



HOKKAIDO UNIVERSITY

Title	複数特徴を利用した映像意味解析とリランキングの導入による検索の高精度化に関する研究〔論文内容及び審査の要旨〕
Author(s)	吉田, 壮
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12197号
Issue Date	2016-03-24
Doc URL	https://hdl.handle.net/2115/61838
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Soh_Yoshida_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士(情報科学) 氏名 吉田 壮

学 位 論 文 題 名

複数特徴を利用した映像意味解析とリランキングの導入による検索の高精度化に関する研究
(Accurate Video Retrieval Using Multiple Feature-Based Video Semantic Analysis and Reranking)

本論文は、複数特徴を利用した映像意味解析とリランキングの導入による検索の高精度化に関する研究の成果をまとめたものである。

高速・大容量な通信回線の普及に伴い、インターネットによる映像配信サービスの利用が増加している。このような状況の中、ユーザが望む映像を効率的に検索し閲覧するための技術開発が急務とされている。一般に、映像検索の結果は、予め映像にメタデータを付与し、それとクエリとして与えるキーワードとの適合度に基づく順位付けに従い提示される。このとき、Web 上の映像は極めて膨大であることから、手動によるメタデータの付与は現実的でなく、自動化が必須である。現状の研究では、1) 映像を同一意味内容のまとまりを持つ区間に分割した後、2) 各区間に何が映っているのかを識別することでメタデータの付与を可能とする。これにより自動で得られたメタデータとクエリを比較して適合度を算出し順位付けすることで検索結果が提示される。しかしながら、特に、2) において行われるメタデータの自動付与に誤りが生じる場合があり、そのような場合に、順位付けに影響を与え、検索精度の劣化を招き、ユーザが望む映像が下位に配置されてしまう。

この問題を解決するための研究として、3) 順位付けにおいて下位に存在する映像の中でユーザが望む映像の順位が高くなるように並べ替えるリランキング手法が存在する。先に示す 2) の手法の精度がある程度高い場合は、このリランキング手法の適用により、望む映像の順位を上げることが可能であり、メタデータ付与の誤りによる検索の精度劣化の低減が期待できる。

このような背景の下、1) および 2)、さらに 3) について、様々な研究が行われている。以下に、各々の従来研究とその問題点を示す。まず、1) 映像をシーンに分割する従来手法について説明する。シーンは「同一もしくは類似する内容を表すショットが連続する区間」として、ショットは「1 台のカメラにより切れ間なく撮像される区間」として定義され、映像のシーン分割は、隣接するショットの意味内容が同一もしくは類似するかを判定して統合することで可能となる。シーンの意味的なまとまりに関する定義は映像のジャンル毎に異なり、各々の手法で個別に設定される場合が多い。代表的な従来研究に、映像の構造に注目したシーン分割手法が挙げられる。これは、編集映像に現れる、類似するショットの繰り返し構造を解析しシーン境界を判定する手法である。しかしながら、シーン境界毎に特徴の変化が大きいジャンルの映像では、画像、音響という特徴を利用するだけでは、映像が有する繰り返し構造の検出は困難となる問題が存在した。次に、2) 映像の各シーンのキーフレーム画像に対して、機械学習を利用し、各々の撮像物体の識別を行うことで、メタデータの付与を可能とする。これまで、画像特徴と撮像物体との対応から識別器を構築する手法が提案されている。しかしながら、背景の多様性や同時に撮像されている他の物体の影響から、撮像物体の識別を行う上で雑音となる特徴が含まれ、識別精度が低下する問題が存在した。最後に、3) リランキング手法は、キーワード検索で得られた映像間の類似性に基づいてその実現を可能とする手法が提案されてきた。一方、映像配信サービスから提供される映像には、映像毎に関連リンクや付随するタグ等、映像の特徴以外にも映

像間の関係を把握する上で有益な情報が含まれている。しかしながら、これらを活用してリランキングの精度を改善する手法はこれまで提案されていない。

そこで、本論文ではまず、1) の問題の解決を目指し、映像の複数特徴から得られるシーン境界の判定結果を協調的に用いるシーン分割手法の提案を行っている。提案手法では、各ジャンルの中でも映像構造が明確に存在する歌謡番組に焦点を当て、例えば、「司会区間」の次に「歌唱区間」が続き、それらが番組内で繰り返し現れるという歌謡番組固有の構造を利用する。この構造に基づいてショットの階層クラスタリングを行い、その結果から取得される有向グラフのカット辺を探索する。なお、ショットの階層クラスタリングを行う際には、歌謡番組に注目した画像、音響特徴を導入する。画像特徴については、シーンが遷移するとき、同時に各シーンの中心となる出演者も変化することを考慮し、新たに出演者の変化を検出する特徴を顔認識法の結果から定義する。各々の特徴から得られる結果を統合する協調処理により歌謡番組固有の繰り返し構造を検出することで、シーン分割の高精度化を実現する。次に、本論文では、2) の高精度化を目指し、撮像物体の自動選定処理を導入した識別器の提案を行っている。具体的に、Support Vector Machine (以降、SVM と呼ぶ) を用いた識別超平面を算出し、特徴空間における特徴ベクトルと超平面との距離に基づいて、撮像物体の存在する領域を内包する矩形を自動検出する処理を提案する。提案手法では、背景領域に存在する特徴を除去することで、撮像物体の領域に限定した特徴抽出を可能とし、新たな識別超平面を算出する。これにより、背景や他の物体による悪影響を抑制した物体識別が実現される。最後に、本論文では、3) を実現する新たな手法として、映像特徴とタグ情報の両者の類似性に注目したリランキング手法の提案を行っている。提案手法では、まず映像のリンク関係を基に、映像とタグそれぞれの類似性を表現可能なネットワークを構築し、そのネットワーク上において、映像とタグ、各々の類似性を伝播させる。さらに、キーワード検索による順位が、類似性の高い映像ノード間で近くなるように目的関数を与え、その最小化問題を解くことで、ネットワークの整合性を保つようなリランキングが可能となり、望む映像が上位に配置される映像検索が実現される。

以下に本論文の構成を示す。第 2 章では、映像意味解析とリランキングに関する従来手法について説明する。さらに、現状の問題点と本研究が解決すべき課題を明確にする。第 3 章では、映像特徴の協調利用による映像の構造に注目したシーン分割手法の提案を行う。歌謡番組固有の構造を明確にするとともに、複数特徴を協調的に用いることで、その構造を高精度に検出可能であることを実験により確認する。さらに、その構造に従うことでシーン分割の精度向上が可能となることを確認する。第 4 章では、物体領域の自動選定処理を導入した SVM に基づく物体識別手法の提案を行う。まず、SVM から得られる識別超平面の特性を利用して撮像物体の領域を自動で選定可能であることを実験により確認する。次に、選定した領域のみから抽出した特徴に基づいて更新した識別器を用いて識別精度の向上が可能となることを確認する。第 5 章では、映像特徴とタグ情報の類似性の両者を考慮したネットワーク構造に基づくリランキング手法の提案を行い、有効性を実験により示す。最後に、第 6 章において、本研究の成果について要約し、論文全体のまとめとする。

以上を要約すると、本論文は、複数特徴を利用した映像の構造に注目したシーン分割手法と物体領域の自動選定に基づく撮像物体の識別手法の提案による映像意味解析の改良、さらには、リランキング手法の新たな提案により映像検索の高精度化を実現している。また、本手法は実データに適用した実験を行うことにより、その有効性を示している。