Chapter 6
Applications of Real-time Cartoon Face Animation to Multi-role Avatar Broadcasting and Web TV through Networks

6.1 Introduction

Nowadays, networked multimedia services prevail in the world. Many applications of networked multimedia services are developed, such as video telephones, teleconferencing, videos on demand, video chat-rooms, distance learning, and web TVs, etc. Therefore, talking cartoon faces can be used on networks. It is possible to use them as tutors to help students study on networks. And they also can be used as broadcasters to announce messages through networks. A real-time cartoon face animation system for use on networks is proposed in this study and described in this chapter. It is separated into the server subsystem and the client subsystem based on the real-time cartoon face animation system described in Chapter 5.

In Section 6.2, a server and client system for animation broadcasting is described. In Section 6.3, a method for synchronization of cartoon videos and speeches in this system is described. In Section 6.4, an application to multi-role avatar broadcasting through networks is described. In Section 6.5, an application to Web TVs by the ActiveX technique is described.

6.2 Server and Client System for Animation Broadcasting

6.2.1 System Organization

For networked applications, a server and client system for real-time talking cartoon face generation is proposed. The system contains two subsystems: a server subsystem and a client subsystem. By the interaction of these two subsystems, the video and audio data in the server can be transmitted to the client site in real time.

The server subsystem includes four major parts, an environment regulator, a facial feature tracker, a sound recorder, and a data transmitter. The environment regulator, the facial feature tracker, and the sound recorder are the components of the real-time cartoon face animation system described in Chapter 5. They are used to generate the image feature points and the speeches. The data transmitter is used to transmit the image feature points and the speeches to the remote client site through networks. The image feature points are transmitted when the facial feature tracking is completed in each frame, and the speeches are transmitted when the audio buffer is full.

The client subsystem includes two major parts: a data acceptor and an animation generator. The data acceptor is used to access the image feature points and the speeches from networks. The audio output device plays the sound when the speeches are accessed. When the image feature points are accessed, the animation generator generates the talking cartoon face in real time with a face model chosen at the client site. A configuration of this system is shown in Figure 6.1.

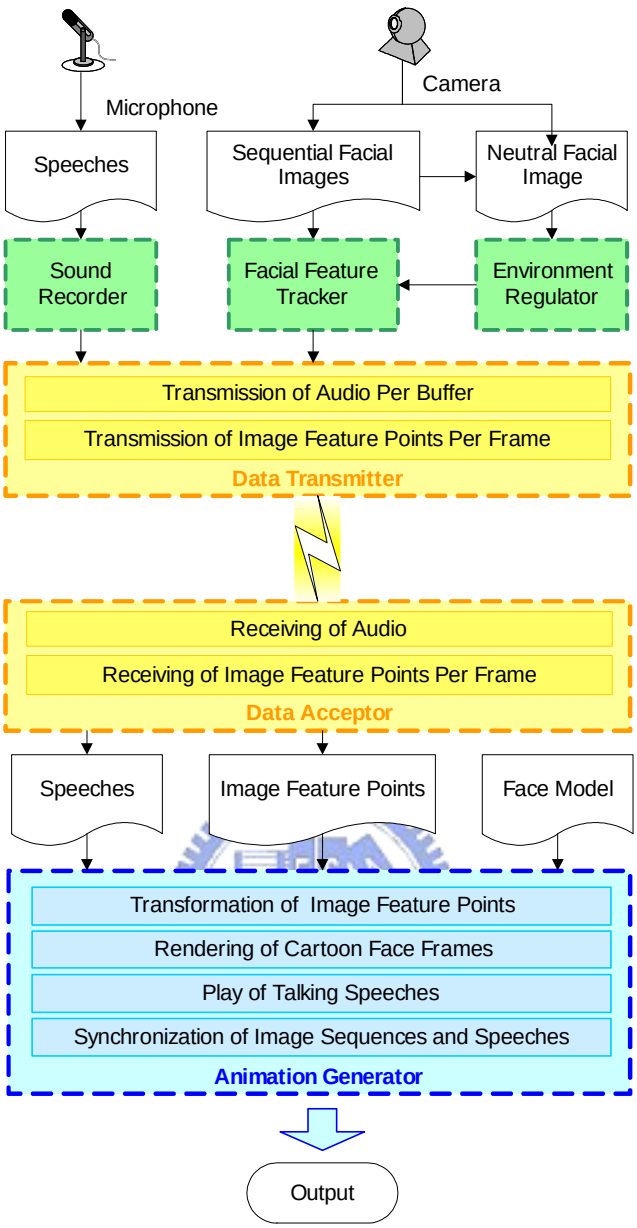
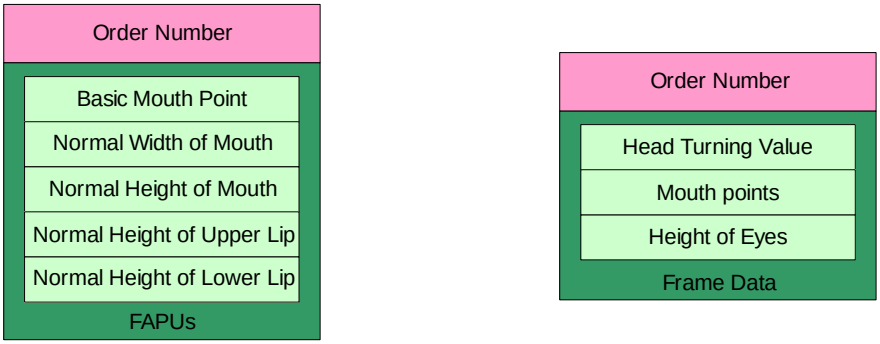


Figure 6.1 A configuration of the real-time cartoon face animation system.

6.2.2 Video and Audio Transmission

A detailed process for video and audio transmission in networks is described in this section. In the video transmission, the data of image features in each frame is packed as a *video unit*. Each video unit has an order number. There are two kinds of video units in the transmission. The first one is a *header unit*. The header unit contains the FAPUs extracted from the neutral facial image of image sequence. They are used

to help the client subsystem transform the image feature points into the face model control points. The second one is a *frame unit*. The frame unit contains the values of the image feature points in a frame.



(a) (b)

Figure 6.2 The structure of the video unit. (a) A header unit. (b) A frame unit.

At the beginning, the server subsystem transmits the header unit to the client site. Then, the client subsystem accesses the parameters in the header unit. If the process of the access is successful, the client subsystem returns a value to notify the server site to start transmitting the frame units sequentially.

In the audio transmission, the data in an audio buffer are packed as an audio unit. Each audio unit has an order number. The server subsystem starts to transmit the audio unit to the client site when the work of the video transmission starts. And the client subsystem accesses the audio data in the audio unit. Then the audio output device plays the audio data at the client site.

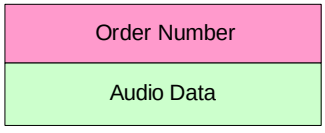


Figure 6.3 The structure of the audio unit.

6.3 Synchronization of Cartoon Videos and Speeches

Sometimes the cartoon videos and the speeches are asynchronous in the client site because of the effect of network congestions. A user can use a friendly interface mentioned in Section 5.5 to synchronize the cartoon videos and the speeches. The interface in the server and client system is designed to be used at the client site. In this interface, a delay value is used to control the delay of images. The subsystem regulates the delay of images according to the delay value. The detailed process for synchronization of videos and speeches is described as follows. A flowchart of the process is shown in Figure 6.4.

Algorithm 6.1. *Real-time synchronization of videos and speeches.*

Input: a delayed value *d*.

Output: a synchronized talking cartoon face.

Steps:

1. Set a value *s* to the order number of the first accessed video unit.
2. Store the accessed video unit into a queue *Q*.
3. If the difference between the order number of the accessed video and *s* equals *d*, pop a video unit from *Q* and render the talking cartoon face image according to the face model control points in the video unit.
4. If the value of *d* is changed in the process of real-time talking cartoon face generation, clear the video units in *Q* and reset the value of *d*. Then go to Step1.
5. Repeat Steps 2 through 4.

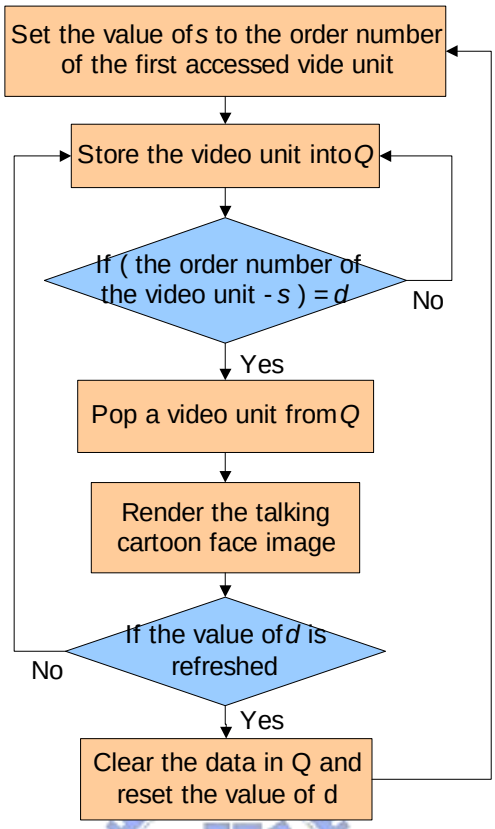


Figure 6.4 A flowchart of the process for synchronization of videos and audios.

6.4 An Application to Multi-role Avatar Broadcasting through Networks

6.4.1 Idea

The networked multi-role avatar broadcasting is a broadcast system that the receiver can choose different avatars to animate talking cartoon faces according the images and the speeches of broadcasters. It can be used in the public place of entertainment. The avatar can be different at different receptors of the broadcast system. So the broadcast system can be variegated. An illustration of multi-role avatar broadcasting is shown in Figure 6.5.

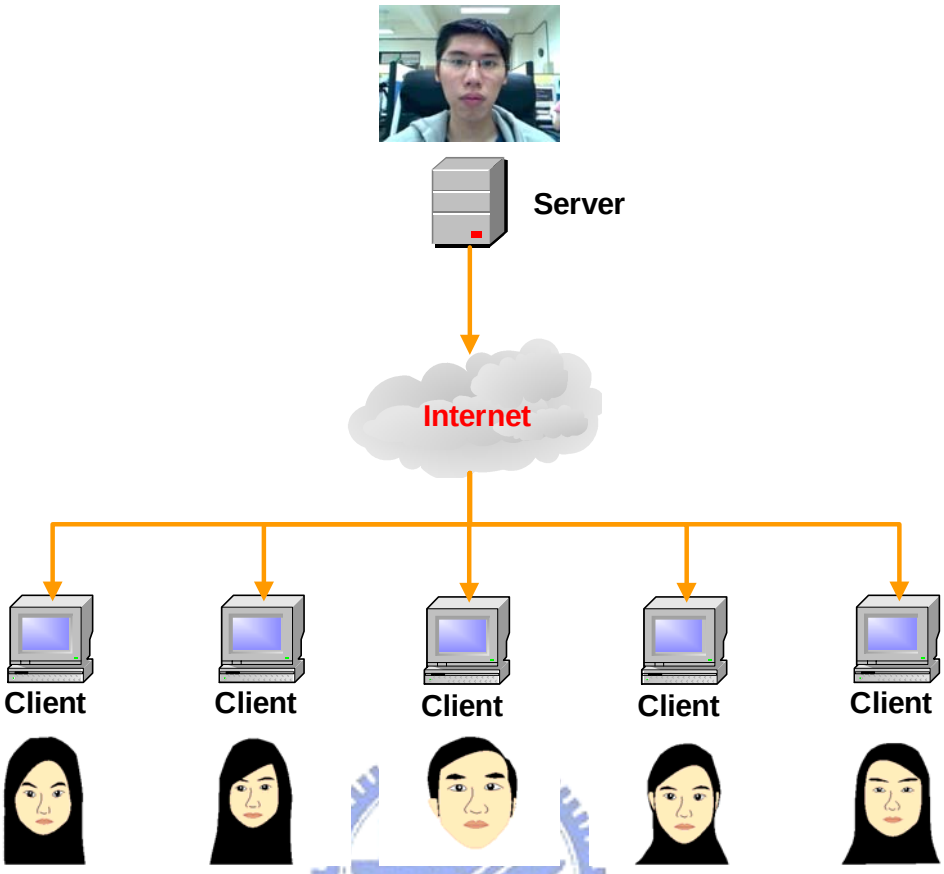


Figure 6.5 An illustration of multi-role avatar broadcasting.

6.4.2 Process of Creation

The process of creating a multi-role avatar broadcasting is shown in Figure 6.6. First, a broadcaster learns the environment at the server site. Then he/she waits for the connections of receivers. Second, a receiver chooses a face model as the avatar before connecting to the server. When one or more receivers connect to the server, the broadcaster can start to broadcast the messages to the receivers through networks. The server will track the facial features of the broadcaster and record the speeches of the broadcaster. Then the video and audio data are transmitted to each client site. The talking cartoon face is rendered in real time at the client site. If the images and speeches are asynchronous, the receiver can change the delay value to synchronize the

images and the speeches.

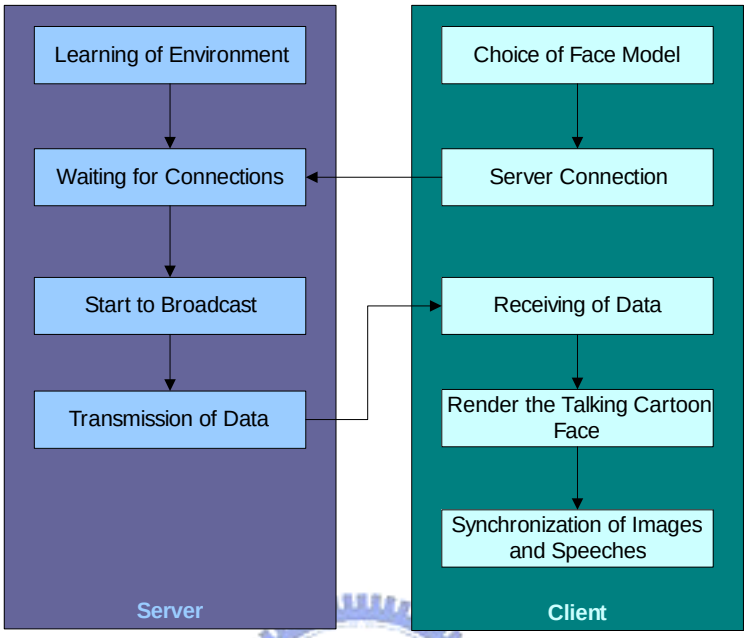


Figure 6.6 The block diagram of creating a multi-role avatar broadcasting.

6.5 An Application to Web TV by ActiveX Technique

6.5.1 Idea

Another application of the server and client system is a Web TV. Similar to multi-role avatar broadcasting, web TVs can be used more conveniently. A user can use the IE browser easily to receive the messages from the server. According to the interface of web TVs, avatar tutors, avatar reporters, and avatar singers can be spread to the world through networks. An illustration of the web TV is shown in Figure 6.7.

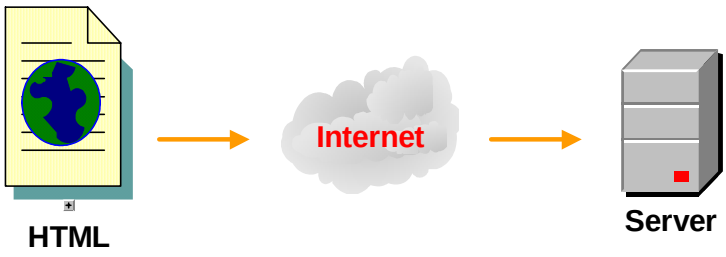


Figure 6.7 An illustration of the web TV.

6.5.2 Process of Creation

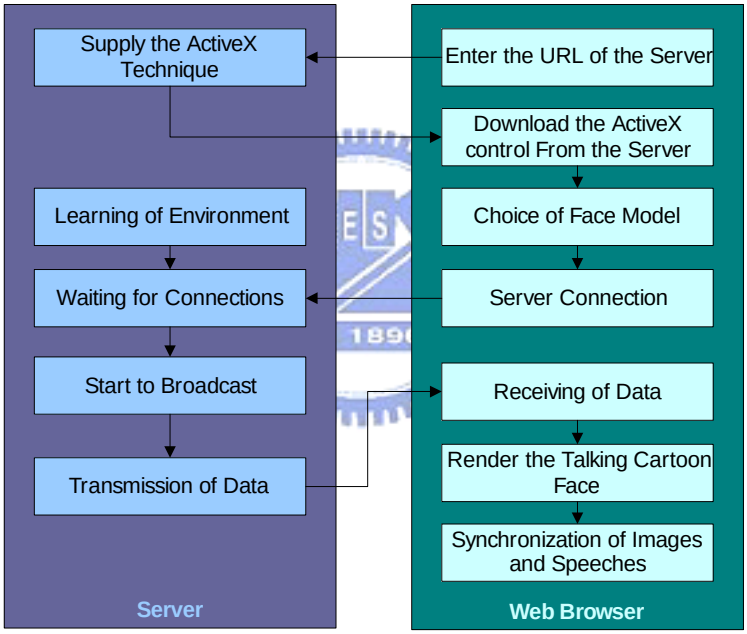


Figure 6.8 The block diagram of creating a web TV.

The process of creating a web TV is shown in Figure 6.8. The process is similar to the process of creating a multi-role avatar broadcasting. The difference is that the operations at the client site are in a Web browser. An ActiveX technique is used in this section. ActiveX is a technology developed by Microsoft for sharing information among different applications. One specific way of implementations of ActiveX

technologies is ActiveX controls. An ActiveX control can be automatically downloaded and executed by a Web browser. The client subsystem is implemented as an ActiveX control in this section. A user must enter the URL of the server in the Web browser. The Web browser will download the ActiveX control from the server. Then a user can operate the client subsystem in the Web browser.

6.6 Experimental Results

Some experimental results for the two applications of real-time talking cartoon face generation system for use on networks are shown in this section.

In the first application, the program interfaces of the multi-role avatar broadcast system are shown in Figure 6.9 and Figure 6.10. First, Figure 6.11 shows the learning of environments in the server site. Then the server starts to wait for the connections from the client site, as shown in Figure 6.12. Second, Figure 6.13 shows the choice of a face model in the client site. Then the client starts to connect to the server site, as shown in Figure 6.14. And the multi-role avatar broadcast system is started to work. The real-time process of facial feature tracking in the server site is shown in Figure 6.15. The resulting real-time talking cartoon face in the client site is shown in Figure 6.16.

In the second application, the process of entering the URL of the server and downloading the ActiveX control from the server is shown in Figure 6.17. Figure 6.18 shows the program interface of the web TV in the Web browser. The process in the server site is similar to the process in the first application. The resulting real-time talking cartoon face in the Web browser is shown in Figure 6.19.

Figure 6.9 The program interface of multi-role avatar broadcast system in the server site.

Figure 6.10 The program interface of multi-role avatar broadcast system in the client site.

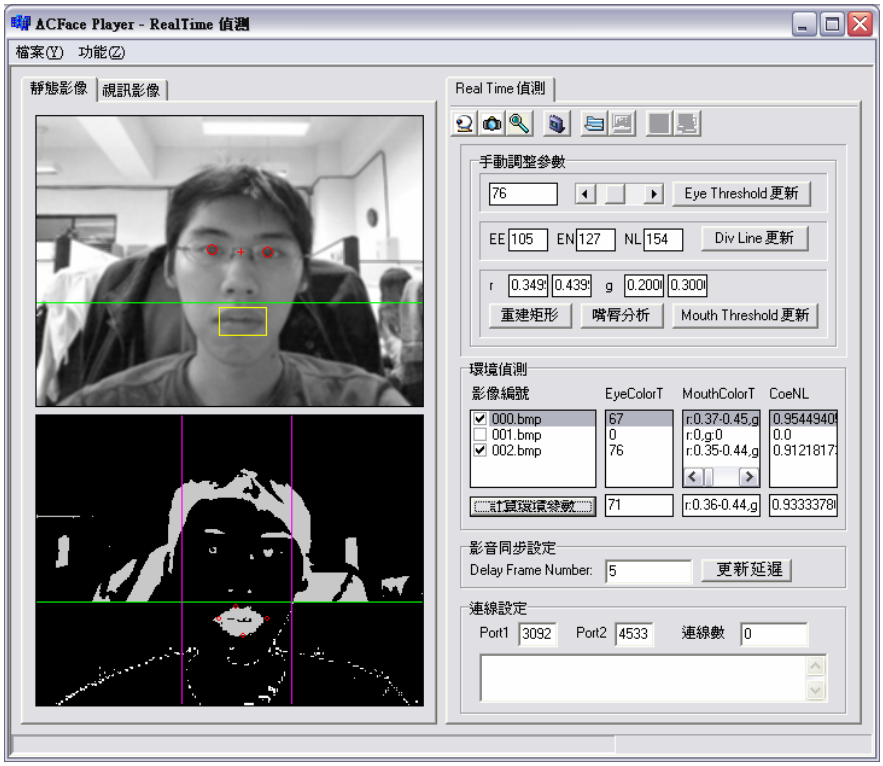


Figure 6.11 The process of learning of environments in the server site.

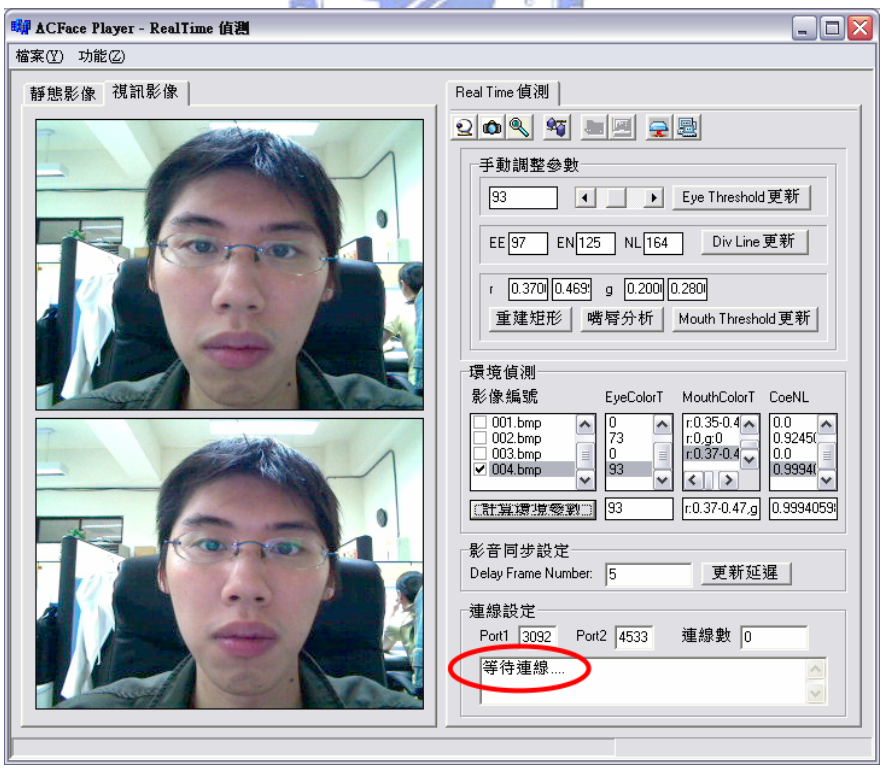


Figure 6.12 The process of waiting for the connections from the client site.

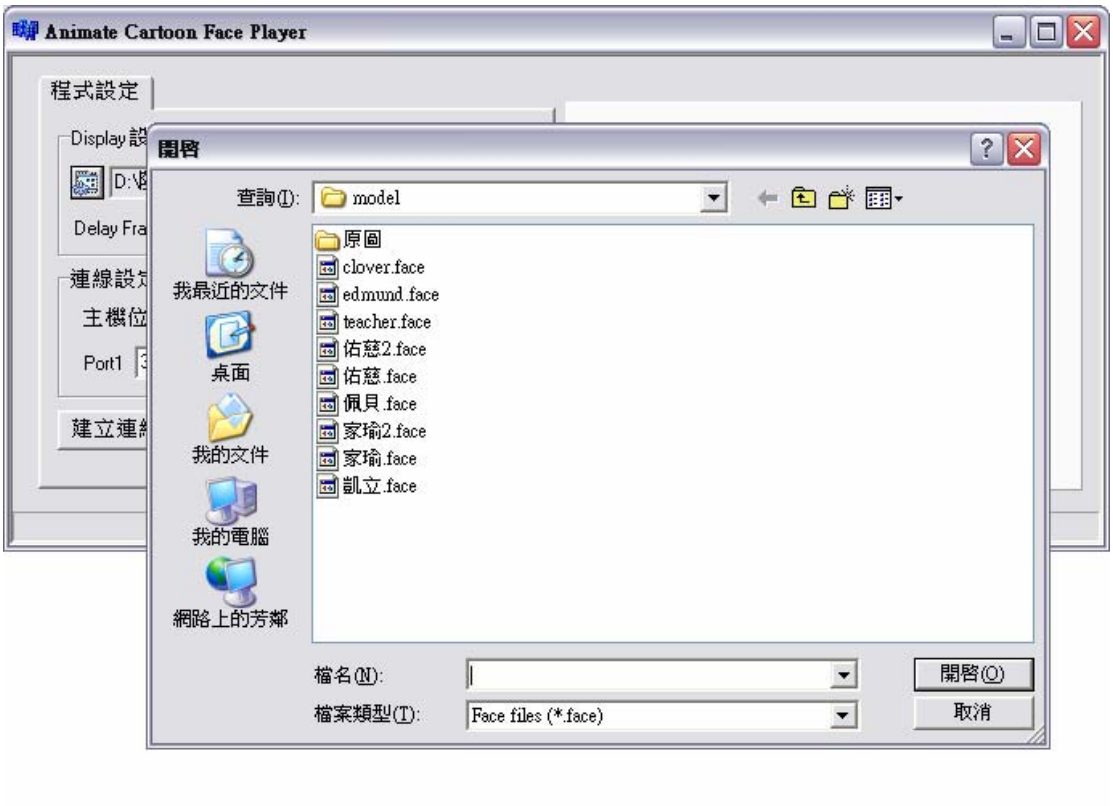


Figure 6.13 The choice of the face model in the client site.

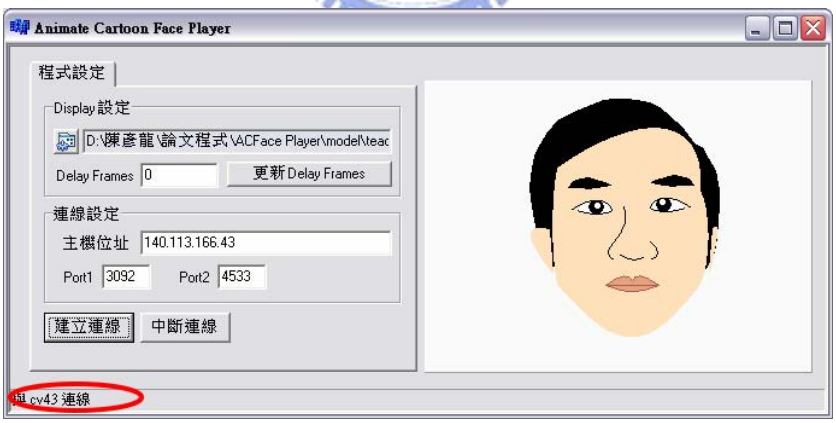


Figure 6.14 The process of connecting to the server in the client site.

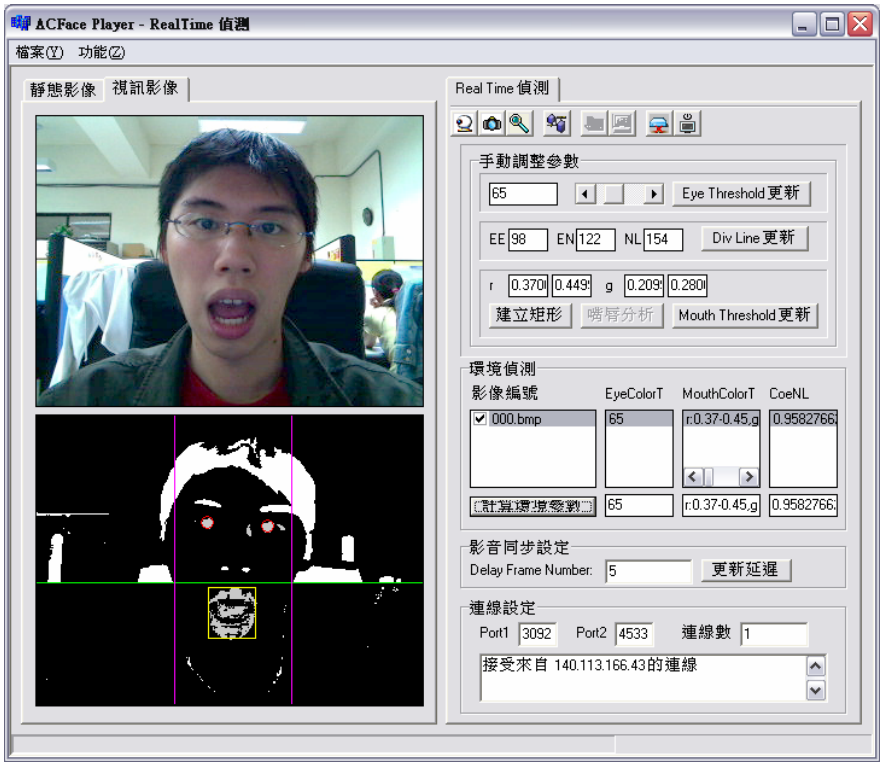


Figure 6.15 The real-time process of facial feature tracking in the server site.

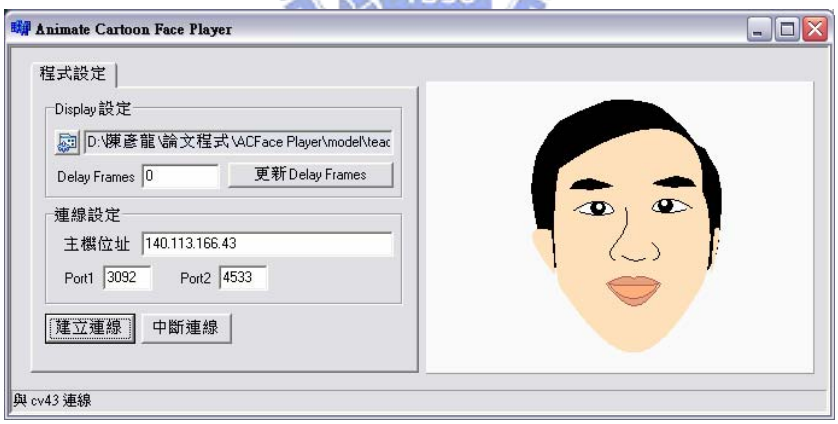


Figure 6.16 The resulting real-time talking cartoon face in the client site.

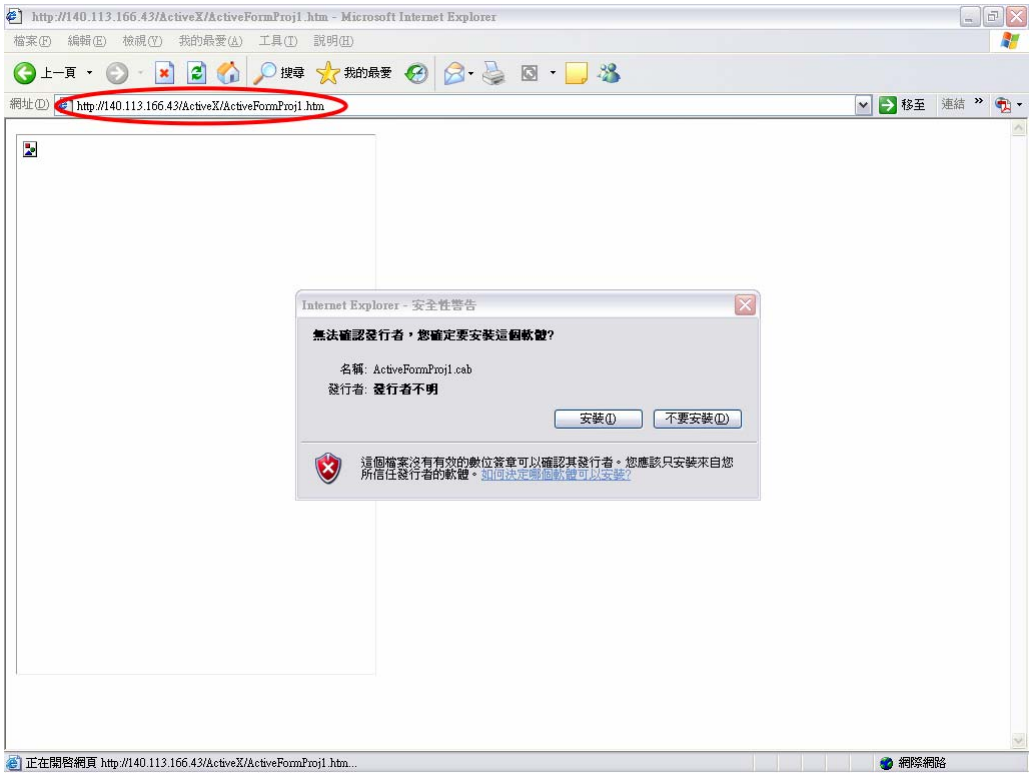


Figure 6.17 The process of entering the URL of the server and downloading the ActiveX control from the server.

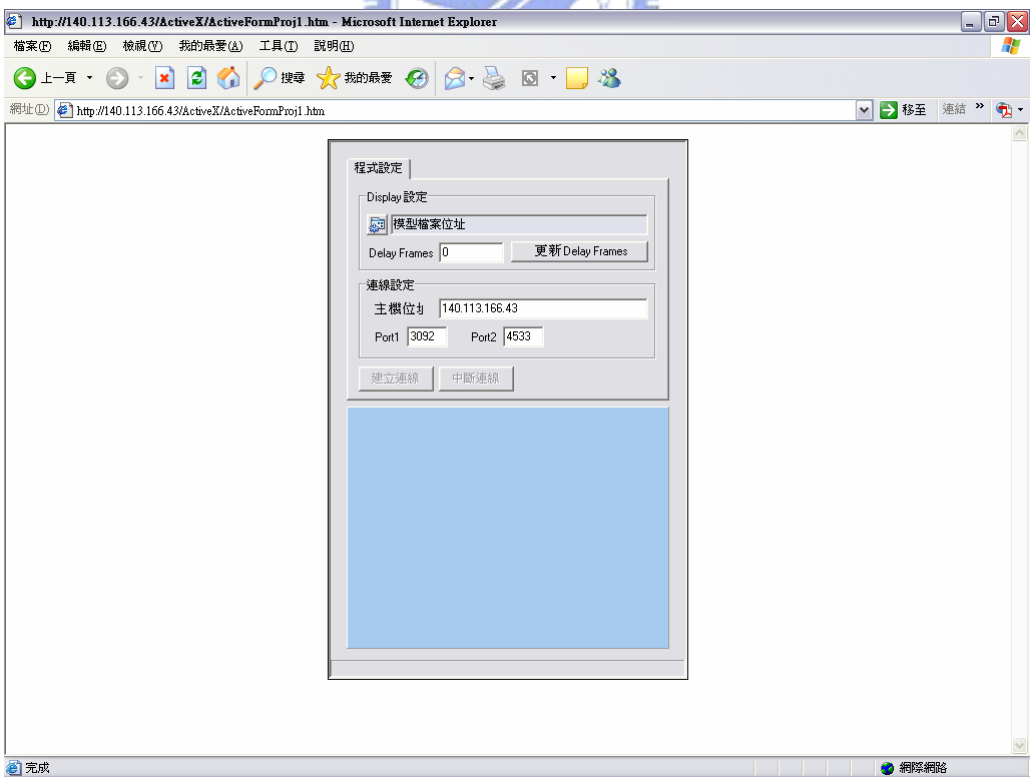


Figure 6.18 The program interface of web TV in the Web browser.

Figure 6.19 The resulting real-time talking cartoon face in Web browser.