International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 1228-1242

https://doi.org/10.70517/ijhsa46395

Research on key technology of visual computing-based animation special effect synthesis and scene reconstruction in the new media era

Jicheng Cong[1,*]
[1] Huanghuai University, Zhumadian, Henan, 463000, China
Corresponding authors: (e-mail: congjicheng618@126.com).

Abstract As the use of 3D modeling becomes more widespread, animation 3D techniques are facing more challenges. Neural radiation field provides a new idea to solve this problem by virtue of its ability to reconstruct a realistic 3D scene from sparse 2D images. In this paper, we use the depth camera (i.e., RGBD camera) to obtain the image depth information and generate the 3D point cloud data, combined with the greedy projection triangulation algorithm to reduce the 2D triangular mesh mapping to 3D space to form a 3D space triangular topology network structure, and obtain the 3D surface model of the object. Improve the neural radiation field for animation scene viewpoint synthesis, and design the neural radiation field framework based on space distortion. Comparison tests are carried out to analyze the reconstruction performance and rendering effect of the two algorithms. The greedy projection triangulation reconstruction algorithm takes only 12.043s to process the instance horse, and the maximum deviation distance, average deviation distance, standard deviation and root mean square error do not exceed 0.1 mm. The viewpoint synthesis method based on the improved neuroradiometric field shows certain algorithmic advantages in the viewpoint synthesis of animated scenes.

Index Terms neural radiation field, RGBD camera, 3D point cloud, greedy projection triangulation algorithm, view synthesis

I. Introduction

Animation as an emerging industry has been increasingly emphasized by countries around the world, especially some developed countries, such as Japan animation industry has become a pillar industry to drive national economic growth [1]-[3]. Compared with television art, animation art has a unique artistic charm, which fully satisfies people's visual freshness and develops people's imagination through vivid and lively cartoon animation images and complex and varied artistic expression methods [4]-[7]. And later animation special effects synthesis and scene reconstruction key technology in the animation production process plays a pivotal role, it is all the animation files and footage synthesized into an organized, sequential production process, but also it is an important factor in determining whether an animated film can attract the audience's attention, whether it can have a visually infectious [8]-[11].

Special effects compositing is the key to add a fantasy color and shocking effect to the animation [12]. Scenes and elements that cannot be realized in the real world are presented through special effects synthesis [13]. For example, magical light, huge monsters, gorgeous starry sky, etc. Special effects synthesis can be divided into physical effects and digital effects [14], [15]. Physical special effects are realized by actual props and devices to achieve the effects, such as explosions, pyrotechnics, etc. [16]. And scene is a specific spatial environment that unfolds the plot unit scenes of an animated film, which is one of the basic concepts in animation art design [17], [18]. Film and television animation scene is an important part of the animated film, but also directly affects the artistic style and artistic level of film and television animation [19], [20].

This paper organizes the technical application requirements of animation special effects, obtains the depth information of the image from the RGBD camera, and understands the systematic error and random error calculation method of the depth image. According to the RGBD camera parameters, coordinate transformation is performed to generate 3D point cloud data. Use greedy projection triangulation algorithm to reconstruct into 3D spatial surface model. Briefly describe the basic theory of neural radiation field, propose the Instant-NGP method as the animation view synthesis algorithm, build the neural radiation field framework based on spatial distortion, and use the view synthesis method based on the improved neural radiation field as a new method for synthesizing new viewpoints of animated large-scale scenes. The 3D reconstruction effect of the greedy projection triangulation algorithm and the view synthesis effect of the improved neural radiation field-based view synthesis method on the animated scene are compared and analyzed respectively.

II. Presentation of key technologies

II. A. Animated special effects techniques

II. A. 1) Technical characteristics

Generally speaking, animation special effects technology can make the design and environment of animation more vivid, and to a certain extent, it breaks through the limitations of traditional animation design, which can bring better visual experience to the audience [21].

Computer animation design needs to strictly abide by the rules of design application and technical standards to make the picture presentation more vivid while enhancing the picture texture. Taking the animated film "Jackie Chan Adventures" as an example, special effects technology is fully utilized in the design process, which effectively integrates the animation software with the real character roles, and enhances the animation effect of characters and images.

In addition, from the characteristics of special effects technology, it is a form of art with rendering, which is based on the global illumination technology, after projecting the light onto the surface of the object, it can make the object image produce different degrees of deformation, forming different visual effects.

In short, animation effects technology is a kind of technology that can break through the limitations of time and geography, and can effectively make up for the shortcomings of traditional animation design, and achieve the purpose of improving animation effect.

II. A. 2) Application of special effects technology

(1) Pre-preparation

Making scripts and planning scenes are the core of the pre-preparation work. Scripting refers to presenting the content of the desired animation screen in the form of words. In addition to the script, the prep work also includes determining the art design style, character modeling and music style. After determining the art design style and character modeling, the director then plans the scenes. In the process of planning the scene, emphasis should be placed on highlighting the character traits of the characters and better shaping the characters through details.

(2) Clip production

In the process of clip production, attention should be paid to the selection of scene models and related parameters to ensure the quality of 3D animation works. In recent years, some advanced three-dimensional production software has been applied in the field of animation design in China. Among these software, MAYA and 3D Studio Max are favored by animation designers because of their flexibility, accuracy and developability.

All the building models presented in the animation are drawn using 3D production software, which requires the use of a large number of lights and stage materials. In order to increase the expressive power of the picture, it is necessary to scientifically match the lighting color and material color. The relationship between different colors can create a realistic and full three-dimensional visual effect.

(3) Post-synthesis

The post-synthesis of animation is, in essence, a comprehensive summary of the preliminary preparatory work and segment production work. Post-composition should control each animation image from a global point of view, establish a big-picture view, grasp the overall effect of the picture, use special effects to coordinate the animated characters and scenes, so that the animation picture and plot has coherence and hierarchy. At present, computer animation technology has been applied to movie production, game development and other fields.

II. B. 3D reconstruction of visual scene based on RGBD data

II. B. 1) Depth image generation principle and error analysis

A depth camera is a camera that acquires both information with RGB colors, which is acquired by a normal color camera sensor, and information about the depth of an object within a certain range from the camera. Therefore, depth cameras can also be called three-dimensional cameras, or RGBD cameras [22], [23].

From the optical wave phase difference $\Delta\varphi$ and modulation frequency f of the depth camera, the transmit and receive time difference t is back-calculated with the following expression:

$$t = \frac{\varphi}{\omega} = \frac{(n \cdot 2\pi + \Delta\varphi)}{2\pi f} \tag{1}$$

And the distance d between the object and the camera is computed with the expression:

$$d = \frac{c}{2} \cdot t \tag{2}$$

The above equation can be obtained by association:

$$d = \frac{c \cdot (2n \cdot \pi + \Delta\varphi)}{4\pi f} \tag{3}$$

where c is the speed of light with a magnitude of 3×10^8 m/s. Due to the short depth distance measured and the optical wave power limitations, in general $n = 0$ and the expression for the distance d that simplifies to:

$$d = \frac{c \cdot \Delta\varphi}{4\pi f} \tag{4}$$

By calculating the above equation, the depth value of the corresponding position in the scene can be obtained and stored in millimeters into the corresponding pixel of the depth image.

The systematic error of the depth image E_{sys} can be expressed as:

$$E_{sys} = \frac{1}{n}\sum_{i=1}^{n} d_i - d_t \tag{5}$$

In the above equation, d_i denotes the depth value captured in the i th frame, where the value of i is in the range of [l, n]. The d_i denotes the true depth value in the scene.

The random error calculation formula of the depth camera can be expressed as:

$$E_r = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})^2} \tag{6}$$

where d_i denotes the depth value of the i th frame captured and $1 \le i \le n$, and $\bar{d}$ denotes the average value of the depth of the corresponding pixel position of the n th frame of the captured image.

II. B. 2) 3D color point cloud conversion of 2D RGBD images

A depth image is a two-dimensional image data that records the distance from an object in a scene to a camera and is stored in pixel points as depth values. Since the depth image is closely spaced, a coordinate transformation based on the camera parameters can be performed to generate three-dimensional point cloud data for subsequent three-dimensional modeling. Since the depth image is a closely spaced set of pixel points in a two-dimensional spatial coordinate system and depth values are stored in the pixel points, the point cloud is a set of points in a three-dimensional coordinate system. Therefore, the process of converting the depth image into a point cloud is actually the process of mapping the 2D pixel points into 3D spatial points under the 3D coordinate system, and the schematic diagram of the coordinate conversion principle is shown in Figure 1.

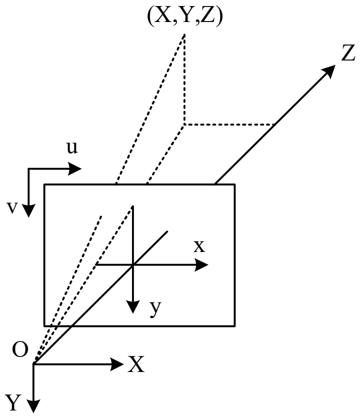


Figure 1: Schematic diagram of coordinate transformation principle

According to the mapping relationship in the previous section and with reference to the principle of pinhole imaging, the following equation can be obtained:

$$z_{dp}\begin{pmatrix}u\\v\\1\end{pmatrix}=\begin{bmatrix}f_{x_{dp}} & 0 & u_0\\0 & f_{y_{dp}} & v_0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{7}$$

where (u,v) and z_{dp} are the row and column numbers and depth values of the depth image pixel points, respectively. $f_{x_{dp}}, f_{y_{dp}}$ is the focal length of the depth sensor in the x,y axis direction. u_0,v_0 is the center coordinate value of the depth image. (X,Y,Z) denotes the corresponding 3D point coordinate values under the transformed 3D coordinate system.

Therefore, the formula for calculating the coordinates corresponding to the x,y,z axes of the converted 3D point cloud can be obtained as follows:

$$\begin{cases}X=z\cdot(u-u_0)/f_{x_{dp}}\\Y=z\cdot(v-v_0)/f_{y_{dp}}\\Z=z\end{cases} \tag{8}$$

The conversion result can be obtained by bringing the camera parameters into the calculation formula.

II. B. 3) Mesh Triangulation Based 3D Surface Reconstruction Algorithm

Point cloud 3D reconstruction is the process of generating a 3D surface mesh model based on the acquired target point cloud processed by an algorithm [24], [25]. Greedy projection triangulation algorithm is an algorithm that quickly establishes the topological relationship of the point cloud through the triangular mesh structure. The specific process of its algorithm is as follows:

STEP1: Establish a 2D plane coordinate system and set the expression of the plane where this 2D plane coordinate system is located in 3D space as:

$$Ax+By+Cz=0 \tag{9}$$

STEP2: Project the 3D point cloud data P into the 2D planar coordinate system to obtain the projected transformed 2D point cloud data $\tilde{P}$, then its projected rotational translation matrix T_p can be expressed as:

$$T_p=T_t\cdot R_x\cdot R_y \tag{10}$$

where T_t denotes the translation transformation matrix in the point cloud projection transformation, i.e:

$$T_t=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\x&y&z&1\end{bmatrix} \tag{11}$$

R_x is the rotation matrix for rotating ω_x by an angle around the x-axis of the two-dimensional planar coordinate system, i.e:

$$R_x=\begin{bmatrix}1&0&0&0\\0&\cos\omega_x&\sin\omega_x&0\\0&-\sin\omega_x&\cos\omega_x&0\\0&0&0&1\end{bmatrix} \tag{12}$$

R_y is the rotation matrix for rotating ω_y by an angle around the y-axis of the two-dimensional planar coordinate system for:

$$R_y = \begin{bmatrix} \cos\omega_y & 0 & -\sin\omega_y & 0 \\ 0 & 1 & 0 & 0 \\ \sin\omega_y & 0 & \cos\omega_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

Then the projection of a 3D point into a 2D planar coordinate system can be expressed as:

$$P^{'} = T_p \cdot P^T \tag{14}$$

where $P^{'}$ is the coordinate value of the 3D point after projection and P is the original coordinate value of the 3D point.

STEP3: Based on the point cloud data $\tilde{P}$ obtained from the projection transformation, a KdTree index is established and a Delaunay-based region growing algorithm is used to construct a 2D triangular topological network. The specific operation is as follows: pick a point p_0 at any point in the point cloud data $\tilde{P}$ and perform a neighborhood search to find and connect its nearest neighbor p_1 to get the line segment $\overline{p_0 p_1}$. Then, the neighborhood of the line segment $\overline{p_0 p_1}$ is searched to obtain the nearest neighbor p_2, and the point p_2 is connected with the line segment $\overline{p_0 p_1}$ end-to-end to form a triangle $P_0 P_1 P_2$. Using this as a base, the algorithm continuously searches outward for the nearest neighboring point of each side of the triangle and forms a new triangle with that side. Until all the points satisfying the requirements of the algorithm are traversed.

STEP4: According to the constructed 2D triangular topological network and the rotation translation matrix T_p obtained from the projection transformation, the 2D triangular network is mapped and reduced to the 3D space to form a 3D spatial triangular topological network structure, and the 3D surface model of the object is obtained.

II. C.New Perspective View Synthesis for Large Scale Scenes

II. C. 1) Instant-NGP

Neural radiation fields are different from traditional 3D reconstruction methods, which usually represent the 3D scene as an explicit form such as a point cloud, voxel, mesh, and so on. Neural radiation field, on the other hand, represents the scene as a continuous function implicitly stored in a multilayer perceptual machine, and the neural network takes the sparse set of pictures with known positional information from multiple viewpoints as input, and is trained to obtain a neural radiation field model that represents the scene, and then it can be synthesized into a clear picture from any given viewpoint based on the neural network.Instant-NGP, as a modified neural radiation field algorithm, dramatically improves the training speed of neural radiation field, and provides a good opportunity for the subsequent reconstruction of neural radiation field. Instant-NGP, as an improved neural radiation field algorithm, greatly improves the training speed of neural radiation field, and lays a solid foundation for the subsequent research on neural radiation field.

(1) Basic theory of NeRF

Neural Radiation Field (NeRF) combines a learnable implicit scene representation with a body rendering ray casting algorithm in computer graphics. Since the body rendering process is microscopic, it can be integrated into the optimization process of neural networks to achieve 2D image supervision.

Implicit neural representation: the NeRF performs an implicit neural representation by optimizing a coordinate-based deep fully connected layer neural network F_0, which takes the low-dimensional coordinates sampled on the ray $x = (x, y, z)$ and the ray direction $d = (\theta, \varphi)$ After position encoding, it is fed into the neural network to be regressed into single-valued bulk density σ and color values c. Coordinate-based MLP is suitable for gradient-based optimization and deep learning, and is orders of magnitude more compact than traditional grid-sampling representations. Coordinate-based MLPs can map low-dimensional coordinates into continuous spatial domains and can be used as implicit shape representations.

To optimize the weights of the multilayer perceptron, the loss function of NeRF is computed by the mean square error between the true color value C_{gt} and the rendered color value $\hat{C}$ for each pixel point as shown in Eq. (15):

$$\mathrm{L} = \sum_{r \in R(P)} \| \hat{C}(r) - C_{gt}(r) \|_2^2 \tag{15}$$

where P denotes the camera pose of the target viewpoint, $R(P)$ denotes the set of all camera rays of the target viewpoint P, r denotes one of the rays, $C_{gt}(r)$ denotes the color value of the ray corresponding to the real image, and $\hat{C}(r)$ denotes the value of rendering color on the ray. The final optimized MLP weights can be used in the implicit neural representation, based on which high quality rendered images can be synthesized at virtual viewpoints different from the input viewpoint location. However, the NeRF algorithm is not generalizable and requires retraining and rendering for new scenes.

Position coding: Standard MLPs are not suitable for coordinate-based low-dimensional vision and graphics tasks, and MLPs themselves have difficulty learning high-frequency functions efficiently, a phenomenon known as “spectral bias”.

The five dimensions of the spatial position sampled on the camera ray and the orientation of the camera ray need to be encoded independently, i.e., the coordinates $x=\left(x,y,z\right)$ and the orientation $d=\left(\theta,\varphi\right)$ need to be encoded in the position, which is shown in Eq. (16)in NeRF:

$$\gamma(x)=[\sin(2\pi\omega_1 x),\cos(2\pi\omega_1 x),\ldots,\sin(2\pi\omega_m x),\cos(2\pi\omega_m x)]^r \tag{16}$$

where ω denotes the frequency of the spatial position after Fourier feature coding, sampled uniformly between 2^0 and 2^{L-1}, where L=10 for spatial position and L=4 for ray direction. This uses sinusoidal positional coding at predefined frequencies to encode each dimension independently, and then combines the resulting vector representations through concatenation. Three-dimensional locations that are otherwise spatially inaccessible to valid information become distinguishable in the Fourier high-dimensional feature space, and the MLP can more easily capture location information in discrete space to infer and learn the structure of the 3D scene.

Image Rendering: NeRF's image rendering is a rendering method that combines classical body rendering and deep neural networks to obtain a more complete and realistic point-of-view image. Instead of using the input point-of-view image after training, the coordinate-based MLP relies on neural networks and microscopic ray projections to synthesize a virtual point-of-view image.

In the NeRF algorithm, firstly, the spatial position x needs to be obtained through camera ray projection and ray sampling, a set of pixels p is randomly sampled from the input 2D image I, and the pixel coordinates (u,v) are transformed to the world coordinate system through the camera model to generate the corresponding camera rays $r(t)$:

$$r(t)=0+td \tag{17}$$

The position of the camera is the origin of the ray $o\in R^3$, and the position of the pixel point on the pixel plane indicates the direction of the camera's line of sight $d\in R^3$ to the scene, and each point on the ray has a volumetric denseness σ and a color value c, and since the wavelength of the ray weakens when it passes through objects of different transparencies, it is necessary to Calculate the cumulative transmittance $T(t)$ corresponding to each sampling point on the ray, as:

$$T(t)=\exp(-\int_{t_0}^{\prime}\sigma(r(s))ds) \tag{18}$$

The cumulative transmittance $T(t)$ represents the cumulative value of the volume density σ of all the objects that passed in front of the ray when the ray reaches a certain object. For objects possessing transparency, the principle of rendering them is to mix the color of transparent objects with the color of opaque objects, and the final pixel color value $\hat{C}(r)$ can be obtained by integrating the cumulative transmittance $T(t)$, volume density σ, and the color value e of all the sample points on the ray, as in:

$$\hat{C}(r)=\int_{t_0}^{t_t}T(t)\sigma(r(t))c(r(t),d)dt \tag{19}$$

In the actual computational process, this continuous integral is numerically approximated using the orthogonal method, with uniform random sampling on the ray according to the stratified sampling, with sampling equations such as:

$$t_i\sim N\left[t_n+\frac{i-1}{N}(t_f-t_n),t_n+\frac{i}{N}(t_f-t_n)\right] \tag{20}$$

where t_n denotes the nearest point on the ray, t_f denotes the farthest point on the ray, and N denotes the number of sampled points. The discretely sampled spatial points t_i are positionally encoded and fed into the MLP to generate the corresponding volume density σ and RGB color value c, and the discrete spatial points on the ray are summed to compute the corresponding color value of the ray, $\hat{C}(r)$, as:

$$\hat{C}(r)=\sum_{i=1}^{N}T_i(1-\exp(-\sigma_i\delta_i))c_i \tag{21}$$

where $T_i=\exp\left(-\sum_{i=1}^{i-1}\sigma_i\delta_i\right)$ denotes the discrete cumulative transmittance, δ_i denotes the distance between the sampling points, and σ_i and c_i denote the bulk density and color values of each sampling point.

(2) Instant-NGP algorithm

Multi-resolution hash coding: Instant-NGP, on the other hand, represents the location information using voxel grid and performs multi-resolution hash coding, which significantly improves the training inference speed of neural radiation field.

The hash encoding resolution is set to a total of 16 layers, i.e., there are a total of 16 levels of voxel resolution changes, and the relationship between the resolution of the corresponding layer and the number of layers is as follows:

$$N_l:=[N_{\min}\cdot b^l] \tag{22}$$

$$b:=\exp\left(\frac{\ln N_{\max}-\ln N_{\min}}{L-1}\right) \tag{23}$$

where b is considered the growth parameter, which is set between 1.38 and 2. For coarser meshes, the T parameter is not needed, and its parameter count is $(N_l)^d\le T$, which ensures a one-to-one mapping relationship, and for finer meshes, the mesh is indexed to the array using a spatial hash function, $h:Z^d\to Z_T$, as shown below. To wit:

$$h(x)=\left(\bigoplus_{i=1}^{d}x_i\pi_i\right)modT \tag{24}$$

where the operator $\oplus$ denotes a different-or operation and π_i denotes a unique large prime, this computation yields a linear congruent arrangement in each dimension to eliminate the effect of dimensionality on the hash value. Finally, based on the relative position of the voxel interior x, the voxel interior corner point eigenvectors are linearly interpolated with the interpolated weights as in Eq:

$$w_l:=x_l-\lfloor x_l\rfloor \tag{25}$$

Image Rendering: The formula for calculating the pixel point color is shown below:

$$C(r)=\int_{t_n}^{t_f}T(t)\cdot\sigma(r(t))\cdot c(r(t),d)dt \tag{26}$$

where $T(t)$ denotes the cumulative light transparency of the ray over the path from t_n to t_f, i.e., the probability that this ray does not hit any particle from t_n to t_f. The formula for $T(t)$ is shown below:

$$T(t)=\exp(-\int_{t_n}^{t}\sigma(r(s))ds) \tag{27}$$

Since continuous integrals are difficult to compute, they are generally approximated by discrete sampling. First, $[t_n,t_f]$ is divided into N equal-length intervals, and then a sample is randomly selected in each interval, and the distance between neighboring samples is denoted by $\delta_i=t_{i+1}-t_i$, and the rendering equation can be expressed as:

$$\hat{C}(r)=\sum_{i=1}^{N}T_i(1-e^{-\sigma_i\delta_i})c_i,\quad T_i=e^{-\sum_{j=1}^{i-1}\sigma_i\delta_i} \tag{28}$$

T_i is the approximate cumulative transmittance.

The difference L between the predicted pixel point color $\hat{C}$ and the real pixel point color $C_{g,t}$ is used as a loss function to train the whole network, and L is computed as:

$$L=\sum_{r\in R}\|\hat{C}(r)-C_{g,t}(r)\|_2^2 \tag{29}$$

II. C. 2) Neural Radiation Field Framework Based on Spatial Distortions

For the borderless large-scale scene under the free trajectory, this paper designs a new neural radiation field framework, which can realize high-quality training rendering effect for the borderless scene under the randomly collected trajectory, and improve the quality of view synthesis of the neural radiation field application in the large-scale scene.

Ⅰ Perspective warping algorithm:

For forward scenes, NeRF uses normalized device coordinate (NDC) warping. Projecting light from the world coordinate system $(o+td)$ to the normalized device coordinate space $(o'+t'd')$ compresses the scene space along the z-axis. The mapping formula for mapping an infinity view into a bounded box, in the NDC space used in NeRF, is shown below:

$$o' = \left(-\frac{f_{cam}}{W/_2} \frac{o_x}{o_z}, -\frac{f_{cam}}{H/_2} \frac{o_y}{o_z}, 1 + \frac{2n}{o_z} \right)^T \tag{30}$$

$$d' = \left(-\frac{f_{cam}}{W/_2} \left(\frac{d_x}{d_z} - \frac{o_x}{o_z} \right), -\frac{f_{cam}}{H/_2} \left(\frac{d_y}{d_z} - \frac{o_y}{o_z} \right), -2n \frac{1}{o_z} \right)^T \tag{31}$$

where W, H are the width and height of the image and f_{cam} is the focal length of the pinhole camera.

After normalized coordinate processing, the original view cone is mapped to the cube [-1, 1] so that t' can be simply sampled linearly from 0 to 1 to obtain a linear sampling of the parallax from n to ∞ in the original space. To wit:

$$contract(x) = \begin{cases} x & \|x\| \le 1 \\ \left(2 - \frac{1}{\|x\|}\right)\left(\frac{x}{\|x\|}\right) & \|x\| > 1 \end{cases} \tag{32}$$

In order to rationally allocate resources for the scene space, this chapter uses perspective warping to divide the space region with the camera viewpoint and project $\square^3 \to \square^3$ on the whole space. Considering the multi-view relationship, using the pixel space of multiple views as a mediator, it can be understood as the projection process of $\square^3 \to \square^{2n} \to \square^3$. Considering n_p sampling points $\{x_j \mid j = 1, 2, \cdots, n_p\}$ uniformly sampled in the original Euclidean space, this paper defines the projection matrix from Euclidean space to pixel space to be G, and the mapping coordinates in the pixel space, $y_j = G(x_j) \in \square^{2n}$, the projection matrix from pixel space to distorted space is M, and the coordinates $z_j = My_j \in \square^3$ in distorted space. Where G is the imaging projection relation, which can be calculated from the camera parameters. And the matrix M is constructed from the first three eigenvectors of the covariance matrix of $\{y_j\}$, i.e., obtained by principal component analysis of the set of projection points $\{y_j\}$.

Ⅱ Scene Space Segmentation

This chapter uses an octree data structure to store the segmented regions, which enables fast searching of regions and retrieval of visible cameras. To construct the octree, this chapter starts with a very large bounding box as the root node. This section sets the size of the bounding box to be 512 times the size of the bounding box at the center of all the cameras, which is able to accommodate very distant skies or other objects. Then, starting from the octree root node, the inspection and subdivision process is performed.

III View Rendering

Based on the above description, in the preparation phase, this section subdivides the raw space according to the camera's view cone and constructs a local warping function based on the selected camera in each sub-region. In the actual rendering phase, this section follows the framework of volumetric rendering by sampling the points on the camera rays and accumulating the colors of the sampled points in a weighted manner. Where the density and color of the sampled points are extracted from a multi-resolution hash grid.

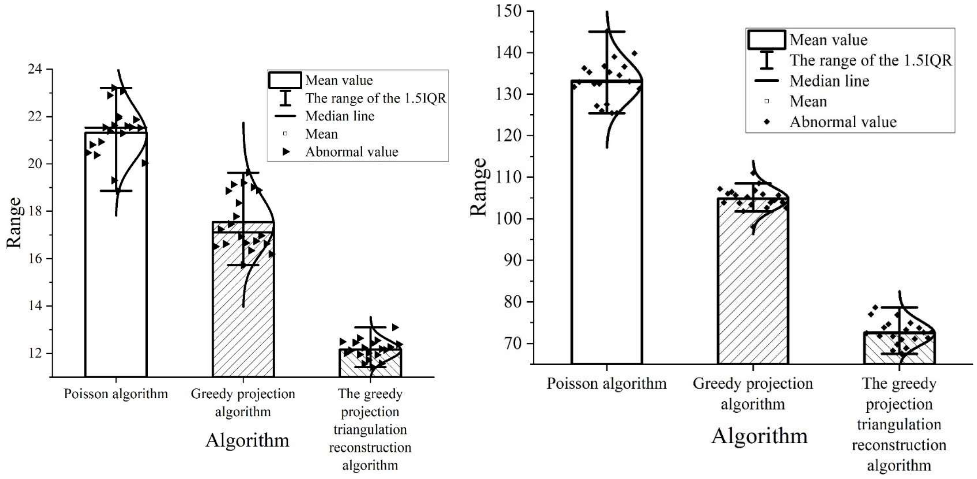
III. Performance analysis of key technologies for animated vision

III. A. Visual 3D reconstruction effect

In order to verify the effectiveness and feasibility of the algorithms in this paper, in this experiment, the software development tools of the experimental platform are Windows 10 operating system, vs2019, pcll.11.0, and CMAKE, VTK, boost libraries required for processing point clouds. Two data sources were chosen for the experimental data:

the standard dataset HORSE from Stanford University and the sculpture data acquired through Trimble TX8 3D laser scanner, respectively.

In order to show the performance of the improved algorithm, the reconstruction time of Poisson algorithm, traditional greedy projection algorithm and greedy projection triangularization reconstruction algorithm are compared. The comparison of the three algorithms' time consumption is shown in Fig. 2, and Figs. (a) and (b) show the reconstruction time consumption of each algorithm for HORSE and sculpture, respectively. The processing time of Poisson algorithm, traditional greedy projection algorithm and greedy projection triangularization reconstruction algorithm for instance HORSE is not more than 25s, and the reconstruction time of each algorithm is lower than the processing time of instance sculpture.In 20 experiments, the reconstruction time of each algorithm for instance sculpture is 132.45s, 105.423s, 72.534s, respectively.Since Poisson algorithm needs to use the moving cubic algorithm principle to extract the equivalent surface, so it takes longer time.



(a) Instance horse processing time

(b) Processing time for instance sculptures

Figure 2: Three algorithms are time-consuming compared

Four quantitative analysis accuracy evaluation indexes were used, i.e., the maximum deviation distance, the average deviation distance, the standard deviation and the root-mean-square error from the point cloud to the model entity were calculated by using Geomagic Studio software, and the number of model facets generated by the three algorithms was displayed using this software, so as to evaluate the accuracy of the reconstruction results of the models generated by the three algorithms.

The reconstruction accuracy of each algorithm is shown in Table 1. Due to the relatively simple structure of HORSE data, the reconstruction accuracy of the three algorithms does not differ much, and the algorithm in this paper is also less than the other two algorithms in the number of triangular facets generated. For the collected sculpture data, due to its complex structure, the accuracy of this paper's algorithm is improved by 1.5mm compared with Poisson algorithm, and 0.71mm compared with the traditional greedy projection algorithm, and it can be concluded that this paper's algorithm meets the accuracy requirements of modeling in most fields.

Table 1: The reconstruction accuracy of each algorithm

Instance	Algorithm	Maximum deviation /mm	Average deviation /mm	Standard deviation /mm	Mean and mean error/mm
Horse model	Poisson algorithm	0.15	0.09	0.07	0.07
	Greedy projection algorithm	0.12	0.04	0.03	0.06
	Improved the rnematic triangulation reconstruction algorithm	0.07	0.03	0.04	0.02
Sculpture model	Poisson algorithm	2.52	0.43	0.36	0.27
	Greedy projection algorithm	1.73	0.25	0.24	0.23
	Improved the rnematic triangulation reconstruction algorithm	1.02	0.16	0.14	0.13

In order to test the effect of the greedy projection triangulation reconstruction algorithm proposed in this paper, the two-by-two aligned point clouds obtained from the improved point cloud alignment algorithm of the Bunny data, the two-by-two aligned point clouds from the Armadillo data, and the two-by-two aligned point clouds from the field view point cloud data are used as the three original point clouds to be reconstructed in this chapter. Comparisons are made to analyze the effect of redundant point removal on surface reconstruction, the effect of point cloud smoothing on surface reconstruction, and the effect of PCA-based normal vector estimation and MLS-based normal vector estimation on surface reconstruction. Finally, we compare and analyze the reconstruction effect of the traditional greedy triangulation algorithm and the improved greedy projection triangulation algorithm in this paper.

The 3D reconstruction times of different point clouds are shown in Table 2.

(1) Analysis of the impact of redundant point removal on the reconstruction effect: the number of point clouds to be reconstructed by Bunny is 81326, and the number of point clouds after removing redundant points by voxelized downsampling algorithm is 48563, which realizes the streamlining of point cloud data.

(2) Analysis of the impact of MLS-based smoothing on the reconstruction effect: After point cloud smoothing, the surface is smoother and more delicate, which better describes the geometric structure of the object surface. Therefore, the MLS smoothing algorithm improves the reconstruction effect.

Table 2: 3D reconstruction time of different clouds

Point cloud	To rebuild the number of clouds	Remove the number of redundant points	Greedy reconstruction time/s	The time of reconstruction of this article
Bunny	81326	48563	190.113	115.632
Armadillo	56903	36952	124.568	72.596
Scene cloud 1	175896	102174	312.854	166.804
Scene cloud 2	176632	113346	336.596	183.541
Scene cloud 2	205307	116894	358.811	204.597

III. B. Animated Scene View Synthesis Effect

III. B. 1) Experimental design and analysis

(1) Experimental Models

In this paper, three models are selected for comparative experiments to verify the effectiveness of the improved neural radiation field-based view synthesis proposed in this paper:

1) instant-NGP

2) instant-NGP 2) TensoRF

3) TensoRF 3) 3D Gaussian Splatting

(2) Training details

The view synthesis algorithm based on improved neural radiation field adopts the loss function and sampling method in NeRF. The Adam optimization algorithm was chosen, and its initial learning rate was set to 0.03, and a learning rate decay factor of 0.1 was set to adjust the training speed. All experiments were performed on an NVIDIA RTX 4090 with 4096 rays processed in each batch for a total of 30,000 iteration steps. The other methods in the comparison experiments also use the above experimental setup.

(3) Experimental data

All models were evaluated using datasets commonly used for new view synthesis, where the Blender dataset is a dataset of single object scenes rendered under absolutely ideal conditions. The ScanNet dataset is a dataset of filmed indoor scenes. In addition, four indoor and outdoor scenes were shot on their own as supplementary experiments. The number of images and descriptions of each dataset are as follows:

nuScenes dataset: 2 scenes, scene-0011 and scene-0931, each scene contains about 240 images, 1 out of every 10 is selected as the test set image, and the images are used as the training set.

Blender dataset: 8 scenes, each scene contains 100 images as training set and 200 images as test set.

ScanNet dataset: 12 scenes were randomly selected for training according to NeRFusion's scene allocation strategy, each scene contains about 1200 images, 1 out of every 8 images is selected as the test set, and the rest of the images are used as the training set.

Customized LiDAR dataset: 4 scenes, the number of images in each scene varies from 200~500 images, 1 out of every 10 images is selected as the test set, and the rest of the images are used as the training set.

(4) Evaluation metrics

Three different metrics, PSNR, SSIM and LPIPS, are used to evaluate the generation quality of new perspective images.PSNR focuses on pixel-level image differences, and the higher the value, the better the quality of the image.SSIM, on the other hand, takes into account the brightness, contrast, and structure, and is more in line with the human eye's visual evaluation standards, and the closer the value is to 1, the higher the quality of the image. LPIPS utilizes deep learning techniques to evaluate the perceptual similarity of images, and the lower the value means the more subtle the perceptual differences between images, which is particularly applicable to the quality evaluation of image processing.

III. B. 2) Comparative analysis of methods

(1) nuScenes dataset

In this section, the view synthesis method based on improved neural radiation field is compared with instant-NGP, TensoRF, and 3D Gaussian Splat-ting.

The PSNR metrics data of each algorithm on the nuScenes dataset are shown in Fig. 3, which demonstrates the three times PSNR evaluation results of different algorithms, respectively. In the figure, green and cyan represent scene-0011 and scene-0931, respectively. Both scene-0011 and scene-0931, the improved neural radiation field-based view synthesis method obtains higher PSNR. In the evaluation of the nuScenes dataset, the improved neural radiation field-based view synthesis method demonstrates the best rendering performance among all the algorithms compared.

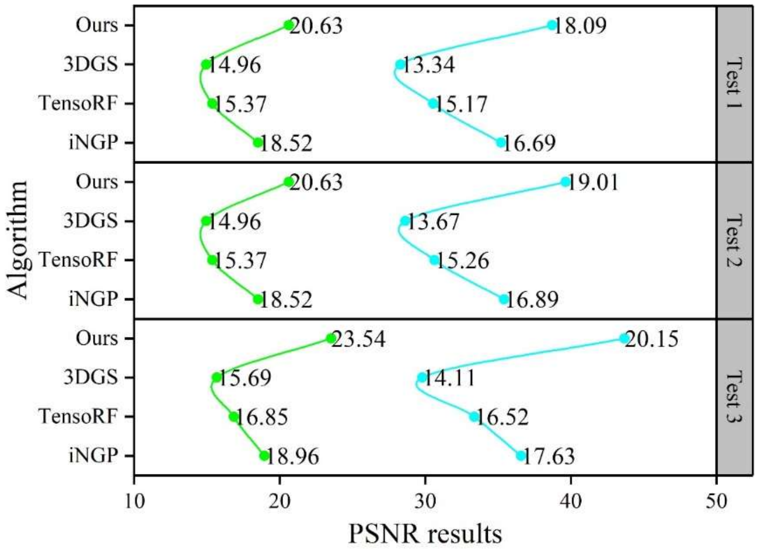


Figure 3: PSNR index data for each algorithm in the nuScenes data set

The SSIM metric data for each algorithm on the nuScenes dataset is shown in Figure 4. Scene-0011 In the first test, the view synthesis method based on the improved neural radiation field obtained a SSIM value of 0.6291, which is an improvement of 0.0719 over the iNGP method.

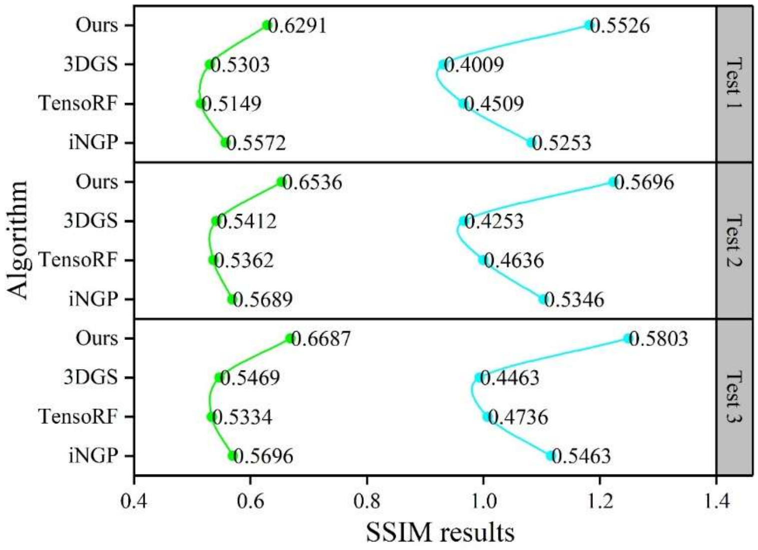


Figure 4: The SSIM index data of each algorithm in the nuscenes data set

The LPIPS metrics data of each algorithm on the nuScenes dataset are shown in Figure 5, from which it can be seen that the LPIPS metrics of the view synthesis method based on the improved neural radiation field are lower than those of the other algorithms. In the scene-0011 test, the LPIPS values of the view synthesis method based on improved neural radiation field are 0.5736, 0.5321, and 0.5967 in that order.

By synthesizing the test results of the three indexes of PSNR, SSIM, and LPIPS, it can be obtained that the view synthesis method based on the improved neural radiation field has the best rendering performance. In particular, the reconstruction task is challenging for objects that are far away from the camera because the actual physical space represented by each pixel is larger, which means that the effective supervision information is relatively reduced. Nevertheless, the view synthesis technique based on improved neural radiation fields accurately reconstructs tall buildings in the distance, maintaining high quality modeling. Similarly, for small details in the near distance, such as tree branches, the modified neural radiation field-based view synthesis method also demonstrates satisfactory reconstruction capability. Tree branches, due to their small size, are very susceptible to camera pose errors, which also illustrates the effectiveness of the pose accumulation correction module.

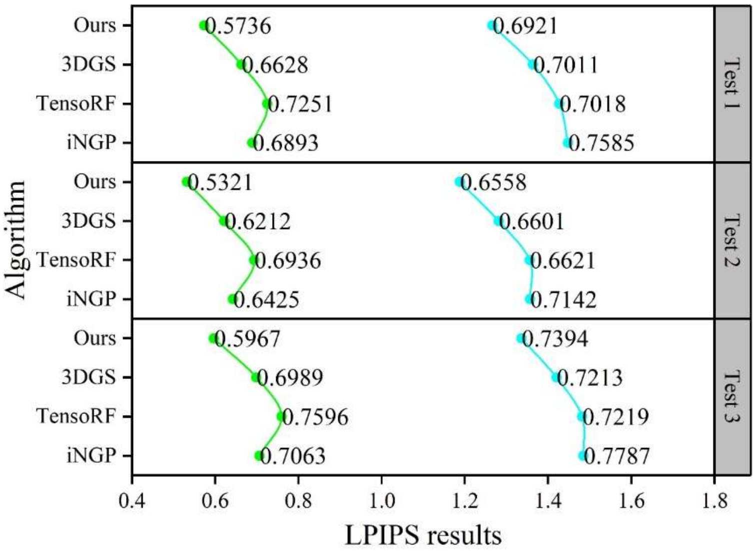


Figure 5: LPIPS index data for each algorithm in the nuscenes data set

(2) ScanNet dataset

In this section, the view synthesis method based on improved neural radiation field is compared with NeRF, TensoRF, and NeRFusion.

The test results of PSNR metrics on the ScanNet dataset are shown in Figure 6. After six tests respectively, it can be obtained that the PSNR value of the view synthesis method based on the improved neural radiation field is larger than that of the NeRF method in different states. And the PSNR values of this paper's method (Ours*30k) are all above 30, and in test 6, Ours*30k=39.19.

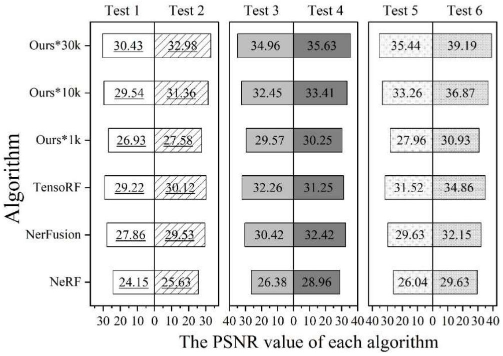


Figure 6: Test results of PSNR metrics in the PSNR data set

The test results of SSIM metrics are shown in Fig. 7, the mean SSIM value of Ours*1k under six tests is 0.8287. It is improved by 0.1819 than the NeRF method and reduced by 0.0308 than the NerFusion method. The mean SSIM value of Ours*10k is 0.8672, which is improved by 0.0077 than the NerFusion method, which indicates that the that the method in this paper is operational.

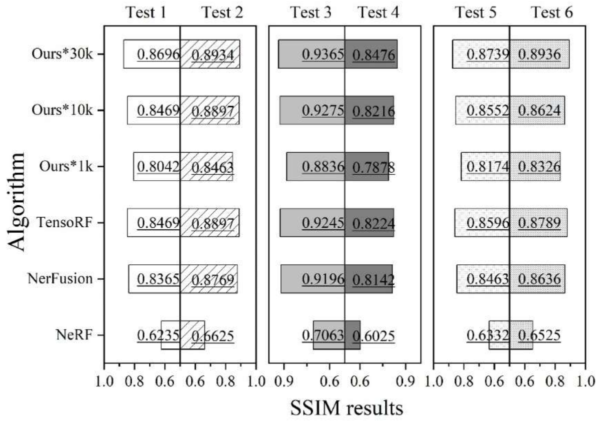


Figure 7: Test results of the SSIM index

The test results of the LPIPS metrics are shown in Fig. 8, and the view synthesis methods based on the improved neural radiation field all outperform the NeRF method in terms of LPIPS metrics.

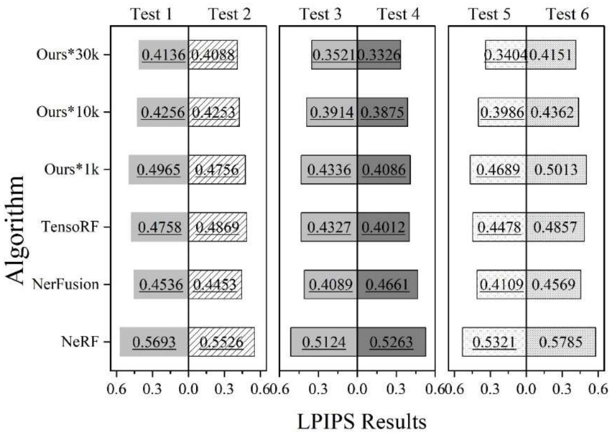


Figure 8: Test results for LPIPS indicators

The comparison with other methods on the ScanNet dataset is shown in Table 3.Ours*1k takes the least amount of time, 1.4min, which is a significant efficiency gain and has an advantage in terms of memory.

Table 3: Contrast with other methods in the scannet data set

Algorithm	Rays ↓	Batch	Mem ↓	Time ↓
NeRF	1100m	4096	12.53GB	>600min
NerFusion	380m	4096	10.31GB	54.2min
TensoRF	380m	2048	9.14GB	115.9min
Ours*1k	6m	4096	4.15GB	1.4min
Ours*10k	50m	4096	4.15GB	16.3min
Ours*30k	110m	4096	4.15GB	47.9min

(3) Blender dataset and custom LiDAR dataset

The comparison experiments on Blender dataset and custom LiDAR dataset are shown in Table 4. The view synthesis method based on the improved neural radiation field also shows a good competitiveness on the Blender dataset and Lidar dataset for small objects.

Table 4: Compare experiments in blender data sets and self-defined lidar data sets

Algorithm	Blender			LiDAR		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
NerFusion	32.63	0.9586	0.0465	19.24	0.5569	0.6503
TensoRF	34.52	0.9689	0.0472	15.87	0.3659	0.7251
Ours*1k	32.59	0.9567	0.0501	21.63	0.5263	0.7012
Ours*10k	35.62	0.9869	0.0463	26.82	0.5667	0.7578
Ours*30k	40.78	0.9921	0.0457	31.26	0.6204	0.8036

IV. Conclusion

In this paper, 2D images are acquired by RGBD camera and converted to 3D point cloud data, and 3D surfaces are reconstructed by applying greedy projection triangulation algorithm. Combined with Instant-NGP setup for spatially distorted scene view synthesis method.

In the time-consuming comparison between Poisson algorithm, traditional greedy projection algorithm and greedy projection triangulation reconstruction algorithm for visual 3D reconstruction effect, Poisson algorithm takes longer time because it needs to use the principle of moving cube algorithm to extract the equivalent surface. The greedy projection triangulation reconstruction algorithm used in this paper takes only 12.043s and 72.534s for horse and sculpture, and the greedy projection triangulation reconstruction algorithm also has better surface reconstruction effect in the three original point clouds to be reconstructed.

In the performance test of the scene view synthesis algorithm, the view synthesis algorithm based on the improved neural radiation field shows certain advantages in the nuScenes dataset, Blender dataset, ScanNet dataset, and customized LiDAR dataset in the three different metrics: SNR, SSIM, and LPIPS, which indicates that the view synthesis algorithm based on the improved neural radiation field can achieve the expected results and the view synthesis algorithm based on the improved neural radiation field can achieve the expected results. This indicates that the view synthesis algorithm based on the improved neural radiation field can achieve the expected effect, and the view rendering performance better meets the demand of animation special effects production.

References

[1] Fan, K. K., & Feng, T. T. (2021). Sustainable development strategy of Chinese animation industry. Sustainability, 13(13), 7235.

[2] Nurjati, E., Rianto, Y., Wulandari, R., & Fatmakartika, O. (2020). Indonesian animation industry: Its mapping and strategy development. International Journal of Business Innovation and Research, 4(VIII), 2454-6186.

[3] Ma, L., Qian, C., Liu, Z., & Zhu, Y. (2018). Exploring the innovation system of the animation industry: Case study of a Chinese company. Sustainability, 10(9), 3213.

[4] Yoon, H. (2017). Globalization of the animation industry: multi-scalar linkages of six animation production centers. International Journal of Cultural Policy, 23(5), 634-651.

[5] Limano, F. (2021). Human and technology in the animation industry. Business Economic, Communication, and Social Sciences Journal (BECOSS), 3(1), 1-7.

[6] Saputra, D. I. S., Manongga, D., & Hendry, H. (2021, November). Animation as a creative industry: State of the art. In 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (pp. 6-11). IEEE.

[7] XIONG, D. (2018). Research on Creative Performance of Animation Art in Advertisement. DEStech Transactions on Social Science, Education and Human Science.

[8] Yusa, I. M. M., Putra, P. S. U., & Putra, I. N. A. S. (2017, November). Sinergy of art and animation technology in multimedia performance art creation entitled sad ripu. In 2017 4th International Conference on New Media Studies (CONMEDIA) (pp. 174-181). IEEE.

[9] Deng, X., Lei, J., & Chen, M. (2021). Application of vr in the experimental teaching of animation art. Mobile Information Systems, 2021(1), 4642850.

[10] Peng, M. (2024). The application of digital media technology in the post-production of film and television animation. Media and Communication Research, 5(2), 129-134.

[11] Junjun, J., & Pillai, M. D. (2025). A STUDY TO UNDERSTAND THE PROBLEMS WITH EFFICIENCY IN THE STAGES OF ANIMATION PRODUCTION. Prestieesci Research Review, 2(1), 343-352.

[12] Liu, Y., Li, L., & Lei, X. (2024). Automatic Generation of Animation Special Effects Based on Computer Vision Algorithms. Computer Aided Design And Applications, 21, 69-83.

[13] Zhang, J., Yang, H., & Zhuang, Z. (2020). Application of Special Effect Technology in Late Synthesis. In International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019: Applications and Techniques in Cyber Intelligence 7 (pp. 197-203). Springer International Publishing.

[14] Peng, S. (2022). Considering Multisensor Data Fusion and Data-Driven Special Effects Animation Simulation. Mobile Information Systems, 2022(1), 2247839.

[15] Li, Y., Feng, C., Yu, H., & Pang, L. (2021). A survey of physics-based character animation synthesis methods. International Core Journal of Engineering, 7(11), 509-521.

[16] Deng, Y. (2021). Fluid Equation-Based and Data-Driven Simulation of Special Effects Animation. Advances in Mathematical Physics, 2021(1), 7480422.

[17] Zhang, J. Q., Xu, X., Shen, Z. M., Huang, Z. H., Zhao, Y., Cao, Y. P., ... & Wang, M. (2021, October). Write-An-Animation: High-level Text-based Animation Editing with Character-Scene Interaction. In Computer Graphics Forum (Vol. 40, No. 7, pp. 217-228).

[18] Jiang, J., & Wang, X. (2024). Animation scene generation based on deep learning of CAD data. Computer-Aided Design and Applications, 21, 1-16.

[19] Song, W., Zhang, X., Guo, Y., Li, S., Hao, A., & Qin, H. (2023). Automatic generation of 3d scene animation based on dynamic knowledge graphs and contextual encoding. International Journal of Computer Vision, 131(11), 2816-2844.

[20] Han, Y. (2024). Research on CG Animation Scene Technology Based on Computer Technology. Procedia Computer Science, 247, 1197-1206.

[21] Mo Bing,Shafilla Subri,Ali Alshehan & Wang Li. (2024). The application and innovation of animation technology in multicultural communication activities that involve Malaysia and China. Cogent Arts & Humanities,11(1).

[22] Hincapié Mauricio,Díaz Christian Andrés,Valencia Arias Alejandro,Güemes Castorena David & Contero Manuel. (2023). Using RGBD cameras for classifying learning and teacher interaction through postural attitude. International Journal on Interactive Design and Manufacturing (IJIDeM),17(4),1755-1770.

[23] Lu Yang,Yu Han,Ni Wei & Song Liang. (2022). 3D real-time human reconstruction with a single RGBD camera.. Applied intelligence (Dordrecht, Netherlands),53(8),11-11.

[24] Bernat Lavaquiol Colell,Alexandre Escolà,Ricardo Sanz Cortiella,Jaume Arnó,Jordi Gené Mola,Eduard Gregorio... & Jordi Llorens Calveras. (2025). A methodology for the realistic assessment of 3D point clouds of fruit trees in full 3D context. Computers and Electronics in Agriculture,232,110082-110082.

[25] Binyu Nie,Wenjie Lu,Yunxuan Feng,Haowen Gao & Kaiyang Lin. (2025). Removing multi-path echoes in underwater 3D reconstruction via multi-view consistency. Pattern Recognition Letters,189,48-55.