論文 / 著書情報
Article / Book Information

Title	Video Semantic Indexing and Localization
Authors	Koichi Shinoda
Citation	5th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, vol. 140, no. 4, p. 3009
Pub. date	2016, 11

MONDAY AFTERNOON, 28 NOVEMBER 2016 CORAL 3, 3:15 P.M. TO 5:35 P.M.

Session 1pSPb

Invited Papers

3:15

3:35

1pSPb2. Video semantic indexing and localization. Koichi Shinoda (Dept. Comput. Sci., Tokyo Inst. of Technol., 2-12-1 W8-81 Okayama, Meguro-ku Tokyo 152-8552, Japan, shinoda@cs.titech.ac.jp)

Nowadays Internet traffic has been largely occupied by consumer video but most of them are not accompanied with text tags for search. Hence, video semantic indexing, which extracts visual concepts such as objects, scenes, and actions directly from video contents, has been intensively studied. Fundamentally, this task consists of two problems: localization and recognition. While until recently these two problems have been studied independently, emerging end-to-end deep learning techniques using convolutional neural networks (CNNs) and recurrent neural networks (RNNs) offer effective ways to solve them simultaneously. These techniques are deeply related to spoken word detection techniques in the speech field. In this talk, we overview the recent progress in this area and discuss potential directions for future research.