Computer Vision (Widzenie Komputerowe)

Module M1:
Introduction and fundamentals

Krzysztof Krawiec
http://www.cs.put.poznan.pl/kkrawiec/

Poznan University of Technology, 2021-2024

Essentials

- Major: Artificial Intelligence
- Volume: Lecture (2 units/week) + labs/project (2 units/week)
- Grading:
  - Lecture: written exam (theory + small tasks)
  - Lab
- IT Platform: Moodle (eKursy)
- Staff:
  - Krzysztof Krawiec, prof. dr hab. inż., krawiec@cs.put.poznan.pl
  - Piotr Wyrwiński, mgr inż.
- Lecture = theory intertwined with use cases and applications
- Relations to other courses
  - Machine learning
  - Deep learning
  - Artificial intelligence

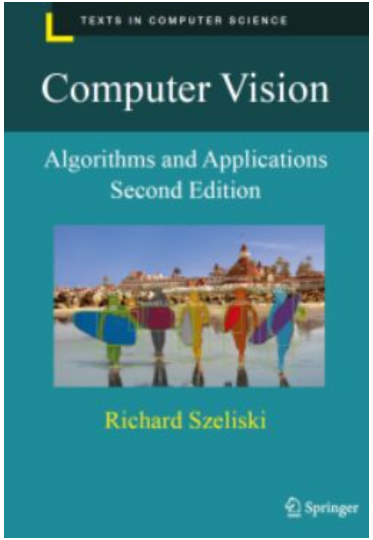
Bibliography

Basic/essential

- Gonzalez, Wintz, Digital Image Processing. Addison-Wesley 2017.
- Szeliski, Computer Vision: Algorithms and Applications, 2nd ed., Springer 2022.
  - Available online (requires registration): https://szeliski.org/Book/

Facultative (in Polish)

- Domański, M., Obraz cyfrowy. WKŁ 2010.
- Zieliński, T.P., Cyfrowe przetwarzanie sygnałów. WKŁ 2009.
- Cyganek, B., Komputerowe przetwarzanie obrazów trójwymiarowych. EXIT 2002.
- Owen, M., Przetwarzanie sygnałów w praktyce. WKŁ 2009.
- Choraś, R. Komputerowa wizja. Metody interpretacji i identyfikacji obiektów. EXIT, 2006.



Bibliography (in Polish)

- Kaehler, A., Bradski, G., OpenCV 3: Komputerowe rozpoznawanie obrazu w C++ przy użyciu biblioteki OpenCV, Helion, 2017
- Watkins, C.D., Sadun, A., Marenka, S.: Nowoczesne metody przetwarzania obrazu. WNT 1995.
- Wojciechowski, K., Rozpoznawanie obrazów, Politechnika Śląska, Gliwice 1997.
- Pavlidis, T., Grafika i przetwarzanie obrazów. Algorytmy, (Seria: Biblioteka Inżynierii Oprogramowania), WNT, Warszawa 1987.
- Tadeusiewicz, R., Systemy wizyjne robotów przemysłowych, WNT, Warszawa 1992.
- Tadeusiewicz, R., Flasiński, M., Rozpoznawanie obrazów, (Seria: Współczesna Nauka i Technika. Informatyka) PWN, Warszawa 1991.
- Ostrowski, M. (red.), Informacja obrazowa. WNT, Warszawa, 1992.
- Nałęcz, M., Problemy biocybernetyki i inżynierii biomedycznej, WKiŁ, Warszawa 1990.
- Nieniewski, M., Morfologia matematyczna w przetwarzaniu obrazów, Akademicka Oficyna Wydawnicza, Warszawa 1998.
- Zieliński, K.W., Strzelecki, M.: Komputerowa analiza obrazu biomedycznego. Wstęp do morfometrii i patologii ilościowej, Wydawnictwo Naukowe PWN, 2002.

Software libraries and tools

Libraries:

- OpenCV, Numpy, Scikit
- (Tensorflow+Keras)

Interactive tools:

- ImageJ/Fiji https://imagej.net/Fiji
- Jupyter notebooks

Other tools:

- ImageMagick

Programming language: Python

- (though notice the usefulness of C/C++ implementations in some contexts)

Course agenda

Modules:

M1: Introduction and fundamentals

M2: Image representation

M3: Image processing

M4: Image description (features, descriptors, etc.)

M5: [Image segmentation]

~~M6: Convolutional Neural Networks~~

M7: Depth (Interpretation of 3D scenes)

M8: Image acquisition*

~~M9: [Motion]~~

*Postponed to facilitate synchronization with the labs.

M1: Introduction

Definition

Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do.

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

Related (?) discipline: Pattern recognition

Pattern recognition:

- *Pattern recognition, a field concerned with machine recognition of meaningful regularities in noisy or complex environments* – Duda & Hart (1973)
- *Pattern recognition can be defined as the categorization of input data into identifiable classes via the extraction of significant features or attributes of the data from a background of irrelevant detail* – Gonzalez and Thomason (1978)
- *Pattern recognition is a search for structure in data* – Bezdek (1981)
- *Pattern recognition is the science that concerns the description or classification (recognition) of measurements* – Schalkoff (1992)

Notice: Image (or image content) is a special case of pattern (spatial pattern).

- But not every pattern is an image (example: time series)
- Historically, CV focused on interpretations of 3D scenes and related aspects (depth, perspective, pose, views, etc.)

Related disciplines/concepts

- Pattern Recognition
  - Historically more focused on rasters/tensors; less into 3D
- Signal Processing:
  - Considers images to be signals defined over certain space (2D or more-dimensional)
- Machine Vision:
  - Focused on industrial applications, in particular Quality Control (QC)
- Computer Graphics:
  - Motivation: vision as inverse graphics
- Artificial Intelligence, including in particular:
  - Computational Intelligence
  - Machine Learning

Relations outside STEM:

- Cognitive Science, Psychology

Motivations

- Vision is the primary source of perception in primates.
  - Estimated contribution: 80%
  - Quick acquisition and learning
- Universality:
  - We live in a 3D space full of objects that radiate/emit/reflect electromagnetic waves
- Wide spectrum of applications
  - And: wide spectrum of problems/tasks that can be approached with CV
- Relatively low cost (currently often ridiculously low) of:
  - Equipment (sensor and optics); e.g. currently in the order of single digit $ in smartphones
  - Transmision
  - Storage
  - Rendering/presentation

Basic concepts

Images are formed as a result of a medium interacting with scenes.

- *Scene*: a set of physical objects
- *Image*: a piece of information reflecting the state of the scene, acquired using:
  - a <u>form</u> of interaction with the scene (emission, transmission, reflection, ...)
  - a <u>medium</u> (electromagnetic waves, mechanical waves, mechanical interaction, ...)
  - a <u>device</u> (sensor, optics, …)
  - from a <u>vantage point</u> (position and orientation of a camera),
  - in some <u>environment</u> (sources of light/radiation, disturbances, noise, ...),
  - at certain point in <u>time</u>.

Implication:

- Many-to-many relationship:
  - We can acquire multiple images from the same scene.
  - Various scenes can give rise to identical images.

Stages of processing in CV

Lighting

Scene

Reflection/transmission/emission

Analog image

Image acquisition

Digital image

Image processing

Digital image

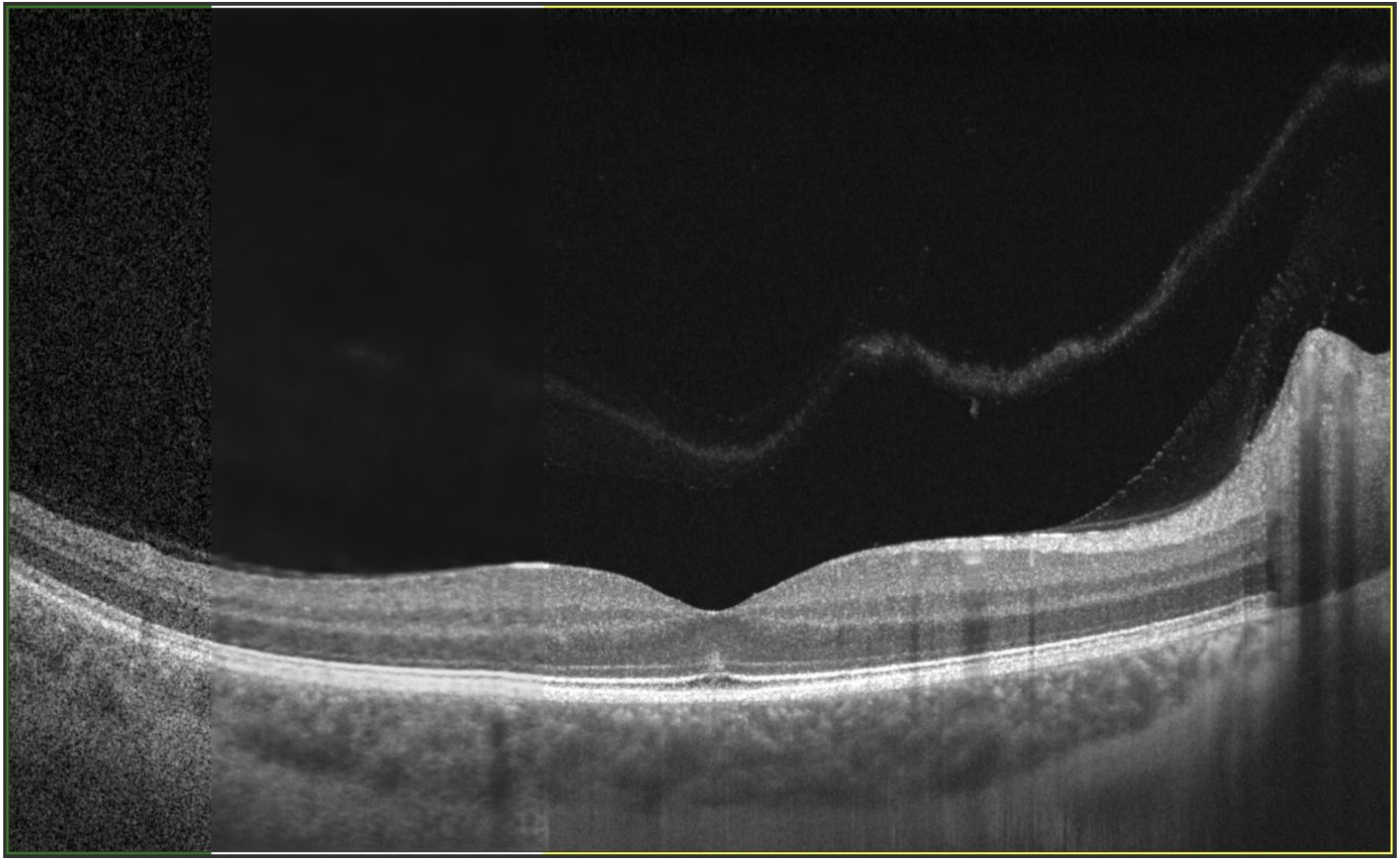
Image analysis (feature extraction)

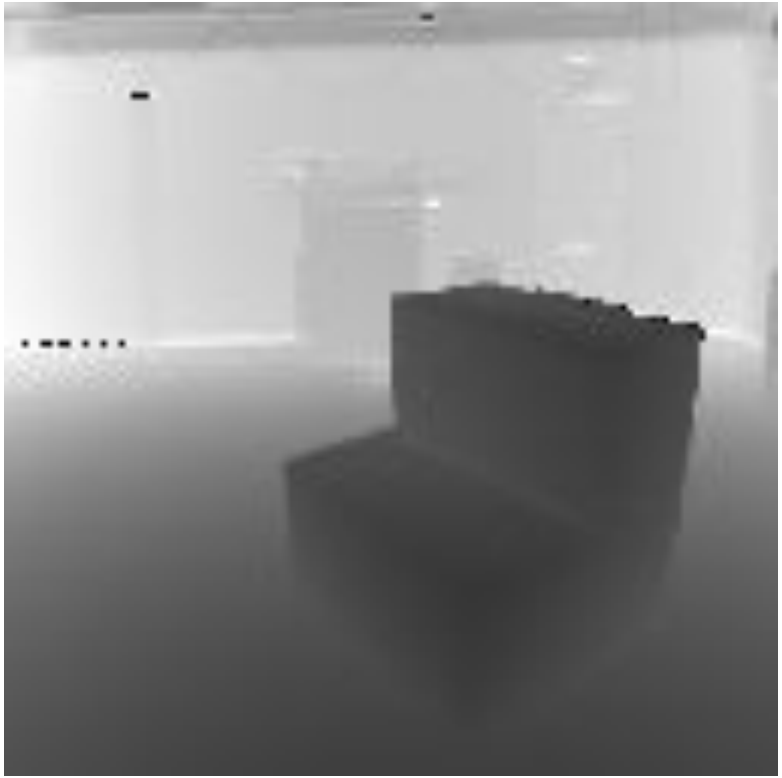
Image/Scene description (feature vector)

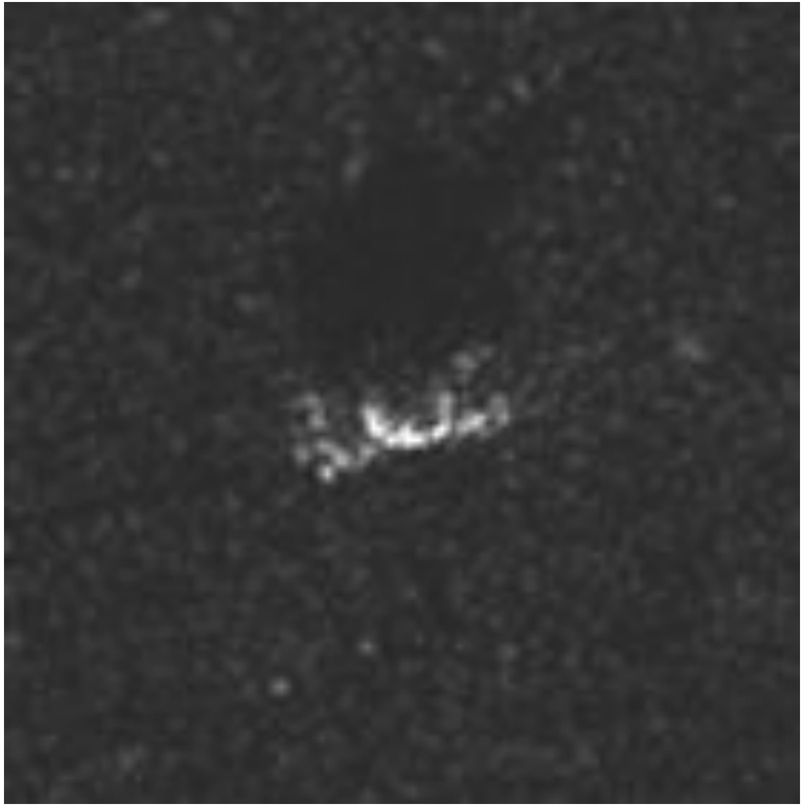
Interpretation/decision making

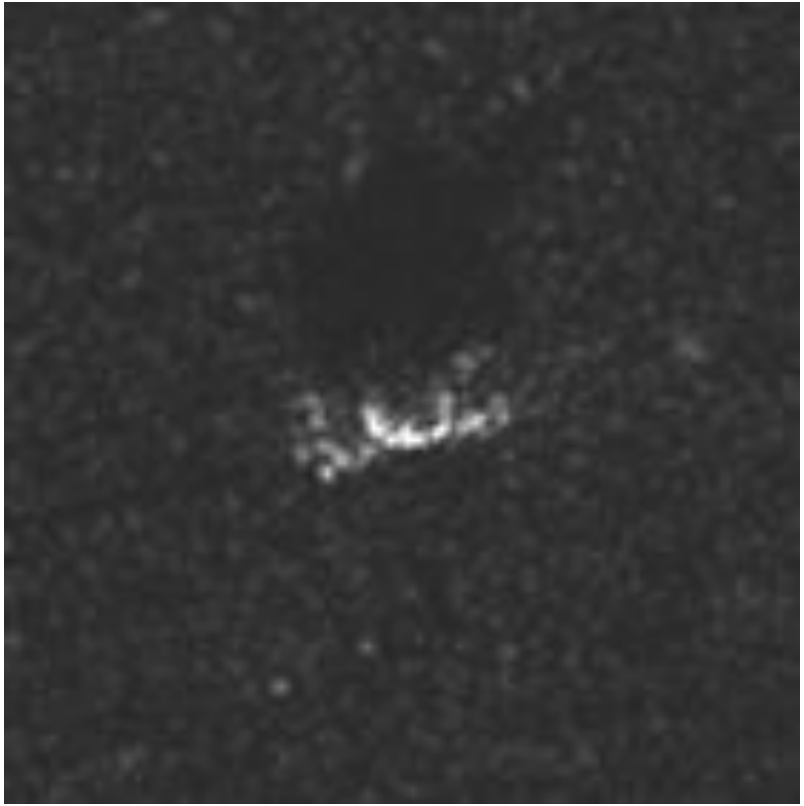
Recognition/decision/measurement

Examples: images, modalities, …

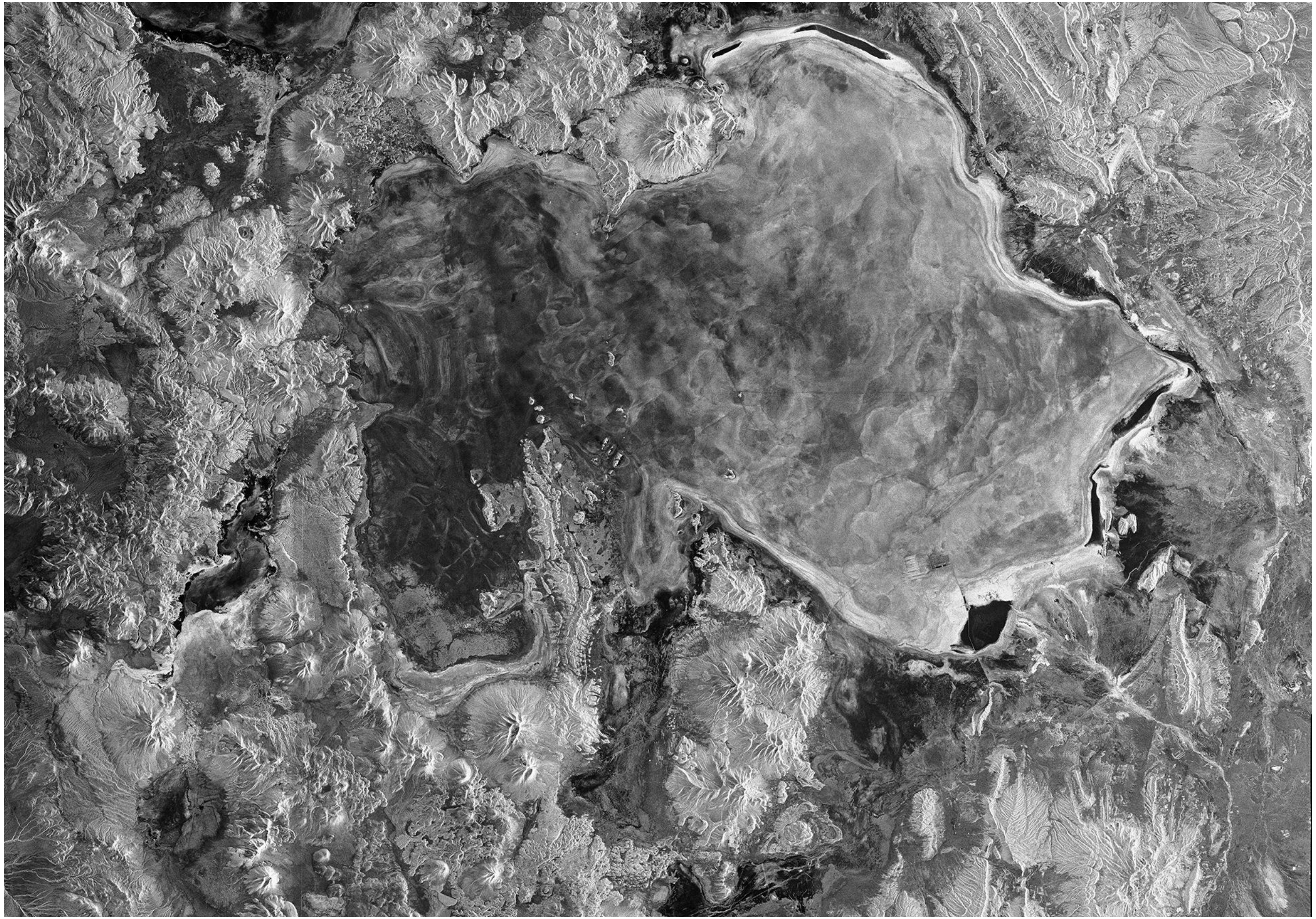


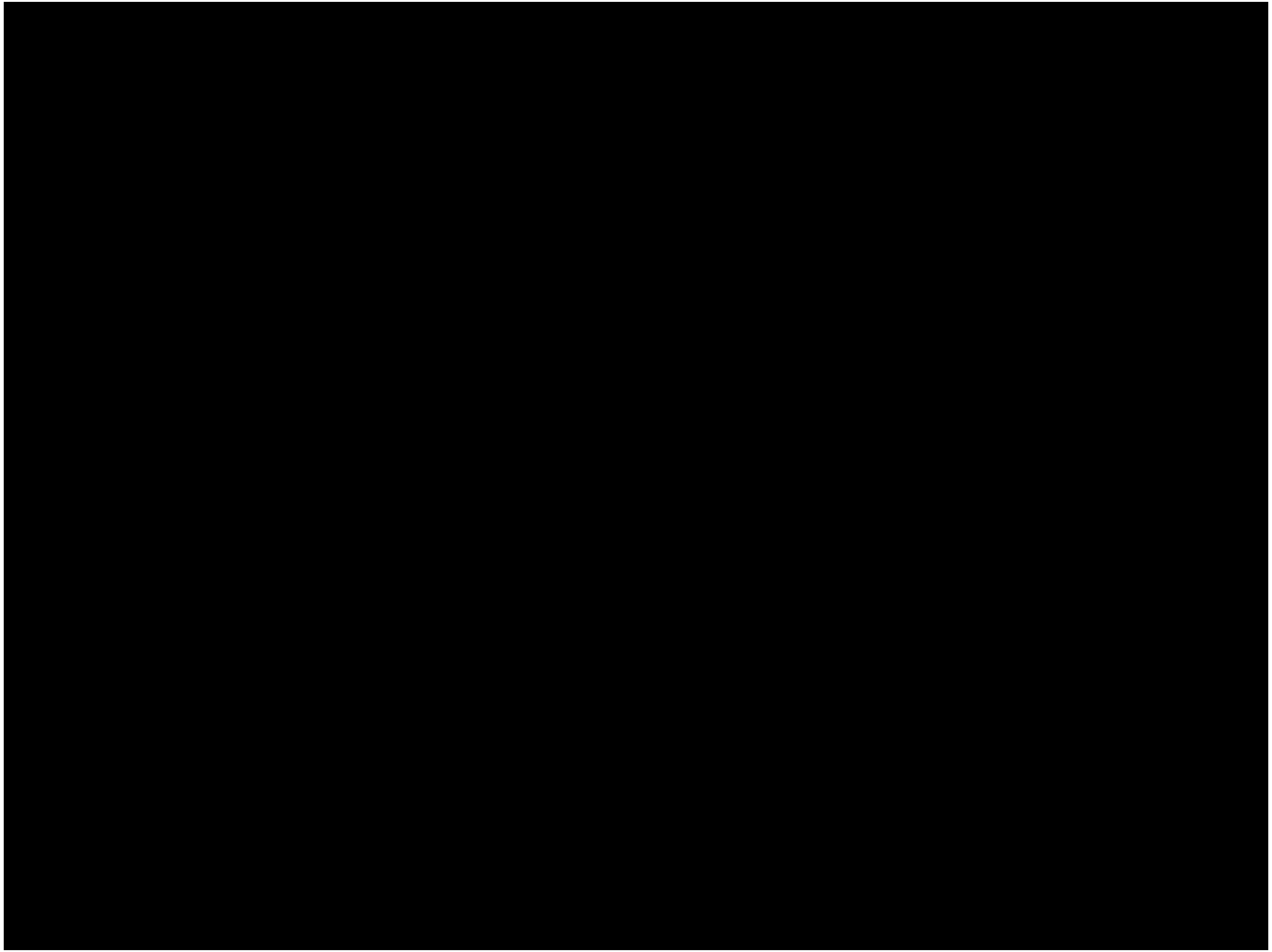




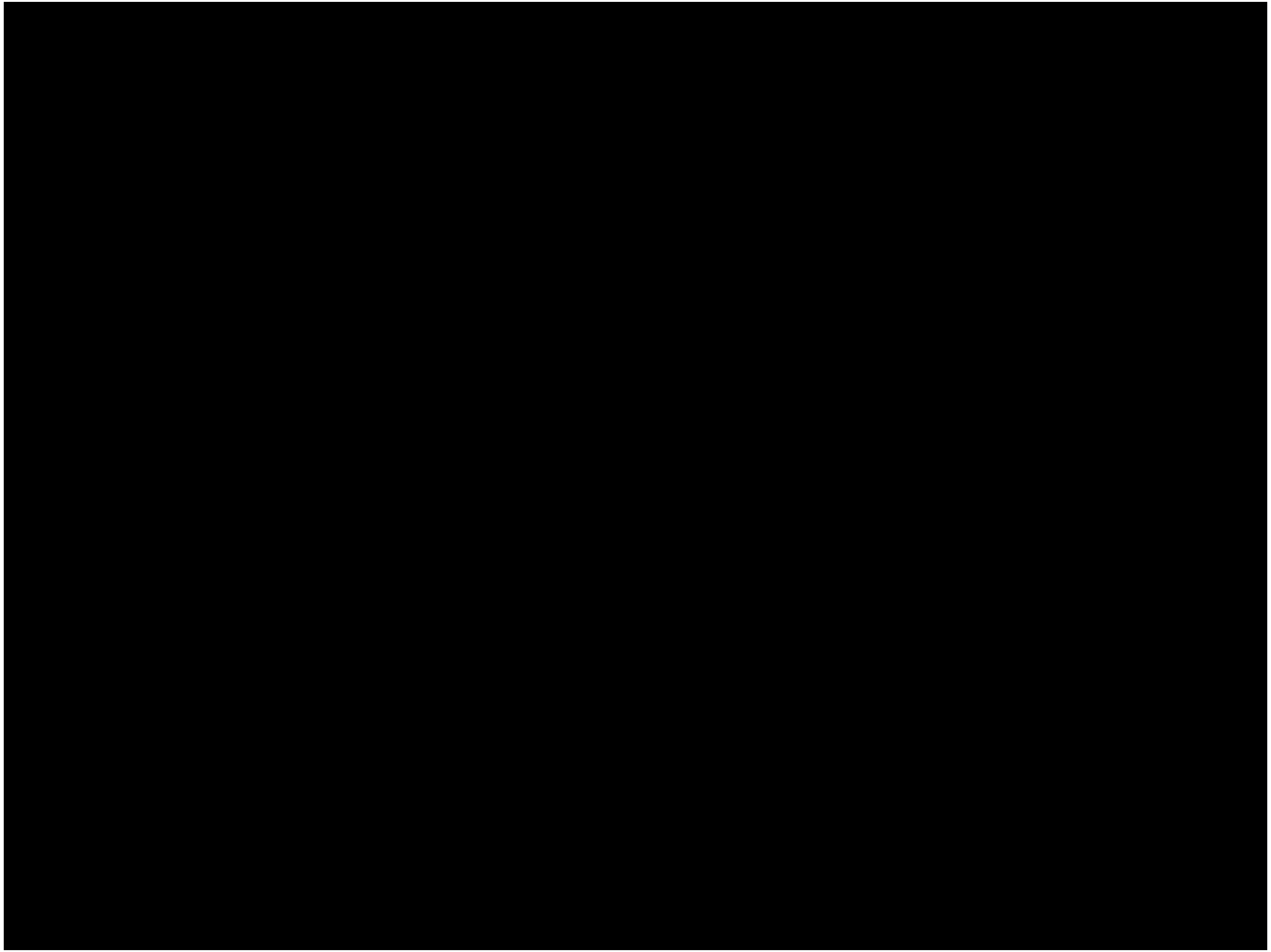
Barcelona. Attribution: Contains modified Copernicus Sentinel data 2019



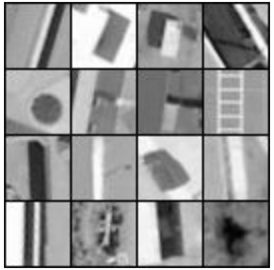
Sentinel-1A satellite (radar), Bolivia. Attribution: Contains modified Copernicus Sentinel data 2019

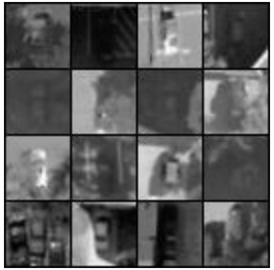


https://www.nature.com/articles/s41598-017-04220-8



https://www.nature.com/articles/s41598-017-04220-8





Main ‘chapters’ of CV

- Image processing
  - Methods of mapping images to images.
- Feature extraction
  - Mapping images to descriptions (which are typically more compact and designed with a specific application and task in mind)
- Image segmentation
  - Delineating regions/classes of objects/object instances in an image.
- Image/object recognition
  - Mapping the image and/or its description to a specific task (close to classification or identification in machine learning, ML).

Types of CV tasks

CV tasks (1)

Related to image processing:

- Improvement of image quality:
    - Image enhancement.
      The objective: Improve human/subjective image perception.
    - Image restoration:
      The objective: Restore the original, undistorted signal.
- Denoising
- Visualization
    - E.g. pseudocoloring
- Compression

CV Tasks (2)

Related to image interpretation:

- Recognition (~classification)
  Does the image belong to/represent a category/class of objects?
  Is this an image of a dog?
- Identification
  Is this an image of a <u>specific object</u>? (instance of a class of objects)
  Is this an image of Rex?
- Detection
  Does the image <u>contain</u> an object with a specific characteristic/class?
  Does the image contain Rex/ a dog?
- Localization
  Determine the location(s) of a given object/class of objects in the image.
  Where are dogs in this image?

CV tasks (3)

Related to scene interpretation (in particular 3D):

- Depth estimation, depth from X
  What are the distances of scene points (all/selected) from the observer?
- Shape estimation, shape from X
  What is the shape of this object? (class, mathematical model, keypoints)
- Pose estimation
  What is the rotation matrix of an object? (the shape of which I may know)
- Measurement
  What is the circumference of this shape/object?
- Scene interpretation
  Provide a possibly complete high-level description of the scene, e.g. list of objects, their categories, locations, spatial relations, ...

CV tasks (4)

Related to motion (video sequences)

- Motion detection
  Is there observable motion in the scene?
  How intense is the motion?
- Motion estimation
  What is the structure of the motion (direction, vector)?
  Does the motion affect the entire scene, or individual objects?
- Object tracking
  Track/monitor the location of object(s) in consecutive frames of the video sequence.

CV tasks (5)

Other:

- Image segmentation
  Partition the image into parts that have relatively uniform perceptual characteristics.
  Separate the object from the background (ang. figure-ground separation)
- Semantic segmentation
  Partition the image into parts with specific semantic content (category).
  Identify tormac, pavement, vegetation and sky in an image acquired by a autonomous robot.
  - Special case: object instance segmentation.
- Image analysis
  Extract other type of information from the image (may lack intuitive/physical interpretation).

CV tasks and practice

- The type of a CV tasks largely determines selection of methods and techniques.
- Correct assignment of a given problem to a category of CV tasks is essential.
- Once the type of a CV task has been determined, one typically needs to pick the <u>metrics</u> (see next slides).

Criteria/metrics for CV systems

Universal:

- computational complexity and memory complexity
- invariance T,S,R (+3D)
- robustness (rate of failure)

Task-specific:

- quality of reconstruction/restoration (e.g. denoising, compression)
- accuracy of classification, sensitivity, specificity, AUC (recognition, detection)
- precision/spatial accuracy (localization)
- conformance with some predefined ground truth (GT), e.g. intersection over union (segmentation, semantic segmentation)

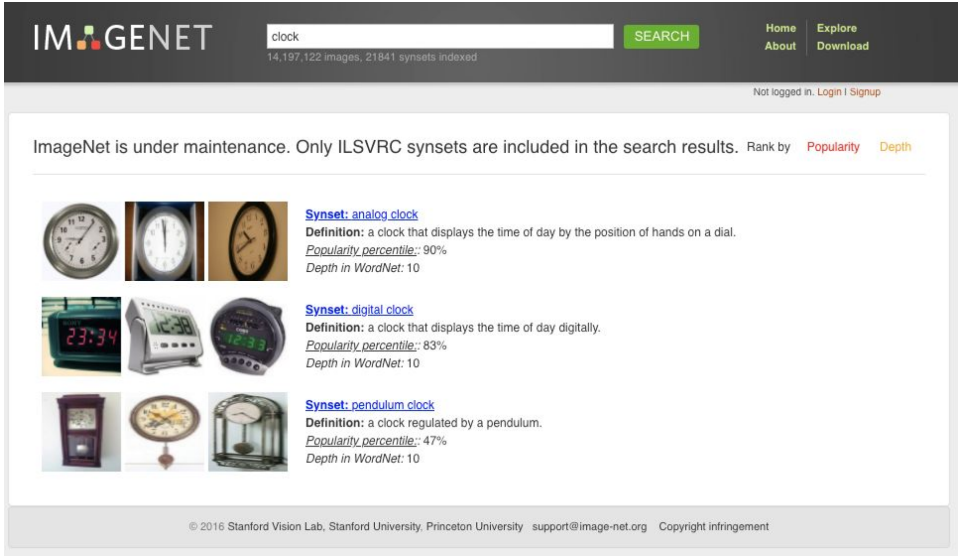
A (definitely incomplete list of applications)

- Cartography, geology, meteorology: satellite and aerial image analysis
- Medicine, e.g. screening, microscopic image analysis,
- Archeology: reconstruction of damaged exhibits,
- Astronomy: automatic pre-classification of celestial bodies,
- Bio-identification based on: fingerprints, iris, retina, hand shape, gait, the mouth-prints, and other biometric features
- Security: supporting monitoring of facilities (CCTV), airport security, luggage control, monitoring, city traffic control,
- Industry: quality control,
- Robotics: industrial robots, autonomous robots and vehicles
- Entertainment: Sony AIBO dog robot, MS Kinect,
- Military

CV applications that are largely ‘solved’ (?)

- Optical character recognition (OCR)
- Person identification based on fingerprints
- Person identification based on face (face-id)
- Classification of common types of objects/images

  See http://image-net.org/

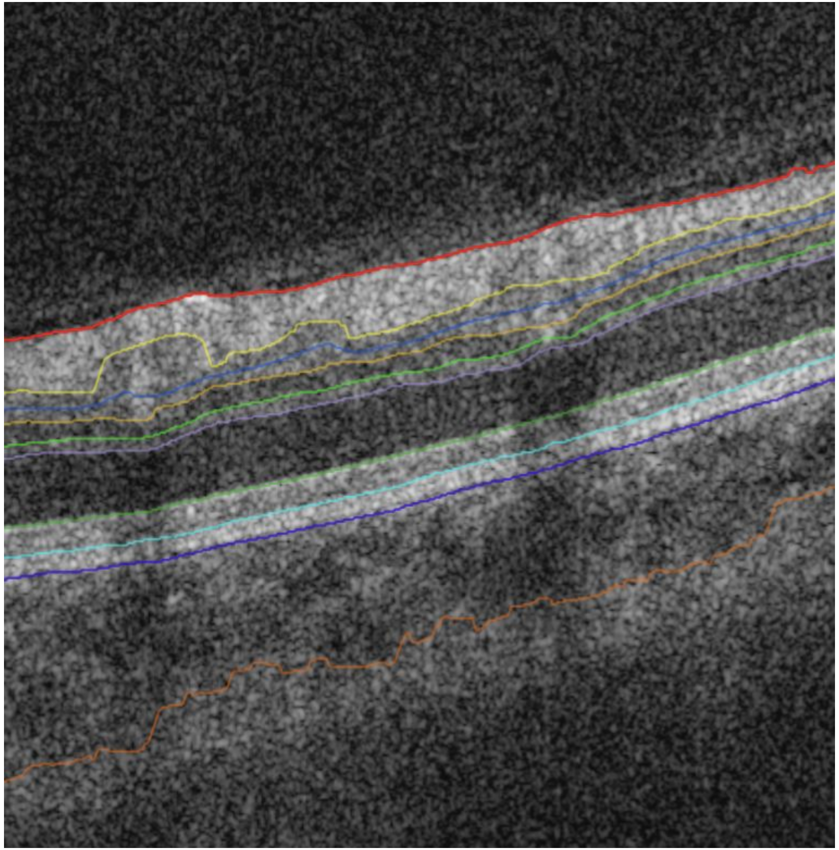


Relations with AI

- AI has always been present in CV.
- The role of Machine Learning has also always been significant, and has recently become fundamental due to advances in ML, especially in deep neural networks (deep learning).
- Nevertheless: it should not be assumed that all CV tasks always require ML, because:
  - Adaptation / learning is not needed/essential in certain usage scenarios.
  - Much of the AI research concerns approximate and heuristic algorithms, while for some [sub] problems in CV we know exact algorithms.

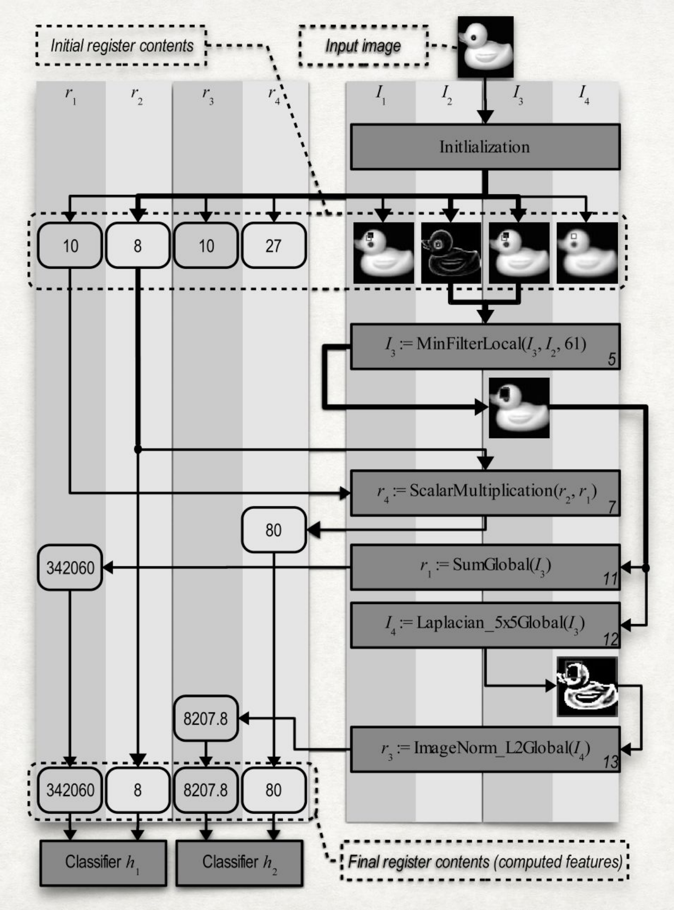
Example 1 (use of an exact algorithm)

- Using Dijkstra's algorithm to determine the location of anatomical boundaries in the posterior of the eye acquired with optical computer tomography (OCT).
- Cost-on-node: The confidence provided by a recognition model (a neural network in this case).



Example 2

- Deep learning is prevalent in CV.
- There are many heuristics and metaheuristics that proved useful in CV.
- The figure: An image processing and analysis program (dataflow/computation graph) synthesized by an evolutionary algorithm (genetic programming) in a domain-specific language (DSL) based on the OpenCV library.



Krzysztof Krawiec. Evolutionary Feature Programming: Cooperative learning for knowledge discovery and computer vision. Poznan University of Technology, Poznan, Poland: 2004. http://www.cs.put.poznan.pl/kkrawiec/pubs/hab/ krawiec_hab.pdf

Challenges in CV

- Needs to take into account:
  - domain knowledge
  - common sense knowledge
- Incomplete/uncertain information resulting from the nature of the image acquisition process:
  - deterioration of signal quality
    - noise, optical distortion, chromatic distortion, etc.
  - information loss: the image does not contain complete information about the scene:
    - projecting three-dimensional objects in the scene onto a two-dimensional sensor,
    - occlusion,
    - incomplete context (e.g. restricted field of view).
- Large volume of processed information
  - consequence: considerable time and memory complexity
  - often needs to be traded for quality (trade-offs between metrics)

Examples

Motivations (market, business, environment, SOTA)

See the Motivations slide, and atop of that:

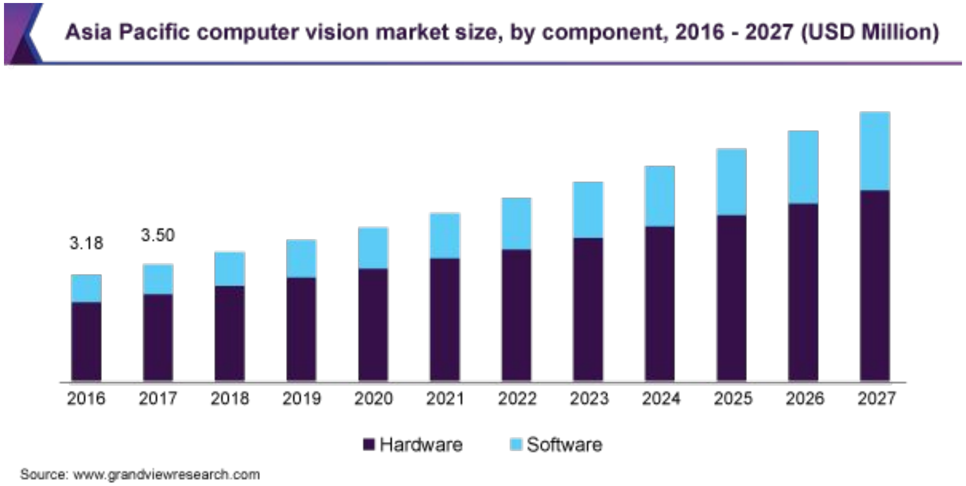
- Significant progress made over the past two decades.
  - 'Tipping point' achieved in many applications.
- Dominance of programming implementations (software)
  - Flexible, easily reprogrammable, extensible, upgradable ...
  - CV is now largely absorbed by computer science.
- Hardware implementations still present, e.g.
  - FPGA (Field-Programmable Gate Array), Graphics Cards (GPU), General-Purpose Computing on GPUs)

Inspiring examples of advanced achievements:

- Google Lens
  https://lens.google.com/
- Automatic generation of image descriptions (image captioning)
  http://cs.stanford.edu/people/karpathy/deepimagesent/

Market incentives

- The size of CV market: [Computer Vision Market Size Worth $19.1 Billion By 2027 | CAGR: 7.6%: Grand View Research, Inc.]
- Computer Vision Market Size, Share & Trends Analysis Report By Component [2020 - 2027]



Source: [www.grandviewresearch.com]

End of module