Segmentation
Segmentation: the problem

Identifying entities in the image, e.g., objects:

- grouping pixels into segments
- crucial and basically unsolved step
Segmentation

1. Thresholding
2. Region based
3. Edge based
Thresholding : principle

For high contrasts between objects and background

Determine intensity threshold that defines 2 pixel categories: object and background

Example image:
Thresholding : result
Threshold selection

- Segmentation is the basis of subsequent decisions
- Example: spine fracture
- Danger: injury of spinal chord
- Vertebrae:

healthy

broken

Bony fragment identification through threshold
Threshold identification on histogram

X-ray attenuation is tissue dependent

- strong overlap
- limited separability on histogram
Statistical pattern recognition

- Feature-based object classification
- General scheme

- Application for threshold determination
- Fracture example
  - objects: pixels in CT image
  - sensor: CT device
  - measurement: CT attenuation (Hounsfield units)
  - feature: measurements used directly
  - object classes: tissue types
- In general feature extraction is essential
  - can be quite complex
  - major influence on classification results
Notation

• \( \Omega \) the set of \( s+1 \) possible classes (\( \omega_r \) rejection)

\[
\Omega = \{\omega_1, \omega_2, \ldots, \omega_s, \omega_r\}
\]

• Extracted features

\[
\vec{\nu} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n
\]

• span a linear space, non deterministic

• Can be characterized by the joint probability density function (PDF): \( p(\vec{\nu}, \omega_i) \)

• Marginal distributions
  • probability of class occurrence
    (a priori probability)

\[
P(\omega_i) = \int p(\vec{\nu}, \omega_i) d\vec{\nu}
\]

• probability of observing a specific feature

\[
p(\vec{\nu}) = \sum_{i=1}^{s} p(\vec{\nu}, \omega_i)
\]
Stochastic characterization

• Feature distribution for class $\omega_i$ (class specific probability distribution)

$$p(\vec{v}|\omega_i) = \frac{p(\vec{v}, \omega_i)}{P(\omega_i)}$$

• Probability of class $\omega_i$ if feature $\vec{v}$ observed

$$P(\omega_i|\vec{v}) = \frac{p(\vec{v}, \omega_i)}{p(\vec{v})}$$

(a posteriori probability)

• Fracture example

  • $\Omega = \{\text{bone, nervous tissue, muscle, air, don’t know}\}$
  • $\vec{V}$: voxel attenuation (HU)
  • $P(\omega_{\text{bone}})$: probability of voxel being in bone
  • $p(\vec{V} = 250)$: probability of observing the given HU in a CT image (approximated by histogram)
  • $p(\vec{V} | \omega_{\text{bone}})$: attenuation distribution of bony tissue
  • $P(\omega_{\text{bone}} | \vec{V} = 250)$: probability of bony tissue for the given attenuation
Maximum a posteriori decision

- Selecting the most probable class for a given feature
- Looking for class $\omega_i$ maximizing $P(\omega_i | \vec{v})$
- Cannot be estimated directly
- Calculated by using the Bayes theorem

$$P(\omega_i | \vec{v}) = \frac{p(\vec{v} | \omega_i) P(\omega_i)}{p(\vec{v})}$$

- estimation of probabilities on the right side
  - $P(\omega_i)$ from prior knowledge
  - $p(\vec{v})$ from specific observation
  - $p(\vec{v} | \omega_i)$ can be learned from examples
Decision theory

- Formal framework for decision making
- Possible decisions $A = \{\alpha_1, \alpha_2, \ldots, \alpha_a\}$
- Decisions associated with costs ($\lambda$)
- Depend on the underlying object class $\omega_k$
- Cost function $\lambda(\alpha_i | \omega_k)$
- The risk of a decision if a specific feature $\vec{v}$ is observed

$$R(\alpha_i | \vec{v}) = \sum_{k=1}^{s} \lambda(\alpha_i | \omega_k) P(\omega_k | \vec{v})$$

- The overall risk over the whole feature space

$$R_{tot} = \int R(\alpha(\vec{v}) | \vec{v}) p(\vec{v}) d\vec{v}$$
Risk minimization

- If feature $\mathbf{v}$ is observed, selecting the decision $\alpha_i$ minimizing $R(\alpha_i | \mathbf{v})$ (cost minimization)
- Discriminant function for every decision: $g_i(\mathbf{v})$
- Decision $\alpha_i$ selected if $g_i(\mathbf{v}) > g_k(\mathbf{v}) \quad \forall \ k \neq i$

straightforward selection $g_i(\mathbf{v}) = -R(\alpha_i | \mathbf{v})$
minimal conditional risk
Application for segmentation
Decision $\alpha_i$: assigning class $\omega_i$ to a voxel
Selection of cost

• Highly problem-specific
• Some general schemes exist
• Most straightforward: minimal error rate
  \[ \lambda(\alpha_i | \omega_k) = \begin{cases} 
    0 & i = k \\
    1 & i \neq k 
  \end{cases} \]

• Conditional risk
  \[ R(\alpha_i | \vec{v}) = \sum_{k \neq i} P(\omega_k | \vec{v}) = 1 - P(\omega_i | \vec{v}) \]

• Discriminant function
  \[ g_i(\vec{v}) = P(\omega_i | \vec{v}) - 1 \]

• Alternatively (monotonic change for all decisions)
  \[ \tilde{g}_i(\vec{v}) = P(\omega_i | \vec{v}) \]
i.e. maximum a posteriori decision
Graphical interpretation

Problem with classes with low a priori probability

These may not be detected at all
Maximum likelihood decision

- Modifying the cost function
- Inverse weighting with a priori probability

\[ \lambda(\alpha_i|\omega_k) = \begin{cases} 
\frac{1}{P(\omega_k)} & i \neq k \\
0 & i = k 
\end{cases} \]

- Conditional risk

\[ R(\alpha_i|\vec{v}) = \sum_{k=1}^{s} \lambda(\alpha_i|\omega_k) P(\omega_k|\vec{v}) = \sum_{k \neq i} \frac{1}{P(\omega_k)} \frac{p(\vec{v}|\omega_k)P(\omega_k)}{p(\vec{v})} \]
Maximum likelihood decision

• By reformatting

\[
R(\alpha_i|\vec{v}) = \frac{1}{p(\vec{v})} \left[ \sum_{k=1}^{s} p(\vec{v}|\omega_k) - p(\vec{v}|\omega_i) \right]
\]

• As \( p(\vec{v}) \) and \( C(\vec{v}) = \sum_{k=1}^{s} p(\vec{v}|\omega_k) \) are not depending on the class \( i \), the discriminant function can be selected as

\[
g_i(\vec{v}) = p(\vec{v}|\omega_i)
\]

which is the a priori class probability
Graphically
Classification schemes

• How to find the necessary PDFs and prior class probabilities?
• May be known a priori in a parametric form e.g. Gaussian distributions

\[
p(\vec{v}|\omega_i) = \frac{1}{\sqrt{(2\pi)^n det(\Sigma)}} e^{-\frac{(\vec{v} - \vec{\mu})^T \Sigma^{-1} (\vec{v} - \vec{\mu})}{2 det(\Sigma)}}
\]

• where \( \vec{\mu} = E[\vec{v}] \) is the mean and \( \Sigma = E[(\vec{v} - \vec{\mu})(\vec{v} - \vec{\mu})^T] \) the covariance matrix of the random variable \( \vec{v} \)
• If the parameters unknown, they can be learned from training data provided by the user

*parametric supervised classification*
Non-parametric classification

- The distributions are not known in parametric form
- Tessellation of the parameter space based on training data

non-parametric supervised classification
Tessellation approaches

- Relying on the mean of the provided samples selecting the class with the nearest mean
- Using empirically determined density functions like histograms
  - large number of samples needed
  - not practicable especially in high-dimensional feature spaces
- Majority voting schemes, e.g. k nearest neighbour classifier (KNN): selecting the class with the most votes
- Direct estimation of class boundaries by fitting (Parsen window)
- Automatically finding structure in feature space
  - cluster analysis
  - non-parametric non-supervised classification
  - seldom successful
Thresholding: alternatives

If area of object(s) and background is known: place threshold at intensity that yields appropriate proportions.

Several techniques that include edge information:

- e.g. maximize sum of gradient magnitudes for the pixels with threshold intensity.
- e.g. use a histogram taking only low gradient magnitude pixels into account.
Otsu criterion

Minimize within-group variance

\[ p_1 \sigma_1^2 + p_2 \sigma_2^2 \]
Thresholding: remarks

- Threshold advantages:
  1. Serious bandwidth reduction
  2. Simplification of further processing
  3. Availability of real-time hardware for shape recognition

- Generally it won’t provide a satisfying segmentation

- Sometimes several thresholds lead to finer labelling, usually infeasible

- Pixel-by-pixel decision:
  - ignores neighbouring pixels
  - structural information lost
Segmentation

- Thresholding
- Edge based
- Region based
Segmentation

- Thresholding
- Edge based
- Region based
A serious challenge!

- The goal is to (precisely) delineate objects
- Traditional techniques are bottom-up
- But actually recognition and segmentation should probably go hand in hand

... a tale of a chicken and an egg?
Edge linking techniques

- 1. Hough: only for predefined shapes
- 2. Elastically deformable contour models Snakes: generic shape priors
- Many other methods for grouping combination with user interaction (dynamic path search)
The Hough transform: principle

Uses parametric shape models to extract objects in lower dimensional spaces

The simplest example: straight lines

Many further possibilities, like circles
Hough transform: straight lines

Suppose we would like to detect straight lines e.g. straight edges in object outlines in all possible positions and orientations.

for general shapes: 3 degrees of freedom

Allowing three-dimensional rotations the situation would even get more complex

Straight lines, however, will remain invariant under several such transformations

The image projection of a straight line is fully characterized by 2 parameters
Hough transform: straight lines

We write the equation of a straight line as

\[ y = ax + b \]

Fixing a point \((x,y)\), all lines through the point:

\[ b = (-x) a + y \]

The Hough transform:

1. Scrutinize all points of interest
2. For each point draw the above line in \((a,b)\) - parameter space
Hough transform: straight lines

Implementation:
1. The parameter space is discretised
2. A counter is incremented at each cell where the lines pass
3. Peaks are detected

\[ b = (-x) a + y \]
Hough transform: straight lines

Problem: unbounded parameter domain

Vertical lines require infinite $a$

Alternative representation:

Each point will add a cosine function in the $(\rho, \theta)$ parameter space
Hough transform : straight lines

Square :

Circle :
Hough transform : straight lines
Hough transform: remarks

1. time consuming

2. robust

3. peak detection is difficult

4. Robustness of peak detection increased by weighting contributions (e.g., in examples weighting with intensity gradient magnitude)
The structure of discrete image spaces

Two main aspects
- Topology
- Distance

Strong relationship
Topology

- The problem of the bridges of Königsberg
- Solution: Euler, 1736
- The birth of graph theory
- Independent of distances
Formal definition

- Topological space

\[(X, \Theta_X)\]

- \(X\) is a set of points,
- \(\Theta_X\) are the open subsets above \(X\)
- Describes the neighbourhood structure of a space
Neighbourhood of the Cartesian image raster

- Defined by the neighbourhood structure
- There is no unique definition
- 4- and 8-connectivity the most popular
- There are other possibilities
Topology of discrete binary images

• Defined by connectivity (Neighbourhood $\Box$)
• Connected pixels through neighbourhood chain

$$X \sim_\Box Y \leftrightarrow \exists \{X_1, X_2, \ldots, X_n\}$$

such that $X = X_1$, $Y = X_n$ and $X_i \Box X_{i+1}$

• Two components: Foreground, Background
• Topology defined by the number of FG/BG components
Determination of topology

- Region growing

\[ \forall X \in K \quad K(X) = 0 \]
\[ P \rightarrow S, \quad K(P) = 1 \]
while \( S \neq \emptyset \)
\[ Q \leftarrow S \]
\[ \forall Q_i \text{ such, that } Q_i \subseteq Q \]
if \( I(Q_i) = I(Q) \quad \land \quad K(Q_i) = 0 \)
\[ Q_i \rightarrow S \]
\[ K(Q_i) = 1 \]
endif

end

- Components can be found one by one until all are labeled
Connected component labeling

- Scanning the image line-by-line (TV scan) enforces an (artificial) causality
- At every pixel the neighbourhood divided into past and future
  
  ![Neighbourhood Diagram]

- All labels on past pixels collected. Possible options
  - No label found: give a new label
  - One label found: propagate it to the central pixel
  - More than one label found: note their equivalence
- At the end equivalent components (connected but labelled as different) have to be re-labeled
Distance on the image raster

- \( d : \Omega \times \Omega \rightarrow \mathbb{R}^+ \) is a metrics over \( \Omega \) if for all \( P, Q \in \Omega \):
  - \( d(P, Q) \leq 0 \) and \( d(P, Q) = 0 \) if and only if \( P \equiv Q \)
  - \( d(P, Q) = d(Q, P) \) (symmetry)
  - for all \( R \in \Omega \) \( d(P, R) + d(R, Q) \geq d(P, Q) \)

- One can define open and closed disks with center \( P \) and radius \( r \):
  \[
  \bar{D}(P, r) = \{ Q \in \Omega \mid d(P, Q) \leq r \}
  \]
  \[
  D(P, r) = \{ Q \in \Omega \mid d(P, Q) < r \}
  \]

which can be used to define neighbourhoods

- Intimate connection between topology and metrics
Topology induced metrics

- Defined based on connectivity
  - Length of connecting paths (number of steps)
  - Minimal length between all connecting paths

- $D_4$ (Manhattan) and $D_8$ distances e.g.
- How to calculate between $P(i,j)$ and $Q(k,l)$?
  \[
  D_4(P, Q) = |i - k| + |j - l|,
  \]
  \[
  D_8(P, Q) = \max(|i - k|, |j - l|)
  \]

- Circles in $D_4$ and $D_8$?

- Euclidean not conform with any discrete neighbourhood
Distance calculation

- Distance transformations: distance maps based on distance propagation along neighbourhoods

- Euclidean distance map
  - True implementation is very cumbersome
  - Approximation by increasingly large neighbourhoods
Enhancement of binary images: problem

Pixels can fall on the wrong side of the threshold

An example:
Binary enhancement : example

Threshold 35 :

Ouch!
Computer Vision

Binary enhancement: sketch

Aspects

- neighbourhood structuring element
- operator $f_t$
  value selection

Other neighbourhoods & functions $f_t$ can also be used.
Morphological operators

original image structure

structuring element

Erosion

Dilation

Opening

Closing
Binary enhancement : rank order op.

New intensity based on rank-ordered neighbourhood values

Important difference with convolution : non-linearity

\[ i_1 \leq i_2 \leq \ldots \leq i_N \]

\[ i_t = f_t(i_1, \ldots, i_N) \]

<table>
<thead>
<tr>
<th>Examples</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ i_t = i_1 ]</td>
<td>erosion</td>
</tr>
<tr>
<td>[ i_t = i_N ]</td>
<td>dilation</td>
</tr>
<tr>
<td>[ i_t = i_{(N+1)/2} ]</td>
<td>median</td>
</tr>
</tbody>
</table>
Binary enhancement: principle

Dilation followed by erosion (closing):
Post-processing chain: example
Binary enhancement: remarks

- **1. Post-processing approach**
  Many alternatives to enforce neighborhood consistency during segmentation

- **2. Erosion + dilation (opening)**
  \[ \text{dilation + erosion (closing)} \]

- **3. Noise in background can be reduced by reversed operation**

- **4. Use same neighbourhood for both steps**

- **5. Reminder:** median filtering useful for edge preserving smoothing
Segmentation

- Thresholding
- Edge based
- Region based
Region growing: principle

On the basis of segment homogeneity rather than inhomogeneity around edges

start with detection of very “homogeneous” regions
these are the “seeds”
e.g. low intensity variance

These are grown as long as homogeneity criterion satisfied

Choice of appropriate homogeneity criteria is not straightforward
Region growing: example

Seeds of low intensity variance are grown, keeping intensity between two slowly floating thresholds and merging overlapping segments.
Region growing : ex. cont’d

will often fail, risk of leakage through low contrast edges
Region growing : remarks

region growing pure is a one-way process: if seeds are wrong, errors cannot be corrected

solution: split-and-merge procedures

merges of similar regions
splitting more difficult: many ways to do it

splitting calls for special decompositions of segments, e.g. "quadtrees"

Region and edge based methods can be combined: hybrid approaches
Watershed algorithm

Finding catchment basins in the image intensity graph

- Starting with local minima
- Different implementations of water filling (usually based on some region growing)
  - Stop propagation at detected ridges (lines)
  - Region competition (distance based)
- Usually strong oversegmentation
Watershed segmentation example