

Image Processing and Computer Vision

Object Detection, Recognition and Tracking

Sanath Thilakarathna

BSc Eng (Hons), MSc (Reading)

CINEC Campus

March 5, 2026



Object Detection, Recognition and Tracking

Definition

Object detection, recognition, and tracking are core tasks in computer vision that allow systems to locate, identify, and follow objects within images or video sequences.

- **Object Detection** – Locating objects in an image.
- **Object Recognition** – Identifying the class or category of the detected object.
- **Object Tracking** – Following an object across multiple frames in a video.

Object Detection Example: Traffic Monitoring

Example

In a traffic monitoring system:

- Detection identifies where vehicles appear in the frame.
- Recognition determines whether the object is a car, bus, or truck.
- Tracking follows each vehicle as it moves through the video.

Important

Splitting these concepts across multiple steps enables efficient and robust computer vision systems.

Detection vs Recognition vs Tracking

Task	Purpose
Object Detection	Locate objects within an image
Object Recognition	Identify the class of the detected object
Object Tracking	Follow the object's movement across frames

Important

In most computer vision systems, detection is performed first, followed by recognition and then tracking in video sequences.

Colour-Based Object Detection

Definition

Colour-based detection identifies objects using distinctive colour information in an image.

- Objects are detected by segmenting specific colour ranges.
- Colour spaces such as RGB or HSV are commonly used.
- Effective when the object has a distinct colour from the background.

Example

A robot can detect a red ball by identifying pixels within a specific red colour range in the HSV colour space.

Limitations of Colour-Based Detection

Note

Colour-based detection may fail when lighting conditions change or when objects have similar colours to the background.

- Sensitive to illumination variations.
- Requires careful calibration of colour thresholds.
- Limited robustness in complex scenes.

Important

Feature-based methods provide better robustness than colour-based approaches for general object detection tasks.

Feature-Based Object Detection

Definition

Feature-based detection identifies objects using distinctive visual features such as edges, corners, or keypoints.

- Visual features are extracted from the image.
- These features are compared with features of known objects.
- Feature matching determines whether the object exists in the image.

Example

Algorithms such as SIFT, SURF, and ORB detect keypoints in an image and match them with reference object features.

Advantages of Feature-Based Detection

Important

Feature-based methods are more robust than colour-based methods because they rely on structural properties of objects.

- Invariant to illumination changes and rotations.
- More effective in complex and cluttered scenes.
- Enables reliable object matching across different viewpoints.

Note

However, feature extraction can be computationally expensive compared to colour-based approaches.

Viola–Jones Object Detection Framework

Definition

The Viola–Jones framework is a classical object detection method designed for real-time detection, particularly for face detection.

- Uses Haar-like features to represent visual patterns.
- Employs a cascade of classifiers for efficient detection.
- Designed for fast real-time performance.

Many early face detection systems used the Viola–Jones algorithm with Haar cascade classifiers to detect faces in webcam images.

Important

The Viola–Jones framework was one of the first methods capable of real-time object detection.

Key Components of the Viola–Jones Framework

- **Haar-like Features** Simple rectangular patterns used to detect edges and intensity changes.
- **Integral Image** A representation that allows very fast computation of Haar features.
- **AdaBoost Learning** A machine learning technique used to select the most useful features.
- **Cascade Classifier** A sequence of classifiers that quickly rejects regions that do not contain the object.

Important

The cascade structure allows the algorithm to reject most non-object regions quickly, enabling real-time detection.

Haar Cascade Classifiers

Definition

A Haar Cascade classifier is a trained object detection model that implements the Viola–Jones framework using Haar-like features and cascade classifiers.

- Widely used for real-time face detection.
- Implemented in OpenCV as pre-trained models.
- Can run efficiently on low-power hardware.

Important

Haar Cascade detectors are practical implementations of the Viola–Jones algorithm.

Haar Cascades in OpenCV

Example

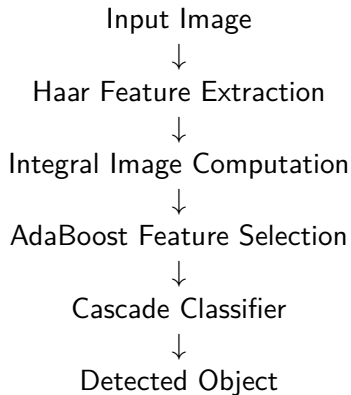
OpenCV provides several pre-trained Haar Cascade models such as:

- `haarcascade_frontalface_default.xml`
- `haarcascade_eye.xml`
- `haarcascade_smile.xml`
- `haarcascade_fullbody.xml`

Note

These models are trained using the Viola–Jones method and can be directly used in computer vision applications.

Viola–Jones Detection Pipeline



This pipeline allows efficient real-time detection of objects such as faces in images and video streams.

- Object detection locates objects within an image.
- Object recognition identifies the class of the object.
- Object tracking follows objects across video frames.
- Colour-based methods detect objects using colour segmentation.
- Feature-based methods detect objects using visual features.
- The Viola–Jones framework introduced real-time object detection.
- Haar Cascades are trained detectors based on the Viola–Jones framework.