

平成22年7月1日現在

研究種目：基盤研究(C)

研究期間：2007～2009

課題番号：19500102

研究課題名（和文） 放送映像からの映像知識の獲得とその活用に関する研究

研究課題名（英文） Research on acquisition and utilization of video knowledge from broadcast videos

研究代表者

孟 洋 (MO HIROSHI)

国立情報学研究所・コンテンツ科学研究系・助教

研究者番号：60312203

研究成果の概要（和文）：本研究では、多量の放送映像を活用し、概念とその概念を説明するような映像の対（映像知識）を獲得する手法の検討を行った。映像で言及された単語や話題、登場する人物や領域の解析を行い、映像内/映像間におけるこれらの共起などの関連構造に基づき、概念と映像の対応関係の抽出を行った。また、放送番組空間において、概念と画像領域/映像区間、及び画像領域/映像区間同士の検索ができる連想検索システムを試作し、映像知識の探索を可能とした。

研究成果の概要（英文）：In this study, I tried to extract "video knowledge", a set of pairs of a concept and a video segment that explains the concept, from a huge collection of broadcast videos. It was realized by finding out temporal coincidences of text keywords and image regions in the video. In order to retrieve the video knowledge from broadcast video archives, I built an associative search system that can find candidates of the pairs in a video archive.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,200,000	360,000	1,560,000
2008年度	1,300,000	390,000	1,690,000
2009年度	600,000	180,000	780,000
年度			
年度			
総計	3,100,000	930,000	4,030,000

研究分野：情報工学

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：マルチメディア情報処理、映像解析、映像情報システム

## 1. 研究開始当初の背景

放送番組は社会生活と密接に関連した多くの情報を保持している。また、映像は文字や画像だけでは表現できない、情景や動作な

ど動的かつ有用な情報を内在し、直感的かつ効果的な情報の伝達を可能とする。このため、従来、視聴という形でしか利用されていなかった放送映像を価値ある事例知識として再

活用することができれば、人間の知的活動を支援する非常に有用な知識形態の一つとなれる。

映像検索という視点からは様々な研究が行われているが、文字情報のみに基づくものが多かった。多くの実例を含む映像アーカイブを構築、活用できるようになったことから、従来困難であった映像情報に基づく構造化や統計処理を実現し、単なる検索とは異なる形で、事例知識となりえる、概念と概念を説明する映像の対の発見が可能になると期待できる。

## 2. 研究の目的

本研究では、日々絶え間なく提供されている放送映像を活用し、「国会議事堂」とはどのような建物か、「首相」とはどのような人物かなど、概念をあらわすキーワードとその概念を説明する映像セグメントの対（「映像知識」と呼ぶ）を自動的に獲得する手法について検討を行うことを目的とする。

また、「映像事典」の構築を目指し、映像知識を整理・体系化する手法、そして映像情報システムへの応用についても検討を行う。

## 3. 研究の方法

映像アーカイブに蓄積された大量の放送映像と映像に付随する文字放送テキストの解析、構造化を行うことで、映像知識の獲得を目指す。本研究は以下の手順で実施した。

### (1) 放送映像アーカイブの構築

放送番組を知識源として活用するため、長期に渡る放送番組の、映像、文字放送テキスト、及び電子番組ガイドの情報を蓄積する。

- 特定ニュース番組（数年分）
- 東京地区地上波全番組（一月分）

### (2) 概念キーワード候補の取得

発話内容が記述された文字放送テキストから形態素解析に基づき構成単語、つまり概念キーワード候補を抽出し、語の品詞や映像内/映像間での出現分布を評価することで、語の重要度を判定する。

### (3) 映像セグメント候補の取得

映像は、フレーム、ショット、シーン、トピック、番組など階層的に構成される。概念を説明するのに適する映像セグメントの粒度は変化するため、画面の連続性から基本単位をショットとし、画像情報からは同一オブジェクトの存在範囲の解析、文字情報からはトピックの境界の解析をとおして、階層的な映像セグメント構造を抽出した。図1にトピック境界の抽出手順を示す。

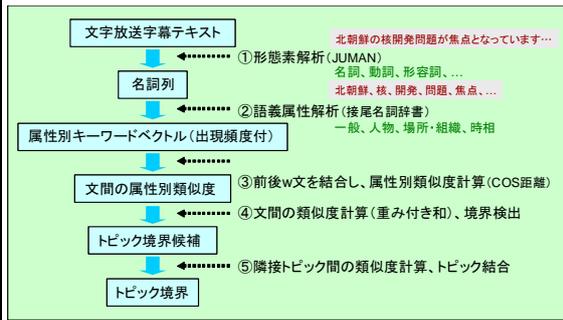


図1 トピック境界の抽出

### (4) 映像特徴の抽出とビジュアルワード化

映像知識の取得のためには、概念と対応する可能性のある映像要素（映像特徴）を抽出し、概念との関連度を評価する。映像をそのままを扱うことは困難であるため、映像要素をビジュアルワード化（記号化）し、映像をその集合体として表現する。ここでは映像要素として、以下のものを利用した。

- 映像ショット（同一映像ショット）
- 顔画像（同一人物）
- SIFT 特徴（類似画像領域）

これにより、概念と映像セグメントを結び付けるにあたり、繰り返し放送されるショット、よく登場する人物、そして類似した画像領域が考慮されることになる。

映像要素のビジュアルワード化は、映像特徴空間上での分布に基づいてクラスタを構成し、個々のクラスタを一つの記号に置き換えることで実現する。顔画像の場合、理想的には個々の人物にそれぞれの人物をあらわす一つの記号が割り当てられることになる。

各映像セグメントは、文字放送テキストから得られる文字単語と、映像要素から得られるビジュアルワードの頻度付ベクトルとして表現することが可能になる。図2に映像情報の記号ベクトル化の枠組を、図3に顔画像の記号化の手順を示す。

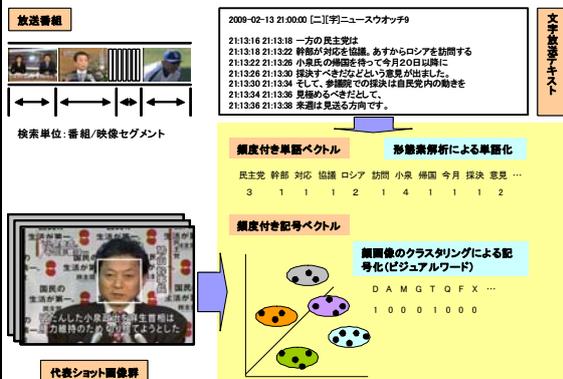


図2 映像情報の記号ベクトル化  
(単語とビジュアルワードによる表現)

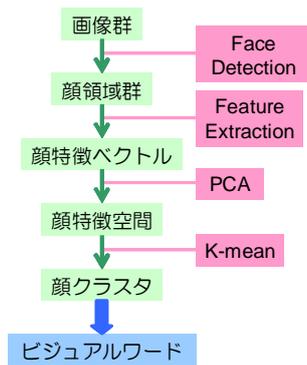


図3 顔画像の記号化

(5) 概念と映像の対応付け

長期間にわたる多量の放送映像を扱う場合、同一あるいは関連する話題や概念を取り上げた映像が複数存在する可能性が高い。このとき、もしある概念を説明する映像要素があると仮定すると、概念キーワードと映像要素の間に共起関係が存在すると考えることができる。このため、文字放送テキストから得られる文字単語と、映像要素をあらわすビジュアルワードの出現パターンを評価することで、概念キーワードと映像要素を結びつけることが可能になる。

映像は時間軸を持つことから、共起を評価する場合、映像要素が画面にあらわれる時間範囲と文字放送テキストが内容をあらわす時間範囲を考慮する必要がある。例えば、ニュース映像においては、映像要素はショットの幅で、文字放送テキストはトピックの幅で有効と考え、共起を評価するのがよい。図4にニュース映像における映像情報と文字情報の対応関係を示す。

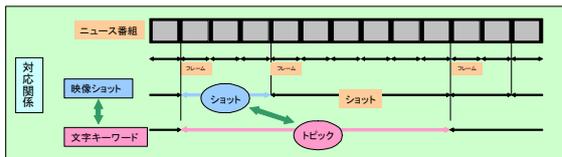


図4 映像情報と文字情報の対応関係

4. 研究成果

本研究では、多量の放送映像を活用して、文字放送テキストから得られる文字単語と映像から得られるビジュアルワードの共起関係を評価することで、映像知識となりえる、概念と映像の対を発見することができることを示した。

また、放送映像空間を探索することで、概念、あるいは要求にあった放送番組や映像セグメントの発見を可能とする放送番組の連想検索システムを試作した。

(1) 概念と映像の対応付け結果

放送映像アーカイブに提案手法を適用することで、映像セグメントに概念キーワードを自動で付与することが可能になった。文字単語とビジュアルワードの共起関係に基づき、人物映像にキーワードを付与した例を図5に示す。



図5 人物映像に付与されたキーワード

(2) 放送番組の連想検索システム

放送映像を活用するための応用システムの一つとして、様々な放送番組から作られる情報空間を、連想検索機能により利用者が自由に渡り歩くことができるシステム「番組マップ」の試作を行った。概念から関連する画像領域/映像区間の検索を行えるのをはじめ、画像領域/映像区間同士の検索も可能である。このシステムにより、利用者は映像空間を探索し、映像知識となりえる、概念と概念を説明する映像の対の候補を発見することができる。

連想検索は「文書群を特徴付ける単語群」、及び「単語群に関連する文書群」を求める二つの基本計算を組み合わせることで行う。検索対象となる文書は、頻度付き単語ベクトルとして表現し、類似尺度は二つのベクトルの距離として定義する。放送番組の場合、番組、話題あるいはショットなど検索単位となる映像セグメントを一つの文書として捉え、それらの画像と文字それぞれの情報から特徴ベクトルを求める。これは3-(4)で述べた文字単語とビジュアルワードの頻度付きベクトルに相当する。連想検索の枠組を図6に示す。

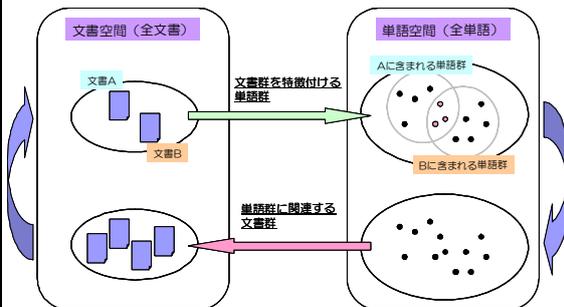


図6 連想検索の枠組

番組マップでは、探したい映像/番組を自由文章で指定することができる。一度検索を行うと、検索された映像/番組の一覧、推定される主要な文字単語、及びビジュアルワードの一覧が表示される。ここで、興味のある情報を選択し、再検索を繰り返すことで、番組空間の探索が可能になり、必要な映像/番組を得ることができる。検索インターフェースの画面を図7に示す。また、映像/番組を閲覧する場合は、番組名、代表画像一覧、文字放送テキスト一覧を表示することができ、指定箇所からの再生も可能である。

試作システムでは、地上派で放送された番組のうち、一部のニュースとドキュメンタリー、数年分が検索可能になっている。このとき、例えば「アメリカ大統領」で検索をかけると、「顔 (オバマ氏相当)」が主要なビジュアルワードとして推定され、関連する番組、例えば就任式の映像/番組などを選び出すことができる。推定されたビジュアルワード(顔)の例を図8に示す。



図7 検索インターフェース



図8 ビジュアルワード(顔)の例

映像知識として適する最適な映像区間の決定手法、及び体系化手法について十分な考察を行うことができず、目標としていた映像事例集、さらにはこれらを体系的に整理した映像事典の構築までには至らなかったが、体系化の戦略を組み込み、構築したシステムを更に発展させることで、これらは実現できるものと考えている。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

- ① 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, vol.23, no.5, pp.282-292, 2008.09

〔学会発表〕(計3件)

- ① 孟洋, 佐藤真一, 画像と文字の両情報を活用した放送番組の連想検索システムの構築, 第14回映像メディア処理シンポジウム, I-6-04, 2009.10
- ② IMAGINE - Federated Associative Search for Spontaneous Learning, Akihiko Takano, Shingo Nishioka, Yuzo Marukawa, Yuji Koike, Takeshi Morimoto, Shin'ichi Satoh, Norio Katayama, Hiroshi Mo, Fuminori Yamagishi, The 3rd Pacific-Rim Symposium on Image and Video Technology (PSIVT2009), 2009.01
- ③ 孟洋, 山岸史典, 小池勇治, 西岡真吾, 佐藤真一, 放送番組の連想的アクセス環境「番組マップ」の構築, 第13回映像メディア処理シンポジウム, I-3-02, 2008.10

## 6. 研究組織

(1) 研究代表者

孟洋 (MO HIROSHI)

国立情報学研究所・コンテンツ科学研究系・助教

研究者番号: 60312203

(2) 研究分担者

なし

(3) 連携研究者

なし