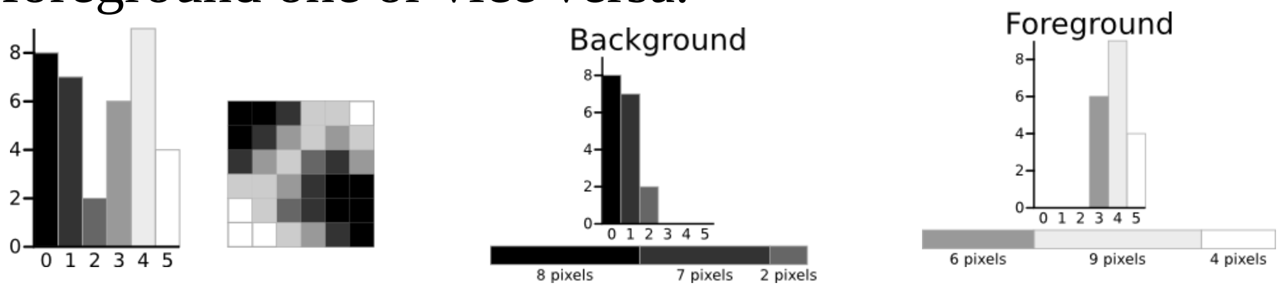
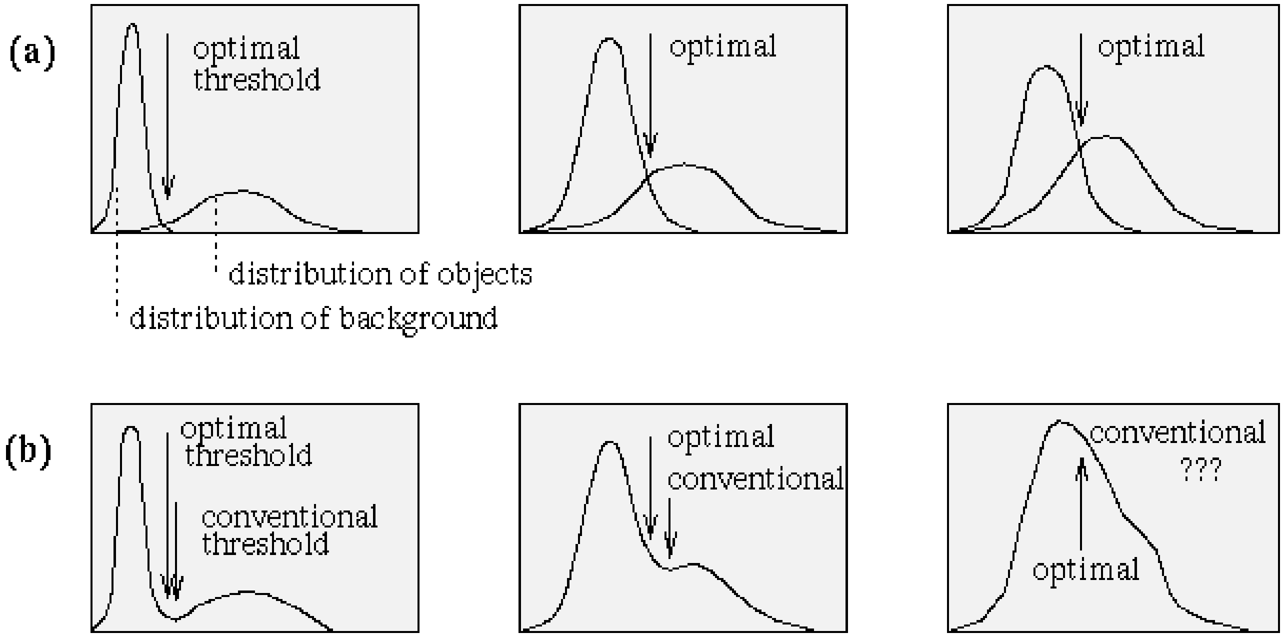
The Otsu Method – one threshold

- Problem statement: we have two groups of pixels, one with one range of values and one with another. Thresholding is difficult because these ranges usually overlap.
- Idea: minimize the error of classifying a background pixel as a foreground one or vice versa.



- Minimize the area under the histogram for one region that lies on the other region’s side of the threshold. Consider the values in the two regions as two *clusters*.
- Set the threshold so as to try to make each cluster as tight as possible, thus (hopefully!) minimizing their overlap.
- **A measure of group homogeneity is variance.** A group with high homogeneity will have low variance. A group with low homogeneity will have high variance.

Adaptive thresholding



Gray level histograms approximated by two normal distributions; the threshold is set to give minimum probability of segmentation error.

(a) Probability distributions of background and objects

(b) Corresponding histograms and optimal threshold

Otsu’s Thresholding Method (1979)

- Find the threshold that ***minimizes the weighted within-class variance*** which turns out to be the same as ***maximizing the between-class variance***.
- Operates directly on the gray level histogram [*e.g.* 256 numbers, P(i)], so it’s fast (once the histogram is computed).
- Histogram (and the image) are *bimodal*.
- Assumes uniform illumination (implicitly), so the bimodal brightness behavior arises from object appearance differences only.

Otsu's Thresholding Method (2)

- The ***weighted within-class variance*** is: $\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)$
- The ***class probabilities are estimated*** as: $q_1(t) = \sum_{i=1}^{t} P(i) \qquad q_2(t) = \sum_{i=t+1}^{I} P(i)$
  - Where $P(i) = \frac{H(i)}{I}$, H(i) the i-th entry in the histogram

    $q_2 = 1 - q_1$
- And ***the class means are given*** by: $\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)} \qquad \mu_2(t) = \sum_{i=t+1}^{I} \frac{iP(i)}{q_2(t)}$
- Finally, ***the individual class variances*** are:

$$\sigma_1^2(t) = \sum_{i=1}^{t} [i - \mu_1(t)]^2 \frac{P(i)}{q_1(t)} \qquad \sigma_2^2(t) = \sum_{i=t+1}^{I} [i - \mu_2(t)]^2 \frac{P(i)}{q_2(t)}$$

- Now, we could actually stop here. All we need to do is just run through the full range of t values [1, 256] and pick the value that **minimizes the *within class variance***.
- But the relationship between the within-class and between-class variances can be exploited to generate a recursion relation that permits a much faster calculation.

Technical University of Cluj Napoca
Computer Science Department

Otsu's Thresholding Method (3)

- The relationship between the total variance and the within group variance can make the calculation of the best threshold less computationally complex.

$$\sigma^2 = \sum_{i=1}^{I}\left[(i-\mu)^2 P(i)\right] \qquad \mu = \sum_{i=1}^{I} iP(i)$$

$$\sigma^2 = \sum_{i=1}^{t}[i-\mu_1(t)+\mu_1(t)-\mu]^2 P(i) + \sum_{i=t+1}^{I}[i-\mu_2(t)+\mu_2(t)-\mu]^2 P(i)$$

$$= \sum_{i=1}^{t}\{[i-\mu_1(t)]^2 + 2[i-\mu_1(t)][\mu_1(t)-\mu] + [\mu_1(t)-\mu]^2\}P(i)$$

$$+ \sum_{i=t+1}^{I}\{[i-\mu_2(t)]^2 + 2[i-\mu_2(t)][\mu_2(t)-\mu] + [\mu_2(t)-\mu]^2\}P(i)$$

But: $\sum_{i=1}^{t}[i-\mu_1(t)][\mu_1(t)-\mu]P(i) = 0 \qquad \sum_{i=t+1}^{I}[i-\mu_2(t)][\mu_2(t)-\mu]P(i) = 0$

$$\Rightarrow \sigma^2 = \sum_{i=1}^{t}[i-\mu_1(t)]^2 P(i) + [\mu_1(t)-\mu]^2 q_1(t) + \sum_{i=t+1}^{I}[i-\mu_2(t)]^2 P(i) + [\mu_2(t)-\mu]^2 q_2(t)$$

$$\sigma^2 = [q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)] + \{q_1(t)[\mu_1(t)-\mu]^2 + q_2(t)[\mu_2(t)-\mu]^2\}$$

$$\mu = q_1(t)\mu_1(t) + q_2(t)\mu_2(t) \qquad 1 - q_1(t) = q_2(t)$$

$$\Rightarrow \sigma^2 = \sigma_w^2(t) + q_1(t)[1-q_1(t)][\mu_1(t)-\mu_2(t)]^2$$

Otsu's Thresholding Method (4)

- ***Between/Within/Total Variance***: For any given threshold, the *total variance* is the sum of the ***within-class variances*** (weighted) and the ***between class variance***, which is the sum of weighted squared distances between the class means.

- As shown in the previous slide, we can express the total variance:

$$\sigma^2 = \underbrace{\sigma_w^2(t)}_{\textbf{Within-class}} + \underbrace{q_1(t)[1-q_1(t)][\mu_1(t)-\mu_2(t)]^2}_{\textbf{Between-class}},$$

- Since the total is constant and independent of *t*, the effect of changing the threshold is simply to move the contributions of the two terms back and forth.

- So, ***minimizing the within-class variance is the same as maximizing the between-class variance.***

Otsu’s Thresholding Method (5)

- For each potential threshold, T:
  1. Separate the pixels into two clusters according to the threshold
  2. Find the mean of each cluster
  3. Square the difference between the means
  4. Multiply by the number of pixels in one cluster times the number in the other.
- The optimal threshold is the one that maximizes the between-class variance (or, conversely, minimizes the within-class variance).
- The nice thing about this is that we can compute the quantities *using a recursion relation* as we run through the range of *t* values.

Initialization... $q_1(1) = P(1) \quad ; \quad \mu_1(0) = 0$

Recursion...

$$q_1(t+1) = q_1(t) + P(t+1)$$

$$\mu_1(t+1) = \frac{q_1(t)\mu_1(t) + (t+1)P(t+1)}{q_1(t+1)}$$

$$\mu_2(t+1) = \frac{\mu - q_1(t+1)\mu_1(t+1)}{1 - q_1(t+1)}$$