International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-13

https://doi.org/10.70917/ijcisim-2026-1007
Article

The Music Contextual Teaching Method: A Study on the Practical Approaches of How Musical Elements Facilitate Immersive Learning of Historical Knowledge from an Interdisciplinary Integration Perspective

Zaichao Zhai [1,*], Asipova Nurbubu Asanalievna [2], Chenjie Li [3], Baoxiang Lu [4] and Yali Huang [5]

[1] Kyrgyz State University named after I. Arabaev, Bishkek, 720001, Kyrgyzstan
[2] Kyrgyz National University named after J. Balasagyn, Bishkek, 720023, Kyrgyzstan
[3] Sichuan University of Culture and Arts, Mianyang, Sichuan, 621000, China
[4] Guangxi Normal University,541004,China
[5] Hunan Vocational College of Foreign Languages,410011,China
* Correspondence author: 18337070670@163.com

Abstract: Music contextual teaching is increasingly emphasized by the educational community. This paper builds a music teaching system architecture for audio processing. Kinect is used to obtain the music teaching scene data, correct the depth camera and color camera, and use OpenNI for coordinate conversion to transform the two-dimensional depth data into a three-dimensional spatial point cloud. The virtual scene 3D modeling data is rendered to generate the free viewpoints of the music teaching 3D classroom. The results show that the method in this paper obtains a good post-alignment positional attitude for the point cloud of the classroom object, the average MAE is only 3.538, and the average alignment time is 31.84 ms. Meanwhile, the classroom reconstruction positional error of this paper's algorithm is reduced by 39.43% on average compared with that of the BundleFusion algorithm. Through the teaching experiment, the historical knowledge achievement using virtual reality teaching improved by 9.23 points on average, which is a significant achievement improvement compared to traditional teaching methods.

Keywords: music contextual teaching; virtual reality; Kinect; 3D scene modeling

1. Introduction

Throughout today's world, is in a period of great development, great change and great adjustment, this is an era of economic globalization, but also an era of cultural diversification, we have to continue Chinese civilization, inherit Chinese culture, carry forward the Chinese spirit, convey the power of China, and promote the common progress of mankind [1-2]. Therefore, China's education department points out the need to give full play to aesthetic education in art classrooms. It is especially important to cultivate students' aesthetic ability and core literacy, and students are the key period to establish the worldview, outlook on life and values, which requires educators to pay attention to aesthetic education in addition to moral and ideological education, and the music curriculum is an important part of aesthetic education, so it is highly valued by all sectors of society [3-5].

The concept of contextual teaching was first proposed in 1989 in the book Contextual Cognition and Learning Culture [6]. There are many different expressions about the concept of contextual teaching in the book, and Chinese academics generally believe that "contextual teaching is a teaching mode created from the dialectical relationship between context and situation, context and rhetoric,

context and reasoning, and context and comprehensive development, creating typical scenes and combining emotional activities with cognitive activities [7-9]. Conceptual definition of music contextual teaching, music contextual teaching refers to the music teacher artificially created in the context of teaching, which is a permeated with the concept of humanistic education, “living space”, is a kind of optimization of the educational environment [10]. In music teaching, teachers take the teaching content as the main body, purposely create a vivid scene with certain emotional characteristics and concrete image of things, so that the students in the environment to enjoy learning, stimulate interest in learning, and help students better understand the content of the teaching and learning methods [11-13]. In addition, in music context teaching, music teachers need to carefully analyze the lyrics, rhythm, tempo, intensity, musical notation and other specific factors in the music content of music teaching materials [14-15]. They should also fully explore the musical mood and emotional characteristics of the teaching materials, choose appropriate contextual teaching methods, mobilize students' emotions and feelings, and make learning an active and conscious activity for students [16].

In this paper, we first build a music teaching system containing access layer, data processing layer, data storage layer, scene management layer and application layer, which contains feature extraction module and feature processing module. Secondly, we obtain virtual 3D depth data, realize RGB-D data conversion, use Kinect sensor to obtain 3D scene data for music teaching, use OpenNI function for coordinate conversion, and obtain 3D surface information of objects through point cloud features. Finally, the 3D coordinates from the model 3D coordinates are converted to 2D coordinates to generate 3D scene free viewpoints. Through the 3D music teaching simulation, the immersive feeling of music context teaching is realized.

2. Virtual reality-based music contextualized teaching system

2.1. Music Teaching System Construction

2.1.1. System architecture

The music teaching system provides a variety of music learning services, online instruction, virtual environment learning and intelligent evaluation. In order to realize its functions, the whole platform adopts a five-layer architecture, from bottom to top: access layer, data processing layer, data storage layer, scene management layer and application layer.

(1) Access layer: The access layer mainly includes audio access and video access. Audio access is realized through voice capture devices (e.g. microphone) for sound input; video access is realized through video capture devices (e.g. camera) for video input.

(2) Data processing layer: It includes audio processing system and teacher guidance system. The audio processing system realizes the extraction of sound characteristic data and the comparison of two sets of base sound data.

(3) Data storage layer: realizes the storage of audio and audio feature data and video storage. The storage interface adopts the commonly used ODBC and JDBC data access methods, and the audio and video file interfaces provide file access services through the encapsulated file IO system of the Java platform.

(4) Scene management system: including virtual scenes and real scenes. The virtual scene is a virtual learning environment and characters created by the software. Realistic scenes are realized by accessing realistic videos.

(5) Application layer: realizes regular online guided learning, virtual environment learning and intelligent evaluation of independent learning.

2.1.2. Audio processing

The audio processing system includes two parts, the feature extraction module and the feature processing module.

The feature extraction module, shown in Figure 1, includes a fundamental frequency recognizer and a tone length acquirer. The fundamental frequency recognizer has built-in fundamental frequency recognition methods such as: cepstrum method, harmonic peak method, cyclic histogram method, wavelet transform method and parallel processing method. The extraction of fundamental frequency data features is performed according to the method selected by the teacher in the teacher guidance system for acquiring fundamental tones. The tone length acquirer uses the algorithm $T = \Delta N \cdot 1/(f_s / 2^q)$ to extract the time value of the tone length, where ΔN is the number of samples between the endpoints of the two tones, q is the depth of the wavelet decomposition, f_s is the

frequency at which the signal is initially sampled.

The feature processing module includes a frequency-to-pitch converter, a pitch comparator, and a tone length comparator. The frequency pitch converter converts by:

$$X = [12 \times \lg(y / 27.5)] / \lg 2 + 1 \tag{1}$$

where y is the fundamental frequency and X is the corresponding pitch.

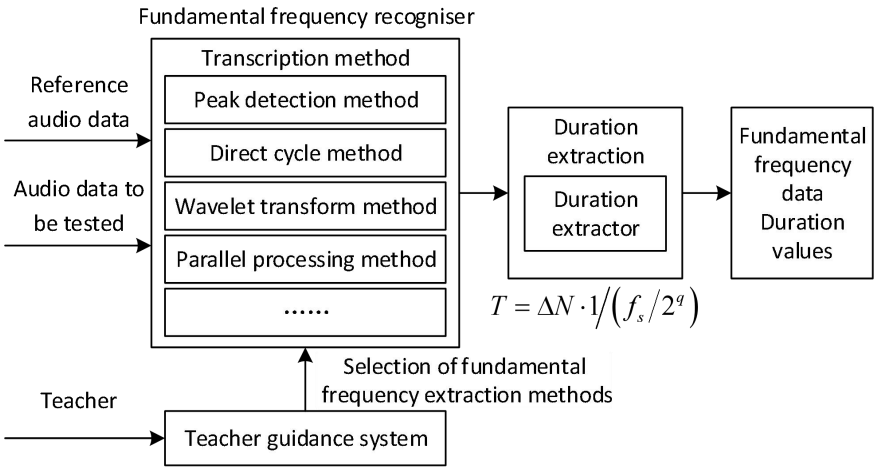


Figure 1. Feature extraction module

2.2. Immersive music teaching virtual 3D scene building

Depth data is the basis of virtual 3D reconstruction technology, its good or bad will affect the accuracy of the reconstructed 3D model and the performance of the scene details. Therefore, the acquisition and processing of depth data in the process of 3D reconstruction is very important.

2.2.1. Acquisition of RGB-D data

Kinect has three cameras, the one in the middle is a color camera and the two on the sides are infrared cameras. The two cameras on both sides work together, one is responsible for transmitting infrared light, the other is responsible for receiving infrared light, so as to obtain the depth image of the surrounding environment.The principle of Kinect data acquisition is shown in Fig. 2.The process of acquiring depth data by Kinect is that the infrared projector projects near-infrared light, and the infrared light encounters the object to be measured or obstacles blocking the scene, which will produce a laser scattering spot on the surface of the object. The laser scattering will be received by the infrared camera after reflection, and after the decoding and calculation of the PS1080 chip, the depth data on the surface of the object or scene will be obtained.

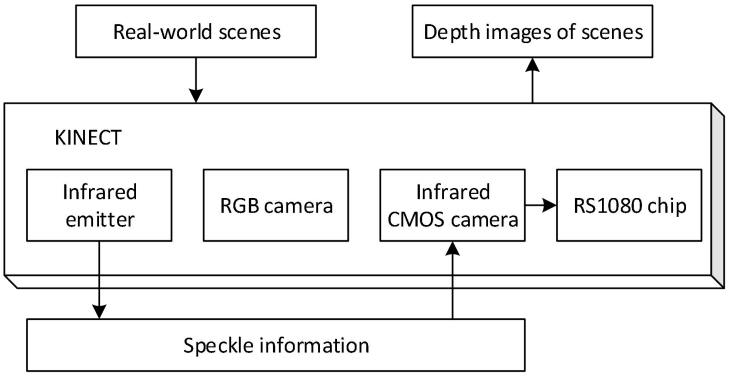


Figure 2. A system for obtaining depth images with Kinect

Since the Kinect has internally corrected the radial distortion of the depth camera and the color camera, the acquired 2D depth information can be directly converted to 3D point cloud data. The pixel

value of the acquired depth image represents the distance of the object from the camera, which is an integer between $[0,255]$, with 0 representing infinity and 255 representing the closest object to the camera.

2.2.2. Camera imaging model

The imaging principle of the Kinect sensor is similar to small hole imaging, and this subsection introduces the Kinect imaging principle using the pinhole model [17]. Assume that P represents any object in space, the pinhole plane is the vertical plane where the camera's point of view is located, and the plane where p is located is called the imaging plane, which is the image of P passing through the pinhole model. The ray passing through and perpendicular to the image plane and the pinhole plane is called the optical axis, which passes through the optical center. f is the distance between the two image planes, the horizontal component of the distance between P and the pinhole plane is Z, and x is the height of the object P after imaging. From the above, the following equation is obtained:

$$\frac{x}{f}=\frac{X}{Z} \tag{2}$$

is the image pixel coordinate system, with the smallest unit being a pixel of the image. The O point represents the origin of the image pixel coordinate system, and u and v are the rows and columns of the image. Take O_0 as the origin of the image coordinate system, x represents the horizontal coordinate and y represents the vertical coordinate.

The coordinate transformation relation between $P(u,v)$ and $P(x,y)$ is:

$$\begin{cases} u=\dfrac{x}{k}+u_0 \\ v=\dfrac{y}{l}+v_0 \end{cases} \tag{3}$$

The origin of the camera coordinate system is the center of light, and its coordinate axes are parallel to those of the image coordinate system. For a spatial point (X,Y,Z) under the camera coordinate system, the coordinate transformation between it and the corresponding image coordinate point is shown in the following equation:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}=\frac{1}{Z_c}\begin{bmatrix} f & -f\cot\theta & 0 & 0 \\ 0 & \dfrac{f}{\sin\theta} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{4}$$

where θ is the angle between the image coordinate system and the camera coordinate system, and f is the focal length.

Combining equations (3) and (4) shows that the transformation formula between the two coordinate systems is as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}=\frac{1}{Z_c}\begin{bmatrix} \dfrac{f}{k} & -\dfrac{f}{k}\cot\theta & u_0 & 0 \\ 0 & \dfrac{f}{l\sin\theta} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}=\frac{1}{Z_c}AP_c \tag{5}$$

where (k,l,u,v,f,θ) is the in-camera parameter and A is the in-parameter matrix.

The transformation relationship between the world coordinate system and the camera coordinate system for the same pixel point is shown in the following equation:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3\times3} & t_{3\times1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{6}$$

where $R_{3\times3}$ is the rotation transformation matrix, $t_{3\times1}$ is the translation matrix, and (R,t) is the outer parameter matrix of the camera.

The combination of the above two equations leads to the following transformation formula between the image plane coordinate system and the world coordinate system:

$$s\tilde{m} = A(R\ t)\tilde{M} \tag{7}$$

where $\tilde{m} = (u\ v\ 1)^T$ are the coordinates of a pixel point in the image pixel coordinate system, $\tilde{M} = (X_w\ Y_w\ Z_w)^T$ the world coordinates of the point, and s is the scale factor.

The mapping point of a point $P(x,y,z)$ in the image plane under the world coordinate system is $P(u,v)$. The coordinate transformation process is as follows: firstly, the point in the world coordinate system is transformed to the camera coordinate system by rotation and translation transformation matrix, then the point P is projected to the image coordinate system according to the principle of small-hole imaging, and finally the coordinate in the image pixel coordinate system, i.e., $P(u,v)$, is obtained by calculating the conversion relationship between pixels and metric units.

2.2.3. Coordinate transformation of depth data

The data captured by the Kinect camera is two-dimensional depth data, and in order to obtain three-dimensional point cloud data, the captured two-dimensional depth data needs to be transformed into a point cloud in three-dimensional space. In this paper, we utilize the function under OpenNI for coordinate transformation on PC to transform the points in the depth image to the camera coordinate system with Kinect depth camera as the coordinate origin.

The transformation relationship of the pixel points on the depth image to under the world coordinate system is shown below:

$$X = (i - w/2) \times (Z - dis) \times scale \times (w/h) \tag{8}$$

$$Y = (j - h/2) \times (Z - dis) \times scale \times (w/h) \tag{9}$$

where w is the height of the image and h is the width of the image. i and j are the horizontal and vertical coordinates of a point on the image. X, Y and Z are the coordinates of the point in the world coordinate system, Z represents the distance from the point on the object to the camera, and dis and $scale$ are constants with values of -11.5 and 0.0032, respectively. The points in the depth image are transformed to the relative spatial points in the scene through the projection transformation, and thus the three-dimensional surface information of the object can be obtained.

Kinect-based scene reconstruction The final reconstructed 3D model of the scene needs to be saved in the form of a point cloud. Point cloud data contains X, Y, Z spatial position information, but also contains R, G, B color information. The operation of point cloud data is actually the operation of the collection of 3D spatial points. The point cloud can be divided into two categories of disordered point cloud and ordered point cloud according to its arrangement characteristics, and the point cloud can be divided into low-density point cloud and high-density point cloud according to the sparseness of the point cloud.

In computer graphics, there are various formats to represent point clouds acquired by mesh models and depth sensors, which roughly include PLY, STL, OBJ, X3D and other formats.

PCD format is a point cloud format under PCL point cloud library, compared with other point cloud data formats, PCL library point cloud file format its advantages are reflected in the following points:

(1) Compared with other point cloud formats, the ability to store and process ordered point clouds is relatively strong, which can be applied to some applications with high real-time performance.

(2) The point cloud data is stored in binary form, and the storage and reading of the point cloud becomes faster.

(3) Suitable for multi-category data format, which makes the point cloud operation and storage more adaptable.

(4) The n -dimensional histogram of the feature descriptors is very important for 3D recognition and computer applications.

(5) Adaptation of point cloud libraries with the help of editing the point cloud's format, so that the best performance of PCL applications can be obtained, avoiding the latency caused when converting different file formats into the PCL internal format.

2.3. Free Viewpoint Generation for Music Teaching 3D Scenes

The 3D model file PLY file is a polygon file format, which is used to store image objects described as a collection of polygons.PLY file mainly consists of two information: vertex and face, where the vertex information usually contains several vertex attributes such as vertex color, normal vector, etc. In this paper, we read the information of PLY file and draw it to the screen to complete the visualization of 3D model. In this paper, we read the PLY file information and draw it to the screen to complete the visualization of the 3D model. The conversion process from the 3D coordinates of the model to the 2D coordinates displayed on the screen is shown in Figure 3. The 3D coordinates are converted to 2D coordinates and displayed on the screen mainly by mode-vision transformation, projection transformation and viewport transformation.

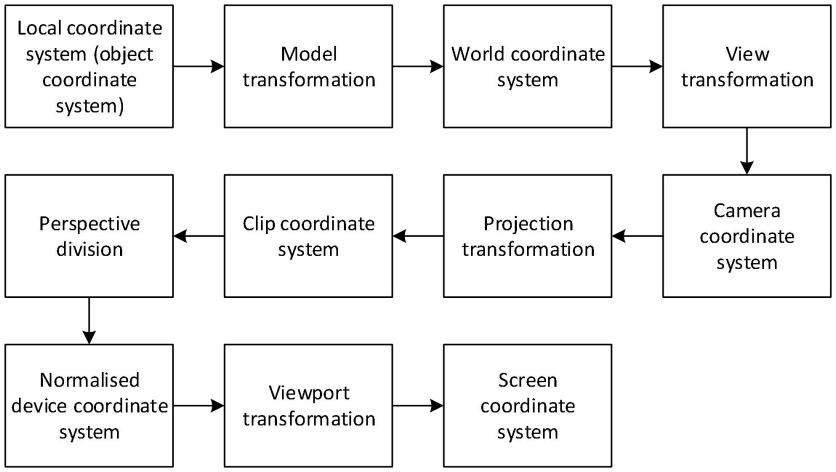


Figure 3. The process of converting 3D coordinates to screen coordinates

Firstly, the local coordinate system is converted to the camera coordinate system by the mode-view transformation, as shown in Equation (10). Where, $\left(X_{obj}, Y_{obj}, Z_{obj}, W_{obj}\right)$ is the coordinate of a point under the object coordinate system, $\left(X_{eye}, Y_{eye}, Z_{eye}, W_{eye}\right)$ are the coordinates in the camera coordinate system, M_{view} is the view matrix, and M_{model} is the model matrix.

The camera coordinate system is then transformed to the cropping coordinate system by a projection transformation, which defines a view body such that the excess outside the view body is cropped off to determine the field of view. The dimensionality is reduced by the use of projection in order to display a three dimensional object on a monitor with a two dimensional image. In this paper, the method of perspective projection is used. Perspective projection is used to obtain a visual effect close to a real three-dimensional object by drawing or rendering on a two-dimensional plane, which realistically reflects the spatial image of the form. The standard cube is the standardized equipment coordinate system. After completing the projection transformation and then converted to the normalized equipment coordinate system through perspective division, as shown in equations (11) to (12). $\left(X_{clip}, Y_{clip}, Z_{clip}, W_{clip}\right)$ is the coordinate under the clipping coordinate system, $\left(X_{ndc}, Y_{ndc}, Z_{ndc}\right)$ is the coordinate under the coordinates in the normalized device coordinate system. $M_{projection}$ is the projection matrix. The formulas are as follows:

$$\begin{pmatrix} X_{eye} \\ Y_{eye} \\ Z_{eye} \\ W_{eye} \end{pmatrix} = M_{modelView} \times \begin{pmatrix} X_{obj} \\ Y_{obj} \\ Z_{obj} \\ W_{obj} \end{pmatrix} = M_{view} \times M_{model} \times \begin{pmatrix} X_{obj} \\ Y_{obj} \\ Z_{obj} \\ W_{obj} \end{pmatrix} \tag{10}$$

$$\begin{pmatrix} x_{clip} \\ y_{clip} \\ z_{clip} \\ w_{clip} \end{pmatrix} = M_{projection} \times \begin{pmatrix} x_{eye} \\ y_{eye} \\ z_{eye} \\ w_{eye} \end{pmatrix} \tag{11}$$

$$\begin{pmatrix} x_{ndc} \\ y_{ndc} \\ z_{ndc} \end{pmatrix} = \begin{pmatrix} \frac{x_{clip}}{w_{clip}} \\ \frac{y_{clip}}{w_{clip}} \\ \frac{z_{clip}}{w_{clip}} \end{pmatrix} \tag{12}$$

Finally, through the viewport transformation, it is converted to the screen coordinate system, which also completes the 2D display of the 3D model on the screen, as shown in Equation (13):

$$\begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} = \begin{pmatrix} \frac{w}{2} x_{ndc} + (x + \frac{w}{2}) \\ \frac{h}{2} y_{ndc} + (y + \frac{h}{2}) \\ \frac{f-n}{2} z_{ndc} + \frac{f+n}{2} \end{pmatrix} \tag{13}$$

After reading and rendering the 3D model data based on OpenGL and completing the visualization of the 3D model, a camera system is defined in space, including the position of the camera in space, the direction of observation, a vector pointing to the right of it, and a vector pointing to the top of it. The super 3D classroom simulation is accomplished by moving the camera to create a sense of moving in the classroom scene with the first viewpoint. The rotation of the camera in 3D space is represented by Euler angles. There are three types of Euler angles: pitch, yaw and roll. The pitch angle indicates how much you look up and down, the yaw angle indicates how much you look left and right, and the roll angle indicates how you roll the camera.

The OpenGL library is commonly used for graphics rendering. The 3D model data is loaded and rendered using OpenGL to increase the speed of 3D model visualization.OpenGL uses a CS model, which is a CPU-GPU model, where the CPU inputs vertex and texture information to the GPU, and the GPU outputs an image to be displayed on the monitor. The vertex data is passed to the graphics rendering pipeline. Vertex data is a collection of vertices, and after defining the vertex data, it is necessary to store these vertices in memory, and this memory is managed through the Vertex Buffer Object (VBO), which will store a large number of vertices in the GPU memory. Since it is very time consuming to send data from the CPU to the graphics card, by using the Vertex Buffer Object to send a large batch of data to the graphics memory at one time, when the data is sent to the graphics card's memory, the vertex shader can access the vertices immediately, thus achieving the effect of optimization and acceleration, and improving the speed of the 3D model visualization.

After reading the 3D model file data through the CPU model, the model is rendered using the OpenGL rendering pipeline. Firstly, the vertex shader will convert the 3D object coordinate system to the camera coordinate system, and compute the vertex color information, normal vector information and some other information, and then assemble all the vertices output from the vertex shader into tuples through the tuple assembly, and then these tuples will be used as the input of the geometry shader to construct new tuples or other tuples to generate other shapes, and then finally the rasterization stage will convert these tuples into the RGBA four-channel tuples, which will be used as the input of the geometric shader. Finally, the rasterization stage converts these tuples into pixels consisting of four channels of RGBA, which are mapped onto the screen for display and are cropped to discard pixels that are out of view, improving loading efficiency. The fragment shader finally performs texture sampling to

assign the texture information of the model to the pixels, which contains 3D scene data such as lighting, shadows, and so on, and is used to calculate the color values of the pixels that are finally displayed on the screen.

3. Test and analysis of virtual scene reconstruction for music contextual teaching

3.1. Point cloud sampling alignment results

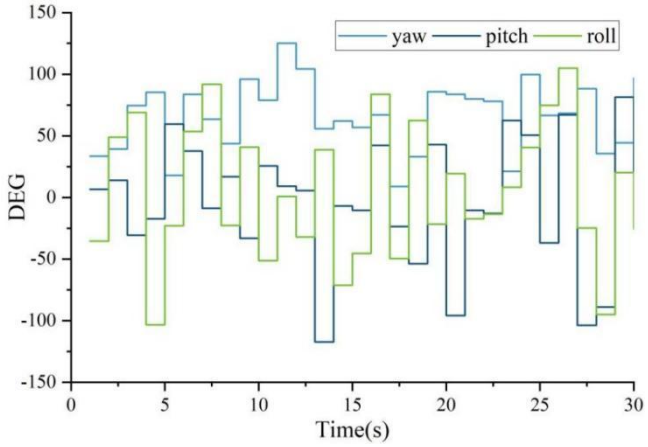
In order to verify the applicability of the point cloud alignment method in this paper to different objects, the collected point cloud data of the indoor environment of a music classroom were tested and preprocessed to be used directly for algorithmic experiments. The downsampled indoor point cloud maps were tested in the experiments by using the weights obtained from the training model of the dataset, and the results of the point cloud alignment accuracy are shown in Table 1. During the alignment test, the overall performance of the model in real data is still evaluated by the mean absolute error and alignment time. The method in this paper can also have good coarse alignment accuracy for sparse point cloud data. At low downsampling ratios, the error keeps increasing as the number of points decreases. When testing in the downsampled hat point cloud, the lack of features in the brim portion of the point cloud causes it to have a larger error in a specific direction, but it still allows the point cloud to obtain a good post-alignment bitmap. The average error is 3.538, and the average alignment time is 31.84 ms. Moreover, after downsampling can make the alignment time reduce significantly, which reduces the time complexity of the algorithm. Therefore, for larger size point clouds, coarse alignment can be performed using downsampling to obtain faster alignment speed.

Table 1. Point cloud registration accuracy results

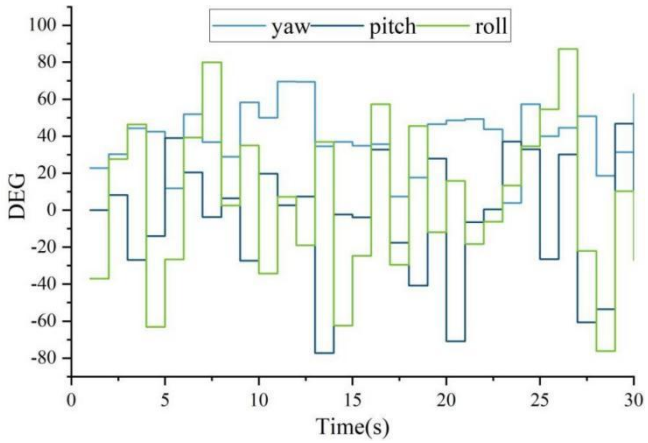
Sampling ratio	Registration error	Registration time (ms)
75%(bag)	2.701	106.37
50%(bag)	3.036	122.4
25%(bag)	1.268	96.52
75%(chair)	3.425	137.46
50%(chair)	2.605	102.67
25%(chair)	2.12	106.84
75%(hat)	3.61	145.14
50%(hat)	3.445	171.51
25%(hat)	9.635	89.04

3.2. Validation of the modeling effect of music teaching scene

In order to verify the 3D reconstruction algorithm on the indoor scene 3D reconstruction of real-time requirements, handheld Kinect device on the music classroom real-time scanning reconstruction experiments, scanning at the same time in the computer 3D reconstruction algorithm processing. The BundleFusion algorithm was selected as a comparison, and the positional error comparison of different algorithms is shown in Figure 4. Mainstream BundleFusion algorithm reconstruction of the camera trajectory is more compact, many areas of the camera trajectory trajectory overlap and drift, in this paper's algorithm results, the camera's trajectory is more stretched, the position information is clear, and greatly avoids the reconstruction of the camera trajectory offset and positional overlap and other problems. Compared with the BundleFusion algorithm, the position error of this paper's algorithm is significantly reduced during reconstruction, and the average error of all positions is reduced by 39.43%, and in the position error map of the BundleFusion algorithm, the curve has more abnormal peaks and the curve smoothness is not good, while the position error curve of this paper's algorithm is obviously more gentle, and the smoothness has been greatly improved.



(a) BundleFusion



(b) The algorithm in this article

Figure 4. Comparison of position errors for different algorithms

The relative position error curves of the two algorithms are shown in Figure 5. It can be concluded that, compared with the BundleFusion algorithm, this paper's algorithm has been improved to carry out the reconstruction of the position error is significantly reduced, and the BundleFusion algorithm's position error graph, the curve of the anomalous peaks are more, the curve smoothness is not good, and this paper's algorithm is processed by the position error curve is obviously more gentle, and the smoothness has been greatly improved.

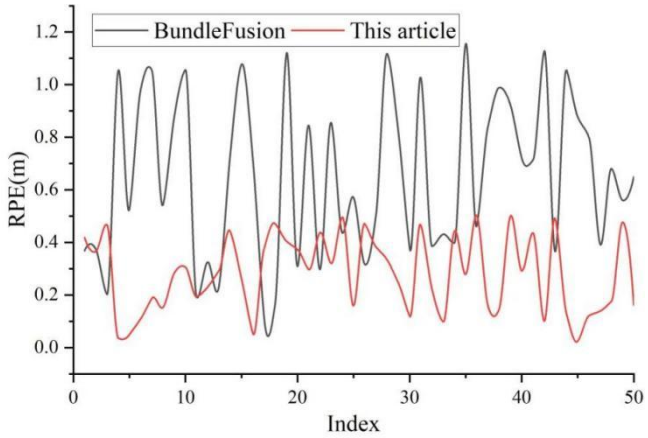


Figure 5. The relative pose error curves of the two algorithms

4. Effectiveness of music contextualized teaching

This paper sets up a music contextual teaching experiment and selects two ninth grade classes in the elementary school of the school as the experimental subjects. Through communicating with the school teachers and referring to the usual performance of the classes, it is understood that these two classes are of similar level and in similar situation, and these two classes are taught by the same teacher. Finally, the experimental subjects were determined as class A and class B. Class A has 45 students and class B has 44 students, and the number of students in the two classes is similar, totaling 99 students. In this study, Class A was set up as the experimental class and Class B as the control class, and Class A was taught with the virtual reality music teaching method, while Class B was taught with ordinary teaching media, and the teaching effectiveness was reflected by the results of the test on the knowledge of music history.

4.1. Analysis of Instructional Pretests

Prior to the lesson, pre-test papers were distributed to Class A (experimental class) and Class B (control class) and the students were guided to actively complete them. The quiz scores were collected and compared and analyzed to test whether there was any difference in the initial ability of the two classes. An independent samples t-test was performed on the scores of the two classes. Comparison of the pretest scores of the experimental and control classes is shown in Figure 6.The mean score of class A (experimental class) was 63.08 with a standard deviation of 16.833 and a standard error mean of 2.657, and the mean score of class B (control class) was 62.89 with a standard deviation of 19.358 and a standard error mean of 3.326.The mean scores of the two classes were similar. An independent samples t-test was conducted on the pre-test scores of the two classes, and the p-value of the probability of significance was 0.507 > 0.05, indicating that there was no significant difference between the pre-test scores of the students in the experimental and control classes. Both classes had essentially the same level of music history knowledge prior to the experiment.

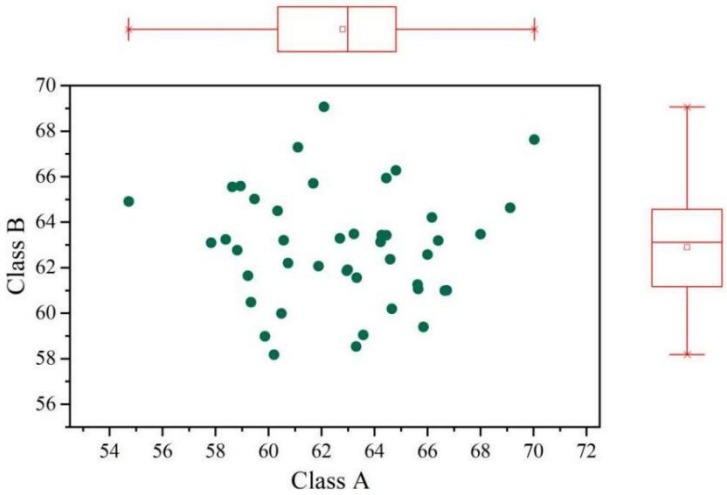


Figure 6. The pre-test scores of the experimental class and the control class

4.2. Analysis of Instructional Post-tests

After the new class was taught, post-test papers were distributed to Class A (experimental class) and Class B (control class), and the students were guided to actively complete them. The quiz scores were collected and compared and analyzed to test whether there was a significant difference in the mastery of knowledge and skills in the experimental class after using the augmented reality virtual musical instrument teaching resources, and the results of the paired samples t-test on the pre- and post-test scores of the experimental class are shown in Figure 7. The p-value of the probability of significance of the pre- and post-test scores of the experimental class is 0.002<0.05, indicating that there is a significant difference between the pre- and post-test scores of the students in the experimental class. Comparison of the mean scores of the pre- and post-tests revealed that the students' scores increased by about 9.23 points on average after using augmented reality virtual musical instrument teaching resources, and the standard deviation was reduced from 16.833 to 11.845, which indicates that there is a significant improvement in the students' mastery of knowledge in the history of music, suggesting that the use of virtual reality teaching has a certain role in promoting the mastery of students'

knowledge of music.

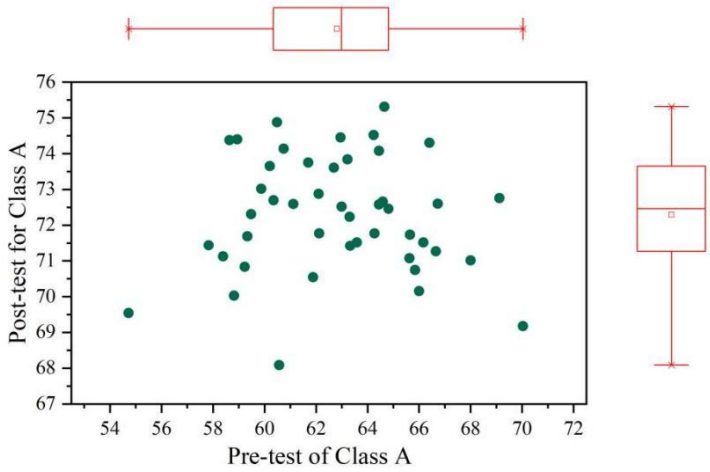


Figure 7. The pre- and post-test scores of the experimental class

At the end of the classroom instruction, an independent samples t-test was conducted on the posttest scores of the two classes, and a comparison of the experimental and control classes is shown in Figure 8. The p-value of the probability of significance of the posttest scores of the experimental and control classes is 0.003<0.05, indicating that there is a statistically significant difference between the posttest scores of the experimental and control classes. In addition, the mean value of the posttest of the experimental class was 72.31 and the mean value of the posttest of the control class was 65.92, which means that the mean score of the posttest of the experimental class was higher than that of the control class.

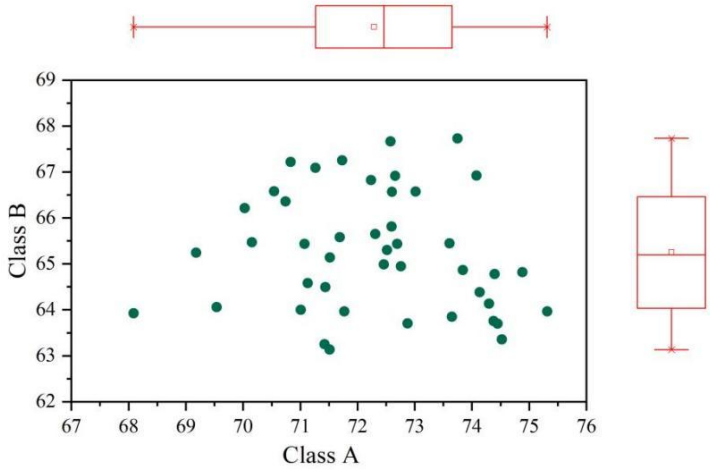


Figure 8. Comparison between the experimental class and the control class

5. Conclusion

In order to realize the music context teaching, improve the historical knowledge immersion to promote the integration of musical elements, the music teaching system is built with the support of virtual reality technology. The research conclusions are as follows:

(1) In this paper, for the point cloud alignment of the music classroom scene, the average MAE for the classroom alignment = 3.538, and the average alignment time = 31.84ms, the alignment method in this paper can reduce the time complexity of the algorithm, and the alignment speed can be obtained. In 3D modeling, this paper's scheme reduces the position modeling error by 39.43% on average than the BundleFusion algorithm, the camera trajectory offset is significantly reduced, and the error curve shows a flat trend.

(2) In the music context teaching experiment, the average scores of class A and class B before the experiment are 62.81 and 62.89 respectively, which are basically similar. In the post-test of the

experimental class, the average score of the experimental class increased by about 9.23 points, and the mastery of music and history knowledge in the experimental class was significantly improved, which indicates that the teaching system established in this paper can effectively enhance the immersion of music teaching and help students strengthen their understanding of historical knowledge.

About the Author

Zaichao Zhai was born in 1999 in Shangcai County, Henan Province, China. He graduated from Shangqiu Normal University with a bachelor's degree and from Kyrgyz State University with a master's degree. Currently, he is pursuing a doctoral degree at Kyrgyz State University, specializing in music and education.

Asipova Nurbubu Asanalievna, Date of birth: 03.02.1946. Academic degree, academic title: Doctor of pedagogical sciences? Full professor. Place of work: Kyrgyz National University named after J. Balasagyn. Work address: Bishkek, Frunze str. 57, Kyrgyz Republic. Professor of the Higher Education Department of the Kyrgyz National University after J. Balasagyn. Professor, Head of the Education Department of the Kyrgyz-Turkish Manas University (KTMU). Supervision of the department activity.

Chenjie Li, highest degree, PhD candidate in Education, title, Assistant Professor, academy of study, Kyrgyz State University (named after I. Arabaev), and current workplace, Sichuan University of Culture and Arts (majoring in music education).

References

1. Liu, Z., & Kalimyllin, D. (2026). Chinese dance education and culture path in the preservation and transmission of cultural heritage to the younger generation. Research in Dance Education, 27(1), 170-183.
2. Jiang, M. (2025). The optimization of curriculum system for music education professionals in the inheritance and transmission of intangible cultural heritage music. Pacific International Journal, 8(2), 139-146.
3. Xiabin, L. (2024). The Strategy and Development of Basic Music Education from the Perspective of Aesthetic Education. Art and Performance Letters, 5(1), 36-41.
4. Jin, X. (2025). Analysis of the Current Situation of Music Education in Primary and Secondary Schools in the Perspective of Aesthetic Education. International Journal of Asian Social Science Research, 2(2), 70-82.
5. Shevtsova, O., Tsarenko, V., Kurkina, S., Voloshyn, P., & Lisovska, T. (2023). The importance of musical and aesthetic education of young people in modern society. Amazonia Investiga, 12(61), 51-60.
6. Shi, S. J., Li, J. W., & Zhang, R. (2024). A study on the impact of Generative Artificial Intelligence supported Situational Interactive teaching on students''flow'experience and learning effectiveness—a case study of legal education in China. Asia Pacific Journal of Education, 44(1), 112-138.
7. Zhang, Z. (2025). The teaching method of STEAM Education-based Audio-visual aesthetics in college vocal music teaching. Frontiers in Educational Research, 8(2).
8. Jiang, H., & Cheong, K. W. (2024). Developing teaching strategies for rural school pupils' concentration in the distance music classroom. Education and Information Technologies, 29(5), 5903-5920.
9. Zhang, S. (2023). Interactive environment for music education: developing certain thinking skills in a mobile or static interactive environment. Interactive Learning Environments, 31(10), 6856-6868.
10. Huanyuan, Z. (2022). Problems in China's college music teaching in recent years. International Journal of Management and Education in Human Development, 2(02), 458-460.
11. Li, Y. (2022). Digital Development for Music Appreciation of Information Resources Using Big Data Environment. Mobile Information Systems, 2022(1), 7873636.
12. Brakhage, H., Gröschner, A., Gläser-Zikuda, M., & Hagenauer, G. (2023). Fostering students' situational interest in physics: Results from a classroom-based intervention study. Research in Science Education, 53(5), 993-1008.
13. Zhan, L., & Hirunrux, S. (2023). The inheritance of national music culture in Chinese university education. Journal of Modern Learning Development, 8(3), 417-425.

14. Shin, J. (2024). Music composition based on creative problem solving: Effects on students' perceptions of creativity and learning music. Bulletin of the Council for Research in Music Education, (240), 65-81.
15. Órbaek, T., & Meen, O. A. "The whole class creates a common unit": a teacher's experience of social inclusion through music. Education as an impulse for social inclusion, 287.
16. Huang, D., Jantharaiit, N., & Thongpanit, P. (2025). Research on the Application of Contextual Improving Literary Knowledge and Attitudes of College Students: Through Situated-Based and Interdisciplinary Instructional Approach. Journal of Education and Learning, 14(3), 221-229.
17. Das, S., Adhikary, A., Laghari, A. A., & Mitra, S. (2023). Eldo-care: EEG with Kinect sensor based telehealthcare for the disabled and the elderly. Neuroscience Informatics, 3(2), 100130.