

# Principes de Compression vidéo

## Plan

- Principes de la vidéo analogique
- Principes de la Vidéo Numérique
- Principes de Compression
- Algorithmes de Compression

# Compression Vidéo

## Pourquoi et Comment ?

Taille des fichiers vidéo numérique très important excède les capacités de stockage et de communication :

Exemple pour 2 heures de TV numérique CCIR 601 :

- 149 Go de stockage requis
- 166 Mbp/s de bande passante

La solution est la compression des images par la :

1. Réduction de la résolution chromatique
2. Réduction de la taille de l'image
3. Réduction de la fréquence d'image
4. Eliminer les informations inutiles et/ou redondantes



## Vidéo Analogique

- Chaque image est représentée par une fluctuation d'un signal électrique qui prend la forme d'une onde électromagnétique.
- Bande passante requise : 6-8 Mhz
- Composition d'un signal vidéo :
  - information représentant la couleur (chromatique)
  - information représentant la luminosité (luminance)
  - information sonore (audio)
    - Vidéo analogique + Audio numérique NICAM (32 Khz) possible
  - information de synchronisation
- On distingue :
  - Vidéo Composite (la plus utilisée) : TV hertzienne
  - Vidéo en Composante : Vidéo VHS, 8 mm, ...



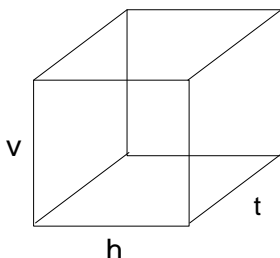
# Vidéo Analogique/Numérique

	Gamme de fréquence	Type d'utilisation
	10 kHz - 150 kHz	Communications radiotélégraphiques
	150 kHz - 300 kHz	Radiodiffusion (grandes ondes)
	510 kHz - 1605 kHz	Radiodiffusion (petites ondes)
	6 MHz - 20 MHz	Radiodiffusion (ondes courtes)
	29,7 MHz - 41 MHz	Radiotéléphonie
VHF	47 MHz - 68 MHz	Télévision
	68 MHz - 87,5 MHz	Liaisons radio en modulation de fréquences
	87,5 MHz - 108 MHz	Radiodiffusion
VHF	108 MHz - 162 MHz	Radiotéléphonie
	162 MHz - 216 MHz	Télévision
	216 MHz - 470 MHz	Radiotéléphonie
UHF	470 MHz - 860 MHz	Télévision et radar
	860 MHz - 960 MHz	Radiotéléphonie
	Autour de 1800 MHz	Radiotéléphonie
SHF Ka, Ku	Entre 6 et 30 GHz	Services satellites en fixe



# Vidéo Numérique

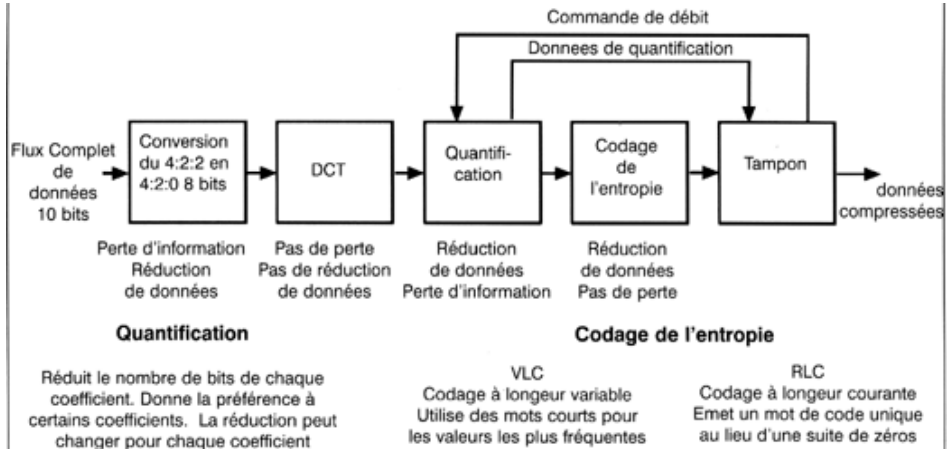
- Objectif :
  1. préserver la qualité lors de copies multiples
  2. faciliter le stockage, la manipulation et la transmission
- Principe : numériser chacun des signaux vidéo analogiques  
(1) échantillonner, (2) quantifier, (3) coder



- Échantillonnage de :
  - axes  $h$  et  $v$  : donne la résolution spatiale
  - axe  $t$  : donne la résolution temporelle
- Ratio d'aspect : 4/3 ou 16/9
- Quantification :
  - 13.5 Mhz et 6.75 Mhz (Vidéo)
  - 32 KHz (Audio)
- Codage de chaque pixel :
  - donne la résolution couleur (15-24 bits)



# Codeur Spatial Type JPEG



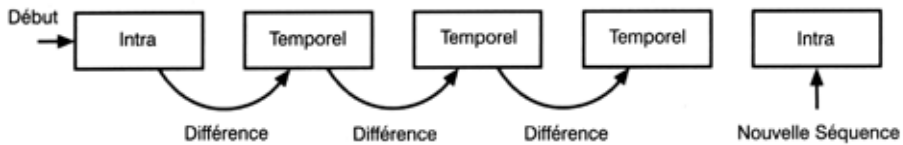
## Compression CBR et VBR

- **Codage CBR : Constante Bit Rate**
  - Débit binaire en sortie constant
  - Qualité visuelle variable
- **Codage VBR : Variable Bit Rate**
  - Débit binaire en sortie variable
  - Qualité visuelle constante
- Dans les systèmes à débit variable, la quantification est fixe mais, dans un système à débit constant, une mémoire tampon est utilisée pour absorber les variations intervenant au cours du codage.
- Si la mémoire tampon est menacée de saturation, les pas de quantification devront être agrandis de façon à atteindre le facteur de compression voulu.

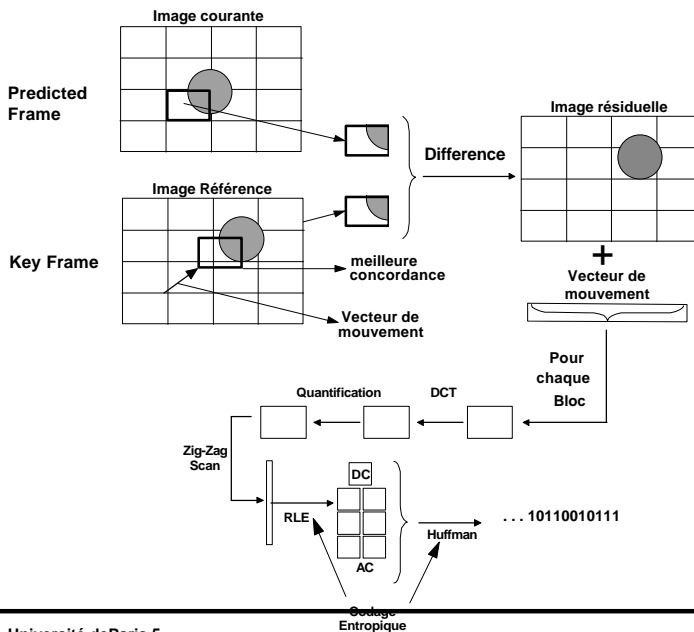


# Codeur Temporelle Type MPEG

- La redondance temporelle peut être exploitée par inter-codage, en transmettant uniquement les différences entre images successives.
- La figure suivante montre que des images complètes sont transmises périodiquement. Elles sont appelées images de références ou « intra-codées » (images I), et elles sont uniquement obtenues par compression spatiale.
  - S'il se produit une erreur ou si un changement de scène intervient, on retrouve un décodage correct à l'image I suivante.



## Codeur temporel type MPEG



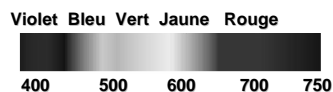
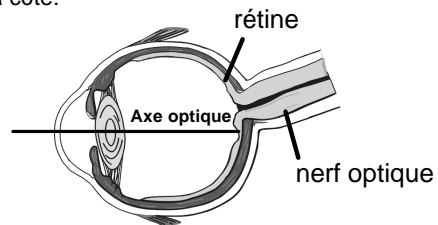
# Erreur de l'estimation de mouvement



# Systeme visuel humain

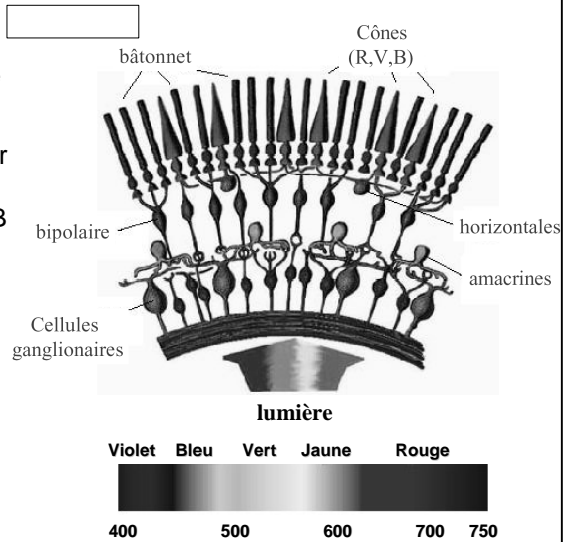
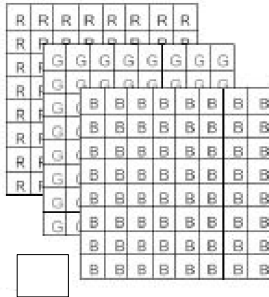
La rétine est une mince couche tapissant le fond de l'œil et constituée de 2 sortes de cellules disposées côte à côte:

- cellules à bâtonnet (vision nocturne)
  - sensibles à l'intensité lumineuse
  - nombreuses : 120 millions
- cellules à cône (vision diurne)
  - Peu sensibles à la luminosité
  - sensibles aux couleurs
  - peu nombreuses : 6 millions
  - se répartissant en trois populations en fonction du pigment qu'elles renferment;
  - ont une réaction électrique maximale quand on les éclaire avec des radiations bleue (435 nm), vert (547 nm) ou rouge (700 nm) ;
  - L'existence de ces 3 types de pigments rend compte du caractère trichrome de la vision humaine et justifie, a posteriori l'utilisation du modèle RVB/RGB.



# Systeme visuel humain (2)

- Les longueurs d'onde de la lumière visible par l'œil sont comprises entre 400 (violet) et 750 (rouge) nm.
- Une image peut être représentée par la superposition de 3 matrices de valeurs d'intensité des couleurs RGB



## Particularité du SAH

- Le système audio visuel humain (SAH) :
  1. est plus sensible aux variations de luminosité qu'aux couleurs,
    1. Choisir un format de représentation des couleurs optimisé (YUV)
  2. ne peut distinguer que 4 millions de nuances de couleurs,
    1. Sous échantillonner les signaux de couleurs
  3. Requiert un rafraîchissement des images > 40 Hz
    1. Utiliser la technique de l'entrelacement des images
  4. Ne perçoit pas les détails dans les zones peu lumineuses
  5. Sensible aux basses fréquences et peu aux hautes fréquences
    1. Représenter les pixels par des coefficients de fréquences (transformé de Fourier)
    2. Appliquer ensuite un filtre numérique (La quantification)



# Formats Chromatique

- Différents formats de couleurs:
  - RGB : le plus répandu – employé par les ordinateurs
    - L'utilisation d'un triplet d'entiers associés à l'intensité de chacune des couleurs primaires (Red, Green, Blue) définit une nuance.
    - 8:8:8 = 24 bits / pixels : 16.7 millions de couleurs
    - 5:5:5 = 15 bits / pixels : 32 768 couleurs
  - YUV et YCbCr : format répandu dans la TV analogique et numérique
    - Dérivé du format RGB
    - mieux représenter la luminosité que les couleurs,
    - Réduction de la bande passante



## RGB vs YUV



Originale



R



Y



G



U



B

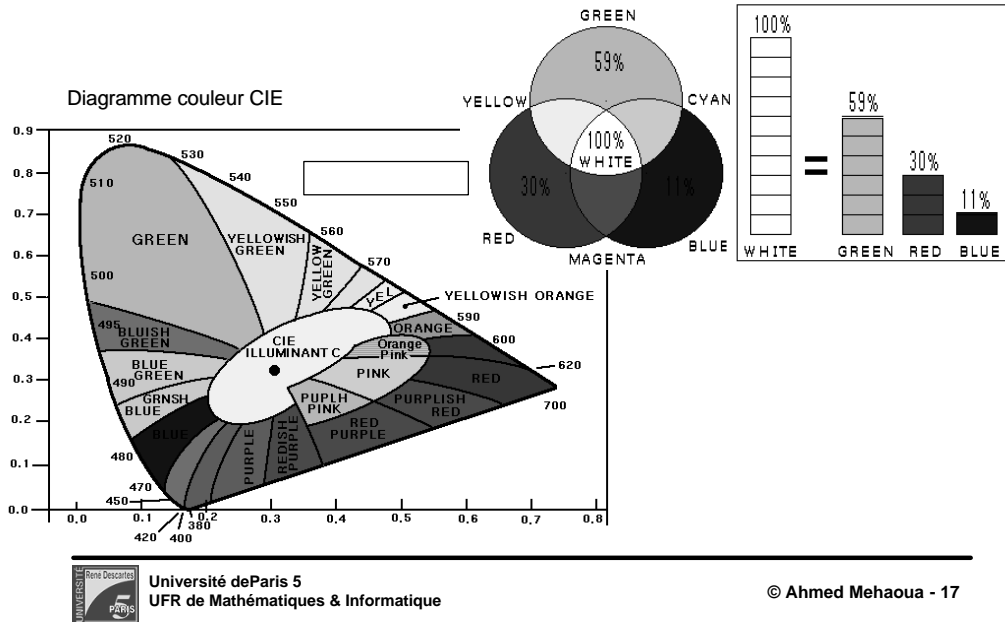


V





# Le format RGB



## Le format YUV - Télé Analogique -

- le format YUV est une transformation par rapport au format RGB où les composantes de chrominance B et R sont corrigées par rapport à la composante de luminance Y de la manière suivante :
  - $Y = 0.299R + 0.587G + 0.114B$
  - $U = -0.147R - 0.289G + 0.436B = 0.492(B - Y)$
  - $V = 0.615R - 0.515G - 0.100B = 0.877(R - Y)$
- Ce format est directement utilisé par les CD-I et la TV analogique PAL (Phase Alternation Line) /SECAM (Systeme Electronique Couleur Avec Memoire).
- Le système TV américain NTSC (National Television Systems Committee) ainsi que les vidéos numériques utilisent des variantes de YUV: YIQ et YCbCr respectivement.

# Le format YCbCr

## - Télé numérique -

Objectif :

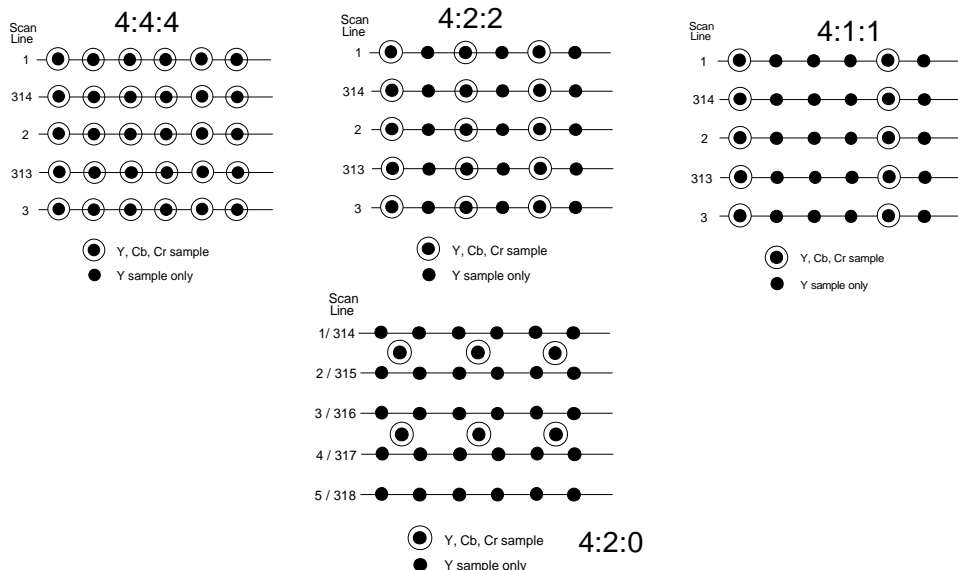
- Faire converger les trois systèmes de TV analogique vers un format numérique UNIQUE.
- Offrir différentes résolutions couleurs pour différentes applications vidéo numériques (DV Motion-JPEG, CDV MPEG-1, SVCD/DVD MPEG2, VidéoConférence H.261 et H.263)

Comment :

- En sélectionnant des facteurs de pondérations optimisées pour la vidéo numérique en composante.
  - $Y = 0.299R + 0.587G + 0.114B$
  - $Cb = 0.169R - 0.331G + 0.500B = 126(B - Y) + 128$
  - $Cr = 0.500 - 0.419G - 0.081B = 160(R - Y) + 128$
- Définir différents ratios d'échantillonnage: 4:N1:N2
  - YCbCr 4:1:1 (DV NTSC)
  - YCbCr 4:2:2 (DVB)
  - YCbCr 4:2:0 (TVHD, DVD, CDV, DV PAL, Video-conférence)



## Les formats YCbCr (2)



# Réduction Chromatique

4:2:0	Y + Cr + Cb = 48 bits	= 32 bits (4 pixels x 8 bits/pixel) + 8 bits (4 pixels partageant 8 bits) + 8 bits (4 pixels partageant 8 bits)
4:1:1	Y + Cr + Cb = 48 bits	= 32 bits (4 pixels x 8 bits/pixel) + 8 bits (4 pixels partageant 8 bits) + 8 bits (4 pixels partageant 8 bits)
4:2:2	Y + Cr + Cb = 64 bits	= 32 bits (4 pixels x 8 bits/pixel) + 16 bits (2 pixels partageant 8 bits) + 16 bits (2 pixels partageant 8 bits)
4:4:4	Y + Cr + Cb = 96 bits	= 32 bits (4 pixels x 8 bits/pixel) + 32 bits (4 pixels x 8 bits/pixel) + 32 bits (4 pixels x 8 bits/pixel)

## Gain :

- 48 bits Total / 4 pixels = 12 bits/pixel (4:1:1 et 4:2:0) 50 %
- 64 bits Total / 4 pixels = 16 bits/pixel (4:2:2) 33 %
- 96 bits Total / 4 pixels = 24 bits/pixel (4:4:4) 0 %



# Réduction Temporelle

- L'oeil humain conserve une copie de l'image durant une courte période après l'avoir interprétée : c'est la période de persistance de la vision.
- Cette propriété est exploitée en vidéo pour transcrire le mouvement en présentant à intervalle régulier plusieurs images fixes;
- Si l'intervalle de présentation des images fixes est supérieure à la période de persistance de la vision, alors l'image scintille.
- On considère qu'une fréquence  $< 40$  images/sec alors scintillement;
- la période de persistance de la vision humaine est inversement proportionnelle à la luminosité de l'image;
  - plus les images sont lumineuses, plus il faut réduire la période de présentation (ou autrement dit accroître la fréquence);

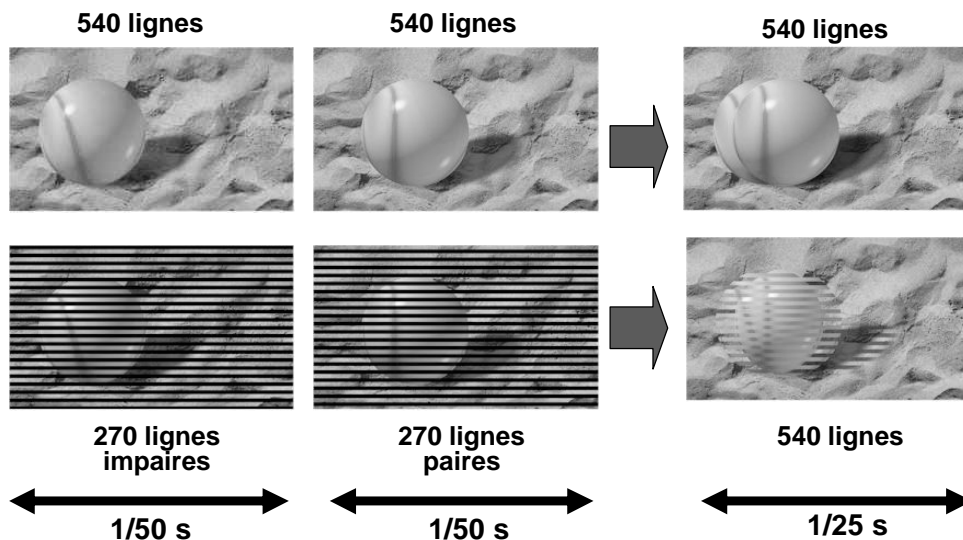


# Réduction Temporelle

- Fréquence d'images mesurée en Hz
  - Cinéma : 24 images seconde
  - TV européenne : 25 images seconde (50 Hz)
  - TV US/Japon : 30 fps (60 Hz)
  - Informatique : 60 - 85 Hz
- Solutions :
  - vidéo entrelacée (TV)
    - découpage de l'images en sous-bandes (trames paires et impaires)
    - Puis transmission et affichage des trames impaires suivie des trames paires
    - 1. Règle le problème de la persistance de la vision en doublant la fréquence d'affichage (des trames);
    - 2. Permet de réduire par 2 la bande passante des canaux de transmission TV
  - vidéo progressive (PC - Cinéma)
    - Affichage ligne par ligne (de haut en bas) sans entrelacement
    - Utilisation d'un doubleur de fréquences (48 ou 72 images /sec)



## Vidéo Progressive vs Entrelacée



# Réduction spatiale

## Les formats vidéo numériques

HDTV

TV numérique

SDTV

PARAMETER	SHD	SMPTE 240M	ITU-R BT.601 NTSC	ITU-R BT.601 PAL/SECAM
Pixels/Line (L)	>2048	1920	720	720
Lines/Image (L)	>2048	1152	484	575
Color Subsampling	4:4:4	4:2:0	4:2:2	4:2:2
Color Format	RGB	YCbCr		
Color Resolution (bpp)	24	24	16	16
Temporal resolution (fps)	>60	30/60	30	25
Raw bit rate (Mbit/s)	> 5700	506	166	166

DVD NTSC

DVD PAL



Université de Paris 5  
UFR de Mathématiques & Informatique

© Ahmed Mehaoua - 25

# Réduction spatiale

## Les formats vidéo numériques

VideoConférence

Video Disc

PARAMETER	SQCIF	QCIF	CIF	4CIF	16CIF	SIF NTSC	SIF PAL/SECAM
Active Pixels/Ligne (L)	128	176	352	704	1408	352	352
Active Lines/Image (L)	96	144	288	576	1152	240	288
Color Format	YCbCr				YCbCr		
Color subsampling	4:1:1				4:2:0		4:2:0
Color Resolution (bits per pixel)	8				8		8
Temporal Resolution	30				30		25
Raw bit rate (Mbit/s)	4.4	9.115	36.45	146.0	583.9	36.45	36.45

CIF : Common Intermediate Format  
SIF : Standard Image Format



Université de Paris 5  
UFR de Mathématiques & Informatique

© Ahmed Mehaoua - 26

# Conversion CCIR-601 -> SIF

720	Résolution horizontale
484	Résolution verticale
348 480	Nombre total de pixels par image
24	Nombre de bits par pixel (résolution)
8 363 520	Nombre total de bits par image
30	Nombre d'images par seconde
250 905 600	Nombre total de bits/seconde
1 048 576	Conversion en Mbps

**239.3** Débit requis en Mégabits/secondes !!!

352	Résolution horizontale (Format SIF)
240	Résolution verticale
84 480	Nombre total de pixels par image
8	Nombre de bits par pixel (résolution)
675 840	Nombre total de bits par image
15	Nombre d'images par seconde
10 137 600	Nombre total de bits/seconde
1 048 576	Conversion en Mbps
<b>9.67</b>	Débit requis en Mégabits/secondes

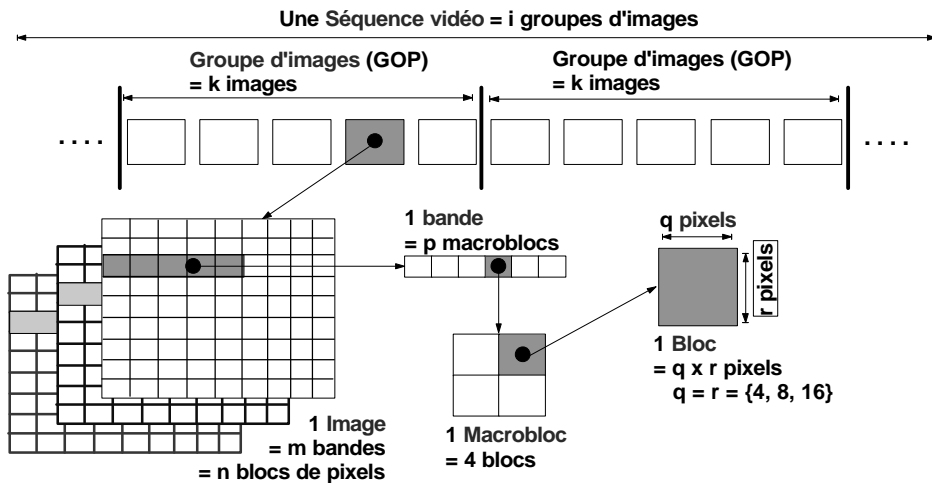


## Les algorithmes de Compression

- Redondances Spatiales (codage intra-image) :
  - codages prédictifs (DPCM, ADPCM)
- Redondances temporelles (codage inter-image)
  - estimation et compensation de mouvement
  - codages prédictifs (DPCM, ADPCM)
- Redondances statistiques
  - codage en entropie (Huffman, arithmétique)
- Redondances psycho-visuelle
  - Sous échantillonnage des couleurs
  - transformation (DCT, FFT)
  - quantification



# Structure des images

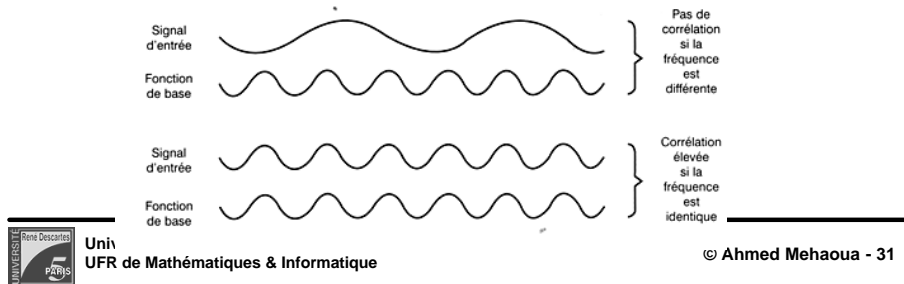


## Transformation

- En codage spatial, la première étape consiste à effectuer une analyse de fréquence spatiale à l'aide d'une transformée.
- Une transformée est un outil mathématique dont l'objectif est de transformer des informations spatio-temporelle (les pixels) issues d'un signal vidéo numérique vers un autre domaine (les fréquences) où les données seront plus structurées et donc susceptibles de faciliter le processus de compression.
- Le résultat d'une transformée est une suite de coefficients décrivant l'amplitude de chaque composante fréquentielle présente dans le signal.
- Cette technique n'est pas réellement une méthode de compression. Ces transformations sont réversibles et n'introduisent pas de perte d'information (conservative).

# Transformé de Fourier

- La transformée la plus répandue est la transformée de Fourier.
- Cette transformée cherche chaque fréquence du signal d'entrée
- Elle caractérise chaque fréquence comprise dans le signal d'entrée en multipliant le signal d'entrée par une fréquence cible appelée fonction de base, et en intégrant le produit obtenu.
- lorsque la forme de signal d'entrée ne contient pas de composante à la fréquence cible, l'intégrale sera nulle mais, s'il en comporte une, l'intégrale constituera un coefficient caractérisant l'amplitude de cette composante : c'est le coefficient de la transformation.



# Discret Cosinus Transform

- La DCT (*Discret Cosinus Transform*) est une version discrète de la transformé de Fourier et s'applique sur des blocs de pixels (8x8 en générale et de valeur comprise entre 0 et 255) et délivre un bloc de 64 coefficients DCT (de valeur comprise entre -1024 et +1024) suivant la formule de transformation bidimensionnelle suivante :

$$F(u, v) = \frac{1}{4} \cdot C(u) \cdot C(v) \cdot \sum_{i=0}^7 \sum_{j=0}^7 f(i, j) \cdot \cos\left(\frac{(2i+1) \cdot u \cdot p}{16}\right) \cdot \cos\left(\frac{(2j+1) \cdot v \cdot p}{16}\right)$$



# Transformation DCT

132	136	138	140	144	145	147	155
136	140	140	147	140	148	155	156
140	143	144	148	150	152	154	155
144	144	146	145	149	150	153	160
150	152	155	156	150	145	144	140
144	145	146	148	143	158	150	140
150	156	157	156	140	146	156	145
148	145	146	148	156	160	140	145

DCT

172	-18	15	-8	23	-9	-14	19
21	-34	24	-8	-10	11	14	7
-9	-8	-4	6	-5	4	3	-1
-10	6	-5	4	-4	4	2	1
-8	-2	-3	5	-3	3	4	6
4	-2	-4	6	-4	4	2	-1
4	-3	-4	5	6	3	1	1
0	-8	-4	3	2	1	4	0

- Basses Fréquences
- Moyennes Fréquences
- Hautes Fréquences



# DCT



1 coef. DCT  
par bloc  
(AC)



3 coef. DCT  
par bloc  
(AC + 2 DC)



6 coef.  
par bloc  
(AC + 5 DC)



64 coef. DCT  
par bloc  
(AC + 63 DC)



# Quantification (suite)

7842	199	448	362	342	112	31	22
198	151	181	264	59	37	14	3
142	291	218	87	27	88	27	12
111	133	159	119	58	65	36	2
49	85	217	50	8	3	14	12
58	120	60	40	41	11	2	1
30	121	61	22	30	1	0	1
22	28	2	33	24	51	44	81

Coefficients d'entrée de la DCT  
(un bloc très complexe)

980	12	23	16	13	4	1	0
12	9	8	11	2	1	0	0
7	13	8	3	0	2	0	1
5	6	6	4	2	1	0	0
2	3	8	1	0	0	0	0
2	4	2	1	1	0	0	0
1	4	2	1	0	0	0	0
0	0	1	0	0	0	0	0

Coefficients de sortie de la DCT  
valeurs indicatives pour l'affichage  
et non les résultats réels

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	48	56	69
27	29	35	38	46	56	69	83

Valeurs de la matrice de quantification  
Les valeurs utilisées correspondent  
à la location du coefficient

Code	Linéaire	Non-Linéaire
1	2	1
8	16	8
16	32	24
20	40	40
24	48	56
28	56	88
31	62	112

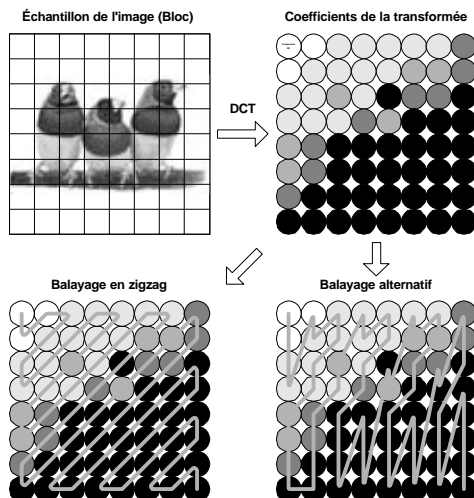
Valeurs de l'échelle de quantification  
Toutes les valeurs de code ne sont pas montrées  
Une valeur est utilisée pour l'ensemble du block 8 X 8

Division par la  
matrice de  
quantification

Division par  
l'échelle de  
quantification

# Lecture en Zig Zag

- Après quantification, les coefficients DCT de faible valeur (hautes fréquences) seront tronqués à zéro.
- On obtiendra une transmission plus efficace si on émet d'abord tous les coefficients non nuls et qu'un code indique ensuite que tous les autres sont à zéro (rôle de l'étape suivante de codage statistique).
- La lecture en Zig Zag constitue une technique qui augmente la probabilité d'obtenir ce résultat car elle émet les coefficients dans l'ordre probable d'amplitude décroissante.



# Codage Statistique

- Malgré la Lecture en Zig-Zag, des termes nuls apparaîtront encore au milieu des coefficients non nuls. Le codage statistique permet de gérer plus efficacement ces coefficients.
- Le codage en entropie (Entropy coding) est une technique de codage statistique qui attribue des mots de code aux valeurs des pixels en tenant compte de certaines de leurs propriétés statistiques (occurrence, régularité, groupement, ...).

Dans cette catégorie on retrouve :

1. le Run-Length Encoding (RLE), et
2. les codages en longueur variable (VLC Variable Length Coding) tels que le codage de Huffman.



## Run Length Encoding

- Les flux vidéo contiennent très souvent des groupes ou blocs de pixels qui possèdent la même valeur, par exemple dans des zones uniformes de l'image. RLE se propose de remplacer chacun de ces blocs par une valeur unique de pixel (jeton) associée au nombre de fois que cette valeur apparaît (*run-length*).
- Ce couple de nombres (valeur, nombre de répétitions) ou (*jeton, run-length*) est séparé par un drapeau qui ne fait pas partie intégrante du flux vidéo. Par exemple le codage de la séquence de pixels '555559992222222000077711' pourra être remplacé par la paire de valeurs '(5,6) (9,3) (2,8) (0,5) (7,3) 11' qui ne requière plus que 12 nombres plutôt que 27 et le flux binaire sera alors représenté par '5!6 999 2!8 0!5 777 11' soit 17 mots au lieu de 27.
- On constate avec cet exemple que le codage ne s'applique qu'aux valeurs apparaissant **plus de trois fois** consécutivement.



# Codage de Huffman

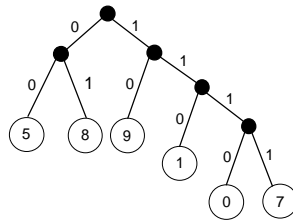
- Codage de Huffman a la propriété d'être optimal car il donne la plus petite moyenne de longueur de code de toutes les techniques de codage statistique
- Par exemple la séquence d'entiers soumise à un codage de Huffman générera la suite binaire qui **minimise la hauteur moyenne de l'arbre**  $\sum_{i=0}^{n-1} p_i n_i$ .

Code initiale : '8888888881111555555555999999997755000'

Résultat dérivé de l'arbre binaire et de la table de VLC ci-dessous : '01 01 01 01 01 01 01 110 110 110 110 110 00 00 00 00 00 00 00 10 10 10 10 10 10 10 1110 1110 00 00 1111 1111 1111'

- La taille des données passe ainsi de 152 bits à 90 bits avec un codage effectué sur 4 bits max, soit un gain de 59 %.

7	2
0	3
1	4
9	8
8	9
5	12

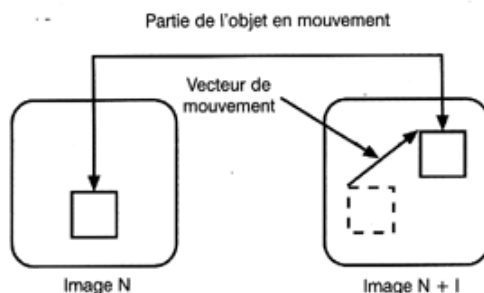


Valeur entière	Code binaire
5	00
8	01
9	10
1	110
0	1110
7	1111



# Estimation de mouvement

- Le mouvement réduit la similitude entre deux images et augmente la quantité de données nécessaires à la création de l'image de différence.
- L'estimation de mouvement est utilisée pour accroître cette similitude. Ce schéma en montre le principe.
- Une image ainsi codée par rapport à une autre est dite **Prédite** ou (image P)
- L'estimation de mouvement n'est calculée que pour la matrice de luminance (pas pour les 2 autres matrices de chrominance),

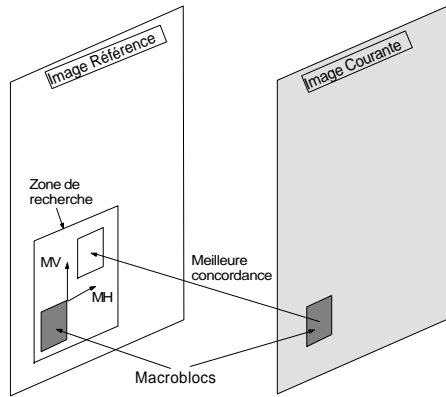


## Actions

1. Calcul du vecteur de mouvement
2. Déplacement des données de l'image N à l'aide du vecteur de mouvement pour construire l'image prédictee N + 1
3. Comparaison de l'image courante avec l'image prédictee
4. Emission de vecteur et de l'erreur de prédiction



# Estimation de Mouvement



Mean Squared Error (MSE) :

$$M_1(i, j) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (X_{m,n} - X_{m+i, n+j}^R)^2 \quad \text{avec } |i| \leq m_2, |j| \leq n_1$$

# Erreur de l'estimation de mouvement



# Vecteurs de mouvement

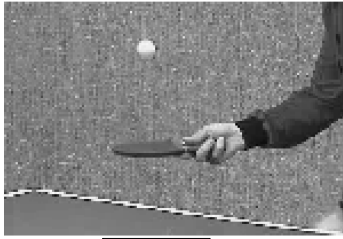


image n

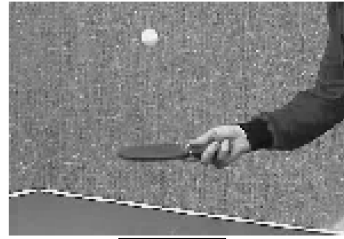
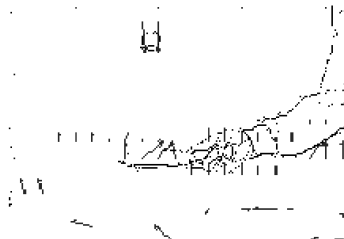


image n+1

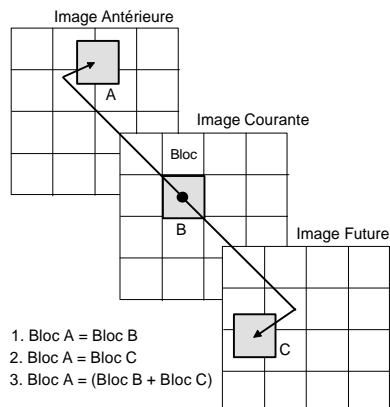


Vecteurs de mouvement

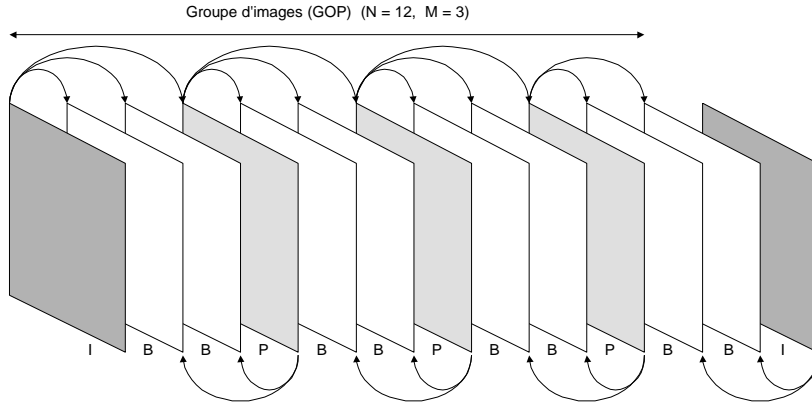


# Estimation de mouvement

- L'estimation de mouvement peut être généralisée à trois images successives. Ce schéma en montre le principe.
- Une image ainsi codée par rapport à deux autres images (une passée et une à venir) est dite bi-directionnellement Prédite ou (image B)



# Groupe d'Images



# Compression : Conclusion

Simple coding	Subsampling	
	Troncation	
	Lookup Table	
	Interpolation	
Entropy coding	Run-Length Encoding (RLE)	
	Huffman coding	
	Arithmetic coding	
Source coding	Transformation	FFT
		DCT
	Prediction	DPCM
		Delta Modulation
		Motion estimation
		Motion compensation
	Layered coding (Subband coding)	
Quantization		
Vector Quantization		
Hybrid coding	JPEG	
	H.261	
	MPEG1	
	MPEG2	
	H.263	

