



Title	放送映像の意味コンテンツ解析に関する研究
Author(s)	河合, 吉彦
Citation	大阪大学, 2010, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/57508">https://hdl.handle.net/11094/57508</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、<a href="https://www.library.osaka-u.ac.jp/thesis/#closed">大阪大学の博士論文について</a>をご参照ください。

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名	河 合 吉 彦
博士の専攻分野の名称	博士(工学)
学位記番号	第 23841 号
学位授与年月日	平成22年3月23日
学位授与の要件	学位規則第4条第1項該当 工学研究科電気電子情報工学専攻
学位論文名	放送映像の意味コンテンツ解析に関する研究
論文審査委員	(主査) 教授 馬場口 登 (副査) 教授 滝根 哲哉 教授 北山 研一 教授 小牧 省三 教授 三瓶 政一 教授 井上 恭 教授 河崎善一郎 教授 鶴尾 隆 教授 溝口理一郎 講師 新田 直子

### 論文内容の要旨

本論文は、放送映像の意味コンテンツ解析に関する研究成果をまとめたものであり、以下の6章により構成される。

第1章は序論であり、本論文の背景となる技術分野に関して現状を述べ、本研究の目的を明らかにした。

第2章では、本論文が対象とする放送映像の構造について説明するとともに、放送映像の意味コンテンツ解析における技術要素である、放送映像のセグメンテーション手法、放送映像の要約手法、放送映像からの高次特徴検出手法のそれぞれについて説明した。また、各手法に関する既存研究を示し、技術的な課題を明らかにすることによって、本研究の位置づけを明確にした。

第3章では、カメラの切り替え点であるショット境界を検出することにより、連続的なデータである放送映像を、意味のある小さな時間単位に分割する手法について論じた。ショット境界検出は、映像解析における最も基本的な処理のひとつであり、高い検出精度と高速な処理の両方が求められるが、既存手法ではこれらを両立させることは困難であった。そこで本論文では、各フレームに対して画像特徴量を逐次的に算出し、「ショット境界でないと判定した時点で残りの処理を省略することにより、高速かつ高精度にショット境界を検出可能な手法を提案した。また、特殊な切り替えであるディジタルビデオ効果を、モデルを利用して検出する手法を提案した。実際の放送映像に対する実験により、本手法の有効性を確認した。

第4章では、元の放送映像から短い要約映像を生成する手法について議論した。具体的には、次の2種類の手法を提案した。第一の手法は、要約対象の番組に対応する電子番組表を利用する手法である。電子番組表に記載される紹介テキストが、番組の見どころを短く紹介する要約情報であることに着目し、紹介テキストの各文に対応する映像区間を抽出して要約映像を生成した。映像区間は、番組音声の書き起こしであるクローズドキャプショ

ンと紹介テキストの類似度に基づいて決定した。類似判定にはベイズ信頼度ネットワークを利用した。第二の手法は、紹介テキストが入手できない番組に対応するためのものであり、電子番組表の紹介テキストで一般的に用いられる特徴的な言い回しに着目した手法である。通常、紹介テキストでは、視聴者の興味を惹くため、具体的な数値や割合を示したり、疑問を投げかけたりといった表現が用いられる。そこで、番組ナレーションにおいても、これらと類似した表現が用いられる場合は同様に、制作者が視聴者の興味を惹くよう意図したシーン、つまり、見どころである可能性が高いと仮定して、それらを要約映像に利用した。文の表現は、過去の様々な番組の電子番組表からアダプースト法を利用して学習した。実際の放送映像に本手法を適用し、有効性を検証した。

第5章では、放送映像からオブジェクトやイベントなどの高次特徴を検出す手法について検討した。本手法は、正解が付与された学習データに基づいて高次特徴を検出すものであり、学習データの変更によって様々な高次特徴に対応できる手法を目指した。既存研究においては、算出された画像特徴ベクトルをサポートベクターマシンによって識別していたが、本手法では、学習や検出に要する計算コストを大きく軽減できるランダムフォレスト法を利用した。しかし、従来のランダムフォレストは、正例と負例の数に大きな偏りがあると、識別精度が著しく低下するという問題があるため、高次特徴検出に適用することは困難であった。そこで本論文では、偏りのある不均衡なデータセットに対しても識別精度を低下させない新たな学習アルゴリズムを提案した。高次特徴を検出すための特徴ベクトルについては、局所領域における輝度勾配を考慮した局所特徴と、色やテクスチャなどの大域特徴を組み合わせることによって算出した。評価用データセットを利用した実験により、本手法の有効性を確認した。

第6章は結論であり、本研究で得られた成果を統括した。

## 論文審査の結果の要旨

画像・映像情報の持つ意味、すなわち画像や映像の意味コンテンツを解析することは、メディア情報処理における究極の研究課題の一つである。特に、映像という時間軸を持つ大規模データに対し、意味コンテンツを解析することは、映像の高速検索、要約、二次コンテンツ作成などの要素技術として重要である。テレビ等で放送された映像からイベントなど意味コンテンツを自動的に抽出できれば、意味的なタグやアノテーションを自動的に付与できるため、近年構築が進められている大規模映像アーカイブの検索に多大なる効用をもたらすことになる。映像の意味コンテンツ解析には、時間軸上でのセグメンテーション、意味コンテンツの検出と認識、その結果に基づく映像の表現変換など様々な課題があるものの、確立された解決法は未だに見出されていない。

本論文は、放送映像の意味コンテンツ解析の手法として、映像データを基本単位に分割するセグメンテーション、元の映像より短い映像表現を作成する映像要約、映像データからイベントやオブジェクトの出現する領域を得る高次特徴抽出についての考察をまとめたものである。主たる研究成果を要約すると以下の通りとなる。

(1) 映像データをカメラのオンオフ区間に相当するショットに分割する手法を提案し、90%を超える分割精度を得ている。提案手法は、カット、ディゾルブ、フェード、ディジタル映像効果など様々なショット境界に対し、逐次的画像特徴抽出に基づく統一的な方式で境界を検出することが特徴であり、トレードオフ関係にある分割精度と処理速度の両面から既存手法より優位であることを実験的に検証している。

(2) 電子番組表 EPG なる外部メタデータと映像に同期して与えられるクローズドキャプション（音声のranscript）との類似性に基づき、EPG の文に対応する映像セグメントを抽出、連結することにより映像要約を生成する新規手法について考察している。直接、番組に対応する EPG が存在する場合のみならず、EPG 一般に用いられる特徴的な表現に着目することにより EPG が存在しない場合にも対応できる応用範囲の広い手法である。専門家の制作した映像との比較を通して、番組内容を短時間で視聴者に伝えるトレー型の映像要約作成の基礎技法として本手法が有効であることを明らかにしている。

(3) 画像の信号的特徴よりも高次な特徴であるオブジェクトやイベントを放送映像より検出する手法を提案、実現している。本手法では、ランダムフォレストなる学習アルゴリズムを、正負例が不均衡な場合に動作するように改良したものを導入し、検出コストと検出精度が良好であることを実験的に確認している。

以上のように本論文は放送映像の意味コンテンツ解析に関する数多くの有用な知見を与えており、情報通信工学、特にメディア情報工学の発展に寄与するところが大きい。よって本論文は博士論文として価値あるものと認める。