

映像の同一性に着目した同一ニュースイベントの言語横断検索

小川 晃[†]高橋 友和[†]井手 一郎^{†‡}村瀬 洋[†][†]名古屋大学大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町[‡]国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-1

1. はじめに

近年、ニュース映像のアーカイビングが重要となり、大容量のニュース映像を効果的に再利用するために話題の検索や追跡といった技術が必要とされている。また、最近では様々な観点からの情勢理解のために国内だけでなく海外のニュース映像を視聴する機会が増えている。そのため、それら映像も検索の対象として有効利用することが求められている。

本研究では、放送国の異なる様々なニュース映像から同一イベントを検索するための技術に注目する。現在、一般にニュースの検索では新聞記事検索などを拡張し、音声の書き下しやテロップなどを用いたテキストのみによる手法が用いられている。しかし、言語横断型の検索では機械翻訳の性能や放送国間の文化的背景の違いなどにより、同一言語・放送局による映像のみを対象にする場合に比べて検索は困難になる。そこで我々は、同一イベント間の映像の同一性に着目し、テキストの類似性に加えてニュース映像間の同一映像区間の存在を用いることで同一ニュースイベントの検索性能を向上することを目指す。

以降、2章で事前処理となるニュースイベントの分割について説明し、3章で同一映像区間の検出方法を述べる。4章で実験とその結果を示し、5章で本研究をむすぶ。

2. ニュースイベントの分割

本研究では照合処理をイベント単位で扱うので、事前処理としてイベント分割を行なう。イベントとは、“ある日ある場所で起こった出来事”のことをいう。一般にニュース映像は複数のストーリーからなることに注目し、本研究では1ストーリーを1イベントとして扱う。以下の実験

では、文字放送字幕(Closed Caption; CC)テキストやアンカーショットの出現などを利用した文献[2, 3]のような既存手法を用いてストーリー分割を行なった結果を利用する。

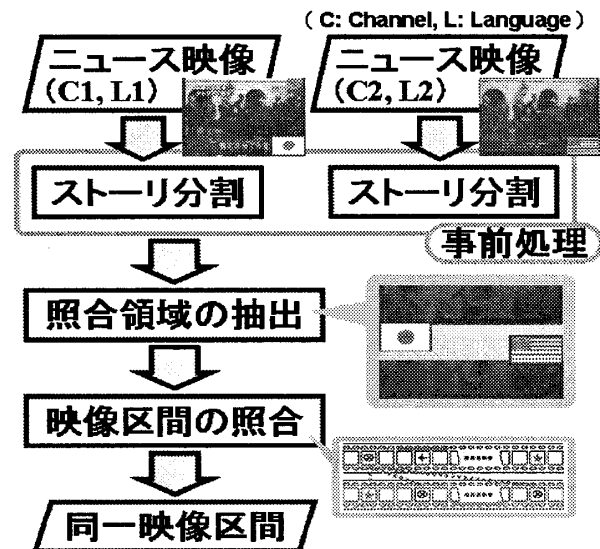


図 1. 同一映像区間の検出手順

3. 同一映像区間の検出

3.1. 照合領域の抽出

ニュース映像には各番組で独自のテロップを挿入することが多く、この影響により素材映像が同一でも映像間の類似度が低下する可能性がある。そこで、本研究ではテロップが挿入されにくい映像の中心部分のみを照合領域として用いる。また、特徴量として RGB 値を用いるため、放送局や録画状況における映像の明るさの違いによる影響も照合時の問題となる。そこで、映像中の各照合領域の輝度値に対してヒストグラム平坦化を施す。

3.2. 映像区間の照合

本研究では、ニュース映像中から同一映像区間を検出する手法として、特徴次元圧縮を用いた 2 段階照合における高速検出手法[1]を用いる。まず、映像照合における計算の高速化のため、各区間映像の特徴ベクトルに対して空間方向(フレーム内)と時間方向(複数フレーム間)の 2 段階の次元圧縮を施す。次に、次元圧縮により低次

Cross-lingual retrieval of identical news events referring to near-duplicate video segments

Akira Ogawa[†] Tomokazu Takahashi[†] Ichiro Ide^{†‡} Hiroshi Murase[†]

[†]Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan

[‡]National Institute of Informatics, 2-1-1, Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan

元化された特徴ベクトルを用いて、照合映像区間の組合せに漏れがないよう参照する映像区間の位置をずらしながら繰り返し照合をすることで同一映像区間の候補を高速に検出する。最後に、候補として得られた映像区間に対し、次元圧縮前の高次元特徴ベクトルによる詳細照合を行い、同一映像区間を正確に検出する。

4. 実験

4.1. 実験対象

2004年11月に放送された日本(NHK:News7)とアメリカ(CNN:NewsNight)のニュース映像を対象に実験を行った。テキストによる検索との比較・併用を行うため、ニュース映像はCCテキストが存在するものを選択し、日本のニュース映像(30分)1本に対し、その前後24時間以内に放送されたアメリカのニュース映像(60分)2本を1つのグループとし、グループ内の映像間でのみ同一イベントの検出を行なった(図2)。これは日米間の時差やニュース伝達の時間を考慮した制約である。以下の実験では10個のグループを用いた。

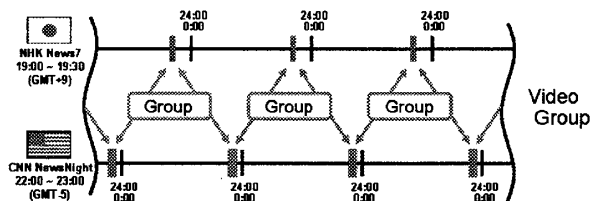


図 2. 日米ニュース映像の放送時間の関係

4.2. 比較実験：テキストによる同一イベント検出

比較のためにテキストのみによる検出を行なった。テキストソースとして各ニュース映像のCCテキストを用い、英文CCテキストは翻訳ソフトウェア「The 翻訳プロフェッショナル v10」[東芝]により和文テキストに変換して用いた。実験方法は、イベント毎にCCテキストから特定の属性をもつ語をキーワードとして抽出し、イベント間の出現頻度ベクトルの内積をテキストの類似度として同一イベントを検出した。

4.3. 実験結果

映像の同一性(画像情報)のみ、テキスト情報のみを用いた同一ニュースイベント検出の結果は表1のようになった。また、各検出処理から画像・テキスト各々でイベント間の類似度を求め、それらを加重和により組み合わせた類似度を用いた検出の結果も表1に示す(画像+テキスト)。なお、テキストの類似度はキーワードの出

現頻度ベクトルの内積、画像の類似度はイベント間で共有する同一映像区間の時間的割合によった。この結果より、画像情報を用いることでテキスト情報では検出できなかった同一イベント対を検出可能となり、両情報を併用することで同一イベント検出の精度を高めることができることが確認された。

表 1. 同一ニュースイベント検出の実験結果

実験条件	画像情報のみ	テキスト情報のみ	テキスト+画像
適合率	72% (13/18)	44% (20/45)	83% (24/29)
再現率	52% (13/25)	80% (20/25)	96% (24/25)

5. むすび

本研究では、言語が異なるニュース映像間から同一イベントを検索する技術に注目し、テキストによる同一イベント検出の不十分さを補うために同一映像区間の存在を利用する検出手法を提案した。従来のテキストによる手法では機械翻訳性能の問題や視点・文化的背景の違いにより検出が困難であった同一イベント対を、同一映像区間検出を用いて対応付けることで検出することが可能になった。

今後は、映像の同一性とテキストの類似性を効果的に併用するための手法を検討し、より高性能な同一ニュースの言語横断検索システムの作成を目指す。

謝辞

本研究の一部は科研情報爆発による。また、実験のデータとして使用したニュース映像を提供して頂いた国立情報学研究所、米国 National Institute of Standards and Technologies による TREC Video 2005 ワークショップに感謝する。

参考文献

- [1] 野田和広ら：“適応的特徴選択を用いた長時間放送映像からの高度な繰り返し区間検出”，信学技報, PRMU2005-289, Mar. 2006.
- [2] 井手一郎ら：“大規模ニュース映像コーパスの意味構造解析”，信学技報, PRMU2003-97, Sept. 2003.
- [3] Yun Zhai et al.：“Story segmentation in news videos using visual and text cues”，Image and Video Retrieval, Fourth Int. Conf., CIVR2005, pp. 92-102, July 2005.