Shin, Jongju; **Lee, Jin**; **Kim, Daijin**
Real-time lip reading system for isolated Korean word recognition. (English) Zbl 1207.68314
Pattern Recognition 44, No. 3, 559-571 (2011).

Summary: This paper proposes a real-time lip reading system (consisting of a lip detector, a lip tracker, a lip activation detector and a word classifier) which can recognize isolated Korean words. Lip detection is performed in several stages: face detection, eye detection, mouth detection, mouth end-point detection, and active appearance model (AAM) fitting. Lip tracking is then undertaken via a novel two-stage lip tracking method, where the model-based Lucas-Kanade feature tracker is used to track the outer lip, and then a fast block matching algorithm is used to track the inner lip. Lip activation detection is undertaken through a neural network classifier, the input for which being a combination of the lip motion energy function and the first dominant shape feature. In the last step, input words are defined and recognized by three different classifiers: HMM, ANN, and K-NN. We combine the proposed lip reading system with an audio-only automatic speech recognition (ASR) system to improve the word recognition performance in the noisy environments. We then demonstrate the potential applicability of the combined system for use within hands-free in-vehicle navigation devices. Results from experiments undertaken on 30 isolated Korean words using the K-NN classifier at a speed of 15 fps demonstrate that the proposed lip reading system achieves a 92.67% word correct rate (WCR) for person-dependent tests, and a 46.50% WCR for person-independent tests. Also, the combined audio-visual ASR system increases the WCR from 0% to 60% in a noisy environment.

MSC:

68T10 Pattern recognition, speech recognition
68T35 Theory of languages and software systems (knowledge-based systems, expert systems, etc.) for artificial intelligence

Keywords:

lip reading; two-stage lip tracking; word classifier; automatic speech recognition; audio-visual ASR

Software:

darch

Full Text: DOI

References:

[1] E. Petajan, Automatic lipreading to enhance speech recognition, in: Proceedings of Global Telecommunications Conference, Atlanta, GA, 1984, pp. 265-272.

[2] Bailly, G.; Vatikiotis-Basteson, E.; Pierrier, P., Issues in visual speech processing, (2004), MIT Press

[3] Yau, W.C.; Kumar, D.K.; Arjunan, S.P., Visual recognition of speech consonants using facial movement features, Integrated computer-aided engineering, 14, 1, 49-61, (2007)

[4] T. Saitoh, K. Morishita, R. Konishi, Analysis of efficient lip reading method for various languages, in: Pattern Recognition 19th International Conference on ICPR, 2008, pp. 1-4.

[5] Ma, W.J.; Zhou, X.; Ross, L.A.; Foxe, J.J.; Parra, L.C., Lip Reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space, Plos one, 4, 3, 1-14, (2009)

[6] K. Kumar, T. Chen, R. Stern, Profile view lipreading, in: Proceedings of IEEE ICASSP, 2007, pp. 29-432.

[7] Y. Kim, S. Kang, S. Jung, Design and implementation of a lip reading system in smart phone environment, in: Proceedings of the 10th IEEE International Conference on Information Reuse and Integration, 2009, pp. 101-104.

[8] Matthews, I.; Cootes, T.F.; Bangham, J.A.; Cox, S.; Harvey, R., Extraction of visual features for lipreading, IEEE transactions on pattern analysis and machine intelligence, 24, 2, 198-213, (2002)

[9] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of the 7th International Joint Conference on Artificial Intelligence, 1981, pp. 674-679.

[10] S. Baker, I. Matthews, Lucas-Kanade 20 years on: a unifying framework: part 1, Technical Report CMU-RI-TR-02-16, Robotics

Institute, Carnegie Mellon University, 2001.

[11] S. Baker, R. Gross, I. Matthews, Lucas-Kanade 20 years on: a unifying framework: part 3, Technical Report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, 2003.

[12] Matthews, I.; Ishikawa, T.; Baker, S., The template update problem, IEEE transactions on pattern analysis and machine intelligence, 24, 6, 810-815, (2004)

[13] Myers, C.S.; Rabiner, L.R., A comparative study of several dynamic time-warping algorithms for connected word recognition, The Bell system technical journal, 60, 7, 1389-1409, (1981)

[14] Rabiner, L.R., A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77, 2, 136-257, (1989)

[15] Dong, L.; Foo, S.W.; Lian, Y., A two-channel training algorithm for hidden Markov model and its application to lip Reading, EURASIP journal of applied signal processing, 9, 1382-1399, (2005) · Zbl 1138.68578

[16] Hinton, G.E.; Salakhutdinov, R.R., Reducing the dimensionality of data with neural networks, Science, 313, 504-507, (2006) · Zbl 1226.68083

[17] Bagai, A.; Gandhi, H.; Goyal, R.; Kohli, M.; Prasad, T.V., Lip Reading using neural networks, International journal of computer science and network security, 9, 4, 108-111, (2009)

[18] W.C. Yau, D.K. Kumar, T. Chinnadurai, Lip reading technique using spatio-temporal templates and support vector machines, in: Proceedings of the 13th Iberoamerican Congress on Pattern Recognition, Lecture Notes in Computer Science, vol. 5197, 2008, pp. 610-617.

[19] C. Rodriguez, F. Boto, I. Soraluze, A. Perez, An incremental and hierachical K-NN classifier for handwritten characters, in: Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), 2002.

[20] Hart, P.E., The condensed nearest neighbor rule, IEEE transactions on information theory, 14, 515-516, (1968)

[21] C. Neti, G. Potamianos, J. Luettin, I. Matthews, H. Glotin, D. Vergyri, J. Sison, A. Mashari, J. Zhou, Audio-visual speech recognition, Johns Hopkins University, CLSP, No. WS00AVSR, 2000.

[22] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2001, pp. 511-518.

[23] B. Froba, A. Ernst, Face Detection with the modified census transform, in: Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, 2004, pp. 91-96.

[24] M. Stegmann, R. Fisker, B. Ersboll, H. Thodberg, L. Hyldstrup, Active appearance models: theory and cases, in: Proceedings of 9th Danish Conference on Pattern Recognition and Image Analysis, 2000, pp. 49-57.

[25] Matthews, I.; Baker, S., Active appearance models revisited, International journal of computer vision, 60, 2, 135-164, (2004)

[26] Viterbi, A.J., Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, IEEE transactions on information theory, 13, 2, 260-269, (1967) · Zbl 0148.40501

[27] Duda, R.O.; Hart, P.E.; Stork, D.G., Pattern classification, (2002), A Wiley-Interscience Publication

[28] Cootes, T.F.; Taylor, C.J.; Cooper, D.H.; Graham, J., Active shape models—their training and application, Computer vision and image understanding, 61, 38-59, (1995)