

ウェブ化ビデオ：テレビ番組からウェブコンテンツへの メディア変換および関連情報との統合的閲覧方式

宮森 恒[†] 田中 克己^{†‡}

[†] 情報通信研究機構けいはんな情報通信融合研究センター 〒619-0289 京都府相楽郡精華町光台 3-5

[‡] 京都大学 大学院情報学研究科 社会情報学専攻 〒565-0456 大阪府吹田市河田 4-5-6

E-mail: [†] miya@nict.go.jp, [‡] ktanaka@i.kyoto-u.ac.jp

あらまし 本稿では、TV 番組をウェブコンテンツに変換し、その結果を関連する情報と統合することで、番組に関する内容をさまざまな視点に基づいて柔軟に閲覧可能とするウェブ化ビデオの概念を提案する。現在、テレビ番組やウェブコンテンツの情報量は増加の一途を辿っているが、人間に限られた時間で情報取得するには限界がある。特に一つ一つのメディアに割ける時間は限られている。そこで、有効に情報取得するには、異なるメディア間を状況や必要性に応じて相互に横断できる仕組みが必要と考えられる。ウェブ化ビデオはそのようなクロスメディア情報統合を実現する一つのツールとして位置づけられる。提案手法による TV 番組からウェブコンテンツへのメディア変換によって、テレビ番組の一覧性が向上し、テレビ番組の概要を早見したり、好きなシーンを効率よく探すことができるようになる。また、関連情報との統合によって、特定の視点に基づいたコンテンツや付加価値のついたコンテンツなどとともに、テレビ番組をさまざまな視点から視聴できるようになる。本稿では、次世代蓄積型テレビを想定したプロトタイプを作成し、提案する概念が有効に働くことを示す。

キーワード 次世代蓄積型テレビ、メディア変換、通信放送コンテンツ融合、トピック分割、メタデータ生成、シーン探索

Webified video: media conversion from TV program to web content and their viewing method integrated with related information

Hisashi MIYAMORI[†] Katsumi TANAKA^{†‡}

[†] National Institute of Information and Communications Technology (NICT)

3-5 Hikari-dai Seika-cho Souraku-gun, Kyoto, 619-0289 Japan

[‡] Kyoto University Yoshida Honmachi, Sakyo, Kyoto, 606-8501 Japan

E-mail: [†] miya@nict.go.jp, [‡] ktanaka@i.kyoto-u.ac.jp

Abstract A method is proposed for viewing broadcast content that converts TV programs into Web content and integrates the results with related information, enabling one to flexibly browse the original content with value-added content and from various viewpoints. Even though the amount of information available via TV programs and the Web is increasing constantly, humans have a definite ceiling on their spare time for each media. To efficiently acquire information, they need a mechanism that can smoothly go back and forth across different media depending on situations and necessity. Our proposed method is a tool to achieve such cross-media information integration. Media conversion from TV programs to Web content enhances the browsability of the former, enabling one to skim a program's outline or to efficiently search for favorite scenes. Integration with related information enriches viewing of TV programs in different ways, such as value-added content and content based on particular viewpoints. Testing of a prototype system for next-generation storage TV validated the approach taken by the proposed method.

Keyword Next-generation storage TV, media conversion, fusion of broadcast and web content, topic segmentation, metadata generation, scene search

1. はじめに

近年、ハードディスク(HD)レコーダの性能向上に伴い、一般の利用者が録画できるテレビ番組のデータ量

も飛躍的に増加している。現在、600GBの容量をもつHDレコーダが登場し、画質モードによっては1070時間以上の録画が可能である。また、複数チャンネル

の同時連続録画が行える HD レコーダも登場しており、6 チャンネル分の番組を 1 週間連続して録画することも可能である。一方、近年の高速なインターネット環境の普及に伴い、一般利用者がアクセスできるウェブコンテンツも増加の一途をたどっている。2000 年に 20~50 テラバイトと見積もられた全世界の Surface Web のコンテンツは、2003 年の試算ではその 3 倍のデータ量にあたる 167 テラバイトに達したとの報告がなされている [1]。

しかし、人間に限られた時間で情報取得するには物理的限界がある。例えば、テレビ番組は実時間で見るように作られているため、2 倍速や 3 倍速などで内容を把握するのは困難である。また、テレビやインターネットなど一つ一つのメディアに割ける時間も限られている。例えば、テレビを視聴していて番組の内容に関する情報をウェブコンテンツから探すには、PC などを用いて検索サイトを開き、適当なキーワード検索を別途行う必要がある。また、検索結果から見たいページを探すまでに手間がかかることも珍しくない。そのため、有効に情報取得するには、異なるメディア間を状況や必要性に応じて相互に横断できる仕組みが重要になると考えられる。

本稿では、このようなクロスメディア情報統合を実現する一つのツールとして、ウェブ化ビデオを提案する。ウェブ化ビデオは、テレビ番組をウェブコンテンツにメディア変換し、その結果を関連する情報と統合することで、付加価値が追加された内容や種々の視点に基づく内容を柔軟に閲覧可能とする方式である。また、テレビ番組とウェブコンテンツという異なるメディアをスムーズに行き来しながら視聴・閲覧できるため、効率よく情報取得をすることができる。本稿では、次世代蓄積型テレビを想定したウェブ化ビデオ生成システムのプロトタイプを実装することで、提案方式の有効性を示す。

本稿の構成は以下の通りである。2 章では、ウェブ化ビデオの概念とウェブ化変換の処理概要について述べる。3 章ではウェブ化ビデオ生成システムの実装について詳しく説明する。4 章では考察を示し、5 章で関連研究を整理する。最後に 6 章でまとめを述べる。

2. ウェブ化ビデオの概念

図 1 に、ウェブ化ビデオの概念を示す。ウェブ化ビデオを生成するには、まず、番組映像とそれに付随する字幕テキストなどのメタデータを用いて番組映像の構造化が行われる。次に、ウェブ上あるいはローカルに利用可能なコンテンツの中から関連情報の検出・取得が行われる。最後に、番組内の構造化データ、および、取得された関連情報がそれぞれハイパーリンクによってオリジナルコンテンツと関連付けられることにより、ウェブコンテンツが生成される。得られたウェブコンテンツを適当なブラウザで閲覧・操作すること

により、番組内のシーン探索を容易に行うことができる。また、その番組に関連したより詳細な、あるいは、多様な情報に円滑にアクセスできるようになる。利用者のフィードバックは、適宜ウェブ化変換に反映され、ウェブコンテンツは利用者の嗜好や状況に応じて再構成することができる。なお、本稿では、ウェブコンテンツとはウェブ技術を用いた電子文書といった意味で用いている。

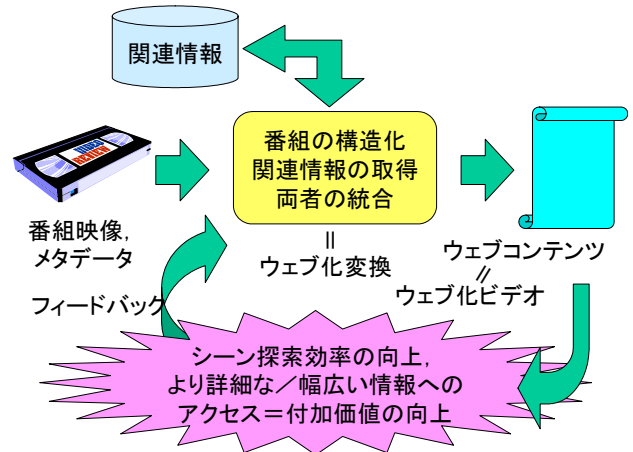


図 1. ウェブ化ビデオの概念

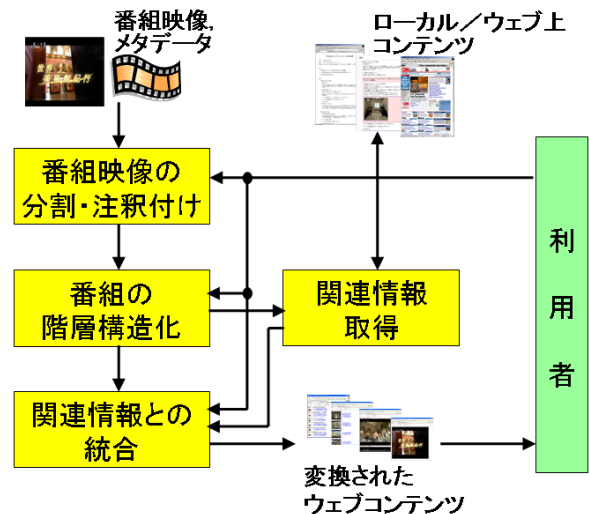


図 2. ウェブ化ビデオ生成処理の概要

図 2 にウェブ化ビデオ生成処理の概要を示す。まず、番組映像とそれに付随するメタデータに対し番組映像の分割が行われる。ここでは、トピックやサブトピック、シーンやショットといった異なる詳細度で分割が行われる。また、分割映像のシーン分類や映像中の人物・イベント検出により番組の注釈付けを行う。得られた分割データおよび注釈データを用いて、詳細度の異なるデータ間で階層構造が作られるように、番組の構造化が行われる。次に、構造化で得られたメタデータを用いて、関連情報の取得が行われる。ここでは、ウェブ上あるいはローカルに利用可能なコンテンツを対象として関連情報の検索や対象番組以外のメタデー

タの取得が行われる。最後に、得られた関連情報とオリジナルの番組映像がハイパーリンクによって関連付けられ、統合されたウェブコンテンツが出力される。

図3に、従来の番組映像とウェブ化された番組映像の比較を示す。録画された従来の番組映像は、単体の映像データとして記録され、ウェブコンテンツを初めとする関連情報との関連付けはなされていない。一方、ウェブ化ビデオは、異なる詳細度で階層構造化され、番組データ内のさまざまな位置へハイパーリンクを介した移動が可能である。さらに、番組データで提供される情報以外の関連情報へのハイパーリンクを介して、さまざまな付加価値をもつコンテンツへスムーズにアクセスすることが可能である。

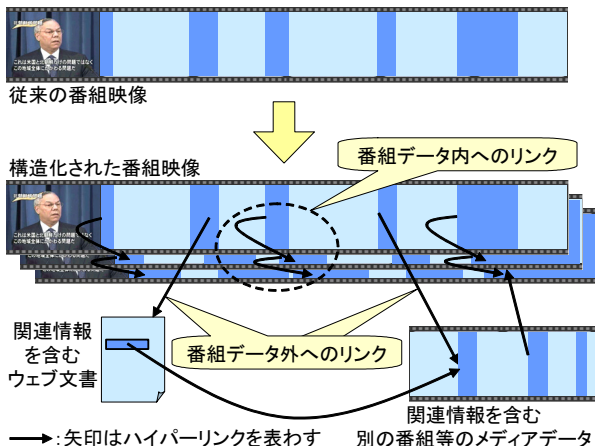


図3. 従来の番組映像とウェブ化された番組映像

3. ウェブ化ビデオ生成システムの実装

ウェブ化ビデオの概念に基づく生成システムのプロトタイプを実装した。

本システムでは、まず、字幕付きのテレビ番組を録画し、字幕テキストを用いたトピック分割を階層的に行う。得られた異なる詳細度の字幕テキストと対応する部分映像をペアとして構造化する。字幕テキストと部分映像の組は、画面上でストーリーボード形式に表示されるようレイアウトされる。次に、補完検索や番組実況チャットを利用して関連情報を取得し、適宜、ストーリーボード上の対応する位置に統合してレイアウトする。出力されたウェブコンテンツは、ズーム操作により、部分映像の大きさが連続的に変化するだけでなく、異なる詳細度データへの遷移が適宜行われる。これにより、通常のテレビ視聴画面とストーリーボード閲覧画面をシームレスに移動することができ、番組中の興味のあるシーンを効率よく探索することができる。また、関連情報へのリンクが統合されているため、状況に応じてより詳しい／幅広い情報へ素早くアクセスすることができる。以下、各手順の処理内容を説明する。

3.1. 番組映像の分割・注釈付け

番組映像の分割・注釈付けは、字幕テキストを用いたトピック分割を階層的に行うことで実現する。

まず、字幕テキストは、番組中の発話がなされた時刻 t と、発話内容を記述したテキスト s のペア (t, s) となるデータであり、文字多重放送の番組であれば適当なデコーダを使うことで簡単に記録することができる。一般に、テキスト s は、文の途中で切れていることがあり、その場合、続きは次の時刻のテキストに含まれている。よって、文の終わりを示す句読点を検出することにより、 (t, s) において s : 単一の文のテキスト、とすることができる。

さて、字幕テキストを用いたトピック分割には大きく分けて2つのアプローチがある。1つは学習に基づくアプローチだが、

- ドメインが限られる可能性があり、汎用的な話題に適用するのが難しい、
- 十分なトレーニングデータが必要、

という問題点がある。もう1つは、統計手法に基づくアプローチである。ここでは、トレーニングデータを特に必要とせず、任意のドメインで利用できるという利点のある、単語分布に基づくトピック分割手法[2]を用いることにした。

このトピック分割手法では、以下の条件を前提としている。

- トピックは、各トピックの単語分布によって定義される。
- 異なるトピックは異なる単語分布をもち、統計的に互いに独立である。
- あるトピックが与えられたとき、そのトピック内の単語は統計的に独立。

いま、 n 個の単語で構成されるテキストを

$$W = w_1, w_2, \dots, w_n$$

とし、 W を m セグメントに分割したテキストを

$$S = S_1, S_2, \dots, S_m$$

とする。分割 S となる確率は、

$$\Pr(S | W) = \Pr(W | S) \Pr(S) / \Pr(W)$$

で表わされる。与えられた W で $\Pr(W)$ は定数であるため、もっともらしい分割 S は、

$$S = \arg \max \Pr(W | S) \Pr(S)$$

で与えられる。

$$\Pr(W | S) = \Pr(W_1, W_2, \dots, W_m | S)$$

$$= \prod_{i=1}^m \Pr(W_i | S)$$

$$= \prod_{i=1}^m \Pr(W_i | S_i)$$

$$= \prod_{i=1}^m \prod_{j=1}^{n_i} \Pr(w_j^i | S_i)$$

ここで、 w_j^i : S_i 内の j 番目の単語を表わす。また、Laplace の法則 [3] から、

$$\Pr(w_j^i | S_i) = (f(w_j^i) + 1) / (n_i + k)$$

が成り立つ。ここで、 $f(w_j^i)$ は、 w_j^i と一致する W_i 内の単語数、 k は、 W 内の異なる単語数、を表わす。

一方、 $l(S) = m \log n$ (bits) とすると、

$$\Pr(S) = n^{-m} = 2^{-l(S)}$$

が成り立つ。

ここで、分割コスト $C(S) = -\log \Pr(W | S) \Pr(S)$ とし、 $C(S)$ を最小化する S を求める。

$$C(S) = -\log \Pr(W | S) \Pr(S)$$

$$= -\sum_{i=1}^m \sum_{j=1}^{n_i} \log \Pr(w_j^i | S_i) - \log \Pr(S)$$

$$= -\sum_{i=1}^m \sum_{j=1}^{n_i} \log(f(w_j^i) + 1) / (n_i + k) + m \log n$$

$$= \sum_{i=1}^m c(w_1^i, w_2^i, \dots, w_{n_i}^i | n, k)$$

ここで、

$$c(w_1^i, w_2^i, \dots, w_{n_i}^i | n, k)$$

$$= \sum_{j=1}^{n_i} \log(n + k) / (f(w_j^i) + 1) + \log n$$

n 個の単語で構成されるテキストを

$$W = w_1, w_2, \dots, w_n$$

とし、 w_i と w_{i+1} の間の位置を g_i とする。

グラフ $G = \langle V, E \rangle$, V : ノードの集合、 E : エッジの集合とし、 $e_{i,j}$ の始点と終点が g_i, g_j とすると、

$$V = \{g_i | 0 \leq i \leq n\}, E = \{e_{i,j} | 0 \leq i < j \leq n\}$$

となるグラフが考えられる (図 4)。

エッジ $e_{i,j}$ のコスト $c_{i,j}$ は、

$$c_{i,j} = c(w_{i+1}, w_{i+2}, \dots, w_j | n, k)$$

となり、 g_0 から g_n までのコスト最小となるパスを DP アルゴリズムで検出することで、もっともらしい分割 S を得ることができる。

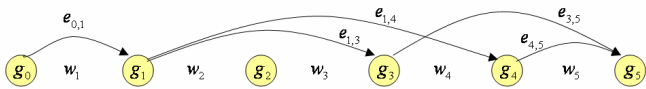


図 4. コスト最小となるパスの検出

以上のトピック分割を字幕テキスト s に適用させることでトピック単位に分割された結果 S を得る。分割された各トピックの字幕テキストを再び s として同様のトピック分割を行うことで、サブトピック単位に分割された結果 S を得る。映像データは、字幕テキスト

の分割結果から得られる時刻に基づいて分割される。

以上より、番組映像は、トピック単位、サブトピック単位、文単位の 3 つの異なる詳細度データに分割された。これを図 5 のように階層的に構造化する。

トピック 1		トピック 2		トピック i	
サブトピック 1	サブトピック 2	サブトピック j	...
文 1	文 2	文 k	...

図 5. 異なる詳細度で分割された字幕および映像の構造化

3.2. 関連情報の取得

関連情報は、補完情報検索あるいは番組実況チャットを利用することで取得する。

まず、補完情報検索について。

補完情報検索は、字幕データから話題構造と呼ばれるデータを抽出し、それに基づく構造化質問を生成した後、ウェブ上の検索エンジンで検索を行うことにより、従来の類似検索とは異なった、より詳細なあるいは別の観点からの情報を検索できる手法である [4]。

話題構造は、subject-term と content-term の集合のペアから構成される。subject-term は、コンテンツに述べられている話題の主題となる語である。ここでは、字幕テキスト中の出現頻度の高い、かつ、その他のキーワードとの共起関係の強いキーワードを subject-term とした。一方、content-term は、同じコンテンツに出現し、subject-term との共起関係の強いキーワードである。言い換えれば、subject-term はその話題のタイトルを表す役割があり、content-term は話題の本体を表し、内容記述の役割がある。

このように、タイトルと内容のキーワードを区別することによって、タイトル/内容についてそれぞれの詳細さを重視した結果を期待する構造化質問、また、タイトル/内容についてそれぞれの話題が広がり重視した結果を期待する構造化質問を記述できる。これにより、主題類似・内容類似といった情報検索が可能となるだけでなく、主題が類似しているが内容が異なるといった傾向をもつ、似て非なる情報の検索も可能となる。

ここでは、以下のようなキーワードの組からなる構造化質問を生成し、Google API を用いた検索を行うこ

とした。

1. 詳細さを重視した質問
(ア) タイトルについての詳細さを重視
(イ) 内容についての詳細さを重視
2. 話題の広がり重視した質問
(ア) タイトルについての広がり重視
(イ) 内容についての広がり重視

次に、番組実況チャットの利用について。

番組実況チャットは、テレビ番組の視聴者がインターネット上のチャットコミュニティに集まり、テレビ番組の内容について感じたことなどやテレビ番組の内容自体のことをリアルタイムに書き込み、会話を楽しむシステムである。利用者は、テレビ番組と並行してチャットを楽しむことにより、視聴者との擬似的な感情の共有による一体感を味わうことができる。

巨大掲示板サイトである2ちゃんねるでは、この番組実況チャットを大規模サポートしている。2ちゃんねるでは、NHK 専用、日本テレビ専用、TBS 専用、フジテレビ専用、テレビ朝日専用、テレビ東京専用、ラジオ専用など各カテゴリ（ここでは放送局）に応じた番組実況チャット用の掲示板が設置されており、その中に番組自体について語り合うスレッド（ある話題に関して投稿されたメッセージの集まりで、書き込まれた時間順に並べられている）が多数存在する。

番組実況チャットは、設定による違いはあるが、基本的に、書き込み時刻、書き込み者の ID、書き込み内容から構成されている（図 6）。

- 80 名前:名無しさん 04/10/19 08:48:22 ID:R9C+YqFS
ものすごいブリッツの中からTD
キタ————(∇)————!!
- 81 名前:名無しさん 04/10/19 08:48:28 ID:+FNfUHkk
50点以上入りそうな勢い
- 82 名前:名無しさん 04/10/19 08:48:30 ID:BZZg6/Cd
1stだけで・・・
- 83 名前:名無しさん 04/10/19 08:48:48 ID:7iYg967m
マニングはこういうとき凄くムキになる。
しかしだでさえ背が高いのにジャンプ投げとは反則だー
- 84 名前:名無しさん 04/10/19 08:49:15 ID:3bLzDEmh
初心者には空中戦がわかりやすくてよいかもな。
- 85 名前:名無しさん 04/10/19 08:49:23 ID:R9C+YqFS
ブリッツ多かったけどバック何人残ってたんだろ？

図 6. 番組実況チャットの書き込み内容の例

著者らは、番組実況チャットにおける書き込み数や書き込み内容から、テレビ番組中のどの場面でどのような感情の変化が生じたのかということ抽出するインデキシング手法を提案している [5]。

ここでは、以下のような数値の組からなるインデックスを取得することとした。

1. 視聴者（書き込み者）全体について
(ウ) 計測時刻、計測単位時間
(エ) 書き込みエントリ数
(オ) 盛り上がり・落胆の感情度数

2. 視聴者（書き込み者）個別について
(カ) 書き込み者 ID
(キ) 計測時刻、計測単位時間
(ク) 書き込みエントリ数
(ケ) 盛り上がり・落胆の感情度数

3.3. ウェブコンテンツの生成と視聴インタフェース

トピック単位、サブトピック単位、文単位の 3 つの異なる階層に構造化された番組データと、補完情報検索／番組実況チャットから得られた関連情報データを統合し、ウェブコンテンツとして表示できるようにする。本稿では、字幕テキストと部分映像の組を、画面の縦方向に並べるストーリーボード形式を採用することとした。映像と字幕テキストは、トピック単位、サブトピック単位、文単位といった異なる単位で分割されているが、ここでは説明の都合上、分割された各部分映像のことを単にシーンと呼ぶ。図 7 に示すように、シーンのサムネイルとシーンに対応する字幕テキストを横に並べ、これを縦方向に時間順に配置する。

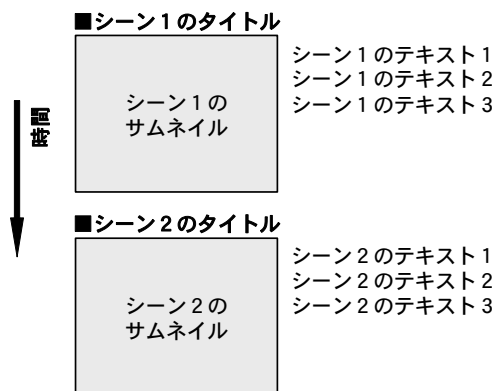


図 7. ストーリーボード表示

図中、シーン 1 がトピック 1 の映像に対応していたとすると、シーン 1 の全テキスト（すなわち、シーン 1 のテキスト 1、2、3）がトピック 1 の字幕テキストに対応する。シーン 1 の各テキスト 1、2、3 は、シーン 1 のサブトピックの字幕テキスト 1、2、3 にそれぞれ対応する。シーン 1 のサムネイルは、トピック 1 の映像を再生するようにハイパーリンクで関連付けられ、シーン 1 のテキスト 1 は、シーン 1 のサブトピック 1 の映像を再生するように同様にハイパーリンクで関連付けられる。実際のストーリーボード表示の画面を図 8 に示す。

さらに、補完情報検索で得られた結果を統合表示した例を図 9 に示す。各トピック、各サブトピック毎に、補完情報検索で得られたウェブページへのハイパーリンクが追加されている。これにより、利用者はより詳細な話題、あるいは、その話題を広げるような情報に必要に応じて効率よくアクセスすることができる。



図 8. トピック単位に分割された部分映像と字幕データのストーリーボード形式での表示

文レベルのストーリーボードに滑らかに表示を切り替える。さらに、ズームインすると、通常のテレビ視聴画面の表示に滑らかに切り替わる。ズームアウト操作は全く逆の効果となる。これにより、通常のテレビ視聴画面とストーリーボード閲覧画面をシームレスに横断することができ、番組中の興味のあるシーンや関連情報を効率よく探索することができる。

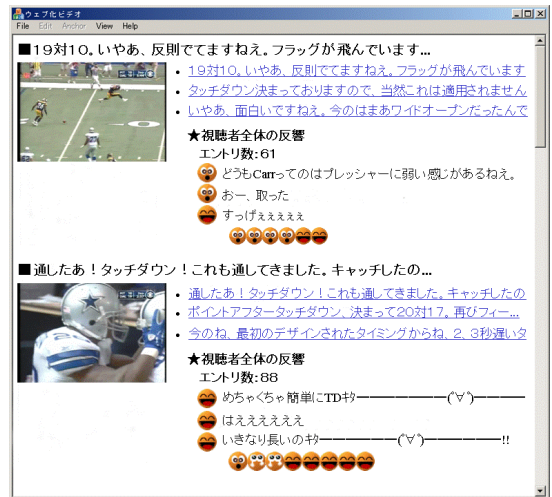


図 10. 番組実況チャットから得られる視聴者全体の反響の様子を統合表示した例

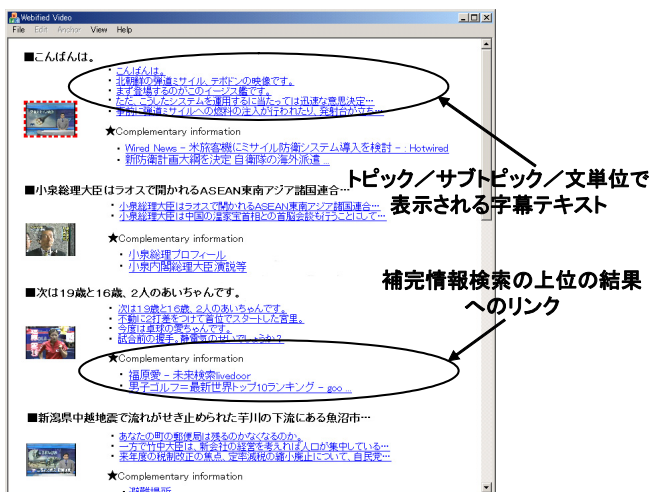


図 9. 補完情報検索の結果を統合表示した例

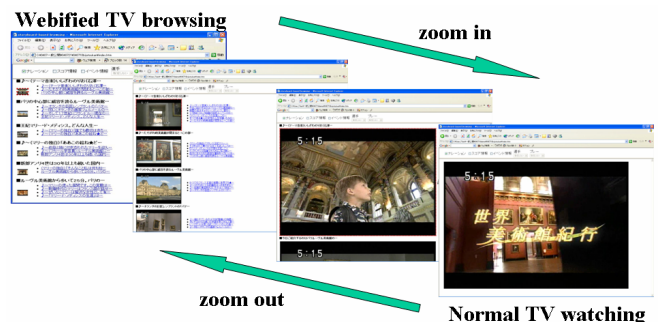


図 11. ズームによる画面遷移

また、番組実況チャットから得られる視聴者全体の反響の様子を統合表示した例を図 10 に示す。書き込みエントリ数や感情を表わす顔アイコンの挿入によって、利用者は視聴者の返す反響の様子をあたかも共有しながら閲覧できるようになる。

図 11 は、ウェブ化ビデオのズームによる画面遷移の様子である。ズーム操作により、表示映像の大きさが連続的に変化するだけでなく、異なる詳細度データへの遷移が適宜行われる。例えば、図 5 のトピック 1、トピック 2、…、トピック i を含むストーリーボードは、ズームインの操作によってサムネイルがある大きさに達すると、サブトピック 1、…、サブトピック j を含むストーリーボードに滑らかに表示を切り替える。さらにズームインの操作を続けてサムネイルがある大きさに達すると、サブトピックのストーリーボードは、

4. 考察

ウェブ化ビデオは、テレビ番組をウェブコンテンツにメディア変換することにより、これまで受動的に視聴するスタイルが一般的だった番組映像を、より積極的に「閲覧」することを可能にした、という意味をもつ。また、オリジナルの番組データと関連するデータがウェブコンテンツ上で統合されていることから、さまざまな情報へのアクセスがスムーズになったといった意味をもつ。

特に、番組映像はウェブコンテンツの形式に変換されているため、図 7 に示した単なる時間順による一覧表示はランキング順で表示しなおすことも可能となる。例えば、番組実況チャットの書き込みエントリ数でソートすることにより、視聴者全体の反響が大きかった順に各シーンを並び替えて表示することができる(図

1 2, 1 3). この表示は、一般的に反響の大きかったシーンを順にチェックしながら視聴する際に役立つと考えられる。

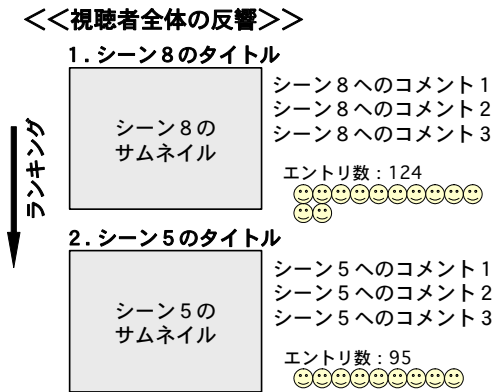


図 1 2. ランキングによる表示



図 1 3. 番組実況チャットから得られる視聴者全体の反響の様子をランキング表示した例

また、同様にウェブコンテンツの特性を活かすことで、複数番組の情報統合も可能になると考えられる。例えば、複数の番組を同一の日付でまとめたグループの一覧を表示したり、同一番組・同一トピックを時系列的にまとめたグループの一覧を表示することなどが実現できる。これらの表示は、その階層レベルによって、動画像と音響データで再生されたり、テキストデータと音響データのみで再生されたりしてもよい。さまざまな提示方法が提案法の枠組みの中で実現できると考えられる。

なお、本稿では字幕データが利用できることを前提としてテキストに基づくトピック分割手法を利用した。しかし、スポーツ中継等の番組では、現在、字幕がほとんど提供されていない。そのような場合、映像のカット検出や映像キャプション認識を利用したテキスト抽出、あるいは、ドメイン知識を利用した分野に特化したイベント認識技術の利用が必要になると考えられる [16][17].

5. 関連研究

従来、膨大な映像データから興味のある部分を効率よく探したり、内容の概要を手軽に理解したり、限られた時間でハイライトだけを見るといったさまざまな視聴方法についての研究は活発に行われてきた。

例えば、映像コーパスから興味のあるショットを探索するために視覚的要素を利用したインタフェースが提案されている [6]. 映像セグメントを表す複数のキーフレームから構成されるストーリーボードと視覚的特徴によるフィルタリング機能を組み合わせることで効率的な探索が可能だとしている。

また、番組概要の把握やシーン探索を効率行うための表示インタフェースとして TV2Web [7] や Video Manga [8] が提案されている。TV2Web では、映像セグメントのサムネイルと字幕テキストから構成されるストーリーボードを異なる詳細度レベル (セグメント、シーン、ショット) で構造化し、これら構造データをズームメタファによってスムーズにつなぎながら (切り替えながら) 表示することで効率的な概要把握やシーン探索が可能だとしている。Video Manga では、クラスタリングされた映像セグメントが長くて稀なもの的重要度を高く計算し、その値に応じて表示するキーフレームの大小を制御する。異なるサイズのキーフレームをマンガのような表示形態に配置しサマリーとして表示することで、より重要な場面を視覚的に把握しやすいインタフェースを実現している。

また、限られた時間でハイライトを視聴するために、テレビ番組のためのダイジェスト視聴システムが提案されている [9]. 番組映像中に配信側で用意された汎用的な番組索引と検索用ルールが含まれることを前提とし、これらを利用して利用者の嗜好を考慮したハイライトをダイジェスト視聴することができる。

これらの手法は、いずれもキーフレームや重要な映像セグメントを空間的あるいは時間的に展開することで、一覧性を向上させたり、短時間で概要把握を可能とするものであり、基本的に、利用者に提示する情報量を削減するといったアプローチであった。一方、ウェブ化ビデオは、このような一覧性の向上や短時間での概要把握を可能とするアプローチは踏襲しているものの、ウェブの特徴であるハイパーリンクの利用によって、番組映像以外のより詳しい情報あるいは幅広い関連情報を統合表示し、利用者に提示する情報量を必要に応じて増やすアプローチもとっている点がユニークであると考えられる。

また、シーン探索や概要把握、ハイライト視聴の仕組みを実現するための基礎技術である映像インデキシング技術の研究も活発に行われてきた。

例えば、色を用いた例としては、色ヒストグラムや特徴的な色の集合、隣接する特徴的な色対などを利用した手法が提案されている [10]. これにより、問合せ

で指定された画像中の物体が、動画中に存在するかどうかの判定などが可能になる。また、動きを用いた例としては、動きベクトルを用いた手法が提案されている[11]。ブロックマッチング法などで得られた動きベクトルを、ズームやパンなどを特徴づけるパラメータでモデル化し分類することにより、各カメラ操作の有無や強さの判定が可能になる。

さらに、マルチモーダルなインデキシングの例としては、シーン分割、カメラ操作検出、顔検出、映像キャプション検出等を利用して、シーンの重要度を判定・選択する手法[12]や、テキスト解析で得られたキーセンテンスと画像解析で得られたキー画像を、動的プログラミングを使って関連付ける手法[13]が提案されている。マルチモーダルな情報に加えて、特定分野のドメイン知識を利用することにより、画像中の人物動作等の、より詳細な内容に関するインデキシングが可能であることも示されている[14][15][16]。

しかし、これら従来手法が利用してきたデータは、番組の配信側（放送局）から提供される情報のみであったため、インデキシングされた情報は、基本的に、番組作成者・配信側の意図に基づく画一的なものであった。よって、従来手法においては、シーン探索や内容の早見、ダイジェスト視聴などの内容に、番組作成者側や自分以外の他の利用者の視点や反応といった要素を取り入れることはできなかった。提案手法では、補完情報検索を利用することで、番組のもつ情報をより詳細に／幅広くするようなインデキシングが可能となる。また、番組実況チャットを利用することで、番組視聴者の視点や反応を反映したインデキシングが可能となった。オリジナルの番組は、番組作成者や配信側以外の利用者が提供する情報とともに付加価値を向上させた様々な見方が可能になったと考えられる。

6. まとめ

本稿では、クロスメディア情報統合を実現する一つのツールとして、ウェブ化ビデオを提案した。また、次世代蓄積型テレビを想定したウェブ化ビデオ生成システムのプロトタイプを実装することで、提案方式の有効性を示した。ウェブ化ビデオでは、テレビ番組を異なる詳細度で構造化し、映像内のさまざまな位置へハイパーリンクを介して移動できるようにする。さらに、番組映像以外の関連情報を取得し、そのコンテンツとオリジナルの番組を適宜ハイパーリンクで関連付ける。これにより、利用者は、テレビ番組中の見たいシーンを探索しやすくなるだけでなく、その番組が提供していた情報よりも詳細な、あるいは、多様な情報に状況に応じて素早くアクセスできるようになった。

今後は、実時間ウェブ化変換や実時間閲覧機能の実現、インタフェースの改善などプロトタイプシステムの改良を行っていく予定である。ウェブ化ビデオ生成システムが手軽に構築でき、まとまった被験者による

利用実験を行うことにより、より多様な番組視聴が可能になることを実証することができると考えられる。

文 献

- [1] Lyman, P., Varian, H.R., Swearingen, K., Charles, P., Good, N., Jordan, L.L., Pal, J.: How much information?2003.
<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/>
- [2] Utiyama, M., Isahara, H.: A Statistical Model for Domain-Independent Text Segmentation. ACL/EACL-2001, pp. 491--498.
- [3] Manning, C.D., Shutze, H.: Foundations of statistical natural language processing. MIT Press, 1999.
- [4] Qiang, Ma., Tanaka, K.: Content Integration Based on Complementary Information Retrieval. IPSJ 2004-DBS-134, pp.337-343, 2004.
- [5] Miyamori, H., Nakamura S., Tanaka, K.: Automatic indexing of broadcast content using its live chat on the web. IEICE NLC2004-123, PRMU2004-205, pp.43-48, 2005.
- [6] Christel, M.G., Huang, C.: Enhanced access to digital video through visually rich interfaces. ICME, MD-L5.1, 2003.
- [7] Sumiya, K., Munisamy, M., Tanaka, K.: TV2Web: generating and browsing web with multiple LOD from video streams and their metadata, ICKS2004, pp.158-167, 2004.
- [8] Uchihashi, S., Foote, J., Girgensohn, A., Boreczky, J.: Video Manga: generating semantically meaningful video summaries. Proc. ACM Multimedia 99, 1999.
- [9] Hashimoto, T., Shiota, Y., Mano, H., Iizawa, A.: Prototype of Digest Viewing System for Television. IPSJ, Vol.41, No.SIG3(TOD6), pp.71-84, 2000.
- [10] Nagasaka, A., Tanaka, Y.: Automatic video indexing and full-video search for object appearances. IPSJ, Vol.33, No.4, pp.543-550, 1992.
- [11] Akutsu, A., Tonomura, Y., Hashimoto, H., Ohba, Y.: Video indexing using motion vectors. In SPIE Proc. VCIP '92, pp.522-530, 1992.
- [12] Smith, M., Kanade, T.: Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques. CVPR, 1997.
- [13] Nakamura, Y., Kanade, T.: Semantic analysis for video contents extraction - spotting by association in news video. ACM Multimedia, pp.393-401, 1997.
- [14] Intille, S.S., Bobick, A.F. Closed-world tracking. Proceedings of the Fifth International Conference on Computer Vision, pp. 672-678, 1995.
- [15] Saur, D.D., Tan, Y-P., Kulkarni, S.R., Ramadge, P.J.: Automated analysis and annotation of basketball video. Storage and Retrieval for Image and Video Databases V, SPIE-3022, pp.167-187, 1997.
- [16] Miyamori, H.: Automatic annotation of tennis action for content-based retrieval by integrated audio and visual information. CIVR2003, LNCS2728, Springer-Verlag, pp.331-341, 2003.
- [17] Miyamori, H., Tanaka, K.: Webification of TV program and its browsing method using extracted metadata and closed captions. FIT, D-036, 2004.