Etude de l'influence du contenu vidéo dans une transmission Pseudo-Analogique de type SoftCast

Anthony Trioux, François-Xavier Coudoux, Patrick Corlay, M Gharbi

▶ **To cite this version:**

Anthony Trioux, François-Xavier Coudoux, Patrick Corlay, M Gharbi. Etude de l'influence du contenu vidéo dans une transmission Pseudo-Analogique de type SoftCast. CORESA 2018, Nov 2018, Poitiers, France. hal-02423790

HAL Id: hal-02423790
https://uphf.hal.science/hal-02423790

Submitted on 25 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Titre :

Etude de l'influence du contenu vidéo dans une transmission Pseudo-Analogique de type SoftCast

Auteurs :

Anthony TRIOUX, François-Xavier COUDOUX, Patrick CORLAY, Mohamed GHARBI

Résumé (*100-200 mots*) :

Récemment, un nouveau schéma *SoftCast* a démontré un potentiel élevé pour la diffusion de vidéo sur des réseaux sans fil avec des caractéristiques de canaux variant dans le temps, en comparaison aux schémas vidéo classiques (H.264/AVC, H.264/SVC, HEVC). Particulièrement utile en situation de Broadcast, il permet de traiter les problèmes rencontrés dans les systèmes de diffusion vidéo sans-fil conventionnels comme le *cliff-effect* [1]. L'objectif de ces travaux est de proposer une amélioration du schéma *SoftCast* originel, en effectuant un encodage adaptatif prenant en considération les contraintes liées à la bande passante ainsi que les caractéristiques des contenus vidéos. Nos résultats de simulations mettent clairement en évidence l'importance de s'adapter au contenu vidéo dans *SoftCast*. Selon les caractéristiques spatio-temporelles du contenu vidéo, une amélioration du PSNR de l'ordre de 2dB est observée pour les contenus vidéos à faible index spatial et temporel (SI, TI) [2], alors qu'un gain en terme de complexité d'encodage allant jusqu'à 40% est observé à qualité quasi-identique pour les contenus vidéos à haut SI, TI.

Mots-clefs (*3 à 5 max.*) :

Transmission vidéo sans-fil, SoftCast, Transmissions Vidéo non Codées, Index Spatio-Temporels

Contexte et état de l'art :

Différent des standards de transmission vidéo actuels, *SoftCast* est un système de codage conjoint source-canal où les pixels sont traités par des opérations linéaires successives et directement transmis sans quantification ni codage. De fait, il permet d'obtenir une dégradation progressive de l'image (graceful degradation [3]) : la qualité reçue est une fonction linéaire du bruit du canal [4]. Le schéma de base de *SoftCast* [3] est présenté dans la Fig. 1. *SoftCast* travaille sur des groupes d'images (GoP) et utilise une transformation DCT-3D. Les plans DCT résultants sont ensuite divisés en blocs rectangulaires de coefficients transformés appelés *chunks*. Ensuite, une mise à l'échelle des coefficients transformés (*scaling*), garantit que les données vidéo sont bien protégées contre le bruit des canaux. La transformée d'Hadamard est par la suite appliquée aux coefficients mis à l'échelle afin de fournir une résilience à la perte de paquets. Ce processus transforme les *chunks* en *slices*. Chaque slice est une combinaison linéaire de tous les *chunks*. Enfin, ces paquets sont transmis d'une manière pseudo-analogique à l'aide d'une modulation dense : Raw-OFDM [16]. Cette étape se fait en ignorant les codes correcteurs d'erreurs (FEC) et la partie de modulation classique d'un système OFDM. Ainsi, dans SoftCast, au lieu de débit binaire, le débit symbole est considéré. Etant donné que les coefficients transformés sont envoyés en paires (plan I et Q en OFDM), la bande passante maximale canal dans SoftCast peut être définie de la manière suivante : BW_{max} = (résolution vidéo * frame rate)/2. Lorsque la bande passante de l'application ne permet pas la transmission de tous les coefficients transformés, SoftCast opère en jetant des *chunks*. Les *chunks* sont préalablement triés par ordre décroissant de variances énergétiques. Dans SoftCast, le niveau de compression CR est défini comme suit : $CR = BW_{ava}/BW_{max}$ où BW_{ava} dénote la bande passante disponible à l'émetteur.

En parallèle du processus de transformation, l'émetteur SoftCast envoie des métadonnées. Ces données sont constituées de la moyenne et de la variance des *chunks* transmis, ainsi que d'une carte binaire, qui indique les positions des *chunks* ignorés dans le GoP. Ces métadonnées sont fortement protégées et transmises de manière robuste pour assurer une réception correcte.

Du côté du récepteur, un décodeur LLSE (Linear Least Square Error) est utilisé afin d'obtenir la meilleure estimation des coefficients DCT reçus. En utilisant les métadonnées et les opérations inverses du décodeur, ces coefficients sont ensuite réassemblés pour reformer les plans DCT. Ces derniers sont ensuite transformés via un processus 3D-DCT inverse. Dans la configuration actuelle, le schéma *SoftCast* utilise une taille de GoP fixe qui ne permet pas de s'adapter aux spécificités spatiotemporelles de chaque contenu vidéo.

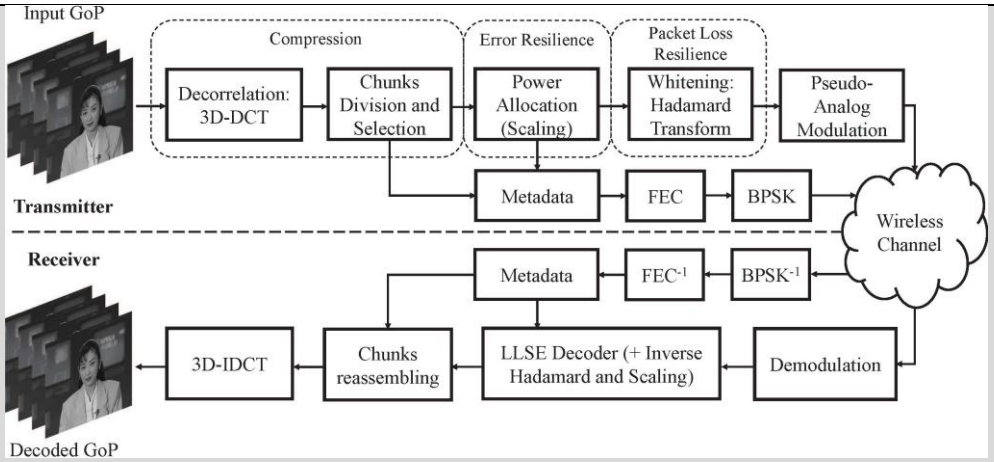

Figure 1 Schéma block du schéma de transmission vidéo SoftCast ([3]).

Travail proposé (*Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié*) :

Dans ces travaux, nous examinons tout d'abord l'impact du contenu vidéo sur la qualité reçue à la réception pour différentes tailles de GoP. Comme mentionné précédemment, le but est de proposer une amélioration du schéma *SoftCast* originel, en effectuant un encodage adaptatif prenant en considération les contraintes liées à la bande passante du canal ainsi que les caractéristiques des contenus vidéos. Ceci dans le but de proposer le meilleur compromis bande-passante canal/reconstruction vidéo. Le défi technique consiste à proposer un algorithme adaptatif pouvant s'inscrire dans un schéma de type *SoftCast* qui détermine à la volée la taille de GoP optimale suivant les variations d'activités spatio-temporelles pouvant avoir lieu dans les séquences vidéos.

Résultats :

La métrique du Peak Signal-to-Noise Ratio (PSNR) est tout d'abord utilisée comme indicateur objectif de la qualité vidéo reçue ceci pour différents rapports signal/bruit (CSNR) et différentes tailles de GoP. Le schéma *SoftCast* est implémenté comme dans [3]. Cependant, afin d'améliorer la qualité en réception, la moyenne de chaque image est soustraite et envoyée en metadata comme décrit dans [5]. Trois tailles de GoP sont ici utilisées : 8, 16 et 32 images. Compte tenu des limitations de place du résumé étendu, 2 séquences CIF parmi l'ensemble des tests effectués sont ici présentées : *Akiyo* et *Husky* (index spatio-temporels très hétérogènes). Pour une comparaison juste, un même bruit est généré et utilisé pour toutes les configurations.

Séquence		CSNR(dB) 0	5	10	15	20	25
Akiyo (SI = 9.57, TI = 5.86)	GoP 8	33,68	38,09	42,69	47,31	51,67	56,69
	GoP 16	35,14	39,50	44,05	48,62	52,86	58,88
	GoP 32	**36,25**	**40,56**	**45,06**	**49,53**	**53,79**	**60,94**
Husky (SI = 20.96, TI = 62.02)	GoP 8	20,57	24,25	28,64	33,39	38,27	43,17
	GoP 16	20,61	24,29	28,67	33,42	38,31	43,20
	GoP 32	20,62	24,28	28,67	33,42	38,31	43,20

Tableau 1. Tableau des PSNR pour différentes tailles de GoP et différents CSNR, CR = 1 (pas de compression).

Puisque le PSNR est une métrique basée sur l'erreur quadratique moyenne (MSE) n'étant pas corrélée avec la perception humaine, nous utilisons également l'index SSIM (Structural SIMilarity) en tant que seconde métrique :

Séquence		CSNR(dB) 0	5	10	15	20	25
Akiyo (SI = 9.57, TI = 5.86)	GoP 8	0,875	0,942	0,976	0,991	0,996	0,998
	GoP 16	0,901	0,954	0,981	0,993	0,997	0,999
	GoP 32	**0,916**	**0,962**	**0,984**	**0,994**	**0,997**	**0,999**
Husky (SI = 20.96, TI = 62.02)	GoP 8	0,671	0,817	0,912	0,964	0,987	0,995
	GoP 16	0,671	0,817	0,912	0,964	0,987	0,995
	GoP 32	0,670	0,816	0,911	0,963	0,987	0,995

Tableau 2. Tableau des index SSIM pour différentes tailles de GoP et différents CSNR, CR = 1 (pas de compression).

Les résultats présentés dans le tableau 1 et 2 montrent que :

- Pour les séquences vidéos présentant une faible activité spatiotemporelle (de type *Akiyo*), une GoP de grande taille permet d'obtenir une meilleure qualité de reconstruction.

- Dans le cas contraire (forte activité spatiotemporelle) comme *Husky*, il n'y a pas d'intérêt à augmenter la taille de la GoP en terme de qualité de reconstruction. Ceci est d'autant plus vrai que la complexité augmente selon *O(K* log(K)) avec K le nombre de pixels dans un GoP [3]. Maintenir une taille de GoP faible permet dans ce cas une réduction de la complexité allant jusqu'à 40%.

La figure 2 illustre également les images reconstruites pour les différentes configurations dans le canal le plus perturbé correspondant à un CSNR égal à 0dB :

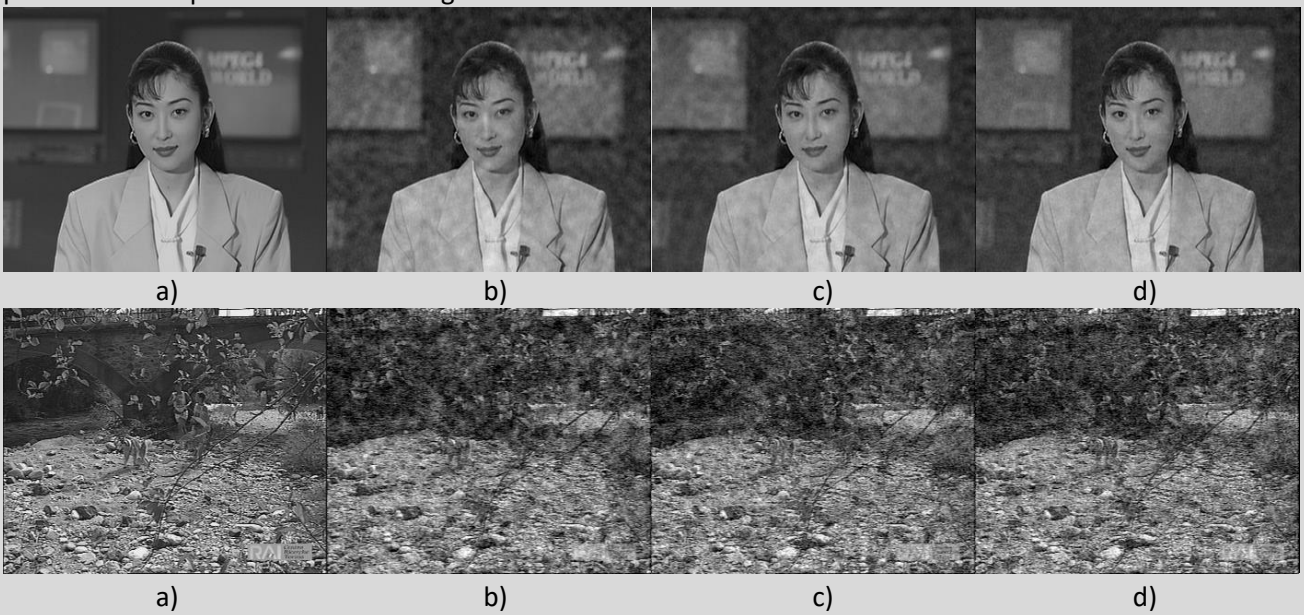

Figure 2 Comparaison visuelle de la qualité à la réception dans un canal à 0dB, CR = 1 (pas de compression). De haut en bas : Séquence Akiyo, Séquence Husky. De gauche à droite : a) première image de la séquence originale, b) première image reconstruite GoP = 8, c) première image reconstruite GoP = 16, d) première image reconstruite GoP = 32.

Des premiers tests concernant les applications à bande passante restreinte ont également été effectués. Le niveau de compression appliqué a été réduit, passant de 1 à 0.25 (25% des coefficients transmis). Les conclusions obtenues dans la première étude (CR = 1, pas de compression) sont également valables lorsque l'application visée requiert l'abandon de certains coefficients. Etant donné les contraintes de place du résumé étendu, ces résultats ne peuvent pas être intégrés ici et seront donc présentés lors de la présentation orale.

Conclusion et perspectives :

Les résultats préliminaires de simulations mettent clairement en évidence l'importance de l'adaptation au contenu vidéo dans un système de transmission vidéo sans fil *SoftCast*.

L'impact du contenu vidéo dans des environnements à bande passante restreinte est en cours d'étude afin de généraliser les travaux actuels. Une proposition de schéma de transmission vidéo de type *SoftCast* effectuant un encodage adaptatif est en cours. Cette version adaptative permet de prendre en considération à la fois les contraintes liées à la bande passante disponible mais aussi les caractéristiques des contenus vidéos et propose le meilleur compromis bande-passante/reconstruction vidéo. Selon les contraintes matérielles de l'application visée, un compromis qualité-complexité pourra être choisi permettant d'obtenir la meilleure qualité disponible pour la ressource matérielle utilisée.

Références (*5 max.*):

[1] S. Kokalj-Filipovi, E. Soljanin, "Suppressing the cliff effect in video reproduction quality," Bell Labs Technical Journal, vol. 16, no. 4, pp.171–185, Mar. 2012.

[2] ITU-T Recommendation, "Subjective video quality assessment methods for multimedia applications," 1999.

[3] S. K. Jakubczak, "SoftCast: exposing a waveform interface to the wireless channel for scalable video broadcast," MIT PhD Thesis, 2011.

[4] R. Xiong, F. Wu, J. Xu, X. Fan, C. Luo, W. Gao, "Analysis of decorrelation transform gain for uncoded wireless image and video communication," IEEE Trans. Image Process., vol. 25, no. 4, pp. 1820–1833, Apr. 2016.

[5] A. Trioux, F.-X. Coudoux, P. Corlay, M. Gharbi, "A Comparative Preprocessing Study for SoftCast Video Transmission", Accepted in IEEE ISIVC Symposium, 27-30 Nov. 2018, Rabat, Morroco.