Principles of grey-level thresholding

Grey-level thresholding is a simple image segmentation technique that assumes the following conditions:

- Scene model: Scene contains uniformly illuminated, flat surfaces.
- Image model: Image is a set of approximately uniform regions.

Goals of thresholding: Set one or more thresholds which split the intensity range into intervals defining intensity classes

- Separate objects from background.
- Label objects by classifying pixel intensities into two or more classes.

Definition of $N$-level thresholding

Set $N - 1$ thresholds $T_k$, $k = 1, \ldots, N - 1$, $N \geq 2$, so that a pixel $f(x, y)$ is classified into class $n$ if

$$T_{n - 1} \leq f(x, y) < T_n, \quad n = 1, \ldots, N,$$

where by definition $T_0 \doteq G_{\min}$ and $T_N \doteq G_{\max} + 1$ are the limits of the intensity range (0 and 256).

Illustration of 4-level thresholding. By definition, $T_0 = 0$ and $T_4 = 256$. The first level is the background.
Examples of automatic thresholding into 2 and 3 levels.

- The case of a single threshold \( (N = 2) \) is called **bilevel** (binary) thresholding, or **binarisation**.
  - The case considered in this course.

- If \( N > 2 \), thresholding is called **multilevel**.
  - Sometimes, the case \( N = 3 \) is called **trilevel** thresholding.

Examples of good and bad threshold selections for a fingerprint image:

- Different thresholds are acceptable.

- A **too low** threshold tends to split the lines.

- A **too high** threshold tends to merge the lines.

Histogram-based thresholding

- **Bimodal histogram** with distinct modes and valley between modes is **most suitable** for threshold selection. **Minimum of valley** separates the 2 classes.

- If a mode lies at limit of intensity range, modelling the histogram is difficult.

- If modes are not distinct, setting a good threshold is not easy.

- Thresholding a **unimodal histogram** is difficult but still possible.

Typical histogram shapes for threshold selection. From left to right: bimodal histogram; mode is too small, peak is too close to limit; unimodal histogram.

Histogram modality analysis

**Algorithm**: Select threshold(s) in valley(s) between peaks.

**Parameters**:

- Minimum height of peak

- Minimum distance between peaks

Thresholding a fingerprint image. In the histogram, positions of good (G), too low (L) and too high (H) thresholds are shown.

Histogram modality analysis: Selecting thresholds in valleys between peaks.
Advantages of modality analysis:

- Natural and easy to understand.
- Multilevel thresholding possible.
- Relatively small populations (classes) can be treated, at least in principle.

Drawbacks of modality analysis:

- Subjective: What is a peak? A valley?
- Several parameters should be preset that specify these histogram features.
- Many histograms are not multimodal
  - Unimodal histograms
  - Histograms having no clear modes
  - Possible solution: Modify histogram to obtain distinct modes (discussed later)

Consider the normalised histogram $P(i)$, $i = 0, 1, \ldots, M$. It has mean $\mu$ and variance $\sigma^2$:

$$\mu = \sum_{i=0}^{M} i \cdot P(i) \quad \sigma^2 = \sum_{i=0}^{M} (i - \mu)^2 \cdot P(i)$$

A candidate threshold $t$ splits the histogram into 2 classes whose means and variances are

$$\mu_k(t) = \frac{1}{q_k(t)} \sum_{i=a_k}^{b_k} i \cdot P(i) \quad \sigma^2_k(t) = \frac{1}{q_k(t)} \sum_{i=a_k}^{b_k} (i - \mu_k(t))^2 \cdot P(i)$$

where $k = 1, 2$, $a_1 = 0$, $b_1 = t$, $a_2 = t + 1$, $b_2 = M$ and

$$q_k(t) = \sum_{i=a_k}^{b_k} P(i) \quad q_1(t) + q_2(t) = 1$$

Maximal separation of classes (N.Otsu, 1978)

Basic idea:

- Consider a candidate threshold $t$. $t$ defines two classes of grayvalues.
- Find the optimal threshold $t = t_{opt}$ as the one that maximises a separation measure for the two classes.

Introduce between-class variance $\sigma_B^2(t)$ and within-class variance $\sigma_W^2(t)$:

$$\sigma_B^2(t) = q_1(t) \cdot [1 - q_1(t)] \cdot [\mu_1(t) - \mu_2(t)]^2$$

$$\sigma_W^2(t) = q_1(t) \cdot \sigma_1^2(t) + q_2(t) \cdot \sigma_2^2(t)$$

It is easy to show that

$$\mu = q_1(t) \cdot \mu_1(t) + q_2(t) \cdot \mu_2(t) \quad \sigma^2 = \sigma_W^2(t) + \sigma_B^2(t)$$

Since $\sigma_W^2(t) + \sigma_B^2(t)$ is constant, we have two equivalent options:

- $\sigma_B^2(t)$ is a measure of class separation $\Rightarrow$ Maximise $\sigma_B^2(t)$
- $\sigma_W^2(t)$ is a measure of class overlap $\Rightarrow$ Minimise $\sigma_W^2(t)$

We use the first option.
To compute $\sigma_B^2(t)$ for any discrete $t > 0$, **recursive formulae** are used:

\[ q_1(t + 1) = q_1(t) + P(t + 1) \quad \text{with} \quad q_1(0) = P(1) \]
\[ \mu_1(t + 1) = \frac{q_1(t) \cdot \mu_1(t) + (t + 1) \cdot P(t + 1)}{q_1(t + 1)} \quad \text{with} \quad \mu_1(0) = 0 \quad (3) \]
\[ \mu_2(t + 1) = \frac{\mu_1(t + 1) - q_1(t + 1) \cdot \mu_1(t + 1)}{1 - q_1(t + 1)} \]

**Algorithm 1: Otsu threshold selection**

1. Compute normalised histogram $P(i)$ of image $I(r, c)$.
2. Starting from $t = 0$ and using (3) and (1), recursively compute $q_1(t), \mu_1(t)$ and $\mu_2(t)$ for each $t < G_{\text{max}}$.
3. For each $0 < t < G_{\text{max}}$, calculate $\sigma_B^2(t)$ by (2).
4. Select threshold as $t_{\text{opt}} = \arg \max_t \sigma_B^2(t)$.

**Properties of Otsu threshold selection**

**Advantages:**
- General: No specific histogram shape assumed.
- Works well, stable.
- Extension to **multilevel thresholding** possible.
  - For $N$ thresholds and $M + 1$ grey levels, optimisation of class separation needs maximum search in a $(M + 1)^N$ array

**Drawbacks:**
- The method assumes that $\sigma_B^2(t)$ is unimodal. This is not always true.
- When optimisation function is flat, false maxima may occur.
- The method tends to artificially **enlarge small classes** to obtain 'better separation': small classes may be merged and missed.

**Histogram modelling by Gaussian distributions**

Basic idea:
- Assume that the histogram $P(i)$ is a **mixture of Gaussian distributions**
- Approximate $P(i)$ by this model and estimate the parameters of the model
- Find optimal threshold(s) analytically as valley(s) in the model function

![Modeling the histogram by a mixture of two Gaussian distributions.](image)

Approximate the histogram $P(i)$ with the model distribution
\[
 f(i; \mathbf{p}) = q_1 \cdot f_1(i; \mathbf{p}_1) + q_2 \cdot f_2(i; \mathbf{p}_2)
\]
\[
 = \frac{q_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2} \left(\frac{i - \mu_1}{\sigma_1}\right)^2\right) + \frac{q_2}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \left(\frac{i - \mu_2}{\sigma_2}\right)^2\right),
\]
where $\mathbf{p} = (q_1, q_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$, $\mathbf{p}_1 = (q_1, \mu_1, \sigma_1)$, $\mathbf{p}_2 = (q_2, \mu_2, \sigma_2)$ are the parameter sets of the functions $f, f_1$ and $f_2$.

$q_1$ and $q_2$ are the weights of the two distributions. Since $q_1 + q_2 = 1$, $f$ has five free parameters. Exclude $q_2$ and denote $\mathbf{p}' = (q_1, \mu_1, \mu_2, \sigma_1, \sigma_2)$.

Introduce the **error function**
\[
 C(\mathbf{p}') = \sum_i [f(i; \mathbf{p}') - P(i)]^2
\]

1. Compute normalised histogram $P(i)$ of image $I(r, c)$.
2. Starting from $t = 0$ and using (3) and (1), recursively compute $q_1(t), \mu_1(t)$ and $\mu_2(t)$ for each $t < G_{\text{max}}$.
3. For each $0 < t < G_{\text{max}}$, calculate $\sigma_B^2(t)$ by (2).
4. Select threshold as $t_{\text{opt}} = \arg \max_t \sigma_B^2(t)$.

![Modeling the histogram by a mixture of two Gaussian distributions.](image)

Theoretically correct estimation is possible in the case of **two** Gaussian distributions, that is, for **bilevel** thresholding.

\[
 f(i; \mathbf{p}) = q_1 \cdot f_1(i; \mathbf{p}_1) + q_2 \cdot f_2(i; \mathbf{p}_2)
\]
\[
 = \frac{q_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2} \left(\frac{i - \mu_1}{\sigma_1}\right)^2\right) + \frac{q_2}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \left(\frac{i - \mu_2}{\sigma_2}\right)^2\right),
\]
where $\mathbf{p} = (q_1, q_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$, $\mathbf{p}_1 = (q_1, \mu_1, \sigma_1)$, $\mathbf{p}_2 = (q_2, \mu_2, \sigma_2)$ are the parameter sets of the functions $f, f_1$ and $f_2$.

$q_1$ and $q_2$ are the weights of the two distributions. Since $q_1 + q_2 = 1$, $f$ has five free parameters. Exclude $q_2$ and denote $\mathbf{p}' = (q_1, \mu_1, \mu_2, \sigma_1, \sigma_2)$.

Introduce the **error function**
\[
 C(\mathbf{p}') = \sum_i [f(i; \mathbf{p}') - P(i)]^2
\]
To approximate \( P(i) \) with \( f(i; \hat{p}) \) and find the optimal parameters, we need to minimize \( C(\hat{p}') \). This means nonlinear minimisation with 5 variables. Any nonlinear minimisation algorithm can be used, for example:

- Newton’s method
- Marquard-Levenberg algorithm

Assume the optimal model function \( f(i; \hat{p}) \) has been obtained and
\[
\hat{p} = (\hat{q}_1, \hat{q}_2, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2)
\]
are the optimal parameters.

The optimal threshold can be derived analytically by minimising the probability of erroneous classification
\[
E(t) = E_1(t) + E_2(t) = \int_{-\infty}^{t} f_2(i) \, di + \int_{t}^{\infty} f_1(i) \, di
\]

- Two solutions for \( t_{opt} \) are possible that minimise classification error.

- If the variances are equal, \( \sigma_1^2 = \sigma_2^2 = \sigma^2 \), a single optimal threshold exists:
\[
t_{opt} = \frac{\hat{\mu}_1 + \hat{\mu}_2}{2} + \frac{\hat{\sigma}^2}{\hat{\mu}_1 - \hat{\mu}_2} \ln \left( \frac{\hat{q}_1}{\hat{q}_2} \right)
\]

Algorithm 2: Gaussian threshold selection

1. Compute normalised histogram \( P(i) \) of image \( I(r, c) \).

2. Using a minimisation algorithm, minimise the error function \( C(\hat{p}') \) defined by (5) and (4) and estimate the optimal parameters \( \hat{q}_1, \hat{q}_2, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2 \).

3. Solve equation (6) for \( t \) and obtain the optimal threshold \( t_{opt} \).

Meaning of \( E(t) \) for a candidate threshold \( t \) (see the above drawing):

- \( E_1(t) \) is the probability that a pixel belonging to class 1 will be classified as belonging to class 2.
- \( E_2(t) \) is the probability that a pixel belonging to class 2 will be classified as belonging to class 1.

Setting the derivative of \( E(t) \) to zero and substituting \( f_1 \) and \( f_2 \) from (4), we obtain that the optimal threshold \( t_{opt} \) is a solution of
\[
A \cdot t^2 + B \cdot t + C = 0,
\]

where
\[
A = \sigma_1^2 - \sigma_2^2
\]
\[
B = 2(\hat{\mu}_1 \sigma_2^2 - \hat{\mu}_2 \sigma_1^2)
\]
\[
C = \sigma_1^2 \sigma_2^2 - \sigma_2^2 \mu_1^2 + 2 \sigma_1^2 \sigma_2^2 \ln \left( \frac{\sigma_2 \hat{q}_1}{\sigma_1 \hat{q}_2} \right)
\]

Properties of the Gaussian mixture approach

Advantages:

- Relatively general histogram model.
- When the model is valid, minimises classification error probability.
- Can be applicable small-size classes.

Drawbacks:

- Many histograms are not Gaussian. In particular, intensities are finite and non-negative.
  - A peak that is close to an intensity limit cannot be approximated by Gaussian.
- Extension to multithresholding requires significant simplification of the model.
- It is difficult to detect close and flat modes.
Examples of thresholding

- Here, both methods give acceptable results.
- The Gaussian algorithm sets lower thresholds in both cases.
  ⇒ Fits object contours better than Otsu.

21

- Here, only Otsu gives a satisfactory result.
- The Otsu algorithm finds the small class of pixels (dark discs).
- The Gaussian algorithm tries to separate two high peaks formed by the background. Noisy valley is selected because the true class is
  - too small
  - too far

22

- Here, both methods still give satisfactory results.
- The Otsu algorithm sets threshold in valley of histogram.
  ⇒ Fingerprint lines are well-separated.
- The Gaussian algorithm sets slightly high threshold.
  ⇒ Some fingerprint lines touch.

23

- Here, only Otsu algorithm produces results.
- Gaussian algorithm gives no results at all.
  - Upper row: unimodal histogram, no approximation obtained
  - Lower row: an approximation obtained, but the threshold equation has no real root.
Improving the histogram for better peak separation

Use of gradient to improve histogram: Combine intensity and gradient information for better separation of objects and background.

- Pixels close to edges have high gradients and medium intensities.
- Pixels of object and background have low gradient and low or high intensities.
- To better separate objects from background, **discard high gradient pixels** when computing the histogram.

\[ T \]

\[ P(i) \]

\[ T \]

\[ i \]

**Principle of histogram peak separation.**

Thresholding versus edge detection

- **Thresholding** with a constant threshold is a global operation.
  - Advantage: Closed contours guaranteed
  - Drawback: Not applicable to images with uneven illumination

- **Edge detection** is a local operation
  - Advantage: Applicable to images with uneven illumination
  - Drawback: Closed contours not guaranteed

\[ f(x) \]

\[ |f_x(x)| \]

**Signal with varying level that cannot be thresholded. Edges can be detected.**

Limits of thresholding

- Merit (quality) of thresholding is **task-dependent**.
- The merit may include **geometric properties**.
- Image histogram does not account for geometry.
  - The crack is detected as set of bright pixels **independently of the crack shape**.

\[ \text{stone crack Otsu method Another method} \]

\[ \text{stone Otsu method Another method} \]

*In this example, the merit of thresholding is uncertain.*

Limits of thresholding:

- No geometric information is taken into account.
  - \[ \Rightarrow \] Compact and connected regions are not guaranteed.
    - Select threshold, then arbitrarily interchange pixels in image, select again \[ \Rightarrow \] same threshold
- Solution: Combine intensity and geometry using **region-oriented** methods.