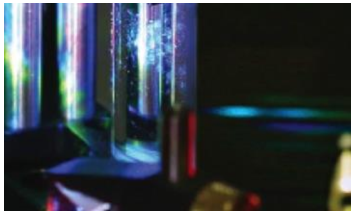
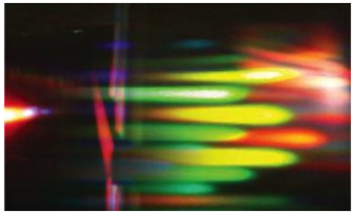
Imperial College London





COMP70058 Computer Vision

Lecture 1 - Introduction to Computer Vision

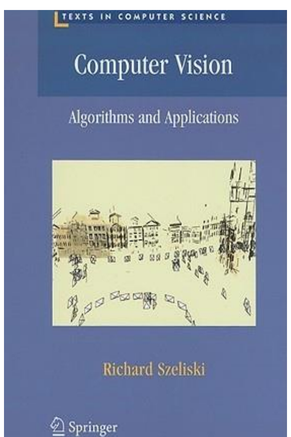
Stamatia (Matina) Giannarou, PhD
The Hamlyn Centre for Robotic Surgery

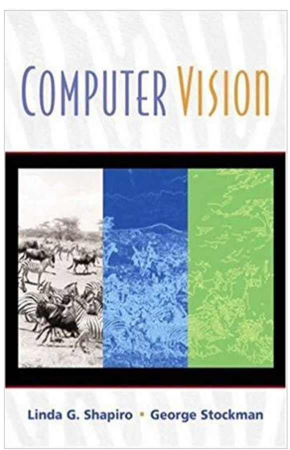
stamatia.giannarou@imperial.ac.uk

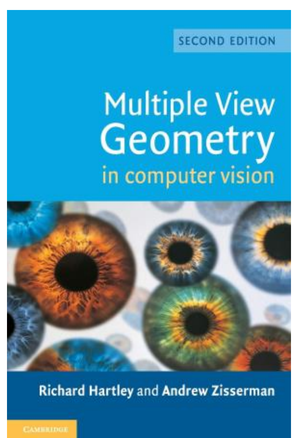
The Hamlyn Centre
for Robotic Surgery

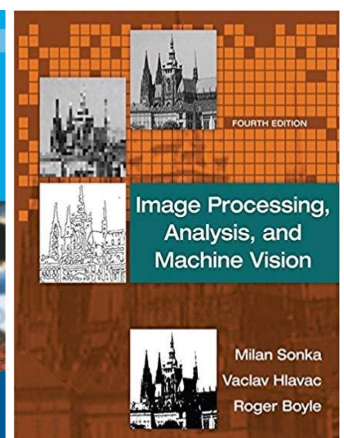
Aims, Structure and Resources

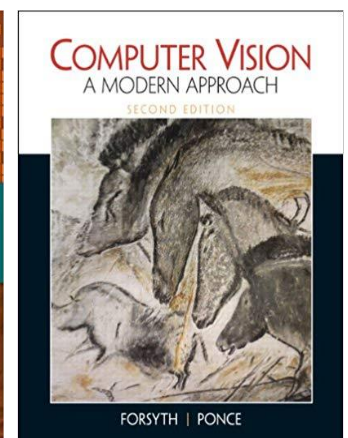
- To introduce the concepts behind computer-based recognition and extraction of features from raster images;
- To illustrate some successful applications of vision systems and their limitations;
- 19 lectures, 7 tutorials, 2 revision lectures;
- Lecture Slides – CATE (https://cate.doc.ic.ac.uk; 70058), which also includes assessed coursework details;
- The coursework will be released in week 5 with submission deadline in week 8.
- Reading list:

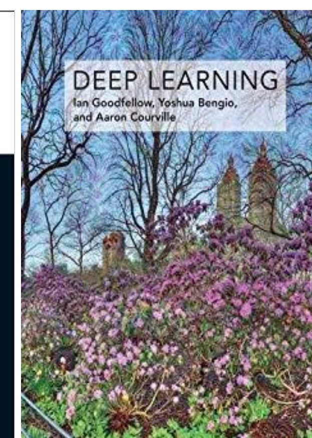












Topics

Image Formation
- Introduction to Computer Vision
- Image Formation

Image Processing
- Image Filtering and Edge Detection
- Hough Transform

Segmentation and Labelling
- Texture and Region-based Segmentation
- Relaxation Labelling

Recognition
- Shape Representation and Matching
- Fourier Methods
- Object Recognition

Image Sequence Processing
- Interest Point Detection
- Feature Descriptors and Matching
- Feature Tracking

3D reconstruction
- Camera Geometry
- Computational Stereo
- Photometric Stereo

Motion and Optical Flow

Deep Learning in Computer Vision

https://disruptionhub.com/computer-vision/

Teaching Assistants

Alfie Roddan

Alistair Weld

Chi Xu

Hadrien Reynaud

Joao Vieira-Cartucho

Michael Tanzer

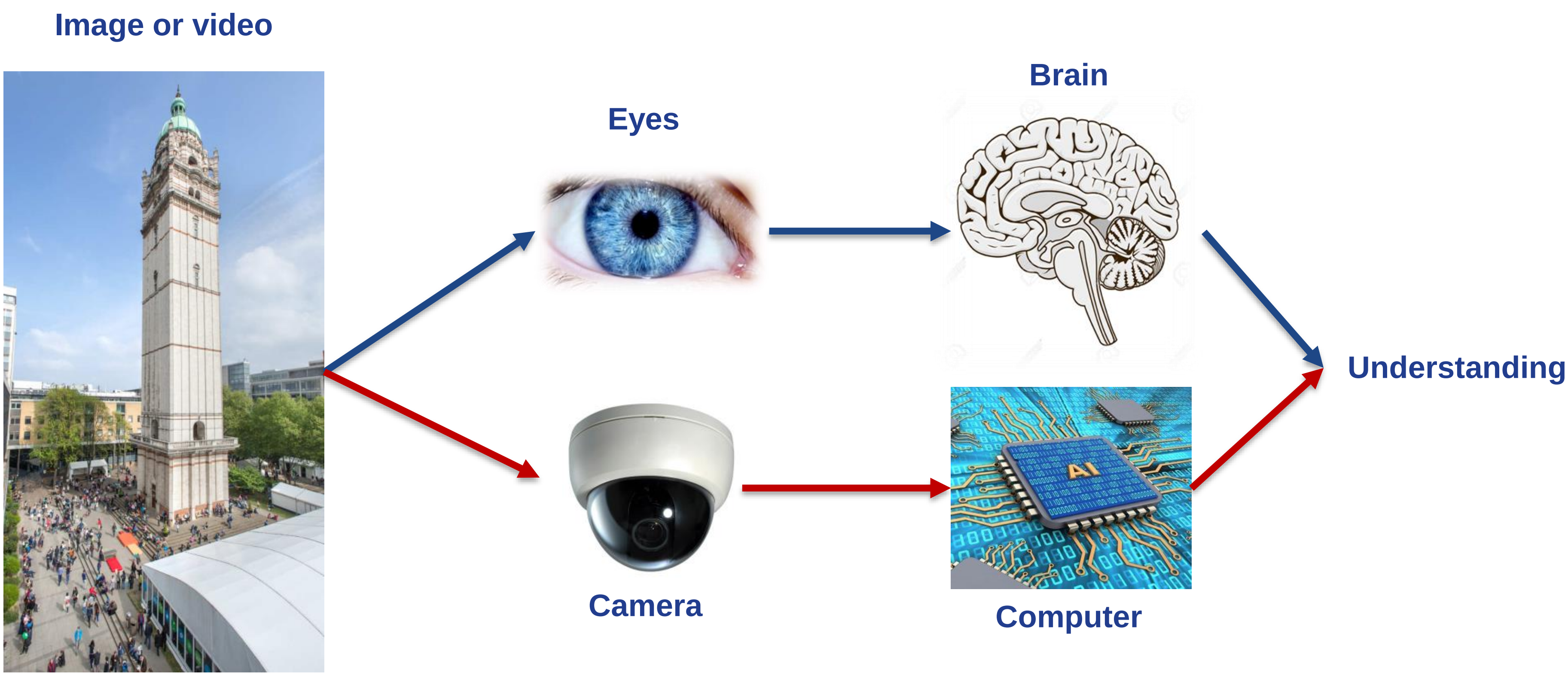
Timetable 2021

	Monday 11:00	Monday 12:00	Friday 09:00	Friday 10:00
Week 2	**Introduction to Computer Vision**	**Image Formation**	**Image Filtering and Edge Detection**	Tutorial: Edge Detection and Pre-Processing
Week 3	**Hough Transform**	**Shape Representation and Matching**	**Fourier Methods**	Tutorial: Fourier Methods and Moments
Week 4	**Texture and Region-based Segmentation**	**Relaxation Labelling**	**Interest Point Detection**	Tutorial: Texture and Region-based Segmentation
Week 5	**Feature Descriptors and Matching**	**Feature Tracking**	**Object Recognition**	Tutorial: Image Sequence Processing
Week 6	**Camera Geometry**	**Computational Stereo – Part 1**	**Computational Stereo – Part 2**	Tutorial: Computational Stereo and Assessed Coursework Clinic
Week 7	**NO LECTURE**	**NO LECTURE**	**NO LECTURE**	**NO LECTURE**
Week 8	**Photometric Stereo**	Tutorial: Photometric Stereo	**Motion and Optical Flow**	Tutorial: Motion and Optical Flow
Week 9	**Deep Learning in Computer Vision – Part 1**	**Deep Learning in Computer Vision – Part 2**	Revision Lecture	
Week 10	Exam Week			
Week 11	Exam Week			

Contents

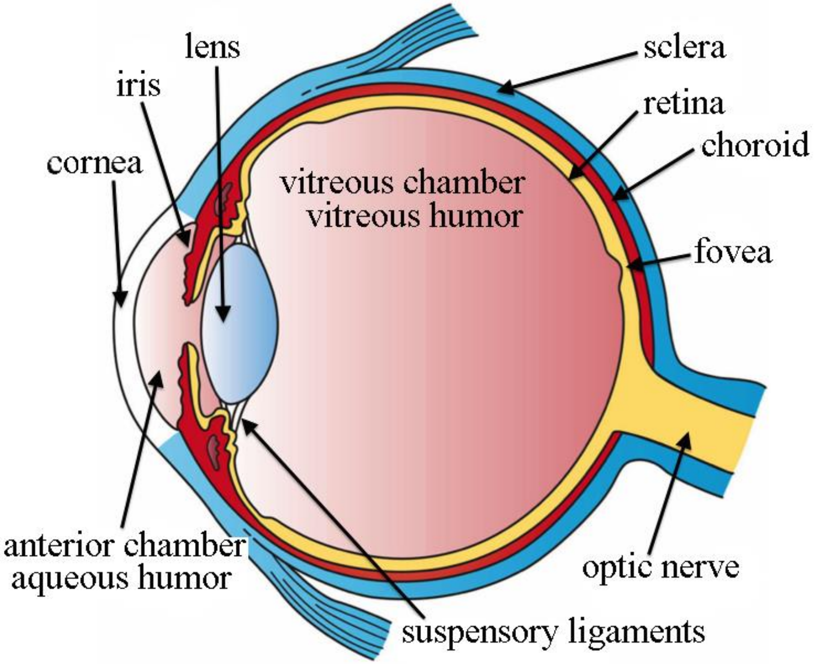
- What is Computer Vision?
- Applications of Computer Vision
- Graphics, image processing and vision
- Computational models for vision
- Levels of vision

What is computer vision?



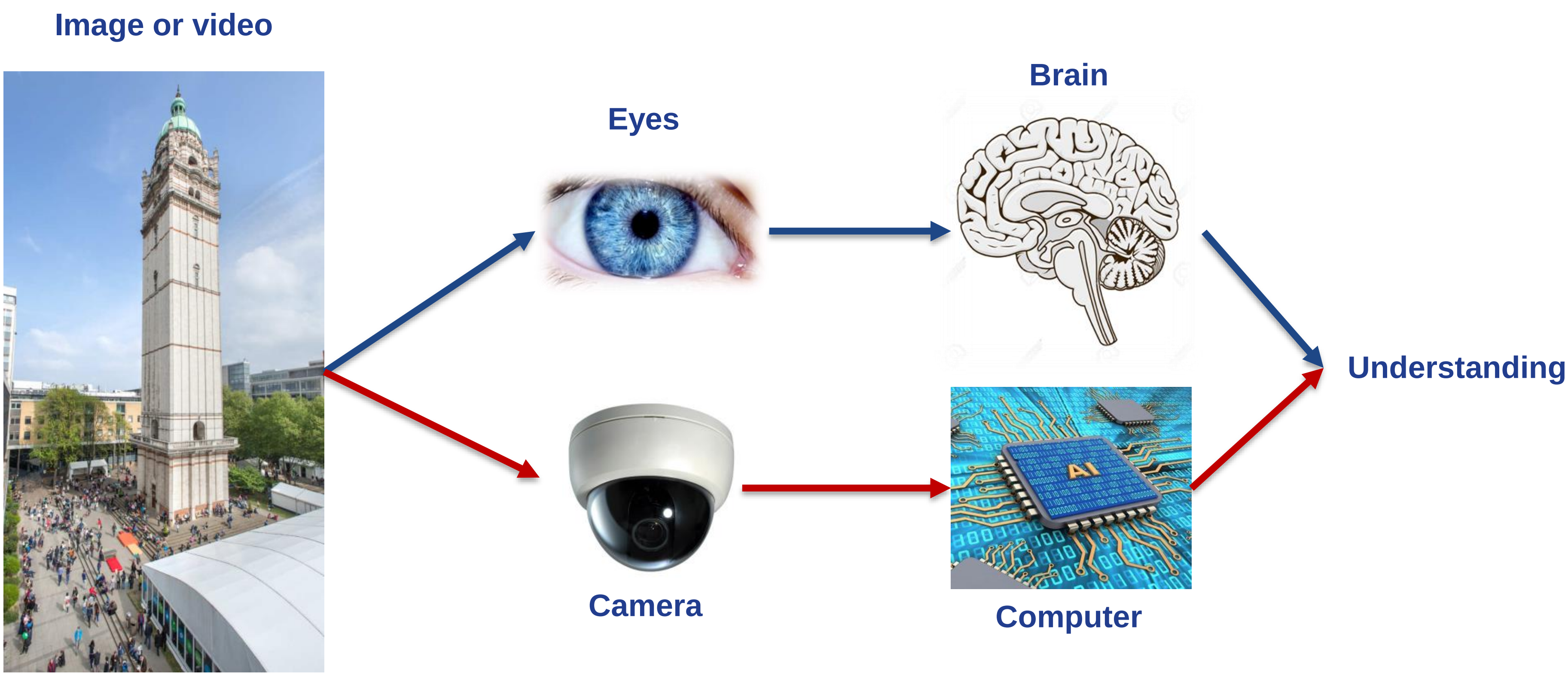
What is computer vision?

Image or video



Understanding

What is computer vision?

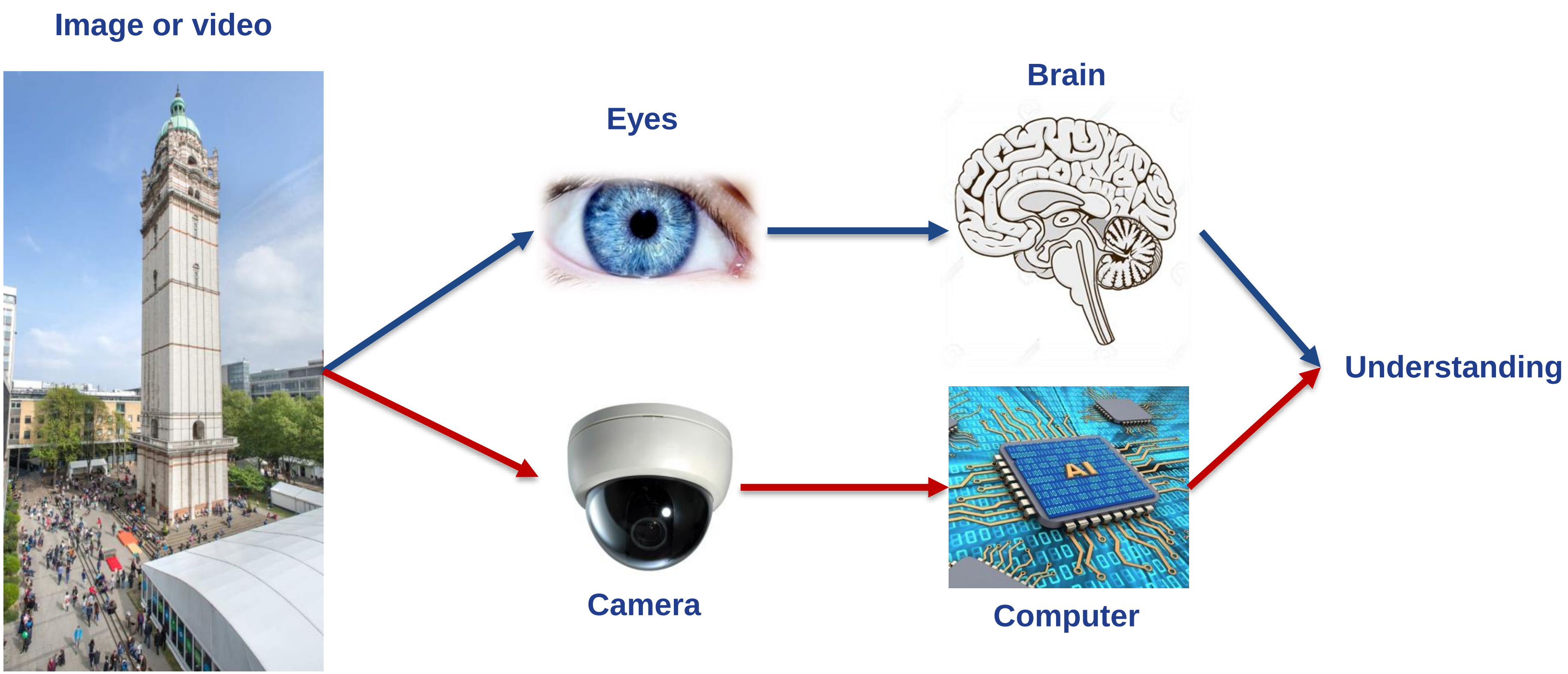


What is computer vision?

Image or video

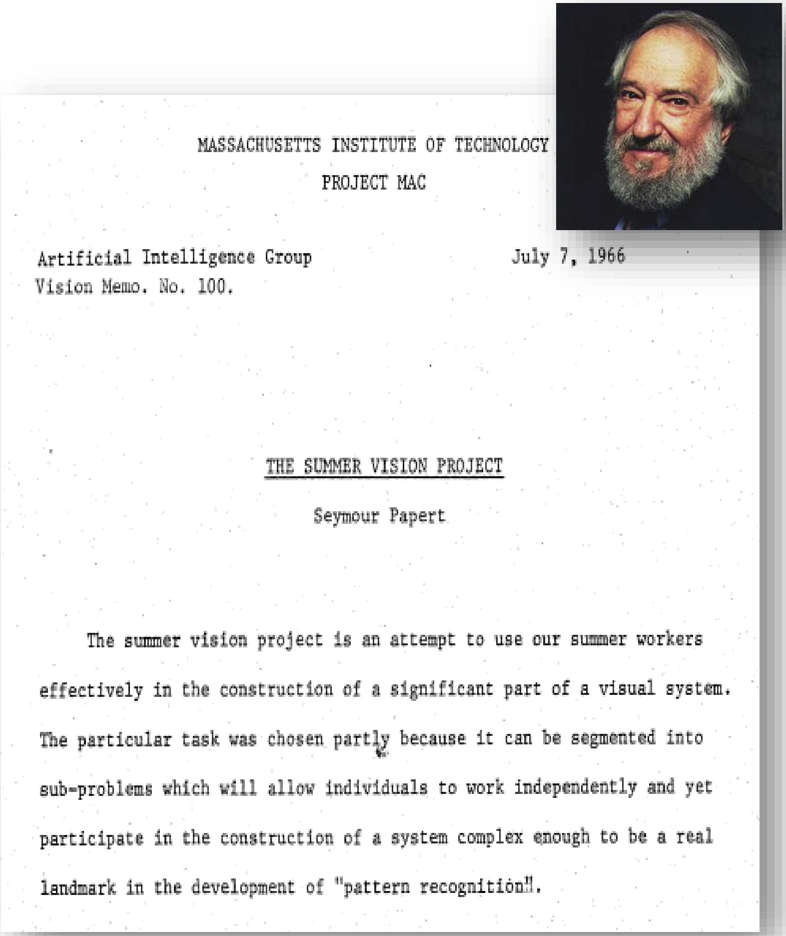
Understanding

What is computer vision?



Brief history of computer vision

- **1960's -** Interpretation of synthetic worlds
- **1970's -** Interpretation of images (edge detection, object representation, motion estimation)
- **1980's -** Shift towards geometry and more rigorous mathematical methods
- **1990's -** Statistical learning techniques, face recognition and 3D geometry applications
- **2000's -** Video processing, computer graphics, machine learning and large annotated datasets available
- **2010's -** Deep learning



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

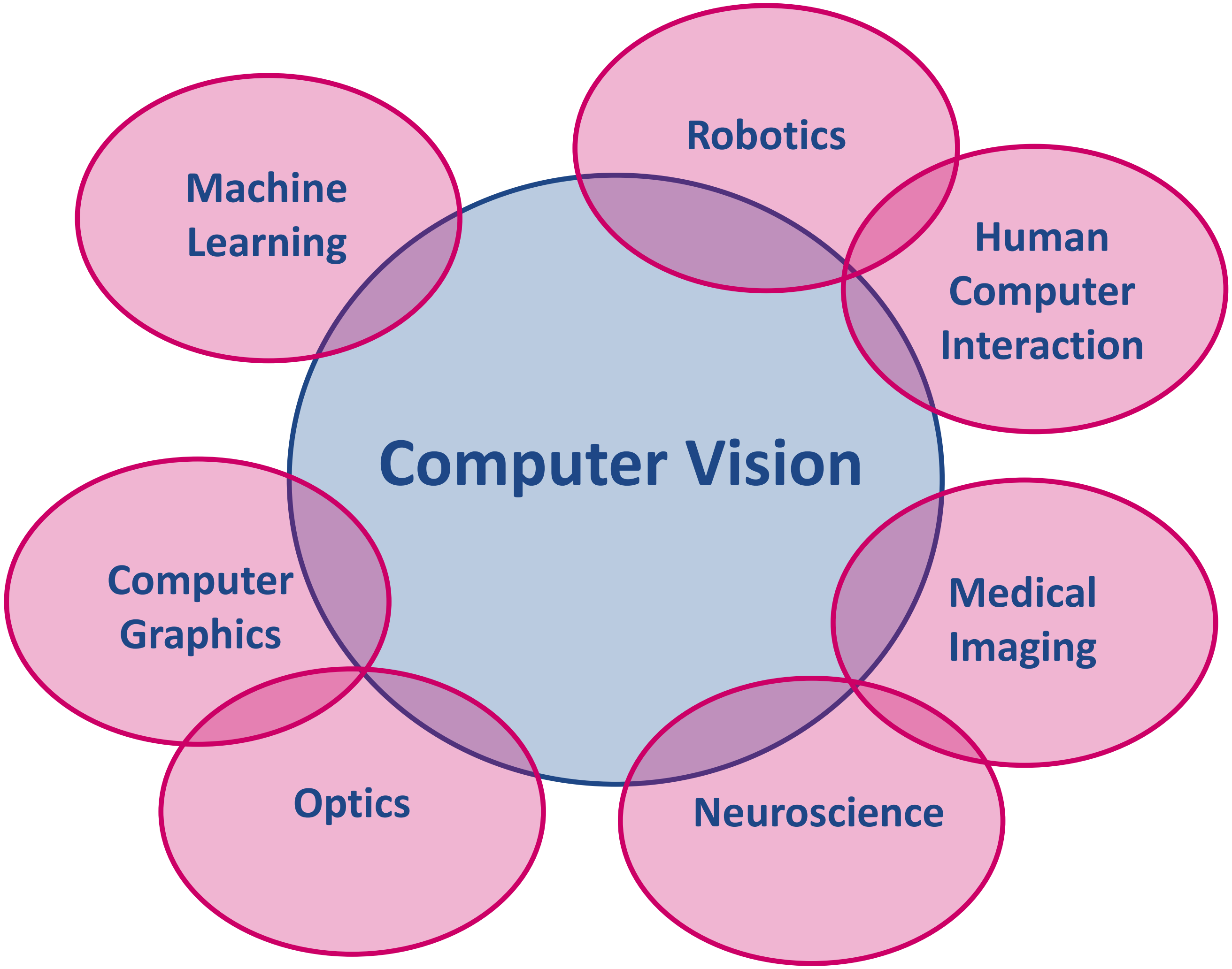
THE SUMMER VISION PROJECT

Seymour Papert

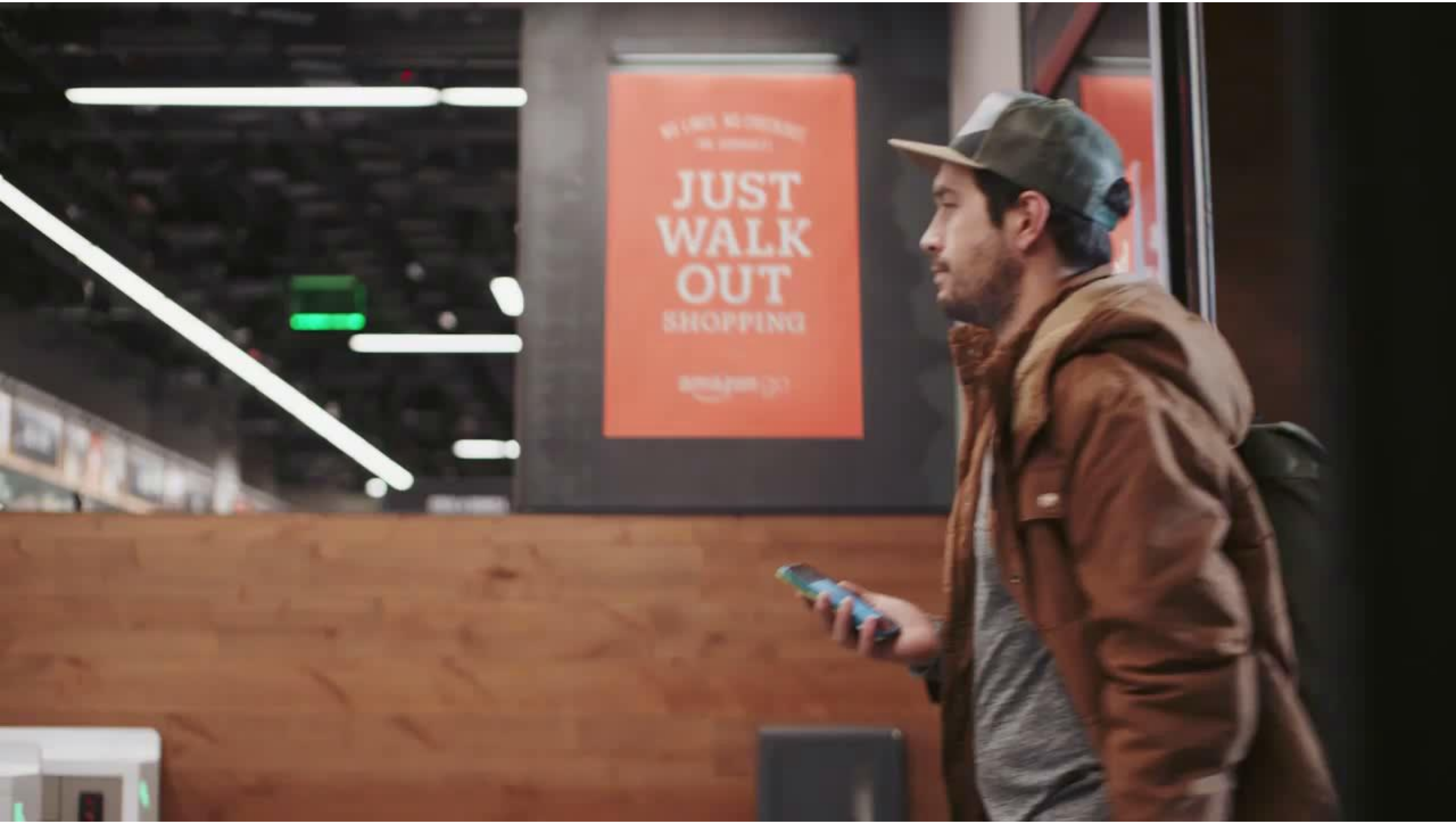
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

Source: http://hdl.handle.net/1721.1/6125

Related fields

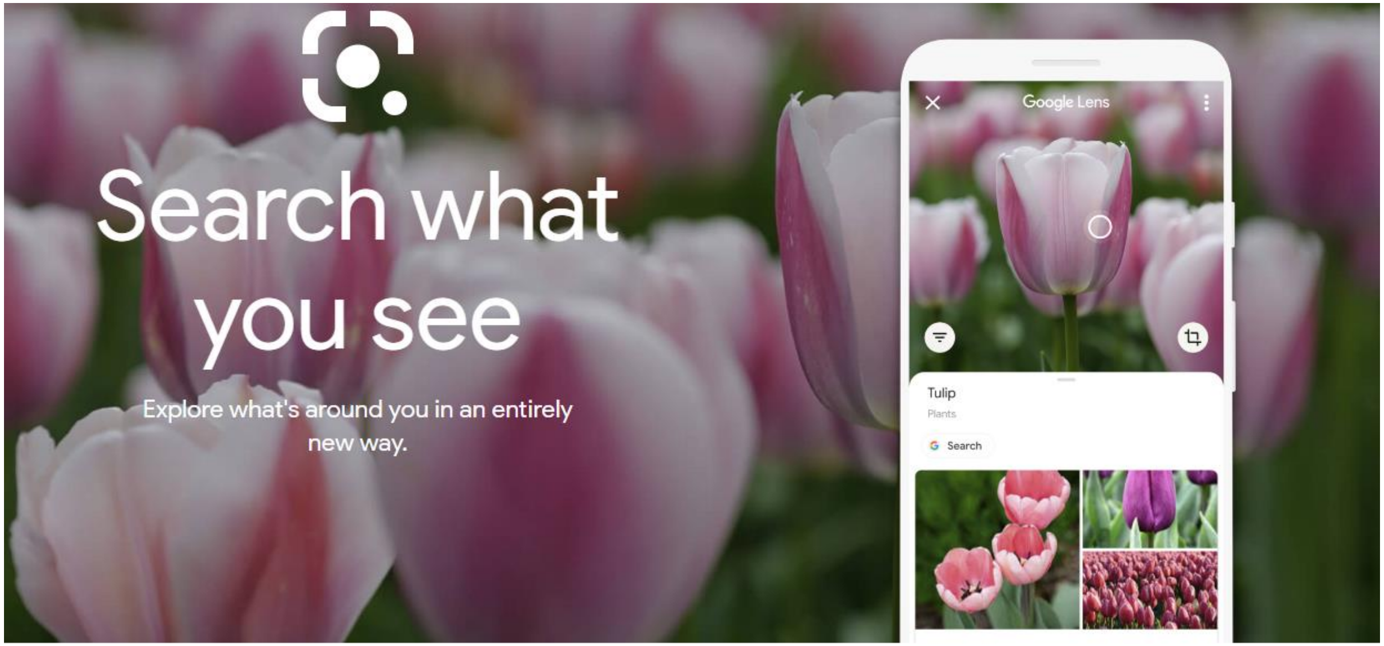


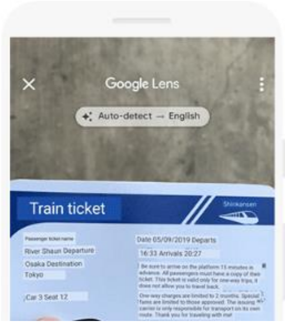
Application: Object Recognition



https://www.amazon.com/b?ie=UTF8&node=16008589011

Application: Object Recognition





Scan and
translate text

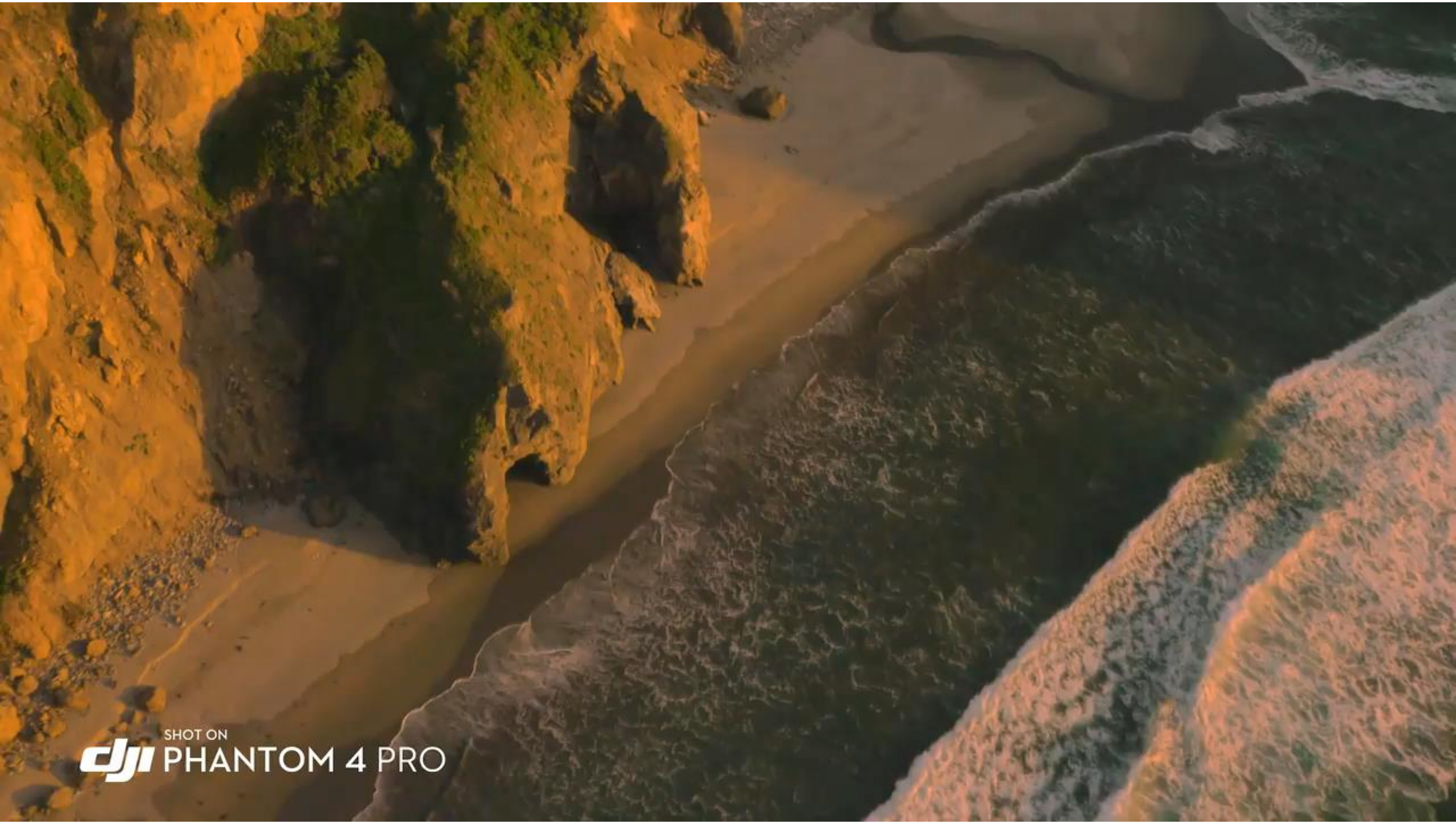
https://lens.google.com/

Application: Vision-based Biometrics

Application: Autonomous driving

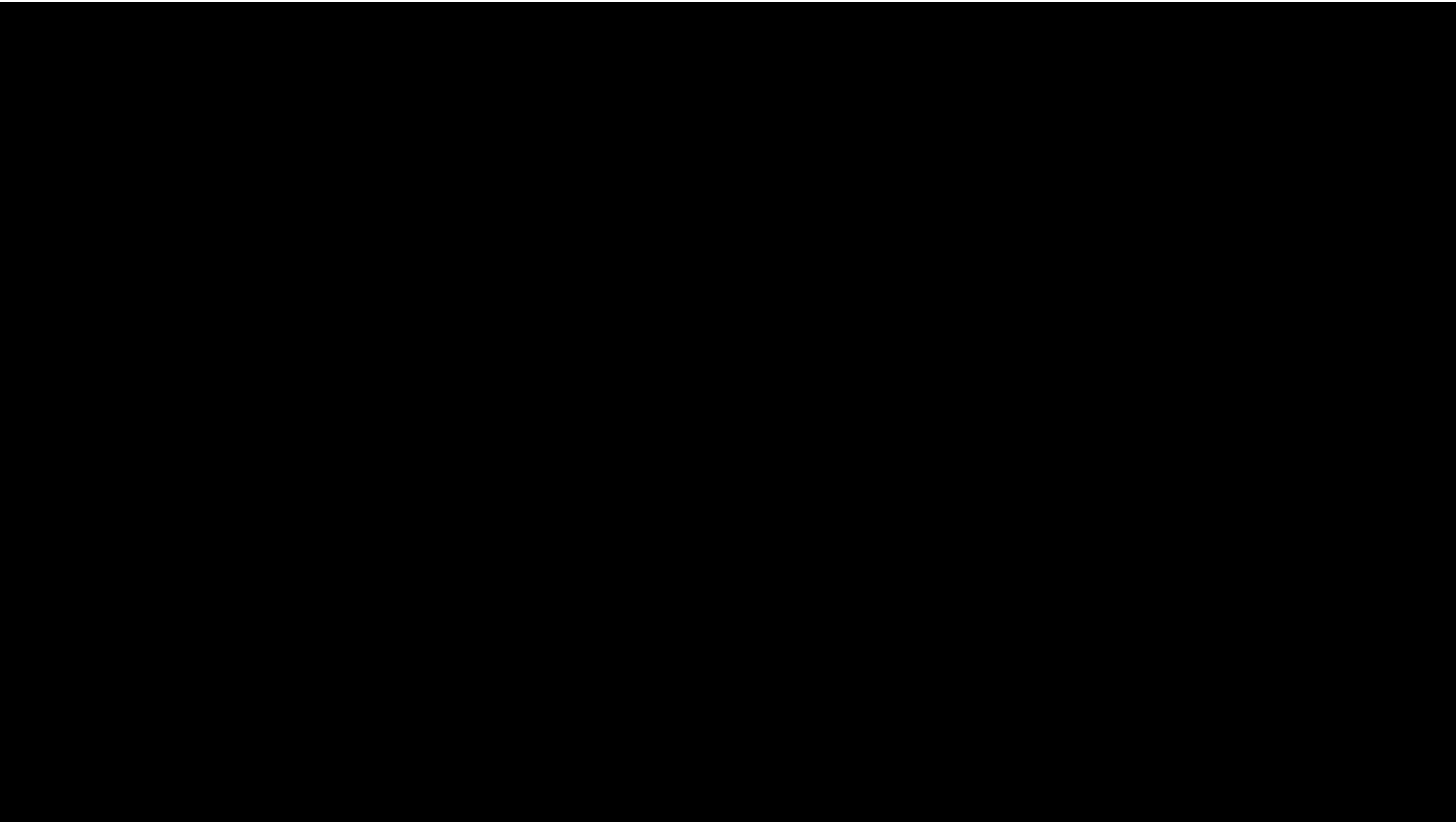
https://www.youtube.com/watch?v=B8R148hFxPw

Application: Drone Navigation



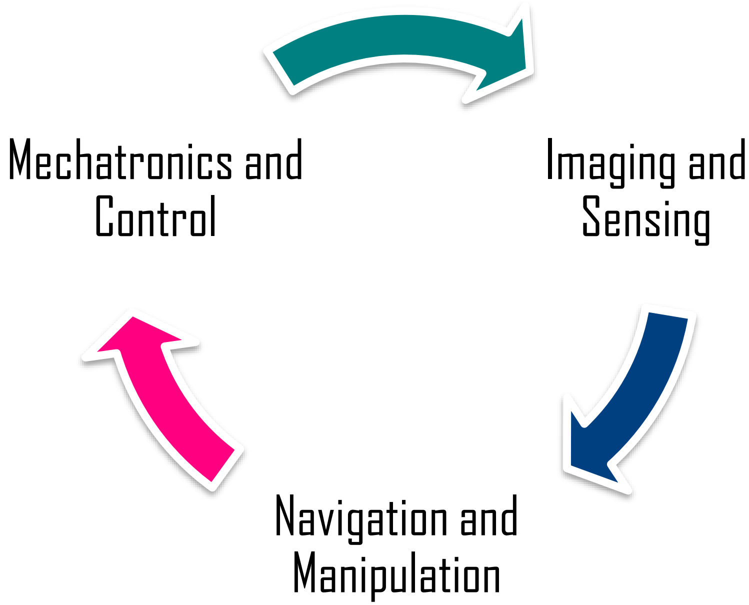
https://www.youtube.com/watch?v=ZVNO-fib6fg

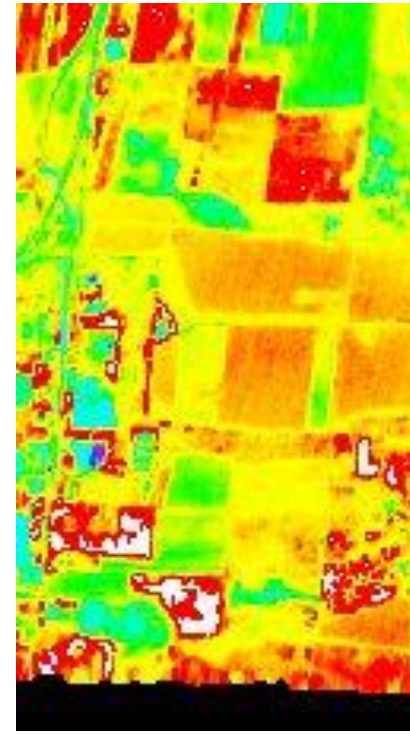
Application: Motion Analysis



https://www.youtube.com/watch?v=oyjNqksc-m8

Key Components of a Robotic Platform





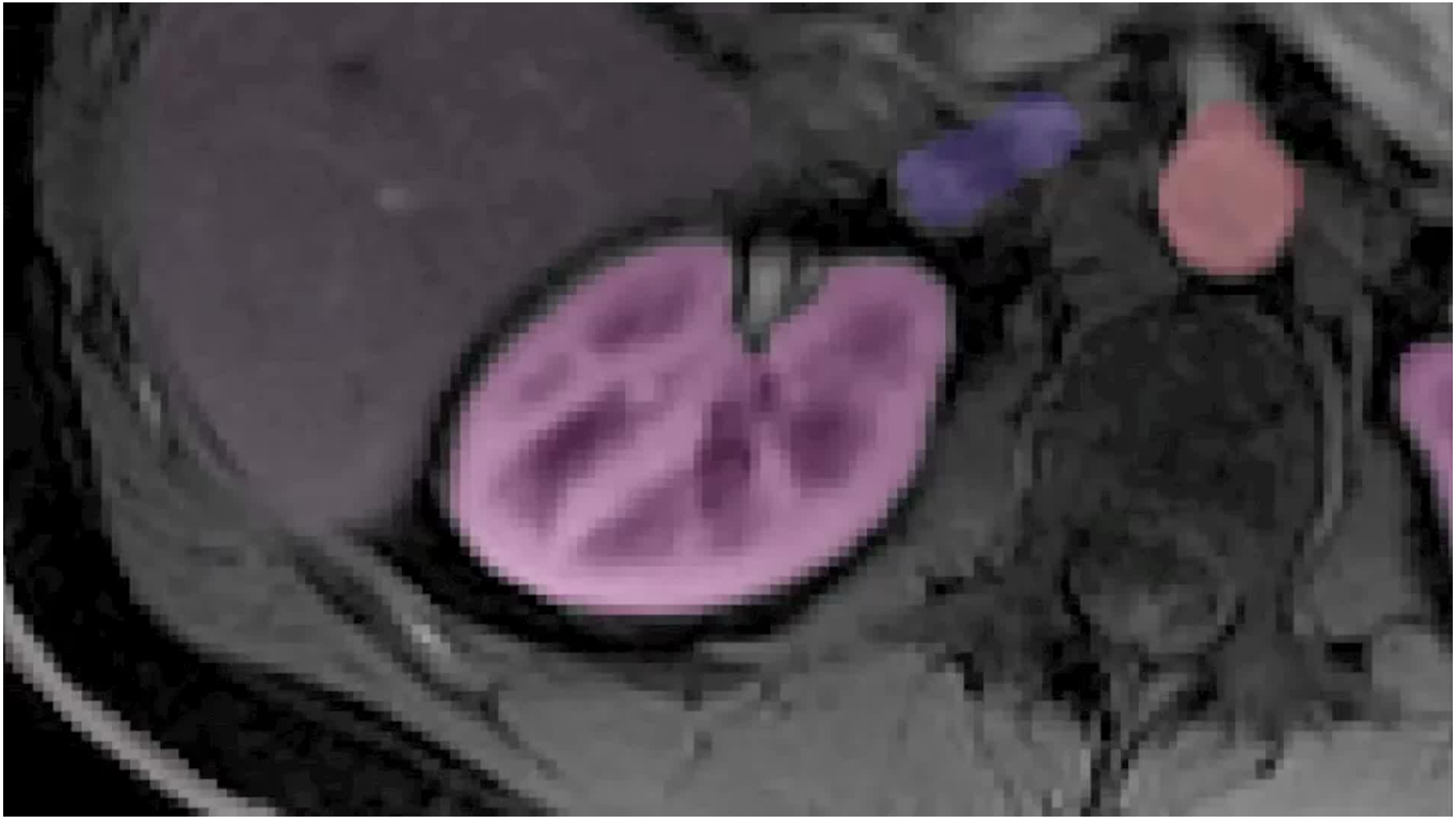
Unibots.com, NARO Japan, MIT CSAIL, and ZDNet

Application: Navigation Inside the Body

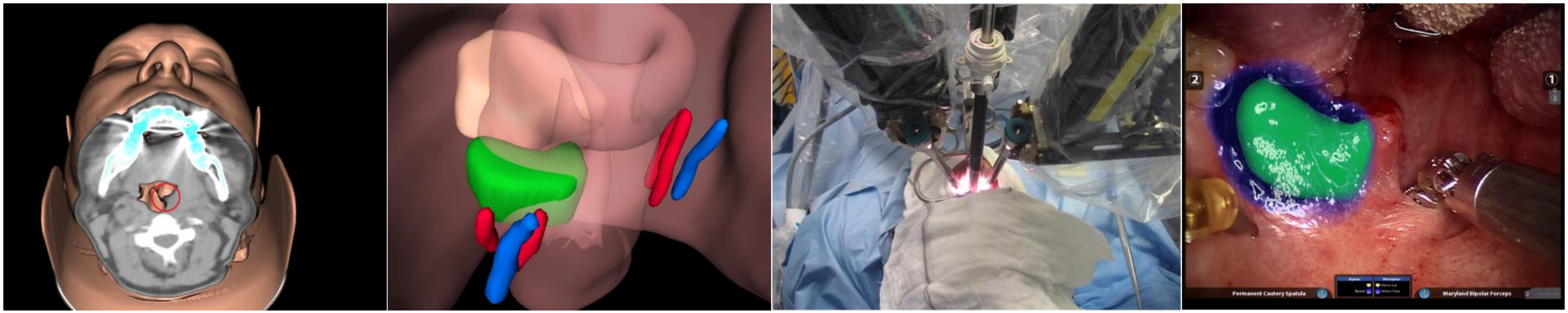
Image and Video Courtesy: Intuitive Surgical Inc

Precision and stability

Application: Navigation Inside the Body



Application: Navigation Inside the Body



- Segmentation
- Shape modelling

- Registration
- Image fusion

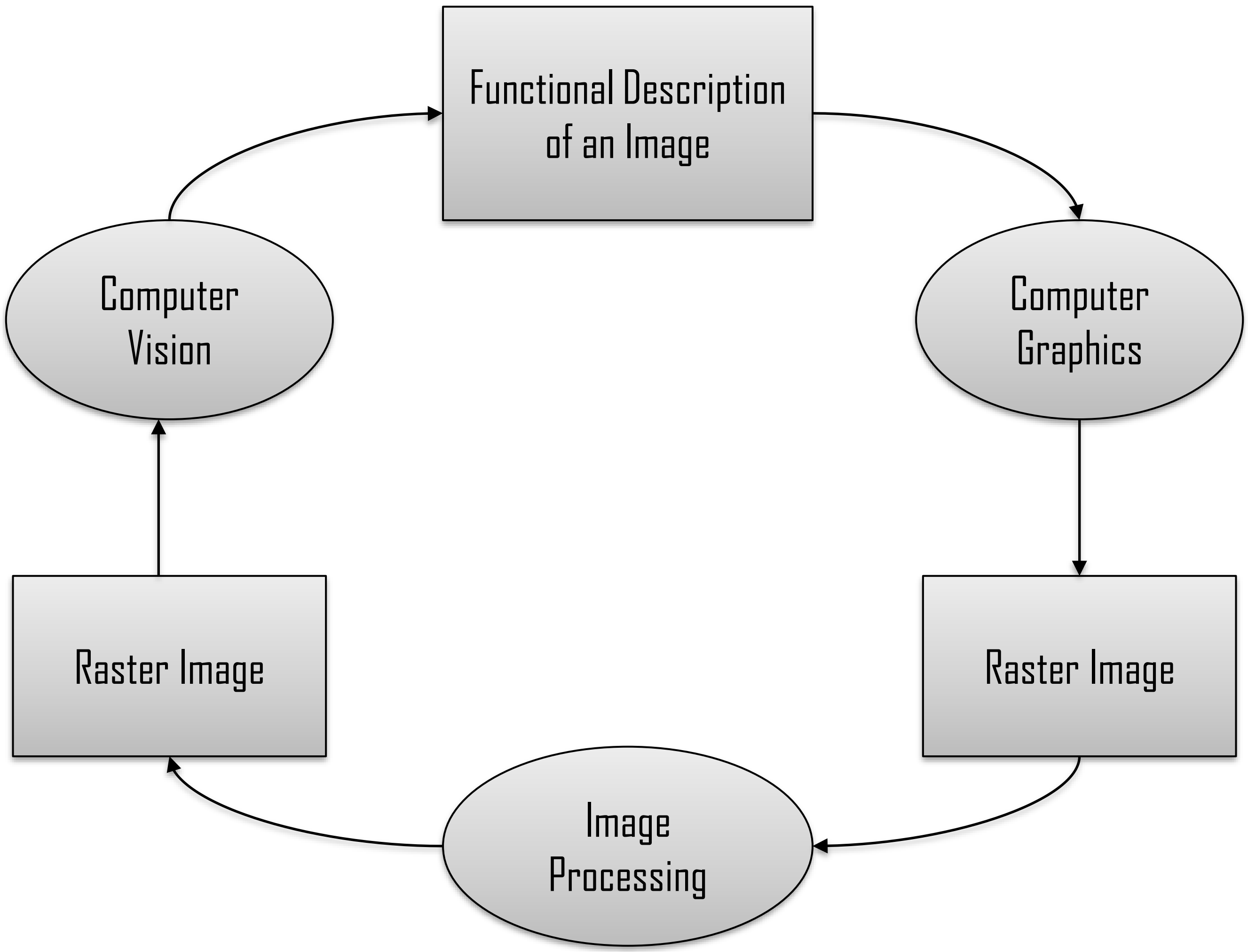
- Physical based modelling
- Simulation and planning

- Tracking
- Surface Reconstruction (shape from X)

- Localisation and mapping
- Intraoperative shape instantiation

- Image fusion and AR visualisation
- Navigation and guidance

Graphics, Image Processing and Vision

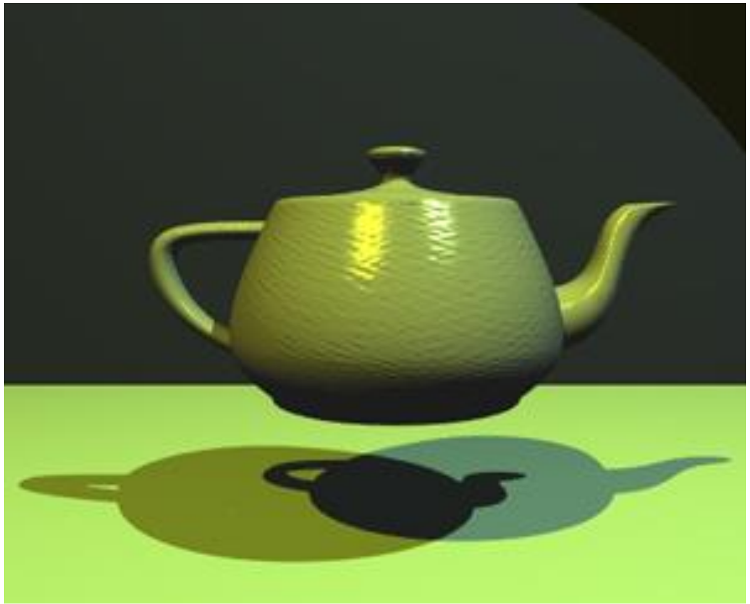


Graphics, Image Processing and Vision

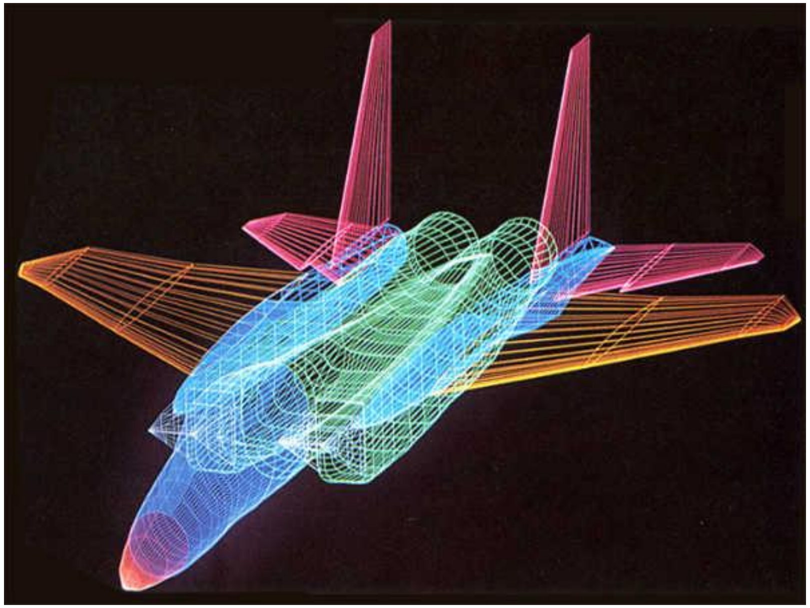
- **Graphics** – image or visual representation of an object, can be in either 2D or 3D
  - Typically in raster or vector format. Raster format uses a grid of pixels where vector format is made up of paths of scalable objects (e.g., lines, shapes). Vector-based graphics can be scaled to a larger size without losing image quality (e.g. EPS, Macromedia Freehand, Adobe Illustrator, SolidWorks);
- **Image Processing** – a form of signal processing for which the input is an array of pixels (image) and the output is also an image describing specific features of the data.
  - It is useful to differentiate commonly used, often interchangeable terms, all using an image as the input - Image Processing: output=image; Image Analysis: output=measurements, e.g., path of an object moving within the image sequence; Image Understanding: output=high-level description of the scene content, which is related to what we discuss in high-level vision.
- **Computer Vision** – from image(s) or image sequence(s), to derive high-dimensional description of the objects or scenes in either numerical or symbolic representations.

Graphics

- 3D representation of the object and its material properties (colour, texture, reflectance) are known
- Full control of lighting and other environmental factors
- Typical rendering methods include ray casting, ray tracing, radiosity
- Photorealistic and non-photorealistic rendering; need to consider perceptual factors to achieve realistic results



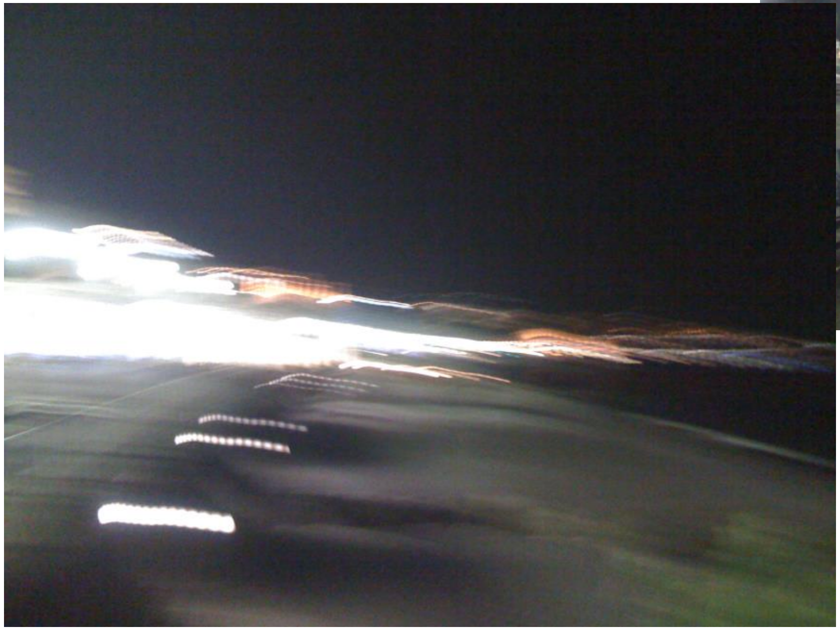
https://www.britannica.com/topic/computer-graphics



https://www.careerindia.com/online-courses/online-course-on-computer-graphics-university-of-california-019339.html

Image Processing

- Removal of defects such as scratches or other noise
- Improvement of contrast
- Removal of camera blur
- Removal of motion blur
- Enhancement of structure
- Enhancement of colour

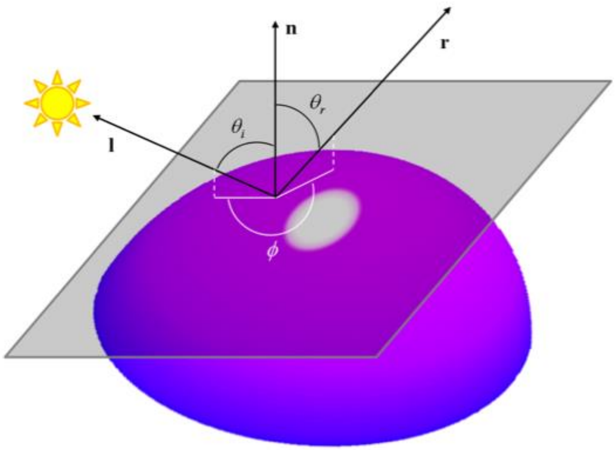


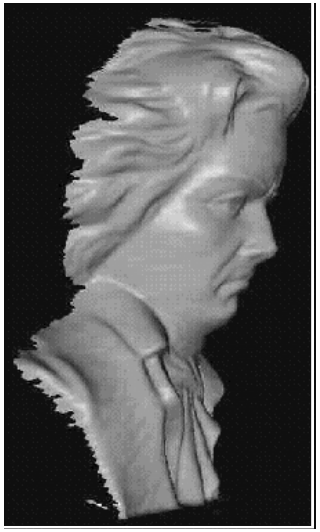
Computer Vision

- Determining the type of an object in the picture
- Assessing an object for quality
- Breaking a picture into different parts
- Constructing a 3D representation of an object
- Extracting a line representation of an object
- Reasoning about a scene to deduce hidden properties

Vision is the exact opposite process to graphics. In graphics, we start with a functional description and end up with a picture representation which is a set of pixels. In the case of graphics, we have complete information and so the problems are largely algorithmic. For vision, we do not have complete information and therefore we must use domain specific knowledge, assumptions or heuristics to achieve our goal.

Computer Vision: Examples





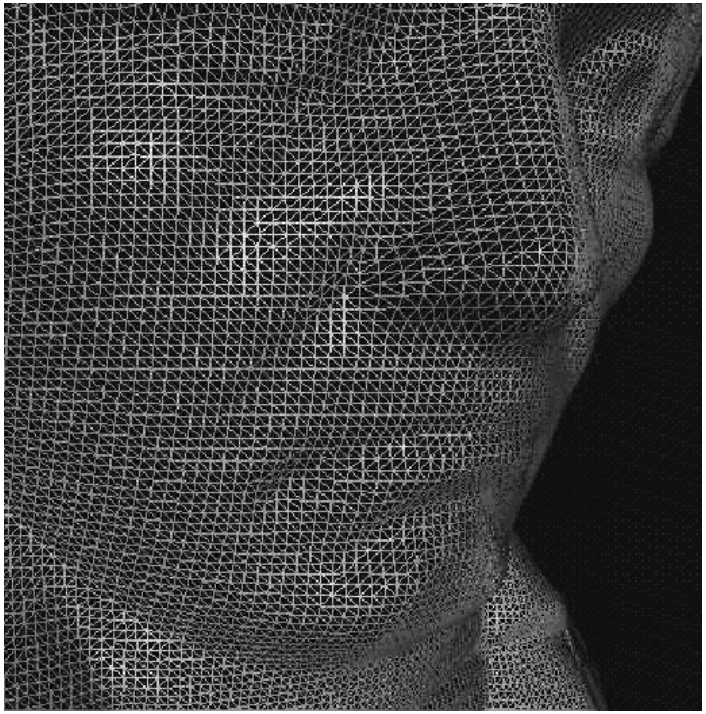
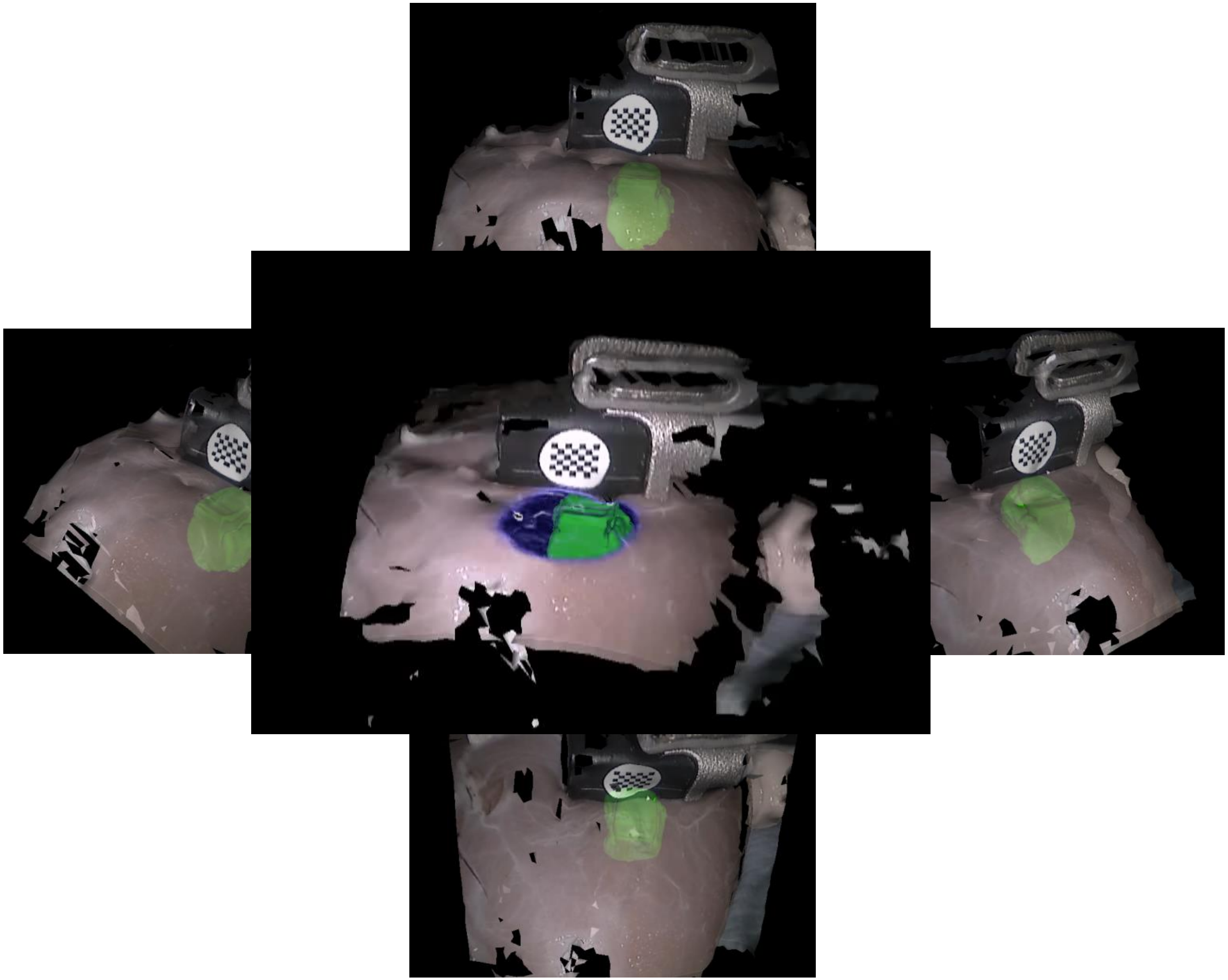


Image courtesy Karsten Schuens

Computer Vision: Examples



Some Questions For Vision System Design

- What information is sought from the image?
- How is it manifested in the image?
- What a-priori knowledge is required to recover the information?
- What is the nature of the computational process?
- How should the required information be represented?

Consider these problems:

Problem: Locate a house among trees
Straight lines are an intrinsic characteristic of houses, not trees

Problem: Separate land from sea in an aerial image
Uniform reflectivity is an intrinsic characteristic of sea, not land

Computational Models for Vision

- **Intrinsic Characteristics**
  - Need to establish a relationship between physical entities and intrinsic characteristics. If, for example we wish to distinguish and extract the position of a house in a scene containing trees, we can use the fact that straight lines are an intrinsic characteristic of a house, but not of trees. Hence we would choose to extract straight lines from the raw images. Conversely, if we wish to separate sea from land for an aerial photograph, we could choose the intrinsic characteristic that water is of uniform appearance, and apply a region based segmentation algorithm tuned to extract large uniform regions. Intrinsic characteristics must be matched to established techniques.

- **Resource Limitations**
  - Consideration of computational resources and the time required to process the image. The exact nature of the computational process will be task dependent, real-time requirements need to be taken into account in most vision systems. For example, if we wish to assess the condition of a motorway surface while driving over it at 60mph, each sampled image must be processed in a fixed time, which in turn places a requirement on the hardware.

Computational Models for Vision

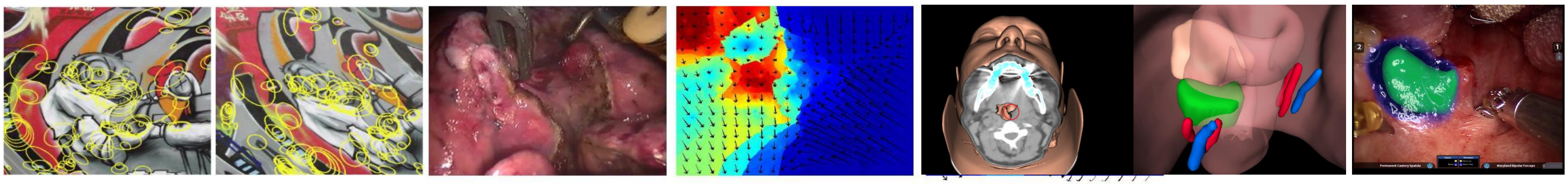
- **Prior Knowledge**
  - It is universally accepted that human vision depends on a vast amount of knowledge. To establish a relationship between pixel brightness and image properties, we will need to have some form of *scene model*, *illumination model* and *sensor model*. The scene model may include such information as the type of features we are trying to detect, or in more general cases make assumptions about properties such as smoothness or convexity. The illumination model will contain information about the position and characteristics of the light source and the surfaces reflectance properties. The sensor model will describe the position and optical performance of the cameras used, and the noise and distortion applied by the digitisation and storage media.

- **Knowledge Representation**
  - Representation is important for vision partly for the encoding of knowledge in a useful form, but also in the presentation of results in an understandable form. Typically, humans find it difficult to describe exactly the visual properties on which they make decisions.

Levels of Vision

Low Level Vision → Intermediate Level Vision → High Level Vision

Edges, corner, depth, optical flow | Contours, regions | Objects and hidden information



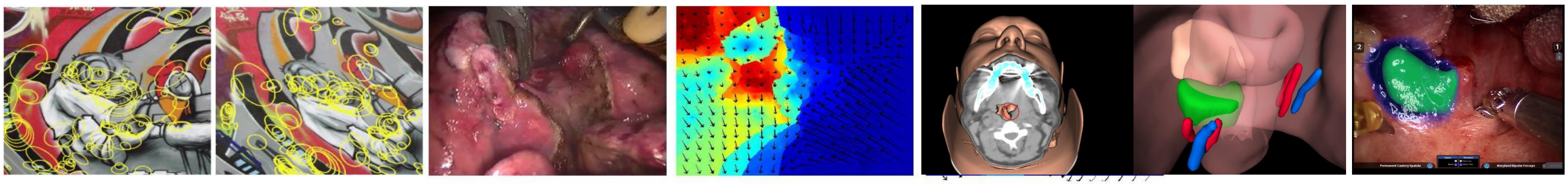
Low Level Processing

Operations carried out on the pixels in the image to extract properties such as the gradient (with respect to intensity) or depth (from the viewpoint) at each point in the image. We may for example be interested in extracting uniform regions, where the gradient of the pixels remains constant, or first order changes in gradient, which would correspond to straight lines, or second order changes which could be used to extract surface properties such as peaks, pits, ridges etc. Low level processing is invariably data driven, sometimes called **bottom up**. It is the area where modelling the visual cortex functioning is most appropriate.

Levels of Vision

Low Level Vision | Intermediate Level Vision | High Level Vision

Edges, corner, depth, optical flow | Contours, regions | Objects and hidden information



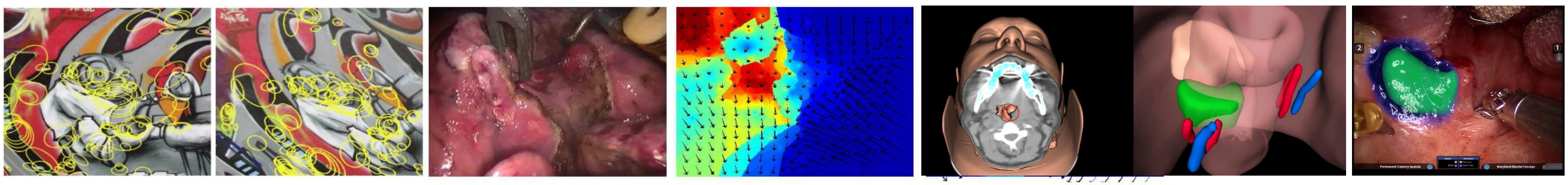
Intermediate Level Processing

The intermediate level of processing is fundamentally concerned with **grouping** entities together. The simplest case is when we group pixels into lines. We can then express the line in a functional form. Similarly, if the output of the low level processing is a depth map, we may further need to distinguish object boundaries, or other characteristics. Even in the simple case where we are trying to extract a single sphere, it is no easy process to go from a surface depth representation to a centre and radius representation. Since intermediate level processing is concerned with grouping, much of the recent work has concentrated on using perceptual grouping methods.

Levels of Vision

Low Level Vision | Intermediate Level Vision | High Level Vision

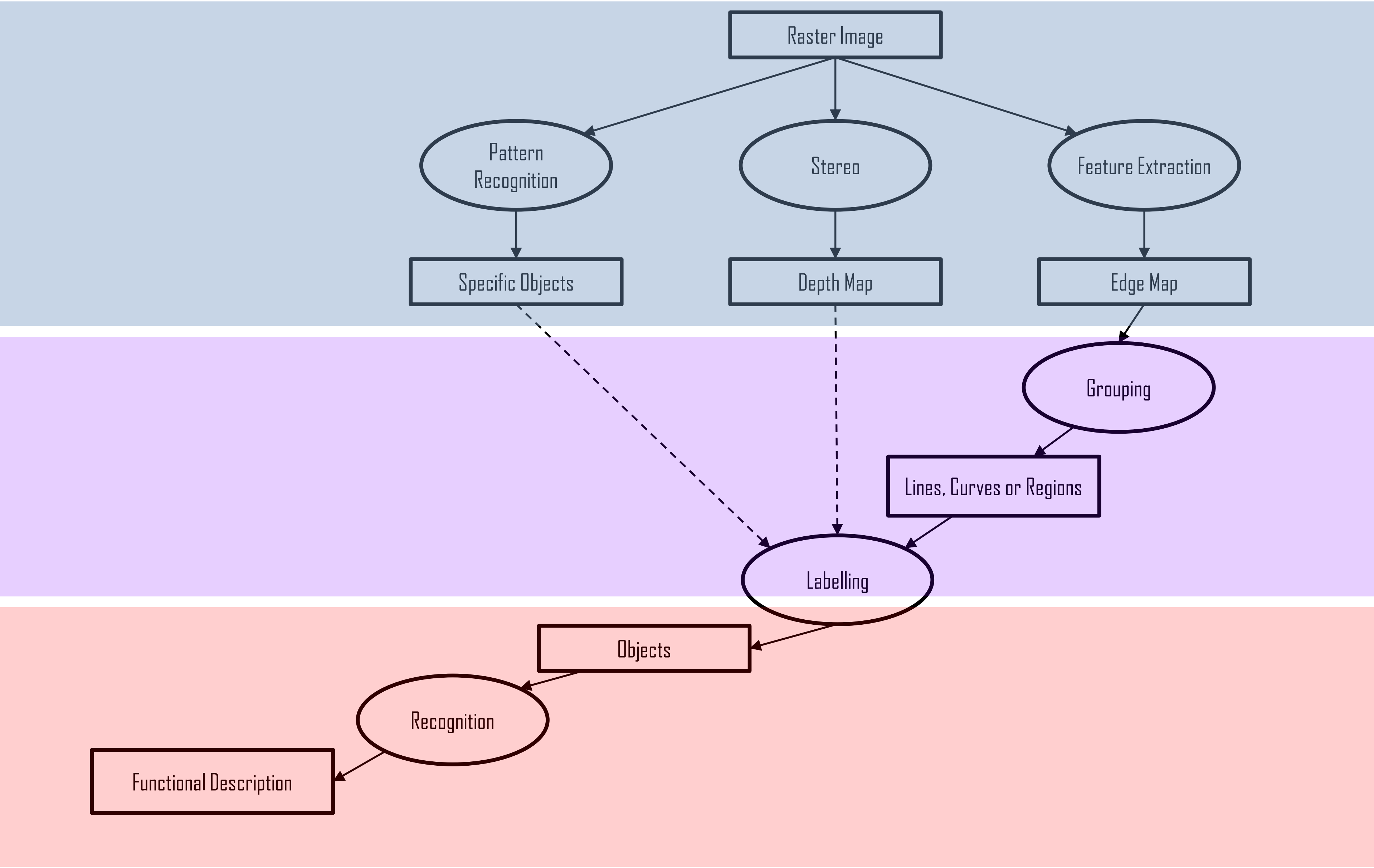
Edges, corner, depth, optical flow | Contours, regions | Objects and hidden information



High Level Processing

Interpretation of a scene goes beyond the tasks of line extraction and grouping. It further requires decisions to be made about types of boundaries, such as which are occluding, and what information is hidden from the user. Further grouping is essential at this stage since we may still need to be able to decide which lines group together to form an object. To do this, we need to further distinguish lines which are part of the objects structure, from those which are part of a surface texture, or caused by shadows. High level systems are therefore object oriented, and sometimes called **top down**. They almost always require some form of knowledge about the objects of the scene to be included.

Computer Vision Processes



Conclusions

- What is computer vision?
- Applications of Computer Vision
- Graphics, image processing and vision
- Computational models for vision
- Levels of vision