

## Retraction

# Retracted: Design of 3D Modeling Face Image Library in Multimedia Film and Television

### Journal of Sensors

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

Copyright © 2023 Journal of Sensors. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] D. Lei and S. Kim, "Design of 3D Modeling Face Image Library in Multimedia Film and Television," *Journal of Sensors*, vol. 2021, Article ID 2221893, 11 pages, 2021.

## Research Article

# Design of 3D Modeling Face Image Library in Multimedia Film and Television

Di Lei and Sae-Hoon Kim 

*Sejong University, 209 Neungdong-ro, Gunja-dong, Gwangjin-gu, Seoul, Republic of Korea*

Correspondence should be addressed to Sae-Hoon Kim; [dlei@nyit.edu](mailto:dlei@nyit.edu)

Received 16 May 2021; Revised 8 September 2021; Accepted 22 September 2021; Published 15 October 2021

Academic Editor: Haibin Lv

Copyright © 2021 Di Lei and Sae-Hoon Kim. 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.

The development of 3D modeling technology has promoted the development of the multimedia film and television industry. This article is aimed at studying the design of 3D modeling facial image library in multimedia film and television, at providing a more comprehensive facial image library for the multimedia film and television industry, at breaking the shackles of the traditional film and television industry with 3D technology, and at continuously surpassing traditional film and television media forms. This article deeply explores the background development of multimedia film and television and the characteristics of the development of new media. Starting from 3D technology, it extracts facial features of characters, transforms image data through deep autoencoders, and uses local binarization mode to perform the original facial image is subjected to texture feature extraction. In this paper, a number of experimental subjects were selected, and the subjects were photographed from the left, front, and right from multiple angles. Through the pinhole camera projection imaging process, the internal and external parameters of the camera were adjusted. In the process of 3D image construction, the image is first selected for feature detection, then the corresponding vector information and geometric conditions are matched to construct a 3D matrix, and the facial structure image is obtained by triangulation. This article compares the 3D production software on the market and selects the Maya platform suitable for building this system. The global constraint information is obtained by training some sample images. When searching the test image, find the appropriate feature point position according to the structural matching degree of the local image. When each search is completed, the global information will be used for constraint, so as to output reasonable feature information. The average residual range of the human face image constructed in this paper is 0.25-0.45, and the maximum residual error does not exceed 4.0. The experimental method in this paper has good stability and robustness. Using the COM transmission model can make experimenters not need to think too much about the underlying details. This face animation-driven simulation scheme can achieve more vivid facial expressions.

## 1. Introduction

With the continuous development of information technology, technological changes have had a profound impact on the development of all walks of life, ranging from a country to a small individual experiencing the innovation brought by information technology. With the continuous development of information technology, the film and television industry has already broken through the development model of traditional theaters and television stations and has entered a

faster development platform with the help of a variety of brand new media forms.

3D face animation is an important part of computer animation research, which mainly studies how to truly imitate human facial expressions and actions [1]. In the process of continuous development and improvement of 3D modeling technology, multimedia film and television have also received more and more influences. Facial expressions, movements, and expressions of characters are more vividly shaped and comprehensive in different scenes and

scheduling the display of orientation. The establishment of a 3D modeling character facial image library will provide more and better choices for multimedia film and television and will also provide the film and television industry with a broader aesthetic perspective.

The combination of multimedia video and 3D technology has aroused more and more scholars' research and discussion. Yaremenko pointed out that the gap between the introduction of digital creativity in the media space and the lack of theoretical argumentation is due to the rapid formation, adaptability, and relevance of the new artistic environment. He aims to analyze the introduction of multimedia into film plots, the exploration of new types of communication spaces, and the changes and adaptations to art forms. He generally described the interactive relationship between the audience and a work of art, and this interactive relationship is reflected in the creative dialogue between the artist and the consumer. Computer-generated 3D face modeling and expression animation technology has been widely used in film production, game entertainment, computer animation, medical diagnosis, man-machine interface, auxiliary teaching, virtual studio, and other fields. However, the design form involves more basic knowledge in the field of film and practical skills when using technology, which is difficult to fully grasp [2]. Yang et al. use a single-axis MEMS mirror as the structured light projector in the three-dimensional modeling system, which has the advantages of small size and low cost. The three-dimensional modeling system based on a single-axis MEMS mirror is unable to project orthogonal graphics and projection distortion, so it is difficult to obtain high accuracy through the existing calibration methods. They proposed a calibration method for a structured light 3D modeling system. The system can only project fringes in one direction, and the projection is distorted. They proposed a curved surface equation, called a curved surface light model, to replace the ideal plane equation as a mathematical model for projecting structured light stripes. Experimental results show that this method can significantly reduce the impact of projection distortion. But the three-dimensional projection is also affected by different light [3]. Annita applied asynchronous learning to a film course in his research. Asynchronous technology was applied in ten conferences uploaded on the university's online platform. At the end of the semester, they conducted a survey to evaluate this approach. His research is aimed at exploring other possible methods of film teaching. This is necessary because teaching films require a lot of physical and social interaction, which is necessary in this era. However, his research did not evaluate the adaptability of students [4].

The innovations of this article are as follows: (1) select the 3D perspective to select the faces of the characters in the film and television industry, and then build an image database to extract more comprehensive features of the characters' images. In the face animation stage, the motion trend of feature points between sequence images is calculated and corresponding to the image feature points adopted in this paper according to the FDP and FAP standards of MPEG-4, so as to drive the animation; (2) use a combination of theoret-

ical research and empirical research method and research on the establishment of the facial image database of multimedia film and television characters.

## 2. Design Method of 3D Modeling Facial Image Library in Multimedia Film and Television

*2.1. Multimedia Film and Television.* Deformation modeling from a single general model is realized according to the deformation model of elasticity. It has the characteristics of simple model, small amount of calculation, and high degree of automation, but the reality of the synthesized specific 3D face model is poor. The so-called multimedia film and television refers to the digital, interactive, and nonlinear transmission mode through the Internet, wireless communication network, and satellite channels, with computers, mobile phones, and digital TV as the main output terminals, under the background of the continuous development of digital technology, network technology, and mobile communication technology in the late 20th century. The three major types include Internet media, mobile phone media, and digital TV media [5, 6].

From the first screening of the movie in 1895 to the birth of the first black and white TV set, the development of film and television art as an art form that integrates audiovisualization has only gone through a short period of more than one hundred years [7]. The progress of science and technology, especially the development of information technology and digital technology, has also brought a new revolution to the film and television industry [8, 9]. Different from the traditional film and television industry that uses film as the carrier, film and television works in the new media era rely on Internet technology and digital technology and huge changes have taken place in the production and distribution links and therefore have a development that the traditional film and television industry cannot have features:

- (1) The low participation rate and wide publishing platform greatly stimulate the participation and creativity of ordinary audiences. Compared with traditional film and television, nowadays, multimedia film and television are edited with a large amount of material and use montage techniques to form a complete narrative structure, and the requirements for editing equipment are not high, which can meet the shooting needs of most people [10, 11]
- (2) Diversified communication channels broke the pattern of media hegemony. The traditional film and television industry has a single communication channel, emphasizing the acceptance of information. However, in the multimedia age, many users have diversified Internet-based publishing platforms, which completely broke the traditional media hegemony pattern, giving viewers the power to evaluate the quality of film and television works and the initiative to obtain content
- (3) The "decentralized" communication model enhances interactivity and satisfies the individual needs of

niche groups. Today, when cultural products are extremely rich, the audience's requirements for cultural products are constantly increasing. The film and television industry is an important source for audiences to obtain cultural products. The mode of film and television communication is also changing to interactive two-way communication. Multimedia provides the film and television industry with a completely different way of development [12, 13]

- (4) The reliance on content is increasing, and "content is king" has become a general consensus in the industry. With the continuous improvement of material living standards, audiences have higher and higher requirements for spiritual life. As the main way to meet people's daily spiritual and cultural needs, the film and television industry is also facing the increasingly critical aesthetic requirements of audiences. Linear combination deformation modeling from multiple general models uses vectors to represent the geometric structure and texture image of each model, respectively. Through linear combination of these vectors, a new 3D face model is obtained, and the coefficients of linear combination are optimized and adjusted to minimize the difference between the two-dimensional projection image of the new model and the input face image

**2.2. 3D Technology.** With the further development of multimedia technology, 3D technology is usually used in various film and television or advertisements to effectively combine text, pictures, sounds, and special effects to bring multiple visual, auditory, and sensory enjoyments to the audience. Multimedia films and televisions add more content and fun. Multimedia films and televisions made with 3D technology usually have better artistic effects and provide audiences with richer choices [14, 15]. Multimedia film and television have interactive and artistic characteristics, which will make it easier for the audience to accept the promoted products and realize the mutual benefit of sellers and consumers. However, based on the analysis of the actual situation of today's new media advertising, 3D technology has not been well used. Also, due to various constraints, many multimedia films and televisions have not achieved the expected results. If these problems are not resolved repeatedly, they will inevitably be affected. The advertisements created by 3D technology have a negative impact and cannot be compensated for benefits [16, 17].

The English abbreviation of 3 dimensions is 3D. Chinese can be interpreted as three dimensions and three coordinates, which is the length, width, and height related to the cube as an example. In other words, 3D is the spatial representation of objects. This statement breaks through the traditional two-dimensional space definition [18, 19]. This virtual technology originated from the United States in the 1950s. Today, whether it is technology or use channels and methods, it is still in the exploratory period. From the perspective of sensory enjoyment, 3D technology is the comprehensive application of multiple technologies to create a

realistic artificial atmosphere for the public and to simulate human senses and behavior in nature to achieve the effect of human and machine communication and interaction [20, 21]. Therefore, whether people are in real life or virtual environment, there will not be much difference. These have not really been realized. They are just the beautiful ideas of scientists. Except for a few 3D technology products, people can enjoy the 3D results that only exist visually, and we have not experienced it personally, but the author believes that with the continuous advancement of science and technology, this fantasy will eventually be realized in the future.

On the other hand, the application of 3D technology is an indispensable link in the creation of current film and television works. To a certain extent, every step of the development and transformation of film and television works is closely related to technological change [22, 23]. Every change has had a certain impact on the creation and design of film and television. Among them, 3D technology has had a new and transformative impact on the special effects, synthesis, and other production links in the film and television production process. The combination of technology and advertising has produced the idea of a new subject we are studying: 3D modeling in the design of the facial image library of film and television characters in the application. At the same time, most of these modeling methods are not fine enough for the description of the model. Relatively speaking, they are not easy to be used as the basic model of facial expression. They are mostly used in applications that do not need too detailed description, such as game scenes.

### 2.3. Character Facial Feature Extraction

**2.3.1. Deep Autoencoder.** Autoencoder (AE) is a neural network used to automatically output the input signal it receives. The self-encoder is a three-layer network structure composed of two information processing units: an encoder and a decoder, and is associated with the input layer, the presentation layer, and the reconstruction layer [24, 25]. In the process of processing network information, the encoder  $g$  first processes the transmission of the input feature  $x$  to generate a representative body of network information with  $k$  features, and then the decoder  $h$  processes the representative body of feature  $k$  to generate such a body. Input a vector of exact reconstruction feature  $t$  with diameter  $x$ . The encoding and decoding steps must be as follows:

$$k = g(x) = d_g(w_g x + b_k), \quad (1)$$

$$t = h(k) = t_h(w_h k + b_t). \quad (2)$$

Among them,  $d_g$  and  $t_h$  are activation functions. The most commonly used activation function is the logistic function, that is,  $d_g(x) = d_h(x) = 1/[1 + e^{-x}]$ ;  $w_g$  and  $w_h$  are the weights of the encoder and the decoder; and  $b_g$  and  $b_h$  are the assumptions of the presentation layer and the reconstruction layer, respectively. The purpose of optimizing autoencoder training is to make the reconstruction  $t$  as close to the input  $x$  as possible, that is, to minimize the error function  $K(x, t)$  [26, 27]. The reconstruction error function is a

measure of the difference between the input  $x$  and the automatic reconstruction  $t$ . The learning goal of AE is to minimize the reconstruction error function corresponding to the input training set  $U$ :

$$I_{AE}(\vartheta) = \sum_{x \in U} I[x, h[g(x)]]. \quad (3)$$

We will set it  $I(\cdot)$  as the reconstruction error function, or it can be set as the square error or cross entropy function of the following two formulas:

$$I(x, y) = \|x - y\|^2, \quad (4)$$

$$I(x, y) = - \sum_{i=1}^{U_x} x_i \log(y_i) + (1 - x_i) \log(1 - y_i). \quad (5)$$

The Sparse Autoencoder (SAE) is obtained by adding a limit  $l_1$  based on AE. By outputting the default mode or the active mode of the hidden layer unit, the inertia expressed by the SAE characteristic can be realized. The Denoising Autoencoder (DAE) adds noise to the training data. Therefore, the autoencoder will learn to remove noise and reconstruct the entrance that is not polluted by noise [28]. DAE learns the strong performance of the entrance signal, and its generation ability is stronger than that of AE. Like DAE, the constructor autoencoder introduces the activation function of the hidden layer into the loss function related to the input of the Jacobi matrix [29]. This can reduce the dependence on information that is irrelevant to the reconstructed data and the training data. Deep Autoencoder (DAE) is a deep network structure model composed of  $n$  autoencoders from bottom to top. DAE includes multiple levels of representation and the process of reconstructing learning characteristics. Using this method, we can obtain fine facial geometry and texture, synthesize various facial deformations, and realize natural and completely controllable facial shape. Uniform mesh and nonuniform mesh are commonly used in polygon modeling technology. Using nonuniform mesh can highlight the details of the face and reduce the computational complexity. First, use the  $x$  input interface to train the lower autoencoder to obtain the corresponding level. The output of the previous autoencoder level is used as the input level of the next autoencoder [30]. After completing all autoencoder training, the output  $g_n$  of the last layer represented by the autoencoder is the vector DAE representing the input  $x$ .

**2.3.2. Local Binary Mode.** “Local” refers to the neighborhood definition of image pixels, “binary” refers to the quantitative relationship between the central pixel and the gray features of the corresponding neighborhood pixels, and “mode” is the texture primitive feature in the image. In the basic LBP, binary codes are used to express the relationship between the grayscale feature value of a central pixel and a local neighborhood point. The binary codes corresponding to all neighborhood points constitute a pattern that describes the local structure information of the central pixel. Here, we study the texture characteristics of a  $3 \times 3$  pixel window with

pixel  $U_c$  as the center, and set the gray feature value of the corresponding central pixel to  $u_c$ , and the gray feature value of the neighboring pixels to be  $u_0, u_1, \dots, u_7$ , respectively. The key research point of generating personalized face model is how to deform the general face model in three-dimensional space or develop a face model that is easy to control the deformation. The local texture feature  $Y$  of the central pixel point can be defined by the function of the central point and the gray value of  $q$  neighboring pixels, namely,

$$Y = y(u_c, u_0, u_1, \dots, u_7). \quad (6)$$

Without losing information, using the gray value of the adjacent pixel to subtract the gray value of the central pixel, you can get the texture image as shown below:

$$Y = y(u_c, u_0 - u_c, u_1 - u_c, \dots, u_7 - u_c). \quad (7)$$

Assuming that the central pixel and the neighborhood difference are independent of each other, the local texture  $U$  can be approximately expressed as

$$Y \approx y(u_c) y(u_0 - u_c, u_1 - u_c, \dots, u_7 - u_c). \quad (8)$$

Among them,  $y(u_c)$  represents the grayscale feature value of the pixel in the local area, and the grayscale difference  $u_q - u_c (q = 0, 1, \dots, 7)$  expresses the grayscale feature value change of the neighborhood point relative to the center point. The gray level  $u_c$  at the center point directly reflects the change of light intensity and has nothing to do with the local texture characteristics of the image. Therefore, in order to remove the influence of illumination changes, the gray distribution  $y(u_c)$  of the center point can be ignored. Then, the texture feature of the image can be directly expressed as a function of the difference:

$$Y \approx y(u_0 - u_c, u_1 - u_c, \dots, u_7 - u_c). \quad (9)$$

The texture defined by the above type is not affected by the change of the gray value, but all adjacent  $Q + 1$  images will be added or deleted to the specified value at the same time, and the represented texture will not be changed. When the values of all pixels are in the same multiple at the same time when zooming in or zooming out, its texture characteristics will change. In order to make the defined texture not affected by the monotonous change of the gray value, a simple method is to binarize the multivalued gray difference value; that is, use the binary expression instead of the gray value difference; then, there is

$$Y \approx y[D(u_0 - u_c), D(u_1 - u_c), \dots, D(u_7 - u_c)]. \quad (10)$$

Among them, where  $D(\cdot)$  is the symbolic function:

$$D(x) = \begin{cases} 1, & x \geq 0, \\ 0, & \text{others.} \end{cases} \quad (11)$$

The sign function describes the two quantization states

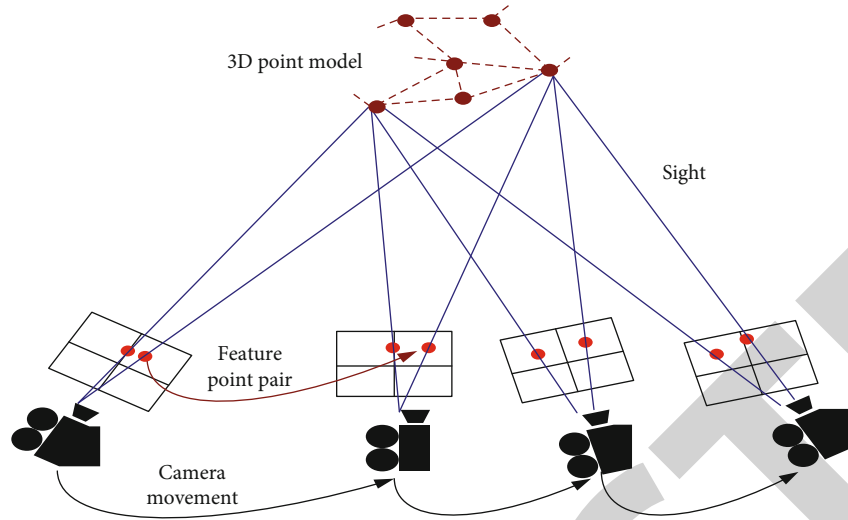


FIGURE 1: Schematic diagram of image sequence and camera movement.

of the gray characteristics of the central pixel and the neighboring pixels. This binary quantization state reduces the range of the joint distribution of the gray difference in the local area, so that the distribution can be used with simple values to represent. Assign a weighting factor of  $2^q$  to each binary expression  $D(u_q - u_c)$ , and sum them, which will obtain the local binary mode texture encoding value representing the central pixel, namely,

$$\text{LBP}_{Q,R} = \sum_{q=0}^{Q-1} D(u_q - u_c) 2^q. \quad (12)$$

In order to enable LBP to obtain texture information of various sizes and structures, adjacent points must be sampled equidistantly within a circle of radius  $R$ . The number of sampling points  $Q$  can be determined according to needs.

### 3. Design Experiment of 3D Modeling Facial Image Library in Multimedia Film and Television

The main content of this section is to design a 3D modeling facial image library experiment in multimedia film and television. The experiment is mainly to select a small number of experimental subjects; first take two-dimensional photos of them, select multiview images, and then compare the experimental subjects. Perform binary feature processing on the facial images, and finally perform 3D modeling of personal faces to establish a database of facial images of people.

**3.1. Camera Parameter Settings.** Thirty subjects were randomly selected in this section, including 14 males and 16 females, aged  $18.59 \pm 3.24$  years old. This section takes a pinhole camera as an example to study the key influencing parameters in the shooting process, which can effectively reduce the “ambiguity” in the results of the 3D modeling image library.

Although the pinhole camera model can represent the basic principle of camera imaging, cameras in the real world are much more complicated than pinhole imaging. The actual camera needs to rely on the lens to converge the light, which will cause a slight deviation of the pixel position between the obtained image and the theoretical result. This kind of phenomenon is called lens distortion. The distortion process arises from the projection process from the camera coordinates to the image plane coordinates, which is the physical limitation of the optical lens itself. The type of distortion can be divided into radial distortion and tangential distortion. The effect of radial distortion is that the part far away from the image center will be visually more distorted than the part close to the image center, while the tangential distortion is mainly caused by the nonparallel placement of the CCD or CMOS photoreceptor and the lens. Generally, the camera model mainly considers the influence of radial distortion, so it is necessary to correct the radial distortion in scenes that require accuracy.

The key research point of generating personalized face model is how to deform the general face model in three-dimensional space or develop a face model that is easy to control the deformation. The pinhole camera simulates the projection imaging process by two parts of the camera's internal parameters and external parameters. The internal parameters represent the parameters of the optical components of the camera, and the external parameters represent the orientation and position parameters of the camera itself. Camera calibration generally needs to use a calibration object of known size as a reference, which has a good corresponding relationship in 3D, and a good internal parameter estimate has a direct effect on the accuracy of reconstruction and the visual effect.

**3.2. Binarized Image Feature Extraction.** The point matching information between different images is the main source of information for multiview 3D reconstruction. Before restoring the three-dimensional structure of the scene, it is necessary to determine which position of the pixels in different

TABLE 1: Comparison of 3D production software in the market.

Name of software	Maya	3D max	Soft image	CAD	Poser	ZBrush	Renderman	Motion builder
Model making	3D	3D	3D	3D, 2D	3D	3D	3D	3D
Flexibility	Good	Good	Good	Good	Good	Good	Good	Good
Texturer	Good	Good	Good	Poor	Good	Good	Poor	Good
Main market positioning	Multimedia design	Industrial, space design	Multimedia design	Industrial and architectural design	Fashion design	Art, multimedia design	Multimedia design	Multimedia
Masterpiece	Star Team, Lord of the Rings	Acquired	Jurassic Park	Nothing	Nothing	Nothing	Toy Story	Nothing

images represents the same point in the three-dimensional scene. Since ordinary images do not have pixel depth information, it is necessary to triangulate geometric element pairs (point pairs, line pairs, etc.) to estimate the three-dimensional coordinates. In order to associate the images of the same scene taken at different angles of view or at different times and to find the “coincidence” between the images, it is necessary to find the image points with the same name in these images. But not all pixels have corresponding points in another image, and not all pixels have enough discrimination to match. Before reconstructing the target structure and camera motion information, it is necessary to select representative, distinguishable, repeatable detection points or local areas in the image as input information for point cloud structure estimation. These special points or small areas are called local features.

Image local features play a very important role in motion detection, image registration, video tracking, image stitching, and 3D reconstruction. They are the first steps in many computer vision algorithms. After the feature points are extracted, the images are matched, and the obtained feature point trajectories can be used as the basis for further camera pose estimation and final 3D modeling.

We selected facial expressions with more relaxed requirements for efficiency and selected binary feature extraction in applications with limited computing resources and high real-time requirements. The generated description vector is used to match the image points of the same name between the images to establish the trajectory of the same point in the image sequence which is regarded as the “coincidence” part between different images.

A well-designed rendering platform can not only completely reflect the images processed by animation technology but also present real-time and smooth dynamic effects to viewers. In this way, the efficiency of rendering platform is directly related to the availability and completeness of the system. In addition to the matching of local features, there are also epipolar geometric constraints that represent the position of the camera between images from different perspectives. Use the point-to-vector matching information of the two images to estimate the basic matrix and the essential matrix, and decompose the relative rotation and translation of the camera for verification, thereby estimating the pose corresponding to the camera in the three-dimensional space.

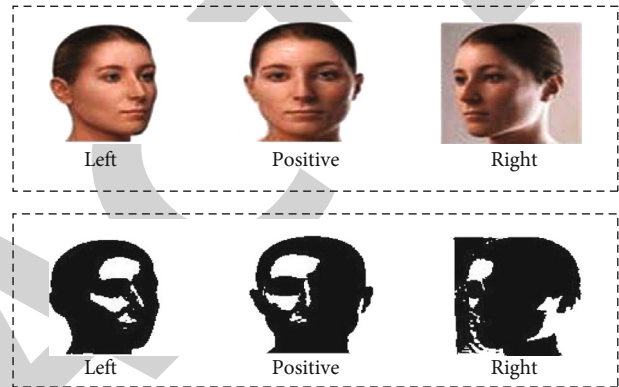


FIGURE 2: Triangular view of the subject (the picture is from Baidu Gallery).

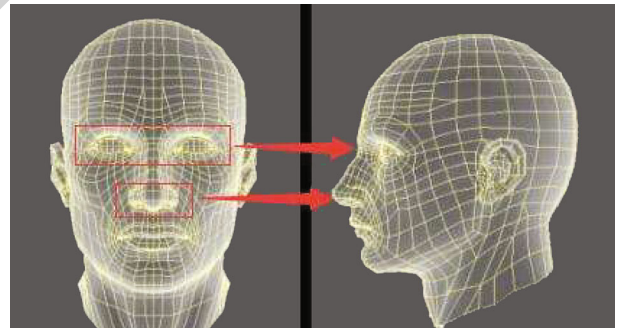


FIGURE 3: 3D facial image construction (the picture is from Baidu Gallery).

**3.3. Three-Dimensional Image Construction.** Combining the previous camera imaging, local feature detection technology, and geometric principles, this article detailed research and design of 3D facial modeling solutions. At present, three-dimensional image construction algorithms using image sequences are mainly divided into two categories: one is to construct the structure of part of the face and the relative expression of the camera as initialization and then gradually add the points observed in other images to the already constructed image, until the entire sequence is constructed. The other is to establish a measurement matrix that reflects the pose relationship between all

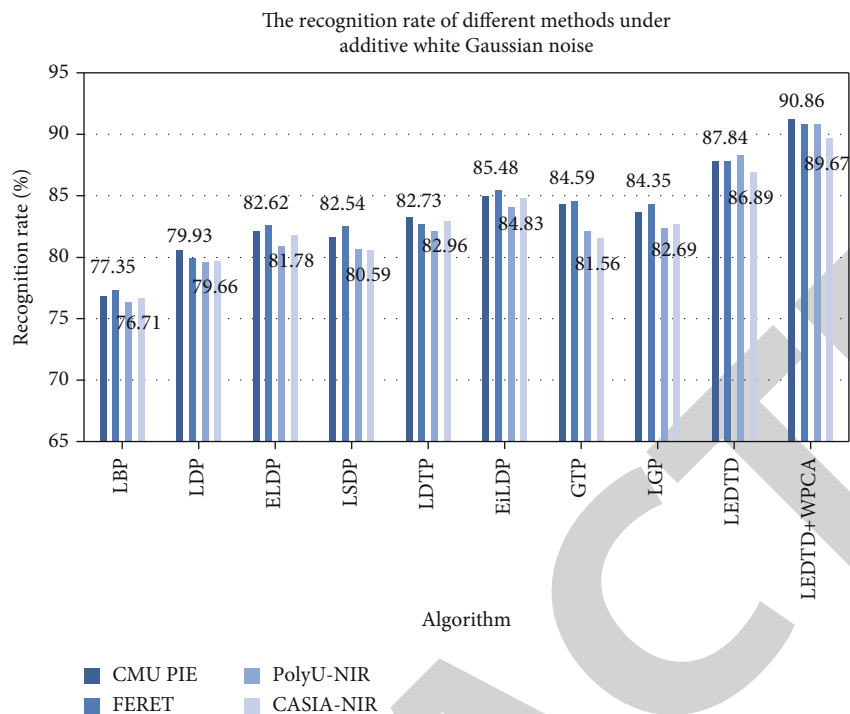


FIGURE 4: The recognition rate of different methods under additive white Gaussian noise ( $\alpha = 0.002$ ).

images in the sequence; firstly solve the camera's orientation position from the global relationship, and then construct a three-dimensional image.

For the face, the movement of some facial organs is a rigid body movement, such as the rotation of the eyeball and the opening and closing of the jaw, without skin deformation. This kind of rigid body motion has some degree and regional independence. The main steps of multiview 3D image construction can be summarized as follows: (1) perform feature detection on the image; (2) match the feature vector between the images, and detect the image points with the same name; (3) use image point matching information and geometric conditions to estimate the orientation and position of the camera; and (4) construct the camera matrix according to the estimated camera motion. The three-dimensional point coordinates of the image point with the same name are calculated by triangulation, and the facial structure is obtained. Figure 1 shows the relationship between camera movement, target point cloud, and image matching points in the image acquisition process, which is an abstract process of multiview 3D image construction.

In this section, we will construct a 3D image library from ordinary 2D images. Ordinary images are composed of pixels and do not contain spatial depth information. We have constructed a multiview 3D image to give a better and more comprehensive appreciation experience visually. And after a lot of experiments on this 3D modeling method, the facial image library of people in multimedia film and television will have a wider range of applications.

## 4. Design Analysis of 3D Modeling Facial Image Library in Multimedia Film and Television

**4.1. 3D Production Platform Comparison.** The 3D production of facial expressions of characters has developed rapidly with the support of computer hardware and software. The following is a simple comparison of 3D production software jobs on the market from the field of computer application (software application), as shown in Table 1. Through the comparison in Table 1, we can see that there is a lot of software for multimedia production, and the production of animation and games is mainly concentrated on the two platforms of Maya and 3D max. Maya focuses more on the production of character animation. Therefore, this article will first use the Maya platform to build the system. At the same time, through the data interpretation design, it can meet the efficiency of graphics rendering and the editability of animation.

**4.2. Face 3D Modeling Image.** Figure 2 shows the real pictures we took of the experimental subjects, and we did partial binarization processing to extract the features of the facial images of people. The image includes not only changes in posture, expression, and lighting but also changes in shooting time and shooting angle. This is a sample image after cropping and resizing a selected personal picture in the image database.

We construct the three-dimensional image of the above picture as shown in Figure 3. We showed the three-dimensional face structure map of the experimental subject from two angles and obtained the coordinates of 45,179



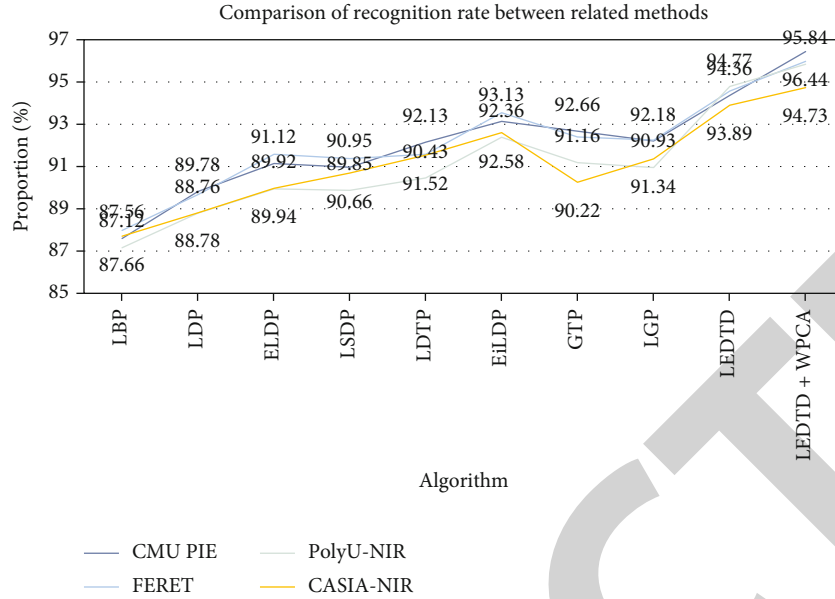


FIGURE 5: Comparison of recognition rates between related methods.

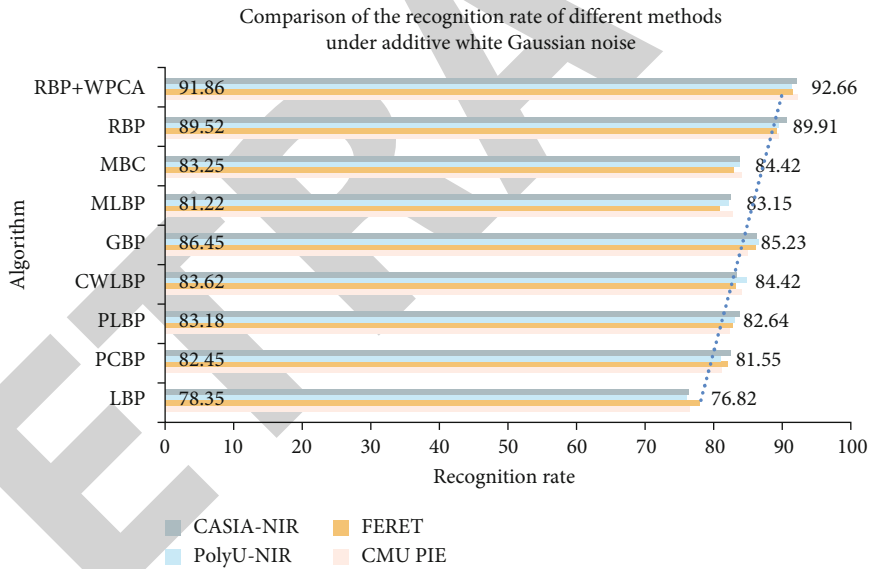


FIGURE 6: Comparison of the recognition rate of different methods under additive white Gaussian noise ( $\alpha = 0.002$ ).

three-dimensional points. The visual effect is good, and the facial shape and some details can be clearly distinguished. The maximum reprojection residual of each image is about 4, and the average residual has reached the subpixel level. The 3D face structure is constructed under the condition of certain changes in illumination, and the experimental results also reflect the stability and robustness of the scheme in this article.

4.3. Noise Robustness Verification of the Proposed Scheme. In the 3D modeling process, noise is inevitable in an uncontrollable environment. Therefore, the noise robustness of image descriptors is very important for the establishment of a facial

image database. This experiment will verify the robustness of the grayscale image descriptor LEDTD to additive white Gaussian noise. In the experiment, the training sample is the face image without noise, and the test sample is the face image with noise. The experimental results are shown in Figure 4.

According to the experimental results, the recognition rate of all methods has decreased in the case of Gaussian white noise and speckle noise. However, among all the compared feature extraction methods, the recognition rate of the grayscale image descriptor LEDTD proposed in this paper is least affected by noise, thus verifying that LEDTD has strong robustness to these two types of noise.

TABLE 2: Characters face construction residual statistics.

Image number/statistical item	1	2	3	4	5	6	7	8	9
Number of observation points	3026	3967	3016	2219	2435	2856	3038	2484	2428
Average residual	0.276	0.293	0.345	0.388	0.341	0.337	0.332	0.388	0.396
Maximum residual	3.365	3.934	3.716	3.235	3.189	3.688	3.465	3.966	3.688
Image number/statistical item	10	11	12	13	14	15	16	17	18
Number of observation points	2428	2561	2081	3634	5256	5571	3380	3029	3176
Average residual	0.396	0.451	0.368	0.382	0.371	0.331	0.396	0.375	0.365
Maximum residual	3.688	3.645	3.225	3.583	3.686	3.218	2.984	3.345	3.341

TABLE 3: Analysis of test cases and expected results.

Serial number	1	2	3	4
Test case	Bone creation test	Cluster selection and soft selection function use	Select to display the original naming and control commands	Bone creation test
Test content	Create arbitrary bones	Set up two selection functions separately to create a controller and switch between functions to create	Arbitrarily select model control points and create controllers for the selected points and name them	Create arbitrary bones
Desired result	Creation success rate is 100%, and response time is less than 10 ms	The creation success rate is greater than 98%, and the response time is less than 15 ms	Naming and display naming success rate is greater than 98%, and response time is less than 8 ms	Creation success rate is 100%, and response time is less than 10 ms

**4.4. Comparison of Recognition Efficiency of Human Facial Image Database.** In order to verify the effectiveness of the grayscale image descriptor LEDTD, this experiment will compare the LEDTD descriptor representation methods in the face image library, using the WPCA dimensionality reduction LEDTD representation method and the face recognition performance of related representative methods. The comparison methods here are LBP, LDP, ELDP, LSDP, LDTP, EiLBP, Gradient Directional Pattern (GDP), and Local Gradient Pattern (LGP). The dimension of the eigenvector of the LEDTD after the WPCA dimensionality reduction is the same as the dimension of the ELDP eigenvector with the least space complexity among similar methods. Figure 5 shows the experimental results of these different image representation methods.

According to the experimental results, the recognition rate of the LEDTD descriptor representation method is higher than other methods. Using the WPCA dimensionality reduction method can further improve the recognition rate of the LEDTD representation method. On the CMUPIE database, the recognition rate of the LEDTD+WPCA method is 96.44%, which is higher than that of the LEDTD method by 2.06%. Similar, on the FERET database, the recognition rate of the LEDTD+WPCA method is 95.96%, while that of the LEDTD method is 94.73%. On the PolyU-NIR and CASIA-NIR databases, using the WPCA method can increase the recognition rate by 1.07% and 0.84%, respectively. All of these indicate that LEDTD's grayscale image feature extraction ability is better than other methods.

Figure 6 verifies the robustness of the RBP descriptor to additive white Gaussian noise. In the experiment, the training sample is the face image without noise, and the test

sample is the face image with noise. The experimental results are shown in Figure 6. According to the experimental results, the recognition rate of all methods decreases in the case of Gaussian white noise. On the other hand, by programming the GPU shader, the vertices are interpolated according to the time series during the animation process. When the animation module can not effectively provide animation frames, the rendering effect is smooth, which reduces the time sensitivity of animation playback to a certain extent. However, among all the methods, the reduction of the recognition rate of the RBP descriptor is the least affected by noise, which proves that the RBP descriptor is the least sensitive to this noise in the compared methods.

**4.5. Test Results of the 3D Modeling Scheme in This Article.** Table 2 shows the residual statistics in the 3D construction of human images. We selected 18 image observation points and compared them from the average residual and the maximum residual. It can be seen from the table that the average residual error range is 0.25-0.45, and the maximum residual error does not exceed 4.0, indicating that the face image constructed in this article has a good correspondence relationship and the residual value is low.

We use the face 3D model created in this paper to simulate and create specific examples, and the results are shown in Table 3. The test content is mainly based on the specific creation and conversion of test cases and comparison with the expected results. The success rate of all our creations is above 95%, and the response time is within 15 ms, which meets the very good 3D modeling requirements.

Based on the analysis of the above test plan, the following test cases are designed: according to the test content designed by the test plan, the plug-in is tested. Make a

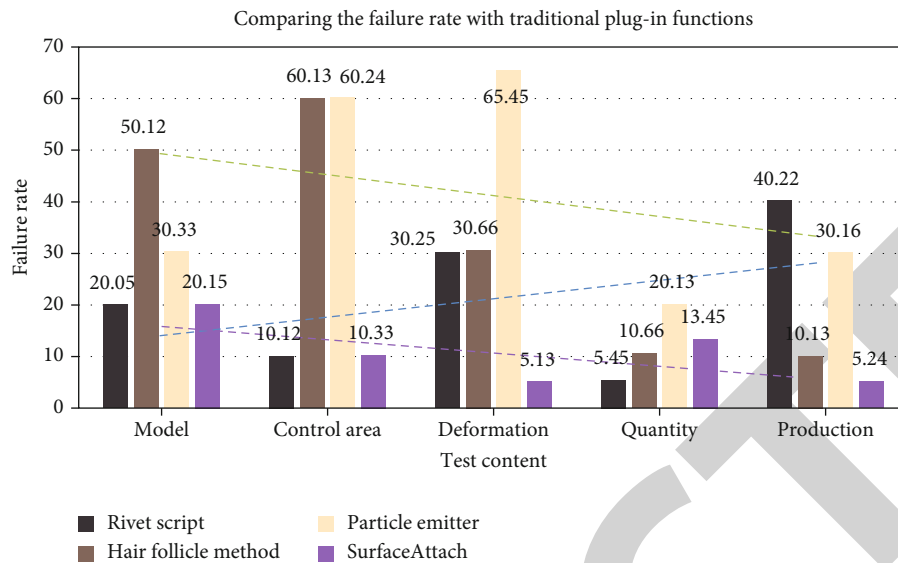


FIGURE 7: Comparing the failure rate with traditional plug-in functions.

comparative analysis of the expected experimental results of all test cases.

Compared with the use of traditional plug-ins, users' feedback that this plug-in solves problems such as being only applicable to NURBS objects; not easy to locate selected areas and points; improper display and inability to adjust actions; and facial deformation which causes displacement of marker points. Figure 7 compares the failure rate with the traditional plug-in. The number of tests is 1000 times. Compared with the traditional face gallery design, this plug-in reduces the plug-in's restrictions on the model creation method. In the operation steps, a large number of repeated operation steps are reduced, and 6 facial expressions of any three-dimensional character are produced. The use of this plug-in improves the efficiency by two to three times compared with the traditional method, and the operation is simple and error-free.

## 5. Conclusion

This article is mainly about the design and research of the facial image library of 3D modeling in multimedia film and television. The 3D modeling character facial image design in the article has good stability and robustness and has good image search performance in the image library. The 3D image quality in the image library is high, which can meet the application of the multimedia film and television industry and promote multimedia the development of the film and television industry. The innovation of this article lies in the application of 3D modeling technology to the establishment of the character image library, which has a better three-dimensional effect and practical effect on the image and highlights the influence of digital technology on the development of the multimedia film and television industry. The shortcomings of this article are that the 3D modeling technology is not perfect enough, and the amount of application data in the facial image database is too large, and

the cost is high. The data sources of both animation deformation and visual presentation are allocated to the scene module, which has a hierarchical relationship. Because the scene module is at the innermost layer of the program architecture, it is particularly important whether the organization of the scene data structure is efficient and applicable. In the future, the application of 3D technology in the multimedia film and television industry can provide ideas for the innovative development of multimedia film and television. Although there is still a big gap, this is still the key content and direction of follow-up research.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] T. Ni, Z. Yang, H. Chang et al., "A novel TDMA-based fault tolerance technique for the TSVs in 3D-ICs using honeycomb topology," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 724–734, 2021.
- [2] E. Yaremenko, "Multimedia learning model: adapting an image in a screen work," *Journal of Film Arts and Film Studies*, vol. 11, no. 4, pp. 90–96, 2019.
- [3] D. Yang, D. Qiao, and C. Xia, "Curved light surface model for calibration of a structured light 3D modeling system based on striped patterns," *Optics Express*, vol. 28, no. 22, pp. 33240–33253, 2020.
- [4] A. Annita, "Teaching film in new normal era at Film Department, Universitas Multimedia Nusantara," *International Journal of Creative and Arts Studies*, vol. 7, no. 2, pp. 89–100, 2020.
- [5] E. Lindasari, K. Ansari, and M. Marice, "Interactive multimedia development in learning of film review text for 8th grade

- students in Senior High School (SMP) 1 Tanjungmorawa,” *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, vol. 2, no. 4, pp. 355–362, 2019.
- [6] Y. Huang, X. Sun, M. Liu et al., “A multimedia fugacity model to estimate the fate and transport of polycyclic aromatic hydrocarbons (PAHs) in a largely urbanized area, Shanghai, China,” *Chemosphere*, vol. 217, pp. 298–307, 2019.
- [7] H. Zhang, “Voice keyword retrieval method using attention mechanism and multimodal information fusion,” *Scientific Programming*, vol. 2021, no. 8, Article ID 6662841, 2021.
- [8] Z. Yan and Z. Lv, “The influence of immersive virtual reality systems on online social application,” *Applied Sciences*, vol. 10, no. 15, p. 5058, 2020.
- [9] M. Jahanbakht, W. Xiang, L. Hanzo, and M. R. Azghadi, “Internet of underwater things and big marine data analytics – a comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 904–956, 2021.
- [10] S. Ventroni and R. Straer, “Filmizenzen im zeitalter von Netflix & Co,” *Multimedia und Recht*, vol. 22, no. 6, pp. 359–364, 2019.
- [11] H. Setiawan, A. Aziz, and D. Kurniadi, “Ideologi patriarki dalam film (semiotika John Fiske pada interaksi ayah dan anak dalam film chef),” *ANDHARUPA Jurnal Desain Komunikasi Visual & Multimedia*, vol. 6, no. 2, pp. 251–262, 2020.
- [12] A. Kharismasari, I. N. Ridho, N. Y. Kinanti, and A. Chusyairi, “Pembuatan film pendek allure of culture dengan teknik edit continuity cutting,” *JTIM Jurnal Teknologi Informasi dan Multimedia*, vol. 1, no. 4, pp. 318–326, 2020.
- [13] S. M. Sutanto, “Dekonstruksi representasi perempuan pada poster film Pahlawan Super Produksi Hollywood,” *ANDHARUPA Jurnal Desain Komunikasi Visual & Multimedia*, vol. 6, no. 1, pp. 1–17, 2020.
- [14] W. P. Kurniawan, “Pembuatan film pendek bertemakan persahabatan “How Bromance Us,”” *JTIM Jurnal Teknologi Informasi dan Multimedia*, vol. 2, no. 2, pp. 109–115, 2020.
- [15] D. Sari, “Pembuatan film pendek IT in millennial daily life dengan teknik handheld dan continuity cutting,” *JTIM Jurnal Teknologi Informasi dan Multimedia*, vol. 1, no. 4, pp. 302–310, 2020.
- [16] Y. B. Lin, M. T. Yang, and Y. W. Lin, “Low-cost four-dimensional experience theater using home appliances,” *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1161–1168, 2019.
- [17] S. Anggelina and A. Trisnadoli, “Analisis efektivitas pesan film animasi 3D bahaya rokok terhadap perokok aktif remaja dan dewasa,” *ANDHARUPA Jurnal Desain Komunikasi Visual & Multimedia*, vol. 6, no. 1, pp. 115–124, 2020.
- [18] A. D. Setiawan and M. Bezaleel, “Perancangan film pendek “Bukan Hak-Ku” untuk menanamkan nilai kejujuran pada anak usia 8-12 tahun,” *ANDHARUPA Jurnal Desain Komunikasi Visual & Multimedia*, vol. 5, no. 1, pp. 127–140, 2019.
- [19] D. Meliala and M. Bezaleel, “Analisis film foror Indonesia Produksi Tahun 2014 (studi kasus: Mall Klender dan Kamar 207),” *ANDHARUPA Jurnal Desain Komunikasi Visual & Multimedia*, vol. 2, no. 1, pp. 1–14, 2018.
- [20] I. Setiawan, I. Adnyana, K. R. Winatha, and I. N. Y. Trisemarawima, “Film dokumenter eksistensi tarian dan kesenian sakral wayang wong desa adat Sidan,” *ANDHARUPA Jurnal Desain Komunikasi Visual & Multimedia*, vol. 5, no. 2, pp. 251–264, 2019.
- [21] S. H. Nugraini, “Dampak model pembelajaran produksi film pendek tentang energi terbarukan pada mata kuliah audio visual terhadap mahasiswa,” *ANDHARUPA Jurnal Desain Komunikasi Visual & Multimedia*, vol. 4, no. 2, pp. 196–206, 2018.
- [22] L. F. Amorim, J. Martins, F. F. Nogueira et al., “Hydrodynamic and ecological 3D modeling in tropical lakes,” *SN Applied Sciences*, vol. 3, no. 4, pp. 1–14, 2021.
- [23] X. Wang, “Application of network protocol improvement and image content search in mathematical calculus 3D modeling video analysis,” *AEJ-Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4473–4482, 2021.
- [24] F. Valivand and H. Katibeh, “Prediction of nitrate distribution process in the groundwater via 3D modeling,” *Environmental Modeling & Assessment*, vol. 25, no. 2, pp. 187–201, 2020.
- [25] V. I. Korenev, “Digital technologies and 3D-modeling of urban planning (The Tomsk Case Study),” *Vestnik Tomskogo gosudarstvennogo arkhitekturno-stroitel'nogo universiteta. JOURNAL of Construction and Architecture*, vol. 22, no. 6, pp. 70–82, 2020.
- [26] A. al-Mnayyis, S. A. Alasal, M. Alsmirat, Q. B. Baker, and S. AlZu'bi, “Lumbar disk 3D modeling from limited number of MRI axial slices,” *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 4101–4108, 2020.
- [27] V. V. Tatarinov and A. S. Kuzakov, “Evaluation of the characteristics of X-ray excitation under the electron-probe effect using 2D and 3D modeling by the Monte Carlo method,” *Journal of Surface Investigation: X-Ray, Synchrotron and Neutron Techniques*, vol. 14, no. 2, pp. 245–252, 2020.
- [28] N. V. Lang and M. M. Shlyakhova, “New directions of using 3D-modeling,” *Interexpo GEO-Siberia*, vol. 6, no. 1, pp. 205–209, 2020.
- [29] H. Xu, J. Xu, and W. Xu, “Survey of 3D modeling using depth cameras,” *Virtual Reality & Intelligent Hardware*, vol. 1, no. 5, pp. 483–499, 2019.
- [30] H. V. Haiko, O. A. Haluzynskyi, and S. V. Burburska, “Using 3D modeling with the manufacture of plastic prototypes in traumatology and orthopedics (clinical examples),” *Trauma*, vol. 20, no. 1, pp. 89–97, 2019.