

GBGI9U07: document multimédia : description et recherche automatique

B. Compression des images et des vidéos

Georges Quénot and Philippe Mulhem

Équipe Modélisation et Recherche d'Information Multimédia



Laboratoire d'Informatique de Grenoble



Février 2018

Le système visuel humain

- Trichromate : un type de bâtonnet (vision nocturne, luminosité faible) et trois types de cônes (vision diurne, luminosité moyenne ou élevée) : la distribution des fréquences lumineuses est échantillonnée suivant trois filtres,
- La couleur, les contours et la texture, et le mouvement sont traités par des sous-systèmes différents ayant des capacités différentes en termes de résolution spatiale et temporelles,
- Focalisé : la résolution est (vraiment) beaucoup plus forte dans la fovéa.
- Stéréoscopique, fonctionne par saccades, ...

Le système visuel humain

- La couleur est perçue avec une résolution spatiale inférieure aux contours et à la texture :
 - Les composantes de chrominance sont séparées des composantes de luminance et encodées suivant un échantillonnage plus faible (1 sur 2 ou 1 sur 4).
- La couleur et les contours et la texture sont perçues avec une résolution temporelle inférieure au mouvement :
 - L'image est décomposée en deux trames présentées alternativement : on dispose de la résolution maximale à la fréquence image (moitié de la fréquence trame) et de la fréquence maximale à la résolution trame (moitié de la résolution image) mais pas des deux à la fois.
- La focalisation n'est pas exploitée.

Compression d'images fixes JPEG :

- Découpage en blocs de 8 x 8 pixels,
- Transformée cosinus (fonction linéaire),
- Quantification suivant une table (pertes),
- Codage différentiel de la composante continue (DC),
- Séquence Zigzag pour les composantes fréquentielles (AC),
- Codage entropique,
- Utilisation de superblocs pour les informations de couleur.

Compression d'images fixes JPEG :

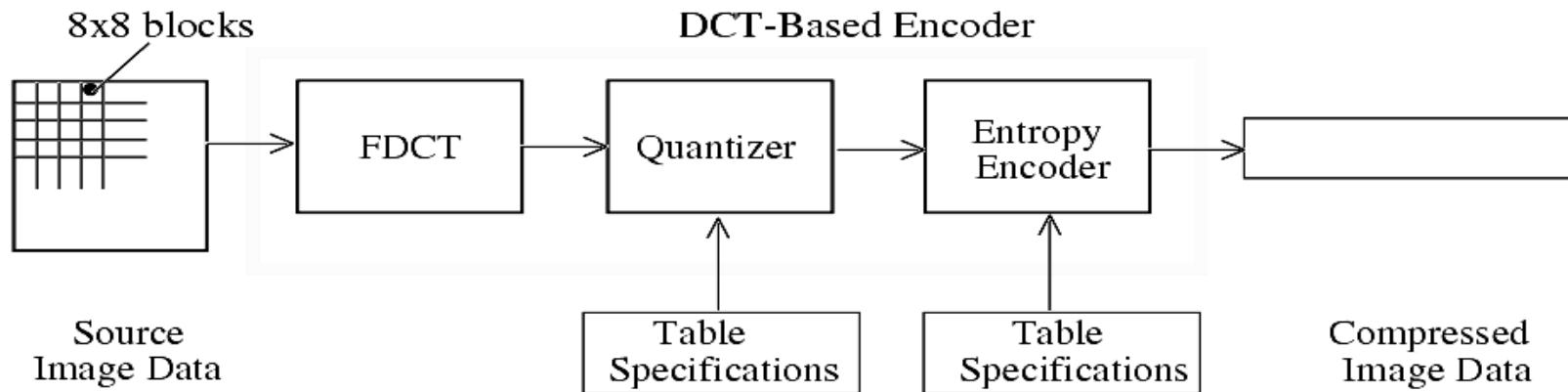


Figure 1. DCT-Based Encoder Processing Steps

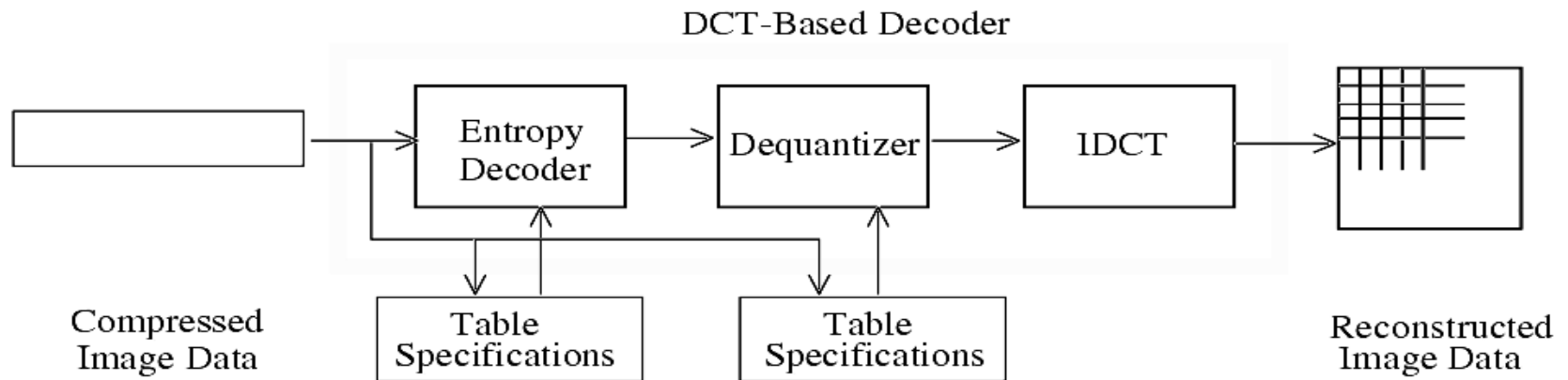


Figure 2. DCT-Based Decoder Processing Steps

Compression d'images fixes JPEG :

Transformées cosinus : applications linéaires bijectives $64D \leftrightarrow 64D$:

$f(x, y)$: bloc d'image initial (représentation spatiale) avec :

$$0 \leq x < 8, \quad 0 \leq y < 8, \quad -2^{p-1} \leq f(x, y) < 2^{p-1}, \quad f(x, y) \text{ entier}$$

$F(u, v)$: bloc d'image transformé (représentation fréquentielle) avec :

$$0 \leq u < 8, \quad 0 \leq v < 8, \quad F(u, v) \text{ réel}$$

Transformée cosinus directe :

$$F(u, v) = \frac{1}{4} C(u) C(v) \left[\sum_{x=0}^{x=7} \sum_{y=0}^{y=7} f(x, y) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right]$$

Transformée cosinus inverse :

$$f(x, y) = \frac{1}{4} \left[\sum_{x=0}^{x=7} \sum_{y=0}^{y=7} C(u) C(v) F(u, v) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right]$$

$$C(u), C(v) = 1/\sqrt{2} \text{ si } u, v = 0, \quad C(u), C(v) = 1 \text{ sinon.}$$

Compression d'images fixes JPEG :

Quantification :

$$F^Q(u, v) = \text{arrondi_entier} \left(\frac{F(u, v)}{Q(u, v)} \right)$$

Déquantification :

$$F^{Q'}(u, v) = F^Q(u, v) \cdot Q(u, v)$$

Compression d'images fixes JPEG :

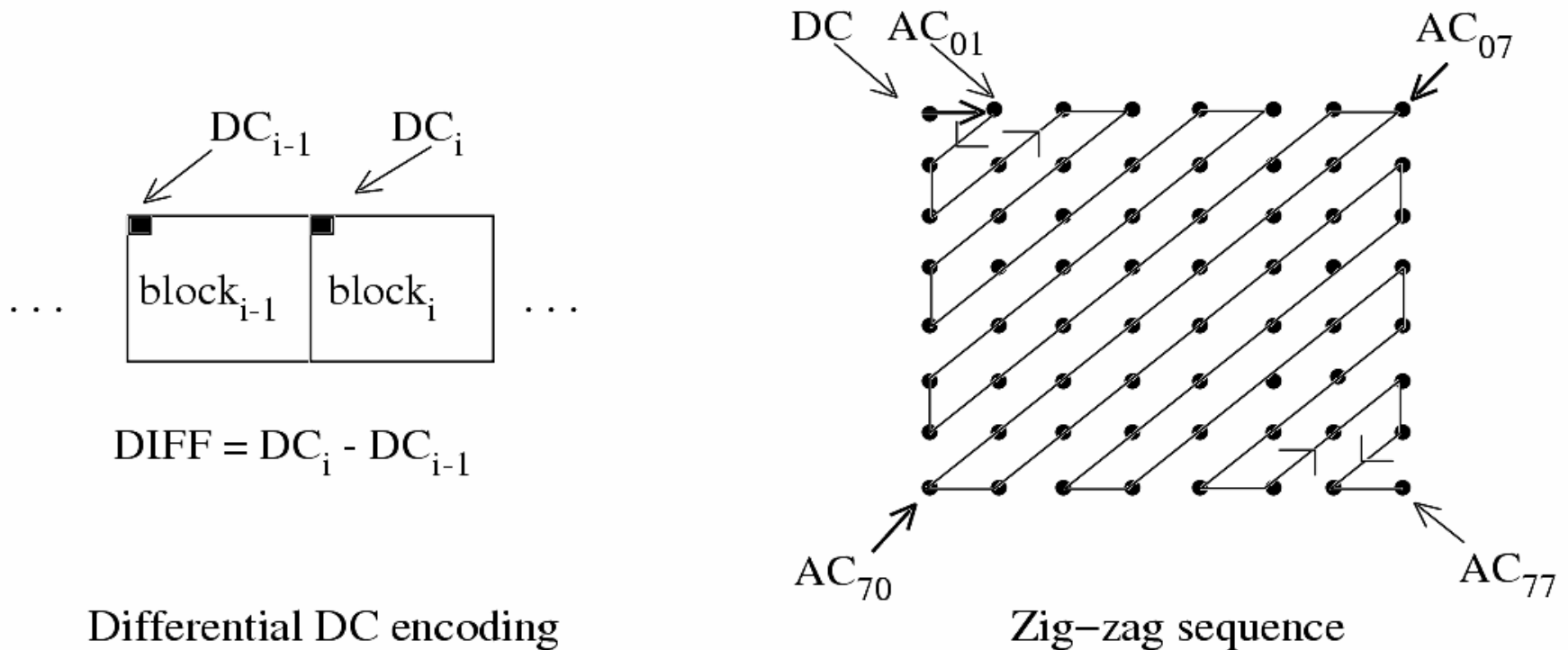


Figure 3. Preparation of Quantized Coefficients for Entropy Coding

Exemple de quantification :

139	144	149	153	155	155	155	155	235.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	1.3	16	11	10	16	24	40	51	61
144	151	153	156	159	156	156	156	-22.6	-17.5	-6.2	-3.2	-2.9	-0.1	0.4	-1.2	12	12	14	19	26	58	60	55
150	155	160	163	158	156	156	156	-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1	14	13	16	24	40	57	69	56
159	161	162	160	160	159	159	159	-7.1	-1.9	0.2	1.5	0.9	-0.1	0.0	0.3	14	17	22	29	51	87	80	62
159	160	161	162	162	155	155	155	-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3	18	22	37	56	68	109	103	77
161	161	161	161	160	157	157	157	1.8	-0.2	1.6	-0.3	-0.8	1.5	1.0	-1.0	24	35	55	64	81	104	113	92
162	162	161	163	162	157	157	157	-1.3	-0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8	49	64	78	87	103	121	120	101
162	162	161	161	163	158	158	158	-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4	72	92	95	98	112	100	103	99

(a) source image samples

(b) forward DCT coefficients

(c) quantization table

15	0	-1	0	0	0	0	0	240	0	-10	0	0	0	0	0	144	146	149	152	154	156	156	156
-2	-1	0	0	0	0	0	0	-24	-12	0	0	0	0	0	0	148	150	152	154	156	156	156	156
-1	-1	0	0	0	0	0	0	-14	-13	0	0	0	0	0	0	155	156	157	158	158	157	156	155
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	160	161	161	162	161	159	157	155
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	163	163	164	163	162	160	158	156
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	163	164	164	164	162	160	158	157
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	160	161	162	162	162	161	159	158
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	158	159	161	161	162	161	159	158

(d) normalized quantized coefficients

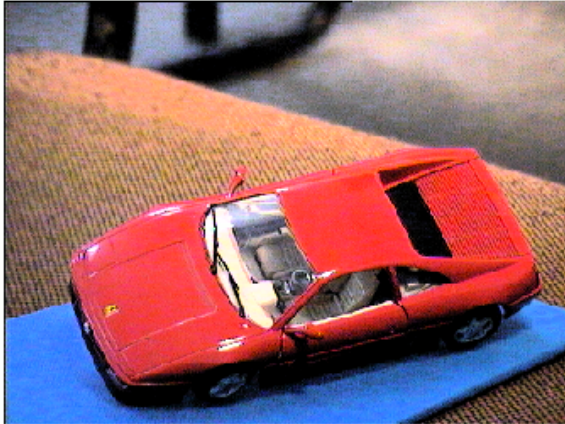
(e) denormalized quantized coefficients

(f) reconstructed image samples

Traitement des images multibandes (couleur) :

- L'information de couleur est perçue subjectivement avec une définition plus faible que l'intensité lumineuse,
- Transformation dans l'espace des couleurs :
 - Linéaire : $(R,G,B) \rightarrow Y,U,V$
 - Non linéaire : $(R,G,B) \rightarrow (I,S,T)$
- Sous-échantillonnage des informations de couleur : même couleur pour des groupes de 2 ou 4 pixels voisins (1 x 2, 2 x 1 ou 2 x 2), seule l'intensité diffère,
- Compression de chaque bande indépendante,
- Tables de quantification différentes (plus de compression et de pertes).

Exemple de compression JPEG :



331920 octets
Taux = 1:1



7930 octets
Taux \sim 42:1



2987 octets
Taux \sim 111:1

Taux d'utilisation typique : 10:1.

En pratique, on spécifie un "niveau de qualité".

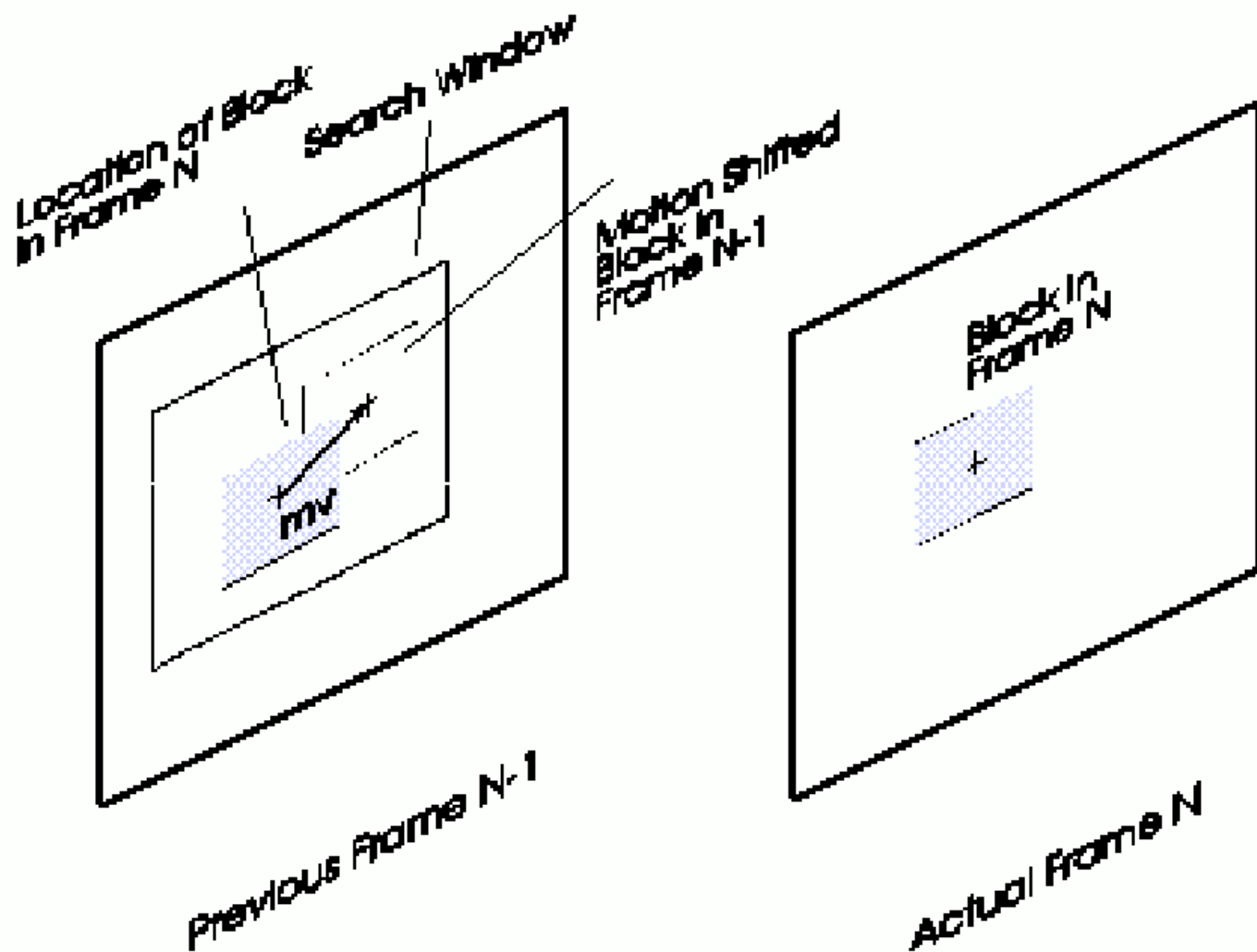
Compression de films ou de vidéos MPEG :

- Comme les images fixes (JPEG) plus : exploitation des redondances *temporelles* statistiques et subjectives.
- MPEG-1 : qualité « VHS » (352 x 288, non entrelacé),
~ 1.5 Mbit/s
- MPEG-2 : qualité « broadcast » (704 x 576, entrelacé),
~ 5.0 Mbit/s
- MPEG-4 : qualité adaptative de quelques kbit/s à plusieurs Mbit/s, codage par « objets »,
- Codage de l'image et du son,
- Compression typique de l'image ~ 50 (MPEG-1 et -2),
- Compression typique de l'audio ~ 6 (MPEG-1 et -2),
- Possibilité d'interaction (MPEG-2 et -4).

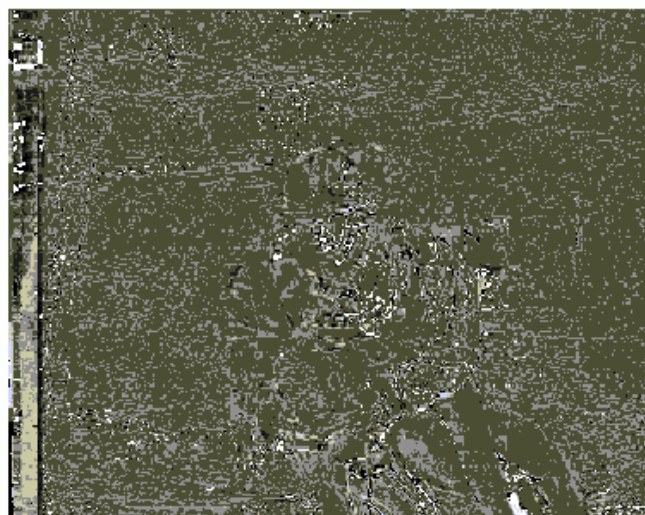
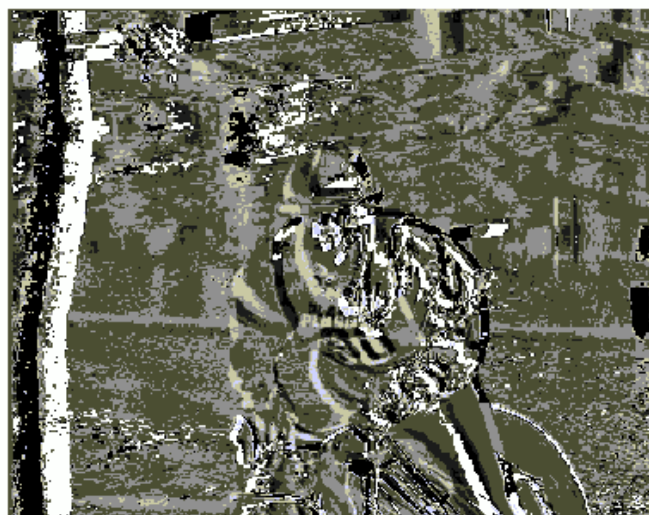
Compression de films ou de vidéos MPEG-1 ou -2 :

- Codage complet de quelques images (méthode ~ JPEG),
- Prédiction des autres par compensation de mouvement,
- Recherche du bloc le plus ressemblant dans une image précédente (et/ou une image suivante),
- Codage du bloc prédit par un vecteur déplacement,
- Codage de la différence entre le bloc prédit et le bloc réel suivant la même méthode (~ JPEG) mais avec moins de bits,
- Codage direct (avec moins de bits) si la prédiction est mauvaise,
- Trois types d'images (frames) :
 - Images « intra-codées » (I), typiquement 1 sur 12,
 - Images « prédites » (P), typiquement 3 sur 12,
 - Images « bidirectionnelles » (B), typiquement 8 sur 12.

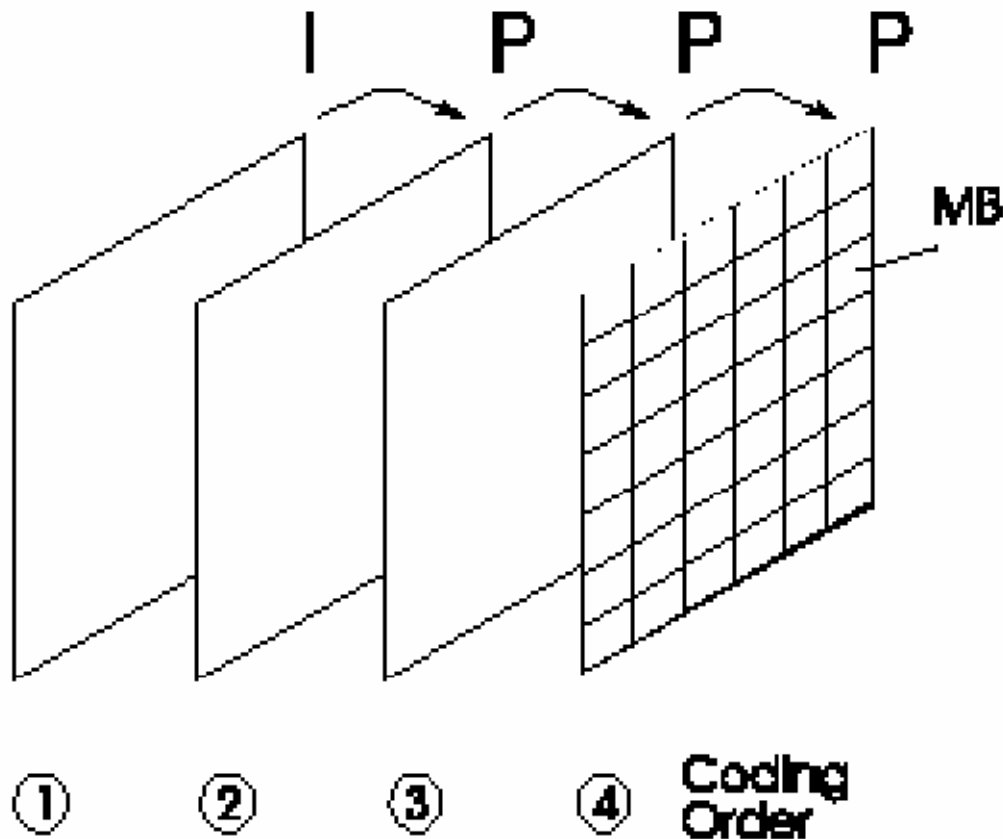
Prédiction par compensation de mouvement :



Prédiction par compensation de mouvement :



Images prédites (frames P) :

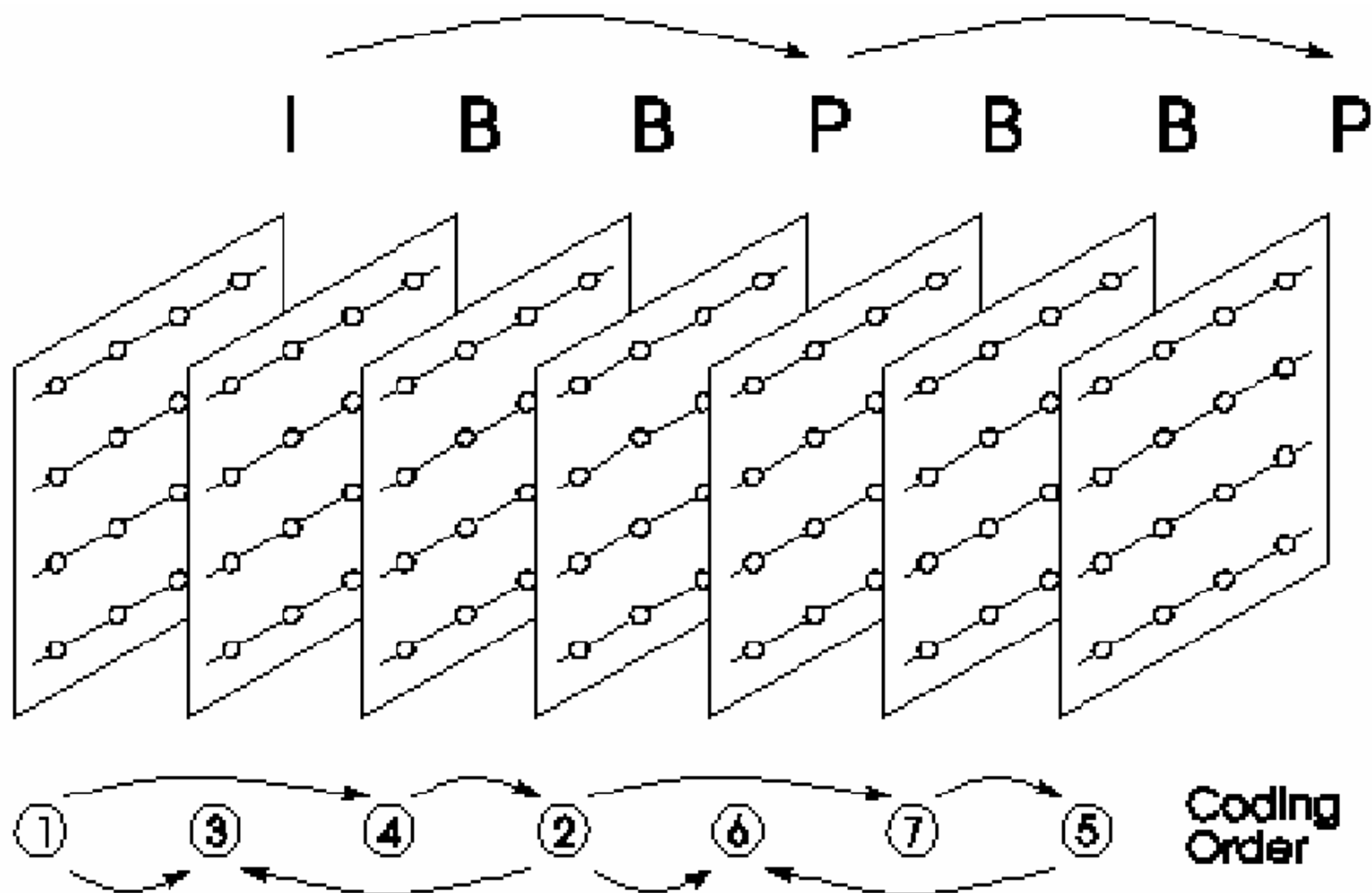


A.) Frame Prediction

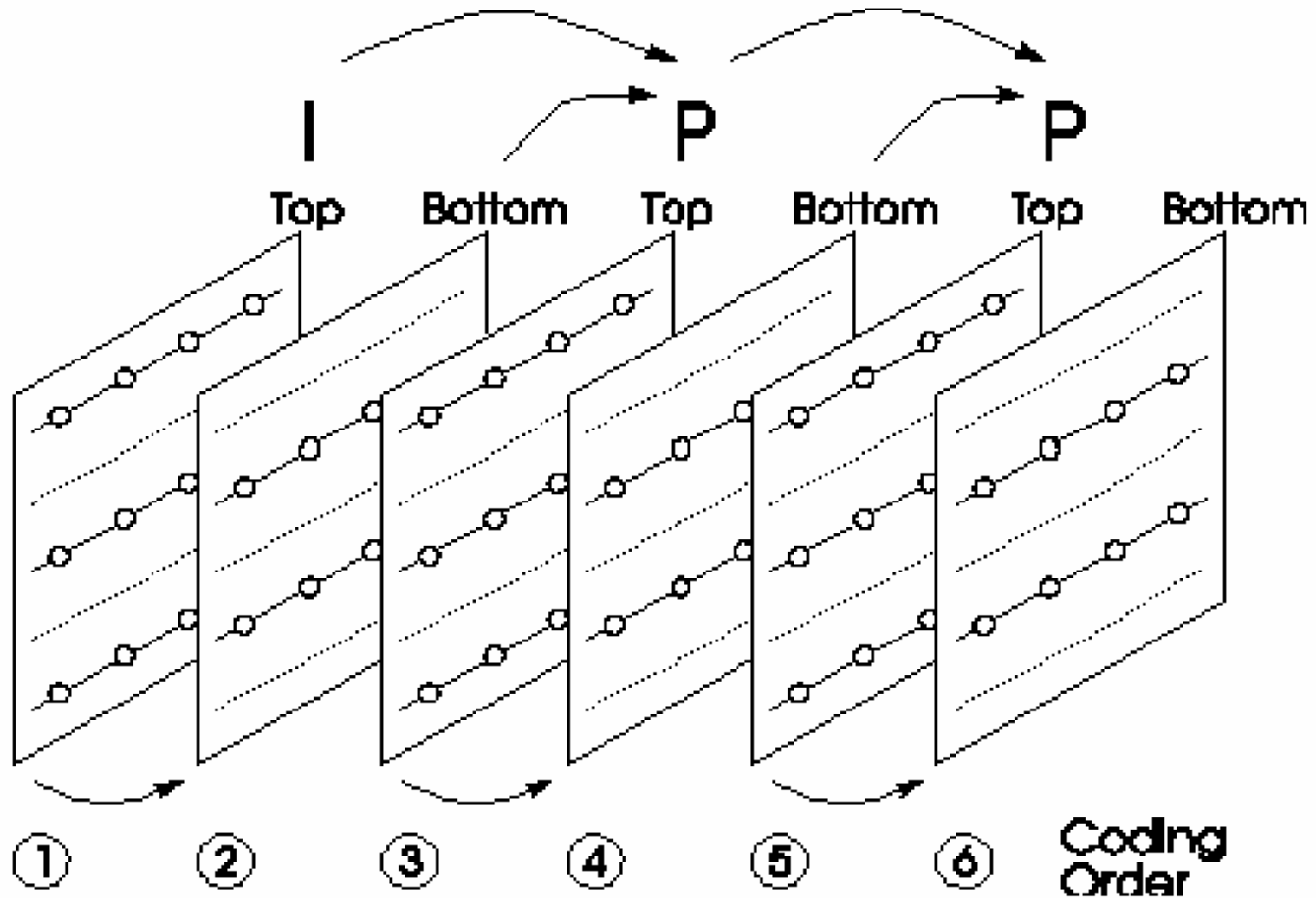


B.) Macroblock

Images bidirectionnelles (frames B) :



Frames entrelacées en MPEG 2 (I et P seulement ici) :



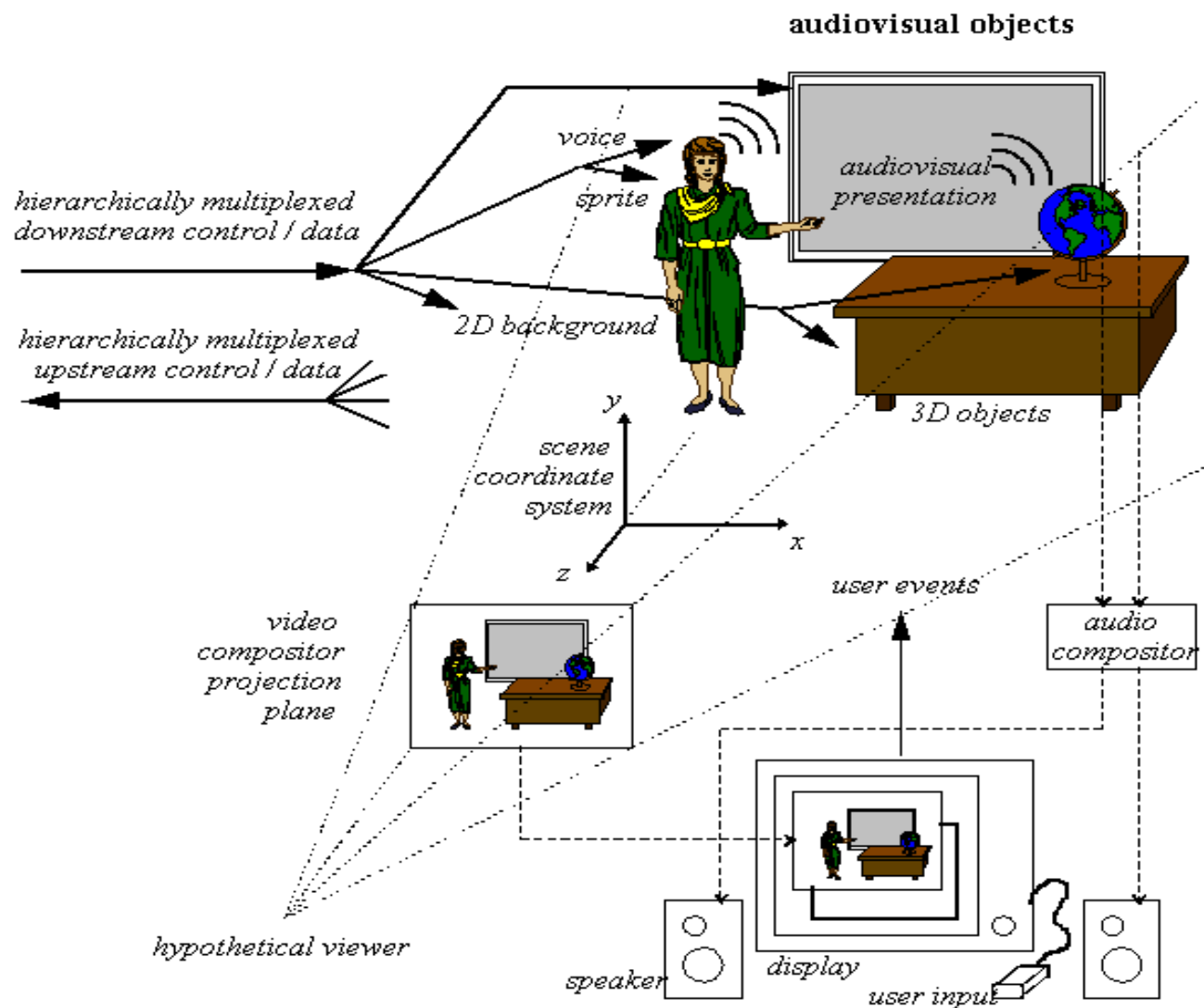
Structure des flux image MPEG-1 ou -2 :

- Vidéo = entête + série de séquences,
- Séquence = entête + série de groupes d'images,
- Groupe d'images = entête + Image intra codée [+ images prédites [+images bidirectionnelles]]
- Image = entête + série de tranches,
- Tranche = entête + série de macroblocs,
- Macrobloc = 1 bloc U + 1 bloc V + 1, 2 ou 4 blocs Y,
- Bloc = Intra-codé ou prédit + correction,
- Tables de quantification, marqueurs et infos diverses dans les entêtes,
- Structure IBBPBBPBBPBBIBBP... : libre et pas forcément régulière.

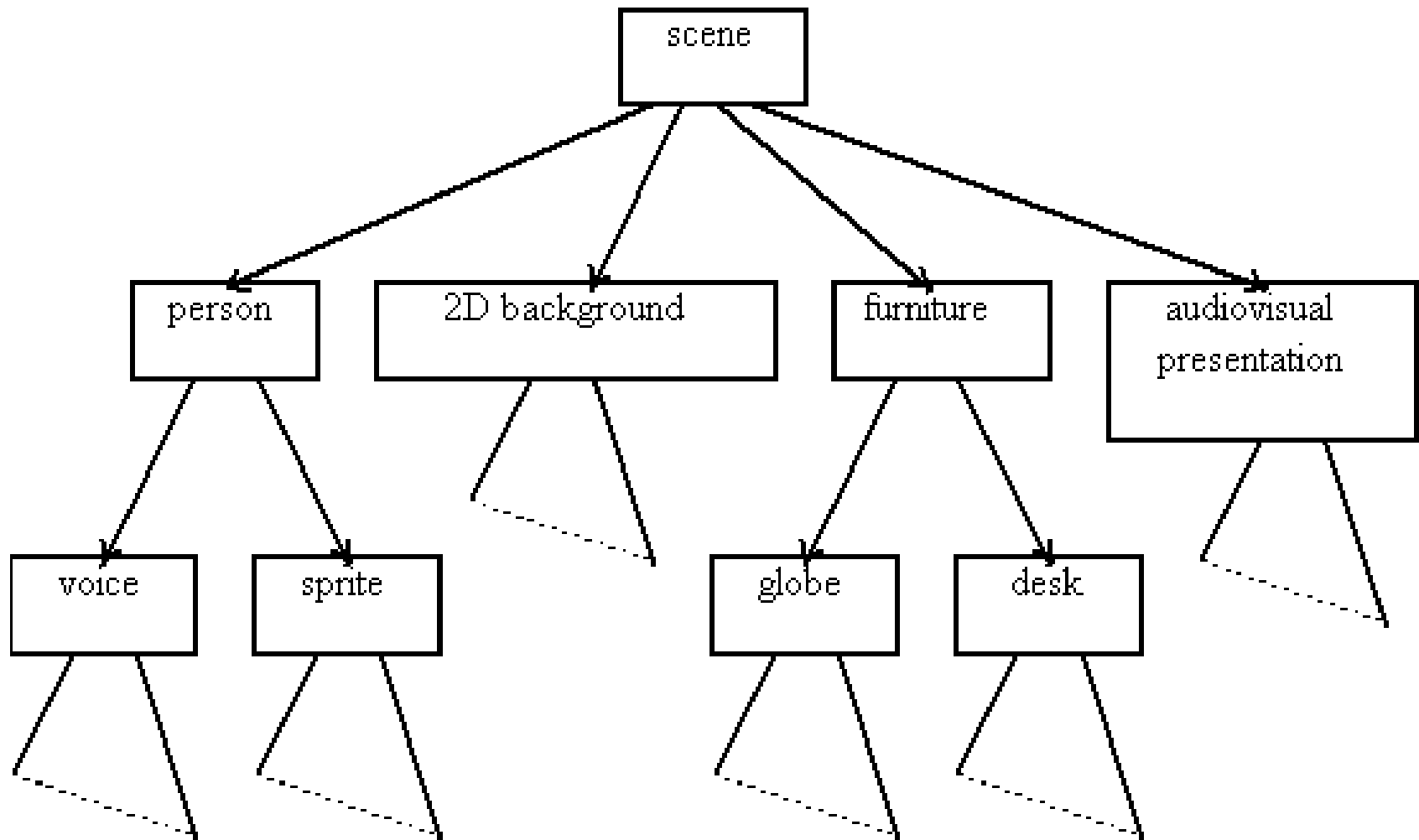
Flux systèmes (MPEG-1 ou -2) et transport (MPEG-2) :

- Système = entête + paquets audio + paquets vidéo + marqueurs + éléments de synchronisation,
- Transport = entête + paquets systèmes de plusieurs flux,
- Extensions pour l'interaction.

Le concept de scène et d'objet en MPEG-4 :



Structure de scène en MPEG-4 :



Les Objets Audio-visuels (AVOs) primitifs en MPEG-4 :

- Sons « naturels » (~ MPEG audio classique +),
- Sons synthétiques (MIDI, synthèse vocale),
- Effets sonores,
- Textures, images et vidéo « naturelles » (~ JPEG et MPEG classiques),
- Objets synthétiques :
 - Animations faciales,
 - Animations corporelles,
 - Réseaux maillés 2D animés,
 - Réseaux maillés 3D génériques,
 - Objets VRML, textes, logos,
- Un flux élémentaire par objet audio-visuel et un codage par type.

Composition des objets audio-visuels en MPEG-4 :

- Objets venant de différentes sources par différents médias,
- Recomposition de la scène à l'arrivée en fonction :
 - de la qualité des transmissions (disponibilité des données),
 - des capacités (de calcul ou algorithmiques) du décodeur,
 - des choix de l'utilisateur (qualité du fond ou d'une personne, changement de point de vue, position des objets),
- Interaction avec la scène dans « les limites prévues pas l'auteur »,
- Informations de propriété intellectuelle et informations d'indexation, (forme non spécifiée) pour chaque objet et chaque composition.

MPEG-7 : « Multimédia Content Description Interface » :

- Standard défini,
- « Langage de Définition des Descriptions » (DDL),
- Extraction ou entrée manuelle de « Descripteurs » (D),
- Construction hiérarchique de « Schémas de Description » (DS),
- Documentation de :
 - images, graphiques, modèles 3D, audio, parole, vidéo, \dots,
 - structuration de ces éléments dans un document composite.
- Informations complémentaires :
 - le format du document,
 - les conditions d'accès (propriété intellectuelle, prix, public),
 - le contexte, autres documents probablement pertinents.

Références, quelques images et infos présentées ici sont issues de :

- <ftp://ftp.uu.net/graphics/jpeg/wallace.ps.gz> : The JPEG Still Picture Compression Standard, Gregory Wallace, Multimedia Engineering, Digital Equipment Corporation.
- <http://www.mpeg.org/> : le site MPEG officiel.
- <http://www.csel.tstet.it/ufv/leonardo/paper/isce96.htm> : MPEG and multimedia communications, Leonardo Chiariglione, CSELT - Torino - Italy,
- <http://wwwam.HHI.DE/mpeg-video/papers/sikora/mpeg12/mpeg12.htm> : MPEG-1 and MPEG-2 Digital Video Coding Standards, Thomas Sikora, Heinrich-Hertz-Intitut Berlin - Image Processing Department.
- <http://wwwam.HHI.DE/mpeg-video/standards/mpeg-4.htm> : Overview of the MPEG-4 Version 1 Standard, Editor: Rob Koenen,
- <http://drogo.csel.tstet.it/mpeg/public/w2460.html> : MPEG-7 Context and Objectives.