

Explorando Poda de Transformada na Compressão de Imagens em 360°

Enzo B. Segala
Escola de Engenharia
Universidade Federal do Rio Grande do Sul
Porto Alegre, Brasil
00335314@ufrgs.br

Thiago L. T. da Silveira
Instituto de Informática
Universidade Federal do Rio Grande do Sul
Porto Alegre, Brasil
tltsilveira@inf.ufrgs.br

Abstract—Omnidirectional media store the whole scene information and are becoming popular nowadays. Such images have high resolution to cope with quality requirements in novel applications such as immersive navigation for virtual reality. This paper proposes a JPEG-like compression method for omnidirectional images represented in equirectangular (ERP) format. Our method proposes to use transform pruning to explore the non-uniform sampling of ERP images, being adapted to the image block latitudes. Experimental results show that our approach can reduce the additive and multiplicative arithmetic costs by 25.58% and 29.96%, respectively, compared to JPEG, while achieving 3.85% higher compression and 4.1% image quality loss.

Resumo— Mídias omnidirecionais armazenam as informações de toda a cena e estão se tornando populares hoje em dia. Tais imagens têm alta resolução para atender aos requisitos de qualidade de novas aplicações, como navegação imersiva para realidade virtual. Este artigo propõe um método de compressão do tipo JPEG para imagens omnidirecionais representadas no formato equirretangular (ERP). Nosso método propõe usar poda de transformada para explorar a amostragem não-uniforme de imagens ERP, adaptando-se às latitudes dos blocos de imagem. Os resultados experimentais mostram que nossa abordagem pode reduzir os custos aritméticos aditivos e multiplicativos em 25,58% e 29,96%, respectivamente, em comparação ao JPEG, ao mesmo tempo em que alcança uma melhora na compressão de 3,85% a uma perda de qualidade de imagem 4,1%.

I. INTRODUÇÃO

Câmeras omnidirecionais – também conhecidas como esféricas, em 360° ou panorâmicas – apresentam um campo de visão (FoV) completo que permite o imageamento de toda uma cena (180° × 360°) de uma só vez [1]. Mídias omnidirecionais estão cada vez mais populares, sendo empregadas em aplicações da robótica [2], no controle de veículos autônomos [1] e, especialmente, na navegação imersiva de ambientes de realidade virtual (VR) [3].

Imagens regulares têm maior resolução espacial por unidade angular comparadas a panoramas visto que estas últimas cobrem o FoV completo. Para que usuários de VR possam experienciar um alto nível de imersão, é necessário que as imagens tenham alta resolução [4], além de que o dispositivo de visualização utilizado proveja seis graus de liberdade (6-DoF) [5].

Imagens em 360° são definidas na superfície da esfera unitária [1] e, portanto, não podem se beneficiar diretamente de algoritmos de computação visual tradicionais independentemente de seu fim [6]. Entretanto, a maioria das empresas fabricantes de câmeras em 360° e aplicações consideram a projeção equirretangular (ERP) da esfera no plano – que lembra um mapa mundi [1], [5] – para armazenamento.

O mapeamento ERP é o formato padrão de imagens esféricas [1], [4], [7], mas introduz distorções severas, proeminentes próximo aos polos [4]. Essa distorção vem da amostragem não-uniforme da esfera

quando mapeada para o plano [1]. Embora outros formatos – como mapas cúbicos – existam, não há mapeamento esfera-plano livre de deformações [8] e a troca de formatos induz perda de informação.

O objetivo desse trabalho é explorar as redundâncias em regiões de alta amostragem para compressão de imagens ERP. A abordagem proposta reduz o número de operações realizadas durante a compressão enquanto mantém tamanho de armazenamento e qualidade de imagem comparáveis as do mecanismo utilizado no padrão JPEG [9].

O restante deste artigo está organizado da seguinte forma. A Seção II comenta os trabalhos relacionados. O método proposto é introduzido na Seção III. Os resultados parciais são discutidos na Seção IV. Considerações finais são apresentadas na Seção V.

II. TRABALHOS RELACIONADOS

Esta seção discute brevemente o padrão de compressão de imagens JPEG [9] e abordagens aplicáveis a mídias em 360° que ajustam a compressão dependendo da taxa de amostragem [7], [10]–[12].

A. Compressão de Imagens Regulares por JPEG

O padrão JPEG [9] se baseia na utilização da transformada discreta do cosseno (DCT) [13] e posterior quantização para compressão de imagens. A DCT tem a propriedade de descorrelacionar os dados em domínio espacial, dadas propriedades estatísticas respeitadas por imagens naturais, agrupando sua energia em poucos coeficientes no domínio espectral [13], [14].

Mais precisamente, o padrão JPEG quebra a imagem de entrada \mathbf{I} de tamanho $A \times L$ pixels em blocos $\mathbf{I}_{i,j}$ de tamanho $N \times N$, $N = 8$:

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{1,1} & \mathbf{I}_{1,2} & \cdots & \mathbf{I}_{1,L/N} \\ \mathbf{I}_{2,1} & \mathbf{I}_{2,2} & \cdots & \mathbf{I}_{2,L/N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_{A/N,1} & \mathbf{I}_{A/N,2} & \cdots & \mathbf{I}_{A/N,L/N} \end{bmatrix}. \quad (1)$$

Cada bloco $\mathbf{I}_{i,j}$ é submetido à transformação 2D direta, utilizando

$$\mathbf{H}_{i,j} = \mathbf{C} \cdot \mathbf{I}_{i,j} \cdot \mathbf{C}^T, \quad (2)$$

onde \mathbf{C} é a matriz da DCT dada por [13]

$$c_{m,n} = \sqrt{\frac{2}{N}} \alpha_{m-1} \cos\left(\frac{\pi(m-1)(2n-1)}{2N}\right), \quad (3)$$

para $m, n = 1, 2, \dots, N$, $\alpha_0 = 1/\sqrt{2}$ e $\alpha_l = 1$ para $l \neq 0$.

O bloco transformado $\mathbf{H}_{i,j}$ é submetido à quantização através de

$$\hat{\mathbf{H}}_{i,j} = \lfloor \mathbf{H}_{i,j} \oslash \mathbf{Q} \rfloor \quad (4)$$

onde \oslash é a divisão matricial aplicada elemento-a-elemento e $\lfloor \cdot \rfloor$, também aplicado elemento-a-elemento, retorna o inteiro mais próximo de seu argumento. A matriz de quantização \mathbf{Q} é dada por

$$\mathbf{Q} = \lfloor (S\mathbf{Q}_0 + 50)/100 \rfloor, \quad (5)$$

onde

$$\mathbf{Q}_0 = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \quad (6)$$

e $S = 5000/QF$ se $QF < 50$ ou $S = 200 - 2QF$ caso contrário [9]. O parâmetro $QF \in \{0, 1, \dots, 100\}$ indica um fator de qualidade e é provido pelo usuário. Se $QF = 50$, considerado o valor padrão, então $\mathbf{Q} = \mathbf{Q}_0$. Após a quantização, os coeficiente são submetidos às codificações diferencial e de Huffman [9].

A decodificação, seguida do processo de dequantização e transformação inversa permitem restaurar uma aproximação da imagem de entrada em domínio espacial. A dequantização do bloco $\tilde{\mathbf{H}}_{i,j}$ é dada por

$$\tilde{\mathbf{H}}_{i,j} = \hat{\mathbf{H}}_{i,j} \odot \mathbf{Q}, \quad (7)$$

onde \odot representa a multiplicação matricial elemento-a-elemento. A transformação 2D inversa do bloco dequantizado $\tilde{\mathbf{H}}_{i,j}$ é dada por

$$\tilde{\mathbf{I}}_{i,j} = \mathbf{C}^\top \cdot \tilde{\mathbf{H}}_{i,j} \cdot \mathbf{C}. \quad (8)$$

O rearranjo adequado dos blocos $\tilde{\mathbf{I}}_{i,j}$ aproxima à imagem original \mathbf{I} .

B. Compressão de Imagens em 360°

Câmeras omnidirecionais capturam intensidades de luz de toda uma cena na superfície da esfera unitária [1]. Como argumentado na Seção I, por questões topológicas, é impossível explorar abordagens para computação visual tradicionais sem aplicar uma função adequada de mapeamento da esfera para o plano.

A projeção ERP mapeia a superfície da esfera unitária em um plano retangular $[-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\pi, \pi]$ usando coordenadas esféricas [1]. Essa transformação resulta em imagens com uma distorção horizontal, proeminente próximo aos polos. A amostragem não-uniforme na esfera no mapeamento ERP é proporcional a $\cos(\phi)$, $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, [7] e resulta nessa deformação (confira a Fig. 1a).

Como os polos em imagens ERP são superamostrados [8], alguns autores propõe explorar as redundâncias da imagem de forma diferenciada, a depender da latitude ϕ [7], [10]–[12]. Lee *et al.* [10] empregam o descarte progressivo de *pixels* em direção aos polos e rearranjo dos demais em imagens quadradas. Essas, ainda mais distorcidas, são submetidas a um processo de codificação tradicional.

De Simone *et al.* [7] observam que quanto mais próximo dos polos, maiores são as frequências espaciais horizontais quando há projeção do panorama para a esfera. Tais autores propõe ajustar as colunas da matriz de quantização \mathbf{Q} de acordo com ϕ utilizando o conceito de deslocamento de frequências horizontais. Na prática, a quantização é aplicada de forma mais intensa aos polos.

Outras abordagens atingem resultados similares em vídeos ERP por ajuste do parâmetro de quantização (QP) [11], [12]. A ideia fundamental desses trabalhos é alterar o QP (que tem relação inversa ao QF) para controlar a intensidade da quantização. Não estamos cientes de abordagens similares para o contexto de imagens estáticas.

O método em [10] descarta informações no domínio espacial, o que reduz o custo da compressão e do armazenamento. As outras abordagens [7], [11], [12] reduzem o custo de armazenamento apenas. Todas as abordagens têm perda de qualidade de imagem se comparadas a abordagens que não se adaptam à latitude (como o padrão JPEG). O método proposto neste artigo, descrito na Seção III, promove a redução de custo de computação e de armazenamento. Como as demais abordagens concorrentes, nosso método gera perda de qualidade de imagem em comparação com o mecanismo do padrão JPEG.

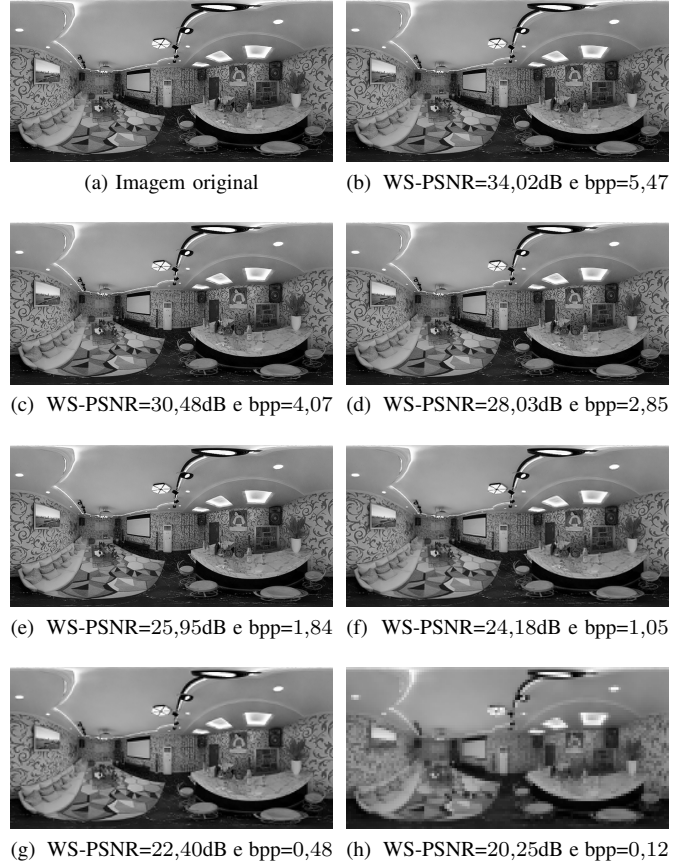


Fig. 1. Qualidade de imagem e de compressão por número de podas. Imagem original em (a) e $K = 1, 2, \dots, 7$ podas em (b)–(h), respectivamente.

III. MÉTODO PROPOSTO

Este trabalho objetiva a aplicação do mecanismo de poda de transformada [15] como forma de reduzir o custo da compressão e tamanho de imagens em 360°. Nossa abordagem explora as diferentes taxas de amostragens (e redundâncias) existentes em imagens ERP.

A poda objetiva extrair uma submatriz de uma transformada enquanto preserva parte de suas propriedades matemáticas [15]. No caso da DCT, a poda de K bases é feita ao descartar as últimas K linhas da matriz de sua transformação, uma vez a maior parte da energia do sinal é agrupada nos primeiros coeficientes espectrais [13].

Portanto, a poda de K bases da DCT resulta em uma matriz de $P = N - K$ linhas e N colunas dada por [15]

$$\mathbf{C}_P = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,N} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{P,1} & c_{P,2} & \cdots & c_{P,N} \end{bmatrix}, \quad (9)$$

onde $c_{m,n}$ são os coeficientes de \mathbf{C} dados pela Eq. (3). A notação adotada assume o subscrito de \mathbf{C} (ex., $\mathbf{C} = \mathbf{C}_N$) como sendo o número de linhas da matriz de transformação.

A aplicação da transformada 2D direta com poda K considera a troca de \mathbf{C} da Eq. (2) por \mathbf{C}_P . Note que os blocos transformados têm tamanho $P \times P$. Assim, a matriz \mathbf{Q} também deve ser podada para acomodar a quantização. A matriz de quantização podada é dada por

$$\mathbf{Q}_{P \times P} = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,P} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ q_{P,1} & q_{P,2} & \cdots & q_{P,P} \end{bmatrix}, \quad (10)$$

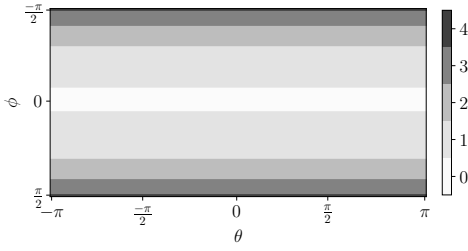


Fig. 2. O número de podas K da transformada dependendo da latitude ϕ .

onde $q_{m,n}$ são os coeficientes de \mathbf{Q} dados pela Eq. (5). Nesse caso, as Eqs. (4) e (7) consideram $\mathbf{Q}_{P \times P}$ ao invés de \mathbf{Q} . A transformada inversa 2D (Eq. (8)) também deve ser ajustada com \mathbf{C}_P , uma vez que esta segue a definição de semi-ortogonalidade [16, p. 104]. Note que o bloco descomprimido tem tamanho $N \times N$, assim como o original.

A Fig. 1 ilustra a aplicação da abordagem de compressão similar à do padrão JPEG considerando $K = 0, 1, \dots, N - 1$ podas e $QF = 100$. Nesse caso, a perda de qualidade é resultante da poda apenas. São exibidos também valores de qualidade de imagem (através da métrica *weighted-to-spherically-uniform peak signal-to-noise ratio* ou WS-PSNR [17]) e tamanho de imagem (pelo cômputo de *bits per pixel* ou bpp [13]). Idealmente, quer-se reduzir o tamanho de imagem (bpp) e aumentar sua qualidade (WS-PSNR). Note, na Fig. 1, que uma maior poda leva a menores tamanhos de imagem mas também resulta em alta degradação. Note também que a degradação não é tão acentuada nos polos se comparada ao equador.

A proposta desse trabalho é considerar um ajuste de poda K por latitude ϕ (altura de bloco). Para tanto, foi proposto um “perfil de podas” construído a partir de 1.000 imagens ERP, não-comprimidas, em escala de cinza e de tamanho 512×1024 pixels obtidas da base de dados pública *Structured3D* [18]. Geralmente, imagens ERP são alinhadas ao equador [1] e o polo sul tem mais informação contextual (textura) que o polo norte [7]. Para criarmos um perfil de podas que desconsidera contexto, rotacionamos aleatoriamente as imagens na esfera e as reprojecemos para o plano via mapeamento ERP [1].

O Algoritmo 1 detalha o procedimento adotado para construção do perfil de podas (\mathcal{K}). O algoritmo recebe um conjunto de imagens ERP rotacionadas \mathcal{I} , onde cada imagem \mathbf{I} tem tamanho $A \times L$ pixels, e um parâmetro λ ($\lambda = 4$ atribuído experimentalmente) que quantifica uma variação na qualidade de imagem resultante da poda em comparação com o método JPEG tradicional. Cada imagem \mathbf{I} é quebrada em blocos (linha 5) e comprimida via JPEG (linha 6). No Algoritmo 1, o marcador “.” indica indexação de todos os elementos naquela dimensão (linha ou coluna). Então, para cada altura de bloco l , aplica-se a compressão sem quantização considerando a DCT com poda K (linha 10). É selecionada a poda K , na altura de bloco l , que tem qualidade superior ao JPEG mais justa à λ . Esse processo é repetido para cada imagem \mathbf{I} em \mathcal{I} . O perfil de podas médio entre todas as imagens é retornado pelo Algoritmo 1.

A Fig. 2 ilustra o perfil de podas obtido. Note que ele é simétrico com relação ao equador e aplica de zero a quatro podas, as quais são intensificadas nos polos.

IV. RESULTADOS EXPERIMENTAIS

Essa seção discute os experimentos e resultados em termos de qualidade e tamanho de imagem e custo computacional (aritmético) em comparação com a metodologia empregada no padrão JPEG.

A. Qualidade de Imagem e Taxa de Compressão

Aqui, foram consideradas as mesmas 1.000 imagens ERP descritas na Seção III, porém com orientação original, isto é sem rotação e alinhadas ao equador. As métricas WS-PSNR [17] e bpp [13] foram utilizadas para quantificar a qualidade e tamanho de imagem.

Algoritmo 1: CONSTRUÇÃO DO PERFIL DE PODAS

Entrada: $\mathcal{I}, A, L, \lambda$
Saída: \mathcal{K}

```

1 início
2    $N \leftarrow 8$ 
3    $\mathcal{L} \leftarrow (1, 2, \dots, A/N)$ 
4   para  $\mathbf{I} \in \mathcal{I}$  faça
5      $\mathbf{I}_{:,l} \leftarrow$  Aplica Eq. (1)
6      $\tilde{\mathbf{I}}_{:,l}^J \leftarrow$  Aplica Eqs. (2), (4), (7) e (8)
7      $\mathcal{K}_l^i \leftarrow (k_l = 0 | l \in \mathcal{L})$ 
8     para  $l \in \mathcal{L}$  faça
9        $K^* \leftarrow 0$ 
10      para  $K \in (0, 1, \dots, N - 1)$  faça
11         $\tilde{\mathbf{I}}_{:,l}^P \leftarrow$  Aplica Eqs. (2), (4), (7) e (8)
12        (com Eqs. (9) e (10) e  $QF = 100$ )
13        se WS-PSNR( $\tilde{\mathbf{I}}_{:,l}^P$ ) – WS-PSNR( $\tilde{\mathbf{I}}_{:,l}^J$ ) >  $\lambda$ 
14          então
15             $K^* = K$ 
16          fim
17        fim
18      fim
19    fim
20 retorna  $[\mu(\mathcal{K}_l^i)]$ 

```

Os experimentos adotam o mecanismo de transformação do padrão JPEG, conforme discutido na Seção II-A, adaptando a transformação e quantização nos casos de poda conforme as Eqs. (9) e (10). Os experimentos associados à abordagem proposta usam o perfil de podas da Fig. 2. Todos os resultados reportados são médios e consideram a totalidade de imagens.

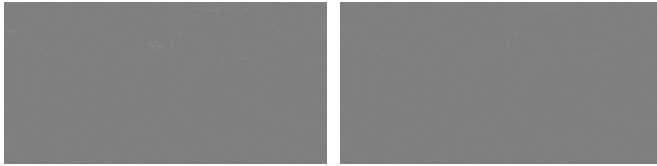
O Experimento 1 desconsidera completamente a quantização. Cada bloco transformado é simplesmente truncado como inteiro e então submetido à transformação inversa. Esse processo é irrealista na prática pois a compressão é mínima e incontrolável. Por outro lado, ele permite avaliar o impacto das podas. Se $K = 0$ para todo bloco (como no padrão JPEG), então obtém-se 58,94dB médios à 6,26bpp. A abordagem proposta, que adapta K à latitude ϕ , atinge, em média, 36,95dB e 4,39bpp quando não há quantização.

A quantização tem um papel vital para a compressão [9]. O Experimento 2 aplica quantização considerando $QF = 50$. Os resultados médios da abordagem proposta são de 33,14dB à 0,75bpp. Em comparação, o mecanismo empregado no padrão JPEG atinge, em média, 34,55dB à 0,78bpp. A abordagem proposta permite um ganho de 3,85% em termos de bpp ao custo de uma perda de qualidade de imagem de 4,10%. É importante ressaltar que a poda reduz o custo de computação, conforme será discutido na Seção IV-B.

A Fig. 3 ilustra o resíduo da compressão de uma imagem ERP (Fig. 1a) dada a aplicação da abordagem proposta e do padrão JPEG. Há, virtualmente, os mesmos artefatos nas duas imagens. Os valores de WS-PSNR e bpp para as imagens compactadas são fornecidos.

B. Redução de Custo Aritmético

A abordagem proposta para compressão de imagens ERP considera a aplicação de K podas sobre a matriz da DCT de ordem $N = 8$. Tal processo resulta em matrizes de transformação não-quadradas, com $P = N - K$ linhas e N colunas. A aplicação da DCT podada reduz o custo aritmético de transformação se comparada ao da DCT exata.



(a) WS-PSNR=29,72dB e bpp=1,28 (b) WS-PSNR=30,33dB e bpp=1,33

Fig. 3. Imagens diferença considerando a imagem original e (a) a abordagem proposta e (b) abordagem empregada no padrão JPEG.

TABELA I
CUSTO ARITMÉTICO DA TRANSFORMADA 2D COM PODA K

K	$\mathcal{M}(\mathcal{C}_P)$	$\mathcal{A}(\mathcal{C}_P)$
0	1024	896
1	840	735
2	672	588
3	520	455
4	384	336

A quantidade de operações aritméticas envolvidas na aplicação da transformada depende de arquitetura e representa uma métrica justa para o custo computacional [14], [15]. O número de multiplicações e de adições envolvidas na transformação 2D direta (ou inversa) por bloco utilizando a DCT com K podas é dado, respectivamente, por

$$\mathcal{M}(\mathcal{C}_P) = PN(N + P) \quad (11)$$

e

$$\mathcal{A}(\mathcal{C}_P) = P(N - 1)(N + P). \quad (12)$$

Note que, se $K = 0$, então $P = N$, e as Eqs. (11) e (12) computam o custo de aplicar a Eq. (2) (ou Eq (8)). A Tabela I mostra o custo aritmético de aplicar a transformada 2D direta (ou inversa) com $K = 0$ (DCT exata, sem poda) e com $K = 1, 2, 3, 4$, que são adotadas em nossa proposta, conforme discutido na Seção III.

O perfil de podas apresentado neste trabalho (ver Fig. 2) combina diferentes valores de K a depender da altura do bloco. Dessa forma, o ganho no custo aritmético aditivo e multiplicativo com relação à aplicação da transformada são de aproximadamente 25,58% e 25,59%, respectivamente. Ainda, nossa abordagem requer quantização de tamanho $P \times P$ ao invés de $N \times N$, resultando em uma real redução de 29,96% no número total de multiplicações.

Destaca-se que, ao variar λ , também varia-se o custo aritmético, bem como qualidade de imagem e taxa de compressão. Também é importante salientar que *todos* os custos aqui apresentados podem ser reduzidos com a adoção de algoritmos rápidos [14], [15], [19], os quais serão investigados no futuro.

V. CONSIDERAÇÕES FINAIS

O presente trabalho propôs o uso de transformadas podadas para compressão de imagens esféricas em formato ERP. Nossa abordagem considerou 1.000 imagens rotacionadas para criação de um perfil de poda dependente de latitude, visando a redução de computação. Nossos resultados experimentais médios, realizados com 1.000 imagens em orientação original, indicaram que a abordagem proposta leva a uma redução média de 3,85% em bpp, uma perda média de 1,41dB em qualidade de imagem e uma redução significativa de 25,58% e 29,96% do número adições e multiplicações, respectivamente, envolvidos na compressão.

No futuro, planejamos construir uma função que mapeie a altura de bloco ao número de podas independentemente de resolução e a investigação do uso de outros valores de QF . Também iremos

comparar resultados com abordagens concorrentes como [7], [10]–[12]. Por fim, experimentos futuros irão explorar abordagens baseadas em algoritmos rápidos e/ou aproximações para a DCT [14], [15], bem como a adoção de transformadas de maior tamanho [19].

AGRADECIMENTOS

Os autores agradecem o financiamento parcial desta pesquisa provido pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Brasil.

REFERENCES

- [1] T. L. T. da Silveira, P. G. L. Pinto, J. Murrugarra-Llerena, and C. R. Jung, “3D scene geometry estimation from 360° imagery: A survey,” *ACM Comput. Surv.*, 2022, Just Accepted.
- [2] M. Karkoub, O. Bouhali, and A. Sheharyar, “Gas pipeline inspection using autonomous robots with omni-directional cameras,” *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15 544–15 553, 2021.
- [3] S. G. Izard, J. A. Juanes, F. J. García Peñalvo, J. M. G. Estella, M. J. S. Ledesma, and P. Ruisoto, “Virtual reality as an educational and training tool for medicine,” *J. Med. Syst.*, vol. 42, no. 3, p. 50, 2018.
- [4] R. G. d. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli, and P. Frossard, “Visual distortions in 360° videos,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2524–2537, 2020.
- [5] S. Schwarz and M. M. Hannuksela, “Perceptual quality assessment of HEVC main profile depth map compression for six degrees of freedom virtual reality video,” in *IEEE International Conference on Image Processing (ICIP)*, no. 17 2, 2017, pp. 181–185.
- [6] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J. P. Thiran, “Scale invariant feature transform on the sphere: Theory and applications,” *International Journal of Computer Vision*, vol. 98, no. 2, pp. 217–241, 2012.
- [7] F. De Simone, P. Frossard, P. Wilkins, N. Birkbeck, and A. Kokaram, “Geometry-driven quantization for omnidirectional image coding,” in *Picture Coding Symposium (PCS)*, 2016, pp. 1–5.
- [8] Y.-C. Su and K. Grauman, “Learning spherical convolution for fast features from 360° imagery,” in *International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 529–539.
- [9] G. K. Wallace, “The JPEG still picture compression standard,” *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
- [10] S.-H. Lee, S.-T. Kim, E. Yip, B.-D. Choi, J. Song, and S.-J. Ko, “Omnidirectional video coding using latitude adaptive down-sampling and pixel rearrangement,” *Electronics Letters*, vol. 53, no. 10, pp. 655–657, 2017.
- [11] S. Jaballah, A. Bhavsar, and M.-C. Larabi, “Perceptual Versus Latitude-Based 360-Deg Video Coding Optimization,” in *IEEE International Conference on Image Processing (ICIP)*, oct 2020, pp. 3423–3427.
- [12] X. Xiu, Y. He, and Y. Ye, “An adaptive quantization method for 360-degree video coding,” in *Applications of Digital Image Processing XLI*, vol. 10752. SPIE, 2018, pp. 317 – 325.
- [13] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic Press, 1990.
- [14] C. Tablada, T. L. T. da Silveira, R. Cintra, and F. Bayer, “DCT approximations based on Chen’s factorization,” *Signal Processing: Image Communication*, vol. 58, pp. 14–23, 2017.
- [15] R. J. Cintra, F. M. Bayer, V. A. Coutinho, S. Kulasekera, A. Madanayake, and A. Leite, “Energy-efficient 8-point DCT approximations: Theory and hardware architectures,” *Circuits Systems Signal Process.*, vol. 35, no. 11, pp. 4009–4029, 2016.
- [16] X.-D. Zhang, *Matrix Analysis and Applications*. Cambridge University Press, 2017.
- [17] Y. Sun, A. Lu, and L. Yu, “Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video,” *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1–1, 2017.
- [18] J. Zheng, J. Zhang, J. Li, R. Tang, S. Gao, and Z. Zhou, “Structured3D: A large photo-realistic dataset for structured 3D modeling,” in *European Conference on Computer Vision*, 2020, pp. 519–535.
- [19] T. L. T. da Silveira, R. S. Oliveira, F. M. Bayer, R. J. Cintra, and A. Madanayake, “Multiplierless 16-point DCT approximation for low-complexity image and video coding,” *Signal, Image and Video Processing*, vol. 11, no. 2, pp. 227–233, 2017.