# Evaluation de la qualité des images

C. CHARRIER[1], C. LARABI[2], H. SAADANE[3]

[1] LUSAC – Saint-Lô – EA 2607

[2] SIC – Poitiers – FRE CNRS 2731

[3] IRCCyN – Nantes – UMR CNRS 6597

c.charrier@chbg.unicaen.fr, chaker.larabi@sic.univ-poitiers.fr

abdelhakim.saadane@polytech.univ-nantes.fr

## Résumé

*Ce cours a pour objectif d'aborder les différentes approches adoptées pour évaluer la qualité des images et des vidéos. Nous utiliserons la compression comme l'application phare afin d'illustrer tous les concepts étudiés. Le cours débutera avec une présentation des principales méthodes d'évaluation de la qualité définies par les différentes recommandations de l'ITU (Union Internationale des Télécommunications. Cette présentation précisera notamment comment sont analysées les données subjectives fournies par les différents observateurs. Les principales métriques de qualité, tant pour les images fixes que les vidéos, seront ensuite abordées en partant des métriques dites « mathématiques » (PSNR, MSE, etc.) pour terminer par les métriques psycho-visuelles intégrant quelques caractéristiques du Système Visuel Humain (SVH). La dernière partie du cours s'intéressera à montrer comment ces caractéristiques sont pris en compte dans les standards de codage tels que JPEG, JPEG2000 et MPEG1,2 pour obtenir les meilleures performances en termes de qualité débit.*

## Mots clefs

Qualité subjective, Métrique perceptuelle, SVH, Carte d'importance perceptuelle, JPEG, JPEG2000, MPEG4.

# 1 Introduction

L'apparition des nouvelles technologies de vidéo numérique, visant des niveaux de compression de plus en plus importants et entraînant la réalisation de multiples structures de codage, pose le problème de la qualité des images restituées. La qualité représente aujourd'hui une des clés du développement des applications et des services multimédias. En effet, la bande passante disponible étant limitée et chère, les opérateurs sont donc amenés à fixer d'abord les ressources disponibles pour le transport et la qualité souhaitée à la réception avant de choisir les algorithmes et les standards de compression dont ils ont besoin. Pour les applications TV, les télédiffuseurs choisissent un débit cible et fixent, en fonction de leur expérience, un certain nombre de paramètre pour l'encodeur. La conséquence de ces choix au niveau de la réception se traduit par une qualité perçue par l'observateur qui peut varier considérablement d'un contenu à un autre. Il en est de même pour les applications multimédias où quand les différents opérateurs parlent de qualité ils se réfèrent en fait à la qualité de service (QoS) qui détermine les caractéristiques du réseau indépendamment de la qualité perçue par les usagers. Même si, dans ce cas, les outils de mesure de la QoS contrôlent la redondance pour palier aux pertes éventuelles des bits et des paquets et surveillent le trafic pour éviter les congestions, ils ne permettent pas de réguler la qualité perçue par l'utilisateur. Des outils fiables de mesure de la qualité perçue sont donc nécessaires à tous les opérateurs pour maintenir la satisfaction de leurs clients et permettront également le développement de nouveaux services.

Il existe deux façons de mesurer la qualité d'une image. La première consiste à mener des tests subjectifs. Ces tests exigent un équipement approprié et des protocoles normalisés pour permettre des coopérations, échanges et comparaisons des résultats fournis par différents laboratoires. Les résultats de ces tests représentent la référence dans l'évaluation de la qualité. Toutefois ces tests sont lourds à mettre en oeuvre, chers et surtout très longs et ne constituent donc pas une solution pratique pour les différents opérateurs. Pour éviter un tel inconvénient, les métriques perceptuelles, qui représentent la deuxième alternative, ont pour objectif de définir des mesures de qualité qui soient fortement corrélées aux notes de qualité qu'auraient donné un ensemble d'observateurs. Jusqu'à un passé récent, ces métriques se limitaient à des métriques simples dont les performances restaient limitées. Le développement technologique de ces dernières années a permis la mise en place d'outils psychophysiques qui ont aidé à mieux comprendre le comportement du SVH et affiner les modèles associés. L'intégration de ces modèles dans les métriques perceptuelles est devenu alors un domaine de recherche et d'application très actif.

L'objectif de ce document est de vous donner une idée plus précise sur les problèmes liés à l'évaluation de la qualité et les solutions apportées à ces problèmes. Ce document est organisé comme suit. Le deuxième paragraphe donne un historique de la qualité et situe la notion de fidélité d'une image par rapport sa qualité. Le troisième paragraphe s'intéresse à l'évaluation subjective de la qualité d'images fixes et de séquences vidéos couleur. Il décrit ainsi les recommandations de l'ITU et tout ce qu'elles supposent comme environnement, méthodes, chronogrammes de test et analyse des données. Le quatrième paragraphe introduit le principe des métriques perceptuelles pour ensuite détailler leur fonctionnement aussi bien dans le cas des images fixes que celui des séquences vidéos. Il précisera notamment la différence des approches selon que l'on dispose des images de référence ou pas. Le cinquième paragraphe décrit les standards. L'accent sera mis JPEG et JPEG2000 pour les images fixes et sur MPEG4 pour les séquences vidéos. Le sixième paragraphe enfin, s'attachera à montrer comment ces standards exploitent les caractéristiques du SVH et les métriques perceptuelles pour améliorer la qualité des images codées.

# 2 Approche philosophique du terme qualité

Au premier coup d'œil posé sur un objet, l'être humain est capable de dire si sa vue lui est plaisante ou non. Il ne fait alors ni plus ni moins qu'une classification de la perception de cet objet en fonction du sentiment procuré et ressenti en deux catégories : « j'aime » ou « je n'aime pas ».

Une telle aptitude à classifier les sensations visuelles est indiscutablement à mettre en relation avec la conscience inhérente à chaque être humain. La conscience est liée à ce que Freud appelle «le système perception-conscience». Il s'agit d'une fonction périphérique de l'appareil psychique qui reçoit les informations du monde extérieur et celles venant des souvenirs et des sensations internes de plaisir ou de déplaisir. Le caractère immédiat de cette fonction perceptive entraîne une impossibilité pour la conscience de garder une trace durable de ces informations. Elle les communique au préconscient, lieu d'une première mise en mémoire. La conscience perçoit et transmet des qualités sensibles. Freud emploie des formules comme « indice de perception, de qualité, de réalité » pour décrire la teneur des opérations du système perception-conscience.

Ainsi la perception est à considérer comme une des échelles internes à un processus menant à une appréciation globale de la qualité d'un objet ou en l'occurrence d'une image.

Il est toutefois à noter que par abus de langage, il est souvent fait un amalgame entre les termes qualité et fidélité. La notion de qualité pourrait être à l'Artiste ce que la notion de fidélité serait au faussaire. L'Artiste travaille généralement à partir de concepts, d'impressions liés à son environnement social et/ou professionnel et se place dans un courant artistique existant (relation maître-élève) ou dans un nouveau courant qu'il créé de toutes pièces. Les œuvres réalisées sont ainsi considérées comme des originaux, et les experts parlent de la qualité de l'œuvre. Derrière cette approche, on se rend compte qu'au terme de qualité est associée la notion d'originalité. Qui ne s'est jamais retrouvé face à une œuvre qui la laissé perplexe tandis que son voisin s'en est émerveillé ? Il suffit de déambuler dans les musées pour constater ce phénomène. Ainsi on qualifie la qualité d'une œuvre en fonction de sa conscience et de sa sensibilité personnelle prédéfinie de par son environnement économique et social.

Le faussaire travaille généralement à partir d'un modèle et essaie de le reproduire avec la plus grande fidélité possible. Dans ce cas le faussaire se doit de fournir une pièce irréprochable et il n'est pas rare qu'il utilise les même techniques employées par l'auteur plusieurs siècles auparavant (combinaison de plusieurs pigments pour réaliser la couleur, utilisation d'une toile de la même époque, etc.). Dans ce cas, la copie doit être fidèle à l'original. D'aucuns pourraient prétendre que, dans certains cas, la copie peut dépasser la qualité de l'original…Pour anecdote, l'un des plus célèbres scandales dans l'histoire du marché de l'art a commencé par une banale petite annonce dans un journal anglais : « Fausses copies authentiques (XIXe et XXe siècles) pour 150 livres. » *John Myatt*, professeur d'arts plastiques aux abois, avait trouvé ainsi le moyen de s'enrichir. Un certain *John Drewe*, escroc génial et bon connaisseur des milieux artistiques, flaira l'affaire du siècle et s'associa sur-le-champ avec l'auteur de l'annonce. Le duo réussit à sévir durant presque dix ans. *Drewe* et *Myatt* trompèrent les plus grands experts, les maisons de vente Sotheby's et Christie's, et des musées de renom comme la *Tate Gallery* ou le *Victoria and Albert Museum*. Copies magistrales de *Giacometti* ou de *Dubuffet*, faux certificats d'authenticité, le système était parfait…

D'un point de vue plus pragmatique, la qualité d'une image est un des concepts sur lequel la recherche en traitement d'image prend de plus en plus une part prépondérante. Tout le problème consiste à caractériser la qualité d'une image, tel que peut le faire l'être humain. Dès lors, il nous faut dissocier les deux types de mesures :

1. Les mesures de fidélité
2. Les mesures de qualité

La mesure de fidélité, comme son nom l'indique, permet principalement de savoir si la reproduction de l'image est fidèle ou non à l'originale. Dans ce cas, on met en place une mesure qui calcule la distance entre les deux images. Cette distance symbolise numériquement l'écart qu'il peut y avoir entre les deux reproductions de l'image.

La mesure de qualité est à rapprocher de ce que fait naturellement et instinctivement l'être humain devant toute nouvelle œuvre : il lui donne une appréciation en fonction de sa conscience. Dès lors, l'être humain ne peut être dissocié de la mesure de la qualité. Ainsi, l'étude des mécanismes permettant de d'appréhender les échelles

internes utilisées lors de l'évaluation de la qualité par un être humain est devenue un domaine de recherche à part entière. Ainsi en 1860, *Gustav Theodor Fechner*[1] proposa de mesurer des événements physiques déclenchés intentionnellement par l'expérimentateur, et des réponses manifestes des observateurs, réponses obtenues selon des modèles spécifiés.

Dans un cadre très général, la psychophysique étudie les relations quantitatives démontrées entre des évènements physiques identifiés et mesurables, et des réponses évoquées selon une règle expérimentale avérée. Ces diverses relations sont ensuite interprétées en fonction de modèles, ce qui contribue à l'approfondissement de nos connaissances sur le fonctionnement de l'organisme par rapport à l'environnement.

Les méthodes psychophysiques permettent, en général, d'aborder l'étude de situations dans lesquelles le stimulus n'est pas définissable a priori mais où sa structure peut être déduite à partir de la structure des jugements des observateurs. L'élaboration de modèles de fonctionnement du système visuel humain est souvent le but recherché lors des expériences psychophysiques.

Lors de ces expériences, la distribution des réponses intègre une partie due aux processus sensoriels et perceptifs, et une partie relative aux processus d'élaboration des réponses. Cette idée de séparer ces deux composantes des réponses reflète l'influence de la théorie de la détection du signal, et de la conception de l'organisme soumis à une expérience, en tant que système de traitement de l'information.

Ces expériences sont couramment utilisées dans le domaine de la compression des images couleur dès lors que l'on souhaite quantifier, à l'aide d'un observateur humain, la qualité d'une image compressée. Dans ce cas, on utilise l'expression de « tests subjectifs de la qualité » qui répondent à un certain nombres de contraintes ; ces dernières étant détaillées en section 3.

Nul doute qu'avec l'avènement dans un futur proche du réseaux numérique hertzien et l'impact indéniable induit sur notre environnement socio-économique, la maîtrise de la qualité des images couleur transmises par le réseau soit devenue un enjeu industriel important pour les prochaines décades.

# 3 Evaluation subjective de la qualité

## 3.1 Images fixes

### *3.1.1 Environnement normalisé*

En concordance avec les standards ISO 3664 – "Viewing Conditions for Graphic Technology and Photography" et ITU-R 500-10 – "Methodology for the subjective assessment of the quality of television pictures", la normalisation de l'environnement d'évaluation subjective de la qualité d'images nécessite un certain nombre d'éléments essentiels. Deux environnements d'évaluation avec différentes conditions d'observations sont décrits dans les normes: l'environnement laboratoire et l'environnement bureau (utilisateur). Voici quelques conditions à contrôler dans l'environnement laboratoire :

- Distance d'environ un mètre entre le fond de la salle et l'écran ;
- Rapport de la luminance de l'écran inactif à la luminance de crête ;
- Luminance de crête de l'écran ;
- Éclairement de la salle (illumination ambiante) ;
- Chromaticité de l'arrière-plan correspondant à l'illuminant D65 ;
- Angle maximal d'observation relatif à la normale (écran CRT) de 30°;
- Moniteur d'évaluation de haute qualité, de taille 50-60 cm (22" – 26").

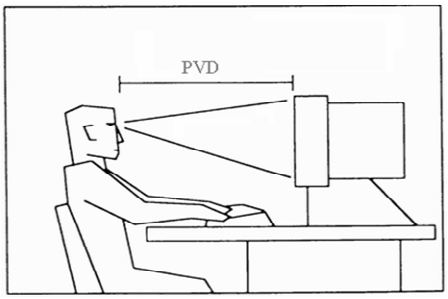



Figure 1 – *Conditions d'évaluation de la qualité*

Salle normalisée

Cette salle doit répondre à un certain nombre de critères, contrôlés par un organisme apte à délivrer un certificat de normalisation. Cet environnement permet d'effectuer des tests psychophysiques liés à la vision sans introduire d'erreurs relatives à l'environnement d'étude.

Figure 2 – *Exemple de salle d'évaluation*

Toutes les sources de lumière, autres que celles utilisées pour l'éclairage de la salle (tubes fluorescentes D65 d'intensité variable contrôlée), doivent être évitées car elles dégradent significativement la qualité d'image. L'écran doit être positionné de telle façon qu'aucune source de lumière, comme une lampe ou une fenêtre, ne soit directement dans le champ visuel de l'observateur, ou

[1] C'est le 22 octobre 1850 qu'il eut l'idée de la loi qui porte son nom et qui marque la naissance de la psychophysique

pouvant causer des réflexions de certaines surfaces sur l'écran.

Le calibrage d'écran

Pour s'assurer que l'appareil de reproduction des couleurs qu'est un moniteur fonctionne dans des conditions optimales, il faut d'abord le calibrer. C'est à dire qu'il faut optimiser son fonctionnement de base et le placer dans des conditions de travail connues. Il s'agit donc de fixer :

- La luminosité maximale (point blanc) de l'écran,
- le gamma,
- la température de couleur (en Kelvins),
- et éventuellement la luminosité minimale (le point noir).

Modélisation du stimulus

Le signal source fournit directement l'image de référence pour le système à évaluer. Sa qualité doit être optimale par rapport aux conditions d'acquisition et d'affichage d'une image. Il est essentiel que l'image de référence n'ait pas de défaut si nous souhaitons obtenir des résultats stables.

Les mesures de seuil et de sensibilité

La notion de seuil est fondamentale en psychophysique puisqu'elle renvoi à l'idée de l'existence d'une limite entre deux états.
En fonction de la tâche effectuée, il existe trois types de seuils :

- **Seuil de détection** : l'observateur répond à une question sur la présence ou non d'un stimulus.
- **Seuil de discrimination** : l'observateur répond à une question portant sur l'existence d'une différence entre deux stimuli.
- **Seuil d'identification** : un processus d'identification consiste à établir une correspondance bijective entre un ensemble de stimuli et un ensemble de réponses qui sont, en général, les étiquettes, les noms des stimuli.

Séance d'évaluation

Une séance ne doit en aucun cas dépasser une demi-heure, car l'observateur commence à présenter des signes de fatigue et/ou d'adaptation, et son jugement ne sera plus fiable. Au début de la première séance, environ un certain nombre de "fausses présentations" doivent être introduites afin de stabiliser l'opinion de l'observateur. Les données résultant de ces présentations ne sont pas considérées pour le résultat final du test. Enfin, Il convient de choisir un ordre aléatoire pour la présentation des images.

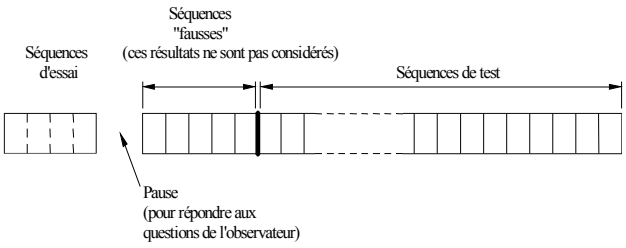



Figure 3 : *Présentation des séquences d'évaluation*.

La gamme des dégradations devra être choisie de telle sorte que toutes les notes soient utilisées par la majorité des observateurs. L'objectif est de converger vers une moyenne générale (moyenne de tous les jugements émis au cours de l'expérience) voisine de 3, pour une échelle de 5 notes.

Les observateurs

Il est recommandé d'avoir un panel d'observateurs le plus large possible, au moins quinze individus. Ils peuvent être experts ou novices, en ce qui concerne le thème de la campagne d'évaluation.

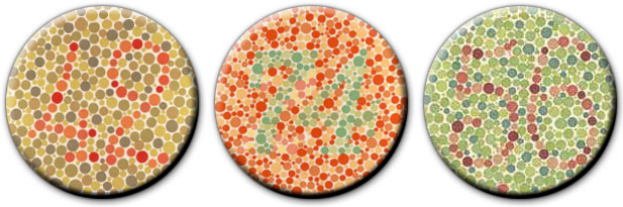


Figure 4: *Test d'Ishihara*.

Avant chaque séance, les observateurs seront sélectionnés pour leur acuité visuelle normale ou rendue normale par correction (tests de *Snellen*) et leur vision normale des couleurs (tests d'*Ishihara*). L'observateur doit avoir ainsi une acuité visuelle de 10/10 pour les deux yeux avec ou sans correction.
Pour la vision normale des couleurs, le test consiste à détecter, grâce à des planches comme celles données par la figure 4, un défaut au niveau de la vision des couleurs (achromatopsie, dont les formes les plus connues sont le daltonisme).
Le panel d'observateurs choisi dépend à la fois des types de traitement à évaluer et de la nature de l'étude à mener. Il faut donc prendre en compte, lors de la sélection de ce panel, plusieurs critères comme l'âge et le sexe mais aussi l'origine ethnique, socioculturelle, professionnelle, etc. Ainsi, la campagne d'évaluation peut être menée sur un panel du genre "femmes européennes de 30 à 40 ans ou hommes asiatiques de plus de 60 ans".

### *3.1.2 Tests psychophysiques*

Par définition, un test d'évaluation psychophysique permet de calculer la sensibilité de l'observateur, par rapport à un environnement normalisé, et à une tâche précise. Ces tests permettent de mettre à contribution la sensibilité de l'être humain, et plus précisément de mesurer cette sensibilité. Il existe différents types de tests psychophysiques proposés dans les normes internationales. Ceux-ci peuvent être réparties génériquement en deux classes:

- **Tests comparatifs** : deux ou plusieurs images sont présentées à l'observateur lequel doit les comparer les unes aux autres,
- **Tests de mesure absolue** : l'observateur doit attribuer une note de qualité à une image présentée seule (sanas référence).

Tests comparatifs

Figure 5 : *Test d'ordonnancement*.

Le test d'ordonnancement permet d'effectuer un classement des images fournies à l'observateur de la meilleure à la plus mauvaise. Ainsi, chaque image sera notée en fonction de son choix et de la configuration retenue (1 à 8 pour la configuration de la figure 5).

Figure 6 : *Test de choix forcé*.

Le test de choix forcé (Figure 6) permet de comparer plusieurs images deux à deux (avec ou sans référence). Une variante de ce test a été baptisée test *Best/Worst*. Ce dernier test (Figure 7) permet de faire sortir à partir d'un groupe d'images, la meilleure et la plus mauvaise d'un point de vue qualitatif.

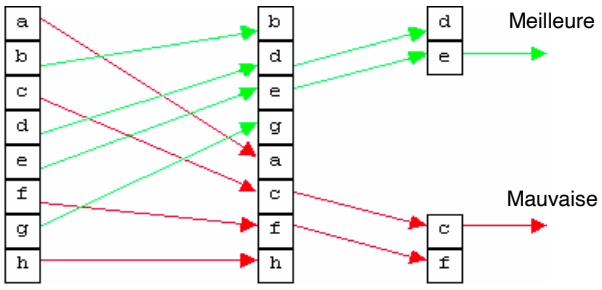



Figure 7 : *Test Best/Worst*.

Tests de mesure absolue

Le but de ces tests n'est pas de déterminer un seuil de sensibilité, mais de noter la qualité d'une image. Il est demandé à l'observateur de quantifier numériquement la qualité de l'image affichée à l'écran. Ce procédé est très largement utilisé pour valider une métrique de qualité, par exemple.
Le Tableau 1 donne un exemple d'une échelle de notation utilisée pour évaluer la qualité d'une image.

| Echelle à cinq notes | | |
|---|---|---|
| Qualité | | Dégradation |
| Excellente | 5 | Imperceptible |
| Bonne | 4 | Perceptible mais non gênant |
| Assez bonne | 3 | Légèrement gênant |
| Médiocre | 2 | Gênant |
| Mauvaise | 1 | Très gênant |

Tableau 1 : *Exemple d'échelle de notation*

### *3.1.3 Analyse des résultats*

A cause de leur variation avec le domaine, il n'est pas approprié d'interpréter les jugements de la plupart des méthodes dans des termes absolus.
Pour chaque paramètre du test, la moyenne et l'intervalle de confiance à 95% de la distribution statistique des notes, doivent être calculés.

MOS : Score d'Opinion Moyen

La première étape de l'analyse des résultats est le calcul de la note moyenne, $\overline{u}_{jkr}$, ou MOS ( Mean Opinion Score) pour chacune des présentations :

$$\overline{u}_{jkr} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr}$$

où $u_{ijkr}$ est la note de l'observateur *i* pour la dégradation *j* de l'image/séquence *k* et la répétition *r*, et *N* est le nombre d'observateurs. D'une manière similaire nous pouvons calculer les notes moyennes globales, $\overline{u}_j$ et $\overline{u}_k$, pour chaque condition de test (dégradation) et chaque séquence/image de test.

Intervalle de confiance

Afin d'évaluer au mieux la fiabilité des résultats, il est préférable d'associer à chaque moyenne (*MOS*) un intervalle de confiance. En général, il est convenu d'utiliser l'intervalle de confiance à 95% donné dans le cas des notes moyennes pour chacune des présentations, par l'équation suivante :

$$\left[\overline{u}_{jkr} - \delta_{jkr},\ \overline{u}_{jkr} + \delta_{jkr}\right]$$

où

$$\delta_{jkr} = 1.95\frac{S_{jkr}}{\sqrt{N}}$$

L'écart-type pour chaque présentation $S_{jkr}$, est donné par :

$$S_{jkr} = \sqrt{\sum_{i=1}^{N}\frac{\left(\overline{u}_{jkr} - u_{ijkr}\right)^2}{(N-1)}}$$

Avec une probabilité de 95%, la valeur absolue de la différence entre le score moyen expérimental et le « vrai » score moyen (pour un très grand nombre d'observateurs) est inférieure à 95% de l'intervalle de confiance, avec la condition que la distribution des scores individuels vérifie certaines conditions.

## 3.2 Séquences vidéos

Pour l'évaluation subjective de la vidéo, l'ITU préconise deux recommandations. La première est la recommandation ITU-R BT.500 et est intitulée « Methodolgy for the subjective assessment of the quality of television pictures ». La deuxième, ITU-T P.910 est intitulée « Subjective video quality assessment for multimédia applications ». Le choix entre les différentes méthodes proposées dans ces deux recommandations dépend de l'application considérée. Ainsi quand la séquence originale (de référence) est disponible, et que les débits en jeu sont supérieurs à 2Mbits/s, la méthode à double stimuli utilisant une échelle de qualité continue DSCQS (Double Stimuli Continuous Quality Scale) est la méthode la plus utilisée. En effet, elle est particulièrement indiquée pour mesurer la qualité des systèmes par rapport à une référence ou bien pour comparer la qualité de plusieurs systèmes entre eux. Le principe de cette méthode cyclique consiste à présenter à l'observateur une paire de séquences vidéos : une séquence de référence et une séquence dégradée. La position de la séquence de référence varie d'une manière pseudo aléatoire. La Figure 8 récapitule la séquence de présentations.

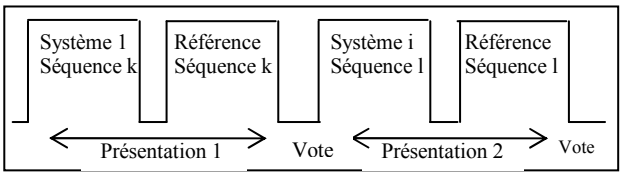



Figure 8 – *Chronogramme de la DSCQS*

A la fin de chaque présentation, les observateurs expriment leurs jugements sur une paire d'échelles verticales continues. Pour permettre de se repérer, ces échelles ont été divisées en cinq intervalles égaux correspondant aux mêmes qualificatifs utilisés dans le cas des images fixes : excellent, bon, assez bon, médiocre et mauvais.

Les méthodes à simple stimulus sont généralement utilisées quand on ne dispose pas de la séquence originale. La plus répandue de ces méthodes est l'ACR (Absolute Category Rating). Dans cette méthode, les séquences sont présentées une par une et sont évaluées indépendamment sur une échelle de catégorie utilisant les mêmes qualificatifs que ci dessus. Le chronogramme associé à cette évaluation par catégories absolues est donné Figure 9.

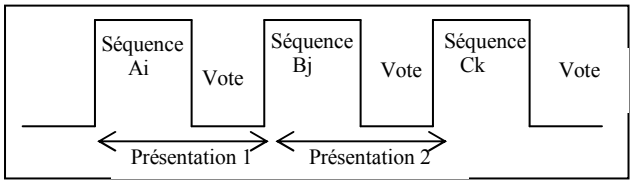



Figure 9 : Chronogramme de l'ACR.

L'analyse des résultats issus des tests subjectifs dépend de la méthode utilisée. Ainsi pour la DSCQS, la cohérence des résultats sera vérifiée en étudiant les notes données par le même observateur à la même séquence pendant la même séance. Si les notes diffèrent de 2 points ou plus (pour une échelle allant de 1 à 5), ces notes seront rejetées. Après chaque séance, il faut calculer les valeurs moyennes E(Xj) et les écarts type associés $\sigma$(Xj) associés à chaque niveau de dégradation ou système de traitement à évaluer (j). La valeur moyenne est donnée par

$$\overline{N}_{jkr} = \frac{1}{N_{obs}} \cdot \sum_{i=1}^{N_{obs}} N_{ijkr}$$

avec $N_{ijkr}$ note de l'observateur i pour la dégradation j de la séquence k et la répétition r. L'écart type utilisé est en général l'intervalle de confiance à 95% et est donné par $\sigma_{jkr} = 1.96 \frac{\left( \frac{1}{N_{obs}-1} \cdot \sum_{i=1}^{N_{obs}} \left( N_{ijkr} - \overline{N}_{jkr} \right)^2 \right)^{1/2}}{\left( N_{obs} \right)^{1/2}}$.

Ces valeurs moyennes reposent sur une distribution dont les deux variables sont les scènes et les observateurs. Il faut vérifier au moyen du test $\beta_2$, si cette distribution est normale ou pas. Pour ce faire il faut calculer le coefficient de kurtosis de la fonction, défini comme le rapport moment du quatrième ordre sur le carré du moment du deuxième ordre, soit

$$\beta_{2jkr} = \frac{\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \left( N_{ijkr} - \overline{N}_{jkr} \right)^4}{\left( \frac{1}{N_{obs}} \cdot \sum_{i=1}^{N_{obs}} \left( N_{ijkr} - \overline{N}_{jkr} \right)^2 \right)^2} .$$

Si $\beta_2$ est compris entre 2 et 4, on peut considérer la distribution comme normale. Les résultats de chaque distribution sont alors à comparer avec la valeur moyenne associée plus l'écart type associé multiplié par 2 (normale) ou par $(20)^{1/2}$ et à la valeur moyenne associée moins ce même écart type multiplié par 2 (normale) ou par $(20)^{1/2}$. Chaque fois que les résultats donnés par un observateur se situent hors de cet intervalle, il faut les enregistrer sur un compteur associé à chaque observateur. Il faut donc deux compteurs séparés, un pour les valeurs supérieures (Pi) et l'autre pour les valeurs inférieures Qi. On calcule ensuite les deux rapports suivants :

$$\frac{P_i + Q_i}{N_{\deg} N_{imr} N_{rep}} > 0.05 \quad et \quad \left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0.3 \cdot$$

Si le premier est supérieur à 5% et le second inférieur à 3%, il faut alors éliminer l'observateur i. La recommandation ITU-R BT.500 récapitule bien la procédure ci dessus et peut être exprimée comme suit :

Si $N_{ijkr} \geq \overline{N}_{jkr} + 2\sigma_{jkr}$ (distribution normale)

ou $N_{ijkr} \geq \overline{N}_{jkr} + \sqrt{20}\sigma_{jkr}$

alors Pi = Pi + 1 ;

si $N_{ijkr} \leq \overline{N}_{jkr} - 2\sigma_{jkr}$ (distribution normale)

ou $N_{ijkr} \leq \overline{N}_{jkr} - \sqrt{20}\sigma_{jkr}$

alors Qi = Qi + 1 ;

si $\frac{P_i + Q_i}{N_{\deg} N_{imr} N_{rep}} > 0.05$ *et* $\left|\frac{P_i - Q_i}{P_i + Q_i}\right| < 0.3$

alors éliminer l'observateur i.

Enfin, deux autres outils peuvent être utilisés en complément. Le premier, le Z-score, permet de minimiser les variations entre les notes individuelles dûes à la non utilisation des échelles entières par les observateurs. Le second, le test de student, précise, d'un point de vue statistique, si les valeurs moyennes associées à un niveau de dégradation j sont discernables ou non entre elles Pour finir signalons que les résultats doivent être donnés avec les informations suivantes :

- Détails de la configuration de l'expérience
- Détails du matériel d'évaluation
- Type de source vidéo et de moniteur d'évaluation
- Nombre et type d'observateurs
- Moyenne générale de l'expérience
- Résultats moyens originaux et corrigés et écarts types si un ou plusieurs observateurs ont été éliminés selon la procédure ci dessus.

# 4 Métriques Objectives

## 4.1 Images fixes avec référence complète

### *4.1.1 Métriques simples*

Ce type de mesures est très largement utilisé pour quantifier numériquement la qualité de reconstruction d'une image.

Parmi le panel de mesures exploitées, l'une des mesures couramment utilisée est le pic du signal sur bruit–*peak signal to noise ratio*–connu sous l'étiquette PSNR. Cette mesure permet de quantifier la distorsion qui existe entre deux images en utilisant la formule suivante :

$$\mathrm{PSNR} = 10 \log_{10} \sum_{(x,y)} \left( \frac{(g(x,y) - f(x,y))^2}{D} \right)$$

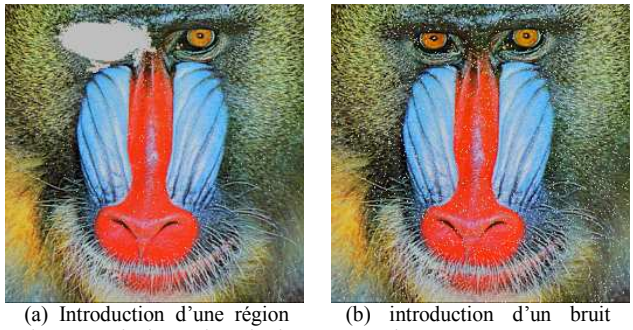
où $D$ représente le domaine de définition de l'image, $f(x,y)$ représente le pixel de coordonnées $(x,y)$ de l'image de référence et $g(x,y)$ le pixel de coordonnées $(x,y)$ de l'image ayant subie un traitement.

Dans le cas des images couleur, aucune définition précise de l'adaptation de l'équation du PSNR n'existe. Plusieurs variantes existent :

5 Le PSNR est calculé sur chacun des trois plans colorimétriques puis une moyenne est ensuite effectuée afin d'obtenir une valeur finale,

6 La distance euclidienne entre deux pixels de couleur est utilisée dans l'équation originale.

Dans tous les cas, il n'est pas très aisé de pouvoir interpréter les résultats obtenus. La Figure 10 illustre les difficultés d'interprétation de la mesure de la qualité en se basant uniquement sur le PSNR.



(a) Introduction d'une région de 1700 pixels et d'un bruit gaussien portant sur 300 pixels

(b) introduction d'un bruit gaussien portant sur 2 000 pixels.

Figure 10 – *Illustration du cas limite PSNR/perception pour une même valeur de PSNR=11.06 dB*

En ce qui concerne les dégradations apportées aux images (a) et (b) de la Figure 8, nous pouvons constater que malgré une même valeur de PSNR, la perception visuelle qu'un être humain peut avoir de ces deux images est très différente Force est de constater que cette mesure de distorsion, même si elle est utilisée très fréquemment en traitement d'image pour quantifier la qualité, ne permet pas de prendre en compte la sensibilité du système visuel humain, et n'apparaît pas être performante.

### *4.1.2 Caractéristiques du SVH*

Comme mentionné dans la section 3, les conditions de visualisation ont une influence significative sur l'apparence d'une image ou d'une vidéo. En effet, elles peuvent amplifier ou au contraire diminuer la visibilité des artéfacts.

Il existe de nombreux phénomènes visuels influençant l'apparence d'une image sur un écran. Da façon plus générale, il peuvent être classés en deux catégories : 1) les phénomènes de vision bas niveau et 2) les phénomènes de vision haut niveau. Dans ce cours, nous nous limiterons à l'étude de quelques phénomènes de vision bas-niveau, à savoir, la sensibilité au contraste, la sélectivité spatio-fréquentielle, le contraste local et le masquage intra et inter-composantes.

La sensibilité au contraste représente l'aptitude du système visuel humain à détecter les changements de luminance (achromatique) ainsi que les changements chromatiques. Toute mesure de sensibilité au contraste dépend du niveau de luminance des stimuli, de leurs fréquences spatiales, de leur chrominance et du niveau d'adaptation de l'observateur humain.

Le contraste permet ainsi de mesurer la variation relative de la luminance par rapport au voisinage. Cette propriété est connue comme étant la loi de Weber-Fechner :

$$C^W = \frac{\Delta L}{L}$$

où $\Delta L$ représente la différence de luminance entre le stimulus et son voisinage, et $L$ la luminance du voisinage. Cette loi n'est bien évidemment qu'une approximation de

la perception sensorielle réelle, mais les mesures de contrastes basées sur ce concept sont largement utilisées en vision. L'allure générale de la courbe ainsi générée dépend des nombreuses caractéristiques des stimuli utilisés ; plus précisément de la couleur ainsi que des fréquences spatiales et temporales. Les fonctions de sensibilité au contraste—Constrast Sensitivity Functions, CFSs—sont généralement utilisées pour quantifier ces dépendances.

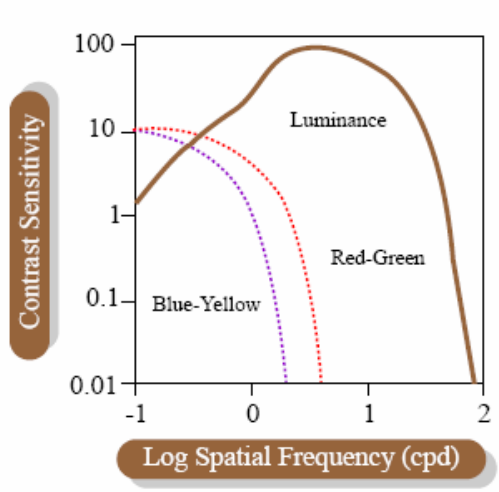



Figure 11 – *Courbes typiques pour les CSFs achromatique et chromatique.*

Tandis qu'il est relativement plus aisé de construire une CSF dans le cas de la perception de la luminance, il est plus difficile de trouver une mesure pour la CSF chromatique. Ceci est en partie dû à l'absence de définition simple du contraste couleur. Cependant cette donnée est primordiale pour aboutir à un schéma de compression ou à une métrique de qualité efficace. Van der Horst et Bouman [1] ont été parmi les premiers à mesurer la CSF chromatique. Ils ont ainsi montré que la CSF chromatique possède des caractéristiques passe-bas.
Les approximations de la CSF chromatique et achromatique sont fournies en Figure 11. Pour les hautes fréquences, la sensibilité au contraste achromatique est plus importante que dans le cas chromatique, alors que dans le cas des basses fréquences, le phénomène inverse est observé.
Cette modélisation de la CSF qui est souvent appelée modélisation mono-canal ne permet pas d'expliquer complètement le comportement du SVH au regard de stimuli complexes. En effet ce dernier utilise plusieurs canaux dont les sensibilités en orientation et en fréquence spatiale sont différentes. Concernant la sélectivité radiale, plusieurs valeurs de largeur de bande existent. Ainsi selon les études menées, pour le canal achromatique, la largeur de bande en fréquence spatiale couvre 1 à 2 octaves et varie de 20 à 60 degrés pour l'orientation.. Concernant les canaux chromatiques, la largeur de bande varie de 60 à 130 degrés. La sensibilité chromatique reste importante pour des très basses fréquences, ce qui nécessite l'emploi de filtre passe-bas. Afin de faciliter leurs implémentations, les mêmes filtres de décomposition sont utilisés pour les canaux chromatique et achromatique. La modélisation en canaux perceptuels est donnée Figure 12.

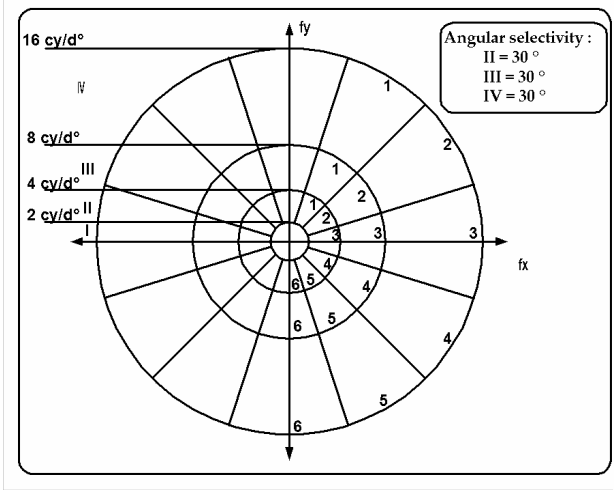



Figure 12 – *Découpage du plan fréquentiel de la composante achromatique selon Daly [2].*

Cette sélectivité radiale et fréquentielle opérée par le SVH, laisse supposer que les signaux interagissent entre eux. Un signal peut dans ce cas modifier la visibilité d'un autre signal. On parle dans ce cas d'effet de masquage. Les tests psychophysiques utilisés permettent de mesurer les effets de masquages entre signaux de différentes orientations, fréquences spatiales, voire entre signaux chromatique et achromatique. Il a ainsi été constaté que le masquage dépendait non seulement de l'énergie dans un canal, mais aussi de l'énergie des canaux adjacents en terme d'orientation. On utilise alors le terme de masquage intra-canal et de masquage inter-canal.
La Figure 13 illustre les effets de masquage pouvant intervenir sur une image. La même quantité de bruit uniforme a été ajoutée dans une zone rectangulaire de l'image qui si situe en haut pour l'image de gauche et en bas pour l'image de droite. Le bruit est clairement visible dans l'image de gauche alors que sa visibilité est plus difficile sur l'image de droite.

Figure 13 – *Illustration de l'effet de masquage.*

Ces diverses caractéristiques seront utilisés dans l'optique de construire des modèles de vision afin d'améliorer les applications en traitement d'images, et plus particulièrement en mesure de qualité.

*4.1.3 Schéma blocs générique d'une métrique perceptuelle et description des blocs*

Les métriques perceptuelles représentent une approche intéressante dans l'évaluation de la qualité des images. Une synthèse des différentes études menées dans ce contexte montrent que ces métriques s'inspirent du fonctionnement du SVH et exploitent les facteurs perceptuels qui ont été décrits dans le paragraphe précédent et qui sont connus pour avoir une influence directe sur la visibilité des distorsions. Un schéma bloc générique de ces métriques est donné par la Figure 14.

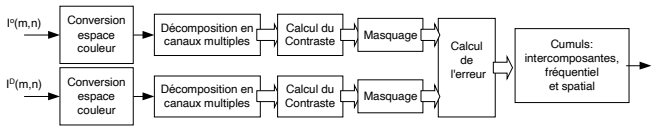



Figure 14: Schéma bloc des métriques perceptuelles

L'image originale $I^0(m,n)$ et l'image dégradée $I^D(m,n)$ subissent une première transformation permettant de passer de l'espace colorimétrique initial (en général RVB) vers un espace antagoniste luminance (A) et chrominance (C1, C2). Ces dernières composantes reflètent les oppositions rouge-verte et jaune-bleue. Le bloc « décomposition en canaux multiples » a pour objectif de rendre compte de la sensibilité spatio-fréquentielle du SVH. Ce dernier analyse les signaux d'entrée moyennant un ensemble de canaux sélectifs en radial et en orientation. Deux questions se posent dans l'implantation de cette sensibilité. La première concerne le choix de la décomposition. Pour la composante luminance, les décompositions les plus utilisées sont celles de Daly [2], Lubin [3] et [4]. Celles de Watson et Daly se caractérisent toutes les deux par une sélectivité radiale dyadique (cinq canaux de largeurs de bande égale à l'octave) et une sélectivité angulaire constante. La différence se situe au niveau de la valeur de la largeur de bande qui est fixée à 30 degrés pour Daly contre 45 degrés pour Watson. La décomposition de Lubin nécessite sept bandes radiales et quatre orientations différentes. Pour les composantes de chrominance, la même décomposition est appliquée avec toutefois une limitation dans la fréquence maximale des composantes C1 et C2. En effet les études de sensibilité au contraste couleur donnent une fréquence de coupure de l'ordre de 3 à 4cpd (cycle/degrés) pour C1 et de 2 à 3cpd pour C2.

La deuxième question concerne le choix de la transformée linéaire à mettre en œuvre. Cette transformée doit satisfaire un certain nombre de propriétés parmi lesquelles il y'a la sélectivité radiale et angulaire bien sur, la phase linéaire, un recouvrement minimum entre canaux adjacents, l'inversibilité, etc. A notre connaissance, seule la transformée Cortex satisfait toutes ces conditions. C'est ce qui explique d'ailleurs pourquoi elle est souvent utilisée.

La sortie de ce bloc résulte donc dans un ensemble d'images de luminance $L_{i,j}$ et de chrominance $C1_{i,j}$ et $C2_{i,j}$ où l'indice i représente le canal radial et l'indice j le canal angulaire. A ce niveau, l'accent est en général mis sur la composante de luminance parce qu'on estime que le gain de performance obtenu est loin de compenser la complexité induite par le traitement des composantes de chrominance.

Pour chacune des images de luminance, un contraste local $c_{i,j}(m,n)$ est calculé en chaque point (m,n). Le bloc « masquage » vise à exploiter les capacités de masquage du SVH. Il est connu pour être, avec le bloc cumul qu'on discutera après, le bloc qui détermine d'une manière prépondérante les performances des métriques perceptuelles. Son rôle est de préciser pour chaque sous bande et pour chaque point la variation du seuil de visibilité quand l'effet de masquage est pris en compte. La connaissance de telles valeurs, rappelons le, permet de ne conserver que les erreurs situées au dessus de leur seuil et contribuant donc à l'élaboration de la qualité finale. Ce bloc a suscité beaucoup d'études et différents résultats sont disponibles dans la littérature. La diversité des modèles proposés dépend du type de masquage considéré et donc des stimuli utilisés. En effet l'augmentation des seuils de visibilité induite par masquage dépend de la nature du stimulus et du signal masquant, de leur phase, de leur orientation et du degré de familiarité de l'observateur.

A la sortie du bloc masquage, seules les erreurs visibles sont donc conservées car contribuant directement à l'élaboration de la qualité (les autres erreurs sont mises à zéro). Le bloc cumul a pour objectif de réduire cette dimensionnalité. Généralement, le cumul s'effectue en trois étapes. Une première où les images d'erreurs de luminance sont combinées linéairement avec les erreurs de chrominance (c'est le cumul inter composantes). La deuxième étape consiste à regrouper les images d'erreurs réparties dans tous les canaux fréquentiels en une seule image d'erreurs (c'est le cumul fréquentiel). La troisième étape enfin est dédiée au cumul spatial et consiste à combiner les erreurs spatiales en une mesure finale représentant la note attribuée par l'algorithme à l'image dégradée. Ces deux cumuls, effectués en général dans cet ordre, utilisent souvent la sommation de Minkowski

$$M = \left( \sum_i |s_i|^\beta \right)^{\frac{1}{\beta}}$$

où l'indice i représente le $i^{ème}$ canal fréquentiel pour un cumul spectral et le $i^{ème}$ pixel pour un cumul spatial.

## 4.2 Images fixes sans référence

Dans cette section, nous présentons une métrique simple de mesure de la qualité (effets de bloc dans ce cas) sans image de référence. Ces approches sont dites *blind* (aveugle ou sans visibilité) car la mesure n'utilise que les informations issues de l'image dégradée.

La méthode décrite a été présentée par *wang* et *al.[5]* et est représentée par le schéma de la Figure 15.

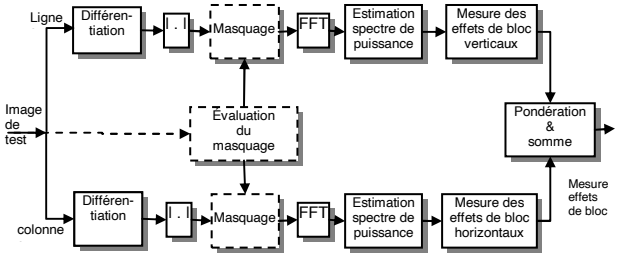



Figure 15 : s*chéma bloc de la métrique d'évaluation des effets de bloc.*

Cette mesure repose sur les étapes suivantes :

- Modélisation du signal contenant des effets de bloc (image originale + signal idéal de bloc)
- Détection et estimation de la puissance du signal de bloc dans l'image contenant des effets de bloc.
  - La puissance spectrale est égale à : $P = \{P[l],\ 0 \le l \le N/2\}$
  - $P_m$ : Filtrage médian de la puissance spectrale $P$
  - Lissage de la puissance spectrale $P_m$

$$P_s[l] = \begin{cases} P[l]\ l = \frac{N}{8}, \frac{2N}{8}, \frac{3N}{8}\ ou\ \frac{4N}{8} \\ P_m[l] \qquad autrement \end{cases}$$

- Mesure des effets de bloc verticaux

$$M_{Bv} = \sum_{i=0}^{4} P\left[\tfrac{iN}{8}\right] - P_M\left[\tfrac{iN}{8}\right]$$

- Mesure globale :

$$M_B = \tfrac{1}{2}\left(M_{Bv} + M_{Bh}\right)$$

- Prise en compte de deux types de masquage : masquage de texture et masquage de luminance.

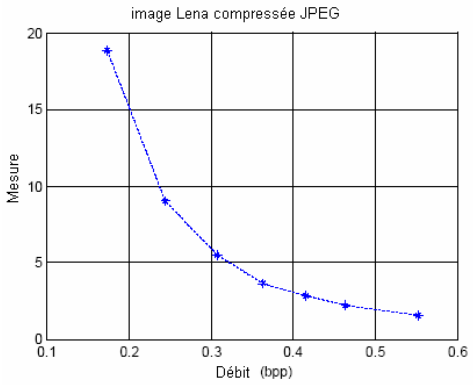



Figure 16 : *Quelques résultats obtenus avec la mesure pour différents taux de compression JPEG.*

## 4.3 Séquences vidéos

### *4.3.1 Schéma générique d'une métrique avec référence complète*

Pour une séquence vidéo, le schéma bloc d'une métrique avec référence complète ressemble à celui des images fixes avec toutefois deux différences fondamentales au niveau des blocs de décomposition et de cumul. Pour les images animées, le bloc de décomposition doit également inclure le comportement des mécanismes temporels du SVH. Deux comportements du SVH sont aujourd'hui largement admis : un comportement statique pour l'analyse des signaux statiques ou de vitesse faible et un comportement dynamique dédié aux signaux ayant un vitesse approximativement supérieure à 1cy/seconde. A ces deux régimes, continu et transitoire, on associe généralement un filtre passe bas et un filtre passe-bande respectivement. A titre d'exemple la réponse impulsionnelle h(t) donnée par

$$h(t) = \exp - \left( \frac{\ln(t/\tau)}{\sigma} \right)^2$$

et sa dérivée seconde ont été utilisées pour modéliser les deux mécanismes. La Figure 17 donne le tracé de la réponse fréquentielle de chacun de ces deux filtres.

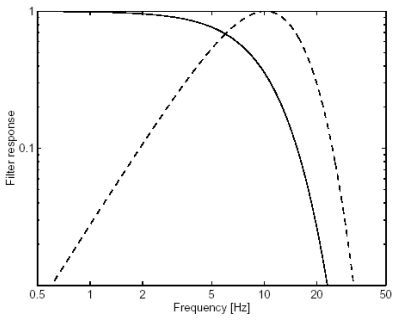



Figure 17 : *Réponses fréquentielles des mécanismes visuels temporels statique (continu) et transitoire (discontinu).*

Le filtrage temporel passe-bas est appliqué aux trois composantes alors que le filtrage passe-bande est en général appliqué qu'à la composante de luminance pour réduire la complexité de la métrique.

Le bloc cumul dans le cas des séquences vidéo doit évidemment inclure le cumul temporel. Après le cumul spatial qui génère une note de qualité pour chaque trame constituant la séquence, le cumul temporel, effectué moyennant une sommation de Minkowski, permet de déterminer la note de qualité globale de la séquence.

Les interactions inter composantes et le masquage temporel constituent encore des domaines de recherche très actifs. Il est par conséquent très difficile de dégager un consensus pour le présenter ici.

### *4.3.2 Schéma générique d'une métrique simple sans référence, carte d'importance*

Plusieurs applications nécessitent d'évaluer la qualité vidéo sans disposer de la vidéo d'origine. Ceci explique donc l'engouement suscité par le développement de métriques de qualité sans référence. Métriques qui restent d'ailleurs très proches de la manière dont un observateur jugerait la qualité d'une vidéo.

La démarche pour la mise en oeuvre de ces métriques est totalement différente de celle des métriques de fidélité puisqu'elle n'exploite plus la comparaison entre la

séquence originale et la séquence dégradée mais se base sur une analyse des dégradations que présente la séquence à évaluer. Toutes les métriques de qualité sans référence supposent donc une connaissance a priori des dégradations mises en jeu.

Dans ce paragraphe, nous nous intéresserons essentiellement aux dégradations induites par les différents standards de codage vidéo (MPEG1, MPEG2 et MPEG 4) dont une liste non exhaustive est donnée.

- Blockiness : distorsion de l'image caractérisée par l'apparition d'une structure de bloc.
- Ringing : distorsion autour des contours.
- Blurring : effet de flou induit par la perte des hautes fréquences spatiales.
- Color bleeding : Mélange de couleurs entre les zones à forte différence de chrominance.
- Temporal edge noise : variations dans le temps de la forme des contours.
- Flickering : bruit de fond.
- Motion jerkiness: perception d'une série d'images discontinues au lieu d'un mouvement continu et fluide.

Une métrique utilisant tout ou une partie de ces dégradations peut être illustrer par la Figure 18.

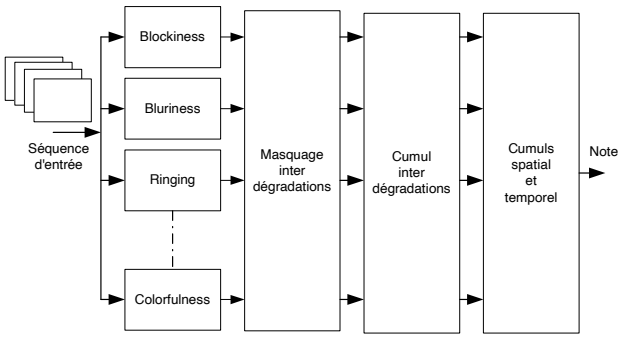



Figure 18: *Schéma bloc d'une métrique simple sans référence.*

Cette structure présente l'avantage d'être modulaire et permet donc en fonction des applications et des codeurs mis en oeuvre de considérer telle dégradation par rapport à telle autre. Un deuxième avantage réside dans la possibilité d'utiliser des algorithmes indépendants pour la mesure des différentes dégradations. Le masquage et cumul inter dégradations permettent ensuite de combiner ces notes individuelles en chaque point de chaque trame. Le cumul spatial génèrera une note par trame alors que le cumul temporel déterminera la note finale attribuée à la séquence d'entrée.

Un exemple de mesure de l'effet de bloc est présenté ci dessous pour donner une idée un peu plus précise sur les métriques de qualité sans référence.

**Mesure de l'effet de bloc sans référence**

Plusieurs méthodes de mesure de l'effet de bloc existent dans la littérature. Un classement rapide permet de distinguer celles qui opèrent dans le domaine spatial, dans le domaine fréquentiel et dans le domaine DCT. Nous nous intéresserons qu'a ce dernier domaine parce que d'une part le temps de calcul est réduit et que d'autre part la majorité des standards de compression vidéo utilise la transformée en cosinus discrète. La métrique décrite ici, est présentée par A.C.Bovik et S.Liu [6]. Cette métrique utilise le fait que l'on peut modéliser l'effet de bloc comme un signal 2-D en créant un nouveau bloc à partir de deux blocs adjacents. Avant de décrire la métrique elle même, on expose d'abord le modèle d'effet de bloc en présentant uniquement l'effet de bloc vertical, sachant que les calculs sont facilement transposables pour l'effet de bloc horizontal. On considère deux blocs adjacents $x_1$ et $x_2$ de taille 8×8, de valeur moyenne respective $\bar{a}$ et $\bar{b}$ et telle que $\bar{a} \neq \bar{b}$. Ainsi, ces deux blocs peuvent être modélisés comme suit :

$$x_1 = \bar{a} + \varepsilon_{i,j} \quad et \quad x_2 = \bar{b} + \delta_{i,j}$$

où $\varepsilon_{i,j}$ et $\delta_{i,j}$ sont des bruits blancs.

On forme le nouveau bloc $x$ (qui va chevaucher les deux blocs considérés) en prenant la moitié inférieure du bloc $x_1$ et la moitié supérieure du bloc $x_2$, comme on le voit sur la figure suivante :

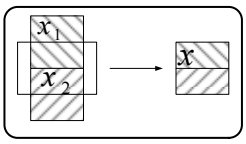



Figure 19 : Méthode de construction du nouveau bloc

On définit maintenant une fonction 2-D sur ce bloc $x$ :

$$u = \begin{cases} -\frac{1}{2} ; \textit{ pour la moitié sup érieure du bloc } x \\ \frac{1}{2} ; \textit{pour la moitié inf érieure du bloc } x \end{cases}$$

Si on note $S = \bar{b} - \bar{a}$, alors le nouveau bloc peut être modélisé par

$$x = S \bullet u + \zeta_{i,j} + B.$$

avec : - $|S|$ représente l'amplitude de la fonction 2-D $u$,

- $B$ est la valeur moyenne du bloc $x$, qui peut être considérée comme la luminosité locale de l'arrière plan,
- $\zeta_{i,j}$ est un bruit blanc considéré comme l'activité locale autour du contour du bloc.

Plus la valeur de $|S|$ est grande, plus l'effet de bloc est important. On peut donc mesurer l'effet de bloc entre les deux blocs $x_1$ et $x_2$ en estimant la valeur de $|S|$. Dans le but d'avoir une mesure plus précise, on suppose que $\bar{a}$ est la valeur moyenne des quatre lignes inférieures du bloc $x_1$ et $\bar{b}$ la valeur moyenne des quatre lignes supérieures de $x_2$.

La Figure 20 donne le schéma bloc, de la mesure de l'effet de bloc, tel que donné par les auteurs.

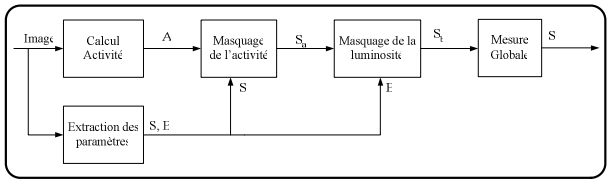



Figure 20: *Mesure de l'effet de bloc dans le domaine DCT.*

**Bloc calcul activité :**
On suppose d'une part que l'activité totale d'un bloc est la somme des activités à toutes les fréquences spatiales et que d'autre part le coefficient DCT représente l'amplitude de la composante attendue.
L'effet de masquage dépend de l'orientation relative. Puisque l'effet de bloc a deux directions particulières dans l'image (verticale et horizontale), une distorsion sur le contour sera d'autant plus masquée si l'activité dans l'arrière plan à la même orientation que l'effet de bloc.
Tenant compte de cette propriété, on définit deux activités orientées distinctes, une activité verticale $A_V$ et une activité horizontale $A_H$ :

$$A_V = \sum_{v=1}^{7} v \sum_{u=1}^{7} |X(u,v)|$$
$$A_H = \sum_{u=1}^{7} u \sum_{v=1}^{7} |X(u,v)|$$

Pour l'effet de bloc vertical, l'activité verticale sera dominante, ce qui implique :

$$A_{total}^{V} = A_V + \alpha A_H \ avec\ \alpha = 0{,}8$$

De même, pour l'effet de bloc horizontal, on obtient :

$$A_{total}^{H} = A_H + \alpha A_V \ avec\ \alpha = 0{,}8 .$$

**Bloc extraction des paramètres :**
Pour expliquer la méthode d'extraction de paramètres, on va considérer uniquement l'effet de bloc vertical, sachant que les calculs sont facilement transposables pour l'extraction des paramètres concernant l'effet de bloc horizontal.
On définit 2 matrices $q_1$ et $q_2$ :

$$q_1 = \begin{bmatrix} 0 & I_{4\times4} \\ 0_{4\times4} & 0 \end{bmatrix} et\ q_2 = \begin{bmatrix} 0 & 0_{4\times4} \\ I_{4\times4} & 0 \end{bmatrix}$$

avec $I_{4\times4}$, la matrice identité et $0_{4\times4}$, la matrice nulle.
On note $X_1$, $X_2$, $X$, $Q_1$, $Q_2$ la DCT respective de $x_1$, $x_2$, $x$, $q_1$, $q_2$. En considérant la linéarité et distributivité de la DCT, on parvient à l'équation suivante :

$$X = Q_1 X_1 + Q_2 X_2 .$$

Si

$$F_+ = Q_1 + Q_2 \ ; F_- = Q_1 - Q_2$$
$$Y_+ = X_1 + X_2 \ ; Y_- = X_1 - X_2$$

alors

$$X = \frac{1}{2}\left(F_+ Y_+ + F_- Y_-\right)$$

Le calcul de cette équation est simple dans la mesure où la moitié des multiplications sont évitées étant donné la dispersion des coefficients DCT. A partir des coefficients DCT de $X$, on peut ainsi obtenir les paramètres caractéristiques $B$ et $S$ (voir modèle de l'effet de bloc) :

$$B = \frac{X(0,0)}{8}$$
$$S = \frac{\left[-Q_2(0,0)Y_-(0,0) + Q_2(1,0)Y_+(1,0) + Q_2(3,0)Y_+(3,0) + Q_2(5,0)Y_+(5,0) + Q_2(7,0)Y_+(7,0)\right]}{4}$$

Sachant que :

$Q_1(0,0) = Q_2(0,0)$ *et* $Q_1(k,0) = Q_2(k,0)$ *pour* $k = 1,3,5,7$ .
$Q_1(k,0) = Q_2(k,0) = 0$ *pour* $k = 2,4,6$

L'étape d'extraction de paramètres se déroule donc comme suit :

1. Etant donnée deux blocs adjacents $X_1$ et $X_2$, on calcule leur somme $Y_+$ et leur différence $Y_-$.
2. On calcule ensuite le nouveau bloc.
3. On déduit enfin la valeur de $B$ et de $S$.

Pour calculer l'effet de bloc horizontal, on peut transposer l'image entière (faire une rotation, par exemple) et utiliser la même méthode que précédemment.

**Bloc masquage de l'activité :**
Le masquage de l'effet de bloc dû à l'activité devrait varier comme une fonction décroissante de l'activité locale. On note $S_a$ la visibilité de l'effet de bloc après masquage de l'activité :

$$S_a = \frac{|S|}{1 + A_{total}}$$

$|S|$ est l'amplitude de la fonction du bloc $x$ (une des sorties du bloc extraction des paramètres), $A_{total}$ prend comme valeur soit $A_{total}^{V}$ soit $A_{total}^{H}$, le choix dépend de la direction de l'effet de bloc considéré (vertical ou horizontal).

**Bloc Masquage de la luminosité**
La visibilité de l'effet de bloc dépend aussi de la luminosité de l'arrière plan local. On introduit donc la carte de visibilité $S_b$ pour une frontière inter-bloc :

$$S_b = \frac{|S_a|}{1 + \left(\frac{B}{150}\right)^2}$$

$|S_a|$ est la visibilité de l'effet de bloc après masquage de l'activité (sortie du bloc masquage de l'activité) et B est la valeur moyenne du bloc $x$ (une des sorties du bloc extraction des paramètres).
$|S_a|$

**Bloc Mesure Globale :**
Pour obtenir une mesure globale de l'effet de bloc dans l'image, on doit combiner toutes les cartes de visibilités $S_b$, pour toutes les frontières inter-bloc dans l'image et ainsi obtenir une valeur numérique pour prédire la qualité de l'image entière comme suit :

$$S_t = \sqrt[p]{\frac{1}{N}\sum_{k=1}^{N}(S_b)_k^p}$$

$N$ est le nombre total de frontière inter-bloc dans l'image.
$S_t$ est la mesure globale de l'effet de bloc dans l'image.
$p$ est une constante choisie égale à 4.

# 5 Standards de compression d'image et de vidéo

## 5.1 Introduction

La compression est un outil qui permet de réduire le coût de stockage et le temps de transmission des données. Concernant les images fixes et la vidéo, il possible de procéder selon deux méthodes : 1) une méthode à reconstruction parfaite dite également compression sans perte, et 2) une méthode durant laquelle des pertes d'informations sont autorisées, communément connue sous l'expression de compression avec perte. L'objectif, dans ce cas, est d'obtenir un signal reconstruit en fonction de contraintes émises par l'utilisateur (volume des données à transmettre, qualité du signal à la reconstruction, etc.). La Figure 21 présente le schéma de compression type utilisé lors d'une compression avec pertes. Les principales étapes sont :

**une transformation**, permettant de représenter les données initiales dans un espace plus adapté à la compression en fonction de critères

**une quantification**, qui permet de réduire de manière significative les coefficients obtenus lors de l'étape précédente en définissant des représentants en nombre restreint,

**un codage**, exploitant la redondance d'information afin d'aboutir à une suite de bits la plus compacte possible.

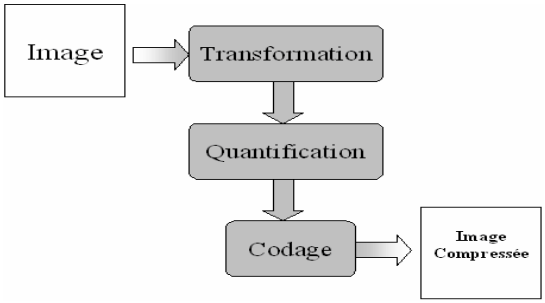



Figure 21 – *Schéma de compression type dans le cas d'une compression avec perte.*

## 6.1 JPEG

Le standard de compression JPEG répond également au schéma de compression type présenté dans la sous-section précédente. On retrouve ainsi une étape de transformation de l'image, suivie d'une étape de quantification. Dans le cas d'une image couleur, les trois plans colorimétriques sont supposés indépendants; chacun étant traités indépendamment les uns des autres selon le schéma appliqué pour les images à niveau de gris. Afin d'obtenir de meilleurs résultats, une transformation des pixels RVB peut être opérée au préalable de façon a obtenir un plan de luminance et deux plans chromatiques.

Selon le document de normalisation, les images sont décomposées en blocs de taille 8×8, sur lesquels est opéré l'étape de transformation des coefficients, suivie de l'étape de quantification des coefficients transformés.

### *4.3.3 Transformation*

L'étape de transformation repose sur la Transformée en Cosinus Discrète (TCD) ou *Discrete Cosine Transform (DCT).* Cette transformation est similaire à celle de Fourier, excepté que la base de décomposition utilise uniquement des cosinus. On a ainsi *f(u,v,k)* le coefficient transformé du bloc *k* correspondant au pixel *x(i,j)* de l'image originale :

$$f(u,v,k) = \frac{1}{4}.C(u).C(v).$$
$$\sum_{i=0}^{7}\sum_{j=0}^{7}\cos\left[\frac{\pi}{N}\,u\left(i+\frac{1}{2}\right)\right]\cos\left[\frac{\pi}{N}\,v\left(j+\frac{1}{2}\right)\right].x(i,j)$$

où i et j représentent les coordonnées du pixel de l'image originale, u et v les coordonnées du pixel transformé, k le bloc considéré, et

$(u,v) \in [0,7] \times [0,7]$
$C(O) = 1/\sqrt{2}$ et $\forall \alpha \neq 0\ C(\alpha) = 1$

Dans le cas d'un bloc de taille 8×8, on obtient 64 vecteurs de bases (*cf.* Figure 22). Les basses fréquences sont localisées dans le coin supérieur gauche. Plus on se déplace vers le coin inférieur droit, et plus les fréquences augmentent.

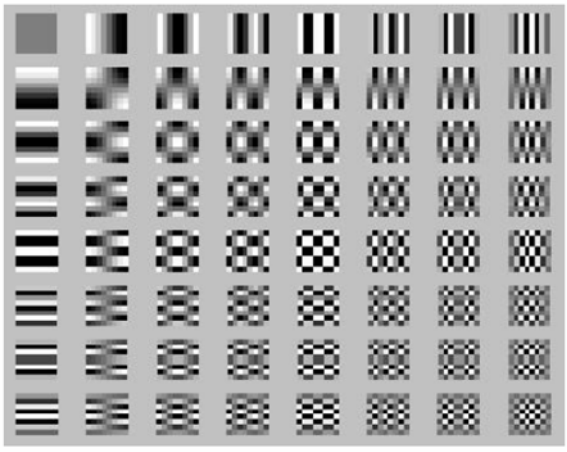


Figure 22 – *Les 64 vecteurs de la base TCD.*

### *4.3.4 Quantification*

Une fois les 64 coefficients *f(u,v)* obtenus, il convient d'effectuer une quantification scalaire. L'objectif étant alors de ne retenir que les coefficients présentant un intérêt certain en fonction de critères prédéfinis. On parle

alors de matrice de quantification. Elle contient les 64 pas de quantification utilisés en fonction de la sensibilité du SVH aux fréquences, puisqu'il est couramment admis qu'il agit comme un filtre passe-bas. Bien évidemment, ces 64 coefficients sont liés au compromis taux de compression / qualité finale de l'image reconstruite. De plus, la définition de ces coefficients de la matrice de quantification dépend de plusieurs autres paramètres tels que le type d'image, le contenu de l'image, la nature des primaires colorimétriques. D'une manière plus pragmatique, l'utilisateur est souvent amené à définir ces propres matrices en fonctions de procédures expérimentales, et principalement en fonction de l'objectif recherché.
Les coefficients TCD quantifiés sont donnés par la relation suivante :

$$\hat{f}(u,v,k) = \mathrm{round}(f(u,v,k)/q(u,v))$$

où *q(u,v)* représente le coefficient de la matrice de quantification à l'indice *(u,v)*. L'erreur de quantification est alors donnée par

$$e(u,v,k) = f(u,v,k) - \hat{f}(u,v,k).q(u,v)$$

## 4.4 JPEG2000

L'appel à proposition pour le nouveau format de compression du troisième millénaire a été lancé en mars 1997. La partie 1, dite le noyau du de JPEG2000, a été publiée comme standard international à la réunion de la Nouvelle-Orléans en décembre 2000. Ce nouveau standard a pour objectif d'offrir de nouvelles fonctionnalités permettant de répondre à une demande croissante, à savoir :

- Obtenir des performances de compression supérieures à son prédécesseur JPEG, notamment pour des débits très faibles.
- Permettre d'organiser le fichier compressé de plusieurs manières, notamment en fonction de la résolution désirée ou de la qualité de reconstruction.
- Avoir un mode de compression sans perte performant
- Fournir la possibilité de coder des régions d'une image avec une meilleure qualité que les autres.

Plusieurs articles de qualité ont été écrits pour décrire le nouveau standard JPEG2000. Le lecteur pourra se référer à celui de *Rabbani* et *Joshi* [7] ou au livre de *Taubman* et *Marcellin* [8].

### *5.1.1 Chaînes de codage et décodage JPEG2000*

Le codage et décodage d'une image au format JPEG2000 s'effectuent en cinq étapes (*cf.* Figure 23): les trois étapes classiques en compression d'image (Transformation, Quantification et Codage), une étape de transformée couleur permettant d'améliorer l'efficacité du codage et une dernière étape d'allocation de débit.

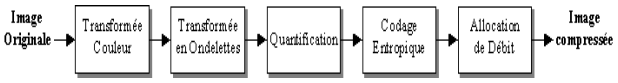



Figure 23 - *Diagramme de la chaîne de codage de l'algorithme JPEG2000.*

### *5.1.2 Prétraitement de l'image*

Dans la norme JPEG2000, chaque image est découpée en un ensemble de tuiles sans recouvrement. Le but de cette opération est de réduire la complexité de l'algorithme pour des images de très grandes tailles mais aussi de faciliter la navigation à l'intérieur de telles images. Ces tuiles sont de tailles carré (64x64) ou (128x128) par exemple.
Comme pour l'algorithme JPEG qui code les coefficients DC de chaque bloc de manière différentielle, l'algorithme retire la valeur moyenne de l'image avant d'effectuer la transformation. Les coefficients de basse-fréquence pourront ainsi être codés sur un nombre de bits moins important. Cette étape intègre également une transformation de l'espace des couleurs RGB en l'espace YCrCb. Cette transformation se justifie par le fait que les dégradations chromatiques ont un effet moindre sur l'image que les dégradations achromatiques. Ce qui permet de sous échantillonnage plus important pour les composantes achromatiques.

### *5.1.3 La transformée en ondelettes discrète(DWT)*

La transformation en ondelettes discrète provient de l'analyse multi-résolution qui a été développée par *Mallat* et *Meyer*. Le but de cette théorie est de décomposer un signal suivant différentes résolutions. Il est procédé ainsi à une décorrélation de l'information qu'il contient. Les basses résolutions représentent la forme grossière du signal tandis que les hautes résolutions encodent les détails du signal. Pour des signaux 2D comme les images, des propriétés topologiques (orientations, agencement du contenu) sont ainsi conservées après la transformation.
JPEG2000 effectue une décomposition en ondelettes discrètes sur plusieurs niveaux, spécifiés jusqu'à 5. Cette transformée peut être configurée à pertes (filtre 9/7 à coefficients non entiers ou de *Daubechies*) ou sans perte (filtre 5/3 à coefficients entiers). Les pixels en entrée ont une dynamique maximale de 12 bits pour la compression sans perte, et de 10 bits autrement.

### *5.1.4 Quantification des sous-bandes*

La Figure 24 représente une décomposition en sous-bandes en utilisant la DWT. Les sous-bandes de résolutions supérieures possèdent un contenu relativement faible alors que les sous-bandes de basses fréquences sont beaucoup plus riches. JPEG2000 adopte une quantification linéaire pour chaque sous-bande. Le pas de quantification utilisé est cependant beaucoup plus faible pour les sous-bandes de basses fréquences.

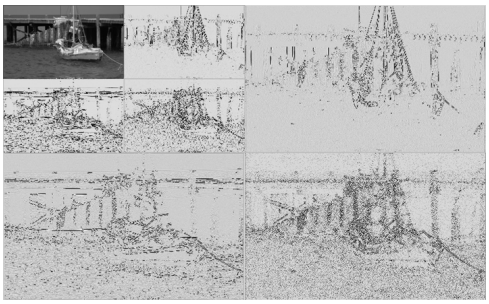


Figure 24 – *décomposition en ondelettes (3 niveaux)*.

### *5.1.5 Codage des sous-bandes*

Le codage de chaque sous-bande s'effectue par plan de bits, cette opération permet d'obtenir une « scalabilité » du fichier généré, car l'information importante (i.e. les bits de poids forts) seront codés dans un premier temps alors que les détails (i.e. les bits de poids faibles) sont codés après.

### *5.1.6 Codage par régions d'intérêt*

L'algorithme JPEG2000 permet également de coder différentes régions de l'image avec des qualités différentes.
Cette fonctionnalité est mise en oeuvre en augmentant le nombre de bits de poids forts des coefficients appartenant à une zone d'intérêt. De part l'orientation du codage, ces coefficients seront alors codés en priorité. Il existe deux manières différentes de coder une région d'intérêt :

- Par insertion d'un masque de forme dans le fichier compressé.
- Par doublement du nombre de bits appartenant aux zones d'intérêts.

Figure 25 – *Codage par ROI d'une image.*

La Figure 25 illustre l'utilité d'une telle technique.

## 4.5 MPEG4

Des millions de DVD, de récepteurs satellites, et de outils de médias vidéo ont utilisé les schémas de compression définis dans les standards de la famille MPEG (Motion Picture Experts Group). La plupart des documents vidéo emploient la norme MPEG-2, mais cela peut bientôt changer car le comité MPEG a approuvé l'année dernière la norme MPEG-4. En plus de quelques améliorations, MPEG-4 fournit un processus de compression appelé AVC (Advanced Video Coding) qui divise par deux le débit pour la même qualité que MPEG-2.
Malgré sa popularité, la norme MPEG-2 a montré quelques défauts dans son implémentation. Les défauts majeurs étant les effets de bloc pour les bas débits de compression et une dérive de la qualité d'image quand l'encodeur et le décodeur perdent la corrélation. De plus, beaucoup d'experts estiment que les algorithmes de base dans MPEG-2 ont atteint leurs limites en termes d'efficacité de compression. En attendant, la demande d'un contenu de résolution plus élevé et de basses largeurs de bande exige l'amélioration continuelle de l'efficacité.
Pour satisfaire ces demandes, le comité MPEG a défini la norme MPEG-4. Cette norme dépasse la compression vidéo puisqu'elle définit un ensemble d'outils pouvant être utilisés pour coder et transmettre le contenu multimédia incluant des vidéos, du texte, du son, et des animations en tant qu'objets indépendants, et permettant à chaque type de supports d'être manipulé indépendamment pour une efficacité maximale. La norme comprend également la qualité scalable d'image, la gestion numérique des droits, et l'interactivité.

### *5.1.7 MPEG-4 : une boite à outils*

Bien que la compression d'images soit ce qui vient à l'esprit en parlant de MPEG-4, elle ne représente qu'un outil parmi d'autres.
La scalabilité est un des outils que fournit MPEG-4, et qui permet à un flux vidéo de s'adapter à des canaux de transmission de différentes bandes passantes ou de s'adapter à la capacité variable d'un canal. La norme permet au flux de données de fournir une image de base avec des améliorations de la qualité que le décodeur peut appliquer.
Pour favoriser la compatibilité entre les implémentations de l'encodeur et du décodeur MPEG-4, le comité a défini une série de « profil » de compression définissant les différentes étapes. Ces profils indiquent les outils de compression devant être utilisés, les niveaux de résolution pour des processus tels que la compensation et la quantification de mouvement, et le type de codage de données à utiliser.

### *5.1.8 Compression MPEG de base*

Avant d'aborder les détails des profils MPEG-4, il est utile d'avoir une idée des étapes de transformation de base utilisées dans le processus de compression vidéo. Ces étapes, commune à toutes les normes MPEG, sont décrites dans la Figure 26. Elles incluent le choix de mode, la compression temporelle, la compression spatiale, et le codage entropique. Bien que cette description se concentre sur la composante de luminance, les étapes de

transformation s'appliquent de la même façon aux composantes de chrominance de l'espace YCrCb. Le traitement des composantes de chrominance utilise cependant une résolution inférieure que celle appliquée à la luminance, traduisant une sensibilité moindre aux dégradations chromatique.

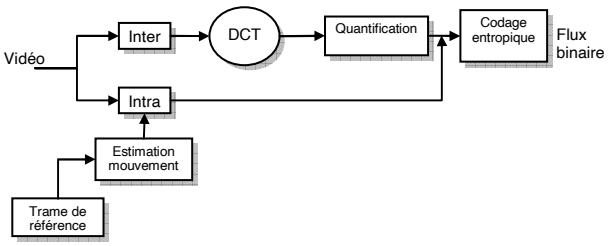



Figure 26 : *Schéma basic de compression de MPEG*

La première étape, le choix du mode, divise la trame entrante en tableaux rectangulaires de 16x16 pixels, appelés macro-blocs. Pour chaque macro-bloc, l'encodeur choisit alors l'usage du codage inter ou intra. Le codage Intra utilise seulement l'information contenue dans la trame d'image courante et produit un résultat compressé appelé « trame-I ». Les données compressées employant seulement des informations des trames précédentes s'appellent « trame-P », tandis que celles qui utilisent des données à la fois avant et après la trame courante s'appellent « trame-B ».

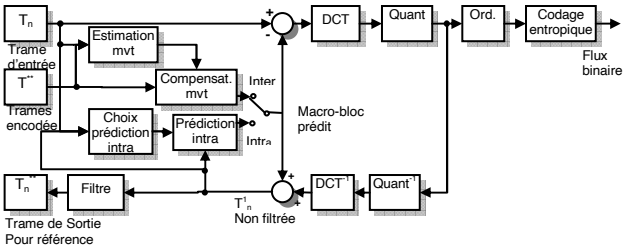



Figure 27 : *Schéma MPEG-4 avec prédiction*

Si le codage inter-trame est sélectionné, la prochaine étape consiste en la compression temporelle. Celle-ci est accomplie en appliquant l'estimation et la compensation de mouvement au macro-bloc. L'encodeur balaye les macro-blocs des trames de référence stockées afin de retrouver des correspondances. Il code alors le macro-bloc courant comme un vecteur décrivant le mouvement du macro-bloc correspondant, plutôt que de coder les valeurs des pixels.
Après la phase temporelle, l'étape de compression spatiale applique un transformée en cosinus discret (DCT) afin de convertir les données de l'image en données de fréquences spatiales. L'encodeur lit alors les entrées du tableau des fréquences spatiales en zig-zag (voir Figure 27), ce qui produit un flux de données périodique. Cette étape est suivie de celle de la quantification où la compression réelle est effectuée.

Enfin, les flux des données image, des vecteurs de mouvement, et des autres informations requises pour la reconstruction d'image sont combinés et leur entropie codée afin d'obtenir le flux binaire final.

### *5.1.9 Codage entropique*

La plupart des implémentations MPEG utilisent des algorithmes de codage à longueur variable (VLC : Variable Length Coding), tels que le codage de *Huffman*, pour réduire le nombre moyen de bits par mot. L'AVC utilise le codage arithmétique binaire à contexte adaptatif (CABAC : Context-Adaptive Binary Arithmetic Coding), ce qui permet d'atteindre des performances de codage très intéressantes.

# 6 Le SVH et les standards de codage.

Etant donné les connaissances acquises sur les comportements du système visuel humain, l'étape suivante a consisté en l'intégration de ces propriétés dans un schéma de compression. Certains systèmes de compression permettent une meilleure exploitation des propriétés du SVH que d'autres. Le
Tableau 2 montre les outils intégrés par JPEG2000 en comparaison à ceux de JPEG.

| **Propriétés SVH** | **JPEG2000** | **JPEG** |
|---|---|---|
| Pondération de fréquence | + | + |
| Pondération de couleur | + | + |
| pondération visuelle progressive | + | - |
| Masquage de voisinage | + | - |
| Auto-masquage | + | - |
| Masquage étendu par paire | + | - |
| Adaptation à lumière locale | + | - |
| Excentricité | + | - |
| Fréquence temporelle | + | - |

Tableau 2 *: Outils d'optimisation visuelle intégrés par JPEG2000 et JPEG.*

## 6.1 JPEG

Le principe habituellement retenu dans la construction de la matrice de quantification (MQ) est guidé par la qualité de reconstruction de l'image. En effet, l'objectif est d'aboutir à une qualité de reconstruction optimale pour un taux de compression donné.
Ainsi la norme ISA/EIC DIS 10918-3 valide une extension de la norme JPEG en autorisant un codage adaptatif limité pour une image. Alors que la MQ reste inchangée pour toute l'image, il est possible d'utiliser un coefficient pondérateur $m_k$ pour chacun des blocs k. On a alors :

$$\hat{f}(u,v,k) = \mathrm{round}(f(u,v,k)/(m_k.q(u,v)))$$

Cependant cette pondération n'est appliquée que sur les coefficients AC, tandis que la quantification des coefficients DC ne change pas.
De manière plus générale, si l'on veut optimiser la qualité des images JPEG, l'étape de définition de la MQ dépendra

de la visibilité des erreurs de quantification aux différentes fréquences TCD, et de l'intégration ou non de caractéristiques du SVH.

Deux approches philosophiques peuvent prétendre aboutir à une telle définition de la MQ : 1) une approche perceptuelle indépendante de l'image [9] et 2) une approche perceptuelle dépendante de l'image [10].

Dans le premier cas, la MQ est construite en fonction de la visibilité des erreurs de quantifications selon les fonctions de base de la TCD. Ainsi, Peterson *et al.* ont mesuré les seuils d'amplitudes des fonctions de base. Pour chaque fréquence (u,v) ils ont détecté, à l'aide de tests psychophysiques, les valeurs de seuils en deçà desquels les erreurs sont invisibles. Cette approche, bien que perceptuelle, est indépendante de l'image car la MQ est construite indépendamment de tout contenu d'une image. Ahumada *et al.* ont amélioré la définition des seuils de détection et introduisant une formule dans laquelle sont prises en compte la luminance de l'écran, la taille de pixels, ainsi que plusieurs autres propriétés liées à l'affichage.

Néanmoins, le principal inconvénient de cette approche concerne la MQ qui est construite indépendamment du contenu des images. Si les seuils visuels des artefacts étaient indépendants du contenu d'une image, cela ne poserait pas de problème. Or ce n'est pas le cas. On peut citer par exemple l'effet de masquage dont le seuil de détection varie nécessairement en fonction des signaux masquant et masqué.

Concernant l'approche perceptuelle dépendante de l'image, la construction de la MQ est liée au contenu de l'image. Les modèles utilisés lors de la construction de la MQ contiennent tous une fonction $t_b(u,v)$ permettant de calculer le seuil de visibilité des 64 fonctions de base comme étant une fonction de la fréquence, de la moyenne de luminance d'un bloc et des canaux couleur perceptuels, où b représente un axe couleur de l'espace initial. Un grand nombre d'espaces couleur ont été proposés afin de prendre en compte la discrimination chromatique. Les axes chromatiques retenus sont ceux préconisés par Boynton, i.e. un axe d'opposition de couleur rouge-vert (que l'on appellera l'axe O) et un canal bleu (axe Z). Le passage de l'espace CIE 1931 XYZ à l'espace YOZ est donné comme suit :

$$[YOZ] = [XYZ]\ {}_{XYZ}M_{YOZ} = [XYZ] \begin{bmatrix} 0 & 0.47 & 0 \\ 1 & -0.37 & 0 \\ 0 & -0.10 & 0 \end{bmatrix}.$$

Afin de prendre en compte les effets de masquage de luminance, les seuils sont ajustés par une fonction puissance des coefficients DC de chaque bloc sur le canal de luminance $f(0,0,k)$ relativement au coefficient DC correspondant à la luminance moyenne $\bar{f}(0,0)$ :

$$a(u,v,k) = t(u,v).\left(f(0,0,k)/\bar{f}(0,0)\right)$$

Puis dans une seconde étape les seuils sont ajustés en fonction du phénomène de masquage au contraste. Le facteur d'ajustement est une fonction puissance définie par :

$$m(u,v,k) = a(u,v,k)\max\left[1, \left|\frac{f(u,v,k)}{a(u,v,b)}\right|^{\omega_{u,v}}\right]$$

où $\varpi_{u,v}$ varie en fonction de la fréquence et du canal perceptuel couleur. Sa valeur est généralement fixée à 0.7.

Les erreurs de quantification, dans l'espace couleur perceptuel, sont ensuite divisées par les seuils de luminance et de contraste de façon a aboutir à une mesure de différence juste notable, connue sous le terme *just noticeable differences (jnd)*. :

$$j(u,v,k) = e(u,v,k)/m(u,v,k)$$

La matrice des jnds j(u,v,k) peut être interprétée comme une mesure simple de la visibilité des artefacts pour une des bandes de fréquence définie par les 64 fonctions de base de la DCT. Cette matrice est également connue sous le terme « matrice d'erreur perceptuelle ». Elle permet de définir les fréquences pour lesquelles l'erreur est la plus importante pour une image donnée et pour une MQ particulière.

Afin d'optimiser cette matrice, il est nécessaire de disposer d'une valeur de la qualité de l'image reconstruite. Une manière simple d'obtenir une mesure de la qualité est de combiner les éléments de la matrice en utilisation la sommation de Minkowski sur chacun des canaux perceptuels

$$p(u,v) = \left(\sum_k |j(u,v,k)|^{\beta_s}\right)^{1/\beta_s}$$

Ces matrices d'erreurs sont ensuite sommées entre-elles de façon à aboutir à une erreur perceptuelle totale pour chaque canal perceptuel couleur :

$$p = \left(\sum_{u,v} p(u,v)^{\beta_f}\right)^{1/\beta_f}$$

Il devient alors possible d'optimiser les valeurs p(u,v,k) de la matrice d'erreur perceptuelle de façon à aboutir à un taux de compression maximal en fonction de la valeur P fixée, ou, inversement, aboutir à la valeur de P minimale (donc qualité maximale) pour un taux de compression donné.

On remarque que si $\beta_f$=0 et $\beta_s$=0, alors la valeur de P est équivalente à la valeur de maximale des $p(u,v,k)$. Intuitivement, si le maximum de $p(u,v,k)=\psi$, alors chacun des élément $p(u,v,k)$ de la matrice d'erreur perceptuelle doit prendre pour valeur de façon a pouvoir atteindre la valeur P fixée sans diminuer le taux de compression. Etant donné que chaque valeur des $p(u,v,k)$ correspond (au moins d'une manière monotone) à la visibilité d'une classe d'artefacts, cette stratégie revient nécessairement à lisser la visibilité de chacune des classes.

Sous les hypothèses $\beta_f$=0 et $\beta_s$=0, chacune des entrées de la matrice d'erreur perceptuelle $p(u,v,k)$ peut être considérée comme une fonction indépendante de la valeur $q(u,v)$ de la matrice de quantification

$$p(u,v,k) = g_{u,v}(q(u,v))$$

Où $g_{u,v}$ est une fonction monotone croissante telle que $g_{u,v}(1)=0,\ \forall\ u,v$.

Il est alors possible d'estimer la valeur $\hat{q}(u,v)$ de la matrice de quantification par :

$$g_{u,v}(\hat{q}(u,v)) = \psi\ \ \forall u,v$$

## 6.2 JPEG2000

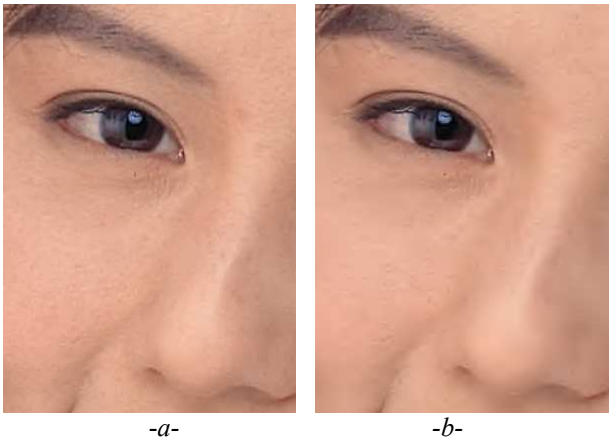
### *6.2.1 Pondération par fréquence visuelle*

La stratégie d'optimisation visuelle commune pour la compression d'images consiste à exploiter la fonction de sensibilité de sensibilité au contraste (CSF) qui caractérise la sensibilité variable du système visuel humain aux fréquences spatiales 2D.
Dans JPEG2000, trois tables de pondération ont été recommandées pour trois distances standard d'observation. Des tables de pondération pour les images couleur ont aussi été recommandées. Un exemple est donné dans le Tableau 3. Il est possible de remarquer que les poids pour les sous-bandes de basses fréquences sont plus importants que ceux des sous-bandes de hautes fréquences.

| N° | Y (LH HL HH) | Cb (LH HL HH) | Cr (LH HL HH) |
|---|---|---|---|
| 1 | 0.275783 0.275783 0.090078 | 0.089950 0.089950 0.027441 | 0.166647 0.166647 0.070185 |
| 2 | 0.837755 0.837755 0.701837 | 0.267216 0.267216 0.141965 | 0.375176 0.375176 0.236030 |
| 3 | 0.999994 0.999994 0.999988 | 0.488887 0.488887 0.348719 | 0.587213 0.587213 0.457826 |
| 4 | 1.000000 1.000000 1.000000 | 0.679829 0.679829 0.567414 | 0.749805 0.749805 0.655884 |
| 5 | 1.000000 1.000000 1.000000 | 0.812612 0.812612 0.737656 | 0.856065 0.856065 0.796593 |

Tableau 3 : *Exemple de facteurs de pondération*

Des pondérations de fréquence et de couleur appropriées permettent en général une meilleure conservation des détails et des textures sans l'introduction de distorsions couleur (voir Figure 28). La pondération par fréquence permet également de réduire les artéfacts de *flickering* dans motion-JPEG2000.



*-a-* *-b-*

Figure 28 : *Portions de l'image "woman" compressée en utilisant JPEG2000 à 0.75 bpp. a- JPEG2000 avec pondération de fréquence et de couleur; b- JPEG2000 sans pondération.*

Pondération par fréquence fixe

Pour une décomposition en ondelettes discrètes, il est admis qu'un poids est attribué à chaque sous-bande afin de simplifier l'implémentation. Cette approche est connue sous le nom de pondération par fréquence fixe. Ces pondérations ne sont pas explicitement transmis au décodeur et peuvent être incorporé de deux façons :

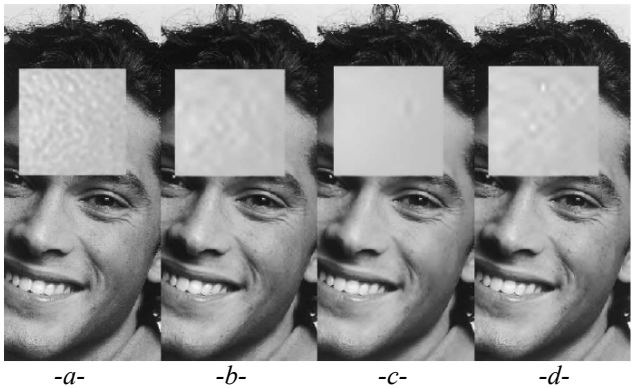
1. **Modification du pas de quantification**: au codage, le pas de quantification $q_i$ des coefficients de la sous-bande $i$ est ajusté pour être inversement proportionnel aux pondérations CSF $w_i$. L'information de pondération CSF est reflétée au niveau des pas de quantification qui sont transmis explicitement pour chaque sous-bande.
2. **Modification de l'ordre de codage** : dans cette implémentation, les pas de quantification ne sont pas modifiés, mais les poids de distorsion utilisé dans l'optimisation débit/distorsion le sont à la place, linéairement proportionnel au poids de CSF pour chaque sous-bande.

Pondération par fréquence progressive

JPEG2000 permet l'implémentation de pondérations visuelles progressives, où différents ensembles de poids de CSF peuvent être appliqués à différentes étapes afin de générer les différentes couches de qualité. En particulier, pour mettre en application les pondérations visuelles progressives, le *VM* (V*erification Model*) de JPEG-2000, change en cours d'exécution, l'ordre dans lequel les plans de bits d'un code-bloc doivent apparaître dans le bitstream global basé sur plusieurs ensembles de pondération de fréquence destinés à différents débits.

### *6.2.2 Masquage visuel*

La pondération de fréquence est en général très efficace pour les applications utilisant soit des supports d'affichage de haute résolution ou des distances d'observation importantes. Dans les deux cas, la distance d'observation exprimée en pixels est supérieure à 1500.
L'avantage de cette technique devient moins évident pour des supports de basses résolutions ou pour des distances très faibles, puisque la courbe de CSF recalée tend à devenir plate dans ces conditions d'observation. Dans ce cas, le masquage visuel est plus approprié pour l'amélioration de la qualité.



*-a-* *-b-* *-c-* *-d-*

Figure 29 : *Résultats de masquage visuel. a- image originale, b- masquage étendu par point, c- masquage de voisinage et d- auto-masquage.*

Dans les applications de compression, l'image joue le rôle du fond qui réduit la visibilité d'un signal généré par la distorsion. JPEG2000 permet l'exploitation de l'auto-masquage, du masquage de voisinage et le masquage étendu par point. Les approches de masquage visuel dans JPEG2000 permettent la scalabilité du flux binaire contrairement aux approches précédentes.
La Figure 29 montre quelques résultats de masquage visuel. Nous pouvons noter que le masquage étendu par point permet une meilleure conservation des détails fins. Dans ce cas, le masquage de voisinage donne les résultats les moins intéressants.

Les divers outils d'optimisation visuelle peuvent être combinés pour maximiser les performances visuelles. Il a été observé que, pour quelques images complexes avec un contenu divers, l'amélioration visuelle peut être équivalente à une économie allant jusqu'à 50% du débit binaire. Enfin, JPEG2000 permet également l'exploitation d'autres propriétés du SVH telles que l'adaptation à la lumière locale, la sensibilité aux fréquences temporelles, etc. Ceux-ci ont pu être certaines des futures matières de recherches.

# 7 Conclusion

Dans ce cours, les différentes approches concernant la mesure de la qualité des images couleur et des vidéos ont été abordées. Les concepts inhérents au système visuel humain ont été présentés et leurs mises en œuvre dans l'élaboration d'une mesure de qualité et l'amélioration des schémas de compression d'images fixes et de vidéos illustrées.
Les avancées scientifiques des métriques de qualité et des standards de compression sont intrinsèquement liées aux résultats des études menées par les chercheurs dans le domaine de la neurophysiologie de la psychophysique et de la psychologie de la vision.

# Références


[1] G. Van der Horst and M. A. Bouman. Spatiotemporal chromaticity discrimination. Journal of Optical Society of America, 59, 1482-1488, 1969.

[2] S.Daly "A visual model for optimizing the design of image processing algorithm" I.C.I.P, Vol. II of III, pp. 16-20, 1994.

[3] J.Lubin, "The use of psychophysical data and models in the analysis of display system performance" A.B.Watson (Ed.), Digital Images and Human Vision, MIT press, Cambridge, MA, pp. 163-178, 1993.

[4] A.B.Watson "The cortex transform: Rapid computation of simulated neural images" Computer Vis. Graphics and image proces. N°. 39, pp. 311-327, 1987.

[5] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind Measurement of Blocking Artifacts in Images", Proc. IEEE Int. Conf. on Image Processing, Sep. 10-13, 2000, vol. III, pp. 981-984, Vancouver, Canada.

[6] Alan C. Bovik and Shizhong Liu, "DCT-Domain Blind Measurement Of Blocking Artifacts In DCT-Coded Images", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, pp. 1725-1728, Salt-lake City, Utah, 07-11 Mai 2001.

[7] M. Rabbani and R. Joshi, "An Overview of the JPEG2000 Still Image Compression Standard," Signal Processing: Image Communication Journal, Volume 17, Number 1, October 2001.

[8] D. Taubman and P. Marcellin, "JPEG2000: Image Compression Fundamentals, Practice and Standards", Kluwer Academic Publishers, 2001.

[9] H. A. Peterson, H. Peng, J. H. Morgan and W. B. Pennebaker, "Quantization of color image components in DCT domain", in SPIE Human Vision, Visual Processing, and Digital Display, 1453, pp 210-222, 1991

[10] A. B. Watson. "Perceptual Optimization of DCT color quantization Matrices", in ICIP94, (Austin, USA), nov. 1994.