



Etat de l'art des critères objectifs de qualité d'images et de vidéos

Fadi Boulos

21 avril 2006

Rapport Bibliographique - Master Recherche
"Architectures Logicielles Distribuées"

Encadreur : Dominique Barba

Ecole Polytechnique de l'Université de Nantes

“I find television very educating. Every time somebody turns on the set, I go into the other room and read a book”

Groucho Marx (1890-1977)

Table des matières

Introduction générale	4
I Qualité visuelle d'images et de vidéos	8
Introduction de la première partie	9
1 Méthodologie de la mesure de la qualité visuelle	11
1.1 Méthodes catégorielles	11
1.1.1 Méthodes à simple stimulus	11
1.1.2 Méthodes à double stimuli	12
1.2 Méthodes comparatives	13
1.3 Vidéo courte vs vidéo longue	13
Conclusion de la première partie	14
II Critères objectifs de qualité	16
Introduction de la seconde partie	17
2 Évaluation des performances d'un critère objectif de qualité	18
2.1 Coefficient de corrélation linéaire	18
2.2 Coefficient de corrélation de rang	19
2.3 Erreur de prédiction de la qualité	20
2.3.1 Écart-type de l'erreur de prédiction de la qualité	20
2.3.2 Écart-type de l'erreur de prédiction de la qualité pondérée par l'intervalle de confiance à 95%	20
2.4 Indicateur de mauvaise prédiction	21
2.5 Coefficient de Kappa	21
2.6 Coefficient de Kurtosis	22

3	Critères de qualité d'images fixes	23
3.1	Fonctionnement du SVH	23
3.1.1	L'œil humain	23
3.1.2	Le cortex visuel primaire	24
3.1.3	Le cortex visuel secondaire	24
3.2	Phénomènes liés au fonctionnement du SVH	25
3.2.1	Le contraste	25
3.2.2	Fonction de sensibilité au contraste (CSF)	26
3.2.3	Effet de masquage	27
3.3	Critères de qualité d'images avec référence complète	27
3.3.1	Critères de qualité non basés sur le SVH	28
3.3.2	Critères de qualité basés sur le SVH	30
3.4	Critères de qualité avec référence réduite	30
3.4.1	Critère de Carnec	31
3.4.2	Critère de Gunawan et Ghanbari	32
3.5	Critères de qualité sans référence	33
3.5.1	Critère de Luo	33
4	Critères de qualité de vidéos	34
4.1	Différences entre critères de qualité d'images fixes et de vidéos	34
4.2	Critères de qualité avec référence complète	35
4.2.1	Critères de qualité non basés sur le SVH	35
4.2.2	Critères de qualité basés sur le SVH	36
4.3	Critères de qualité avec référence réduite	36
4.3.1	Critère de Webster <i>et al.</i>	36
4.3.2	Critère de Wolf et Pinson	37
4.4	Critères de qualité sans référence	37
4.4.1	Critère de Le Callet <i>et al.</i>	37
4.4.2	Critère de Sugimoto <i>et al.</i>	38
4.5	L'obtention d'une note de qualité	38
4.5.1	Méthodes de cumul	39
4.5.2	Réseaux de neurones	40
	Conclusion de la seconde partie	41
III	Évaluation objective de qualité de vidéos améliorées par des post-traitements d'affichage	43
	Introduction de la troisième partie	44

5	Les nouvelles technologies d'écrans	45
5.1	Les écrans Plasma	45
5.2	Les écrans LCD	46
5.3	Les écrans CRT	46
6	Amélioration de la qualité de vidéos : l'exemple intégré de Pixel Plus 3 HD	47
6.1	L'augmentation de la netteté	47
6.2	La réduction du bruit	48
6.3	La réduction du flou	48
6.4	L'augmentation du contraste	48
6.5	L'augmentation de la saturation des couleurs	49
7	La mesure d'amélioration de qualité de vidéos	50
7.1	Mesures de la netteté des contours	50
7.2	Mesures du contraste	51
7.3	Mesures de la qualité des couleurs	53
	Conclusion de la troisième partie	53
	Conclusion générale et perspectives	54
	Bibliographie	56

Introduction générale

La Télévision numérique Haute Définition (*High Definition Television* : HDTV) est une nouvelle technique de diffusion visant à produire une qualité d'image supérieure à celle de la télévision numérique classique (*Standard Definition* : SDTV). Les premières chaînes de HDTV sont nées aux Etats-Unis fin 1998 et ont été vite suivies par de nombreuses chaînes dans beaucoup de pays au monde (Canada, Brésil, Japon, Australie...). En Europe, la chaîne de diffusion Euro1080 (Belgique) a commencé à diffuser en HD début 2004 et la chaîne anglaise NTL a annoncé en mars 2006 la possibilité de réception de la HDTV sur tout le territoire anglais. Le projet européen "High Definition TV For Europe" (HD4U) sur lequel travaille l'équipe IVC du laboratoire IRCCyN en collaboration avec plusieurs partenaires (TF1, Thomson, Philips, Euro1080, Nextream, Newtec, MaxiSat) a pour objectif d'établir une chaîne complète de diffusion HD.

La qualité remarquable de la TVHD s'exprime surtout par la visualisation de détails plus fins et un champ de vision élargi tel que le téléspectateur se sent davantage immergé dans l'image. Pour obtenir cette qualité d'image, nettement supérieure à celle obtenue dans le cas de la SDTV, les écrans utilisés dans les téléviseurs HD sont des écrans LCD (Liquid Crystal Display) ou des écrans Plasma dont la définition varie de 1280 pixels par ligne x 720 lignes jusqu'à 1920 x 1080 au format 16 : 9.

Le prix de ces écrans étant trop élevé et le nombre de chaînes diffusant en HD limité (à l'échelle mondiale), l'EICTA (*European Information & Communications Technology Industry Association*) a défini la norme *HDTV-Ready* [1] qui standardise les écrans "potentiellement" capables d'afficher une qualité HD. En effet, tout écran portant le logo *HD-Ready* (appelé aussi *HDTV-Compatible*) peut afficher une vidéo en HD même si initialement elle a été diffusée en SD. La résolution minimale de tels écrans est 720 lignes en mode progressif (720p) pour 1280 pixels/ligne et peut atteindre 1080 lignes en mode entrelacé pour 1920 pixels/ligne.

D'autre part, la norme *Full HDTV*, appelée aussi *True HDTV* représente la

meilleure qualité HDTV visualisée sur des écrans dont la résolution est de 1080 lignes en mode progressif (1080p) pour 1920 pixels/ligne.

Dans le cadre du projet HD4U, la nouvelle norme de codage vidéo H.264/AVC développée conjointement par l'ITU-T et l'ISO/CEI a été adoptée. Ce codage permet un taux élevé de compression de la vidéo tout en préservant une très bonne qualité ce qui rend son utilisation compatible avec de nombreuses applications (visiophonie, visioconférence...).

L'amélioration de la qualité d'une vidéo avant de la diffuser est une étape nécessaire pour obtenir la meilleure qualité visuelle. Cette amélioration peut se faire à deux niveaux : au niveau du décodage de la vidéo (c'est le cas du filtrage pour réduire les effets de bloc dans le décodage H.264/AVC [2]) ou bien au niveau de l'affichage (c'est le cas des traitements intégrés dans le système "Pixel Plus 3 HD" de Philips [3]). Quand l'amélioration de la qualité d'une vidéo se fait au niveau du décodage, elle vise surtout à diminuer autant que possible (voire à éliminer) les dégradations liées au codage. Les dégradations les plus fréquentes sont les effets de blocs, le flou, les effets "d'oscillations" (*ringing artifacts*) et le bruit. Comme ces dégradations ne sont pas indépendantes les unes des autres (par exemple en essayant d'éliminer le flou on peut augmenter le bruit), la mesure de leur effet gênant est une tâche difficile.

D'autre part, quand l'amélioration de la qualité se fait au niveau de l'affichage, elle s'effectue généralement en plusieurs étapes (appelées post-traitements) visant chacune à améliorer un aspect de la vidéo : ses couleurs, sa netteté ou son contraste par exemple. Elle peut aussi tenter de réduire le bruit. Cette amélioration est apportée par une modification de paramètres internes à l'écran. De nouveau se pose un problème de mesure mais cette fois-ci celui de la mesure de l'amélioration amenée par ces traitements.

Pour évaluer la qualité d'une vidéo, différentes approches ont été adoptées dans la littérature dont la majorité avait comme objectif l'évaluation d'une version dégradée de la vidéo originale plutôt que l'évaluation d'une version améliorée. Nous pouvons classer ces approches en deux grandes classes : l'approche "subjective" où des observateurs humains évaluent la qualité de la vidéo et l'approche "objective" où un algorithme appliqué à une vidéo donnée lui attribue une note, dite note de qualité. Mais les performances de ces critères (ou algorithmes) ne sont pas toujours à la hauteur des objectifs visés, ceci pour plusieurs raisons : la complexité du traitement de vidéos, la difficulté de modéliser le processus humain de perception et la complexité

calculatoire des différents critères objectifs élaborés (ce qui empêche leur utilisation dans des applications temps réel).

Le besoin de nouveaux critères de qualité est évident ; en particulier, celui de critères pouvant être appliqués à une vidéo améliorée par des traitements au niveau de l’affichage (vu la rareté de tels critères). Dans le cas d’une vidéo améliorée, pour juger si sa qualité a bien été améliorée et dans quelle proportion, nous avons besoin d’une métrique qui mesure la qualité d’une vidéo avant et après l’amélioration et en mesure ainsi le gain. Nous allons dans ce mémoire dresser un état de l’art des différentes approches d’évaluation de la qualité d’images, étape préliminaire à une bonne compréhension des approches concernant la vidéo que nous allons détailler par la suite.

Ce mémoire est composé de trois grandes parties. Dans la première partie, nous présentons la méthodologie de mesure de la qualité visuelle d’images et de vidéos appelée aussi évaluation subjective de qualité.

Dans la seconde partie, intitulée “Critères objectifs de qualité”, nous dressons un véritable état de l’art des critères objectifs de qualité présents dans la littérature. Nous commençons cette partie par la description des méthodes d’évaluation des performances de tels critères. Ensuite, nous décrivons les principaux critères objectifs de qualité d’images de la littérature ; dans ce chapitre, une description du Système Visuel Humain (SVH) et des principaux phénomènes le caractérisant est proposée. Les différents types de critères sont ensuite détaillés et les critères objectifs de qualité de vidéos sont proposés. Enfin, une explication des différentes méthodes adoptées dans l’obtention d’une note de qualité est donnée.

La troisième partie de ce mémoire traite l’évaluation objective de l’augmentation de qualité d’une vidéo par des post-traitements d’affichage. Pour ce faire, les nouvelles technologies d’écrans sont expliquées et un exemple d’amélioration de qualité par les post-traitements d’affichage (“Pixel Plus 3 HD”) est donné. Enfin, nous présentons la méthodologie de l’évaluation objective de la qualité de vidéos améliorées.

Le mémoire est terminé par une conclusion générale et la proposition de perspectives pour orienter les travaux futurs.

Première partie

Qualité visuelle d'images et de vidéos

Introduction

La mesure de la qualité visuelle d'images et de vidéos, appelée classiquement "méthode subjective d'évaluation de la qualité d'images et de vidéos" implique des observateurs humains qui ont comme tâche de donner une note de qualité à une image (ou vidéo) donnée lors d'une séance de tests. La note attribuée à l'image (ou vidéo) est choisie parmi une échelle définie a priori. Les observateurs subissant ce test peuvent être des "experts" c'est-à-dire des personnes travaillant dans le domaine du traitement d'images ou des observateurs "naïfs" c'est-à-dire des personnes n'ayant aucune expérience dans le domaine du traitement d'images ou vidéos. Nous donnons dans la figure 1 un exemple d'image et de vidéo utilisées dans les tests subjectifs d'évaluation de qualité visuelle.

Les facteurs influençant le jugement d'un observateur sont essentiellement psychologiques : ils sont liés à son propre processus cognitif c'est-à-dire la façon dont il acquiert les informations visuelles en se basant sur ses connaissances antérieures. Ils dépendent aussi du contexte d'observation [4] : l'observateur porte son attention sur des détails différents selon le contenu de la vidéo (les visages dans une visioconférence, les articles dans un "*internet shopping*"...). Comme le processus d'évaluation de qualité est assez compliqué et les types de tests possibles sont nombreux, l'organisme international de standardisation ITU-R a défini les normes de l'évaluation subjective de la qualité d'une vidéo [5].

Généralement, les méthodes subjectives d'évaluation de qualité peuvent être classées en deux grandes classes : les méthodes catégorielles et les méthodes comparatives. Nous allons dans cette première partie présenter ces deux classes de méthodes (valables pour les images et vidéos) et les deux méthodes possibles d'évaluation de qualité de vidéos.

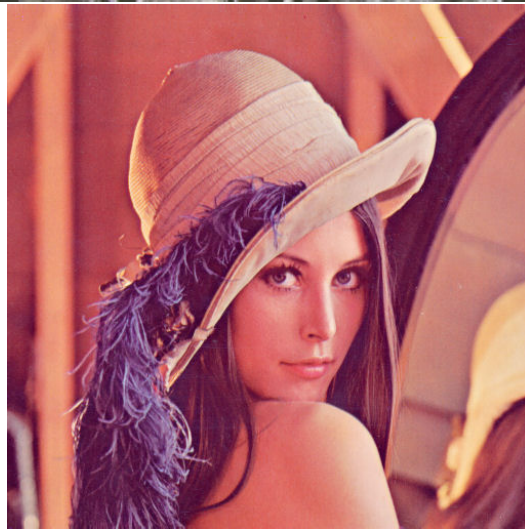


FIG. 1 – L'image "Lena" (en bas) et une trame de la vidéo "Parkrun" (en haut) utilisées dans les tests d'évaluation de qualité visuelle.

Chapitre 1

Méthodologie de la mesure de la qualité visuelle

Nous présentons dans ce chapitre les différentes méthodes d'évaluation de qualité visuelle d'images (ou vidéos). Ces méthodes sont normalisées par l'ITU-R.

1.1 Méthodes catégorielles

Les méthodes catégorielles d'évaluation de la qualité d'une image (ou vidéo) offrent à l'observateur une échelle de catégories : celui-ci doit attribuer à l'image (ou la vidéo) présentée une note de qualité située dans la catégorie qui lui paraît la plus proche de son jugement personnel. Cette échelle peut être explicitée sous la forme de vocabulaire qualificatif (adjectifs) allant de "mauvaise qualité" à "excellente qualité" par exemple ou bien sous une forme numérique, allant de 0 à 100. Dans ce dernier cas, l'échelle est dite continue. Un exemple d'échelles de catégories est présenté dans la figure 1.1.

Selon le mode de présentation des images (ou vidéos) à juger, nous pouvons classer les méthodes catégorielles en deux groupes : méthodes à simple stimulus et méthodes à double stimuli.

1.1.1 Méthodes à simple stimulus

Ces méthodes consistent à présenter à l'observateur l'image (ou vidéo) dont la qualité est à évaluer : il lui attribue une note de qualité puis après une période dite de latence lui sera présentée l'image (ou vidéo) suivante et ainsi de suite. L'échelle de catégories peut être élargie ou réduite selon la dynamique des dégradations à évaluer en se conformant toujours aux normes

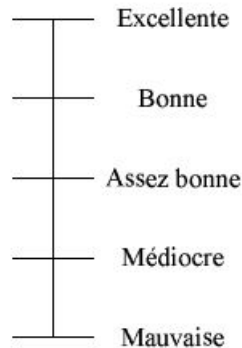


FIG. 1.1 – Un exemple d'échelle de catégories de qualité visuelle.

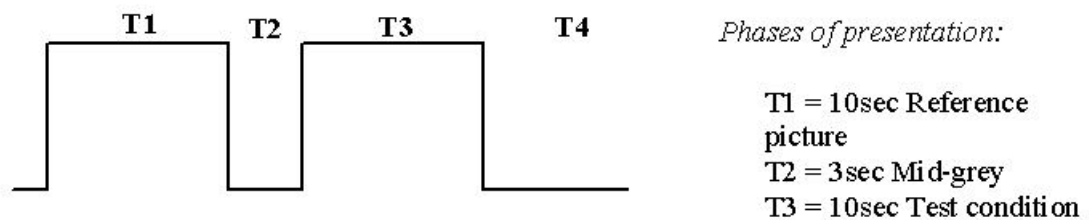


FIG. 1.2 – Le modèle d'un test à double stimuli [6].

de l'ITU. Ces méthodes sont surtout utilisées lorsque l'image (ou vidéo) de référence n'est pas disponible.

1.1.2 Méthodes à double stimuli

Ces méthodes consistent à présenter à l'observateur la version originale, dite de référence, de l'image (ou vidéo) à évaluer puis la version modifiée (dégradée ou améliorée) : l'observateur lui attribue une note de qualité durant la période de temps qui suit, appelée période de vote. Le déroulement d'un tel test est montré à la figure 1.2. La méthode DSCQS (*"Double-Stimulus Continuous Quality Scale"*) est recommandée par le VQEG (*"Video Quality Experts Group"*) pour les tests subjectifs en HDTV car elle permet l'évaluation d'un grand nombre de conditions de test en une seule session [6].

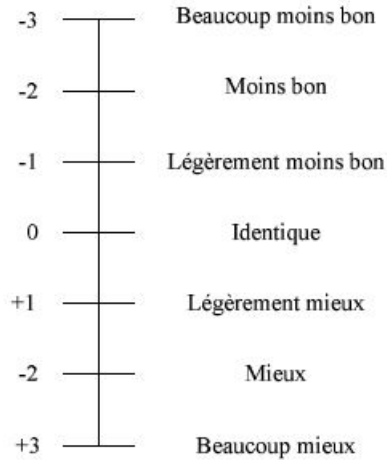


FIG. 1.3 – Une échelle comparative de qualité visuelle.

1.2 Méthodes comparatives

Les méthodes comparatives d'évaluation de la qualité d'une image (ou vidéo) consistent à noter la différence entre deux versions d'une même image (ou vidéo) à l'aide d'une échelle : les deux images (ou vidéos) sont montrées en même temps et l'observateur juge laquelle est de meilleure qualité. Un exemple d'échelle comparative est donné à la figure 1.3.

1.3 Vidéo courte vs vidéo longue

Les méthodes expliquées ci-dessus sont valables pour l'évaluation de la qualité visuelle d'images et de vidéos. Mais nous distinguons entre deux types de vidéos pouvant être utilisés lors des tests d'évaluation de la qualité visuelle : la vidéo courte et la vidéo longue. La durée d'une vidéo courte varie de quelques secondes à une dizaine de secondes ; elle est notée une seule fois à sa fin. Dans ce genre de test, l'observateur doit donner une note de qualité finale à l'ensemble de la vidéo.

D'autre part, la vidéo longue dont la durée est de quelques minutes voire même d'une dizaine de minutes est notée plusieurs fois ; en effet, l'observateur donne une note de qualité d'une façon périodique durant toute la durée d'affichage de la vidéo. Ce type de test est appelé "évaluation au fil de l'eau"

ou “*on the fly assessment*”. Dans ces tests, chaque note de qualité évalue la partie de vidéo affichée entre l’instant d’attribution de cette note et celui de la note précédente : cela permet une évaluation de plusieurs types de dégradations pour des contenus de scènes différents.

Conclusion

Les méthodes subjectives d'évaluation de la qualité d'une vidéo semblent être un moyen fiable pour évaluer l'information visuelle car dans la plupart des applications, l'œil humain est le terminal final qui va traiter cette information. Une nouvelle approche, non exploitée encore est celle de Pereira [7] qui étend la notion de qualité de service à la notion de qualité d'expérience englobant trois dimensions : les sensations, les perceptions et les émotions. Mais ces méthodes sont coûteuses en temps (le temps d'explication et de test), en ressources humaines (au moins 18 observateurs pour des tests d'évaluation de qualité et au moins huit observateurs pour des tests de perception) et en matériel (luminosité de la salle bien définie, écrans de très bonne qualité...). De plus, ces tests ne peuvent pas être reproduits (et par conséquent vérifiés) et dépendent fortement de la nature des vidéos évalués (scènes texturées ou uniformes, peu ou très animées...). C'est la raison pour laquelle d'autres méthodes d'évaluation ont été mises en œuvre : les critères objectifs de qualité que nous allons présenter dans la seconde partie.

Deuxième partie

Critères objectifs de qualité

Introduction

Un critère objectif de qualité est un traitement particulier qui produit, à partir des données particulières d'une image (ou vidéo), une note de qualité de cette image (ou vidéo). Les données en entrée varient selon le type du critère de qualité qui peut être de type référence complète FR (*Full Reference*), de type référence réduite RR (*Reduced Reference*) ou de type sans référence NR (*No Reference*). Les critères objectifs de qualité peuvent aussi être classés suivant la nature de l'information qu'ils traitent, c'est-à-dire s'ils se basent sur un modèle du SVH (Système Visuel Humain) ou bien s'ils n'intègrent aucune propriété de ce modèle. Une autre approche de classification est proposée dans [8] où les critères objectifs de qualité sont classés en boîte noire (seulement la note de qualité de la vidéo est calculée) ou boîte blanche (la note de qualité et les traits caractéristiques la composant sont calculés). Les critères objectifs de qualité peuvent être utilisés pour contrôler le gain des algorithmes d'amélioration de qualité, comme par exemple dans [9].

Dans cette partie, nous expliquons brièvement le fonctionnement du SVH. Ensuite, les principaux phénomènes (affectant la vision) liés à ce modèle sont présentés. Enfin, nous détaillons les différents types de critères objectifs (FR, RR, NR) et nous les classifions selon leur nature (basés sur le SVH ou sur le signal) tout en établissant un état de l'art des critères objectifs d'évaluation de la qualité d'une vidéo présents dans la littérature.

Chapitre 2

Évaluation des performances d'un critère objectif de qualité

Les performances d'un critère objectif d'évaluation de qualité sont mesurées en comparant les notes de qualité données par ce critère, quand il est appliqué sur un groupe donné d'images (ou vidéos), aux notes fournies par des observateurs humains lors de tests subjectifs d'évaluation de qualité (sur ce même groupe d'images ou vidéos). Le procédé de comparaison entre notes subjectives et notes objectives est de plusieurs formes [10] que nous explicitons dans cette partie.

2.1 Coefficient de corrélation linéaire

Le coefficient de corrélation linéaire CC exprime la dépendance linéaire entre les mesures objectives M_{obj} et les notes subjectives moyennes MOS ¹ (*Mean Opinion Score*). Il est donné par :

$$CC = \frac{1}{NbJK} \frac{\sum_{j=1}^J \sum_{k=1}^{K(j)} (MOS_{jk} - \overline{MOS})(M_{obj_{jk}} - \overline{M_{obj}})}{\sqrt{\sigma_{MOS}^2 \cdot \sigma_{M_{obj}}^2}} \quad (2.1)$$

avec :

- $NbJK = \sum_{j=1}^J K(j)$ le nombre d'images (ou vidéos) testées (on suppose que J versions originales ont subi chacune $K(j)$ dégradations différentes

¹La note subjective moyenne d'une image (ou vidéo) est la moyenne des notes données par les observateurs pour cette image (ou vidéo)

- $\overline{MOS} = \frac{1}{NbJK} \sum_{j=1}^J \sum_{k=1}^K (j) MOS_{jk}$ le MOS moyen
- $\overline{Mobj} = \frac{1}{NbJK} \sum_{j=1}^J \sum_{k=1}^K (j) Mobj_{jk}$ la mesure objective moyenne
- $\sigma_{MOS}^2 = \frac{1}{NbJK - 1} \sum_{j=1}^J \sum_{k=1}^K (j) (MOS_{jk} - \overline{MOS})^2$ la variance des MOS
- $\sigma_{Mobj}^2 = \frac{1}{NbJK - 1} \sum_{j=1}^J \sum_{k=1}^K (j) (Mobj_{jk} - \overline{Mobj})^2$ la variance des mesures objectives

La valeur du coefficient CC est toujours comprise entre -1 et 1. Une valeur proche de 1 ou de -1 indique une bonne corrélation.

2.2 Coefficient de corrélation de rang

Le coefficient de corrélation de rang CCR décrit combien les mesures objectives et subjectives de qualité, après remise en ordre propre à chaque type, suivent un ordre similaire c'est-à-dire évoluent dans le même sens. Il est donné par la relation :

$$CCR = 1 - \frac{6 \cdot \sum_{j=1}^J \sum_{k=1}^{K(j)} d_{jk}^2}{NbJK^3 - NbJK} \quad (2.2)$$

avec d_{jk} la différence de classement de l'image (ou vidéo) j ayant subi la dégradation k entre $Mobj$ et MOS . Une valeur de CCR proche de 1 montre que le classement des images (ou vidéos) par le critère objectif de qualité est similaire à celui fait par les observateurs humains.

2.3 Erreur de prédiction de la qualité

2.3.1 Écart-type de l'erreur de prédiction de la qualité

C'est la mesure de la différence entre MOS et $Mobj$ pour une même image (ou vidéo). L'écart-type est donné par :

$$\sigma(\epsilon_{MOS}) = \sqrt{\frac{\sum_{j=1}^J \sum_{k=1}^{K(j)} (MOS_{jk} - Mobj_{jk})^2}{NbJK - 1}} \quad (2.3)$$

2.3.2 Écart-type de l'erreur de prédiction de la qualité pondérée par l'intervalle de confiance à 95%

L'intervalle de confiance à 95% est l'intervalle $[MOS_{jk} - e_{jk}, MOS_{jk} + e_{jk}]$ dans lequel se trouvent 95% des réponses des observateurs humains avec $e_{jk} = 1.96 * \sigma_{jk}$ en supposant que MOS_{jk} suit une loi Gaussienne. L'écart-type σ_{jk} des notes fournies par les observateurs est donné par la formule :

$$\sigma_{jk} = \sqrt{\frac{1}{N_{obs} - 1} \sum_{i=1}^{N_{obs}} (Nsubj_{jk}(i) - MOS_{jk})^2} \quad (2.4)$$

avec N_{obs} le nombre d'observateurs, $Nsubj_{jk}$ la moyenne des notes fournies par un seul observateur pour l'image (ou vidéo) originale j ayant subi la dégradation k et MOS_{jk} la moyenne des notes fournies par tous les observateurs pour l'image (ou vidéo) j ayant subi la dégradation k . L'intervalle de confiance à 95% permet d'augmenter la fiabilité des résultats des tests subjectifs.

L'écart-type de l'erreur de prédiction de la qualité pondérée par l'intervalle de confiance à 95% est alors donné par l'équation :

$$\sigma(\epsilon_{MOS}^p) = \sqrt{\frac{1}{NbJK - 1} \sum_{j=1}^J \sum_{k=1}^{K(j)} \left(\frac{MOS_{jk} - Mobj_{jk}}{IC_{jk} + 0.025} \right)^2} \quad (2.5)$$

avec IC_{jk} l'intervalle de confiance associé au MOS_{jk} . Cet indicateur prend en compte la relation entre l'écart-type et l'intervalle de confiance associé au MOS pour augmenter encore plus la précision de la mesure de la performance du critère objectif de qualité.

2.4 Indicateur de mauvaise prédiction

L'indicateur de mauvaise prédiction, appelé "*Outliers Ratio*" mesure le taux de "fausses notes" données par le critère objectif de qualité. On entend ici par "fausses notes" toutes les notes situées en dehors de l'intervalle $[MOS_{jk} - 2 * \sigma_{jk}, MOS_{jk} + 2 * \sigma_{jk}]$. Il est mesuré par :

$$OR = \frac{\text{Nombre de fausses notes}}{\text{Nombre total de notes}} \quad (2.6)$$

Le critère objectif de qualité est performant lorsque ce taux est faible.

2.5 Coefficient de Kappa

Le coefficient de Kappa indique si les images (ou vidéos) sont notées de la même manière par les observateurs humains et le critère de qualité. Pour illustrer ce coefficient, nous donnons dans la figure 2.5 un exemple de tableau regroupant les probabilités d'obtention des notes subjectives et objectives (l'exemple est donné pour une échelle d'évaluation de qualité de cinq catégories). Le coefficient de Kappa est calculé à partir de P_{bon} et P_{err} les probabilités de l'obtention de "bonnes" notes² et de concordance aléatoire, respectivement. Les deux probabilités sont données par :

$$P_{bon} = \sum_{k=1}^n P_{kk} \quad (2.7)$$

$$P_{err} = \sum_{k=1}^n P_{S_k} P_{O_k} \quad (2.8)$$

où

$$P_{S_k} = \sum_{i=1}^n P_{ik} \quad \text{pour } k = 1, 2 \dots n \quad (2.9)$$

et

$$P_{O_k} = \sum_{j=1}^n P_{kj} \quad \text{pour } k = 1, 2 \dots n \quad (2.10)$$

représentent les probabilités de l'obtention de notes subjectives et objectives dans la catégorie k , respectivement avec :

- n le nombre de catégories de l'échelle d'évaluation de qualité

²Si le critère objectif de qualité et les observateurs humains donnent la même note à une image (ou vidéo), cette note est alors considérée "bonne".

OS	1	2	3	4	5
1	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}
2	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}
3	P_{31}	P_{32}	P_{33}	P_{34}	P_{35}
4	P_{41}	P_{42}	P_{43}	P_{44}	P_{45}
5	P_{51}	P_{52}	P_{53}	P_{54}	P_{55}

FIG. 2.1 – Le tableau de probabilités d’obtention des notes subjectives et objectives.

- P_{kk} la probabilité d’obtenir d’une “bonne” note dans la catégorie k
- P_{ij} la probabilité que l’image (ou vidéo) soit classée dans la catégorie i par le critère objectif et dans la catégorie j par les observateurs humains.

Le coefficient de Kappa est donné alors par :

$$C_{Kappa} = \frac{P_{bon} - P_{err}}{1 - P_{err}} \quad (2.11)$$

Sa valeur varie entre 0 et 1. Une valeur égale à 1 indique une concordance parfaite entre les notes subjectives et objectives (dans ce cas $P_{bon} = 1$) et une valeur égale à zéro implique que les “bonnes” notes sont obtenues par pur hasard. En général, un coefficient de Kappa supérieur à 0.7 indique une bonne corrélation entre les notes subjectives et objectives.

2.6 Coefficient de Kurtosis

Le coefficient de Kurtosis mesure l’aplatissement d’une distribution c’est-à-dire à quel point elle s’approche d’une distribution normale. Il est donné par :

$$C_{Kurtosis} = \frac{\frac{1}{NbJK} \sum_{j=1}^J \sum_{k=1}^{K(J)} (MOS_{jk} - Mobj_{jk})^4}{\left(\frac{1}{NbJK} \sum_{j=1}^J \sum_{k=1}^{K(J)} (MOS_{jk} - Mobj_{jk})^2\right)^2} \quad (2.12)$$

Un coefficient de Kurtosis très proche de zéro correspond à une distribution normale alors qu’un coefficient négatif (respectivement positif) correspond à une distribution aplatie (respectivement pointue) par rapport à une distribution normale.

Chapitre 3

Critères de qualité d'images fixes

Certains critères objectifs de qualité sont basés sur le SVH. Il paraît donc nécessaire de commencer ce chapitre par une description du fonctionnement du SVH et des principaux éléments le constituant. De plus, les phénomènes liés à ses propriétés caractéristiques sont expliqués car ce sont ces aspects qui sont intégrés dans l'élaboration d'un critère objectif de qualité basé sur une modélisation du SVH.

3.1 Fonctionnement du SVH

Le SVH est l'un des mécanismes fonctionnels les plus complexes de notre corps humain ; en effet, 50% de notre cerveau est dédié à l'analyse de l'information visuelle. Le SVH se compose de trois grandes parties : l'œil et les nerfs optiques, le cortex visuel primaire, le cortex visuel secondaire. Un schéma montrant ce modèle est donné à la figure 3.1.

Une explication détaillée de ces parties est fournie dans [10]. Nous nous contentons d'une explication générale du fonctionnement de chaque partie.

3.1.1 L'œil humain

L'œil humain est l'organe de base du SVH : il capte l'information visuelle à traiter à l'aide de ses propres éléments qui le constituent et la transmet au cerveau. Ses éléments constitutifs sont :

- l'iris, muscle entourant la pupille et contrôlant son ouverture et par suite la quantité de lumière pénétrant dans l'œil
- la cornée, membrane transparente protégeant le globe oculaire

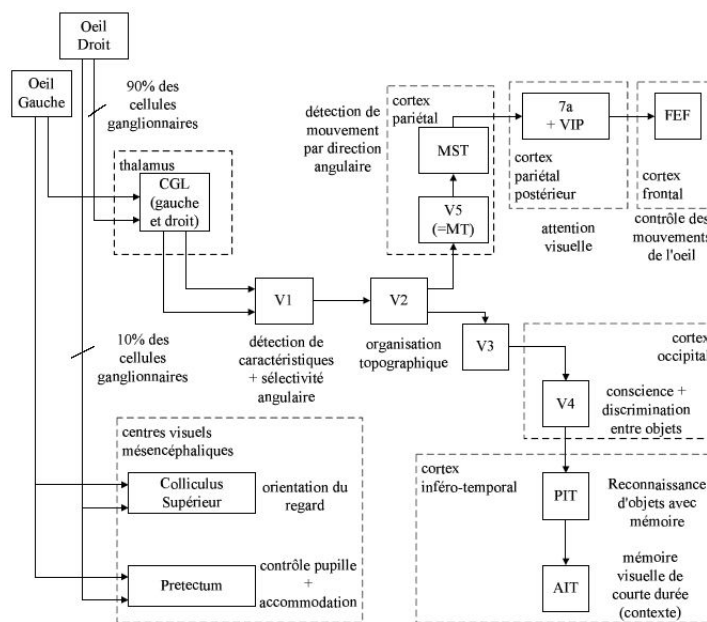


FIG. 3.1 – Le modèle du SVH extrait de [10].

- le cristallin, lentille convergente dont la courbure varie pour accomoder
- la rétine, formée de plusieurs couches de neurones et sur laquelle se forment les images provenant de l'extérieur
- le nerf optique, responsable de l'acheminement de l'information vers le cerveau

La figure 3.2 montre une coupe de l'œil humain.

3.1.2 Le cortex visuel primaire

Le cortex visuel primaire est essentiellement formé de l'aire V1, appelée aussi "aire visuelle primaire"; c'est la partie du SVH qui reçoit les informations de l'œil à travers le nerf optique et en sélectionne les traits caractéristiques liés aux couleurs, à la forme et au mouvement.

3.1.3 Le cortex visuel secondaire

Le cortex visuel secondaire comprend les aires V2, V3, V4 et V5. Ces aires sont réparties sur deux voies : une voie ventrale regroupant les aires V2, V3 et V4, responsable de la reconnaissance des objets et des formes et

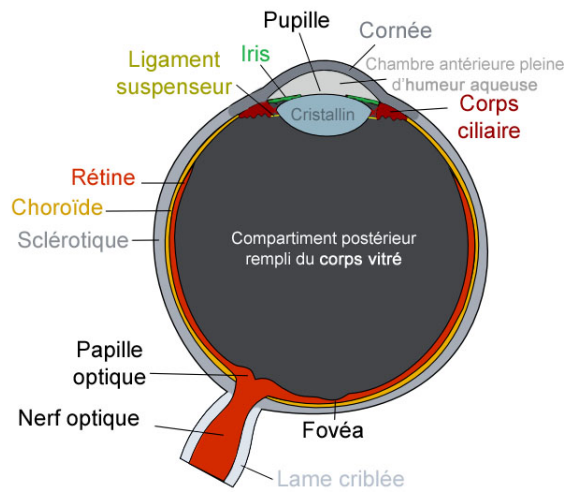


FIG. 3.2 – Une coupe de l’œil humain extraite de [11].

une voie dorsale regroupant les aires V2 et V5, responsable du guidage des mouvements. Au niveau de l’aire V2, l’information reçue en provenance de l’aire V1 est projetée selon la voie de traitement correspondante.

3.2 Phénomènes liés au fonctionnement du SVH

Du fonctionnement assez complexe du SVH, naissent plusieurs phénomènes qui affectent le processus de perception d’images (ou vidéos). Le plus important est l’effet de masquage, phénomène étroitement lié à la notion de contraste. Nous donnons dans ce qui suit plusieurs définitions du contraste présentes dans la littérature puis nous définissons la fonction de sensibilité au contraste “*Contrast Sensitivity Function*” (CSF). Ces deux étapes sont nécessaires pour une bonne compréhension de l’effet de masquage.

3.2.1 Le contraste

Le contraste est défini comme étant la variation relative de luminance entre une région donnée et le fond où se trouve cette région. Dans la littérature, les auteurs ont interprété cette définition de plusieurs façons ; nous en retenirons les deux les plus utilisées pour le moment car une explication détaillée de la notion de contraste sera donnée ultérieurement :

Contraste de Weber

Weber a défini le contraste comme étant le rapport de la variation de luminance ΔL d'une région donnée par rapport à la luminance du fond divisée par la valeur de luminance L_0 du fond :

$$C_{Weber} = \frac{\Delta L}{L_0} \quad (3.1)$$

Cette définition est surtout utilisée pour mesurer le contraste d'un stimulus simple sur un arrière-plan uniforme.

Contraste de Michelson

Michelson définit le contraste comme étant le rapport de la différence entre les luminances maximales L_{max} et minimales L_{min} de la région considérée divisée par la somme des luminances maximales et minimales :

$$C_{Michelson} = \frac{L_{max} - L_{min}}{L_{min} + L_{max}} \quad (3.2)$$

La définition de Michelson est utilisée pour caractériser les images à variation spatiale de luminance de type sinusoïdal.

3.2.2 Fonction de sensibilité au contraste (CSF)

Les cellules du cortex visuel sont sensibles au contraste. Donc pour modéliser le phénomène de perception, c'est-à-dire la sensibilité du SVH au contraste, la fonction de sensibilité au contraste CSF est utilisée. Cette fonction représente la sensibilité de l'œil humain aux différentes fréquences spatiales de l'image : généralement, le SVH est plus sensible aux basses fréquences spatiales qu'aux hautes fréquences (détails fins). La fréquence spatiale est mesurée en cycles par degré d'angle visuel c'est-à-dire à quelle vitesse le stimulus change en fonction de l'espace.

L'équation de la variation de sensibilité de perception du contraste en fonction de sa fréquence spatiale pour Mannos et Sakrison [12] est donnée par la relation 3.3 ; la figure 3.3 montre son allure.

$$A(f) = 2.6 * (0.0192 + 0.114 * f) * \exp(-0.114 * f)^{1.1} \quad (3.3)$$

Cette fonction montre que la sensibilité de l'œil humain est maximale pour une fréquence spatiale de 8 cycles/deg ; aux alentours de cette valeur, la sensibilité reste grande mais diminue au fur et à mesure qu'on s'éloigne de l'ordre de $f = 8$ cycles/deg : elle devient négligeable à partir de $f = 50$ cycles/deg (détails très fins de l'image).

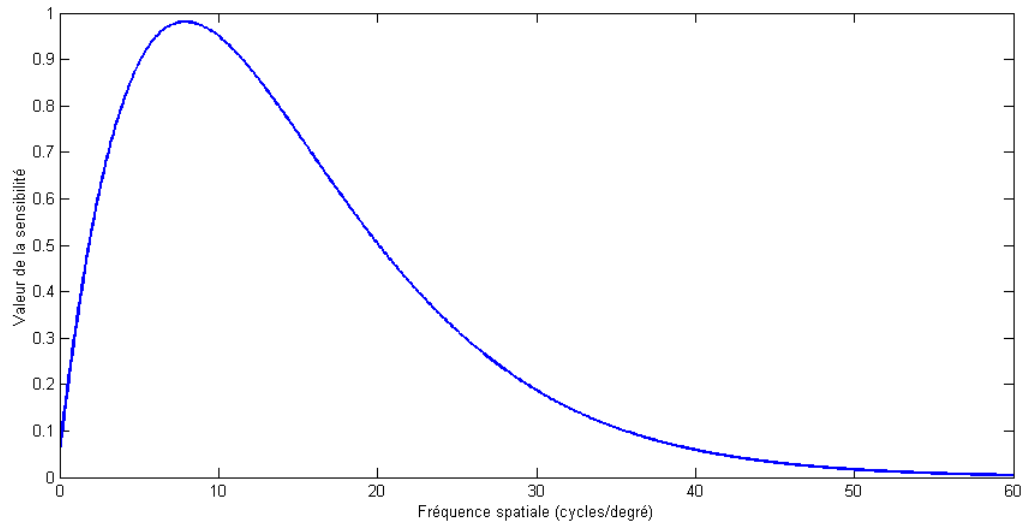


FIG. 3.3 – L’allure de la CSF de Mannos et Sakrison.

3.2.3 Effet de masquage

L’effet de masquage se produit lorsqu’un signal, appelé signal masquant fait varier la visibilité d’un autre signal, appelé signal masqué. Cette variation peut être décroissante (on parle dans ce cas de phénomène de masquage) ou croissante (phénomène de facilitation). Généralement, ce phénomène implique des stimuli de même fréquence mais peut, dans des cas particuliers, impliquer des stimuli de fréquences différentes. Cet effet de masquage est important dans la mesure où un défaut visuel peut être caché ou mis en relief par un autre ; il doit être pris en compte dans l’élaboration d’un critère de qualité basé sur une modélisation du SVH.

3.3 Critères de qualité d’images avec référence complète

Les critères de qualité avec référence complète sont des critères qui évaluent la qualité d’une image dégradée en utilisant toute l’image originale comme référence. Ces critères sont surtout utilisés dans les systèmes introduisant des dégradations, par exemple les systèmes de compression avec pertes, pour évaluer la quantité de distorsions introduites par la compression et la qualité de l’image en résultant. Ces critères sont majoritairement basés sur une

modélisation du SVH et donc intègrent une ou plusieurs propriétés de ce modèle bien que d'autres, basés sur les outils "classiques" de traitement du signal, sont encore utilisés. Nous explicitons dans ce qui suit plusieurs critères d'évaluation de qualité d'images présents dans la littérature.

3.3.1 Critères de qualité non basés sur le SVH

Ce type de critères traite l'image comme un pur signal numérique : il en extrait les informations nécessaires à l'obtention de la note de qualité sans se baser sur aucune information liée au fonctionnement du SVH.

Critères basés sur le "signal"

PSNR

Le PSNR (*Peak Signal-to-Noise Ratio*) est une métrique utilisée pour évaluer la fidélité d'une image (par rapport à celle de référence). C'est un critère qui traite l'image purement comme un signal numérique. Il se base sur l'erreur quadratique moyenne MSE (*Mean Squared Error*) liée à la différence entre l'image dégradée et celle de référence :

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (3.4)$$

avec N le nombre de pixels dans l'image ou la vidéo, x_i et y_i les i èmes pixels de l'image originale et dégradée, respectivement.

Nous en déduisons le rapport signal à bruit crête :

$$PSNR = 10 \log_{10} \frac{L^2}{MSE} \quad (3.5)$$

avec L la dynamique des valeurs des pixels ($L = 255$ pour un codage des pixels sur 8 bits).

Une valeur de PSNR supérieure à 34 dB indique une assez bonne qualité d'image. MSE et PSNR sont largement utilisés car leur implémentation est facile et leur complexité est faible. Mais plusieurs tests ont démontré qu'ils ne sont pas corrélés suffisamment avec la perception humaine, par exemple dans [10], où l'image visuellement plus dégradée a un PSNR plus élevé et une MSE plus petite que l'image dont la qualité visuelle est meilleure.

Critères de qualité basés sur les dégradations

Ces critères de qualité mesurent l'importance des dégradations introduites depuis le codage de la vidéo jusqu'à son affichage. La combinaison de ces mesures résulte en une note de qualité de l'image.

Critère de Wang et Bovik [13]

Le critère de Wang et Bovik appelé UIQI (*Universal Image Quality Index*) fait partie des critères généralistes de qualité. En effet, l'UIQI est indépendant des conditions d'observation des images et de leur contenu. De plus, son implémentation est facile et il peut être appliqué à différents types d'images (d'où l'appellation "*universal*").

L'UIQI mesure la décorrélation entre l'image originale et l'image dégradée et la dégradation de la composante de luminance et du contraste entre les deux versions de l'image. Le produit de ces trois mesures donne la note finale de qualité. Les mesures sont faites de la manière suivante : une "fenêtre glissante" est appliquée au coin supérieur gauche de l'image puis cette fenêtre est déplacée d'un pixel horizontal et vertical jusqu'à atteindre le coin inférieur droit de l'image. La note de qualité du contenu de la fenêtre est calculée à chaque étape et la moyenne de toutes ces notes donne la note de qualité finale de toute l'image.

Les performances de l'UIQI sont évaluées dans [13] par une comparaison des notes fournies par le MSE et celles fournies par l'UIQI. De plus, les tests montrent que les images sont classées par les observateurs dans le même ordre que l'UIQI les a classés. Mais ce critère généraliste n'est pas totalement adapté à la mesure de qualité d'images pour deux raisons principales : premièrement, le contraste et la luminance sont mesurés comme de simples données mathématiques en négligeant leur aspect perceptuel. Ensuite, la note de qualité finale est obtenue en calculant simplement la moyenne des notes de qualité de chaque région de l'image : cette approche est simpliste car elle ne prend pas en compte aucune information structurelle ou spatiale.

Les critères exposés dans cette partie sont loin d'être les meilleurs en termes d'évaluation de la qualité d'une vidéo car le processus s'étendant de la vision au jugement personnel ne peut pas être modélisé par un modèle mathématique simple. Des critères intégrant les propriétés du SVH ont été élaborés pour essayer de modéliser le plus fidèlement le fonctionnement du SVH.

3.3.2 Critères de qualité basés sur le SVH

Critère de Daly *et al.*

Dans [14], Daly *et al.* proposent un critère objectif de qualité d'images intégrant quelques propriétés du SVH. En réalité, ce critère est une extension du “*Visual Difference Predictor*” de Daly (proposé en 1993) pour les images à grande échelle dynamique (*High Dynamic Range : HDR*). Les images HDR se distinguent des images numériques classiques par leur reproduction précise de la quasi-totalité des couleurs présentes dans une image de scène naturelle (de l'ordre des millions). Le principe du critère de Daly *et al.* (appelé HDR VDP) est présenté ci-dessous.

Le HDR VDP modélise tout d'abord l'œil humain et la rétine à l'aide de l'OTF (*Optical Transfer Function*), d'une fonction non linéaire de l'amplitude des réponses du SVH aux stimuli et la CSF. Le filtrage de l'image par l'OTF représente la diffusion de la lumière dans la cornée, le cristallin et la rétine. La compression non linéaire de l'amplitude des réponses du SVH simule la réponse non linéaire des photo-récepteurs à leur stimulation par la lumière. Finalement, l'image est filtrée par la CSF.

Ensuite, le HDR VDP modélise le cortex visuel et l'effet de masquage. Durant cette phase, les deux images sont décomposées en canaux spatio-directionnels et comparées.

Enfin, une série de traitements subis par les deux images résulte en une “carte de probabilités d'erreur” qui génère une note de qualité. Ces traitements sont : la décorrélation entre l'effet de masquage et la phase du signal, l'application d'une fonction psychométrique et la combinaison pondérée des probabilités d'erreur à travers tous les canaux.

Dans [14], des tests d'évaluation de la qualité visuelle sont menés pour calibrer le HDR VDP et non pas pour évaluer ses performances : nous ne disposons donc pas de matière suffisante pour juger si les notes fournies par le HDR VDP corrélaient bien avec les notes subjectives de qualité.

3.4 Critères de qualité avec référence réduite

Les critères objectifs de qualité avec référence réduite diffèrent des critères avec référence complète par le fait que l'image originale n'est pas disponible et/ou utilisée. La référence est alors remplacée par une représentation formée de descripteurs pertinents qui peuvent être comparés dans les deux versions de l'image. Ces critères sont surtout utilisés dans un contexte de transmission

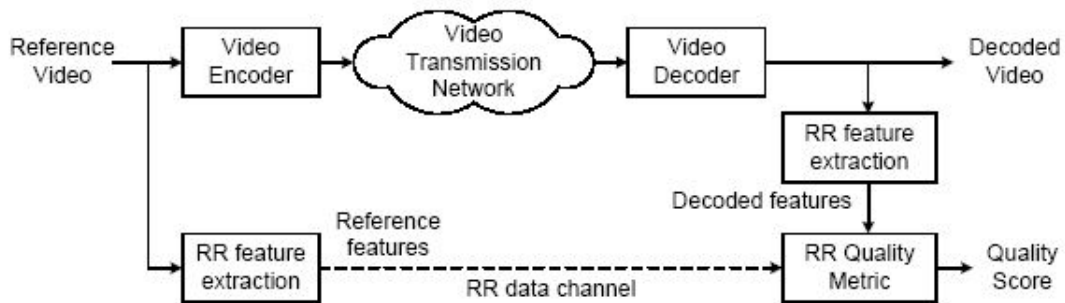


FIG. 3.4 – Le schéma de fonctionnement des critères RR (extrait de [15]).

où il est impossible de transmettre toute l'information liée aux deux versions. La référence réduite est codée et transmise avec la version dont la qualité est à évaluer.

Dans la littérature, rares sont les critères d'évaluation de la qualité d'une image avec référence réduite. Ceci est dû au choix difficile de descripteurs qui soient une fidèle représentation de la vidéo et à la complexité de leur extraction.

Le modèle général de l'évaluation de la qualité d'une vidéo avec référence réduite est donné dans la figure 3.4, tirée de [15].

3.4.1 Critère de Carnec

Dans [10], Mathieu Carnec (membre de l'équipe IVC du laboratoire IRC-CyN) présente un critère de qualité avec référence réduite perceptuelle. Ce critère, fortement basé sur une modélisation du SVH, extrait l'information structurelle de l'image dégradée et de l'image de référence et produit une note de qualité. Il procède comme suit : tout d'abord, l'espace de couleurs (RVB) de l'image est projeté dans un espace perceptuel (ACr1Cr2) après l'application d'une fonction gamma.

Ensuite, l'image est décomposée en sous-bandes perceptuelles desquelles sont extraits les traits caractéristiques structurels composant la référence réduite de l'image. L'extraction des traits caractéristiques de l'image est basée sur plusieurs propriétés du SVH dont la localisation des points de fixation des yeux sur l'image et leur phénomène d'adaptation à la lumière. Enfin, les références réduites des deux images sont comparées à l'aide de plusieurs mesures de similarité (mesures de correspondance, mesures de similarité locale et globale) .

L'application du critère objectif de Carnec sur une large gamme d'images (bases d'images notées de l'IVC et du laboratoire LIVE-Texas) a montré une bonne corrélation entre les notes du critère et celles d'observateurs humains.

3.4.2 Critère de Gunawan et Ghanbari

Dans [16], Gunawan et Ghanbari présentent un critère objectif de qualité d'images et de vidéos avec référence réduite. Ce critère analyse les effets de blocs et le flou dans l'image (ou vidéo) de référence et dans l'image (ou vidéo) dégradée et attribue à cette dernière une note de qualité. Cette analyse, basée sur l'amplitude des harmoniques fréquentielles, permet la détection et la localisation de ces deux types de dégradations. En effet, les effets de blocs produisent un signal pseudo-périodique qui génère des harmoniques dans le domaine fréquentiel : l'importance de ces harmoniques est proportionnelle à celle des effets de blocs.

La référence réduite de ce critère est formée des harmoniques. Pour obtenir cette référence réduite, le gradient de l'image est calculé par le filtrage de l'image par un filtre de Sobel 3x3. Ensuite, ce gradient est décomposé en blocs de 32x32 pixels. Une transformée rapide de Fourier (*Fast Fourier Transform : FFT*) appliquée aux deux images (ou vidéos) permet l'extraction des harmoniques locales. Cette référence réduite peut être interprétée comme une forme d'activité spatiale constituée des contours verticaux et horizontaux de l'image. L'indicateur de qualité est une "carte" (matrice) représentant la différence entre l'intensité et la position des harmoniques locales dans chacune des deux versions de l'image (ou vidéo). La différence peut être positive c'est-à-dire qu'il y a eu une augmentation de l'activité spatiale (dûe notamment à l'apparition des effets de blocs) ou négative c'est-à-dire qu'il y a eu une diminution de l'activité spatiale (dûe à la perte de détails causée par le flou). Finalement, la note de qualité est obtenue en moyennant les gains et les pertes calculés séparément.

Les tests effectués par Gunawan et Ghanbari ont montré une bonne corrélation entre les notes de leur critère et les notes subjectives. Le critère s'est montré efficace en détectant les effets de blocs et le flou dans la partie uniforme de l'image et en les négligeant dans la partie fortement texturée. En effet, la présence de texture dissimule ce type de dégradations et les rend moins gênantes.

Mais l'application de ce critère sur un nombre limité d'images n'est pas suffisante pour évaluer ses performances.

3.5 Critères de qualité sans référence

Les critères objectifs de qualité de vidéos sans référence évaluent la qualité d'une vidéo sans se référer à la vidéo originale (de référence). Ce type de critères est utilisé surtout pour l'évaluation de la qualité de vidéos dégradées par un système connu a priori. Ces critères sont difficiles à élaborer car ils doivent intégrer, en plus des propriétés du SVH, des données statistiques sur le système de compression (introduisant les dégradations). Un contexte de transmission comme la diffusion télévision par exemple est le cadre idéal d'utilisation de ce type de critères car ils ne nécessitent pas la transmission de la version originale pour évaluer la qualité de la vidéo transmise. Vu la rareté de ces critères dans la littérature, nous nous contentons de présenter un seul critère de qualité d'images fixes sans référence.

3.5.1 Critère de Luo

Le critère de Luo [17] se base sur l'information spatiale de l'image pour évaluer sa qualité. Cette information est obtenue par la décomposition et la reconstruction des fréquences spatiales dans des bandes spectrales à l'aide de la transformée ondelette. Au niveau de la décomposition, plusieurs analyses sont effectuées pour extraire les traits caractéristiques dont la combinaison forme la note de qualité finale. Ces traits caractéristiques sont : la quantité de bruit présent dans l'image, la netteté des contours, la luminosité et le contraste de l'image. Une combinaison linéaire de ces traits, à l'aide de coefficients obtenus lors de la phase d'apprentissage du critère donne la note finale de qualité.

Les résultats publiés dans [17] montrent une bonne corrélation avec les résultats obtenus par des observateurs humains.

Chapitre 4

Critères de qualité de vidéos

Les critères de qualité pour images fixes, présentés dans le chapitre précédent, sont la base des critères de qualité pour vidéos. Mais ces derniers, peu nombreux dans la littérature, prennent en considération quelques facteurs en plus de ceux pris en compte dans l'élaboration des critères de qualité pour images fixes. Nous présentons dans ce chapitre les principaux éléments permettant le passage des critères de qualité pour images fixes à des critères pour vidéos et nous détaillons des critères de qualité pour vidéos appartenant aux trois types FR, RR et NR. Enfin nous expliquons le processus de l'obtention d'une note de qualité à partir des données mesurées par le critère objectif.

4.1 Différences entre critères de qualité d'images fixes et de vidéos

La méthodologie générale des critères objectifs de qualité est la même pour les images fixes et les vidéos. Mais une vidéo est différente d'une image fixe : comme on dit couramment, *“une image vaut des milliers de mots et une vidéo vaut des milliers d'images”*. Par conséquent, les critères de qualité de vidéos doivent prendre en considération la différence entre les images fixes et les vidéos ; ceci peut être fait par la modélisation des deux aspects suivants : le mouvement du contenu des scènes (ou même de la prise de vue) et la dimension temporelle de la vidéo. En effet, le mouvement des objets d'une scène influe énormément la qualité visuelle de la vidéo donc l'intégration d'une mesure de ce mouvement est indispensable à l'élaboration d'un critère de qualité performant. Ce mouvement peut causer une dégradation de la qualité visuelle de la vidéo (accentuation des effets de blocs, du flou...) ou même améliorer cette qualité (augmentation de la netteté par exemple [18]).

D'autre part, la succession des trames de la vidéo introduit une dimension temporelle qui doit être prise en compte lors de l'élaboration du critère de qualité pour vidéos. Il est donc important d'ajouter un cumul temporel au cumul spatial des traits caractéristiques. Ainsi, l'estimation de la qualité d'une vidéo est moyennée sur tout le temps de son affichage.

4.2 Critères de qualité avec référence complète

De même que pour les images fixes, les critères de qualité de vidéos avec référence complète utilisent la vidéo originale pour attribuer une note de qualité à la vidéo sous test. Ils sont aussi classés de la même façon.

4.2.1 Critères de qualité non basés sur le SVH

Critères basés sur le “signal”

PSNR

Le critère objectif de qualité d'images fixes PSNR est utilisé aussi dans le cas de la vidéo. Il est appliqué pour chaque image de la vidéo : la moyenne de ses valeurs sur toute la vidéo représente la note finale de qualité de la vidéo. Cette approche est très simpliste car elle ne prend en compte aucun aspect de la vidéo et la traite en tant qu'image fixe.

Critères de qualité basés sur les dégradations

AVQM (*Artifacts-based Video Quality Metric*)

Le critère AVQM [19] se base sur des travaux antérieurs de mesure de qualité de vidéos. Il mesure la quantité de dégradations présente dans la vidéo et utilise la logique floue pour attribuer une note de qualité à la vidéo. Son principe est le suivant : tout d'abord, la vidéo est segmentée en blocs spatio-temporels de 8 x 8 pixels par 3 secondes. Ensuite, une TCD-2D (Transformée en Cosinus Discrète suivant 2 dimensions) leur est appliquée. Puis les coefficients des composantes continues et non continues sont calculés pour chaque bloc et comparés à des valeurs prédéterminées (au cours de la phase d'apprentissage du critère) pour obtenir le taux T de blocs qui “flashent” (*flashing blocks*) défini par :

$$T = \frac{\text{nombre de blocs qui flashent}}{\text{nombre total de blocs}} \quad (4.1)$$

Les blocs qui flashent sont les blocs qui apparaissent comme le “flash” d’une caméra photographique lorsque la vidéo est affichée. La note de qualité est obtenue par la combinaison de plusieurs variables (variables spatiales issues de la comparaison et T) en utilisant le schéma flou de Takagi-Sugeno. Ce critère montre une bonne corrélation avec les résultats obtenus lors de tests subjectifs d’évaluation de la qualité de vidéos.

4.2.2 Critères de qualité basés sur le SVH

PVQM (*Perceptual Video Quality Measure*)

Le critère objectif d’évaluation de la qualité d’une vidéo PVQM [20] se base, comme son nom l’indique, sur une représentation perceptuelle de la vidéo pour évaluer sa qualité. Selon [20], un critère de qualité basé sur la perception doit modéliser quelques aspects principaux dont le filtrage spatio-temporel, le phénomène d’adaptation de l’œil, la saturation des couleurs, l’effet de masquage, le mouvement des yeux, le contraste et la luminance. En effet, PVQM utilise une combinaison linéaire de trois indicateurs pour obtenir la note finale de qualité. Ces trois indicateurs sont : la netteté des contours, la décorrélation temporelle et l’erreur normalisée des couleurs. La note de qualité finale $DMOS_p$ (p pour prédite) est donnée par :

$$DMOS_p = 3.95E' + 0.74N[c] - 0.78D - 0.4 \quad (4.2)$$

avec E' l’indicateur de netteté des contours (cumulée dans l’espace), $N[c]$ l’indicateur d’erreur sur la couleur c (cumulée dans le temps) et D l’indicateur de décorrélation temporelle.

4.3 Critères de qualité avec référence réduite

Les critères de qualité de vidéos avec référence réduite ont le même principe que celui des critères de qualité (avec référence réduite) d’images fixes tout en intégrant les propriétés additionnelles de la vidéo. Ces critères sont rares dans la littérature. Nous en exposons deux dans ce qui suit.

4.3.1 Critère de Webster *et al.*

Dans [21], Webster *et al.* proposent une évaluation de la qualité d’une vidéo à partir de trois mesures effectuées sur la vidéo. Ces mesures, déterminées par une analyse statistique, sont : la mesure de la distorsion spatiale m_1 et les mesures de distorsions temporelles m_2 et m_3 . Les distorsions spatiales et

temporelles sont basées sur les informations spatiales SI et temporelles TI respectivement.

La note de qualité finale \hat{s} est donnée par :

$$\hat{s} = 4.77 - 0.992m_1 - 0.272m_2 - 0.356m_3 \quad (4.3)$$

La comparaison entre les résultats obtenus par ce critère et ceux obtenus lors de tests subjectifs d'évaluation de qualité donne un coefficient de corrélation satisfaisant (0.94). Mais le principal manque dans ce critère est l'omission totale des deux composantes de chrominance ; en effet, le critère a été appliqué à la seule composante de luminance du signal vidéo lors des tests d'évaluation.

4.3.2 Critère de Wolf et Pinson

Dans [22], Wolf et Pinson apportent des extensions au critère élaboré précédemment dans [21] pour améliorer ses performances. Par exemple, un filtrage augmentant la netteté des contours est effectué suivi de l'extraction des descripteurs. Le premier descripteur mesure la netteté des contours dans les deux directions horizontales et verticales et le second la mesure dans les autres directions. La combinaison linéaire des deux descripteurs donne la note de qualité finale.

De même que le critère de base, ce critère est performant mais ne prend toujours pas en considération les composantes de chrominance dans le calcul de la qualité de la vidéo. De plus, les tests ayant été faits à bas débit, les résultats sont susceptibles de décroître pour des transmissions à bande passante limitée.

4.4 Critères de qualité sans référence

Les critères de qualité de vidéos sans référence n'utilisent aucune information de la vidéo originale pour attribuer une note de qualité à la vidéo testée. En fait, ils sont similaires aux critères de qualité d'images sans référence mais encore moins nombreux.

4.4.1 Critère de Le Callet *et al.*

Patrick Le Callet et d'autres membres de l'équipe IVC du laboratoire IRCCyN proposent dans [23] un critère objectif de qualité de vidéos qui peut être utilisé dans deux contextes : avec référence réduite et sans référence. Ce critère suit le principe suivant : la vidéo dégradée subit tout d'abord une

transformation des composantes de luminance et chrominance dans l'espace perceptuel ACr1Cr2 pour modéliser la perception du SVH. Ensuite, des traits caractéristiques sont calculés pour chaque composante dans toutes les trames de la vidéo. Enfin, la combinaison des traits caractéristiques et le cumul temporel sont effectués à l'aide de réseaux de neurones pour aboutir à une note de qualité de la vidéo.

Les traits caractéristiques extraits pour chaque composante sont les suivants :

- la mesure de la netteté des contours
- la mesure du flou
- la mesure de la décorrélacion temporelle entre les trames
- la mesure des effets de blocs

Les trois premiers traits caractéristiques dépendent du contenu de la vidéo alors que le quatrième est générique. Les réseaux de neurones utilisés ici sont de type convolutionnel intégrant la notion de corrélation temporelle entre les trames de la vidéo. Les performances de ce critère, mesurées sur différentes vidéos aux contenus variés, montrent une bonne corrélation entre les notes fournies par le critère et les notes subjectives.

4.4.2 Critère de Sugimoto *et al.*

Le critère de Sugimoto *et al.* [24] (cité dans [15]) se base sur une transmission de bits, dits de “marquage” entre les trames de la vidéo. Ces bits de “marquage” sont des séquences aléatoires de bits qui serviront à l'évaluation de la qualité de la vidéo transmise; en effet, le taux d'erreur dans les bits reçus indique la perte de qualité de la vidéo transmise.

En réalité, ce critère se place à un niveau intermédiaire entre les critères avec référence réduite et ceux sans référence car il y a transmission d'information avec la vidéo à évaluer mais cette information n'est pas relative à la vidéo de référence (donc cette dernière n'est pas utilisée).

4.5 L'obtention d'une note de qualité

Comme nous l'avons déjà expliqué, la note de qualité d'une image (ou vidéo) est la combinaison de plusieurs mesures effectuées sur cette image (ou vidéo). Cette combinaison comprend un cumul spatial sur toute l'image et, dans le cas de la vidéo, un cumul temporel sur toute la durée de son affichage. Mais cette combinaison peut se faire de plusieurs façons, notamment en utilisant la formule de Minkowski ou même les réseaux de neurones. Nous donnons dans ce qui suit les principales méthodes de cumul utilisées dans les

critères objectifs de qualité de la littérature.

4.5.1 Méthodes de cumul

La majorité des méthodes de cumul des mesures de traits caractéristiques se base sur la somme pondérée de Minkowski qui est de la forme :

$$N_{obj} = \left(\sum_{i=1}^n a_i T_i^p \right)^{1/p} \quad (4.4)$$

avec N_{obj} la note de qualité fournie par le critère objectif, n le nombre total de traits caractéristiques mesurés, a_i le coefficient de Minkowski du trait caractéristique T_i) et p l'exposant de Minkowski.

Les coefficients de Minkowski sont calculés lors de la phase d'apprentissage du critère. Pour $p = 1$, la formule devient une simple combinaison linéaire des traits caractéristiques.

Par exemple, le critère PVQM [20] utilise une combinaison linéaire simple de trois traits caractéristiques : la netteté des contours, la décorrélation temporelle et l'erreur normalisée des couleurs pour attribuer une note de qualité à la vidéo. Engeldrum propose dans [25] d'utiliser la somme pondérée de Minkowski, en lui intégrant une mesure de la vivacité des couleurs de l'image, pour obtenir une note de qualité de l'image.

D'autre part, la somme pondérée de Minkowski peut être utilisée pour calculer des variables qui ne sont pas directement liées aux critères de qualité. Par exemple, dans [26], une étude de la gêne¹ introduite par quatre types de dégradations (effets de blocs, flou, bruit et effets "d'oscillations") sur une vidéo utilise la formule de Minkowski pour calculer la gêne totale et la contribution de chaque type de dégradations dans cette gêne.

Dans [27], un travail similaire est effectué pour étudier la gêne de trois types de dégradations (effets de blocs, flou et effets "d'oscillations") sur des vidéos ayant subi une compression MPEG-2.

Une approche nouvelle, présentée dans [19], utilise la logique "floue" pour combiner les différents traits caractéristiques formant la note de qualité de vidéos. Deux attributs principaux ("petit" et "grand") sont utilisés pour classer les valeurs des mesures de traits caractéristiques. Selon leur classification, les traits caractéristiques suivent le schéma flou de Takagi-Sugeno

¹La gêne est la dégradation de la qualité d'une image (ou vidéo) qui est perceptuellement visible et gênante pour un observateur humain

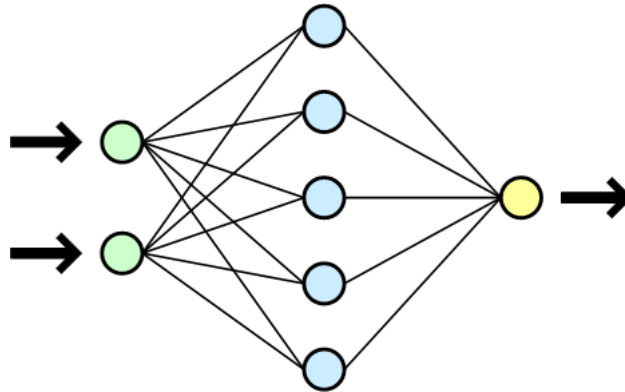


FIG. 4.1 – Un exemple simple de réseaux de neurones (extrait de [11]).

dans lequel ils subissent des opérations arithmétiques : ceci résulte en une note de qualité de la vidéo.

4.5.2 Réseaux de neurones

Une approche complexe d'obtention d'une note de qualité est celle de l'utilisation des réseaux de neurones. Les réseaux de neurones artificiels (*“Artificial Neural Networks”* : *ANNs*, comme leur nom l'indique, simulent le fonctionnement des réseaux de neurones biologiques présents dans notre système nerveux. Ce sont des groupes de neurones interconnectés qui produisent à partir des traits caractéristiques d'une image (ou vidéo), une note de qualité de cette image (ou vidéo). La figure 4.1 illustre un exemple de réseaux de neurones. Le modèle mathématique sur lequel se basent les ANNs définit une fonction de coût que des données acquises lors de tests subjectifs essaient de minimiser : c'est la phase d'apprentissage du réseau de neurones. Le rôle principal des neurones est l'affectation d'un coefficient à chaque trait caractéristique et le calcul de la somme pondérée de ces nouvelles valeurs (les valeurs des traits caractéristiques multipliées par les coefficients correspondants) pour produire la note de qualité finale.

Cette approche n'est pas très utilisée dans la littérature mais nous retrouvons quand même quelques critères basés sur les réseaux de neurones. Par exemple dans [28], Gastaldo *et al.* utilisent les réseaux de type CBP (*“Circular Back Propagation”*) dans l'élaboration de leur critère de qualité d'images améliorées. Les réseaux de neurones CBP sont formés de plusieurs couches de

neurones et sont parfaitement adaptés à la modélisation de processus complexes tels la perception humaine. Les traits caractéristiques sont extraits de l’histogramme de premier ordre, de la matrice de co-occurrence et de la TCD de l’image. Ensuite, les valeurs de ces traits caractéristiques subissent une série de transformations non-linéaires dans les neurones qui aboutit à une note de qualité finale. Une phase d’apprentissage précède la phase de prédiction de la qualité des images : des images dont la note de qualité est connue sont utilisées pour optimiser les coefficients de pondération des neurones. Les résultats publiés dans [28] montrent une bonne corrélation entre les notes fournies par des observateurs humains et celles prédites par le critère de qualité. Cependant, la fiabilité de ces résultats est contestée car les tests subjectifs ont été faits sur internet : dans ce genre de tests, les conditions d’illumination de la salle où se trouve l’observateur ne sont pas contrôlées ce qui remet en question la validité des tests d’évaluation de la qualité visuelle. De plus, les tests ont été effectués pour des images monochromatiques et l’extension de ce critère pour les images couleurs est laissée aux travaux futurs.

Un autre exemple d’utilisation des réseaux de neurones dans un critère objectif de qualité est celui de [23]. En effet, Le Callet *et al.* proposent un critère de qualité de vidéos basé sur les réseaux de neurones de type TDNN (“*Time Delay Neural Network*”) qui modélisent fidèlement le processus de cumul temporel des dégradations effectué par le SVH durant la durée d’affichage de la vidéo. Les bonnes performances du critère de qualité avec référence réduite montrent l’efficacité de l’approche des réseaux de neurones dans l’élaboration des critères de qualité de vidéos.

Conclusion

Nous avons présenté dans cette seconde partie les principaux critères objectifs de qualité d'images et de vidéos présents dans la littérature. Nous avons également détaillé le fonctionnement du SVH et montré la méthodologie d'évaluation des performances des critères de qualité. De plus, nous avons expliqué le principe de l'obtention d'une note de qualité à partir de traits caractéristiques mesurés sur l'image ou la vidéo. Le principal outil de comparaison entre les performances des différents critères de qualité est le taux de corrélation entre les notes de qualité fournies par des observateurs humains et celles fournies par les critères de qualité.

D'une manière générale, aucun critère de qualité n'a prouvé sa "généricité" c'est-à-dire les bonnes performances d'un critère de qualité sont toujours liées à des conditions particulières (nature des images testées ou système de compression utilisé par exemple). De plus, la majorité des critères de qualité présentés estiment la qualité d'images (ou vidéos) dégradées par rapport à l'image (ou vidéo) originale.

La mesure de la qualité de vidéos améliorées fait donc l'objet de la troisième partie.

Troisième partie

Évaluation objective de qualité de vidéos améliorées par des post-traitements d'affichage

Introduction

L'amélioration de la qualité d'une vidéo avant de l'afficher peut s'effectuer de plusieurs façons et à plusieurs niveaux. Les post-traitements d'affichage représentent une forme d'amélioration de qualité au dernier niveau : l'affichage sur l'écran. En effet, ces post-traitements, comme leur nom l'indique, interviennent après le décodage de la vidéo pour garantir sa bonne qualité. Leur importance majeure réside dans le fait que l'observateur tend à préférer une image "parfaite" même si elle n'est pas réelle. Les post-traitements d'affichage se distinguent des méthodes classiques d'amélioration de qualité d'images (égalisation et normalisation d'histogrammes) par leur précision : ils visent chacun à améliorer un aspect particulier de l'image ou la vidéo. Nous commençons cette partie par une description des nouvelles technologies d'écrans. Ensuite, nous présentons un exemple de post-traitements d'affichage et nous terminons par une explication de la méthodologie de mesure d'amélioration de qualité de vidéos.

Chapitre 5

Les nouvelles technologies d'écrans

Comme nous l'avons déjà mentionné dans l'introduction générale, les deux types d'écrans les plus utilisés dans le cadre de la HDTV sont les écrans LCD et Plasma. A ces deux types d'écrans s'ajoutent les écrans classiques CRT (*"Cathode Ray Tube"*) mais qui sont moins utilisés pour l'affichage de la HDTV. Nous présentons dans la suite les principales caractéristiques de ces trois types d'écrans et nous citons leurs avantages et leurs inconvénients.

5.1 Les écrans Plasma

Les écrans Plasma sont des écrans plats et minces qui contiennent des "cellules" (d'où le nom Plasma) de gaz nobles (Xénon et Néon) situées entre deux plaques en verre reliées à des électrodes ; l'application d'une haute tension électrique permet aux particules de phosphore d'illuminer l'écran. Les principaux avantages des écrans Plasma sont : un angle de vue très grand (160°), une reproduction "parfaite" des couleurs (surtout le noir) et leurs prix raisonnables. Mais des études ont montré qu'une même image laissée longtemps (pause d'une image, bande d'informations...) peut causer la brûlure de l'écran c'est-à-dire la persistance de l'ombre de cette image quand une autre est affichée. Avec l'avancement de la recherche dans ce domaine, des améliorations sont amenées aux écrans Plasma pour réduire cet effet de *"burn in"*.

5.2 Les écrans LCD

D'autre part, les écrans LCD sont aussi minces et plats que les écrans Plasma mais différent de ces derniers par leur constitution ; en effet, les écrans LCD contiennent une source de lumière qui éclaire les pixels, colonnes de molécules de cristal liquide situés entre deux électrodes et deux filtres polarisants, et donc illumine l'écran. Les principaux avantages de l'utilisation d'un écran LCD sont : le support d'une résolution très grande, une espérance de vie élevée (60 000 heures) et une consommation d'énergie diminuée par rapport aux écrans Plasma et aux écrans CRT. Mais les écrans LCD souffrent du "*motion blur*", défaut visuel qui survient lors de mouvements rapides dans un film ou une émission sportive. En effet, dû à l'illumination périodique de l'écran par la source d'éclairage, un flou apparaît si la scène est animée d'un mouvement rapide. Plusieurs améliorations ont été menées (et d'autres sont toujours en cours) pour diminuer cet effet gênant, parmi elles la technologie ClearLCD de "Royal Philips Electronics" [3] qui diminue le temps de réponse des écrans LCD (presque 1000 fois plus grand que celui des écrans Plasma).

5.3 Les écrans CRT

Les écrans CRT sont les écrans des premiers téléviseurs et donc les plus anciens. Leur structure interne comprend un tube qui émet des électrons à une très grande vitesse ; ces électrons sont déviés par un champ électrique ou magnétique (créé par une différence de tensions) et bombardent l'écran formé de particules de phosphore qui émettent de la lumière à l'instant du choc. L'affichage de l'image nécessite finalement le balayage de l'écran d'une manière uniforme et la modulation du faisceau d'électrons par le signal vidéo. Peu à peu, les écrans CRT laissent leur place aux nouveaux écrans LCD et Plasma qui sont plus fins, pèsent moins et affichent une qualité d'image supérieure. Mais ils sont toujours maintenus par quelques fabricateurs d'écrans car leur temps de réponse est faible et leur reproduction des couleurs et du contraste est fidèle. De plus, leurs prix sont les plus abordables parmi les trois types d'écrans.

Chapitre 6

Amélioration de la qualité de vidéos : l'exemple intégré de Pixel Plus 3 HD

La technologie Pixel Plus de “Royal Philips Electronics” vise à améliorer la qualité de la vidéo affichée (surtout sur des écrans LCD) par une série de post-traitements. Dans le cadre de la HDTV, Philips présente sa technologie Pixel Plus 3 HD [3] qui est la dernière version de Pixel Plus comprenant les traitements les plus performants ajoutés à ceux déjà existants. Cette nouvelle technologie comprend une série de post-traitements visant à améliorer les aspects suivants de la vidéo : la netteté, le contraste, le bruit, le mouvement et les couleurs. Nous décrivons dans ce qui suit les techniques utilisées dans ces post-traitements.

6.1 L'augmentation de la netteté

Pour augmenter la netteté de l'image, Pixel Plus 3 HD intègre la technique LTI (“*Luminance Transient Improvement*”) qui augmente le nombre de lignes et de pixels par ligne (pour occuper toute la résolution de l'écran) et modifie la valeur de la composante de luminance de chaque pixel. La technique LTI consiste à interpoler des pixels c'est-à-dire ajouter des pixels dont la valeur de la composante de luminance est fixée à une valeur proche de celles des pixels correspondants. Cette technique LTI a pour objectif l'augmentation de la pente du gradient des contours : ceci résulte en une image plus nette.

6.2 La réduction du bruit

Pour diminuer le bruit, Pixel Plus 3 HD effectue les post-traitements d’affichage intégralement dans le domaine numérique pour éviter le bruit résultant des conversions analogique-numérique et inversement. De plus, la quantité de bruit est mesurée et la technique DNR (*“Digital Noise Reduction”*) réduit le bruit dans la proportion correspondante. Les visages et les régions uniformes sont surtout ciblés par cette réduction.

6.3 La réduction du flou

Pixel Plus 3 HD tente de réduire le “motion blur” des écrans LCD à l’aide de la technique ClearLCD qui diminue le temps de réponse des écrans LCD. Son principe est le suivant : l’objet en mouvement est illuminé seulement quand il est dans la position désirée c’est-à-dire qu’il n’est perçu qu’au moment où il atteint sa position correcte. De plus, la technique “Reverse 3-2 Pulldown” élimine le flou né potentiellement du passage de vidéos filmées en 24 images/seconde à une diffusion de vidéos en 30 images/seconde. Ceci réduit le flou créé par le mouvement et confère à l’image une bonne qualité.

6.4 L’augmentation du contraste

Comme l’accentuation du contraste des images augmente leur qualité, Pixel Plus 3 HD utilise la technique “Dynamic Contrast” pour assombrir les régions initialement sombres et éclaircir les régions initialement claires. Elle comprend aussi une correction “gamma”¹ positive (respectivement négative) est appliquée aux régions sombres (respectivement claires) de l’image pour permettre la visualisation des détails les plus fins. Ces améliorations sont nécessaires surtout dans le cas des écrans LCD qui souffrent d’une reproduction “imparfaite” de la couleur noire et d’une mauvaise visualisation des détails dans les régions sombres de l’image.

¹La correction gamma est une fonction non-linéaire liant la valeur de la composante de luminance du signal vidéo à la luminosité de l’écran.

6.5 L'augmentation de la saturation des couleurs

La dernière amélioration apportée par Pixel Plus 3 HD est l'augmentation de la saturation des couleurs c'est-à-dire les rendre plus vives. Pour ce faire, la technique "Color Booster" est utilisée. Elle consiste à saturer la couleur verte pour obtenir un vert aussi naturel que possible, à normaliser les valeurs de pixels de couleur bleue pour augmenter la "profondeur" de l'image, à augmenter la netteté des objets de couleur rouge et à préserver la naturalité de la couleur de la peau humaine. Cette combinaison d'améliorations de couleurs résulte en une qualité d'image qui parfois peut impressionner.

La technologie Pixel Plus 3 HD est un exemple d'amélioration de qualité de vidéos qui s'inscrit dans le cadre de la HDTV. Mais il existe aussi d'autres post-traitements qui visent à améliorer la qualité d'une vidéo avant son affichage. Cependant, le besoin d'une mesure de l'importance de ces améliorations implique la nécessité de l'élaboration d'un critère de qualité de vidéos améliorées. La méthodologie de mesure de cette amélioration fait l'objet du chapitre suivant.

Chapitre 7

La mesure d'amélioration de qualité de vidéos

La majorité des critères objectifs de qualité présentés dans la seconde partie sont dédiés à l'évaluation de la qualité d'images (ou vidéos) dégradées. Pourtant, nous retrouvons dans la littérature quelques critères de qualité qui mesurent l'amélioration de la qualité d'une image (ou vidéo). Le principe de la mesure effectuée par de tels critères est similaire à celle des critères exposés à la seconde partie : à partir de traits caractéristiques particuliers, une note de qualité de l'image (ou vidéo) est calculée. Les principaux traits caractéristiques mesurés pour déterminer l'amélioration apportée par un traitement particulier (ou série de traitements) sont la netteté des contours, le contraste et les couleurs. Nous exposons dans ce chapitre quelques exemples de mesures de ces traits caractéristiques.

7.1 Mesures de la netteté des contours

La netteté des contours d'une image est définie comme étant l'acuité des contours des objets de cette image. Elle représente aussi la profondeur de l'image, synonyme de sa troisième dimension (dans la direction perpendiculaire au plan de l'image). L'augmentation de la netteté des contours d'une image est un facteur important dans l'amélioration de la qualité d'une image. En effet, plusieurs études ont montré que les observateurs tendent à préférer une image plus nette qu'elle ne l'est en réalité. Nous présentons dans ce qui suit quelques mesures de netteté de la littérature.

Dans [29], Zhang *et al.* proposent trois métriques qui mesurent la netteté des contours d'une image. Ils font aussi des expériences pour déterminer

quel est le niveau de netteté préféré des observateurs humains. La première métrique, appelée DSS (*“Digital Sharpness Scale”*) mesure le contraste local et les contours de l’image. Les deux autres métriques, AETW (*“Average Edge Transition Width”*) et AETS (*“Average Edge Transition Slope”*) se basent sur l’extraction des contours de l’image et de leurs profils.

Les tests effectués montrent que les prédictions des métriques DSS et AETS corrélaient bien avec le jugement des observateurs humains. De plus, ils montrent que le niveau de netteté des contours d’une image dépend fortement du contenu de cette image c’est-à-dire que le niveau de netteté préféré (par les observateurs) d’une image texturée est supérieur à celui d’une image contenant des scènes humaines.

Dans [30], Caviedes et Oberti présentent une mesure de la netteté basée sur le coefficient de Kurtosis calculé dans les deux dimensions x et y . Le coefficient de Kurtosis est calculé pour chaque contour dans chaque bloc de l’image et la moyenne de ces coefficients sur toute l’image donne la mesure de netteté finale. L’application de cette métrique sur les vidéos donne de bons résultats mais en présence de dégradations (bruit, effets de blocs...), ces performances diminuent beaucoup. Caviedes et Oberti proposent finalement l’intégration de leur mesure de netteté dans un critère de qualité.

Ferzli *et al.* proposent aussi dans [31] l’utilisation du coefficient de Kurtosis dans la mesure de la netteté des contours d’une image. La méthode proposée opère dans le domaine des ondelettes (*“Discrete Dyadic Wavelet Transform”* : DDWT) qui permet la séparation du bruit et du signal de l’image. La mesure de netteté est appliquée aux deux sous-bandes résultant de la transformée en ondelettes et leur moyenne donne la mesure finale de netteté. Cette métrique est testée sur une seule image flouée et les résultats montrent que la métrique indique une diminution de la netteté quand la quantité de flou augmente. Mais l’application de cette métrique à des images de différents contenus n’est pas fructueuse et implique donc la nécessité de l’élaboration d’une métrique indépendante du contenu des images.

7.2 Mesures du contraste

La définition du contraste d’une image est donnée au paragraphe 3.2.1. Une étude approfondie du contraste est menée par Calabria et Fairchild dans [32, 33] dans laquelle ils définissent la relation entre le contraste et les autres caractéristiques perceptuelles de l’image (chrominance, netteté et luminance) et tentent de déterminer les niveaux de contraste préférés des observateurs hu-

maines. Les résultats de l'étude montrent une liaison étroite entre le contraste et ces caractéristiques perceptuelles et concluent donc que l'élaboration d'une métrique de contraste doit prendre en considération l'importance de ces caractéristiques perceptuelles. Dans la seconde partie de l'étude, Calabria et Fairchild proposent une modélisation de la perception du contraste et des préférences des observateurs.

Montrucchio *et al.* proposent dans [34] une mesure de contraste appelée CMM ("*Circular Mask Metric*") qui mesure le contraste d'images améliorées par les techniques d'égalisation d'histogrammes ("*Histogram Equalization*" : HE), d'égalisation adaptative d'histogrammes ("*Adaptive Histogram Equalization*" : AHE) et d'égalisation par blocs d'histogrammes ("*Partially Overlapping Sub-block Histogram Equalization*" : POSHE). Le principe de la CMM est le suivant : tout d'abord, l'image est filtrée par un filtre passe-bande. Ensuite, pour chaque pixel de l'image filtrée, les valeurs maximales et minimales des pixels dans un voisinage circulaire de rayon r (lié à la fréquence de coupure du filtre) sont calculées. Leur différence forme une nouvelle image et finalement la moyenne des valeurs de tous les pixels de cette nouvelle image donne le contraste local moyen correspondant à la fréquence spatiale choisie. Les résultats des tests effectués sur trois images montrent l'infériorité de la méthode HE qui est globale par rapport aux méthodes AHE et POSHE qui sont adaptées au contexte de l'image : ceci est en effet souligné par la CMM.

Une approche pour la mesure de l'amélioration du contraste d'une image est proposée par Carbonaro et Zingaretti dans [35]. En réalité, Carbonaro et Zingaretti exposent dans cet article leur algorithme d'amélioration du contraste mais proposent aussi une piste pour mesurer la quantité d'amélioration apportée par cet algorithme. Le principe de base de cet algorithme est la mesure de l'activité spatiale ("*Spatial Activity*" : SA), définie comme étant la somme (sur un voisinage) des différences entre les valeurs de pixels adjacents. Pour évaluer les performances de la technique d'amélioration, la SA est divisée tout d'abord en trois classes : faible SA (le bruit de l'image de fond), moyenne SA (les détails de l'image) et grande SA (les contours nets). Idéalement, la classe moyenne SA doit subir la plus grande augmentation de contraste. Ensuite, la différence d'activité spatiale entre l'image originale et celle améliorée est calculée pour les trois classes. Enfin, une somme pondérée de ces trois différences donne la mesure finale.

La validation de la technique d'amélioration est effectuée à l'aide de tests subjectifs de perception. L'optimisation de la mesure proposée est laissée pour des travaux futurs.

7.3 Mesures de la qualité des couleurs

Les couleurs d'une image représentent un aspect important dans l'évaluation de sa qualité. Cependant, la majorité des critères de qualité de la littérature ne prennent pas en compte cet aspect de l'image. Dans le cadre d'une amélioration de qualité d'images, ceci semble un impératif car les techniques d'améliorations visent généralement à augmenter la saturation des couleurs de l'image. Nous présentons dans ce qui suit deux mesures de la qualité des couleurs d'une image.

Janssen introduit dans [36] les notions d'utilité (*"usefulness"*) et de naturalité (*"naturalness"*) des images. L'utilité d'une image est définie comme étant la précision de sa représentation interne dans le SVH. La naturalité d'une image est la similarité de sa représentation interne avec son éventuelle représentation stockée dans la mémoire. En manipulant la chromacité des images, Janssen démontre que la qualité d'une image dépend fortement de la naturalité et de l'utilité de ses couleurs. Il propose ainsi deux métriques complexes basées sur la théorie des métriques flexibles¹ qui mesurent la naturalité et l'utilité des couleurs d'une image. Leur principe de base est le même : tout d'abord, l'image est transformée dans l'espace $Y'u'v'$; ensuite, les valeurs des trois composantes mesurées sont correspondues à une échelle de valeurs prédéfinies. Enfin, la différence entre la représentation interne et la représentation "éternelle" est calculée.

Les résultats des tests effectués sur des images dont la chrominance est manipulée montrent une bonne corrélation entre les jugements d'observateurs humains et les mesures des deux métriques.

Des mesures de la saturation des couleurs d'une image sont présentées dans [37]. Hasler et Süssstrunk utilisent l'espace de couleurs *Lab* et proposent 12 paramètres parmi lesquels plusieurs sous-groupes sont combinés pour former plusieurs métriques. En effet, la saturation des couleurs est définie comme étant une combinaison linéaire de paramètres comme les écart-types des valeurs suivant les axes de l'espace de couleurs par exemple ou la valeur moyenne de la saturation. Ainsi, plusieurs métriques sont formées et testées et une comparaison est établie entre les performances de ces métriques.

¹Une métrique flexible est une métrique dont l'échelle de mesure varie en fonction des conditions.

Conclusion

Nous avons présenté dans cette partie les nouvelles technologies d'écrans et leurs principaux avantages et inconvénients : les écrans LCD et Plasma sont favorisés dans le cas de la HDTV. Ensuite, nous avons expliqué les améliorations de qualité de vidéos intégrées dans la technologie Pixel Plus 3 HD : l'augmentation de la netteté des contours, la réduction du bruit, la réduction du flou, l'augmentation du contraste et l'augmentation de la vivacité des couleurs de l'image. D'une manière générale, ces améliorations sont communes à toutes les technologies mais leur implantation diffère d'un système à l'autre. Enfin, nous avons donné des exemples de métriques mesurant des traits caractéristiques de l'amélioration de qualité. Ces principaux traits caractéristiques sont la netteté des contours, le contraste et la qualité des couleurs de l'image.

Conclusion générale et perspectives

Nous avons présenté dans ce mémoire un véritable état de l'art des critères objectifs de qualité d'images et de vidéos. Pour ce faire, nous avons abordé les principaux thèmes en liaison avec les critères de qualité. La première partie a détaillé la méthodologie d'évaluation de la qualité visuelle d'images et de vidéos. Cette évaluation subjective de qualité est primordiale dans toute tentative d'élaboration d'un critère objectif de qualité car les résultats de ces tests subjectifs sont comparés aux notes fournies par le critère objectif de qualité.

Dans la seconde partie, qui est l'essence même de notre travail, nous nous sommes intéressés tout d'abord aux méthodes d'évaluation des performances des critères objectifs de qualité. Ensuite, nous avons présenté les principaux critères de qualité d'images de la littérature appartenant aux trois types (FR, RR et NR) et une synthèse des avantages et des inconvénients de chaque critère a été faite. Puis les principales différences entre un critère de qualité d'images et un critère de qualité de vidéos ont été expliquées. Ainsi, les critères de qualité de vidéos ont été introduits et les différentes méthodes d'obtention d'une note de qualité ont été présentées.

Enfin, pour contourner tous les aspects du thème de la recherche, nous avons présenté dans la troisième partie les nouvelles technologies d'écrans sur lesquelles vont être affichées les vidéos améliorées. A titre d'illustration d'un système intégré de post-traitements d'affichage, nous avons présenté l'exemple Pixel Plus 3 HD. Cet exemple donne une idée claire et globale de la méthodologie d'amélioration de qualité de vidéos. Enfin, nous avons terminé cette partie par la présentation des principales mesures d'améliorations dans la littérature. En effet, la qualité des couleurs, la netteté des contours et le contraste de l'image représentent les aspects les plus importants visés par l'amélioration de qualité de vidéos.

Dans la perspective de notre travail futur, qui consiste à élaborer un critère objectif de qualité de vidéos améliorées par des post-traitements d'affichage, nous pensons qu'une combinaison raisonnée de mesures de traits caractéristiques permet d'obtenir une note de qualité pour la vidéo améliorée. Probablement, les traits caractéristiques à mesurer seront la netteté des contours, le contraste de l'image et la qualité des couleurs. L'impact de ces traits caractéristiques sur la qualité de la vidéo devra être identifié ainsi que la façon de les combiner. Les performances du critère doivent être évaluées par la comparaison de ces notes fournies pour des vidéos et les notes fournies par les observateurs humains pour ces mêmes vidéos.

Bibliographie

- [1] European Information Communications Technology Industry Association. Conditions for High Definition Labelling of Display Devices, January 2005.
- [2] R. Schäfer, T. Wiegand, and H. Schwarz. The emerging H.264/AVC standard. *EBU Technical Review*, January 2003.
- [3] Philips Consumer Electronics. Pixel Plus 3 HD Technology Backgrounder. Technical report, Royal Philips Electronics, 2006.
- [4] K. Yamagishi and T. Hayashi. Analysis of psychological factors for quality assessment of interactive multimodal service. In *Proceedings of SPIE-IS&T Electronic Imaging*, pages 130–138, 2005.
- [5] ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures, June 2002.
- [6] VQEG-HDTV Group. Test Plan 1.0. Draft Paper, September 2005.
- [7] F. Pereira. Sensations, perceptions and emotions :Towards quality of experience evaluation for consumer electronics video adaptations. In *VPQM (Invited talk)*, 2005.
- [8] J. Caviedes and W. Ibrahim Ali. Principles of closed-loop video processing systems for objective quality optimization. In *Proceedings of VPQM*, 2005.
- [9] I. Heynderickx and R. Muijs. How perceptual knowledge might affect video processing. In *Proceedings of VPQM*, 2005.
- [10] Mathieu Carnec. *Critères de qualité d'images couleur avec référence réduite perceptuelle générique*. PhD thesis, Ecole Polytechnique de l'Université de Nantes, July 2004.
- [11] <http://fr.wikipedia.org>.
- [12] J.L. Mannos and D.J. Sakrison. The effects of a visual fidelity criterion on the encoding of images. In *IEEE Transactions on Information Theory*, pages 525–535, 1974.

- [13] Z. Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, pages 81–84, March 2002.
- [14] S. Daly *et al.* Predicting visible differences in high dynamic range images-Model and its calibration. In *Proceedings of SPIE-IS&T Electronic Imaging*, pages 204–214, 2005.
- [15] Z. Whang, H.R. Sheikh, and A.C. Bovik. *The handbook of video databases :Design and Applications*, chapter Chapter 41 :Objective Video Quality Assessment, pages 1041–1078. CRC Press, September 2003.
- [16] I.P. Gunawan and M. Ghanbari. Reduced-reference picture quality estimation by using local harmonic amplitude information. In *Proceedings of London Communications Symposium*, pages 137–140, September 2003.
- [17] G. Luo. A novel technique of image quality objective measurement by wavelet analysis throughout the spatial frequency range. In *Proceedings of SPIE-IS&T Electronic Imaging*, pages 173–184, 2005.
- [18] T. Takeuchi and K. De Valois. Sharpening image motion based on the spatio-temporal characteristics of human vision. In *Proceedings of SPIE-IS&T Electronic Imaging*, pages 83–94, 2005.
- [19] W. Dai, Z. Cai, and W.E. Lynch. An artifacts-based video quality metric using fuzzy logic. In *Proceedings of SPIE-IS&T Electronic Imaging*, pages 546–553, 2005.
- [20] A.P. Hekstra *et al.* PVQM- A perceptual video quality measure. *Journal of Signal Processing : Image Communication*, Vol 17 :781–798, November 2002.
- [21] A.A. Webster *et al.* An objective video quality assessment system based on human perception. In *Proceedings of SPIE*, volume 1913, pages 15–26, 1993.
- [22] S. Wolf and H. Pinson. Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system. In *Proceedings of SPIE*, volume 3845, pages 266–277, 1999.
- [23] Patrick Le Callet *et al.* No reference and reduced reference video quality metrics for end to end qos monitoring. *IEICE Transactions on Communications*, E89-B(2) :289–296, February 2006.
- [24] O. Sugimoto *et al.* Objective measurement scheme for perceived picture quality degradation caused by MPEG encoding without any reference pictures. In *Proceedings of SPIE*, volume 4310, pages 932–939, 2001.
- [25] Peter G. Engeldrum. Extending image quality models. In *Proceedings of IS&T's PICS*, pages 65–69, 2002.

- [26] M.C.Q. Farias, J.M. Foley, and S.K. Mitra. Perceptual analysis of video impairments that combine blocky, blurry, noisy, and ringing synthetic artifacts. In *Proceedings of SPIE-IS&T Electronic Imaging*, pages 107–118, 2005.
- [27] J.M. Foley *et al.* Annoyance of individual artifacts in MPEG-2 compressed video and their relation to overall annoyance. In *Proceedings of SPIE-IS&T Electronic Imaging*, pages 595–606, 2005.
- [28] I. Heynderickx *et al.* Objective quality assessment of displayed images by using neural networks. *Journal of Signal Processing :Image Communication*, Vol 20 :643–661, August 2005.
- [29] B. Zhang, Z. Pizlo, and J.P. Allebach. An investigation of perceived sharpness and sharpness metrics. In *Proceedings of SPIE-IS&T Electronic Imaging*, volume 5668, pages 98–110, 2005.
- [30] J. Caviedes and F. Oberti. A new sharpness metric based on local kurtosis, edge and energy information. *Journal of Signal Processing :Image Communication*, Vol 19 :147–161, February 2004.
- [31] R. Ferzli, L.J. Karam, and J. Caviedes. A robust image sharpness metric based on kurtosis measurement of wavelet coefficients. In *VPQM (Invited talk)*, 2005.
- [32] M.D. Fairchild and A.J. Calabria. Perceived image contrast and observer preference II. Empirical modeling of perceived image contrast and observer preference data. *Journal of Imaging Science and Technology*, pages 479–493, 2003.
- [33] M.D. Fairchild and A.J. Calabria. Perceived image contrast and observer preference I. The effects of lightness, chroma and sharpness manipulations on contrast perception. *Journal of Imaging Science and Technology*, pages 494–508, 2003.
- [34] B. Montrucchio *et al.* Measuring isotropic local contrast :A circular mask based approach. In *WSCG (Short Papers)*, pages 83–90, 2002.
- [35] A. Carbonaro and P. Zingaretti. A comprehensive approach to image-contrast enhancement. In *10th ICIAP*, pages 241–246, 1999.
- [36] T.J.W.M. Janssen. *Computational Image Quality*. PhD thesis, Technische Universiteit Eindhoven, 1999.
- [37] D. Hasler and S. Süssstrunk. Measuring colourfulness in natural images. In *Proceedings of SPIE-Human Vision and Electronic Imaging VIII*, volume 5007, pages 87–95, June 2003.