3D Point-Cloud Quality Assessment Using Color and Geometry Texture Descriptors

Rafael Diniz
Advisor: Mylène C.Q. Farias
University of Brasília, Brazil
http://www.gpds.ene.unb.br/rafael/

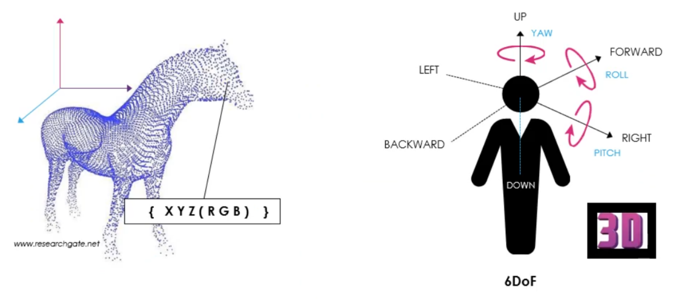
Defesa de Tese de Doutorado, 26 de Julho de 2021, Brasília - DF

Universidade de Brasília

Summary Of The Presentation

Introduction: Beyond The Window To The World (2D)

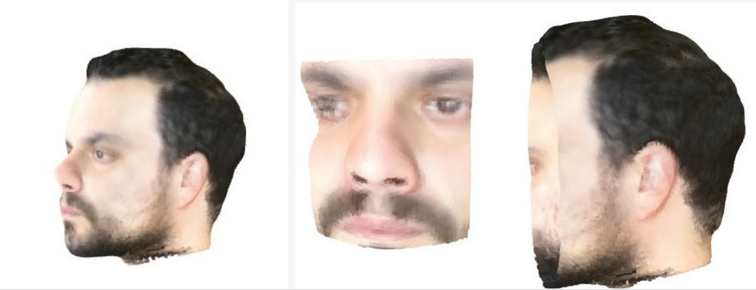
- Visual 3D Representations
- Voxels Instead Of Pixels
- 6 Degrees of Freedom



Introduction

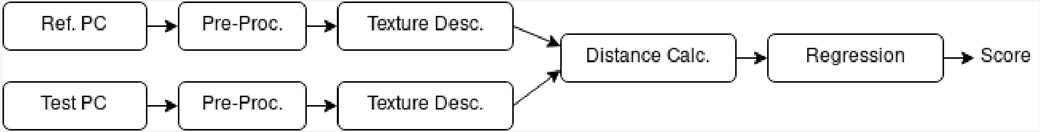
Problem Description

- Point Cloud (PC) units contain 3D spatial information (eg. x, y, z) and color information (eg. R, G, B)
- Points (or voxels) are sparsely distributed in the 3D space
- New methods for automatic assessment the quality of 3D content are needed!



Introduction: Proposed Method

- Use color-based and geometry-based texture descriptors to estimate the quality of a degraded PC.
- The proposed objective full-reference PCQA method is based on the following steps:
  1. Pre-processing (Voxelization Methodology / Normals Calculation)
  2. PC texture descriptors (LBP, LLP, LCP and GEO) application
  3. Descriptors histogram distance calculation (reference vs test)
  4. Quality prediction model based on a regression algorithm



Introduction: Summary Of Contributions

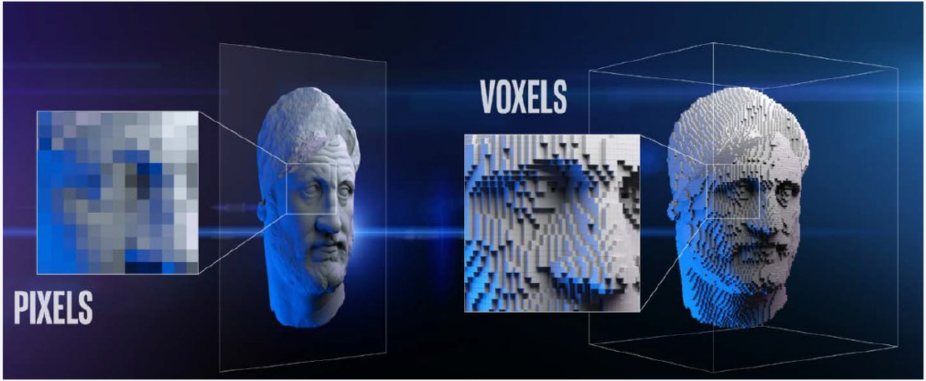
- Parameterized voxelization method
- 4 novel PC texture descriptors based on local PC neighborhoods
- Statistical analysis of the proposed texture descriptors on different data-sets
- A model for PC quality assessment based on texture descriptors

Introduction: Publications

- **Color and Geometry Texture Descriptors for Point-Cloud Quality Assessment**. IEEE Signal Processing Letters, 2021.
- **A novel point cloud quality assessment metric based on perceptual color distance patterns**. Electronic Imaging, 2021.
- **Towards a Point Cloud Quality Assessment Model using Local Binary Patterns**. International Conference on Quality of Multimedia Experience, 2020.
- **Multi-Distance Point Cloud Quality Assessment**. IEEE International Conference on Image Processing, 2020.
- **Local Luminance Patterns for Point Cloud Quality Assessment**. IEEE International Workshop on Multimedia Signal Processing, 2020.
- **Real-time 3D volumetric human body reconstruction from a single view RGB-D capture device**. Electronic Imaging, 2019.

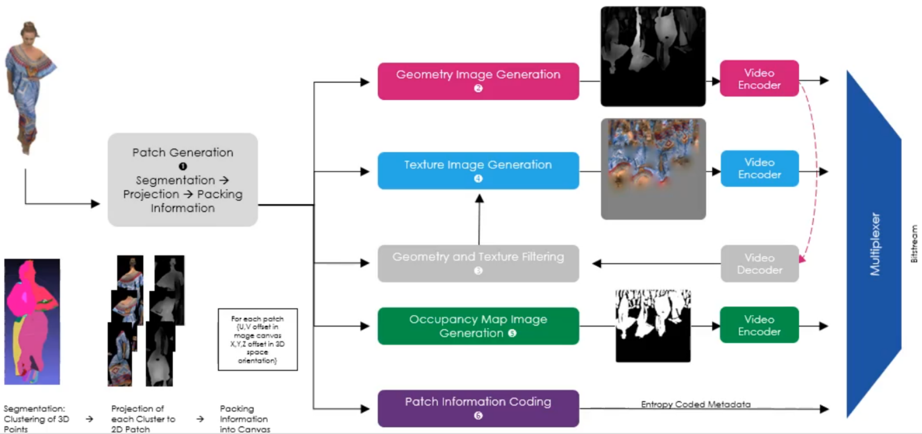
Overview: More Degrees Of Freedom...

Overview: From Pixels to Voxels ...



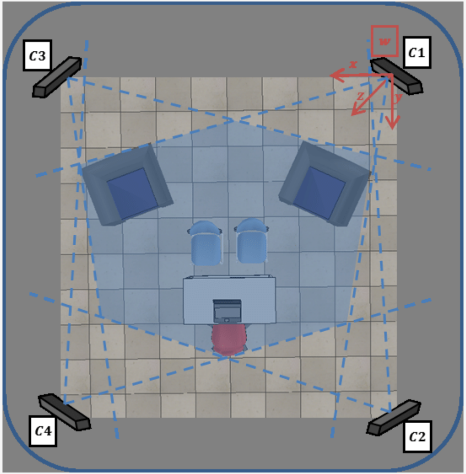
Overview: Immersive Media Formats

- Examples: Point Cloud, Mesh, Light Field and Hologram.
- Preferred format for live real-time workflows: Point Cloud
- New standards: ISO/IEC 23090-5:2021 (V-PCC) and 23090-9 DIS (G-PCC)



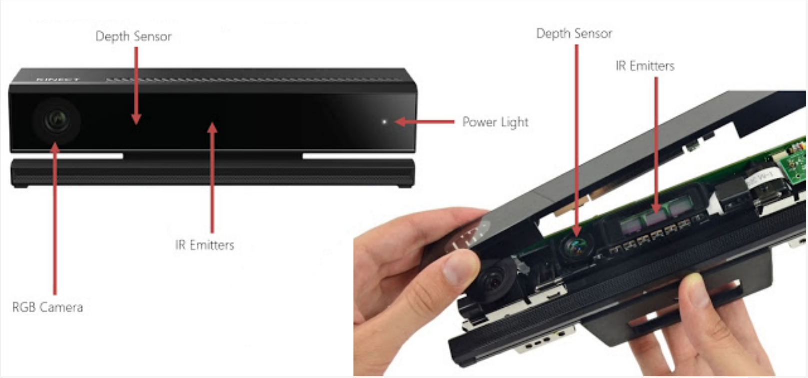
Overview: Capture

Typical setup: Array of RGB+D cameras



Overview: Capture

RGB+D Hardware Eg.: Kinect v2



Overview: Display

Typical display: HMD with environment and head-tracking sensors

Overview: Point Cloud Quality Assessment

- Objective: Predict the quality in an automatic way
- Subjective: Humans evaluate and rate the content

Subjective X Objective

Feature	Subjective measure (SM)	Objective measure (OM)
Human involvement (observer)	Yes	No
Automatic	No	Yes
Mathematically defined algorithms	No	Yes
Expensive evaluation	Yes	No
Computational complexity	No	Yes
Inconvenient	Yes	No
Time consuming	Yes	No

Overview: PC Subjective Quality Assessment

- Very important for the development of Objective metrics
- Typically just extend the protocols for 2D image/video to PC
- ITU-R BT.500-14: "Methodologies for the subjective assessment of the quality of television images"
- Subjective evaluation can be passive (no user interaction) or active (with 6dof interaction)

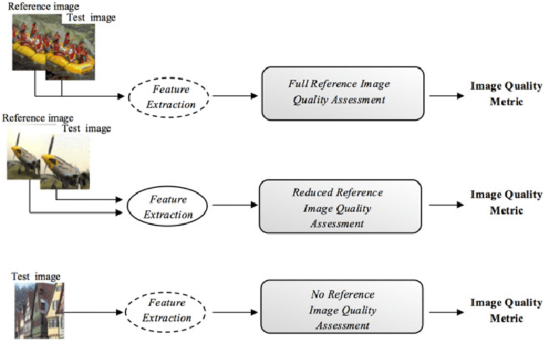
MOS Scale

Level	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, but not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

Overview: Objective Quality Assessment

- Full-reference (FR): Uses all the information of the reference
- Reduced reference (RR): Uses partial information of the reference
- No-reference (NR): Uses no reference information

Objective QA Types

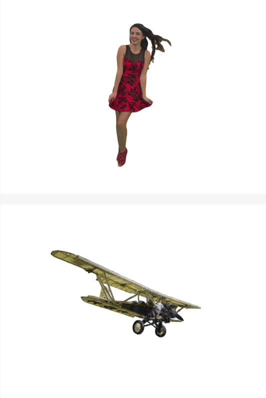


Overview: State-of-the-art in Subjective PCQA

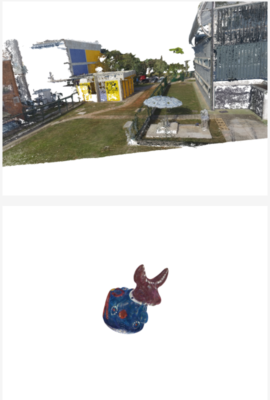
PC subjective experiments can evaluate different types of PC degradation. Some work evaluate only geometry distortions, other just color distortions, while other jointly evaluate both.

- Torlig et al., **A novel methodology for quality assessment of voxelized point clouds**, Applications of Digital Image Processing, International Society for Optics and Photonics, 2018.
- Alexiou et al., **A comprehensive study of the rate-distortion performance in mpeg pointcloud compression**, Transactions on Signal and Information Processing (APSIPA), 2019.
- Stuart et al., **Quality evaluation of static point clouds encoded using MPEG codecs**, IEEE International Conference on Image Processing (ICIP), 2020.
- Yang et al. **Predicting the Perceptual Quality of Point Cloud: A 3D-to-2D Projection-Based Exploration**, IEEE Transactions on Multimedia, 2020.

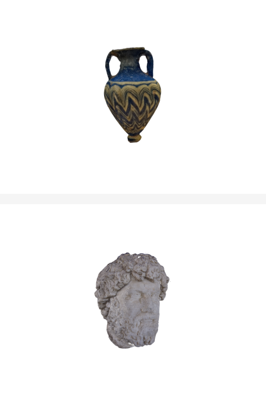
Overview: Data Sets



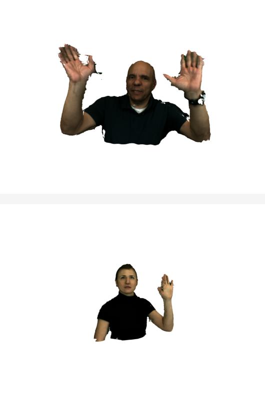
D1



D2



D3



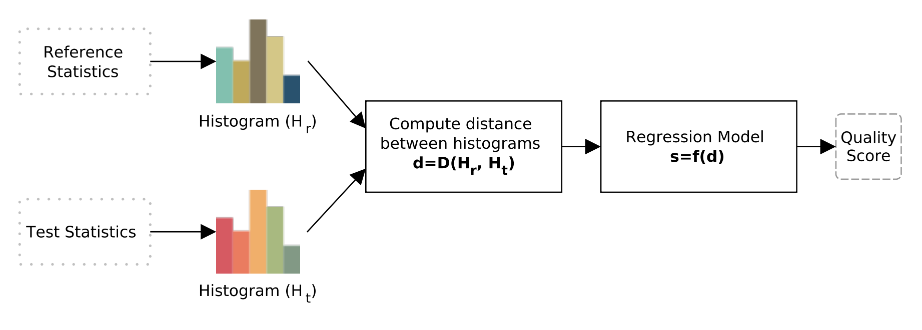
D4

Overview: PC Objective Quality Assessment

- Point-based metrics: Based on the error between attributes of correspondent points in the impaired and reference content. Different types: Po2Po, Po2Pl and Pl2Pl. By Tian, Mekuria, Alexiou and others (MPEG metrics);
- PointSSIM: **Towards a point cloud structural similarity metric**, by Alexiou et al;
- PCQM: **A full-reference quality metric for colored 3D point clouds**, by Meynet et al;
- PCM_RR: **A reduced reference metric for visual quality evaluation of point cloud contents**, by Viola et al;
- **Graph**-based encoding by Yang et al.;
- **Machine learning**-based metrics were also proposed, by Liu et al.

Color And Geometry Textures For Point Cloud Quality Assessment

- Pre-Processing
- Color and Geometry Texture Descriptors For PC
- Texture Histogram Distances
- Quality Prediction Model



Voxelization Pre-Processing

- PC points are sparsely distributed in the 3D space without volumetric meaning
- Voxelization convert point(s) to discrete volumetric units (voxels)
- One alternative to the voxels are meshes, but the mesh representation leads to higher computational cost, as capture devices capture RGB+Depth information, which are easily converted to PC, but have no point connectivity information needed by a mesh
- The definition of the voxel size is important, if too small, neighboring voxels may not touch each other, leaving visual “holes” between PC elements, while oversized voxels creates a swollen visual effect

Voxelization With Different Voxel Sizes

Voxelization Pre-Processing

Considering a cube-shaped voxel, the following heuristic is defined to obtain a the voxel size, where ES is the edge size of the cube (voxel).

$$\mathsf{ES} = \frac{k}{S} \cdot \sum_{n=1}^{S} \left(\frac{1}{k_{nn}} \cdot \sum_{i=1}^{k_{nn}} \mathbf{d}\left(N_i(P_n), P_n\right) \right)$$

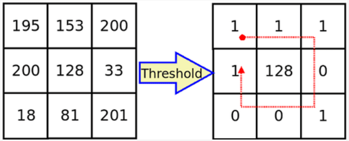
- S is the number of points of the PC
- k is a constant tested with different values (a multiplier of ES)
- P_n is the *n*-th point of a PC
- $N_i(P_n)$ gives the coordinates of the i-th nearest point to P_n
- $\mathbf{d}\left(P_a, P_b\right)$ gives the Euclidean distance of points P_a and P_b
- k_{nn} is the k-nearest neighbors and is set to 8 in this work.

Local Binary Patterns (LBP) for 2D images

$$\mathsf{LBP}_R^N(P_c) = \sum_{n=0}^{N-1} \theta(P_n - P_c) \cdot 2^n,$$

where

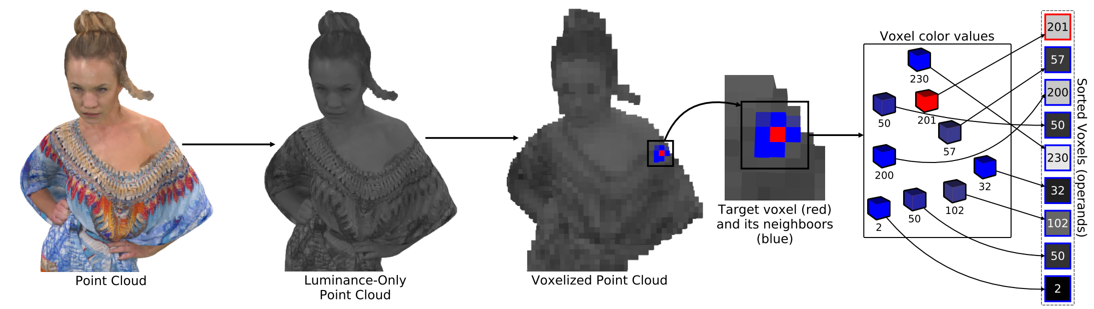
$$\theta(u) = \begin{cases} 1 & \text{if } u \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



- 2D images have pixels equally distributed in a dense 2D grid.

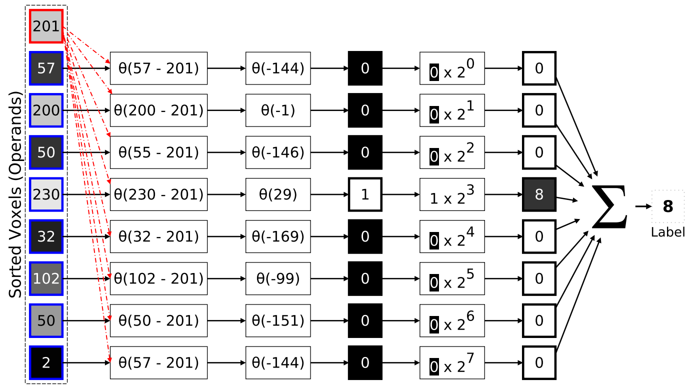
Local Binary Patterns for PC (LBP)

- The LBP is a texture descriptor proposed by Ojala *et al.* to improve the accuracy of texture recognition tasks in 2D images. This work adapted the LBP to work with PCs
- PCs have no dense and uniform neighborhood to traverse, which is a problem, so the LBP for PCs traverse the neighbors considering the distance to a target point



Local Binary Patterns for PC (LBP)

Diagram of the LBP descriptor label attribution for PCs.



Local Luminance Patterns (LLP)

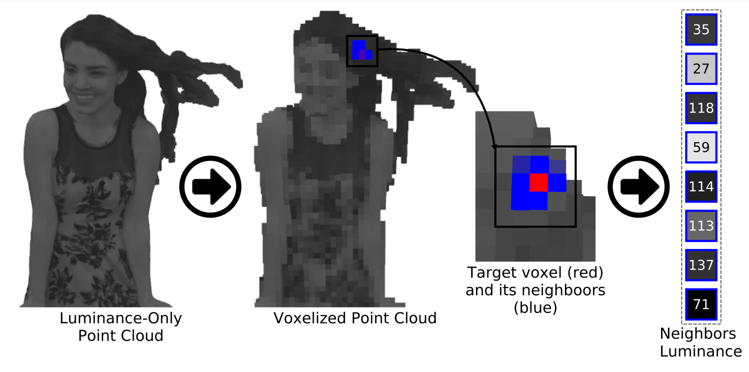
- A PC color-based texture descriptor based luminance patterns that are representative of intrinsic PC texture (RGB is converted to Y)
- Voxelization is used to improve the performance
- The texture descriptor label is calculated in an iterative way, for each neighbor:

$$(16-bit)L = \begin{cases} L \vee (1 \ll \lfloor \frac{Y[i]-15}{15} \rfloor), & \text{if } 15 \leq Y[i] < 240; \\ L \vee (1 \ll 15), & \text{if } 240 \leq Y[i] \geq 255. \end{cases}$$

$$(12-bit)L = \begin{cases} L \vee (1 \ll \lfloor \frac{Y[i]-20}{20} \rfloor), & \text{if } 20 \leq Y[i] < 240; \\ L \vee (1 \ll 11), & \text{if } 240 \leq Y[i] \geq 255. \end{cases}$$

Local Luminance Patterns (LLP)

Diagram of the LLP label computation with a set of neighbor voxels



Local Luminance Patterns (LLP)

Example of LLP label calculation

Neighbor (i)	Y[i]	Bit Set	Label (accumulated)
0	35	1	00000000 00000010
1	27	0	00000000 00000011
2	118	6	00000000 01000011
3	59	2	00000000 01000111
4	114	6	00000000 01000111
5	113	6	00000000 01000111
6	137	8	00000001 01000111
7	71	3	**00000001 01001111**

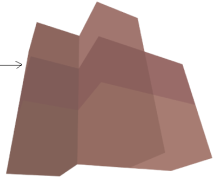
Local CIEDE2000 Patterns (LCP)

- CIELab color space - intended as a perceptually uniform space - has 3 channels: $L*$ for perceptual lightness, $a*$ is relative to green-red opponent colors, and $b*$ for the blue-yellow axis.
- CIELab color space is not really uniform - CIEDE2000 (CIELab Delta-E 2000) distance was introduced to fix CIELab perceptual non-linearities
- A numerical change corresponds to similar perceived change in color

$$(8-bit)L = \begin{cases} L \vee (1 \ll \lfloor \frac{C[i]-2.5}{2.5} \rfloor), & \text{if } 2.5 \leq C[i] < 20.0; \\ L \vee (1 \ll 7), & \text{if } C[i] \geq 20. \end{cases}$$

$$(12-bit)L = \begin{cases} L \vee (1 \ll \lfloor \frac{C[i]-1.5}{1.5} \rfloor), & \text{if } 1.5 \leq C[i] < 18.0; \\ L \vee (1 \ll 11), & \text{if } C[i] \geq 18, \end{cases}$$

Local CIEDE2000 Patterns (LCP)



	RED	GREEN	BLUE	delta-E	label bit
TARGET	130	91	88	0	/
	139	96	92	2.533617	0
	128	92	88	1.0756	/
	153	107	104	7.109899	1
	161	118	111	11.038319	3
	149	108	98	7.247803	1
	130	97	88	3.977761	0
	141	102	94	4.633231	0
	167	124	114	13.56139	4

LABEL: 00011011b
bit position: 76543210

Geometry-based Texture Descriptor

- In order to address PC geometric distortions, the geometry-based texture descriptor considers only the geometric information of each PC point (X, Y, Z)
- For this, normal vectors are calculated for all points, considering a PC local surface
- Since typical PC capture devices do not capture normal vectors (just RGB+D), the normal vectors need to be computed
- The normal vectors are normalized and set to a arbitrary point direction in order to remove duality (two normal vectors can correctly represent a surface)

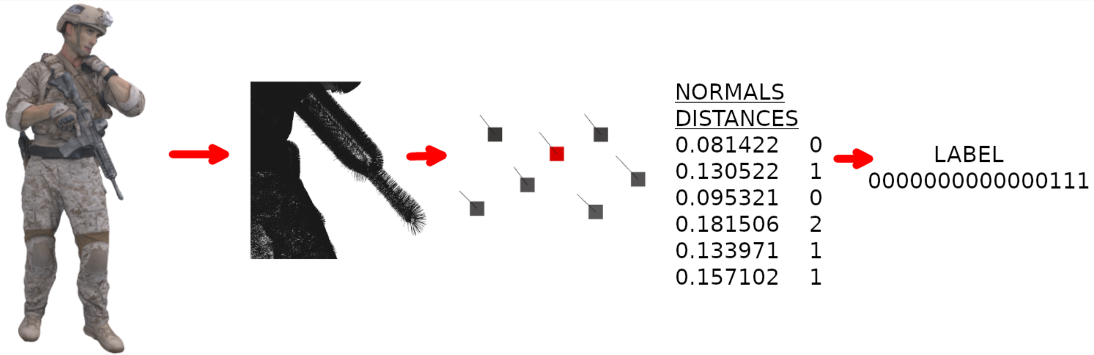
Normal vector distances is defined as:

$$G = \sqrt{\sum_{d=1}^{3} (n_{t_d} - n_{i_d})^2}$$

Where G ranges from 0 to 2.

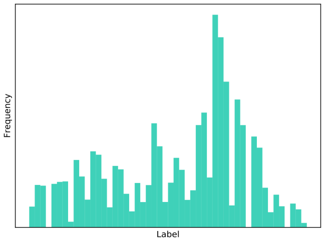
Geometry-based Texture Descriptor (GEO)

Diagram of the geometric texture label computation, with the normal vectors represented as black lines:

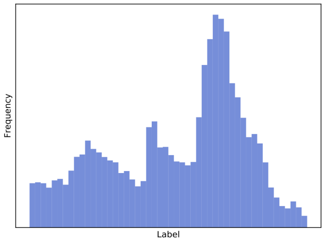


Texture Histogram Distances

- The histograms are obtained from the statistics of the labels produced by the texture descriptors, for reference and test PCs
- Different histogram distance measures were evaluated, namely: Bray-Curtis, Canberra, Chebyshev, City Block, Cosine, Euclidean, Jensen-Shannon, Wasserstein and Energy.



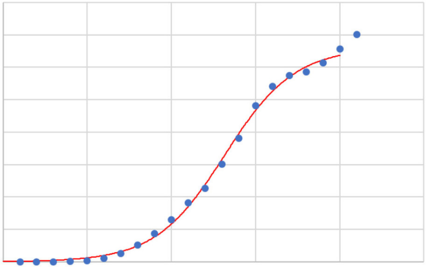
Histogram of labels $(\mathbf{h_i})$



Histogram of labels $(\mathbf{h_r})$

PC Quality Prediction Modeling

- Quality predition model based on a regression method
- The regression algorithm takes as input the distance $\mathbf{D}$ of the histograms and maps it into an objective (predicted) quality score using subjective scores as ground-truth values
- Different regression models were evaluated: Extra Trees, Gradient Boosting, Random Forest and the Logistic function



Experimental Setup: Data-Sets

We used the following data-sets, named D1 to D4 and subjective scores as follows:

- D1: Torlig 2018 [1]
- D2: Cruz 2019 [2]
- D3: Alexiou 2019 [3]
- D4: Stuart 2020 [4]

[1] A novel methodology for quality assessment of voxelized point clouds
[2] Point cloud quality evaluation: Towards a definition for test conditions
[3] A comprehensive study of the rate-distortion performance in mpeg pointcloud compression
[4] Quality evaluation of static point clouds encoded using MPEG codecs

Experimental Setup: Metrics

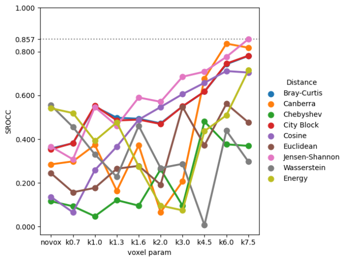
- We opted to use the MPEG-released PC metrics plus PointSSIM (Alexiou et al.) and PCQM (Meynet et al.) as benchmark. The MPEG metrics (Tian, Mekuria, Alexiou et al.) are the following:

po2point$_{MSE}$	PSNR-po2point$_{MSE}$	po2point$_{Haus}$	PSNR-po2point$_{Haus}$
Color-YCbCr$_{MSE}$	PSNR-Color-YCbCr$_{MSE}$	Color-YCbCr$_{Haus}$	PSNR-Color-YCbCr$_{Haus}$
po2plane$_{MSE}$	PSNR-po2plane$_{MSE}$	po2plane$_{Hausdorff}$	PSNR-po2plane$_{Haus}$

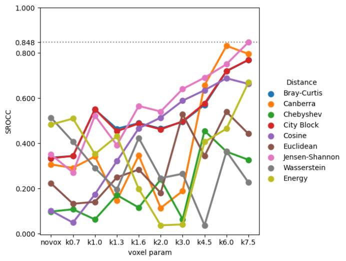
Experimental Setup

- 4 Data-sets: D1, D2, D3 and D4;
- Performance metrics: PCC, SROCC and RMSE;
- 'k' voxelization parameter: novox, 0.7, 1.0, 1.3, 1.6, 2.0, 3.0, 4.5, 6.0, 7.5;
- Histogram distances: Bray-Curtis, Canberra, Chebyshev, City Block, Cosine, Euclidean, Jensen-Shannon, Wasserstein and Energy;
- Logistic function as regressor.

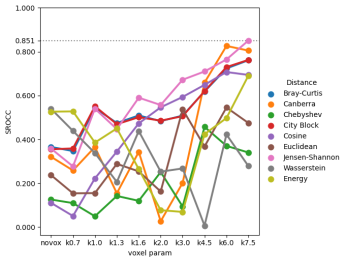
Regressors Evaluation - D1, LCP 8-bit, 12 N, SROCC



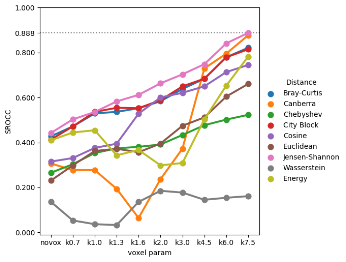
(a) Extra Trees



(b) Gradient Boosting

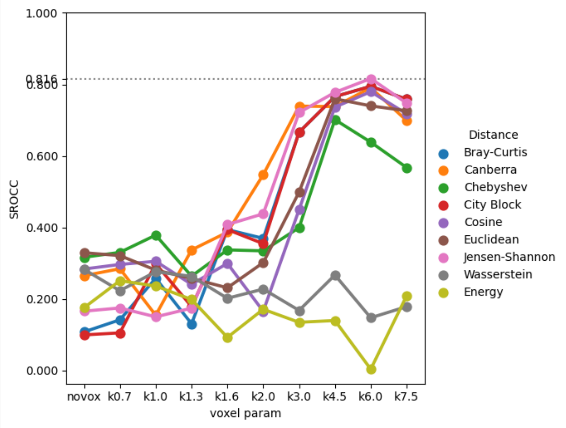


(c) Random Forest

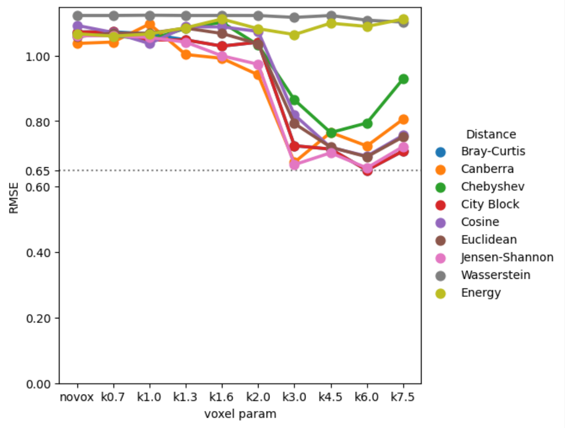


(d) Logistic

Simulation Results - Dataset D1 - LBP, 8 neighbors

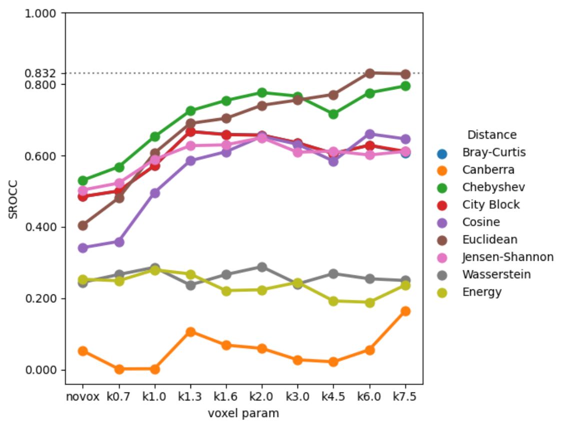


SROCC

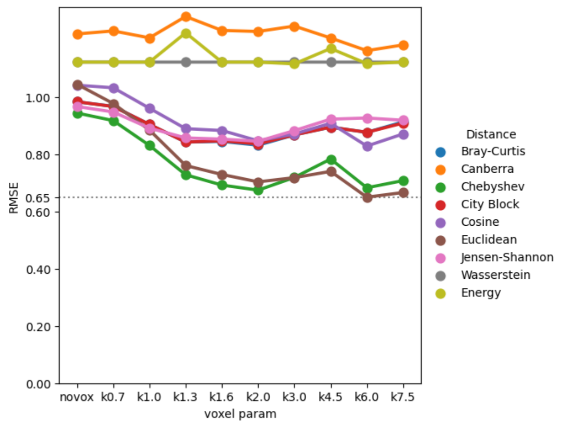


RMSE

Simulation Results - Dataset D1 - LLP 12-bit, 8 neighbors

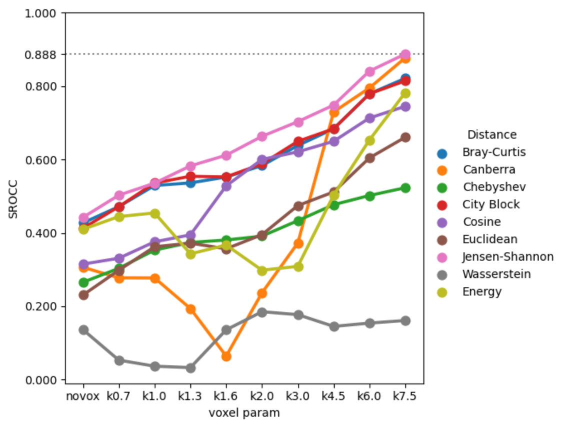


SROCC

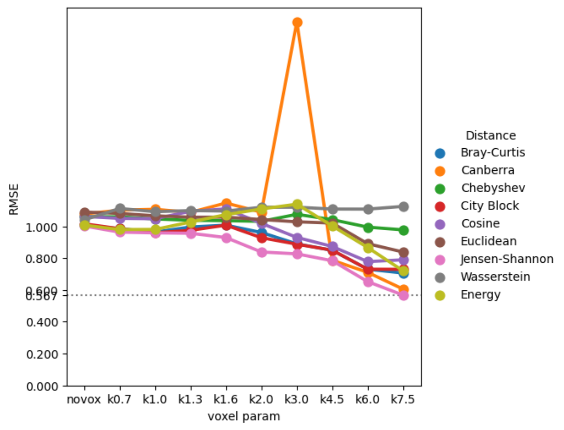


RMSE

Simulation Results - Dataset D1 - LCP 8-bit, 12 neighbors

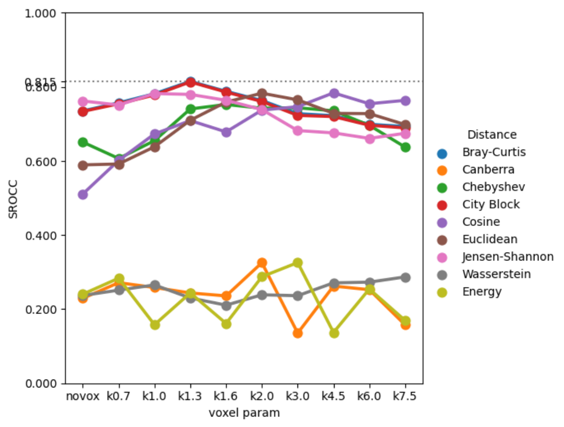


SROCC

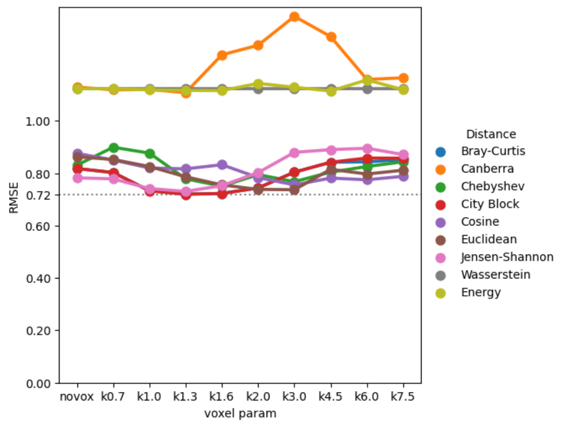


RMSE

Simulation Results - Dataset D1 - GEO, 6 neighbors

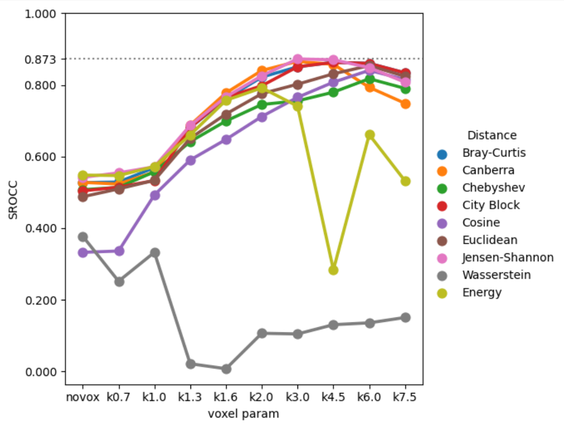


SROCC

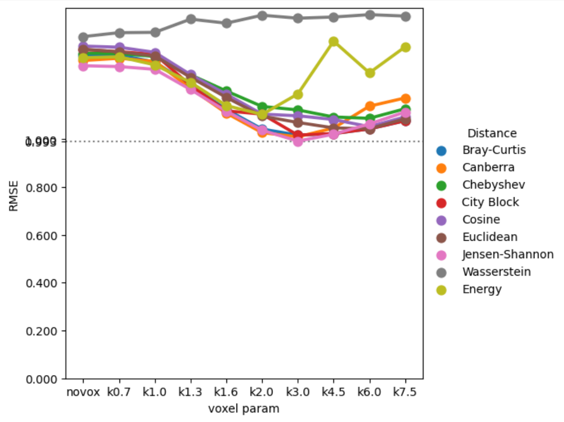


RMSE

Simulation Results - Dataset D2 - LBP 8 neighbors

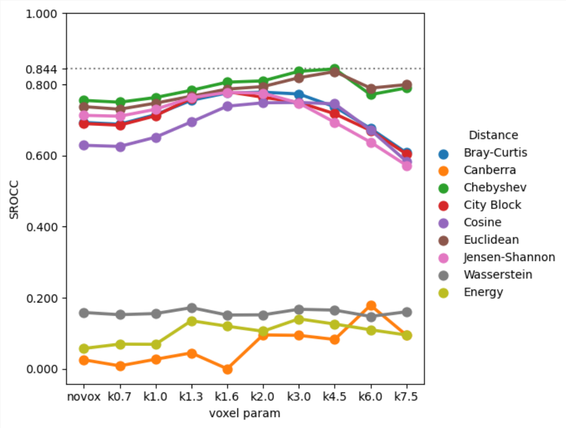


SROCC

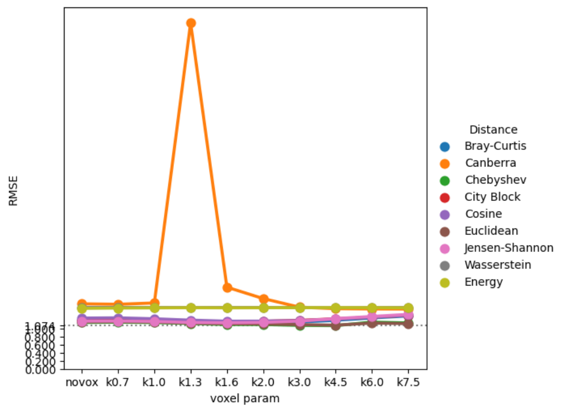


RMSE

Simulation Results - Dataset D2 - LLP 12-bit, 8 neighbors

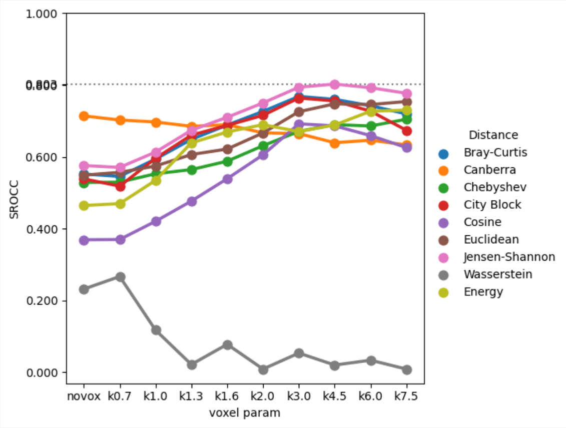


SROCC

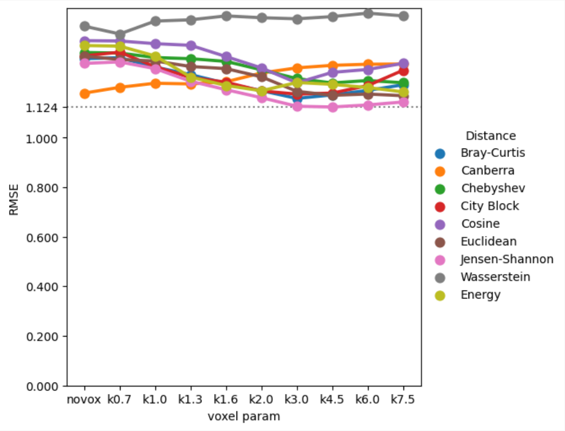


RMSE

Simulation Results - Dataset D2 - LCP 8-bit, 12 neighbors

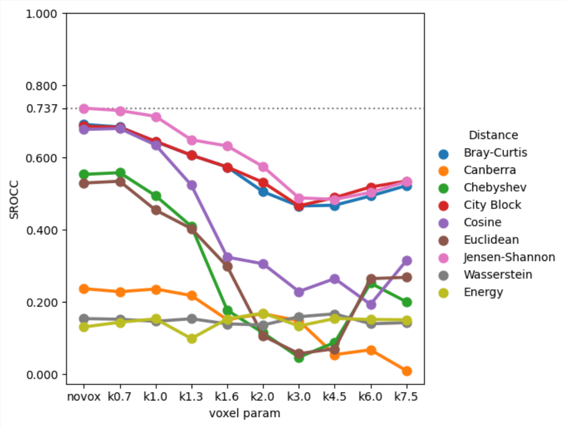


SROCC

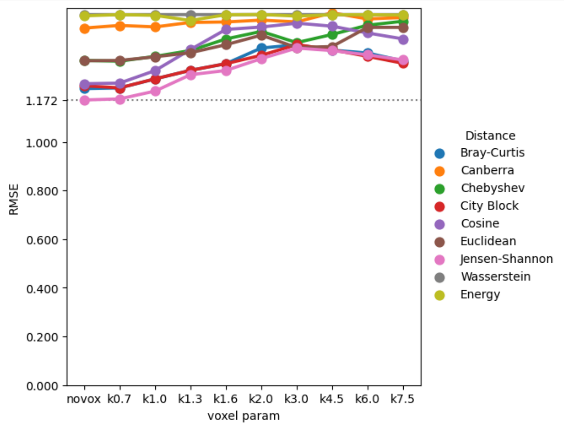


RMSE

Simulation Results - Dataset D2 - GEO, 6 neighbors

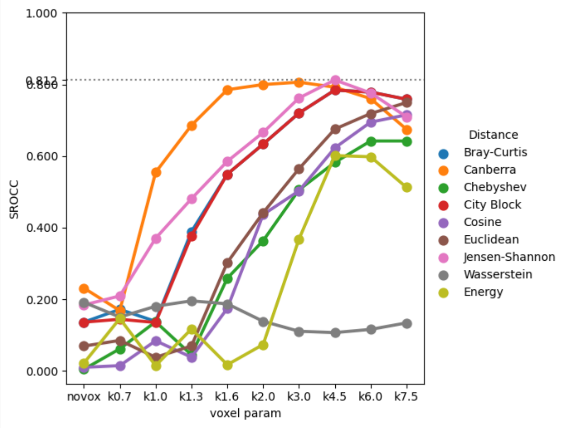


SROCC

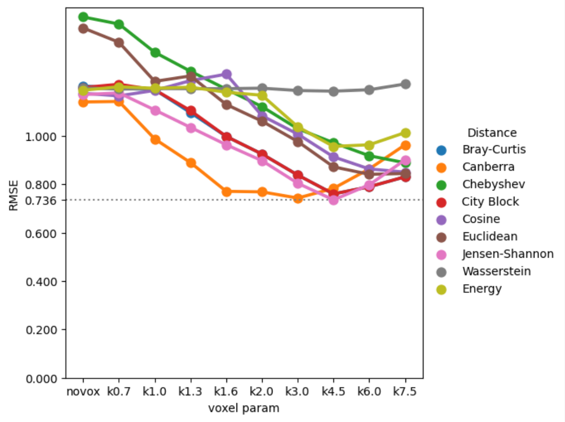


RMSE

Simulation Results - Dataset D3 - LBP 8 neighbors

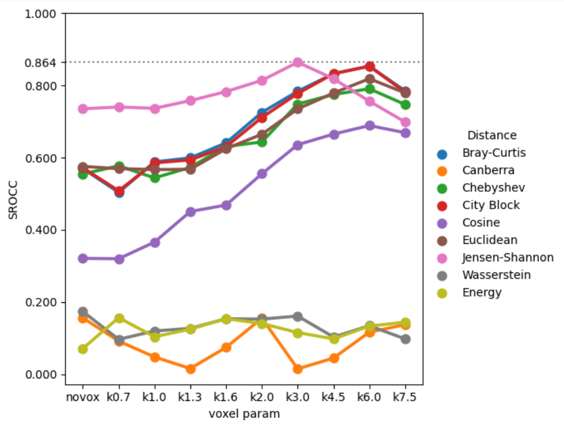


SROCC

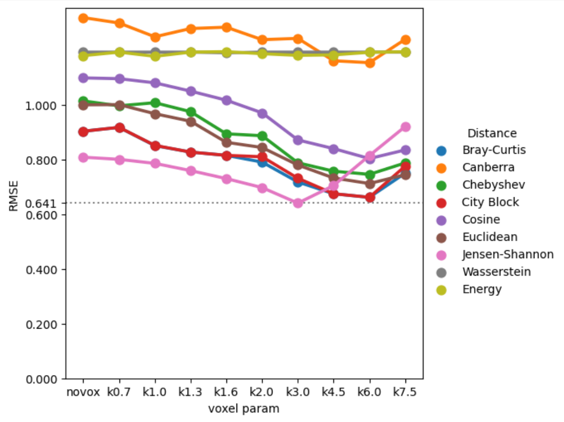


RMSE

Simulation Results - Dataset D3 - LLP 12-bit, 8 neighbors

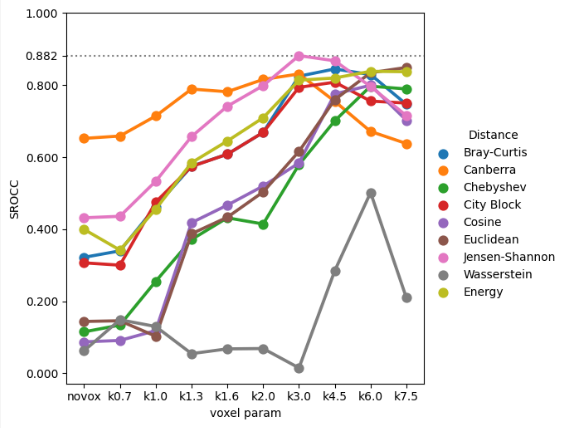


SROCC

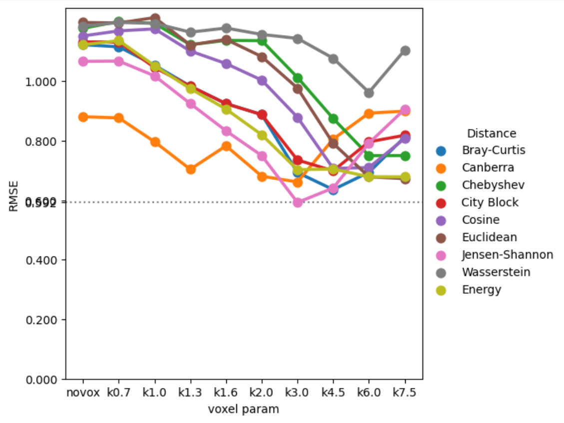


RMSE

Simulation Results - Dataset D3 - LCP 8-bit, 12 neighbors

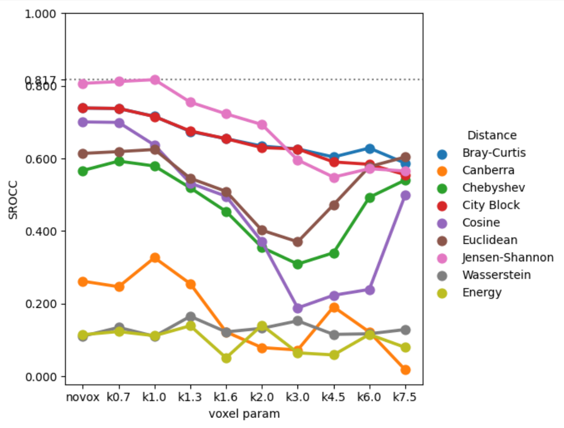


SROCC

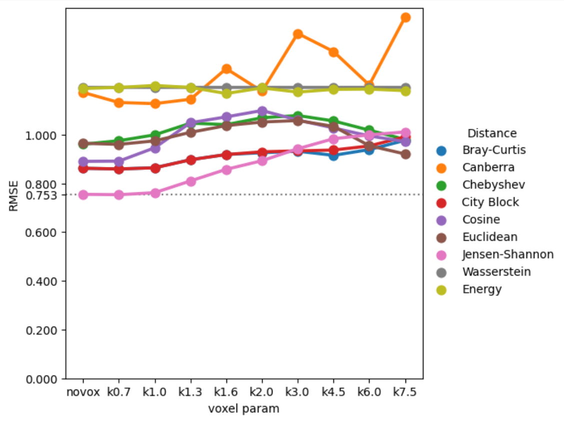


RMSE

Simulation Results - Dataset D3 - GEO, 6 neighbors

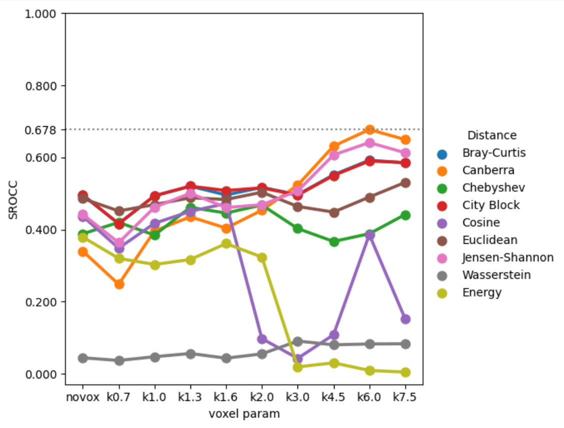


SROCC

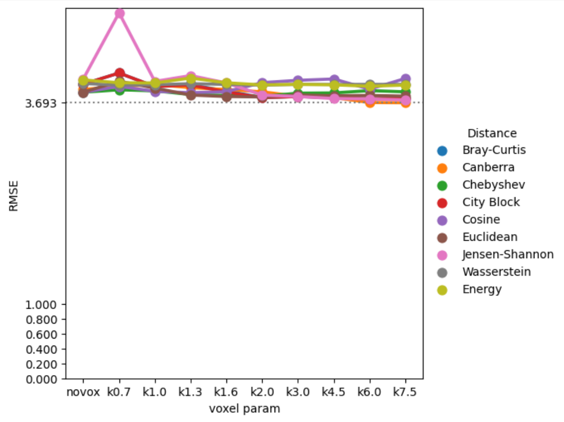


RMSE

Simulation Results - Dataset D4 - LBP 8 neighbors

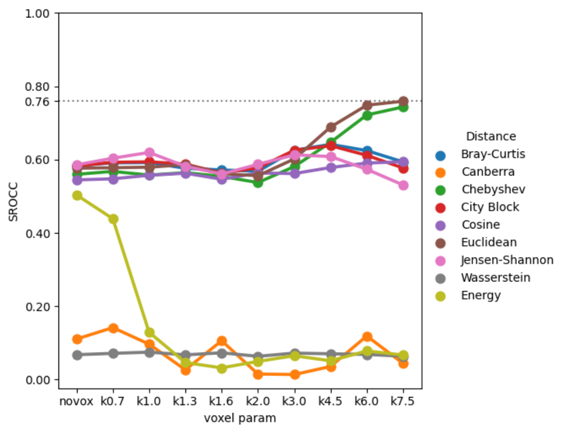


SROCC

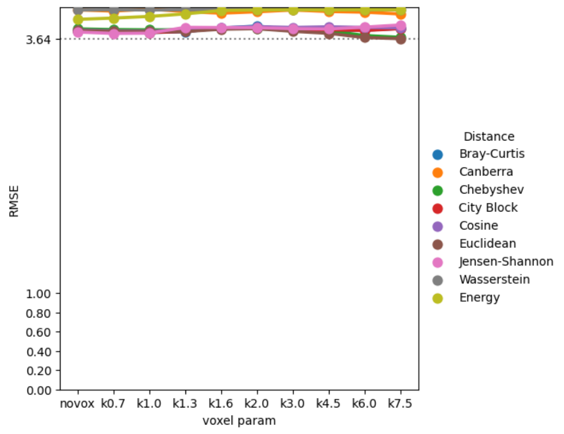


RMSE

Simulation Results - Dataset D4 - LLP 12-bit, 8 neighbors

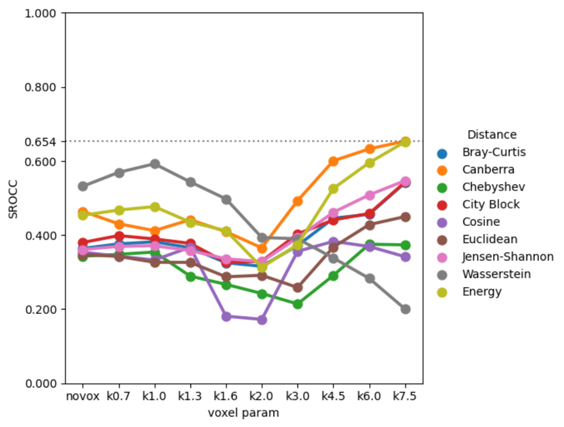


SROCC

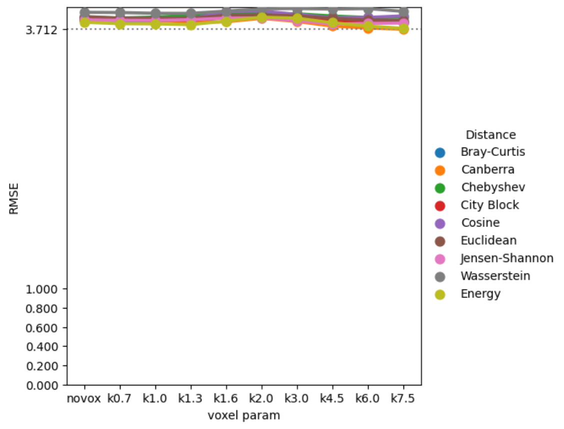


RMSE

Simulation Results - Dataset D4 - LCP 8-bit, 12 neighbors

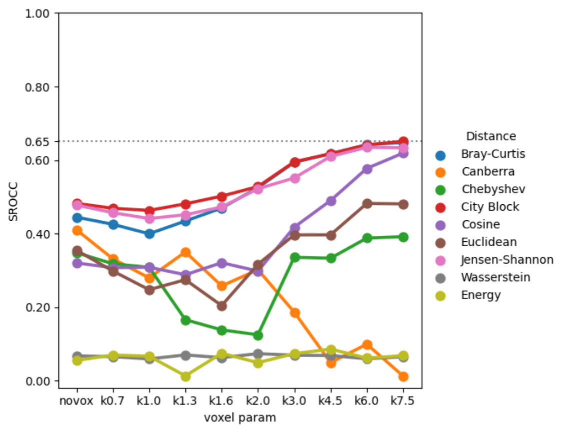


SROCC

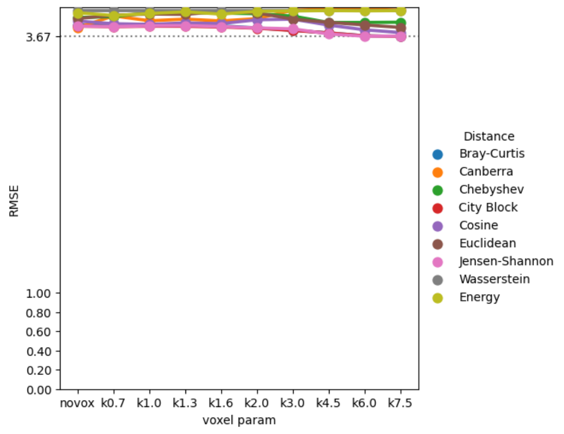


RMSE

Simulation Results - Dataset D4 - GEO, 6 neighbors



SROCC



RMSE

Conclusions: LBP

- Highly influenced by the voxelization;
- k voxelization parameter between 2 and 6 are best;
- Canberra and Jensen-Shannon distances are best performing distances;
- Neighborhood size does not influence much for D1, D2 and D3, for D4 6 neighbors perform a bit better;
- PCC performance peaks at 0.877, 0.878 and 0.907 in D1, D2 and D3 respectively, while in D4, LBP PCC peaks at 0.724.

Conclusions: LLP

- 16-bit and 12-bit tested;
- The Euclidean best for the 16-bits variant. Euclidean, Jensen-Shannon and Chebyshev alternating best performance for 12-bits version;
- ’k’ best between 4.5 and 7.5 for the 16-bits version;
- no clear best ’k’ among all datasets for the 12-bits version;
- 16-bits peaks of PCC: 0.880, 0.839, 0.870 and 0.728; the 12 bits version peaked at 0.834, 0.820, 0.884 and 0.762.

Conclusions: LCP

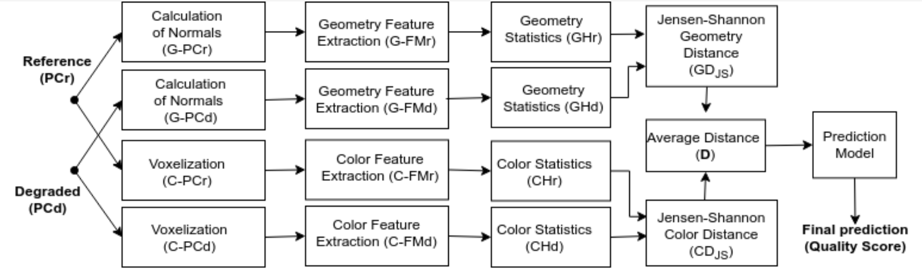
- 12 bits and a 8 bits variations;
- Voxelization improves performance, but not homogeneously across data-sets;
- Jensen-Shannon is best for D1, D2 and D3, but not by wide margin; for D4 Energy is best;
- LCP 8-bits is slightly better than the 12 bits;
- 12 neighbors slightly better performance to both variants;
- LCP 12-bits has PCC peak values of 0.802, 0.780, 0.881 and 0.660, while the LCP 8 bits PCC peaked at 0.880, 0.775, 0.912 and 0.660.

Conclusions: GEO

- 16-bit;
- Voxelization degrades the performance in most scenarios; D4 is exception;
- Voxelization is not suitable for the geometry-based texture descriptor;
- Best histogram distance is the Jensen-Shannon;
- PCC peaked at 0.777, 0.731, 0.814 and 0.713;
- Neighborhood size does not influence so much.

Use of two or more descriptors for quality assessment

- Color and geometry texture descriptors alone, cannot capture all types of distortions;
- This proposed framework allows the use of two or more descriptors to predict PC quality;
- The histogram distances can also be combined and then used as input to the regressor.



Selected descriptors for joint use

- Geometry-based descriptor (GEO) fixed with 6 neighbors, no voxelization, Jensen-Shannon distance;
- LBP+GEO, LBP with 8 neighbors, $k = 1.6$, Jensen-Shannon;
- LLP+GEO, LLP 12-bits, with 8 neighbors, $k = 2.0$, Euclidean;
- LCP+GEO, LCP 8-bits, with 12 neighbors, $k = 6.0$, Jensen-Shannon.

Results and Comparisons

Metrics	Data Sets														
	D1			D2			D3			D4			AVG		
	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE
po2point_MSE	0.270	0.250	1.122	0.808	0.835	1.095	*0.941*	0.920	*0.534*	0.418	0.350	3.857	0.609	0.589	1.652
PSNR-po2point_MSE	0.518	0.484	0.953	0.494	0.430	1.352	0.538	0.549	1.025	0.470	0.376	3.832	0.505	0.460	1.791
po2point_Haus	0.270	0.215	1.122	0.627	0.421	1.282	0.496	0.446	1.024	0.261	0.224	3.900	0.414	0.327	1.832
PSNR-po2point_Haus	0.512	0.469	0.968	0.454	0.396	1.379	0.549	0.527	1.008	0.481	0.455	3.833	0.500	0.462	1.797
Color-YCbCr_MSE	0.383	0.367	1.039	0.553	0.571	1.333	0.755	0.682	0.921	0.500	0.512	3.822	0.548	0.533	1.779
PSNR-Color-YCbCr_MSE	0.368	0.337	1.097	0.536	0.565	1.351	0.793	0.801	0.797	0.504	0.503	*3.805*	0.550	0.552	1.763
Color-YCbCr_Haus	0.147	0.172	1.131	0.413	0.375	1.380	0.377	0.306	1.122	0.191	0.095	3.955	0.282	0.237	1.897
PSNR-Color-YCbCr_Haus	0.386	0.320	1.059	0.435	0.391	1.417	0.445	0.449	1.100	0.344	0.270	3.875	0.403	0.358	1.863
po2plane_MSE	0.270	0.275	1.122	*0.845*	0.858	**1.031**	**0.958**	*0.945*	**0.492**	0.432	0.370	3.859	0.626	0.612	*1.626*
PSNR-po2plane_MSE	0.484	0.421	0.984	0.499	0.495	1.361	0.542	0.579	1.021	0.380	0.390	3.893	0.476	0.471	1.815
po2plane_Hausdorff	0.270	0.247	1.122	0.604	0.427	1.267	0.586	0.418	0.981	0.223	0.188	3.990	0.421	0.320	1.840
PSNR-po2plane_Haus	0.440	0.408	1.016	0.428	0.367	1.394	0.497	0.463	1.034	0.464	0.451	3.836	0.457	0.422	1.820
PCQM	0.797	**0.898**	2.656	0.607	*0.915*	2.899	0.738	**0.970**	3.123	0.271	0.708	5.786	0.603	**0.873**	3.616
PointSSIM-Color	0.842	0.823	2.234	**0.910**	**0.918**	2.436	0.869	0.865	2.697	*0.676*	*0.682*	5.354	*0.824*	0.822	3.180
PointSSIM-Geometry	0.804	0.820	2.102	0.784	0.834	2.321	0.849	0.905	2.534	0.527	0.560	5.323	0.741	0.780	3.070
LCP + GEO	**0.876**	*0.896*	**0.572**	0.819	0.839	1.068	0.936	0.932	0.544	**0.730**	**0.714**	**3.663**	**0.840**	*0.845*	**1.462**
LBP + GEO	*0.845*	0.837	*0.620*	*0.845*	0.850	*1.037*	0.863	0.869	0.672	0.579	0.543	3.764	0.783	0.775	1.523
LLP + GEO	0.790	0.795	0.702	0.812	0.822	1.077	0.873	0.877	0.651	0.672	0.660	3.705	0.787	0.789	1.534

Conclusions

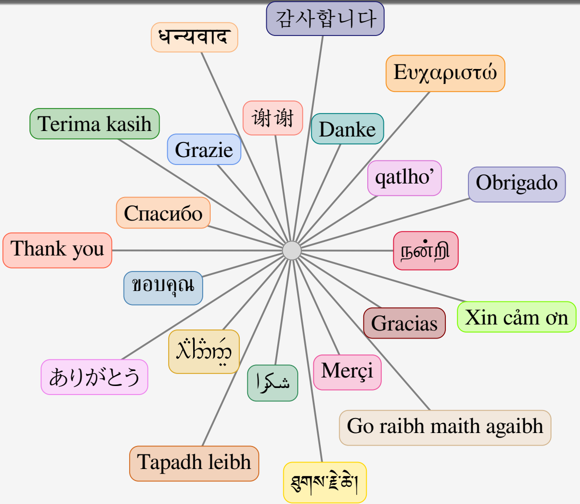
- The voxelization process improves the performance of the color-based texture descriptors;
- Data-sets ground-truth were obtained using different types of rendering techniques, implying the voxelization affects differently the performance of the descriptors among the data-sets;
- Descriptors are scale and rotational invariant;
- The Jensen-Shannon distance presented the best overall correlation results, but not by wide margin;

Conclusions

- Color-based texture descriptors perform better than geometry-based texture descriptors;
- LCP needs less bits than LLP for the same performance accuracy;
- Joint use of LCP and the geometry-based strategies outperformed single texture descriptors;
- MPEG metrics work better when the content is degraded with the MPEG PC encoders, with test conditions that degrade geometry and color with similar intensities;
- Overall, the proposed PC quality assessment metric framework proposed by this work outperforms all other state of the art PCQA metrics.

Future Work

- Adaptive voxel size selection;
- Optimization of current texture descriptor through a detailed statistical analysis of the available PC content;
- No-Reference PCQA method that uses the histogram distances as input to a data-driven quality assessment model.



Questions?

rafael@riseup.net,
http://www.gpds.ene.unb.br/rafael/
http://www.ene.unb.br/mylene/databases.html