ADAPTIVE ARITHMETIC CODING FOR POINT CLOUD COMPRESSION

Ismael Daribo #, Ryo Furukawa †, Ryusuke Sagawa ‡, Hiroshi Kawasaki ⋆

National Institute of Informatics, Tokyo, Japan
† Hiroshima City University, Hiroshima, Japan
‡ National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan
⋆ Kagoshima University, Kagoshima, Japan

ABSTRACT

Recently, structured-light-based scanning systems have gain in popularity and are capable of modeling entire dense shapes that evolve over time with a single scan (*a.k.a.* one-shot scan). By projecting a static grid pattern onto the object surface, one-shot shape reconstruction methods can scan moving objects while still maintaining dense reconstruction. However, the amount of 3D data produced by these systems grows rapidly with point cloud of millions of points. As a consequence, effective point cloud compression scheme is required to face the transmission need. In this paper we propose a new approach to compress point cloud by taking advantage of the fact that arithmetic coding can be split into two parts: an encoder that actually produces the compressed bitstream, and a modeler that feeds information into the encoder. In particular, for each position point and normal, we propose to calculate the distribution of probabilities based on their spatial prediction as modeler, while classical point cloud coder mainly focus on the reduction of the prediction residual. Experimental results demonstrate the effectiveness of the proposed method.

***Index Terms* —** Point cloud, arithmetic coding, adaptive model, probability, space curve

1. INTRODUCTION

An increasing number of applications require a dense and precise shape acquisition of moving objects, giving rise to a large amount of data to be stored, transmitted, and rendered. Dense shape acquisition consists mainly of laser-based systems, wherein a laser beam physically moves in one or two dimensions for scanning. Nonetheless, such systems can only reconstruct a single slit per frame, which makes impossible a full dense shape acquisition in a single scan. Currently, 3D scanners that use structured light are also actively researched for scanning 3D shapes [1]. Especially within our scope of interest, one-shot structured-light-based systems that uses a static grid-pattern are capable of modeling dense moving shape in one single scan (*a.k.a.* one-shot scan) [2, 3]. The object surface is sampled by a grid pattern formed by straight lines distinguishable only as horizontal and vertical lines (see Fig. 1). As long as the lines are extractable, the pattern can be as dense as needed. Although, such a pattern is normally not enough to reconstruct the shape, new techniques have been introduced utilizing either coplanarity constraints [2] or probabilistic graphical models [3]. Grid-pattern-based systems are then capable of scanning moving objects while still maintaining dense reconstruction,

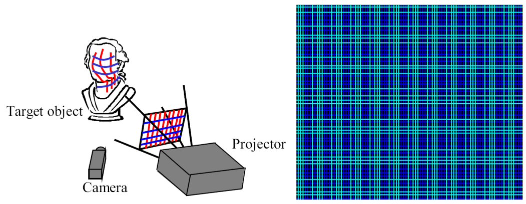


Figure 1. (left) Grid-pattern-based scanning system: a grid pattern is projected from the projector and captured by the camera. (right) Example of projected grid pattern.

which results in 3D models with millions of points.

Nonetheless, the huge amount of raw point data has to be stored or transmitted by efficient compact means. Noise and incompleteness, however, make the process more difficult to achieve. A common way to compress point cloud is to: i) spatially predict each point in a defined order and, ii) entropy code the prediction residual by representing frequently occurring patterns with fewer bits. One of the earliest entropy coding technique is Huffman's algorithm [4], which has been the subject of many studies. Huffman coding is, however, now superseded by the important breakthrough of arithmetic coding [5] that can achieves vastly superior compression performance. In addition, arithmetic coding clearly encourages a clear separation between the statistical model for representing data and the coding of information with respect to that model. It also accommodates adaptive models easily and is computationally efficient. Current point cloud compression methods are broadly based on the before-mentioned scheme, where points are spatially predicted and, the residual is entropy coding. In the followings, we denote these methods as residual-based compression schemes. They broadly assume that the prediction residual follows a Laplacian distribution. The problem, however, is to design a model that at the same time: i) covers a large variety of data and, ii) is accurate enough to fit real probability distribution of the current data to encode.

In this paper, we propose an alternative point compression without residual-based coding. Specifically, we propose to explicitly incorporate the spatial prediction into the arithmetic coder in the form as a statistical model. The model is a way of calculating the distribution of probabilities for the next point. In particular, we first made the observation that structured-light-based scanning systems output points along the measuring direction, which naturally orders groups of points along the same direction: the scan lines. Especially, when the projected grid pattern is extracted from the captured image, points are naturally fitted into a series of 3D

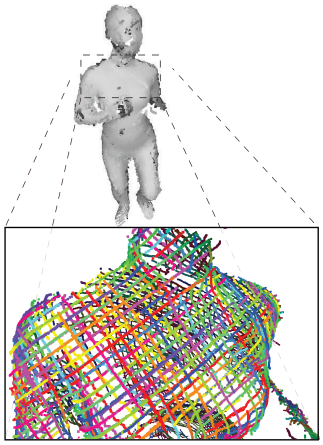


Figure 2. Sampled point cloud partitioned into series of curves wrt to the projected grid pattern. Curves are discriminate by different colors.

space curves. Then, we represent each 3D space curve as a 3D chain code, which is a 3D extension of Freeman chain code [6], where each direction is represented by the curvature of the curve in space. After, each direction along the space curve is entropy coded with respect to his spatial prediction, resulting in an accurate statistical model and, thus, better compression efficiency.

The rest of the paper is organized as follows. We introduce some related work in Section 2. Section 3 addresses the problem of efficiently compressing a point cloud using a spatial-prediction-based entropy coding. Experimental results are presented in Section 4, and finally, our final conclusions are drawn in Section 5.

2. RELATED WORK

The problem of 3D geometry compression has been extensively studied for more than a decade and many compression schemes were designed. Since many important concepts have been introduced in the context of mesh compression, several point cloud compression schemes apply beforehand triangulation and mesh generation, and then use algorithms originally developed for mesh compression , but at the cost of also encoding mesh connectivity. Instead of generating meshes from the point cloud, other approaches propose partitioning the point cloud in smooth manifold surfaces close to the original surface, which are approximated by the method of moving least squares (MLS) [7]. On the other hand, an augmentation of the point cloud by a data structure has been proposed to facilitate the prediction and entropy coding. The object space is then partitioned based on a data structure, *e.g.* octree [8].. Each point is then predicted using a linear combination of the ancestor vertices wrt the data structure. Although not strictly a compression algorithm, the QSplat rendering system offers a compact representation of the hierarchy structure [9]. A high quality rendering is obtained despite a strong quantization.

3. PROPOSED FRAMEWORK

Let us consider the point cloud defined by $\mathcal{S} = \{(p_k, n_k) \mid k \leq N\}$ as a collection of N pairs of 3D points $\{p_k\}$ and normals $\{n_k\}$. As mentioned earlier, structured-light-based 3D scanning systems output points into series of space curves. The point cloud $\mathcal{S}$ can then be represented as a set of M curves $\mathcal{C}^{l_{1\leq l \leq M}}$ as

$$\mathcal{S} = \{\mathcal{C}^1, \mathcal{C}^2, \cdots, \mathcal{C}^M\} \tag{1}$$

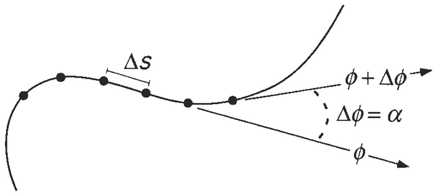


Figure 3. 2D example of plane curve sampled at intervals of arc length ΔS. Each point has a turning angle α as the angle between two consecutive segments.

where a l-ieme curve $\mathcal{C}^l$ is expressed as

$$\mathcal{C}^l = \{(p_r, n_r), (p_{r+1}, n_{r+1}), \cdots, (p_s, n_s)\} \tag{2}$$

with $1 \leq r < s < N$. The organization of the point cloud into series of space curve can be straightforwardly obtained from the lien detection algorithm during the acquisition process [11].

For simplicity of notation, let $\mathcal{C}$ be the current curve to encode. Each point p_k and normal n_k in $\mathcal{C}$ are spatially predicted with respect to the previous reconstructed points $\widetilde{p}_i$ with $i < k$.

3.1. 3D Chain Code Representation

We propose to parameterize a space curve by its length ΔS and its turning angle α (the angle between two consecutive segments), as illustrated in Fig. 3. Given the vector $\mathbf{v_k} = p_k - p_{k-1}$ between two consecutive point. the turning angle in 3D space between two vectors is defined by the triplet angles α_{v_k} $(\alpha_x, \alpha_y, \alpha_z)$ as the difference of their direction angles. Given a vector $\mathbf{v_k}(v_{k\,x}, v_{k\,y}, v_{k\,z})$, his direction angles $\theta\left(\theta_{v_{k\,x}}, \theta_{v_{k\,y}}, \theta_{v_{k\,z}}\right)$ are expressed by

$$\begin{aligned} \cos\left(\theta_{v_{k\,x}}\right) &= \frac{v_{k\,x}}{\|\mathbf{v_k}\|}, \\ \cos\left(\theta_{v_{k\,y}}\right) &= \frac{v_{k\,y}}{\|\mathbf{v_k}\|}, \\ \cos\left(\theta_{v_{k\,z}}\right) &= \frac{v_{k\,z}}{\|\mathbf{v_k}\|}. \end{aligned} \tag{3}$$

Then, the space curve is represented by a starting point followed by a list of turning angles, where after quantization each turning angle can be replaced by a code in the same manner as the Freeman chain code [6].

3.2. Spatial Prediction

Let us describe the prediction approach used in the proposed framework for point location and normal. Though we chose the following specific prediction approached for concreteness, it is important to note that our proposed framework extends beyond those specific prediction schemes.

3.2.1. Point Location Prediction

The prediction problem of the current point location p_k is reformulated as the prediction problem of its associated turning angle α_k. Considering that consecutive points are closed enough, it is reasonable to assume that the curve is locally turning constantly, *i.e.*, the difference between two consecutive turning angles is near zero. The predicted turning angle is then expressed as the linear combination of its two previous ancestors as follows

$$\widetilde{\alpha}_k = 2 \cdot \widetilde{\alpha}_{k-1} - \widetilde{\alpha}_{k-2} \tag{4}$$

where the reconstructed point location $\widetilde{p}_k$ is defined by

$$\widetilde{p}_k = \widetilde{p}_{k-1} + \mathcal{R}\left(\widetilde{\alpha}_k\right) \circ \mathbf{v_{k-1}}, \tag{5}$$

with $\mathcal{R}\left(\widetilde{\alpha}_k\right) \circ \mathbf{v_{k-1}}$ being the 3D rotation of the vector $\mathbf{v_{k-1}}$ wrt the predicted turning angle $\widetilde{\alpha}_k$.

3.2.2. Point Normal Prediction

The surface of many objects consist of repetitive patterns and similar structure. To exploit this repetitiveness, we propose to search for a similar portion of curve in terms of shape prior to a close Euclidean invariant signature. Let us first define the signature of a space curve, up to a Euclidean transform, by its curvature function $\kappa(n{\cdot}\Delta S)$ and torsion function $\tau(n{\cdot}\Delta S)$, both functions of the parameter n. It was shown in [12] that κ and τ can be approximated at the point p_k by

$$\kappa(p_k) = \pm 4 \cdot \frac{\sqrt{s\cdot(s-a)\cdot(s-b)\cdot(s-c)}}{a\cdot b\cdot c}, \tag{6}$$

$$\tau(p_k) = \pm 6 \cdot \frac{H}{d\cdot e\cdot f\cdot \kappa(p_k)}. \tag{7}$$

where H being the height of the tetrahedron form by $p_{i-1}, p_i, p_{i+1}, p_{i+2}$ of base p_{i-1}, p_i, p_{i+1}, and

$$a = d(p_{k-1}, p_k), \quad b = d(p_k, p_{k+1}), \quad c = d(p_{k-1}, p_{k+1}),$$
$$d = d(p_{k+1}, p_{k+2}), \quad e = d(p_k, p_{k+2}), \quad f = d(p_{k-1}, p_{k+2}),$$

and $s = \frac{1}{2}(a+b+c)$. Since κ and τ only depends on the Euclidean distance $d(.,.)$ between points, they provide a completely Euclidean invariant numerical signature approximation. Once the best similar portion of curve is selected, the geometric transformation (*i.e.*, translation, rotation) from the best candidate and the current portion of curve is estimated by least-squares fitting [13]. Finally, the current normal is predicted.

3.3. Spherical Quantization

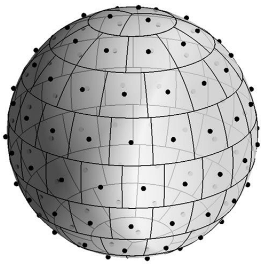


Figure 4. Quantization cells covering the surface of a 3D unit sphere and the corresponding vectors in black dot.

Point locations through their turning angle are expressed in spherical coordinates. We also propose to express normals in spherical coordinates. The quantization is then achieved through an uniform distribution of the quantization cells on the surface of an unit sphere as shown in Fig. 4, at a given quantization parameter interval.

3.4. Adaptive Statistical Model

Under the assumption of a piecewise smooth space curve, the turning angle distribution can be modeled by a von Mises probability distribution, defined below, to assign probability to the current turning angle α_k:

$$p\left(\alpha_k | \alpha_{k-1}, \alpha_{k-2}\right) = \frac{1}{2\pi \cdot I_0\left(\kappa\right)} \cdot e^{\kappa \cos\left(\widetilde{\alpha}_k - \mu\right)} \tag{8}$$

where $I_0\left(.\right)$ is the modified Bessel function of order 0. The parameters μ and $1/\kappa$ are respectively the mean and variance in the circular normal distribution; we set $\mu = \alpha_{k-1}$ in our case. The von Mises distribution is the natural Gaussian distribution for angular measurements. We argue this is an appropriate choice because: i) it maximizes the probability when the curve turn constantly ($\widetilde{\alpha}_k = \alpha_{k-1}$), and ii) it decreases symmetrically in left / right directions as the turning angle deviates from the predicted one. We apply the same strategy for the normal vector in the quantization sphere space, where the distance between two normal vector is the number of distant cells on the surface of the quantization sphere: *e.g.* if they belongs to the same sphere, the distance is zero; if they belongs to neighboring cells, the distance is equal to one; and so on.

3.5. Adaptive Arithmetic Coding

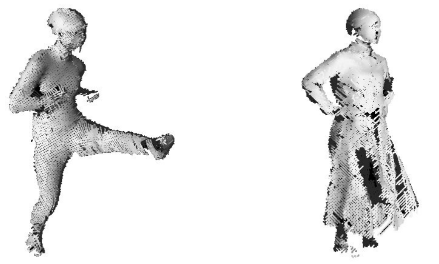
Arithmetic coding is a powerful entropy coding, and it represents the current state-of-the-art in lossless compression. One important feature of arithmetic coding is that the actual encoding and modeling of the source can be completed separately. Thus, we can design our own statistical model that fits our particular application and use arithmetic coding in a straight-forward manner. It has been used in image and video coding for entropy coding for well over a decade. Unlikely Huffman coding that replaces each symbol by a variable-length code, arithmetic coding associates the entire curve $\mathcal{C}$ with a sub-interval $[a, b)$ inside the original interval of $[0, 1)$, such that:

$$\begin{aligned} b - a &= p\left(\mathcal{C}\right) \\ &= p\left(\alpha_1, \alpha_2, \ldots, \alpha_N\right) \\ &= p^{(1)}\left(\alpha_1\right) \cdot p^{(2)}\left(\alpha_2|\alpha_1\right) \cdot \ldots \\ &\qquad \ldots \cdot p^{(N)}\left(\alpha_N|\alpha_{N-1}, \alpha_{N-2}\right) \end{aligned} \tag{9}$$

The conditional probability reflects the prediction of the current point prior to the previous ones. We then used the statistical model previously designed to adaptively estimate the probability of each point in the point cloud.

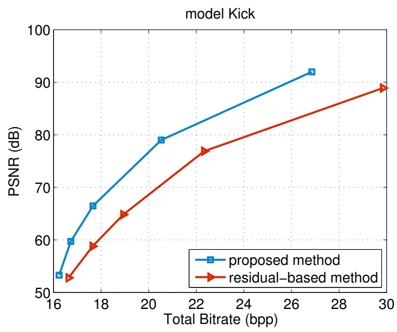
4. EXPERIMENTAL RESULTS

The performance of the proposed framework is evaluated using the two models shown in Figure 5. The objective compression performance of the proposed method is investigated in the rate-distortion (RD) curves plotted in Figure 6 through the average number of bits per points (bpp), in relation to the loss of quality, measured by the peak signal to noise ratio (PSNR). The PSNR is evaluated using the Euclidean distance between points. The peak signal is given by the length of the diagonal of the bounding box of the original model. The bitrate corresponds to the total rate of the position points and the normals. We compare our model-based approach against classical residual-based approach by using the same prediction techniques described in this paper. We see that our new model-based approach results in significant compression gain. Specifically, an average bitrate reduction up to 7% are observed.

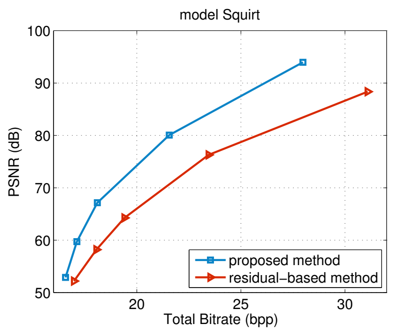


(a) Model Kick (b) Model Squirt

Figure 5. Test models captured with a setup of multiple projectors and cameras [2].



(a)



(b)

Figure 6. Rate-distortion performance of the proposed encoder.

5. CONCLUSION

We designed and implemented a model-based predictive single-rate compression for the positions points and normals outputted by a grid-pattern-based 3D scanning system. The compression is achieved by directly incorporate the spatial prediction into the arithmetic coder as a statistical model. Several issues remain that warrant further research. In future studies, we intend to extend this work to arbitrary dense point cloud.

Acknowledgment

This work is supported in part by Strategic Information and Communications R&D Promotion Programme (SCOPE) No.101710002 (Ministry of Internal Affairs and Communications, Japan), Funding Program for Next Generation World-Leading Researchers No.LR030 (Cabinet Office, Government Of Japan), and the Japan Society for the Promotion of Science (JSPS) Program for Foreign Researchers.

6. REFERENCES

[1] Joan Batlle, El Mustapha Mouaddib, and Joaquim Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," *Pattern Recognition*, vol. 31, pp. 963–982, 1998.

[2] R. Furukawa, R. Sagawa, H. Kawasaki, K. Sakashita, Y. Yagi, and N. Asada, "One-shot entire shape acquisition method using multiple projectors and cameras," in *Proc. of the Pacific-Rim Symposium Image and Video Technology (PSIVT)*, Singapore, Dec. 2010, pp. 107–114.

[3] A. O. Ulusoy, F. Calakli, and G. Taubin, "Robust one-shot 3D scanning using loopy belief propagation," in *Proc. of the IEEE Computer Vision and Pattern Recognition Workshops (CVPRW)*, San Francisco, CA, USA, June 2010, pp. 15–22.

[4] D.A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098 –1101, sept. 1952.

[5] Ian H. Witten, Radford M. Neal, and John G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 520–540, June 1987.

[6] Herbert Freeman, "On the encoding of arbitrary geometric configurations," *IRE Transactions on Electronic Computers*, , no. 2, pp. 260–268, 1961.

[7] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva, "Computing and rendering point set surfaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, no. 1, pp. 3–15, 2003.

[8] Ruwen Schnabel and Reinhard Klein, "Octree-based point-cloud compression," in *Proc. of the IEEE Eurographics Symposium on Point-Based Graphics*, M. Botsch and B. Chen, Eds. July 2006, Eurographics.

[9] Szymon Rusinkiewicz and Marc Levoy, "QSplat: A multiresolution point rendering system for large meshes," in *Proc. of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, July 2000, pp. 343–352.

[10] Ryusuke Sagawa, Yuichi Ota, Yasushi Yagi, Ryo Furukawa, Naoki Asada, and Hiroshi Kawasaki, "Dense 3D reconstruction method using a single pattern for fast moving object," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, Sept. 2009, pp. 1779–1786.

[11] E. Calabi, P.J. Olver, C. Shakiban A. Tannenbaum, and S. Haker, "Differential and numerically invariant signature curves applied to object recognition," *International Journal of Computer Vision*, vol. 26, pp. 107–135, 1998.

[12] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, May 1987.