MULTI-RESOLUTION INTRA-PREDICTIVE CODING OF 3D POINT CLOUD ATTRIBUTES

Eduardo Pavez[⋆]*, André L. Souto*[#]*, Ricardo L. De Queiroz*[#]*, and Antonio Ortega*[⋆]

[⋆]University of Southern California, Los Angeles, California, USA
[#] Universidade de Brasilia, Brasilia, Brazil

ABSTRACT

We propose an intra frame predictive strategy for compression of 3D point cloud attributes. Our approach is integrated with the region adaptive graph Fourier transform (RA-GFT), a multi-resolution transform formed by a composition of localized block transforms, which produces a set of low pass (approximation) and high pass (detail) coefficients at multiple resolutions. Since the transform operations are spatially localized, coefficients at a given resolution may still be correlated. To exploit this phenomenon, we propose an intra-prediction strategy, in which decoded approximation coefficients are used to predict uncoded detail coefficients. The prediction residuals are then quantized and entropy coded. For the 8i dataset, we obtain gains up to 0.5db as compared to intra predicted point cloud compresion based on the region adaptive Haar transform (RAHT).

***Index Terms*—** 3D point clouds, intra prediction, multiresolution transform, graph fourier transform

1. INTRODUCTION

3D point clouds (3DPC) have become the preferred representation of 3D scenes, people and objects [1, 2, 3]. They consist of a list of points coordinates in 3D space along with color attributes. Recent advancements in real time 3D capture, along with potential applications to entertainment and immersive communications, have prompted research on 3DPC compression [1, 2].

This paper focuses on 3DPC attribute compression. Earlier approaches for compression of 3DPC attributes were based on transform coding techniques, that is, transformation followed by quantization and entropy coding, similar to modern image codecs. For 3DPCs, a popular approach is based on the region adaptive hierarchical (or Haar) transform (RAHT) [4, 5]. In image and video coding, intra and inter prediction are often combined with transform coding. Predictive methods for 3DPC compression have only recently become popular, which may be explained by the fact that good spatial and temporal predictors are harder to obtain for 3DPCs, since these i) represent complex surface and non-surface geometries, ii) have spatial frame to frame irregularity, and iii) lack temporal consistency. There has been substantial recent work to apply transform coding to inter frame (temporal) prediction residuals [6, 7, 8, 3, 9]. However, intra frame prediction is less explored. In [10, 11], block based intra-prediction similar to video was considered. More recently, MPEG adopted an intra prediction strategy for the RAHT (I-RAHT) [12], which uses a multi-resolution prediction, instead of traditional directional block based prediction.

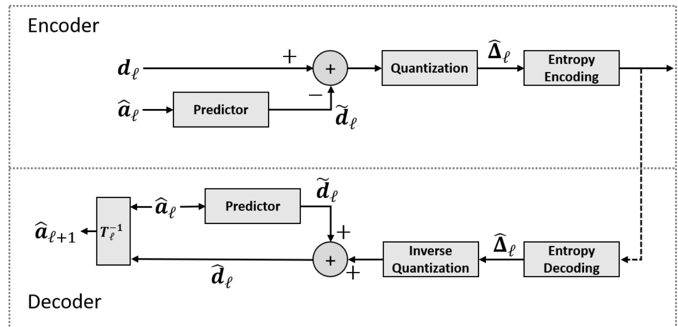
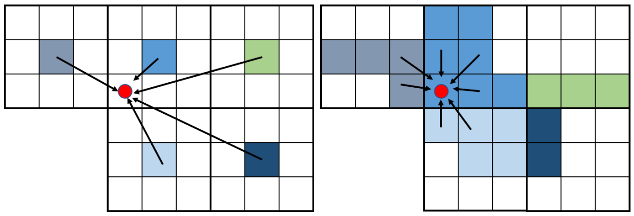


Fig. 1: Multi-resolution intra-predictive coding

In I-RAHT, low pass decoded RAHT coefficients are used to predict high pass RAHT coefficients. To achieve this, a higher resolution predictor point cloud is constructed from lower resolution decoded coefficients. Each predictor attribute in the high resolution point cloud is obtained as a linear combination of the nearest attributes in the lower resolution point cloud. Then the detail coefficients of this predictor signal are used to predict the target detail coefficients. While this approach to construct a higher resolution point cloud is suitable for the RAHT, since it is formed by a composition of 2×2 orthogonal transforms, it fails to take into account the more complex geometry of the higher resolution point cloud. This becomes even more important when considering larger block transforms with more points used by the region adaptive graph Fourier transform (RA-GFT) [13], as shown in Figure 2.

In this paper we propose an intra-predictive coding framework for RA-GFT (I-RAGFT). We use the same multi-resolution prediction strategy used in I-RAHT, where de-

This work was funded in part by NSE under grant CNS-1956190. Author's email: pavezcar@usc.edu, andre@image.unb.br, queiroz@ieee.org, ortega@sipi.usc.edu.



(a) I-RAHT (b) Proposed

Fig. 2: Comparison of multi-resolution predictors. (2a) Predictor attribute (red dot) is a linear combination of neighbors in lower resolution point cloud ($\mathbf{V}_\ell$). (2b) Predictor attribute is a linear combination of neighbors in current resolution point cloud ($\mathbf{V}_{\ell+1}$).

coded approximation coefficients are used to predict and code detail coefficients, as depicted in Figure 1. However, different from the prediction algorithm used by I-RAHT, our proposed predictors exploits the higher resolution point cloud geometry. We start by projecting the low resolution signal onto the higher resolution geometry by zero padding and applying a one level inverse RA-GFT. The resulting interpolated signal is piece-wise constant, as depicted by the colored voxels in Figure 2b. Then a smoothing graph filter [14, 15, 16] is applied using a graph constructed on top of the higher resolution point cloud. While our approach is new for 3DPCs, similar ideas have been used to improve image coding by predicting wavelet coefficients with learning based super resolution algorithms [17].

The difference between our proposed approach and I-RAHT is illustrated in Figure 2. I-RAHT uses a *block-level* predictor, since the approximation coefficients of the nearest neighboring blocks are used (Figure 2a), while our approach uses a *point-level* predictor, where fine resolution point values are interpolated, then filtered (across block boundaries), so that only nearby points (instead of nearby blocks) are used to compute the predictor. We show through compression experiments that the proposed approach can outperform I-RAHT, when using uniform quantization and adaptive RLGR entropy coding [18].

The rest of the paper is organized as follows. In Section 2 we introduce the multi-resolution intra-prediction coding framework. The proposed predictors for the RA-GFT are described in 3. We show numerical experiments and conclusions in Sections 4 and 5, respectively.

2. MULTI RESOLUTION PREDICTIVE CODING

Consider a point cloud with point coordinates stored in the $N \times 3$ matrix $\mathbf{V} = [\mathbf{v}_i]$, and attributes $\mathbf{a} = [a_i]$. We will assume the attributes are processed with a multi-resolution transform, which at each resolution level takes an attribute vector $\mathbf{a}_{\ell+1}$, and produces approximation and detail coefficients at resolution ℓ

$$\begin{bmatrix}\mathbf{a}_\ell\\ \mathbf{d}_\ell\end{bmatrix} = \mathbf{T}_\ell \mathbf{a}_{\ell+1}. \tag{1}$$

The transform matrix $\mathbf{T}_\ell$ is an orthonormal matrix, thus $\mathbf{T}_\ell^{-1} = \mathbf{T}_\ell^\top$. We assume that the original 3DPC attributes are stored at the highest resolution, L, so that $\mathbf{a}_L = \mathbf{a}$. After applying (1) L times, we obtain transform coefficients

$$\left[\mathbf{a}_0^\top, \mathbf{d}_0^\top, \mathbf{d}_1^\top, \cdots, \mathbf{d}_{L-1}^\top\right]^\top. \tag{2}$$

Several orthonormal transforms for 3DPC attributes can be described this way, including the block based graph Fourier transform [19], RAHT [4] and RA-GFT [13]. Since RA-GFT is a composition of spatially localized block transforms, there may be additional redundancy between transformed coefficients, similar to what is observed for the RAHT [12]. While previous approaches code the coefficients in (2) independently ignoring their dependencies across blocks, in this work we exploit them to improve coding efficiency.

Denote by $\widehat{\mathbf{\Phi}}_0 = \mathcal{Q}(\mathbf{a}_o)$, the quantized low pass coefficients, where $\mathcal{Q}(\cdot)$ is a quantization operator, and denote by $\hat{\mathbf{a}}_0 = \mathcal{Q}^{-1}(\widehat{\mathbf{\Phi}}_0)$ the corresponding decoded coefficients. Now we define several quantities recursively. The decoded approximation coefficient at resolution ℓ is given by $\hat{\mathbf{a}}_\ell$, while the corresponding detail coefficient at the same resolution is $\mathbf{d}_\ell$. We will assume there is function $\mathcal{P}_\ell(\hat{\mathbf{a}}_\ell) = \tilde{\mathbf{d}}_\ell$, that predicts detail coefficients from approximation coefficients. This predictor will be explained in detail in the next section. Using $\mathcal{P}$ we compute a residual and quantize it obtaining

$$\widehat{\mathbf{\Delta}}_\ell = \mathcal{Q}(\mathbf{d}_\ell - \tilde{\mathbf{d}}_\ell). \tag{3}$$

If the predictor is good enough, coding $\widehat{\mathbf{\Delta}}_\ell$ is more efficient than coding the transform coefficients $\mathcal{Q}(\mathbf{d}_\ell)$ directly. The decoded details coefficients are given by

$$\hat{\mathbf{d}}_\ell = \tilde{\mathbf{d}}_\ell + \mathcal{Q}^{-1}(\widehat{\mathbf{\Delta}}_\ell). \tag{4}$$

The decoded approximation and detail coefficients at resolution ℓ are used to obtain approximation coefficients at resolution $\ell+1$ with

$$\hat{\mathbf{a}}_{\ell+1} = \mathbf{T}_\ell^{-1}\begin{bmatrix}\hat{\mathbf{a}}_\ell\\ \hat{\mathbf{d}}_\ell\end{bmatrix}. \tag{5}$$

The quantities $\widehat{\mathbf{\Delta}}_\ell$ can be computed recursively as depicted in Figure 1, starting from the lowest resolution coefficients $\mathbf{a}_0$. A typical transform coding strategy would quantize and entropy encode (2). In this work we encode

$$\left[\widehat{\mathbf{\Phi}}_0^\top, \widehat{\mathbf{\Delta}}_0^\top, \widehat{\mathbf{\Delta}}_1^\top, \cdots, \widehat{\mathbf{\Delta}}_{L-1}^\top\right]^\top. \tag{6}$$

3. MULTI RESOLUTION PREDICTION

In this section we describe the function $\mathcal{P}(\hat{\mathbf{a}}_\ell) = \tilde{\mathbf{d}}_\ell$, used to predict the high pass coefficients $\mathbf{d}_\ell$.

3.1. Graph representation of point clouds

While forming transform coefficients (2), many transforms [4, 13], either explicitely or implicitely, produce sets of point coordinates at various resolutions (e.g., by down-sampling), thus for each resolution ℓ, we have a point cloud $(\mathbf{V}_\ell, \mathbf{a}_\ell)$, where $\mathbf{V}_L = \mathbf{V}$ and $\mathbf{a}_L = \mathbf{a}$.

For each of these point clouds consider a graph $\mathcal{G}_\ell = (\mathbf{V}_\ell, \mathbf{W}_\ell, \mathcal{E}_\ell)$. The matrix $\mathbf{W}_\ell$ is the adjacency matrix and $\mathbf{D}_\ell = \mathrm{diag}(\sum(\mathbf{W}_\ell))$ is the degree matrix. The graph has edge set $\mathcal{E}_\ell$, where $ij \in \mathcal{E}_\ell$ if point $\mathbf{v}_{i,\ell}$ is "near" to point $\mathbf{v}_{j,\ell}$. Edge weights are given by

$$(\mathbf{W}_\ell)_{ij} = \frac{1}{\|\mathbf{v}_{i,\ell} - \mathbf{v}_{j,\ell}\|}. \tag{7}$$

3.2. Constructing the predictor

The first step in constructing our predictor is interpolation of low resolution point cloud $(\mathbf{V}_\ell, \hat{\mathbf{a}}_\ell)$ by zero padding, and inverse transformation, leading to the attribute signal

$$\mathbf{b}_{\ell+1} = \mathbf{T}_\ell^{-1} \begin{bmatrix} \hat{\mathbf{a}}_\ell \\ \mathbf{0} \end{bmatrix}, \tag{8}$$

and higher resolution point cloud $(\mathbf{V}_{\ell+1}, \mathbf{b}_{\ell+1})$. We propose applying a smoothing (low pass) filter to the point cloud $(\mathbf{V}_{\ell+1}, \mathbf{b}_{\ell+1})$, and then applying the transform $\mathbf{T}_\ell$. The *predictor* block from Figure 1 is depicted in Figure 3, with the caveat that the output $\tilde{\mathbf{a}}_\ell$ is ignored.

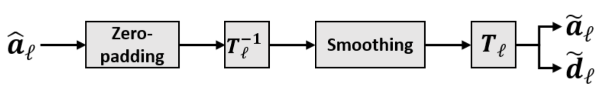


Fig. 3: Multi-resolution predictor.

3.3. Graph filtering of RA-GFT coefficients

In the RA-GFT [13], each transform coefficient has an importance weight that depends on the size (number of points) of the region of the point cloud they represent. The ith point of the point cloud of resolution ℓ is denoted by $\mathbf{v}_{i,\ell}$. The point $\mathbf{v}_{i,\ell}$ has a set of children at resolution $\ell+1$ denoted by $\mathcal{C}_{i,\ell}$. The index set of the children is denoted by $\mathcal{I}_{i,\ell} = \{j : \mathbf{v}_{j,\ell+1} \in \mathcal{C}_{i,\ell}\}$. We denote the weight of point $\mathbf{v}_{i,\ell}$ by $q_{i,\ell}$. Based on this parent-child relation, the weights can be computed recursively as

$$q_{i,\ell} = \sum_{j \in \mathcal{I}_{i,\ell}} q_{j,\ell+1}, \tag{9}$$

and the weights at full resolution are $q_{i,L} = 1$ for all i. For the RA-GFT, the interpolation equation (8) has a closed form, thus the jth entry of $\mathbf{b}_{\ell+1}$ is equal to

$$b_{j,\ell+1} = \sqrt{\frac{q_{j,\ell+1}}{q_{i,\ell}}}\hat{a}_{i,\ell}, \tag{10}$$

where point $\mathbf{v}_{i,\ell}$ is the parent of $\mathbf{v}_{j,\ell+1}$. Note that if we define the diagonal matrix of weights by $\mathbf{Q}_{\ell+1}$, with jj entry equal to $q_{j,\ell+1}$, then the interpolated signal is

$$\mathbf{b}_{\ell+1} = \mathbf{Q}_{\ell+1}^{1/2} \left(\sum_i \frac{\hat{a}_{i,\ell}}{\sqrt{q_{i,\ell}}} \mathbf{1}_{\mathcal{I}_{i,\ell}} \right), \tag{11}$$

where $\mathbf{1}_{\mathcal{I}_{i,\ell}}$ is the indicator of the set $\mathcal{I}_{i,\ell}$. Thus after normalization by the square root of the point weights, the interpolated signal is piece-wise constant. In fact, this signal is constant within cube shaped regions, because the RA-GFT is is a composition of localized "block" transforms. Figure 2b depicts this piece-wise constant signal. Given that $\mathbf{b}_{\ell+1}$ has this form, we take into account the normalization matrix $\mathbf{Q}_{\ell+1}$, and the piece-wise constant structure, when designing our smoothing operator. The first step is to normalize the entries of $\mathbf{b}_{\ell+1}$ by their point weights using $\mathbf{Q}_{\ell+1}^{-1/2}$. The resulting piece-wise constant signal is filtered to smooth out boundaries between regions. Finally, we scale back each attribute using the matrix $\mathbf{Q}_{\ell+1}^{1/2}$. The proposed filtering operation has the form

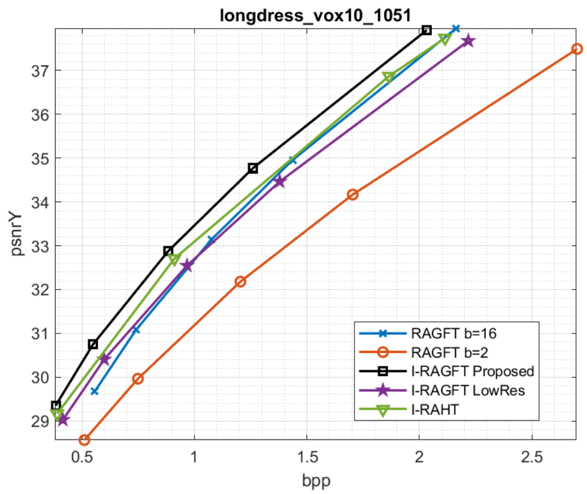
$$\tilde{\mathbf{b}}_{\ell+1} = \mathbf{Q}_{\ell+1}^{1/2} \mathbf{D}_{\ell+1}^{-1} \mathbf{W}_{\ell+1} \mathbf{Q}_{\ell+1}^{-1/2} \mathbf{b}_{\ell+1}. \tag{12}$$

At resolution ℓ, the adjacency matrix $\mathbf{W}_\ell$ is constructed using k nearest neighbors on the point cloud geometry $\mathbf{V}_\ell$. The resulting signal $\tilde{\mathbf{b}}_{\ell+1}$ is transformed using (1), resulting in $\tilde{\mathbf{a}}_\ell$ and $\tilde{\mathbf{d}}_\ell$.

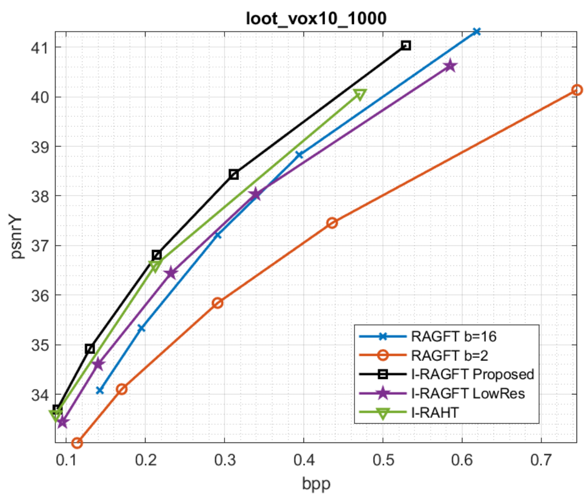
4. NUMERICAL RESULTS

We integrate intra prediction with the RA-GFT with $2 \times 2 \times 2$ blocks at all resolution levels, thus each localized block transform processes at most 8 points. The proposed predictors are implemented using graph filters, that use K nearest neighbor graphs, with $K = 7$ neighbors per point. We denote this approach by "I-RAGFT". We also implement predictors similar to those used by the G-PCC implementation of I-RAHT using KNN from the lower resolution graph (see Fig. 2a). For this approach we set $K = 5$, found after optimizing for best performance. This approach is denoted by "I-RAGFT LowRes". We also compare against the RA-GFT with high resolution block size equal to 16, which achieves highest coding performance compared to RAHT, and the RA-GFT with high resolution block size equal to 2, which provides a mild improvement over RAHT (see [13] for details). For RA-GFT and intra RA-GFT, we uniformly quantize coefficients and entropy code them using adaptive run-length Golomb-Rice (RLGR) algorithm [18].

The state of the art in non-video-based compression of point cloud attributes, called G-PCC [5] uses I-RAHT. Several techniques are implemented in G-PCC to improve coding efficiency beyond that obtained through intra prediction. Some of them include jointly encoding YUV coefficients that



(a) Longdress



(b) Loot

Fig. 4: Rate distortion curves for color compression.

are equal to zero, and adaptive quantization schemes of AC coefficients. These techniques could also be applied to the RA-GFT but their implementation goes beyond the scope of this paper and we leave them for future work. In order to obtain a performance comparison between our method and the current state-of-the-art, under the same conditions, we changed the source code of G-PCC modifying both the quantizer and the entropy encoder schemes. The adaptive quantizer was replaced with a uniform quantizer and the entropy encoder was modified to an adaptive RLGR algorithm. Moreover, we encode YUV coefficients independently. In Figure 4 we report rate distortion curves for the "longdress" and "loot" point clouds of the "8iVFBv2" dataset [20].

Incorporating intra prediction into the RA-GFT improves coding performance significantly. The difference between "RAGFT $b = 2$" and "I-RAGFT LowRes" is about 1.5db, while the gain obtained by using better predictors ("I-RAGFT Proposed") can reach up to 2.5db. In [13], the RA-GFT implemented with larger block sizes ($b_L = 16$) led to the best results, since spatial redundancy can be removed more efficiently using graph transforms on larger blocks. Our results show that by combining low resolution intra prediction and RA-GFT with small block size ("I-RAGFT $b = 2$"), we can outperform the best RA-GFT at low bitrates. However this approach is still inferior to I-RAHT. The proposed predictor based on the higher resolution point cloud, combined with the RA-GFT achieves the best rate distortion performance among all methods considered, and outperforms I-RAHT by up to 0.5db.

5. CONCLUSION

We studied the use of intra prediction for compression of 3D point cloud attributes with the RA-GFT. As with the RAHT, we showed that it is possible to remove redundancy between RA-GFT coefficients at different resolutions. In the RAHT, a high resolution point cloud is predicted from a lower resolution point cloud. While this approach is efficient for predicting RAHT coefficients, it is less effective for the RA-GFT. This is because the RA-GFT uses larger block transforms, that makes prediction more challenging. To overcome this issue, we proposed a different prediction based on interpolation and graph signal filtering, that takes into account the higher resolution geometry, thus adapting better to the RA-GFT. Compression experiments show the proposed approach outperforms intra predictive coding of RAHT coefficients at all rates by up to 0.5db.

6. REFERENCES

[1] Sebastian Schwarz, Marius Preda, Vittorio Baroncini, Madhukar Budagavi, Pablo Cesar, Philip A Chou, Robert A Cohen, Maja Krivokuća, Sébastien Lasserre, Zhu Li, et al., "Emerging MPEG standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2018.

[2] D Graziosi, O Nakagami, S Kuma, A Zaghetto, T Suzuki, and A Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, 2020.

[3] Eduardo Pavez, Philip A Chou, Ricardo L De Queiroz, and Antonio Ortega, "Dynamic polygon clouds: representation and compression for VR/AR," *APSIPA Transactions on Signal and Information Processing*, vol. 7, 2018.

[4] Ricardo L De Queiroz and Philip A Chou, "Compression of 3D point clouds using a region-adaptive hierarchical transform," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, 2016.

[5] 3DG, "G-PCC codec description," Tech. Rep. W19620, ISO/IEC JTC1/SC29/WG11 MPEG, 2020.

[6] Yiqun Xu, Wei Hu, Shanshe Wang, Xinfeng Zhang, Shiqi Wang, Siwei Ma, Zongming Guo, and Wen Gao, "Predictive generalized graph Fourier transform for attribute compression of dynamic point clouds," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[7] André L Souto and Ricardo L de Queiroz, "On predictive RAHT for dynamic point cloud coding," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 2701–2705.

[8] Ricardo L de Queiroz and Philip A Chou, "Motion-compensated compression of dynamic voxelized point clouds," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3886–3895, 2017.

[9] Dorina Thanou, Philip A Chou, and Pascal Frossard, "Graph-based compression of dynamic 3D point cloud sequences," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1765–1778, 2016.

[10] Robert A Cohen, Dong Tian, and Anthony Vetro, "Point cloud attribute compression using 3-d intra prediction and shape-adaptive transforms," in *2016 Data Compression Conference (DCC)*. IEEE, 2016, pp. 141–150.

[11] Yiting Shao, Qi Zhang, Ge Li, Zhu Li, and Li Li, "Hybrid point cloud attribute compression using slice-based layered structure and block-based intra prediction," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 1199–1207.

[12] Sébastien Lasserre and David Flynn, "On an improvement of RAHT to exploit attribute correlation," input document m47378, ISO/IEC JTC1/SC29/WG11 MPEG, 2019.

[13] Eduardo Pavez, Benjamin Girault, Antonio Ortega, and Philip A. Chou, "Region adaptive graph Fourier transform for 3D point clouds," in *International Conference on Image Processing (ICIP)*, 2020.

[14] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[15] Peyman Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE signal processing magazine*, vol. 30, no. 1, pp. 106–128, 2012.

[16] Gene Cheung, Enrico Magli, Yuichi Tanaka, and Michael K Ng, "Graph spectral image processing," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 907–930, 2018.

[17] Antoni Dimitriadis and David Taubman, "Augmenting JPEG2000 with wavelet coefficient prediction," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 1276–1280.

[18] Henrique S Malvar, "Adaptive run-length/Golomb-Rice encoding of quantized generalized gaussian sources with unknown statistics," in *Data Compression Conference (DCC'06)*. IEEE, 2006, pp. 23–32.

[19] Cha Zhang, Dinei Florencio, and Charles Loop, "Point cloud attribute compression with graph transform," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2066–2070.

[20] Eugene d'Eon, Bob Harrison, Taos Myers, and Philip A. Chou, "8i voxelized full bodies-a voxelized point cloud dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006*, 2017.