

Research Article

Metaverse-Empowered Music and Dance: Experience of Emotion and Scene Unification

Fengfei Wang 

Xinxiang University, No. 191, Jinsui Road, Xinxiang City, 453003 Henan, China

Correspondence should be addressed to Fengfei Wang; wangfenglei8909@xxu.edu.cn

Received 25 February 2022; Accepted 22 March 2022; Published 4 April 2022

Academic Editor: Jianhui Lv

Copyright © 2022 Fengfei Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Music and dance are closely related and symbiotic. On the one hand, dance often requires music accompaniment. On the other hand, dance can enrich the melody and style of music. The emergence of the metaverse has taken the experience of music and dance to a new level. This paper studies the three-dimensional situational experience of music and dance in Virtual Reality (VR) empowered by metaverse to feel the beauty of situational integration. After the spherical video is projected onto a two-dimensional plane to form a panoramic video, the two-dimensional panoramic video needs to be converted into a spherical video for users to watch. Therefore, it is more reasonable to take spherical video distortion as the distortion measure of panoramic video coding. In this paper, spherical video distortion is taken as the measurement standard of video quality, and the panoramic video coding technology is optimized. Furthermore, the corresponding weights are introduced to change the distortion ratio of different interpolation regions in the calculation process of rate distortion cost, and a rate distortion optimization technology based on spherical distortion measurement is proposed. The equal weight feature of spherical pixel is realized on two-dimensional plane, which improves the coding efficiency of panoramic video. Experimental results show that compared with the three benchmarks, the proposed algorithm can achieve 1.6157% bit saving on average and achieve a good Quality of Experience (QoE) when other processes are the same.

1. Introduction

Music contains rich ideological connotations, which can be reflected through vivid musical artistic images [1]. In order to highlight and embody the thoughts and emotions of dance works, it is necessary to analyze and express the musical elements. Music is an art form of expressing rich emotions with sound, rhythm, and melody, while dance is an art form of expressing the emotional connotation of music with human body and movement [2–4]. Therefore, music and dance are closely related. For example, many ethnic dances in China have their own unique styles, and dance music also contains different cultural customs. Therefore, when appreciating the related works, it is necessary to understand not only the skills of dance movements but also the national cultural connotation of the dance, so as to correctly understand the emotional connotation and style rhythm contained in the folk dance music, and improve the appreciation of dance music.

During the dance performance, the music starts first and then the dance performance catches the eye. Therefore, music is the soul of dance art and an important prerequisite to lead people to appreciate dance [5]. Music can touch the deep emotions of human being through beautiful melodies and different rhythms and harmony forms and realize the ideological resonance between the audience and the connotation of dance. Hence, the use of music melody in dance is an important basis for the full expression of dance art. By feeling the music, it can stimulate people's deep desire of emotional expression, so as to realize the real emotional expression of dance with the integration of body and mind, and exercise people's ability of dance expression [6–8]. In addition, music can also be used as a basis for dance, so that people can show the beauty of their body movements to a greater extent in music. In dance, people show the corresponding dance movements by mastering the rhythm and tone of music and grasp the emotional expression of dance through in-depth appreciation of music, so that the dance

shown by people can be more closely connected with music, and at the same time, it can also make people's dance with emotion and richer connotation [9, 10]. Such training for a long time can make people through the mastery of music to innovate dance movements and improve their own dance performance.

Music is an important support of dance art and an important medium to help dance present rich and three-dimensional expressive effects. At present, the experience of dance is no longer the training of body movements and skills, but how to guide people to fully show the thoughts and feelings contained in dance with the help of body movements. Music is used to stimulate artistic emotions so that people can complete dance movements through emotional stimulation while dancing, and the dance performance is richer and more three-dimensional [11]. But when importing music, we should pay attention to whether the emotion of music is expressed clearly enough and also pay attention to the feedback, so as to help people better feel music and show the emotion of dance.

The metaverse is a virtual world created by humans that parallels but does not simply copy the real world [12]. As an emerging concept, it is gradually formed on the basis of Virtual Reality (VR), augmented reality, blockchain, cloud computing, 5G, digital twin, and other technologies [13]. Currently, most of the technologies and products that embody the metaverse are primarily focused on digital entertainment, treating the metaverse as an immersive gaming world [14]. Most people tend to focus on the entertainment properties of the metaverse rather than the potential of edutainment. As we all know, metaverse is the natural expansion of virtual environment. VR determines the manifestation of the metaverse, and virtual human-computer interaction closely related to VR is the symbol of life in the metaverse [15]. It can be said that without VR, there would be no metaverse.

This paper studies the three-dimensional situational experience of music and dance in VR empowered by metaverse to feel the beauty of situational integration. With the rapid development of multimedia technology, VR has become more and more popular due to its immersive experience. The video information in the mainstream VR is spherical video information. Spherical video requires high definition due to user experience. If uncompressed, large amount of data will bring great challenges to common hard disk storage and network transmission. Therefore, it is very important to improve the coding efficiency of spherical video [16]. Due to the limitations of existing video coding standards, spherical video needs to be projected to a two-dimensional plane to form panoramic video before codec operation. At the receiving end, the two-dimensional panoramic video is converted into a spherical video for viewing by the user. Therefore, it is more reasonable to use spherical video distortion as a distortion measure for panoramic video coding. Based on this, this paper takes the spherical video distortion as the measurement standard of video quality to optimize the panoramic video coding technology. Since panorama video is encoded by an existing encoder, rate distortion optimization (RDO) is still used in the mode selection process.

The contribution of this paper is that the corresponding weights are introduced for different projection format designs to change the proportion of distortion in different areas of interpolation in the rate distortion cost calculation process, and a rate distortion optimization technique based on spherical distortion measurement is proposed. The feature of equal weights of spherical pixels is realized in two-dimensional plane, and the coding efficiency of panoramic video is improved.

The remainder of this paper is organized as follows. Section 2 reviews related work. In Section 3, spherical distortion measurement-based ROD is presented. Experimental results are presented in Section 4. Section 5 concludes this paper.

2. Related Work

The rapid development of computer information technology and the rapid update of modern educational technology means make computer technology play an increasingly important role in music and dance. In [17], the authors designed and implemented the dance movement of the basic design idea of the automatic generation algorithm based on genetic algorithm. In [18], the authors proposed to reproduce the scene of closing eyes and listening to music in a computer vision system. In [19], the authors proposed a generative adversarial network-based cross modal association framework, which associated dance movements with music to create the required dance sequence according to the input music. In [20], the authors discussed the examples of electronic dance music from three different angles. In [21], the authors proposed a framework for generating a series of three-dimensional human dance postures for a given music. In [22], the principle and implementation of a digital audio workstation plug-in for chord sequence generation was described. In [23], a novel probability autoregression architecture was proposed to simulate future posture distribution using a standardized flow based on previous posture and music background. In [24], a graph convolution networks based automatic dance generation method were proposed.

As a new video mode, the impact of VR panorama video is beyond doubt. At present, great progress has been made in various technologies of panoramic video, especially in coding. In [25], a new motion model based on spherical coordinate transformation to explicitly solve the deformation problem was proposed to reduce the decoding time. In [26], a new octagonal mapping scheme was proposed to reduce the oversampling area and arrange the points into an octagon. In [27], the authors presented and classified the latest advances in projection methods, video quality evaluation indexes, and transmission optimization methods for video coding.

RDO is a key technology in video coding system. The traditional RDO system measures the distortion of reconstructed video from the perspective of signal processing but does not fully consider the characteristics of human vision and video call. The purpose of making full use of human visual characteristics is to maximize the visual quality of video in the area of concern under the condition of limited

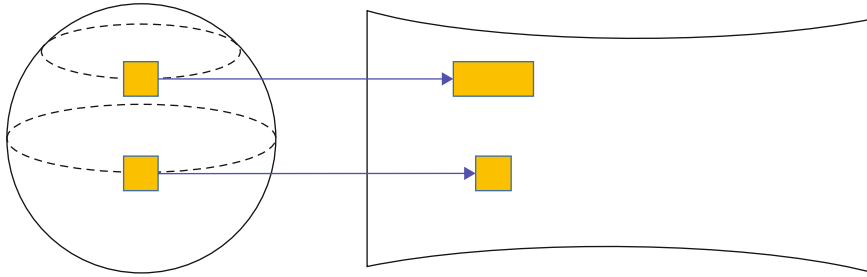


FIGURE 1: Diagram of spherical video and ERP format video.

bandwidth. Therefore, it is of great significance to study rate distortion model based on human visual perception for video coding. In [28], the authors proposed a method to generate encoded video stream. This method introduced the mathematical theory of decoding-energy-rate-distortion optimization (DERDO), which required less decoding energy than traditional encoded video stream. In [29], the authors proposed a new Lagrange multiplier determination model, which is a key part of rate distortion optimization (LM-RDO). In [30], the authors proposed an accurate coding tree unit level distortion structural similarity and distortion mean squared error (D-SSIM-D-MSE) model to obtain better video coding quality.

3. Spherical Distortion Measurement-Based RDO

Panoramic video coding still adopts the traditional block-based hybrid coding technology, so combined with the theory of encoder's block mode, prediction, rate distortion, quantization, and other operations, we can find ways to optimize panoramic video of different formats. Video distortion is as small as possible to achieve a sufficiently clear video content, and the rate of the video will be high. However, video compression coding hopes to find a way between distortion and bit rate to make the distortion small and the bit rate does not exceed the maximum allowable bit rate. Due to the constraints and contradictions between the two, RDO technology arises at the historic moment. For all lossy compression systems, RDO throughout the video coding system is a very important technology to balance the relationship between distortion and bit rate.

After the completion of the encoding and decoding of panoramic video, the spherical video is actually an output for users to watch. Therefore, for spherical video, regardless of the user viewing window, each pixel in the spherical video is equally important, that is, all pixels on the sphere are equally weighted. However, corresponding to different projection formats, the importance of pixels will change in the process of projection, which will affect the selection process of rate distortion, and further affect the final codec result, so that the video quality loses a certain degree of accuracy under the measurement of spherical video distortion.

In this paper, RDO in coding is improved according to different features of different projection formats of panoramic video. Taking Equirectangular Projection (ERP) as

an example, a RDO model of panoramic video based on spherical distortion measurement is proposed.

In panoramic video, ERP format is taken as an example, the video content is stretched at different latitudes due to different interpolation operations in the transformation process. The pixel stretching is the smallest or even nonexistent on the equatorial path. From the equator to the two poles, the horizontal stretching of the pixels becomes more and more serious, and the stretching of the two poles is the largest.

The north and south pole contents of ERP format generally differ greatly in the encoding process due to the interpolation operation of stretching. The larger the stretch, the flatter the area is, resulting in a larger optimal block size after the rate distortion optimization process. For the same mode, selection near the equator obviously has a greater impact on the final video quality than the north and south pole options.

Theoretically, distortion in the coding process should be equally important for content near the equator on the sphere and near high latitudes, as shown in Figure 1. However, when converted to ERP format, the coding process calculates the rate distortion cost by applying the same weight to the distortion of the corresponding content of two blocks in ERP format. After back-projection to spherical surface, the distortion importance of the content in the actual high latitude area becomes greater.

Therefore, this paper adopts the method of introducing weight to distortion to reduce the proportion of distortion d when calculating the cost of rate distortion and to increase the proportion of distortion near the equator, so as to further improve the coding performance.

As for the panoramic video projection format except ERP format, although the severe stretching in ERP format is alleviated, there is still a certain degree of pixel bending. Therefore, when calculating the rate distortion cost, the corresponding weight is given to the distortion. In this paper, Cube Map Projection (CMP), Compact Octahedron Projection (COHP), and Segmented Sphere Projection (SSP) projection formats are used for study.

The rate distortion optimization model proposed in this section assigns different weights to the distortion when calculating the rate distortion cost of coding tree unit (CTU) at different positions of each projection format.

$$d'_k = w_k \times d_k, \quad (1)$$

where w_k and d_k are the weighted factor and distortion of the k th CTU, respectively. Thus, the rate distortion cost function of each CTU can be expressed as follows.

$$f_{(RDC)}' = d_k' + \alpha r_k = w_k \times d_k + \alpha r_k, \quad (2)$$

where α is the slope, and r_k is the bitrate of the k th CTU.

Assuming that a video frame is divided into N CTU during encoding, then its rate distortion cost function is defined as follows.

$$f'(RDC) = \sum_{k=0}^{N-1} (f_{(RDC)}') = \sum_{k=0}^{N-1} (d_k' + \alpha r_k) = \sum_{k=0}^{N-1} (w_k \times d_k + \alpha r_k), \quad (3)$$

where w_k is the weight of distortion d at the calculation of rate distortion cost of the k th CTU, which is obtained by calculating the average weight of all pixels of the current CTU, that is

$$w_k = \frac{1}{s_{CTU}^2} \sum_{m=0}^{s_{CTU}-1} \sum_{n=0}^{s_{CTU}-1} wp_{m,n}, \quad (4)$$

where s_{CTU} is the size of CTU, and $wp_{m,n}$ is the pixel weight of the m th row and n th column of the video frame, which is obtained according to the conversion algorithm of spherical video to each projection format.

3.1. Weight for ERP Format. For ERP format, $wp_{m,n}$ is the scale factor from the ERP region to the spherical region, that is, the weight, which is only related to the height of the current CTU row. Let W be the width of a frame of video, H be its height, and $cWidth$ be the height of the current pixel. Assuming that the radius of the sphere is R , the latitude of the current pixel is θ , and the radius of cross-section with angle θ is r .

According to the projection algorithm in ERP format, its weight can be obtained as follows.

$$wp_{m,n} = \frac{2\pi r}{2\pi R} = \frac{r}{R} = \cos \theta. \quad (5)$$

At the same time, θ can be calculated as follows.

$$\theta = \frac{W-1-2n}{2} \times \frac{\pi}{W}. \quad (6)$$

Since sine function is even function, which can be obtained as follows.

$$wp_{m,n} = \cos \left(\frac{2n-W+1}{2} \times \frac{\pi}{W} \right). \quad (7)$$

As can be seen from the Equations (5) to (7), at the two ends of the ERP format, that is, the north and south pixels of the spherical video, have little weight, but the middle pixel row of the ERP format has the largest weight.

3.2. Weight for CMP Format. The CMP format is composed of six square faces, and the pixel weight of the corresponding position of each square face is the same, so only one face should be considered. Let each square face width and height be S .

From the microscopic point of view, the projection algorithm is the change from $dmdn$ to $d\varphi d\theta$, and the relationship between them can be described as $dmdn = J(\varphi, \theta) d\varphi d\theta$, where $J(\varphi, \theta)$ is the Jacobian determinant.

$$J(\varphi, \theta) = \frac{\partial(m, n)}{\partial(\varphi, \theta)} = \begin{vmatrix} \frac{\partial m}{\partial \varphi} & \frac{\partial m}{\partial \theta} \\ \frac{\partial n}{\partial \varphi} & \frac{\partial n}{\partial \theta} \end{vmatrix}. \quad (8)$$

At this point, according to the projection relationship between the sphere and a plane in CMP, we can conclude that

$$m = \tan \varphi, \quad (9)$$

$$n = \frac{\tan \theta}{\cos \varphi}. \quad (10)$$

Since the area projected on the square surface is equal to $dmdn$, while the corresponding area on the sphere is actually $\cos \theta |d\varphi d\theta|$, so the weight of a point (m, n) on the surface of CMP format can be obtained as follows.

$$wp_{m,n} = \frac{\cos \theta}{|J(\varphi, \theta)|} = \frac{\cos \theta}{\cos^2 \varphi \cos^3 \theta} = (\cos \varphi \cos \theta)^3 = \left(1 + \frac{d^2}{R^2} \right)^{-3/2}, \quad (11)$$

where R is the radius of the circle tangent to the cube, and $R = S/2$ by this time, and the distance d between the points on the surface (m, n) and the center of the surface is defined as follows.

$$d^2 = \left(m + \frac{1-S}{2} \right)^2 + \left(n + \frac{1-S}{2} \right)^2. \quad (12)$$

So the weight of each pixel position is defined as follows.

$$wp_{m,n} = \left(1 + \frac{(m+1-S/2)^2 + (n+1-S/2)^2}{(S/2)^2} \right)^{-3/2}. \quad (13)$$

3.3. Weight for COHP Format. COHP format is composed of triangular faces, and the weight of the corresponding position of each triangular face is the same. Therefore, it is the same as CMP format to calculate one surface weight. Suppose the side length of each triangular face is T , and the weight of COHP is defined as follows.

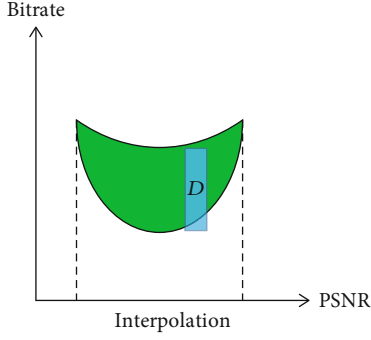

 FIGURE 2: Schematic diagram of $\Delta\text{Bitrate}$.

 TABLE 1: Results for ERP format with weight w_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	-2.3608%	-2.2318%	-2.2441%	-2.4912%
Carphone	-4.1459%	-4.1765%	-4.1789%	-5.0126%
Foreman	-6.0678%	-6.4791%	-6.5110%	-6.8817%
Gaslamp	-0.4624%	-0.4972%	-0.4338%	-0.5241%
Highway	-0.6324%	-0.6223%	-0.6765%	-0.6004%
KiteFlite	-0.1580%	-1.5002%	-1.3900%	-0.1326%
Average	-2.3046%	-2.5845%	-2.5724%	-2.6071%

 TABLE 2: Results for ERP format with weight w'_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	-2.6775%	-2.7609%	-2.5573%	-2.8699%
Carphone	-4.2016%	-4.4195%	-4.0142%	-4.1277%
Foreman	-5.7500%	-6.5608%	-6.1569%	-6.8102%
Gaslamp	-0.4325%	-0.5496%	-0.5874%	-0.6636%
Highway	-0.5998%	-0.6338%	-0.8012%	-0.7890%
KiteFlite	-1.7631%	-1.6584%	-1.7433%	-1.8867%
Average	-2.5708%	-2.7638%	-2.6434%	-2.8579%

 TABLE 3: Results for SSP format with weight w_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	-0.0284%	-0.0456%	-0.0897%	-0.7446%
Carphone	-0.1918%	0.1189%	0.3012%	-0.4058%
Foreman	0.5462%	-0.2165%	0.0218%	-0.6915%
Gaslamp	0.0610%	0.0299%	0.0347%	-0.0705%
Highway	-0.1209%	-0.0448%	0.0459%	-0.1202%
KiteFlite	0.2008%	0.3267%	0.4167%	0.4468%
Average	0.0778%	0.0281%	0.7306%	-0.2643%

 TABLE 4: Results for SSP format with weight w'_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	-0.0454%	-0.0398%	-0.1919%	-0.6989%
Carphone	-0.0741%	-0.0574%	-0.1215%	-0.4365%
Foreman	-0.1563%	0.1550%	0.1384%	-0.5898%
Gaslamp	-0.5156%	0.0012%	0.0301%	-0.0574%
Highway	-0.0754%	0.0899%	0.1257%	-0.1002%
KiteFlite	-0.1887%	0.3418%	0.3352%	-0.3878%
Average	-0.1759%	0.0818%	0.0527%	-0.3784%

$$R = \frac{T}{\sqrt{6}},$$

$$d^2 = \left(m + \frac{1-T}{2}\right)^2 + \left(n + \frac{3-2\sqrt{3}T}{6}\right)^2. \quad (14)$$

The weight of a point (m, n) on each triangular surface is defined as follows.

$$wp_{m,n} = \left(1 + \frac{(m+1-T/2)^2 + (n+3-2\sqrt{3}T/6)^2}{(T/\sqrt{6})^2}\right)^{-3/2}. \quad (15)$$

Since COHP format contains operations such as rotation, segmentation, and combination of triangles, the weight of the whole frame can be obtained by rotation and segmentation and combination of the triangular face with the weight obtained.

3.4. Weight for SSP Format. The SSP format contains invalid fields and the weight is 0. Except invalid regions, the weight of other regions is defined as follows.

$$wp_{m,n} = \begin{cases} \frac{2 \cos \theta}{\pi - 2|\theta|}, \theta \in \left(\frac{\pi}{4}, \frac{\pi}{2}\right] \cup \left[-\frac{\pi}{2}, -\frac{\pi}{4}\right) \\ \frac{2\sqrt{2}}{\pi}, \theta \in \left[-\frac{\pi}{2}, -\frac{\pi}{4}\right] \end{cases}. \quad (16)$$

In this paper, the weight w_k multiplied by calculating distortion for each CTU of each projection format is modified, and the expected value of the weight is changed to 1 by dividing each weight by the expected value. In this paper, the distortion of rate distortion calculation process is directly modified without the operation of adjusting Lagrange multiplier. If the distortion becomes smaller or too large, the bit rate and video quality cannot be well balanced, so the deviation degree of data should be moderate. After a lot of experiments, the variance of the weight is controlled within 0.15 in order to achieve a good balance between bits and video quality without modifying the Lagrange multiplier.

TABLE 5: Results for CMP format with weight w_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	0.4248%	0.5164%	0.3979%	0.2306%
Carphone	-0.1589%	-0.1236%	-1.1265%	-1.1189%
Foreman	1.4259%	1.1866%	1.2876%	1.0075%
Gaslamp	0.1898%	0.0900%	0.1267%	0.1048%
Highway	-0.2046%	-0.1923%	-0.2687%	-0.6789%
KiteFlite	0.3744%	0.5551%	0.4367%	0.1007%
Average	0.3419%	0.3387%	0.1423%	-0.0590%

TABLE 6: Results for CMP format with weight w'_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	-0.3596%	-0.5215%	-0.4662%	-0.8745%
Carphone	-2.2712%	-1.9846%	-0.2370%	-2.6574%
Foreman	-2.1095%	-2.1005%	-2.2689%	-2.2461%
Gaslamp	-0.0667%	-1.2708%	-1.2227%	-1.3201%
Highway	-1.0909%	-1.0909%	-0.9016%	-1.2107%
KiteFlite	-1.2110%	-1.1234%	-0.9647%	-1.2680%
Average	-1.1848%	-1.3486%	-1.0102%	-1.5961%

TABLE 7: Results for COHP format with weight w_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	-1.1394%	-1.1795%	-1.1516%	-1.0678%
Carphone	-1.4500%	-1.2461%	-1.5438%	-2.3684%
Foreman	-0.2268%	-0.3364%	-0.2323%	-0.2941%
Gaslamp	-0.2963%	-0.1919%	-0.3067%	-0.1945%
Highway	0.0634%	0.0361%	0.1891%	-0.1662%
KiteFlite	-0.0091%	-0.0502%	-0.0846%	-0.1108%
Average	-0.5097%	-0.4947%	-0.5217%	-0.7003%

TABLE 8: Results for COHP format with weight w'_k .

Test sequences	$\Delta\text{Bitrate}$			
	DERDO	LM-RDO	D-SSIM-D-MSE	Ours
Basketball	-1.5080%	-1.5398%	-1.5237%	-2.5256%
Carphone	-2.1367%	-2.1397%	-2.0083%	-2.3765%
Foreman	-1.6632%	-1.7665%	-1.6266%	-1.6349%
Gaslamp	-1.0130%	-0.9840%	-1.0487%	-1.0897%
Highway	-1.2464%	-1.2647%	-1.3167%	-1.3136%
KiteFlite	-0.7266%	-0.7787%	-0.8718%	-0.8415%
Average	-1.3823%	-1.4122%	-1.3993%	-1.6303%

TABLE 9: Comparison of w_k and w'_k with four projection formats.

Projection format	w_k	w'_k	Comp
ERP	-2.6071%	-2.8579%	-0.2508%
SSP	-0.2643%	-0.3784%	-0.1141%
CMP	-0.0590%	-1.5961%	-1.5371%
COHP	-0.7003%	-1.6303%	-0.9300%
Average	-0.9077%	-1.6157%	-0.7080%

$$w'_k = \frac{w_k}{E} \times \sqrt{\frac{0.15}{d}}. \quad (17)$$

To verify the above reasoning, all format weights w_k and w'_k are encoded in the experiment, so as to draw relevant conclusions.

4. Experiment Results and Performance Analysis

4.1. Performance Evaluation Metrics. $\Delta\text{Bitrate}$ is used to measure coding performance in this paper, and the calculation method is shown in Figure 2. Different quantization parameters were used for video sequences, and several sets of corresponding Peak Signal to Noise Ratio (PSNR) and bit rate data were obtained after encoding with the proposed method and three benchmarks, respectively. Their curves were plotted in the coordinate system in the figure, respectively. The integral area difference D and the maximum interpolation over PSNR of the two curves is the final $\Delta\text{Bitrate}$ value. A positive $\Delta\text{Bitrate}$ indicates an increase in bitrate, while a negative $\Delta\text{Bitrate}$ indicates a saving in bitrate and improved performance.

4.2. Experimental Parameter Settings. This paper uses the 360lib-HM16.14 as a test platform to integrate the improved RDO technology into 360lib-HM16.14 code for performance testing, where 360Lib-HM16.14 is a 360-degree video tool provided by joint video experts team and integrated with HM or JEM for 360-degree video encoding and decoding. To better verify the quality of panoramic video, spherical video distortion is used to measure the quality. The configuration in coding adopts the given configuration [31] in the universal test environment CTC of panoramic video, and the quantization parameters are 22, 27, 32, and 37, respectively. The experiment is based on the RandomAccess (RA) structure suggested by high efficiency video coding general test conditions and the RandomAccessMain profile (RA-main). The test sequences are selected as Basketball, Carphone, Foreman, Gaslamp, Highway, and KiteFlite.

4.3. Performance Analysis. This section compares the performance of the improved distortion algorithm. Three the-state-of-the-art algorithms are selected for comparison, which are DERDO [28], LM-RDO [29], and D-SSIM-D-MSE [30]. We compare the projection format of the four algorithms with the weighted w_k and the modified w'_k in the calculation of rate distortion in 360lib-HM16.14.

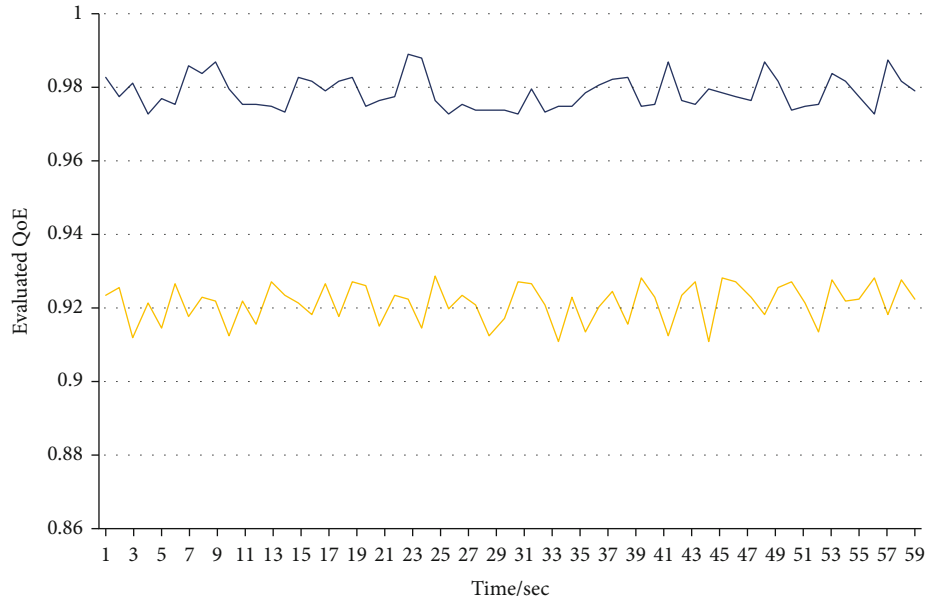


FIGURE 3: Evaluated QoE with w_k and w'_k .

When weighted w_k is added into ERP format, the results of the proposed algorithm compared with the three benchmarks are shown in Table 1. After adding the optimized weight w'_k , the results of the algorithm proposed in this paper compared with the three benchmarks are shown in Table 2.

From the above data, it can find that the introduction of weight w_k in the rate distortion cost calculation of ERP format can improve the coding performance by 2.6071%, while the introduction of optimized weight w'_k can improve the coding performance by 2.8579%.

When weighted w_k is added into SSP format, the results of the proposed algorithm compared with the three benchmarks are shown in Table 3. After adding the optimized weight w'_k , the results of the algorithm proposed in this paper compared with the three benchmarks are shown in Table 4.

From the above data, it can find that the introduction of weight w_k in the rate distortion cost calculation of ERP format can improve the coding performance by 0.2643%, while the introduction of optimized weight w'_k can improve the coding performance by 0.3784%.

Summarizing the results of encoding ERP and SSP formats with weighted w_k and w'_k in the process of rate distortion, we can see that the performance improvement is not significant because the values of w'_k and w_k are close to each other.

When weighted w_k is added into CMP format, the results of the proposed algorithm compared with the three benchmarks are shown in Table 5. After adding the optimized weight w'_k , the results of the algorithm proposed in this paper compared with the three benchmarks are shown in Table 6.

From the above data, it can find that the introduction of weight w_k in the rate distortion cost calculation of CMP format can improve the coding performance by 0.0590%, while

the introduction of optimized weight w'_k can improve the coding performance by 1.5961%, showing a large performance improvement.

When weighted w_k is added into COHP format, the results of the proposed algorithm compared with the three benchmarks are shown in Table 7. After adding the optimized weight w'_k , the results of the algorithm proposed in this paper compared with the three benchmarks are shown in Table 8.

From the above data, it can find that the introduction of weight w_k in the rate distortion cost calculation of COHP format can improve the coding performance by 0.7003%, while the introduction of optimized weight w'_k can improve the coding performance by 1.6303%, showing a large performance improvement.

It can be seen from the results that weight w_k and w'_k are used to code CMP format and COHP format in the process of rate distortion. As the modified weight w'_k changes greatly compared with the original w_k , the modified weight improves the coding performance of these two formats especially significantly. Compared with the original weights, the modified weight can achieve a better balance between the bit rate and video quality.

Table 9 compares the bitrate of the four algorithms in the two weight results of the four projection formats and compares the performance of the two algorithms. Comp is obtained by subtracting the performance change of w'_k from the performance change of w_k . If the value is negative, the performance of w'_k is better than w_k .

As can be seen from the above results, for the four projection formats mentioned in this paper, w'_k can improve the coding performance to a certain extent compared with w_k , up to 0.7080% on average. However, the performance of CMP and COHP formats is significantly improved by

1.5371% and 0.9300%, respectively, compared with w'_k . Therefore, when calculating the rate distortion cost of different projection formats, the weight w'_k of the corresponding format for the distortion at different positions of the image improves the coding performance, which can reach 1.6157% on average.

Furthermore, this paper constructs the metaverse-empowered experience of music and dance, so the Quality of Experience (QoE) is evaluated. As can be seen from Figure 3, the projection format with w'_k (blue line) in VR music and dance scenario experience always has a good QoE, which is always higher than the projection format with w_k (orange line), indicating the effectiveness of the RDO model based on spherical distortion measurement proposed in this paper.

5. Conclusions

In this paper, a rate distortion optimization technique for panoramic video based on spherical distortion measurement is proposed for metaverse-empowered music and dance. In high efficiency video coding, rate distortion optimization plays a decisive role in mode selection. However, for different projection formats of panoramic video, the weight of the blocks with equal weight on the sphere will change after the projection algorithm corresponds to different formats. Therefore, this paper optimizes rate distortion technology according to different features of different projection formats and introduces corresponding weights in the process of distortion calculation to restore the properties of equal weights of video content on the spherical surface. Compared with the original distortion algorithm, the encoding performance of the proposed panoramic visual frequency distortion optimization model based on spherical distortion measure is improved by 1.6157% on average.

With the gradual development of VR, people have higher expectations on the quality of panoramic video, and video coding becomes more and more important for panoramic video. How to encode and transmit panoramic video is still a problem that cannot be ignored. In this paper, the operation of panoramic video coding optimization still has some limitations. The optimization of rate distortion cannot only add the corresponding weight to the distortion but also modify the Lagrange multiplier to find a better model. Inter-frame prediction can also be further optimized and improved.

Data Availability

All data used to support the findings of the study is included within the article.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] T. C. Chin, E. Coutinho, K. R. Scherer, and N. S. Rickard, "MUSEBAQ: a modular tool for music research to assess musicianship," *MUSICAL CAPACITY, MUSIC PREFERENCES, AND MOTIVATIONS FOR MUSIC USE, MUSIC PERCEPTION*, vol. 35, no. 3, pp. 376–399, 2018.
- [2] Y. Qi, Y. Liu, and Q. Sun, "Music-driven dance generation," *IEEE ACCESS*, vol. 7, pp. 166540–166550, 2019.
- [3] N. F. Bernardi, A. Bellemare-Pepin, and I. Peretz, "Dancing to "groovy" music enhances the experience of flow," *Annals of the New York Academy of Sciences*, vol. 1423, no. 1, pp. 415–426, 2018.
- [4] R. T. Solberg and A. R. Jensenius, "Group behaviour and interpersonal synchronization to electronic dance music," *Musicae Scientiae*, vol. 23, no. 1, pp. 111–134, 2019.
- [5] Z. Wang, "Music Choreography Algorithm Based on Feature Matching and Fragment Segmentation," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 9274098, 9 pages, 2022.
- [6] J. Geng, "Personalized analysis and recommendation of aesthetic evaluation index of dance music based on intelligent algorithm," *Complexity*, vol. 2021, 15 pages, 2021.
- [7] P. Yao, "Key frame extraction method of music and dance video based on multicore learning feature fusion," *Scientific Programming*, vol. 2022, 8 pages, 2022.
- [8] F. A. Raposo, D. M. de Matos, and R. Ribeiro, "Assessing kinetic meaning of music and dance via deep cross-modal retrieval," *Neural Computing & Applications*, vol. 33, no. 21, pp. 14481–14493, 2021.
- [9] K. Chen, Z. Tan, J. Lei et al., "ChoreoMaster," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 1–13, 2021.
- [10] R. T. Solberg and N. Dibben, "Peak experiences with electronic dance music: subjective experiences," *PHYSIOLOGICAL RESPONSES, AND MUSICAL CHARACTERISTICS OF THE BREAK ROUTINE, MUSIC PERCEPTION*, vol. 36, no. 4, pp. 371–389, 2019.
- [11] C. Howlin, S. Vicary, and G. Orgs, "Audiovisual aesthetics of sound and movement in contemporary dance," *Empirical Studies of the Arts*, vol. 38, no. 2, pp. 191–211, 2018.
- [12] S. M. Park, Y. G. Kim, and A. Metaverse, "A metaverse: taxonomy, components, applications, and open challenges," *Components, Applications, and Open Challenges, IEEE ACCESS*, vol. 10, pp. 4209–4251, 2022.
- [13] M. S. Smith and R. Brooks, "Meta offers nothing new to the metaverse," *IEEE Spectrum*, vol. 59, no. 1, pp. 22–22, 2022.
- [14] A. Jovanovic and A. Milosavljevic, "VoRtex Metaverse platform for gamified collaborative learning," *Electronics*, vol. 11, no. 3, p. 317, 2022.
- [15] S. Wu, X. Chen, J. Fu, and Z. Chen, "Efficient VR video representation and quality assessment," *Journal of Visual Communication and Image Representation*, vol. 57, pp. 107–117, 2018.
- [16] J. Lv, X. Wang, K. Ren, M. Huang, and K. Li, "ACO-inspired information-centric networking routing mechanism," *Computer Networks*, vol. 126, pp. 200–217, 2017.
- [17] Y. He and Q. Zhang, "Automatic generation algorithm analysis of dance movements based on music-action association," *CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND APPLICATIONS*, vol. 22, no. S2, pp. 3553–S3561, 2019.

- [18] X. Guo, Y. Zhao, and J. Li, "Dancelt: music-inspired dancing video synthesis," *IEEE Transactions on Image Processing*, vol. 30, pp. 5559–5572, 2021.
- [19] G. Sun, Y. Wong, Z. Cheng, M. S. Kankanhalli, W. Geng, and X. Li, "DeepDance: music-to-dance motion choreography with adversarial learning," *IEEE Transactions on Multimedia*, vol. 23, pp. 497–509, 2021.
- [20] B. Burger and P. Toiviainen, "Embodiment in electronic dance music: effects of musical content and structure on body movement," *Musicae Scientiae*, vol. 24, no. 2, pp. 186–205, 2020.
- [21] H. Ahn, J. Kim, K. Kim, and S. Oh, "Generative autoregressive networks for 3D dancing move synthesis from music," *IEEE ROBOTICS AND AUTOMATION LETTERS*, vol. 5, no. 2, pp. 3501–3508, 2020.
- [22] D. Conklin, M. Gasser, and S. Oertl, "Creative chord sequence generation for electronic dance music," *APPLIED SCIENCES-BASEL*, vol. 8, no. 9, p. 1704, 2018.
- [23] G. Valle-Perez, G. Henter, J. Beskow, A. Holzapfel, P. Y. Oudeyer, and S. Alexanderson, "Transflower," *ACM Transactions on Graphics*, vol. 40, no. 6, pp. 1–14, 2021.
- [24] J. P. Ferreira, T. M. Coutinho, T. L. Gomes et al., "Learning to dance: a graph convolutional adversarial network to generate realistic dance motions from audio," *COMPUTERS & GRAPHICS-UK*, vol. 94, pp. 11–21, 2021.
- [25] Y. Wang, D. Liu, S. Ma, F. Wu, and W. Gao, "Spherical coordinates transform-based motion model for panoramic video coding," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 98–109, 2019.
- [26] C. Wu, H. Zhao, and X. Shang, "Octagonal mapping scheme for panoramic video encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2402–2406, 2018.
- [27] Z. Chen, M. Li, and Y. Zhang, "Recent advances in omnidirectional video coding for virtual reality: projection and evaluation," *Signal Processing*, vol. 146, pp. 66–78, 2018.
- [28] C. Herglotz, A. Heindel, and A. Kaup, "Decoding-energy-rate-distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 171–182, 2019.
- [29] F. Zhang and D. R. Bull, "Rate-distortion optimization using adaptive Lagrange multipliers," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3121–3131, 2019.
- [30] Y. Li and X. Mou, "Joint optimization for SSIM-based CTU-level bit allocation and rate distortion optimization," *IEEE Transactions on Broadcasting*, vol. 67, no. 2, pp. 500–511, 2021.
- [31] Y. Liu, L. Yang, M. Xu, and Z. Wang, "Rate control schemes for panoramic video coding," *Journal of Visual Communication and Image Representation*, vol. 53, pp. 76–85, 2018.