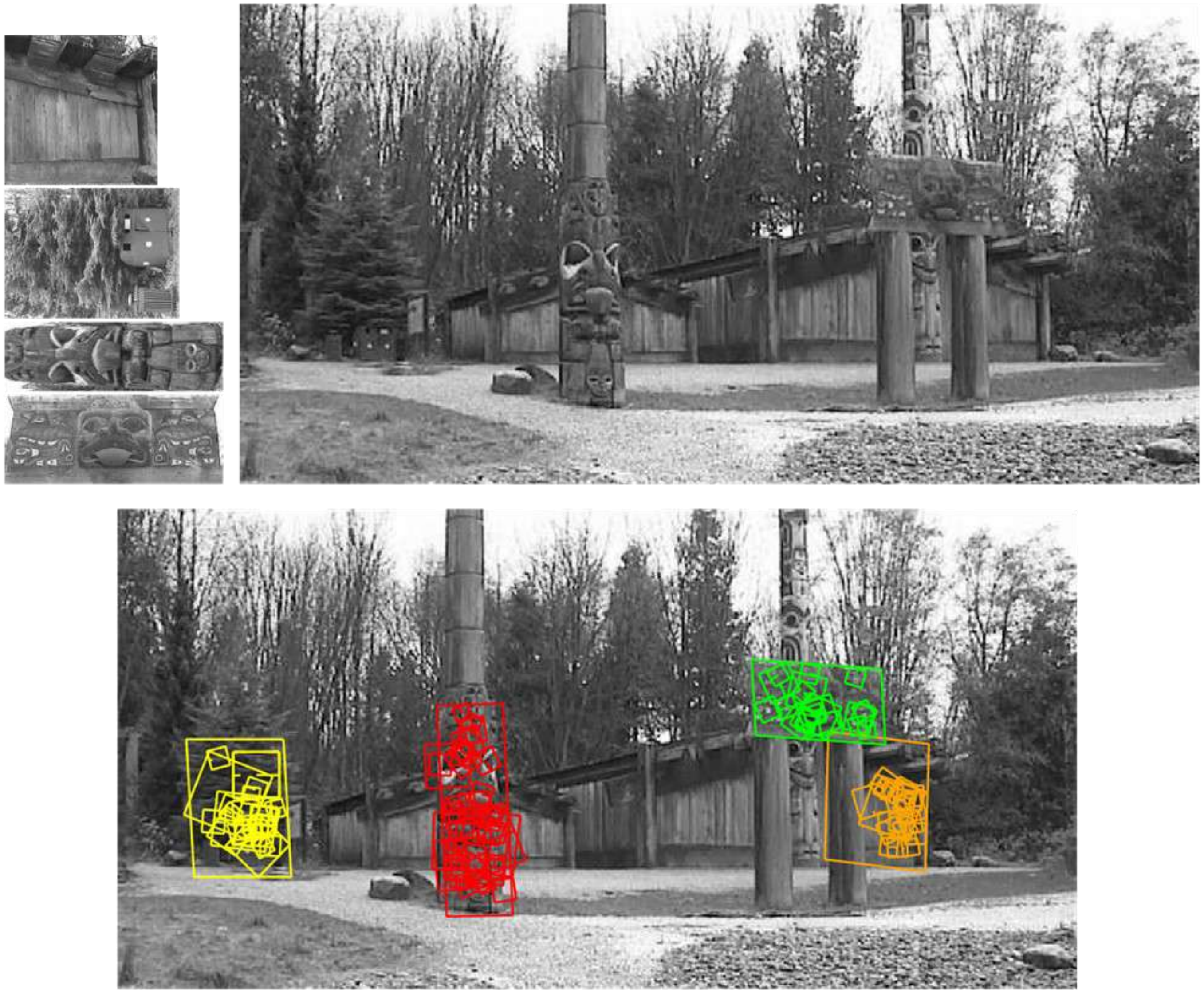
- Divide into subwindows (4x4)
- Bin gradients within subwindow, get histogram
    - Normalize to unit length
    - Clamp at maximum .2
    - Normalize again
    - Helps with lighting changes!



SIFT is great!

- Find good keypoints, describe them
- Finding objects, recognition, panoramas, etc.

SIFT is great!



Today’s Agenda

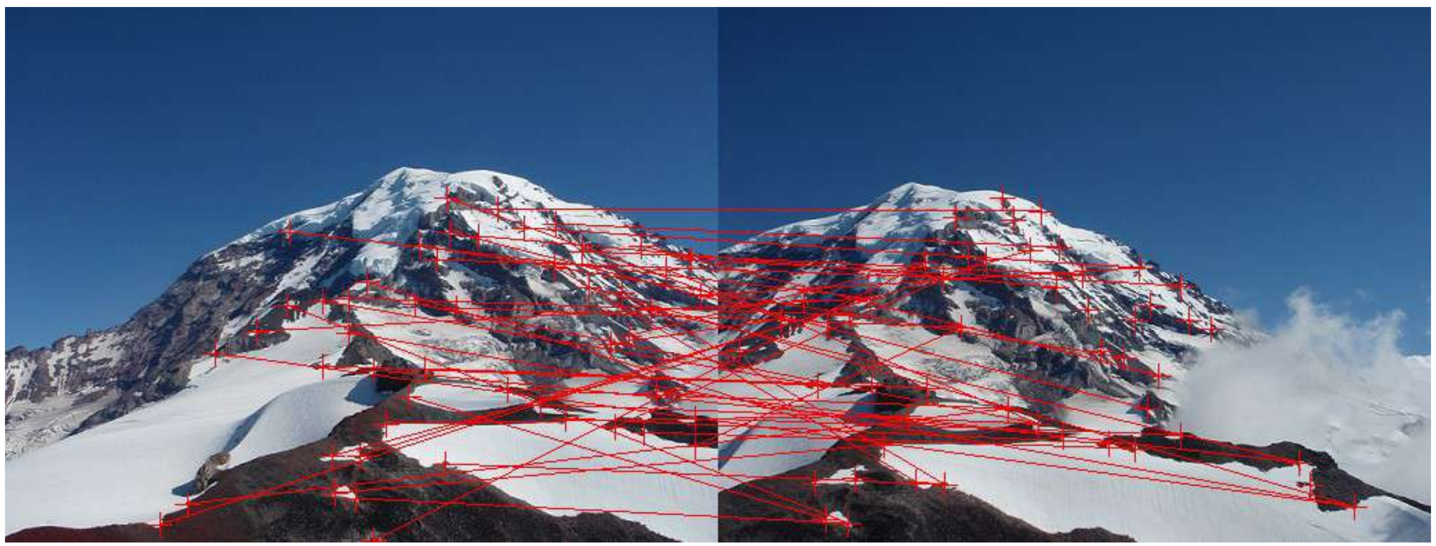
- Basic feature descriptor and matching
- Histogram of Oriented Gradients
- SIFT
- Image transformations
- Estimate transformations

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Matching patches: descriptors!

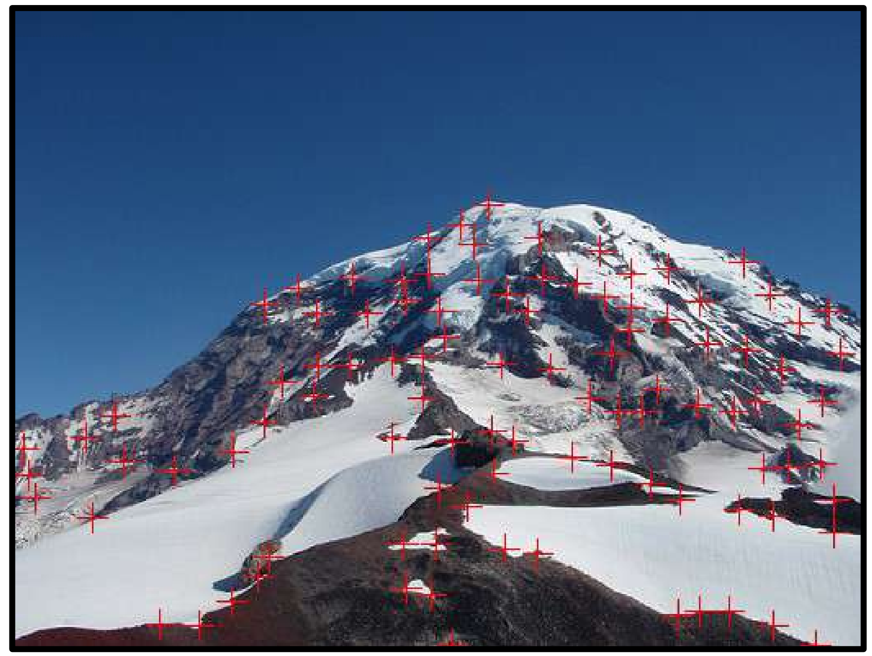
- Already have our patches that are likely "unique"-ish
- Loop over good patches in one image
  - Find best match in other image
- Do something with them?



This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ok, we found corners, now what?

- Need to match image patches to each other
- Need to figure out transform between images

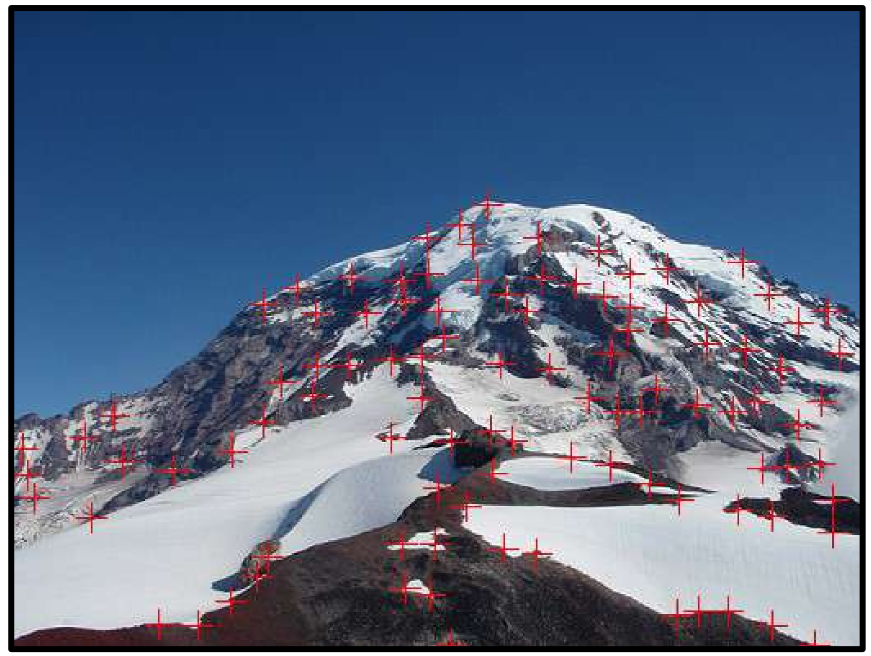


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ok, we found corners, now what?

- Need to match image patches to each other
- Need to figure out transform between images
  - How can we transform images?
  - How do we solve for this transformation given matches?



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How can we transform images?

- Need to warp one image into the other
- Many different image transforms
  - Nested hierarchy of transformations

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How can we transform images?

- **x** is a point in our image where:
  - **x** = (x, y) or in matrix terms

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Say we want new coordinate system

- Map points from one image into another
- Often we can use matrix operations
- Given a point x, map to new point x' using M

$$\mathbf{x}' = \mathbf{M}\,\mathbf{x}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Scaling is just a matrix operation

- Map points from one image into another
- Often we can use matrix operations
- Given a point x, map to new point x' using M

$$\mathbf{x'} = \mathbf{S}\,\mathbf{x}$$

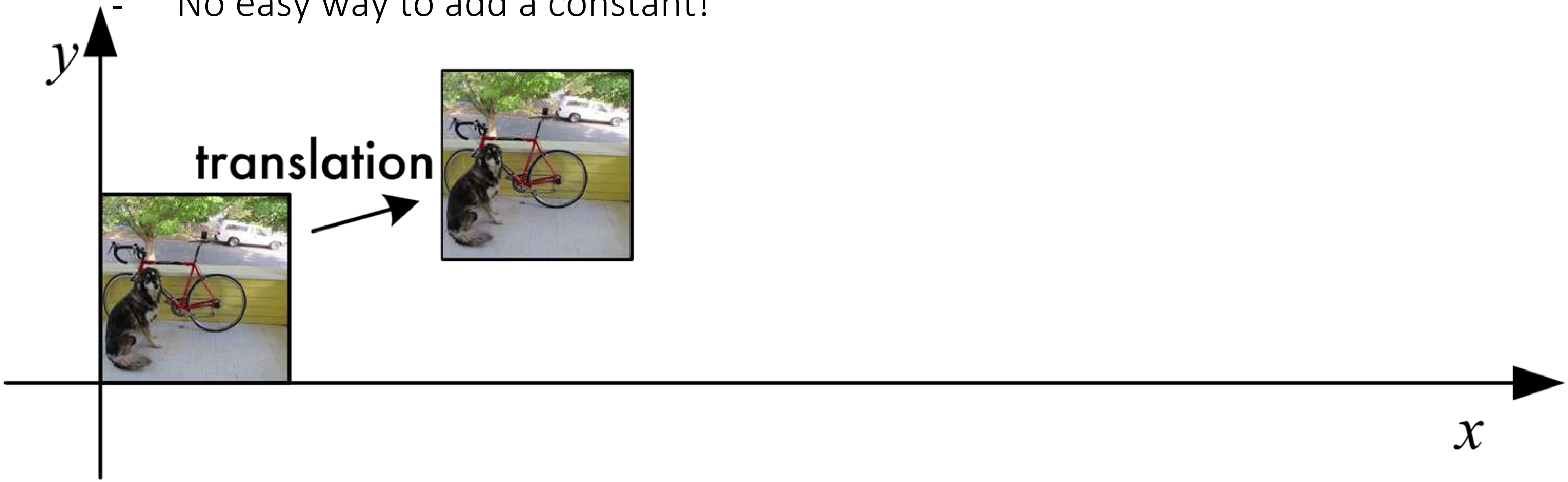
$$\mathbf{x'} = \begin{bmatrix} S & 0 \\ 0 & S \end{bmatrix} \mathbf{x}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Translation is harder...

- $x' = M\,x$
  - Want to move x' by dx and y' by dy
  - How do we pick **M**?
  - Can only add up multiples of x or y
  - No easy way to add a constant!



Translation: add another row

- $\bar{\mathbf{x}}$ is $\mathbf{x}$ but with an added 1
- *Augmented vector*

$$\bar{\mathbf{x}} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Translation: add another row

- $\bar{\mathbf{x}}$ is $\mathbf{x}$ but with an added 1
- *Augmented vector*
- Now translation is easy

$$\bar{\mathbf{x}} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{x}' = \begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Reminder, **I** = Identity

Common to just use **I** as a generic, whatever size identity fits here.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ & & & \ldots & \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

$$\mathbf{I}_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{I}_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Translation: add another row

- $\bar{\mathbf{x}}$ is $\mathbf{x}$ but with an added 1
- *Augmented vector*
- Now translation is easy

$$\bar{\mathbf{x}} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{x}' = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{x}' = [\,\mathbf{I} \quad \mathbf{t}\,]\,\bar{\mathbf{x}}$$

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Translation: add another row

- $\bar{\mathbf{x}}$ is $\mathbf{x}$ but with an added 1
- *Augmented vector*
- Now translation is easy
- x' = 1*x + 0*y + dx*1
- y' = 0*x + 1*y + dy*1

$$\bar{\mathbf{x}} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{x}' = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{x}' = \begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Translation: add another row

- $\bar{\mathbf{x}}$ is $\mathbf{x}$ but with an added 1
- *Augmented vector*
- Now translation is easy
- x' = 1*x + 0*y + dx*1
- y' = 0*x + 1*y + dy*1

$$\bar{\mathbf{x}} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

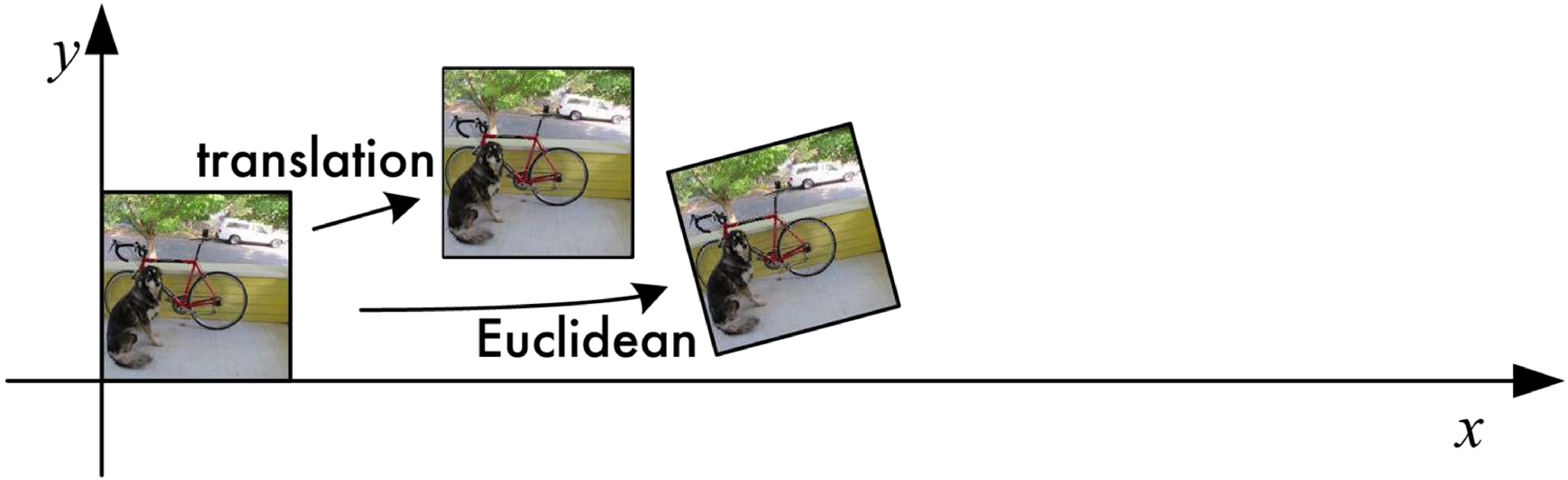
$$\mathbf{x}' = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{x}' = [\,\mathbf{I} \quad \mathbf{t}\,]\, \bar{\mathbf{x}}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Euclidean: rotation + translation

- Want to translate and rotate at same time
- Still just matrix operation



Euclidean: rotation + translation

- Want to translate and rotate at same time
- Still just matrix operation

$$\mathbf{x}' = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Euclidean: rotation + translation

- Want to translate and rotate at same time
- Still just matrix operation
- **R** is rotation matrix, **t** is translation

$$\mathbf{x}' = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}}$$

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Euclidean: rotation + translation

- Want to translate and rotate at same time
- Still just matrix operation
- **R** is rotation matrix, **t** is translation

$$\mathbf{x}' = \begin{bmatrix} \cos\theta & -\sin\theta & dx \\ \sin\theta & \cos\theta & dy \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

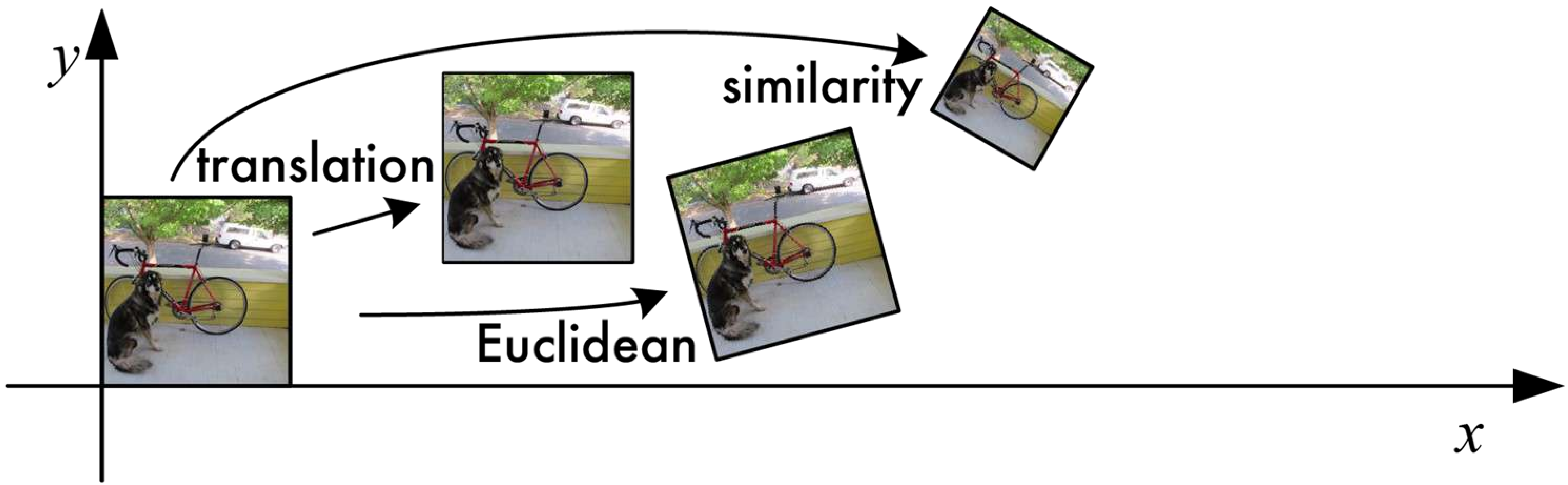
$$\mathbf{x}' = [\,\mathbf{R} \;\; \mathbf{t}\,]\,\bar{\mathbf{x}}$$

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility

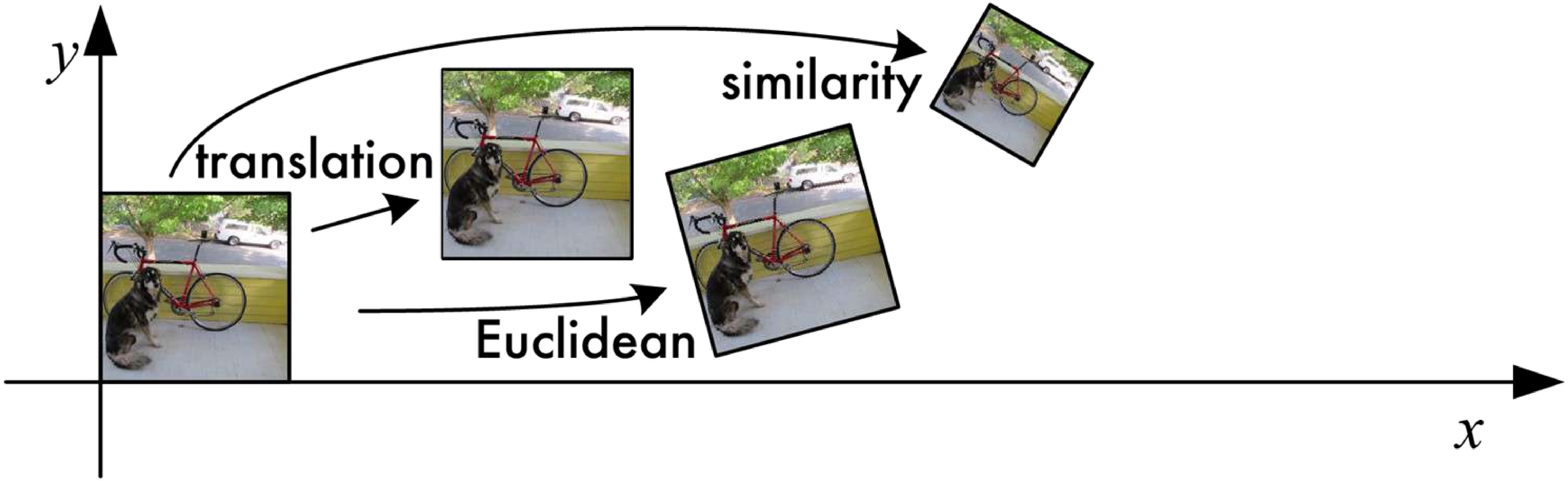
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Similarity: scale, rotate, translate



Similarity: scale, rotate, translate

$$\mathbf{x}' = [\, s\mathbf{R} \quad \mathbf{t} \,]\, \bar{\mathbf{x}}$$

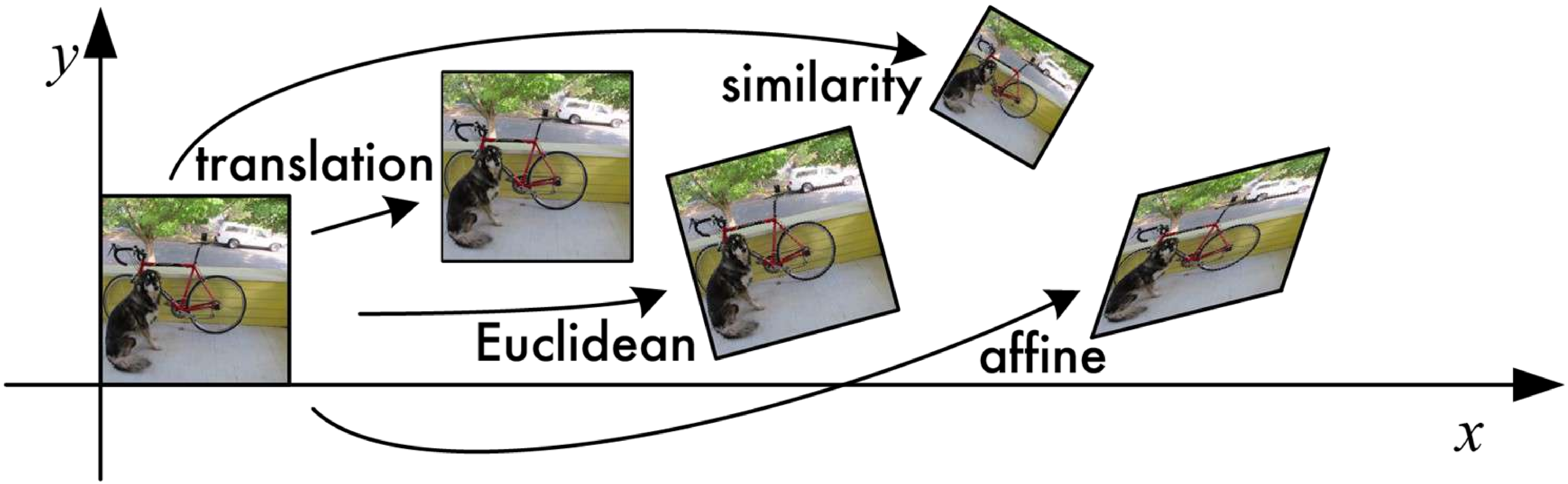


Similarity: scale, rotate, translate

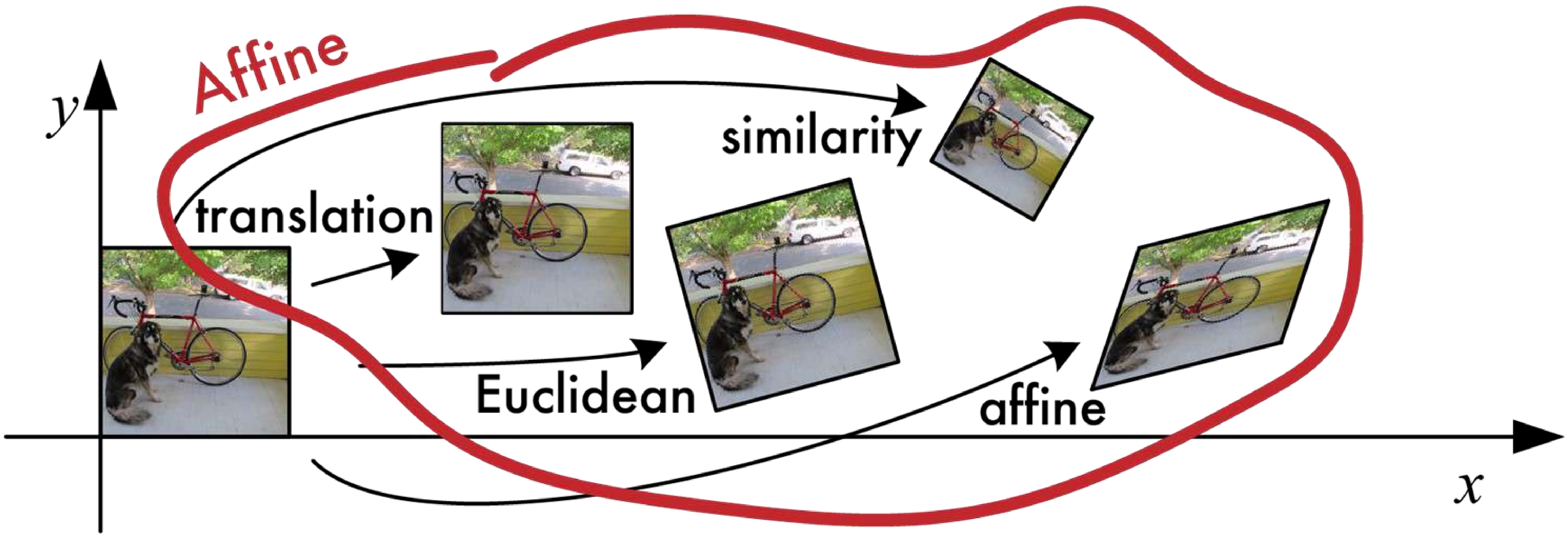
$$\mathbf{x}' = [\,s\mathbf{R} \quad \mathbf{t}\,]\,\bar{\mathbf{x}}$$

$$\mathbf{x}' = \begin{bmatrix} a & -b & dx \\ b & a & dy \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Affine: scale, rotate, translate, shear



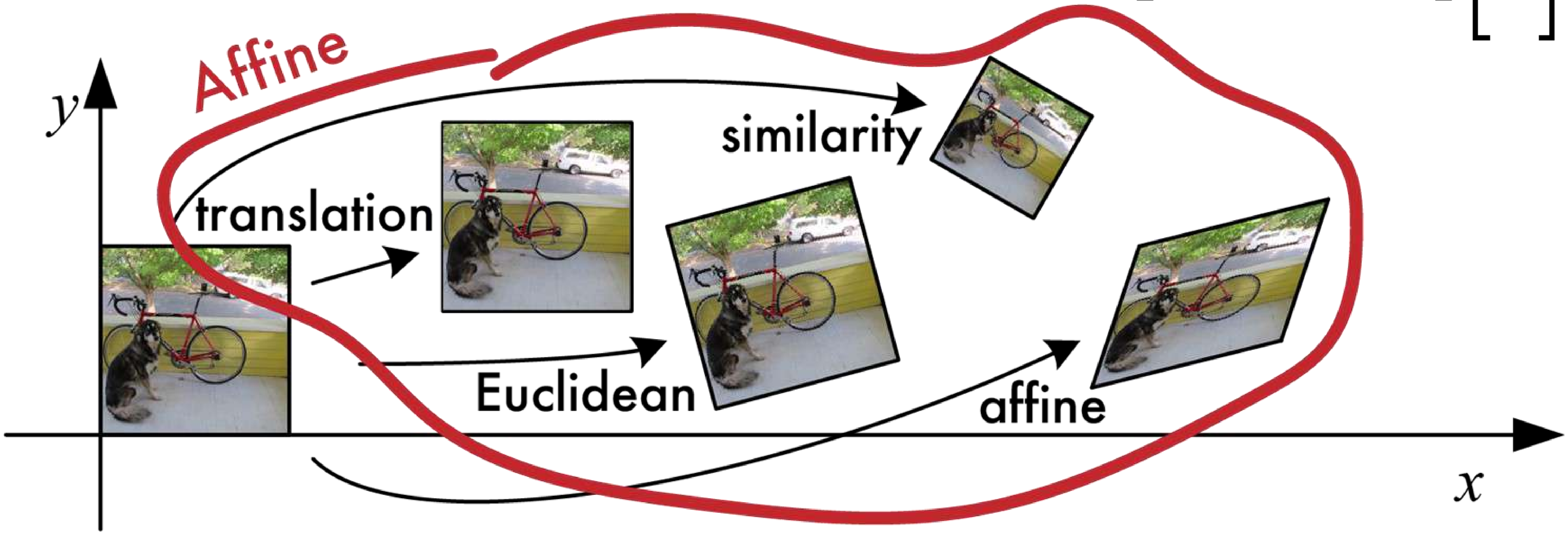
Affine: scale, rotate, translate, shear



Affine: scale, rotate, translate, shear

General case of 2x3 matrix

$$\mathbf{x}' = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$



Combinations are still affine

Say you want to translate, then rotate, then translate back, then scale.

$\mathbf{x}' = \mathbf{S}\,\mathbf{t}\,\mathbf{R}\,\mathbf{t}\,\bar{\mathbf{x}} = \mathbf{M}\,\bar{\mathbf{x}}$,

If $\mathbf{M} = (\mathbf{S}\,\mathbf{t}\,\mathbf{R}\,\mathbf{t})$

$\mathbf{M}$ is still affine transformation

Wait, but these are all 2x3, how to we multiply them together?

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Added row to transforms

$$\bar{\mathbf{x}}' = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\bar{\mathbf{x}}' = \begin{bmatrix} \cos\theta & -\sin\theta & dx \\ \sin\theta & \cos\theta & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

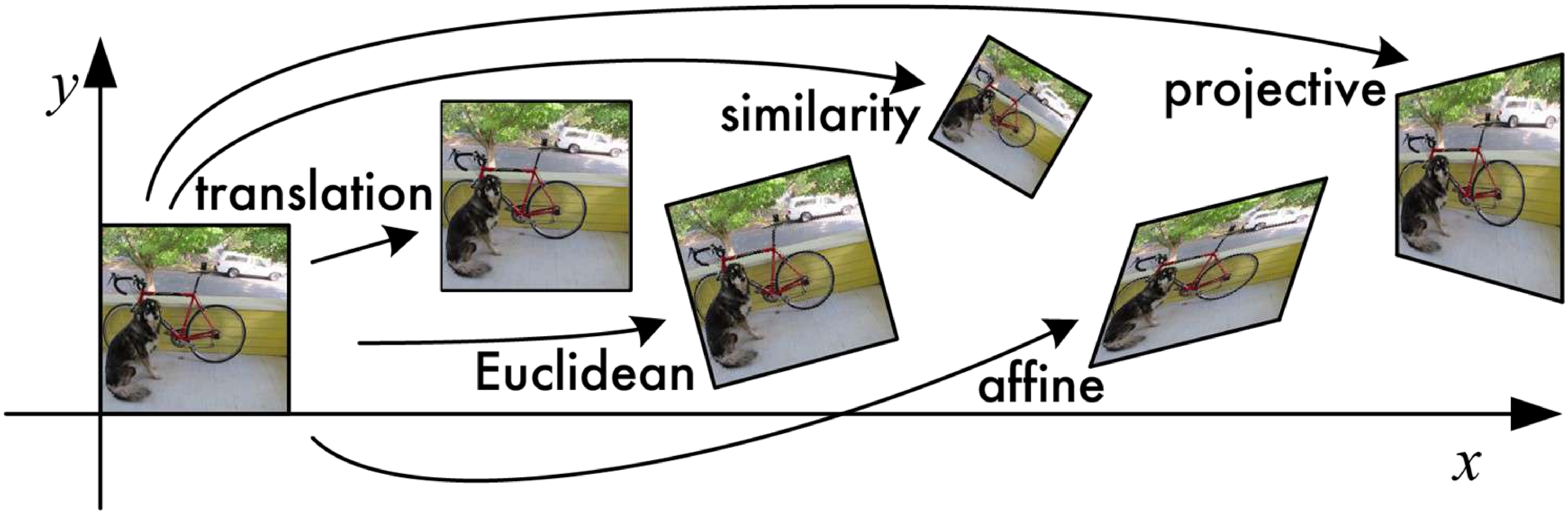
$$\bar{\mathbf{x}}' = \begin{bmatrix} a & -b & dx \\ b & a & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\bar{\mathbf{x}}' = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Projective transform

- Also known as homography
- Wait but affine was any 2x3 matrix...



Need some new coordinates!

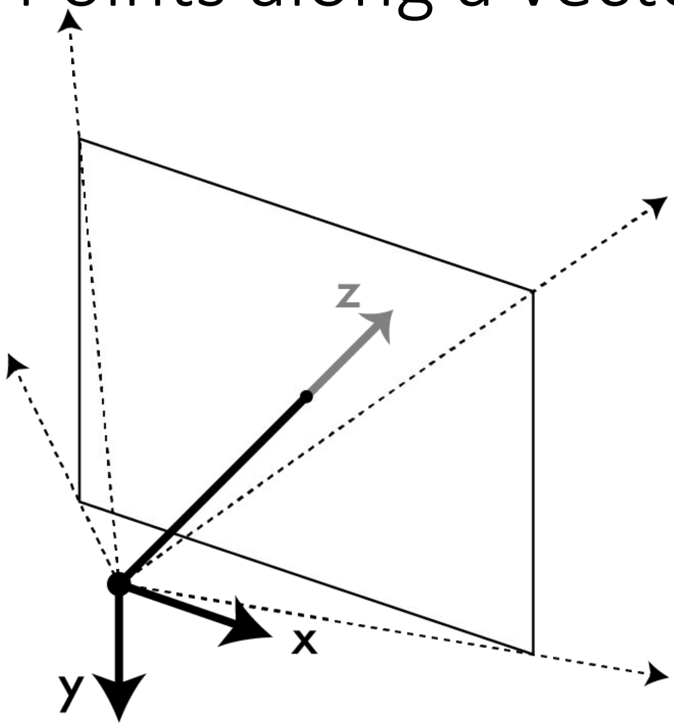
- Homogeneous coordinate system
- Each point in 2d is actually a vector in 3d
- Equivalent up to scaling factor
- Have to normalize to get back to 2d

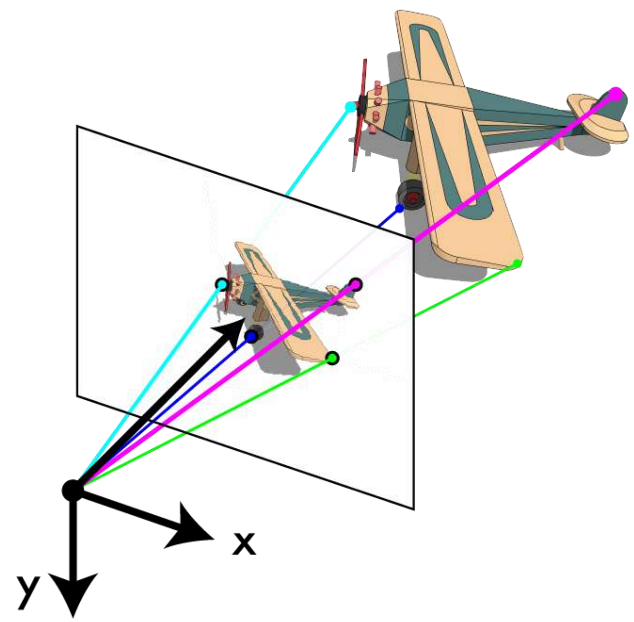
$$\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{bmatrix} \qquad \bar{\mathbf{x}} = \tilde{\mathbf{x}} / \tilde{w}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Why does this make sense?

- Remember our pinhole camera model
- Every point in 3d projects onto our viewing plane through our aperture
- Points along a vector are indistinguishable





Projective transform

- Also known as homography
- Wait but affine was any 2x3 matrix…
- Homography is general 3x3 matrix
- Multiplication by scalar is equivalent

$$\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\,\tilde{\mathbf{x}}$$

Projective transform

- Also known as homography
- Wait but affine was any 2x3 matrix…
- Homography is general 3x3 matrix
- Multiplication by scalar: equivalent projection
  - 3*H ~ H

$$\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\,\tilde{\mathbf{x}}$$

$$\tilde{\mathbf{x}}' = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Using homography to project point

- Multiply **x~** by **H~** to get **x'~**
- Convert to **x'** by dividing by **w'~**

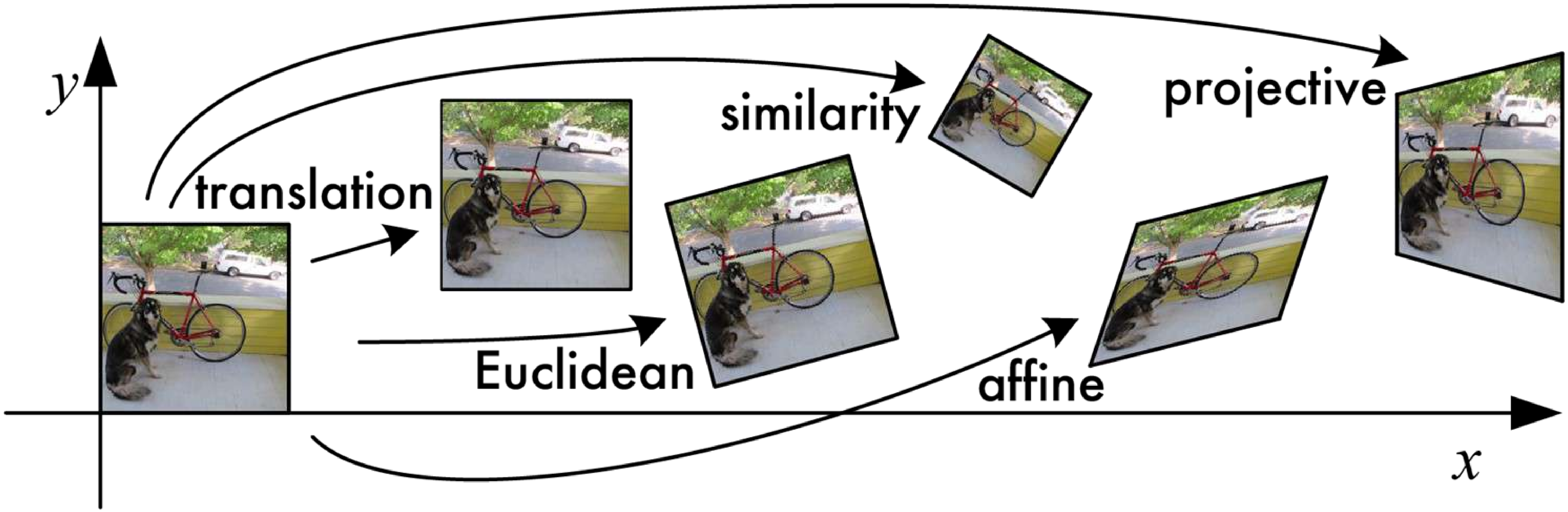
$$\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\,\tilde{\mathbf{x}}$$

$$\begin{bmatrix} \tilde{x}' \\ \tilde{y}' \\ \tilde{w}' \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{bmatrix}$$

$$\bar{\mathbf{x}} = \tilde{\mathbf{x}} \,/\, \tilde{w}$$

Lots to choose from

- What do each of them do?
- Which is right for panorama stitching?



Today’s Agenda

- Basic descriptor and matching
- Image transformations
- Estimate transformations

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How hard are they to recover?

$$\bar{\mathbf{x}}' = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\bar{\mathbf{x}}' = \begin{bmatrix} \cos\theta & -\sin\theta & dx \\ \sin\theta & \cos\theta & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

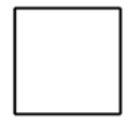
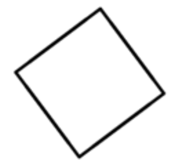
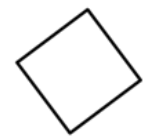
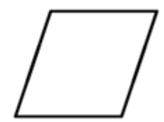
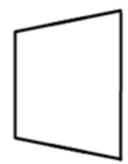
$$\bar{\mathbf{x}}' = \begin{bmatrix} a & -b & dx \\ b & a & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\bar{\mathbf{x}}' = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\tilde{\mathbf{x}}' = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{bmatrix}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

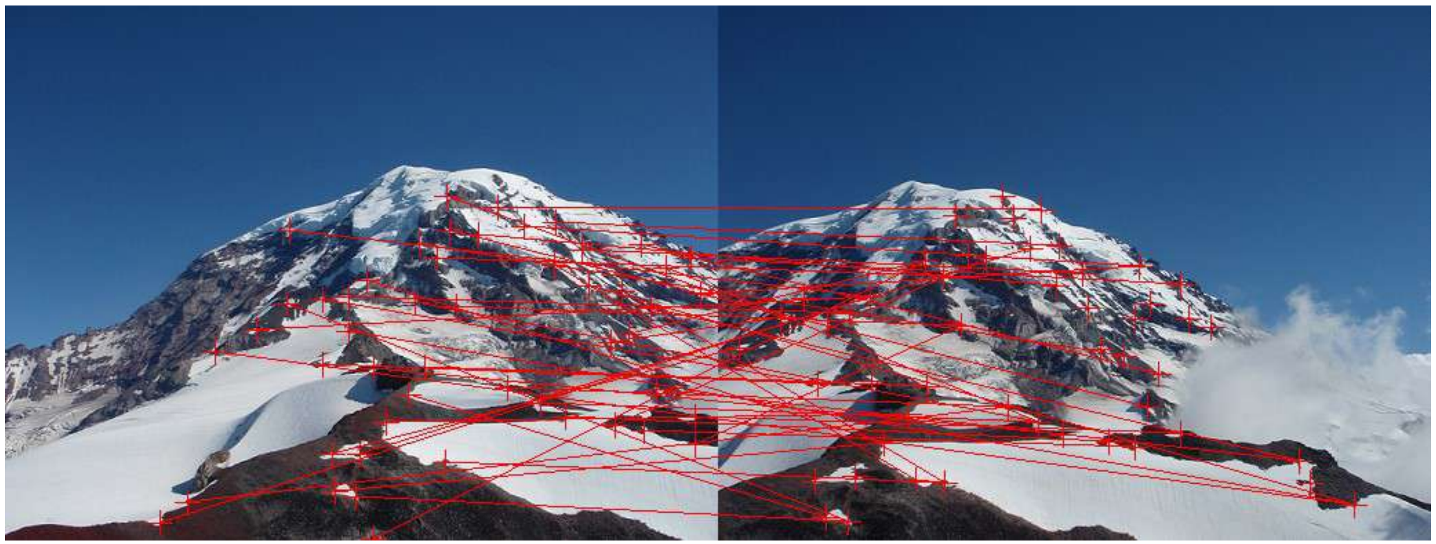
Lots to choose from

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} \mathbf{I} & \| & \mathbf{t} \end{bmatrix}_{2\times 3}$	2	orientation	
rigid (Euclidean)	$\begin{bmatrix} \mathbf{R} & \| & \mathbf{t} \end{bmatrix}_{2\times 3}$	3	lengths	
similarity	$\begin{bmatrix} s\mathbf{R} & \| & \mathbf{t} \end{bmatrix}_{2\times 3}$	4	angles	
affine	$\begin{bmatrix} \mathbf{A} \end{bmatrix}_{2\times 3}$	6	parallelism	
projective	$\begin{bmatrix} \tilde{\mathbf{H}} \end{bmatrix}_{3\times 3}$	8	straight lines	

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Say we want affine transformation

- Have our matched points
- Want to estimate **A** that maps from **x** to **x'**
- **Ax = x'**



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Say we want affine transformation

- Have our matched points
- Want to estimate **A** that maps from **x** to **x'**
- **Ax** = **x'**
- How many degrees of freedom?

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Say we want affine transformation

- Have our matched points
- Want to estimate **A** that maps from **x** to **x'**
- **A**x = x'
- How many degrees of freedom?
  - 6
- How many knowns do we get with one match?

$$\mathbf{x'} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Say we want affine transformation

- Have our matched points
- Want to estimate **A** that maps from **x** to **x'**
- **A****x** = **x'**
- How many degrees of freedom?
  - 6
- How many knowns do we get with one match?
  - 2
  - $n_x = a_{00} * m_x + a_{01} * m_y + a_{02} * 1$
  - $n_y = a_{10} * m_x + a_{11} * m_y + a_{12} * 1$

$$\mathbf{x'} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Say we want affine transformation

- How many knowns do we get with one match?
  - $n_x = a_{00} * m_x + a_{01} * m_y + a_{02} * 1$
  - $n_y = a_{10} * m_x + a_{11} * m_y + a_{12} * 1$
  - Solve linear system of equations **M a = b**
    - $M^{-1} M a = M^{-1} b \Rightarrow a = M^{-1} b$
    - But M^{-1} does not exist in general - Why?
  - Still works if overdetermined
    - Why???
    - Pseudoinverse – least squares solution
    - $M^T M a = M^T b$
    - $(M^T M)^{-1} (M^T M) a = (M^T M)^{-1} M^T b$
    - $\Rightarrow a = (M^T M)^{-1} M^T b$

$$
\begin{array}{cccc}
\mathbf{M} & \mathbf{a} & & \mathbf{b} \\
\begin{bmatrix}
m_{x1} & m_{y1} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & m_{x1} & m_{y1} & 1 \\
m_{x2} & m_{y2} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & m_{x2} & m_{y2} & 1 \\
m_{x3} & m_{y3} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & m_{x3} & m_{y3} & 1 \\
 & & \cdots & & & \\
 & & \cdots & & &
\end{bmatrix}
&
\begin{bmatrix}
a_{00} \\ a_{01} \\ a_{02} \\ a_{10} \\ a_{11} \\ a_{12}
\end{bmatrix}
& = &
\begin{bmatrix}
n_{x1} \\ n_{y1} \\ n_{x2} \\ n_{y2} \\ n_{x3} \\ n_{y3}
\end{bmatrix}
\begin{matrix} \\ \\ \\ \\ \\ \\ \cdots \\ \cdots \end{matrix}
\end{array}
$$

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 9: RANSAC, Panorama Stitching

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Basic feature descriptor and matching
- Histogram of Oriented Gradients
- SIFT
- Image transformations
- Estimate transformations

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Linear least-squares
- RANSAC
- Panorama Stitching

[material based on Joseph Redmon’s course]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Linear least-squares
- RANSAC
- Panorama Stitching

Linear least squares

Want to solve overdetermined linear system:

- $M a = b$

Want to minimize squared error:

$|| b - M a ||^2 =$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Linear least squares

Want to solve overdetermined linear system:

- $M a = b$

Want to minimize squared error:

$|| b - M a ||^2 =$

$(b - M a)^T(b - M a)$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Linear least squares

Want to solve overdetermined linear system:

- $M a = b$

Want to minimize squared error:

$|| b - M a ||^2 =$

$(b - M a)^T (b - M a) =$

$b^T b - a^T M^T b - b^T M a + a^T M^T M a$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Linear least squares

Want to solve overdetermined linear system:

- $M\, a = b$

Want to minimize squared error:

$|| b - M\, a ||^2 =$

$(b - M\, a)^T(b - M\, a) =$

$b^Tb - a^TM^Tb - b^TMa + a^TM^TMa =$

$b^Tb - 2a^TM^Tb + a^TM^TMa$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Linear least squares

Want to minimize squared error: $|| b - M a ||^2 =$

$b^Tb - 2a^TM^Tb + a^TM^TMa$

This is convex and minimized when gradient = 0. So we take the derivative wrt a and set = 0.

$-M^Tb + (M^TM)a = 0$

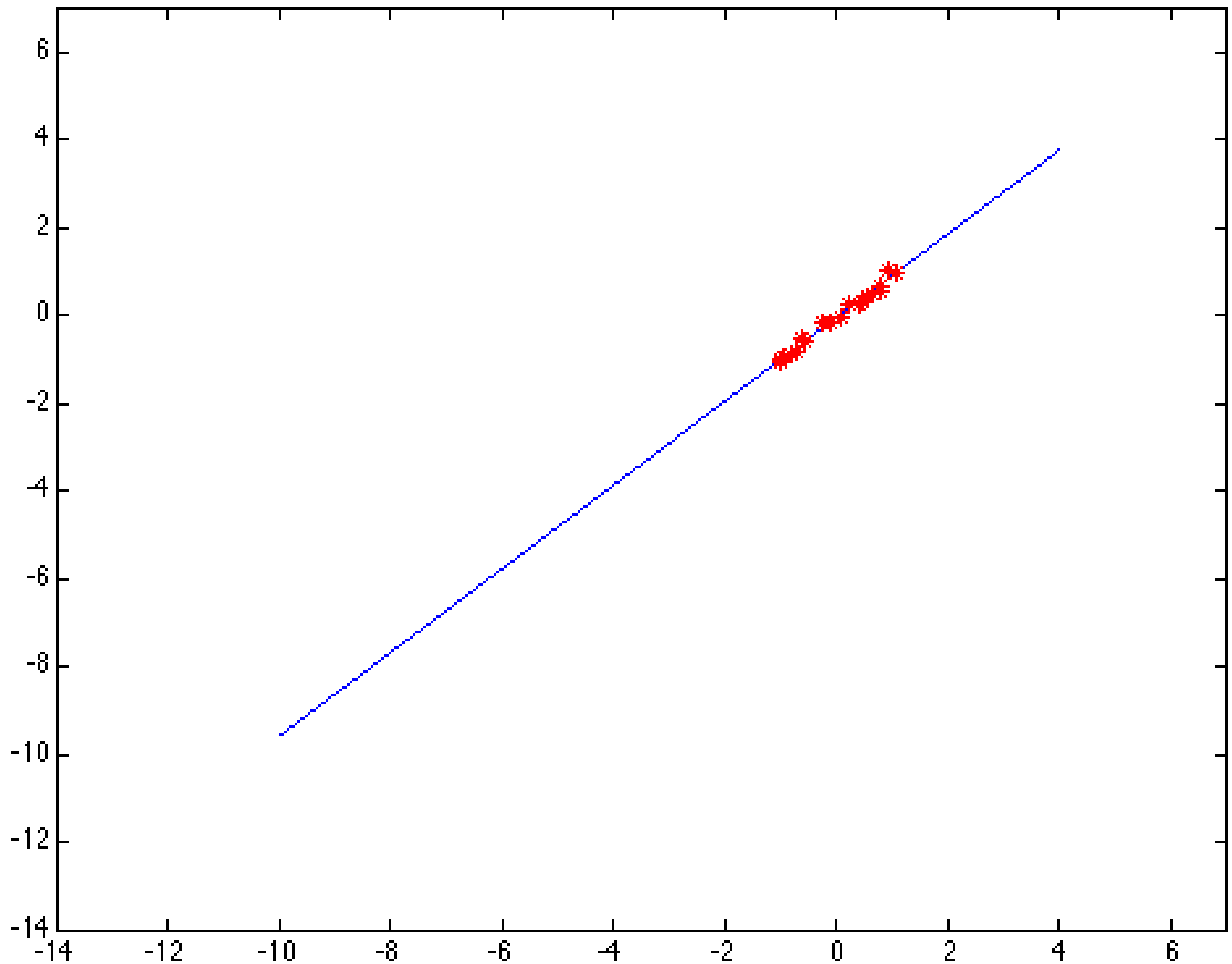
$(M^TM)a = M^Tb$

$a = (M^TM)^{-1}M^Tb$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

So what does linear least squares do?

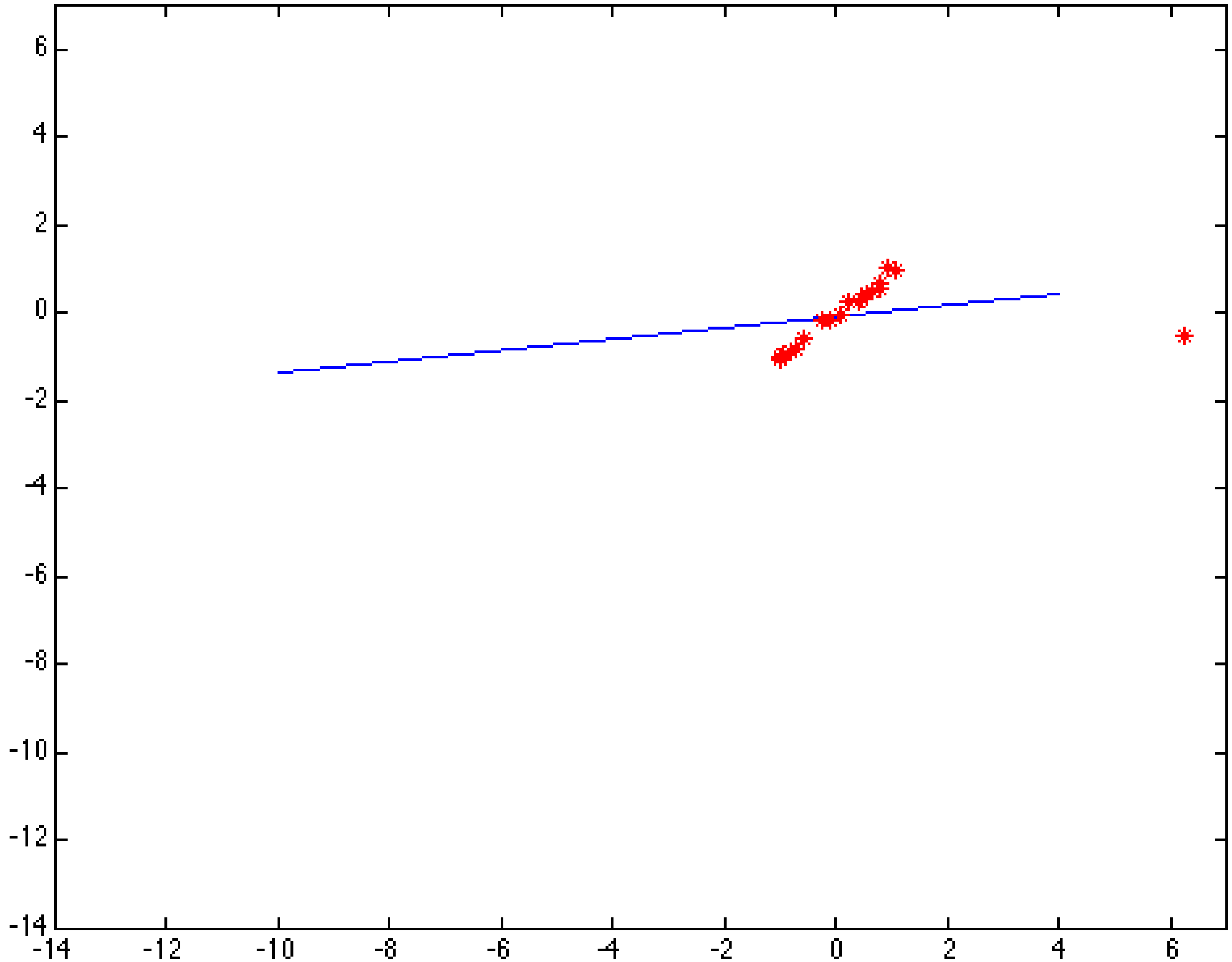


So what does linear least squares do?

Error based on **squared** residual

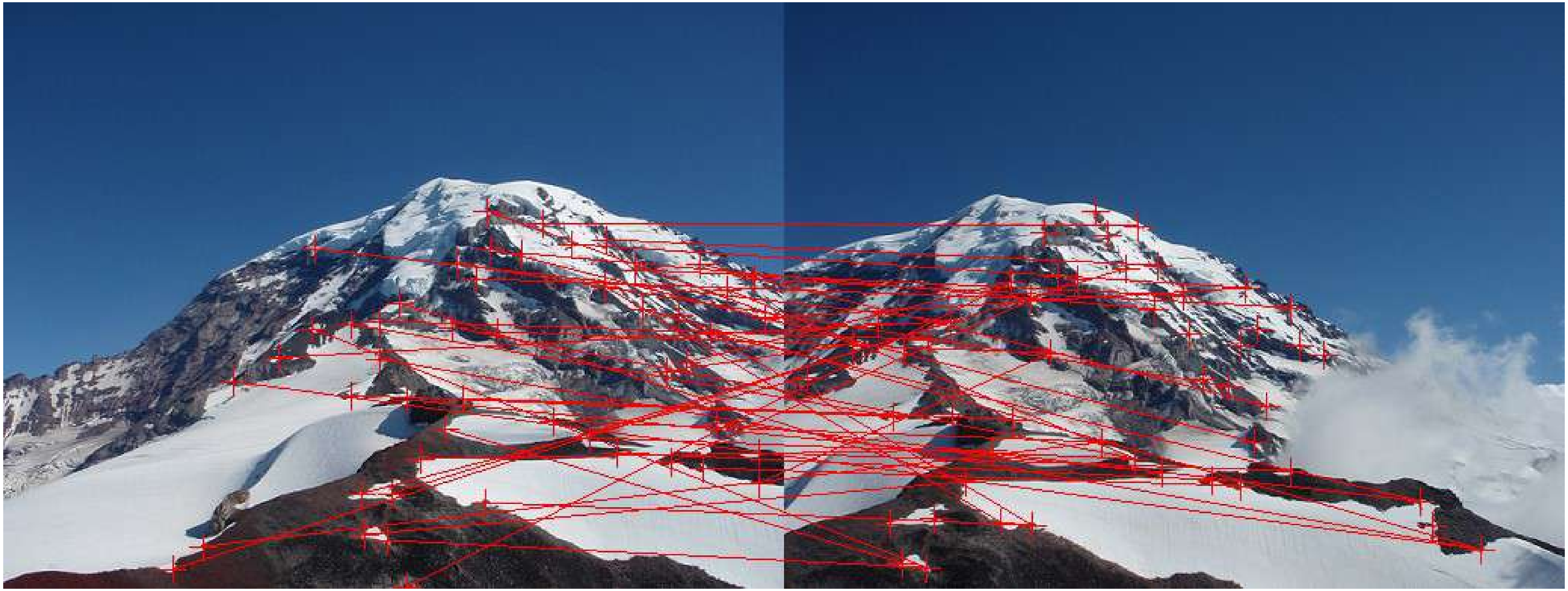
Very scared of being wrong, even for just one point

Very bad at handling outliers in data



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Not a problem for us, our data is perfect…



Not really …

Today’s Agenda

- Linear least-squares
- RANSAC
- Panorama Stitching

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

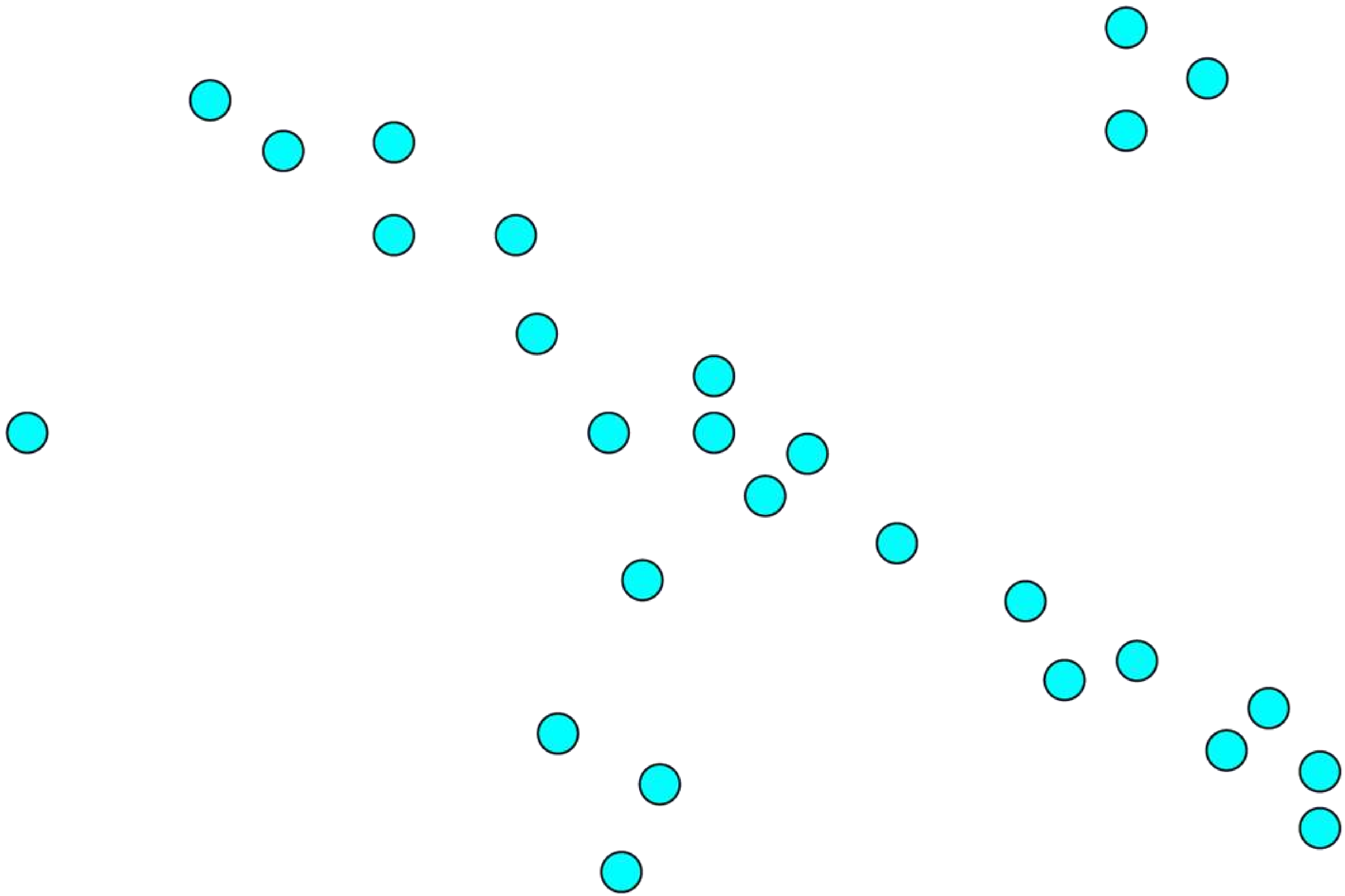
RANSAC: RANdom SAmple Consensus

- How can we fit model to inliers but ignore outliers?
- Try a bunch of models, see which ones are best!
- Inliers will all agree on a model
- Outliers are basically random, will not agree

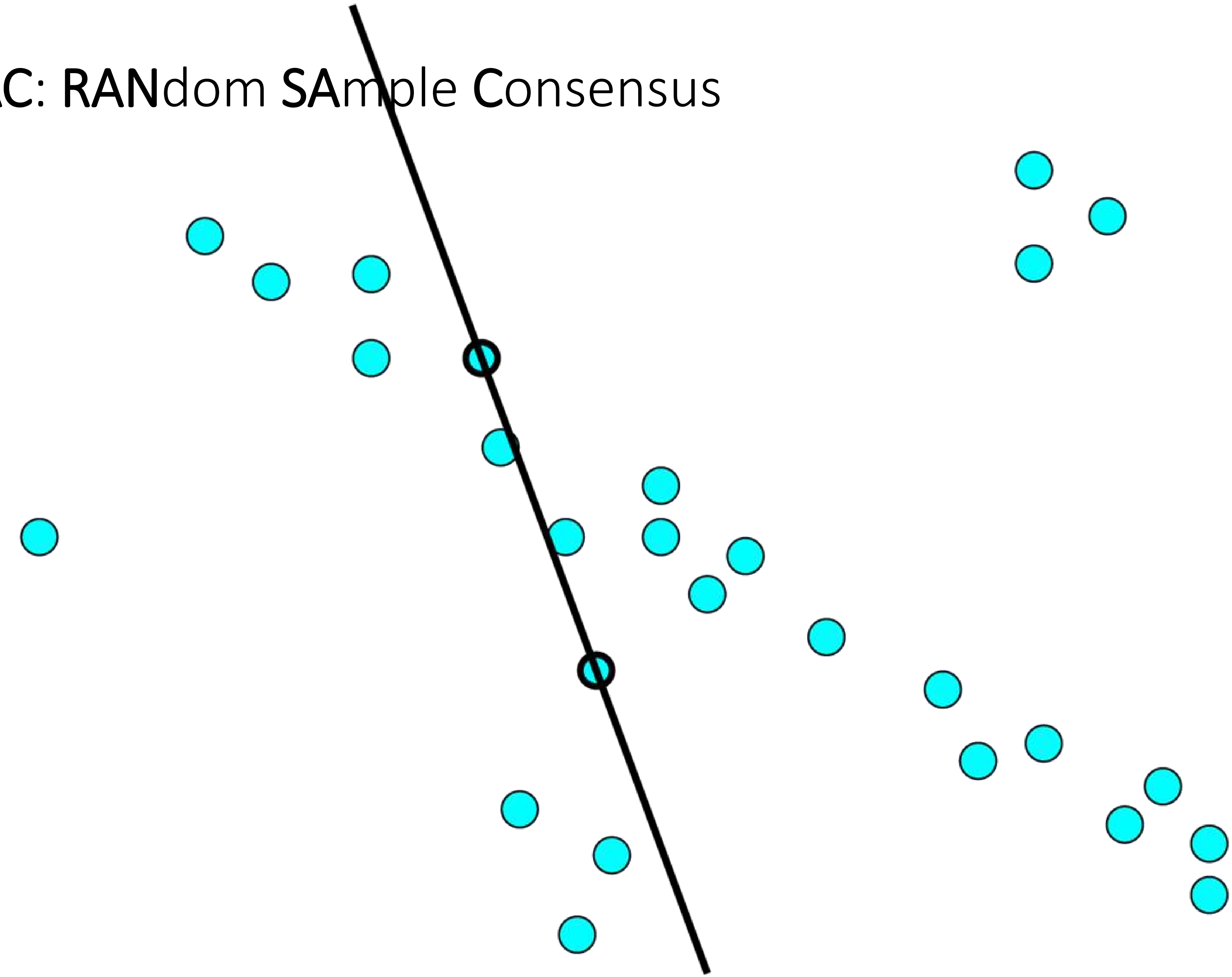
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

RANSAC: RANdom SAmple Consensus



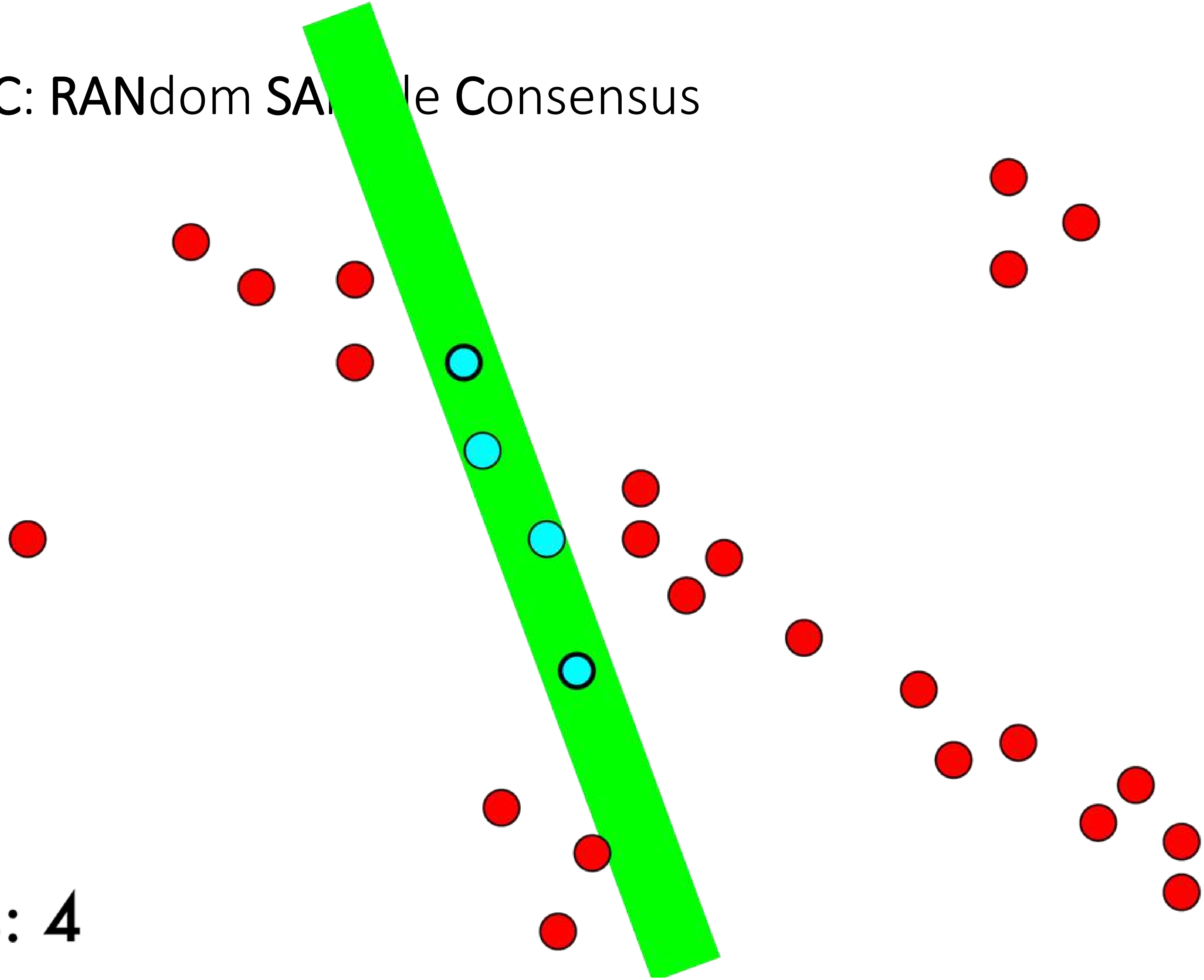
RANSAC: RANdom SAmple Consensus



Co-financed by the European Union
Connecting Europe Facility

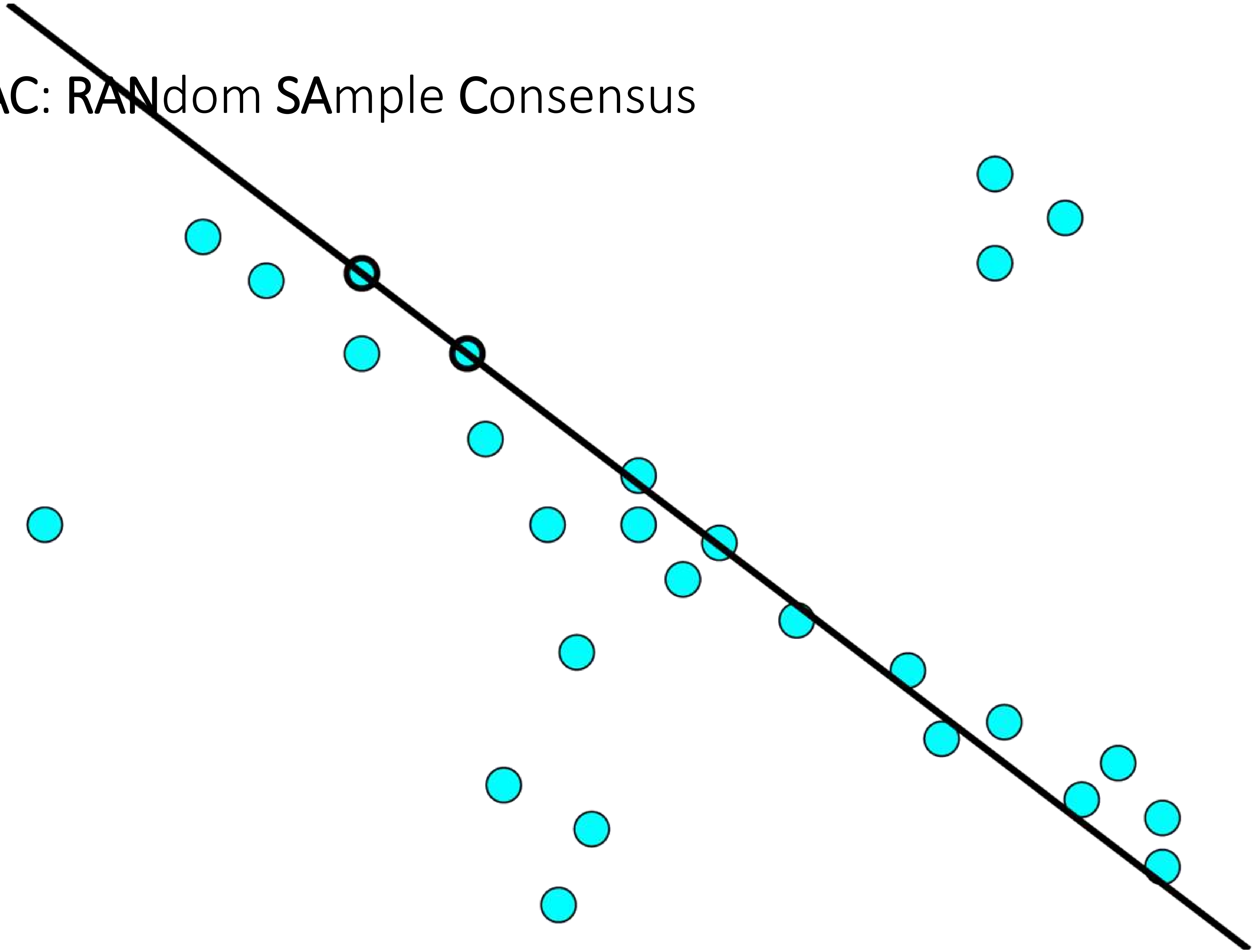
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

RANSAC: RANdom SA le Consensus

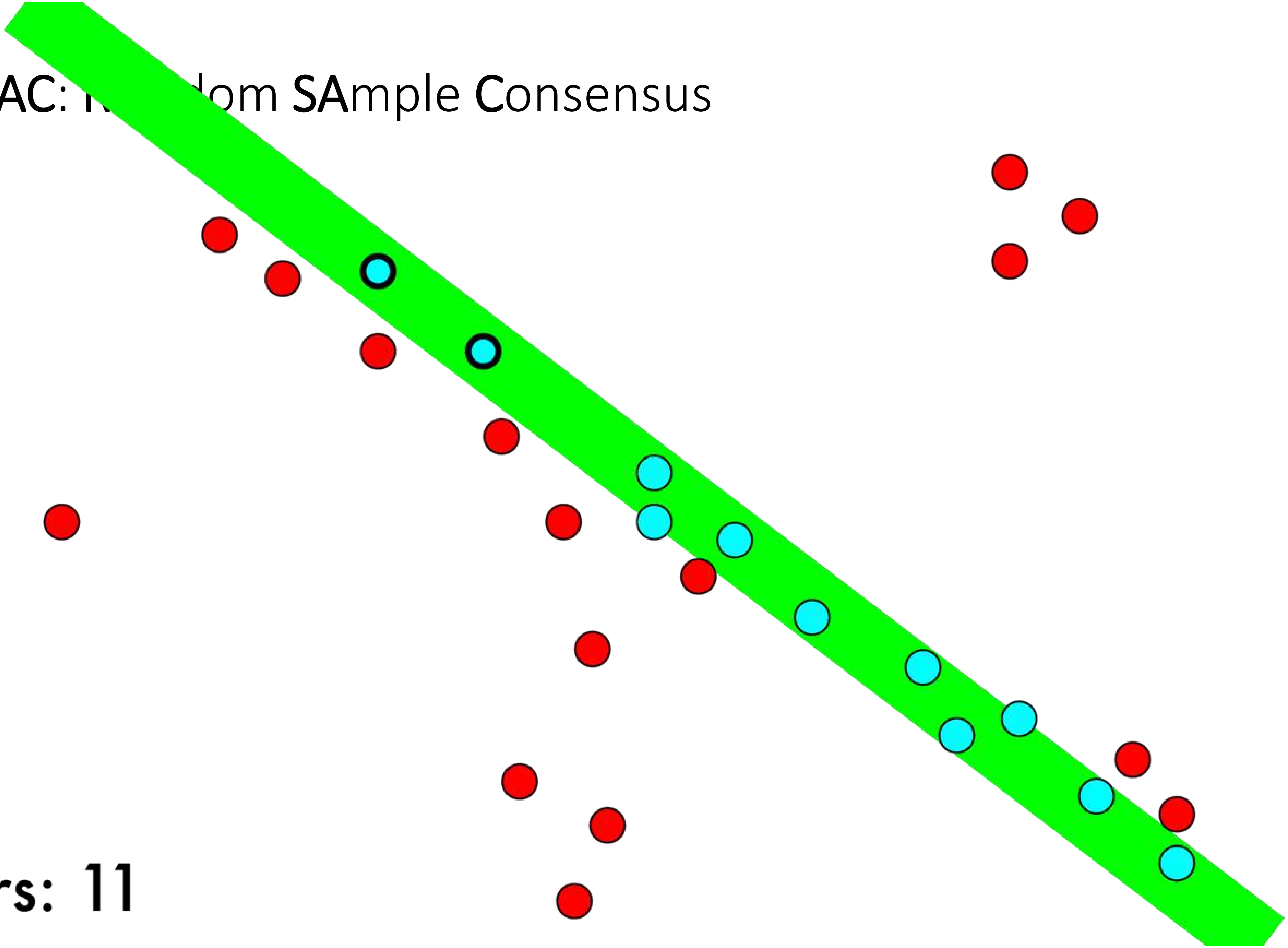


Inliers: 4

RANSAC: RANdom SAmple Consensus



RANSAC: Random SAmple Consensus

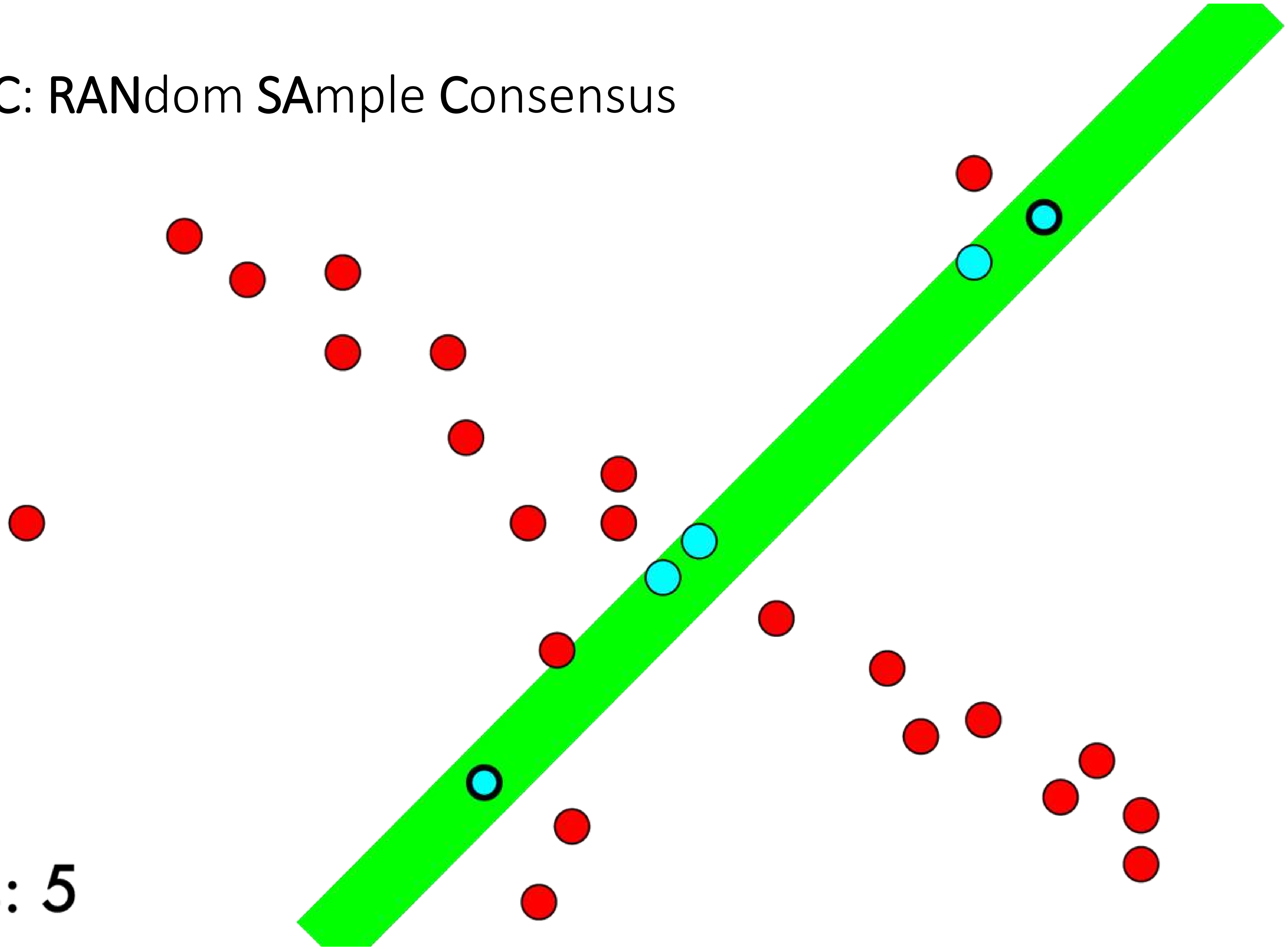


Inliers: 11

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

RANSAC: RANdom SAmple Consensus

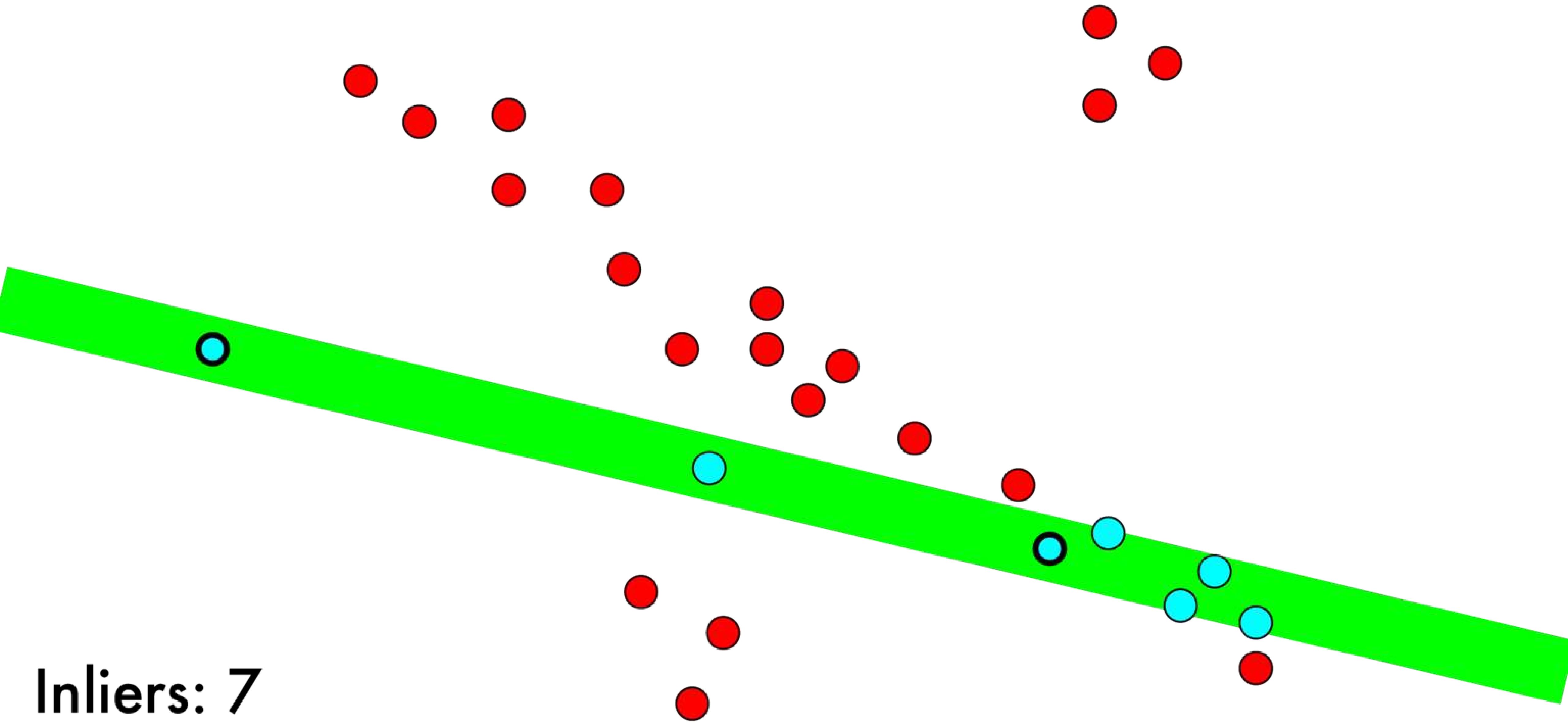


Inliers: 5

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

RANSAC: RANdom SAmple Consensus

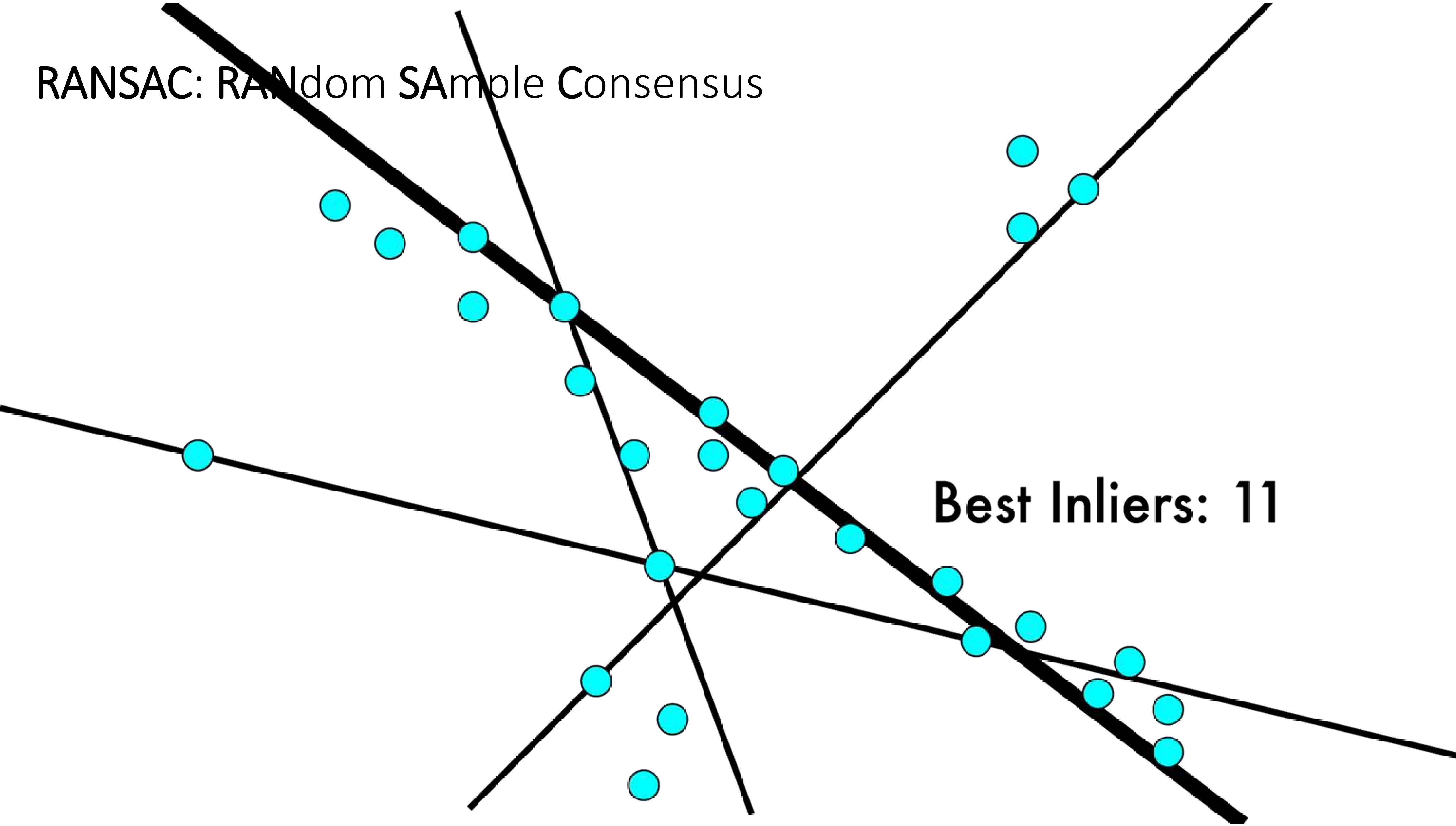


Inliers: 7

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

RANSAC: RANdom SAmple Consensus



RANSAC: RANdom SAmple Consensus

- Parameters: data, model, n points to fit model, k iterations, t threshold, d “good” fit cutoff

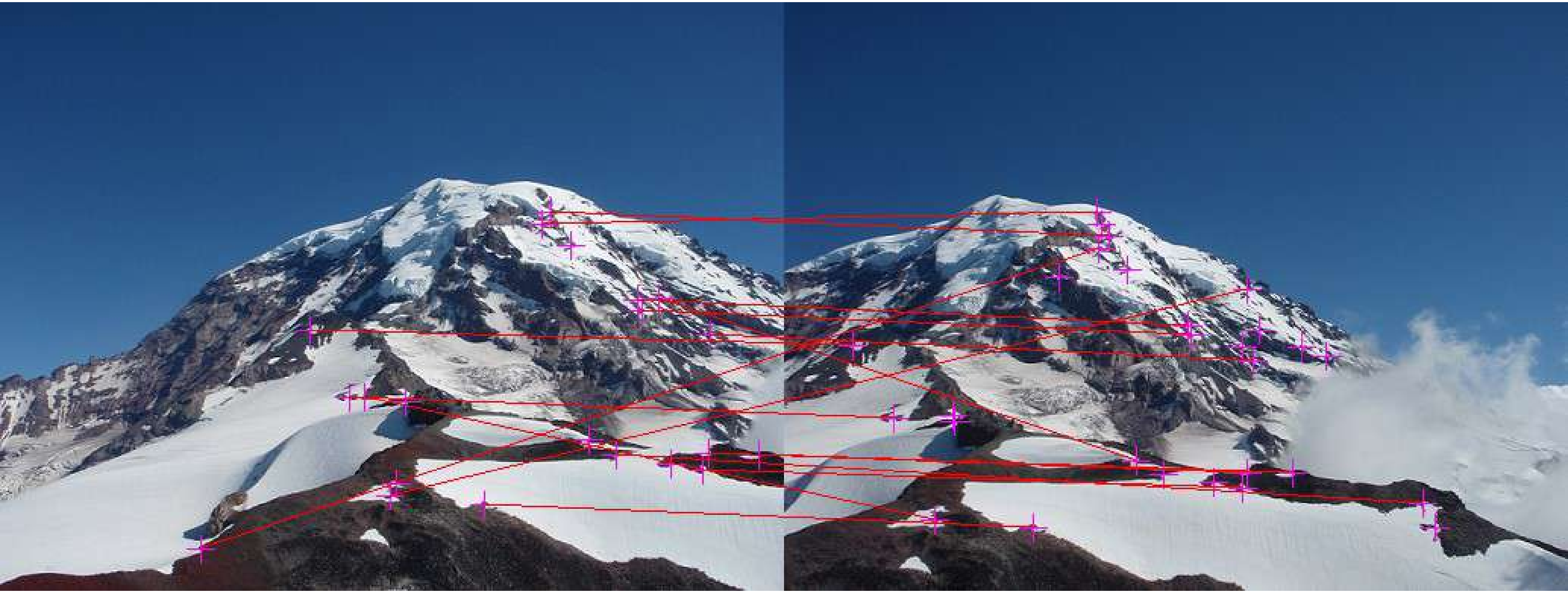
```
bestmodel = None
bestfit = -INF
While i < k:
        sample = draw n random points from data
        Fit model to sample
        inliers = data within t of model
        if inliers > bestfit:
                Fit model to all inliers
                bestfit = fit
                bestmodel = model
                if inliers > d:
                        return model
return bestmodel
```

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

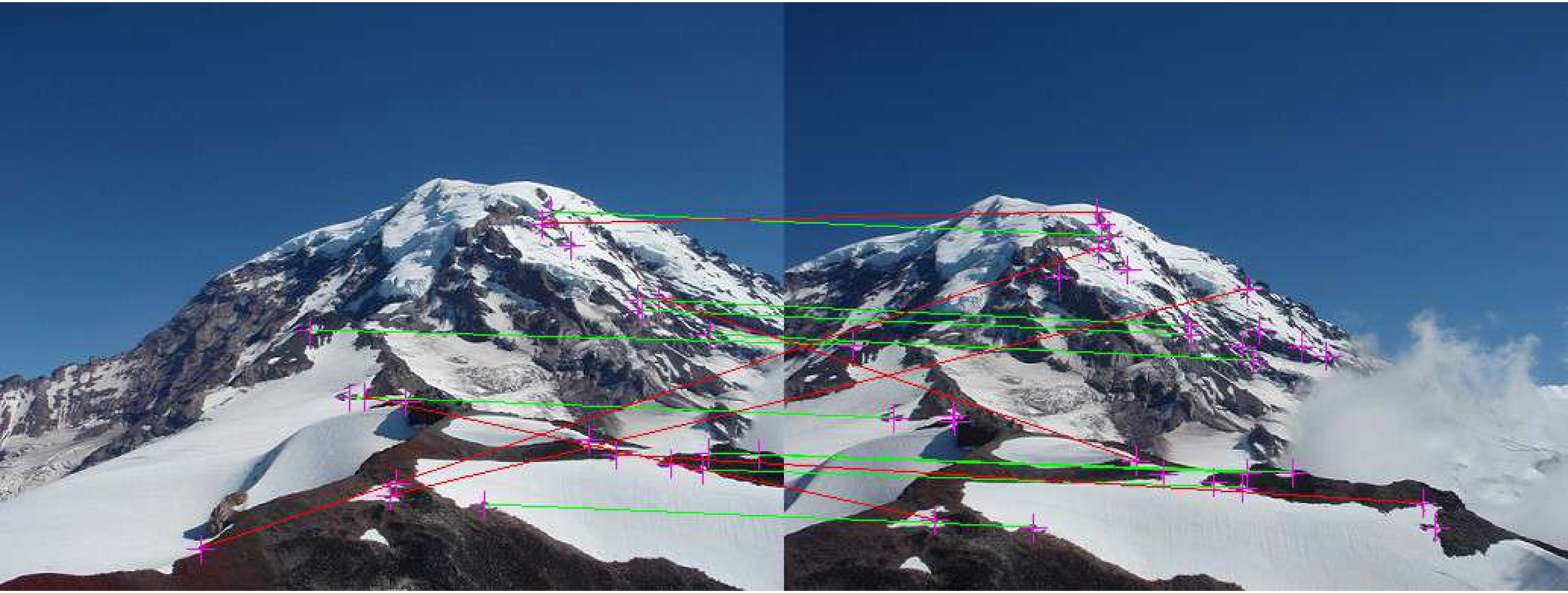
RANSAC: RANdom SAmple Consensus

- Works well even with extreme noise.



RANSAC: RANdom SAmple Consensus

- Works well even with extreme noise.



RANSAC: RANdom SAmple Consensus

- Parameters: data, model, n points to fit model, k iterations, t threshold, d “good” fit cutoff
- Lots of tunable parameters
- Want high probability of recovering “right” model
- t: often quite small, assume “good” inliers
- n: should be just enough to fit model, no extra
- k: can be very high
- d: should be >> n

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We can estimate affine..

- How many knowns do we get with one match?
  - $n_x = a_{00} * m_x + a_{01} * m_y + a_{02} * 1$
  - $n_y = a_{10} * m_x + a_{11} * m_y + a_{12} * 1$
  - Solve linear system of equations **M a** = **b**
    - $M^{-1} M a = M^{-1} b \Rightarrow a = M^{-1} b$
    - But M^{-1} does not exist in general - Why?
  - Still works if overdetermined
    - Why???
    - Pseudoinverse – least squares solution
    - $M^T M a = M^T b$
    - $(M^T M)^{-1} (M^T M) a = (M^T M)^{-1} M^T b$
    - $\Rightarrow a = (M^T M)^{-1} M^T b$

$$
\begin{array}{ccc}
\mathbf{M} & \mathbf{a} & \mathbf{b} \\
\begin{bmatrix}
m_{x1} & m_{y1} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & m_{x1} & m_{y1} & 1 \\
m_{x2} & m_{y2} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & m_{x2} & m_{y2} & 1 \\
m_{x3} & m_{y3} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & m_{x3} & m_{y3} & 1 \\
 & & & \dots & & \\
 & & & \dots & &
\end{bmatrix}
&
\begin{bmatrix}
a_{00} \\ a_{01} \\ a_{02} \\ a_{10} \\ a_{11} \\ a_{12}
\end{bmatrix}
=
&
\begin{bmatrix}
n_{x1} \\ n_{y1} \\ n_{x2} \\ n_{y2} \\ n_{x3} \\ n_{y3}
\end{bmatrix} \\
 & & \dots \\
 & & \dots
\end{array}
$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We want projective (homography)

- What are our equations now?
    - $n_x = (h_{00}*m_x + h_{01}*m_y + h_{02}*m_w) / (h_{20}*m_x + h_{21}*m_y + h_{22}*m_w)$
    - $n_y = (h_{10}*m_x + h_{11}*m_y + h_{12}*m_w) / (h_{20}*m_x + h_{21}*m_y + h_{22}*m_w)$

$$\begin{bmatrix} \tilde{x}' \\ \tilde{y}' \\ \tilde{w}' \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{bmatrix}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We want projective (homography)

- What are our equations now?
  - $n_x = (h_{00}*m_x + h_{01}*m_y + h_{02}*m_w) / (h_{20}*m_x + h_{21}*m_y + h_{22}*m_w)$
  - $n_y = (h_{10}*m_x + h_{11}*m_y + h_{12}*m_w) / (h_{20}*m_x + h_{21}*m_y + h_{22}*m_w)$
- Assume h_{22} and m_w are 1, now 8 DOF
  - $n_x = (h_{00}*m_x + h_{01}*m_y + h_{02}) / (h_{20}*m_x + h_{21}*m_y + 1)$
  - $n_y = (h_{10}*m_x + h_{11}*m_y + h_{12}) / (h_{20}*m_x + h_{21}*m_y + 1)$

$$\begin{bmatrix} \tilde{x}' \\ \tilde{y}' \\ \tilde{w}' \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{bmatrix}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We want projective (homography)

- What are our equations now?
  - $n_x = (h_{00}*m_x + h_{01}*m_y + h_{02}*m_w) / (h_{20}*m_x + h_{21}*m_y + h_{22}*m_w)$
  - $n_y = (h_{10}*m_x + h_{11}*m_y + h_{12}*m_w) / (h_{20}*m_x + h_{21}*m_y + h_{22}*m_w)$
- Assume h_{22} and m_w are 1, now 8 DOF
  - $n_x = (h_{00}*m_x + h_{01}*m_y + h_{02}) / (h_{20}*m_x + h_{21}*m_y + 1)$
  - $n_y = (h_{10}*m_x + h_{11}*m_y + h_{12}) / (h_{20}*m_x + h_{21}*m_y + 1)$
- More algebra on n_x
  - $n_x * (h_{20}*m_x + h_{21}*m_y + 1) = (h_{00}*m_x + h_{01}*m_y + h_{02})$
  - $n_x*h_{20}*m_x + n_x*h_{21}*m_y + n_x = h_{00}*m_x + h_{01}*m_y + h_{02}$
  - $n_x = h_{00}*m_x + h_{01}*m_y + h_{02} - n_x*h_{20}*m_x - n_x*h_{21}*m_y$
- Similar for n_y

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We want projective (homography)

- What are our equations now?
  - $n_x = h_{00}*m_x + h_{01}*m_y + h_{02} - n_x*h_{20}*m_x - n_x*h_{21}*m_y$
  - $n_y = h_{10}*m_x + h_{11}*m_y + h_{12} - n_x*h_{20}*m_x - n_x*h_{21}*m_y$
- New matrix equations:

$$\underbrace{\begin{bmatrix} m_{x1} & m_{y1} & 1 & 0 & 0 & 0 & -m_{x1}n_{x1} & -m_{y1}n_{x1} \\ 0 & 0 & 0 & m_{x1} & m_{y1} & 1 & -m_{x1}n_{y1} & -m_{y1}n_{y1} \\ m_{x2} & m_{y2} & 1 & 0 & 0 & 0 & -m_{x2}n_{x2} & -m_{y2}n_{x2} \\ 0 & 0 & 0 & m_{x2} & m_{y2} & 1 & -m_{x2}n_{y2} & -m_{y2}n_{y2} \\ m_{x3} & m_{y3} & 1 & 0 & 0 & 0 & -m_{x3}n_{x3} & -m_{y3}n_{x3} \\ 0 & 0 & 0 & m_{x3} & m_{y3} & 1 & -m_{x3}n_{y3} & -m_{y3}n_{y3} \\ m_{x4} & m_{y4} & 1 & 0 & 0 & 0 & -m_{x4}n_{x4} & -m_{y4}n_{x4} \\ 0 & 0 & 0 & m_{x4} & m_{y4} & 1 & -m_{x4}n_{y4} & -m_{y4}n_{y4} \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} h_{00} \\ h_{01} \\ h_{02} \\ h_{10} \\ h_{11} \\ h_{12} \\ h_{20} \\ h_{21} \end{bmatrix}}_{\mathbf{a}} = \underbrace{\begin{bmatrix} n_{x1} \\ n_{y1} \\ n_{x2} \\ n_{y2} \\ n_{x3} \\ n_{y3} \\ n_{x4} \\ n_{y4} \end{bmatrix}}_{\mathbf{b}}$$

We want projective (homography)

- New matrix equations:

$$
\underbrace{\begin{bmatrix}
m_{x1} & m_{y1} & 1 & 0 & 0 & 0 & -m_{x1}n_{x1} & -m_{y1}n_{x1} \\
0 & 0 & 0 & m_{x1} & m_{y1} & 1 & -m_{x1}n_{y1} & -m_{y1}n_{y1} \\
m_{x2} & m_{y2} & 1 & 0 & 0 & 0 & -m_{x2}n_{x2} & -m_{y2}n_{x2} \\
0 & 0 & 0 & m_{x2} & m_{y2} & 1 & -m_{x2}n_{y2} & -m_{y2}n_{y2} \\
m_{x3} & m_{y3} & 1 & 0 & 0 & 0 & -m_{x3}n_{x3} & -m_{y3}n_{x3} \\
0 & 0 & 0 & m_{x3} & m_{y3} & 1 & -m_{x3}n_{y3} & -m_{y3}n_{y3} \\
m_{x4} & m_{y4} & 1 & 0 & 0 & 0 & -m_{x4}n_{x4} & -m_{y4}n_{x4} \\
0 & 0 & 0 & m_{x4} & m_{y4} & 1 & -m_{x4}n_{y4} & -m_{y4}n_{y4}
\end{bmatrix}}_{\mathbf{M}}
\underbrace{\begin{bmatrix} h_{00} \\ h_{01} \\ h_{02} \\ h_{10} \\ h_{11} \\ h_{12} \\ h_{20} \\ h_{21} \end{bmatrix}}_{\mathbf{a}}
=
\underbrace{\begin{bmatrix} n_{x1} \\ n_{y1} \\ n_{x2} \\ n_{y2} \\ n_{x3} \\ n_{y3} \\ n_{x4} \\ n_{y4} \end{bmatrix}}_{\mathbf{b}}
$$

- Same procedure, Solve **M a = b**
  - Exact if #rows of M = 8
  - Least squares if #rows of M > 8

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Are there any problems with this??

- New matrix equations:

$$
\begin{array}{ccc} \mathbf{M} & \mathbf{a} & & \mathbf{b} \\
\begin{bmatrix}
m_{x1} & m_{y1} & 1 & 0 & 0 & 0 & -m_{x1}n_{x1} & -m_{y1}n_{x1} \\
0 & 0 & 0 & m_{x1} & m_{y1} & 1 & -m_{x1}n_{y1} & -m_{y1}n_{y1} \\
m_{x2} & m_{y2} & 1 & 0 & 0 & 0 & -m_{x2}n_{x2} & -m_{y2}n_{x2} \\
0 & 0 & 0 & m_{x2} & m_{y2} & 1 & -m_{x2}n_{y2} & -m_{y2}n_{y2} \\
m_{x3} & m_{y3} & 1 & 0 & 0 & 0 & -m_{x3}n_{x3} & -m_{y3}n_{x3} \\
0 & 0 & 0 & m_{x3} & m_{y3} & 1 & -m_{x3}n_{y3} & -m_{y3}n_{y3} \\
m_{x4} & m_{y4} & 1 & 0 & 0 & 0 & -m_{x4}n_{x4} & -m_{y4}n_{x4} \\
0 & 0 & 0 & m_{x4} & m_{y4} & 1 & -m_{x4}n_{y4} & -m_{y4}n_{y4}
\end{bmatrix}
&
\begin{bmatrix} h_{00} \\ h_{01} \\ h_{02} \\ h_{10} \\ h_{11} \\ h_{12} \\ h_{20} \\ h_{21} \end{bmatrix}
& = &
\begin{bmatrix} n_{x1} \\ n_{y1} \\ n_{x2} \\ n_{y2} \\ n_{x3} \\ n_{y3} \\ n_{x4} \\ n_{y4} \end{bmatrix}
\end{array}
$$

- Same procedure, Solve **M a = b**
  - Exact if #rows of **M** = 8
  - Least squares if #rows of **M** > 8

Today's Agenda

- Linear least-squares
- RANSAC
- Panorama Stitching

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Panorama algorithm

Find corners in both images

Calculate descriptors

Match descriptors

RANSAC to find homography

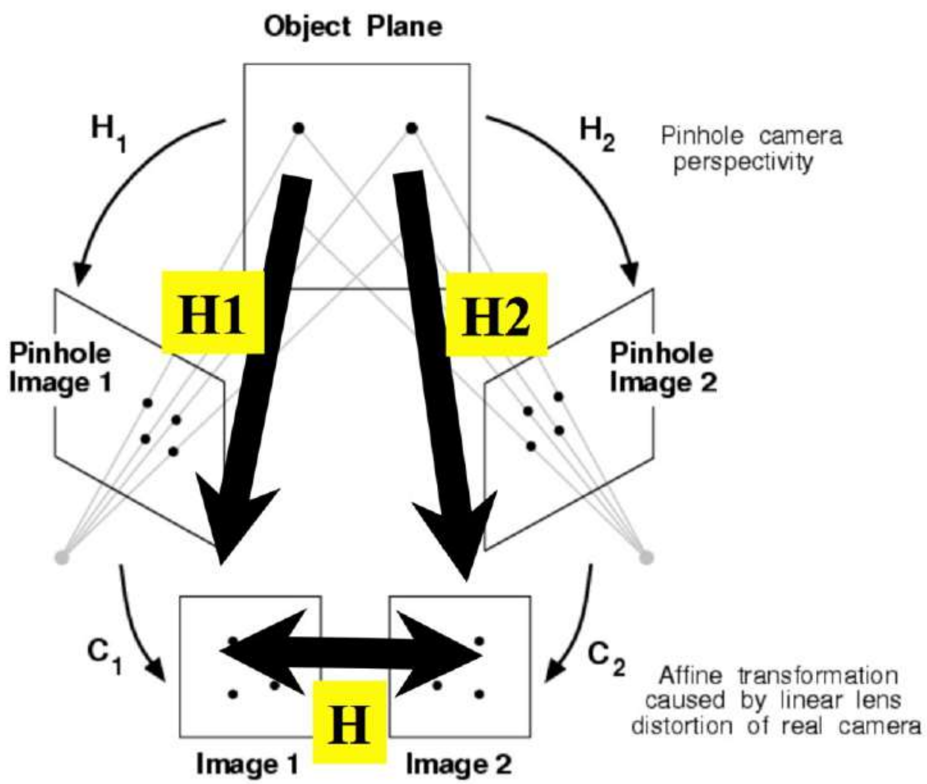
Stitch together images with homography

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Stitching panoramas

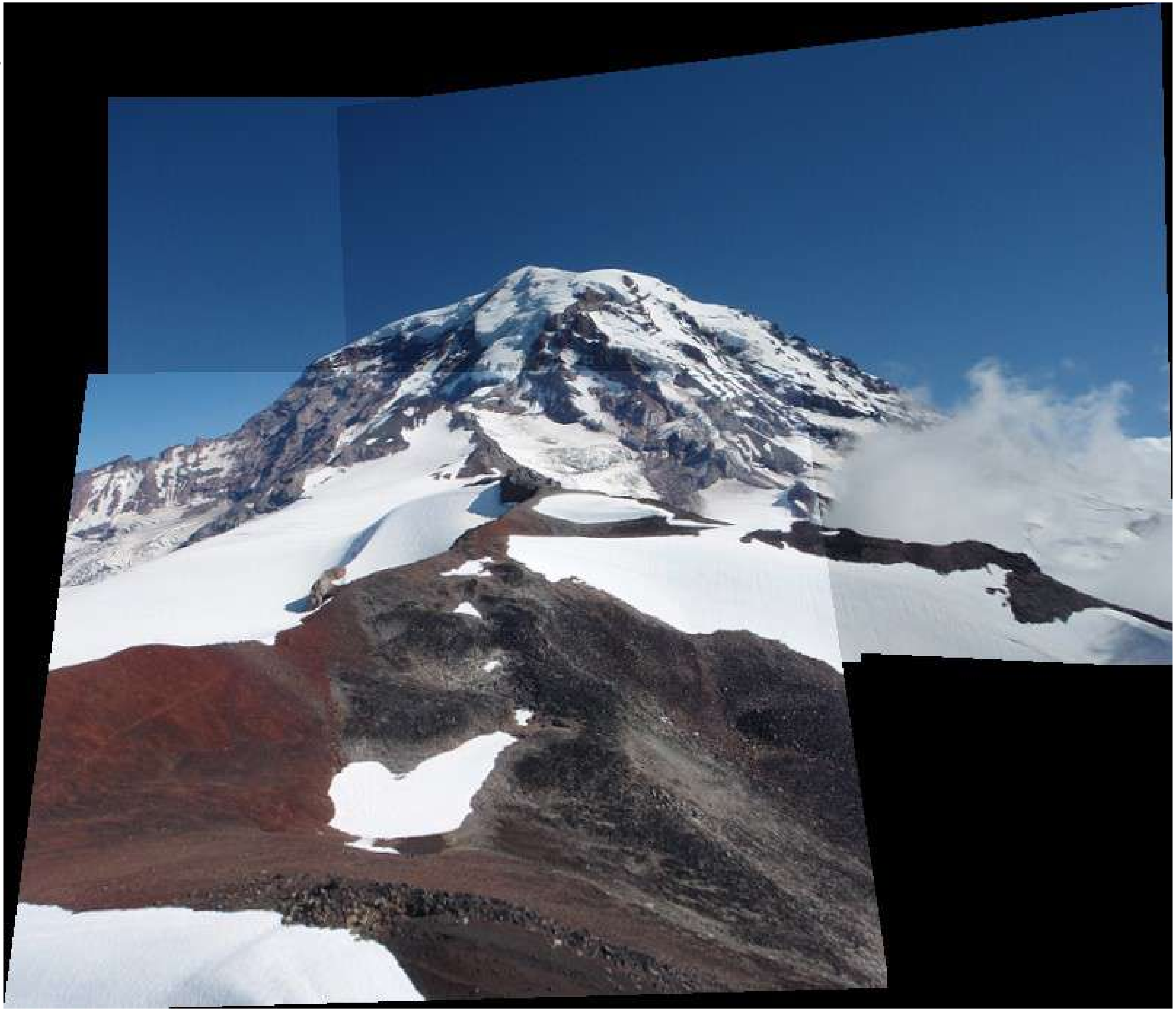
- We know homography is right choice under certain assumption:
    - Assume we are taking multiple images of planar object



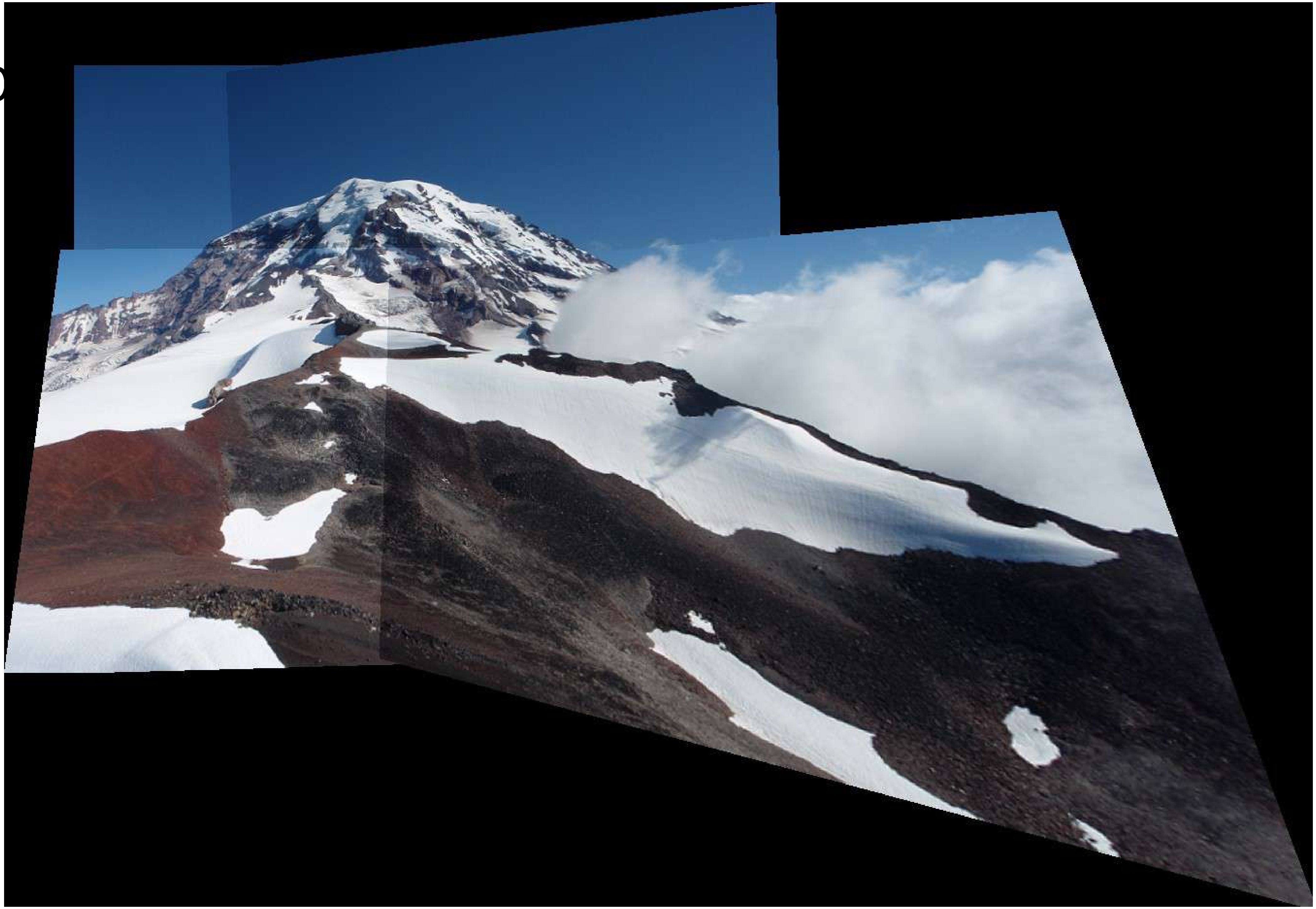
In practice

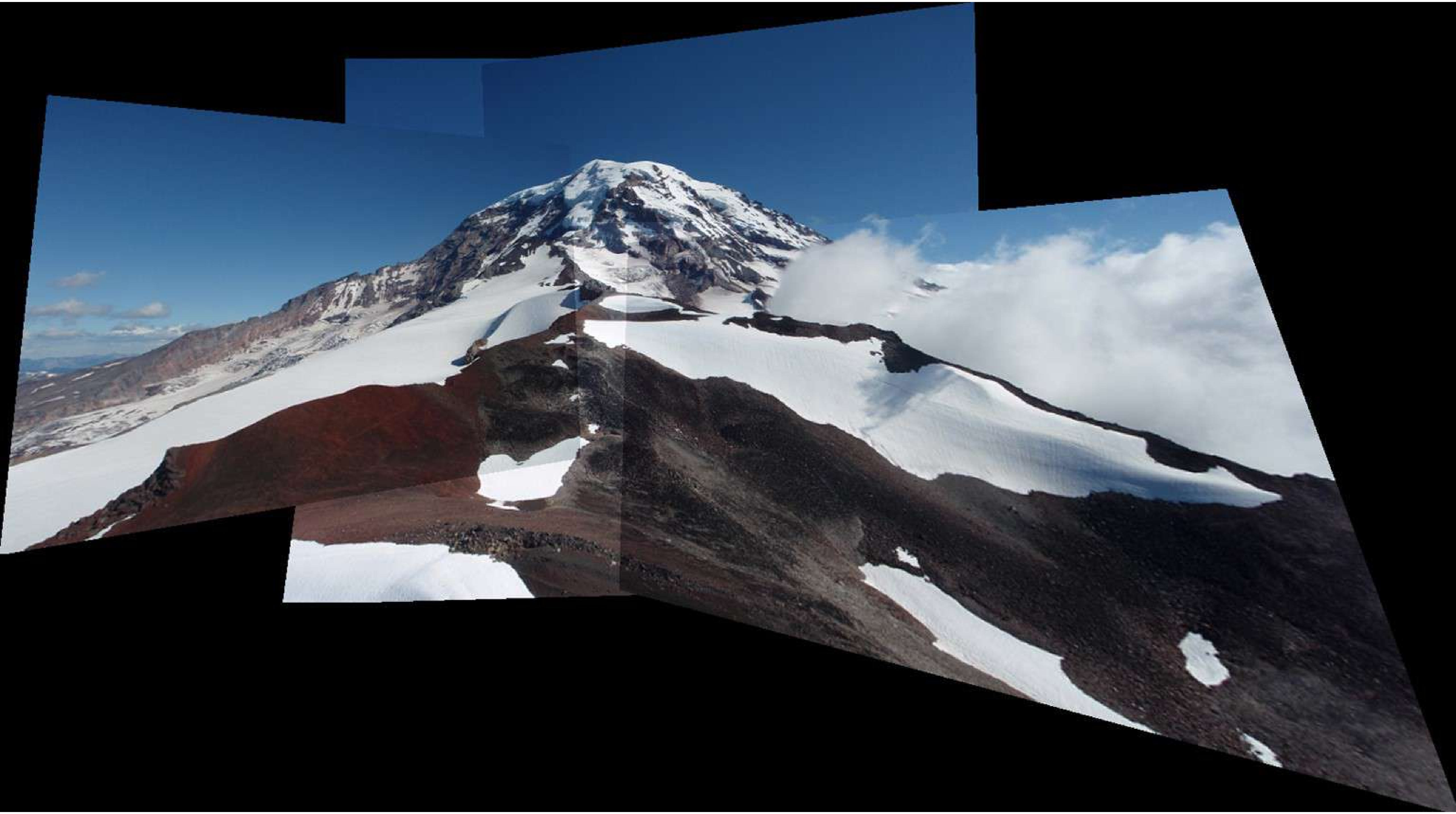
In practice

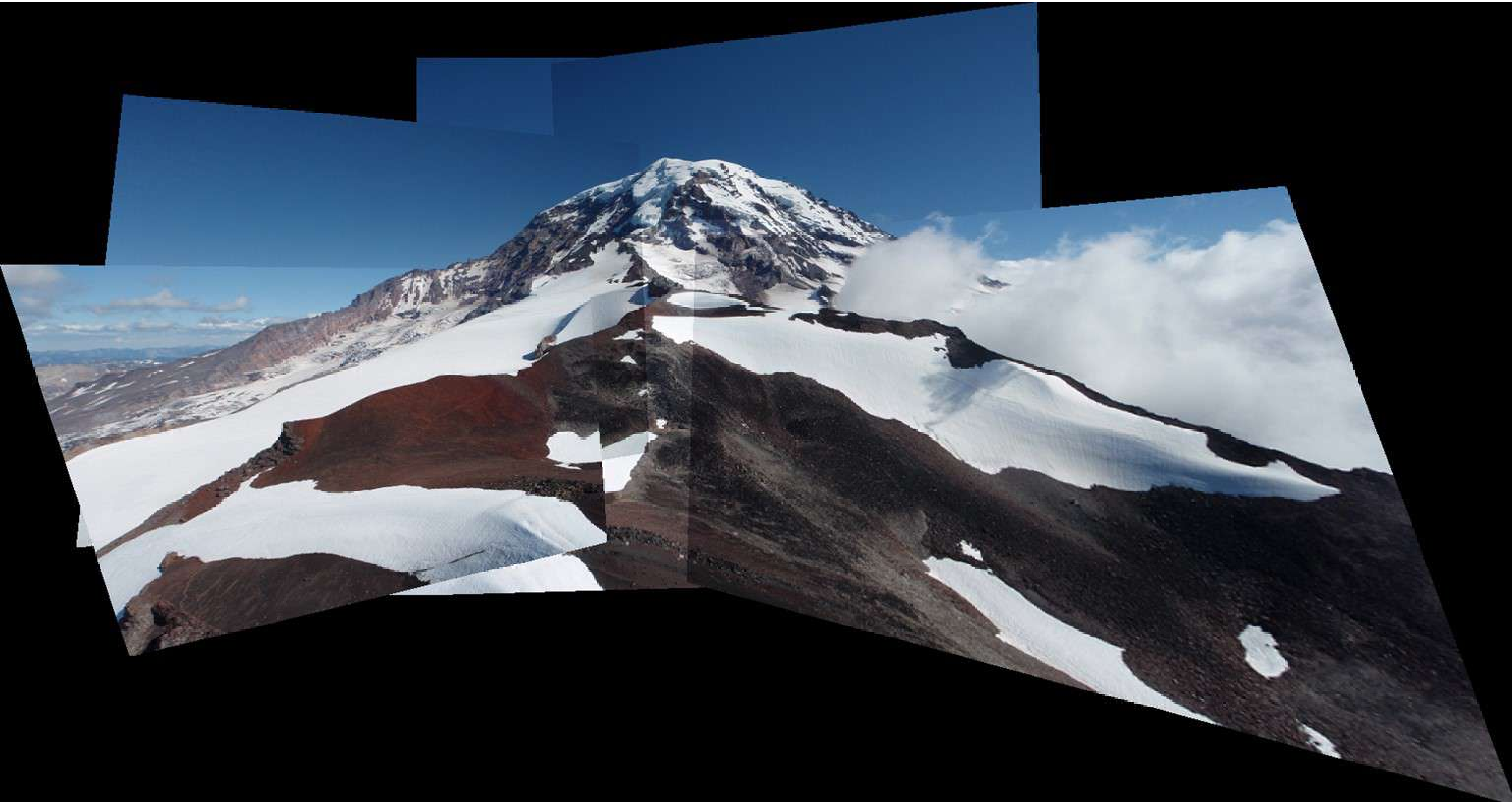
In practi



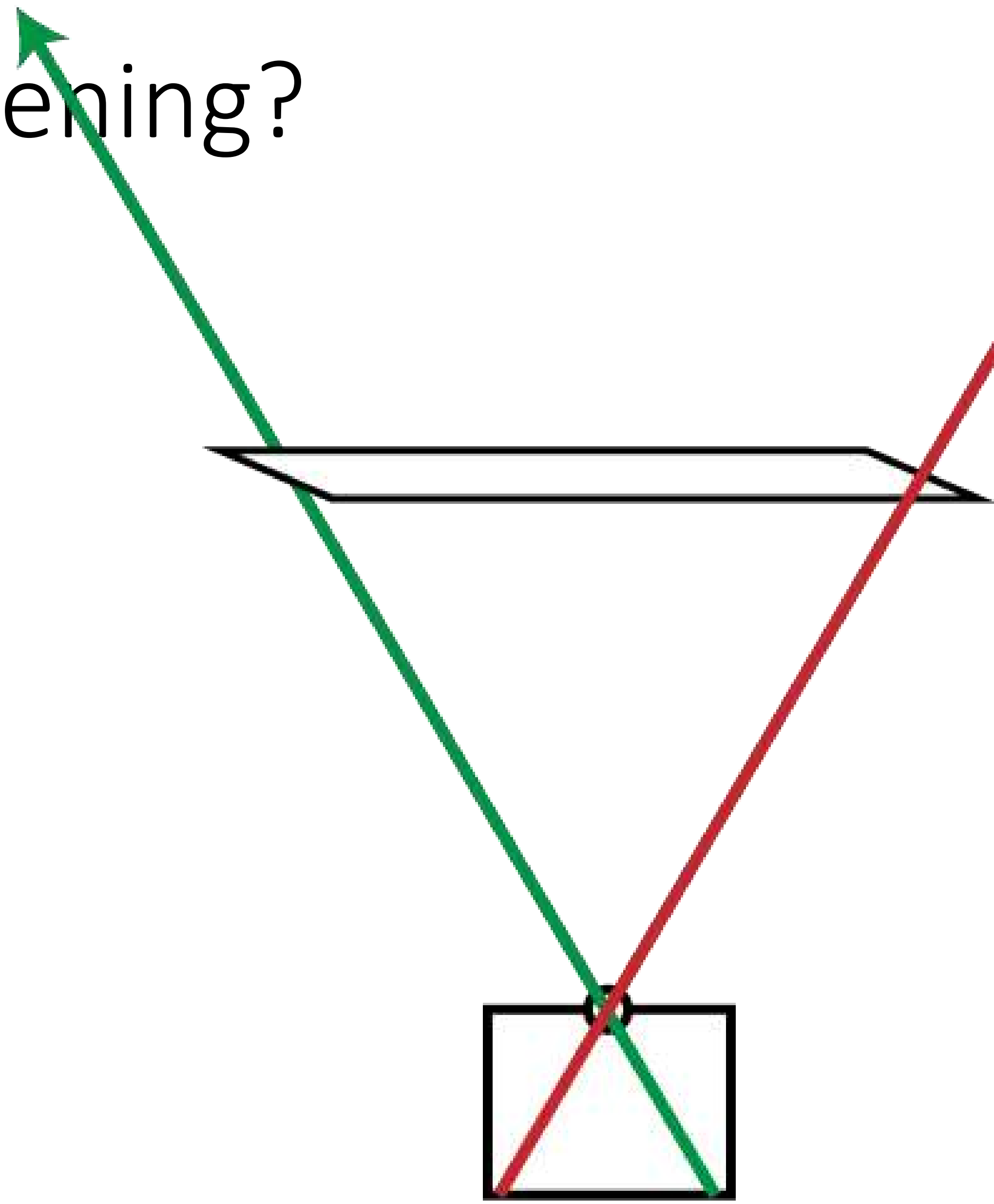
In p



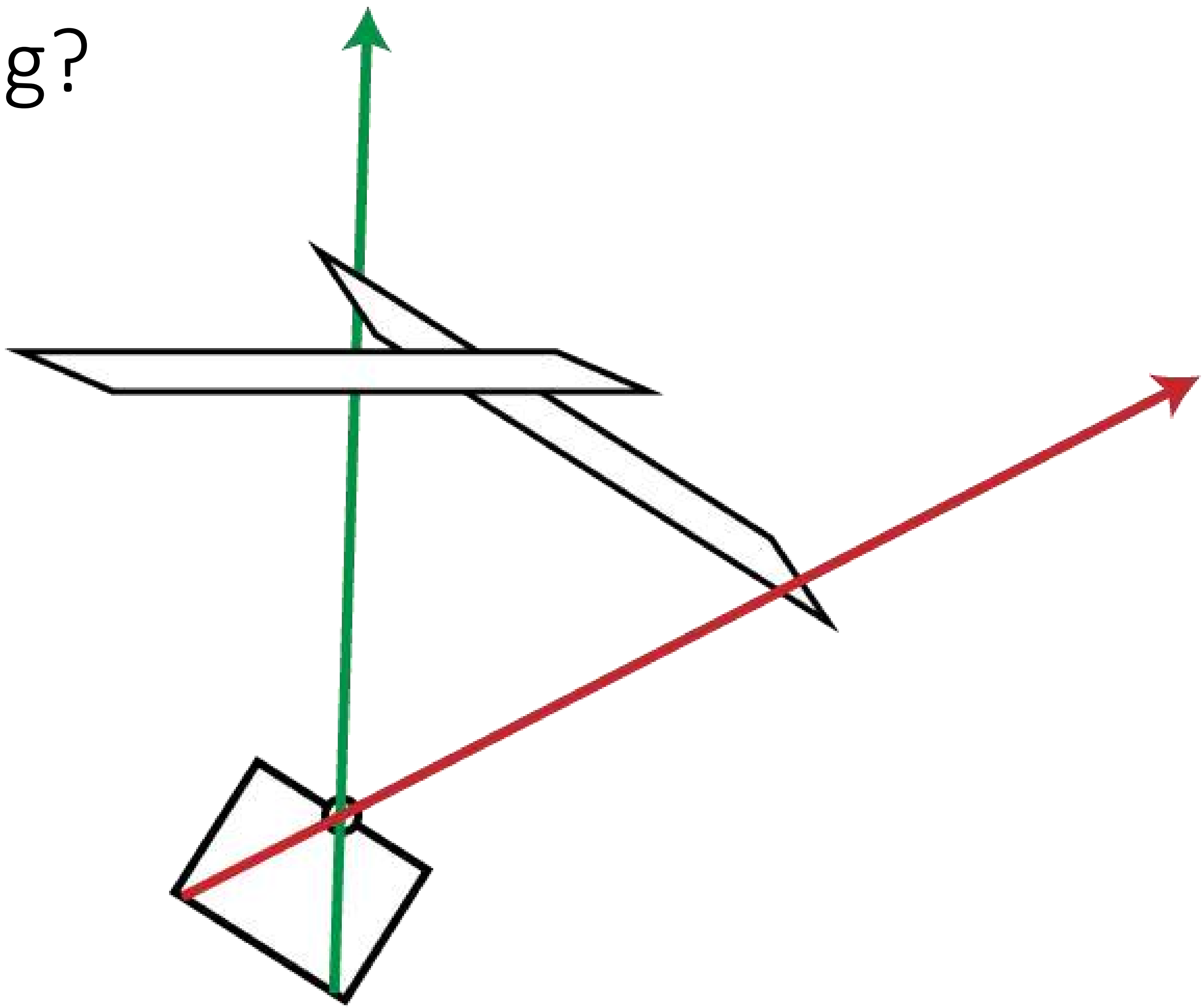




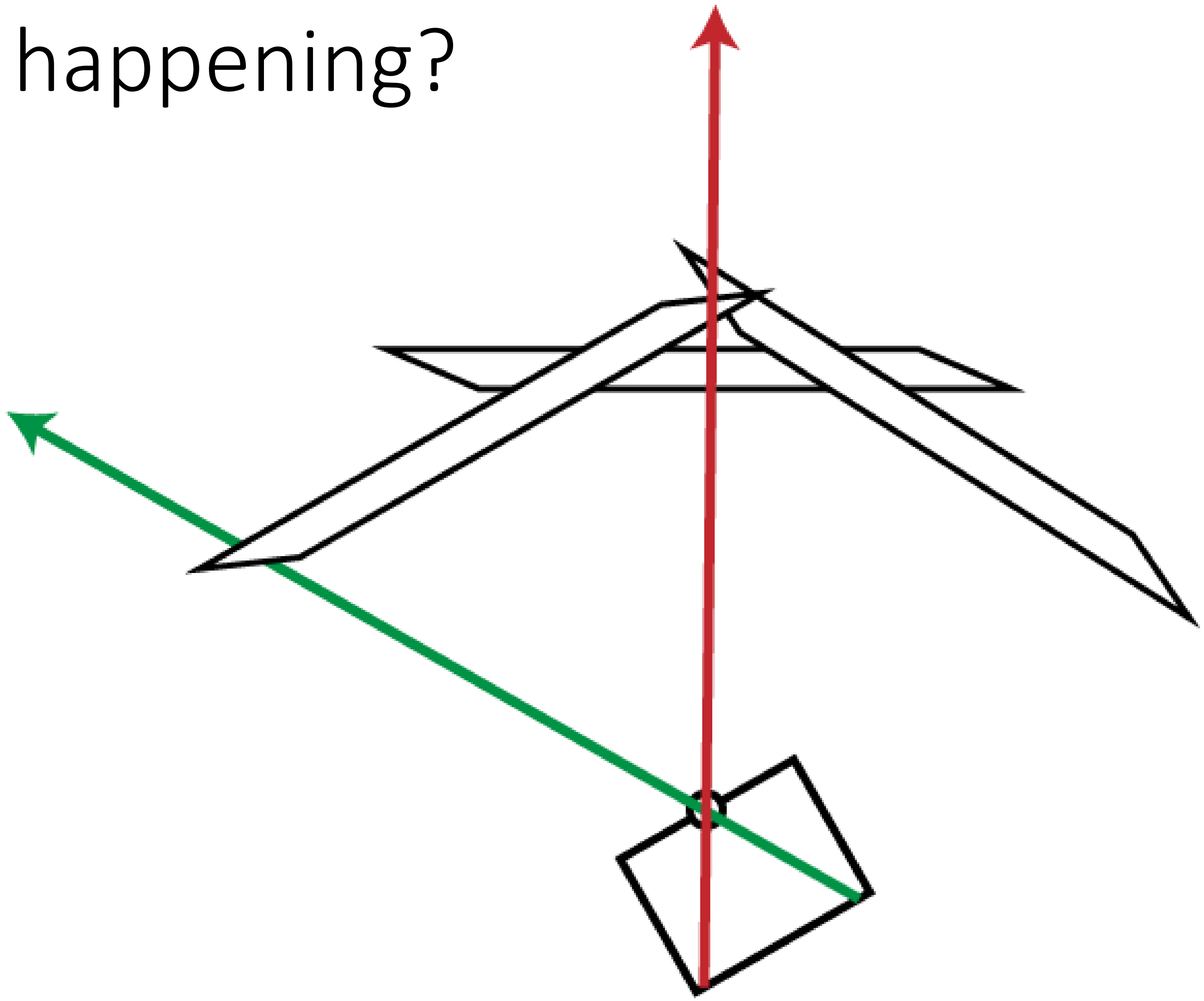
What's happening?



What’s happening?



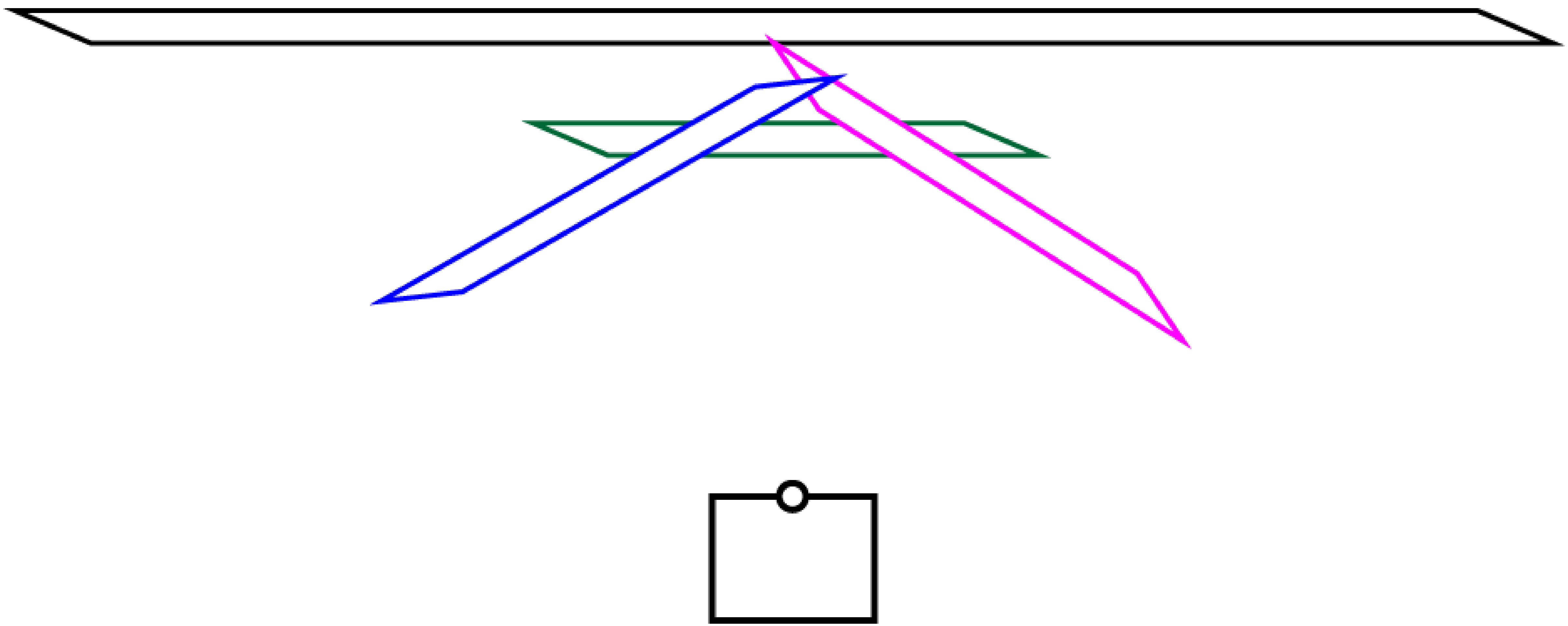
What’s happening?



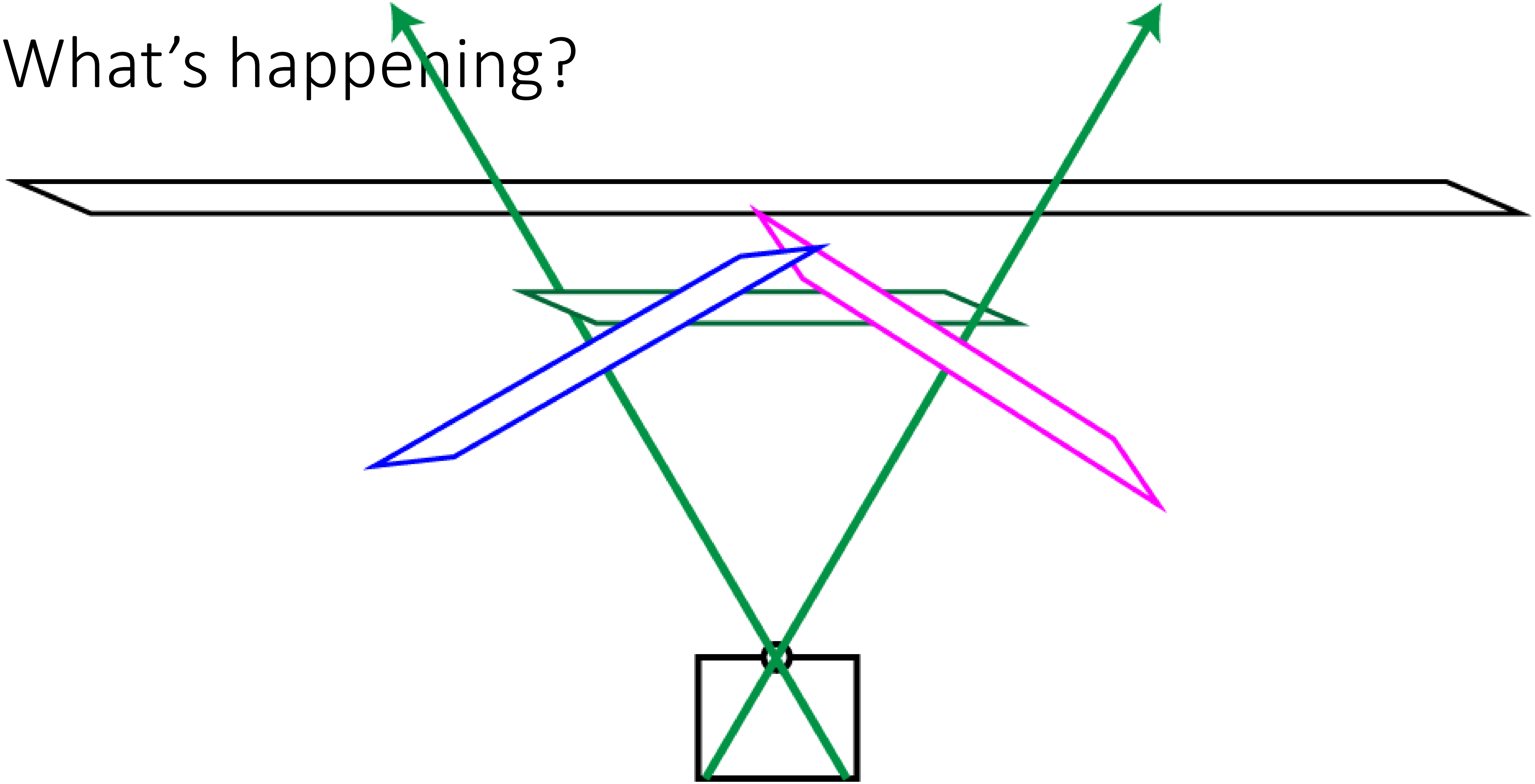
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

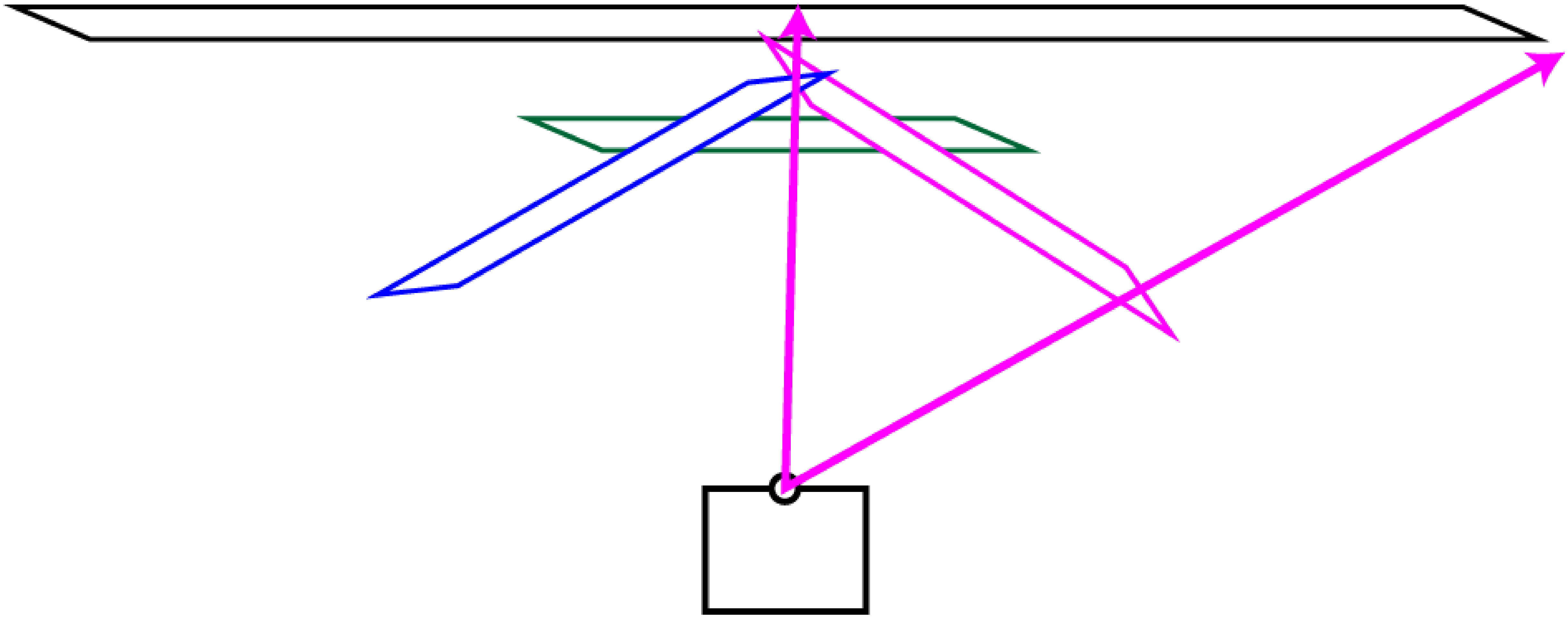
What’s happening?



What's happening?



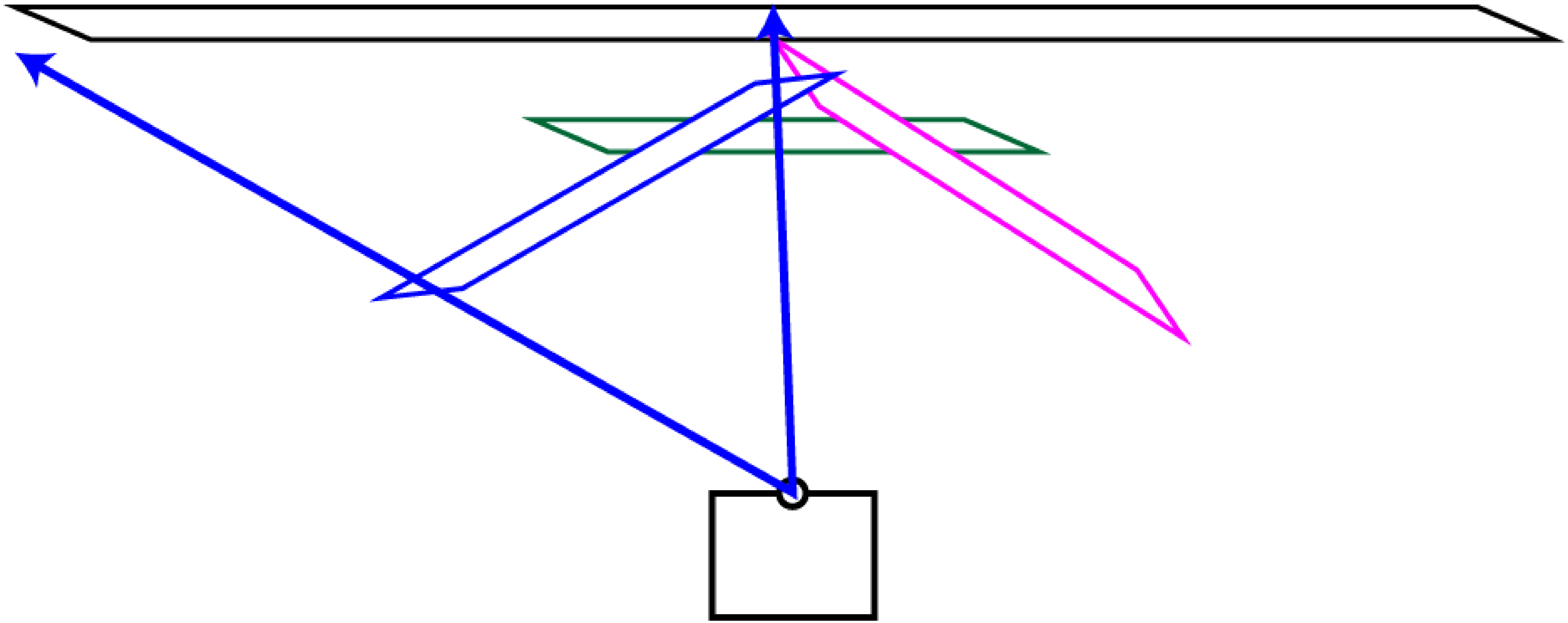
What's happening?



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

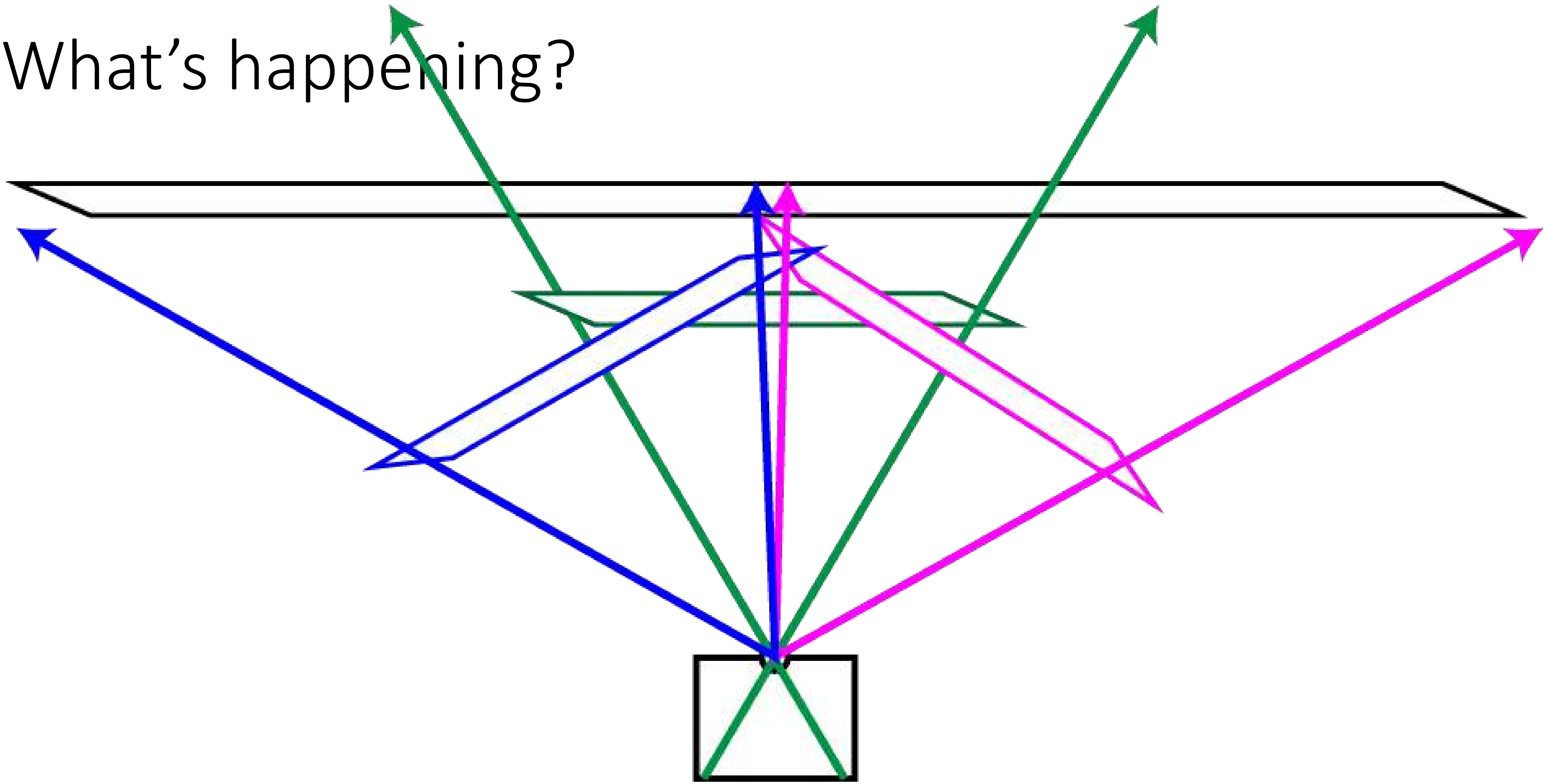
What’s happening?



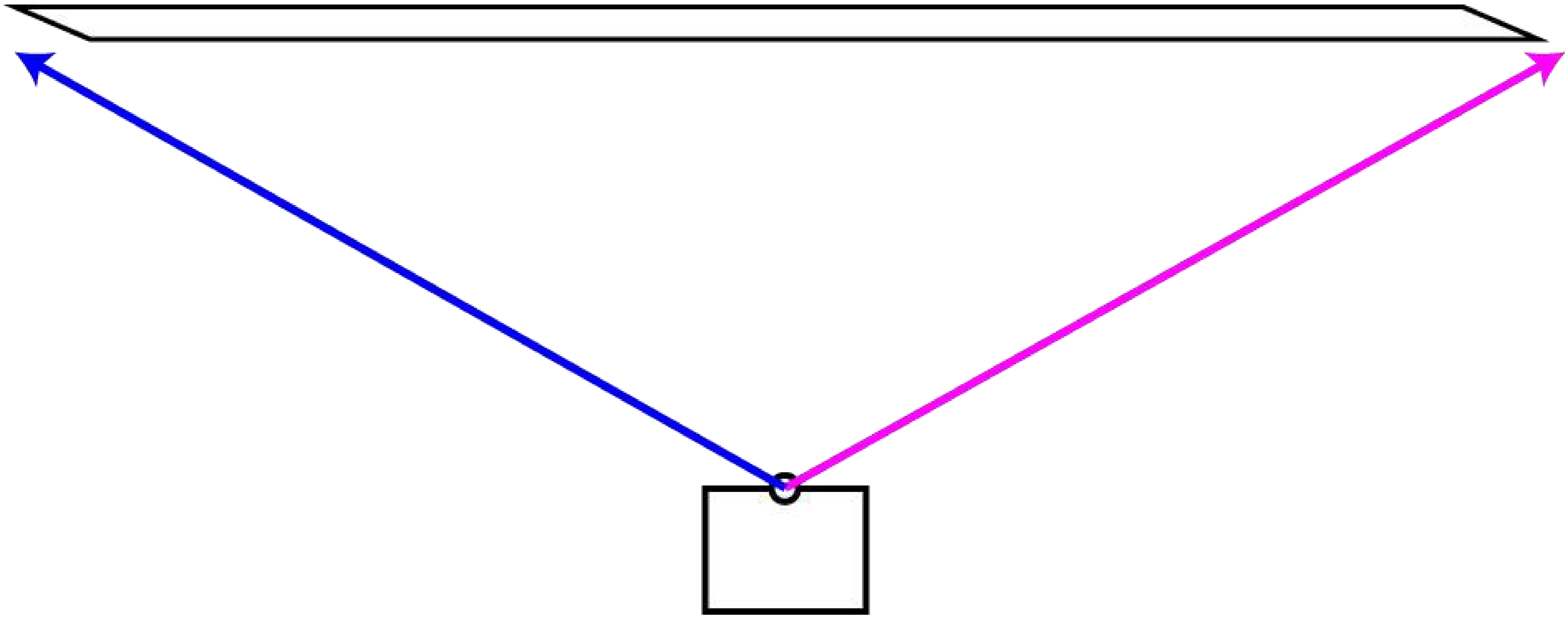
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What's happening?



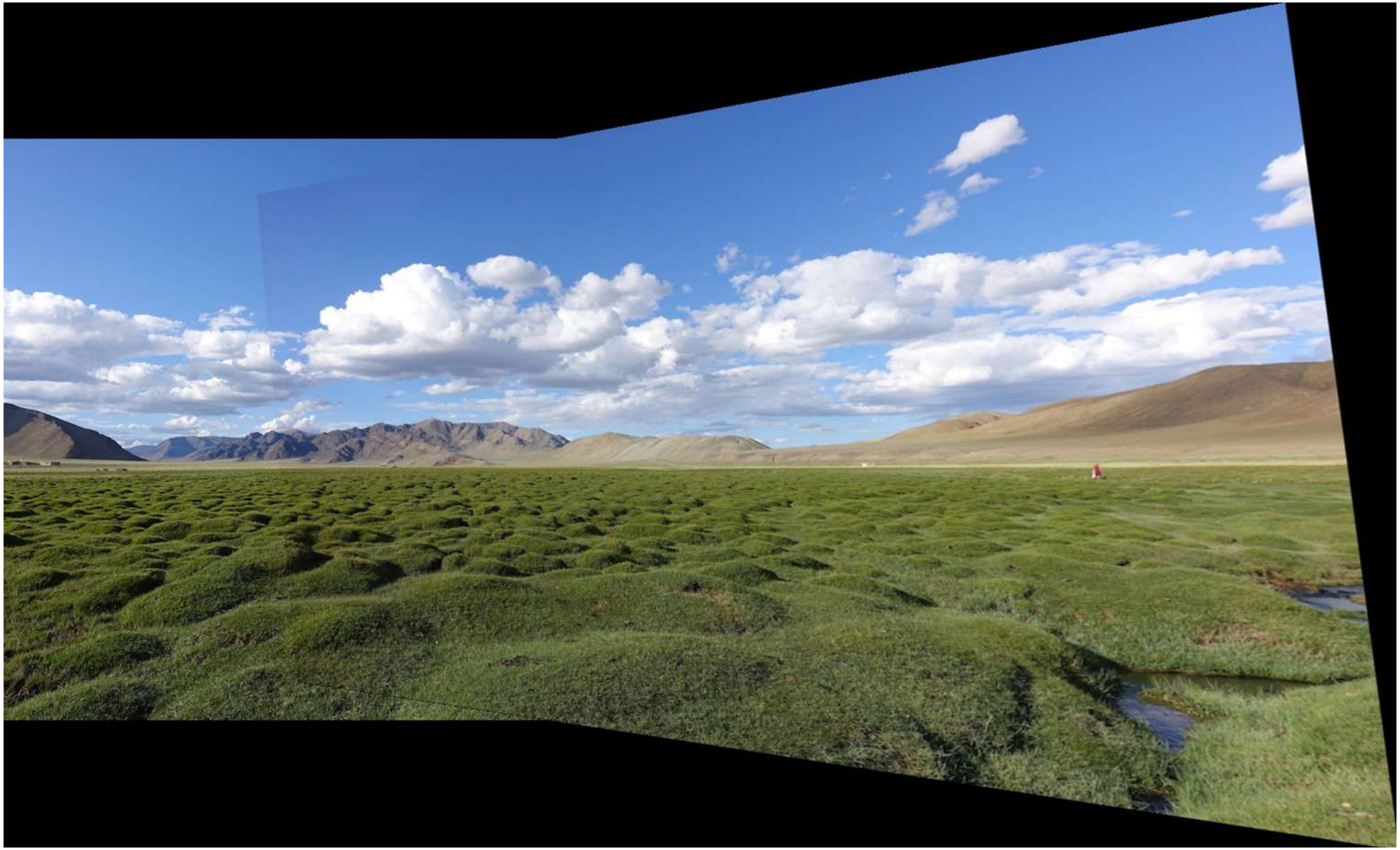
What’s happening?



Very bad for big panoramas!

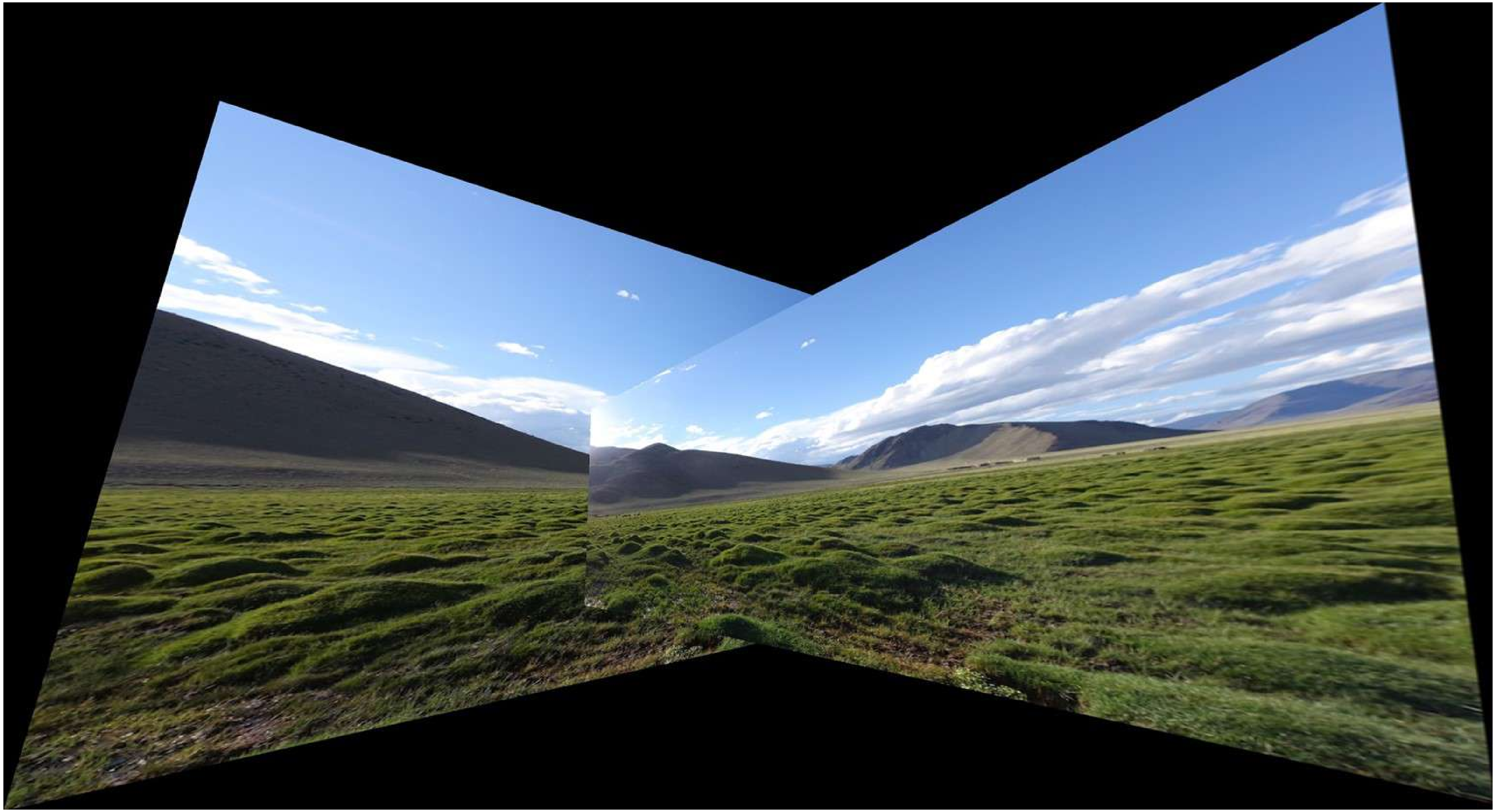
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Very bad for big panoramas!



Very bad for big panoramas!

Fails :-(

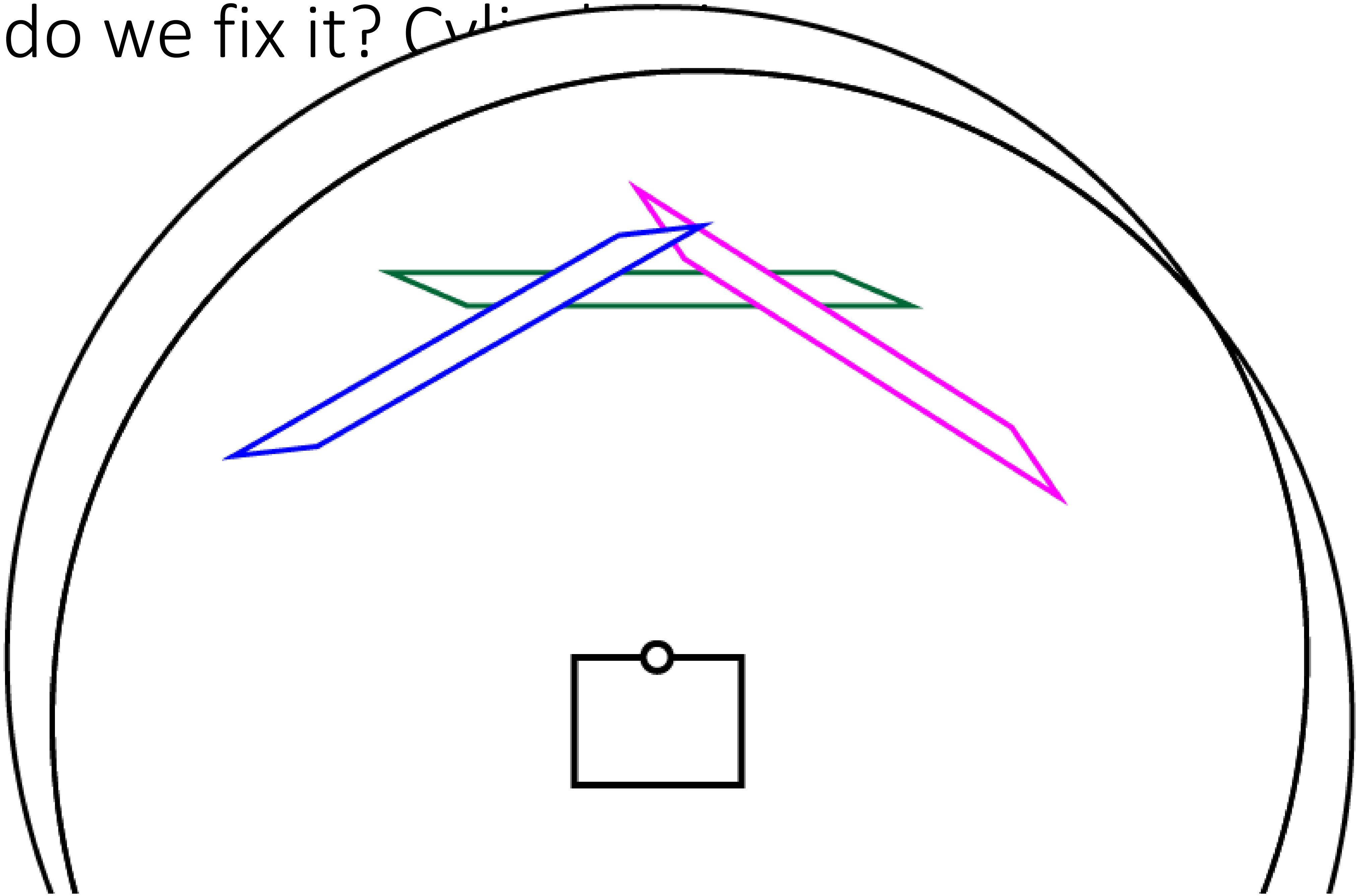


This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How do we fix it? Cylinders!

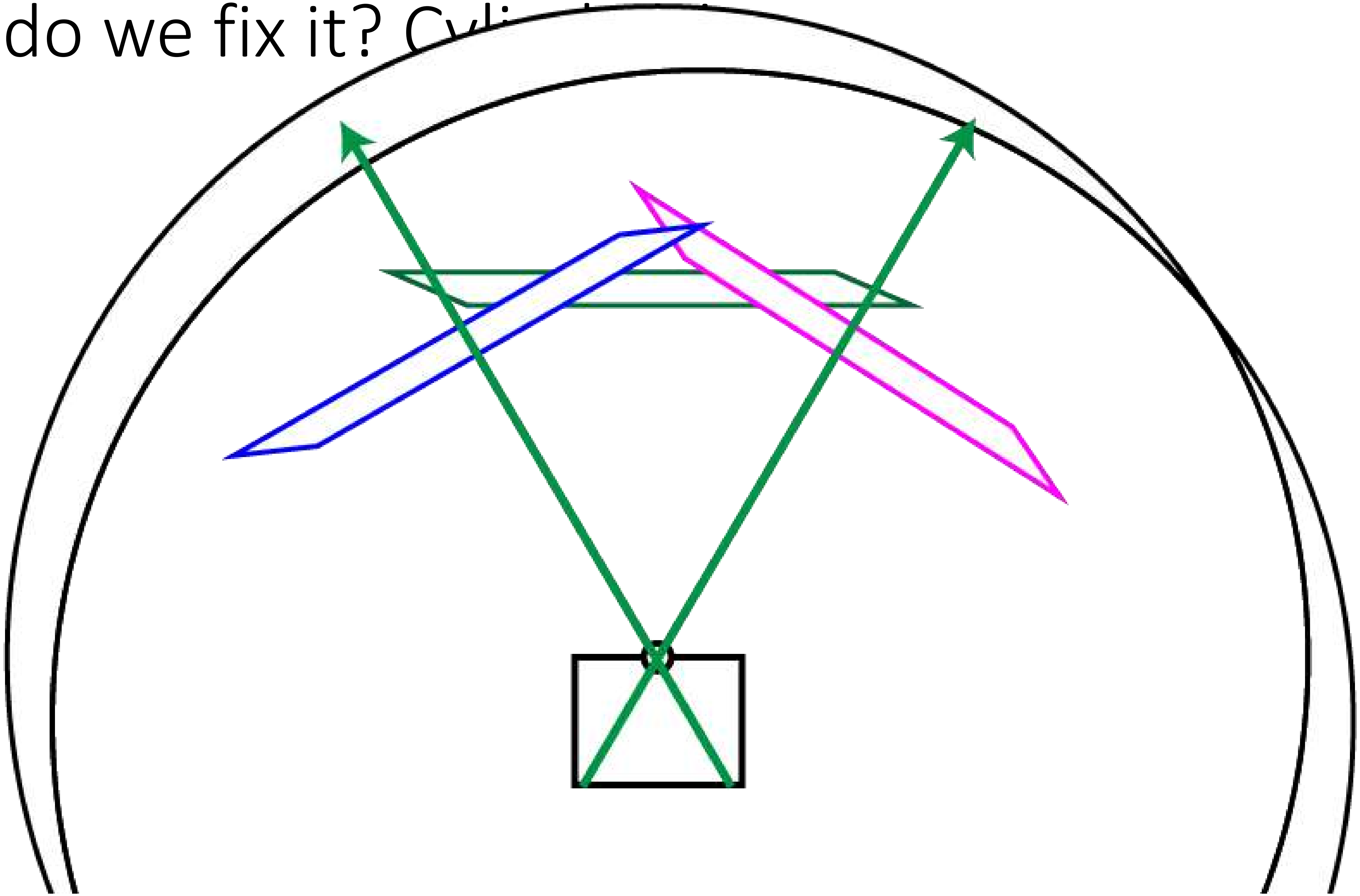
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

How do we fix it? Cyl

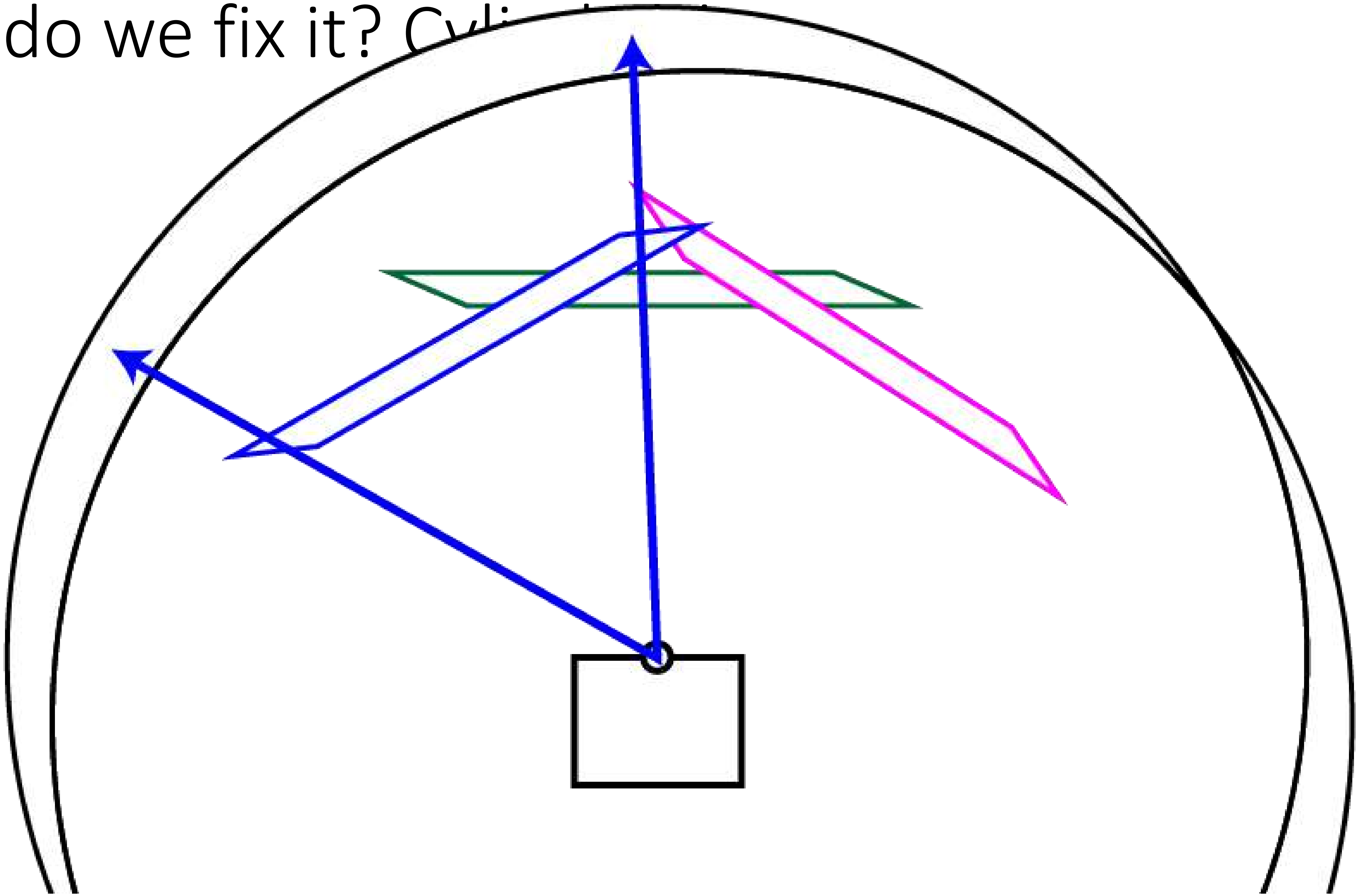


Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

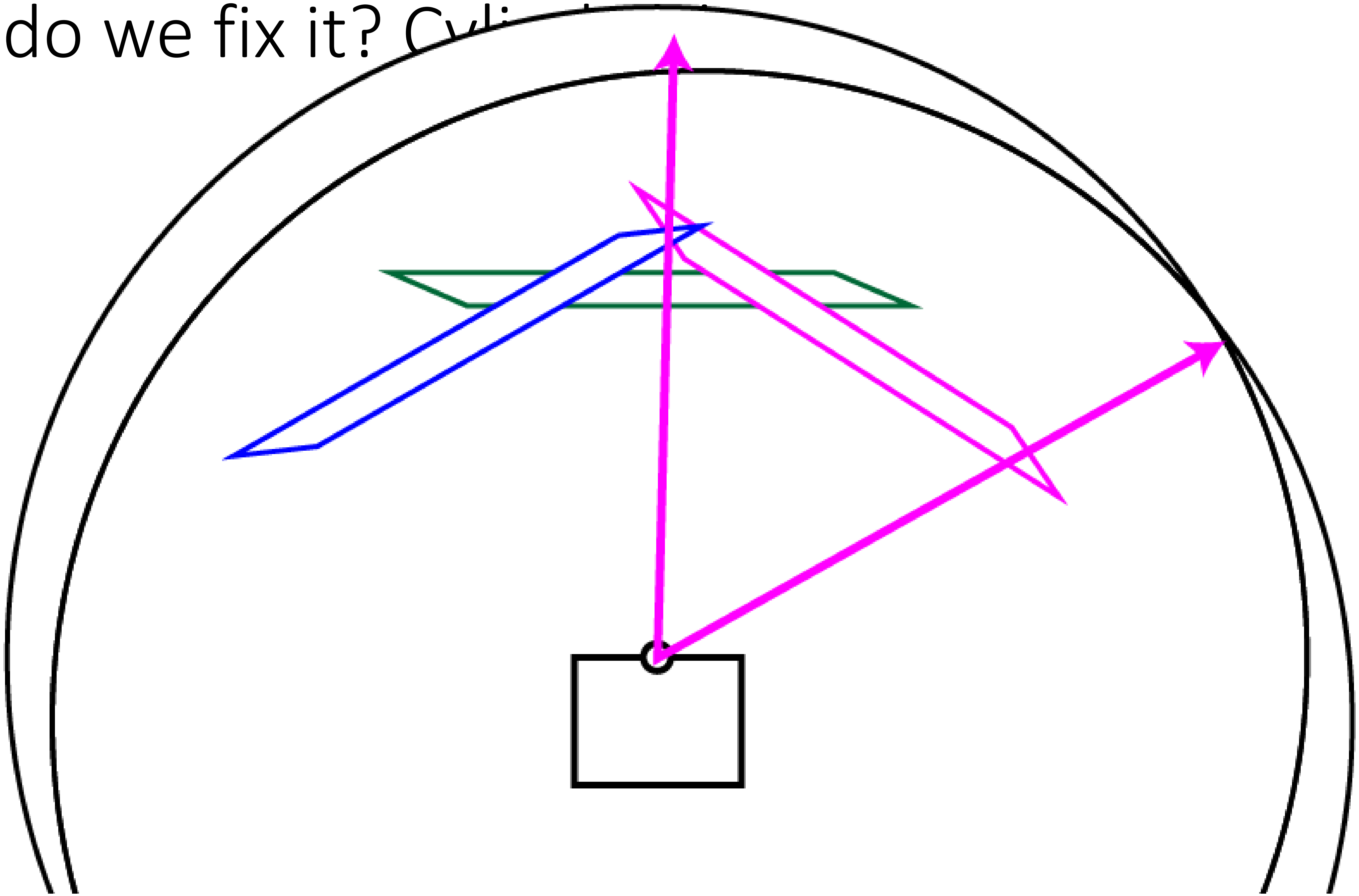
How do we fix it? Cyl



How do we fix it? Cyl



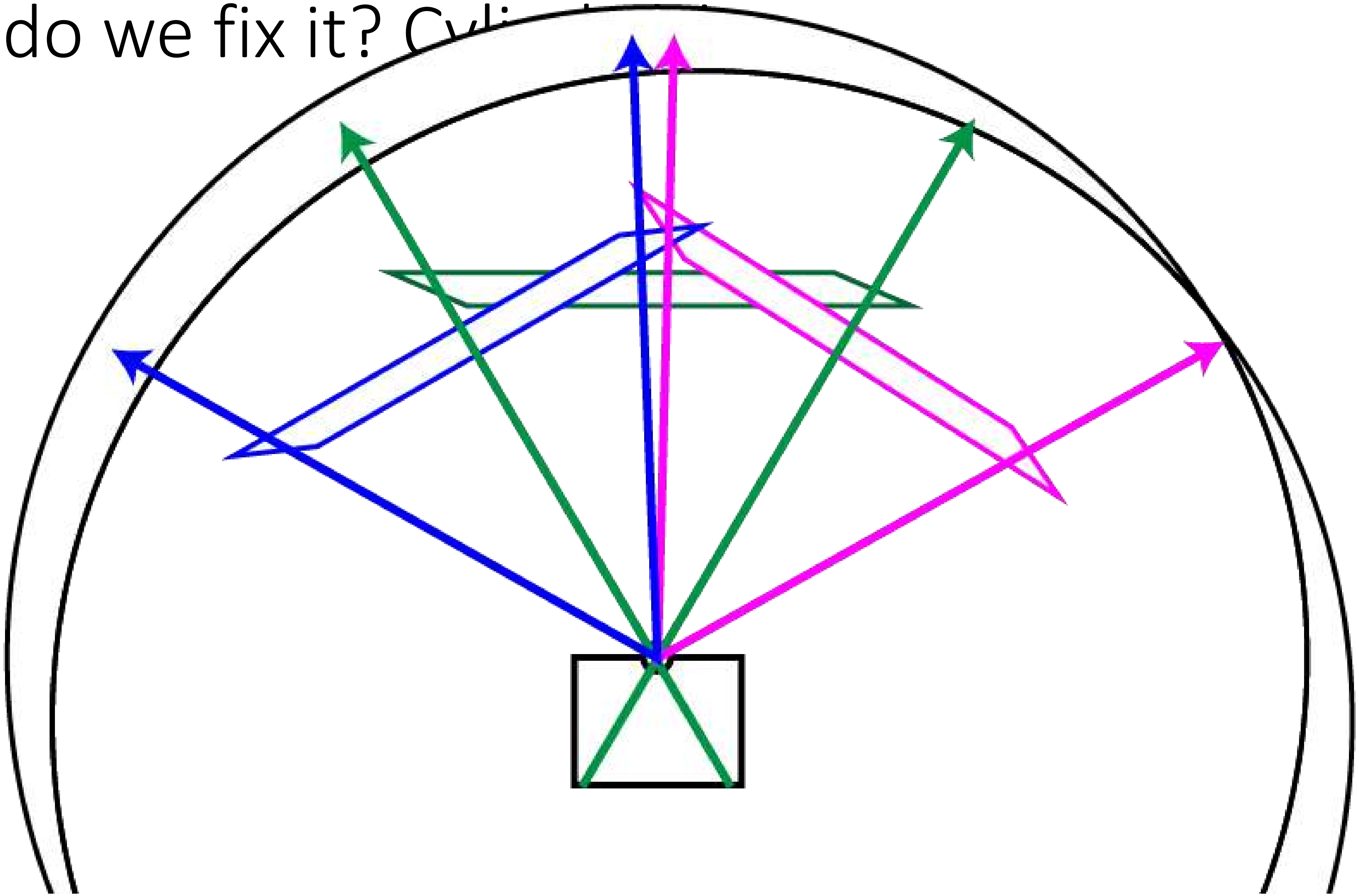
How do we fix it? Cyl



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How do we fix it? Cyl



How do we fix it? Cylinders!

Calculate angle and height:

$\theta = (x - xc) / f$

$h = (y - yc) / f$

Find unit cylindrical coords:

$X' = \sin(\theta)$

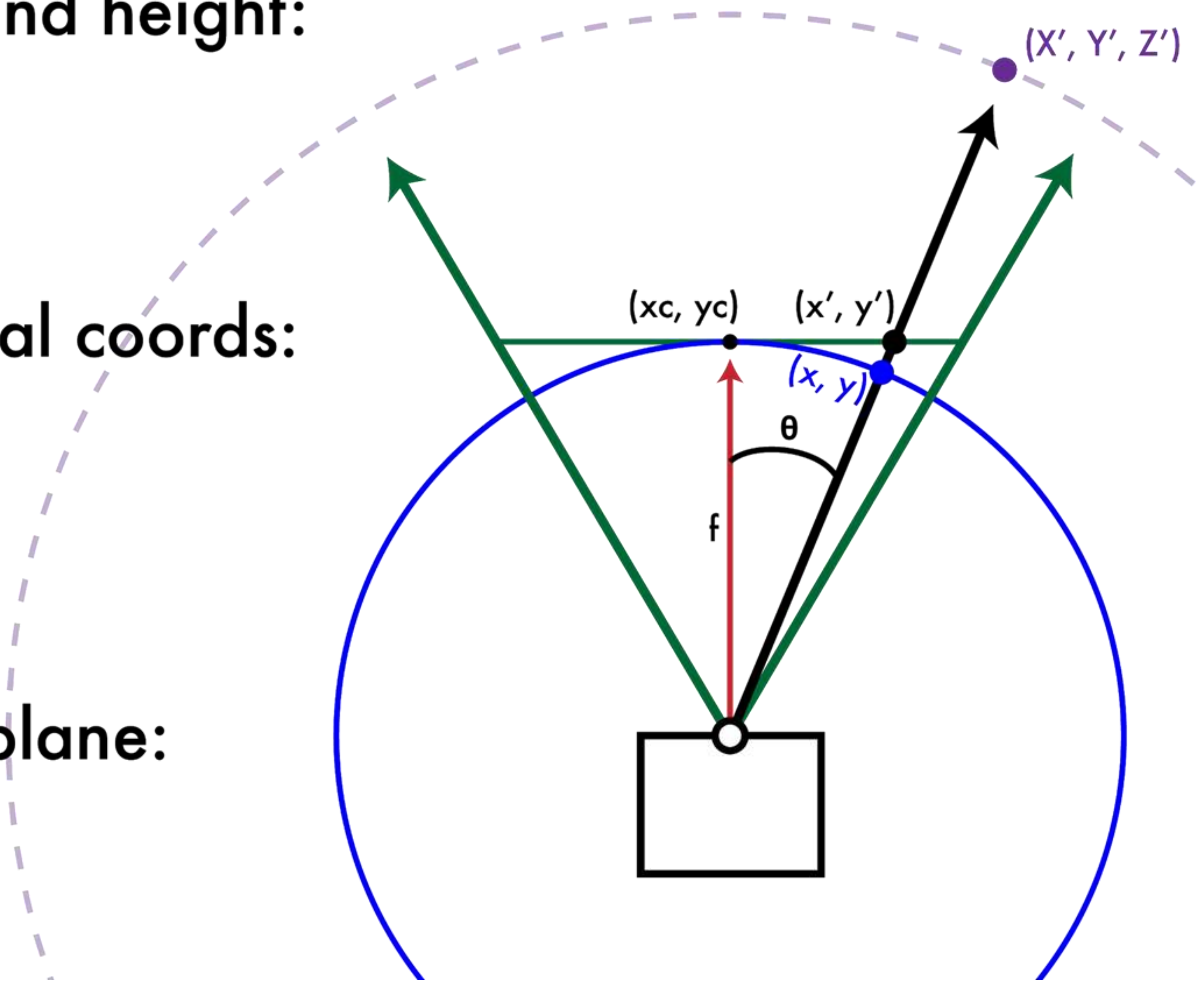
$Y' = h$

$Z' = \cos(\theta)$

Project to image plane:

$x' = f\, X'/Z' + xc$

$y' = f\, Y'/Z' + yc$



Dependant on focal length!

f = 300

f = 500

f = 1000

f = 1400

f = 10,000

f = 10,000

Does it work?

Does it work?

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Does it work?

Does it work?

Does it work?

Yes! Assuming camera is level and rotating around its vertical axis

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 10: Visual Recognition – Segmentation

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Linear least-squares
- RANSAC
- Panorama Stitching

Today’s Agenda

- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

Reading: Forsyth Chapter 9

D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, TPAMI 2002

[material based on Niebles-Krishna]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

Reading: Forsyth Chapter 9

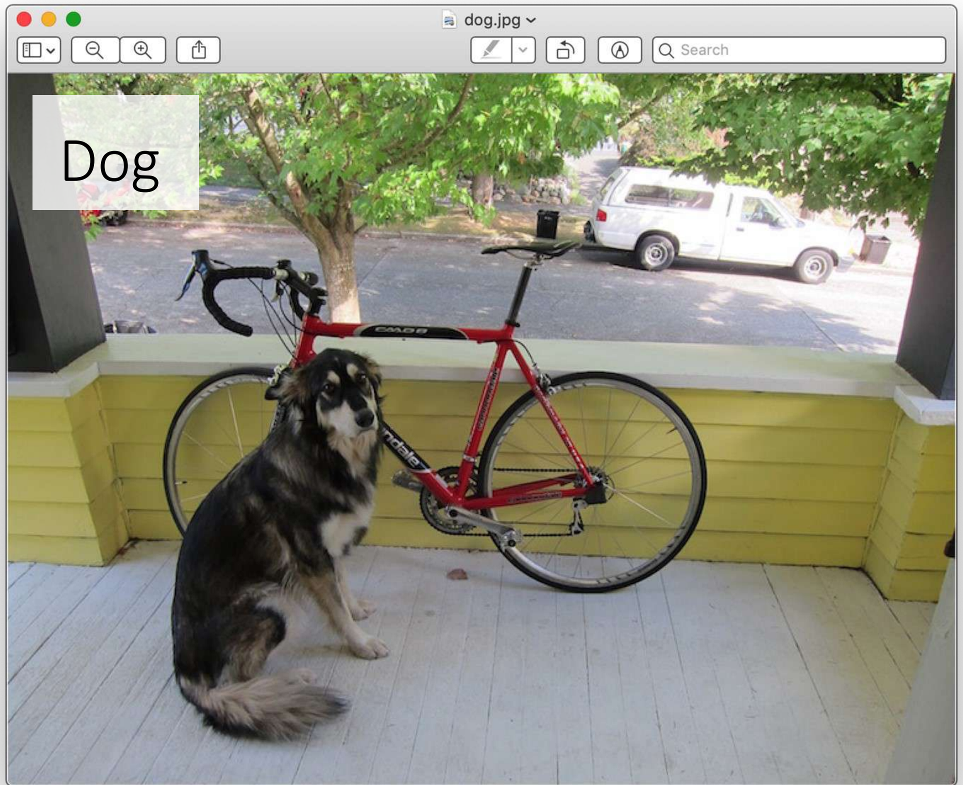
D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, TPAMI 2002

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Classification

- What is in the image?

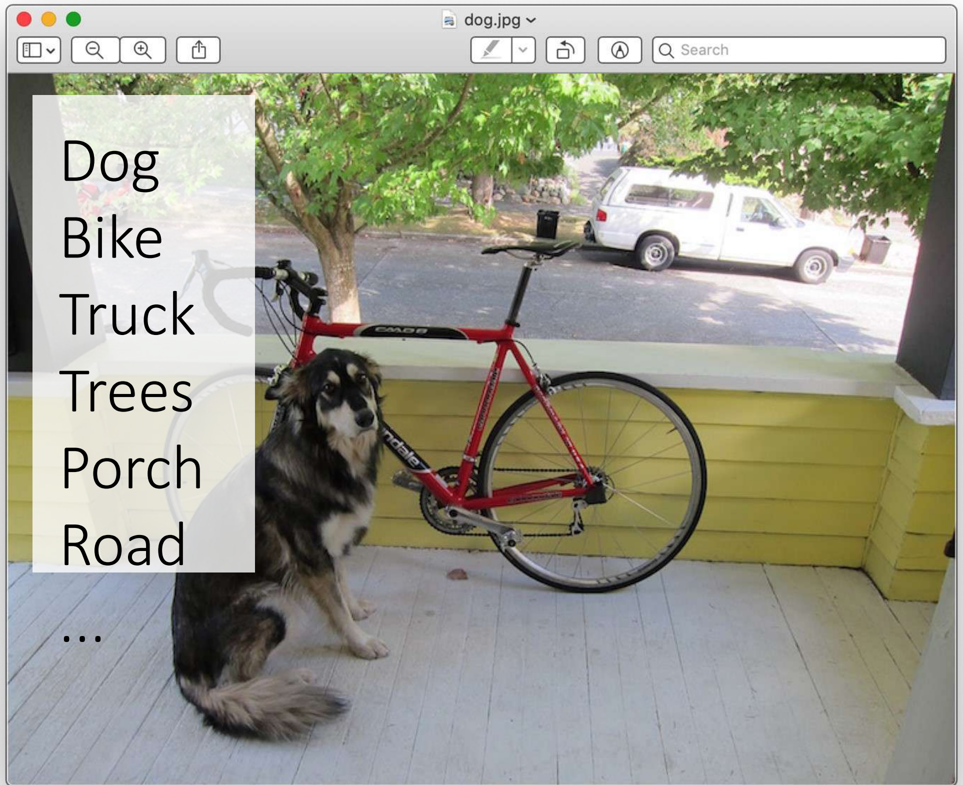


[Redmon]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Tagging

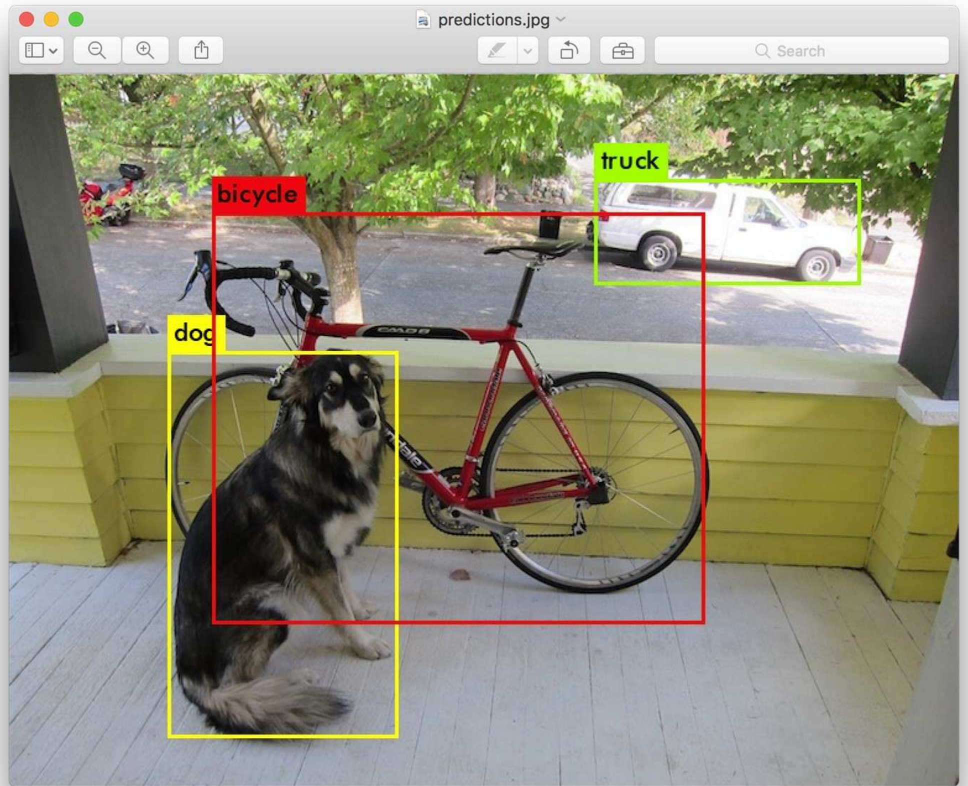
- What are ALL the things in the image?



[Redmon]

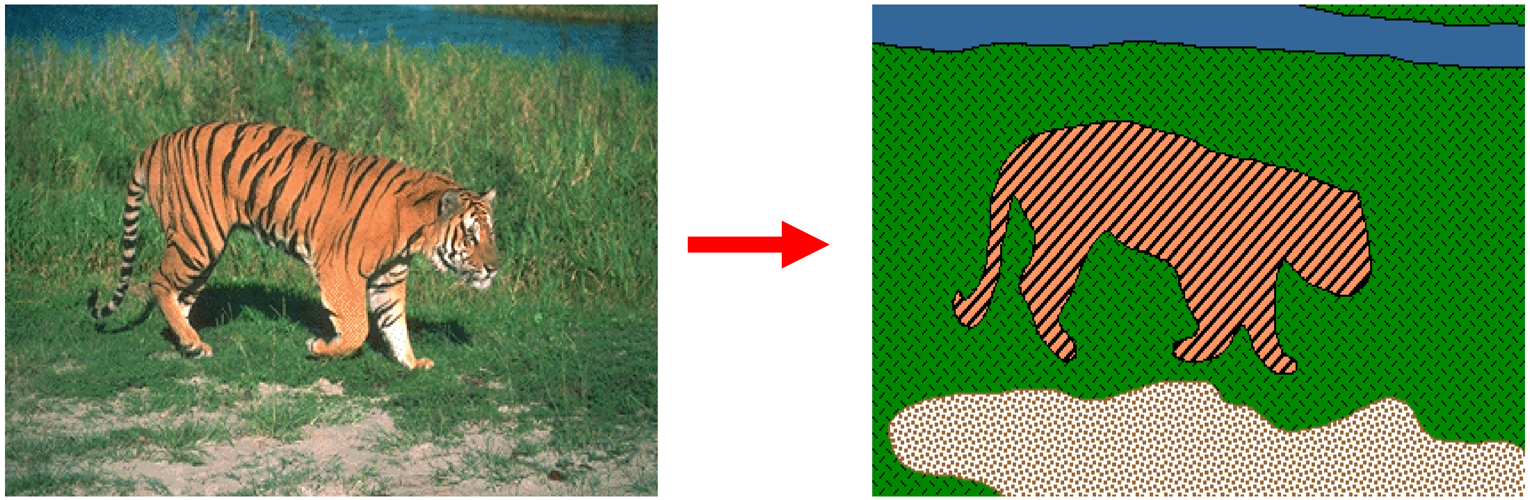
Detection

- What are ALL the things in the image?
- Where are they?



[Redmon]

Segmentation



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

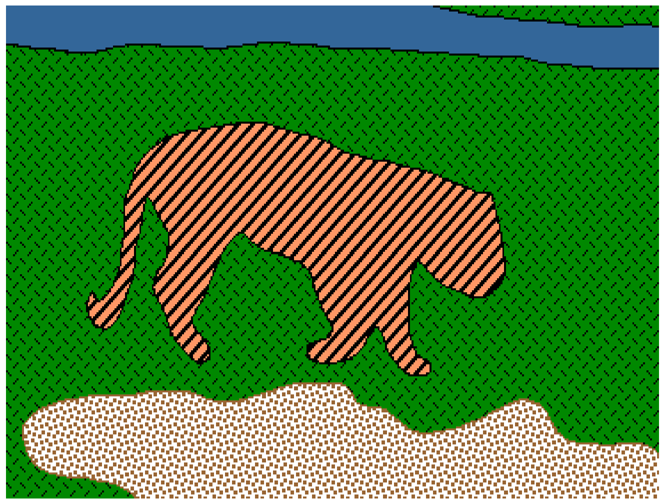
- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

Reading: Forsyth Chapter 9

D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, TPAMI 2002

Image Segmentation

Goal: identify groups of pixels that go together



Slide credit: Steve Seitz, Kristen Grauman

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

The Goals of Segmentation

- Separate image into coherent “objects”

Image | Human segmentation

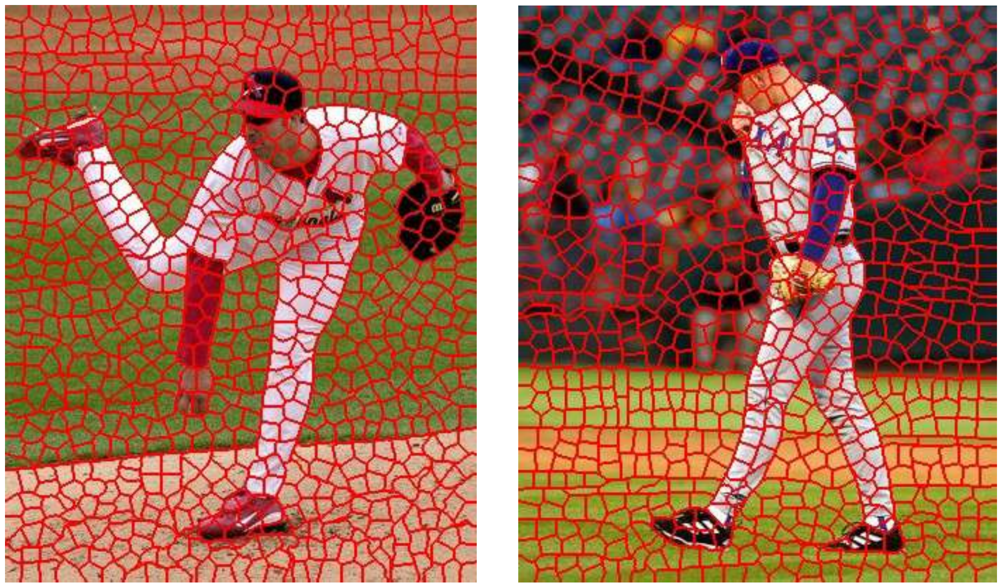
Slide credit: Svetlana Lazebnik

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The Goals of Segmentation

- Separate image into coherent “objects”
- Group together similar-looking pixels for efficiency of further processing

“superpixels”



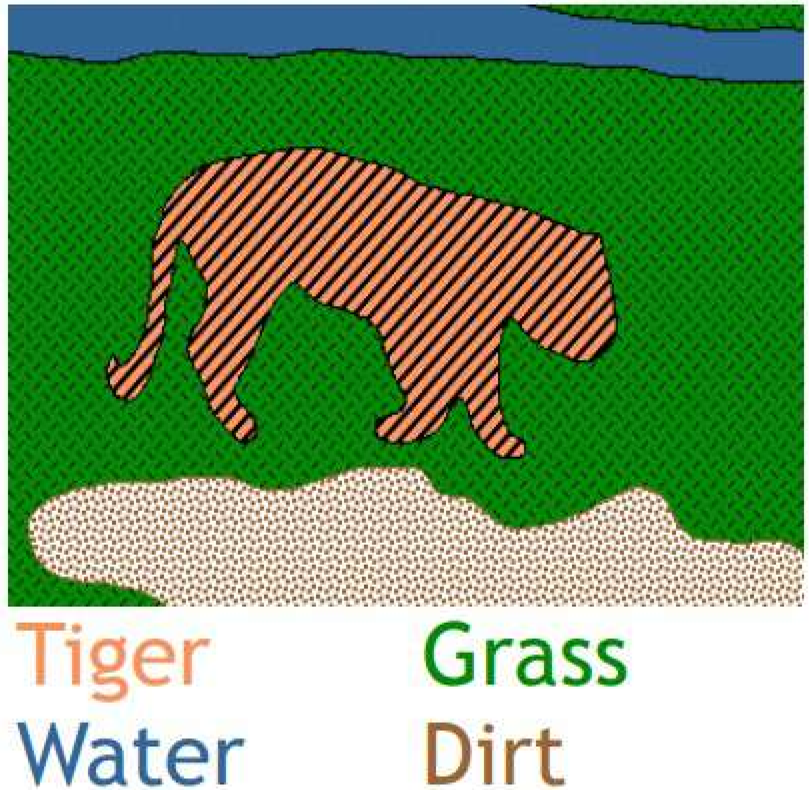
X. Ren and J. Malik. [Learning a classification model for segmentation.]() ICCV 2003.

Slide credit: Svetlana Lazebnik

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

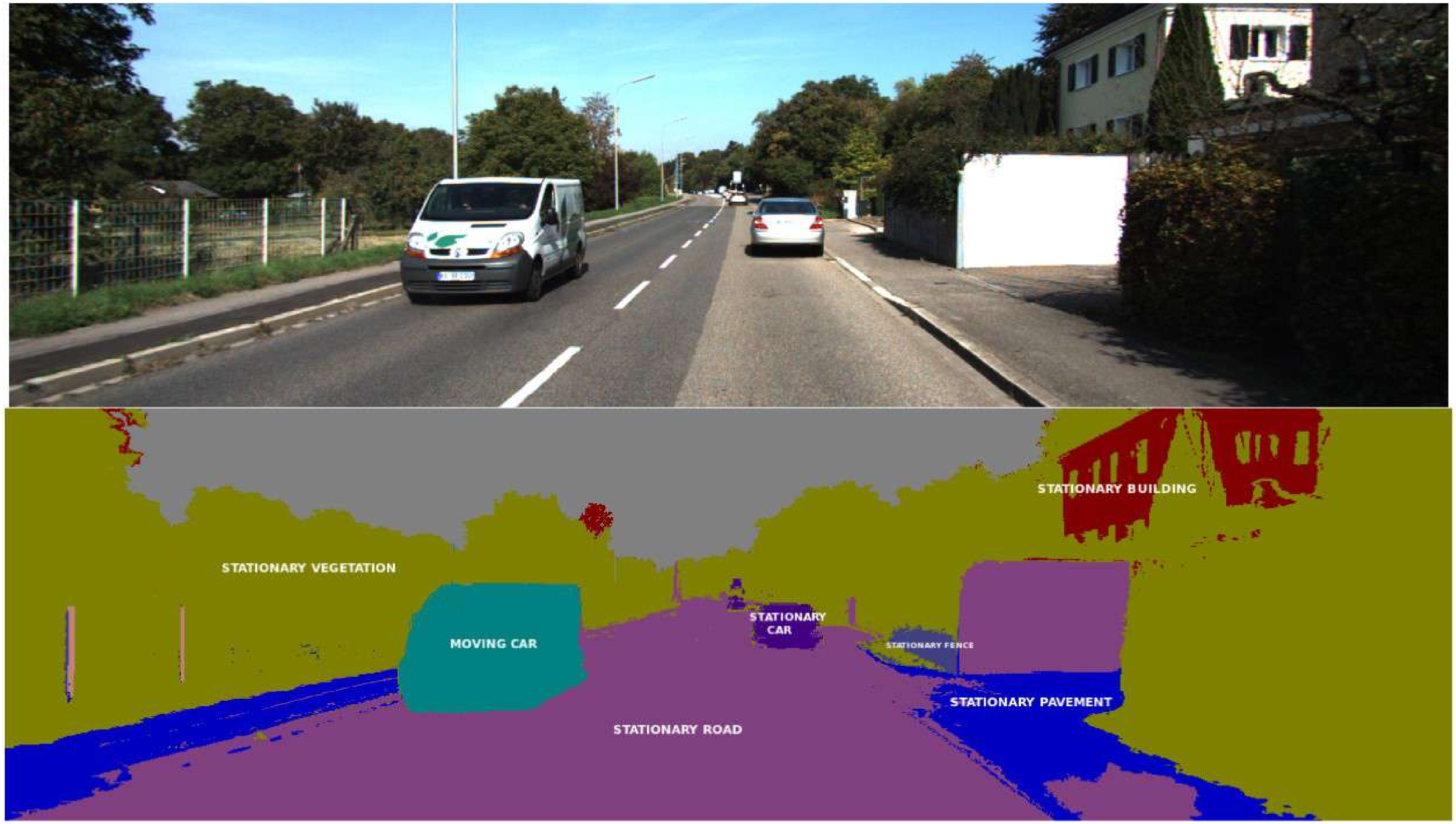
Types of Segmentation

- Semantic segmentation: Assign labels



Types of Segmentation

- Semantic segmentation: Assign labels



Types of Segmentation

- Instance segmentation: Assign labels per object

http://www.youtube.com/watch?v=OOT3UIXZztE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Types of Segmentation

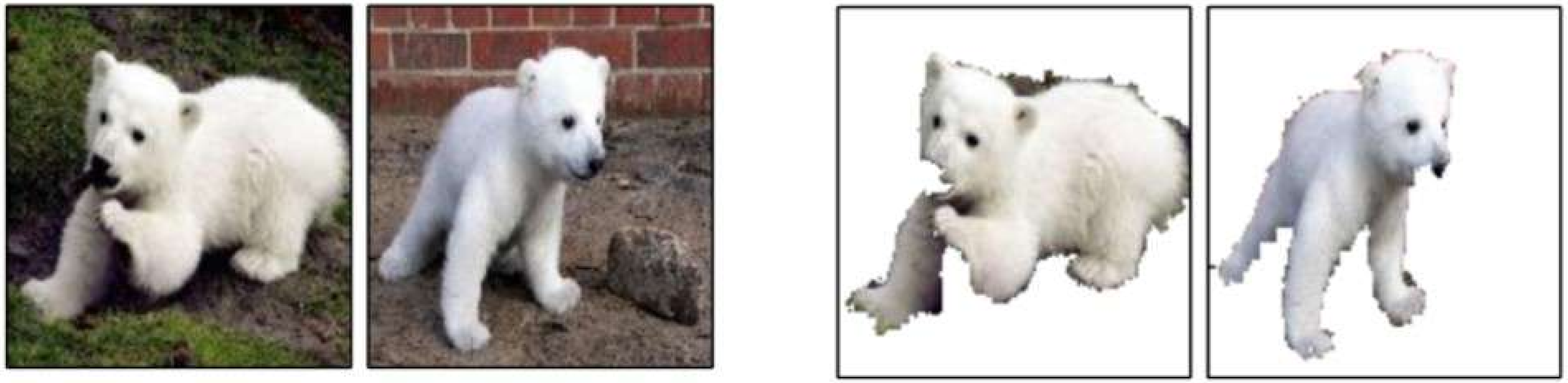
- Foreground / background segmentation

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

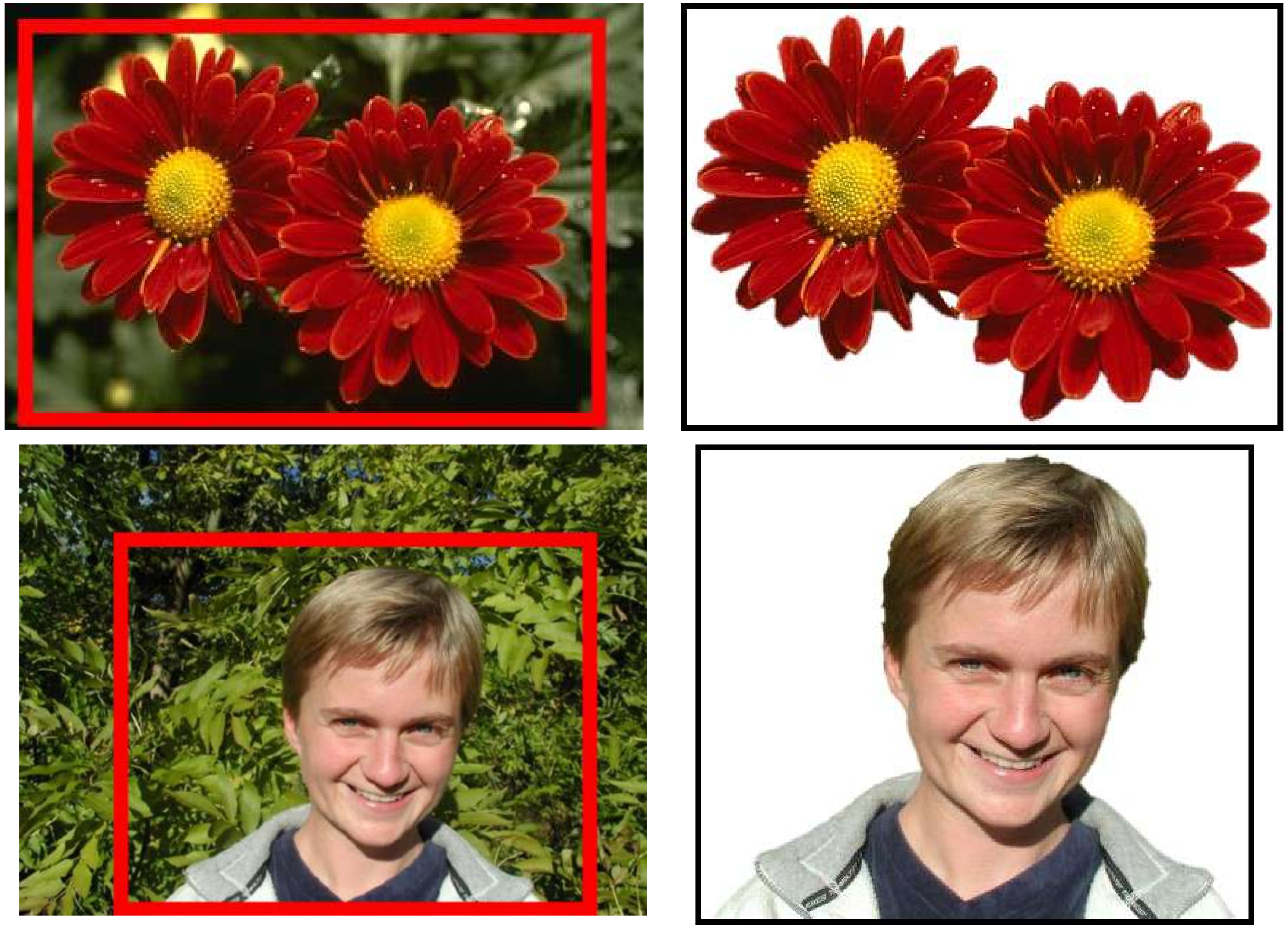
Types of Segmentation

- Co-segmentation: Segment common object in multiple images



This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Application: as a result

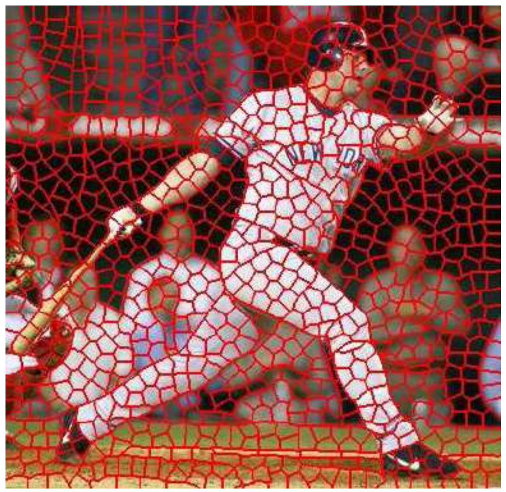


GrabCut: Rother et al. 2004

Application: for efficiency – e.g., speed up recognition

[Felzenszwalb and Huttenlocher 2004]

[Hoiem et al. 2005, Mori 2005]



[Shi and Malik 2001]

Slide: Derek Hoiem

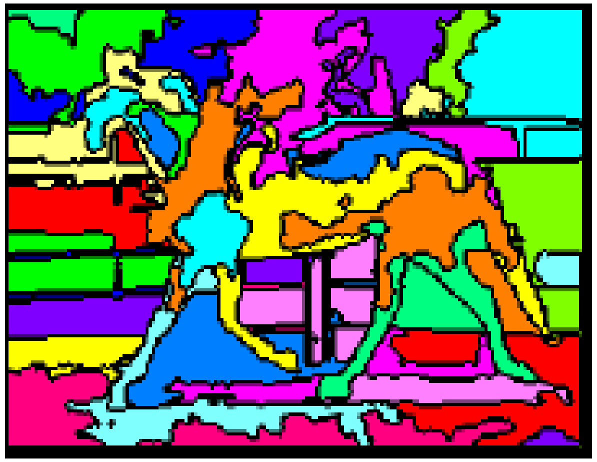
Co-financed by the European Union
Connecting Europe Facility

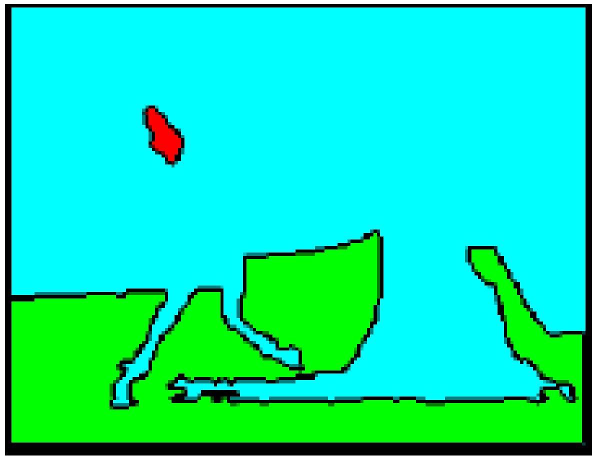
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

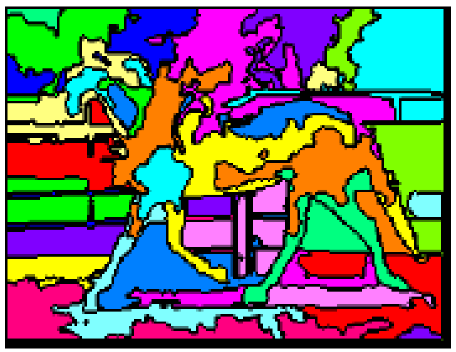
Application: better classification

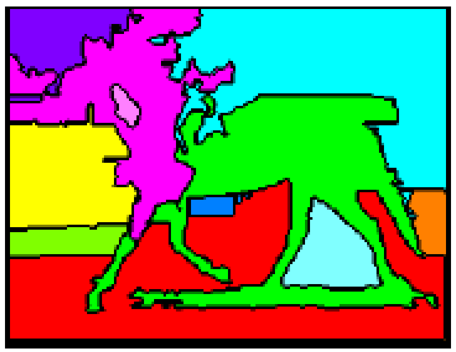
Angelova and Zhu, 2013

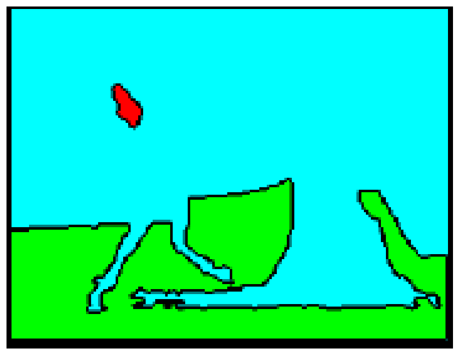
Over/under segmentation


Oversegmentation


Undersegmentation





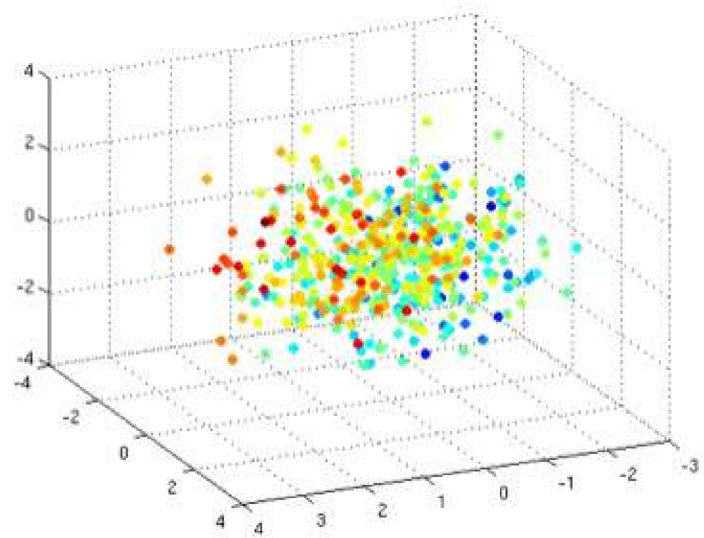

Multiple Segmentations

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

One way to think about segmentation is **Clustering**

- Pixels are points in a (high-dimensional) feature space, e.g.
  - color: 3D
  - color + location: 5D
- Cluster pixels into segments



One way to think about segmentation is **Clustering**

Clustering: group together similar data points and represent them as a single *entity*

Clustering is an unsupervised learning method

Key Challenges:

1) What makes two points/images/patches similar?

2) How do we compute an overall grouping from pairwise similarities?

Slide: Derek Hoiem

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Distance vs Similarity Measures

Let x and x' be two objects from the dataset.
The distance or similarity between x and x' is a real number, dist(x, x') or sim(x, x')

- The Euclidian distance is defined as

$$dist(x, x') = \sqrt{\sum (x_i - x'_i)^2}$$

- In contrast, cosine similarity measure would be

$$\begin{aligned} sim(x, x') &= \cos(\theta) \\ &= \frac{x^\top x'}{\|x\| \cdot \|x'\|} \\ &= \frac{x^\top x'}{\sqrt{x^\top x}\sqrt{x'^\top x'}}. \end{aligned}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Desirable Properties of a Clustering Algorithm

1. Scalability in terms of both time and space
2. Ability to deal with different data types
3. Minimal requirements for domain knowledge to determine algorithm parameters
   - Don't need to know how many objects there are or what those object categories will be.
4. Interpretability and usability are optional
   - Incorporation of user-specified constraints

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

General ideas

- Bottom-up clustering
  - pixels belong together because they are locally coherent
- Top-down clustering
  - pixels belong together because they lie on the same visual entity (object)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Clustering algorithms

- Agglomerative clustering
- K-means
- Mean-shift clustering

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

Reading: Forsyth Chapter 9

D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, TPAMI 2002

Co-financed by the European Union
Connecting Europe Facility

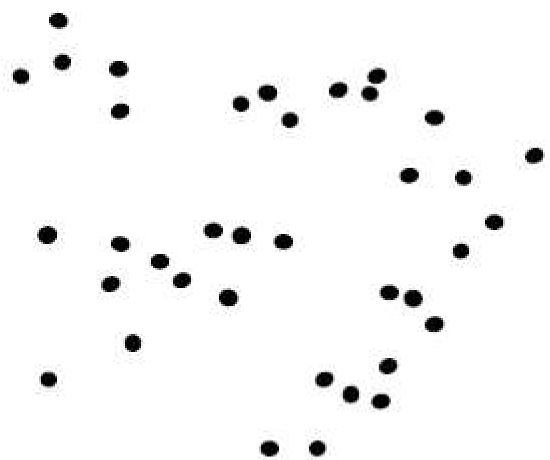
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Agglomerative Hierarchical Clustering - Algorithm

1. Initially each item $x_1, \ldots, x_n$ is in its own cluster $C_1, \ldots, C_n$.
2. Repeat until there is only one cluster left:
3. Merge the nearest clusters, say C_i and C_j.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Agglomerative Hierarchical Clustering - Algorithm



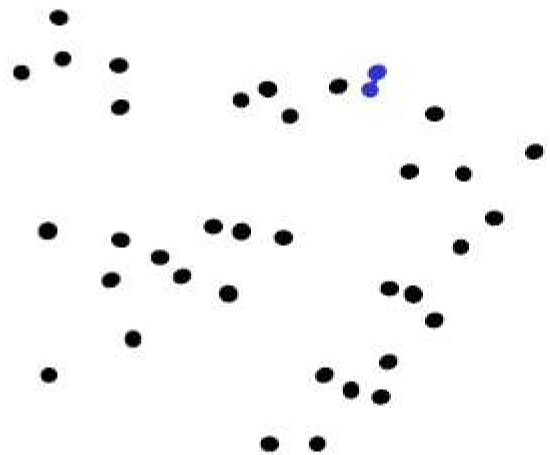
1. Say "Every point is its own cluster"

Slide credit: Andrew Moore

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Agglomerative Hierarchical Clustering - Algorithm



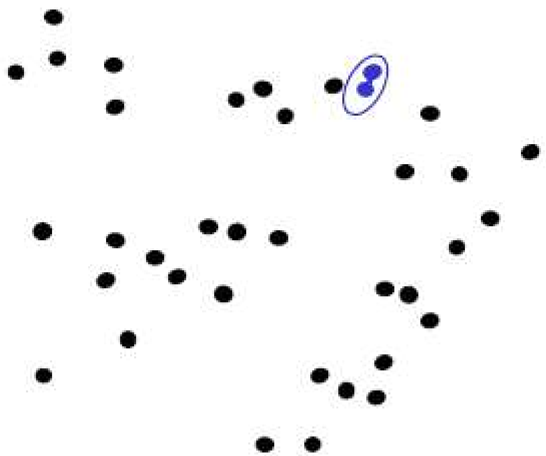
1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters

Slide credit: Andrew Moore

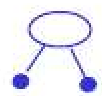
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Agglomerative Hierarchical Clustering - Algorithm



1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster

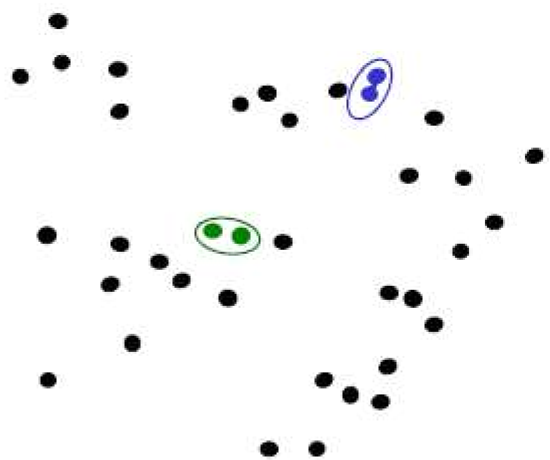


Slide credit: Andrew Moore

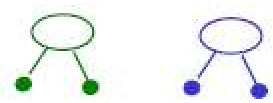
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Agglomerative Hierarchical Clustering - Algorithm



1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat

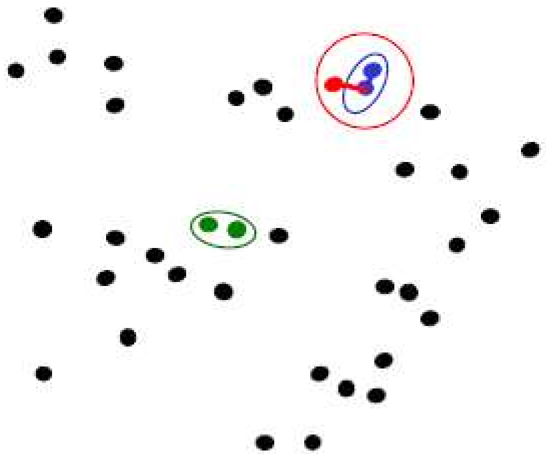


Slide credit: Andrew Moore

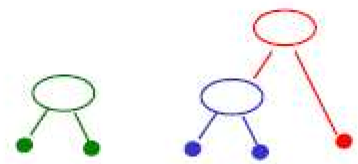
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Agglomerative Hierarchical Clustering - Algorithm



1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat

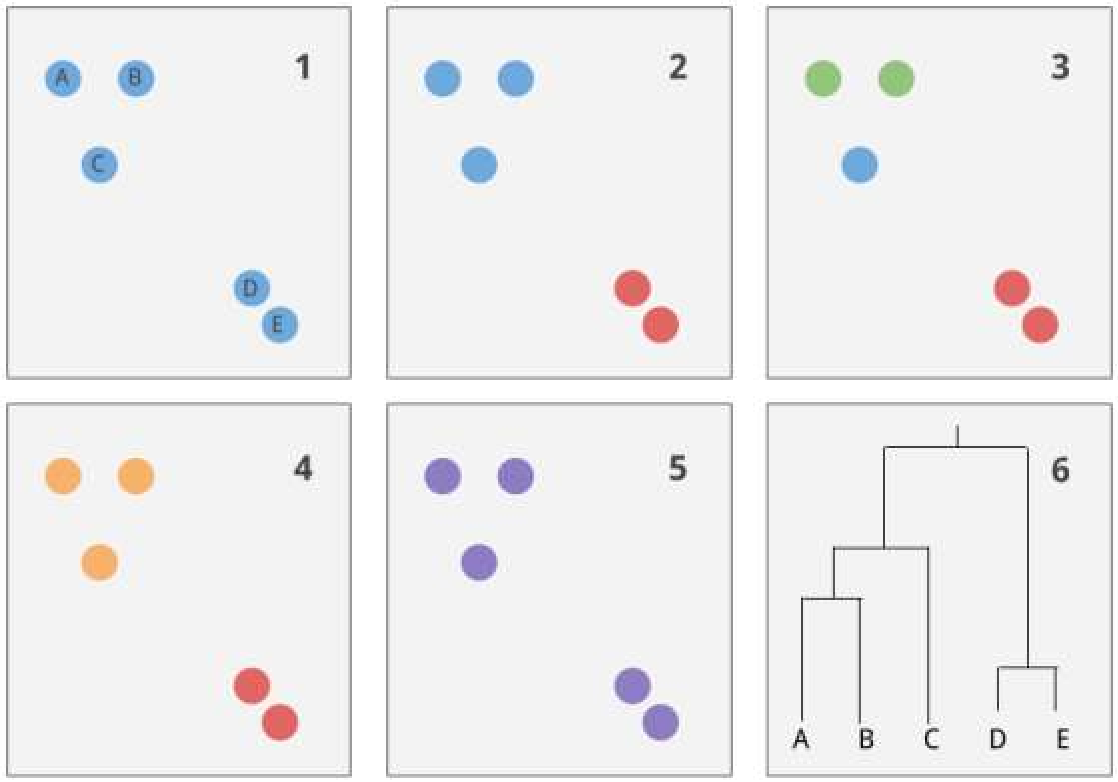


Slide credit: Andrew Moore

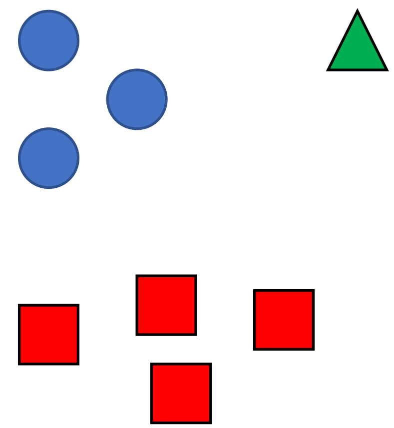
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Agglomerative clustering example



Agglomerative clustering

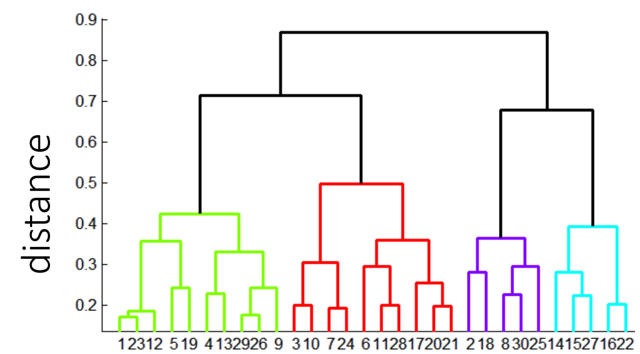


How to define cluster similarity?

- average distance between points
- maximum distance
- minimum distance
- distance between means or medoids

How many clusters?

- Clustering creates a dendrogram (a tree)
- Threshold based on max number of clusters or based on distance between merges

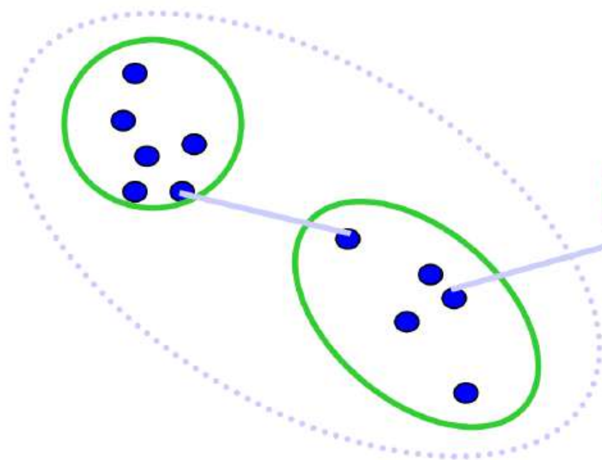


This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Different measures of nearest clusters

Single Link

- Distance between clusters is the minimum distance between their points



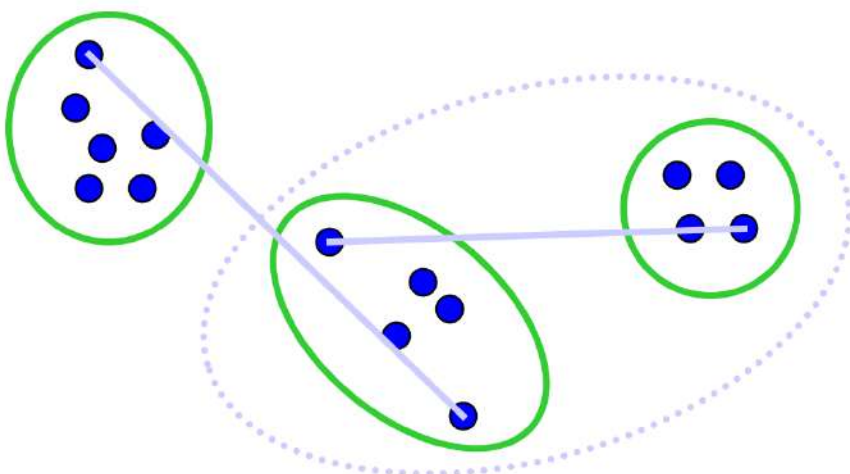
Long, skinny clusters

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Different measures of nearest clusters

Complete Link

- Distance between clusters is the maximum distance between their points



Tight clusters

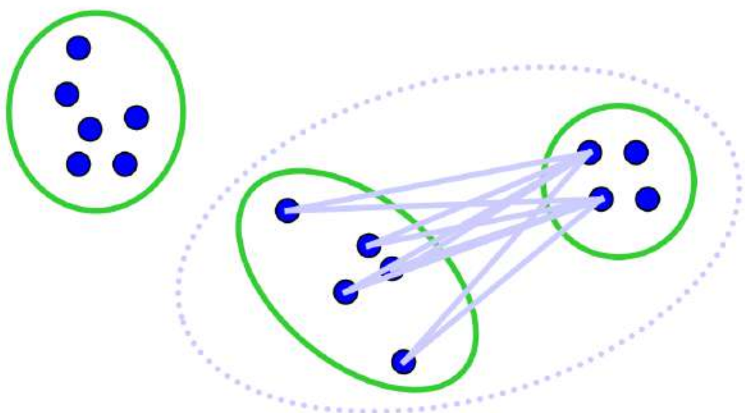
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Different measures of nearest clusters

Average Link

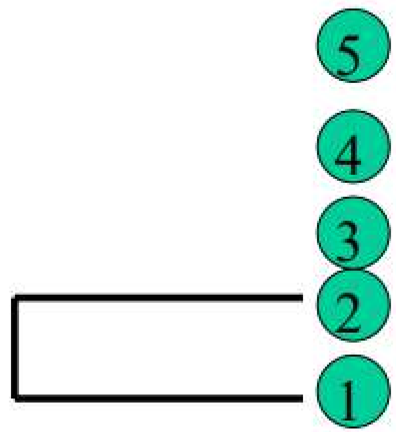
- Distance between clusters is the average distance between their points



Robust against noise

Example – single link

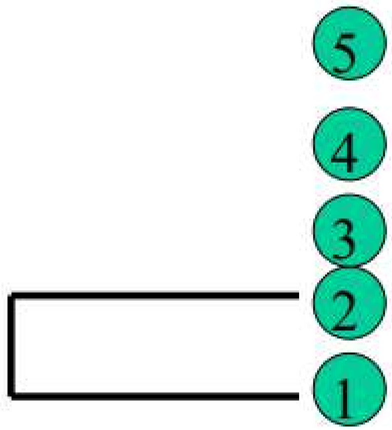
$$
\begin{array}{c c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\begin{bmatrix}
0 & & & & \\
2 & 0 & & & \\
6 & 3 & 0 & & \\
10 & 9 & 7 & 0 & \\
9 & 8 & 5 & 4 & 0
\end{bmatrix}
\end{array}
$$



Example – single link

$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0 & & & & \\
2 & 2 & 0 & & & \\
3 & 6 & 3 & 0 & & \\
4 & 10 & 9 & 7 & 0 & \\
5 & 9 & 8 & 5 & 4 & 0
\end{array}
\Rightarrow
\begin{array}{c|cccc}
 & (1,2) & 3 & 4 & 5 \\
\hline
(1,2) & 0 & & & \\
3 & 3 & 0 & & \\
4 & 9 & 7 & 0 & \\
5 & 8 & 5 & 4 & 0
\end{array}
$$

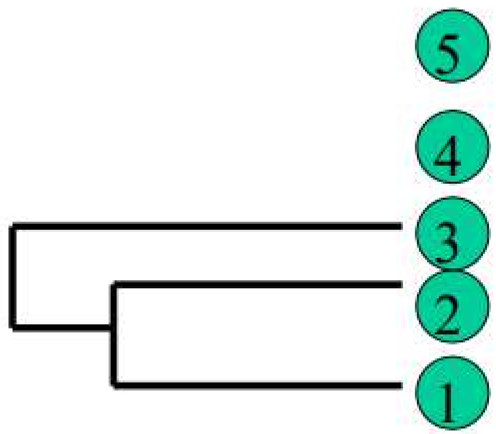
$$d_{(1,2),3} = \min\{d_{1,3}, d_{2,3}\} = \min\{6,3\} = 3$$
$$d_{(1,2),4} = \min\{d_{1,4}, d_{2,4}\} = \min\{10,9\} = 9$$
$$d_{(1,2),5} = \min\{d_{1,5}, d_{2,5}\} = \min\{9,8\} = 8$$



Example – single link

$$\begin{array}{c|ccccc|} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 0 & & & & \\ 2 & 2 & 0 & & & \\ 3 & 6 & 3 & 0 & & \\ 4 & 10 & 9 & 7 & 0 & \\ 5 & 9 & 8 & 5 & 4 & 0 \end{array} \Rightarrow \begin{array}{c|cccc|} & (1,2) & 3 & 4 & 5 \\ \hline (1,2) & 0 & & & \\ 3 & 3 & 0 & & \\ 4 & 9 & 7 & 0 & \\ 5 & 8 & 5 & 4 & 0 \end{array} \Rightarrow \begin{array}{c|ccc|} & (1,2,3) & 4 & 5 \\ \hline (1,2,3) & 0 & & \\ 4 & 7 & 0 & \\ 5 & 5 & 4 & 0 \end{array}$$

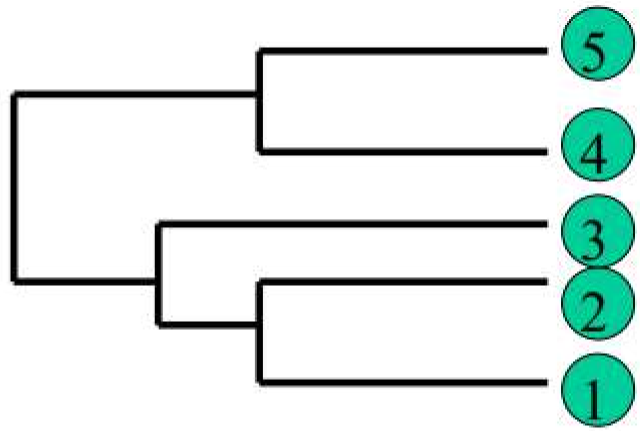
$$d_{(1,2,3),4} = \min\{d_{(1,2),4}, d_{3,4}\} = \min\{9,7\} = 7$$
$$d_{(1,2,3),5} = \min\{d_{(1,2),5}, d_{3,5}\} = \min\{8,5\} = 5$$



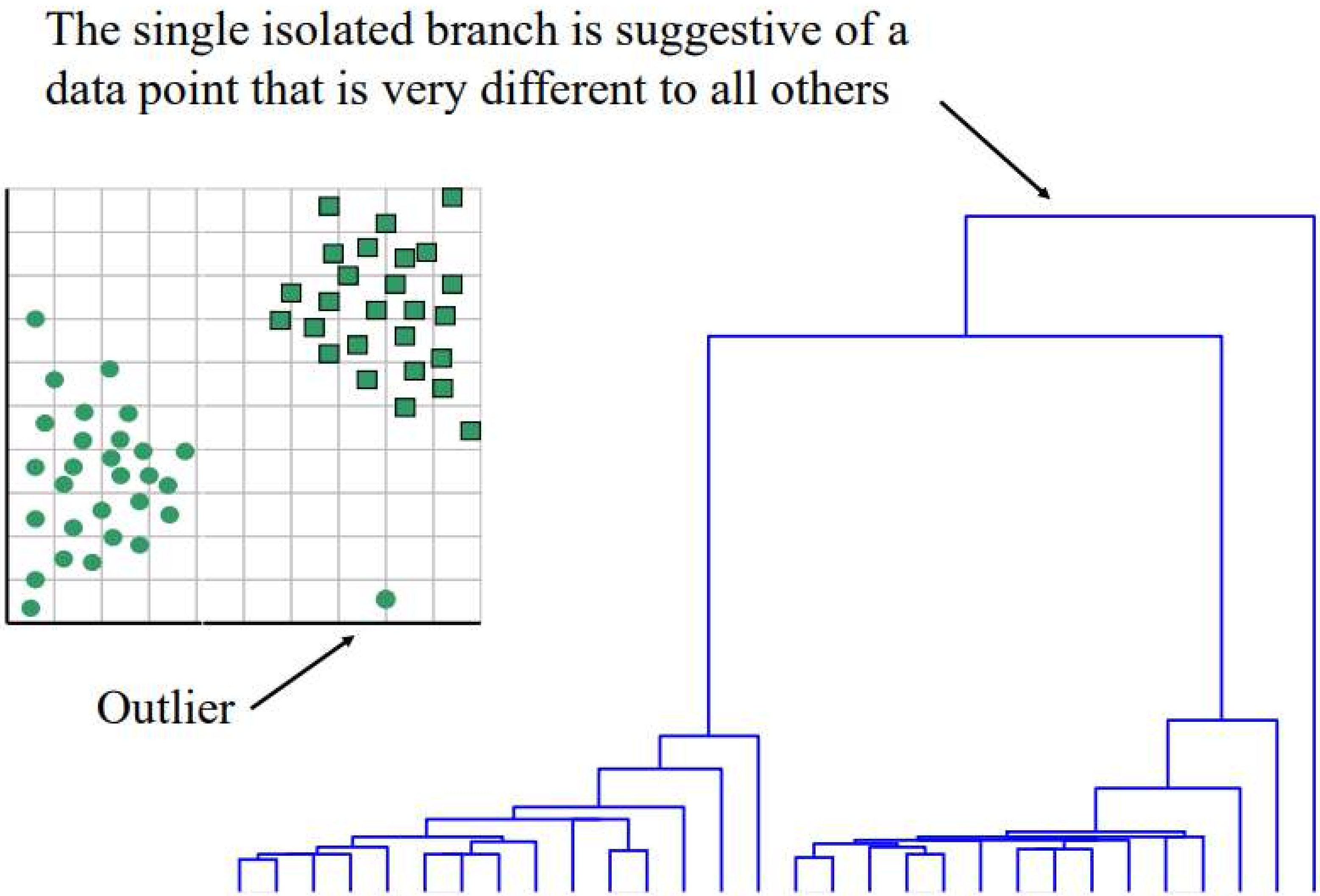
Example – single link

$$
\begin{array}{cc} & \begin{matrix}1&2&3&4&5\end{matrix}\\ \begin{matrix}1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}0&&&&\\2&0&&&\\6&3&0&&\\10&9&7&0&\\9&8&5&4&0\end{bmatrix}\end{array}
\Rightarrow
\begin{array}{cc} & \begin{matrix}(1,2)&3&4&5\end{matrix}\\ \begin{matrix}(1,2)\\3\\4\\5\end{matrix} & \begin{bmatrix}0&&&\\3&0&&\\9&7&0&\\8&5&4&0\end{bmatrix}\end{array}
\Rightarrow
\begin{array}{cc} & \begin{matrix}(1,2,3)&4&5\end{matrix}\\ \begin{matrix}(1,2,3)\\4\\5\end{matrix} & \begin{bmatrix}0&&\\7&0&\\5&4&0\end{bmatrix}\end{array}
$$

$$d_{(1,2,3),(4,5)} = \min\{d_{(1,2,3),4}, d_{(1,2,3),5}\} = 5$$



Outliers


The single isolated branch is suggestive of a data point that is very different to all others
Outlier

Conclusions: Agglomerative Clustering

Good

- Simple to implement, widespread application.
- Clusters have adaptive shapes.
- Provides a hierarchy of clusters.
- Can avoid specifying number of clusters in advance.

Bad

- May have imbalanced clusters.
- Still have to choose number of clusters or threshold to use them.
- Does not scale well. Runtime of $O(n^3)$.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

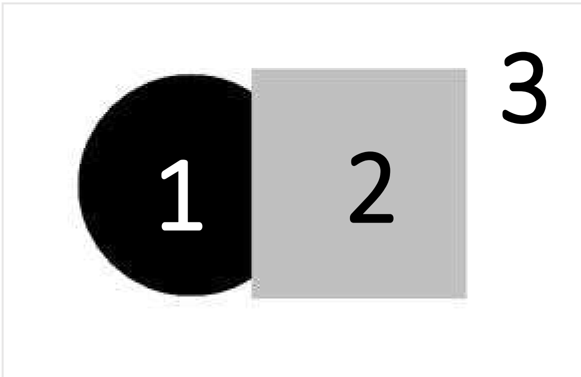
- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

Reading: Forsyth Chapter 9

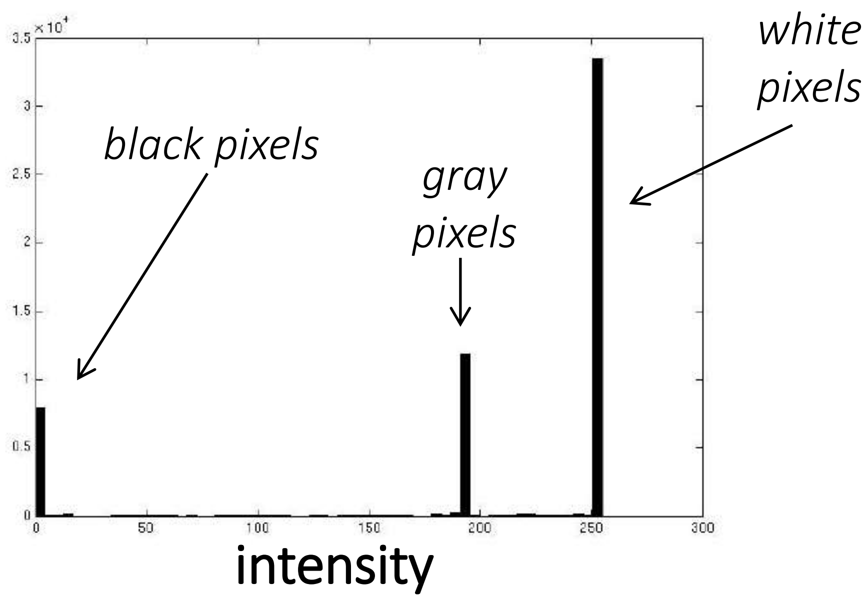
D. Comaniciu and P. Meer, [Mean Shift: A Robust Approach toward Feature Space Analysis](), TPAMI 2002

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image Segmentation: Toy Example

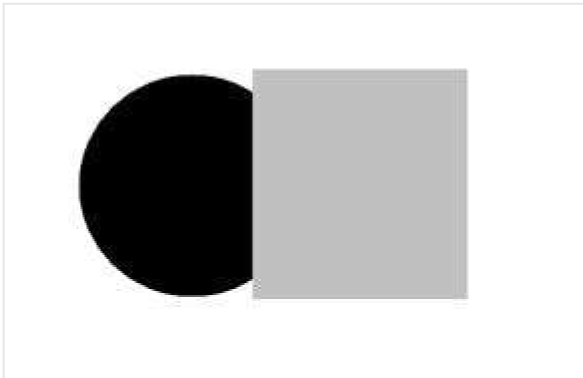


input image

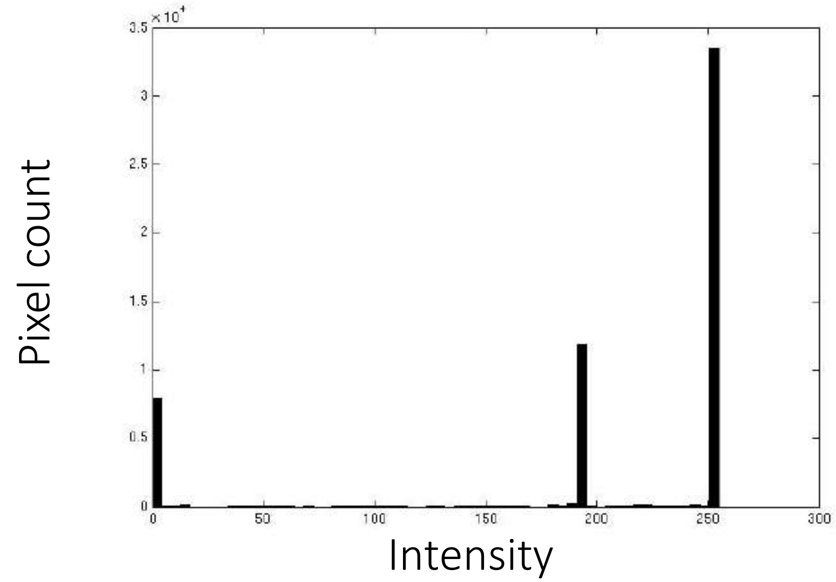


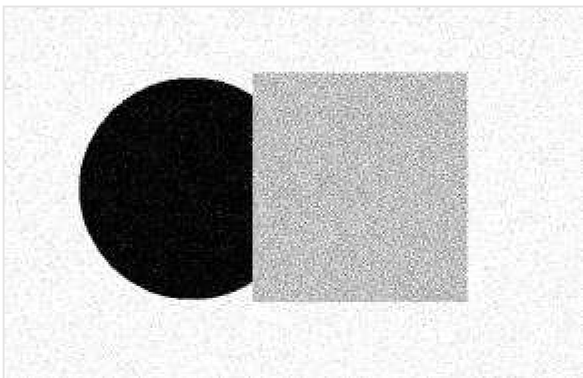
- These intensities define the three groups.
- We could label every pixel in the image according to which of these primary intensities it is.
  - i.e., segment the image based on the image intensity feature.
- What if the image isn't quite so simple?

Slide credit: Kristen Grauman

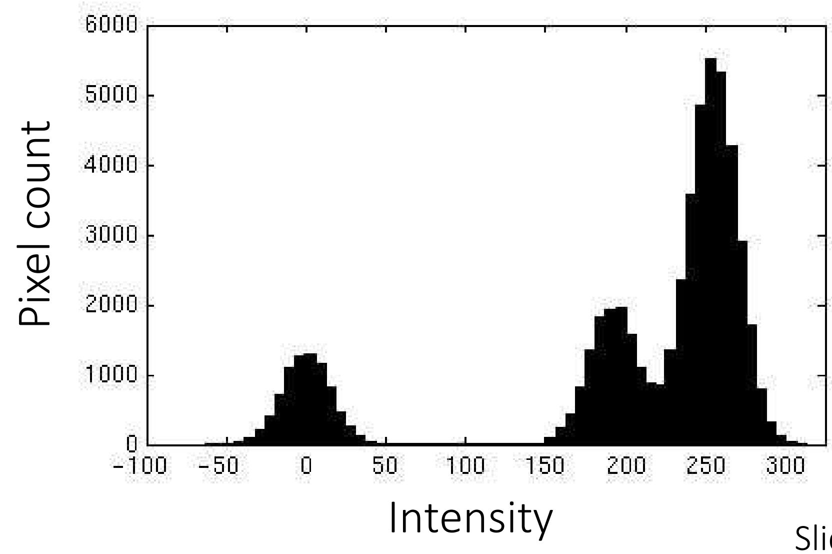


Input image



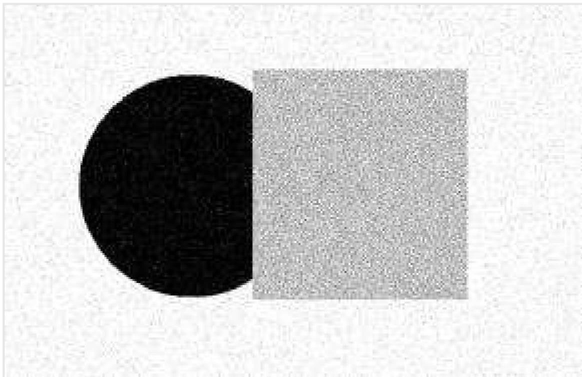


Input image

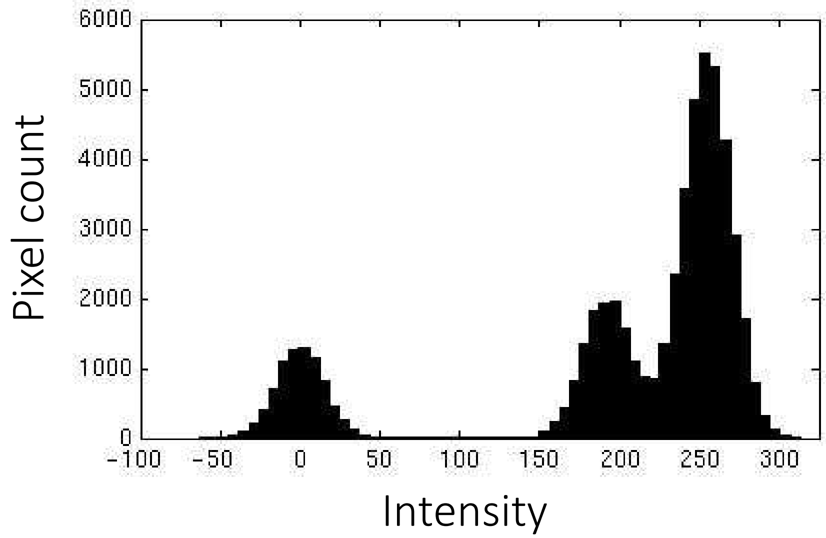


Slide credit: Kristen Grauman

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

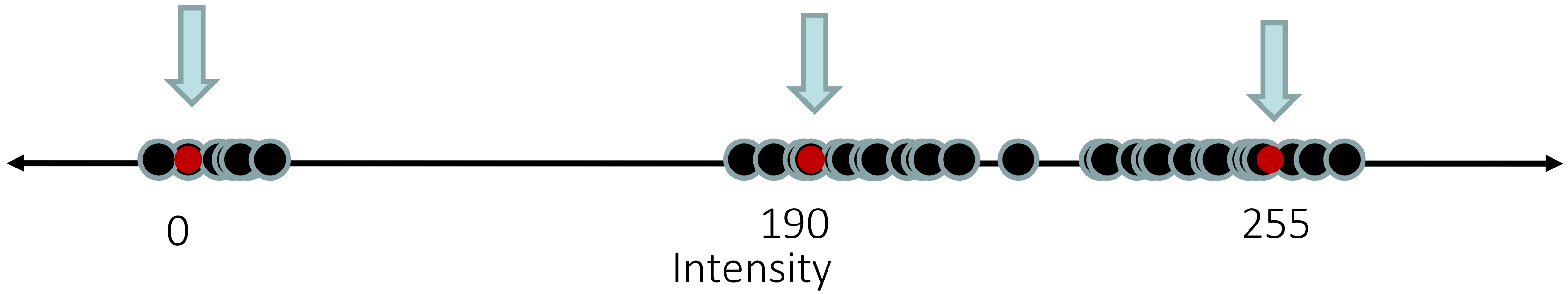


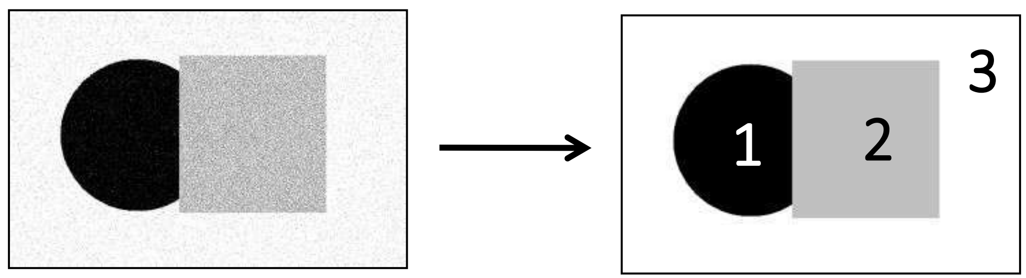
Input image



- Now how to determine the three main intensities that define our groups?
- We need to cluster

Slide credit: Kristen Grauman





- Goal: choose three “centers” as the representative intensities and label every pixel according to which of these centers it is nearest to.
- Best cluster centers are those that minimize Sum of Square Distance (SSD) between all points and their nearest cluster center c_i:

$$SSD = \sum_{cluster\, i} \sum_{x \in cluster\, i} (x - c_i)^2$$

Slide credit: Kristen Grauman

Clustering for Summarization

Goal: cluster to minimize variance in data, given clusters

Cluster center

Data

$$c^*, \delta^* = \arg\min_{c,\delta} \frac{1}{N} \sum_j^N \sum_i^K \delta_{ij} \left(c_i - x_j\right)^2$$

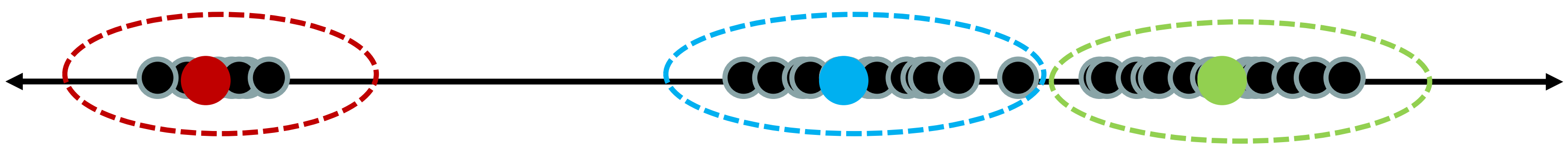
Whether x_j is assigned to c_i

Slide credit: Derek Hoiem

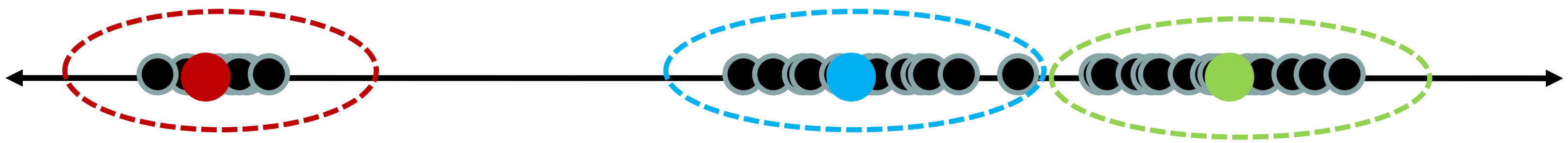
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Clustering

- With this objective, it is a “chicken and egg” problem:
  - If we knew the *cluster centers*, we could allocate points to groups by assigning each to its closest center.



  - If we knew the *group memberships*, we could get the centers by computing the mean per group.



Slide credit: Kristen Grauman

K-means clustering

1. Initialize ($t = 0$): cluster centers $c_1, \ldots, c_K$

2. Compute δ^t : assign each point to the closest center
   - δ^t denotes the set of assignments for each x_j to cluster c_i at iteration t

$$\delta^t = \underset{\delta}{\operatorname{argmin}} \frac{1}{N} \sum_{j}^{N} \sum_{i}^{K} \delta_{ij}^{t-1} \left(c_i^{t-1} - x_j\right)^2$$

3. Compute c^t: update cluster centers as the mean of the points

$$c^t = \underset{c}{\operatorname{argmin}} \frac{1}{N} \sum_{j}^{N} \sum_{i}^{K} \delta_{ij}^{t} \left(c_i^{t-1} - x_j\right)^2$$

4. Update $t = t + 1$, Repeat Step 2-3 till stopped

Slide credit: Derek Hoiem

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

K-means clustering

1. Initialize ($t=0$): cluster centers $c_1, \ldots, c_K$
   - Commonly used: random initialization
   - Or greedily choose K to minimize residual
2. Compute δ^t : assign each point to the closest center
   - δ^t denotes the set of assignments for each x_j to cluster c_i at iteration *t*
   - Typical distance measure:
     - Euclidean
     - Cosine

$$\delta^t = \underset{\delta}{\operatorname{argmin}} \frac{1}{N} \sum_{j}^{N} \sum_{i}^{K} \delta_{ij}^{t-1} \left(c_i^{t-1} - x_j\right)^2$$

3. Compute c^t: update cluster centers as the mean of the points

$$c^t = \underset{c}{\operatorname{argmin}} \frac{1}{N} \sum_{j}^{N} \sum_{i}^{K} \delta_{ij}^{t} \left(c_i^{t-1} - x_j\right)^2$$

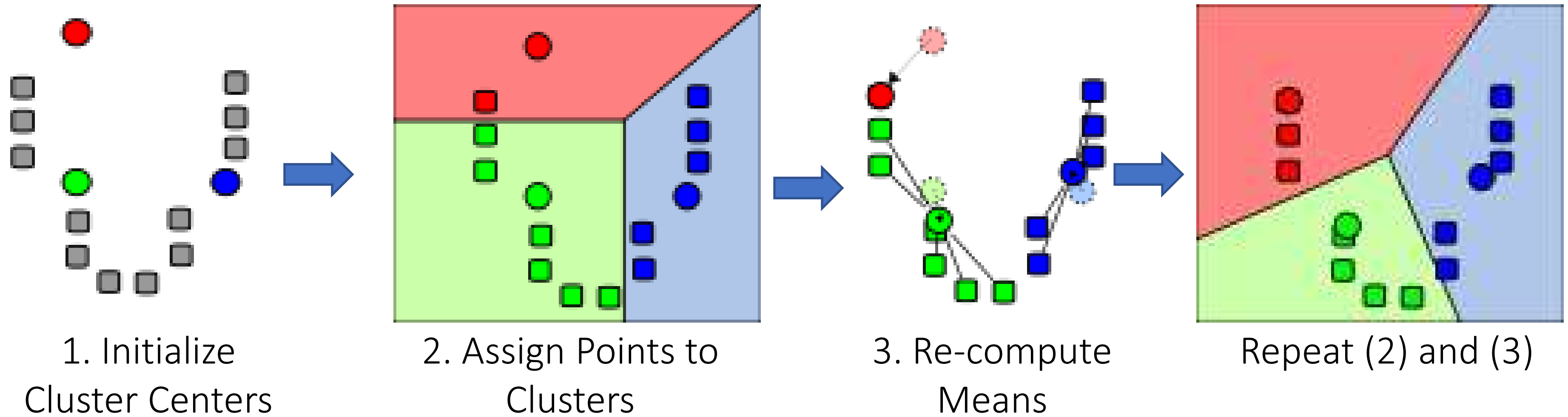
4. Update $t=t+1$, Repeat Step 2-3 till stopped
   - c^t doesn't change anymore.

Slide credit: Derek Hoiem

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

K-means clustering



Demo

https://stanford.edu/class/engr108/visualizations/kmeans/kmeans.html

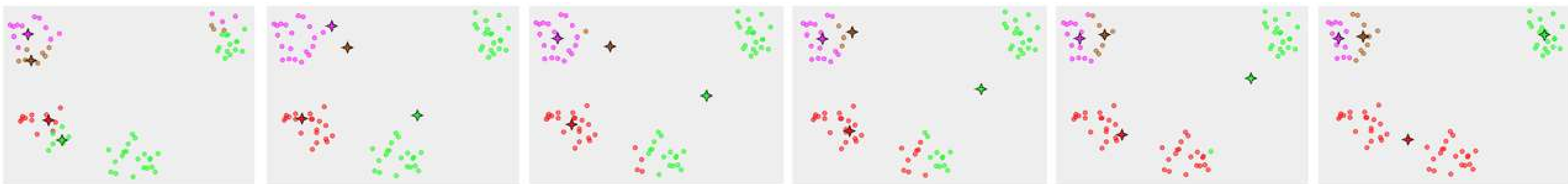
Illustration Source: wikipedia

Co-financed by the European Union
Connecting Europe Facility

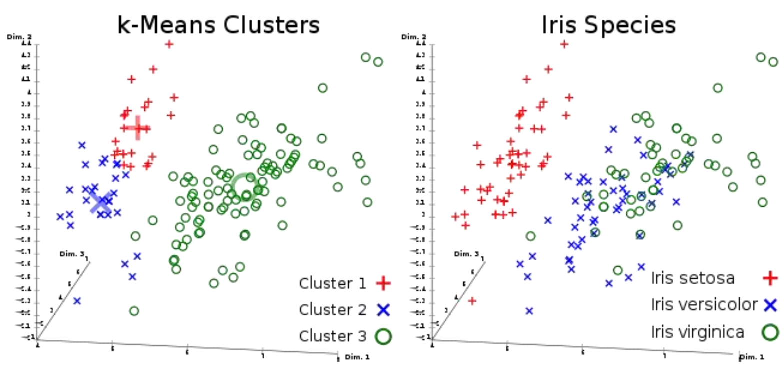
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

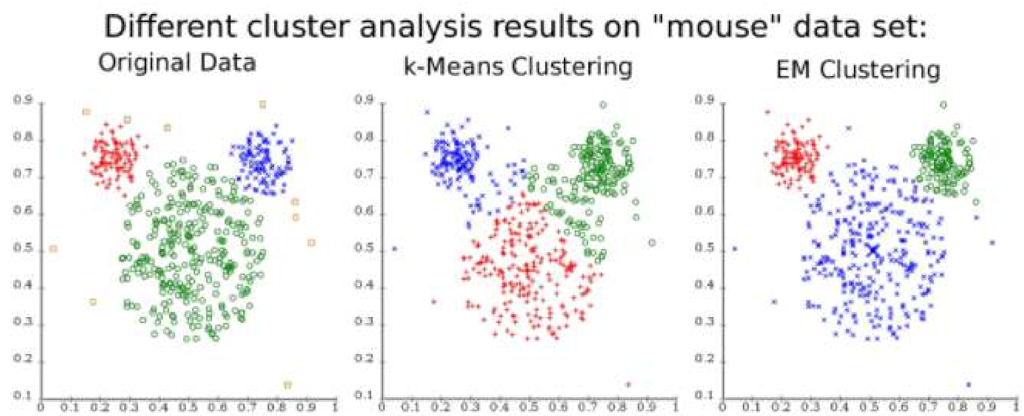
K-means clustering

- Converges to a *local minimum* solution
    - Initialize multiple runs



- Better fit for spherical data



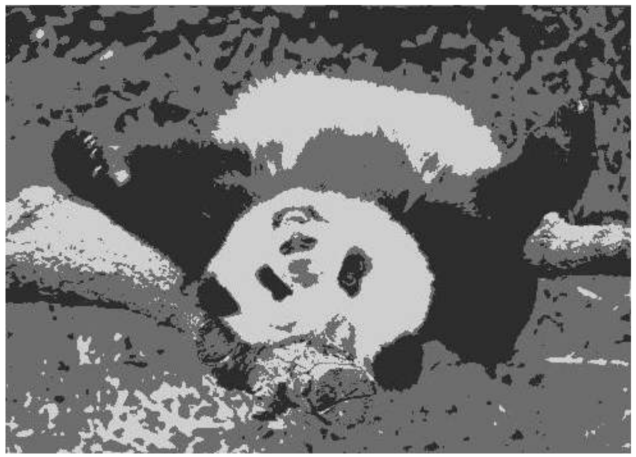


- Need to pick K (# of clusters)

Segmentation as Clustering

Original image

2 clusters



3 clusters

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Feature Space

- Depending on what we choose as the *feature space*, we can group pixels in different ways.

- Grouping pixels based on intensity similarity

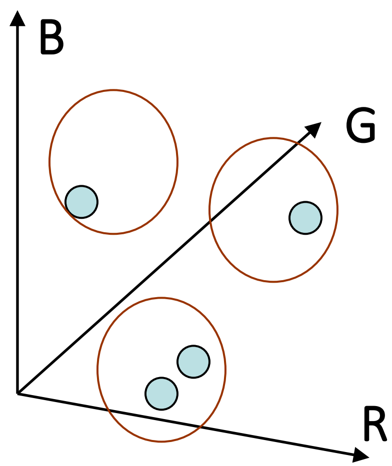
- Feature space: intensity value (1D)

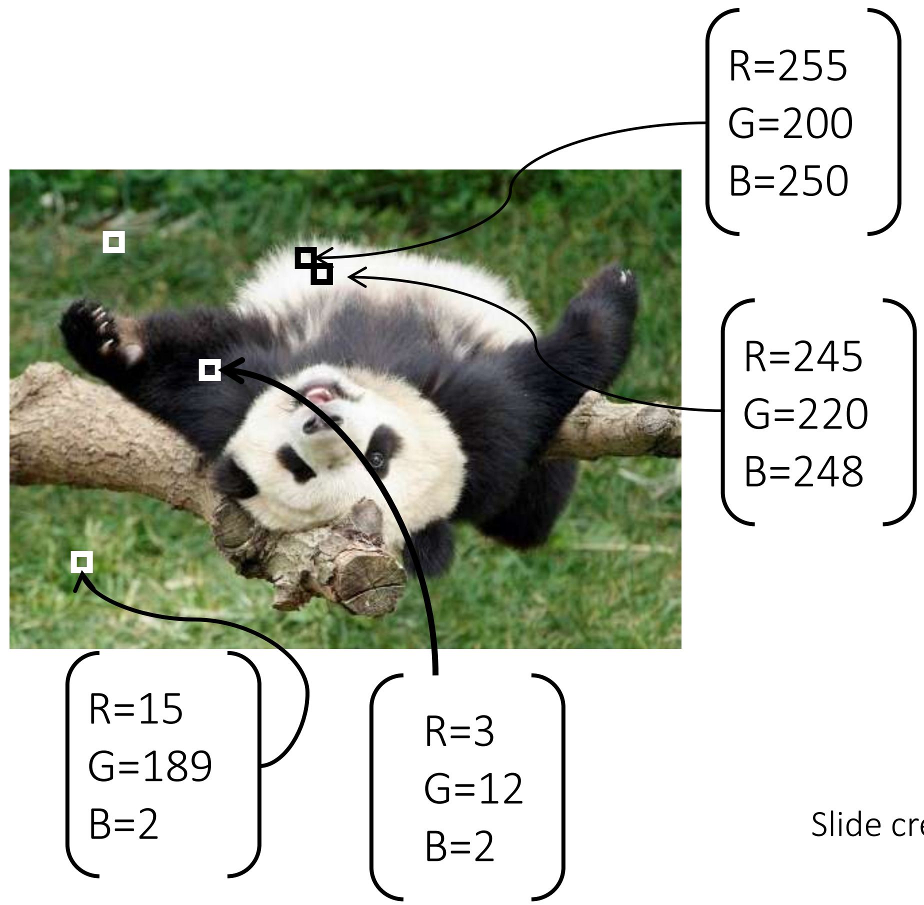
Slide credit: Kristen Grauman

Feature Space

- Depending on what we choose as the *feature space*, we can group pixels in different ways.

- Grouping pixels based on color similarity





- Feature space: color value (3D)

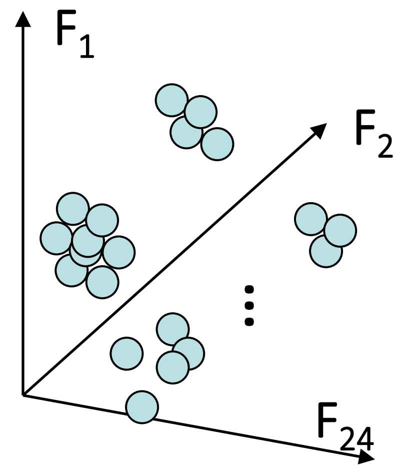
Slide credit: Kristen Grauman

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Feature Space

- Depending on what we choose as the *feature space*, we can group pixels in different ways.

- Grouping pixels based on texture similarity



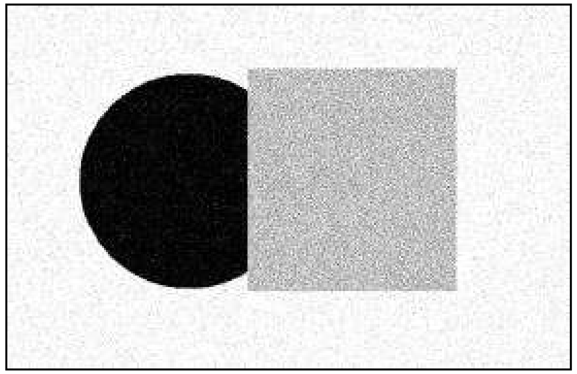
Filter bank of 24 filters

- Feature space: filter bank responses (e.g., 24D)

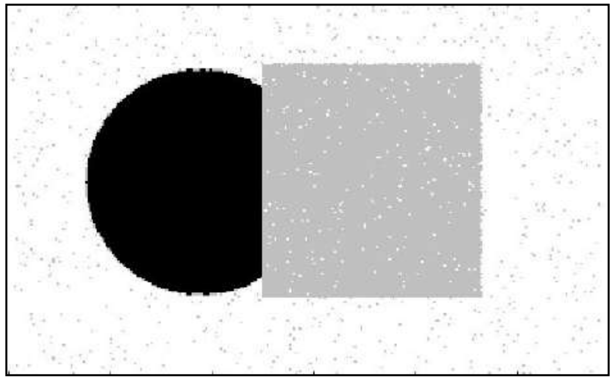
Slide credit: Kristen Grauman

Smoothing Out Cluster Assignments

- Assigning a cluster label per pixel may yield outliers:

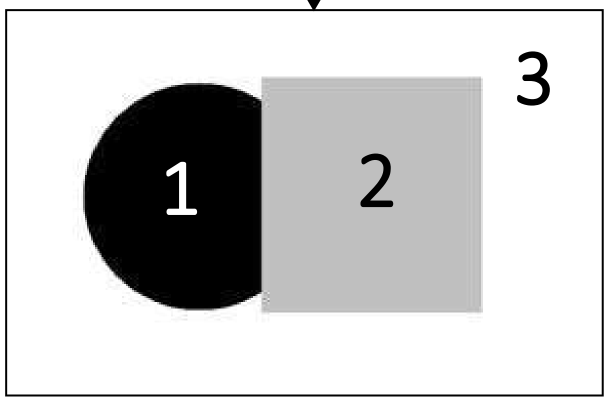


Original



Labeled by cluster center's intensity

?



- How can we ensure they are spatially smooth?

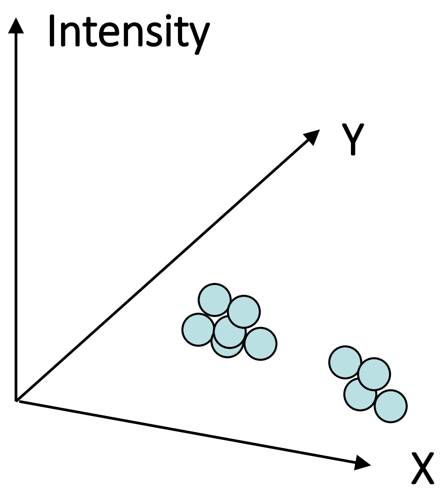
Slide credit: Kristen Grauman

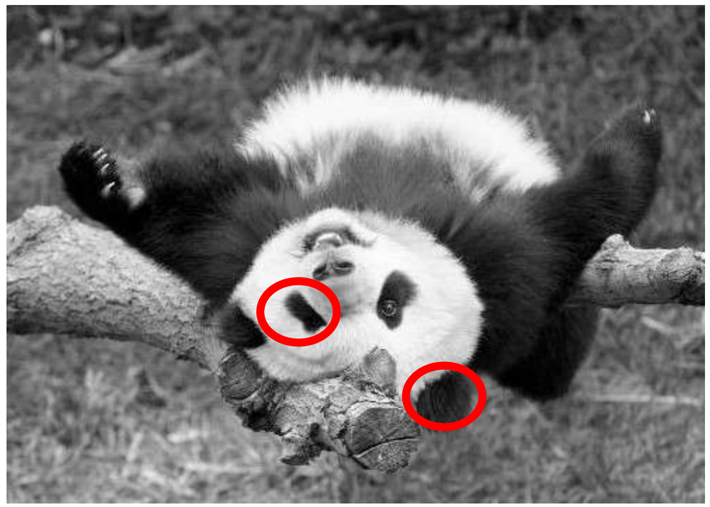
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Segmentation as Clustering

- Depending on what we choose as the *feature space*, we can group pixels in different ways.

- Grouping pixels based on *intensity+position* similarity





⇒ Way to encode both *similarity* and *proximity.*

Slide credit: Kristen Grauman

K-Means Clustering Results

- K-means clustering based on intensity or color is essentially vector quantization of the image attributes
  - Clusters don’t have to be spatially coherent

Image

Intensity-based clusters

Color-based clusters

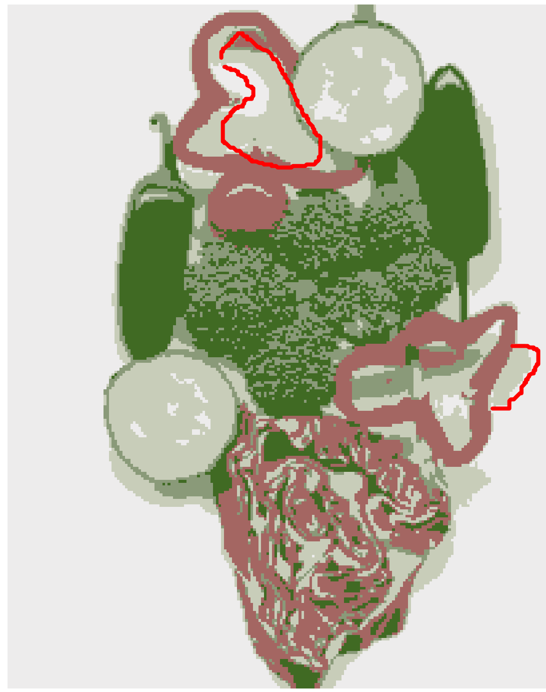


Image source: Forsyth & Ponce

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

K-Means Clustering Results

- K-means clustering based on intensity or color is essentially vector quantization of the image attributes
  - Clusters don't have to be spatially coherent
- Clustering based on (r,g,b,x,y) values enforces more spatial coherence

Co-financed by the European Union
Connecting Europe Facility

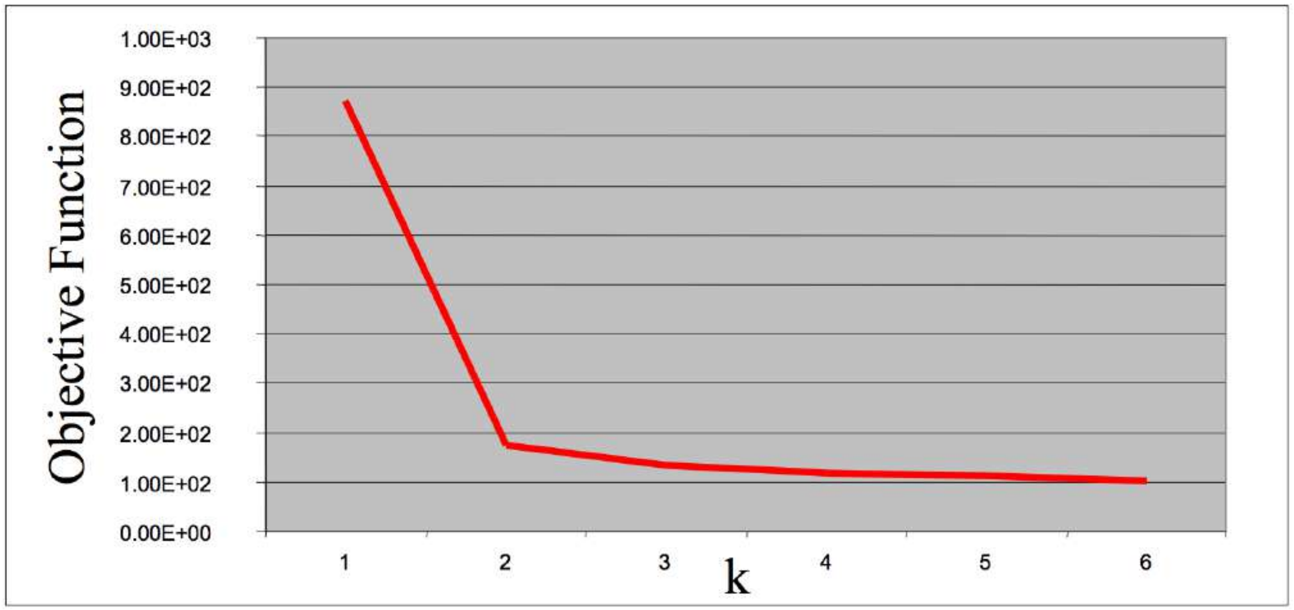
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How to choose the number of clusters?

Try different numbers of clusters in a validation set and look at performance.

We can plot the objective function values for k equals 1 to 6…

The abrupt change at k = 2, is highly suggestive of two clusters in the data. This technique for determining the number of clusters is known as "knee finding" or "elbow finding".

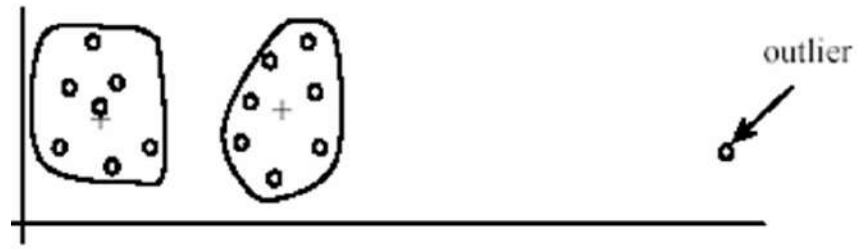


Slide credit: Derek Hoiem

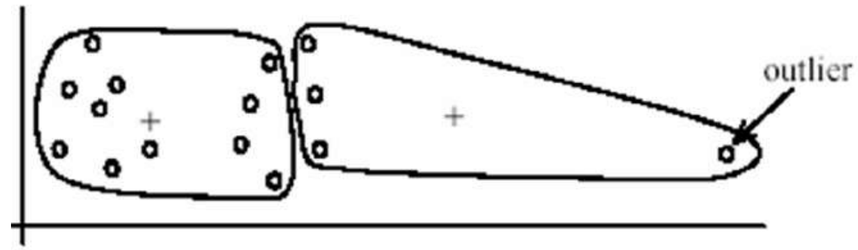
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

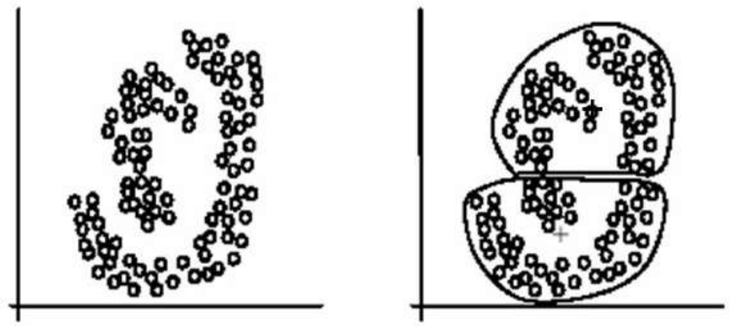
K-Means pros and cons

- Pros
  - Finds cluster centers that minimize conditional variance (good representation of data)
  - Simple and fast, Easy to implement
- Cons
  - Need to choose K
  - Sensitive to outliers
  - Prone to local minima
  - All clusters have the same parameters (e.g., distance measure is non-adaptive)
  - Distance computation in N-dimensional space could be slow



(B): Ideal clusters




(A): Two natural clusters (B): *k*-means clusters

Today's Agenda

- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

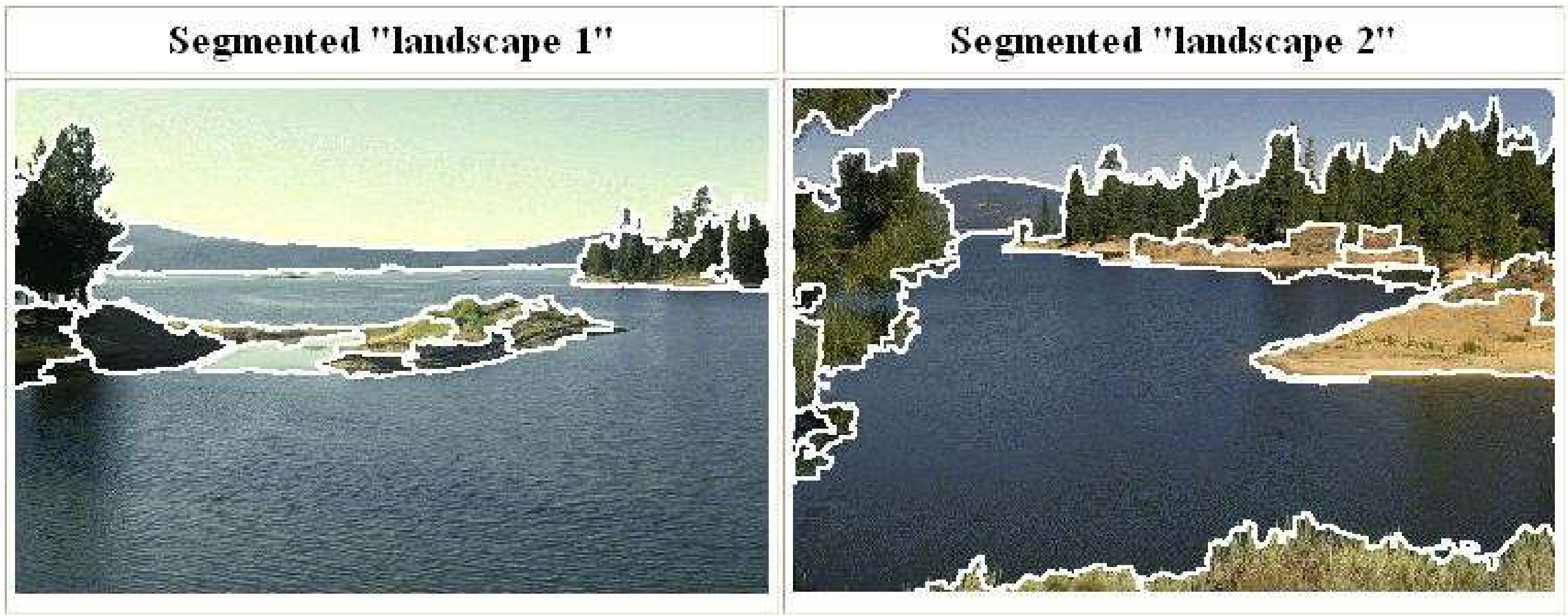
Reading: Forsyth Chapter 9

D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, TPAMI 2002

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mean-Shift Segmentation

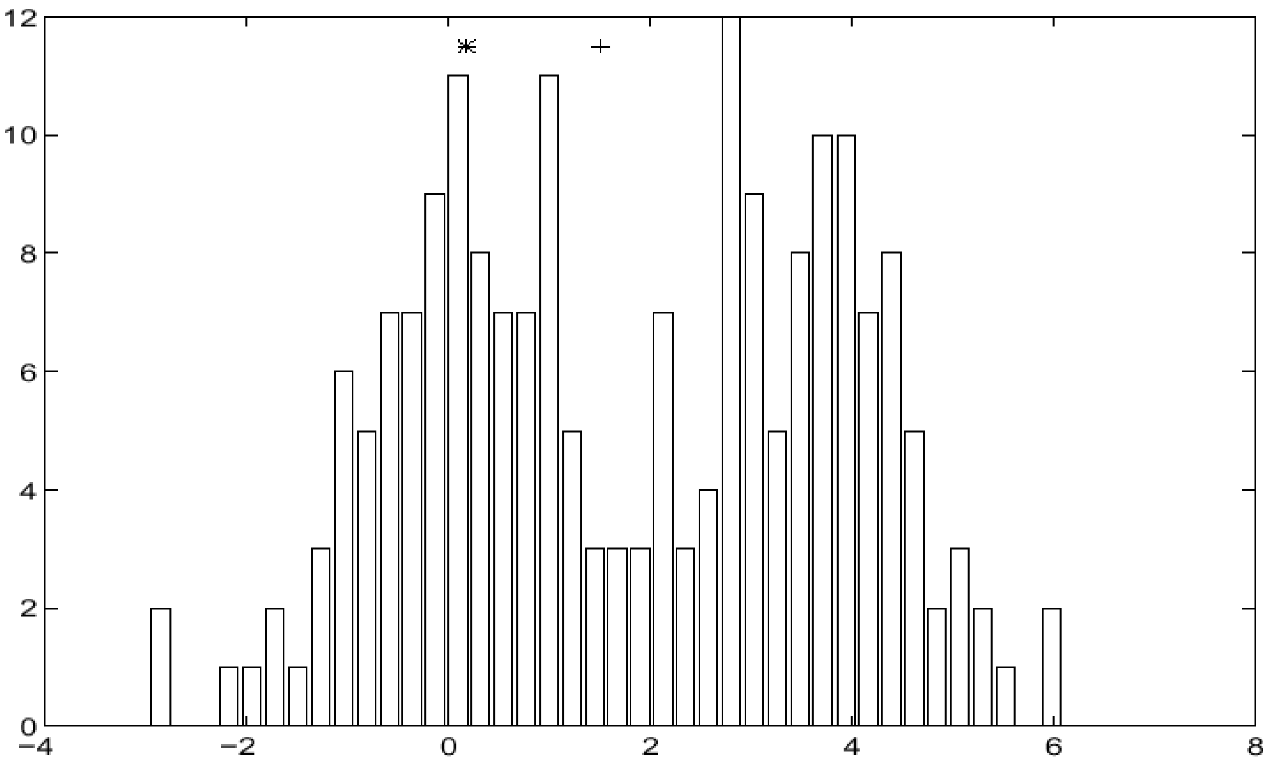
- An advanced and versatile technique for clustering-based segmentation



D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, TPAMI 2002

Slide credit: Svetlana Lazebnik

Mean-Shift Algorithm



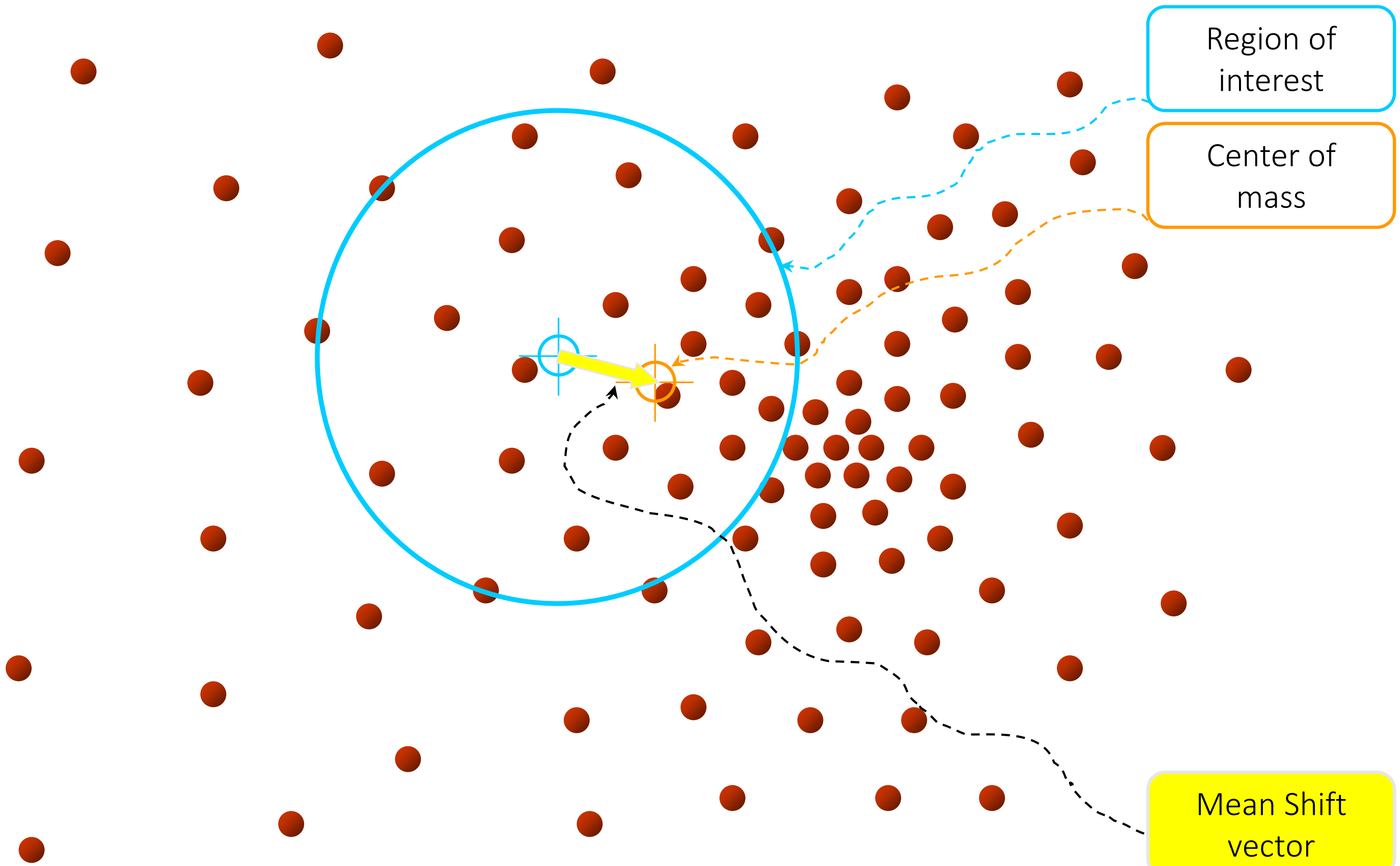
- Iterative Mode Search
  1. Initialize random seed, and window W
  2. Calculate center of gravity (the "mean") of W: $\sum_{x \in W} x H(x)$
  3. Shift the search window to the mean
  4. Repeat Step 2 until convergence

Slide credit: Steve Seitz

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

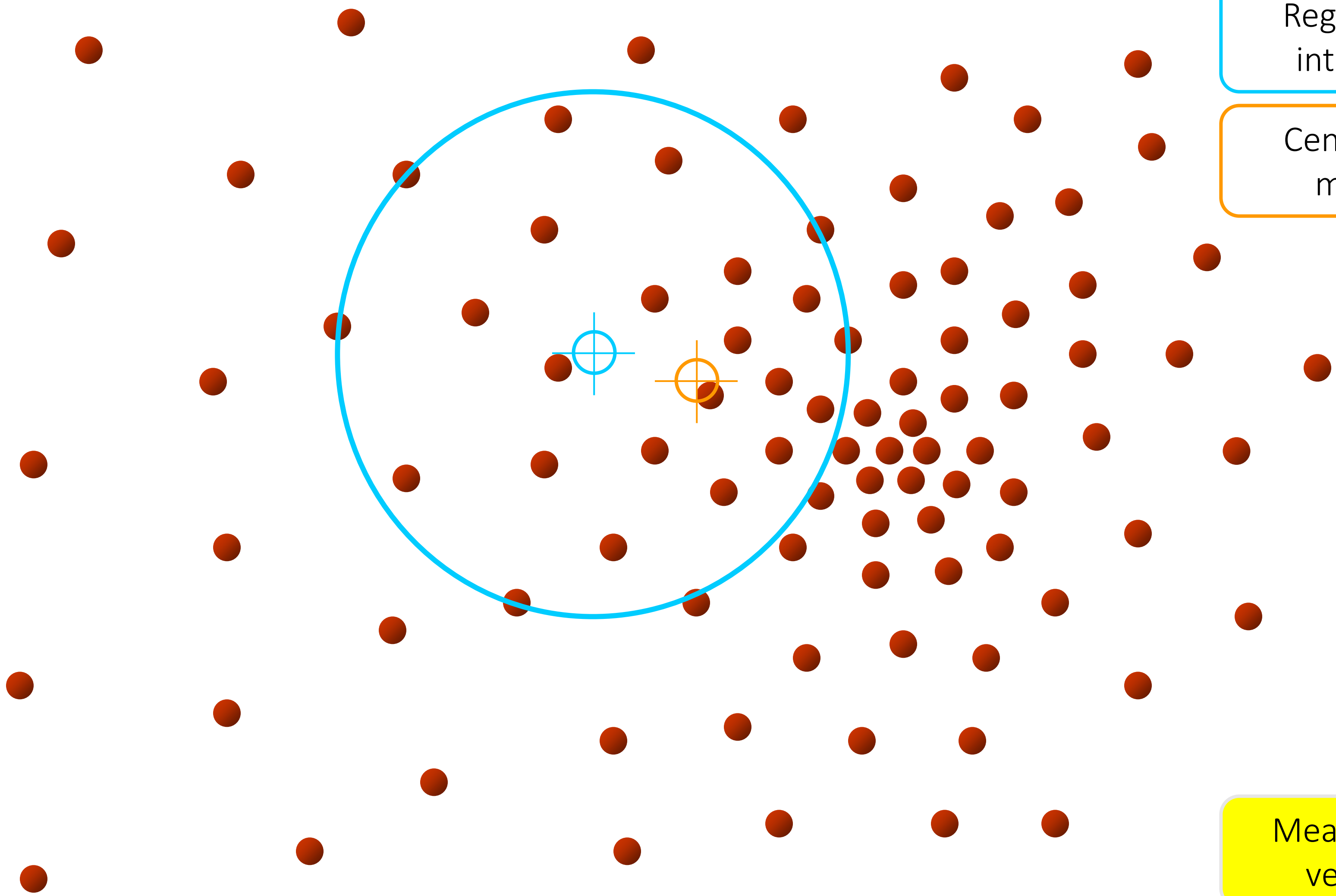
Mean-Shift



Slide credit: Y. Ukrainitz & B. Sarel

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mean-Shift



Region of interest

Center of mass

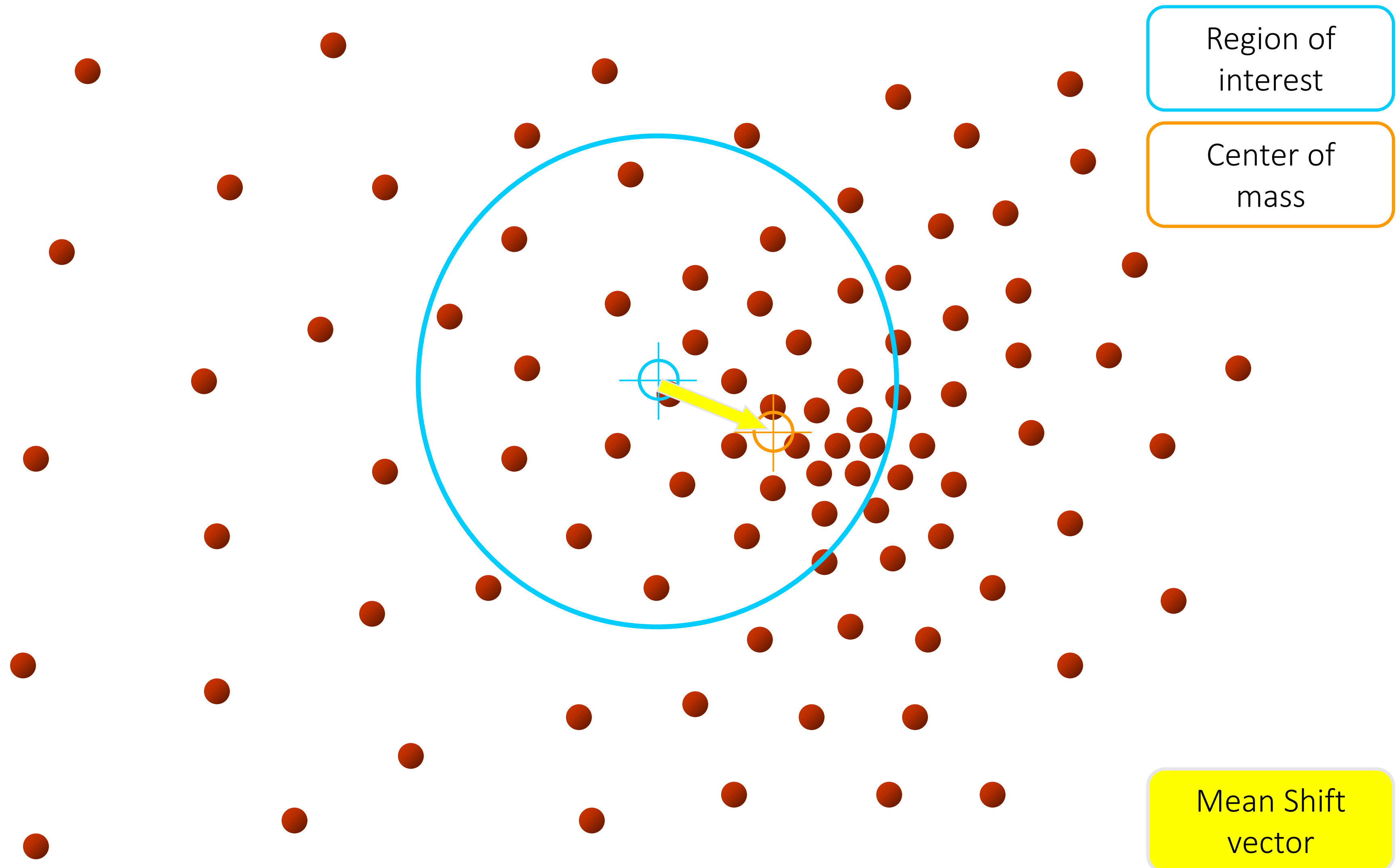
Mean Shift vector

Slide credit: Y. Ukrainitz & B. Sarel

Co-financed by the European Union
Connecting Europe Facility

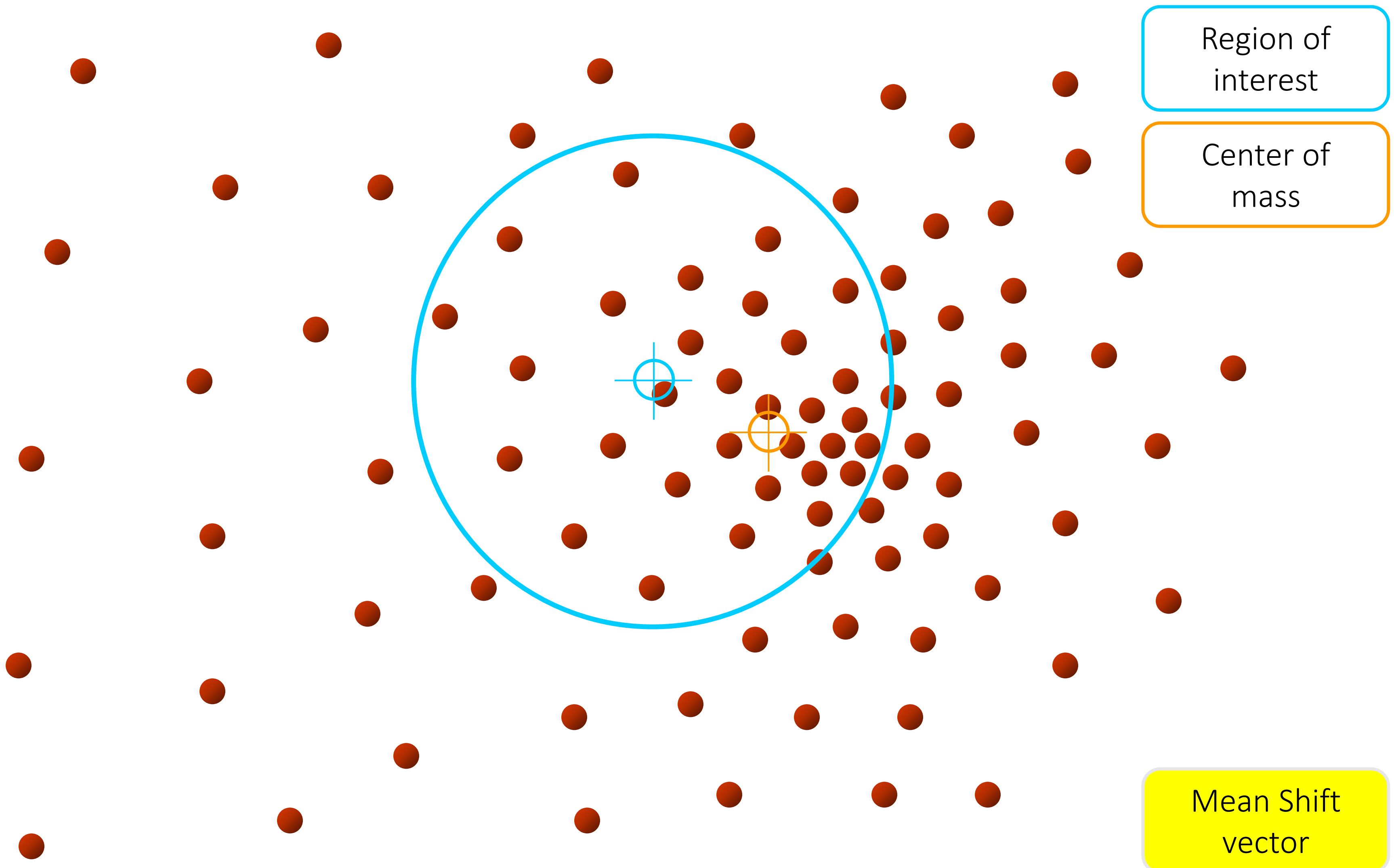
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mean-Shift



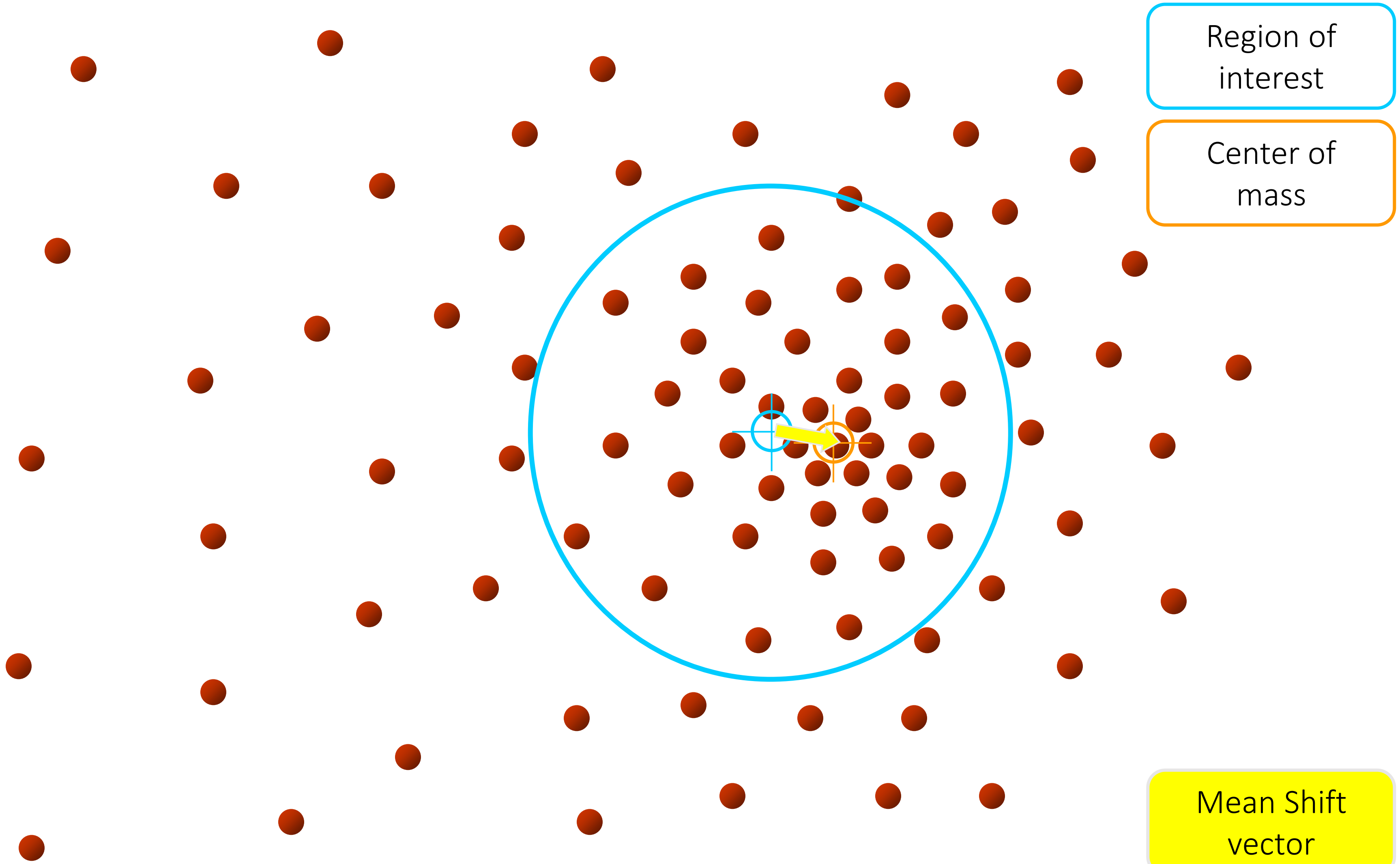
Slide credit: Y. Ukrainitz & B. Sarel

Mean-Shift



Slide credit: Y. Ukrainitz & B. Sarel

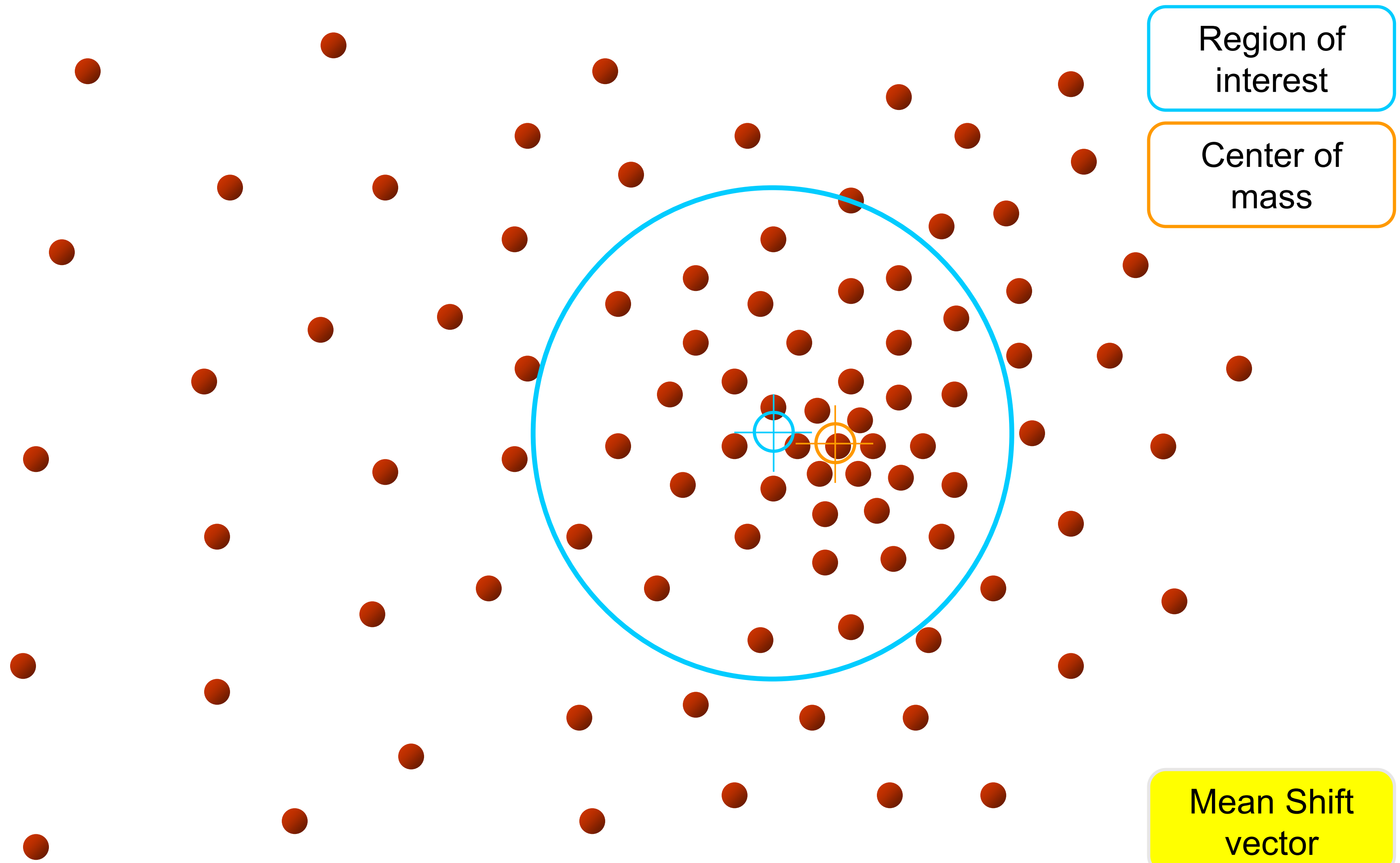
Mean-Shift



Slide credit: Y. Ukrainitz & B. Sarel

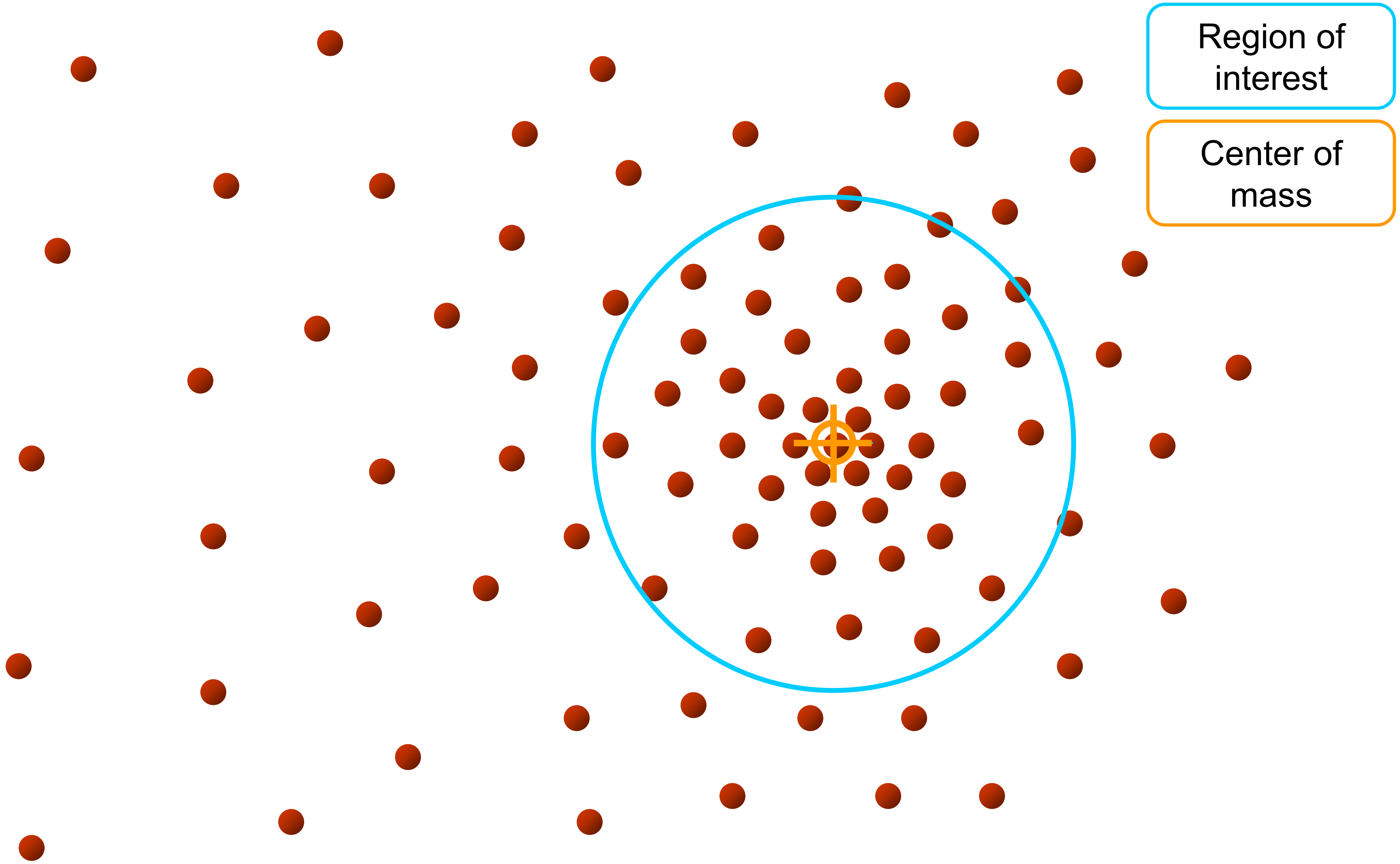
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mean-Shift



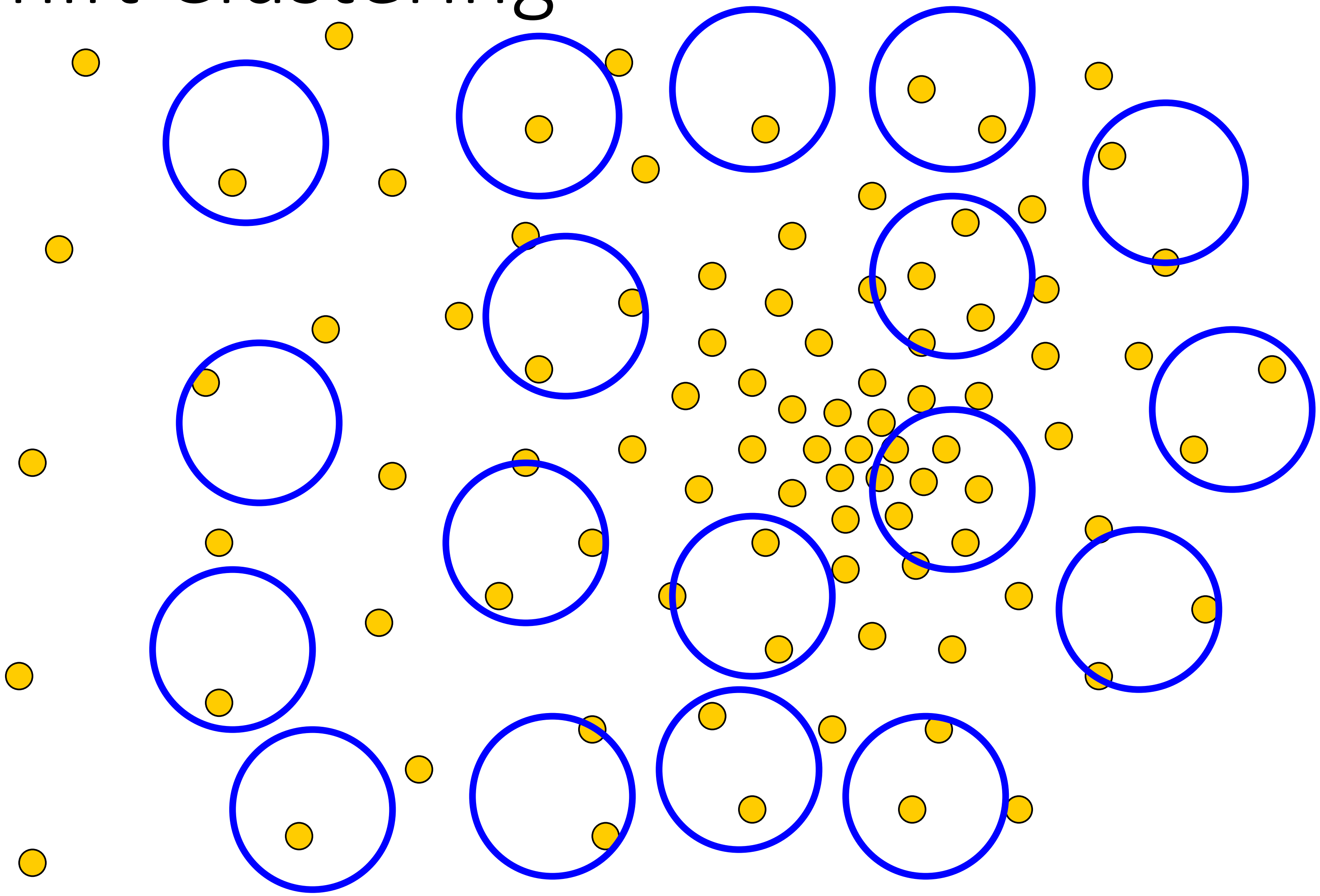
Slide credit: Y. Ukrainitz & B. Sarel

Mean-Shift



Slide credit: Y. Ukrainitz & B. Sarel

Mean-Shift Clustering



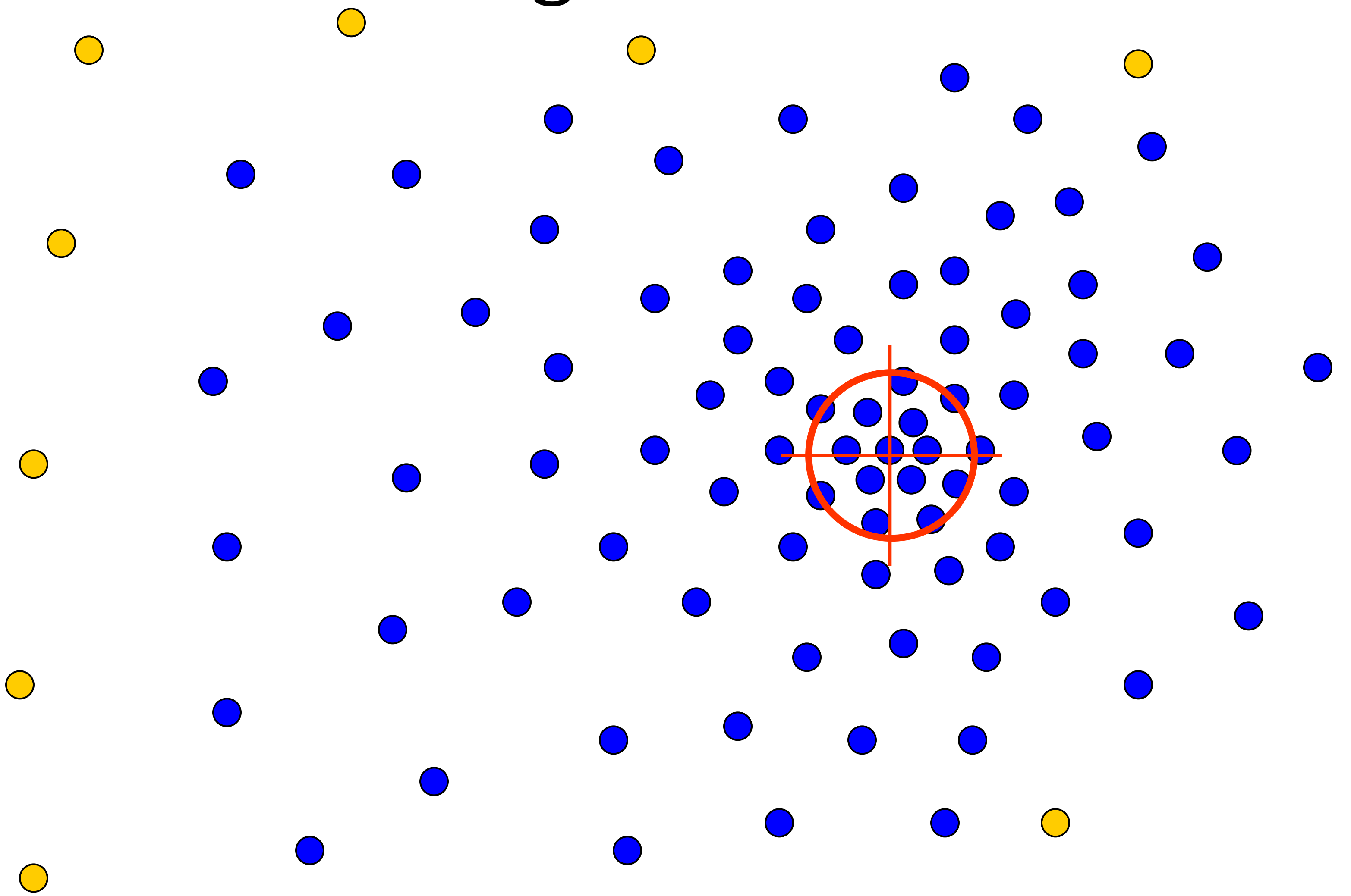
Initialize multiple windows in the space

Run the procedure in parallel

Slide credit: Y. Ukrainitz & B. Sarel

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mean-Shift Clustering



The blue data points were traversed by the windows towards the mode.

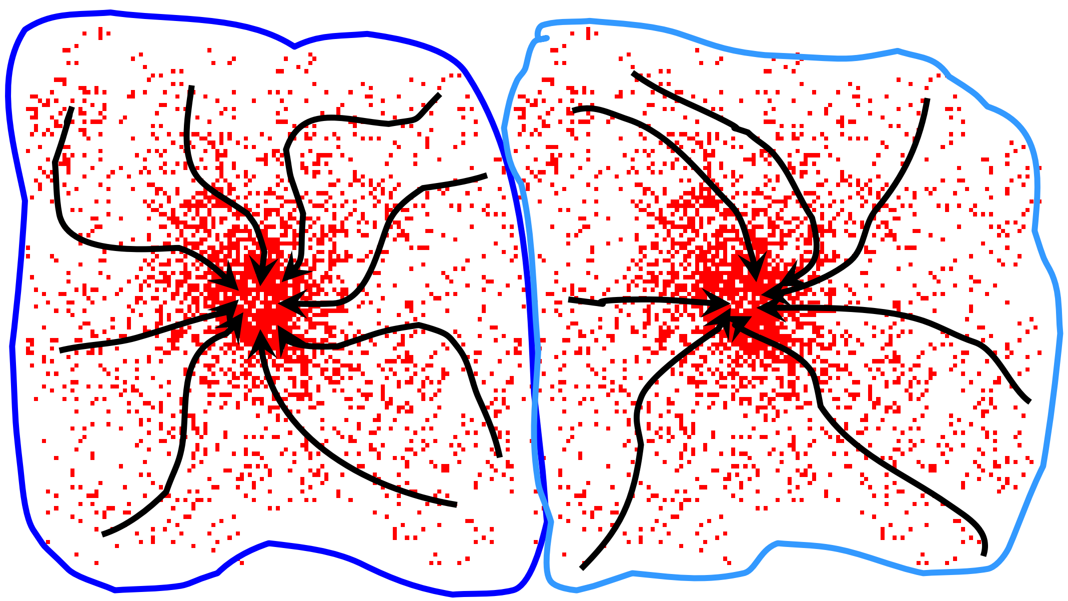
Slide credit: Y. Ukrainitz & B. Sarel

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mean-Shift Clustering

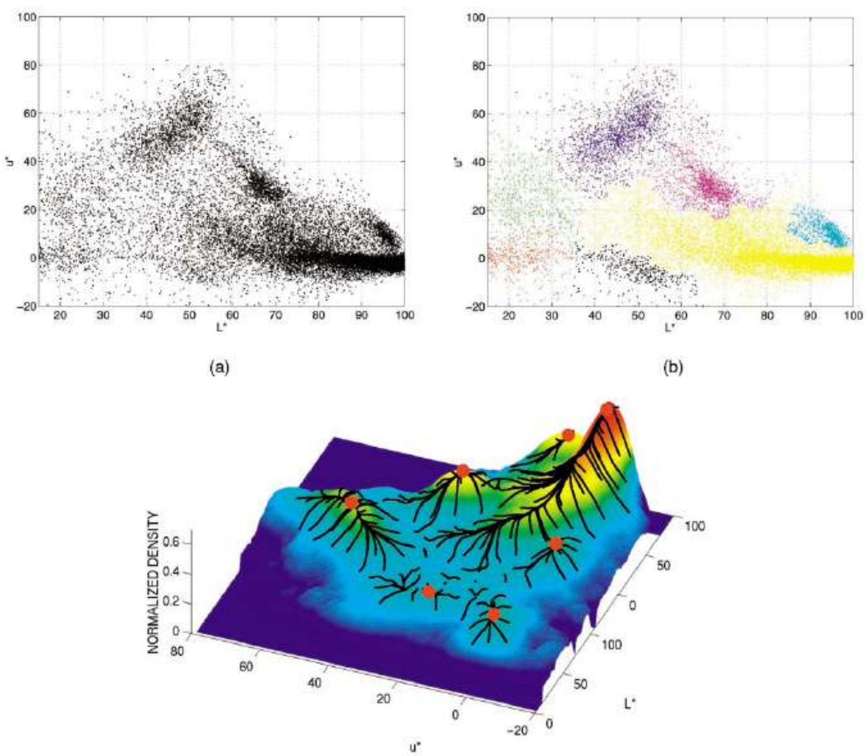
- Cluster: all data points in the attraction basin of a mode
- Attraction basin: the region for which all trajectories lead to the same mode



Slide credit: Y. Ukrainitz & B. Sarel

Mean-Shift Clustering/Segmentation

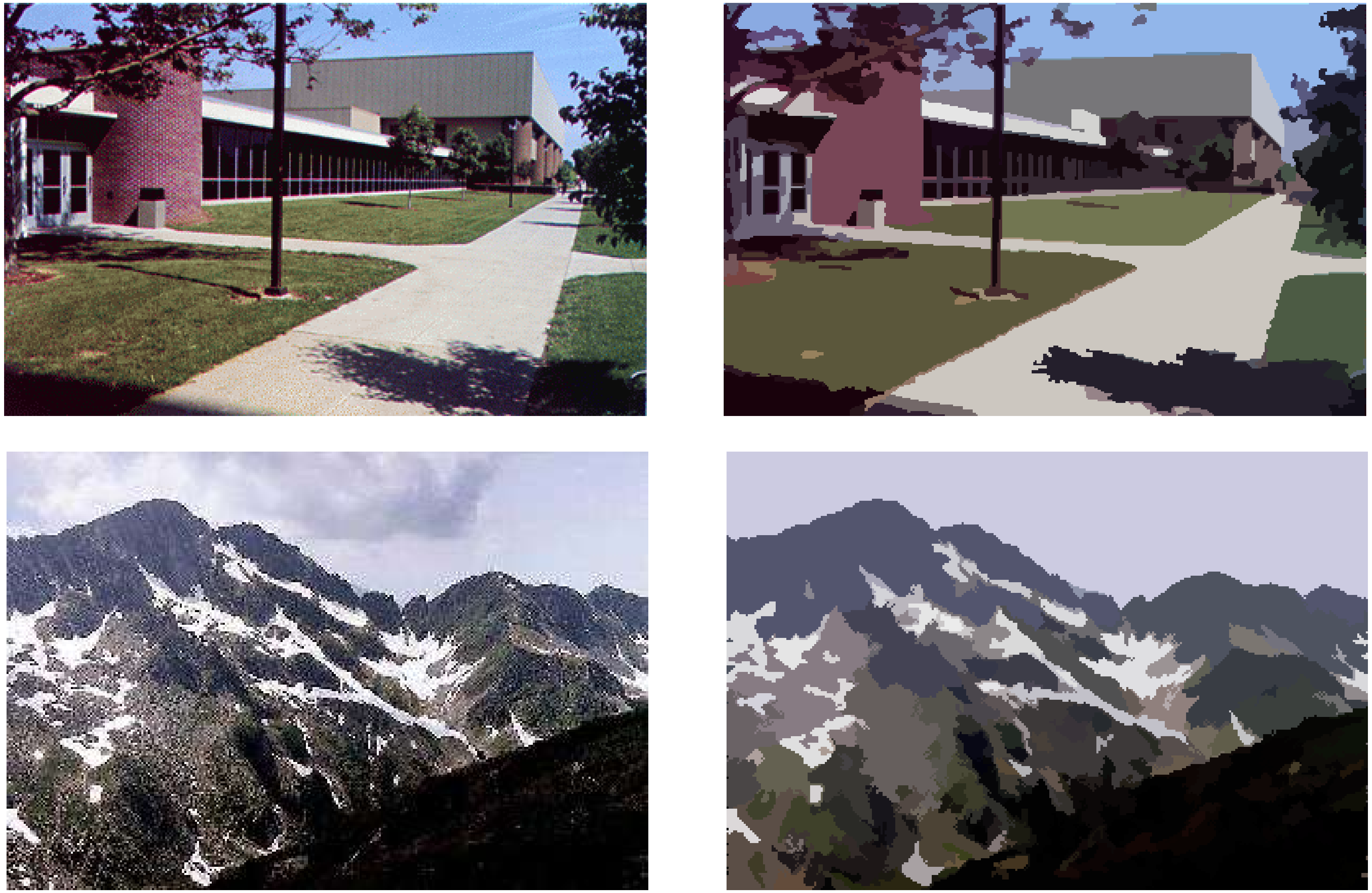
- Find features (color, gradients, texture, etc.)
- Initialize windows at individual pixel locations
- Perform mean shift for each window until convergence
- Merge windows that end up near the same “peak” or mode



Slide credit: Svetlana Lazebnik

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mean-Shift Segmentation Results



Slide credit: Svetlana Lazebnik

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

More Results




Slide credit: Svetlana Lazebnik

More Results

Mean-Shift pros and cons

- Pros
  - General, application-independent tool
  - Model-free, does not assume any prior shape (spherical, elliptical, etc.) on data clusters
  - Just a single parameter (window size h)
    - h has a physical meaning (unlike k-means)
  - Finds variable number of modes
  - Robust to outliers
- Cons
  - Output depends on window size
  - Window size selection is not trivial
  - Computationally (relatively) expensive (~2s/image)
  - Does not scale well with dimensionality of feature space

Slide credit: Svetlana Lazebnik

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

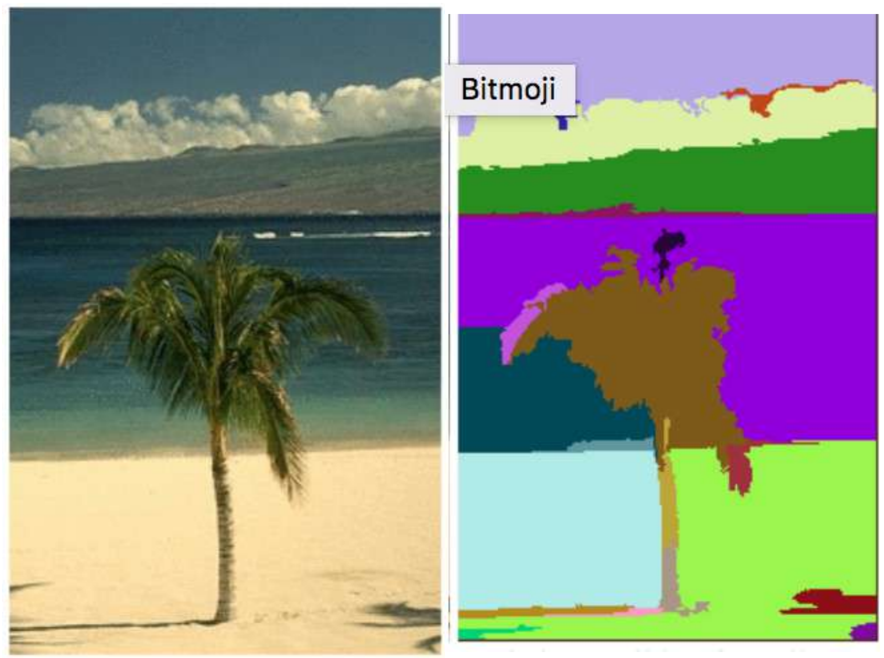
Reading: Forsyth Chapter 9

D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, TPAMI 2002

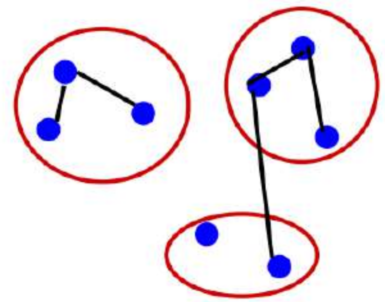
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Efficient Graph-based Image Segmentation

Oversegmentation algorithm introduced by *Felzenszwalb and Huttenlocher* in the paper titled *Efficient Graph-Based Image Segmentation*



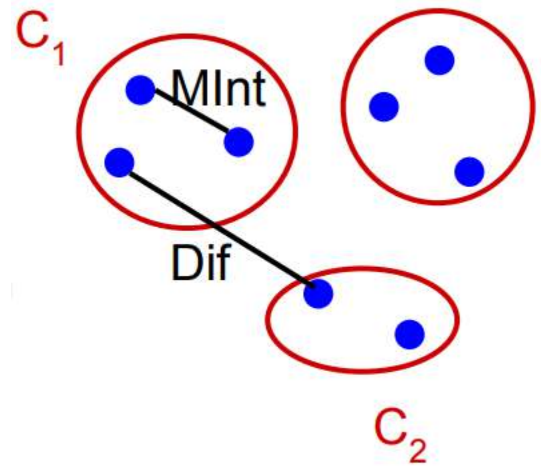
Problem Formulation



- Graph G = (V, E)
- V is set of nodes (i.e. pixels)
- E is a set of undirected edges between pairs of pixels
- w(vi, vj) is the weight of the edge between nodes vi and vj .
- S is a segmentation of a graph G such that G' = (V, E') where E' ⊂ E.
- S divides G into G' such that it contains distinct clusters C.

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Predicate for segmentation



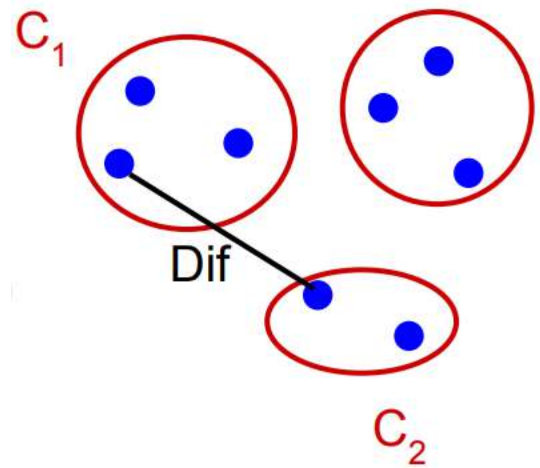
- Predicate D determines whether there is a boundary for segmentation.

$$Merge(C_1, C_2) = \begin{cases} True & \text{if } dif(C_1, C_2) < in(C_1, C_2) \\ False & otherwise \end{cases}$$

Where

- dif(C1 , C2) is the difference between two clusters.
- in(C1 , C2) is the internal difference in the clusters C1 and C2

Predicate for Segmentation



- Predicate D determines whether there is a boundary for segmentation.

$$Merge(C_1, C_2) = \begin{cases} True & \text{if } dif(C_1, C_2) < in(C_1, C_2) \\ False & otherwise \end{cases}$$

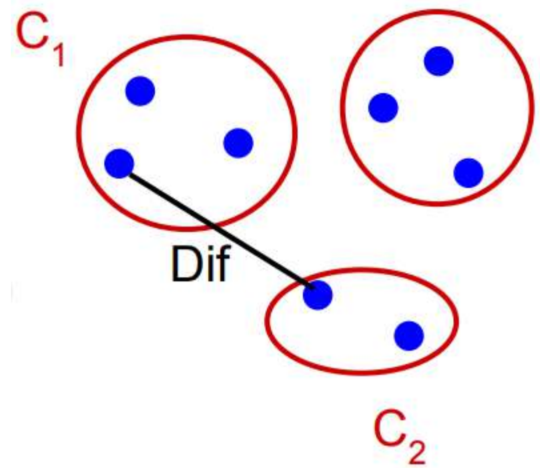
$$dif(C_1, C_2) = \min_{v_i \in C_1, v_j \in C_2, (C_1, C_2) \in E} w(v_i, v_j)$$

The difference between two components is the minimum weight edge that connects a node vi in cluster C1 to node vj in C2

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Predicate for Segmentation



- Predicate D determines whether there is a boundary for segmentation.

$$Merge(C_1, C_2) = \begin{cases} True & if\ dif(C_1, C_2) < in(C_1, C_2) \\ False & otherwise \end{cases}$$

$$dif(C_1, C_2) = \min_{v_i \in C_1, v_j \in C_2, (C_1, C_2) \in E} w(v_i, v_j)$$

$$in(C_1, C_2) = \min_{C \in \{C_1, C_2\}} \left[\max_{v_i, v_j \in C} \left[w(v_i, v_j) + \frac{k}{|C|} \right] \right]$$

in(C1, C2) is the maximum weight edge that connects two nodes in the same component.

Predicate for Segmentation

- k/|C| sets the threshold by which the components need to be different from the internal nodes in a component.
- Properties of constant k:
  - If k is large, it causes a preference for larger components.
  - k does not set a minimum size for components.

Algorithm for Segmentation

The input is a graph $G = (V, E)$, with n vertices and m edges. The output is a segmentation of V into components $S = (C_1, \ldots, C_r)$.

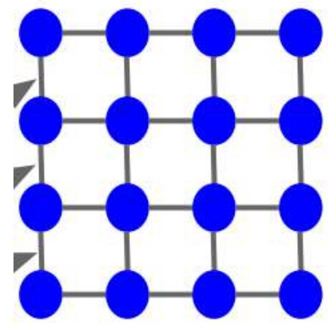
0. Sort E into $\pi = (o_1, \ldots, o_m)$, by non-decreasing edge weight.
1. Start with a segmentation S^0, where each vertex v_i is in its own component.
2. Repeat step 3 for $q = 1, \ldots, m$.
3. Construct S^q given S^{q-1} as follows. Let v_i and v_j denote the vertices connected by the q-th edge in the ordering, i.e., $o_q = (v_i, v_j)$. If v_i and v_j are in disjoint components of S^{q-1} and $w(o_q)$ is small compared to the internal difference of both those components, then merge the two components otherwise do nothing. More formally, let C_i^{q-1} be the component of S^{q-1} containing v_i and C_j^{q-1} the component containing v_j. If $C_i^{q-1} \neq C_j^{q-1}$ and $w(o_q) \leq MInt(C_i^{q-1}, C_j^{q-1})$ then S^q is obtained from S^{q-1} by merging C_i^{q-1} and C_j^{q-1}. Otherwise $S^q = S^{q-1}$.
4. Return $S = S^m$.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Features and weights

How to build the graph ? Two options:

1. Grid-graph: Every pixel is connected to its 8 neighboring pixels and the weights are determined by the difference in intensities.



2. NN-graph: Project every pixel into **feature space** defined by
   - (x, y, r, g, b).
   - Weights between pixels are determined using L2 (Euclidian) distance in feature space.
   - Edges are chosen for only top ten nearest neighbors in feature space to ensure run time of O(n log n) where n is number of pixels.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Results

With 8-neighbor grid graph
Edge weight: intensity difference

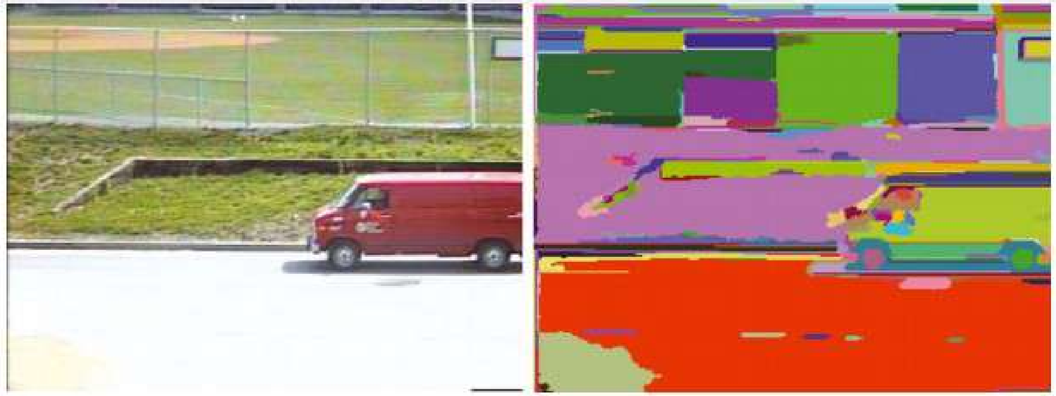


Figure 2. A street scene (320 × 240 color image), and the segmentation results produced by our algorithm ($\sigma = 0.8, k = 300$).

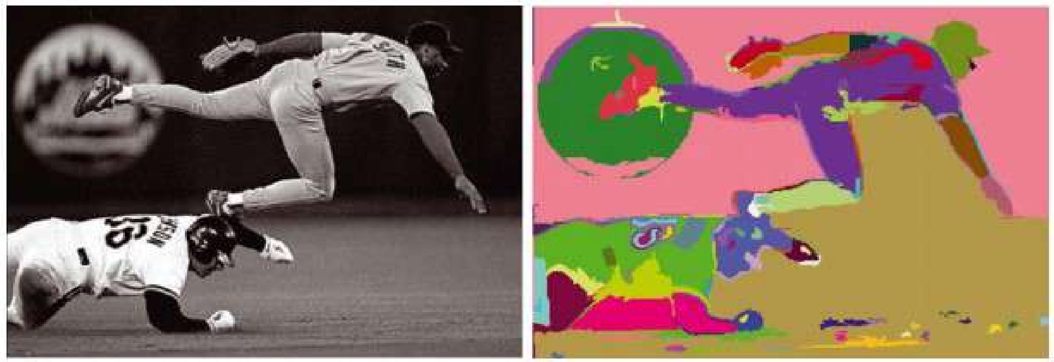


Figure 3. A baseball scene (432 × 294 grey image), and the segmentation results produced by our algorithm ($\sigma = 0.8, k = 300$).

Figure 4. An indoor scene (image 320 × 240, color), and the segmentation results produced by our algorithm ($\sigma = 0.8, k = 300$).

Results

With nearest neighbor graph
Edge weight: L2 distance in feature space

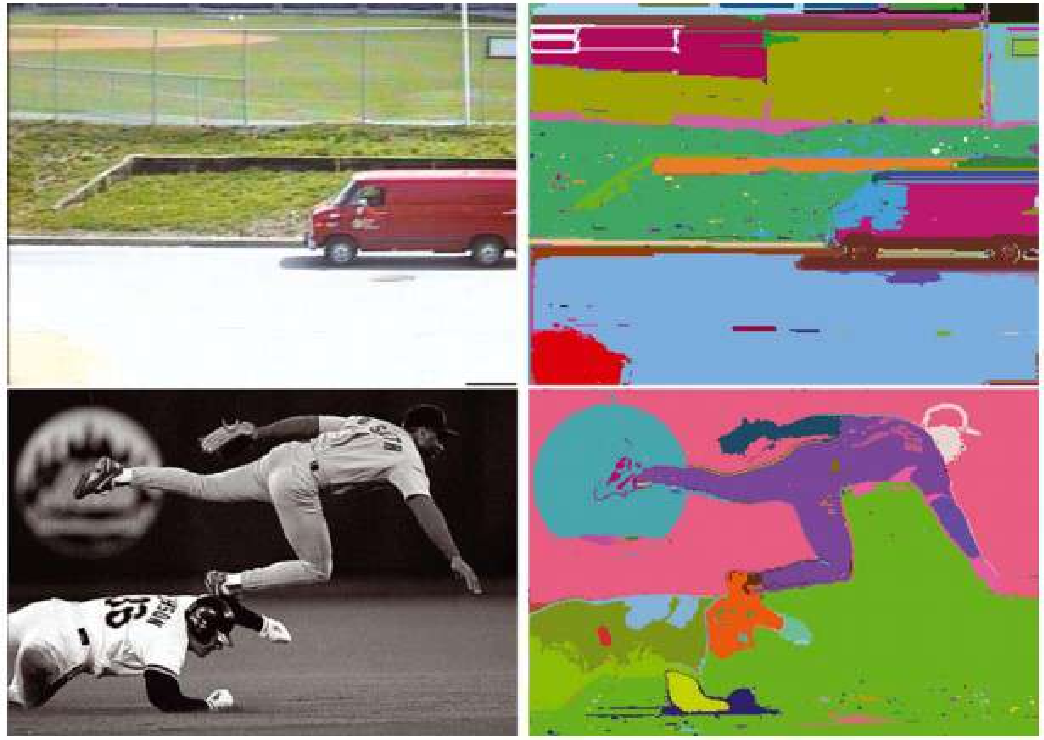


Figure 7. Segmentation of the street and baseball player scenes from the previous section, using the nearest neighbor graph rather than the grid graph ($\sigma = 0.8$, $k = 300$).

Figure 8. Segmentation using the nearest neighbor graph can capture spatially non-local regions ($\sigma = 0.8$, $k = 300$).

Results – close up

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 11: Visual Recognition – Image Classification

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Visual Recognition Tasks
- Introduction to segmentation and clustering
- Agglomerative clustering
- K-means clustering
- Mean-shift clustering
- Efficient Graph-based image segmentation

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- A simple Image Classification pipeline
  - Classification overview
- K-nearest neighbor algorithm
  - kNN: algorithm
  - kNN: analysis

[material based on Niebles-Krishna]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

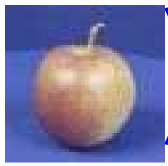
Today's Agenda

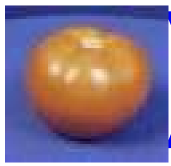
- A simple Image Classification pipeline
  - Classification overview
- K-nearest neighbor algorithm
  - kNN: algorithm
  - kNN: analysis

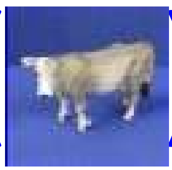
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Visual recognition: a classification framework

- Apply a prediction function to a feature representation of the image to get the desired output:

f() = "apple"

f() = "tomato"

f() = "cow"

Dataset: ETH-80, by B. Leibe Slide credit: L. Lazebnik

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The machine learning framework

$$y = f(\mathbf{x})$$

output   prediction function   Image feature

- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- **Testing:** apply f to a never-before-seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

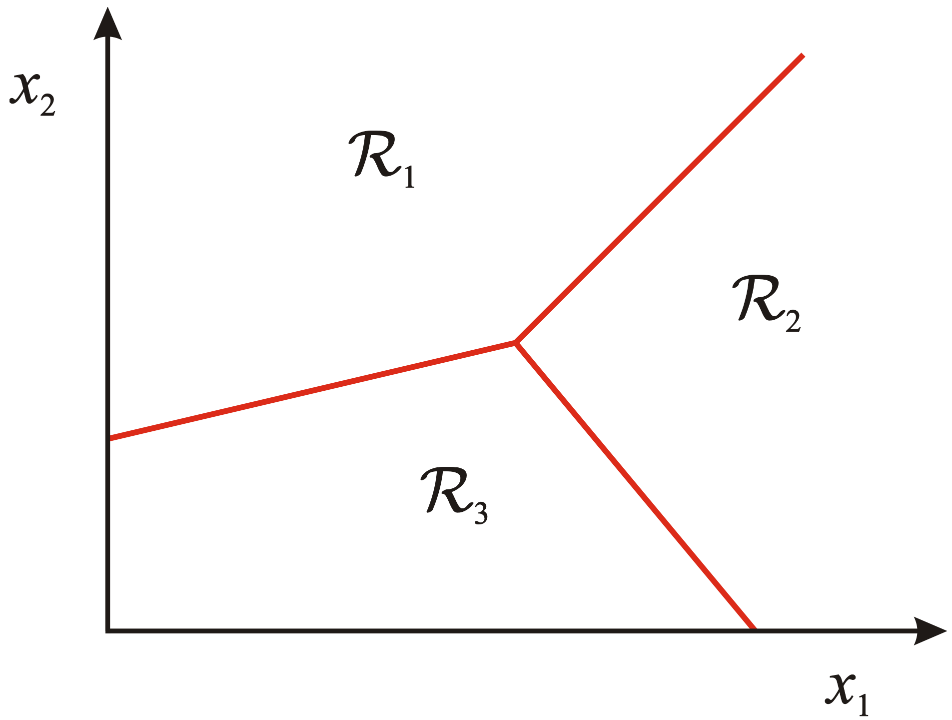
Slide credit: L. Lazebnik

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Classification

- Assign input feature vector to one of two or more classes
- Any decision rule divides input space into *decision regions* separated by *decision boundaries*



Slide credit: L. Lazebnik

A simple pipeline - Training

Training Images

Image Features

A simple pipeline - Training

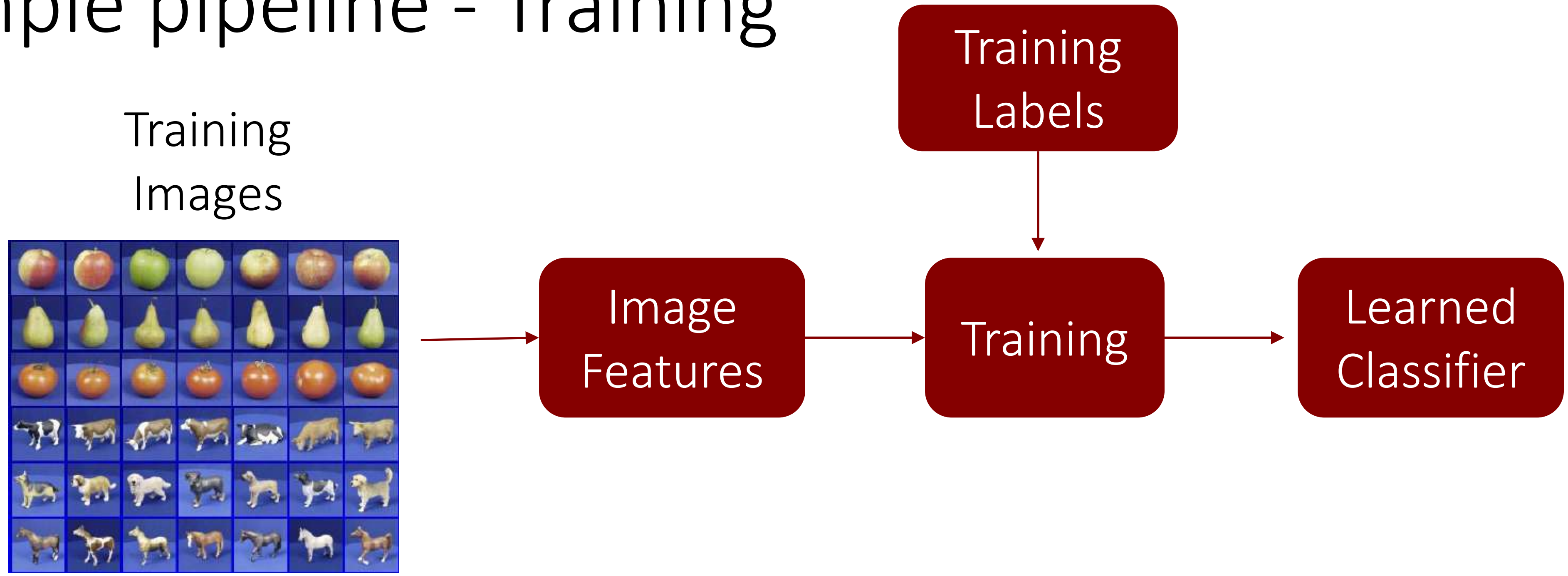
Training Images

Image Features

Training Labels

Training

A simple pipeline - Training



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A simple pipeline - Testing

Training Images

Training Labels

Image Features

Training

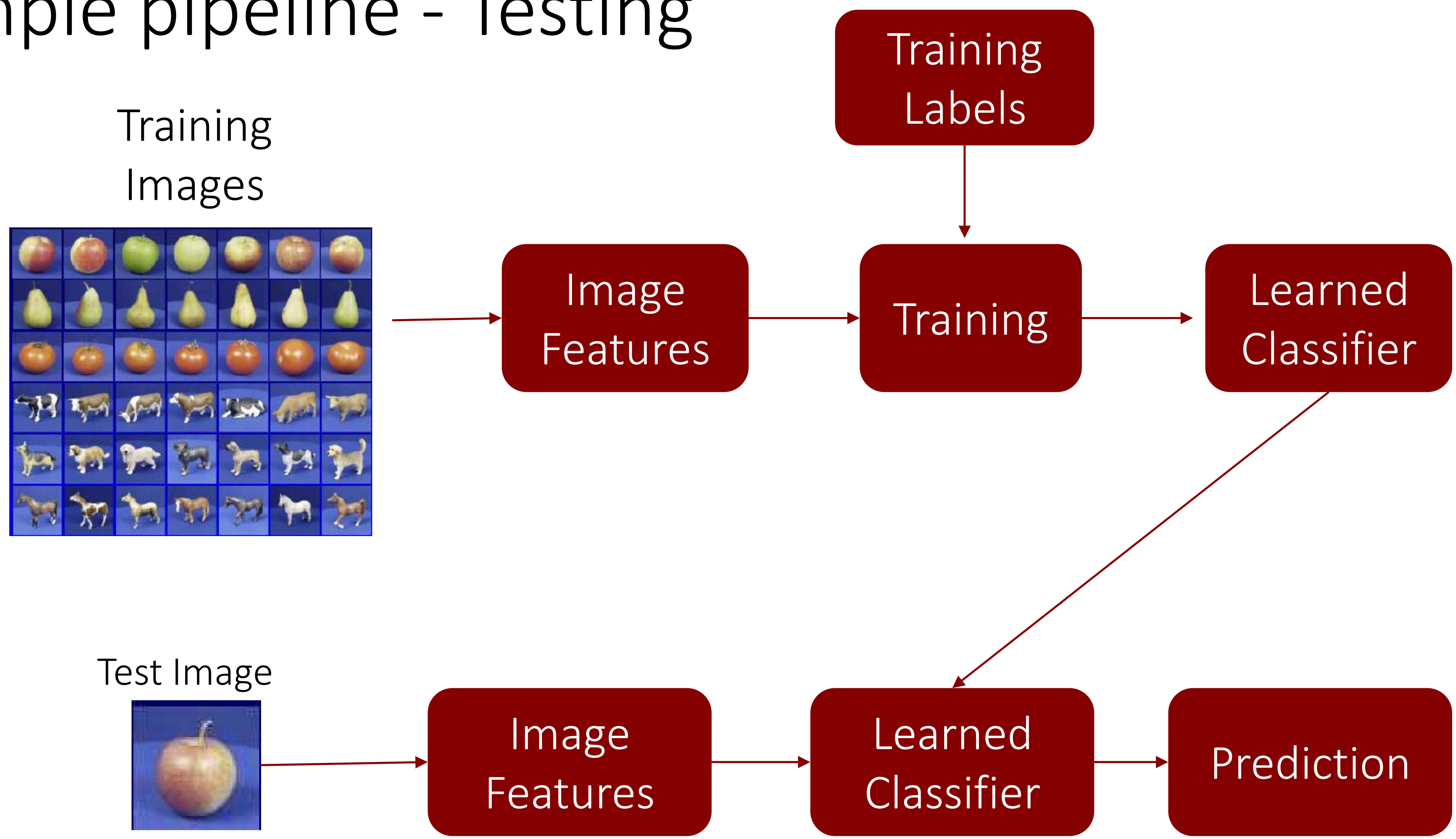
Learned Classifier

Test Image

Image Features

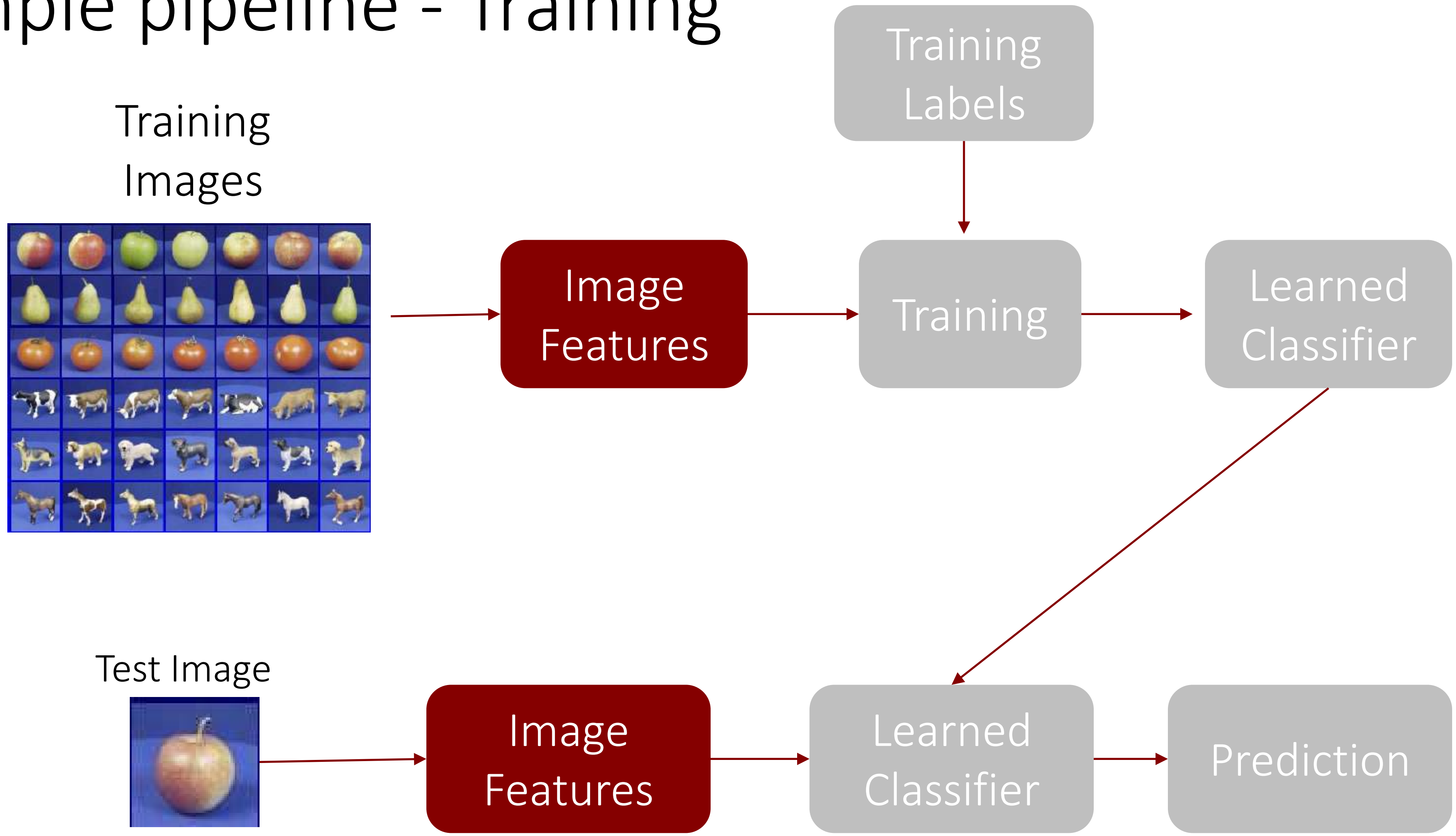
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A simple pipeline - Testing



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A simple pipeline - Training



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image Features

Input image

Many features to choose from

- Histogram of Color
- Histogram of Gradients
- SIFT
- Bag of words
- etc...

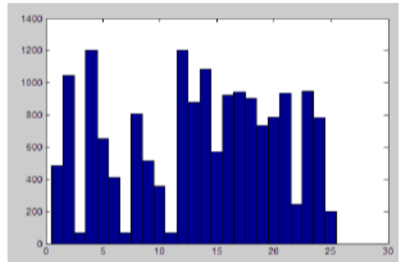
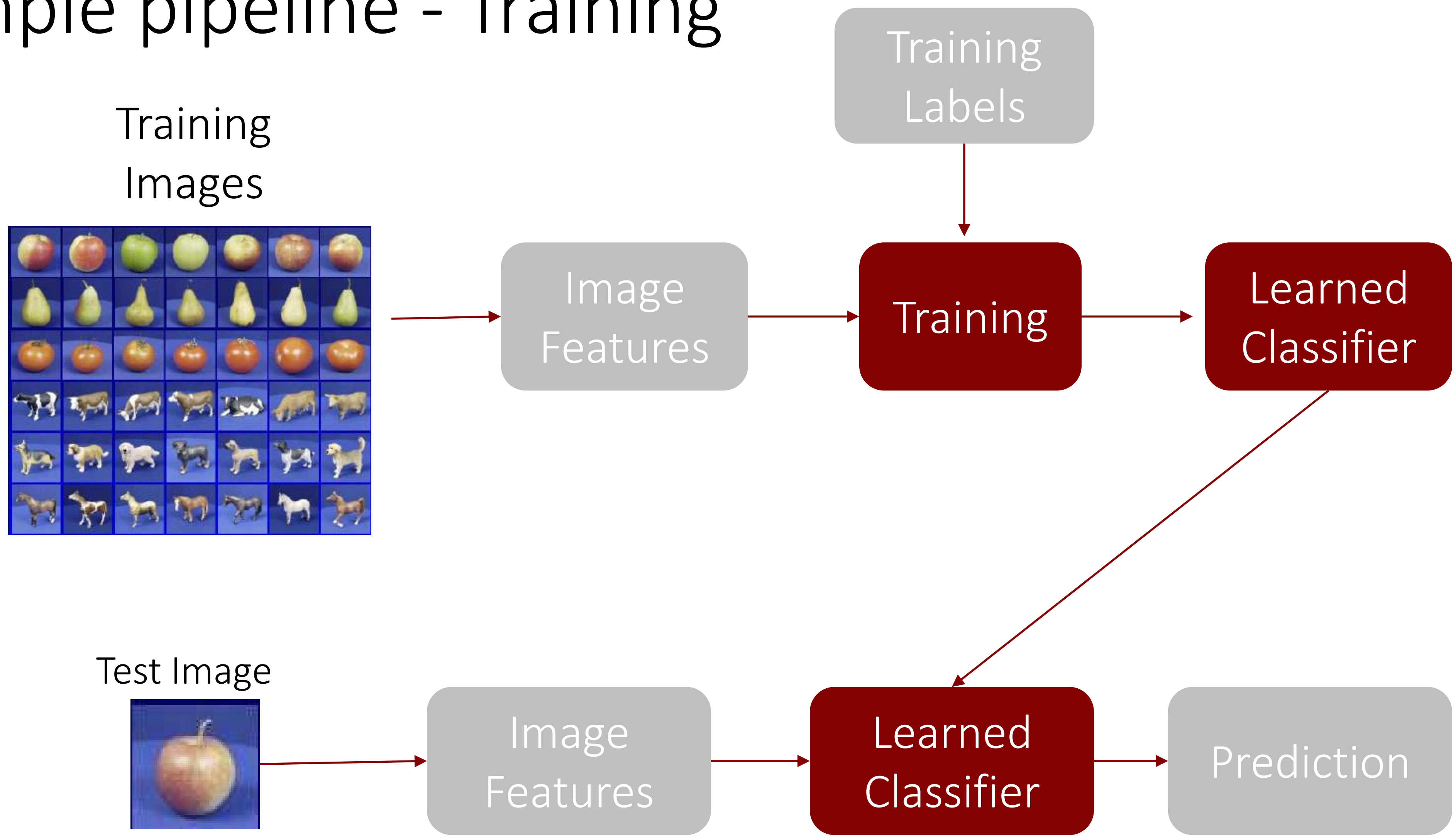


Image From ETH-80 dataset: Analyzing Appearance and Contour Based Methods for Object Categorization - by Leibe et al 2003

A simple pipeline - Training



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Many classifiers to choose from

- K-nearest neighbor
- SVM
- Neural networks
- Naïve Bayes
- Bayesian network
- Logistic regression
- Randomized Forests
- Boosted Decision Trees
- Restricted Boltzmann Machines
- Etc.

Which is the best one?

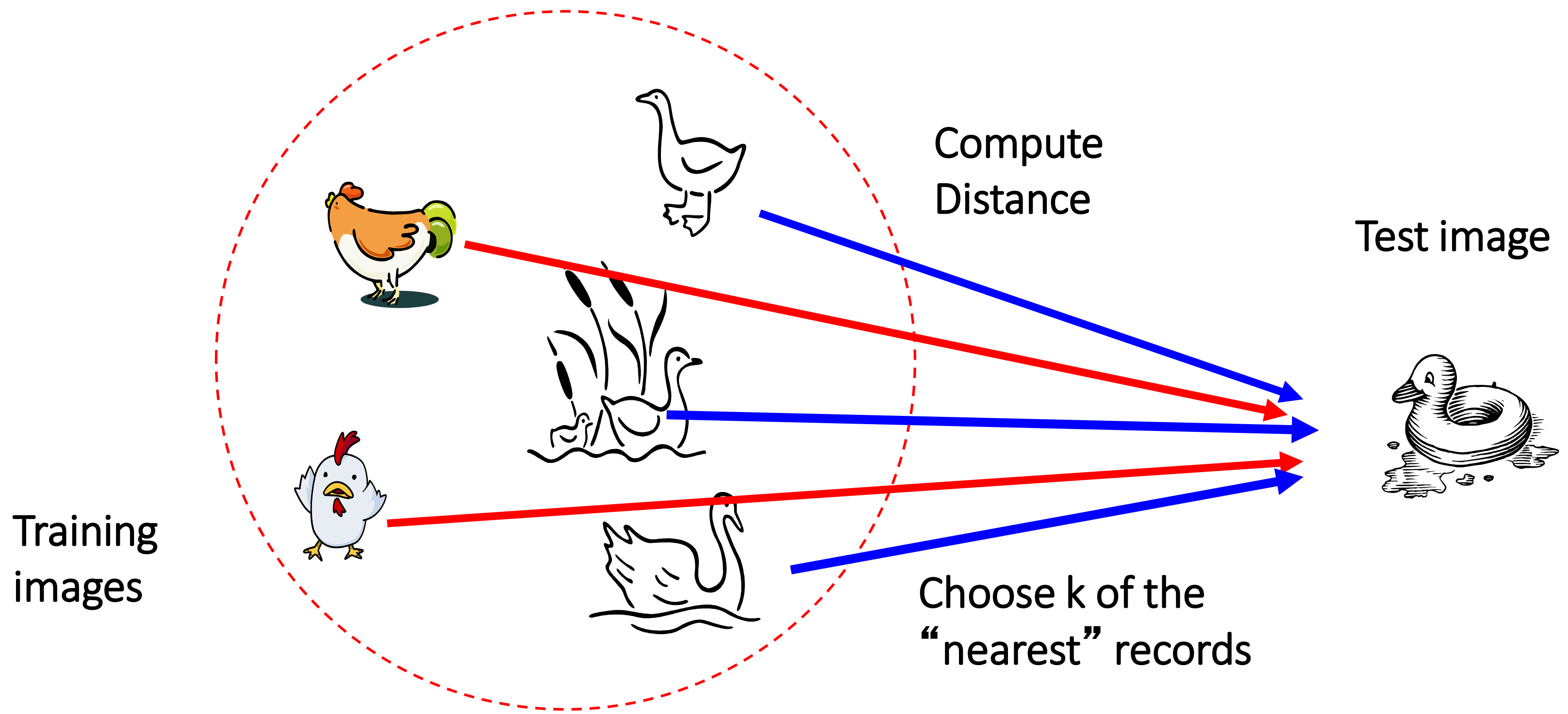
Slide credit: D. Hoiem

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

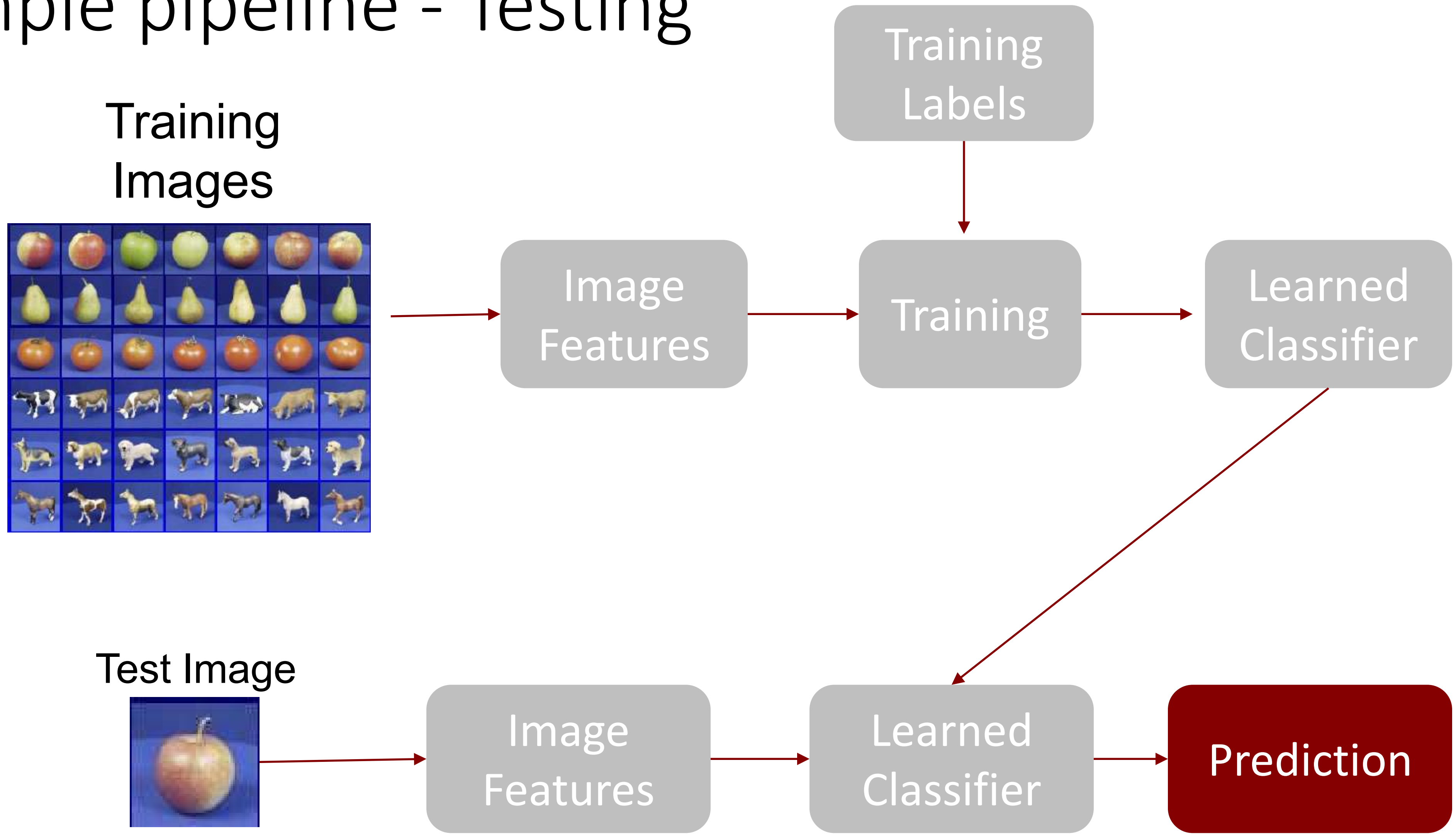
Classifiers: Nearest neighbor

Assign label of nearest training data point to each test data point

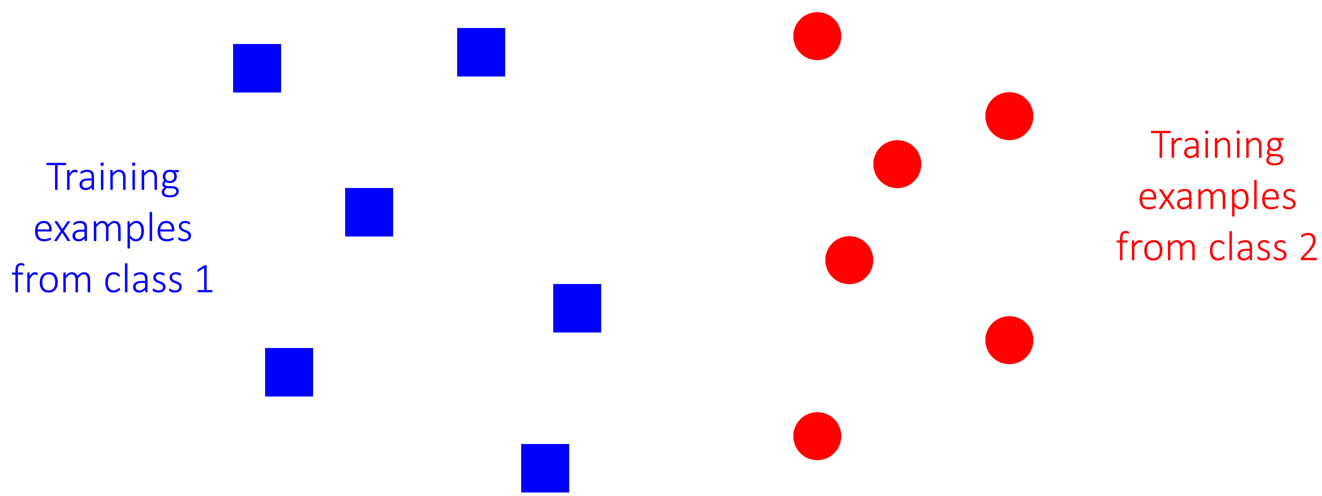


Source: N. Goyal

A simple pipeline - Testing

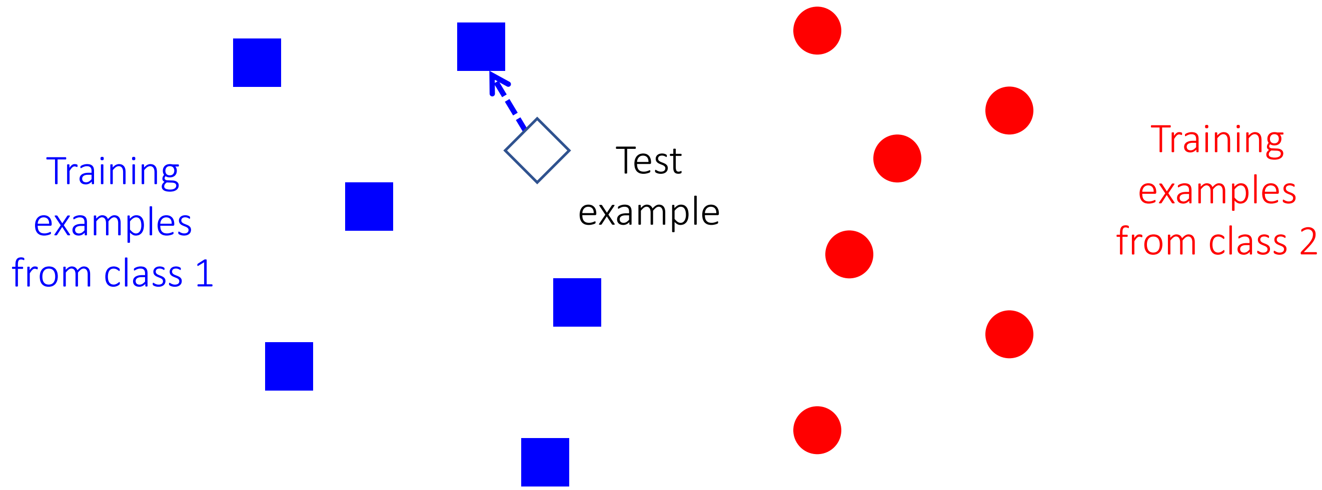


Classifiers: Nearest neighbor



Slide credit: L. Lazebnik

Classifiers: Nearest neighbor



Slide credit: L. Lazebnik

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- A simple Image Classification pipeline
  - Classification overview
- K-nearest neighbor algorithm
  - kNN: algorithm
  - kNN: analysis

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

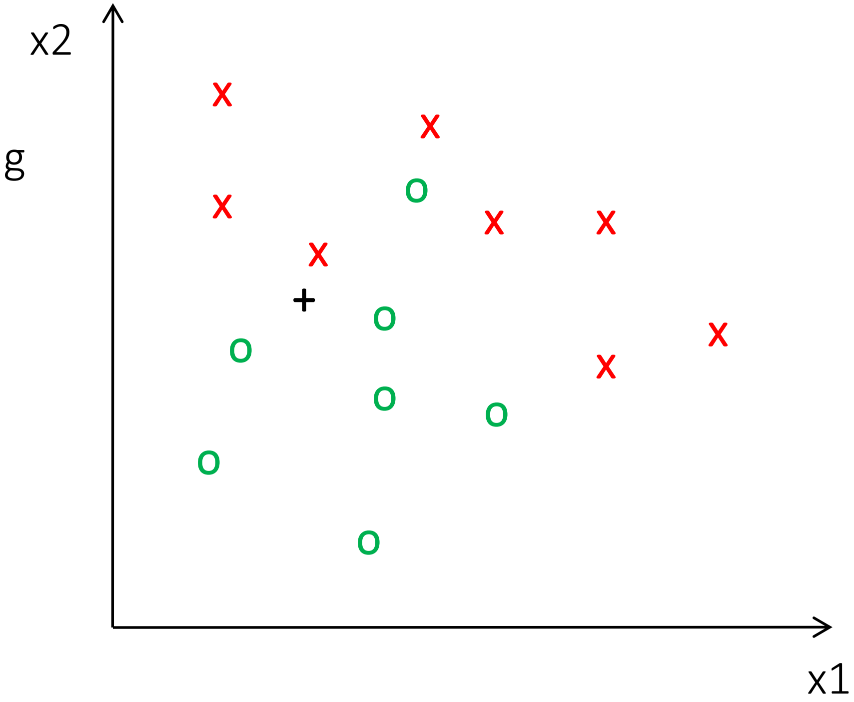
K-nearest neighbor

- Algorithm (training):

1. Store all training data points $\boldsymbol{x}_i$ with their corresponding category labels y_i

- Algorithm (testing):

1. We are given a new test point $\boldsymbol{x}$
2. Compute distance to all training data points
3. Select k training points closest to $\boldsymbol{x}$
4. Assign $\boldsymbol{x}$ to label y that is most common among the k nearest neighbors.

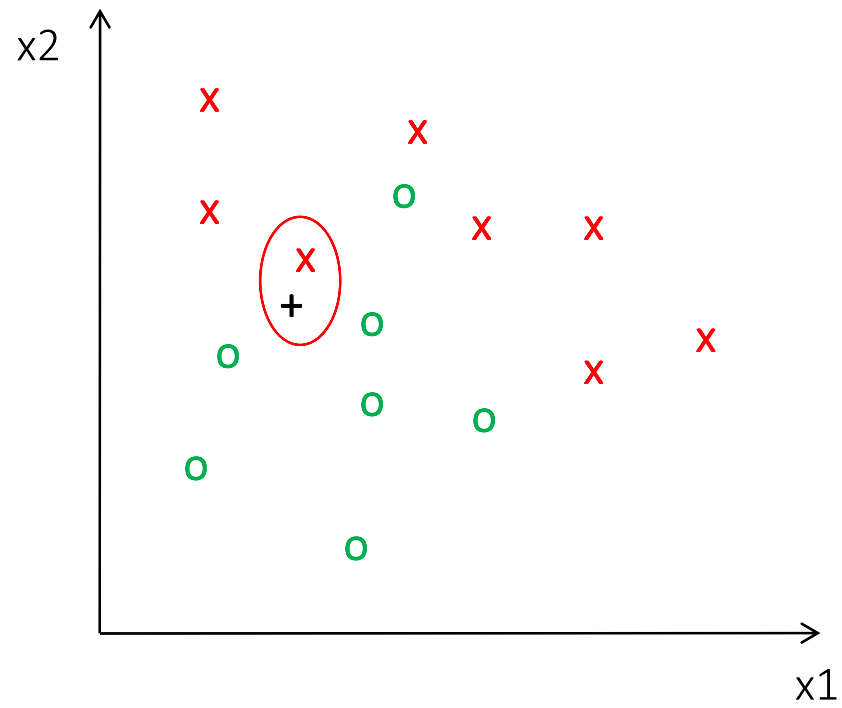


- Distance measurement: – Euclidean

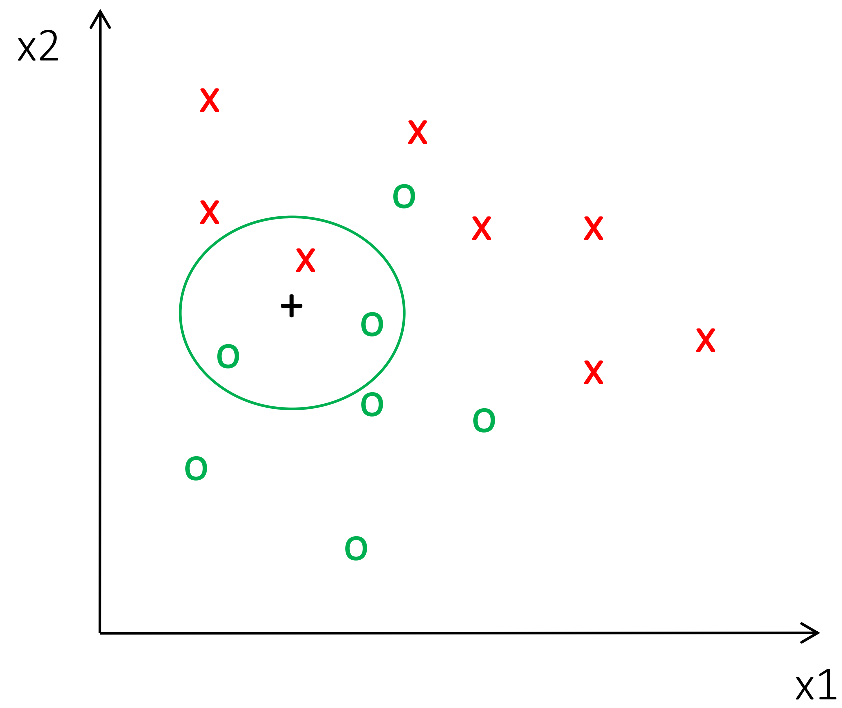
$$Dist(X^n, X^m) = \sqrt{\sum_{i=1}^{D} (X_i^n - X_i^m)^2}$$

Where X^n and X^m are the n-th and m-th data points

1-nearest neighbor



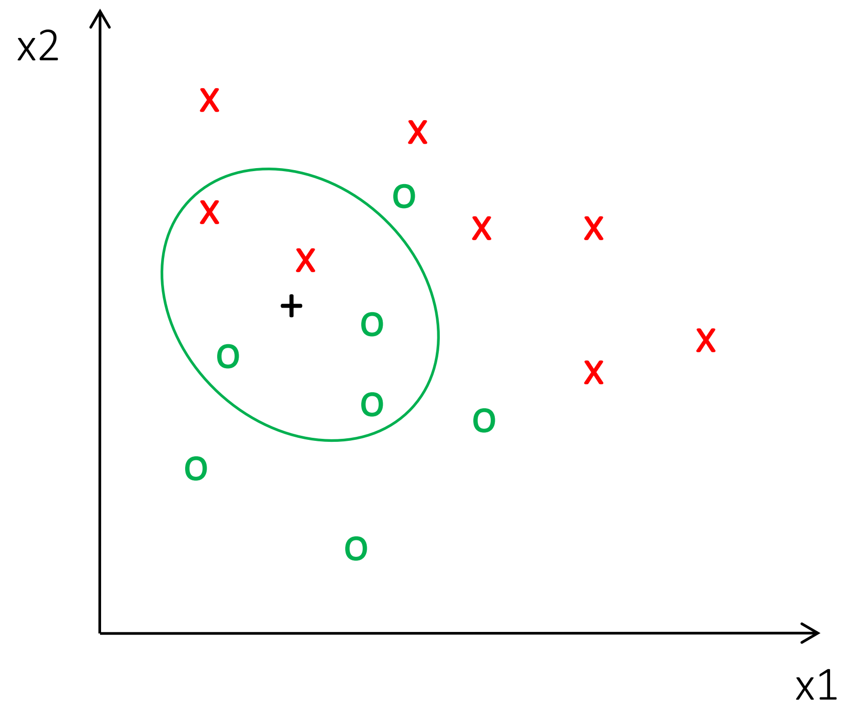
3-nearest neighbor



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

5-nearest neighbor



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- A simple Image Classification pipeline
  - Classification overview
- K-nearest neighbor algorithm
  - kNN: algorithm
  - kNN: analysis

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

K-NN: a very useful algorithm

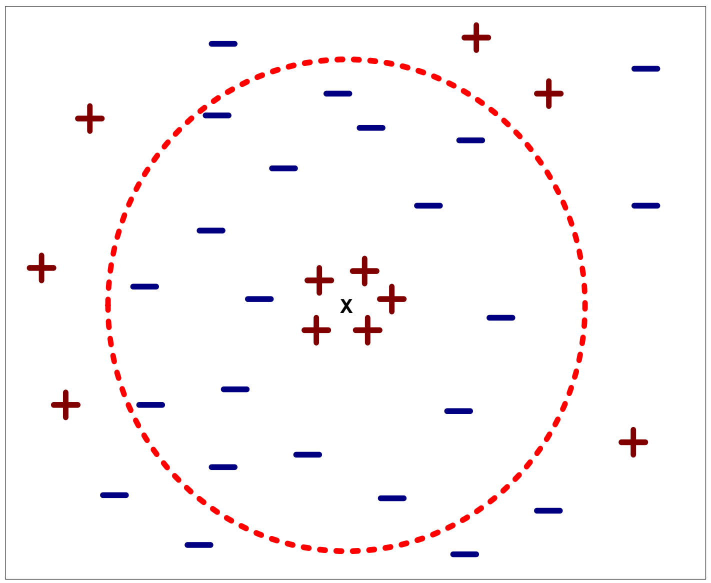
- Simple, a good one to try first
- Very flexible decision boundaries

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

K-NN: issues to keep in mind

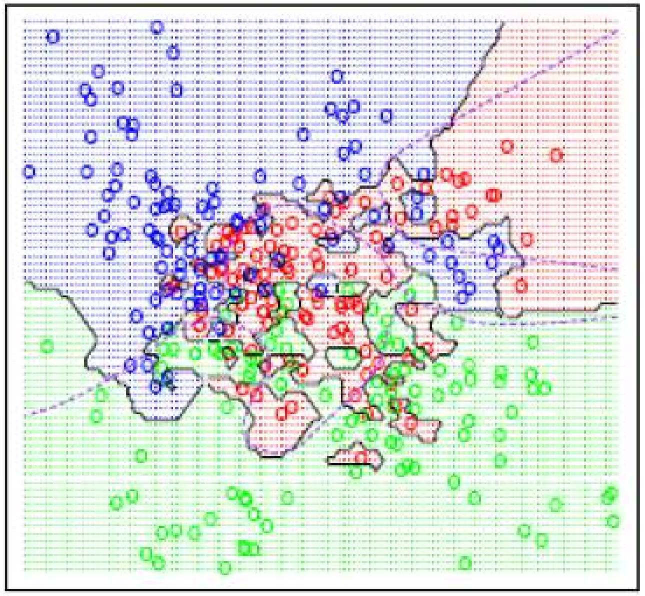
- Choosing the value of k:
  - If too small, sensitive to noise points
  - If too large, neighborhood may include points from other classes



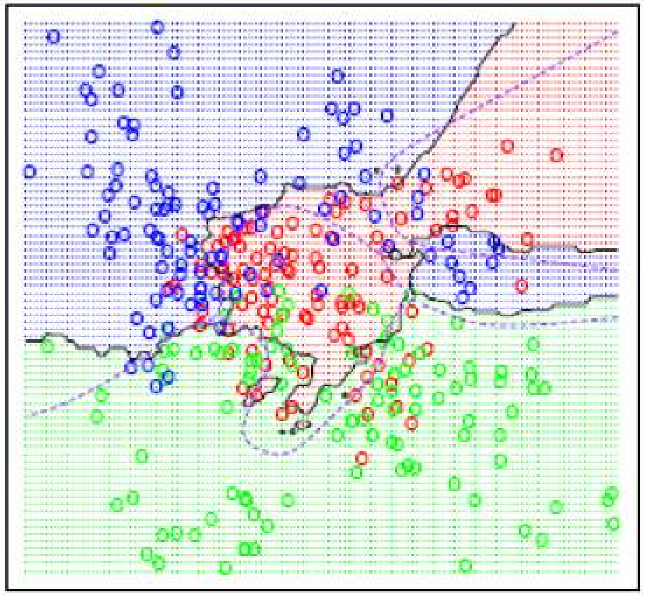
K-NN: issues to keep in mind

- Choosing the value of k:
  - If too small, sensitive to noise points
  - If too large, neighborhood may include points from other classes

K=1



K=15

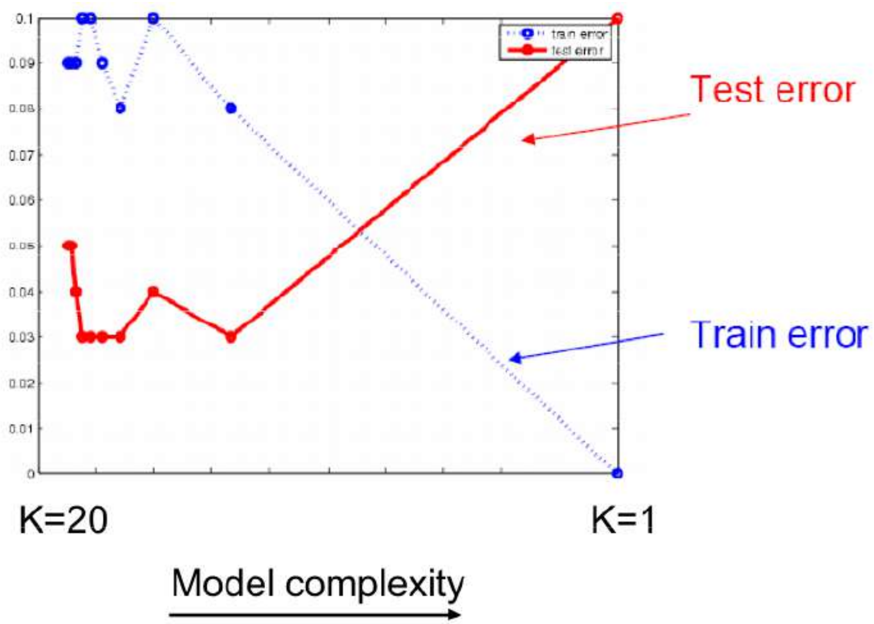


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

K-NN: issues to keep in mind

- Choosing the value of k:
  - If too small, sensitive to noise points
  - If too large, neighborhood may include points from other classes
  - Solution: cross validate!

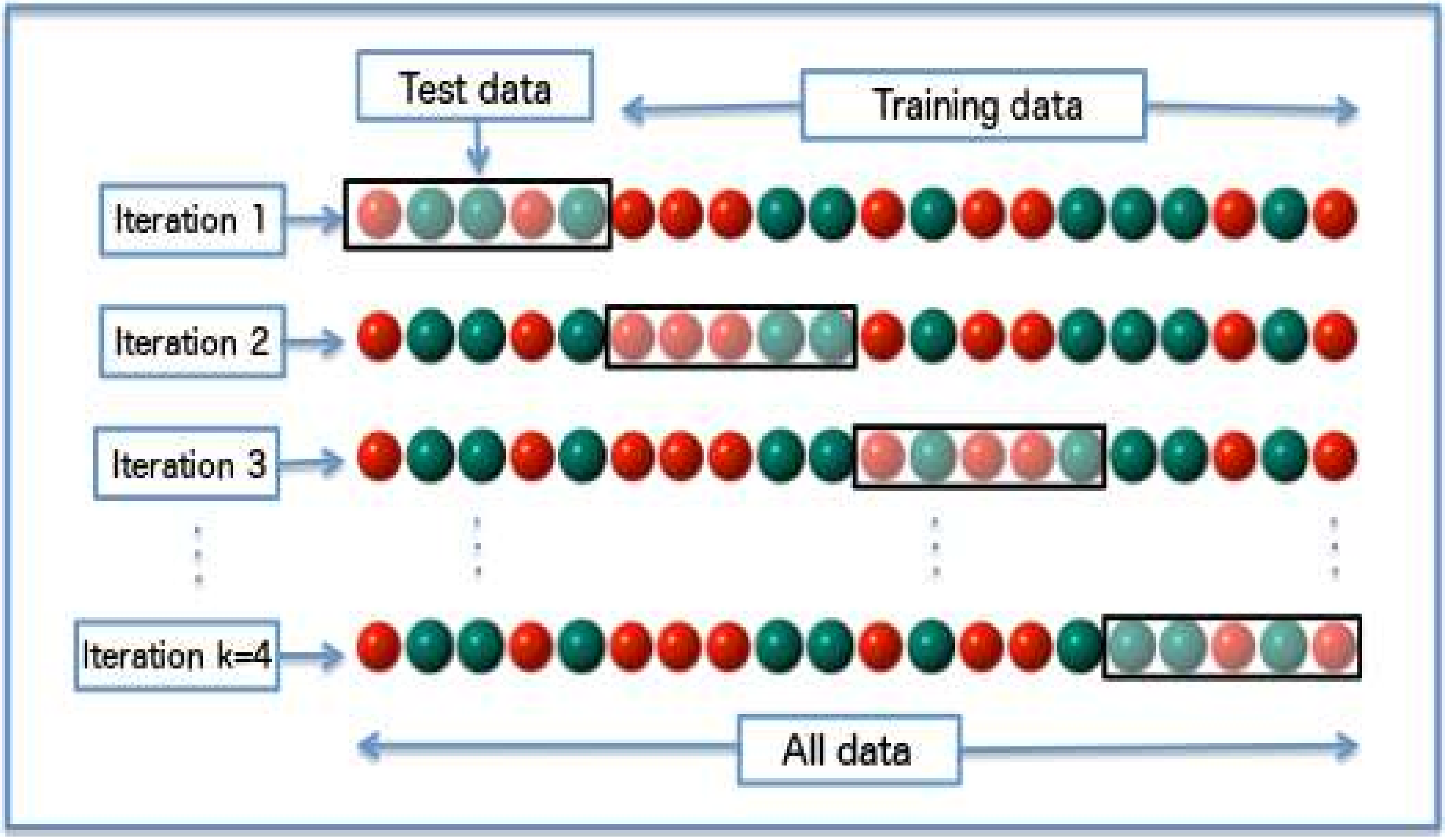


This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Cross validation

- For each value of k in the nearest neighbors algorithm:
- Create multiple train/test splits
    - For each split:
        - Measure performance
- Average performance over all splits

- Select k with best average performance



K-NN: issues to keep in mind

- Choosing the value of k:
  - If too small, sensitive to noise points
  - If too large, neighborhood may include points from other classes
  - Solution: cross validate!

- Can produce counter-intuitive results (using Euclidean measure)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Euclidean measure

1 1 1 1 1 1 1 1 1 1 1 1 0
0 1 1 1 1 1 1 1 1 1 1 1 1

d = 1.4142

VS

1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1

d = 1.4142

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

K-NN: issues to keep in mind

- Choosing the value of k:
  - If too small, sensitive to noise points
  - If too large, neighborhood may include points from other classes
  - Solution: cross validate!
- Can produce counter-intuitive results (using Euclidean measure)
  - Solution: normalize the vectors to unit length

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

K-NN: issues to keep in mind

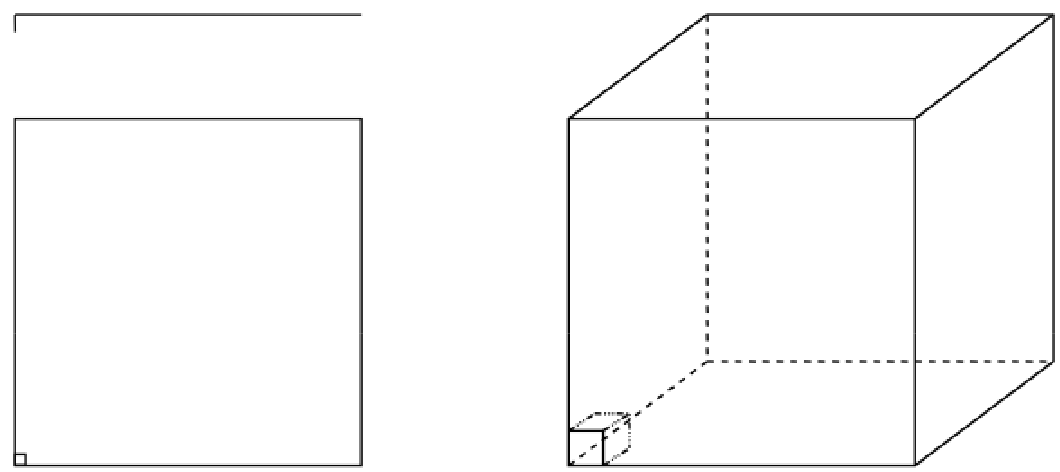
- Choosing the value of k:
  - If too small, sensitive to noise points
  - If too large, neighborhood may include points from other classes
  - Solution: cross validate!
- Can produce counter-intuitive results (using Euclidean measure)
  - Solution: normalize the vectors to unit length
- Curse of Dimensionality

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

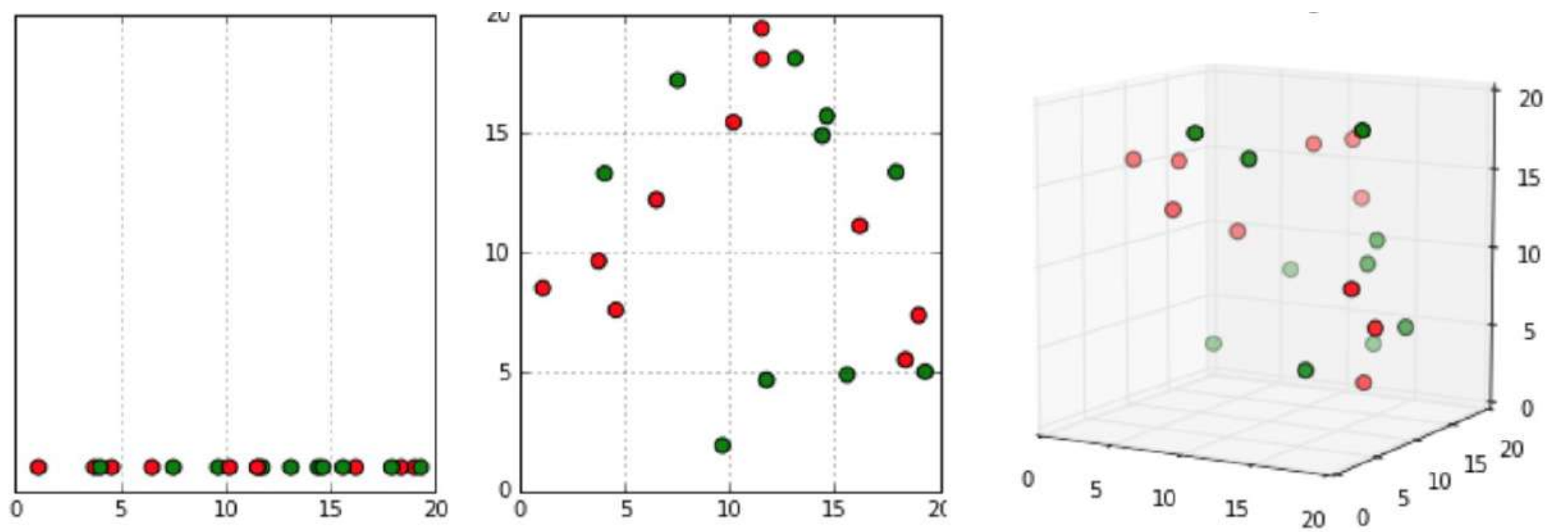
Curse of dimensionality

- Assume 5000 points uniformly distributed in the unit (hyper)cube and we want to apply 5-NN. Suppose our query point is at the origin.
  - In 1-dimension, we must go a distance of 5/5000=0.001 on average to capture 5 nearest neighbors.
  - In 2 dimensions, we must go $\sqrt{0.001}$ to get a square that contains 0.001 of the volume.
  - In d dimensions, we must go $(0.001)^{1/d}$



Curse of dimensionality

- Assume 5000 points uniformly distributed in the unit hypercube and we want to apply 5-NN. Suppose our query point is at the origin.
  - In 1-dimension, we must go a distance of 5/5000=0.001 on average to capture 5 nearest neighbors.
  - In 2 dimensions, we must go $\sqrt{0.001}$ to get a square that contains 0.001 of the volume.
  - In d dimensions, we must go $(0.001)^{1/d}$



K-NN: issues to keep in mind

- Choosing the value of k:
  - If too small, sensitive to noise points
  - If too large, neighborhood may include points from other classes
  - Solution: cross validate!
- Can produce counter-intuitive results (using Euclidean measure)
  - Solution: normalize the vectors to unit length
- Curse of Dimensionality
  - Solution: no good one – need to get more data

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 12: Visual Bag of Words

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union

Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- A simple Image Classification pipeline
  - Classification overview
- K-nearest neighbor algorithm
  - kNN: algorithm
  - kNN: analysis

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Visual bag of words (BoW)
  - Background
  - Algorithm
- Applications
  - Image search
- Spatial Pyramid Matching

[material based on Niebles-Krishna]

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

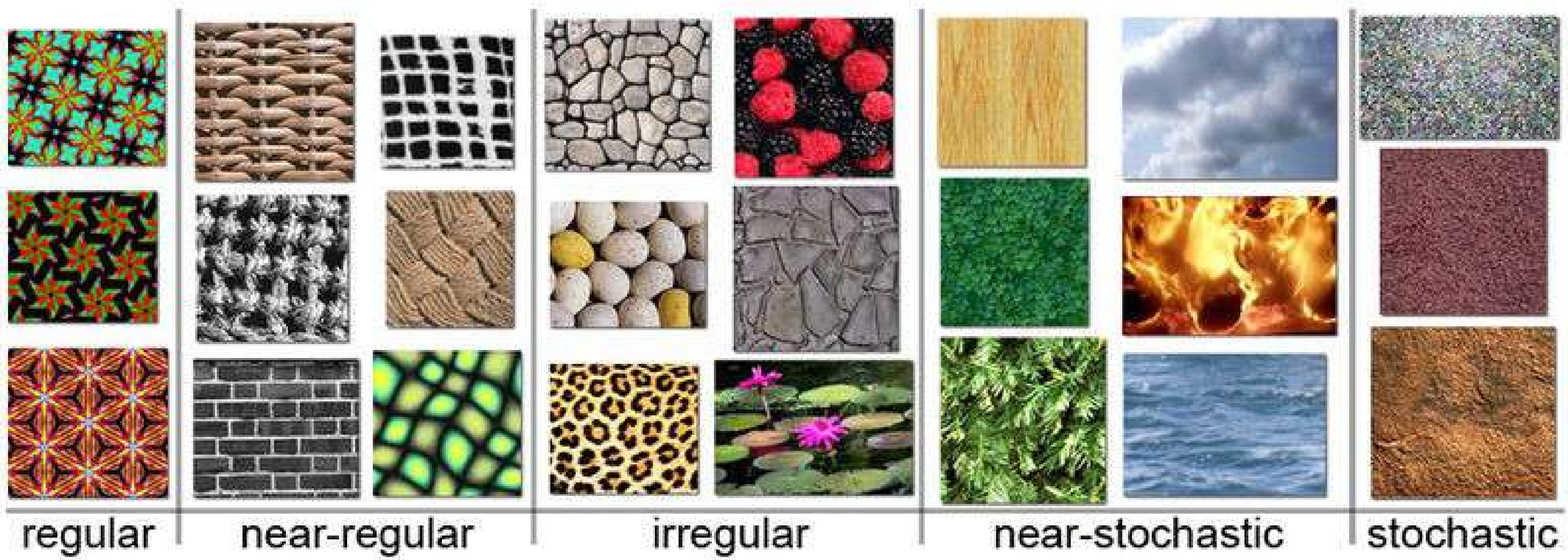
- Visual bag of words (BoW)
  - Background
  - Algorithm
- Applications
  - Image search
- Spatial Pyramid Matching

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Object → Bag of 'words'

Origin 1: Texture Recognition

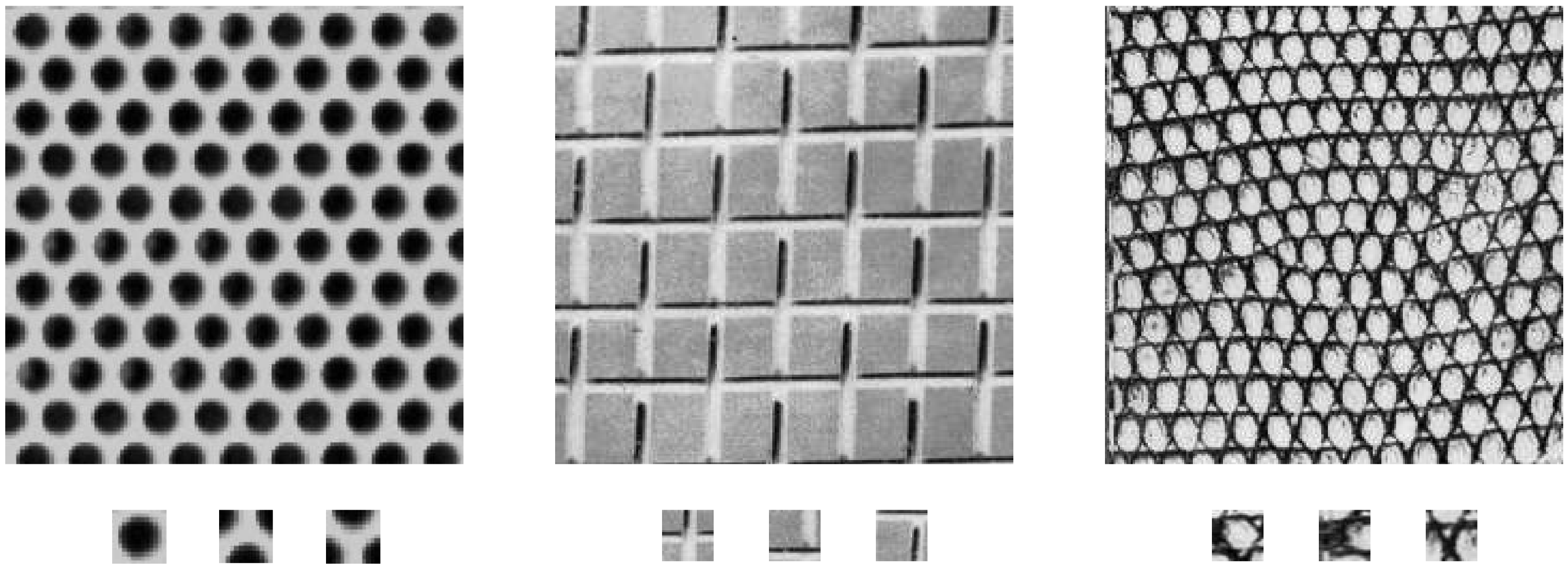


Example textures (from Wikipedia)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Origin 1: Texture Recognition

Texture is characterized by the repetition of basic elements or ***textons***



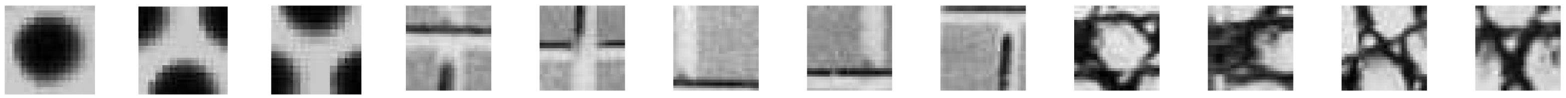
Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Origin 1: Texture Recognition

Recognition based on identity of the textons, not their spatial arrangement (although that is very important too!)

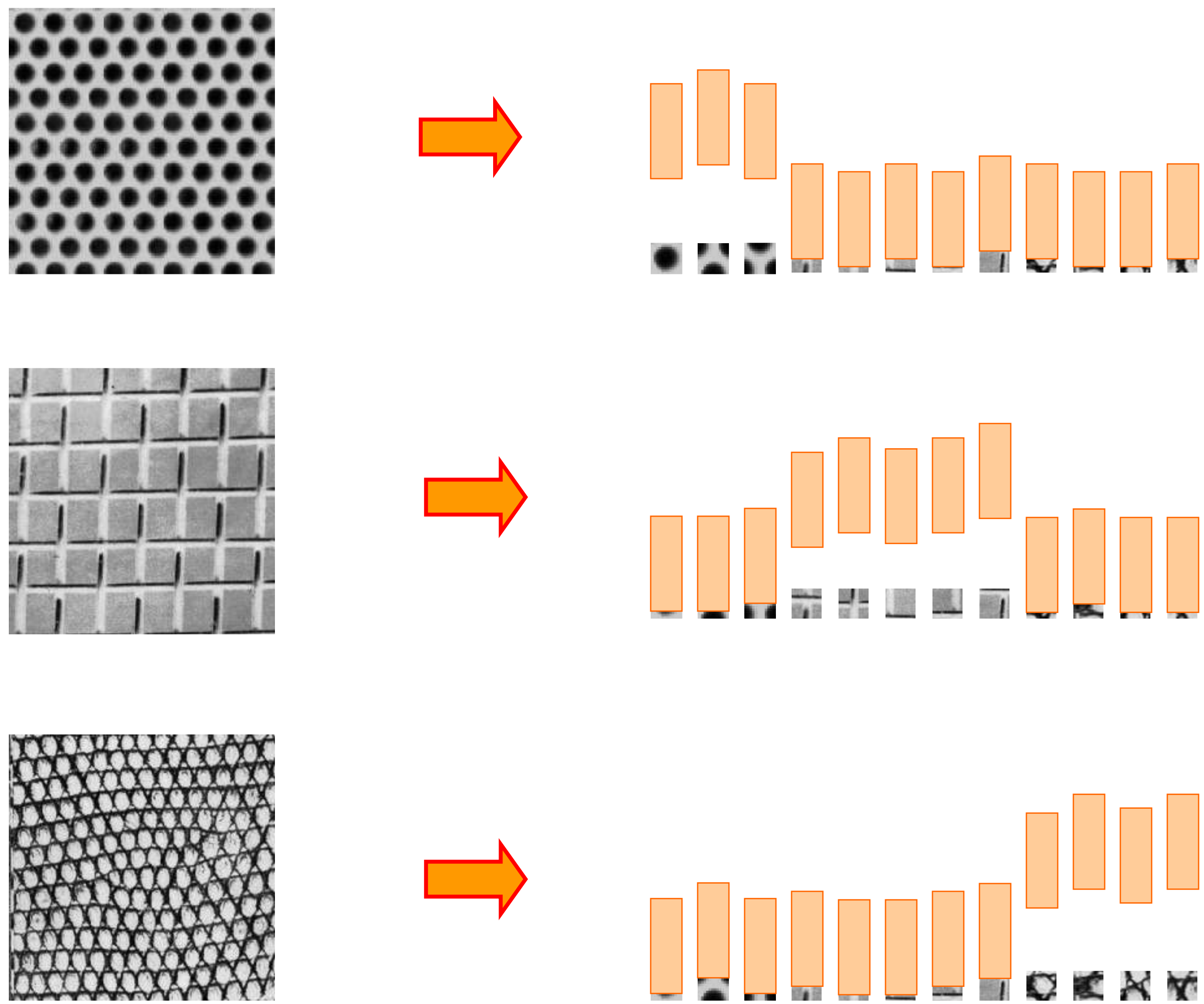


Universal texton dictionary

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Origin 1: Texture Recognition



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary
Salton & McGill (1983)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary

Salton & McGill (1983)

US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary

Salton & McGill (1983)

US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

Origin 2: Bag-of-words models

Orderless document representation: frequencies of words from a dictionary
Salton & McGill (1983)

US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bags of features for object recognition

face, flowers, building

Works pretty well for image-level classification and for recognizing object *instances*

Csurka et al. (2004), Willamowski et al. (2005), Grauman & Darrell (2005), Sivic et al. (2003, 2005)

Bags of features for object recognition

class	bag of features	bag of features	Parts-and-shape model
	Zhang et al. (2005)	Willamowski et al. (2004)	Fergus et al. (2003)
airplanes	**98.8**	97.1	90.2
cars (rear)	98.3	**98.6**	90.3
cars (side)	**95.0**	87.3	88.5
faces	**100**	99.3	96.4
motorbikes	**98.5**	98.0	92.5
spotted cats	**97.0**	—	90.0

Today’s Agenda

- Visual bag of words (BoW)
  - Background
  - Algorithm
- Applications
  - Image search
- Spatial Pyramid Matching

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of features

- First, take a bunch of images, extract features, and build up a “dictionary” or “visual vocabulary” – a list of common features

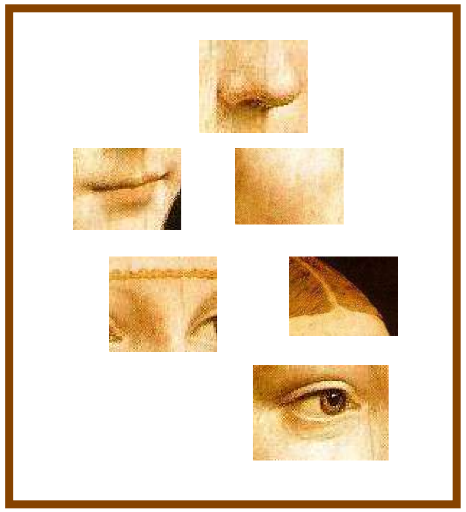
- Given a new image, extract features and build a histogram – for each feature, find the closest visual word in the dictionary

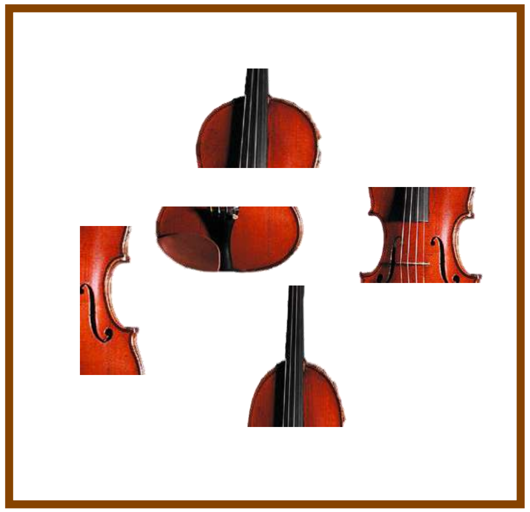
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of features: outline

1. Extract features





Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of features: outline

1. Extract features
2. Learn "visual vocabulary"

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

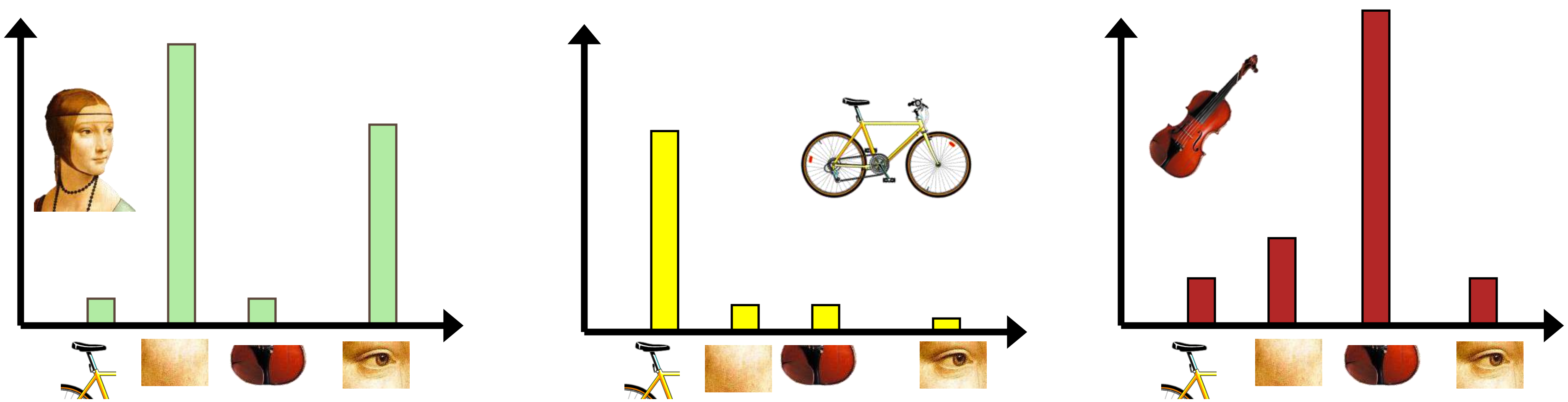
Bag of features: outline

1. Extract features
2. Learn "visual vocabulary"
3. Quantize features using visual vocabulary

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of features: outline

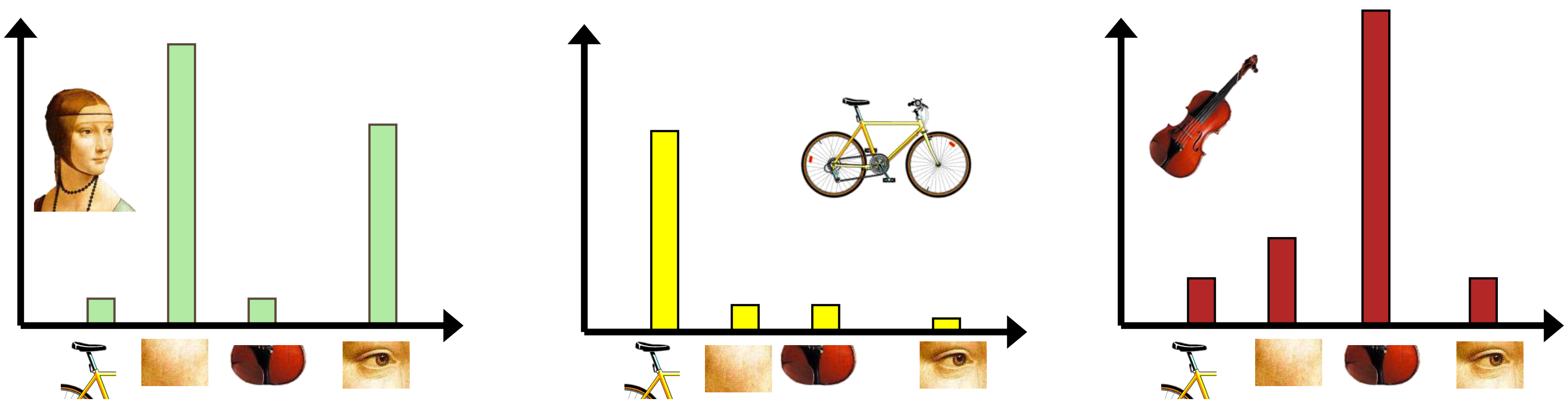
1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”

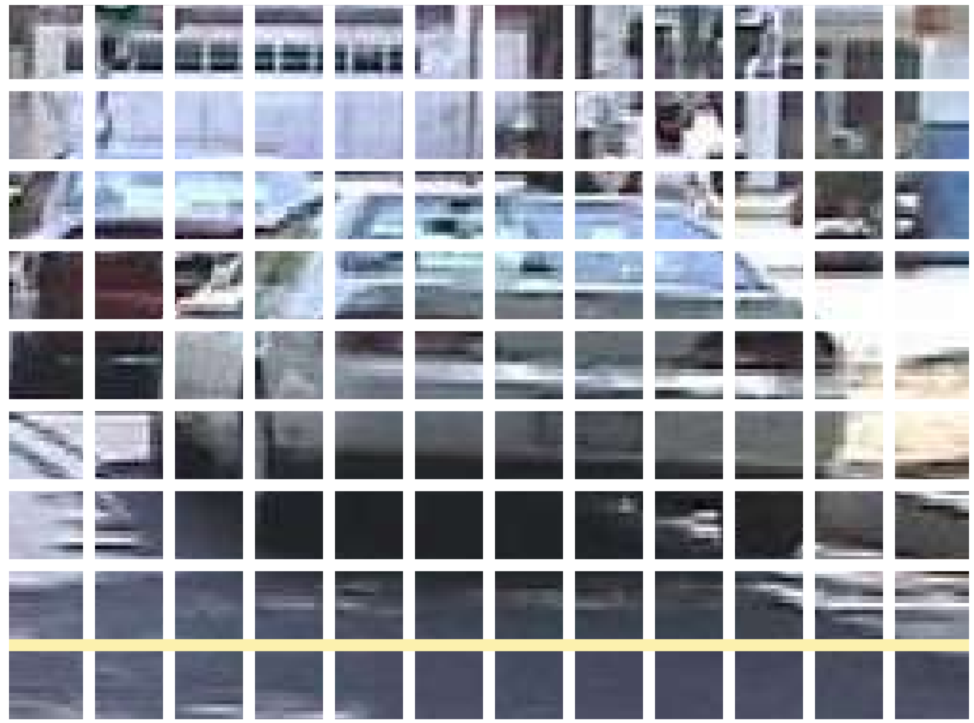


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

1. Feature extraction

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005

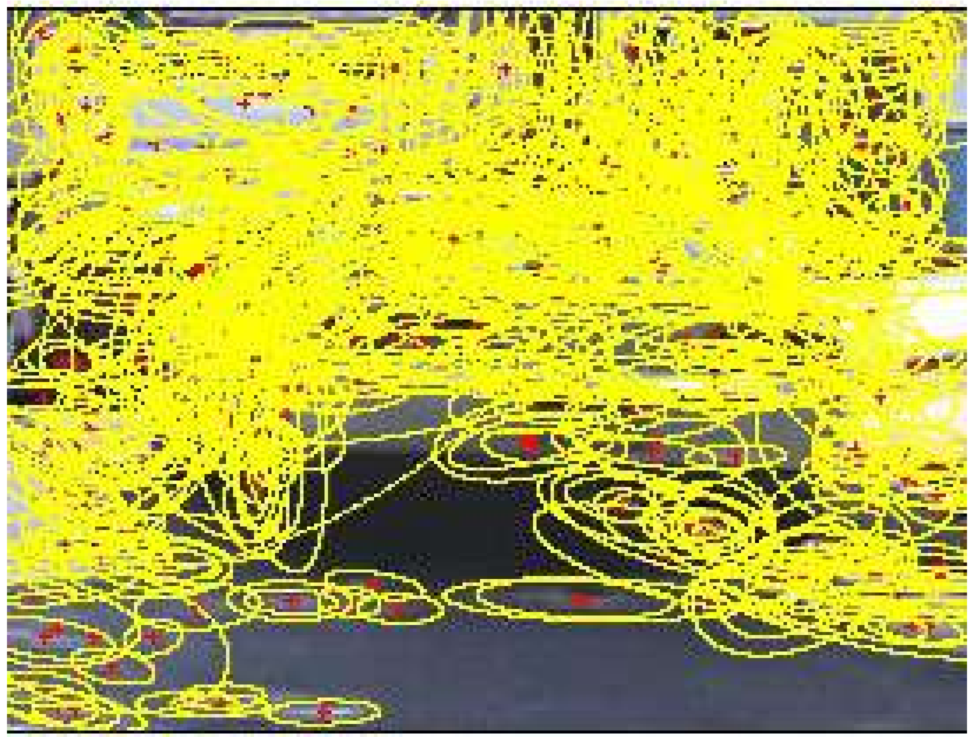


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

1. Feature extraction

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic et al. 2005



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

1. Feature extraction

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic et al. 2005
- Other methods
  - Random sampling (Vidal-Naquet & Ullman, 2002)
  - Segmentation-based patches (Barnard et al. 2003)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of features: outline

1. Extract features
2. Learn "visual vocabulary"
3. Quantize features using visual vocabulary
4. Represent images by frequencies of "visual words"

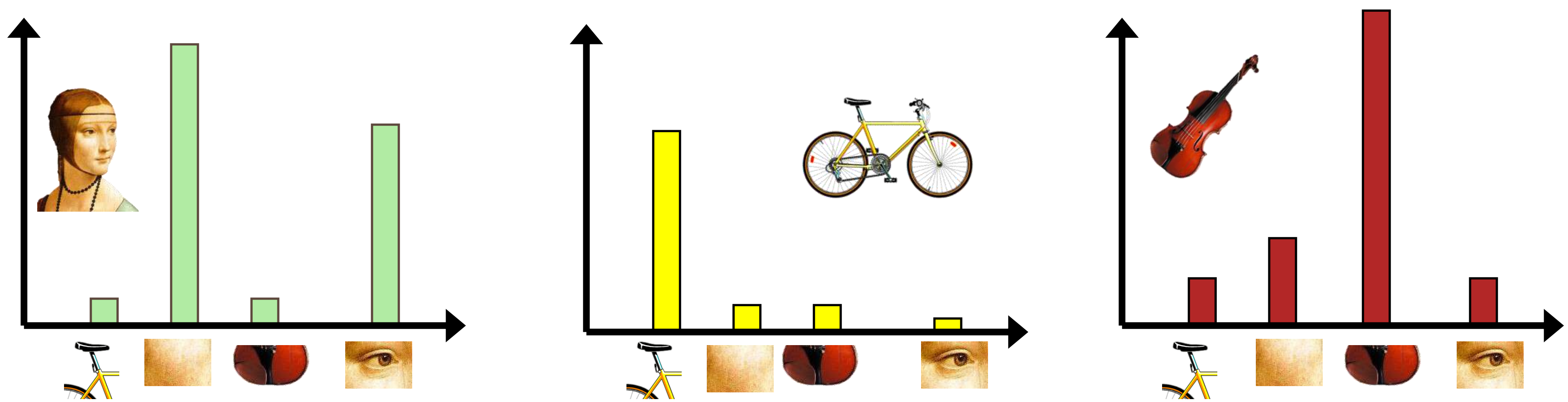
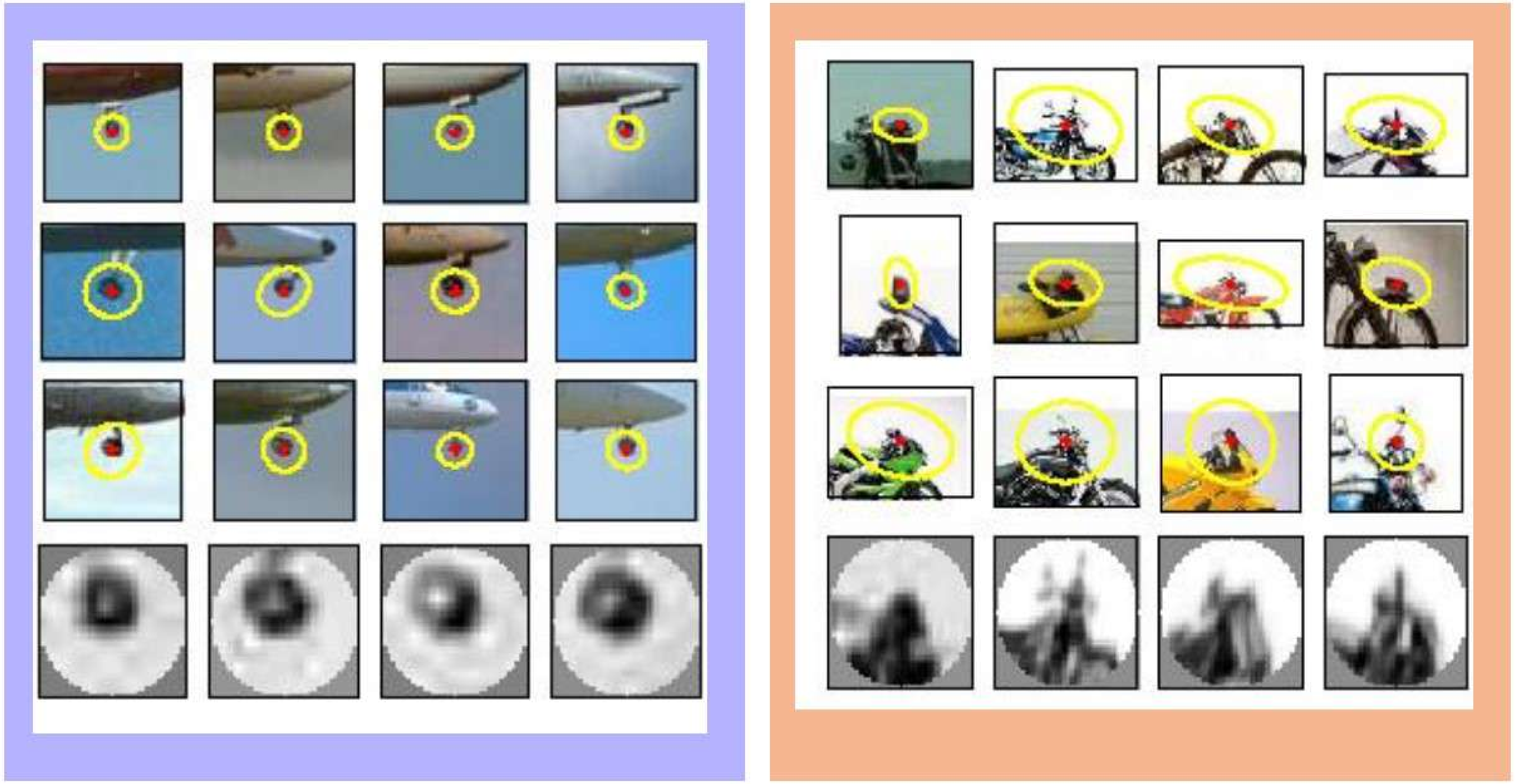


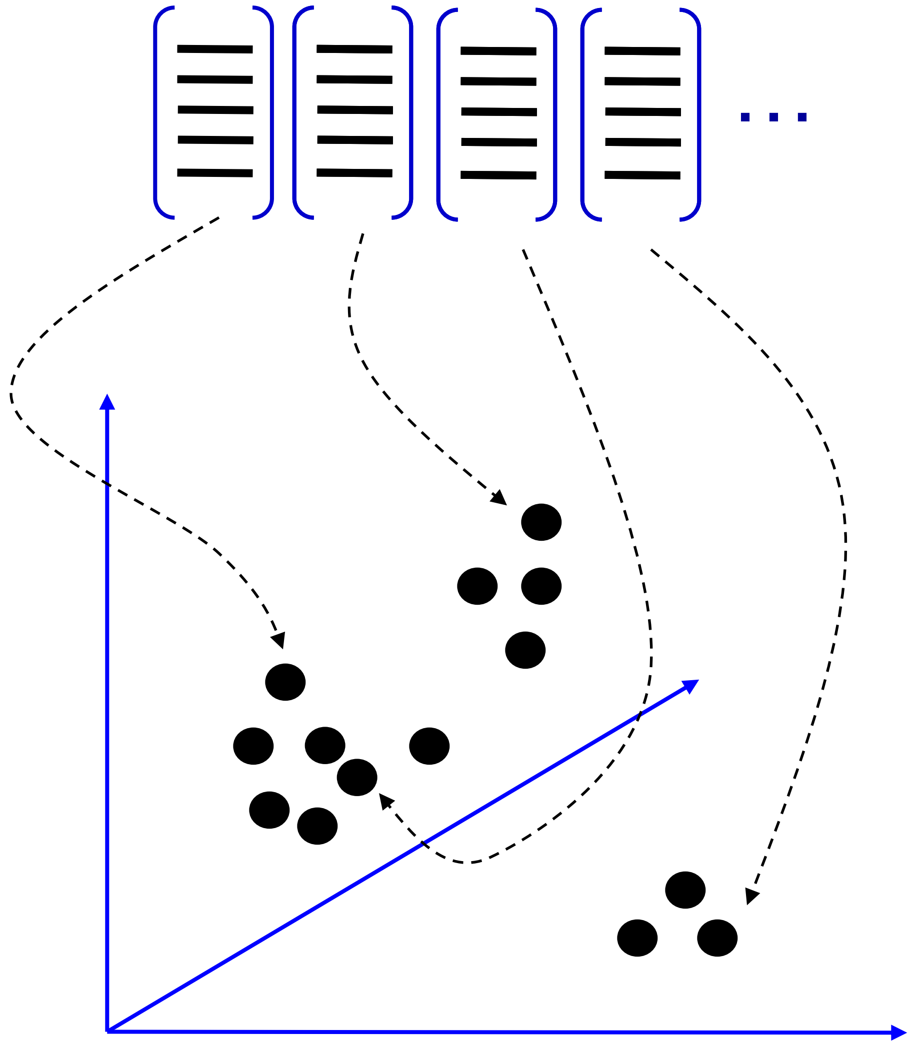
Image patch examples of visual words



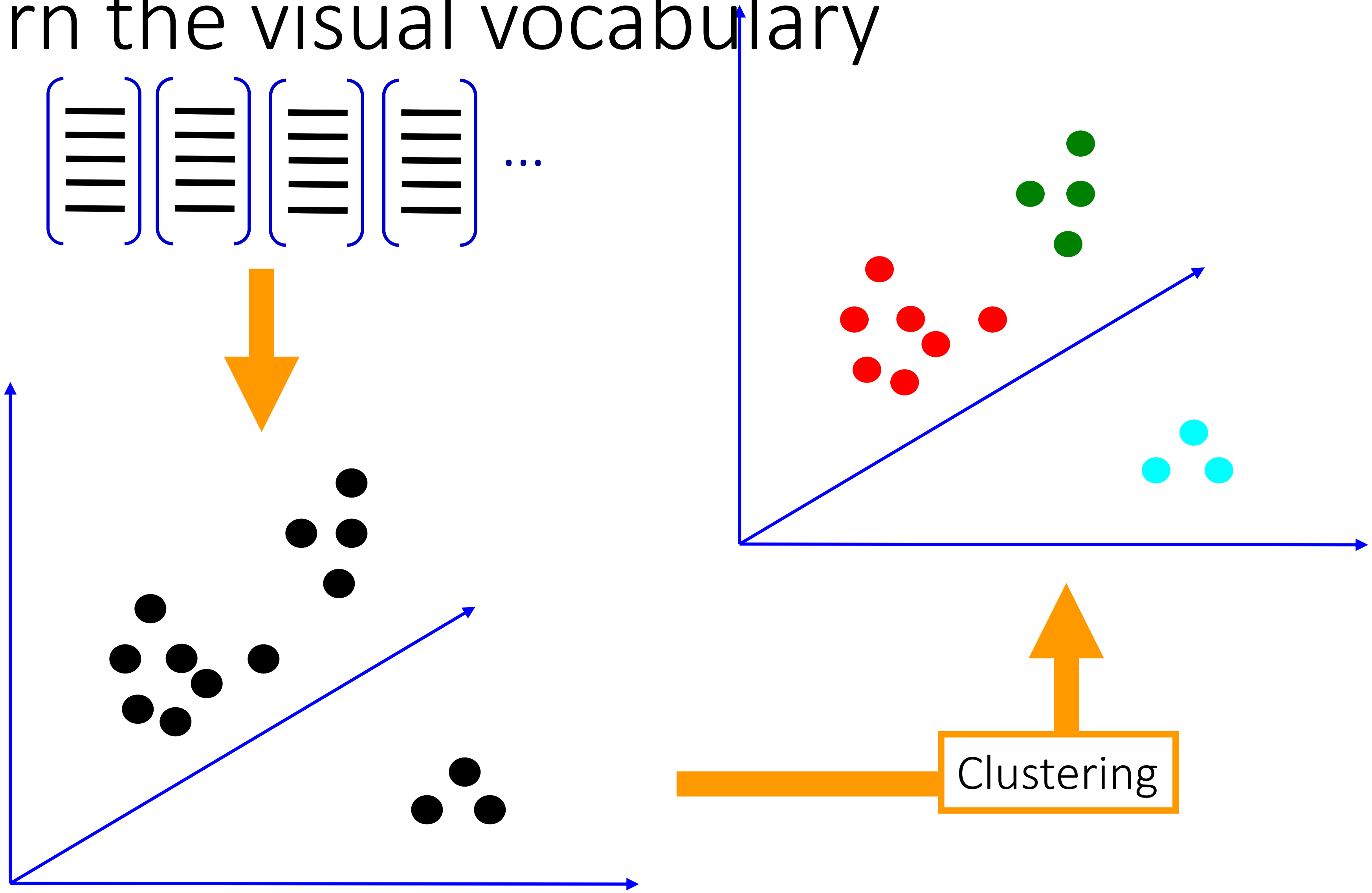
Sivic et al 2005

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

2. Learn the visual vocabulary

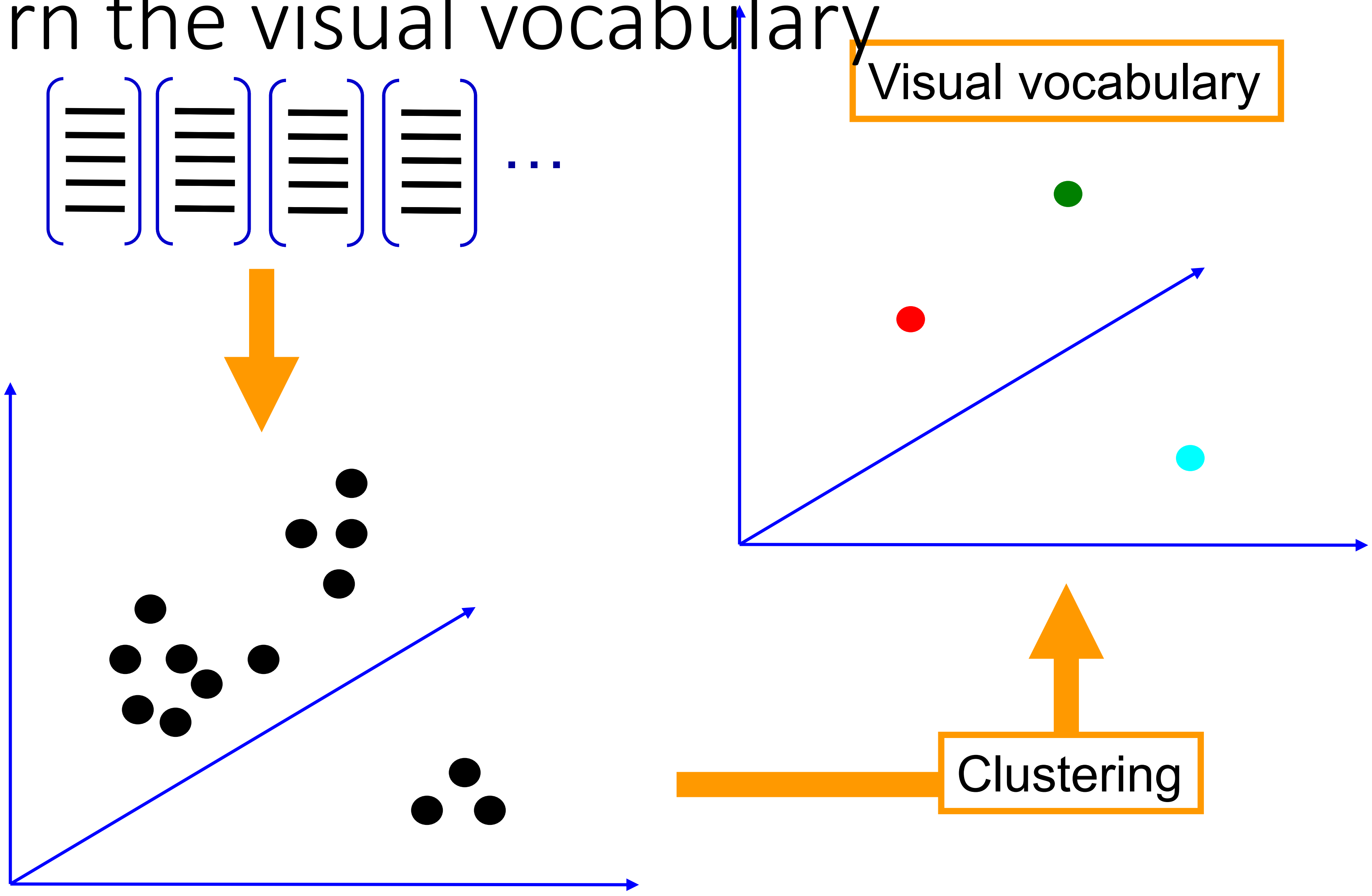


2. Learn the visual vocabulary



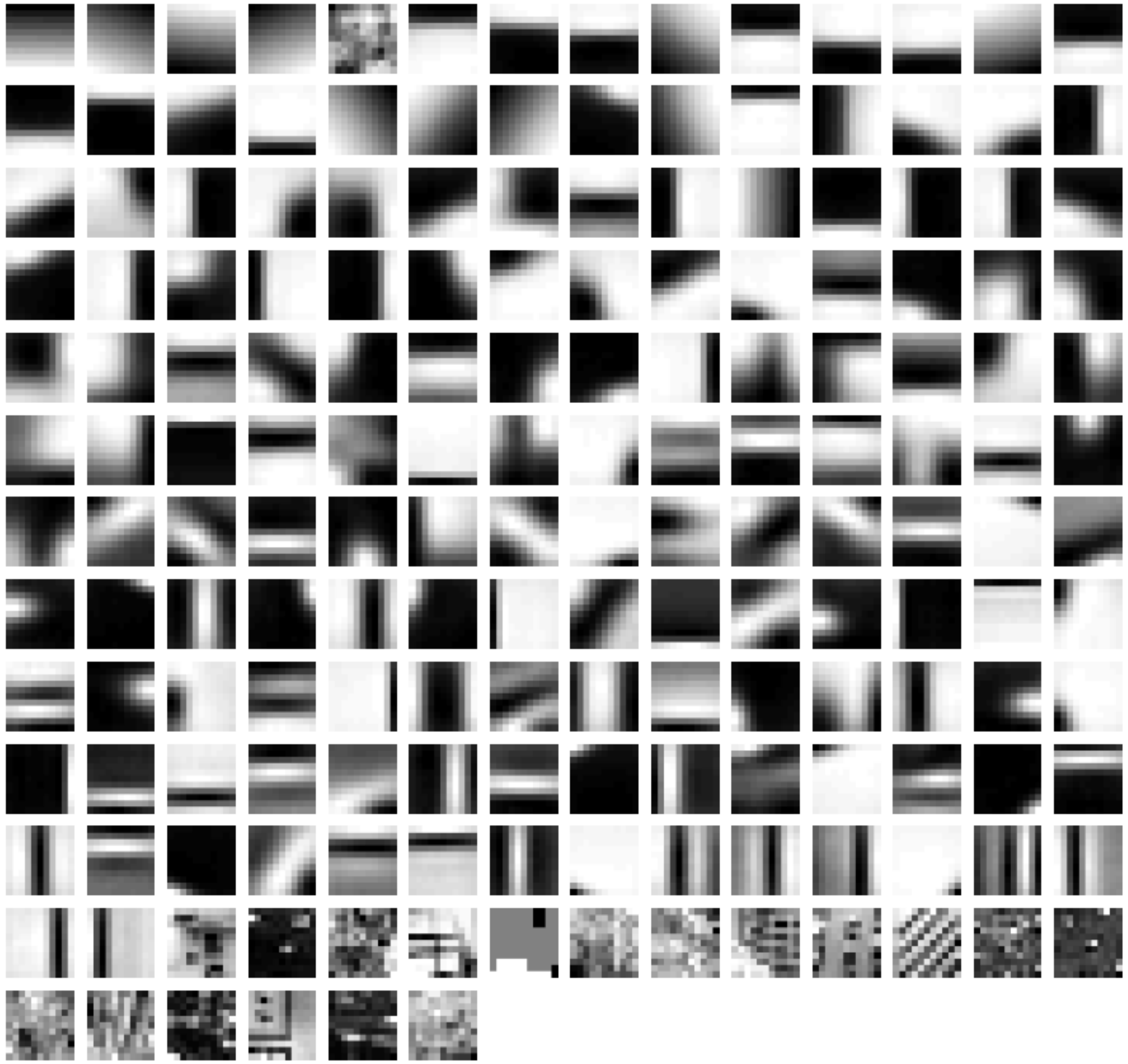
Slide credit: Josef Sivic

2. Learn the visual vocabulary



Slide credit: Josef Sivic

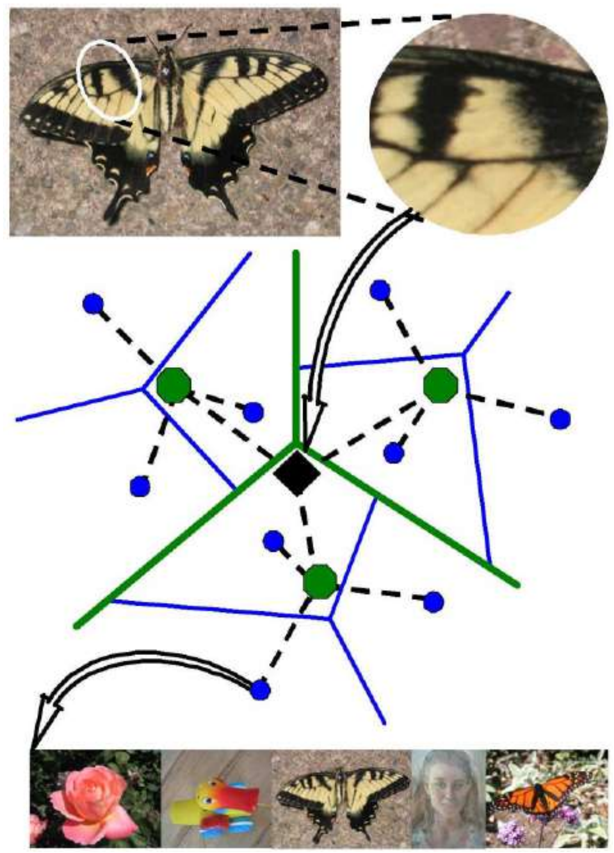
Example visual vocabulary



Fei-Fei et al. 2005

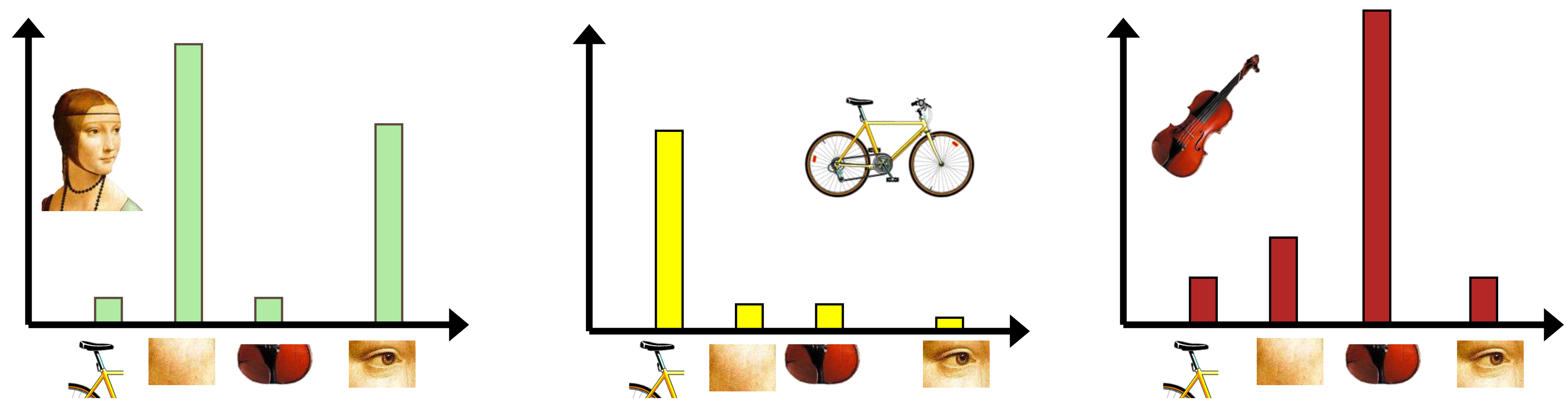
Visual vocabularies: issues

- How to choose vocabulary size?
    - Too small: visual words not representative of all patches
    - Too large: quantization artifacts, overfitting
- Computational efficiency
    - Solution: Vocabulary trees (Nister & Stewenius, 2006)



Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

3. From clustering to vector quantization

- Clustering is a common method for learning a visual vocabulary or codebook
  - Unsupervised learning process
  - Each cluster center produced by k-means becomes a codevector
  - Codebook can be learned on separate training set
  - Provided the training set is sufficiently representative, the codebook will be “universal”

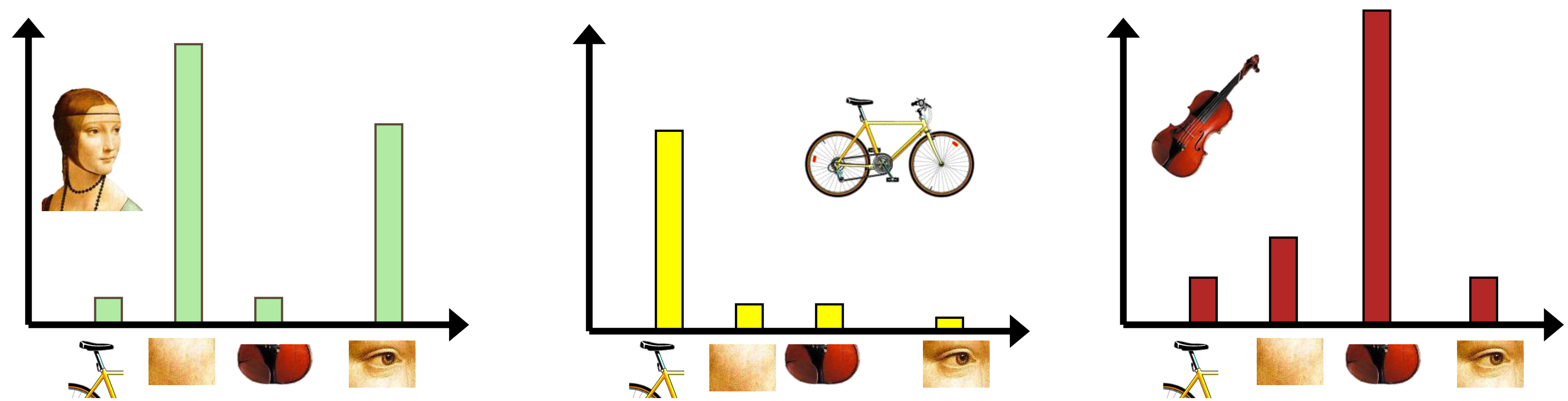
- The codebook is used for quantizing features
  - A *vector quantizer* takes a feature vector and maps it to the index of the nearest codevector in a codebook
  - Codebook = visual vocabulary
  - Codevector = visual word

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of features: outline

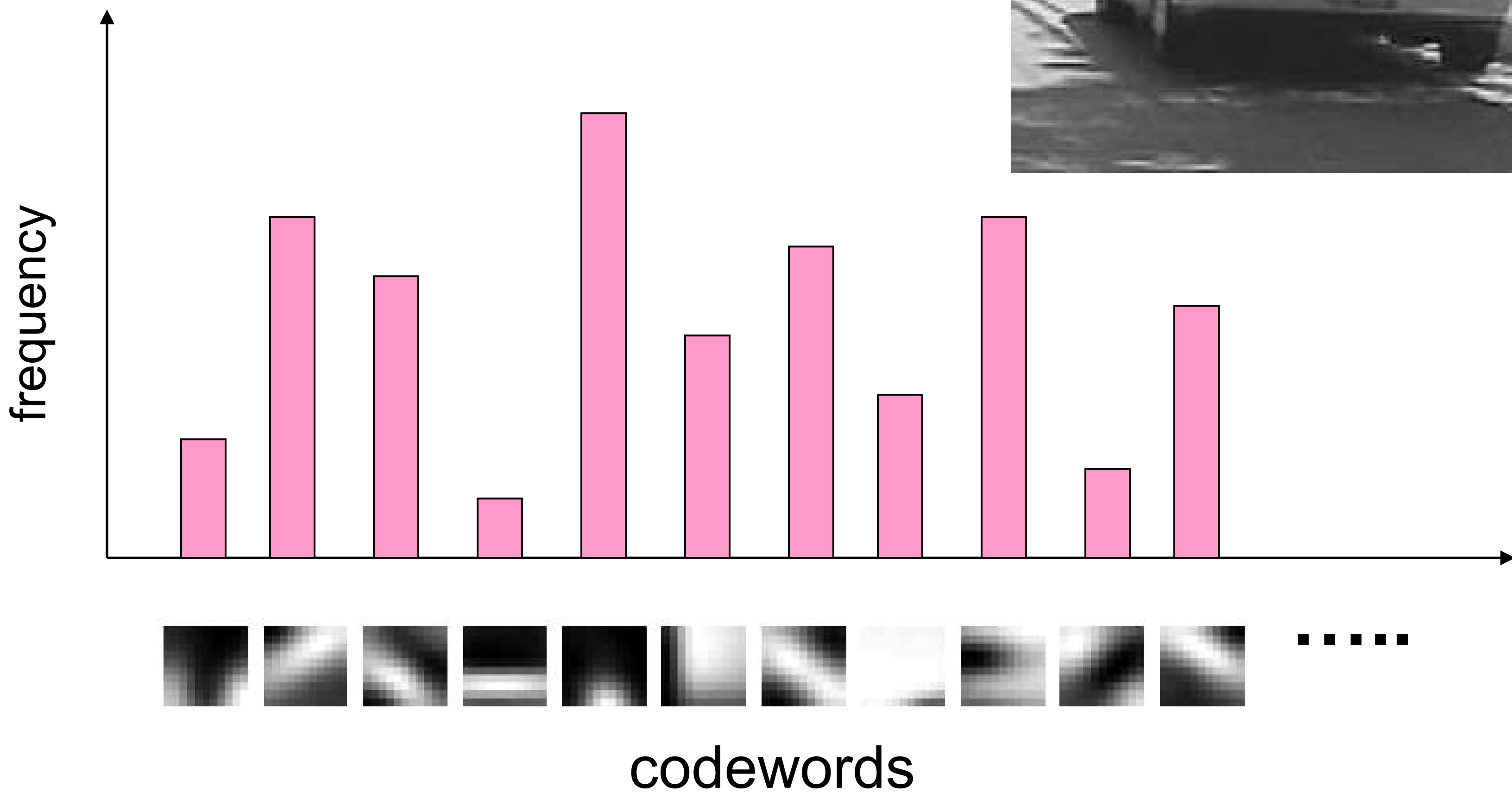
1. Extract features
2. Learn "visual vocabulary"
3. Quantize features using visual vocabulary
4. Represent images by frequencies of "visual words"



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

4. Image representation



Today's Agenda

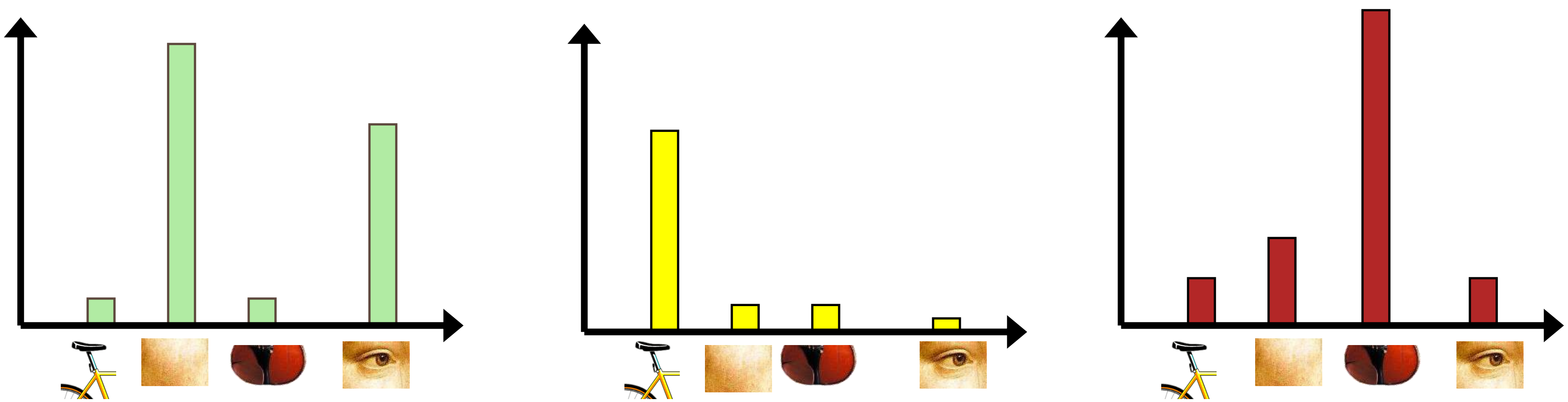
- Visual bag of words (BoW)
  - Background
  - Algorithm
- Applications
  - Image search
- Spatial Pyramid Matching

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Image classification

- Given the bag-of-features representations of images from different classes, how do we learn a model for distinguishing them?



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Uses of BoW representation

- Treat as feature vector for standard classifier
  - e.g k-nearest neighbors, support vector machine
- Cluster BoW vectors over image collection
  - Discover visual themes

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

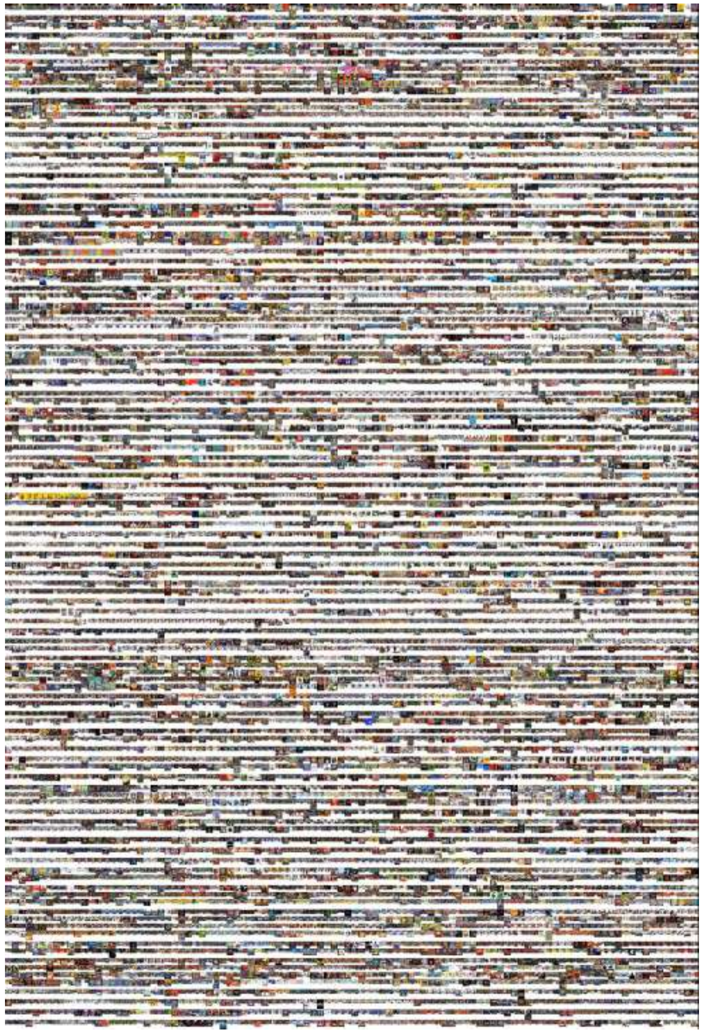
Large-scale image search

11,400 images of game covers
(Caltech games dataset)

Bag-of-words models have been useful in matching an image to a large database of object *instances*

How do I find this image in the database?

Large-scale image search



Build the database:

- Extract features from the database images
- Learn a vocabulary using k-means (typical k: 100,000)
- Compute *weights* for each word
- Create an inverted file mapping words → images

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Weighting the words

- Just as with text, some visual words are more discriminative than others

the, and, or vs. ***cow, AT&T, Cher***

- The bigger fraction of the documents a word appears in, the less useful it is for matching
  - e.g., a word that appears in *all* documents is not helping us

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

TF-IDF weighting

- Instead of computing a regular histogram distance, we'll weight each word by its *inverse document frequency*

- Inverse Document Frequency (IDF) of word *j* =

$$\log \frac{\text{number of documents}}{\text{number of documents in which } j \text{ appears}}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

TF-IDF weighting

- Term Frequency (TF) of word *j is the number of times it appears in the 'document', i.e., the image*

- To compute the value of bin *j* in image *I, compute TF-IDF*:

Term frequency of *j* in *I* **x** *Inverse Document Frequency* of *j*

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Inverted file

- Each image has ~1,000 features
- We have ~100,000 visual words
  →each histogram is extremely sparse (mostly zeros)
- Inverted file
  - mapping from words to 'documents', i.e., images

```
"a":      {2}
"banana": {2}
"is":     {0, 1, 2}
"it":     {0, 1, 2}
"what":   {0, 1}
```


Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

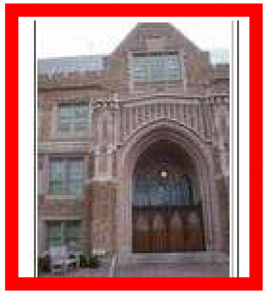
Inverted file

- Can quickly use the inverted file to compute similarity between a new image and all the images in the database
  - Only consider database images whose bins overlap the query image

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Large-scale image search

query image top 6 results

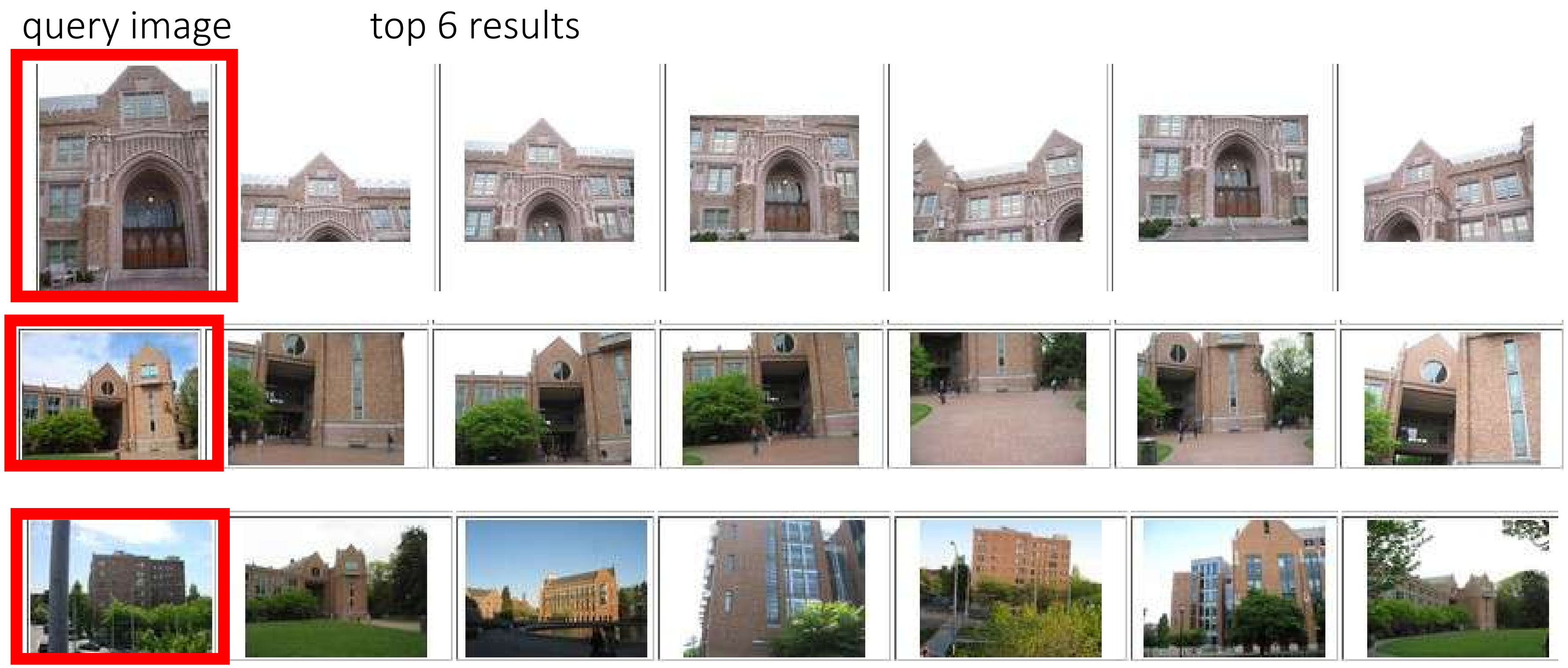


- Cons:
  - performance degrades as the database grows

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Large-scale image search



- Cons:
    - performance degrades as the database grows

Large-scale image search

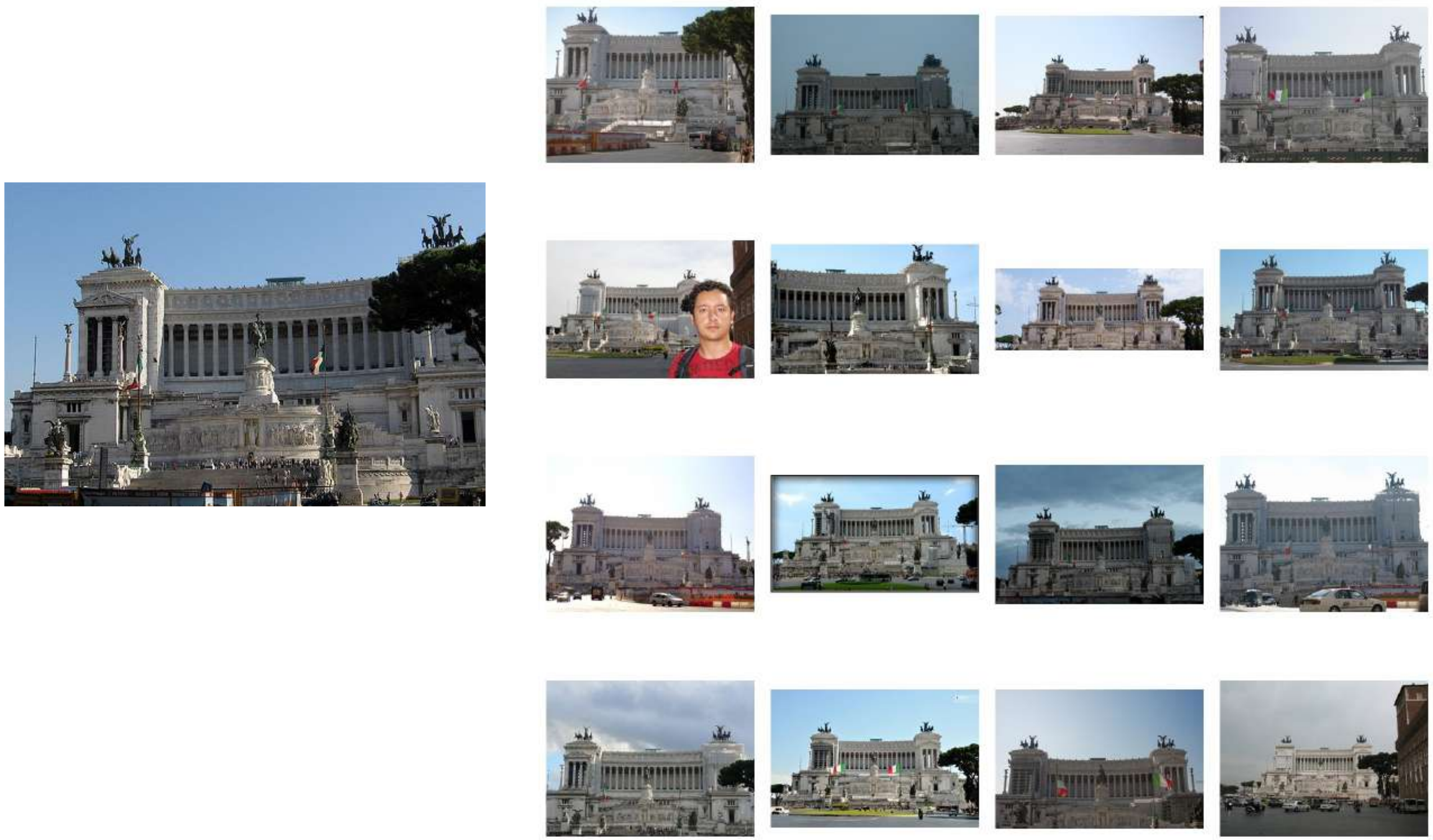
- Pros:
  - Works well for CD covers, movie posters
  - Real-time performance possible

Real-time retrieval from a database of 40,000 CD covers

Nister & Stewenius, **Scalable Recognition with a Vocabulary Tree**

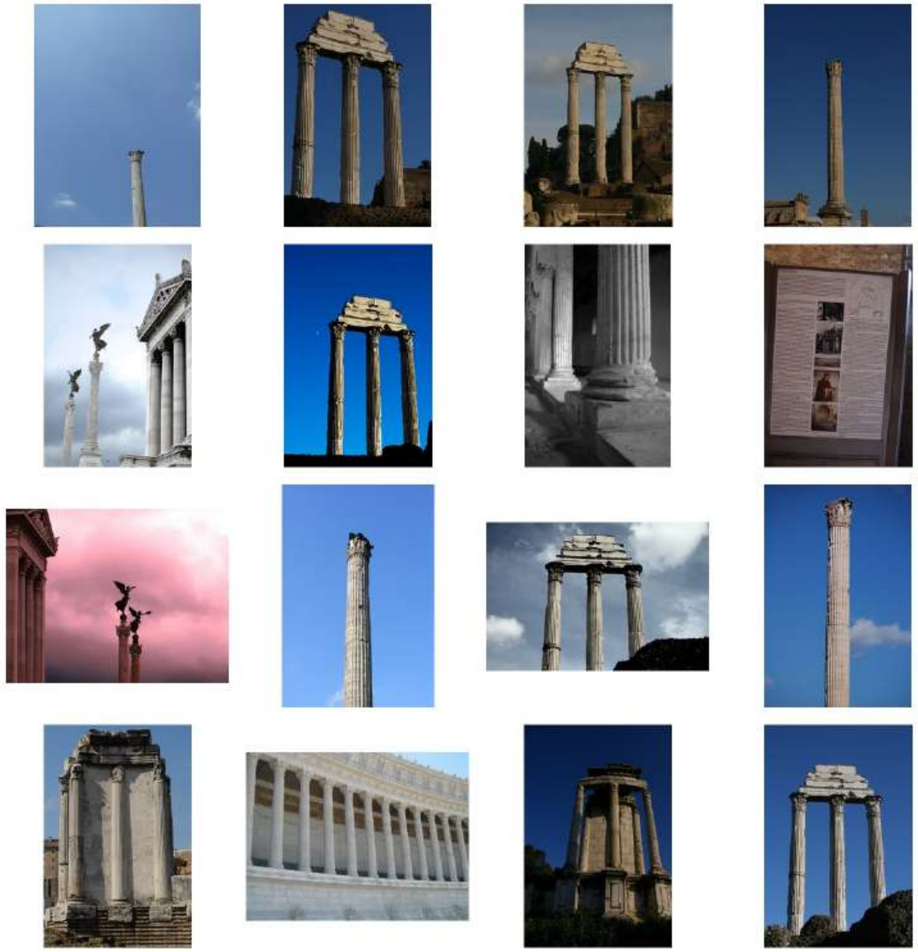
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Example bag-of-words matches



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Example bag-of-words matches



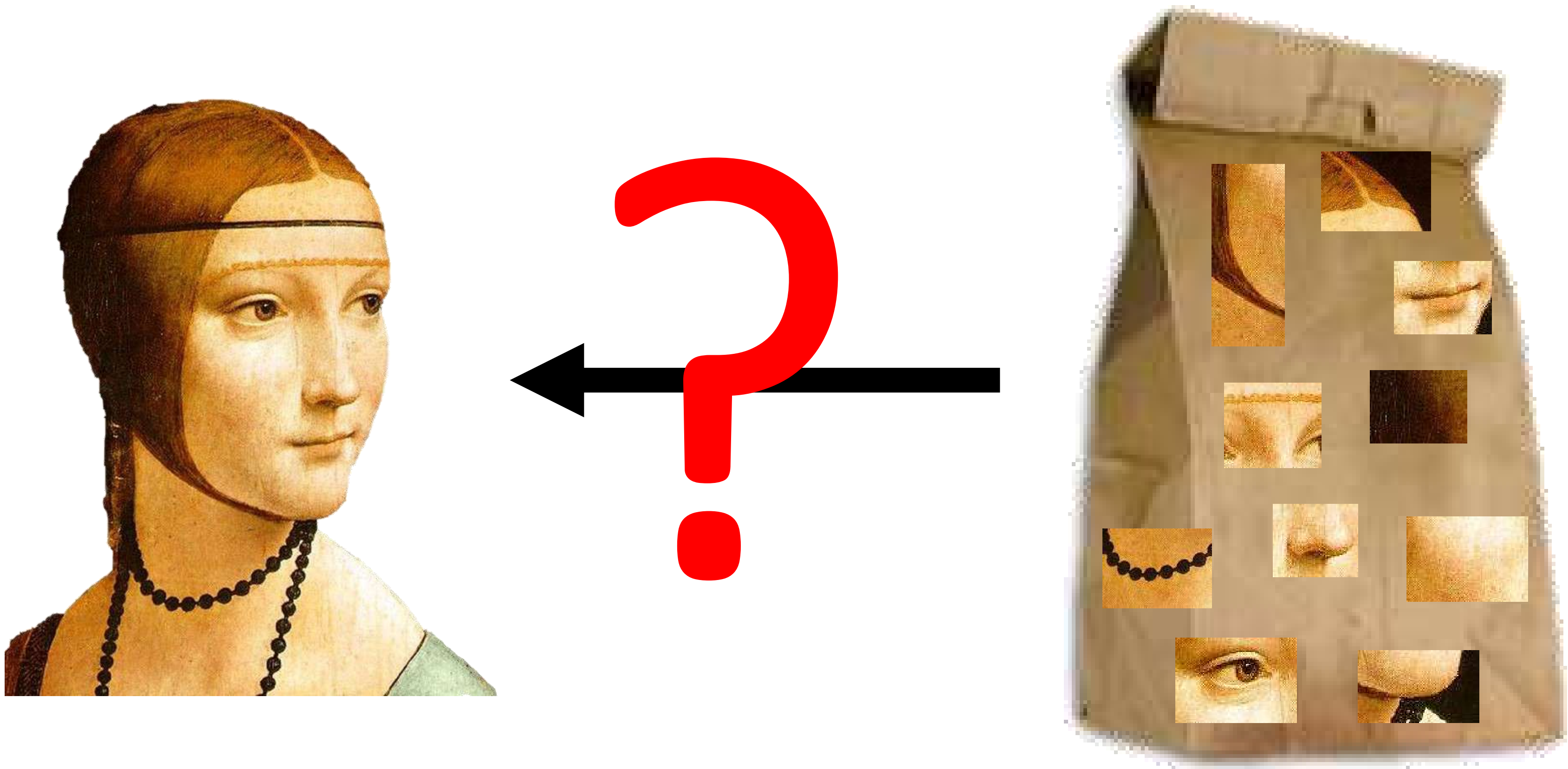
Today's Agenda

- Visual bag of words (BoW)
  - Background
  - Algorithm
- Applications
  - Image search
  - Action recognition
- Spatial Pyramid Matching

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What about spatial info?



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

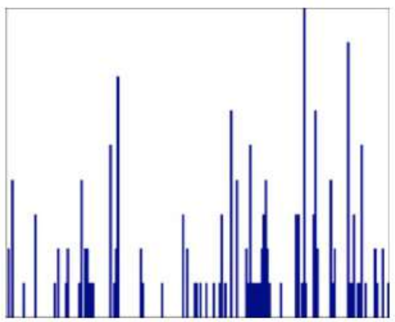
Pyramids

- Very useful for representing images.
- Pyramid is built by using multiple copies of image.
- Each level in the pyramid is 1/4 of the size of previous level.
- The lowest level is of the highest resolution.
- The highest level is of the lowest resolution.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of words + pyramids

Locally orderless representation at several levels of spatial resolution

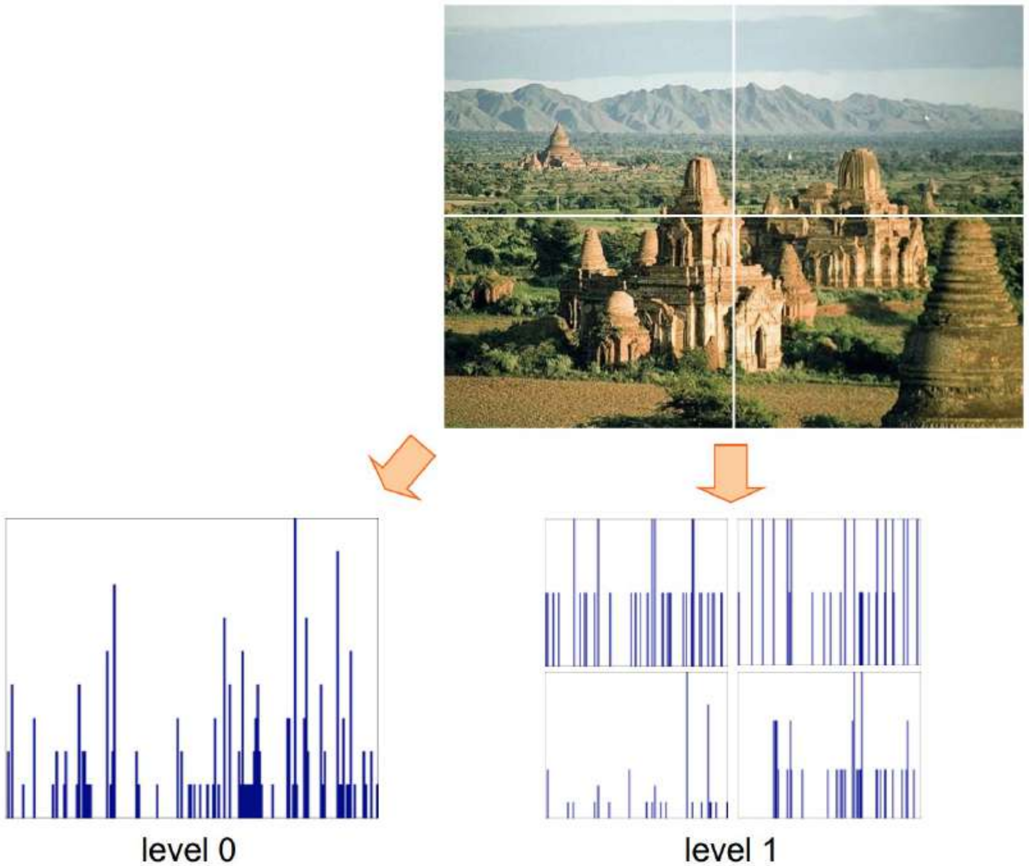


level 0

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

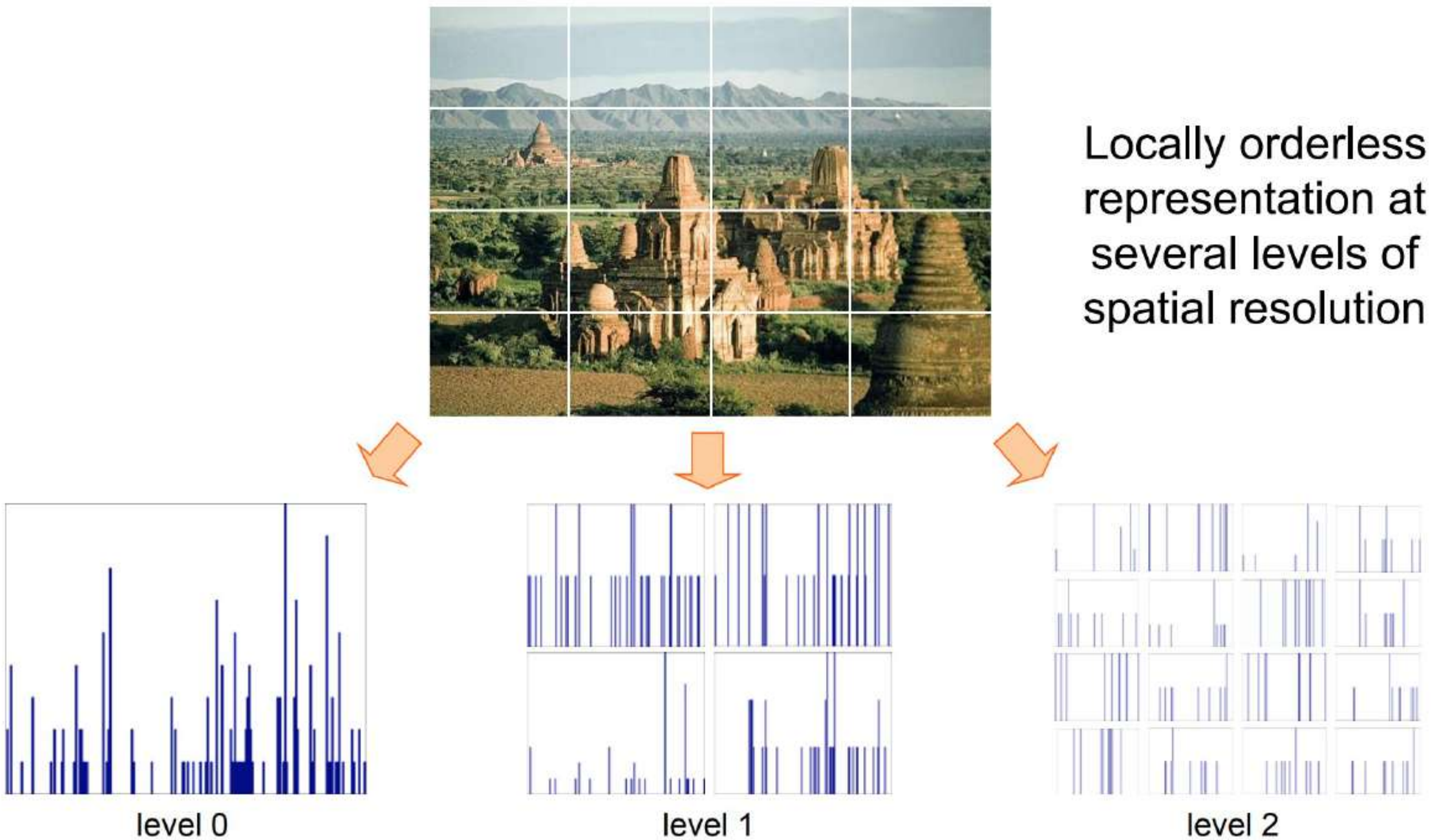
Bag of words + pyramids



Locally orderless representation at several levels of spatial resolution

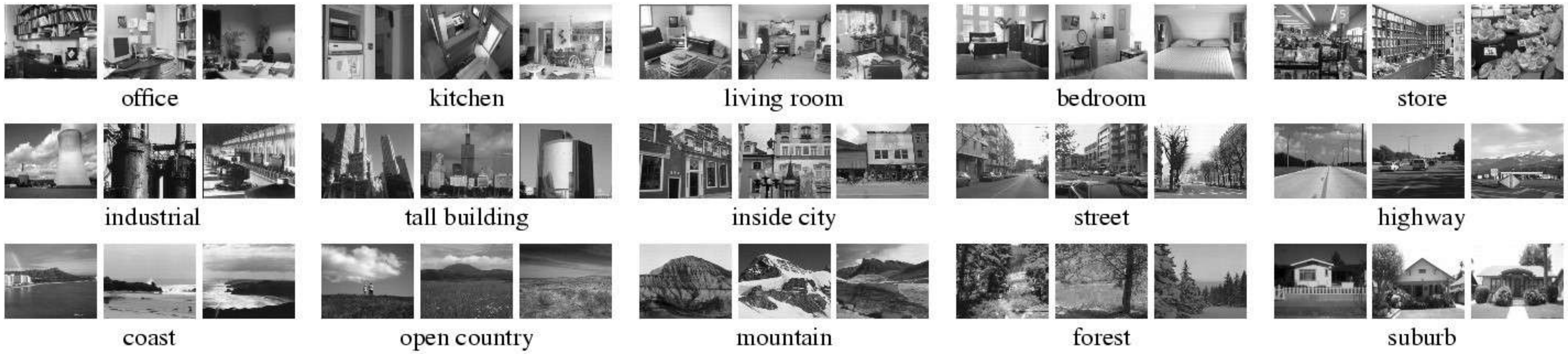
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bag of words + pyramids



Locally orderless representation at several levels of spatial resolution

Results: Scene category dataset



Multi-class classification results (100 training images per class)

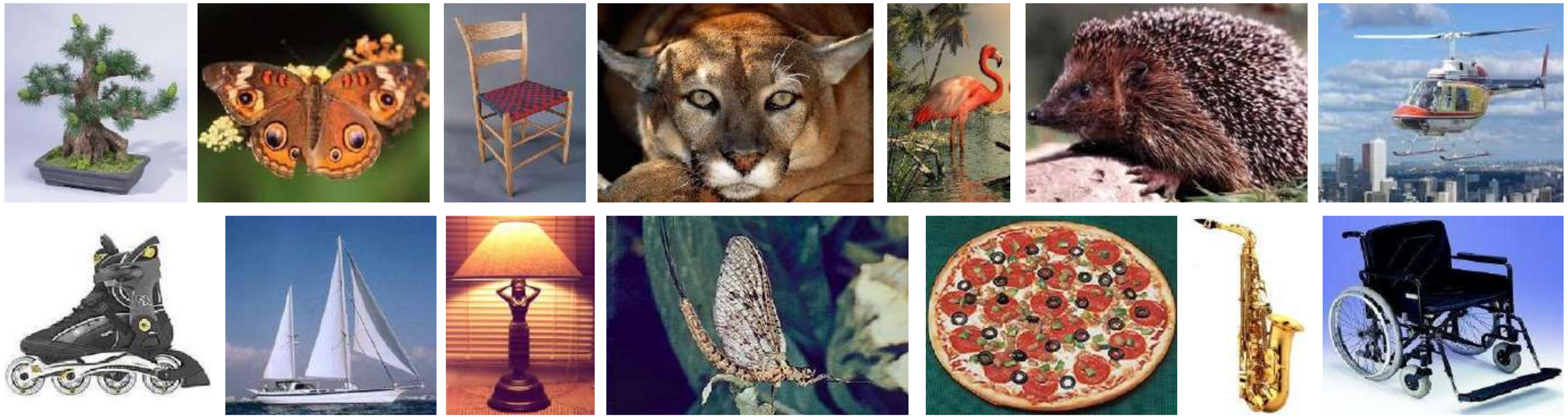
	Weak features (vocabulary size: 16)		Strong features (vocabulary size: 200)	
Level	Single-level	Pyramid	Single-level	Pyramid
$0\ (1 \times 1)$	45.3 ± 0.5		72.2 ± 0.6	
$1\ (2 \times 2)$	53.6 ± 0.3	56.2 ± 0.6	77.9 ± 0.6	79.0 ± 0.5
$2\ (4 \times 4)$	61.7 ± 0.6	64.7 ± 0.7	79.4 ± 0.3	$\mathbf{81.1} \pm 0.3$
$3\ (8 \times 8)$	63.3 ± 0.8	$\mathbf{66.8} \pm 0.6$	77.2 ± 0.4	80.7 ± 0.3

Lazebnik, Schmid & Ponce (CVPR 2006)

Slide credit: Svetlana Lazebnik

Results: Caltech101 dataset

http://www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html



Multi-class classification results (30 training images per class)

	Weak features (16)		Strong features (200)	
Level	Single-level	Pyramid	Single-level	Pyramid
0	15.5 ± 0.9		41.2 ± 1.2	
1	31.4 ± 1.2	32.8 ± 1.3	55.9 ± 0.9	57.0 ± 0.8
2	47.2 ± 1.1	49.3 ± 1.4	63.6 ± 0.9	$\mathbf{64.6} \pm 0.8$
3	52.2 ± 0.8	$\mathbf{54.0} \pm 1.1$	60.3 ± 0.9	64.6 ± 0.7

Lazebnik, Schmid & Ponce (CVPR 2006)

Slide credit: Svetlana Lazebnik

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 13: Object Detection

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union

Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Visual bag of words (BoW)
  - Background
  - Algorithm
- Applications
  - Image search
- Spatial Pyramid Matching

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Object detection
  - Task definition
  - Benchmarks
  - Evaluation
- A simple object detector

[material based on Niebles-Krishna]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

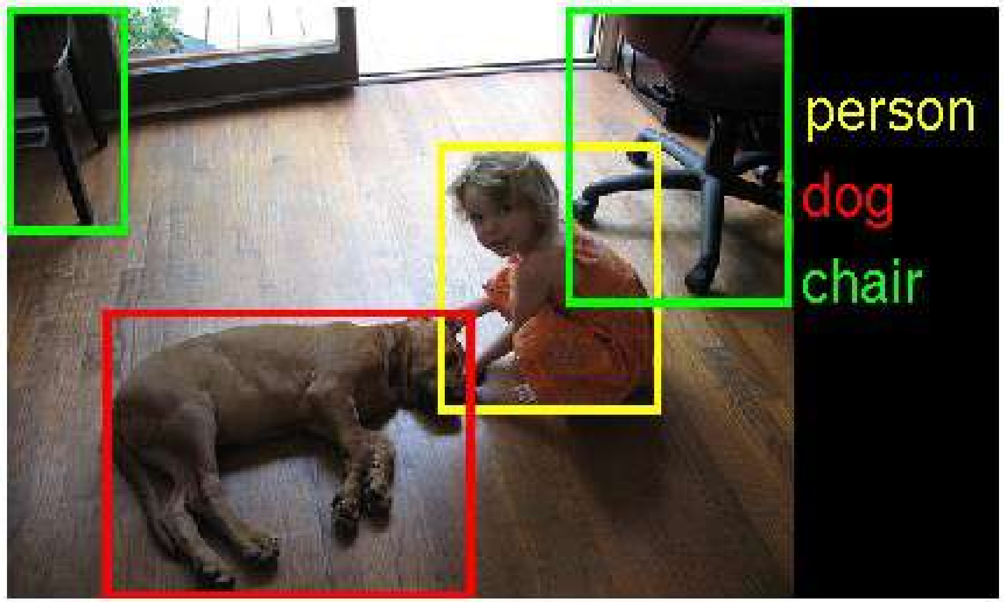
Today's Agenda

- Object detection
  - Task definition
  - Benchmarks
  - Evaluation
- A simple object detector

Object Detection

- **Problem**: Detecting and localizing generic objects from various categories, such as cars, people, etc.

- Challenges:
  - Illumination,
  - viewpoint,
  - deformations,
  - Intra-class variability



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

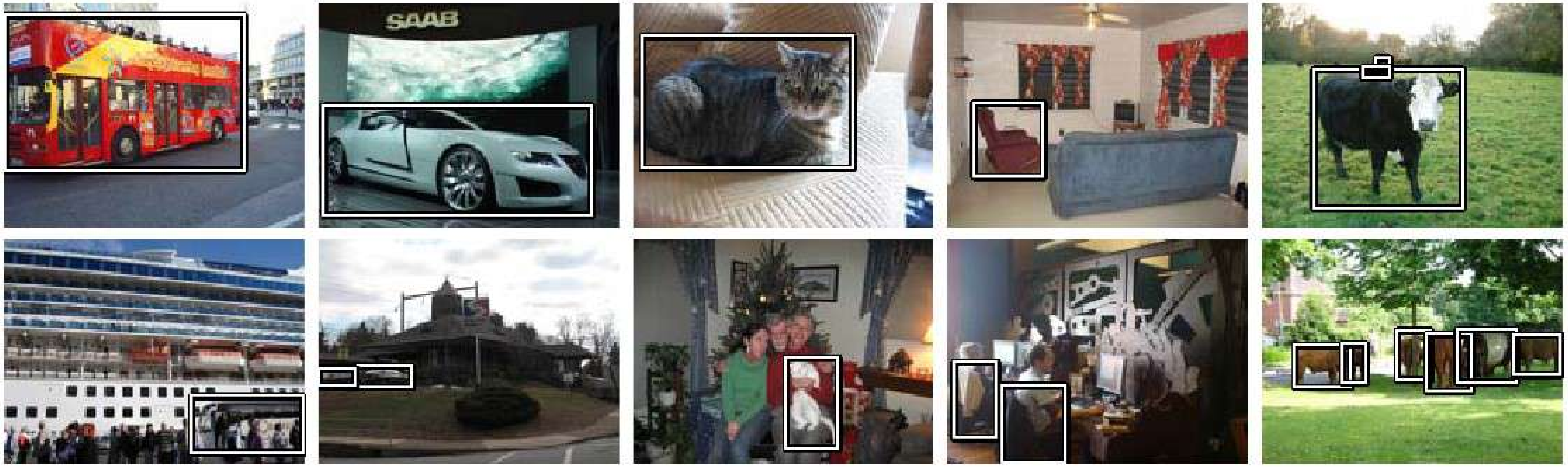
Today’s Agenda

- Object detection
  - Task definition
  - Benchmarks
  - Evaluation
- A simple object detector

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Object Detection Benchmarks

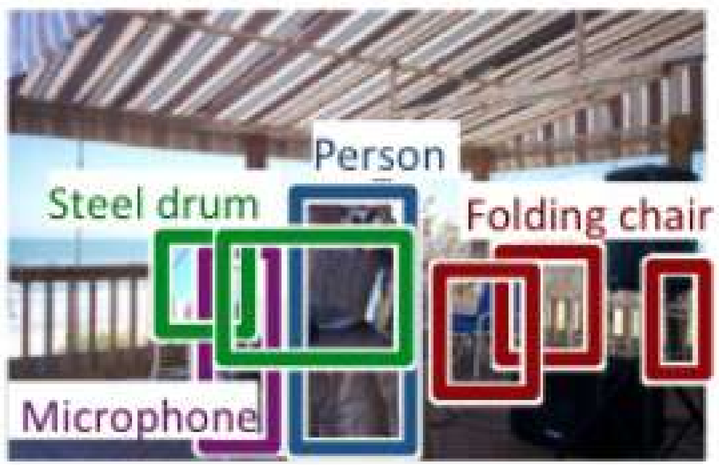
- PASCAL VOC Challenge



- 20 categories
- Annual classification, detection, segmentation challenges

Object Detection Benchmarks

- PASCAL VOC Challenge
- ImageNet Large Scale Visual Recognition Challenge (ILSVR)
  - 200 Categories for detection



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Object Detection Benchmarks

- PASCAL VOC Challenge
- ImageNet Large Scale Visual Recognition Challenge (ILSVR)
- Common Objects in Context (COCO)
  - 80 Object categories

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Object detection
  - Task definition
  - Benchmarks
  - Evaluation
- A simple object detector
- Deformable parts model
  - Overview
  - Method
  - Pipeline
  - Results and analysis

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

How do we evaluate object detection?

How do we evaluate object detection?

predictions

ground truth

True positive:

- The overlap of the prediction with the ground truth is MORE than 0.5

How do we measure overlap ?

Intersection over Union (IoU)

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

How do we evaluate object detection?

predictions

ground truth

True positive:

False positive:

- The overlap of the prediction with the ground truth is LESS than 0.5

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How do we evaluate object detection?

predictions

ground truth

True positive:

False positive:

False negative:

- The objects that our model doesn't find

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How do we evaluate object detection?

predictions

ground truth

True positive:

False positive:

False negative:

- The objects that our model doesn't find

What is a True Negative?

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

	Predicted 1	Predicted 0
True 1	TP	FN
True 0	FP	TN

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

	Predicted 1	Predicted 0
True 1	TP	FN
True 0	FP	TN

	Predicted 1	Predicted 0
True 1	hits	misses
True 0	false alarms	correct rejections

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

	Predicted 1	Predicted 0
True 1	true positive	false negative
True 0	false positive	true negative

	Predicted 1	Predicted 0
True 1	TP	FN
True 0	FP	TN

	Predicted 1	Predicted 0
True 1	hits	misses
True 0	false alarms	correct rejections

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How do we evaluate object detection?

predictions

ground truth

True positive: 1
False positive: 2
False negative: 1

So what is the
- precision?
- recall?

Precision versus recall

- Precision:
  - how many of the object detections are correct?

- Recall:
  - how many of the ground truth objects can the model detect?

In reality, our model makes a lot of predictions with varying scores between 0 and 1

predictions

ground truth

Here are all the boxes that are predicted with score > 0.

This means that our:
- Recall is perfect!
- But our precision is BAD!

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

In reality, our model makes a lot of predictions with varying scores between 0 and 1

predictions

ground truth

There are no boxes that are predicted with score = 1.

This means that our
- Precision is undefined!
- And our recall is BAD!

How do we evaluate object detection?

predictions

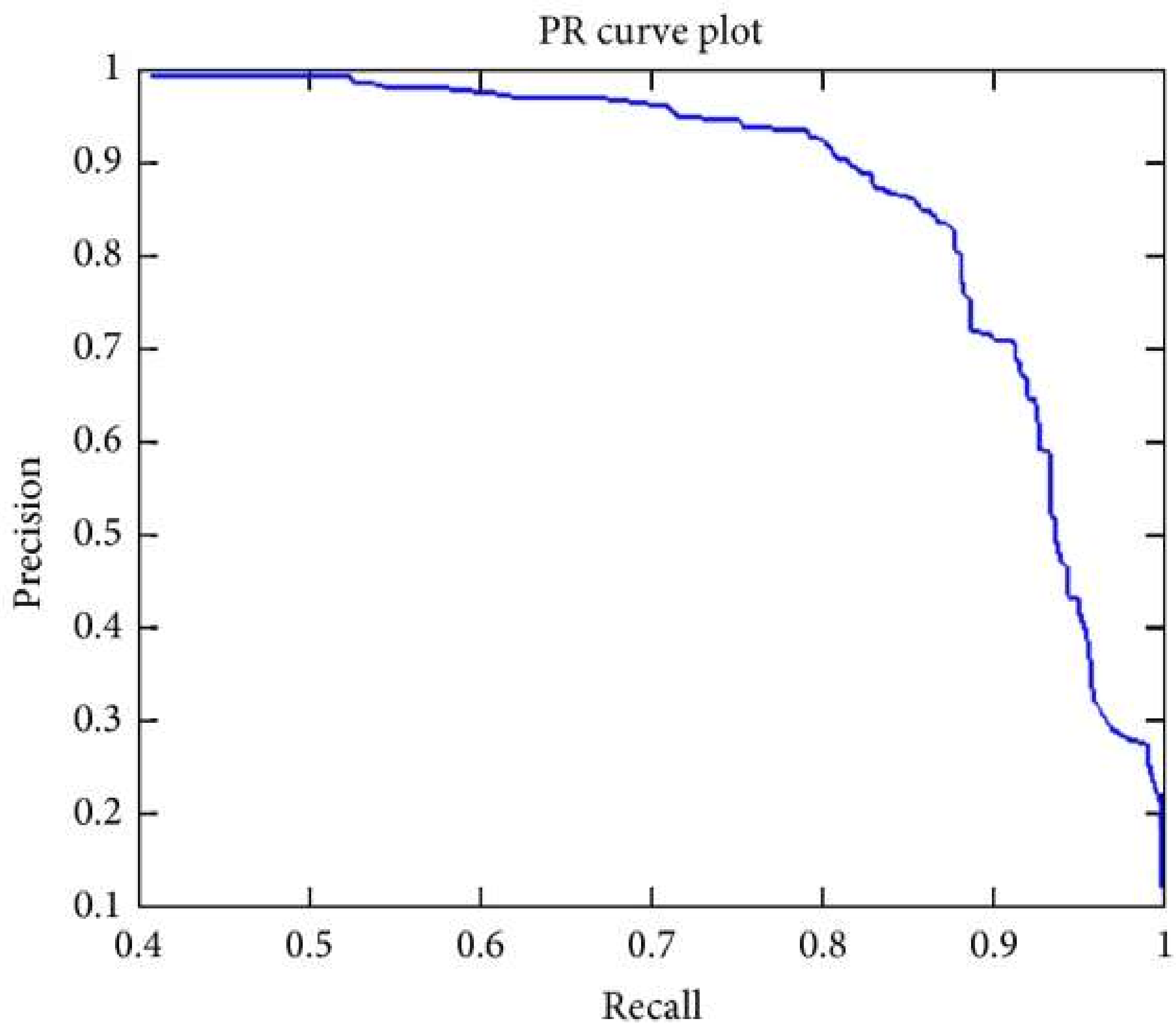
ground truth

Here are all the boxes that are predicted with score > 0.5

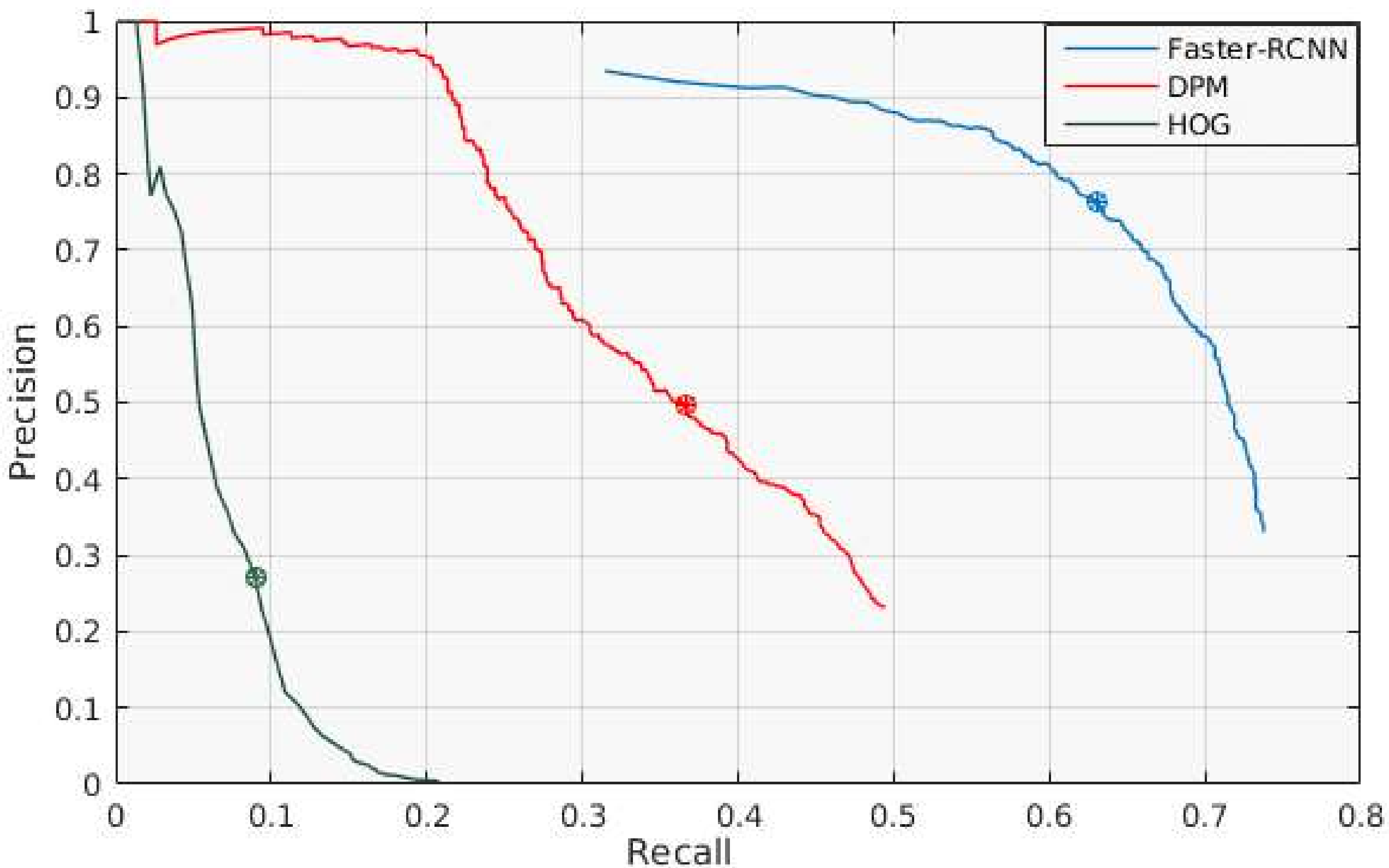
We are setting a threshold of 0.5

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Precision – recall curve (PR curve)



Which model is the best?



True Positives - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

NECUIUC_CLS-DTCT

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

False Positives - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

NECUIUC_CLS-DTCT

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

"Near Misses" - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

NECUIUC_CLS-DTCT

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

True Positives - Bicycle

UoCTTI_LSVM-MDPM

OXFORD_MKL

NECUIUC_CLS-DTCT

False Positives - Bicycle

UoCTTI_LSVM-MDPM

OXFORD_MKL

NECUIUC_CLS-DTCT

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Object detection
  - Task definition
  - Benchmarks
  - Evaluation
- A simple object detector

Co-financed by the European Union
Connecting Europe Facility

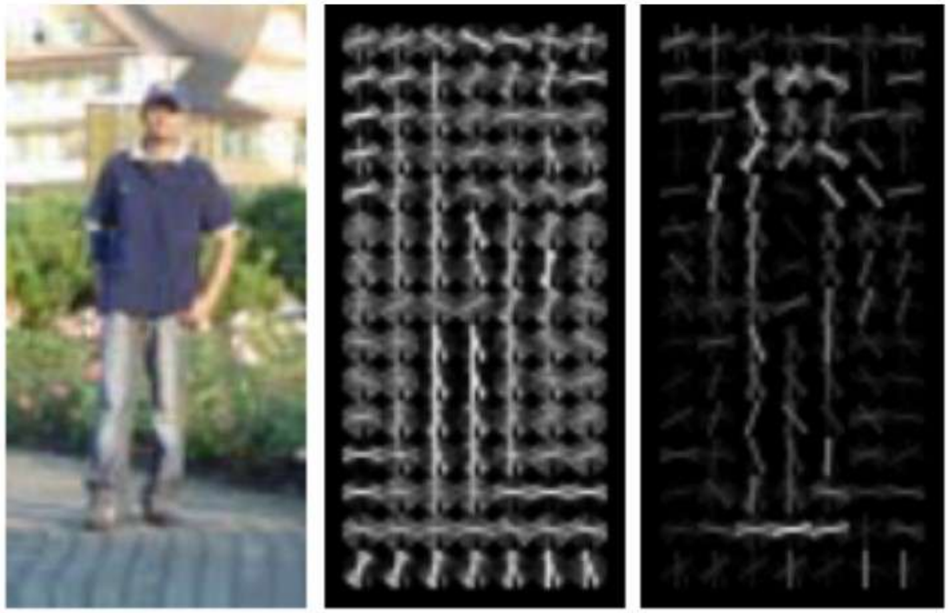
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

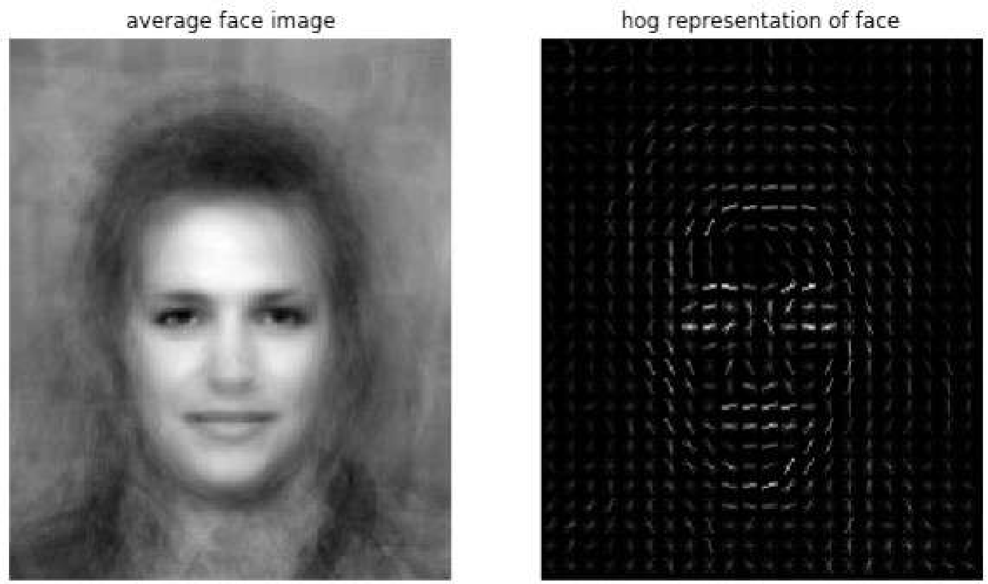
Dalal-Triggs method

sliding window

Recap – HoG features

Find a HoG template and use as filter





Train a linear SVM classifier on HoG

Take the average image and extract HoG

Sliding window + HoG features

No person here

- Slide through the image and check if there is an object at every location
- Compare HOG feature template to HOG features from each location in the image.
  - Use dot product

Sliding window + HoG features

YES!! Person match found

- Slide through the image and check if there is an object at every location
- Compare HOG feature template to HOG features from each location in the image.
- If a comparison produces a high score, output detection at the corresponding location

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Sliding window + HoG features

No bus found

- But what if we were looking for buses?

Sliding window + HoG features

No bus found

- But what if we were looking for buses?

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Sliding window + HoG features

No bus found

- We will never find the object if we don't choose our window size wisely!

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

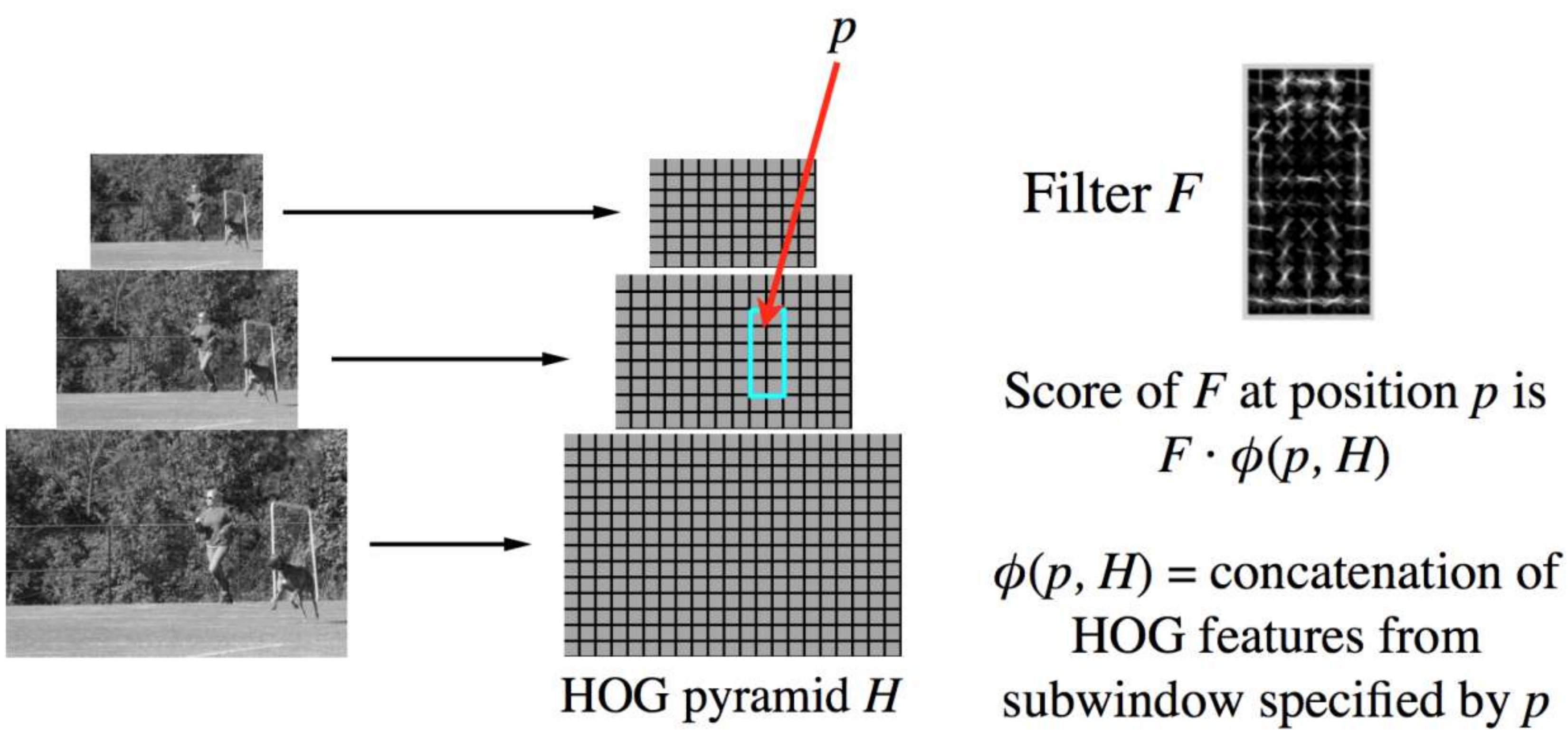
Sliding window + HoG features

We need do a multi scale sliding window search

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Create a feature pyramid



Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 14: Camera Models

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Object detection
  - Task definition
  - Benchmarks
  - Evaluation
- A simple object detector

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Perspective projection
- Vanishing points
- Full camera model

[material based on Paris Kaimakis]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

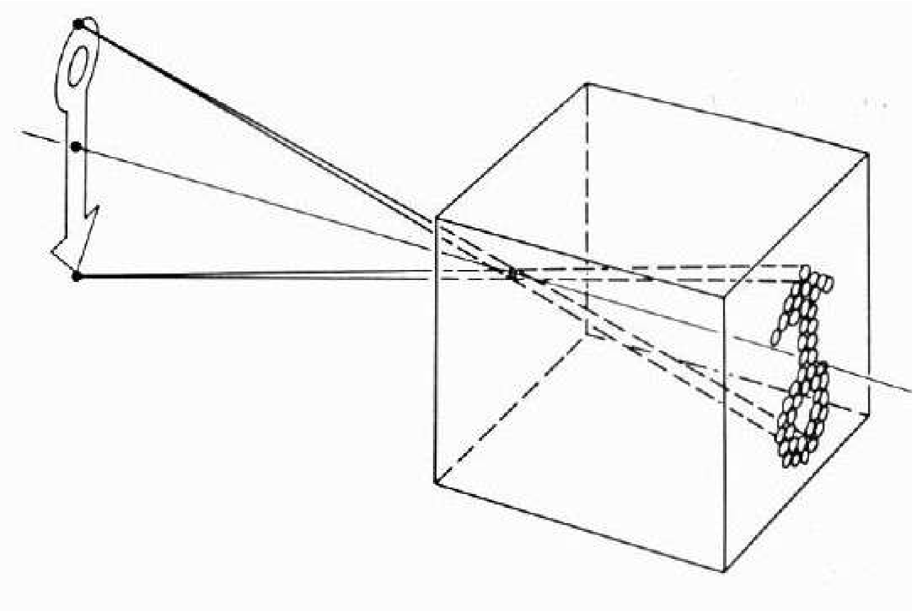
Today's Agenda

- Perspective projection
- Vanishing points
- Full camera model

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

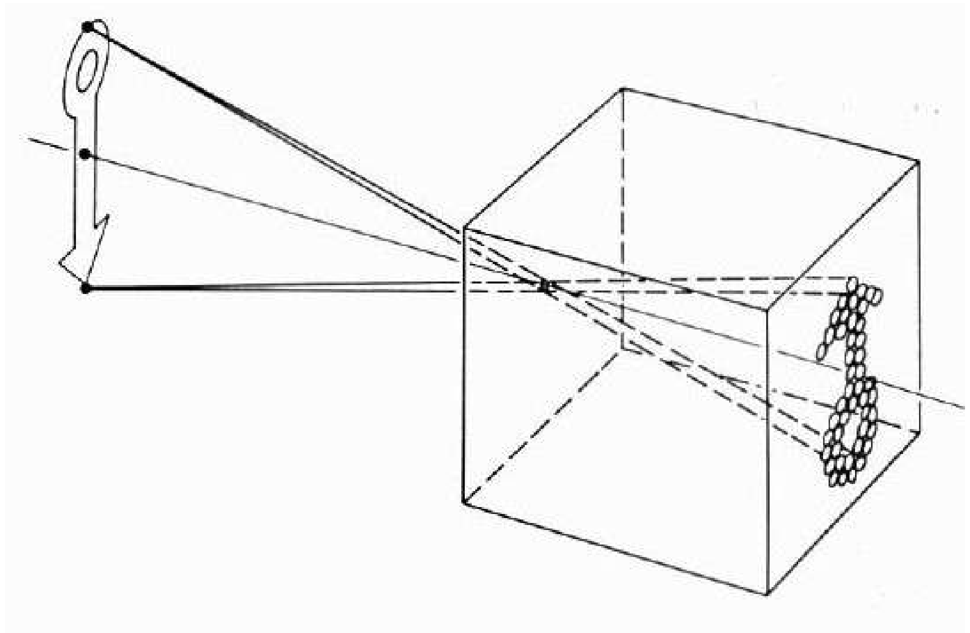
Perspective Projection

- The process of capturing 2D images of the 3D world is modelled by **perspective projection**
- We model perspective projection using the **pin-hole camera**



Pinhole Camera model

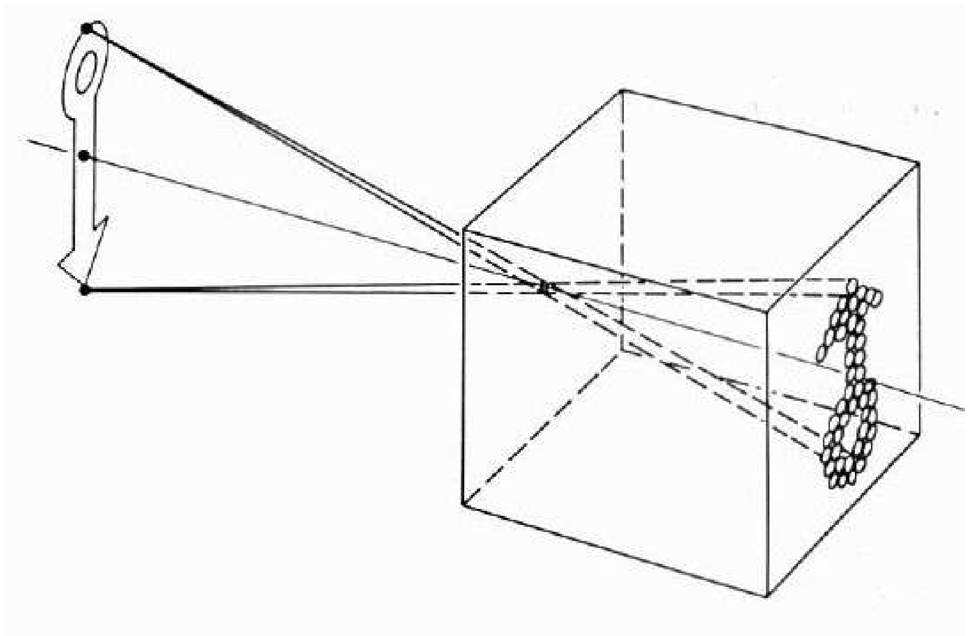
- Captures **pencil of rays** – all rays through a single point
- The point is called **Center of Projection**
- The image is formed on the **Image Plane**



[Slide by Steve Seitz]

Pinhole Camera - Camera Obscura

The first camera
- Known to Aristotle
- How does the aperture size affect the image?

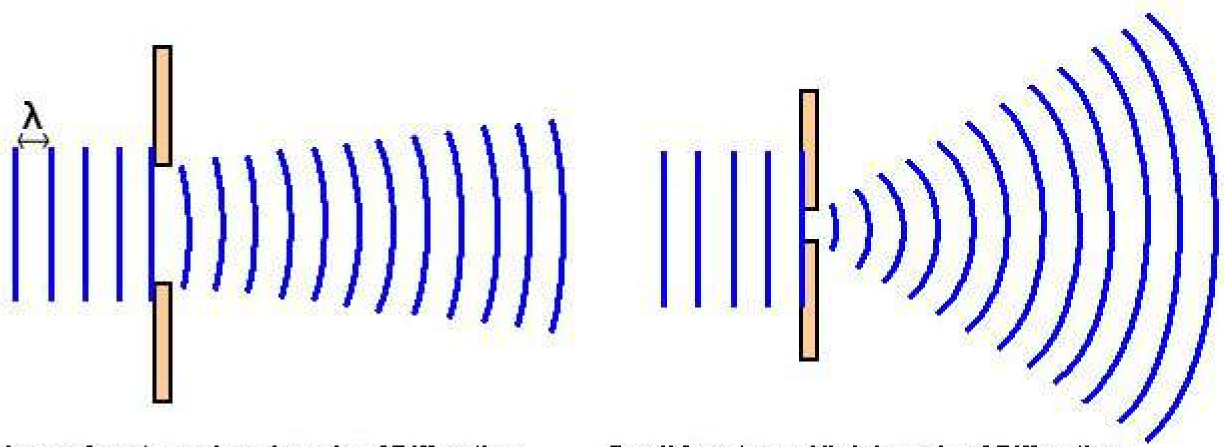


[Slide by Steve Seitz]

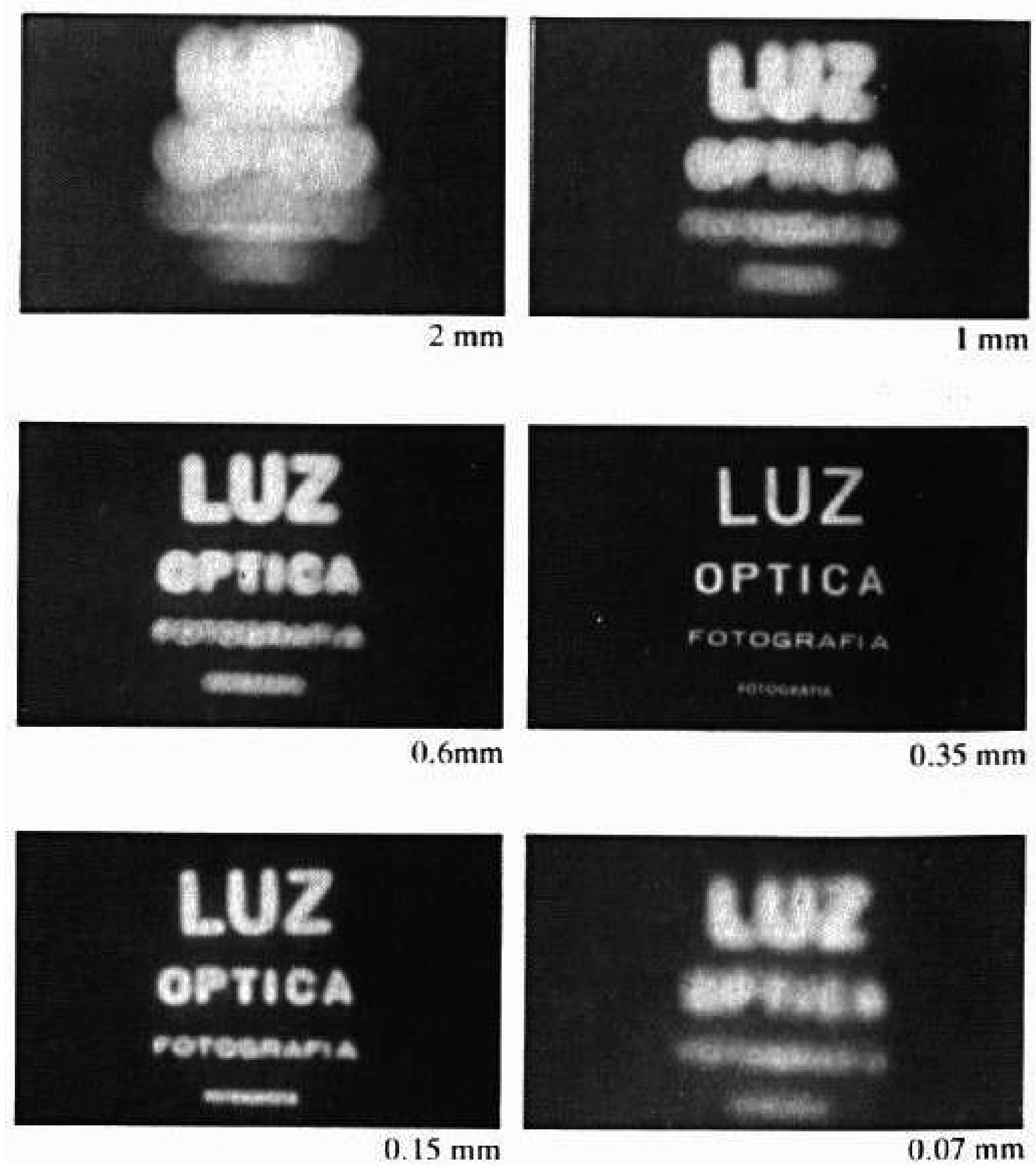
Shrinking the aperture

Why not make the aperture as small as possible?

- Less light gets through
- Diffraction effects



Large Aperture - Low Levels of Diffraction Small Aperture - High Levels of Diffraction

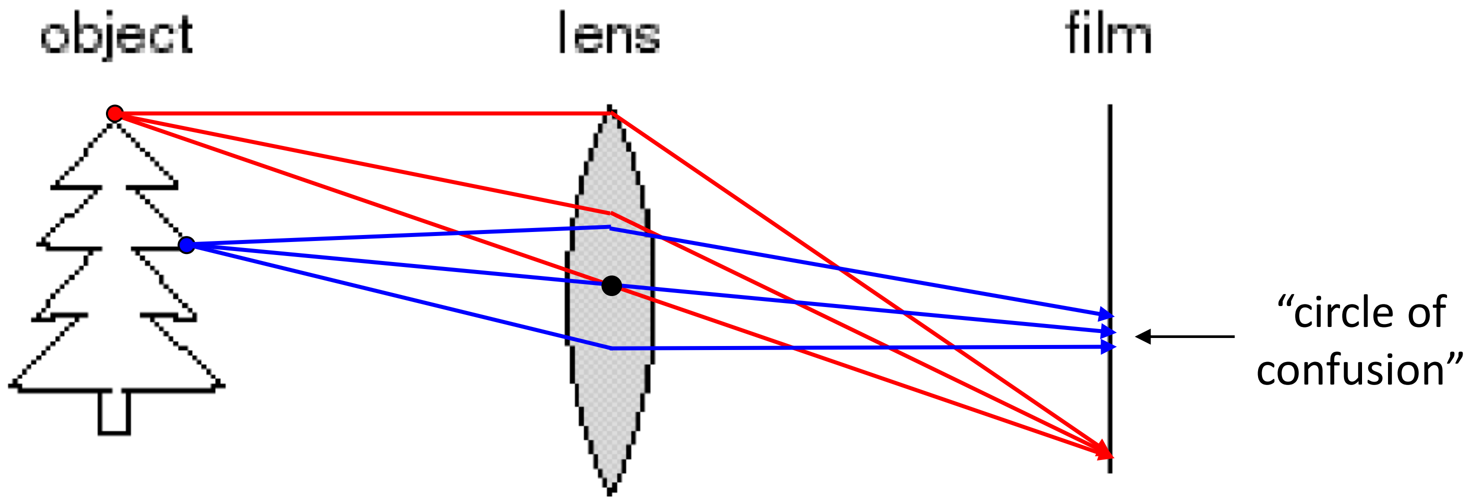


[Slide by Steve Seitz]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Adding a lens

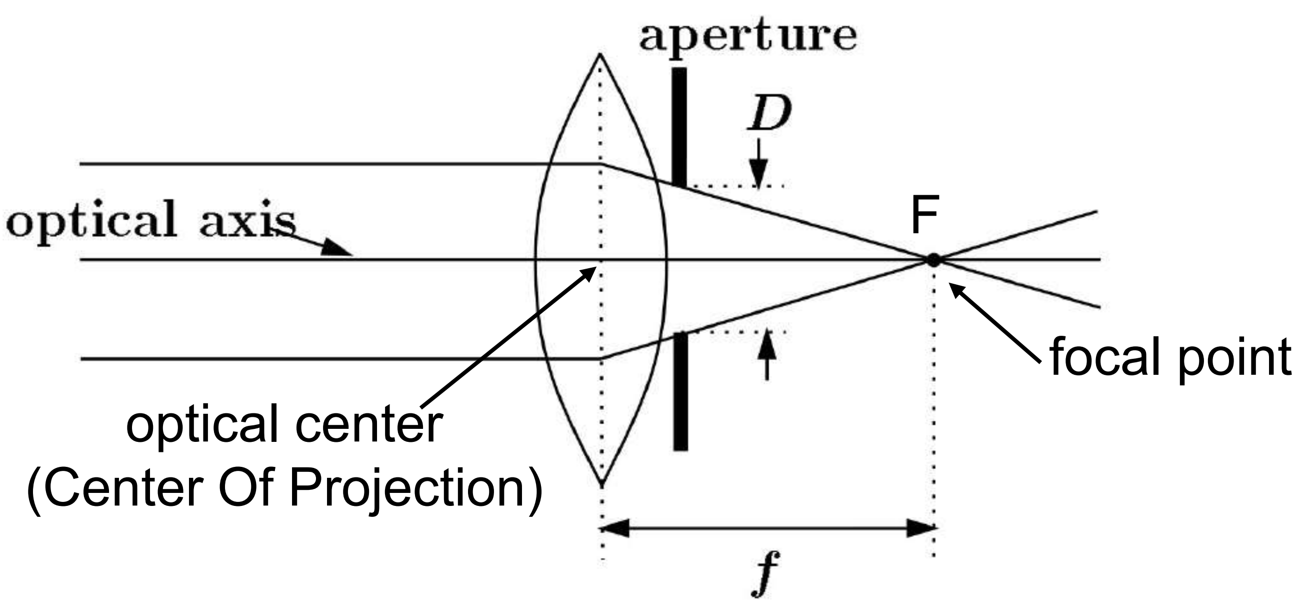


A lens focuses light onto the film

- Rays passing through the center are not deviated
- There is a specific distance at which objects are “in focus”
  - other points project to a “circle of confusion” in the image

[Slide by Steve Seitz]

Lenses



A lens focuses parallel rays onto a single **focal point**

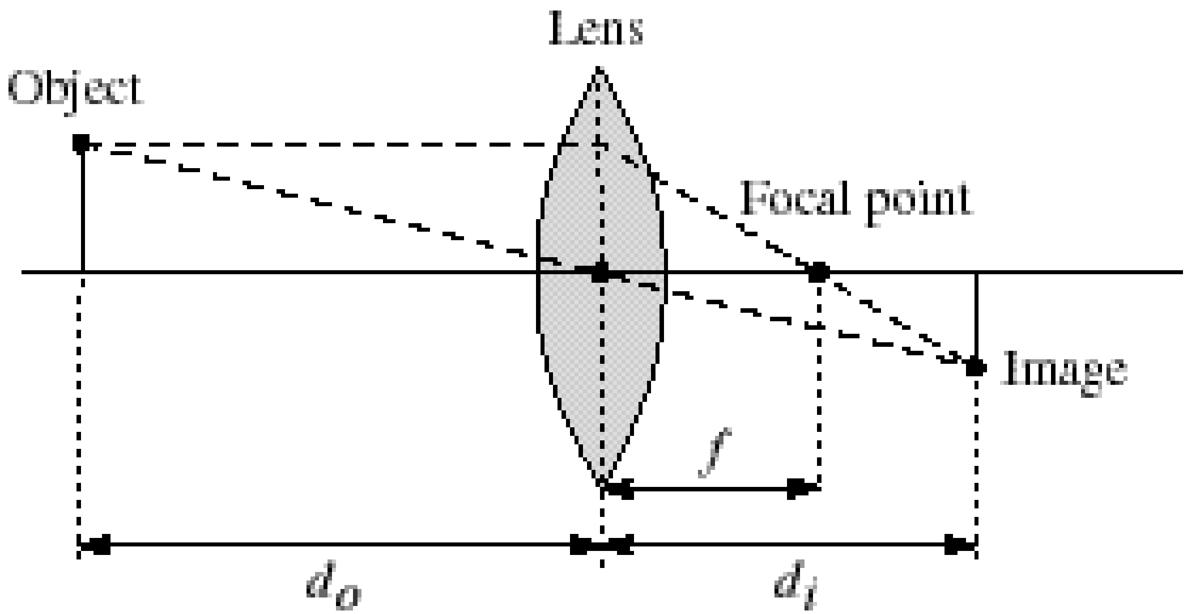
- Focal point is on a plane located at a distance **f** (**focal length**) beyond the plane of the lens
  - f is a function of the shape and index of refraction of the lens
- Aperture of diameter **D** restricts the range of rays
  - aperture may be on either side of the lens
- Lenses are typically spherical (easier to produce)

[Slide by Steve Seitz]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thin lenses



Thin lens equation: $$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

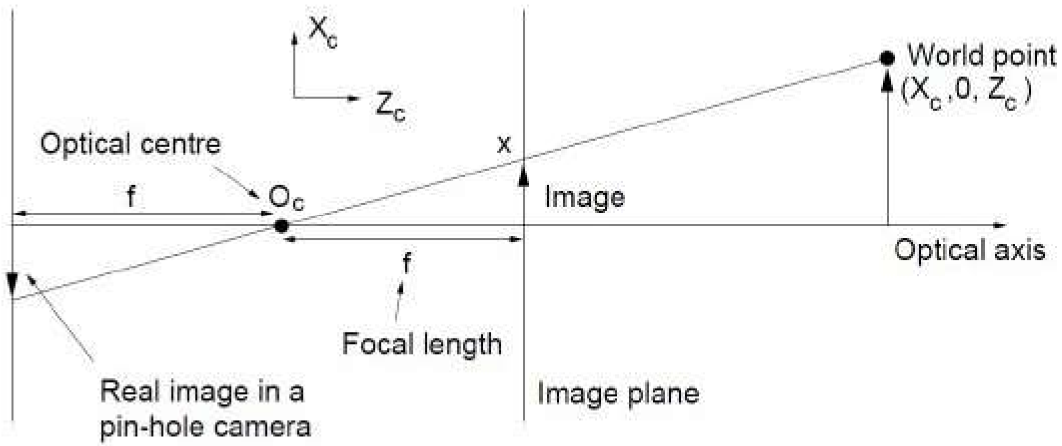
- Any object point satisfying this equation is in focus
- Thin lens applet (needs flash player): http://www.phy.ntnu.edu.tw/java/Lens/lens_e.html (by Fu-Kwun Hwang)

[Slide by Steve Seitz]

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Perspective Projection

- Objects closer to the camera appear larger, those far away from the camera appear smaller
- We can place the image plane in front on the camera to avoid flipping

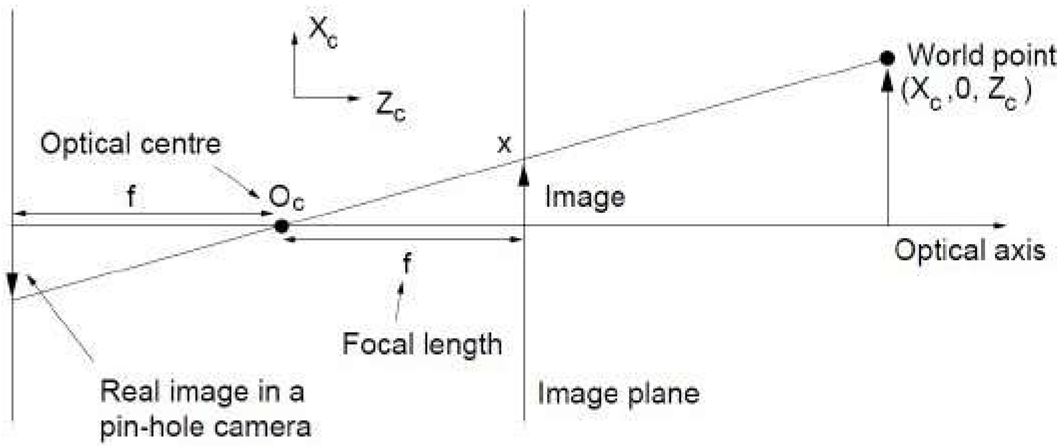


Perspective Projection equation

- Using similar triangles we find that:

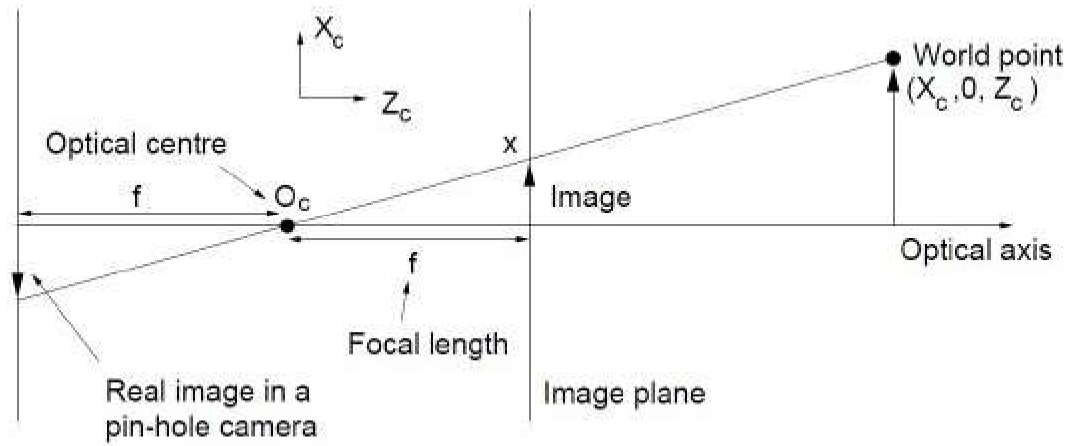
$$\frac{x}{f} = \frac{X_c}{Z_c} \Longleftrightarrow x = \frac{f}{Z_c} X_c$$

- Same for y



Perspective Projection equation

- Therefore, under perspective projection every 3D point **X**(X_C, Y_C, Z_C) projects to a 2D point **x**($\frac{f}{Z_C} X_C, \frac{f}{Z_C} Y_C$) on the camera's image plane



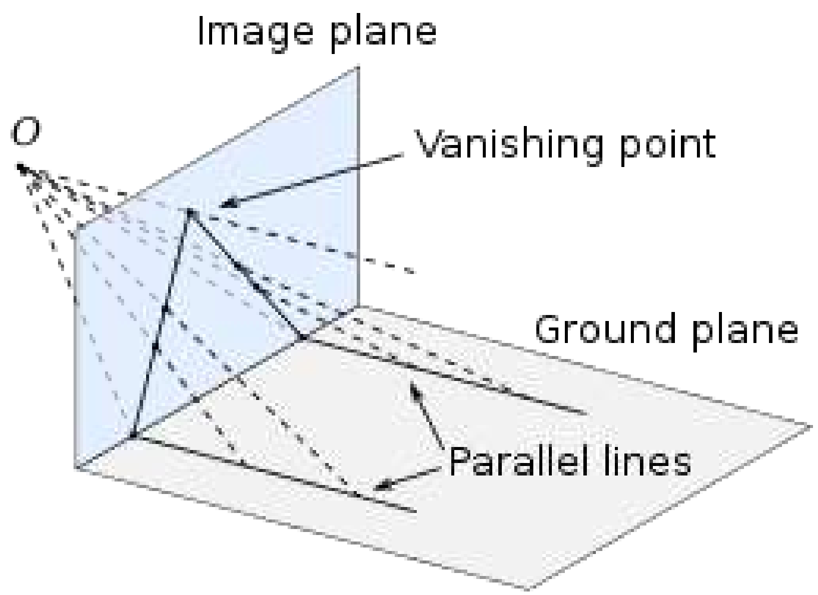
Today’s Agenda

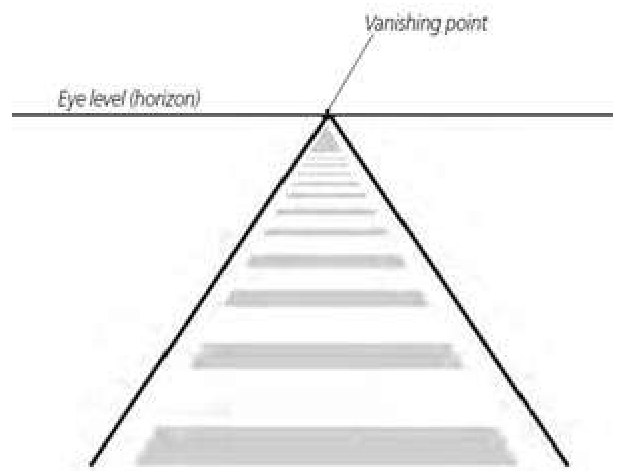
- Perspective projection
- Vanishing points
- Full camera model

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Vanishing points

- Vanishing points are points in the image where parallel lines appear to meet.
- Each set of parallel lines in the world (that are not parallel to the image plane) will have a different vanishing point in the image.

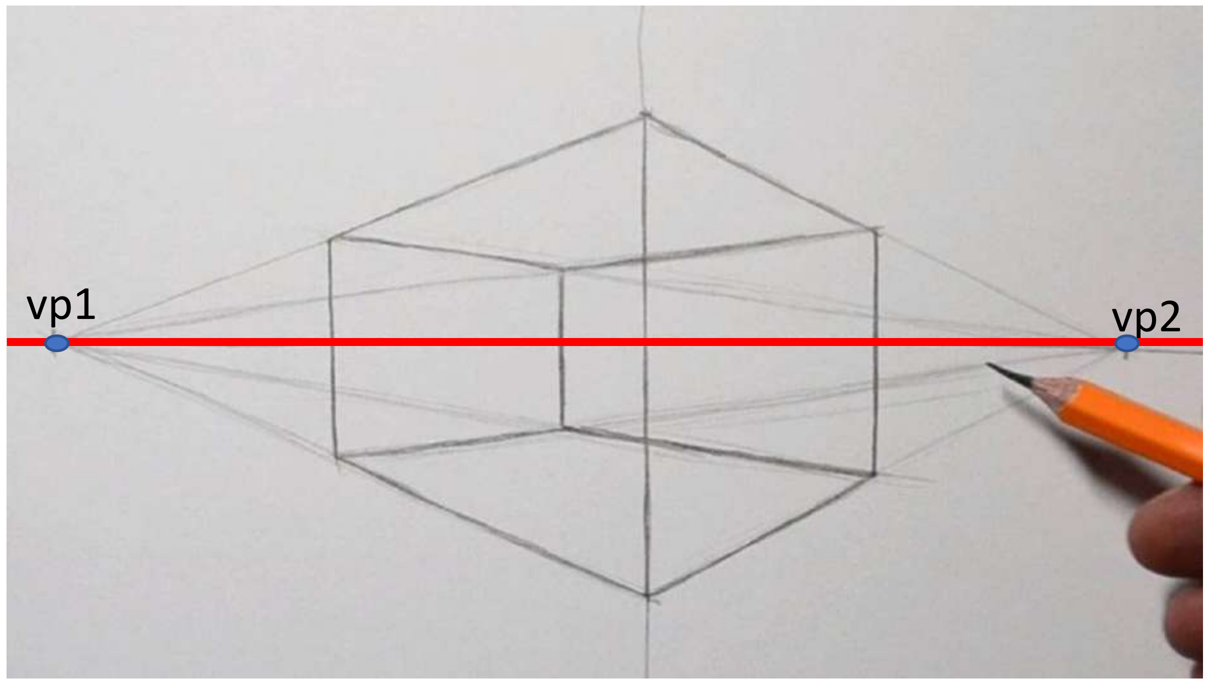




Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Vanishing points

- In the same way, parallel planes in the world meet in a line in the image, often called a **horizon line**.
- Any set of parallel lines lying on these planes in the 3D world will have a vanishing point on the horizon line.



Vanishing points

Heavily used in Renaissance paintings to introduce 'depth' into the scene

Pietro Perugino, Christ Handing the Keys to St. Peter, 1481-1482. Image via Wikimedia Commons.

Vanishing point example

- Let's find the image location of the vanishing point for a line in the world

$$\mathbf{X}_c = \mathbf{a} + \lambda\mathbf{b} = (a_x + \lambda b_x,\ a_y + \lambda b_y,\ a_z + \lambda b_z)$$

$$\Rightarrow \mathbf{x} = \frac{f}{Z_c}(X_c,\ Y_c) = f\left[\frac{a_x + \lambda b_x}{a_z + \lambda b_z},\ \frac{a_y + \lambda b_y}{a_z + \lambda b_z}\right]$$

- As λ goes to infinity we move down the line and x converges to the vanishing point

$$\mathbf{x}_{\mathrm{vp}} = \lim_{\lambda \to \infty} \mathbf{x} = f\left[\frac{b_x}{b_z},\ \frac{b_y}{b_z}\right]$$

Vanishing point example

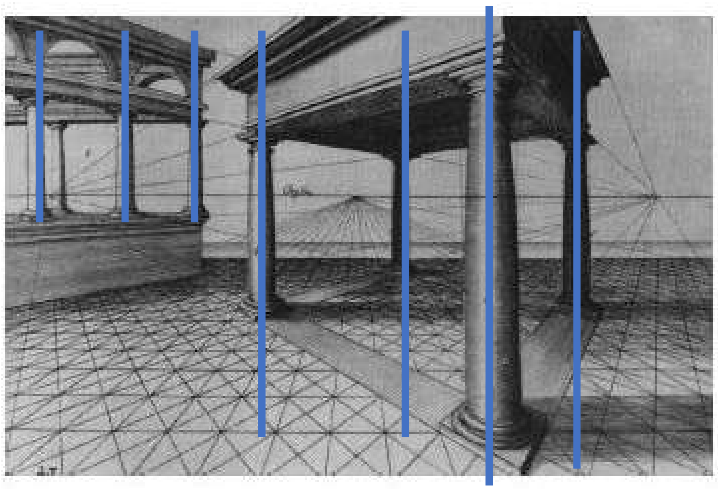
- As λ goes to infinity we move down the line and x converges to the vanishing point

$$\mathbf{x}_{\mathrm{vp}} = \lim_{\lambda \to \infty} \mathbf{x} = f \left[\frac{b_x}{b_z}, \frac{b_y}{b_z} \right]$$

- The vanishing point depends only on the line's orientation (the vector b in the line equation), and not its position.

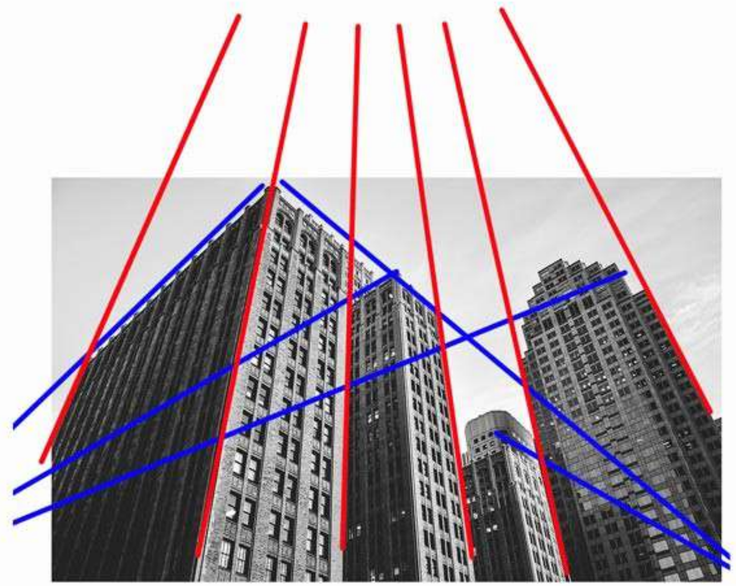
- When b_z=0, the line is parallel to the image plane, and the vanishing point is at infinity. Remember:
  - Each set of parallel lines in the world (that are not parallel to the image plane) will have a different vanishing point in the image.

Vanishing point example



$b_z=0$

vs



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

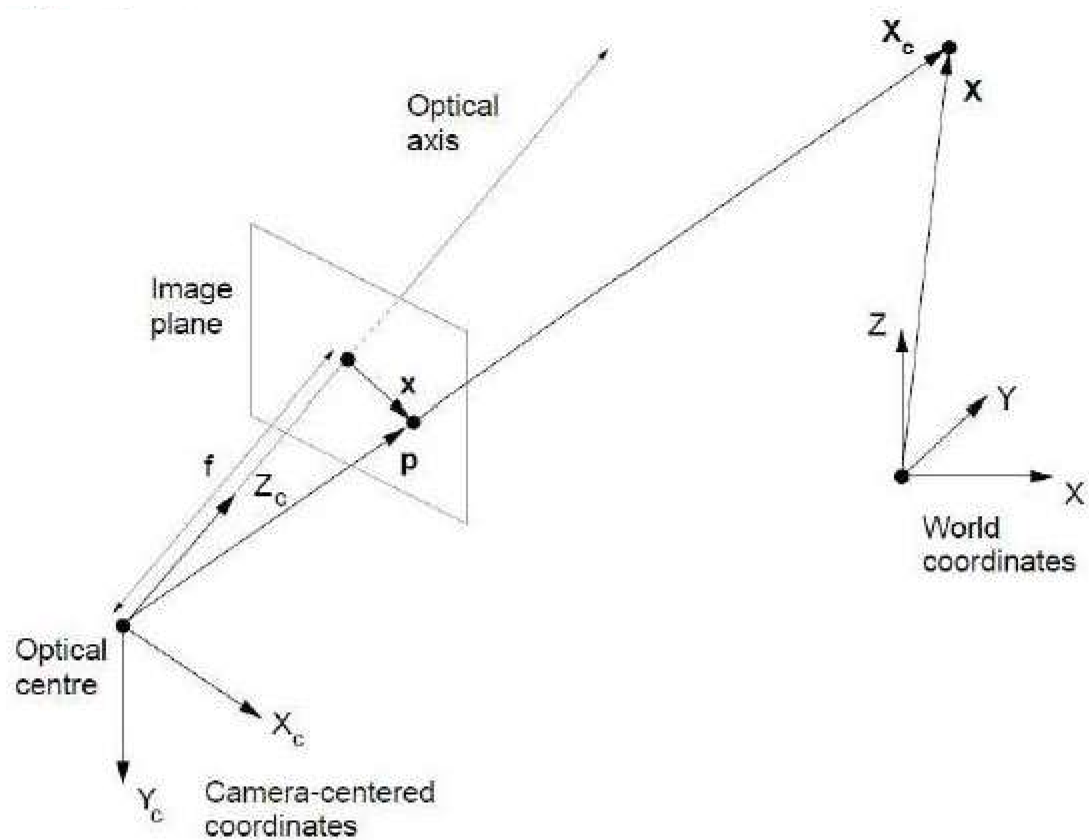
Today’s Agenda

- Perspective projection
- Vanishing points
- Full camera model

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model

A full camera model describes the mapping from 3D world to 2D pixel coordinates



Full camera model

It consists of three transformations:

1. The **Euclidean** (**Rigid**) **transformation** between the camera and the world, i.e., the translation and rotation of the camera with respect the world origin – takes points from 3D world coordinates to 3D camera coordinates

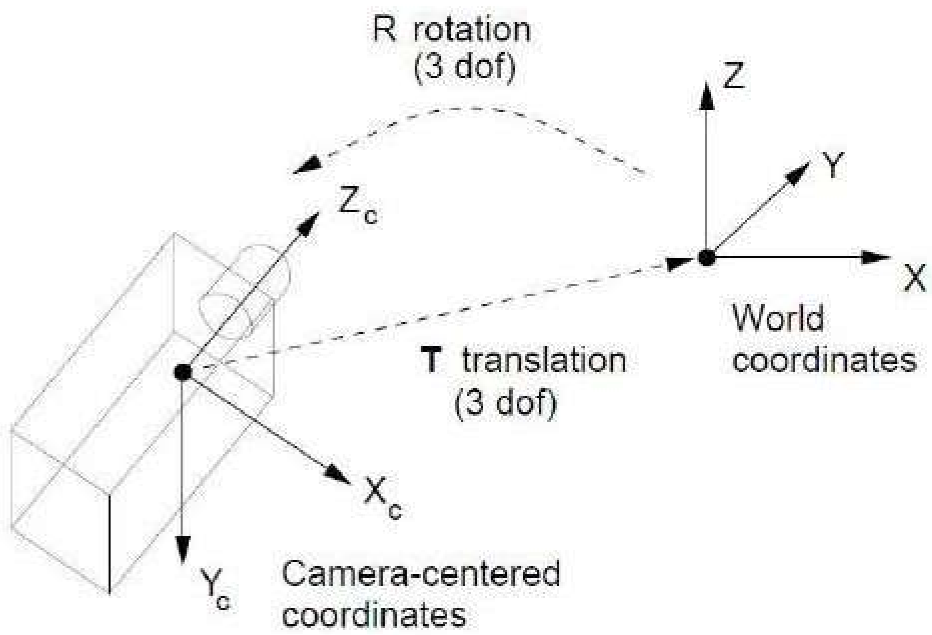
2. The **perspective projection** onto the camera plane – takes points from 3D camera coordinates to 2D image coordinates

3. **CCD imaging**, i.e., the geometry of the CCD array (the size and shape of the pixels) and its position with respect to the optical axis – takes points from 2D image coordinates to 2D pixel coordinates

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model – Euclidean transformation

We attach a coordinate system **X**(X, Y, Z) to the world and another coordinate system to the camera $\mathbf{X_C}$ (X_C, Y_C, Z_C)

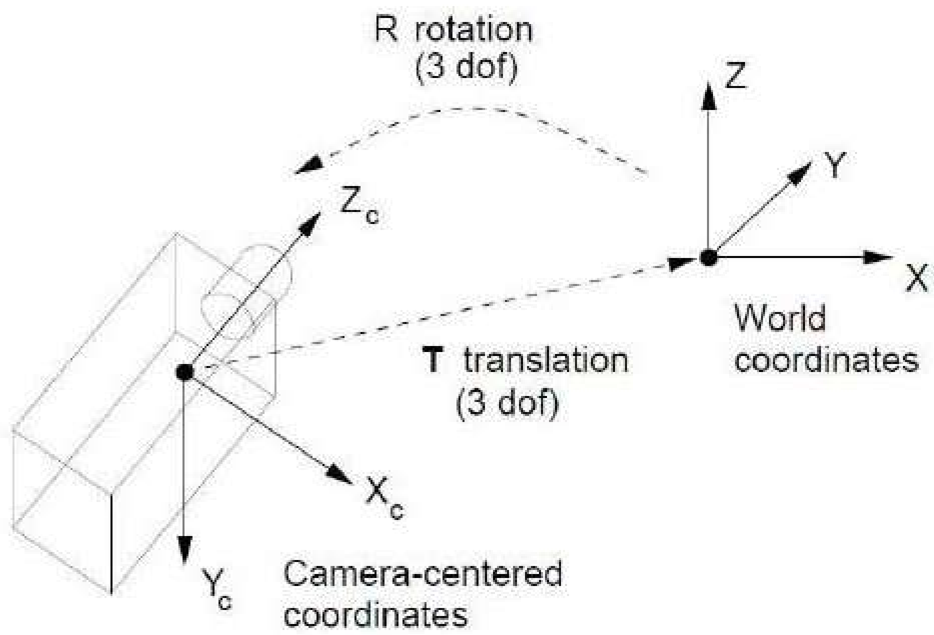


Full camera model – Euclidean transformation

The Euclidean transformation can be described by a rotation matrix R and a translation vector T

$$\mathbf{X}_c = \mathbf{R}\mathbf{X} + \mathbf{T}$$

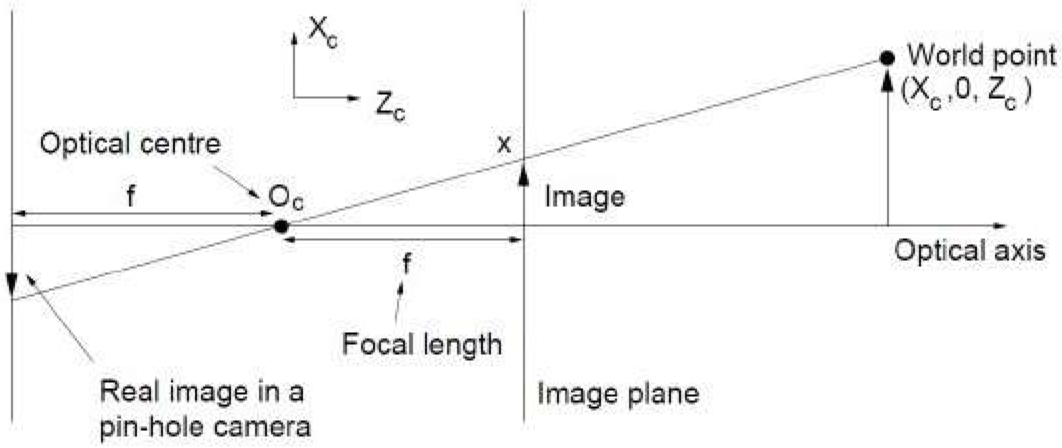
$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$



Full camera model – Perspective projection

Perspective projection is modelled by

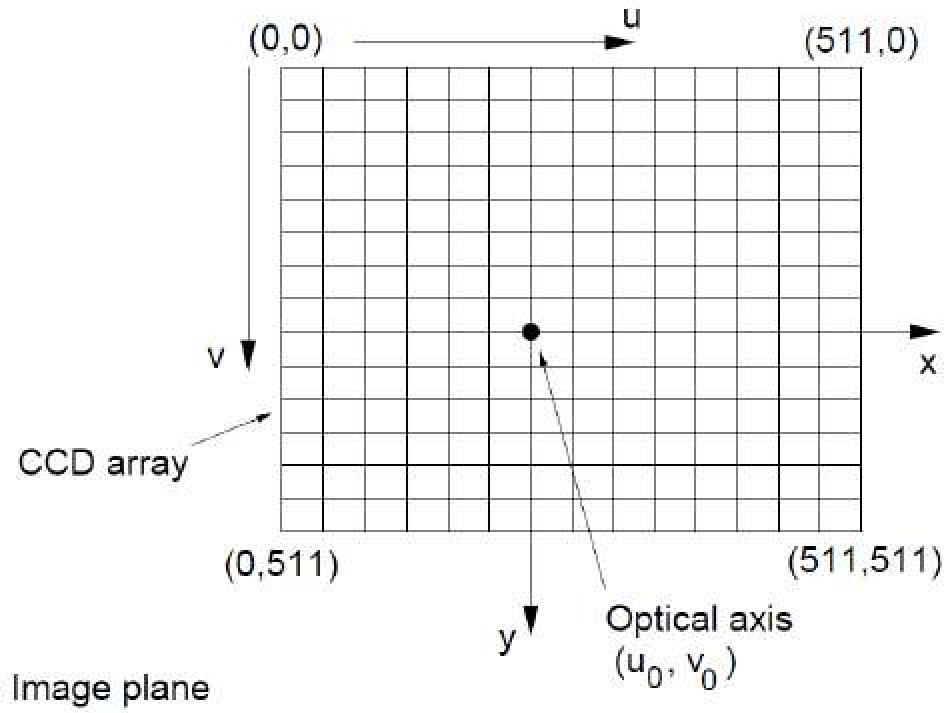
$$\left.\begin{aligned} x &= \frac{f}{Z_c} X_c \\ y &= \frac{f}{Z_c} Y_c \end{aligned}\right\}$$



This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model – CCD Imaging

To model CCD Imaging we define pixel coordinates **w**(u,v) in addition to the image coordinates **x**(x,y)

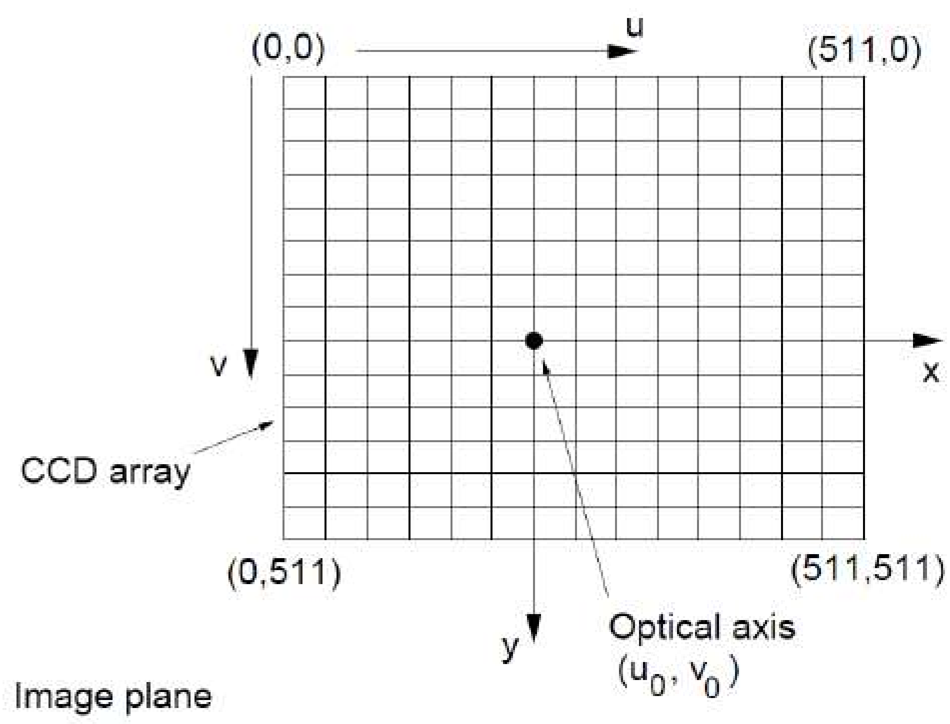


This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model – CCD Imaging

w(u,v) and **x**(x,y) are related as follows

$$\left.\begin{aligned} u &= u_0 + k_u x \\ v &= v_0 + k_v y \end{aligned}\right\}$$



Full camera model – Putting it all together

So here is the full camera model combining all three transformations

$$\left.\begin{array}{rcl} u & = & u_0 + k_u x \\ v & = & v_0 + k_v y \end{array}\right\} \Rightarrow \left.\begin{array}{rcl} u & = & u_0 + k_u \dfrac{f}{Z_c} X_c \\ v & = & v_0 + k_v \dfrac{f}{Z_c} Y_c \end{array}\right\}$$

$$\Rightarrow \left.\begin{array}{rcl} u & = & u_0 + k_u f \dfrac{r_{11}X + r_{12}Y + r_{13}Z + T_x}{r_{31}X + r_{32}Y + r_{33}Z + T_z} \\ v & = & v_0 + k_v f \dfrac{r_{21}X + r_{22}Y + r_{23}Z + T_y}{r_{31}X + r_{32}Y + r_{33}Z + T_z} \end{array}\right\}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 15: Camera Calibration

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union

Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Perspective projection
- Vanishing points
- Full camera model

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Full camera model in matrix form
- Camera calibration
- Calibration – Projective camera model
- Calibration – Affine camera model

[material based on Paris Kaimakis]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

- Full camera model in matrix form
- Camera calibration
- Calibration - Projective camera model
- Calibration – Affine camera model

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Remember - Full camera model

It consists of three transformations:

1. The **Euclidean** (**Rigid**) **transformation** between the camera and the world, i.e., the translation and rotation of the camera with respect the world origin – takes points from 3D world coordinates to 3D camera coordinates

2. The **perspective projection** onto the camera plane – takes points from 3D camera coordinates to 2D image coordinates

3. **CCD imaging**, i.e., the geometry of the CCD array (the size and shape of the pixels) and its position with respect to the optical axis – takes points from 2D image coordinates to 2D pixel coordinates

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Remember - Full camera model – Putting it all together

So here is the full camera model combining all three transformations

$$\left.\begin{aligned} u &= u_0 + k_u x \\ v &= v_0 + k_v y \end{aligned}\right\} \Rightarrow \left.\begin{aligned} u &= u_0 + k_u \frac{f}{Z_c} X_c \\ v &= v_0 + k_v \frac{f}{Z_c} Y_c \end{aligned}\right\}$$

$$\Rightarrow \left.\begin{aligned} u &= u_0 + k_u f \frac{r_{11}X + r_{12}Y + r_{13}Z + T_x}{r_{31}X + r_{32}Y + r_{33}Z + T_z} \\ v &= v_0 + k_v f \frac{r_{21}X + r_{22}Y + r_{23}Z + T_y}{r_{31}X + r_{32}Y + r_{33}Z + T_z} \end{aligned}\right\}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model in matrix form - Euclidean

- How can we express all three transformations in the form of matrices ?
  - Homogeneous coordinates
- Let's start from the Euclidean (rigid) transformation connecting the world and camera coordinate systems.
- This is composed of a rotation (3DOF) expressing the camera pose with respect to the world coordinate frame, and a translation (3DOF) expressing the camera location with respect to the world origin.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

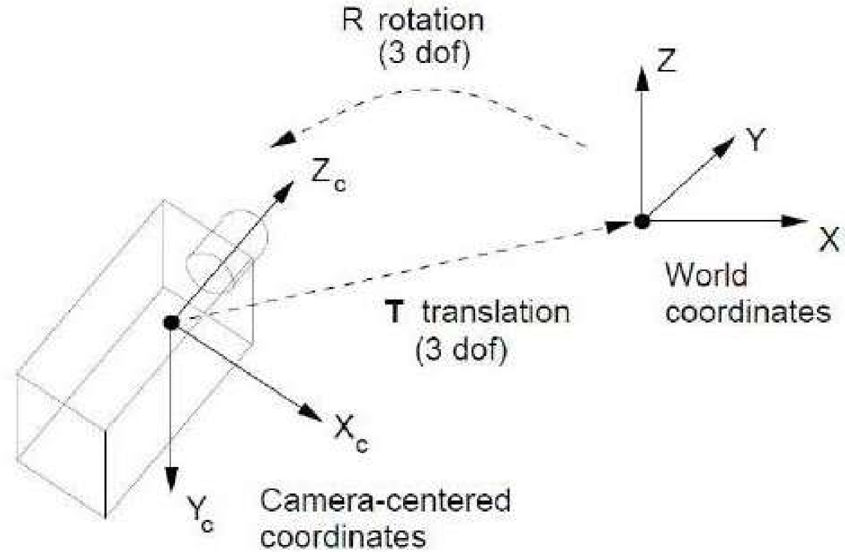
Full camera model in matrix form - Euclidean

- We want to express a 'world' 3D point **X**(X, Y, Z) as a 3D point $\mathbf{X_C}$ (X_C, Y_C, Z_C) in camera coordinates
- In homogeneous coordinates we have $\tilde{X}$(λX,λY,λZ,λ) and $\widetilde{X_C}$(λX_C,λY_C,λZ_C,λ) – λ can be set to 1 as it has no effect on $\mathbf{X_C}$, the cartesian equivalent of $\widetilde{X_C}$
- The Euclidean transformation can now be expressed as:

$$\begin{bmatrix} \lambda X_C \\ \lambda Y_C \\ \lambda Z_C \\ \lambda \end{bmatrix} = \begin{bmatrix} r11 & r12 & r13 & T_x \\ r21 & r22 & r23 & T_y \\ r31 & r32 & r33 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda X \\ \lambda Y \\ \lambda Z \\ \lambda \end{bmatrix}$$

- Or equivalently:

$$\tilde{\mathbf{X}}_c = \mathbf{P}_e \tilde{\mathbf{X}} \quad \text{where} \quad \mathbf{P}_e = \left[\begin{array}{ccc|c} & \mathbf{R} & & \mathbf{T} \\ \hline 0 & 0 & 0 & 1 \end{array}\right]$$



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model in matrix form - Perspective

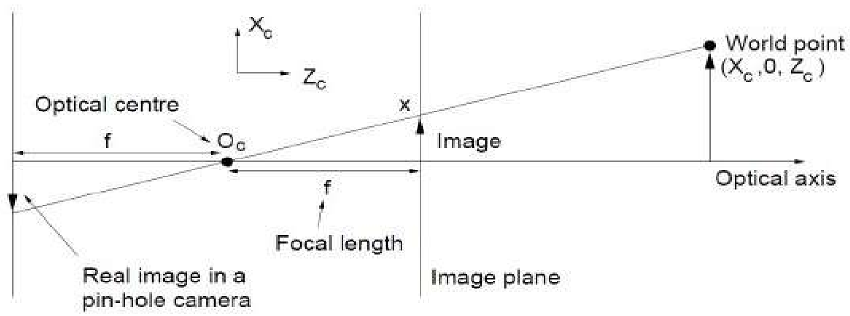
- We want to compute the projection of a 3D point $\mathbf{X_C}$ (X_C, Y_C, Z_C) expressed in camera coordinates onto a 2D point $\mathbf{x}(x,y)$ lying on the image plane
- In homogeneous coordinates we have $\widetilde{X_C}(\lambda X_C,\lambda Y_C,\lambda Z_C,\lambda)$ and $\tilde{x}(sx,sy,s)$ – again λ and s can be set to 1
- Perspective projection can now be expressed as:

$$\begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda X_c \\ \lambda Y_c \\ \lambda Z_c \\ \lambda \end{bmatrix}$$

$$\left.\begin{aligned} x &= \frac{f}{Z_c} X_c \\ y &= \frac{f}{Z_c} Y_c \end{aligned}\right\}$$

- Or equivalently:

$$\tilde{\mathbf{x}} = \mathbf{P}_p \, \tilde{\mathbf{X}}_c$$

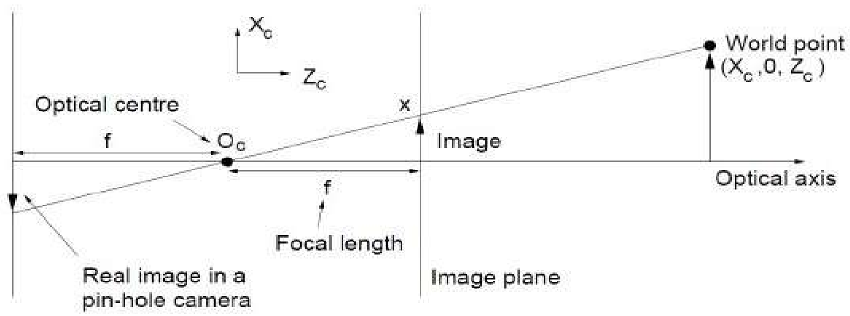


Full camera model in matrix form - Perspective

- We can confirm that the expression in the previous slide is equivalent to the perspective equation by recovering **x**, the cartesian equivalent of $\tilde{x}$

$$\tilde{x} = \begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} f\lambda X_C \\ f\lambda Y_C \\ \lambda Z_C \end{bmatrix} \Rightarrow x = \begin{bmatrix} sx/s \\ sy/s \end{bmatrix} = \begin{bmatrix} f\lambda X_C/\lambda Z_C \\ f\lambda Y_C/\lambda Z_C \end{bmatrix} = \begin{bmatrix} fX_C/Z_C \\ fY_C/Z_C \end{bmatrix}$$

$$\boxed{\left.\begin{aligned} x &= \frac{f}{Z_c} X_c \\ y &= \frac{f}{Z_c} Y_c \end{aligned}\right\}}$$



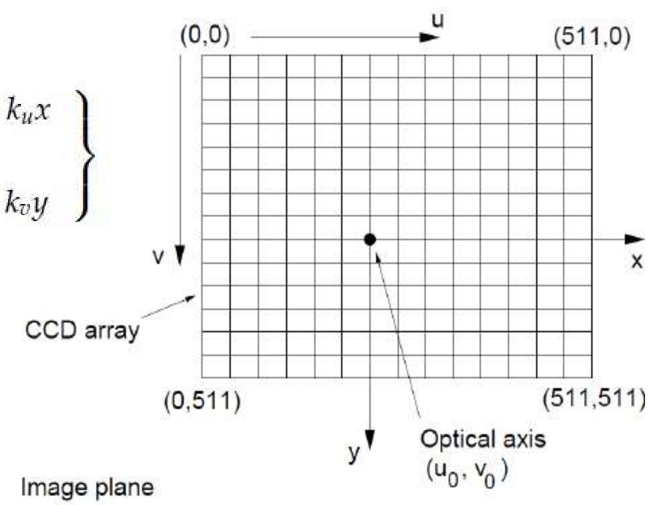
Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model in matrix form – CCD Imaging

- Last, wee want to express 2D point **x**(x,y) lying on the image plane in pixel coordinates **w**(u,v)
- In homogeneous coordinates we have $\tilde{x}$(sx,sy,s) and $\widetilde{w}$(su,su,s) – again s can be set to 1
- This is a translation and scaling which can be expressed as:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} sx \\ sy \\ s \end{bmatrix}$$

$$\left.\begin{aligned} u &= u_0 + k_u x \\ v &= v_0 + k_v y \end{aligned}\right\}$$



- Or equivalently:

$$\tilde{\mathbf{w}} = \mathbf{P}_c \tilde{\mathbf{x}}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model in matrix form – all together

We can now express the whole process, from $\tilde{X}$ to $\widetilde{w}$ as a single transformation P_{ps}

$$
\begin{aligned}
\tilde{\mathbf{w}} &= \mathbf{P}_{ps}\,\tilde{\mathbf{X}} \\
\text{where}\quad \mathbf{P}_{ps} &= \mathbf{P}_c\,\mathbf{P}_p\,\mathbf{P}_e \\
&= \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
\left[\begin{array}{ccc|c} & \mathbf{R} & & \mathbf{T} \\ \hline 0 & 0 & 0 & 1 \end{array}\right]
\end{aligned}
$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model in matrix form – all together

- The transformation P_{ps} is not a general 3x4 matrix, because it has a special structure composed of P_e, P_p,and P_c.
- We can simplify P_{ps} into an upper-triangular matrix K composed of P_e and P_p, and a matrix representing the Euclidean transformation.

$$\mathbf{P}_{ps} = \mathbf{K}\,[\mathbf{R}\,|\mathbf{T}\,]$$

$$= \begin{bmatrix} m_u & 0 & u_0 & 0 \\ 0 & m_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r11 & r12 & r13 & T_x \\ r21 & r22 & r23 & T_y \\ r31 & r32 & r33 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$m_u = k_u\, f$$

$$m_v = k_v\, f$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Full camera model in matrix form – all together

- The matrix **K** is called the camera calibration matrix and it contains all the viewing parameters (**intrinsics**) coming from inside the camera.
- The Euclidean transformation matrix contains all the viewing parameters (**extrinsics**) that are not controlled by the camera lens or sensor.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

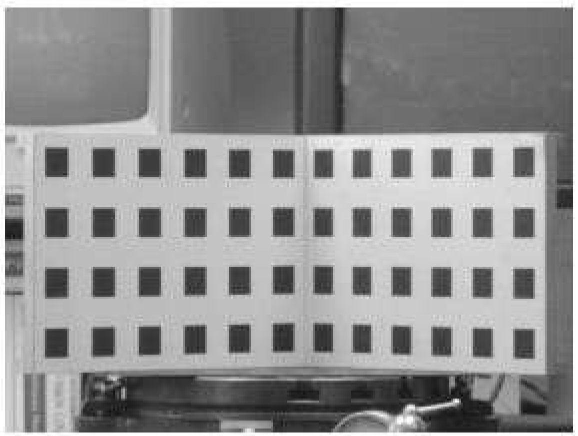
- Full camera model in matrix form
- Camera calibration
- Calibration - Projective camera model
- Calibration – Affine camera model

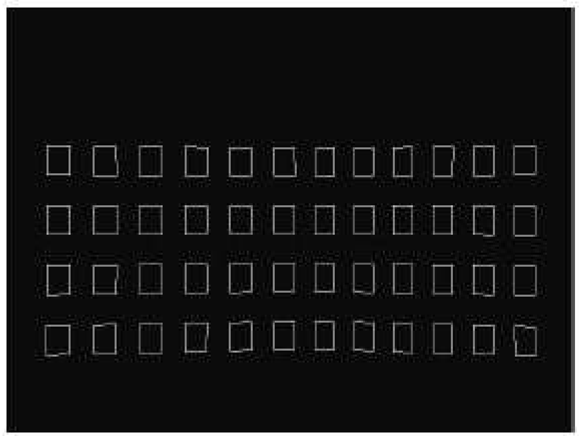
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Camera calibration

- Camera calibration is the name given to the process of discovering the parameters inside our camera model, i.e., the values inside the camera projection matrix.
- This is done by using an image of a controlled scene.
- We may want to use a scene with some sort of regular pattern.



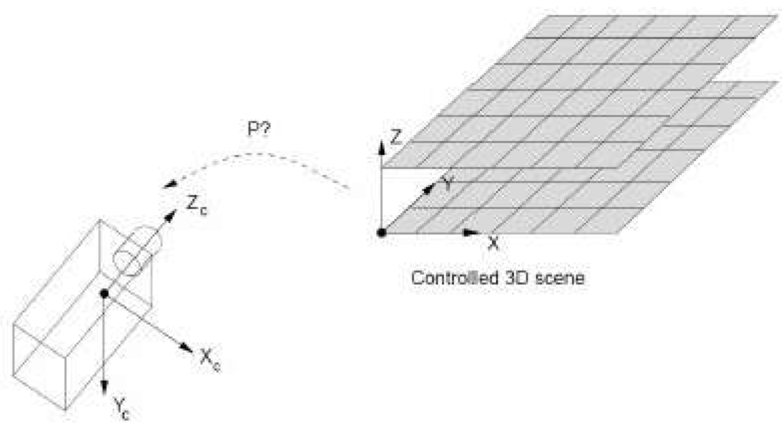


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Camera calibration

- Camera calibration is the name given to the process of discovering the parameters inside our camera model, i.e., the values inside the camera projection matrix. This is done by using an image of a controlled scene.
- We may want to use a scene with some sort of regular pattern



Today's Agenda

- Full camera model in matrix form
- Camera calibration
- Calibration - Projective camera model
- Calibration – Affine camera model

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Camera calibration – the projective camera

Remember that the **perspective** camera projection matrix P_{ps} is not a general 3x4 matrix, as it has a special structure composed of P_e, P_p,and P_c.

$$
\begin{aligned}
\tilde{\mathbf{w}} &= \mathbf{P}_{ps}\tilde{\mathbf{X}} \\
\text{where} \quad \mathbf{P}_{ps} &= \mathbf{P}_c\,\mathbf{P}_p\,\mathbf{P}_e \\
&= \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
\left[\begin{array}{ccc|c} & \mathbf{R} & & \mathbf{T} \\ \hline 0 & 0 & 0 & 1 \end{array}\right]
\end{aligned}
$$

Camera calibration – the projective camera

- Calibrating the perspective camera model may therefore be difficult. We can instead use the projective camera model which is described by a general 3x4 matrix.

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$

- The projective camera has **11 degrees of freedom**, since the overall scale of P does not matter when using homogeneous coordinates.
- It is more convenient to deal with a projective camera than a perspective one, since we don't have to worry about placing nonlinear constraints on the elements of P

Camera calibration – the projective camera

- Using a projective camera, the whole imaging process is described by

$$\tilde{\mathbf{w}} = \mathbf{P}\tilde{\mathbf{X}}$$

$$\Rightarrow \begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- We must estimate 11 parameters (the overall scale does not matter, so let's set $p_{34} = 1$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Camera calibration – the projective camera

- Because the scene is controlled, we **know** the location of the world points $\mathbf{X}_i$
- We can also find the location of their projections $\mathbf{w}_i$ on the image, e.g., using corner detection.
- Each point we observe gives us two equations

$$u_i = \frac{su_i}{s} = \frac{p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + 1}$$

$$v_i = \frac{sv_i}{s} = \frac{p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + 1}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Camera calibration – the projective camera

Rearranging them gives us two linear equations in the unknown parameters of the projection matrix

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & u_iX_i & u_iY_i & u_iZ_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & v_iX_i & v_iY_i & v_iZ_i \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \end{bmatrix} = \begin{bmatrix} -u_i \\ -v_i \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Camera calibration – the projective camera

- As there are 11 unknowns, we need at least 6 points to calibrate this camera model to get enough equations (rows) in the linear system Ax=b.
- Each observed points adds two equations (rows) to the matrix A
- We can solve the system of equations using linear least squares (pseudo-inverse of A)

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \Rightarrow \quad \mathbf{x} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \end{aligned}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today's Agenda

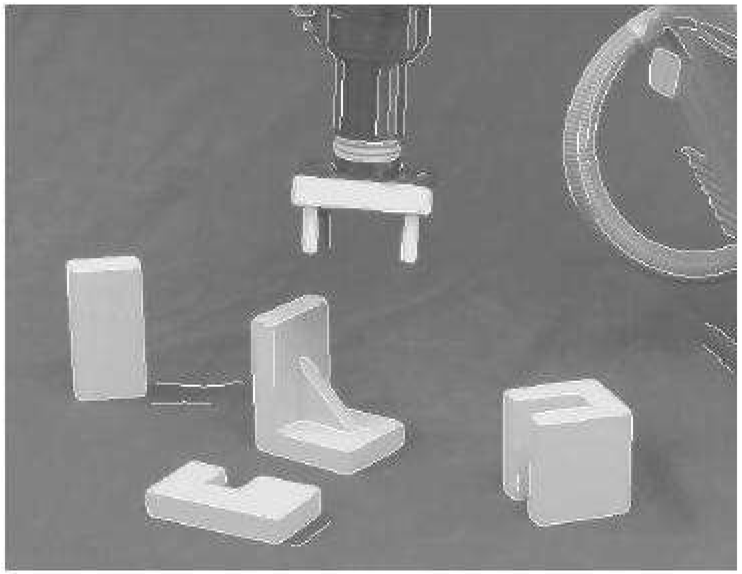
- Full camera model in matrix form
- Camera calibration
- Calibration – Projective camera model
- Calibration – Affine camera model

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

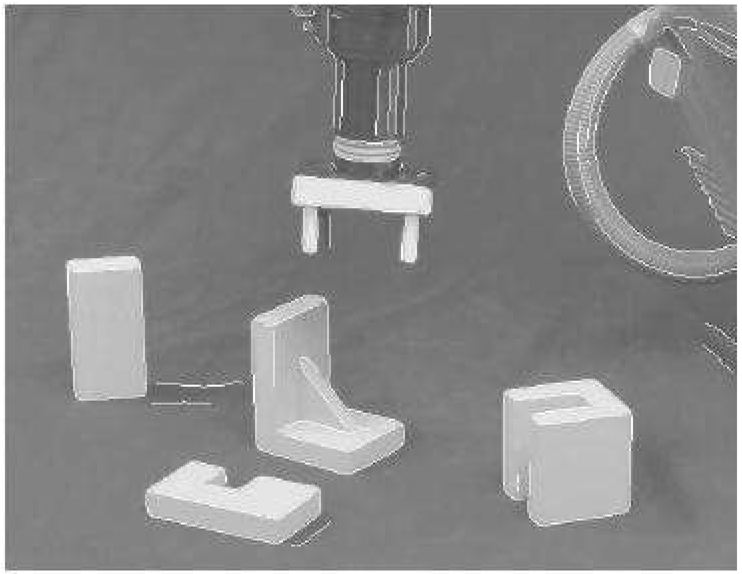
Parallel Projection

When the depth difference ΔZ_C of objects in the scene is small compared to their distance Z_C to the camera, the resulting image is not described well by perspective projection.



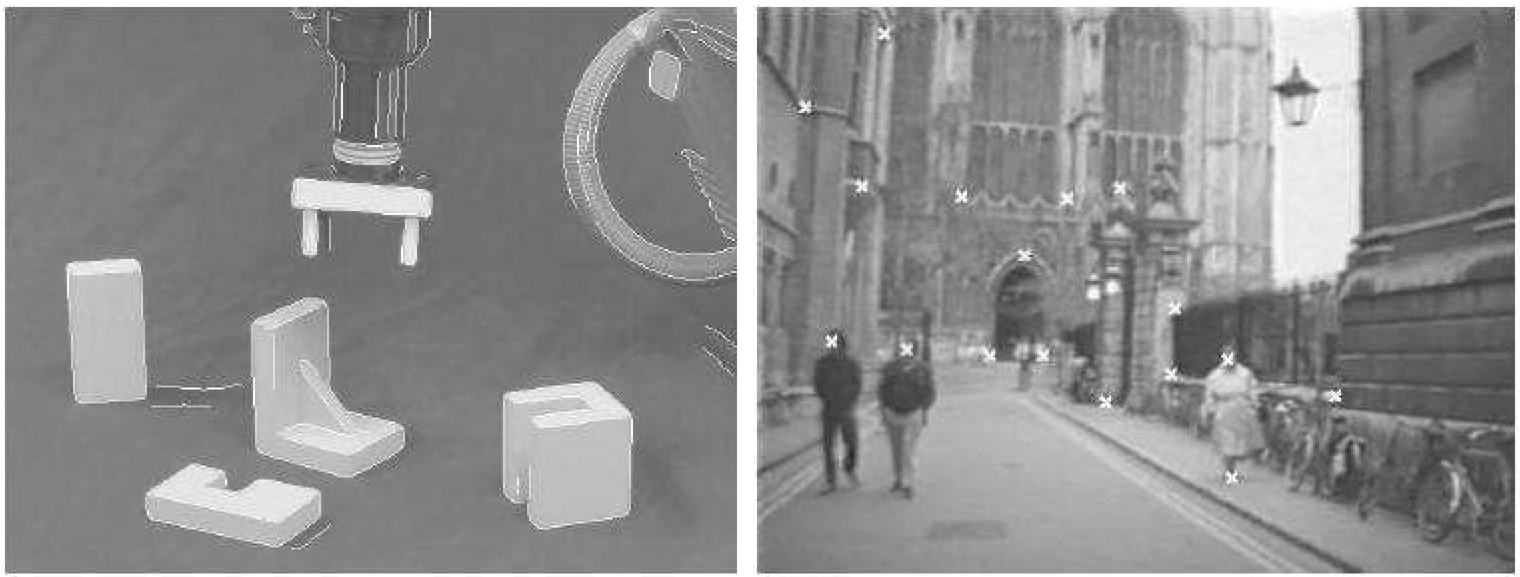
Parallel Projection

For example, parallel lines on the left image remain parallel after projection. In this case, the 3D to 2D projection is better described by **parallel projection**.



Parallel Projection

A camera which creates images like the one on the left is known as a **weak-perspective camera.**



Parallel Projection

What changes in our imaging transformations is the perspective projection matrix. Remember P_p:

$$\begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Parallel Projection

What changes in our imaging transformations is the perspective projection matrix.
If the depth variation in the scene is small, then $Z_C \approx Z_C^{avg}$ and the perspective projection matrix P_p can be changed to the parallel projection matrix P_{pll}

$$\begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & Z_c^{\text{avg}} \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

$$\tilde{\mathbf{x}} = \mathbf{P}_{pll} \tilde{\mathbf{X}}_c$$

Weak-Perspective Camera Model

This leads to the **weak perspective** camera model:

$$\tilde{\mathbf{w}} = \mathbf{P}_{wp}\,\tilde{\mathbf{X}}$$

where

$$\begin{aligned}
\mathbf{P}_{wp} &= \mathbf{P}_c\,\mathbf{P}_{pll}\,\mathbf{P}_e \\
&= \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & Z_c^{\text{avg}} \end{bmatrix} \left[\begin{array}{ccc|c} & \mathbf{R} & & \mathbf{T} \\ \hline 0 & 0 & 0 & 1 \end{array}\right]
\end{aligned}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Weak-Perspective Camera Model

P_{wp} is the projection matrix for a weak-perspective camera. It is a 3x4 matrix with a special structure composed of P_e, P_{pll},and P_c.

$$\tilde{\mathbf{w}} = \mathbf{P}_{wp}\,\tilde{\mathbf{X}}$$

where

$$\begin{aligned} \mathbf{P}_{wp} &= \mathbf{P}_c\,\mathbf{P}_{pll}\,\mathbf{P}_e \\ &= \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & Z_c^{\text{avg}} \end{bmatrix} \left[\begin{array}{ccc|c} & \mathbf{R} & & \mathbf{T} \\ \hline 0 & 0 & 0 & 1 \end{array}\right] \end{aligned}$$

Camera calibration – the affine camera

- The special structure of the weak-perspective camera model makes it difficult to calibrate. This is why the affine camera model is often used instead:

$$\mathbf{P}_{aff} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{bmatrix}$$

- P_{aff} is the projection matrix for the affine camera. It has 8 degrees of freedom (remember the overall scale does not matter so we can set p_{34} to 1).
- It can be calibrated in the same way as the projective camera. There are eight degrees of freedom, so we need a minimum of 4 points.

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

University of Cyprus – MSc Artificial Intelligence

MAI644 – COMPUTER VISION

Lecture 16: Stereo Vision

Melinos Averkiou

CYENS Centre of Excellence

University of Cyprus - Department of Computer Science

m.averkiou@cyens.org.cy

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union

Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Last time

- Full camera model in matrix form
- Camera calibration
- Calibration – Projective camera model
- Calibration – Affine camera model

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Recovery of world position
- Triangulation
- Epipolar Geometry

[material based on Paris Kaimakis]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Today’s Agenda

- Recovery of world position
- Triangulation
- Epipolar Geometry

Recovery of world position

- Previously we saw that the imaging process can be described as a transformation in homogeneous coordinates.
- If we can invert this transformation, the world coordinates of each pixel in the image can be computed.
- Recovering world coordinates of objects based on the projection on an image is known as **shape recovery** or **depth recovery**.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Recovery of world position

- Is this possible using a single camera ?
- The camera needs to be calibrated, i.e. we know all its parameters
- Remember the projective camera model:

$$\tilde{\boldsymbol{w}} = \boldsymbol{P}\tilde{\boldsymbol{X}}$$

$$\Leftrightarrow \begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Unfortunately the transformation described in $\boldsymbol{P}$ is not invertible and the world point $\boldsymbol{X}$ cannot be uniquely determined

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

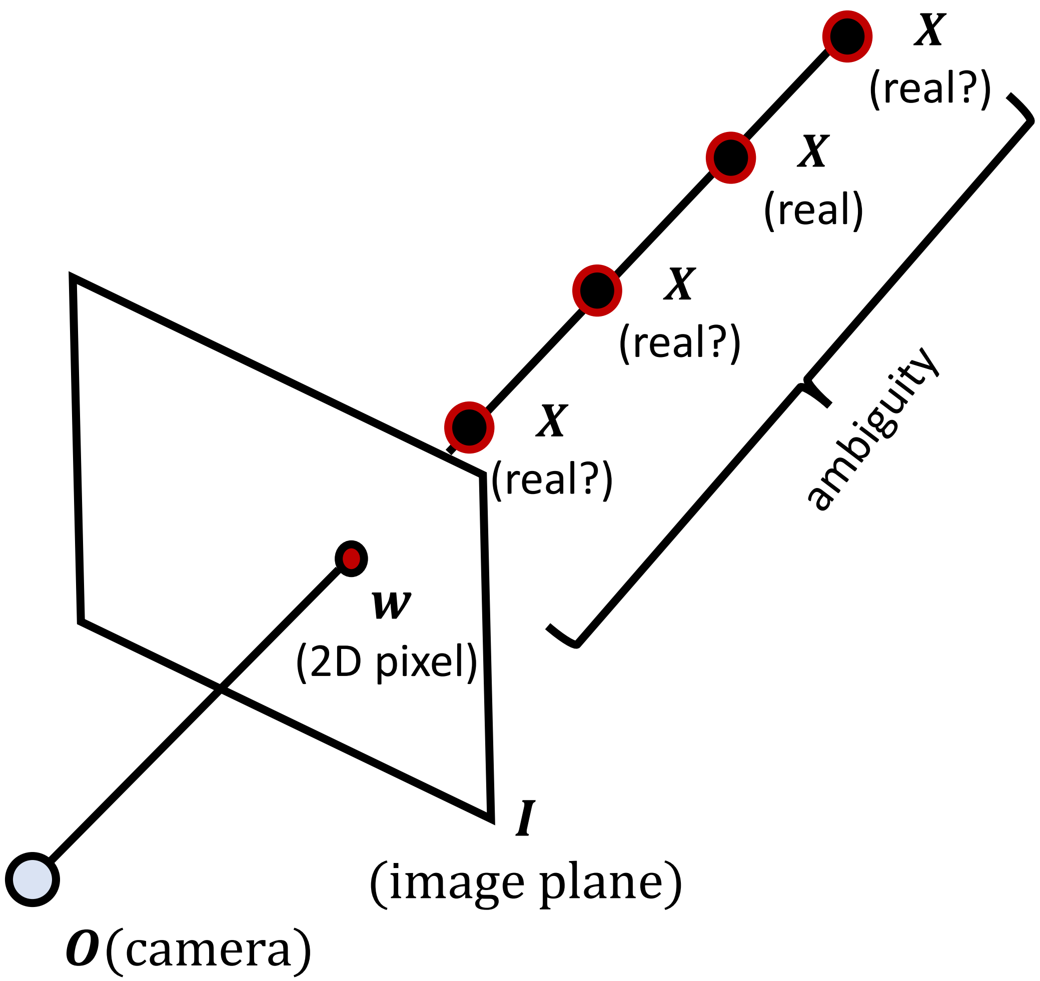
Recovery of world position

- Depth ambiguity

Courtesy slide S. Lazebnik

Recovery of world position

- Each observed feature on the image gives 2 equations with 3 unknowns and therefore defines a line (a ray) of solutions for $\boldsymbol{X}$



Recovery of world position

- Each observed feature on the image gives 2 equations with 3 unknowns and therefore defines a line (a ray) of solutions for $\boldsymbol{X}$
- This system of equations is under-constrained.
- This can be seen by the size of $\boldsymbol{P}$. There are more columns than rows.

$$\widetilde{\boldsymbol{w}} = \boldsymbol{P}\widetilde{\boldsymbol{X}}$$

$$\Leftrightarrow \begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Recovery of world position

- Under-constrained problems never have a unique solution.

- To uniquely recover $\boldsymbol{X}$, additional views must be used, so that the transformation between $\boldsymbol{w}$ (pixel coordinates) and $\boldsymbol{X}$ (world coordinates) is *forced* to become invertible.

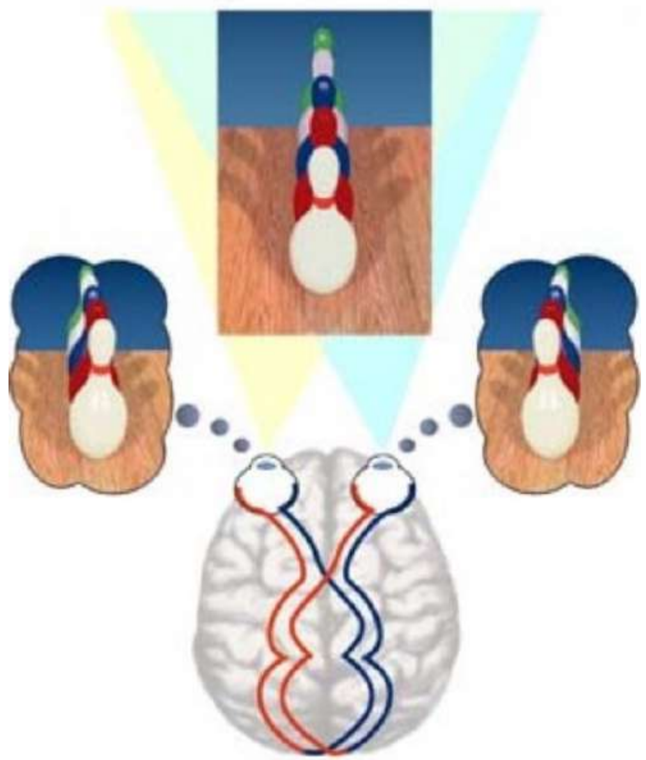
- This is the subject of **stereo vision**.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Recovery of world position

- **Two** eyes/**cameras** help.



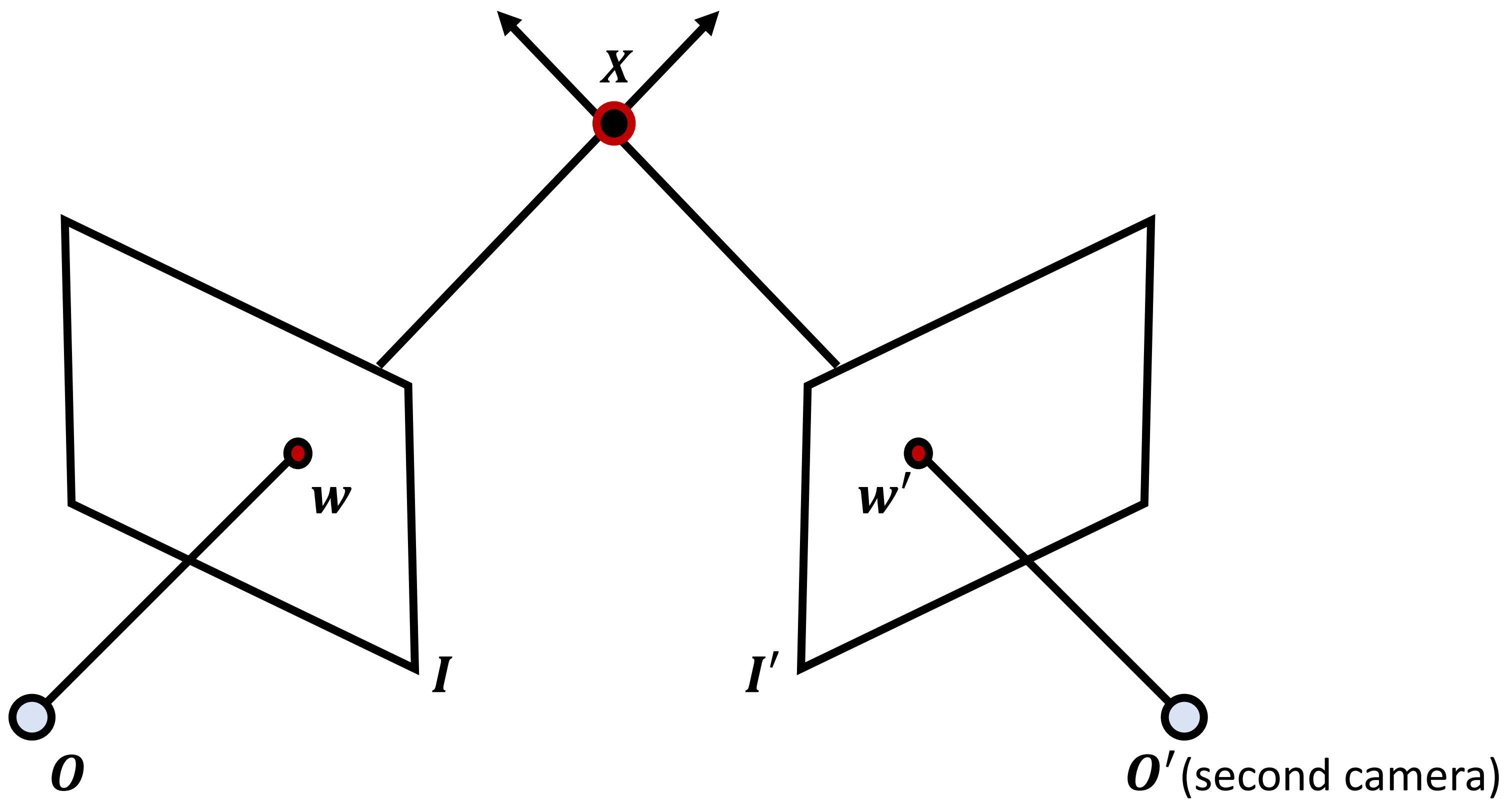
Today's Agenda

- Recovery of world position
- Triangulation
- Epipolar Geometry

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

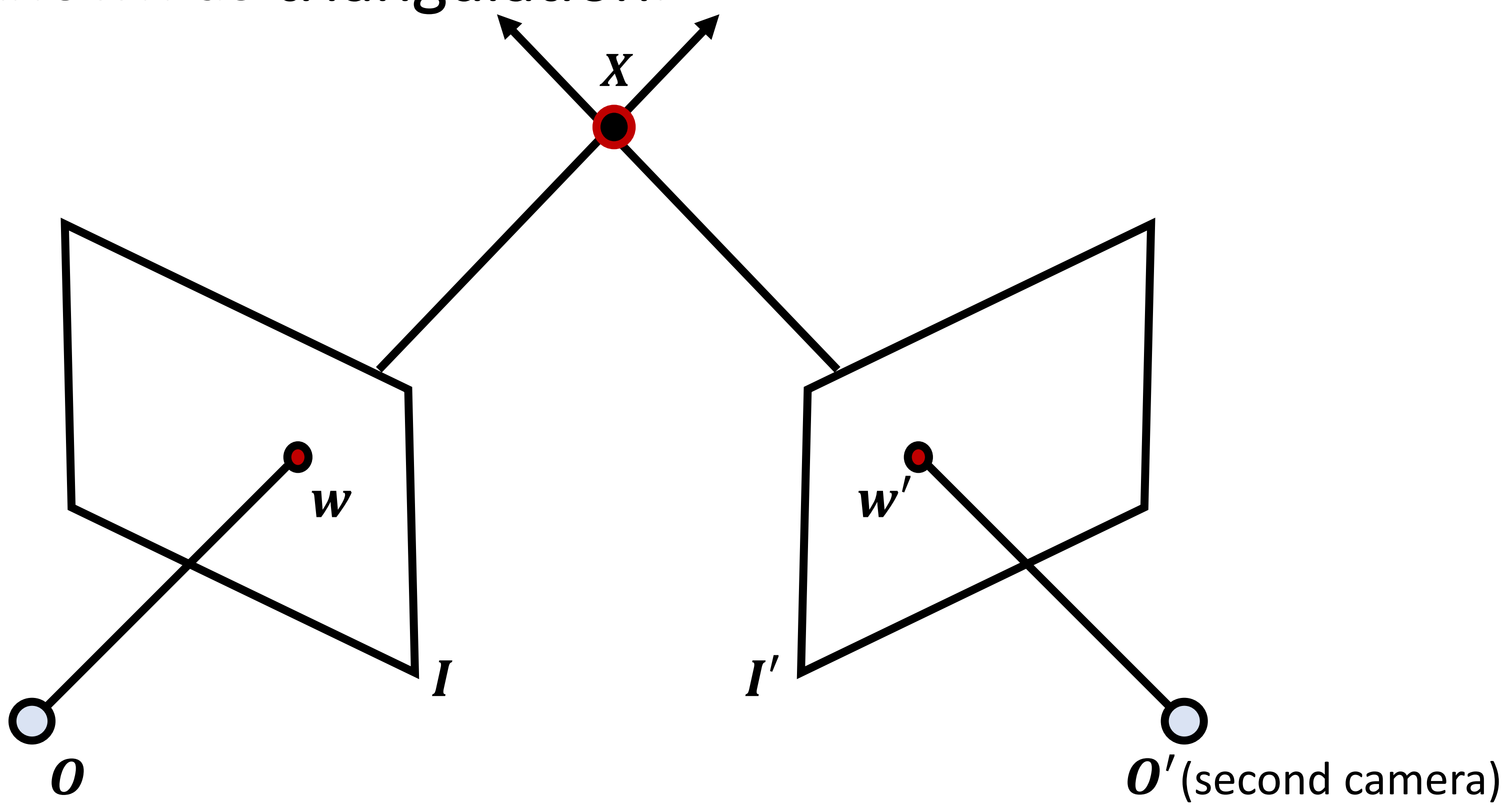
Triangulation

- In stereo vision at least two cameras are set up to view the $3D$ scene.
- Each $3D$ world location $\boldsymbol{X}$ projects to pixel $\boldsymbol{w}$ on camera 1 ($\boldsymbol{O}$) and to pixel $\boldsymbol{w}$' on camera 2 ($\boldsymbol{O}'$).



Triangulation

- If both cameras are calibrated, the $3D$ world location $\boldsymbol{X}$ projected on the pair of corresponding pixel locations $\boldsymbol{w}$ and $\boldsymbol{w}'$ can be estimated via a process known as **triangulation**.



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Triangulation

- Consider the projection of $\boldsymbol{X}$ onto $\boldsymbol{w}$:

$$\widetilde{\boldsymbol{w}} = \boldsymbol{P}\widetilde{\boldsymbol{X}}$$

$$\Leftrightarrow \begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Compute the equation for $\boldsymbol{u}$ and rearrange so our unknowns X, Y, Z are on the left:

$$u = \frac{su}{s} = \frac{p_{11}X + p_{12}Y + p_{13}Z + p_{14}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}}$$

$$\Rightarrow p_{11}X + p_{12}Y + p_{13}Z + p_{14} = p_{31}uX + p_{32}uY + p_{33}uZ + p_{34}u$$

$$\Rightarrow (p_{11} - p_{31}u)X + (p_{12} - p_{32}u)Y + (p_{13} - p_{33}u)Z = p_{34}u - p_{14}$$

Triangulation

- Compute the equation for u and rearrange so our unknowns X, Y, Z are on the left:

$$u = \frac{su}{s} = \frac{p_{11}X + p_{12}Y + p_{13}Z + p_{14}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}}$$

$$\Rightarrow p_{11}X + p_{12}Y + p_{13}Z + p_{14} = p_{31}uX + p_{32}uY + p_{33}uZ + p_{34}u$$

$$\Rightarrow (p_{11} - p_{31}u)X + (p_{12} - p_{32}u)Y + (p_{13} - p_{33}u)Z = p_{34}u - p_{14}$$

- Put this in matrix form:

$$\begin{bmatrix} p_{11} - p_{31}u & p_{12} - p_{32}u & p_{13} - p_{33}u \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{34}u - p_{14} \end{bmatrix}$$

$$\Leftrightarrow \boldsymbol{Ax} = \boldsymbol{b}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Triangulation

- Put this in matrix form:

$$[p_{11} - p_{31}u \quad p_{12} - p_{32}u \quad p_{13} - p_{33}u]\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [p_{34}u - p_{14}]$$
$$\Leftrightarrow \boldsymbol{Ax} = \boldsymbol{b}$$

- $\boldsymbol{A}$ is a $1x3$ matrix

- The resulting system is under-constrained

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Triangulation

- Put this in matrix form:

$$[p_{11} - p_{31}u \quad p_{12} - p_{32}u \quad p_{13} - p_{33}u]\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [p_{34}u - p_{14}]$$

$$[p_{21} - p_{31}v \quad p_{22} - p_{32}v \quad p_{23} - p_{33}v]\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [p_{34}v - p_{24}]$$

$$\Leftrightarrow \boldsymbol{Ax} = \boldsymbol{b}$$

- By computing $\boldsymbol{v}$ in the same way as $\boldsymbol{u}$ and rearranging we can add a new row in $\boldsymbol{A}$ and in $\boldsymbol{b}$, making it $2x3$

Triangulation

- Put this in matrix form:

$$[p'_{11} - p'_{31}u' \quad p'_{12} - p'_{32}u' \quad p'_{13} - p'_{33}u'] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [p'_{34}u' - p'_{14}]$$

$$[p'_{21} - p'_{31}v' \quad p'_{22} - p'_{32}v' \quad p'_{23} - p'_{33}v'] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [p'_{34}v' - p'_{24}]$$

$$\Leftrightarrow \boldsymbol{Ax} = \boldsymbol{b}$$

- By also considering the projection of $\boldsymbol{X}$ onto $\boldsymbol{w'}$, a new pair of rows will be added to $\boldsymbol{A}$, thus forcing it to be $4x3$, i.e. over-constrained

Triangulation

- Here is the resulting system

$$\begin{bmatrix} p_{11} - p_{31}u & p_{12} - p_{32}u & p_{13} - p_{33}u \\ p_{21} - p_{31}v & p_{22} - p_{32}v & p_{23} - p_{33}v \\ p'_{11} - p'_{31}u' & p'_{12} - p'_{32}u' & p'_{13} - p'_{33}u' \\ p'_{21} - p'_{31}v' & p'_{22} - p'_{32}v' & p'_{23} - p'_{33}v' \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{34}u - p_{14} \\ p_{34}v - p_{24} \\ p'_{34}u' - p'_{14} \\ p'_{34}v' - p'_{24} \end{bmatrix}$$

- Where p'_{11} etc. are the parameters inside the camera projection matrix for camera 2 ($\boldsymbol{O}'$), and $\boldsymbol{w}' = (u', v')$ are the pixel coordinates of $\boldsymbol{X}$ projected on the image plane $\boldsymbol{I}'$ of the second camera

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Triangulation

- Using an additional camera has forced $\boldsymbol{A}$ to become over-constrained.

$$\begin{bmatrix} p_{11}-p_{31}u & p_{12}-p_{32}u & p_{13}-p_{33}u \\ p_{21}-p_{31}v & p_{22}-p_{32}v & p_{23}-p_{33}v \\ p'_{11}-p'_{31}u' & p'_{12}-p'_{32}u' & p'_{13}-p'_{33}u' \\ p'_{21}-p'_{31}v' & p'_{22}-p'_{32}v' & p'_{23}-p'_{33}v' \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{34}u-p_{14} \\ p_{34}v-p_{24} \\ p'_{34}u'-p'_{14} \\ p'_{34}v'-p'_{24} \end{bmatrix}$$

- Therefore we can now find a least-squares solution:

$$\boldsymbol{Ax} = \boldsymbol{b}$$
$$\Leftrightarrow \boldsymbol{x} = \left(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}\right)^{-\boldsymbol{1}}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{b}$$

Today’s Agenda

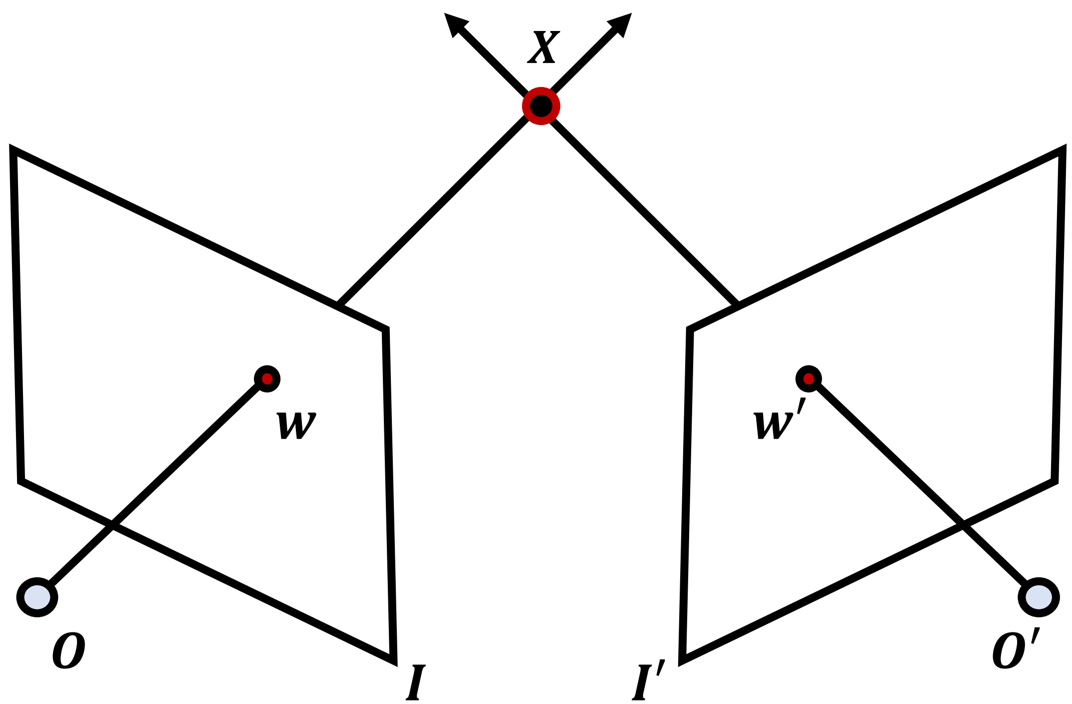
- Recovery of world position
- Triangulation
- Epipolar Geometry

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

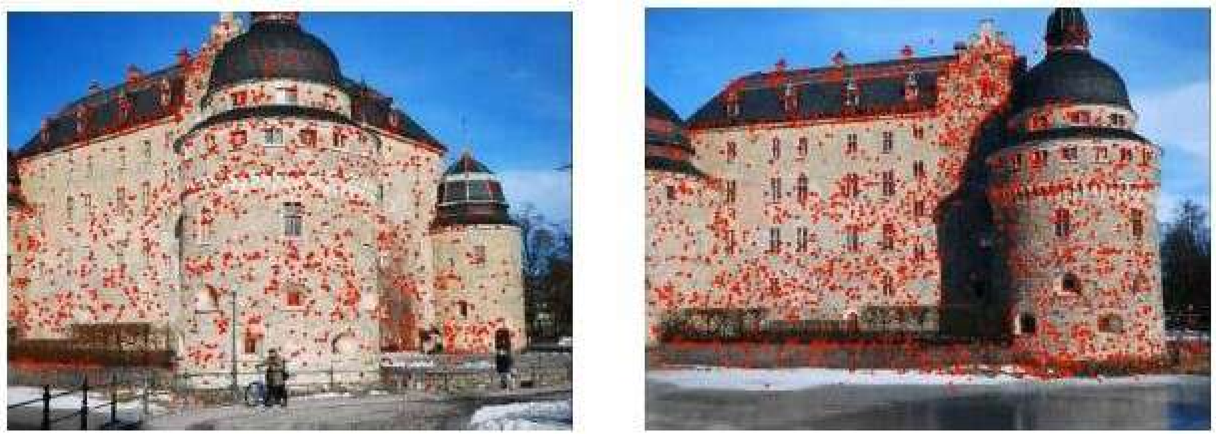
Beyond triangulation

We have seen the simplest form of stereo vision: Given a pair of *calibrated* cameras observing a single feature at *corresponding* pixel locations $\boldsymbol{w} \leftrightarrow \boldsymbol{w}'$, the $3D$ position of the corresponding world location $\boldsymbol{X}$ can be estimated via triangulation



Beyond triangulation

How is the correspondence problem solved if there are several points $\{\boldsymbol{w_i}\}_{i=1}^{N_1}$ in image 1 and several points $\{\boldsymbol{w'_j}\}_{j=1}^{N_2}$ in image 2 ?



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Beyond triangulation

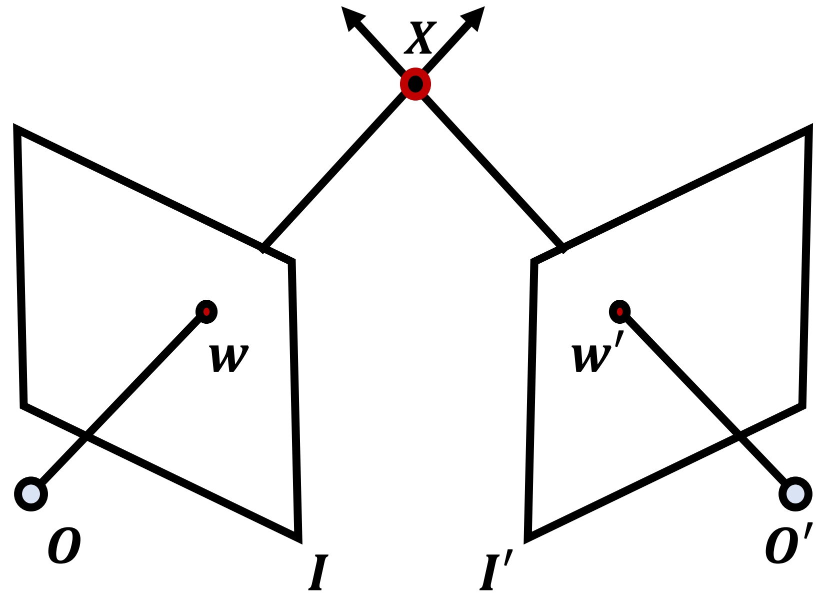
SIFT will give us a set of proposed correspondences $\{\boldsymbol{w_i} \leftrightarrow \boldsymbol{w'_j}\}$ but there will be **many** outliers in these proposals

The question is then how can we remove outliers ?

Using the **epipolar constraint**

Epipolar Geometry

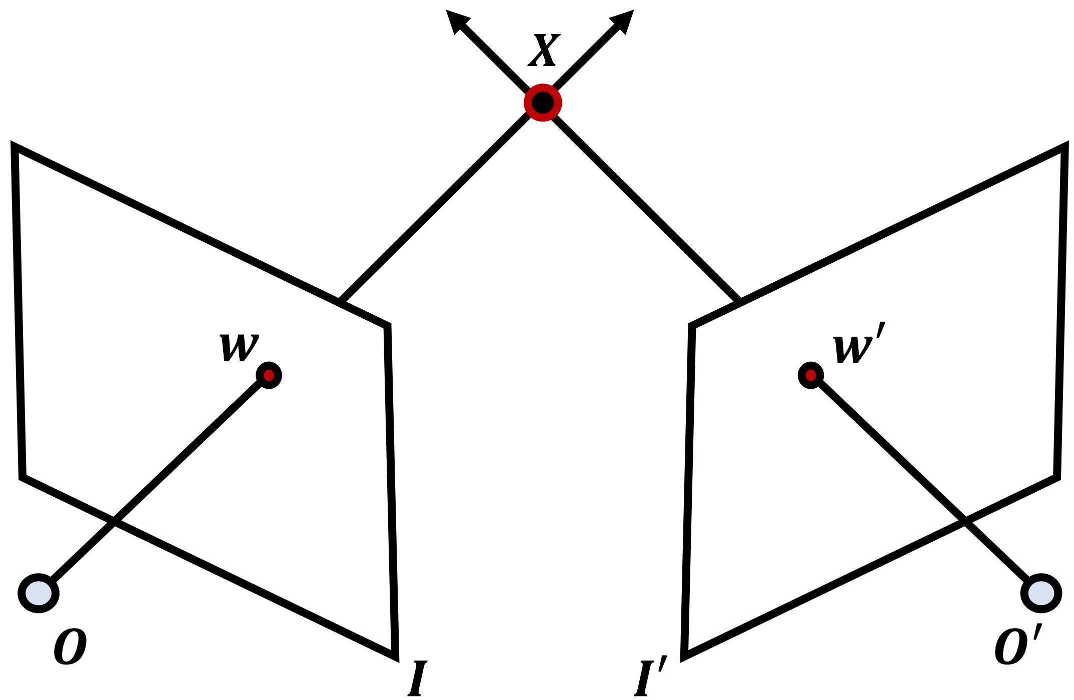
- To understand the **epipolar constraint**, we first need to understand the **geometry** that relates the
  - **cameras**
  - **points in $\mathbf{3D}$** space
  - and their corresponding observations $\{\mathbf{w}_i \leftrightarrow \mathbf{w'}_j\}$



- This type of geometry is referred to as the **epipolar geometry**

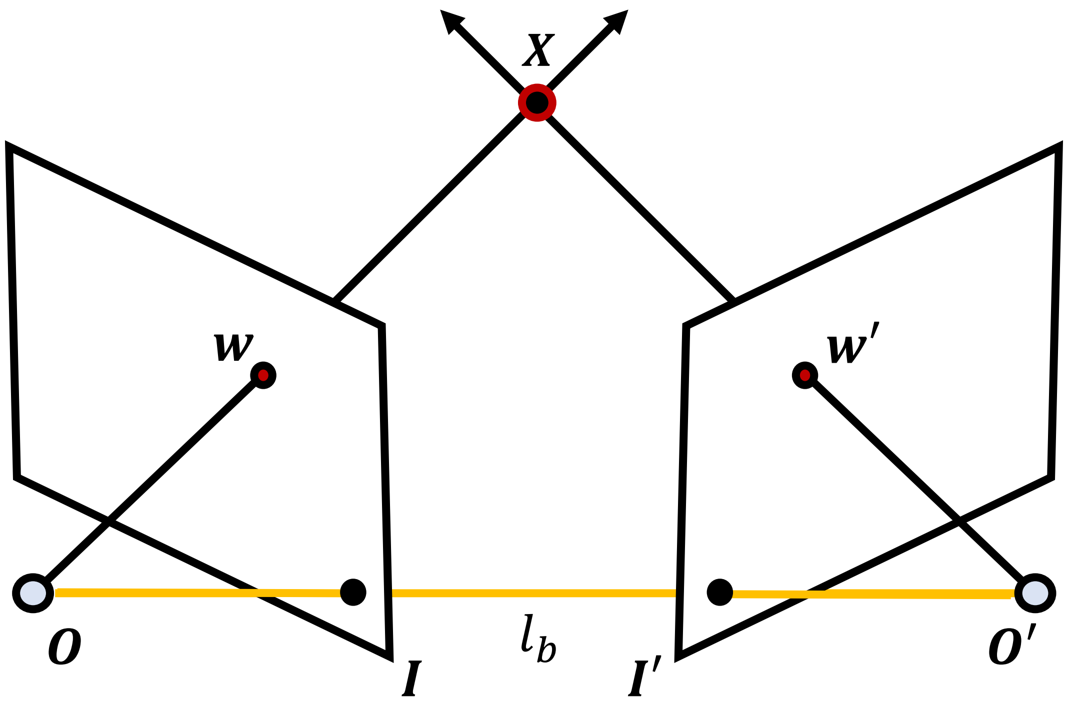
Epipolar Geometry - fundamentals

Lets revisit our **stereo pair**



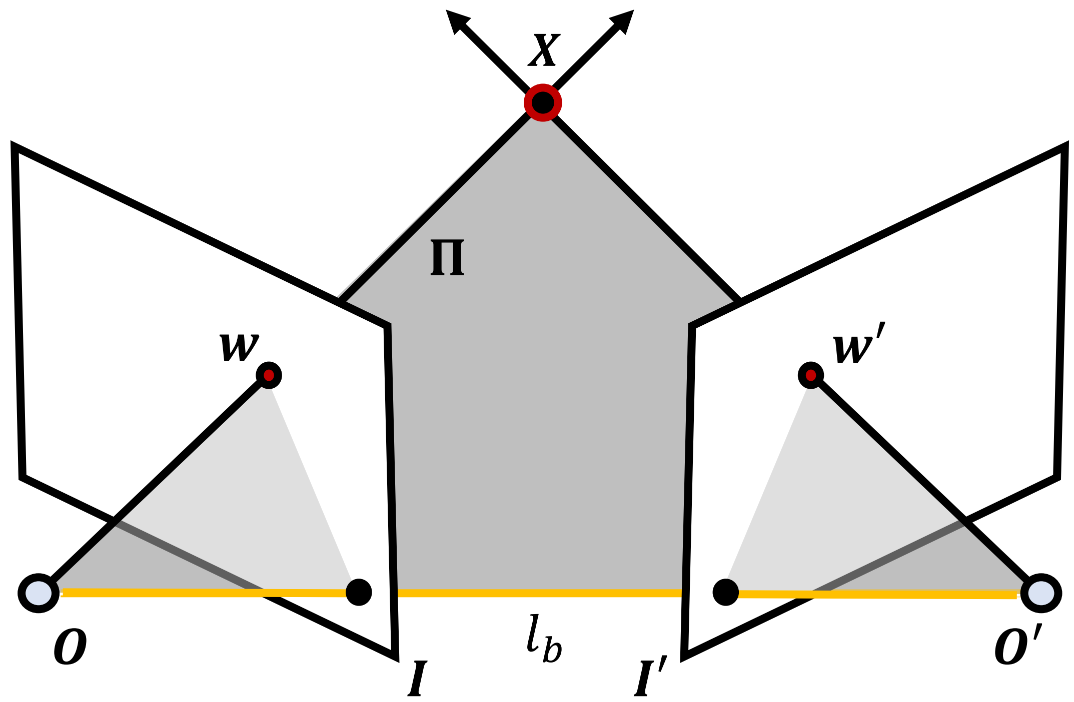
Epipolar Geometry - fundamentals

The baseline l_b is the line joining the two optical centers.



Epipolar Geometry - fundamentals

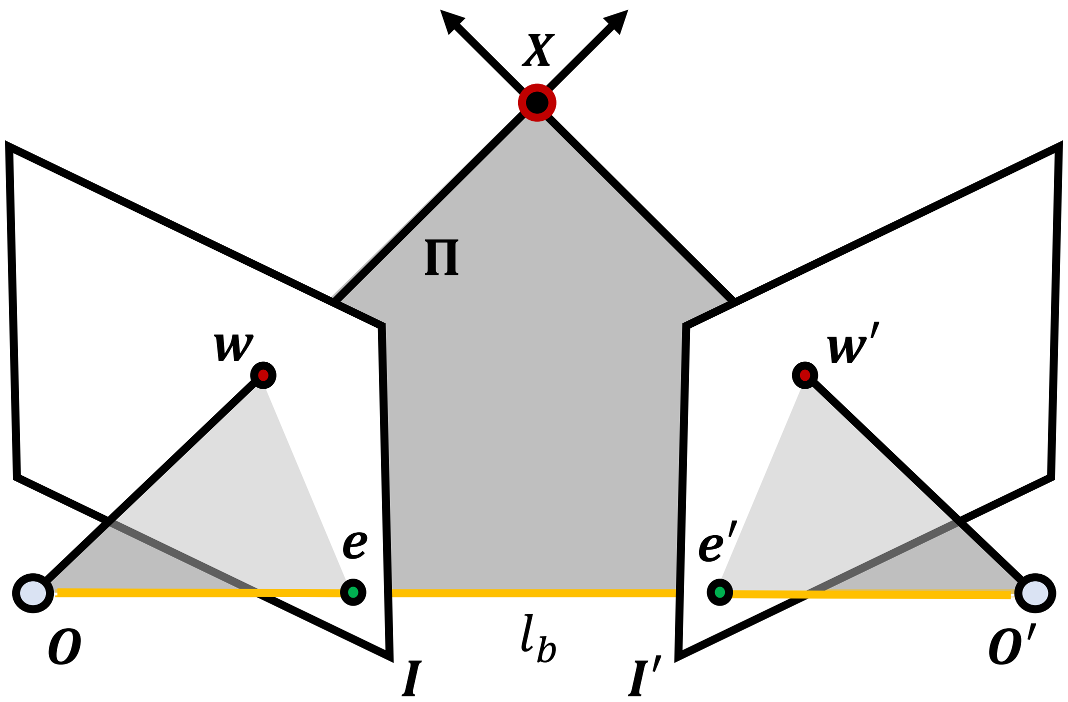
The epipolar plane $\boldsymbol{\Pi}$ is the plane defined by the $3D$ point $\boldsymbol{X}$ and the optical centers of the cameras.



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

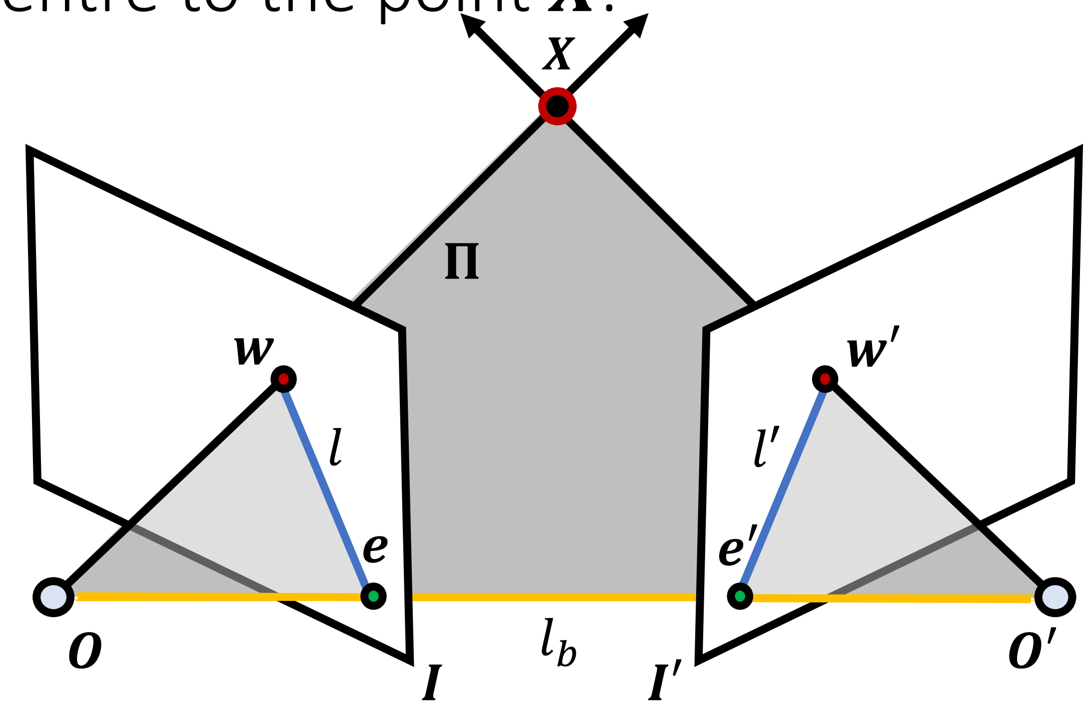
Epipolar Geometry - fundamentals

An epipole is the point of intersection of the baseline with the image plane. There are two epipoles $\boldsymbol{e}$ and $\boldsymbol{e'}$, one for each image.



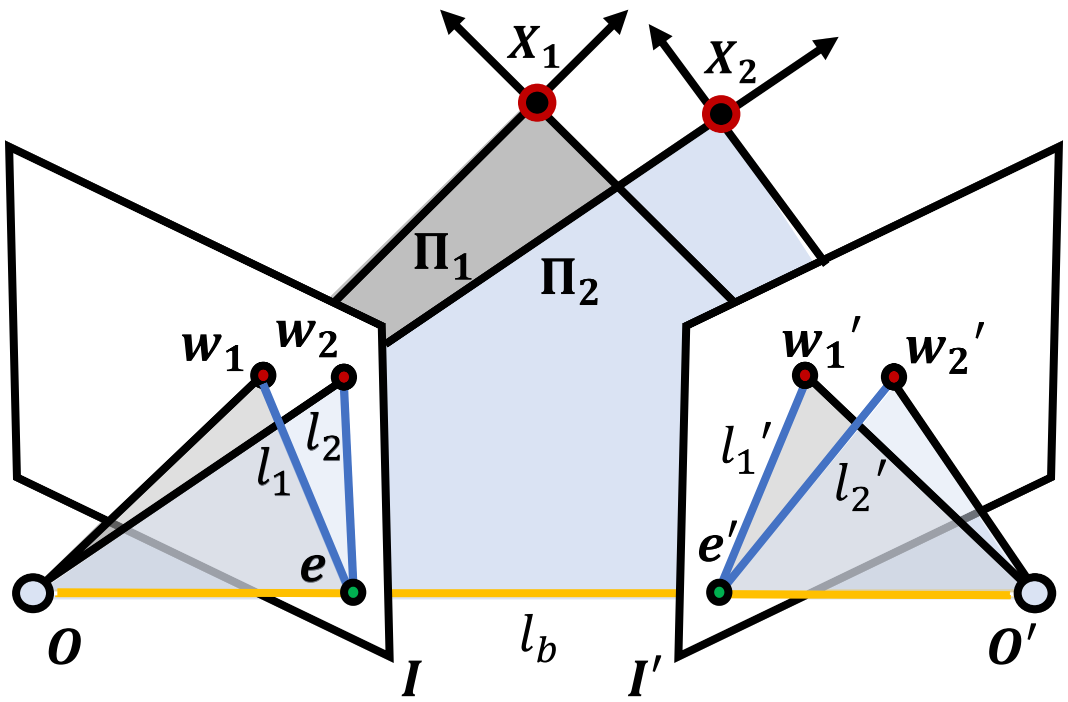
Epipolar Geometry - fundamentals

An epipolar line is a line of intersection of the epipolar plane with an image plane. It is the image, in one camera, of the ray from the other camera's optical centre to the point $\boldsymbol{X}$.



Epipolar Geometry - fundamentals

For different world points $\boldsymbol{X}$, the epipolar plane rotates about the baseline. All epipolar lines intersect at their corresponding epipole.

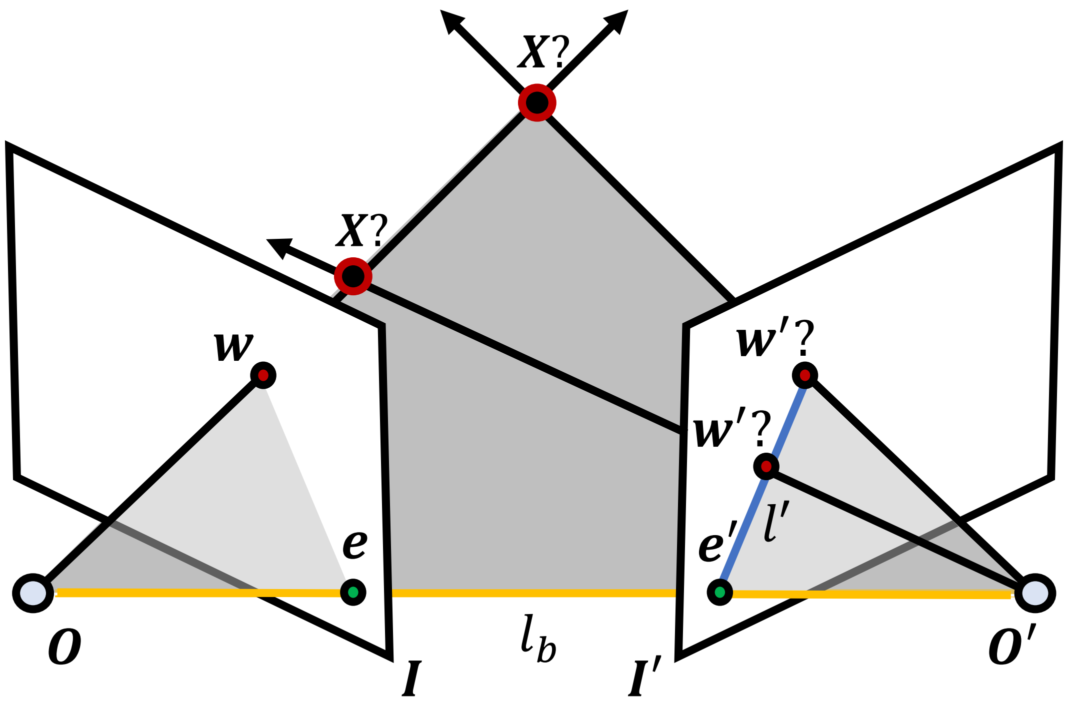


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

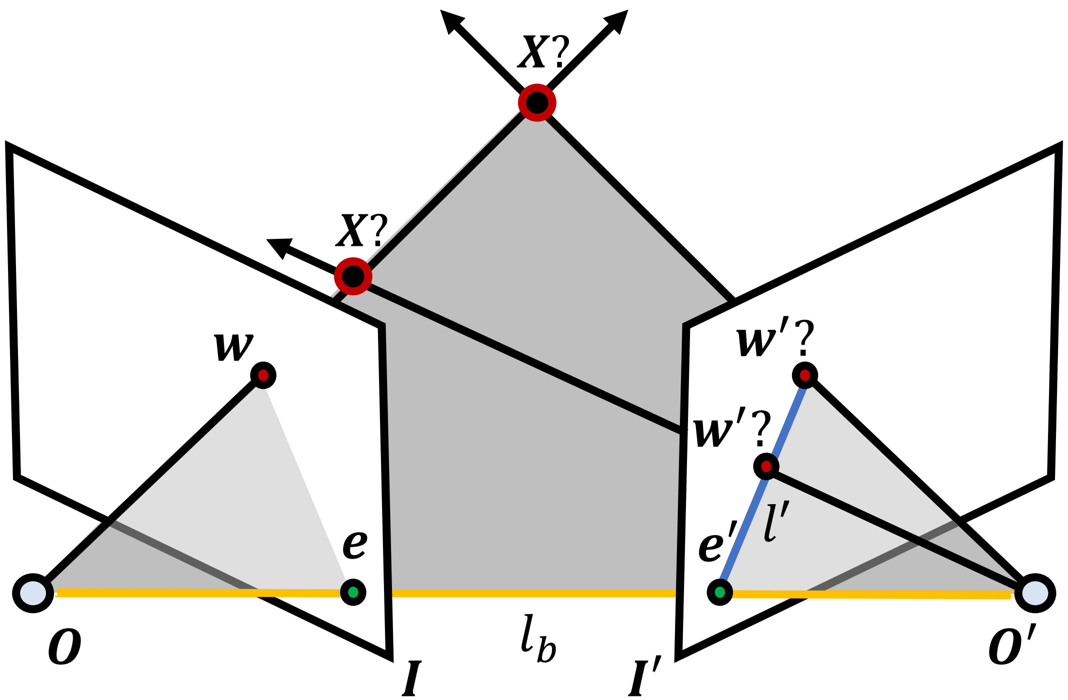
Epipolar Geometry - fundamentals

The **epipolar constraint** limits the search for correspondences, from the region of the whole image, to only the pixels spanned by the epipolar line.



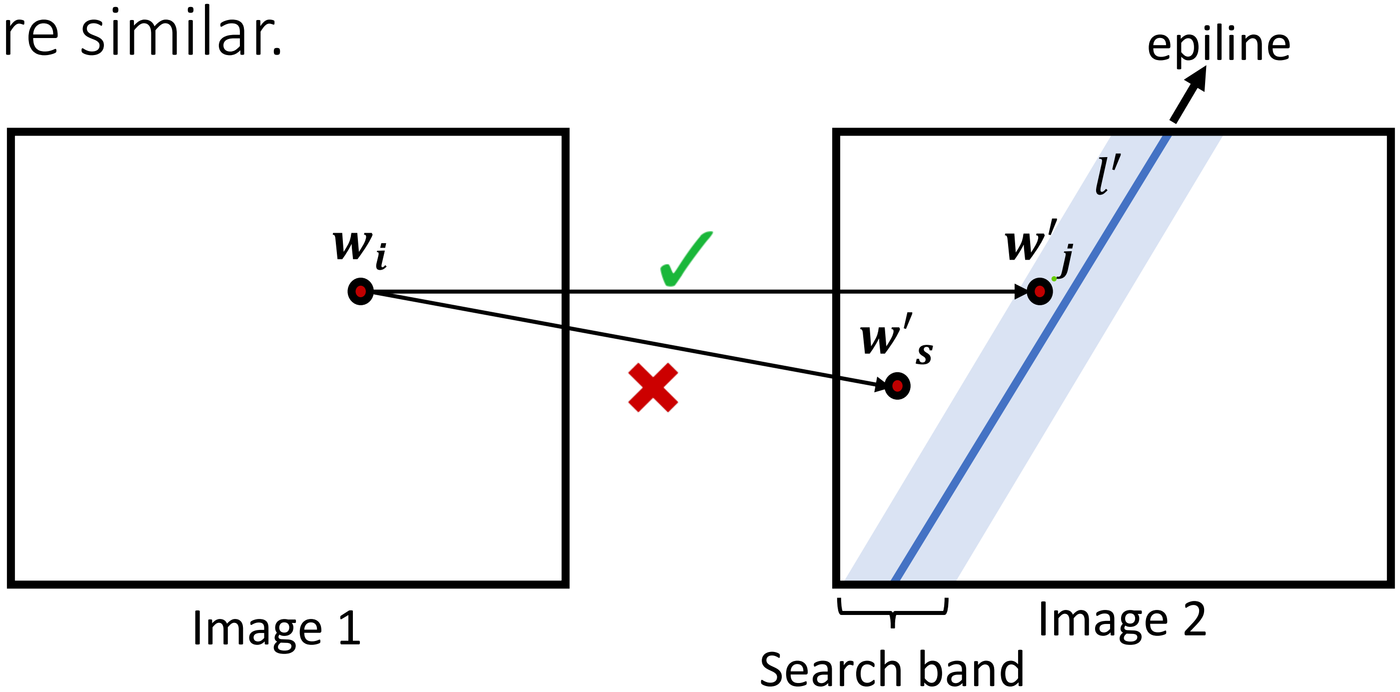
Epipolar Geometry - fundamentals

If a point feature $\boldsymbol{w}$ is observed in one image, then its location $\boldsymbol{w}'$ in the other image must lie on its corresponding epipolar line l'



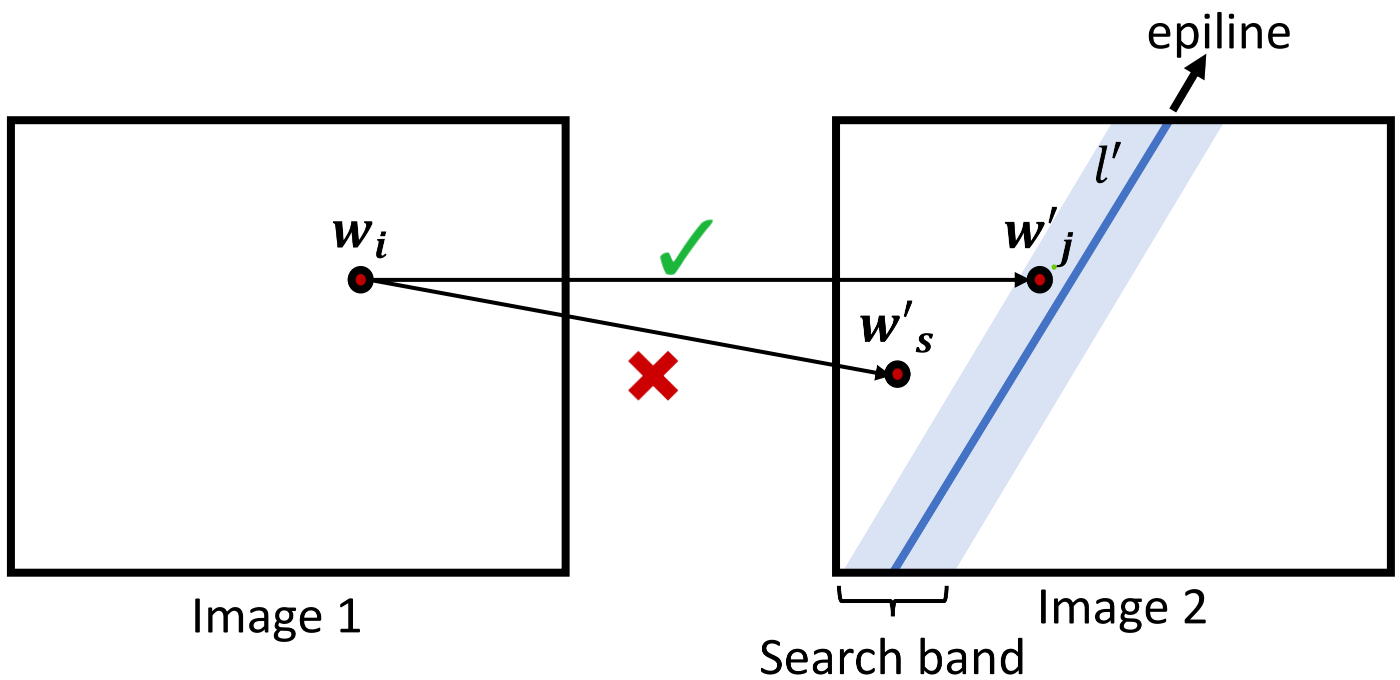
Epipolar Geometry - fundamentals

So a simple algorithm for determining correspondences is to match each feature $\boldsymbol{w_i}$ from camera 1 to a feature $\boldsymbol{w'_j}$ from camera 2 which is close to the epipolar line of camera 2, provided that the SIFT descriptors for $\boldsymbol{w_i}$ and $\boldsymbol{w'_j}$ are similar.



Epipolar Geometry - fundamentals

Therefore, we must calculate the equation of the epipolar lines.



Essential Matrix

- Firstly, lets assume that the 1st camera is located at the **world origin $\boldsymbol{O}$**

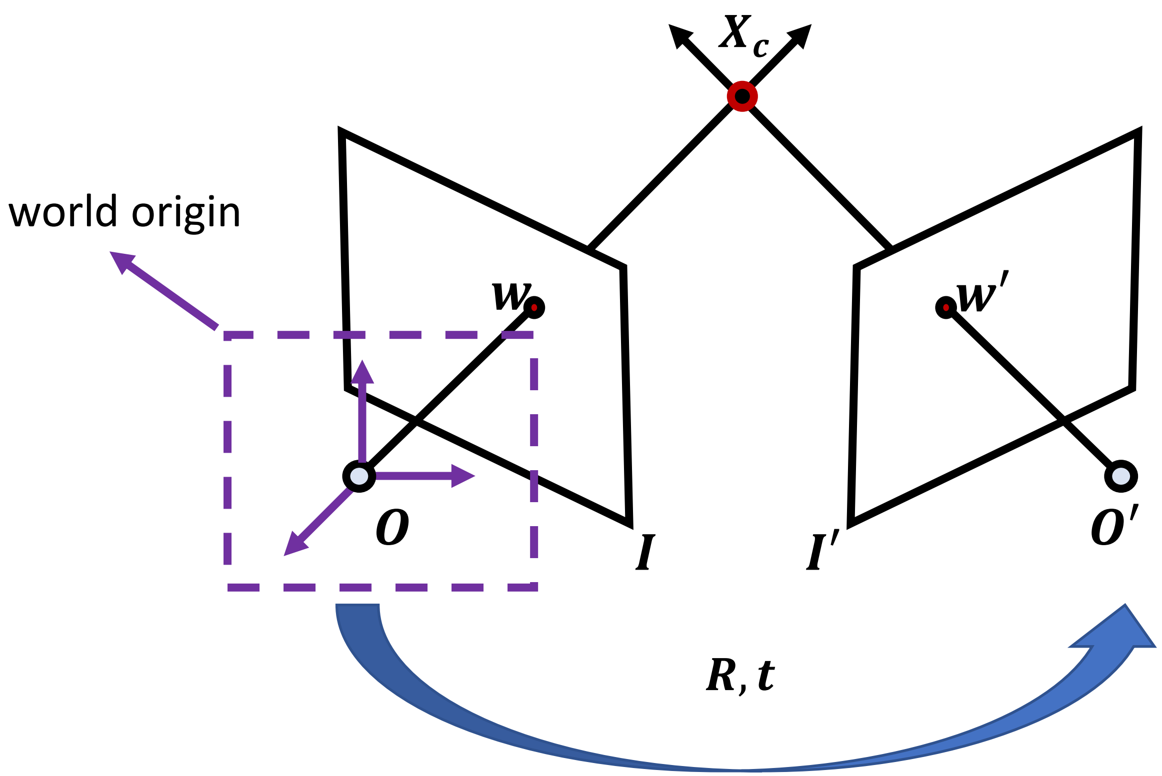
- This means that every point $\boldsymbol{X}$ is expressed in the camera coordinates of the 1st camera, since its optical center coincides with the world origin $\Rightarrow \boldsymbol{X} \rightarrow \boldsymbol{X_c}$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Essential Matrix

- If this is the case, the 2nd camera is at a location $\boldsymbol{O'}$ in world coordinates, which can be expressed by a **translation $\boldsymbol{t}$** and **rotation $\boldsymbol{R}$**, w.r.t. 1st camera, i.e., the world origin



Co-financed by the European Union
Connecting Europe Facility
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Essential Matrix

- Every $3D$ point $\boldsymbol{X}$ can be expressed as $\boldsymbol{X_c}$, which is the camera-centered coordinate system of the 1st camera, and as $\boldsymbol{X'_c}$, which is the camera-centered coordinate system of the 2nd camera.

- Since $\boldsymbol{X} = \boldsymbol{X_c}$, we can relate $\boldsymbol{X} \leftrightarrow \boldsymbol{X'_c}$ or $\boldsymbol{X_c} \leftrightarrow \boldsymbol{X'_c}$ using a **Euclidean transformation** composed by the **translation** $\boldsymbol{t}$ and **rotation** $\boldsymbol{R}$ of the 2nd camera, w.r.t. the 1st camera

Essential Matrix

Here is how to find an expression for the epipolar line:

$$\widetilde{X'_c} = P_e\widetilde{X_c}$$ ← This is in homogeneous coordinates

$$\Leftrightarrow X'_c = RX_c + t$$ ← Express it in cartesian coordinates

$$\Leftrightarrow t \times X'_c = t \times RX_c + \cancelto{0}{t \times t}$$ ← Apply cross product with t to both sides

$$\Leftrightarrow X'_c \cdot (t \times X'_c) = X'_c \cdot (t \times RX_c)$$ ← Apply dot product with X'_c to both sides

$$\Leftrightarrow 0 = X'_c \cdot (t \times RX_c)$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Essential Matrix

This can be rewritten in matrix form:

$$\boldsymbol{X}_{\boldsymbol{c}}' \cdot (\boldsymbol{t} \times \boldsymbol{R}\boldsymbol{X}_{\boldsymbol{c}}) = \boldsymbol{0}$$

$$\Leftrightarrow \boldsymbol{X}_{\boldsymbol{c}}'^{T} \boldsymbol{E} \boldsymbol{X}_{\boldsymbol{c}} = \boldsymbol{0}$$

Where $\boldsymbol{E} = \boldsymbol{T}_{\times}\boldsymbol{R}$ is the **essential matrix,** and

$$\boldsymbol{T}_{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

is a matrix representing the cross product with $\boldsymbol{t}$ such that $\mathbf{t} \times \boldsymbol{v} = \boldsymbol{T}_{\times}\mathbf{v}$

Fundamental Matrix

Recall that for $\boldsymbol{X_c} \leftrightarrow \boldsymbol{w}$: $\widetilde{\boldsymbol{w}} = \boldsymbol{K}\boldsymbol{X_c} \Leftrightarrow \boldsymbol{X_c} = \boldsymbol{K}^{-\mathbf{1}}\widetilde{\boldsymbol{w}}$

and similarly, for $\boldsymbol{X}'_{\boldsymbol{c}} \leftrightarrow \boldsymbol{w}'$: $\widetilde{\boldsymbol{w}}' = \boldsymbol{K}'\boldsymbol{X}'_{\boldsymbol{c}} \Leftrightarrow \boldsymbol{X}'_{\boldsymbol{c}} = \boldsymbol{K}'^{-\mathbf{1}}\widetilde{\boldsymbol{w}}'$

Combining the two equations yields the equation of **the two epipolar lines in pixel coordinates**:

$$\boldsymbol{X}'^{T}_{\boldsymbol{c}}\,\mathbf{E}\,\mathbf{X_c} = 0$$

$$\Rightarrow (\boldsymbol{K}'^{-\mathbf{1}}\widetilde{\boldsymbol{w}}')^{T}\,\boldsymbol{E}\,(\boldsymbol{K}^{-\mathbf{1}}\widetilde{\boldsymbol{w}}) = 0$$

$$\Rightarrow \widetilde{\boldsymbol{w}}'^{T}(\mathbf{K}'^{-T}\,\mathbf{E}\,\mathbf{K}^{-\mathbf{1}})\,\widetilde{\boldsymbol{w}} = 0$$

$$\Rightarrow \widetilde{\boldsymbol{w}}'^{T}\boldsymbol{F}\,\widetilde{\boldsymbol{w}} = \mathbf{0}$$

where $\boldsymbol{F} = \boldsymbol{K}'^{-T}\boldsymbol{E}\boldsymbol{K}^{-\mathbf{1}}$ is the **fundamental matrix**.

Fundamental Matrix

This is the equation of the epipolar line in either camera:

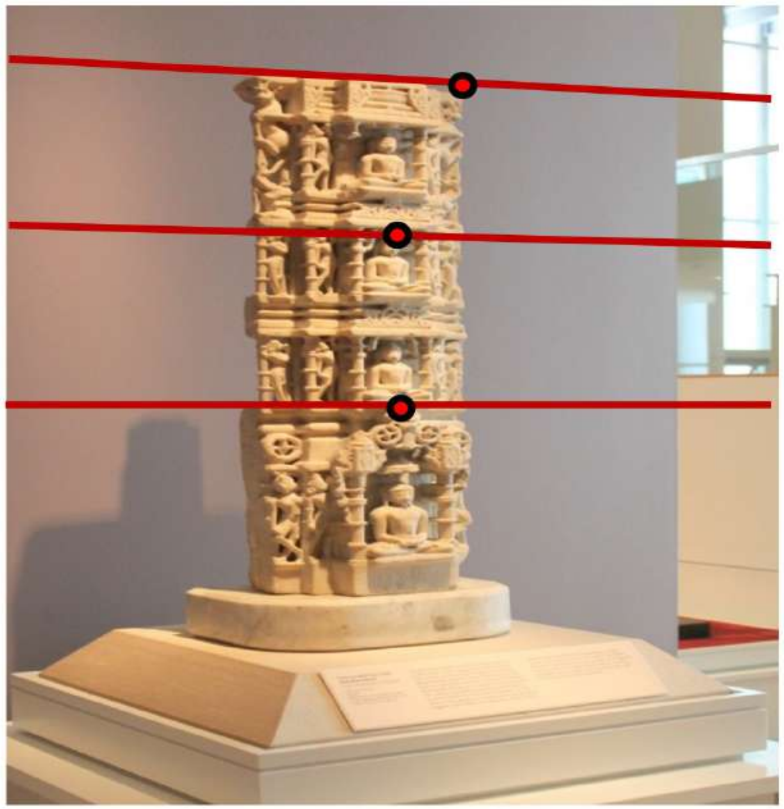
$$\tilde{\boldsymbol{w}}'^T \boldsymbol{F}\, \tilde{\boldsymbol{w}} = \boldsymbol{0}$$

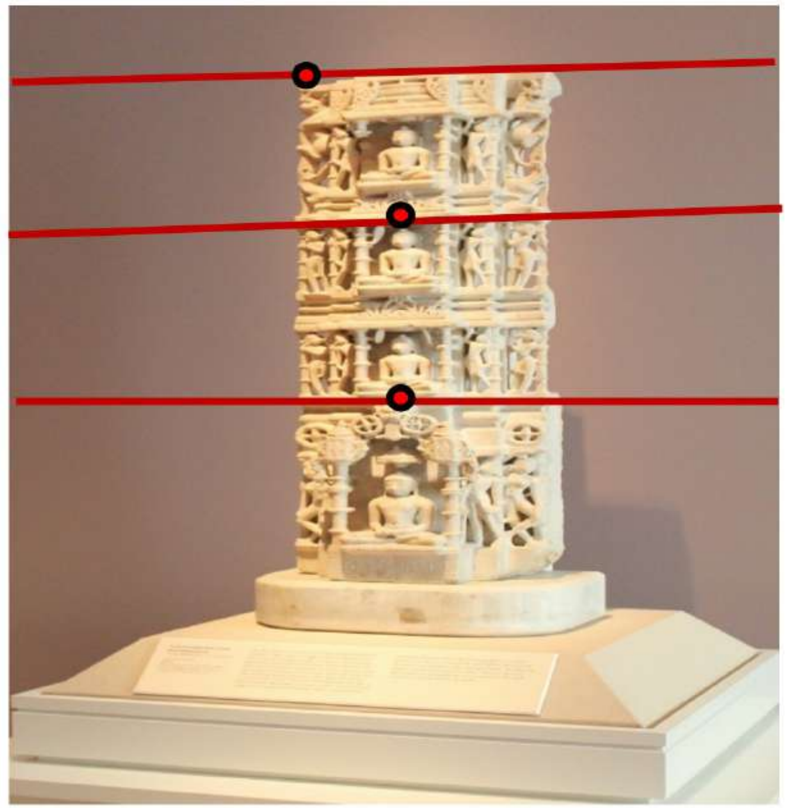
Assuming we know the fundamental matrix, for every point $\boldsymbol{w}$ in image 1, this expression gives us the line in image 2 on which the corresponding $\boldsymbol{w}'$ must lie, and *vice versa.*

- $l' = \boldsymbol{F}\tilde{\boldsymbol{w}}$ is the epipolar line in the 2nd image, associated with $\boldsymbol{w}$
- $l = \boldsymbol{F}^T\tilde{\boldsymbol{w}}'$ is the epipolar line in the 1st image, associated with $\boldsymbol{w}'$
- $l_i: ax + by + c = 0$

Examples

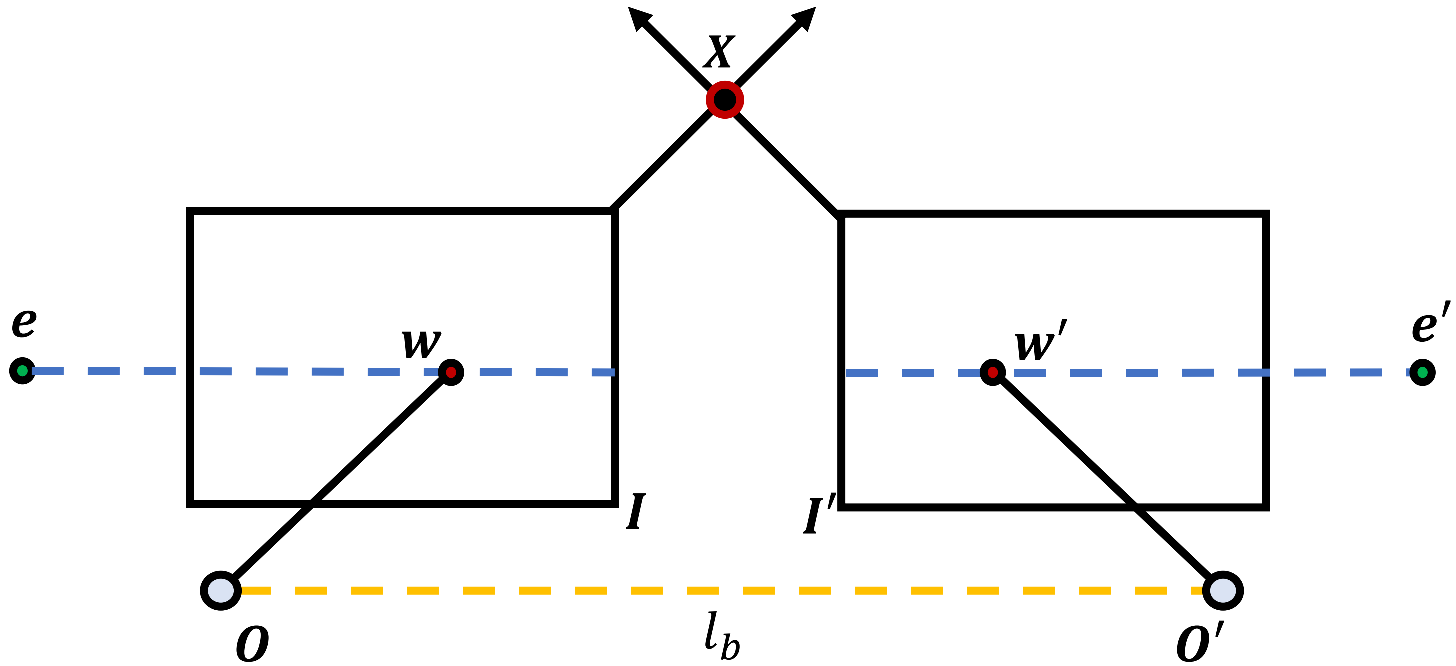
Here are a few examples of the epipolar constraint





Examples

Epipolar constraint examples: **Parallel image planes**



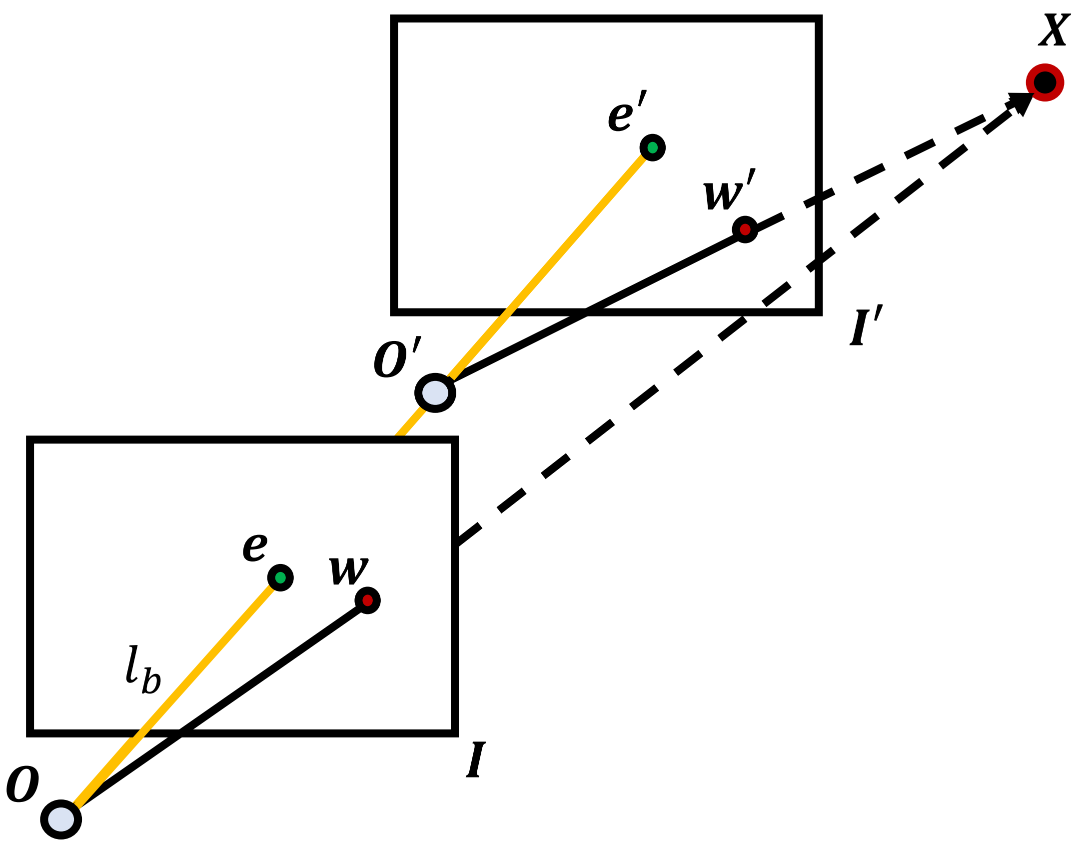
- Baseline intersects the image planes at infinity
- Epipoles are at infinity
- Epipolar lines are parallel to the $\boldsymbol{u}$-axis of each image plane

Examples

Epipolar constraint examples: **Parallel image planes**

Examples

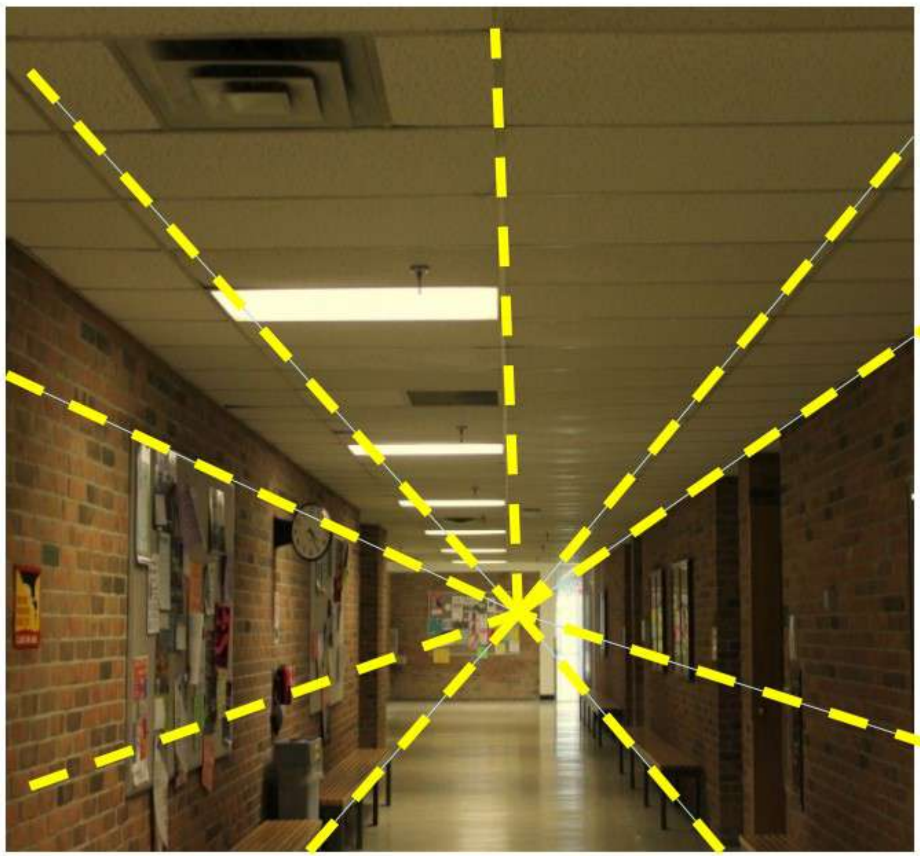
Epipolar constraint examples: **Forward translation**

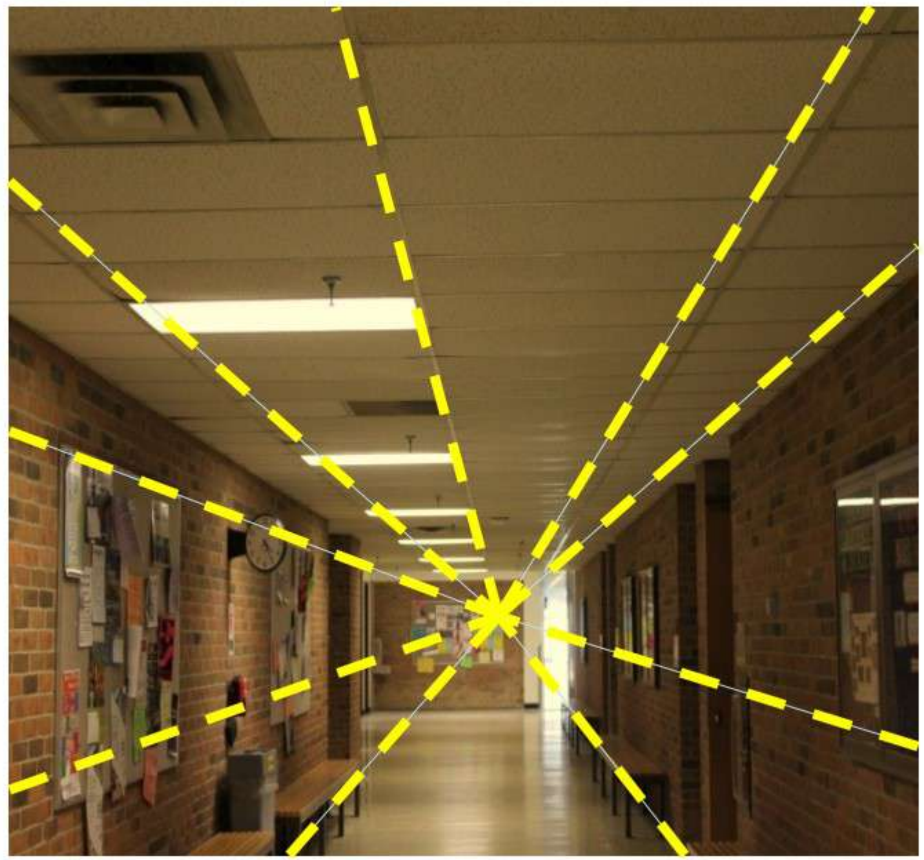


- The epipoles have the same position in both images

Examples

Epipolar constraint examples: **Forward translation**





Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Fundamental Matrix

In order to apply the epipolar constraint we need to know the fundamental matrix:

$$\begin{aligned}\boldsymbol{F} &= \boldsymbol{K}'^{-T}\boldsymbol{E}\boldsymbol{K}^{-1}\\ &= \boldsymbol{K}'^{-T}\boldsymbol{T}_{\times}\boldsymbol{R}\boldsymbol{K}^{-1}\end{aligned}$$

All the parameters of $\boldsymbol{F}$ come from the calibration of the two cameras.

Remember that the perspective camera model $\boldsymbol{P_{ps}} = \boldsymbol{K}[\boldsymbol{R}|\boldsymbol{T}]$ contains this information.

If, however, these are not available (e.g., because we used a projective camera model to calibrate our cameras), $\boldsymbol{F}$ must be estimated using known image correspondences.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Estimating the fundamental matrix

To estimate the fundamental matrix, we follow a similar approach as in camera calibration

$$\widetilde{\boldsymbol{w}}'^{T} \boldsymbol{F}\, \widetilde{\boldsymbol{w}} = 0 \Rightarrow [u' \ \mathrm{v}' \ 1] \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

Each point correspondence $\boldsymbol{w} \leftrightarrow \boldsymbol{w}'$ generates a single equation for estimating the parameters inside $\boldsymbol{F}$.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Estimating the fundamental matrix

Here are N such correspondences:

$$\begin{bmatrix} u_1u'_1 & v_1u'_1 & u'_1 & u_1v'_1 & v_1v'_1 & v'_1 & u_1 & v_1 \\ & & & \vdots & & & & \\ u_Nu'_N & v_Nu'_N & u'_N & u_Nv'_N & v_Nv'_N & v'_N & u_N & v_N \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \end{bmatrix} = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}$$

The minimal number for N is 8, because $\boldsymbol{F}$ has 8 degrees of freedom (we can set $f_{33} = 1$)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Estimating the fundamental matrix

- In practise, 8 clearly indicated correspondences are never available

- What is available is a set of correspondences proposed by SIFT, for a large set of features, in which there are **always** errors

Estimating the fundamental matrix

- We can use RANSAC to solve this problem
  1. Obtain 8 random SIFT correspondences. Use them to estimate $\boldsymbol{F}$.
  2. For every feature $\boldsymbol{w_i}$ in image 1 calculate the epipolar line in image 2. Check if the corresponding feature $\boldsymbol{w'_j}$ in image 2 proposed by SIFT falls **"close"** to the epipolar line. Count the number S of such "inliers".
  3. If $S \geq T$ where T is a threshold, then there is consensus with the random sample taken in the first step. Calculate $\boldsymbol{F}$ for all inliers and terminate here.
  4. If $S < T$ then no consensus is reached. Repeat from step 1.
  5. If after N iterations no consensus is reached, select the model that gave the highest S, calculate $\boldsymbol{F}$ using all inliers in S and terminate

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU Master programmes in Artificial Intelligence 4 Careers in Europe

Visual Computing Group

CYENS CENTRE OF EXCELLENCE

Thank you.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423