



45 sec

## Probability Theory

✓ Video: Probability Theory  
35 sec

## Algorithms

✓ Video: Algorithms  
38 sec

✓ Video: Using Algorithms  
1 min

✓ Practice Quiz:  
Algorithms  
1 question

Mathematics for  
Computer Vision  
Resources and  
Evaluation

✓ Reading: Resources  
(Optional): Mathematics  
for Computer Vision  
30 min

✓ Video: Aligning RGB  
channels  
2 min

🔗 Graded External Tool:  
MATLAB: Aligning RGB  
Channels  
1h

Computer Vision Basics -  
Key Takeaways

📖 Reading: Computer  
Vision Basics - Key  
Takeaways  
10 min

# Computer Vision Basics - Key Takeaways



## Computer Vision Basics – Key Takeaways

Below you will find a number of key points from this course. Defined terms are in bold.

### Week One: Overview

The **objective of Computer Vision** is to make computers see and interpret the world like humans or possibly even better than. However, recreating human vision isn't just a hard problem, it's a set of problems- each which relies on the other.

**Computer Vision** is concerned with the automatic extraction, analysis, and understanding of useful information from a single image or sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding.

**Dual goal of Computer Vision:** From the biological science point of view, Computer Vision aims to come up with computational models of the human visual system. From the engineering point of view, Computer Vision aims to build autonomous systems to perform some tasks the human visual system can perform and it even surpasses human capabilities in many cases.

Computer Vision systems contain these basic elements:

1. A power source
2. At least one camera
3. A processor
4. Control and communication cables or some kind of wireless interconnection mechanism
5. Configurable software
6. A display in order to monitor the system

Some of the applications of Computer Vision include: facial recognition in smartphone cameras, analysis of surroundings in self-driving cars, and factory robots that navigate around coworkers.

The field of Computer Vision heavily incorporates concepts from the areas of digital processing, neuroscience, computer graphics, solid-state physics, photogrammetry, and artificial intelligence.

The field of **Digital Image Processing** predominantly deals with image-to-image transformations. Typical image processing operations include image compression, image restoration, and image enhancement.

**Computer Graphics** studies the techniques that produce image data from 3D models, whereas computer vision works to produce 3D models from image data.

**Machine Vision** is the process of applying a range of technologies and methods to provide imaging-based automatic inspection, process control, and robot guidance in industrial applications.

**Photogrammetry** is the science of making measurements from photographs, especially for recovering the exact positions of surface points captured in the image.

The field of Computer Vision emerged in the 1950s, with research along three distinct lines: replicating the eye, replicating the visual cortex, and replicating the rest of the brain.

**Optical Character Recognition (OCR)** makes typed, handwritten, or printed text intelligible for computers.

**Convolutional Neural Networks (CNNs)** have been applied to identify faces, objects, and traffic signs, as well as

powering vision in robots and self-driving cars.

Computer Vision is outperforming humans on certain restricted real world tasks such as circuit board inspections and facial recognition under controlled conditions.

## Week Two: Color, Light, & Image Formation

A **point light source** originates at a single location in a 3 dimensional space, like a small light bulb, or potentially at infinity like the sun.

The key factors that affect the “color” of a pixel are:

1. The light sources
2. Object surface properties
3. Emittance and reflective spectrum
4. Relative position and orientation

The most basic camera model is a **pinhole camera model**. In this model, conceptually, all light passes through a vanishingly small pinhole placed at the origin and illuminates an image plane beneath it. When using a pinhole camera model, this geometric mapping from 3D to 2D is called a **perspective projection**.

Perspective projection makes parallel lines in the real world appear that they might be converging. The point of convergence is called a **vanishing point**.

The two main kinds of sensors used in still and video cameras today are *charge-coupled devices (CCD)* and *complementary metal oxide on silicon (CMOS)*.

The main factors affecting the performance of a digital image sensor are:

1. Shutter speed
2. Sampling pitch
3. Chip size
4. Analog gain
5. Sensor noise
6. The quality of the analog-to-digital converter

The retina of the human eye has *two types of photo-receptors*: Rods and Cones. **Rods** are sensitive to light intensity, while **cones** are color sensitive.

When incoming light hits an imaging sensor, light from different parts of the spectrum is integrated into the discrete red, green, and blue (RGB) color values that we see in a digital image.

## Week Three: Low, Mid, & High-Level Vision

Three levels in David Marr’s paradigm:

1. *Computational theory* – describes what the device is supposed to do
2. *Representation and algorithm* – addresses precisely how the computation may be carried out
3. *Implementation* – includes physical realization of the algorithm, programs, and hardware

A **bottom-up** reasoning approach that mimics what is found in the brain is most promising for research in Computer Vision. A computer can apply a series of transformations to an image and discover the edges and the objects they imply. The process amounts to the computer trying to match the shapes it sees with shapes it has been trained to recognize.

The **paradigm of 3R’s** requires us to study the interaction among the processes of *recognition*, *reconstruction*, and *reorganization* and work towards the goal of a unified framework for computer vision.

Computer Vision concepts can be broadly categorized as low, mid and high-level vision techniques:

- **Low-level vision** constitutes of image processing techniques, feature detection, image matching, and early segmentation.
  - Research on low-level vision is concentrated in discovering what information about the world can be initially extracted from the image. The low-level image processing techniques involve extracting fundamental image primitives like edges and corners and performing filtering, and morphology etc.
- **Mid-level vision** is where things start to come together attributing meaning.

- o The two major aspects in mid-level vision are inferring the scene geometry and inferring the camera and object motion (answering the question of “how does the object move?”). Some fundamental concepts of geometric vision include: multi-view geometry, stereo, and structure from motion (SfM), all of which infer 3D scene information from 2D images.

- **High-level vision** tasks are the algorithms which make sense of the visual content and make computer vision live up to the capabilities of human vision.

- o The objective of High-level vision is to infer the semantics, for example, object recognition and scene understanding. A challenging question for many decades has been “how do you recognize 3D objects from different view directions?” There have been two approaches for recognition: *model-based recognition* and *learning-based recognition*.

**Image segmentation** is the process of identifying which areas in the image belong to the object.

## Week Four: Mathematics for Computer Vision

Computer Vision is used to solve vital problems in a vast array of fields including medical imaging, surveillance, face and object detection and identification. The techniques that *linear algebra* provides for solving complicated mathematical models are essential to solve problems in each of these fields.

- **Singular value decomposition (SVD)** is the most common and useful linear algebra technique in Computer Vision because it helps to achieve the goal of Computer Vision, which is to explain the three dimensional world through two dimensional pictures.

*Calculus* has two major branches:

1. **Differential calculus** - concerning instantaneous rates of change and slopes of curves
  2. **Integral calculus** - concerned with the theory and applications of integrals. It deals with total size or value, such as lengths, areas, and volumes.
- Computer Vision uses derivatives, integrals and partial differential equations extensively in several low and mid-level vision tasks.

**Artificial intelligence** deals with making decisions in the real world, often in the presence of great uncertainty. Therefore, we can conjecture that the visual world is uncertain and should be described through the language of *probabilities*.

Computer Vision benefits from Computer Science algorithms and numerical methods for mathematical optimizations.



Mark as completed

