論文内容の要旨

Video Affective Content Modeling for Analysis, Search, and Editing

(情動的印象のモデル化に基づく映像の解析、検索と編集)

氏名　入江　豪

（本文）　The goal of video content modeling is to realize methods that enable computers to understand video content like humans do. Recent researchers focus on “bridging the semantic gap” issue and have explored models for capturing the relationship between semantic meaning and low-level features of videos. On the other hand, human understanding does not only rely on semantic meaning of videos, but also affective evaluation. The affective preferences play an important role in video selection. Affective scenes will often be search targets, so catching the emotional highlights are clearly interesting. Moreover, affect has a fundamentally important impact on the viewers' attention and memory. Therefore, modeling affective content of videos will contribute to extending the potential of video content modeling and improving the performance of existing applications such as video analysis, search and editing.

Motivated by these observations, this thesis proposes a series of video affective content modeling methods. Essential questions are: (a) what features of videos cause affects/emotions to the viewers? (b) how can we model various categories of affects based on the features? (c) how can we apply video affective content modeling to the real world applications? This thesis explores answers to these questions on each of the following five different scenarios:

- Interest-oriented video search ranking: Today, consumers are required to search large-scale user generated video (UGV) databases in order to find out interesting UGVs, but this is not an easy task. In Chapter 2, we propose Degree-of-Edit (DoE), a novel content-based UGV search ranking measure that assists users to find interesting UGVs. The core concept of DoE is based on the idea that “a highly edited video is more interesting”. For each UGV registered in a video database, our method first estimates its DoE score (level of editing) based on audio-visual features, and then ranks UGVs depending on estimated DoE scores. We show the effectiveness of our method through a series of experiments on over 70,000 UGVs in the context of UGV search.

- Impressive face key-frame extraction: Home videos contain often imagery of people, so human faces are important. Provided an application context where the objective is to extract impressive keyframes from videos, we propose a method to extract “good shot of the person(s)” from a home video. We investigate the influence of facial parameters on the subjective impression that is created when looking at photographs containing people. Based on the findings from the user study, an impression-oriented image ranking function is designed. We evaluate its effectiveness in terms of correlation between the ranking generated by our ranking function and that by ground truth data.

- Joyful, sad and exciting video segments extraction: Pleasure (e.g. levels of joy or sadness) and arousal (level of excitement) are two key factors for representing human's affect. In Chapter 4, we propose a method to extract joyful, sad, and exciting video segments. The key idea of our approach is that emotional audio events (EAEs) are closely related to viewers' affects. The proposed method first detects EAEs, and then estimates levels of joy, sadness, and excitement of video segments by utilizing correlations between EAEs and affects. We show the effectiveness of our method by conducting several experiments.

- Scene classification into basic emotions: Basic emotions indicate the elemental emotion categories and can be blended together in various ways (secondary emotions) to form the full spectrum of human emotional experience. In Chapter 5, we focus on classifying video scenes into basic emotion categories. There are two main issues to be considered: one is “how to extract features that are strongly related to viewer's emotions”, and the other is “how to map the extracted features to the emotion categories”. For the former issue, we propose affective audio-visual words (AAVW), efficient representation of audio-visual features that strongly related to viewer's affects. For the latter issue, we present a model named latent topic driving model (LTDM), that considers the relationship between latent topic of video scenes and human affects. We show the promising performance of the method that combines AAVW with LTDM.

- Emotionally impactful trailer generation: Since a trailer is ad of a movie, it is expected to be impactful to viewers. In Chapter 6, we explore an automatic movie trailer generation based on video affective content modeling approach. We propose a method to extract impressive speech and video segments. Furthermore, we propose a computational method for estimating affective impact of a shot sequence, and provide an algorithm to arrange a set of shots by maximizing the affective impact. A series of experiments show effectiveness of our method.

According to these scenarios and proposed methods, this thesis will contribute to extending video content modeling and to improving performance of its applications. It is expected that this thesis also provides a promising direction of future video affective content modeling research.