Edge Computing System applying Integrated Object Recognition based on Deep Learning

Kwihoon Kim*, Sewon Oh**

*KNUE, 250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do, 28173, Korea

** ETRI, 218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, Korea

kimkh@knue.ac.kr, sewonoh@etri.re.kr

Abstract— *With the development of personal broadcasting such as YouTube, the demand for editing the video he filmed himself is steadily increasing. Traditional media video editing solutions help edit the results of the filming. However, there is a problem that requires a lot of editing time, as people usually edit videos that are from tens to hundreds of times longer than the final result of a personal broadcasting after filming. To overcome this problem, this paper automatically classifies images with specific scenes in the whole media image editing process, and secondly proposes automatic media editing solution technology in which people intervene. In particular, personal broadcasting focuses on the use of images that include characters, specific objects, and cue sign gestures among the entire. While the existing deep learning techniques such as faces, objects and gestures are advanced, integrated recognition technologies that simultaneously deal with special requirements for editing videos are still in the early stages of research. In this paper, the automatic composite recognition technology for editing video based on deep learning is proposed. The proposed technology was implemented with python and tensorflow software based on edge computing equipment. Using actual youtube videos, it took 0.1 second to process five-person recognition, 63-food recognition, or cue sign recognition using clapping or V poses at the same time. The recognized results are divided into timestamps of the entire movie, recognition results, and locations of objects on the screen, and are output to the json file. In addition, this solution was developed on an edge computing in order to increase real-time reliability. We expect to provide automatic video editing based on perceived json results as well as shorter editing times based on this implementation.*

Keywords — Automatic media editor, Object recognition, Context recognition, Deep learning, Edge computing

Kwihoon Kim is currently an Assistant Professor in the Department of Artificial Intelligence Convergence Education, Korea National University of Education (KNUE), South Korea. He received the B.S, M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea in 1998, 2000 and 2019, respectively. He worked in LG DACOM 2000~2005. From 2005 to 2020, he was a Principle Researcher with Electronics and Telecommunications Research Institute (ETRI). He is an editor and rapporteur of ITU-T SG11 since 2006. His interested fields are Fog/edge computing, Internet of Things, 5G/IMT2020, deep learning, machine learning, reinforcement learning, GAN and knowledge-converged intelligent service.

Sewon OH is currently a Principal Researcher working for Knowledge-converged Super Brain(KSB) Convergence Research Department, Electronics and Telecommunications Research Institute(ETRI), South Korea. He received the B.S.(1999) and the M.S. degrees(2001) from Pohang University of Science and Technology (POSTECH), and received the Ph.D.(2018) from Chungnam National University (CNU), Daejeon, South Korea, respectively. Since 2001, he has been involved in several research projects on software platform (such as RFID Event Management, USN Middleware, IoT Platform) which integrates legacy applications with various data resources and sensors. His recent interested areas include anomaly detection, machine learning application, and knowledge-converged intelligent service.