

A System for Teaching Sign Language using Live Gesture Feedback

Daniel Kelly, John McDonald, Charles Markham
N.U.I. Maynooth
Ireland
dankelly@cs.nuim.ie

Abstract

This paper presents a computer vision based virtual learning environment for teaching communicative hand gestures used in Sign Language. A virtual learning environment was developed to demonstrate signs to the user. The system then gives real time feedback to the user on their performance of the demonstrated sign. Gesture features are extracted from a standard web-cam video stream and shape and trajectory matching techniques are applied to these features to determine the feedback given to the user.

1. Background

In today's world, communication is predominantly expressed in the form of voice sounds, but for anyone with an inability to communicate using voice an alternative form of communication such as sign language is needed. It is a general rule of thumb that 1 in every 1000 people are deaf. The majority of deaf people rely on Sign Language as their alternative form of communication; however, because the number of people, in the general community, who understand sign language is small, communication with people outside of the immediate deaf community can be extremely difficult.

The practical application of the application detailed in this paper is to provide the general community with a virtual environment for learning sign language. The virtual teaching environment provides an immersive learning experience while also retaining the flexibility of learning at the users convenience on a standard PC. The system was developed with consultation and help from the Irish Deaf Society.

2. Modeling Sign Language

Sign language is often regarded as the most structured of the various gesture categories [3]. This work follows the basic ideas of a segmental model known as the Movement-Hold model [2]. The movement-hold model breaks a sign

into a sequence of segments, where a segment is either a movement or a hold. Movements are defined as those segments in which some aspect of the hand configuration changes, such as a change in hand shape, hand movement or hand orientation. Holds are defined as those segments during which all aspects of the signs configuration remain unchanged for a brief period of time. Signs are made up of sequences of movements and holds.

The system described here extracts features describing hand shape, position and movement from a standard web-cam video stream and then applies Sizemap [4][5] and Principal Component Analysis (PCA) techniques to these features to give real-time feedback on sign performance.

3. System Overview

The key feature of the system is the live sign performance feedback. As the user progresses through a lesson, a virtual teacher, who is animated using pre-recorded motion capture data, performs the sequence of signs defined in that lesson. After performing a sign, the virtual teacher monitors and gives immediate feedback on the users performance of that sign. This is achieved by tracking and analysing the users hand poses and movements through a webcam. Tracking of the hands is performed by tracking coloured gloves (see Fig. 1-a) using the Mean Shift algorithm [1]. The external contour is then extracted from the image and used as the key feature for analyzing the hand pose (see Fig. 1-c).



Figure 1. Hand External Contours Extracted from WebCam Stream

4. Hand Shape Analysis

As stated in [2], the structure of the signs can be described as the combination of the shape, position and movement of the hand. In order to give feedback on sign performance, analysis on these three sign components must be performed.

Hand shape analysis is performed by applying Size Function [4][5] pattern matching techniques on the external hand contour.

5. Implementation and Performance

The system described in this work was developed as a robust software application that can be used on any standard Microsoft Windows based PC. The vision feature extraction component of the system was built using C++ and utilizes OpenCV functions. The user interface (see Figure 2) and feature analysis engine was developed using C# and using a Direct 3D avatar to demonstrate signs to the user. The system was built using object oriented programming and is a working application that has been demonstrated live in lecture halls.

Tests were performed to calculate the total speed of the system. Results show that the average time, of feature extractions, feature analysis and feature comparison was 60 milliseconds, which results in the average frames per second being 16.7 fps.

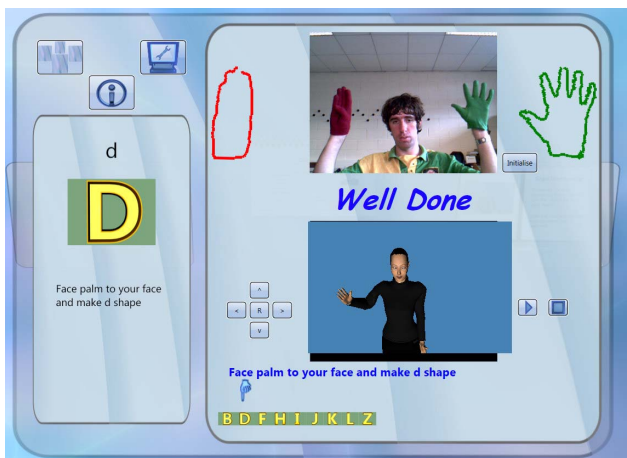


Figure 2. Front End of Sign Language Teaching Software with 3D avatar

6. The Demo

Demonstration of the application consists of 14 sample signs being performed by the virtual teacher. Before the virtual teacher performs the next sign in the lesson the user must correctly perform the current sign. The demo is in-

teractive and the author encourages interested conference attendees to use the application themselves.

The typical demonstration scenario for a user will be:

1. Put on colored gloves for hand tracking.
2. Sit in front of the WebCam and calibrate tracking.
3. Follow instructions and feedback from virtual teacher and aim to progress to the end of the current sign language lesson.

To test the accuracy of the system, an experiment was carried out in which two subjects were asked to perform the signs b, d, f, h, i, j, k, l and z from the Irish Sign Language alphabet. Each subject was asked to perform each of the 9 signs 7 times. It is to be noted that the signs for the letter j and the letter z are signs which contain both hold and movement segments. Results of the experiment show the system classifies 93.8% of correct signs as correct while only classifying 9.1% of incorrect signs as correct. The aim of the system is to give feedback on sign performance and learning, and from the results of the above experiment it can be seen the system performs at a promising level in giving correct feedback.

References

- [1] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2:142–149 vol.2, 2000.
- [2] J. R. Liddell, S.K. American sign language: The phonological base. *Sign Language Studies*, 64.
- [3] S. C. W. Ong and S. Ranganath. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):873–891, 2005.
- [4] A. Uras, C.; Verri. On the recognition of the alphabet of the sign language through size functions. *Pattern Recognition, 1994. Vol. 2 - Conference B: Computer Vision & Image Processing., Proceedings of the 12th IAPR International. Conference on*, 2:334–338 vol.2, 9-13 Oct 1994.
- [5] A. Verri, C. Uras, P. Frosini, and M. Ferri. On the use of size functions for shape analysis. *Qualitative Vision, 1993., Proceedings of IEEE Workshop on*, pages 89–96, 14 Jun 1993.