Modeling Generalized Rate-Distortion Functions

Zhengfang Duanmu, *Student Member, IEEE,* Wentao Liu, *Member, IEEE,*
Zhuoran Li, *Student Member, IEEE,* and Zhou Wang, *Fellow, IEEE*

***Abstract*—Many multimedia applications require precise understanding of the rate-distortion characteristics measured by the function relating visual quality to media attributes, for which we term it the generalized rate-distortion (GRD) function. In this study, we explore the GRD behavior of compressed digital videos in a two-dimensional space of bitrate and resolution. Our analysis on a large-scale video dataset reveals that empirical parametric models are systematically biased while exhaustive search methods require excessive computation time to depict the GRD surfaces. By exploiting the properties that all GRD functions share, we develop an Robust Axial-Monotonic Clough-Tocher (RAMCT) interpolation method to model the GRD function. This model allows us to accurately reconstruct the complete GRD function of a source video content from a moderate number of measurements. To further reduce the computational cost, we present a novel sampling scheme based on a probabilistic model and an information measure. The proposed sampling method constructs a sequence of quality queries by minimizing the overall informativeness in the remaining samples. Experimental results show that the proposed algorithm significantly outperforms state-of-the-art approaches in accuracy and efficiency. Finally, we demonstrate the usage of the proposed model in three applications: rate-distortion curve prediction, per-title encoding profile generation, and video encoder comparison.**

***Index Terms*—Quality-of-experience (QoE); rate-distortion theory; content distribution; Clough-Toucher interpolation; quadratic programming; statistical sampling.**

I. Introduction

RATE-DISTORTION (RD) theory provides the theoretical foundations for lossy data compression and are widely employed in image and video compression schemes. One of the most profound outcomes from the theory is the so-called RD function [1], which describes the minimum bitrate required to encode a signal when a fixed amount of distortion is allowed (*i.e.*, the highest achievable quality given limited bitrate resources). Many multimedia applications require precise measurements of RD functions to characterize source signal and maximize user Quality-of-Experience (QoE). Examples of applications that explicitly use RD measurements are codec evaluation [2], RD optimization [3], video quality assessment (VQA) [4], encoding representation recommendation [5]–[8], and QoE optimization of streaming videos [9], [10].

Digital videos usually undergo a variety of transforms and processes in the content delivery chain, as shown in Fig. 1. To address the growing heterogeneity of display devices, contents, and access network capacity, source videos are encoded into different bitrates, spatial resolutions, frame rates, and bit depths before transmitted to the client. In an adaptive streaming video distribution environment [11], based on the bandwidth, buffering, and computation constraints, client devices adaptively select a proper video representation on a per-time segment basis to download and render. Each process influences the visual quality of a video in a different way, which can be jointly characterized by a generalized rate-distortion (GRD) function. In general, this attribute-distortion mapping comprises several complex factors, such as source content, operation mode/type of encoder, rendering system, and human visual system (HVS) characteristics. In this work, we assume the GRD surface is a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, where the input of the function is the video representation consisting of bitrate and spatial resolution, and the output of the function is the perceptual video quality under a specific viewing condition. Furthermore, the GRD function is content- and encoder-dependent.

Despite the tremendous growth in computational multimedia over the last few decades, estimating a GRD function is difficult, expensive, and time-consuming. Specifically, probing the quality of a single sample in the GRD space involves sophisticated video encoding and quality assessment, both of which are expensive processes. For example, the recently announced highly competitive AV1 video encoder [12] and video quality assessment model VMAF [13] could be over 100 times and 10 times slower than real-time for full high-definition (1920×1080) video content. Given the massive volume of multimedia data on the Internet, the real challenge is to produce an accurate estimate of the GRD function with a minimal number of samples.

We aim to develop a GRD function estimation framework with three desirable properties:

- Accuracy: It produces asymptotically unbiased estimation of GRD function, independent of the source video complexity and the encoder mechanism.
- Speed: It requires a minimal number of samples to reconstruct a full GRD function.
- Mathematical soundness: The GRD model has to be mathematically well-behaved, making it readily applicable to a variety of computational multimedia applications.

To achieve *accuracy*, we analyze the properties that all GRD functions share, based on which we formulate the GRD function approximation problem as a quadratic programming problem. The solution of the optimization problem provides an optimal interpolation model lying in the theoretical GRD function space. To achieve *speed*, we propose an efficient sampling algorithm that constructs a set of queries to maximize the expected information gain. The sampling scheme results in a unique sampling sequence invariant to source contents, enabling parallel encoding and quality assessment processes. To achieve *mathematical soundness*, the GRD

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: {zduanmu, w238liu, zhou.wang}@uwaterloo.ca).

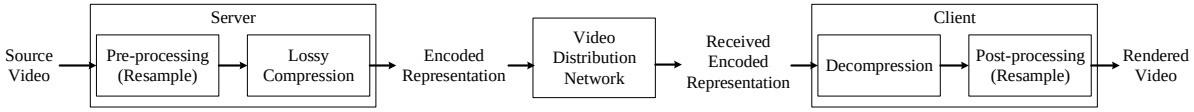
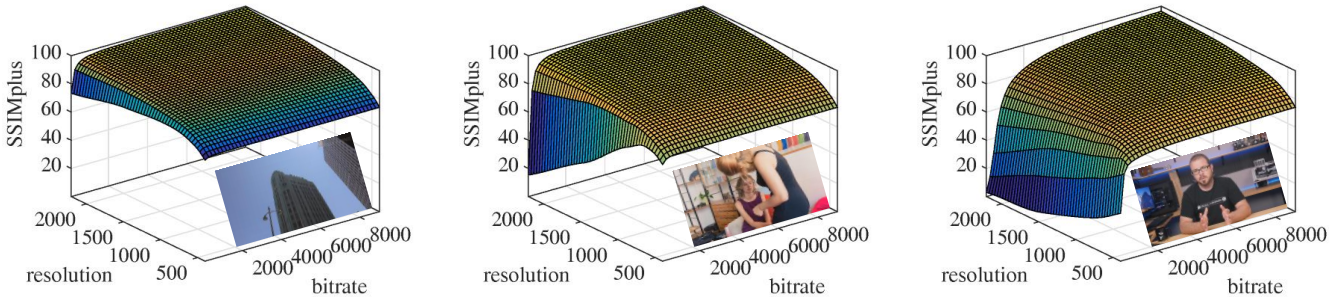


Fig. 1. Flow diagram of video delivery chain.



(a) (b) (c)

Fig. 2. Samples of generalized rate-distortion surfaces for different video content.

model is inherited from the Clough-Toucher (CT) interpolation method, and the function is differentiable everywhere on the domain of interests. Extensive experiments demonstrate that the resulting GRD function estimation framework achieves consistent improvement in speed and accuracy compared to the existing methods. The superiority and usefulness of the proposed algorithm are also evident by three applications.

The remainder of this paper is organized as follows. Section II reviews existing methods of estimating 1D RD or 2D GRD functions. The theoretical analysis of GRD functions and the proposed GRD model are introduced in Section III. The information-theoretic sampling method is elaborated in Section IV. Performances of the proposed GRD model and the sampling method are evaluated on a large-scale GRD function database in Section V, followed by three practical applications of the GRD model in Section VI. Finally, Section VII concludes the paper.

II. Related work

Although the RD theory has been successfully employed in many multimedia applications, the research in the practical GRD surface modeling has only become a scientific field of study in the past decade. Existing methods can be roughly categorized based on their assumptions about the shape of a GRD function. The first model class only makes weak assumptions about the properties of the GRD functions. For example, [7] assumes the continuity of GRD functions and applies linear interpolation to estimate the response function after densely sampling the video representation space. However, the exhaustive search process is computationally expensive, not to mention the number of samples required increases exponentially with respect to the dimension of input space.

By contrast, the second class of models make strong *a priori* assumptions about the form of the GRD function to alleviate the need of excessive training samples. For example, [4] assumes the video quality exhibits an exponential relationship with respect to the quantization step, spatial resolution, and frame rate. Alternatively, Toni *et al.* [6], [14] derived a reciprocal function to model the GRD function. Similarly, [8] models the rate-quality curve at each spatial resolution with a logarithmic function. A significant limitation of these models is that domains of the analytic functional forms are restricted only to the bitrate dimension and several discrete resolutions, lacking the flexibility to incorporate other dimensions such as frame rate and bit depth, and the capability to predict the GRD behaviors at novel resolutions.

In addition to the specific limitations the two kinds of models may respectively have, they suffer from the same problem that the training samples in the GRD space are either manually picked or randomly selected, neglecting the difference in the informativeness of samples. While many recent works acknowledge the importance of GRD function [5]–[8], a careful analysis and modeling of the response has yet to be done. We wish to address this void. In doing so, we seek a good compromise between 1) global and rigid models depending on random training samples and 2) local and indefinite models requiring exhaustive search in the video representation space.

III. Modeling Generalized Rate-Distortion Functions

We begin by stating our assumptions. Our first assumption is that the GRD function is smooth. In theory, the Shannon lower bound, the infimum of the required bitrate to achieve a certain quality, is guaranteed to be continuous with respect to the target distortion [15]. On the other hand, successive change in the spatial resolution would gradually deviate the frequency component and entropy of the source signal, resulting in smooth transition in the perceived quality. In practice, many subjective experiments have empirically shown the smoothness of GRD functions [4], [16]. We further assume that GRD functions are C^1 continuous, which is not only consistent with

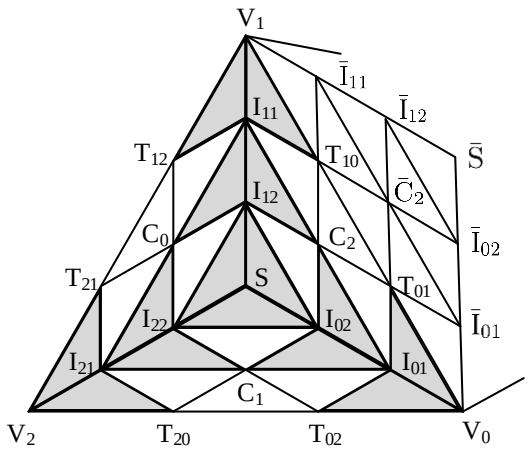


Fig. 3. Top view of one triangle of the triangulation, showing its three microtriangles, and the control net with 19 Bézier ordinates associated with the vertices and center points of microtriangles and the trisection points of the microtriangle edges.

most existing RD function models [4], [6], [8], [17], but also desired in many applications [6], [10]. One example of such applications is shown in Section VI-A.

Our second assumption is that the GRD function is axial-monotonic with respect to the bitrate[1]. This is a theoretical result from the RD theory [15], which assumes that the minimal possible rate is achieved for a given distortion constraint. It should be noted that there is a significant gap between theory and practice in video coding. There is no guarantee that a video encoder or a VQA model would not make "wrong" decisions that undermined the theoretical axial-monotonicity assumption. However, the monotonicity behavior is expected for any reasonably useful pair of encoder and quality measure, and should be generally true for state-of-the-art encoders and quality measures. We have empirically verified the assumption by observing all test cases collected in our proposed large-scale GRD function database.

It is worth mentioning that such monotonicity constraint may not apply to the spatial resolution. It has been demonstrated that encoding at high spatial resolution may even result in lower video quality than encoding at low spatial resolution under the same bitrate combined with upsampling and interpolation [7]. To be specific, encoding at high resolution with insufficient bitrate would produce artifacts such as blocking, ringing, and contouring, whereas encoding at low resolution with upsampling and interpolation would introduce blurring. The resulting distortions are further amplified or alleviated by the characteristics of the viewing device and viewing conditions, which interplay with HVS features such as the contrast sensitivity function [18]. A few sample GRD surfaces with their corresponding source videos are illustrated in Fig. 2.

Our third assumption is that the quality measurement is precise. Because the HVS is the ultimate receiver in most applications, subjective evaluation is a straightforward and reliable approach to evaluate the quality of digital videos. Traditional subjective experiment protocol models a subject's perceived quality as a random variable, assuming the quality labeling process to be stochastic. Because subjective experiment is expensive and time consuming, it is hardly used in the GRD function approximation process. In practice, objective VQA methods that produce deterministic quality predictions are often employed to generate ground truth samples in the GRD function. Therefore, a GRD function should pass through the quality scores of objective VQA evaluated on the encoded video representations.

Under these assumptions, we define the space of GRD functions as:

$$\mathcal{W}_{GRD} := \{f | f(x_n, y_n) = z_n, \forall n \in N, f \in C^1 : \mathbb{R}^2 \to \mathbb{R} \text{ and } \forall x_a < x_b, f(x_a, y) < f(x_b, y)\},$$

where N, x_n, y_n, and z_n represent the total number of training samples, bitrate, spatial resolution, and quality of the n-th training sample, respectively.

In the subsequent sections, we introduce the proposed GRD model. Section III-A and III-C review the traditional CT method and the monotonicity condition of cubic polynomial Bézier function, which the proposed model relies on. The proposed C^1 continuity condition, optimization framework, and robust axial-monotonic CT algorithm are novel contributions that are detailed in Section III-B, III-D, and III-E, respectively.

A. Review of Clough-Tocher Method

Since first introduced in 1960's [19], the CT method has been the most widely used multi-dimensional scattered data interpolant, thanks to its C^1 continuity and low computational complexity [20], [21]. Consider the scattered points (x_n, y_n) located in the x, y plane and their values z_n over the plane, the triangulation of the scattered points in the $x-y$ plane induces a piecewise triangular surface over the plane, whose nodes are the points (x_n, y_n, z_n). Fig. 3 conceptually illustrates one such triangle from the top view. In the CT method, each triangle is further divided from its center point S into three equivalent subtriangles, $\Delta_{V_0V_1S}$, $\Delta_{V_1V_2S}$, $\Delta_{V_2V_0S}$. Hereafter, we refer to the overall triangle as the macrotriangle and its subtriangles as microtriangles. The CT method estimates a cubic function in the form of Bézier surface on each microtriangle, so the whole CT interpolant is a piecewise cubic function. Mathematically, a cubic Bézier surface in $\Delta_{V_0V_1S}$ can be formulated as

$$\begin{aligned} z(\alpha,\beta,\gamma) =& c_{V_0}\alpha^3 + 3c_{T_{01}}\alpha^2\beta + 3c_{I_{01}}\alpha^2\gamma + c_{V_1}\beta^3 + \\ & 3c_{T_{10}}\alpha\beta^2 + 3c_{I_{11}}\beta^2\gamma + c_S\gamma^3 + 3c_{I_{02}}\alpha\gamma^2 + \\ & 3c_{I_{12}}\beta\gamma^2 + 6c_{C_2}\alpha\beta\gamma. \end{aligned} \quad (1)$$

From (1), we can see two major differences between a normal cubic function and a Bézier one. First, it represents a point by barycentric coordinates (α, β, γ) instead of Cartesian coordinates. Specifically, the barycentric coordinates of a point P with regard to $\Delta_{V_0V_1S}$ can be defined as

$$\alpha_P = \frac{A_{PV_1S}}{A_{V_0V_1S}}, \beta_P = \frac{A_{PSV_0}}{A_{V_1SV_0}}, \gamma_P = \frac{A_{PV_0V_1}}{A_{SV_0V_1}},$$

[1]In this work, we use RD function and rate-quality function interchangeably. Without loss of generality, we assume the function f to be monotonically increasing. If f is decreasing, we replace the given response with the function $f_{max} - f$, where f_{max} is the maximum value of quality.

where A_{UVW} means the directional area of the triangle formed by points U, V, W and is positive when U, V, W is counter-clockwise. The conversion from Cartesian coordinates to barycentric coordinates is lengthy and thus omitted here. Interested readers may refer to [20] for more details. Second, the 10 Bézier parameters in (1) can be associated with 10 specific points of the microtriangle $\Delta_{V_0V_1S}$. We have illustrated the parameter-point correspondence by the parameter subscripts in (1) and the marked points in Fig. 3. Specifically, the parameters c_{V_0}, c_{V_1}, c_{C_2} and c_S are associated with V_0, V_1, C_2 and S, respectively, and the remaining parameters are associated with the 6 trisection points on the three edges of $\Delta_{V_0V_1S}$. By building up the correspondence, the Bézier cubic function is controlled by a net of parameters, which is often referred to as the *control net*.

In a similar way, two more Bézier cubic functions are defined on the other two microtriangles, and thus we have to determine the 30 parameters only knowing the function values at three macrotriangle vertices. The CT method solves this highly underdetermined problem by effectively exploiting certain continuity constraints. Under the C^0 assumption within macrotriangles, each two Bézier surfaces should share their parameters at the their common boundaries V_0S, V_1S, and V_2S, leaving 19 free parameters in the macrotriangle $\Delta_{V_0V_1V_2}$. The inner-macrotriangle C^1 continuity removes 7 additional degree of freedoms by enforcing the shaded neighboring microtriangles in Fig. 3 to be coplanar [22]. To ensure inter-macrotriangle C^1 continuity, a standard approach is to assume the cross-boundary derivatives of the neighboring macrotriangles to be collinear, which further reduces the degree of freedom to 9. Taking into account the three known values at V_0, V_1, and V_2, we eventually have 6 unknown parameters in each macrotriangle. Although the gradients at vertices is not always available in practice, in most cases they can be estimated by considering the known values not only in the vertices of the triangle in question, but also in its neighbors. The most commonly used method is to estimate the gradients by minimizing the second-order derivatives along all Bézier curves [23]. Readers who are interested in the details of the CT method may refer to [20], [21], [23], [24].

The original CT method suffers from at least three limitations in approximating GRD functions. First, it uses the normal derivative of macrotriangle edges to guarantee inter-macrotriangle C^1 continuity. However, this choice gives an interpolant that is not invariant under affine transforms. This has some undesirable consequences: for a very narrow triangle, the spline can develop huge oscillations [24]. Second, the interpolant composite of piece-wise Bézier polynomials is not axial-monotonic, even when the given points are axial monotonic. Third, the CT algorithm achieves inter-macrotriangle C^1 continuity by imposing a linear assumption on normal derivatives at macrotriangle boundaries. Such an assumption is somewhat arbitrary and may violate monotonicity we want to achieve. We will address the three limitations in the subsequent sections.

B. Affine-Invariant C1 Continuity

In this section, we propose an affine invariant CT interpolant. For clarity and brevity, we would like to denote the macrotriangle edge that is opposite to the vertex $V_i, i = 0, 1, 2$ by E_i, and the internal microtriangle edge that connects V_i and S by $\hat{E}_i$. Instead of the normal derivative at the triangle boundary E_i, we consider $d^e_{E_i}$ to be parallel to $C_i\vec{\bar{C}}_i$, *i. e.*

$$c_{P_i} = (x_{P_i} - x_{V_i})d^x_{V_i} + (y_{P_i} - y_{V_i})d^y_{V_i} + z_{V_i} \tag{2a}$$

$$c_{C_i} = \theta_{kj}c_{T_{jk}} + \theta_{jk}c_{T_{kj}} + \eta_i d^e_{E_i} \tag{2b}$$

$$\begin{aligned} c_{I_{i2}} =& \frac{1}{3}[(x_{I_{i1}} - x_{V_i}) + (x_{T_{ki}} - x^*_j) + (x_{T_{ji}} - x^*_k)]d^x_{V_i} + \\ & \frac{1}{3}[(y_{I_{i1}} - y_{V_i}) + (y_{T_{ki}} - y^*_j) + (y_{T_{ji}} - y^*_k)]d^y_{V_i} + \\ & \frac{1}{3}(x_{T_{ij}} - x^*_k)d^x_{V_j} + \frac{1}{3}(y_{T_{ij}} - y^*_k)d^y_{V_j} + \frac{1}{3}\eta_k d^e_{E_k} + \\ & \frac{1}{3}(x_{T_{ik}} - x^*_j)d^x_{V_k} + \frac{1}{3}(y_{T_{ik}} - y^*_j)d^y_{V_k} + \frac{1}{3}\eta_j d^e_{E_j} + \\ & \frac{1}{3}[z_{V_i} + (\theta_{ki}z_{V_i} + \theta_{ik}z_{V_k}) + (\theta_{ij}z_{V_j} + \theta_{ji}z_{V_i})] \end{aligned} \tag{2c}$$

$$\begin{aligned} c_S =& \frac{1}{9}\sum_{i=0}^{2}[(x_{I_{i1}} - x_{V_i}) + 2(x_{T_{ki}} - x^*_j) + 2(x_{T_{ji}} - x^*_k)]d^x_{V_i} + \\ & \frac{1}{9}\sum_{i=0}^{2}[(y_{I_{i1}} - y_{V_i}) + 2(y_{T_{ki}} - y^*_j) + 2(y_{T_{ji}} - y^*_k)]d^y_{V_i} + \\ & \frac{2}{9}\sum_{i=0}^{2}\eta_i d^e_{E_i} + \frac{1}{9}\sum_{i=0}^{2}[(1 + 2\theta_{ji} + 2\theta_{ki})z_{V_i}], \end{aligned} \tag{2d}$$

where

$$P_i \in \{T_{ij}, T_{ik}, I_{i1}\},$$

$$x^*_i = \frac{(x_{\bar{C}_i} - x_{C_i})(x_{V_j}y_{V_k} - x_{V_k}y_{V_j}) - (x_{V_k} - x_{V_j})(x_{C_i}y_{\bar{C}_i} - x_{\bar{C}_i}y_{C_i})}{(x_{\bar{C}_i} - x_{C_i})(y_{V_k} - y_{V_j}) - (y_{\bar{C}_i} - y_{C_i})(x_{V_k} - x_{V_j})},$$

$$y^*_i = \frac{(y_{\bar{C}_i} - y_{C_i})(x_{V_j}y_{V_k} - x_{V_k}y_{V_j}) - (y_{V_k} - y_{V_j})(x_{C_i}y_{\bar{C}_i} - x_{\bar{C}_i}y_{C_i})}{(x_{\bar{C}_i} - x_{C_i})(y_{V_k} - y_{V_j}) - (y_{\bar{C}_i} - y_{C_i})(x_{V_k} - x_{V_j})},$$

$$\eta_i = \sqrt{(x_{C_i} - x^*_i)^2 + (y_{C_i} - y^*_i)^2},$$

$$\theta_{kj} = \frac{x_{T_{kj}} - x^*_i}{x_{T_{kj}} - x_{T_{jk}}},$$

$$\theta_{jk} = \frac{x_{T_{jk}} - x^*_i}{x_{T_{jk}} - x_{T_{kj}}},$$

$d^x_{V_i}$ and $d^y_{V_i}$ are partial derivatives of the Bézier surface at V_i and $\{i, j, k\}$ is a cyclic permutation of $\{0, 1, 2\}$. Since this quantity transforms similarly as the gradient under affine transforms, the resulting interpolant is affine-invariant [24].

We also lift the unwanted linear constraints on the cross-boundary derivatives, elevating the number of parameters in a macrotriangle back to 9. In summary, the equality constraints in (2) can be factorized into the matrix form for simplicity

$$\mathbf{c} = \mathbf{R}\mathbf{d} + \mathbf{f}, \tag{3}$$

where $\mathbf{c} \in \mathbb{R}^{16\times1}$, $\mathbf{R} \in \mathbb{R}^{16\times9}$, $\mathbf{d} \in \mathbb{R}^{9\times1}$, $\mathbf{f} \in \mathbb{R}^{16\times1}$, $\mathbf{c}$ and $\mathbf{d}$ represent the values of control net and unknown derivatives, respectively. Therefore, finding the interpolant of the macrotriangle corresponds to determining the 9 unknown parameters in $\mathbf{d}$.

Besides the inner macrotriangle constraints, we also want to keep $d^e_{E_i}$ consistent across the triangle boundary to ensure external C^1 smoothness. As a result, the following equality constraints need to be added for each edge with adjacent triangles

$$d^e_{E_i} + d^e_{\bar{E}_i} = 0. \tag{4}$$

Combining (3) and (4), we conclude that the resulting function is C^1 continuous and affine-invariant.

C. Axial Monotonicity

This section aims to derive the sufficient constraints on $\mathbf{d}$ for the Bézier surface in the macrotriangle $\Delta_{V_0V_1V_2}$ to be axial-monotonic. In general, the interpolant composite of piece-wise Bézier polynomials is not monotonic even though the sampled points are monotonic. Several works have been done to derive sufficient conditions for a univariate or bivariate Bézier function [25], [26]. We adopt the sufficient condition proposed in [26], where it was proved that the cubic Bézier surface in a microtriangle is axial-monotonic when all the 6 triangular patches of its control net (*e.g.* $\Delta_{I_{02}I_{12}S}$, $\Delta_{C_2I_{11}I_{12}}$, $\Delta_{C_2I_{01}I_{02}}$, $\Delta_{V_1T_{10}I_{11}}$, $\Delta_{T_{10}T_{01}C_2}$, and $\Delta_{T_{01}V_0I_{01}}$ in $\Delta_{V_0V_1S}$) are axial-monotonic. By combining the sufficient conditions in all three microtriangles and the inner triangle continuity, we obtain

$$(y_{V_i} - y_{V_k})c_{T_{ij}} + (y_{V_j} - y_{V_i})c_{T_{ik}} \le (y_{V_j} - y_{V_k})z_{V_i} \tag{5a}$$
$$(y_{V_k} - y_{V_j})c_{I_{i1}} + (y_{V_i} - y_{V_k})c_{C_k} + (y_{V_j} - y_{V_i})c_{C_j} \le 0 \tag{5b}$$
$$(y_{V_2} - y_{V_1})c_{I_{02}} + (y_{V_0} - y_{V_2})c_{I_{12}} + (y_{V_1} - y_{V_0})c_{I_{22}} \le 0 \tag{5c}$$
$$(y_S - y_{V_j})c_{T_{ij}} + (y_{V_i} - y_S)c_{T_{ji}} + (y_{V_j} - y_{V_i})c_{C_k} \le 0. \tag{5d}$$

We can summarize the monotonicity constraint in matrix form

$$\mathbf{Gc} \le \mathbf{h}, \tag{6}$$

where $\mathbf{G} \in \mathbb{R}^{10\times 16}$ and $\mathbf{f} \in \mathbb{R}^{10\times 1}$. Further substitute (3) into (6), we obtain the monotonicity constraint in terms of $\mathbf{d}$

$$\mathbf{GRd} \le \mathbf{h} - \mathbf{Gf}. \tag{7}$$

More details on how we construct $\mathbf{G}$ and $\mathbf{h}$ are given in the Appendix.

D. Optimization-based Solutions

To determine the unknown derivatives, we propose to minimize the total curvature of the interpolated surface under the smoothness assumption. Directly computing the total curvature is computationally intractable. Alternatively, we minimize the curvature of Bézier curves at the edges of each microtriangle as its approximation. Specifically, in $\Delta_{V_0V_1V_2}$, the objective function is written as

$$L_{V_0V_1V_2} = \tfrac{1}{2}\sum_{i=0}^{2}\int_{E_i}\left[\tfrac{\partial^2 z}{\partial E_i^2}\right]^2 ds_{E_i} + \sum_{i=0}^{2}\int_{\hat{E}_i}\left[\tfrac{\partial^2 z}{\partial \hat{E}_i^2}\right]^2 ds_{\hat{E}_i}, \tag{8}$$

where the weight $\frac{1}{2}$ is introduced to cancel the double counting of the external edges.

Consider an external boundary E_i, whose Bézier control net coefficients are z_{V_j}, $c_{T_{jk}}$, $c_{T_{kj}}$, and z_{V_k}. The integral of the second order derivative of the Bézier curve on E_i can be represented in terms of the four coefficients as

$$\begin{aligned}
&\int_{E_i}\left[\frac{\partial^2 z}{\partial E_i^2}\right]^2 ds_{E_i} = \frac{1}{\|E_i\|^3}\int_0^1\left[z''_{E_i}(t)\right]^2 dt\\
=&\frac{18}{\|E_i\|^3}(2c^2_{T_jk} + 2c^2_{T_kj} - 2c_{T_jk}c_{T_kj}) +\\
&\frac{-36}{\|E_i\|^3}(z_{V_j}c_{T_jk} + z_{V_k}c_{T_kj}) + \frac{12}{\|E_i\|^3}(z^2_{V_j} + z^2_{V_k} + z_{V_j}z_{V_k})\\
=&\begin{bmatrix} c_{T_{jk}} & c_{T_{kj}}\end{bmatrix}\begin{bmatrix}\frac{36}{\|E_i\|^3} & \frac{-18}{\|E_i\|^3}\\ \frac{-18}{\|E_i\|^3} & \frac{36}{\|E_i\|^3}\end{bmatrix}\begin{bmatrix} c_{T_{jk}}\\ c_{T_{kj}}\end{bmatrix} +\\
&\begin{bmatrix}\frac{-36z_{V_j}}{\|E_i\|^3} & \frac{-36z_{V_k}}{\|E_i\|^3}\end{bmatrix}\begin{bmatrix} c_{T_{jk}}\\ c_{T_{kj}}\end{bmatrix} +\\
&\frac{12}{\|E_i\|^3}(z^2_{V_j} + z^2_{V_k} + z_{V_j}z_{V_k}),
\end{aligned} \tag{9}$$

where

$$\|E_i\| = \sqrt{(x_{V_j} - x_{V_k})^2 + (y_{V_j} - y_{V_k})^2}$$

is the length of E_i.

Similarly, we get the other part of the objective function from an internal boundary $\hat{E}_i$, whose coefficients are $z_{V_i}, c_{I_{i1}}, c_{I_{i2}}$, and c_S.

$$\begin{aligned}
&\int_{\hat{E}_i}\left[\frac{\partial^2 z}{\partial \hat{E}_i^2}\right]^2 ds_{\hat{E}_i} = \frac{1}{\|\hat{E}_i\|^3}\int_0^1\left[z''_{\hat{E}_i}(t)\right]^2 dt\\
=&\frac{6}{\|\hat{E}_i\|^3}(6c^2_{I_{i1}} + 6c^2_{I_{i2}} + 2c^2_S - 6c_{I_{i1}}c_{I_{i2}} - 6c_{I_{i2}}c_S) +\\
&\frac{12z_{V_i}}{\|\hat{E}_i\|^3}(-3c_{I_{i1}} + c_S) + \frac{12}{\|\hat{E}_i\|^3}z^2_{V_i}\\
=&\begin{bmatrix} c_{I_{i1}} & c_{I_{i2}} & c_S\end{bmatrix}\begin{bmatrix}\frac{36}{\|\hat{E}_i\|^3} & \frac{-18}{\|\hat{E}_i\|^3} & 0\\ \frac{-18}{\|\hat{E}_i\|^3} & \frac{36}{\|\hat{E}_i\|^3} & \frac{-18}{\|\hat{E}_i\|^3}\\ 0 & \frac{-18}{\|\hat{E}_i\|^3} & \frac{12}{\|\hat{E}_i\|^3}\end{bmatrix}\begin{bmatrix} c_{I_{i1}}\\ c_{I_{i2}}\\ c_S\end{bmatrix}\\
&+\begin{bmatrix}\frac{-36z_{V_i}}{\|\hat{E}_i\|^3} & 0 & \frac{12z_{V_i}}{\|\hat{E}_i\|^3}\end{bmatrix}\begin{bmatrix} c_{I_{i1}}\\ c_{I_{i2}}\\ c_S\end{bmatrix} + \frac{12z^2_{V_i}}{\left\|\hat{E}_i\right\|^3},
\end{aligned} \tag{10}$$

where

$$\|\hat{E}_i\| = \sqrt{(x_S - x_{V_i})^2 + (y_S - y_{V_i})^2}$$

is the length of $\hat{E}_i$.

Substitute (9) and (10) into (8), we obtain the loss function for $\Delta_{V_0V_1V_2}$ in matrix form

$$L_{V_0V_1V_2} = \mathbf{c}^T\mathbf{U}_{V_0V_1V_2}\mathbf{c} + \mathbf{w}^T_{V_0V_1V_2}\mathbf{c} + const, \tag{11}$$

where $\mathbf{U}_{V_0V_1V_2} \in \mathbb{R}^{16\times 16}$ and $\mathbf{w}_{V_0V_1V_2} \in \mathbb{R}^{16\times 1}$.

Substituting $\mathbf{c} = \mathbf{Rd} + \mathbf{f}$ into (11), we get

$$\begin{aligned}
L_{V_0V_1V_2} =&(\mathbf{Rd} + \mathbf{f})^T\mathbf{U}_{V_0V_1V_2}(\mathbf{Rd} + \mathbf{f}) +\\
&\mathbf{w}^T_{V_0V_1V_2}(\mathbf{Rd} + \mathbf{f}) + const\\
=&\mathbf{d}^T(\mathbf{R}^T\mathbf{U}_{V_0V_1V_2}\mathbf{R})\mathbf{d} +\\
&(\mathbf{f}^T\mathbf{U}_{V_0V_1V_2} + \mathbf{w}^T_{V_0V_1V_2})\mathbf{Rd} + const.
\end{aligned} \tag{12}$$

In summary, finding the axial-monotonic interpolant corresponds to solving the following optimization problem

$$\begin{aligned} \underset{\mathbf{d}}{\text{minimize}} \quad & \mathbf{d}^T(\mathbf{R}^T\mathbf{U}_{V_0V_1V_2}\mathbf{R})\mathbf{d} + (\mathbf{f}^T\mathbf{U}_{V_0V_1V_2} + \mathbf{w}_{V_0V_1V_2}^T)\mathbf{R}\mathbf{d} \\ \text{subject to} \quad & \mathbf{G}\mathbf{R}\mathbf{d} \le \mathbf{h} - \mathbf{G}\mathbf{f}, \\ & d_{E_i}^e + d_{\bar{E}_i}^e = 0. \end{aligned} \tag{13}$$

Note that the constraints are linear with respect to $\mathbf{d}$ and $\mathbf{R}^T\mathbf{U}_{V_0V_1V_2}\mathbf{R}$ is positive-semidefinite. Thus, finding $\mathbf{d}$ turns into a standard problem of quadratic programming, which can be efficiently solved by the existing convex programming packages [27].

E. Robust Axial-Monotonic Clough-Tocher Method

Here we propose our Robust Axial-Monotonic Clough-Tocher (RAMCT) method. The inequality constraints in (6) are sufficient conditions for x-axial monotonicity. However, the sufficient conditions excessively shrink the solution space in some extreme cases, making the primary solution infeasible. To relax these constraints, we introduce hinge loss to some of these inequalities, motivated by the success of Support Vector Machine [28]. Specifically, the modified inequality constraints are formulated as

$$(y_{V_i} - y_{V_k})c_{T_{ij}} + (y_{V_j} - y_{V_i})c_{T_{ik}} \le (y_{V_j} - y_{V_k})z_{V_i} \tag{14a}$$
$$(y_{V_k} - y_{V_j})c_{I_{i1}} + (y_{V_i} - y_{V_k})c_{C_k} + (y_{V_j} - y_{V_i})c_{C_j} + \xi_{i1} \le 0 \tag{14b}$$
$$(y_{V_2} - y_{V_1})c_{I_{02}} + (y_{V_0} - y_{V_2})c_{I_{12}} + (y_{V_1} - y_{V_0})c_{I_{22}} \le 0 \tag{14c}$$
$$(y_S - y_{V_j})c_{T_{ij}} + (y_{V_i} - y_S)c_{T_{ji}} + (y_{V_j} - y_{V_i})c_{C_k} + \xi_{C_k} \le 0, \tag{14d}$$

where $\boldsymbol{\xi} = [\xi_{11}, \xi_{21}, \xi_{31}, \xi_{C_1}, \xi_{C_2}, \xi_{C_3}]$ are auxiliary variables and $\boldsymbol{\xi} \le \mathbf{0}$. Note that (14a),(14c) are identical to (5a),(5c) because they are necessary conditions of axial monotonicity (See Appendix for proof). Rewriting these constraints in the matrix form, we obtain

$$\begin{bmatrix} \mathbf{G} & \mathbf{J}_1 \\ \mathbf{O} & \mathbf{J_2} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \boldsymbol{\xi} \end{bmatrix} \le \begin{bmatrix} \mathbf{h} \\ \mathbf{0} \end{bmatrix},$$

where $\mathbf{G}$ and $\mathbf{h}$ are the same as in (6),(7). $\mathbf{J}_2$ is a 6×6 identity matrix, while $\mathbf{J}_1 \in \mathbb{R}^{10\times6}$ is obtained by padding $\mathbf{J}_2$ with 3 rows of zeros to its top and inserting a row of zeros between the 3rd and 4th rows of $\mathbf{J}_2$.

By substituting (3) into the inequality above, we finally obtain the inequality constraints in terms of the unknowns $\mathbf{d}$ and the auxiliary variables $\boldsymbol{\xi}$ as

$$\begin{bmatrix} \mathbf{G}\mathbf{R} & \mathbf{J}_1 \\ \mathbf{O} & \mathbf{J}_2 \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \boldsymbol{\xi} \end{bmatrix} \le \begin{bmatrix} \mathbf{h} - \mathbf{G}\mathbf{f} \\ \mathbf{0} \end{bmatrix}. \tag{15}$$

The objective function is then modified accordingly,

$$L_{V_0V_1V_2} = \mathbf{c}^T\mathbf{U}_{V_0V_1V_2}\mathbf{c} + \mathbf{w}_{V_0V_1V_2}^T\mathbf{c} - \boldsymbol{\lambda}^T\boldsymbol{\xi} + const, \tag{16}$$

where $\boldsymbol{\lambda} = [\lambda, \lambda, \cdots, \lambda]^T$ is the weighting parameter. Substituting $\mathbf{c} = \mathbf{R}\mathbf{d} + \mathbf{f}$ into (16), we get

Algorithm 1: Uncertainty Sampling

Initialize $S = \emptyset$; $\bar{\boldsymbol{\Sigma}}^{(1)} = \boldsymbol{\Sigma}$;
for $k := 1$ *to* K **do**
  $i^{(k)} = \underset{i}{\text{minimize}} \quad \text{tr}(\bar{\boldsymbol{\Sigma}}_{ii}^{(k)} - \frac{\bar{\boldsymbol{\sigma}}_i^{(k)T}\bar{\boldsymbol{\sigma}}_i^{(k)}}{\bar{\sigma}_{ii}^{(k)}})$;
  $x^{(k)}$ = VQA(Encode($\mathbf{r}_i^{(k)}$));
  Set $S = S \cup x^{(k)}$;
  $\bar{\boldsymbol{\Sigma}}^{(k+1)} = \bar{\boldsymbol{\Sigma}}_{ii}^{(k)} - \frac{\bar{\boldsymbol{\sigma}}_i^{(k)T}\bar{\boldsymbol{\sigma}}_i^{(k)}}{\bar{\sigma}_{ii}^{(k)}}$;
  if $\text{tr}(\bar{\boldsymbol{\Sigma}}_{ii}^{(k)} - \frac{\bar{\boldsymbol{\sigma}}_i^{(k)T}\bar{\boldsymbol{\sigma}}_i^{(k)}}{\bar{\sigma}_{ii}}) \le T$ **then**
    Break ;
  end
end

$$\begin{aligned} & L_{V_0V_1V_2} = \mathbf{d}^T(\mathbf{R}^T\mathbf{U}_{V_0V_1V_2}\mathbf{R})\mathbf{d} + (\mathbf{f}^T\mathbf{U}_{V_0V_1V_2} + \\ & \mathbf{w}_{V_0V_1V_2}^T)\mathbf{R}\mathbf{d} - \boldsymbol{\lambda}^T\boldsymbol{\xi} + const \\ & = \begin{bmatrix} \mathbf{d}^T & \boldsymbol{\xi}^T \end{bmatrix} \begin{bmatrix} \mathbf{R}^T\mathbf{U}_{V_0V_1V_2}\mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \boldsymbol{\xi} \end{bmatrix} + \\ & \begin{bmatrix} (\mathbf{f}^T\mathbf{U}_{V_0V_1V_2} + \mathbf{w}_{V_0V_1V_2}^T)\mathbf{R} & -\boldsymbol{\lambda}^T \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \boldsymbol{\xi} \end{bmatrix} + const. \end{aligned} \tag{17}$$

Replacing (12),(6) with (17),(15) in (13), we find that the original interpolation problem remains to be a quadratic programming problem.

IV. Information-theoretic sampling

In this section, we first explore the informativeness of samples in the GRD space via a probabilistic model. We then present an information-theoretic sampling strategy that optimally selects the samples, offering enormous savings in time and computational resources.

Let $\mathbf{x} = (x_1, ..., x_N)$ be a vector of discrete samples on a GRD function uniformly distributed in the bitrate-resolution space, where N is the total number of sample points on the grid. Given that the GRD function is smooth, when the sampling grid is dense, these discrete samples provide a good description of the continuous GRD function. In particular, when the GRD function is band-limited, it can be fully recovered from these samples when the sampling density is larger than the Nyquist rate. Assuming $\mathbf{x}$ is created from GRD functions of real-world video content, we model $\mathbf{x}$ as an N-dimensional random variable, for which the probability density function $p_{\mathbf{x}}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ follows a multivariate Normal distribution. The total uncertainty of $\mathbf{x}$ is characterized by its joint entropy given by

$$H_{\mathbf{x}}(\mathbf{x}) = \frac{1}{2}\log|\boldsymbol{\Sigma}| + const, \tag{18}$$

where $|\cdot|$ is the determinant operator. If the full vector $\mathbf{x}$ is further divided into two parts such that $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, and the $\mathbf{x}_2$ portion has been resolved by $\mathbf{x}_2 =$

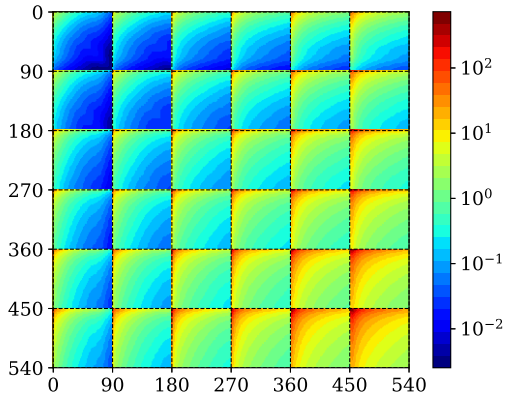


Fig. 4. Empirical covariance matrix of the GRD functions. Bitrate and spatial resolution are presented in ascending order, and spatial resolution elevates every 90 samples.

$\mathbf{a}$, then the remaining uncertainty is given by the conditional entropy

$$H_{\mathbf{x}_1|\mathbf{x}_2}(\mathbf{x}_1|\mathbf{x}_2=\mathbf{a})=\frac{1}{2}\log|\bar{\boldsymbol{\Sigma}}|)+const, \tag{19}$$

where

$$\bar{\boldsymbol{\Sigma}}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{20}$$

As a special case, we aim to find one sample that most efficiently reduces the uncertainty of GRD estimation. This is found by minimizing the log determinant of the conditional covariance matrix [29]

$$\underset{i}{\text{minimize}} \quad \log|\bar{\boldsymbol{\Sigma}}| = \underset{i}{\text{minimize}} \quad \log|\bar{\boldsymbol{\Sigma}}_{ii}-\frac{\bar{\boldsymbol{\sigma}}_i^T\bar{\boldsymbol{\sigma}}_i}{\bar{\sigma}_{ii}}|, \tag{21}$$

where $\bar{\boldsymbol{\Sigma}}=\begin{bmatrix}\bar{\boldsymbol{\Sigma}}_{ii} & \bar{\boldsymbol{\sigma}}_i\\ \bar{\boldsymbol{\sigma}}_i^T & \bar{\sigma}_{ii}\end{bmatrix}$ and i is the row index of $\bar{\boldsymbol{\Sigma}}$.

Minimizing (21) directly is computationally expensive, especially when the dimensionality is high. Alternatively, we minimize the upper bound of the conditional entropy

$$\underset{i}{\text{minimize}} \quad \text{tr}(\bar{\boldsymbol{\Sigma}}_{ii}-\frac{\bar{\boldsymbol{\sigma}}_i^T\bar{\boldsymbol{\sigma}}_i}{\bar{\sigma}_{ii}}), \tag{22}$$

where $\log|\bar{\boldsymbol{\Sigma}}_{ii}-\frac{\bar{\boldsymbol{\sigma}}_i^T\bar{\boldsymbol{\sigma}}_i}{\bar{\sigma}_{ii}}|\leq\text{tr}(\bar{\boldsymbol{\Sigma}}_{ii}-\frac{\bar{\boldsymbol{\sigma}}_i^T\bar{\boldsymbol{\sigma}}_i}{\bar{\sigma}_{ii}}-\mathbf{I})$ and $\mathbf{I}$ denotes identity matrix. The sample with the minimum average loss in (22) over all viewing devices is most informative. Once the optimal sample index is obtained, we encode the video at the i-th representation, evaluate its quality with objective VQA algorithms, and update the conditional covariance matrix in (20). The process is applied iteratively until the overall uncertainty in the system is reduced below a certain threshold T. We summarize the proposed uncertainty sampling method in Algorithm 1, where $\mathbf{r}_i$ represents the bitrate and spatial resolution at the i-th representation.

Remark: To get a sense of what type of samples will be chosen by the proposed algorithm, we analyze several influencing factors in the objective function (22):

- By the basic properties of trace, the objective function in the uncertainty sampling can be factorized as

$$\begin{aligned}&\text{tr}(\bar{\boldsymbol{\Sigma}}_{ii}-\frac{\bar{\boldsymbol{\sigma}}_i^T\bar{\boldsymbol{\sigma}}_i}{\bar{\sigma}_{ii}})\\ =&\text{tr}(\bar{\boldsymbol{\Sigma}}_{ii})-\frac{\text{tr}(\bar{\boldsymbol{\sigma}}_i^T\bar{\boldsymbol{\sigma}}_i)}{\bar{\sigma}_{ii}}\\ =&\text{tr}(\bar{\boldsymbol{\Sigma}})-(\bar{\sigma}_{ii}+\frac{1}{\bar{\sigma}_{ii}}\sum_{j\neq i}\bar{\sigma}_{ij}^2).\end{aligned}$$

  Thus, $\text{tr}(\bar{\boldsymbol{\Sigma}}_{ii}-\frac{\bar{\boldsymbol{\sigma}}_i^T\bar{\boldsymbol{\sigma}}_i}{\bar{\sigma}_{ii}})$ is a decreasing function with respect to $\bar{\sigma}_{ii}$ when $\bar{\sigma}_{ii}>\sqrt{\sum_{j\neq i}\bar{\sigma}_{ij}^2}$. This indicates that samples with large uncertainty are more likely to be selected than those with small uncertainty.
- According to (20), $\forall j\neq i$,

$$\bar{\sigma}_{jj}^{(k+1)}=\bar{\sigma}_{jj}^{(k)}-\frac{\bar{\sigma}_{ij}^{(k)^2}}{\bar{\sigma}_{ii}^{(k)}},$$

  suggesting the rate of reduction in the uncertainty of sample j is proportional to its squared correlation with the selected sample i in the k-th iteration. Fig. 4 shows an empirical covariance matrix $\bar{\boldsymbol{\Sigma}}$ estimated from our video dataset that will be detailed in the next section, from which we observe that the GRD functions typically exhibit high correlation in a local region. Combining the first observation above, we conclude that the next optimal choice of sample should be selected from the region where labeled samples are sparse.
- Note that knowing that $\mathbf{x}_2=\mathbf{a}$ alters the variance, though the new variance does not depend on the specific value of $\mathbf{a}$. The independence has two important consequences. First, the proposed sampling scheme is general enough to accommodate GRD estimators from all classes. More importantly, the algorithm results in a unique sampling sequence for all GRD functions. In other words, we can generate a lookup table of optimal querying order, making the sampling process fully parallelizable.

V. Experiments

In this section, we first describe the experimental setups including our GRD function database, the implementation details of the proposed algorithm, and the evaluation criteria. We then compare the proposed algorithm with existing GRD estimation methods.

A. Experimental setups

GRD Function Database: We construct a new video database which contains 250 pristine videos that span a great diversity of video content. An important consideration in selecting the videos is that they need to be representative of the videos we see in the daily life. Therefore, we resort to the Internet and elaborately select 200 keywords to search for creative common licensed videos. We initially obtain more than 700 4K videos. Many of these videos contain significant distortions, including heavy compression artifacts,

TABLE I
MSE PERFORMANCE OF THE COMPETING GRD FUNCTION MODELS WITH DIFFERENT NUMBER OF LABELED SAMPLES SELECTED BY RANDOM SAMPLING (RS) AND THE PROPOSED UNCERTAINTY SAMPLING (US). SMALLEST ERRORS ARE HIGHLIGHTED WITH BOLDFACE.

sample #	Reciprocal [14]		Logarithmic [8]		PCHIP		CT		RAMCT	
	RS	US	RS	US	RS	US	RS	US	RS	US
20	N.A.	N.A.	23.07	**13.33**	68.76	26.49	88.54	56.04	135.27	16.10
30	62.27	83.34	13.08	10.56	30.95	**2.06**	37.99	22.78	10.98	3.29
50	38.11	73.88	9.43	6.77	8.64	0.07	11.75	12.16	4.70	**0.06**
75	30.27	48.85	5.15	4.92	3.08	**0**	4.84	3.26	1.01	**0**
100	27.44	38.46	4.60	4.18	1.77	**0**	2.75	1.26	0.13	**0**
540	24.51	24.51	2.76	2.76	**0**	**0**	**0**	**0**	**0**	**0**

TABLE II
l_∞ PERFORMANCE OF THE COMPETING GRD FUNCTION MODELS WITH DIFFERENT NUMBER OF LABELED SAMPLES SELECTED BY RANDOM SAMPLING (RS) AND THE PROPOSED UNCERTAINTY SAMPLING (US). SMALLEST ERRORS ARE HIGHLIGHTED WITH BOLDFACE.

sample #	Reciprocal [14]		Logarithmic [8]		PCHIP		CT		RAMCT	
	RS	US	RS	US	RS	US	RS	US	RS	US
20	N.A.	N.A.	19.40	**16.56**	38.87	28.11	36.50	29.51	45.15	21.88
30	48.32	45.36	17.85	12.28	33.04	11.07	29.84	18.70	27.07	**6.13**
50	52.48	45.48	15.75	12.37	24.33	**2.10**	21.82	14.30	23.99	2.13
75	54.49	49.08	14.59	13.53	18.22	0.47	17.89	7.76	16.51	**0.11**
100	55.54	51.26	14.22	14.44	16.00	0.26	15.59	5.84	14.23	**0**
540	58.04	58.04	18.33	18.14	**0**	**0**	**0**	**0**	**0**	**0**

noise, blur, and other distortions due to improper operations during video acquisition and sharing. To make sure that the videos are of pristine quality, we carefully inspect each of the videos multiple times by zooming in and remove those videos with visible distortions. We further reduce artifacts and other unwanted contaminations by downsampling the videos to a size of 1920 $\times$ 1080 pixels, from which we extract 10 seconds semantically coherent video clips. Eventually, we end up with 250 high quality videos.

Using the aforementioned sequences as the source, each video is distorted by the following processes sequentially:

- Spatial downsample: We downsample source videos using bi-cubic filter to six spatial resolutions (1920 $\times$ 1080, 1280 $\times$ 720, 720 $\times$ 480, 512 $\times$ 384, 384 $\times$ 288, 320 $\times$ 240) according to the list of Netflix certified devices [7].
- H.264/HEVC/VP9 compression: We encoded the downsampled sequences using the three commonly used video encoders with two-pass encoding. Specifically, the x264 [30], x265 [31], and vpx-vp9 [32] libraries are employed for H.264, HEVC and VP9 encoding, respectively. In the experiment, we used two-pass encoding to demonstrate the usefulness of RAMCT. However, the proposed model can also be applied to one-pass encoding when low latency is of major concern. Detailed encoding specifications are given in the Appendix. The target bitrate ranges from 100 kbps to 9 Mbps with a step size of 100 kbps.

In total, we obtain 540 (hypothetical reference circuit) $\times$ 250 (source) $\times$ 3 (encoder) = 405,000 video representations (currently the largest in the VQA community). We evaluate the quality of each video representation at five commonly used display devices including cellphone, tablet, laptop, desktop, and TV using SSIMplus [33] for the following reasons. First, SSIMplus is currently the only HVS motivated spatial resolution and display device-adapted VQA model that is shown to outperform other state-of-the-art quality measures in terms of accuracy and speed [33], [34]. Second, a simplified VQA model SSIM [35] has been demonstrated to perform well in estimating the GRD functions [8]. The resulting dense samples of SSIMplus are regarded as the ground truth of GRD functions (The range of SSIMplus is from 0 to 100 with 100 indicating perfect quality). However, our GRD modeling approach does not constrain itself on any specific VQA methods. When other ways of generating dense ground-truth samples are available, the same GRD modeling approach may also be applied.

Implementation Details: We initialize the scattered network with delaunay triangulation [36], inherited from CT method [19]. The balance weight λ in (16) is set to 10^{-4}. In our current experiments, the performance of the proposed RAMCT is fairly insensitive to variations of the value. We employed OSQP [27] to solve the quadratic programming problem, where the maximum number of iterations is set to 10^6. The stopping criteria threshold T is set to 540 (the total number of representation samples in the discretized GRD function space) $\times$ 10 (the standard deviation of mean opinion score in the LIVE Video Quality Assessment database), resulting in an average sample number of 38. When $tr(\Sigma)$ is below the threshold, we conclude that the uncertainty in the system can be explained by the disagreement between subjects. Therefore, further improvement in prediction accuracy may not be as meaningful. Since a triangulation only covers the convex hull of the scattered point set, extrapolation beyond the convex hull is not possible. In order to make a fair comparison, we initialize the training set S as the representations with maximum and minimum bitrates at all spatial resolutions. To construct the covariance matrix described in Section IV as well as test the proposed algorithm, we randomly segregated the database into a training set of 200 GRD functions and a testing set with 50 GRD functions. The random split is repeated 50 times and the median performance is reported.

Evaluation Criteria: We test the performance of the GRD

estimators in terms of both accuracy and rate of convergence. Specifically, we used two metrics to evaluate the accuracy. The mean squared error (MSE) and l_∞ norm of the error values are computed between the estimated function and the actual function for each source content. The median results are then computed over all testing functions. All interpolation models can fit increasingly complex GRD functions at the cost of using many parameters. What distinguishes these models from each other is the rate and manner with which the quality of the approximation varies with the number of training samples.

B. Performance

We test five GRD function models including reciprocal regression [6], logarithmic regression [8], 1D piecewise cubic Hermite interpolating polynomial (PCHIP), CT interpolation, and the proposed RAMCT on the aforementioned database. To evaluate the performance of the uncertainty sampling algorithm, we apply it on the five GRD models above and compare its performance with random sampling scheme as the baseline. For random sampling, the initial set of training sample S is set as the representations with the maximum and minimum bitrates at all spatial resolutions to allow fair comparison. The training process with random sampling was repeated 50 times and the median performance is reported.

Table I and II show the prediction accuracy on the database, from which the key observations are summarized as follows. First, the models that assume a certain analytic functional form are consistently biased, failing to accurately fit GRD functions even with all samples probed. On the other hand, the existing interpolation models usually take more than 100 random samples to converge, although they are asymptotically unbiased. By contrast, the proposed RAMCT model converges with only a moderate number of samples. Second, we analyze the core contributors of RAMCT with deliberate selection of competing models. Both the RAMCT and the PCHIP models outperform the traditional CT model, suggesting the importance of axial monotonicity. Besides, the RAMCT model achieves better performance than the PCHIP model by exploiting the 2D structure and jointly modeling the GRD functions. Third, we observe strong generalizability of the proposed uncertainty sampling strategy evident by the significant improvement over random sampling on all models. The performance improvement is most salient on the proposed model. In general, RAMCT is able to accurately model GRD functions with only 30 labeled samples, based on which the reciprocal model merely have sufficient known variables to initialize fitting. To gain a concrete impression, we also recorded the execution time of the entire GRD estimation pipeline including video encoding, objective VQA, and GRD function approximation with the competing algorithms on a computer with 3.6GHz CPU and 16G RAM. RAMCT with uncertainty sampling takes around 10 minutes to reduce l_∞ below 5, which is more than 100 times faster than the tradition regression models with random sampling.

C. Performance with other VQA models

The proposed RAMCT model does not constrain itself to a specific VQA measure. To demonstrate this, we use the RAMCT model to construct GRD functions measured by another widely used VQA model, peak signal-to-noise ratio (PSNR). We follow similar experimental setups as described in Section V-A and V-B except employing PSNR as the VQA measure, and that only the uncertainty sampling strategy is used. The experimental results are summarized in Table III, from which we find that the RAMCT method well generalizes to the PSNR metric. This is because the RAMCT model is based on the common properties of GRD functions rather than the peculiarities of a specific VQA model.

D. Performance with High Dynamic Range Videos

In recent years, high dynamic range (HDR) videos are becoming increasingly popular [39]. It is desired know how the RAMCT model generalizes to GRD functions of HDR videos. We test the RAMCT model with the uncertainty sampling method on the Waterloo UHD-HDR-WCG database [37], which consists of 15 different HDR video contents. Specifically, we downsample the 15 videos to 1080P, and treat them as the reference. These videos are then downsampled and compressed into 540 representations following the same procedures as described in Section V-A. We select SR-SIM [38] as the VQA model, which exhibits the highest correlation with subjective opinions in previous experiments [37], and re-scale the SR-SIM scores from $[0, 1]$ to $[0, 100]$ for better presentation. The experimental results are listed in Table IV, from which we can see that the RAMCT model performs consistently well in reconstructing GRD functions of HDR videos with a new VQA metric. The results not only further attest that the proposed RAMCT model is robust to different VQA metrics, but also indicate that RAMCT generalizes well to HDR videos. This may be ascribed to the fact that the constraints of the RAMCT model are primarily derived from the RD theory, which applies to any kinds of signal.

VI. Applications

The application scope of GRD model is much broader than VQA. Here we demonstrate three use cases.

A. Rate-Distortion Curve at Novel Resolutions

Given a set of RD curves at multiple resolutions, it is desirable to predict the RD performance at novel resolutions, especially when there exists a mismatch between the supported viewing device of downstream content delivery network and the recommended encoding profiles. Traditional methods linearly interpolate the RD curve at novel resolutions [7], neglecting the characteristics of GRD functions. Fig. 5 compares the linearly interpolated and RAMCT-interpolated RD curves at 960×540 with the ground truth SSIMplus curve, from which we have several observations. First, the linearly interpolated curve shares the same intersection with the neighboring curves at 740×480 and 1280×720, inducing consistent bias to the prediction. The proposed RAMCT model is able to accurately predict the quality at the intersection of the neighboring curves by taking all known RD curves into consideration. Second, the linearly interpolated RD curve always lies between its

TABLE III
MSE AND l_∞ PERFORMANCES OF THE COMPETING MODELS WITH THE PSNR METRIC. SMALLEST MSE AND l_∞ ERRORS ARE HIGHLIGHTED WITH ITALICS OR BOLDFACE, RESPECTIVELY.

sample #	Reciprocal [14]		Logarithmic [8]		PCHIP		CT		RAMCT	
	MSE	l_∞	MSE	l_∞	MSE	l_∞	MSE	l_∞	MSE	l_∞
20	N.A.	N.A.	*0.08*	**1.30**	0.34	2.20	0.47	2.28	0.18	2.36
30	4.56	10.28	0.05	**0.86**	0.07	1.09	1.33	2.20	*0.04*	1.08
50	5.07	9.74	0.04	0.85	*0*	0.53	0.11	1.20	*0*	**0.38**
75	3.85	10.85	0.03	0.94	*0*	0.27	0.05	0.85	*0*	**0.24**
100	3.38	11.47	0.03	1.02	*0*	**0.14**	0.02	0.66	*0*	0.16
540	2.46	12.98	0.02	1.30	*0*	**0**	*0*	**0**	*0*	**0**

TABLE IV
MSE AND l_∞ PERFORMANCE OF THE COMPETING MODELS ON HDR VIDEOS [37] WITH THE SRSIM METRIC [38]. SMALLEST MSE AND l_∞ ERRORS ARE HIGHLIGHTED WITH ITALICS OR BOLDFACE, RESPECTIVELY.

sample #	Reciprocal [14]		Logarithmic [8]		PCHIP		CT		RAMCT	
	MSE	l_∞	MSE	l_∞	MSE	l_∞	MSE	l_∞	MSE	l_∞
20	N.A.	N.A.	*0.49*	**1.88**	3.63	4.79	3.64	5.75	2.06	3.33
30	3.67	7.47	0.36	**1.75**	1.34	3.99	4.39	7.28	*0.34*	2.26
50	2.97	7.52	0.18	1.41	*0.03*	0.99	1.27	3.45	*0.03*	**0.46**
75	2.04	7.44	0.14	1.43	*0*	**0.26**	0.05	0.73	*0*	**0.26**
100	1.57	7.62	0.09	1.53	*0*	**0.09**	0.03	0.69	*0*	0.21
540	0.89	9.28	0.06	2.01	*0*	**0**	0	**0**	*0*	**0**

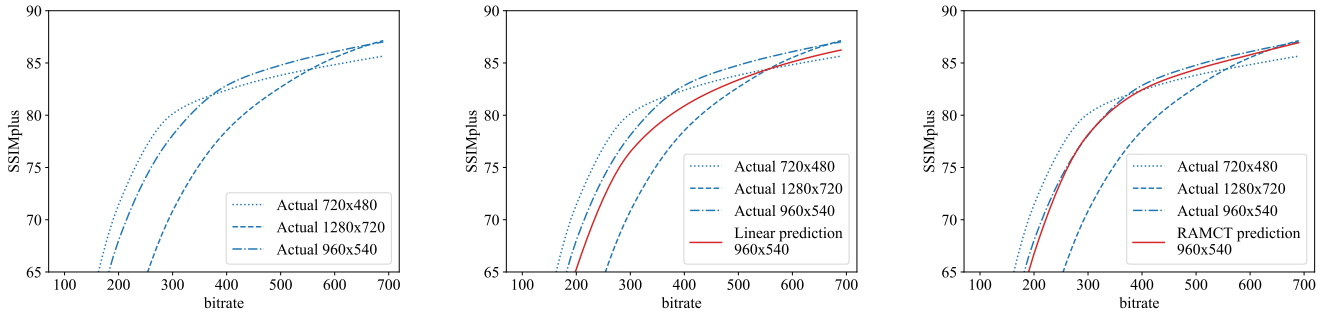


Fig. 5. Prediction of RD curve of novel resolution from known RD curves of other resolutions. (a) Ground truth RD curves; (b) Prediction of 960×540 RD curve from 720×480 and 1280×720 curves using linear interpolation; (c) Prediction of 960×540 RD curve using the proposed method.

TABLE V
PERFORMANCE OF LINEAR INTERPOLATION AND RAMCT ON PREDICTING THE RD FUNCTION AT A NOVEL RESOLUTION.

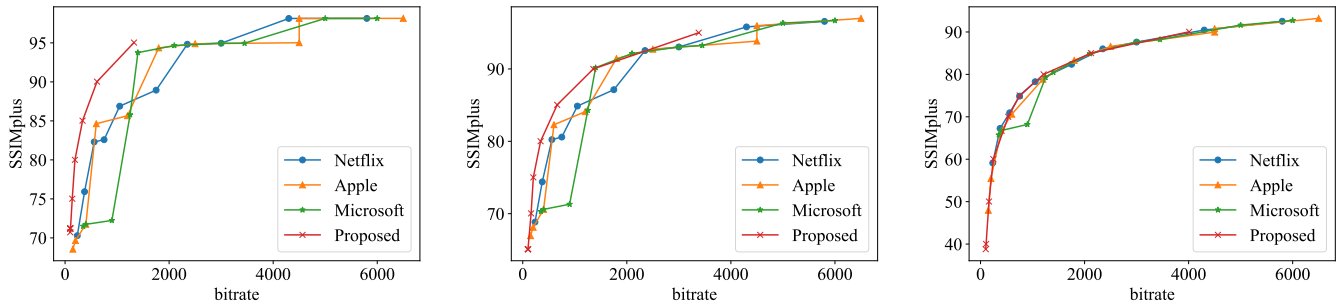
Resolution	l_∞		MSE	
	Linear	RAMCT	Linear	RAMCT
640×360	7.68	3.12	7.89	1.56
960×540	7.66	4.83	6.61	3.14
1600×900	8.77	7.87	5.18	4.98
Average	8.04	5.27	6.56	3.23

neighboring curves, suggesting that the predicted quality at any bitrate is lower than the quality on one of its neighboring curves. This behavior contradicts the fact that each resolution may have a bitrate region in which it outperforms other resolutions [7]. On the contrary, RAMCT better preserves the general trend of resolution-quality curve at different bitrate, thanks to the regularization imposed by the C^1 condition at given nodes. Third, RAMCT outperforms the linear interpolation model in predicting the ground truth RD curve across all bitrates. The experimental results also justify the effectiveness of the C^1 and smoothness prior used in RAMCT.

To further validate the performance of the proposed GRD model at novel spatial resolutions, we predict the RD curves of 20 randomly selected source videos from the dataset at three novel resolutions (640×360, 960×540, and 1600×900). The evaluated bitrate ranges from 100 kbps to 9 Mbps with a step size of 100 kbps. The results are listed in Table V. We can observe that RAMCT outperforms the linear model [7] with a clear margin at novel resolutions.

B. Per-Title Encoding Profile Generation

To overcome the heterogeneity in users' network conditions and display devices, video service providers often encode videos at multiple bitrates and spatial resolutions. However, the selection of the encoding profiles are either hard-coded, resulting in sub-optimal QoE due to the negligence of the difference in source video complexities, or selected based on interactive objective measurement and subjective judgement that are inconsistent and time-consuming. To deliver the best quality video to consumers, each title should receive a unique bitrate ladder, tailored to its specific complexity characteristics. This process is known as per-title optimization. We will show



(a) Title with low complexity (b) Title with moderate complexity (c) Title with high complexity

Fig. 6. Bitrate ladders generated by the recommendations and the proposed algorithm for three contents.

TABLE VI
AVERAGE BITRATE SAVING OF ENCODING PROFILES. NEGATIVE VALUES INDICATE ACTUAL BITRATE REDUCTION.

	Microsoft	Apple	Netflix	Proposed
Microsoft	0	-	-	-
Apple	-25.3%	0	-	-
Netflix	-29.3%	-5.6%	0	-
Proposed	-62.0%	-48.9%	-46.8%	0

the pivotal role of the proposed RAMCT model in a quality-driven per-title optimization framework.

Content delivery networks often aim to deliver videos at certain quality levels to satisfy different viewers. It is beneficial to minimize the bitrate usage in the encoding profile when achieving the objective. Mathematically, the quality-driven bitrate ladder selection problem can be formulated as a constrained optimization problem. Specifically, for the i-th representation,

$$\begin{aligned} &\underset{\{x,y\}}{\text{minimize}} \quad x \\ &\text{subject to} \quad f(x,y) \geq C_i,\ i=1,\ldots,m, \end{aligned}$$

where x, y, $f(\cdot,\cdot)$, C_i and m represent the bitrate, the spatial resolution, the GRD function, the target quality level of video representation i, and the total number of video representations, respectively. Solving the optimization problem requires precise knowledge of the GRD function. Thanks to the effectiveness and differentiability of RAMCT, the proposed model can be incorporated with gradient-based optimization tools [40] to solve the per-title optimization problem. (Interested readers may refer to the Appendix for more details on how we solve the optimization problem).

To validate the proposed per-title encoding profile selection algorithm, we apply the algorithm to generate bitrate ladders using H.264 [41] for 50 randomly selected videos in the aforementioned dataset. We set the target quality levels $\{C_i\}_{i=1}^{10}$ as $\{30, 40, 50, 60, 70, 75, 80, 85, 90, 95\}$ to cover diverse quality range and to match the total number of representations in standard recommendations [6]. For simplicity, we optimize the representation sets for only one viewing device (cellphone), while the procedure can be readily extended to multiple devices to generate a more comprehensive representation set. In Fig. 6, we compare the rate-quality curve of representation sets generated by the proposed algorithm to the hard-coded bitrate ladders recommended by Netflix [42], Apple [43], and Microsoft [44] for three videos with different complexities, from which the key observations are as follows. First, contrasting the hand-crafted bitrate ladders, the encoding profile generated by the proposed algorithm is content adaptive. Specifically, the encoding bitrate increases with respect to the complexity of the source video as illustrated from Fig. 6(a) to Fig. 6(c). Second, the proposed method achieves the highest quality at all bitrate levels. The performance improvement is mainly introduced by the encoding strategy at the convex hull encompassing the individual per-resolution RD curves [7]. Table VI provides a full summary of the Bjøntegaard-Delta bitrate (BD-Rate) [45], indicating the required overhead in bitrate to achieve the same SSIMplus values. We observe that the proposed framework outperforms the existing hard-coded bitrate ladders by at least 47%. Since the Netflix [42], Apple [43], and Microsoft [44] recommendations may be tuned with PSNR rather than SSIMplus [33], we also conduct a similar experiment using PSNR as the quality measure for fair comparison. Similar results are observed, showing the robustness of the RAMCT model in the context of per-title optimization. Interested readers may refer to the Appendix for more details.

C. Encoder Comparison

In the past decade, there has been a tremendous growth in video compression algorithms and implementations, thanks to the fast development of computational multimedia. With many video encoders at hand, it becomes pivotal to compare their performance, so as to find the best algorithm as well as directions for further advancement. Bjøntegaard-Delta model [17], [45] has become the most commonly used objective coding efficiency measurement. Bjøntegaard-Delta PSNR (BD-PSNR) and BD-Rate are typically computed as the difference in bitrate and quality (measured in PSNR) based on the interpolated rate-distortion curves

$$Q_{BD} = \frac{\int_{x_L}^{x_H} [z_B(x) - z_A(x)]dx}{\int_{x_L}^{x_H} dx}, \tag{23a}$$

$$R_{BD} \approx 10^{\frac{\int_{z_L}^{z_H} [x_B(z) - x_A(z)]dz}{\int_{z_L}^{z_H} dz}} - 1, \tag{23b}$$

where x_A and x_B are the logarithmic-scale bitrate, z_A and z_B are the quality of the interpolated reference and test bitrate

curves, respectively. $[x_L, x_H]$ and $[z_L, z_H]$ are the effective domain and range of the rate-distortion curves. However, BD-PSNR and BD-Rate do not take spatial resolution into consideration. Fig. 7 shows two GRD surfaces generated by x264 [30] and x265 [31] encoders for a source video. Although H.264 performs on par with HEVC at low resolutions, it requires higher bitrate to achieve the same target quality at high resolutions. Therefore, applying BD-Rate on a single resolution is not sufficient to fairly compare the overall performance between encoders. To this regard, we propose generalized quality gain (Q_{gain}) and rate gain (R_{gain}) models as

$$Q_{gain} = \frac{\int_U \int_{y_L}^{y_H} \int_{x_L}^{x_H} p(u)[z_B(x,y,u) - z_A(x,y,u)]dxdydu}{\int_{y_L}^{y_H} \int_{x_L}^{x_H} dxdy}, \tag{24a}$$

$$R_{gain} \approx 10^{\frac{\int_U \int_{y_L}^{y_H} \int_{z_L}^{z_H} p(u)[x_B(z,y,u) - x_A(z,y,u)]dzdydu}{\int_U \int_{y_L}^{y_H} \int_{z_L}^{z_H} p(u)dzdydu}} - 1, \tag{24b}$$

where $p(u)$, U, and $[y_L, y_H]$ represent the probability density of viewing devices, the set of all device of interests, and the domain of video spatial resolution, respectively. The generalized Q_{gain} and R_{gain} models represent the expected quality gain and the expected bitrate gain (saving when R_{gain} negative) across all spatial resolutions and viewing devices, leading to a more comprehensive evaluation of practical encoders. It should be noted that $z_A(x, y, u)$ is essentially the GRD function of codec A, which can be efficiently approximated by the proposed model. $x_A(z, y, u)$ can also be estimated numerically from the interpolated surface. (Interested readers may refer to the Appendix for more details on how we compute Q_{gain} and R_{gain}.) Therefore, RAMCT is a natural fit to the generalized Q_{gain} and R_{gain} models. The effect of any individual influencing factor can be obtained by taking the marginal expectation in the corresponding dimension, which is more robust than BD-PSNR and BD-Rate at a single resolution.

In summary, we show that the proposed RAMCT model can help compare two video encoders more comprehensively than BD-PSNR and BD-Rate [17], [45]. However, it should be noted that such comparison is only restricted to the specific encoder implementations and configurations, and does not take into consideration time complexities.

VII. Conclusions

GRD functions represent the critical link between multimedia resource and perceptual QoE. In this work, we proposed a learning framework to model the GRD function by exploiting the properties all GRD functions share and the information redundancy of training samples. The framework leads to an efficient algorithm that demonstrates state-of-the-art performance, which we believe arises from the RAMCT model for imposing axial monotonicity, the joint modeling of the multi-dimensional GRD function for exploiting its functional structure, and the information-theoretic sampling algorithm for improving the quality of training samples. Extensive experiments have shown that the algorithm is able to accurately model the function with a very small number of training samples. Furthermore, we demonstrate that the proposed GRD model plays a central role in a great variety of visual communication applications.

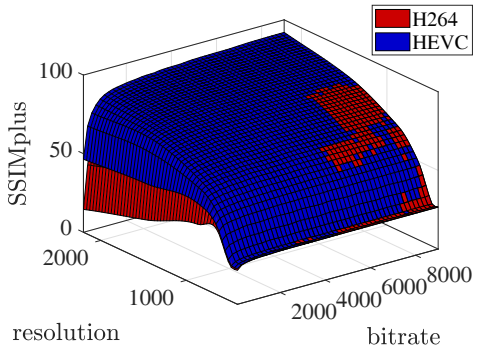


Fig. 7. GRD surfaces of H.264 [30] and HEVC [31] encoders for a sample source video.

The current work can be extended in many ways. As a basis for future work, we note that the interpolant can be readily extended to higher dimensions [24], making it applicable to more general applications. For example, in the fields of machine learning [29] and data visualization [46], flexible monotonic interpolation can provide regularization and makes the model more interpretable. Another promising future direction is to develop models that can predict GRD functions without sampling the GRD space.

References

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *Institute of Radio Engineers, International Convention Record*, vol. 4, no. 4, pp. 142–163, Mar. 1959.

[2] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," in *Picture Coding Symposium*, 2013, pp. 394–397.

[3] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

[4] Y. Ou, Y. Xue, and Y. Wang, "Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," *IEEE Trans. Image Processing*, vol. 23, no. 6, pp. 2473–2486, Jun. 2014.

[5] W. Zhang, Y. Wen, Z. Chen, and A. Khisti, "QoE-driven cache management for HTTP adaptive bit rate streaming over wireless networks," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1431–1445, Oct. 2013.

[6] L. Toni, R. Aparicio-Pardo, K. Pires, G. Simon, A. Blanc, and P. Frossard, "Optimal selection of adaptive streaming representations," *ACM Trans. Multimedia Computing, Communications, and Applications*, vol. 11, no. 2, pp. 1–43, Feb. 2015.

[7] J. De Cock, Z. Li, M. Manohara, and A. Aaron, "Complexity-based consistent-quality encoding in the cloud," in *Proc. IEEE Int. Conf. Image Proc.*, 2016, pp. 1484–1488.

[8] C. Chen, S. Inguva, A. Rankin, and A. Kokaram, "A subjective study for the design of multi-resolution ABR video streams with the VP9 codec," in *Electronic Imaging*, 2016, pp. 1–5.

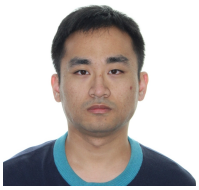
[9] Z. Wang, K. Zeng, A. Rehman, H. Yeganeh, and S. Wang, "Objective video presentation QoE predictor for smart adaptive video streaming," in *Proc. SPIE Optical Engineering+Applications*, 2015, pp. 95 990Y.1–95 990Y.13.

[10] C. Chen, Y. Lin, A. Kokaram, and S. Benting, "Encoding bitrate optimization using playback statistics for HTTP-based adaptive video streaming," *arXiv preprint arXiv:1709.08763*, Sep. 2017.

[11] D. I. Forum. (2013) For promotion of MPEG-DASH 2013. [Online]. Available: http://dashif.org.

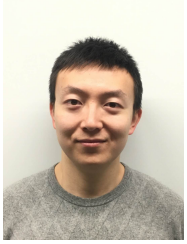
[12] Alliance for Open Media. (2018) AV1 bitstream and decoding process specification. [Online]. Available: https://aomedia.org/av1-bitstream-and-decoding-process-specification/.

[13] Z. Li, A. Aaron, L. Katsavounidis, A. Moorthy, and M. Manohara. (2016) Toward a practical perceptual video quality metric. [Online]. Available: http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html.
[14] C. Kreuzberger, B. Rainer, H. Hellwagner, L. Toni, and P. Frossard, "A comparative study of DASH representation sets using real user characteristics," in *Proc. Int. Workshop on Network and OS Support for Digital Audio and Video*, 2016, pp. 1–4.
[15] T. Berger, "Rate distortion theory and data compression," in *Advances in Source Coding*. Springer, 1975, pp. 1–39.
[16] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316–1324, Nov. 2008.
[17] G. Bjøntegaard, "Improvements of the BD-PSNR model," Tandberg, Oslo, Norway, Tech. Rep. VCEG-AI11, ITU-T SG 16/Q6, 35th VCEG Meeting, Jul. 2008.
[18] J. Robson, "Spatial and temporal contrast-sensitivity functions of the visual system," *Journal of Optical Society of America*, vol. 56, no. 8, pp. 1141–1142, Aug. 1966.
[19] R. Clough and T. J., "Finite element stiffness matrices for analysis of plates in bending," in *Proceedings of Conf. on Matrix Methods in Structural Analysis*, 1965.
[20] P. Alfeld, "A trivariate Clough-Tocher scheme for tetrahedral data," *Computer Aided Geometric Design*, vol. 1, no. 2, pp. 169–181, Jun. 1984.
[21] I. Amidror, "Scattered data interpolation methods for electronic imaging systems: A survey," *Journal of Electronic Imaging*, vol. 11, no. 2, pp. 157–177, Apr. 2002.
[22] G. Farin, "Bézier polynomials over triangles and the construction of piecewise C^R polynomials," *Brunel University Mathematics Technical Papers collection*, 1980.
[23] G. M. Nielson, "A method for interpolating scattered data based upon a minimum norm network," *Mathematics of Computation*, vol. 40, no. 161, pp. 253–271, 1983.
[24] G. Farin, "A modified Clough-Tocher interpolant," *Computer Aided Geometric Design*, vol. 2, no. 1-3, pp. 19–27, Sep. 1985.
[25] N. Fritsch and R. Carlson, "Monotone piecewise cubic interpolation," *SIAM Journal on Numerical Analysis*, vol. 17, no. 2, pp. 238–246, 1980.
[26] L. Han and L. Schumaker, "Fitting monotone surfaces to scattered data using c1 piecewise cubics," *SIAM Journal on Numerical Analysis*, vol. 34, no. 2, pp. 569–585, Apr. 1997.
[27] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, "OSQP: An operator splitting solver for quadratic programs," *ArXiv preprint arXiv:1711.08013*, Nov. 2017.
[28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[29] C. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Heidelberg: Springer-Verlag, 2006.
[30] FFmpeg team. (2018) FFmpeg v.2.8.15. [Online]. Available: https://trac.ffmpeg.org/wiki/Encode/H264
[31] ——. (2018) FFmpeg v.2.8.15. [Online]. Available: https://trac.ffmpeg.org/wiki/Encode/H.265
[32] ——. (2018) FFmpeg v.2.8.15. [Online]. Available: https://trac.ffmpeg.org/wiki/Encode/VP9
[33] A. Rehman, K. Zeng, and Z. Wang, "Display device-adapted video Quality-of-Experience assessment," in *Proc. SPIE*, 2015, pp. 939 406.1–939 406.11.
[34] Z. Duanmu, K. Ma, and Z. Wang, "Quality-of-Experience of adaptive video streaming: Exploring the space of adaptations," in *Proc. ACM Int. Conf. Multimedia*, 2017, pp. 1752–1760.
[35] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[36] B. Delaunay, "Sur la sphere vide," *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, vol. 7, no. 793-800, pp. 1–2, Oct. 1934.
[37] S. Athar, T. Costa, K. Zeng, and Z. Wang, "Perceptual quality assessment of UHD-HDR-WCG videos," in *Proc. IEEE Int. Conf. Image Proc.*, Sep. 2019, pp. 1740–1744.
[38] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in *Proc. IEEE Int. Conf. Image Proc.*, Sep. 2012, pp. 1473–1476.
[39] A. Chalmers and K. Debattista, "HDR video past, present and future: A perspective," *Signal Processing: Image Communication*, vol. 54, pp. 49–55, May 2017.
[40] M. Grant and S. Boyd. (2014) CVX: Matlab software for disciplined convex programming, version 2.1. [Online]. Available: http://cvxr.com/cvx.
[41] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[42] A. Aaron, Z. Li, M. Manohara, D. J. Cock, and D. Ronca. (2015) Per-Title encode optimization. [Online]. Available: https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2.
[43] Apple. (2016) Best practices for creating and deploying HTTP live streaming media for iPhone and iPad. [Online]. Available: http://is.gd/LBOdpz.
[44] G. Michael, T. Christian, H. Hermann, C. Wael, N. Daniel, and B. Stefano. (2013) Combined bitrate suggestions for multi-rate streaming of industry solutions. [Online]. Available: http://alicante.itec.aau.at/am1.html.
[45] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," Telenor Satellite Services, Oslo, Norway, Tech. Rep. VCEG-M33, ITU-T SG 16/Q6, 13th VCEG Meeting, Apr. 2001.
[46] M. Sarfraz and M. Hussain, "Data visualization using rational spline interpolation," *Journal of Computational and Applied Mathematics*, vol. 189, no. 1, pp. 513–525, May 2006.



Zhengfang Duanmu (S15) received the B.A.Sc. and M.A.Sc. degrees in electrical and computer engineering from the University of Waterloo in 2015 and 2017, respectively, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests include perceptual image processing and quality-of-experience.

Wentao Liu (S15-M'20) received the B.E. and M.E. degrees from Tsinghua University, Beijing, China in 2011 and 2014, respectively, and the Ph.D. degree in electrical and computer engineering from University of Waterloo, ON, Canada, in 2019, where he is currently a Postdoctoral Fellow. His research interests include image and video processing; perceptual quality assessment; computational vision; and multimedia communications.



Zhuoran Li (S14) received the B.A.Sc. degree from the McMaster University at Hamilton, ON, Canada, in 2017. He is currently pursuing the Ph.D. degree in electrical and computer engineering at the University of Waterloo, Waterloo, ON, Canada. His research interests include perceptual image processing, video quality of experience, and video coding.

Zhou Wang (S99M02SM12F14) received the Ph.D. degree from The University of Texas at Austin in 2001. He is currently a Canada Research Chair and Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image and video processing and coding; visual quality assessment and optimization; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has more than 200 publications in these fields with over 50,000 citations (Google Scholar).

Dr. Wang serves as a member of IEEE Image, Video and Multidimensional Signal Processing Technical Committee (2020-2022) and IEEE Multimedia Signal Processing Technical Committee (2013-2015), a Senior Area Editor of IEEE Transactions on Image Processing (2015-2019), an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology (2016-2018), IEEE Transactions on Image Processing (2009-2014), IEEE Signal Processing Letters (2006-2010), and a Guest Editor of IEEE Journal of Selected Topics in Signal Processing (2013-2014 and 2007-2009), among other journals. He was elected a Fellow of Royal Society of Canada, Academy of Science in 2018, and a Fellow of Canadian Academy of Engineering in 2016. He is a recipient of 2016 IEEE Signal Processing Society Sustained Impact Paper Award, 2015 Primetime Engineering Emmy Award, 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, 2013 IEEE Signal Processing Magazine Best Paper Award, and 2009 IEEE Signal Processing Society Best Paper Award.