Geometry based lip reading system using Multi Dimension Dynamic Time Warping

M. Z. Ibrahim[ab]*; D. J. Mulvaney*[a]
[a]School of Electronic, Electrical and Systems Engineering, Loughborough University, LE11 3TU, United Kingdom
[b]Faculty of Electrical and Electronics Engineering, University Malaysia Pahang, 26300 Pahang, Malaysia

ABSTRACT
This paper describes an automatic lip reading system consisting of two main modules 1) a pre-processing module able to extract lip geometry information from the video sequence and 2) a classification module to identify the visual speech based on dynamic lip movements. The recognition performance of the proposed system has been assessed in the recognition of the English digits 0 to 9 as spoken by the speakers in the video sequences available in the CUAVE database. Extraction of lip geometry features was carried out using a combination of a skin color filter, a border following algorithm and a convex hull approach. The proposed method was compared with the popular `snake' technique and was found to improve lip shape extraction performance for the database studied. Lip geometry features including height, width, ratio, area, perimeter and various combinations of these features were evaluated to determine which performs the best when representing speech in the visual domain in the application of three separate classification methods, namely optical flow, Dynamic Time Warping (DTW) and a new approach termed Multi-Dimensional DTW. Experiments show that the proposed system is capable of a recognition performance of 68% just using lip height, lip width and the ratio of these features demonstrating that the system has the potential to be incorporated in a multimodal speech recognition system for use in noisy environments.

KEYWORDS:
border following; convex hull; Lip reading; lip shape

REFERENCES

1. S. W. Chin, K. P. Seng, and L.-M. Ang, "Audio-Visual Speech Processing for Human Computer Interaction," in Advances in Robotics and Virtual Reality, vol. 26, Springer Berlin Heidelberg, 2012, pp. 135-165.

2. G. Potamianos, C. Neti, J. Luettin, and I. Matthews, "Audio-visual automatic speech recognition: An overview," Issues in Visual and Audio-Visual Speech Processing. MIT Press, 2004.

3. W. Yau, D. Kumar, and S. Arjunan, "Voiceless speech recognition using dynamic visual speech features," in HCSNet Workshop on the Use of Vision in HCI, 2006.

4. A. A. Shaikh, D. K. Kumar, W. C. Yau, M. Z. C. Azemin, and J. Gubbi, "Lip Reading using Optical Flow and Support Vector Machines," in 3rd International Congress on Image and Signal Processing (CISP), 2010, pp. 327-330.

5. E. K. Patterson, S. Gurbuz, Z. Tufekci, and J. N. Gowdy, "Moving-Talker, Speaker-Independent Feature Study, and Baseline Results Using the CUAVE Multimodal Speech Corpus," EURASIP Journal on Advances in Signal Processing, vol. 2002, no. 11, pp. 1189-1201, Jan. 2002.