Bayesian analysis in image processing

Florence Tupin
Télécom ParisTech - LTCI
Image classification

- **Introduction**

- **Bayesian classification**
  - Image modeling
  - Mono-spectral case
  - Multi-spectral case
  - Punctual / Contextual

- **Kmeans classification**
  - Unsupervised case
  - Algorithm
Introduction

- **Classification objectives**
  - identification of the different classes in the image
  - preliminary step of pattern recognition methods (object detection)

- **Hypotheses**
  - grey-level images
  - classes = peaks in the histogram
  - low grey-level variations in the same class
  - punctual classification: each pixel is classified separately
  - supervised learning: samples of each class are available

- **Possible extensions**
  - multi-channel images
  - contextual classification: markovian framework
Image classification

○ **Introduction**

○ **Bayesian classification**
  - Image modeling
  - Mono-spectral case
  - Multi-spectral case
  - Punctual / Contextual

○ **Kmeans classification**
  - Unsupervised case
  - algorithm
Image modeling

- **Probabilistic model**

\[ S \text{ set of sites (pixels = localization } (i,j) \text{ in the image)} \]
\[ y_s \in \{0,\ldots,255\} = E \quad \rightarrow \quad x_s \in \{1,\ldots,K\} = \Lambda \]
\[ Y_s \text{ random variable of grey-level} \quad \quad \quad \quad \quad \quad \quad \quad X_s \text{ random variable of label} \]
Example: brain image
Example : brain image
Image classification

- **Introduction**

- **Bayesian classification**
  - Image modeling
  - Mono-spectral case
  - Multi-spectral case
  - Punctual / Contextual

- **Kmeans classification**
  - Unsupervised case
  - algorithm
Mono-spectral case (grey-level image) (1)

- **Maximum A Posteriori criterion**

you know a grey-level $y_s$ for pixel $s$

$\Rightarrow$ take the “best” class $x_s$ knowing $y_s$

$\Rightarrow$ find the $i$ which maximizes $P(X_s = i|Y_s = y_s) \ \forall i \in \Lambda$

- **Can we compute $P(X_s = i|Y_s = y_s)$?**

Bayes rule

$$P(X_s = i|Y_s = y_s) = \frac{P(Y_s = y_s|X_s = i)P(X_s = i)}{P(Y_s = y_s)}$$

$$X_s = \text{argmax}_{i \in \{1,\ldots,K\}} P(Y_s = y_s|X_s = i)P(X_s = i)$$
Mono-spectral case (grey-level image) (2)

\[ P(X_s = i|Y_s = y_s) = \frac{P(Y_s = y_s|X_s = i)P(X_s = i)}{P(Y_s = y_s)} \]

- Example of brain image

\( \Lambda = \{0 = \emptyset; 1 = \text{skin}; 2 = \text{bone}; 3 = \text{GrayMatter}; 4 = \text{WhiteMatter}; 5 = \text{LCR}\} \)

- \( P(X_s = i) = \) apparition probability of class \( i \)
- \( P(Y_s = y_s|X_s = i) = \) grey-level distribution knowing that the pixels belong to class \( i \)
Mono-spectral case (grey-level image) (3)

How can we learn these probabilities ?

- **Learning of** $P(X_s = i)$
  - frequencies of apparition for each class
  - no knowledge : uniform distribution ($P(X_s = i) = \frac{1}{\text{Card}(\Lambda)}$) ⇒ Maximum Likelihood criterion

- **Learning of** $P(Y_s = y_s | X_s = i)$

grey-level histogram for class $i$
Example (brain image)
Mono-spectral case (grey-level image) (4)

○ **Supervised learning**
  
  • samples selection in an image
  
  • histogram computation
  
  • histogram filtering

○ **Parametric case**

If there exists a parametric model for the grey-level distribution, compute the model parameters!

**Ex:**

• Gaussian distribution: _mean, standard deviation_

• Gamma distribution: _mean, knowledge of the sensor parameter_
Mono-spectral case (grey-level image) (5)

○ Case of a Gaussian distribution

each class \( i \in \Lambda \) is characterized by \((\mu_i, \sigma_i)\)

● the conditional probability is:

\[
P(Y_s = y_s | X_s = i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{(y_s - \mu_i)^2}{2\sigma_i^2}\right)
\]

● if the classes are equiprobable:

\[
P(Y_s | X_s = i) \quad \text{maximum} \Leftrightarrow \frac{(y_s - \mu_i)^2}{2\sigma_i^2} + \ln(\sigma_i) \quad \text{minimum}
\]

● if the classes are equiprobable and have the same standard deviation (gaussian noise):

\[
P(Y_s | X_s = i) \quad \text{maximum} \Leftrightarrow (y_s - \mu_i)^2 \quad \text{minimum}
\]
Image classification

- **Introduction**

- **Bayesian classification**
  - Image modeling
  - Mono-spectral case
  - Multi-spectral case
  - Punctual / Contextual

- **Kmeans classification**
  - Unsupervised case
  - algorithm
Multi-spectral case

- **Vectorial observations**

\[ \bar{y}_s \in \{0, ..., 255\}^d \]

\[ \Downarrow \]

\[ \bar{X}_s \]

- **Bayes rule**

\[
P(X_s = i | \bar{Y}_s = \bar{y}_s) = \frac{P(\bar{Y}_s = \bar{y}_s | X_s = i)P(X_s = i)}{P(\bar{Y}_s = \bar{y}_s)}
\]

\[
x_s = \text{argmax}_{i \in \{1, ..., K\}} P(\bar{Y}_s = \bar{y}_s | X_s = i)P(X_s = i)
\]

\[ P(\bar{Y}_s = \bar{y}_s | X_s = i) \] multidimensionnal histogram for class i
Multi-spectral case

- Multi-variate Gaussian distribution

Each class \( i \in \Lambda \) is characterized by \((\mu_i, \Sigma_i)\) (mean vector and variance-covariance matrix)

- The conditional probability is:

\[
P(Y_s = y_s | X_s = i) = \frac{1}{\sqrt{2\pi}^d \sqrt{\det(\Sigma_i)}} \exp\left(-\frac{1}{2}(y_s - \mu_i)^t \Sigma_i^{-1} (y_s - \mu_i)\right)
\]

- If the classes are equiprobable with the same covariance matrix:

\[
P(Y_s | X_s = i) \text{ maximum} \Leftrightarrow (y_s - \mu_i)^t \Sigma_i^{-1} (y_s - \mu_i) \text{ minimum}
\]

\[\Rightarrow \text{linear classifier}\]
Image classification

- Introduction

- Bayesian classification
  - Image modeling
  - Mono-spectral case
  - Multi-spectral case
  - Punctual / Contextual

- Kmeans classification
  - Unsupervised case
  - algorithm
Contextual / punctual classification

- **Global classification**

  \[ y = \{y_s\}_{s \in S} \text{(observed image)}, \quad Y = \{Y_s\}_{s \in S} \text{(random field)} \]

  \[ x = \{x_s\}_{s \in S} \text{(searched classification)}, \quad X = \{X_s\}_{s \in S} \text{(random field)} \]

  \[
P(X = x|Y = y) = \frac{P(Y = y|X = x)P(X = x)}{P(Y = y)}
  \]

- **Independance assumption for** \( P(Y = y|X = x) \)

  \[
P(Y = y|X = x) = \prod_{s \in S} P(Y_s = y_s|X_s = x_s)
  \]

- **Independance assumption for** \( P(X = x) \)

  \[
P(X = x) = \prod_{s \in S} P(X_s = x_s)
  \]

  \[\Rightarrow P(X = x|Y = y) \propto \prod_{s \in S} P(Y_s = y_s|X_s)P(X_s = x_s) \Rightarrow \text{punctual classif.}!\]
Contextual / punctual classification

- **Prior knowledge on** $P(X)$
  - independence assumption not verified in practice: images are smooth with strong spatial coherency (image description = smooth areas)
  - BUT the coherency is at a local scale $\Rightarrow$ introduction of contextual knowledge

Markov random fields $\Rightarrow$ smoothness of the solution, local spatial coherency in the result!
Image classification

○ Introduction

○ Bayesian classification
  ● Image modeling
  ● Mono-spectral case
  ● Multi-spectral case
  ● Punctual / Contextual

○ Kmeans classification
  ● Unsupervised case
  ● algorithm
Unsupervised case

○ Bayesian case with gaussian distribution and same std

\[ P(Y_s|X_s = i) \ MAXIMUM \Leftrightarrow ||y_s - \mu_i||^2 \ MINIMUM \]

○ Unsupervised classification

\( \mu_i \) unknown \( \Rightarrow \)

• classify the pixels using some initial \( \mu_i^0 \)

• compute new \( \mu_i^1 \) with the empirical mean of the classified pixels

• iterate until no modification in \( \mu_i^k \)

\( \Rightarrow \)kind of bayesian classification with changing means
Unsupervised case

- **K-means algorithm**
  
  - choose $\mu_1^0, \mu_2^0, \ldots, \mu_K^0$

  At iteration $k$:
  
  - $\forall s \in S \quad l_s = \arg\min_{i \in \Lambda} ||y_s - \mu_i^k||^2$
  - $\forall i \in \{1, \ldots, K\} \quad \mu_{i}^{k+1} = \frac{1}{\text{card}(R_i)} \sum_{s,l_s=i} x_s$
  - if $\mu_{i}^{k} \neq \mu_{i}^{k+1}$ iterate

- **Drawbacks**
  
  - no proof of convergence to the optimal solution
  - influence of the initial means
Unsupervised case

- **K-means algorithm for grey-level images**

  In 1D, everything can be done with the histogram!

  \( f_n \) frequency of grey-level \( n \) in the image

  Ex for 2 classes:

  initial values of the centers \( m_1^0 = 10 \); \( m_2^0 = 120 \)

  initial classification: thresholding with \( t = 65 \) (histogram)

  computation of new means:

  \[
  \mu_1^1 = \frac{1}{\sum_{n=t}^{n=t} f_n} \sum_{n=0}^{n=t} n f_n
  \]

  \[
  \mu_2^1 = \frac{1}{\sum_{n=t}^{n=N} f_n} \sum_{n=t}^{n=N} n f_n
  \]

- **multi-channels** : in nD, high cost of storage \( \Rightarrow \) updating of a classified image

- Application to high dimensional space : reduction with ACP
Brain image
SAR image
Imagerie radar