

Introduction to Computer Vision

Introduction to Computer Vision

Key concepts, terminology, and foundational techniques

1. What Is Computer Vision?

Aspect	Description
Definition	A field of artificial intelligence that enables computers to interpret and act upon visual data (images, video, depth maps).
Goal	Replicate or surpass human visual perception for tasks such as recognition, detection, segmentation, and scene understanding.
Typical Applications	• Facial recognition • Autonomous driving • Medical imaging analysis • Retail (product counting, checkout) • Augmented/virtual reality • Robotics and industrial inspection

2. Core Pipeline of a Vision System

1. **Image Acquisition** – Capture raw pixel data via cameras, scanners, LiDAR, or other sensors.
 2. **Pre-processing** – Noise reduction, normalization, resizing, color space conversion.
 3. **Feature Extraction / Representation** – Transform raw pixels into a more informative form (edges, textures, keypoints, deep features).
 4. **Modeling / Inference** – Apply algorithms (classical or deep learning) to solve a specific task.
 5. **Post-processing** – Refine output (e.g., non-maximum suppression, morphological operations).
 6. **Decision / Action** – Use the result for downstream processes (alert, control, UI update).
-

3. Classical (Pre-Deep-Learning) Techniques

Introduction to Computer Vision

Technique	What It Does	Typical Use?Cases
Edge Detection (Sobel, Canny)	Highlights abrupt intensity changes ? outlines of objects.	Shape analysis, lane detection.
Corner/Keypoint Detectors (Harris, Shi-Tomasi, FAST)	Finds distinctive points that are repeatable across views.	Image stitching, SLAM.
Descriptors (SIFT, SURF, ORB)	Encodes local patch appearance into a vector.	Feature matching, object retrieval.
Histogram of Oriented Gradients (HOG)	Captures gradient orientation distribution in spatial cells.	Pedestrian detection.
Scale-Space & Blob Detection (Laplacian of Gaussian, DoG)	Detects structures at multiple scales.	Multi-scale object detection.
Morphological Operations (erosion, dilation)	Alters binary shapes for noise removal or gap filling.	Document binarization, object separation.
Color Space Transformations (HSV, Lab)	Separates chromaticity from intensity.	Skin detection, illumination-robust segmentation.

4. Deep Learning Revolution

4.1 Convolutional Neural Networks (CNNs)

Component	Role
Convolutional layers	Learn spatial filters (edges ? textures ? object parts).
Pooling (max/average)	Down-sample, provide translation invariance.
Activation (ReLU, LeakyReLU)	Introduce non-linearity.
Fully-connected / Global-average pooling	Aggregate features for classification or regression.
Batch Normalization	Stabilize training, accelerate convergence.

4.2 Popular Architectures

Architecture	Notable Feature	Typical Tasks
LeNet-5	Early small-scale CNN (handwritten digit recognition).	Digit classification.

Introduction to Computer Vision

Architecture	Notable Feature	Typical Tasks
AlexNet	Deep (8 layers) + ReLU + dropout, won ImageNet 2012.	General image classification.
VGG	Very deep (16/	