

Lip Motion Recognition for Indonesian Vowel Phonemes Using 3D Convolutional Neural Networks

Maxalmina Satria Kahfi
School of Computing
Telkom University
Bandung, Indonesia
maxalmina.kahfi@gmail.com

Kurniawan Nur Ramadhani
School of Computing
Telkom University
Bandung, Indonesia
kurniawannr@telkomuniversity.ac.id

Anditya Arifianto
School of Computing
Telkom University
Bandung, Indonesia
anditya@telkomuniversity.ac.id

Abstract—Lip motion recognition is a technique for interpreting visual data that focuses on the mouth area and aims to recognize lip movement. The development of lip motion recognition is expected to be used to develop communication tools with deaf people and to automate the speech-to-text process visually. In the Indonesian language, the existence of vowel phonemes is needed to produce sounds so that words and sentences in the Indonesian language can be formed. This paper proposes a model that can recognize Indonesian vowel phonemes (/a/, /i/, /u/, /e/, and /o/) in lip movements. We proposed a model that uses 3D Convolutional Neural Networks. The data in this paper were processed by resizing into 112x56 pixel resolution then, proceed to the data augmentation by reversing the data horizontally and add blur to the data. The results of the testing of the vowel phoneme recognition model on lip motion show the highest accuracy rate of 84%.

Keywords—lip motion recognition, convolutional neural networks, Indonesian vowel phonemes, 3D convolution