Projet BARRACUDA: Etude Audio/Vidéo

Pierre Jouvelot
CRI, Ecole des Mines de Paris
jouvelot@cri.ensmp.fr

Action de R&D G/3042/1 BARRACUDA du CNES,
concernant les Banques de Données Multimédia

9 mars 1995

Table des matières

Chapitre 1

Présentation

Le présent document est le rapport de l'*Etude Audio/Vidéo* du projet BARRACUDA engagé par le CNES auprès de la CISI.

1.1 Exigences

Cette étude Audio/Vidéo répond au besoin, exprimé par le CNES dans le document BAR-SB-0100-001-CN du projet BARRACUDA, d'un survol des techniques actuelles et standards permettant:

- d'acquérir,
- de stocker,
- de compresser,
- d'éditer,
- et de restituer

sur stations de travail et PC des flux d'informations non textuelles de type son et image.

Ce rapport, suite aux besoins exprimés par le CNES, propose également un ensemble significatif de produits du marché permettant de réaliser les objectifs évoqués ci-dessus dans le cadre de l'expérience ALICE. Parmi ceux-ci, le produit sélectionné pour le prototype du projet devra pouvoir être intégré au SGBD choisi pour le stockage des données du projet BARRACUDA.

1.2 Structure du document

Pour répondre à ces exigences, ce rapport est divisé en deux parties:

- La première partie (chapitre 2), intitulée *Etat de l'Art en Compression Multimédia*, présente en détail un survol de l'existant dans le domaine de la compression de données multimédia, en insistant sur les techniques et standards utilisés aussi bien dans le domaine du son que de l'image. On y décrit les principes fondamentaux de la manipulation (acquisition/numérisation, compression, décompression/restitution) des données non-textuelles.

– La seconde partie (chapitre 3), intitulée *Application au Projet BARRACUDA*, est plus spécifiquement dédiée à l'application du CNES. Une première section rappelle les exigences du projet en ce qui concerne l'utilisation des données de type audio et vidéo. Elle est suivie par une évocation des produits significatifs disponibles sur le marché et susceptibles de répondre à ces exigences. La troisième section discute des problèmes d'adéquation entre ces produits et les exigences du projet, du prototype ALICE.

Pour des raisons de généralité, nous avons préféré découpler la présentation de l'état de de l'art en compression de données non-numériques des préoccupations de l'expérience BARRACUDA.

Prière de contacter l'auteur pour toute erreur présente dans ce document.

Chapitre 2

Etat de l'Art en Compression Multimédia

2.1 Introduction

Ce chapitre présente un survol des techniques d'acquisition par digitalisation, de compression et enfin de restitution par décompression des informations non-textuelles, de type son et image. L'accent est mis ici sur une présentation aussi exhaustive que possible des concepts essentiels utilisés, dans le monde du multimédia, pour la manipulation de son et d'images. On insiste également sur l'existence des nombreux standards internationaux qui assurent la pérennité des développements engagés dans ce secteur.

La structure de ce document est la suivante:

- La première section (section 2.2) montre en quoi les données non-textuelles se différencient des informations alphanumériques plus courantes. En particulier, des ordres de grandeur clefs en terme de volume et de temps de traitement sont introduits. Ceux-ci permettent de motiver l'élaboration de techniques de compression qui autorisent la manipulation des grands volumes de données que génèrent ces média.

- Un rapide rappel (section 2.3) des notions fondamentales de la théorie de l'information permet d'évoquer les bases nécessaires à une bonne compréhension des techniques présentées par la suite. On introduit ici les notions d'information (section 2.3.1), de compression entropique (section 2.3.2), de prédiction (section 2.3.3) et de quantification (section 2.3.4). Ces concepts sont nécessaires pour apprécier les techniques décrites par la suite;

- De part sa riche nature, le son s'avère complexe à compresser. On présente tout d'abord les techniques de codage du son (section 2.4.1). Puis les techniques de compression par prédiction linéaire sont introduites (section 2.4.2), suivie du standard MPEG-Audio (section 2.4.3). Notons tout de suite que, dans un système mêlant audio et vidéo, la bande passante que demande le son est faible par rapport à celle demandée par l'image.

- Les techniques fondamentales de compression des séquences vidéo sont dérivées de celles utilisées dans la compression des images fixes: celles-ci font l'objet de la sec-

tion 2.5. Après avoir présenté les techniques de codage des images (section 2.5.1), on étudie les images noir-et-blanc (section 2.5.2), les grisées (section 2.5.3), avec en particulier le standard JBIG, et enfin les images fixes couleur (section 2.5.4) avec le standard JPEG;

- Les techniques abordées dans la section précédente sont étendues dans la section 2.6 à la gestion de séquences animées. On étudie tout d'abord le standard MPEG (section 2.6.1), en présentant en particulier les standards émergents que sont MPEG-2 et MPEG-4 (section 2.6.2). Pour la vidéoconférence, le stantard antérieur H.261 est rapidement évoqué (section 2.6.3);

- Pour atteindre des taux de compression supérieurs à ceux obtenus avec les technologies précédentes, des approches nouvelles ont été suggérées comme les ondelettes ou la compression fractale. Celles-ci sont évoquées dans la section 2.7;

2.2 Motivations

Les techniques de traitement de l'information sont adaptées à la manipulation de données appartenant à un domaine discret tel que l'alphabet d'un langage. Ces données, dites *digitales*, sont à opposer aux valeurs *analogiques* provenant de signaux sonores ou visuels. Celles-ci ne sont manipulables aisément par l'outil informatique [1] qu'après une étape de numérisation, ou *digitalisation*, qui correspond à une phase de codage de ces données. Quelque soit le type de support utilisé, le traitement de l'information se ramène ainsi à la manipulation d'information digitale, c'est-à-dire de caractères dans un alphabet.

Si l'on voit que données analogiques et digitales peuvent être apparemment manipulées de manière identique, cela n'est vrai que tant que l'on ne prend pas en compte des aspects plus pragmatiques, tels que les contraintes de volume ou de temps de traitement. Ainsi, une séquence d'images télévision couleur comprend de l'ordre de 0.5 millions de points toutes les 30-ièmes de secondes environ. Le débit nécessaire à sa visualisation dépasse les 200 Mb/s, à comparer par exemple aux 10 Mb/s d'Ethernet. Si l'on désire traiter des images de meilleure qualité comme des photographies d'art, on considère qu'une résolution de 4000 points par inch, avec une précision de l'ordre de 30 bits par point, est nécessaire. A titre comparatif, les imprimantes laser courantes propose généralement une résolution de 300 points par inch. Les techniques de compression et de décompression de données permettent de résoudre les problèmes de manipulation de ces très gros volumes d'information.

La *compression de données* permet de transformer un volume de données D en un volume plus petit C; on utilise pour cela une factorisation des redondances présentes dans D. Le rapport $|C|/|D|$ des tailles respectives est le *taux de compression*. Cette compression peut être faite avec ou sans *perte* (*loss* en anglais) d'information. Si perte il y a (par exemple, par élimination de variations minimes de couleur dans une image), les dégradations du signal original reconstitué par l'étape inverse de *décompresion* devront être minimisées. Les méthodes de compression avec perte permettent, au prix d'une diminution de la qualité du signal, d'augmenter très sensiblement le taux de compression. Ceci est particulièrement utile pour les images, l'oeil étant peu sensible

1. On fait ici abstraction des approches anciennes, et remises au gout du jour, de VLSI analogiques.

aux aspects de détail du champ visuel. Du fait du fort volume de données à traiter et de la nécessité d'implémenter des algorithmes sophistiqués de recherche de redondance, une simple implémentation logicielle peut ne pas suffir; des circuits et cartes spécialisées ont été développés dans le but d'accelérer les routines de base des algorithmes de compression.

On voit ainsi que la compression de données se caractérise alors par le choix d'un compromis entre les aspects suivants:

- L'importance de la réduction des volumes de stockage et des temps de transmission de l'information;
- La qualité de l'information restituée après passage dans le pipeline compression/décompression;
- L'accroissement des temps de traitement du fait des étapes intermédiaires de compression/décompression;
- Les couts du stockage, de transfert et de traitement de l'information;
- Le type de configuration informatique disponible (carte de compression intégrée ou traitement logiciel).

Ainsi, il n'est pas envisageable d'utiliser des méthodes de compression avec perte pour des textes ou des images où chaque détail compte (par exemple, des scanners médicaux, des images satellite). Par contre, ces techniques sont bien adaptées au son (musique, parole) ou aux images de qualité moyenne (télévision, vidéoconférence). On passera alors d'un taux de compression sans perte de l'ordre de 1:2, pour une image couleur de complexité moyenne, à plus de 1:100, voir 1:1000 pour les techniques les plus avancés avec perte.

De par l'importance commerciale des marchés concernés, la large palette de technologies existantes et les couts de mise en place des infrastructures, le choix de la méthode de "codage-compression" à utiliser dans un domaine donné est stratégique pour l'industrie. Si les premiers algorithmes de compression étaient développés sans concertation, il n'en n'est plus de même aujourd'hui; les organismes de standardisation comme l'ISO, l'IEC ou le CCITT (appelé maintenant ITU-T) ont un rôle majeur dans la mise en application des techniques développées dans les centres de recherche. L'existence de standards permet d'augmenter l'interopératibilité des systèmes, de diminuer les couts de développement et d'industrialisation (en particulier, en rendant économiquement viables des approches VLSI en grande quantité) et de pérériser les techniques utilisées. On jugera la pertinence de cette analyse par l'omniprésence de la télécopie (fax), dont les algorithmes de compression ont été parmi les premiers à être normalisés.

Parmi les conséquences de cette concertation accrue entre industries des médias, on trouve la convergence en cours entre les mondes du téléphone, de la télévision, des stations de travail et des télécommunications (télévision haute définition HDTV, vidéo à la demande). Enfin, directement lié à l'intervention croissante des industriels dans ces standards, se pose le problème des droits d'accès à ces technologies de codage/compression. Le domaine du traitement d'image, en particulier, est celui qui génère le plus de brevets (patents). Ainsi, à titre d'exemple, l'US Patent Office a enregistré, en 1994, 623 brevets dans le domaine du traitement d'image, le plus haut score

parmi tous les domaines regroupés dans la rubrique "logiciel". Il est donc important de bien vérifier les limites d'utilisation des logiciels ou technologies utilisés, comme le montre l'exemple récent (Janvier 1995) de la controverse à propos du format GIF utilisé libéralement sur Internet et soumis de fait à un brevet Unisys.

2.3 Rappels

Nous présentons ci-dessous les données de base nécessaires à une bonne compréhension de la suite du rapport. Ces informations ne sont pas spécifiques à la compression multimédia, mais intéressent également la compression de texte, sans perte.

2.3.1 Information

L'information est une mesure du caractère aléatoire d'un *message* échangé, sur un *canal*, entre un *émetteur* et un *récepteur*. Un message ici est vu comme une suite de *caractères* pris dans un *alphabet* donné. Intuitivement, la notion d'information (Claude E. Shannon, Bell Labs, 1940) s'établit autour de la trilogie: "Plus d'aléatoire, plus d'information, plus de bits". Ainsi, un message complétement déterminé, par exemple constant, ne nécessite la transmission d'aucun bit puisqu'il peut être reconstruit par le récepteur.

Historiquement, la compression de données a été longtemps vue comme une partie de la théorie du codage. Suivant la loi de distribution des probabilités d'émission par la source des caractères de l'alphabet, il s'agit de choisir le plus efficacement possible la manière de représenter (coder) messages et/ou caractères en bits. Le théorème fondamental de Shannon montre que, pour une source d'ordre 1 (les probabilités p_i des n caractères de l'alphabet sont indépendantes), le codage binaire d'un caractère nécessite, en moyenne, $H(n)$ bits. $H(n)$ est l'*entropie* de la source:

$$H(n) = \sum_{i=1}^{n} p_i log_2(1/p_i)$$

Un certain nombre de conséquences, importantes pour une utilisation pratique d'algorithmes de compression, peuvent être déduites de cette notion d'information:

- Une source aléatoire est non compressible; son entropie est $log_2(n)$. Ainsi, une image complètement bruitée est non compressible;
- Les données compressées par un compresseur optimal pour le modèle de la source ne peuvent être compressées. Il est donc inutile de pipeliner plusieurs compresseurs optimaux. Attention, ceci n'est plus vrai dès que les algorithmes de compression utilisés ne sont plus optimaux;
- On ne peut garantir qu'un compresseur aura une performance donnée sur toute donnée. En particulier, pour tout algorithme de compression sans perte, il existe au moins un message dont la taille, après compression, est supérieure à celle qu'il avait avant codage. Attention, ceci n'est plus vrai dès que la compression est avec perte! Et, en pratique, pipeliner plusieurs algorithmes de compression est souvent utilisé.

2.3.2 Compression Entropique

Les premiers algorithmes de compression, qui peuvent aussi être vu comme des techniques de codage, utilisent les soubassements fondamentaux de la théorie de l'information. Nous nous limitons à décrire ici ceux qui sont effectivement utilisés dans les algorithmes de compression d'image et de son.

Le codage le plus simple est appelé *codage RLE* pour "Run-Length Encoding". Toute répétition $c...c$ de n occurences d'un même caractère c est simplement remplacée par le couple (n, c), formé du compte n de caractères suivi de sa valeur c. Du fait de sa simplicité, le codage RLE est très souvent utilisé, en particulier dans les phases finales des standards JPEG (section 2.5.4) et MPEG (section 2.6.1).

On trouve ensuite toute la famille des codes préfixes. Dans un code *préfixe*, on associe à tout caractère une chaine de bits telle qu'aucun code n'est préfixe d'un autre. La construction d'un tel code peut être une simple procédure récursive batie sur les probabilités des caractères de l'alphabet (code d'*Huffman*) ou peut être réalisée dynamiquement lors de l'envoi des messages (code *arithmétique*). Si le codage arithmétique est optimal pour toute loi de distribution, son implémentation efficace est beaucoup plus complexe que celui d'Huffman qui, lui, a une optimalité restrainte aux lois de probabilités en puissance de 2. De plus, le codage arithmétique est soumis à de nombreux brevets (une licence coute de l'ordre de $5000 auprès d'IBM, AT&T et Mitsubishi), tandis que le codage d'Huffman est libre d'utilisation. On considère généralement que l'amélioration relative de performance apportée par le codage arithmétique ne justifie pas ces inconvénients, mais celui-ci apparait néanmoins dans le standard JBIG (section 2.5.3).

Si la loi de distribution des caractères de l'alphabet varie dans le temps, les codeurs entropiques présentés ci-dessus, dits *statiques*, ne sont plus adaptés. Il existe des versions dites *adaptatives* de ces compresseurs qui modifient les codes en fonction de ces fluctuations temporelles (par exemple, le nombre de bits utilisés pour coder les caractères). A nouveau, leur complexité n'est pas toujours justifiée, bien qu'on en trouve souvent des versions simplifiées, par exemple dans le codage RLE des standards JPEG (section 2.5.4) ou MPEG (section 2.6.1).

2.3.3 Prédiction

Si la loi de distribution des caractères présente de fortes corrélations entre émissions successives de caractères, il est possible d'obtenir un codage plus efficace en ayant recours à la notion de *prédiction*. Ici, au lieu de coder chaque caractère indépendamment les uns des autres, on s'attache à prévoir la valeur d'un caractère en fonction de ceux déjà émis par l'émetteur. Seule la différence, faible dans les bons cas, entre cette valeur prédite et la valeur effective du caractère est codée. Cette technique de prédiction est utilisée largement dans les standards audio comme CELP (section 2.4.2) ou images comme JPEG (section 2.5.4).

2.3.4 Quantification

La *quantification* ("quantization") est une technique similaire à la technique de Shannon: on associe à une suite de n caractères (et non plus un caractère unique) un code stocké dans une table partagée par l'émetteur et le récepteur. La compression

vient de ce que cette table, qui n'a pas toujours â être représentée comme telle, mais peut être implémentée par une simple fonction, est de taille plus faible que l'ensemble des n chaines possibles. Une telle suite de n caractères peut être vue comme un vecteur dans un espace à n dimensions; on parle alors de quantification *vectorielle*. Si le vecteur est limité à une seule composante, il s'agit de quantification *scalaire*, dont *PCM* ("Pulse Code Modulation") est l'exemple typique, utilisé dans la conversion par échantillonnage d'un signal analogique en un signal numérique.

Par sa nature même, la quantification entraine une compression avec perte. Pour en accroitre les performances, on peut y rajouter la possibilité d'adaptativité et de prédiction. Cette technique de quantification est à la base de la compression des images utilisée dans JPEG (section 2.5.4) et MPEG (section 2.6.1).

2.4 Compression du son

On présente ici les techniques de codage du son (section 2.4.1). Puis les techniques de compression par prédiction linéaire sont introduites (section 2.4.2), suivies du standard MPEG-Audio (section 2.4.3).

2.4.1 Codage du son

La technique la plus simple de codage consiste à utiliser un échantillonage PCM des données continues représentant le son. La fréquence correspond à la hauteur du son, l'amplitude à son volume. Généralement, la fréquence est un paramètre plus sensible que l'amplitude. Une résolution type en amplitude est de 8 à 16 bits. En fréquence, selon le théorème de Nyquist, l'échantillonage doit se faire à une fréquence au moins double de celle que l'on désire préserver, soit environ 40 kHz puisque la plage de sensibilité de l'oreille humaine est [20 Hz, 20 kHz]. En pratique, pour la voix seule, on utilise un codage dit "96kb" de 8000 échantillons de 12 bits par seconde; pour la musique, on trouve 44100 paires de 16 bits par seconde pour le Compact Disc et 48000 pour le DAT.

L'échantillonage simple de type PCM peut etre utilisé en différentiel (*DPCM*) pour lequel on ne code que les différences, éventuellement quantifiées, entre échantillons, et/ou en mode prédictif *ADPCM*, dans lequel on change dynamiquement le pas de quantification. Cette dernière technique est à la base du standard G.721 et G.723 du CCITT pour le codaghe de la voie à 32 et 24 kb/s.

Un autre mode de codage, qui peut également être vu comme une technique de compression, s'appelle *μ-law* aux US et *A-law* dans le reste du monde. Il s'agit du standard CCITT G.711 qui est une simple quantification non-linéaire (logarithmique) du signal échantilloné. Dans la A-law, on code les 12 bits du signal échantilloné sous forme de 8 bits formés respectivement:

- du bit de signe;
- de la position, codée sur 3 bits, du premier bit non-nul dans les 7 premiers bits de poids fort du signal;
- des 4 bits suivants.

Le comportement logarithmique de l'oreille, que l'on retrouve dans la définition des *décibels*[2], justifie un tel codage. A titre indicatif, les fichiers audio sur Sun utilisent la μ-law.

2.4.2 Prédiction Linéaire

Cette technique est bien adaptée au codage de la voix. La prédiction est effectuée par rapport à un modèle analytique des cordes vocales. On parle alors de *vocoder* ou "Voice Coder". On utilise typiquement un modèle linéaire de la forme:

$$t_n = Gu_n - \sum_{k=1}^{p} (a_n t_{n-k})$$

où G est un paramètre supplémentaire de gain et u_n de hauteur. Les coefficients a_k sont déterminés par analyse des moindres carrés E entre le signal r'eel s_n et le signal prédit $s'_n = -\sum_{k=1}^{p} (a_n s_{n-k})$:

$$E = \sum_{k=1}^{p} (s_n - s'_n)^2$$

G est calculé en maintenant l'énergie constante entre le signal original et le signal synthétisé ("analysis-by-synthesis"), tandis que u_n dépend du mode de codage de la hauteur de l'échantillon et varie selon la méthode de codage choisie. Par différentiation ($dE/da_k = 0$), a_k est calculable en fonction de la matrice d'autocorrélation du signal $R(i) = \sum_n (s_n s_{n+i})$.

Le message envoyé sur le canal contient (1) les paramètres a_k, u et G du modèle, ce qui permet de transmettre le son fondamental, et (2) le signal résiduel $s_n - s'_n$ (correspondant aux caractéristiques personnelles des cordes vocales). Il est possible d'ajouter un facteur d'aptabilité en augmentant p dans les plages de fréquences les plus importantes comme [0, 5 kHz].

La possibilité de compression provient de la limitation, généralement entre 8 et 14, de p et de la quantification des coefficients. En fait, pour des raisons de stabilité, les coefficients a_k ne sont pas directement quantifiés, mais le sont les coefficients de réflection, calculant l'autocorrélation entre s_n et s_{n+i}, toutes choses étant égales par ailleurs. Obtenir une bonne compression est toujours un problème de recherche ouvert, en particulier dans les très bas débit (quelques centaines de b/s) utiles dans les applications militaires sécurisées.

Il existe de nombreux standards utilisant ces techniques de prédiction linéaire:

LPC-10 "Linear Predictive Coding" (DoD Federal Standard-1015 à 2400 bps). La prédiction est d'ordre 10; u_n correspond à un simple *bruit blanc*, i.e. suite d'échantillons décorrélés de moyenne nulle et de variance 1. Les informationss de fréquence et d'amplitude sont transmises toutes les 10-20 ms;

CELP Code Excited Linear Predictor" (DoD Federal Standard-1016 à 4800 bps, et aussi ITU G.728). Ici, u_n est un index dans une table de codage ("Code Excited") de 1024 échantillons de 5 ms changeant tous les 40 échantillons. CELP

2. Le *bel* est proportionnel au logarithme en base 10 du rapport de deux intensités ou puissances (signal, bruit). Le *décibel* est un dixième de bel.

est meilleur, moins synthétique, que ADPCM et est de qualité comparable à ADPCM-32;

VSELP "Vector-Sum Excited LP". Il est spécialisé pour des liaisons à faible débit (8 kb/s) et est utilisé dans le téléphone cellulaire US. On utilise 3 codebooks différents (pitch-adaptative, 2 structurés);

RPE-LTP "Regular Pulse Excitation-Long Term Predictor". Il est utilisé dans le téléphone mobile GSM (Groupement Spécial Mobile) à un débit 13 kb/s. Il est basé sur la modulation d'impulsions (5 ms) régulièrement espacées, d'amplitude variable par trame (20 ms, 160 échantillons de 13 bits à 8 kHz). La prédiction à court terme d'ordre 8 est effectuée toutes les 20 ms.

2.4.3 MPEG Audio

Le standard MPEG Audio fait partie d'une suite de trois standards MPEG décrite plus complètement dans la section 2.6.1. Le "Motion Picture Experts Group" s'est attaché à développer un standard de compression d'informations sonores adapté au type de "bandes son" utilisées sur les produits audiovisuels grand public (TV, film, vidéo). Une des caractéristiques des standards MPEG, que l'on retrouve dans de nombreux autres standards de compression, comme JPEG, est que seul le format de représentation des données est spécifié. L'implémentation du codeur/décodeur, ou *codec*, est laissée libre et peut donc varier en fonction de l'introduction de nouveaux algorithmes.

Le signal initial traité par MPEG Audio est un échantillonage PCM à 705 kb/s (soit un canal de qualité CD). Le signal compressé doit lui être amené à un débit compris entre 32 et 384 kb/s, soit un taux de compression allant jusqu'à 1:22. On constate en pratique qu'il n'y a pas de perte audible de qualité jusqu'à un facteur 1:7. Pour atteindre cet objectif, le standard est bati sur 3 niveaux de compression, compatibles ascendants, caractérisés par des débits croissants (32 et 64 kb/s pour le niveau III, 128 kb/s pour le niveau II et 192 à 384 kb/s pour le niveau I). En pratique, on trouve essentiellement des implémentations du niveau II. Quand on passe d'un niveau au suivant, à qualité sonore identique, on accroit (1) la complexité du codeur, (2) le retard entre le signal et son équivalent compressé à puissance de machine constante et (3) la qualité du son, à complexité de compression constante.

MPEG Audio utilise une méthode de compression dite *psycho-acoustique*, basée sur les défauts de perception de l'oreille humaine. Ainsi, par exemple, un signal de 1 kHz, superposé à un signal de puissance plus faible de 18 dB et de fréquence 1.1 kHz, sera équivalent au seul signal à 1000 Hz car l'écart entre fréquence est trop faible pour être senti; on peut compresser ce signal en éliminant la composante à 1.1 kHz. Par contre, si le signal plus faible est de fréquence 2 kHz, il ne pourra plus être négligé car l'oreille est sensible à la différence de fréquence, malgré l'atténuation; aucune compression n'est possible. En revanche, si l'atténuation est amenée à 45 dB, le signal à 2 kHz cessera d'être percu; le signal peut, à nouveau, être comprimé. Une des idées clef de MPEG Audio est donc de prendre en compte les *effets de masquage* entre fréquences dans les forts niveaux sonores.

Pour utiliser cette idée, le standard MPEG Audio spécifie que la bande sonore de 20 Hz à 20 kHz est découpée en 32 sous-bandes; on parle de *subband coding*. L'effet de masquage est calculé entre chaque sous-bande et ses sous-bandes adjacentes. Par

exemple, la sous-bande 8, autour de 1 kHz, masque la sous-bande 7 quand la différence de niveau est de 25 dB. Pour prendre en compte la sensibilité accrue de l'oreille autour de 2 à 4 kHz, plus de bits de résolution sont affectés à cette tranche. Outre le masquage de niveau, MPEG Audio offre aussi la possibilité d'utiliser le *masquage temporel*. Un son plus faible de 30 ou 40 dB est négligeable en présence d'un son plus fort qui le suit de de quelques millisecondes ou le précède jusqu'à un dixième de seconde. En fait, puisque MPEG Audio ne spécifie pas l'algorithme de compression, mais simplement le format de représentation en subband coding est d'autres informations peuvent être utilisées pour augmenter le taux de compression: traitement différencié des informations tonales et atonales, couplage entre canaux, etc ... Un codage d'Huffman est enfin appliqué au résultat.

En ce qui concerne l'implémentation de cette spécification, citons certains ordres de grandeur intéressants. Au niveau II, les sous-bandes sont traitées par tranches de 23 ms. Le calcul du masquage se fait de manière itérative, avec arrêt au bout d'un temps limite en vue de préserver la synchronisation. On considère qu'un compresseur temps-réel MPEG Audio de niveau III nécessite une puissance de l'ordre de 20 MIPS, alors qu'un simple circuit de type DSP peut suffir au niveau II.

La version suivante de MPEG, dite MPEG-2 (voir section 2.6.2), propose également une évolution de l'aspect sonore de la suite de standards MPEG. MPEG-2 Audio est compatible avec MPEG (dit aussi MPEG-1), mais permet des taux d'échantillonage plus réduits (de 16 à 24 kHz) pour s'adapter aux lignes à faible débit. MPEG-2 propose également des canaux supplémentaires par rapport aux deux voies stéréo de MPEG. On parle de *5.1 canaux*: ils correspondent aux 2 canaux principaux gauche et droit, aux 2 canaux avant et arrière, à un canal central et à une faible bande passante ("0.1 canal") réservée aux codages des effets spéciaux (effets "surround-sound" par exemple).

La technologie MPEG Audio est utilisée dans de nombreux produits: au niveau II dans le project européen de radio digitale DAB ("Digital Audio Broadcasting"), au niveau III pour le transport de la musique via satellite et réseaux type RNIS. A noter que MPEG Audio est similaire au protocole MUSICAM et aussi PASC (voir section 2.4.4).

2.4.4 Autres Protocoles

Nous évoquons rapidement ci-dessous les systèmes de compression de son présents dans le système PASC utilisé dans la DCC par Philips et ATRAC concu par Sony pour le MiniDisc. Il conviendrait de citer également le Dolby AC-3, utilisé dans la future télévision HDTV américaine (les européens comptent utiliser MPEG-Audio). Néanmoins, les informations sur ce système sont très limitées du fait du caractère propriétaire de l'algorithme.

PASC . "Precision Audio Subband Coding" a été concu par Philips pour servir de noyau de compression dans la cassette DCC ("Digital Compact Cassette"). Cette nouvelle cassette soutient un débit de 384 kb/s, obtenu grâce à une vitesse de défilement de la bande de 4.76 cm/s. Cette vitesse est celle des lecteurs K7 audio classiques; un lecteur DCC intégre une tête de lecture analogique permettant de relire d'anciennes cassettes. Une seconde tête de lecture à film mince lit les cassettes digitales, de type vidéo 1/2 pouce. Pour obtenir un son proche de la

qualité d'un disque compact, DCC utilise la compression sonore avec un taux de 1:4. Celle-ci est basée sur une variante de MPEG Audio, utilisant le codage sur 32 sous-bandes uniformes (filtrage FIR unique passe-bande défini dans le standard) et les masquages de niveau et temporel. Ce masquage, effectué sur des trames de 12 échantillons stéréo, peut être réalisé avec différents types d'outils (filtre sous-bande, FFT, ...) suivant le support ciblé (enregistrement en direct, bandes préenregistrées). Le débit est réparti dynamiquement entre chaque sous-bande en fonction du niveau de chacune d'entre elles, avec priorité aux basses fréquences.

Un flux codé PASC contient plus de 30% de codes servant à la correction d'erreurs (codage Reed-Solomon), ainsi qu'à des informations annexes ("Auxiliary Channel" pour du texte, et "System Channel" réservé aux cassettes préenregistrées).

ATRAC "Adaptative TRansform Acoustic Coding" est utilisé dans le MiniDisc de Sony. Ce système est basé sur un disque laser à écriture magnéto-optique (dans un boitier similaire à une disquette magnétique). Le facteur de compression que nécessite le MD est de 1:5. La technologie Sony utilise un codage par sous-bandes (0-5.5 kHz, 5.5-11 kHz et 11-22 kHz) mais, au lieu d'appliquer une approche à la MPEG Audio, Sony effectue une transformation en fréquence voisine de la DCT utilisée dans JPEG (section 2.5.4) avec des tailles de bloc adaptatives. Une version sans perte du MiniDisc, appelée MD-Data, est adaptée au stockage informatique (140 Mo).

Du fait de la perte d'information créée par ces standards, il est déconseillé de les utiliser pour élaborer des copies maitres, en studio.

2.5 Compression d'Images Fixes

Après avoir présenté les techniques de codage des images (section 2.5.1), on étudie ici les images noir-et-blanc (section 2.5.2), les images grisées (section 2.5.3), avec en particulier le standard JBIG, et enfin les images fixes couleur (section 2.5.4) avec le standard JPEG.

2.5.1 Codage des Images

Avant d'aborder les techniques de compression propres aux images, il convient d'évoquer rapidement les techniques de codage numérique des images. L'élément de base d'une image est le *pixel* ("picture element"), dont le regroupement en grille forme l'image. La valeur associée à un pixel dépend, entre autres, du mode de l'image (noir-et-blanc, couleur) et de la précision (nombre de niveaux d'intensitée). En mode noir-et-blanc, une valeur binaire, ou sur quelques bits, permet de préciser l'intensité du point. En mode couleur, le choix de la valeur, souvent un triplet, à affecter à chaque pixel est une caractéristique de l'*espace de couleur* utilisé dans l'appareil. Il en existe de nombreux, dont ceux qui nous concernent plus directement:

RGB ("Red Green Blue"). Il est utilisé sur les écrans informatiques couleur et indique 3 niveaux de potentiel pour chacune des couleurs. Il est mal adapté à la compression car il y a peu de corrélation entre les niveaux RGB de pixels voisins;

YUV Dans cet espace, surtout utilisé en télévision et vidéo, on distingue l'information achromatique (intensité ou niveau de gris, dite *luminance*), codée en 1 dimension, de l'information chromatique (couleur, dite *chrominance*), codée elle sur 2 dimensions. Cette dichotomie permet de maintenir la compatibilité entre les vieux postes (et films) noir-et-blanc et les émissions en couleur. Il est possible de passer d'un vecteur YUV à un vecteur RGB par une simple multiplication matricielle. Cet espace est utilisé dans le standard télévision PAL, ainsi que dans le standard de l'IUT CCIR-601 de codage vidéo. Il est aussi bien adapté à la compression car les informations de chrominance et de luminance sont fortement correlées spatialement; JPEG (section 2.5.4) et MPEG (section 2.6.1) l'utilisent.

A noter le standard YIQ, pendant de YUV pour les Etats-Unis.

L'oeil étant moins sensible aux variations de couleurs, les signaux numérique vidéo et télévision utilisent la notion de *sous-échantillonage*, ou *décimation*, en couleur ("chrominance subsampling"). Ainsi, si l'intensité de chaque pixel est bien codée, seule une paire (U,V) sur 2 est préservée dans le signal; on parle alors de codage 4:2:2 (chaque champ correspond aux dimensions Y, U et V, avec une valeur de 4 pour l'absence de sous-échantillonage). Pour une précision de 8 bits par dimension, on obtient alors des débits, avant compression, de l'ordre de 200 Mb/s.

Le Comité Consultatif International de Radiodiffusion de l'UIT (Union International des Télécommunications) a proposé le standard CCIR-601 spécifiant les paramètres de codage en télévision numérique pour studios (de meilleur qualité, donc, que les télévisions grand public). Ce standard sert de référence dans le standard MPEG-2 (section 2.6.2) utilisable en HDTV. Ses caractéristiques sont les suivantes (les valeurs entre parenthèses concernant le continent américain):

- 576 (480) lignes de 720 pixels à 50 (60) trame/s entrelacées, soit 25 (30) image/s non entrelacées, dans la bande Y. Avec le sous-échantillonage en couleur, on obtient donc 576 (480) par 360 valeurs en U et V;
- Type 4:2:2;
- Transmission par groupe (U-Y-V Y) codés sur 8 bits.

Il est important de constater que les différences entre US et Europe tiennent essentiellement aux fréquences différentes du courant électrique disponible localement, les débits d'information restant égaux, et donc compatibles, dans les deux systèmes.

Rappelons que les signaux télévision sont *entrelacés*, c'est-à-dire que les images sont affichées par moitiés, appelées *trames*. Enfin, il est intéressant de comparer ces valeurs avec les caractéristiques actuelles des écrans de télévisions, soit 625 (525) lignes à 50 (60) trame/s.

2.5.2 Noir et Blanc

Dès le début des années 1980 le comité SG XIV du CCITT a travaillé à améliorer les techniques de transmission de documents noir-et-blanc, essentiellement manuscripts, envoyés sur ligne téléphonique commutée. Les systèmes de télécopie (fax) précédents, appelés Groupe 1 et Groupe 2, utilisaient des techniques analogiques limitées en débit

et résolution. Les télécopieurs de type Groupe 3 et 4 sont basés sur des méthodes numériques et utilisent des techniques de compression sans perte permettant d'améliorer par un facteur 10 les performances des faxs préexistants.

Les télécopieurs de type Groupe 3 et 4 procèdent tout d'abord à la digitalisation, puis quantification en 0/1, de feuilles au format A4, à la résolution de 200 points par inch et 1188 lignes. Le protocole est élaboré pour permettre de transmettre une telle page en 1 minute environ à 4.8 kb/s. Ceci est possible grâce à un algorithme de compression original.

La compression des télécopieurs Groupe 3 est basée sur un codage statique de type Huffman d'un codage RLE unidimensionnel. Cette compression est donc sans perte. On associe à toute séquence de points unicolorés un code binaire dont la longeur est inversement proportionnelle à la probabilité d'apparence d'une telle suite. Ces probabilités ont été estimées une fois pour toutes en utilisant un jeu de tests spécifiés dans le standard. Par exemple, un séquence de 3 blancs successifs correspond au code 1000, tandis qu'une suite de 63 noirs est codée 000001100111. Par cette simple méthode, des facteurs de compression de 5 à 15 peuvent être observés sur des documents typiques.

Parmi les expansions proposées (dont celles utilisées dans les télécopieurs de type Groupe 4), on trouve la prise en compte du format A3, des vitesses de transfert plus importantes et, surtout, une extension de la méthode de compression à 2 dimensions (on prend en compte les valeurs de la ligne précédente).

2.5.3 JBIG

Une nouvelle génération de norme pour télécopieurs a été proposée au début des années 1990 dans le cadre d'un effort commun par le groupe JTC1/SC2/SW9 de l'ISO, regroupant également l'IEC et le CCITT (SG VIII). L'objectif est de permettre la transmission d'images en niveau de gris, sans perte, avec une résolution meilleure que les fax CCITT Groupe 3 et 4 (section 2.5.2); on estime que le gain sur texte et dessin est de l'ordre de 10 à 50 % par rapport aux télécopieurs Groupe 4 (jusqu'à 300 % sur certaines images).

L'approche utilisée dans le standard JBIG (Joint Bi-Level Image expert Group) consiste à offrir des représentations multiples, avec des résolution différentes, sans surcout majeur. Ceci permet aux émetteurs/récepteurs de s'adapter au type de l'image à transmettre: télécopie standard à basse résolution, images médicales à haute résolution. La réalisation est faite par codage séquentiel d'une image progressivement codée. Ce codage *progressif* consiste à envoyer des approximations successives de l'image à transmettre, le récepteur décidant de celles à conserver en fonction de la résolution souhaitée. Ceci peut être utile dans un système multimédia pour faire une recherche d'image dans une base d'images, ou pour créer automatiquement des icones. Cette notion de codage progressif se trouve également dans JPEG (section 2.5.4).

Pour une utilisation en niveau de gris (ou même en couleur), JBIG suggère d'utiliser un code de Gray des champs de valeur. Ce type de codage, dans lequel deux niveaux adjacents sont codés par deux nombres qui diffèrent d'un bit uniquement, est moins sensible aux erreurs de transmission. On travaille ensuite par plan de bits. Pour limiter la précision, JBIG conseille d'utiliser le *"dithering"* qui consiste à approximer une couleur ou un niveau de gris inexistant par l'utilisation dans les pixels du voisinage de couleurs appropriées à l'effet global; ceci diminue néanmoins la résolution spatiale.

En fait, on conseille d'utiliser le mode sans perte de JPEG (section 2.5.4) dès que les amplitudes à coder dépassent 6 bits.

Pour un plan de bits donné, l'algorithme de codage séquentiel progressif travaille par bandes dont la hauteur est définissable par l'utilisateur (typiquement 8 mm). Pour chaque bande, on envoie successivement des codes correspond à une résolution qui double à chaque étape. L'entrelacement entre bandes, couches de résolution et plans est paramétrable par l'utilisateur. On effectue ainsi un envoi graduel d'approximations successives de l'image. Les petites bitmaps binaires ainsi déterminées sont enfin compressées par un "Q-coder".

Ce type de compresseur, breveté par IBM, est très voisin d'un compresseur arithmétique. Il utilise 11 pixels de contexte, en 2 dimensions, pour estimer la probabilité d'un point. En fonction de cette prédiction, un code est élaboré et transmis, si besoin est. Cette loi de probabilité est mise à jour au fur et à mesure de la transmission du document. Pour éviter d'effectuer des calculs flottants, les probabilités sont arrondies à la puissance de 2 la plus proche. Outre la mise à jour de la loi de probabilité, JBIG offre la possibilité d'effectuer une compression adaptative par modification de la topologie des 11 pixels de contexte.

2.5.4 JPEG

Si JBIG permet de traiter les images en niveau de gris ou couleur avec un minimum de qualité, la restriction à une technique de compression sans perte limite sérieusement les performances de compression que l'on peut espérer atteindre. L'oeil humain étant peu sensible à de faibles variations locales, il est possible, en modifiant légèrement le signal original, d'obtenir une image très ressemblante mais nécessitant beaucoup moins de place mémoire. Le groupe joint JTC1/SC2/WG10 de l'ISO, regroupant l'IEC et le groupe SG VIII du CCITT, s'est attelé dès le milieu des années 1980 à définir un standard propre à la compression d'images fixes.

Introduction

Les objectifs du standard JPEG (Joint Photographic Expert Group) sont:

- La compression d'images fixes à niveau de gris ou couleur. Chaque image est vue par JPEG comme une grille monovaluée. Les images couleur sont traitées comme 3 images indépendantes correspondant aux trois valeurs de l'espace de couleur choisi;
- L'utilisation de critères psychovisuels, liés aux imperfections de la vision humaine, pour améliorer la quantité de compression. En particulier, les techniques proposées ont été choisies par des jugements subjectifs de testeurs, jugeant en aveugle[3] la qualité des images traitées;
- L'avancement de l'état de l'art dans le domaine de la compression. En particulier, JPEG utilise la technique de transformation orthogonale DCT (section 2.5.4);
- La possibilité d'être utilisable dans de nombreux domaines. Ceci justifie la présence d'un mode de compression dans perte disponible dans JPEG;

3. Désolé!

- La paramétrisation, en particulier la possibilité de jouer sur le compromis cout vs. qualité dans la quantité de compression à effectuer. Par exemple, une image de qualité dégradée pourra être obtenue si l'on souhaite profiter d'un facteur de compression élevé;

- Le cout d'implémentation. En particulier, les choix techniques ont été faits en prenant en compte le type de technologie VLSI disponible à terme pour faciliter les réalisations matérielles, seules garantes d'économie d'échelle et de performance acceptables.

Pour remplir ces objectifs, JPEG propose quatre (4) *modes* différents de codage des images:

- Le mode *séquentiel*. Dans ce mode, un balayage unique séquentiel, de haut en bas et de gauche à droite, de l'image est effectué en vue de sa compression;

- Le mode *progressif*. Des codages et balayages successifs sont effectués en améliorant la précision de l'image, par exemple en augmentant au fur et à mesure le nombre de chiffres significatifs envoyés;

- Le mode *pyramidal* ou *hiérarchique*. Des codages multiples sont effectués en améliorant successivement la résolution spatiale, par exemple par pas de 2, de l'image. Il s'agit d'un sous-échantillonage;

- Le mode *sans perte*. Cette addition n'utilise pas les mêmes algorithmes que ceux développés dans les autres modes, mais permet de considérer JPEG comme un outil réellement généraliste.

La majorité des codeurs/décodeurs disponibles sont néanmoins généralement limité au mode séquentiel uniquement; on parle alors de codeurs "baseline". Leur performance en terme de compression varie entre 10 et 50, suivant le type d'image, la qualité souhaitée, la vitesse de compression/décompression désirée, le type d'implémentation utilisé ... On classifie typiquement ces taux de compression en quatre groupes (en bit par pixel):

0.08 Pour un taux de 1:200, l'image reste reconnaissable;

0.25 La qualité de l'image est moyenne, pour ce taux de 1:60;

0.75 Excellente qualité, avec un taux de 1:20;

2.25 L'image est visuellement identique à l'image originale (taux de 1:7).

Il est important de savoir que JPEG est plus particulièrement adapté aux images naturelles, i.e. différentes du dessin ou de texte. Pour ce dernier type d'image, des algorithmes sans perte utilisés généralement sur des données de type texte, comme *LZW*, qui reconnait des sous-chaines identiques dans un texte, appliqués au fichier image même, sont plus efficace. Cette dernière technique est utilisée dans le format *GIF* ("Graphical Interchange Format), largement utilisé dans Internet pour de images

quantifiées sur 8 bits. La précision de JPEG va elle jusqu'à 24 bits par pixel, 8 et 16 étant les plus communs.

La structure générale du compresseur JPEG avec perte est présentée dans la figure 1. Une première étape effectue une transformation orthogonale de l'image, par bloc de taille 8×8 pixels. Les matrices 8x8 obtenues par cette transformation sont alors quantifiées, transformées par un codage RLE, avant d'être enfin codées par un codeur de type Huffman. Les opérations inverses sont effectuées dans la phase de décompression.

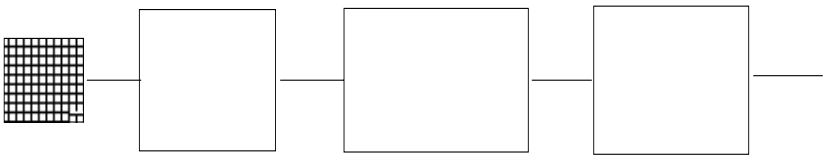

Figure 1: Principes de la chaine JPEG

La suite de cette section présente la transformation orthogonale (section 2.5.4) utilisée dans JPEG: la transformée en cosinus discret (DCT). Ensuite on détaille la procédure de quantification utilisée (section 2.5.4), le codage entropique (section 2.5.4), le mode sans perte (section 2.5.4) et les formats utilisés (section 2.5.4).

Transformée en Cosinus Discrète (DCT)

La Transformée en Cosinus Discrète DCT fait partie de la famille des transformations orthogonales du type de la transformée de Fourier. Il existe de nombreuses transformations orthogonales, telle la transformée de Fourier discrète DFT, la transformée de Haar, qui joue un rôle dans la compression fractale (section 2.7.2) ou la transformée de Karhunen-Loeve KLT. Elles permetttent de passer d'une fonction spatiale $f(x, y)$ à une fonction en fréquences $F(u, v)$.

La DCT offre un certain nombre d'avantages par rapport aux autres qui justifient son adoption pour la compression d'images; elle sert également de base à MPEG (section 2.6.1):

- Bien que la KLT représente l'optimum en terme de variance de distribution, calculée par la méthode des moindres carrés, et de taux de distorsion, la DCT en est proche;
- Contrairement à la DFT, la DCT ne nécessite pas de travailler avec des nombres complexes. En fait, on peut voir une DCT sur un signal comme une DFT effectuée sur le même signal augmenté de son signal miroir. Ainsi, un signal rampe est analysée comme un signal triangle, dont la replication (les transformées orthogonales de ce type travaillent sur des signaux infinis) présente un caractère plus "continu" que celui dérivé d'une rampe;
- Alors que la KLT ne possède pas d'algorithme rapide, la DCT, comme la DFT avec la FFT ("Fast Fourier Transform"), peut être effectuée en un temps moindre que $O(N^3)$. Néanmoins sa complexité temporelle théorique est moins bonne que la FFT.

Soit $f(x,y)$ une matrice de $N \times N$ éléments. La DCT de f est représentée par la matrice $F(u,v)$ de dimension $N \times N$ telle que:

$$F = CfC^t$$

où C est la matrice de transformation DCT, de dimension $N \times N$, définie par:

$$\begin{aligned} C(0,n) &= \frac{1}{\sqrt{N}} \\ C(k,n) &= \sqrt{2/N} cos(\frac{\pi(2n+1)}{2N}) \end{aligned}$$

Cette matrice de passage est unitaire, orthogonale et réelle: $CC^t = 1$. On en déduit que la DCT inverse (IDCT) de F est donnée par $f = C^tFC$.

Explicitement, on obtient:

$$\begin{aligned} F(u,v) &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} C(u,x) f(x,y) C(v,y) \\ &= \frac{1}{4} C(u) C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} cos(\frac{\pi(2x+1)u}{2N}) f(x,y) cos(\frac{\pi(2y+1)v}{2N}) \end{aligned}$$

avec:

$$\begin{aligned} C(0) &= \frac{2}{\sqrt{N}} \\ C(n) &= 2\sqrt{\frac{2}{N}} \end{aligned}$$

Réciproquement, pour l'IDCT:

$$\begin{aligned} f(x,y) &= \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u,x) F(u,v) C(v,y) \\ &= \frac{1}{4} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u) C(v) cos(\frac{\pi(2x+1)u}{2N}) F(u,v) cos(\frac{\pi(2y+1)v}{2N}) \end{aligned}$$

La complexité du calcul de F, en utilisant la multiplication naive matricielle, est $O(N^3)$. Il existe néanmoins des algorithmes rapides ("Fast DCT") pour lesquels la complexité temporelle passe à $O(N^2 log_2(N))$.

Il est important de constater que la transformation DCT n'induit pas de perte d'information. C'est une transformée complète et le signal original peut être retrouvé par transformation inverse. En particulier, l'énergie est préservée.

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j)^2 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} F(i,j)^2$$

La DCT n'est donc pas, en elle-même, source de compression.

Codage par DCT

L'utilisation de la DCT dans JPEG est limitée à des blocs de pixels de taille 8×8. La DCT $F(u,v)$ avec u et v dans $[0,7]$ d'un tel bloc est alors:

$$F(u,v) = \frac{1}{4} C(u) C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} base(u,x) f(x,y) base(v,y)$$

avec:

$$\begin{aligned} base(a,t) &= cos(\frac{(2t+1)a\pi}{16}) \\ C(0) &= \frac{1}{\sqrt{2}} \\ C(a) &= 1 \text{ sinon} \end{aligned}$$

L'intuition derrière l'utilisation de la DCT est qu'elle permet de décorréler une image, ou un bloc de celle-ci; elle détermine les informations indépendantes. De manière similaire à une transformation de Fourier sonore, qui est plus usuelle, et dans laquelle on détermine les harmoniques fondamentale et secondaires d'un son, la DCT réalise un analyseur harmonique spatial. L'IDCT, utilisée à la décompresion, réalise elle un synthétiseur harmonique.

Un bloc source est un signal discret, en 2 dimensions, de 64 points. La DCT le décompose en 64 signaux de base orthogonaux; chaque signal contient une des 64 "fréquences spatiales" qui forment le spectre de l'entrée. Le coefficient $F(0,0)$ de fréquence 0, appelé *coefficient DC*, correspond au signal moyen d'un bloc. Les autres coefficients, appelés *coefficients AC*, rafinnent ce coefficient en précisant les détails d'autant plus fins que (u,v) se rapproche de $(7,7)$. Le maximum de signal est présent dans les fréquences basses; l'image change graduellement dans l'espace. Insistons sur le fait que cette transformation est, mathématiquement, sans perte; la compression sera réalisée par quantification des coefficients de la DCT (section 2.5.4). Bien évidemment, des considérations d'implémentation interviennent pour, dès cette étape, introduire de la perte d'information. En effet, les fonctions transcendantes ne peuvent être calculées de manière exacte.

Plusieurs raisons motivent l'utilisation de blocs pour effectuer la DCT. La première est la complexité de l'algorithme de transformation. En se limitant à des blocs de coté 8, la complexité polynomiale reste performante. JPEG ne spécifie pas l'algorithme à utiliser (DCT, FFT sur 128 points, ...); des implémentations plus efficaces de la DCT restent un domaine ouvert de recherche. On considère que 25 % du temps de compresion JPEG est dépensé dans la calcul de la DCT.

Une autre raison du principe de blocage est que cela permet de restreindre l'influence des détails d'une image. Ainsi, une zone relativement uniforme sera codée avec peu de coefficients, tandis qu'une zone à grain fin pourra utiliser beaucoup de coefficients sans pénaliser le taux de compression sur l'ensemble de l'image.

Quantification

La quantification est le premier étage de JPEG responsable de compression. Cette quantification est effectuée sur les coefficients de la DCT par bloc à l'aide d'une table de 64 (8×8) entrée (fournie par l'application) à valeurs entre 0 et 255. Le choix de la table est fonction des pertes acceptées et se fait expérimentalement. C'est un des paramètres de compression et son choix est fonction de la distorsion considérée comme acceptable.

Cette quantification pourrait se faire sur l'image initiale, mais le passage par la DCT permet de déterminer les informations inutiles facilement. En effet, la majorité de l'information est présente dans les coefficients de basse fréquence. Rappelons également que le découpage en bloc permet de localiser les effets de la quantification et de diminuer la complexité. Un des inconvénients de ce découpage arbitraire en bloc est l'impression "cubiste" créée sur l'image décompressée quand il est nécessaire de dégrader la qualité de l'image pour maintenir le débit et la synchronisation.

Codage des coefficients DC et AC

Une fois les coefficients quantifiés, il reste à les transmettre. Le codage utilisé distingue les coefficients DC (c'est-à-dire $F(0,0)$, représentant le signal moyen du bloc) des coefficients AC (les autres, codant pour les aspects de plus en plus précis de l'image au fur et à mesure que l'on se rapproche de $F(N,N)$). Les coefficients DC, fortement corrélés d'un bloc à l'autre, puisque représentant le niveau moyen du bloc, sont codés en *différentiel*; on code non la valeur du DC, mais la différence avec le bloc précédent.

Les coefficients AC sont ordonnés en *zig-zag* pour faciliter le codage entropique. Les coefficients de basse fréquence apparaissent avant ceux de haute fréquence. En particulier, les zéros apparaissant fréquemment dans les hautes fréquences sont ainsi transmis en séquence. Ceci justifie le dernier étage du compresseur JPEG: un codeur RLE adaptatif, suivi d'un étage Huffman statique dont la table de codage est fournie par l'application.

Si le codage correspond au mode progressif, on peut envisager deux implémentations, apriori non exclusives:

- N'envoyer que les coefficients de basse fréquence, ce qui correspond à un simple filtre passe-bas;
- Limiter les valeurs des coefficients AC aux seuls bits de poids fort, puis, dans un second balayage, envoyer les bits restants.

Mode "compression sans perte"

Le mode de compression sans perte, introduit dans JPEG pour préserver la généralité du standard, utilise une technique de codage indépendante de la chaine vue précedemment. Cette méthode est prédictive et utilise les 3 pixels environnants (haut B, gauche A et diagonal gauche haut X) compatibles avec un parcours de type séquentiel. Diverses sélections parmi 8 sont possibles, comme par exemple le mode 1 dans lequel seul le pixel A sert de prédicteur ou 7 dans lequel on utilise la moyenne $\frac{A+B}{2}$. C'est au codeur à choisir le meilleur prédicteur pour un point donné. La différence entre le pixel prédit et la valeur réelle est encodée par un simple codeur entropique.

Ce mode est essentiellement utilisé dans les applications temps-réel, avec des images dont l'amplitude peut varier entre 2 et 16 bits. La qualité de la compression est meilleure que GIF sur les images à tons variables, mais moins bonne que ce dernier pour des dessins animés par exemple.

Formats d'Images

Le standard JPEG définit un cadre général de compression d'image ainsi qu'un format logique de représentation de données appelé "Source Image Format" (*SIF*). Pour JPEG, l'unité de base ou "data unit" est un bloc de 8×8 pris sur une image. Une image est formée de 1 à 255 plans de couleur ou canaux codés avec P bits: 8 ou 16 pour les modes utilisant la DCT; 2 à 16 pour le mode sans perte. Chaque canal a un taux relatif de sous-échantillonage par rapport à une valeur maximale. Le format précise également la table de quantification utilisée; une table unique est appliquée à toute l'image.

Si JPEG ne précise pas de format physique de fichiers, on trouve divers "pseudo" standards dans l'industrie. On donne ci-dessous quelques exemples de ces formats:

- JFIF (JPEG File Interchange Format). Il s'agit d'un format très simple; il n'y a pas d'en-tête ajoutée, mais uniquement le flux de données. Il est standard sur Usenet et est aussi utilisé dans l'outil de visualisation Quicktime de chez Apple;
- TIFF/JPEG (TIFF 6.0, extension de Alduss TIFF). Il s'agit ici d'un format plus complexe permettant de décrire complètement une image.

En ce qui concerne l'étendue de JPEG, celui-ci est un must dans de nombreuses applications:

- Adobe l'utilise dans le stockage de certaines fontes de type Postscript;
- Il sert de support à l'Office Document Architecture (ODA) spécifié par l'ISO;
- Le CCITT le considère pour le futur standard de télécopie couleur;
- De même, JPEG est un candidat pour le futur standard ETSI de videotex.

On le voit, JPEG est promis à un développement important dans un proche avenir. Il est d'ailleurs question de développer une nouvelle version de JPEG, tentativement appelée JPEG-2, qui, en empruntant des idées présentes dans MPEG-2 (quantification adaptative), permettrait d'améliorer encore les taux de compression déjà obtenus.

2.6 Compression d'Images Mobiles

Passer d'une image fixe compressée par JPEG à une séquence compressée d'images peut sembler simple. De fait, de nombreux produits sur le marché proposent l'extension naive de JPEG appelée *MJPEG*. Dans ce système, la compression d'une séquence d'images est simplement vue comme la séquence d'images compressées par JPEG, indépendamment les unes des autres. Ce système, s'il est simple à mettre en oeuvre et s'il présente des avantages comme la simplicité d'édition (postproduction), ne permet pas d'extraire toute la redondance présente dans le flux d'information. Les formats MPEG (section 2.6.1) et H.261 (section 2.6.3) prennent en compte la redondance *temporelle* entre une image et celles que l'entourent pour améliorer le taux de compression. On estime que l'utilisation de MJPEG est au moins trois fois moins efficace que MPEG, à qualité équivalente.

2.6.1 MPEG

Le groupe joint de travail JTC1 SC29 WG11, communément appelé *MPEG* (Motion Picture Experts Group), regroupant l'ISO et l'IEC, s'est attelé à la fin des années 1980 à proposer un standard de codage/compression adapté à la vidéo digitale qui est apparue dans le début des années 1990. Le standard est devenue une norme internationale (ISO/IEC IS 11172) au cours de 1993. De nouvelles extensions sont déjà en cours d'élaboration dans le cadre des projets MPEG-2 et MPEG-4 (section 2.6.2).

Introduction

Le standard MPEG (ou MPEG-1) recouvre trois aspects fondamentaux:

- MPEG-Video (11172-2). Est abordée dans cette partie la compression d'images couleur mobiles vidéo à 1.5 Mb/s (taux compatible avec les CD et DAT non compressés) sans entrelacement, de "qualité VHS";
- MPEG-Audio (11172-3). A été décrit dans la section 2.4.3;
- MPEG-System (11172-1). Ce standard spécifie la couche de niveau supérieure responsable de la synchronisation et du multiplexage de flot de données (son/image) compressés.

Plus récemment est apparu le "MPEG Conformance Testing" qui s'attache à spécifier les procédures de tests de conformité des outils MPEG proposés sur la marché.

Les applications du standard MPEG sont nombreuses, que ce soit pour les systèmes *symétriques* (i.e., où les deux équipements communicants ont besoin de compression) comme le vidéophone ou le téléphone, ou pour les systèmes assymétriques comme la publication électronique, les jeux vidéo et, bien entendu, les films (Video on Demand, satellite DBS,...).

De part les applications envisagées, les spécifications fonctionnelles d'un système utilisant MPEG sont classiques. Un tel système doit supporter, outre les opérations naturelles d'enregistrement (pour les systèmes symétriques) et de restitution, l'accès direct aux images, le déroulement rapide avant et arrière, la synchronisation audio (d'où le volet MPEG-Sytems), l'édition de courtes séquence, ... Il doit de plus être robuste aux erreurs de stockage et de transmission et être temps-réel (environ 150 ms). A noter que l'ensemble de ces spécifications est relativement contradictoire. Ainsi, obtenir un fort taux de compression nécessite de diminuer de manière importante la redondance temporelle, d'où une forte compression inter-images ... ce qui nuit à la facilité d'accéder directement à une image, demandant essentiellement une compression intra-image. MPEG permet d'opérer un compromis en utilisation à la fois prédiction (codage causal) et interpolation (non causal).

En combinant compression spatiale, basée sur une transformation orthogonale DCT (à la JPEG), et compression temporelle par compensation de mouvement, MPEG atteint une performance de l'ordre de 0.35 bits par pixel (1.2 en MJPEG). Comme dans JPEG, seul un format de codage est spécifié par MPEG; le meilleur algorithme permettant de l'obtenir est laissé à l'implémentation.

Dans la suite de cette section, on abordre successivement les techniques de compression temporelle (section 2.6.1), avant d'évoquer la compression spatiale (section 2.6.1) et les formats de représentation (section 2.6.1). Cette partie s'achève avec une présentation des futurs standards MPEG-2 et MPEG-4 (section 2.6.2).

Compression temporelle

La technique de base de MPEG pour diminuer la redondance entre images successives s'appelle la *compensation de mouvement* ("block-based motion compensated prediction" ou MCP). Celle-ci est faite par analyse du canal Y.

MPEG distingue trois types d'images:

- *Intra (I).* Les images I, utilisées pour faciliter l'accès direct, aussi appelées "key frames", utilisent uniquement la compression intra-image (sectio 2.6.1). Le taux de compression est donc faible.

- *Prédites (P).* Les images P codent la différence entre une image et une prédiction basée sur l'image précédente (de type I ou P). Seule cette différence est transmise.

- *Interpolées (B).* Les images B, ou bidirectionnelles, déterminent cette différence en utilisant deux images, une passée et une future, de type I ou P.

A noter que l'existence des pages B nécessite des réordonnancements complexes au sein du flux de données MPEG. En effet, il est parfois nécessaire d'envoyer des images futures avant les images courantes pour permettre au récepteur de calculer celles-ci quand elles sont de type B.

Les images sont analysées pour l'estimation de mouvement par *macroblocs* de taille 16. Les informations sont codées par différence avec le macrobloc précédent, puis compressées par un code de longueur variable comme Huffman. La taille des macroblocs (16) a de nombreux avantages comme:

- La compatibilité avec celle des blocs de DCT (8);

- Elle est déjà utilisée dans H.261 (section 2.6.3);

- Elle offre un bon compromis entre cout de l'algorithme de détection de mouvement et efficacité de la compression;

- C'est le plus petit commun multiple entre la taille des blocs Y et blocs UV, du fait de la décimation de la couleur;

- Enfin, elle est expérimentalement satisfaisante. Comme dans JPEG, des tests de qualité ont été effectués sur des populations d'évaluateurs qui jugeaient sans apriori les performances des différents algorithmes.

Le type d'algorithme de prédiction (images P) à utiliser dans MPEG est laissé à l'implémentation. De nombreuses techniques existent, que ce soit par filtrage ("bloc matching") ou recherche de points significatifs. Toute latitude est laissée à l'implémentation et le domaine reste ouvert.

Pour les images de type B, l'algorithme de calcul est spécifié dans le standard. On trouve ainsi, par exemple, des macroblocs "forward predicted", pour lesquels on utilise simplement l'image précédente, ou "average", où l'on utilise la moyenne entre l'image précédente et l'image future. La différence entre cette prédiction et l'image réelle est, comme dans les images de type P, transmise pour codage.

Il est utile, en particulier pour assurer la robustesse aux erreurs de transmission, faciliter l'édition (accès direct toutes les 0.4 secondes) et améliorer la compression, d'insérer de manière fréquente des images de type I. Aux US, on considère qu'une référence fixe toutes les 12 images est un bon compromis.

Compression spatiale

La technique de compression spatiale de MPEG est très fortement inspiré de JPEG. On retrouve tous les ingrédients d'une technique de compression basée sur une transformation orthogonale: DCT, quantification vectorielle, parcours en zig-zag, codage RLE et codage entropique à la Huffman.

Indiquons ici quelques améliorations propres à MPEG:

- La matrice de quantification est paramétrable et peut donc être adaptée à l'application (type d'affichage, distance de vision, quantité de bruit dans l'image originale,...);
- La quantification, uniforme, est différente pour les images I et les images P et B;
- La quantification est adaptative; le pas (intervalle de quantification uniforme) peut être changé à chaque bloc;
- Le codage RLE introduit des codes spéciaux pour indiquer la fin d'un bloc, diminuant ainsi les données à encoder;
- Le codage entropique final n'est pas Huffman, mais est plus proche d'un système "à la Fax Groupe 3" (section 2.5.2). Dans cette techique, utilisée aussi dans H.261 (rection 2.6.3), les tables entropiques sont uniques et optimisées à partir d'un nombre limité d'applications types.

Formats

MPEG ne fournit pas, de manière similaire à JPEG, de format physique de fichier, mais uniquement une organisation logique générique. Celle-ci permet de spécificier de multiples configurations, en variant des paramètres comme la fréquence des images de type I et P, l'entrelacement, ...

Un flux de données MPEG est formé de sept (7) couches fonctionnelles:

- *Séquence*. Elle permet de spécificier l'accès direct, de coder les dates des images (format SMPTE en heures:minutes:secondes:trames), ...;
- *Groupe d'images (GOP)*. Il s'agit des suites d'images encadrées entre deux images I;
- *Image*. Unité de base du codage intra-image;
- *Tranche* (ou "slice"). Unité de resynchronisation, typiquement correspondant à quelques lignes de l'image (une ligne de macroblocs);
- *Macrobloc*. Unité de compensation de mouvement;
- *Bloc*. Unité pour la DCT.

A noter que la notion de tranche est absente de JPEG; elle permet ici de récupérer les erreurs, de réinitialiser les codeurs entropiques et d'effectuer du "macrobloc stuffing" permettant de maintenir un taux de transmission (en bits/seconde) constant. Les GOP sont les unités typiques d'édition sur les systèmes MPEG; ceci pose des problèmes

(augmentation de la complexité des outils d'édition devant traiter les images de type I, P et B) quand il s'agit d'effectuer des éditions fines, à l'image près, des séquences enregistrées.

MPEG spécifie un certain nombre de paramètres décrivant une image et indique des bornes pour ces paramètres. Tout flux conforme à ces bornes, le "Constrained Parameters Bitstream" (CPB), doit pouvoir être traité efficacement avec les concepts MPEG. Parmi ceux-ci on note:

- Largeur et hauteur de l'image ($\leq 720 \times 576$);
- "Pixel aspect ratio". Typiquement 4:3 (pour la télévision) ou 16:9 (télévision HDTV);
- Taux de trame (≤ 30 image/seconde);
- Taux de transmission (typiquement, 1.5 Mb/s, et en tout cas, ≤ 1.86);
- Taille du buffer de décodage ($\leq 376kb$);
- Limites sur les vecteurs d'estimation de mouvement ($[-128, +127.5]$ en position et $[-64, +63.5]$ en luminance)

Enfin, une image ne doit avoir plus de 396 macroblocs. Ce jeu de paramètres a été déterminé pour tenir compte des performances des implémentations VLSI avec la technologie de 1992 (0.8 micron).

Un format commun pour MPEG, compatible avec le CPB, est le "Source Interchange Format" (SIF). Il existe en deux variantes, US et Europe. Nous indiquons ci-dessous celui commun à l'Europe:

- 50 Hz;
- 288 lignes de 352 pixels. Il s'agit donc d'un format CCIR-601, auquel on a enlevé 4 pixels de chaque coté de chaque ligne pour avoir un multiple de 16, décimé par un facteur 2 horizontalement, 2 dans le temps et 2 supplémentaire dans la chrominance. On nomme, illogiquement, ce format 4:2:0;
- Les coefficients DC et AC sont codés sur 8 bits;
- Les tables de quantification en luminance et chrominance sont identiques;

La qualité du format SIF est donc voisine d'une vidéo VHS. A noter que le modèle de codeur n'est pas imposé par MPEG.

2.6.2 MPEG-2 et MPEG-4

Si MPEG définit le Constrained Parameters Bitstream comme le cadre préférentiel d'utilisation des flux vidéo compressés, rien n'empêche de s'y limiter. Evidemment, MPEG n'est alors plus aussi efficace que l'on pourrait le souhaiter. C'est dans la perspective d'augmenter la performance de l'approche MPEG quand on sort des limites du CPB que de nouvelles versions, dites MPEG-2 et MPEG-4, ont été étudiées par le groupe d'experts de l'ISO.

MPEG-2, qui est sur le point d'aboutir (IS 13818 en Novembre 1994), est le fruit d'un groupe élargi à l'ITU (International Telecommunication Union), le SMPTE (Society of Motion Picture Technical Engineers) et la communauté HDTV aux US. Le but de cette association est d'étendre le protocole MPEG aux débits audio et vidéo entre 3 et 15 Mb/s. En somme, il s'agit de prendre en compte le format CCIR-601 sans décimation, i.e., 720 $\times$ 480 à 30 images/s, c'est-à-dire 4$\times$ SIF avec un débit jusqu'à 60 trames/seconde. Le codage doit prendre en compte l'entrelacement (ce qui prohibe l'utilisation du MPEG "de base") et le mode progressif. La qualité visée est celle du LaserDisc vidéo ou de la télévision de type "broadcast' ou "studio". Bien évidemment, MPEG-2 se doit d'être compatible ascendant avec MPEG-1.

Les améliorations aux principes de base de MPEG pour admettre ces spécifications sont essentiellement mineures. Nous les évoquons rapidement ici:

- Les algorithmes de prédiction de mouvement peuvent être plus sophistiqués car la résolution est descendue à 1/2 pixel;
- Plusieurs modes de parcours des coefficients AC sont utilisables. En particulier, cela permet de prendre en compte l'entrelacement;
- La précision des coefficients est accrue (paramétrable entre 8 et 11 bits);
- La quantification des blocs peut ne pas être linéaire;
- Les tables de codage entropique sont paramétrables (une par image I);
- On rajoute un quatrième type, les images DC, qui, en ne transmettant que les coefficients DC des images, permettent d'effectuer des "fast forward" en maintenant une lisibilité suffisante. Elles ne sont pas stockées dans le fichier;
- Par une meilleure gestion des buffers, le "macrobloc stuffing" n'est plus nécessaire;

Outre ces améliorations mineures, MPEG-2 introduit la notion nouvelle d'*échelle* ("scalability"). Elle consiste à découper le signal vidéo en couches de priorité qui peuvent être gérées de manière indépendantes, par exemple pour rajouter des codes correcteurs d'erreur pour les couches importantes. Il existe quatre (4) modes d'échelle:

- *Spatial.* Plusieurs résolutions sont disponibles sur le même canal. Une application de cette approche est le *simulcasting* dans lequel un signal TV classique est mixé à un signal HDTV);
- *Partionnement des données.* Il s'agit d'un mode similaire au mode progressif de JPEG. On peut par exemple découper les 64 coefficients en deux groupes (basses fréquences, plus DC, pour la couche prioritaire, puis le reste ensuite);
- *Rapport Signal/Bruit.* Ce mode est similaire au mode hiérarchique de JPEG.
- *Temporel.* On utilise ici des taux de trame différents pour le même signal. Ceci est par exemple utile dans le canal audio où la voie gauche peut être prédite de, et donc nécessite moins de débit que, la voie droite.

Si MPEG-1 définit le CPB, MPEG-2 classifie ses paramètres en *profil* (pour les algorithmes et la syntaxe) et *niveau* (paramètres). Sans rentrer dans les détails, les cas les plus fréquents correspondent, d'une part, au "Main Profile - Main Level" qui correspond au format CCIR-601 utilisé en vidéo de studio et, d'autre part, "Main Profile - High 1440 Level" qui permet la télévision HD de type broadcast.

De nouvelles extensions à moyen terme (1997) concernent les protocoles de communication interactive, le codage vidéo de haute qualité (10 bits) et l'audio.

Alors que MPEG-2 prenait en compte une amélioration de la bande passante, MPEG-4[4], prévu pour 1997, étend MPEG-1 dans les bas débits. Les applications concernées sont le vidéophone sur lignes téléphoniques analogiques ou le multimédia interactif mobil. Les débits visés sont de l'ordre de 4.8 à 64 kb/s, soit des images du type 176 $\times$ 144 à 10 Hz. Il est évident que des techniques de compression agressives seront nécessaires (voir section 2.7).

2.6.3 H.261

MPEG tire une grande partie de ses spécificités (transformée DCT, quantification, codage entropique) d'un produit plus ancien et plus simple appelé *H.261*, aussi appelé *P×64*. Ce standard de transmission vidéo numérique à des débits multiples de 64 kb/s (le débit unitaire des réseaux de type RNIS comme Numeris) a été élaboré par le Study Group XV du CCITT dès le mileu des années 1980. Cette norme de codeurs/décodeurs, appelés *codec*, fait partie de la suite de standard H.320, qui inclut:

- H.216, pour le codec vidéo;
- G.711, G.722 et G.728, pour les codecs audio (sections 2.4.1 et 2.4.2) ;
- H.221, pour la description de la structure des trames du canal de données commun.

Elle est en cours de normalisation par l'*ANSI*, l'American National Standard Institute, alors qu'un nouvelle version est en cours d'élaboration par l'IUT-TS (Telecommunication Sector).

De par les débits sélectionnés et la nécessité de travailler en temps-réel., H.261 est bien adapté aux systèmes de vidéophone (e.g., sur station de travail) ou de téléconférence. On considère que qu'un débit de 64 ou 128 kb/s (P = 1 ou 2) correspond à la première application, tandis qu'un P supérieur à 6 est nécessaire à la seconde. De fait, le marché actuel utilise essentiellement H.261 comme passerelle entre des systèmes propriétaires comme ceux de PictureTel (Massachussets, USA), qui possède 70 % du marché, ou de CompressionLabs (Californie, USA).

Le format promulgé par H.261 est le *CIF* ou Common Intermediate Format, que l'on voit également décliné sous sa forme Quarter-CIF ou *QCIF*. Un système ompatible H.261 doit au moins admettre le QCIF. QCIF est caractérisé par 144 lignes et 180 pixels par lignes, avec un sous-échantillonage de 2 en chrominance. Le taux maximum est de 29.97 trames par secondes, ce qui correspond à un débit, non compressé, de plus de 9 Mb/s en QCIF. En pratique, QCIF à 10 trames par secondes nécessite encore un facteur de compression de 1/47 pour pouvoir passer sur 64 kb/s!.

4. Il y a eu un bref MPEG-3 qui visait le marché de la HDTV. Il a été éliminé quand il s'est avéré que MPEG-2 s'adaptait déjà bien à ce type de média.

H.261 utilise des techniques très voisines de celles présentes dans MPEG. L'image est découpée en blocs de 8×8 pixels, regroupés en macroblocs, eux même regroupés en *GOB* ("Group Of Blocks") qui forment l'image (3 GOB pour QCIF). La compression H.261 n'utilise que des trames I et P. Les trames I sont codées en DCT, avec quantification linéaire de pas variable pour s'adapter au débit. La compression inter-trames utilise un codage DPCM (section 2.4.1) d'une estimation de mouvement entre macroblocs. En phase finale, un codage de type Huffman est utilisé. On voit donc que si H.261 n'est pas compatible avec MPEG, il est basé sur une technique trés similaire.

2.7 Techniques Avancées

Si les techniques à base de transformation DCT sont à la base des standards photo et vidéo utilisés à l'heure actuelle ou dans un avenir très proche, de telles méthodes peuvent s'avérer trop limitées pour obtenir des taux de compression encore supérieurs comme ceux envisagés pour MPEG-4. En vue d'atteindre ces objectifs, il convient de trouver des méthodes permettant, soit d'augmenter l'élimination de la redondance encore présente après traitement par JPEG ou MPEG, soit de perdre plus d'information, tout en essayant de préserver l'essentiel de ce qui est pertinent pour l'observateur. On considère qu'il faut alors se tourner vers des approches radicalement différentes comme les *ondelettes* ou la *compression fractale*. Ces deux techniques sont évoquées rapidement dans cette section.

2.7.1 Ondelettes

La compression par ondelettes peut être vue comme une analyse multispectrale d'une image. On applique une unique transformation orthogonale (à la DCT ou FT) sur une image complète, en variant hiérarchiquement la résolution de l'image.

Contrairement à la DCT (cosinus) ou la FT (exponentielle), la fonction définissant la transformée orthogonale *DWT* ("Discrete Wavelet Transform") appelée *ondelette* est de durée (ou d'étendue spatiale) finie. La DWT étend cette définition élémentaire $W_{0.0}(t)$ en faisant varier la fréquence (i.e., résolution) et le temps (i.e., espace) sur laquelle elle opère. Si i est la fréquence et j le temps, $W_{i,j}(t)$ est un signal défini par:

$$W_{i,j}(t) = W_{i,0}(t-j)$$
$$W_{i+1,j}(t) = W_{i,j}(2t)$$

Cette localité intrinsèque des ondelettes permet de coder l'image sans avoir à se résoudre à une découpe préalable en blocs de celle-ci comme dans la DCT. Un exemple typique d'ondelettes est la *transformée de Haar*, qui utilise des ondelettes carrées:

$$W_{0,0}(t) = 1 \quad \text{pour } -1 \leq t \leq 0$$
$$W_{0,0}(t) = -1 \quad \text{pour } 0 \leq t \leq 1$$
$$W_{0,0}(t) = 0 \quad \text{sinon}$$

Malgré son caractère simple, on peut d'ailleurs utiliser la transformée de Haar directement sur des images, avec des résultats meilleurs que la DCT sur des images à la JBIG. En pratique, une famille d'ondelettes discrètes ressemble à une base de Haar, mais plus adoucie avec évanouissement dans chaque direction, proche de la fonction

sinc $(sin(x)/x)$. On adapte le choix de la base au type d'images à coder; les ondelettes sont alors rarement connues de manière analytique. Elles sont définies comme des filtres sur le voisinage d'un point, dont on spécifie la matrice (orthogonale) des coefficients. La compression est alors obtenue par quantification, et élimination, des ondelettes de faible amplitude.

Des expériences préliminaires montre que la vitesse de traitement est proche de celle de la DCT, mais que la compression est meilleure d'un facteur 2. L'avantage essentiel de la compression par ondelettes est que la localité inhérente des fonctions $W_{i,j}$ permet d'éviter qu'un besoin très ponctuel de haute résolution, par exemple dans une petite région d'une image, entraine l'adoption de cette résolution pour l'ensemble de l'image. Les techniques par DCT aboutissent à un résultat similaire par découpage préliminaire, et sujet à artefacts, de l'image en blocs.

2.7.2 Fractals

La compression fractale consiste à représenter une (portion d') image par un *système de fonction itérées* ou IFS. Ces fonctions w_i doivent être *contractive*, i.e. telles $d(w(P_1), w(P_2)) < sd(P_1, P_2)$. L'image est alors le *point fixe*, ou attracteur, calculé par itération de Klenne, de l'union de ces fonctions w_i:

$$fix(\cup_i w_i) = \lim_{n \to \inf} \cup_i wn(x_0)$$

La compression vient du caractère concis des descriptions fonctionnelles (i.e., intensives) par rapport à l'image elle-même (i.e., extensive) et de l'approximation de l'image avec un IFS.

Le problème essentiel de la compression fractale est la détermination des w_i, qui se fait par recherche de similarité, modulo homothétie, entre diverses portions d'image. Une image est alors vue comme un ensemble de copies transformées d'elle-meme. Une approche (on parle alors de PIFS pour "Partitioned IFS") consiste à découper en blocs l'image et à restreindre les recherches de similarité à ces blocs, en y ajoutant un facteur de luminosité pour éliminer à terme les résidus. Cette recherche est en fait voisine de celle effectuée dans la quantification vectorielle (section 2.3.4) et on y retrouve des structures de données classiques comme par exemple les "quadtrees". Il s'avère néanmoins difficile de trouver de bons partitionnements.

C'est pour cette raison que les travaux sur la compression fracatale évoluent de plus en plus vers une méthode plus simple de "Fractal Transform", proche de la quantification vectorielle dans laquelle l'image elle-même sert de dictionnaire de codes: on découpe l'image en zones non recouvrantes et on détermine quelles transformations permettent de représenter les portions d'une image à partir d'un sous-ensemble. En pratique, on peut atteindre des taux de compression de 1/20 à 1/60, mais avec une qualité généralement moins bonne qu'avec JPEG ou la DWT. A noter également que ces techniques de compression font l'objet de brevets qui en limitent la diffusion.

Chapitre 3

Application au Projet BARRACUDA

Après avoir présenté l'expérience ALICE (section 3.1) et précisé les exigences du CNES en ce qui concerne ce segment Audio/Vidéo, nous détaillons dans la section 3.2 les divers systèmes disponibles sur le marché des stations de travail (ou terminaux X) et PC, en insistant sur l'évolution prévisible de ce marché à court et moyen terme. Cette étude du marché permet de suggérer une première configuration pour le projet BARRACUDA (section 3.3) que nous discutons.

3.1 ALICE

L'expérience ALICE se propose d'étudier, en conditions de microgravité dans la station MIR, l'évolution du comportement d'un liquide dans un enceinte thermostatée à température proche de sa température critique. Cette évolution est filmée sur une dizaine de cassettes vidéo Hi8 lors du vol; celles-ci doivent être analysées une fois le cosmonaute retourné sur terre. Le but de l'analyse, faite à l'aide d'un magnétoscope, est de corréler les images visionnées avec les mesures de températures effectuées de manière synchrone.

L'informatisation de cette phase d'analyse, objet en partie du projet BARRACUDA, va entrainer un couplage entre l'information vidéo projetée sur station de travail ou PC et la base de données techniques de mesures. Ceci se fera grâce à l'utilisation d'une base de données objet multimédia permettant de manipuler facilement données techniques et images. De manière plus précise, il est nécessaire de pouvoir:

- acquérir/numériser,
- compresser,
- décompresser/restituer

des images. Par ailleurs, et de manière optionnelle, il pourrait être avantageux de pouvoir effectuer les opérations suivantes:

- arrêt sur image,
- ralenti,

– avance et retour lent et rapide,

– zoom sur image.

A noter que l'aspect audio ne sera pas pris en compte lors de cette première expérience.

Du fait de la masse d'informations que représente de telles séquences d'images animées, il est nécessaire de faire appel aux techniques de compression vidéo présentées dans le chapitre 2 pour permettre stockage et manipulation en temps réel de ces données.

3.2 Les produits du marché

Si nous avons présenté dans le chapitre 2 les bases théoriques de la compression d'images et de son, il nous reste à découvrir quels types de produits sont proposés sur le marché pour implémenter ces techniques. Nous introduisons ici certains des produits actuellement disponibles sur le marché PC et Sun. Ces équipements sont relativement similaires et ne font généralement qu'ajouter un facteur (certes parfois important) de performance aux méthodes de compression décrites précedemment. Les besoins de calculs induits par les techniques de compression sont en effet importantes. Ceci justifie l'existence de produits performants, souvent accompagnés d'ajouts matériels.

L'essentiel des produits vise le marché de la vidéoconférence, des produits multimédia et des jeux. L'accent est donné sur l'aspect image de ces produits, sachant que les systèmes permettant de présenter des images mobiles offrent généralement la possibilité d'y insérer du son, vu le relativement faible besoin de bande passante de l'audio par rapport à la vidéo.

A coté des formats qui s'appuient sur une norme ISO (ou CCITT) et que nous avons décrit précédemment (JFIF et TIFF pour JPEG, MJPEG, SIF pour MPEG), on trouve sur le marché, essentiellement dans le monde PC d'ailleurs, d'autres formats propriétaires:

– Indeo d'Intel, utilisé surtout dans les système de videoconférence de la firme;

– Quicktime d'Apple, très voisin d'Indeo. Des implémentations compatibles sur PC et Sun (Xanim) sont disponibles;

– RTV (Real Time Video) et PLV (Production Level Video) sont utilisés sur les systèmes DVI (Digital Video Interactive) pour CD-ROM;

– AVI (Audio Video Interleaved) de Microsoft, utilisé dans Video for Windows de Microsoft.

– TrueMotion (ou TrueMotion S) de Duck Corporation Inc. pour CD-ROM. uniquement.

Ces formats de fichiers offrent parfois la possibilité d'utiliser plusieurs codecs de compression/décompression comme, en particulier, AVI qui peut accepter Indeo, Cinepak ou MPEG (en fait, trames I uniquement). En outre, des compatibilités (parfois avec pertes) sont possibles entre ces différents formats. Par exemple, XingCD est un compresseur et décompresseur logiciel offrant en outre la possibilité de coversion AVI/MPEG.

Ces formats propriétaire, dont l'influence a été relativement grande jusqu'à un proche passé, devraient voir une décroissance de leur part de marché, tout au moins professionnel, avec l'adoption de plus en plus massive de standards comme MPEG, et l'arrivée concomittante de cartes et progiciels dédiés à leur implémentation. Ainsi, la version 2 de Quicktime est compatible avec MPEG.

Ces divers techniques et formats de compression sont manipulables, soit par logiciels, soit par cartes matérielles. Ainsi, des codeurs MPEG logiciels sont disponibles en freeware pour évaluation sur Internet:

- `ftp.uu.net:/graphics/jpeg/jpegsrc.v5.tar.gz` est la version standard proposée par l'Independent JPEG Group;
- `ftp.netcom.com:/pub/cf/cfogg/mpeg!/vmpeg12a.zip` est une implémentation de MPEG-1 fournie par le MPEG Software Simulation Group;
- `ftp.netcom.com:/pub/cfogg/mpeg2` offre un codeur MPEG-2 fourni également par le MPEG Software Simulation Group.

Du fait du cout élevé en temps CPU des routines de manipulation, compression et décompression de séquences d'images, il reste encore souvent nécessaire, en phase de production, d'utiliser des produits matériels (processeurs DSP spécialisés de traitement du signal, cartes d'interface, ...). De nombreuses cartes d'acquisition et de traitement vidéo sont disponibles sur le marché. Il convient de bien y distinguer les cartes offrant la possibilité de compression, beaucoup plus chères, même de plusieurs ordres de grandeur, que les simples cartes de décompression. De plus, en ce qui concerne la compression, la qualité de celle-ci (taux de compression, rapidité de calcul, qualité de l'image, ...) est un facteur déterminant dans le cout final de la carte.

Pour les implémentations matérielles, on distingue plusieurs niveaux d'intégration:

Circuit C-Cube Microsystems, un des leaders du marché OEM, propose le circuit CL 450 (CPB jusqu'à 300 Ko/s) et CL 950 (CCIR-601 - 720x576, 30 fps - jusqu'à 1.2 Mo/s) pour effectuer les noyaux DCT de M(J)PEG. Ces circuits sont essentiellement utilisés sur des lecteurs de disque optiques compacts. Le dernier processeur de la firme, le CL 480, devrait donner lieu à un CL 480 PC orienté vers l'utilisation microinformatique. Le CLM4100 VideoRISC, juste annoncé à $495 en OEM, permet quatre modes de travail: acquisition et compression (jusqu'à 24 images/s) en temps-réel, décodage en temps-réel, vidéoconférence et "MPEG éditable", facilitant la manipulation de séquence MPEG (avance, arrière) avant d'effectuer un codage MPEG standard. Voir ci-dessous le système Realmagic Producer de Sigma Design.

Thomson propose la famille STI dans laquelle le 3400, voisin de CL-450, et le 3500, voisin de CL-950, sont compatibles MPEG-2.

LSI Logic fournit les L647*0 pour H.261 et MPEG (DCT, estimation, quantification, encodage);

Codeur/Décodeur Motorola propose le MCD250 un encodeur MPEG pour CD-I jusqu'à 5 Mb/s.

C-Cube Microsystems offre le CL 550 qui effectue une compression JPEG à 44.1 Mo/s.

Processeurs L'intel i750 sur la carte Action Media II est utilisé pour des cartes complètes d'acquisition, compression (RTV et PLV) et diffusion. On obtient des performances de l'ordre de 30 trame/s en 256x240 pour l'Action Media II.

Cartes Chez Sun, on trouve la carte SunVideo (circuit DSP sur S-bus, MJPEG et MPEG-1 à 30 trames/s, $1000) ou VCA-1 (DCT, estimation de mouvement, sur S-bus, $2500).

Optibase offre divers produits sur le marché professionnel: (1) MPEG ELS, temps réel, $5000. 24i/s vers VGA ; (2) MPEG Lab Pro, $19500, SIF, 25i/s PAL ou 30i/s NTSC; (3) MPEG Lab Suite, $18000, associé à la carte d'acquisition MDI-500 (incluant le Betacam numérique).

Optivision propose l'Optivideo Input Processor/Filter (avec préfiltrage), à $9995, ainsi qu'un décodeur Optivideo VGA MPEG, $895 (couleur 24 bits, compatible PAL/NTSC);

Pour l'utilisateur final, seules les cartes ou systèmes complets sont d'un intérêt immédiat. Celles-ci sont encore essentiellement dédiées à AVI, Indeo, JPEG ou MJPEG, quoique les produits purement MPEG voient leur nombre croitre rapidement. Sans souci d'exhaustivité (il y a plusieurs dizaines de produits disponibles), on a sélectionné ci-dessous certains de celles qui sont le plus cités dans la littérature et le milieu de l'image numérique, que ce soit sur le marché grand-public (PC) ou professionel (Sun):

Nom de carte/système	Constructeur	Format	Machine
Smart Video Recorder	Intel	Indeo	PC
Smart Video Recorder Pro	Intel	Indeo R3.0	PC
DC1	Miro	MJPEG	PC
Action Media II	IBM	RTV, PLV	PC
Xvideo 700	Parallax Graphics	MJPEG	Sun
Sun Video	Sun	MJPEG, MPEG	Sun
Video NT	Vitec	MPEG	PC
Realmagic Producer	Sigma Design	MPEG-2	PC

Exemples de cartes vidéo

Les couts moyen varient de $50 à plus de $10000 suivant les caractéristiques. Le système Realmagic Producer ($3995), outre sa compatibilité AVI, permet également d'effectuer des éditions préalables (enregistrement en images I uniquement), avant d'effectuer une compression en MPEG complet (images I, P et B); la vitesse de compression est de 3 minutes par minute de vidéo finale.

En terme de performance, les implémentations logicielles faites par les constructeurs sont généralement moins performantes que celles utilisant des accélérateurs matériels. A noter néanmoins que TrueMotion S, même à 640$\times$ 380 pour 30 images/s, est uniquement à base logicielle. La première version de TrueMotion utilisait un i750 pour accélerer les opérations. La license de TrueMotion S pour les applications PC non-ludiques a été acquise par Horizons Technology Inc. (San Diego). Egalement, le logiciel XingCD est équivalent en performance de décodage, sur un Pentium 90, à une carte RealMagic.

3.3 Solution proposée

Pour faire le lien entre les produits proposés sur le marché et les besoins introduits par l'expérience ALICE, il convient de bien noter que le choix du type de matériel utilisé influe peu sur les caractéristiques fonctionnelles du système final. Celui-ci joue essentiellement sur les performances, en terme de qualité d'image, de taux de compression et de débit de restitution.

Par contre, plus crucial est le choix du format de compression utilisé pour représenter les images et du système logiciel de manipulation d'images. Le marché vidéo professionel, et son évolution prévisible à court et moyen terme, fait sans aucun doute préférer des solutions standards (MPEG-1, MPEG-2, MJPEG) aux formats propriétaires rapidement évoqués ci-dessus, et ce d'autant que ces derniers sont souvent orientés vers une utilisation amateur ou seulement semi-professionnelle. Parmi ces standards, MPEG-1 semble à priori offrir le plus de souplesse ou de généralité, MPEG-2 ne se justifiant pas vu la qualité des images enregistrés sur magnétoscope. Rappelons d'ailleurs que MPEG-2 est essentiellement une extension de MPEG-1 à la diffusion d'images de haute qualité, de type TVHD. Néanmoins, MPEG pose encore des problèmes d'éditabilité, du fait de la présence des images P et B. Une solution qui semble donc attrayante est d'utiliser une carte MJPEG qui permet de concilier compression efficace et éditabilité à l'image près, comme la carte Xvideo 700 sur station de travail Sun. De plus, puisque l'adaptation du format MJPEG au format MPEG est simple (MJPEG est en fait déjà très voisin d'un format MPEG avec GOP à 1 trame), l'évolution vers MPEG sera facilitée si besoin est, par exemple avec un système comme RealMagic Producer.

Toutefois, le type d'images disponibles pour la seule opération ALICE, caractérisées par de forts contrastes, peu de mouvements discernables entre images et une absence de couleur, laisse à penser que le taux de compression avec MJPEG sera sans doute assez faible. En effet, d'après les besoins exprimés par le CNES, il peut être nécessaire de préserver une très bonne résolution; le phénomène d'intérêt prend en effet la forme d'une ligne très fine, à une échelle de temps de l'ordre de quelques millisecondes. MJPEG, du fait de artefacts créés par le traitement par blocs 8×8 des images, devra être utilisé en maintenant une qualité quasi-parfaite, du type de celle requise par les images médicales. Pour information, il est question d'autoriser MPEG et JPEG pour le stockage de ces images médicales, en maintenant un taux de compression de l'ordre de 1:4 pour être certain de ne pas perdre de phénomènes importants et de taille minime.

Il serait donc judicieux, si le taux de compression avec MJPEG s'avérait insuffisant, d'envisager d'autres techniques de compression comme GIF, JBIG ou même un simple fax de type Groupe 3 ou Groupe 4. La compagnie PixelMagic propose des implémentations logicielle (PC) et matérielle (ASIC PM-2) de JBIG (et aussi G3 et G4) permettant d'obtenir des débits de 2Mb/s sur Pentium et 33 Mb/s avec carte spécialisée. Malheureusement, comment ces standards sont seulement adaptés à des images fixes, ceci se fera sans prise en compte de la redondance inter-images qui est importante dans les échantillons visualisés. Il pourrait donc être avantageux, toujours dans le cas où les taux obtenus avec ces standards s'avéraient insuffisants, de développer des algorithmes ad-hoc qui combineraient compression d'images noir-et-blanc à fort contraste et hautre résolution avec de l'estimation de mouvement. Il ne semble pas exister à l'heure actuelle de système de compression de ce type.

Un autre conséquence de cette non-unicité probable du système de compression est

qu'un système comme Vidéo for Windows, qui permet de varier le type de codeur utilisé suivant la séquence, s'avérerait bien adapté au type de problème évoqué ci-dessus. Un tel système ne tourne pas sur stations Sun.