# ON PREDICTIVE RAHT FOR DYNAMIC POINT CLOUD CODING

*André L. Souto and Ricardo L. de Queiroz*

Universidade de Brasília, Brasília, Brazil
Email: andre@image.unb.br, queiroz@ieee.org

## ABSTRACT

We studied predictive coding applied to the region-adaptive hierarchical transform (RAHT) which is used for point cloud compression (PCC). RAHT is part of MPEG's geometry-based PCC test model and an intra-frame prediction scheme for RAHT (URAHT), wherein the prediction residual is encoded rather than the voxel attributes themselves, has been shown to deliver large gains. We extend the scheme to inter-frame prediction and show that a combination of simple zero-motion-vector (ZMV) inter-frame and intra-frame predictions can provide sizeable gains over pure RAHT or over intra-frame-only prediction when compressing dynamic point clouds. An adaptive method is used such that sections where ZMV does not yield good prediction switch to intra-frame prediction, assuring the performance to be at least that of the intra-frame case. Be the gains large (in steady parts) or very small (where there is rapid motion) results show consistent positive gains coming from a simple inter- and intra-frame prediction combination.

***Index Terms***— inter-frame prediction, RAHT transform, attribute coding, zero-motion-vector, intra-frame prediction

## 1. INTRODUCTION

Point clouds (PC) have enjoyed a growth in popularity in the past few years due to the increase in 3D applications, such as telepresence, virtual reality and autonomous driving [1]. PCs are defined as a set of points with proper geometry and a list of attributes. Geometry consists of the 3D position coordinates (x, y, z) of each point in the set. The attributes are usually color components (RGB or YUV), but may also include reflectance, motion vectors and so forth, for each point in the set [2].

PCs can be grouped into three different sets: static, dynamic and dynamically acquired [3]. Static PCs consist of a single static frame. Dynamic PCs are represented by multiples frames as a temporal sequence, as shown in Fig. 1. Dynamically acquired PCs mostly pertain to autonomous driving applications and are typically obtained by LiDAR technology. This work is mainly focused on improving the compression performance of dynamic PCs.

**Fig. 1**. Example of projection of successive frames of dynamic PC "Redandblack" [4].

Similar to images and videos, PCs represent a large amount of data. Thus, for their availability in practical applications, compression is required. For this reason, the Moving Picture Experts Group (MPEG) is currently working towards standardization of point cloud compression (PCC) technologies [3].

The region-adaptive hierarchical transform (RAHT) [5] was initially adopted in MPEG's test model for geometry-based point cloud compression (G-PCC) [2]. RAHT was developed to compress color signals and other attributes, such as reflectance [5, 6]. It consists of a hierarchical orthogonal sub-band transform that resembles an adaptive variation of a Haar wavelet transform. The basic idea behind RAHT is to follow the octtree scan backwards, from voxels towards the entire PC space, in a bottom-up approach, at each step, recombining voxels into larger ones by transforming them along each direction until the root is reached. Each transformation generates low- and high-pass coefficients. The low-pass coefficients are transmitted to the octtree's upper level and the high-pass coefficients are quantized and encoded [5].

In the latest G-PCC test models, the original RAHT has been modified. The current version consists of a fixed-point implementation [7]. Also, an intra-frame prediction step was introduced resulting in performance improvements [8]. However, the core of the RAHT transform remains the same.

For dynamic PCC, motion estimation (ME) has a major role in developing an inter-frame predictive coder. ME al-

Work partially supported by CNPq under grant 301647/2018-6.

lows to exploit inter-frame redundancy in scenes with significant object motion. Efforts have been made to develop a robust algorithm for point cloud ME [9–13]. In [9–11], graph transforms and block-based partitions were used, while other works [12, 13], have focused on reducing the computational cost of MEs. However, despite those efforts, point cloud ME algorithms remain inadequate for real-time applications due to their high computational cost. This work studies predictive RAHT and proposes the use of a low computational cost zero-motion-vector (ZMV) approach as an alternative for performance improvement of intra-frame predictive RAHT for attribute coding of dynamic PCs.

## 2. PREDICTIVE RAHT

In order to illustrate the process and benefits of a prediction step in RAHT, we begin with a theoretical experiment using the first frame of the dynamic PC Longdress [4]. Let $\mathcal{F}_{cur}$ be the current PC attribute frame and $\mathcal{F}'_{cur}$ its estimation. In predictive RAHT the residual error $\mathcal{F}_{cur} - \mathcal{F}'_{cur}$ is encoded instead of $\mathcal{F}_{cur}$. In this experiment, we generate $\mathcal{F}'_{cur}$ by adding Gaussian noise with standard deviation $\sigma$ to $\mathcal{F}_{cur}$. The rate-distortion curves for each experiment (value of $\sigma$) are presented in Fig. 2. It can be seen that, indeed, there may be gains in encoding the residual (predictive RAHT), but only if the prediction is good enough. We can see from Fig. 2 that, in this example for $\sigma < 4$, there may be gains, which roughly translates to an approximation yielding around 34-35 dB PSNR. In the contrary case, with poor approximation, one would be better off encoding $\mathcal{F}_{cur}$ instead of the prediction residual. Note that PSNR$_Y$ is obtained as point-to-point PSNR on Y color channel.

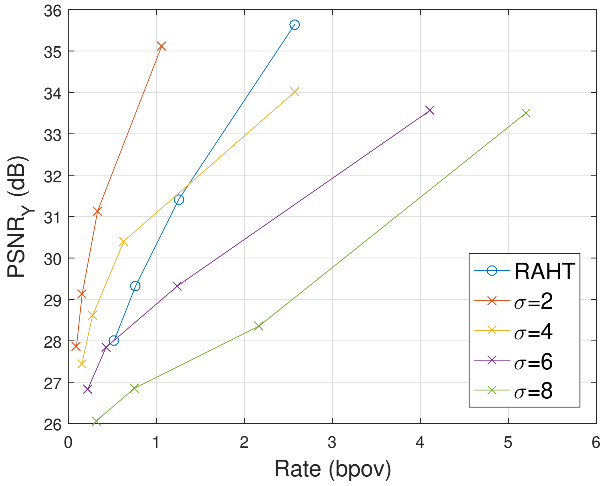


**Fig. 2**. Theoretical experiment to illustrate the potential gains of the predictive approach to the RAHT process.

### 2.1. MPEG Intra-predictive RAHT

Currently implemented in MPEG's TMC13 test model, URAHT [8] is an intra-frame predicted RAHT. Its prediction step aims to explore attribute correlation among neighboring voxels to generate a predictive model. The $\mathcal{F}'_{cur}$ is obtained through an intra-predictive method consisting of an inter-depth upsampling similar to a weighted average procedure [8]. The differences between $\mathcal{F}_{cur}$ and $\mathcal{F}'_{cur}$ are taken as residual errors and encoded as attributes within the conventional RAHT coder. Fig. 3 illustrates the potential coding gains of an intra-predictive RAHT in comparison to conventional RAHT for PC "Ricardo" [14]. Intra-predictive RAHT currently represents the state of the art in attribute compression of PCs.

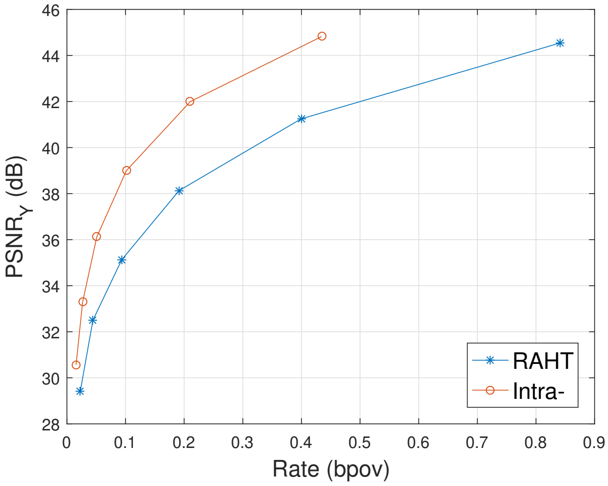


**Fig. 3**. Intra-predictive RAHT performance relative to RAHT for PC "Ricardo".

## 3. INTER-FRAME PREDICTION

Given two subsequent frames of a specific dynamic PC, wherein $\mathcal{F}_{cur}$ is to be compressed and its immediately preceding frame ($\mathcal{F}_{past}$) has already been encoded and locally decoded. An inter-frame prediction algorithm, that uses the lowest euclidean distance criteria among voxels, is applied to $\mathcal{F}_{past}$ resulting in estimated frame $\mathcal{F}'_{cur}$. The attribute differences of $\mathcal{F}_{cur}$ and $\mathcal{F}'_{cur}$ are computed producing a residual frame. Thus, one may eliminate redundancies that can be obtained at the decoder side by performing simple inter-frame prediction. The residual differences can be further subject to intra-frame prediction before RAHT, thereby merging both intra- and inter-frame prediction approaches.

The compression performance of the combination of inter- and intra-frame predictions against intra-frame-only prediction is illustrated with the tests in Fig. 4. We picked two example PC pairs of consecutive frames from popular

PC datasets [4, 14]. The particular pair of consecutive frames of PC sequence "Sarah" has very little motion, while the pair of consecutive frames of sequence "Loot" contains a more intense motion. Hence, we expect the inter-frame prediction to be fairly acurate for "Sarah" and poor for "Loot". That information, combined with the tests in Fig. 2, tell us that the inter-frame prediction approach should outperform the intra one for "Sarah", and to be outperformed for "Loot". That is precisely what the results illustrate in Fig. 4. The correspondence among PCs using simple proximity will lead to good approximations and to inter-frame prediction improvement if there is not much local motion in between frames. If there is, it may be better to use intra-frame-only prediction.

In order to improve inter-frame prediction reliability on regions with more intense object motion, a ME solution might be the first choice. However, a robust ME solution for PC frames remains elusive [9, 11–13]. Current ME approaches of PC are excessively expensive for real-time communications and not much reliable. Since a robust algorithm for inter-frame compensation is still an open issue, we opted here to use simple ZMV approaches like the one in the above example.

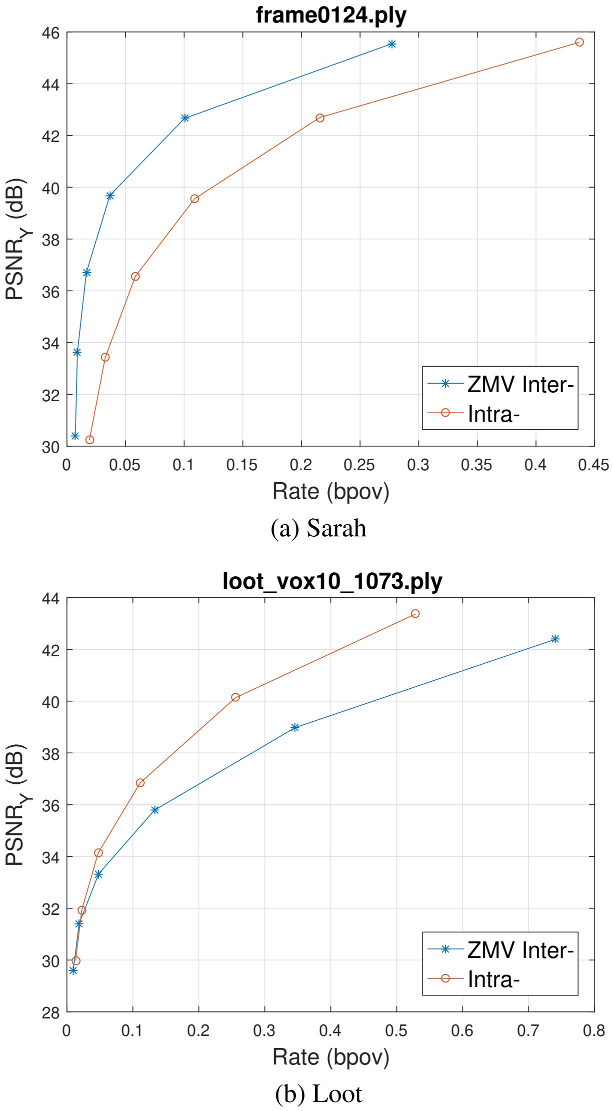
## 4. ZERO-MOTION-VECTOR APPROACHES

In order to overcome the weakness of the ZMV approach for regions with fast motion, we devised two adaptive methods that divide the PC into parts to only carry inter-frame prediction in parts where the ZMV-based prediction is sufficiently accurate.

### 4.1. Single-thresholding segmented decision

The Morton-code-sorted PC is fragmented into sections of 1500 voxels. For each fragment of $\mathcal{F}_{cur}$ ($f_{cur}$), we estimate its prediction using voxels near the fragment in frame $\mathcal{F}_{past}$. The distortion $f'_{cur}$-$f_{cur}$ (in terms of PSNR in dB) is compared to a threshold $t$. If the fragment's inter-frame prediction $f'_{cur}$ is sufficiently reliable, with PSNR above $t$, $f_{cur}$ is discarded. Otherwise, inter-frame prediction is deemed unreliable and $f_{cur}$ is subject to intra-frame prediction only. Hence, in this adaptive method, inter-frame predictions are only used in regions which can be well predicted by $\mathcal{F}_{past}$. Otherwise, we resort to intra-frame-only prediction. The threshold $t$ can be determined as the distortion measure (PSNR in dB) of locally decoded $\mathcal{F}_{past}$. In our tests, the use of the distortion measure of the previous frame locally decoded ($\mathcal{F}_{past}$) was verified to represent a good approximation of the quality expected from each fragment to be considered reliable. Hence, $t$ may vary from frame to frame according to the previous one.

Rate-distortion curves are presented in Fig. 5 for the same PCs as in Fig. 4. This time, for PC "Loot", instead of the large negative gains over the intra-only approach, the inter-frame approach now slightly outperforms it. However, compared to Fig. 4, the performance improvement of the inter-frame mode for "Sarah" has significantly decreased. In other words, with this adaptation method, the inter-frame approach is always advantageous, but the gains in the best case scenario are smaller than in the non-adaptive case.



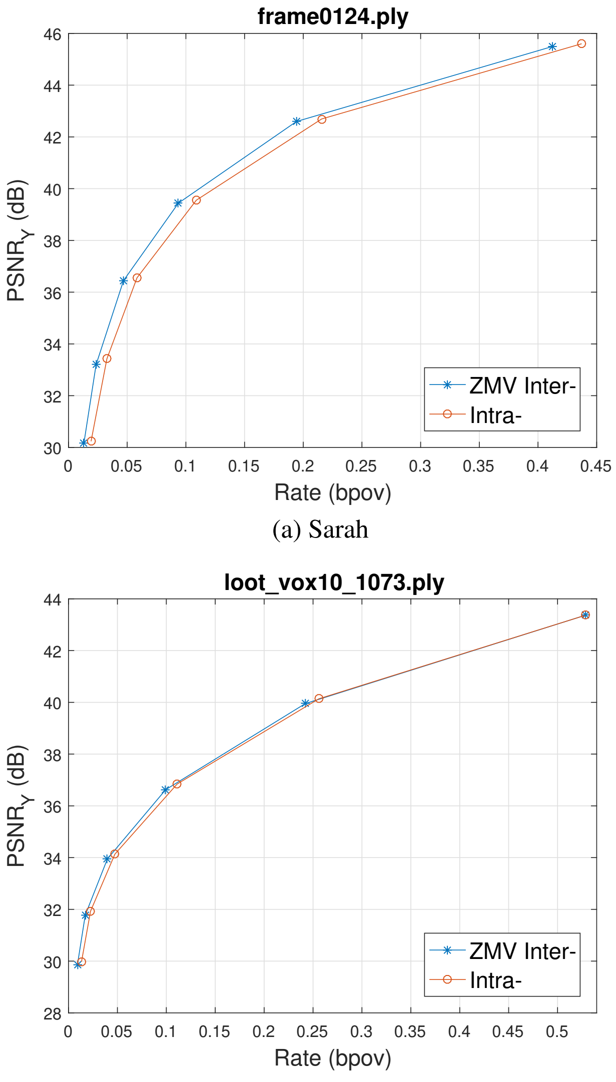
(a) Sarah

(b) Loot

**Fig. 4.** Results of the inter-frame and intra-frame predictive approach compared to intra-predictive approach for PCs "Sarah" and "Loot".

### 4.2. Fragment-based multiple decision

The distortion of the current frame $\mathcal{F}_{cur}$ to its prediction frame $\mathcal{F}'_{cur}$ (in terms of PSNR dB) is compared to a threshold $T$. If $\mathcal{F}'_{cur}$ is not sufficiently reliable, with PSNR below $T$, the method described in section 4.1 is performed. Otherwise, frame prediction is considered reliable and the following is



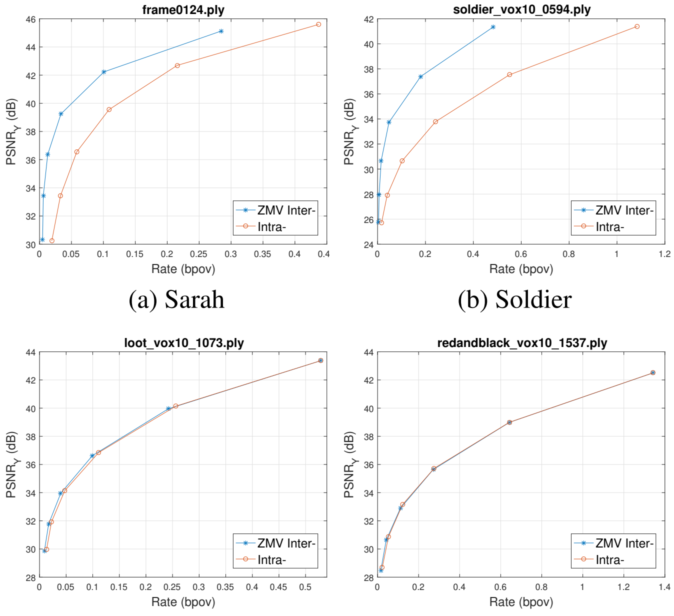
(a) Sarah

(b) Loot

**Fig. 5**. Results of the inter-frame and intra-frame predictive approach compared to intra-predictive approach for PCs "Sarah" and "Loot".

performed: the Morton-code-sorted PC is fragmented into sections of 1500 voxels. For each fragment of $\mathcal{F}_{cur}$ ($f_{cur}$), its prediction is estimated using voxels near the fragment in frame $\mathcal{F}_{past}$. The distortion $f'_{cur}$-$f_{cur}$ (in terms of PSNR in dB) is compared to a threshold $t$. If the fragment's inter-frame prediction $f'_{cur}$ is sufficiently reliable, with PSNR above $t$, $f_{cur}$ is discarded. Otherwise, inter-frame prediction is deemed unreliable, the residual error $f'_{cur} - f_{cur}$ is computed and subjected to intra-frame prediction. Hence, this adaptive method merges both sections 3 and 4.1. The threshold $T$ is defined as 34 dB, based on the theoretical experiment described in section 2 and the threshold $t$ can be determined as the distortion measure (PSNR in dB) of locally decoded $\mathcal{F}_{past}$, as previously defined in 4.1.

Results are presented in Fig. 6 for PCs "Sarah", "Soldier", "Loot" and "Redandblack". For "Loot", the performance of the inter-frame approach remains the same as in section 4.1. Similar to "Loot", the performance obtained for PC "Redandblack" remains unaltered. However, for PC "Sarah" and "Soldier" high gains are achieved. In summary, in this adaptive method, the inter-frame approach is always advantageous and the gains, in the best scenario, are as high as in the non-adaptive case.



(a) Sarah

(b) Soldier

(c) Loot

(d) Redandblack

**Fig. 6**. Results of the inter-frame and intra-frame predictive approach compared to intra-predictive approach for PCs "Sarah", "Soldier", "Loot" and "Redandblack".

## 5. CONCLUSIONS

This work studied the use of predictive RAHT for dynamic PC attribute coding. The combination of simple ZMV inter- and intra-frame predictions have shown positive gains over pure RAHT transform and over intra-frame-only prediction. As an alternative to high computational cost ME algorithms for inter-frame prediction, an adaptive ZMV-based low-computation method, suitable for real-time communications, was used. Future work may include efficient motion estimation and compensation for the inter-frame prediction algorithm, that may be suitable to real-time applications.

## 6. REFERENCES

[1] C. Tulvan, R. Mekuria, Z. Li, and S. Laserre, "Use cases for point cloud compression (pcc)," ISO/IEC JTC1/SC29/WG11 MPEG, output document N16331, Jun. 2016.

[2] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging MPEG standards for point cloud compression," *IEEE J. Emerging Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, March 2019.

[3] "Call for proposals for point cloud compression v2," ISO/IEC JTC1/SC29/WG11 MPEG, output document N16763, April 2017.

[4] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i voxelized full bodies — a voxelized point cloud dataset," ISO/IEC JTC1/SC29/WG1 & WG11 JPEG & MPEG, input documents M74006 & m40059, Jan. 2017.

[5] R. L. de Queiroz and P. A. Chou, "Compression of 3D point clouds using a region-adaptive hierarchical transform," *IEEE Trans. Image Process.*, vol. 25, no. 8, Aug. 2016.

[6] G. Sandri, R. L. de Queiroz, and P. A. Chou, "Comments on "Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform"," *ArXiv e-prints*, May 2018.

[7] G. P. Sandri, P. A. Chou, M. Krivokuća, and R. L. de Queiroz, "Integer alternative for the region-adaptive hierarchical transform," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1369–1372, Sep. 2019.

[8] S. Lasserre and D. Flynn, "On an improvement of raht to exploit attribute correlation," ISO/IEC JTC1/SC29/WG11 MPEG, input document m47378, Jul. 2019.

[9] D. Thanou, P. A. Chou, and P. Frossard, "Graph-based compression of dynamic 3d point cloud sequences," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1765–1778, April 2016.

[10] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 828–842, April 2017.

[11] R. L. de Queiroz and P. A. Chou, "Motion-compensated compression of point cloud video," in *IEEE Int'l Conf. Image Processing (ICIP)*, Sep. 2017, pp. 1417–1421.

[12] C. Dorea and R. L. de Queiroz, "Block-based motion estimation speedup for dynamic voxelized point clouds," in *IEEE Int'l Conf. Image Processing (ICIP)*, Oct 2018, pp. 2964–2968.

[13] C. Dorea, E. M. Hung, and R. L. de Queiroz, "Local texture and geometry descriptors for fast block-based motion estimation of dynamic voxelized point clouds," in *IEEE Int'l Conf. Image Processing (ICIP)*, Sep. 2019, pp. 3721–3725.

[14] C. Loop, Q. Cai, S. O. Escolano, and P. A. Chou, "Microsoft voxelized upper bodies – a voxelized point cloud dataset," ISO/IEC JTC1/SC29 & WG11/WG1 (MPEG/JPEG), input document m38673 & M72012, May 2016.