POSTER: Caracterização Estatística do Tempo de Decodificação de Ladrilhos de Vídeos 360°

Henrique Domingues Garcia
Departamento de Engenharia Elétrica
Universidade de Brasília
Brasília, Brazil
henriquedgarcia@gmail.com

Mylène C.Q. Farias
Departamento de Engenharia Elétrica
Universidade de Brasília
Brasília, Brazil
mylene@ene.unb.br

Marcelo M. Carvalho
Departamento de Engenharia Elétrica
Universidade de Brasília
Brasília, Brazil
mmcarvalho@ene.unb.br

***Resumo*—A popularização de vídeos 360° trouxe consigo desafios significativos para sua transmissão eficiente pela Internet via Dynamic Adaptive Streaming over HTTP (DASH) utilizando a técnica de *ladrilhamento* ("tiling") e segmentação temporal do vídeo em diferentes qualidades. Em particular, as exigências significativas de largura de banda e tempo exíguo para predição do movimento da cabeça do usuário apontam para a necessidade de uma modelagem acurada do tempo de decodificação dos ladrilhos de vídeos 360°. Esta modelagem permite a requisição otimizada de ladrilhos de vídeo, considerando a limitação de largura de banda e espaço de armazenamento no lado cliente da aplicação. Neste sentido, este trabalho apresenta uma caracterização estatística preliminar do tempo de decodificação de segmentos (*chunks*) de diferentes conteúdos de vídeos 360°, particionados em ladrilhos de diferentes dimensões e codificados com diferentes taxas de bits.**

***Index Terms*—Vídeo omnidirecional, Vídeo 360°, streaming de vídeo, HEVC, DASH.**

I. Introdução

A visualização de vídeos 360° (i.e., esférico ou omnidirecional) é uma das aplicações mais populares de realidade virtual (VR) dos últimos anos. Sua adoção tem crescido rapidamente com a evolução dos aparelhos de telefonia móvel, a disseminação de dispositivos *head-mounted display* (HMD) e de câmeras de captura em 360°. De fato, estima-se que as aplicações VR gerem aproximadamente 254 Petabytes de dados por mês em 2022, quase doze vezes mais do que em 2017 [1]. Atualmente, a transmissão de vídeo 360° pela Internet faz uso da arquitetura tradicional para transmissão de vídeo 2D, ou seja uma aplicação cliente solicita ao servidor segmentos (ou *chunks*) pré-armazenados do vídeo de uma certa qualidade via *Dynamic Adaptive Streaming over HTTP* (DASH) [2]. Para a codificação, primeiramente, toda a esfera do vídeo 360° é projetada em um plano e, em seguida, um codificador de vídeo como HEVC é usado para comprimir o vídeo. Na maioria dos sistemas atuais, toda a esfera é enviada para o cliente. No entanto, o usuário não consegue visualizar todas as direções ao mesmo tempo. Tipicamente, apenas 16,6% da esfera é visualizada a cada instante no *viewport* [3]. Consequentemente, uma fração significativa de largura de banda e armazenamento são desperdiçados. Esta situação é ainda agravada pelo fato de que o vídeo 360° precisa ser codificado em alta resolução para que a cena apresentada no *viewport* tenha uma qualidade aceitável.

Para atenuar este problema, uma técnica denominada *ladrilhamento* espacial ("*tiling*") dos vídeos foi proposta com objetivo de reduzir a largura de banda necessária para esta aplicação. Cada quadro do vídeo é particionado em ladrilhos que são decodificados e reproduzidos de forma independente. Desta forma, o cliente requere apenas os ladrilhos correspondentes a uma região de interesse (ROI – *Region of Interest*) para compor o *viewport*. Baseados nesta técnica, diversos trabalhos propuseram técnicas adaptativas transmissão dos ladrilhos referentes ao *viewport* do HMD [4]. Os segmentos solicitados e recebidos são armazenados no *buffer* do cliente para decodificação e apresentação.

Infelizmente, os atuais preditores de direção de visualização do usuário apresentam desempenho aceitável (92%) apenas para janelas curtas de predição (aprox. 1 segundo). Desta forma, o tempo de decodificação dos ladrilhos é um parâmetro fundamental para o projeto de decodificação escalonada nos atuais sistemas de transmissão adaptativa de vídeos 360° pela Internet [4]. Mais especificamente, o tempo de decodificação dos ladrilhos tem impacto significativo na quantidade e qualidade dos ladrilhos solicitados para uma dada largura de banda. Atualmente, este tempo de decodificação tem sido estimado grosseiramente, sem maior discussão e aprofundamento do seu comportamento estatístico para as diferentes escolhas de dimensionamento de ladrilhos, taxas de codificação e natureza do conteúdo 360°. Neste sentido, nosso trabalho apresenta uma primeira caracterização estatística do tempo de decodificação de segmentos de vídeos 360° particionados em ladrilhos.

II. Objetivos

Este trabalho tem como objetivo apresentar uma caracterização estatística preliminar do tempo de decodificação de segmentos (*chunks*) de vídeos 360° particionados em ladrilhos de diferentes dimensões, codificados com diferentes taxas pelo HEVC e sobre diferentes conteúdos de vídeos 360°.

III. Materiais e Métodos

Foram selecionados 12 vídeos 360°, em projeção equirretangular e com duração de 60 segundos. Para garantir a diversidade de conteúdo e, consequentemente, a complexidade de codificação, os vídeos foram selecionados, de acordo com a mediana da sua informação espacial (SI) e informação

Tabela I
TAXA DE CODIFICAÇÃO MÉDIA E SI/TI DOS VÍDEOS SELECIONADOS

Taxa (Mbps) - CRF25	Vídeo	SI	TI
1.380	om_nom	103.37	1.81
2.491	pluto	24.64	1.95
3.3989	masquerade_ball	73.46	2.81
3.867	super_mario	117.95	1.18
19.180	lions	39.44	14.34
17.016	elephants	78.55	9.88
8.068	ski	59.82	4.27
12.433	clans	30.26	14.54
12.640	surf	67.14	15.30
12.945	manhattan	29.08	14.26
18.224	rollercoaster	72.17	23.51
12.508	venice	24.10	12.13

temporal (TI). A Tabela I apresenta um resumo dos valores de SI/TI e das taxas de bit médias correspondentes a um valor comum de CRF (*Constant Rate Factor*) igual a 25 (que reflete a complexidade de codificação) para todos os vídeos considerados. Os vídeos foram codificados utilizando HEVC (i.e., H.265) com resolução 4K, razão 2:1, 30 fps, e 4 valores de CRF: 15, 25, 35 e 45. Além disso, os vídeos foram particionados em 4 padrões de ladrilhos (1×1, 3×2, 6×4, 12×8) e *chunks* de 1 segundo, totalizando 365.760 *chunks*.

Foram realizadas três medições do "*user time*" da decodificação de cada *chunk*, com uma única *thread*. "*User time*" é o tempo gasto em operações fora do *kernel*, como operações matriciais, por exemplo, sem considerar alocação dinâmica de memória e outras chamadas de sistema. Para decodificação dos vídeos, utilizamos um computador *desktop* com processador i7-4770 3.4 GHz com 16 GB de RAM, Ubuntu 18.04.2 LTS e FFmpeg 4.1. Embora não seja a plataforma mais comum de visualização de vídeos 360°, esta primeira caracterização visa obter o comportamento estatístico "típico". Medições futuras serão baseadas em plataformas específicas.

IV. RESULTADOS PRELIMINARES

A Figura 1 apresenta histogramas dos tempos de decodificação de cada *chunk* de todos os vídeos considerados, contendo um único ladrilho para cada tipo de particionamento. Para caracterização do tempo de decodificação, avaliamos diversas distribuições de probabilidade, ajustadas aos dados segundo o critério de máxima verossimilhança. O critério de seleção das distribuições baseou-se na soma de erros quadráticos (*Sum of Squared Error* - SSE). As distribuições com melhores resultados foram a Gaussiana Inversa (indicada por "invgauss" nos gráficos), 12ª distribuição de Burr ("bur12"), Birnbaum-Saunders ("fatiguelife") e Log-Normal ("lognorm"). O tempo médio de decodificação dos formatos 1×1, 3×2, 6×4, 12×8 foram 0,843 s, 0,137 s, 0,032 s e 0,009 s respectivamente e a variância foi de 0,432 s, 0,090 s, 0,025 s e 0,007 s respectivamente. Todos os formatos apresentam correlação linear com a taxa de bits de aproximadamente 0,95.

Observa-se que, para *tiles* de formato (1×1 e 6×4), a distribuição Gaussiana Inversa apresenta o melhor ajuste. No entanto, para os ladrilhamentos (3×2 e 12×8), a

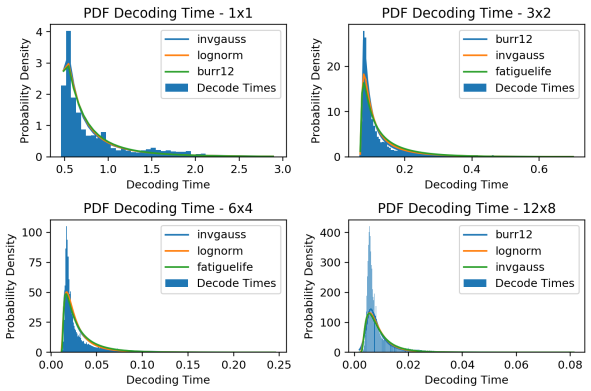

Figura 1. Histograma dos tempos de decodificação e três distribuições com menores SSE

12ªdistribuição de Burr ajusta-se melhor aos dados. Pela menor dimensão, os *tiles* dos particionamentos 6×4 e 12×8 possuem um tempo médio de decodificação que é 94,2% e 98,4% menores do que o vídeo sem ladrilhamento (1×1), respectivamente. Considerando que o usuário visualiza apenas uma fração do vídeo esférico a cada instante, e que um maior ladrilhamento resulta em crescimento apenas modesto na taxa média do vídeo codificado (ex: *tiles* 12×8 possui taxa média apenas 19,17% maior do que vídeo sem ladrilhamento), é possível então esperar ganhos significativos no tempo de decodificação total e ocupação de *buffer* se apenas os *tiles* relativos ao *viewport* predito do usuário forem solicitados do servidor.

V. CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresentou uma caracterização estatística preliminar do tempo de decodificação de ladrilhos de segmentos de vídeos 360° em diferentes dimensões e taxas de compressão. Verificou-se que a distribuição Gaussiana Inversa ajusta-se melhor ao ladrilhamento 3×2 e 12×8, enquanto que a 12ªdistribuição de Burr representa melhor o tempo de decodificação de ladrilhamentos mais refinados. Como trabalho futuro, caracterizaremos a estatística do tempo de decodificação de ladrilhos de vídeos para diferentes conteúdos de vídeo.

REFERÊNCIAS

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper," Tech. Rep., 2019. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html

[2] C. Concolato, J. Le Feuvre, F. Denoual, F. Mazé, E. Nassor, N. Ouedraogo, and J. Taquet, "Adaptive Streaming of HEVC Tiled Videos Using MPEG-DASH," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 8, pp. 1981–1992, 2018.

[3] D. He, C. Westphal, and J. J. Garcia-Luna-Aceves, "Joint Rate and FoV adaptation in immersive video streaming," in *Proc. of the Workshop on Virtual Reality and Augmented Reality Network*, 2018, pp. 27–32.

[4] F. Qian, B. Han, Q. Xiao, and V. Gopalakrishnan, "Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices," in *Proc. MobiCom*, 2018.