# Segmentation using Bayesian Decision Theory

J. K. Aggarwal
The University of Texas at Austin
Austin, TX 78712

## Segmentation

- Segmentation of images is the separation of pixels into different categories depending upon their intensities and/or other contextual information. We will pose this problem as Background Vs. Foreground
- Segmentation process is fairly simple for black and white or gray scale images, using a "threshold" one is able separate the foreground from the background.

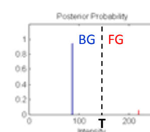


## Example



## Represent FG & BG by Single Values

- Simple thresholding can do the separation




$I(u_i,v_i) > T$



## Segmentation Contd.

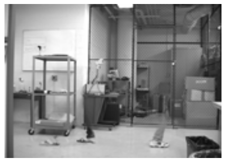
- When one is considering a sequence of images, and one is interested in separating the foreground and the background, one may use the mean at the pixel or median of the pixel to get a good estimate of the intensity at the given pixel. This process works for simple cases.



## Original Video

## Mean & Median Images

- Mean
- Median





## Segmentation cont.

- However, for more robust segmentation, one may assume that the intensity for the background and the foreground each is described by a probability density function.
- One may use Bayesian Decision Theory for separating the foreground and background.



## Prior probability

- In this B vs. F example, let $\omega$ denote the *state of nature*
  - $\omega_1$ = Background
  - $\omega_2$ = Foreground
- Prior (*a priori*) probabilities $P(\omega_1), P(\omega_2)$
  - Reflects knowledge of what the next pixel might be before the pixel appears
  - $P(\omega_1) + P(\omega_2) = 1$, assuming 2 classes



## Decision Using Only Priors

- Decide on the type of the "pixel" without being allowed to know the intensity
- Decision Rule
  - Decide $\omega_1$ if $P(\omega_1) > P(\omega_2)$
  - Decide $\omega_2$ if $P(\omega_1) < P(\omega_2)$
- This rule decides on the same class for all pixels!
- But under these conditions, no other classifier can perform better.



## Density Functions

$p_l(x \mid \omega_1)$ $\qquad p_m(x \mid \omega_2)$

- Probability density functions, where $x$ denotes the intensity, $l$ and $m$ indicate pixel position. In the most general case one can assume that each pixel has different probability density.



## A Practical Decision Scenario

- Classify based on some feature, say intensity, of the pixel samples $x$
- We will capture the variability of this feature using a continuous class-conditional probability distribution $p(x \mid \omega)$

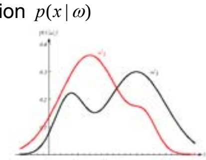


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value $x$ given the pattern is in category $\omega_i$. If $x$ represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

## Bayes Rule

- Bayes Rule states:

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j)P(\omega_j)}{p(x)}$$

  - $p(x \mid \omega_j)$ is the *likelihood* of $x$ being in class $\omega_j$
  - $P(\omega_j)$ is the *prior* probability of class $\omega_j$
  - $p(x) = \sum_{j=1}^{2} p(x \mid \omega_j)P(\omega_j)$ ensures that $P(\omega_j \mid x)$ is a valid posterior probability function that sums to one.



## Bayes Decision Rule

- Bayes decision rule:

  Select $\omega_1$ if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$

  i.e. $p(x \mid \omega_1)P(\omega_1) > p(x \mid \omega_2)P(\omega_2)$

  - If $p(x \mid \omega_1) = p(x \mid \omega_2)$ decision is based entirely on priors
  - If $P(\omega_1) = P(\omega_2)$ decision is based entirely on likelihoods
  - Bayes rule combines both to achieve minimum probability of error



## Levels of Difficulty

- One knows probability density functions and a priori probabilities.
- One estimates these probabilities from samples. One may assume normal distributions or more general forms
- You do not have a way of estimating these probabilities, you pose it as an optimization problem or a clustering problem.



## About Bayes Rule

- Bayes Rule is derived from the joint distribution

$$p(\omega_j, x) = P(\omega_j \mid x)p(x) = p(x \mid \omega_j)P(\omega_j)$$

$$\therefore P(\omega_j \mid x) = \frac{p(x \mid \omega_j)P(\omega_j)}{p(x)}$$

- In words, Bayes rule says

$$posterior = \frac{likelihood * prior}{evidence}$$



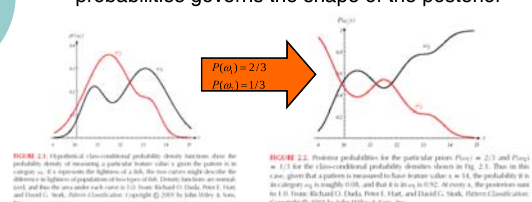
## Bayes Rule (cont.)

- Likelihood: $p(x|\omega_j)$ simply denotes that all other things being equal, the category $\omega_j$ for which $p(x|\omega_j)$ is large is more "likely" to be the category
- Evidence: $p(x)$ is simply a scale factor to ensure that $P(\omega_j|x)$ is a valid probability function.
- Bayes rule converts the prior and the likelihood to a posterior probability, which can now be used to make decisions



## Bayes Rule (cont.)

- Note that the product of likelihood and prior probabilities governs the shape of the posterior




- Decision Rule: $\omega_1$ if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ else $\omega_2$

## Error Analysis

- $P(error \mid x) = \begin{cases} P(\omega_1 \mid x) \text{ if we decide } \omega_2 \\ P(\omega_2 \mid x) \text{ if we decide } \omega_1 \end{cases}$
- Average probability of error is

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error \mid x)p(x)dx$$

- If $P(error|x)$ is as small as possible for every $x$, the above integral will be minimized
- Hence using Bayes rule

$$P(error \mid x) = \min\left[P(\omega_1 \mid x), P(\omega_2 \mid x)\right]$$



## Variations of the Bayes Rule

- Bayes decision rule:

  Select $\omega_1$ if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$

  i.e. $p(x \mid \omega_1)P(\omega_1) > p(x \mid \omega_2)P(\omega_2)$

  - If $p(x \mid \omega_1) = p(x \mid \omega_2)$ decision is based entirely on priors
  - If $P(\omega_1) = P(\omega_2)$ decision is based entirely on likelihoods
  - Bayes rule combines both to achieve minimum probability of error



## Univariate Density

- A Univariate normal density $p(x) \sim N(\mu, \sigma^2)$

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

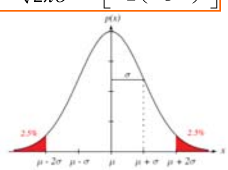


FIGURE 2.7. A univariate normal distribution has roughly 95% of its area in the range $|x - \mu| \leq 2\sigma$, as shown. The peak of the distribution has value $p(\mu) = 1/\sqrt{2\pi}\sigma$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.



## Univariate Denisty (cont.)

- Expected value $\quad \mu \equiv E[x] = \int_{-\infty}^{\infty} xp(x)dx$
  - Points tend to cluster around the mean
- Variance $\quad \sigma^2 \equiv E\left[(x-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$
  - Measure of spread of values
- Entropy $\quad H(p(x)) = -\int_{-\infty}^{\infty} p(x)\ln p(x)dx$
  - Measure of uncertainty
  - Normal has max. entropy of all distributions given mean and variance



## Multivariate Density

- *d*-dimensional normal density $p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boxed{p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]}$$

- $\boldsymbol{x}$ is a *d*-component column vector
- $\boldsymbol{\mu}$ is the *d*-component mean vector
- $\boldsymbol{\Sigma}$ is the *d*-by-*d* covariance matrix and $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\Sigma}^{-1}$ are its determinant and inverse, respectively

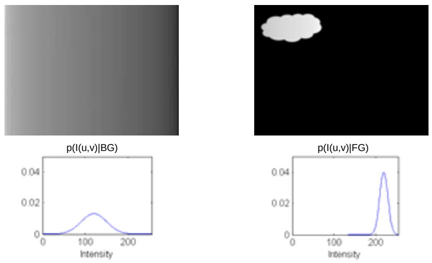


## Parameters of the multivariate normal

- Mean
  - $\boldsymbol{\mu} \equiv E[\boldsymbol{x}] = \int \boldsymbol{x}p(\boldsymbol{x})d\boldsymbol{x}$
  - Computed component-wise; i.e. $\mu_i = E[\boldsymbol{x}_i]$
- Covariance Matrix
  - $\boldsymbol{\Sigma} \equiv E\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t\right] = \int (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t p(\boldsymbol{x})d\boldsymbol{x}$
  - $\sigma_{ij} = E\left[(x_i-\mu_i)(x_j-\mu_j)\right]$
  - Always symmetric and positive semidefinite
  - $\sigma_{ii}$ is variance of $x_i$
  - $\sigma_{ij} = 0$ implies that $x_i$ and $x_j$ are statistically independent

## Represent FG & BG by Single Distributions

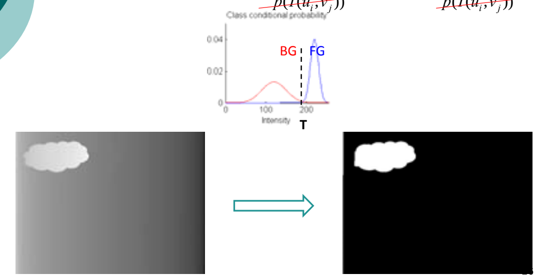
- FG and BG are generated from single 1D normal distributions
- We are able to estimate the parameters ($\mu_i$, $\sigma_i$) from the training sequences




25

## Represent FG & BG by Single Distributions

- Assume the prior probabilities are equal
- $I(u_i, v_j)$ belongs to FG if $\dfrac{p(I(u_i,v_j)\,|\,FG)\cancel{P(FG)}}{\cancel{p(I(u_i,v_j))}} > \dfrac{p(I(u_i,v_j)\,|\,BG)\cancel{P(BG)}}{\cancel{p(I(u_i,v_j))}}$




## Multivariate Normal Density (cont.)

- Loci of points of constant density are hyper-ellipsoids of the form $(\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$
- $r^2 = (\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is the Mahalanobis distance (squared) from $\boldsymbol{x}$ to $\boldsymbol{\mu}$
- Volume of hyperellipsoid

$$V = V_d\,|\boldsymbol{\Sigma}|^{1/2}\, r^d$$

$$V_d = \begin{cases} \pi^{d/2}/(d/2)! & d \text{ even} \\ 2^d \pi^{(d-1)/2}\left(\dfrac{d-1}{2}\right)!/d! & d \text{ odd} \end{cases}$$

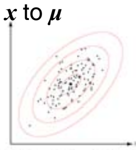
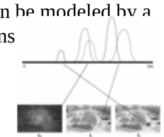


FIGURE 2.9. Samples drawn from a two-dimensional Gaussian lie in a cloud centered on the mean μ. The ellipses show lines of equal probability density of the Gaussian. From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification. Copyright © 2001 by John Wiley & Sons, Inc.

27

## The Real Situations

- In stead of the entire background, the values of a background pixel over time can be modeled by a single or a mixture of Gaussians



- Due to the motion of objects, by pixel foreground model is usually not available

28

## The Real Situations (cont'd)

- Given a controlled sequence with only background values, we can train a Gaussian for each pixel location

*Color image:*

$$p(\boldsymbol{I}(u_i,v_j)\,|\,BG) = \frac{1}{(2\pi)^{3/2}\left|\boldsymbol{\Sigma}_{ij}\right|^{1/2}} \exp[-\frac{1}{2}(\boldsymbol{I}(u_i,v_j)-\boldsymbol{\mu}_{ij})^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{I}(u_i,v_j)-\boldsymbol{\mu}_{ij})]$$

*Grayscale image:*

$$p(I(u_i,v_j)\,|\,BG) = \frac{1}{\sqrt{2\pi\sigma_{ij}^{\,2}}} \exp[-\frac{1}{2}(\frac{I(u_i,v_j)-\mu_{ij}}{\sigma_{ij}})^2]$$

- Without the knowledge of priors and foreground conditional probability, we can threshold on the

Z-value to perform background subtraction

$$z = \frac{I(u_i,v_j)-\mu_{ij}}{\sigma_{ij}}$$

29

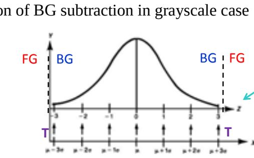
## The Real Situations (cont'd)

- In the context of color image, the Mahalanobis distance is defined as:

$$D_M(\boldsymbol{I}(u_i,v_i)) = \sqrt{(\boldsymbol{I}(u_i,v_j)-\boldsymbol{\mu}_{ij})^T \boldsymbol{\Sigma}_{ij}^{-1}(\boldsymbol{I}(u_i,v_j)-\boldsymbol{\mu}_{ij})}$$

- The Mahalanobis distance implies the probability of the test pixel value belonging to the background model

ex: Illustration of BG subtraction in grayscale case

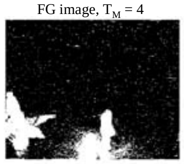



30

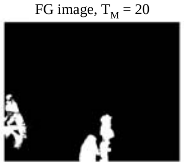
## Results

Test image

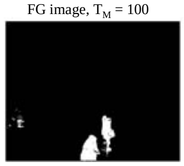
FG image, $T_M = 4$



FG image, $T_M = 20$



FG image, $T_M = 100$



## Example 1

Consider the given class conditional probability density functions:

$$p(x\,|\,\omega_1) = \frac{1}{2.5}[u(x) - u(x-2.5)]$$
$$p(x\,|\,\omega_2) = \frac{1}{2}[u(x-2) - u(x-4)]$$

where u(x) is the unit step function:

$$u(x) = 1, \quad x \geq 0$$
$$\quad = 0, \quad otherwise$$

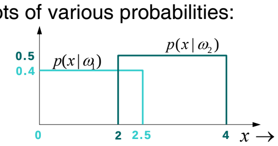
A priori probabilities:

$$P(\omega_1) = P(\omega_2) = 1/2$$

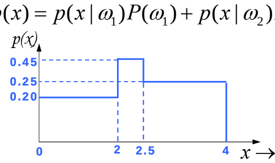


## Example 1 (contd.)

- Plots of various probabilities:



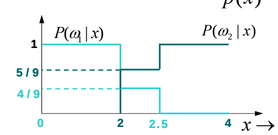

$$p(x) = p(x\,|\,\omega_1)P(\omega_1) + p(x\,|\,\omega_2)P(\omega_2)$$






## Example 1 (contd.)

Posterior Probabilities: $P(\omega_i\,|\,x) = \dfrac{p(x\,|\,\omega_i)P(\omega_i)}{p(x)}$




Observe that there will be error in classification for $2 \leq x \leq 2.5$, if the true class is $\omega_1$, because the Bayes classifier will decide $\omega_2$ since $P(\omega_2\,|\,x) > P(\omega_1\,|\,x)$ for this region



## Example 1 (contd.)

- $\therefore P(error) = P(2 \leq x \leq 2.5,\ \omega = \omega_1)$

$$= \int_2^{2.5} p(x\,|\,\omega_1)P(\omega_1)dx$$
$$= \int_2^{2.5} 0.4 \times \tfrac{1}{2}\, dx$$
$$= 0.1$$

- So the probability of error via a Bayes classifier is 0.1
- What is the a priori probability of error?

  $\because P(\omega_1) = P(\omega_2) = \tfrac{1}{2}$, it is 0.5!



## Example 2

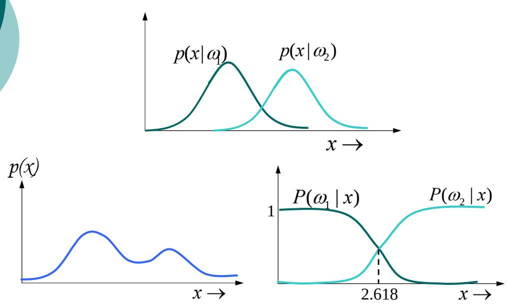
- $$p(x\,|\,\omega_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$$p(x\,|\,\omega_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

$$where \quad \mu_1 = 0 \quad \sigma_1 = 1.5$$
$$\mu_2 = 4 \quad \sigma_2 = 1.5$$

$$and \quad P(\omega_1) = \tfrac{3}{4} \quad P(\omega_2) = \tfrac{1}{4}$$

## Example 2 (Plots)






## Example 2 (contd.)

- Decision Boundary

$$P(\omega_1|x) = P(\omega_2|x)$$

$$\frac{p(x|\omega_1)P(\omega_1)}{p(x)} = \frac{p(x|\omega_2)P(\omega_2)}{p(x)}$$

$$\frac{3}{4}\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = \frac{1}{4}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

$$3e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

$$\log_e 3 - \frac{x^2}{4.5} = -\frac{(x-4)^2}{4.5}$$

$$x = 2.618$$



## Example 2 (contd.)

- Error Analysis:

$$P(error) = P(x<2.618,\ \omega=\omega_2) + P(x>2.618,\ \omega=\omega_1)$$

$$= \int_{-\infty}^{2.618} p(x|\omega_2)P(\omega_2)dx + \int_{2.618}^{\infty} p(x|\omega_1)P(\omega_1)dx$$

$$= \int_{-\infty}^{2.618} \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} . \frac{1}{4}dx + \int_{2.618}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} . \frac{3}{4}dx$$

$$= 7.4957\times 10^{-2}$$



## Example 3

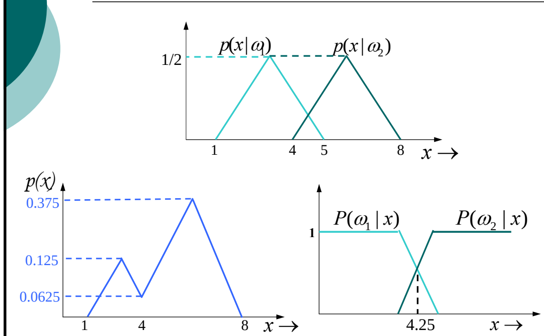
$$p(x|\omega_1) = \begin{cases} \frac{1}{4}x - \frac{1}{4} & 1 \le x \le 3 \\ -\frac{1}{4}x + \frac{5}{4} & 3 \le x \le 5 \\ 0 & otherwise \end{cases}$$

$$p(x|\omega_2) = \begin{cases} \frac{1}{4}x - 1 & 4 \le x \le 6 \\ -\frac{1}{4}x + 2 & 6 \le x \le 8 \\ 0 & otherwise \end{cases}$$

$$P(\omega_1) = \tfrac{1}{4}, \quad P(\omega_2) = \tfrac{3}{4}$$



## Example 3 (Plots)






## Example 3 (contd.)

- Decision Boundary:

$$P(\omega_1|x) = P(\omega_2|x)$$

Notice that the optimal boundary lies between $4 \le x_0 \le 5$

$$p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2)$$

$$(-\frac{1}{4}x + \frac{5}{4})\frac{1}{4} = (\frac{1}{4}x - 1)\frac{3}{4}$$

$$-\frac{1}{16}x + \frac{5}{16} = \frac{3}{16}x - \frac{3}{4}$$

$$x = 4.25$$

# Example 3 (contd.)

- Error Analysis:

$$P(error) = \int_{-\infty}^{4.25} p(x \mid \omega_2) P(\omega_2)\, dx + \int_{4.25}^{\infty} p(x \mid \omega_1) P(\omega_1)\, dx$$

$$= \int_{4}^{4.25} \left( \frac{1}{4}x - 1 \right) \frac{3}{4} dx + \int_{4.25}^{5} \left( -\frac{1}{4}x + \frac{5}{4} \right) \frac{1}{4} dx$$

$$= \frac{3}{4} \left( \frac{x^2}{8} - x \right) \Bigg|_{4}^{4.25} + \frac{1}{4} \left( -\frac{x^2}{8} + \frac{5}{4}x \right) \Bigg|_{4.25}^{5}$$

$$= 2.3438 \times 10^{-2}$$