UNIT I

Computer Vision

ECSCI24402

Syllabus

Unit	Content	Hrs.
1	Overview of computer vision and its applications: Computer Vision and Computer Graphics: What is Computer Vision – Low level, Mid-level, High-level, Overview of Diverse Computer Vision Applications: Document Image Analysis, Biometrics, Object Recognition, Tracking, Medical Image Analysis, Content-Based Image Retrieval, Video Data Processing, Multimedia, Virtual Reality and Augmented Reality.	5

Topics to cover

- Overview of computer vision and its applications:
- Computer Vision and Computer Graphics:
  - What is Computer Vision – Low level, Mid-level, High-level,
- Overview of Diverse Computer Vision Applications:
  - Document Image Analysis, Biometrics, Object Recognition, Tracking, Medical Image Analysis, Content-Based Image Retrieval, Video Data Processing, Multimedia, Virtual Reality and Augmented Reality.

What is Computer Vision?

- Definition: Computer Vision (CV) is an interdisciplinary field of Artificial Intelligence (AI) focused on enabling computers and systems to extract meaningful information from images, videos, and other visual inputs.
- Analogy to Human Vision:

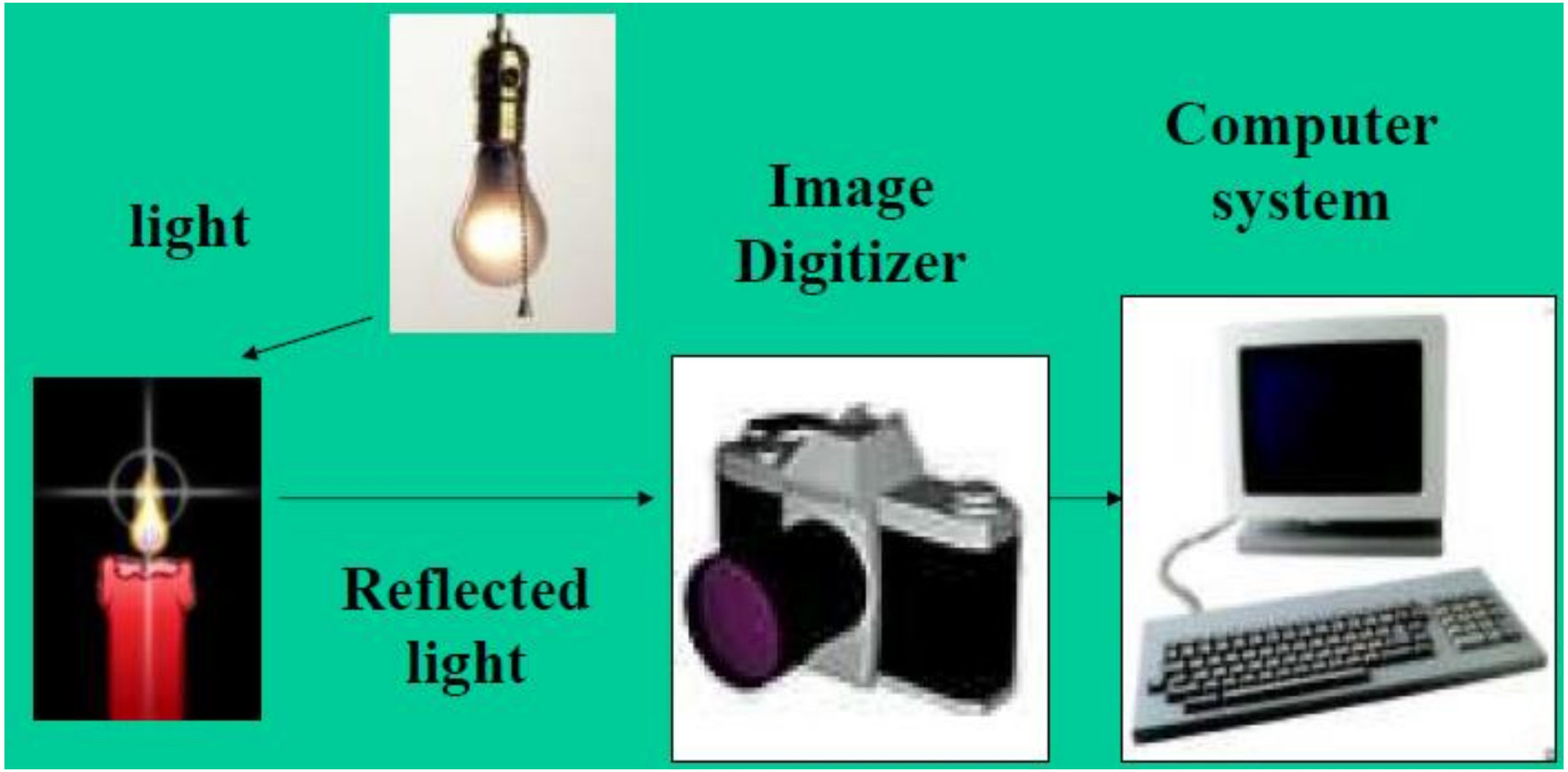
  Human vision: Eyes + brain (retina, optic nerve, cortex).

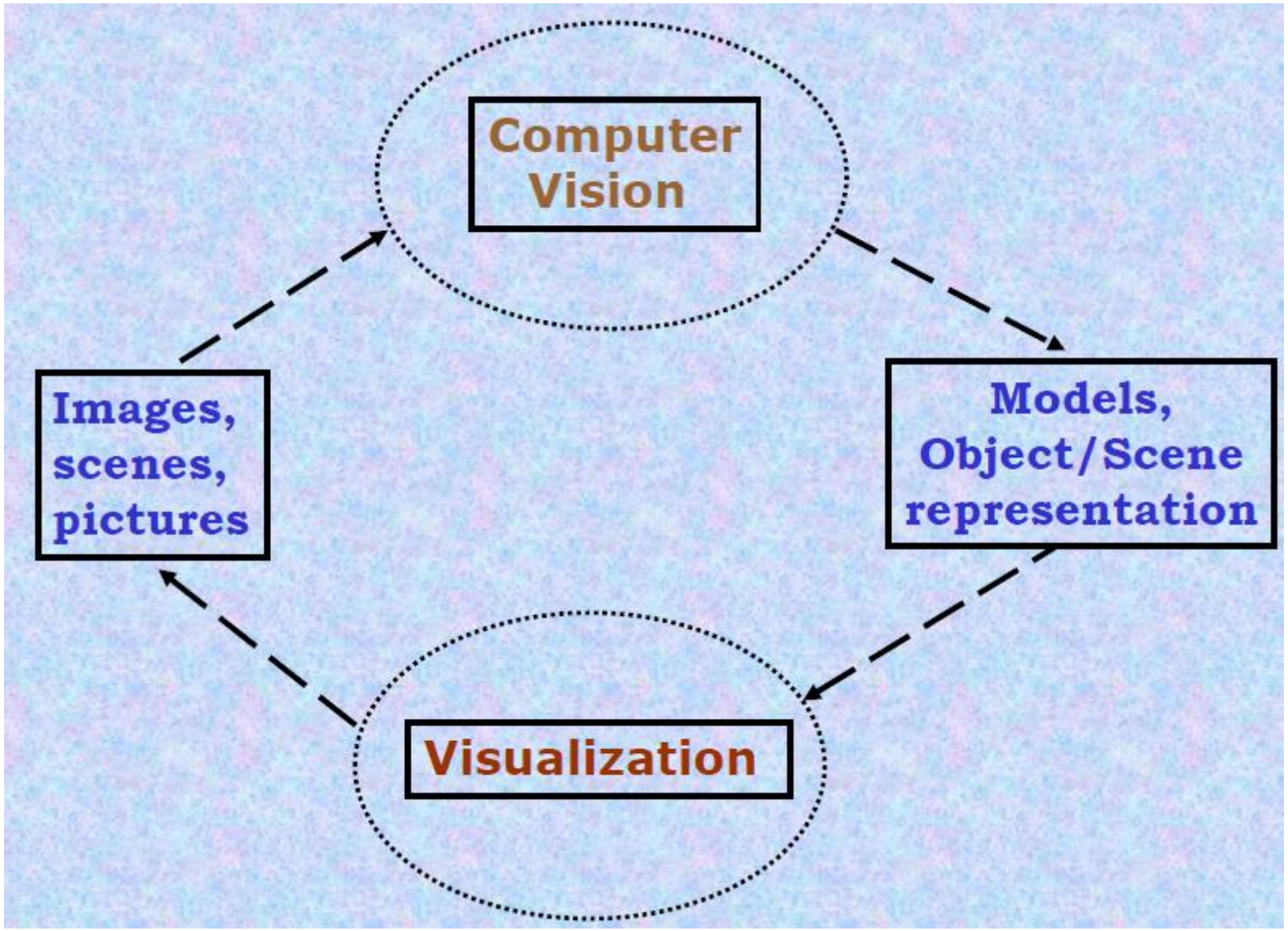
  Computer vision: Cameras + algorithms + data.
- Objective: Mimic human perception to recognize, interpret, and act on visual inputs.

Machine Vision

Human Vision

Computer Vision System




Computer Vision
Models, Object/Scene representation
Visualization
Images, scenes, pictures

Why is Computer Vision Important?

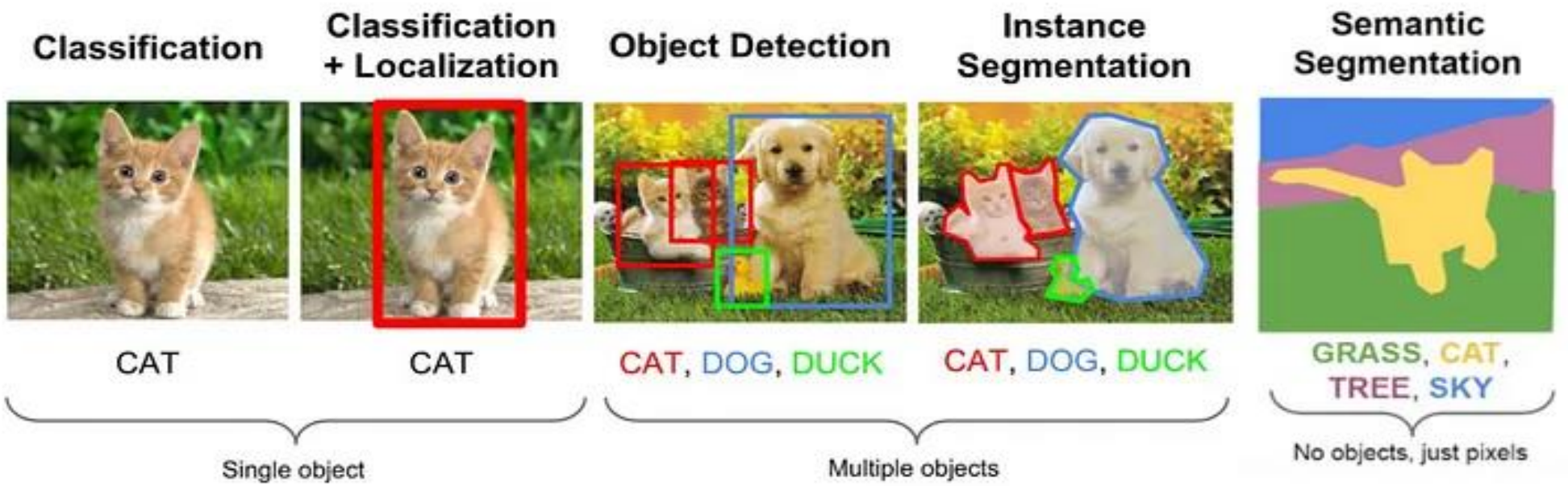
- Automation & Efficiency: Enhances decision-making and process automation in real-time environments.

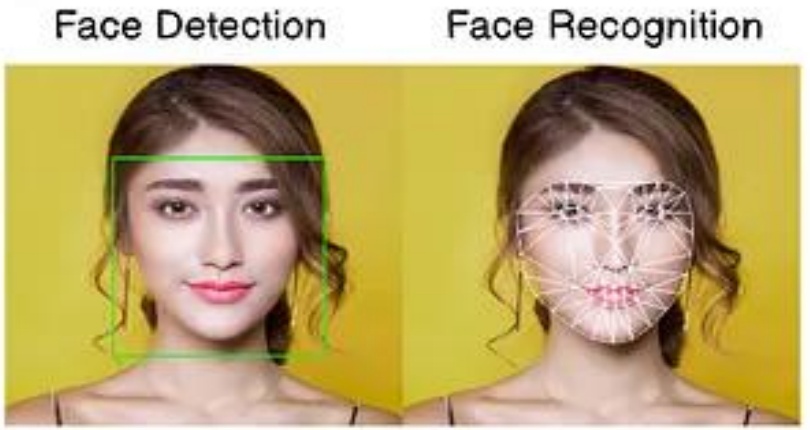
- Scalability: Can deploy “digital eyes” at any scale—drones, satellites, surveillance systems, etc.

Wide Applications

- **Medical Imaging**: MRI, CT, X-ray analysis for disease detection.
- **Autonomous Vehicles**: Lane detection, obstacle avoidance, pedestrian tracking.
- **Retail & Fashion**: Visual search, virtual try-on systems, Amazon StyleSnap.
- **Security & Surveillance**: Intrusion detection, facial recognition.
- **Agriculture**: Crop health monitoring using drones.
- **Industry 4.0**: Automated defect detection in manufacturing.
- **Self-driving cars** (Tesla Autopilot).
- **Social media filters** (Instagram, Snapchat).

Common Computer Vision Tasks

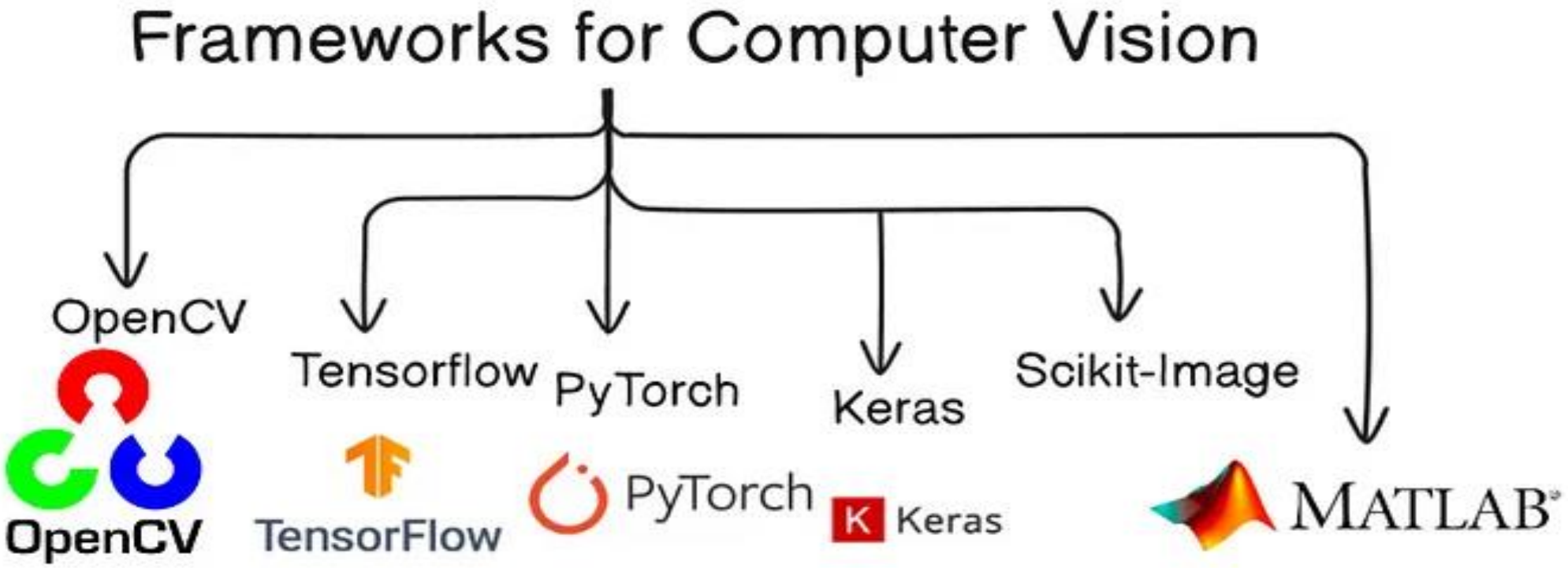




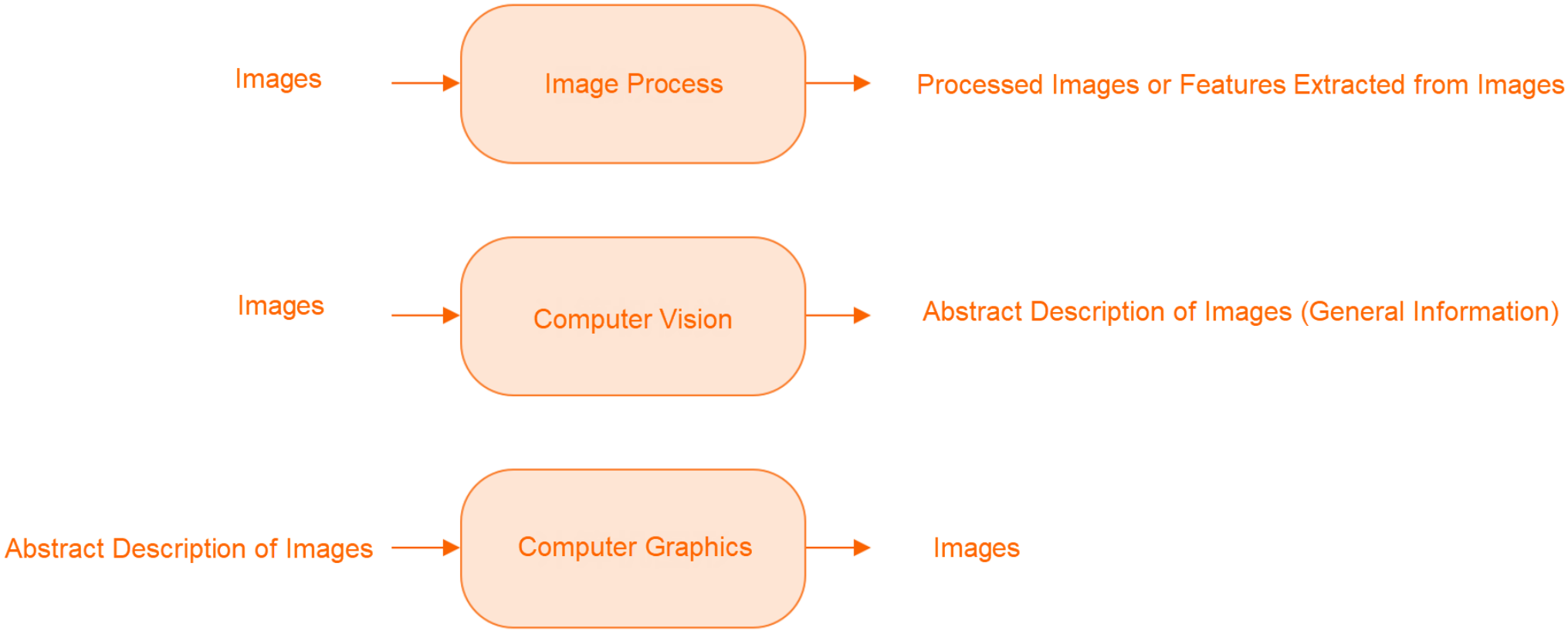
Common Computer Vision Tasks

- Image Classification: Assign a label to an entire image.
- Object Detection: Locate and classify multiple objects within an image.
- Semantic Segmentation: Assign a class label to each pixel (e.g., background vs object).
- Instance Segmentation: Distinguish between different instances of the same object class.
- Pose Estimation: Detect body keypoints (skeleton tracking).
- OCR (Text Recognition): Extract printed/handwritten text from images.
- Image Captioning: Generate textual description of an image.

Frameworks for Computer Vision



Computer Vision and Computer Graphics



The Three Stages of Computer Vision

- low-level (image processing)

  image → image

- mid-level (feature extraction)

  image → features

- high-level (the intelligent part)

  features → analysis

Computer Vision Hierarchy

Low-Level Vision

- **Goal**: Process raw image data to extract **basic visual features**.
- **Tasks**:
  - Edge detection
  - Corner detection
  - Optical flow estimation
- **Output**: Primitive features such as edges, motion vectors, or gradients.
- **Use**: Acts as the foundation for higher-level interpretation.

Computer Vision Hierarchy

Mid-Level Vision

- **Goal**: Use features from low-level vision to perform **structured analysis**.
- **Tasks**:
  - Object recognition
  - Motion analysis
  - 3D reconstruction
- **Output**: Geometrical or grouped information like object boundaries, shapes, or trajectories.
- **Use**: Critical in tasks like tracking, detection, and environment modeling.

Computer Vision Hierarchy

High-Level Vision

- **Goal**: **Interpret and reason** about the visual content semantically.
- **Tasks**:
  - Scene understanding
  - Activity recognition
  - Behavior/intention prediction
- **Characteristics**:
  - Directs mid and low-level vision operations dynamically.
  - Adds **semantic context** to the observed scene.
- **Use**: Surveillance, autonomous systems, cognitive robotics.

Low-Level

original image

Canny
edge
operator

edge image

Mid-Level (Lines and Curves)

edge image

ORT
line &
circle
extraction

data
structure

circular arcs and line segments

Low- to High-Level

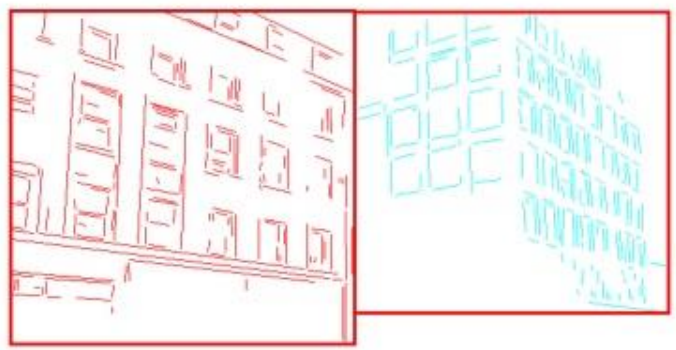
low-level

edge image

mid-level

consistent line clusters

high-level



Building Recognition

Diverse Computer Vision Applications

Document Image Analysis

- Extracts text and structure from scanned documents using OCR, layout detection.
- Used in digitization, archiving, and automation.

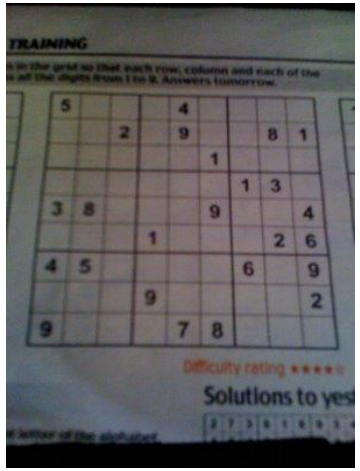
Optical character recognition (OCR)

- If you have a scanner, it probably came with OCR software

Digit recognition, AT&T labs (1990's)
http://yann.lecun.com/exdb/lenet/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Automatic check processing

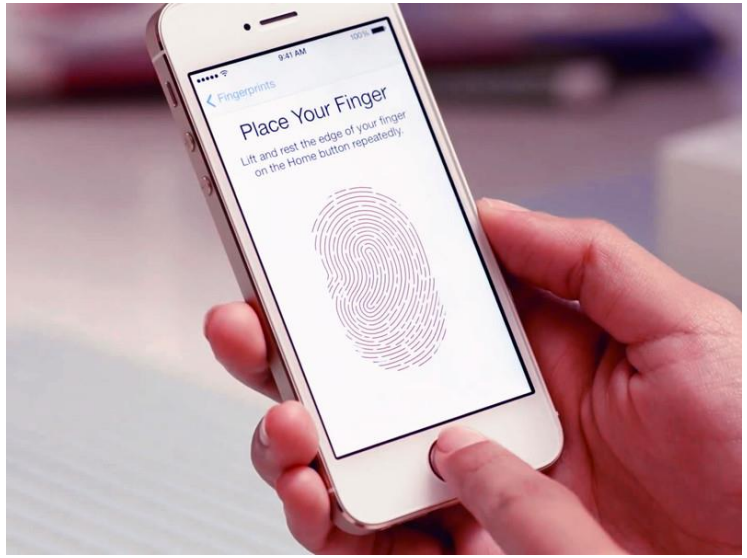


Sudoku grabber
http://sudokugrab.blogspot.com/

Biometrics

- Uses facial, iris, fingerprint or gait recognition for identification and verification in secure systems.

Login without a password

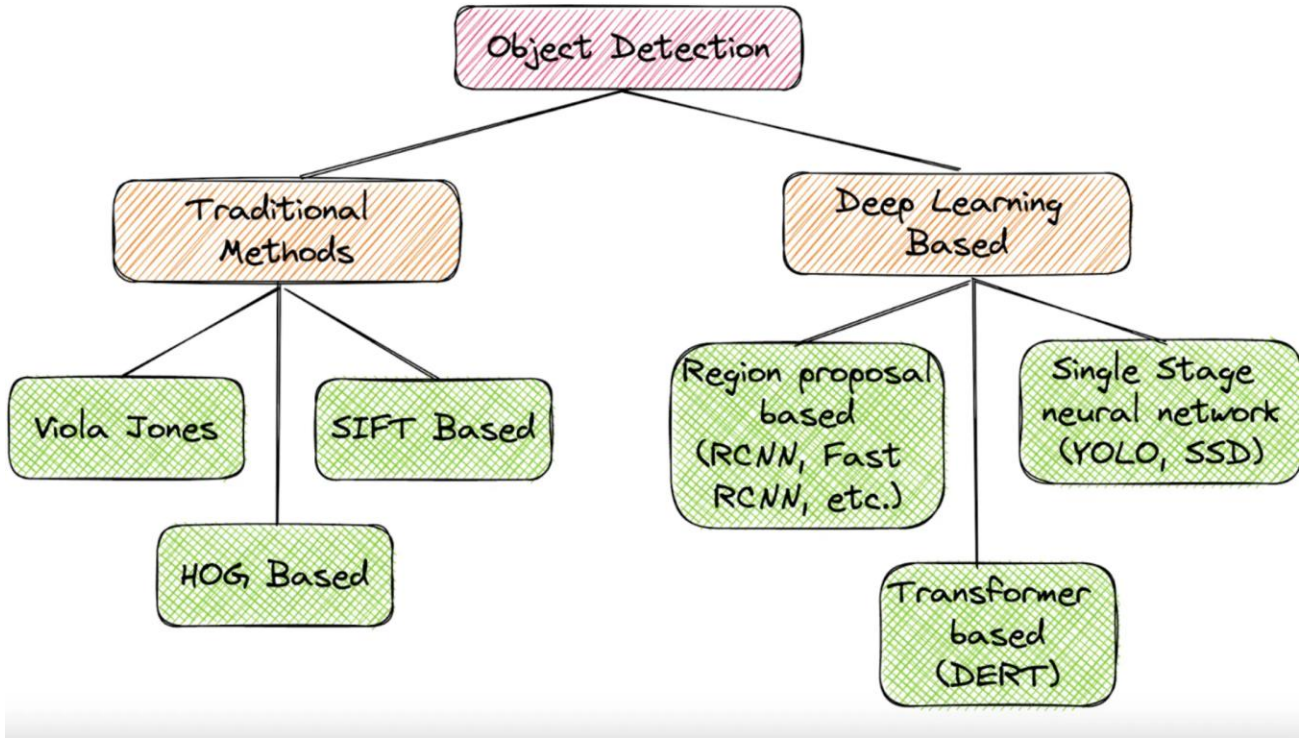


Fingerprint scanners on many new smartphones and other devices

Face unlock on Apple iPhone X
See also http://www.sensiblevision.com/

Object Recognition

- Classifies and locates objects in images using deep learning models.
- Applied in robotics, retail, and self-driving vehicles.




SELF DRIVING
MODE
40

Tracking

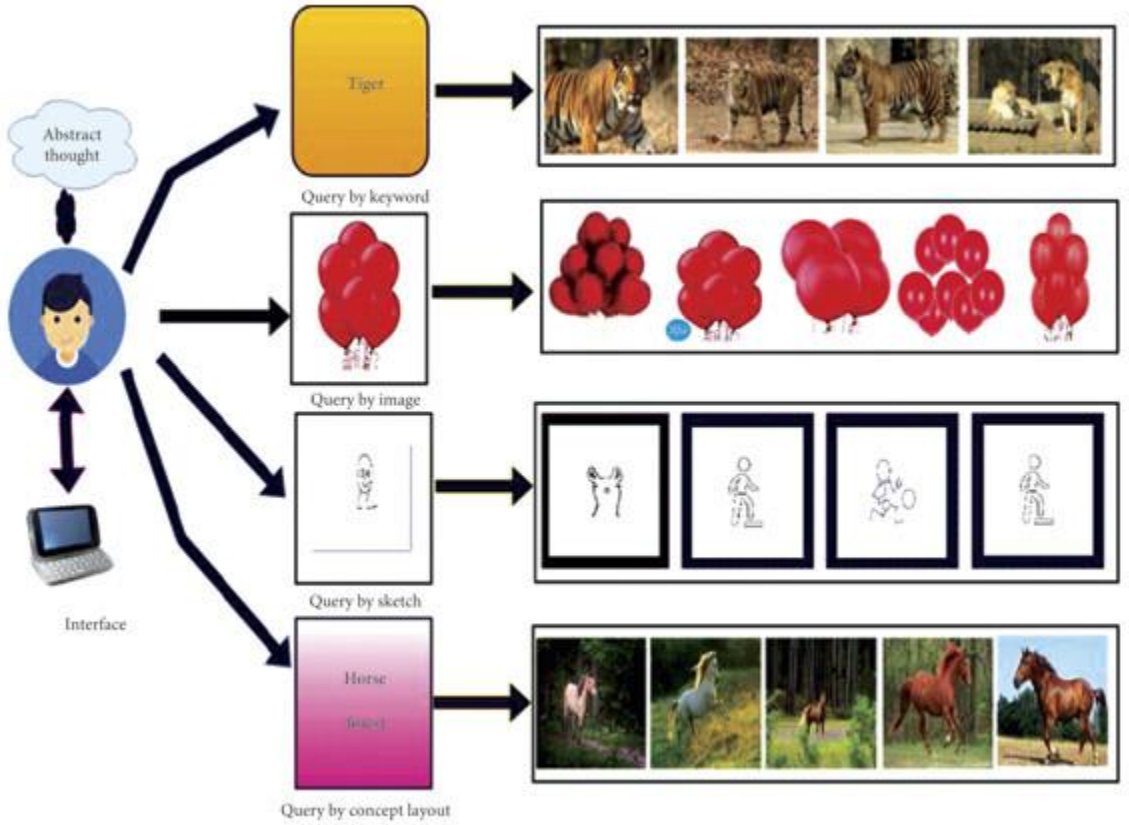
- Monitors moving objects across video frames. Used in surveillance, sports analytics, and UAV navigation.

Medical Image Analysis

- Analyzes MRI, CT, and X-rays for disease detection and surgical planning. Enhances diagnostic accuracy.

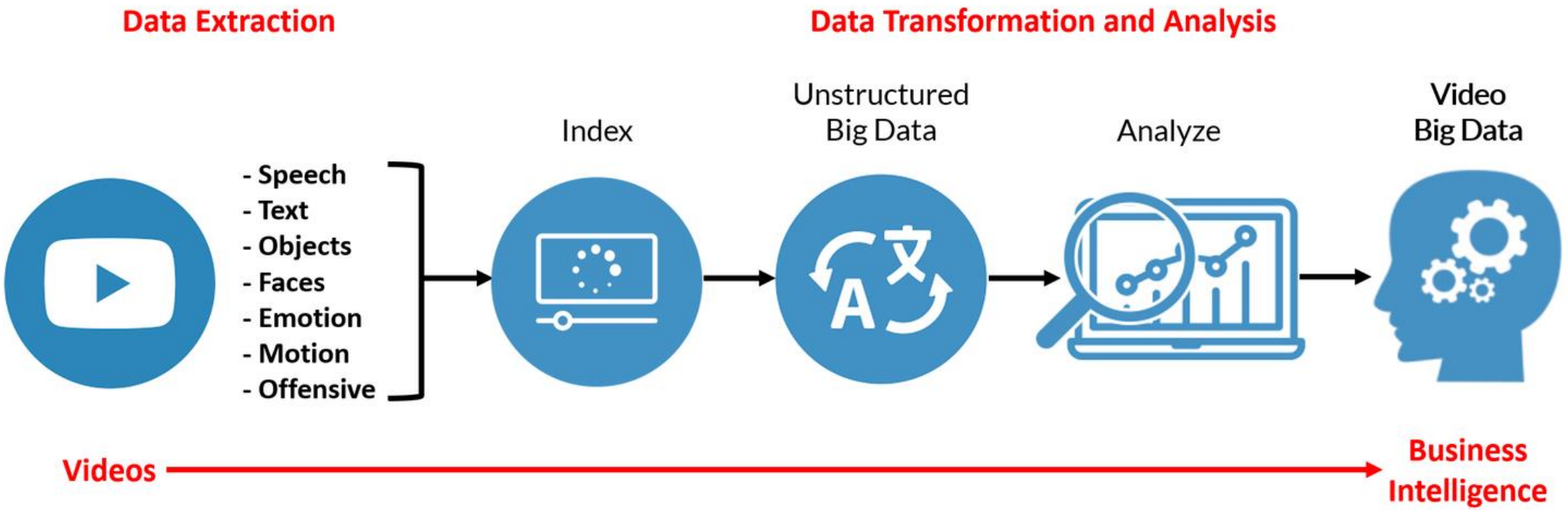
Content-Based Image Retrieval (CBIR)

- Searches for images using visual features like color and shape rather than text metadata.



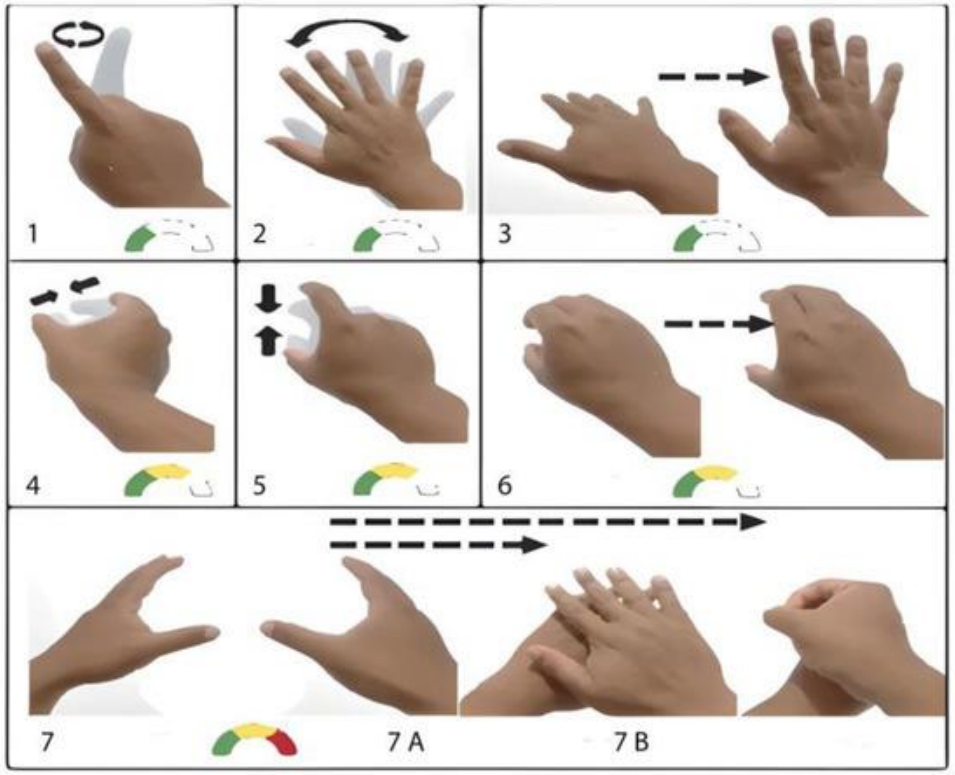
Video Data Processing

- Processes frames in videos for activity recognition, summarization, and surveillance event detection.



Multimedia

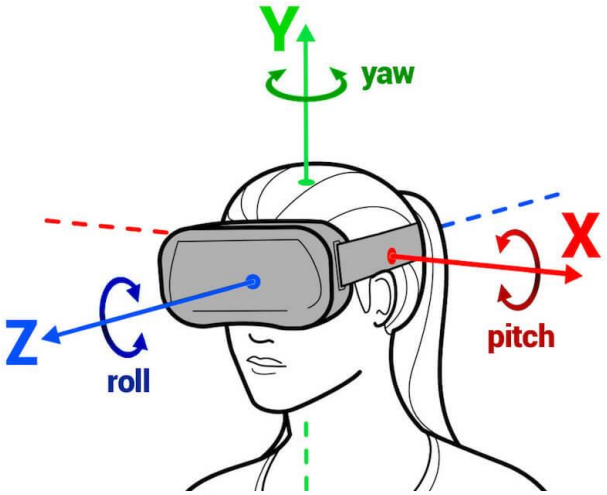
- Applies computer vision in media applications like filters, effects, gesture control, and video enhancement.



Virtual Reality (VR) & Augmented Reality (AR)

- VR offers immersive 3D scenes, AR overlays information on real views. Used in gaming, education, and virtual try-ons.

Virtual & Augmented Reality



6DoF head tracking

Hand & body tracking

3D scene understanding

3D-360 video capture

Quiz!!!

Computer Vision Quiz (MCQ)

Choose the correct option from the given alternatives.

1. Which of the following best defines Computer Vision?
A. Study of digital sound
B. Automatic interpretation of images and videos
C. Creating 3D models manually
D. Writing text descriptions of algorithms

✅ **Answer:** B

2. What is the main difference between Computer Graphics and Computer Vision?
A. Graphics uses images, vision uses text
B. Computer Graphics generates images, Computer Vision analyzes images
C. Vision uses only grayscale, Graphics uses color
D. They are the same

✅ **Answer:** B

Which level of computer vision involves edge detection and corner detection?

A. High-level

B. Mid-level

C. Low-level

D. All of the above

Which of the following is primarily a Mid-level vision task?

A. Scene interpretation

B. 3D reconstruction

C. Edge filtering

D. Object categorization

What is the primary focus of high-level vision?

A. Feature extraction

B. Motion detection

C. Scene interpretation

D. 3D modeling

Which technique is essential for extracting text from scanned documents?

A. Biometrics

B. Object Tracking

C. Document Image Analysis

D. Augmented Reality

Which application uses facial geometry and fingerprint analysis?

A. Medical Image Analysis

B. Biometrics

C. Object Detection

D. Multimedia

Which application uses feature matching across video frames?

A. Document Analysis

B. Tracking

C. Content Retrieval

D. VR

CBIR stands for:

A. Computer-Based Image Rendering

B. Content-Based Image Retrieval

C. Common Biometric Image Recognition

D. Camera-Based Inference Recognition

Which application is heavily used in diagnostic imaging?

A. Object Recognition

B. Medical Image Analysis

C. Tracking

D. Virtual Reality

Which of the following enhances user experience through artificial interaction?

A. Object Detection

B. CBIR

C. Virtual and Augmented Reality

D. Medical Imaging

Which of the following links computer vision with 3D rendering techniques?

A. Biometrics

B. Computer Vision and Graphics

C. Content-Based Retrieval

D. Medical Analysis