A lip geometry approach for feature-fusion based audio-visual speech recognition

M. Z. Ibrahim[a,b], , D. J. Mulvaney[b]
[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 a feature-fusion audio-visual speech recognition (AVSR) system that extracts lip geometry from the mouth region using a combination of skin color filter, border following and convex hull, and classification using a Hidden Markov Model. By defining a small number of highly descriptive geometrical features relevant to the recognition task, the approach avoids the poor scalability (termed the `curse of dimensionality') that is often associated with featurefusion AVSR methods. The paper describes comparisons of the new approach with conventional appearance-based methods, namely the discrete cosine transform and the principal component analysis techniques, when operating under simulated ambient noise conditions that affect the spoken phrases. The experimental results demonstrate that, in the presence of audio noise, the geometrical method significantly improves speech recognition accuracy compared with appearance-based approaches, despite the new method requiring significantly fewer features.

KEYWORDS
Lip geometry; Feature fusion; Audio-visual speech recognition; OpenCV

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

2. R. E. Bellman,
    Adaptive Control Processes: A Guided Tour. Princeton, NJ: Princeton Univ. Press, 1961.

3. J. Lee and C. H. Park,
    "Robust Audio-Visual Speech Recognition Based on Late Integration," IEEE Trans. Multimed., vol. 10, no. 5, pp. 767–779, Aug. 2008.

4. G. Bradski and A. Kaehler,
    Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, 2008.

5. S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D.
    Ollason, D. Povey, V. Valtchev, and P. Woodland,
    "The HTK Book" Cambridge University, 2006.