

氏 名 FENG Yunlong

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 1720 号

学位授与の日付 平成26年9月29日

学位授与の要件 複合科学研究科 情報学専攻
学位規則第6条第1項該当

学位論文題目 Human Interactive System for Networked Video Streaming

論文審査委員 主 査 教授 計 宇生
教授 山田 茂樹
教授 曾根原 登
准教授 福田 健介
准教授 CHEUNG GENE
准教授 高橋 桂太 名古屋大学

論文内容の要旨
Summary of thesis contents

In this thesis, we discussed the application and development using human interactive technology and visual saliency map analysis for the networked video streaming. As we know, for current video streaming systems, there are more and more devices to record videos, more and more powerful servers to process encoding, larger and larger databases to store them, and also faster and faster network speed to transmit video packages to viewers. However, viewers still feel discontented on passively transmitted videos, for preferring a real time interactive system. In this thesis, we developed a Human Interactive System to make network video streaming more efficient, not only to benefit the network traffic, but also to adapt video content based on viewers' interests and real-time requirement.

For interactive media applications, eye gaze is now used as a content adaptation trigger, such as customized advertisement in video, and bit allocation in streaming video based on region-of-interest (ROI). The reaction time of a gaze-based networked system, however, is lower-bounded by the network round trip time (RTT). Furthermore, only low-sampling-rate gaze data is available when commonly available webcam is employed for gaze tracking. To realize responsive adaptation of media content even under non-negligible RTT and using common low-cost webcams, we propose a Hidden Markov Model (HMM) based gaze-prediction system that utilizes the visual saliency of the content being viewed. Due to the strong prior of likely gaze locations offered by saliency information, accurate runtime gaze prediction is possible even under large RTT and using common webcam. And region-of-interest (ROI) bit allocation is real-time performed based on predicted future gaze location to adapt the video content for reducing bit size and perceived visual quality.

By the latter half of the thesis, a novel method using saliency map for detecting video busyness, which is called visual attention deviation (VAD) is improved, to develop the gaze prediction system. We all know that analyzing human perception is time-consuming, for subjective evaluation is necessary. The experiment always cost much time and need a lot of subjects. However, saliency map analysis is able to find out the most salient region by low-level figures using given picture or frames. Based on the existed analysis methods, we proposed our metric VAD to detect video busyness by analyzing the saliency regions along the whole timeline with the presented HMM in first half work. Through experiments, we show that VAD was able to detect the video busyness by analyzing the saliency objects transition probability, using trained HMM. And our comparison results show that it's much sensitive than other metrics, and the most important is that VAD result is matching subjective evaluation, which means it's reflecting human perception while the video is play-backed.

博士論文の審査結果の要旨
Summary of the results of the doctoral thesis screening

本博士論文は「Human Interactive System for Networked Video Streaming (ネットワーク上における映像ストリーミングのための対話型システム)」と題し、視聴者の注視点と映像の視覚的顕著性マップを利用した対話型映像ストリーミングシステムに関する研究内容である。

従来の映像ストリーミングシステムでは、視聴者が受動的に映像の配送を受けるのみであるため、不満を感じることもある。一方で、視聴者の視覚的注意に関する予測ができれば、視聴者の要求と視聴好みに合わせた映像コンテンツの提供が可能になる。また、視聴者の目の注視点をカメラで追跡する技術が成熟し、すでに商用のレベルに達している。本論文は、視聴者の注視点の予測ならびに映像の視覚的顕著性マップを利用して、人間の受ける視覚的品質を保ちながらネットワーク帯域を節約し、視覚的注意に基づく適応的映像ストリーミングを可能にする方法の提案に関するものである。

論文は4章から構成され、第1章の研究の背景、関連研究に続いて、第2章と第3章では、注視点の予測に基づくビット割当てとそれによるビット削減方式、視覚的注意変動指標とその応用に関する研究内容と結果について述べている。

第2章では、視聴者の目の注視点の予測に基づくビット割当ての方法について述べている。ここでは目の注視点をコンテンツ適応のきっかけとし、予測された注目画像領域 (Region Of Interest, ROI) により多くのビットを割当てていく。そのためにはまず、目の動きを追跡状態と不規則な動きの2つの状態に分けて、視覚的顕著性マップを利用して状態間の遷移パラメータを隠れマルコフ過程として推定を行う。さらに、追跡状態における視点の予測を行い、注目画像領域を決める。注目画像領域以外の領域のビット数を削減することによって、ネットワークの帯域を節約する。このようにできる対話型システムについて評価を行った結果、視覚効果が変わらない条件で、ラウンドトリップ時間 (RTT) が200ミリ秒の場合でも最大で29%のビット数の削減を確認した。

第3章では、顕著性マップを使った視覚的注意の解析として、映像の「忙しさ」 (Busyness) を表すための視覚的注意変動指標 (Visual Attention Deviation, VAD) の提案とその応用について述べている。人間の感覚を解析するための主観評価には時間がかかるため、画像フレームに対する解析で得られる顕著性マップを利用することを考え、VADを使って映像の忙しさ、すなわち動きの度合いと変化の速さを表し、場面の変化を検出する。ロゴ挿入という応用における主観評価テストによって、VADは他の既存指標よりも映像の忙しさをより正確に表すことができることを確かめた。

また、第4章では結論として、本論文のコントリビューションと考察についてまとめ、今後の研究課題を提示した。

なお、研究成果として、出願者は主著でIEEEトランザクション誌論文を1篇、査読付き国際会議論文を3篇、その他の学会発表を6件行っている。

以上を要するに、本論文は視聴者の注視点に対する予測方法と映像の忙しさと場面の変化を検出するための指標を提案し、その有効性を示したものであり、ネットワーク上の映像ストリーミングの効率性と視聴者の受ける視覚的品質を高めるための技術の実現に貢献するところが大きい。よって、本論文は博士の学位請求論文として合格と認められる。