Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9999113, 11 pages
https://doi.org/10.1155/2021/9999113

Research Article

3D Virtual Animation Instant Network Communication System Design

Jing Liu,[1] Qixing Chen,[2] and Xiaoying Tian[1]

[1]*College of Culture and Art, Chengdu University of Information Engineering, Chengdu 610225, China*
[2]*College of Communication, Chengdu University of Information Engineering, Chengdu 610225, China*

Correspondence should be addressed to Xiaoying Tian; txy@cuit.edu.cn

Received 9 March 2021; Revised 15 June 2021; Accepted 18 June 2021; Published 1 July 2021

Academic Editor: Wei Wang

Copyright © 2021 Jing Liu 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.

This project uses Openfire to implement a virtual 3D animation instant messaging system, which is easier to use and more expandable. The main work of the client is to implement the Extensible Messaging and Presence Protocol (XMPP) and use XMPP to transmit data to the server side and receive data from the server side, while Openfire is built by the server side to use. To address the problem that the current mainstream face key point localization model is less robust to complex environments, this project adopts a deep learning-based approach to design and implement the face key point localization model, through data preprocessing, model design, and model training, to achieve a robust model that can locate 68 face key points and complete the migration of the model to mobile. The current video communication often suffers from delay and lag, so this project uses face key point data instead of video stream data transmission to reduce the pressure on the network. This topic also uses voice coding and decoding, noise reduction, echo cancellation, and other processing to solve the problems of noise interference and echo interference in voice transmission. This paper also introduces the creation, import, and loading of 3D virtual models, and explains how to use face key point association to drive 3D animation models, how to make the drive smoother and more natural, and using individual face key points as an example.

1. Introduction

Video communication in instant messaging systems usually requires high real time and stability; otherwise, it is prone to data delay, playback lag, and other instability. Due to the influence of an unstable network environment, the data are easily disturbed by various factors during transmission, resulting in the data not being broadcasted properly at the receiving end [1]. The goal of this topic is to design a virtual video chat system combined with virtual reality, which needs to be based on application scenarios, transforming from the original transmission of video data to the transmission of user's face key point data and handling transmission abnormalities [2]. At the same time, because this topic is not based on video streaming instant messaging, but 3D virtual animation video chat, the user sees the expression animation of the virtual animation model during the chat, so the data format transmitted in the network is the dataset of face key points and voice data, and this topic is based on the actual application scenario; there are high requirements for the noise reduction and echo cancellation of voice; the synchronization of voice and animated expressions and speech optimization become the urgent problems in this project [3, 4].

With the rapid development of computer network technology, the advantages of applying virtual reality technology in the field of instant communication have become greater and greater [5]. In a comprehensive and detailed exploration of the virtual reality technology applications that exist in computer communications, virtual reality technology plays an important role in communication networks, including enhancing resource utilization, device redundancy, immersion, interactivity, conceptualization, and holography [6]. Research of combining instant communication with virtual reality is still challenging due to the current problems of poor computing power, unstable data transmission, and large transmission delay on the mobile side. The ability to integrate communication and computing is used in the future 5G mobile communication network with multilevel

computing to solve the problem of the limited computing power of mobile terminals, efficient spectrum sensing, layered coding, and airport adaptive transmission to overcome the problem of unstable mobile channel transmission and delay guarantee mechanism to ensure the delay of mobile ARIVR services. However, research on the application of combining instant messaging with virtual reality technology is relatively routed and still in the exploration stage [7]. The system targets autistic patients and can measure patients' gaze-related index during their interaction with virtual companions, and this index can be mapped to their corresponding anxiety level. At the same time, the system can influence the patient's task performance and gaze-related index in response to the virtual companion's emotions [8]. Apple's release of Animoji enables the communication between users using 3D animated emesis. The implementation principle is to use facial recognition sensors to detect changes in user expressions and record user speech with a microphone to generate cartoonist 3D animated demos that can be shared by users with each other via iMessage [9]. The release of this feature has reaped a strong response from the community, stimulating interest and raising the interest of researchers in related fields. This topic will combine virtual reality and instant messaging, with the client driving 3D virtual models by parsing key point positioning data of faces to achieve virtual animated real-time communication [10].

In recent years, research in face key point localization has become increasingly abundant and mature, while research in deep learning has also made many breakthroughs, bringing better innovative approaches and more opportunities for other related research fields. The key point localization is the foundation of face recognition and other research, and the application scenarios are very broad [11]. As shown in Figure 1, the researchers have proposed many algorithms for face key point localization and achieved good results in related fields, but in practical applications, faces are often affected by various internal and external factors such as expression, posture, lighting, and occlusion, making it very difficult to achieve accurate face key point localization, which remains a great challenge [12]. This topic will address the design and optimization of the face key point positioning model and its application in mobile based on the actual application scenarios.

This topic is based on the XMPP to implement an instant messaging system combined with virtual reality. The system includes a client and a server. The client side realizes to build and maintain an instant communication system: based on the basic instant communication system, it locates the face key points in each frame of the video stream through the face key point positioning model, records, noise reduction, echo cancellation, encoding, and packaging the user's voice in real-time through the client microphone, and transmits the face key point positioning data and voice data to the friend user in real time. After receiving the face key point data, the 3D animation model is driven to enable the virtual model to make real-time accurate expressions, while decoding and playing the received voice data.

2. 3D Virtual Animation Key Point Model Design and Implementation

2.1. 3D Virtual Animation Is Raw Data Preprocessing. The design and implementation of the face key point localization model are developed using TensorFlow, and the model training is completed before exporting the file and porting it to the mobile application [13, 14]. The model design and implementation are divided into three parts: data preprocessing, model building, and model training. The main purpose of data preprocessing is to prepare data for model building, and the effect of modeling is greatly affected by good or bad data preparation; the model building is to research and analyze relevant algorithms, design a model that meets the subject matter, and make improvements in experimental analysis; model training is to learn the preprocessed data through relevant machine learning algorithms to get a model that can be used for predicting new data, which is a process of continuous iteration and exploration.

The data for this project is based on the 300 W challenge data, a well-known benchmark for evaluating key point detection algorithms. The challenge datum is a combination of image compilation from five datasets: LFPW, HELEN, AFW, IBUG, and the private 300 W test set. The last private 300 W dataset was originally used to evaluate eligibility for the competition and was then privately owned by the competition organizers, hence the name. Each image in the dataset has 68 markers with bounding boxes generated by the face detector. In this project, the 300 W challenge data are divided into a training set part and a test set part [15]. The training part includes the AFW dataset, the training subset of the LFPW dataset, and the training subset of the HELEN dataset, containing a total of 3148 images. The test section includes the remaining datasets: the IBUG dataset, the private 300 W test set, the best subset of the LFPW dataset, and the best subset of the HELEN dataset. In all datasets, each image corresponds to a text file with the coordinates of 68 key points of the face [16].

In the data preprocessing of this project, the dataset is firstly expanded: combined with the specific application scenario of this project, the dataset is flipped, panned, rotated, and scaled to improve the coupling between the dataset and the application scenario while increasing the number of the dataset. For each image in the dataset, we first flip the image left and right and then flip the data of 68 coordinate points corresponding to the image, and the dataset becomes twice the original one after the flip is completed. After the flip is completed, each image and its corresponding face key point coordinates are randomly panned, rotated, and scaled, and the final output is $112 * 112$. Since the original image is the input of the convolutional neural network in the model, the size needs to be equal, so here the image size is unified into $112 * 112$, and finally, each image is expanded into 20 images after the above operation. For the translation transformation, the translation is relatively simple, which is a simple coordinate addition and subtraction process:

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} = \begin{bmatrix} dx & dy & 1 \end{bmatrix}. \quad (1)$$

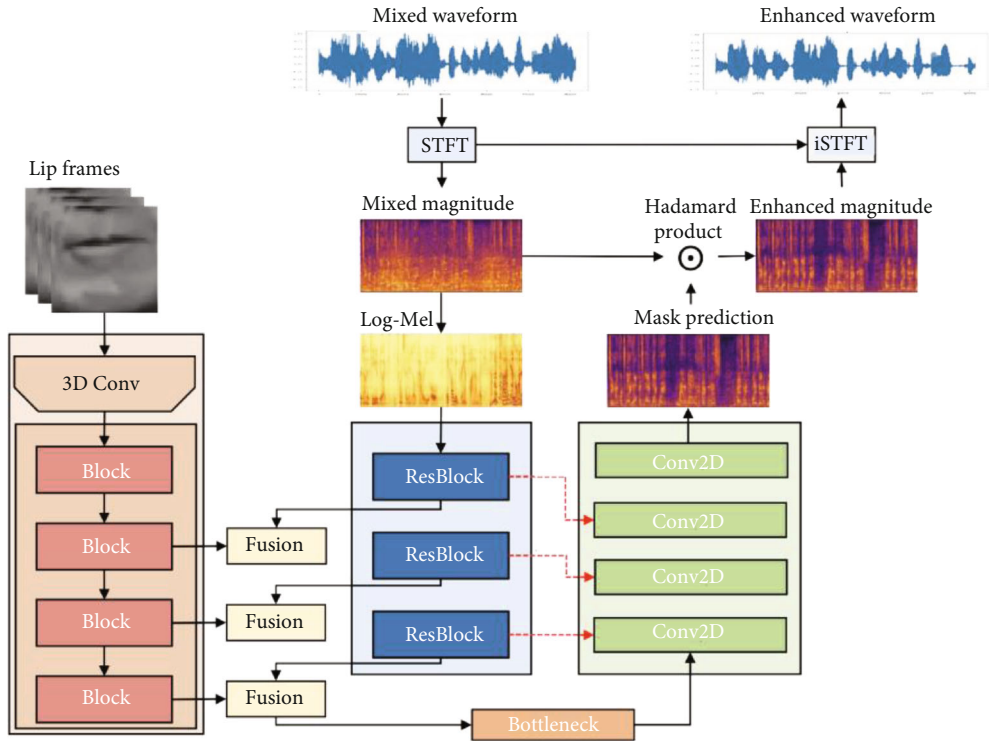

Figure 1: Framework diagram of combined virtual reality instant messaging system.

Calculation with the transformation matrix is as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \begin{bmatrix} dx \\ dy \\ 0 \end{bmatrix}^{\mathrm{T}} = x'dx + y'dy. \tag{2}$$

For the rotational transformation, the formula is calculated to obtain the following relation:

$$\begin{bmatrix} x' \\ 1 \\ y' \end{bmatrix} \begin{bmatrix} dx \\ 1 \\ dy \end{bmatrix} = \begin{bmatrix} \cos\theta & 1 & 0 \\ x & \tan\theta & 1 \\ \tan\theta & y & \sin\theta \end{bmatrix}. \tag{3}$$

For the scaling transformation, if the original coordinates are scaled directly, the center point of the image will be shifted.

$$\begin{bmatrix} x' \\ 0 \\ y' \end{bmatrix} = \begin{bmatrix} S_x & 1 & 0 \\ x & \tan\theta & dy \\ S_y & dx & \sin\theta \end{bmatrix} \begin{bmatrix} x \\ 1 \\ y \end{bmatrix}. \tag{4}$$

If you want to keep the center point unchanged, you can first translate the center point to the origin, then scale it, and then translate it back, i.e.,

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & 1 & 0 \\ x & 1 & dy \\ S_y & 0 & dy \end{bmatrix} \begin{bmatrix} x - x_0 \\ 1 \\ y - y_0 \end{bmatrix} \begin{bmatrix} x \\ 1 \\ y \end{bmatrix}. \tag{5}$$

2.2. 3D Virtual Animation Key Point Model Construction Design. The face key point localization model is based on a cascaded convolutional neural network, which mainly includes three parts (convolutional neural network, connection layer, and point coordinate similar inversion), and the original image is passed through two convolutional neural networks connected by the connection layer, and then, point coordinate similar inversion is performed to obtain the final 68 face key point coordinate results.

As shown in Figure 2, the convolutional neural network used in this model includes a total of four convolutional pooling operations and two fully connected layers; each convolutional pooling operation contains two convolutional layers and one pooling layer, the pooling layer uses the maximum pooling method, and the fully connected layer uses ReLU as the activation function. The key point coordinates initialized by the key point offset are summed up after the original image passes through the convolutional neural network to obtain the key point coordinates of the face in the first layer of the network.

The main purpose of the connection layer is to similarly transform the original image and its corresponding face key point coordinates, so that the image head pose changes to the standard position to reduce the impact of the different image head poses on the model; meanwhile, the connection layer generates the key point heat map according to the face

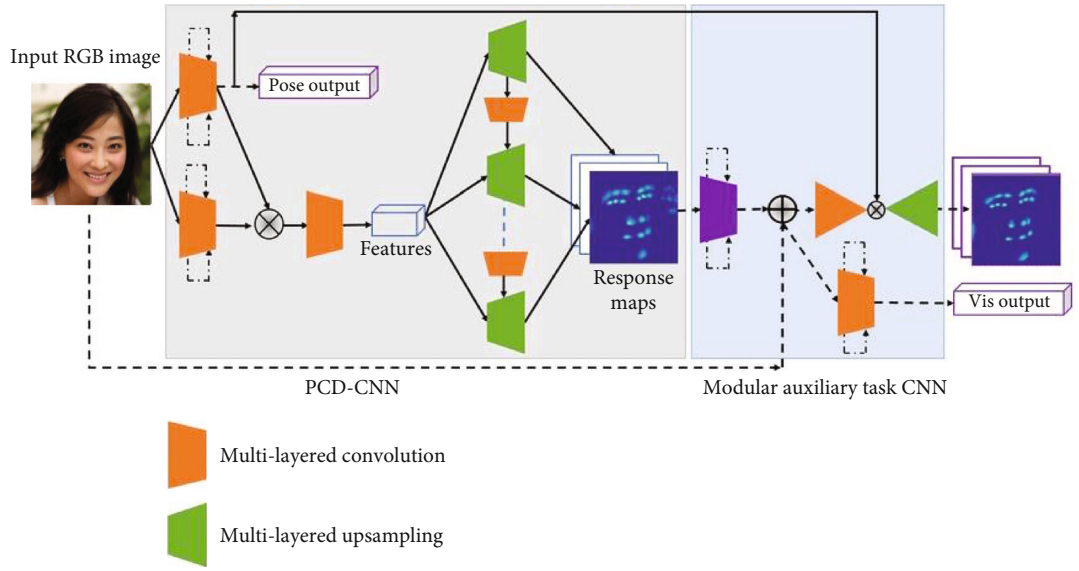

Figure 2: Block diagram of face key point localization model.

key point coordinates as the input of the second layer of the convolutional neural network; the key point heat map can better provide the face key point information and reduce the connection layer which mainly consists of four parts: similar transformation parameter calculation, image similar transformation, heat map generation, and feature map generation [17]. Among them, the similar transformation parameter calculation part calculates the parameters to similarly transform the image and key points to the standard position based on the initialized key point coordinates, the output value of the first layer, and the formula introduced in the data preprocessing and obtains the parameters of the similar transformation inverse transformation [18–20]. The heat map generation section firstly transforms the key point coordinates of the face by similar transformation and then generates the key point heat map according to the transformed key point coordinates as the second input of the second layer network. The feature map generation section takes the output of the first fully connected layer of the convolutional neural network as the input and passes through a fully connected layer to obtain the feature map as the third input of the second layer network, which is used to add feature information to the heat map.

The three outputs of the connection layer are used as the input of the second layer network, and after the second layer convolutional neural network, the key point offset of the second layer network is obtained, and the result of the similar transformation with the key point coordinate result of the first layer is added to obtain the face key point coordinate, which is the result of the similar transformation; therefore, it is necessary to use the inverse transformation parameters obtained from the connection layer to perform the similar inverse transformation of the point coordinates to obtain the result of the second layer network, which is the final face key point coordinate result.

For model training, the two layers of networks are trained separately, and after the training of the first layer is completed, the training of the second layer of networks is continued. The error rate is calculated as the Euclidean distance between the model prediction result and the real 68 key points divided by the distance between the two pupils of the face of that image [21]. The training results are shown in Figure 3. The error rate of the first layer network is higher and harder to converge because the first layer network only contains a convolutional neural network, and the separate convolutional neural network is difficult to achieve better results for the face key point localization problem, and the input of this model is the whole picture, which increases the difficulty of key point localization; therefore, the first layer network is less effective. The results of the second layer network training show that the second layer network converges faster and has a lower error rate due to the introduction of the face heat map, which again verifies the effectiveness of the cascade structure in the model and the advantage of the face heat map.

2.3. 3D Virtual Animation Is a Basic Instant Module Design. Functional requirements are the most basic conditions that a system can meet, but a system designed only based on functional requirements usually cannot be used directly by users, because in the process of use, users usually require the system to meet certain performance standards. The unit learning theme is the structured part of "association and structure". The goal of unit learning is to grasp the essence of knowledge. The unit learning activity is a part of activity and experience, transfer, and application. Therefore, the theme of unit learning is to move from "knowledge units" to "learning units", based on students' learning and development, and to organize the "learning" units in a big conceptual way. The logic of the subject reflects a rich, three-dimensional activity and openness. In the past, subjects were usually closed, but now we want to turn them into something open and unfinished, with unfinishedness and openness, providing space for students to explore and rediscover. We randomized the data according to an 8 : 2 distribution. Performance requirements, that is, nonfunctional requirements, refer to the design of the

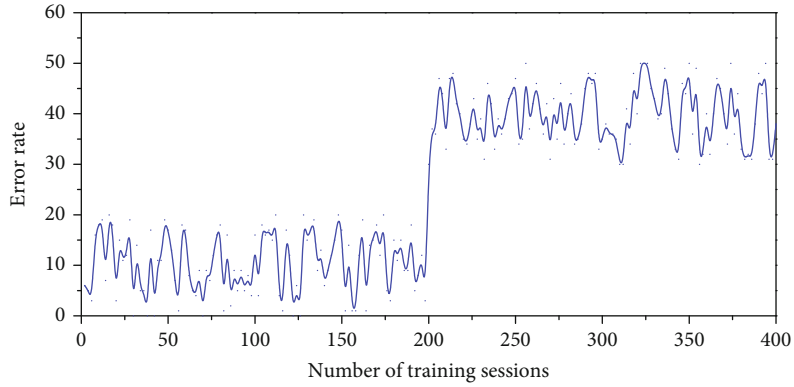

Figure 3: 3D virtual animation key point model training results.

system should meet some principles and functional minimum restrictions, performance indicators to be greater than these minimum requirements, and constraints to meet the needs of users. The good or bad performance of a system determines the user experience of this system, that is, the success or otherwise, so we should consider all aspects of the system's performance as much as possible during the preparation of the system. For performance requirements, the system considers a total of five major aspects, which are the simplicity of the system, the response time of the system, the compatibility of the system, the maintainability and expandability of the system, and the ease of use of the system.

For the basic instant messaging module, it mainly needs to implement a user system, buddy system, and communication system, specifically user registration, user login, buddy management, and real-time communication with the client. For user registration, the client fills in the user's name information and password information and transmits them to the server side for processing. The server side processes the registration request, generates a unique identifier on the server side upon success, and writes this user information to the database. When a user logs in, username and password information are transmitted to the server side for processing, and the server side processes the log-in request. If the user account or password is entered incorrectly, the prompt "Account or password error" is returned; otherwise, the user is redirected to the client communication homepage.

Friendly management mainly includes adding friends and friend lists [22]. After the user logs in successfully, clicking View Buddy List will request the buddy data from the server, and when it succeeds, all the buddies will be returned, and the client will display them in the form of a list. When you click Add Buddy, the requested data is saved in the database, and when the other party accepts the request, the server saves the buddy relationship and updates the data in the client's buddy list. In real-time communication, the client clicks on a buddy for real-time communication and sends a message, which is transmitted to the other buddy by the server in real time, and similarly, if a buddy message is received, the client receives it in real time and displays it in the conversation interface [23]. The client saves each chat log to the local database and loads the chat log every time it opens a conversation with a friend. If a friend sends a message to this client when this client is not online, the message data is saved in the server's offline message database and the offline message is loaded immediately when this client logs in. The design of the basic instant messaging module is shown in Figure 4.

With the sound module, it mainly needs to realize voice recording, voice coding and decoding, and sound playback. For voice recording, the client collects the voice data in real time through the microphone and decomposes the voice data into multiple voice packets of the same size according to the specific size. Since the actual application scenario of this topic requires simultaneous recording and playback of voice and the use of voice playback, it generates large noise interference and echoes interference, so the recorded voice needs to be coded noise reduction, echo cancellation, and other processing. In the voice coding and decoding, for the recorded voice, each recorded voice packet needs to be coded and rewritten, then saved as a file and transmitted; for the received voice, each received voice packet needs to be decoded and played back. When the voice is played, the voice packets need to be cached and processed, and the voice needs to be played in order, and at the same time, the synchronization process with the animation model driver needs to be performed when the voice is played.

3. 3D Virtual Animation Instant Network Communication System Design and Test Analysis

3.1. Instant Messaging System Design and Development. The main process of user registration is first to initialize XMPPStream, then connect to the server, and then send password authorization after connecting to the service successfully. In this project, the client's registration interface is designed with four input boxes for entering a username, nickname, password, and password confirmation, and two button controls for canceling registration and initiating a registration request. To protect the sensitive data of passwords, the password and password confirmation input boxes are controlled in the interface in a way that can be hidden.

The main process of user login is first to initialize XMPPStream, then connect to the server, and after a

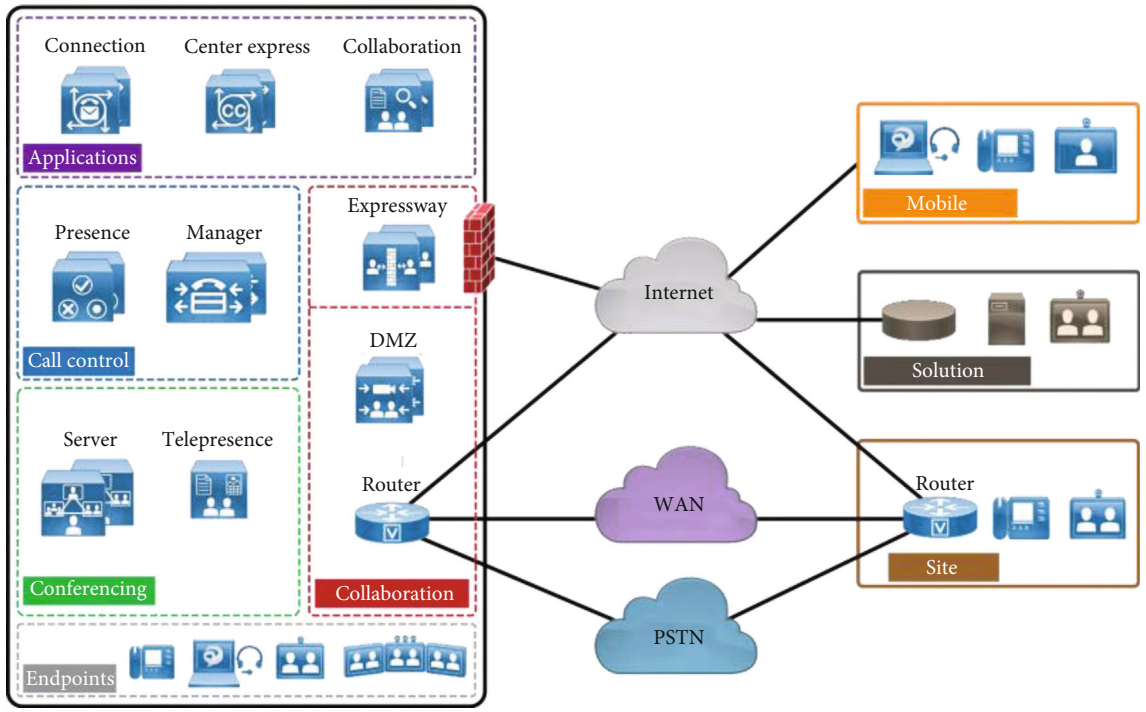

Figure 4: Basic instant messaging module flow chart.

successful connection to the service, send the password authorization, and after successful authorization, send the "online" message. After the connection is successfully established, the client will request offline message data and buddy data from the server, and then jump to the chat history interface and buddy list interface. In this project, the client's buddy management interface adopts the interface design like the buddy list, click the expand button in the upper right corner of the interface to select the add buddy menu, click the pop-up input box to add buddy, enter the user's name, and click the confirmation to send the request to the server to add buddy. The middle part of the friendly interface is designed with a fragment, which contains a list control that sorts the list of friends by their first alphabetical order, and implements a quick search by clicking on the letters on the right side. When you click on a friend in the buddy list, a pop-up window appears to indicate whether to initiate a chat, and when you click to confirm, you jump to the real-time communication interface with that friend, and when you click to cancel, the pop-up window disappears [24]. At the bottom of the friends, the screen is two buttons that can be clicked to switch between the friend's screen and the chat screen.

The main process of real-time communication is that client which sends messages to the server, which forwards messages to client B. At the same time, client B can also reply to messages to the server, which are forwarded to client A. When client B is not online, the messages sent by client A to client B are temporarily saved in the offline message database of the server, and then when the client blogs in, it will read the offline data from the server when the client blogs in. The chat data loading does not go to the server side to read the data in real time but synchronizes the offline data to the client local after each successful user login (each time synchronizing the data that is not available locally), and then, the interface layer of the client single goes to the local (SQLite) to read the data. As in Figure 5, in an instant messaging system with hand-free enabled, the microphone captures the echo of the other parts played by the amplifier, and the remote user receives it and hears his voice, creating an echo effect. The echo canceling can eliminate the echo before it is transmitted back to the other party, thus improving the quality of the sound received by the other user.

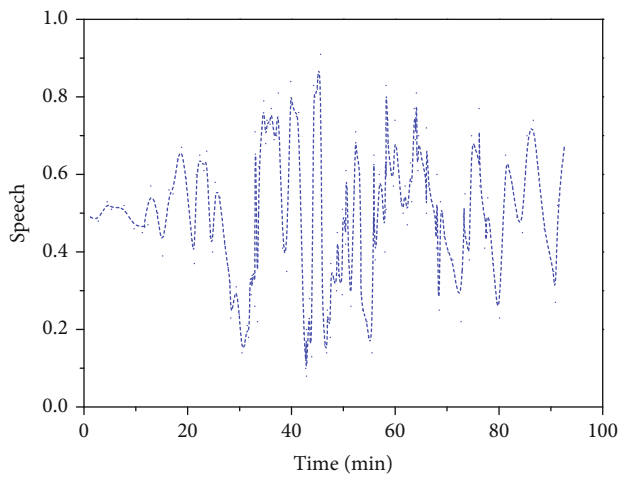

Figure 5: Effect of echo canceled on speech quality.

3.2. Design and Analysis of Facial Key Point Capture and Processing Functions. This project uses FaceTracker, an OpenCV-based facial tracking library. Firstly, FaceTracker needs to be integrated into the project and then tuned

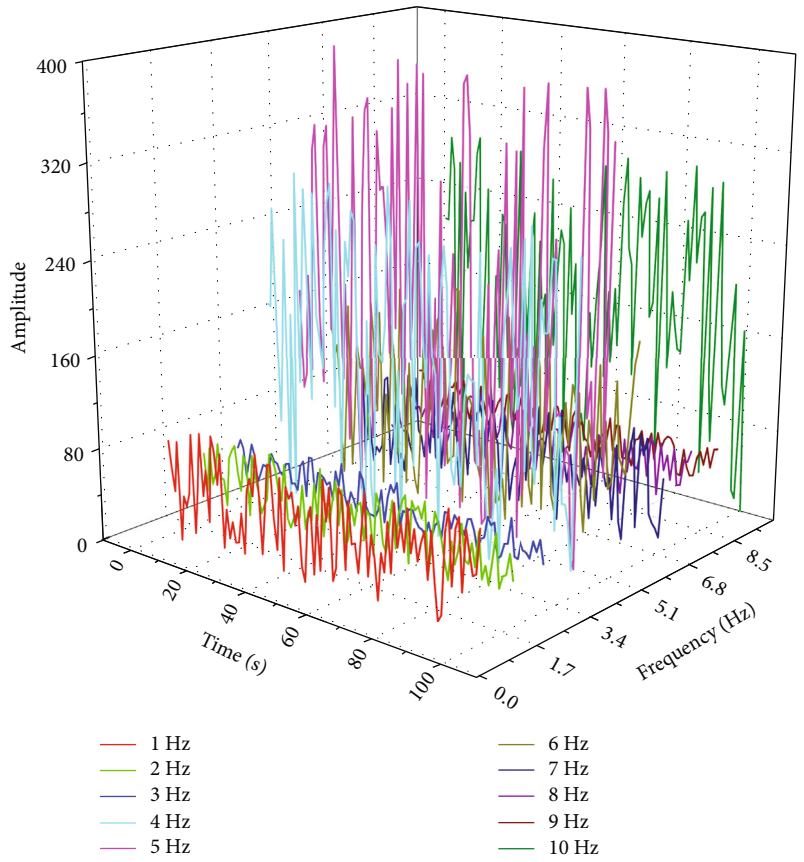

Figure 6: Face positioning timing diagram.

according to the actual situation. After the key facial recognition, a set of data is thrown every other frame, and the structure of the thrown data is shown in the key facial data structure table. After receiving the fatal key point data, it is forwarded to the key facial smoothing module for smoothing. The SG smoothing filter has a good smoothing and denoting effect on two-dimensional points, combined with the advantages of SG smoothing filter and the application scenario of the real-time nature of this topic; the five-point three-time smoothing algorithm of SG filter is used for smoothing, taking five nearby points, and determining a three-time curve, so that the square sum of the difference between the points on the curve and the original points' vertical coordinates is minimized. The coefficients of the cubic curve are determined, and by continuously adjusting the coefficients of the cubic curve, the curve is made smoother and closer to the actual situation.

In this project, different gaps are set for different facial key points to ensure that the movements of the key facial driving the model are effective, to achieve the role of dejittering [25]. The 5 key points show the threshold values corresponding to the facial key points, and the thresholds need to be constant of the adjustment by experiment, due to the large range of motion of the corner of the mouth, so the threshold is larger; taking the left corner of the mouth as an example, the threshold is 0.2. As shown in Figure 6, this paper compares whether the range of activity of a certain frame is greater than 0.2; the displacement is valid and enters the driving module; if not, this displacement is regarded as an invalid displacement. The movement coefficient is required to be adjusted continuously during the driving process of facial key points, and it plays the role of adjustment and optimization in the feature extraction and rereduction algorithm. After the blendShape is calculated by the feature extraction and rereduction algorithm, the driving effect is optimized by continuously adjusting the movement coefficient to drive the 3D virtual model accurately.

3.3. Design and Test Analysis of Driven 3D Virtual Animation Models. In this paper, we calculate the current degree of mouth opening of the user. Then, depending on the size of the user's open mouth, the maximum degree of mouth opening is multiplied by the movement parameter. This weight is assigned to blendShape, and we successfully control the opening of the character. Similarly, there are key points around the mouth, the left eyebrow in the corner of the mouth, the middle of the eyebrow, the opening and closing of the eyes, and the sliding calculation, each with a different calculation algorithm. This algorithm traverses all resources in the resource package. If the type of resource is a game object type, then the bone is instantiated into the scene. The blendShape is a masked mesh rendering property attached to the bone; so when the bone is loaded, it is also loaded. Then, iterating through all the bones and taking the blendShape name can be easily invoked and a face model is loaded into the scene.

Table 1: Virtual portrait key point expression reproduction degree.

Frame rate	Corner of the mouth	Corner of the eye	Eyes	Eyebrows
1	92.47	38.66	17.63	17.54
2	84.43	37.83	16.09	16.52
3	76.35	39.21	15.54	17.68
4	85.33	42.01	17.35	17.89

The encoding part gets data from the sound module and the expression recognition module. Then, key facial packets and voice packets are packaged into the basic data structure synchronously through the key facial queue and sound queue in the expression sound synchronization module. Since the communication of this topic is based on XMPP, which does not support the direct transmission of packet structures and needs to be converted into string format first, this part mainly implements this serialization function. The first thing to consider is to separate the fields into space and write them as strings. As shown in Table 1, the approach is feasible, but the compression rate is too low because the characters used are too low compared to all character sets, and if floating-point numbers are used to express key points, only the numeric part of the string can be used, and the alphabetic part is wasted and the effective utilization is too low.

To improve the compression ratio, we use the following method: First, the grouped sequences are converted into binary data, and the binary strings are converted into strings. This topic is implemented using the Base64 algorithm. Base64 algorithm is one of the most common encoding methods to convert arbitrary invisible binary data into visible text data. Usually, the algorithm is used to transmit long identification information, such as URLs. The input to the Base64 encoding algorithm is a stream of 8-bit bytes, i.e., byte arrays and strings. The advantage of the Base64 algorithm is that the encoding method is simple and easy to understand and the compression ratio is greatly improved compared to the direct conversion method of numeric strings. Its disadvantage is that the encoded strings are not readable, which can cause a certain degree of inconvenience during debugging.

In the process of open-mouth check and shut-mouth check, the user faces the camera and the facial recognition module locates the key to the face by clicking on OpenMouthCheck, the red dot in the upper left corner of the picture. After positioning, the user opens his mouth wide and checks it through the button. Through this check, it is equivalent to preprocessing the user's face. As shown in Figure 7, the parameters of each part are from the beginning of the facial key point identification and positioning, then to the smoothing module to denoise, smoothing, and then to the drive module of blendShape smoothing and feature extraction and then restore algorithm drive, etc.; the 3D virtual model can restore the user's expression more accurately.

3.4. 3D Virtual Animation Instant Network Communication System Performance Test. Functional testing refers to the testing of a product's features and operable behaviors to determine that they meet the design requirements based on

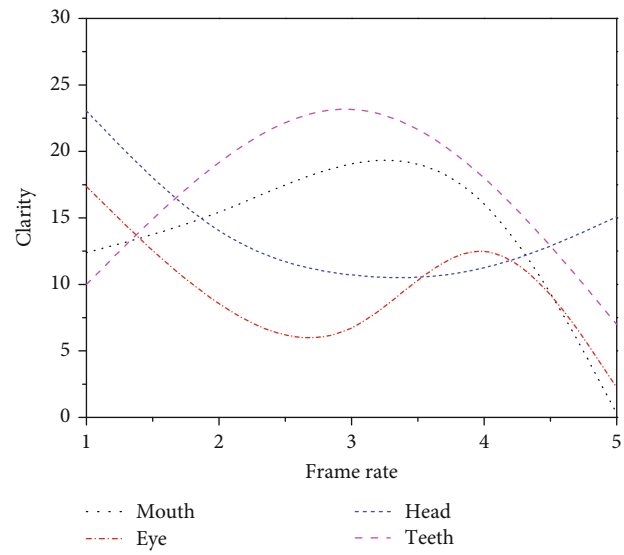

Figure 7: Comparison of facial features before and after virtual animation processing.

product features, operational descriptions, and user scenarios. The ultimate goal of functional testing is to ensure that the program operates in the desired manner, so the software should be tested according to functional requirements, by testing all features and functions of a system to ensure compliance with requirements and specifications. Functional testing can be divided into manual testing and automated testing, and this system uses a manual approach to test the main functions of the system. Since the system is divided into five modules (user management, friend management, instant messaging management, video management, and file storage management, and among them), the file storage management module can be tested by other modules.

Performance testing is the testing of performance indicators to improve the user experience when the functional requirements of the system have been met. The purpose of performance testing is to verify whether the software system can meet the performance indicators proposed by the user and to discover the performance bottlenecks in the software system, optimize the software, and finally play the purpose of optimizing the system. Performance testing mainly includes compatibility, security, stress testing, and load testing.

Security testing refers to the software system in the late stages of development to undergo rigorous security testing to meet the security needs of users using the process. Only products that pass the test can be put online and run. Security testing is to protect the system and users in case of malicious

Table 2: Priority queue for communication network packet processing.

ID	Sending time (s)	Acceptance time (s)	Play time (s)	Loss of condition	Hold time (s)
265	10.45	15.32	10.65	No	0
266	10.66	11.32	11.32	No	0
267	11.32	12.69	5.67	No	0.22
268	11.21	13.42	5.45	No	0.12

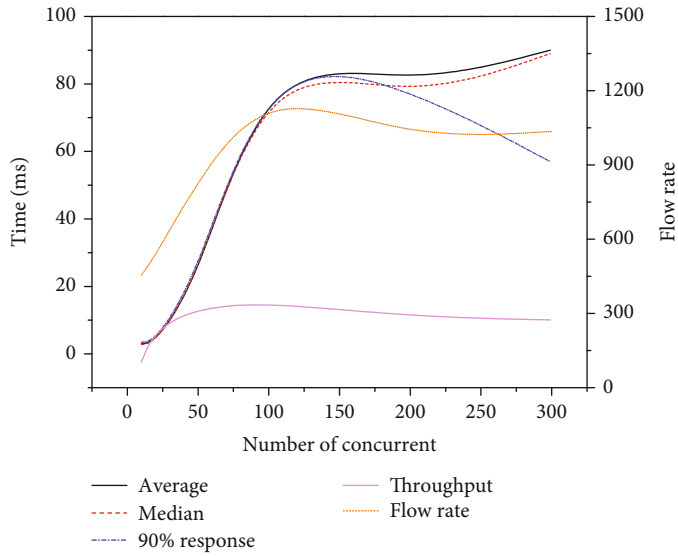

Figure 8: Stress test results of instant network communication system.

and wrong input. Network information security includes two layers of meaning: one is the security of external data exchange and the other is the security of internal LAN.

Web itself has good security monitoring measures; compared to traditional web real-time communication, users need to install plug-ins or applications in the browser; the process itself has potential risks, such as plug-ins and applications with their own Trojan horse virus, video transmission when the unscripted video stream is intercepted, or plug-ins and applications themselves which have the function of extracting user information, while the emergence of WebRTC shield exists. As shown in Table 2, in the video transmission process, the web uses secure communication protocols such as DTLS and SRTP to provide data security to both sides of the peer-to-peer video, which can prevent the leakage of video data on the web and video encryption at the sender and receiver side to encrypt video data; the key is negotiated by the video; both sides negotiate. The overall web components are encrypted, especially the codec, audio, and video transmission module; eliminating the video transmission process was intercepted.

As shown in Figure 8, the results of the stress test show that the average response time of the system can be kept within 100 ms, which meets the performance requirements, while when the number of concurrencies is high, a sudden increase in error rate can be found, so the current number of concurrency that the system can handle is around 200, while the throughput is low compared to large websites. Overall, the system can meet the current user requirements and can provide stable instant messaging and video communication.

The instant messaging management function includes seven main functions: text message, picture message, expression message, voice message, video clip, history message management, and text color change. The seven functions are highly independent, so test cases are designed for each of the seven functions to meet the testing requirements of the instant messaging management module. The instant messaging management module is the main module of the system, and the communication methods such as text, picture, expression, voice, and small video realized by the module greatly facilitate the users to communicate in real time, and each function is crossed and combined in the process of use, which can express the users' wishes completely and clearly and meet their needs. At the same time in the public chat room and private chat process, instant messaging functions are reused, reducing the workload of testing. Our method improves the accuracy by about 10.5% and the efficiency by about 14% compared to other studies.

4. Conclusion

This project combines the development status of cell phone operating systems and instant messaging software, based on the results of end-to-end instant messaging system requirements analysis, and combined with virtual reality technology,

basically realizes an instant messaging system based on XMPP for virtual video chatting, using 3D virtual animation models instead of real faces in the chat, and expression changes are reflected in the model animation, making the chat interaction more interesting and it also makes the transmission process lighter, reduces the burden on the server, and improves the fluency. In the model-driven, the subject first conducts in-depth research on algorithms related to face key point localization, compares several classical algorithms with better results, and implements a face key point localization model based on deep learning, and in the process of model implementation, through in-depth analysis of the actual application scenarios of the subject, proposes and implements improvements to the model, including optimization of data preprocessing and improvement of model algorithms, to improve the robustness of the model is improved. After the training of the model is completed, the migration of the model to run on mobile applications is realized, and each frame of the face video stream captured by the camera is processed in real-time in the subject project to output the relevant face key point localization data, thus driving the subject to further drive the 3D animation model. After acquiring face key point localization data, the system performs denoting and smoothing on the data, drives the 3D animation model by blendShape weight, and continuously tunes the parameters to make the model perform more smoothly and naturally. In this system, real-time voice calls are implemented for aiding virtual video chat. In voice processing, due to the specificity of the subject, the phone needs to open the external playback function for voice recording and playback, so the voice needs to be related to noise reduction, echo cancellation, and codec processing; the subject uses the third-party speed library to realize the voice noise reduction processing, echo cancellation, and voice codec. In the problem of voice delay, by constantly balancing the transmission time and voice recording time as well as the smoothness of playback, the voice delay is minimized. In the future, we will optimize and improve the model based on this.

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 known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Ministry of Education Industry University Cooperation Collaborative Education Project: Research on innovation and entrepreneurship education of digital art design major based on competition project construction (the second batch in 2020) and the Ministry of Education Industry University Cooperation Collaborative Education Project: Research on the construction of online advertising course in the new media environment (the second batch in 2020).

References

[1] M. G. Goodman, M. Alvarez, and S. B. Halstead, "Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: an historical perspective and role of antibody-dependent enhancement of infection," *Archives of Virology*, vol. 158, no. 7, pp. 1445–1459, 2013.

[2] P. Jackson and M. T. Raiji, "Evaluation and management of intestinal obstruction," *American Family Physician*, vol. 83, no. 2, pp. 159–165, 2011.

[3] J. Park and Y. Cho, "Design and implementation of automated steganography image-detection system for the KakaoTalk instant messenger," *Computers*, vol. 9, no. 4, pp. 103–110, 2020.

[4] X. Liu, T. Zhang, N. Hu, P. Zhang, and Y. Zhang, "The method of Internet of Things access and network communication based on MQTT," *Computer Communications*, vol. 153, pp. 169–176, 2020.

[5] K. Morris, O. Sugiyama, G. Yamamoto et al., "Towards a medical oriented social network service: analysis of instant messaging communication among emergency physicians," *Advanced Biomedical Engineering*, vol. 9, pp. 35–42, 2020.

[6] F. H. Chen and S. Y. Yang, "A balance interface design and instant image-based traffic assistant agent based on GPS and linked open data technology," *Symmetry*, vol. 12, no. 1, pp. 1–10, 2020.

[7] T. Koonen, K. A. Mekonnen, F. Huijskens, N. Q. Pham, Z. Cao, and E. Tangdiongga, "Fully passive user localization for beam-steered high-capacity optical wireless communication system," *Journal of Lightwave Technology*, vol. 38, no. 10, pp. 2842–2848, 2020.

[8] F. Lamberti, G. Paravati, V. Gatteschi, A. Cannavo, and P. Montuschi, "Virtual character animation based on affordable motion capture and reconfigurable tangible interfaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 5, pp. 1742–1755, 2018.

[9] Y. Qiu, J. Xie, H. Lv et al., "FULL-KV: flexible and ultra-low-latency in-memory key-value store system design on CPU-FPGA," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 8, 2020.

[10] G. Liu, Y. Huang, N. Li et al., "Vision, requirements and network architecture of 6G mobile network beyond 2030," *China Communications*, vol. 17, no. 9, pp. 92–104, 2020.

[11] C. M. Chen, M. C. Li, and Y. L. Huang, "Developing an instant semantic analysis and feedback system to facilitate learning performance of online discussion," *Interactive Learning Environments*, vol. 3, no. 5, pp. 1–19, 2020.

[12] F. Li, X. Wang, Z. Wang et al., "A local communication system over Wi-Fi direct: implementation and performance evaluation," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5140–5158, 2020.

[13] H. Wang, Y. Wang, G. Zhuang, and J. Lu, "Asynchronous passive dynamic event-triggered controller design for singular Markov jump systems with general transition rates under stochastic cyber-attacks," *IET Control Theory & Applications*, vol. 14, no. 16, pp. 2291–2302, 2020.

[14] H. Song, J. Thiagarajan, P. Sattigeri, and A. Spanias, "Optimizing kernel machines using deep learning," *IEEE Transactions*

on Neural Networks and Learning Systems, vol. 29, no. 11, pp. 5528–5540, 2018.

[15] S. De, L. Bruzzone, A. Bhattacharya, F. Bovolo, and S. Chaudhuri, "A novel technique based on deep learning and a synthetic target database for classification of urban areas in PolSAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 1, pp. 154–170, 2017.

[16] L. Jiang, L. Yan, Y. Xia, Q. Guo, M. Fu, and K. Lu, "Asynchronous multirate multisensor data fusion over unreliable measurements with correlated noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 5, pp. 2427–2437, 2017.

[17] H. Wu, Z. Zhang, C. Jiao, C. Li, and T. Q. S. Quek, "Learn to sense: a meta-learning-based sensing and fusion framework for wireless sensor networks," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8215–8227, 2019.

[18] H. Zhang, X. Zhou, Z. Wang, H. Yan, and J. Sun, "Adaptive consensus-based distributed target tracking with dynamic cluster in sensor networks," *IEEE Transactions on Cybernetics*, vol. 49, no. 5, pp. 1580–1591, 2019.

[19] X. Yuan and Y. Pu, "Parallel lensless compressive imaging via deep convolutional neural networks," *Optics Express*, vol. 26, no. 2, pp. 1962–1977, 2018.

[20] D. Nada, M. Bousbia-Salah, and M. Bettayeb, "Multi-sensor data fusion for wheelchair position estimation with unscented Kalman filter," *International Journal of Automation and Computing*, vol. 15, no. 2, pp. 207–217, 2018.

[21] V. Rahu, C. Tong, S. Bhattacharya et al., "Multimodal deep learning for activity and context recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–27, 2018.

[22] Q. Zhou and Y. Zheng, "Long link wireless sensor routing optimization based on improved adaptive ant colony algorithm," *International Journal of Wireless Information Networks*, vol. 27, no. 2, pp. 241–252, 2020.

[23] J. Hülsmann, J. Traub, and V. Markl, "Demand-based sensor data gathering with multi-query optimization," *Proceedings of the VLDB Endowment*, vol. 13, no. 12, pp. 2801–2804, 2020.

[24] P. Ghamisi, R. Gloaguen, P. M. Atkinson et al., "Multisource and multitemporal data fusion in remote sensing: a comprehensive review of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 1, pp. 6–39, 2019.

[25] H. A. Pierson and M. S. Gashler, "Deep learning in robotics: a review of recent research," *Advanced Robotics*, vol. 31, no. 16, pp. 821–835, 2017.