

Title	Semantic Content Analysis of Broadcasted Sports Videos with Intermodal Collaboration
Author(s)	新田, 直子
Citation	
Issue Date	
oaire:version	VoR
URL	https://hdl.handle.net/11094/547
rights	
Note	

Osaka University Knowledge Archive : OUKA

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

Osaka University

氏名	新田直子
博士の専攻分野の名称	博士(工学)
学位記番号	第 17919 号
学位授与年月日	平成 15 年 3 月 25 日
学位授与の要件	学位規則第 4 条第 1 項該当 基礎工学研究科情報数理系専攻
学位論文名	Semantic Content Analysis of Broadcasted Sports Videos with Intermodal Collaboration (インターモーダル協調による放送型スポーツ映像の意味内容解析)
論文審査委員	(主査) 教授 宮原 秀夫 (副査) 教授 萩原 兼一 教授 竹村 治雄 教授 馬場口 登

論文内容の要旨

本論文は、放送型スポーツ映像を対象とした意味内容解析に関する研究をまとめたものであり、次の 5 章から構成されている。

第 1 章は緒論であり、マルチメディア情報であるテレビ映像の意味内容解析の重要性とその問題点について触れ、本研究の目的及び位置付けについて述べている。

第 2 章では、スポーツのテレビ番組、スポーツの試合の 2 つの観点からスポーツ映像の構造について議論し、両構造をもとに、スポーツ映像の意味内容記述モデルを提案している。このモデル化により、スポーツ映像の意味内容記述に際して必要な情報を明らかにした上で、映像内容の実際の記述に用いられる MPEG-7 との関わりについて述べている。

本論文の後半部分では、第 2 章で提案した意味内容記述の自動獲得法について議論し、映像を構成するクローズドキャプションと呼ばれるテキスト情報と画像情報の相互的な利用による手法を提案している。

まず第 3 章では、スポーツ映像中のプレイシーンに対し、その中で行われたプレイ、またそのプレイを行った選手に関する自動記述生成法を提案している。まずクローズドキャプションと呼ばれる言語ストリームから、キーワード列探索によりプレイ・選手に関する情報を獲得する。次に獲得した情報を付与する映像位置を決定するため、画像ストリームに対するテンプレートマッチングによりプレイシーンを抽出する。最後に両ストリームの時間的同期をとることによって、映像内の各プレイシーンに対するプレイ・選手に関するテキスト記述の自動付与を実現する。本手法を実際のスポーツ映像に対し適用した結果、アメリカンフットボール映像 2 本に対し再現率 86%、適合率 95%、野球映像 1 本に対し再現率 69%、適合率 96%で、プレイ・選手に関する記述が可能となっている。

さらに第 4 章では、映像中のストーリー構成要素に相当するセグメントに対して、意味内容を含むクローズドキャプション部分をテキスト情報として付与する手法を提案している。提案手法では、より汎用性の向上を目指し、まずクローズドキャプション中のスポーツの種類に依存性の低い特徴をペイジアンネットワークの利用により学習することでストーリー構成要素への分割を試みる。また、画像ストリーム上の特徴の利用により、映像に対して同様にストーリー構成要素への分割を行い、これらの結果の時間的対応付けにより、映像の意味内容に関するテキスト自動記

述を実現する。本手法を実際のスポーツ映像に対し適用した結果、アメリカンフットボール映像 10 本に対して再現率 92%、適合率 89%、野球映像 5 本に対して再現率 93%、適合率 94%で、相当するクロズドキャプション部分を付与することが可能となっている。

第 5 章は結論であり、本研究で得られた成果を総括すると共に、その意義、及び今後の課題について述べている。

論文審査の結果の要旨

映像メディアの内容記述書式として MPEG-7 が標準化されているものの、各応用に最適な記述枠組の設定、及び意味内容記述の効率的な獲得法の確立が重要な課題として顕在化しつつある。本論文は、スポーツ放送映像に対し、意味内容の記述の一形式を提案すると共に、インターモーダル協調と呼ぶ映像の協調解析により、意味内容に基づく記述(注釈)を自動獲得する手法に関する研究成果について論じている。主たる研究成果は以下のように要約される。

(1)対象とするスポーツに対してテレビ番組並びに試合としての構造を考慮した上で、テキストによる検索が可能な意味内容の記述形式、及び映像単位への分割法を定式化している。本記述形式の特徴は階層構造にあり、種々のスポーツに広く適用できるものとなっている。

(2)インターモーダル協調を用いた記述獲得の第一の手法として、テキスト情報からのキーワード列抽出、及び画像情報からの特徴的シーン抽出を相補的に実行し、試合進行部分へのプレイ名・選手名による注釈付け法を実現し、複数種類のスポーツ映像に対して良好な結果が得られることを実験的に示している。

(3)第二の手法は第一の手法を発展させたもので、テキスト情報に対しベイジアンネット学習によるストーリー分割を行い、画像セグメントとのマッチングに基づきストーリー単位ごとの意味内容記述法を具体化している。300 分の放送型スポーツ映像に対し、記述獲得についての再現率、適合率が共に 90%を超え、提案手法の有効性並びにドメイン非依存性を検証している。

以上のように、本論文は映像のもつテキスト情報と画像情報の協調解析に基づく意味内容解析法について多くの知見を含んでおり、映像メディア処理技術の発展に寄与するところが大きい。よって本論文は博士(工学)の学位論文として価値あるものと認める。