

Retraction

Retracted: 3D Film Animation Image Acquisition and Feature Processing Based on the Latest Virtual Reconstruction Technology

Complexity

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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) Manipulated or compromised peer review

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.

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] S. Wu, S. Xiao, Y. Di, and C. Di, "3D Film Animation Image Acquisition and Feature Processing Based on the Latest Virtual Reconstruction Technology," *Complexity*, vol. 2021, Article ID 2331306, 11 pages, 2021.

Research Article

3D Film Animation Image Acquisition and Feature Processing Based on the Latest Virtual Reconstruction Technology

Siwei Wu ^{1,2}, Shan Xiao,³ Yihua Di,⁴ and Cheng Di ^{1,2}

¹School of Arts and Communication, China University of Geosciences, Wuhan 430074, China

²Faculty of Fine and Applied Arts, Burapha University, Bang Saen, Chonburi 20131, Thailand

³International Education College, China University of Geosciences, Wuhan 430074, China

⁴College of Mechanical and Electrical Engineering, Wuhan Textile University, Wuhan 430074, China

Correspondence should be addressed to Cheng Di; dicheng@cug.edu.cn

Received 10 April 2021; Revised 15 May 2021; Accepted 28 May 2021; Published 14 June 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Siwei Wu et al. 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.

In this paper, the latest virtual reconstruction technology is used to conduct in-depth research on 3D movie animation image acquisition and feature processing. This paper firstly proposes a time-division multiplexing method based on subpixel multiplexing technology to improve the resolution of integrated imaging reconstruction images. By studying the degradation effect of the reconstruction process of the 3D integrated imaging system, it is proposed to improve the display resolution by increasing the pixel point information of fixed display array units. According to the subpixel multiplexing, an algorithm to realize the reuse of pixel point information of 3D scene element image gets the element image array with new information; then, through the high frame rate light emitting diode (LED) large screen fast output of the element image array, the human eye temporary retention effect is used, so that this group of element image array information go through a plane display, to increase the limited display array information capacity thus improving the reconstructed image. In this way, the information capacity of the finite display array is increased and the display resolution of the reconstructed image is improved. In this paper, we first use the classification algorithm to determine the gender and expression attributes of the face in the input image and filter the corresponding 3D face data subset in the database according to the gender and expression attributes, then use the sparse representation theory to filter the prototype face like the target face in the data subset, then use the filtered prototype face samples to construct the sparse deformation model, and finally use the target faces. Finally, the target 3D face is reconstructed using the feature points of the target face for model matching. The experimental results show that the algorithm reconstructs faces with high realism and accuracy, and the algorithm can reconstruct expression faces.

1. Introduction

With the continuous progress of image technology, optical device processing process, and display system, 3D stereo display technology has now been realized in the laboratory and developed some stereo imaging systems, and the preliminary 3D display effect has been obtained and applied to some fields [1]. As is known to all, technology of three-dimensional display mainly has a holographic three-dimensional display and nonholographic three-dimensional display [2]. The nonholographic 3D display can be further divided into 3D display mode with an auxiliary device and 3D display mode without an auxiliary device.

Divided by the nature of the reconstructed 3D images, these systems can be divided into three types: binocular, multiview, and spatial image reconstruction 3D display systems [3]. Each of these systems is a complex technology that includes image acquisition optical structures, image information recording, and image information reconstruction. The combination of virtual reality and human motion capture techniques can be used for the simulation of virtual scenes. In crowd evacuation simulation of buildings, based on the simulation of human motion in obstructed and narrow passages, it helps to design safer and reasonable escape routes, thus reducing casualties and losses [4]. Also, virtual simulation technology is important for the study of

ergonomics. Ergonomics simulation is used to examine the coordination between humans, machines, and the environment, to guide and examine the design and modification of machines and equipment and thus to reduce design costs. In simulation scenarios, virtual humans are used to verifying the practicality and accessibility of workstations and tooling equipment operations.

The rise and development of human motion capture technology have also provided a new way of thinking for medical diagnosis and rehabilitation treatment. The first aspect is motion monitoring, where human motion capture systems are used to analyse and estimate the energy expenditure, muscle fatigue, and skeletal and joint damage of athletes during training [5]. And it is more difficult to get started in the process of using and requires experienced professionals for human-computer interactive operation, which limits the population of users, making it difficult to promote and apply these software [6]. Skeletal skinning technology is the main research direction in the field of 3D character animation. It creates animation by adjusting the skeletal motion of the model to drive the skin motion attached to the skeleton. The skeletal skinning technique is applicable to simulate the skeletal and surface shapes of humans and other living creatures. Also, the skeleton skinning technique is highly controllable for characters in animation provides advanced control and allows easy editing of animation data; i.e., the motion pose of the character model can be adjusted by directly adjusting the motion state of the skeleton with real-time interactivity. However, in practice, the effect of skinning is not very perfect.

The main body of research in this paper is material appearance, and the main research is divided into two parts, which are oriented towards the efficient processing of surface materials and the rendering of media-based materials with realism. Among them, the former work tries to solve the compression problem of measured materials and their related applications, and the latter work proposes a method for rendering media with granularity. Measurement data using real surface reflection properties is one of the most effective methods for drawing realistic materials. A measurement material uses measurement data to accurately reproduce the appearance of real-world material. One way to use this measurement data is to fit several parametric BRDFs so that parameters can be set like those of a parametric BRDF model such as Torrance-Sparrow [7]. The results may also be poor if the data errors are too large: especially for highlight terms and surfaces with various anisotropies, the 4-dimensional BRDF must be densely sampled, which leads to a large amount of data to be stored and many measurements to be taken. To alleviate these problems, this paper proposes an efficient method for handling measured materials and a practical framework for material editing.

2. Current Status of Research

Due to the nonrigid nature of the human body, the human action posture, clothing style and colour, background lighting, and occlusion can affect the final detection results.

Especially, the occlusion problem is a major difficulty in human body detection at present [8]. The detection model based on human variable parts can deal with the problem of occlusion and pose change very well, and the deformable parts model proposed by Xia et al. does not require image preprocessing, but by dividing the human model into several pieces and then fusing the results of each part detection to determine the presence of the human body [9]. This has a strong detection performance but requires the construction of many templates to achieve a good matching effect in the case of occlusion and pose change [4]. The most important step of the R-CNN algorithm based on the region suggestion frame proposed by Bladin et al. is to perform a selective search algorithm to get the candidate regions, then normalize the candidate domains by convolutional neural network and perform the complex convolutional computation to extract the convolutional features in the image target, and finally use support vector machine to classify the data according to the features [1]. Finally, the data is divided into maximum intervals according to the features to achieve the expected classification effect [10]. The generative tracking algorithm extracts features and learns the appearance model of the target based on the target image, then extracts samples in the subsequent frames, calculates the similarity error between the samples and the target, and selects the sample with the smallest error as the predicted target for the current image [11]. Generative algorithms can be divided into mean offset-based tracking algorithms, subspace-based tracking algorithms, and sparse coding-based tracking algorithms [12]. Pfarr-Harfst et al. first introduced the idea of sparse coding in target tracking algorithms and proposed an L1 tracker. L1 tracker is to model occlusion and noise by trivial templates and use L1 parametric coefficients to calculate the candidate samples of sparse representation, and then select the one with the least reconstruction error of this sparse representation as to the target particle of the current frame. However, the L1 tracker has high computational complexity and is slow [13].

3D modelling software human modelling is the use of 3D software to build human models, such as the use of Maya and Drax to build human models. Now many industries use 3D modelling software, such as games and 3D animation [14]. The effect of the human body model built by 3D software is very intuitive, the operator can directly on the software interface to fully observe the model they create and can rotate and scale the model to create detailed textures. However, for 3D modelling software professionals, all consume a lot of time. Even if designers use relevant plug-ins to assist modelling or optimize modelling results, it is still difficult to reduce modelling time. 3D scanning modelling is to obtain point cloud data information on the surface of an object by scanning devices [15]. 3D scanning modelling can be divided into two categories according to the principle of device implementation: laser scanners and structured light scanners [16].

Firstly, the development history and basic structure of convolutional neural networks are studied and summarized, and several typical lightweight convolutional neural networks are learned, and the Mobile Net network is selected as

the base network and optimized. Then we introduce the null convolution, which can expand the convolution kernel to get the scale bounded by the null parameter and then fill the unoccupied area in the original convolution kernel with “0”, to expand the perceptual field with the same number of parameters. Therefore, several methods for applying null convolution in different convolutional layers of Mobile Net networks are proposed and analysed in this paper. In the human motion tracking stage, the network structure of Open Pose is optimized after a detailed introduction of tracking-related algorithms and the working principle of Open Pose. Then, the original VGG network is replaced by the optimized mobile network. For most of the computations to obtain the feature maps and the juxtaposition between some of the affinity fields, all the layers in it except for the last two layers of each block are shared. Finally, because the 7×7 convolution kernel is computationally intensive, the complexity is reduced by using convolutional blocks with the same sensory field for a replacement.

3. Analysis of 3D Film Animation Image Acquisition and Feature Processing by the Latest Virtual Reconstruction Technology

3.1. Virtual Reconstruction Technology Image Acquisition and Feature Processing Design. A video-based 3D reconstruction method of the human body adjusts the shape and pose parameters of SMPL based on the 2D key point information extracted from the video frames to reconstruct a 3D human body model like the human body in the video frames [17]. A Skinned Multi-Person Linear Model (SMPL) is a masked vertex-based model. The SMPL is a model that accurately represents the body shape and poses various human bodies by using human body data generated by large 3D scanning equipment. It is a data-driven parametric model with parameters that include a series of parameters that control shape and parameters that control pose, and as soon as these parameters are determined, the SMPL model can generate a specific human model with different poses as shown in Figure 1 for the SMPL model.

The SMPL 3D mesh model constructs a skeleton model with a parent-child relationship. It has 6890 vertices and 23 skeletal key points. The SMPL model also gives the skeleton a tree-like structure, with the pelvis joint (Pelvis) as the root node of the tree, and the pose of the child key points is influenced by the change in pose of the parent key points. The model vertices are divided into 24 blocks and the vertices in each block can be categorized into key points that are like it. In other words, in the SMPL 3D mesh model, the SMPL model performs regression analysis on the vertices in the 3D mesh to obtain the locations of the key points in the model. It is also possible to adjust the vertex position in the model according to the position of the key points.

In the SMPL model, each key point has three degrees of freedom, i.e., each key point can be rotated around three axes of its local coordinates. Also, the key points control the fixed key points of the model offset. The parameter that controls the pose of the mesh model is θ , which is 1×75 vectors. The

SMPL model performs principal component analysis on the training set of the 3D human mesh sequence to obtain the corresponding morphological parameters, each of which corresponds to a coefficient value [18]. The coefficient values can be adjusted if the SMPL model shape is edited. The parameter controlling the shape of the grid model is β . The length of β is not fixed, but it is a 1×10 vector by default. The initial template of SMPL is T . The shape modified SMPL is denoted as T , the number of vertices in the template is denoted as N , and the length of vector β is set as n . The whole model shape modification equation is shown in equation (1). The entire model shape correction equation is given as

$$T = \frac{\bar{T} + \sum_{i=1}^n \beta_i S_i}{\bar{T} - \sum_{i=1}^n \beta_i S_i}, \quad (1)$$

where β_i is the i -th value of vector β , and each value in β corresponds to a morphological parameter; $S_i \in R^{3N}$ is a vector unit of length $3N$, the i -th morphological parameter of the 3D human network module sequence training set after PCA analysis under the SMPL method. In the shape correction, the value in β needs to be modified to adjust the weight of its shape basis, and the new model shape obtained after the correction is then linearly summed with the position values of each vertex of the model template to obtain the new model. In essence, the pose correction process of SMPL is the process of the rigid motion of each module in the model, and the rigid motion of each module is based on the state form of its parent key point. After setting any vertex in the SMPL model template as v_i and the attitude parameter as θ , the vertex obtained after the attitude deformation is v'_i , and its attitude deformation process satisfies

$$v'_i = \sum_{k=1}^K \left(w_{k,i} \otimes \left(\prod_{j \in A(k)} \begin{bmatrix} 1 & \nabla \bar{j}_j \\ 0 & e^{-\bar{w}_j} \end{bmatrix} \right) \right), \quad (2)$$

where $w_{k,i}$ denotes the weight of the influence of the k th key point on the i -th vertex; $A(k)$ denotes the parent key point of the k th key point; a vector of length 3 \bar{w}_j in which the values denote the rotation angles of the key points is around the X , Y , and Z axes, respectively; a 3×3 rotation matrix corresponding to the vector \bar{w}_j , where \bar{j}_j is a vector of length 3 , denotes the position of the j -th key point.

Equation (2) points out that the position of any vertex in the model is influenced not only by the key point corresponding to that vertex but also by the parent node of the key point corresponding to it; i.e., in the process of model correction for attitude deformation, the change of attitude starts from the root node of the skeleton and affects the position of each key point in the next sequel, while each joint also determines the position of the vertex inside the module in which it is located. In SMPL, the positions of the parameters related to the rotation of each key point are fixed in θ . For example, the angle of rotation of key point 0 around the X , Y , and Z axes are represented by the first three values of θ . Therefore, only the corresponding parameters need to be changed during the posture correction of the human model in SMPL, and there is no need to edit the skeleton manually, which greatly optimizes the convenience of the

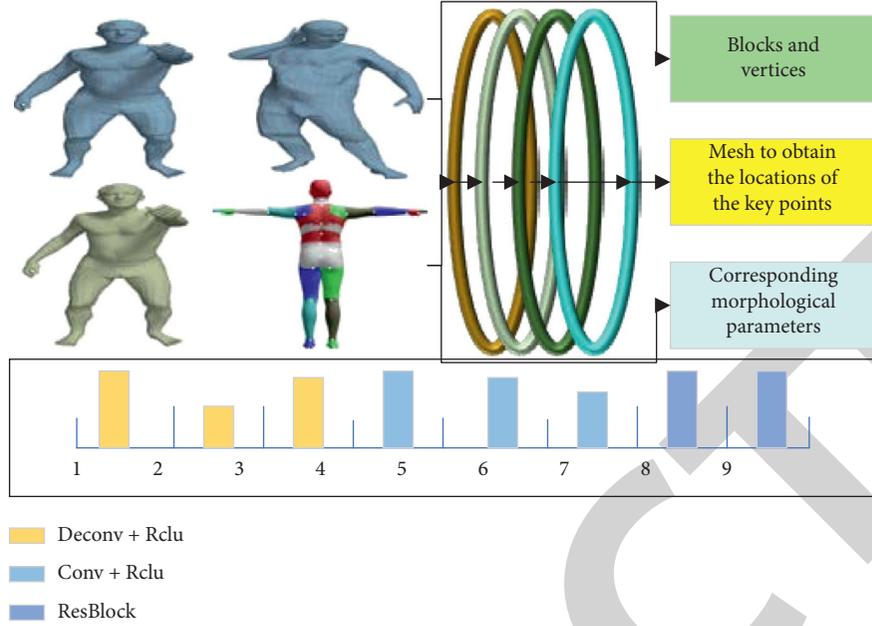


FIGURE 1: SMPL model.

posture correction of the model. For example, the shape of the model can be edited by adjusting only the value of β . SMPL separates the shape of the model from its pose, so the SMPL model can exhibit a greater variety of shapes and

poses. In this study, the SMPL model is used to reconstruct the human 3D model, and the Simplify algorithm establishes the energy function in reconstructing the 3D human pose and shape:

$$E(\beta, \theta) = \frac{E_f(\beta, \theta; K, J_{est}) + \lambda_\theta E_\theta(\theta) - \lambda_a E_a(\theta) + \lambda_{sp} E_{sp}(\theta; \beta) + \lambda_\beta E_\beta(\beta)}{2} \quad (3)$$

The function contains five blocks: key point-based data, preferred pose energy, limb pose energy, collision energy, and preferred shape energy. Where J_{est} is the 2D key point data in the image obtained by Deep Cut, θ is the parameter that controls the pose of the mesh model, and β is the parameter that controls the shape of the mesh model. K is the parameter of the camera that is used to calculate the projection matrix of the 3D model projected onto the image. λ_θ , λ_a , λ_{sp} , and λ_β are the weight parameters corresponding to the latter four blocks, respectively. From (3), the algorithm reconstructs the shape and pose of the 3D human model mainly by adjusting the shape parameter β and the pose parameter θ of the SMPL model.

When the camera performs 3D object photography, the micro lens array-based integrated imaging moves perpendicular to the plane of the camera's optical axis and has a different axial distribution perception than the micro lens array integrated imaging. It collects 3D information along the optical axis of the camera, which has the advantage of adding parallax information to the scene information along the optical axis of the camera. The axial distribution perception structure acquisition process is shown in Figure 2; after determining the object distance in the camera, a camera is placed at the starting point, and then the camera is moved in the opposite direction along the optical axis. One image is

taken for each determined distance moved, and the element images are obtained when the shooting is completed. Because each element image is taken by the camera at different distances, the element images have different magnifications, as can be seen in the element images from the images. In the centre of micro lens array or synthetic aperture array technology, the ability to record the information of each micro lens or each camera is the same, but in the axial distribution because the camera shooting is carried out along the optical axis, so the field of view angle increases continuously in a radial shape, and the ability to record the optical axis position is zero, which can be derived from Figure 2, assuming that the vertical distance from the camera in space is l . The whole system can be acquired in the image. The achieved angular range ϕ is

$$\phi = \sec^{-2} \left(\frac{lR}{l^2 + s_0^2 + s_0 R} \right), \quad (4)$$

where R is the distance between the closest distance of the camera to the object and the farthest distance of the camera from the object, and $0s$ is the distance between the initial position of the camera and the object. From this equation, the information perception angle of the axially distributed sensing structure is related to the vertical distance between

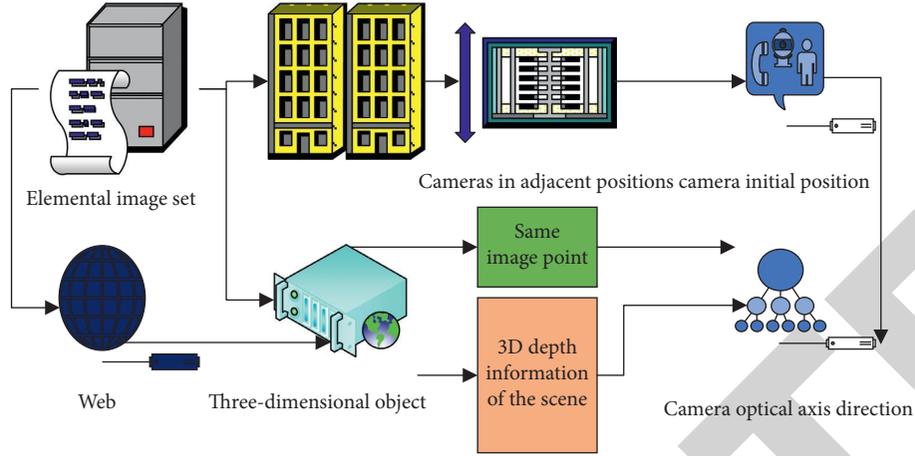


FIGURE 2: Axial distribution perception structure.

the 3D object and the camera optical axis. Therefore, if the object is within the camera angle, changing the vertical distance between the object point and the optical axis of the camera can obtain an image of each point containing different 3D information about the element. And the farther the vertical distance is, the larger the ϕ value is, the more 3D information can be obtained; on the contrary, the smaller the vertical distance is, the less 3D information can be obtained when the vertical distance is zero; that is, when the 3D object is on the camera optical axis, the perception system cannot capture any 3D object information, and the obtained 3D information is zero.

The camera in the axial distribution perception technique has a problem in performing the movement process; that is, there is a minimum step in the movement process, because the camera movement step is of pixel level, then the distance of the cameras in adjacent positions may be very close, which may lead to the same image point of an object point in space between two adjacent element images so that there is no parallax information. There is no way to get the 3D depth information of the scene, and the formula for the minimum camera movement step is given:

$$\max \Delta z_{i,j} = \frac{\mu z_i^2 + r g}{u z_i - r g}, \quad (5)$$

where μ represents the pixel size, g is the distance between the lens of the camera and the CCD plane, it denotes the distance between the i -th camera and the 3D object, and $\Delta z_{i,j}$ denotes the minimum distance between the i -th camera position and the j -th camera position. Different from synthetic aperture integrated imaging which can only obtain parallax information in the vertical direction of the camera optical axis and different from axial distribution perception and structure which can only obtain parallax information in the parallel direction of the camera optical axis, a method is proposed which can make the element image obtain both horizontal view information and vertical view information called the integrated imaging system with off-axis distribution structure.

To address the above problems, this section defines the topological loss term and adds it to the objective function of NICP, so that the algorithm does not need to maintain the topology by drastically reducing the distance loss weights, thus improving the dense correspondence accuracy. The 3D face is connected by the triangular meshing of the scattered points to form a mesh model of the 3D face [19]. The vertex index of the triangles, the shape of the triangles, and the number of triangles in the mesh model determine the topology of the 3D face. To quantitatively measure the change of the topology of the 3D face, this section quantifies the change of the topology by using the sum of the differences of the corresponding triangular face slices between the 3D faces. The difference between the triangular faces is measured by the sum of the differences of the vectors between the centres of gravity of the two triangles to their respective three vertices. Then, according to the above definition, the difference between the two triangles can be expressed as

$$H_i^2 = \frac{\|\vec{DA} + \vec{HE}\|^2 + \|\vec{DB} + \vec{HF}\|^2 + \|\vec{DC} - \vec{HG}\|^2}{3}. \quad (6)$$

As mentioned above, the topological loss term is defined as the sum of the differences of the triangular face slices corresponding to the 3D face templates after deformation during two adjacent iterations, and the differences of the corresponding triangular face slices are measured by the differences of the vectors between the two triangles to their respective three vertices; then, the topological loss term $E_T(F)$ is expressed in the following form:

$$E_T(F) = \sum_{i=1}^c H_i^2, \quad (7)$$

where c denotes the number of triangular faces on the 3D face template, and using H_i^2 and g to denote the three vertices and the centre of gravity of the i -th triangle during the k th iteration, respectively, we have

$$E_T(F) = \sum_{i=1}^c \frac{\| (r_i^k + r_{\text{grav}}^k) + (r_i^{k-1} + r_{\text{grav}}^{k-1}) \|_F^2}{\| (r_i^k + r_{\text{grav}}^k) - (r_i^{k-1} + r_{\text{grav}}^{k-1}) \|_F^2} \quad (8)$$

The part in parentheses represents the vector from the centre of gravity to the three vertices of the i -th triangle on the face template after the k th deformation. The traditional NICP objective function contains a distance loss term, a smoothing loss term, and a feature point loss term. After adding the topological loss term, the improved NICP objective function is shown as follows:

$$E(F) = \frac{E_d(F) - \alpha E_s(F) - \beta E_m(F) + \lambda E_T(F)}{2} \quad (9)$$

The adaptive weighting strategy is a strategy that automatically adjusts the distance loss weights of some points during the solution process of NICP according to the size of the distance between the last deformation result and the target face [20–22]. The 3D face aligned using NICP algorithm has a very high accuracy of dense correspondence in the central region of the face, but it is easy to produce a small number of points with large errors around the corners of the eyes, mouth, and nose, and the presence of these points greatly reduces the detail quality of the 3D face after dense correspondence. After adding the topological loss term to the objective function and increasing the distance loss weight of the edge part of the face, the accuracy of the dense correspondence of the edge part of the face is improved, but this strategy cannot reduce the error of the points in the centre region of the face. The adaptive weighting strategy controls the distance loss weights of the points to control their dense correspondence errors, thus improving the dense correspondence accuracy of the point clouds in specific parts of the face (around the corners of the mouth, eyes, and nose).

3.2. Experimental Design of 3D Film Animation. Today's Internet is rich in images containing a variety of natural and artificial objects, and it is easy for users to photograph real-world objects [23–25]. Two-dimensional images provide a rich resource for the research in this paper. Computer image processing technology has also been a research hotspot in the field of computer vision in the past decades, where researchers have achieved fruitful research results in filtering, sharpening, target recognition, segmentation, etc. of digital images [26, 27]. The improvement of image processing technology has also provided a solid theoretical basis for the research on obtaining image shape features in this paper. The method is implemented by having the micro lens array vibrate rapidly and synchronously in the horizontal direction during the process of acquiring or reconstructing the 3D scene, and controlling the vibration frequency during the image dwell time, which needs to be kept within the human eye dwell time. The amplitude of the micro lens array is to be fixed at the size of a single micro lens, while the rest of the hardware is fixed throughout the system. While the micro lens array used in acquisition and reconstruction vibrates, the observer, equivalent to a sensor, receives a stationary, pulsation-free image, and the resolution of this

reconstructed image is improved relative to the reconstructed image resolution of conventional integrated imaging systems.

The display resolution of the display device is an important indicator of the display system, the size and resolution of the display device affects the number of pixels displayed, and the existing devices cannot achieve many pixels displayed, so how to use the existing display devices to achieve high-resolution display is the focus of research. As shown in Figure 3, a high-resolution stereo display projection method based on space-division multiplexing is proposed, the main way is to realize the whole element collection through many display panels, the whole element image is divided into several subimage sets in space, and the 2D projection equipment is used to project each subimage set onto the recording board, which plays the role of the display panel after the lens array in the traditional stereo imaging system. This plate acts as a display panel behind a lens array in a conventional stereo imaging system. This method is based on the principle of using space division to achieve large-scale elemental image display with a small number of display panels.

The second is the use of a projection device through the time-division multiplexing method, each display panel in time order. The subimage will be displayed in the corresponding area, and the projection is faster than the human eye visual inertia. Increasing the system pixel density or sampling frequency by the above methods will increase the complexity of the device, causing system errors and noise, increasing the difficulty of implementation, and affecting the system effect. The following proposes a method to improve the reconfiguration resolution of the integrated imaging display without increasing the system equipment and reducing the hardware difficulty. The method is to achieve the purpose of increasing the pixel information of the element image by swapping subpixel components between a single image point and its neighbouring image points and then realize the high-resolution display of the integrated imaging by effectively realizing the fast playback of the element image through the playback characteristics of LEDs.

In the acquisition phase of the integrated imaging system, the acquisition of 3D scenes can be achieved by two methods: the optical system-based acquisition process and the computer system-based acquisition process. The optical-based acquisition process is to record 3D scene information through micro lens arrays, which greatly depends on the parameters of the micro lenses. The computer-based acquisition process is an ideal image acquisition process in which the camera simulates the micro lens array and records the 3D scene information through virtual micro lenses, avoiding the errors caused by the optical system. In this paper, 3Ds Max, a 3D modelling and animation software, is used to simulate the optical environment to achieve the 3D scene information acquisition, and a virtual camera is used to simulate the lens array and acquire the element image array by rendering. The acquisition of element images by a virtual system can overcome the crosstalk and error problems between adjacent elements of the pure optical system. 3Ds Max is easy to operate, powerful, and widely used in the

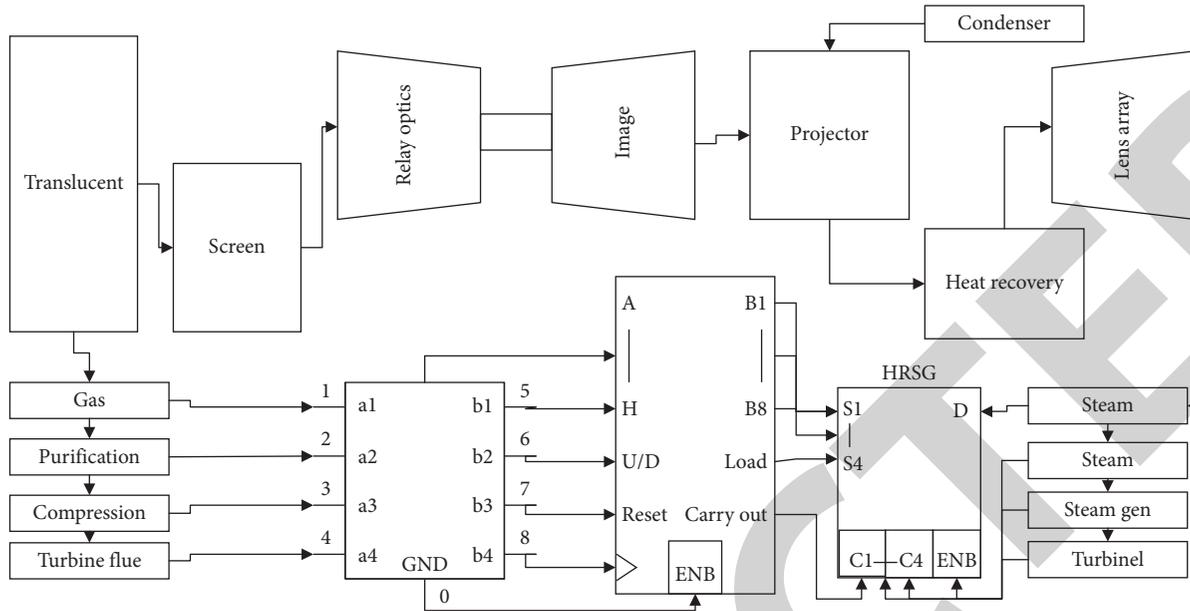


FIGURE 3: Projection method of air-separation multiplexing.

design, video production, animation, and other visualization fields with many basic users.

Create a 3D scene; create 3D objects; adjust the shape, size, coordinates, and colour rendering of the objects; set virtual camera parameters and array; calculate the parameters of the lens array used in reconstructing the scene through the formula to get the virtual camera parameters and then set the virtual camera parameters, number, shooting distance, etc. in 3Ds Max simulation software; shoot and render the image and output, shoot through the virtual camera, and then render to get a small local 3D scene and output to the corresponding folder. The output image is flipped and stitched, and the output image is flipped and stitched to get the element image array by scripting the output image first and then output the element image array to the specified folder, as shown in Figure 4.

In this experiment, two sets of experiments were conducted for verification. To facilitate the setting of the model position and camera position, the 3Ds Max workspace interface was set up with four display screens, namely front view, left view, top view, and camera field of view, so that the captured model could be clearly and intuitively seen when shooting. The first set of experimental model information acquisition is carried out below. The parameters of the micro lens array used for reconstruction in this group of experiments are determined, so the parameters of the free camera set in 3Ds Max should match the parameters of the micro lens array, then set a dice model, adjust the model to a certain position and angle, adjust the camera angle to align with the target model, and set the camera parameters.

4. Analysis of Results

4.1. Virtual Reconstruction Technique Image Acquisition and Feature Processing Results. Figure 5 shows the comparison of the dense correspondence accuracy of the central region

of the face after using different methods for alignment on BJUT-3D and Face-lab, respectively. As can be seen from the table, on the BJUT-3D dataset, the average error of our method is improved by about 0.023 mm or about 4%, compared with Booth's method, and the average error of Liang's method is improved by about 0.02 mm or about 3%, compared with Liang's method. On the Face-Lab dataset, the average error is improved by about 0.02 mm or 4% compared to Booth's method, and the average error is improved by about 0.19 mm or 3% compared to Liang's method.

The topological loss is defined according to the 3D face mesh model to quantify the topological structure change of the 3D face, and then the topological loss constraint is introduced into the objective function of NICP so that the topological structure change during the 3D face deformation can be strongly constrained without maintaining the topological structure by significantly reducing the distance loss weight. Also, a weight adaptation strategy is used in the centre region of the face for the problem that a small number of points with large errors exist in the centre region of the face (around the eyes, near the nose, and at the corners of the mouth) after the dense correspondence of the 3D face. The experimental results show that our proposed algorithm has a significant improvement in dense correspondence accuracy compared with similar algorithms while maintaining the topology of the face well.

In the stationary state, the attitude estimation algorithms proposed in this paper and the Quaternion-EKF algorithm are both very stable, and the mean error and root mean square error of the attitude estimation in all three axes are less than 1° . The algorithm shows a large fluctuation in the yaw angle, and the mean error and root mean square error are more than 1° . Due to the relatively poor stability of the magnetometer measurements, the estimation errors of the three algorithms on the yaw axis are significantly higher than those on the other two axes. Under static experiments, the

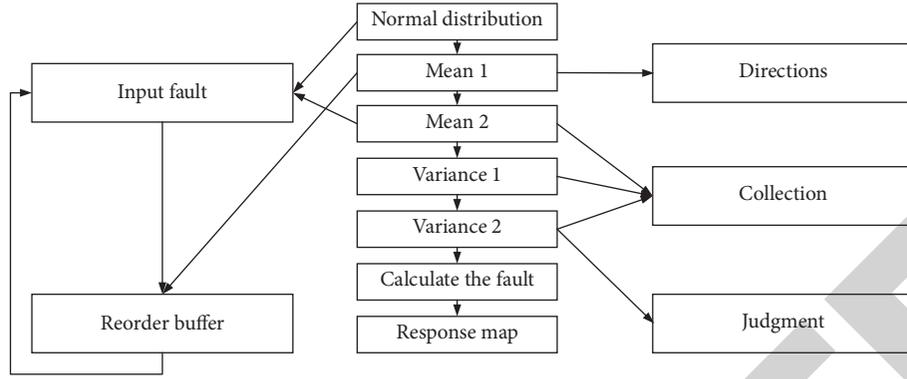


FIGURE 4: Experimental step design.

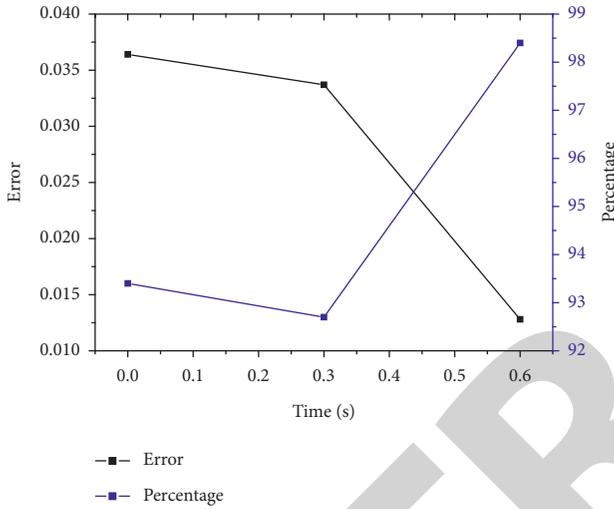


FIGURE 5: Comparison of the alignment accuracy of the central region of the face under different methods.

performance of the attitude estimation algorithm proposed in this paper is comparable to that of the Quaternion-EKF algorithm and slightly better than that of the algorithm. The normal rotation experiment was conducted to test the estimation accuracy of the algorithm in normal rotation (no acceleration/deceleration motion). Figure 6 plots the pose estimation values of the three algorithms against the pose reference values in the normal rotation motion.

Human body movements can be divided into static movements and dynamic movements. Static movements can be represented by the posture or joint angles of the limbs, which are relatively simple to identify, and only need to compare the posture or joint angles of the limbs. Dynamic movements usually involve only part of the limb movement, which can be composed of a single limb movement, such as head twisting or head bowing, or the coordinated movement of multiple limbs, such as hand gestures that involve the upper arm, forearm, and hand of the human upper limb, which is relatively more complex to recognize. To reduce the amount of data and processing time, we only focus on the

motion of the participating limbs in human action recognition; i.e., we use the relative motion of the participating limbs to describe the human action.

4.2. Analysis of Experimental Results. The processing effectiveness of the proposed octree-based convolutional neural network segmentation algorithm in this paper is experimentally evaluated. The experiments in this paper were conducted on 10 different types of partial objects randomly selected from the Princeton Segmentation Benchmark (PSB) dataset, including the human body, airplane, cup, chair, and table, and compared with the shape diameter function segmentation algorithm, Rand Cuts, and other segmentation algorithms to evaluate the effectiveness of the proposed segmentation algorithm. Figure 7 shows the results of the segmentation of airplane and cup models using Rand Cuts, shape diameter function-based segmentation algorithm, and the octree-based convolutional neural network segmentation algorithm proposed in this paper, which qualitatively evaluates the performance of various segmentation algorithms including the segmentation algorithm proposed in this paper. The first row of Figure 7 shows the cut result model obtained by the algorithm Rand Cuts; the second row of Figure 7 shows the cut result model obtained by the algorithm Shape Diameter Function segmentation algorithm; the third row of Figure 7 shows the resulting model obtained by the octree-based convolutional neural network segmentation algorithm proposed in this paper. By analysing the results of Figure 7, we can see that the segmentation algorithm proposed in this paper has less noise and smoother edges when segmenting adjacent regions and boundary segmentation, which has better results compared with the other two segmentation algorithms. The IoU criterion is mainly used to calculate the correlation between the actual label value and the predicted label value, and the higher the correlation, the better the segmentation effect. The comparison of the results in Figure 7 shows that the 3D model segmentation method proposed in this paper gets better segmentation results when dealing with models such as airplanes, chairs, and cups.

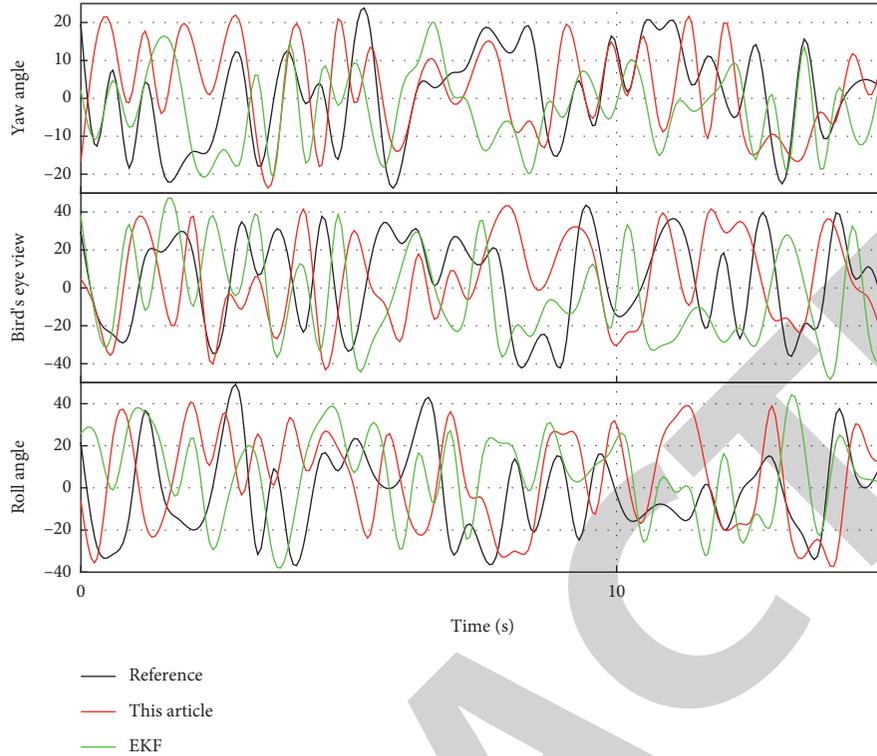


FIGURE 6: Comparison of attitude estimation errors of three attitude algorithms during normal rotation.

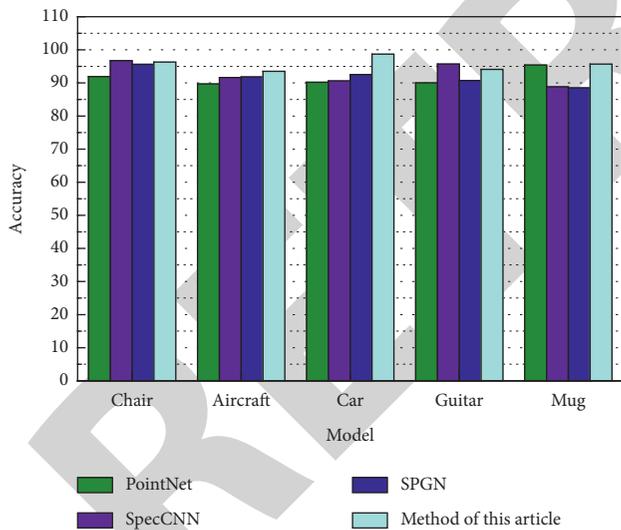


FIGURE 7: Comparison results of segmentation methods for IoU metric.

The second set of 3D reconfiguration optical experiments is carried out below. The optical experimental platform includes an LED screen, micro lens array, and supporting slide table. The parameters of the LED screen are 384×384 pixels, 1.25 mm pixel size, and the parameters of the micro lens array for the second group of experiments are

single-lens diameter d_{10} mm, lens focal length f_7 mm, lenses placed adjacent to each other without interval, the overall number of the lens array is 48×48 , the lens material is acrylic (optical PMMA), the light transmission rate is 93%, and the scientific working temperature is $-30^\circ \pm 80^\circ$. The 3D image obtained from the single image reconstruction is shown in Figure 8 below. By using the 3D deformation model to create a personalized template that adaptively affects coarse-to-fine reconstruction schemes, we can effectively create a more accurate model than the previous work, as shown in the synthetic and realistic data experiments. There are many avenues for future work, such as fusing 3D deformation models and photometric stereo-based reconstructions so that it can gracefully downscale to individual images and automatically identify the level of detail possible to reconstruct from an arbitrary collection of photographs.

Subjectively, the outline of the kettle reconstructed by the time-division multiplexing method is clearer, the green circle is imaged, the edges are neat, the edges are clearer than the single element image reconstructed image, the image three-dimensional effect is obvious, the resolution is higher, and the resolution of the reconstructed image is improved. Analysis of the above experimental results: the method of increasing the resolution by increasing the pixel information around the pixel points through super-resolution reconstruction of the element images after extracting the element image array and then by quickly putting it on the LED large

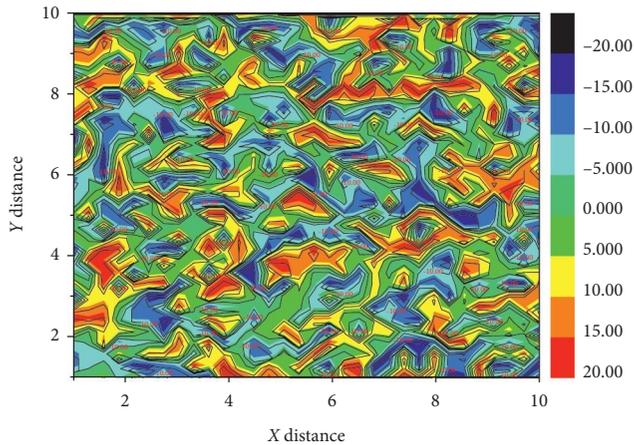


FIGURE 8: Reconstructed image of single element image.

screen is feasible, and in the experiment, the reconstructed image has clear edge contour, complete shape, high imaging resolution, and obvious stereo effect.

5. Conclusion

In this paper, we propose an improved NICP-based dense correspondence algorithm for 3D faces, which allows the algorithm to maintain the face topology without significantly sacrificing the alignment accuracy. The topology of the 3D face is first quantified, and the topological loss term is constructed by using the difference between triangles in the mesh model, and then the topological loss term is added to the objective function of NICP, so that the topology of the face template is strongly constrained during the deformation process, without the need to maintain the topology by significantly reducing the distance loss weight of the point cloud. The affine transformation matrix of each point is obtained to realize the deformation of the face template. A weight adaptive strategy is used in the central region of the face for the problem of large errors in a small number of points in the corners of the mouth and the corners of the eyes. The experimental results show that the algorithm proposed in this paper improves the dense correspondence accuracy while maintaining the 3D face topology. Through the projection of 3D key point coordinates to 2D key point coordinates in two directions, the distance between the projected 2D key point coordinates and the acquired 2D skeletal key point coordinates is calculated separately, and the direction with the smaller distance is finally selected as the direction of the target in the image.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study was funded by the National Social Science Foundation Art Project: Comparative Study on Cultural Value Orientation and Communication Effect of Contemporary Tibetan Film and Television Works Domestic and Overseas (No. 17CC188).

References

- [1] K. Bladin, E. Axelsson, E. Broberg et al., "Globe browsing: contextualized spatio-temporal planetary surface visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 802–811, 2017.
- [2] H. Rahaman and E. Champion, "To 3D or not 3D: choosing a photogrammetry workflow for cultural heritage groups," *Heritage*, vol. 2, no. 3, pp. 1835–1851, 2019.
- [3] Y. Zhang, C. Chen, Q. Wu et al., "A Kinect-based approach for 3D pavement surface reconstruction and cracking recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 3935–3946, 2018.
- [4] A. Salim Rasheed, R. Hasan Finjan, A. Abdulsahib Hashim, and M. Murtdha Al-Saeedi, "3D face creation via 2D images within blender virtual environment," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 1, pp. 457–464, 2021.
- [5] S. Zhang, H. Yu, T. Wang, L. Qi, J. Dong, and H. Liu, "Dense 3D facial reconstruction from a single depth image in unconstrained environment," *Virtual Reality*, vol. 22, no. 1, pp. 37–46, 2018.
- [6] A. Haleem and M. Javaid, "3D scanning applications in medical field: a literature-based review," *Clinical Epidemiology and Global Health*, vol. 7, no. 2, pp. 199–210, 2019.
- [7] G. Zhou, H. Hu, and L. Ma, "Progress and review of 3D image feature reconstruction," *Recent Patents on Computer Science*, vol. 10, no. 1, pp. 43–52, 2017.
- [8] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering," *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 1–12, 2019.
- [9] S. Xia, L. Gao, Y.-K. Lai, M.-Z. Yuan, and J. Chai, "A survey on human performance capture and animation," *Journal of Computer Science and Technology*, vol. 32, no. 3, pp. 536–554, 2017.
- [10] J.-R. Chang and P.-Y. Yu, "Weighted-fuzzy-relations time series for forecasting information technology maintenance cost," *Granular Computing*, vol. 4, no. 4, pp. 687–697, 2019.
- [11] C. Li and F. Cohen, "In-home application (App) for 3D virtual garment fitting dressing room," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5203–5224, 2021.
- [12] I. Ferrari and A. Quarta, "The Roman pier of San Cataldo: from archaeological data to 3D reconstruction," *Virtual Archaeology Review*, vol. 10, no. 20, pp. 28–39, 2019.
- [13] M. Pfarr-Harfst, "Digital 2D and 3D visualisations as iconic epistemological models," *SCIRES-IT-SCIENTIFIC RESEARCH and Information Technology*, vol. 10, no. 1, pp. 19–30, 2020.
- [14] M. Cao, L. Zheng, W. Jia et al., "Accurate 3-D reconstruction under IoT environments and its applications to augmented reality," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2090–2100, 2020.
- [15] G. M. Diaz, D. Mohr-Bell, M. Garrett, L. Muñoz, and J. D. Lencinas, "Customizing unmanned aircraft systems to reduce forest inventory costs: can oblique images substantially improve the 3D reconstruction of the canopy?" *International Journal of Remote Sensing*, vol. 41, no. 9, pp. 3480–3510, 2020.

- [16] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," *Sports Medicine-Open*, vol. 4, no. 1, pp. 24–15, 2018.
- [17] W. N. Deng, S. Wang, J. Ventrici de Souza, T. L. Kuhl, and G.-Y. Liu, "New algorithm to enable construction and display of 3D structures from scanning probe microscopy images acquired layer-by-layer," *The Journal of Physical Chemistry A*, vol. 122, no. 26, pp. 5756–5763, 2018.
- [18] C. Mousas and C.-N. Anagnostopoulos, "Real-time performance-driven finger motion synthesis," *Computers & Graphics*, vol. 65, no. 3, pp. 1–11, 2017.
- [19] N. A. Bushmeleva, N. I. Isupova, E. A. Mamaeva et al., "Peculiarities of engineering thinking formation using 3D technology," *European Journal of Contemporary Education*, vol. 9, no. 3, pp. 529–545, 2020.
- [20] C. Kapeller, B. Sespede, M. Nezveda et al., "Objective and subjective evaluation of a multi-stereo 3D reconstruction system," *Electronic Imaging*, vol. 2020, no. 2, pp. 138–141, 2020.
- [21] W. W, X. Xia, M. Wozniak, X. Fan, R. Damaševičius, and Y. Li, "Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels," *Computer Networks*, vol. 161, pp. 210–219, 2019.
- [22] X. Zenggang, T. Zhiwen, C. Xiaowen et al., "Research on image retrieval algorithm based on combination of color and shape features," *Journal of Signal Processing Systems*, vol. 93, pp. 1–8, 2019.
- [23] Q. Jiang, F. Shao, W. Gao et al., "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2018.
- [24] M. Chen, S. Lu, and Q. Liu, "Uniform regularity for a Keller-Segel-Navier-Stokes system," *Applied Mathematics Letters*, vol. 107, Article ID 106476, 2020.
- [25] K. Sim, J. Yang, W. Lu, and X. Gao, "Blind stereoscopic image quality evaluator based on binocular semantic and quality channels," *IEEE Transactions on Multimedia*, vol. 2021, Article ID 3064240, 1 page, 2021.
- [26] J. S. Almeida, P. P. Rebouças Filho, T. Carneiro et al., "Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques," *Pattern Recognition Letters*, vol. 125, pp. 55–62, 2019.
- [27] J. Yang, M. Xi, B. Jiang et al., "FADN: fully connected attitude detection network based on industrial video," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2011–2020, 2020.