

Département d'informatique et de génie logiciel
Compression de données
IFT-4003/IFT-7023

Notes de cours
Préface et introduction

Édition Hiver 2012

Mohamed Haj Taieb

Local: PLT 2113

Courriel: mohamed.haj-taieb.1@ulaval.ca

Faculté des sciences et de génie
Département de génie électrique et de
génie informatique



Plan

- Préface et introduction:
 - Importance de la compression
 - Exemples historiques de compression
 - Compression sans perte
 - Compression avec perte
 - Modélisation et codage des données
 - Étude des données pour établir un modèle

Pourquoi a-t-on besoin de la compression? (1)

Techniques de transmissions ne cessent d'évoluer.

- **Fibre optique:** communications à très longue distance: plusieurs centaines, voire milliers, de kilomètres. La capacité de transmission **de 1 à 10 Gigabits par seconde.**



- **Asymmetric digital subscriber line (ADSL):** débit ascendant 3.3 Mbit/s pour un débit descendant de 24.0 Mbit/s .

Pourquoi a-t-on besoin de la compression? (2)

Techniques de de stockage de plus en plus performantes offrant des volumes et des espace mémoire de plus en plus larges.



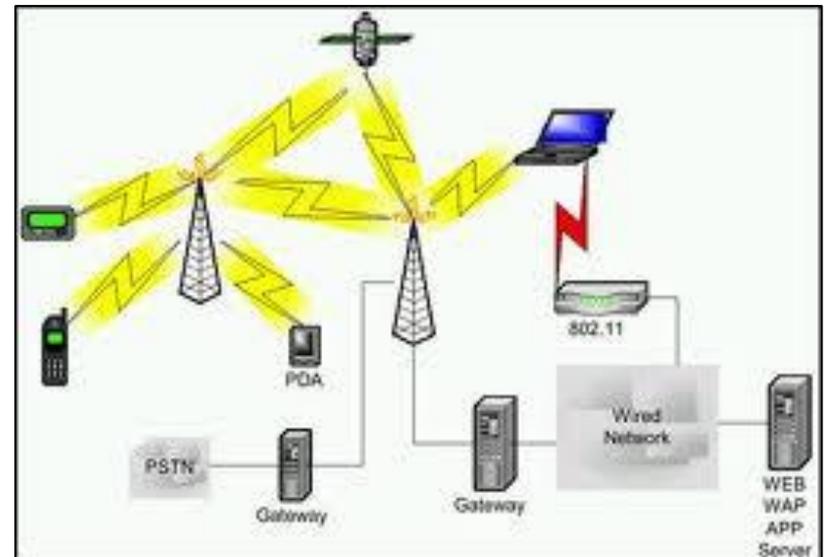
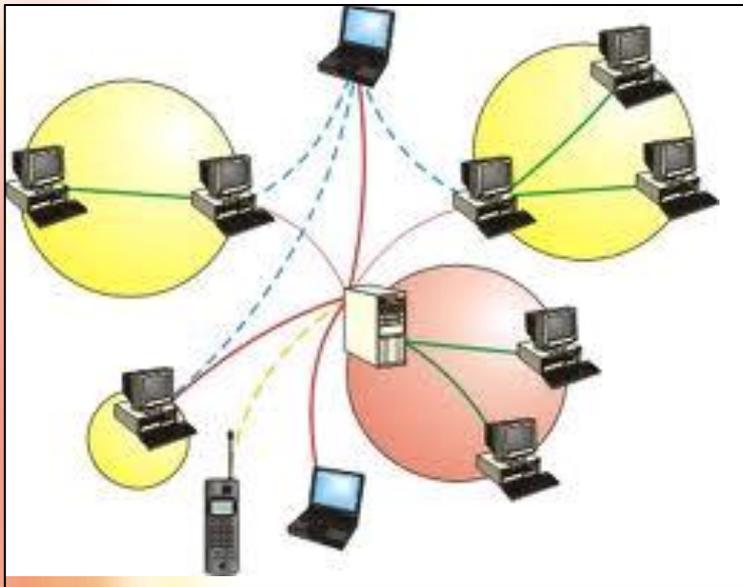
Question: Si les techniques de transmission et de stockage évoluent exponentiellement pourquoi compresser les données ?

Réponse: Les données sont générées de façon encore plus rapide.

Transmission des données

- Révolution dans le domaine de la communication via internet et via le réseau cellulaire.
- Transfert de la voix, échanges de données, communication vidéo dans des canaux à capacité limitée.

➔ Il faut réduire la taille du contenu à véhiculer.



Taille des données

- 1 seconde d'une vidéo CCIR 601 sans compression (858 colonnes x 525 lignes, 50Hz, 8bits=1 octet par pixel).

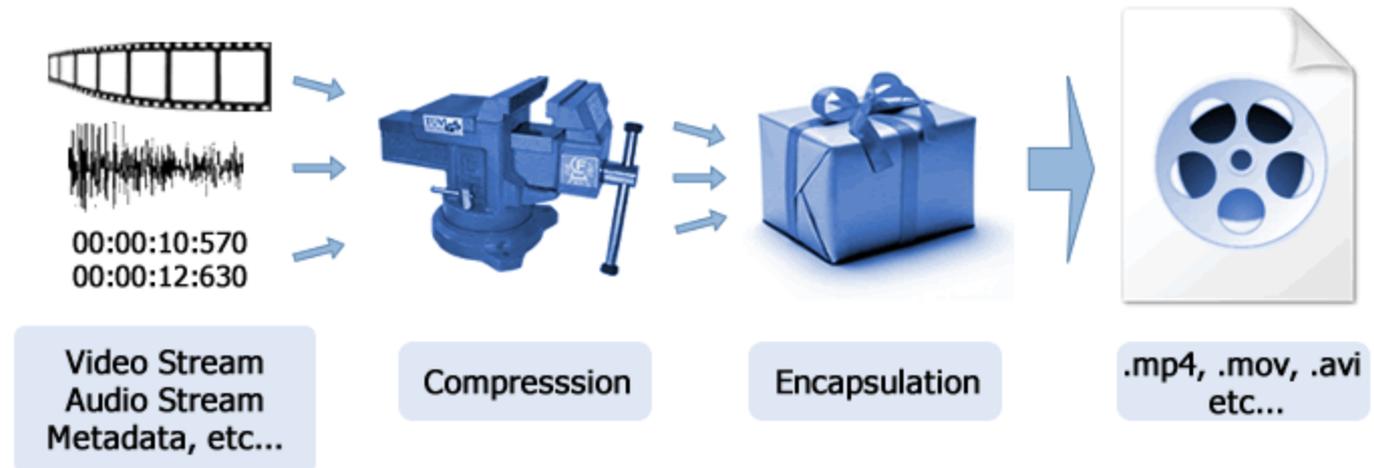
$858 \times 525 \times 50 = 22522500$ octets = 22 mega-octets = 176 megabits

1 heure = 3600×22 Mo = 80 Go.

CD-ROM: 700 Mo \rightarrow 114 CD-ROMs

DVD: 4.7 Go \rightarrow 17 DVDs

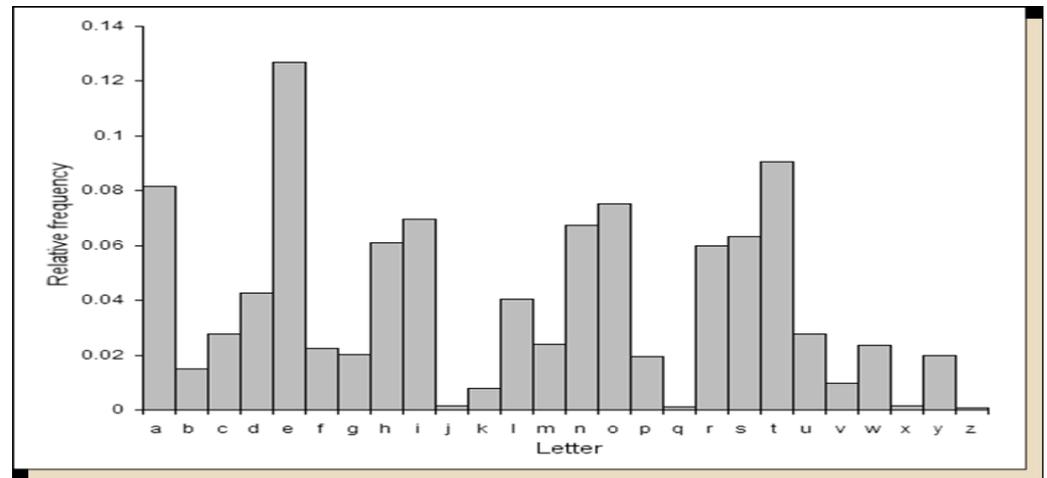
Nécessité de la
compression



Exemples historiques de compression

- Code de Morse [milieu du 19^{ème} siècle]:
- Les lettres les plus fréquentes ont un code court : e, a, t
- Les lettres moins fréquentes ont un code long: j, q, z
- Exploitation de la redondance statistique non optimale: l vs m

A	• —	N	— •
B	— • • •	O	— — —
C	— • • —	P	• — — •
D	— • •	Q	— — • —
E	•	R	• — •
F	• • — •	S	• • •
G	— — •	T	—
H	• • • •	U	• • —
I	• •	V	• • • —
J	• — — —	W	• — —
K	— • —	X	— • • —
L	• — • •	Y	— • — —
M	— —	Z	— — • •



Exemple: Thank you = -- -. -.- / -.- - - - ..-

Traducteur du code de Morse en ligne: <http://morsecode.scphillips.com/jtranslator.html>

Exemples historiques de compression

- Code de Braille[milieu du 19^{ème} siècle]:

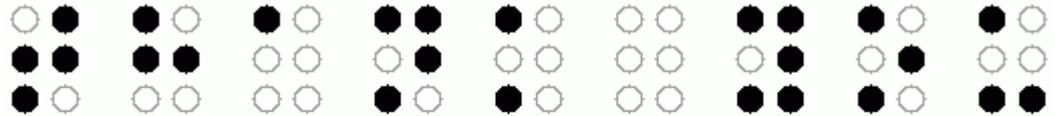
Caractère à 6 points= tableau de taille 2 x 3

Grade 1: chaque caractère a un tableau.

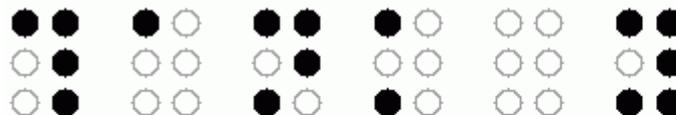
Grade 2: 2⁶ possibilités=64. Seulement 26 lettres. Les 38 combinaisons restantes représente des mots et des syllabes communément utilisés: and, you ...

A	B	C	D	E	F	G	H	I	J
K	L	M	N	O	P	Q	R	S	T
U	V	X	Y	Z	and	for	of	the	with
ch	gh	sh	th	wh	ed	er	ou	ow	W

Grade 1: Thank you =

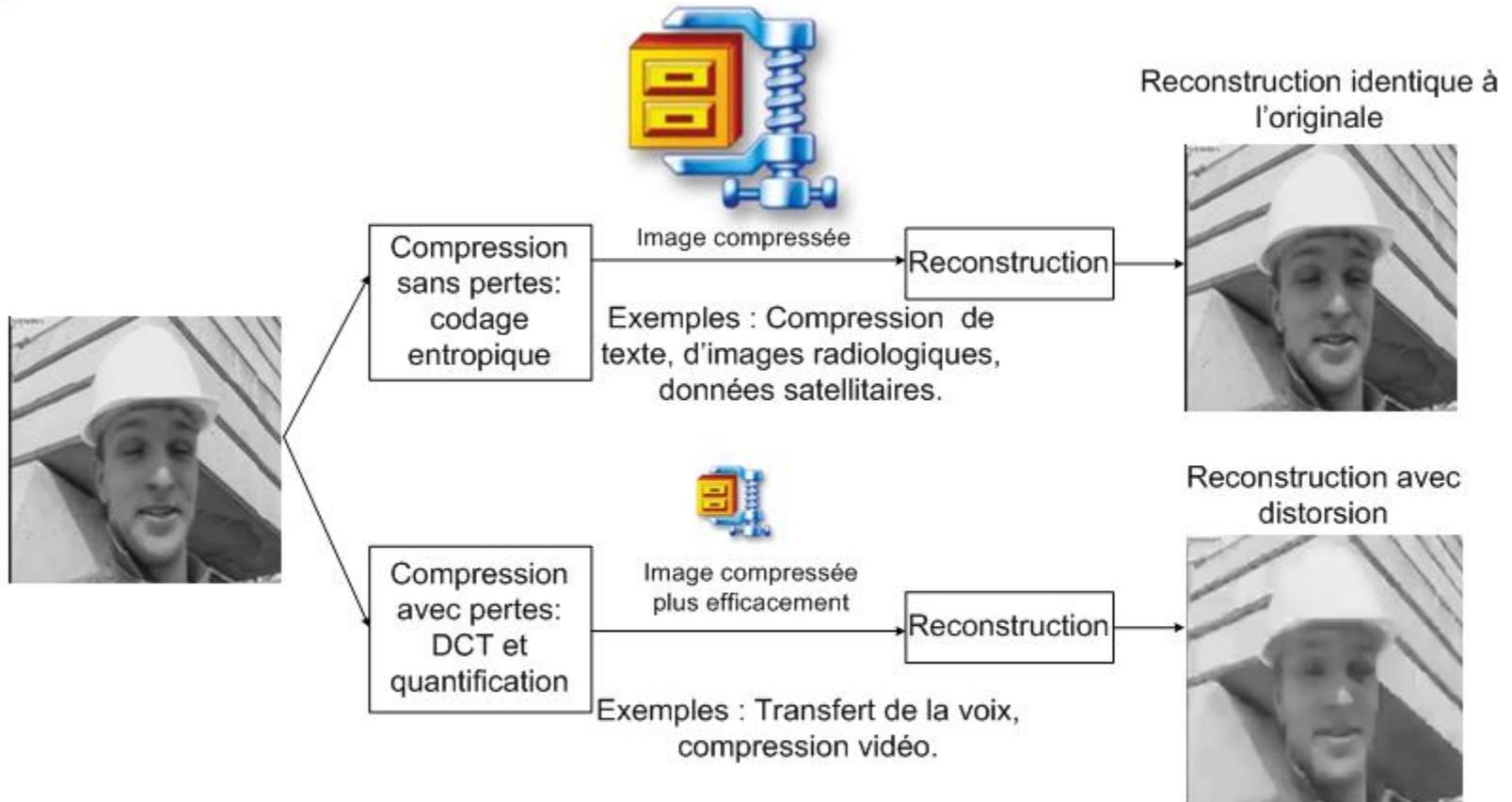


Grade 2: Thank you =



Traducteur de Braille en ligne: <http://www.mathsisfun.com/braille-translation.html>

Technique de compression



Performance d'un algorithme de compression

- Critères de performance d'un algorithme:
 - Complexité calculatoire: rapidité d'exécution et mémoire requise
 - Taux de compression=taille des données originale/tailles après compression
 - Ressemblance entre les données reconstruites avec l'originales
- Exemple [Taux de compression]
 - Image 256x256 pixels: 65 536 octets. Après compression: 16 384 octets.
 - Taux de compression est de 4:1 ou taux =75% ou taux= 2 bits/pixels

- Critères de distorsion objectifs:

$$MSE = \frac{1}{N} \sum_{n=1}^N x_n - \hat{x}_n$$

$$SNR = \frac{Var \ x}{MSE}$$

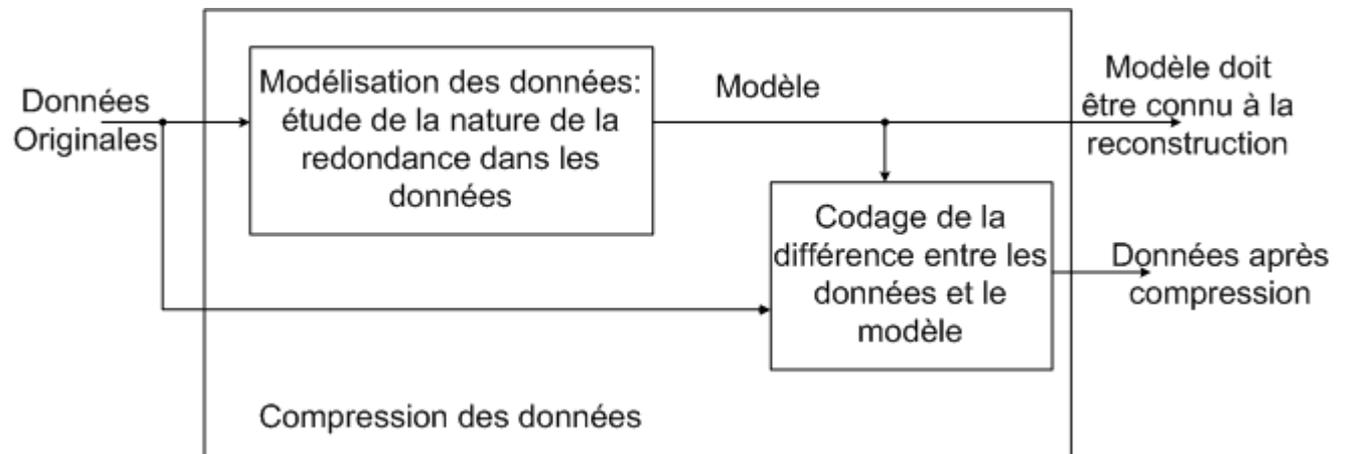
$$SNR_{dB} = 10 \log_{10} SNR$$

- Critères de distorsion subjectifs: basé sur la perception humaine.

$$PSNR = 10 \log_{10} \left(\frac{\max x^2}{MSE} \right)$$

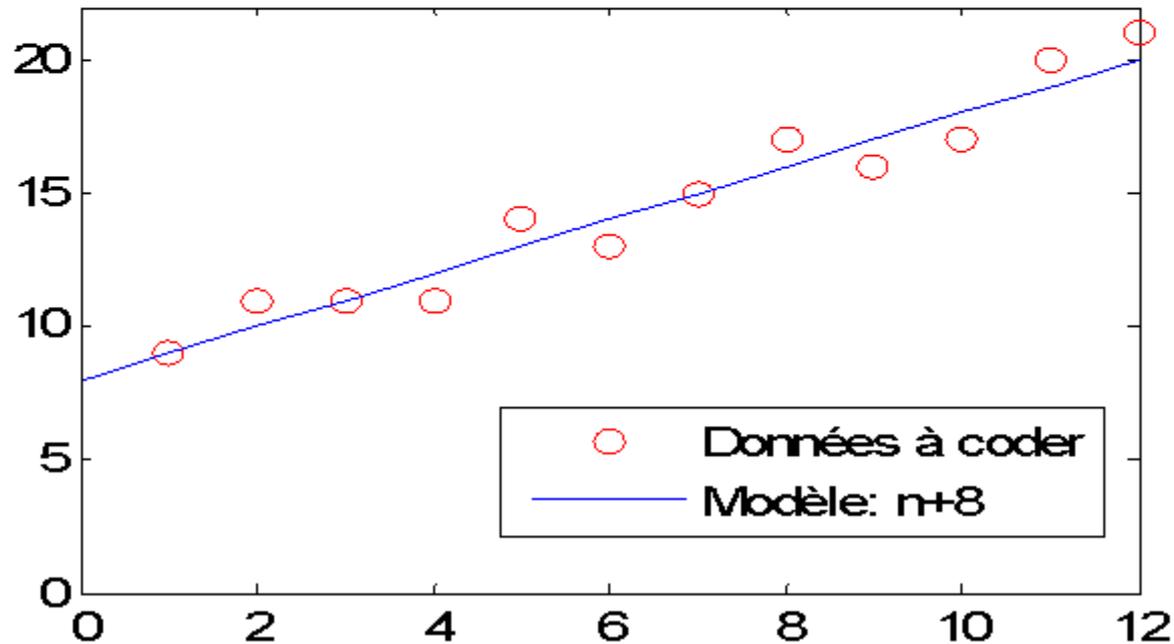
Modélisation et codage

- Exploitation des caractéristiques des données pour optimiser le schéma de compression.
- Les techniques de compression diffèrent selon les données: les techniques de compression du texte ne sont pas les mêmes techniques de compression des images.
- L'efficacité de la compression dépend de la redondance se trouvant dans les données.
- La nature de la redondance dans un fichier texte n'est pas la même que celle d'une image.



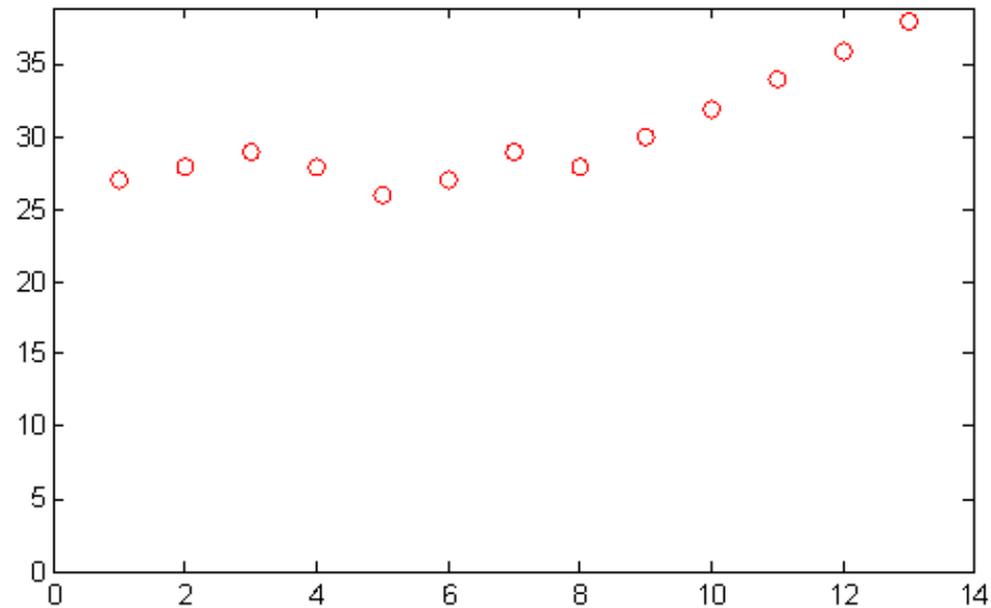
Exemples de modèle (1)

Données originales	9	11	11	11	14	13	15	17	16	17	20	21	→ 5 bits/échantillon
Estimation par le modèle	9	10	11	12	13	14	15	16	17	18	19	20	Taux de compression 5:2
Différence / résidu	0	1	0	-1	1	1	0	1	-1	-1	1	1	
Données codées	01	10	01	00	10	10	01	10	00	00	10	10	→ 2 bits/échantillon



Exemples de modèle (2)

Données originales	27	28	29	28	26	27	29	28	30	32	34	36	38
Estimation par le modèle	0	27	28	29	28	26	27	29	28	30	32	34	36
Différence / résidu	27	1	1	-1	-2	1	2	-1	2	2	2	2	2



Récapitulation

- La motivation pour le développement des techniques de compression: élimination de la redondance dans les données pour faciliter leur transfert et leur stockage.
- Présentation de quelques techniques anciennes de compression.
- Deux types majeurs de la compression selon le besoin et le contexte: compression sans perte et compression avec pertes.
- Les techniques de compression sont conçues après une étude des statistiques inhérents des données pour établir un modèle.
- On s'intéresse en premier lieu au technique de compression sans pertes.