

映像アーカイブとファインダビリティ

井手 一郎*

近年、ハードディスクなどの記憶装置の高密度化に伴い、大量の映像を電子的にアーカイブする環境の構築が容易になった。しかし、視聴に時間を要することや、内容の正確かつ網羅的な記述が難しいことから、単純なキーワード検索や画像特徴に基づく検索により、欲しい映像を求めるのは容易ではない。本稿では、筆者らがこれまで取り組んできた、放送映像アーカイブにおける効率的な検索のための研究を中心に、内容解析およびそのための構造化の視点から、さまざまな研究事例について外部情報源との関連づけの有無に分けて、映像ジャンル別に紹介する。

キーワード：映像アーカイブ、放送映像、映像検索、構造化、内容解析、ニュース映像、料理映像

1. はじめに

従来、放送局を中心に、テレビ映像やその素材映像がアーカイブされてきた。わが国では、日本放送協会のNHKアーカイブスをはじめ、各事業者による自主的なアーカイブが中心だが、海外では、フランスの国立視聴覚研究所(INA: Institut National de l'Audiovisuel) やイギリスの英国映画協会(BFI: British Film Institute) などにおいて、社会で共有すべき資産として、早くから公的にアーカイブされてきた国もある。

一方、近年、ハードディスクなどの記憶装置の高密度化に伴い、大量の映像を電子的に蓄積し、オンラインでアクセスできる環境の構築が容易になった。前掲の各映像アーカイブにおいても、現在は電子的にアーカイブされつつある。我が国では、2011年に予定されているデジタル放送への完全移行を機に、今後はオンライン映像アーカイブの構築がより簡単になるとと思われる。

しかし、大量の映像を電子的にアーカイブしてオンラインでアクセスできるようになっても、視聴に時間を要することや内容の正確かつ網羅的な記述が難しいことから、単純なキーワード検索や画像特徴に基づく検索により、欲しい映像を求めるのは容易ではない。

このような点を見越して、1994年に米国Carnegie Mellon大学(CMU)において、Informediaプロジェクト²⁾が始まった。これは、大規模な放送映像アーカイブを想定した世界初の研究プロジェクトである。

このような流れをうけ、筆者らもこれまで、国立情報学研究所において放送映像アーカイブNII-TVRECSを共同で構築し³⁾、効率的な検索を実現するために、映像の索引付け、映像間の意味構造の解析、さらに利用方法の検討を行ってきた。本稿ではこれらの研究を中心に関連する研究

に触れながら、様々な形態による検索をふまえた、放送映像の構造化・内容解析に関する研究事例を紹介する。

2. 放送映像の内容解析

放送映像の内容を解析する方法としては、純粹に受信した映像のみを手がかりにする場合と、関連する外部情報を利用する場合がある。以下、これらの方法について簡単に説明する。

2.1 映像に含まれる情報

一般に映像とは、高速に連続して撮影された画像の系列と、それに同期した(主・副)音声を組み合わせたデータである。また、わが国の放送映像の場合、総務省の指針に従い、可能な番組に対して、字幕放送により音声書き下しテキスト(以下、字幕放送テキスト)を付すこととされている⁴⁾。その結果、2009年度時点で、一部の生放送番組を含め、全放送の約40%に字幕が付随している⁵⁾。また、米国をはじめ、他の国でも、closed-captionの形で同種のデータが付随することが多い。放送映像の内容を解析する際には、これらの情報のうち特に字幕放送テキストを手がかりとすることが多い。

一方、画像特徴量から画像の内容を推定する技術が発展しつつある⁶⁾ため、最近では、画像特徴量も積極的に利用されるようになった。このような画像特徴量による画像・映像検索はCBIR/CBVR(Content-Based Image/Video Retrieval)と呼ばれている。特に映像検索については、米国標準技術研究所(NIST: National Institute of Standard and Technologies)が主催して2001年に始まった、評価型ワークショップTRECVID(当初はテキスト検索の評価型ワークショップTRECの一セッション)⁷⁾を中心に、大量の画像から画像特徴量と画像内容の対応関係を統計的に学習する方法が盛んに研究されている⁸⁾。

2.2 放送映像に関連する情報

一方、放送映像には、関連する様々な外部情報も存在す

*いで いちろう 名古屋大学 大学院情報科学研究科
〒464-8601 愛知県名古屋千種区不老町
Tel. 052-789-3313 (原稿受領 2010.8.30)

る。

直接関連する情報源としては、以下のようなものがある。

- ・出版物：台本、副読本、教材
- ・番組情報：新聞や雑誌に掲載されているもの、電子番組表（EPG：Electronic Program Guide）
- ・ウェブ：番組公式ページなど

また、間接的に関連する情報源としては、以下のようなものがある。

- ・出版物：新聞、雑誌、書籍など
- ・ウェブ：ブログ、掲示板など
- ・映像：予告・総集編や、他放送局・視聴者が撮影した映像

これらの情報は、関連する内容について第三者が作成したもののほか、視聴者が当該映像を承けて制作したものである。

3. 映像特徴による構造化と内容解析

3.1 放送映像の構造と構造化

一般に放送映像は高度な編集作業を経て製作されるため、時系列的な構造をもっている。これには、映像の内容によらず信号処理的に検出し得る「物理構造」(図1)のほか、撮影者・編集者の意図に基づく、「意味構造」がある。各構造に関する用語を次のように定義する。

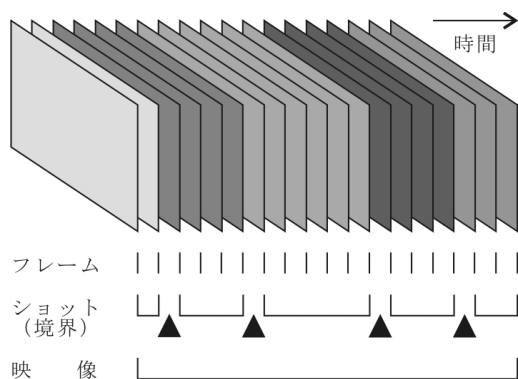


図1 映像の物理構造

(1)物理構造

- ①フレーム：映像を構成する最小単位の静止画像
- ②ショット：連続して撮影された一連のフレーム系列
- ③ショット境界：ショット間で画像が不連続になる箇所
ショット境界には、カット、ワイプ、フェードイン・アウト、ディゾルブなど様々な編集効果を加えられる。

(2)意味構造

- ①シーン：特定の構図や役割をもつ映像区間。たとえば、

ニュース映像の「キャストシーン」、スポーツ映像の「得点シーン」、料理映像の「手元シーン」のように、対象映像ジャンル毎に重視する内容に応じて設定されることが多い。なお、シーンは必ずしもショットから構成されるとは限らない。

- ②映像ジャンル固有の構造：ニュース映像の「トピック」、スポーツ映像の「攻守交代」など、対象映像ジャンル固有の様々な構造が存在する。

一方、映像アーカイブのように大量の映像がある場合は、映像間に存在する関連性に基づく構造も考える必要がある。

放送映像の内容を解析するためには、これらの映像内・映像間の構造を明らかにする必要がある。以下では構造化の方法に応じて、様々な内容解析の事例について紹介する。

3.2 一般の放送映像の例

一般の放送映像を対象とした構造化を考える際には、映像ジャンルによらない特徴を考える必要がある。初期の研究として佐藤らによる「InfoStripes」⁹⁾、最近では(株)東芝による映像閲覧ツール「顔 de ナビ」¹⁰⁾などがある。前者は、ショット境界の種類、動き、テロップの出現など低次の映像特徴の出現パターンを可視化した。後者は、顔の出現、音量、歓声・拍手による盛り上がり、音楽・通常音響の分類を可視化した。

一方、筆者らは、準同一 (near-duplicate) 映像と呼ばれる極めて類似した映像区間を、映像アーカイブから高速に全て検出する技術¹¹⁾を用いて、その出現パターンを可視化した(図2)¹²⁾。このインターフェースは、指定した番組中に出現する全ての準同一映像の一覧を表示する(図2左下部)ほか、指定した準同一映像の時系列(横方向)、放送局間(縦方向)の出現パターンを表示する(同上部)。図2は、東京地区地上波6局(横方向に伸びる帯が各放送局の映像)において3日間にわたり、特定のCM映像が出現する(帯中の太線)パターンを表示した例である。

これらの構造化・可視化技術は、直感的には把握しにくい大規模映像アーカイブ中の映像データ間の関係を利用者に分かりやすく提示するとともに、構造に基づいて検索対

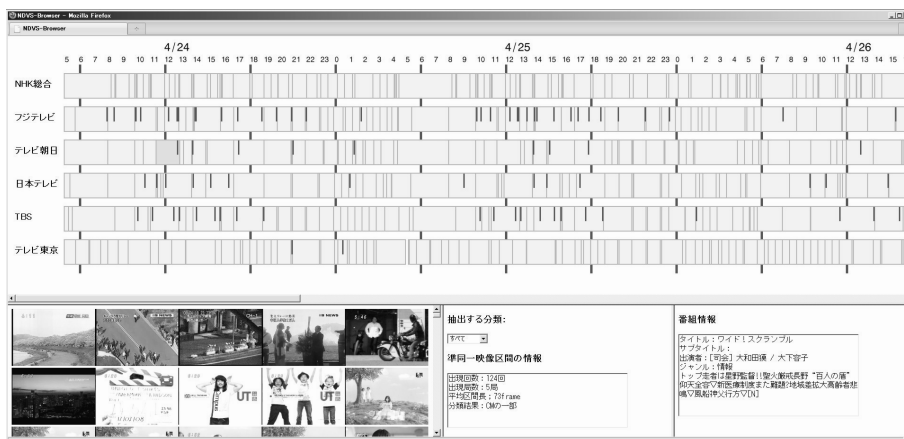


図2 準同一映像の出現パターン可視化インターフェース¹²⁾

象を絞り込めるようになる点で重要である。

3.3 ニュース映像の例

一般にニュース報道では、いわゆる 5W1H (When, Where, Who, What, Why, How) が重要と言われていいる。これをふまえて、ニュース映像の構造化・内容解析においても、これらの属性に注目した研究が多い。ここでは、このなかでも特に、Where (場所)、Who (人物) に注目した研究を紹介する。なお、When (時間) については、ニュース映像に関する多くの研究で、暗黙のうちに構造化に組み込まれている。

(1) 場所に注目した構造化 前出の Informedia プロジェクト²⁾において開発されたニュース映像閲覧インターフェース「Video Collage」¹³⁾の「Map Collage」インターフェースでは、字幕放送テキスト中に出現する地名を手がかりにして地図上に映像のサムネイル画像を配置し、地理的範囲を指定した検索を実現した。

(2) 人物に注目した構造化 佐藤らが開発した「Name-It」¹⁴⁾は、映像中に登場する人物の名前を字幕放送テキストから抽出し、顔画像と自動で対応づけ、人物名から顔画像を、顔画像から人物名を、と双方向に検索できるようにした。また、大量の映像から、人物名と顔画像の組を大量に収集し、顔の類似度に基づいてより正確に人物名と顔画像を検索できるようにした研究¹⁵⁾もある。

一方、筆者らは、登場する人物名の共起に基づく人間関係の強さに基づいて映像を閲覧するインターフェース「trackThem」(図 3)¹⁶⁾を開発した。このインターフェースでは、指定した人物を中心に放射状に配置された、関係が強い人物を次々と選択していくことで人間関係をたどりながら(図 3 左上部)、指定した二人の人物が共起した映像を閲覧することができる。

一方、特に 5W1H に注目しない構造化も提案されている。例えば、Duygulu ら¹⁷⁾や Wu ら¹⁸⁾は、準同一映像や同一ロゴの出現など、画像特徴の共起を手がかりに、映

像を時系列に連ねた構造を構築し、特定のトピックを追跡しながら映像を閲覧できるようにした。また、瀧本らは、記者会見などにおけるフラッシュの発光パターンが撮影位置によらないことを利用して、異なる放送局の映像に出現する同一イベントを関連付けて検索する手法を提案した¹⁹⁾。

一方、筆者らは、字幕放送テキスト中に出現する単語の類似度に基づいて、複雑なトピックの変遷過程を追跡しながら映像を閲覧するインターフェース「mediaWalker」を開発した(図 4)²⁰⁾。このインターフェースにより、膨大な映像の中から、興味があるトピックおよびそれから派生した別のトピックについて、個々の映像を検索することなく、芋づる式にたどって次々に閲覧することができる。

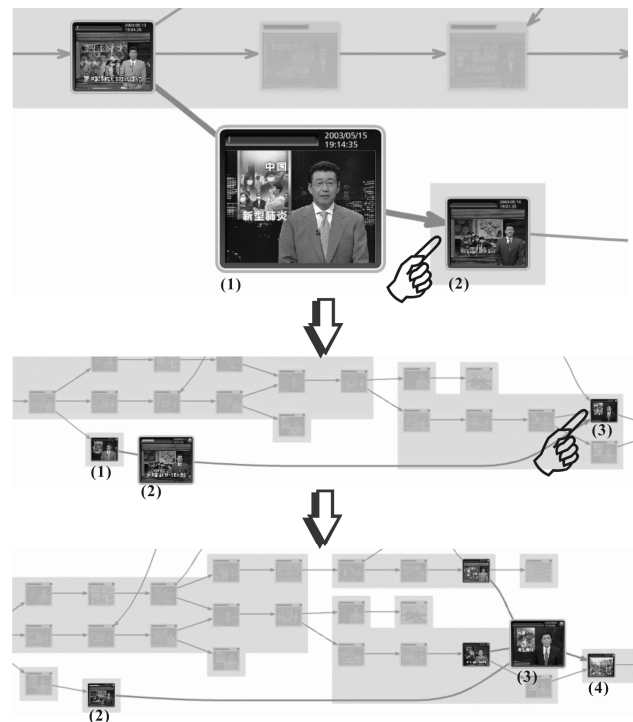


図 4 mediaWalker インターフェースによりニューストピックの時間的変遷をたどりながら映像を閲覧する様子²⁰⁾

なお、ニュース映像検索の詳細については、4.1 のほか、文献²¹⁾を参照されたい。

3.4 料理映像の構造化の例

筆者らは料理映像を対象とし、効率的に調理過程を閲覧することを想定した構造化に取り組んだ。具体的には、調理という対象をふまえ、手元が大写しになる「手元シーン」において、動作に注目して重要箇所を検出する方法を提案し、閲覧インターフェース(図 5)を開発した²²⁾。ここで、重要箇所として、調理動作に多く含まれる反復動作(図 5 中▲印)、フレーム全体の激しい動き(同■印)、料理の状態が提示されれば動きがない(同●印)箇所を検出した。これにより短時間で料理映像を閲覧することができるようになり、大量の料理映像について、内容を視覚的に把握し



図 3 trackThem インターフェース¹⁶⁾



図5 動きに基づく料理映像の要約閲覧インタフェース²²⁾

ながらレシピを検索できるようになった。

4. 関連情報による構造化と内容解析

2.2 で挙げたようなさまざまな関連情報と関連付けた構造化について紹介する。個人が撮影した映像 (CGM : Consumer Generated Media) や監視映像などと異なり、放送映像においてはこれらの関連情報を利用できることが多い。効率的かつ的確な内容解析のために、手がかりとすることが多い。

4.1 ニュース映像の例

ニュース映像と外部情報源を関連づける研究は、初期の新聞記事への関連づけに関する研究²³⁾に始まり、様々なウェブページと関連づけ、双方向に閲覧できるインタフェース²⁴⁾などがある。

筆者らも、言語や字幕放送テキストの有無によらず、準同一映像の出現を手がかりに、他言語によるニュース映像と関連づける方法²⁵⁾や、ニュース映像を Wikipedia 項目と関連づける方法²⁶⁾に関する研究に取り組んでいる。

前者は、日本の映像と、前出の TRECVID⁷⁾の配布データから得た米国、中国、台湾、レバノンにおいて他言語で放送された映像について、前後 24 時間以内のものを照合し、同一素材映像を編集して放送された準同一映像を検出することで、同一イベントに関する映像を検索する方法である。機械翻訳可能な字幕放送テキストがある場合のみテキストの類似度も併用して、この結果と組み合わせて検索できるようにした閲覧イ

ンタフェース「mediaTraveller」を開発した (図 6)²⁵⁾。日本の映像 (図 6 左側) に対応する米国の映像 (同右側「同一ニュース映像」2 つ目) が検索されたことが分かる。この場合はテキストの類似度も加味されたが、中国やレバノンの映像も、準同一映像の出現のみを手がかりに検出された (検索結果の下位に存在するため、図 6 には表示されていない)。

この方法により、同一の映像素材が利用されている限り、字幕放送テキストの有無、音声認識・機械翻訳の有無や精度、さらに、各国において異なる視点で放送され、音声・テキスト上はあまり似ていないなどの差異によらず、同一イベントに関する映像を検索できるようになる。

一方、後者は、映像の字幕放送テキストと、全ての Wikipedia 記事のテキストを照合し、類似度が高い Wikipedia 項目を映像に対応づける。これにより、Wikipedia 項目で正規化された索引を映像に付与でき、検索性が向上する。現在、この結果を利用して、3.3 で紹介した mediaWalker インタフェース中の各映像から Wikipedia を始め、様々なウェブページを閲覧できるように改修しているところである²⁷⁾。

また、逆に Wikipedia 記事に関連映像を付与することができる考えると、既存のテキストによる記述を映像情報で補えるようになることもできる。



図6 準同一映像区間の出現を手がかりにした言語横断ニュース映像検索インタフェース「mediaTraveller」²⁵⁾

4.2 料理映像の例

一般に料理番組にはレシピを掲載したウェブページや教材のテキストが販売されている。筆者らは、レシピ中の調理手順と料理映像を対応づけることで、映像を調理手順毎

に分割し、レシピに沿って構造化する手法を提案し、得られた映像の構造を可視化しながら、実際に調理を支援するインタフェースを開発した²⁸⁾。この構造化手法は、文字放送字幕中の出現単語と画像特徴を組み合わせ、木構造で表される調理手順(図7右上部)中の各手順と映像を対応づける。これにより、映像を最初から最後まで通して見ずとも、分かりにくい箇所のみ閲覧しながら調理することができるようになる。

5. おわりに

本稿では、映像アーカイブの中でも中心となる放送映像を想定し、主に内容解析およびそのための構造化の視点から、さまざまな研究事例について、外部情報源との関連づけの有無に分けて、ジャンル別に紹介した。このほか、スポーツ映像を対象とした構造化・内容解析に関する研究²⁹⁾も活発に行われてきた。一方、本稿で言及しなかった話題として、大規模な映像アーカイブを対象とした高速な映像間照合・検索に関する研究や、画像・映像そのものをクエリとした検索に関する研究がある。また、近年急速に増加している監視カメラ映像について、相互の地理的配置も考慮して複数の映像にまたがった統合的な検索技術についても研究が進んでいる。これらは、今後重要になるとと思われる技術だが、現時点ではまだ一般利用者を想定したものではないことから、割愛した。

さらに、近年急速に普及しているインターネット上の動画投稿・共有サービスも大規模な映像アーカイブとみなすことができるため、今後は従来型の統制されたものだけでなく、統制されていない映像アーカイブを対象とした研究も重要になるとと思われる。

謝辞

本稿で紹介した研究のうち、筆者が関係するものは名古屋大学の村瀬洋教授、出口大輔助教および研究室の学生諸氏、岐阜聖徳学園大学の高橋友和准教授、国立情報学研究所の佐藤真一教授、片山紀夫准教授、孟洋助教ら、東京大学の田中英彦教授(現在、情報セキュリティ大学院大学教授)、坂井修一教授および後輩諸氏との共同研究の成果である。

参考文献

- 1) 江原学. 特集, デジタルアーカイブ: 放送映像アーカイビング. 映像情報メディア学会誌. 2007, vol.61, no.11, p.1567-1572.
- 2) 金出武雄, 佐藤真一. 特集, デジタル図書館: Informedia:



図7 料理映像と教材レシピとの対応付けによる調理支援インタフェース「Happy Cooking」²⁸⁾

CMU デジタルビデオライブラリプロジェクト. 情報処理. 1996, vol.37, no.9, p.841-847.

- 3) 佐藤真一. 特集, デジタルアーカイブ: 映像解析研究のためのアーカイブ. 人工知能学会誌. 2006, vol.18, no.3, p.230-235.
- 4) 総務省. 視聴覚障害者向け放送普及行政の指針. 平成 19 年 10 月 30 日. http://www.soumu.go.jp/menu_news/s-news/2007/071030_2_bs1.html [accessed 2010-8-19].
- 5) 総務省. 平成 21 年度の字幕放送等の実績. 平成 22 年 9 月 3 日. http://www.soumu.go.jp/menu_news/s-news/01ryutsu05_01_000002.html [accessed 2010-09-15].
- 6) 井手一郎, 柳井啓司. セマンティックギャップを越えて—画像・映像の内容理解に向けて—. 人工知能学会誌. 2009, vol.24, no.5, p.691-699.
- 7) 佐藤真一. 映像内容解析における TRECVID の取組み. 電子情報通信学会誌. 2008, vol.91, no.1, p.55-59.
- 8) 柳井啓司. 一般物体認識の現状と今後. 情報処理学会論文誌: コンピュータビジョンとイメージメディア. 2007, vol.48, no.SIG16(CVIM9), p.1-24.
- 9) 佐藤隆, 児島治彦, 阿久津明人, 外村佳伸. 映像コーパスの構築と分析. 電子情報通信学会論文誌 (D-II). 1999, vol.J82-D-II, no.10, p.1552-1560.
- 10) 久保田英俊, 桃崎浩平, 青木恒, 風間久. 映像シーンを簡単に検索できる“顔 de ナビ”. 東芝レビュー. 2008, vol.63, no.11, p.54-57.
- 11) Ide, I.; Suzuki, S.; Takahashi, T.; Murase, H. “Adaptive division of feature space for rapid detection of near-duplicate video segments”. Proc 2009 IEEE Int. Conf. on Multimedia and Expo, New York, NY, USA, 2009-6-28/7-3. 2009, p.694-697.
- 12) Ide, I.; Shamoto, Y.; Deguchi, D.; Takahashi, T.; Murase, H. “Classification of near-duplicate video segments based on their appearance patterns”. Proc. 20th IAPR Int. Conf. on Pattern Recognition. Istanbul, Turkey, 2010-8-23/26. 2010, p.3129-3133.
- 13) Christel, M.G.; Hauptmann, A.G.; Wactler, H.D.; Ng, T.D. “Collages as dynamic summaries for news video”. Proc. 10th ACM Int. Multimedia Conf. Juan-les-Pins,

- Alpes-Maritimes, France, 2002-12-1/6. 2002, p.561-569.
- 14) Satoh, S.; Nakamura, Y.; Kanade, T. Name-It: Naming and detecting faces in news videos. *IEEE Multimedia*. 1999, vol.6, no.1, p.22-35.
 - 15) Ozkan, D.; Duygulu, P. "Finding people frequently appearing in news". *Proc. 5th Int. Conf. on Image and Video Retrieval*. Tempe, AZ, USA, 2006-7-13/15. Springer-Verlag, 2006, p.173-182, (Lecture Notes in Computer Science, 4071)
 - 16) Ide, I.; Kinoshita T.; Mo, H.; Katayama, N.; Satoh, S. "trackThem: Exploring a large-scale news video archive by tracking human relations". *Proc. 2nd Asia Information Retrieval Symposium*. Jeju, Korea, 2005-10-13/15. Springer-Verlag, 2005, p.510-515, (Lecture Notes in Computer Science, 3689)
 - 17) Duygulu, P.; Forsyth, D.; Pan, J.-Y. "Towards auto-documentary: Tracking the evolution of news stories". *Proc. 10th ACM Int. Multimedia Conf. New York, NY, USA, 2004-10-10/16*. 2004, p.820-827.
 - 18) Wu, X.; Ngo, C.-W.; Li, Q. Threading and aut documenting news videos. *IEEE Signal Processing Mag.* 2006, vol.23, no.2, p.59-68.
 - 19) 瀧本政雄, 佐藤真一, 坂内正夫. 大容量放送映像アーカイブからの同一フラッシュシーン映像の発見. *電子情報通信学会論文誌 (D)*. 2006, vol.J89-D, no.12, p.2699-2709.
 - 20) 井手一郎, 木下智義, 高橋友和, 孟洋, 片山紀生, 佐藤真一, 村瀬洋. 大量ニュース映像を対象とした時系列意味構造に基づく情報編纂手法の提案. *人工知能学会論文誌*. 2008, vol.23, no.5, p.282-292.
 - 21) 井手一郎. 連載講座, マルチメディア検索の最先端: ニュース映像の検索. *映像情報メディア学会誌*. 2010, vol.64, no.3, p.306-311.
 - 22) 三浦宏一, 浜田玲子, 井手一郎, 坂井修一, 田中英彦. 動きに基づく料理映像の自動要約. *情報処理学会論文誌: コンピュータビジョンとイメージメディア*. 2003, vol.44, no.SIG9 (CVIM7), p.21-29
 - 23) 渡辺靖彦, 岡田至弘, 角田達彦, 長尾真. TV ニュースと新聞記事の対応づけ. *人工知能学会誌*. 1997, vol.12, no.6, p.921-927
 - 24) 宮森恒, 馬強, 田中克己. WA-TV: 次世代蓄積型テレビのための番組コンテンツのウェブ化と情報補強. *情報処理学会論文誌: データベース*. 2006, vol.47, no.SIG8(TOD30), p.71-80.
 - 25) 井手一郎, 小川晃, 高橋友和, 村瀬洋. 画像とテキストの利用による同一ニュースの言語横断検出. *画像ラボ*. 2008, vol.19, no.12, p.22-26.
 - 26) Okuoka, T.; Takahashi T.; Deguchi, D.; Ide, I.; Murase, H. "Labeling news topic threads with Wikipedia entries". *Proc. 11th IEEE Int. Symposium on Multimedia*. San Diego, CA, USA, 2009-12-14/16. 2009, p.501-504.
 - 27) 井手一郎, 高橋友和, 木下智義, 奥岡知樹, 出口大輔, 佐藤真一, 村瀬洋. "mediaWalker II: ウェブコンテンツと連携したニュース映像アーカイブ閲覧インタフェース". *画像の認識・理解シンポジウム (MIRU) 2010 論文集*. 北海道釧路市, 2010-7-27/29. 2010, p.1320-1321.
 - 28) 三浦宏一, 高野求, 浜田玲子, 井手一郎, 坂井修一, 田中英彦. 料理映像の構造解析による調理手順の対応付け. *電子情報通信学会論文誌 (D-II)*. 2003, vol.J86-D-II, no.11, p.1647-1656.
 - 29) 新田直子. 連載講座, マルチメディア検索の最先端: ニュース映像の検索: スポーツ映像の検索. *映像情報メディア学会誌*. 2010, vol.64, no.4, p.495-501.

Special feature: Digital archiving of images. Video archive and findability. Ichiro IDE (Nagoya University, Graduate School of Information Science. Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601 JAPAN)

Abstract: Recently, it has become easier to digitally archive a large amount of video data, thanks to the increase in storage densities of storage devices such as HDDs. However, since it requires time to view and also is difficult to describe the contents precisely and thoroughly, it is not easy to obtain a required video as a result of simple retrieval methods based on keywords or image features. This article introduces various examples including the author's work on efficient retrieval for broadcast video archives, from the view-point of video structuring and its contents analysis. It will be introduced according to the association with external information source, and per video genre.

Keywords: video archive / broadcast video / video retrieval / structuring / contents analysis / news video / cooking video