Computer vision

José Mennesson

IMT Nord Europe

jose.mennesson@imt-nord-europe.fr

December 29, 2024

Definition

Computer vision, what is it??

Definition from Wikipedia : "Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, for example in the forms of decisions."

State of the art

- Berger, L. *Traitement d'images et de vidéos avec OpenCV 4 en Python (Windows, Linux, Raspberry)*. Éditions D-BookeR, 2020. ISBN: 9782822707954.
- Duda, Richard O, Peter E Hart, et al. *Pattern classification*. John Wiley & Sons, 2006.
- Gonzales, Rafael C and Richard E Woods. *Digital image processing*. 2002.

Some tools

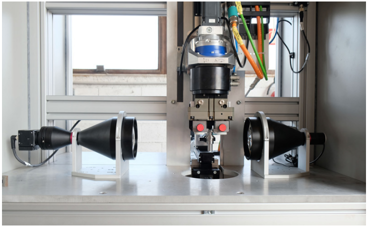
Image processing :

- openCV library (multi-language, multi-platform, opensource);
- ITK/VTK;
- ImageJ;
- numpy;
- CImg, Imagemagick, ...

Machine learning :

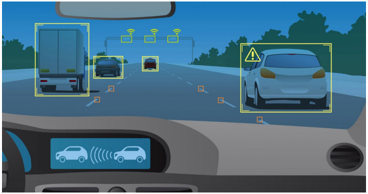
- libSVM;
- Weka;
- TensorFlow/ Keras;
- Caffe;multi-plateforme opensource
- PyTorch, Theano ...

Some applications

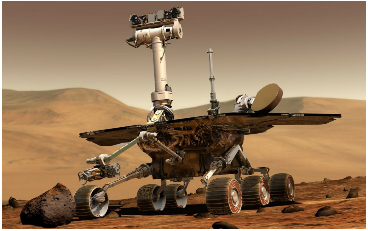


Industrial vision

Medical image processing



autonomous vehicles

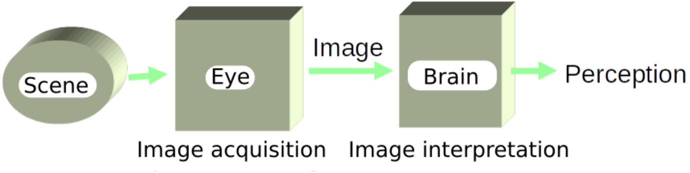


robotics

Goals

$->$ Imitate human or animal vision through electronic components.

Human vision



Computer Vision

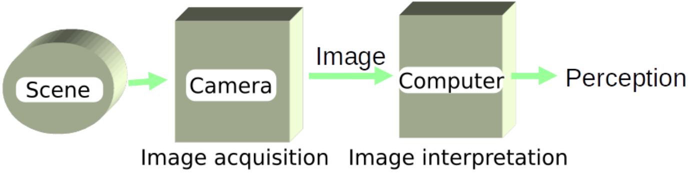
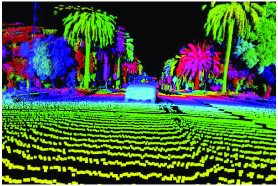


Image acquisition devices

Photo camera

Color camera

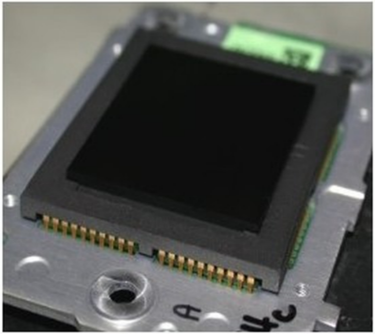
Infrared/thermal camera

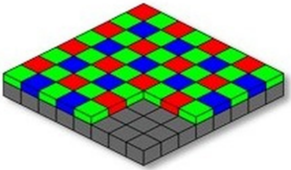


and many others...

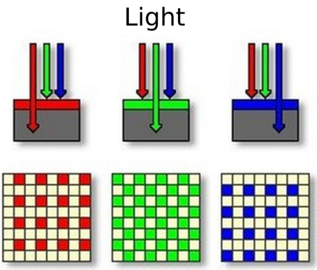
The color CCD image sensor

The color CCD image sensor





the photosensor is covered
with a Bayer grid



source : http://infographie-heaj.eu

An image

Definition

A picture is a 2D matrix, each cell contains a number (resp. a triplet of numbers) corresponding to the gray level (resp. the color). Each cell is called a pixel (picture element).

source : http://images.math.cnrs.fr

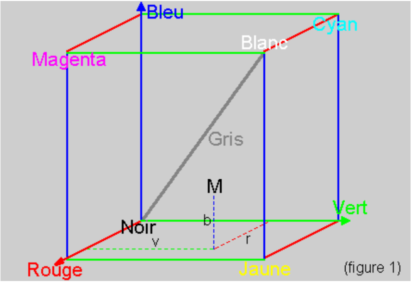
An image
Grayscale pixels

Pixels in grayscale images represent a level of intensity. Pixel values range from 0 (black) through all shades of gray to 255 (white).

An image

Color pixels

Color pixels are triplets of numbers representing the intensities of Red, Green and Blue (R,G,B).



An image

openCV Library

OpenCV (Open Source Computer Vision Library) is an open-source library which allows to do :

- computer vision;
- machine learning.

Applications are :

- detect and recognize faces;
- recognize objects;
- classify human actions in videos;
- tracking moving objects;
- recognize scenes for augmented reality, etc.

Over 47000 users and over 18 million downloads.
Interfaces C++, Python, Java and MATLAB and can be installed on Windows, Linux, Android and Mac OS.

An image

Demo openCV/Python

File LARM_BASE.py

- Open a picture file;
- display an image;
- display pixels;
- get a pixel;
- change of pixel value.

Image processing

Images are matrices. You can then perform operations on these images, such as addition, subtraction, division...

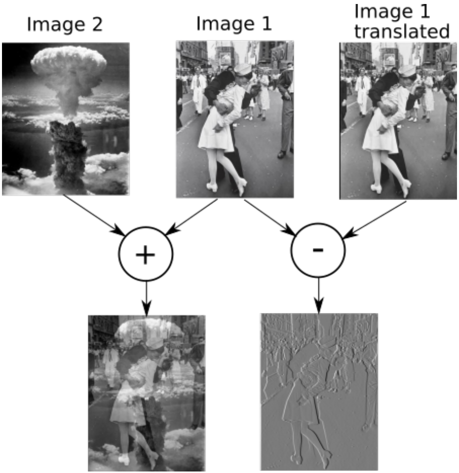


Image processing

Image gradients

The x-derivatives (on the columns) and y-derivatives (on the rows) of the image are called image gradients. This gives us the image's vertical and horizontal contours.

Original image	Gradient in the x direction	gradient in the y directior
		

Demo openCV/Python

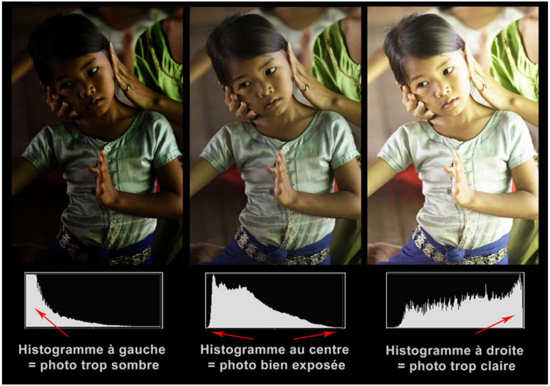
File LARM_OP.py

- Addition;
- Substraction;
- Division;
- Gradients.

Histogram of an image

Definition

The histogram of an image is a graphical representation of the distribution of pixel values in the image.



source : https://avecunphotographe.fr

Histogram of an image

Histogram computation

We count the number of pixels for each intensity between 0 and 255. For example, considering the image :

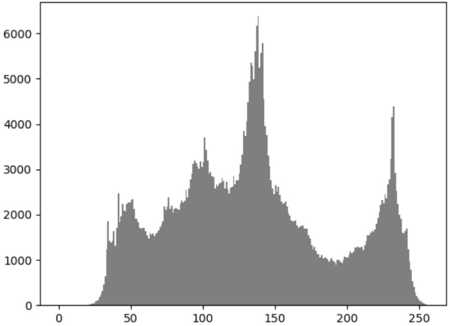
5	2	3	6
4	2	4	7
1	4	6	7
5	2	4	6

The associated histogram is :

Intensity	0	1	2	3	4	5	6	7	...	255
Number of pixels	0	1	3	1	4	2	3	2	...	0

Histogram of an image

Histogram of a natural image



Histogram of an image
Image thresholding

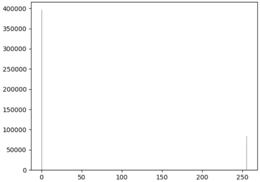
$\rightarrow$ Used to convert a grayscale image into a binarized image using a Threshold.

$$binarized\_image(i,j) = \begin{cases} 255 & \text{if } image(i,j) >= Threshold \\ 0 & \text{else.} \end{cases}$$

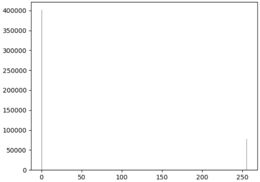
Histogram of an image

Grayscale thresholding

$(p > 200)\ ?\ 255 : 0$



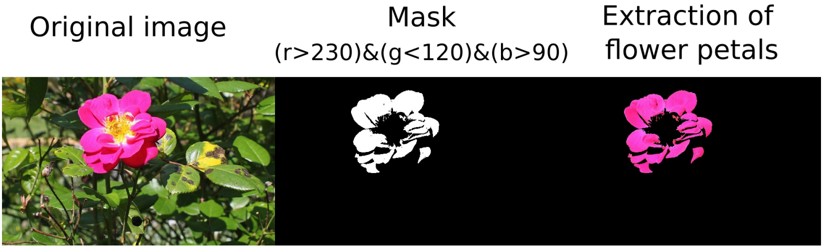
$(p < 75)\ ?\ 255 : 0$



Histogram of an image

Color thresholding

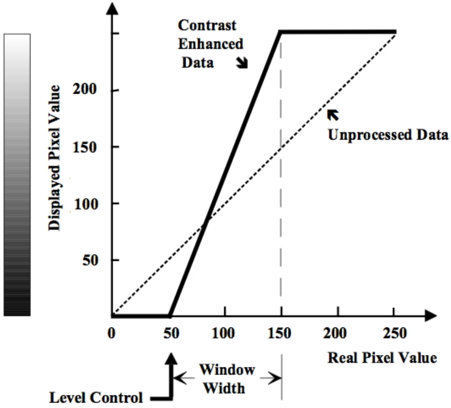
It's the same ! But we're going to threshold the three color channels (R,G,B). This can be used to extract an object from the scene.



Histogram of an image

The Lookup Table (LUT)

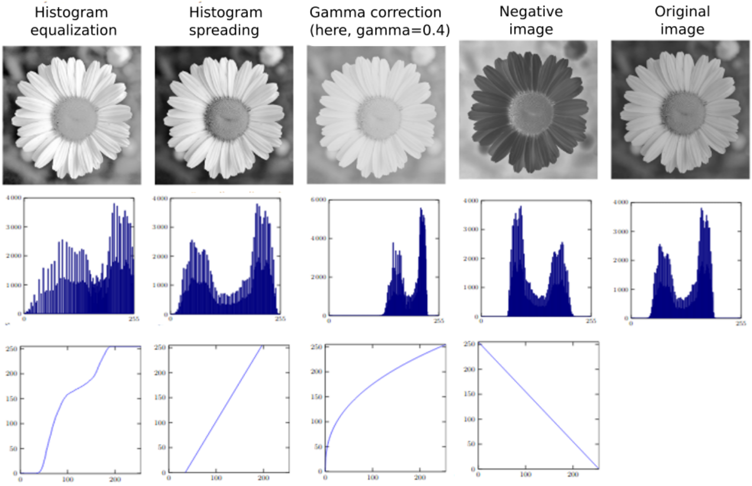
A LUT is a mapping table between an original pixel intensity and a desired pixel intensity. By varying this mapping table, we can influence image intensity.



source : Wikipédia

Histogram of an image

Intensity transformations

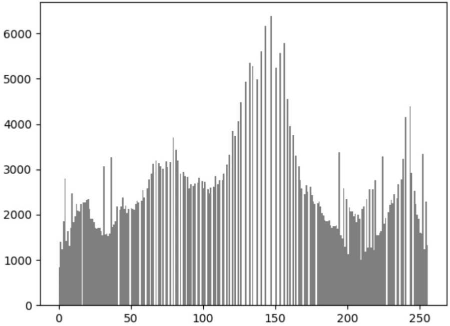


source : http://miv.u-strasbg.fr

Histogram of an image

Histogram equalization [Gonzalez, Woods, and Eddins 2004]

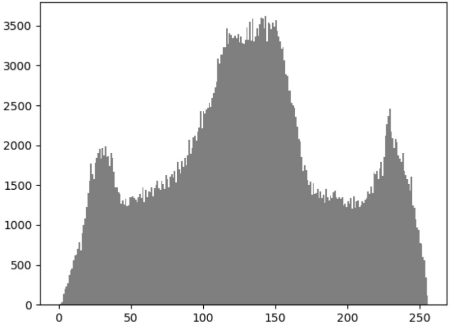
The aim is to spread the histogram so that grayscale takes up the entire grayscale range between 0 and 255. This results in an overall improvement in image contrast.



Histogram of an image

Contrast Limited Adaptive Histogram Equalization (CLAHE) [Pizer et al. 1987]

In this method, the aim is to increase contrast locally (in areas of size 8 x 8). In these areas, the histogram is equalized in the conventional way.



Histogram of an image

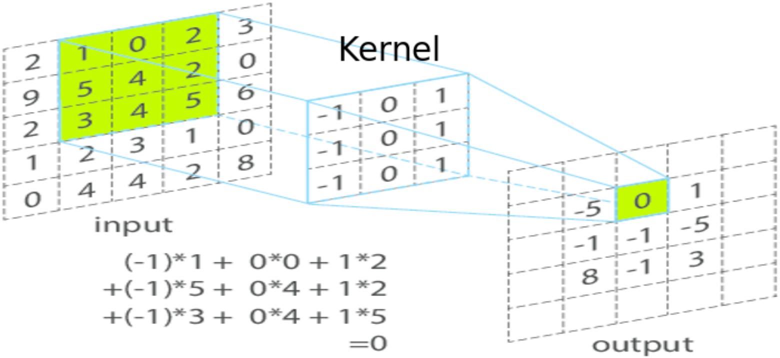
Demo openCV/Python

File LARM_HISTO.py

- Histogram computation;
- Grayscale thresholding;
- Color Thresholding;
- Applying a LUT;
- Histogram equalization;
- CLAHE.

Convolution

How to ?



source : https://perso.esiee.fr/perretb

Convolution

Different convolution kernels

Kernel	$\begin{array}{\|c\|c\|c\|}\hline 1/9 & 1/9 & 1/9 \\ \hline 1/9 & 1/9 & 1/9 \\ \hline 1/9 & 1/9 & 1/9 \\ \hline \end{array}$	$\begin{array}{\|c\|c\|c\|}\hline -1 & -1 & -1 \\ \hline -1 & 8 & -1 \\ \hline -1 & -1 & -1 \\ \hline \end{array}$	$\begin{array}{\|c\|c\|c\|}\hline 0 & -1 & 0 \\ \hline -1 & 5 & -1 \\ \hline 0 & -1 & 0 \\ \hline \end{array}$
Convoluted image			

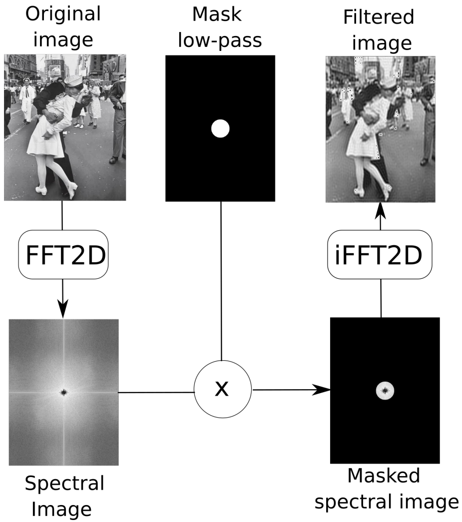
The spectral domain: the Fourier transform

Definition

The Fourier transform [Fourier 1822] allows you to switch to the spectral domain and manipulate "frequencies" in the image.
High frequencies (FT edges) → object contours
Low frequencies (FT center) → flat tints, soft changes, ...

The spectral domain: the Fourier transform

Spectral domain filtering



The spectral domain: the Fourier transform

Demo openCV/Python

File LARM_AVANCE.py

- Filtering by convolution
- Filtering by Fourier transform

Image analysis: a need for semantics

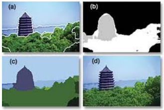
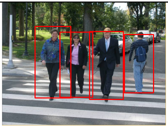
What's in the picture?

Image analysis: a need for semantics

What's in the picture?

Detection/segmentation/recognition

- **Image segmentation** is the action of partitioning the image into several zones according to pre-defined criteria.
- **Object detection** is the operation of searching for the position of a specific type of object in the image. The result is often in the form of bounding boxes.
- **Object recognition** is the action of determining the class (car, person, building, etc.) of an object in the image.
- **Semantic segmentation** is the fusion of segmentation and object recognition.

Segmentation	Detection	Recognition	Semantic segmentation
			

The challenge

- Point of view variations;
- Lighting variations;
- Scale variations;
- Deformations;
- Occlusions;
- Intra-class object variation.

Image segmentation

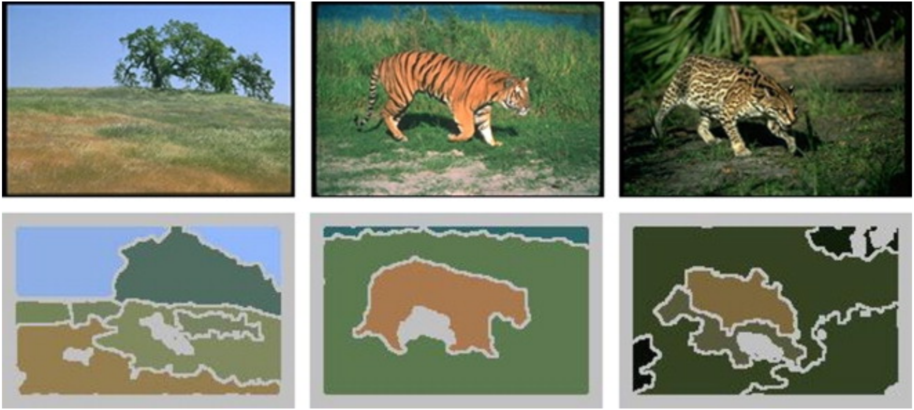


Image segmentation

Image thresholding

Very simple to implement but... very sensitive and not very selective!

Otsu's method [Otsu 1979] allows to automatically threshold grayscale images without setting a prior threshold.

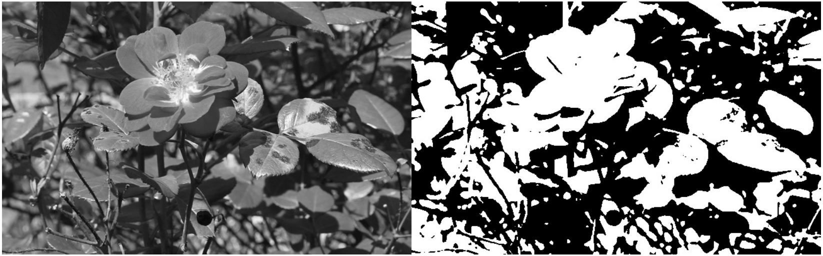


Image segmentation

K-means algorithm [MacQueen et al. 1967; Tou and Gonzalez 1974]

The idea is to divide pixel colors into K color classes. The algorithm will then automatically partition the pixel colors into K groups. In the example below, we took K=4.

Image segmentation

K-means algorithm

1. Initialization: partition objects by any method (random partition for example)
2. Step 1: for each class, calculate the centroid.
3. Step 2: assign each object to the nearest centroid. If at least one object has changed class, return to step 1 else: END

In the case of color image segmentation, each object is a pixel of the image.

Image segmentation

K-means algorithm

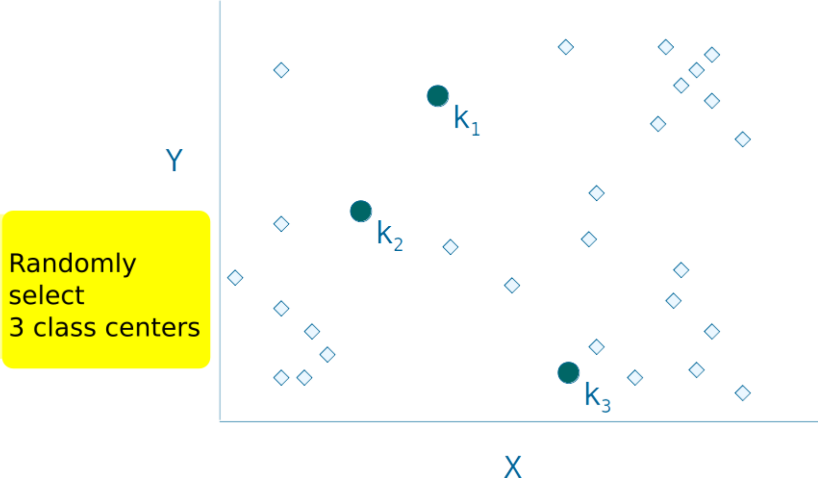
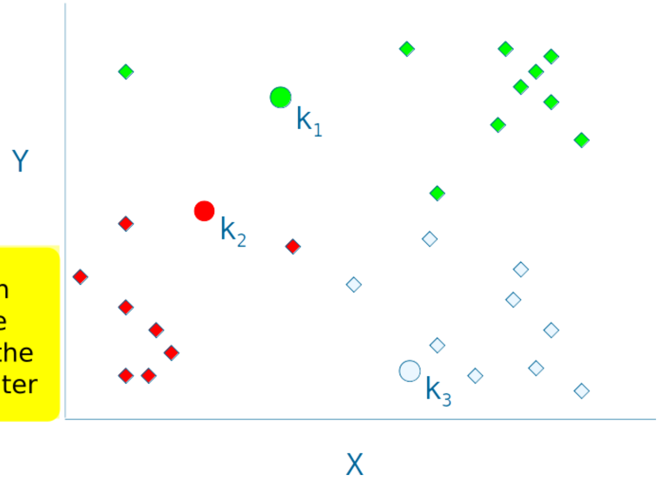


Image segmentation

K-means algorithm



assign each point to the class with the closest center

Image segmentation

K-means algorithm

Move each class centers to the class mean

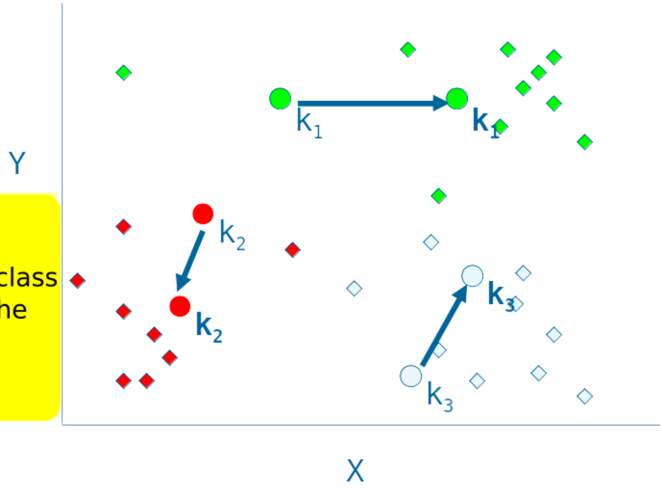


Image segmentation

K-means algorithm

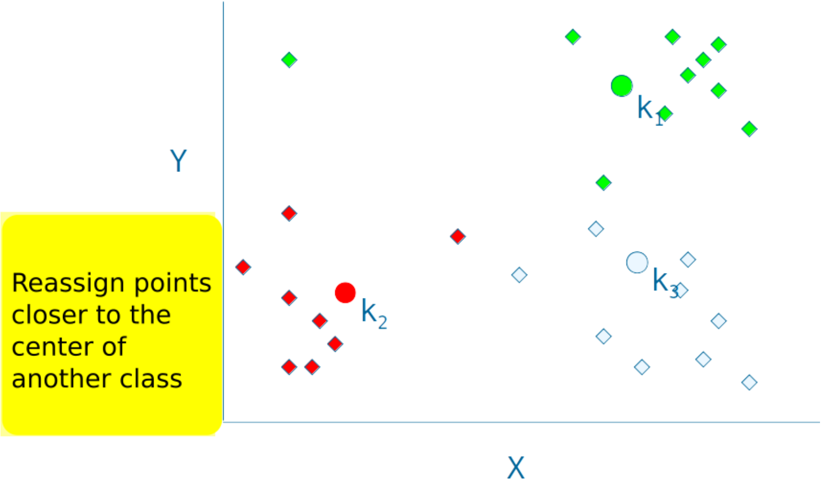


Image segmentation

K-means algorithm

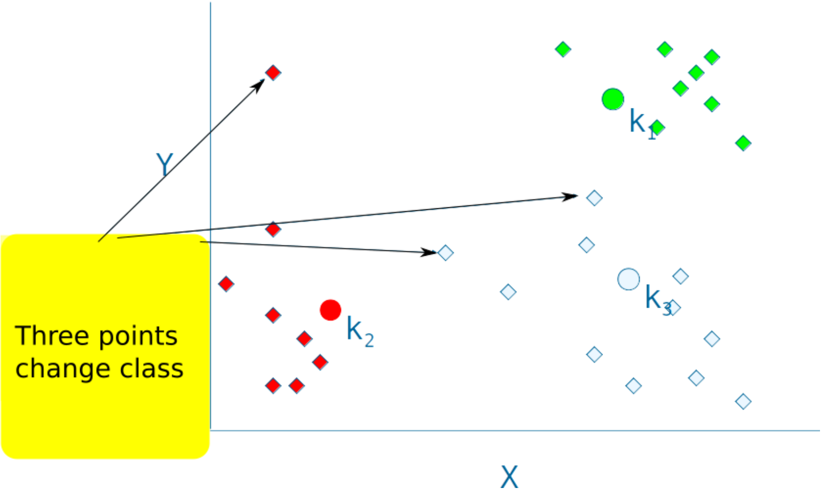


Image segmentation

K-means algorithm

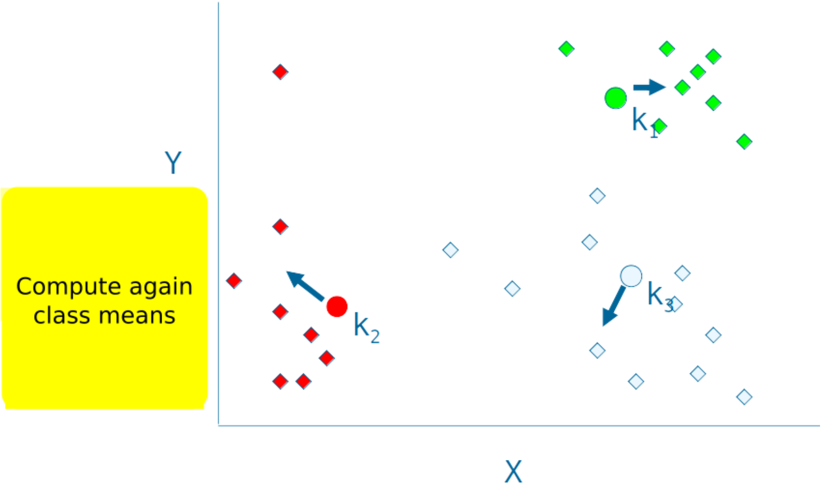
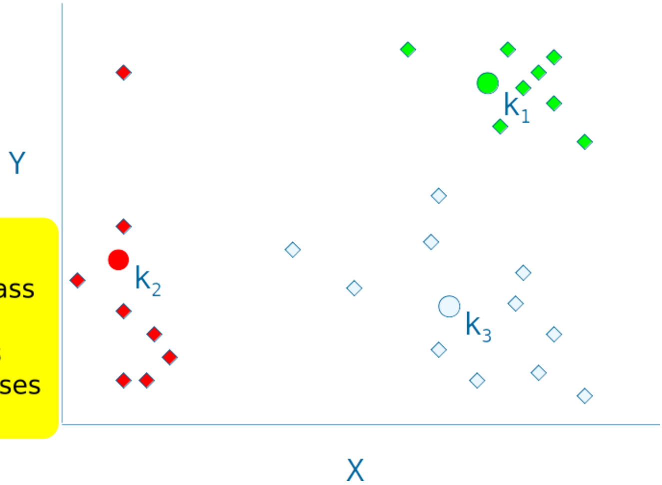


Image segmentation

K-means algorithm



Shifting class centers to the means of the classes

Image segmentation

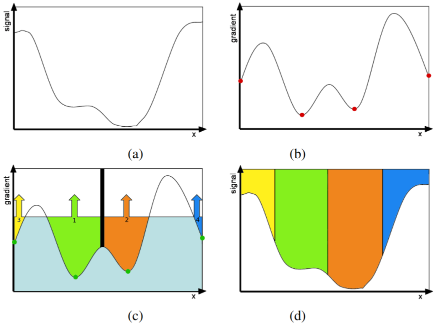
Properties of the K-means algorithm

- The advantage of this method is that it converges quickly to a local optimum.
- On the other hand, the local optimum encountered is highly dependent on the initial choice of class centers, so this is not necessarily a good solution.
- In general, in order to improve the solution, you need to try out several initial conditions.
- K-means, as described earlier, works quite well if the number of classes required is moderate.
- As the number of classes increases, the solution found by the algorithm is mediocre, or even very bad (2 or 3 times less good than the optimal solution).

Image segmentation

The watershed method

We consider a grayscale image as a topographic relief, and we simulate flooding.
One-dimensional illustration of Watershed.



(a) the signal to be segmented. (b) The corresponding signal gradient. (c) Gray-scale flooding around each local minimum. When two different floods meet, a dam is built. (d) The resulting signal segmentation.

source : [Roudier et al. 2008]

Image segmentation

Interactive segmentation : color watershed [Meyer 1992]

In practice, we need to annotate "seeds" on the objects we want to segment. These seeds can be defined automatically by pre-processing.

Image segmentation

Interactive segmentation : the Grab-cut [Rother, Kolmogorov, and Blake 2004]

1. User annotates objects to be segmented
2. The algorithm estimates the color distribution of the target object and the background using a Gaussian mixture model.
3. An optimization based on graph cutting (GraphCut) is performed.
4. This two-step procedure is repeated until convergence is reached.

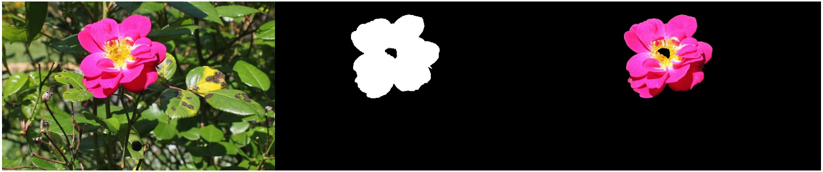
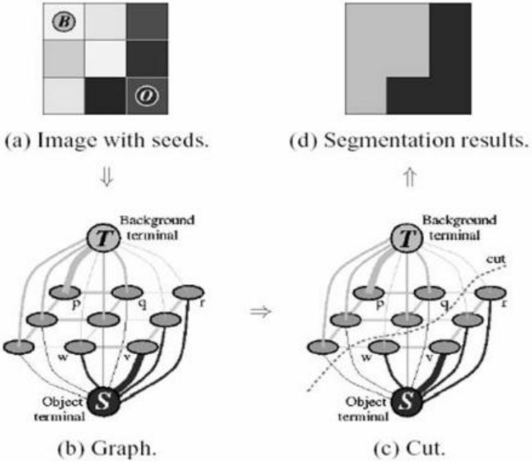


Image segmentation

Interactive segmentation : the Grab-cut [Rother, Kolmogorov, and Blake 2004]



source : http://www.cs.ru.ac.za/research/g02m1682/

Image segmentation

Démo openCV/Python

File LARM_SEG.py

- Otsu
- K-Means
- Watershed
- GrabCut

And now ?

Once the image has been partitioned into several "consistent" areas, how can we tell which objects are contained in these areas?

Objects must be characterized by extracting invariant (to illumination, geometric transformations, occlusions, etc.) features. In other words, features calculated for an object do not vary (or vary only slightly) if its appearance in the image changes.

We call these features **image or shape descriptors** !

Feature extraction : descriptors

Originally, descriptors were defined “Handmade” by making assumptions about the type of object to be recognized.

For example, if we want to recognize a type of flower, we will extract the following descriptors:

- the mean color
- the number of petals
- the number of sepals
- the stem length
- etc.

Feature extraction : descriptors

Two types of descriptors can be calculated:

- Global descriptors (computed over the entire object or image)
- Local descriptors (computed near points of interest)

To be relevant, descriptors must have minimum intra-class variance and maximum inter-class variance!

Feature extraction : descriptors

Global descriptors

Here is a list of global descriptors from the literature:

- Color histograms [Van De Sande, Gevers, and Snoek 2009]
- Texture histograms [Manjunath et al. 2001]
- Hue and Zernike moments [Khotanzad and Hong 1990; Liao and Pawlak 1996]
- Fourier descriptors [Smach et al. 2008]
- Local Binary Patterns (LBP) [Ahonen, Hadid, and Pietikäinen 2004]
- etc.

Please note! This only works if the object to be recognized has been segmented beforehand!

Feature extraction : descriptors

Local descriptors

Local descriptors are features computed near points of interest. This method has several advantages! One can recognize an object when

- the background is not uniform;
- the object of interest cannot be easily segmented;
- the image contains several objects;
- there are wide variations in the object's point of view (3D rotation, affine transformations, etc.).

Feature extraction : descriptors

Extraction of points of interest

Points of interest are points in the image that respect some properties:

- It has a clear definition, preferably mathematically based;
- It has a well-defined position in image space;
- The local structure of the image around the point of interest is rich in local information content;
- It is stable under local and global disturbances in the image domain like illumination/brightness variations, so that points of interest can be reliably calculated with a high degree of repeatability.

Point of interest detectors

We can extract points of interest of several types:

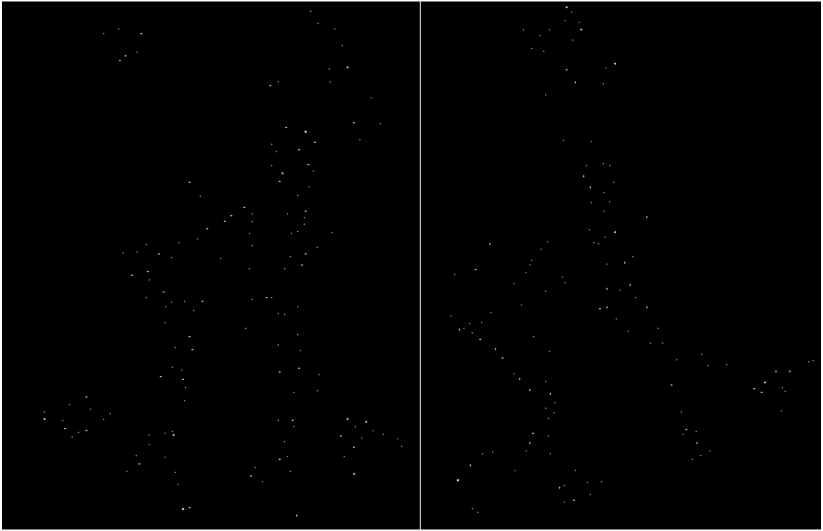
- The Moravec detector[Moravec 1980]
- The Harris detector [Harris, Stephens, et al. 1988]
- The DoG (Difference of Gaussians) [Lowe 2004], the LoG (Laplacian of Gaussian) [Lindeberg 1998] or the DoH (Determinant of Hessians) [Lindeberg 1998].

Point of interest detectors

The Harris detector [Harris, Stephens, et al. 1988]

Point of interest detectors

The Harris detector [Harris, Stephens, et al. 1988]



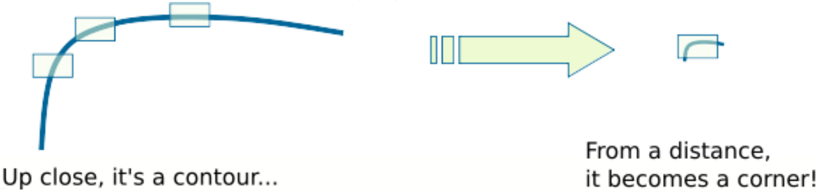
Point of interest detectors

The Harris detector [Harris, Stephens, et al. 1988]

The points are invariant to:

- 2D rotations;
- Lighting changes;
- Orientation of the object;
- Point of view.

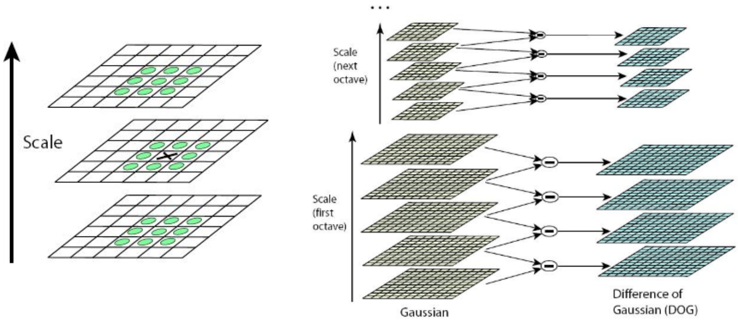
But no invariance when changing scale !!!



Point of interest detectors

DoG detector [Lowe 2004]

Scale invariance is obtained with DoG !



source : [Lowe 2004]

Local descriptors

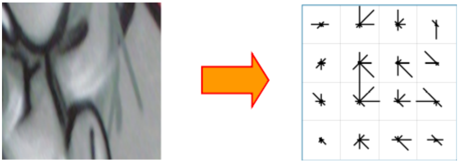
Some local descriptors of the State of the Art :

- the SIFTs [Lowe 2004]
- the SURFs [Bay et al. 2008]
- the ORBs [Rublee et al. 2011]
- compute global descriptors in the neighborhood of points of interest (Fourier, Moments, LBP, etc.)

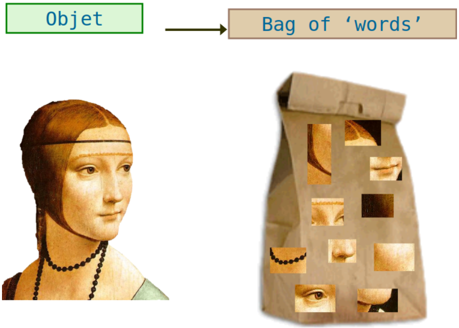
Local descriptors

SIFT descriptors[Lowe 2004]

- Get the scale (by DoG maximization in scale and space)
- Determine local focus (dominant gradient direction)
- Compute the histogram
- Divide the patch into 4x4 sub-patches
- Compute the histogram of gradient orientations (on 8 reference angles) within each sub-patch 16 windows, i.e. 128 values for each point (4x4 histograms of 8 bins)
- Normalize descriptor to be invariant to intensity change



A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]



A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

Method

First, take a set of images,

- extract visual features;
- build a "dictionary" or "visual vocabulary" → a list of common features.

Given a new image,

- extract features;
- for each feature, build a histogram;
- find the nearest visual word in the dictionary.

A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

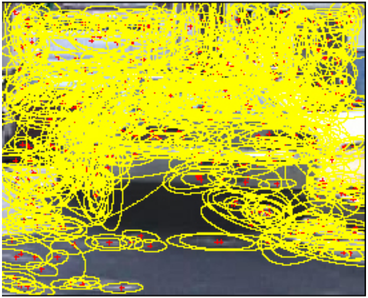
1 - Extract features

A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

1 - Extract features

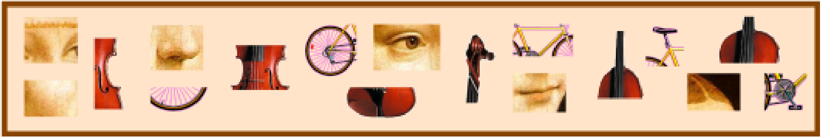
On a regular grid

With a point-of-interest detector



A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

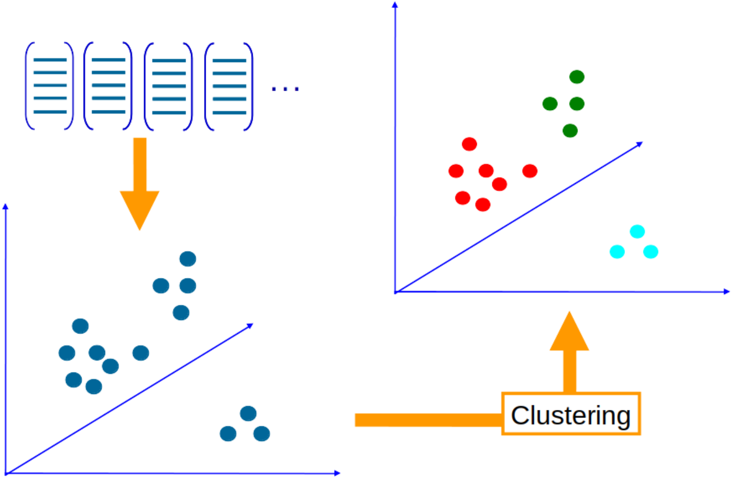
2 - Learning a "visual vocabulary"



A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

2 - Learning a "visual vocabulary"

Using the K-means algorithm seen above

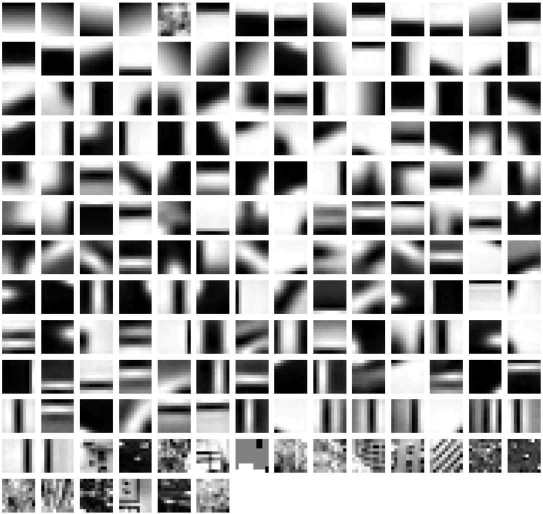


The visual vocabulary corresponds to the K centers of the classes obtained!

A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

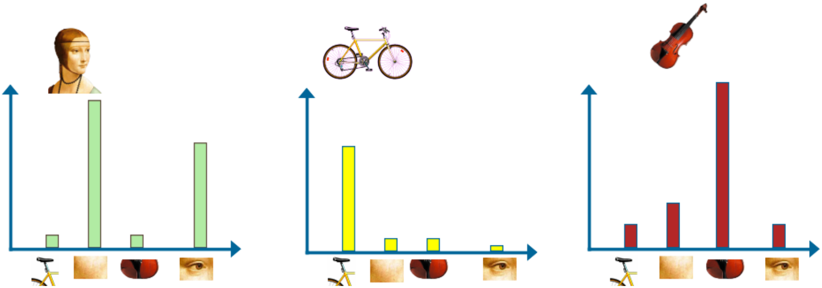
2 - Learning a "visual vocabulary"

Example of visual vocabulary:



A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

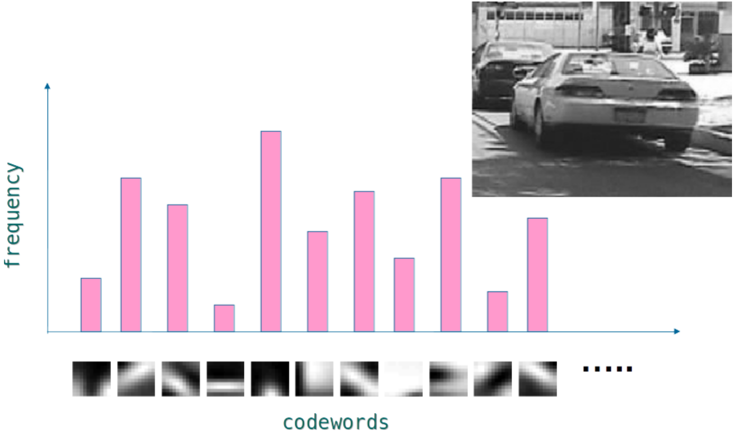
3 - Quantify features using "visual vocabulary"
4 - Represent each image by the frequencies of "visual words".



A representation of the image: the BAG OF WORDS [Sivic and Zisserman 2008]

3 - Quantify features using "visual vocabulary"

4 - Represent each image by the frequencies of "visual words".



From descriptors to object recognition

First solution $\rightarrow$ **Image matching** :

On dataset images :

- Detect points of interest;
- Compute visual features around these points;
- Export features in a usable format.

For a test image :

- Detect points of interest;
- Compute visual features around these points;
- Computes matches between these features and those in the database
- Find the corresponding image(s) and rank them their similarity score (from closest to farthest).

From descriptors to object recognition

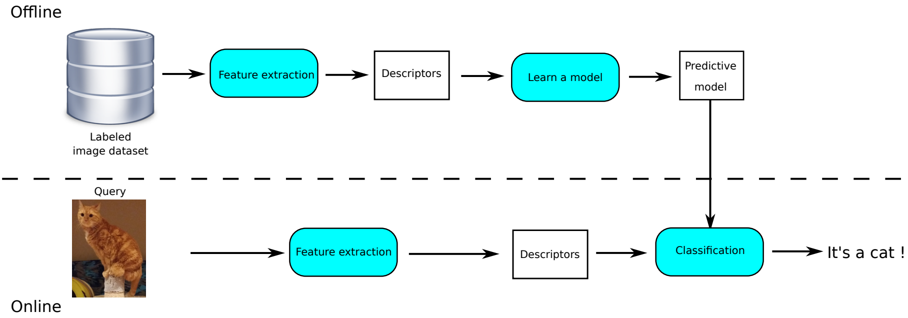
Second solution → Machine learning

1. Learn a model from a learning dataset
2. Matching the features of the test image to the model

Machine learning

Machine learning, artificial learning or statistical learning is a field of study in artificial intelligence that relies on mathematical and statistical approaches to give computers the ability to learn from data, i.e. to improve their performance in solving tasks without being explicitly programmed for each one. More broadly, it concerns the design, analysis, optimization, development and implementation of such methods.

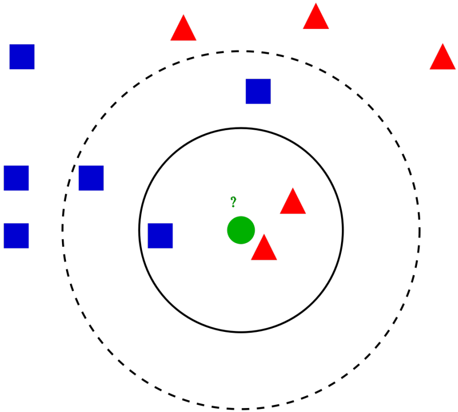
Machine learning
Supervised learning



Machine learning

The KNN (K - nearest neighbors) [Fix 1951]

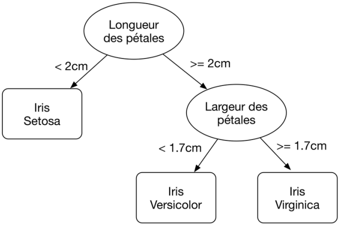
The method is simple: we compute the distance between the descriptors of the query and the descriptors in the database. Then, we consider the classes of the K-closest observations. The most represented class will be chosen.



Machine learning

Decision Trees

- One of the input variables is selected at each internal node of the tree.
- Each edge to a child node corresponds to a set of values of an input variable.
- Each leaf represents either a value of the target variable, or a probability distribution of the various possible values of the target variable.
- The combination of input variable values is represented by the path from the root to the leaf.



For exemple, one can use Random Forest [Breiman 2001].

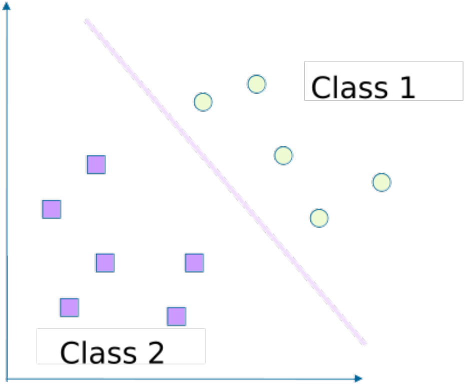
Machine learning

Support vector machine : SVM [Duda, Hart, et al. 2006]

- Dealing with non-linear discrimination problems, and reformulating the classification problem as a quadratic optimization problem
- Notion of maximum margin :
  - the distance between the separation boundary and the nearest support vector samples must be as great as possible.
- The input data are embedded in a higher-dimensional space in which a linear separator is likely to exist.
  - Kernel functions !

Machine learning

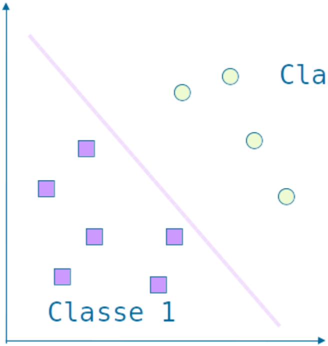
Linearly separable two-class problem

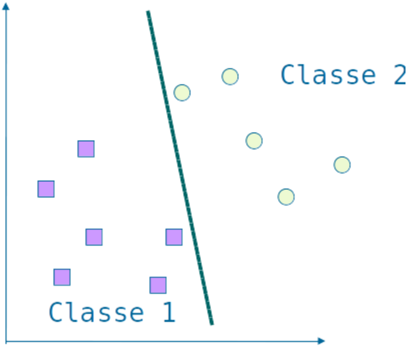


- Several decision surfaces exist to separate the classes
- Which one to choose?

Machine learning

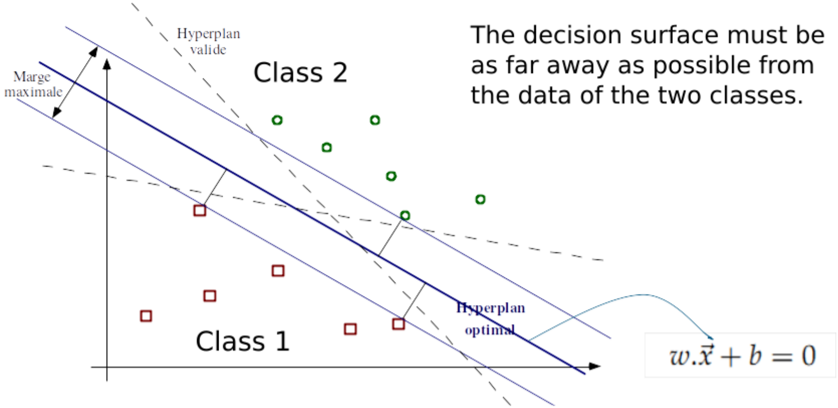
Example of a bad choices





Machine learning

Hyperplane with wider margin

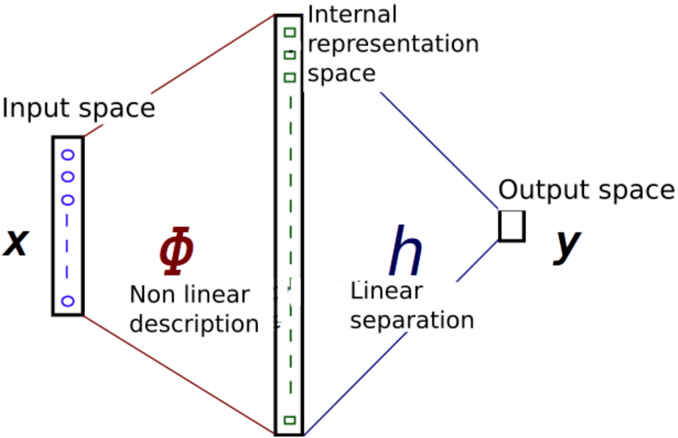


The decision surface must be as far away as possible from the data of the two classes.

Machine learning

Extension to a non-linear separation surface

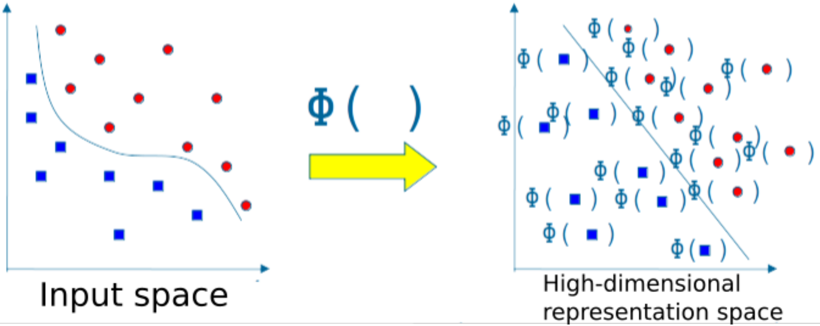
- Idea ! simplify things by transforming X into a higher-dimensional space



Machine learning

Extension to a non-linear separation surface

- Potential problems with the transformation
  - High computation effort and difficulty in obtaining a good estimate
- SVM with kernel functions solves both problems
  - Using kernel functions for computational efficiency



Machine learning

The kernel trick

- Find $\phi(\ )$ is very hard !
- Definition of a kernel function :

$$K(x, y) = \langle \phi(x), \phi(y) \rangle$$

- K(x,y) intuitively represents the similarity between x and y, obtained from our a priori knowledge
- However, K(x,y) must satisfy certain conditions for the corresponding $\phi(\ )$ to exist.

Machine learning

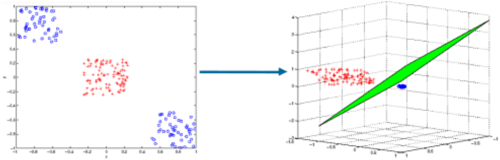
The kernel trick

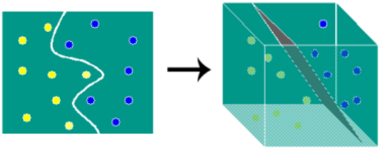
Solving a non-linear case

We change dimension using a kernel function

- Linear $K(x, y) = x * y$
- Polynomial $K(x, y) = (x * y + \theta)^d$
- Gaussian $K(x, y) = exp(-\alpha||x - y||^2)$
- Perceptron $K(x, y) = -\|x - y\|_2$

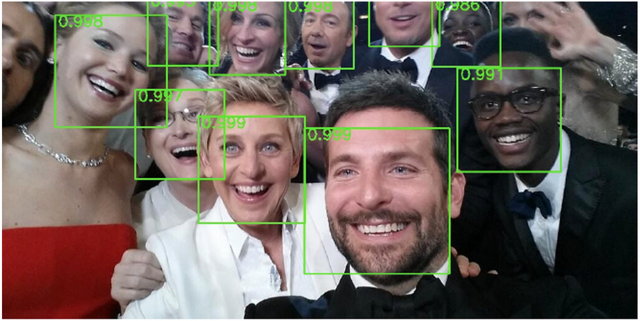
The new space is called "description space".





Object detection

Object detection involves estimating the position (often a bounding box) of some type of object in an image. One or more objects can be detected in the same image.

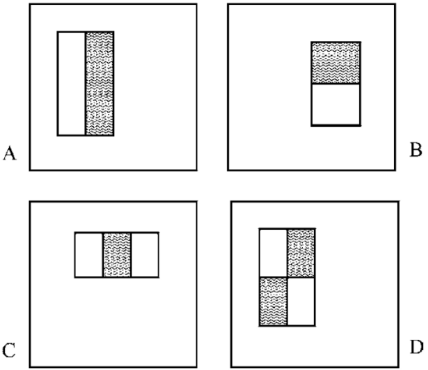


Face detection is now a common feature of digital cameras and cell phones.

Object detection

Viola and Jones algorithm[Viola, Jones, et al. 2001]

The method is to exhaustively search within the image for image parts containing the desired object. This exhaustive search can be very time-consuming. To make this search as quick as possible, features that can be computed quickly have been proposed (pseudo-Haar features).



Object detection
Classification

For learning, Viola-Jones uses a variant of Adaboost classifier.

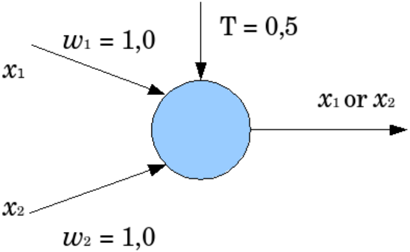
The principle comes from the combination of weak classifiers (also called hypotheses). By successive iterations, the knowledge of a weak classifier is added to the final classifier (strong classifier).

Originally dedicated to face detection, the Viola-Jones algorithm can be trained using the openCV library to detect any object.
See https://docs.opencv.org/2.4/doc/user_guide/ug_traincascade.html

Neural Networks

What is a neuron?

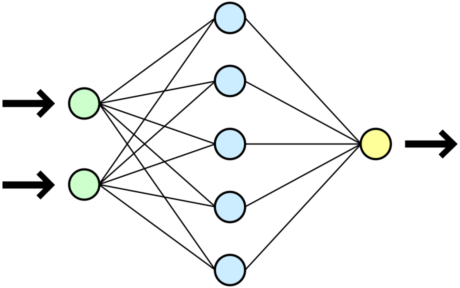
A formal neuron is a mathematical and computational representation of a biological neuron. In its simplest version, a formal neuron calculates the weighted sum of the inputs received, then applies an activation function, generally non-linear, to this value. The final value obtained is the output of the neuron.



Neural Networks

What is a neural network ?

A neural network is a computer system that enables learning to be carried out by the cooperation of a set of neurons structured in layers. Data is passed to the input layer of the neural network. At the output of the network, one or more outputs are retrieved.

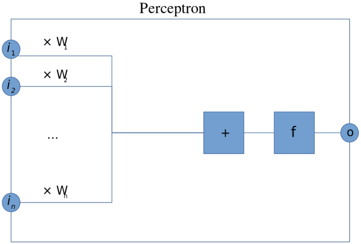


Learning corresponds to calculating the weights of each input connected to the neuron.

Neural Networks

The perceptron [Rosenblatt 1958]

This is the simplest type of neural network. It enables linear classification. Descriptors can be given as input and, as output, we can have one or more outputs corresponding to each of the classes considered.



$$o = f\left(\sum_{k=1}^{n} i_k \cdot W_k\right)$$

Source : Wikipédia

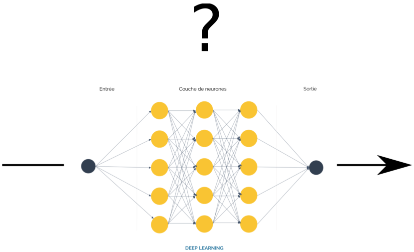
For multi-layer perceptrons, weights can be learned using a backpropagation algorithm. [Kröse et al. 1993].

Neural Networks

The backpropagation

Neural Networks

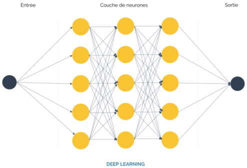
The backpropagation



Neural Networks

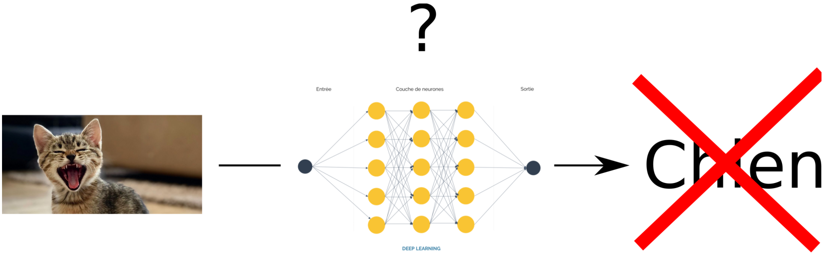
The backpropagation

?



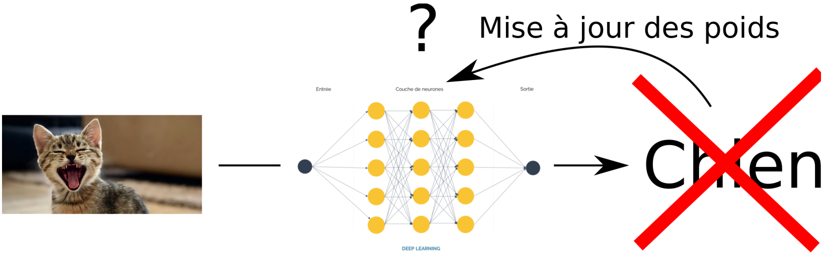
Chien

Neural Networks
The backpropagation



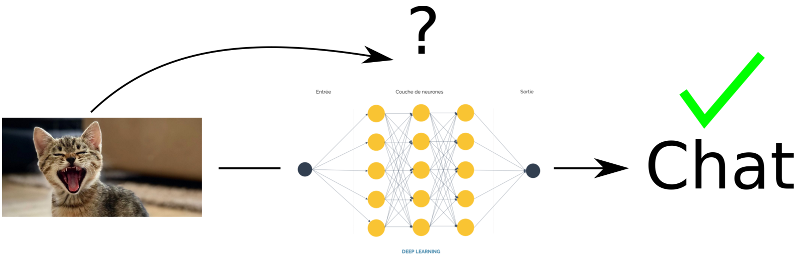
Neural Networks

The backpropagation



Neural Networks

The backpropagation



Neural Networks

Deep-Learning

Nowadays, most computer vision processing is carried out using deep neural networks (Deep Learning). These are neural networks containing many layers of neurons.

In computer vision, one type of neural network is used in particular for its properties: $\rightarrow$ convolutional neural networks.

Neural Networks

Convolutional neural networks [LeCun et al. 1998]

The aim of convolutional neural networks is to learn how to extract the most discriminating features from images in order to carry out a task, unlike previous "handcrafted" approaches.

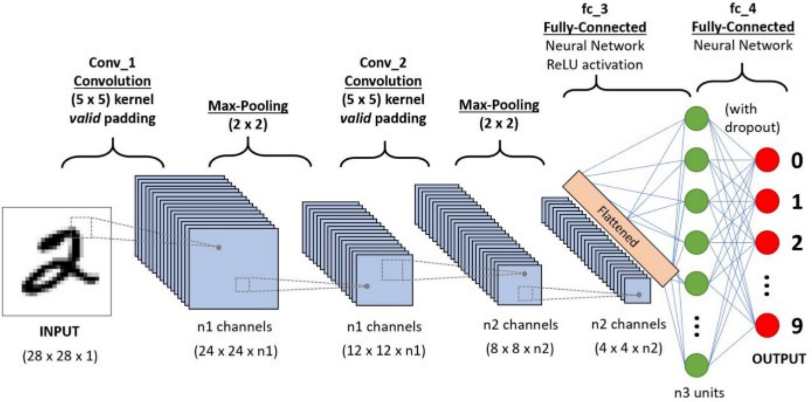
To do this, they perform convolutions on the images to extract discriminating features. To extract features at different "scales", pooling (reducing the image size) is performed between each convolution.

At initialization, the convolution filters are set at random. During training, they evolve according to the database and the task to be performed.

Neural Networks

Convolutional neural networks [LeCun et al. 1998]

Example of a convolutional neural network for recognizing handwritten digits.



Source : https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Neural Networks

Networks are becoming increasingly complex...

One of the most widely used networks is called an auto-encoder. It consists of a convolutional network (encoder) followed by a so-called "deconvolution" network (decoder).

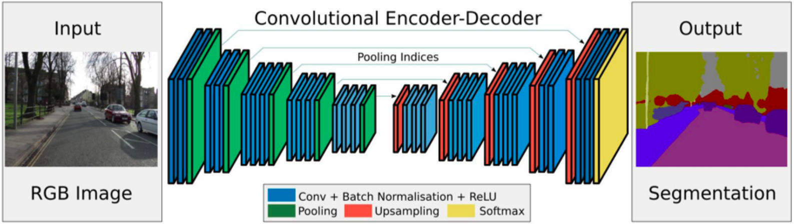
The encoder compresses the signal to obtain only the most important information.

The decoder allows to return to the original configuration (an image with the same dimensions as the original).

Neural Networks

... for increasingly complex tasks

Example of a encoding-decoding neural network for semantic image segmentation. Semantic segmentation corresponds to image segmentation and recognition of segmented objects at the same time. Each pixel in the image is assigned a class.



Source : [Badrinarayanan, Kendall, and Cipolla 2017]

References I

- Ahonen, Timo, Abdenour Hadid, and Matti Pietikäinen. "Face recognition with local binary patterns". In: *European conference on computer vision*. Springer. 2004, pp. 469–481.
- Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation". In: *IEEE transactions on pattern analysis and machine intelligence* 39.12 (2017), pp. 2481–2495.
- Bay, Herbert et al. "Speeded-up robust features (SURF)". In: *Computer vision and image understanding* 110.3 (2008), pp. 346–359.
- Berger, L. *Traitement d'images et de vidéos avec OpenCV 4 en Python (Windows, Linux, Raspberry)*. Éditions D-BookeR, 2020. ISBN: 9782822707954.
- Breiman, Leo. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

References II

- Duda, Richard O, Peter E Hart, et al. *Pattern classification*. John Wiley & Sons, 2006.
- Fix, Evelyn. *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF School of Aviation Medicine, 1951.
- Fourier, Jean Baptiste Joseph baron. *Théorie analytique de la chaleur*. F. Didot, 1822.
- Gonzales, Rafael C and Richard E Woods. *Digital image processing*. 2002.
- Gonzalez, Rafael C, Richard Eugene Woods, and Steven L Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.
- Harris, Christopher G, Mike Stephens, et al. “A combined corner and edge detector.”. In: *Alvey vision conference*. Vol. 15. 50. Citeseer. 1988, pp. 10–5244.

References III

- Khotanzad, Alireza and Yaw Hua Hong. “Invariant image recognition by Zernike moments”. In: *IEEE Transactions on pattern analysis and machine intelligence* 12.5 (1990), pp. 489–497.
- Kröse, Ben et al. “An introduction to neural networks”. In: (1993).
- LeCun, Yann et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.
- Liao, Simon Xinmeng and Miroslaw Pawlak. “On image analysis by moments”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18.3 (1996), pp. 254–266.
- Lindeberg, Tony. “Feature detection with automatic scale selection”. In: *International journal of computer vision* 30.2 (1998), pp. 79–116.
- Lowe, David G. “Distinctive image features from scale-invariant keypoints”. In: *International journal of computer vision* 60.2 (2004), pp. 91–110.

References IV

- MacQueen, James et al. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA. 1967, pp. 281–297.
- Manjunath, Bangalore S et al. "Color and texture descriptors". In: *IEEE Transactions on circuits and systems for video technology* 11.6 (2001), pp. 703–715.
- Meyer, Fernand. "Color image segmentation". In: *1992 International Conference on Image Processing and its Applications*. IET. 1992, pp. 303–306.
- Moravec, Hans. "Obstacle avoidance and navigation in the real world by a seeing robot". PhD thesis. Ph. D. Thesis, Stanford University, Stanford, CA, USA, 1980.
- Otsu, Nobuyuki. "A threshold selection method from gray-level histograms". In: *IEEE transactions on systems, man, and cybernetics* 9.1 (1979), pp. 62–66.

References V

- Pizer, Stephen M et al. "Adaptive histogram equalization and its variations". In: *Computer vision, graphics, and image processing* 39.3 (1987), pp. 355–368.
- Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain.". In: *Psychological review* 65.6 (1958), p. 386.
- Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. """ GrabCut" interactive foreground extraction using iterated graph cuts". In: *ACM transactions on graphics (TOG)* 23.3 (2004), pp. 309–314.
- Roudier, Pierre et al. "Management zone delineation using a modified watershed algorithm". In: *Precision Agriculture* 9.5 (2008), p. 233.
- Rublee, Ethan et al. "ORB: An efficient alternative to SIFT or SURF". In: *2011 International conference on computer vision*. Ieee. 2011, pp. 2564–2571.

References VI

- Sivic, Josef and Andrew Zisserman. "Efficient visual search of videos cast as text retrieval". In: *IEEE transactions on pattern analysis and machine intelligence* 31.4 (2008), pp. 591–606.
- Smach, Fethi et al. "Generalized Fourier descriptors with applications to objects recognition in SVM context". In: *Journal of mathematical imaging and Vision* 30.1 (2008), pp. 43–71.
- Tou, Julius T and Rafael C Gonzalez. "Pattern recognition principles". In: (1974).
- Van De Sande, Koen, Theo Gevers, and Cees Snoek. "Evaluating color descriptors for object and scene recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 32.9 (2009), pp. 1582–1596.
- Viola, Paul, Michael Jones, et al. "Robust real-time object detection". In: *International journal of computer vision* 4.34-47 (2001), p. 4.