



Parameter Estimation

Biometrics
CSE 190
Lecture 7

CSE190, Winter11

- HW 1 due thursday
- Today's lecture was on the blackboard. These slides are an alternative presentation of the material.

CSE190, Winter11

Camera's sensor

- Measured pixel intensity is a function of irradiance integrated over
 - pixel's area
 - over a range of wavelengths
 - For some time

$$I = \iiint_{t \lambda x y} E(x, y, \lambda, t) s(x, y) q(\lambda) dy dx d\lambda dt$$

CSE190, Winter 2011

Biometrics

Light at surfaces

Many effects when light strikes a surface -- could be:

- transmitted
 - Skin, glass
- reflected
 - mirror
- scattered
 - milk
- travel along the surface and leave at some other point
- absorbed
 - sweaty skin

Assume that

- surfaces don't fluoresce
 - e.g. scorpions, detergents
- surfaces don't emit light (i.e. are cool)
- all the light leaving a point is due to that arriving at that point

CSE190, Winter 2011

Biometrics

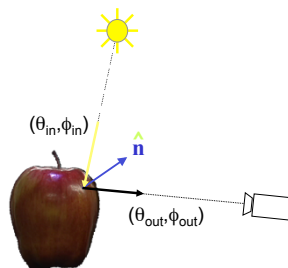
BRDF

- Bi-directional Reflectance Distribution Function

$$\rho(\theta_{in}, \phi_{in}; \theta_{out}, \phi_{out})$$
- Function of
 - Incoming light direction:

$$\theta_{in}, \phi_{in}$$
 - Outgoing light direction:

$$\theta_{out}, \phi_{out}$$
- Ratio of incident irradiance to emitted radiance



CSE190, Winter 2011

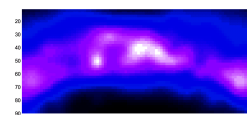
Biometrics

Surface Reflectance Models

Common Models

- Lambertian
- Phong
- Physics-based
 - Specular [Blinn 1977], [Cook-Torrance 1982], [Ward 1992]
 - Diffuse [Hanrahan, Kreuger 1993]
 - Generalized Lambertian [Oren, Nayar 1995]
 - Thoroughly Pitted Surfaces [Koenderink et al 1999]
- Phenomenological [Koenderink, Van Doorn 1996]

Arbitrary Reflectance

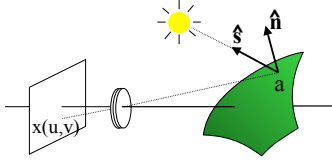


- Non-parametric model
- Anisotropic
- Non-uniform over surface
- BRDF Measurement [Dana et al, 1999], [Marschner]

CSE190, Winter 2011

Biometrics

Lambertian Surface



At image location (u,v) , the intensity of a pixel $x(u,v)$ is:

$$x(u,v) = [a(u,v) \hat{n}(u,v)] \cdot [s_0 \hat{s}] \\ = b(u,v) \cdot s$$

where

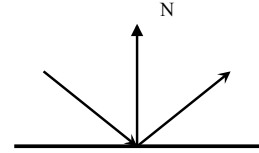
- $a(u,v)$ is the albedo of the surface projecting to (u,v) .
- $\hat{n}(u,v)$ is the direction of the surface normal.
- s_0 is the light source intensity.

$\hat{s}$ is the direction to the light source.

CSE190, Winter 2011

Biometrics

Specular Reflection: Smooth Surface

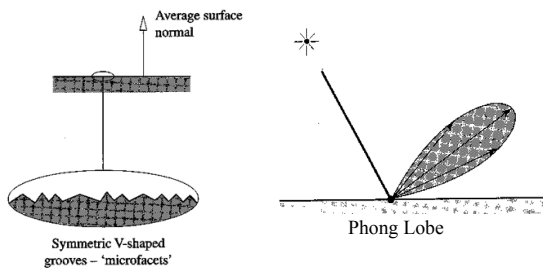


Phong – rough, specular

CSE190, Winter 2011

Biometrics

Rough Specular Surface



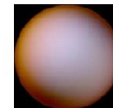
CSE190, Winter 2011

Biometrics

Phong Model



Mirror



Diffuse



CS348B Lecture 10

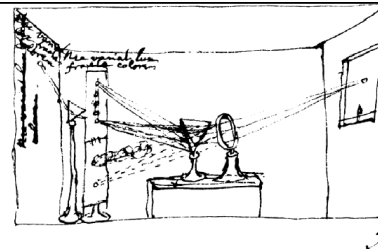
Pat Hanrahan, Spring 2002

Illumination

- Point Sources
- Point Sources at Infinity
- Distributed Sources
- Distributed distant sources
- Getting fancy (spherical harmonics).
- Color of illuminant

CSE190, Winter 2011

Biometrics



4.1 NEWTON'S SUMMARY DRAWING of his experiments with light. Using a point source of light and a prism, Newton separated sunlight into its fundamental components. By reconverging the rays, he also showed that the decomposition is reversible.

From Foundations of Vision, Brian Wandell, 1995, via B. Freeman slides



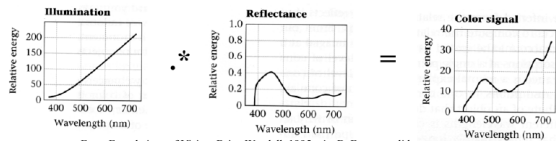
CSE190, Winter 2011

Biometrics

Color Reflectance



Measured color spectrum is a function of the spectrum of the illumination and reflectance

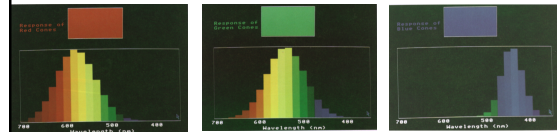


From Foundations of Vision, Brian Wandell, 1995, via B. Freeman slides

CSE190, Winter 2011

Biometrics

Color receptors



"Red" cone

"Green" cone

"Blue" cone

Response of k'th cone =
$$\int \rho_k(\lambda) E(\lambda) d\lambda$$

CSE190, Winter 2011

Biometrics

Color Cameras

Eye:

Three types of Cones

Cameras:

1. Prism (with 3 sensors)
 2. Filter mosaic
 3. Filter wheel
- ... and X3

CSE190, Winter 2011

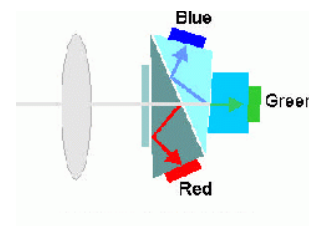
Biometrics

Prism color camera

Separate light in 3 beams using dichroic prism

Requires 3 sensors & precise alignment

Good color separation

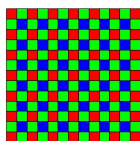
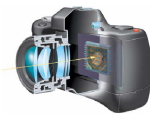


CSE190, Winter 2011

Biometrics

Filter mosaic

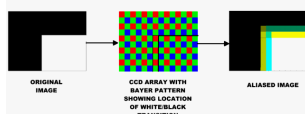
Coat filter directly on sensor



Bayer filter



Demosaicing (obtain full colour & full resolution image)



CSE190, Winter 2011

Biometrics

Chapter 3: Maximum-Likelihood & Bayesian Parameter Estimation (part 1)

- Introduction
- Maximum-Likelihood Estimation
 - Example of a Specific Case
 - The Gaussian Case: unknown μ and σ
 - Bias
-

● Introduction

- Data availability in a Bayesian framework
 - We could design an optimal classifier if we knew:
 - $P(\omega_i)$ (priors)
 - $P(x | \omega_i)$ (class-conditional densities)
- Unfortunately, we rarely have this complete information!

- Design a classifier from a training sample
 - No problem with prior estimation
 - Samples are often too small for class-conditional estimation (large dimension of feature space!)

Pattern Classification, Chapter 3

18

- A priori information about the problem

- Normality of $P(x | \omega_i)$

$$P(x | \omega_i) \sim N(\mu_i, \Sigma_i)$$

- Characterized by 2 parameters
- Estimation techniques
 - Maximum-Likelihood (ML) and the Bayesian estimations
 - Results are nearly identical, but the approaches are different

Pattern Classification, Chapter 3

19

- Parameters in ML estimation are fixed but unknown!
- Best parameters are obtained by maximizing the probability of obtaining the samples observed
- Bayesian methods view the parameters as random variables having some known distribution
- In either approach, we use $P(\omega_i | x)$ for our classification rule!

Pattern Classification, Chapter 3

20

● Maximum-Likelihood Estimation

- Has good convergence properties as the sample size increases
- Simpler than any other alternative techniques

- General principle

- Assume we have c classes and

$$P(x | \omega_i) \sim N(\mu_i, \Sigma_i)$$

$$P(x | \omega_i) = P(x | \mu_i, \Sigma_i) \text{ where:}$$

$$\theta = (\mu_j, \Sigma_j) = (\mu_j^1, \mu_j^2, \dots, \mu_j^{11}, \mu_j^{22}, \text{cov}(x_j^m, x_j^n) \dots)$$

Pattern Classification, Chapter 3

21

- Use the information provided by the training samples to estimate $\theta = (\theta_1, \theta_2, \dots, \theta_c)$, each θ_i ($i = 1, 2, \dots, c$) is associated with each category

- Suppose that D contains n samples, $x_1, x_2, \dots, x_n$

$$P(D | \theta) = \prod_{k=1}^{k=n} P(x_k | \theta) = F(\theta)$$

$P(D | \theta)$ is called the likelihood of θ w.r.t. the set of samples

- ML estimate of θ is, by definition the value that $\hat{\theta}$ maximizes $P(D | \theta)$

"It is the value of θ that best agrees with the actually observed training sample"

Pattern Classification, Chapter 3

22

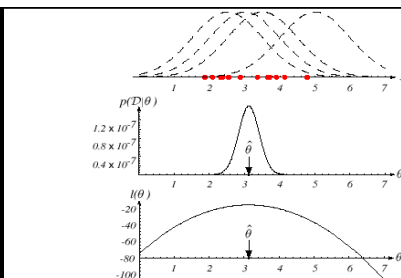


FIGURE 3.1. The top graph shows several training points in one dimension, known or assumed to be drawn from a Gaussian of a particular variance, but unknown mean. Four of the infinite number of candidate source distributions are shown in dashed lines. The middle figure shows the likelihood $p(D|\theta)$ as a function of the mean. If we had a very large number of training points, this likelihood would be very narrow. The value that maximizes the likelihood is marked $\hat{\theta}$; it also maximizes the logarithm of the likelihood—that is, the log-likelihood $l(\theta)$, shown at the bottom. Note that even though they look similar, the likelihood $p(D|\theta)$ is shown as a function of θ whereas the conditional density $p(x|\theta)$ is shown as a function of x . Furthermore, as a function of θ , the likelihood $p(D|\theta)$ is not a probability density function and its area has no significance. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

23

● Optimal estimation

- Let $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ and let ∇_θ be the gradient operator

$$\nabla_\theta = \left[\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_p} \right]^T$$

- We define $l(\theta)$ as the log-likelihood function
 $l(\theta) = \ln P(D | \theta)$
- New problem statement:
determine θ that maximizes the log-likelihood

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} l(\theta)$$

Pattern Classification, Chapter 23

24

Set of necessary conditions for an optimum is:

$$(\nabla_\theta l = \sum_{k=1}^{k=n} \nabla_\theta \ln P(\mathbf{x}_k | \theta))$$

$$\nabla_\theta l = 0$$

Pattern Classification, Chapter 23

25

● Example of a specific case: unknown μ , Σ known

- $P(\mathbf{x}_i | \mu) \sim N(\mu, \Sigma)$
(Samples are drawn from a multivariate normal population)

$$\ln P(\mathbf{x}_i | \mu) = -\frac{1}{2} \ln \left[(2\pi)^d |\Sigma| \right] - \frac{1}{2} (\mathbf{x}_i - \mu)^T \Sigma^{-1} (\mathbf{x}_i - \mu)$$

and $\nabla_\mu \ln P(\mathbf{x}_i | \mu) = \Sigma^{-1} (\mathbf{x}_i - \mu)$

$\theta = \mu$ therefore:

- The ML estimate for μ must satisfy:

$$\sum_{i=1}^{k=n} \Sigma^{-1} (\mathbf{x}_i - \hat{\mu}) = 0$$

Pattern Classification, Chapter 23

26

- Multiplying by Σ and rearranging, we obtain:

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{k=n} \mathbf{x}_k$$

Just the arithmetic average of the samples of the training samples!

Conclusion:

If $P(\mathbf{x}_i | \omega_j)$ ($j = 1, 2, \dots, c$) is supposed to be Gaussian in a d -dimensional feature space; then we can estimate the vector $\theta = (\theta_1, \theta_2, \dots, \theta_c)^T$ and perform an optimal classification!

Pattern Classification, Chapter 23

27

● ML Estimation:

- Gaussian Case: *unknown μ and σ*
 $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$

$$l = \ln P(\mathbf{x}_k | \theta) = -\frac{1}{2} \ln 2\pi\theta_2 - \frac{1}{2\theta_2} (\mathbf{x}_k - \theta_1)^2$$

$$\nabla_\theta l = \begin{pmatrix} \frac{\sigma}{\sigma\theta_1} (\ln P(\mathbf{x}_k | \theta)) \\ \frac{\sigma}{\sigma\theta_2} (\ln P(\mathbf{x}_k | \theta)) \end{pmatrix} = 0$$

$$\begin{cases} \frac{1}{\theta_2} (\mathbf{x}_k - \theta_1) = 0 \\ -\frac{1}{2\theta_2} + \frac{(\mathbf{x}_k - \theta_1)^2}{2\theta_2^2} = 0 \end{cases}$$

Pattern Classification, Chapter 23

28

Summation:

$$\begin{cases} \sum_{k=1}^{k=n} \frac{1}{\hat{\theta}_2} (\mathbf{x}_k - \hat{\theta}_1) = 0 & (1) \\ -\sum_{k=1}^{k=n} \frac{1}{\hat{\theta}_2} + \sum_{k=1}^{k=n} \frac{(\mathbf{x}_k - \hat{\theta}_1)^2}{\hat{\theta}_2^2} = 0 & (2) \end{cases}$$

Combining (1) and (2), one obtains:

$$\hat{\theta}_1 = \hat{\mu} = \sum_{k=1}^{k=n} \frac{\mathbf{x}_k}{n} ; \quad \hat{\theta}_2 = \hat{\sigma}^2 = \frac{\sum_{k=1}^{k=n} (\mathbf{x}_k - \hat{\mu})^2}{n}$$

Pattern Classification, Chapter 23

29

- Bias

- ML estimate for σ^2 is biased

$$E\left[\frac{1}{n}\sum (x_i - \bar{x})^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

- An elementary unbiased estimator for Σ is:

$$\mathbf{C} = \frac{1}{n-1} \sum_{k=1}^{k=n} (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^t$$

Sample covariance matrix

- Appendix: ML Problem Statement

- Let $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$

$$P(\mathbf{x}_1, \dots, \mathbf{x}_n | \theta) = \prod_{k=1}^n P(\mathbf{x}_k | \theta); |D| = n$$

Our goal is to determine $\hat{\theta}$ (value of θ that makes this sample the most representative!)