COMP9517

Computer Vision

2025 Term 1 Week 1

A/Prof Yang Song

Introduction

What is computer vision?

Computer science perspective

Computer vision is the interdisciplinary field that develops theories and methods to allow computers extract relevant information from digital images or videos

Computer engineering perspective

Computer vision is the interdisciplinary field that develops algorithms and tools to automate perceptual tasks normally performed by the human visual system

Every picture tells a story

"A picture is worth a thousand words"

Computer vision

automates and integrates many information processing and representation approaches useful for visual perception

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

Can computers match (or beat) humans?

Yes and **No**

- Humans are still better at “hard” tasks

  Ambiguous data, leveraging prior knowledge, continual learning, working across applications

- Computers can be better at “easy” tasks

  High-quality data, using mathematical models, consistent training set, single well-defined application

Granny Smith	85.6%
iPod	0.4%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.1%

Granny Smith	0.1%
iPod	99.7%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.0%

State-of-the-art methods

Can you describe this image

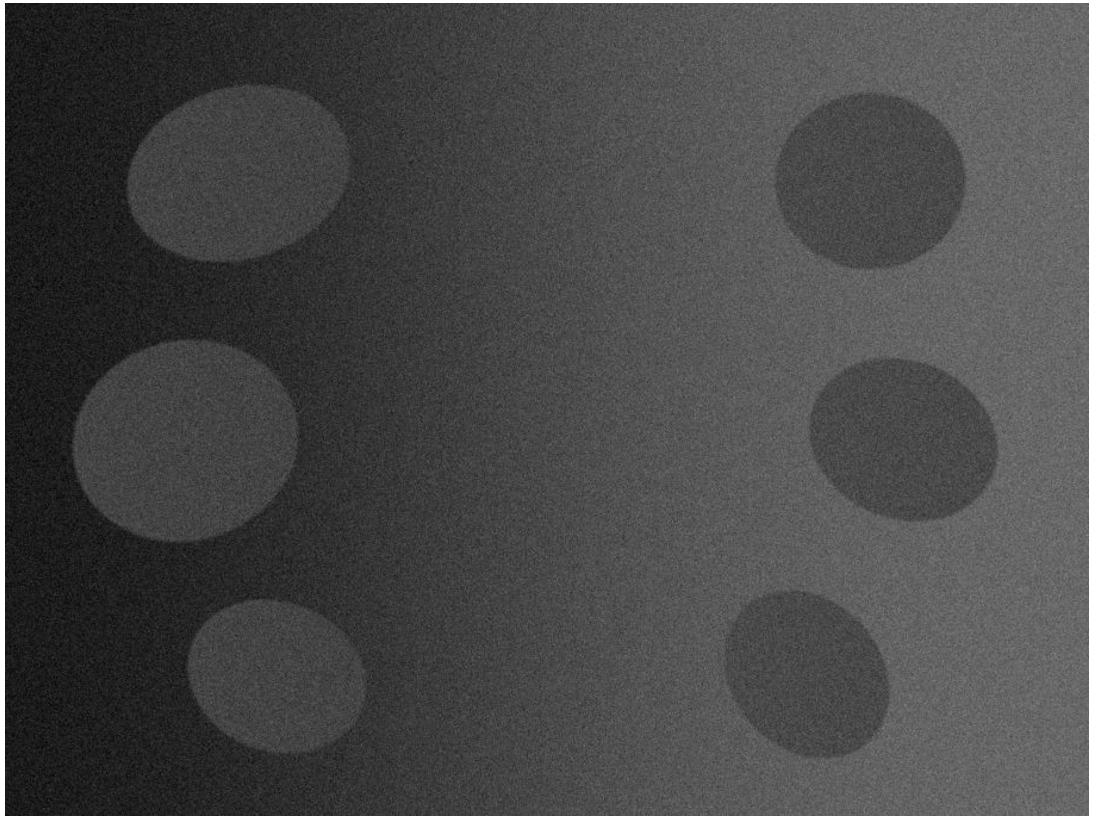
This is a famous historical photograph of the **Montparnasse derailment**, which occurred on **October 22, 1895**, at the **Gare Montparnasse** railway station in **Paris, France**. The image captures the dramatic moment when a steam locomotive overran the buffer stop, crashed through the station wall, and plunged onto the street below.

The train, operated by **Chemins de fer de l'Ouest**, failed to stop in time due to excessive speed and possibly a failure in the braking system. As a result, it smashed through the station's facade and landed

Can you describe this image:

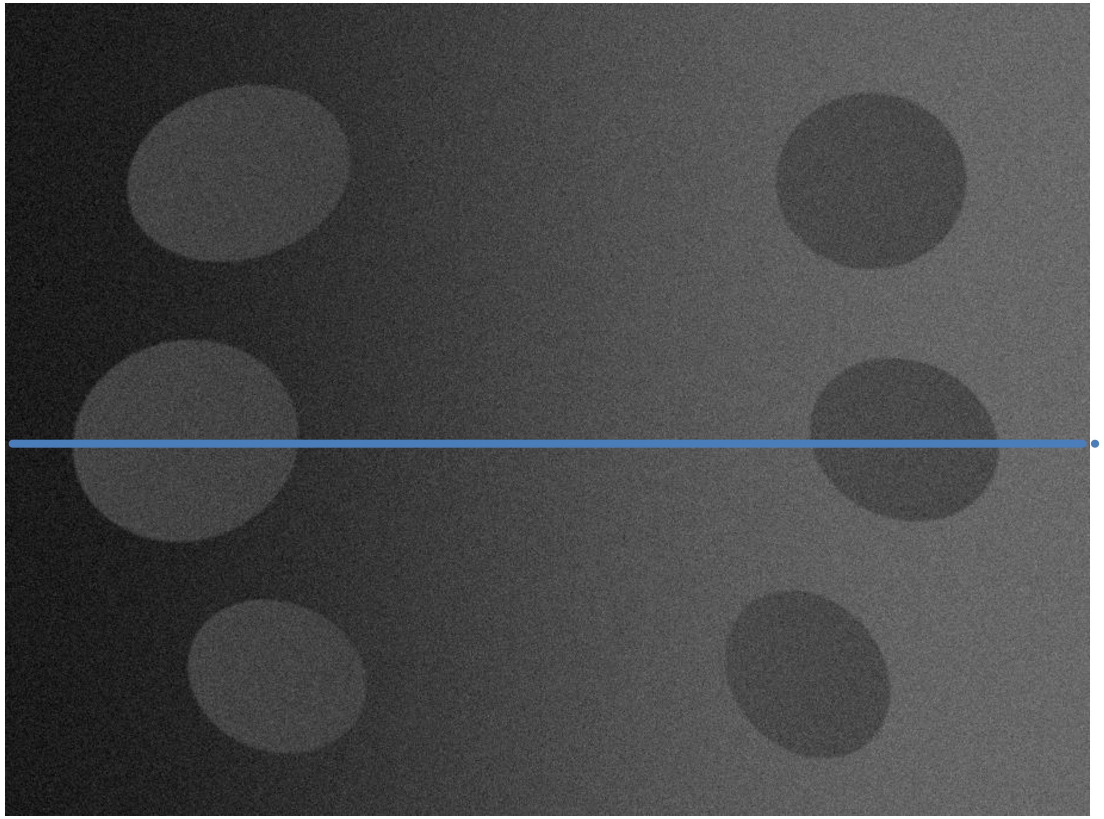
The image shows an apple with a piece of paper taped to it. The paper has "iPod" written on it in handwritten text. This appears to be a humorous play on words, referencing Apple's famous iPod device by combining the brand name with a literal apple (the fruit). The background includes an outdoor setting with a wooden fence and some greenery.

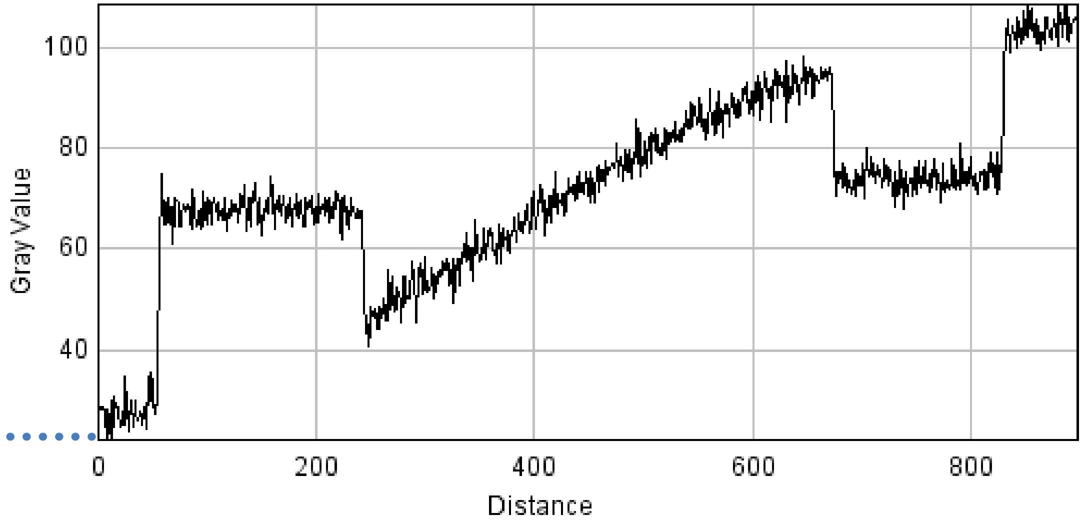
Human vision has its limitations…



Which objects are brighter?

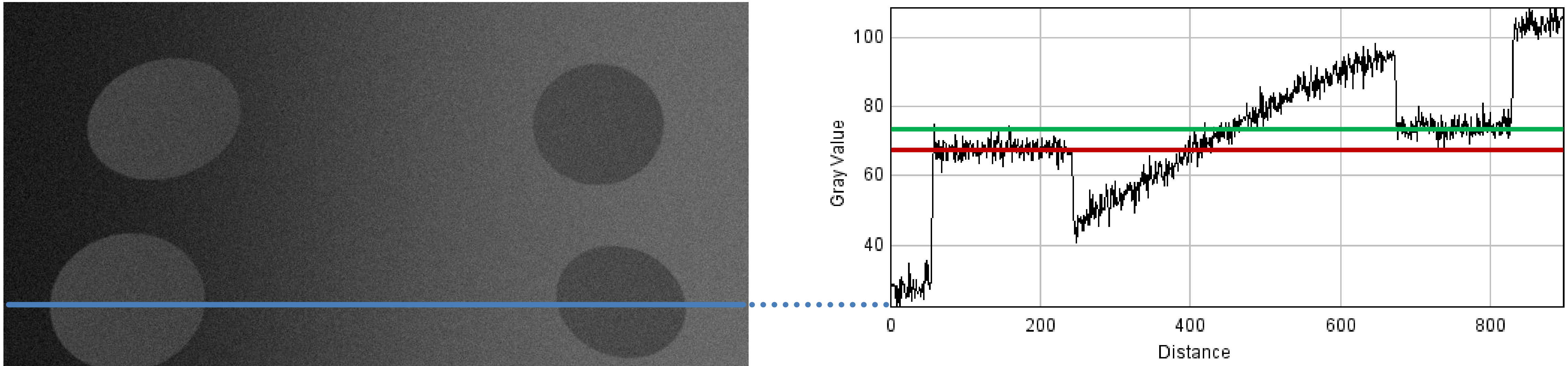
Human vision has its limitations…





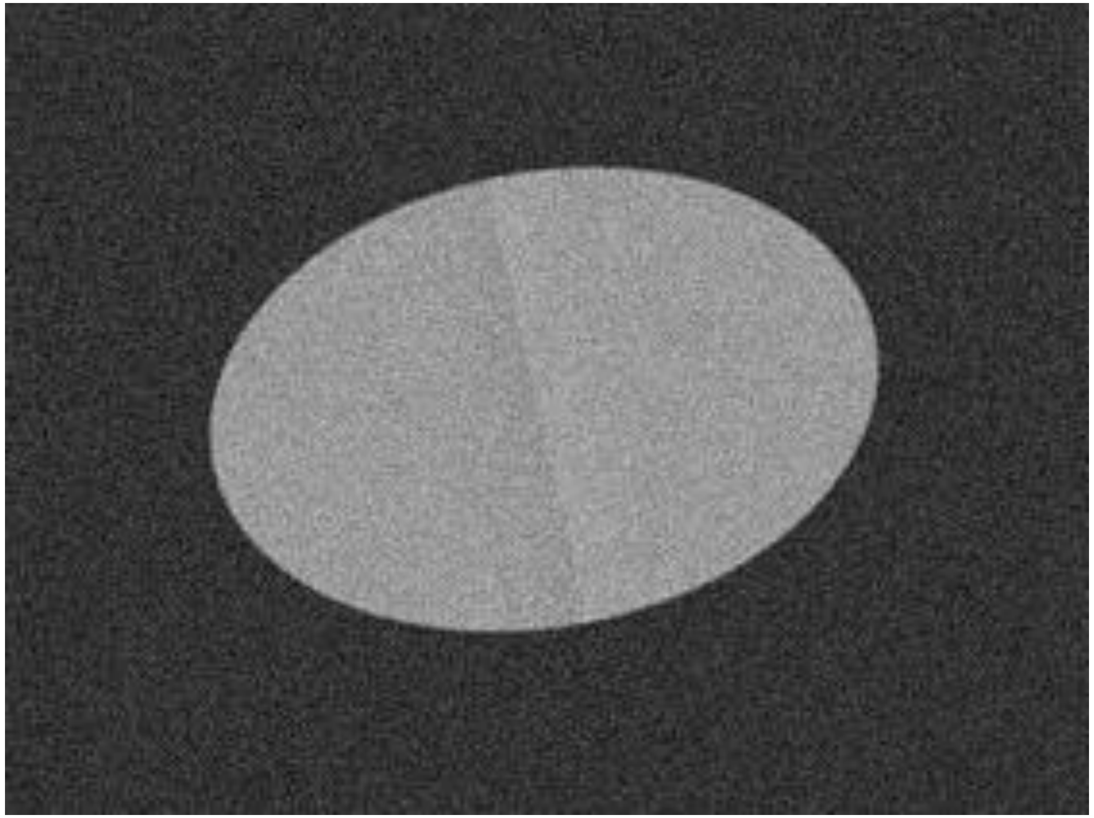
Which objects are brighter?

Human vision has its limitations…



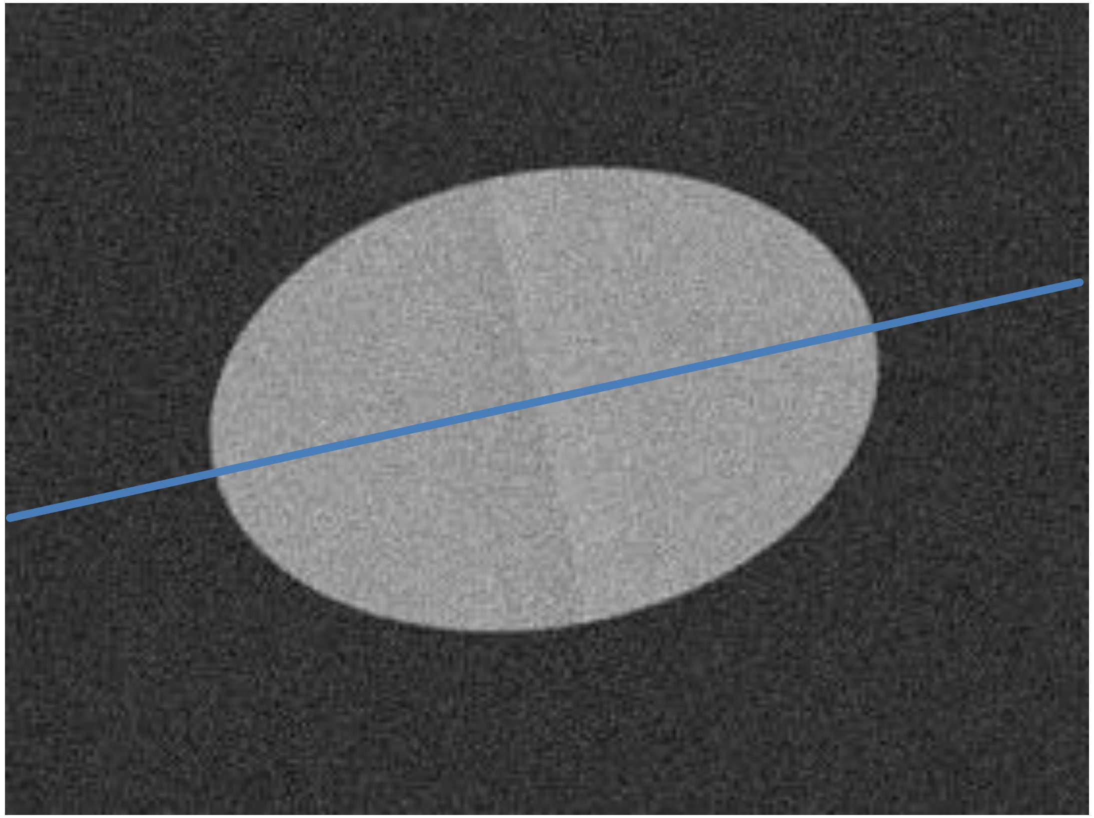
Which objects are brighter?

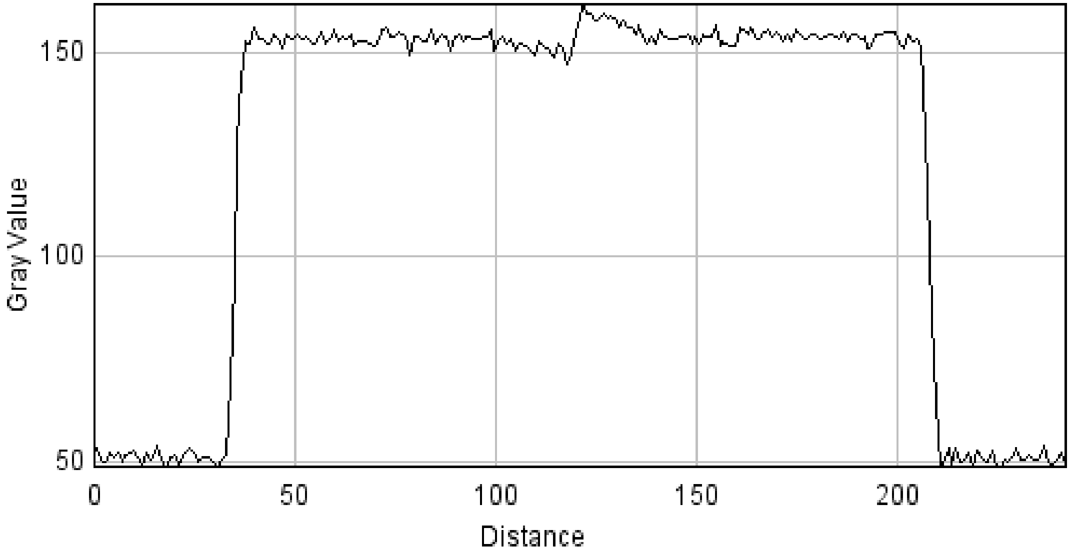
Human vision has its limitations…



Which side of this object is brighter?

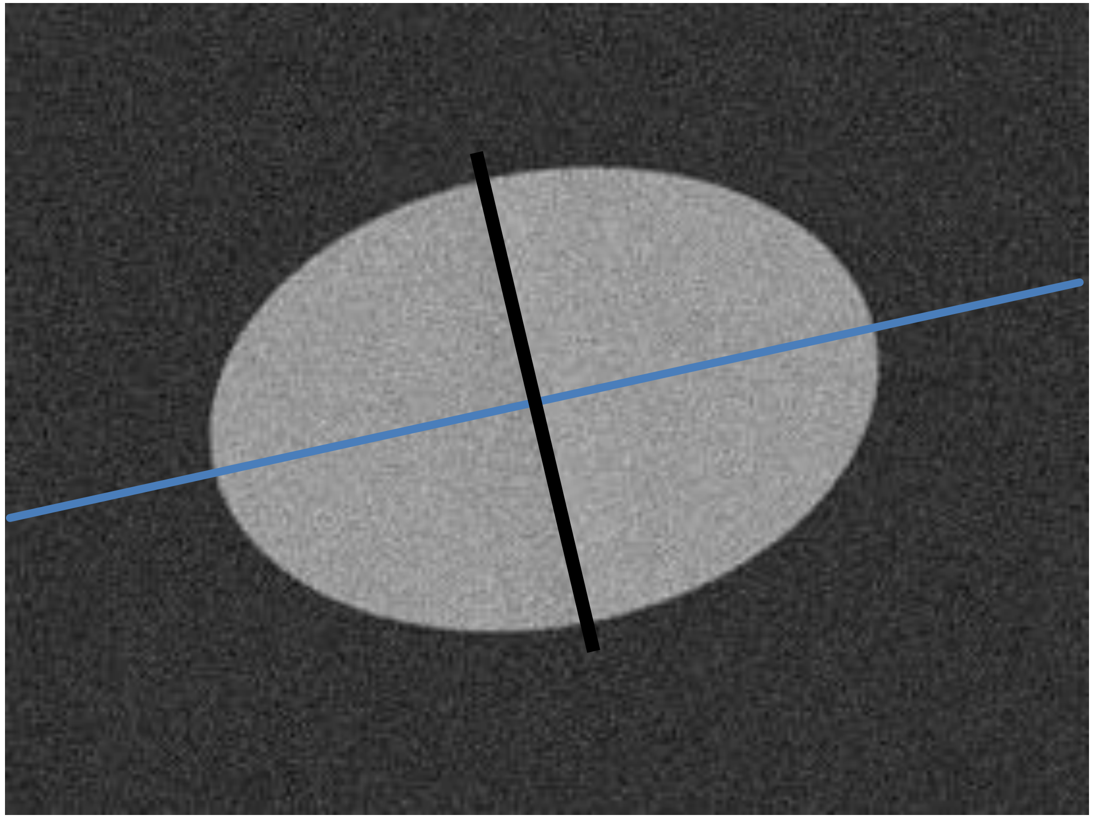
Human vision has its limitations…

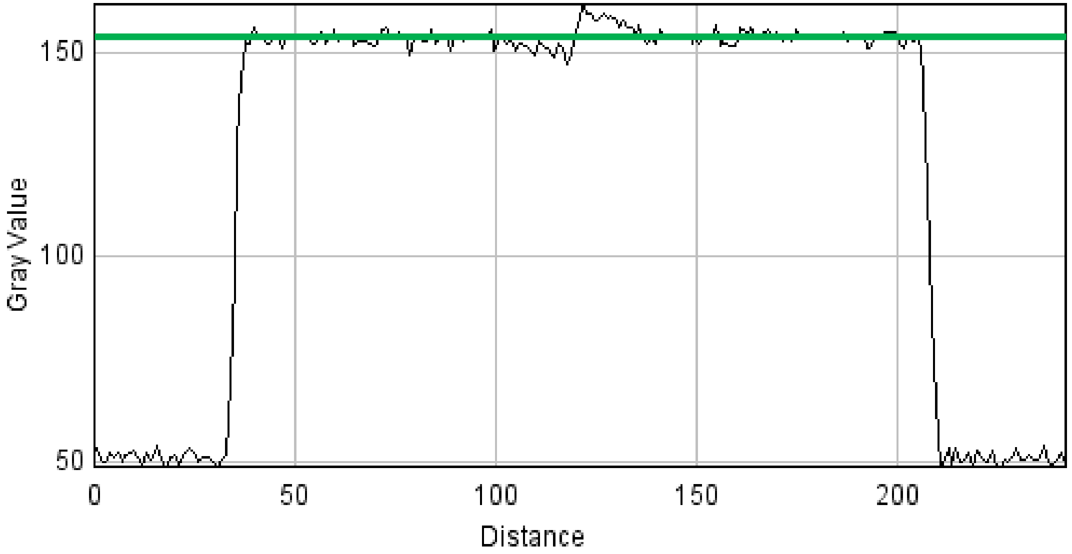




Which side of this object is brighter?

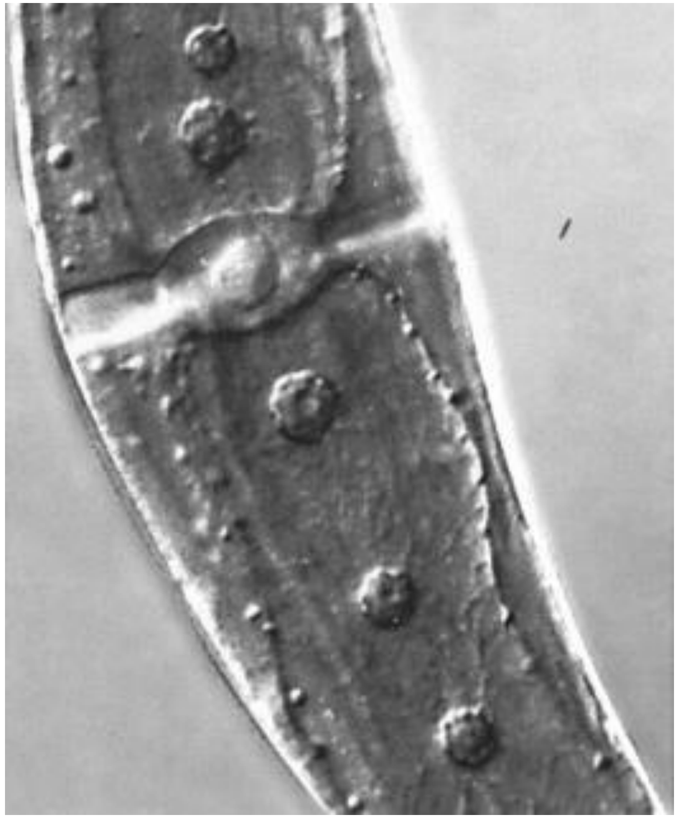
Human vision has its limitations…





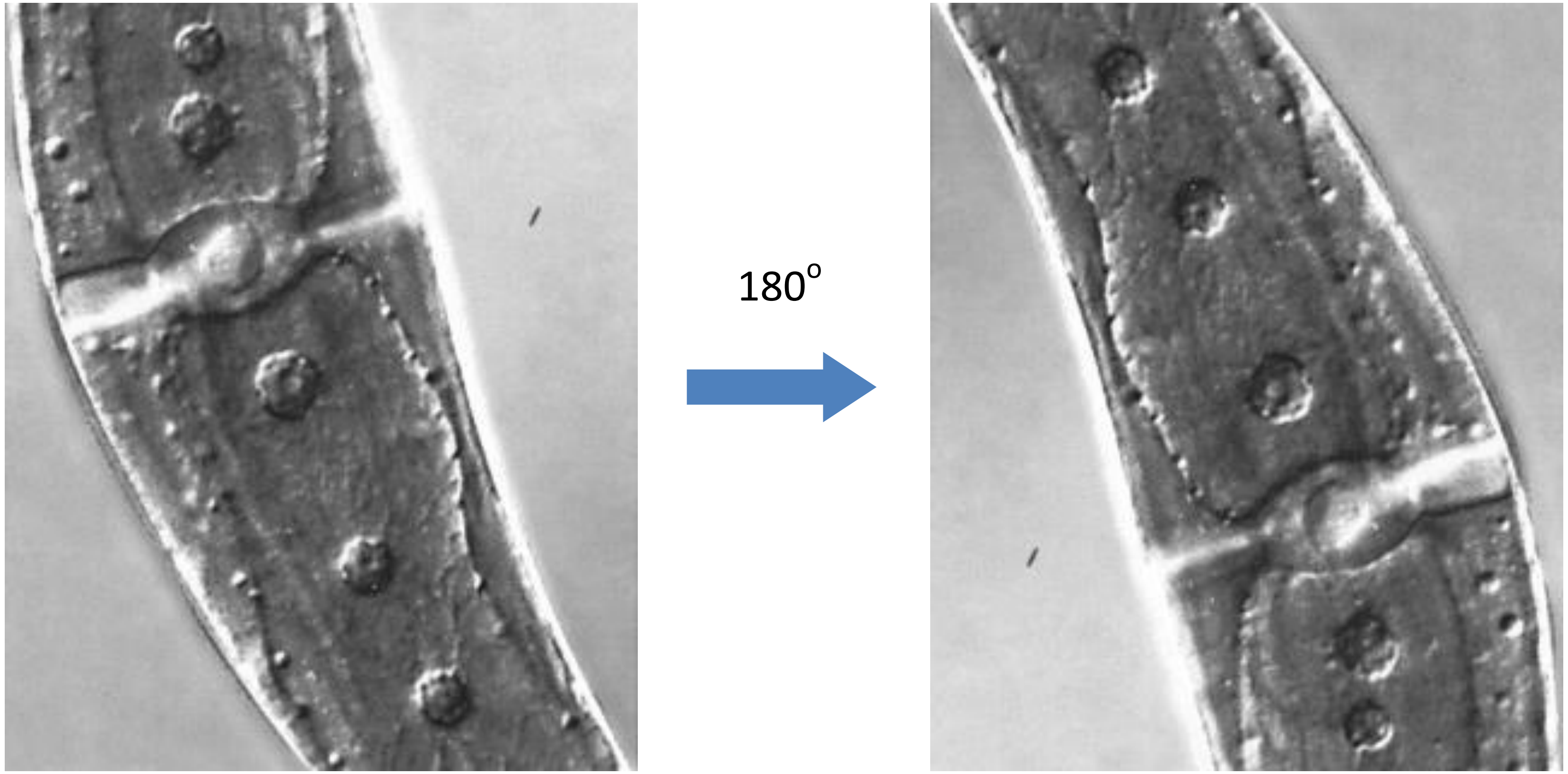
Which side of this object is brighter?

Human vision has its limitations…

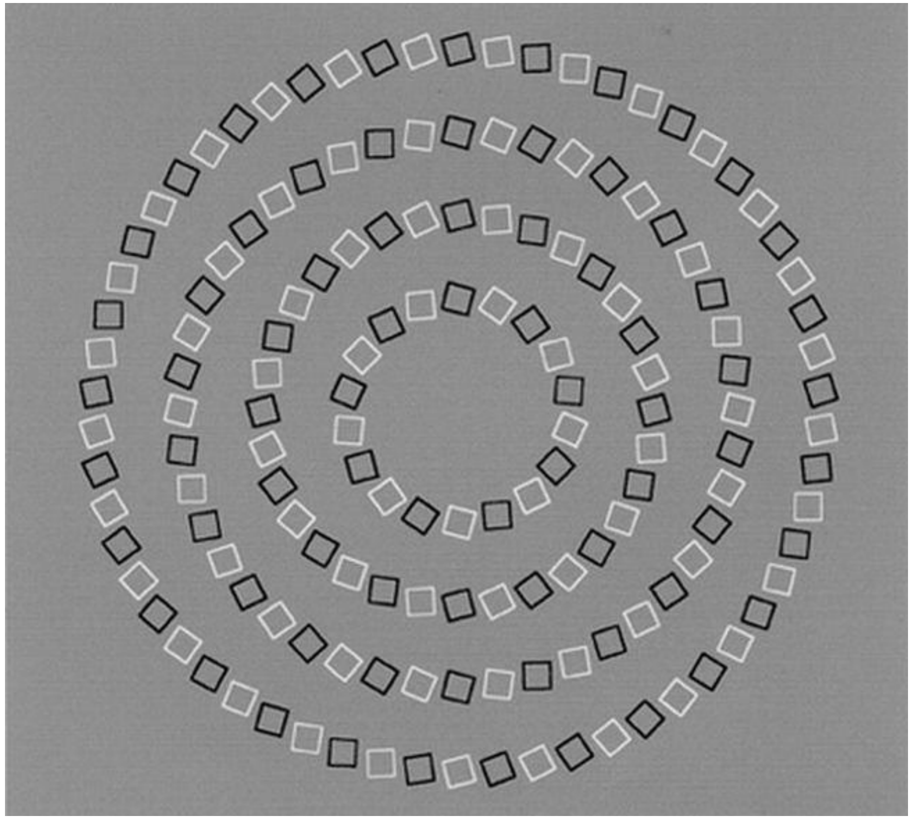


Are the cells “popping out” or “popping in”?

Human vision has its limitations…

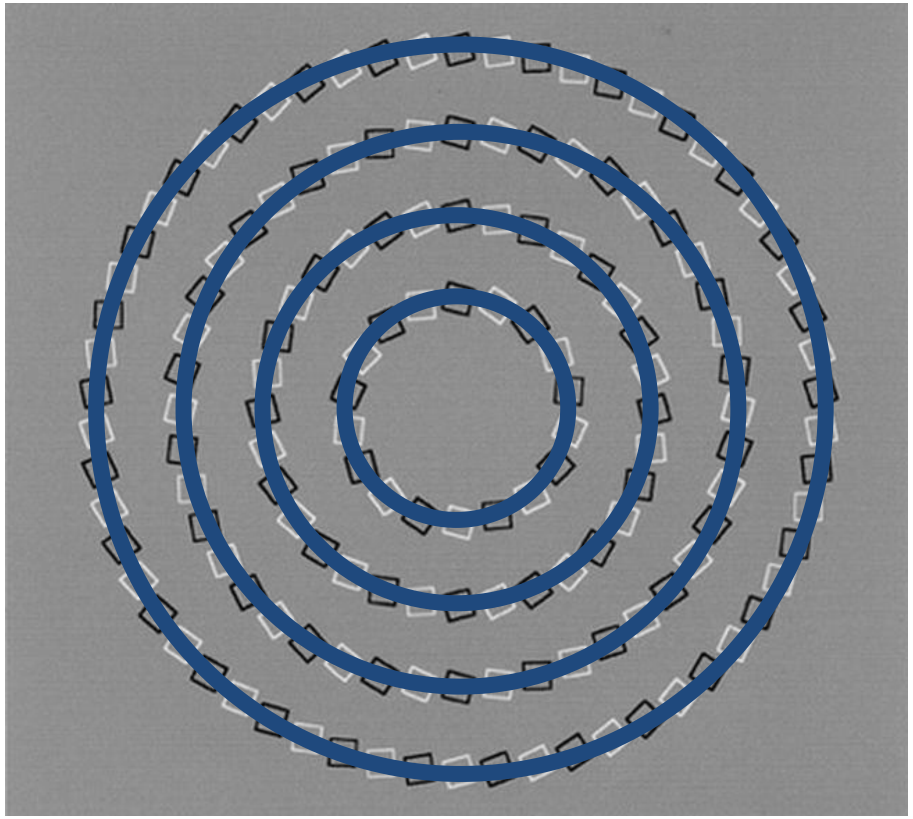


Human vision has its limitations…



What pattern do the squares form?

Human vision has its limitations…

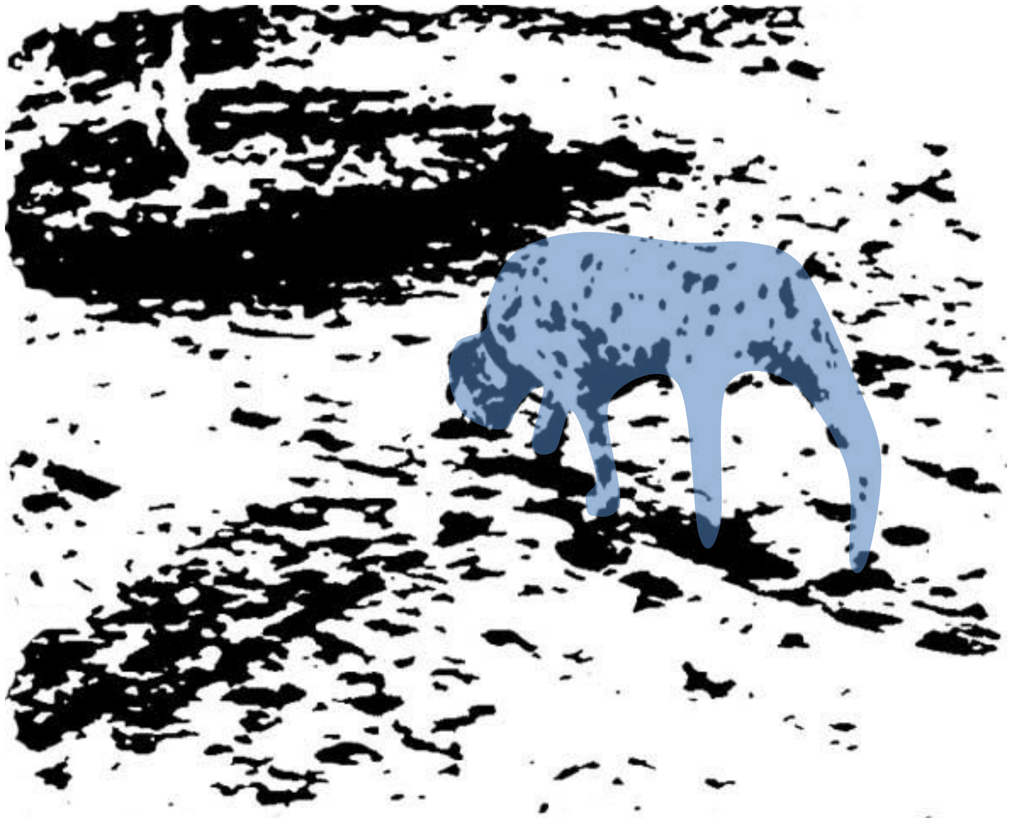


What pattern do the squares form?

Human vision has its limitations…

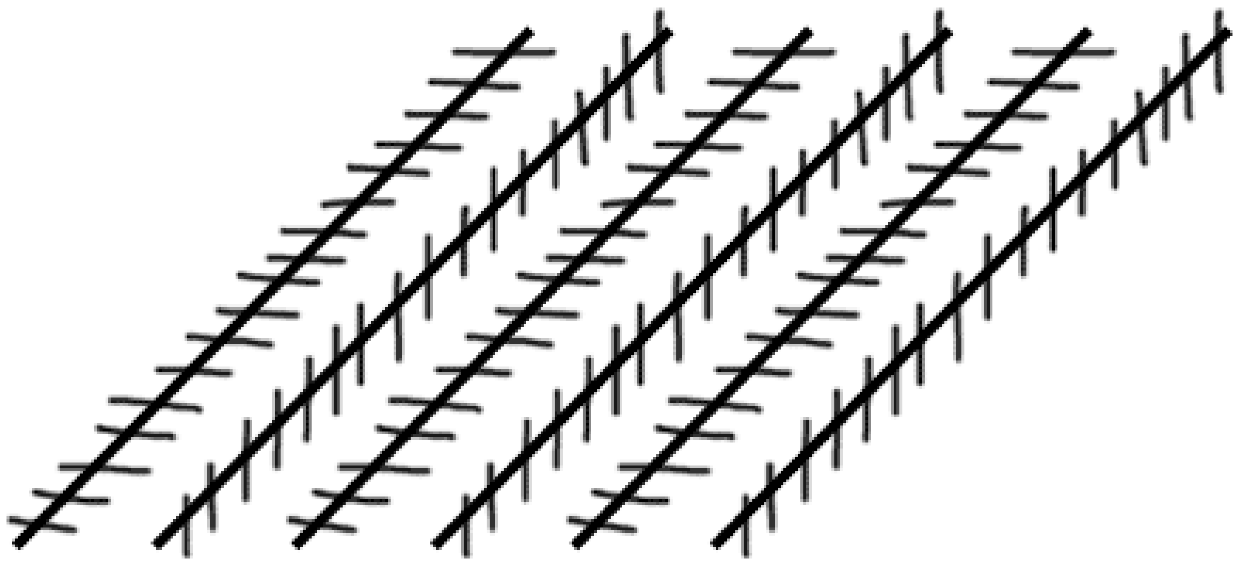
What object do you see in this image?

Human vision has its limitations…



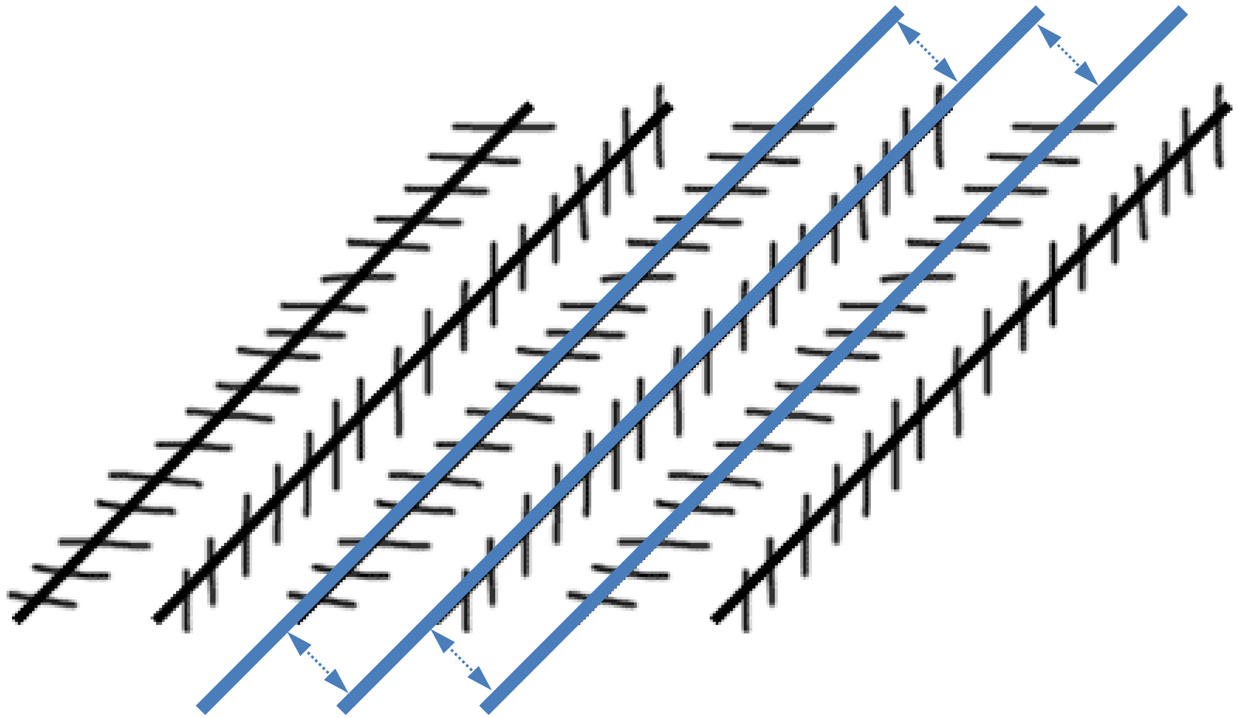
What object do you see in this image?

Human vision has its limitations…



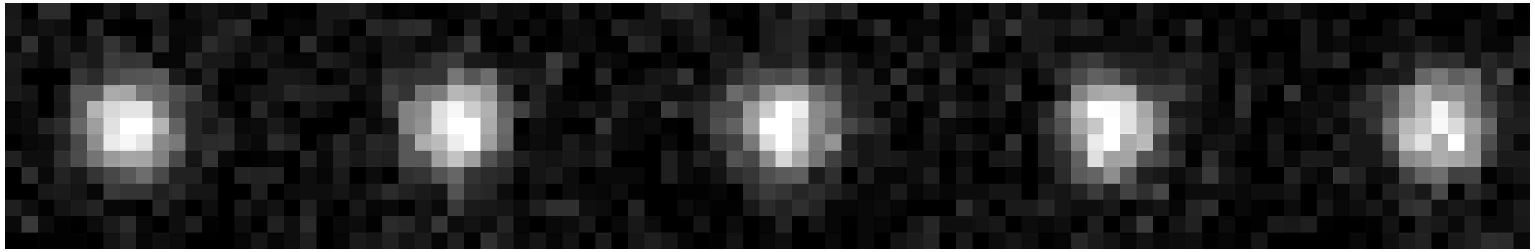
How do the main lines run with respect to each other?

Human vision has its limitations…



How do the main lines run with respect to each other?

Human vision has its limitations…

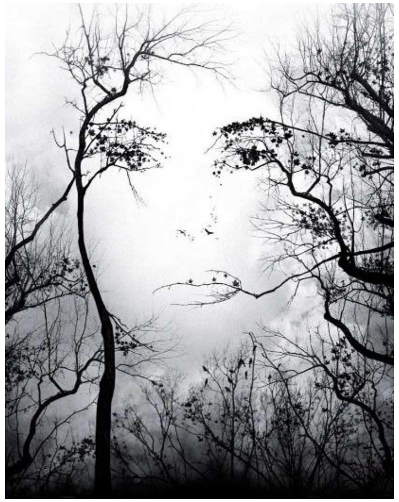


In which direction are these particles moving ?

Human vision has its limitations…

https://www.youtube.com/watch?v=a7efEqgpIrE

Course rationale



Human vision has its limitations

- Intensities, shapes, patterns, motions can be misinterpreted
- Is labor intensive, time-consuming, subjective, error-prone

Computer vision can potentially improve this

- Computers can work day and night without getting tired
- Analyze information quantitatively and objectively
- Potentially more accurate, precise, reproducible

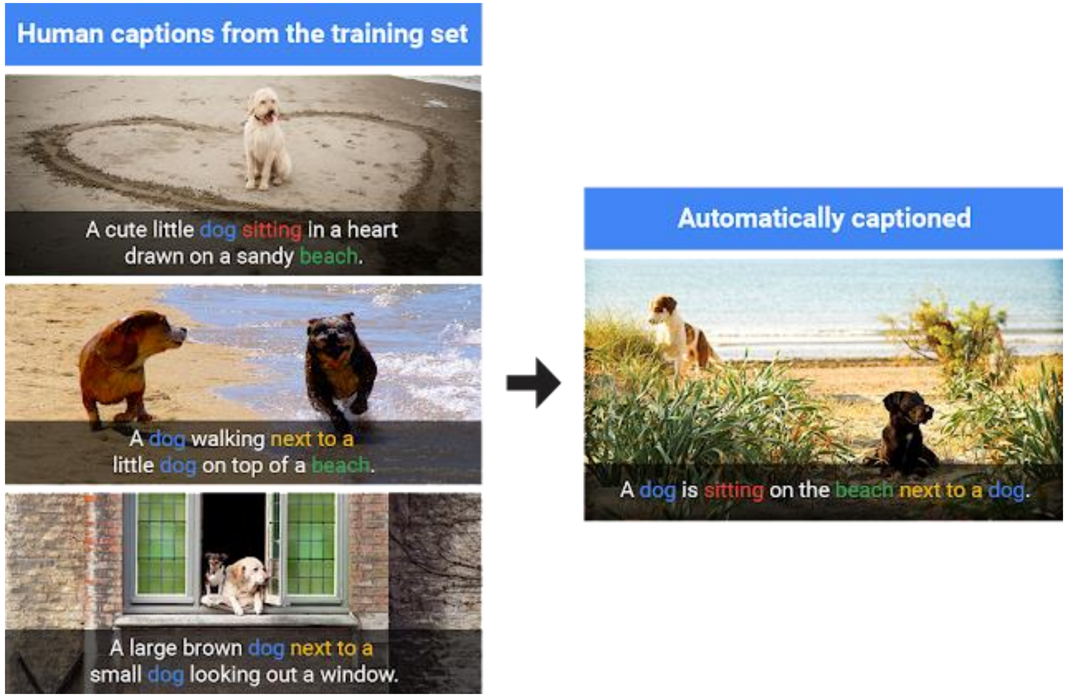
If the methods and tools are well designed!

Application: 3D shape reconstruction

Project [VarCity]() recreates 3D city models using social media photos

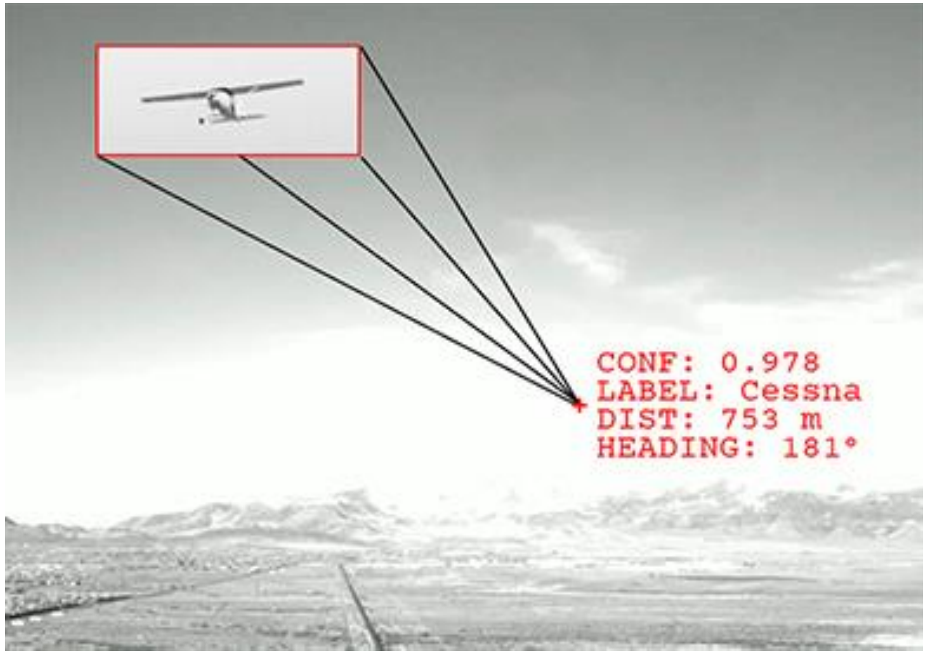
Application: image captioning

Google's Show and Tell open-source image captioning model in TensorFlow



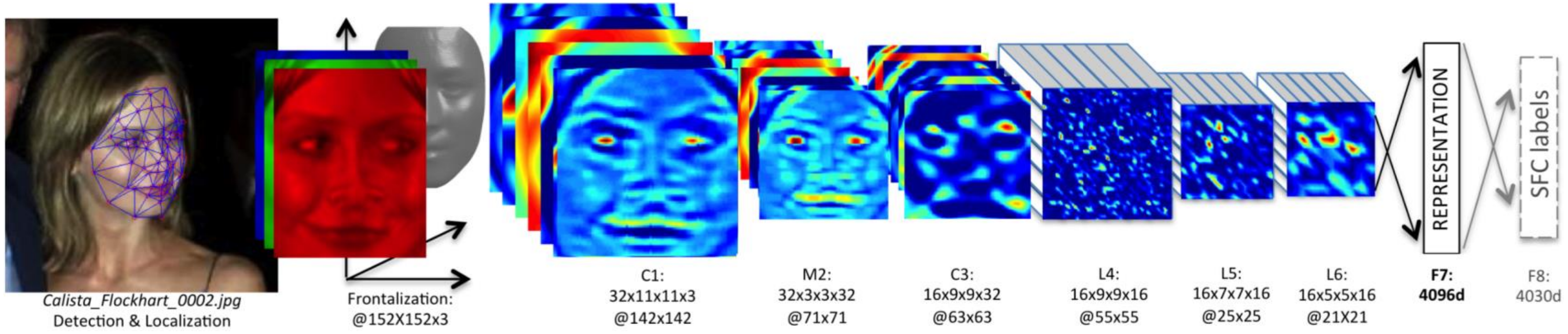
Application: intelligent collision avoidance

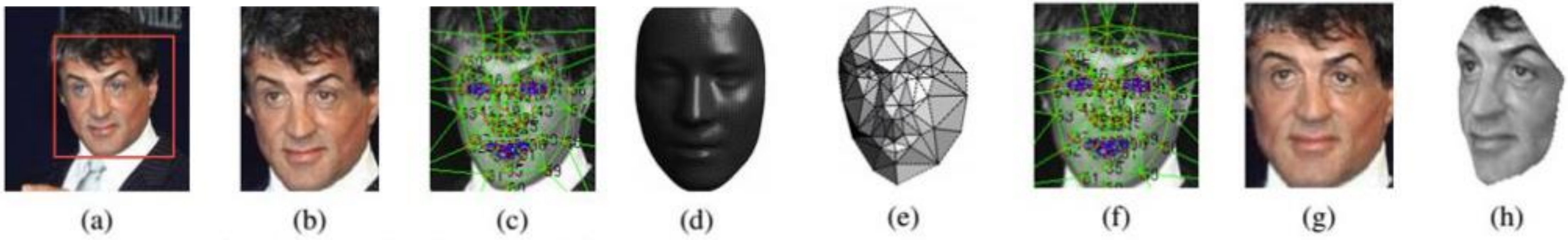
Iris Automation provides safer drone operation with intelligent collision avoidance



Application: face detection and recognition

Facebook's DeepFace project nears human accuracy in identifying faces

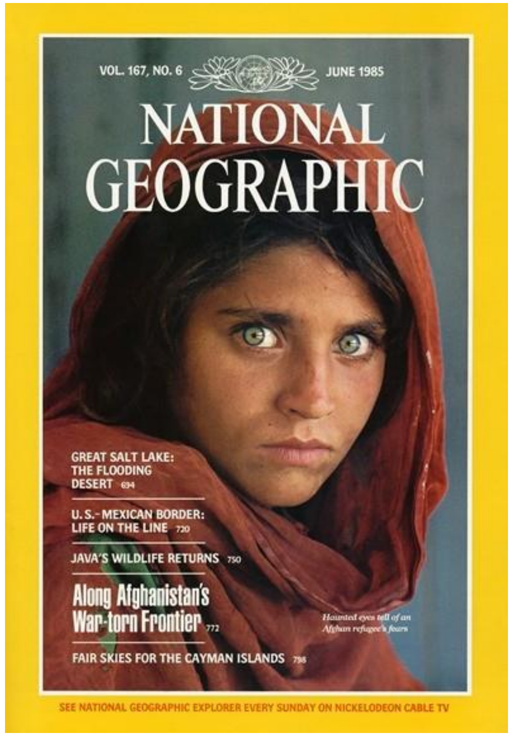




Application: face detection and recognition

For improving image capture on digital cameras

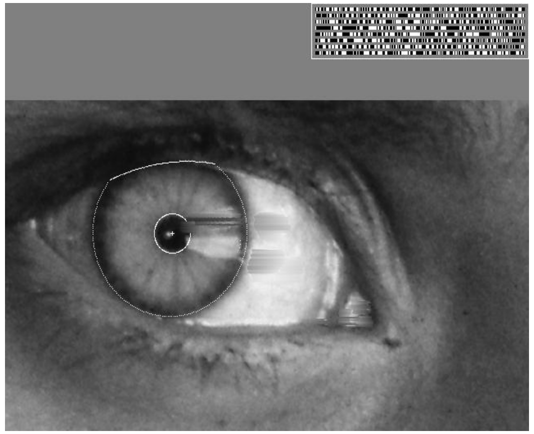
Application: vision-based biometrics

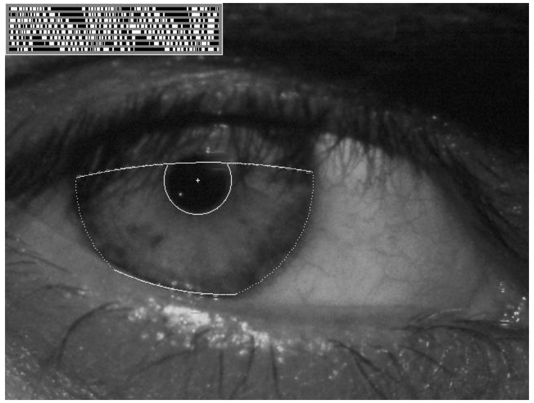


Who is she?

[How the Afghan girl was identified by her iris patterns]

The remarkable story of Sharbat Gula, first photographed in 1984 aged 12 in a refugee camp in Pakistan by National Geographic photographer Steve McCurry, and traced 18 years later to a remote part of Afghanistan where she was again photographed by McCurry…





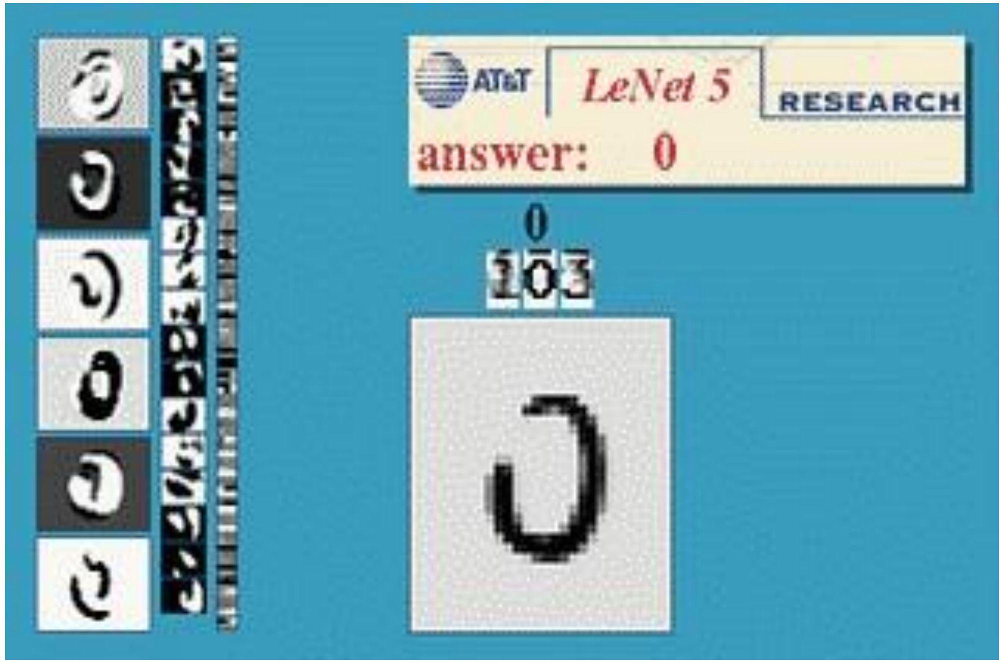
Application: logging in without a password

Fingerprint scanners
on modern laptops
and other devices

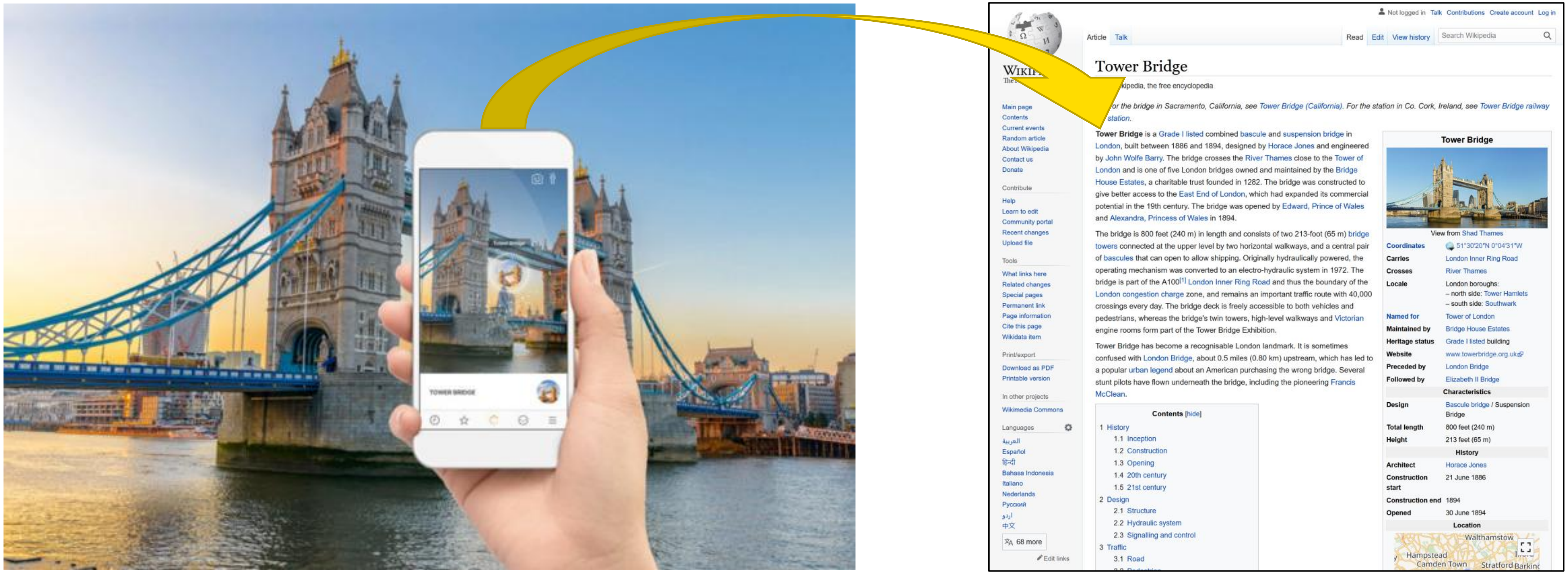
Windows Hello makes
logging in as easy as
looking at your PC

Application: optical character recognition (OCR)

Converting scanned documents or number plates to processable text

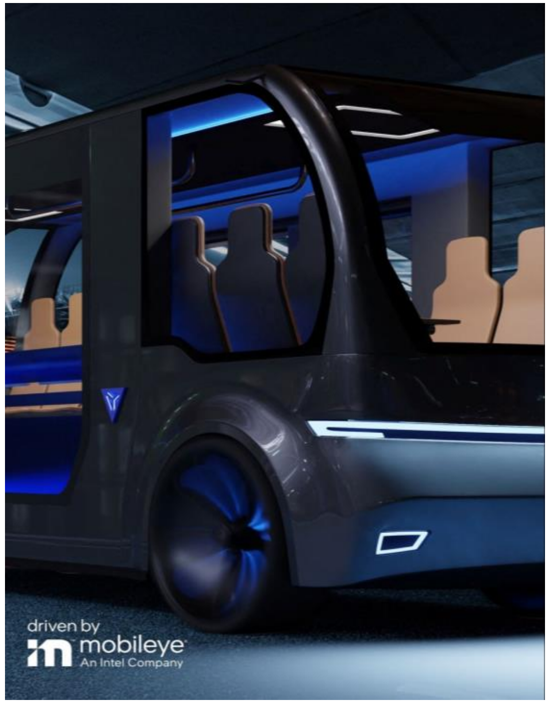


Application: landmark recognition



Application: autonomous vehicles

Intel's Mobileye makes cars safer and more autonomous



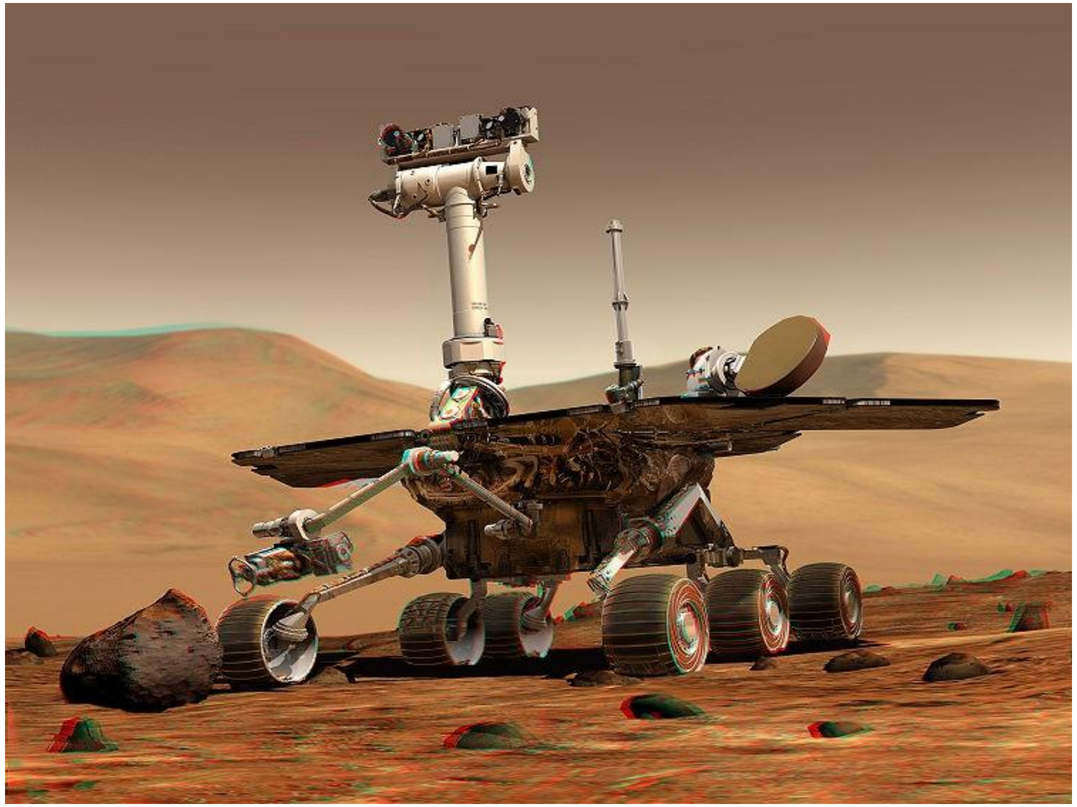
Application: space exploration

NASA's Mars Exploration Rover Spirit autonomously captured this picture in 2007

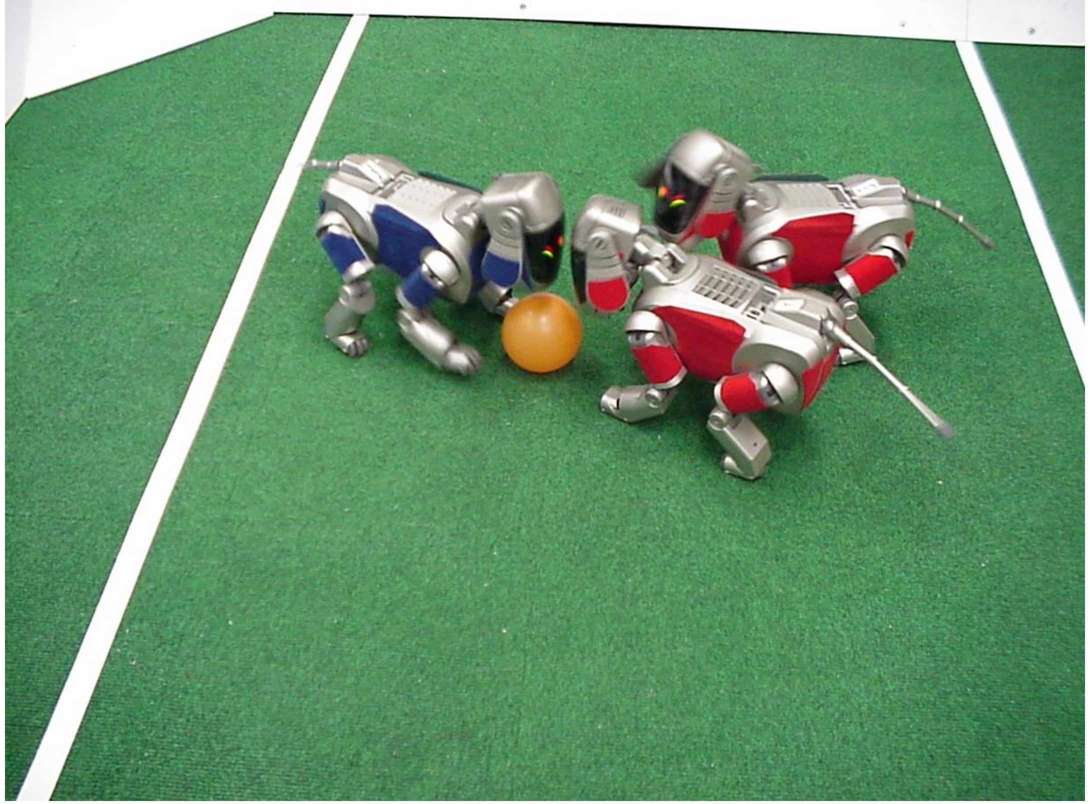
Vision systems used for panorama stitching, 3D terrain modeling, obstacle detection, position tracking

See [Computer Vision on Mars] for more information

Application: machine vision in robotics



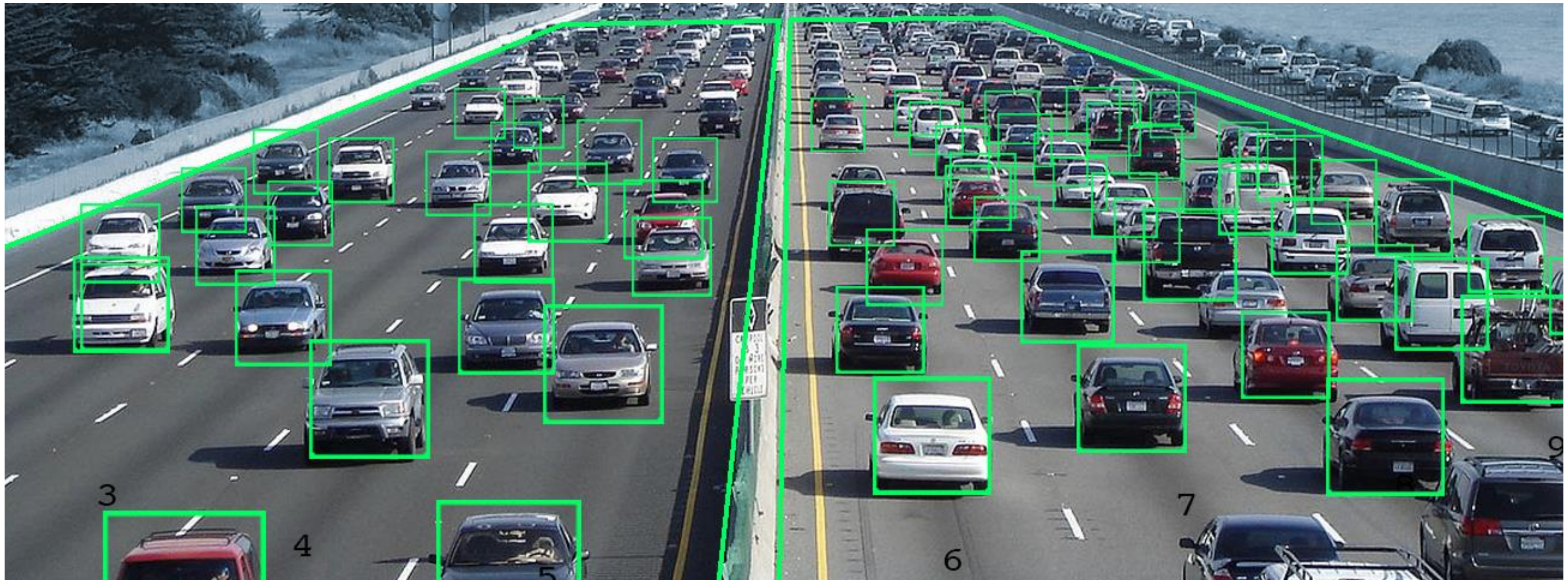
NASA's Mars Spirit Rover



RoboCup

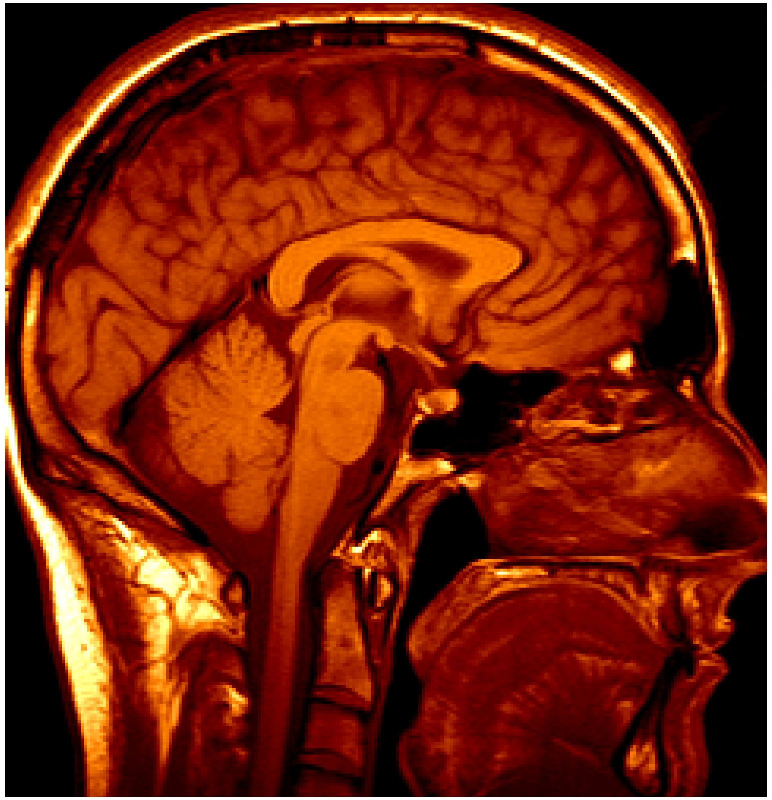
Application: video surveillance

Software from [TrafficVision] turns traffic cameras into an intelligent sensors



- Traffic monitoring
- Action recognition
- Incident detection
- Speed estimation
- Vehicle counting
- ...

Application: medical imaging



Computer Aided Diagnosis

Image Guided Surgery

Goals and challenges of computer vision

- Extract useful information from images, both metric and semantic
- Data ambiguity, heterogeneity, and complexity are a big challenge
- Significant progress in recent years due to various improvements:
  - Processing power
  - Memory capacity
  - Storage capacity
  - Data availability
- Careful design of every step in the computer vision workflow:

  images > measurements > representation > algorithms for learning and inference

Computer vision tasks

- Obtain simple inferences from individual pixel values
- Group pixels to separate object regions or infer shape information
- Recognise objects using geometric or statistical pixel information
- Combine information from multiple images into a coherent whole

Requires understanding of the **physics of imaging** and the use of **mathematical and statistical models** for information extraction

Low-level computer vision

This concerns mostly **image processing** (image in > image out)

- **Sensing**: image capture and digitization
- **Preprocessing**: suppress noise and enhance object features
- **Segmentation**: separate objects from background and partition them
- **Description**: compute feature maps which differentiate objects
- **Labeling**: assign labels to image segments (regions of interest)

High-level computer vision

This concerns deeper **image analysis** (image in > knowledge out)

- **Detection**: detect, localize, count objects of interest
- **Recognition**: identify object types based on low-level information
- **Classification**: assign unique labels to recognized objects
- **Interpretation**: assign meaning to groups of recognized objects
- **Scene analysis**: complete understanding of the captured scene

Assumed knowledge

To do this course successfully you should:

- Be able to program well in **Python** or willing to learn it independently
- Be familiar with **data structures and algorithms** and basic **statistics**
- Be able/learn to **use software packages** (OpenCV, Scikit-Learn, Keras)
- Be familiar with **vector calculus and linear algebra** or willing to learn it

Please self-assess **before** deciding to stay/enroll in the course

Student learning outcomes

After completing this course, you will be able to:

- **Explain** basic scientific and engineering approaches to computer vision
- **Implement and test** computer vision algorithms using existing software
- **Build** larger computer vision applications by integrating software modules
- **Interpret and comment** on articles in the computer vision literature

Course topics and lecturers

Week	Topic	Lecturer
1	Introduction & Image Formation	A/Prof Yang Song
2	Image Processing	A/Prof Yang Song
3	Feature Representation	A/Prof Yang Song
4	Pattern Recognition	A/Prof Yang Song
5	Image Segmentation	Dr Hao Xue
6	Flexible Week	(potentially an additional Project Info session online)
7	Deep Learning I, Project Introduction	Dr Hao Xue, A/Prof Yang Song
8	Deep Learning II	Dr Hao Xue
9	Motion and Tracking, Project consultation	Dr Hao Xue, A/Prof Yang Song
10	Extra Topics & Revision	A/Prof Yang Song

Weekly class structure

- **Lectures**: Tuesdays 2-4pm & Wednesdays 2-4pm (hybrid mode)

  Both lectures will be on campus and online at the same time (Echo360 links in Moodle)

- **Tutorial Sessions**: One hour per week in Weeks 2-5, 7-10 (via BB Collab)

  Lab and project consultations online with your assigned tutor (links in Moodle)

  Wednesdays 4-5pm, or Thursdays 4-5pm or 5-6pm, or Fridays 12-1pm (as signed up)

  Tutorial groups will be assigned by Week 1 – each group will have two tutors

Assessments

Assessment	Marks	Release	Due	Where
Lab Work (4x)	10%	Weeks 2, 3, 4, 5	Weeks 3, 4, 5, 7	Online
Group Project	40%	Week 5	Week 10	Online
Exam	50%	Exam Day	Exam Day	On Campus (CSE Labs)

Important – Form project groups of 5 members by Week 5

- Create “Project Group” in WebCMS3 (we can assign you to a group as well)
- Submission: a presentation recording, a report, and code

Final Exam:

- On campus, CSE labs, 2 hours, multiple choice questions

Communication modes and etiquette

- **Online forum (Ed) is your first port of call** for queries of wider interest on lectures, labs, project, exam, and general administrative things
- **Contact the LiC** for late submission, absence, and more specific questions about the labs, project, and assessment contents (yang.song1@unsw.edu.au)
- **Contact the course admin** for issues with enrolment, file submission, group enrolment, or other administration related matters (cs9517@cse.unsw.edu.au)
- **Team is committed to respond quickly** to queries with a maximum turnaround of 24 hours
- **Do observe standards of equity and respect** in dealing with all students and staff, in person, emails, forum posts, and all other communication
- **Language of communication is English**

Special Consideration

- If your work in this course is affected by unforeseen adverse circumstances, you should **apply for Special Consideration via the UNSW website**
- UNSW handles Special Consideration requests centrally, so **use the website and do not email the Lecturer in Charge about Special Consideration requests**
- Special Consideration **requests must be accompanied by documentation**
- **Marks are calculated the same way** as other students who sat the original assessment
- If you are awarded a Supplementary Exam and do not attend, your exam mark will be zero

Late submission penalty:

- Unless you have been granted Special Consideration, work submitted after the deadline **during term** will incur a penalty of 5% per day, capped at 5 days, after which submissions are no longer accepted. **For the final examination, university exam rules apply.**

Plagiarism Policy

READ the UNSW Policy and Procedure on this (links in the course outline on WebCMS3)

For the purposes of COMP9517, plagiarism includes copying or obtaining all, or a substantial part, of the material for your assignment, whether written or graphical report material, or software code, **without written acknowledgement** in your assignment from:

- A location on the internet (including ChatGPT, GitHub Copilot, Google Bard etc.)
- A book, article or other written document (published or unpublished) in any form
- Another student, whether in your class or another class, at UNSW or elsewhere
- Someone else (for example someone who writes assignments for money)

Plagiarism Policy

- If you copy material from another student or non-student **with acknowledgement**, you will not be penalized for plagiarism, but the marks you get for this will be **at the marker's discretion** and will reflect the marker's perception of the amount of work you put into finding and/or adapting the code/text
- If you use text found in a publication (on the internet or elsewhere), the marks you get for this will be **at the marker's discretion** and will reflect the marker's perception of the amount of work you put into finding and/or adapting the text

Assessments provide opportunities for you to develop important skills

Use these opportunities

Copyright Notice

- All course materials made available to you are copyrighted by UNSW
- Reproducing, publishing, posting, distributing, or translating is a copyright infringement
- Infringements will be reported to UNSW Student Conduct and Integrity for action

Further information on WebCMS3

Please be sure you are familiar with:

- Communication Etiquette
- Special Consideration
- Student Conduct
- Plagiarism Policy
- Academic Integrity

Further reading on discussed topics

In the lectures we will be referring to various online resources for further reading:

- Richard Szeliski, Computer Vision: Algorithms and Applications, 2nd Edition, Springer, 2021
- Dana H. Ballard and Christopher M. Brown, Computer Vision, Prentice Hall, 1982
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016
- David A. Forsyth and Jean Ponce, Computer Vision: A Modern Approach, Prentice Hall, 2011
- Simon J. D. Prince, Computer Vision: Models, Learning and Inference, CUP, 2012

And other books, articles, and resources online or via the UNSW Library

Further reading on discussed topics

- Chapter 1 of Szeliski for a general introduction to computer vision
- Appendix A of Szeliski for a recap of linear algebra and numerical techniques