

COMP9517

Computer Vision

2025 Term 3 Week 1

Professor Erik Meijering



UNSW
SYDNEY



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 application

Misclassification example
from OpenAI in 2021:

<https://openai.com/index/multimodal-neurons/>



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 (ChatGPT)



Can you describe this image?

This is a famous historical photograph depicting the aftermath of a dramatic train accident. The image shows a steam locomotive that has crashed through the wall of a railway station and fallen nose-first to the ground below.

The scene is from the **Montparnasse derailment**, which occurred on **October 22, 1895**, at the **Gare Montparnasse** railway station in **Paris, France**. The train overran the buffer stop at the end of the track due to a combination of excessive speed and brake failure. The locomotive plowed through the station wall and fell to the street below, resulting in one fatality — a woman on the street who was struck by falling debris.

The dramatic nature of the crash and the surreal angle of the locomotive make this image one of the most iconic and widely recognized photos of a train accident in history.

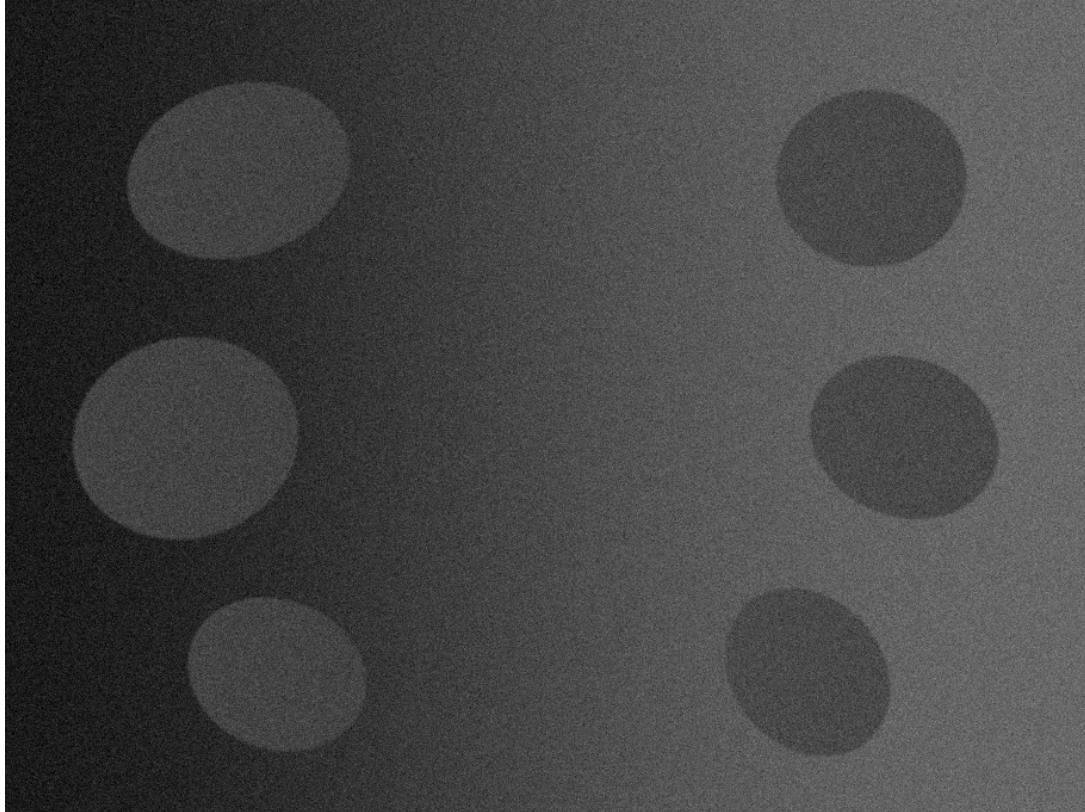


Can you describe this image?

This image is a playful visual pun. It shows a **green apple** with a piece of white paper taped around it. On the paper, the word "**iPod**" is handwritten in black ink, mimicking Apple's iconic product branding.

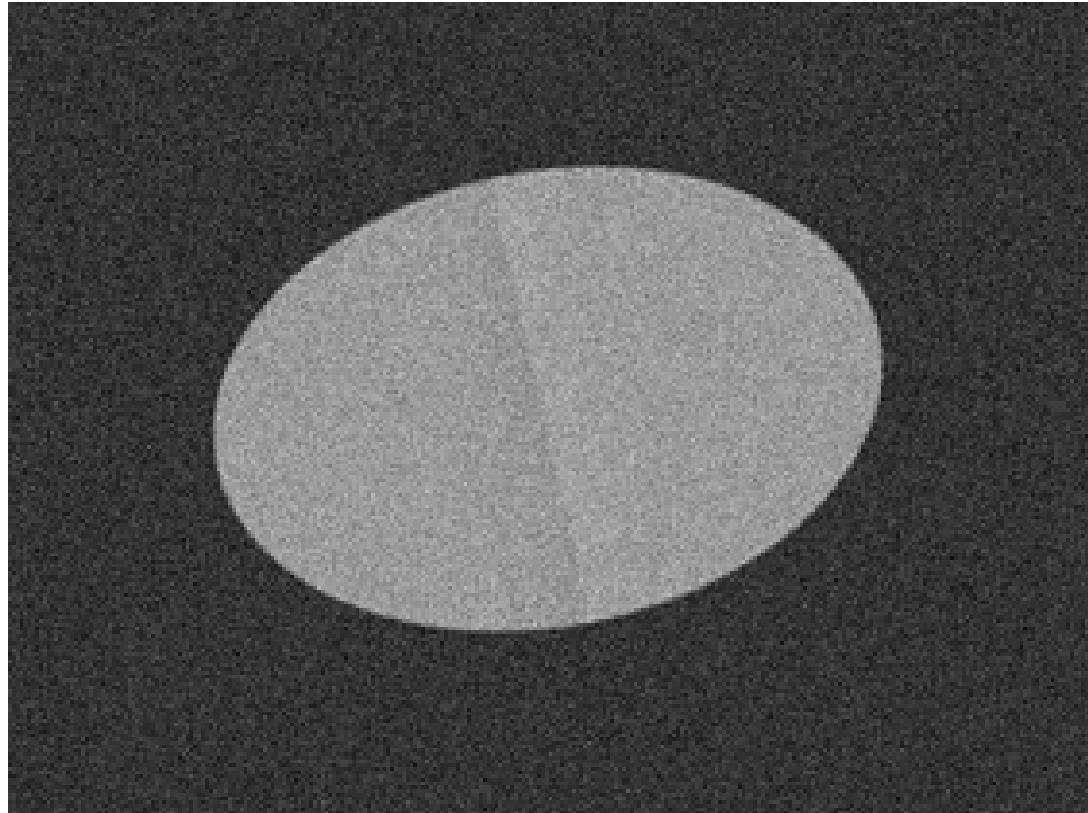
The humor comes from the fact that it's literally an "**iPod**" — a combination of the Apple product name and a **literal apple**, making a pun out of the fruit and the famous digital music player. It's a simple, clever joke blending wordplay and visual humor.

Human vision has its limitations...



Which objects are brighter?

Human vision has its limitations...



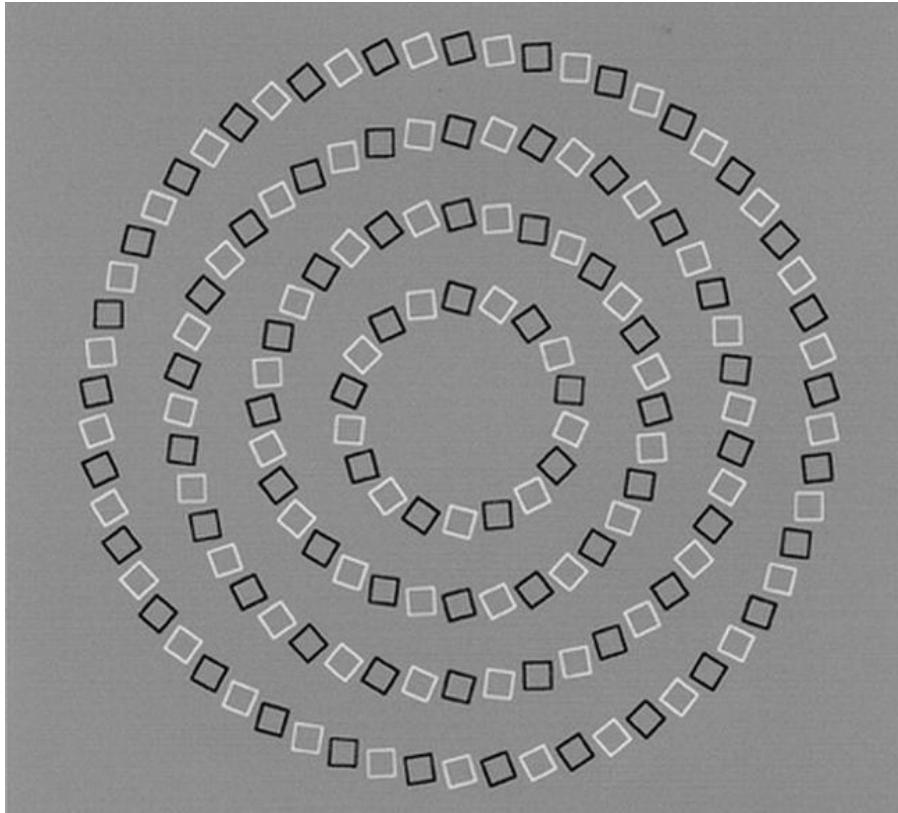
Which side of this object is brighter?

Human vision has its limitations...



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

Human vision has its limitations...



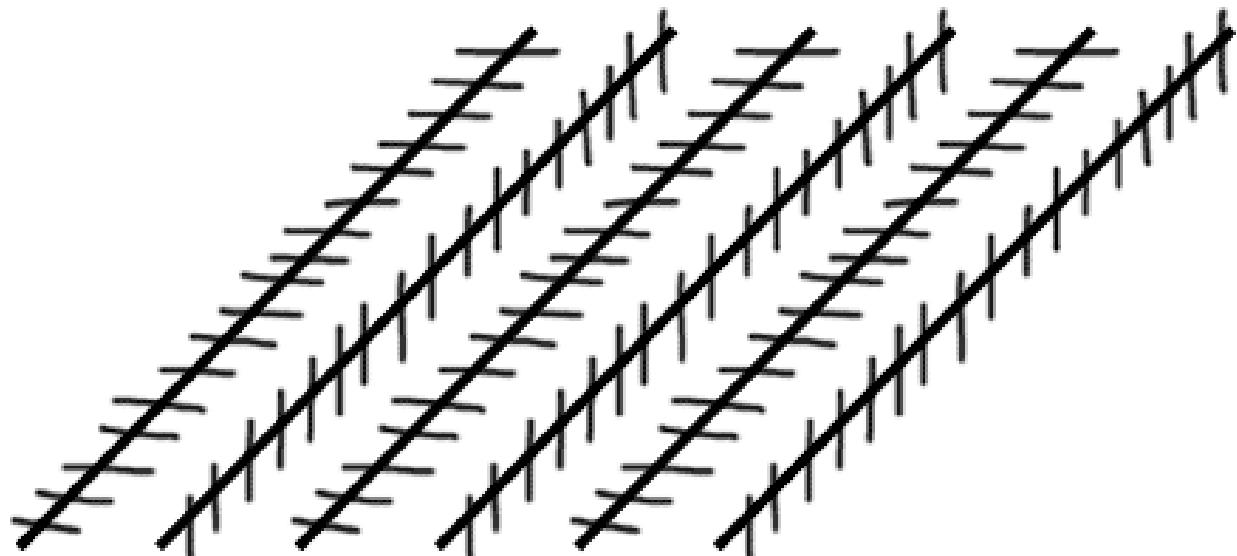
What pattern do the
squares form?

Human vision has its limitations...



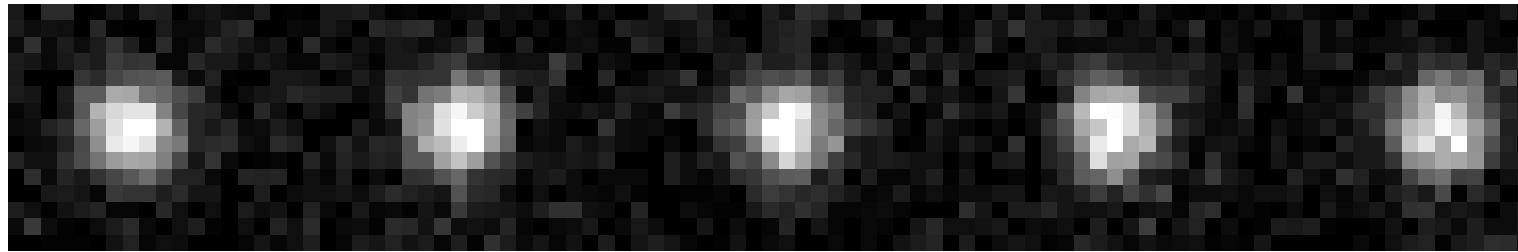
What object do you
see in this image?

Human vision has its limitations...



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=a7efEqgplrE>

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!



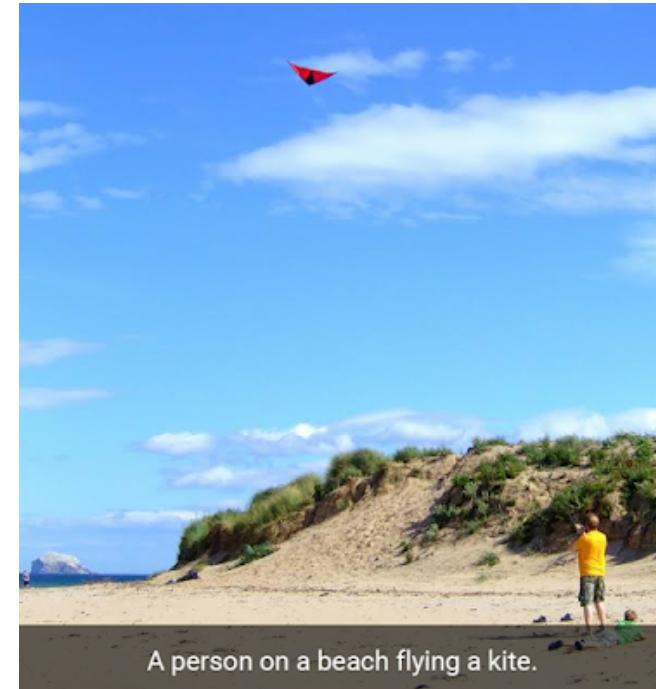
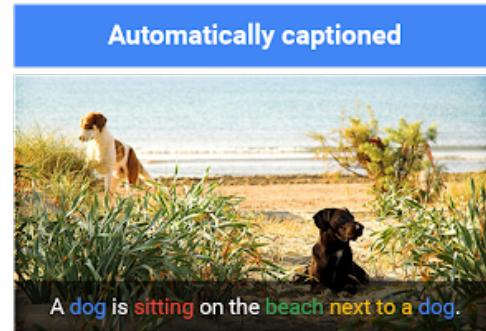
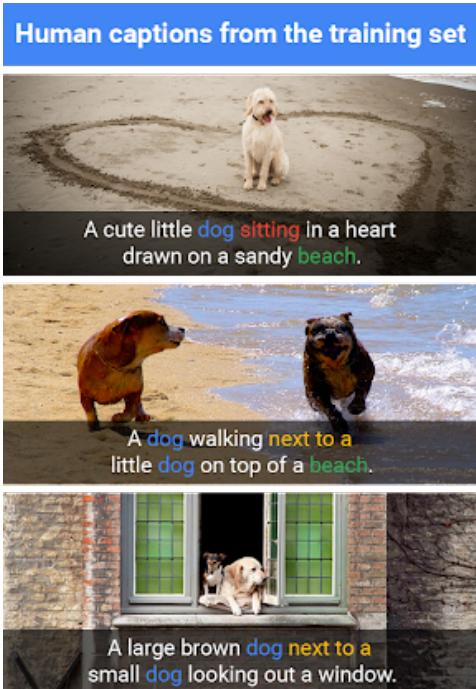
Application: 3D shape reconstruction

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



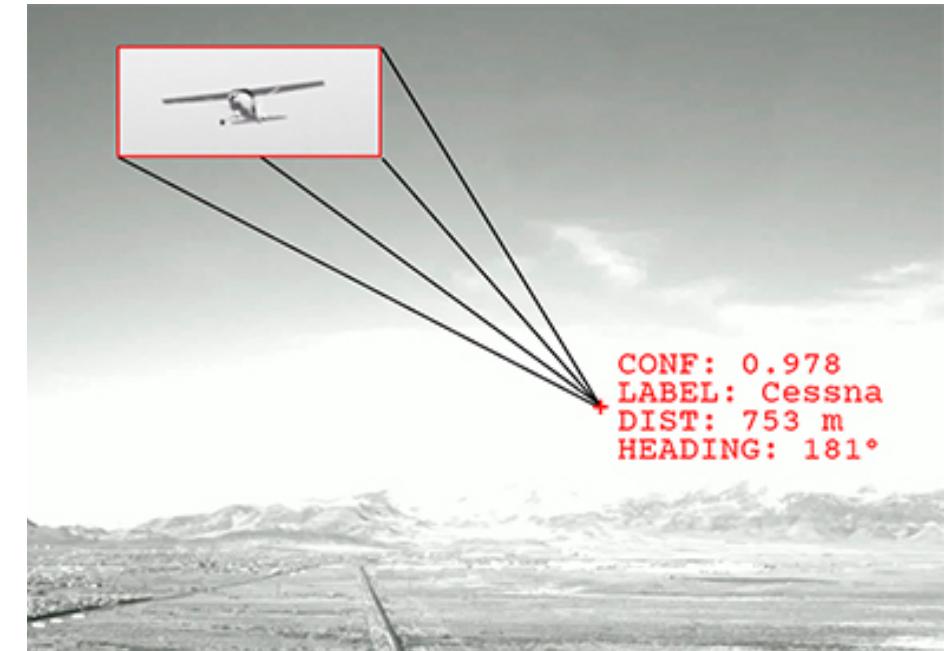
Application: image captioning

[Google's Show and Tell](#) open-source image captioning model in TensorFlow



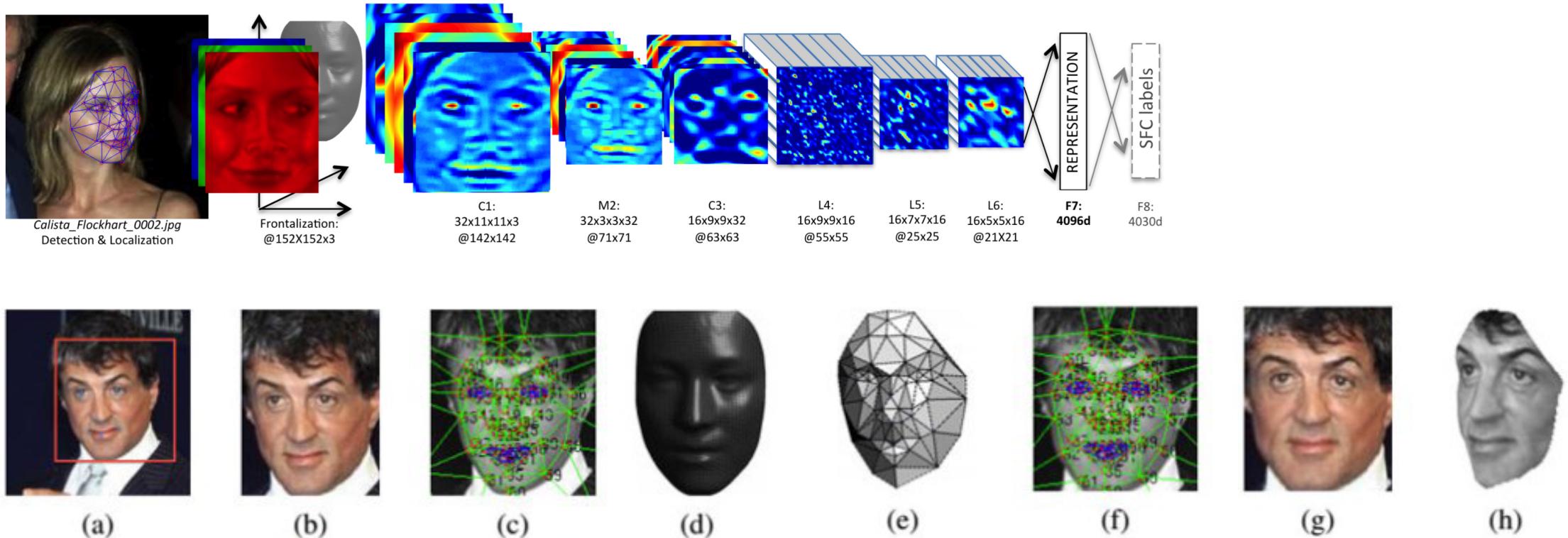
Application: intelligent collision avoidance

[Iris Automation](#) provides safer drone operation with intelligent collision avoidance



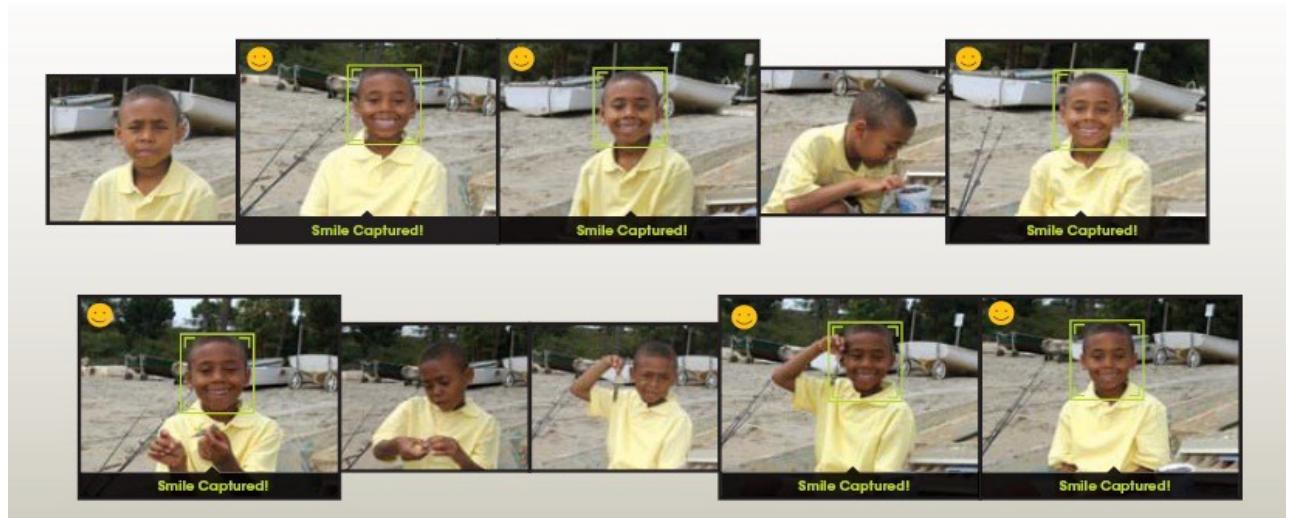
Application: face detection and recognition

Facebook's [DeepFace](#) project nears human accuracy in identifying faces

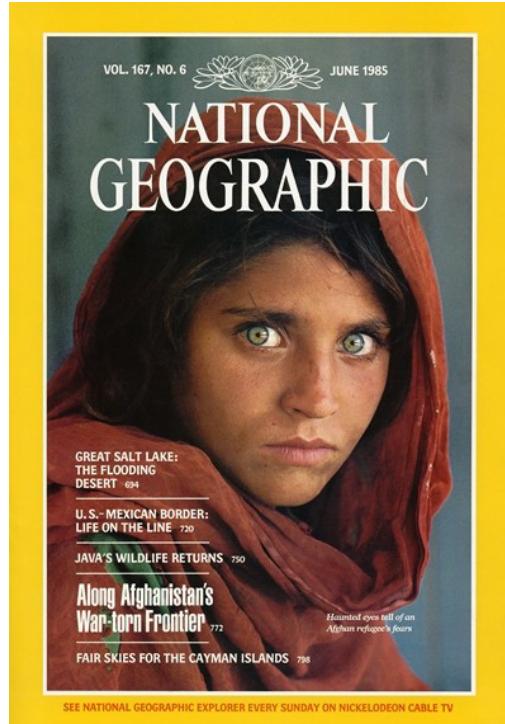


Application: face detection and recognition

For improving image capture on digital cameras



Application: vision-based biometrics

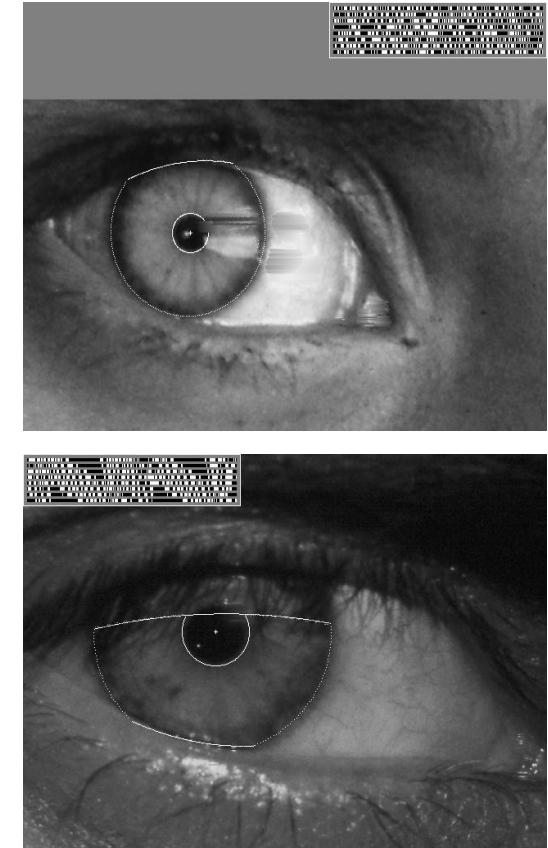


Who is she?



[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



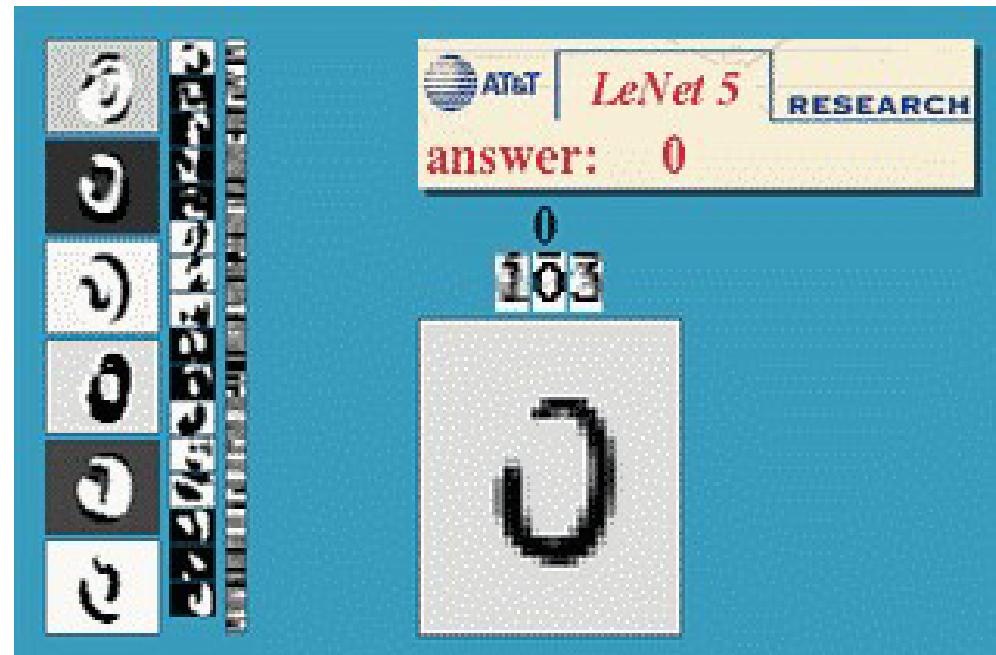
Fingerprint scanners
on modern laptops
and other devices



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



The screenshot shows the Wikipedia article for "Tower Bridge". The page includes a sidebar with navigation links like "Main page", "Contents", and "Help". The main content area starts with a brief summary and then delves into detailed sections about the bridge's history, design, and physical characteristics. It also lists its "Carries", "Crosses", and "Locality". A map at the bottom indicates the bridge's location in London, spanning the River Thames between Walthamstow to the north and Stratford to the south. The page is well-structured with headings, sub-headings, and a "Contents" section.

Not logged in | Talk | Contributions | Create account | Log in

Article | Talk | Read | Edit | View history | Search Wikipedia

Tower Bridge

View the page as it appears to a user in IP address 128.110.113.111.

Tower Bridge is a Grade I listed combined bascule and suspension bridge in London, built between 1886 and 1894, designed by Horace Jones and engineered by John Wolfe Barry. The bridge crosses the River Thames close to the Tower of London and is one of five London bridges owned and maintained by the Bridge House Estates, a charitable trust founded in 1282. The bridge was constructed to give better access to the East End of London, which had expanded its commercial potential in the 19th century. The bridge was opened by Edward, Prince of Wales and Alexandra, Princess of Wales in 1894.

The bridge is 800 feet (240 m) in length and consists of two 213-foot (65 m) bridge towers connected at the upper level by two horizontal walkways, and a central pair of bascules that can open to allow shipping. Originally hydraulically powered, the operating mechanism was converted to an electro-hydraulic system in 1972. The bridge is part of the A100^[1] London Inner Ring Road and thus the boundary of the London congestion charge zone, and remains an important traffic route with 40,000 crossings every day. The bridge deck is freely accessible to both vehicles and pedestrians, whereas the bridge's twin towers, high-level walkways and Victorian engine rooms form part of the Tower Bridge Exhibition.

Tower Bridge has become a recognisable London landmark. It is sometimes confused with London Bridge, about 0.5 miles (0.8 km) upstream, which has led to a popular urban legend about an American purchasing the wrong bridge. Several stunt pilots have flown underneath the bridge, including the pioneering Francis McClean.

Contents [hide]

1 History

- 1.1 Inception
- 1.2 Construction
- 1.3 Opening
- 1.4 20th century
- 1.5 21st century

2 Design

- 2.1 Structure
- 2.2 Hydraulic system
- 2.3 Signalling and control

3 Traffic

- 3.1 Road
- 3.2 Rail

Coordinates: 51°30'20"N 0°43'1"W
Carries: London Inner Ring Road
Crosses: River Thames
Locale: London boroughs:

- north side: Tower Hamlets
- south side: Southwark

Named for: Tower of London
Maintained by: Bridge House Estates
Heritage status: Grade I listed building
Website: www.towerbridge.org.uk
Preceded by: London Bridge
Followed by: Elizabeth II Bridge

Characteristics:

- Design: Bascule bridge / Suspension Bridge

Total length: 800 feet (240 m)
Height: 213 feet (65 m)

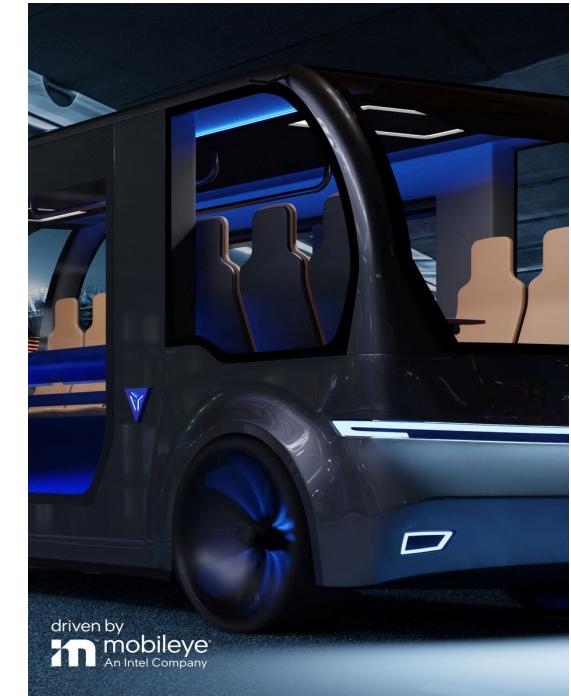
History:
Architect: Horace Jones
Construction start: 21 June 1886
Construction end: 1894
Opened: 30 June 1894

Location:
Walthamstow
Hampstead
Camden Town
Stratford
Barking

Application: autonomous vehicles

[Intel's Mobileye](#) makes cars safer and more autonomous

The screenshot shows the Mobileye website interface. At the top, there are two tabs: "manufacturer products" on the left and "consumer products" on the right. Below the tabs is a large banner with the slogan "Our Vision. Your Safety." and an illustration of a car from above with three types of cameras labeled: "rear looking camera" (top), "forward looking camera" (front), and "side looking camera" (sides). Below the banner are three main product sections: "EyeQ Vision on a Chip" (with an image of a chip), "Vision Applications" (with an image of a pedestrian crossing), and "AWS Advance Warning System" (with an image of a display screen). To the right of these sections is a "News" sidebar with a list of articles and a "Events" sidebar with links to "Mobileye at Equip Auto, Paris, France" and "Mobileye at SEMA, Las Vegas, NV".



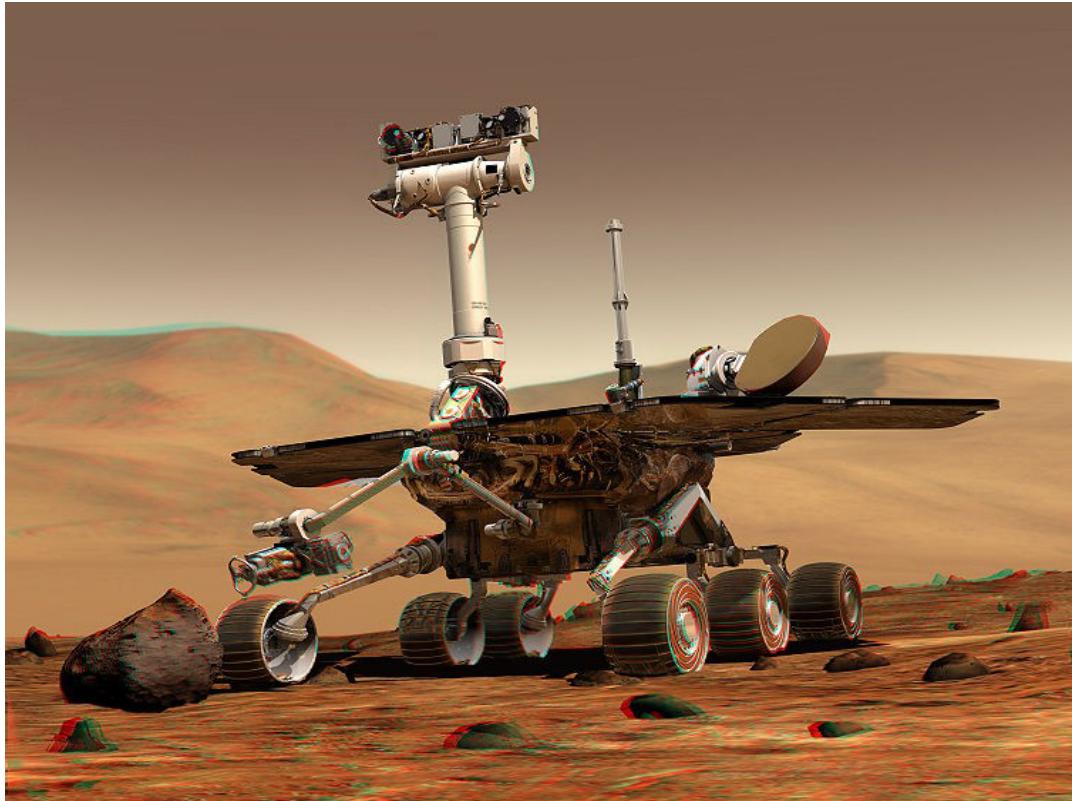
Application: space exploration

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

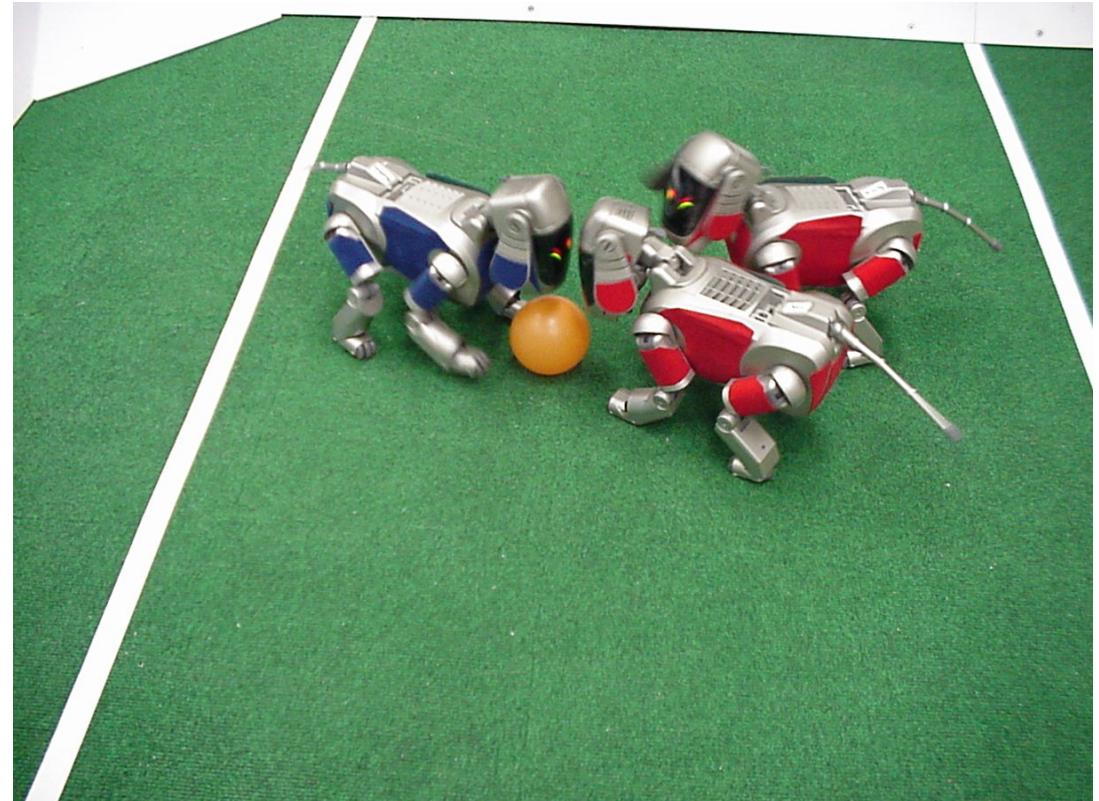


Vision systems used for panorama stitching, 3D terrain modeling, obstacle detection, position tracking
See [Computer Vision on Mars](#) for more information

Application: machine vision in robotics



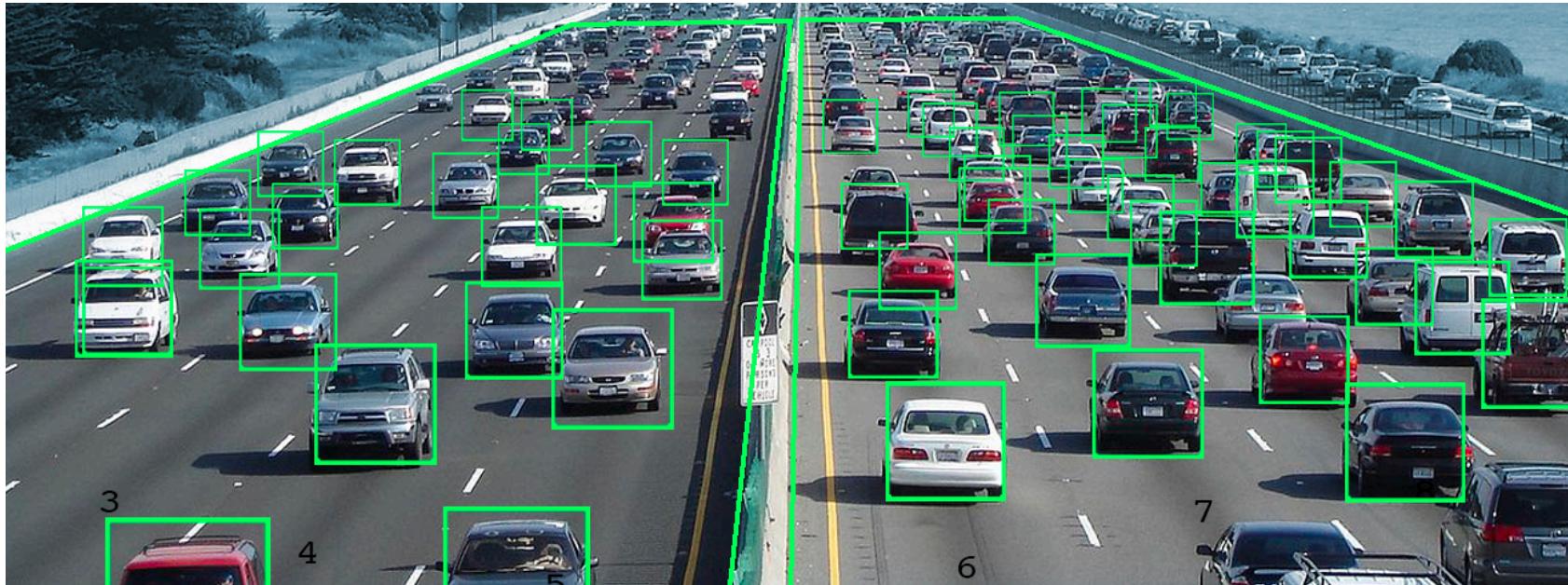
[NASA's Mars Spirit Rover](#)



[RoboCup](#)

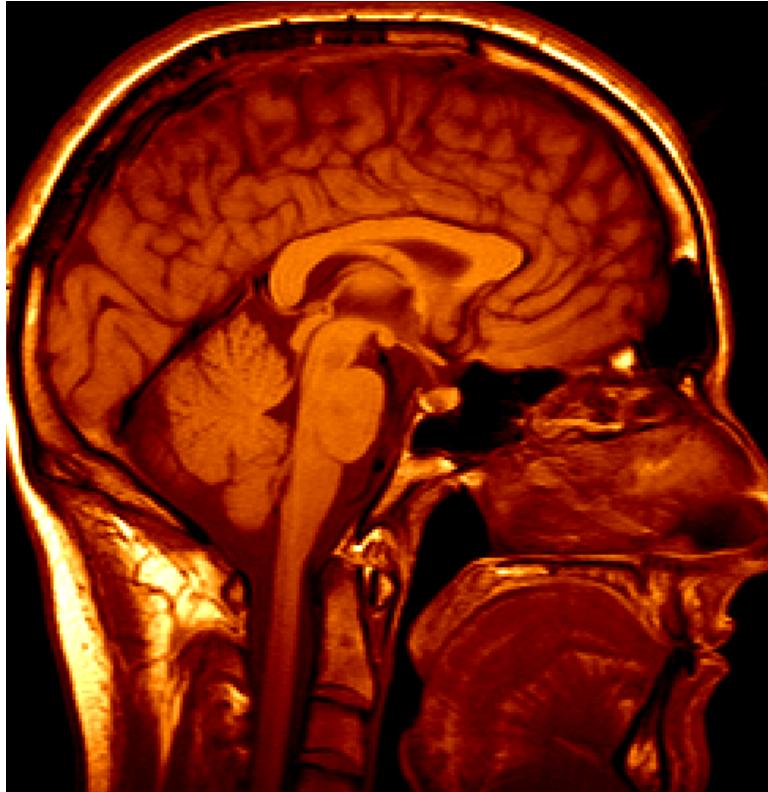
Application: video surveillance

Software from [TrafficVision](#) turns traffic cameras into 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/storage capacity
 - Data availability
 - Code/library 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 familiar with **data structures and algorithms** and basic **statistics**
- Be familiar with **vector calculus and linear algebra** or willing to learn it
- Be able to program well in **Python** or willing to learn it independently
- Be able/learn to **use software packages** (OpenCV, Scikit-Learn, Keras)

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	Professor Erik Meijering
2	Image Processing	Professor Erik Meijering
3	Feature Representation	Dr Sonit Singh
4	Pattern Recognition	Dr Sonit Singh
5	Image Segmentation	Professor Erik Meijering
6	Flexible Week (No Lectures)	
7	Deep Learning I	Dr Sonit Singh
8	Deep Learning II	Dr Sonit Singh
9	Motion and Tracking	Professor Erik Meijering
10	Applications & Revision	TBC

Weekly class structure

- **Lectures:** Mondays & Thursdays 11am–1pm (hybrid mode)
Both lectures will be on campus and online at the same time (Echo360 links in Moodle)
- **Tutorials:** Weeks 2-10 (in person / online via Blackboard Collaborate)
Lab consultations in weeks 2-5 with your assigned tutors (links in Moodle)
Project consultations in weeks 6-10 with your assigned tutors (same links in Moodle)
Tutorial groups will be assigned in week 1 (each group will have two tutors)
More details will follow later in week 1 (posted on WebCMS3)

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

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.**

Communication modes and etiquette

- **Online forum (Discourse) 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 (cs9517@cse.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** (<https://student.unsw.edu.au/special-consideration>)
- 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

See the course webpage on WebCMS3 for more detailed information and links

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