

Vers un contrôle de la consommation de décodage VVC

Pierre-Loup Cabarat

Daniel Ménard

Oussama Hammami

Hafssa Boujida

Université de Rennes, INSA Rennes, CNRS, IETR - UMR 6164

{pcabarat, dmenard, ohammami, hboujida}@insa-rennes.fr

Résumé

Tandis que l'impact de l'homme sur le climat est désormais indéniable, les projections de l'évolution de la consommation de contenus vidéo sont toujours à la hausse. L'objectif de cet article est de présenter les dernières avancées effectuées vers un décodage VVC logiciel basse consommation. Dans un premier temps nous justifions l'attention portée vers le décodage logiciel. Nous présentons ensuite une première construction théorique d'un simple régulateur de fréquence de décodage, puis nous en projetons les performances en appliquant à des mesures concrètes de temps de décodages obtenues avec openVVC.

Cet article fait suite à l'article intitulé "Pictures Decoding Time Estimation for Low-Power VVC Software Decoding" présenté à la 32ième conférence European Signal Processing Conference à Lyon

Mots clefs

VVC, Décodage vidéo, basse consommation, DVFS, Obsolescence

1 Introduction

Tandis que l'impact de l'homme sur le climat est désormais indéniable [1], les projections de l'évolution de la consommation de contenus vidéo sont toujours à la hausse [2]. Outre l'appétit croissant des humains pour des services vidéo allant de la vidéo à la demande aux applications de visio-conférence, le dimensionnement des réseaux de télécommunications doit aussi faire face à l'arrivée d'usages nouveaux tel que le codage pour les machines, les vidéos volumétriques, la conduite autonome et beaucoup d'autres. Afin de s'attaquer à la problématique de saturation de la bande passante, les organismes de normalisation poursuivent l'étude et la publication de nouvelles normes de codage toujours plus performantes en compression, mais également plus complexes [3, 4]. En parallèle des évolutions apportées aux méthodes "traditionnelles", l'essor récent des méthodes d'apprentissage artificiel a permis l'émergence de nouvelles méthodes de codage. Désormais, nous avons d'un côté des méthodes d'apprentissage artificielles extrêmement performantes en compression mais très énergivores, et d'un autre côté des méthodes traditionnelles à la consommation plus raisonnable, mais moins efficace en terme de compression.

S'il peut arriver de coder une vidéo à des fins d'archivages sans jamais être décodée par la suite, il apparaît raisonnable de supposer que ce n'est pas la norme, ou que l'utilité de ce stockage est discutable. Par conséquent il semble acceptable de supposer qu'une vidéo est en général décodée de multiples fois. Ceci est d'autant plus vrai si on considère la présence de vidéos à la demande ciblant des millions d'utilisateurs ainsi que les usages rapportés dans [2]. Dès lors quand bien même les décodeurs sont conçus pour être bien moins complexes que les encodeurs il semble intéressant de se situer du côté du décodeur quant à l'optimisation de la consommation énergétique des méthodes de compression vidéos. Un gain même infime mis à l'échelle du nombre d'utilisateurs des services vidéos il semble intéressant de chercher à réduire du côté du récepteur et donc du décodeur. En ce qui concerne l'évolution des méthodes de codage deux tendances se dessinent. Tandis que les codeurs IA peuvent se permettre le sacrifice d'un peu d'efficacité de codage, afin de gagner en efficacité énergétique, les méthodes de codage traditionnelles gagnent en efficacité de codage au détriment de leur efficacité énergétique. Si les fabricants de puces s'intéressent de près à la conception de composants spécialisés à l'utilisation de réseaux de neurones profonds. Les résultats quoiqu'encourageant semblent encore très éloignés des besoins pour une utilisation à l'échelle du grand public. Au final ces systèmes semblent pour le moment soit trop spécialisés et par définition peu versatiles, soit trop peu performants pour une utilisation adaptée aux méthodes de l'état de l'art. C'est pourquoi le reste de l'étude se concentre sur les méthodes de codage "traditionnelles" et plus particulièrement la norme VVC.

Les premières analyses sur le cycle de vie des smartphones [5] semblent présager de la domination de la quantité de gaz à effet de serre par rapport à celle induite par son utilisation. La durée moyenne de conservation d'un smartphone en France est actuellement de 3 ans et demi [6], avec une part importante des renouvellements dus à l'obsolescence des technologies utilisées. En complément des adaptations socio économiques à adopter [7], il apparaît nécessaire d'étudier les stratégies de lutte contre l'obsolescence des matériels afin de diminuer l'impact des Technologies de l'Information et de la Communication sur le climat. La souplesse offerte par le décodage logiciel semble une alternative intéressante et immédiate contre l'obsolescence des

terminaux de réception. Elle permet de surcroît de continuer à garantir une utilisation de la bande passante réduite. Par conséquent elle permet donc également de limiter les opérations de redimensionnement des infrastructures réseaux dans un contexte de hausse des échanges.

Dès lors il convient de s'intéresser aux technologies et plus spécifiquement à VVC qui est l'état de l'art des normes en vigueur. La problématique est donc : Peut-on proposer un décodage efficace sur des technologies de compression telles que la norme VVC ?

2 État de l'art

[8, 9, 10, 11] s'intéressent à l'optimisation des stratégies et configurations de codage afin de diminuer la complexité du décodage des vidéos. Plus récemment [12] tente d'estimer le nombre de coeurs nécessaire à la tâche de décodage.

[13, 14], cherchent à estimer avec précision et parcimonie le temps nécessaire au décodage d'une image afin de régler au plus finement la

Le reste de cet article propose et projette les résultats d'une méthode en complément à [14].

3 Experimentation

Pour une lecture fluide de la vidéo la consigne principale est donnée par la fréquence d'affichage f_p , qui contraint le temps T_c dont le système dispose pour présenter l'image suivante dans l'ordre d'affichage.

$$T_c = \frac{1}{f_p} \quad (1)$$

Cette contrainte temporelle n'impose nullement de décoder chacune des images en un temps inférieur à T_c . Il est en effet possible d'ajuster une mémoire tampon pouvant contenir N images afin d'absorber les temps de décodage supérieurs à la moyenne. Pour une fréquence d'affichage f_p donnée, la somme S_k des temps nécessaires au décodage des images d'un tampon de taille N ne doit pas excéder :

$$S_k = \sum_{l=k-N}^k T_l \leq N \times T_c \quad (2)$$

Afin de proposer un ordonnancement efficace il convient de pouvoir estimer finement le temps nécessaire à la réalisation des tâches de décodage. Outre l'intérêt porté à la précision de l'estimation du temps de décodage d'une image il convient également de considérer le temps réservé à l'inférence de cette dernière. En effet l'inférence doit également être réalisée dans le temps imparti.

$$\sum_{l=k-N}^k T_k(l) + \sum_{l=k-N}^k T_{ik}(l) \leq N \times T_c \quad (3)$$

Pour une séquence de 60 Hz la prochaine image à afficher doit être disponible en moins de 16.7 ms. Arbitrairement la limite a été fixé à 1 %, ce qui nous donne un temps d'inférence cible moyen de l'ordre de 167 μ s. Les temps

d'inférences obtenus dans [14] semblent indiquer qu'il est possible d'obtenir une inférence correcte en un temps de l'ordre de 1 μ s, nous négligerons donc par la suite le temps nécessaire à l'inférence, d'où l'approximation :

$$\sum_{l=k-N}^k T_k(l) \leq N \times T_c \quad (4)$$

Tandis que T_c correspond à un temps de référence indépendant de la fréquence de décodage, les temps de décodages T_k dépendent de la fréquence d'horloge du processeur durant le décodage de l'image k . Par conséquent la somme S_k varie en fonction de l'évolution des fréquences f_k de décodages utilisées pour chacune des images k d'où la notation :

$$S_k(F_k) \leq N \times T_c \quad (5)$$

où $F_k = (f_{k-N}, \dots, f_k)$ correspond aux fréquences utilisées pour les décodages de chaque image k .

Notons T_k^* le temps résultant du décodage de l'image k à la fréquence f_k^* . Ou f_k^* correspond à la fréquence qui minimise la différence Δ_{S_k}

$$\Delta = N \times T_c - S_k(F_k) \quad (6)$$

Au final on cherche un vecteur $F_k^* = (f_{k-N}^*, \dots, f_k^*)$ qui minimise Δ_{S_k} . Une solution triviale serait de faire en sorte que $T_k(f_k) = T_c$ c'est-à-dire,

$$f_k^* = f_n \times \frac{\hat{T}_k}{T_c} \quad (7)$$

Cependant on peut réécrire :

$$S_k = \sum_{l=k-N}^{k-1} T_l + T_k = S_{k-1} + T_k - T_{k-(N+1)} \quad (8)$$

Dans 9 on remarque que seul T_k est inconnu au moment de l'ajustement de fréquence. On peut donc chercher à répartir l'estimation \hat{T}_k , au plus proche de

$$N \times T_c - (S_{k-1} - T_{k-(N+1)}) \quad (9)$$

Le potentiel de diminution de fréquence du processus de décodage sera d'autant plus élevé que l'estimation \hat{T}_k sera proche de T_k .

Annexe

Puisqu'il n'est pas possible de le déposer à côté du premier, le caméra ready de l'article [14] intitulé "Pictures Decoding Time Estimation for Low-Power VVC Software Decoding" présenté à la 32^{ème} conférence European Signal Processing Conference peut-être transmis sur demande aux auteurs.

Références

- [1] Hoesung Lee, Katherine Calvin, Dipak Dasgupta, Gerhard Krinner, Aditi Mukherji, Peter Thorne, Christopher Trisos, José Romero, Paulina Aldunce, Ko Barrett, et al. *Ipcc, 2023 : Climate change 2023 : Synthesis report, summary for policymakers. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change [core writing team, h. lee and j. romero (eds.)]. ipcc, geneva, switzerland. 2023.*
- [2] U Cisco. Cisco annual internet report (2018–2023) white paper. *Cisco : San Jose, CA, USA*, 10(1) :1–35, 2020.
- [3] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J Sullivan, et Jens-Rainer Ohm. Overview of the versatile video coding (vvc) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10) :3736–3764, 2021.
- [4] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, et Thomas Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12) :1649–1668, 2012.
- [5] Mine Ercan, Jens Malmödin, Pernilla Bergmark, Emma Kimfalk, et Ellinor Nilsson. Life cycle assessment of a smartphone. Dans *Proceedings of ICT for Sustainability 2016*, pages 124–133. Atlantis Press, 2016/08.
- [6] Léa Mosesso, Nolwenn Maudet, Edlira Nano, Thomas Thibault, et Aurélien Tabard. Obsolescence Paths : living with aging devices. Dans *ICT4S 2023 - International Conference on Information and Communications Technology for Sustainability*, Rennes, France, Juin 2023.
- [7] Ilan Noy. Economists are not engaged enough with the ipcc. *npj Climate Action*, 2(1) :33, 2023.
- [8] Matthias Kränzler, Adam Wieckowski, Geetha Ramasubbu, Benjamin Bross, André Kaup, Detlev Marpe, et Christian Herglotz. Optimized decoding-energy-aware encoding in practical vvc implementations. Dans *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1536–1540, 2022.
- [9] Matthias Kränzler, Christian Herglotz, et André Kaup. Decoding energy modeling for versatile video coding. Dans *2020 IEEE International Conference on Image Processing (ICIP)*, pages 3144–3148, 2020.
- [10] Christian Herglotz, Andréas Heindel, et André Kaup. Decoding-energy-rate-distortion optimization for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(1) :171–182, 2019.
- [11] Matthias Kränzler, Christian Herglotz, et André Kaup. Energy efficient video decoding for vvc using a greedy strategy-based design space exploration. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7) :4696–4709, 2022.
- [12] O. Le Gonidec, M. Chavarrías, A. Saha, G. Rosa, et F. Pescador. Energy efficient versatile video coding decoder using lightweight regression models. Dans *2023 26th Euromicro Conference on Digital System Design (DSD)*, pages 46–52, 2023.
- [13] Hafssa Boujida, Pierre-Loup Cabarat, et Daniel Menard. Decoding time prediction for versatile video coding. Dans *2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2023.
- [14] Pierre-Loup Cabarat, Houssama Hammani, Hafssa Boujida, et Daniel Menard. Pictures decoding time estimation for low-power vvc software decoding. Dans *2024 European Signal Processing Conference (EUSIPCO)*, pages 1–6, 2024.