A Comparison of Nonparametric Histogram-Based Thresholding Algorithms

Presentation for 8002202 Digital Image Processing III

Antti Niemistö

antti.niemisto@tut.fi

October 27, 2004

TAMPERE UNIVERSITY OF TECHNOLOGY
Institute of Signal Processing

Goals of the presentation

- Explore different types of nonparametric histogram-based thresholding algorithms.
- Send the message that there are several different approaches and algorithms out there.
- Talk about the general principles of different algorithms without going too much into the details.
- Show how the algorithms can be implemented.
- To understand the formulae presented, you need to refer to the original papers or other literature.

Thresholding

- Thresholding converts a gray-level image into a binary one.
- The binary levels may represent objects and background.
- Pixels whose value exceeds a critical value are assigned to one category, and the rest to the other.
- Global thresholding: the same threshold is used across the whole image.
- Algorithms for automatically choosing the threshold are needed.

Histogram-based thresholding

- Histogram-based algorithms have been studied extensively.
  - simple
  - easy to implement
  - fast
- Here we study 12 different nonparametric histogram-based thresholding algorithms.
- The algorithms have been chosen such that a wide range of approaches is represented.
- Algorithms that use contextual information are outside the scope of this study.

Notation

- The histogram is denoted by $y_0, y_1, \ldots, y_n$, where y_i is the number of pixels in the image with the gray-level i, and n is the maximum gray-level n (255 in an 8-bit image).
- The threshold is denoted by t.
- The following partial sums are also used:

$$A_j = \sum_{i=0}^{j} y_i, \;\; B_j = \sum_{i=0}^{j} i y_i, \;\; C_j = \sum_{i=0}^{j} i^2 y_i,$$

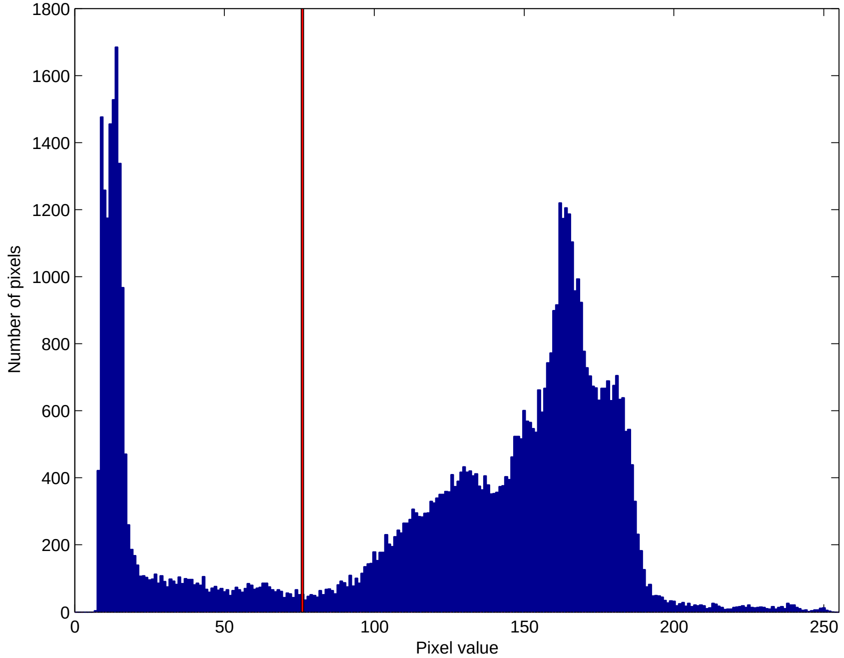
for $j = 0, \ldots, n$.

MINIMUM algorithm

J. M. S. Prewitt and M. L. Mendelsohn, "The analysis of cell images," in *Ann. New York Acad. Sci.*, vol. 128, pp. 1035-1053, 1966.

- Assumes a bimodal histogram.
- The histogram needs to be smoothed (using the three-point mean filter) iteratively until the histogram has only two local maxima.
- Choose t such that $y_{t-1} > y_t \leq y_{t+1}$.
- Unsuitable for images that have a histogram with extremely unequal peaks or a broad and flat valley.

MINIMUM result (1)



Threshold: $t = 76$

MINIMUM result (2)

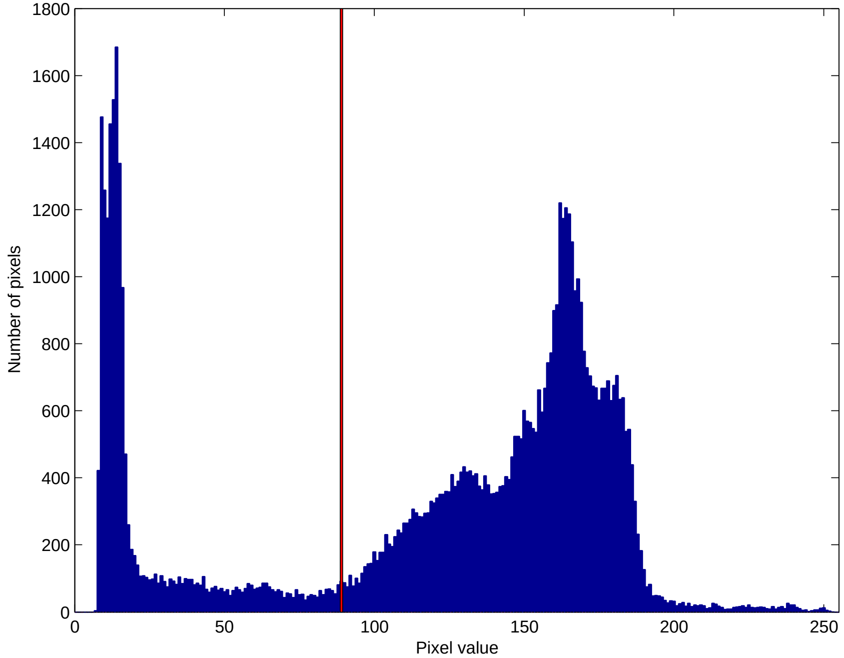
Threshold: $t = 76$

INTERMODES algorithm

J. M. S. Prewitt and M. L. Mendelsohn, "The analysis of cell images," in *Ann. New York Acad. Sci.*, vol. 128, pp. 1035-1053, 1966.

- An alternative to MINIMUM.
- Assumes a bimodal histogram.
- Find the two peaks (local maxima) y_j and y_k.
- Set t to $(j+k)/2$.
- Still unsuitable for images that have a histogram with extremely unequal peaks.

INTERMODES result (1)



Threshold: $t = 89$

INTERMODES result (2)

Threshold: $t = 89$

CONCAVITY algorithm (1)

A. Rosenfeld and P. De La Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Trans. Systems Man Cybernet.*, vol. 13, pp. 231-235, 1983.

- If the image does not have distinct objects and background, the MINIMUM and INTERMODES algorithms are not suitable.
- A good threshold may be found at the shoulder of the histogram.
- The shoulder location can be found based on the concavity of the histogram.

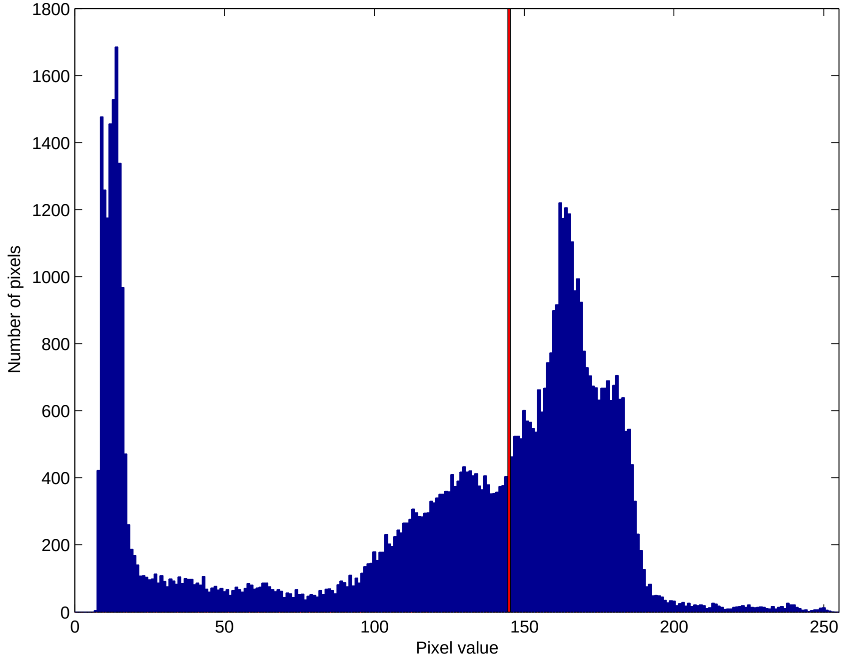
CONCAVITY algorithm (2)

- Construct the convex hull H of the histogram y.
- Find the local maxima of $H - y$.
- Set t to the value of j at which the balance measure

$$b_j = A_j(A_n - A_j)$$

  is maximized.
- The algorithm seems to work well in many cases, but in some cases it gives thresholds that are clearly unusable.

CONCAVITY result (1)



Threshold: $t = 145$

CONCAVITY result (2)

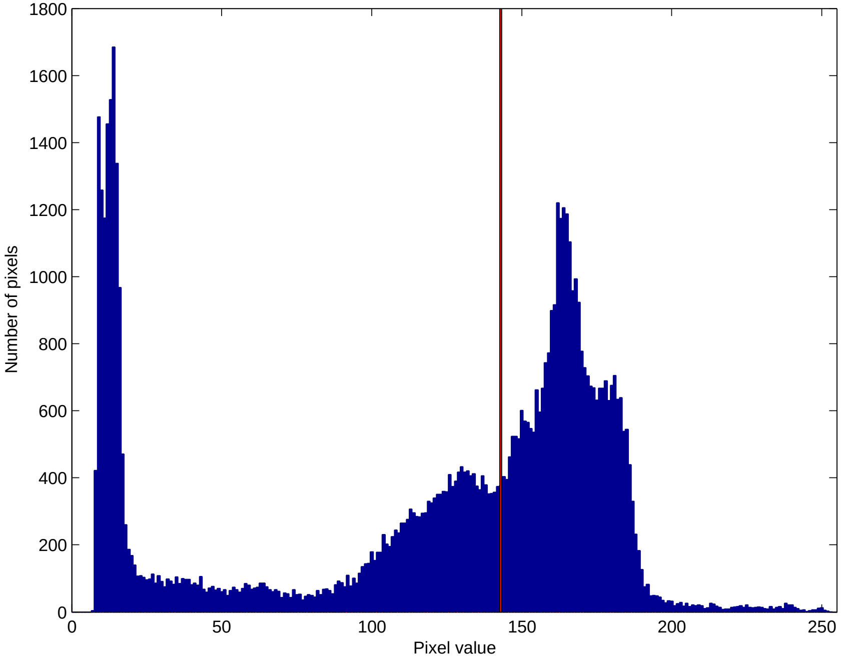
Threshold: $t = 145$

PTILE and MEDIAN algorithms

W. Doyle, "Operation useful for similarity-invariant pattern recognition," *J. Assoc. Comput. Mach*, vol. 9, pp. 259-267, 1962.

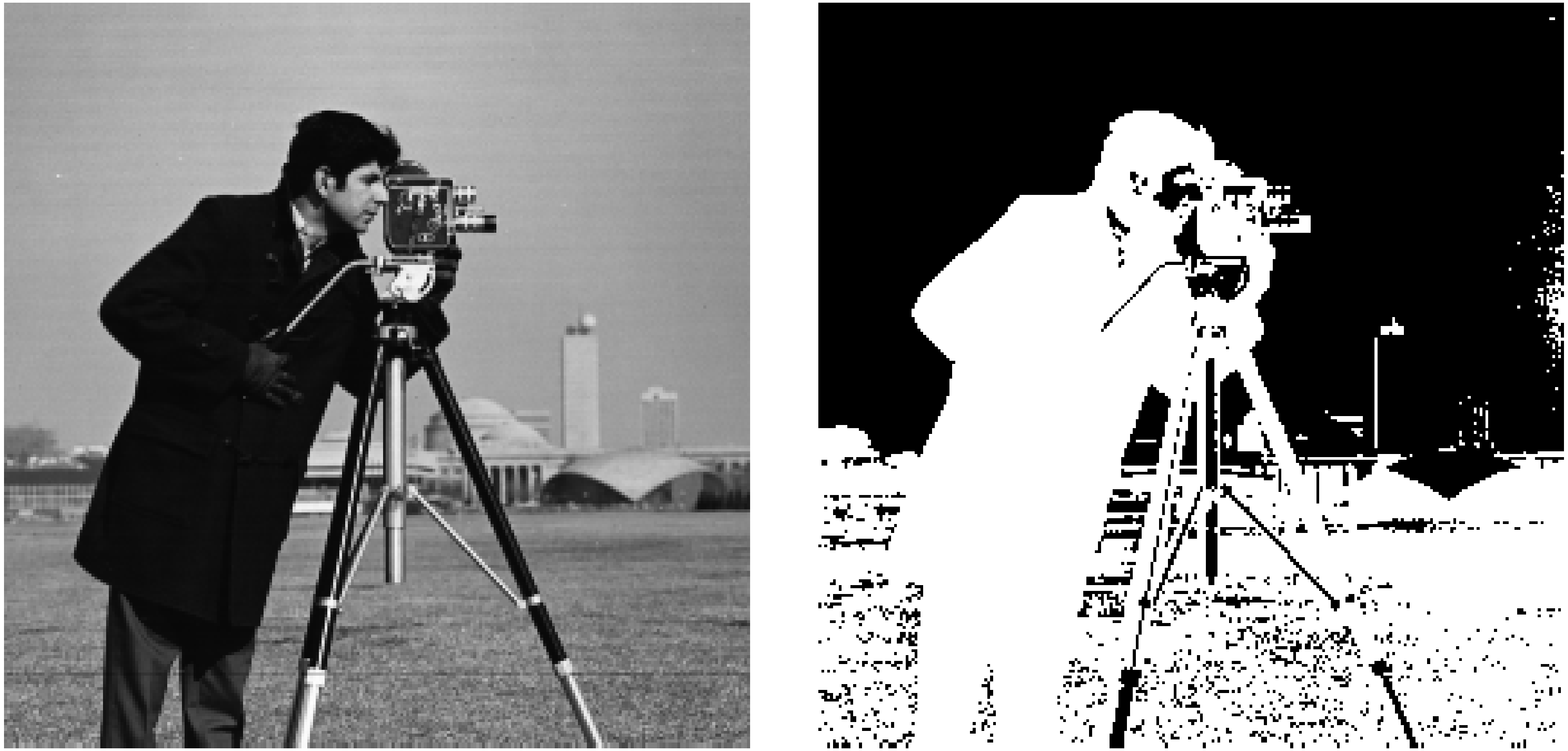
- Assumes that the percentage of object pixels is known.
- Set t to the highest gray-level which maps at least $(100 - p)\%$ of the pixels into the object category.
- Not suitable if the object area is not know.
- Problem: the algorithm is parametric.
- Solution: set $p = 50$ so that t is the median of the distribution of pixel values.

MEDIAN result (2)



Threshold: $t = 143$

MEDIAN result (2)

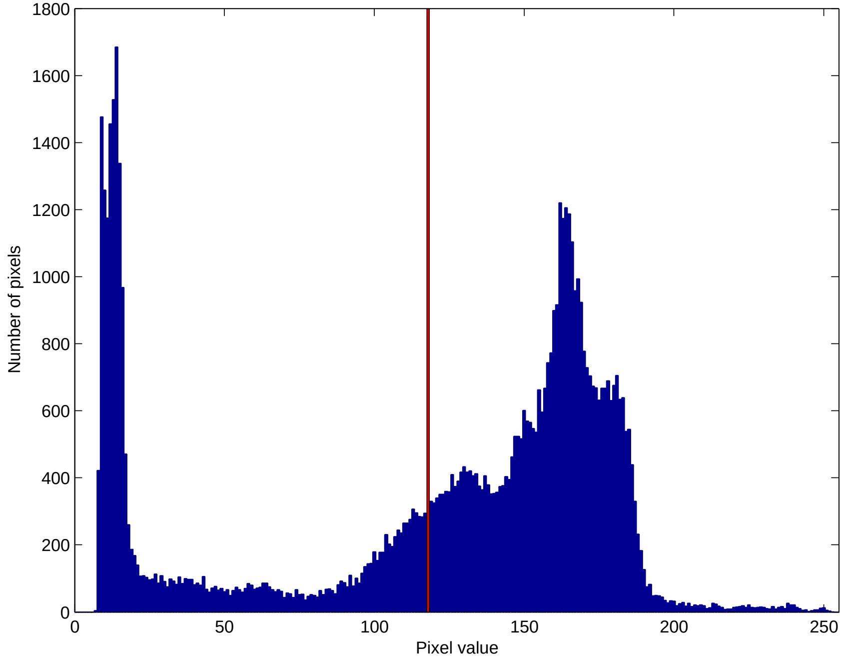


Threshold: $t = 143$

MEAN algorithm

- Similar to the MEDIAN algorithm.
- Instead of median, set t such that it is the integer part of the mean of all pixel values.
- With the partial sum notation, $t = \lfloor B_n / A_n \rfloor$.
- Does not take into account histogram shape, so obviously the results are suboptimal.

MEAN result (1)



Threshold: $t = 118$

MEAN result (2)

Threshold: $t = 118$

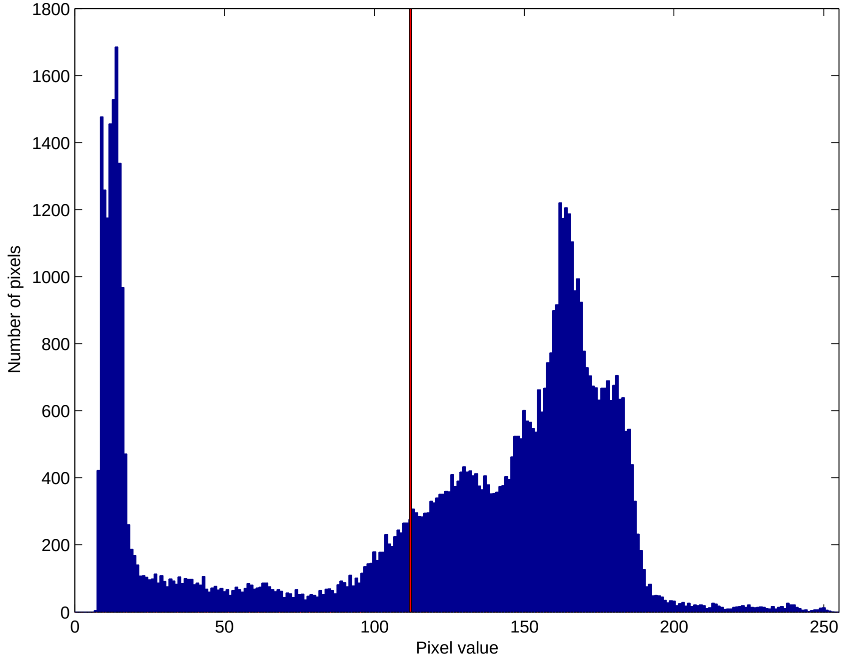
MOMENTS algorithm

W. Tsai, "Moment-preserving thresholding: a new approach," *Comput. Vision Graphics Image Process.*, vol. 29, pp. 377-393, 1985.

- Choose t such that the binary image has the same first three moments as the gray-level image.
- This is achieved by setting t such that A_t/A_n is the value of the fraction nearest to x_0, where

$$
\begin{aligned}
x_0 &= \frac{1}{2} - \frac{B_n/A_n + x_2/2}{\sqrt{x_2^2 - 4x_1}}, \quad x_1 = \frac{B_n D_n - C_n^2}{A_n C_n - B_n\, 2}, \\
x_2 &= \frac{B_n C_n - A_n D_n}{A_n C_n - B_n^2}, \qquad D_n = \sum_{i=0}^{n} i^2 y_i.
\end{aligned}
$$

MOMENTS result (1)



Threshold: $t = 112$

MOMENTS result (2)

Threshold: $t = 112$

ENTROPY algorithm (1)

J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, “A new method for gray-level picture thresholding using the entropy of the histogram,” *Comput. Vision Graphics Image Process.*, vol. 29, pp. 273-285, 1985.

- One of several maximum entropy algorithms.
- Divides the histogram of the image into two probability distributions, one representing the objects and one representing the background.
- Choose t such that the sum of the entropies of these probability distributions is maximized.

ENTROPY algorithm (2)

- Define the partial sums

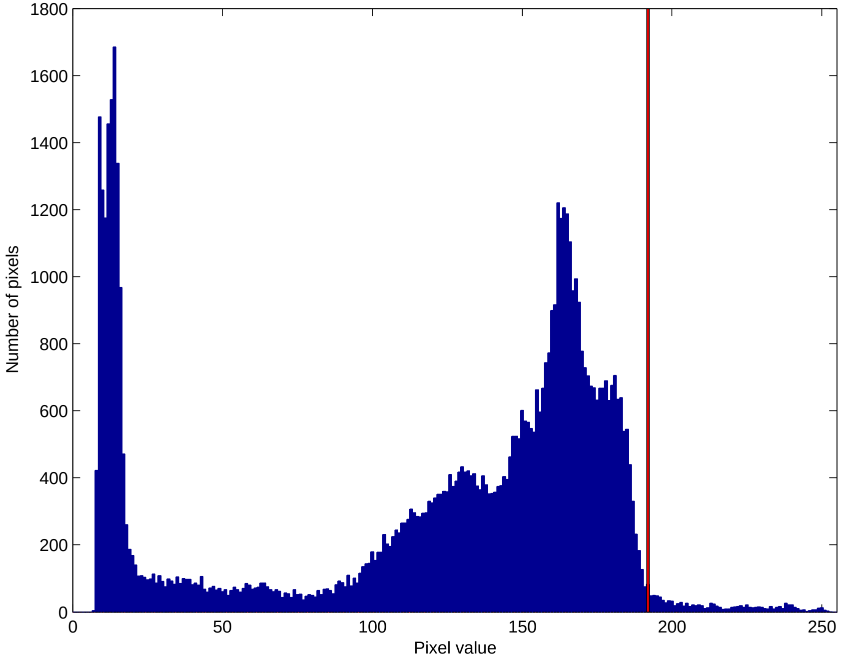
$$E_j = \sum_{i=0}^{j} y_i \log y_i, \quad \text{for } j = 0, \ldots, n.$$

- Set t to the value of j at which

$$\frac{E_j}{A_j} - \log A_j + \frac{E_n - E_j}{A_n - A_j} - \log A_n - A_j$$

is maximized.

ENTROPY result (1)



Threshold: $t = 192$

ENTROPY result (2)

Threshold: $t = 192$

INTERMEANS algorithm (1)

N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Systems Man Cybernet.*, vol. 9, pp. 62-66, 1979.

- Choose t such that the between-class variance is maximized and the intra-class variance is minimized.
- The algorithm positions t midway between the means of the two classes.
- Widely used in many applications.
- Available in MATLAB with the `graythresh` function.

INTERMEANS algorithm (2)

- Define the means of the gray levels in the two classes

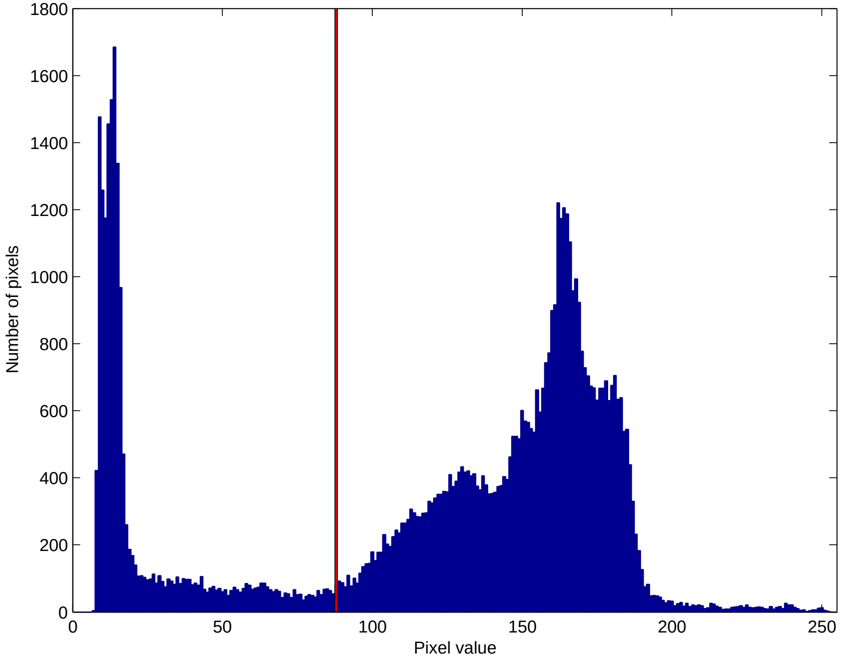
$$\mu_t = \frac{B_t}{A_t}, \quad \nu_t = \frac{B_n - B_t}{A_n - A_t}.$$

- Set t to the value of j at which

$$A_j(A_n - A_j)(\mu_j - \nu_j)^2$$

is maximized.

INTERMEANS result (1)



Threshold: $t = 88$

INTERMEANS result (2)

Threshold: $t = 88$

INTERMEANS(I) algorithm (1)

T. Ridler and S. Calvard, “Picture thresholding using an iterative selection method,” *IEEE Trans. Systems Man Cybernet.*, vol. 8, pp. 630-632, 1978.

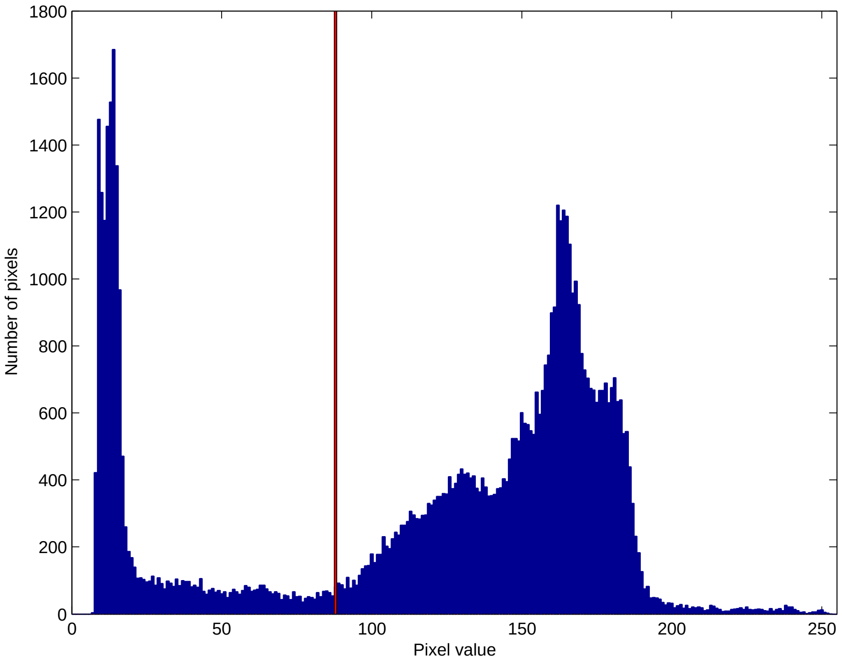
H. J. Trussell, “Comments on ’Picture thresholding using an iterative selection method’,” *IEEE Trans. Systems Man Cybernet.*, vol. 9, p. 311, 1979.

- An iterative algorithm that gives similar results as the INTERMEANS algorithm.
- Computationally less intensive than INTERMEANS.
- The algorithm starts with an initial guess for t.

INTERMEANS(I) algorithm (2)

- Define the means μ_t and ν_t of the two classes.
- Set $t = \lfloor (mu_t + \nu_t)/2 \rfloor$ and recalculate μ_t and ν_t.
- Repeat until t has the same value in two consecutive iterations.
- The obtained t may strongly depend on its initial value.
- If the objects and background occupy comparable areas, use MEAN.
- If the objects are small compared to the background, use INTERMODES.

INTERMEANS(I) result (1)



Threshold: $t = 88$

INTERMEANS(I) result (2)

Threshold: $t = 88$

MINERROR algorithm (1)

J. Kittler and J. Illingworth, “Minimum error thresholding,” *Pattern Recognition*, vol. 19, pp. 41-47, 1986.

- Similar to the INTERMEANS algorithm.
- Views the histogram as an estimate of the probability density function of the mixture population (objects and background).
- Assumes a Gaussian mixture model, that is,
  - the pixels in the two categories come from a normal distribution and
  - the normal distributions may have different means as well as variances.

MINERROR algorithm (2)

- Define the following statistics:

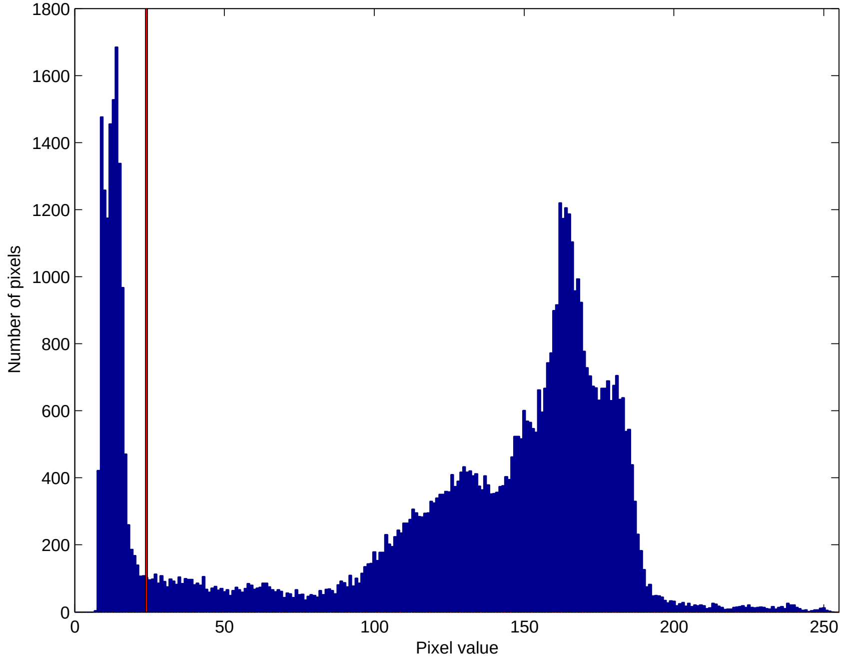
$$\begin{aligned} p_t &= \frac{A_t}{A_n}, & q_t &= \frac{A_n - A_t}{A_n}, \\ \sigma_t^2 &= \frac{C_t}{A_t} - \mu_t^2, & \tau_t^2 &= \frac{C_n - C_t}{A_n - A_t} - \nu_t^2. \end{aligned}$$

- Set t to the value of j at which

$$p_j \log\left(\frac{\sigma_j}{p_j}\right) + q_j \log\left(\frac{\tau_j}{q_j}\right)$$

is minimized.

MINERROR result (2)



Threshold: $t = 24$

MINERROR result (2)

Threshold: $t = 24$

MINERROR(I) algorithm (1)

J. Kittler and J. Illingworth, “Minimum error thresholding,” Pattern Recognition, vol. 19, pp. 41-47, 1986.

- The iterative version of the MINERROR algorithm is computationally less intensive.
- Find initial value for t using MEAN.
- The integer part of the larger solution of

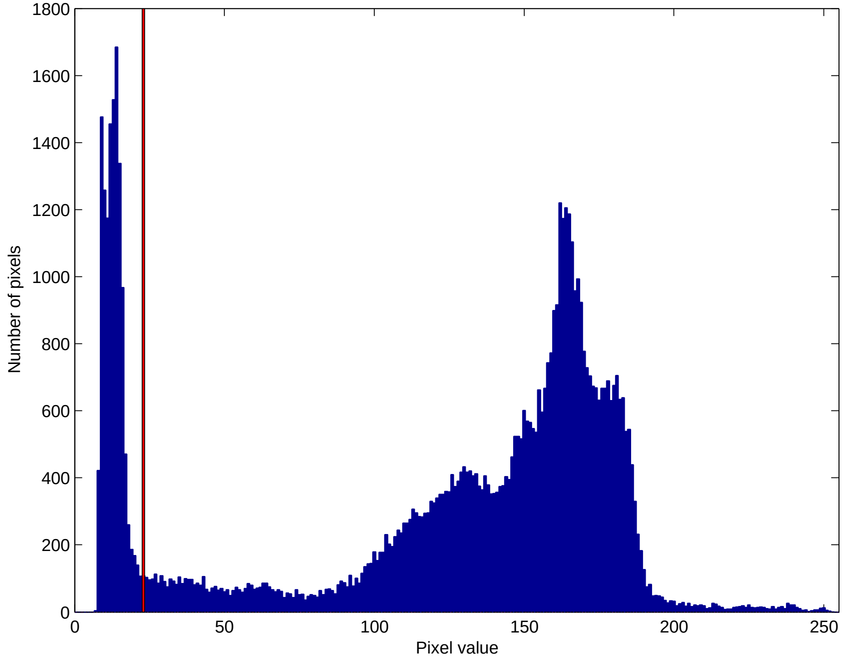
$$x^2 \left\{ \frac{1}{\sigma^2} - \frac{1}{\tau^2} \right\} - 2x \left\{ \frac{\mu}{\sigma^2} - \frac{\nu}{\tau^2} \right\} + \left\{ \frac{\mu^2}{\sigma^2} - \frac{\nu^2}{\tau^2} + \log\left(\frac{\sigma^2 q^2}{\tau^2 p^2} \right) \right\} = 0.$$

  provides a new value for t.
- Let w_0, w_1 and w_2 denote the three terms.

MINERROR(I) algorithm (2)

- Set $t = \lfloor (w_1 + \sqrt{(w_1^2 - w_0 w_2)} / w_0 \rfloor$.
- Recalculate all the terms using the new value of t and re-derive t.
- Repeat until convergence.
- This minimizes the number of misclassifications between the two normal distributions with the given means, variances, and proportions.
- The algorithm fails to converge if the quadratic equation does not have a real solution.

MINERROR(I) result (2)



Threshold: $t = 23$

MINERROR(I) result (2)

Threshold: $t = 23$

MAXLIK algorithm (1)

A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” J. Royal Statist. Soc. Series B, vol. 39, pp. 1-38, 1977.

- The problem with the INTERMEANS algorithm is that the estimators for the statistics are biased, because they do not allow overlapping distributions.
- The MAXLIK algorithm takes an expectation maximization approach for fitting mixtures of distributions.
- Initial estimates of the statistics are obtained from MINERROR(I).

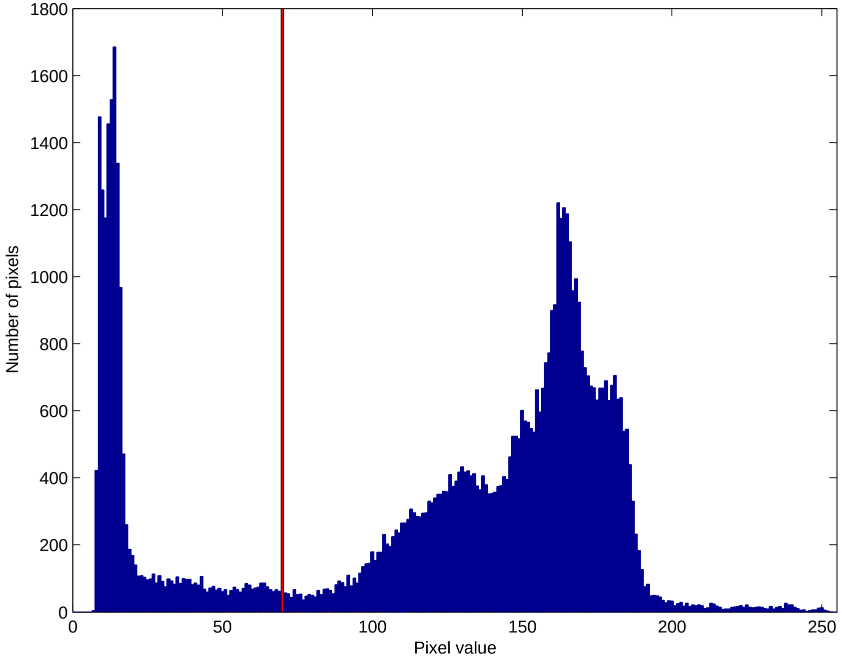
MAXLIK algorithm (2)

- The statistics are updated iteratively using:

$$
\begin{aligned}
\phi_i &= \frac{p}{\sigma}\exp\left[-\frac{(i-\mu)^2}{2\sigma^2}\right] \Big/ \left(\frac{p}{\sigma}\exp\left[-\frac{(i-\mu)^2}{2\sigma^2}\right] + \frac{q}{\tau}\exp\left[-\frac{(i-\nu)^2}{2\tau^2}\right]\right), \\
\gamma_i &= 1-\phi_i, \\
F &= \sum_{i=0}^{n}\phi_i y_i, \qquad G = \sum_{i=0}^{n}\gamma_i y_i, \\
p &= F/A_n, \qquad q = G/A_n, \\
\mu &= \sum_{i=0}^{n} i\phi_i y_i/F, \qquad \nu = \sum_{i=0}^{n} i\gamma_i y_i/G, \\
\sigma^2 &= \sum_{i=0}^{n} i^2\phi_i y_i/F-\mu^2, \quad \tau^2 = \sum_{i=0}^{n} i^2\gamma_i y_i/G-\nu^2.
\end{aligned}
$$

- After convergence t is determined as in INTERMEANS(I).

MAXLIK result (2)



Threshold: $t = 70$

MAXLIK result (2)

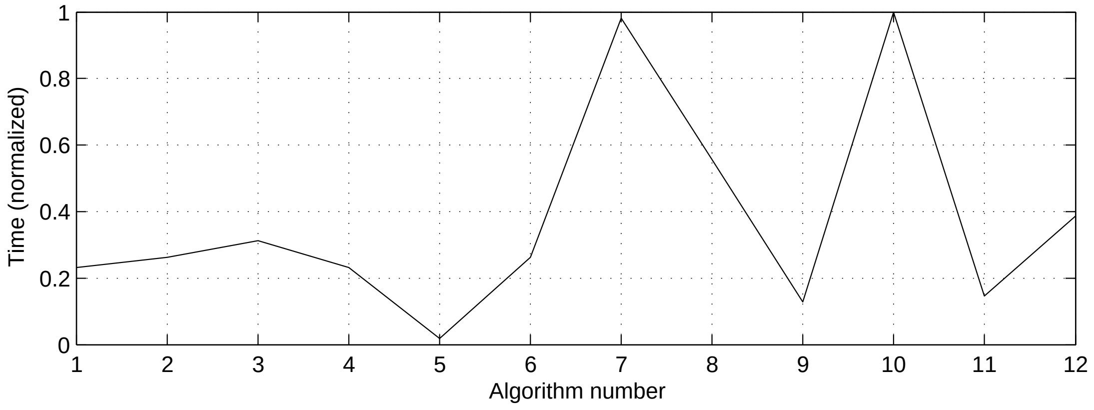
Threshold: $t = 70$

Comparison

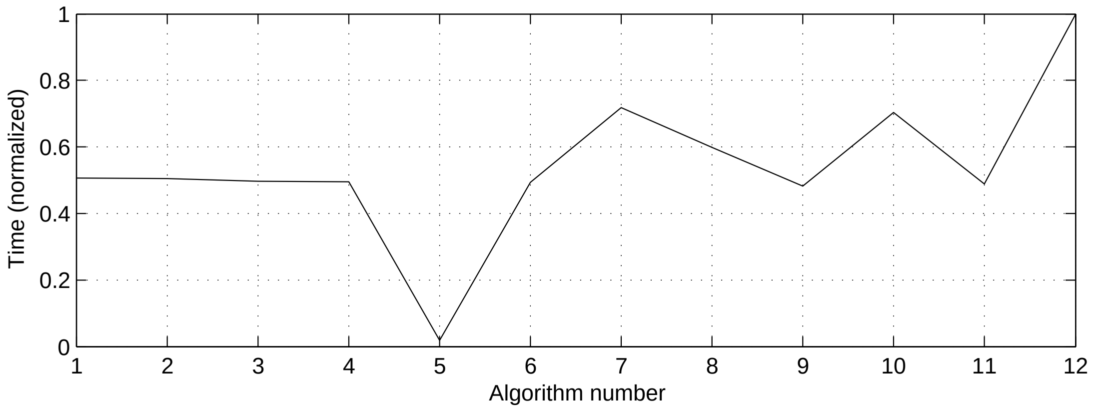
- Glasbey (1993) compared the presented algorithms under the assumption of a Gaussian mixture model (GMM).
- Not surprisingly, the algorithms that assume a GMM were found to provide the best performance.
- INTERMEANS was found to be better than INTERMEANS(I).
- ENTROPY and MOMENTS were found to sometimes give very poor results.

Running times

Small images (256 × 256)



Large images (1024 × 1024)



References

C. A. Glasbey, “An analysis of histogram-based thresholding algorithms,” *CVGIP: Graphical Models and Image Processing*, vol. 55, pp. 532-537, 1993.

- formulae for all algorithms except CONCAVITY
- analysis in the case of two normally distributed classes

P. K. Sahoo, S. Soltani, and A. K. C. Wong, “A survey of thresholding techniques,” *Computer Vision, Graphics, and Image Processing*, vol. 41, pp. 233-260, 1988.

- some of the algorithms presented here
- many other algorithms