

# ANÁLISE DE COMPLEXIDADE ESCALÁVEL PARA A TRANSCODIFICAÇÃO DE VÍDEO DE HEVC PARA AV1

ALEX BORGES<sup>1</sup>; BRUNO ZATT<sup>1</sup>; MARCELO PORTO<sup>1</sup>; GUILHERME CORREA<sup>1</sup>

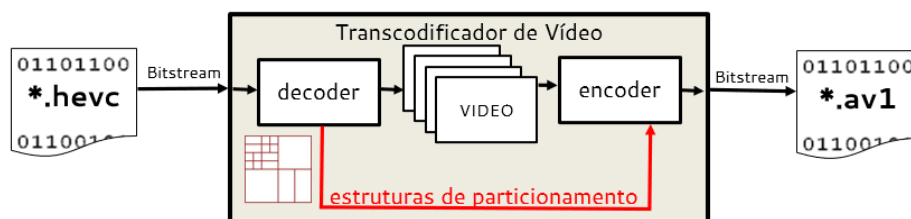
<sup>1</sup>Universidade Federal de Pelotas – {amborges, zatt, porto, gcorrea}@inf.ufpel.edu.br

## 1. INTRODUÇÃO

Ao longo das últimas décadas, conteúdo em vídeo tem crescido em popularidade, principalmente após o advento da internet. O H.264 (ITU, 2003) é o formato de codificação de vídeo mais utilizado na indústria, estando presente em 92% dos dispositivos e sistemas (BITMOVIN, 2020). Contudo, essa é uma tecnologia de 2003 e, desde então, muitos outros formatos de codificação de vídeo surgiram. Como o HEVC (ITU, 2013), que foi lançado em 2013 e é o sucessor do H.264, e o AV1 (HAN, 2021), lançado em 2018 com proposta livre de royalties. De acordo com alguns trabalhos publicados na literatura, como CHEN (2018), o AV1 apresenta uma eficiência de compressão de dados 22,75% maior que o HEVC.

Redução de custos de royalties ou maior eficiência de compressão de dados são razões para migração de tecnologias de vídeo. Essa migração se dá com um processo em cascata, conforme a Figura 1. Nesse processo, um software decodifica de vídeos codificados de um formato, obtendo-se o vídeo bruto e, depois, recodifica-se esse vídeo com o codificador desejado. No entanto, devido ao elevado custo computacional para realizar esse processo, formas de acelerar essa transcodificação são encontradas na literatura, tais como reaproveitar informações provenientes do decodificador para acelerar tomadas de decisão do novo codificador. A Figura 1 também representa a transcodificação acelerada, através dos fluxos destacados em vermelho.

Neste resumo iremos apresentar um artigo que foi publicado no início deste ano em uma revista internacional de Qualis B2 (BORGES, 2021), de forma mais simplificada e de fácil compreensão para leitores não especializados.



**Figura 1.** Visão em alto nível de um transcodificador de vídeo

## 2. METODOLOGIA

Há diversas formas de se realizar o reaproveitamento de informações, todavia, percebe-se pela literatura científica a preferência pelo reaproveitamento de estruturas de particionamento. Um vídeo é dividido em quadros e que, por sua vez, é dividido em blocos de codificação. Cada um desses blocos pode ser subdividido em uma estrutura menor que permite o processamento das previsões e a codificação geral do vídeo. Esse tipo de divisão permite que o codificador aplique diferentes técnicas de codificação para regiões diferentes do vídeo, a fim de se obter a maior taxa de compressão de dados possível. Há vantagens em se

reaproveitar esse tipo de informação ao se lidar com formatos de codificação diferentes, pois as estruturas de particionamento são mais dependentes das informações contidas na imagem em si do que das técnicas de codificação.

Cada formato de codificação de vídeo possui a sua própria forma de distribuir e organizar os blocos, mas todas elas possuem formatos de blocos quadráticos de tamanho  $2^N \times 2^N$  píxeis, onde  $N$  pode assumir qualquer inteiro positivo. No HEVC a estrutura de particionamento contém o seguinte conjunto de valores de  $N$ : 3, 4, 5, 6. Já no AV1, o  $N$  pode assumir os valores: 2, 3, 4, 5, 6, 7. Desta forma, é possível associar diretamente os blocos quadráticos que o HEVC escolheu para que o AV1 também opte por esses tamanhos, acelerando o processo de busca pela melhor distribuição de blocos no quadro. No entanto, o AV1 possui outros tamanhos de blocos quadráticos ( $2^2 \times 2^2$  e  $2^7 \times 2^7$ ) que não são observáveis no HEVC. Em BORGES (2021) é explicitado as relações existentes entre as escolhas de tamanhos de blocos feitas pelo HEVC e pelo AV1, deixando claro que se o AV1 utilizar combinações de valores de  $N$  próximos ao observado no HEVC pode trazer vantagens na transcodificação. Em outras palavras, caso uma região do vídeo o HEVC optou por utilizar o tamanho  $N=4$ , permitir ao AV1 utilizar as combinações  $N=\{4, 5\}$  apresentará resultados diferentes se optarmos por  $N=\{2, 3, 4\}$ .

Há algumas combinações de conjuntos de blocos quadráticos que o AV1 poderá testar baseado no  $N$  observado no HEVC. Cada um desses conjuntos é identificado por um limitante superior (*Levels above,  $L_a$* , no trabalho original) e um limitante inferior (*Levels below,  $L_b$* ). Tanto o  $L_a$  como o  $L_b$  partem de zero, onde nenhum  $N$  maior ou menor que o observado no HEVC é permitido no AV1. Por outro lado, quatro é o valor máximo de  $L_a$  e  $L_b$ , pois é o valor da diferença entre o menor  $N$  do HEVC com o maior  $N$  do AV1, para  $L_a$ , e o maior  $N$  do HEVC subtraído do menor  $N$  do AV1, para o  $L_b$ . Assim, utilizando a definição  $L_a:L_b$  podemos identificar 25 conjuntos limitantes, onde o conjunto de blocos quadráticos que o AV1 poderá utilizar é definido por  $N+L_a$  até  $N-L_b$ . Por exemplo, ao utilizar os limitantes 2:1, se for observado na decodificação do HEVC a utilização de um bloco  $N=4$ , isso significa que o AV1 estará habilitado para utilizar algum dos blocos cujos  $N$  sejam  $\{6, 5, 4, 3\}$ . É preciso esclarecer que o AV1 também poderá utilizar os seus blocos não-quadráticos, desde que estejam pertencentes ao conjunto limitante.

### 3. RESULTADOS E DISCUSSÃO

De forma a possibilitar a verificação dessa análise de vários  $L_a:L_b$  em uma transcodificação de vídeo de HEVC para AV1, um experimento foi realizado utilizando a transcodificação em cascata para contrapor a transcodificação acelerada. Para tanto, foram utilizados os softwares referência do HEVC e AV1, conforme podem ser vistos no trabalho original (BORGES, 2021).

As transcodificações foram realizadas sobre 60 quadros de quatro vídeos de alta definição 1080p. O resultado desse experimento pode ser observado na Tabela I, que apresenta a variação  $L_a:L_b$  utilizada e os valores do percentual de perda da eficiência de codificação, conforme a métrica Bjøntegaard Delta (BD)-rate (BJØNTEGAARD, 2001), e organizados pelo percentual de aceleração obtido na transcodificação ao utilizar aquele  $L_a:L_b$ . Observa-se que diferentes resultados são observados para diferentes variações de  $L_a:L_b$ , inclusive para valores próximos, como 1:1 e 1:2, que ao permitir um nível a mais de  $N$  no limite inferior, tanto a aceleração obtida como a perda na eficiência de codificação são reduzidos em quase a terça parte. Como era esperado, o se permitir todos os  $N$  no AV1 (4:4) não se observa diferença quando comparado à transcodificação em cascata, contudo,

**Tabela I.** Média dos resultados das transcodificações aceleradas

<b><i>La:Lb</i></b>	<b>BD-Rate (%)</b>	<b>Redução de Tempo</b>
0:0	14.6688	61.22
2:0*	5.0935	40.81
3:0*	5.0628	40.70
4:0*	5.0215	40.02
1:0	6.1817	39.82
0:1	6.1688	31.98
1:1*	1.1950	21.02
0:2	3.9088	19.95
2:1*	0.8708	17.98
3:1*	0.8703	17.70
4:1	0.9020	17.69
0:3	3.5850	15.43
0:4	3.5125	15.11
1:2*	0.4263	8.02
2:2*	0.1145	5.00
3:2*	0.0645	4.65
4:2*	0.0583	4.59
1:3	0.3308	3.53
1:4	0.3180	3.50
2:3*	0.0070	0.48
3:3*	0.0043	0.36
2:4	0.0050	0.34
4:3	0.0493	0.28
3:4*	-0.0128	0.23
4:4*	0.0000	0.00

\* selecionada pela curva de Pareto

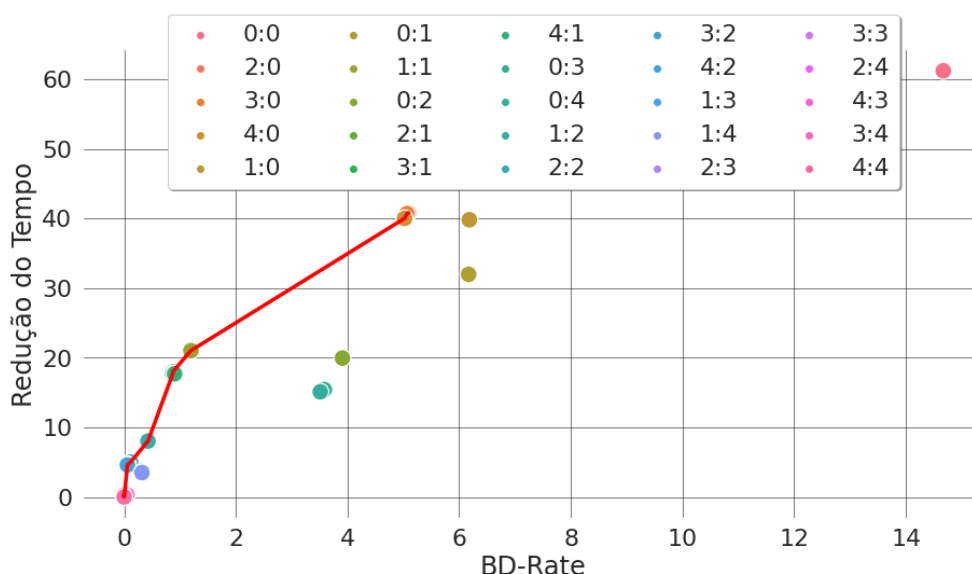
permitir somente o mesmo valor de  $N$  observado (0:0) a aceleração da transcodificação atinge uma média de 61% a um custo de perder quase 15% da eficiência de codificação. Assim sendo, enquanto a solução 4:4 não se distingue do codificador original, a 0:0 deve ser descartada como opção válida de transcodificação, dado o histórico que se é observado na literatura científica sobre custos toleráveis de BD-Rate.

Com esse experimento é possível identificar a eficiência de cada um dos casos, todavia, não nos informam corretamente as variações de  $La:Lb$  que merecem serem considerados como uma opção de transcodificação rápida. Para isso, colocamos esses dados em um gráfico de dispersão e aplicamos a curva de Pareto sobre esses valores. Basicamente, a curva de Pareto seleciona aqueles casos que apresentam a maior redução de tempo com o menor custo na perda de eficiência de codificação para cada grau de intensidade testada. Isso fica mais evidente ao olharmos a Figura 2, ligando aquelas soluções que se encaixam na definição da curva de Pareto. Na Tabela I foram destacados os  $La:Lb$  dessa curva.

## 4. CONCLUSÕES

Este trabalho visou apresentar um resumo do artigo publicado em BORGES (2021), que propôs uma transcodificação de vídeo acelerada do formato HEVC para o AV1 fazendo reaproveitamento de estruturas de particionamento através do conceito de  $La:Lb$ , que determina quantos níveis (acima e abaixo) de profundidade da árvore de particionamento o AV1 poderá utilizar dado o nível de profundidade observado na árvore de particionamentos do HEVC. Além de identificar o custo e benefício de cada uma das variantes de  $La:Lb$ , também foi necessário utilizar a curva de Pareto para identificar quais soluções são válidas de serem utilizadas. Desta forma, das 25 possibilidades de reaproveitamento, sugerem-se apenas 14, que varia entre não utilizar a versão acelerada da transcodificação até possibilitar uma aceleração de 40,81% a um custo de 5,09% na perda da eficiência de codificação do AV1.

No trabalho completo apresentado em BORGES (2021) também foi apresentado uma solução escalável para alternar entre as soluções de  $La:Lb$



**Figura 2.** Relação BD-Rate x Redução de Tempo, a Curva de Pareto está em vermelho.

destacadas pela curva de Pareto, além de apresentar essas o impacto dessas variações de La:Lb sob outras sequências de vídeos de mesma resolução.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

BITMOVIN, I. **Bitmovin's 4th Annual Video Developer Report 2020**. Bitmovin, 2020. Acessado em 22 de jul. 2021. Disponível em: <https://go.bitmovin.com/video-developer-report-2020>

BJØNTEGAARD, G. **Calculation of average PSNR differences between RDcurves**. ITU, 28 mar. 2001. Acessado em 22 de jul. 2021. Disponível em: [http://wftp3.itu.int/av-arch/video-site/0104\\_Aus/VCEG-M33.doc](http://wftp3.itu.int/av-arch/video-site/0104_Aus/VCEG-M33.doc)

BORGES, A.; ZATT, B.; PORTO, M.; CORREA, G. Complexity-scalable HEVC-to-AV1 video transcoding based on partition inheritance. **Journal of Real-Time Image Processing**, 2 abr. 2021. DOI: 10.1007/s11554-021-01101-2

CHEN, Y., et al. An overview of core coding tools in the AV1 video codec. **Picture Coding Symposium** (PCS 2018), p. 41–45, 2018. DOI: 10.1109/PCS.2018.8456249

HAN, J. et al. A Technical Overview of AV1. **Proceedings of the IEEE**, p.1–28, 2021. doi: 10.1109/JPROC.2021.3058584.

INTERNATIONAL TELECOMMUNICATION UNION. **Recommendation H.264: Advanced Video Coding for Generic Audiovisual Services**. ITU, mai. 2003. Acessado em 22 de jul. 2021. Disponível em: <https://www.itu.int/rec/T-REC-H.264>

INTERNATIONAL TELECOMMUNICATION UNION. **Recommendation H.265: High Efficiency Video Coding**. ITU, abr. 2013. Acessado em 22 de jul. 2021. Disponível em: <https://www.itu.int/rec/T-REC-H.265>