

Réponses aux questions des experts

Nous remercions les experts pour leur relecture et commentaires sur notre article. Nous répondons à leurs différentes questions ci-dessous.

Commentaires de l'expert 1

Cet article est un extrait du papier [3] paru en mars 2017 dans *Signal Processing : Image Communication*.

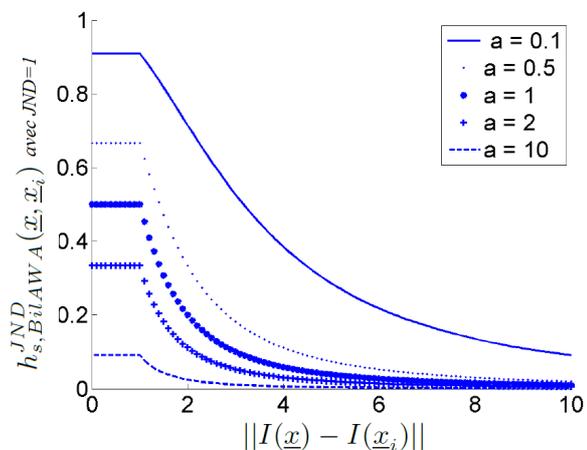
Il est intéressant et la démarche expérimentale est relativement bien expliquée. Les résultats sont probants et bien expliqués.

- Il faudrait cependant donner des compléments d'information pour l'équation (1): Comment est ajusté le paramètre a ?

Le paramètre a contrôle la vitesse de décroissance des poids du support de filtrage. Plus le paramètre a est faible et plus rapidement les pixels différents du pixel courant seront rejetés du support de filtrage (Cf figure ci-à-côté). Ainsi le paramètre a permet de régler la force de filtrage, plus a est faible et plus la force de filtrage est importante.

Le paramètre a a été fixé à 1 comme le suggèrent les auteurs du filtre AWA.

Cette information a été ajoutée à l'article, section 2 sous les équations 1 à 3.



Influence du paramètre a sur le noyau photométrique du filtre BilAWA

- Il serait pertinent de justifier le choix du modèle JND spatial décrit par [9] et expliquer pourquoi il est seulement basé sur la luminance. Pourquoi ne pas retenir un modèle JND basé sur les espaces couleur CIELAB ou CIELUV qui ont été conçus pour cela?

Le modèle JND décrit par Yang a été choisi pour le bon compromis qu'il offre entre résultat et complexité. Ces travaux ayant été réalisés dans le cadre d'un partenariat industriel, nous avons choisi un modèle pouvant être implémenté en temps réel sur un processeur industriel standard. Toutefois les préfiltres peuvent être guidés par d'autres modèles perceptuels décrivant un seuil par pixel. Ces justifications ont été ajoutées à l'article dans la section 3.1. Protocole expérimentale, deuxième paragraphe.

Nous avons dans un premier temps utilisé la version du modèle décrit uniquement dans la luminance. Aux vues des bons résultats, nous prévoyons d'utiliser la version du modèle de Yang incluant les informations chromatiques comme précisé dans la section 4 « Conclusion et travaux futurs ».

Commentaires de l'expert 2

Il s'agit dans cet article d'une expérimentation sur la compression de vidéos 4K de filtres perceptuels déjà développés pour du contenu HD dans une précédente communication par les mêmes auteurs .
- Quelles modifications sont/doivent être apportées à ces filtres afin qu'ils soient adaptés à la résolution 4K ?

- Ces filtres sont-ils appliqués à la fois sur la luminance et les chrominances ?

Commentaires de l'expert 3

Il s'agit dans cet article d'une expérimentation sur la compression de vidéos 4K de filtres perceptuels déjà développés pour du contenu HD dans une précédente communication par les mêmes auteurs.

- Quelles modifications sont/doivent être apportées à ces filtres afin qu'ils soient adaptés à la résolution 4K ?

Suite à des tests, nous avons conservé la même taille de support de filtrage pour les vidéos 4K et les vidéos HD. Le support 11x11 offre un bon compromis entre efficacité de filtrage et complexité algorithmique. De prochains travaux tiendront compte du passage en UltraHD (une image UltraHD contient 4 fois plus de pixels qu'une image FullHD) mais aussi des tailles de CTUs pour adapter la taille de support de filtrage.

- Ces filtres sont-ils appliqués à la fois sur la luminance et les chrominances ?

Les filtres ont uniquement été appliqué sur la luminance. De prochains travaux considéreront un modèle JND et un filtrage en chrominance.

Ces informations sont données dans la section 4 « Conclusion et travaux futurs ».

Commentaires de l'éditeur

Il faut absolument répondre aux remarques et questions des experts.

Prétraitement perceptuel pour l'optimisation du codage HEVC

Eloïse VIDAL¹, François-Xavier COUDOUX¹, Patrick CORLAY¹, Christine GUILLEMOT²

¹Univ. Valenciennes, CNRS, Univ. Lille, ISEN, Centrale Lille, UMR 8520 - IEMN, DOAE, F-59313 Valenciennes, France

²INRIA, Campus Universitaire de Beaulieu, 35042 Rennes, France

eloise.vidal@lgmail.com, francois-xavier.coudoux@univ-valenciennes.fr
patrick.corlay@univ-valenciennes.fr

Résumé – Dans cet article, deux préfiltres perceptuels sont considérés pour l'optimisation de l'encodage HEVC de séquences UHD, dans le but de réduire le débit utile en préservant la qualité perçue. Ces deux préfiltres, nommés BilAWA et TBil, combinent les propriétés des filtres AWA (Adaptive Weighted Averaging) et Bilatéral. Un modèle perceptuel JND (Just Noticeable Distortion) permet de contrôler localement la force de filtrage en fonction des propriétés du système visuel humain. Des tests psychovisuels ont été conduits sur des séquences UHD au contenu spatio-temporel (SI, TI) varié. Les résultats de cette évaluation subjective, confortés par une métrique objective, ont démontré une réduction de débit jusqu'à 23% à même qualité perçue que l'encodage HEVC.

Abstract – In this paper, two perceptual filters are considered as pre-processing techniques to reduce the bitrate of HEVC compressed Ultra high-definition (UHD) video contents at constant visual quality. The proposed perceptual filters rely on two novel adaptive filters (called BilAWA and TBil) which combine the good properties of the bilateral and Adaptive Weighting Average (AWA) filters. Moreover, these adaptive filters are guided by a just-noticeable distortion (JND) model to adaptively control the strength of the filtering process, taking into account the properties of the human visual system. Extensive psychovisual evaluation tests were conducted on various spatio-temporal content (SI, TI) and were confirmed by an objective metric. Results show that applying the proposed pre-filters prior to HEVC encoding of UHD video contents lead to bitrate savings up to 23% for the same perceived visual quality.

1 Introduction

Le passage des formats de diffusion HD (High Definition) vers l'UHD (Ultra High Definition) impose l'utilisation d'algorithmes de codage de plus en plus performants pour rendre réalisable la transmission de quantités de données croissantes à travers l'infrastructure réseau existante. Dans ce contexte, la norme de codage HEVC [1] a permis de diviser par deux le débit nécessaire à la transmission de contenus vidéo UHD à même qualité que son prédécesseur H.264/AVC.

La mise en oeuvre de systèmes de prétraitement perceptuel permettent d'améliorer encore ces performances en garantissant la conformité du flux encodé à la norme HEVC [4], [2], [5], [6]. Les auteurs de [3] proposent deux préfiltres perceptuels, nommés BilAWA et TBil, pour optimiser les encodages H.264/AVC et HEVC. Les filtres AWA [7] et Bilatéral [8] ont été choisis comme base pour la définition de ces deux nouveaux filtres, de part leur bon compromis entre performance de réduction de bruit et simplicité algorithmique permettant une implémentation temps réel. Un modèle JND (Just Noticeable Distortion) est introduit dans le contrôle des préfiltres BilAWA et TBil pour réduire les informations perceptuellement non-significatives des images d'une séquence vidéo. Les préfiltres sont indépendants du schéma de codage.

Ces deux préfiltres perceptuels ont montré une réduction de débit moyen proche de 20% pour des encodages HD H.264/AVC

et HEVC, ce qui les positionne au niveau de l'état de l'art. Nous proposons ici d'étendre l'étude de ces préfiltres au cas de l'encodage HEVC de contenus vidéo 4K. Tout d'abord, les filtres perceptuels proposés dans [3] sont décrits en détail. Puis les expériences et résultats obtenus sur des contenus vidéo UHD sont présentés. L'évaluation des performances des préfiltres est basée principalement sur des tests psychovisuels faisant intervenir un large panel d'observateurs. L'évaluation subjective constitue la vérité de terrain : elle permet de qualifier de manière fiable la qualité visuelle d'une séquence vidéo. Toutefois, nous les complétons par l'utilisation de la métrique objective bien connue SSIM [12]. Les résultats montrent une réduction de débit jusqu'à 23% pour une qualité perçue identique à l'encodage HEVC.

2 Filtre perceptuel

A la manière du filtre Bilatéral, les filtres BilAWA et TBil sont composés d'un noyau géométrique et d'un noyau photométrique. Les poids du noyau géométrique sont fonction de la distance géométrique des pixels du support de filtrage au pixel courant. Les poids du noyau photométrique sont fonction de la similarité des pixels du support de filtrage au pixel courant. Le noyau géométrique des filtres BilAWA et TBil est identique, cependant ces derniers diffèrent par leur noyau pho-

tométrie illustrés par la Figure 1.a. A la manière du filtre AWA, les noyaux photométriques des filtres BilAWA et TBil comprennent un seuil au dessous duquel les différences photométriques au sein du support de filtrage sont considérées peu significatives et plus fortement filtrées. Au-delà de ce seuil, les poids décroissent progressivement. La décroissance exponentielle du filtre TBil, est plus rapide que celle du filtre BilAWA qui est en $1/x$. Cette propriété confère une force de filtrage plus importante au filtre BilAWA qu’au filtre TBil à même paramétrage. Par la combinaison des deux noyaux, plus un pixel du support de filtrage est spatialement proche et ressemblant au pixel courant, plus son poids dans le masque de filtrage est important (Cf. Figure 1.b).

Dans le but de simplifier le contenu haute fréquence des vidéos sans introduire de perte de qualité, un modèle JND est introduit dans le contrôle des préfiltres. Un modèle JND définit un seuil, en fonction des propriétés locales de l’image, en dessous duquel les différences photométriques ne sont pas perceptibles par le système visuel humain. En intégrant un tel modèle décrit ici dans le domaine pixel pour contrôler le seuil du noyau photométrique des filtres BilAWA et TBil, les informations perceptuellement non-significatives sont réduites.

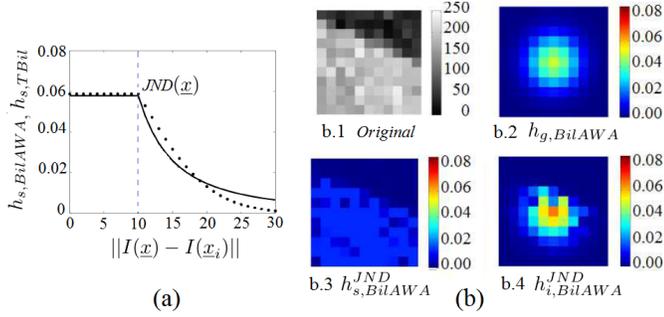


FIGURE 1 – Illustration du comportement des préfiltres proposés BilAWA et TBil. (a) Evolution des noyaux photométriques BilAWA (ligne) et TBil (points) en fonction de la distance photométrique. (b) Illustration de l’impact des noyaux photométrique (b.3) et géométrique (b.2) sur les poids BilAWA (b.4) pour un pixel particulier et son voisinage 11x11 (b1).

Le filtre BilAWA est défini par l’équation suivante,

$$h_{i,BilAWA}^{JND} = h_{g,BilAWA}(\underline{x}, \underline{x}_i) h_{s,BilAWA}^{JND}(\underline{x}, \underline{x}_i), \quad (1)$$

$$h_{g,BilAWA}(\underline{x}, \underline{x}_i) = \exp\left(-\frac{\|\underline{x} - \underline{x}_i\|^2}{2\sigma_g^2}\right), \quad (2)$$

$$h_{s,BilAWA}^{JND}(\underline{x}, \underline{x}_i) = \frac{1}{1 + \max(JND^2(I(\underline{x})), \|I(\underline{x}) - I(\underline{x}_i)\|^2)}, \quad (3)$$

Avec $h_{g,BilAWA}$ le noyau géométrique de variance σ_g^2 , le noyau photométrique $h_{s,BilAWA}^{JND}$ est guidé par un modèle JND de valeur $JND(I(\underline{x}))$, avec $I(\underline{x})$ la luminance d’un pixel à la position $\underline{x} = (x, y)$. Le paramètre a contrôle la force globale du filtre, il est fixé à 1 suivant les recommandations des auteurs du filtre AWA [7].

Le filtre TBil est défini par l’équation suivante,

$$h_{i,TBil}^{JND} = h_{g,TBil}(\underline{x}, \underline{x}_i) h_{s,TBil}^{JND}(\underline{x}, \underline{x}_i), \quad (4)$$

$$h_{s,TBil}^{JND}(\underline{x}, \underline{x}_i) = \exp\left(-\frac{\|I(\underline{x}) - I(\underline{x}_i)\|^2}{2 * JND^2(I(\underline{x}))}\right), \quad (5)$$

Avec $h_{g,TBil}$ le noyau géométrique décrit dans l’équation 2.

3 Expériences

3.1 Protocole expérimental

Afin de mettre en évidence les performances des préfiltres perceptuels étudiés, ils sont appliqués à des séquences vidéo 4K avant un encodage HEVC. Les séquences ainsi encodées sont ensuite comparées aux séquences encodées sans préfiltres.

Les préfiltres TBil et BilAWA sont appliqués sur la luminance de chaque image avec un support de filtrage 11x11. Le modèle JND spatial en luminance décrit par [9] a été choisi, toutefois d’autres modèles décrivant un seuil JND par pixel peuvent être utilisés. La variance du noyau géométrique des deux préfiltres a été fixée expérimentalement à 3.24. Le choix du modèle JND et le paramétrage des filtres ont été guidés par le meilleur compromis entre performance du prétraitement et complexité algorithmique permettant une implémentation temps réel sur une architecture parallèle.

L’implémentation x265 de la norme HEVC a été utilisée pour encoder les séquences tests, de part sa reconnaissance dans le milieu industriel. Le profil Main est utilisé avec les filtres de deblocking et SAO activés. L’encodage est réalisé à QP constant. Quatre valeurs de QP ont été sélectionnées [27, 32, 38, 41], correspondant à des débits entre 20Mbit/s à 280Kb/s. Ces débits correspondent à une utilisation de streaming temps réel.

Quatre séquences 4K ont été sélectionnées pour cette étude, elles sont présentées à la Figure 2. Les séquences *Artic*, *Tahiti* et *Boat* ont été fournies par des professionnels du broadcast. La séquence *Tears* est un extrait de la vidéo *Tears of Steel* disponible librement [10]. Ces quatre séquences ont une résolution de 3840x2160 pixels, 25 images par seconde, 250 images, et sont au format 4 : 2 : 0. Elles représentent une forte diversité de contenu comme le montre les indices d’information temporelle et spatiale [11] (Figure 5.b).

Les performances des préfiltres proposés sont analysées en termes de réduction de débit, de qualité subjective et de métrique objective (SSIM) [12]. Les séquences encodées avec les préfiltres sont comparées à la version encodée sans préfiltre selon ces trois critères.

Le protocole de comparaison par paire (PC) a été sélectionné pour l’évaluation subjective parmi les différents protocoles préconisés par la recommandation ITU P.911 [13]. Un environnement de test respectant les recommandations a été mis en place (distance d’observation, luminosité ambiante et de l’écran, ainsi que la couleur de la pièce). Quinze observateurs novices, cinq femmes et dix hommes, ont participé à l’évaluation. L’échelle de notation à sept niveaux décrite ci-dessous a été utilisée.



FIGURE 2 – Extrait des quatre séquences 4K utilisées au cours des test subjectifs.

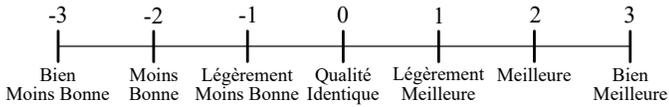


FIGURE 3 – Echelle de notation utilisée pour l'évaluation de qualité subjective.

3.2 Résultats

3.2.1 Analyse globale

Le Tableau 1 présente les trois critères d'évaluation pour les schémas de codage testés : x265 avec les préfiltres BilAWA, TBil et sans préfiltre. La Figure 4 représente les notes du test subjectif appelées CMOS (Comparison Mean Opinion Scores) \pm l'intervalle de confiance, en fonction de la réduction de débit Δ Bitrate apportée par les deux préfiltres. Pour chaque QP les résultats sont moyennés sur les quatre séquences tests.

Premièrement, les deux préfiltres apportent des réductions de débit significatives, 13.9% pour le filtre TBil et 17.7% pour le filtre BilAWA à QP 27. Ces résultats sont comparables à ceux obtenus sur des séquence HD avec l'encodeur x264 (10.3% et 15.9% pour les filtres TBil et BilAWA respectivement) et x265 (11.1% et 17.6%) à QP 27 et GOP IBBP12 [3].

Deuxièmement, ces réductions de débit sont obtenues pour une qualité perçue quasiment identique à l'encodage sans préfiltre. Les notes CMOS varient de -0.29 ± 0.19 et 0.15 ± 0.19 . Nous rappelons que la note CMOS de 0 indique une qualité identique à la séquence encodée sans préfiltre et la note -1 légèrement moins bonne. On note toutefois que l'effet du filtre TBil est encore moins perçue que celle du filtre BilAWA pour des réductions de débit moindre. Ces observations sont cohérentes avec la définition des filtres (Section 2). En effet, le noyau photométrique du filtre TBil rejette plus rapidement les pixels du support de filtrage différents du pixel courant, il en résulte une force de filtrage moindre.

Troisièmement, les notes de SSIM montrent une différence très faible entre les versions encodées avec et sans préfiltre, entre -0.0009 à QP 41 et -0.0041 à QP 27. Pour plus de lisibilité, les notes de Δ SSIM sont multipliées par 100 dans le Tableau 1.

Enfin, La réduction de débit est d'autant plus importante que le QP est faible. En effet, lorsque la quantification est importante, elle réduit d'elle même le contenu haute fréquence, laissant peu de marge de manoeuvre aux préfiltres. Des réductions de débit plus importantes pourraient être obtenues pour des débits plus importants (QP plus faibles).

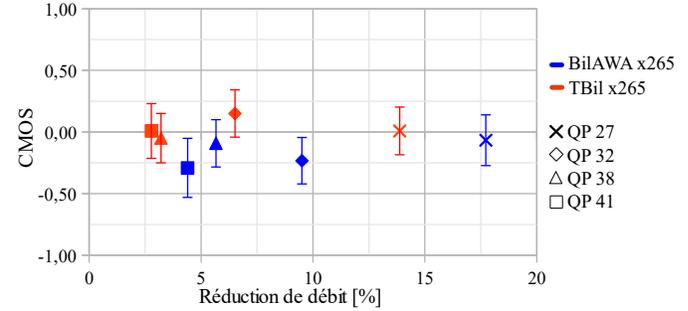


FIGURE 4 – Evolution de la réduction de débit en fonction du QP

TABLE 1 – Comparaison des performances du codec x265 avec et sans préfiltre perceptuel. Analyse des performances des deux préfiltres proposés en termes de variation de débit (Δ Débit), de qualité subjective (CMOS avec son intervalle de confiance $\delta_{[95\%]}$) et de variation de mesure objective (δ SSIM). Les résultats sont donnés par QP en moyennant les résultats sur toutes les séquences. Le codec x265 est utilisé avec les filtres de deblocking et SAO, avec un GOP IBBP12.

| | | QP 27 | QP 32 | QP 38 | QP 41 |
|--------|-----------------------|---------------|--------------|--------------|--------------|
| | Débit [Mbs] | 7.92 | 3.44 | 1.46 | 1.03 |
| | SSIM | 0.964 | 0.957 | 0.946 | 0.939 |
| BilAWA | Δ Débit [%] | -17.73 | -9.51 | -5.67 | -4.40 |
| 11x11 | CMOS | -0.07 | -0.23 | -0.09 | -0.29 |
| JND | $\pm \delta_{[95\%]}$ | 0.19 | 0.19 | 0.20 | 0.22 |
| +x265 | Δ SSIM * 100 | -0.410 | -0.272 | -0.201 | -0.167 |
| TBil | Δ Débit [%] | -13.88 | -6.51 | -3.21 | -2.78 |
| 11x11 | CMOS | 0.01 | 0.15 | -0.05 | 0.01 |
| JND | $\pm \delta_{[95\%]}$ | 0.21 | 0.19 | 0.19 | 0.24 |
| +x265 | Δ SSIM * 100 | -0.290 | -0.168 | -0.109 | -0.085 |

3.2.2 Analyse en fonction du contenu

La Figure 5 présente les performances du préfiltre BilAWA pour chaque QP et chaque séquence (a), ainsi que les indices d'information spatiale et temporelle des quatre séquences tests (b). Des résultats similaires sont obtenus avec le préfiltre TBil. L'étendue des réductions de débit en fonction du QP varie en fonction des séquences, de 4.4% à 10.6% pour la séquence *Artic* et de 3.0% à 22.8% pour la séquence *Boat*. On constate une corrélation entre la quantité d'information spatiale et temporelle contenue dans la séquence et la réduction de débit apportée par le préfiltre BilAWA. Cette corrélation est spécialement notable à QP 27 où le classement de la réduction de débit et de

la quantité d'information est identique : *Artic* 10.6% pour un couple d'indices SI TI de [9.2, 2.9], *Tahiti* 17.5% [17.6, 6.0], *Tears* 20.0% [17.7, 30.2] et *Boat* 22.8% [21.5, 29.4]. En revanche, lorsque la quantification augmente, la quantité d'information contenue dans les séquences ne permet pas de prévoir le classement des réductions de débit. Ainsi à QP 41, les plus fortes réductions de débit sont obtenues pour les séquences *Tahiti* et *Artic*.

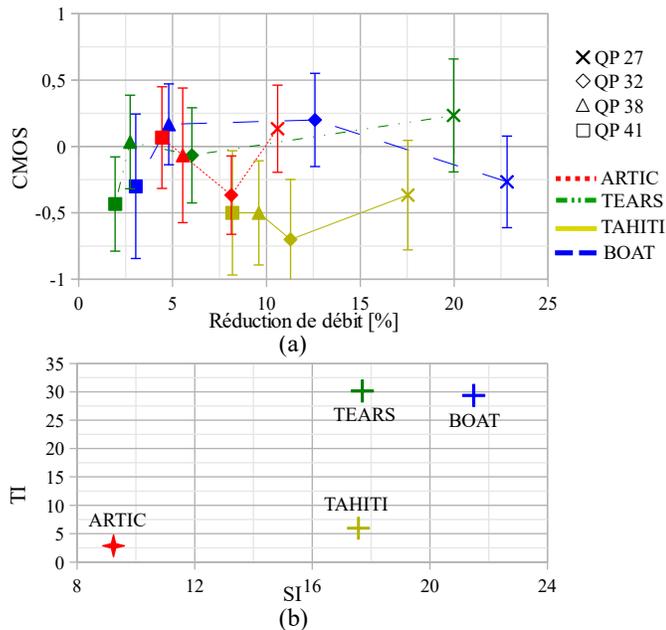


FIGURE 5 – Analyse des performances des préfiltres en fonction du contenu. (a) Performances du filtre BilAWA pour chaque QP et chaque séquence. (b) Indice spatial et temporel de chaque séquence.

4 Conclusion et travaux futurs

Dans cet article nous avons présenté les performances des préfiltres perceptuels BilAWA et TBil pour l'optimisation du codage HEVC de séquences 4K. L'introduction d'un modèle perceptuel JND dans le contrôle des filtres a permis de réduire le contenu perceptuellement non-significatif des séquences vidéo. Les expériences ont montré une réduction de débit jusqu'à 23% pour une qualité perçue identique au codage HEVC sans préfiltre. Dans de futurs travaux nous considérerons un modèle perceptuel JND spatio-temporel intégrant les informations colorimétriques ainsi qu'une adaptation du support de filtrage aux CTUs HEVC.

Les auteurs souhaitent remercier Elie de Rudder pour les expériences conduites durant son stage. Nicolas Braud de TF1 pour les contenus UHD et Prof. Sylvie Merviel-Leleu pour le prêt d'équipements audiovisuels. Ces travaux ont été en partie réalisés en collaboration avec Digigram et l'ANRT (Cifre #

1098/2010).

Références

- [1] G. J. Sullivan, J.-R. Ohm, W. Han, and T. Wiegand, *Overview of the High Efficiency Video Coding (HEVC) Standard*. IEEE Trans. On Circuits and Systems for Video Technology, vol.22, no.12, December 2012.
- [2] S.H. Bae, J. Kim, M. Kim, *HEVC-Based Perceptually Adaptive Video Coding Using a DCT-Based Local Distortion Detection Probability Model*. IEEE Trans. on Image Processing, vol.25, no.7, pp. 3543–3357, Jul. 2016.
- [3] E. Vidal, N. Sturmel, C. Guillemot, P. Corlay, and F. Coucoudoux, *New adaptive filters as perceptual preprocessing for Rate-Quality Performance Optimization of Video Coding*. Signal Processing : Image Communication, vol.52, pp. 124–137, March 2017.
- [4] R. Vanam, L. Kerofsky, and Y. Reznik, *Perceptual preprocessing filter for video on demand content delivery*. IEEE ICIP, pp. 2537–2541, October 27-30 2014.
- [5] L. Ding, G. Li, R. Wang, and W. Wang, *Video preprocessing with JND-based Gaussian filtering of superpixels*. Proc. SPIE 9410, Visual Information Processing and Communication VI, 941004, March 2015.
- [6] H. Oh and W. Kim, *Video Processing for Human Perceptual Visual Quality-Oriented Video Coding*. IEEE Transactions on Systems and Circuits for Video Technology, vol.22, no. 4, pp. 1526-1535, 2016.
- [7] M. Ozkan, I. Sezan, and M. Tekalp, *Adaptive motion-compensated filtering of noisy image sequences*. IEEE Trans. on Circuits and Systems for Video Technology, vol.3, no.4, pp. 277–290, Aug. 1993.
- [8] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*. ICCV, pp. 836–846, 1998.
- [9] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, *Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile*. IEEE Transactions on Circuits and Systems for Video Technology, vol.15, no.6, pp. 742–752, June 2005.
- [10] Test video sequences available at the Xiph web site : <https://media.xiph.org/>
- [11] M. H. Pinson and S. Wolf., *A New Standardized Method for Objectively Measuring Video Quality*. IEEE Trans. On Broadcasting, vol.50, no.3, September 2004.
- [12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment : from error visibility to structural similarity*. IEEE Trans. Image Process, vol.13, no.4, pp. 600–612, 2004.
- [13] ITU-T Recommendation P.911, *Subjective audiovisual quality assessment methods for multimedia applications*. Series P, December 1998.