H. Gao, M. Lim, E. Lin, R. Green, 'Marker Based Facial Tracking Application in Communication Disorder Research', *Proceedings of Image and Vision Computing New Zealand 2007*, pp. 244–247, Hamilton, New Zealand, December 2007.

Marker based Facial Tracking Application in Communication Disorder Research

Hongzhi Gao, Dr. Marilyn Lim, Dr. Emily Lin, Dr. Richard Green

University of Canterbury
Email: hongzhi.gao@canterbury.ac.nz

Abstract

An infrared reflective marker tracking system is introduced in this paper. This system is designed to measure facial muscle movement for communication disorder research. The extent and timing of jaw opening and lip spreading are monitored while subjects are performing a variety of speech production tasks with different speech or voice facilitating strategies. This facial tracking device is used in combination with acoustic and electroglottographic recordings to allow for an investigation of the oral-laryngeal coordination. A calibrator with known geometry properties is used to compensate for errors resulted from a subject's head movement and convert pixel distance to real distance in millimeters.

Keywords: communication disorder, infrared reflective marker, marker based tracking

1 Introduction

Facial muscle movement is valuable information in communication disorder research. In this paper, we present a computer vision based motion capture system to measure a subject's facial muscle movement accurately in real-time.

Motion capture (MoCap) gains a lot attention in computer vision research literature. MoCap research works in the past can be classified into two categories: marker based MoCap [1] [2] [3] and markerless MoCap [4] [5]. The advantages of a marker based MoCap system is highly efficient and accurate. Therefore, such systems are widely applied in movie and gaming industry. Nowadays, computer scientists are trying to make MoCap become markerless. However, most of markerless MoCap systems have strict restrictions on environment and the tracking is not reliable.

Marker based mocap algorithm is applied in this research to produce high accuracy real-time measurement on facial muscle movement. The system designed in this research is represented in the next section followed by a quantitative experiment on its accuracy.

2 System Description

Figure 1 shows the physical setup of the marker based facial tracking system proposed in this paper. A subject (not shown on this photo) sits on the chair in figure 1 during the tracking process. The webcam shown in figure 1 will be pointed towards the subject's face. As shown in this photo, a black background is used to minimize the chances of unexpected highlight spots being observed.

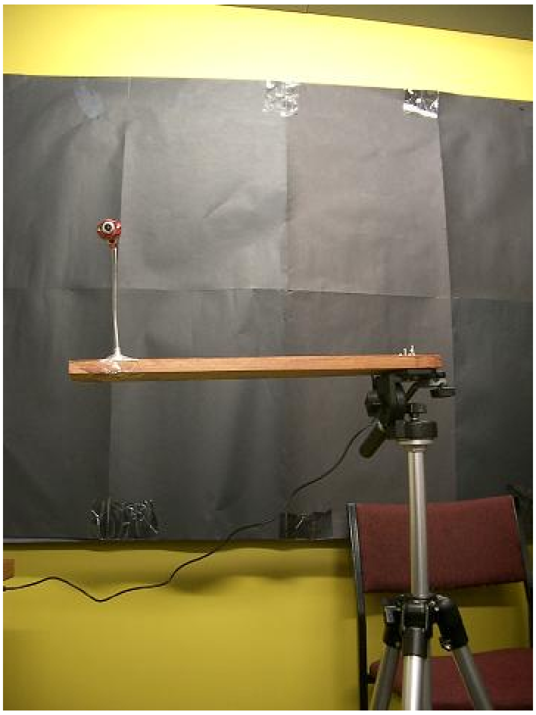

Figure 1: Physical setup of the marker based facial tracking system.

Figure 2 shows a close look at the webcam used in this project. The two LEDs placed on both side of the webcam lens are infrared LEDs. They are used to provide infrared illumination during the tracking process.

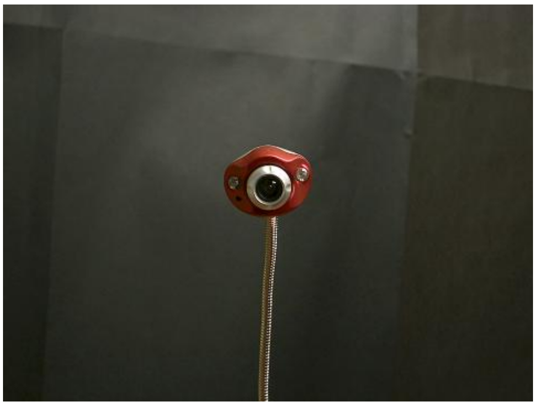

Figure 2: The webcam with infrared backlight used in this project.

Figure 3 shows the infrared reflective markers and their placement on a human face. These markers are round dots shape with seven millimeters diameter. The four markers placed on a black card board and attached onto the subject's forehead area are for calibration purpose. The distance between the left calibration dot and right one is exactly 30 millimeters, which is the same as the distance between the top calibration dot and the bottom one on the black card board. The four measurements infrared markers are placed on the nose, on both lips spreading and under the bottom lip. When the subject is speaking, the movement of his / her facial muscles around mouth area will be disclosed by the movement of these measurements dots.

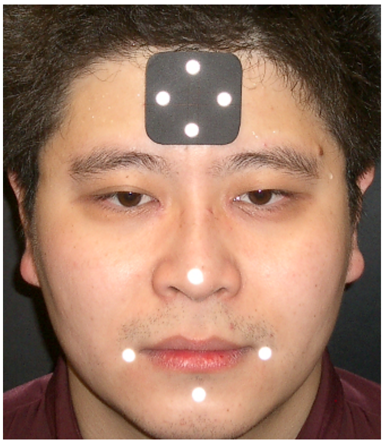

Figure 3: The infrared reflective markers and their placement on a human face.

The software part of this facial tracking system contains two major modules: low level computer vision module and infrared reflective marker alignment module. The low level computer vision module converts color image grabbed from the webcam to an array of connected components (blobs). These blobs will be further processed by the infrared reflective marker alignment module to disclose quantitative facial muscle movement information.

Figure 4 illustrates the algorithm used in the low level computer vision module. Since the color information in an image may lose when infrared illumination is enabled, the computer vision operations are performed over single channel gray scale image. As shown in figure 4: firstly a three channels BGR color image is converted to a single channel gray scale image. This gray scale image is filtered with a predefined threshold value, which generates a binary mask that indicates highlight areas in the original image. Then, morphological operation (open) is performed on tthis binary mask to remove noises. This binary mask is converted to an array of connected components. Then, these connected components are filtered with their area properties to remove extra large or small size blobs that are not possible to be an infrared reflective marker. This array of blobs is the final output of the low level computer vision module and is sent to the infrared reflective marker alignment module for further processing.

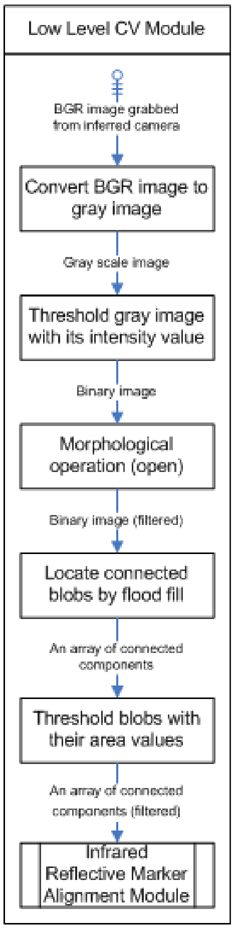

Figure 4: The flowchart of the low level computer vision module.

Figure 5 shows the algorithm used in the infrared reflective marker alignment module. As shown, it takes an array of connected components as input.

Firstly, this module checks whether there are exactly eight dots (markers) in the array. If affirmative, these dots will be aligned to the four calibration markers and four measurement markers according to their relative positions. Pixel distances between measurement markers are calculated and then converted to real distances in millimeter with prior knowledge of the calibration markers' geometry properties. In the case that more than eight dots are found in the input array, an extra filtering process is performed to remove dots that far from any markers in the pervious frame.

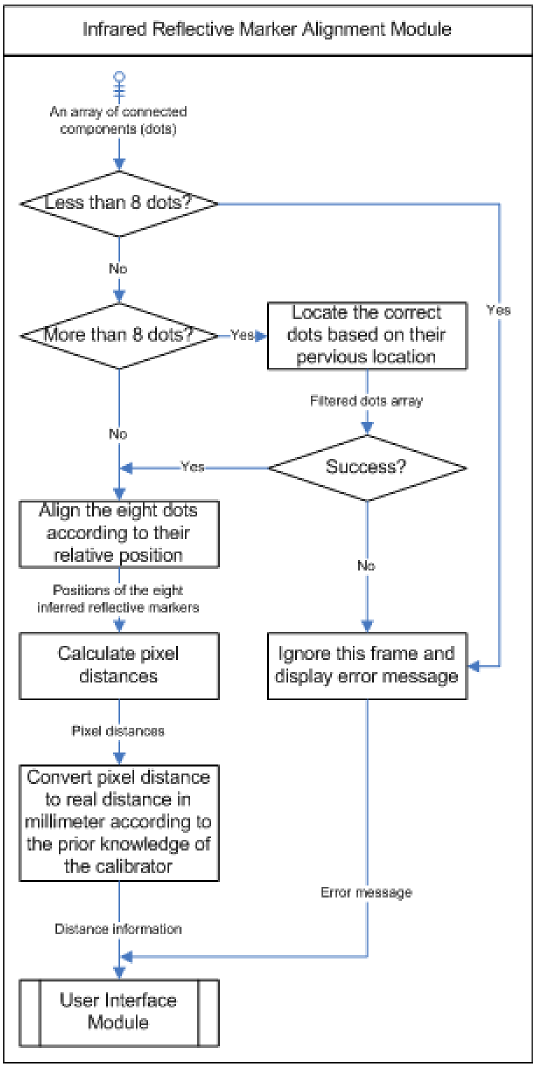

Figure 5: The flowchart of the infrared reflective marker alignment module.

Figure 6 shows an example that the extra dots filtering process is performing. As shown, there are more than eight dots being detected in this image. The extra dots are the results of high reflective teeth surface. Yellow circles in this image illustrate acceptable areas of the markers in the current frame according to the positions of these markers in the pervious frame. As shown, this filtering algorithm successfully finds out the eight markers from all detected high reflection areas.

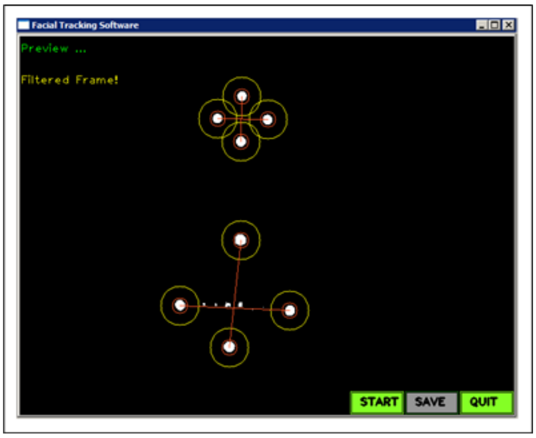

Figure 6: The extra dots are detected because of high reflection of teeth area.

3 Experiment

Figure 7 shows the experiment results of still tracking test. The goal of this test is to evaluate the maximum tracking error when the subject is not speaking. During this experiment, we asked the subject try to keep mouth muscle relax and still. The subject can move their head as long as all of the reflective markers are in sight of the webcam.

As shown in figure 7, the maximum vertical distance between the marker placed on the subject's noise and the marker placed under the subject's lower lip is 63.21mm and the minimum vertical distance between these two markers is 61.54mm. Therefore, the maximum vertical error during this experiment is 1.68mm (2.69 percent of the average 62.30mm in vertical distance). The maximum horizontal distance between the two markers placed on both side of the lips spreading is 72.01mm and the minimum horizontal distance detected in this test is 71.05mm. Therefore, the maximum horizontal error during this experiment is 0.96mm (1.38 percent of the average 69.77mm in horizontal distance).

4 Conclusion and Future Work

In this paper, we presented an infrared reflective marker based facial tracking system. This system is designed to support further research works in communication disorder domain. A webcam with infrared backlight is used in this system to track eight infrared reflective markers placed on subject's face. Experiment shows that in still condition, the maximum error in vertical direction is 1.68mm and the maximum error in horizontal direction is 0.96mm.

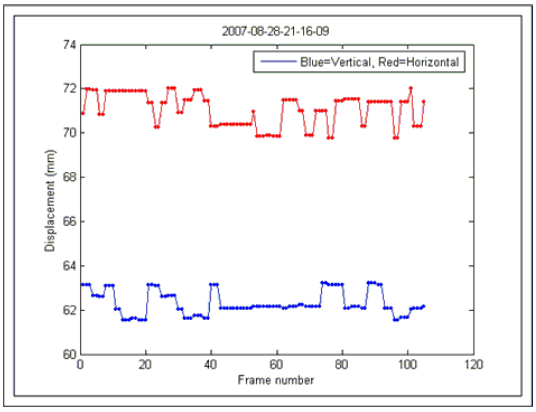

Figure 7: The accuracy experiment results.

In the future, we will increase the number of markers that can be tracked by this system so that it can be used as high accuracy human face expression software.

References

[1] F. Steinicke, C. Jansen, K. H. Hinrichs, J. Vahrenhold, and B. Schwald, "Generating optimized marker-based rigid bodies for optical tracking," in *2nd International Conference on Computer Vision Theory and Applications*, INSTICC Press, 2007, pp. 387–395.

[2] M. Kaltenbrunner and R. Bencina, "reactivision: A computer-vision framework for table-based tangible interaction," in *the first international conference on Tangible and Embedded Interaction*, Louisiana, USA, 2007.

[3] R. Budiman, M. Bennamoun, and D. Q. Huynh, "Low cost motion capture," in *International Conference on Image and Vision Computing New Zealand*, Dunedin, New Zealand, November 2005.

[4] A. Sundaresan and R. Chellappa, "Markerless motion capture using multiple cameras," *Computer Vision for Interactive and Intelligent Environments*, vol. 25, pp. 1337–1342, 2006.

[5] S. Corazza, L. Muendermann, A. Chaudhari, T. Demattio, C. Cobelli, and T. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach," *Annals of Biomedical Engineering*, vol. 34, pp. 1019–1629, 2006.