ISO /IEC JTC 1/SC 29 /WG 1 **M78030**
78th Meeting – Rio de Janeiro, Brazil – 29 Jan-2 Fev 2018

ISO/IEC JTC1/SC29/WG1
(ITU-T SG16)

Coding of Still Pictures

JBIG JPEG

Joint Bi-level Image Experts Group	Joint Photographic Experts Group

TITLE: Objective Metrics and Subjective Tests for Quality Evaluation of Point Clouds

SOURCE:
Ricardo L. De Queiroz
Eric Torlig
Tiago A. Fonseca

Universidade de Brasília, Brazil
e-mails: queiroz@ieee.org , {eric , tiago}@image.unb.br

PROJECT: JPEG Pleno

STATUS: For information

REQUESTED ACTION: For review

DISTRIBUTION: WG1

Contact:
ISO/IEC JTC 1/SC 29/WG 1 Convener – Prof. Touradj Ebrahimi
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2606, Fax: +41 21 693 7600, E-mail: Touradj.Ebrahimi@epfl.ch

Abstract: In order to evaluate compression methods for a point-cloud-based 3D representation, we propose an objective metric to evaluate distance (distortion) in between two point clouds. The method needs to be simple and to capture some visual properties of the object. We propose to convert points into volumetric elements (voxels) which are orthographically projected onto N image planes, which are combined into a single image. The resulting images (one for each cloud) are then compared using for example peak signal-to-noise ratio (PSNR), which results in the metric we call *projection* PSNR or P-PSNR. (It can be trivially extended so P-SSIM or P-VQM.) The more views, the better coverage of the possible viewpoints. However, $N = 6$ to $N = 14$ may be sufficient. Any arbitrary number N of uniformly distributed directions can be chosen using a spherical Fibonacci point set. Preliminary subjective tests corroborate the correlation with the proposed metric.

1) Introduction

The definition of 3D representation of objects for AR/VR applications would eventually demand compression methods, which need to be evaluated. A definition of distortion in between point clouds (PCs) that would somehow relate to the visual distance between two PCs is still elusive and largely unexplored. It has been observed an effort to create objective metrics to assess the quality of processed point clouds [5]-[7]. However, the approaches tend to separate the analysis in two evaluations: geometric quality assessment and colour quality assessment. Our proposal embeds both geometric and colour attributes assessment in one objective metric. We propose a simple yet efficient metric to compute distortion between two PCs that avoids full searches and does not require the PCs to have the same number of points. It just relates to its appearance under many viewpoints.

Instead of using a dense volumetric signal with all possible RGB samples for each frame, a 3D signal can be represented as a list of points. Each point can then be considered as a combination of a spatial address (x,y,z) and point attributes such as normal coordinates or colour. Hence, a PC can be described by a list of points as:

$$(x, y, z, R, G, B, n_x, n_y, n_z) .$$

Instead of processing infinitely small points that are usually rendered with view dependent splatters, we use the notion of volumetric element, or voxel, which is said to be occupied or unoccupied. Furthermore, for each occupied position, the surface colour (RGB) is recorded, and we ignore other attributes.

2) The method

Voxelization

Here, voxels regularly tile the 3D space. Much like pixels represent an image in a discrete regular grid, voxels represent the volume in a discrete grid. Voxels are the smallest units to be individually addressed and can be represented by their integer position on the grid. If the space into consideration has dimensions $W_x \times W_y \times W_z$ and the voxels have dimensions $D_x \times D_y \times D_z$, which are the spatial sampling intervals, the grid of voxels has size $N_x \times N_y \times N_z$ where $N_x = W_x/D_x$, $N_y = W_y/D_y$, $N_z = W_z/D_z$.

Voxels contrast with points which can have any precision, be located anywhere in space, and encompass no volume. For that, as we want to work with voxels, a points list need to be converted into a list of occupied voxels. A simple way to do so is to list all points that fall within a voxel and average their attributes (Figure

1). The resulting average is used as attribute for that voxel which is marked occupied. Usually, making the voxel resolution high enough, most voxels would contain at most one point. In other words, if v_x = `floor`(x/D_x), v_y = `floor`(y/D_y) and v_z = `floor`(z/D_z), then the point cloud is represented as a list of occupied voxels as

$$\{v_x, v_y, v_z, R, G, B\},$$

where the attributes of co-located voxels are averaged to produce a unique voxel attribute.

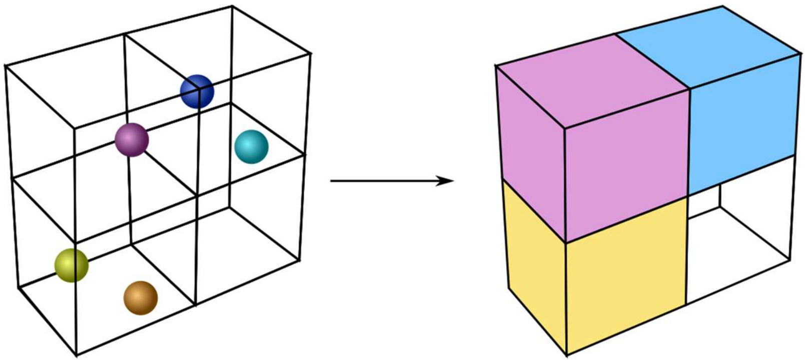

Figure 1: Voxelizing a PC.

Voxels as an abstraction remove one issue with PC rendering which is the dependence on the point of view. As a volume in itself, each voxel can be uniquely rendered from any viewpoint. In essence, the choice of point or splatter size is replaced with the voxel size or grid resolution.

Projection

For many applications using PC and AR/VR one should be concerned with the visual appearance of the object. The problem is that it depends on the point of view, and there is obviously an infinite number of them. Consider a simple orthographic projection, a form of parallel projection, taken from a determined view direction, assuming the observer is at an infinite distance, such that light rays are assumed parallel in that direction. A parallel projection is a two-dimensional representation of the surface of a three-dimensional object, and represents a viewpoint (Figure 2).

Assuming an orthographic camera is a good-enough surrogate for an actual observer at that orientation, we remove a degree of freedom (distance) and in polar coordinates ($0 \leq \theta \leq 2\pi$, $-\pi/2 \leq \phi \leq \pi/2$) we need to cover the ($\theta,\phi$) plane. We do not need an infinite number of views to visualize all voxels and if one has to sample the (θ,ϕ) plane two immediate options are to regularly sample the sphere coordinates or to draw random positions. We opt for the former.

Figure 2: 3 parallel projections of a PC.

Assume we would use N projections (directions). Regular spherical sampling relates to perfect solids and an arbitrary number of uniformly distributed directions can be generated using a spherical Fibonacci point set [10]. We, however, propose N=6 (Figure 3), such that projections are taken from the vertices of an octahedron, or the faces of a cube, starting with the point cloud rotated by 0 degrees in the polar and azimuthal directions. From there, rotations are taken in sequence in orthogonal directions. In (θ,ϕ) coordinates this means projections are taken at (0, 0), (0, π/2), (0, π), (0, -π/ 2), (π/ 2, 0) and (-π/ 2, 0). Figure 4 also shows the projections in the case of N = 14.

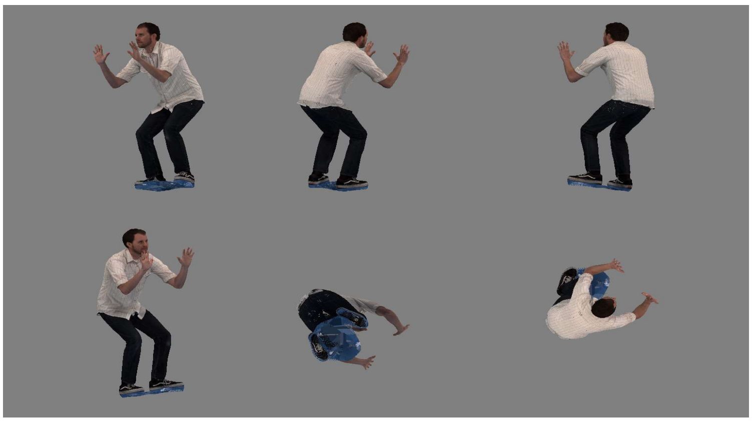

Figure 3: PC projections for *N*=*6*.

Figure 4: PC projections for $N=14$.

The great advantage of using the 6-views approach is computational, since it does not require interpolation or rotation of the voxel data. It can be made to fit the voxel walls directly onto image pixels. The colour of the closest voxel to the projection plane can be transferred directly to a pixel at that position in the projection image. The 6 resulting images have dimensions $N_x \times N_y$, $N_x \times N_z$, and $N_y \times N_z$. For example if we use a grid of 512 x 512 x 512 voxels, we end up with 6 easily computable 512x512-pixel images.

Note that if at a given position (v_x, v_y) all voxels at (v_x, v_y, v_z) for all v_z are unoccupied, then the scene is transparent and the pixel at coordinates (v_x, v_y) at the corresponding projection images should be empty as well. Since we cannot assign empty pixels in regular images we assign a default colour. For example, we assumed neutral grey, i.e. all RGB values are 128 in a scale from 0 to 255.

Image metrics

Once we derive an image from a PC with the voxel projection views, we assume we can use typical image distance metrics in between them. Let we have PCs 1 and 2, from which we derive two projection images:

PC1 → Projection Image 1 (PI1)
PC2 → Projection Image 2 (PI2)

Then, for 8-bit color attributes, we compute

$$\text{P-PSNR} = 10 \log_{10} \frac{255^2}{MSE(PI1, PI2)}$$

One can replace the PSNR by other metric such as VQM [8] or SSIM [9], thus creating P-VQM or P-SSIM.

3) Subjective Tests

Despite its expected correlation with the visual quality, we ran subjective tests in order corroborate the P-PSNR metric. An informal subjective evaluation was conducted. A set of 4 point clouds were assessed by 12 people under the home viewing room conditions [3] using a 47-inch LCD TV and computer to control the free viewpoint.

Each point cloud was analyzed under 4 conditions:
- original version (no processing);
- aggressive colour compression [4];
- light colour compression [4];
- low-pass (downsampled and upsampled) version.

For each point cloud and condition, each test subject was required to assess:
- Geometric quality;
- Colour quality;
- Overall quality.

The evaluation considered ITU recommendations [3] and followed a multi-stimulus continuous quality-scale method to assess the quality. Figure 5 shows the overall quality aspect obtained from the subjective quality assessment. The values 1-5 correspond to the Mean Opinion Scores (MOS) of the results the overall quality. The error bars represent one standard deviation around the mean opinion score.

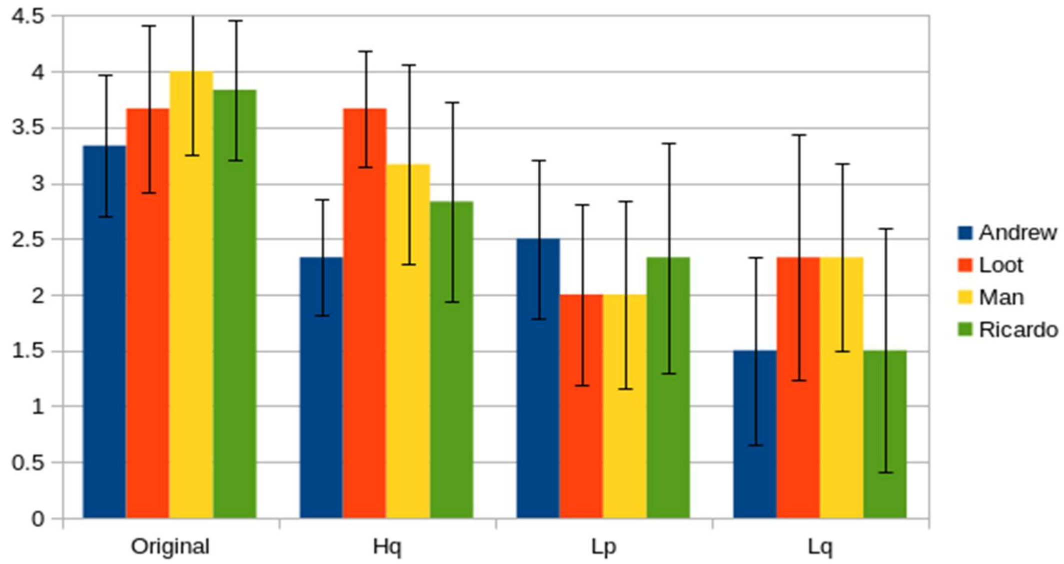

Figure 5: Subjective Results, overall quality of the 3D scenes (MOS). Hq stands for high-quality version (light colour compression), Lp represents low-pass version meanwhile Lq refers to low-quality version (aggressive colour compression). Andrew and Ricardo are upper body scenes [2].

Table 1 presents the P-PSNR values for the test set.

Table 1: P-PSNR (dB) values for the test set.

Sequence	Hq	Lp	Lq
Andrew	51.76	34.59	40.87
Loot	55.49	44.04	47.99
Man	53.89	36.47	44.32
Ricardo	56.13	37.14	49.53

A preliminary analysis demonstrates that P-PSNR can discriminate a high-quality signal from others. Of the two other quality aspects assessed in the preliminary tests, colour quality was the one most correlated with overall quality MOS. Results also show a positive correlation between overall MOS and P-PSNR, or a negative correlation between differential mean opinion score (DMOS), with respect to the original signals, and P-PSNR.

The relative relevance of geometric noise and of colour noise to the overall quality assessment is still an open question. Our exploratory study does not have an adequate number of observers to derive further conclusions. A greater number of observers should undergo subjective evaluation to determine the relative relevance of geometric defects over colour defects on the quality assessment. What is important to us at this point is to see how P-PSNR correlates with overall subjective assessment.

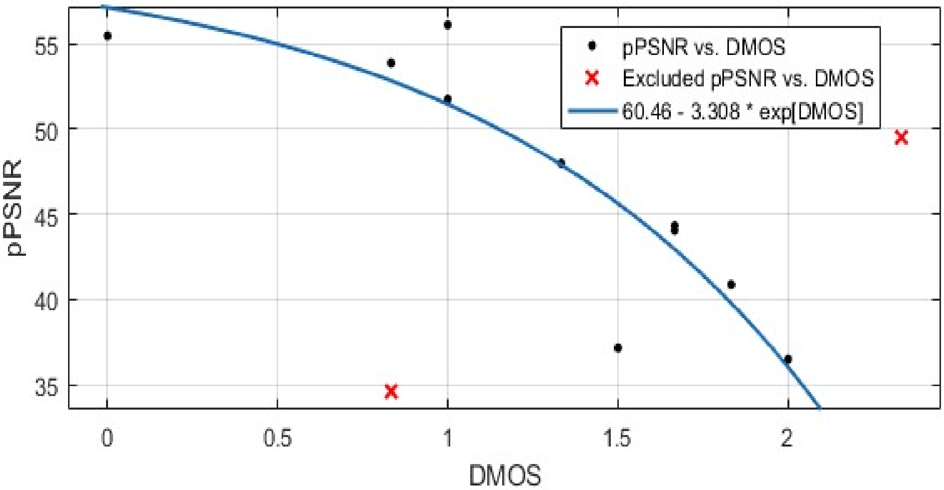

Figure 6: Relation between DMOS and P-PSNR. Preliminary experimental results indicate a negative proportionality.

4) Conclusions

An objective metric for quality evaluation was proposed and an informal subjective evaluation verified a significant correlation between the proposed framework and subjective quality evaluation.

The results in Figure 6 relating P-PSNR and DMOS are very much in line with the results relating DMOS and PSNR in conventional image compression and that encourages us to propose P-PSNR as a simple metric to evaluate point cloud distortion with aim at standardizing a compression scheme for point clouds.

5) References

[1] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan, "High-quality streamable free-viewpoint video," *ACM Trans. Graph.*, Vol. 34, No. 4, pp. 69:1–69:13, July 2015.

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

[3] Methodology for the Subjective Assessment of the Quality of Television Pictures, *document ITU-R BT.500-13*, Geneva, Switzerland, Jan. 2012.

[4] R. L. de Queiroz and P. 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, Aug. 2016.

[5] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *Proc. IEEE Intl. Conf. Image Processing*, September 2017.

[6] E. Alexiou and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Applications of Digital Image Processing XL*, Vol. 10396, p. 103961H, International Society for Optics and Photonics, September 2017.

[7] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," in *Proc. IEEE Intl. Conf. Multimedia & Expo Workshops (ICMEW)*, 2017.

[8] M. Pinson and S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting*, Vol. 50, No.3, pp. 312-322, September, 2004.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, Vol. 13, No. 4, pp. 600-612, 2004.

[10] Á. González, "Measurement of Areas on a Sphere Using Fibonacci and Latitude–Longitude Lattices", *Mathematical Geosciences*, Vol. 42, No. 1, pp. 49-64, 2010. https://doi.org/10.1007/s11004-009-9257-x