



# TECHNIQUES MULTIMÉDIA

Mehdi Louizi

Mehdi.louizi@gmail.com

# INTRODUCTION

## ○ Multimédia?

- Un média est un mode de représentation de l'information clairement caractérisable
- Il peut être
  - **Discret** (indépendant du temps)
    - Graphique
    - Image fixe
  - **Continu** (dépendant du temps)
    - Son
    - Images animées

# INTRODUCTION

- Multimédia?

- *Digital Content*
- Toute application utilisant ou servant à travailler sur au moins un média spécifique
- En informatique :
  - Réunion sur un même support de plusieurs média.

# INTRODUCTION



# INTRODUCTION

- Intérêt

- Des études portant sur l'interactivité montrent que l'être humain ne retient que :
  - 10 % de ce qu'il LIT
  - 50 % de ce qu'il ENTEND et LIT
  - 70 % de ce qu'il ENTEND, LIT et VOIT

# INTRODUCTION

- Domaines :
  - Production audio visuelle numérique (CD, film, vidéo)
  - Edition Hors ligne (CD, DVD, Blu Ray)
  - Edition En ligne (site web, télévision interactive, WAP, 3G...)
  - ....

# INTRODUCTION

## ○ Métiers

- Journaliste
- Graphiste
- Webdesigner
- Animateur 3d
- Monteur vidéo
- Infographiste
- Art numérique
- Développement
- Impression...

# INTRODUCTION

- Technologie matérielle

- Numérisation

- Son
- Vidéo
- Image

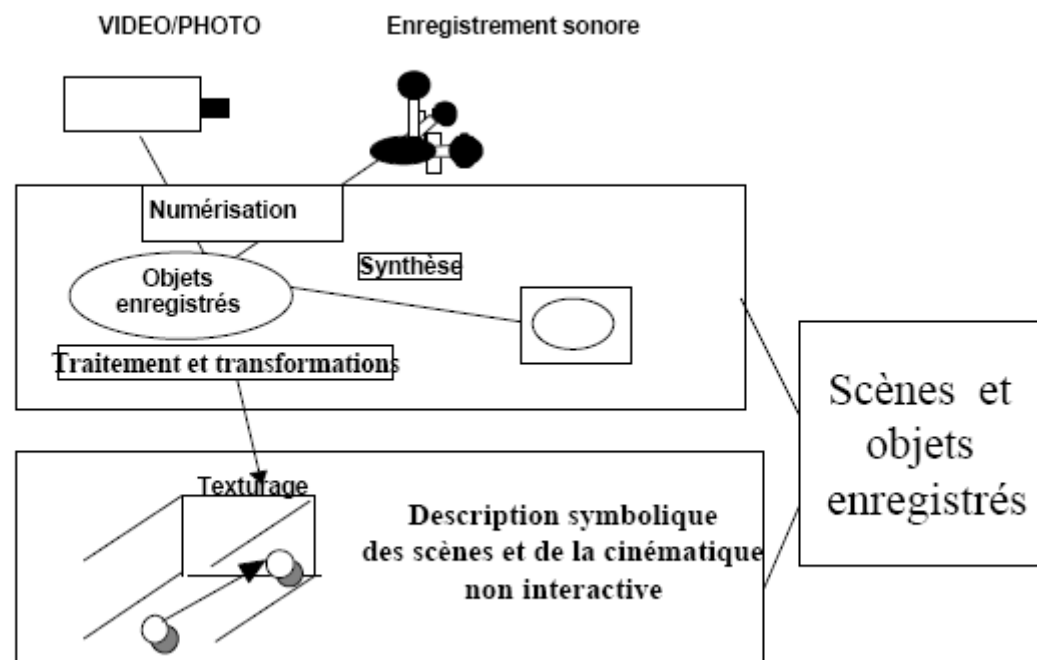
- Accroissement des capacités de stockage et des débits de transfert
- Dispositifs de compression et de décompression
- Accroissement des puissances de calcul

# INTRODUCTION

- Technologie logicielle
  - Représentation de données complexes (Hypertexte, réseau sémantiques)
  - Programmation orientée objet
  - Synthèse du son et de l'image

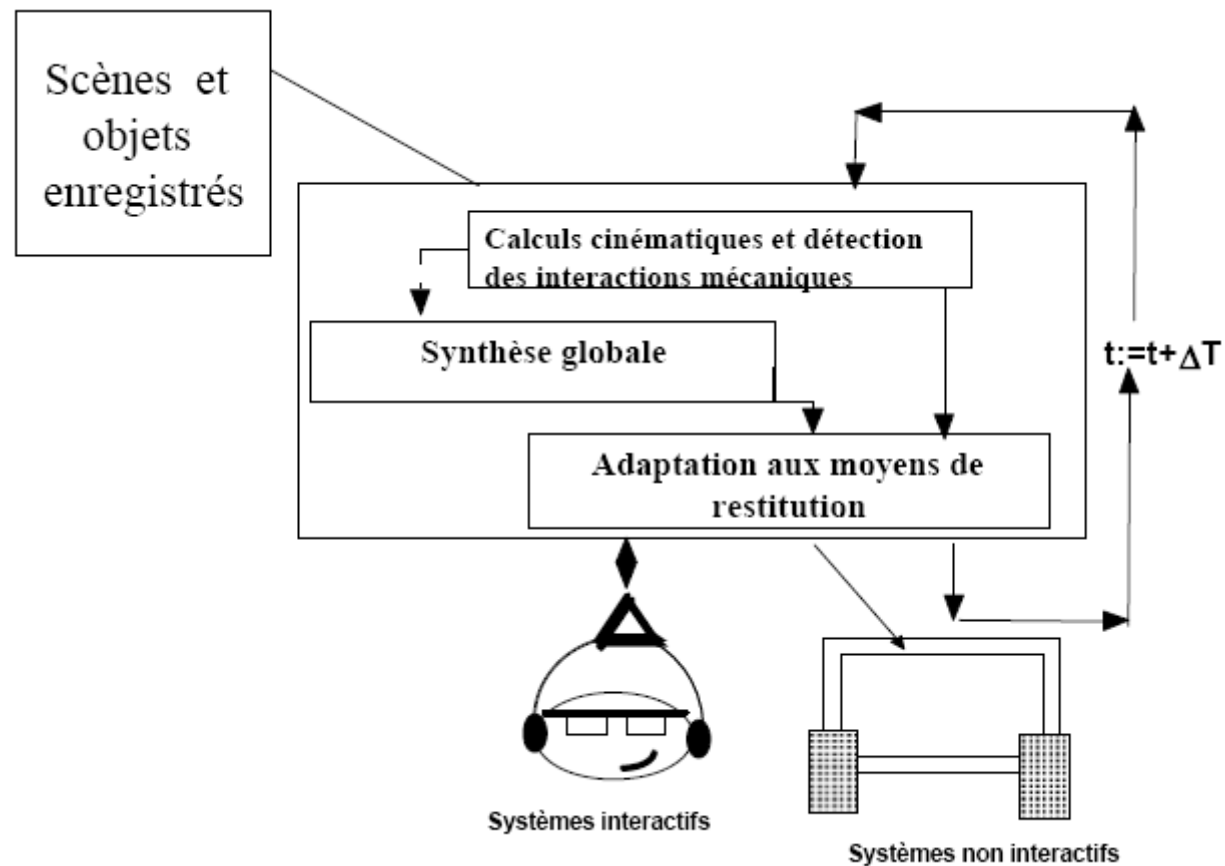
# INTRODUCTION

- Processus de création d'un objet multimédia (1)



# INTRODUCTION

- Processus de création d'un objet multimédia (2)



# INTRODUCTION

## ○ La scène

- C'est une description symbolique d'un univers virtuel
- Interprétable par des outils de synthèse et de visualisation
- Qui se présente comme un document structuré

# INTRODUCTION

- Que faut-il représenter ?
  - Un référentiel spatial et temporel
  - Du texte
  - Des objets virtuels
  - Des textures
  - Des comportements dynamiques
  - Des indications pour les outils de synthèse et de visualisation
- **Scène sonore** : position dans l'espace, acoustique des pièces, transformations à appliquer...

# LES IMAGES

## ○ Domaines d'application

- Vision industrielle
- Imagerie médicale
- Imagerie satellite
- Microscopie
- Télécommunications
- Animations, Images de synthèse...

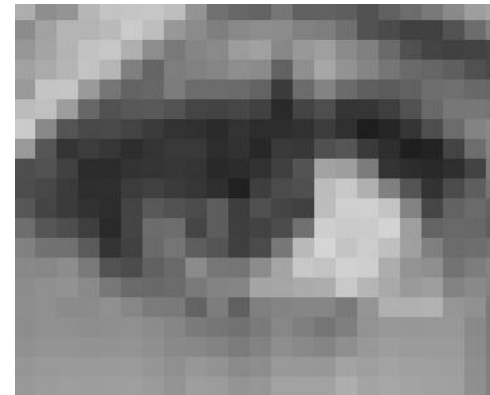
# L'IMAGE ET SES CARACTÉRISTIQUES

- Qu'est-ce qu'une image?



Réalité analogique

Discrétisation

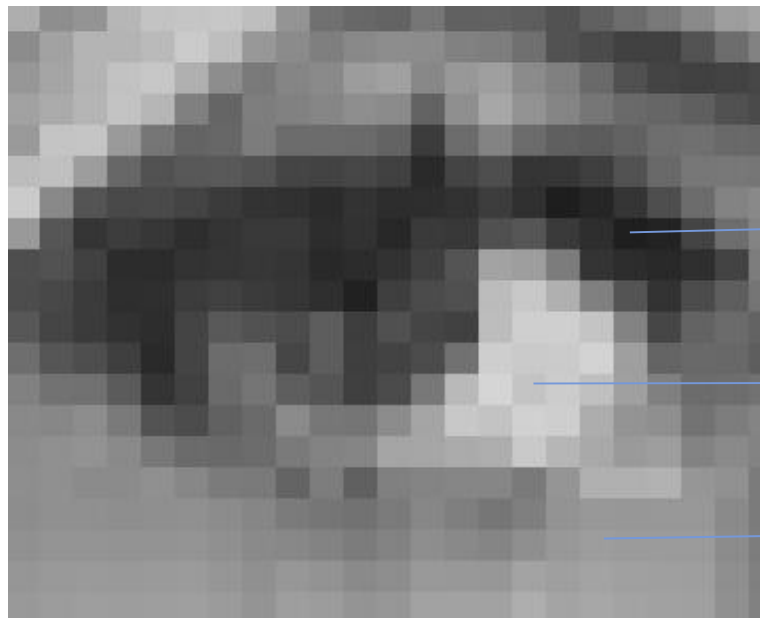


Représentation numérique

# L'IMAGE ET SES CARACTÉRISTIQUES

- Contenu de l'image

1 Image = 1 Ensemble de pixels  
1 valeur (ou plusieurs) par pixel



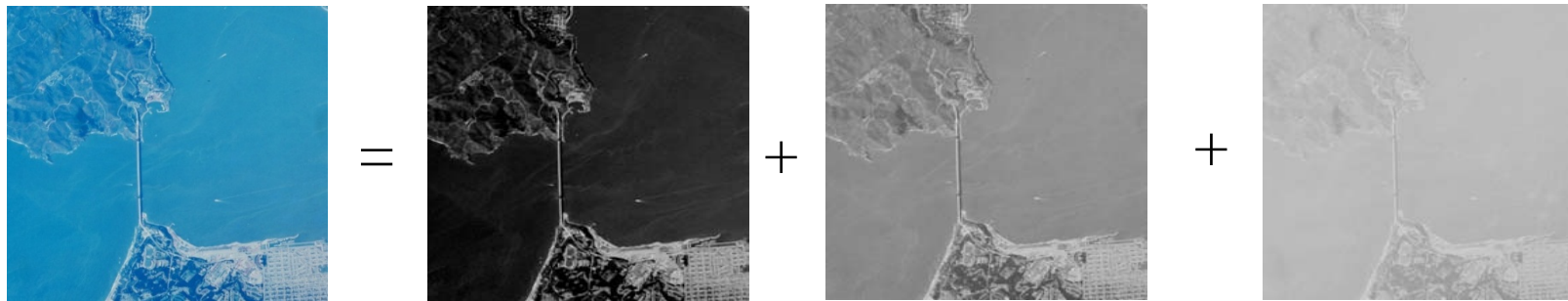
31

196

157

# IMAGES MULTIVALUÉES

- Plusieurs valeurs par pixel
  - Images couleurs

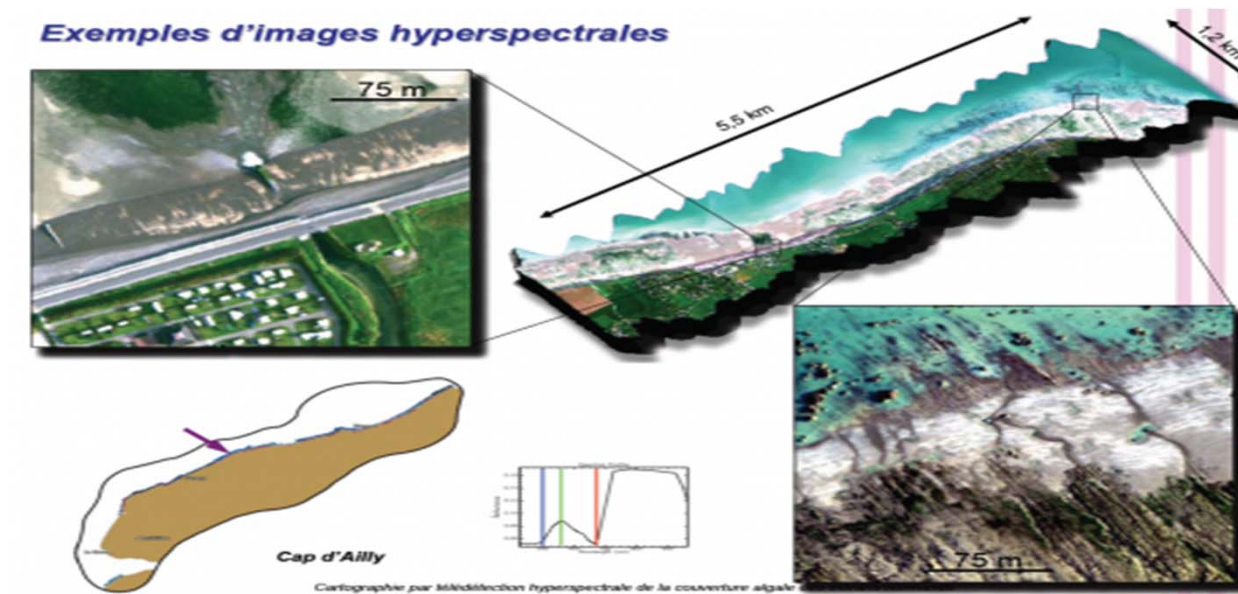


- Images multispectrales, hyperspectrales

# IMAGES MULTIVALUÉES

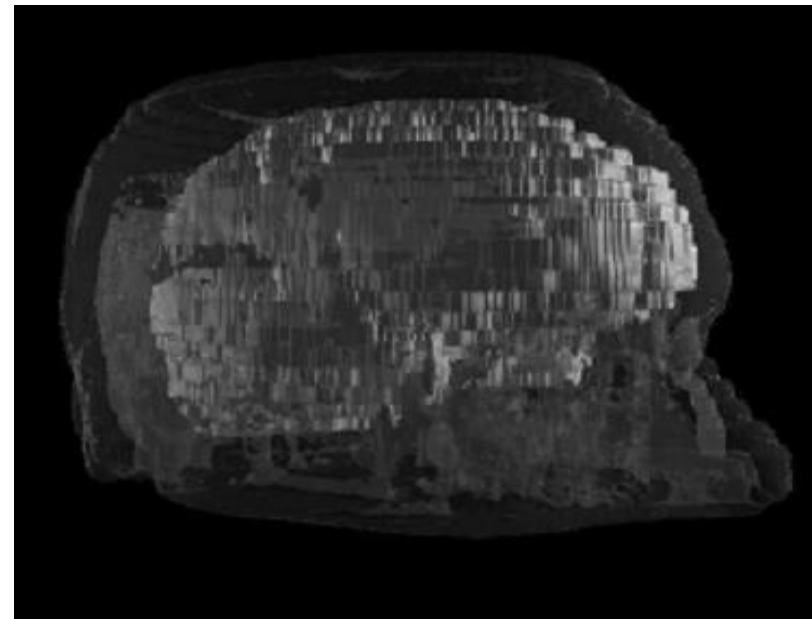
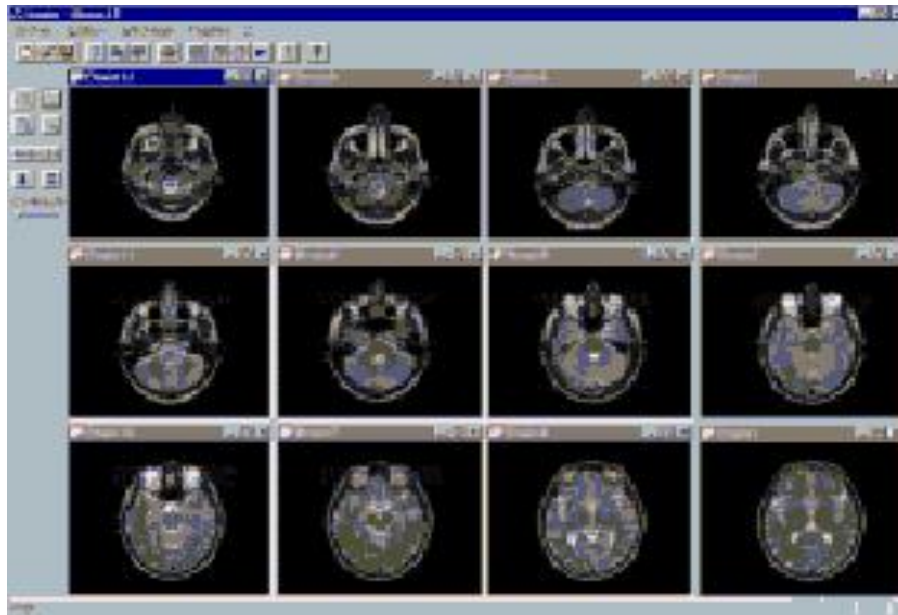


← Exemple d'images multispectrales



# VOLUMES 3D

- Empilement de coupes (images)



- Epaisseur de chaque coupe

# DES EXEMPLES D'APPLICATION



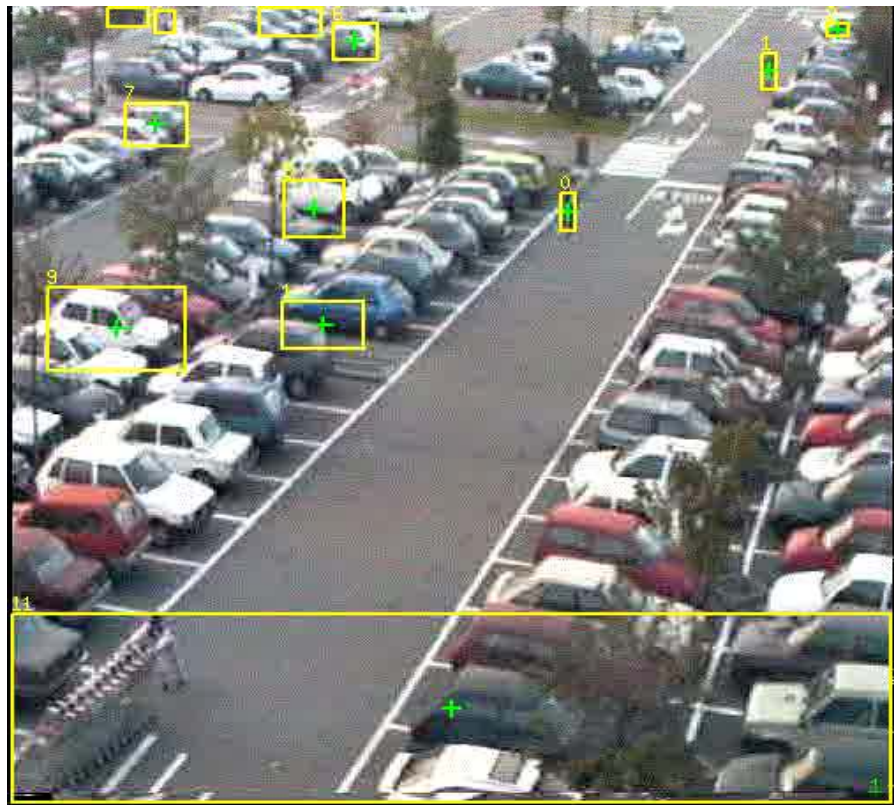
## DES EXEMPLES D'APPLICATION

- INA, Archives, ...
- Navigation et recherche dans les documents
- Identifier :
  - Changements de plan
  - Changements de scène
  - Mouvement de caméra
  - Mouvements des objets
  - ...



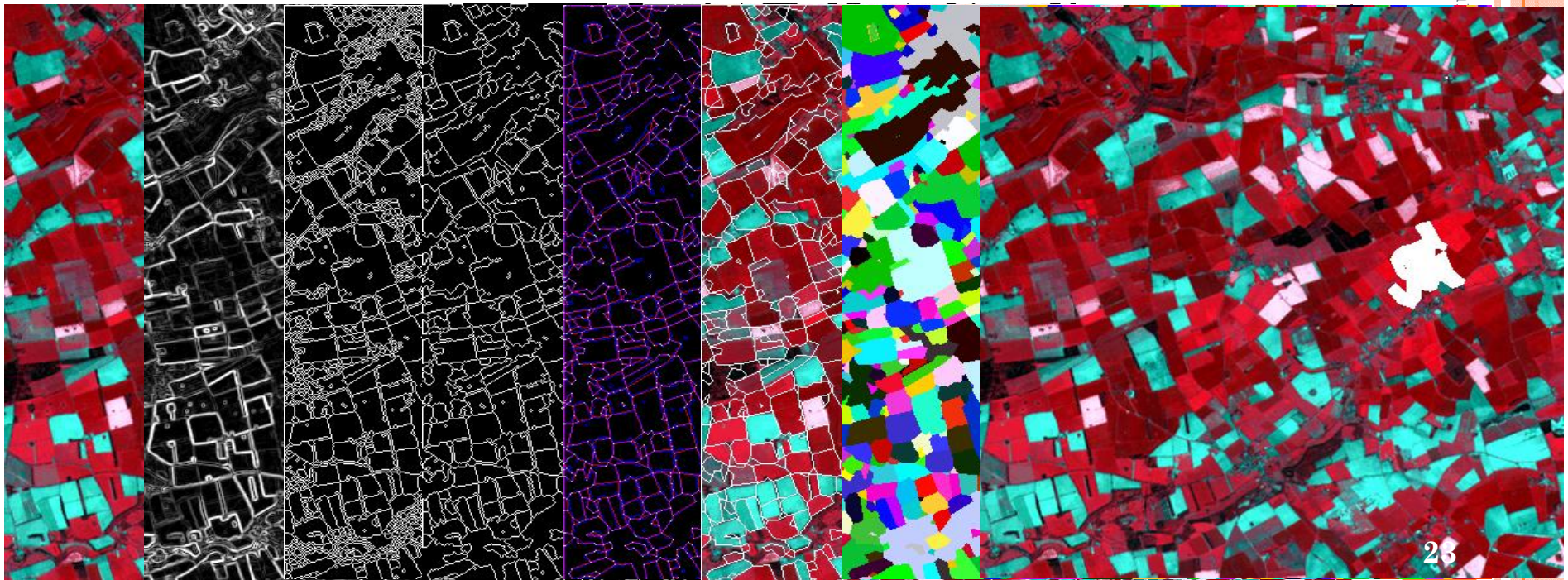
# DES EXEMPLES D'APPLICATION

- Vidéo-surveillance
  - Lieux publics
  - Détection d'activités « anormales »



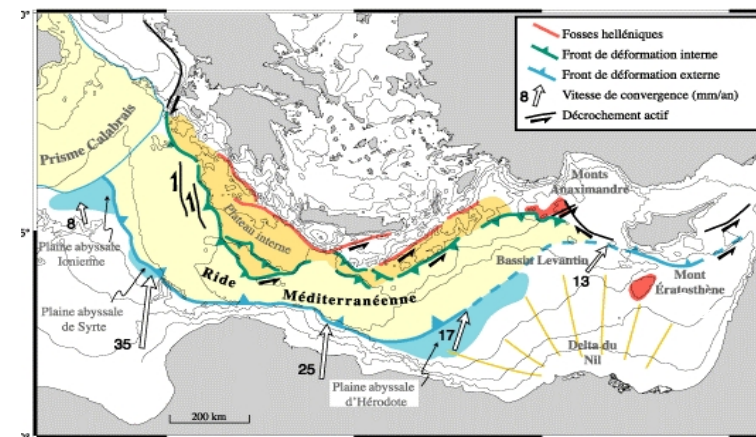
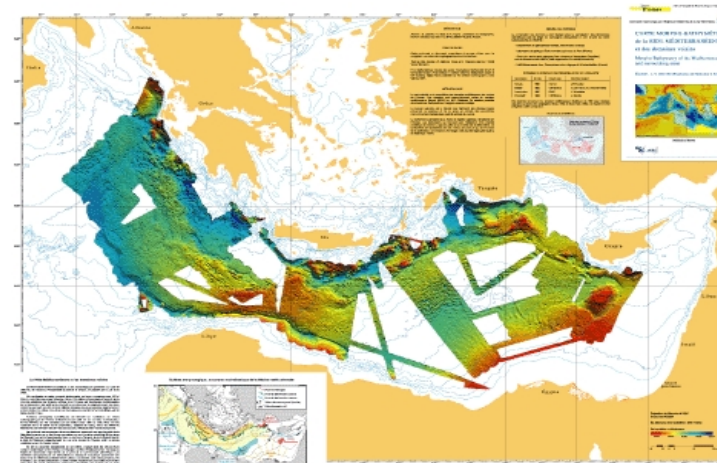
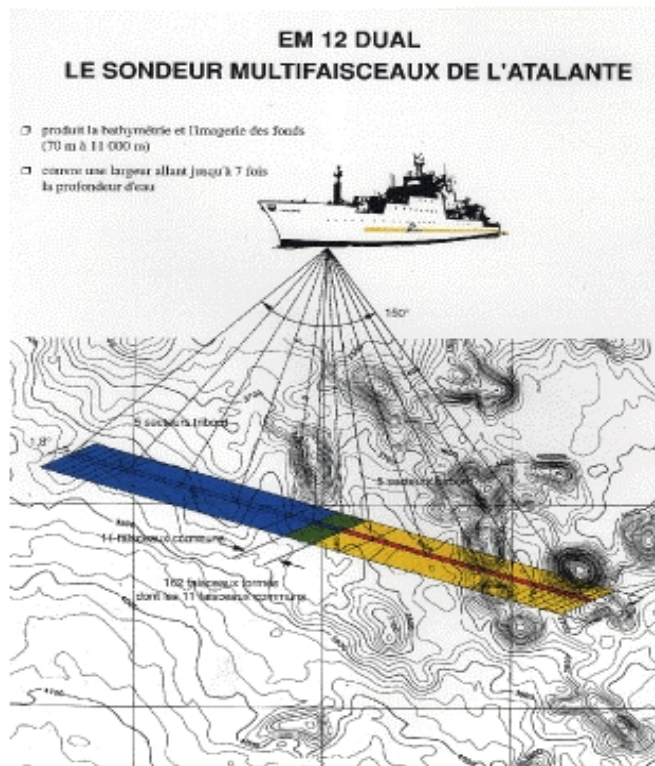
# DES EXEMPLES D'APPLICATION

- Imagerie satellite
  - Généralement multispectrale, hyperspectrale...



# DES EXEMPLES D'APPLICATION

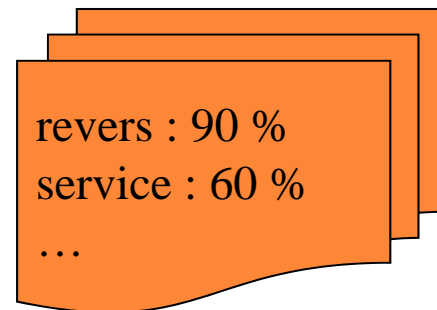
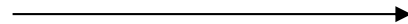
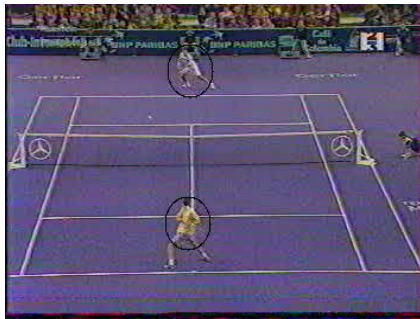
## ○ Imagerie sonar



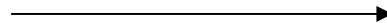
# DES EXEMPLES D'APPLICATION

- Analyse sportive

- Satisfaire l'entraîneur



- Satisfaire le supporter



# DES EXEMPLES D'APPLICATION

- Imagerie médicale

- Différentes modalités :
  - IRM
  - Scanner
  - Radiologie
  - Echographie
- De nombreux problèmes à résoudre
- But : aide au praticien (localisation, mesure)

# DES EXEMPLES D'APPLICATION

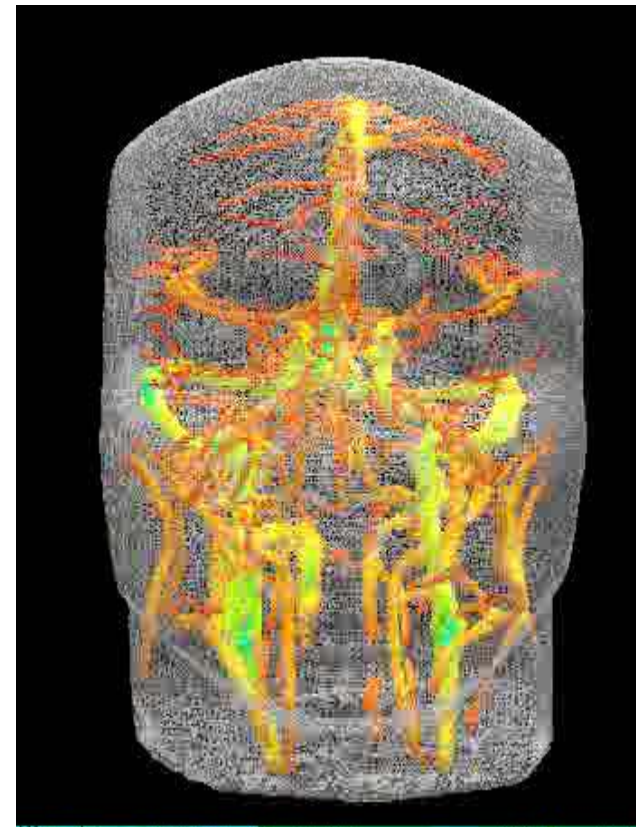
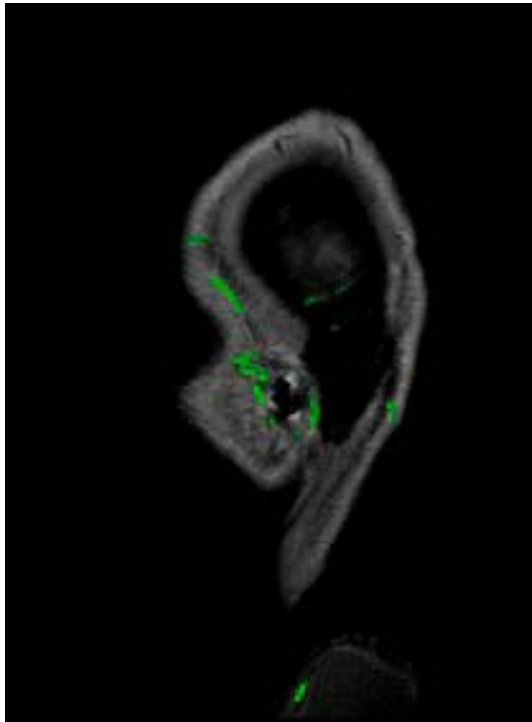
- Imagerie médicale
  - Mesures automatiques



Techniques Multimédia

## DES EXEMPLES D'APPLICATION

- Imagerie médicale
  - Analyse et reconstruction



# DES EXEMPLES D'APPLICATION

Titre, haut de page

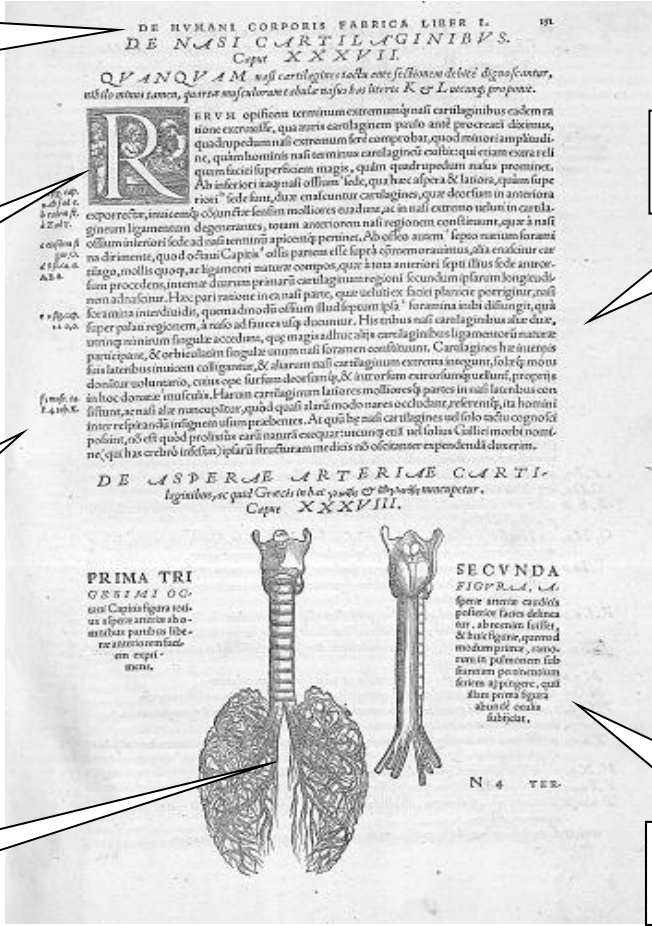
Lettrine

Note en marge

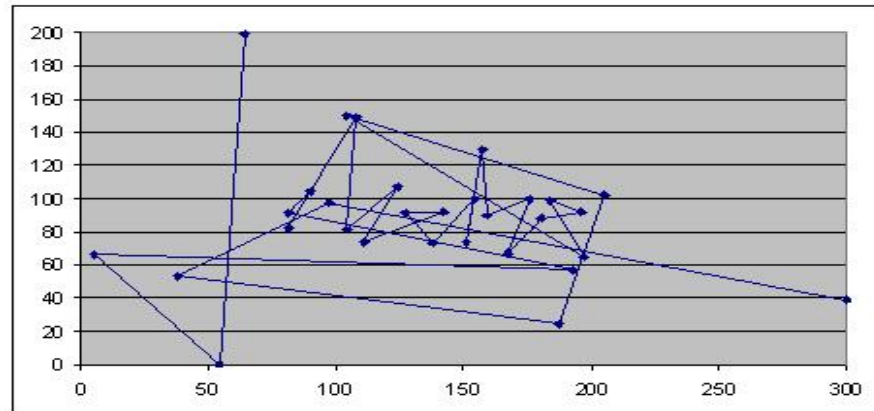
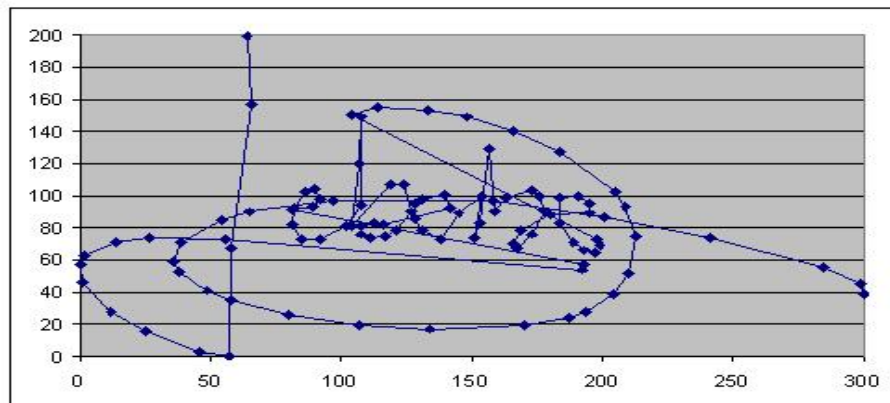
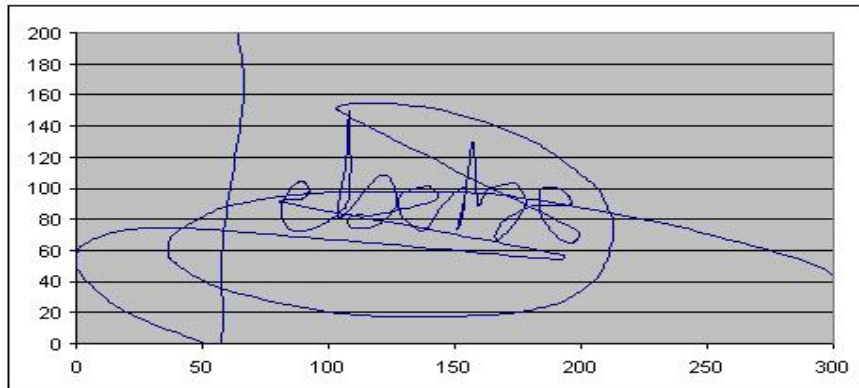
Illustration

Zone de texte

Légende

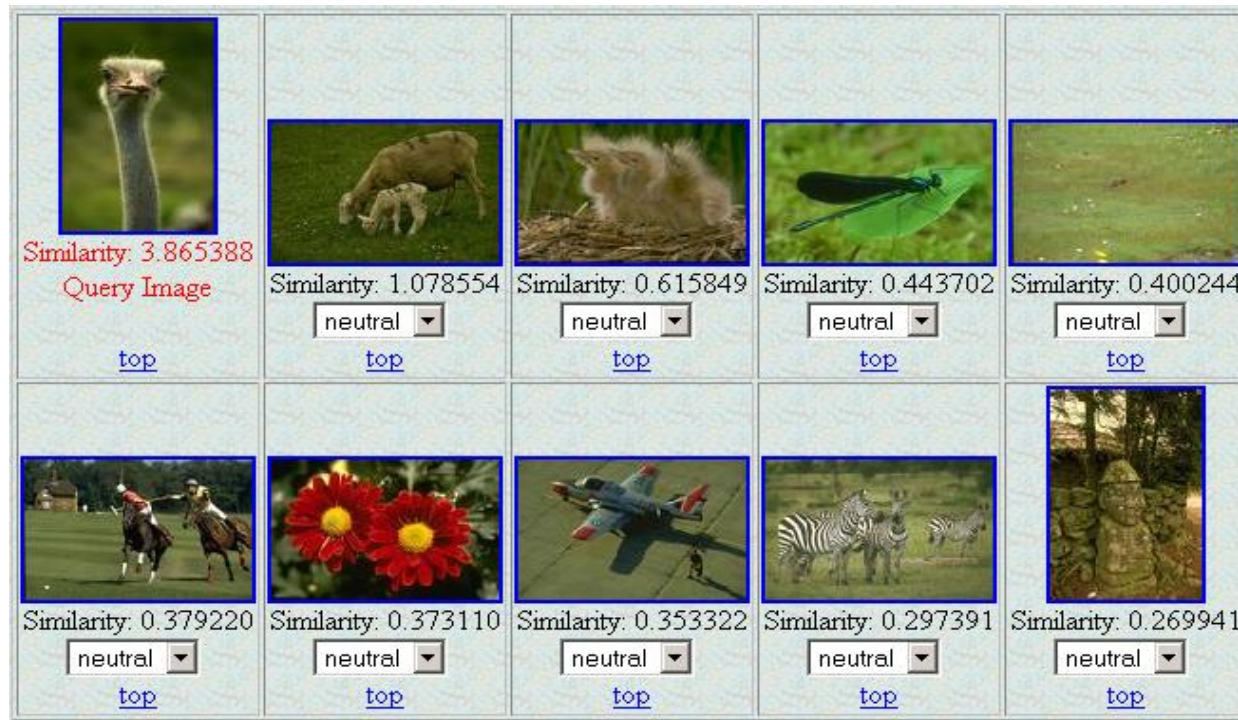


# DES EXEMPLES D'APPLICATION



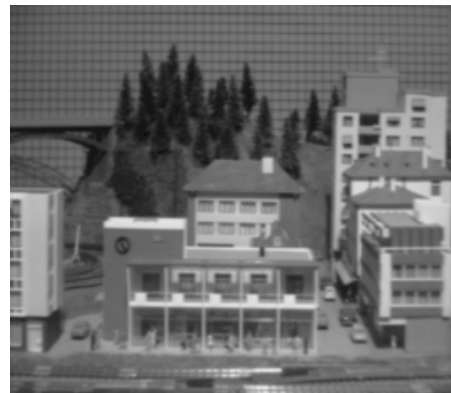
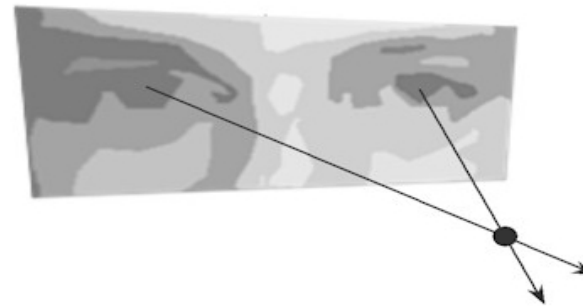
# DES EXEMPLES D'APPLICATION

- Bases de données
  - BDD image, BDD multimédia



# DES EXEMPLES D'APPLICATION

- Robotique
  - Vision stéréoscopique
  - Reconnaissance de scène



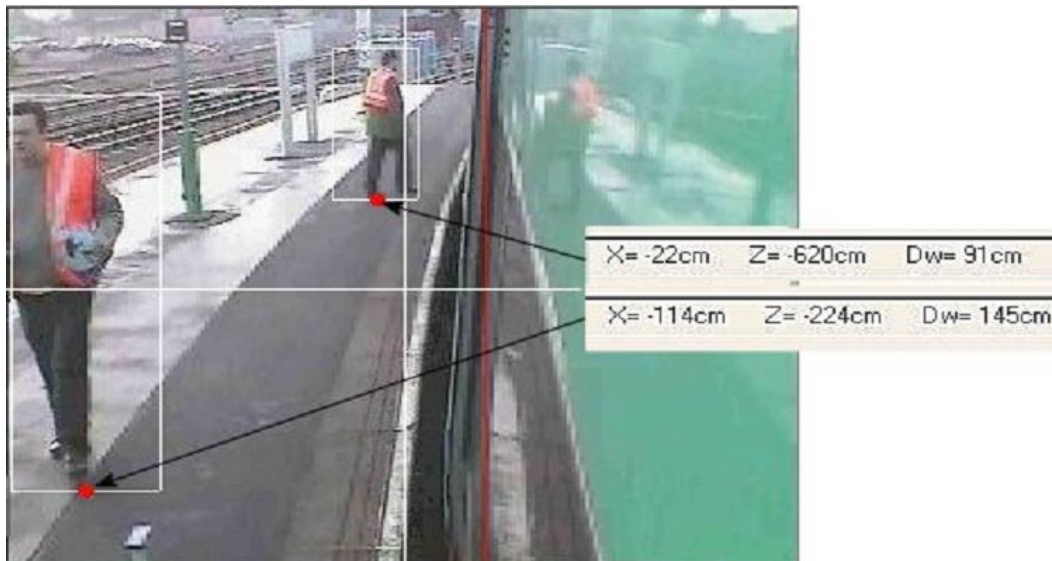
## DES EXEMPLES D'APPLICATION

- Véhicules intelligents
  - Aide au conducteur
  - Conduite automatique



# DES EXEMPLES D'APPLICATION

- Chemins de fers
  - Aide au conducteur : surveillance de quai



# DES EXEMPLES D'APPLICATION

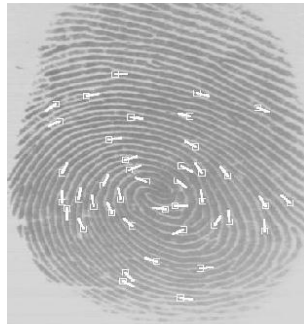
- Réalité augmentée
  - Analyse/Traitement + Synthèse



# DES EXEMPLES D'APPLICATION

## ○ Biométrie

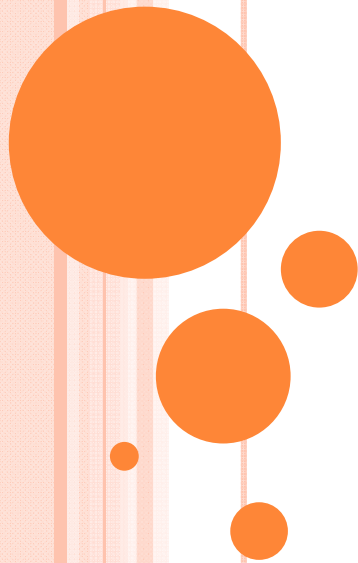
- Différentes modalités : iris, empreinte, visage, signature, ...



## DES EXEMPLES D'APPLICATION

- Et bien d'autres encore...
  - Vision industrielle
  - Aide aux handicapés
  - Nouveaux modes d'interaction
  - Représentation de données
  - Physique, biologie...

# CODAGE DES IMAGES



## RAPPELS – GÉNÉRALITÉS SUR LE CODAGE

- En informatique, toute information (texte, image, son...) est codée sous *forme binaire*, c'est à dire composée de 0 et de 1. L'unité d'information la plus petite est appelée un **bit** (« **binary digit** »),
- une série de 8 bits est appelée un **octet** (en anglais **byte**). **Un octet permet de stocker un caractère**, telle qu'une lettre, un chiffre... Ce regroupement de nombres par série de 8 permet une **lisibilité plus** grande, au même titre que l'on apprécie, en base décimale, de regrouper les nombres par trois pour pouvoir distinguer les milliers.
  - Par exemple, le nombre 1 256 245 est plus lisible que 1256245.

## *COMMENT L'INFORMATION EST-ELLE CODÉE EN BINAIRE ?*

- Pour ce qui est des nombres, cette opération s'effectue via une reconversion en base 2.
- Un entier naturel est un entier positif ou nul. Le nombre de bits à utiliser dépend de la fourchette des nombres que l'on désire utiliser.
- Avec un bit, il est possible d'obtenir 2 ( $= 2^1$ ) valeurs : 0 et 1
- Avec 2 bits, il est possible de représenter 4 ( $= 2^2$ ) valeurs différentes : 00, 01, 10 et 11
- Avec un octet (8 bits), il est possible de représenter 256 ( $= 2^8$ ) valeurs, soit des nombres entiers compris entre 0 et 255

## COMMENT L'INFORMATION EST-ELLE CODÉE EN BINAIRE?

- Pour un groupe de  $n$  bits, il est possible de représenter \_\_\_ valeurs, soit des nombres entiers compris entre 0 et \_\_\_\_

→ *Jusqu'à combien peut-on compter avec 4 bits ?  
Avec 24 bits ?*

## *COMMENT L'INFORMATION EST-ELLE CODÉE EN BINAIRE?*

- La base-2 fonctionne exactement de la même façon que la base-10, excepté bien entendu pour son unité de mesure. Par exemple, en base-10, « onze » s'écrit « 11 » soit « $10^1 + 10^0$ ».
- En base-2, « onze » s'écrit « 1011 » soit « $2^3 + 2^1 + 2^0$ » ( **$1*2^3 + 0*2^2 + 1*2^1 + 1*2^0$** )
- La valeur d'un octet est comprise entre 0 et 255.

# EXEMPLE

255 sera noté : \_\_\_\_\_

0 sera noté : \_\_\_\_\_

1 sera noté : \_\_\_\_\_

129 sera noté : \_\_\_\_\_

$2^7=128$	$2^6=64$	$2^5=32$	$2^4=16$	$2^3=8$	$2^2=4$	$2^1=2$	$2^0=1$
1	1	1	1	1	1	1	1
$2^7=128$	$2^6=64$	$2^5=32$	$2^4=16$	$2^3=8$	$2^2=4$	$2^1=2$	$2^0=1$
0	0	0	0	0	0	0	0
$2^7=128$	$2^6=64$	$2^5=32$	$2^4=16$	$2^3=8$	$2^2=4$	$2^1=2$	$2^0=1$
0	0	0	0	0	0	0	1
$2^7=128$	$2^6=64$	$2^5=32$	$2^4=16$	$2^3=8$	$2^2=4$	$2^1=2$	$2^0=1$
1	0	0	0	0	0	0	1

# EXERCICE

⇒ Donner la notation de 15 et de 160

15 : \_\_\_\_\_

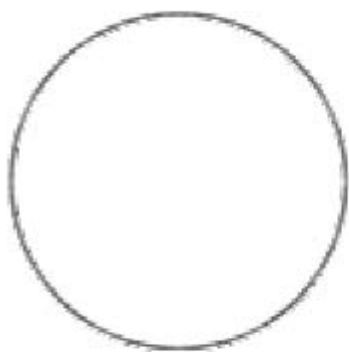
160 : \_\_\_\_\_

$2^7=128$	$2^6=64$	$2^5=32$	$2^4=16$	$2^3=8$	$2^2=4$	$2^1=2$	$2^0=1$

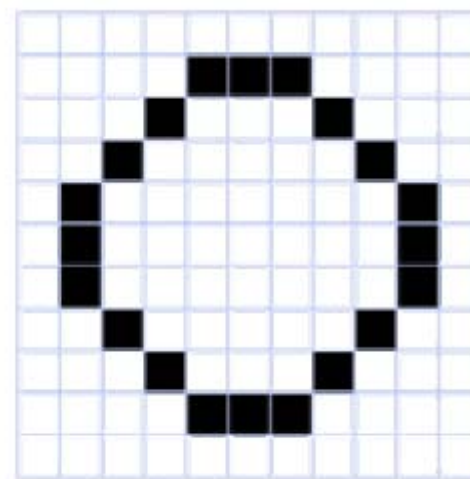
Comment écrit-on 01101101 en décimal ? \_\_\_\_\_

## CODAGE DES IMAGES

- Il existe actuellement une très grande diversité de formats d'images. Chaque logiciel ne peut traiter qu'un nombre limité de ces formats. On peut les classer en deux grandes catégories :
  - *Les images vectorisées*
  - *Les images codées par points (codage Bitmap ou matriciel)*



*image vectorielle*



*image bitmap*

## LES IMAGES VECTORISÉES

- Description géométrique de l'image.
- → suite de formules mathématiques décrivant les formes élémentaires constituant l'image (carré, cercle,...).
  - Chaque forme possède un certain nombre d'attributs tels que la couleur, l'épaisseur du trait,...
- Stockage plus économique que celui d'une image par points. La reconstitution de l'image par points au moment de l'affichage se fera sur base de ces formules.
- La taille du fichier varie en fonction de la complexité de l'image.
- Type de codage est particulièrement adapté pour les dessins techniques qui sont essentiellement composés de formes géométriques ou pour les schémas de toutes sortes.

# LES IMAGES CODÉES PAR POINTS

## ○ Notions élémentaires

- **Définition** : représente le nombre de points (pixels) constituant l'image. C'est-à-dire sa « dimension informatique »

- Une image de 640x480



8 Dpi



16 Dpi



72 Dpi

et 480 en hauteur  
soit 480, notée

- Les définitions de résolution (8...)

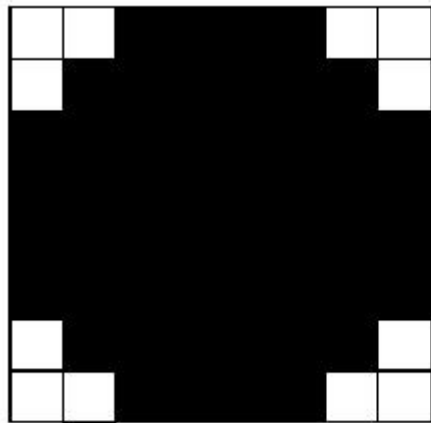
- **Résolution** : détermine le nombre de points par unité de surface, exprimé en *points par pouce* (**PPP**, en anglais **DPI** pour *Dots per Inch*)

- Un pouce représentant 2.54 cm
- Une résolution de 300 dpi signifie donc 300 colonnes et 300 rangées de pixels sur un pouce carré -> 90000 pixels sur un pouce carré

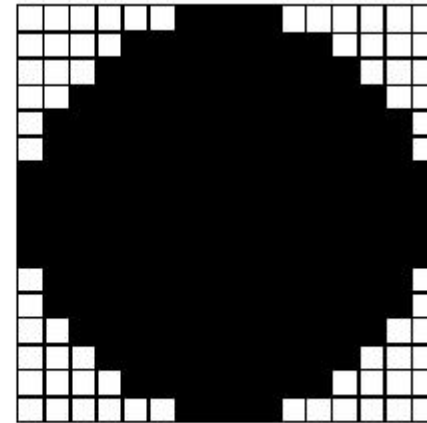
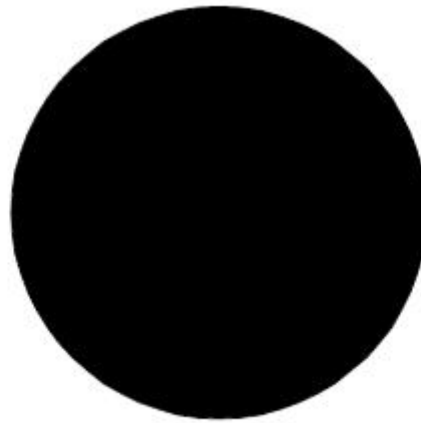
## LES IMAGES CODÉES PAR POINTS

- Les images utilisées dans les présentations multimédias sont des images destinées à être vues sur un écran.
  - → L'image sur écran est faite de pixels (picture element), c'est-à-dire de points auxquels correspondent des bits de mémoire.
  - → La finesse du tramage de l'image ou **résolution** (découpe de l'image en points) peut varier.

## LES IMAGES CODÉES PAR POINTS



Taille : 64 bits

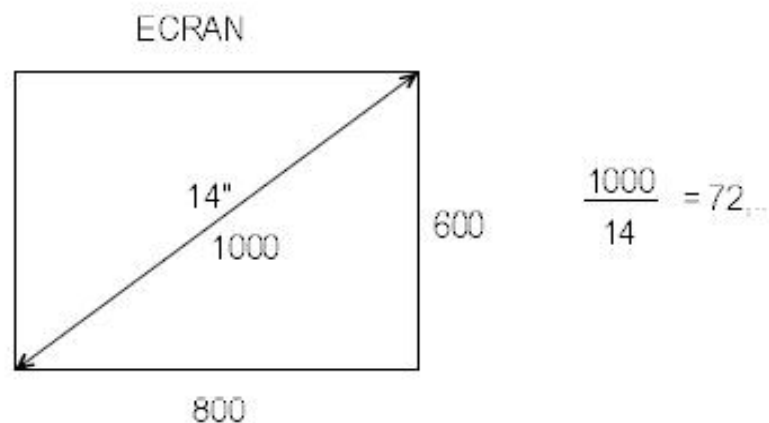


Taille : 256 bits

- Plus la trame est fine, plus la qualité de l'image restituée est bonne. La taille de l'image et donc du fichier qui la contient si elle est enregistrée est évidemment directement proportionnelle à la résolution.
- Si une image que l'on veut capturer occupe l'entièreté de l'écran dont la résolution est de 800x 600, elle est formée de 480.000 points ou pixels.

## LES IMAGES CODÉES PAR POINTS

- Affichage → **pixels**
- Nous parlons plutôt de **dpi** (*dot per inch*) ou **ppp** (*points par pouce*) lorsqu'il s'agit d'imprimer ou de capturer une image.
- → La qualité de l'impression sera fonction de la densité des points imprimés ou des signaux lumineux captés.



# LES IMAGES CODÉES PAR POINTS

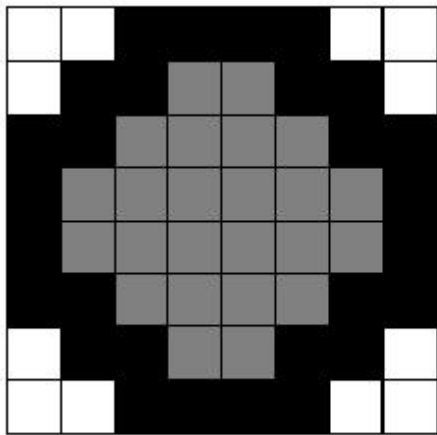
- Lors d'une capture, il faut choisir la résolution la plus adaptée.
  - Visualisation sur écran ou impression sur papier?
  - 1<sup>er</sup> cas : résolution plus basse que le 2<sup>ème</sup>
  - **Ne pas augmenter le nombre lumineux d'un écran au-delà de certaines limites**

## LES IMAGES CODÉES PAR POINTS

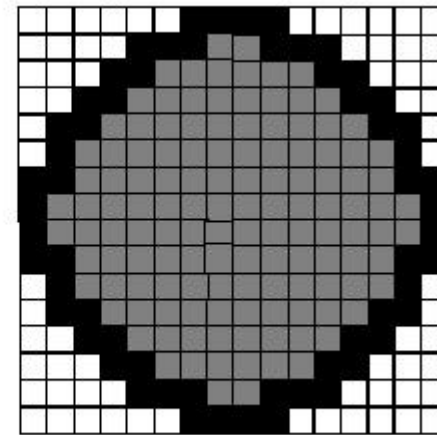
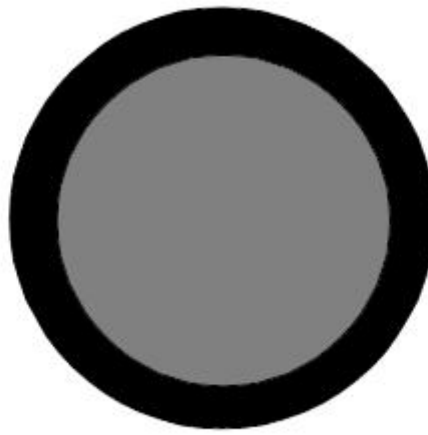
- Le nombre de couleurs à coder est un facteur déterminant quand à la **qualité** de l'image restituée mais également quant à la **taille** du fichier.
- Suivant le nombre de couleurs, le codage de chaque pixel se fait sur
  - un bit (noir et blanc)
  - 2 bits (quatre nuances de gris)
  - 4 bits (16 couleurs)
  - 8 bits (256 couleurs)
  - 16 bits (65536 couleurs)
  - True color :
    - 24 bits (plus de 16 millions de couleurs)
    - Rajouter une information de transparence : 32 bits

# LES IMAGES CODÉES PAR POINTS

- La taille du fichier correspondant est forcément chaque fois en proportion.



Taille : 128 bits



Taille : 512 bits

# LES IMAGES CODÉES PAR POINTS

- Taille de l'image en fonction du nombre de couleurs choisies

Nombre de couleurs	Codage sur	Taille
noir et blanc	1 bit	480 000 bits ou 60 Ko
4 nuances de gris	2 bits	960 000 bits ou 120 Ko
16 couleurs	4 bits	1 920 000 bits ou 240 Ko
256 couleurs	8 bits	480 Ko
65 536 couleurs	16 bits	960 Ko
plus de 16 millions de couleurs	24 bits	1,92 Mo

# IMAGES CODÉES PAR POINTS

## ○ Exemple :

- Image de dimension 640x480 codée en 256 couleurs
- →  $640 * 480 * 1$  (256 couleurs correspondent à 1 octet)
- = 307200 Octets  $\leftrightarrow$  300 Ko

# IMAGES CODÉES PAR POINTS

## ○ Formats de fichiers

<b>extension</b>	<b>nombre de couleurs</b>	<b>Compression</b>	<b>Commentaires</b>
BMP	16 M	Non	format standard Windows
JPG	16M	Oui	format courant sur Internet
Gif	256	Oui	permet les animations (Gif animés) ainsi que le mode transparence. Très utilisé sur Internet modèle de compression déposé ©
ICO	16 ou 256	Non	format des icônes sous Windows
Tiff	16M	Oui	(Tagged image format) utilisé en gestion de document, supporte différents formats.
PCX	16M	Non	ancien format (Paintbrush)
PNG	16M	Oui	concurrent libre du Gif
TGA	16M	Oui / Non	
...			

# IMAGES CODÉES PAR POINTS

- Quelques formats de fichiers
  - **GIF** (Graphic Interchange Format) (Format d'échange graphique)
    - Format propriétaire (CompuServe).
    - GIF stocke les images comme une séquence de pixel en valeur de couleur RGB.
    - Chaque image est précédée
      - d'une signature (no de version, etc.),
      - une définition d'écran (permet l'ajustage à l'écran de visualisation)
      - et une échelle de couleur.
    - Les données sont compressées avec l'algorithme LZW, ce qui engendre une perte de données lors de l'agrandissement.
    - Facilité d'emploi + largement diffusé

# IMAGES CODÉES PAR POINTS

## ○ GIF

- Chaque image possède une palette de 256 couleurs maximum, chaque couleur de la palette est codée sur 3 octets (RVB), elle est donc choisie dans 16 millions de nuances.
- Chaque point de l'image est ensuite codé par le n° d'ordre de la couleur dans la palette (index).

# IMAGES CODÉES PAR POINTS

- **PNG** Portable Network Graphics (Graphiques transmissibles en réseau)
  - Projet Norme internationale (W3C)
  - En plus de fonctionnalités de GIF, PNG devrait supporter:
    - des images en couleur vraie de plus de 48 bits/pixel
    - des images en grisé de plus de 16 bits/pixel
    - un canal de texte (masque transparent général)
    - des infos sur le gamma de l'image (=affichage inaltérable)
    - une détection de la corruption de fichier
    - un affichage progressif rapide

## IMAGES CODÉES PAR POINTS

- **TIFF** Tag Image File Format (format de fichier d'image “ étiquetées ”)
  - Format propriétaire (Adobe).
  - définit une séries de description de fichiers permettant la reconnaissance de la plupart des formats de données en deux dimensions.
  - TIFF défini des types de classe selon les données numérisées.
  - Des “ marqueurs ” (tags) privés peuvent être ajoutés pour définir des types d'image particuliers.

# IMAGES CODÉES PAR POINTS

- **JPEG** Joint Photographic Expert Group standard
  - Norme internationale
  - Format de compression variable permettant la compression sans perte ou avec perte d'information
  - Gains de place mémoire et de vitesse d'affichage (internet) pour les images fortement compressées.

## IMAGES CODÉES PAR POINTS

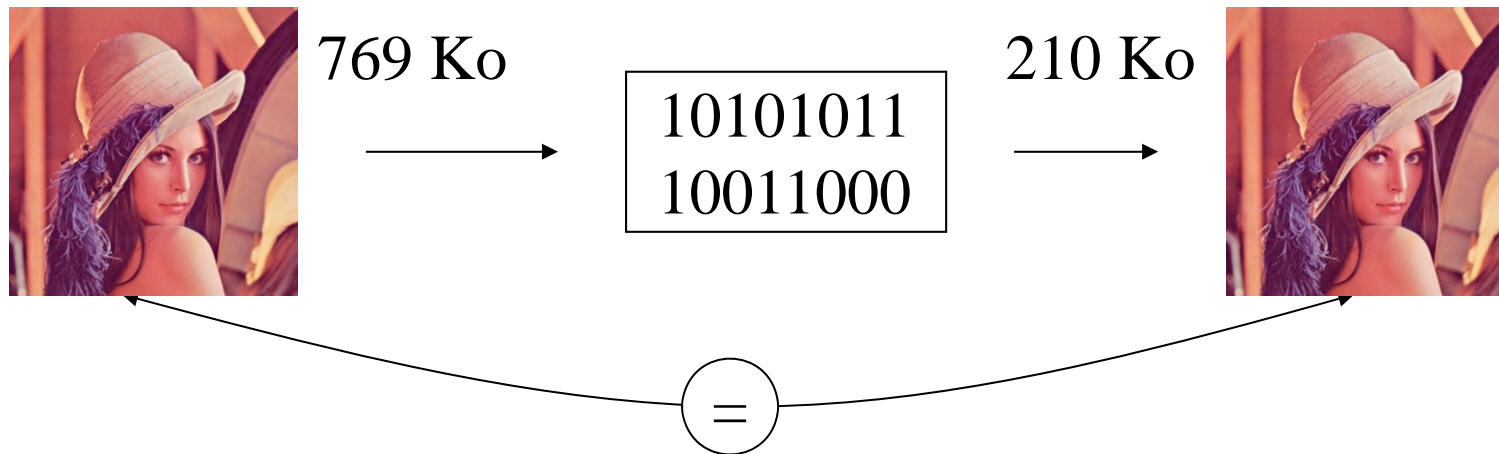
- Parmi toutes les extensions de format d'images codées par point selon ce principe, la plus connue est *BMP*.
- → *Nécessité de recourir à la compression*

# COMPRESSION

- Un exemple : Image couleur 512\*512 pixels
  - $512 * 512 * 3 * 1 \text{ octet} = 768 \text{ Ko}$
- Autre exemple : Film couleur 320\*200 d'1h
  - $320 * 200 * 3600 * 25 * 3 * 1 \text{ octet} = 16 \text{ Go}$
- Intéressant pour les images
- Obligatoire pour les séquences vidéo

# COMPRESSION SANS PERTE

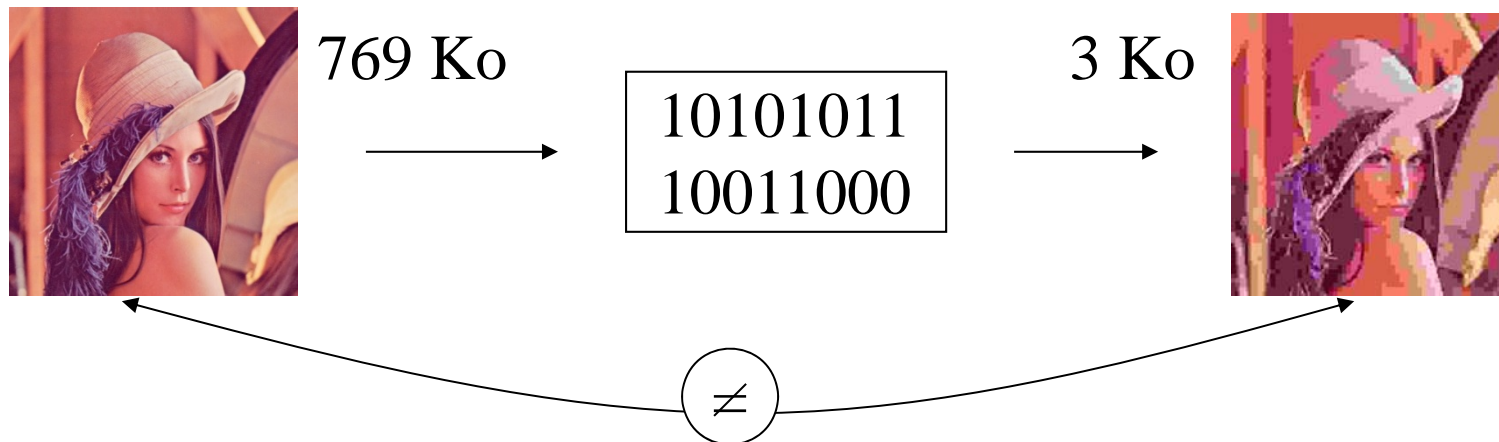
- On peut récupérer entièrement les données initiales



- Adapté aux images dont les détails sont très importants (imagerie médicale)

# COMPRESSION AVEC PERTE

- Les données originales ne peuvent être complètement restaurées



- Adapté aux autres images et aux séquences vidéo

# STANDARDS DE COMPRESSION

- Compression d'image
  - sans perte : TIFF, GIF...
  - avec perte : JPEG, *JPEG 2000* ...
- Compression de vidéo
  - MPEG-1, MPEG-2, *MPEG-4 (DivX)*

## TECHNIQUES DE COMPRESSION

- Suppression des répétitions : **RLE** (Run Length Encoding)
  - n caractères successifs c sont remplacés par c suivi d'un caractère spécial et du nombre d'occurrence
  - N A M E : x x x x x x x x x x x x x
  - N A M E : x R 13
- **Statistiques** : même parties qui reviennent souvent
  - on utilise un dictionnaire
  - on recode les ensembles en fonction de leur fréquence d'occurrence : les plus grandes sont recodées par des codes plus petits

# TECHNIQUES DE COMPRESSION

## ○ Huffman :

- généralisation pour les images des techniques "statistiques"
- On calcule les fréquences d'apparition de chaque octet.
- Même système de dictionnaire et de recodage
- utilisé pour des ensembles d'images ou des images animées

## ○ Transformations :

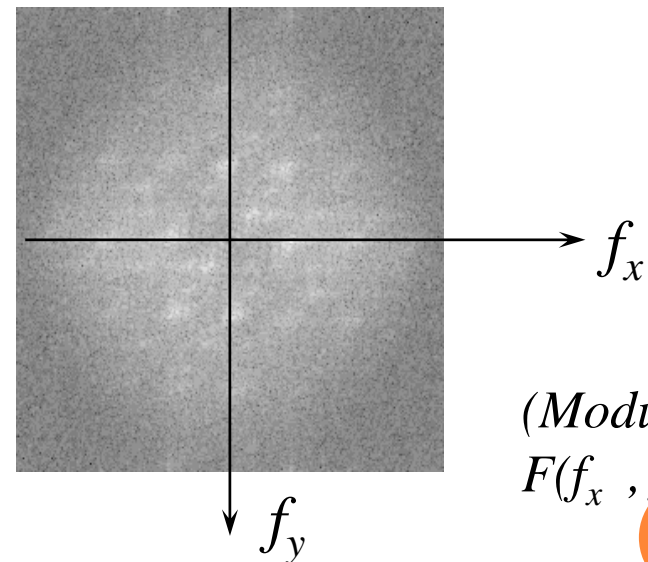
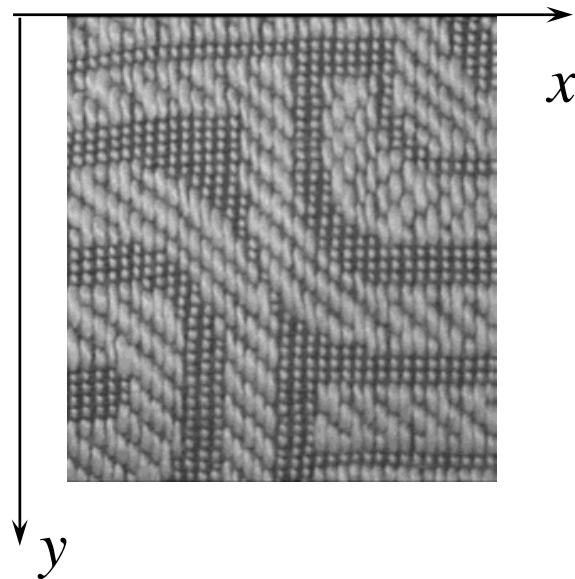
- On passe du domaine temporel ou spatial au domaine fréquentiel
- On élimine les coefficients les moins significatifs
- Fourier, Cosinus,...

# Transformée de **Fourier** 2D

- Image =  $\Sigma$  images sinusoïdales (A,f, $\phi$ )

$$F(f_x, f_y) = \iint f(x, y) \exp(-2j\pi(f_x x + f_y y)) dx dy$$

- $F$  = image complexe (module & phase)



(Module de  
 $F(f_x, f_y)$ )

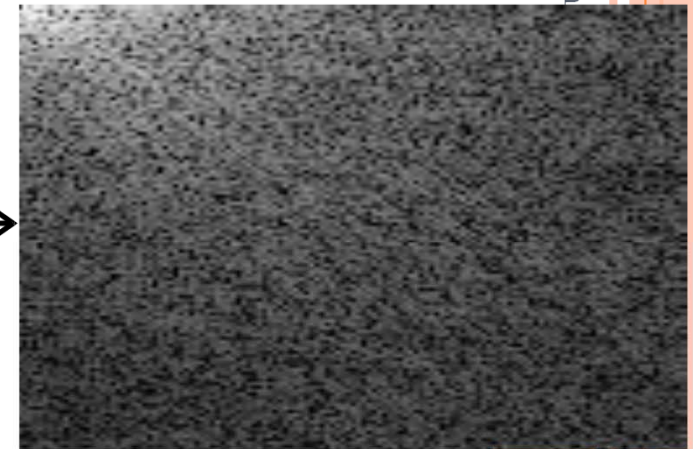
# Transformée **Cosinus** Discrète

$$C(u, v) = \frac{4.c(u).c(v)}{M.N} \cdot \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} f(i, j) \cdot \cos\left(\frac{(2i+1)u.\pi}{2N}\right) \cdot \cos\left(\frac{(2j+1)v.\pi}{2M}\right)$$

$$\text{Avec} \begin{cases} c(u) = \sqrt{2/N} & \text{pour } u \neq 0 \\ c(u) = \sqrt{1/N} & \text{pour } u = 0 \end{cases}$$



Image originale



TCD de l'image

# TECHNIQUES DE COMPRESSION

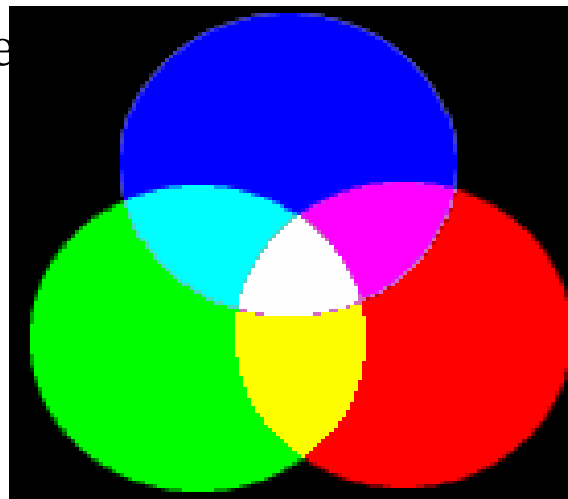
## ○ LZW : *Lempel-Ziv-Welch*

- Algorithme de compression des données sans perte. Il s'agit d'une amélioration des algorithmes LZ77 (1977) et LZ78 (1978), tous les deux écrits par Abraham Lempel et Jacob Ziv. LZW fut créé en 1984 par Terry Welch, d'où son nom.
- Breveté par Unisys
- L'algorithme a été conçu de manière à être rapide à implémenter, mais n'est la plupart du temps pas optimal car il effectue une analyse limitée des données à compresser.
- Utilisation d'une table de traduction

# CODAGE DES COULEURS

## ○ RGB

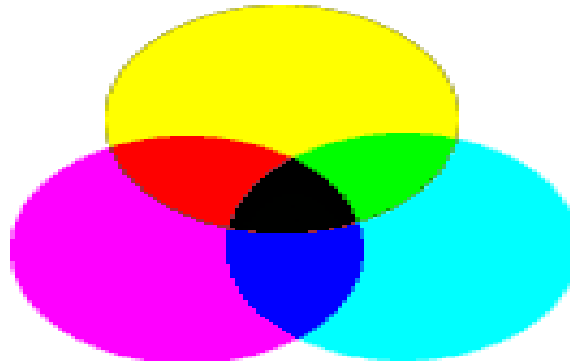
- Sur les ordinateurs et en télévision on utilisera le plus souvent un codage RVB (*Rouge, Vert, Bleu*).
- *Ces trois* couleurs primaires permettent par synthèse additive la recombinaison de toutes les couleurs visibles.
- L'affichage sur un écran d'ordinateur est par principe additif, l'utilisation des 3 couleurs au maximum de luminosité donne



# CODAGE DES COULEURS

## ○ CMJN

- En imprimerie, la synthèse des couleurs étant soustractive, on utilise les primaires de la peinture (*Cyan, Magenta,Jaune*) auxquels on ajoute le Noir car *l'utilisation des 3 primaires ne donne pas en général un noir satisfaisant.*
- On travaille alors en quadrichromie. Le codage de base est le plus souvent en 32 bits (4x8).



## CODAGE DES COULEURS

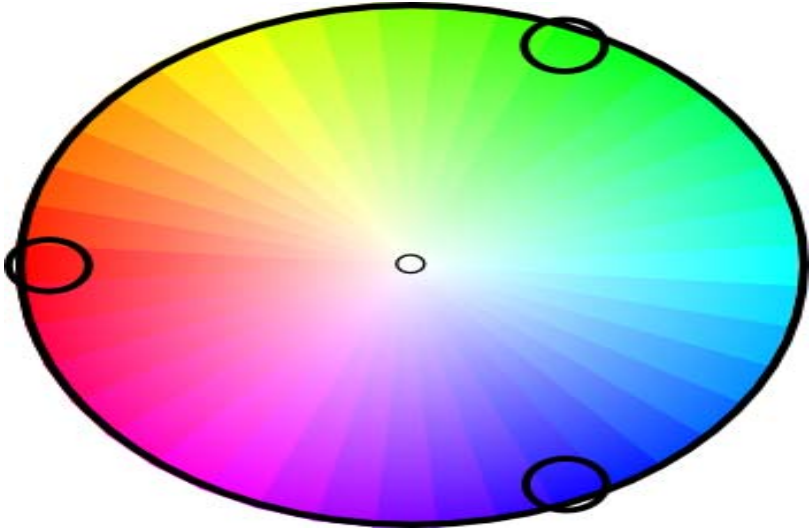
- Peu de logiciels permettent le codage des couleurs directement en CMJN.
  - Corel Draw
- La conversion RVB vers CMJN peut réserver des surprises car certaines couleurs RVB ne pourront être correctement imprimées.
- → Pour éviter ces désagréments, il est conseillé d'utiliser des nuanciers de type « PANTONE » pour la création des images à imprimer.
  - Pour des impressions de luxe, on utilise parfois un séparateur Hexachromatique (avec les couleurs du nuancier Pantone).

# CODAGE DES COULEURS

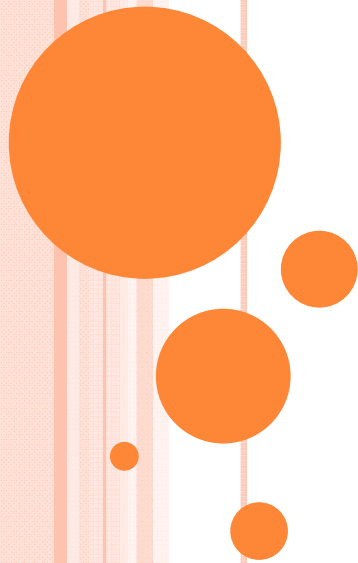


**PANTONE Matching System — EXAMPLE**

PMS 111	PMS 110	PMS 109	PMS 108	PMS 107
PMS 174	PMS 175	PMS 172	PMS 171	PMS 170
PMS 201	PMS 200	PMS 199	PMS 198	PMS 197
PMS 349	PMS 348	PMS 347	PMS 346	PMS 345
PMS 296	PMS 295	PMS 294	PMS 293	PMS 292
PMS 527	PMS 528	PMS 529	PMS 520	PMS 521



# TRAITEMENT D'IMAGES

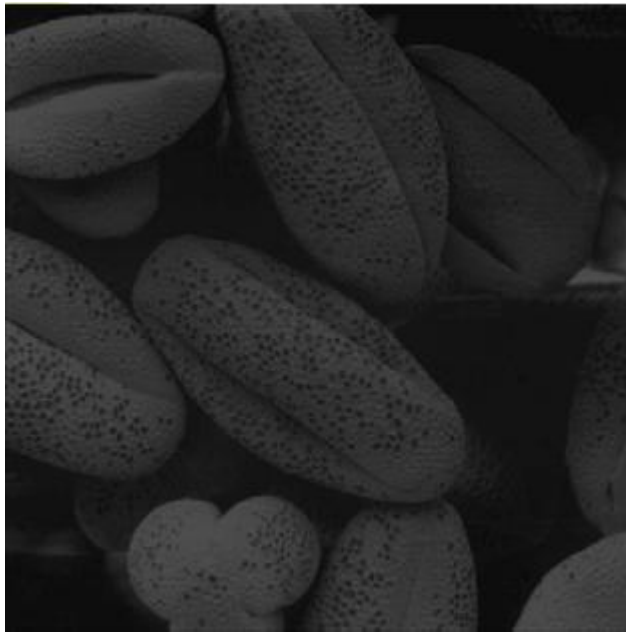


## POURQUOI TRAITE-T-ON DES IMAGES?

- Moins cher
- Quand il n'y a pas d'autres choix
- Quand on veut utiliser pleinement les données
- Grande-profondeur, capteurs de haute résolution,
- Images multi-spectrales,
- Rendering de données 3D,
- Et plus encore.

## POURQUOI TRAITE-T-ON DES IMAGES?

- On traite des images pour avoir des images **subjectivement** plus « jolies ».



- Quelle image est meilleure ?

# POURQUOI TRAITE-T-ON DES IMAGES?

- Before / After



# POURQUOI TRAITE-T-ON DES IMAGES?



# POURQUOI TRAITE-T-ON DES IMAGES?



# POURQUOI TRAITE-T-ON DES IMAGES?



# POURQUOI TRAITE-T-ON DES IMAGES?



# POURQUOI TRAITER DES IMAGES PAR ORDINATEUR?

## ○ Humain

- Identification
- Reconnaissance
- Voir et décrire les relations
- Interprétation expérimentée

## ○ Ordinateur

- Mesure des valeurs absolues
- Calculs
- Infatigable
- Pas cher
- Objectif

# TRAITER UNE IMAGE, EN QUOI ÇA CONSISTE ?

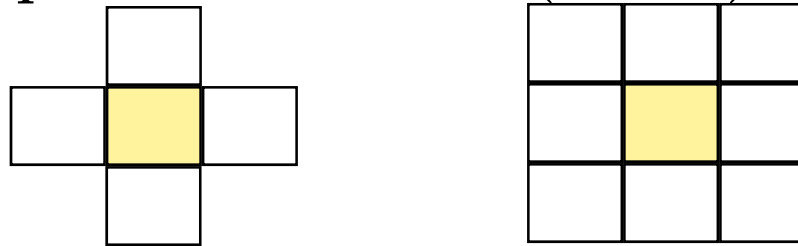
- Coder
- Améliorer
- Simplifier
- Analyser
- Interpréter
- ...

# CARACTÉRISTIQUES USUELLES

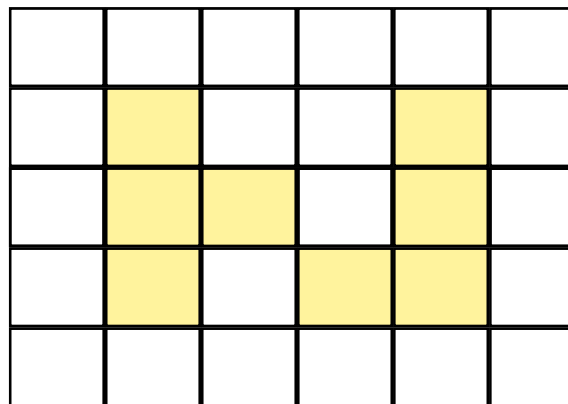
- Des pixels, des valeurs, qu'en fait-on ?
- Notions globales
  - ex : l'histogramme
- Notions locales
  - ex : le voisinage

# NOTIONS DE VOISINAGE, CONNEXITÉ ET DISTANCE

- Beaucoup de traitements font intervenir la notion de **voisinage**
- Un pixel possède plusieurs voisins (4 ou 8)



- On parlera de **connexité** 4 ou 8



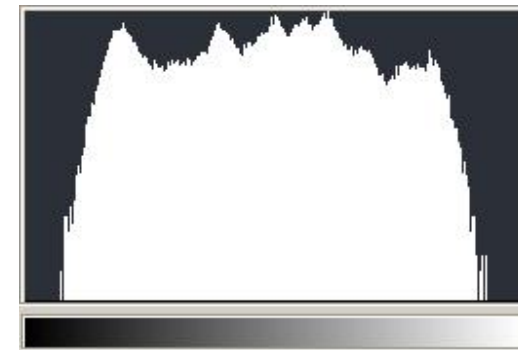
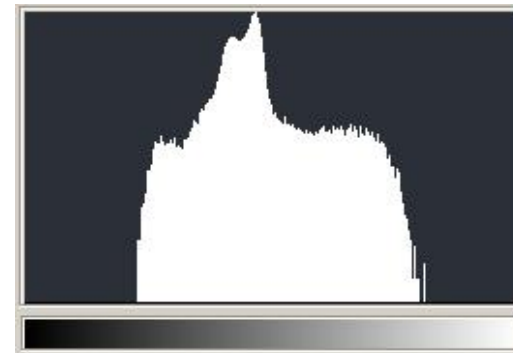
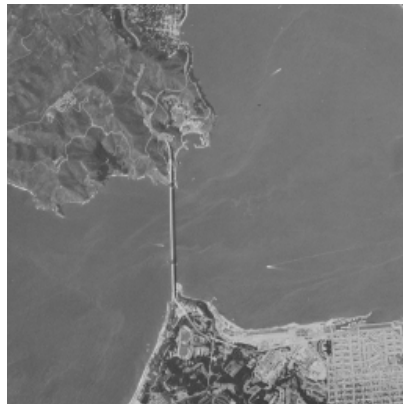
La région jaune forme :  
UN seul objet en connexité  
8  
DEUX objets en connexité  
4

## Distance entre deux pixels $f [i,j]$ et $f '[k,l]$

- Distance Euclidienne  $d_e(f, f') = \sqrt{(i-k)^2 \Delta x^2 + (j-l)^2 \Delta y^2}$
- Distance *City-Block*  $d_c(f, f') = |i-k|\Delta x + |j-l|\Delta y$   
longueur du chemin en connexité 4
- Distance de l'échiquier  $d_b(f, f') = \max(|i-k|\Delta x, |j-l|\Delta y)$

# HISTOGRAMME

- Permet de mesurer la répartition globale des valeurs des pixels de l'image (probabilités)



# BON OU MAUVAIS TRAITEMENT ?

- Besoin de comparer des méthodes...  
... mais pas toujours facile !
- Il faut tenir compte de :
  - qualité
  - efficacité
  - adaptabilité
  - facilité

# QUALITÉ D'UNE SOLUTION

- Comment quantifier la qualité ?
- Mesures de qualité
  - rappel
  - précision
- Comparaison avec résultat idéal
  - obtenu à la main
  - obtenu automatiquement
- Corpus de test

# EFFICACITÉ D'UNE SOLUTION

- Temps de calcul
- Estimable a priori
- Peut dépendre d'un matériel spécifique
  - processeur embarqué (DSP...)
  - architecture parallèle

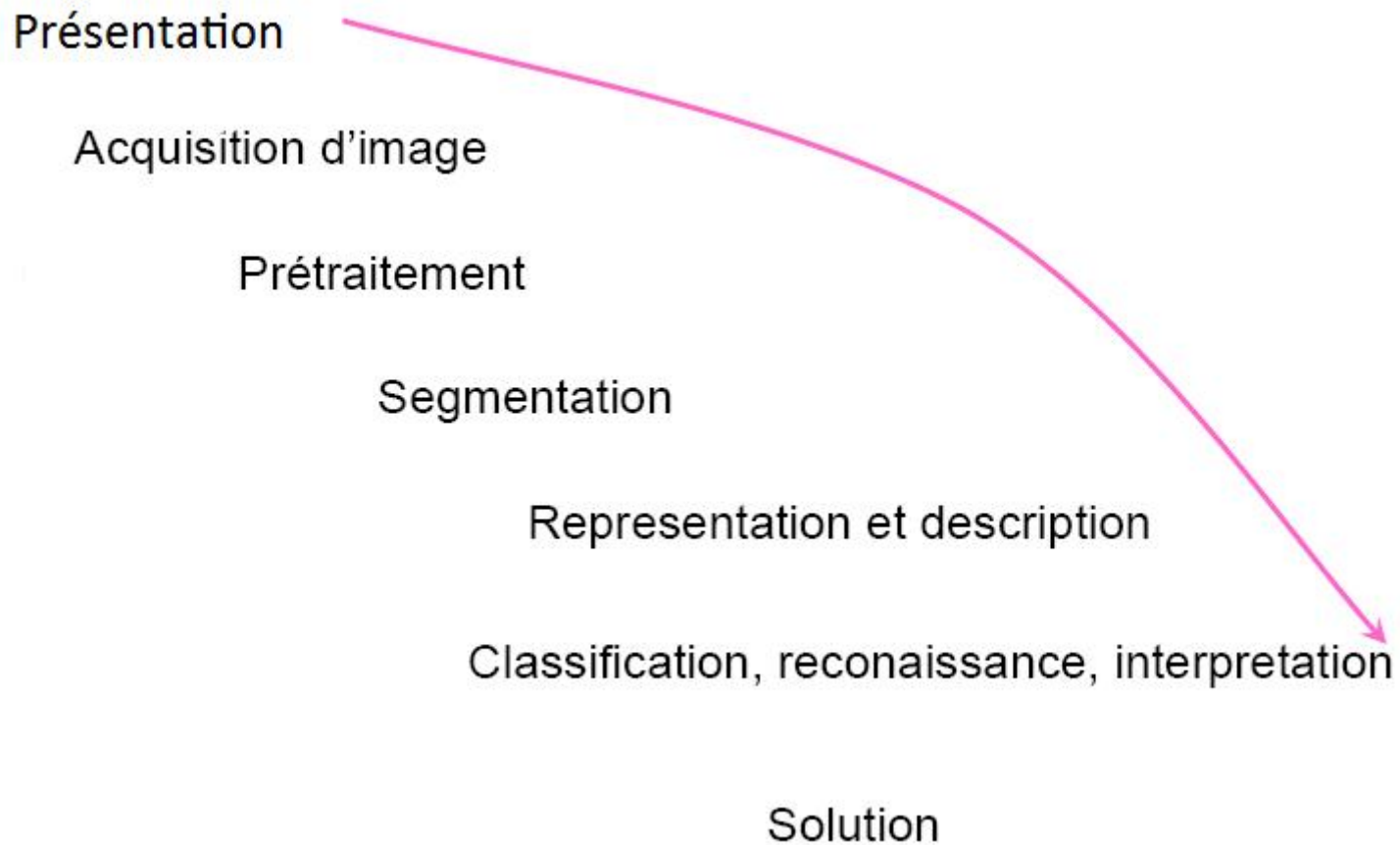
# ADAPTABILITÉ D'UNE SOLUTION

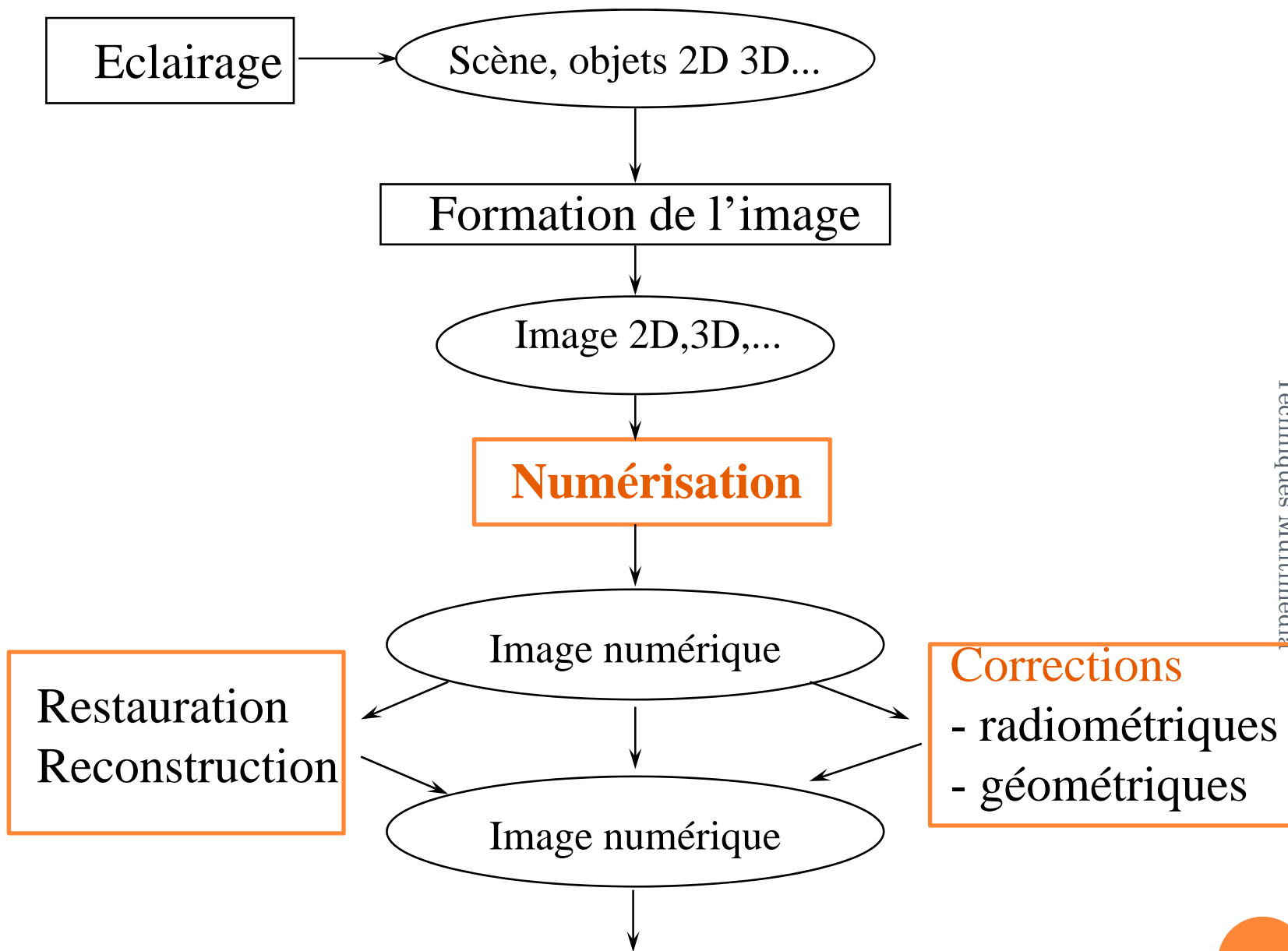
- La situation :
  - Méthodes spécifiques à un contexte
- L'idéal :
  - Méthodes génériques
- L'objectif à atteindre :
  - Réutilisation !

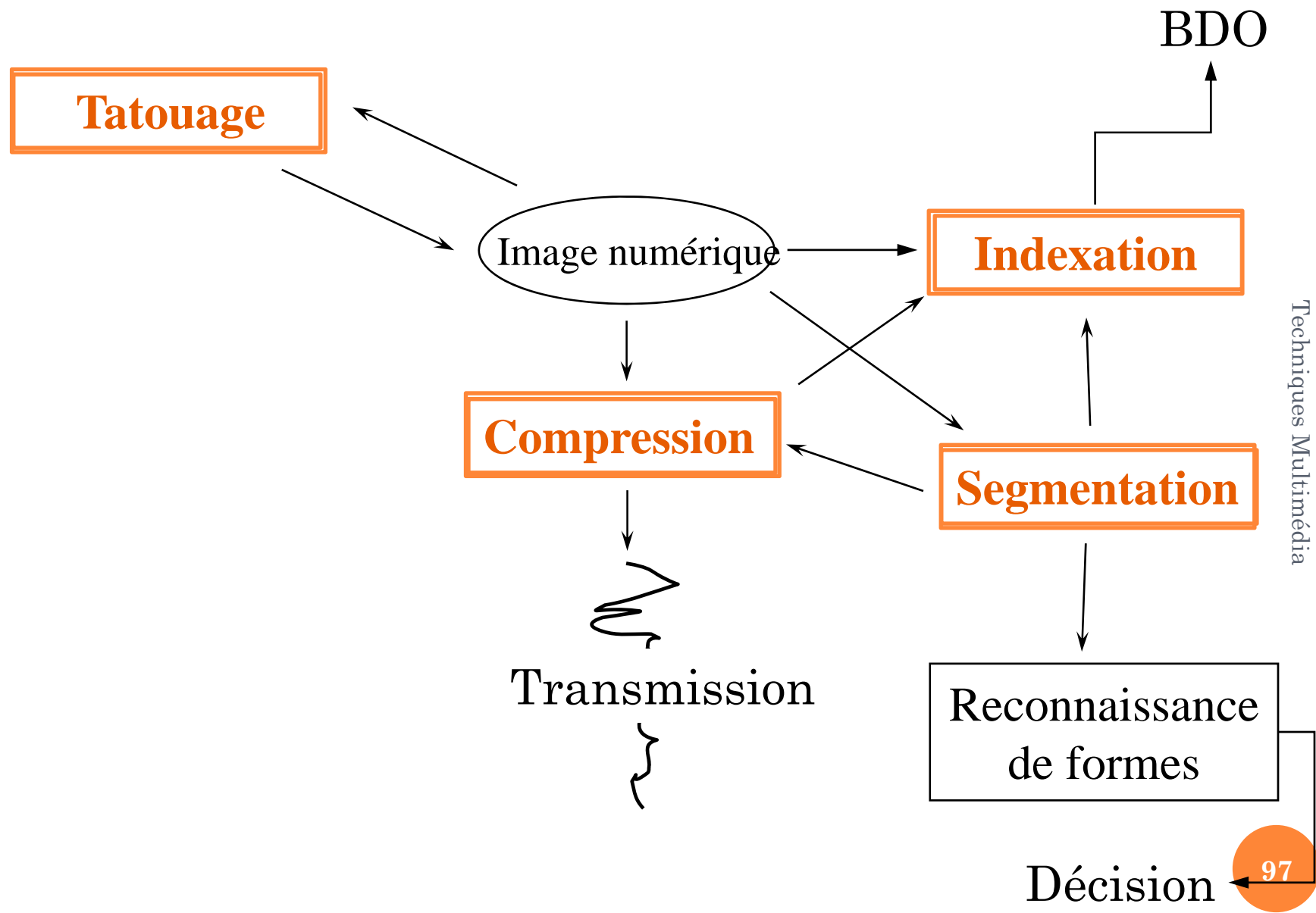
# FACILITÉ D'UNE SOLUTION

- Pour bien fonctionner, chaque méthode a besoin d'être configurée
- Evaluation de la facilité :
  - Nombre de paramètres
  - Facilité de réglage

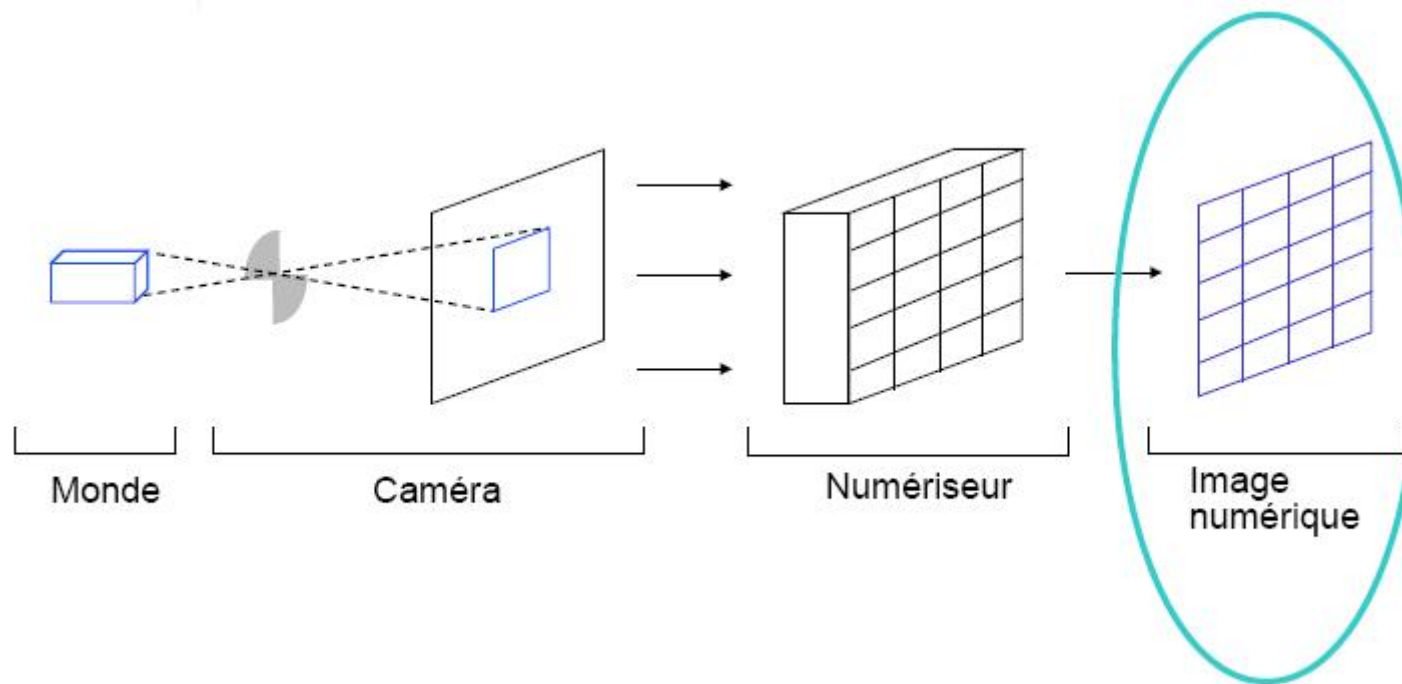
# ETAPES DE TRAITEMENT D'IMAGES





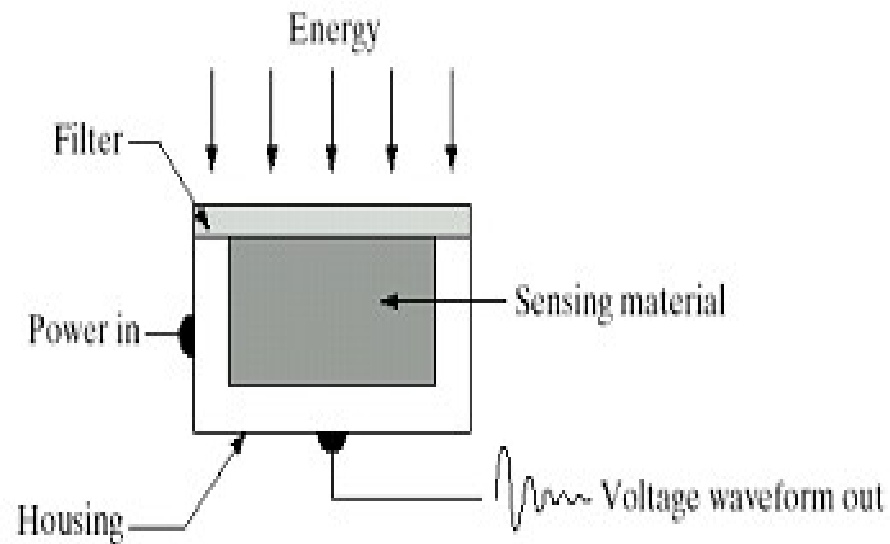


# ACQUISITION D'UNE IMAGE

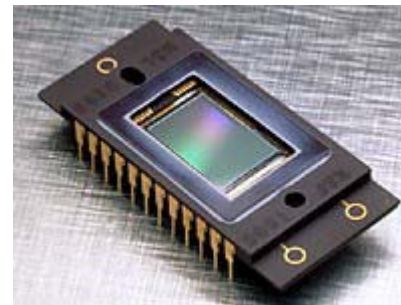
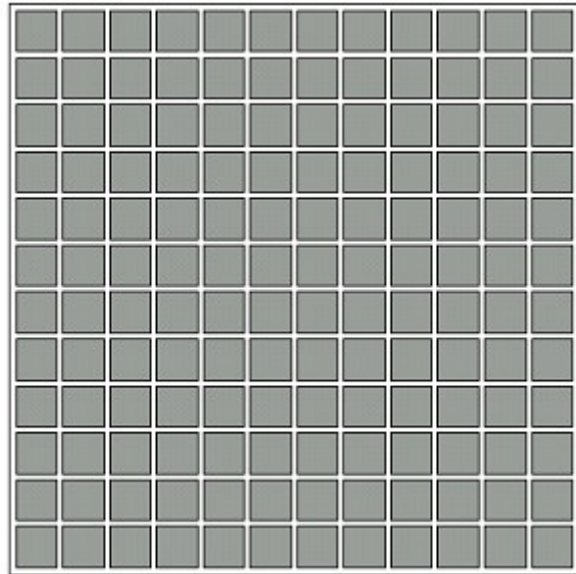


# CAPTEUR : PRINCIPE GÉNÉRAL

- Principe général (ex: photodiode)
  - L'énergie incidente est convertie en signal électrique
  - Sortie est proportionnelle à la lumière
  - Filtre pour augmenter la sélectivité



# CAPTEURS – MATRICE 2D

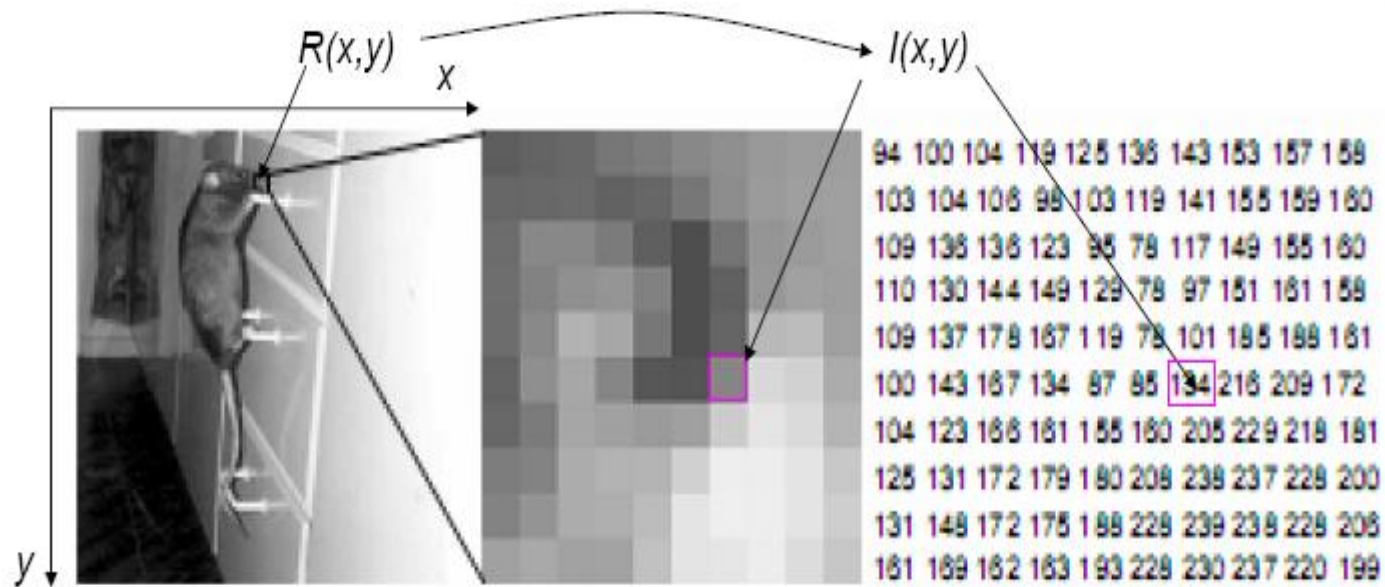


*KAF-1600* - Kodak

## CAPTEUR CCD

- Caméras numériques CCD
- Matrice CCD (Charged Coupled Devices)
- Système d'acquisition numérique 2D le plus utilisé
- La réponse est proportionnelle à l'intégrale de l'énergie lumineuse qui atteint chaque élément
- Pour la couleur, on utilise trois capteurs par pixel réagissant à des longueurs d'ondes différentes (rouge, vert et bleu)
- L'image d'entrée ne devra pas contenir trop de hautes fréquences ( *Ne passez pas à la télé avec un costume rayé !* )

# RAPPEL : QU'EST-CE QU'UNE IMAGE NUMÉRIQUE?

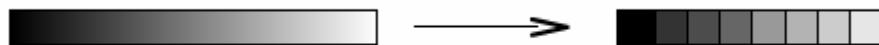


## IMAGE NUMÉRIQUE

- Les valeurs de  $f(x,y)$  sont la réponse du capteur au phénomène observé
- Les valeurs de  $f(x,y)$  sont des valeurs de « voltage » continu
- Les valeurs de  $f(x,y)$  doivent être converties vers le domaine numérique
  - Conversion Analogique/Numérique (A/N)
- Deux procédés sont impliqués pour numériser une image :
  - **Numérisation** = **Échantillonnage** + **Quantification**

## RAPPEL : QU'EST-CE QU'UNE IMAGE NUMÉRIQUE ?

- L'**échantillonnage** est le procédé de *discrétisation spatiale d'une image consistant à associer à chaque zone rectangulaire  $R(x,y)$  d'une image continue une unique valeur  $I(x,y)$ .*
- On parle de **sous-échantillonnage** lorsque *l'image est déjà discrétisée et qu'on diminue le nombre d'échantillons.*
- La **quantification** désigne *la limitation du nombre de valeurs différentes que peut prendre  $I(x,y)$ .*



- Une **image digitale** est une image *échantillonnée et quantifiée.*

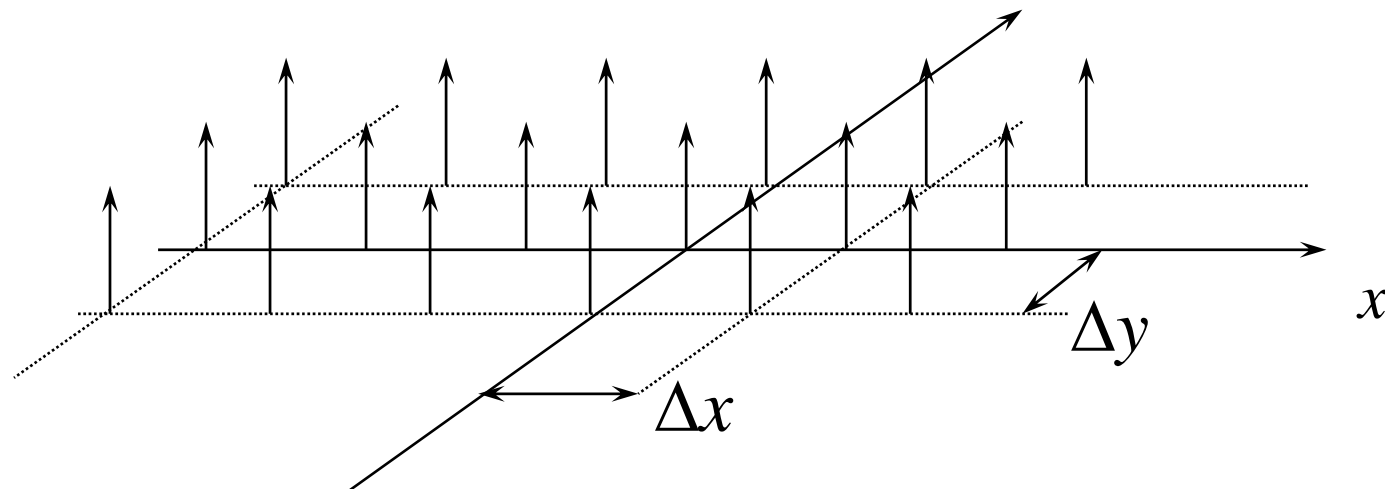
# REPRÉSENTATION ÉCHANTILLONNÉE

- Echantillonnage d'une fonction  $f(x,y)$

$$f_e(x,y) = f(x,y) \cdot \sum_i \sum_j \delta(x - i \Delta x, y - j \Delta y)$$

$\Delta x$  pas d'échantillonnage dans la direction  $x$

$\Delta y$  pas d'échantillonnage dans la direction  $y$



$$\sum_i \sum_j \delta(x - i \Delta x, y - j \Delta y) \text{ Peigne de Dirac 2D}$$

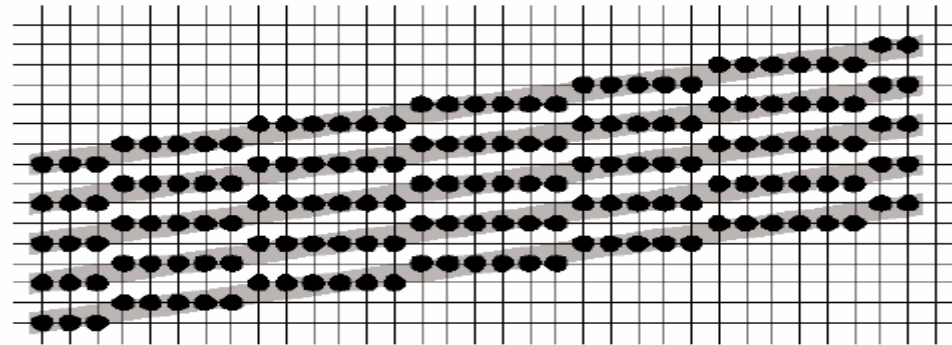
# THÉORÈME DE SHANNON

→ La fréquence d'échantillonnage d'un signal doit être égale ou supérieure au double de la fréquence maximale contenue dans ce signal, afin de convertir ce signal d'une forme analogique à une forme numérique. Ce théorème est à la base de la conversion numérique des signaux.

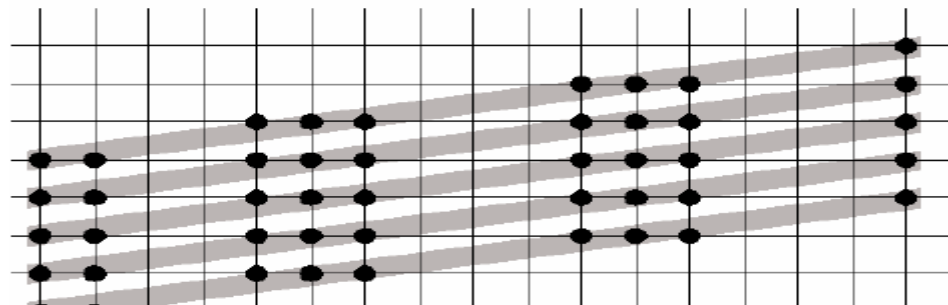
→ La meilleure illustration de l'application de ce théorème est la détermination de la fréquence d'échantillonnage d'un CD audio, qui est de 44,1 kHz. En effet, l'oreille humaine peut capter les sons jusqu'à 16 kHz, quelquefois jusqu'à 20 kHz. Il convient donc, lors de la conversion, d'échantillonner le signal audio à au moins 40 kHz. 44,1 kHz est la valeur normalisée par l'industrie

# ÉCHANTILLONNAGE ET INFORMATION

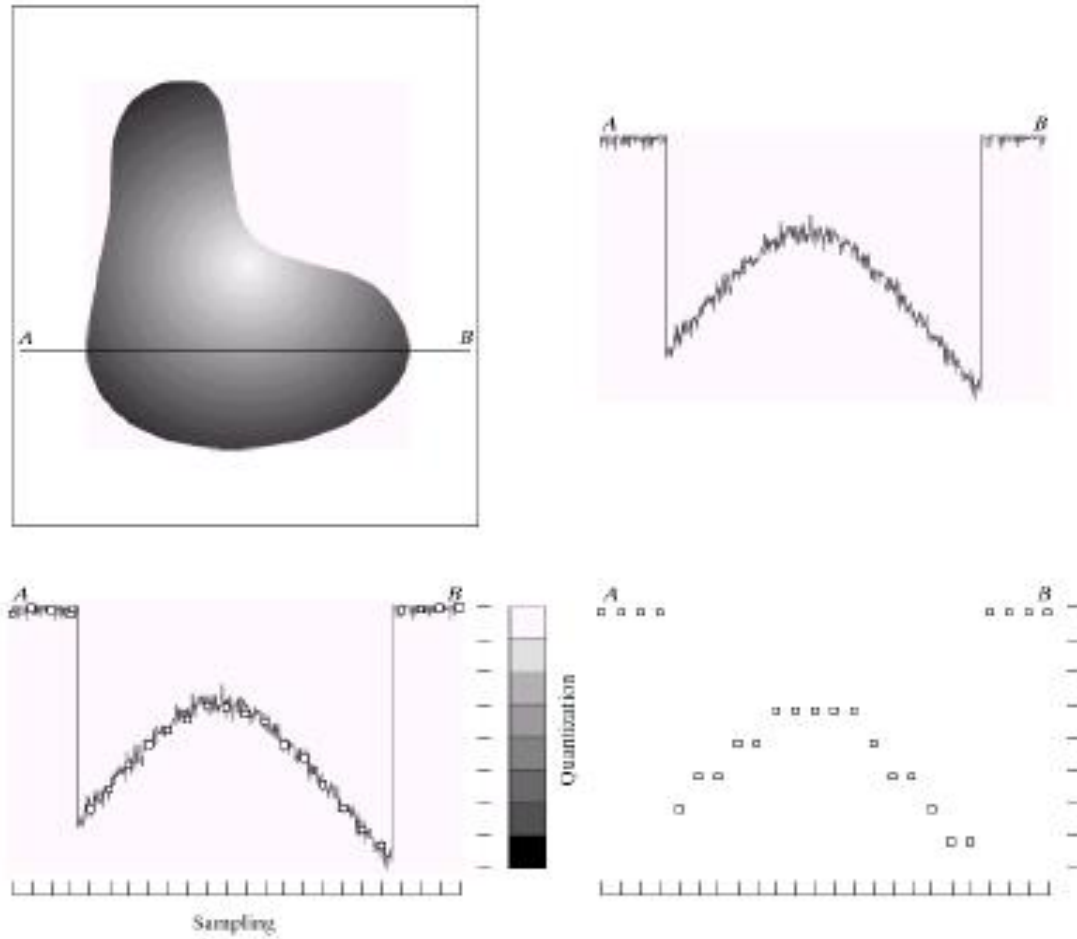
- Avec un échantillonnage adapté, l'image numérique fait apparaître des structures conformes à l'information présente dans l'image



- Mais en considérant seulement 1 échantillon sur 2, une structure différente apparaît, dont l'analyse (ici des bandes verticales, plus épaisses) ne sera pas conforme à la réalité de l'objet.

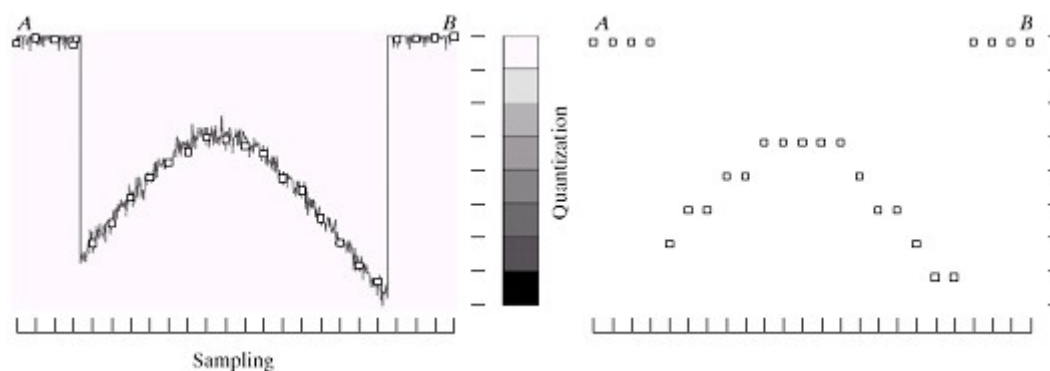


# ÉCHANTILLONNAGE ET QUANTIFICATION



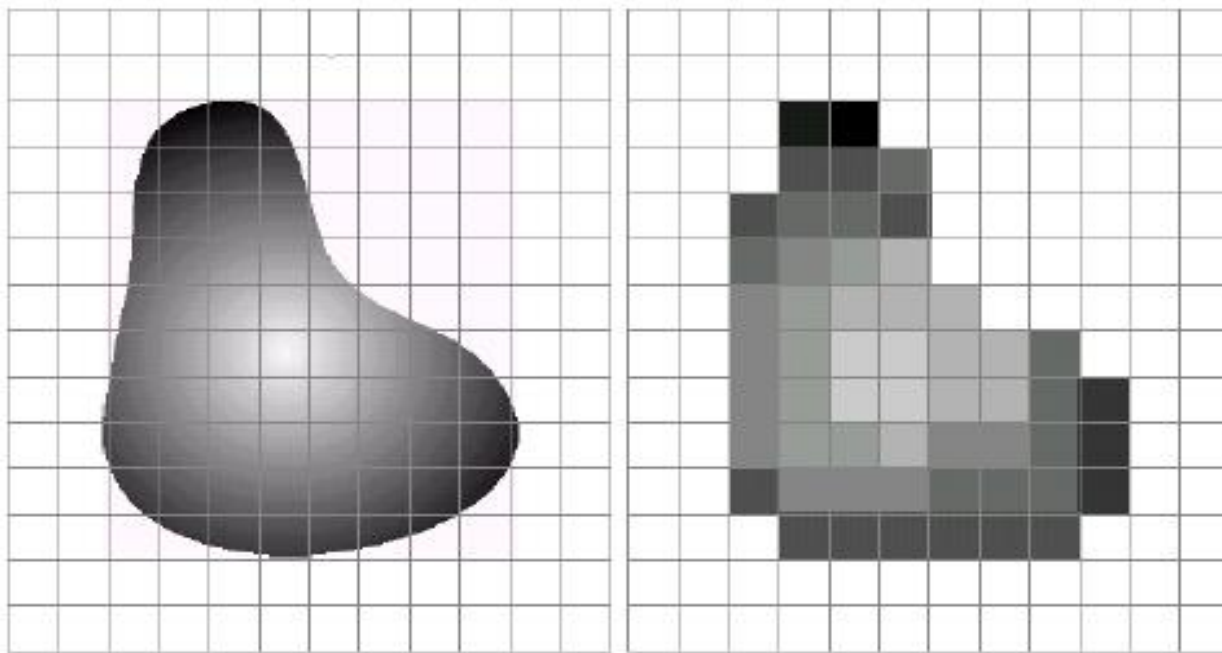
# ÉCHANTILLONNAGE ET QUANTIFICATION

- L'**échantillonnage** est limité par la capacité du capteur, donc le nombre de pixels disponible (ou autre limite imposée)
- La **quantification** est limitée par la quantité de tons (de gris) définie dans l'intervalle



# ÉCHANTILLONNAGE ET QUANTIFICATION

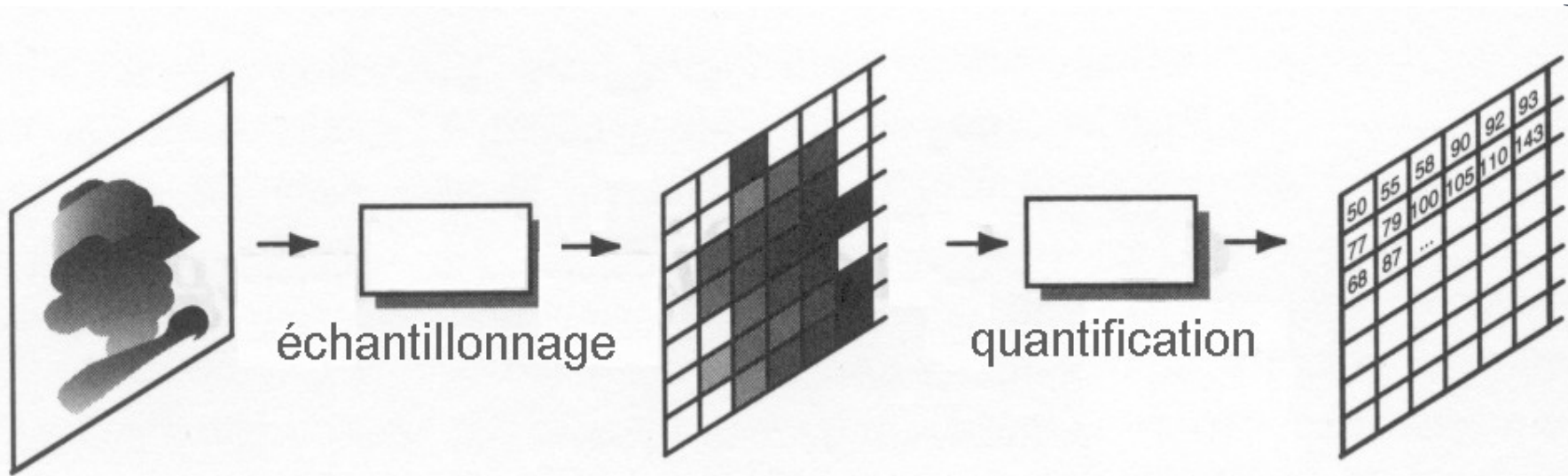
- Avec un capteur à matrice



a b

**FIGURE 2.17** (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.

# ÉCHANTILLONNAGE ET QUANTIFICATION



## REPRÉSENTATION DES IMAGES

- Matrice de dimension  $M \times N$
- Chaque élément a une valeur entière dans l'intervalle  $[L_{\min}, L_{\max}]$
- Le nombre de « bits » requis pour représenter les niveaux de gris dans l'intervalle «  $L$  » est «  $K$  »
- La relation entre «  $K$  » et «  $L$  » est :
  - $L = 2^K$
- Le nombre de bit pour entreposer un image est donc :
  - $b = M \times N \times K$

# RÉSOLUTION DES IMAGES

- Résolution **spatiale**
  - Le plus petit détail discernable
- Résolution **tonale** (de tons de gris)
  - Le plus petit changement discernable
- Une image a donc une résolution spatiale de **M X N** pixel et une résolution de tons de gris de **K** bits ou de **L** niveaux ou tons

# RÉSOLUTION DES IMAGES

## Résolution spatiale : échantillonnage



256x256



128x128



64x64



32x32

## Résolution tonale : quantification



6 bits



4 bits



3 bits



2 bits



1 bit

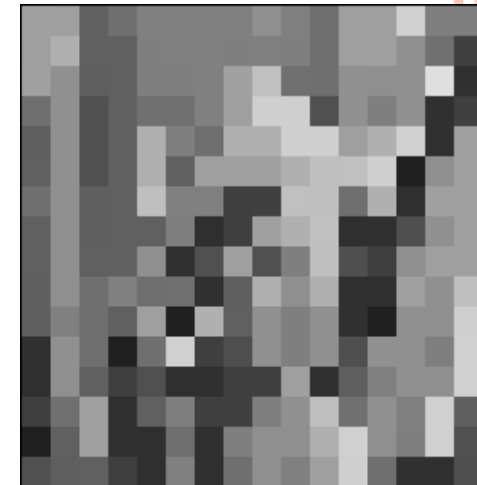
## Effets de l'échantillonnage : pixelisation



256 x 256 pixels



64 x 64 pixels



16 x 16 pixels dia

- Contours en marche d'escalier
- Perte de netteté
- Détails moins visibles/ moins précis
- Perte de résolution

# Effets de la **quantification** à l'acquisition

- Codage de la valeur de chaque pixel sur N bits (En général 8 bits)



8 bits (256 niv.)



4 bits (16 niv.)

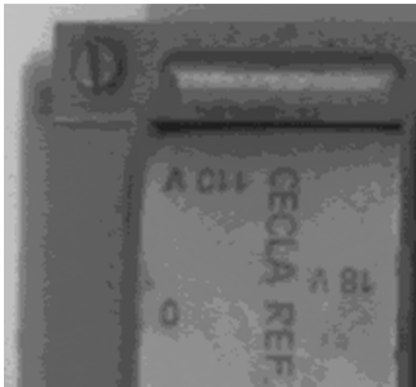


2 bits (4 niv.)

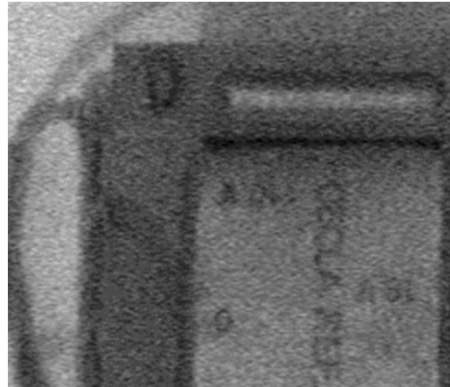
- Apparition de faux contours
- Bruit de quantification
- Effet visible à l'œil en dessous de 6/7 bits

## Bruits liés à l'acquisition

Les images sont souvent entachées de bruit, parfois non visible à l'œil, et qui perturberont les traitements



Diaphragme  
F/4

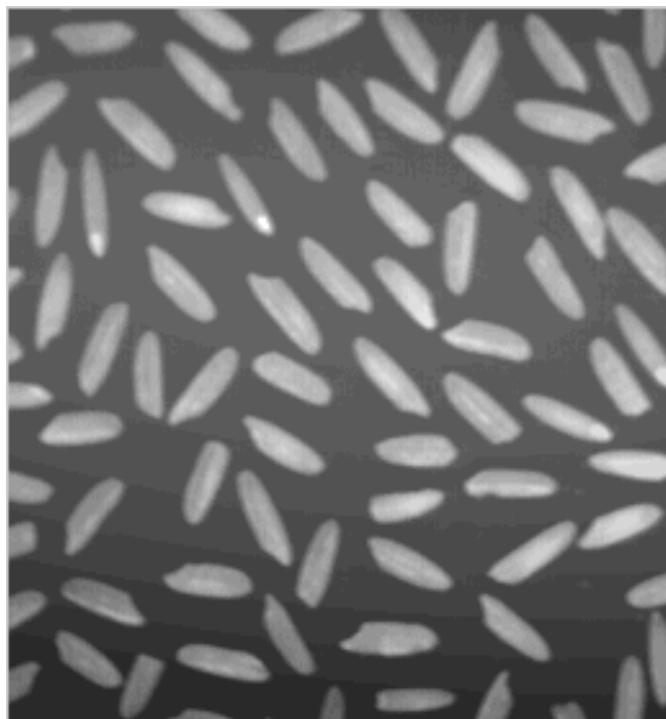


F/8

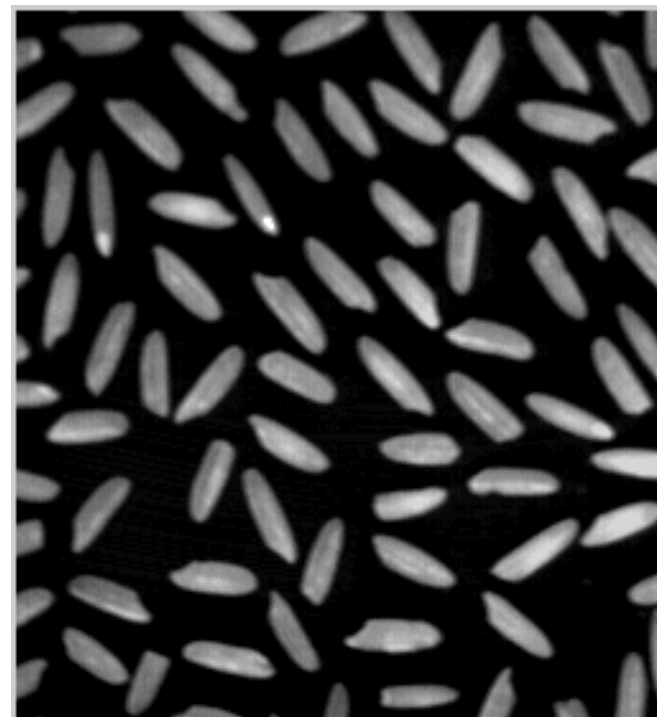
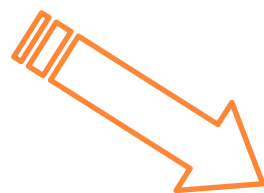


F/16

- Optimiser les conditions d'éclairage
- Attention à l'éclairage ambiant
- Mais... diaphragme ouvert = faible profondeur de champ
- Mais... éclairage important = dégagement de chaleur

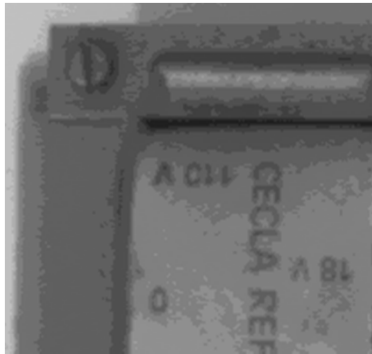


Eclairage non uniforme !



Correction de l'éclairage

- Flou de bougé/filé dû à un temps de pose/intégration trop long



Cet effet est limité par l'usage d'obturateur rapide et/ou d'éclairage flash

- Effet de lignage dû au balayage entrelacé des caméras vidéo

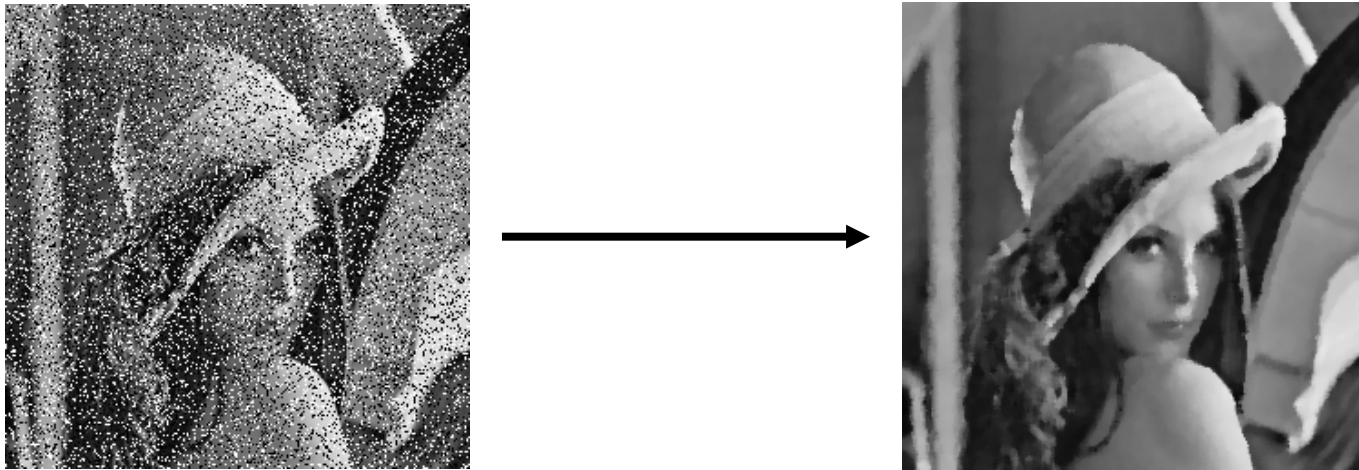


Cet effet disparaît avec les caméras à balayage progressif non entrelacé

Une bonne acquisition  Des traitements facilités

# FILTRAGE : RÉDUCTION DE BRUIT

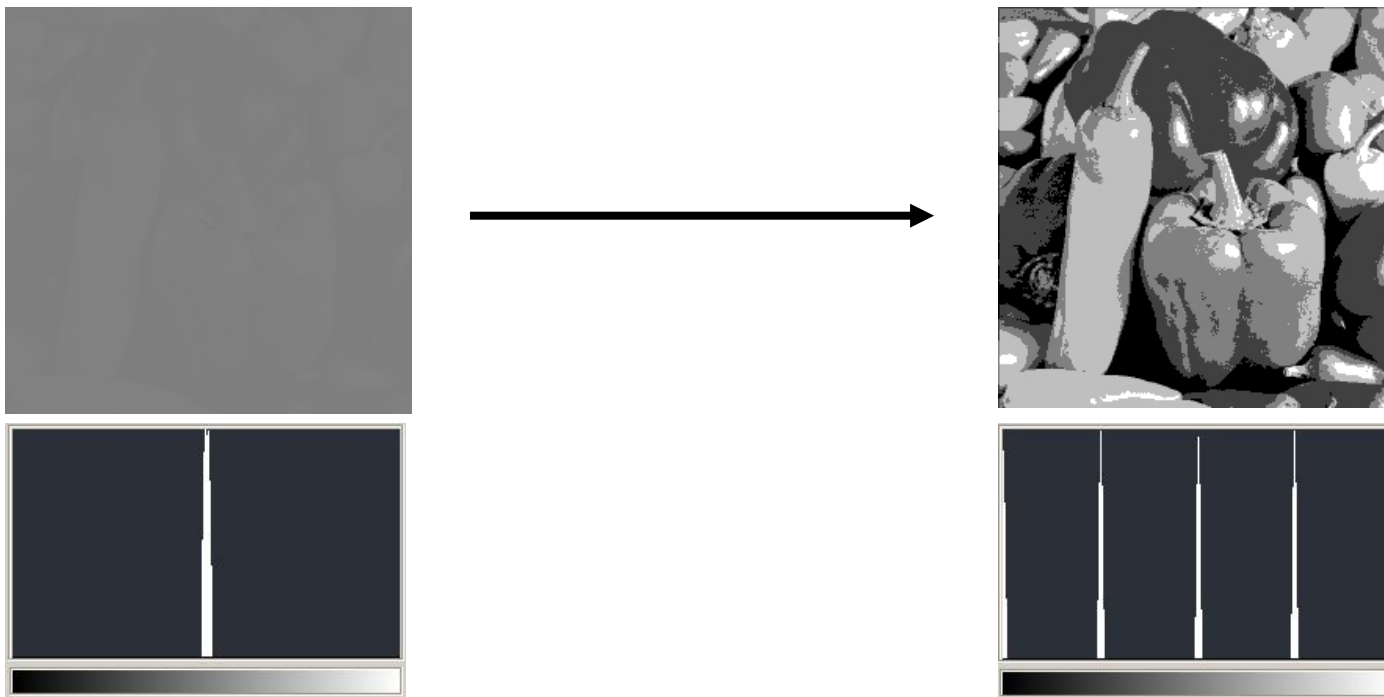
- Image originale rarement parfaite



- Dans certains cas, on ne connaît même pas le type de bruit

# FILTRAGE : REHAUSSEMENT DE CONTRASTE

- Améliorer la lisibilité de l'image



# SEGMENTATION

- Simplification qui permet de mieux comprendre le contenu de l'image

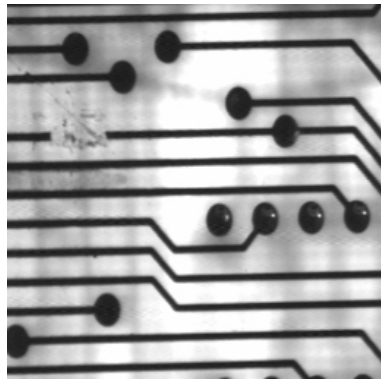
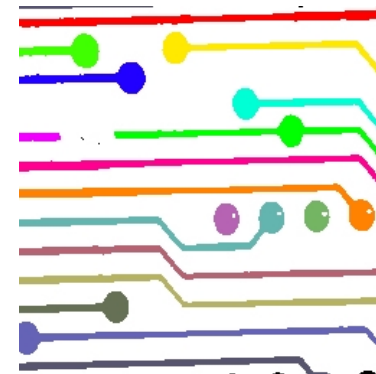


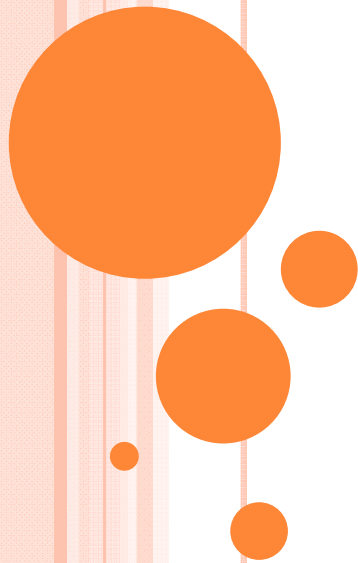
Image en entrée



Carte en sortie

- Segmentation = Partition de l'image

# LES VIDÉOS



# LES DIFFÉRENTES ÉTAPES DE LA PRISE DE DONNÉES À L'ANALYSE

- 1- Préparation du dispositif d'enregistrement Audio-vidéo :  
Choix du matériel d'enregistrement / vérification de son fonctionnement / Préparation des accessoires et consommables
- 2- Séance d'enregistrement = prise de vue + prise de son  
Recueil de données complémentaires : documents / notes...
- 3- Numérisation ou acquisition des données vidéo / audio  
sur l'ordinateur – opérations de montage - compression
- 4- Analyse des données avec logiciels

# ACQUISITION et compression

Source numérique  
ou analogique



Cordons  
analogiques  
(et carte d'acquisition)

*Acquisition /  
Numérisation*

Montage



*Compression*



Support externe (CD,  
DVD, disque dur,  
cassette DV...)



Fichiers  
compressés

# L'ACQUISITION

- Également nommée transfert, capture ou numérisation
- Consiste à transférer des contenus stockés sur un support audiovisuel vers un ordinateur
- La source de l'enregistrement peut être :
  - analogique (magnétoscope VHS, caméscope hi8...)
  - numérique (caméscopes mini-DV, DVD, à disque dur...)
- Intérêts de la numérisation :
  - permettre des traitements informatiques (montage, retouche, incrustation, synchronisation...)
  - faciliter la diffusion et l'échange des données (internet, supports informatiques...)
  - conservation des données, copie ultérieure sans dégradation du signal

# LES LOGICIELS NÉCESSAIRES À L'ACQUISITION

L'acquisition sur ordinateur nécessite un logiciel intégrant une fonction de capture. La plupart des logiciels de montage vidéo proposent ce type de fonction :

## ■ Les logiciels de montage « grand public »

- Caractéristiques : simple à utiliser et généralement gratuit. Fonctions limitées pour le montage et l'enregistrement des données
  - » Imovie (Mac) ; Windows Movie Maker (Windows XP et Vista) ;
  - » Les logiciels livrés avec les caméscopes

## ■ Les logiciels de montage semi-professionnels :

- Caractéristiques : relativement complets et ergonomiques ; coût abordable
- Ex : Premiere Elements, FinalCut Express, Pinnacle Studio...

## ■ Les logiciels de montage professionnels :

- Caractéristiques : très complets, complexes à utiliser ; onéreux ; lourds
- Ex : Adobe Premiere Pro, FinalCut Pro, Avid Xpress, SonyVegas...

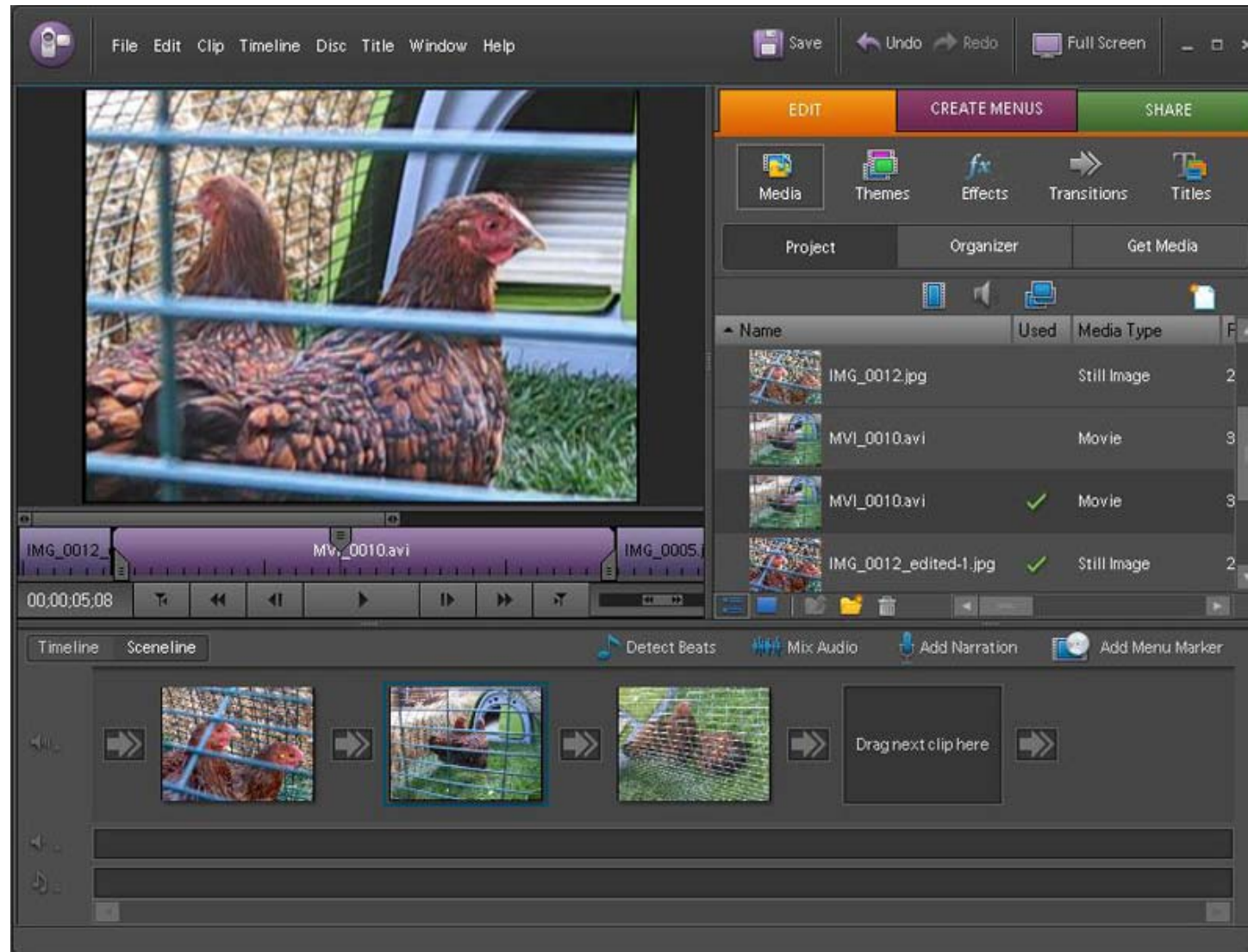
# LES LOGICIELS NÉCESSAIRES À L'ACQUISITION

## ○ iMovie



# LES LOGICIELS NÉCESSAIRES À L'ACQUISITION

- Adobe Premiere Elements



# LES LOGICIELS NÉCESSAIRES À L'ACQUISITION

- Adobe Premiere Pro



# COMPOSITION ET LECTURE DES FICHIERS VIDÉOS

- Un fichier vidéo se compose généralement de 2 éléments :
    - Le **conteneur**. Il correspond généralement au format du fichier. Son rôle est de rassembler et d'organiser dans un fichier, différents types de données (flux audio, vidéos, sous-titres, meta-données...). Exemple de conteneurs :
      - L'Audio Video Interleave (.AVI), Quicktime (.MOV), Real Media (.RM), MP4
      - FLV (Flash Vidéo), MKV (Matroska), le WMV (Windows Media Video),
    - **Chaque conteneur possède ses spécificités** en termes de nombre de pistes acceptées pour la vidéo et l'audio, de codecs reconnus...
    - Le(s) **contenu(s)** : Ils se composent essentiellement de **flux audios et/ou vidéos**.  
Ceux-ci sont généralement compressés à l'aide d'un codec (**algorithme de compression/décompression**) comme le Divx, le H264, le mp3...  
Ex : un flux vidéo au format Divx peut être « encapsulé » dans un conteneur AVI ou Quicktime.
  - Un fichier vidéo nécessite un « lecteur multimedia » (« player » en anglais) pour le lire
    - ex: Windows Media Player pour les fichiers AVI, Quicktime pour les MOV, Real Player pour les RM, VLC, MediaPlayer Classic...
- Remarque : les lecteurs ne savent lire qu'un nombre restreint de conteneurs (ainsi on ne pourra pas lire les fichiers Quicktime avec Windows Media Player...)

# INTÉRÊT DE LA COMPRESSION

- Constat : une fois capturées les vidéos sont très volumineuses :
  - 1 heure de vidéo non compressée occupe environ 85 Go (dimension 720/576 ; 24 img/s ; couleurs 24 bits)
  - 1 heure au format DV occupe environ 13 Go sur le disque dur

- Solution : compresser les données à l'aide d'un CoDec

Définition : ensemble d'algorithmes mathématiques implémentés dans un environnement logiciel et/ou matériel qui permettent de réduire l'encombrement d'un fichier.

# LES DIFFÉRENTS TYPES DE COMPRESSIONS

- Compression non destructives (lossless):
  - Des algorithmes mathématiques compressent les données **sans aucune perte d'information**
  - La décompression **restitue intégralement les données initiales**
  - Le gain de taille par rapport à l'original est moindre  
*Exemple de compression non destructive : le format ZIP*
- Compression destructives (lossy) :
  - Supprime définitivement les informations peu représentatives (détails...) ou non perceptibles (fréquences non audibles, nuances de couleurs...)
  - Le gain de taille est très important. Le ratio de compression peut être très important (ex. 10 : 1 voire beaucoup plus)  
*Exemples de compressions destructives : le JPEG, le MPEG...*

# LES TECHNIQUES DE COMPRESSION

## ■ Compression spatiale

- Dans une image fixe non compressée, les pixels sont codés individuellement
- La compression spatiale recherche sur l'image les éléments redondants : par un exemple un ciel de couleur uni. On code ensuite la description du bloc correspondant au ciel, plutôt que chaque élément individuellement.



## ■ Compression temporelle

- Dans une suite d'images (vidéo), seuls sont codés les éléments qui varient d'une image à l'autre.  
Ex. : personnage qui bouge dans un plan fixe

# LES CODECS

## (COMPRESSEUR / DECOMPRESSEUR)

- Il existe différents algorithmes de compression plus ou moins performants. Chaque codec possède ses spécificités et ses domaines de prédilection.
- Les plus anciens : Le MPEG 1, Cinepak, Sorenson Video...
- Les plus récents, ils sont tous dérivés de la norme MPEG 4 :
  - Le DIVX : format très populaire sur Internet
  - XVID : similaire au DIVX en termes de performances mais gratuit (format « open source »)
  - WMV : le format propriétaire de Microsoft
  - H264 - MPEG4-AVC : résulte du travail du groupe MPEG et IUT-T

# TEMPS DE COMPRESSION ET NOTION DE DÉBIT

- Le temps de compression dépend de plusieurs **facteurs** :
  - facteurs **matériels** : caractéristiques du micro-processeur équipant l'ordinateur, quantité de mémoire installée...
  - facteurs **logiciels** : codec et logiciel utilisés, options choisies...
  - **qualité du rendu** final : nombre d'images par seconde, dimension de l'image (résolution), débit (bitrate), nombre d'images clés...

- Notion de débit (bitrate en anglais) :

Le débit désigne la quantité d'informations transférée en l'espace d'une seconde. L'unité de mesure est le bit par seconde (bit/s) et ses différents multiples (Kilo-bits par seconde, Mega-bits par seconde).

Le choix du débit aura un impact direct sur la taille et la qualité du fichier final

- le poids du fichier est directement proportionnel au débit utilisé
- à partir d'un certain débit, l'amélioration de la qualité n'est plus perceptible

- on estime qu'un débit de 1 500 Kbit/s (~150 Ko/s) représente un bon compromis pour obtenir une vidéo de qualité à partir d'un codec dérivé du MPEG 4 (DIVX, WMV, XVID, H264...). Attention, la notation anglaise du Ko est KByte (ne pas confondre avec le Kilobit)

# TABLEAU COMPARATIF (CODECS)

	Format	1 seconde (Ko) *	15s en Mo	1mn Mo	1 heure (Mo)	Sur 1 CD 700 (en mn)	Evaluation Qualité/Poid de 1 à 5	Points faibles
Audio	Son 44Khz, 16bits, Stereo (qualité CD audio <b>non compressé</b> )	150	2,2	8,8	527	80	5	
	Son 44Khz, Stereo (qualité CD audio <b>compressé MP3</b> , 128 kbits)	16		0,9	56	747	4,5	
Audio + Video	<b>Format DV</b> (Video et Son)	3600	54,3	217,2	13032	3	5	Encombrement
	<b>Format MPEG</b> (Pal : 352*288 + Audio 224 Kbits)	170	2,5	10,0	600	70		
	<b>XVID</b> debit 900 Kbits - 720 * 576 - (Son MP3 128 Kbits)	100	1,5	6,0	360	117	4,5	aucun
	<b>DIVX</b> debit 900 Kbits - 720 * 576 - (Son MP3 128 Kbits)	100	1,5	6,0	360	117	4,5	codec sous licence (apparition d'un logo en bas à droite durant les premières secondes)
	<b>Quicktime</b> (DSL/Cable "élevé"; 320*240), 25ips	100	1,5	6,0	360	117	2,5	taille (320*240)
	<b>Quicktime</b> (Mpeg 4, 25ips, 720*576, 1image clef pour 100 images, debit 150Ko, qualité optimale son Mpeg4)	165	2,3	9,2	552	76	3,5	encombrement, rendu durant des séquences de mouvement
	<b>Windows Media Video</b> (video pour réseau local 1Mbits)	125	1,8	7,2	432	97	4,5	Format propriétaire (Microsoft)

Techniques Multimédia

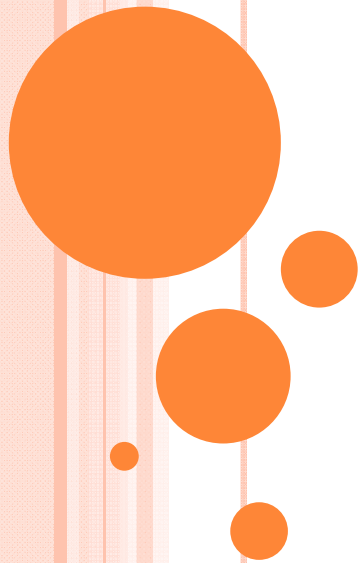
## LES CODECS EN RÉSUMÉ

- Si l'on désire travailler avec un format « universel » :  
MPEG 1 ou 4, le DIVX (du fait de son intégration progressive dans les équipements multimédias)
- Pour sa simplicité d'utilisation : Le Windows Media Video (intégré à Windows Movie Maker ou téléchargeable sur le site de Microsoft)
- Si la qualité est prépondérante : le DIVX, le XVID ou le H264 ;
- Pour le montage : le format DV
- Pour l'archivage pérenne : le format DV ou le MPEG 2



Les logiciels d'analyse et d'annotation de vidéo n'acceptent que certains types de fichiers (MPEG 1, MOV, AVI...).

# LES VIDÉOS SUR INTERNET



## EVOLUTION : L'EXEMPLE US

- Le nombre de foyers équipés du Haut débit a augmenté de 227 % pour atteindre 71 millions de foyers (2008)
- 78,5 % de l'audience US sur Internet regardent les vidéos en ligne régulièrement
- La moyenne d'un spectateur de vidéo en ligne est de plus de **5 heures (309 mn)** par mois
- Les spectateurs Internet ont regardé un record de 14,3 milliard de vidéos en ligne en Décembre 2008

● Source: comScore. Chiffres de 2009, non encore actualisés

# EVOLUTION

Over 60% Internet Traffic is **Video...**

Over 90% In 2012



David Martinon présente l'agenda du Président de la République du 9 au 15 ...

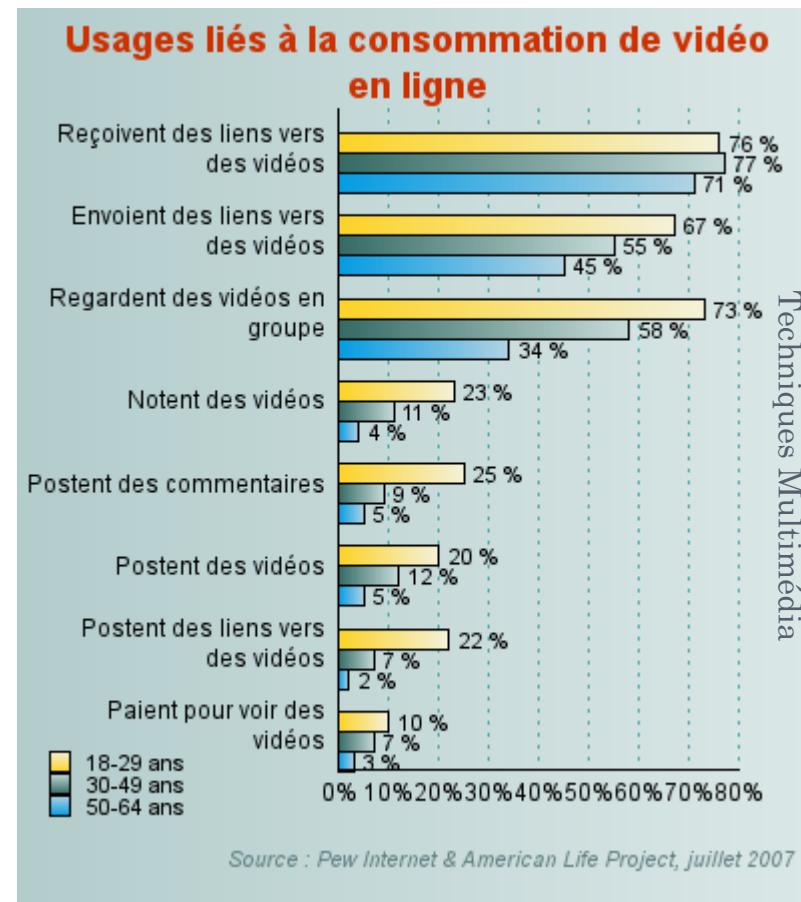
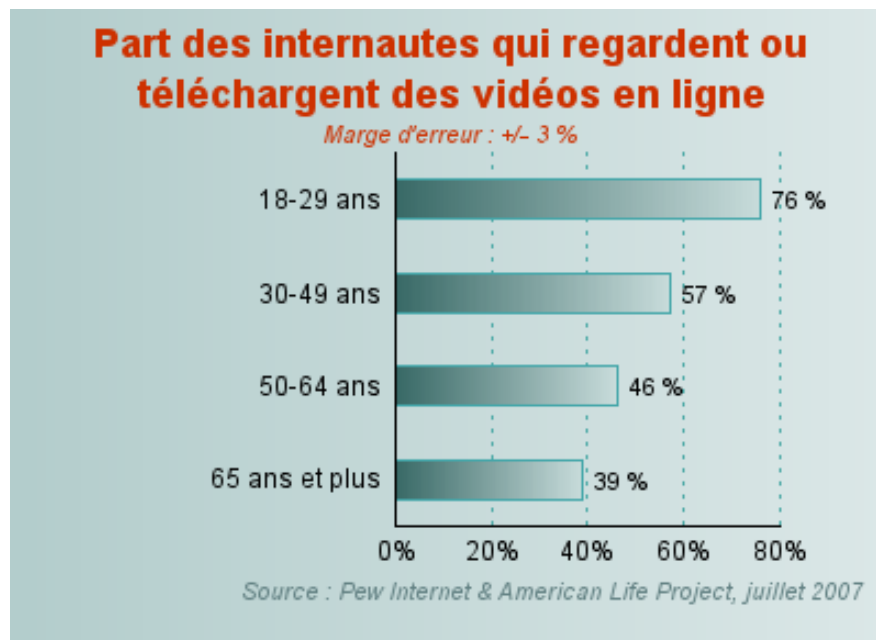


Durée : 11mn25. - Téléchargements : .wmv .mpeg

Techniques Multimédia



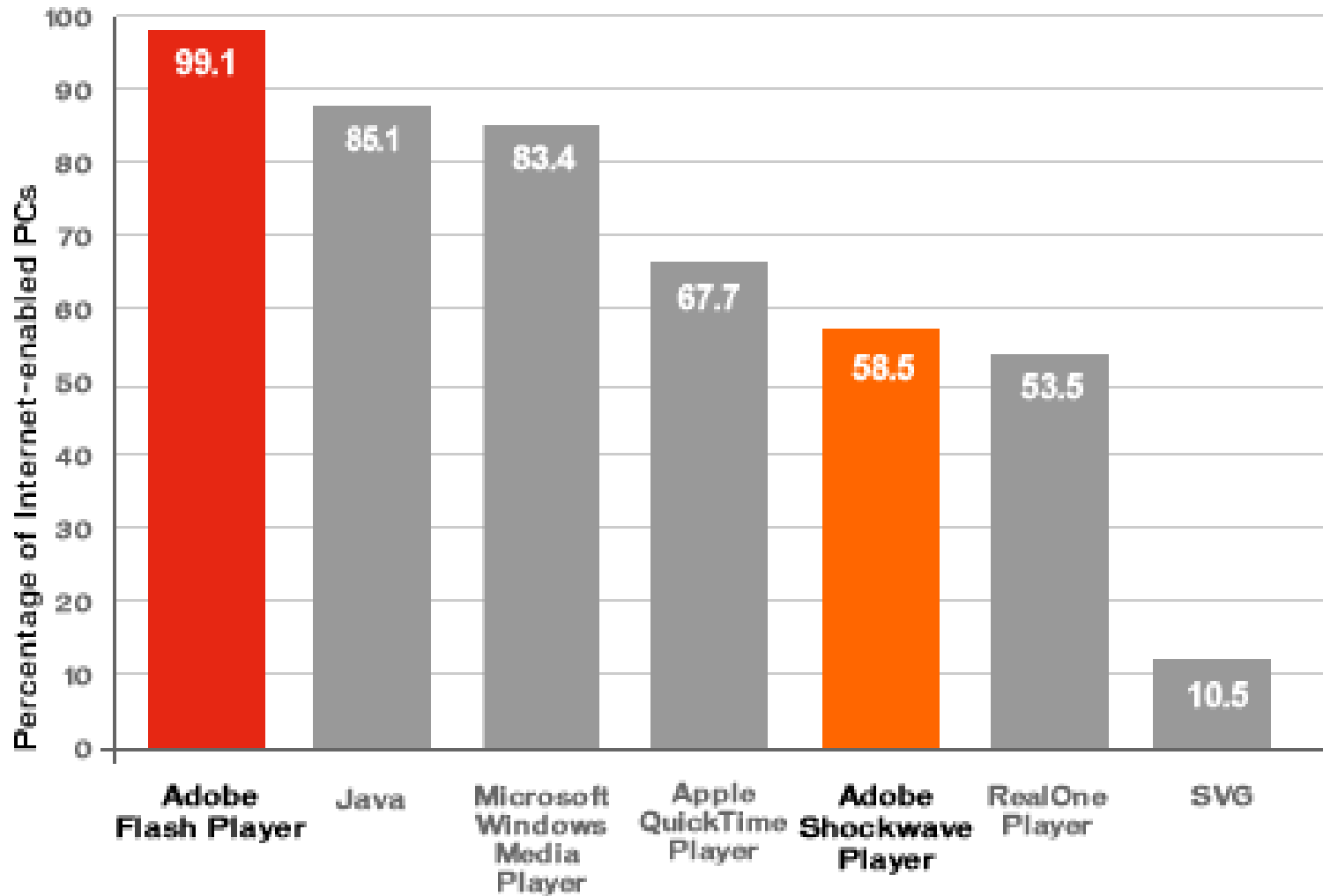
# LES INTERNAUTES CONSOMMENT ET PRODUISENT DE PLUS EN PLUS DE CONTENU VIDÉO



Sources : [http://www.pewinternet.org/PPF/r/219/report\\_display.asp](http://www.pewinternet.org/PPF/r/219/report_display.asp)

Journal du Net

# QUEL LECTEUR UTILISENT LES INTERNAUTES ?



## CONTRAINTES TECHNIQUES

- Palier à l'incompatibilité des lecteurs.
- Permettre une lecture instantanée et multi plate-forme.
- Associer la vidéo à une interface intuitive.
- Naviguer par chapitre à l'intérieur d'une vidéo.
- Automatiser le renseignement et la manipulation de méta-données.

## FORMATS DE VIDÉO PC ET MOBILES

- Parmi les formats vidéo les plus utilisés aujourd'hui sur ordinateurs:
  - QuickTime, AVI, Mpeg, Real Video et Windows Media.
  - H264 (nouvelle génération de codec vidéo)
    - amélioration du taux de compression meilleure
    - qualité d'affichage
- Vidéo pour mobiles
  - 3GPP, MP4.

# MÉTHODES DE DIFFUSION DE FLUX VIDÉO

## 1. Les fichiers vidéo inclus

- Clips vidéos de petite taille
- Lecture différée. (téléchargement complet)

## 2. Téléchargement progressif

- Début de la lecture avant la fin du téléchargement
- Lecture de haute qualité quelle que soit la vitesse de connexion des utilisateurs.

## 3. Diffusion de flux vidéo (video streaming)

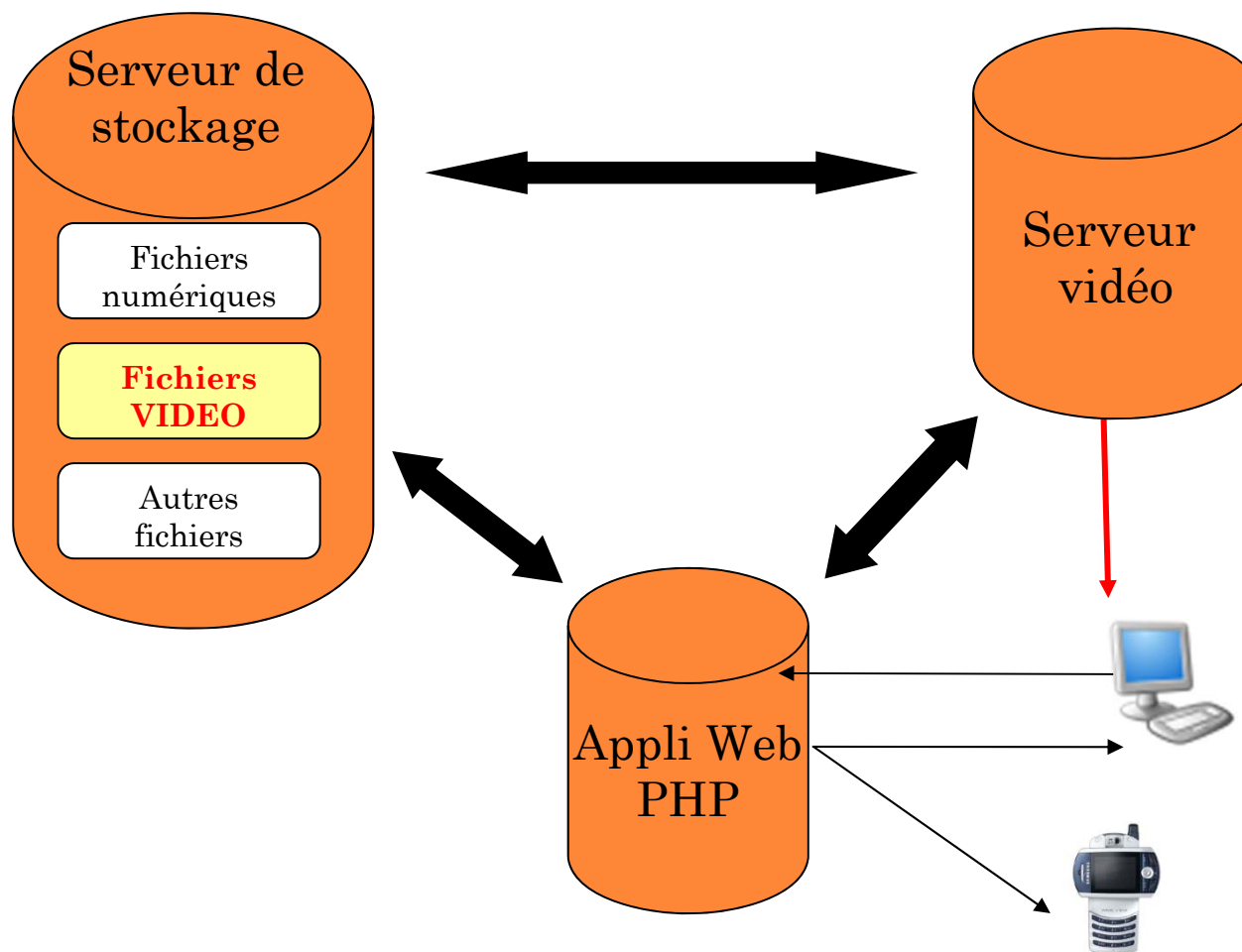
- Diffusion en tant réel depuis un serveur spécialisé
- Flux de données joué en continu
- Flux non stocké chez le client

# SOLUTIONS DE GESTION DE VIDÉO

- Choix : téléchargement progressif ou video streaming
- Intégration d'un serveur dédié à la vidéo
  - Serveur performant (encodage)
  - Bande passante assez élevée (flux simultanés, qualité)
  - Capacité de stockage importante (serveurs spécialisés)
  - Encodage durant l'upload

# SOLUTIONS DE GESTION DE VIDÉO

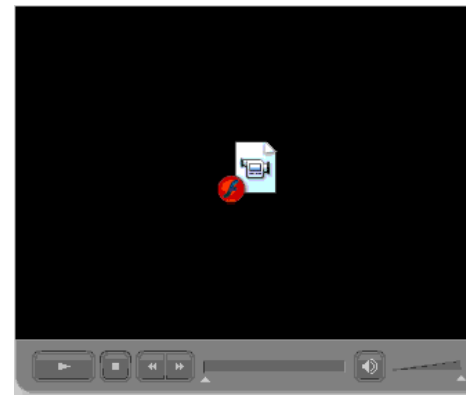
- Intégration d'un serveur dédié à la vidéo



# SOLUTIONS DE GESTION DE VIDÉO

## ○ **Macromedia FLASH**

- Format FLV (Flash video)
- Ne Supporte pas les fichiers vidéos de téléphones
- Serveur d'application « Flash Media Server »
- FLASH 10-11 Video Encoder
- Plugin Flash (98,3 %)



# SOLUTIONS DE GESTION DE VIDÉO

## ○ VIDEOLAN

- Solution open source
- Streaming vidéo
- Support des formats vidéos (PC et mobiles)
- Gestion des contrôles via javascript
- Inconvénients : plugin VLC disponible uniquement sous Firefox (12%)



# SOLUTIONS DE GESTION DE VIDÉO

## ○ Outils open source :

- Encodeur et serveur de streaming : FFMPEG
- Encodeur FLV : Riva Video Encoder



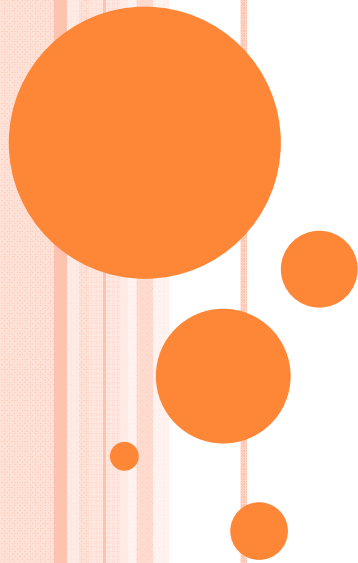
## ○ Solutions commerciales :

gestion de contenu vidéo, diffusion et stockage

- VitalStream (Flash et Windows Media)
- MediaBox (Flash video)



# LA VIDÉO NUMÉRIQUE



# LA VIDÉO NUMÉRIQUE

- Sur l'image numérique, lorsqu'un rayon lumineux frappe notre rétine, des terminaisons nerveuses spécialisées (ou **bâtonnets**) réagissent à la quantité de lumière alors que d'autres cellules (ou **cônes**) réagissent sélectivement au rouge, au vert et au bleu, permettant ainsi la perception des couleurs grâce à l'**analyse colorimétrique** de la lumière blanche.
- Nos yeux décomposent donc la lumière en **intensité** et en **couleurs**, ce qui signifie la décomposition en informations de luminance et de chrominance :

# SIGNAL VIDÉO

- Sur un récepteur de télévision, une image est obtenue par balayage électronique successif des **lignes de luminophores** constituant l'écran.
- Un écran au standard **PAL** ou **SECAM** comprend 625 lignes, alors qu'un écran au standard **NTSC** n'en possède que 525.
- On sait que, pour donner à l'œil humain une impression d'éclairement continu à l'aide d'une source lumineuse émettant de façon discontinue (vidéo projecteur, projecteur de cinéma, écran TV, ... ), il faut au moins 40 éclats par seconde :
  - Une fréquence de **20 à 40 éclats** provoque une impression de scintillement
  - Une fréquence au-dessous de **20 éclats**, l'éclairement paraît saccadé
  - Pour obtenir une impression de continuité parfaite avec 25 images/s, on divise chaque image en **2 demi-images** projetées successivement, provoquant ainsi 50 sensations différentes par seconde.

# LA VIDÉO NUMÉRIQUE

- **NTSC :**

standard de diffusion TV du continent nord-américain et du Japon, la fréquence du courant électrique est 60Hz, la fréquence d'affichage est de **30 images/s (exactement 29,97 im/s) sur 525 lignes.**

- **PAL :**

standard de diffusion TV couleur, qui a été mis point en Allemagne vers 1960. Son principal avantage relativement au standard américain dont il est très proche, est de remédier aux erreurs de phase dont souffre ce dernier.

- **SECAM :**

standard français, il comporte **625 lignes et 25 images/s.** la fréquence du courant électrique distribué en Europe est en effet de 50Hz.

- **Standard CCIR 601 :**

comité international des radiocommunications. Cet organisme a été remplacé par le secteur de radiocommunication de l'union internationale des télécommunications (UIT) : **standard concernant la vidéo numérique.**

# LA VIDÉO NUMÉRIQUE

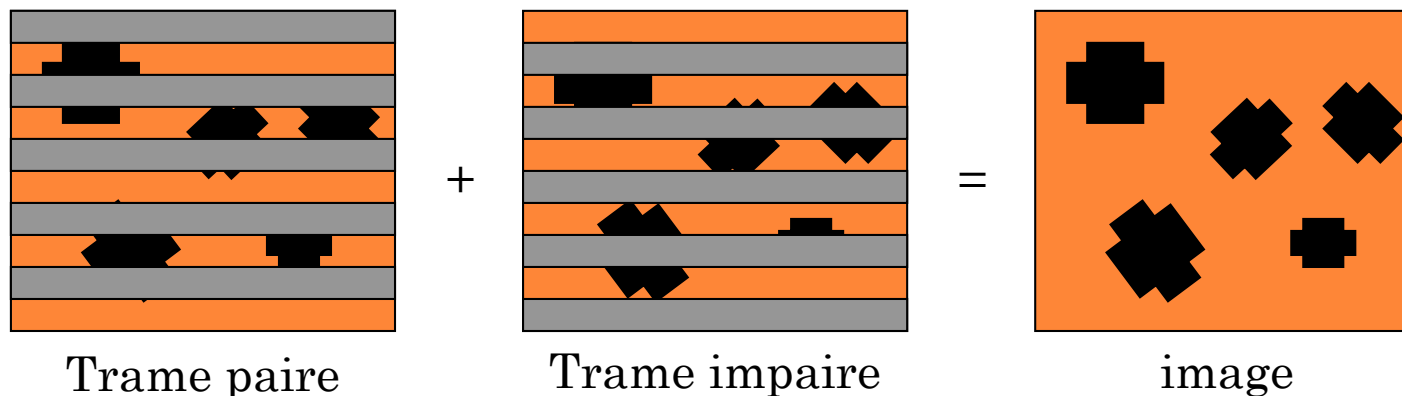
- Image en continuité parfaite :

Les demi-images ou trames sont obtenues par balayage des seules **lignes paires** dans un 1er temps, puis des **lignes impaires** dans un 2e temps.

→ *l'image complète est obtenue par **entrelacement des 2 trames**.*

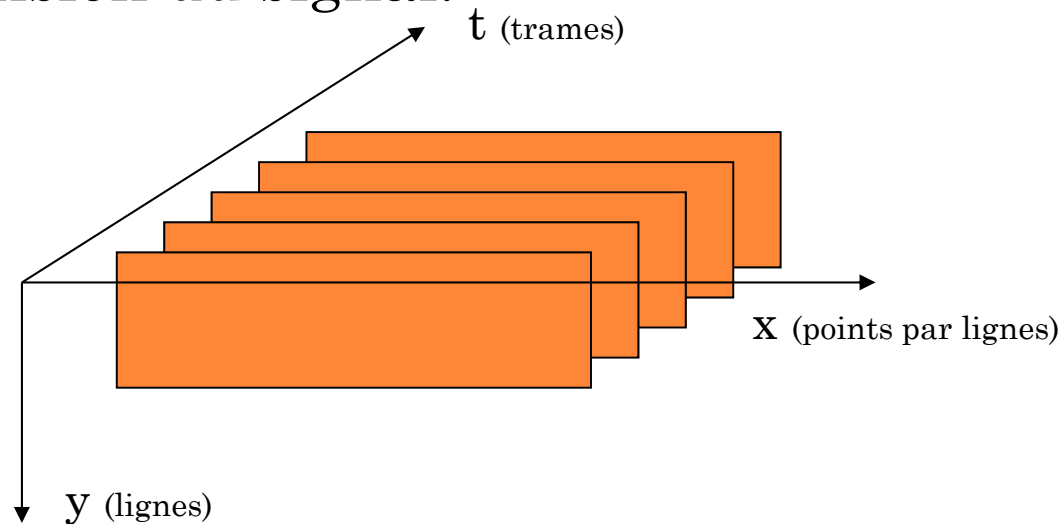
- Exemple :

*le signal vidéo analogique est donc découpé en 25 ou 29,97 images/s (en. frames), elles-mêmes divisées en  $25 \times 2 = 50$  ou en  $29,97 \times 2 = 59,94$  trames (en. fields), celles-ci étant à leur tour analysées en 312,5 ou 262,5 lignes.*



# LA VIDÉO NUMÉRIQUE

- Ce découpage est essentiel car il constitue un échantillonnage préalable du signal en **lignes** et **trames**.
- Cela signifie que, pour numériser le signal vidéo, il suffira d' **échantillonner la ligne** qui constitue la 3e dimension du signal.



# ÉCHANTILLONNAGE VIDÉO

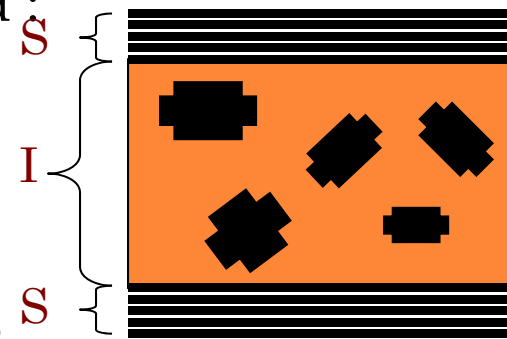
- Échantillonner la ligne :

en vidéo analogique, l'unité minimale est la ligne, car le signal est conçu pour piloter en continu les déplacements du triple faisceau d'électrons (RVB) et NON pour exprimer la couleur spécifique de chaque point de l'image.

- **Résolution verticale :**

le **nombre de lignes** sur un écran (résolution verticale) est exactement déterminé par chaque standard :

- **PAL et SECAM** distinguent 625 lignes dont 576 seulement sont destinées au balayage réel de l'image, le reste étant affecté à divers « services » de synchronisation et de positionnement des faisceaux (changement de ligne et de trame)
- **NTSC** n'affecte quant à lui que 480 lignes sur 525 à l'affichage de l'image



Résolution verticale 160

# LA VIDÉO NUMÉRIQUE

- **Résolution horizontale :**

la résolution horizontale quant à elle n'est pas un paramètre fixe du signal vidéo analogique: elle dépend de la bande passante allouée à la vidéo et à la qualité du moniteur sur laquelle elle viendra s'afficher.

- L'image télévisée standard obéit à un **ratio de 4:3** (ie. largeur de l'image est 1,33 fois plus grande que sa hauteur), on peut 'estimer' le nombre de points par ligne (ie. estimer la largeur de l'image en nombre de points) :

- **PAL, SECAM :**

L'image ayant une hauteur de 575 lignes, sa largeur sera :

$$575 \times 1,33 = 768 \text{ points}$$

- **NTSC :**

L'image ayant une hauteur de 480 lignes, sa largeur sera de 640 points.

- La résolution théorique du standard NTSC est à l'origine du célèbre affichage **VGA de 640 x 480** sur les moniteurs informatiques.

# LA VIDÉO NUMÉRIQUE

- Au cours des années 1970, les industriels commencent à prendre conscience de l'avenir de la vidéo en passant par le numérique.
- En conséquence, ils commencent à s'intéresser aux techniques de numérisation applicables au secteur des télécommunications, télédiffusions, télédétections, etc.
- Pour éviter l'anarchie dans les formats propriétaires, en 1982, le **standard CCIR 601** a été adapté par l'UIT pour la télévision numérique :
  - Le nombre de points actifs par ligne est de 720 (pour tous les autres standards)
  - Le nombre total des points par ligne (ppl) est de :
    - 858 ppl en NTSC (**525 lignes / 59,94 trames par seconde**)
    - 864 ppl en PAL et SECAM (**575 lignes / 50 trames par seconde**).

## EXEMPLES :

- Dans le standard PAL, à raison de 625 lignes par image, de 864 points par ligne et de 25 images/seconde, on obtient :

$$625 \text{ lignes/image} \times 864 \text{ points/ligne} \times 25 \text{ images/s} \\ = 13\,500\,000 \text{ points/s}$$

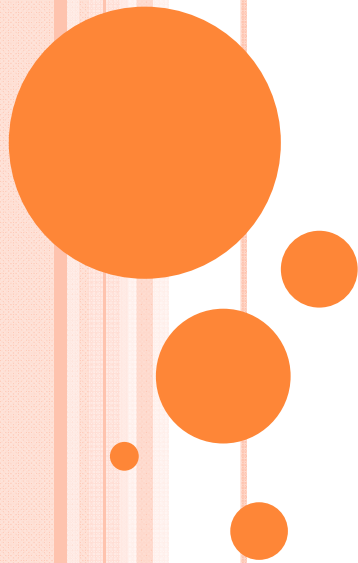
- Dans le standard NTSC, on obtient exactement le même nombre de points par seconde :

$$858 \text{ points/ligne} \times 525 \text{ lignes/s} \times 30 \text{ images/s} \times 1000/1001 \\ = 13\,500\,000 \text{ points/s}$$

→ *si 1 seconde de vidéo correspond à 13 500 000 points et si l'on veut conserver toute l'information concernant ces points alors il est nécessaire de les décrire individuellement.*

→ *pour 1 seconde vidéo, on doit définir 13 500 000 échantillons : la fréquence d'échantillonnage requise dans ces conditions est donc de 13,5 MHz*

# LE STANDARD MPEG



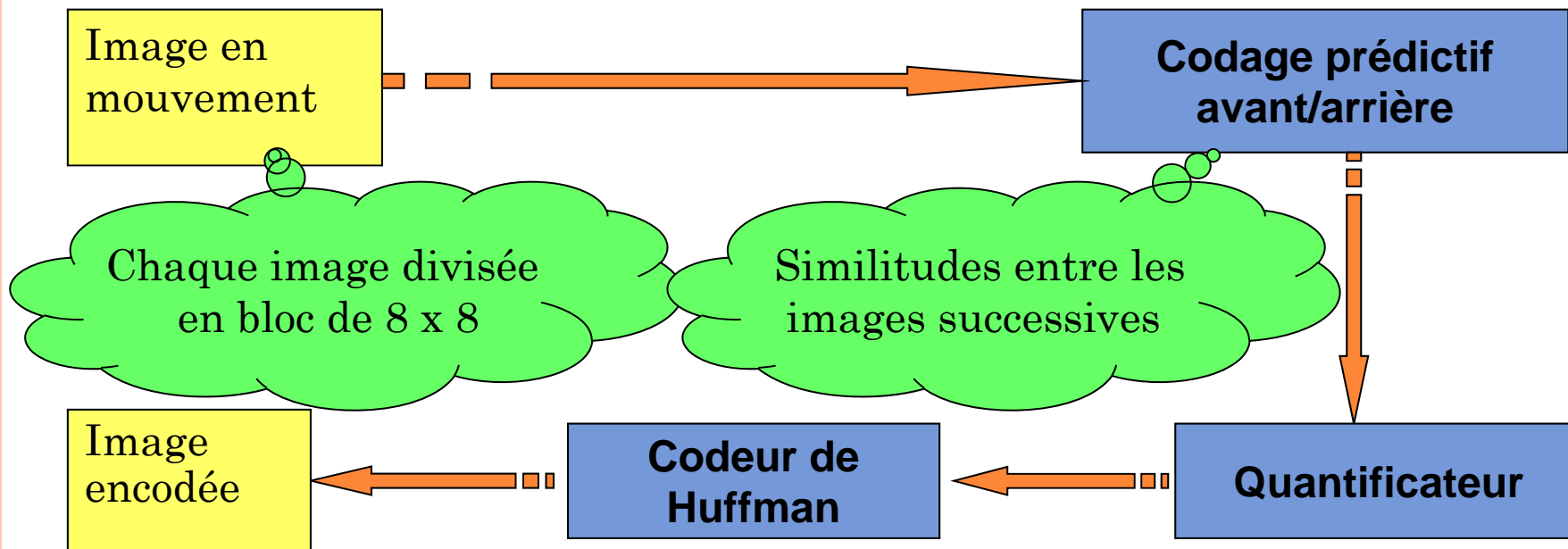
# LE MPEG

(*MOVING PICTURES EXPERTS GROUP*)

- Dans de nombreuses séquences vidéos, de nombreuses scènes sont fixes ou bien changent très peu, c'est ce que l'on nomme la **redondance temporelle**.
- Lorsque seules les lèvres de l'acteur bougent, presque seuls les pixels de la bouche vont être modifiés d'une image à l'autre, il suffit donc de ne décrire seulement le changement d'une image à l'autre. C'est là la différence majeure entre le MPEG (*Moving Pictures Experts Group*) et le M-JPEG. Cependant cette méthode aura beaucoup moins d'impact sur une scène d'action.

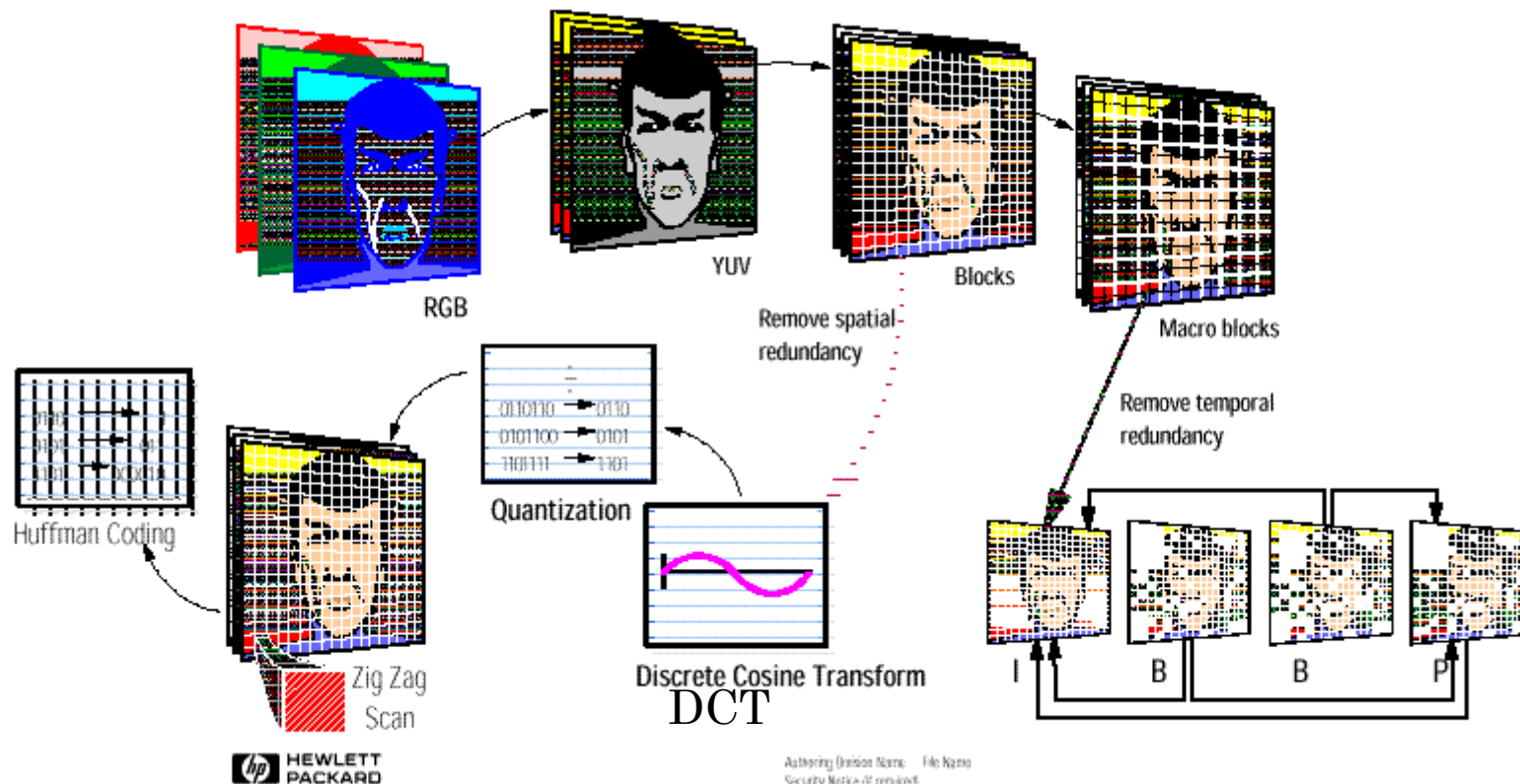
# LE STANDARD MPEG

- Schéma de traitement - (transmission)



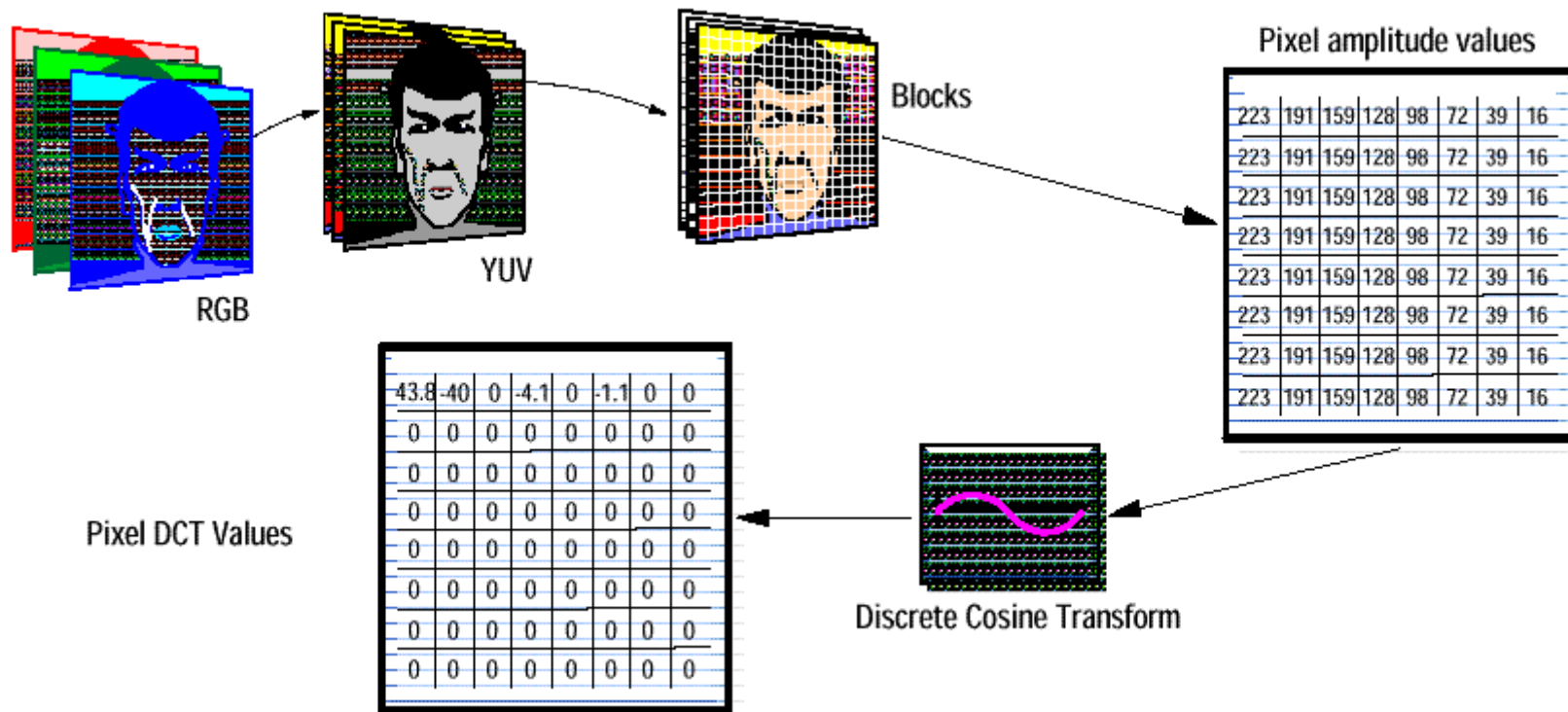
# LE STANDARD MPEG

- Redondance spatiale et temporelle:



# LE STANDARD MPEG

- Codage des pixels avec les DCT:

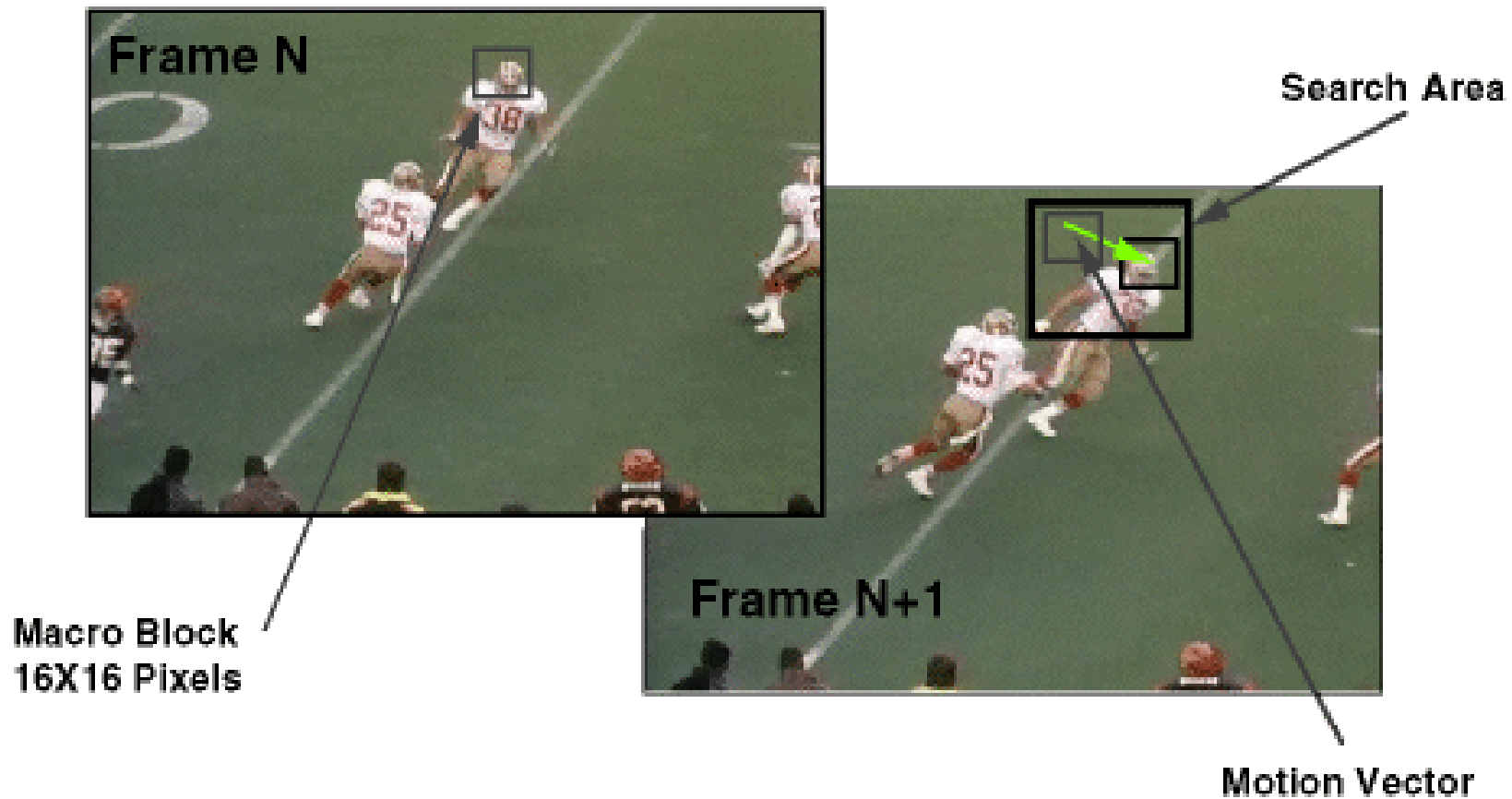


# STANDARDS ET NORMES VIDÉOS

- le **MPEG-1**, développé en 1988, est un standard pour la compression des données vidéos et des canaux audio associés (jusqu'à 2 canaux pour une écoute stéréo). Il permet le stockage de vidéos à un débit de 1.5Mbps dans une qualité proche des cassettes VHS sur un support CD appelé VCD (*Vidéo CD*).
- le **MPEG-2**, un standard dédié originalement à la télévision numérique (*HDTV*) offrant une qualité élevée à un débit pouvant aller jusqu'à 40 Mbps, et 5 canaux audio surround. Le MPEG-2 permet de plus une identification et une protection contre le piratage. Il s'agit du format utilisé par les DVD vidéos.
- le **MPEG-4**, un standard destiné à permettre le codage de données multimédia sous formes d'objets numériques, afin d'obtenir une plus grande interactivité, ce qui rend son usage particulièrement adapté au Web et aux périphériques mobiles.
- le **MPEG-7**, un standard visant à fournir une représentation standard des données audio et visuelles afin de rendre possible la recherche d'information dans de tels flux de données. Ce standard est ainsi également intitulé *Multimedia Content Description Interface*.
- le **MPEG-21**, en cours d'élaboration, dont le but est de fournir un cadre de travail (en anglais *framework*) pour l'ensemble des acteurs du numérique (producteurs, consommateurs, ...) afin de standardiser la gestion de ces contenus, les droits d'accès, les droits d'auteurs, ...

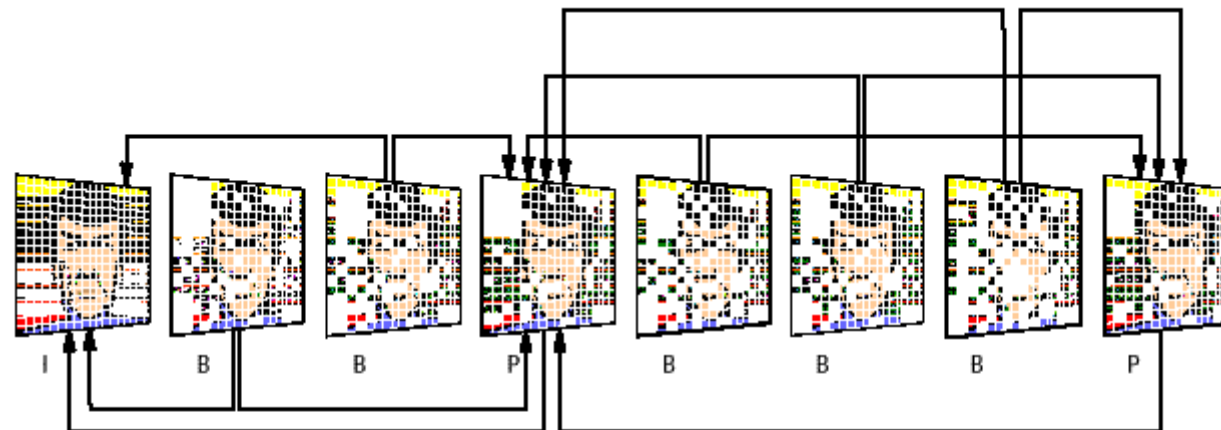
# LE STANDARD MPEG

- Estimation du mouvement:



# LE STANDARD MPEG

- Les trames I, P et B:



- **I-frames: contain full picture information**
- **P-frames: predicted from past I or P frames**
- **B-frames: use past and future I or P frames**
- **Transmit I frames every 12 frames or so.**

## FRAMES I :

Ces images sont codées uniquement en utilisant le codage JPEG, sans se soucier des images qui l'entourent.

De telles images sont nécessaires dans une vidéo MPEG car ce sont elles qui assurent la cohésion de l'image (puisque les autres sont décrites par rapport aux images qui les entourent), elles sont utiles notamment pour les flux vidéo qui peuvent être pris en cours de route (télévision), et sont indispensables en cas d'erreur dans la réception.

Il y en a donc une ou deux par seconde dans une vidéo MPEG.

## FRAMES P :

- Ces images sont définies par différence par rapport à l'image précédente. L'encodeur recherche les différences de l'image par rapport à la précédente et définit des blocs, appelés macroblocs (16x16 pixels) qui se superposeront à l'image précédente.
- L'algorithme compare les deux images bloc par bloc et à partir d'un certain seuil de différence, il considère le bloc de l'image précédente différent de celui de l'image en cours et lui applique une compression JPEG.
- C'est la recherche des macroblocs qui déterminera la vitesse de l'encodage, car plus l'algorithme cherche des "bons" blocs, plus il perd de temps...  
Par rapport aux frames-I (compressant directement), les frames-P demandent d'avoir toujours en mémoire l'image précédente.

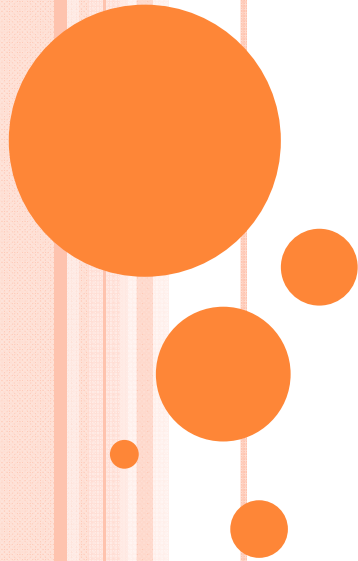
## FRAMES **B** :

- De la même façon que les frames P, les frames B sont travaillées par différences par rapport à une image de référence, sauf que dans le cas des frames B cette différence peut s'effectuer soit sur la précédente (comme dans les cas des frames P) soit sur la suivante, ce qui donne une meilleure compression, mais induit un retard (puisque'il faut connaître l'image suivante) et oblige à garder en mémoire trois images (la précédente, l'actuelle et la suivante).

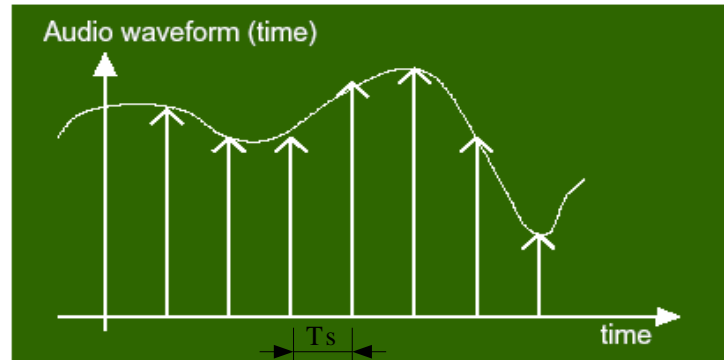
## FRAMES D :

- Ces images donnent une résolution de très basse qualité mais permettent une décompression très rapide, cela sert notamment lors de la visualisation en avance rapide car le décodage "normal" demanderait trop de ressources processeur.
- Dans la pratique :
  - Afin d'optimiser le codage MPEG, les séquences d'images sont dans la pratique codées suivant une suite d'images I, B, et P (D étant réservé à l'avance rapide) dont l'ordre a été déterminé expérimentalement. La séquence type appelée *GOP* (*Group Of Pictures* ou en français *groupes d'images*) est la suivante: `IBBPBBPBBPBBBI`
  - Une image I est donc insérée toutes les 12 frames.

# LE SON



# REPRÉSENTATION NUMÉRIQUE DU SIGNAL (ENCODAGE CD)



## → Définitions

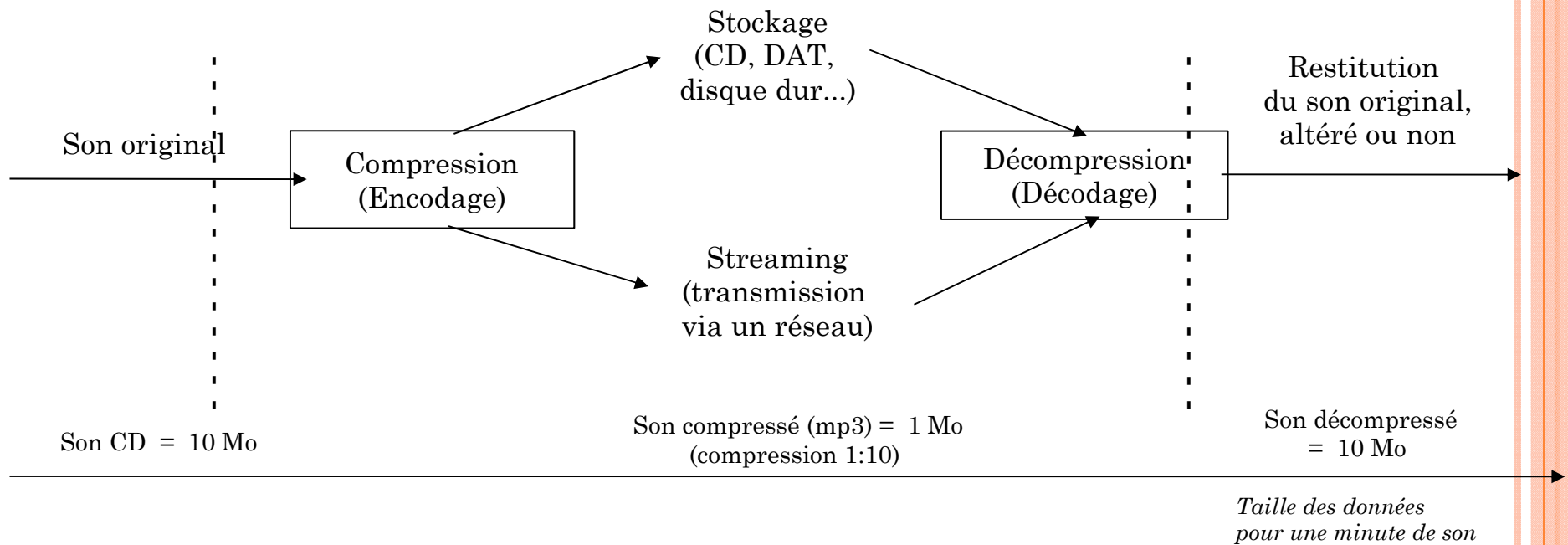
- **Fréquence d'échantillonnage** (*sampling rate*) :  $F_s = 1/T_s$ .  
= nombre d'échantillons prélevés pour une seconde du signal.
- **Résolution** = nombre de bits utilisés pour coder un échantillon,
- **Débit** (*bit rate*) = nombre de bits utilisés pour coder 1 seconde de son.

## ○ Qualité CD :

- $F_s = 44100$  Hz (44100 éch/s de son)
- Résolution 16 bits
- Stéréo

$$\left. \begin{array}{l} - F_s = 44100 \text{ Hz (44100 éch/s de son)} \\ - \text{Résolution 16 bits} \\ - \text{Stéréo} \end{array} \right\} 44100 * 16 * 2 = \mathbf{1.41 \text{ Mbits/sec}}$$

# LA COMPRESSION AUDIO



*Processus de compression/décompression audio*



# DEUX MODES DE COMPRESSION AUDIO

## 1) Compression non destructive (sans perte) :

= le signal reconstruit est exactement identique au signal original

## 2) Compression destructive (avec perte) :

= le signal reconstruit est différent du signal original mais cette différence est imperceptible/peu perceptible par l'oreille humaine.

= permet une meilleure réduction de la taille des données

= au détriment de la qualité

=> **compromis taille du fichier/qualité audio**



## LA COMPRESSION AUDIO

- But : réduire la taille des données représentant un son original, sans dégrader la perception de ce son.
- Encodage des composantes fréquentielles (analyse par bancs de filtres)
- La compression se base sur des modèles psychoacoustiques (psychoacoustique = analyse de la perception auditive humaine).
- La compression s'effectue en supprimant de l'information :
  - Suppression de l'information redondante
  - Suppression de l'information "secondaire" du point de vue perceptif ("perceptually irrelevant")

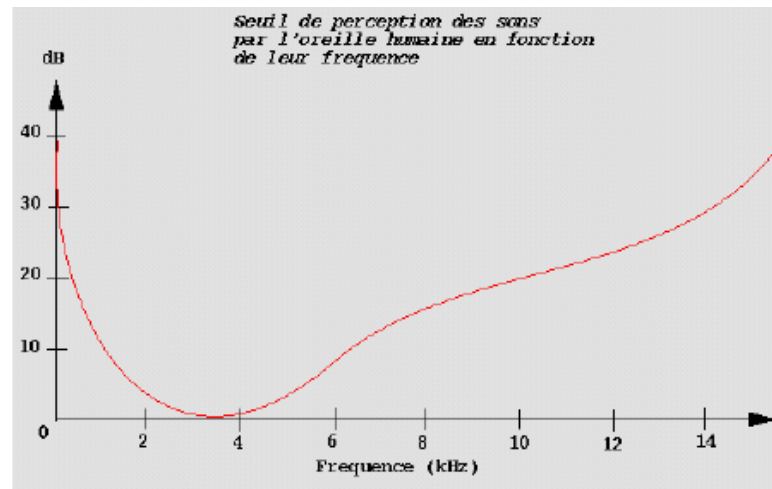
## MODÈLES PSYCHOACOUSTIQUES

- Les sons que l'on retrouve dans la nature sont complexes (constitués de nombreuses composantes fréquentielles).
- L'oreille humaine ne perçoit pas chacune de ces composantes de la même façon, certaines sont prépondérantes pour la perception globale d'un son alors que d'autres ne sont quasiment pas entendues.
- L'encodeur va ainsi supprimer les composantes les moins déterminantes pour la perception par l'oreille humaine.



## LE SEUIL D'AUDITION

- Seuil d'audition = niveau sonore à partir duquel un son est détecté  
=> dépend de la fréquence et de la durée du son



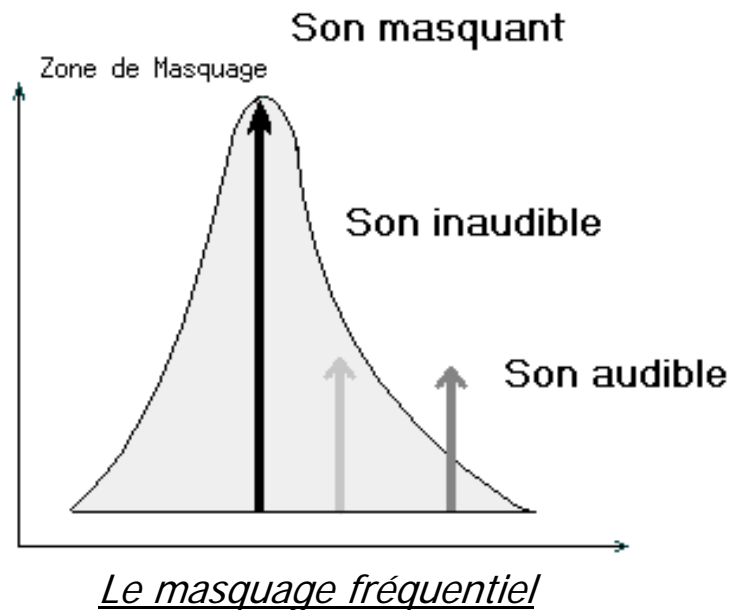
*Exemple d'audiogramme humain*  
(Seuil d'audition en fonction de la fréquence du son)

- L'encodeur supprime les composantes fréquentielles :
  - dont le niveau est inférieur au seuil d'audition correspondant.
  - dont la durée est trop courte pour être perçues.



## LE MASQUAGE FRÉQUENTIEL

- Masquage fréquentiel: baisse d'audibilité d'un son causée par la présence simultanée d'un autre son de fréquence proche.

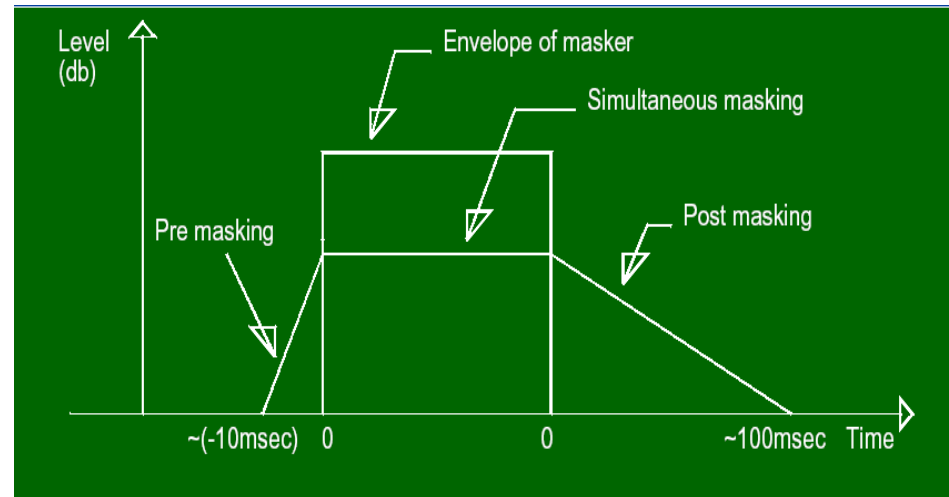


=> compression audio : suppression des fréquences masquées



## LE MASQUAGE TEMPOREL

- Masquage temporel (non simultané): baisse d'audibilité d'un son causée par la présence antérieure ou postérieure d'un autre son.



*Le masquage temporel*  
*(non simultané)*

=> compression audio : suppression des fréquences masquées



## EVALUATION DE LA COMPRESSION AUDIO

- Indices de performance d'un algorithme de compression :
  - Taux de compression/bitrate
  - Qualité du son restitué (indice perceptif)
  - Délai algorithmique (streaming, applications temps réel)

Bitrate	Qualité	Taux de compression	Taille pour 1 minute de son
1,4 Mbits/s	CD Audio	Aucune	10 Mo
192 kbits/s	CD Audio conservée	T= 1:7	1,37 Mo
128 kbits/s	Qualité limite	T= 1:11	915 Ko
96 kbits/s	Moyenne	T= 1:15	703 Ko
64 kbits/s	Mauvaise	T= 1:22	469 Ko

*Evaluation de la compression mp3  
à différents bitrates*

Mp3 "classique"

# LES DIFFÉRENTS FORMATS DE COMPRESSION AUDIO

## ➤ MPEG1- Layer 3 (mp3)

- 1er format de compression (1992) => plus de 19 ans !!
- Aujourd'hui dépassé en terme de performances
- Avantage : le plus utilisé (Internet, Hifi, baladeurs...)
- Mp3 pro (2001): version améliorée du mp3 (préservation des HF)

## ➤ WMA (format Microsoft, 1994)

- Meilleure qualité audio que mp3 à bitrate équivalent
- Encodage plus rapide
- Protection contre le piratage (DRM)

## ➤ AAC (Format audio du MPEG-2 et MPEG-4, 1997)

- Le plus performant de tous (qualité CD à un débit de 96 kbits/s)
- Spatialisation (encodage possible sur 48 canaux)

## ➤ OGG Vorbis (Format ouvert, "Linux" de la compression audio)

- Proche de AAC en terme de performance



## LE FORMAT MIDI

- Musical Instrument Digital Interface (MIDI) : format d'échange destiné aux instruments numériques (synthétiseurs, contrôleurs...).
- Encode une description des actions du musicien, pas le son en lui-même  
=> **MIDI n'est pas un format de compression audio !!**
- Représentation note par note d'un extrait musical. Chaque note est décrite par un chiffre qui donne sa hauteur, sa durée et son intensité.
  - => MIDI format très léger (quelques Ko pour plusieurs minutes de musique)
  - => Représentation paramétrique du son (cf. MPEG-4 Audio)



# MPEG-4 AUDIO

DE L'ENCODAGE

À LA COMPOSITION DE SCÈNES SONORES



# HISTORIQUE MPEG-AUDIO

1992 : MPEG-1 Audio => 3 qualités disponibles :

- Layer 1 : taux de compression 1:4 (384 kb/s en stéréo)
- Layer 2 : taux de compression 1:8 (192 kb/s en stéréo)
- Layer 3 : taux de compression 1:10 (128 kbps en stéréo)

=> **MP3** = MPEG-1 Layer-3

1994 : MPEG-2 Audio : extension de MPEG-1 Audio, compression multicanal avec des débits plus faibles.

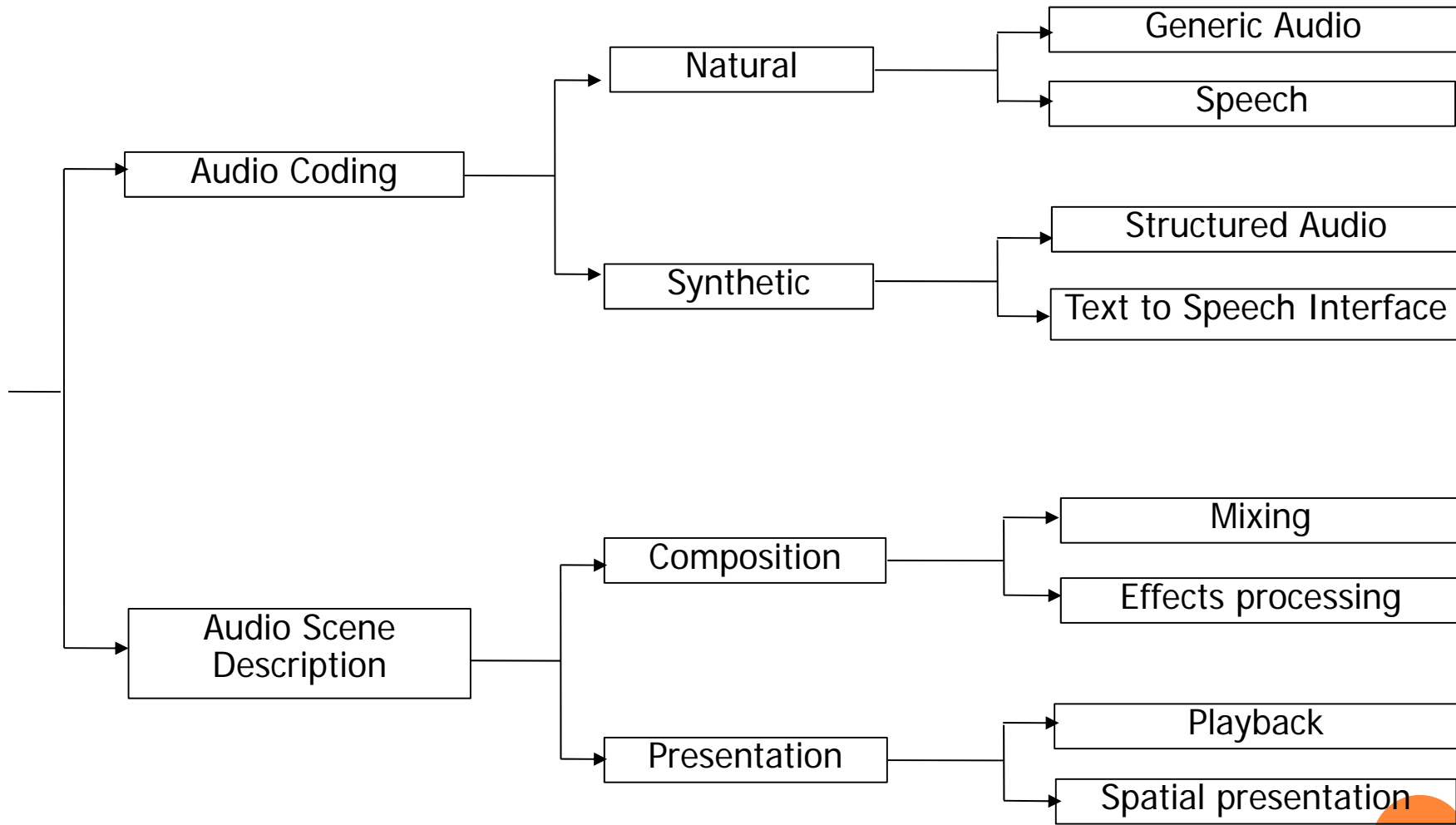
1997 : MPEG-2 Audio Advanced Coding (AAC), encodeur complètement nouveau, compression deux fois plus efficace.

1999 : MPEG-4 Audio version 2 => extension de AAC +ajout de nouvelles fonctionnalités pour la composition de scènes sonores.

# INTRODUCTION MPEG-4 AUDIO

- MPEG-4 Audio = **description d'une scène audio** :
  - Décomposition de la scène en objets sonores
  - Intégration dans la scène d'objets "naturels" ou synthétiques
  - Adaptation de la méthode de codage selon la nature du son (parole, musique...)
  - Fonctions de traitements des objets sonores (effets)
  - Présentation spatialisée de la scène audio
  - Reproduction des effets de salles (réverbération) et de la directivité des sources

# MPEG-4 AUDIO TOOLS



## NATURAL CODING

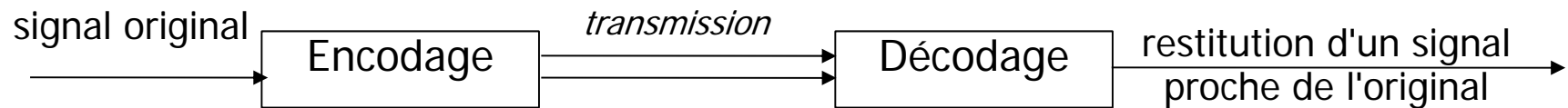
- › Version optimisée de Advanced Audio Coding (AAC, MPEG-2)
- › Utilisation de méthodes de codage spécifiques adaptées à la nature du son à encoder (parole, son musical...)
- › Compression plus efficace (jusqu'à 4kbits/s), meilleure qualité audio
- › Scalability : possibilité d'adapter le bitrate, utiliser différents bitrates au cours de la transmission



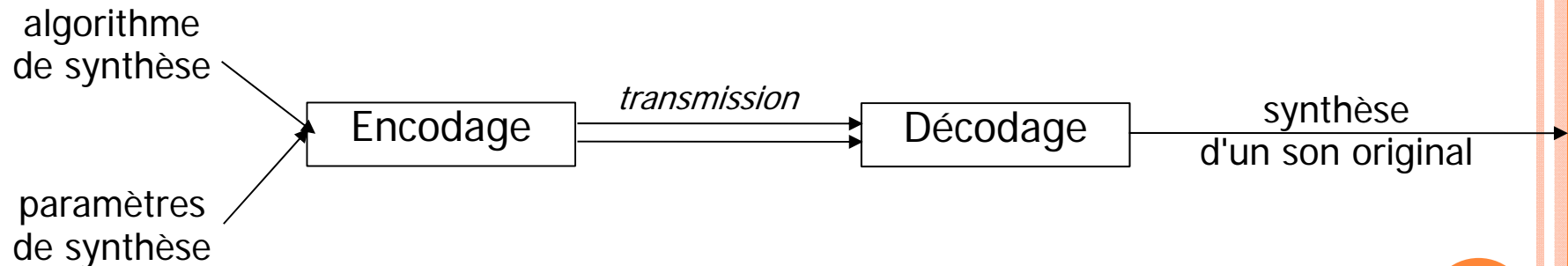
# SYNTHETIC CODING

› Démarche totalement différente du codage "naturel" :

› Natural Coding :

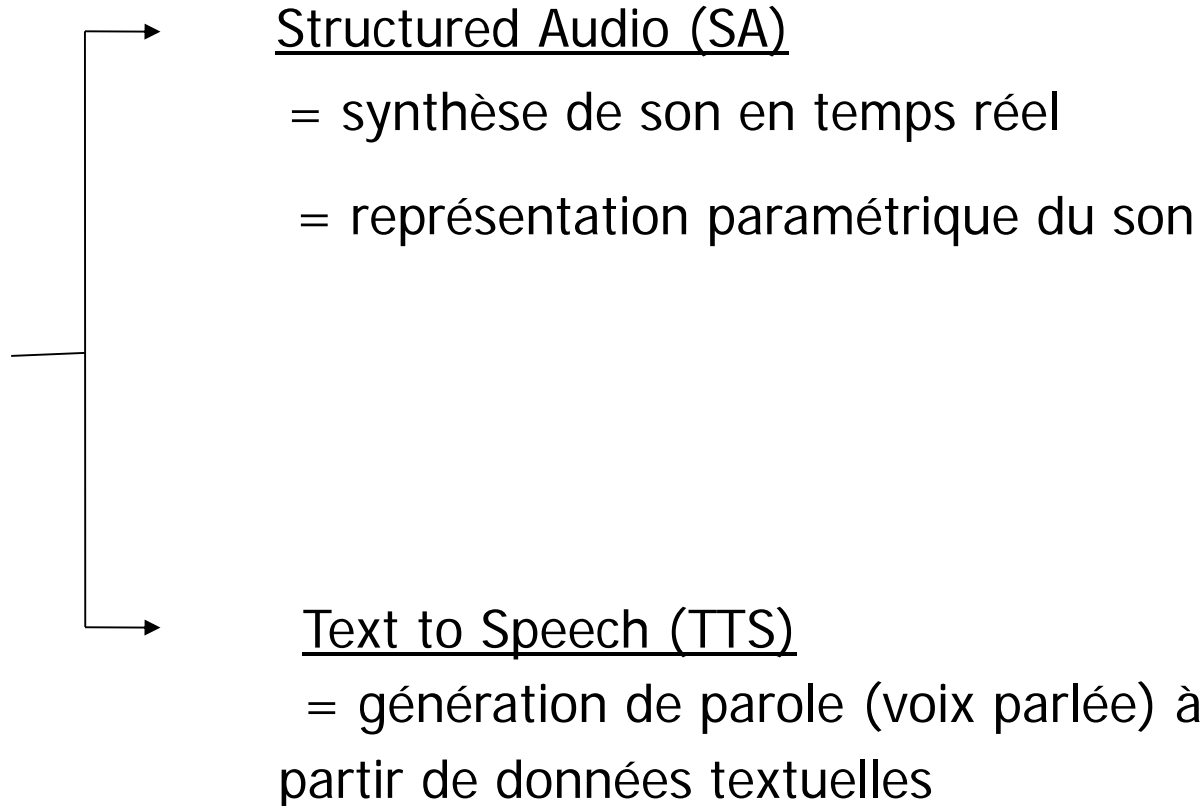


› Synthetic coding



=> **représentation paramétrique du son**

# SYNTHETIC CODING



# Structured Audio (SA)

- Un son est généré à partir d'un algorithme à qui sont fournis des paramètres pour la synthèse.
- Extension de CSound (langage de synthèse et traitements sonores)
- Distinction Orchestre / Partition :
  - Structured Audio Orchestra Language (SAOL)  
= algorithmes de synthèse (ex : filtrages...)
  - Structured Audio Score Language (SASL)  
= paramètres de la synthèse (ex : fréquences de coupure des filtres...)

## TEXT TO SPEECH INTERFACE (TTSI)

- Génération de la parole à partir de données textuelles
  - Utilisation à très bas débit : 200 bits/s à 1.2 kbits/s
  - Spécification de paramètres additionnels pour la synthèse :
    - informations sur le locuteur (genre, age, vitesse d'allocution)
    - paramètres prosodiques ("ligne mélodique" de la voix parlée)
    - paramètres liés à l'animation faciale (mouvement des lèvres...)
- => amélioration du rendu sonore (meilleure intelligibilité)



# AUDIO SCENE DESCRIPTION

## Composition

= mixage des objets sonores  
dans la scène (multicanal)

= traitements sonores  
(postprocessing)

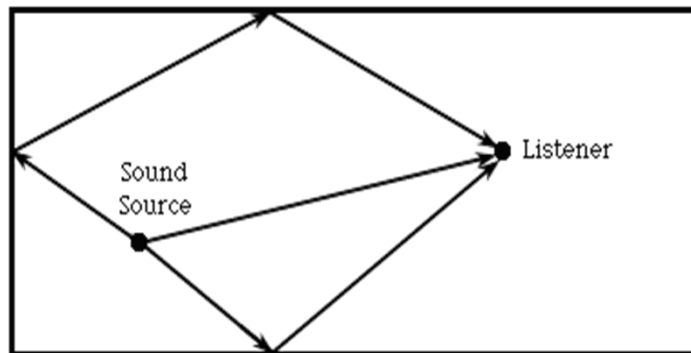
## Présentation

= Reproduction spatialisée de la  
scène (enceintes, headphones)

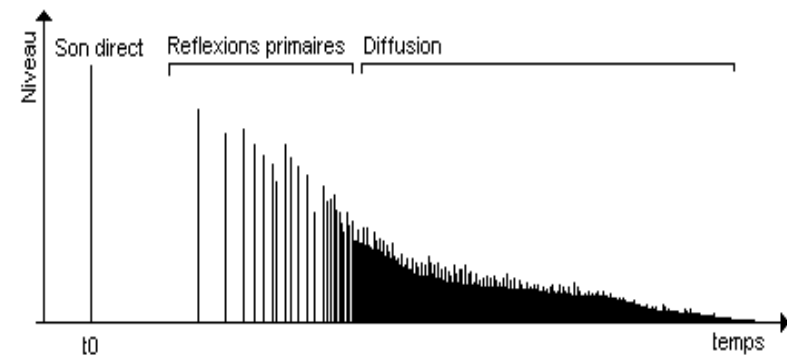
AudioBIFS  
& AdvancedAudioBIFS

# AUDIO BIFS

- Traitements : filtrage, effets (delay, chorus, flanger...)
- Spatialisation 2D et 3D (position des objets sonores dans la scène, position de l'auditeur, distance source/auditeur)
- Reproduction de la réverbération naturelle



Phénomène de réverbération



Décomposition  
du champ réverbéré



## ADVANCED AUDIO BIFS

- Specification de paramètres additionnels pour améliorer le réalisme du rendu de la sonore:
  - Propriétés acoustiques du matériau (reflectivité et transmission sur les surfaces)
  - Propagation du son (directivité des sources, absorption de l'air, effet Doppler...)
  - Attributs perceptifs (chaleur, brillance...)



## CONCLUSION MPEG-4 AUDIO

- Haut niveau de performance de compression (grande qualité audio à très bas débits)
- Fonctionnalités de **synthèse** et traitement sonores
- Description de **scènes sonores** (décomposition en objets sonores, reproduction de l'espace sonore)
- Nombreuses applications : réalité virtuelle, jeux vidéos, web, ...
- Avenir : MPEG-7, MPEG-21...



# MP3

## QUELLE TECHNIQUE DE COMPRESSION?

- **Ensemble de techniques**
  - **Utilise les défauts de l'oreille humaine**
    - **Compression destructive : supprime les sons que nous n'entendons pas**
  - **Compression non destructive**
    - **Huffman**

# MP3

## L'EFFET DE MASQUE

- **Suppression des sons faibles  
« cachés » par des sons plus forts**
- **Exemple :**
  - Le pépiement des oiseaux couvert par le passage d'un avion dans le ciel**
  - ⇒ **On entend plus les oiseaux**
  - ⇒ **Son des oiseaux inutile**
  - ⇒ **Son des oiseaux supprimé**

# MP3

## LE RESERVOIR D'OCTETS

- **Permet d'encoder certains passage complexes sans altérer la qualité (ou de façon moindre)**
- **Utilisation de passages qui peuvent être encodé à un taux inférieur au autres**

# MP3

## LE JOINT STÉRÉO

- **Défaillance de l'oreille :**

**Ne peut localiser l'origine du son en dessous d'une certaine fréquence**

**⇒ Son codé en mono (au lieu de stéréo)**

# MP3

## CODE HUFFMAN

- **Algorithme de codage qui agit à la fin de la compression**
  - **Code les fréquence selon leur nombre d'apparition**
  - **Codes de longueur variable**
  - **Permet un gain de ~ 20%**
  - **Efficacité accrue sur les sons « purs » ou sons digitalisés**

# MP3

## RÉSULTAT & ALTERNATIVES

- **Une minute d'un CD-audio (à une fréquence de 44.1 kHz, 16 bits, stéréo) ne prendra qu'un seul Mo**
- **Limite quantitative et qualitative du MP3**
  - ⇒ Apparition de nouveaux formats
    - MP3 Pro
    - AAC (Advanced Audio Coding)
    - Ogg Vorbis
    - Windows Media Audio
    - *Vqf*

# Ogg Vorbis

- Comparable à MP3
- Compression avec perte
- Débit entre 30 et 500 kbits/s
- Format libre : [www.vorbis.com](http://www.vorbis.com)
- Compression à débit variable
- Taux de compression > MP3
- Mais nécessite un traitement plus complexe

# FLAC

- Compression sans perte
- Format libre
- Encodage et décodage rapide
- Compression d'environ 60%