

M2RI, Parcours Images et Données

Module Acquisition et Représentation des Données

Extraction de caractéristiques

Features Extraction - 1

Bertrand Couasnon

Irisa/Insa

Imadoc

Bureau C314

<http://www.irisa.fr/imadoc>

couasnon@irisa.fr

■ Sensors and Digitization

◆ Data capture

- ✦ Sound signal
- ✦ Image
- ✦ Video

■ Analysis and transformation

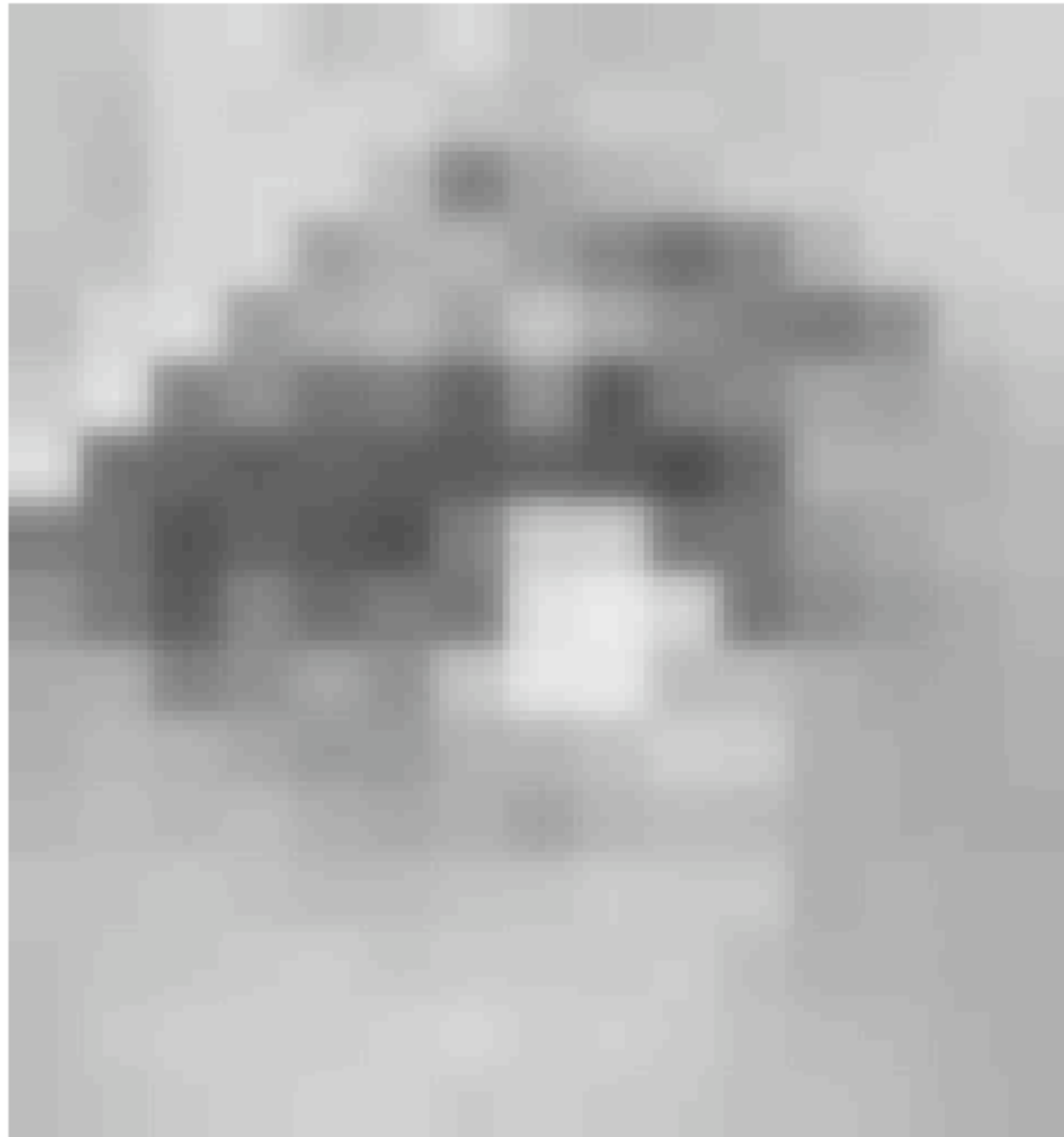
- ◆ Space transformation
- ◆ Improve signal quality

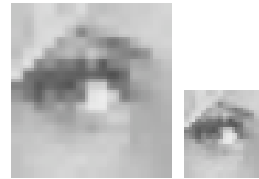
■ Objective

- ◆ Information extraction

■ What is it ?

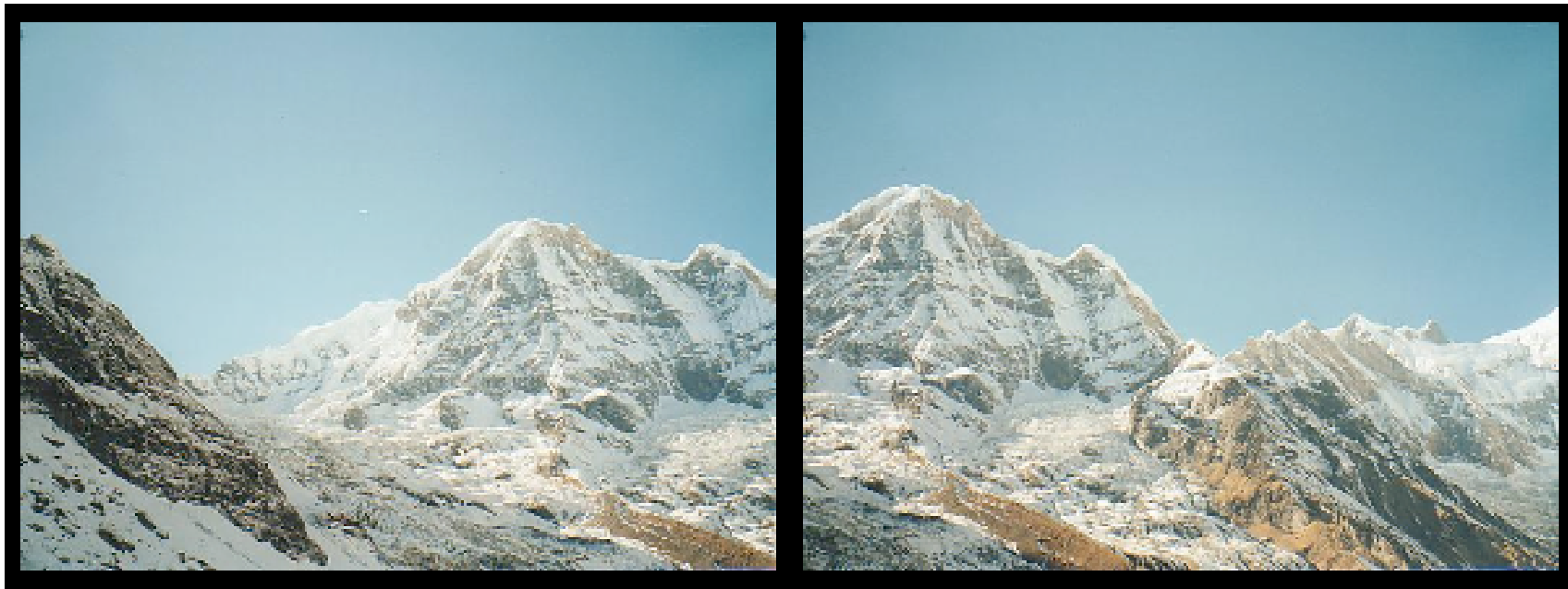
255 255 255 255 255 255 185 188 208 210 183 192 211 182 177 182 189
194 193 199 204 255 255 255 255 185 179 206 203 205 204 185 178 193
192 187 195 196 202 199 255 255 255 255 188 174 205 210 209 176 113
150 168 178 195 195 202 202 198 255 255 255 255 184 183 206 210 146
164 168 143 118 096 118 166 197 199 202 255 255 255 255 175 207 216
141 160 172 147 187 155 125 114 105 125 187 193 255 255 255 255 194
220 112 130 101 108 083 130 073 112 122 151 147 165 182 255 255 255
255 216 096 087 078 085 078 076 085 076 064 089 163 161 162 173 255
255 255 255 108 100 069 085 075 067 112 195 198 099 102 139 153 167
173 255 255 255 255 132 103 075 119 092 109 100 220 231 205 100 124
141 155 168 255 255 255 255 158 163 114 129 152 131 186 226 229 177
175 162 153 156 163 255 255 255 255 163 173 168 153 139 141 165 168
181 199 195 161 160 156 163 255 255 255 255 172 178 172 177 162 155
163 149 170 178 174 161 160 158 158 255 255 255 255 183 183 189 182
176 177 184 186 195 190 192 161 168 162 163 255 255 255 255 179 186
188 194 196 188 194 194 192 195 178 166 167 163 160 255 255 255 255
179 190 197 198 196 202 206 198 200 194 182 176 173 165 160 255 255
255 255 179 185 194 196 203 202 194 197 193 189 181 181 176 167 163



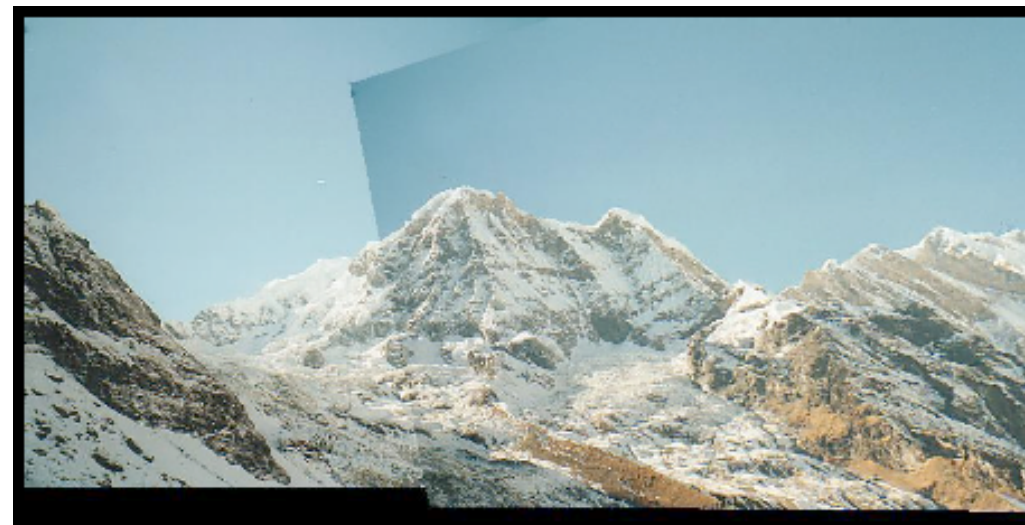
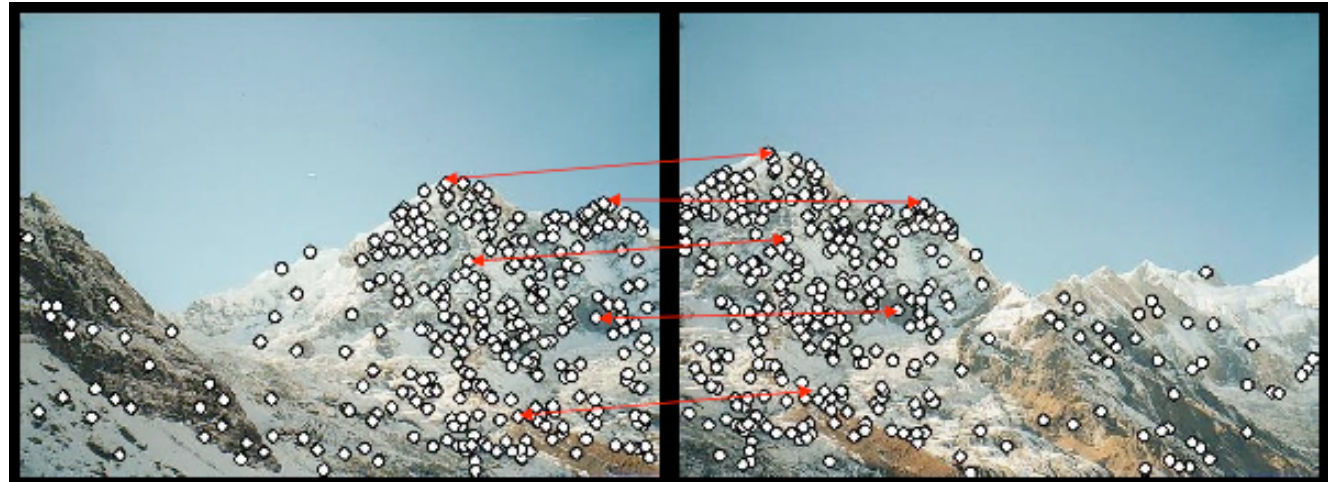


- **Original signal/image**
 - ◆ **Data array**
 - ◆ **Too much information**
- **Need for less information**
- **Information levels**
 - ◆ **Image of handwritten text --> text transcription (in ASCII for ex.)**
- **Information reduction**
 - ◆ **More concise representation**
 - ◆ **Maintains the meaningful**
 - ◆ **According to the needs**
 - ✦ Kind of information
 - ✦ Application
- **First reduction level**
 - ◆ **Features detection/extraction**

- **Speech recognition**
 - ◆ **What is said**
- **Speaker recognition**
 - ◆ **Who speaks**
- **Image registration**
 - ◆ **Align two images**



- Detect feature points in both images
- Find corresponding pairs
- Use these pairs to align images



Images from: M. Brown and D. G. Lowe. Recognising Panoramas. In Proceedings of the International Conference on Computer Vision (ICCV 2003)

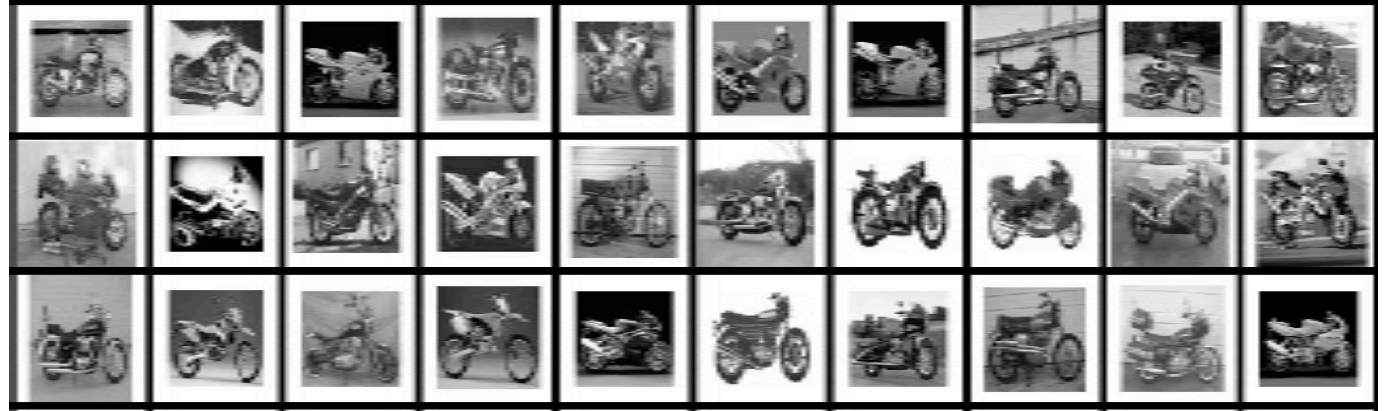
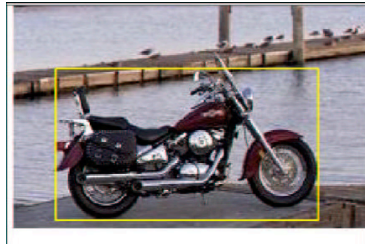
- Recognition in an image

- ◆ Find book on the left in image on the right



■ Find images with an object

◆ Request



■ Locate objects in images

◆ Locate cars



■ Robotique / Vision industrielle :

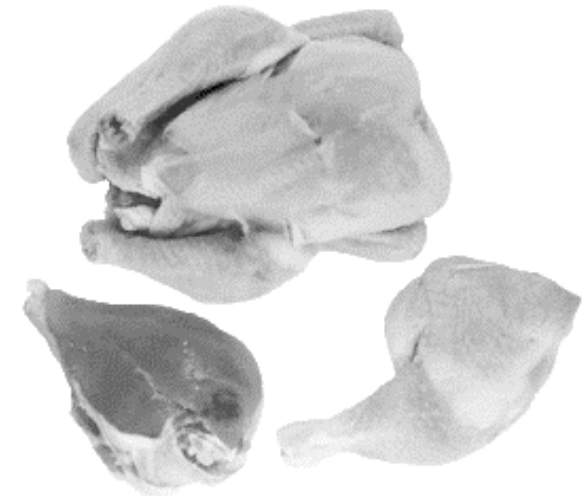
- ◆ véhicules (semi-)autonomes
- ◆ saisie guidée par l'image
- ◆ contrôle qualité



Circulation :
gestion du
comportement du
véhicule



Mise en bouteille :
Vérification du
remplissage et de
l'étiquetage



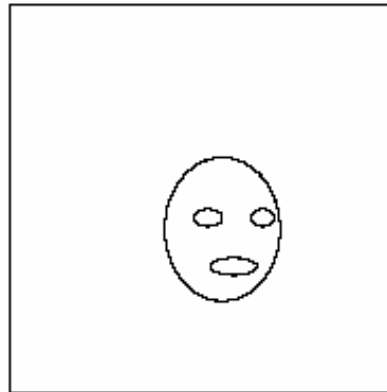
Agro alimentaire :
Contrôle qualité

■ 3D Object Recognition



■ Face Detector

- ◆ Détection-localisation
- ◆ Analyse des expressions
- ◆ Reconnaissance des visages



Expression : sourire (?)

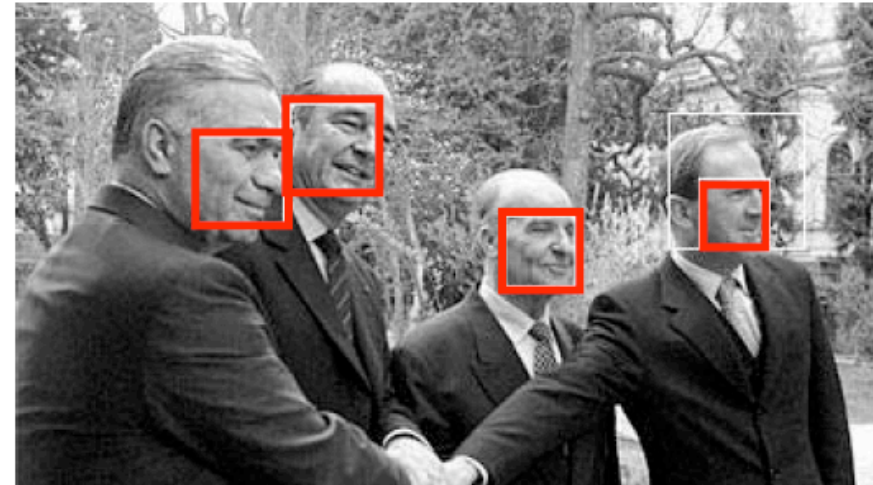
Identification : Lena

- **Domaines** : systèmes d'identification, recherche dans des bases de données, visiophonie

■ Face Detector

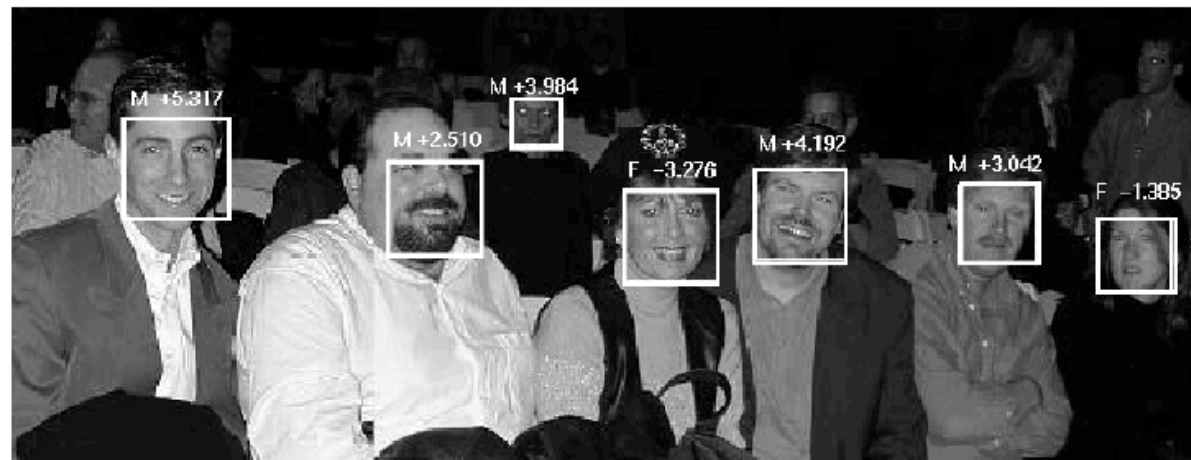


Facial Feature Localization



Profile Detection

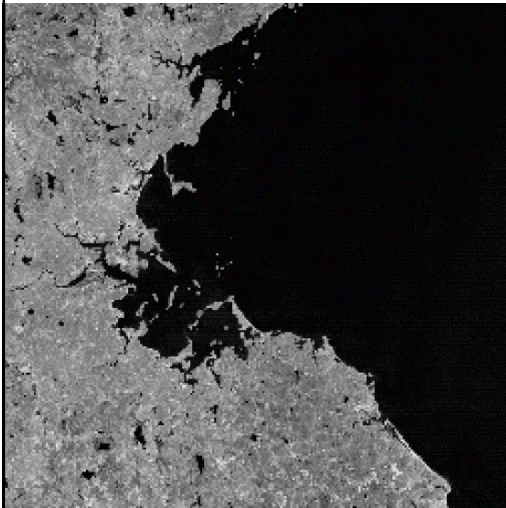
Demographic Analysis



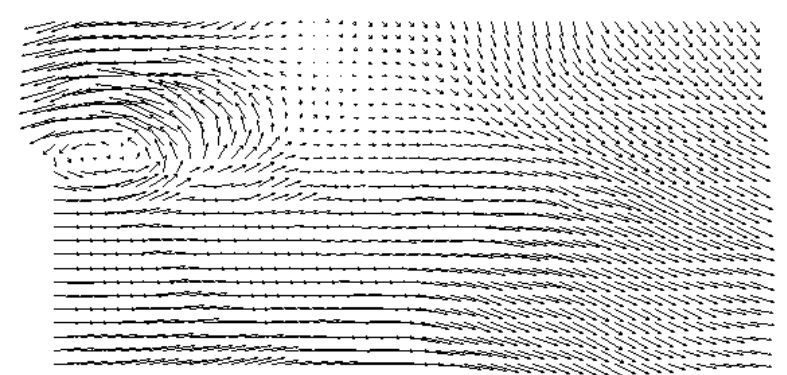
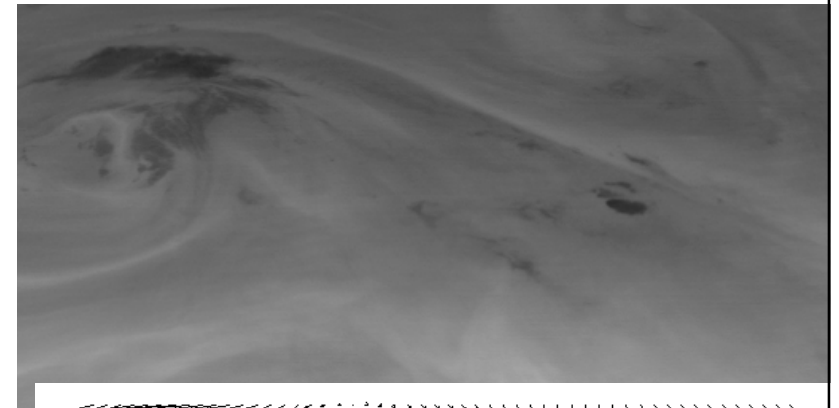
Slide credit: Frank Dellaert, Paul Viola, Foryth&Ponce

■ Télédétection :

- ◆ classification automatique (agronomie, urbanisme)
- ◆ détection de réseaux routiers (cartographie, SIG)
- ◆ restitution du relief (MNT)
- ◆ Estimation de champs de vents
- ◆ ...

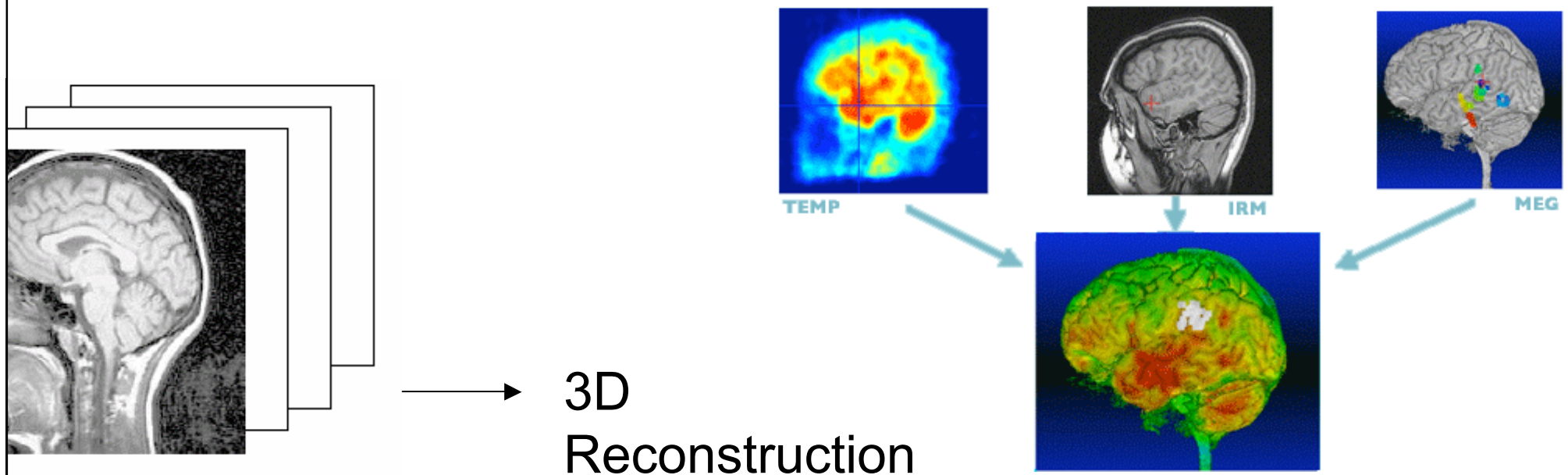


55 % étendues d'eau
15 % zones urbaines
30 % zones cultivées



■ Traitement d'images médicales

- ◆ Reconstruction de volumes 3D à partir de coupes
- ◆ Aide au diagnostic (détection, localisation, suivi)
- ◆ Aide au geste chirurgical (localisation, guidage, visualisation)



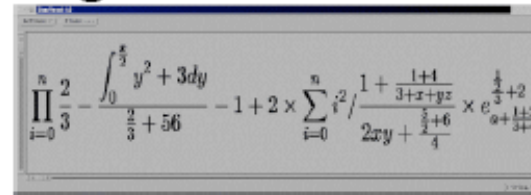
Introduction: Examples of application

- Document structure recognition

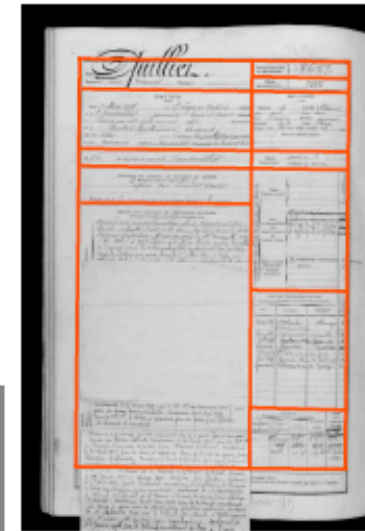
- Optical Music Recognition



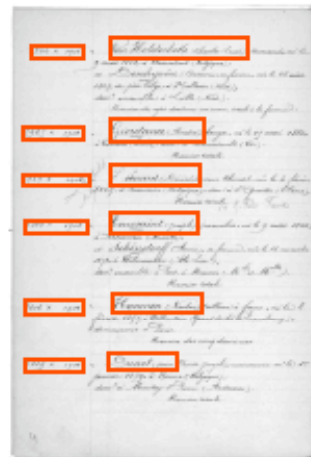
- Mathematical Formulae Recognition



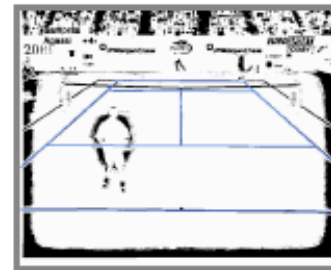
- Old Forms Recognition



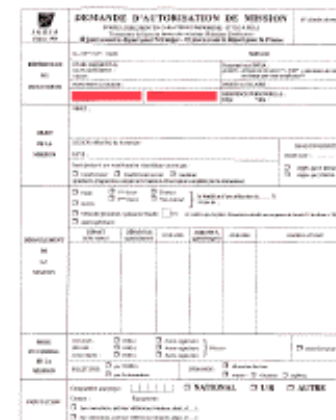
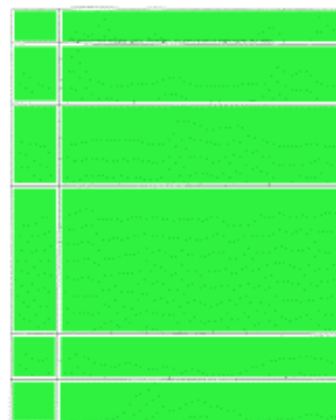
- Handwritten Structure Recognition



- Tennis Court Recognition



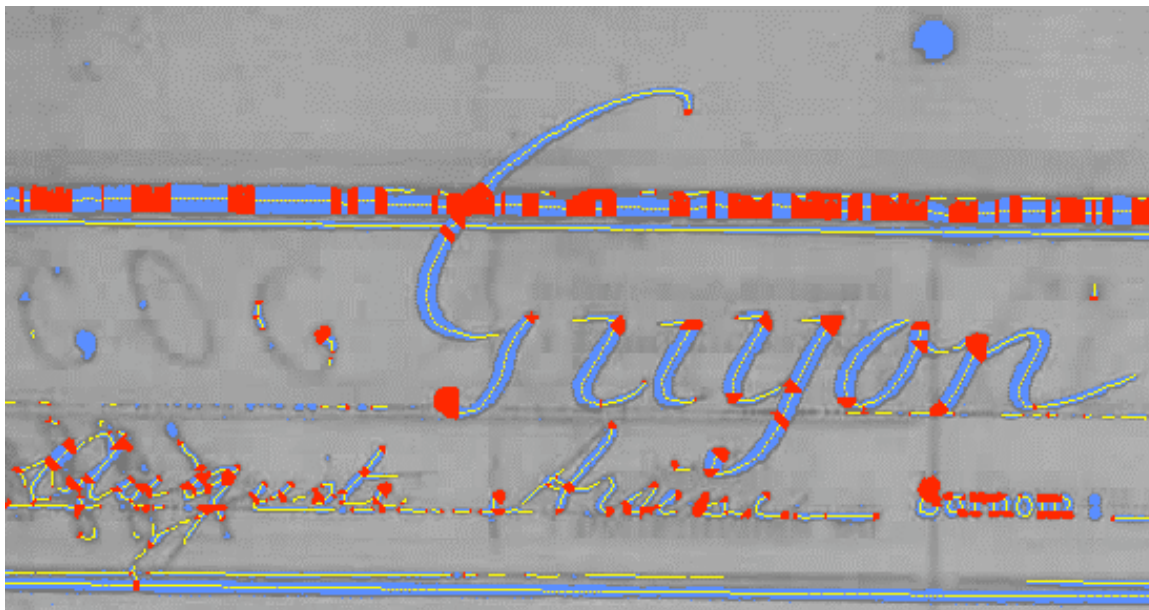
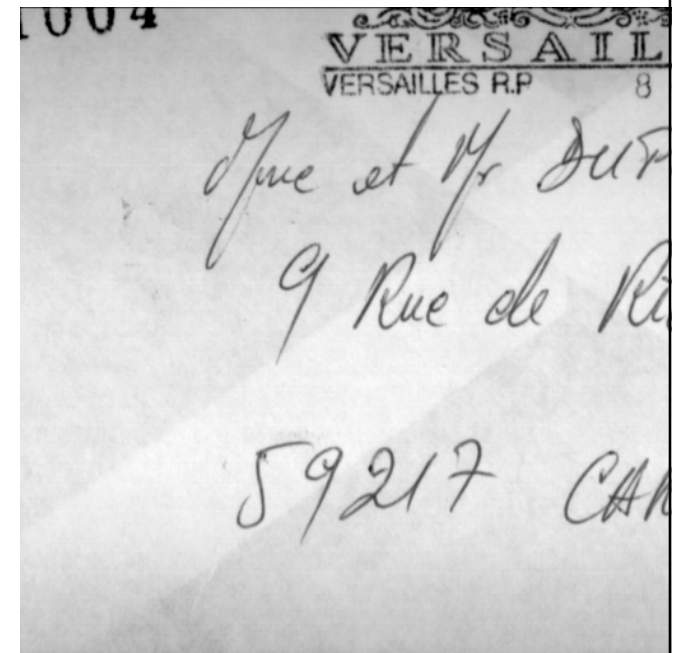
- Hierarchical Table Structure Recognition



∇ # columns,
∇ # rows

■ Off-line Handwriting recognition

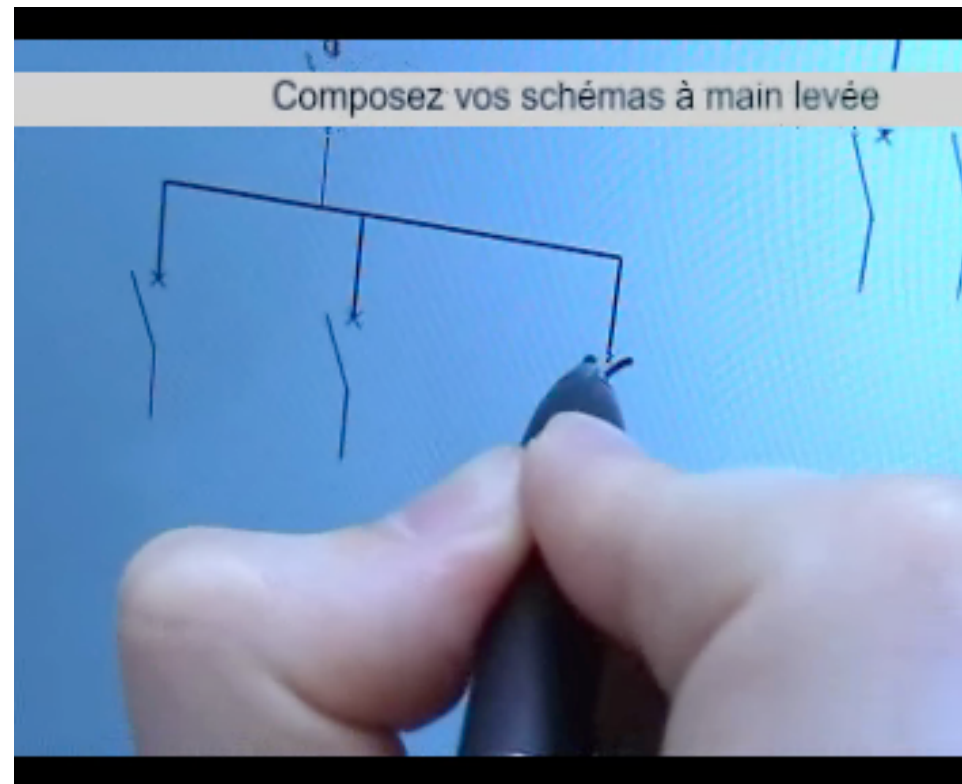
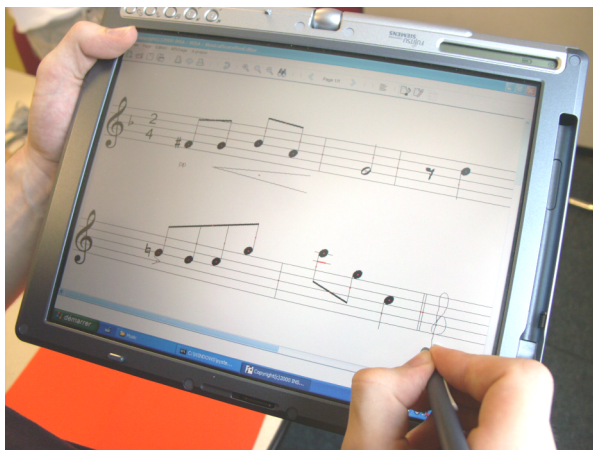
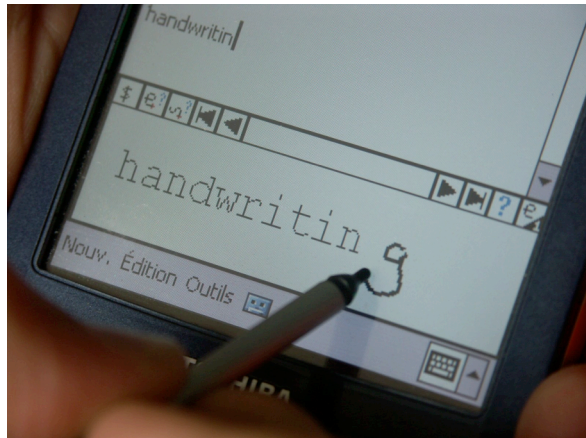
- ◆ Bank Check
- ◆ Postal Address
- ◆ Forms
- ◆ Historical documents
- ◆ ...



--> Guyon, Auguste Arsène

■ On-line Handwriting Recognition

- ◆ Tablet PC
- ◆ PDA
- ◆ Smartphone



Introduction: Examples of application

Interface d'annotation collective

Fiches

Case Etat Civil

Date naissance: 27/8/1860/Grégorien | Lieu naissance: ponthévrard | Canton naissance: | Département naissance: | Résidence: villers | Canton résidence: | Département résidence: | Profession: | Nom père: | Prénoms père:

Recherche

Nom à rechercher: **v a l l i e r**

chevallier

Registre RM0108
Tri des 30 premiers résultats

1 Chevallier
2 Chevallier
3 Chevallier
4 Chevallier
5 Hénager
Cardeller

Image 138 / 503 Afficher Texte

Case Nom
Nom: Chevallier
Prénoms: Louis

Case Masculin (1)
Date de naissance: 27/8/1860

Case Classe (1)
Classe: 1880

Case Etat Civil (1)
ÉTAT CIVIL
Né le 27 août 1860 à Ponthévrard, canton de Ponthévrard, département de la Somme, résident à Villers, canton de Chyennes, département de la Somme, profession de charpentier.
Fils de Louis Chevallier, en son domicile à Ponthévrard, canton de Ponthévrard, département de la Somme.

Case Signalement
Cheveux: bruns, yeux: bleus, nez: droit, bouche: moyenne, taille: 1 m 52, poids: 52 kg, tatouage: 1 m.

Case Trage

Case Conseil de Révision

Case Etat de Service

Case Instruction
Instruction: militaire (a) 1880

Case Affectation
Affectation: 30 Régiment de Chasse

Case Domicile

- Text Detection/Recognition in real scenes

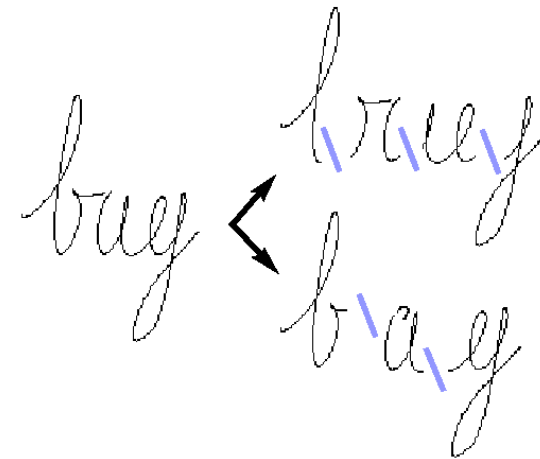
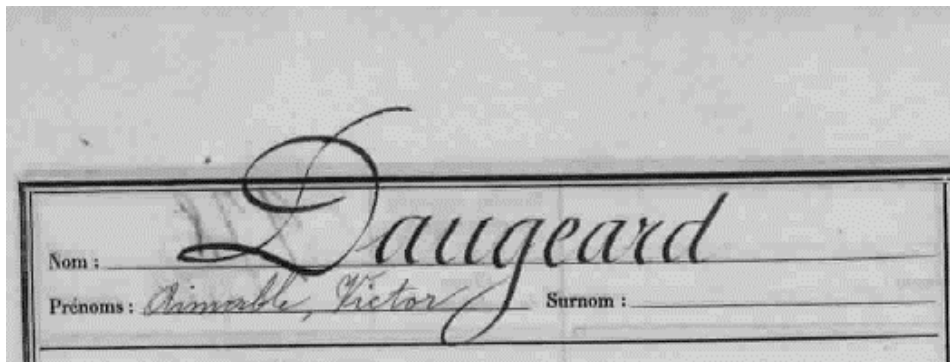


- All those applications
 - ◆ Need some features to build a recognition
 - ✦ Specific points in image
 - ✦ Specific regions
 - ✦ Describe Object Shapes
 - ✦ Texture

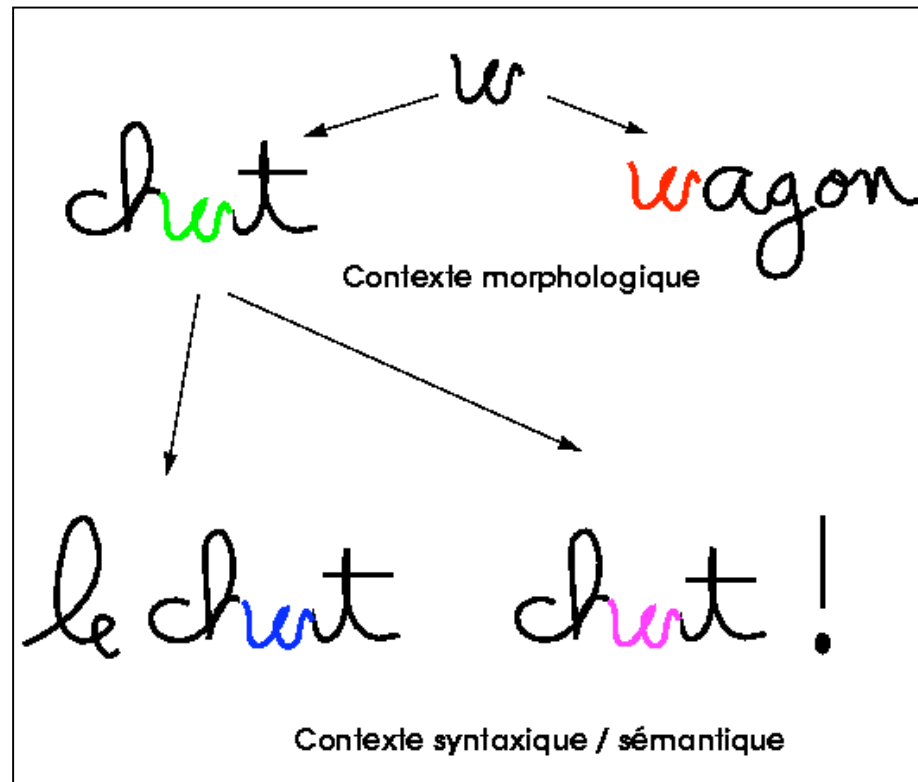
- **Difficulties of Feature Extraction**
 - ◆ Segmentation
 - ◆ Recognition in Images
 - ◆ Recognition in Document Images
- **Feature Extraction/Detection on the whole Signal**
 - ◆ Global Descriptor
 - ◆ Local Descriptor
- **Feature Extraction/Detection on Objects**
 - ◆ Object Localization / Extraction
 - ◆ Object Characterization
- **Interest of Multi-Resolution**
- **Using Features**

- **Difficulties of Feature Extraction**
 - ◆ **Segmentation**
 - ◆ **Recognition in Images**
 - ◆ **Recognition in Document Images**
- **Feature Extraction/Detection on the whole Signal**
 - ◆ **Global Descriptor**
 - ◆ **Local Descriptor**
- **Feature Extraction/Detection on Objects**
 - ◆ **Object Localization / Extraction**
 - ◆ **Object Characterization**
- **Interest of Multi-Resolution**
- **Using Features**

- To recognize an object, it needs to be segmented
- But to segment it correctly, it should have been recognized before



■ Contextual segmentation



■ Speech analysis

- ◆ Segment sentences in words
- ◆ Separate speakers
- ◆ Recognition in noisy environment
 - ✦ Separate speaker from noise

■ Sound Analysis

- ◆ Separate musical instruments
- ◆ Separate each note for polyphonic instruments

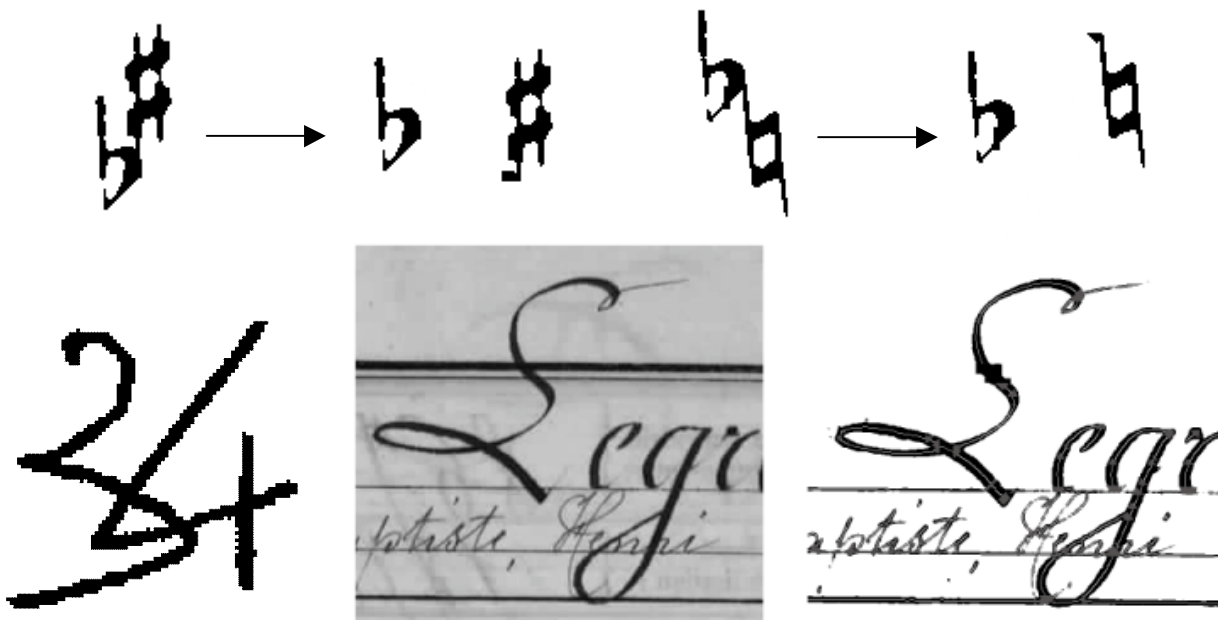
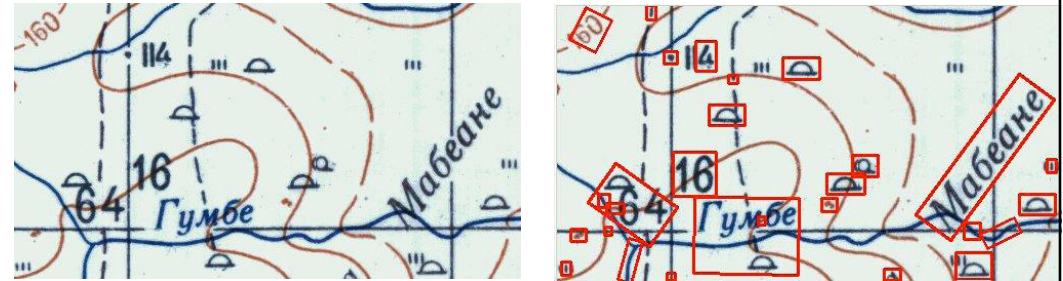
■ Locate/Detect Objects

◆ Separate cars from the rest of the image

- ◆ On one isolated image
- ◆ Using the sequence



- Document Images
- Difficult to separate
 - ◆ Text / Graphics
 - ◆ Text / Pictures
 - ◆ Connected Characters
 - ◆ Handwriting touching objects
 - ◆ ...



- **Difficulties of Feature Extraction**
 - ◆ Segmentation
 - ◆ Recognition in Images
 - ◆ Recognition in Document Images
- **Feature Extraction/Detection on the whole Signal**
 - ◆ Global Descriptor
 - ◆ Local Descriptor
- **Feature Extraction/Detection on Objects**
 - ◆ Object Localization / Extraction
 - ◆ Object Characterization
- **Interest of Multi-Resolution**
- **Using Features**

■ Recognition in different conditions

◆ Size

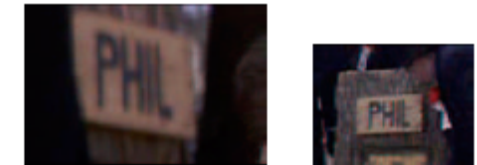


◆ Illumination

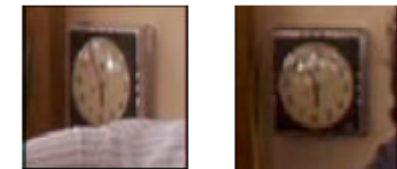


■ Recognition in different conditions

◆ Point of view



◆ Partial occultation

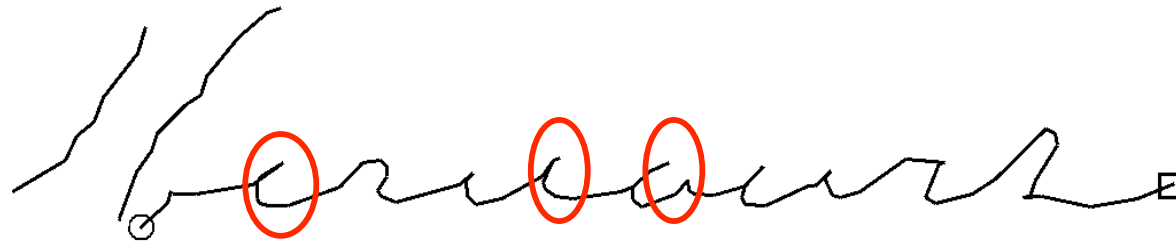
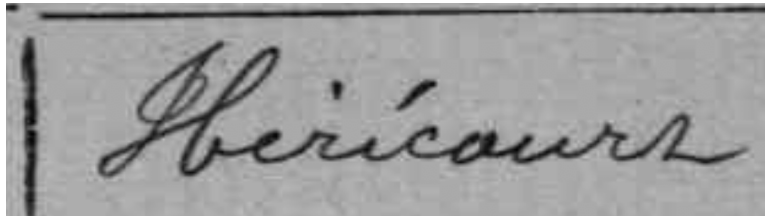


■ 3D objects projected in 2D

■ Background very heterogeneous

- **Difficulties of Feature Extraction**
 - ◆ Segmentation
 - ◆ Recognition in Images
 - ◆ Recognition in Document Images
- **Feature Extraction/Detection on the whole Signal**
 - ◆ Global Descriptor
 - ◆ Local Descriptor
- **Feature Extraction/Detection on Objects**
 - ◆ Object Localization / Extraction
 - ◆ Object Characterization
- **Interest of Multi-Resolution**
- **Using Features**

- A lot of information in details



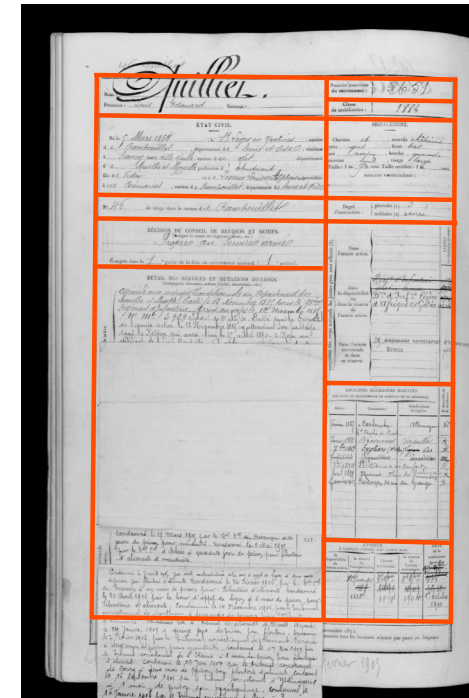
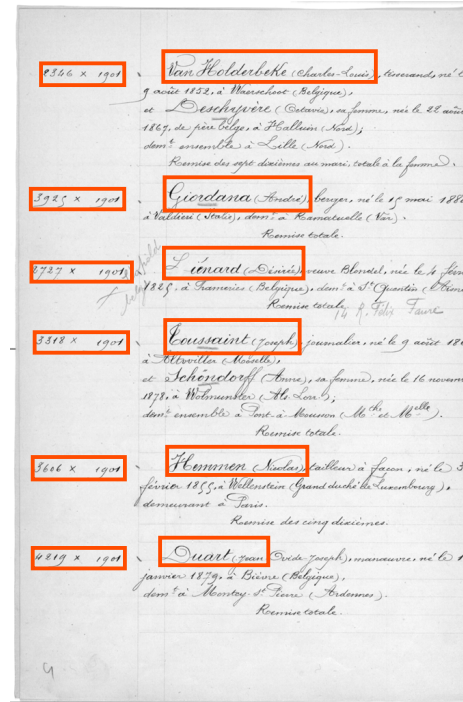
- Segmentation problems between objects

$$\frac{f''(0)}{2!} x^2$$

$$\sqrt{\frac{(v+k+1)(k-v+1)}{(2k+1)(2k+2)}}$$

■ Strong Structure can help

◆ Document Structure



◆ Handwritten / Printed Text Structure

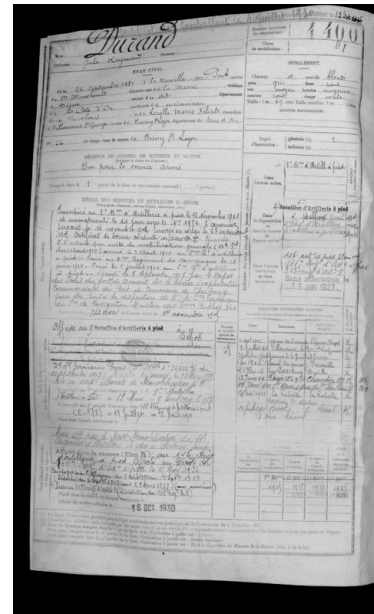
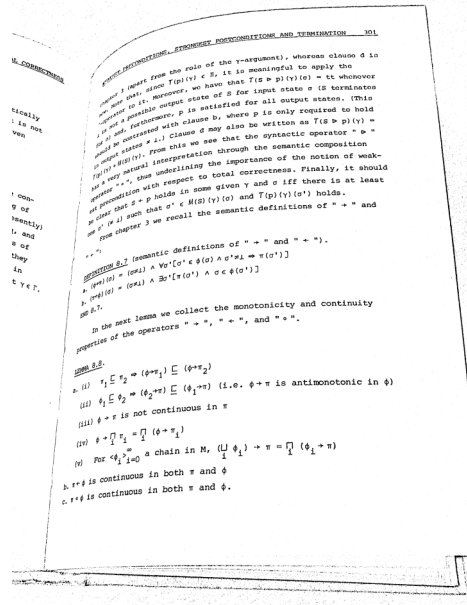
◆ Writing Order / Linear Structure

◆ Reading Order

◆ Usually one orientation

◆ Formalization needed

- Camera-based
 - ◆ Strong Geometric Distortion



- Video
 - ◆ Titles
 - ◆ Text in a Scene
 - ◆ Street Names

- **Difficulties of Feature Extraction**
 - ◆ Segmentation
 - ◆ Recognition in Images
 - ◆ Recognition in Document Images
- **Feature Extraction/Detection on the whole Signal**
 - ◆ Global Descriptor
 - ◆ Local Descriptor
- **Feature Extraction/Detection on Objects**
 - ◆ Object Localization / Extraction
 - ◆ Object Characterization
- **Interest of Multi-Resolution**
- **Using Features**

■ Similarity between Images

- ◆ Use image descriptors
- ◆ Compute similarity between descriptors

■ Algorithm

- ◆ Descriptor for each image
- ◆ Descriptors indexing
- ◆ Image request -> descriptor
 - ✦ Comparison with descriptors in database

■ Interest

- ◆ Simple
- ◆ Robust

■ Classical Descriptors for images

- ◆ Color
- ◆ Texture
- ◆ Shape

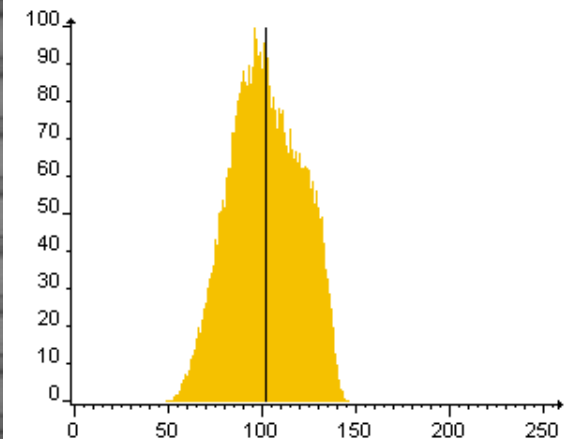
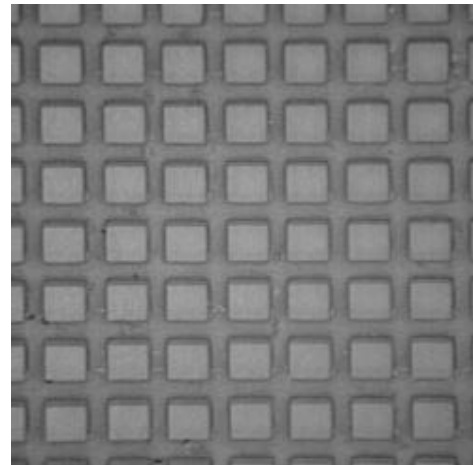
■ Color description using histogram

■ Black and white histogram

- ◆ # black pixels, # white pixels
- ◆ Not enough information

■ Grey level histogram

- ◆ # pixels for each level



■ Color histogram

- ◆ # pixels for each color $(r,g,b) \mapsto \text{Card}\{\text{pixel} = (r,g,b)\}$
- ◆ $256^3 = 16\,777\,216$ levels
- ◆ Reduction
 - ◆ 3 histograms for each canal (3x256)
 - ◆ 16 values by canal ($16^3 = 4\,096$)

■ Drawbacks of histograms

- ◆ Loss of geometry

- ◆ No invariance to

 - ✦ Translation

 - ✦ Scale

 - ✦ Illumination

■ What distance use ?

- ◆ To detect the closest image

■ Some examples of simple distances between histograms

◆ **2 histograms h and l** : $h = (h_i)_{i \in 1, n}$, $l = (l_i)_{i \in 1, n}$. $N = \sum_{i=1}^n h_i = \sum_{i=1}^n l_i$

◆ **Euclidian distance**

$$d(h, l) = \sqrt{\sum_{i=1}^n (h_i - l_i)^2}$$

◆ **Manhattan distance**

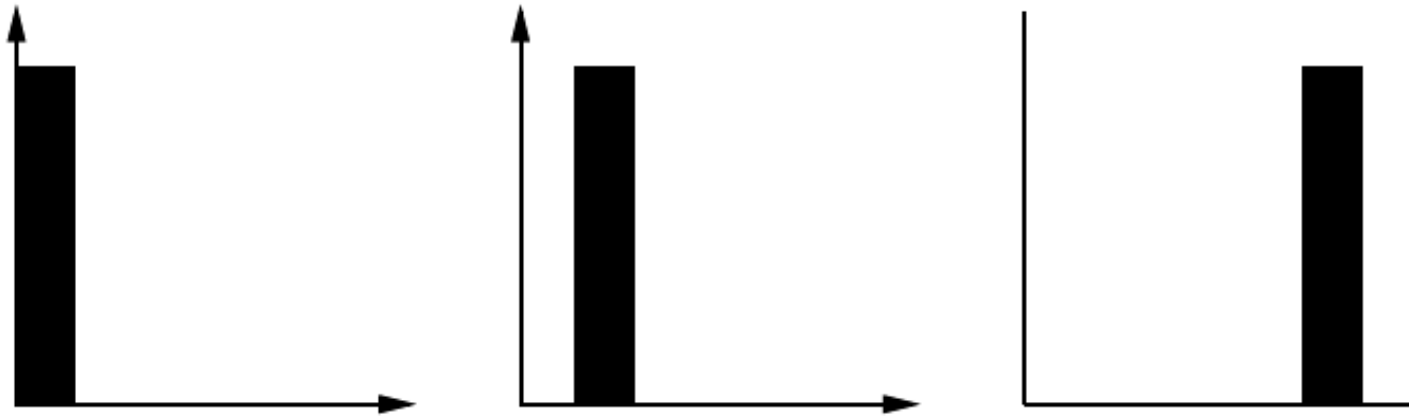
$$d(h, l) = \sum_{i=1}^n |h_i - l_i|$$

◆ **Max distance**

$$d(h, l) = \max_i |h_i - l_i|$$

◆ ...

■ Some problems



◆ Which one is the closest?

$$d(H_1, H_2) = d(H_1, H_3) \quad !!!$$

■ Evolved Distance

- ◆ Various versions: statistical, quadratic

- ◆ Most classical

 - ✦ EMD: Earth Mover Distance

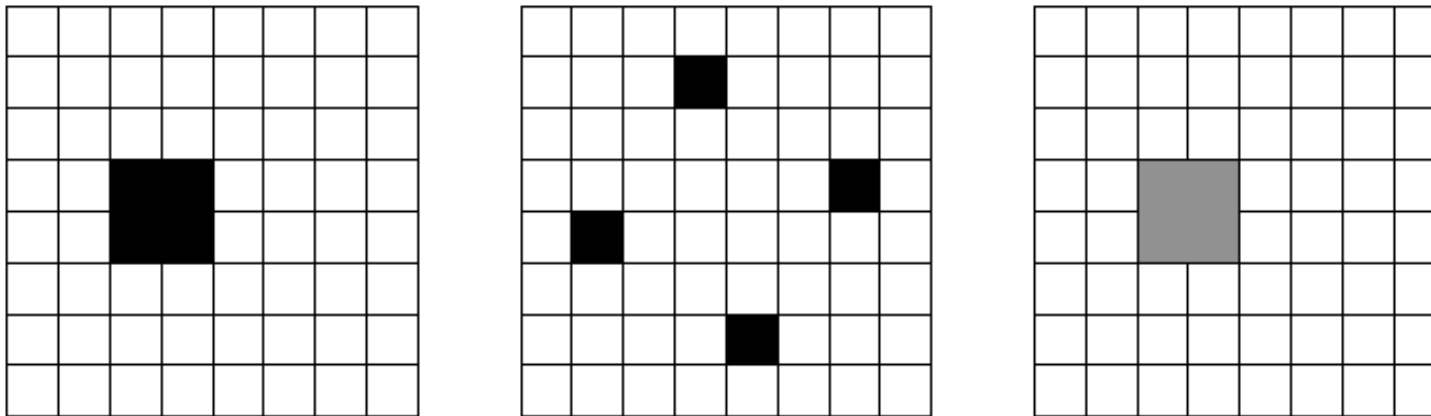
$d(h,l) = \text{minimal effort to go from } (h_i)_i \text{ to } (l_i)_i$

 - ✦ Needs an optimization phase

■ Performances of histograms

- ◆ Simple, fast, bad

- ◆ Left and middle images are closest for histograms



■ How characterize this?



■ Texture Descriptors

- ◆ Fractals coefficients
- ◆ Co-occurrences

■ Which distance?

■ Which scale?

- Haralick Co-occurrence matrix

- Image : $(m,n) \in D \mapsto I(m,n) \in 0,1,\dots,255$

- Counting co-occurrences (co-occurrences matrix)

$$P(i,j,d,0^\circ) = \# \left\{ ((k,l),(m,n)) \mid k=m, |l-n|=d, I(k,l)=i, I(m,n)=j \right\}$$

$$P(i,j,d,45^\circ) = \# \left\{ ((k,l),(m,n)) \mid (k-m=d, l-n=-d) \text{ ou } (k-m=-d, l-n=d), I(k,l)=i, I(m,n)=j \right\}$$

$$P(i,j,d,90^\circ) = \# \left\{ ((k,l),(m,n)) \mid |k-m|=d, l=n, I(k,l)=i, I(m,n)=j \right\}$$

$$P(i,j,d,135^\circ) = \# \left\{ ((k,l),(m,n)) \mid (k-m=d, l-n=d) \text{ ou } (k-m=-d, l-n=-d), I(k,l)=i, I(m,n)=j \right\}$$

- Normalization by the number of pixels couples with a d distance in the concerned direction

■ Haralick Co-occurrence matrix

■ Descriptor

- ◆ 14 formulas computed for each matrix for a distance d (usually 1 or 2)
 - ✦ Second moment angulaire, Entropie, Contraste, Correlation, Somme des carré, Moment de la différence inverse, Moyenne de la somme, Entropie de la somme, Variance de la différence, Entropie de la différence, Mesures d'information de la corrélation, Coefficient de corrélation maximale.
- ◆ 14 means
- ◆ 14 standard deviations

- Haralick's co-occurrence matrix have been published in 1973!
- Thousands of papers proposed since texture descriptors
- Which one use?

- What is a shape in an image?
- How to extract it?



- **Shape extraction**

- ◆ See Object Localization / Extraction section
- ◆ Very difficult in real scene images

- **Shape Characterization**

- ◆ See Object Characterization section

- **Limited to objects on plain background**