Basic Algorithms for Digital Image Analysis

Dmitrij Csetverikov

Eötvös Loránd University, Budapest, Hungary
csetverikov@inf.elte.hu
csetverikov@sztaki.hu

Faculty of Informatics

Thresholding

1. Principles of greyvalue thresholding
   - Histogram-based thresholding
2. Two methods for threshold selection
   - Otsu's method
   - Histogram modelling by Gaussian distributions
3. Examples and analysis of thresholding
   - Examples of thresholding
   - Analysis of thresholding

Grey-level thresholding

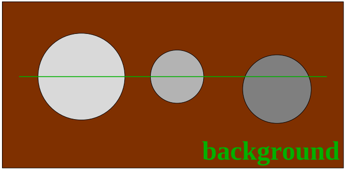
- Basic **image segmentation** technique
- Assumes following **conditions**
  - scene contains uniformly illuminated, flat surfaces
  - image is set of approximately uniform regions
- **Goal**
  - set one or more **thresholds** which split intensity range into intervals
  - $\Rightarrow$ define **intensity classes**
- **Result**
  - objects labelled by classifying pixel intensities into classes
  - $\Rightarrow$ objects separated from background

Definition of N-level thresholding

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

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

- By definition, $T_0 = 0$ and $T_N = G_{max} + 1 = 256$



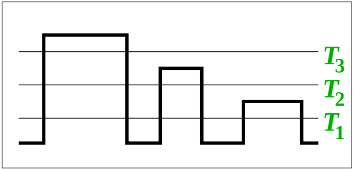
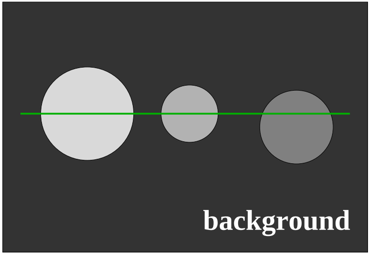


Illustration of 4-level thresholding. $T_0 = 0$ *and* $T_4 = 256$*. First level is background.*

Illustration of four-level thresholding



image

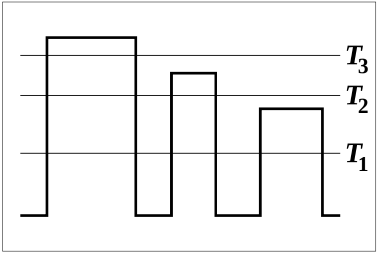
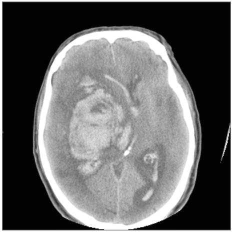


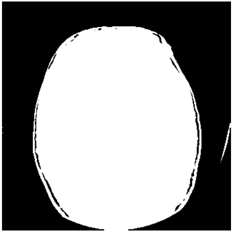
image profile along the line

- $T_0 = 0$ and $T_4 = 256$.
- The first level is dark background
- The fourth level is the brightest disc

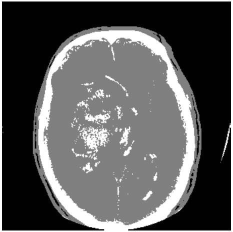
Examples of automatic thresholding into 2 and 3 levels



original image



bilevel thresholding



trilevel thresholding

- Single threshold: $N = 2$
  - **bilevel** (binary) thresholding, or **binarisation**
  - $\Rightarrow$ considered in this course
- **Multilevel** thresholding: $N > 2$
  - case $N = 3$ often called **trilevel**

Outline

1. Principles of greyvalue thresholding
   - Histogram-based thresholding
2. Two methods for threshold selection
   - Otsu's method
   - Histogram modelling by Gaussian distributions
3. Examples and analysis of thresholding
   - Examples of thresholding
   - Analysis of thresholding

Grey-level histogram

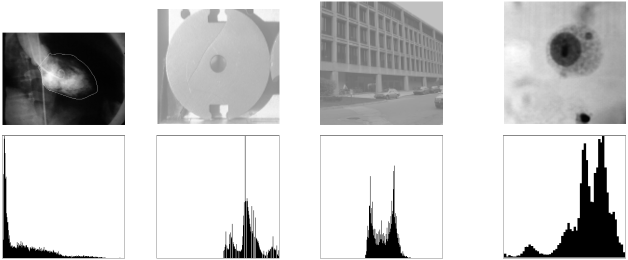
- Occurrence probability of greyvalue k in image

$$P(k) = \frac{n_k}{n}$$

  - n_k is number of pixels with greyvalue $k = 0, 1, \ldots, 255$
  - n is total number of pixels in image

  $\Rightarrow$ $P(k)$ shows how frequently k occurs in image
- Calculation simple and fast
  - initialise $p[k] = 0$
  - scan image, for greyvalue k set $p[k] \leftarrow p[k] + 1$
  - after scan, normalise $P[k] = p[k]/n$

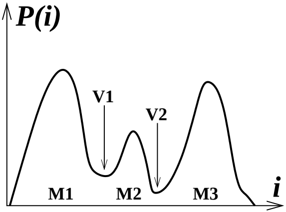
Basic types of histograms



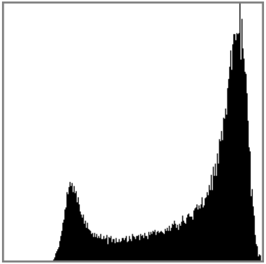
too dark | too light | low contrast | good contrast

- Grey-levels
  - concentrate at dark end: too dark
  - concentrate at light end: too light
  - concentrate in middle: narrow dynamic range
  - spread over histogram: good contrast

Good histograms to threshold



ideal trimodal

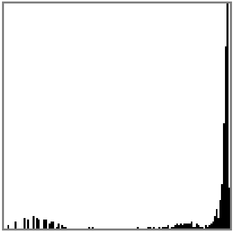


real bimodal

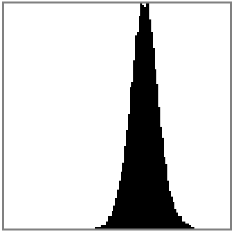
- Two or more distinct modes (M)
- Definite minima in valleys between modes (V)

⇒ Intensity classes easy to separate

Bad histograms to threshold



hard to threshold



unimodal

- Mode at limit of intensity range
  - $\Rightarrow$ histogram hard to model
- Mode not distinct
- Unimodal
  - $\Rightarrow$ hard to threshold, but not hopeless

Examples of good and bad threshold selections

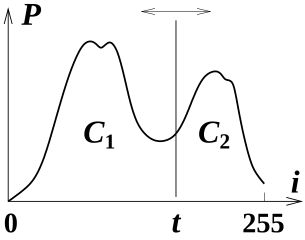
image | good 1 | good 2 | too low | too high | histogram

- Several thresholds are acceptable
  - near valley (G) in histogram
- Bad thresholds have different effects
  - too low threshold (L) tends to split lines
  - too high threshold (H) tends to merge lines

Outline

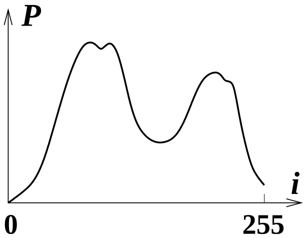
1. Principles of greyvalue thresholding
   - Histogram-based thresholding
2. Two methods for threshold selection
   - Otsu's method
   - Histogram modelling by Gaussian distributions
3. Examples and analysis of thresholding
   - Examples of thresholding
   - Analysis of thresholding

Maximal separation of classes



- Proposed by N.Otsu (Japan), 1978
- Consider a **candidate threshold** t
  - t defines two classes of grayvalues
- Define measure of **separation of classes**
  - distance between classes as function of t
- Find optimal threshold t_{opt} that **maximises separation**

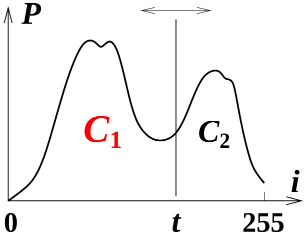
Mean and variance of entire histogram



$$\mu = \sum_{i=0}^{255} iP(i) \qquad \sigma^2 = \sum_{i=0}^{255} (i-\mu)^2 P(i)$$

- Range of entire histogram: $[0, 255]$
- $P(i)$ is normalised: $\sum_{i=0}^{255} P(i) = 1$

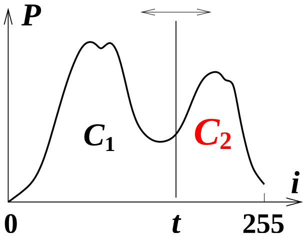
Mean and variance of class C_1



$$\mu_1(t) = \frac{1}{q_1(t)} \sum_{i=0}^{t} iP(i) \qquad \sigma_1^2(t) = \frac{1}{q_1(t)} \sum_{i=0}^{t} \left[i - \mu_1(t)\right]^2 P(i)$$

- Range of class C_1: $[0, t]$
- Weight (relative size) of C_1: $q_1(t) = \sum_{i=0}^{t} P(i)$

Mean and variance of class C_2



$$\mu_2(t) = \frac{1}{q_2(t)} \sum_{i=t+1}^{255} iP(i) \qquad \sigma_2^2(t) = \frac{1}{q_2(t)} \sum_{i=t+1}^{255} [i - \mu_2(t)]^2 P(i)$$

- Range of class C_1: $[t+1, 255]$
- Weight of C_1: $q_2(t) = \sum_{i=t+1}^{255} P(i)$; $q_2(t) = 1 - q_1(t)$

Within- and between-class variances

- For each t, total variance σ^2 has two components:
  - **within-class variance**
    - $\Rightarrow$ weighted sum of two class variances
  - **between-class variance**
    - $\Rightarrow$ distance between classes
- Within-class variance is

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

$\Rightarrow$ but $\mu = q_1(t)\mu_1(t) + q_2(t)\mu_2(t)$

- Between-class variance is the rest of σ^2

$$\sigma_B^2(t) = \sigma^2 - \sigma_W^2(t)$$

Threshold selection as optimisation problem

- It is easy to show that

$$\begin{aligned} \sigma_B^2(t) &= q_1(t) q_2(t) \left[\mu_1(t) - \mu_2(t)\right]^2 \\ &= q_1(t) \left[1 - q_1(t)\right] \left[\mu_1(t) - \mu_2(t)\right]^2 \end{aligned} \tag{1}$$

- Optimal threshold t_{opt} best separates the two classes
- $\sigma_W^2(t) + \sigma_B^2(t)$ is constant $\longrightarrow$ two equivalent options:
  - *minimise* $\sigma_W^2(t)$ as *overlap of classes*
  - *maximise* $\sigma_B^2(t)$ as *distance between classes*

$\Rightarrow$ Will use second option

Obtaining $\sigma_B^2(t)$

- $\sigma_B^2(t)$-t can be obtained **by definition**:
  - for each t, calculate the components of eq. (1): $q_1(t), \mu_1(t), \mu_2(t)$

  $\Rightarrow$ for each t, scan histogram

- But faster **recursive solution** is available:

$$
\begin{aligned}
q_1(t+1) &= q_1(t) + P(t+1), & q_1(0) &= P(0) \\
\mu_1(t+1) &= \frac{q_1(t)\mu_1(t) + (t+1)P(t+1)}{q_1(t+1)}, & \mu_1(0) &= 0 \qquad (2) \\
\mu_2(t+1) &= \frac{\mu - q_1(t+1)\mu_1(t+1)}{1 - q_1(t+1)}
\end{aligned}
$$

Summary of Otsu's algorithm

Algoritmus: Otsu threshold selection

1. Compute image histogram $P(i)$, calculate μ
2. For each $0 < t < G_{max}$, recursively compute $q_1(t)$, $\mu_1(t)$ and $\mu_2(t)$ by eq.(2) [1]
3. Calculate $\sigma_B^2(t)$ by eq.(1)
4. Select threshold as $t_{opt} = \arg\max_t \sigma_B^2(t)$

[1] Skip possible zeroes at beginning of histogram!

Properties of Otsu algorithm

- **Advantages**
  - general: no specific histogram shape assumed
  - works well, stable
  - extension to *multilevel thresholding* possible

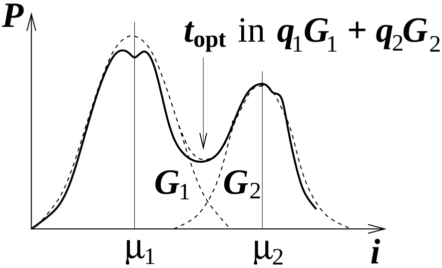
    ⇒ for N thresholds and $M = G_{max} + 1$ grey levels, maximum search in array of M^N size
- **Drawbacks**
  - assumes that $\sigma_B^2(t)$ is unimodal: not always true
  - $\sigma_B^2(t)$ is often flat, false maxima may occur
  - tends to artificially enlarge small classes
    - prefers balance of class weights, see cost function 1

    ⇒ small classes may be merged and missed

Outline

1. Principles of greyvalue thresholding
   - Histogram-based thresholding
2. Two methods for threshold selection
   - Otsu's method
   - Histogram modelling by Gaussian distributions
3. Examples and analysis of thresholding
   - Examples of thresholding
   - Analysis of thresholding

Modeling histogram by mixture of two Gaussians



- Assume histogram $P(i)$ is mixture of **two Gaussian distributions**

$$P(i) \approx G(i) \doteq q_1 G_1(i) + q_2 G_2(i)$$

- Fit this model to $P(i)$, estimate parameters of model
- Find optimal threshold **analytically** as valley in model function

Parameterised model distribution

- Model distribution is weighted sum of two Gaussians

$$G(i, \mathbf{p}) = \frac{q_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{i-\mu_1}{\sigma_1}\right)^2} + \frac{q_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{i-\mu_2}{\sigma_2}\right)^2} \qquad (3)$$

⇒ Frequently used Multiple Gaussian Model, or Gaussian Mixture

- Has six parameters: $\mathbf{p} = (q_1, q_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$
- q_1 and q_2 are weights of two Gauss distributions
  - $q_1 + q_2 = 1$

  ⇒ **five free parameters** (degrees of freedom, dof)

  ⇒ exclude q_2, denote $\mathbf{p}' = (q_1, \mu_1, \mu_2, \sigma_1, \sigma_2)$

Fitting model distribution to histogram

- To fit $f(i, \mathbf{p}')$ to $P(i)$, minimise 5-parameter error function

$$C(\mathbf{p}') = \sum_{i=0}^{G_{max}} \left[f(i, \mathbf{p}') - P(i) \right]^2 \quad (4)$$

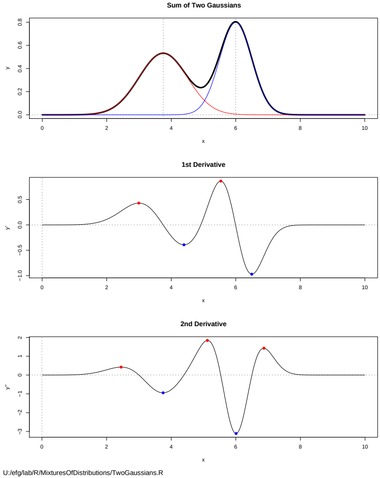
$\Rightarrow$ estimate optimal parameters $\hat{\mathbf{p}}$

- A nonlinear minimisation algorithm can be used
  - Newton
  - Marquard-Levenberg
  - stochastic
- Iterative minimisation algorithm can *fail* to give any result
  - e.g., does not converge

  $\Rightarrow$ Gaussian mixture model not applicable

  $\Rightarrow$ no solution for fitting, no threshold

Estimation of initial values of parameters



Source: Earl F. Glynn, Stowers Inst. for Med. Res.

- In case of distinct modes
  - $\Rightarrow$ moderate overlap of modes
- Mean: μ_1, μ_2
  - positions of dominant maxima in $P(i)$, **or**
  - positions of dominant valleys in $P''(i)$
- Variance: σ_1, σ_2
  - distance between maximum and valley in $P'(i)$ is 2σ
- Weight: $q_1 = 0.5$

Optimal threshold

- Assume model fitting was successful

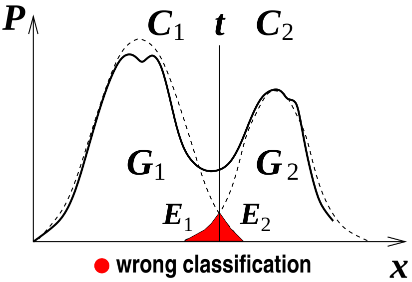
⇒ Optimal parameters were obtained:

  - $(\hat{q}_1, \hat{q}_2, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2)$

- Now, hogy to calculate optimal threshold?

Probability of wrong classification

$$E(t) = E_1(t) + E_2(t) = \int_{-\infty}^{t} G_2(x)\, dx + \int_{t}^{\infty} G_1(x)\, dx$$



- $E_1(t)$: pixel from C_1 classified as C_2
- $E_2(t)$: pixel from C_2 classified as C_1

Minimisation of classification error

- The error is minimal when

$$\frac{\partial E}{\partial t} = 0$$

- It can be proved that **optimal threshold** t_{opt} is solution of

$$At^2 + Bt + C = 0, \tag{5}$$

where

$$\begin{aligned} A &= \hat{\sigma}_1^2 - \hat{\sigma}_2^2 \\ B &= 2(\hat{\mu}_1\hat{\sigma}_2^2 - \hat{\mu}_2\hat{\sigma}_1^2), \\ C &= \hat{\sigma}_1^2\hat{\mu}_2^2 - \hat{\sigma}_2^2\hat{\mu}_1^2 + 2\hat{\sigma}_1^2\hat{\sigma}_2^2 \ln\left(\frac{\hat{\sigma}_2\hat{q}_1}{\hat{\sigma}_1\hat{q}_2}\right) \end{aligned}$$

Cases for optimal threshold

- Equation has **two real roots** $\in [0, 255]$
  $\Rightarrow$ select root for which error $E(t)$ is smaller
- Equation has **no real root** $\in [0, 255]$
  $\Rightarrow$ no optimal threshold available
- If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, single optimal threshold exists:

$$t_{opt} = \frac{\hat{\mu}_1 + \hat{\mu}_2}{2} + \frac{\hat{\sigma}^2}{\hat{\mu}_1 - \hat{\mu}_2} \ln\left(\frac{\hat{q}_1}{\hat{q}_2}\right)$$

Summary of Gaussian Mixture-based threshold selection

Algoritmus: Gaussian threshold selection

1. Calculate normalised histogram $P(i)$
2. Minimise fitting error function $C(\mathbf{p}')$ defined by (4) and (3)

   $\Rightarrow$ obtain optimal parameter estimates $\hat{q}_1, \hat{q}_2, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2$
3. Solve equation (5) for t, obtain two roots
4. Consider real roots $\in [0, 255]$ only
   - if single root exists, use it as t_{opt}
   - if two roots exist, select root with smaller $E(t)$

Properties of Gaussian mixture approach

Advantages

- Reasonably general histogram model
- When model is valid, minimises classification error probability
- May work for small-size classes, as well

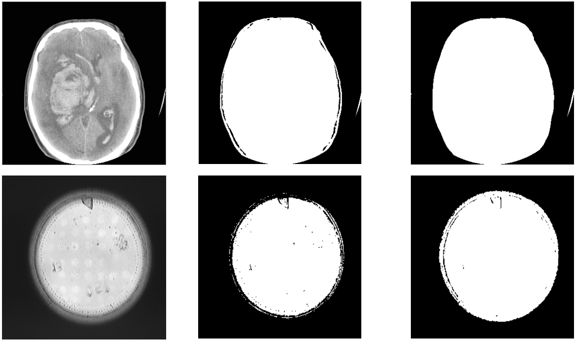
Drawbacks

- Real histograms cannot always be modelled by Gaussian mixtures
  - ⇒ greyvalues are **finite** and **non-negative**
  - ⇒ peak close to intenisity limit do not fit Gaussian
- Extension to multithresholding needs simplification
  - ⇒ e.g., assumption of well-separated modes
- Difficult to detect close or flat modes

Outline

1. Principles of greyvalue thresholding
   - Histogram-based thresholding
2. Two methods for threshold selection
   - Otsu's method
   - Histogram modelling by Gaussian distributions
3. Examples and analysis of thresholding
   - Examples of thresholding
   - Analysis of thresholding

Example 1: Good thresholding

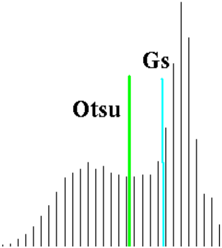


images Otsu results Gaussian results

- Gaussian algorithm sets lower thresholds in both cases
  $\Rightarrow$ fits object contours slightly better than Otsu

Example 2: Satisfactory results

image



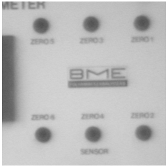
histogram

Otsu

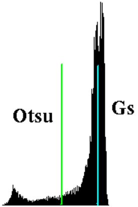
Gauss

- Otsu threshold $T = 158$: exactly in valley
  $\Rightarrow$ lines are well-separated
- Gaussian threshold $T = 199$: a bit too high
  $\Rightarrow$ some lines are not well-separated

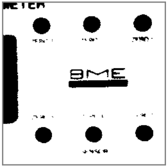
Example 3: Gaussian mixture gives poor result

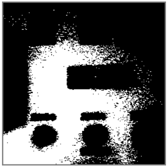


image



histogram


Otsu $T = 159$


Gauss $T = 201$

- **Otsu algorithm** finds small class of pixels (dark discs)
- **Gaussian algorithm** tries to separate two high peaks formed by background

⇒ Selects noisy valley because true class is
  - too small
  - too far away

Example 4: Gaussian mixture gives no result

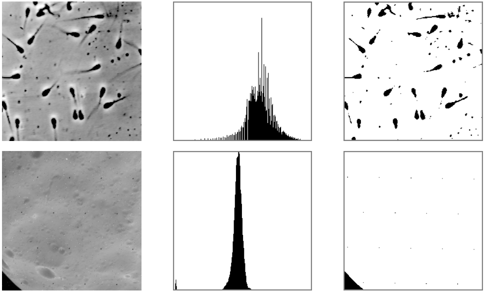


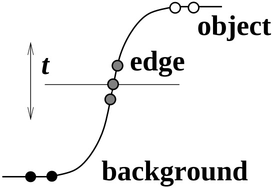
image histogram Otsu result

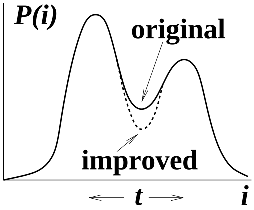
- Only Otsu algorithm produces results
- Gaussian algorithm gives **no results** at all
  - upper row: unimodal histogram, model fitting failed
  - lower row: fitting done, threshold equation has no real root

Outline

1. Principles of greyvalue thresholding
   - Histogram-based thresholding
2. Two methods for threshold selection
   - Otsu's method
   - Histogram modelling by Gaussian distributions
3. Examples and analysis of thresholding
   - Examples of thresholding
   - Analysis of thresholding

Use of gradient for better mode separation





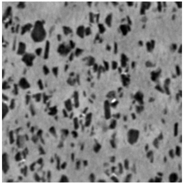
- Combine intensity and gradient data to improve histogram

⇒ Better separation of objects and background

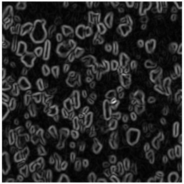
  - pixels close to edges have high gradients
  - pixels of object and background have low gradients

  ⇒ **discard high-gradient pixels** when computing histogram

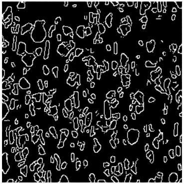
Real-data example of histogram improvement



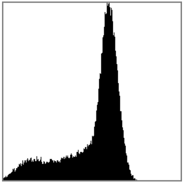
image



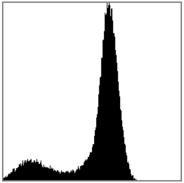
gradient



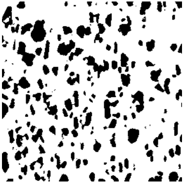
edges



original histogram

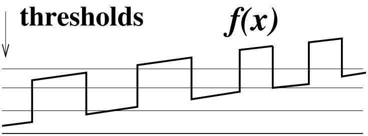


improved histogram

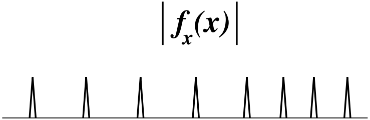


thresholding

Thresholding versus edge detection



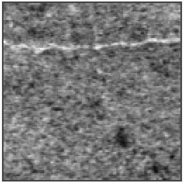
signal with trend cannot be thresholded



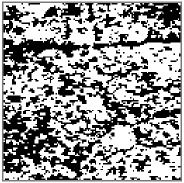
edges can still be detected

- **Thresholding** with constant threshold is non-adaptive op.
  - advantage: closed contours guaranteed
  - drawback: not applicable to images with uneven illumination
- **Edge detection** is local operation adaptive to slow variation of background intensity
  - advantage: applicable to images with uneven illumination
  - drawback: closed contours not guaranteed

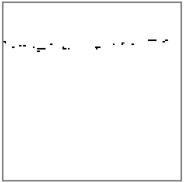
Limits of thresholding 1/2



stone crack



Otsu



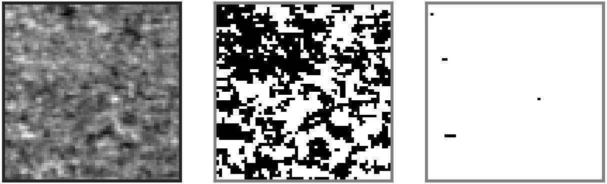
method X

- Merit of thresholding is **task-dependent**
  - merit is measure of quality
- Merit may involve **geometric properties**
  - histogram does not account for geometry

  ⇒ crack is detected as bright pixels

  ⇒ detection is independent of crack shape

Limits of thresholding 2/2



stone Otsu method X

- In this example, quality of result is **not obvious**
- Limits of thresholding
  - no structural informaton taken into account
  - same threshold for arbitrarily swapped pixels
  - ⇒ connected regions not guaranteed
- Solution
  - **region-oriented** methods that use intensity and structure