Université de Nantes

–

École polytechnique de l'université de Nantes

Projet RIAM ArchiPEG

(convention ANR05RIAM01401)

Lot 4.4 : Rapport sur les tests des algorithmes de filtrage et de pré-analyse du flux vidéo

O. Brouard, F. Delannay, V. Ricordel et D. Barba

Laboratoire IRCCyN - équipe IVC

septembre 2008

Table des matières

Table des figures

Introduction générale

Les travaux présentés dans ce rapport ont été réalisés dans le cadre du projet RIAM ArchiPEG qui relève de la convention ANR05RIAM01401. Ils correspondent à la quatrième tâche du sous-projet 4 intitulé : Pré-analyse et conditionnement du flux vidéo en haute définition.

Le dernier standard de codage vidéo développé par le JVT (Joint Video Team) regroupant les experts MPEG et ITU, à savoir MPEG-4 Part 10 (ou encore AVC ou H.624), vise à gagner jusqu'à 50% de la bande passante actuellement utilisée par MPEG-2 pour une qualité visuelle équivalente. On s'accorde donc à décrire ce standard [**??**, **??**] comme le futur de la compression des signaux TV capable de transmettre un programme HD[1] à des débits allant de 6 à 9 Mbits/s. Le schéma du codeur H.264 est présenté en figure 1.

De telles performances ne peuvent être atteintes qu'au prix d'une estimation et d'une compensation de mouvement complexes, afin d'exploiter de façon optimale les redondances spatiales et temporelles présentes au sein des vidéos. Le standard H.264 offre donc une palette large et complexe de possibilités pour l'estimation et la compensation de mouvement, notamment au niveau de :

- la précision des vecteurs déplacement : elle peut aller jusqu'au au quart de pixel pour la luminance et jusqu'au huitième de pixel pour la chrominance ;
- la taille variable des blocs estimés : 7 modes pour la prédiction inter (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4) et 2 modes pour la prédiction intra (16×16, 4×4) ;
- la sélection des images de référence : le choix de l'image de référence intervient au niveau macrobloc et sous-macrobloc contrairement aux normes précédentes telles que MPEG-2.

Le codeur H.264 réalise, lors de la phase de codage d'une séquence vidéo, une optimisation débit-distorsion pour chaque macrobloc afin d'obtenir le meilleur mode de codage (intra ou inter, taille des sous-partitions de macrobloc). Lors de cette optimisation débit-distorsion, le codeur doit réaliser une estimation de mouvement sur tous les modes inter en testant toutes les images de référence précédemment codées-décodées stockées dans un buffer. Cette phase est donc très coûteuse en temps de calcul, alors qu'elle ne garantie pas la cohérence avec le contenu spatio-temporel de la séquence vidéo.

Cette observation indique qu'une connaissance *a priori* sur le contenu spatio-temporel de la séquence vidéo à coder, permettrait de réduire significativement la charge de calculs du codeur. Il apparaît donc judicieux de placer, en amont du codeur, un phase de pré-analyse dédiée au mouvement au sein de la vidéo. Il sera ainsi possible d'appréhender de façon plus juste le mouvement des objets[2] et leur ancrage temporel. Cette analyse doit pouvoir caractériser le mouvement physique ainsi que la complexité locale de l'image dans le but d'accélérer le codage, en choisissant la meilleure stratégie offerte par le codeur H.264 (i.e. le meilleur jeu de paramètres du codeur). La connaissance approfondie des objets (cycle de vie, suivi spatio-temporel, texture, ...) présents dans une scène permettra notamment de décider, pour chacun d'entre eux, quelles sont les meilleures images de référence pour la prédiction et les modes les mieux adaptés à leur codage.

Ce document présente les résultats de l'outil de pré-analyse de flux vidéo haute définition en vue d'un encodage en temps réel sous le standard H.264. L'objectif est donc de fournir au codeur H.264 un jeu de paramètres adapté au codage d'une séquence vidéo et présentant une cohérence spatio-temporelle fonction des objets présents dans la scène. Le premier chapitre décrit les séquences vidéo

[1] TVHD : télévision haute-définition.

[2] Un objet désigne un ensemble de macroblocs dont le mouvement, la couleur et la texture sont homogènes.

Haute Définition utilisées lors des tests. Le deuxième chapitre de ce rapport présente les métriques utilisées lors des tests pour mesurer la qualité des vidéos codées. Le dernier chapitre présente les résultats obtenu après pré-analyse de la vidéo, ceux-ci sont comparés aux résultats obtenus à l'aide d'une approche classique de codage.

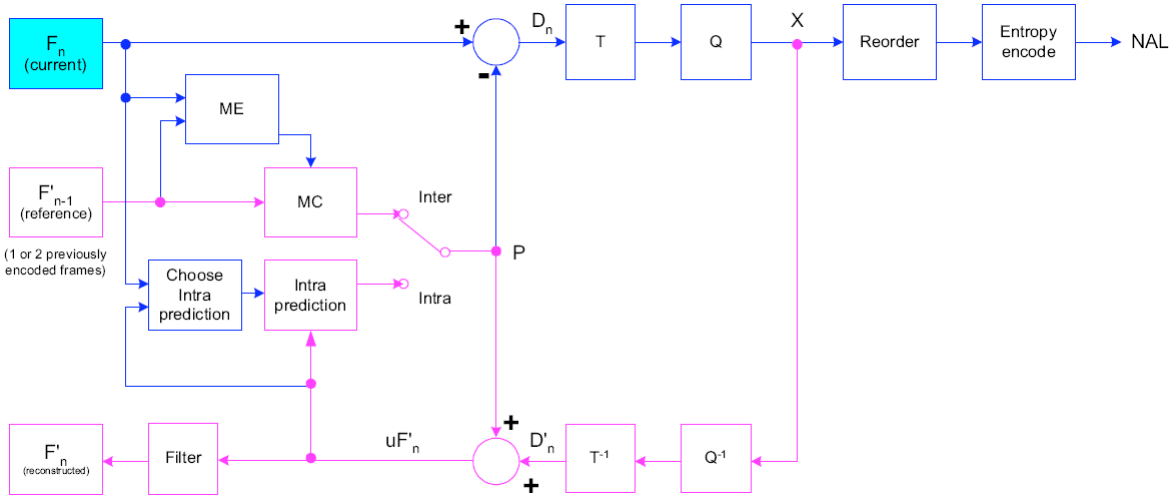

FIG. 1 – Schéma du codeur H.264/AVC [**??**].

Chapitre 1

Présentation des séquences vidéo utilisées lors des tests

Les séquences utilisées lors des tests réalisés pour les besoins de ce rapport sont disponibles via le serveur ftp `ftp://ftp.ldv.e-technik.tu-muenchen.de/pub/test_sequences/`. Ces séquences ont été filmées à une fréquence de 50 images par seconde avec l'équipement du SVT en octobre 2004. La plus grande attention a été donnée à la conversion des films vers un format numérique. Les détails concernant les conditions de prise de vue et les post-traitements sont présentés dans la documentation fournie par le SVT [1].

1.1 Les séquences 720p

Les séquences 720p utilisées ici sont des vidéos progressives de 720 lignes par 1280 colonnes, cadencées à 50 images par seconde, la structure d'échantillonnage couleur des composantes YUV est 4 :2 :0.

1.1.1 New mobil and calendar

La séquence comporte 500 images filmées en plan rapproché. La caméra, qui subit un mouvement translationnel puis de zoom arrière, filme un calendrier avec du texte et une photo détaillée du Vasa[1]. Á partir de la 355ème image apparaît un train en mouvement translationnel avec des jouets très colorés. Le fond est composé de deux types de papiers peints, le premier est jaune, uniforme avec quelques figures dessinées et le second est très texturé. La figure 1.1 présente une image extraite de la séquence *New mobil and calendar*.

1.1.2 Knightshields

La séquence comporte 500 images filmées en plan rapproché. Un homme avec une barbe et une veste très texturée marche devant un mur composé de boucliers de chevaliers détaillés. Á la fin de la séquence, le capteur effectue un zoom avant de la scène. La figure 1.2 présente une image extraite de la séquence *Knightshields*.

1.1.3 Parkrun

La séquence comporte 500 images filmées en plan éloigné. La scène représente un homme, avec un parapluie à la main, qui court le long d'un canal dans un parc puis s'arrête et reste immobile vers la 340ème image. L'arrière plan est composé d'arbres, de neige et d'une source d'eau. Le contenu est très détaillé. La figure 1.3 présente une image extraite de la séquence *Parkrun*.

[1] Le Vasa est un vaisseau de guerre scandinave du 17ème siècle.

1.2 La séquence 1080p

Les séquences 1080p utilisées ici sont des vidéos progressives de 1080 lignes par 1920 colonnes, cadencées à 25 images par seconde, la structure d'échantillonnage couleur des composantes YUV est également 4 :2 :0.

1.2.1 Tractor

La séquence comporte 690 images qui présentent un tracteur dans un champ. La séquence entière contient des zones sur lesquelles un très fort zoom arrière est appliqué de manière à en obtenir une vue totale. La caméra suit le tracteur, avec un mouvement chaotique, sur la structure du champ de récolte. La figure 1.4 présente une image extraite de la séquence *Tractor*.

FIG. 1.1 – Image 478 de la séquence *New mobil and calendar*.

FIG. 1.3 – Image 160 de la séquence *Parkrun*.

FIG. 1.2 – Image 1 de la séquence *Knightshields*.

FIG. 1.4 – Image 60 de la séquence *Tractor*.

Chapitre 2

Métriques utilisées

Nous utilisons deux métriques de qualité, la première est le PSNR (*Peak Signal Noise Ratio*) et la seconde est issue du logiciel VQA[2] (*Video Quality Analyzer*).

Le seul avantage de cette première métrique est sa simplicité et de ce fait son large déploiement. Les inconvénients sont nombreux et bien connus : faible corrélation avec les tests visuels de qualité, incapacité de la métrique à prendre en compte des effets de masquage des dégradations...

VQA est une solution de l'état de l'art pour mesurer la qualité perçue d'une vidéo et pour obtenir des rapports d'analyses détaillées sur la qualité vidéo et les distorsions visuelles. Entièrement basée sur la modélisation du système visuel humain (SVH), VQA produit des évaluations de qualité qui sont fortement corrélées avec le jugement humain de la qualité visuelle. La technologie de mesure de qualité vidéo intégrée dans VQA possède une grande précision. En effet, le coefficient de corrélation linéaire, entre les évaluations de qualité vidéo (calculées par VQA) et les évaluations de qualité subjective (données par les observateurs au cours de tests d'évaluation subjective de qualité dans des conditions normalisées), est en moyenne supérieur à 0,924.

2.1 Vue d'ensemble de la technologie de mesure de qualité

VQA produit des notes de qualité qui sont fortement corrélées avec le jugement humain de qualité visuelle, grâce à l'architecture de mesure de la qualité vidéo qui reproduit le comportement du dispositif d'affichage et du système visuel humain. Cette technologie est basée sur l'architecture représentée dans la figure 2.1.

Dans cette architecture, les principaux traitements sont la modélisation du SVH et l'extraction de caractéristiques perceptuelles. La modélisation du SVH transforme la vidéo dans un domaine de représentation perceptuelle. Cette étape correspond principalement à la transformation effectuée dans la rétine et dans la zone V1 du cortex visuel.

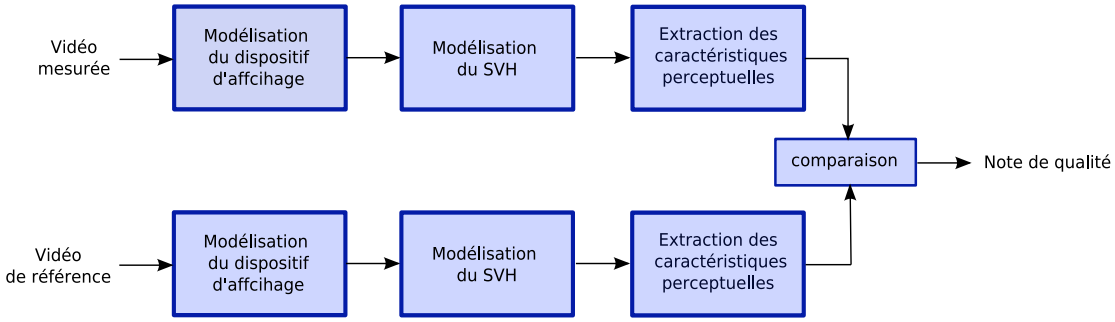

Fig. 2.1 – Architecture de la technologie de mesure de la qualité vidéo (VQA).

Ensuite, les caractéristiques perceptuelles sont extraites. Elles sont similaires aux caractéristiques extraites dans les zones V1 et V2 du cortex visuel. Elles décrivent principalement l'information structurelle, le contraste visuel, l'activité et la couleur, à un niveau local.

Enfin, les caractéristiques perceptuelles extraites de la vidéo mesurée sont comparées aux caractéristiques perceptuelles extraites de la vidéo de référence. Ces distorsions produisent une note de qualité qui est ensuite projetée sur une échelle de DMOS pour constituer le "DMOS prédit".

2.2 MOS et DMOS

MOS signifie "*Mean Opinion Score*" (moyenne des notes subjectives). Le MOS d'une séquence vidéo est, la valeur moyenne de toutes les évaluations subjectives recueillies, lors de tests d'évaluation subjective de la qualité vidéo de cette séquence vidéo. Lorsque le MOS augmente, la qualité perçue augmente. Au cours des tests d'évaluation subjective de la qualité vidéo, les observateurs jugent la qualité de vidéos dégradées mais ils peuvent également juger la qualité des vidéos de référence. Comme les vidéos de référence ne sont pas dégradées, elles obtiennent de très bonnes notes en terme de qualité par les observateurs humains. Cependant elles obtiennent rarement la note maximale. Par exemple, sur une échelle de vote allant de 0 (la plus faible qualité) à 100 (qualité supérieure), une vidéo de référence peut obtenir un MOS de 95, ce qui signifie que la plupart des observateurs ont trouvé sa qualité visuelle excellente, mais certains observateurs (ou la totalité d'entre eux) lui ont donné une note inférieure à 100. Cela peut être dû au fait que ces observateurs n'ont pas apprécié le contenu vidéo (par exemple, la vidéo représentait du football et certains observateurs n'aiment pas ça, ou bien le film avait un faible bruit d'acquisition, mais visible).

Le DMOS est le "*Differential Mean Opinion Score*" (différence des moyennes des notes subjectives). Le DMOS d'une vidéo est obtenu en calculant la différence entre le MOS de la vidéo de référence moins le MOS de cette vidéo. Par conséquent, si le DMOS est de 0, cela signifie que la vidéo mesurée à la même qualité que la vidéo de référence. Et lorsque le DMOS augmente, la qualité diminue. Lors de la mesure de la qualité d'une vidéo à l'aide de sa vidéo de référence, le niveau de qualité à prendre en considération est donc le DMOS plutôt que le MOS. En effet, utilisé le DMOS au lieu du MOS permet de diminuer l'effet de familiarité avec le contenu des vidéos ou des artéfacts dans les vidéos de référence dans les votes des observateurs.

Chapitre 3

Tests réalisés

3.1 Description des tests réalisés

Cette approche a été testée sur trois séquences ayant une résolution de 1280×720 pixels (*New mobile and calendar*, *Knightshields* et *Parkrun*) et sur une séquence ayant une résolution de 1920×1080 pixels (*Tractor*). Il s'agit ici de comparer la qualité globale lorsqu'on utilise un codage classique ou un codage perceptuel.

Les séquences ont été codées en utilisant les paramètres suivants :
- images I et P seulement (IPPP... IPPP...),
- cinq images de référence,
- tous les modes de prédiction INTER : 16×16, 8×16, 16×8, 8×8, 8×4, 4×8 et 4×4,
- les deux modes de prédiction INTRA : 16×16 et 4×4,
- fréquence de 50 images par seconde.

Un exemple de ligne de commande pour coder la séquence *tractor* au débit de $1000 kBits$ par seconde est présenté ci-dessous :

```
1 x264.exe -o "tractor_1000kbs.264" "tractor.yuv" 1920x1080 --
      partitions "p8x8,p4x4,i4x4" -r 5 --verbose --bitrate 1000 --
      fps 50 --frames 450 2> \tractor_1000kbs.txt
```

Afin de réaliser le codage perceptuel, nous devons modifier le coeur de codage. Cela consiste à déterminer le pas de quantification de chaque macrobloc. Le pas de quantification est déterminé en fonction de la carte de saillance spatio-temporelle. Plus une zone est saillante et plus elle sera quantifiée finement et inversement. La carte de saillance spatio-temporelle nous indique la saillance de chaque macrobloc situé sur l'image centrale d'un segment temporel (de neuf images). Les cartes de saillance des autres images du segment temporel (les quatre images précédentes et les quatre images suivantes) sont déduites à partir des informations de mouvement (projection compensée en mouvement de la carte de saillance de l'image centrale).

L'indice de saillance calculé pour un macrobloc varie entre 0 (saillance nulle) et 1 (très saillant). Afin de quantifier les macroblocs en fonction de leur indice de saillance, le pas de quantification doit être modifié par rapport à une stratégie classique de codage. Pour ce faire, on modifie la valeur du pas de quantification calculé par le codeur, de la façon suivante :

$$\begin{aligned} QP^{final}(i) &= QP^{initial}(i) + \frac{\overline{S}-S(i)}{\overline{S}} \times \alpha \times (QP_{max} - QP^{initial}(i)) \quad \text{si} \quad 0 \leq S(i) \leq \overline{S}, \\ QP^{final}(i) &= QP^{initial}(i) + \frac{\overline{S}-S(i)}{1-\overline{S}} \times \alpha \times (QP^{initial}(i) - QP_{min}) \quad \text{si} \quad \overline{S} \leq S(i) \leq 1, \end{aligned} \tag{3.1}$$

où $S(i)$ est l'indice de saillance du macrobloc i et $\overline{S}$ est la saillance moyenne de l'image à laquelle appartient le macrobloc i. $QP^{initial}(i)$ et $QP^{final}(i)$ sont respectivement, le pas de quantification initial déterminé par le codeur et le pas de quantification final modifié en fonction de l'indice de saillance du

macrobloc i. QP_{max} et QP_{min} sont respectivement les pas de quantification maximal et minimal du codeur. Lors de nos tests, nous avons fixé le paramètre α égal à $\frac{1}{2}$.

3.2 Résultats expérimentaux

3.2.1 PSNR

Les résultats obtenus en terme de PSNR en fonction du débit, par un codage classique et un codage perceptuel, sont donnés dans les figures 3.1, 3.2, 3.3 et 3.4. Pour les séquences *Tractor*, *New mobil and calendar* et *Parkrun*, les résultats en terme de PSNR sont très similaires pour les deux approches testées (codage classique et codage perceptuel). Par contre, pour la séquence *Knightshields*, notre méthode de codage perceptuel obtient un gain significatif en terme de PSNR. Le gain maximal obtenu est alors de $6,4dB$ pour un débit de $20Mbits$ par seconde. Il varie de $0,25dB$ à $3,3dB$ entre $6Mbits$ et $14Mbits$ par seconde. Notre méthode de codage adapté semble améliorer la qualité en terme de PSNR pour la séquence vidéo *Knightshields*. Cependant, nous savons que le PSNR est une métrique faiblement corrélée avec les résultats de tests visuels de qualité et nous devons tenir compte de ce défaut. C'est pourquoi nous avons également testé par la suite les deux approches avec une métrique de qualité corrélée avec les résultats d'évaluations subjectives de qualité.

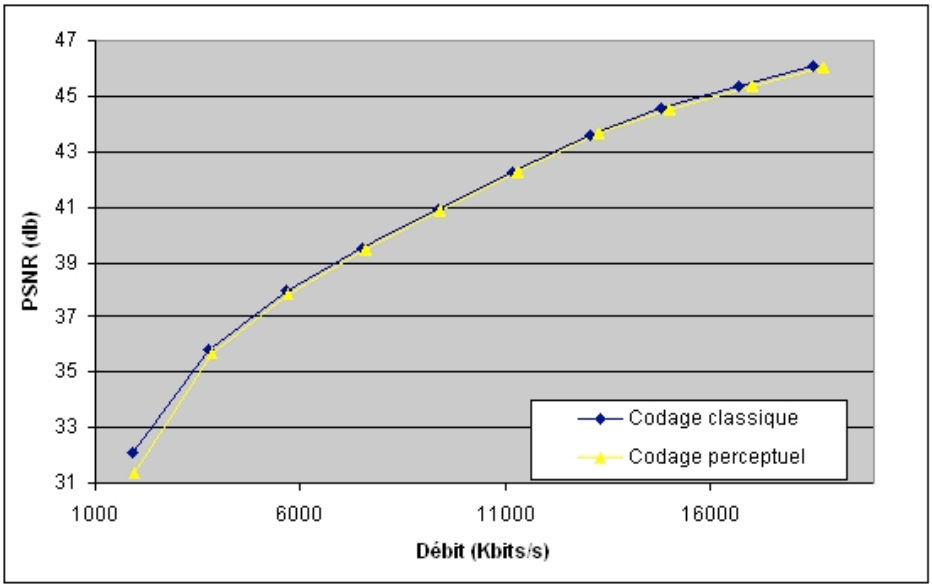

FIG. 3.1 – PSNR en fonction du débit pour la séquence *Tractor*.

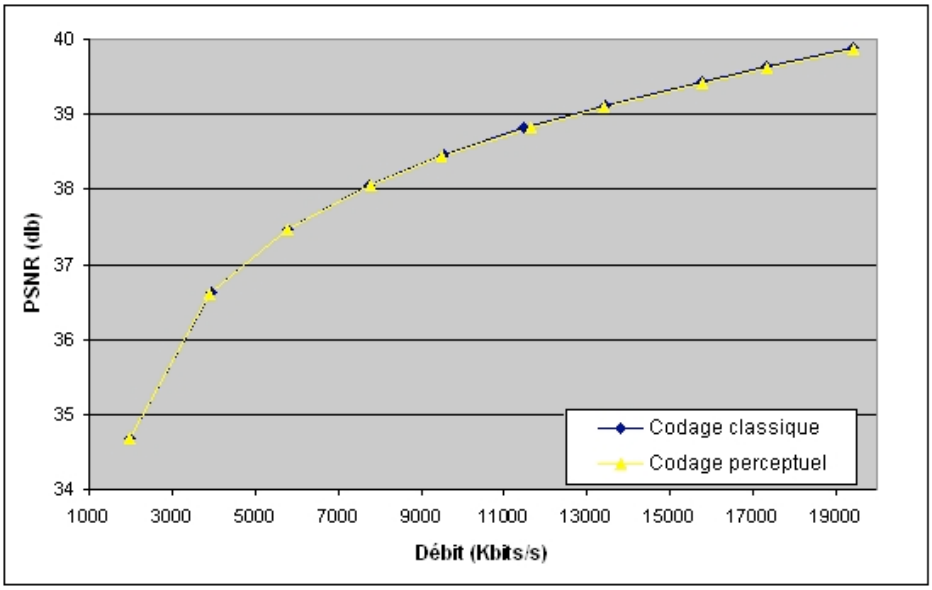

FIG. 3.2 – PSNR en fonction du débit pour la séquence *New mobile and Calendar*.

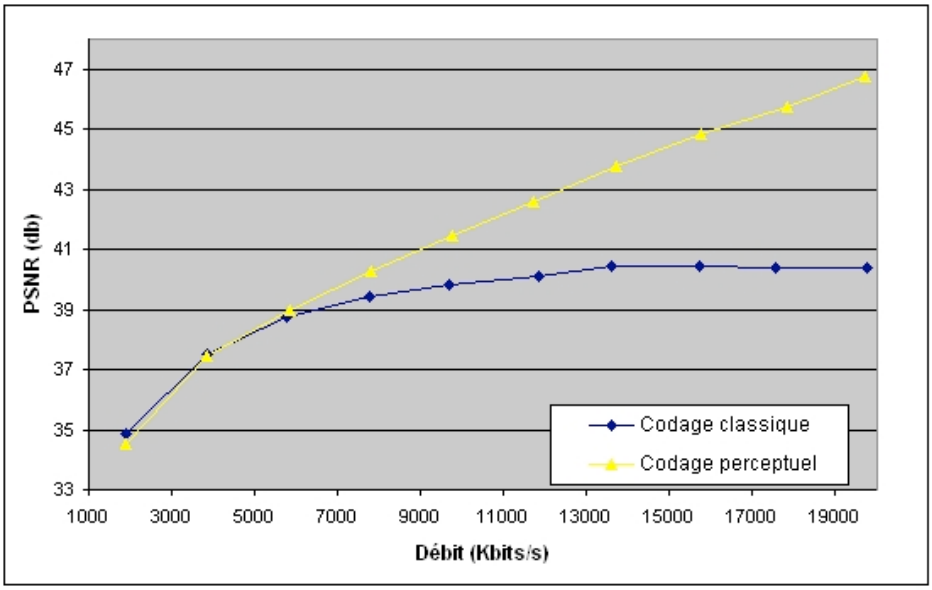

FIG. 3.3 – PSNR en fonction du débit pour la séquence *Knightshields*.

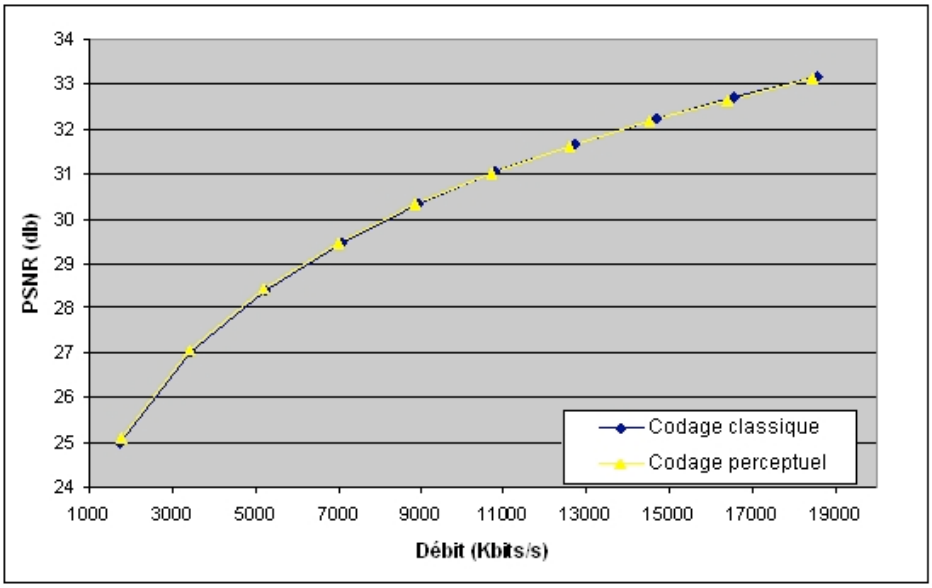

FIG. 3.4 – PSNR en fonction du débit pour la séquence *Parkrun*.

3.2.2 Qualité de la zone saillante

Afin d'évaluer notre méthode de codage perceptuel, qui détermine le pas de quantification d'un macrobloc en fonction de son indice de saillance, nous mesurons la qualité en terme de PSNR pour la zone la plus saillante (région d'intérêt) dans chaque image de la séquence vidéo. La région d'intérêt de chaque image est obtenue à partir de la valeur de saillance moyenne. Pour chaque macrobloc, si l'indice de saillance est supérieur à la saillance moyenne de l'image, alors le macrobloc appartient à la région d'intérêt, comme cela est illustré dans la figure 3.5.

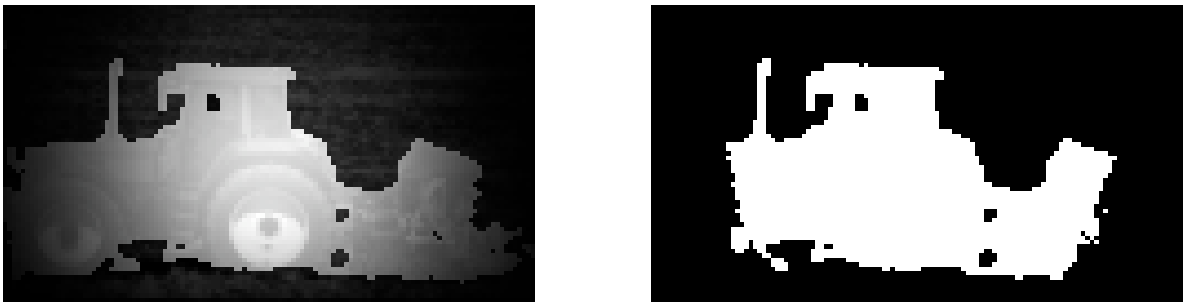

FIG. 3.5 – Région d'intérêt pour l'image 143 de la séquence *Tractor*, à gauche la carte de saillance et à droite la région d'intérêt obtenue.

Les figures 3.6, 3.7, 3.8 et 3.9 présentent les résultats en terme de PSNR pour les quatre séquences vidéo testées pour différents débits. Chaque figure présente la qualité globale et la qualité de la zone d'intérêt en terme de PSNR pour les deux approches de codage.

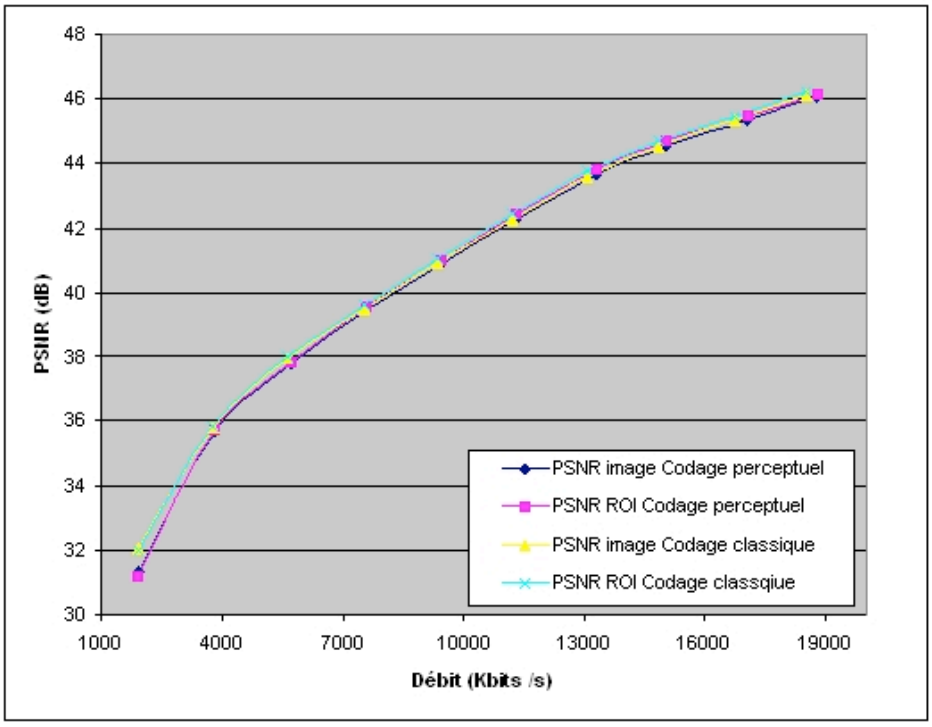

FIG. 3.6 – PSNR (global et zone saillante) en fonction du débit pour la séquence *Tractor*.

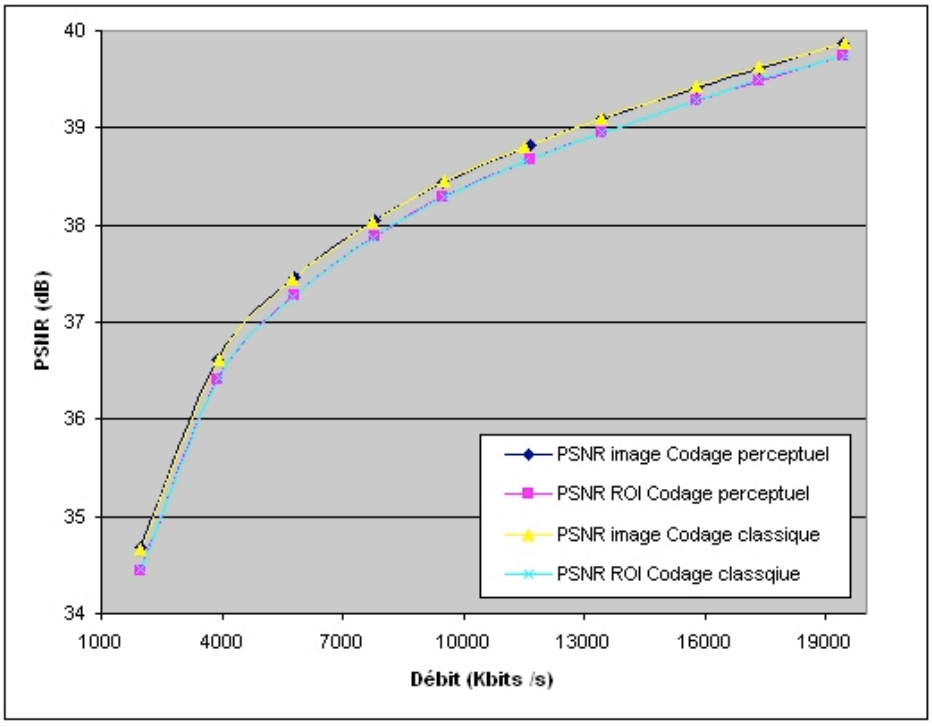

FIG. 3.7 – PSNR (global et zone saillante) en fonction du débit pour la séquence *New mobile and Calendar*.

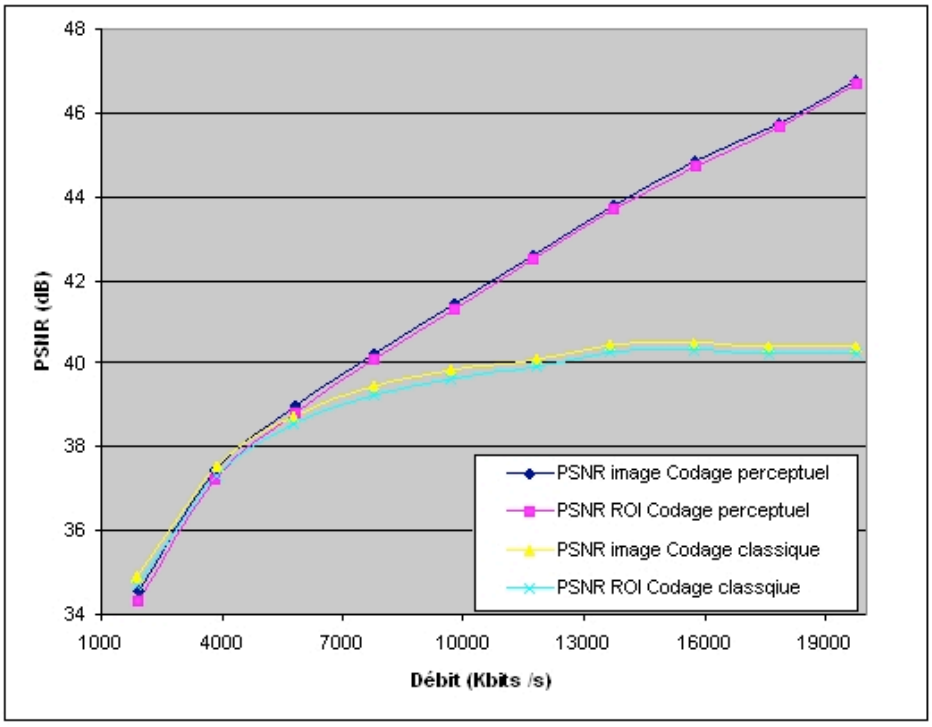

FIG. 3.8 – PSNR (global et zone saillante) en fonction du débit pour la séquence *Knightshields*.

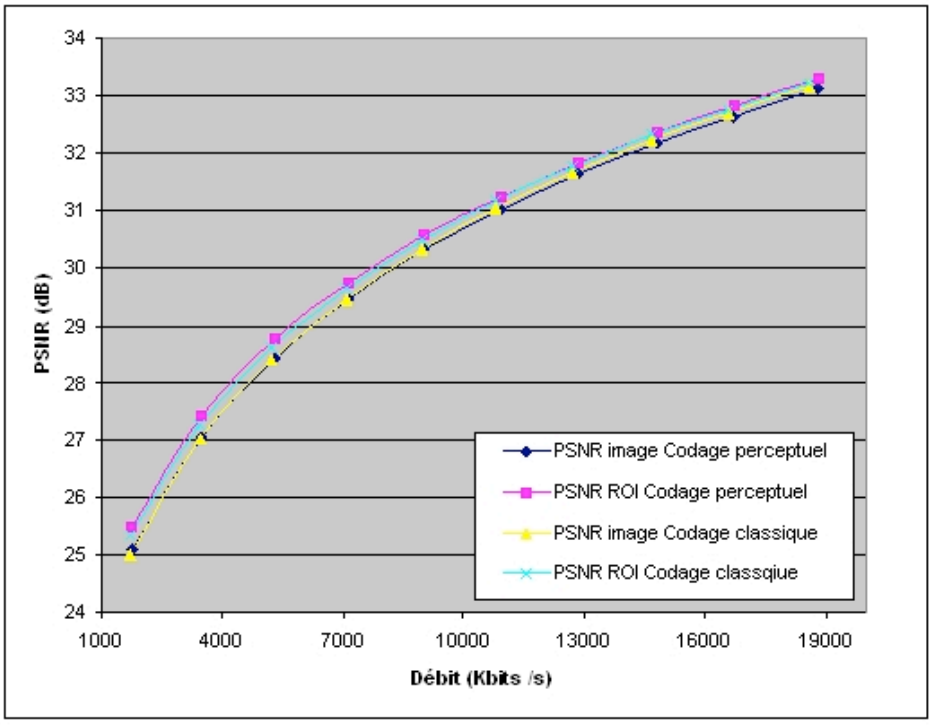

FIG. 3.9 – PSNR (global et zone saillante) en fonction du débit pour la séquence *Parkrun*.

Comparativement aux résultats obtenus avec un codage classique, notre méthode de codage per-

ceptuel n'améliore pas la qualité en terme de PSNR pour la région d'intérêt. Cependant, ces résultats doivent être nuancés. En effet, les résultats présentés ci-dessus indiquent la qualité moyenne pour toutes les images des séquences vidéos codées. Par la suite (cf section 3.2.4), nous présentons des résultats en terme de PSNR et DMOS pour chaque image des différentes séquences vidéos encodées à différents débits.

3.2.3 DMOS

Les résultats obtenus en terme de DMOS par rapport au débit par un codage classique et un codage perceptuel sont donnés dans les figures 3.10, 3.11, 3.12 et 3.13.

La figure 3.10 montre que notre approche obtient des résultats légèrement inférieurs en terme de DMOS pour la séquence *Tractor*. En effet, la perte en DMOS est de $0,23$ en moyenne sur tous les débits testés et de $0,09$ pour les débits allant de 6 à $12Mbits$ par seconde. Le gain maximal est de $2,2$ et est obtenu pour un débit cible de $4Mbits$ par seconde.

Pour la séquence *New mobile and Calendar* (cf figure 3.11), les résultats obtenus avec une approche de codage classique sont meilleurs que ceux obtenus avec notre approche de codage perceptuel. En effet, la perte de DMOS est de $2,59$ en moyenne et de $2,91$ pour les débits cibles allant de 6 à $12Mbits$ par seconde.

Les résultats obtenus pour la séquence vidéo *Knightshields* (cf figure 3.12) montrent que notre approche de codage perceptuel permet d'obtenir une meilleure qualité pour des débits importants (supérieurs à $14Mbits$ par seconde). Par contre, pour les débits allant de 6 à $12Mbits$ par seconde, la perte de DMOS est de $1,58$ en moyenne. Ces résultats illustrent parfaitement le fait que le PSNR est une métrique faiblement corrélée avec les tests subjectifs de qualité visuelle. En effet, pour la séquence vidéo *Knightshields*, nous obtenions toujours en terme de PSNR de meilleurs résultats avec notre approche de codage adapté.

Pour la séquence Parkrun (cf figure 3.13), les résultats obtenus avec notre méthode de codage perceptuel sont légèrement meilleurs en terme de DMOS. En effet, le gain en DMOS est de $0,68$ en moyenne sur tous les débits testés et de $0,20$ pour les débits allant de 6 à $12Mbits$ par seconde. Le gain maximal est de $3,85$ et est obtenu pour un débit cible de $2Mbits$ par seconde.

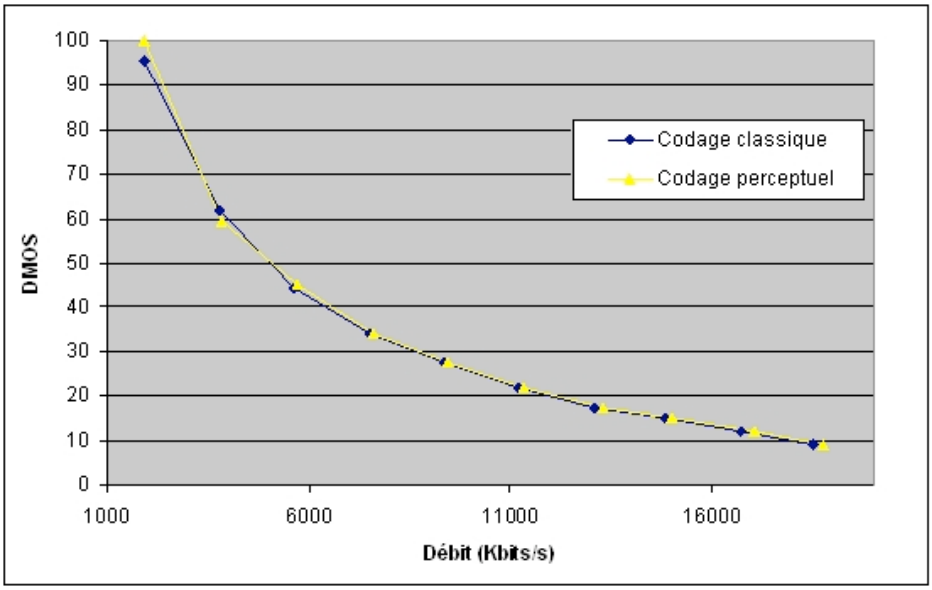

FIG. 3.10 – DMOS en fonction du débit pour la séquence *Tractor*.

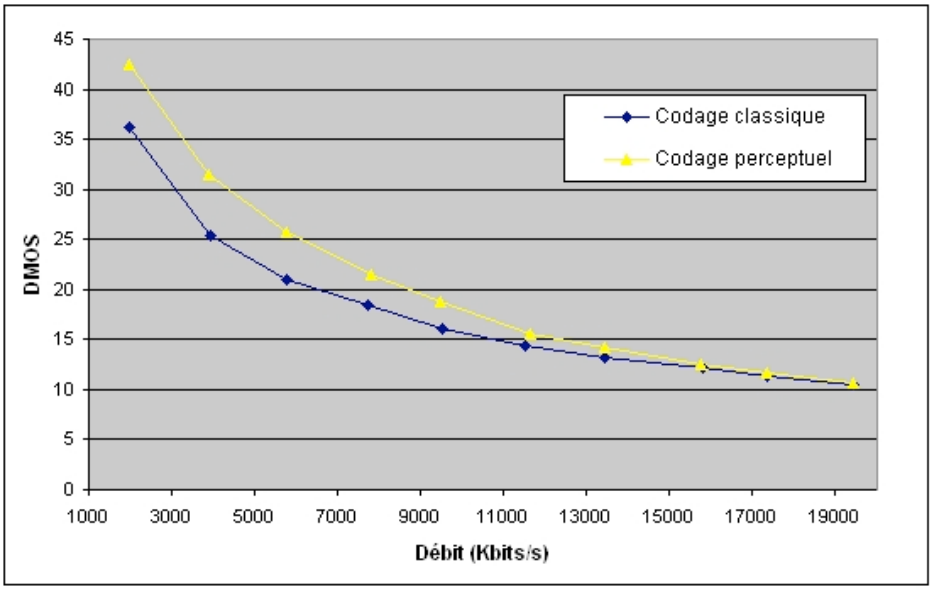

FIG. 3.11 – DMOS en fonction du débit pour la séquence *New mobile and Calendar*.

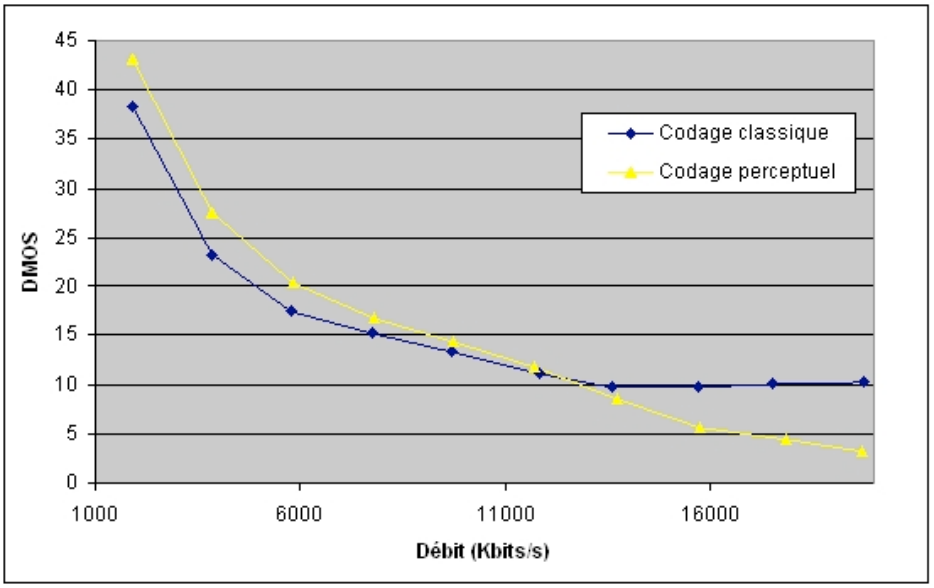

FIG. 3.12 – DMOS en fonction du débit pour la séquence *Knightshields*.

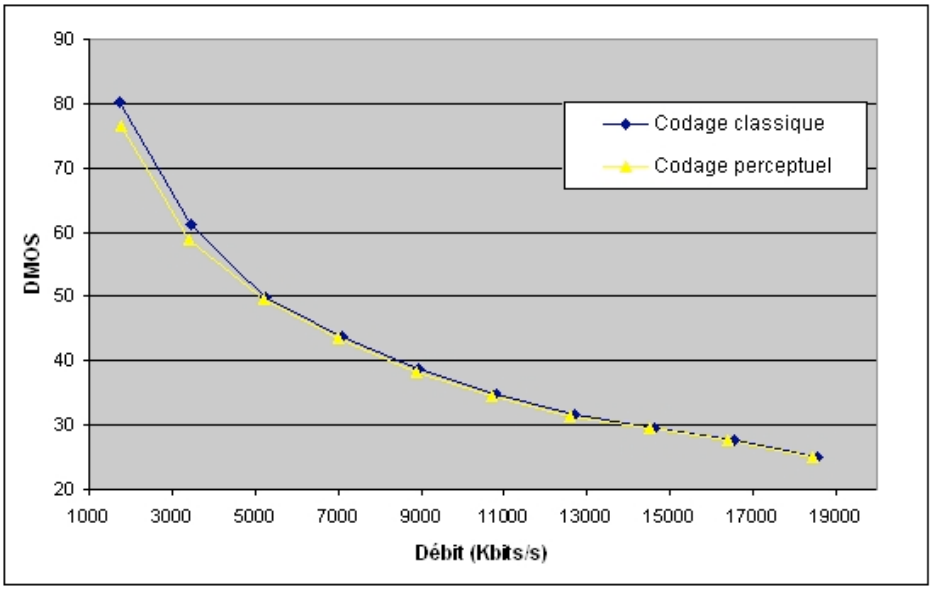

FIG. 3.13 – DMOS en fonction du débit pour la séquence *Parkrun*.

3.2.4 Qualité par image au cours du temps

Dans cette section, nous présentons les résultats en terme de PSNR et de DMOS au cours du temps (une valeur pour chaque image) pour les quatre séquences vidéo et pour trois débits différents.

Quelque soit la séquence vidéo et quelque soit le débit, on observe une perte de qualité pour les premières images de la séquence avec notre méthode de codage perceptuel. Cela semble indiquer, qu'un temps d'adaptation, de l'ordre d'une dizaine d'images, est nécessaire au codeur lorsqu'on lui transmet nos directives de codage pour adapter l'optimisation débit - distorsion. Ceci pénalisent donc nos résultats par rapport au codage classique.

3.2.4.1 Séquence *Tractor*

Les figures 3.14, 3.15 et 3.16 présentent les résultats en terme de PSNR au cours du temps pour trois débits cibles différents ($4000Kbits/s$, $10000Kbits/s$ et $20000Kbits/s$). Les figures 3.17, 3.18 et 3.19 présentent quant à elles les résultats en terme de DMOS au cours du temps pour ces trois débits.

Les résultats obtenus pour la séquence *Tractor* sont proches entre les deux approches de codage (classique et perceptuel). La différence notable est obtenue avec notre méthode de codage perceptuel pour un débit cible fixé à $4000Kbits/s$. En effet, on observe deux pics dans la courbe illustrant la qualité en terme de PSNR (cf figure 3.14) pour les images 184 et 389. Cela est dû au fait que le codeur a détecté des (faux) changements de scène et à coder ces images en Intra.

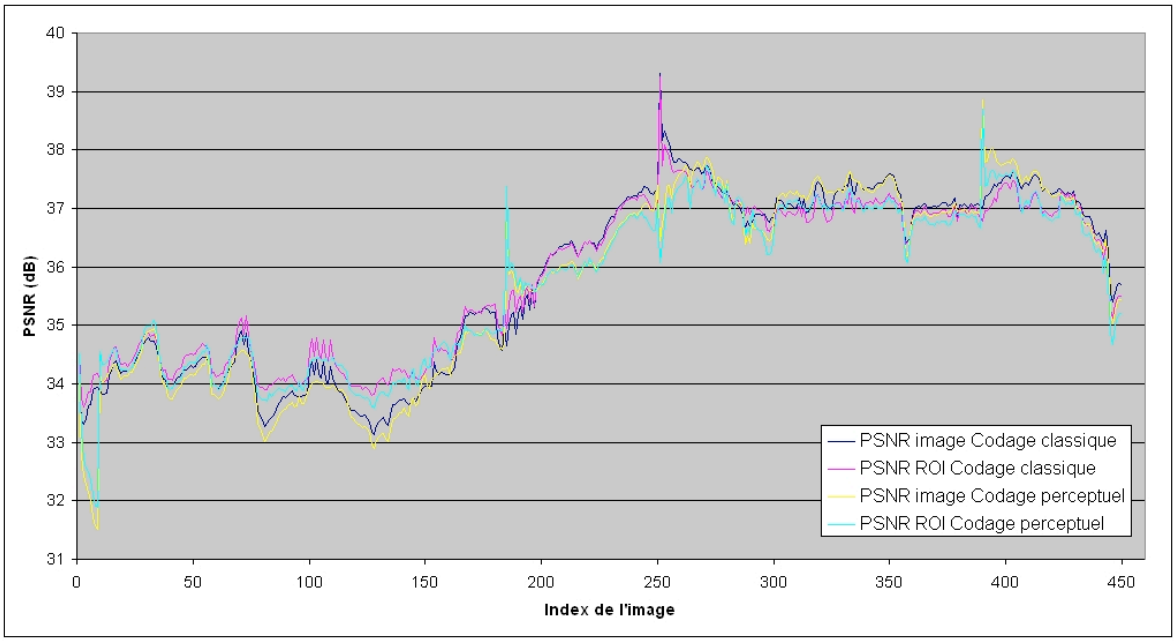

FIG. 3.14 – PSNR au cours du temps pour la séquence *Tractor* codée à *4000 Kbits/s*.

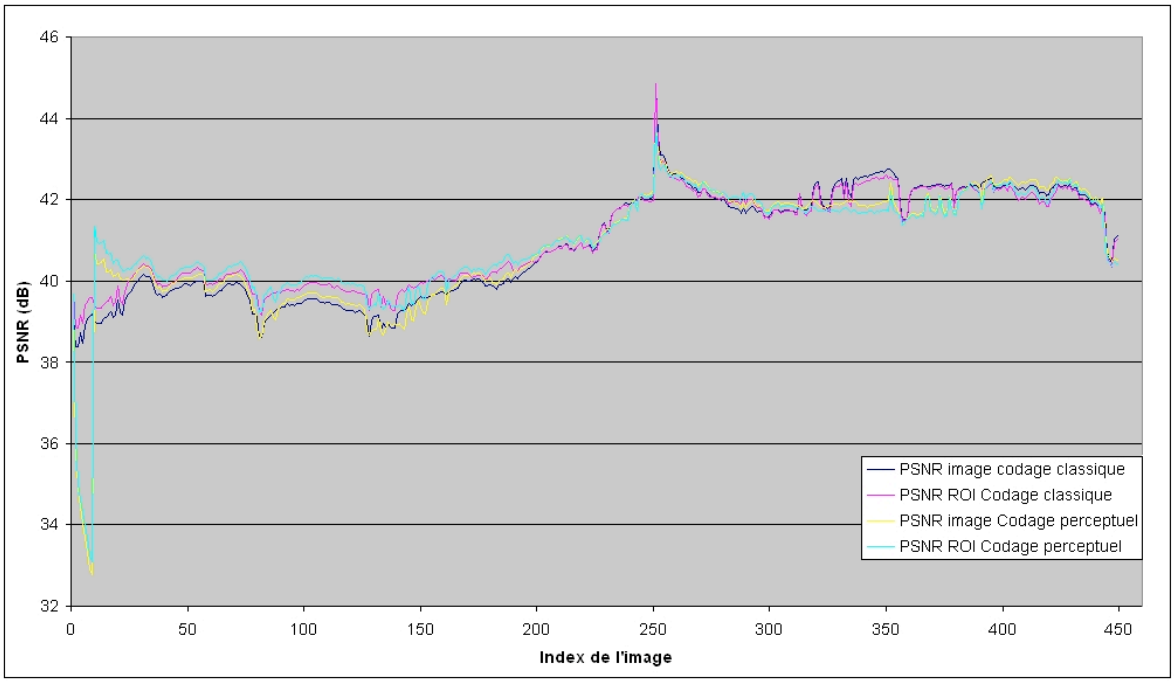

FIG. 3.15 – PSNR au cours du temps pour la séquence *Tractor* codée à *10000 Kbits/s*.

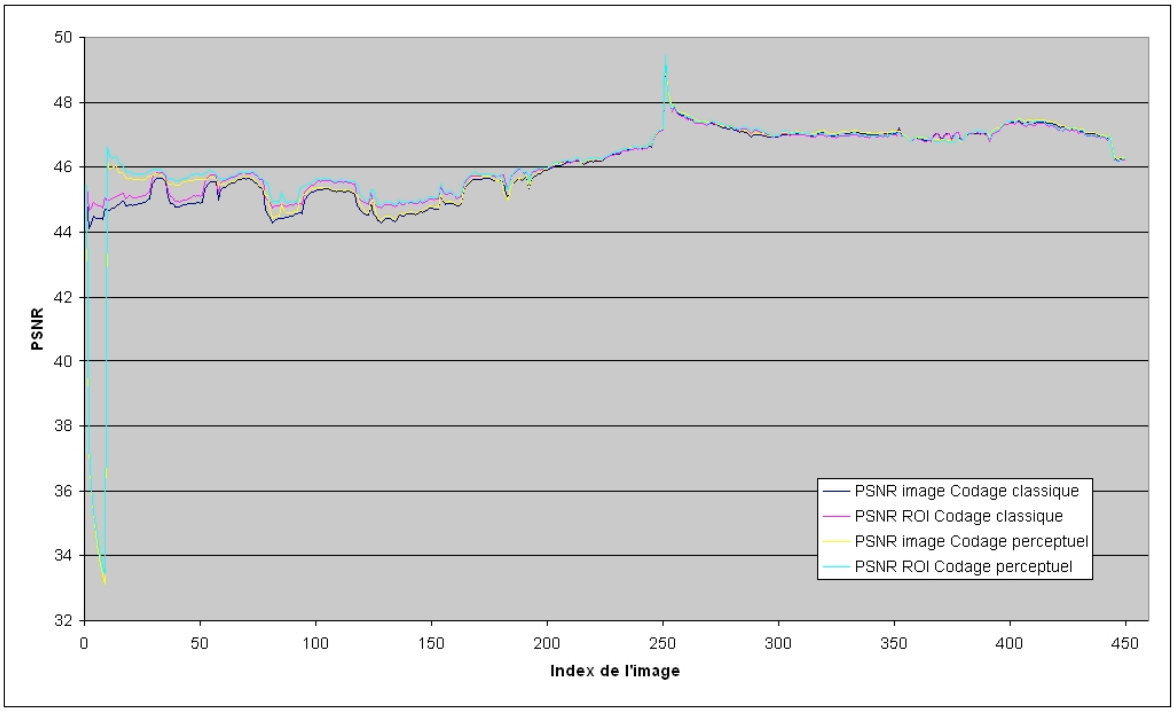

FIG. 3.16 – PSNR au cours du temps pour la séquence *Tractor* codée à *20000 Kbits/s*.

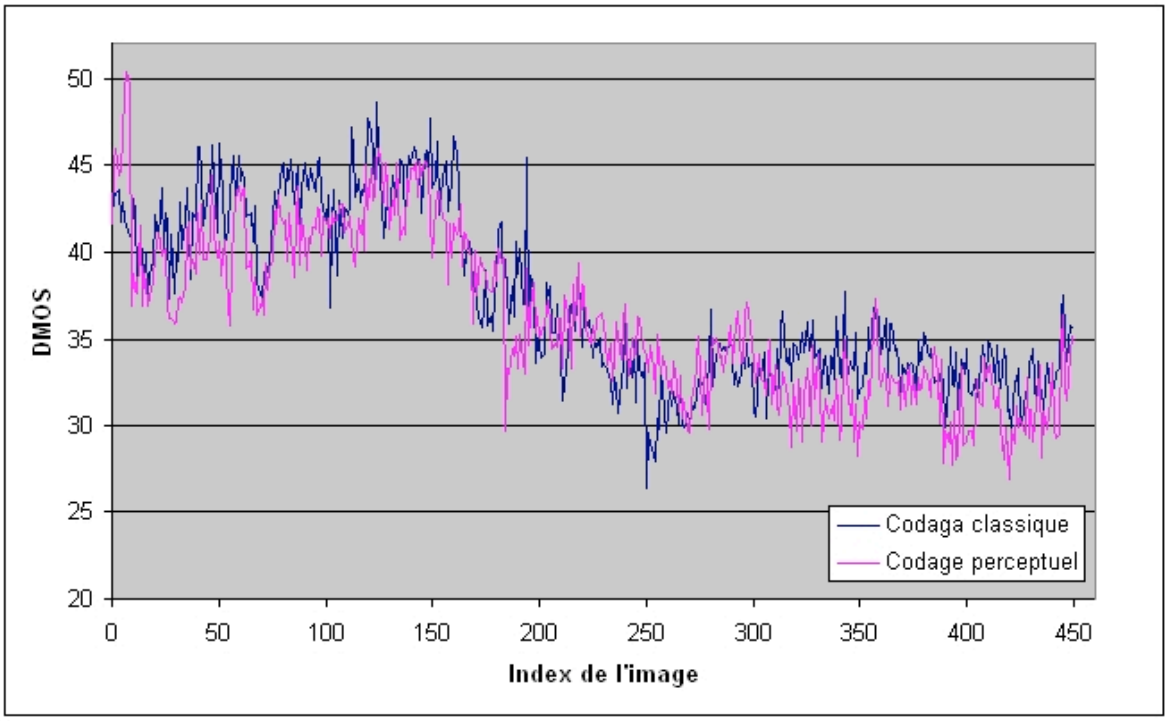

FIG. 3.17 – DMOS au cours du temps pour la séquence *Tractor* codée à *4000 Kbits/s*.

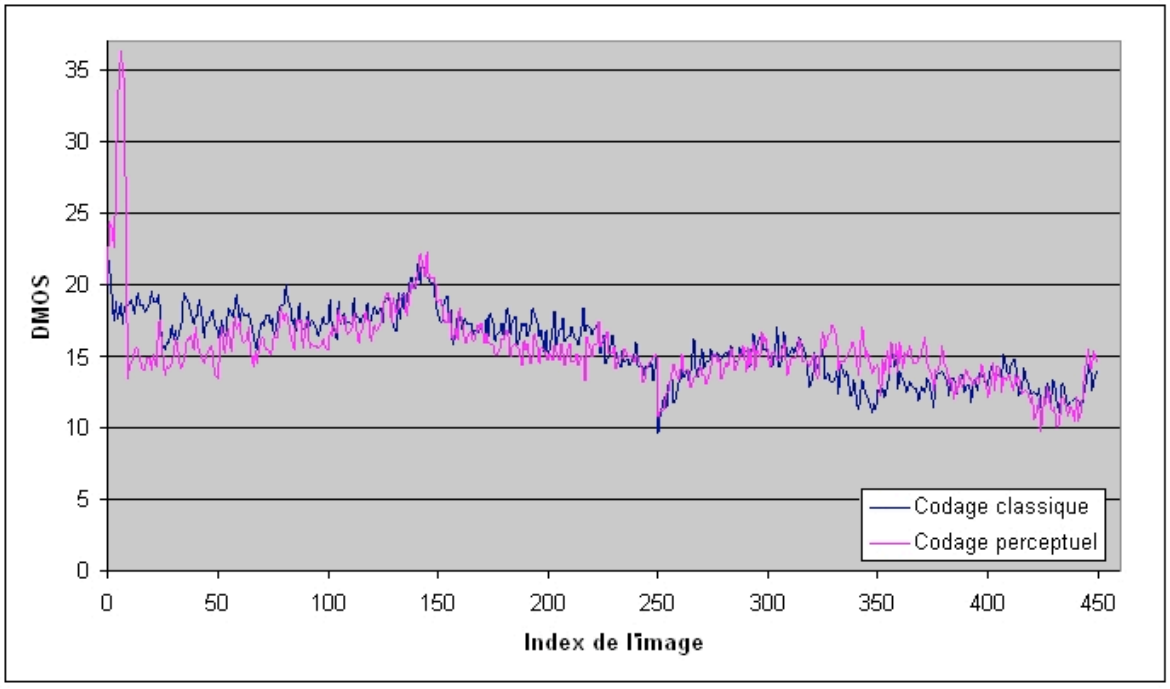

FIG. 3.18 – DMOS au cours du temps pour la séquence *Tractor* codée à *10000 Kbits/s*.

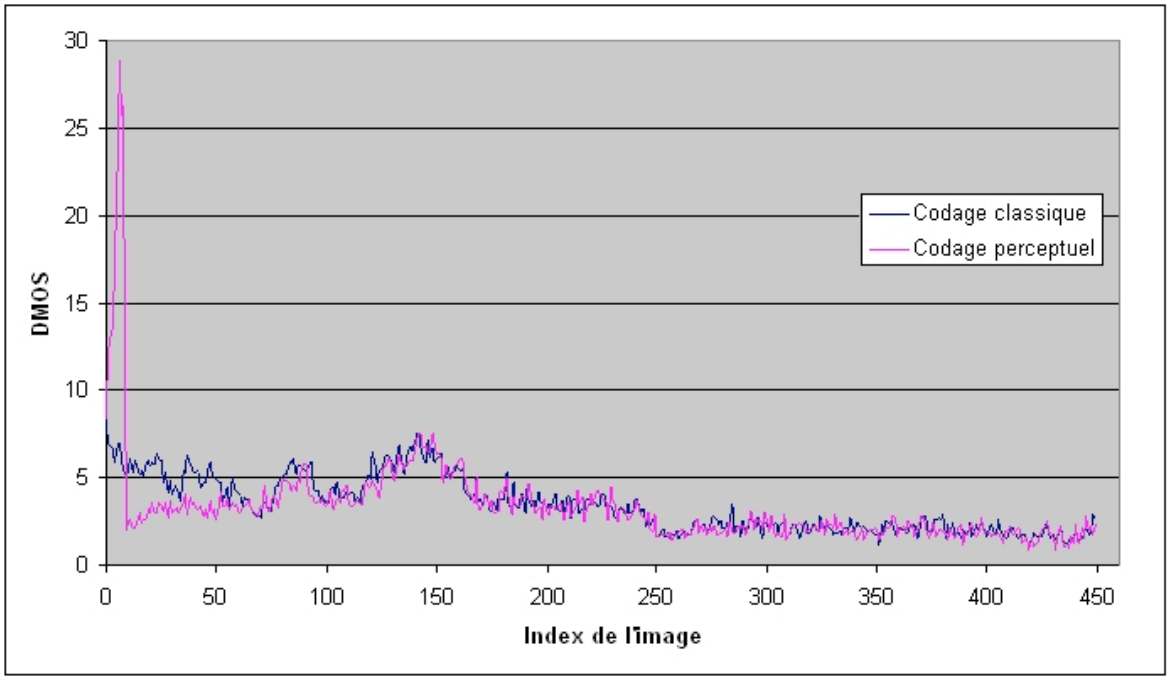

FIG. 3.19 – DMOS au cours du temps pour la séquence *Tractor* codée à *20000 Kbits/s*.

3.2.4.2 Séquence *New Mobile and Calendar*

Les figures 3.20, 3.21 et 3.22 présentent les résultats en terme de PSNR au cours du temps pour trois débits cibles différents ($2000 Kbits/s$, $10000 Kbits/s$ et $20000 Kbits/s$). Les figures 3.23, 3.24 et

3.25 présentent quant à elles les résultats en terme de DMOS au cours du temps pour ces trois débits.

Les résultats en terme de PSNR pour les deux approches de codage sont très similaires. Par contre, les résultats en terme de DMOS pour le débit cible fixé à $2000 Kbits/s$ (cf figure 3.23) sont différents. En effet, à partir de l'image 200, le DMOS obtenu avec notre méthode de codage perceptuel devient plus important que le DMOS obtenu avec le codage classique. Cela coïncide avec l'apparition dans la séquence d'un fond uniforme peu saillant.

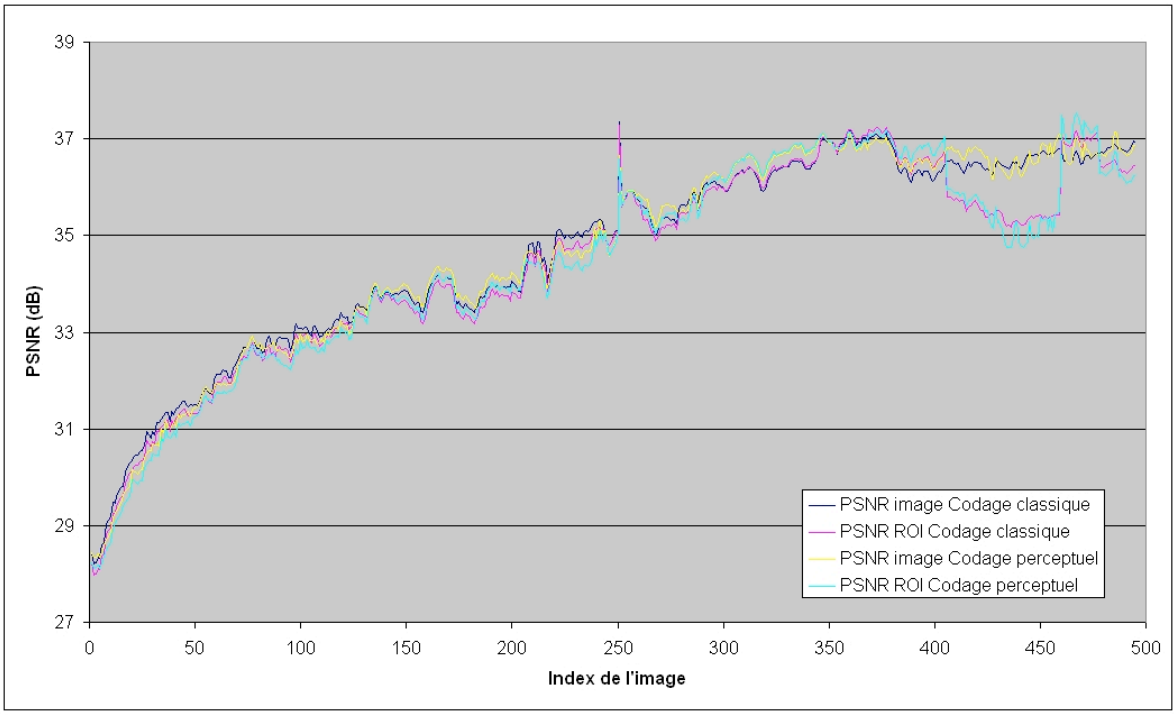

FIG. 3.20 – PSNR au cours du temps pour la séquence *New Mobile and Calendar* codée à *2000 Kbits/s*.

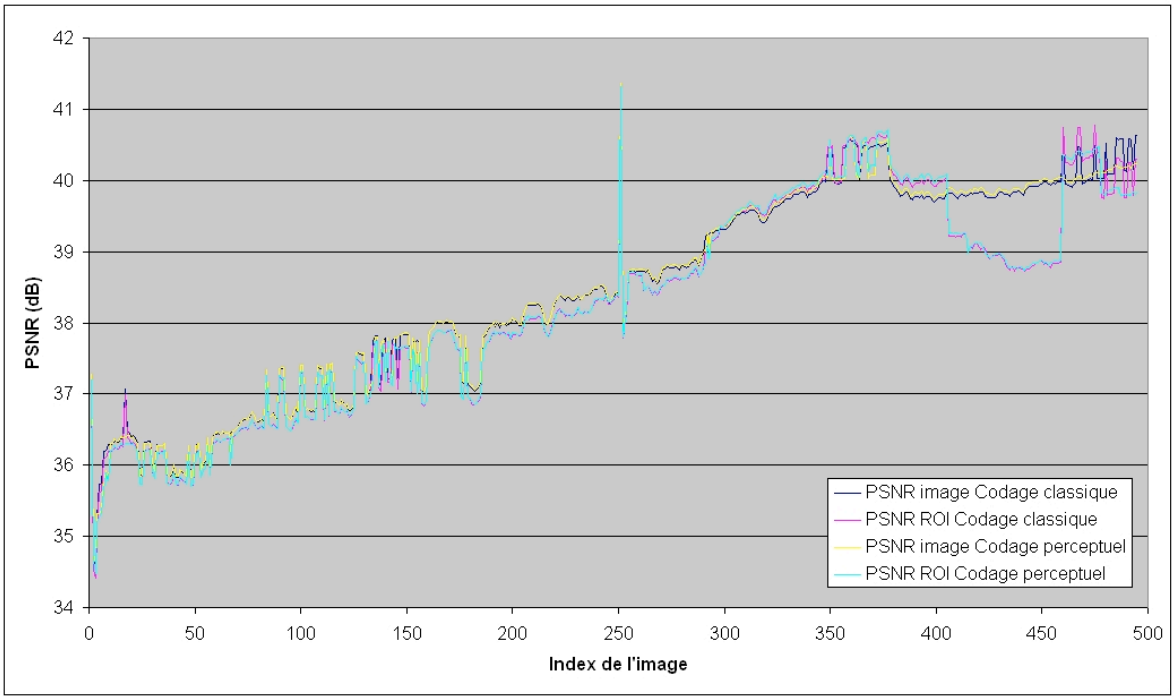

FIG. 3.21 – PSNR au cours du temps pour la séquence *New Mobile and Calendar* codée à *10000 Kbits/s*.

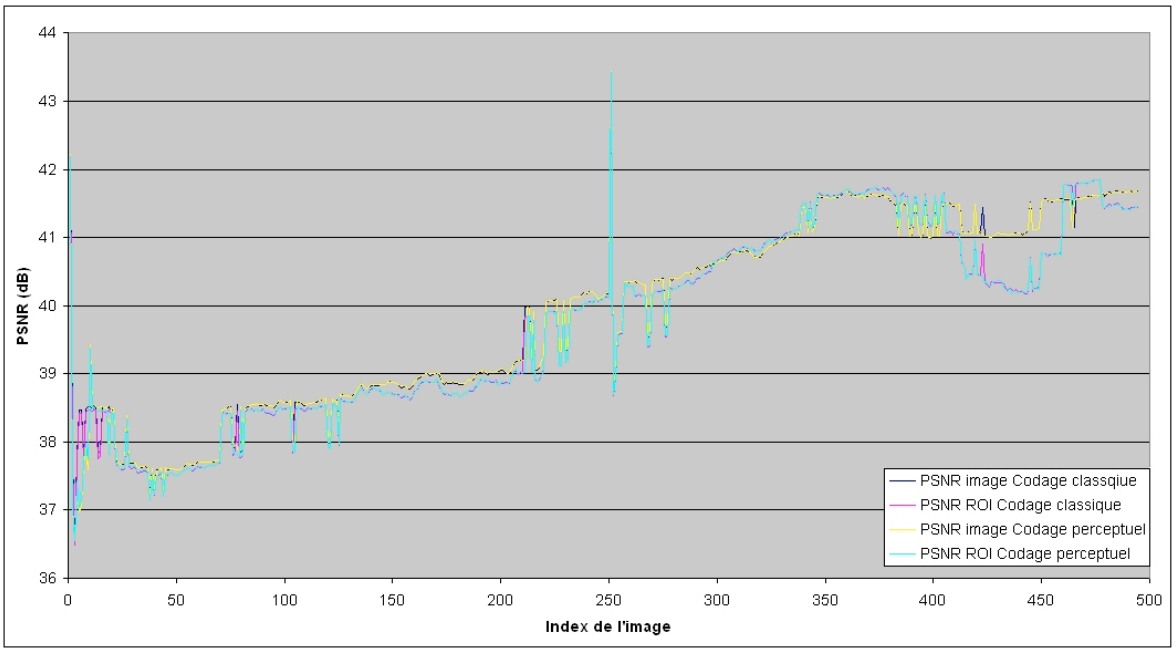

FIG. 3.22 – PSNR au cours du temps pour la séquence *New Mobile and Calendar* codée à *20000 Kbits/s*.

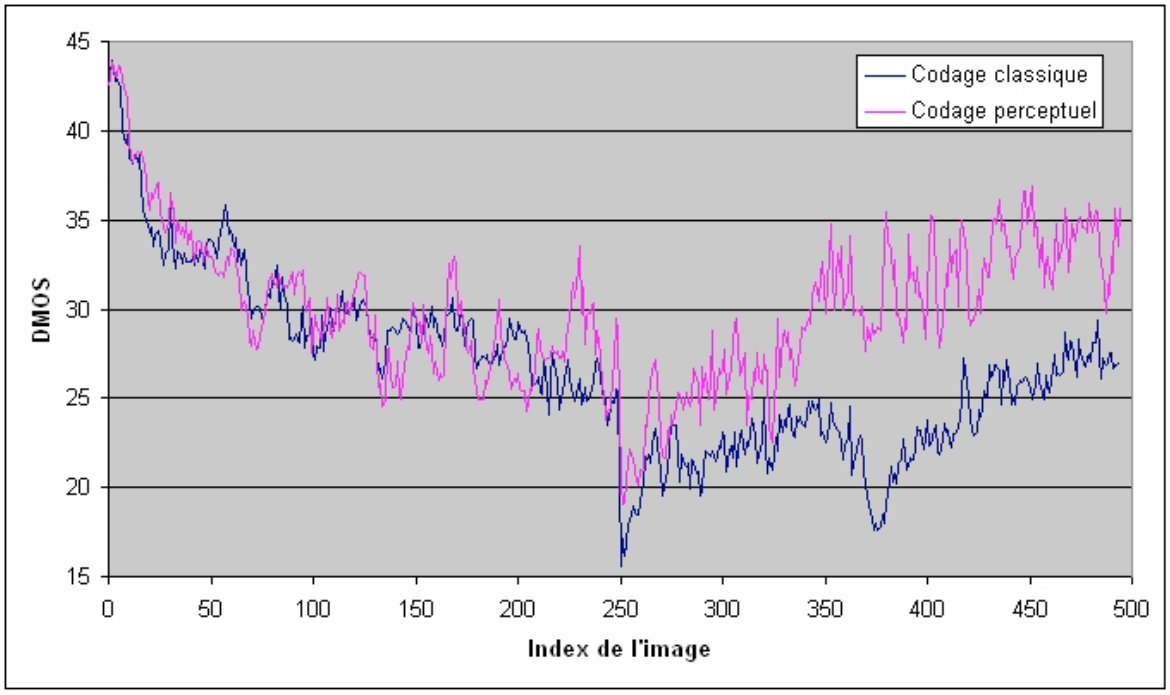

FIG. 3.23 – DMOS au cours du temps pour la séquence *New Mobile and Calendar* codée à *2000 Kbits/s.*

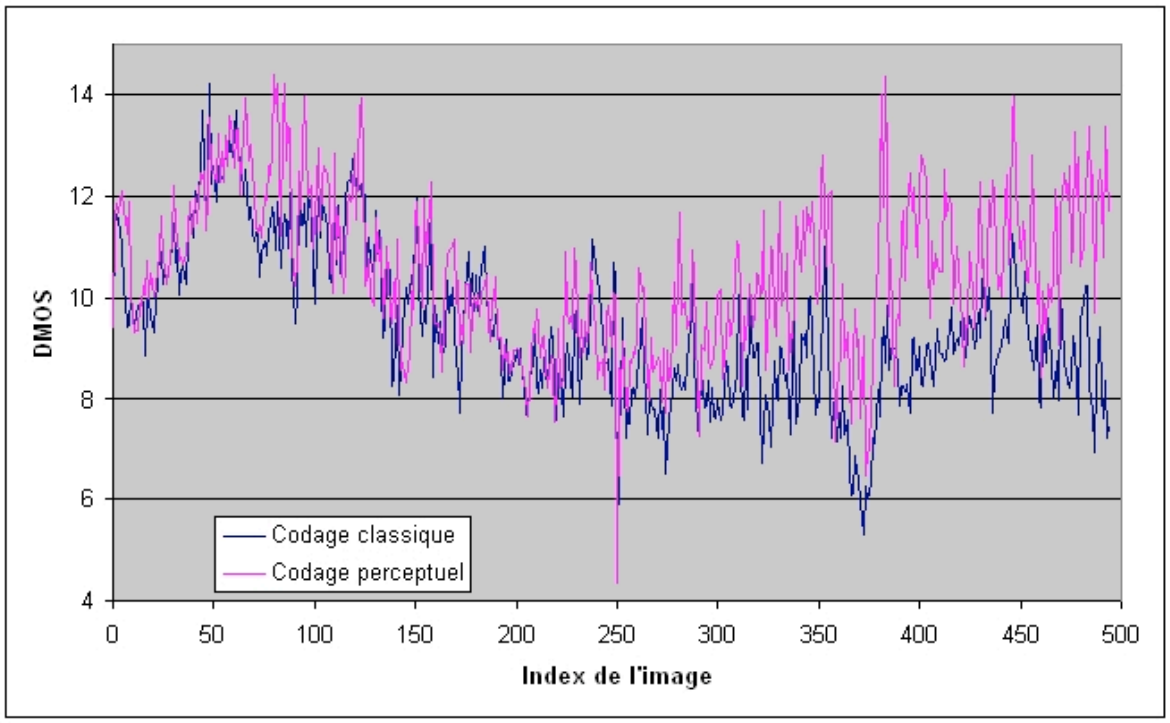

FIG. 3.24 – DMOS au cours du temps pour la séquence *New Mobile and Calendar* codée à *10000 Kbits/s.*

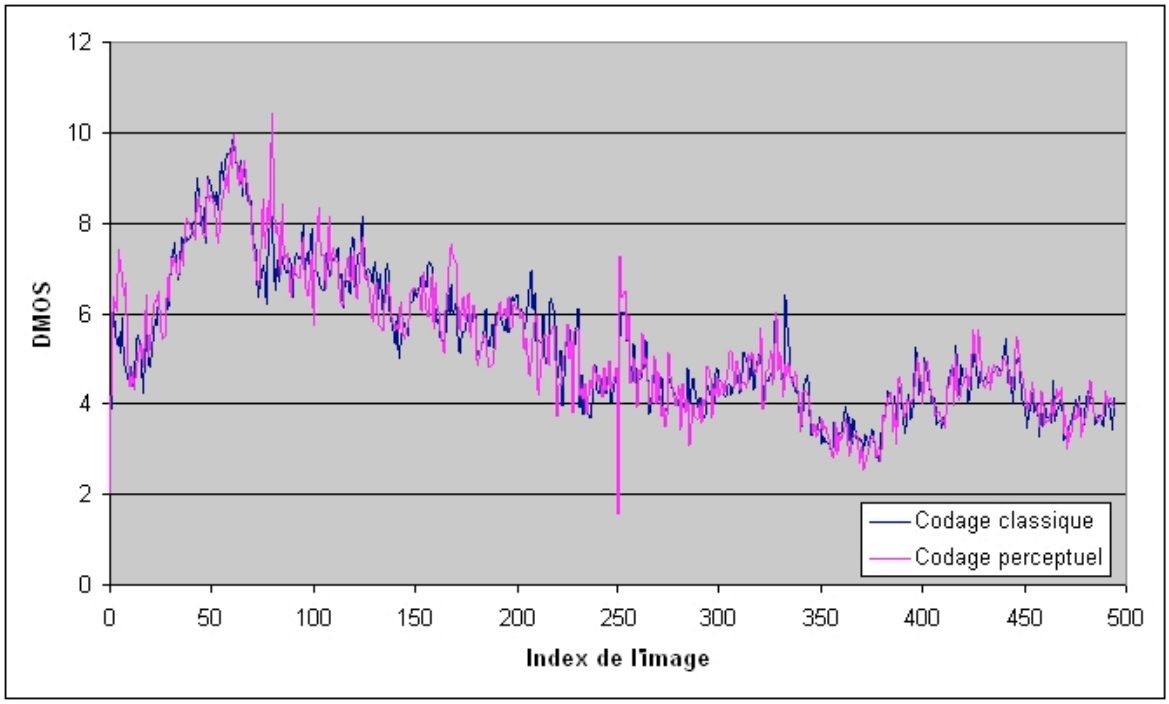

FIG. 3.25 – DMOS au cours du temps pour la séquence *New Mobile and Calendar* codée à *20000 Kbits/s*.

3.2.4.3 Séquence *Knightshields*

Les figures 3.26, 3.27 et 3.28 présentent les résultats en terme de PSNR au cours du temps pour trois débits cibles différents ($2000 Kbits/s$, $6000 Kbits/s$ et $14000 Kbits/s$). Les figures 3.29, 3.30 et 3.31 présentent quant à elles les résultats en terme de DMOS au cours du temps pour ces trois débits.

Les résultats obtenus en terme de PSNR illustrent le bon comportement de notre méthode de codage perceptuel lorsque que le débit augmente ($6000 Kbits/s$ et $14000 Kbits/s$). La figure 3.31 illustre les résultats en terme de DMOS et indique que notre méthode de codage perceptuel semble être plus performante sur la première partie de la vidéo. Dans la deuxième partie de la vidéo, l'homme s'immobilise ainsi que la caméra qui le suivait. La scène reste immobile quelques instants avant qu'un zoom avant de la caméra ne soit réalisé. Ces conditions semblent être favorables à une approche de codage classique plutôt qu'à notre méthode de codage perceptuel qui utilise une carte de saillance basée en partie sur le mouvement relatif des objets présents dans la séquence vidéo.

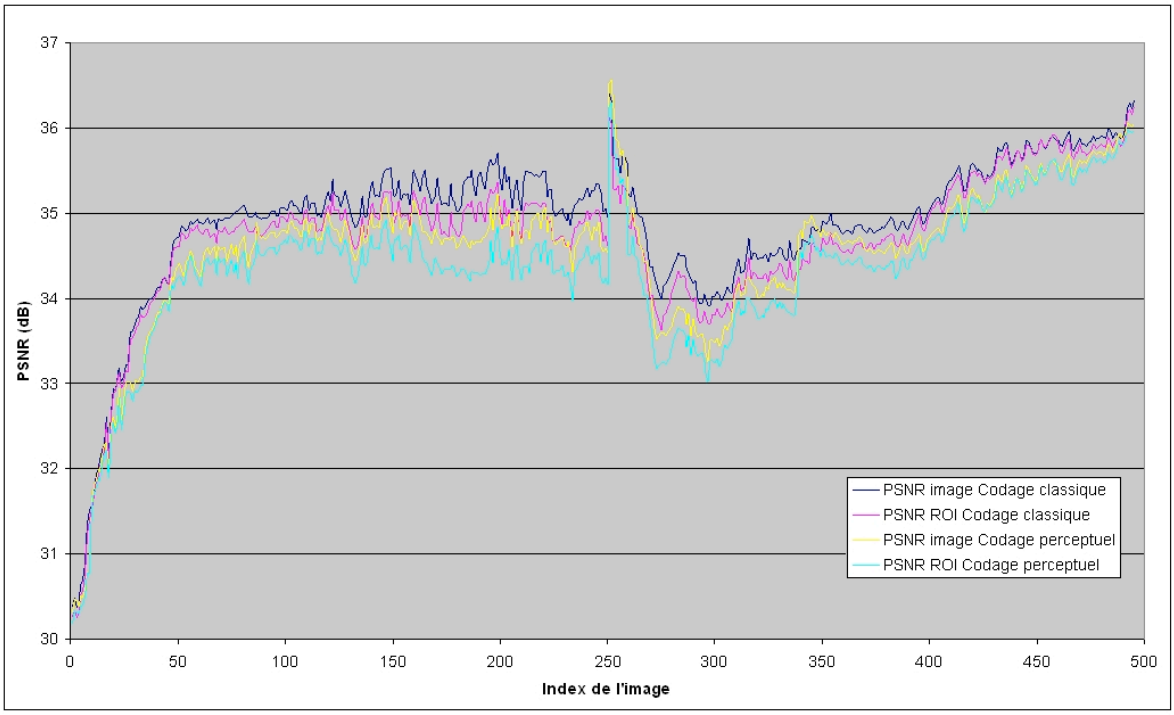

FIG. 3.26 – PSNR au cours du temps pour la séquence *Knightshields* codée à *2000 Kbits/s*.

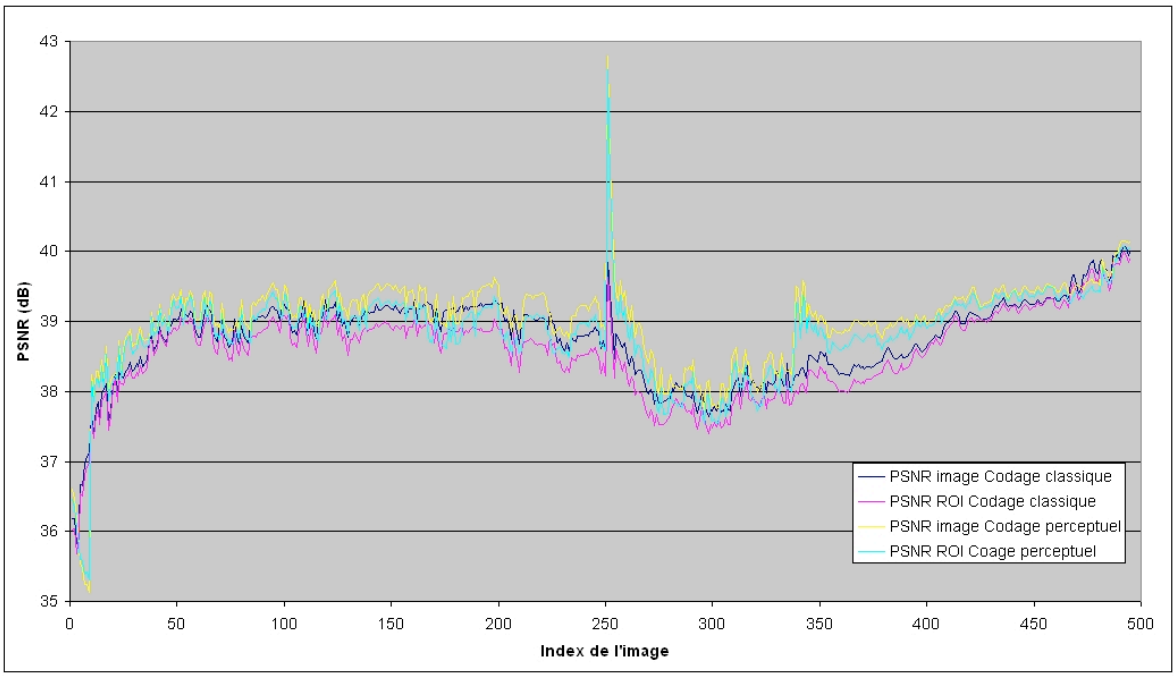

FIG. 3.27 – PSNR au cours du temps pour la séquence *Knightshields* codée à *6000 Kbits/s*.

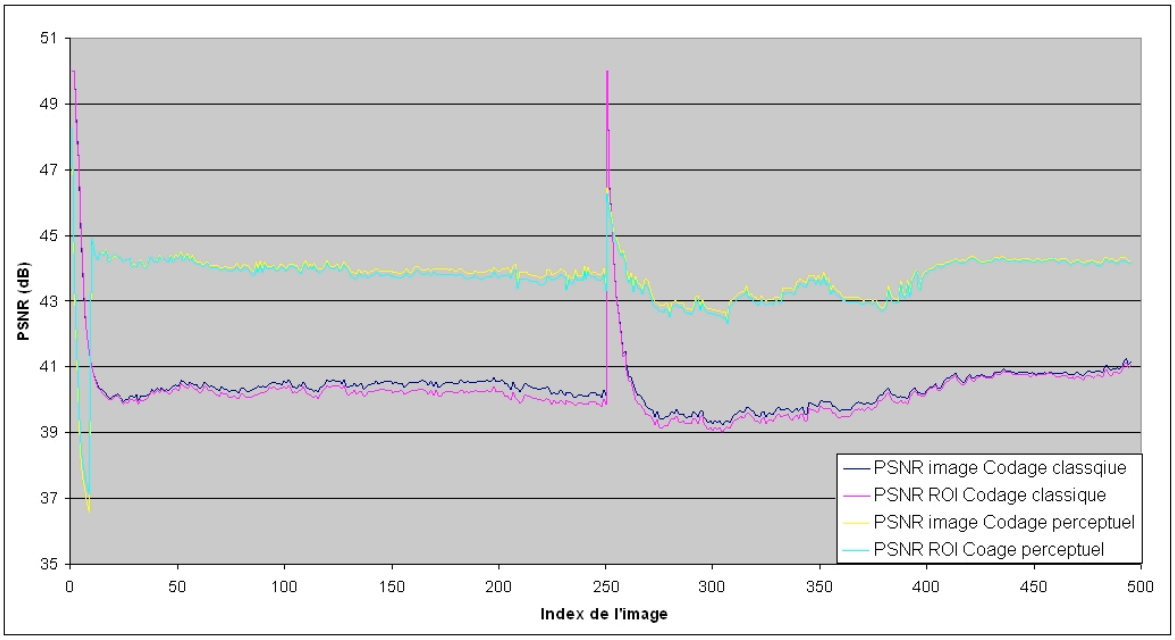

FIG. 3.28 – PSNR au cours du temps pour la séquence *Knightshields* codée à *14000 Kbits/s*.

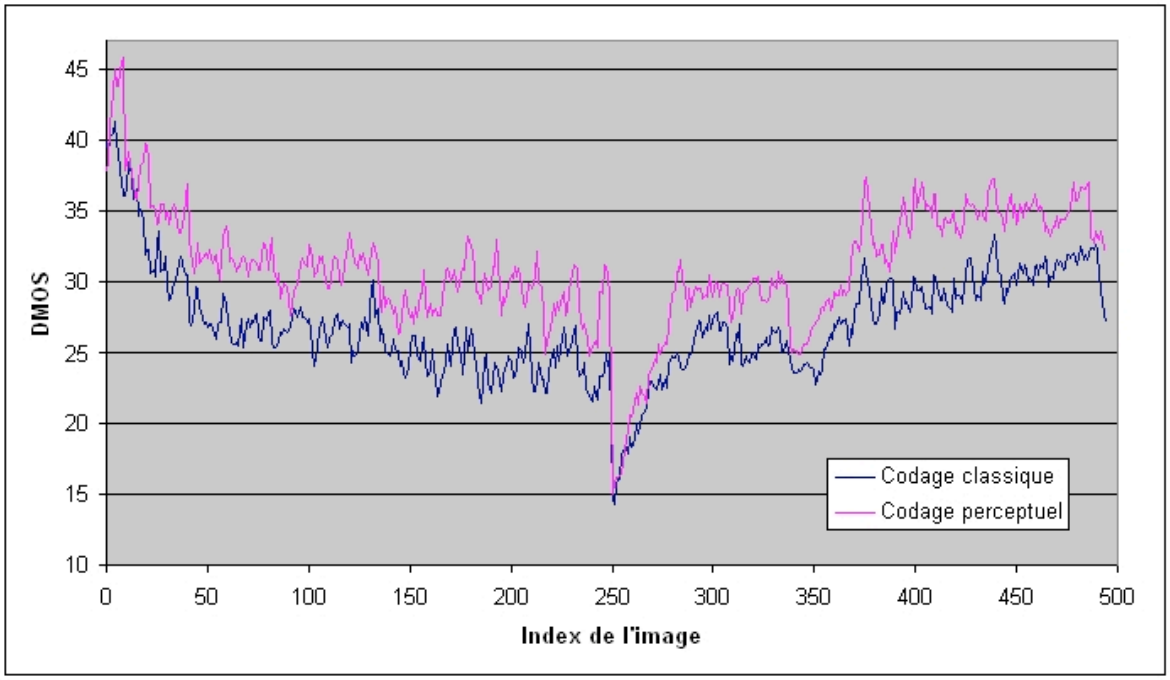

FIG. 3.29 – DMOS au cours du temps pour la séquence *Knightshields* codée à *2000 Kbits/s*.

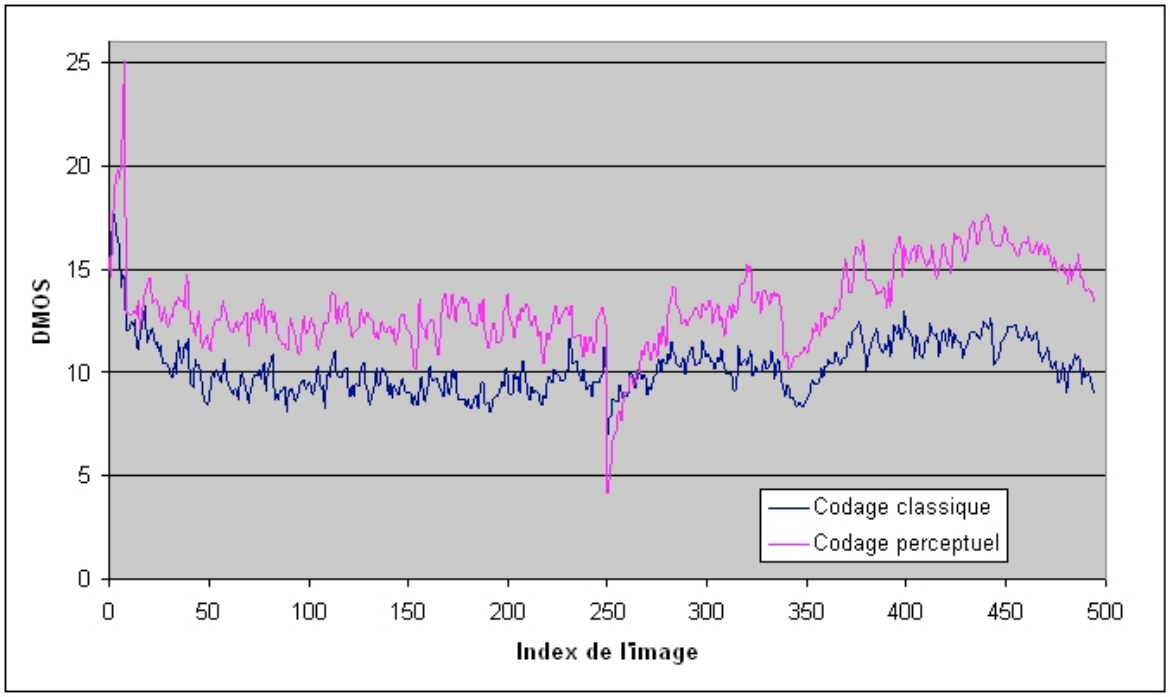

FIG. 3.30 – DMOS au cours du temps pour la séquence *Knightshields* codée à *6000 Kbits/s*.

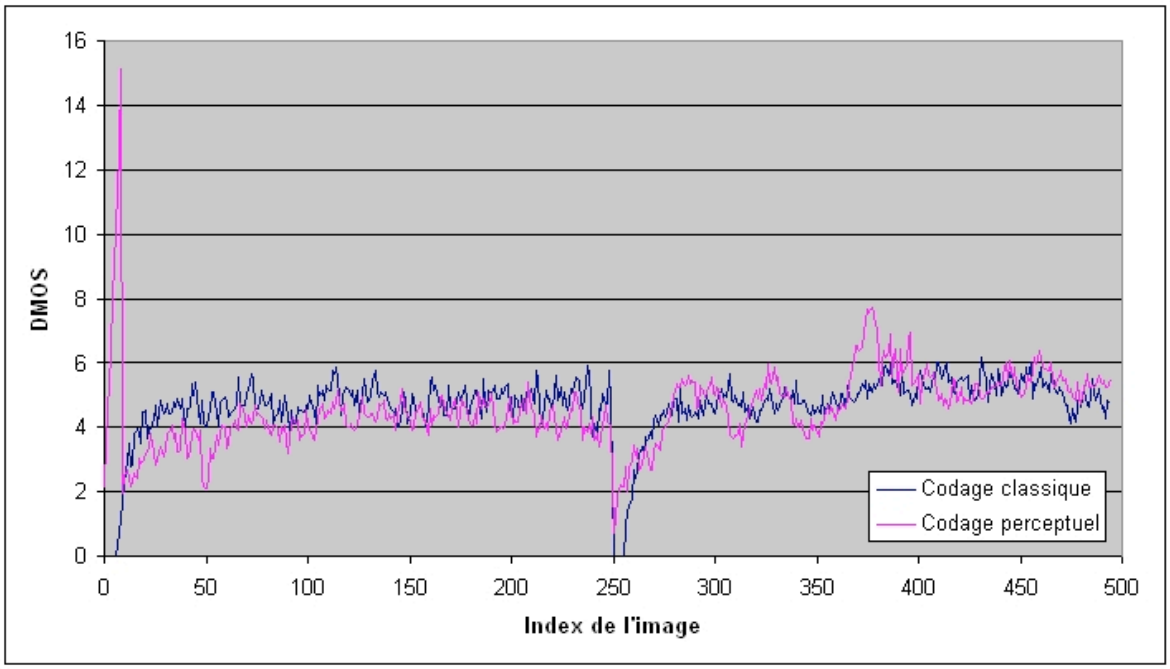

FIG. 3.31 – DMOS au cours du temps pour la séquence *Knightshields* codée à *14000 Kbits/s*.

3.2.4.4 Séquence *Parkrun*

Les figures 3.32, 3.33 et 3.34 présentent les résultats en terme de PSNR au cours du temps pour trois débits cibles différents ($4000 Kbits/s$, $10000 Kbits/s$ et $20000 Kbits/s$). Les figures 3.35, 3.36 et

3.37 présentent quant à elles les résultats en terme de DMOS au cours du temps pour ces trois débits.

Les résultats obtenus illustrent un meilleur comportement de notre méthode de codage perceptuel en terme de qualité pour la première partie de la séquence vidéo. En effet, à la fin de la séquence vidéo, la scène devient immobile (l'homme s'immobilise et les mouvements de caméra deviennent nuls) ce qui est plus propice au codage classique.

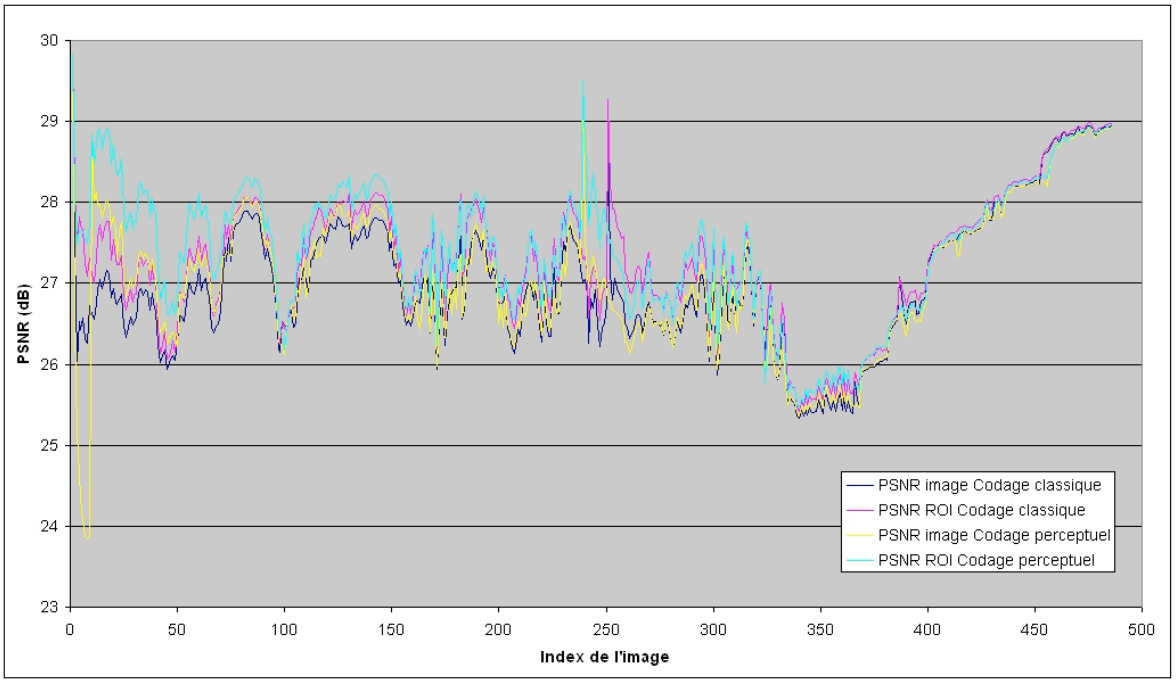

FIG. 3.32 – PSNR au cours du temps pour la séquence *Parkrun* codée à *4000 Kbits/s*.

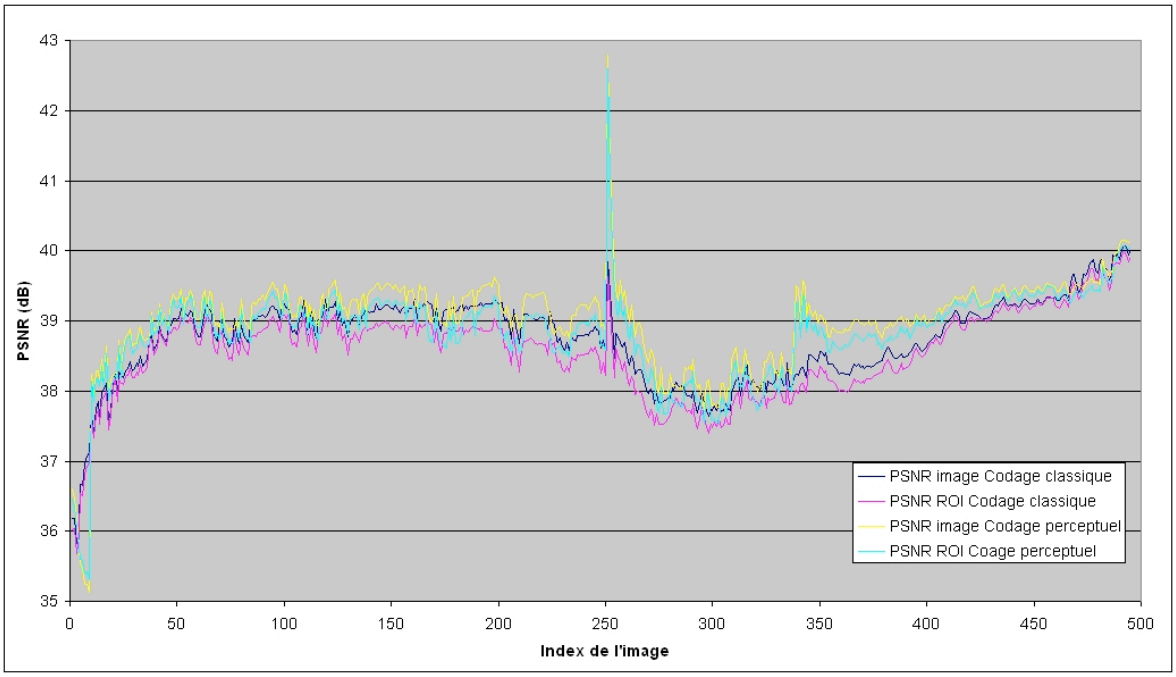

FIG. 3.33 – PSNR au cours du temps pour la séquence *Parkrun* codée à *10000 Kbits/s*.

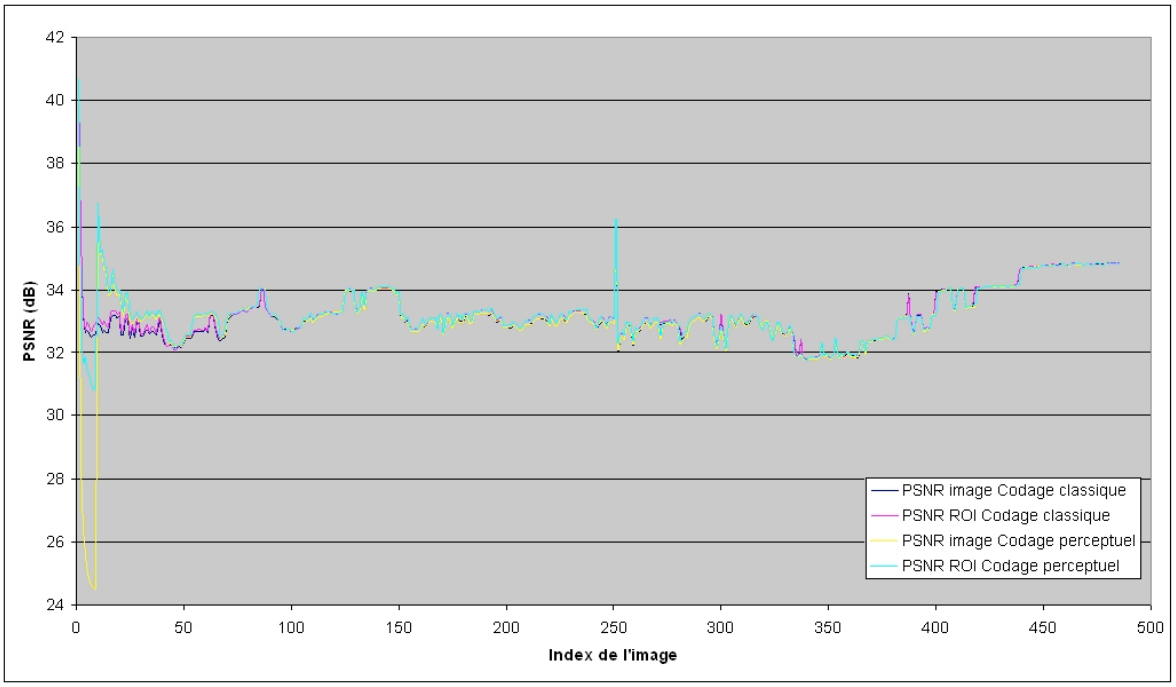

FIG. 3.34 – PSNR au cours du temps pour la séquence *Parkrun* codée à *20000 Kbits/s*.

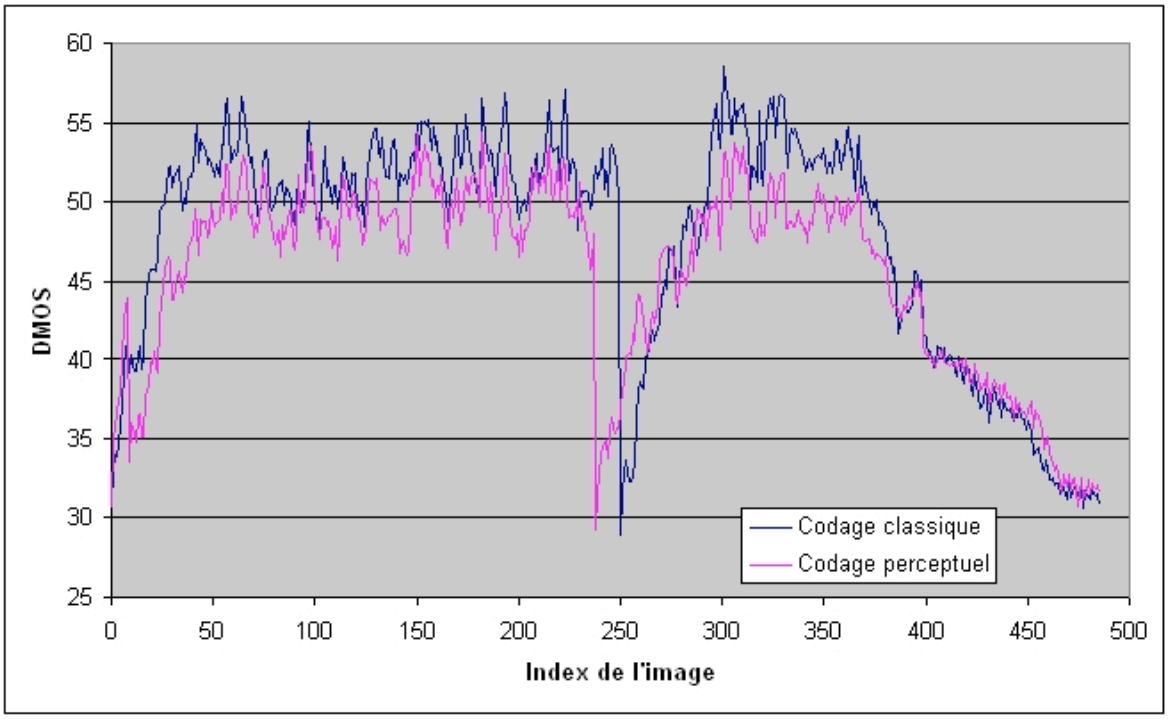

FIG. 3.35 – DMOS au cours du temps pour la séquence *Parkrun* codée à *4000 Kbits/s*.

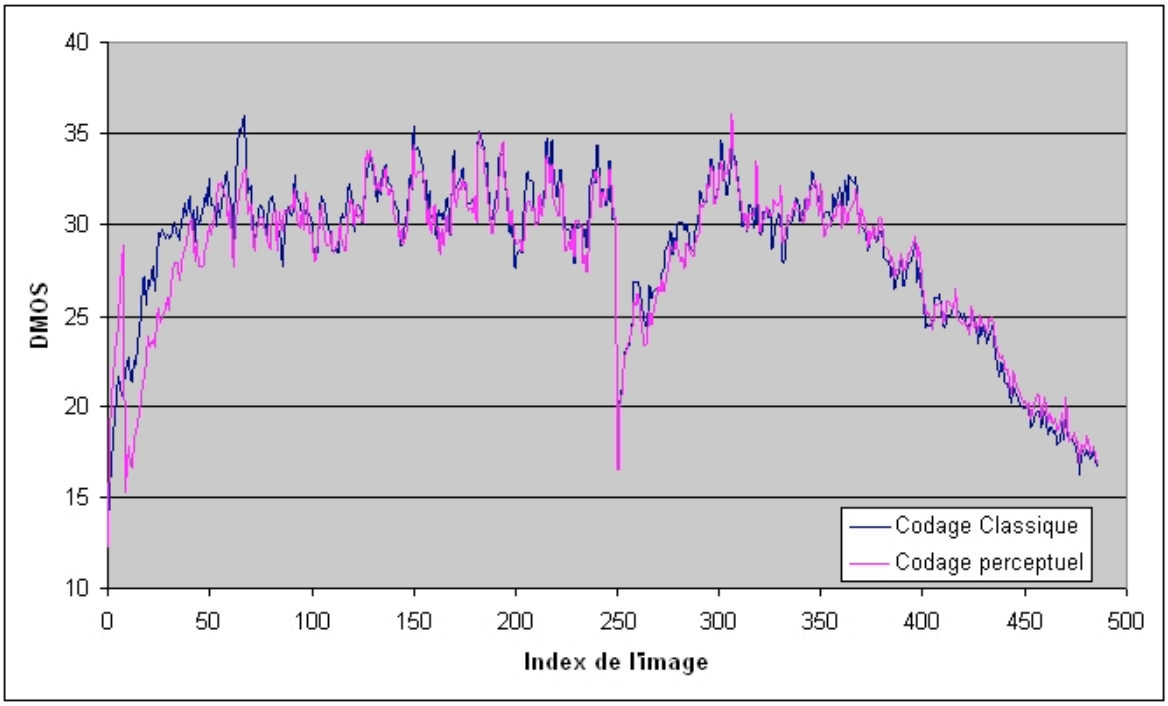

FIG. 3.36 – DMOS au cours du temps pour la séquence *Parkrun* codée à *10000 Kbits/s*.

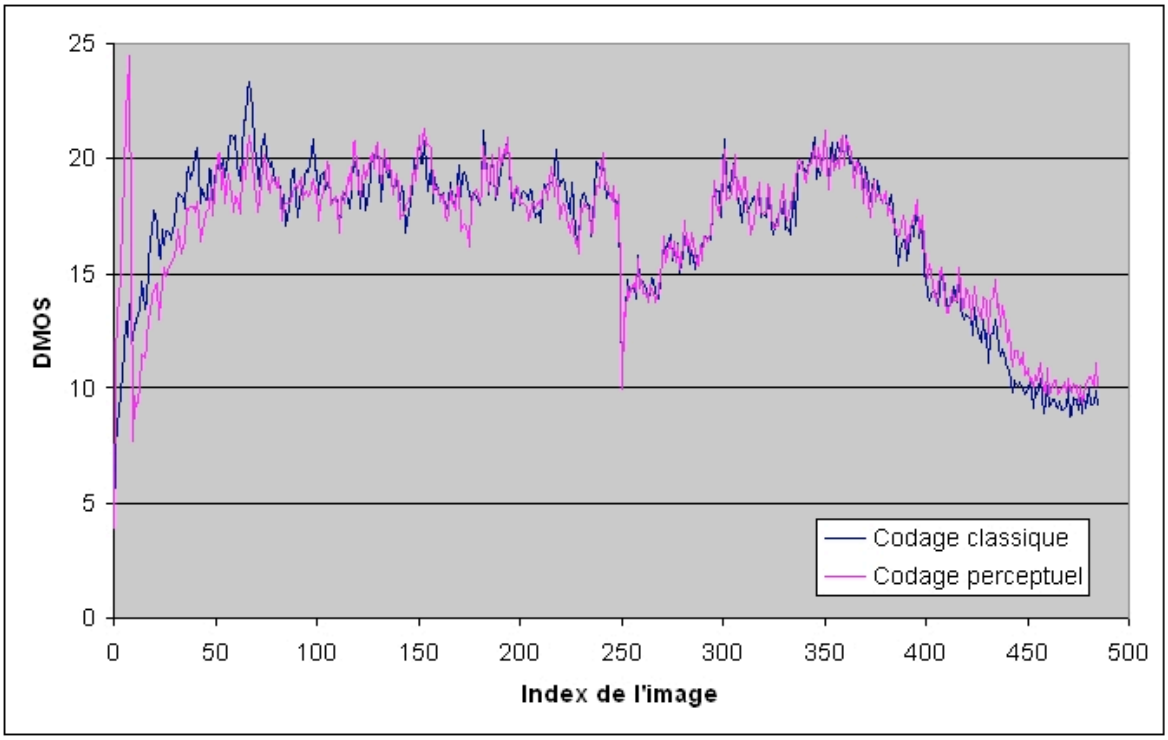

FIG. 3.37 – DMOS au cours du temps pour la séquence *Parkrun* codée à *20000 Kbits/s*.

3.2.5 Résultats qualitatifs

Les résultats sont ici obtenus via une observation directe des vidéos en salle de test dont l'environnement est normalisée, l'écran est full HD (1920 x 1080 points) et observé à une distance de 3H (c.a.d 3 fois la hauteur de l'écran).

Les figures 3.38 et 3.39 présentent une partie de l'image 495 extraite de la séquence *New Mobile & Calendar,* et obtenues avec une approche de codage classique ou notre méthode de codage perceptuel. La consigne de débit était fixée à $2000 Kbits/s$ dans les deux cas. Nous observons des effets de blocs au niveau du fond uniforme pour la séquence obtenue avec le codage classique (cf figure 3.38). Lors de la lecture de la séquence vidéo obtenue via le codage classique, on observe aussi un effet de papillotement dû à ces effets de blocs. Ce papillotement est très gênant et nuit à la qualité perçue. Alors que pour la séquence vidéo obtenue avec notre méthode de codage perceptuel, cet effet de papillotement n'est pas perceptible. La qualité semble alors être supérieure avec notre méthode de codage perceptuel, contrairement aux résultats obtenus avec les métriques utilisées précédemment. En effet, le DMOS obtenu avec le codage classique est de 36.3 contre 42.5 pour notre méthode de codage perceptuel. De plus, cet effet de papillotement apparaît encore sur le fond uniforme présent dans la deuxième partie de la séquence vidéo. Or si l'on observe les résultats de la figure 3.23, le DMOS obtenu avec le codage classique est nettement inférieur à celui obtenu avec notre méthode de codage perceptuel pour la deuxième partie de la séquence vidéo (à partir de l'image 250). Bien qu'étant corrélés avec le jugement humain de qualité visuelle, les résultats obtenus avec VQA pour ces deux séquences ne semblent pas être corrects, puisque visuellement la séquence vidéo codée avec notre méthode de codage perceptuel semble de meilleure qualité que celle obtenue avec le codage classique.

Ces effets de papillotement sont également visibles pour la séquence vidéo *Knightshields* encodée via le codage classique pour un débit cible de $2000 Kbits/s$. Ces défauts ont pour effet de nuire à la qualité perçue. Or les notes de qualité en terme de DMOS pour la séquence *Knightshields* à $2000 Kbits/s$ sont de $38,3$ pour le codage classique et de $43,1$ pour le codage perceptuel. Ceux-ci ne sont pas illustrés dans ce rapport, car ils ne sont vraiment visibles que lors de la visualisation de la séquence vidéo.

FIG. 3.38 – Zoom sur une partie d'image de la séquence *New Mobile & Calendar* obtenue avec le codage classique à *1980 Kbits/s*.

FIG. 3.39 – Zoom sur une partie d'image de la séquence New *Mobile & Calendar* obtenue avec le codage perceptuel à *1973 Kbits/s*.

3.3 Conclusion

Ce chapitre a présenté les résultats en terme de qualité pour une approche de codage classique et notre méthode de codage perceptuel. Pour ce type de compression, c'est le cœur et la stratégie de codage qui sont modifiés. L'objectif est d'améliorer la qualité perçue comparativement à une approche classique de codage.

Nous avons utilisé deux métriques de qualité pour évaluer notre méthode. La première métrique est le PSNR, celle-ci est simple à mettre en œuvre mais est faiblement corrélée avec les résultats de tests subjectifs de qualité visuelle. La deuxième métrique utilisée est issue du logiciel VQA2 qui note la qualité visuelle d'une vidéo en calculant le DMOS par rapport à la vidéo de référence. Cette métrique est très corrélée avec les tests subjectifs de qualité visuelle et permet d'obtenir des résultats proches de ceux obtenus avec des observateurs.

Les résultats obtenus diffèrent entre les deux métriques, illustrant parfaitement la faible corrélation du PSNR avec les tests subjectifs de qualité visuelle. Notre approche de compression sélective directe obtient de meilleurs résultats en moyenne à haut débit. Cependant pour les débits dédiés à la télévision Haute Définition (6 à 13$Mbits$ par seconde), les résultats obtenus ne sont pas satisfaisants.

Des tests subjectifs devront être effectués afin de confirmer l'intérêt ou non d'une compression sélective directe et d'améliorer notre modèle.

Conclusion

Ce rapport présente les résultats de l'outil de pré-analyse de flux vidéo haute définition en vue d'un encodage en temps réel sous le standard H.264. L'objectif est donc de fournir au codeur H.264 un jeu de paramètres adapté au codage d'une séquence vidéo et présentant une cohérence spatio-temporelle fonction des objets présents dans la scène. Le premier chapitre a présenté les séquences vidéos Haute Définition utilisées lors des tests. Le deuxième chapitre a présenté les métriques de qualité utilisées lors de nos tests. La première est le PSNR dont l'avantage est sa simplicité de ce fait son déploiement. Le principal défaut de cette métrique étant d'être faiblement corrélée avec les tests visuels de qualité. La deuxième métrique utilisée est une solution de l'état l'art basée sur le comportement du système visuel humain. Cette métrique produit des évaluations de qualité qui sont fortement corrélées avec le jugement humain de qualité visuelle. Le dernier chapitre présente les résultats de codage entre une approche de codage classique et notre méthode de codage perceptuel.

Les premiers résultats de codage montrent un gain en qualité pour les débits supérieurs à $15Mbits$ par seconde. Cependant les débits dédiés à la télévision Haute Définition seront sûrement inférieurs à $15Mbits$ par seconde. Ces premiers résultats sont encourageant même s'ils restent insuffisants pour le moment. Plusieurs stratégies de codage adaptatives sont envisagées dans la suite du travail (cf dernière année thèse d'Olivier Brouard) :

- choix des images I,
- fréquence des images bi-prédictives (B) entre images I et P,
- amélioration de la quantification adaptative.

Bibliographie

[1] The SVT High Definition Multi Format Test Set, SVT corporate technology, février 2006, ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/SVT_MultiFormat_v10.pdf

[2] http ://www.acceptv.com/