UC Riverside
UC Riverside Previously Published Works

Title

MPED: Quantifying Point Cloud Distortion based on Multiscale Potential Energy Discrepancy

Permalink

https://escholarship.org/uc/item/53x1q1v9

Authors

Yang, Qi
Zhang, Yujie
Chen, Siheng
et al.

Publication Date

2021-03-03

Peer reviewed

Point Cloud Distortion Quantification based on Potential Energy for Human and Machine Perception

Qi Yang, Siheng Chen, Yiling Xu, *Member, IEEE*, Jun Sun, M. Salman Asif, *Senior Member, IEEE*, and Zhan Ma, *Senior Member, IEEE*

Abstract—Distortion quantification of point clouds plays a stealth, yet vital role in a wide range of human and machine perception tasks. For human perception tasks, a distortion quantification can substitute subjective experiments to guide 3D visualization; while for machine perception tasks, a distortion quantification can work as a loss function to guide the training of deep neural networks for unsupervised learning tasks. To handle a variety of demands in many applications, a distortion quantification needs to be distortion discriminable, differentiable, and have a low computational complexity. Currently, however, there is a lack of a general distortion quantification that can satisfy all three conditions. To fill this gap, this work proposes multiscale potential energy discrepancy (MPED), a distortion quantification to measure point cloud geometry and color difference. The proposed MPED is able to capture both geometrical and color impairments by quantifying the total distortion between reference and distorted samples. By evaluating at various neighborhood sizes, the proposed MPED achieves global-local tradeoffs, capturing distortion in a multiscale fashion. We further theoretically show that classical Chamfer distance is a special case of our MPED. Extensive experimental studies validate MPED's superiority for both human and machine perception tasks. For human perception tasks, the proposed MPED works as subjective score predictor on two large subjective database. For machine perception tasks, the proposed MPED is plugged in as the loss function to enable the training of deep neural networks for three tasks, including point cloud reconstruction, shape completion and upsampling. The experimental results reveal that the proposed MPED produces better results than the point-wise Chamfer distance and Earth Mover's distance under the same network architecture. Our code is avaliable at https://github.com/Qi-Yangsjtu/MPED.

Index Terms—Distortion quantification, objective quality assessment, human perception, machine perception, potential energy, point cloud.

I. INTRODUCTION

Three-dimensional point clouds have become an important geometric data structure, which precisely present the external surfaces of objects or scenes in the 3D spatial space. They can be widely used in augmented reality [1], autonomous driving [2], industrial robotics [3], art documentation [4] and many others. To provide immersive user experience and reliable performance, a high-quality point cloud is in high demand. For example, point clouds used in augmented reality provide rich color and detailed geometric structures to better serve human perception; and point clouds used in autonomous driving present informative depth information for accurate 3D object detection (e.g., pedestrians, trees and buildings).

To evaluate the distortion of a point cloud based on a reference point cloud, we need a reliable, accurate and objective distortion quantification that can be generally applied for a wide range of scenarios. To reflect human perception, a distortion quantification should be sensitive to both geometric and color distortion. For example, the point cloud "LongDress" [5] provided by the moving picture experts group (MPEG) involves the RGB color as additional attributions. After compressed by video-based point cloud compression (V-PCC), the number of points might increase, and points' colors might be deteriorated due to the quantization [6].

Besides being used for quality assessment in human perception tasks, an objective distortion quantification could also serve for machine perception, which is a loss function to guide the training of deep neural networks. Specifically, unsupervised tasks of machine perception generally aim to extract informative features from point clouds without using human supervision [7]. Some typical tasks include reconstruction [8], [9], completion [10], [11], and upsampling [12]. To train those networks, a distortion quantification usually works as the loss function to supervise the output of the network towards the ground-truth. Therefore, a distortion quantification should be differentiable to enable backpropagation and has low computational complexity. Since machine perception tasks need to train networks with a huge amount of data, a distortion quantification needs to have a low resource overhead to facilitate training.

Based on the above analysis, an ideal point cloud distortion quantification need to satisfy three basic conditions, including *distortion discrimination*, *differentiability* and *low complexity*. Currently, however, there is a lack of a generic distortion quantification for point clouds that can satisfy the above three conditions at the same time. For example, Chamfer distance (CD) [8], Earth mover's distance (EMD) [13], and $PSNR_{YUV}$ [14] use point-wise distance to quantify distortion, which presents unstable performance when faced with multiple
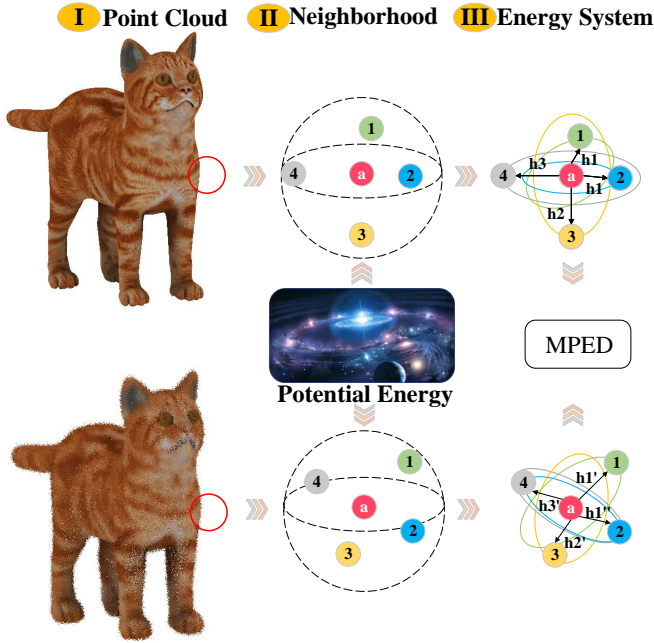
Fig. 1: MPED compares the total point potential energies of multiple neighboring point clouds in the source and the target.

distortions. Besides, as the most widely used loss function in machine perception tasks, CD has a shortcoming named "Chamfer's blindness" [8] or "point-collapse" [15] which can lead to unexpected results (Fig. 2 (b)). Point cloud quality metric (PCQM) [16] and GraphSIM [5] present obvious biases in distortion detection [17] (e.g., both PCQM and GraphSIM present excellent performance on local offset, while poor performance on Gamma noise), and surface fitting and graph construction used in these two methods are time-consuming. This work aims is to fill this gap and design a universal distortion quantification that can present robust and reliable performance on point clouds distortion measurement.

Proposed distortion quantification. In this paper, we present a new 3D point cloud distortion quantification; called multiscale potential energy discrepancy (MPED). The proposed MPED is inspired by the potential energy used in classical physics. Consider that a lossless 3D point cloud is a stable spatial system, the distortion can be regarded as system disturbance caused by the external force [18]. Assuming each 3D point has potential energy corresponding to the zero potential energy plane set manually, the system disturbance can be quantified by the total energy variation. Moreover, the standard the perceived distortion quantification for evaluating 2D images, MSSIM [19], also inspires us that multiscale features present more robust performance in terms of 3D distortion detection. Specifically, the multiscale features provide a more comprehensive measurement of the overall impact of distortion [20]. Therefore, we fuse the thoughts about distortion-energy inspiration and multiscale representation to design our MPED.

Based on the above intuitions, we design MPED in the following four steps. Specifically, Fig. 1 illustrates the scheme of the proposed MPED, i) we first divide a point cloud into

multiple spatial neighborhoods and set the center of each neighborhood at zero potential energy plane; ii) we assign the potential energy to each point in the neighborhood using point mass and spatial fields (e.g., gravitational acceleration); iii) we compare the total point potential energies of neighborhoods in two point clouds, the geometric and attributive differences between two point clouds are converted to the total energy variation; and iv) finally, MPED captures the distortion of a point cloud at multiple scales by selecting various neighborhood sizes.

We analyze the properties of MPED via several toy examples. The proposed MPED is differentiable, the complexity of the proposed MPED is $O(N^2)$, which is the same as CD and much lower than EMD ($O(N^3)$), where N is the number of points in a point cloud, and the proposed MPED is sensitive to multiple types of distortion, including isometrical distortion. We also prove that CD is a special case of the proposed MPED.

We validate the performance of MPED based on two main applications: human and machine perception tasks.

In terms of human perception tasks, the proposed MPED presents better performance on perception prediction. Specifically, MPED provides better Pearson linear correlation (PLCC), Spearman rank-order correlation (SROCC) and root mean squared error (RMSE) than point-to-point (p2point), p2plane, GraphSIM and PCQM on two large databases (e.g., SJTU-PCQA [21] and LSPCQA [17]).

In terms of machine perception tasks, the proposed MPED achieves a better tradeoff between effectiveness and efficiency. Compared with CD, MPED solves the "point-collapse" issue; compared with EMD, MPED reduces the computational cost from $O(N^3)$ to $O(N^2)$. Specifically, MPED provides better performance compared to CD and EMD on three unsupervised machine perception tasks (point cloud reconstruction, shape completion, and upsampling);

The main contributions of this paper are summarized as:

• We present a new universal distortion quantification, MPED, for point clouds. Inspired by classic mechanics, we use potential energy variation to quantify point cloud distortion.

• MPED satisfies the three desired conditions for a universal objective distortion quantification; that is, MPED is differentiable, low-complexity, and distortion discriminative.

• MPED shows reliable performance for both human and machine perception tasks.

The rest of the paper is organized as follows. Section II presents the related work of point cloud quality assessment. Section III introduces point cloud representation and distortion types. Section IV demonstrates the proposed potential-energy-based feature used for point cloud distortion quantification. Section V presents the generic implementation of the proposed MPED. Section VI presents the experimental results of MPED on both human and machine perception tasks. Section VII concludes the paper.

II. RELATED WORK

Point cloud distortion quantifications are first used in human vision tasks. To evaluate the distortion introduced by lossy compression, MPEG first uses p2point as the distortion quantification. Specifically, p2point shares the same calculation

method as the CD. However, according to a serial experiments [14], [22]–[27], p2point presents unstable results in terms of multiple types of distortion. Studies [19], [28] in image distortion quantification have proven that the human visual system is more sensitive to structural features, rather than point-wise features. Therefore, recently proposed distortion quantifications for human perception tasks, such as p2plane [29], PCQM [16], and GraphSIM [5] use the distortion of structural features to replace the distortion of point-wise features. Experimental results demonstrate that these methods provide better performance than the original p2point. For example, experiments in [5] show that the utilization of structure features (e.g., gradients) presents better performances when predicting human perception in terms of compression-related distortion (e.g., V-PCC and geometry-based point cloud compression (G-PCC)). However, PCQM and GraphSIM need to convert scattered points to inerratic representations, such as surface and spatial graph, which are time-consuming.

Besides, point cloud distortion quantifications are commonly used as loss functions to assist in unsupervised machine perception tasks (such as point cloud reconstruction, shape completion, and upsampling). There are two mainly used quality quantifications in current machine perception tasks: CD and EMD. EMD is the solution of a transportation problem that attempts to transform one set to the other.

The advantage of CD is low resource overhead. Since one point can be used multiple times during point matching, this causes a significant drawback for CD, known as "point-collapse" [15] or "Chamfer's blindness" [8]. Specifically, the generated samples that use CD as a loss function suffer from obvious structure deformations. Fig. 2 (b) shows an example in which the points of a chair arm in the reconstruction are sparser than those in the ground truth, while the points of the seat in the reconstruction are denser than those in the ground truth. Points are over-populated around collapse centers [15]. The reason is that a reference point may be used multiple times in the point-nearest-neighbor search, and the final results fall into a local optima .

EMD is a method that computes the distance between two distributions as the minimum mass that needs to transfer to match two distributions. Compared to CD, the point matching in EMD is stricter as any point that has been used in the point matching cannot be used to match with other points anymore. It means that the pairing of two points considers their spatial distance as well as the global transfer costs. Therefore, EMD is more reliable than CD in machine perception tasks. Comparison of reconstruction results with CD and EMD as loss functions in [8] reveal that CD can only provide partially good reconstructed results while EMD can provide better global reconstruction. Despite the better performance, many methods avoid using EMD because of its computational complexity [9], [15], [30]. The computational complexity of EMD is $O(N^3)$ [13] and that of CD is $O(N^2)$ for N points. The method in [10] provides a lightweight version of EMD with complexity as $O(N^2k)$, where k is the number of iteration to search for the best-matched point pair. By sacrificing a certain amount of computational accuracy, the computation time can be reduced. The selection of k, however, can be different
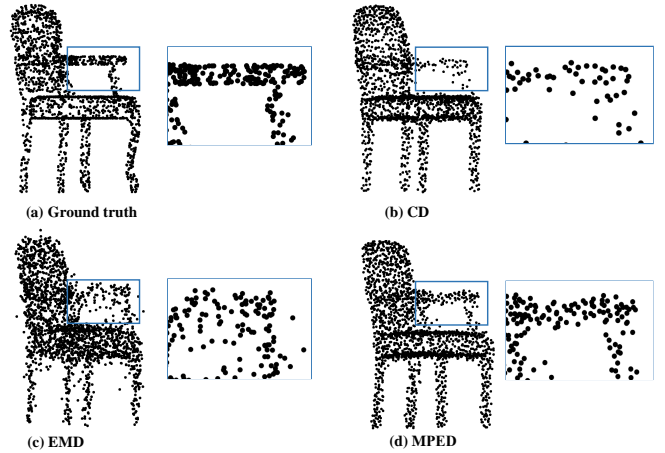

Fig. 2: Point cloud reconstruction with LatentNet [8]. (a): baseline; (b) CD as loss function; (c) EMD as loss function; (d) MPED as loss function. MPED provides the best reconstruction.

for different scales of point clouds, which is inconvenient in practical application.

III. PROBLEM FORMULATION

In this section, we introduce some basic properties of 3D point clouds and distortion quantification criteria.

A. 3D Point Cloud

Let $\mathcal{P}$ be a 3D point cloud with N points: $\mathcal{P} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, where each $\mathbf{x}_i \in \mathbb{R}^6$ is a vector with 3D coordinates and three-channel color attributes. We only consider color attributes of point cloud in this paper; therefore, $\mathbf{x}_i = [x, y, z, R, G, B] \equiv [\mathbf{x}_i^O, \mathbf{x}_i^I]$, where $\mathbf{x}_i^O = [x, y, z]$ and $\mathbf{x}_i^I = [R, G, B]$. The superscript "O" stands for geometric *occupancy*, and "I" stands for color *intensity*.

B. Point Cloud Distortion

Different from image/video distortion, point cloud distortion can be divided into three categories, including i) geometry-lossy and attribute-lossless, ii) geometry-lossless and attribute-lossy, iii) and geometry-attribute superimposed distortion. We now introduce each of the three distortions.

Geometry-lossy and attribute-lossless distortion. In this case, the number of points and points' coordinates are changed, while points' attributes keep unchanged. Let $\mathbf{n}^O \in \mathbb{R}^{N \times 3}$ be the additive geometrical noise and $\mathcal{D}(\cdot)$ be a generic measurement operator that produces a partial or enhance 3D point cloud. The geometry-lossy and attribute-lossless (noted as Φ_{D1}) distortion is

$$\Phi_{\mathrm{D1}}(\mathcal{P}) = \mathcal{D}([\mathrm{X}^O + \mathbf{n}^O, \mathrm{X}^I]) \in \mathbb{R}^{M \times 6},$$

where M is the number of point for the distorted sample. Generally, the value of M depends on the type of $\mathcal{D}(\cdot)$. For example, when $\mathcal{D}(\cdot)$ represents downsampling, $M < N$; when $\mathcal{D}(\cdot)$ represent point cloud reconstruction (e.g., LatentNet [8]), $M = N$; when $\mathcal{D}(\cdot)$ represents upsampling, $M > N$.

Geometry-lossless and attribute-lossy distortion. In this case, the points' attributes are changed, while the number of points and points' coordinates keep unchanged. Let $\mathbf{n}^I \in \mathbb{R}^{N\times 3}$ be the additive color noise. The geometry-lossless and attribute-lossy (noted as Φ_{D2}) distortion is

$$\Phi_{\mathrm{D2}}(\mathcal{P}) = [\mathrm{X}^O, \mathrm{X}^I + \mathbf{n}^I] \in \mathbb{R}^{M\times 6}.$$

The typical geometry-loss and attribute-lossless distortion includes color noise [21] and chromatic aberration [17].

Geometry-attribute superimposed distortion. In this case, the number of points, points' coordinates, and points' attributes are all changed. Geometry-attribute superimposed distortion (Noted as Φ_{D3}) is the superposition of above two distortions. Mathematically,

$$\Phi_{\mathrm{D3}}(\mathcal{P}) = \mathcal{D}([\mathrm{X}^O + \mathbf{n}^O, \mathrm{X}^I + \mathbf{n}^I]) \in \mathbb{R}^{M\times 6}.$$

The typical geometry-attribute superimposed distortion includes video-based point cloud compression (V-PCC) and geometry-based point cloud compression (G-PCC).

IV. POINT POTENTIAL ENERGY

Before presenting our objective distortion quantification between two 3D point clouds, we first introduce a novel feature to quantify the spatial distribution of 3D points. This feature is inspired by the gravitational potential energy in physics and lays a foundation for the proposed distortion quantification. Let $\mathcal{P}_{\mathbf{x}_0} = \{\mathbf{x}_i \in \mathbb{R}^6\}_{i=1}^K$ be a set of K 3D points with $\mathbf{x}_0$ being the origin. We aim to propose a function $f(\cdot)$ so that $E_{\mathcal{P}_{\mathbf{x}_0}} = f(\mathcal{P}_{\mathbf{x}_0})$ reflects the spatial distribution of $\mathcal{P}_{\mathbf{x}_0}$.

In classic mechanics, the potential energy is the energy that is stored in an object due to its position relative to some zero position, which is a quantitative measure of the object's physical state. For example, the gravitational potential energy relative to the Earth's surface is defined as $E = mgh$, where m is the mass, h is the Euclidean distance to the origin $\mathbf{x}_0$ and g reflects the gravitational field. We consider each 3D point as an object with a certain mass and make a mathematical analogy to define the potential energy of a point $\mathbf{x}_i \in \mathbb{R}^6$ as

$$E_{\mathbf{x}_i} = m_{\mathbf{x}_i} g_{\mathbf{x}_i} h_{\mathbf{x}_i}, \tag{1}$$

where the zero potential plane is at $\mathbf{x}_0$.

As discussed in Section III-B, point cloud distortion might be related to both geometric and attribute information. Therefore, for a generalized potential energy field, we consider that the mass m_i is related to the point attributes, e.g., $m_{\mathbf{x}_i} = f_1(\mathbf{x}_0^I, \mathbf{x}_i^I)$; and the gravitational field $g_{\mathbf{x}_i}$ and distance $h_{\mathbf{x}_i}$ is related to the point geometric coordinates, reflecting the spatial location relative to the zero potential plane, e.g., $g_{\mathbf{x}_i} = f_2(\mathbf{x}_0^O, \mathbf{x}_i^O)$, $h_i = f_3(\mathbf{x}_0^O, \mathbf{x}_i^O)$. Note for a generalized field, $h_{\mathbf{x}_i}$ might not be limited to the Euclidean distance. We will give specific formulations for $m_{\mathbf{x}_i} g_{\mathbf{x}_i} h_{\mathbf{x}_i}$ in Section V-B and Section V-C.

For the entire point cloud $\mathcal{P}_{\mathbf{x}_0}$, we define the total point potential energy as

$$E_{\mathcal{P}_{\mathbf{x}_0}} = \sum_{\mathbf{x}_i \in \mathcal{P}_{\mathbf{x}_0}} E_{\mathbf{x}_i}, \tag{2}$$

which aggregates the the potential energies of all the points in $\mathcal{P}_{\mathbf{x}_0}$. The proposed total point potential energy is a quantitative measure of the state of a 3D point cloud, reflecting the geometric, attributive, and contextual information in this point cloud.

An advantage of the potential energy is sensitive to the isometrical distortion, which plays an import role in machine vision tasks to generate high-fidelity samples (Please refer to appendix for more details).

Definition 1: isometrical distortion distinguishability. Let $\mathcal{P}$ be a reference point cloud and $\mathcal{P}' = \mathcal{P} \oplus \Delta$ be a distorted point cloud with $\oplus$ the element-wise addition and Δ some perturbation that distorts the 3D coordinates or attributes of one or multiple points. A measure $\mathbf{H}$ is isometrical distortion indistinguishable when we can find a mapping function $\mathbf{Q}$, so that $\mathbf{H}(\mathcal{P}, \mathcal{P}') = \mathbf{Q}(\Delta)$ for an arbitrary perturbation Δ.

Intuitively, the isometrical distortion indistinguishability reflects that the measure $\mathbf{H}$ is only aware of the distortion and is invariant to the ego property of a point cloud. This is unfavorable because in this case multiple perturbations could lead to the same measure value; that is, $\mathbf{H}$ might not distinguish a specific perturbation applied on a point cloud. A good distortion quantification between two point clouds should be isometrical distortion distinguishable.

Toy example. We illustrate a toy example of the isometrical distortion in Fig. 3. Assuming $\mathbf{x}_0$ is the origin, the yellow line, note as $\mathcal{Y}$, is a coordinate axis, which represents the spatial distance to the origin or color intensity difference compared with the origin. When the axis means the Euclidean distance to the origin, for a point $\mathbf{x}_1 \in \mathcal{Y}$, $\mathbf{H}_1$ is a measure that calculates the Euclidean distance differences after applying perturbation Δ. Mathematically,

$$\begin{aligned}\mathbf{H}_1(\mathbf{x}_1, \mathbf{x}_1 \oplus \Delta) &= |\|\mathbf{x}_1 - \mathbf{x}_0\|_2 - \|\mathbf{x}_1' - \mathbf{x}_0\|_2| \\ &= |\|\mathbf{x}_1 - \mathbf{x}_0\|_2 - \|\mathbf{x}_1'' - \mathbf{x}_0\|_2| = \Delta.\end{aligned}$$

We see that $\mathbf{H}_1$ is not sensitive to the direction of perturbation. The distorted locations $\mathbf{x}_1'$ and $\mathbf{x}_1''$ play the same role for $\mathbf{H}_1$.

Moreover, for a point $\mathbf{x}_2 \in \mathcal{Y}, \mathbf{x}_2 \neq \mathbf{x}_1$,

$$\begin{aligned}\mathbf{H}_1(\mathbf{x}_2, \mathbf{x}_2 \oplus \Delta) &= |\|\mathbf{x}_2 - \mathbf{x}_0\|_2 - \|\mathbf{x}_2' - \mathbf{x}_0\|_2| \\ &= |\|\mathbf{x}_2 - \mathbf{x}_0\|_2 - \|\mathbf{x}_2'' - \mathbf{x}_0\|_2| \\ &= \Delta = \mathbf{H}_1(\mathbf{x}_1, \mathbf{x}_1 \oplus \Delta).\end{aligned}$$

We see that $\mathbf{H}_1$ is not sensitive to the initial location of perturbation. The initial locations $\mathbf{x}_1$ and $\mathbf{x}_2$ play the same role for $\mathbf{H}_1$.

When the axis means the color intensity difference compared with the origin, and $\mathbf{H}_1$ is a measure that captures the color differences after perturbation, the above derivations are still true.

We consider six point clouds, $\mathcal{P}_1 = [\mathbf{x}_0, \mathbf{x}_1]$, $\mathcal{P}_2 = [\mathbf{x}_0, \mathbf{x}_1']$, $\mathcal{P}_3 = [\mathbf{x}_0, \mathbf{x}_1'']$, $\mathcal{P}_4 = [\mathbf{x}_0, \mathbf{x}_2]$, $\mathcal{P}_5 = [\mathbf{x}_0, \mathbf{x}_2']$, and $\mathcal{P}_6 = [\mathbf{x}_0, \mathbf{x}_2'']$. Each point cloud has two points, and CD as $\mathbf{H}_1$. Note the specific calculation method of CD is as follows:

$$d_{\mathrm{CD}}(\mathcal{P}_1, \mathcal{P}_2) = \sum_{\mathbf{p}\in\mathcal{P}_1} \min_{\mathbf{q}\in\mathcal{P}_2} \| \mathbf{p} - \mathbf{q} \|_2^2 + \sum_{\mathbf{q}\in\mathcal{P}_2} \min_{\mathbf{p}\in\mathcal{P}_1} \| \mathbf{p} - \mathbf{q} \|_2^2. \tag{3}$$
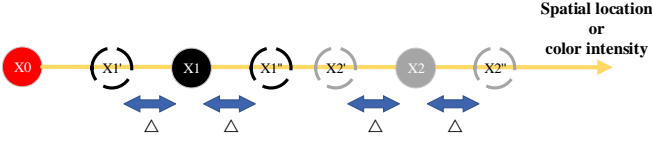
Fig. 3: Toy example of isometrical distortion.

We have

$$d_{\mathrm{CD}}(\mathcal{P}_1, \mathcal{P}_2) = d_{\mathrm{CD}}(\mathcal{P}_1, \mathcal{P}_3) = d_{\mathrm{CD}}(\mathcal{P}_4, \mathcal{P}_5)$$
$$= d_{\mathrm{CD}}(\mathcal{P}_4, \mathcal{P}_6) = 2\Delta^2.$$

We see that CD cannot detect the influence of the same perturbation on different point clouds. The reason is that the CD only considers individual point features.

On the other hand, point potential energy takes contextual information into consideration, which can distinguish isometrical distortion. Specifically, we use $m_{\mathbf{x}_i}$, $g_{\mathbf{x}_i}$ and $h_{\mathbf{x}_i}$ to describe the unique role of each point for the whole point cloud. Under the influence of the same perturbation, each point will give a different response to facilitate discrepancy quantification. We will validate the isometrical distortion distinguishability of point potential energy through multiple toy examples in Section V-D2.

V. MPED: MEASURING POINT CLOUD DISTORTION VIA MULTISCALE POTENTIAL ENERGY DISCREPANCY

As illustrated in Fig. 1, we present the proposed MPED in several steps: i) neighborhood construction. Both source and target point clouds are divided into multiple local neighborhoods based on a set of points as the neighborhood centers; ii) for each neighborhood, the neighborhood center is set to be the zero-potential-energy plane. We then compute the potential energy for each point, and the total potential energy of a neighborhood is the summation of the potential energies of all the points insides the neighborhood; iii) we propose the potential energy discrepancy to quantify the difference between the source and the target point clouds, leading to the single-scale potential energy discrepancy (SPED); and iv) we finally extend SPED to a multiscale form, e.g., MPED.

A. General Form of MPED

Here we first present the general form of MPED, then we will present the implementation details of MPED for machine and human perception tasks in Section V-B and V-C.

Neighborhood construction. Let $\mathcal{S} = \{\mathbf{s}_i \in \mathbb{R}^6\}_{i=1}^N$ and $\mathcal{T} = \{\mathbf{t}_i \in \mathbb{R}^6\}_{i=1}^M$ be source and target 3D point clouds with N and M points, respectively. Assuming a set of points as neighborhood center $\mathcal{C} = \{\mathbf{c}_i \in \mathbb{R}^6\}_{i=1}^L$, its neighboring point cloud in the source point cloud is $\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{source}} \subset \mathcal{S}$ with $\mathbf{c}_i$ being the origin, which collects $\mathbf{c}_i$'s K closest points in $\mathcal{S}$. Similarly, $\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{target}} \subset \mathcal{T}$ denotes $\mathbf{c}_i$'s K closest points in $\mathcal{T}$. Now, each center point has two local neighboring point cloud that reflects its contextual roles in source and target point clouds, respectively. We then can use the total point potential energy, Eq. (2) to quantify the spatial distribution

properties in two neighboring point clouds; that is, $E_{\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{source}}}$ and $E_{\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{target}}}$, respectively.

Point potential energy discrepancy. As we demonstrate in Section III-B, the distortion of point clouds is related to both geometric and attributive information. To capture this distortion, we can leverage the point potential energy proposed in Section IV, which quantifies the spatial distribution of 3D points via Eqs. (1) and (2). Specifically, we consider the mass, the spatial field, and the distance as follows,

$$m_{\mathbf{x}_i} = \begin{cases} f_1(\mathbf{x}_0^I, \mathbf{x}_i^I), & \text{if } \mathbf{x}_0^I, \mathbf{x}_i^I \neq \phi \\ 1, & \text{otherwise.} \end{cases} \tag{4}$$

$$g_{\mathbf{x}_i} = f_2(\mathbf{x}_0^O, \mathbf{x}_i^O), \tag{5}$$

$$h_{\mathbf{x}_i} = f_3(\mathbf{x}_0^O, \mathbf{x}_i^O). \tag{6}$$

Here the mass $m_{\mathbf{x}_i}$ is related to the point color information. Considering some point clouds do not have color information, we set: i) when $\mathbf{x}_0^I, \mathbf{x}_i^I \neq \phi$, $m_{\mathbf{x}_i} = f_1(\mathbf{x}_0^I, \mathbf{x}_i^I)$; ii) when $\mathbf{x}_0^I, \mathbf{x}_i^I = \phi$, $m_{\mathbf{x}_i} = 1$ to avoid point potential energy becomes 0. The distance $h_{\mathbf{x}_i}$ is related to point coordinate information. The spatial field $g_{\mathbf{x}_i}$ can be regarded as a weighting factor when pooling all the point energies via Eq. (2). A benefit of introducing $g_{\mathbf{x}_i}$ is to distinguish isometrical distortion. We will validate this property in Section V-D2. We will give specific formulations of $m_{\mathbf{x}_i}$, $g_{\mathbf{x}_i}$ and $h_{\mathbf{x}_i}$ for human perception tasks in Section V-B and for machine perception tasks in Section V-C. Based on above three components, we use Eq. (2) to calculate $E_{\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{source}}}$ and $E_{\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{target}}}$, respectively.

We first set the distortion of point with greater energy to contribute more to the total potential energy variation. Consider in classic mechanics, for the object with the same mass, the farther the object is from the zero potential energy plane, the greater the energy. We need g_i to satisfy the following requirement:

Requirement 1. For two points $\mathbf{x}_1$ and $\mathbf{x}_2$ with same mass, if $h_{\mathbf{x}_1} \leq h_{\mathbf{x}_2}$, g_i should satisfy $g_{\mathbf{x}_1} \leq g_{\mathbf{x}_2}$.

With $g_{\mathbf{x}_i}$ satisfies **Requirement 1**, potential energy can better detect the isometrical distortion, which will be proved in Section V-D1 **Theorem 1**.

We now propose a potential energy discrepancy to measure the difference between the source and the target point cloud. Specifically, the discrepancy is defined as

$$\mathrm{SPED}_K = \mathbf{A}\left(\{\mathbf{M}^i(E_{\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{source}}}, E_{\mathcal{N}_{\mathbf{c}_i,K}^{\mathrm{target}}})\}_{i=1}^L\right), \tag{7}$$

where L is the number of neighborhood centers, $\mathbf{M}^i(\cdot)$ represents the neighborhood-wise operation for the ith neighborhood, and $\mathbf{A}(\cdot)$ represents an aggregation method that generates the final objective score. We will give the specific formulations of $\mathbf{M}^i(\cdot)$ and $\mathbf{A}(\cdot)$ in Section V-B and Section V-C.

Multiscale point potential energy discrepancy: MPED. A hyperparameter K is introduced in SPED to establish a spatial neighborhood. Inspired by MSSIM [19], we extend SPED to a multiscale form, e.g., MPED via

$$\mathrm{MPED}_\Psi = \frac{1}{|\Psi|} \sum_{K \in \Psi} \mathrm{SPED}_K, \tag{8}$$

in which Ψ represents a collection of K, $|\Psi|$ represents the number of elements in Ψ.

B. MPED Implementation Details on Human Perception Tasks.

In this section, we present the detailed implementation of MPED on human perception tasks. The point clouds used for human-oriented applications target for visualizations, which usually require dense point distributions and contain rich color information. For example, MPEG compression standardization, "RedandBlack" [31] data has 729,133 points rendered inside a cube of size $393 \times 977 \times 232$. Such dense sampling provides realistic visual experiences of 3D objects observed in a human visual system.

Neighborhood center selection.

Human visual system is more sensitive to high-frequency information [32] [20], such as edge and contour. Therefore, human perception tasks need to pay more attention to the distortion of high-frequency features. In Section V, we establish multiples neighborhoods to calculate point potential energy via setting a set of neighborhood centers. Therefore, we use high-pass filters to select a set of high-frequency points as neighborhood centers. The total point potential energy of these neighborhoods can reflect the characteristics of point cloud high-frequency structures.

Refer to the proposal in [5], we use the method proposed in [33] to filter high-frequency points as element of $\mathcal{C}$. Specifically,

$$\mathcal{C} = \lfloor \Upsilon(\mathcal{S}, \beta) \rfloor_L \in \mathbb{R}^{L \times 6}, L \ll N,$$

where $\Upsilon(\cdot)$ is a Haar-like filter, β is filter length, and L is the number of points for $\mathcal{C}$. For more details, please check [33] [5].

Point potential energy. In Section V, we give a generic formulation of point potential energy. In this part, we present the specific implementations of Eqs. (5), (4) and (6). Specifically, we consider the mass, the spatial field, and the distance as follows,

$$m_{\mathbf{x}_i} = f_1(\mathbf{x}_0^I, \mathbf{x}_i^I) = \sum_{j}^{3} k_j |(\mathbf{x}_0^I)_j - (\mathbf{x}_i^I)_j| + 1,$$

$$g_{\mathbf{x}_i} = f_2(\mathbf{x}_0^O, \mathbf{x}_i^O) = \left(1 + e^{-\frac{\|\mathbf{x}_i^O - \mathbf{x}_0^O\|_2^2}{\sigma^2}} \Gamma_K \right)^{-1}, \quad (9)$$

$$h_{\mathbf{x}_i} = f_3(\mathbf{x}_0^O, \mathbf{x}_i^O) = \|\mathbf{x}_i^O - \mathbf{x}_0^O\|_p^p,$$

where k_i represents the weighting factors between different color channel, e.g., if $\mathbf{x}^I \in \text{RGB}$, $k_R : k_G : k_B = 1 : 2 : 1$; if $\mathbf{x}^I \in \text{YUV}$, $k_Y : k_U : k_V = 6 : 1 : 1$ [5]. We have presented in Section V that $g_{\mathbf{x}_i}$ is used to pooling element which in the same neighborhood, therefore we introduce $\Gamma_K = 0$ if $K = 1$, $\Gamma_K = 1$ if $K > 1$. $m_i \geq 1$ to avoid energy become 0 when $\mathbf{x}_0^I = \mathbf{x}_i^I$. When $p = 1$, $\| \cdot \|_1^1$ represents 1 norm; when $p = 2$, $\| \cdot \|_2^2$ represents the square of 2 norm. The form of $g_{\mathbf{x}_i}$ in Eq. (9) satisfies **Requirement 1**.

Discrepancy calculation. For the human visual system, the final score is derived based on the whole object, and the score is a normalized value. Therefore, we first aggregate the energy

of multiple neighborhoods and obtain the overall energy of source and target point clouds. We formulate the functions proposed in Eqs. (7) and (8) as follows,

$$\{\mathbf{M}^i(E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{source}}}, E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{target}}})\}_{i=1}^{L} = [\sum_{\mathbf{c}_i \in \mathcal{C}} E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{source}}}, \sum_{\mathbf{c}_i \in \mathcal{C}} E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{target}}}]$$

$$= [E_{\text{source}}, E_{\text{target}}].$$

For E_{source} and E_{target}, the discrepancy is defined as

$$\text{SPED}_K = \mathbf{A}([E_{\text{source}}, E_{\text{target}}])$$

$$= \frac{2 \times E_{\text{source}} \times E_{\text{target}} + T}{E_{\text{source}}^2 + E_{\text{target}}^2 + T},$$

where T is a small constant to prevent the denominator from being zero. Obviously, the range of SPED_K is $[0, 1]$. Subjective scores distributed in a certain range [21], objective scores also need to be normalized into a fixed range. Refer to SSIM [32] that normalizes the objective score into [0,1], we adopt the same method to calculate the discrepancy.

C. MPED Implementation Detials on Machine Perception Tasks.

In this section, we present the detailed implementation of MPED on machine perception tasks. The point clouds used for machine-oriented applications target for 3D machine perception and the point distributions are usually sparser than those for human perception. For instance, samples in ShapeNet [34] and ModelNet [35] consist of thousands of points that only reflect rough shapes of objects (e.g., Fig. 2 (a)). These samples are often used in machine perception tasks, including supervised learning and unsupervised learning. Specifically, unsupervised machine perception tasks extract informative features from point cloud samples without human supervision [7]. Some typical tasks include reconstruction [8], [9], completion [10], [11], and upsampling [12]. Recently, deep neural networks are common tools to achieve those tasks, and an objective point cloud distortion quantification is usually needed as the supervision to train deep neural networks.

Neighborhood center selection. Different from human perception tasks, machine perception tasks usually aim to generate samples that exactly the same as the reference samples. Therefore, distortion quantifications used in machine perception tasks need to equality detect the shape deformation at any location.

We make MPED equally detect the shape deformations at any location via applying the following neighborhood center selection strategy: both source and target samples are used as references to calculate the geometrical distortion. Therefore, we set

$$\mathcal{C} = \mathcal{S} \bigcup \mathcal{T}.$$

Point potential energy. Specifically, we consider the mass, the spatial field, and the distance as follows,

$$m_{\mathbf{x}_i} = 1,$$

$$g_{\mathbf{x}_i} = f_2(\mathbf{x}_0^O, \mathbf{x}_i^O) = \left(1 + e^{-\frac{\|\mathbf{x}_i^O - \mathbf{x}_0^O\|_2^2}{\sigma^2}} \Gamma_K \right)^{-1}, \quad (10)$$

$$h_{\mathbf{x}_i} = f_3(\mathbf{x}_0^O, \mathbf{x}_i^O) = \|\mathbf{x}_i^O - \mathbf{x}_0^O\|_p^p.$$

Same to human perception tasks, we set $\Gamma_K = 0$ if $K = 1$, $\Gamma_K = 1$ if $K > 1$. $m_{\mathbf{x}_i} = 1$ to avoid energy become 0 when $\mathbf{x}_0^I = \mathbf{x}_i^I = \phi$. When $p = 1$, $\|\cdot\|_1^1$ represents 1 norm; when $p = 2$, $\|\cdot\|_2^2$ represents the square of 2 norm. The form of $g_{\mathbf{x}_i}$ in Eq. (10) satisfies **Requirement 1**.

Discrepancy calculation. The networks in machine perception tasks solve the optimization problem by minimizing the differences between original and reconstructed samples. We need to quantify the discrepancy in each neighborhood. Therefore, we propose a machine perception tasks oriented discrepancy calculation method.

We formulate the functions proposed in Eqs. (7) and (8) as follows. Specifically, We first calculate the neighborhood-wise difference, e.g.,

$$\mathbf{M}^i(E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{source}}}, E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{target}}}) =$$
$$|E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{source}}} - E_{\mathcal{N}_{\mathbf{c}_i,K}^{\text{target}}}| = d_{\mathbf{c}_i,K}.$$

Then, the discrepancy is defined as

$$\text{SPED}_K = \mathbf{A}(\{d_{\mathbf{c}_i,K}\}_{i=1}^L) = \sum_{i=1}^L d_{\mathbf{c}_i,K}.$$

D. Property

In this section, we discuss the properties of the proposed MPED. Specifically, we first demonstrate that the proposed MPED is **monotonic** when encountering isometrical distortion (i.e., **Theorem 1** and **Theorem 2**). Then, we prove that CD is a special case of the proposed MPED (i.e., **Theorem 3**). Finally, we illustrate several toy examples to validate that the proposed MPED satisfies all three conditions for an ideal objective distortion quantification (e.g., differentiability, low complexity, and distortion discrimination).

1) Theorem:

Theorem 1. *When* $m_1 = m_2$, $\|\mathbf{x}_1^O - \mathbf{x}_0^O\|_2 \leq \|\mathbf{x}_2^O - \mathbf{x}_0^O\|_2$, *g satisfies* **Requirement 1**. *Then,*

$$E_{\mathbf{x}_1} \leq E_{\mathbf{x}_2}.$$

For an isometrical perturbation $\epsilon \in \mathbb{R}$, *we further have*

$$E_{\mathbf{x}_1+\epsilon(\mathbf{x}_1-\mathbf{x}_0)} - E_{\mathbf{x}_1} \leq E_{\mathbf{x}_2+\epsilon(\mathbf{x}_2-\mathbf{x}_0)} - E_{\mathbf{x}_2}.$$

Proof. $F = ma = mg_{\mathbf{x}}$, F represents force, and a represents accelerated speed. Assuming in Euclidean distance space, we plot the variation of F corresponding to distance h in Fig. 4, $\because \|\mathbf{x}_1 - \mathbf{x}_0\|_2 \leq \|\mathbf{x}_2 - \mathbf{x}_0\|_2$, ϵ is an isometrical perturbation, we use $h = 0$ represents $\mathbf{x}_0$, $h = 1$ represents $\mathbf{x}_1$, $h = 2$ represents $\mathbf{x}_2$, and $\epsilon = 1$. $\therefore E_{\mathbf{x}_1+\epsilon(\mathbf{x}_1-\mathbf{x}_0)} - E_{\mathbf{x}_1} - (E_{\mathbf{x}_2+\epsilon(\mathbf{x}_2-\mathbf{x}_0)} - E_{\mathbf{x}_2}) = \Delta E_1 - \Delta E_2 \leq 0$. $\square$

Theorem 1 explains that with two points share the same mass while different distances to zero potential plane, MPED can detect the geometrical isometrical distortion via total energy variation.

Theorem 2. *When* $\|\mathbf{x}_1^O - \mathbf{x}_0^O\|_2 = \|\mathbf{x}_2^O - \mathbf{x}_0^O\|_2$, $m_1 \leq m_2$, *g satisfies* **Requirement 1**. *Then,*
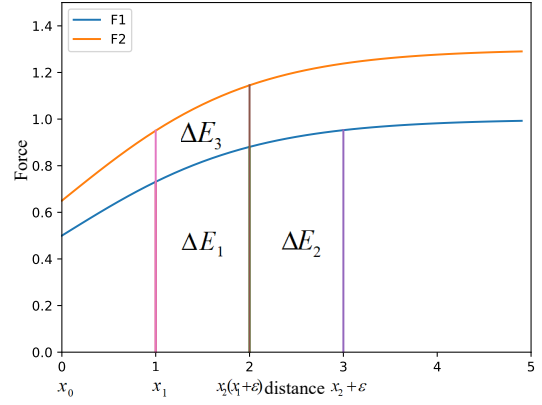
$$E_{\mathbf{x}_1} \leq E_{\mathbf{x}_2}.$$

Fig. 4: Force-distance curves.

For an isometrical perturbation $\epsilon \in \mathbb{R}$, *we further have*

$$E_{\mathbf{x}_1+\epsilon(\mathbf{x}_1-\mathbf{x}_0)} - E_{\mathbf{x}_1} \leq E_{\mathbf{x}_2+\epsilon(\mathbf{x}_2-\mathbf{x}_0)} - E_{\mathbf{x}_2}.$$

Proof. $F_1 = m_1 a = m_1 g_{\mathbf{x}_1} \leq F_2 = m_2 a = m_2 g_{\mathbf{x}_2}$. Assuming in Euclidean distance space, we plot the variation of F_1 and F_2 corresponding to distance h in Fig. 4. $\because \|\mathbf{x}_1 - \mathbf{x}_0\|_2 = \|\mathbf{x}_2 - \mathbf{x}_0\|_2$, ϵ is an isometrical perturbation, we use $h = 0$ represents $\mathbf{x}_0$, $h = 1$ represents $\mathbf{x}_1$ and $\mathbf{x}_2$, and $\epsilon = 1$. $\therefore E_{\mathbf{x}_1+\epsilon(\mathbf{x}_1-\mathbf{x}_0)} - E_{\mathbf{x}_1} - (E_{\mathbf{x}_2+\epsilon(\mathbf{x}_2-\mathbf{x}_0)} - E_{\mathbf{x}_2}) = \Delta E_1 - (\Delta E_3 + \Delta E_1) \leq 0$ $\square$

Theorem 2 explains that with two points share the same distance to the zero potential planes while different masses, MPED can detect the geometrical isometrical distortion via total energy variation.

Theorem 3. *For* $\mathcal{P}_1$, $\mathcal{P}_2$, *if* $\forall \mathbf{p}_i \in \mathcal{P}_1$, $\mathbf{q}_i \in \mathcal{P}_2$ *satisfy* $\mathbf{p}_i^I, \mathbf{q}_i^I = \phi$ ($\mathcal{P}_1$, $\mathcal{P}_2$ *only have coordinate information*), $\Psi = [1]$, $p = 2$, $\mathcal{C} = \mathcal{P}_1 \bigcup \mathcal{P}_2$, *MPED* $= d_{\text{CD}}$.

Proof. Refer to Eq. (4), if point clouds do not have additional attributes except for spatial coordinate, we have $m_{\mathbf{x}_i} = 1$. Considering $c_i \in \mathcal{P}_1$, the nearest neighbor for $\mathbf{c}_i$ in $\mathcal{P}$ is exactly $\mathbf{c}_i$. Therefore, $\mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_1}} = 0$. The same when $c_i \in \mathcal{P}_2$, e.g., $\mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_2}} = 0$. Therefore,

$$\begin{aligned}
\text{MPED} &= \sum_{\mathbf{c}_i \in \mathcal{C}} d_{\mathbf{c}_i,1} \\
&= \sum_{\mathbf{c}_i \in \mathcal{P}_1} |\mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_1}} - \mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_2}}| + \sum_{\mathbf{c}_i \in \mathcal{P}_2} |\mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_1}} - \mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_2}}| \\
&= \sum_{\mathbf{c}_i \in \mathcal{P}_1} |0 - \mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_2}}| + \sum_{\mathbf{c}_i \in \mathcal{P}_2} |\mathrm{E}_{\mathcal{N}_{\mathbf{c}_i,1}^{\mathcal{P}_1}} - 0| \\
&= \sum_{\mathbf{c}_i \in \mathcal{P}_1, \mathbf{q}_i \in \mathcal{P}_2} \|\mathbf{c}_i - \mathbf{q}_i\|_2^2 + \sum_{\mathbf{c}_i \in \mathcal{P}_2, \mathbf{q}_i \in \mathcal{P}_1} \|\mathbf{c}_i - \mathbf{q}_i\|_2^2 \\
&= d_{\text{CD}}.
\end{aligned}$$

$\square$

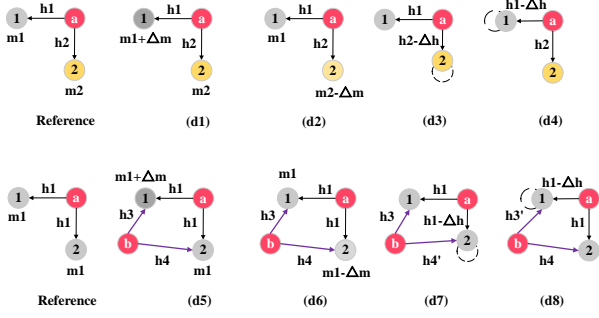**Theorem 3** explains that CD is a special case of MPED.

Fig. 5: Toy examples of geometrical and attributive distortion sensitivity.

2) Qualification: We further check the three conditions for an ideal distortion quantification proposed in Section I.

• Differentiability. Refer to specific implementations illustrated in Section V-B and Section V-C, all the steps adopted in MPED are clearly differentiable;

• Low complexity. The complexity of MPED is $O(N^2)$, which is the same as CD. Note the complexity of EMD is $O(N^3)$;

• For distortion discrimination, we use several toy examples to validate. Specifically, we use CD, EMD and PSNR_{YUV} [14] as comparison. For PSNR_{YUV}, it first matches points via Euclidean distance like CD, e.g., $\mathbf{p}_i \in \mathcal{P}_1$, $\mathbf{q}_i \in \mathcal{P}_2$, $\min(\|\mathbf{p}_i^O - \mathbf{q}_i^O\|_2)$. Then calculating the color difference between paired $\mathbf{p}_i$ and $\mathbf{q}_i$. Specifically,

$$\text{PSNR}_{\text{YUV}} = (6 \times \text{PSNR}_\text{Y} + \text{PSNR}_\text{U} + \text{PSNR}_\text{V})/8,$$
$$\text{PSNR}_{\text{j},\text{j} \in \{Y,U,V\}}$$
$$= 10 \log_{10}(\frac{255^2}{\text{mean}(\sum_{\mathbf{p}_i \in \mathcal{S}_1, \mathbf{q}_i \in \mathcal{S}_2}^{\min(\mathbf{p}_i^O - \mathbf{q}_i^O)}[(\mathbf{p}_i^I)_j - (\mathbf{q}_i^I)_j]^2)}), \quad (11)$$

where $\text{mean}(\cdot)$ represents average operator.

Case 1: Sensitivity to attributive isometrical distortion. As illustrated in Fig. 5 d1 and d2, Δm means a perturbation on three color channels simultaneously.

◇ $d_{\text{CD}_{\text{d1}}} = 0 = d_{\text{CD}_{\text{d2}}}$;
◇ $d_{\text{EMD}_{\text{d1}}} = 0 = d_{\text{EMD}_{\text{d2}}}$;
◇ $\text{PSNR}_{\text{YUV}}^{\text{d}_1} = \text{PSNR}_{\text{YUV}}^{\text{d}_2}$;
◇ Because of $h_1 \neq h_2$, $\text{MPED}^{d_1} \neq \text{MPED}^{d_2}$.

Case 2: Sensitivity to geometrical isometrical distortion. As illustrated in Fig. 5 d3 and d4, Δh means a perturbation on distance.

◇ $d_{\text{CD}_{\text{d3}}} = \Delta h = d_{\text{CD}_{\text{d4}}}$;
◇ $d_{\text{EMD}_{\text{d3}}} = \Delta h = d_{\text{EMD}_{\text{d4}}}$;
◇ $\text{PSNR}_{\text{YUV}}^{\text{d}_3} = 0 = \text{PSNR}_{\text{YUV}}^{\text{d}_4}$;
◇ Because of $h_1 \neq h_2$, $m_1 \neq m_2$, $\text{MPED}^{d_1} \neq \text{MPED}^{d_2}$ under most cases.

Case 3: Sensitivity to attributive distortions at various locations. As illustrated in Fig. 5 d5 and d6,

◇ $d_{\text{CD}_{\text{d5}}} = 0 = d_{\text{CD}_{\text{d6}}}$;
◇ $d_{\text{EMD}_{\text{d5}}} = 0 = d_{\text{EMD}_{\text{d6}}}$;
◇ $\text{PSNR}_{\text{YUV}}^{\text{d}_5} = \text{PSNR}_{\text{YUV}}^{\text{d}_6}$;
◇ For neighborhood a, $\text{MPED}^{a,d_5} = \text{MPED}^{a,d_6}$; For neighborhood b, because of $h_3 \neq h_4$, $\text{MPED}^{b,d_5} \neq$

Requirement	CD	EMD	PSNR_{YUV}	MPED
differentiability	✓	✓	✓	✓
low complexity	✓	✗	✓	✓
distortion discrimination	✗	✗	✗	✓

TABLE I: MPED satisfies all the three conditions.

MPED^{b,d_6}, therefore, $\text{MPED}^{d_5} = \text{MPED}^{a,d_5} + \text{MPED}^{b,d_5} \neq \text{MPED}^{a,d_6} + \text{MPED}^{b,d_6} = \text{MPED}^{d_6}$.

Case 4: Sensitivity to geometrical distortions at various locations. As illustrated in Fig. 5 d7 and d8,

◇ $d_{\text{CD}_{\text{d5}}} = \Delta h = d_{\text{CD}_{\text{d6}}}$;
◇ $d_{\text{EMD}_{\text{d5}}} = \Delta h = d_{\text{EMD}_{\text{d6}}}$;
◇ $\text{PSNR}_{\text{YUV}}^{\text{d}_5} = 0 = \text{PSNR}_{\text{YUV}}^{\text{d}_6}$;
◇ For neighborhood a, $\text{MPED}^{a,d_7} = \text{MPED}^{a,d_8}$; For neighborhood b, because of $h_3 \neq h_4$, $\text{MPED}^{b,d_7} \neq \text{MPED}^{b,d_8}$, therefore, $\text{MPED}^{d_7} = \text{MPED}^{a,d_7} + \text{MPED}^{b,d_7} \neq \text{MPED}^{a,d_8} + \text{MPED}^{b,d_8} = \text{MPED}^{d_8}$.

We compare the properties of four typical distortion quantifications in Table I.

VI. EXPERIMENT

In this section, we test the proposed MPED on both human and machine perception tasks.

A. Experimental Evaluations on Human Perception Tasks

For human perception tasks, we test MPED on two fairly large database: i) the first is SJTU-PCQA **People** [21]; ii) the second is LSPCQA [17].

Subjective point cloud assessment database

• SJTU-PCQA database. There are 5 high-quality human body point cloud samples in SJTU-PCQA **People** category. Each native point cloud sample is augmented with 7 different types of impairments under 6 levels, including four individual distortions, Octree-based compression (OT), Color noise (CN), Geometry Gaussian noise (GGN), Downsampling (DS), and three superimposed distortions, such as Downsampling and Color noise (D+C), Downsampling and Geometry Gaussian noise (D+G), Color noise and Geometry Gaussian noise (C+G). Please refer to [21] for more details.

• LSPCQA database. There are 104 high-quality point cloud samples in LSPCQA, each reference sample is processed with 34 types of impairments under 7 levels. In all there are 24,752 samples in LSPCQA, 1,020 of them provide mean opinion scores. The distortion types include Quantization noise, Contrast change, V-PCC, G-PCC, Local rotation, Luminance noise, and so on. For more details about LSPCQA, please check [17].

Parameters of MPED. i) σ^2. For fairness, when $\mathbf{c}_i$ belongs to source sample, we use the average square of Euclidean distance of $\mathbf{c}_i$ and its neighbors collected from source sample as σ^2, e.g., $\sigma^2 = \sum_{\mathbf{p}_i \in \mathcal{N}_{\mathbf{c}_i,K}^{\text{source}}} \|\mathbf{p}_i - \mathbf{c}_i\|_2^2/K$, and vice versa; ii) Ψ. For human perception tasks, we set Ψ as [10, 5] for both $p = 1$ and $p = 2$; iii) L and β. Refer to [5], we set $L = N/1000$ and $\beta = 4$; iv) T. Refer to [5], we simply set $T = 0.001$; v)k_j. We first using RGB color space to calculate $m_{\mathbf{x}_i}$, therefore, we set $k_R : k_G : k_B = 1 : 2 : 1$.

			PLCC								SROCC								RMSE							
		quantification:	OT	CN	GGN	DS	D+C	D+G	C+G	ALL	OT	CN	GGN	DS	D+C	D+G	C+G	ALL	OT	CN	GGN	DS	D+C	D+G	C+G	ALL
People (ave)	M	p2po	0.88	-	0.97	0.96	0.97	0.99	0.99	0.89	0.80	-	0.95	0.92	0.97	0.97	0.97	0.79	0.84	-	0.67	0.64	0.61	0.38	0.40	1.11
		p2pl	0.90	-	0.96	0.80	0.77	0.99	0.99	0.74	0.80	-	0.94	0.56	0.66	0.98	0.98	0.66	0.78	-	0.71	1.36	1.55	0.34	0.39	1.66
	H	p2po	0.86	-	0.96	0.91	0.84	0.99	0.99	0.80	0.80	-	0.95	0.89	0.81	0.97	0.97	0.70	0.92	-	0.66	0.95	1.32	0.38	0.40	1.49
		p2pl	0.89	-	0.97	0.82	0.86	0.99	0.99	0.71	0.81	-	0.95	0.77	0.87	0.97	0.97	0.71	0.82	-	0.67	1.31	1.23	0.37	0.40	1.83
		PSNR$_{YUV}$	0.54	0.97	0.86	0.74	0.98	0.85	0.98	0.71	0.52	0.94	0.82	0.74	0.97	0.77	0.97	0.71	1.52	0.48	1.34	1.56	0.49	1.37	0.52	1.74
		PCQM	0.81	0.92	0.97	0.91	0.97	0.98	0.99	0.89	0.77	0.88	0.95	0.83	0.97	0.97	0.98	0.89	1.06	0.72	0.66	0.94	0.57	0.47	0.42	1.12
		GraphSIM	0.81	0.90	0.97	0.97	0.95	0.99	0.98	0.89	0.71	0.82	0.96	0.91	0.95	0.96	0.97	0.88	1.05	0.78	0.62	0.55	0.79	0.43	0.52	1.13
		MPED($p=1$)	0.82	0.92	0.97	0.97	0.96	0.99	0.97	**0.92**	0.74	0.86	0.96	0.93	0.97	0.96	0.94	**0.93**	1.03	0.70	0.60	0.56	0.65	0.40	0.58	**0.95**
		MPED($p=2$)	0.83	0.92	0.97	0.97	0.97	0.99	0.97	**0.92**	0.76	0.85	0.96	0.92	0.97	0.97	0.96	0.91	1.01	0.71	0.65	0.54	0.62	0.39	0.56	1.00

TABLE II: Model performance (PLCC, SROCC and RMSE) for point clouds samples in SJTU-PCQA database in terms of different impairments. **People (ave)** represents treat all the samples in SJTU-PCQA as whole. The proposed MPED presents robust performance in terms of multiple types of distortion.

Performance evaluation. We compare our MPED with another 7 state-of-the-art distortion quantifications, e.g.,

- PSNR-MSE-P2point (M-p2po)
- PSNR-MSE-P2plane (M-p2pl)
- PSNR-Hausdorff-P2point (H-p2po)
- PSNR-Hausdorff-P2plane (H-p2pl)
- PSNR$_{YUV}$
- PCQM
- GraphSIM

To ensure the consistency between subjective scores (e.g., mean opinion scores) and objective predictions from various models, we map the objective predictions of different models to the same dynamic range following the recommendations suggested by the video quality experts group (VQEG) [36], [37], to derive popular PLCC for prediction accuracy, SROCC for prediction monotonicity, and RMSE for prediction consistency for evaluating the model performance. The larger PLCC or SROCC comes with better model performance. On the contrary, the lower RMSE is better. More details can be found in [36]. Note we use RGB color space in this part to calculate point mass.

• SJTU-PCQA database. Table II presents the performance of MPED and other state-of-the-art distortion quantifications on SJTU-PCQA. We see that: i) MPED presents best overall performance on SJTU-PCQA. Specifically, the overall PLCC, SROCC and RMSE of MPED($p=1$) is (0.92, 0.93, 0.95) and MPED($p=2$) is (0.92, 0.91, 1.00). While M-p2po is (0.89, 0.79, 1.11), M-p2pl is (0.74, 0.66, 1.66), H-p2po is (0.80, 0.70, 1.49), H-p2pl is (0.71, 0.66, 1.83), PSNR$_{YUV}$ is (0.71, 0.89, 1.12), PCQM is (0.89, 0.89, 1.12), and GraphSIM is (0.89, 0.88, 1.13); ii) the performance of MPED under $p = 1$ or $p = 2$ is close, which means the proposed MPED is robust on multiple selection of spatial distance.

For different distortion quantifications in terms of distortion types. We see that: i) M-p2po, M-p2pl, H-p2po, and H-p2pl cannot handle CN because they only consider point-wise geometrical features; ii) PSNR$_{YUV}$ presents the best performance on CN, while the worst performance on OT. Referring to Eq. (11), PSNR$_{YUV}$ first matches two points via nearest neighbor searching, then uses the color difference of point pairs as distortion measurement. Essentially, OT uses a central point to replace all the points within a spatial cube. The central points usually share similar attributes with replaced points, and can be used multiple times during point matching. Therefore, PSNR$_{YUV}$ are not sensitive to OT.

		quantification:	PLCC	SROCC	RMSE
LSPCQA	M	p2po	0.46	0.26	0.73
		p2pl	0.42	0.24	0.75
	H	p2po	0.36	0.21	0.77
		p2pl	0.36	0.21	0.77
		PSNR$_{YUV}$	0.50	0.48	0.72
		PCQM	0.32	0.42	0.75
		GraphSIM	0.33	0.31	0.78
		MPED($p=1$)	0.64	**0.61**	0.63
		MPED($p=2$)	**0.66**	0.60	**0.62**

TABLE III: Model performance (PLCC, SROCC and RMSE) for point clouds samples in LSPCQA database. The proposed MPED presents best PLCC, SROCC and RMSE on LSPCQA.

Color Space	PLCC	SROCC	RMSE
RGB	0.6569	0.6015	0.6229
YUV	0.6558	0.6030	0.6237
GCM	0.6575	0.6048	0.6225

TABLE IV: Model performance with various color spaces on LSPCQA. The proposed MPED presents stable performance for multiple types of color space.

• LSPCQA database. Table III presents the performance of MPED and other state-of-the-art distortion quantifications on LSPCQA. We see that: i) MPED is obviously superior to other distortion quantifications on LSPCQA. Specifically, the PLCC and SROCC of MPED are above 0.6, other distortion quantifications are all lower than 0.5; ii) the SROCCs of M-p2po, M-p2pl, H-p2po and H-p2pl are lower than 0.3. The reason is that there are several types of geometry lossless but color lossy distortion, such as color quantization dither (CQD). These four distortion quantifications cannot detect color distortion.

Ablation Study: Color Space.

We have exemplified the effectiveness of MPED based on the RGB-based color channel decomposition. In this part, we test the performance of MPED in terms of other color channel space, e.g., YUV and Gaussian color model (GCM) [38]. YUV and GCM are two color channel spaces that consist of one luminance component and two chrominance components. Given that human visual system is more sensitive to the luminance component [14], we set $k_l = 6$, $k_{c1} = k_{c2} = 1$ [5]. k_l represents the weighting factors of luminance component, and k_{c1}, k_{c2} for chrominance components. We use LSPCQA as the test database and the results shown in Table IV.

In Table IV, we see that the performance of MPED is very close for multiple color spaces, which proves the robustness of MPED.
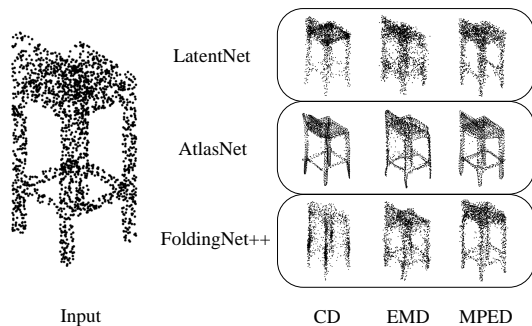
Fig. 6: Illustration of point cloud reconstruction. The proposed MPED presents the best reconstruction quality.

Fig. 7: Illustration of point cloud completion. The proposed MPED presents the best completion quality.

B. Experimental Evaluation on Machine Perception Tasks

For machine perception, we test MPED on three typical unsupervised learning tasks, e.g., point cloud reconstruction, shape completion, and upsampling.

Parameters of MPED. i) σ^2. Same to human perception tasks; ii) Ψ. For machine perception, we set Ψ as $\Psi_1 = [5, 1]$ and $\Psi_2 = [10, 5, 1]$ for both $p = 1$ and $p = 2$;

Performance evaluation. To validate the proposed MPED as a loss function for machine perception tasks, we compare it with CD [39] and EMD [40]. For each task, with various supervisions, we test several state-of-the-art networks under the same training settings, e.g., same network, same epochs, and same network parameters. To compare the final performance, we use CD, EMD, and JSD as objective quantitative evaluation. JSD measures the marginal distributions between the generated and reference samples [8]. To distinguish CD as loss function and quantitative metrics, we use **CD** as loss function while CD as the quantitative metric. The EMD in the same way.

• 3D Point Cloud Reconstruction

Here we adopt ShapeNet dataset used in [8] and validate the proposed MPED on three typical networks for point cloud reconstruction with default network parameter setting expect training epochs, e.g., LatentNet [8], AtlasNet [30], and FoldingNet++ [9]. For a fair comparison, we randomly set training epochs of LatentNet, AtlesNet and FoldingNet++ as 500, 300, and 100 due to the default training epochs selected by presenting best CD value on the test sequences. Four sub-categories from ShapeNet are used as test sequences, e.g., "Chair", "Airplane", "Table" and "Rifle". Specifically, for each sub-category, we set 75% as training set and 25% as testing set. Table V shows the reconstruction performance of three networks with various supervisions. We use **red**, **blue** and **cyan** to highlight best performance for three networks, respectively. Besides the individual performance, we also present average performance treating four categories as a whole.

We see that: i) for LatentNet, **CD** presents best CD, worst EMD and JSD; **EMD** presents best EMD, worst CD, and passable JSD; MPED presents best JSD, while passable CD and EMD; ii) for AtlasNet and FoldingNet++, MPED presents dominated performance while only several EMDs are awarded by **EMD**; iii) the proposed MPED presents close performance under different scales and norms, which prove the robustness
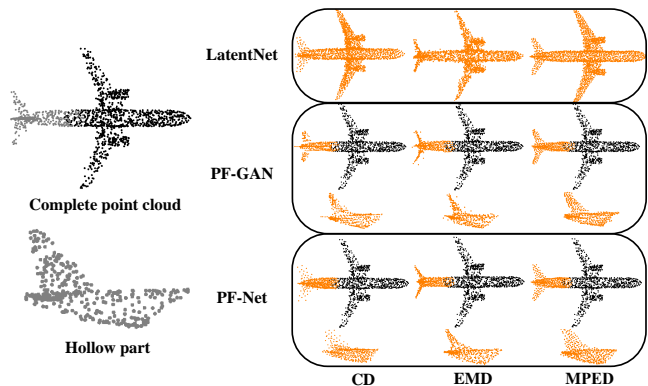
of the proposed model; iv)considering JSD as the third party metric that was not involved in training, the proposed MPED is more capable. When using **CD** as the supervision, it is less objective to prove its superiority by providing CD is best.

Fig. 6 illustrates several reconstruction results for three networks under various supervisions. MPED($p = 2$, Ψ_2) are set as representation. We see that: i) the results of **CD** present structure losing or blurred, especially finer parts, e.g., the leg of a chair; ii) **EMD** presents some "noise" points around sample surface, which means the coordinate of a single point is not accurate enough. iii) the proposed MPED can reconstruct each part of samples with high quality.

• 3D Shape Completion

Same to reconstruction, we use the ShapeNet dataset proposed in [8]. Four sub-categories from ShapeNet are used as the test sequences, e.g., "Chair", "Airplane", "Table" and "Car". The incomplete samples are generated via the method proposed by [11]. Specifically, for each sub-category, we set 75% as training set and 25% as testing set. We use LatentNet [8], PF-Net and PF-GAN [11] with default network parameter setting to test the empirical performances of various loss functions on the task of shape completion. Besides the individual performance, we also present average performance treating four categories as a whole. All the results are shown in Table VI. We see that: i) for three networks, **EMD** presents the best EMD, while other top performances are provided by MPED; ii) similar to point cloud reconstruction, the four versions of MPED show similar performances under most cases, which also proves the robustness of the proposed method.

The qualitative examples of shape completion are illustrated in Fig. 7. MPED($p = 1$, Ψ_2) is selected as the MPED representation. For LatentNet, the output is the whole sample, while PF-GAN and PF-Net only generate the missing part. We use black points to represent the input samples, grey points represent the hollow part (e.g., airplane tail), and orange points to highlight generated points. We see that: i) the tail of **CD** is quite blurred, while **EMD** is too tight with several scattered points far away from the reconstructed surface; and ii) the proposed MPED generates the most realistic hollow part.

• 3D Point Cloud Upsampling

We use PU-Net [41] to test the performance of three loss

Model	Loss	CD(×10e-4)					EMD(×10e1)					JSD(×10e-2)				
		Chair	Air	Table	Rifle	Ave	Chair	Air	Table	Rifle	Ave	Chair	Air	Table	Rifle	Ave
LatentNet	CD	6.84	2.35	6.47	2.82	4.62	4.88	2.28	5.78	4.68	4.41	3.13	1.84	3.35	3.21	2.88
	EMD	11.74	3.51	12.57	4.03	7.96	2.36	1.21	2.43	2.41	2.10	1.57	1.73	1.38	1.80	1.62
	MPED($p=1$, Ψ_1)	6.86	2.42	6.85	2.84	4.74	3.26	1.65	3.16	3.62	2.92	0.55	0.70	0.47	0.96	0.67
	MPED($p=1$, Ψ_2)	7.25	2.43	7.13	2.87	4.92	2.92	1.43	2.79	3.26	2.60	0.47	0.49	0.30	0.72	0.50
	MPED($p=2$, Ψ_1)	7.35	2.57	7.13	2.80	4.96	2.98	1.44	2.87	3.28	2.64	0.51	0.47	0.30	0.63	0.48
	MPED($p=2$, Ψ_2)	7.35	2.57	7.13	2.80	4.96	2.98	1.44	2.87	3.28	2.64	0.51	0.47	0.29	0.63	0.48
AtlasNet	CD	20.67	3.08	29.46	3.39	8.98	3.76	3.21	4.19	4.94	4.03	2.93	9.98	2.61	12.70	7.06
	EMD	22.74	3.59	8.99	6.89	10.55	2.79	0.73	1.65	2.90	2.02	3.18	2.44	0.82	6.12	3.14
	MPED($p=1$, Ψ_1)	13.84	3.04	7.70	4.81	7.35	2.22	0.96	2.19	1.93	1.83	0.92	0.98	0.78	1.35	1.01
	MPED($p=1$, Ψ_2)	11.12	4.25	8.01	6.26	7.41	1.98	0.96	1.79	1.78	1.63	0.95	1.67	0.66	1.46	1.19
	MPED($p=2$, Ψ_1)	15.25	2.82	7.76	2.99	7.21	2.48	0.89	2.05	1.80	1.81	1.55	1.37	0.67	1.06	1.16
	MPED($p=2$, Ψ_2)	8.18	2.97	8.20	3.73	5.77	1.69	0.79	1.90	1.98	1.59	0.69	1.51	0.62	1.69	1.13
FoldingNet++	CD	9.11	3.17	8.29	3.43	7.43	7.10	2.98	13.14	5.25	7.12	11.18	9.46	8.03	10.57	9.81
	EMD	10.97	4.67	11.86	6.61	11.49	2.00	1.23	2.01	2.67	1.98	1.92	4.37	1.98	5.13	3.35
	MPED($p=1$, Ψ_1)	6.74	2.33	6.36	2.65	4.52	2.68	1.52	2.31	2.68	2.30	0.94	1.06	0.40	1.08	1.37
	MPED($p=1$, Ψ_2)	6.65	2.33	6.26	2.62	4.47	2.35	1.11	2.01	2.76	2.06	0.58	0.95	0.42	0.92	0.72
	MPED($p=2$, Ψ_1)	7.52	2.33	6.36	2.65	4.72	2.92	1.52	2.31	2.68	2.36	1.24	1.06	0.40	1.08	0.95
	MPED($p=2$, Ψ_2)	6.79	2.48	7.02	3.03	4.83	2.43	1.16	2.19	2.95	2.18	0.66	0.82	0.62	1.67	0.94

TABLE V: Point cloud reconstruction on ShapeNet. The proposed MPED grants most of the top-level performance on different networks.

Model	Loss	CD(×10e-3)					EMD(×10e1)					JSD(×10e-1)				
		Chair	Air	Table	Car	Ave	Chair	Air	Table	Car	Ave	Chair	Air	Table	Car	Ave
LatentNet	CD	8.05	5.00	4.99	5.68	5.93	5.87	4.32	3.50	3.44	4.28	5.79	4.39	4.21	3.97	4.59
	EMD	5.04	2.58	5.01	5.41	4.51	1.44	1.17	1.69	1.53	1.46	5.22	3.78	4.16	3.50	4.17
	MPED($p=1$, Ψ_1)	3.93	2.43	4.05	4.33	3.69	1.89	1.88	2.07	1.60	1.86	4.73	3.49	3.64	3.20	3.77
	MPED($p=1$, Ψ_2)	3.90	2.35	4.26	4.33	3.71	1.67	1.65	1.91	1.47	1.68	4.72	3.49	3.68	3.16	3.76
	MPED($p=2$, Ψ_1)	3.86	2.33	3.91	4.33	3.61	1.87	1.69	2.19	1.52	1.82	4.95	3.63	3.84	3.20	3.91
	MPED($p=2$, Ψ_2)	3.82	2.16	3.95	4.16	3.52	1.60	1.40	1.88	1.34	1.58	4.82	3.51	3.80	3.18	3.83
PF-Net	CD	3.98	2.32	4.28	4.56	3.79	4.35	2.50	4.65	3.26	3.69	5.66	4.34	4.14	3.84	4.50
	EMD	5.57	3.22	5.57	5.57	4.98	1.40	1.08	1.59	1.47	1.39	5.35	4.04	4.29	3.42	4.28
	MPED($p=1$, Ψ_1)	3.72	2.17	4.17	4.32	3.60	1.53	1.21	1.83	1.40	1.49	4.62	3.34	3.56	3.10	3.66
	MPED($p=1$, Ψ_2)	3.96	2.27	4.51	4.49	3.81	1.83	1.50	2.24	1.43	1.75	4.75	3.50	3.64	3.19	3.77
	MPED($p=2$, Ψ_1)	4.89	3.03	5.31	5.02	4.56	1.78	1.47	2.13	1.55	1.73	5.16	4.04	3.98	3.40	4.15
	MPED($p=2$, Ψ_2)	5.44	3.51	5.97	5.40	5.08	2.29	1.77	3.20	1.77	2.26	5.39	4.35	4.23	3.49	4.37
PF-GAN	CD	4.80	2.60	4.99	5.03	4.36	2.15	1.60	2.60	1.68	2.01	5.22	4.10	4.13	3.35	4.20
	EMD	5.43	2.93	5.44	5.70	4.88	1.37	1.01	1.58	1.41	1.33	5.36	3.99	4.20	3.60	4.29
	MPED($p=1$, Ψ_1)	4.25	2.46	5.13	4.82	4.17	1.83	1.42	2.22	1.55	1.76	4.78	3.53	3.74	3.17	3.81
	MPED($p=1$, Ψ_2)	3.94	2.11	4.43	4.50	3.75	1.57	1.29	1.92	1.42	1.55	4.64	3.38	3.64	3.16	3.71
	MPED($p=2$, Ψ_1)	7.24	5.02	8.19	7.34	6.95	2.13	1.85	2.63	1.68	2.07	5.99	5.29	4.88	3.98	4.29
	MPED($p=2$, Ψ_2)	5.88	3.60	6.50	6.10	5.52	1.89	1.64	2.29	1.59	1.85	5.53	4.59	4.39	3.68	4.55

TABLE VI: Shape completion on ShapeNet. The proposed MPED grants most of the top-level performance on different networks.
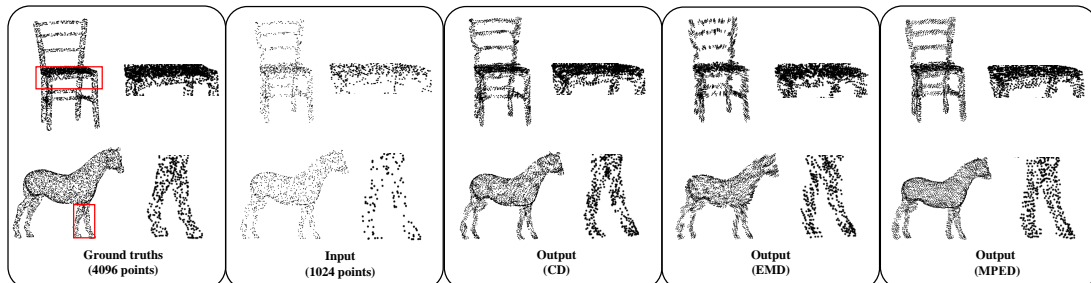

Fig. 8: Illustration of point cloud upsampling. The proposed MPED presents the best upsampling results.

Model	Ratio	Loss	Upsampling		
			CD	EMD	JSD
PU-Net	2	CD	6.50	5.93	10.12
		EMD	9.17	4.92	13.53
		MPED($p=1$, Ψ_1)	**5.82**	4.49	**8.36**
		MPED($p=1$, Ψ_2)	6.23	**4.41**	9.05
		MPED($p=2$, Ψ_1)	6.21	4.59	9.13
		MPED($p=2$, Ψ_2)	6.64	4.47	9.66
	4	CD	**2.78**	8.64	4.57
		EMD	5.56	8.26	7.85
		MPED($p=1$, Ψ_1)	2.89	7.97	**3.55**
		MPED($p=1$, Ψ_2)	3.63	7.85	4.41
		MPED($p=2$, Ψ_1)	3.03	7.90	3.76
		MPED($p=2$, Ψ_2)	3.48	**7.70**	4.19

TABLE VII: Upsampling results on PU-Net (CD ($\times$10e-4), EMD ($\times$10e1), and JSD ($\times$10e-2)). The proposed MPED presents most of the top-level performance.
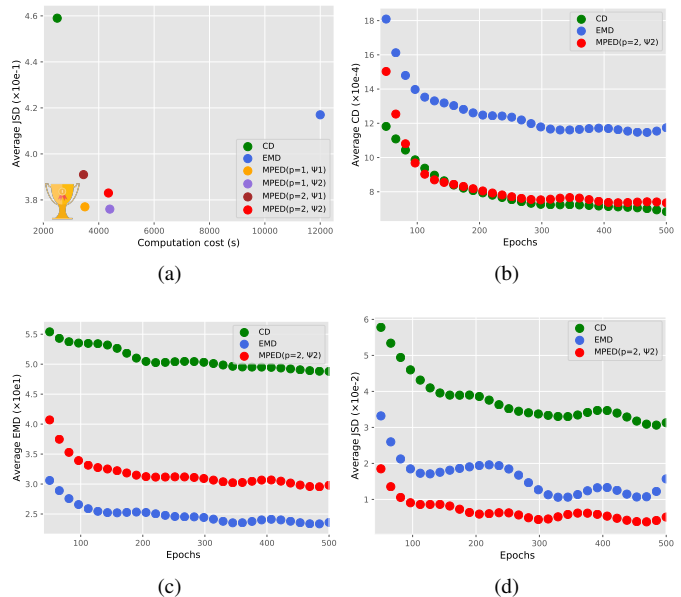

Fig. 9: (a) Performance vs. time. The proposed MPED presents best performance with low time consuming; (b)-(c): reconstruction loss under different training epochs (b) CD; (c) EMD; (d) JSD.

functions for the upsampling task. We vary the upsampling ratio as 2 and 4, respectively. And the we use the database provided by the [41], all the experiment parameters are default parameters refer to the implementation proposed by [42]. The results are shown in Table VII. Fig. 8 illustrates some upsampling examples under **ratio** = 4, and MPED($p = 1$, Ψ_1) is used as representation. We zoom in some details for better observation.

We see that: i) for both **ratio** = **2** and **4**, the MPED presents dominant performance; and ii) compared with MPED, **CD** presents more discontinuities, such as holes in horse's leg and chair's seat; **EMD** generates some uneven texture, which damages the viewing experience to some extent.

Ablation Study: Convergence Rate.

To illustrate the effectiveness of the proposed MPED, Fig. 9 (a) shows the tradeoff between performance and computational cost. We use LatentNet as an example. The y-axis shows the performance of JSD under different loss functions, and the x-axis shows the training time. We see that MPED achieves lower JSD while reducing the computational time by 3 times compared to **EMD**. Fig. 9 (b)-(d) show the reconstruction loss of CD, EMD and JSD under different training epochs.

For CD, e.g., Fig. 9(b), MPED presents obvious faster convergence rate in the early stage, after which it realizes close performance with **CD**; for EMD, e.g., Fig. 9(c), the variation of **EMD** and MPED curves show similar trends. They cost around 300 epochs to reach stable and best results; for JSD, e.g., Fig. 9(d), 200 epochs is enough for MPED to converge to stable results, while CD and EMD need 500 and 300 epochs, respectively.

VII. Conclusions

In this paper, we propose a universal point cloud distortion quantification named multiscale potential energy discrepancy (MPED). The proposed MPED is differentiable, has low computational complexity, and can discriminate distortions. MPED presents robust performances for both human perception and machine perception tasks. Specifically, for human perception tasks, MPED reveals the best performance on two fairly large databases, e.g., SJTU-PCQA and LSPCQA; for machine perception tasks, MPED shows better performance

than Chamfer distance and Earth mover's distance on three typical unsupervised tasks, e.g., point cloud reconstruction, shape completion and upsampling. We further demonstrate the robustness of MPED in ablation study, the results show that the proposed MPED exhibits stable and reliable performance in terms of various color spaces, and presents high optimization efficiency as loss function.

VIII. Acknowledgment

References

[1] S. Lim, M. Shin, and J. Paik, "Point cloud generation using deep local features for augmented and mixed reality contents," in *2020 IEEE Int. Conf. Consumer Electronics (ICCE'20)*, 2020, pp. 1–3.

[2] S. Chen, B. Liu, C. Feng, C. Vallespi-Gonzalez, and C. Wellington, "3D point cloud processing and learning for autonomous driving," *IEEE Signal Processing Magazine*, 2020.

[3] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (pcl)," in *2011 IEEE Int. Conf. Robotics and Automation (ICRA'11)*, 2011, pp. 1–4.

[4] C. Dore and M. Murphy, "Integration of historic building information modeling (hbim) and 3d gis for recording and managing cultural heritage sites," in *2012 18th Int. Conf. Virtual Systems and Multimedia (ICVSM'12)*, 2012, pp. 369–376.

[5] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

[6] A. Javaheri, C. Brites, F. M. B. Pereira, and J. M. Ascenso, "Point cloud rendering after coding: Impacts on subjective and objective quality," *IEEE Trans. Multimedia*, pp. 1–1, 2020.

[7] L. Zhang and Z. Zhu, "Unsupervised feature learning for point cloud understanding by contrasting and clustering using graph convolutional neural networks," in *2019 Int. Conf. 3D Vision (3DV'19)*, 2019, pp. 395–404.

[8] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas, "Learning representations and generative models for 3d point clouds," in *Int. conf. Machine Learning*. PMLR, 2018, pp. 40–49.

[9] S. Chen, C. Duan, Y. Yang, D. Li, C. Feng, and D. Tian, "Deep unsupervised learning of 3D point clouds via graph topology inference and filtering," *IEEE Trans. Image Processing*, vol. 29, pp. 3183–3198, 2019.

[10] M. Liu, L. Sheng, S. Yang, J. Shao, and S.-M. Hu, "Morphing and sampling network for dense point cloud completion," *arXiv preprint arXiv:1912.00280*, 2019.

[11] Z. Huang, Y. Yu, J. Xu, F. Ni, and X. Le, "Pf-net: Point fractal network for 3D point cloud completion," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR'20)*, 2020, pp. 7662–7670.

[12] G. Qian, A. Abualshour, G. Li, A. Thabet, and B. Ghanem, "Pu-gcn: Point cloud upsampling using graph convolutional networks," *arXiv preprint arXiv:1912.03264*, 2019.

[13] H. W.Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, pp. 83–97, 1955.

[14] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Applications of Digital Image Processing XLI*, vol. 10752. Int. Soc. Optics and Photonics, 2018, p. 107520I.

[15] J. Pang, D. Li, and D. Tian, "Tearingnet: Point cloud autoencoder to learn topology-friendly representations," *arXiv preprint arXiv:2006.10187*, 2020.

[16] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "Pcqm: A full-reference quality metric for colored 3D point clouds," in *12th Int. Conf. Quality of Multimedia Experience (QoMEX 2020)*, 2020.

[17] Y. Liu, Q. Yang, and Y. Xu, "Point cloud quality assessment: Large-scale dataset construction and learning-based no-reference approach," *arXiv preprint arXiv:2012.11895*, 2020.

[18] M. Balsera, S. Stepaniants, S. Izrailev, Y. Oono, and K. Schulten, "Reconstructing potential energy functions from simulated force-induced unbinding processes," *Biophysical journal*, vol. 73, no. 3, pp. 1281–1287, 1997.

[19] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The 37th IEEE Asilomar Conf. Signals, Systems & Computers (ACSS'2003)*, vol. 2, 2003, pp. 1398–1402.

[20] Q. Yang, Z. Ma, Y. Xu, L. Yang, W. Zhang, and J. Sun, "Modeling the screen content image quality via multiscale edge attention similarity," *IEEE Trans. Broadcasting*, vol. 66, no. 2, pp. 310–321, 2019.

[21] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun, "Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration," *IEEE Trans. Multimedia*, 2020.

[22] E. Alexiou, "On subjective and objective quality evaluation of point cloud geometry," in *9th IEEE Int. Conf. Quality of Multimedia Experience (QoMEX'17)*, 2017, pp. 1–3.

[23] 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, 2017, p. 103961H.

[24] L. A. da Silva Cruz, E. Dumić, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *11th IEEE Int. Conf. Quality of Multimedia Experience (QoMEX'19)*, 2019, pp. 1–6.

[25] E. Alexious, A. M. Pinheiro, C. Duarte, D. Matković, E. Dumić, L. A. da Silva Cruz, L. G. Dmitrović, M. V. Bernardo, M. Pereira, and T. Ebrahimi, "Point cloud subjective evaluation methodology based on reconstructed surfaces," in *Applications of Digital Image Processing XLI*, vol. 10752. Int. Soc. Optics and Photonics, 2018, p. 107520H.

[26] E. Alexiou, T. Ebrahimi, M. V. Bernardo, M. Pereira, A. Pinheiro, L. A. D. S. Cruz, C. Duarte, L. G. Dmitrovic, E. Dumic, D. Matkovics *et al.*, "Point cloud subjective evaluation methodology based on 2d rendering," in *10th IEEE Int. Conf. Quality of Multimedia Experience (QoMEX'18)*, 2018, pp. 1–6.

[27] E. Alexiou, "Exploiting user interactivity in quality assessment of point cloud imaging," in *11th IEEE Int. Conf. Quality of Multimedia Experience (QoMEX'19)*, 2019, pp. 1–6.

[28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[29] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *IEEE Int. Conf. Image Processing (ICIP'17)*, 2017, pp. 3460–3464.

[30] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "A papier-mâché approach to learning 3d surface generation," in *Pro. IEEE conf. Computer Vision and Pattern Recognition (CVPR'18)*, 2018, pp. 216–224.

[31] Mpeg people datasets. [Online]. Available: http://mpegfs.int-evry.fr/MPEG/PCC/DataSets/pointCloud/CfP/decoded/Dynamic_Objects/People/8i/

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[33] S. Chen, D. Tian, C. Feng, A. Vetro, and J. Kovačević, "Fast resampling of three-dimensional point clouds via graphs," *IEEE Trans. Signal Processing*, vol. 66, no. 3, pp. 666–681, 2017.

[34] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su *et al.*, "Shapenet: An information-rich 3D model repository," *arXiv preprint arXiv:1512.03012*, 2015.

[35] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D shapenets: A deep representation for volumetric shapes," in *Pro. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'15)*, 2015, pp. 1912–1920.

[36] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *[online]. Availabel: http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx*.

[37] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[38] J.-M. Geusebroek, R. Van Den Boomgaard, A. W. Smeulders, and A. Dev, "Color and scale: The spatial structure of color images," in *European Conf. Computer Vision (ECCV'00)*. Springer, 2000, pp. 331–341.

[39] N. Ravi, J. Reizenstein, D. Novotny, T. Gordon, W.-Y. Lo, J. Johnson, and G. Gkioxari, "Accelerating 3D deep learning with pytorch3D," *arXiv:2007.08501*, 2020.

[40] H. Fan, K. Mo, and J. Gu, "Pytorch wrapper for point-cloud earth-mover-distance (emd)," 2019, address: https://github.com/daerduoCarey/PyTorchEMD.

[41] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "Pu-net: Point cloud upsampling network," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'18)*, 2018, pp. 2790–2799.

[42] Pytorch unofficial implementation of punet and pugan. [Online]. Available: https://github.com/UncleMEDM/PUGAN-pytorch