Earth Observation and Remote Sensing:

Why AI is needed?

Master AIC (Apprentissage, Information et Contenu) and D&K (Data & Knowledge) – Université Paris Saclay

Henri Maître - Télécom Paris
henri.maitre@telecom-paris.fr

October 2020

Objectives of this course

- The objective of this lecture is to give an introduction to the professional domain of Remote Sensing for engineers and researchers in the fields of Computer Science, Artificial Intelligence, Image Processing or Pattern Recognition,
- This objective will be reached by several sub-objectives:
  - To show how important Remote Sensing is. To show how diverse the application domains are with a survey of the most important fields of interest: agriculture, climate, environment survey and monitoring, defense, cartography, land use planning, …
  - To present the scientific context around Earth observation from satellite: positioning w.r.t. Earth, satellite trajectory, sensor capacities, acquisition rate, etc.
  - To inform about the diversity of satellite images: resolution, size, spectral bands, radiometric accuracy,

Objectives of this course (2/2)

- **This objective will be reached by several sub-objectives (continued)**
  - To show that Remote Sensing image mining is not the same problem that image retrieval on the web.
  - To present the main characteristics of satellite images which are used for most of the applications: textures, contours, lines and networks of lines, areas ...
  - To enlighten the role of scale and the role of semantics in the context of satellite image processing,
  - To clarify the role of time series
  - To show some early results obtained with Machine Learning and handcrafted primitive classification
  - To present the modern approach using deep neural networks
  - To show where difficulties and perspectives are.

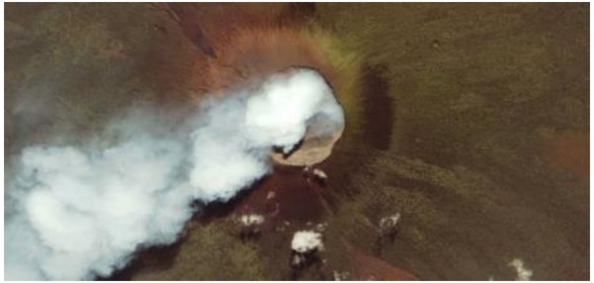
Content

Part I - Remote Sensing and Remote sensing images

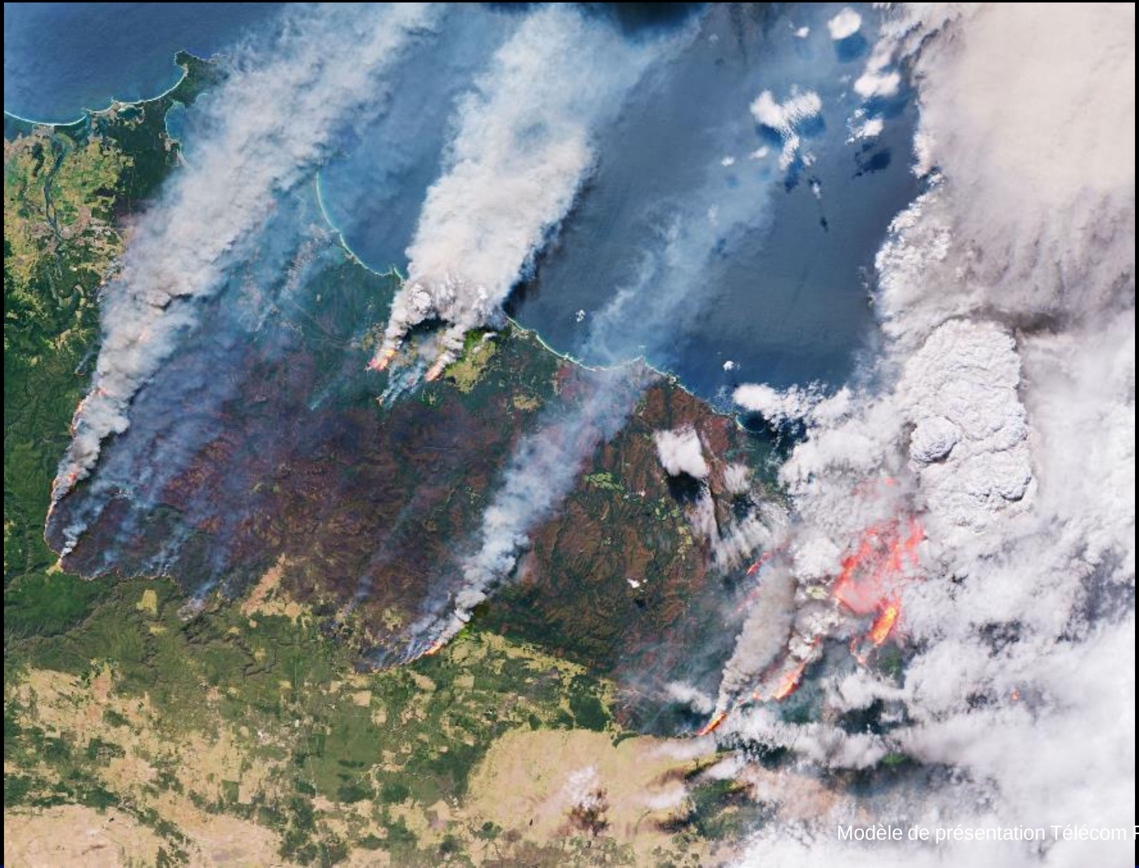
Why? How? For Whom?

Why do we need Remote Sensing

- **Environnement:**
  - **Meteorology: short-term weather prediction**
  - **Climate: long-term monitoring**
  - **GMES = Global Monitoring for Environment and Security: survey of natural and man-made catastrophies**
    - volcanos
    - earthquake, tsunamis, floods
    - Industrial hazards
    - Marine pollution



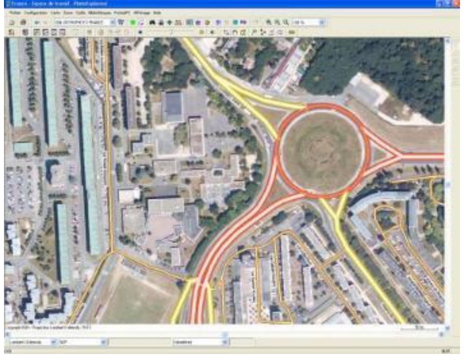
Australie : 13 décembre 2019, Sentinel 2



Why do we need Remote Sensing

- **Agriculture :**
  - Survey and evaluation of crop & farming production
  - Fish & Aquaculture resources management
  - Forestry resources planning
  - Water management, dams, watering
  - Desertification & urban pressure

Why do we need Remote Sensing



- **Town & country planning:**
  - Mapping and inventories
  - Constructions & public work: railways, airports, harbours, dams, …
  - Cities and Mega-cities management
  - Management of moving populations, displacements, installation
  - Climatic impact management
  - Crisis management: fires, floods, …

TELECOM Paris

IP PARIS

Why do we need Remote Sensing

- **Defense & Security applications:**
  - **Military deployement preparation**
  - **Military mission debriefing / damage survey**
  - **Intelligence and survey of national/foreign territory**

How is prepared a remote sensing program

Remote sensing mission/program

- Where the vocabulary is given: launcher, control station, ground station, altitude, orbit, geostationary, traveling, revisit time, spectral range, atmosphere window,
- The image parameters: resolution, swath, channel number,
- Difference between passive (optical wavelength range) and active (Radar) sensors

How is prepared a remote sensing program

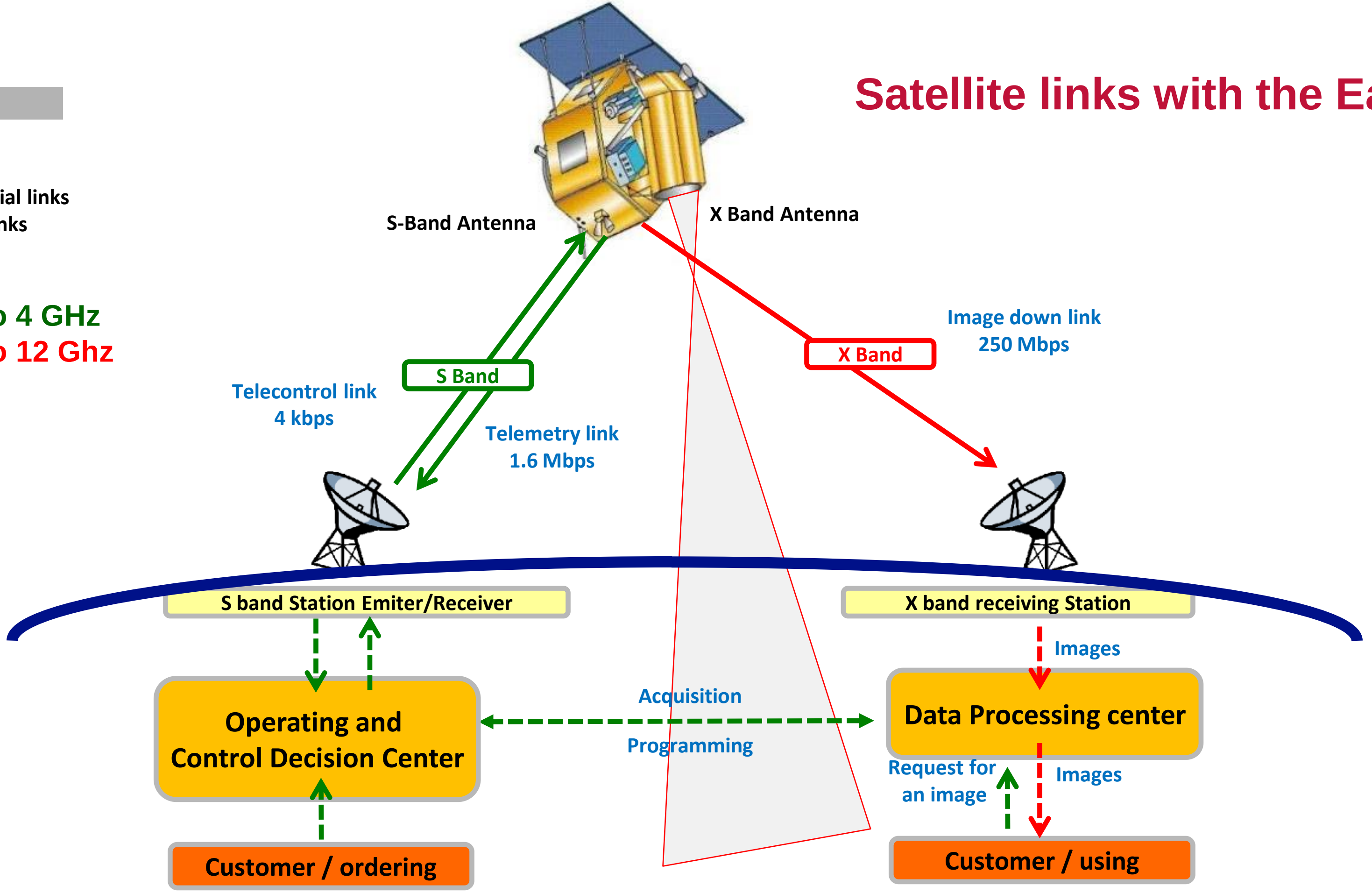
- **Conceive the sensor:** application, customers, scientific and technological issues, financial issues
- **Determine which satellite / which launcher**
- **Conceive the ground-station and the data management process :** economical, social and technical issues

→ **15 to 20 years**

Satellite links with the Earth

- - - - Terrestrial links
——— Aerial links

S-band : 2 to 4 GHz
X-band : 8 to 12 Ghz

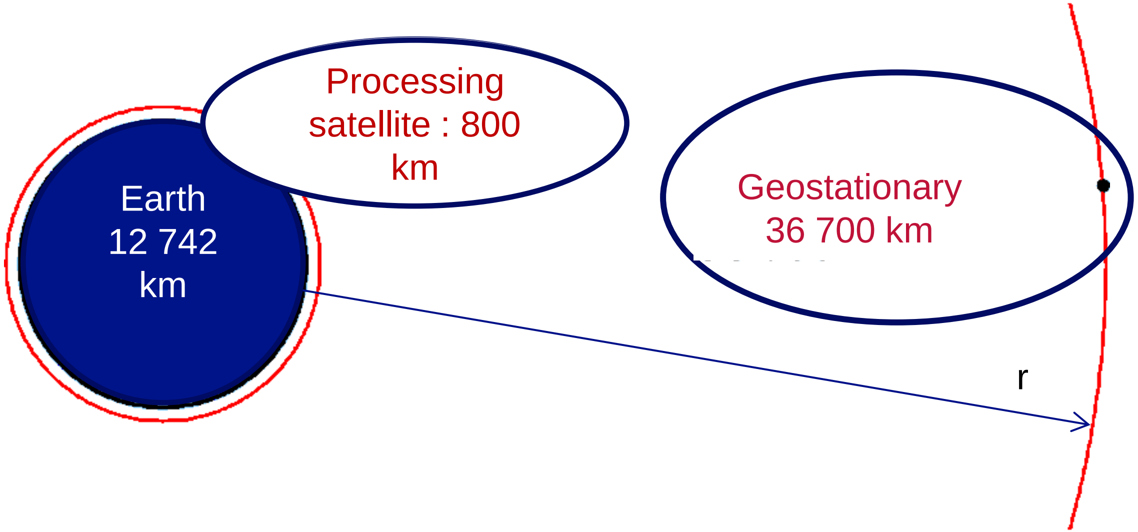


Satellite : orbit choice

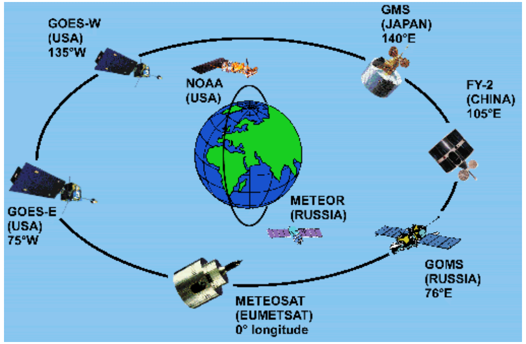
$$\vec{F} = -\frac{\mu\, m}{r^2}\frac{\vec{r}}{r}$$

- **Mecanics laws:**
  - **Newton = centripetal force**
  - **Satellite speed = driving force**

    ➔ **elliptical or circular trajectory (Kepler)**

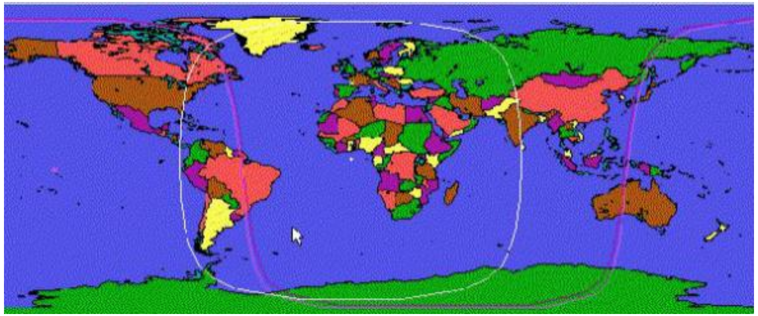


Choice of Orbit



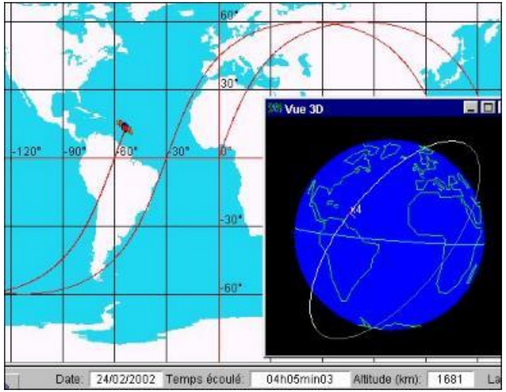
- **1) Geostationary**
  - **Always in the Equator Plane**
  - **Always at vertical of the same point on the Equator**
  - **Altitude ~ 36 700 km**
  - **Field of view: ~1/3 Earth: always the same**

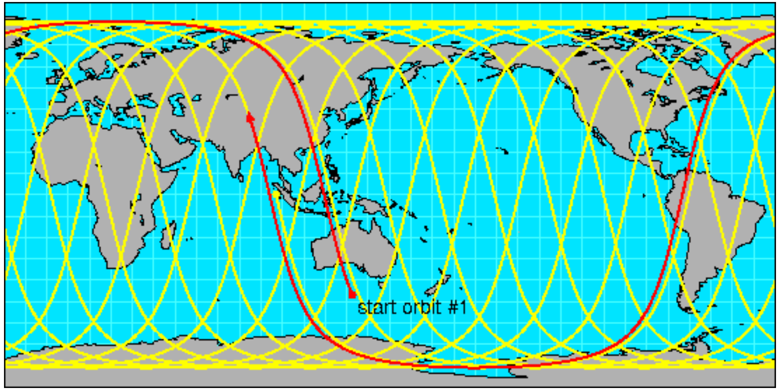
  - **Applications : meteo, survey of catastrophies, telecoms, TV**



Orbit choice

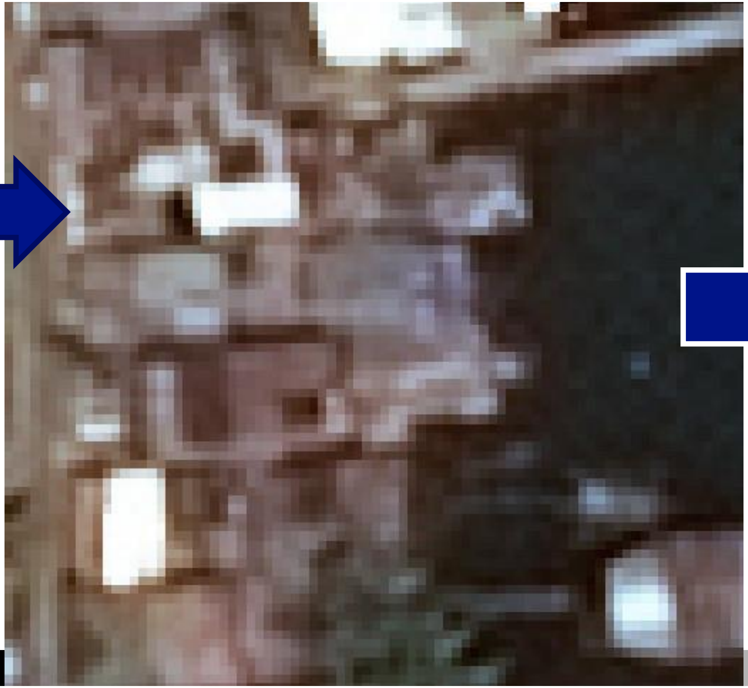
- **2) Processing satellite (low orbit)**
  - **Altitude ~ 800 km (down to 250 km)**
  - **Circular ~ N/S**
  - **Trajectory : ± polar**
  - **~ 15 revolutions / day**
  - **Helio-synchronous**





Choice of resolution

- **Pixel size = smallest measured terrain on the ground**
  - **from 30 cm to 10 km**



SPOT 5

Δx = 2,5m

On Ground resolution

- **Depends on:**
  - **Sensor :**

    Photosites size: δx

    $G = \frac{f}{D}$ = **enlargement**

    $\Delta x = \frac{\delta x}{G}$ = smallest detail

  - **The camera lens**

    $\delta' x = \frac{\lambda f}{d}$ = **diffraction limited resolution**

$$\Delta x_{min} = \frac{\lambda f}{Gd} = \frac{\lambda D}{d}$$

➔ Smallest detail

D = satellite-Earth distance
~ 1 000 km = 10^6 m

λ = wave length
= 0,5 . 10^{-6} m

d = lens diameter
~ 0,5 m

$$\boldsymbol{\Delta x_{min} = 1\ m}$$

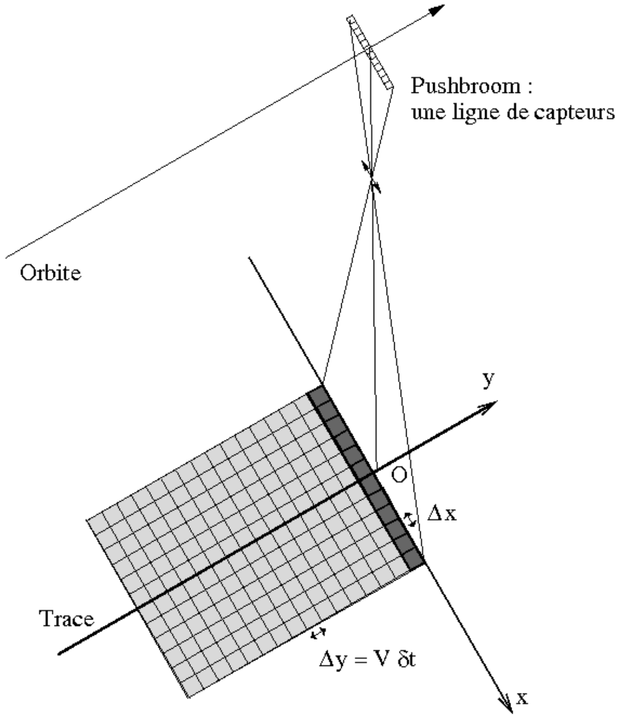
Possible with : f = 1 m

if $\delta' x = \frac{\lambda f}{d}$ = 1µm

the photosite measures 10^{-6} m

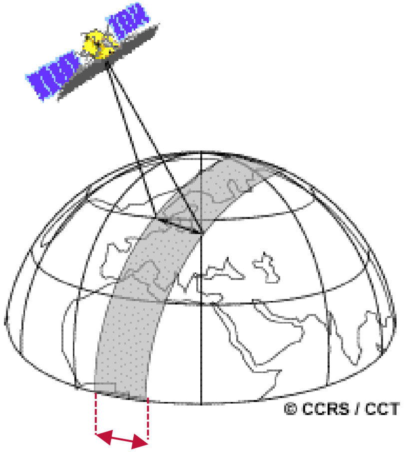
Often *push-broom* *sensor*

- **Sensor size along track:**
  - **On line sensor**
  - **= speed x aperture time**
- **In the other direction**
  - **Number of sensors on a line**
  - **from 6 000 to 40 000**
- **Resolution :**
  - **Depends on the lens**

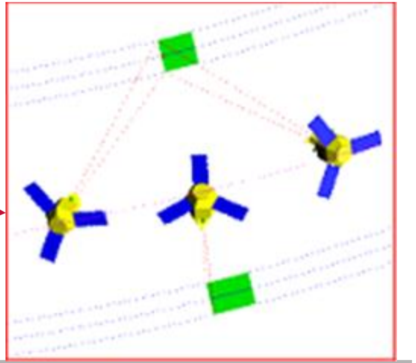


Swath choice

- **Swath = image width**
  - **from 10 km to 10 000 km**
  - **= from 3 000 to 40 000 pixels / line**
  - **Given by the sensor size**
  - **Limited by the communication link with Earth**

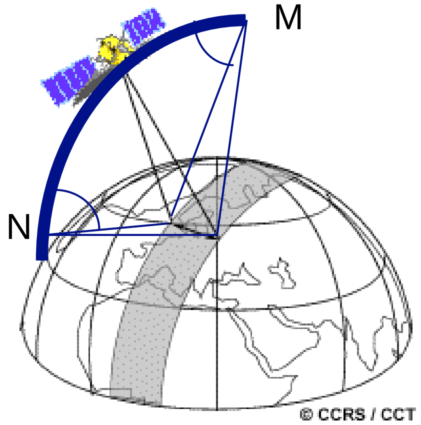


- **Revisit delay**
- **15 min for geostationnary sat. (to dump the memory)**
  - **from 1h30 (min) to 1 month for processing satellites**
  - **But … sensor agility!**



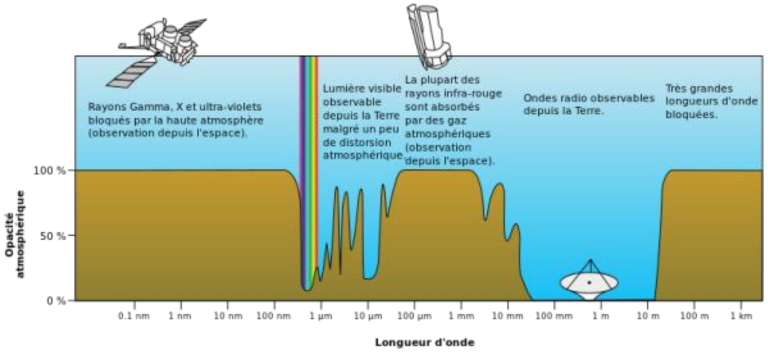
Video facilities ?

- **Angle of view ~ + or – 50 degrees:**
  - **MN ~ 2000 km**
  - **1 rotation around the Earth = 90 min**
    ~ 40 000 km
  - **Time to go from M to N**
    = 90*2000/40000 = 4 min 30 s

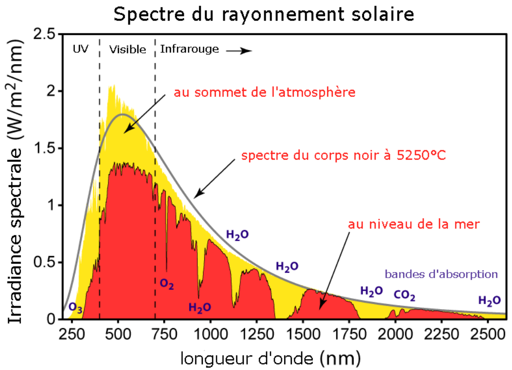


Which wavelength?

- **1 – Passive sensors: measure the energy sent back from Sun by Earth or the energy radiated by Earth**
  - Emitted from the Sun (Wien's law) x   Atmosphere transparency x Ground Reflection
  - Black and White (Panchromatic)
  - Visible   =   Blue - Green - Red
  - Visible and Near Infra-Red : G - R - IR = false colors
  - Multispectral   :   7 ➔   20 channels
  - Hyperspectral   :   64 ➔ 512 channels



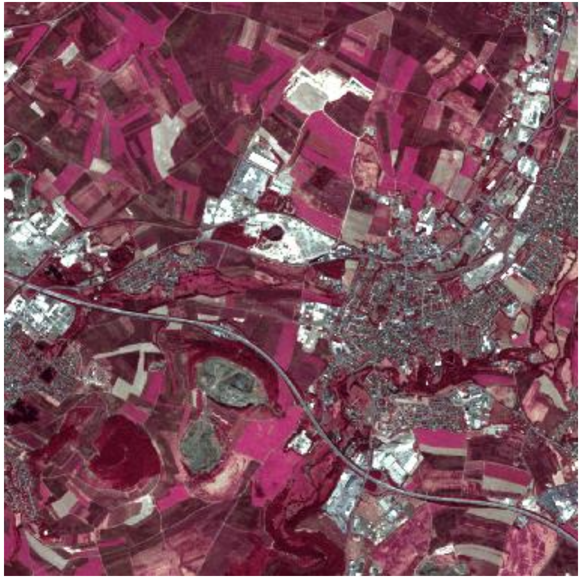
© Wikipedia



© Wikipedia

False colors : NIR-R-G → R-G-B

vegetation = red



False colors

True colors

Multispectral image visualisation: pseudo colors

Landsat = 7 channels

1
2 R
3
4 G
5
6 + B
7

41(7+5)

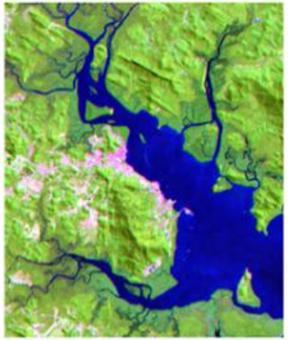
321

(a) combinaison 321

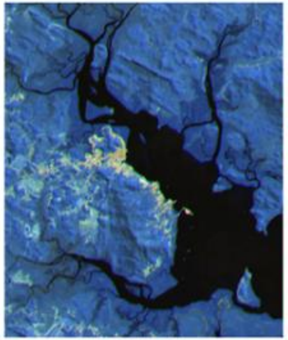
432

(b) combinaison 432

542

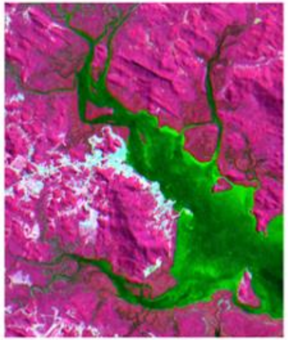


(c) combinaison 542



(d) combinaison 754

754



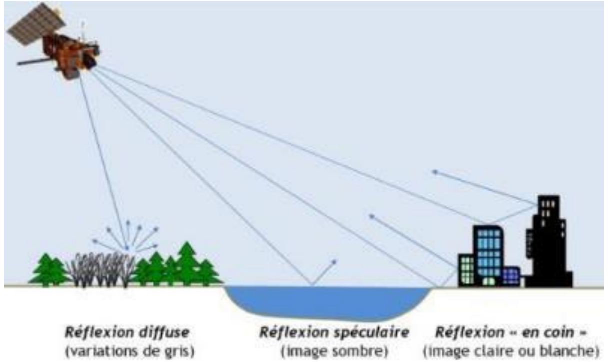
(e) combinaison 435

435

Which wave length?

- **2 – Active sensors: EM emitter + receiver**

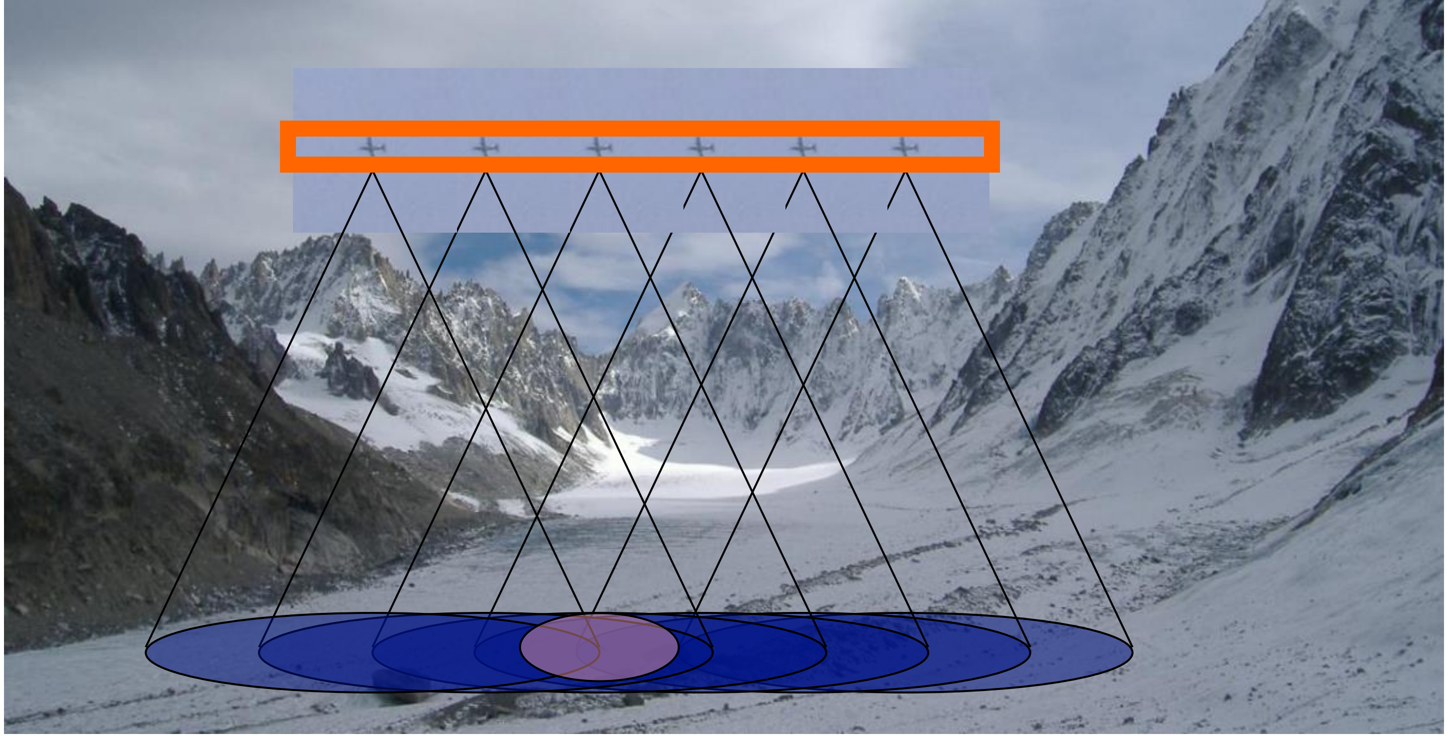
  radar = Micro waves: λ= 1 cm to 10 m



- But low resolution : $\Delta x = \frac{\lambda f}{Gd}$
- **With complex processing: SAR = Synthetic Aperture Radar ➔ hi resolution**

Real antenna is too small, it covers a very large field

When moving, superposition of the seen areas



One point is seen from several antenna positions

From computation we obtain an accurate information = synthetic antenna

Satellite images = big data !

Television HD	1 280 x 720 pixels
Television 4k	4 000 x 2 000 pixels
PC display screen	1 600 x 1 200 pixels
Photo camera	5 000 x 4 000 pixels
Spot 1 … 4	**6 000 x 6 000 pixels**
SPOT 5	**24 000 x 24 000 pixels**
Quickbird	**40 000 x 40 000 pixels**

1 600 000 000 pixels = 1,6 Gpixels
= 800 PC display screens

1 SPOT 5 image = 10 s of satellite observation

Satellite images for the customer

- " 1 image" =
  - several images (1 image = 1 channel) or 1 image (1 pixel = several values, ! For each channel)
    - ✓ 1 channel = panchromatic
    - ✓ 3, 4, …7 = multispectral
    - ✓ 32 … 256 = hyperspectral
  - ancillary data
    - ✓ Date & time, sun position
    - ✓ Geographic position of image center, Satellite position
    - ✓ Cloud cover, atmospheric conditions
    - ✓ Sensor calibration

Satellite image for the customer

- **Several levels of processing (de-pends on the satellite) for instance**
  - **Level 0**
    - ✓ Raw data as issued from the satellite, on board geometry (equi angle from the satellite positioning), no photometric correction, correction of satelite mvt
  - **Level 1**
    - ✓ Registration by projection on the geoid, Equalisation of sensors,
  - **Level 2**
    - ✓ Accurate registration on a map using a Digital Terrain Model (DTM), Correction of atmospheric effects
  - **Ortho correction**
    - ✓ Very accurate registration on a map using a Digital elevation model (DEM)

Image corrections

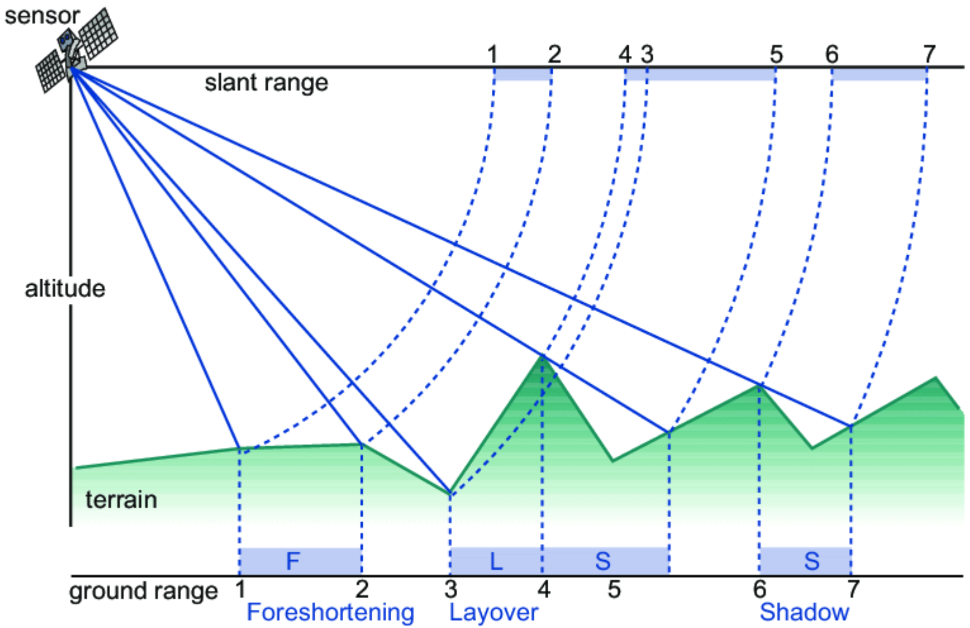
Radiometric

- **Sensor homogeneity or time drift:**
  - Calibration on known areas: Nevada, Atacama, Sahara, Crau)
  - Use of target stars
- **Atmospheric corrections**
  - Depending or not on meteorological data
  - Taking into account the position of the pixel in the swath
- **Radiometric compensation of Sun/Satellite angle**
  - Using a reflectance terrain model

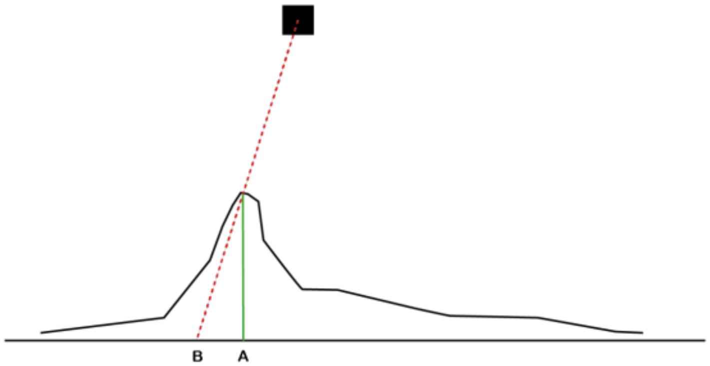
Geometric corrections

- **Roll, pitch and tossing of the satellite**
  - Internal consistency of the image
- **Projection of the image on the average altitude geoid**
  - Using the X,Y,Z,t positions of the satellite
  - Using Ground control points
- **Using a DTM to correct the projection from the terrain altitude**
  - ➔ georeferenced images
- **Using a DEM to take into account the man-made constructions**
  - ➔ Ortho image

The role of geometric corrections

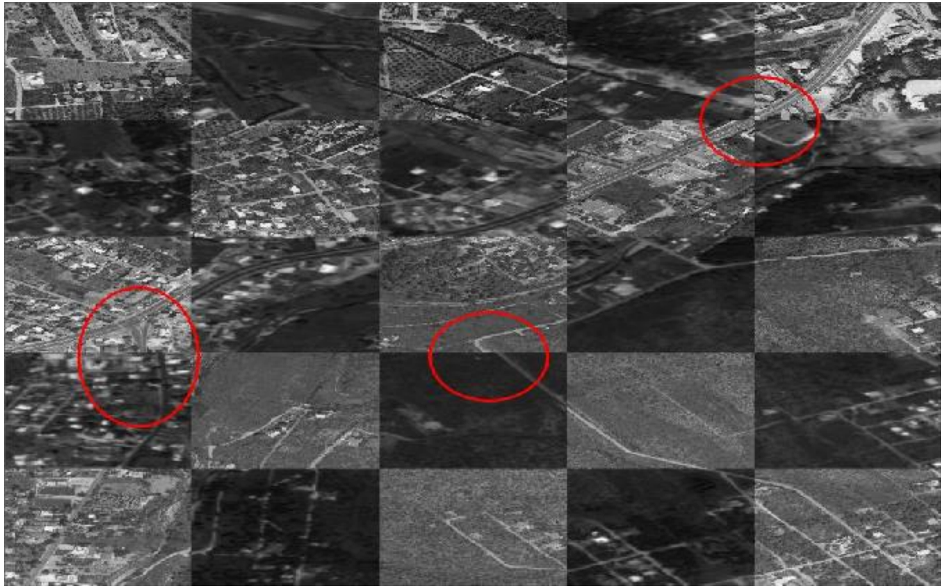


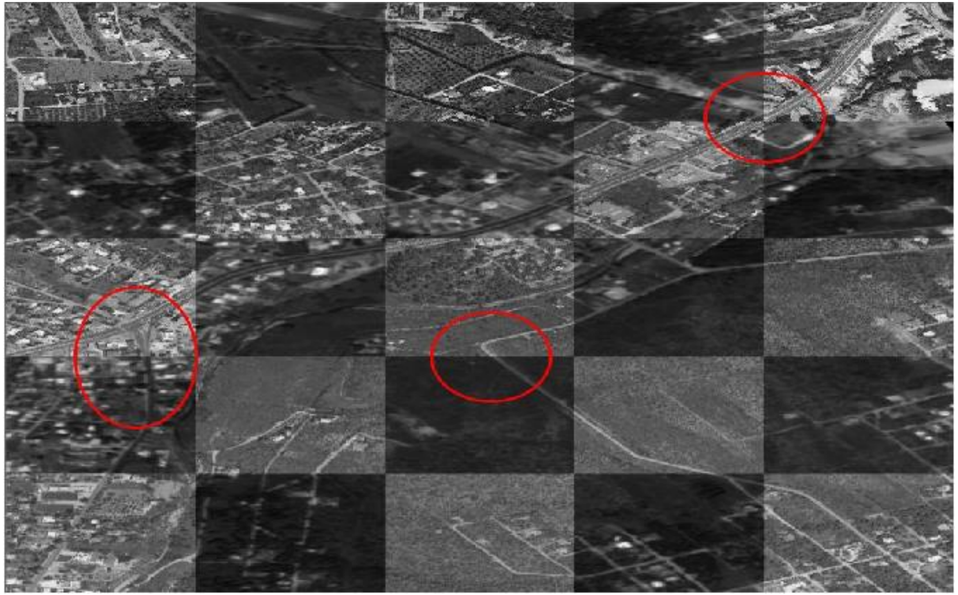
Defects of raw satellite geometry



Use of a DTM to correct an image

Coarse vs. fine registration – mosaic presentation





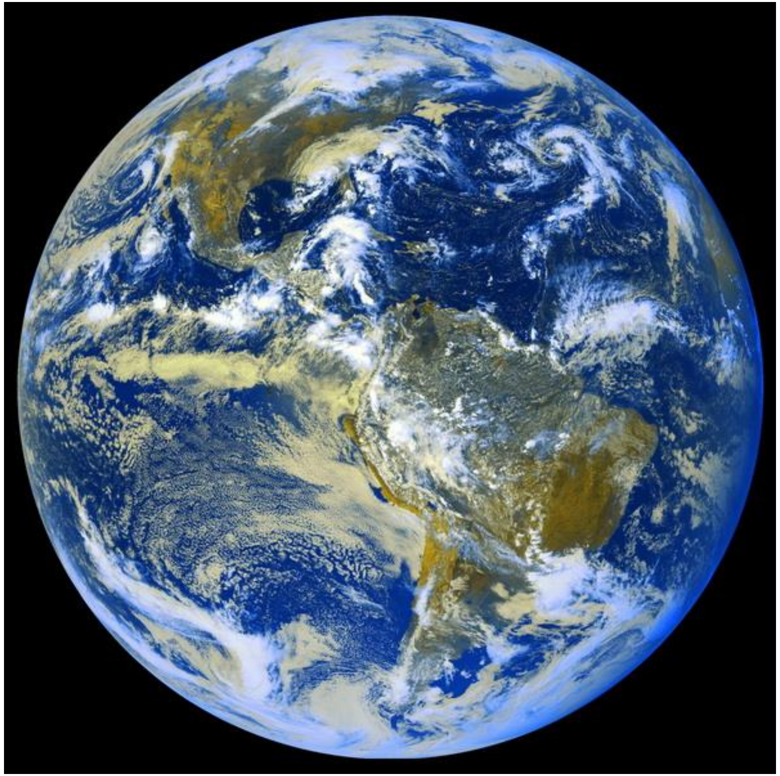
Worldview-2 image & aerial photography before and after fine registration
Copyright Karantzalos et al.

Diversity of Remote Sensing Images

Diversity of images

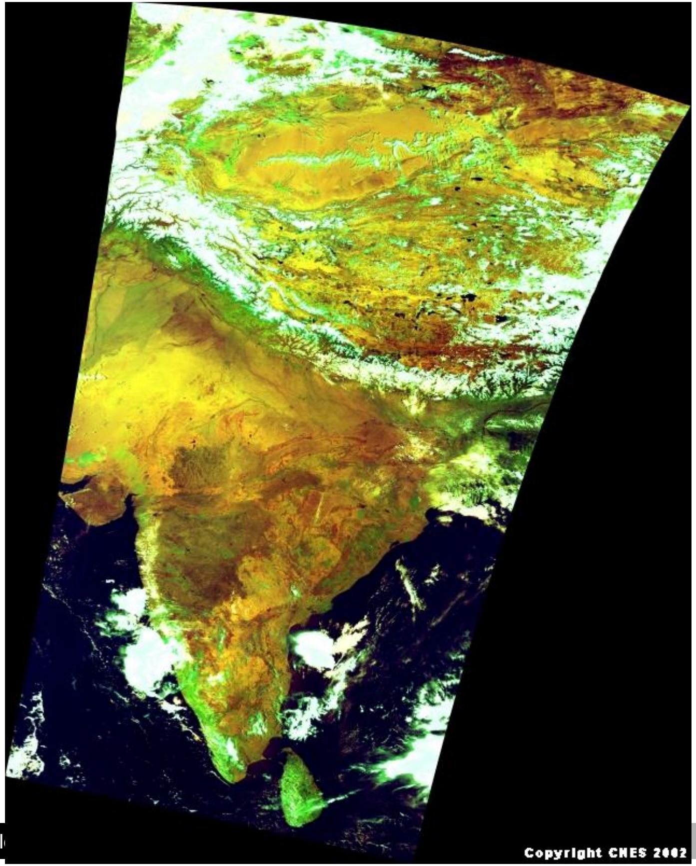
- We present several images issued from different satellites with very different characteristics.
  - The main difference comes from resolution and field of view
  - Another difference comes from the functional objective of the images: agriculture, meteorology, defense, land use planning
- As a result of technology evolution, the surveyed data change from clouds, crops, forests to cities and buildings, from highways to small streets.

Meteo satellites: very low resolution



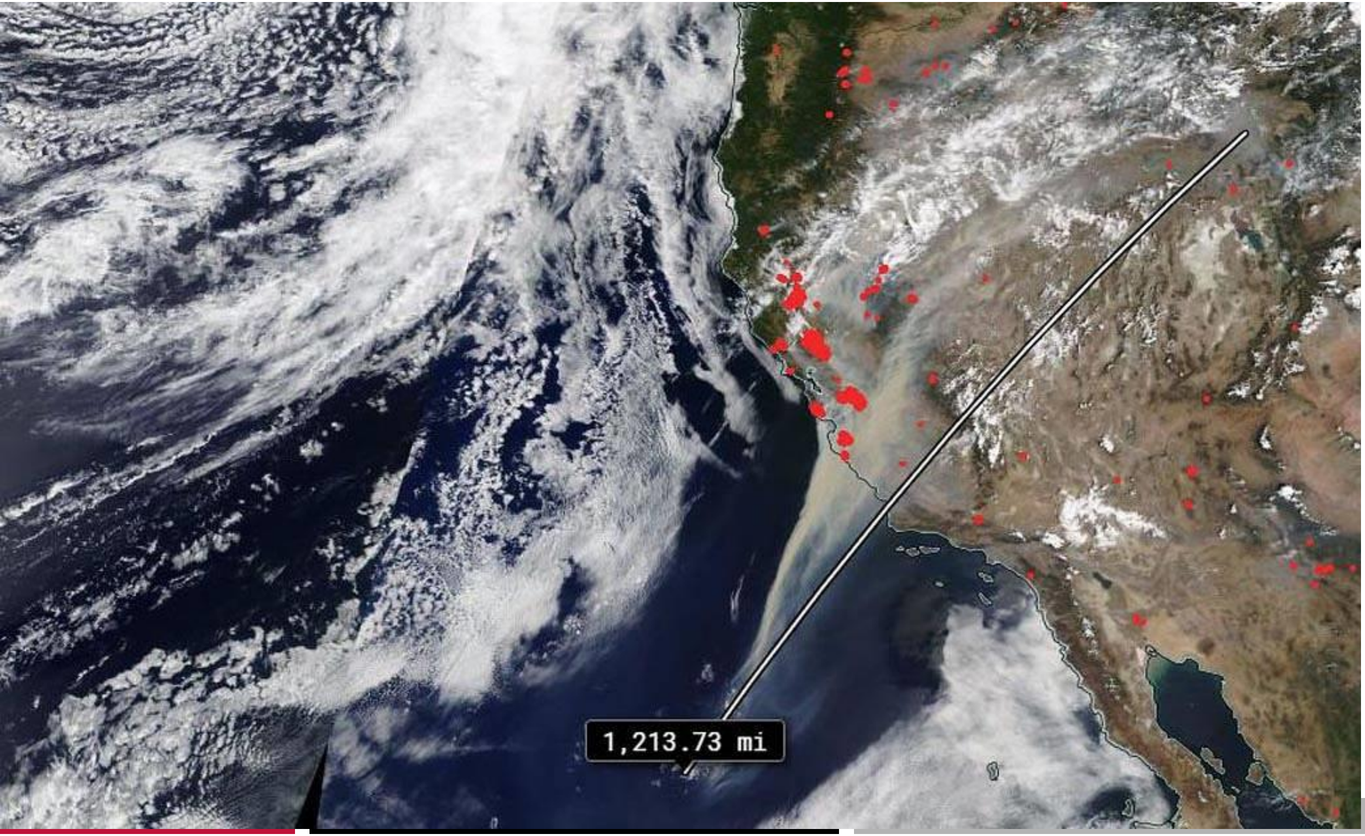
Meteosat = 3 km

Climate/environnement: low resolution



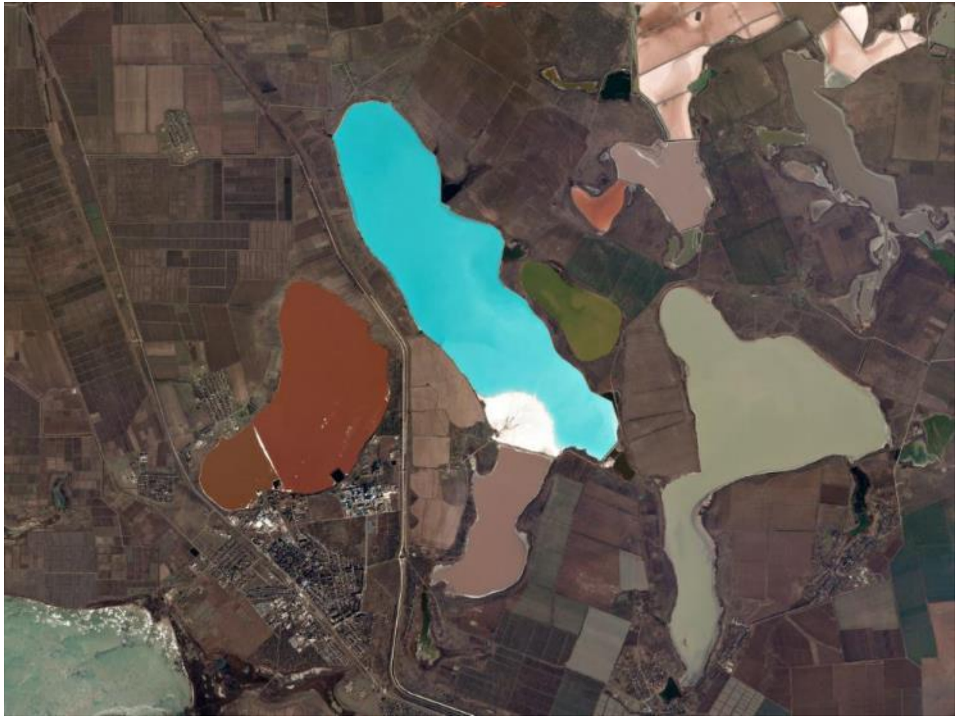
INSAT = 2,2 km

Climate/environnement: middle resolution



Modis Terra Images = 1 km
From Idaho to Pacific Ocean
Aug. 20, 2020

High resolution: Planetscope

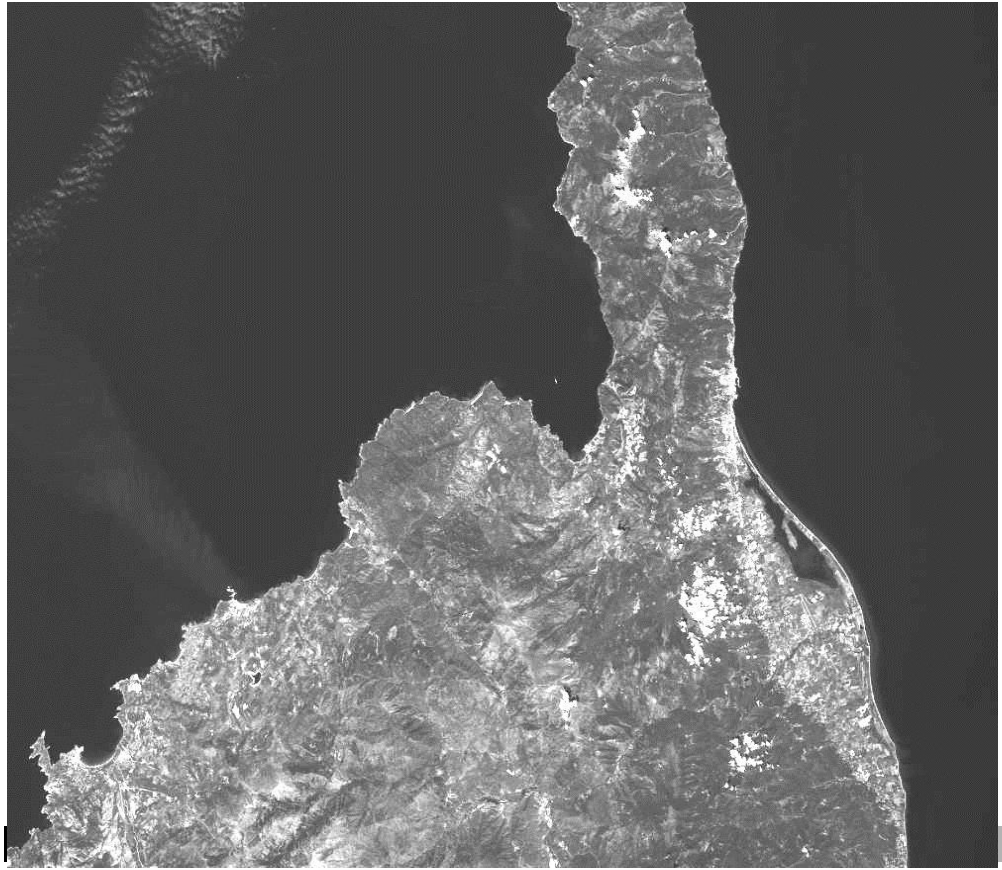


Planetscope :
Krasne Hypersalted Lake
- Crimea

Multispectral
= 4 m

175 satellites
300 Mkm^2 / day
= 2/3 Earth

SPOT 5 : high resolution ; pixel = 2,5 m



Very high resolution

Pléïades :
Bora-Bora

Panchro
= 0,70 m

Multispectral
= 2, 8 m

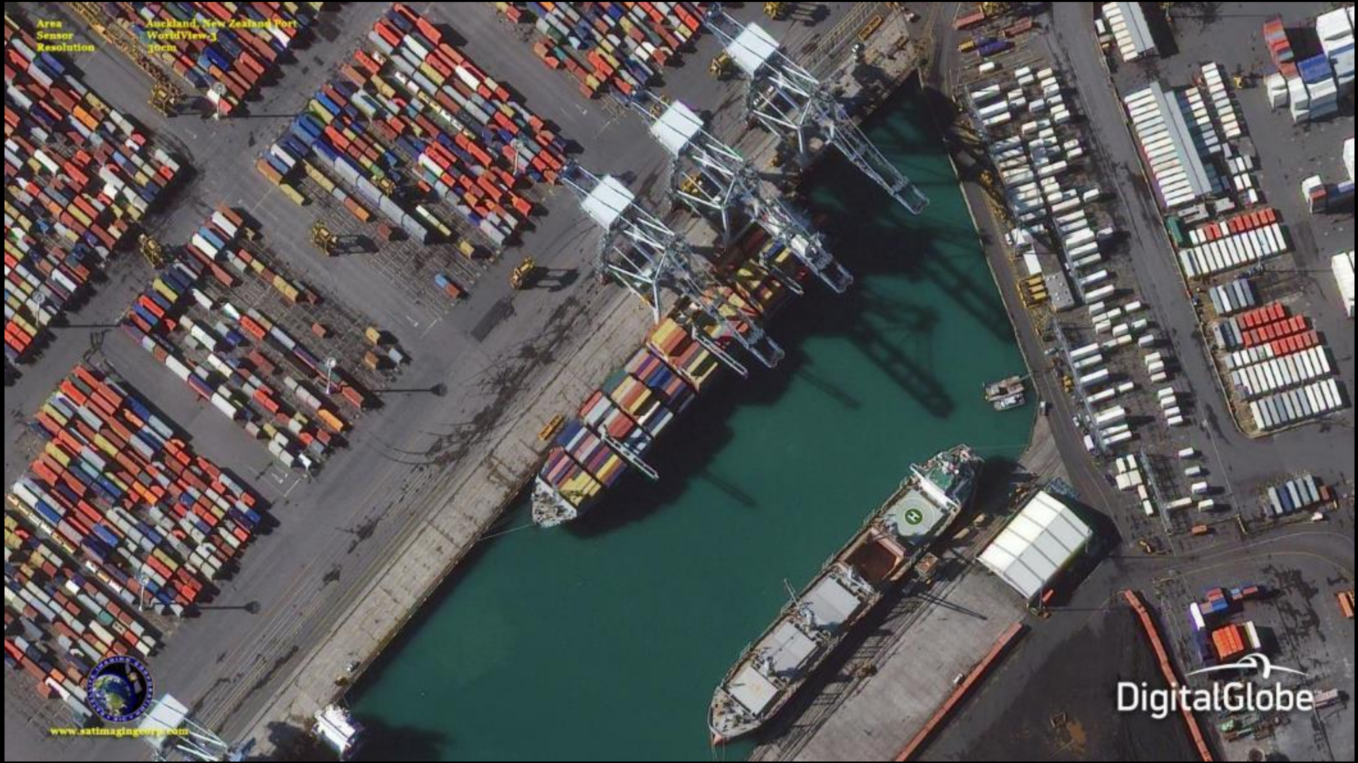
Very high resolution: Quickbird

Panchro
= 0,61 m

Multispectral
= 2, 4 m

Pléïades : Mont Saint Michel

Auckland New-Zealands

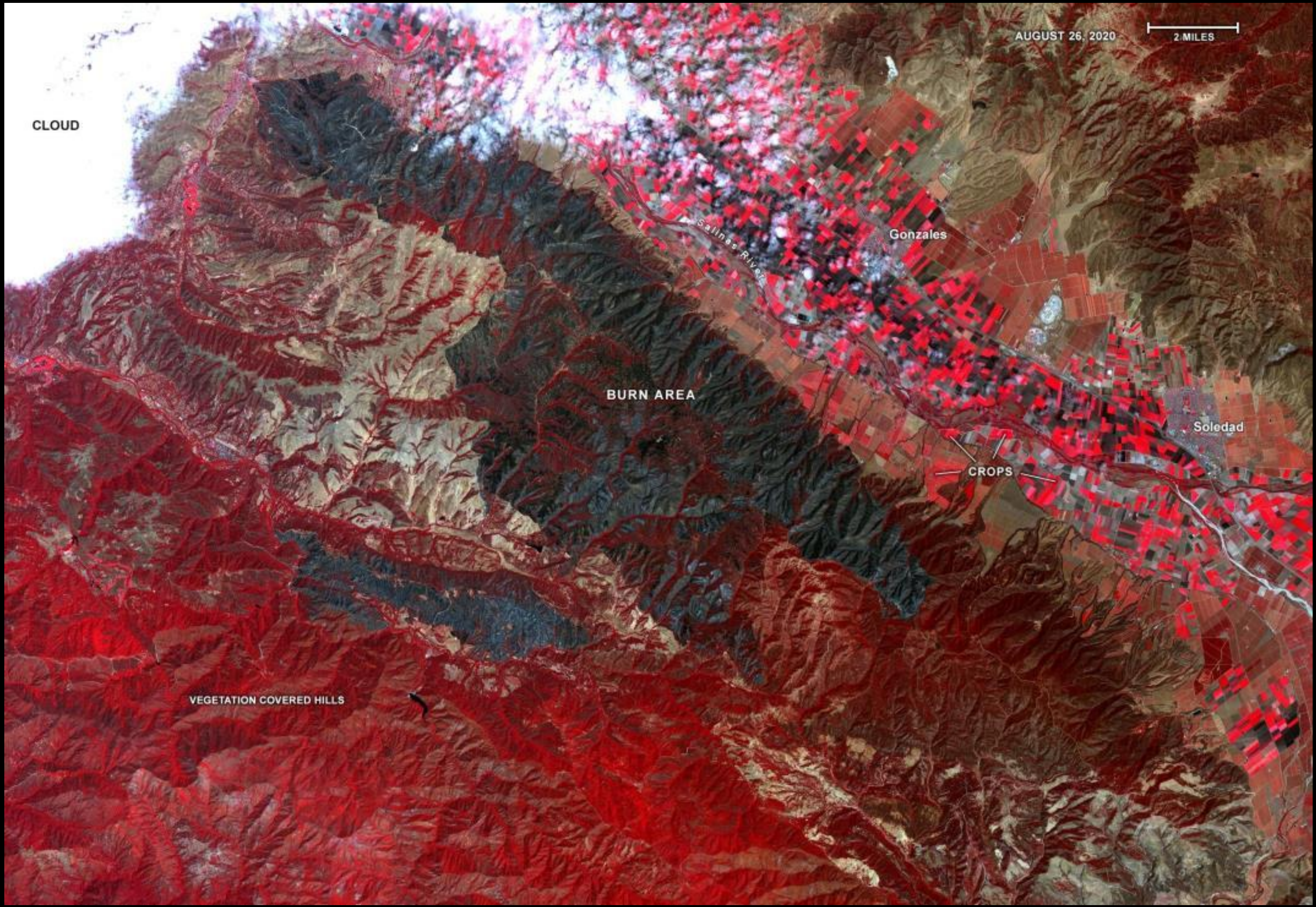


WorldView : King Abdullah Petroleum Center

Ikonos – Lebanon: agriculture

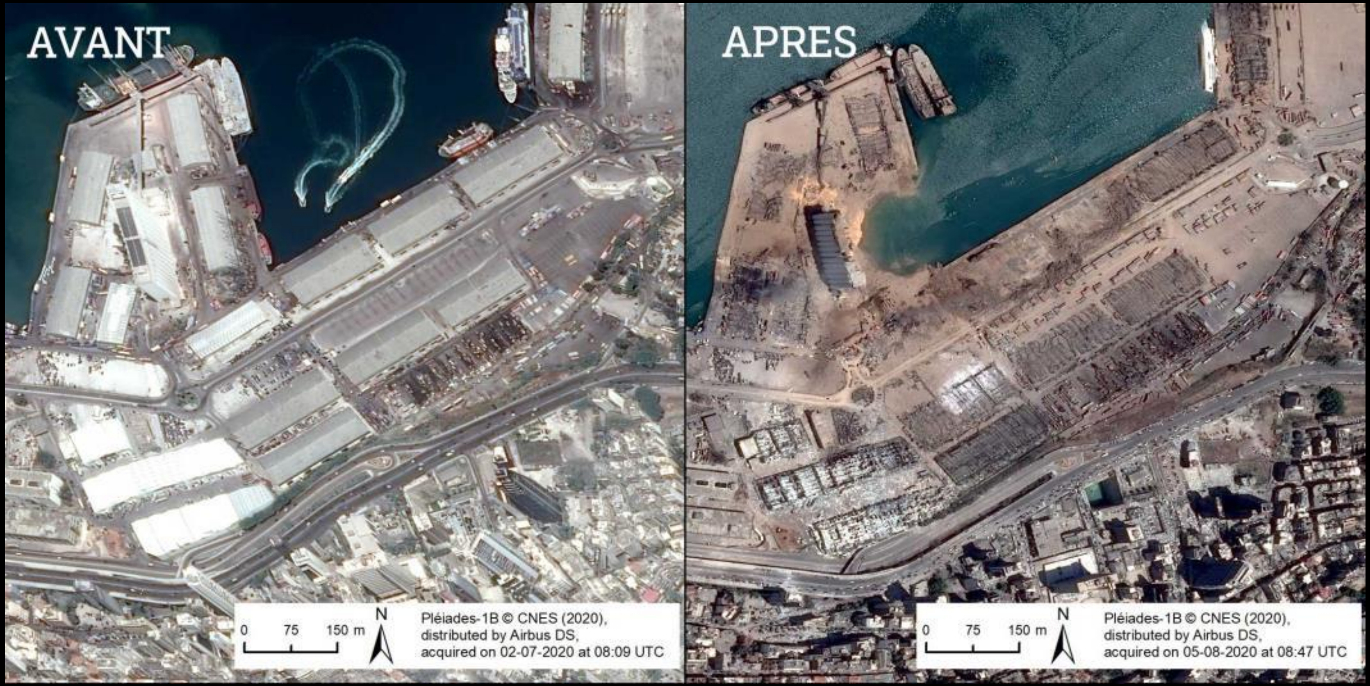
WorldView : Bayan Mines (China)

Thermal sensor ASTER – California - 26 aug 2020



Temporal evolution: Baikal Lake with SPOT

Pleiades – Beyrouth – Liban - 8 aug 2020



Part II – Remote Sensing Image Mining

Remote Sensing Imaging: Archiving Problems and Issues

Remote sensing imaging IS big data

- Hundreds of satellites, each with tenth of thousands of images, each with tenth of millions of pixels
- A huge problem … storage of data ➔ refreshing storage subject to technological evolutions: tapes, discs, VLSI
- Additional problem: where is information?
- Solution: Image mining
  - Has been developed since about 2000, firstly with classification of handmade features, then more successfully with deep neural networks (DNN)
  - DNN are end-to-end solutions ➔ Blind techniques, not yet ‘’explainable” . They are still under development and far from being stabilized for remote sensing applications.
  - Handmade features are much more ‘’explainable”, they are well adapted to man machine interaction and human supervision. We will spend more time with them

Satellite Image archives

- How can we store millions of images?
- How can we ensure durability of storage?
- How knowing that information exists?
- How retrieving information?
- How exploiting information?

Not treated here

➔ Data Mining directly on image files

When searching in a small set of images

➔ Indexing images when received
➔ data mining on index

When searching in large sets

RS Image mining IS NOT MultiMedia Image Mining

Multimedia vs. Satellite

- Image retrieval on the web (Google-like) is very efficient and most used. Is it possible to use it for satellite images?
- Efficient techniques for image retrieval on the web (called here ‘’Multimedia images”) are based on semantic descriptors attached to the image. These descriptors do not exist for satellite images.
- Multimedia image retrieval looks for ‘’exact” retrieval. Satellite image retrieval looks for ‘’similar configurations”. → specific techniques with specific metrics have to be developed.

Mining in Multimedia Image databases

- **Multimedia information retrieval :**
  - **Either from semantic information: name, description, caption, text**
    (90 % of Google-like retrieval)
  - **Or from instance (i.e. with a reference image)**
    (Face or fingerprint recognition) ➔ converted to symbolic (list of nodes)

- **I – ''Classical'' Machine Learning techniques (2000-2012)**
  - **Hand-crafted feature detection and/or salient point detection**
  - **Classification in p-dimensional space (Bayes, k-NN, hierarchical clustering, Random Forrest, SVM, …)**
    - ➔ few parameters
    - ➔ few learning images (groundtruth) ~ 1000

- **II – Deep neural networks (2012 - …)**
  - **Directly with images as input and/or with extracted features**
  - **Several +/- linear classifiers in cascade**
    - ➔ thousands of parameters
    - ➔ hundred of thousands of images as groundtruth

Multimedia image mining: handcrafted features + classification

- **Multimedia information retrieval from instances:**
  - **Choices: to be robust wrt possible differences**
    - scale, lighting, orientation, color, … ➔ invariance
  - **Strategy: detect invariant features**
    - Histograms, color distribution, area-based segmentation, graph description, …
    - Textures
    - Salient point detection: Harris, SIFT, SURF, …
  - **Represent the image as a vector in a *p* dimensional space $\mathbb{R}^p$**
  - **Classification : Bayès, k-NN, dynamic clustering, SVM (Support Vector Machine), Graph-tree, random forrest…**

Salient points: SIFT

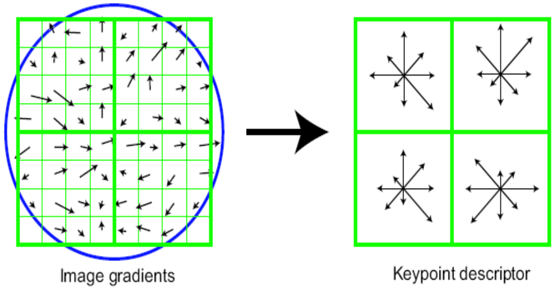
$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y),$$

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G.$$

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma). \end{aligned}$$

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$

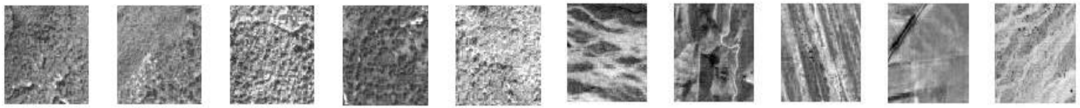
$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}. \qquad \frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r+1)^2}{r}$$



Specificities of RS Image mining

Category-based retrieval in specific data-bases

- Mostly attached to specific domains:
  - Biomedical
  - Biology
  - Astronomy
  - Remote sensing and satellite images
- Goal: to retrieve images « looking the same » as a given sample in very specialized data-bases



- Different from : retrieving the exact object in a very broad data-base

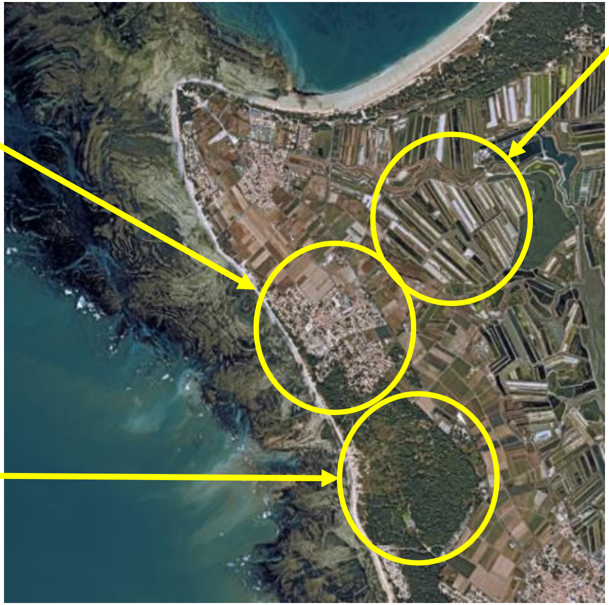
A satellite image as a mosaïc of textures

- **A very specific content**

Fields

City

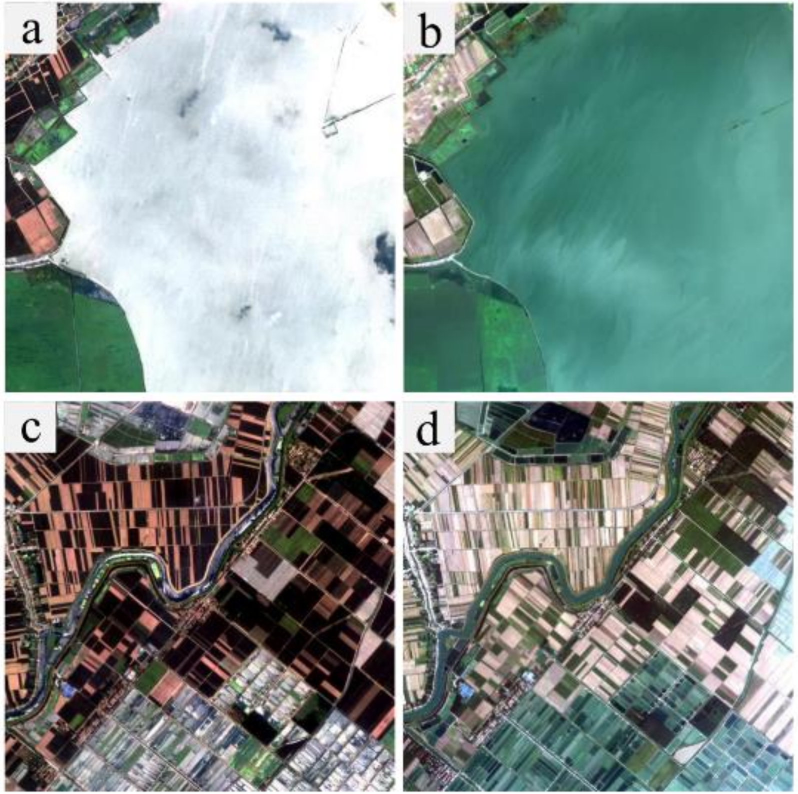
Forest



But … a same region may provide different images

Meteorological variations

Seasonnal agricultural variations



From : Tong et al.
arXiv 1807.05713 - 2018

The role of scale

15 m

1 m

High-Badakchan, Tadjikistan - Ikonos

Main scales

- **< 1 meter = Very high resolution :** fine details in urban context, roofs, chemneys, cars, pedestrians, zebra crossings, containers, fences, small boats, … Ikonos, Pleiades, QuickView
- **1 m < … < 5 m = High resolution :** urban structures, houses, streets, gardens, individual trees, railway & road networks, … SPOT 5
- **5 m < … < 30 m = Middle resolution:** fine landcover, coarse urban structure: dense urban, residential or commercial areas, Landsat, Spot 1-3
- **> 30 m = low resolution:** global landcover

Available additional information on satellite images (semantic information) = Ancillary data

- **Accurate positionning in universal geographical references:** UTM, Mercator, Lambert, etc.
- **Precise time referencing:** seasonal variations (vegetation, insolation, agricultural production, …), sun positionning (shadows), tide effects (precise coast-line, harbours and fishering activities), meteorological conditions (snow, floods, …)
- **Satellite parameters:** resolution, spectral sensitivity, noise, on-board callibration, roll pitch
- **Often:** Image quality: cloud cover, smokes, …

Satellite image indexing is difficult

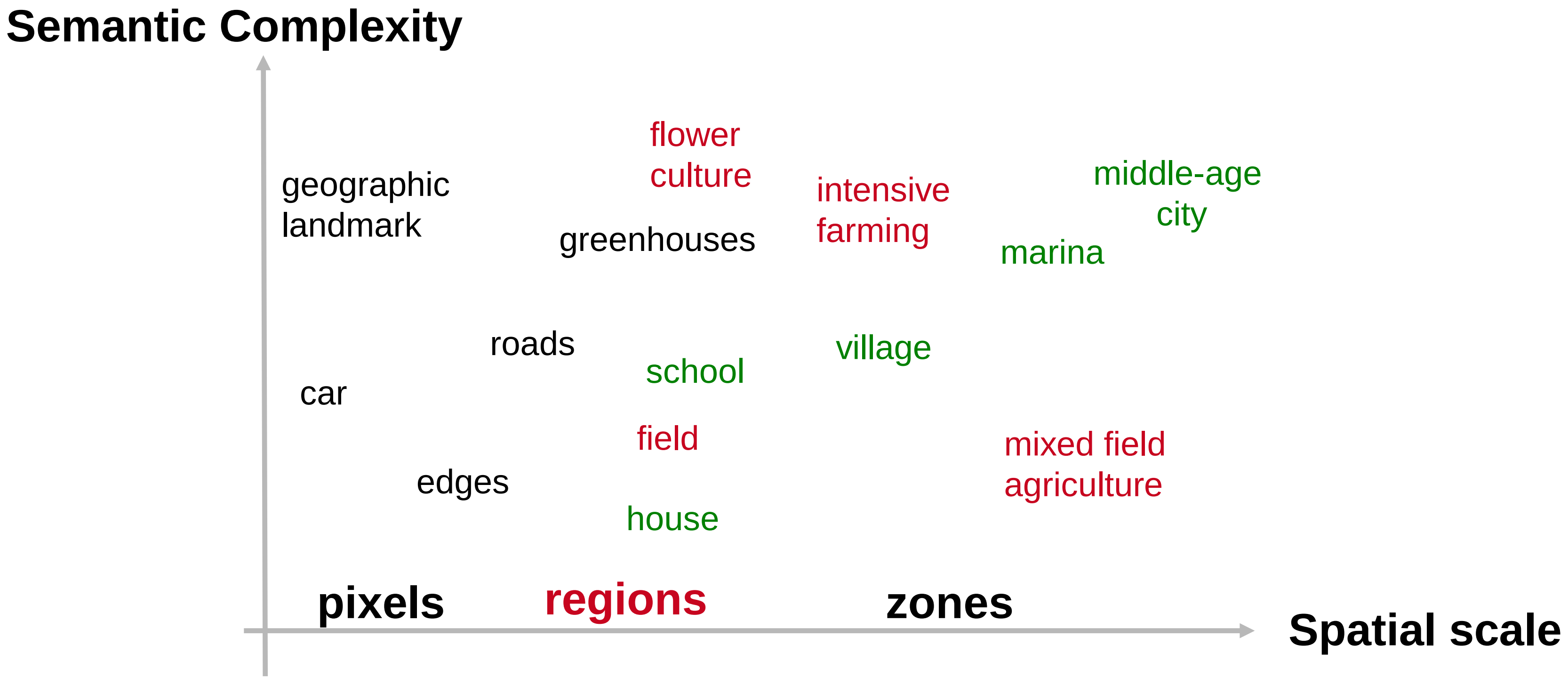
- **What are we looking for?** **It is not clear!** **(image production and image use are 2 different jobs)**

  - **Precise objects:**
    - Boat
    - Building
    - Road-crossing
    - Airplane landing area
    - Troops movements

  - **Generic objects:**
    - Marina
    - Greenhouse cultures
    - Oil pipeline
    - Geological synclinal
    - Forest fire
    - refugee camps
    - typhoon hazards

  - **Specific terrain configurations:**
    - Conducive to: … floods, … desertification, … urban pollution, …
    - Conducive to: … build a factory, … plan a bombing, … cultivate marijuana

Spatial scale vs. Semantic complexity

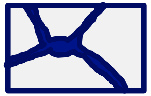


Hierarchical representation

- **Pixel**
  - **spectral properties (R,G,B,IR)**
  - contrast / texture
  - **edges, contours**


house

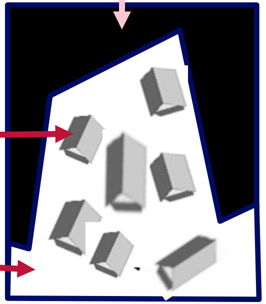
network



sea

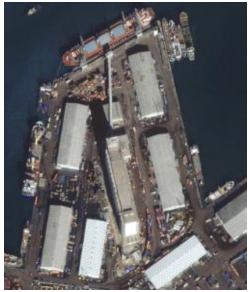
warehouse

wharf



➔ **Objects**

➔ **Scene**




fields

- **Region**
  - **form / shape**
  - content (spectral : textural)

Increasing semantics

RS image processing & hand-crafted feature detection

Handcrafted features

- **Handcrafted features are chosen by the user to reflect what is known about the object under investigation.**
  - **It may be positive: reflecting a property which is strongly associated with the looked for object**
    - (for instance swimming pools in residential areas)
  - **It may be negative if we know that its presence is not possible in the looked for object**
    - (for instance gas cisterns in residential areas)
- **Handcrafted features are issued from application expertise**
- **Handcrafted features are detected using image processing expertise**

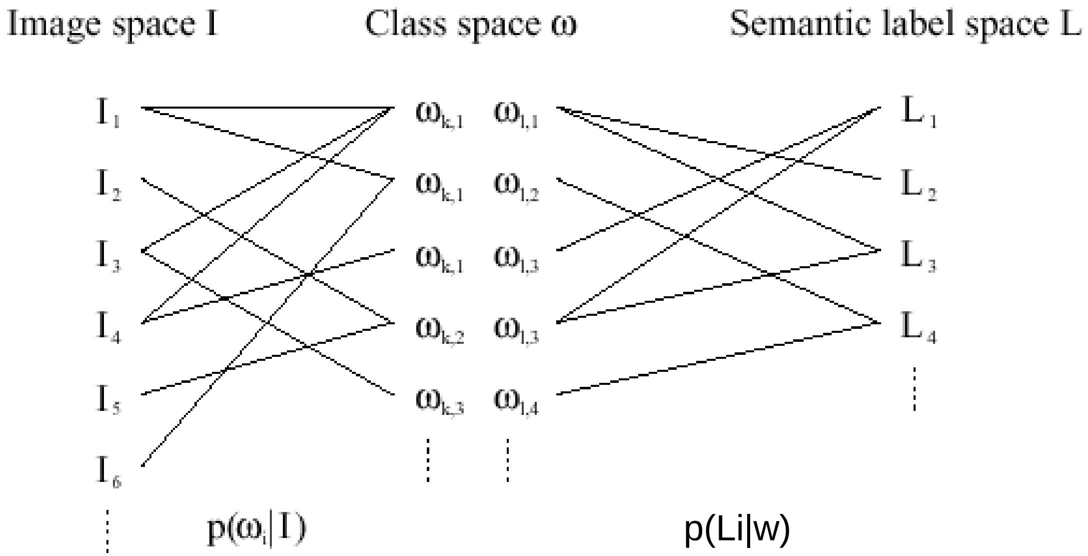
Mining in RS Image databases

- Semantic information retrieval :
  - From ancillary data

- I – Classical Machine Learning techniques (2000-2012)
  - Image Processing
  - Hand-crafted feature detection and/or salient point detection
  - Classification in p-dimensional space
    - ➔ few parameters
    - ➔ few learning images (groundtruth) ~ 1000

- II – Deep neural networks (2012 - …)
  - Directly with images as input and/or with extracted features
  - Several +/- linear classifiers in cascade
    - ➔ thousands of parameters
    - ➔ hundred of thousands of images as groundtruth

Probabilistic evaluation



Hand crafted features

Radiometry

- Multispectral : channels
- Specific combinations for remote sensing : NDVI (= $\frac{NIR-red}{NIR+red}$) , IB , ISU

Textures

- Gabor Filters
- Haralick cooccurrence matrices and their descriptors
- Quadratic Mirror Filters (wavelets)
- Contourlet decomposition
- Steerable wavelets
- Markov random fields parameters (Gaussian, Laplacian, Log-laplacian …)

Structures

- Contours & edges (coastline, deserts, …), regions (lakes, forests, …)
- Objects : roads, buildings, rivers, lakes
- Roads, Railways or River networks

Some efficient choices

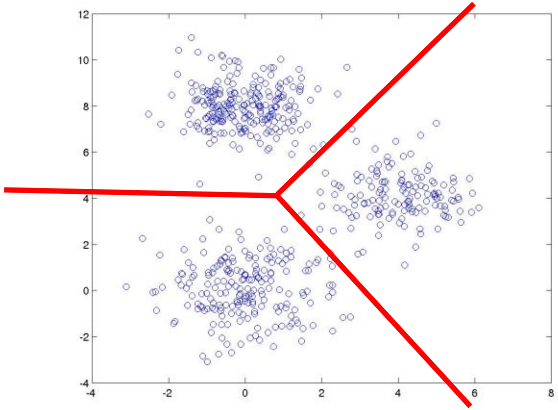
- **Indexing:** small subimages: (~ 64 x 64 pixels) = 320 m x 320 m on the ground for SPOT 5 images
- **Mixed features:**
  - **Radiometry (Panchro only)**
  - **Structure (contours)**
  - **wavelets : 2 directions, 4 scales**
- **Automatic feature selection (supervised:** ReliefF, Fisher FS, SVM-RFE or **unsupervised**: MIC (*Max Information Compression*), k-means FS**)**

  ~ 100 features with redundancy **or** ➔ **10 to 20 features without redundancy**
- **Give names to classes (***from label to name***)**
  - **Waste fields**
  - **Cultures**
  - **Housing**
  - **Road and river networks**

Classification

Many different classifiers:

- MAP & Bayes decision
- K-nearest neigbours
- Graph tree, Random Forest
- Kernel methods (SVM = Support Vector Machine)
- Hierarchical clustering



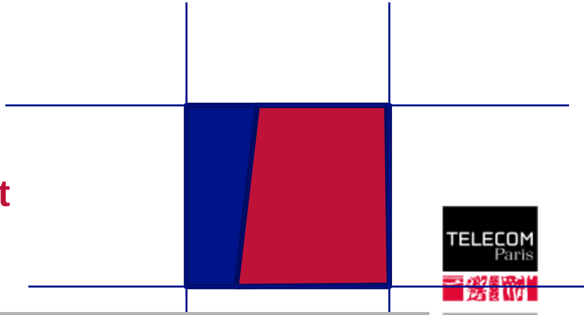
or

label = 24

Semantic labelling
name = « Corn field »

Supervised
or
Unsupervised

Partial volume effect



Support Vector Machine

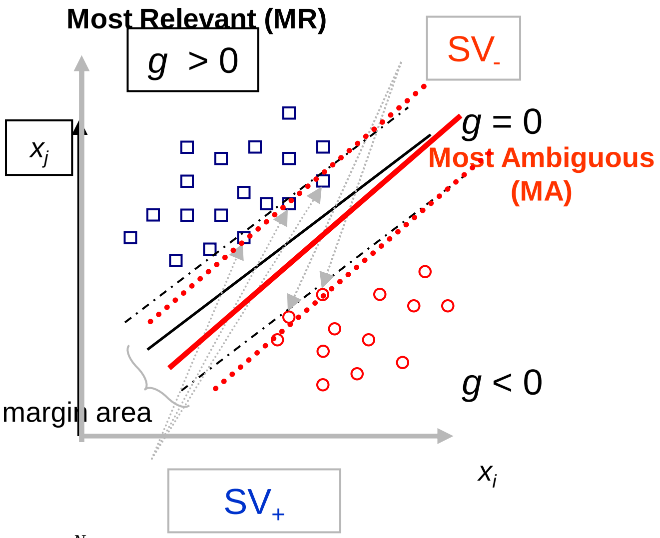
- **Linear separation case**
  - **Labeled data training set**

    $$(x_i, y_i),\ x_i \in F = \Re^d,\ y_i \in \{-1,+1\},\ i = 1..N$$

  - **Find a separation surface**

    $$g(x) = w \cdot x + b = 0 \qquad y_i(w \cdot x_i + b) \geq 1$$

  - **Decision function *f = sign(g(x))***
  - **d_+ = distance from *g* to closest {+1}**
  - **d_- = distance from *g* to closest {-1}**
  - **Margin area = $d_+ + d_-$ =** $2/\|w\|$
- **Find a separating hyperplane with largest margin**



$$L = \frac{1}{2}\|w\| - \sum_{i=1}^{N} \alpha_i (y_i (w \cdot x_i) - 1) \rightarrow \frac{\partial L}{\partial w} = 0 \ and\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \ and\ w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\max_{\alpha} \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j x_i x_j \right) \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0\ , \alpha_i \geq 0$$

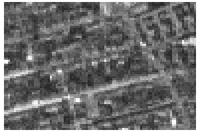
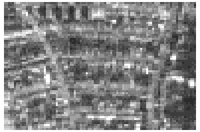
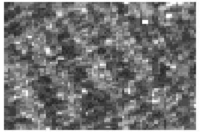
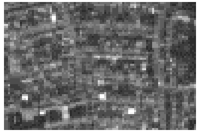
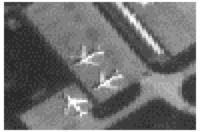
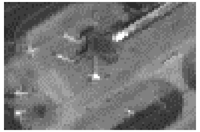
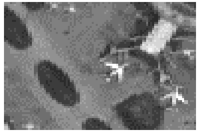
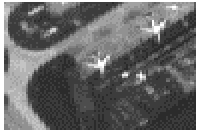
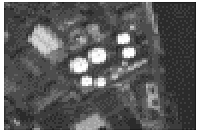
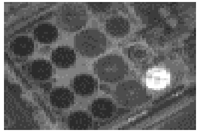
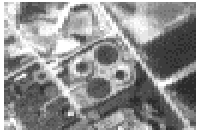
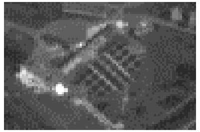
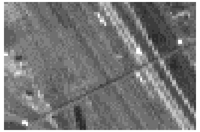
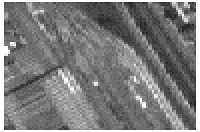
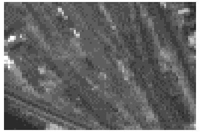
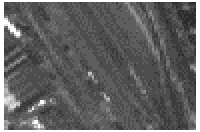
How to introduce semantics?
Where are words coming from?

- **Supervised methods**
  - Fully manual indexing (experts or crowd sourcing)
  - Partly: learning (relevance feedback)
- **Contextual analysis of the document**
  - Tittle, caption, text, web site
- **Use of external data-bases**
  - Corine Land Cover (to learn classes and categories)
  - Maps and GIS (annotation)
- **Semantics inference**
  - Bayesian Modelling
  - Latent Models = Dirichlet, Blei & Jordan
  - « Ontological » deduction
  - Spatial reasonning

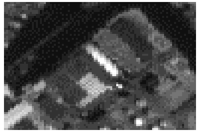
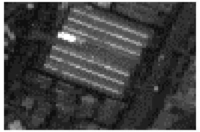
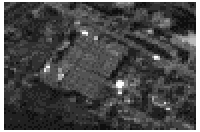
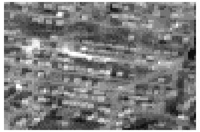
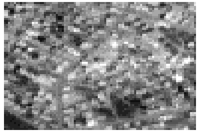
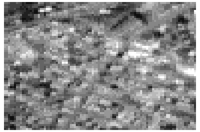
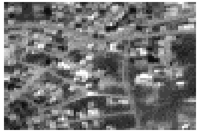
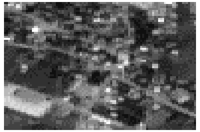
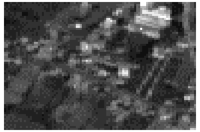
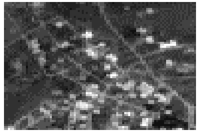
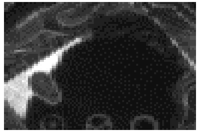
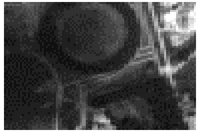
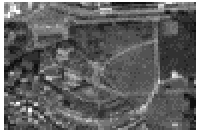
Example : CorineLandCover ontology

- 111: Continuous urban fabric
- 112: Discontinuous urban fabric
- 121: Industrial or commercial units
- 122: Road and rail networks and associated land
- …
- 211: Non-irrigated arable land
- 221: Vineyards
- 222: Fruit trees and berry plantations
- …
- 231: Pastures
- 242: Complex cultivation patterns
- 243: Land principally occupied by agriculture with significant areas of natural vegetation
- 311: Broad-leaved forests
- 312: Coniferous forests
- 313: Mixed forests
- …
- 411: Inland marshes
- …
- 511: Water courses
- …

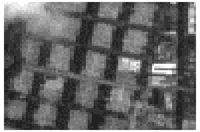
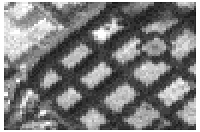
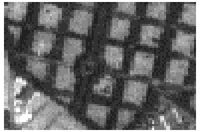
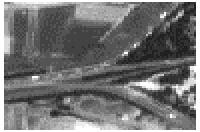
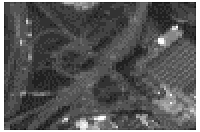
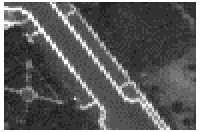
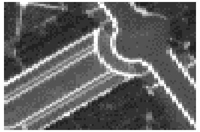
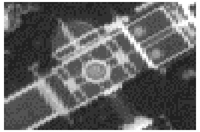
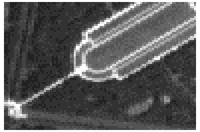
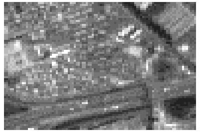
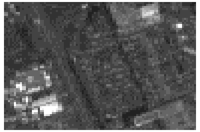
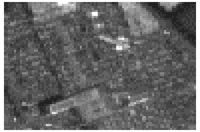
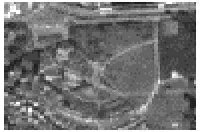
Supervised classes

Residential areas				
Planes				
Industrial tanks & cisterns				
Railway marshalling yard				

Supervised classes

factories				
Dense urban area				
villages				
Urban parks				

Supervised classes

Graveyards				
Road interchange				
Castle parks				
Parking lots				

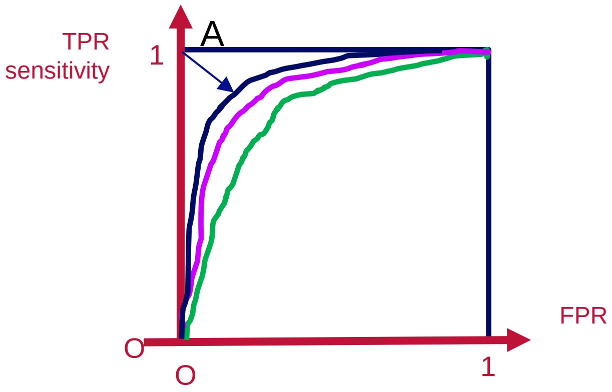
How to express results?

- **Classification rate** 97.3 % (or **error rate**: 2.7 %)
- **Confusion matrix**

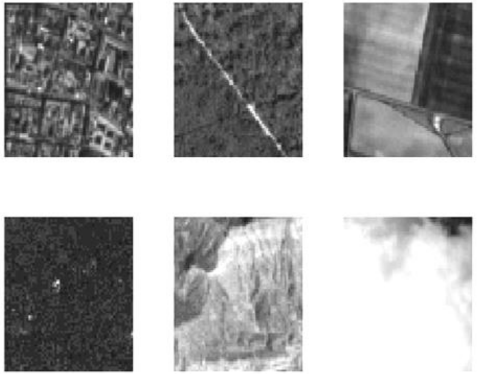
	Present object	**Absent object**
Positive detection	True positive (TP)	False positive (FP) (type I error)
Negative detection	False negative (type II error)	True Negative

- **Receiver Operating Characteristic (ROC Curve)**

Convert TP and FP into FPR and TPR ϵ [0,1]
Plot TPR = *f*(FPR) for many different parameters
Without specific instruction, take the closest point from A = (0,1) as working condition



Typical performances of algorithms



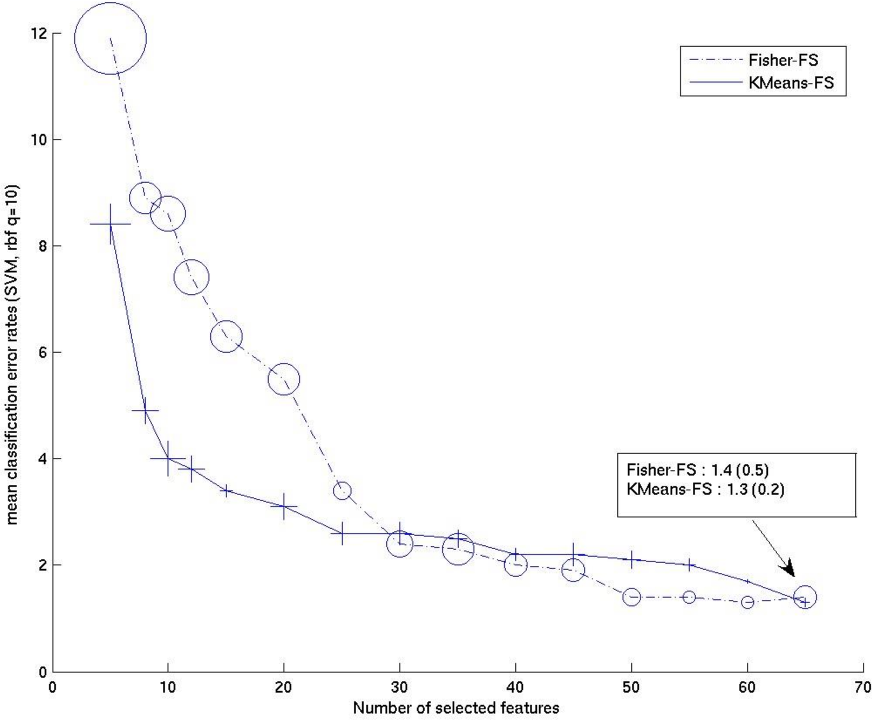
Sub image classification (128 x 128) :

city, wood, fields, sea, desert & clouds

600 images for each class
Results: Gaussian SVM,
Mean error 1.4% ± 0.4%
(147 features, cross validated)

True\Found (%)	city	clouds	desert	fields	woods	sea
city	98.8	0	0	0.5	0	0
cloud	0	99.3	0.2	0	0	0
desert	0	0	99.0	0.3	0	0
fields	0.5	0.2	0.8	98.1	0.3	0.4
woods	0	0.2	0	0	98.0	1.4
sea	0.7	0.3	0	1.0	1.7	98.2

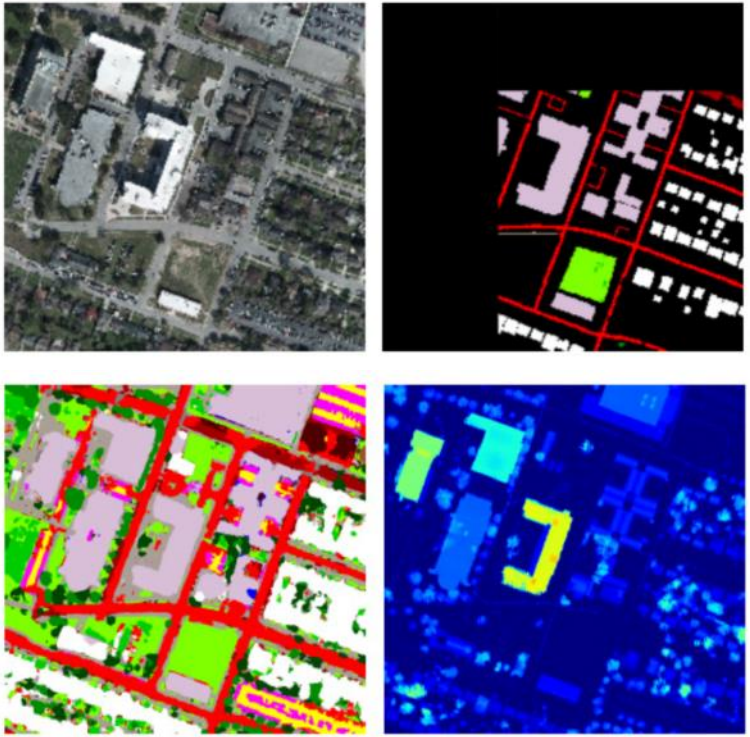
How many features?



Automatic feature selection

- Wrappers
- Filters (mutual information)
- Embedded (Lasso)

Different ground truthes



- Obtained from manual delineation

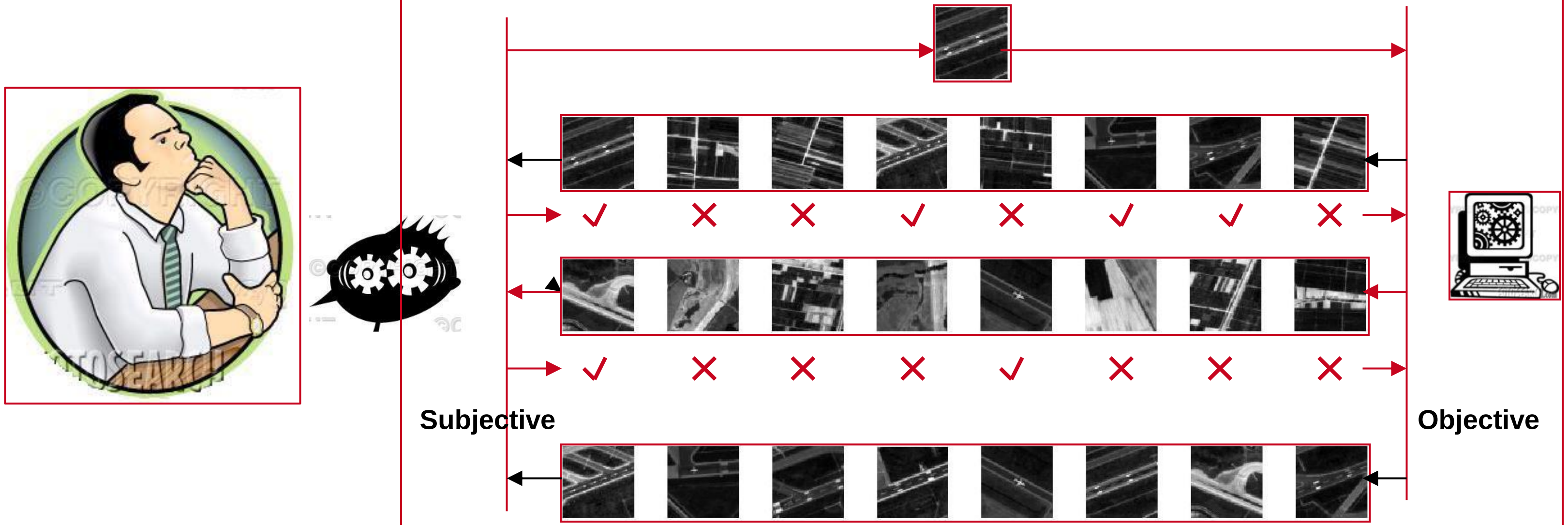
- Obtained by image processing
  - Edge detection, road detection area classification,
  - stereovision

Using a human expert to improve learning

« A man (woman ?) in the loop »

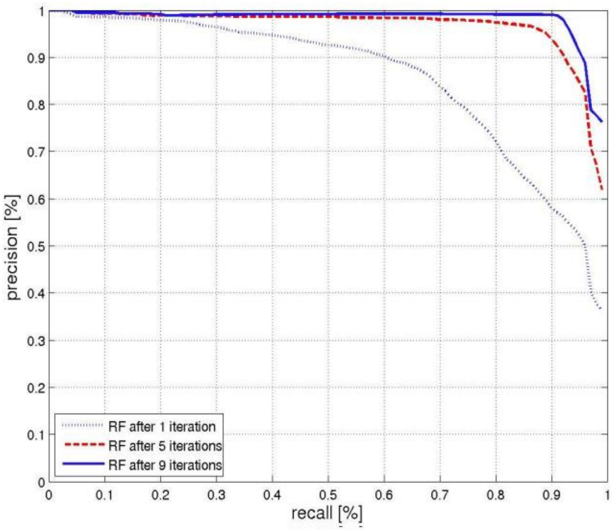
Learning with Relevance feedback

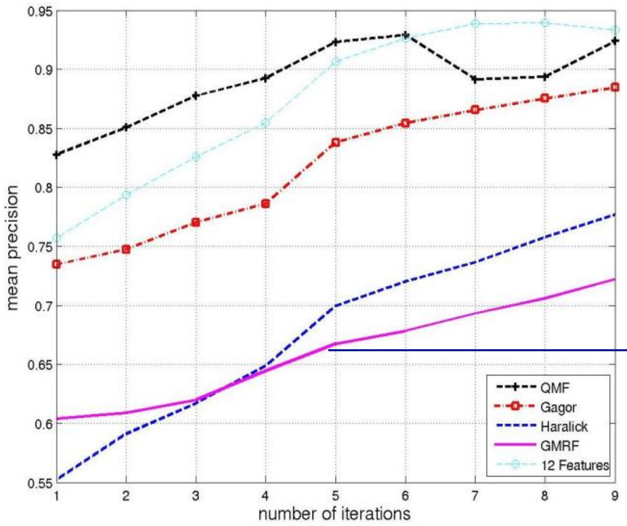
- **Man Machine dialog**



Learning with Relevance feedback

- Database composed of 600 SPOT5 images divided in 6 classes
- Used features: Gabor, Haralick, QMF and GMRF
- Gaussian Kernel
- System evaluation: Precision-Recall graphs





Deep Neural Networks

Deep Neural Networks

- **As for many other Pattern Recognition problems, DNN is one of the most efficient solution for Remote Sensing applications.**
- **Solutions take benefit of the development of efficient architectures in the field of Pattern Recognition**
- **Softwares and Architectures are not yet stabilized and are still under investigations**
- **Domain application expertise is required to build the annotated ground data set.**

Mining in RS Image databases

- **Semantic information retrieval :**
  - **From ancillary data**

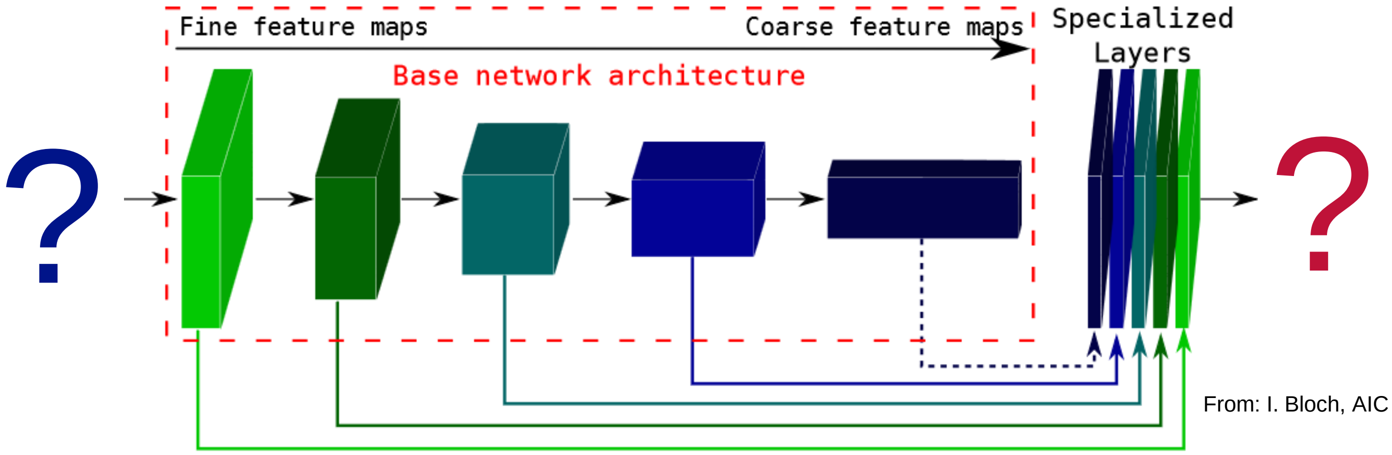
- **I – Classical Machine Learning techniques (2000-2012)**
  - **Image processing**
  - **Hand-crafted feature detection and/or salient point detection**
  - **Classification in p-dimensional space**
    - ➔ few parameters
    - ➔ few learning images (groundtruth) ~ 1000

- **II – Deep neural networks (2012 - …)**
  - **Directly with images as input and/or with extracted features**
  - **Several +/- linear classifiers in cascade**
    - ➔ thousands of parameters
    - ➔ hundred of thousands of images as groundtruth

Some references (dated 01/10/2020)

- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). **Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9), 2784-2817.**
- Holloway, J., & Mengersen, K. (2018). **Statistical machine learning methods and remote sensing for sustainable development goals: a review. Remote Sensing, 10(9), 1365.**
- Carter, C., & Liang, S. (2019). **Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. International Journal of Applied Earth Observation and Geoinformation, 78, 86-92**
- Li, J., Huang, X., & Gong, J. (2019). **Deep neural network for remote-sensing image interpretation: Status and perspectives. National Science Review, 6(6), 1082-1086.**
- Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). **Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing, 11(2), 196.**
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., ... & Gao, J. (2020). **Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment, 241, 111716.**
- Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 22-40.
- Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 645-657. Tong, X. Y., Lu, Q., Xia, G. S., & Zhang, L. (2018). Large-scale Land Cover Classification in GaoFen-2 Satellite Imagery. *arXiv preprint arXiv:1806.00901*.
- Boualleg, Y., & Farah, M. (2018, July). Enhanced Interactive Remote Sensing Image Retrieval with Scene Classification Convolutional Neural Networks Model. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4748-4751). IEEE.
- Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2016). Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, *13*(1), 105-109.
- Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 778-782.
- Penatti, O. A., Nogueira, K., & dos Santos, J. A. (2015). Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 44-51).
- Pelletier, C., Webb, G. I., & Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5), 523.
- Zhang, S., He, G., Chen, H. B., Jing, N., & Wang, Q. (2019). Scale adaptive proposal network for object detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, *16*(6), 864-868.

Deep Neural Network



From: I. Bloch, AIC

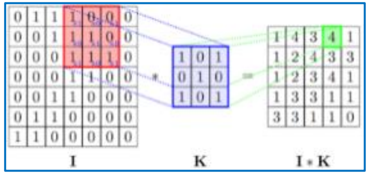
- **Which input?**
- Raw image
- Processed image (filtered, segmented …)
- Feature detected image (classified, edge detected, …)
- Features

- **Which architecture?**
  - # layers,
  - type of layers
- **Which protocole?**
  - Feature learning
  - Fine tuning

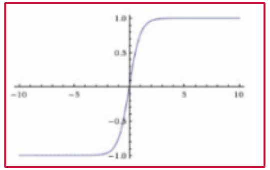
- **Which output?**
- Densely classified image
- Detected targets
- List of targets
- List of Features

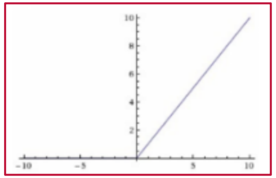
CNN basic components

- **Convolutional layer**: with *rxr* kernel – down scaling

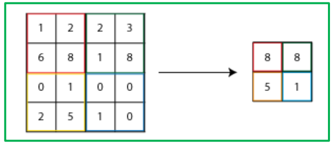


- **Nonlinearity**: sigmoïd or RELU (rectified linear unit)

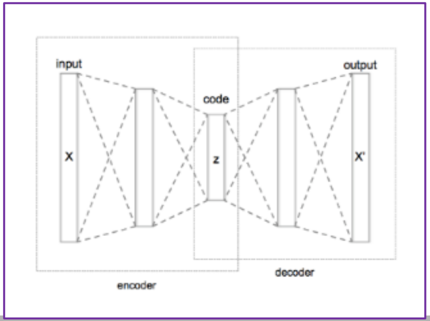




- **Pooling layer**: single value taken from a set of values - ex: *max* on a *rxr* patch

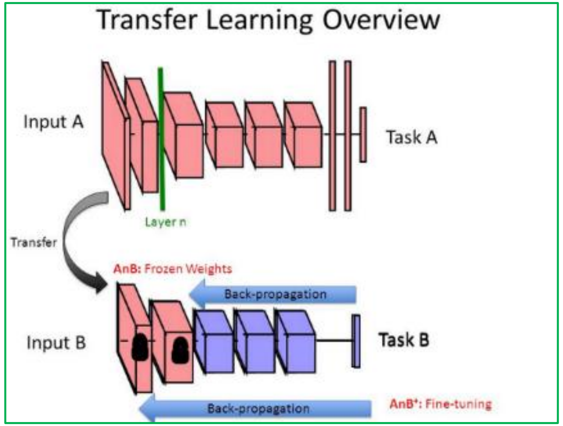


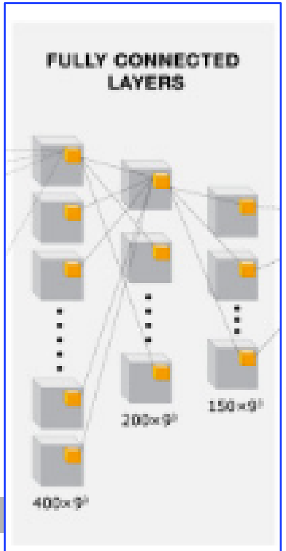
- **Autoencoder**: symetrical NN to reduce the model dimensionality

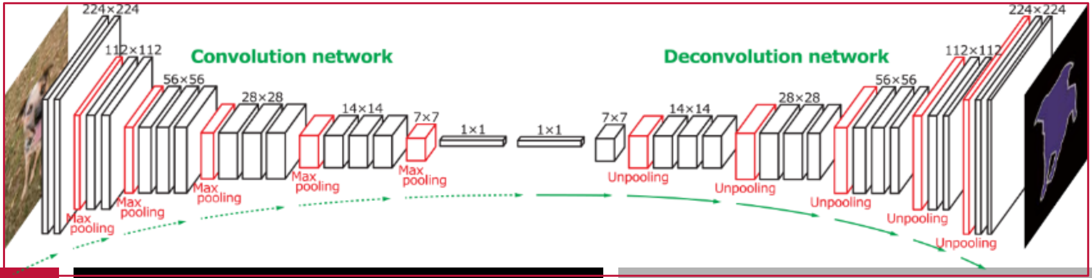


CNN basic components

- **Fully convolutional layer**: to perform a large distance context dependance
- **Transfer coding:** to learn from a database and use for another one
- **Fine Tuning**: to specify a network to a given task after training on a general purpose data base
- **Yoyo architecture** : downsampling for feature extraction then upsampling for fine positioning of targets







Most used components for RS-CNN (2019)

- **CNN from the Pattern Recognition community**
  - AlexNet
  - GoogleNet
  - VGGNet
  - ResNet
  - Inception
- **Training sets (specific or not to Remote Sensing community)**
  - ImageNet (General purpose image library for pattern recognition)
  - UC Merced DataSet (Aerial images / 21 classes)
  - OSM - OpenStreetMap (Aerial Image Database)
  - Google Street Map (hi level semantic)
  - NLCD - USGS data Base (Geological survey)
  - Corinne Landcover (Agriculture & vegetation)
  - Gaofen Image Dataset (GID) (Hi Resolution Satellite)
  - …

Marmanis et al. IEEE TGRS, Jan 2016

Instance # 1 : Basic CNN (DLR)

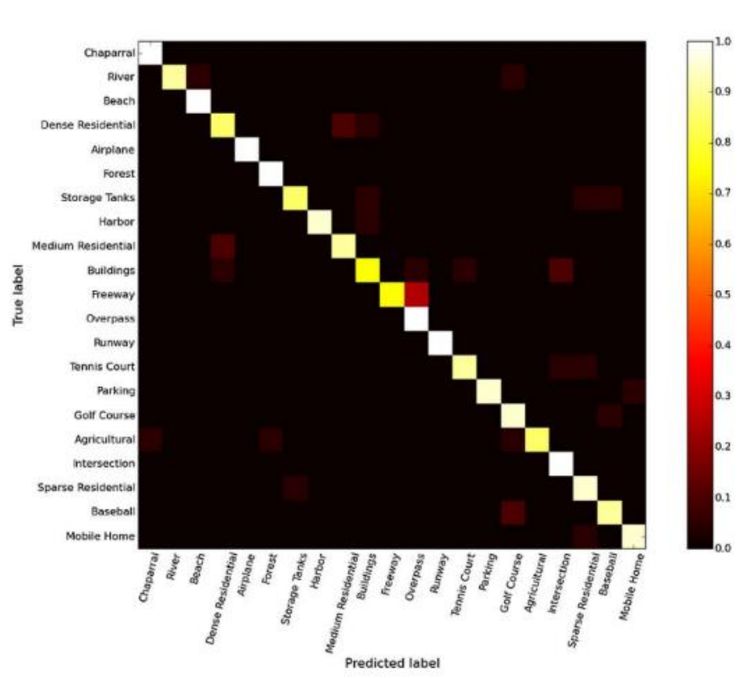
- With UC Merced Land database (aerial / 21 classes)
- With pre-trained CNN (Imagenet)
- Fine-tuned full convolutional layers with enhanced data

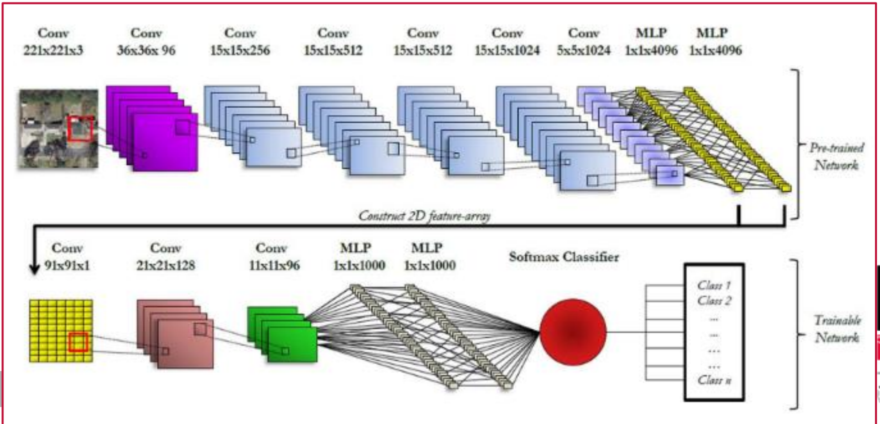
TABLE II
CLASSIFICATION COMPONENTS AND ALGORITHM COMPARISON

Method & Algorithm	Test-set Accuracy
Random Forest with RGB feature	44%
CNN with RGB feature	44.5%
Random Forest with Overfeat features	86.9%
CNN with Overfeat feature	**92.4 %**

TABLE III
METHOD COMPARISON OVER THE UCML BENCHMARK

Method & Algorithm	Test-set Accuracy
BOVW [2]	71.8%
SPMK [1]	74%
SPCK++ [2]	76%
Sparce Coding [4]	81.7%
Salient Unsupervised Learning [6]	82.8% ± 1.18%
MinTree + KD-Tree [3]	83.1% ± 1.2%
CNN with Overfeat feature	**92.4 %**

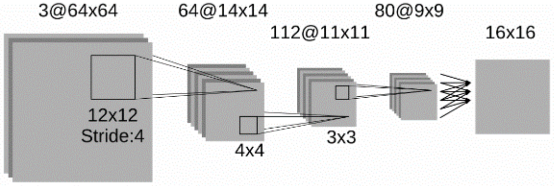




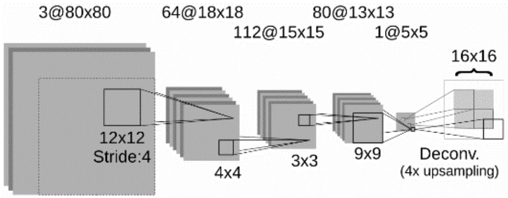
Instance # 2 : fully CNN (Inria)

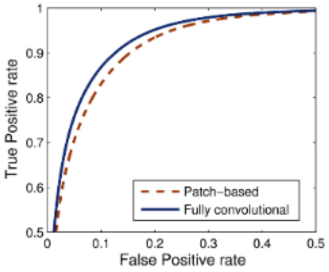
Maggiori et al. IEEE TGRS, feb 2017

Patch-based CNN



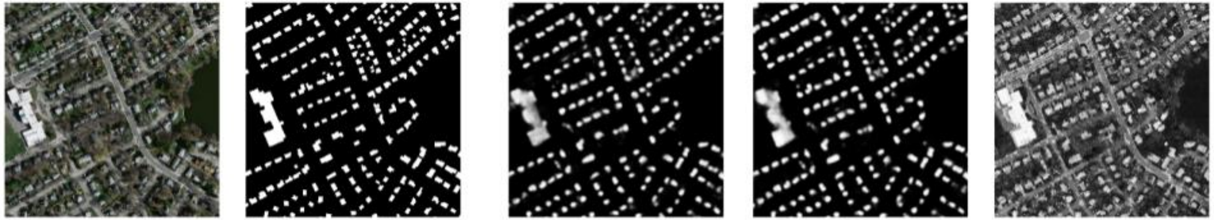
Fully convolutional Patch -based CNN





Detection of buildings

Image ground truth patch based fully convolutional SVM



(a) (b) (c) (d) (e)

Instance # 3 : RS CNN (Liemars/Wuhan)

Pretrained with ResNet

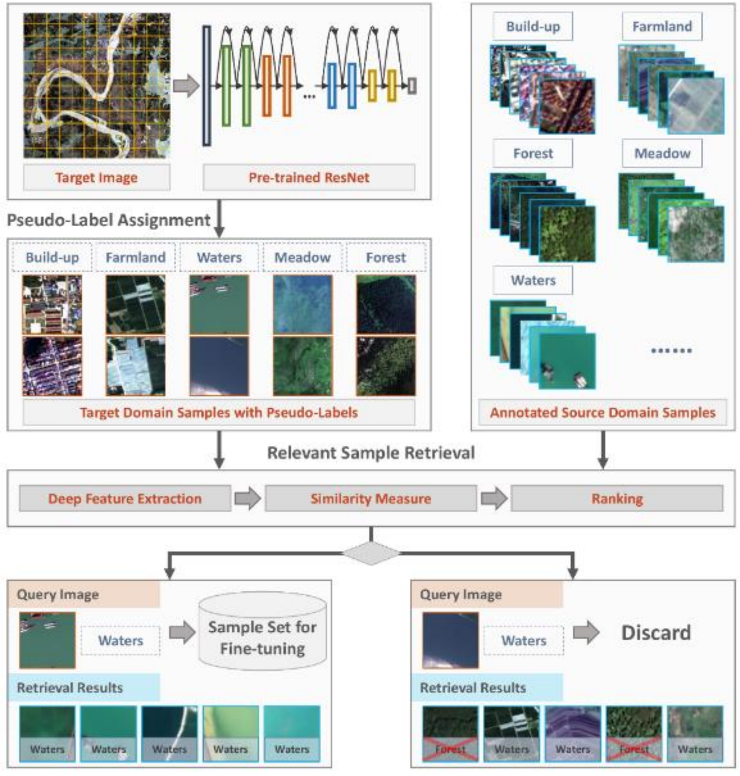
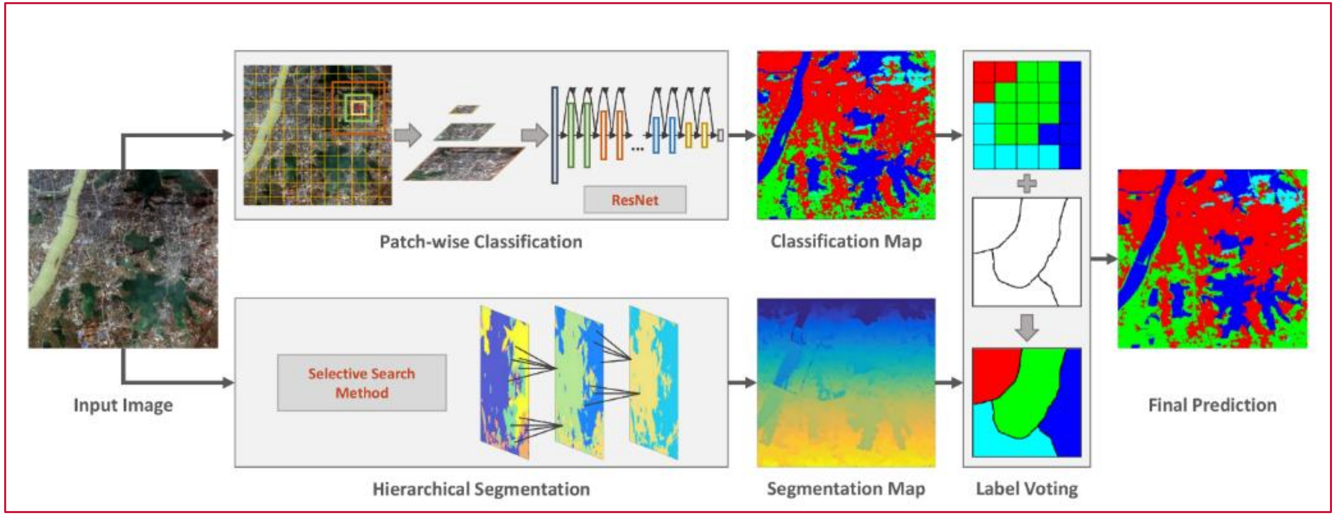


Figure 1: Sample selection for model fine-tuning.

Instance # 3 : RS CNN (Liemars/Wuhan)

Cooperation between classifying (sparse) and segmenting (dense)



From : Tong et al.
arXiv 1807.05713 - 2018

Instance # 3 : RS CNN (Liemars/Wuhan)

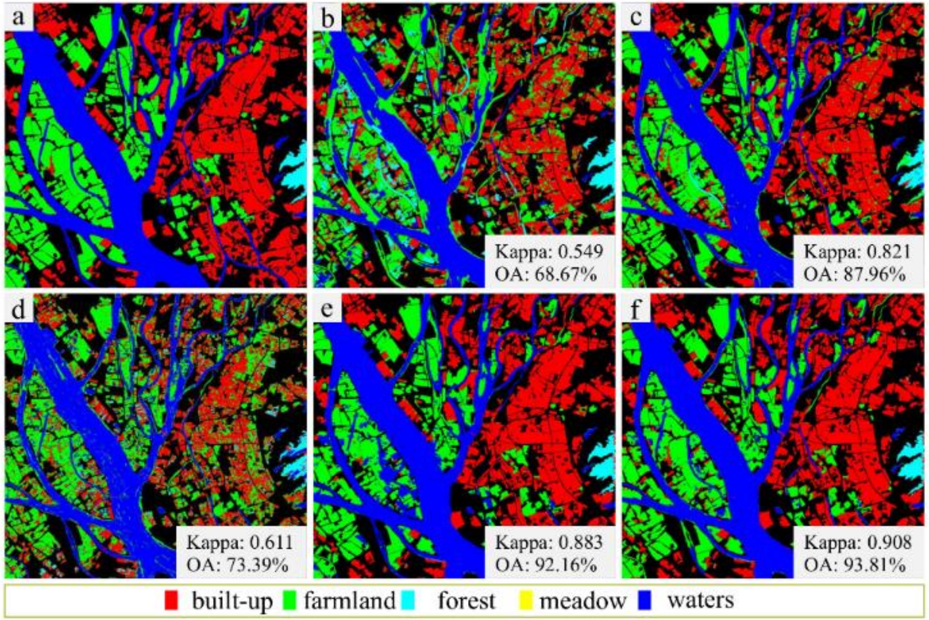


Figure 8: Land-use classification maps of the GF-2 image obtained in Dongguan, Guangdong Province on January 23, 2015. (a) Ground truth. (b)-(f) Results of eCognition, RF+Fusion, SVM+Fusion, PT-GID, and FT-$\mathbf{U}_{tg}$.

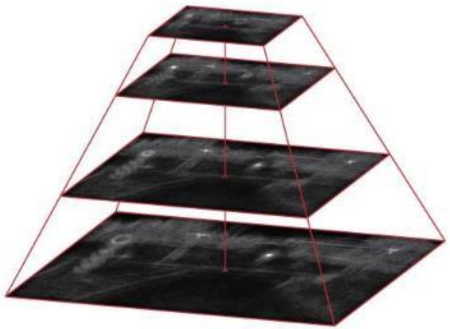
From Low to High Level - Changing the scale

Complexity of images

Analysis window : real size
128 x 128 pixels

Analysis window : enlarged

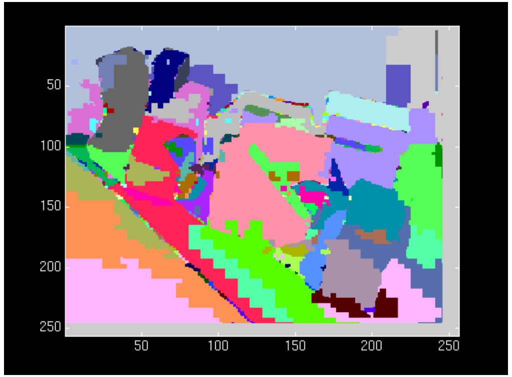
Scale enlargement strategy



Pyramid

Sliding window

Growing and Merging

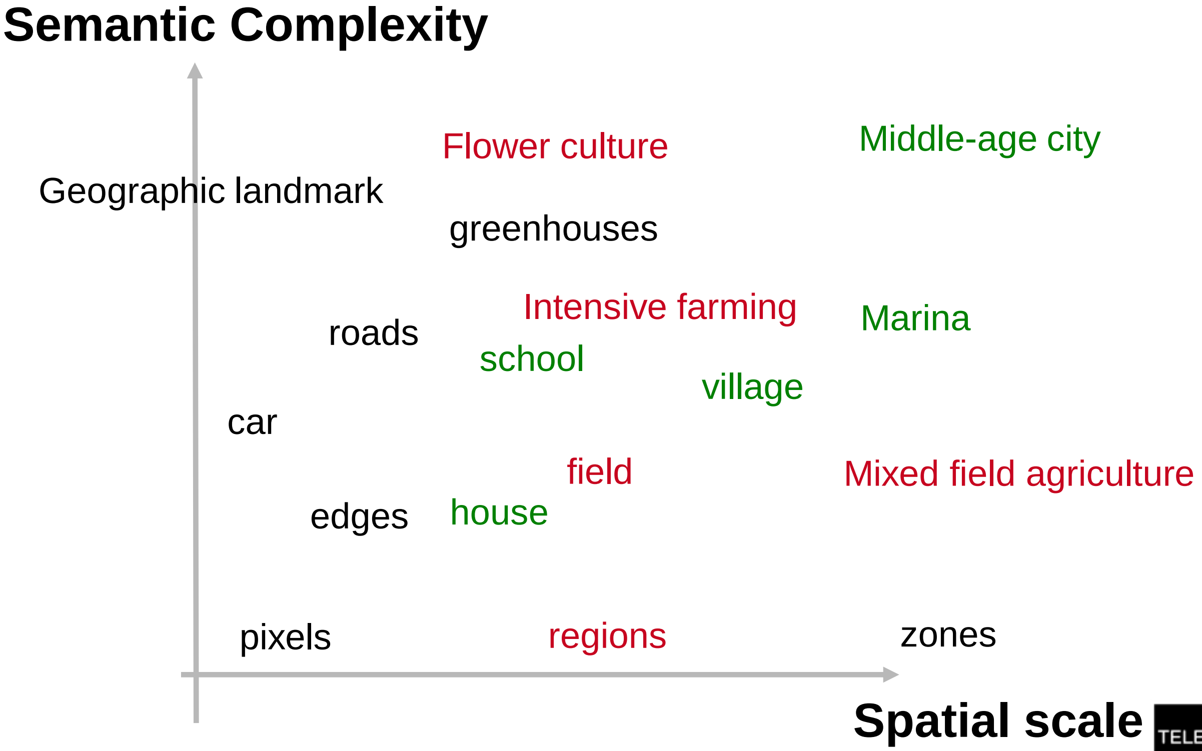


Hierarchical representation

Two goals:
- Enlarge the field of view
- Increase the semantic level

Grouping strategy:
- Sliding window
- Pyramid
- Growing and Merging

Decision strategy:
- Bag of Visual Words (BOVW)



Increasing the semantics

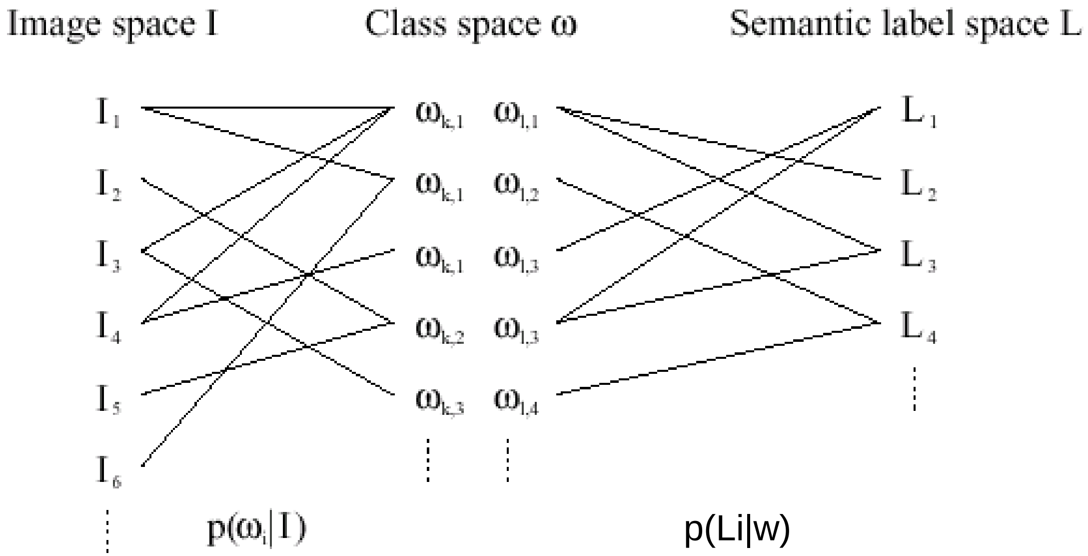
Park = {trees+fields+tracks}

Residential area = {houses + lawns + pools + roads}

Waste area ={waste+lawns+trees+roads}

Commercial area = {buildings+houses+parking lots+ waste

Probabilistic evaluation



Decision making: Bag of Words

- **2 levels** ➔ **H=high** (unknown) **L = low** (known)
- **List of N classes at H** = $\{c_1, c_2, \ldots c_N\}$
- **At H : 1 super-region with n objects, each** ∈ **1 class = n labels** described by the ordered list of the probability (or the occurrence) of each class**:**

  $$R_k = \{p_1, p_2, \ldots p_n\}$$

- **Classify H according to the** R_k
  - **Naïve Bayes :** $\boldsymbol{c}^* = \operatorname{argmax} \boldsymbol{p}(\boldsymbol{c}|\boldsymbol{x}) = \operatorname{argmax} \boldsymbol{p}(\boldsymbol{c}) \prod_{\boldsymbol{k}=1}^{\boldsymbol{n}} \boldsymbol{p}(\boldsymbol{x_k}|\boldsymbol{c})$
  - **Improving Naïve Bayes:**
    - pLSA = Probabilistic Latent Semantic Analysis
    - LDA = Latent Dirichlet Analysis