

A proposed reconstruction method of a 3D animation scene based on a fuzzy long and short-term memory algorithm

Ming Zhou

Zhou Ming Anhui Sanlian College, Hefei, China

ABSTRACT

With the development of computer technology leading to a broad range of virtual technology implementations, the construction of virtual tasks has become highly demanded and has increased rapidly, especially in animation scenes. Constructing three-dimensional (3D) animation characters utilizing properties of actual characters could provide users with immersive experiences. However, a 3D face reconstruction (3DFR) utilizing a single image is a very demanding operation in computer graphics and vision. In addition, limited 3D face data sets reduce the performance improvement of the proposed approaches, causing a lack of robustness. When datasets are large, face recognition, transformation, and animation implementations are relatively practical. However, some reconstruction methods only consider the one-to-one processes without considering the correlations or differences in the input images, resulting in models lacking information related to face identity or being overly sensitive to face pose. A face model composed of a convolutional neural network (CNN) regresses 3D deformable model coefficients for 3DFR and alignment tasks. The manuscript proposes a reconstruction method for 3D animation scenes employing fuzzy LSMT-CNN (FLSMT-CNN). Multiple collected images are employed to reconstruct 3D animation characters. First, the serialized images are processed by the proposed method to extract the features of face parameters and then improve the conventional deformable face modeling (3DFDM). Afterward, the 3DFDM is utilized to reconstruct animation characters, and finally, high-precision reconstructions of 3D faces are achieved. The FLSMT-CNN has enhanced both the precision and strength of the reconstructed 3D animation characters, which provides more opportunities to be applied to other animation scenes.

Submitted 28 November 2023

Accepted 18 January 2024

Published 29 February 2024

Corresponding author

Ming Zhou, shanyuzm@163.com

Academic editor

Junaid Shuja

Additional Information and
Declarations can be found on
page 15

DOI 10.7717/peerj-cs.1864

© Copyright
2024 Zhou

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Algorithms and Analysis of Algorithms, Artificial Intelligence, Computer Vision, Data Mining and Machine Learning, Neural Networks

Keywords Temporal neural network, 3D animation, Scene reconstruction, Deep learning

INTRODUCTION

One way human beings access external information is through vision, which is screened, integrated, and analyzed through the brain system so that humans understand and feel the world around them. Research on computer vision, which enables machines to extract and analyze information from vision-based data quickly, has received great attention, similar to how humans achieve a reasonable understanding of external environment information.

A three-dimensional (3D) reconstruction is one of the substantial fields in computer vision, and its related technologies are widely implemented in film animation, robot navigation, medical imaging, relic protection, and many other fields (Wu et al., 2021). Scene simulation is an important research field of computer graphics, which is widely implemented in 3D video games, special effects, computer animation, military simulation, virtual reality, and augmented reality. 3D face reconstruction technologies and face animation are important research in computer graphics and focus on constructing 3D animation scenes (Casas et al., 2019). The conventional 3DFR technology mainly relies on costly 3D scanning devices, requires post-manual processes that take a long time, and necessitates scanned character subjects to maintain a fixed posture for a significant time. Moreover, hardware systems that enable running computations on a large scale have been developed recently, especially the development of distributed computing, which enables GPUs to generate efficient solutions.

In recent years, academic research and industry-based implementations have made great breakthroughs in detection and identification processes. The related research on the human body, especially the human face, has become a widely concerned research topic, and the scenes are also commonly employed in real life. The demand for 3D face models in face recognition, ARR technology, game animation, medical beauty, and other fields has increased. For example, 3DFR and face animation are two important research fields in computer graphics. Due to the complex spatial structure and rich fine attributions, 3D face reconstruction and face animation technologies have grown to become pivotal areas in the generation of computer graphics.

A face is the most recognizable part of an individual. Still, the characteristics of human faces in 2D space are generally constrained by uncertain factors such as lighting, facial expression, angle of face, occlusion, etc. Therefore, studying human faces in 3D space has become a breakthrough in the next step of development (Wang & Zhai, 2019).

Keeping in mind that face videos help reconstruct 3D faces in 3D face animation scenes by making full use of potential face geometric structure and apparent information, and considering the correlation of different images in time, thus mastering the properties of face features, finally recovering the real 3D face geometric model as close as possible is an open problem. The 3D face model generated by employing the proposed network (FLSTM-CNN) could reflect certain personalized features and result in certain effectiveness and robustness. Moreover, it can bring significant convenience to music users, so it presents a substantial research area and wide implementation possibilities.

The manuscript is sectioned as follows: 'Introduction' introduces the research background and motivation, the predominant methods, and the effects of the reconstruction. 'Preliminary' introduced the sequential deep neural network and the 3D reconstruction algorithm. 'The proposed method' introduced the proposed model, modeled the human face, and evaluated the reconstruction effect. 'Experimental results and discussion' mainly summarizes the conducted research and presents the future direction.

PRELIMINARY

A 3DFR has been one of the popular study areas in computer vision. Many enterprises, institutes, and universities have invested a lot of funds and resources and generated a lot of distinct algorithms and methods to improve the impact of a 3DFR. The robustness of the algorithms has advanced greatly (Mai et al., 2018). Moreover, continuous research progress and technological innovations have enhanced several methods of 3DFR that could be split into two classes: 3DFR employing conventional approaches and deep learning algorithms (Jin et al., 2017).

Parke (1982) was the first to propose a face parameterization model that employed the grid representation of a face utilizing about 400 vertices and 250 polygons by changing the structure and action of the grid surface, implementing rotation, translation, and scale, which makes face shapes change. Thus, the characteristics of personalized faces were initially realized between different individuals (Becker & Fregosi, 2017). Muscle models have emerged to cope with the problem that the proposed parametric models cannot accurately represent facial features and expressions. Zhao (2001) connected the elastic features of the facial skin vertices, utilized force on the facial muscles to realize the representation of the facial features, and established a 3D facial model (Hernandez et al., 2017). A muscle model that can be employed to simulate the process of facial expression changes was also proposed by Waters (1987). The model caused the deformation of muscles by utilizing the force when the expression changed and made the face grid represented by the polygon deformation. This model could align the expression information with the real facial expression (Merras et al., 2017). Lee et al. (2022) proposed to employ a high-precision 3D face scanner to attain 3D point cloud data and facial texture information with depth information. Afterward, the corresponding 3D face algorithm is constructed according to the obtained information. The reconstruction results obtained by this method were relatively accurate, but the equipment was expensive, and the data acquisition was limited (Shirowzhan et al., 2020). Blanz & Vetter (1999) first proposed the famous 3D deformation model that led to the implementation of statistical analysis, which is of landmark significance for 3DFR (Turani, 2017).

Methods employing deep learning algorithms (DLAs) have advanced swiftly and provided superior performance in both computer vision and image processing. Many scholars have also started to employ DLAs to examine 3DFR. Zhu et al. (2019) proposed a 3D face alignment network that converted the dense 3D face algorithm into a 2D face image through CNNs to improve face alignment performance at large poses. To resolve the issue of the unnoticed significant points of the face under the big posture (Tu et al., 2020), a robust approach for regression-separable 3D deformation models (3DFDM) was designed by Tran et al. (2017). The researchers suggested employing a CNN to directly regress the shape and texture parameters from the input photos and then fit the 3D face approach. The accurate 3D face approach could be reconstructed by inputting a single image and was faster when 3DFDM parameters were estimated. Tran & Liu (2018) proposed a nonlinear 3DFDM method to model shapes or textures through deep neural networks (DNNs) (Guo et al., 2018). The conventional 3DFDM employed linear basis

approaches for shapes or textures, which were trained on 3D face scans and associated 2D images. However, this method did not require 3D scanning by implementing weak supervision. It could be trained from naturally unconstrained 2D face images that align with nonlinear changes such as facial expressions and gestures. So, the original image could be better reconstructed with a stronger representation, and the performance of related tasks was further improved (Feng et al., 2018b). The voxel regression network (VRN) proposed by Jackson et al. enabled the 3DFR of 2D face images utilizing arbitrary facial poses and expressions without constructing or fitting 3DFDM parameters during training and testing stages (Yueyang, Wan & Junkaiyi, 2019), respectively. This method could obtain an end-to-end representation of 3D facial geometry, while it had the problem of a large overall computational burden and limited resolution in reconstructions. Feng et al. suggested an end-to-end position mapping regression network (PRN) that could realize 3DFR and dense face alignment in two tasks by designing a UV position mapping as a 2D representation method, which was implemented to record the UV space. Thus, it enabled 2D image storage for 3D face cloud coordinates. Each UV polygon contained the corresponding semantic information (Zhang et al., 2021). Richardson, Sela & Kimmel (2016) reconstructed the 3D face geometry by CNNs regression to the 200-dimensional identity vector and 84-dimensional expression vector. They then aligned the recovered face mask to the input image and optimized the face shape by utilizing the network to obtain the reconstruction results with rich, detailed features. With the utilization of an unsupervised training method combined with the 3DFDM proposed by Genova et al. (2018) the face recognition network was innovatively introduced into the training loss function based on the encoder and decoder models so that the generated 3D faces could retain the individual face features of the input pictures (Trần et al., 2018).

The literature suggests that the demand for implementations constantly increases, and employing 2D images to recover 3D face data is a substantial qualitative problem (Peng & Tian, 2017). Therefore, the current research work still faces many challenges. This is mainly due to the serious problems of available single-view 3DFR methods when applied in real scenes. The 3D face shape estimation implementing a single image is either unstable or could be different when different photos of the same object are presented, or the estimated 3D face shape is overgeneralized but not applicable. However, collecting large amounts of 3D scanned facial data simultaneously is laborious and impractical (Rui et al., 2018). Instead, the large quantity of 2D face data could help better. Therefore, to improve the learning capability and strength of the algorithms, it is important to examine the 2D face reconstruction model. First, a 3D scene reconstruction algorithm based on temporal deep learning, whose basic framework is shown in Fig. 1, is presented. With the utilization of the temporal deep neural network, the video-sampled image sequence is learned, and then the face is shaped by the CNN to reconstruct the 3D face animation.

THE PROPOSED METHOD

A brief introduction to RNN and LSTM algorithms

When using a video to extract face images, a series of time series images are obtained. In deep learning algorithms, neural networks that can crunch time-dependent data are the

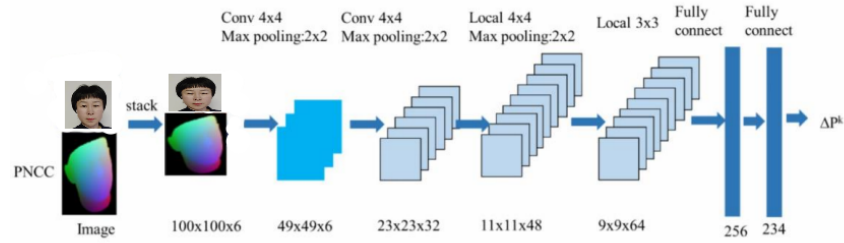


Figure 1 A reconstruction method of a 3D animation scene based on time series deep learning.

Full-size [DOI: 10.7717/peerjcs.1864/fig-1](https://doi.org/10.7717/peerjcs.1864/fig-1)

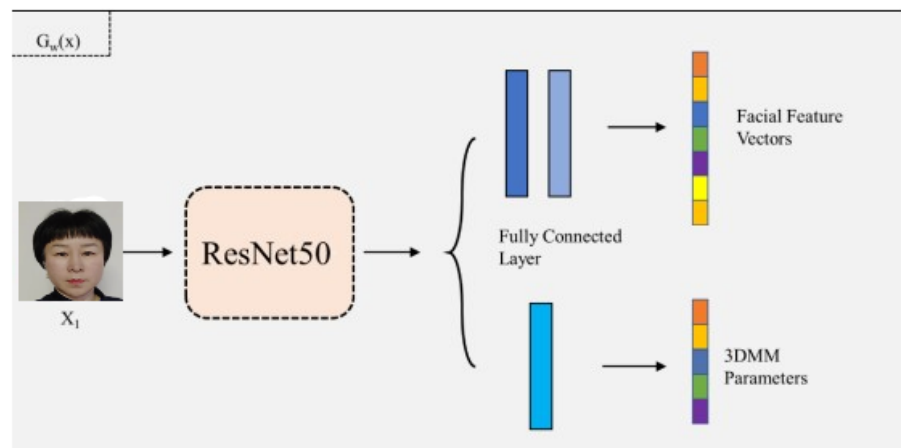


Figure 2 The structure of an RNN.

Full-size [DOI: 10.7717/peerjcs.1864/fig-2](https://doi.org/10.7717/peerjcs.1864/fig-2)

best way to process time series data. A recurrent neural network (RNN), a fully connected neural network with self-loop feedback processing time series data, is presented in Fig. 2. The result of its neuron is passed onto itself at the next time point, and it also outputs a hidden layer state for the current layer to be implemented when the sampled observations are processed at the next time point.

The input, hidden, and output layer vectors are denoted by x , s , and o , respectively. U , V , and W denote the coefficient matrices of the input layer and hidden layer, the hidden layer to the output layer, and weights, respectively. The score s of the hidden layer is obtained from the current moment input x and the hidden layer s of the previous moment. As shown in Fig. 1, the s_t at time t is decided by x_t and s_{t-1} .

The forward propagation of the RNN is employed to predict the next value. For example, y_t represents the true value at time t , and x_t denotes the input for time t , then \hat{y}_t represents the predicted value at time t . s_t is the hidden state at time t . The output is denoted by o_t .

$$s_t = \varnothing(W_x x_t + W_x x_{t-1} + b_s) \quad (1)$$

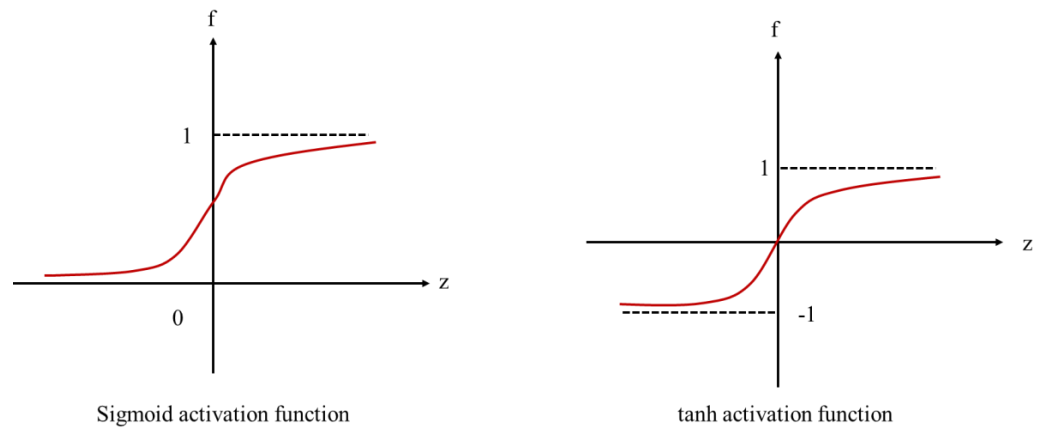


Figure 3 Activation functions used in the RNN-LSTM.

Full-size  DOI: [10.7717/peerjcs.1864/fig-3](https://doi.org/10.7717/peerjcs.1864/fig-3)

$$o_t = W_0 x_t + b_0 \quad (2)$$

$$\hat{y}_t = \sigma(o_t) \quad (3)$$

where σ represents the activation function that is tanh, W_X denotes the coefficient matrix of the input layer to the hidden layer; W_S denotes the coefficient matrix between the hidden layers; W_O is the weight matrix from the hidden layer to the output layer; b_S is the offset item; b_O is the bias item; σ is an active function, generally called the softmax function. For example, at time t_3 , the loss function is defined by

$$L_3 = \frac{1}{2}(y_3 - o_3)^2 \quad (4)$$

The RNN's backpropagation process utilizes the stochastic gradient descent method to guide the bias term in the network and continuously update it to minimize the loss value L . However, when the RNN gets trained, gradient disappearance or explosion (GDE) could occur. So, the long and short-term memory network (LSTM) is implemented to overcome this issue.

The LSTM, initially suggested by [Hochreiter & Schmidhuber \(1997\)](#), is based on the improvement of the RNN and can recall both long and short-term information. The LSTM is composed of an LSTM unit with a chain structure and adds thresholds to the algorithm's input, forgetting, and output layers, respectively so that the coefficient of its cycle alters, permitting the network to forget the currently gathered information, thereby preventing GDE. Forget, input, and output gates denoted by f , i , and o , respectively, in the LSTM unit are input from the hidden state h_{t-1} of the prior moment and the current time step input x_t , and their activation mappings are depicted in [Fig. 3](#).

The threshold mechanism of the LSTM network allows it to stash information for a long time, thus preventing GDE. Input, forgetting, and output gates are placed in each unit,

and all employ the sigmoid activation function to control the transmission of information in the network, so they assign certain information at the current moment and then to the information required by the network at the following time. The value domain of all the gates is $\{0,1\}$, where 0 denotes all representative information forgotten, and 1 is the opposite. Input, forget, and output gates control which information remains at the current time step, which prior time step information must be forgotten, and which current time step information must be output, respectively.

$$C_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$\hat{C}_t = C_{t-1} * f_t + \delta_t * i_t \quad (6)$$

$$h_t = \tanh(C_t) * o_t \quad (7)$$

The predicted score of the model at time t is delineated by

$$\hat{y}_t = \sigma(W_{h_t} + C). \quad (8)$$

when compared with the RNN's solution process to derive bias parameters, multiple branches appear in the LSTM, accumulating in every successive step, and the calculation parameters are directly or indirectly related to the matrix W_{xf} . As it increases, the cumulative term increases along with the gradient calculation. Due to these additive terms, the vanishing gradient problem in the RNN is effectively resolved by LSTM.

3D face deformation model (3DFDM)

Network regression parameters are usually employed as the backbone of a time-dependent neural network. The coefficients of a 3D face deformation model (3DFDM) complete 3DFR by utilizing a CNN to input extracted images' face features and then map the feature vector to 3DFR parameters to finalize the whole reconstruction process (Lium et al., 2021). A 3DFDM-based 3DFR method is implemented to determine the shape robustness to any face posture, namely, the different face postures of the same identity employing the network regression parameters are based on similar shape parameters. In other words, the 3D face form of the same object recovering different attitudes of the 3D face shapes should be close to its original 3D face shape. If a deviation occurs, the algorithm constraints are employed to make it as close as possible to the 3D face shape. Therefore, the problem is converted into a clustering problem. To address the problem, two different objects were selected from the 300W-LP dataset, reducing the problem to dichotomous problems, as depicted in Fig. 4.

The 3D face shapes recovered from the 3DFDM parameters predicted by the two objects may differ. With the inclusion of the scale, shape, or expression aspects, the 3DFDM parameters are projected into a 2D space. Then, the 3DFDM parameter distribution of

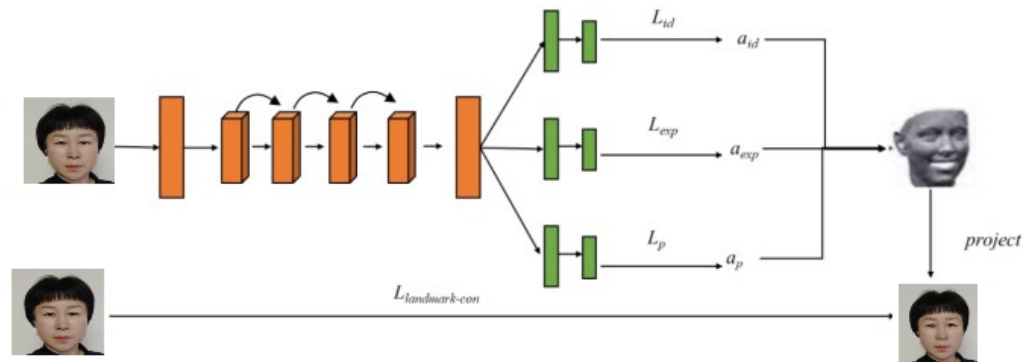


Figure 4 The schematic diagram of the 3DFDM parameter attribute decomposition network.

Full-size [DOI: 10.7717/peerjcs.1864/fig-4](https://doi.org/10.7717/peerjcs.1864/fig-4)

different poses would become more discrete. However, the 3D face shapes reconstructed by 3DFDM parameters are more consistent after shape constraints are embedded. The projection of the 3DFDM parameters of different poses in the 2D plane is also relatively concentrated with increasing interclass distance. The robustness of 3D face shapes is effectively resolved with the research results of 2D face images; that is, the predicted 3DFDM parameter vectors are clustered while reducing the intra-class length and increasing the inter-class difference.

When constructing a 3DFDM, the modeling is based on the available database of faces, and all the 3D face models in the database are represented by a triangular grid composed of the same number of vertices and triangular face slices so that the same index point has the same semantic information. The 3D facial geometry of a human face can be expressed in the vector form in Eq. (9).

$$S = (x_1, y_1, z_1, L x_n, y_n, z_n)^T \in R^{3n} \quad (9)$$

where S and n represent the shape vector, and the total vertex numbers that constitute the face model, respectively. There are m numbers of different 3D face samples in the known face database. Each of these can be represented by the shape vector S . Therefore, a 3D face geometry of a new face can be obtained through a linear combination of the shape vectors of the m different personal face samples in the knowledge database presented in Eq. (10).

$$S_{new\ model} = \sum_{i=1}^m \alpha_i S_i. \quad (10)$$

With the alteration of the shape coefficient of a human face's mouth structure, a new geometric model of a human face can be generated.

Deep learning algorithm based on fuzzy temporal neural network

The process of face reconstruction implements CNN to master features. Feature extraction is performed by employing a fuzzy LSTM (FLSTM) network proposed by [Pattanayak, Sangameswar Vodnala & Das \(2022\)](#) as the backbone network of the FLSTM-CNN while replacing the last two layers of the network with two parallel fully connected layers (FCL). The FLSTM's algorithm is given as follows:

Algorithm: The suggested FLSTM

Input the time series observations as $y = [y_1, y_2, \dots, y_n]$

T , length of training set ($\ell t r$), length of validation set (ℓ), and length of test set (ℓt).

Output Forecasted score.

1. Calculate θ_0 and θ_1 for the time series observations $\theta_1 = \frac{\sum (t_i - \sum_{i=1}^n t_i/n)}{\sum_{i=1}^n (y_i - \sum_{i=1}^n y_i/n) \sum_{i=1}^n (t_i - \sum_{i=1}^n t_i/n)}$ $\theta_0 = \frac{1}{n} (\sum_{i=1}^n y_i - \theta_1 \sum_{i=1}^n t_i)$
2. Calculate the trend score of each observation in the time series data. *for* $t = 1$ *to* n (*size of TS*) $u_t = \theta_0 + \theta_1 t$ *end of for*
3. Compute the RTV scores *i.e.*, V of each observation in the time series data *for* $t = 1$ *to* n (*Size of TS*) $V_t = (y_t - u_t) \gamma * 100$ *end of for*
4. By using the RTV data's median, it is split into P and Q subgroups.
5. Calculate the absolute first difference Pd and Qd in both P and Q subgroups.
6. Attain the mean of the Pd and Qd as Pv and Qv for both the subsets P and Q , respectively.
7. Attain the deciding factor ϑ , $\vartheta = 10 \text{ floor}(e) (Pv - Qv)$
8. Attain the NOIs *i.e.*, k as follows $k = ((\min(V) - \vartheta) - (\max(V) + \vartheta) \vartheta) + \ell t \times n$
9. Find the UOD as: $\mathcal{U} = (\min(y) - \delta, \max(y) + \delta)$ where δ denotes the standard deviation.
10. Attain the NOIs k of the time series data.
11. Split the UOD $\mathcal{U} = \{u_{e1}, u_{e2}, u_{e3}, \dots, u_{ew}\}$ into k equal interval lengths and Calculate the midpoint m_i ($i = 1, 2, \dots, w$) of each interval. // The time series data is fuzzified.
12. *for* $r = 1$ *to* $\lfloor \text{length of time series} \rfloor$ *if* y_r *belongs to the partition* *uei fr* $(t) = i$ *end of if* *end of for* // Construct the FLR by employing LSTM
13. FLSTM's order is attained by utilizing the autocorrelation function.
14. Employ Min-Max normalization to attain the normalized in sample time series observations $f'(t)$.
15. Convert the $f(t)$, into $n - r$ patterns using sliding window protocol.
16. Split the time series observations into training, validation, and test sets.
17. Attain normalized forecasting outcome $f'(t)$ utilizing the LSTM's optimized parameters.
18. Denormalize the forecasting outcome $f(t)$
19. Attain the forecasted score using defuzzification of $f(t)$ by employing a middle point fuzzy interval and attain the real forecasted score \hat{y} .

The CNN implemented for feature learning is shown in Fig. 5. The FLSTM network is employed as the backbone of the FLSTM-CNN to extract features, while the last two layers of the network are replaced with two parallel fully connected layers (FCLs). One FCL represents a 62-dimensional vector of the 3DFDM parameters; the other fully connected layer outputs a 512-dimensional face feature vector for face recognition. As a sub-network

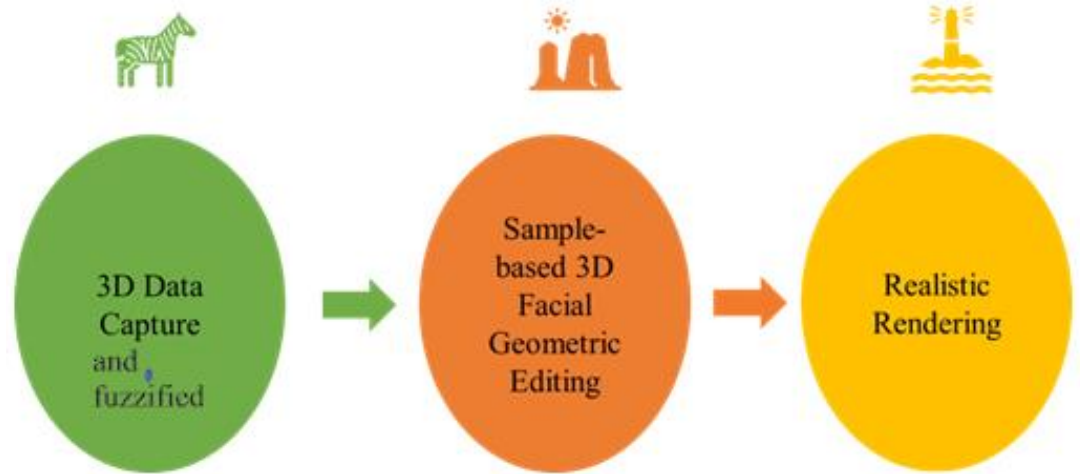


Figure 5 The production planning process of the proposed algorithm.

Full-size DOI: [10.7717/peerjcs.1864/fig-5](https://doi.org/10.7717/peerjcs.1864/fig-5)

of the FLSTM-CNN, the network structure has the same network structure as its symmetric CNN, sharing the network parameters mastered by the model.

In the manuscript, we assume the time series data to be fuzzy or fuzzified, so the fuzzy time series method is applied to DLAs. The generation operation of the proposed FLSTM-CNN algorithm is depicted in Fig. 5. First, the 3D capture software is employed to attain the 3D model observations of the face and the observations are fuzzified. Then, the data processing software is employed to extract attribution points of the 3D animations by implementing the FLSTM-CNN. The modeling software realizes the modeling realization of a 3D animation. The entire model is trained by minimizing the three loss functions.

The first loss function is employed to measure the precision by predicting the 3DFDM parameters, called the weighted prediction of loss function coefficients. The second loss function is employed to maintain the consistency of the face shape. When the image is input, the reconstruction parameters of similar face shapes are output. The third loss function is called the identity loss function, which ensures that the facial images with the same identity have a similar distribution in the feature space. Therefore, the loss function of the overall training is defined by.

$$t = t_{3d} + t_{shp} + t_{id}. \quad (11)$$

Overall, we observed that currently popular 3D face deformation methods, namely, 3DFDM and DeFA, considered only the one-to-one reconstruction process but not the correlation or differential information of the input images in the training set, which led to the reconstruction results being unreliable for face identity representations.

The proposed fuzzy temporal CNN (FLSTM-CNN) implements a many-to-one training strategy to restrict the face geometry of the same person in different poses to the same shape while preserving the identity information of the recovered shape. The proposed algorithm could learn more robust 3DFDM parameters and obtain more discriminative features for identity representations.

The proposed algorithm employs the classical 3DFDM to recapture the 3D face shape from a single image, the same as the 3DFDM. Modifications are made on a theoretical basis to satisfy the requirements of the proposed algorithm. Based on the regression task algorithm utilizing the mean square error criterion to evaluate the accuracy of the predicted values, the failure of weighing the loss is considered a deficiency since network optimization is slow, and the effect of the model prediction is not good. The improved WPDC loss function is chosen as the evaluation function of the 3DFR accuracy combined with the contrast loss function, finalizing the 3D face model training and optimization processes. The heavy-distance cost function is trained for the accuracy of the reconstruction of the 3D face model according to the importance of each predicted 3DFDM parameter.

EXPERIMENTAL RESULTS AND DISCUSSION

Experimental dataset and preprocessing

In the construction process of the 3D face model database employing the FLSTM-CNN, the video data of a user is employed to gather different 2D face images with distinct facial gestures and facial expressions. Then the 3D face model corresponding to the image is obtained as training data. A scanning device of a 3D model (hardware equipment: Structure Sensor) is employed to scan the 3D shape of the user's face and edit the 3D face model grid (3D-Mesh). The face template database for the user's face features is implemented to run further smooth processing. The 3D scanning device can easily and quickly obtain the 3D face model in real-world applications. Still, the scanned 3D model contains large noise, and the topology cannot be effectively unified, which causes some difficulties in the automation process (Feng et al., 2018a).

The process of feature acquisitions first implements the template model and the rigid registration of the target model, namely, rotation, translation, and zoom, which leads to overlapping the centers of the two models. In the 3D face registration model, a model often has certain semantic feature information regarding faces. For the 3D face template model and the same semantic information, a simple semantic correspondence is established between the template and target models based on the corresponding rigid alignment template model of the rotation transformation matrix and target model, respectively. Multiple iterations are run to optimize it. The matching registration of the two 3D models means finding the mapping of the 3D template model and the 3D target model, respectively. Therefore, the mapping must fully express the corresponding vertices with the same semantic information as the two models. The matching registration outcomes show that the 3D template model has the individual characteristics of the 3D target model and the corresponding facial expression features, respectively (Thakrar et al., 2017). Considering the type of target model and registered scene, different matching registration algorithms often show a completely different effect. To build a scene model, the matching registration process is relatively simple. In contrast, the 3D face model has richer facial detailed features and scans the 3D model whose noise is larger, so it needs multi-step continuous iterative optimization.

Although the above 3D face model database and the corresponding data construction method contain different individuals and expressions, the database expresses detailed facial

expressions that could be utilized in computer graphics, vision, and many other fields. The database, composed of 3D facial expressions of 150 individuals of different ages, races, and genders, utilized in the manuscript is called Face-Warehouse (*Liu et al., 2018*). Each has a variety of rich expressions, such as mouth, smile, anger, frown, pout, mouth, and so on. Since the employed 3D face model has the same topology, a 3D tensor containing the database in the experimentation is implemented to construct a bi-linear 3D face model with both individual and expression properties.

The 3D face model data plays a key role in developing practical applications. With the utilization of the available 3D model data, 3D reconstruction and expression animation can be developed. An accurate 3D model generally needs to go through four stages to extract face feature points, generate a face model, adjust the face model, and display the model. However, the available 3D model datasets often have certain limitations in the breadth of data. Systematic production based on available data expands the character library in the breadth of the dataset and enables fast and precise production of individual virtual characters. At the same time, it enriches the facial gesture expressions of the character database in the depth of the data set, providing rich data for computer graphics and visions. Hence, virtual reality, interactive education, and face-driven animation technology utilizing virtual characters provide directly feasible technology solutions for the industry.

Time-dependent deep learning-based scene reconstruction algorithm

Several different algorithms mentioned previously are implemented to reconstruct face scenes, and the effect is shown in [Fig. 6](#). Although the face shape reconstructed by the PRN method is inconsistent with the mouth shape, the suggested algorithm results in the constructed mouth shape with no change. The comparison outcome suggests that the suggested algorithm is more robust than PRN reconstruction when reconstructing 3D face shapes based on the different poses. The implemented deep learning framework is called Pytorch, employing the SGD as the optimizer and implementing the exponential decay learning rate strategy expressed by [Eq. \(12\)](#).

$$lr = \lambda_x * 0.001^{\left(\frac{epoch+1-sd}{te+sd}\right)}. \quad (12)$$

Due to the limitations of the test data, the K-fold cross-validation algorithm is employed to test the accuracy of the face recognition algorithm. However, when the dataset is small, this method can be employed to test the accuracy of the face recognition algorithm to improve data utilization effectively. K-fold cross-validation first splits the observations into K subsets, as shown in [Fig. 7](#). For example, when $K = 10$, the observations are divided into ten equal subclasses, and nine subsets are employed for the training set; the last one is kept as the test set. Then, the average of all recognition accuracy rates based on the recognition accuracy rate of each subset is utilized to compute the final recognition rate.

The manuscript verifies the performance of the suggested attribute decomposition learning network for the 3D face alignments on two datasets, called AFLW and AFL, and W2000.3D, respectively. The test results and the quantitative comparison with other

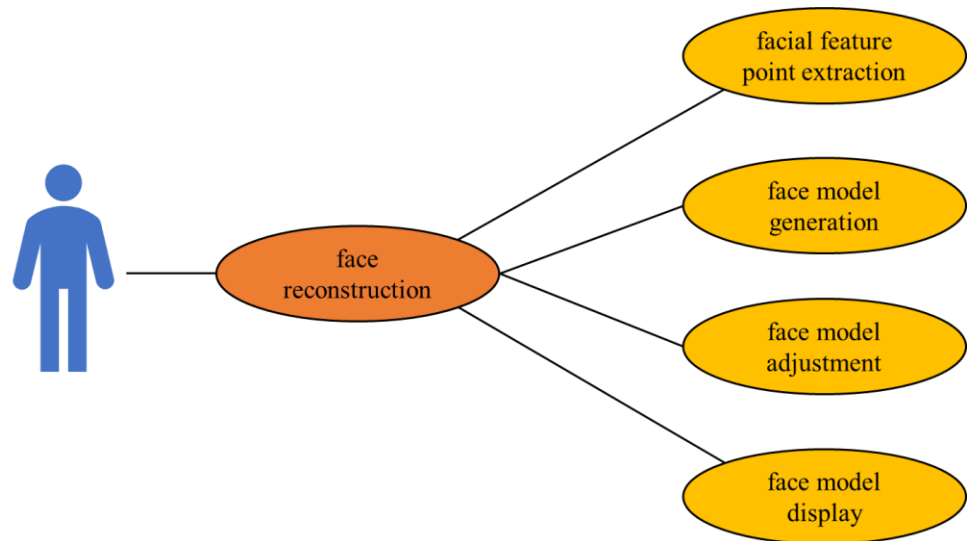


Figure 6 Data renderings.

Full-size DOI: 10.7717/peerjcs.1864/fig-6



Figure 7 The outcomes of k-fold cross-validation.

Full-size DOI: 10.7717/peerjcs.1864/fig-7

conventional approaches are shown by utilizing the cumulative error distribution curve (CED) presented in Figs. 8 and 9, respectively.

The experimental results show that the 3D face scene reconstruction method based on FLSTM-CNN could generate better outcomes.

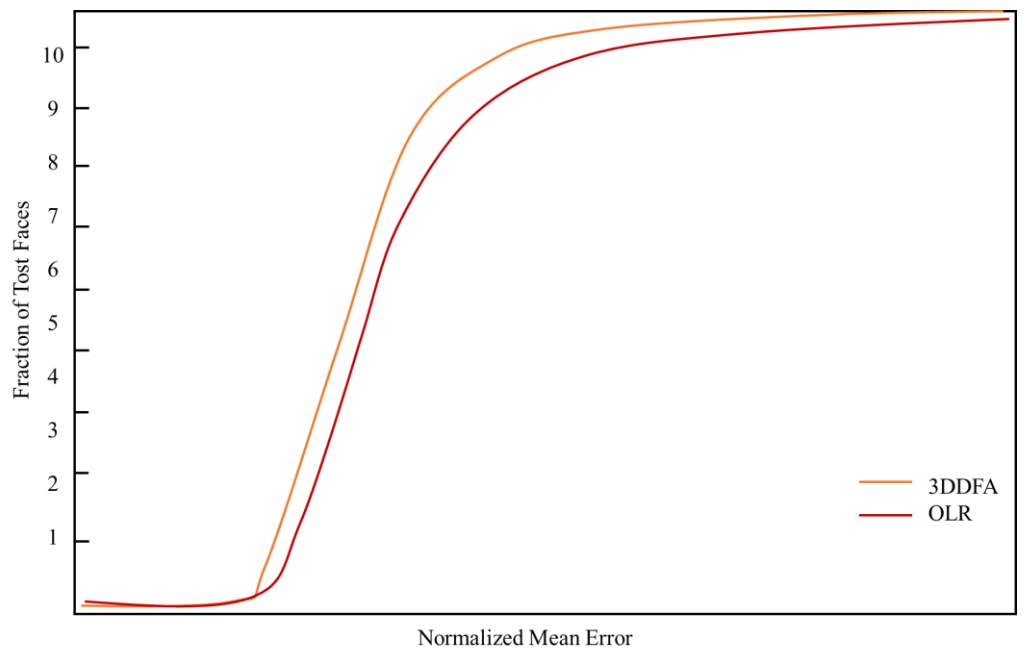


Figure 8 The CED curves of the 3D face alignment outcomes on the AFLW dataset.

Full-size  DOI: 10.7717/peerjcs.1864/fig-8

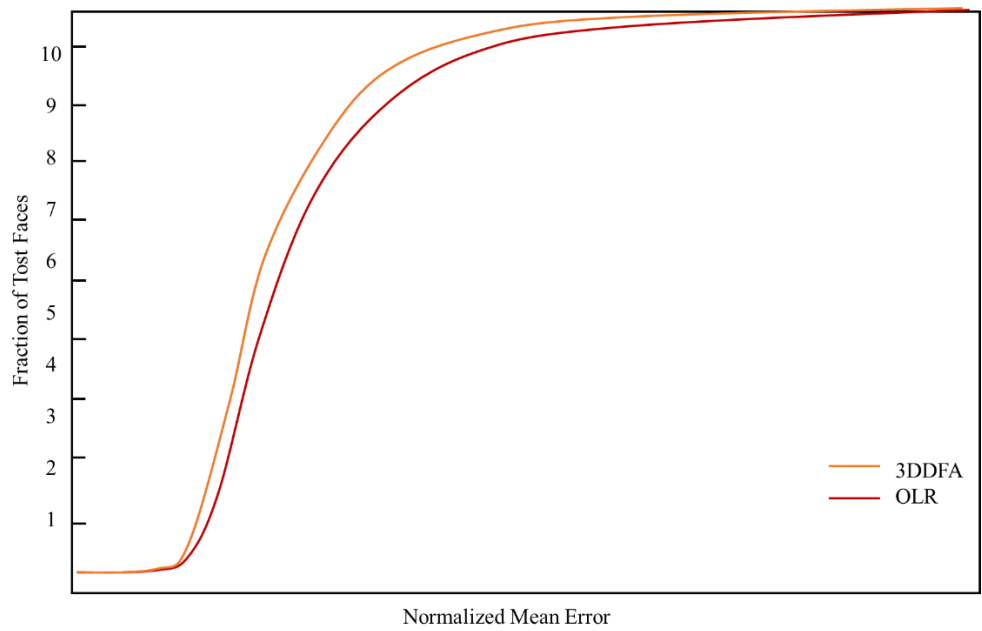


Figure 9 The CED curves of the 3D face alignment outcomes on the AFLw2000 dataset.

Full-size  DOI: 10.7717/peerjcs.1864/fig-9

CONCLUSION

A 3DFR employing single pictures makes it difficult to recognize faces when computer graphics and vision technologies are implemented. Face-positive and face-animation applications are of great significance and suffer from this issue. Recent studies have mainly focused on the CNN-based face regression models implementing the coefficient of 3DFDM to reconstruct 3D face and alignment tasks. However, reconstruction methods only consider a one-to-one reconstruction process, regardless of the correlation or differences in the input image. This leads to the model's lack of sensitivity to identify face information or deal with excessive sensitivity to face posture. In addition, the lack of a 3D face data set limits the performance of the mentioned methods that lack the robustness of natural face images.

Based on the problems encountered in the literature, the fuzzy temporal neural network (FLSTM) is proposed to process the sequence of images and improve the original 3DFR algorithm (3DFDM). The video image frame algorithm is used to realize the 3DFR algorithm. Experiments designate that the suggested method could lower the number of images needed even though it has a higher accuracy and presents more robustness. Moreover, it can not only be applied to the reconstruction of the 3-D face animation but also to the reconstruction of other 3D animation scenes, which has strong practical value.

Time-dependent data could be treated with other fuzzy time series methods available in the literature to search for better outcomes. Future research will concentrate on the implementation and comparison of those methods.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Ming Zhou conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code is available in the [Supplementary File](#).

The PK image is copied and edited from <https://en.photo-ac.com/photo/24455124/a-boy-appealing-for-environmental-problems>

One of the images used is available from <https://en.photo-ac.com/photo/24455124/a-boy-appealing-for-environmental-problems>

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.1864#supplemental-information>.

REFERENCES

- Becker H, Fregosi N. 2017.** The impact of animation deformity on quality of life in post-mastectomy reconstruction patients. *Aesthetic Surgery Journal* **37(5)**:531–536 DOI [10.1093/asj/sjw264](https://doi.org/10.1093/asj/sjw264).
- Blanz V, Vetter T. 1999.** A morphable model for the synthesis of 3D faces. In: *SIG-GRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*.
- Casas L, Fauconneau M, Kosek M, Mclister K, Mitchell K. 2019.** Enhanced shadow retargeting with light-source estimation using flat fresnel lenses. *Computers* **8(2)**:29 DOI [10.3390/computers8020029](https://doi.org/10.3390/computers8020029).
- Feng Y, Wu F, Shao X, Wang Y, Zhou X. 2018a.** Joint 3d face reconstruction and dense alignment with position map regression network. In: *Proceedings of the european conference on computer vision (ECCV)*. 534–551.
- Feng ZH, Huber P, Kittler J, Hancock P, Wu XJ, Zhao Q, et al, Räscht M. 2018b.** Evaluation of dense 3D reconstruction from 2D face images in the wild. In: *2018 13th IEEE International conference on automatic face & gesture recognition (FG 2018)*. Piscataway: IEEE, 780–786.
- Genova K, Cole F, Maschinot A, Sarna A, Vlastic D, Freeman WT. 2018.** Unsupervised Training for 3D Morphable Model Regression. In: *2018 IEEE/CVF conference on computer vision and pattern recognition*. Piscataway: IEEE, 8377–8386 DOI [10.1109/CVPR.2018.00874](https://doi.org/10.1109/CVPR.2018.00874).
- Guo Y, Cai J, Jiang B, Zheng J. 2018.** Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41(6)**:1294–1307.
- Hernandez M, Hassner T, Choi J, Medioni G. 2017.** Accurate 3D face reconstruction via prior constrained structure from motion. *Computers & Graphics* **66**:14–22 DOI [10.1016/j.cag.2017.05.008](https://doi.org/10.1016/j.cag.2017.05.008).
- Hochreiter S, Schmidhuber J. 1997.** Long short term memory. *Neural Competition* **9(8)**:1735–1780 DOI [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- Jin H, Wang X, Zhong Z, Hua J. 2017.** Robust 3D face modeling and reconstruction from frontal and side images. *Computer Aided Geometric Design* **50**:1–13 DOI [10.1016/j.cagd.2016.11.001](https://doi.org/10.1016/j.cagd.2016.11.001).
- Lee JD, Nguyen O, Lin Y-C, Luu D, Kim S, Amini A, Lee SJ. 2022.** Facial scanners in dentistry: an overview. *Prosthesis* **4**:664–678 DOI [10.3390/prosthesis4040053](https://doi.org/10.3390/prosthesis4040053).
- Liu F, Zhao Q, Liu X, Zeng D. 2018.** Joint face alignment and 3D face reconstruction with application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42(3)**:664–678.

- Lium O, Kwon YB, Danelakis A, Theoharis T. 2021.** Robust 3D face reconstruction using one/two facial images. *Journal of Imaging* 7:169 DOI [10.3390/jimaging7090169](https://doi.org/10.3390/jimaging7090169).
- Mai G, Cao K, Yuen PC, Jain AK. 2018.** On the reconstruction of face images from deep face templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41(5):1188–1202.
- Merras M, Hazzat SEL, Saaidi A, Satori K, Nazih AG. 2017.** 3D face reconstruction using images from cameras with varying parameters. *International Journal of Automation and Computing* 14:661–671 DOI [10.1007/s11633-016-0999-x](https://doi.org/10.1007/s11633-016-0999-x).
- Parke FI. 1982.** Parameterized models for facial animation. *IEEE Computer Graphics and Applications* 2(9):61–68 DOI [10.1109/MCG.1982.1674492](https://doi.org/10.1109/MCG.1982.1674492).
- Pattanayak RM, Sangameswar Vodnala D, Das H. 2022.** Fuzzy Time Series Forecasting Approach using LSTM Model. *Computación y Sistemas* 26(1):485–492 DOI [10.13053/CyS-26-1-4192](https://doi.org/10.13053/CyS-26-1-4192).
- Peng S, Tian JL. 2017.** Real-time virtual animation reconstruction algorithm based on the perception of random human behavior and collision prediction. *Journal of Communications* 12(2):111–117.
- Richardson E, Sela M, Kimmel R. 2016.** 3D face reconstruction by learning from synthetic data. In: *Proceedings, 2016 4th international conference on 3D vision, 3DV 2016*. Piscataway: IEEE, 460–467.
- Rui T, Zhang S, Zou J, Zhou Y, Tang J. 2018.** Deep auto-encoder based on supervised learning for damaged face reconstruction. In: *Internet multimedia computing and service: 9th international conference, ICIMCS 2017, Qingdao, China, August (2017) 23-25, revised selected papers 9*. Singapore: Springer, 290–299.
- Shirowzhan S, Tan W, Sepasgozar SM (eds.) 2020.** *Spatial big data, BIM and advanced GIS for smart transformation: city, infrastructure, and construction*. Basel: MDPI.
- Thakrar RM, Dickson J, Ahmed A, George J, Sandhu G, Janes SM, Navani N. 2017.** 149: virtual animation and 3-dimensional airway reconstruction using Lungpoint[®] to guide tracheobronchial stenting. *Lung Cancer* 1(103):S67–S68.
- Trần AT, Hassner T, Masi I, Paz E, Nirkin Y, Medioni G. 2018.** Extreme 3d face reconstruction: seeing through occlusions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 3935–3944.
- Tu X, Zhao J, Xie M, Jiang Z, Balamurugan A, Luo Y, Feng J, et al. 2020.** 3D face reconstruction from a single image assisted by 2D face images in the wild. *IEEE Transactions on Multimedia* 23:1160–1172.
- Tran AT, Hassner T, Masi I, Medioni G. 2017.** Regressing robust and discriminative 3D morphable models with a very deep neural network. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR)*. Piscataway: IEEE, 493–1502 DOI [10.1109/CVPR.2017.163](https://doi.org/10.1109/CVPR.2017.163).
- Tran L, Liu X. 2018.** Nonlinear 3D face morphable model. In: *2018 IEEE/CVF conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 2018*. Piscataway: IEEE, 7346–7355 DOI [10.1109/CVPR.2018.00767](https://doi.org/10.1109/CVPR.2018.00767).

- Turan C. 2017.** Robust face recognition via sparse reconstruction vector. In: *proceedings of the 2017 13th International conference on electronics, computer, and computation (ICECCO)*.
- Wang YL, Zhai JL. 2019.** 3D animation fusion technology based on block matching in color space. *Computer and Modernization* **290(10)**:112–116
DOI [10.3969/j.issn.1006-2475.2019.10.021](https://doi.org/10.3969/j.issn.1006-2475.2019.10.021).
- Waters K. 1987.** A muscle model for animating three-dimensional facial expression. *Computer Graphics* **21(4)**:17–24.
- Wu S, Xiao S, Di Y, Di C. 2021.** 3D film animation image acquisition and feature processing based on the latest virtual reconstruction technology. *Complexity* **2021**:1–11.
- Yueyang SU, Wan J, Junkaiyi A. 2019.** 3D face reconstruction based on a single frontal image and a statistical model. *Journal of Beijing University of Chemical Technology (Natural Science Edition)* **46(1)**:104–109 DOI [10.13543/j.bhxbzr.2019.01.016](https://doi.org/10.13543/j.bhxbzr.2019.01.016).
- Zhang J, Wu S, Li L, Lin K, Zheng X, Cao H. 2021.** Towards rich-detail 3D face reconstruction and dense alignment via multi-scale detail augmentation. In: *2021 IEEE International conference on multimedia and expo (ICME)*. Piscataway: IEEE, 1–6.
- Zhao L. 2001.** Synthesis and acquisition of Laban movement analysis qualitative parameters for communicative gestures. PhD thesis, UPENN, Philadelphia, PA, USA. Available at <https://repository.upenn.edu/server/api/core/bitstreams/65077895-dbf6-4920-badd-76b4a1f4646e/content>.
- Zhu X, Liu X, Lei Z, Li SZ. 2019.** Face alignment in full pose range: a 3D total solution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41(1)**:78–92
DOI [10.1109/TPAMI.2017.2778152](https://doi.org/10.1109/TPAMI.2017.2778152).