Image Segmentation

1. Introduction / Definition

Segmentation subdivides an image into its constituent regions or objects. The level to which subdivision is carried depends on the problem being solved — segmentation should stop when the objects of interest have been isolated.

Let the entire image region be R. Segmentation partitions R into n sub-regions $R_1, R_2, \ldots, R_n$ such that:

1. $\bigcup_{i=1}^{n} R_i = R$ (union of all regions = whole image)
2. R_i is a connected set, for $i = 1, 2, \ldots, n$
3. $R_i \cap R_j = \emptyset$ for all i and j, $i \neq j$ (regions are disjoint)
4. $P(R_i) = \text{TRUE}$ for $i = 1, 2, \ldots, n$ (all pixels in a region satisfy a property, e.g., same intensity)
5. $P(R_i \cup R_j) = \text{FALSE}$ for any adjacent regions R_i and R_j (regions are maximally different)

Segmentation algorithms for gray-scale images are based on two principal properties of intensity values:

- Discontinuity — partitioning the image based on abrupt changes in intensity (edges).
- Similarity — partitioning based on grouping pixels that share similar characteristics (thresholding, region growing, region splitting/merging).

2. Detection of Discontinuities

Three basic types of gray-level discontinuities: points, lines, and edges. Detected by running a mask over the image and using the response:

$$R = w_1 z_1 + w_2 z_2 + \ldots + w_9 z_9 = \Sigma\, w_i z_i$$

where z_i are gray levels of pixels under the mask and w_i are mask coefficients.

(a) Point Detection

A point is detected at location (x, y) if $|R| \geq T$, where T is a nonnegative threshold. The mask used is the Laplacian-type mask with center coefficient 8 and surrounding coefficients −1.

(b) Line Detection

Uses four masks to detect lines in horizontal, vertical, +45°, and −45° directions. If, at a point, $|R_i| > |R_j|$ for all $j \neq i$, that point is more likely to be associated with a line in direction i.

(c) Edge Detection

An edge is a set of connected pixels lying on the boundary between two regions with different gray-level properties. Edge detection is the most common approach.

Gradient-based operators — use first-order derivatives:

$$\nabla f = [Gx, Gy] \text{ where } Gx = \partial f/\partial x,\ Gy = \partial f/\partial y$$

$$\text{Magnitude: } M(x,y) = \sqrt{(Gx^2 + Gy^2)} \approx |Gx| + |Gy|$$

$$\text{Direction: } \alpha(x,y) = \tan^{-1}(Gy/Gx)$$

Common masks:

- Roberts cross-gradient operators (2×2)
- Prewitt operators (3×3)

- Sobel operators (3×3) — gives smoothing + differentiation, more noise-resistant

Second-order derivative — Laplacian:

$$\nabla^2 f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2$$

Used mainly for finding zero-crossings.

Laplacian of Gaussian (LoG / Marr-Hildreth operator): Smooths the image with a Gaussian first, then applies the Laplacian, to reduce noise sensitivity; edges are located at zero crossings.

Canny Edge Detector (optimal edge detector):

1. Smooth image with Gaussian filter.
2. Compute gradient magnitude and direction.
3. Apply non-maxima suppression to thin edges.
4. Use double thresholding (Th, Tl) and edge linking by hysteresis to detect and connect edges.

Edge Linking:

- Local processing (analyzing neighborhood pixels for similar magnitude/direction)
- Global processing using the Hough Transform (maps edge points to parameter space to detect lines/curves)
- Global processing via graph-theoretic techniques

3. Thresholding

Simplest segmentation method — separates objects from background using intensity value T.

$$g(x,y) = 1 \text{ if } f(x,y) > T \text{ (object)}$$
$$g(x,y) = 0 \text{ if } f(x,y) \leq T \text{ (background)}$$

Types:

- Global thresholding — single T for entire image (works well when histogram has two clear peaks separated by a valley — bimodal histogram).

Basic Global Thresholding Algorithm:

1. Select initial estimate for T.
2. Segment image using T into G1 (>T) and G2 (≤T).
3. Compute average intensities μ1 and μ2.
4. Compute new T = ½(μ1 + μ2).
5. Repeat until ΔT is smaller than a predefined limit.

- Otsu's Method — chooses threshold to maximize between-class variance $\sigma^2_B(k)$, statistically optimal, based on the histogram.
- Adaptive/Local thresholding — image divided into sub-images and threshold computed separately for each (useful for uneven illumination).
- Multiple thresholding — used when histogram is multimodal.

Factors affecting thresholding: noise, non-uniform illumination — improved by using illumination correction or edge information combined with histograms.

4. Region-Based Segmentation

Groups pixels into regions based on predefined similarity criteria.

(a) Region Growing

Starts with a set of "seed" points, and grows regions by appending neighboring pixels having similar properties (e.g., intensity within a threshold range).

Steps:

1. Choose seed point(s).
2. Define similarity criterion (e.g., $|f(x,y) - \text{seed value}| \leq T$).
3. Define connectivity (4- or 8-connected).
4. Append qualifying neighboring pixels to region.
5. Stop when no more pixels satisfy the criterion.

(b) Region Splitting and Merging

Uses quadtree representation. Let R represent entire image and predicate Q.

1. Split image into four disjoint quadrants whenever $Q(R_i) = \text{FALSE}$.
2. Merge any adjacent regions R_j and R_k if $Q(R_j \cup R_k) = \text{TRUE}$.
3. Stop when no further splitting or merging is possible.

(c) Region Segmentation using Watersheds

Based on visualizing an image in 3-D — two spatial coordinates and intensity as elevation. The method identifies:

- Regional minima treated as catchment basins,
- Water rises from minima, and dams (watershed lines) are built where waters from different basins would merge — these dams form the segmentation boundaries.

Often, the gradient image is used instead of raw intensity to avoid over-segmentation, and marker-controlled watershed segmentation is used to reduce noise-induced over-segmentation by using internal and external markers.

5. Segmentation Using Motion

Motion is a powerful cue used to segment moving objects from stationary backgrounds.

Differencing approach: Compute difference image between two frames:

$$d(x,y) = 1 \text{ if } |f(x,y,t_i) - f(x,y,t_j)| > T, \text{ else } 0$$

Non-zero entries indicate motion/changes.

Accumulative differences over multiple frames build motion history to detect direction and speed.

6. Summary Table

Approach	Basis	Techniques
Discontinuity-based	Abrupt intensity changes	Point, Line, Edge detection (Sobel, Prewitt, Canny, LoG)
Similarity-based (Thresholding)	Pixel intensity grouping	Global, Otsu, Adaptive, Multiple thresholding
Region-based	Spatial + intensity similarity	Region growing, Splitting & merging, Watershed
Motion-based	Temporal change	Frame differencing

7. Conclusion

Image segmentation is a critical intermediate step in image analysis that converts raw pixel data into meaningful regions or objects. The choice of technique (discontinuity-based, similarity-based, region-based, or motion-based) depends on the nature of the image, noise level, and application requirements (medical imaging, object recognition, industrial inspection, etc.).