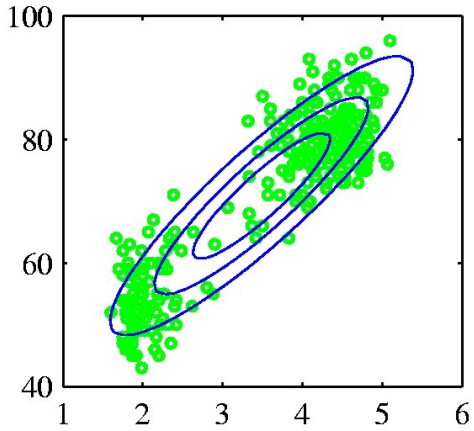
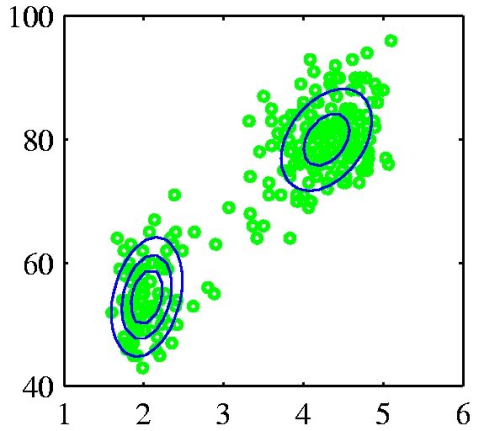
Gaussian Mixture Model

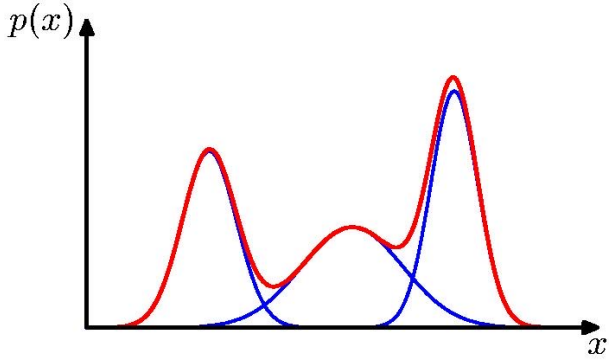
Wei-Ta Chu

Introduction

- By using a sufficient number of Gaussians, and by adjusting their means and covariances as well as the coefficients in the linear combinations, almost any continuous density can be approximated to arbitrary accuracy.







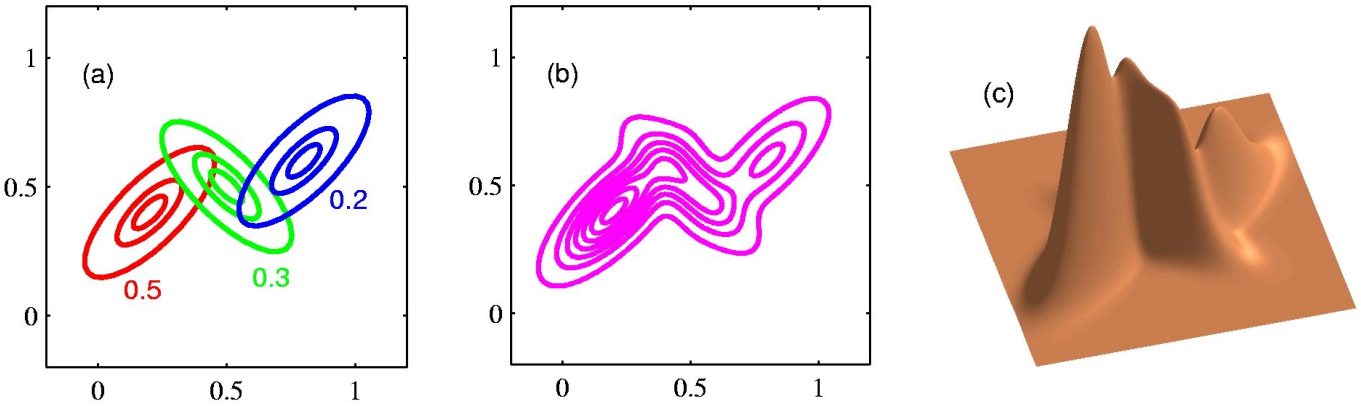
C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006

Introduction

- Consider a superposition of *K* Gaussian densities

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

- Each Gaussian density $\mathcal{N}(x|\mu_k, \Sigma_k)$ is called a component of the mixture and has its own mean μ_k and covariance Σ_k.



Introduction

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

- From the sum and product rules, the marginal density is given by

$$p(x) = \sum_{k=1}^{K} p(k)p(x|k)$$

- We can view $\pi_k = p(k)$ as the prior probability of picking the *k*th component, and the density as the probability of *x* conditioned on *k*:

$$\mathcal{N}(x|\mu_k, \Sigma_k) = p(x|k)$$

Introduction

- From Baye's theorem, the posterior probability $p(k|x)$ is given by

$$p(k|x) = \frac{p(k)p(x|k)}{\sum_l p(l)p(x|l)} = \frac{\pi_k \mathcal{N}(x|\mu_k,\Sigma_k)}{\sum_l \pi_l \mathcal{N}(x|\mu_l,\Sigma_l)}$$

- Gaussian mixture distribution is governed by parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$. One way to set these parameters is to use maximum likelihood.

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}$$

$$\mathbf{X} = \{x_1, ..., x_N\}$$

Assume that different mixtures are independent and identically distributed

Introduction

- In case of a single variable x, the Gaussian distribution is in the form

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

- For a D-dimensional vector $\boldsymbol{x}$, the multivariate Gaussian distribution takes the form

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

Maximizing Likelihood

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}$$

- Setting the derivative of $\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to the means $\boldsymbol{\mu}$ of the Gaussian components to zero

$$0 = -\sum_{n=1}^{N} \underbrace{\frac{\pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}}_{\gamma(z_{nk})} \boldsymbol{\Sigma}_k(\boldsymbol{x}_n - \boldsymbol{\mu}_k)$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_n \qquad N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

- The mean $\boldsymbol{\mu}_k$ for the *k*th Gaussian component is obtained by taking a weighted mean of all of the points in the data set, in which the weighting factor for data point $\boldsymbol{x}_n$ is given by the posterior probability $\gamma(z_{nk})$

Maximizing Likelihood

- Setting the derivative of $\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to the covariance $\boldsymbol{\Sigma}_k$ of the Gaussian components to zero

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T$$

- Each data point weighted by the corresponding posterior probability

Maximizing Likelihood

- Maximize $\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to the mixing coefficients π_k
- Constraint: the sum of mixing coefficients is one
- Using Lagrange multiplier and maximizing the following quantity

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right) \Rightarrow 0 = \sum_{n=1}^{N} \frac{\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \lambda$$

- If we multiply both sides by π_k and sum over *k* making use of the constraint $\sum_{k=1}^{K} \pi_k = 1$, we find $\lambda = -N$. Using this to eliminate λ and rearranging we obtain $\pi_k = \frac{N_k}{N}$
  - Mixing coefficient of the *k*th component is given by the average responsibility which that component takes for explaining the data points

Expectation-Maximization (EM) Algorithm

- We first choose some initial values for the means, covariances, and mixing coefficients.
- Expectation step (E step)
  - Use the current parameters to evaluate the posterior probabilities
- Maximization step (M step)
  - Re-estimate the means, covariances, and mixing coefficients
- Each update to the parameters resulting from an E step followed by an M step is guaranteed to increase the log likelihood function.

EM for Gaussian Mixtures

Given a Gaussian mixture model, the goal is to maximize the likelihood function with respect to the parameters (comprising the means and covariances of the components and the mixing coefficients).

1. Initialize the means $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ and mixing coefficients π_k, and evaluate the initial value of the log likelihood.
2. **E step**. Evaluate the responsibilities using the current parameter values

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \tag{9.23}$$

3. **M step**. Re-estimate the parameters using the current responsibilities

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \tag{9.24}$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^{\mathrm{T}} \tag{9.25}$$

$$\pi_k^{\text{new}} = \frac{N_k}{N} \tag{9.26}$$

where

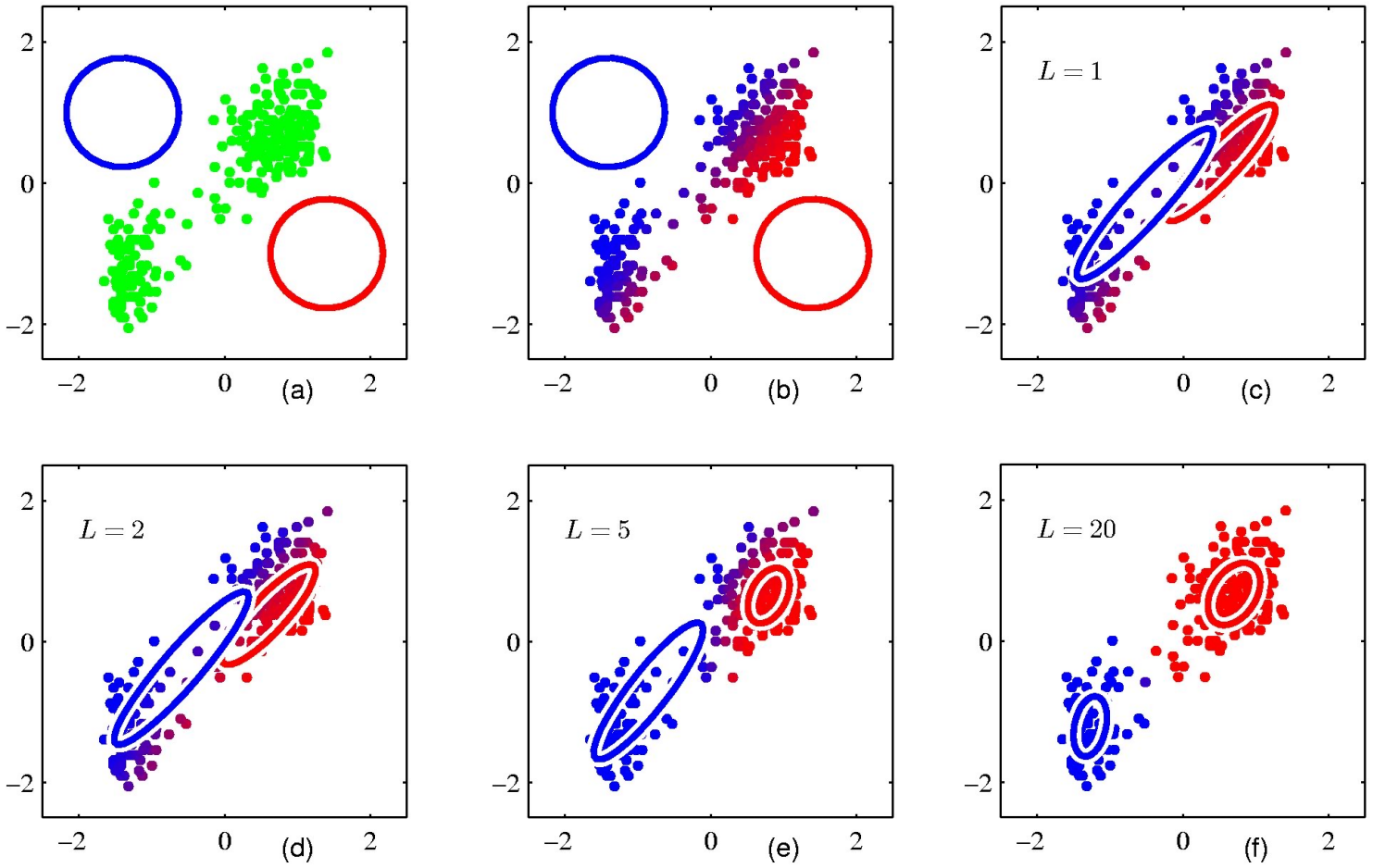
$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}). \tag{9.27}$$

4. Evaluate the log likelihood

$$\ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \tag{9.28}$$

and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to step 2.

Example



Case Study

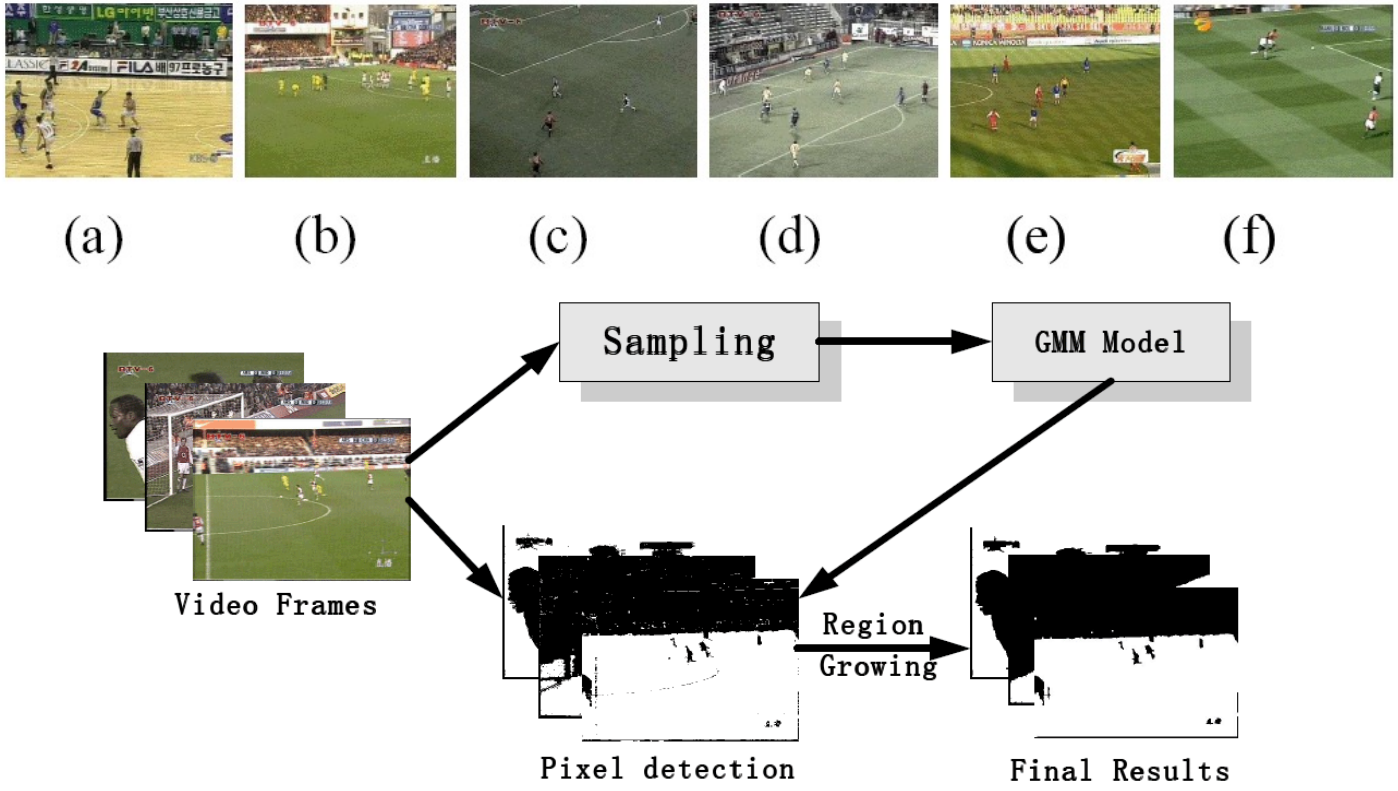


Figure 2: Flowchart of segmentation Algorithm

Jiang, et al. "A new method to segment playfield and its applications in match analysis in sports video," In Proc. of ACM MM, pp. 292-295, 2004.

Case Study

- The condition density of a pixel ξ belongs to the playfield region Φ is modeled with M Gaussian densities: $p(\xi|\Phi) = \sum_{i=1}^{M} w_i b_i(\xi)$

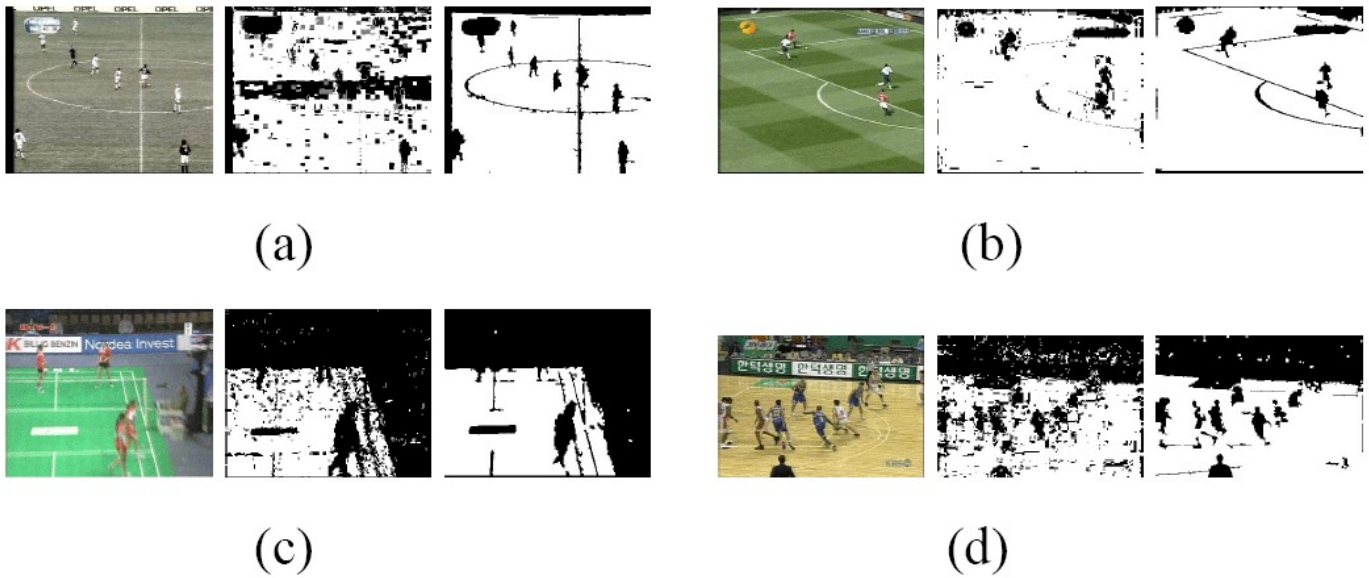


Figure 4: Segmentation results of some frames