

答辩后归档材料:

硕士学位论文题目: 基于用户关注空间与注意力分析的音视频情感理解
研究

答辩时间: 2008年5月12日

答辩委员会主席: 焦健彬教授

答辩委员会成员: 程健副研究员 卿来云副教授

毕业时间: 2008年7月

研究方向: 多媒体技术

导师: 黄庆明教授

毕业去向: 甲骨文(中国)软件系统有限公司

曾获奖项: 研究生院三好学生

毕业论文的摘要

摘 要

音视频数据作为一种动态、直观、形象的数字媒体承载了大量丰富的语义信息,越来越多地出现在各类信息服务和应用场合,如何自动而高效地从大量音视频数据中挖掘出有效信息,理解其中的语义内涵已成为当前基于内容的音视频分析领域一个前沿问题。

目前尽管一些研究机构在音视频情感分析、内容摘要等问题中针对“语义鸿沟”(底层媒体特征与高层语义信息之间的差异)展开大量研究工作,但却忽略了将人类的感知因素考虑到计算机分析系统中。本文通过对音视频文件中蕴含的情感内容进行分析从而获得该文件所表达的主题并以摘要等形式将这些用户关注信息传递给观众,实现选择性快速浏览整个音

视频文件及个性化定制其关注内容的要求。本课题所研究的情感分析技术既包含面向多种视频的通用解决方案如在访谈节目、电影、广播剧等类型上的音视频分析，也有面向特定视频类型的特殊分析方法如在体育视频中常见的精彩事件检测与排序技术，即通过引入相关领域知识来检测体育比赛中的精彩事件并按照其激烈程度排序来方便用户快速浏览其感兴趣内容。

本文首先提出一种新颖的基于二叉层次型结构与分类器选择的音频分类算法将媒体文件中的主要声音类型进行有效分类，然后依据心理学相关理论将音视频文件中影响观众注意力的主要因素定义为用户关注空间，分别从视觉、听觉、时序等角度使用不同的情感特征在该空间上对用户注意力进行表示，从而在音视频底层特征与高层语义概念之间建立有机过渡，为缩小“语义鸿沟”提供一条可行的途径。

在视觉方面我们使用平均运动向量(Average Motion Vector)和镜头切换率(Shot Change Rate)对用户视觉关注度进行表示；在听觉方面使用经音频分类获得的主要声音类型的短时平均能量、音调、过零率等特征对用户听觉关注度进行表示。我们设计了顺序决策融合算法融合视觉与听觉关注度，生成关注度时序变化曲线并获得此文件的精彩摘要片段。最后使用支持向量回归模型(Support Vector Regression Model)并引入相关反馈机制来实现用户个性化的精彩片段排序。

将用户关注空间与注意力分析方法引入音视频情感理解领域并在此基础上实现精彩摘要、排序等技术，是通过构建中层特征来描述音视频文件中蕴含的语义信息，获取反映人类认知机理的情感信息，在音视频底层特征与高层语义概念之间建立有机的过渡。本文的特点是采用统计方法使计算机建立起符合人类认知行为的用户关注模型并结合相关反馈技术，对音视频内容进行类人理解，使分析结果更贴近用户要求，最终使音视频情感分析技术更符合用户个性化特点。因此，我们采用基于用户关注空间与注意力分析的多模态融合技术对音视频情感内容进行理解。

关键词：用户关注空间；注意力分析；关注模型；音视频摘要；精彩排序

User Attention Analysis Based Audio-Video Affective Content Understanding

Zheng Yijia

Advised By Huang Qingming

Since the amount of the multimedia data grows rapidly, the audio and video content analysis

technology has been widely investigated. How to mine specific semantic information from the audio and video data automatically and effectively is becoming a hot research point.

This paper describes our process on how to get the proper subject of the audio/video files and generate the summary of them by analyzing the affective content of these files. Also, its main content has been studied in the work of affective analysis and attention analysis. Our work of affective content analysis not only focus on the research of the general audio/video types such as talk show videos, movies, radio plays and so on, but also face to the summarization and highlight ranking on the specific audio/video type such as sports videos.

Nowadays although some approaches have been investigated on audio/video affective content analysis, there is still no favorable solution that considers the human perception factors in the audio/video computing systems.

In this paper, we first define the factors that affect the human perception in audio/video files as “user attention space” which have three subspaces in aural, visual and temporal based on the psychological theories. The user attention space and attention modeling based affective content analysis approach present an effective way of establishing the bridge between the low-level audio/video features and the high-level semantic content. Then we use the average motion vector and shot change rate features to describe the visual modality, and the audio energy, pitch, zero cross rate etc. are used to describe the aural modality. An Ordinal-Fusion strategy is exploited to construct the visual-aural model and give the user attention change curve. We could find the highlight segments from this curve. Last, relevance feedback mechanism is employed to make the ranking result more suitable for human personalization.

The method that introduces the user attention into the affective content analysis field could effectively realize the summarization, highlight ranking etc. work. It combines the visual-aural attention to construct the mid-level affective features for semantic mining.

This paper mainly investigates the nonlinear user attention modeling approach and the relevance feedback technology. Compared with existing methods, our proposed approach is based on the changes of human attention while watching TV rather than the simple content changes of videos, which is more consistent with human understanding.

Keywords: User Attention Space, Attention Analysis, Attention Model, Audio-Video Summarization, Highlight Ranking

[以上材料同时发送给冯春丽 clfeng@jd.com](mailto:clfeng@jd.com)