

# Evaluation de la qualité des codeurs vidéo dans le contexte de la vidéo-surveillance.

L. Quintard

M.C. Larabi

C. Fernandez-Maloigne

SIC Signal Image Communication

Université de Poitiers  
Blvd Marie & Pierre Curie, BP30179  
86962 Futuroscope Cedex

{quintard, larabi, maloigne}@sic.univ-poitiers.fr

Concours Jeune Chercheur : Oui

## Résumé

*Nous proposons une étude complète afin de fournir des recommandations de débits pour des codeurs de la famille MPEG-x et le MJPEG2000. Le but étant d'avoir une qualité perceptuelle identique à celle obtenue avec un codeur hardware MJPEG dont le débit est fixé à 5.6 Mbit/s. Cette étude s'applique dans le cadre de la vidéo surveillance. En premier, une étude objective est réalisée. Pour ce faire, trois métriques sont utilisées, le PSNR<sup>1</sup>, l'index Universel de qualité ainsi qu'une métrique que nous avons développée. Par la suite, nous proposons une étude subjective en accord avec les recommandations ITU<sup>2</sup> : le test DSIS<sup>3</sup> pour mesurer la dégradation globale de la vidéo par rapport à l'originale, le test DSCQS<sup>4</sup> pour mesurer l'impression visuelle globale de chaque vidéo. Finalement, l'étude de la corrélation entre les évaluations subjectives et objectives est réalisée.*

## Mots clefs

Evaluation subjective, évaluation objective, qualité vidéo, corrélation.

## 1 Introduction

La compression joue un rôle très important dans le contexte de la vidéo surveillance. En effet, la vidéo nécessite des capacités de stockage énorme. Pour pallier à ce point, la compression permet de réduire la taille des données et donc de diminuer la capacité de stockage nécessaire. De nombreux codecs sont utilisés, certains utilisent une méthodologie de compression image par image comme le MotionJPEG (MJPEG) et le MJPEG2000 [1] quand d'autres travaillent sur l'aspect temporel présent au sein d'une vidéo, c'est notam-

ment le cas des codecs de la famille MPEG<sup>5</sup> [2, 3, 4]. Dans tous les cas, les codecs introduisent des artefacts dans la vidéo [5]. Ces derniers affectent la qualité visuelle de la vidéo compressée.

Cependant, dans le cadre de la vidéo surveillance, il est nécessaire que le codec fournisse une qualité vidéo suffisante et, ce afin de pouvoir reconnaître le visage des personnes par exemple. La meilleure solution pour mener à bien ce problème, est d'évaluer objectivement et subjectivement les vidéos.

Les méthodes objectives sont basées pour la plupart sur la mesure de différence entre la vidéo originale et la vidéo compressée. Cette mesure peut utiliser un simple calcul mathématique comme le PSNR ou peut intégrer des propriétés du SVH<sup>6</sup>.

Les méthodes subjectives exploitent le jugement humain et nécessite des conditions d'expérimentations spécifiques comme par exemple une salle normalisée. Quand l'expérimentation psychophysique est faite, et que les analyses statistiques sont effectuées, les données issues de l'évaluation sont considérées comme cohérentes et peuvent être utilisées.

Une fois ces deux tests effectués, il est nécessaire d'étudier la corrélation existant entre les deux. Pour ce faire, le groupe VQEG<sup>7</sup> fournit des recommandations afin de corréliser les tests subjectifs et objectifs [6].

Ce papier décrit l'expérimentation que nous avons conduite pour différents codeurs vidéo. Le propos est de pouvoir fournir des recommandations de débits en fonction du codeur vidéo et ce pour obtenir une qualité visuelle équivalente à celle obtenue avec un codeur hardware MJPEG dont le débit est fixé par l'exploitant à 5.36 MBit/s. Un autre objectif est de pouvoir fournir des seuils de métrique dans le cas où il serait utile de tester différentes implémentations de codeurs hardware. Les codecs testés sont le MJPEG2000, MPEG-1, MPEG-2 et MPEG-4. Les

<sup>1</sup>Peak Signal Noise Ratio

<sup>2</sup>International Communication Union

<sup>3</sup>Double Stimulus Impairment Scale

<sup>4</sup>Double Stimulus Continuous Quality Scale

<sup>5</sup>Moving Picture Expert Group

<sup>6</sup>Système Visuel Humain

<sup>7</sup>Video Quality Expert Group

vidéos sélectionnées sont représentatives et de format CIF<sup>8</sup>. La figure 1 illustre un exemple de vidéos utilisées.



Figure 1 – Les figures -a- et -b- sont représentatives des vidéos utilisées pour l'étude.

Les résultats subjectifs sont obtenus en utilisant deux méthodologies définies dans [7], nommées DSIS et DSCQS. Les résultats objectifs sont obtenus en utilisant trois métriques. Le PSNR et deux autres métriques intégrant quelques propriétés du SVH, celles-ci sont l'Index Universel de Qualité [8], et une métrique développée lors de cette étude [9].

Ce papier est organisé de la manière suivante : La section 2 présente l'évaluation objective. La méthodologie de l'évaluation subjective est présentée en section 3. Ensuite, la section 4 est dédiée à l'étude de corrélation entre évaluation subjective et objective. Finalement, nous concluons en section 5.

## 2 Evaluation objective

Lorsque qu'une vidéo est compressée, elle subit des modifications entraînant des artefacts. Ces artefacts sont plus ou moins gênants d'un point de vue perceptuel [10]. Ils sont inhérents à la compression et peuvent prendre plusieurs formes. La figure 2 illustre un artefact de bloc et de débordement de couleurs.



Figure 2 – Exemple d'artefacts, effet de bloc (gauche), débordement de couleurs (droite).

Le but d'un test objectif est donc de pouvoir évaluer ces différents artefacts. Ces métriques peuvent être avec référence, sans référence ou avec référence réduite. Les métriques sans référence sont peu nombreuses dans la littérature. Le fait de ne pas avoir d'image originale

implique que l'on ne se trouve plus dans une mesure de fidélité mais bien dans une évaluation absolue de qualité. Cette dernière se calcule en général sur des attributs de l'image pour lesquels nous savons que le SVH est sensible, comme le contraste. Il faut de plus, considérer deux sortes de métriques, celles sans pondération, basées sur une mesure de distance comme le PSNR, et les autres sur critères pondérés qui prennent en compte dans leurs calculs quelques propriétés du SVH [11, 12].

Dans cette étude, nous nous proposons de tester trois métriques devant avoir un temps de calcul rapide. Pour ce faire, nous avons opté pour l'utilisation du PSNR, de l'index universel de qualité [8] ainsi qu'une métrique développée lors de cette étude [9].

Ces trois métriques sont avec référence. Le PSNR compare les différences entre les pixels de chaque image, quand deux images sont fidèles sa valeur se situe aux alentours de 40dB. Certes cette mesure n'est pas toujours corrélée avec le jugement humain, mais, dans le cadre de la compression les résultats sont parfois satisfaisants.

Le système visuel humain est un détecteur de contraste [13], l'index universel de qualité [8] et la nouvelle métrique [9] travaillent sur le contraste local d'une image. Elles ont respectivement une échelle de distance comprise entre  $[-1; 1]$  et  $[0; 1]$ . Avec 1 quand l'image de référence et l'image compressée sont fidèles d'un point de vue perceptuel. A l'inverse, les valeurs -1 ou 0 sont obtenues quand la fidélité perceptuelle est mauvaise.

L'index de qualité universel suit l'approche suivante :

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]} \quad (1)$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (2)$$

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2, \sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 \quad (3)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (4)$$

Les équations 1, 2, 3 et 4 expriment la corrélation existant entre deux images. La formule comprend en fait trois mesures. La première est la corrélation linéaire entre deux images  $x$  et  $y$ . La seconde est la différence entre les valeurs moyennes de  $x$  et  $y$ . La dernière mesure la similarité du contraste.

La fenêtre de calcul de taille  $A * A$  se déplace horizontalement et verticalement pixel par pixel.

Quant à la métrique que nous avons proposée elle travaille sur le contraste mais avec une approche différente. Le contraste est calculé par une fenêtre de taille  $A * A$  par l'équation 5.

<sup>8</sup>Common Intermediare Format (352 \* 288)

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (5)$$

$$R_i = \bar{x} - \bar{y} \quad (6)$$

$$R = 1 - \left( \left[ \frac{1}{M} \sum_{i=1}^M M |R_i| \right] * \theta \right) \quad (7)$$

La fenêtre se déplace horizontalement et verticalement par pas de dimension  $A$ . La différence est donnée fenêtre par fenêtre entre l'image de référence et l'image compressée comme le décrit l'équation 6. Le résultat final est calculé selon 7 où  $\theta$  représente le maximum de distorsion admissible.

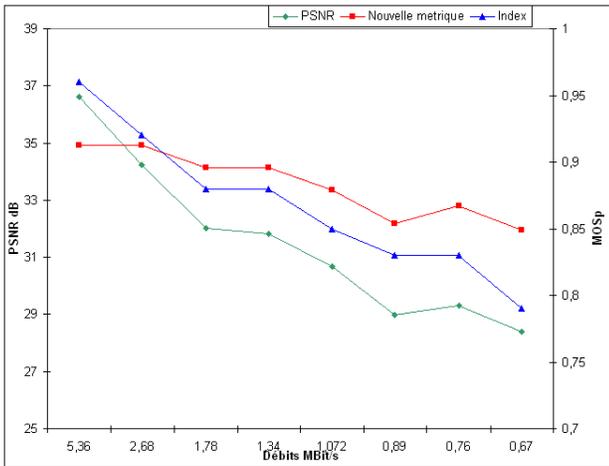


Figure 3 – Exemple de résultats pour les différentes métriques dans le cas du codeur MPEG-2.

La figure 3, montre un exemple de résultats pour les différentes métriques dans le cas du codeur MPEG-2.

A partir de cette figure, il est possible d'obtenir des seuils qui pourront servir par la suite à l'évaluation de codec. Néanmoins, ces seuils ne pourront réellement servir que si lors de l'étude de la corrélation entre les tests subjectifs et objectifs, les résultats sont concluants.

### 3 Evaluation psychophysique

#### 3.1 Conditions

L'évaluation subjective a été réalisée avec des vidéos représentant des conditions de vidéo surveillance (cf. figure 1). Les quatre codecs testés sont MJPEG2000, MPEG-1, MPEG-2, MPEG-4. Différents débits ont été choisis et ce de manière linéaire. La table 2 donne les échelles de débits évalués pour chaque codec.

Les conditions d'observation respectent les normes ITU décrites dans [7, 14]. La figure 4 (a) illustre la configuration

Tableau 1 – Débits utilisés pour les différents codecs.

MPEG1	0.76 à 5.36 Mbit/s
MPEG2	0.67 à 5.36 Mbit/s
MPEG4	0.357 à 5.36 Mbit/s
MJPEG2000	0.76 à 5.36 Mbit/s

de laboratoire utilisée. La distance entre l'écran et l'observateur est de 60 centimètres afin de respecter les conditions réelles du superviseur. Le moniteur est à tube cathodique, calibré (24" Sony). Chaque session est limitée à 25 minutes afin de ne pas dépasser les capacités de concentration de l'observateur. Le mur est d'un gris neutre. Les vidéos sont visualisées décompressées à 25 images/seconde.

#### 3.2 Méthode d'évaluation

Les deux méthodes utilisées lors de l'évaluation psychophysique sont spécifiées dans ITU-R Recommandation BT.500 [7]. La qualité perceptuelle doit être la même que celle obtenue pour un codec MJPEG à 5.36MBit/s. Pour ce faire, nous devons comparer les vidéos obtenues par l'utilisation de ce codec avec des vidéos obtenues par d'autres codecs. Ceci est en réalité le sujet du test DSIS. Nous utiliserons aussi le test DSCQS afin d'avoir une comparaison entre les codecs.

- **Le test DSIS** : l'observateur visualise différentes séquences vidéo par paire. La première vidéo étant la référence. Il doit juger la dégradation existant entre la vidéo compressée et la vidéo de référence. Pour ce faire, il dispose d'une échelle discrète de jugement allant de *imperceptible* à *très gênant*.
- **Le test DSCQS** : l'observateur visualise des vidéos deux à deux. Il donne son jugement sur chaque vidéo à partir d'une échelle linéaire allant de *mauvais* à *excellent*.

La figure 4 (b), montre l'interface développée pour cette étude. Chaque séquence vidéo dure 12 secondes.

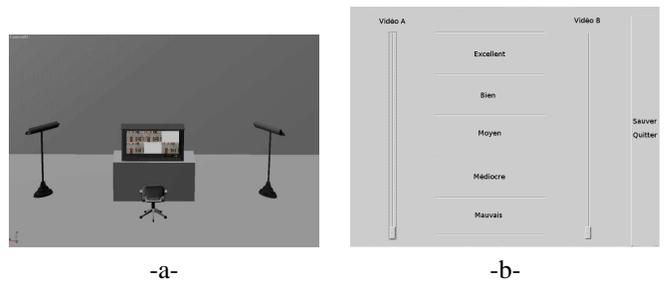


Figure 4 – Exemple d'une salle de laboratoire pour l'évaluation psychophysique (a). Interface développée pour le test DSCQS (b)

### 3.3 Les observateurs

Pour cette étude, 23 observateurs non expert ont participé à la session de test. L'acuité visuelle ainsi que la perception des couleurs ont été testés sur chaque observateur par l'intermédiaire des tests de *Snellen* et d'*Ishihara*.

### 3.4 Analyse de l'évaluation psychophysique

Une fois les tests réalisés, il est de nécessaire de calculer le MOS<sup>9</sup> et l'intervalle de confiance à 95% [7]. Ces résultats ne peuvent être calculés qu'à partir des observateurs dont le jugement est cohérent. Pour ce faire une étude statistique est réalisée en s'appuyant sur le test du *kurtosis*. Afin de faciliter l'analyse, une valeur numérique linéaire est attribuée selon l'échelle de valeur du test.

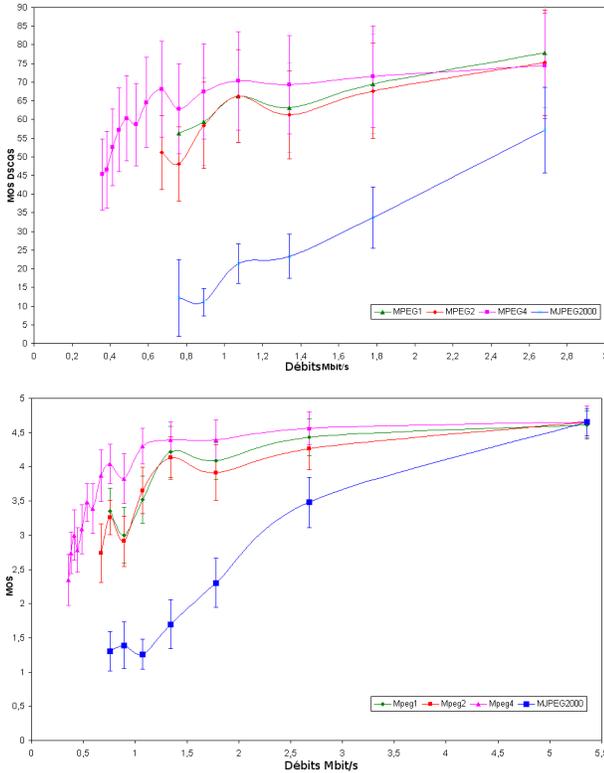


Figure 5 – MOS DSCQS (haut) MOS DSIS (bas) vs débits. Les barres verticales indiquent l'intervalle de confiance à 95%

Les figures 5 et 6 sont riches en informations et viennent confirmer des résultats certes prévisibles comme la supériorité du codeur MPEG-4 qui permet d'atteindre des débits records. Cependant, nous avons ici une information sur les débits permettant d'avoir une même fidélité perceptuelle (test DSIS). De plus, étant donné la bonne corrélation existant entre les tests DSIS et DSCQS (coefficient de Pearson de 94%) comme le montre la figure 6, nous pouvons ainsi connaître par l'intermédiaire du test DSCQS la qualité perceptuelle pour ces différents débits.

<sup>9</sup>Mean Opinion Score / score moyen des opinions

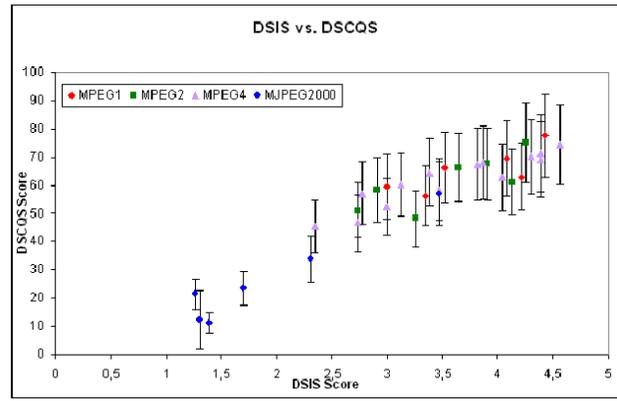


Figure 6 – Comparaison entre les MOS du test DSIS et DSCQS.

Tableau 2 – Débits minimum pour la même qualité perceptuelle obtenue avec un codec MJPEG à 5.36 MBit/s

CODEC	DEBITS
MPEG1	1 Mbit/s
MPEG2	1 Mbit/s
MPEG4	0.76 Mbit/s
MJPEG2000	3 Mbit/s

Le tableau 2 donne les débits minimum auxquels nous pouvons prétendre pour une qualité perceptuelle équivalente.

## 4 Etude de la corrélation

Les évaluations subjectives sont très contraignantes. C'est pourquoi une évaluation objective, qui utilise une métrique est bien plus intéressante. Néanmoins, il est nécessaire que la métrique fournisse des informations en concordance avec le jugement humain. Les attributs qui permettent de caractériser la performance d'une métrique objective par rapport aux données subjectives sont :

- Prédiction de l'exactitude
- Prédiction de la monotonie
- Prédiction de l'uniformité

Nous ne détaillerons ici qu'une partie de la prédiction de la monotonie, pour de plus amples informations le lecteur pourra se référer à [15] par exemple.

### 4.1 Modèle de prédiction de la monotonie

L'analyse de la corrélation nous indique le degré par lequel les valeurs de la variable  $Y$  peuvent être prédites, ou expliquées, par les valeurs de la variable  $X$ . Une forte corrélation implique qu'il est possible d'effectuer une inférence sur  $Y$  en partant de  $X$ .

La force et la direction du rapport entre  $X$  et  $Y$  sont données par le coefficient de corrélation. Il est souvent facile de prévoir s'il y a une corrélation, simplement en examinant les données en utilisant des nuages de points (voir

la fig. 7).

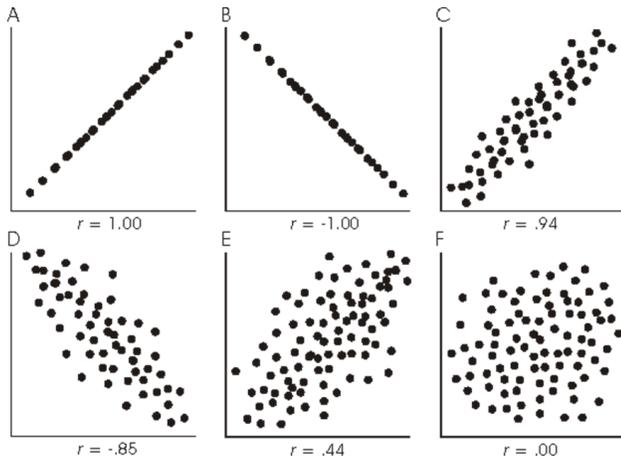


Figure 7 – Nuage de points tracé indique divers degrés de corrélation linéaire.

**Coefficient de corrélation Pearson** Le coefficient de corrélation Pearson  $r$  est utilisé pour des données sur des échelles d'intervalle ou de rapport, et est basé sur le concept de la covariance. Quand des échantillons  $X$  et  $Y$  sont corrélés il est possible de dire qu'ils varient conjointement ; ou qu'ils sont dans des modèles similaires. Le produit statistique du moment  $r$  est donné par :

$$r = \frac{n \sum_{i=0}^n x_i y_i - (\sum_{i=0}^n x_i)(\sum_{i=0}^n y_i)}{\sqrt{[n \sum_{i=0}^n x_i^2 - (\sum_{i=0}^n x_i)^2] [n \sum_{i=0}^n y_i^2 - (\sum_{i=0}^n y_i)^2]}} \quad (8)$$

Où  $n$  est le nombre de paires de scores. Le degré de liberté est  $df = n - 2$ .

Pour une valeur de  $df = 5$ , toute valeur de corrélation supérieure à 0.754, nous permettra de conclure que  $X$  et  $Y$  varie conjointement pour un intervalle de confiance à 95%. Les figures 8, 9 et 10 illustrent les corrélations obtenues selon la métrique considérée. Le coefficient de Pearson est proche de 1 dans tous les cas et au dessus de la valeur critique. Nous avons donc une forte corrélation entre les évaluations subjectives et objectives. Ceci nous permettra dans le futur de n'utiliser que les métriques pour une comparaison des qualités perceptuelles.

## 5 CONCLUSIONS

Cette étude a permis de fournir des recommandations de débits en fonction d'une qualité perceptuelle. Des tests subjectifs et objectifs ont été menés. Quatre codeurs ont été testés : MJPEG2000, MPEG-1, MPEG-2, MPEG-4. Les tests subjectifs ont permis de définir le tableau de recommandations 2. L'étude de la corrélation existant entre les tests subjectifs et objectifs autorise d'utiliser les seuils de métriques comme seuil de qualité sen restant dans la même base de vidéo.

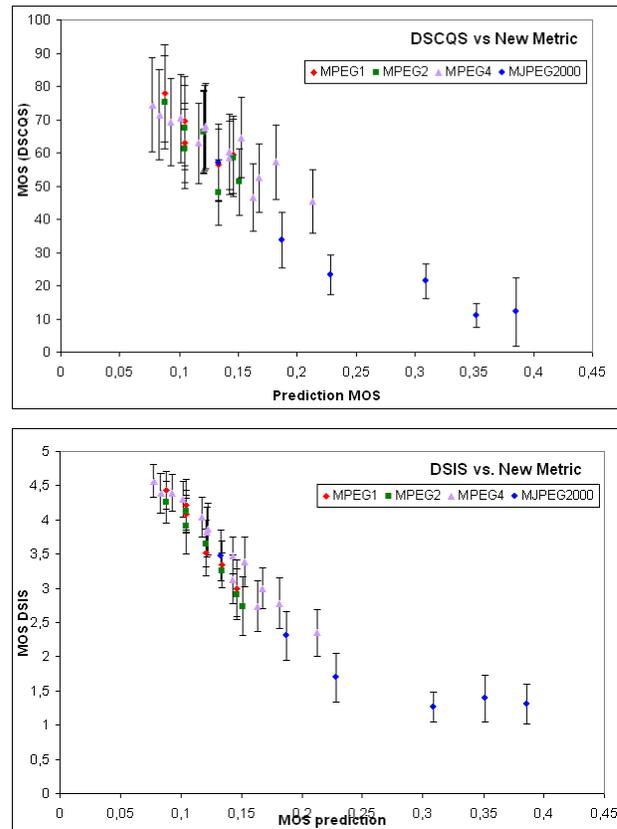


Figure 8 – MOS prédit vs. MOS subjectif, nouvelle métrique.

## Références

- [1] MJPEG2000 ISO/IEC JPEG committee. Information technology-jpeg 2000 image coding system. Rapport technique, Motion JPEG 2000-ISO IEC 15444-3 :2002, 2005.
- [2] MPEG-1 ISO IEC. Mpeg-1 standard. Rapport technique, ISO IEC, 1993.
- [3] MPEG-2 ISO IEC. Mpeg-2 standard. Rapport technique, ISO IEC, 1993.
- [4] MPEG-4 ISO IEC MPEG-98. Mpeg-4 overview. Rapport technique, Requirements Group, 1998.
- [5] Peter Symes. *Video compression demystified*. Mc Graw-Hill, New York, 2000.
- [6] VQEG. Final report from the video quality expert group on the validation of objective models of video quality assessment. Rapport technique, VQEG, 2000.
- [7] ITU-R Recommendation BT.500-10. Methodology for the subjective assessment of the quality of television picture. Rapport technique, International Communication Union, 2002.
- [8] Zhou Wang et Alan C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 2002.

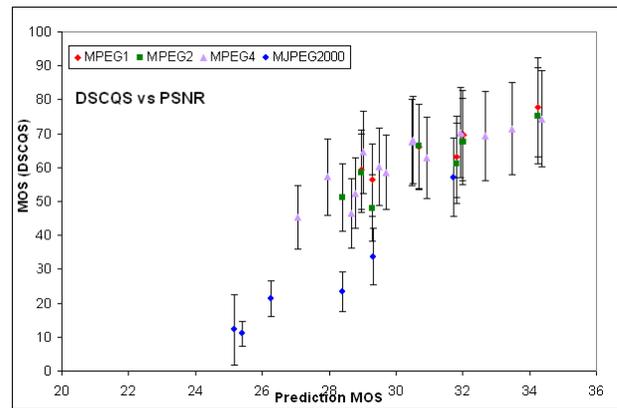
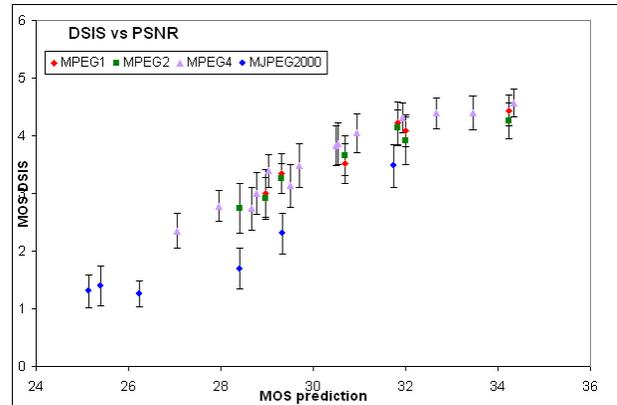
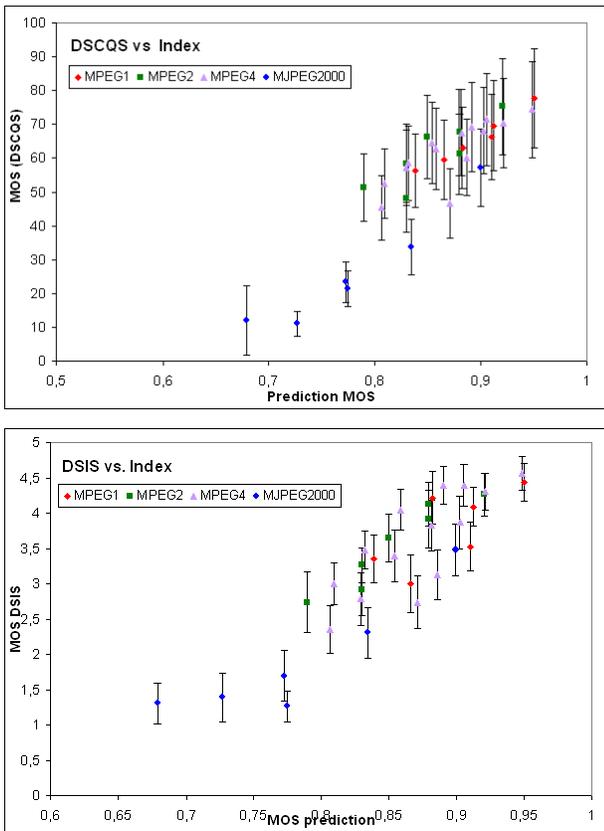


Figure 9 – MOS prédit vs. MOS subjectif, index universel de qualité.

Figure 10 – MOS prédit vs. MOS subjectif, PSNR.

- [9] ITU-R Recommendation P.910. Subjective video quality assessment methods for multimedia applications. Rapport technique, International Communication Union, 1996.
- [10] Z. Wang, H.R. Sheikh, et A.C. Bovik. *Handbook of video databases : design and applications*, chapitre Objective video quality assessment, pages 1041–1078. CRC press, Florida, september 2003.
- [11] C. Larabi, A. Stoica, et C. Fernandez-Maloigne. La qualité d'images entre mesures quantitatives, qualitatives et expérience psychovisuelles. *Ecole de printemps - images numériques couleurs*, mars 2003.
- [12] B.W Keelan. *Handbook of image quality : characterization and prediction*. Marcel Dekker, Inc, New York, 2002.
- [13] Dr. Salmon's. National board part 1 review. Rapport technique, July 2000.
- [14] L. Quintard et M.C. Larabi. Qualité vidéo : évaluation subjective et objective. Rapport technique, Université de poitiers, 2005.
- [15] D.Freedman, R.Pisani, et R. Purves. *Statistics*. W.W.Norton & Company, New York, 1998.