

INTRODUCTION AUX IMAGES NUMERIQUES ET AUX VIDEOS

2.1 Introduction

Ce chapitre ne prétend pas contenir une présentation complète sur l'imagerie numérique, sur la vidéo numérique et sur l'audio numérique, mais vise plutôt à donner une notion sur chacun de ces thèmes dans le but de nous familiariser avec les termes et pour mieux appréhender le chapitre suivant.

Trois grandes parties composent donc ce chapitre. La première consiste à dresser une analyse sur la structure d'une image numérique. La deuxième se focalise sur la vidéo numérique et pour terminer, on dédiera la troisième à l'audio numérique.

2.2 Les images numériques et le système visuel humain

L'étude du système visuel humain est intéressante pour le traitement d'images pour deux raisons principales :

- La première est qu'elle peut nous être utile nous mettre sur la voie de nouveaux algorithmes reflétant les mécanismes naturels.
- Et la seconde est qu'elle nous permet de connaître les limites de nos perceptions. Ainsi il est, par exemple inutile de représenter plus de couleurs que nous pouvons percevoir lors d'une application de visualisation.

Dans un système d'analyse d'images, on distingue la lumière captée par un récepteur (camera), transmise par des transmetteurs (câbles ou autres) à l'analyseur (l'ordinateur). On peut effectuer la même décomposition avec la perception visuelle.

- La lumière est captée par l'œil, l'information visuelle est transmise via les nerfs optiques vers l'analyseur qui est le cerveau.
- La perception visuelle est un mécanisme complexe qui met en jeu plusieurs structures : l'œil, la rétine et le cerveau.

- La compréhension de ce mécanisme repose sur la modélisation du SVH en vue d'en simuler son fonctionnement.

Le SVH est un système sophistiqué qui détecte et agit sur des stimuli visuels. Intuitivement, la vision par ordinateur et la vision humaine semblent avoir la même fonction. Le but des deux systèmes est d'interpréter des données spatiales. Même si l'ordinateur et la vision de l'homme sont fonctionnellement similaires, on ne peut pas s'attendre à un système de vision par ordinateur pour reproduire exactement la fonction de l'œil humain.

Cela s'explique en partie parce que nous ne comprenons pas entièrement comment l'œil fonctionne. En fait, certaines des propriétés de l'œil humain sont utiles pour élaborer des techniques de vision par ordinateur, alors que d'autres sont en fait pas souhaitables dans un système de vision par ordinateur.

Mais il existe des techniques de vision par ordinateur qui peuvent être reproduites dans une certaine mesure et, dans certains cas, améliorées même sur le SVH. Pour mieux comprendre ce qu'est une image numérique, voyons d'abord ce qu'est une image et comment fonctionne le SVH.

Dans le SVH, l'élément sensible est l'œil à partir duquel les images sont transmises via le nerf optique au cerveau, pour un traitement ultérieur. Le nerf optique a une capacité insuffisante pour transporter toutes les informations perçues par l'œil. En conséquence, il doit y avoir de prétraitement avant que l'image ne soit transmise par le nerf optique.

Le SVH peut être modélisé en trois parties :

- L'œil: il s'agit d'un modèle physique puisqu'une grande partie de sa fonction peut être déterminée par pathologie ;
- Le système nerveux: il s'agit d'un modèle expérimental, puisque sa fonction peut être modélisée, mais ne peut pas être déterminée avec précision ;
- Le traitement par le cerveau: c'est un modèle psychologique puisque nous ne pouvons pas modéliser le traitement directement, mais nous pouvons seulement déterminer le comportement par l'expérience et la déduction.

Tout d'abord, pour obtenir une image, il faut de la lumière. Cette dernière est émise d'une ou plusieurs sources telles que le soleil, des spots, des néons, etc. Cette lumière est représentée par des rayons qui partent de la source dans toutes les directions.

Généralement, lorsqu'un rayon de lumière rencontre un objet, ce dernier en absorbe une partie correspondant à sa couleur, et disperse le reste en une infinité de rayons qui peuvent éventuellement être captés par un œil annonçant la présence de l'objet ainsi que sa couleur. Pour recevoir ces rayons, l'œil est équipé d'un appareil optique complet illustré par la Figure 2.01 :

- L'*Iris* sert de diaphragme il s'ouvre et se ferme pour accepter plus ou moins de lumière.
- Le *Cristallin* fait la mise au point en fonction de la distance de l'objet [17] [18].

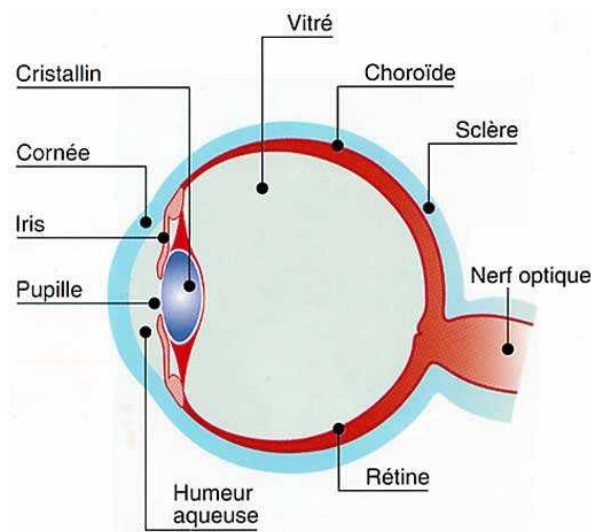


Figure 2.01 : *L'œil, notre capteur*

Finalement, cette lumière arrive sur des capteurs placés sur la rétine appelés cellules à cônes et cellules à bâtonnets du fait de leur forme. Les cellules à bâtonnets, plus sensibles, sont spécialisées dans la vision nocturne. Les cellules à cônes, plus précises, sont séparées en trois types, chacun étant plus sensible à une couleur qu'aux autres. C'est ce découpage de l'image en trois couleurs primaires qui permet la vision des couleurs.

Ces informations sont ensuite transmises au cerveau par le nerf optique. C'est le cerveau qui réalise ensuite la partie la plus complexe de regroupement de toutes ces informations pour former une image mentale en couleur de notre environnement.

Du point de vue fonctionnel, l'œil peut être comparé à un appareil photo et la rétine à la pellicule photographique comme illustré sur la figure 2.02. En effet, le rôle de l'appareil photo est de concentrer sur le film une image nette ni trop sombre ni trop lumineuse. On y parvient grâce à la bague de mise au point qui met l'objet au foyer et au diaphragme qui s'ouvre et se ferme pour laisser passer juste la bonne quantité de lumière pour la sensibilité du film [20].

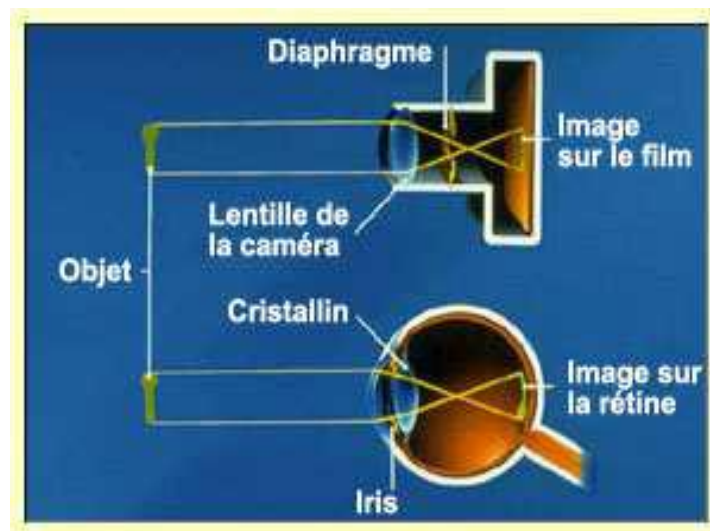


Figure 2.02 : *Analogie entre l'œil et l'appareil photo*

L'animation des images est basée sur le phénomène suivant : Lorsqu'une cellule capte de la lumière, l'impression lumineuse persiste pendant environ 1/50s. En effet, quand l'image change rapidement, l'œil n'est pas assez rapide pour percevoir une succession d'images fixes et croit voir un mouvement continu ([21], [19]).

Image réelle/image numérique

Voyons maintenant comment transformer une image réelle en une suite de chiffres compréhensible par un ordinateur.

En fait, il suffit de s'inspirer de l'œil humain avec ses cellules à bâtonnets et ses cellules à cônes. L'image ne pouvant être analysée de façon continue, son intensité est analysée à intervalles réguliers. Plus il y a de capteurs, plus l'image est précise. L'image numérique fonctionne sur ce principe. Elle est découpée en de nombreux petits points appelés pixels comme le montre la figure 2.03. Pour chaque élément, on attribue une intensité lumineuse. La qualité de l'image dépend d'une part du nombre de pixels, et d'autre part du nombre de valeurs possibles pour l'intensité [22].

2.2.2 Numérisation des images

Le terme d'image numérique désigne, dans son sens le plus générale, toute image qui a été acquise, traitée et sauvegardée sous une forme codée représentable par des valeurs numériques.

La numérisation est le processus qui permet le passage de l'état d'image réelle qui est caractérisée par l'aspect continu du signal qu'elle représente, à l'état d'image numérique qui est caractérisée par l'aspect discret, c'est à dire, l'intensité lumineuse ne peut prendre que des valeurs quantifiées en un nombre fini de points distincts.

C'est cette forme numérique qui permet une exploitation ultérieure par des outils logiciels sur ordinateur. Du point de vue mathématique, une image réelle est généralement représentée par une fonction bidimensionnelle représentant des caractéristiques particulières du signal lumineux de l'image en chaque point de son espace (intensité, couleur, etc.).

2.2.3 Représentation mathématique sous forme matricielle

Une image numérique 2D est représentée par un tableau f de n lignes et m colonnes. Le pixel est désigné par un couple (i, j) où j est l'indice de colonne $j \in \llbracket 0, m-1 \rrbracket$, et i est l'indice de ligne $i \in \llbracket 0, n-1 \rrbracket$, m est la largeur, n est la hauteur de l'image f . Par convention le pixel d'origine $(0,0)$ est

en général en haut à gauche. Le nombre $f(i, j)$ est la valeur du pixel (i, j) , $f(i, j) \in \llbracket 0, N_{\max}-1 \rrbracket$, $N_{\max}$ est le nombre de niveaux de gris. On appelle dynamique de l'image le logarithme en base de $N_{\max}$, .i.e, le nombre de bits utilisés pour coder l'ensemble des valeurs possibles [6].

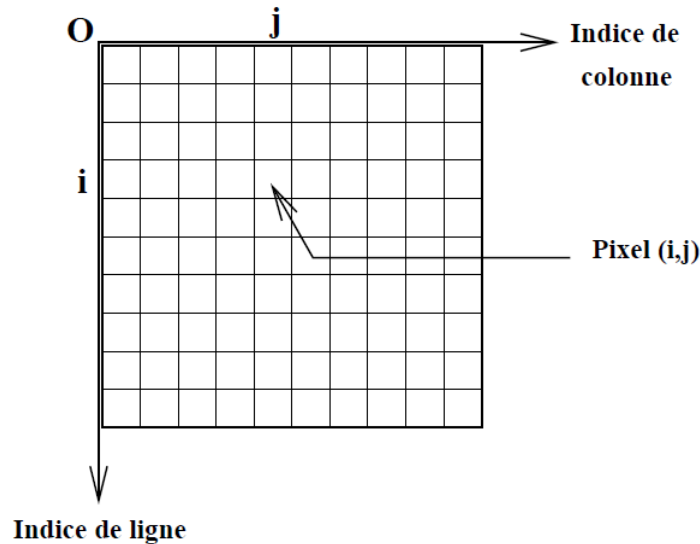


Figure 2.03 : *Représentation mathématique d'une image sous forme matricielle*

2.2.4 Processus de numérisation

La représentation informatique d'une image est nécessairement discrète, alors que l'image est de nature continue : le monde est continu. Si on regarde un peu plus près, la transformation d'un signal analogique 2D nécessite à la fois une discrétisation de l'espace : c'est l'échantillonnage, et une discrétisation des couleurs : c'est la quantification.

Le processus de numérisation d'une image suit les étapes suivantes :

–**Echantillonnage** : l'échantillonnage est le procédé de discrétisation spatiale d'une image consistant à associer à chaque pixel $R(x, y)$ une valeur unique $I(x, y)$ illustré par la figure 2.04. On parle de sous échantillonnage lorsque l'image est déjà discrétisée et qu'on diminue le nombre de pixels [6].

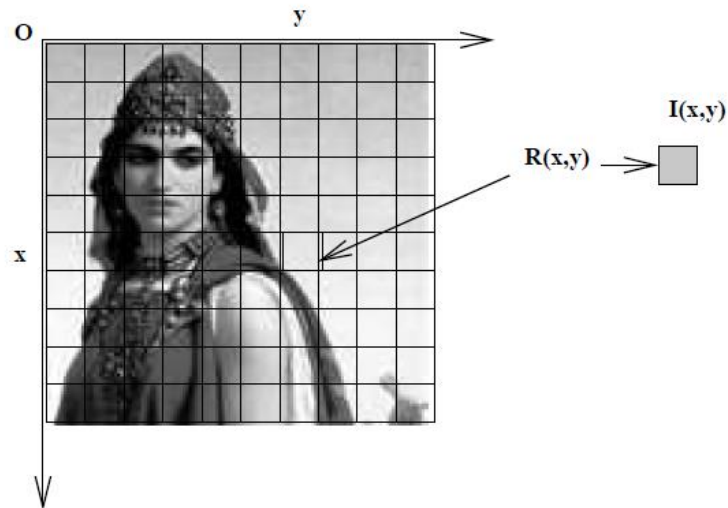


Figure 2.04 : *Echantillonnage, discrétisation spatial*

– **Quantification** : la quantification désigne la discrétisation totale correspondant à la limitation du nombre de valeurs différentes que peut prendre chaque pixel. Idéalement, le nombre de valeurs différentes devrait dépendre de l’amplitude des grandeurs observées (réflectance de la lumière visible, luminance infrarouge, ...) dans la scène. Mais en pratique, le nombre de valeurs utilisées pour coder une image lors de son acquisition dépend de la capacité effective du capteur à observer des signaux de grandeurs différentes, qui s’assimile au rapport signal sur bruit ([24], [25]).

2.2.5 Fidélité de la numérisation

La fidélité de la représentation de l’image numérique par rapport à l’image modèle analogique dépend de nombreux paramètres comme : la résolution, la définition de l’image ou la dimension de l’image, l’échantillonnage et la qualité de stockage [23].

2.2.5.1 La résolution

La résolution est le nombre de pixels par unité de longueur dans cette image. Plus la résolution est élevée, mieux les détails seront représentés. Ce paramètre dépend principalement des caractéristiques du matériel utilisé. Pour les moniteurs d’ordinateurs, la résolution est exprimée en nombre de pixels par unité de mesure (pouce ou centimètre). On utilise aussi le mot résolution

pour désigner le nombre total de pixels affichables horizontalement ou verticalement sur un moniteur.

La résolution standard pour l’affichage Web est de 72 dpi : c’est la résolution maximale supportée par un écran jusqu’à maintenant. A titre informatif, la résolution des images destinées à l’impression est de 150 dpi pour l’impression qualité journal. Cette résolution peut facilement monter jusqu’à 1200 dpi pour une impression en qualité photo ([26], [23]).

2.2.5.2 Définition de l’image ou dimension

La définition de l’image est le nombre fixe de pixels qui est utilisé pour représenter l’image dans ses deux dimensions. Pour une image analogique donnée, plus la définition est grande plus la précision des détails seront élevée. Ce nombre de pixels détermine directement la taille des informations nécessaire au stockage de l’image (du fichier numérique brut). La dimension, en pixels, détermine le format d’affichage à l’écran (la taille des pixels de l’écran étant fixe) [27].

2.2.5.3 Echantillonnage

La quantification détermine la qualité de l’échantillonnage du signal. Celui-ci se mesure en nombre de bits par pixel de l’image (bpp). La précision du rendu colorimétrique de l’image dépend du nombre de niveaux du signal pouvant être codés pour chaque pixel. Les valeurs les plus courantes sont 8 bits/pixel pour les images en niveaux de gris (256 niveaux de gris) et 24 bits/pixels, c’est à dire 8 bits par composante primaire, pour les images en couleur (plus de 16 millions de couleurs distinctes) [23].

2.2.5.4 Qualité de stockage

Le volume des informations qu’il est nécessaire pour stocker une image peut être très important, surtout dans le cas de l’utilisation d’images en haute résolution. Des techniques de compression doivent souvent être mises en place pour diminuer ce volume tout en conservant une certaine qualité de représentation.

Il existe des techniques de compression non destructives (basées sur des compressions de données sans perte d’informations et qui conservent l’intégralité du signal) et des techniques destructives

qui augmentent le taux de compression au prix d'une dégradation (généralement paramétrable) de la qualité de l'image. Un exemple de technique de compression destructive couramment utilisée est la compression JPEG [23].

2.2.6 Codage des images numériques

2.2.6.1 Codage en noir et blanc

Pour ce type de codage, chaque pixel est soit noir, soit blanc. Il faut un bit pour coder un pixel (0 pour noir, 1 pour blanc). Ce type de codage peut convenir pour un plan ou un texte mais on voit ses limites lorsqu'il s'agit d'une photographie.



Figure 2.05 : *Image codée en noir et blanc*

2.2.6.2 Codage en niveau de gris

Si on code chaque pixel sur 2 bits on aura 4 possibilités (noir, gris foncé, gris clair, blanc). L'image codée sera très peu nuancée.

En général, les images en niveaux de gris renferment 256 teintes de gris. Par convention la valeur zéro représente le noir (intensité lumineuse nulle) et la valeur 255 le blanc (intensité lumineuse maximale). Le nombre 256 est lié à la quantification de l'image. En effet chaque entier représentant un niveau de gris est codé sur 8 bits. Il est donc compris entre 0 et $2^8 - 1$. C'est la quantification la plus courante. On peut coder une image en niveaux de gris sur 16 bits ou sur 1 bit : dans ce dernier cas le «niveau de gris» vaut 0 ou 1 : il s'agit alors d'une image binaire (Noir et Blanc) [26].



Figure 2.06 : *Image codée en niveaux de gris*

2.2.6.3 Codage en couleurs 24 bits ou « couleurs vraies »

Il s'agit d'une appellation trompeuse car le monde numérique (fini, limité) ne peut pas rendre compte intégralement de la réalité (infinie). Le codage de la couleur est réalisé sur trois octets, chaque octet représentant la valeur d'une composante couleur par un entier de 0 à 255. Ces trois valeurs codent généralement la couleur dans l'espace RVB, c'est-à-dire, le rouge de 0 à 255, le vert de 0 à 255 et le bleu de 0 à 255. Le nombre de couleurs différentes pouvant être ainsi représenté est de $256 \times 256 \times 256$ possibilités, soit près de 16 millions de couleurs.

Comme la différence de nuance entre deux couleurs très proches mais différentes dans ce mode de représentation est quasiment imperceptible pour l'œil humain, on considère commodément que ce système permet une restitution exacte des couleurs, c'est pourquoi on parle de « couleurs vraies ». C'est ce codage de la couleur qui est utilisé par la plupart des écrans d'ordinateurs actuellement[28].



Figure 2.07 : *Image codée en 24 bits*

2.2.6.4 Codage en couleurs sur 8 bits

Pour réduire la place occupée par l'information de couleur, on utilise une *palette de couleurs* « attachée » à l'image. On parle alors de couleurs indexées : la valeur associée à un pixel ne véhicule plus la couleur effective du pixel, mais renvoie à l'entrée correspondant à cette valeur dans une table ou palette de couleurs appelée look-up table ou LUT en anglais, dans laquelle on dispose de la représentation complète de la couleur considérée.

Selon le nombre de couleurs présentes dans l'image, on peut ainsi gagner une place non négligeable : on considère en pratique que 256 couleurs parmi les 16 millions de couleurs 24 bits sont suffisantes.

Pour les coder, on aura donc une palette occupant $24 \text{ bits} \times 256 \text{ entrées}$, soit $3 \times 256 \text{ octets}$, et les pixels de l'image seront associés à des index codés sur un octet. L'occupation d'une telle image est donc de 1 octet par pixel plus la LUT, ce qui représente un peu plus du tiers de la place occupée

par une image en couleurs 24 bits (plus l'image contient de pixels, plus le gain de place est important, la limite étant le tiers de la place occupée par l'image en couleurs vraies) [28].

2.2.7 La représentation de la couleur

L'espace des couleurs primaires RGB est calqué sur notre perception visuelle. Il utilise trois couleurs de base : le rouge ($\lambda = 700\text{nm}$), le vert ($\lambda = 546\text{nm}$) et le bleu ($\lambda = 435,8\text{nm}$) ; où λ est la longueur de l'onde.

2.2.7.1 Synthèse additive de la lumière (mode RGB)

L'image est obtenue par superposition de trois rayonnements lumineux : le rouge (R), le vert (G) et le bleu (B). Dans le cas d'un écran cathodique, ces trois rayonnements sont obtenus en bombardant les luminophores photosensibles de l'écran.

Une image RGB est composée de la somme de trois rayonnements lumineux rouge, vert, et bleu dont les faisceaux sont superposés. A l'intensité maximale, ils produisent un ray de lumière blanche, et à l'extinction une zone aussi noire que l'éclairage ambiant le permet ([25], [28]).

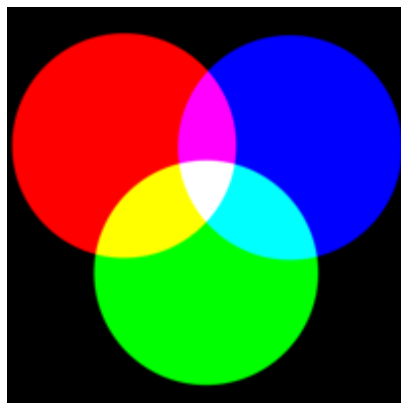


Figure 2.08 : *Synthèse additive ou mode RGB*

2.2.7.2 Synthèse soustractive ou mode CMJN

La synthèse soustractive permet de restituer une couleur par soustraction, à partir d'une source de lumière blanche, avec des filtres correspondant aux couleurs complémentaires :

- Cyan (C)
- Magenta (M)
- Jaune (J).

Ce procédé est utilisé en photographie et pour l'impression des couleurs.

Si on soustrait la lumière Magenta de la lumière blanche (par exemple par un filtre), on obtient de la lumière verte. Si on soustrait la lumière Cyan, on obtient de la lumière rouge et si on soustrait la lumière jaune, on obtient de la lumière bleue.

Si on soustrait à la fois la lumière magenta, Cyan et jaune (par exemple en superposant trois filtres), on n'obtient plus de lumière, donc du noir (que l'on note donc en toute logique : "N", comme Noir).

La gamme des couleurs reproductibles par le mode CMJN est plus restrictive que celle de la gamme RGB. Elle est, de surcroît, particulièrement sensible aux variations inévitables dues aux conditions mécaniques et physiques de l'impression en machine([25], [28]).

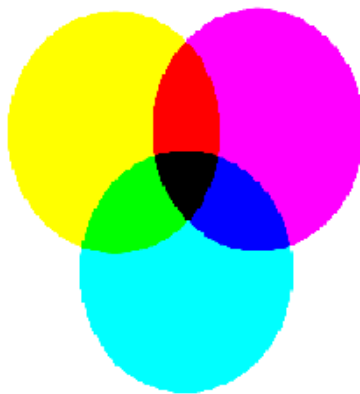


Figure 2.09 : *Synthèse soustractive de la lumière ou mode CMJN*

2.2.8 *Stockage des images numériques*

Il existe de nombreux formats plus ou moins performants et ne permettant pas de faire les mêmes choses. Par ailleurs, certains éditeurs de logiciel créent leur format propriétaire, l'interopérabilité n'étant souvent pas assurée.

Techniquement, on peut distinguer les images matricielles (bitmap) et les images vectorielles.

Les premières sont composées d'une matrice de points à plusieurs dimensions. En deux dimensions, cas le plus fréquent, les points sont nommés pixels tout comme sur un moniteur d'ordinateur.

Les images vectorielles de leur côté utilisent des formules géométriques décrivant le contenu de l'image à afficher. Ainsi au lieu de mémoriser un ensemble de points comme c'est le cas pour l'image matricielle, seront mémorisées les opérations conduisant au résultat. Si cette méthode présente de nombreux avantages, il n'en faut pas moins passer par une conversion de l'image vectorielle en représentation matricielle pour l'afficher sur les moniteurs d'ordinateur actuels.

Les applications des images vectorielles sont multiples. Elles sont en effet très utilisées pour des applications de visualisation scientifique ainsi que pour la création Web (format flash), la PAO (Publication Assistée par Ordinateur) et surtout l'illustration.

Ceci est en effet dû à plusieurs raisons. La première vient de la taille des fichiers. Ceux-ci sont en effet très peu volumineux en comparaison des images bitmap. La seconde vient de la qualité et de la précision des images. Cela vient de la manière dont sont créées ces images. Comme son nom l'indique une image vectorielle est faite de vecteurs. Ainsi, pour créer une droite, il suffit de déterminer les coordonnées d'un des points de la droite ainsi que son orientation.

Pour créer un segment, les coordonnées de début et de fin de segment suffisent. Un cercle sera défini par son centre et son rayon, etc. De même, les couleurs sont réparties en fonction d'équations mathématiques. Si l'on veut faire un dégradé, le principe est le même. Une autre chose très intéressante en dessin vectoriel, c'est que les objets ne s'écrasent pas entre eux. Chaque objet crée existe. Il faut alors définir pour chaque objet sur quelle couche il se situe, les zones dessinées des couches les plus élevées masquant les zones des couches les plus basses. Ceci a comme énorme avantage que si l'on veut modifier des objets ou modifier la taille de l'image, la

qualité restera la même. En effet, il suffit de recalculer les dimensions de chaque objet et les zones de couleur. Ainsi, il n'y a pas de perte d'information ([28], [29]).

Nous s'abstiendrons d'élaborer les formats d'image vectorielle car on s'intéresse surtout aux images matricielles dont voici quelques exemples de formats les plus utilisés :

2.2.8.1 Le format JPEG

Ce format est l'un des plus complexes, son étude complète nécessite de solides bases mathématiques, cependant malgré une certaine dégradation il offre des taux de compressions plus qu'intéressants.

JPEG est la norme internationale (ISO 10918-1) relative à la compression d'images fixes, notamment aux images photographiques. La méthode de compression est "avec pertes" et s'appuie sur l'algorithme de transformée en cosinus discrète DCT. Un mode "sans perte" a ensuite été développé mais n'a jamais été vraiment utilisé. Cette norme de compression a été développée par le comité JPEG (*Joint Photographic Experts Group*) et normalisée par l'ISO/JTC1 SC29. Ce type de compression est très utilisé pour les photographies, car il est inspiré des caractéristiques de perception visuelles de l'œil humain.

Le JPEG2000 est la norme internationale (ISO 15444-1). Elle apporte quelques améliorations au JPEG classique et notamment permet un réglage autorisant une compression sans perte ou encore la résistance aux erreurs de transmission. JPEG 2000 est relative à la compression d'images qui s'appuie sur un mécanisme de compression par ondelettes [30].

2.2.8.2 Le format GIF

Le format GIF pour *Graphical Interchange Format* été créé en 1987 par CompuServe pour que les utilisateurs puissent s'échanger des images de façon efficace et moins onéreuse. Ce format permet une compression sans perte (algorithme LZW).

Quelques problèmes juridiques avec la société Unisys détenant un brevet sur le LZW et donc revendiquant des royalties sur le GIF ont favorisés le développement de nouveau format à l'instar du PNG. Ces brevets ont aujourd'hui expiré faisant tomber le GIF dans le domaine public.

Ce format fonctionne sur la base d'une palette de 256 couleurs indexées (8 bits), le GIF est de fait limité à seulement 256 couleurs. Il autorise une bonne compression et une décompression très rapide grâce à la méthode LZW. Cette compression est plus efficace pour les dessins et graphiques que pour les photographies numériques ([25], [30]).

2.2.8.3 Les formats PNG et MNG

Le PNG pour Portable Network Graphic (ISO 15948) a été développé par le W3C pour remplacer le GIF. Il surpasse ce dernier en ce qu'il n'est notamment pas limité à 256 couleurs. De même, le format est ouvert et permet une bonne compression sans perte. Son utilisation est recommandée à l'instar du GIF pour les petits logos. Côté photo, s'il permet une compression sans perte, le poids de la photo n'est pas compétitif avec les formats JPEG. Précisons que le PNG ne gère pas l'animation mais un format dérivé, le MNG, y est destiné. Ces formats ne sont pas encore démocratisés, et le MNG notamment nécessite l'adjonction de plugins [30].

2.2.8.4 Le format TIFF

Le TIFF pour *Tagged Image File* a été mis au point en 1987 par la société Aldus (appartenant désormais à Adobe). Les dernières spécifications (Révision 6.0) ont été publiées en 1992.

Le format TIFF est un ancien format graphique, permettant de stocker des images bitmap (raster) de taille importante (plus de 4 Go compressées), sans perte de qualité et indépendamment des plates-formes ou des périphériques utilisés (Device-Indépendant Bitmap, note DIB). Il supporte différents types de compression autant avec que sans perte de données.

Le format TIFF permet de stocker des images en noir et blanc, en couleurs réelles (Truecolor, jusqu'à 32 bits par pixels) ainsi que des images indexées, faisant usage d'une palette de couleurs [31].

2.3 La vidéo numérique

Une vidéo est une succession d'images à une certaine cadence. L'œil humain a comme caractéristique d'être capable de distinguer environ 20 images par seconde. Ainsi, en affichant plus de 20 images par seconde, il est possible de tromper l'œil et de lui faire croire à une image animée. On caractérise la fluidité d'une vidéo par le nombre d'images par secondes (en anglais *frame rate*), exprimé en *FPS* (*Frames per second*, en français *trames par seconde*).

La vidéo numérique représente l'information sous la forme d'un flux vidéo composé d'une succession d'images numériques.

Le format vidéo décrit l'ordre et la structure de ces images. Les données du flux vidéo, qui peuvent être accompagnées de sons sous la forme de flux audio, sont très volumineuses : elles doivent impérativement être compressées (codées) à l'aide d'un codec pour être stockées (sur disque dur ou sur les supports d'enregistrement : CD, DVD) ou/et transmises (et donc être adaptées au débit des réseaux).

Les flux vidéo (et le(s) flux audio éventuellement associé(s)), une fois encodés, sont généralement encapsulés dans des fichiers conteneurs : ces derniers permettent, notamment, leur lecture simultanée.

2.3.1 La compression

Comme on a vu précédemment que la vidéo accompagnée de son est très volumineuse. C'est pour cela que la compression est impérativement requise que ce soit pour le stockage ou pour la diffusion.

La compression est le processus par lequel des fichiers volumineux sont réduits en tailles en supprimant de ces fichiers toutes les informations redondantes. La compression va donc consister à déterminer ces redondances et à les éliminer en préservant la capacité de reproduire les images originales intactes ou avec le moins de dégradation possible.

2.3.2 *Les codecs*

Un codec désigne un moyen, souvent un logiciel et peut être une pièce d'équipement ou la combinaison des deux. Un codec (compresseur-décompresseur) est utilisé pour diminuer la taille et le débit de fichier qui serait autrement encombrant. Les codecs peuvent compresser les fichiers et les sauvegarder sous différents formats.

2.3.3 *Différents formats vidéo*

2.3.3.1 Formats non compressés

- Le standard S-Vidéo: Le standard S-Vidéo (pour « Sony Vidéo»), parfois appelé Y/C, est un mode de transmission vidéo à composantes séparées utilisant des câbles distincts pour faire transiter les informations de luminance (luminosité) et de chrominance (couleur). Une connexion S-Vidéo permet d'offrir une qualité de vidéo optimale en envoyant simultanément les 576 lignes de l'image, sans entrelacement (en une seule fois). Le signal S-Vidéo est habituellement transporté via un câble comportant un connecteur mini-DIN 4 broches (deux broches distinctes pour chaque composante vidéo). On trouve ainsi ce type de connecteurs sur les caméscopes ou magnétoscopes haut de gamme S-VHS, les caméras Hi8 ou tout simplement sur la majorité des cartes graphiques possédant une sortie TV. Cependant, le signal S-Vidéo ne permettant de transporter que le signal vidéo, il est nécessaire d'utiliser un câble audio séparé pour le transport des données audio. Il faut noter toutefois que les prises péritel récentes permettent également de transporter un signal S-Vidéo. Il existe d'ailleurs des adaptateurs permettant de connecter un câble S-Vidéo et des connecteurs RCA audio sur une prise péritel, au prix d'une dégradation supplémentaire du signal. Néanmoins, étant donné que le signal S-Vidéo ne faisait pas partie des spécifications de la prise péritel à l'origine, certains équipements possédant une telle prise ne supportent parfois pas ce type de signal.
- Le standard YUV (appelé aussi CCIR 601), auparavant baptisé YCrCb, est un modèle de représentation de la couleur dédié à la vidéo analogique. Il se base sur un mode de transmission vidéo à composantes séparées utilisant trois câbles différents pour faire transiter les informations de luminance (luminosité) et les deux composantes de chrominance (couleur). Il s'agit du format utilisé dans les standards PAL (Phase

Alternation Line) et SECAM (Séquentiel Couleur avec Mémoire). Le paramètre Y représente la luminance (c'est-à-dire l'information en noir et blanc), tandis que U et V permettent de représenter la chrominance, c'est-à-dire l'information sur la couleur. Ce modèle a été mis au point afin de permettre la transmission des informations colorées aux téléviseurs couleurs, tout en s'assurant que les téléviseurs noir et blanc existants continuent d'afficher une image en tons de gris. Voici les relations liant Y à R, G et B, U à R et à la luminance, et enfin V à B et à la luminance :

$$\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ U = -0.147R - 0.289G + 0.436B = 0.492(B - Y) \\ V = 0.615R - 0.515G - 0.100B = 0.877(R - Y) \end{cases} \quad (2.01)$$

Ainsi U est parfois noté Cr et V noté Cb, d'où la notation YCrCb.

- Le modèle YIQ est très proche du modèle YUV. Il est notamment utilisé dans le standard vidéo NTSC (utilisé entre autres aux États-Unis et au Japon). Le paramètre Y représente la luminance. I et Q représentent respectivement l'Interpolation et la Quadrature. Les relations entre ces paramètres et le modèle RGB sont les suivantes :

$$\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ I = 0.596R - 0.275G - 0.321B \\ Q = 0.212R - 0.523G + 0.311B \end{cases} \quad (2.02)$$

2.3.3.2 Les formats compressés

- MJPEG

Le codec vidéo MJPEG compresse la vidéo image par image, en utilisant la technologie JPEG appliquée à l'image fixe, et réunit ces images en mouvement et le son dans un même format de

fichier. Le MJPEG est le codec le plus utilisé pour les captures vidéo des ensembles cartes d'acquisition et logiciels d'édition vidéo. La conservation d'une bonne qualité d'image produit toutefois de gros fichiers. Le format MJPEG est un format non normalisé. Les solutions M-JPEG, au début des années 90, ont été développées sans concertations par des fabricants, conduisant à des solutions propriétaires et à des fichiers très souvent incompatibles entre eux.

- DV (Digital Vidéo)

Format vidéo mis au point en 1996 par un large consortium. Format d'enregistrement numérique utilisé par la plupart des caméscopes numériques. Il permet d'enregistrer des vidéos sur des cassettes en numérique avec une faible compression pour chaque image (et donc, une perte de qualité négligeable). Les cassettes DV existent en sept formats : DV, MiniDV, DVCAM, Digital8, DVCPRO, DVCPRO50 et DVCPRO HD. Le format DV utilise une compression de type M-JPEG, mais dont l'algorithme a été complètement normalisé à l'échelle mondiale.

- MPEG (Moving Picture Experts Group)

Les formats MPEG sont des formats de compression avec pertes pour les séquences vidéo. Le groupe MPEG (Moving Picture Experts Group), est un groupe d'experts créé en 1988 et chargé du développement de normes internationales pour la compression, la décompression, le traitement et le codage de la vidéo, de l'audio et de leur combinaison, de façon à satisfaire un large panel d'applications.

Les formats produits par MPEG sont ouverts, mais non libres : leur utilisation est soumise au paiement de redevances.

Ce groupe a développé les standards suivants :

- MPEG-1

Extension : .mpeg, .mpe, .mpg, .mpv, .dat. Norme de compression pour la vidéo numérique apparue en 1992 et utilisée pour stocker les films sur VCD ou bien pour diffuser de la vidéo sur internet. Qualité proche de la VHS. But visé : faire tenir 74mn de vidéo sur un VCD. Ce format offre une résolution à l'écran de 352×240 pixels (1/4 d'écran) à 30 images par seconde ou de 352

× 288 à 25 images par seconde avec un débit d'environ 1,5 Mbit/s. Ne code qu'un frame sur 2 (l'autre étant répétée) et donc pas d'entrelacement. Compression vidéo spatiale (de type MJPEG). Compression temporelle. Compression du son : Mpeg1 layer 2 :MP2. Compression du son : Mpeg1 layer 3 : MP3.

- MPEG-2

Extension : .mpeg, .mpe, .mpg, .mpv, .mp2, .m2p, .vob. Normalisé en 1994, le MPEG-2 est devenu le standard de compression de référence pour tous les secteurs de diffusion/distribution audiovisuelles. Utilisé par les SVCD et les DVD, le montage numérique, la télévision numérique terrestre (TNT, ou DVB-T) et la diffusion numérique par satellites (DVB-S) et le câble (DVB-C). Format orienté vers la qualité et le professionnalisme mais les fichiers MPEG-2 sont souvent volumineux. Résolution allant de 352x288 à 1920x1152. Compression du son : AAC

- MPEG-4

Extension: .mp4, .mov, .avi, .asf, .wmv. Normalisé en 2000, MPEG-4 est destiné à permettre le codage de données multimédia sous formes d'objets numériques (codage orienté objet), afin d'obtenir une plus grande interactivité, ce qui rend son usage particulièrement adapté au Web et aux périphériques mobiles. MPEG4 est la norme décrivant la gestion des contenus multimédia pour le Web. MPEG-4 autorise des taux de compression très élevés, mais pas très adaptés à un usage professionnel car la qualité de l'image n'est pas aussi bonne que celle du MPEG-2, DV, MJPEG, RAW, etc... Le Mpeg-4 a pour but de stocker et de diffuser des éléments tels que de la vidéo, de l'audio, de la 2D ou de la 3D.

Les usages de MPEG-4 englobent toutes les nouvelles applications multimédias comme le téléchargement et le streaming sur Internet, le multimédia sur mobile, la radio numérique, les jeux vidéo, la télévision et les supports haute définition.

- DivX

Extension : .avi. Codec vidéo propriétaire et fermé proposé par DivX Inc., conçu à partir de MPEG-4 part 2, ce dernier ayant été modifié afin d'y ajouter la possibilité de compresser le son au format MP3.

Cela permet ainsi d'obtenir des vidéos compressées très peu volumineuses avec une perte de

qualité raisonnable. Ainsi le format DivX permet de stocker un film complet de plusieurs heures sur un CD-ROM de 650 ou 700 Mo

- XviD

Extension :.avi. Codec MPEG-4 part 2. Format propriétaire et ouvert : le format XviD est une implémentation Open Source du codec Divx, développée à partir de 2001, à l'occasion du passage du format DivX original (OpenDivX porté par le groupe Project Mayo) à un format propriétaire.

- H264 ou MPEG-4 AVC

Advanced Vidéo Coding. Extension : dans un fichier conteneur .mp4, .avi. Format propriétaire et ouvert. Norme internationale de codage vidéo développée conjointement par l'UIT-T (Union internationale des télécommunications) et MPEG. Elle comprend de nombreuses techniques nouvelles qui lui permettent de compresser beaucoup plus efficacement les vidéos que les normes précédentes (H.261, MPEG-1, MPEG-2, MPEG-4 Part 2/ASP) et fournit plus de flexibilité aux applications dans un grand nombre d'environnements réseau. Le codec H.264/AVC est donc adapté à une très grande variété de réseaux et de systèmes (par exemple, pour la diffusion de la télévision, le stockage HD DVD et Blu-ray, le streaming RTP/IP, et des systèmes de téléphonie propres à l'UIT-T).

- Real vidéo

Extension: .rv, .rmvb (Realmedia), .rf (RealFlash) Codec vidéo propriétaire et fermé développé par Real Networks. Il est très utilisé pour délivrer des vidéos en streaming sur Internet.

- Theora

Codec vidéo ouvert et libre promu par la fondation Xiph.org. Theora est un des composants de leur projet de format d'encapsulation Ogg, qui a pour but de créer un ensemble de standards ouverts concernant le traitement de signaux multimédia (son, vidéo).

2.3.4 Les conteneurs vidéo

Un format conteneur (*wrapper* ou *container* en anglais) est un format de fichier qui peut contenir divers types de données et celles-ci sont compressées à l'aide de codecs normalisés. Le fichier conteneur est utilisé pour pouvoir identifier et classer les différents types de données. Les formats conteneur les plus simples peuvent contenir différents types de codec audio, tandis que les formats conteneur les plus avancés sont capables de gérer de l'audio, de la vidéo, des sous-titres, des chapitres et des métadonnées (ou tags) et de façon synchronisée pour que les différents flux soient bien lus en même temps [4].

Un conteneur vidéo est un format de fichier permettant de rassembler en un seul fichier : un ou des flux vidéo et un ou des flux audio. Les principaux conteneurs vidéo sont :

2.3.4.1 AVI

Extension : .avi. Format propriétaire et ouvert. Le format AVI (Audio Vidéo Interleave), développé par Microsoft, très répandu et lisible sur tous les lecteurs vidéo, c'est le format d'encapsulation le plus populaire. Dans un fichier AVI, chaque piste audio et/ou vidéo peut théoriquement être compressée par n'importe quel codec. Il est possible de créer des fichiers AVI n'utilisant pas de codec spécifique, le fichier contiendra simplement des données non-compressées, on appelle cela le format "RAW" (données brutes) : garantit un stockage de données sans perte de qualité mais génère des fichiers très lourds. Format vieillissant dont on ressent rapidement les limites, informations limitées, manque de flexibilité, problèmes de taille, etc. Une seule piste vidéo et jusqu'à 99 pistes audio (ce qui permet le doublage multilingue). Ne peut pas contenir de texte (pas de chapitres). Conteneur de flux audio (mp3, WAV, mp2) et vidéo (DivX, XviD, MPEG).

2.3.4.2 QuickTime

Extension : .mov, .qt, .qt, .qtx, .qtr, .qt3. Format propriétaire et fermé développé par Apple. QuickTime n'est pas un codec mais un environnement de développement multimédia. Il désigne à la fois un codec audio, un codec vidéo et un conteneur. QuickTime était à l'origine un format réservé aux Mac, mais a été accepté depuis par Windows. Format plus ancien qu'AVI mais plus

flexible. Il Peut contenir une ou plusieurs pistes, chacune comportant un type de données : des textes (sous-titres), de l'audio ou de la vidéo (animation, graphique, 3D, réalité virtuelle).

Une piste peut être également un stream (diffusion en temps réel par internet). Caractéristiques techniques proches des spécifications du MPEG-4. Le format QuickTime est lisible sur la majorité des plates-formes. Nécessite le lecteur QuickTime Player (gratuit) sur Mac et Windows. Supporte de très nombreux formats audio, image et vidéo. Notamment : formats audio (WAV, Midi, MPEG-1, AAC) et vidéo (DV, H.261, H.263, H.264, MPEG-2, MPEG-4).

2.3.4.3 ASF

Extension : .asf(Advanced Streaming Format) : Format récent développé par Microsoft. Format propriétaire et fermé. Format utilisé sur les successeurs du DVD : Blu-Ray et HD-DVD. Compatible avec les DRM (gestion des droits numérique). Très utilisé pour le streaming par internet. AVI amélioré : plusieurs pistes audio, vidéo et texte. Compression vidéo basée sur la norme MPEG-4 V2. Conteneur de fichiers audio (WMA) et vidéo (MPEG4, WMV).

2.3.4.4 WMV

Format conteneur utilisé pour la diffusion de la vidéo haute définition sur supports optiques (Blu-Ray, HD-DVD) et via les réseaux (téléchargement, streaming, VoD). Lisible sur MS Windows et MacOS. Il est très difficile de lire du contenu Windows Media sans Windows Media Player. Toutefois, certains lecteurs multimédias (comme MPlayer) sont capables d'exploiter les codecs Windows Media créés pour MS Windows, et rendre ainsi la lecture possible sur les systèmes Libres. Conteneur de flux audio (WMA) et vidéos (ASF, WMV).

2.3.4.5 MP4

Extension : .mp4, .mp4a, .mp4v, .m4P. Conteneur officiel pour la norme MPEG-4 : MPEG-4 ASP, MPEG-4 AVC (vidéo) et AAC (audio).

2.3.4.6 Flash Vidéo

Extension : .flv. Format propriétaire et fermé développé par Adobe System. Il est très utilisé pour délivrer des vidéos en streaming sur Internet. Peut mélanger vidéo et interface graphiques

interactives. Le contenu FLV peut être incorporé aux fichiers SWF. Solution de diffusion vidéo multiplateforme facilement intégrable via le plugin Flash Player dans son outil de conception d'applications et d'animations dynamiques pour le web.

Format utilisé notamment par Google Vidéo, You tube ou DailyMotion. Lisible avec les lecteurs: Macromedia Flash Player, VLC Media Player, Media Player Classic, Mplayer. Conteneur de flux audio (mp3, PCM, ADPCM) et vidéo (codec basé sur une évolution du codec H.263).

2.3.4.7 OGG

Extension : .ogg, .ogv, .oga. Format conteneur récent ouvert et libre créé par la fondation Xiph.Org, organisation à but non lucratif qui s'oppose au mouvement de privatisation dans le domaine des formats multimédia en développant des logiciels libres et des algorithmes de compression multimédia libres de brevet. Le format Ogg peut être stocké sous forme de fichier ou diffusé sous forme de flux. Il est lisible sur la majorité des plates-formes. Conteneur de flux audio (Flac, Musique : Vorbis, Voix : Speex) et vidéo (Theora, Dirac).

2.3.4.8 OGM

Extension : .ogm Format ouvert et libre. OGM est capable de gérer un flux vidéo, une ou plusieurs pistes son, des sous-titres et un chapitrage. Il permet la multipiste, le multi chanel, gère le multiplexage ainsi que la correction d'erreur. Il résulte d'une modification du conteneur Ogg, qui ne supporte que les formats issus du projet éponyme (Theora et Vorbis), lesquels sont libres. Mais ce format reste limité. Conteneur de flux audio (OGG Vorbis, MP3, WAV, ACC, FLAC, WAV) et vidéo (Theora, Xvid ou DivX).

2.3.4.9 Matroska

Extension : .mkv (1 piste vidéo + éventuellement autres pistes), .mka (audio uniquement), .mks (une seule piste). Conteneur vidéo ouvert et libre. Le format MKV est basé sur une structure dérivée de XML, appelée EBML (Extensible Binary Meta Language). Ainsi grâce au format Matroska, il est notamment possible de réaliser des fonctions de chapitrage, de créer des menus, de faire des recherches dans le fichier, de sélectionner une source sonore ou bien de choisir un sous-titrage. Flexible et multiplateformes. Conteneur de flux audio (Vorbis,AAC, MP3, AC3, PCM, WV, FLAC, MPC) et vidéo (DivX, XviD, Real Vidéo, H.264, Theora, etc.).

2.4 L'audio numérique

Un format de fichier audio est un format de données utilisé en informatique pour stocker des sons, (de la musique, des voix, etc.) sous forme numérique. De nombreux standards existent; certains s'appliquent à la production, au stockage et à la diffusion, d'autres (ceux qui utilisent des algorithmes de compression de données ou de débit), sont destinés, en principe, uniquement à la diffusion. Actuellement, le format le plus utilisé est de loin le mp3, suivi du wma, et de l'aac. [8]

2.4.1 La compression audio numérique

La compression audio numérique vise à réduire la quantité de données pour stocker la musique. Deux formats de compression sont possibles : la compression de donnée sans perte et la compression avec perte de donnée.

2.4.2 Le format de compression audio sans perte

La compression sans perte (lossless) signifie qu'on utilise un algorithme tel qu'on peut toujours retrouver les données d'origine. Dans l'absolu, il existe toujours un fichier d'origine tel que l'algorithme ne ferait pas gagner d'espace disque.

Typiquement, la compression sans perte permet de diviser la taille des fichiers par deux ou trois. Elle est relativement peu utilisée, car ce gain est très faible en comparaison de ceux permis par la compression avec perte (ce qui est un gros handicap pour les échanges de fichiers), et assez gourmande en temps de calcul. Aucun standard n'a donc suffisamment convaincu pour devenir universellement lisible. Voici quelques exemples de ces formats de compression sans perte :

2.4.2.1 ATRAC

(Adaptive TransformAcousticCoding) est une technique de compression audio avec et sans pertes développée par Sony en 1992. Ce format a subi plusieurs évolutions : ATRAC3, ATRAC3plus (familièrement écrit ATRAC3+) et ATRAC Advanced Lossless se sont succédé respectivement en 1999, 2002 et 2006.

2.4.2.2 FLAC

Le format FLAC (Free Lossless Audio Codec), est un format libre de compression audio sans perte. Maintenu par la fondation Xiph.org, il est apprécié pour conserver la qualité des fichiers sonores originaux en alternative aux formats de compression avec perte type MP3.

2.4.3 *Le format de compression audio avec perte*

La compression audio avec perte (lossy) se base sur des algorithmes spécialisés pour déterminer quelles transformations simplifient la représentation du son tout en étant perçue quasiment de la même manière par l'oreille humaine. Elle diminue la taille du fichier en éliminant les nuances perçues comme les moins utiles. L'élimination est définitive, créer un fichier dans un format de haute qualité à partir d'un fichier compressé avec perte ne sert strictement à rien. [8] en voici quelques exemples des formats de compression audio les plus utilisés :

2.4.3.1 Le Mp3

MP3 est l'abréviation de MPEG-1/2 Audio Layer 3. Cet algorithme de compression prend naissance en 1987. L'ISO en fera un standard dans les années 92-93. La couche (*Layer*) III est la couche la plus complexe. Elle est dédiée à des applications nécessitant des débits faibles (128 kbit/s) d'où une adhésion très rapide du monde Internet à ce format de compression. Les taux de compression (*ratio*) sont d'ordinaire de 1 pour 10 (1:10) (1:4 à 1:12). Très rapide à l'encodage. Des royalties importantes sont à payer pour exploiter la licence MP3. Utiliser l'encodeur MP3 LAME dernière version, encodé à 130 kbit/s (V5) permet d'obtenir une *qualité comparable* au AAC (Advanced Audio Coding) encodé à 48 kbit/s⁶.

Le suffixe des fichiers créés est .mp3

Type de compression : constant ou variable (VBR).

2.4.3.2 mp3PRO

Le format mp3PRO, fruit de la collaboration entre Thomson Multimédia et l'Institut Fraunhofer, combine l'algorithme MP3 et un système améliorant la qualité des fichiers comprimés appelé SBR pour Spectral Bandwidth Replication.

Ce format a été publié à la fin de 2001 ; un fichier MP3pro 64 kbit/s a une qualité équivalente à celle d'un MP3 à 128 kbit/s.

Le suffixe des fichiers créés est .mp3.

2.4.3.3 OggVorbis

Le format OggVorbis est un format libre, fruit de la fondation Xiph.org. Vorbis se différencie des MP3, WMA et autre AAC par son algorithme. Il segmente les sources audio en paquets successifs, l'algorithme de compression agissant dans un premier temps sur chaque paquet indépendamment des autres. Cela lui permet d'avoir très peu de faiblesses sur certaines fréquences et de conserver la même qualité quel que soit le type de musique.

Le suffixe des fichiers créés est .ogg ou parfois .oga.

Par abus de langage, on appelle fichier Ogg des fichiers musicaux compressés par l'algorithme Vorbis. Ceci peut être particulièrement déroutant à l'ère des baladeurs numériques supportant audio et vidéo. En fait, Ogg est un conteneur qui peut contenir des pistes sonores (Vorbis), audio sans perte (FLAC), audio parlées (Speex) et vidéo (Theora). Un fichier Ogg peut donc contenir l'un ou l'autre (ou une combinaison) de pistes. Pour être plus clair, nous devrions parler de fichier OggVorbis lorsque nous mentionnons un fichier .ogg qui ne contient qu'une piste sonore au format Vorbis.

2.4.3.4 WMA

Le format WMA (Windows Media Audio), créé par Microsoft à partir des recommandations MPEG-4 en 1999, est utilisé par le logiciel Windows Media Player. Ce format est lié à une gestion pointue des droits d'auteurs (Gestion numérique des droits, en anglais Digital Right Management ou DRM) qui permet de définir par exemple une durée de vie limitée pour les fichiers ou d'interdire les possibilités de gravure.

Il existe plusieurs versions du codec (wma7.1, wma9, wma pro).

Le suffixe des fichiers créés est .wma.

2.5 Conclusion

Dans ce chapitre nous avons présenté les images numériques, l'audio et la vidéo d'une manière générale. Ce chapitre nous a permis de connaître quelques notions et terminologies dans le domaine de la vidéo et de l'image telles que la numérisation, le codage et le stockage.