同一映像区間を手がかりとした同一ニュースイベントの言語横断検索

小川　晃[†]　野田　和広[1,†]　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　〒 464–8603 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 2–1–2
E-mail: †{aogawa,knoda,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp

あらまし　近年、HDD の進歩などにより映像資源を大量に蓄積し、利用することが可能になり、それら映像資源の再利用・検索などの技術が望まれている。特に、重要性や利用価値の高さから、ニュース映像における同一イベント検索に対する期待は大きい。現在、ニュース映像の検索には一般にテキスト情報のみによる手法が用いられているが、言語横断型検索においては機械翻訳性能の問題のほか、視点の違いに対処するための高度な自然言語理解が必要となる。そこで、同一イベント検索のためにニュース映像中の画像情報に注目する。ニュース映像間で同一素材映像が使用される場合、それらは同一、または関係の深いニュースイベントである可能性が高い。本研究では、画像情報からニュース映像間の同一映像区間を検出することで、テキスト情報による同一ニュースイベントの言語横断検索を補完することを目標とし、テキスト情報による検出と比較することで画像情報の有効性を示した。
キーワード　言語横断検索, 映像検索, 放送映像, 同一ニュースイベント

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

Akira OGAWA[†], Kazuhiro NODA[1,†], Tomokazu TAKAHASHI[†], Ichiro IDE[†,††], and Hiroshi MURASE[†]

† Nagoya University Graduate School of Information Science　Furo-cho, Chikusa-ku, Nagoya, 464–8603 Japan
†† National Institute of Informatics　2–1–2, Hitotsubashi, Chiyoda-ku, Tokyo, 101–8430 Japan
E-mail: †{aogawa,knoda,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp

Abstract　Recently, video resources are accumulated in large quantities by the progress of storage media Technologies to reuse and search these video resources, are desired. Especially, identical event detection in the news video is important. Currently, text information is generally used for the retrieval of news video. However, advanced natural language understanding is required for dealing with the difference of viewpoint, besides the problem of machine translation performance is needed in cross-lingual retrieval. In this paper, we make use of images information for the task. When an identical source video is used in different news programs, it is highly probable that, they discuss the identical or a closely related news event. We aim at complementing cross-lingual retrieval of identical news events by text information referring to the existence of identical video segments.
Key words　Cross-lingual retrieval, video retrieval, broadcast video, identical news events

1. はじめに

1.1 背景と目的

近年、HDD の進歩などにより映像資源を大量に蓄積して利用することが可能になっている。それに伴い、それら大容量の映像資源を要約するための技術や効率的に検索して利用するための技術が求められている [1]〜[4], [7]。なかでも、重要性や利用価値の高さから、ニュース映像における同一イベント検索に対する需要は高いと考えられる。

現在、このような検索処理は新聞記事などにおけるテキスト情報のみを用いた従来手法の拡張により行われている。これらの手法は同一言語のニュース映像間では比較的容易に実現する

(注1)：現在、(株) デンソー

ことができ、検索性能も高いものになっている。しかし、映像資源としては国内のものだけでなく海外の放送映像も存在しており、それら映像資源も再利用や検索などに有効利用することが望まれる。これに対して従来通りテキスト情報のみによる手法で検索する場合、機械翻訳の難しさや語の多義性などに対処するための高度な自然言語理解の必要性や、同一ニュースイベントに対しても国や放送局による視点の違いがあり、一般的に同一言語を対象としたときより検索性能が低下してしまう。

そこで本研究では、複数言語間での同一ニュースイベント検索のための情報として、テキスト情報のほかにニュース映像中の画像情報を利用することにより、検索性能の向上を目指す。ニュース映像では、長期間に及びある特定の話題を報道する際に、同一映像を象徴的に使いまわす傾向がある。また、海外の話題においてはその国の放送局などから映像の提供をうけることも多々あり、稀少な映像は放送局・国を問わずに配信されて放映される。これらの性質を考慮すると、異なるニュース映像間で同一映像が含まれる場合には、それらニュース映像間には何らかの関連性があり、同一ニュースイベントを扱っている可能性があると考えられる。また、そうでないとしても探索範囲を絞り込むことに大きく効いてくる。

以上のことより我々は、複数言語のニュース映像間から同一映像区間を検出することで、同一ニュースイベントの検出を試みた。検証にあたっては、同一映像区間を含むニュースストーリ間の文字放送字幕テキストの類似性とを比較し、また、性能を比較するために、テキスト情報のみにより複数言語のニュース映像間からの同一ニュースイベント検出も試みた。このようにして、画像情報・テキスト情報から検出された結果を比較することで、言語横断型の同一ニュースイベント検出において画像情報を用いることの効果と有用性を確認する。

1.2 用語の整理

本研究に関連する用語について整理をする。

本研究において、ニュース映像とは動画像と音声、文字放送字幕テキストの集合体である。ここで、文字放送字幕テキストとは、映像中の音声を書き下したテキストであり、クローズドキャプション (Closed Caption) とも呼ばれる。

また、本論文中において、映像の画像的構成は

- フレーム：動画像の最小構成単位である静止画像
- ショット：画像的に連続するフレーム群

となっているとし、情報的構成は

- ストーリ：映像を話のつながりで分割した単位
- イベント：ストーリにおける具体的な出来事

となっているとする。

2. 関 連 研 究

テキスト情報を用いたニュース映像の検索・追跡に関連した研究は、テレビニュース映像と新聞記事の対応付けによる双方向検索 [1] や、文字放送字幕テキストを用いたトピック分割・追跡・スレッド抽出手法 [2] など、多くなされている。しかし、そのほとんどは日本語のみを対象にしたものであり、これらを応用して複数言語間を対象とした言語横断型検索に適用するとしても、機械翻訳性能や視点の違いに関する問題から、性能は低下してしまう。

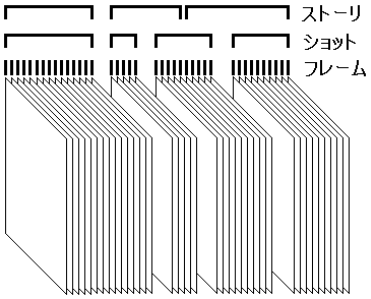

図 1　映像の構成

そこで、本研究ではニュース映像中の動画像に注目し、画像情報を用いてニュース映像の検索を行なうことを提案する。その際、重要だと考えられるのは同一映像の存在である。ニュース映像間で同一映像が存在していれば、そのニュースイベント間には何らかの関係があるので、本手法ではこの同一映像区間を使用する手法を提案する。現在、特定の区間映像に対し、同一映像を長時間映像群から探し出す映像検索手法として時系列アクティブ探索法 [3] をはじめ、いくつかの手法が存在する。しかし、ニュース映像間から任意の同一映像区間を検出する技術はいまだ確立されていない。そこで本研究では、(長時間) ニュース映像中における同一映像区間の高速検出の手法として、我々がこれまで提案してきた主成分分析を用いた特徴次元圧縮による同一区間映像の高速検出手法 [4] を用いることにした。

3. 画像情報による同一ニュースイベントの検出

3.1 手法の概要

図 2 の手順に従い、画像情報による同一ニュースイベント検出を行なう。処理は大きくテキスト処理と画像処理の 2 つに分かれる。テキスト処理では、文字放送字幕テキストを機械翻訳し、その結果をストーリ分割して複数のストーリを取得する。一方、画像処理では、まず、ニュース映像から比較領域を切り出し、明るさに関する正規化を行なう。ここで比較領域とは、テロップなどの影響を受けにくいように設定したフレーム中の特定の領域である。次に、比較領域から特徴ベクトルを抽出し、同一映像区間を検出する。この処理を単純に行なうと計算量が爆発するが、4 章で紹介する特徴次元圧縮を用いた検出手法により高速に実現する。最後に、検出された同一映像区間をストーリへ対応付けることで、同一ニュースイベントを検出する。以下の各節でそれぞれの処理について詳しく説明する。

3.2 テキスト処理

ニュース映像は、視聴者に情報をよりわかりやすく伝えるために、複数のストーリにより構成されている。そこで、ニュース映像をストーリごとに分割し、同一ニュースイベントを検出する単位とする。ニュース映像におけるストーリ分割としては、キャスタショットを用い　た手法や文字放送字幕テキストを用いた手法があるが、以下の実験では日本語ニュース映像に関し

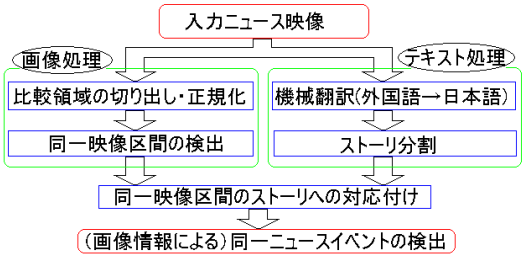

図 2　同一映像区間を手がかりとした同一ニュースイベント検出の処理手順

ては、テキストにおける文間のキーワードベクトルの類似性に基づく映像内構造化手法を用いる。一方、海外のニュース映像に対するストーリ分割は、比較的高性能とされる翻訳ソフトウェア東芝製「The 翻訳プロフェッショナル v10」により、文字放送字幕テキストの全文を日本語に機械翻訳した結果から人手により行なう。

3.3　画像処理

3.3.1　比較領域の切り出し

ニュース映像には各番組で独自のテロップなどが挿入されることがあり、同じ素材映像が使用されているにもかかわらず、番組間で類似度が低下するおそれがある。単純に類似度の閾値を緩くすることも考えられるが、それでは誤検出が増加してしまう。そこで本研究では、比較的テロップ等が挿入されることが少ないフレームの中心部分のみを比較領域とする（図 3）。

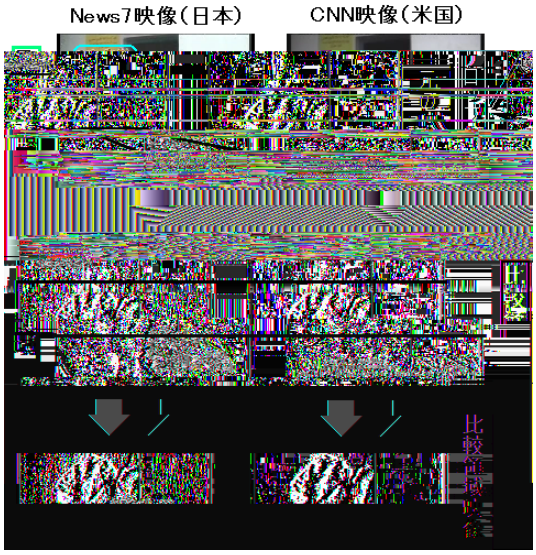

図 3　比較領域の切り出し

3.3.2　比較領域の正規化

ニュース映像の照合時の問題は映像の編集によるものだけでなく、放送局による映像の明るさや色相の違いに関する問題もある。この影響により同一素材映像であっても、画像として同一とみなせないことがある。本研究では、照合に RGB 値を使用しているため、この変化により検出漏れや誤検出が生じる可能性がある。そこで、ニュース映像全体に対し、ヒストグラム平坦化 [5] によって色相の正規化を行なうことで、この問題に対処する。

ヒストグラム平坦化の処理は、以下の手順で行なう。

（1）　フレーム画像の RGB 値に対してそれぞれ独立したヒストグラムを作成

（2）　作成した各ヒストグラムから、RGB 値それぞれについて値の小さい画素順にランク（番号）付け

（3）　各画素において、RGB 値のそれぞれ作成されたランクに range 値（256 階調なら 256）をかけ合わせ、対象画像の全画素数で正規化

上記によって得られた値が、正規化された新たな画素特徴となる。

3.3.3　同一映像区間の検出

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、特徴次元圧縮による長時間映像中における同一区間映像の高速検出手法 [4] を用いる。この手法は、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元化された特徴ベクトルを用いて、照合映像の組合せに漏れがないよう参照する映像の位置をずらしながら繰り返し照合することで同一映像区間の候補を高速に検出する。そして、候補として得られた映像区間に対し、次元圧縮する前の高次元の特徴ベクトルによる詳細照合を行なうことで、同一映像区間を正確に検出する。

本研究では、正規化した RGB 値 $b_i (i = 1, 2, \ldots, I)$ を変数とした I 次元ベクトルを 1 フレームの画像の特徴ベクトルとして扱う。また、複数フレームによる照合を行なう際には、各フレーム画像の特徴ベクトルを並べたものを照合時の特徴ベクトルとして扱う。

4.　特徴次元圧縮による高速検出手法の解説

前述のとおり、本研究ではニュース映像から同一映像区間を検索するために、我々がこれまで提案してきた特徴次元圧縮による高速検出手法 [4] を用いた。ここでは、この手法の処理手順について簡単に説明する。

4.1　準備処理

長時間映像中の各区間の次元圧縮を行なうために、事前に多くの映像サンプルを訓練データとして収集し、それらの画像からなるベクトルの主成分分析により、次元圧縮の基底となる固有値ベクトルを生成する。次元圧縮は空間方向（フレーム内）と時間方向（複数フレーム間）の 2 段階で行なうので、固有ベクトルもそれぞれに対応した 2 種類を生成する。

4.2　フレームの抽出

入力された映像から全てのフレームを抽出し、それらを低解像度化する。これは、映像中に含まれる細かな雑音を除去すると同時に、入力映像が長時間映像となるため、後の処理における計算量とメモリ領域の削減を目的としたものである。本実験では、16 × 16 画素を平均化して 1 つの画素とし、解像度を 352 × 240pixel から 22 × 15pixel に低下した。

次に、低解像化した各フレーム画像について、画素の RGB

値の平均が 0、ノルムが 1 となるように正規化する。正規化において、全画素の RGB 値をそれぞれ独立した値と考えて計算する。画素数 I_0 の画像の場合は $I = I_0 \times 3$ として次式により RGB 値 $\dot{a}_i (i = 1, 2, 3, \ldots, I)$ の正規化値 a_i を算出する。

$$a_i = \frac{\dot{a}_i - \dot{a}}{\sqrt{\sum_{i=1}^{I} (\dot{a}_i - \dot{a})^2}}; (i = 1, 2, \ldots, I); (\dot{a} = \frac{1}{I} \sum_{i=i}^{I} \dot{a}_i) \quad (1)$$

このように正規化された RGB 値を各フレームの特徴ベクトルの要素とする。

4.3 特徴次元圧縮

前述のとおり、この手法では空間方向と時間方向の 2 段階で、主成分分析による特徴ベクトルの次元圧縮を行なう（図 4）。各段階において、各フレームの特徴ベクトルとして準備処理で作成した固有ベクトルのうち、固有値の大きいものを上位 M_1, M_2 個選んで特徴空間の基底とする。まず、第 1 段階の圧縮は空間方向の圧縮として、映像の各フレームの特徴ベクトルに対しての次元圧縮を行なう。次に、第 2 段階の圧縮は時間方向の圧縮として、第 1 段階の次元圧縮が施されたフレームを時系列に並べ、照合する区間映像の時間（照合窓時間）分まとめた複数フレームの特徴ベクトルを 1 つの特徴ベクトルとして次元圧縮を行なう。なお、区間映像は事前に定めた時間長分のフレーム群を 1 フレームずつずらして切り出し、次元圧縮を繰り返す。

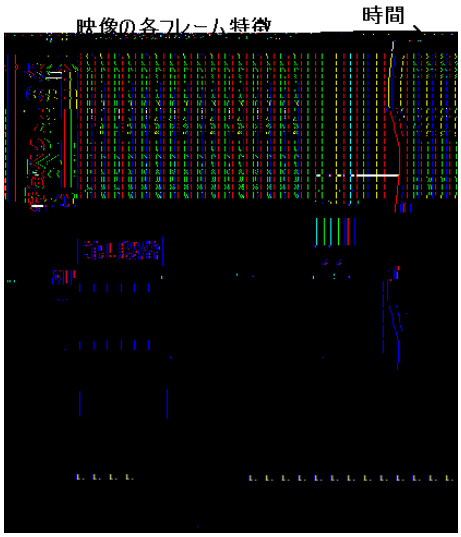

図 4　2 段階の特徴次元圧縮

4.4 低次元繰り返し照合

前節の処理により生成された低次元の特徴ベクトルを用いて、繰り返し照合を行なう。この処理により、入力長時間映像中の同一映像区間の候補を絞込む。

まず、映像の先頭から照合窓時間分の区間映像を参照区間とし、その他の全区間映像と照合する。この処理を参照区間をずらしながら繰り返すことにより、全ての同一映像区間の候補を検出する。ただし、本研究では同一ニュース映像中の同一映像区間を検出する必要はないので、同一ニュース映像間の照合は行なわないようにした。

また、照合の際に映像間の類似度を表す尺度として、正規化相互相関（Normalized Cross Correlation; NCC）を用いることにした。NCC は次式 2 によって与えられ、−1 から 1 までの値をとる。

$$NCC(a, b) = \frac{\frac{1}{n} \sum_i^n (a_i - \dot{a})(b_i - \dot{b})}{\sqrt{\frac{1}{n} \sum_i^n (a_i - \dot{a})^2} \sqrt{\frac{1}{n} \sum_i^n (b_i - \dot{b})^2}} \quad (2)$$

ただし、a, b は画像を、a_i, b_i は画像を構成する n 個の画素の輝度を、$\dot{a} = \sum_i^n a_i, \dot{b} = \sum_i^n b_i$ を表している。全く同一の画像同士では $NCC = 1$ となるので、本研究では、相関値が極めて 1 に近い（0.9 以上）場合を同一（near-duplicate）であるとした。

4.5 詳細照合

前節の低次元繰り返し照合により、同一映像区間の候補が検出されたが、この中には誤検出も多く含まれている。そこで、この絞り込んだ候補の対に対し、特徴ベクトルを特徴次元圧縮を行なう前の高次元の特徴ベクトルに戻して照合を行なう。ここでも類似度の尺度として NCC を使用し、閾値 0.9 以上となった候補を同一映像区間の対として検出する。

4.6 後処理

詳細照合により検出された映像区間のより正確な区間範囲を出力するための区間範囲調整処理を行なう。

5. 実験と考察

以上の手順に従って実験を行ない、その結果について考察した。

5.1 実験対象

本実験で使用する入力ニュース映像として、2004 年 11 月に実際に放送された日本と米国のニュース映像（日本は NHK の News7、米国は CNN の NewsNight AARON BROWN）を対象とした。比較対象として、日本語ニュース映像（News7 の映像 30 分）1 本に対し、そのニュースの前後 24 時間以内の英語ニュース映像（CNN の映像 60 分）2 本を比較対象として 1 つのグループとして扱う（図 5）。これは、日米間の時差やニュース伝達の時間を考慮したためである。さらに、実験で使用した全てのニュース映像には文字放送字幕テキストが存在している。以下に記す実験では、このようなグループを 2 つ用意し、その結果について検証・考察する。

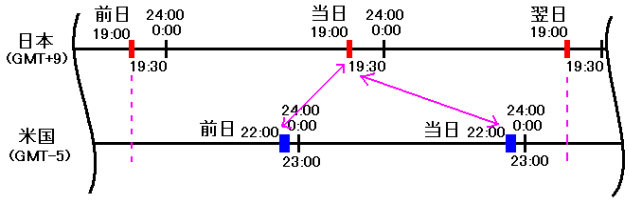

図 5　実験に使用した映像グループにおける日米ニュース映像の放送時間の関係

5.2 実験手順

実験として、評価実験と比較実験を行ない、最終的にそれぞれの検出結果を比較することで同一ニュースイベント検出における画像情報の効果を確認する。

5.2.1　評価実験

図2に示した手順で、実験対象となるニュース映像のストーリを分割し、比較領域を切り出し・正規化した後に同一映像区間を検出する。本実験では照合する際の映像長を3秒とし、参照する映像のシフトする幅を2秒とする。そして、検出された同一映像区間をストーリごとにまとめ、同一映像区間を含むストーリ対は同一ニュースイベントを扱っているものとして検出する。しかし、日本語ニュース映像にはその日の放送中に取り扱うニュースイベントを象徴した複数の短い映像をつなぎ合わせた1分30秒程度のオープニング映像が存在する。この映像部には対応する文字放送字幕テキストが存在せず、1つのニュースストーリとして扱うのは適切ではないと考える。そこで、本実験ではこのオープニング映像中に存在する同一映像区間の組を評価の対象外とする。

5.2.2　比較実験

比較実験として、テキスト情報のみから同一ニュースイベントを検出する。これは、映像情報により得られる結果がテキスト情報のみにより得られる結果に対し、どのような効果があるかを検証するためである。比較実験の処理は、以下に示す手順に従う。

（1）　日本語形態素解析システムJUMAN[6]により、各ストーリからキーワードとなる文字列を抽出

- 抽出するキーワードはJUMANにより「名詞」、または「未定義語」と判断された文字列とする
- 名詞：「人称名詞」や「形式名詞」は対象外, 前後に「接頭辞」や「接尾辞」がある際は結合, 名詞が連続して並んでいる場合は結合する
- 未定義語：「地名」や「人名」のみを対象とする

（2）　キーワードの抽出後、ストーリ間で出現頻度ベクトルの内積を算出

- ストーリ中の出現回数が2回以上のキーワードのみを対象とする
- 内積がある閾値以上の場合に、同一ニュースイベントとして検出する

5.3　実験結果

5.3.1　評価実験

提案手法の結果、各グループの日本語ニュース映像と英語ニュース映像2本の間に、グループ1では25組（前日）と0組（当日）、グループ2では4組（前日）と1組（当日）の同一映像区間が検出された。

ここで、グループ2で当日の英語ニュース映像との間に存在している1組の映像区間は同一の映像ではなく、比較領域がそ

沆僨僨市 2 まｷ井脈語い逖醗/P 諤膝ゾ

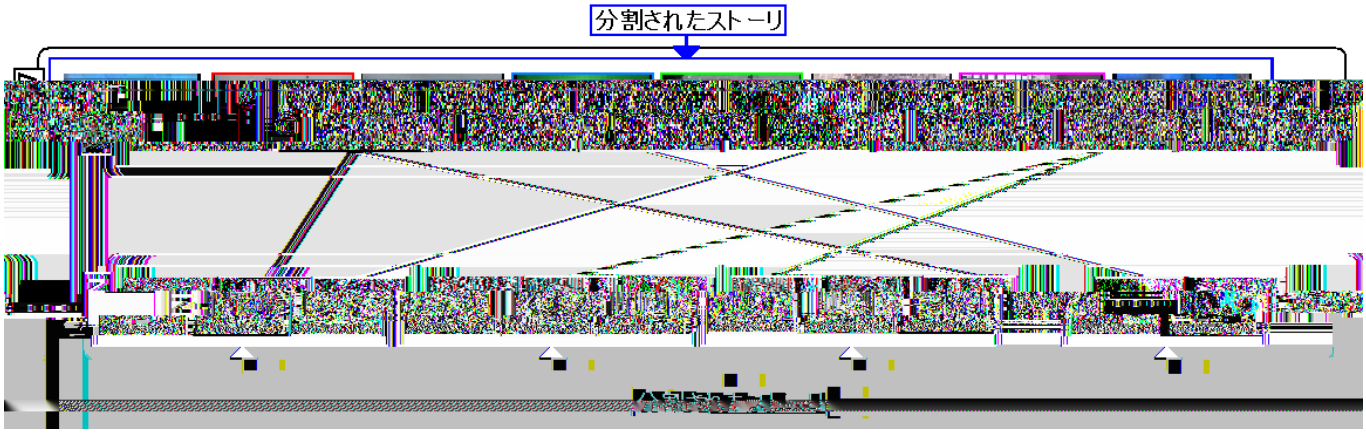

図 6　画像情報による同一ニュースイベントの検出結果の例

画像情報によっても検出することができた 3 組に関しては、目視によるストーリ内容比較の結果、テキスト情報のみにより検出された組よりもニュースイベントの同一性は高かった。

以上のことから、画像情報による検出結果はテキスト情報による検出結果と比較して少数であり、単体では検出性能が高いとはいえないが、テキスト情報では検出することができないイベントを検出することができることをふまえると、テキスト情報による検出と併せることで、より正確に同一イベントを検出できそうである。すなわち、言語横断型の同一ニュースイベント検出における画像情報の利用は、テキスト情報を補完する効果があり、検出性能の向上に有効である可能性があると考えられる。

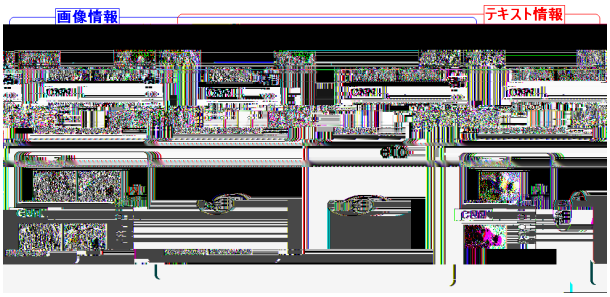

図 7　画像情報・テキスト情報の検出結果の比較

6. む す び

本研究では、テキスト情報のみによる検出では充分な結果を得ることができない、言語横断型のニュース映像中の同一ニュースイベント検出に対し、その検出精度を向上させるための 1 つの手法として、画像情報を用いる手法が効果的であることを、ニュース映像中から同一映像区間を検出することで示した。日本と米国で放送されたニュース映像に対し適用した実験により、テキスト情報のみの検出と比較して、画像情報を用いた検出の有効性を示した。画像情報による同一映像区間の検出の際に若干誤検出が出たり、テキストによる検出と比較して検出数は少なかったが、テキストでは検出できなかったニュースイベントを検出することができた。また、テキストによる検出と併せることで、より正確に同一ニュースイベントを検出できそうなことも確認した。

今後の課題として、英語以外の言語のニュース映像なども含め、より大量のニュース映像に対する適用とその評価が挙げられる。また、画像情報とテキスト情報の連携に関する検討のために、従来のテキストによるニュースイベント検索手法の併用、画像とテキストの重要度を考慮した重み付けなどを考えている。さらに、文字放送字幕テキストの存在しないニュース映像に対する適用についても考慮していく。

謝　辞

本研究の一部は 21 世紀 COE プログラムおよび科学研究費補助金による。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。

文　献

[1] 渡辺靖彦, 岡田至弘, 金地健吾, 阪元慶隆: “TV ニュースと新聞記事を対象にしたマルチメディアデータベースシステム”, 信学技報, PRMU97–257, pp.47–54, Mar. 1998.

[2] 井手一郎, 孟洋, 片山紀生, 佐藤真一: “大規模ニュース映像コーパスの意味構解析”, 信学技報, PRMU2003–97, pp.13–18, Sept. 2003.

[3] 柏野邦夫, 黒住隆行, 村瀬洋: “ヒストグラム特徴を用いた音や映像の高速 AND/OR 探索”, 信学論, vol.J83-D-II, no.12, pp.2735–2744, Dec. 2000.

[4] 野田和広, 高橋友和, 井手一郎, 目加田慶人, 村瀬洋: “適応的特徴選択を用いた長時間放送映像からの高速な繰り返し区間検出”, 信学技報, PRMU2005–289, Mar. 2006.

[5] Graham Finlayson, Steven Hordley, Gerald Schaefer, Gui Yun Tian: “Illuminant and device invariant colour using histogram equalisation”, Pattern Recognition, Volume 38, Issue 2, pp.179–190, Feb. 2005.

[6] 黒橋禎夫, 河原大輔: “日本語形態素解析システム JUMAN version5.1”, 東京大学大学院情報理工学系研究科, http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html より入手, Sept. 2005.

[7] 田村秀行編著: “コンピュータ画像処理”, オーム社, 2003.