Server-Aided Recognition of Finger Shapes in a Portable System

Cheng-Wei Lin, Lin-Fei Sung, Shyh-In Hwang, and Wei-Ting Chou
Department of Computer Science and Engineering
Yuan-Ze University
Taoyuan, Taiwan, ROC
Email: desty@mmlab.cse.yzu.edu.tw, floyd@mmlab.cse.yzu.edu.tw, shyhin@cs.yzu.edu.tw, s976056@mail.yzu.edu.tw

Abstract

Sign language recognition had been developed for many years, but none of these methods consider about the convinence and availability for deaf community. We propose a client-server system which is designed for easy to use with low-cost equipments. This system consists of a hand-held device client and a computer server, the hand-held device is used to capture the gestures of a user, transmit data to the computer server through wireless network, and receive the result of a recognition for multi-lingual sound or text output. In this system, deaf people with this client could communicate with normal people more easily and efficiently with mobility. This paper implemented the system and examined about the power consumption of the client, the usage of wireless network bandwidth and a set of recognitions in Taiwan sign language.

Keywords: sign language, finger, distributed, recognition

1. Introduction

As the improvement of the hardware of hand-held devices like a cell-phone or a PDA, the abilities of these equipments are more and more powerful. But the deaf community still can not use these technology to ameliorate their life. Transmitting the sign language video through 3G network is not afford, because the quality, the popularity, and the price is not accepted. The most communicating methods they used is limit to the sign language and the hand writing. But the sign language is hard to understand for normal people, and the hand writing is effectlessly to each other.

The most of sign language recognitions require a powerful computer to resolve the postures signed by people, this does not consider about the usability and mobility of a user. If the recognition procedure is implemented on the hand-held devices, the battery will be used drastically.

We use the client-server architecutre to distribute the heavy process to a recognition server, and propose the key-frame generating to save the bandwidth usage. In this paper, we also improve the finger shape recognition, implement our system and examine about its performance.

2. Related Works

2.1. Distributed Recogniton Computing

George Nagy had proposed a CAVAIR[5] system which could compare and recognize for some images, this system is divided into a client and a server, the client is a PDA system which is used to capture and transmit an image to the server, the server recognizes an image and search for relative information of the recognition on network.

2.2. Sign Language Recognition

The two main appraoches of the sign language recognition[8] are image process and the sensor glove. The image process is using the color information of a picture to detect the hand region for recognizing, these kind of methods mainly have some steps, the skin color detection, the motion detection[12], [13], and the recognition. The skin color detection has to choose the color space which is sensitive to skin color[11], generally, it is necessary to convert the color space. The motion detection could be archived by calculate the difference of two pictures, by this way, the motion of a gesture could be resolved. The recognitions of a hand shape and a gesture are two different researches, the hand shape recognition focuses on the fingers, the palm, and its meaning, while the gesture recognition is to detect the movement. J.N. Huang divided the hand shapes of the Taiwan sign language into 49 shapes[9], and using the angle normalization to classify the detected shapes. C.S. Cheng used the trajectory representation to a sequence of video frames[2] to find the motion of a gesture, and then use the hidden markov model to recognize.

The sensor glove is to use the sensors of the glove to capture the gesture data and recognize. The sensor glove could avoid the luminace problem, Jun Park and Y.L. Yoon proposed a interaction system based on its LED glove[10], and the gesture are recognized in hidden markov model[6]. In this system, the positions of LEDs are captured by camera, and using these positions to analyse the gesutre of a user. A. Togenetti developed a data-glove[1] which equips 20 sensors, the research is focus on how to design a comfortable sensor glove. The gesture recognition method they used to examine is least square and the neural network[4]. Using the sensor glove could avoid the interference of enviroment, but it has higher cost and not suitable for daily use.

3. System

The hand held device like a cell phone or a PDA has become more and more powerful recently, most of them equip a camera and a wireless interface. But these device still can not handle the recognizing process because of the limitation of battery. The recognizing process would cause a lot of computing, so we distribute this heavy loading to a personal computer server, like the CAVAIR[5] do. The server of CAVAIR has to search the network for furthermore information of its recognition, therefore the client has to wait until the server transmit its result. This would increase the waiting time of the client, so it is not suitable for real-time approach. In our system, the server will send the result right after the recognition finished. The communication flow is like Figure1, by this way, we could decrease the waiting time of the client.

To recognize a gesture or a hand shape via a video requires a lot of network bandwidth in our system, in order to save the bandwidth usage, we propose the key-frame generating to eliminate this problem. The concept of key-frame generating is to fetch the significant postures of a sentence in the Taiwan sign language. A sentence in the Taiwan sign language is composed in phrases, and its grammar is lossely, so a sentence could be represented in a number of certain independent postures. The client could utilitze this property to generate a set of pictures, which are the key-frames of one sentence, and transmit it to the server to recognize. As the result, the key-frame generating could reduce the usage of bandwidth, the transmitting time of a client could be decreased, and the life time of the battery of a client could be extended. The trigger of a key-frame is used by the client to capture a posture, we use the sound as the trigger because it is a nature way to operate. The sound trigger is not a voice, it is a throat sound like cough, so it could be used by anyone including the deaf people. Using the sound trigger could save the two hands of a user, and one could make any posture freely. Figure2 shows the concept of the key-frames generating.

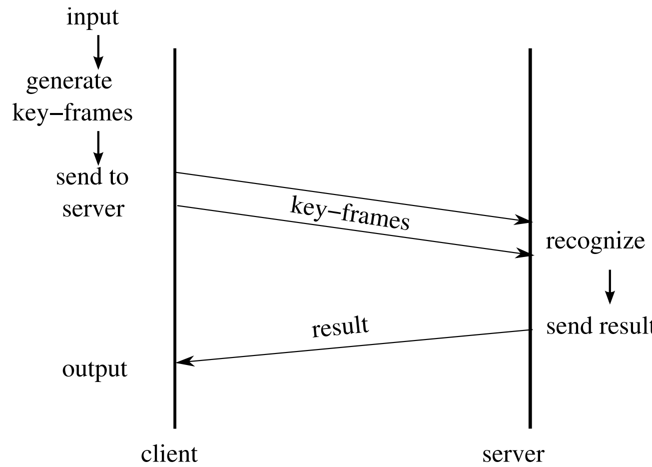

Figure 1. Communication Flow

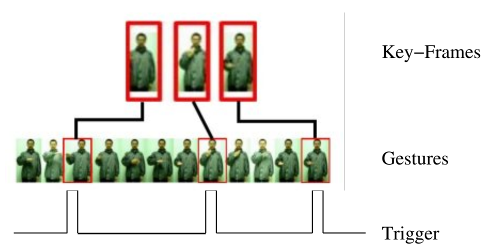

Figure 2. The Concept of Key-Frames Generating

Our goal is to recognize that how many fingers in a key-frame, the recognition steps are color space convertion, skin color detection, and finger shape recognition. The content of a key-frame is compressed into JPEG format which is widely used in many applications, the default color space of JPEG is RGB which is easily affected by luminance, so the RGB color space is not suitable for skin color detection.

Although various color spaces are used for different purpose[11], but only some of these color spaces are sensitive to the skin color[2], [9], [18], [15], [17], [14], in these color spaces, the skin color are limit to a certain interval. The YIQ color space is one of these color spaces which are sensitive to skin color[21]. In YIQ color space, the luminance are separated and stored in the Y-component, and the color is stored in the I-component and the Q-component. The I-component stores the color from orange to blue and the Q-component stores from purple to green. The skin color is limited when the I-component is 20 to 90 in YIQ color space, so it could be used to detect the region of skin color in a key-frame. Another advantage of YIQ color space is its convertion is a linear transformation, which is more faster than other color spaces, therefore we choose the YIQ color space to detect the region of skin color.

The fingers of human hand could be recognized by the fingertip and the finger shape[19], [20]. Hardenberg proposed a rounded fingertip model to recognize a finger of the hand, but it has less accuracy of recognition because it lacks the condition of finger shape. P. Song uses that fingertip model to detect the fingertips and then calculate the width orthgonally for a certain length[16]. Since their method calculate the width orthgonally, it can not recognize the curve finger shapes. Figure3 is an example of the failed recognition of curve finger shape, If the width j is much bigger than i, then it will consider its length is k, and this finger would not be recognized correctly.

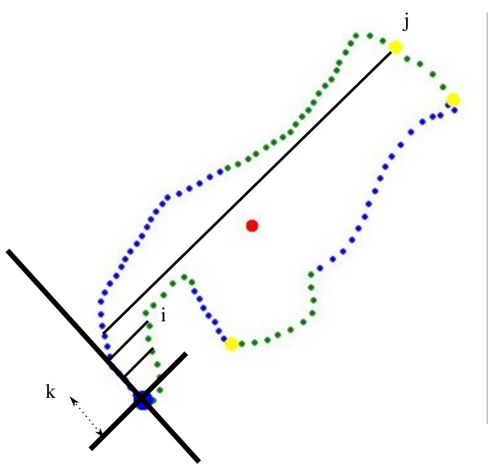

Figure 3. Failed Recognition of Curve Finger Shape

Our method is to find the centroid of the outline coordinates of hand region, and then calculate the distance between the centroid and the outline coordinates. We could use these distances to find where the fingertips are, and use the width of the symmetric coordinates to recognize a finger. By this way, we could handle more finger shapes and recognize more precisely. Figure4 shows that how to find the fingertip, if c is the centroid of the outline coordinates, $cd2 > cd1$ and $cd2 > cd3$, then we could consider that coordinate $d2$ is the fingertip candicate.

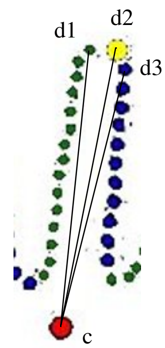

Figure 4. Find the Fingertip

The width of symmetric coordinates is calculated from the fingertip, so its direction is not orthgonal to the fingertip, and it is more flexible for any shape of a finger. Figure5 is an example of a curve shape recognition, we consider a, b, c, and d as the fingertip candicates, and calculate the width of symmetric coordinates of these candicates. If the average width of a fingertip candicates is twice bigger than the total average width, or if the maximum width of fingertip candicates is twice bigger than its minimum width, then this fingertip candicate is not possible to be a finger. So there has only one finger d in Figure5.

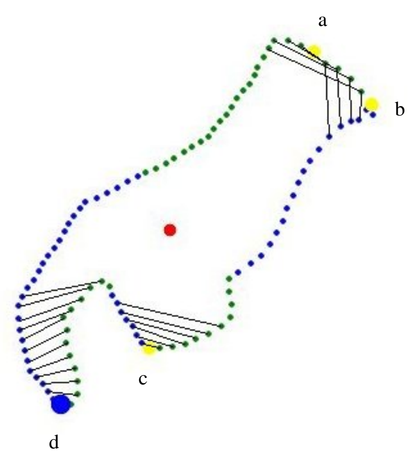

Figure 5. Curve Finger Shape Recognition

4. Experimental Results

We implement our client system in a notebook with the Logitech QuickCam Pro which has a built-in microphone, and using a personal computer as the server system. Our software development platform is the GNU/Linux Debian distribution, the library used are listed in Table1.

library	description
ImageMagick C++ library	image process
SDL	used to access the video data
ALSA	sound process and driver
LUVC	quickcam driver
berkeley sockets	network communication

Table 1. The Libraries Used

In the portable system, the input data size of the client would affect the bandwidth usage, we use the size of the video capturing and the key-frame generating to campare. The frame size is 640∗480, the capturing time is 20 second, the content is one hand shows zero to five fingers, and test for 100 times. Table2 shows that the key-frame generating results in more files than the video capturing, but the total file size is much smaller than the video capturing.

	video	key-frame
file number	1	12
file format	avi	jpg
average single file size	6,931,854 bytes	15,281 bytes
total file size	6,931,854 bytes	183,372 bytes

Table 2. The File Size Comparision

The trigger of key-frame generating is used to capture the most important postures of a sentence, its functionality likes a shutter of a camera. The average capturing time of one key-frame is about 0.5 second.

Since there are 3 different hand size 150∗150, 250∗250, and 350∗350 in the Taiwan sign language, we test each size of left and right hand shows zero to five fingers with different genders. The total number of test key-frames are 360. As the Table3 shows, the recognition accuracy is imporved with bigger hand size. This is because that the bigger hand size in a key-frame has more information of the finger shapes, and has less possibility of the fail recognition.

Accuracy(%)			
fingers number	150∗150	250∗250	350∗350
0	90.1	93.0	95.4
1	82.3	91.7	93.8
2	83.2	90.5	94.1
3	85.3	92.1	95.9
4	84.5	91.6	95.6
5	83.1	91.1	93.7
average	84.75	91.65	94.75

Table 3. The Accuracy of Different Frame Size

The fail recognitions come from two reasons, the wrong detection of the thumb, and the error detection of the finger width. Figure6 is an example of the error recognition, it shows that the carpus of up-right corner is recognized as a finger, and the thumb has not been recognized. The carpus is recognized as a finger because of the diffenence of width has grow up smoothly, this could be conquer by calculate the average width and the maximum width of fingers in a key-frame. As for the thumb, it is more hard to detect becaue of the face of fingertip of the thumb is different from other fingers, so the fingertip detection must be improved to solve this problem.

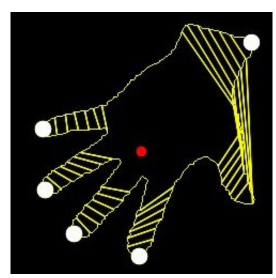

Figure 6. Failed Recognition

The time used for a key-frame recognition is listed in Table4, it shows that the more time is used to recognize with more fingers at the same frame size, this is because the more fingers increase the more skin region and the length of the outline. The bigger hand size also increase the processing time for the same reason. The average recognition time of a key-frame is about 1.497 second.

time used(second)			
fingers number	150∗150	250∗250	350∗350
0	0.74	1.17	1.77
1	0.82	1.29	1.83
2	0.87	1.36	1.96
3	0.94	1.44	2.14
4	1.06	1.58	2.31
5	1.12	1.69	2.54
avarage	0.925	1.421	2.091

Table 4. The Recognition Time of Different Frame Size

The waiting time of a client includes it sends a key-frames to the server, waits for the recognition, and receives the result. The client sends and receives data through IEEE 802.11g, the distance of the client and the server is about 15 meter, and the average number of routers is 3. The average transmitting time of a key-frame is about 0.1 second in our experiments, so the total average waiting time is about 1.5 to 2.0 second.

5. Conclusion

We propose a distributed recognition system which improve the existing finger recognition, and we examine its usability and recognition accuracy. This system uses the key-frame generating to decrease the bandwidth usage, and distribute the heavy load of a recognition to a server.

The key-frame generating proposed in this paper is a creative and useful method, which could be used for many applications in multi-media and wireless approach. It aims to increase the preformance of a client, decrease the power comsumption, and less network bandwidth usage. The sound trigger of key-frame generating may be affected by the environment. When the sound of backgound is louder than the trigger threshold, the system will produce incorrect key-frame, which will cause wrong result. This problem could be solved by human sound recognition, we leave this issue for future work.

The preformance of the finger shape recognition still have to improve for real-time approach. The system implementation on a embeeded board is necessary for more flexible, and the schedule of multi-clients is another important issue for future research.

References

[1] A. Tognetti, N. Carbonaro, G. Zupone and D. D. Rossi, *Characterization of a Novel Data Glove Based on Textile Integrated Sensors*, Engineering in Medicine and Biology Society, 2006.

[2] C. S. Cheng, P. F. Hsieh and C. H. Wu, *Hand Motion Recognition for the Vision-Based Taiwanese Sign Language Interpretation*, ACII 2005 LNCS 3784, 2005.

[3] C. von. Hardenberg and F. Berard, *Bare-Hand Human-Computer Interaction*, PUI, 2001.

[4] E. Stergiopoulou and N. Papamarkos, *A New Technique for Hand Gesture Recognition*, ICIP, 2006.

[5] George Nagy, *Interactive, Mobile, Distributed Pattern Recognition*, ICIAP, 2005.

[6] H. K. Lee and J. H. Kim *An HMM-Based Threshold Model Approach for Gesture Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21, 10, October, 1999.

[7] H. K. Lee and J. H. Kim, *Gesture spotting from continuous hand motion*, Pattern Recognition Letters 19, 1998.

[8] J. Joseph and LaViola Jr., *A Survey of Hand Posture and Gesture Recognition Techniques and Technology*, CS-99-11, 1999.

[9] J. N. Huang, P. F. Hsieh and C. H. Wu, *Vision-Based Recognition of Hand Shapes in Taiwanese Sign Language*, ACII 2005 LNCS 3784, 2005.

[10] J. Park and Y. L. Yoon, *LED-Glove Based Interactions in Multi-Modal Displays for Teleconferencing*, Artificial Reality and Telexistence–Workshops, 2006.

[11] J. Terrillon and S. Akamatsu, *Comparative Performance of Different Chrominance Spaces for Color Segmentation and Detection of Human Faces in Complex Scene Images*, J. Terrillon and S. Akamatsu, Comparative Performance of Different Chrominance Spaces for Color Segmentation and Detection of Human Faces in Complex Scene Images, International Conf on Face and Gesture Recognition, pp54-61, 2000.

[12] K. K. Kim and K. C. Kwak and S. Y. Chi, *Gesture Analysis for Human-Robot Interaction*, Advanced Communication Technology, 2006.

[13] M. K. Bhuyan and D. Ghoah, *A Framework for Hand Gesture Recognition with Applications to Sign Language*, Annual India Conference, 2006

[14] M. W. Schwarz, W. B. Cowan and J. C. Beatty, *An Experimental Comparison of RGB, YIQ, LAB, HSV, and Opponent Color Models*, ACM Transactions on Graphics, 6, 2, April, 1987.

[15] N. Habili, C. C. Lim and A. Moini, *Segmentation of the Face and Hands in Sign Language Video Sequences Using Color and Motion Cues*, IEEE Transaction on Circuits and Systems for Video Technology, 14, 8, August, 2004.

[16] P. Song, S. Winkler, S. O. Gilani and Z. Y. Zhou, *Vision-Based Projected Tabletop Interface for Finger Interactions*, HCI 2007 LNCS 4796, 2007.

[17] S. L. Phung, D. Chai and A. Bouzerdoum, *Adaptive Skin Segmentation in Color Images*, ICASSP, 2003.

[18] S. Jayaram, S. Schmugge, M. C. Shin and L. V. Tsap, *Effect of Colorspace Transformation, the Illuminance Component and Color Modeling on Skin Detection*, CVPR, 2004.

[19] T. Rhee, U. Neumann and J. P. Lewis, *Human Hand Modeling from Surface Anatomy*, I3D, 2006.

[20] Z. Mo and U. Neumann, *Real-time Hand Pose Recognition Using Low-Resolution Depth Images*, CVPR, 2006.

[21] X. He, Z. M. Liu and J. L. Zhou, *Real-Time Human Face Detection in Color Image*, Proceeding of the Second Internatinal Conference on Machine Learning and Cybernetics, 2003.