

D-036

抽出メタデータと字幕データを用いたテレビ番組のウェブ化とその閲覧方式 Webification of TV program and its browsing method using extracted metadata and closed captions

宮森 恒†
Hisashi Miyamori

田中 克己† ‡
Katsumi Tanaka

1. まえがき

本稿では、映像から抽出したメタデータと映像に付随する字幕データを利用して、テレビ番組などの映像を、テキスト情報とサムネイル画像から構成されるストーリーボード形式のウェブページに変換し閲覧する手法について検討する。従来、字幕データに基づきストーリーボード形式のウェブページを構成し、閲覧する手法 [1] が提案されていたが、スポーツ映像などの場合、ある時刻の字幕データと同時刻の映像とは、必ずしもその内容が一致していないという問題点があった。本手法では、映像解析によるメタデータを利用することで、より映像内容にマッチしたシーンを表すテキストを生成し、直接それらのシーンと対応付けできるという利点が期待できる。応用例としてテニス番組をウェブ化・閲覧する試作システムについて報告する。

2. ウェブ化ビデオ：映像とメタデータのウェブ化の概念

筆者らは、録画された長時間にわたるテレビ番組を、そのメタデータを利用して、内容や閲覧機能に付加価値をつけた上でウェブ形式に変換するウェブ化ビデオの概念を提案している (図 1)。

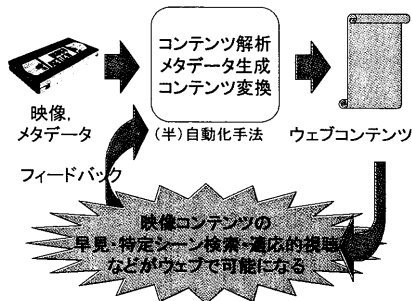


図 1: 映像とメタデータのウェブ化の概念

まず、入力となる映像およびメタデータを用いてコンテンツ解析などを行い、必要なメタデータ生成を行う。次に、それらを用いて意味内容や閲覧機能をウェブ形式のデータに変換する処理が行なわれる。得られたウェブコンテンツを適当なブラウザで閲覧することにより、従来のテレビ的視聴では得られない、早見や特定シーン検索、適応的視聴などが、ウェブページ閲覧のインタラクティブ性を活かした形で可能になる。利用者のフィード

†情報通信研究機構, NICT
‡京都大学, Kyoto University

バックは、適宜新たなコンテンツ変換処理に反映され、ウェブコンテンツは適応的に再構成されることになる。

よって、ウェブ化により、映像閲覧の効率が向上し、付加価値が与えられ、さまざまな視点からコンテンツ閲覧することが可能になると考えられる。

3. 抽出メタデータと字幕データに基づくウェブ化

本稿では、入力映像を、テキスト情報とサムネイル画像から構成されるストーリーボード形式のウェブページに変換して閲覧する。図 2 に、ウェブページへの変換に必要なメタデータの抽出手順を示す。

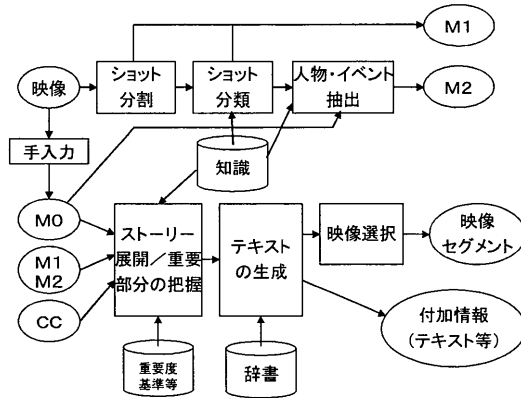


図 2: メタデータの抽出手順

まず、入力映像はショット分割され、時間方向のセグメント情報が得られる。さらに、ショット分類により、各ショットの種別を表すデータが得られる (M1)。例えば、テニス映像の場合、コート斜め上から縦方向にコートが写るようなアングルで撮影されたショット (以下、コートショットと呼ぶ)、選手や審判、観客のアップとなるショットなどの内容に分類される。分類処理は、コートなどの背景色を利用した選択手法 [2] とドメイン知識を組み合わせる方法により行っている。

次に、映像内に存在する人物・イベント抽出を行う。例えば、テニス映像の場合、コートショットに対して、動画像と音響、ドメイン知識を利用したイベント抽出手法が利用できる [3]。これにより、フォアハンドストローク、スマッシュ、サービス等のイベントが抽出される (M2)。

映像のタイトルなどの一部の情報は手入力により取得する (M0)。本稿では、テニスの試合スコアおよびその開始・終了時刻を手入力で与えることとした。スコア情報については、映像中のテロップ領域の意味を特定し、

他のメディアと関連付ける解析手法 [4] などを利用することで抽出する方法も考えられる。

また、CC は放送データから直接得られる字幕データ (クローズドキャプション) を表示。

次に、これら M[0-2] および CC データを基にして、ストーリー展開や重要部分の把握を行う。例えば、スコア情報を基に、ある時刻でどちらの選手がより優勢であったかを示す優勢度を定義し [5]、この時間変化から試合展開を把握する。また、セットや試合の最後のラリーは他のラリーより重要、サービスエースやパッシングのシーンはより重要、といった基準から重要部分を把握する。これらを基にして、ウェブページ上に表示する表層テキストデータを生成する。この際、特定ドメインで使われる用語やイベントなどを対応付けた辞書を利用する。結果として、映像の重要な部分を示す付加情報 (アンカーテキスト) と対応する映像セグメントのデータが抽出される。これらのデータは、時間詳細度に応じて階層データとして構成される。

4. ウェブ化に基づく閲覧システムの試作

抽出メタデータと字幕データを用いたウェブ化による閲覧システムを試作した (図 3,4)。

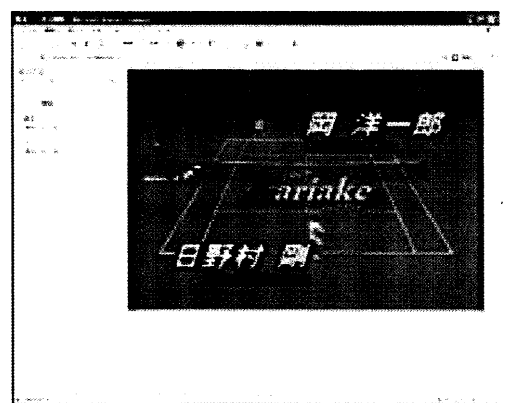


図 3: 試作システムの番組視聴画面

起動すると、まず通常のテレビ番組を視聴するスタイルで映像が表示される (図 3)。ここで、マウスをドラッグすると、通常のテレビ視聴をしていた画面が図 4 のように、サムネイルとテキストからなるウェブページに変換された形で表示され、視聴スタイルではなく閲覧スタイルとして番組を楽しむことができるようになる。

ここでストーリーボード形式として表示されているのは、図 2 の処理で得られたメタデータの、ある階層に相当するデータである。この階層データは、マウスの上下方向のドラッグ量に応じて拡大縮小される。また、ある大きさに別階層のデータに段階的に切り換えられ、その表示が連続的にズームするように制御される。

図 4 の左側フレームからは、右側フレーム内に表示するテキストの種類を選択できるようになっている。例えば、スコア情報のみ表示、イベント情報のみ表示、スコア情報+イベント情報などを選択することで、それに対応

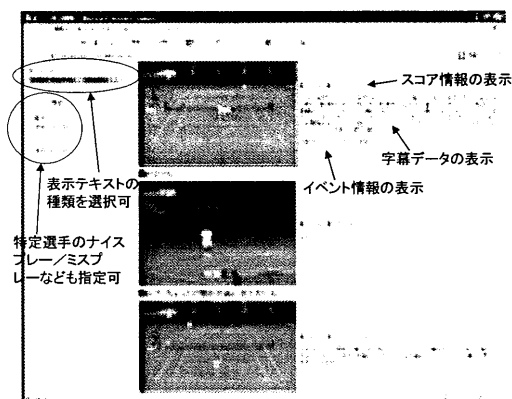


図 4: 試作システムの番組閲覧画面

するテキストのみが右側フレームに表示される。これにより、自分の興味のあるイベントや内容に対応するシーンをブラウザしながら効率よく探すことが可能となる。

また今回、ナレーションに相当する字幕データだけでなく、映像解析に基づく抽出メタデータを利用したことにより、スマッシュやパッシングといったイベントに直接対応した映像シーンをより正確に簡単に参照することができるようになったことを確認した。

5. まとめ

映像から抽出したメタデータと映像に付随する字幕データを利用して、テキスト情報と映像サムネイル画像からなるウェブページを生成・閲覧する手法について検討した。映像解析に基づく抽出メタデータを利用してテキスト情報を生成することにより、字幕データのみを利用して生成した場合と比べ、スポーツなどの映像において全体の内容把握や興味あるシーンの探索がより正確に簡単に行えることを確認した。また、表示テキストの種類を適宜選択できるようにすることで、より多角的な閲覧・探索ができるようになることを確認した。メタデータ抽出処理の改良、精度向上などが今後の課題である。

参考文献

- [1] K. Sumiya, et. al.: TV2Web: generating and browsing web with multiple LOD from video streams and..., ICKS2004, pp.158-167, 2004.
- [2] Sudhir, G., et. al.: Automatic classification of tennis video... CAIVD'98, 1998.
- [3] Miyamori, H.: Automatic annotation of tennis action for content-based retrieval by integrated... CIVR2003, LNCS2728, pp.331-341, 2003
- [4] Y.Watanabe, et al.: Image Analysis Using Natural Language..., JSAI, Vol.13, No.1, pp.66-74, 1998.
- [5] Miyamori, H.: Automatic generation of personalized video summary based on..., VLBV, 2003