



Codage vidéo à description multiple basé sur HEVC pour le pilotage de véhicules semi-autonomes

Trung Hieu Le, Marc Antonini, Marc Lambert, Karima Alioua

► To cite this version:

Trung Hieu Le, Marc Antonini, Marc Lambert, Karima Alioua. Codage vidéo à description multiple basé sur HEVC pour le pilotage de véhicules semi-autonomes. XXVIIIème Colloque Francophone de Traitement du Signal et des Images (GRETSI'22), Sep 2022, Nancy, France. pp.1125-1128. hal-04208550

HAL Id: hal-04208550

<https://hal.science/hal-04208550>

Submitted on 15 Sep 2023

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.

Codage vidéo à description multiple basé sur HEVC pour le pilotage de véhicules semi-autonomes

Trung Hieu LE¹, Marc ANTONINI¹, Marc LAMBERT², Karima ALIOUA²

¹Laboratoire I3S - Université Côte d'Azur et CNRS, UMR 7271, Sophia Antipolis, France

²Lextan SAS, Gemenos, France

thle@i3s.unice.fr, am@i3s.unice.fr

marc.lambert@lextan.fr, karima.alioua@lextan.fr

Résumé – Les véhicules semi-autonomes nécessitent la transmission d'une grande quantité de données. L'une des sources de données les plus critiques pour le conducteur provient de la vidéo, qui est cruciale pour assurer le contrôle du véhicule. Dans le cadre d'un projet piloté par la compagnie LEXTAN, le même flux vidéo codé est transmis simultanément sur plusieurs réseaux radio 5G pour assurer une transmission vidéo fiable. Cependant, cette solution présente un coût de transmission élevé. En raison des caractéristiques du canal sans fil, qui n'est pas fiable et présente des pertes de bits aléatoires, il est nécessaire d'utiliser des méthodes d'encodage vidéo plus efficace qui rendent le flux vidéo robuste aux bruits. Parmi toutes les solutions possibles pour des méthodes de codage vidéo robustes sur des canaux bruités, le codage à description multiple (MDC) est un bon candidat qui peut satisfaire les contraintes de la transmission vidéo pour la conduite telles qu'une faible latence et une bonne qualité d'expérience (QoE). Dans cet article, nous appliquons une solution MDC à 2 canaux adaptée au codage HEVC "All-Intra". Cette solution montre des performances de codage élevées avec des gains de BD-PSNR allant jusqu'à 5 dB pour une séquence CIF et jusqu'à 9 dB pour une séquence Full HD par rapport à un codage à description unique en présence de bruit de canal.

Abstract – Remote-control vehicles require the transmission of a vast amount of data. One of the most critical data sources for the driver comes from the video, which is crucial to ensure the control of the vehicle. Under the LEXTAN's project scope, the same encoded video stream is transmitted simultaneously over several 5G radio networks to ensure reliable video transmission. However, this solution has a high transmission cost. Due to the characteristics of the wireless channel, which is unreliable and has random bit loss, it is necessary to use more efficient video encoding methods that make the video stream robust to noises. Among all the possible solutions for robust video coding methods on noisy channels, multiple description coding (MDC) is a good candidate that can satisfy remote-control constraints such as low latency and good quality of experience (QoE). In this paper, we use a 2-channel MDC solution adapted to "All-Intra" HEVC encoding. The solution shows high coding performance with BD-PSNR gains up to 5 dB for CIF sequence and up to 9 dB for Full HD sequence compared to a single description coding in the presence of channel noise.

1 Introduction

Le pilotage à distance de véhicules semi-autonomes nécessite une transmission constante de données vidéos par le biais d'un canal radio-mobile. Ces données permettent au conducteur distant de percevoir l'environnement du véhicule. Dans une telle application, les images reçues doivent respecter les contraintes sur la latence et la qualité afin de garantir la sécurité. Le réseau 5G est caractérisé par une haute performance et une faible latence, cependant, ce réseau est plus sensible aux bruits causés par l'environnement. Pour résoudre ce problème, plusieurs approches ont été développées afin d'améliorer la résistance au bruit de la transmission vidéo, comme l'Automatic Repeat Request (ARQ), qui demande à l'émetteur de renvoyer les paquets perdus. Cependant, cette méthode n'est pas adaptée à la communication en temps réel, car elle introduit beaucoup de délais. D'autres solutions utilisent les codes correcteurs d'erreurs (*Forward Error Correction -FEC-*) qui ajoutent des données dans le flux binaire pour corriger les paquets perdus. Cependant, la performance des FEC est limitée par un taux maximal de perte de paquets. Or, comme la caractéristique du canal 5G varie fortement dans le temps, le taux de perte de paquet toléré peut être dépassé et qui rends la méthode de FEC inadaptée.

Dans notre contexte des véhicules semi-autonomes, la so-

lution de transmission robuste proposée par la société LEXTAN¹ consiste à transmettre la même séquence vidéo simultanément sur deux canaux sans fil indépendants afin de pouvoir prendre le relais lorsqu'un des deux canaux est bruité. Cependant, la transmission de la même séquence vidéo entraîne un gaspillage de la bande passante. Par conséquent, le codage par descriptions multiples (MDC) introduit en 1979 par Gersho, Witsenhausen, Wolf, Wyner, Ziv et Orarow à IEEE Information Theory Workshop en Septembre 1979 [1] apparaît comme une solution pour réduire le coût de la transmission et la robustifier.

Dans le cas du MDC optimisé pour deux canaux, le codeur MDC produit deux descriptions différentes S1, S2 de la vidéo avec les débits binaires R1 et R2, respectivement, à partir de la source vidéo originale. Ensuite, deux descriptions seront transmises sur deux canaux indépendants par deux transmetteurs. S'il n'y a qu'une seule description disponible au décodeur MDC, soit S1 ou S2, le décodeur latéral sera utilisé pour produire la séquence vidéo avec le niveau de distorsion D1 ou D2, respectivement. Sinon, si deux descriptions sont disponibles au niveau du décodeur MDC, le décodeur central fusionne ces deux descriptions afin de construire la description

1. <https://www.lextan.co>

dite centrale en supprimant les informations redondantes et en ne conservant que la principale. Par conséquent, la qualité vidéo sera plus élevée avec une distorsion centrale D0 plus petite. Le principe du MDC est décrit dans la figure 1.

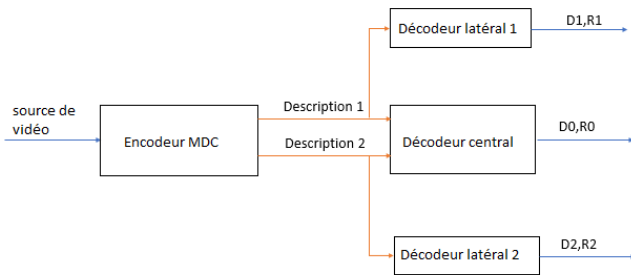


FIGURE 1 – Le principe de codage par descriptions multiples sur deux canaux de transmission.

Selon la manière dont les informations de redondance sont ajoutées dans chaque description, le MDC est classé en trois catégories principales : MDC dans le domaine spatial, MDC dans le domaine fréquentiel et MDC dans le domaine temporel. Plusieurs approches de MDC sont développées comme [2], une méthode de MDC proposée dans le domaine fréquentiel, qui est basée sur l'allocation de différents niveaux de quantification des différentes sous-bandes de la transformée en ondelette selon une contrainte de débit-distorsion est développé. Au décodeur, les sous-bandes ayant le plus petit pas de quantification sont réunies pour reconstruire la description centrale. Cependant, la compression avec des coefficients d'ondelette nécessite une capacité de mémoire et une complexité élevées. Il existe plusieurs études de MDC dans le domaine spatial, la plupart étant compatibles avec la norme d'encodage H.264 comme dans [3], Tilo *et al* ont proposé une méthode de codage à descriptions multiples compatible avec la norme H.264 basée sur l'ajustement du niveau de redondance des différentes "slices". Dans l'article [4], les auteurs ont proposé un schéma MDC H.264 basé sur la permutation des blocs et la division des coefficients DCT.

En comparaison avec la norme H.264, la norme H.265 est plus performante jusqu'à 50 % du débit binaire par rapport à la norme précédente H.264 [5]. Cela est dû au fait que sa structure de codage, l'unité d'arbre de codage (CTU), peut être divisée en unités de codage (CU) selon l'objectif de débit binaire et de qualité. Plus la taille du bloc de codage est grande, plus les performances de compression sont élevées au détriment de la qualité et vice versa.

Inspirés par cette caractéristique de la norme HEVC, nous proposons un codeur à descriptions multiples dans le domaine spatial, adapté au codage intra-frame de HEVC, en attribuant différents niveaux du paramètre de quantification (QP) pour chaque CTU dans une image. Cette solution garantit une robustesse et une faible latence pour la transmission vidéo sur des réseaux sans fil bruités. Cet article s'organise en deux parties. La première partie présente la solution proposée. Ensuite,

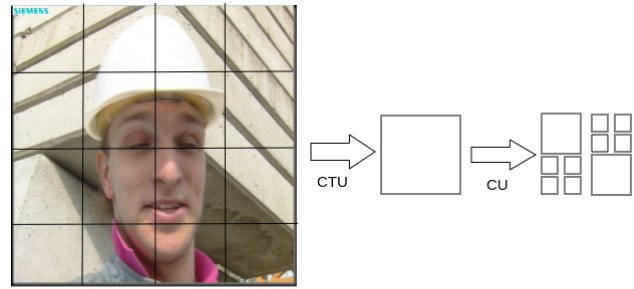


FIGURE 2 – Unité de codage HEVC. Dans cette illustration, une image donnée est d'abord divisée en CTU de 64x64 pixels. Ensuite, chaque CTU peut ensuite être divisée en plusieurs CUs de taille 32x32 et 16x16 pixels.

la deuxième partie montre les performances expérimentales de la solution dans le cas d'un canal bruité.

2 MDC basé sur HEVC

2.1 La norme de codage HEVC

Dans la norme HEVC, une image est d'abord divisée en CTUs de taille variant de 64x64 pixels à 16x16 pixels. Un CTU est un arbre à quatre feuilles. Chaque feuille du CTU est un CU dont la taille varie de 64x64 pixels à 8x8 pixels (voir figure 2).

La prédiction permet d'améliorer les performances de la compression. Il existe deux processus de prédiction : la prédiction *Intra* et la prédiction *Inter*. Bien que le processus de prédiction *Inter* permet d'économiser plus de débit binaire que celui de prédiction *Intra*, les dépendances temporelles entre les images réduisent la résistance du flux binaire au bruit. Par conséquent, dans ce travail, nous ne considérons que le mode de codage HEVC "All-Intra" qui n'utilise pas d'images prédites dans la séquence vidéo.

Les résidus obtenus à l'issue de la prédiction *Intra* sont transformés en coefficients DCT. Ils sont ensuite quantifiés selon un pas de quantification donné. Plus le pas de quantification augmente, plus la qualité du bloc DCT décodé diminue, ainsi que son débit binaire. Les coefficients quantifiés sont ensuite codés avec le codeur entropique CABAC en flux binaire. De plus, la norme HEVC définit le "slice" comme une unité décodable indépendante, un slice pouvant contenir un CTU ou regrouper plusieurs CTUs. Un slice est contenu dans une unité NAL (Network Abstraction Layer). On peut noter que les slices de petite taille sont plus résistantes au bruit, mais au prix d'un coût élevé à cause de la taille de l'entête NAL par rapport à celle de la donnée utile ("Payload"). Par conséquent, il est important de trouver un compromis entre la taille de la slice, la robustesse et le coût de l'entête.

2.2 La solution de MDC spatial proposée

Dans cet article, nous utilisons la méthode de MDC à 2 canaux adaptée à la norme HEVC. Comme mentionné précédemment, la solution proposée est appliquée à l'ensemble du mode intra (I-frames) pour éviter la propagation d'erreurs dues à la prédiction temporelle des P-frames. Ainsi, nous créons deux descriptions redondantes de telle sorte que pour chaque image

d'une description donnée, le codeur entrelace des CTUs de haute qualité (valeur QP basse) avec des CTUs de qualité basse (valeur QP élevée). Pour ce faire, il affecte différentes valeurs QP pour chaque CTU dans deux descriptions de manière complémentaire. Ensuite, ces deux descriptions sont envoyées à travers deux réseaux sans fil indépendants par deux émetteurs. Comme le montre la figure 3, la description S1 est construite en alternant la valeur de quantification QP_p pour les CTUs principales et QP_r pour les CTUs redondantes avec $QP_p < QP_r$. L'autre description S2 est aussi construite avec le set de valeurs de quantification $\{QP_p, QP_r\}$ mais de manière complémentaire comparée à S1. HEVC traite les CTUs d'une image ligne par ligne ; donc, l'affectation des CTUs principaux et des CTUs redondants doit également se faire ligne par ligne. Puis, les deux descriptions peuvent ensuite être décodées indépendamment en utilisant deux décodeurs HEVC côté décodeur. En absence de bruit, il est évident que le décodeur MDC central assemble les CTUs ayant le pas de quantification le plus faible (QP_p) et rejette les CTUs ayant le pas de quantification le plus élevé (QP_r) afin de reconstruire la description centrale avec la meilleure qualité possible.

Cependant, la sélection des CTUs n'est pas évidente lorsque le flux binaire est bruité, car les CTUs ayant une valeur QP plus petite peuvent aussi être bruités. Une expérience a été menée en utilisant la décision "oracle", autrement dit en utilisant l'image de référence pour différencier les CTUs non-bruités des bruités. Cependant, cette méthode reste purement théorique et nous sert uniquement à évaluer la borne supérieure de la performance de notre MDC. Afin d'identifier les pertes, nous pouvons nous baser sur l'entête du slice dans la trame binaire. Si l'entête du slice n'est pas présent à la bonne position, on peut en déduire que le slice est corrompu. Par conséquent, celui-ci sera remplacé par le slice de l'autre description. Cependant, cette méthode est seulement appliquée lorsque la taille du slice est petite. Si la taille du slice est grande, il est plus probable que les deux slices de chacune des descriptions soient à la fois bruités et qu'on ne puisse pas trouver la bonne position du header dans aucune des deux descriptions. Dans ce cas de figure, pour décoder un CTU courant, le décodeur central se base sur les CTUs de son voisinage causal \mathcal{V} , spatial et temporel, afin de sélectionner le CTU de meilleure qualité entre les deux descriptions (voir figure 4). L'idée est de choisir la description pour laquelle le CTU courant a la plus forte "ressemblance" avec ses voisins en utilisant la divergence Kullback-Leiber (KL) symétrisée. Cette mesure donne la meilleure estimation de niveau ressemblance par rapport aux autres indices comme MSE et coefficient de corrélation. L'efficacité de cette mesure est évaluée expérimentalement. La formule de d est donnée par la formule suivante :

$$d = \frac{KL(P||Q) + KL(Q||P)}{2} \quad (1)$$

avec KL la divergence de Kulback-Leibler donnée par :

$$KL(P||Q) = \sum_{i=0}^{255} P_i \log \left(\frac{P_i}{G_i} \right). \quad (2)$$

P_i et G_i sont les distributions de probabilités des intensités i définies dans chacun des CTUs considérés dans le voisinage.

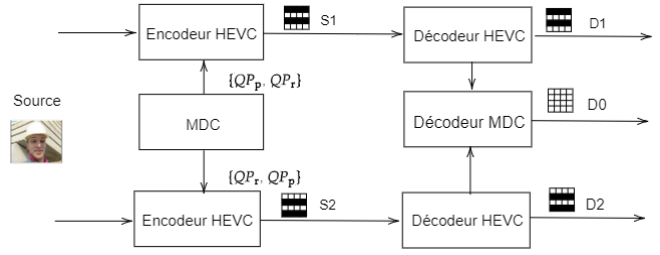


FIGURE 3 – Le schéma MDC proposé. Une séquence vidéo est dupliquée en deux descriptions et chaque description est codée en HEVC avec des paramètres de quantification (QP_p et QP_r) définis par le MDC, générant deux flux binaires différents $S1$ et $S2$. Le décodage est assuré grâce à trois décodeurs : deux latéraux (HEVC) et un central (MDC), fournissant trois vidéos décodées possibles $D1$, $D2$ et $D0$.

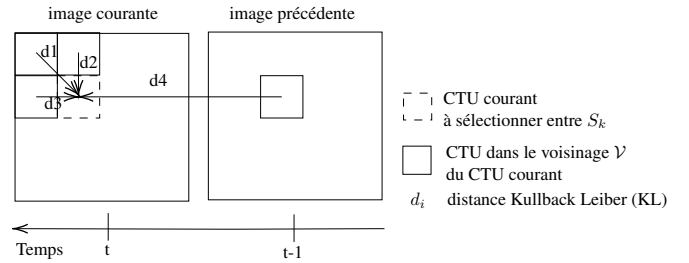


FIGURE 4 – La sélection entre les deux CTUs de deux descriptions différentes.

Ainsi, pour chaque description $S_k \forall k \in \{1, 2\}$, le décodeur central calcule la divergence moyenne $\bar{d}(k)$ entre le CTU courant et les CTUs appartenant à son voisinage \mathcal{V} tel que défini dans la figure 4. Il choisit alors le CTU de la description S_k qui présente la plus petite divergence moyenne $\bar{d}(k)$ avec son voisinage.

3 Résultats expérimentaux

Dans cette partie, l'objectif est de mesurer l'impact de la perte de paquets sur notre solution MDC comparée à la solution HEVC à une seule description, notée SDC. Le protocole de transmission vidéo choisi est Real-Time Protocol (RTP) et le "payload" dans un RTP est égal à 1446 octets pour être compatible avec le réseau sans fil. Par conséquent, si la taille de l'unité de NAL est plus grande que la taille du payload, cette unité sera divisée en petits fragments. Ici, nous avons choisi de simuler la perte de paquets RTP sur un réseau IP au moyen du modèle de canal de Gilbert Elliot [6]. Le réseau est configuré pour simuler le bruit avec différents taux de perte de paquets p_r . Dans nos expériences, les valeurs des taux de perte sont telles que $p_r = \{0.05, 0.1, 0.15, 0.2\}$. La probabilité de récupération de l'état de perte de paquets est de 1. Dans nos expériences, tous les CTU dans une ligne de l'image sont contenus dans un slice autrement dit dans une unité NAL.

Pour le décoder les flux HEVC bruités, nous avons utilisé le décodeur lib265 [7] qui tolère des informations de perte pour décoder les flux binaires bruités. Pour évaluer l'impact du bruit, nous avons utilisé trois séquences dans cette expérience : FO-

PLR (%)		5			10			15			20		
Résolution	Séquence	KL	S	KL+S	KL	S	KL+S	KL	S	KL+S	KL	S	KL+S
352x288	Soccer	81	91	92	79	85	87	77	79	83	75	76	81
	Foreman	86	93	94	84	87	90	82	83	87	78	80	85
1920x1080	Pedestrian Area	84	90	92	83	87	88	81	83	85	79	81	83

TABLE 1 – Précision des solutions Kullback-Lleibler (KL) et "Marqueur de début du slice" (S), exprimée en % de la méthode "Oracle". KL+S correspond à la combinaison des deux méthodes. 100% signifie que la solution donne les mêmes résultats que l'Oracle.

PLR (%)		5	10	15	20
Resolution	Sequence	BD-PSNR (dB)			
352x288	Soccer	1.94	4.37	5.24	5.57
	Foreman	1.65	4.15	4.9	5.91
1920x1080	Pedestrian Area	6.3	8.18	9.09	6.57

TABLE 2 – Gain BD-PSNR du MDC par rapport à SDC/R pour différentes séquences sous différents taux de perte de paquets (PLR)

REMAN et SOCCER à la résolution CIF (300 images) et PEDESTRIAN AREA à la résolution FHD (100 images). Afin d'évaluer la performance des solutions de décodage MDC proposées (sélection de slice et distance de Kullback-Leibler), nous utilisons comme référence une méthode "Oracle" pour choisir dans quelle description récupérer le CTU courant pour la description centrale.

Les résultats sont présentés dans le tableau 1. Ils correspondent à un pourcentage de performance par rapport à la solution Oracle. Prenons exemple le cas de la séquence FOREMAN, les performances pour de l'approche basée sur KL varient entre 78% et 82% en fonction du taux de perte de paquet (PLR) alors que pour la solution basée sur la position de l'entête du slice nous obtenons un résultat qui varie entre 78% et 93%. La combinaison des deux approches améliore encore la performance qui atteint 86% à 94% selon le PLR. Par conséquent, la combinaison des deux approches est utilisée à la suite pour évaluer la performance de notre méthode avec SDC.

Le tableau 2 montre le gain en BD-PSNR (PSNR calculé par la méthode de Bjøntegaard [8]) produit par la solution combinée par rapport à la solution SDC/R, pour les différentes séquences vidéos. D'après le tableau 2, notre méthode MDC peut améliorer de 1,65 dB le BD-PSNR pour un taux de perte $p_r = 0,05$ et jusqu'à 6 dB pour $p_r = 0,2$ par rapport au SDC/R pour la séquence CIF comme FOREMAN. De plus, avec la séquence FHD PEDESTRIAN AREA, l'impact de la perte de paquets est plus important; notre MDC permet de gagner de 6,3 dB avec $p_r = 0,05$ à 9,09 dB avec $p_r = 0,15$ de gain BD-PSNR par rapport au SDC.

4 Conclusion

Dans cette étude, nous avons proposé une solution de codage à descriptions multiples dans le domaine spatial adaptée à la norme HEVC. Cette solution consiste à attribuer les différents paramètres de quantification pour les CTU dans une image en fonction du niveau de redondance souhaité. Notre MDC montre une performance supérieure comparé à l'approche SDC pour un taux de perte de paquets élevé. De plus, la méthode montre

une meilleure capacité de résistance aux bruits pour du contenu HD, qui est plus sensible aux pertes. Cette capacité de résilience rend cette solution adaptée à la transmission vidéo en temps réel pour les véhicules semi-autonomes pilotés à distance.

Dans nos travaux futurs, nous souhaitons intégrer dans le schéma MDC la prédiction temporelle, ce qui permettrait d'obtenir de meilleures performances débit/distorsion que la solution basée sur les images Intra uniquement. Il est aussi important de noter que le flux vidéo ne subit pas seulement des pertes de paquets, mais aussi des retards qui peuvent influencer les performances du décodage MDC. Par conséquent, une conception de tampon au niveau d'un protocole de bas niveau est envisagée pour améliorer le système actuel.

Références

- [1] Gamal A. et Cover T. *Achievable rates for multiple descriptions*. IEEE Transactions On Information Theory, 1982
- [2] Pereira M., Antonini M. et Barlaud M. *Multiple description coding for Internet video streaming*. ICIP, 2003
- [3] Tillo T., Grangetto M. et Olmo G. *Redundant Slice Optimal Allocation for H.264 Multiple Description Coding*. IEEE Trans. Circuits Syst. Video Technol, 2008
- [4] Hsiao, C. & Tsai, W. *Hybrid Multiple Description Coding Based on H.264*. IEEE Trans. Circuits Syst. Video Technol, 2010
- [5] Ohm J., Sullivan G., Schwarz, H., Tan T. et Wiegand T. *Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)*. IEEE Transactions On Circuits And Systems For Video Technology, 2012
- [6] Hasslinger G. et Hohlfeld O. *The Gilbert-Elliott Model for Packet Loss in Real Time Services on the Internet*. 14th GI/ITG Conference - Measurement, Modelling And Evaluation Of Computer And Communication Systems, 2008
- [7] strukturag/libde265 <https://github.com/strukturag/libde265>, visité le 20 Mars 2022
- [8] Bjøntegaard G. *Calculation of average PSNR differences between RD-curves (VCEG-M33)*. VCEG Meeting (ITU-T SG16 Q. 6). (2001)