画像情報を用い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、情報通信技術の進歩により大量の映像資源を蓄積・利用することが可能になり、それら映像資源を効果的・効率的に利用する技術が求められている。その中で重要性や利用価値の高さから、我々は多種多様なニュース映像中から同一イベントを検索する技術に注目している。現在のニュース映像検索は一般にテキスト情報のみを利用した手法が用いられているが、言語横断検索においては機械翻訳性能の問題や異なる視点・文化的背景などの制約により、十分な検索は困難である。そこで本研究では、ニュース映像中の画像情報に注目し、同一映像区間を検出することでテキスト情報による検索を補完し、より精度の高い同一イベントの言語横断検索を目指す。
キーワード　言語横断情報検索, 映像検索, 放送映像, 同一ニュースイベント, 同一映像区間検出

Cross-lingual retrieval of identical news events using image information

Akira OGAWA[†], Kazuhiro NODA[†], 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 Information and Communication Technology. Media processing technologies to use these video resources effectively and efficiently are needed. We focus on identical event detection in various news programs. Currently, text information is generally used for the retrieval of news video. However, cross-lingual retrieval is difficult due to the problem of machine translation performance and different view-points and cultures. In this paper, we introduce a method which makes use of image information for the task. We aim at complementing cross-lingual retrieval of idential news events by text information referring to the existence of near-duplicate video segments.
Key words　Cross-lingual information retrieval, video retrieval, broadcast video, identical news events, near-duplicate video segments

1. はじめに

1.1 背景と目的

近年、情報通信技術の進歩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効果的・効率的に利用するために要約・解析・検索などの技術が求められている。なかでも、重要性や利用価値が高いとされるニュース映像における様々な検索技術に対する需要は高いと考えられる。そこで、我々は多種多様なニュース映像中から同一イベントを検索する技術に注目する。

現在、ニュース映像を対象とした検索処理は一般に、新聞記事検索 [9] 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の検索においては比較的容易に実現することができ、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するため、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。そのような複数の国の放送映像からなる映像資源に対して従来

(注1)：現在、(株) デンソー

通りテキスト情報のみによる手法で検索する場合、機械翻訳の性能における問題や高度な自然言語理解の必要性、国や放送局による異なる視点・文化的背景があり、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する。また、全てのニュース映像に必ずしも文字放送字幕テキスト（Closed Caption とも呼ばれる）のような音声を書き下したテキストが付随しているわけではないので、音声認識によるテキスト情報を用いなければならない場合もある。しかし、現在の音声認識の性能では正しく音声を書き下すことが困難なため、十分なテキスト情報を得ることができない。

そこで本研究では、テキスト情報を補完するための情報としてニュース映像中の画像情報を利用することにより、複数言語間での同一ニュースイベント検索の性能を向上させることを目標とする。多くのニュース映像では、長期間にわたって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に同一映像が含まれる場合には、それらの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。

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る。検証にあたっては、本実験ではストーリ単位でニュースイベントの検出を行ない、画像情報により検出された同一イベント対の正誤は実際の映像と文字放送字幕テキストの内容から目視で行なう。また、画像情報による同一イベント検出の性能を示すため、テキスト情報による複数言語のニュース映像間からの同一イベント検出結果との比較を行なう。このようにして、画像情報による同一イベント検出の性能を確認するとともに、テキスト情報による同一イベント検索を補完し、検索性能の向上に利用できるかを検証する。

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の集合体であり、存在する場合には文字放送字幕テキストもこれに含む。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

とし、内容的構成は

- ストーリ：ニュース映像中で 1 つのイベントを扱う単位
- イベント：ある日ある場所で起こった出来事

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からなるテキスト情報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場合、機械翻訳・音声認識の性能に関する問題や異なる視点・文化的背景などにより性能が低下してしまう。

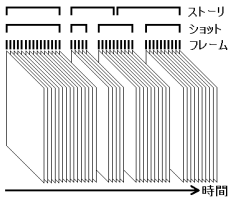

図 1　映像の構成

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でテキスト情報による検索を補完することを提案する。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る場合、そのニュースイベント間には何らかの関係があると考えられるので、本手法ではこの同一映像区間を使用した同一イベント検出の手法を提案する。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、（長時間の）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. 画像情報を用いた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、画像情報を用いた同一ニュースイベント検出を行なう。処理は大きく準備処理と主処理の 2 つに分かれる。準備処理では、既存の手法によりストーリ分割を行なう。一方、主処理では、まず、ニュース映像から照合領域を切り出し、明るさに関する正規化を行なう。ここで照合領域とは、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、照合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 準 備 処 理

ニュース映像は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 [2] を用いる。一方、海

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

外のニュース映像に対するストーリ分割は、米国 NIST 主催の TrecVid(注1)参加者共有データによる分割結果(注2)を用いる。

3.3　主　処　理

3.3.1　照合領域の切り出し

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多々あり、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の類似度が低下する恐れがある。単純に類似度の閾値を低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各フレームの中心部分（図 3）のみを照合領域として用いる。これは、大量のニュース映像から目視で判断した結果である。

図 3　照合領域の切り出し

3.3.2　照合領域の正規化

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、照合時における特徴量として 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高い。そこで、ニュース映像全体に対し、ヒストグラム平坦化 [5] により明るさの正規化を行なうことで、この問題に対処する。以下の実験では、RGB 値を YUV 値に変換し、輝度値である Y に対して平坦化を行い、再び RGB 値に戻すことで正規化を行っている。この正規化により得られた RGB 値を新たな色特徴とし、以降の処理を行なう。

(注1)：http://www-nlpir.nist.gov/projects/trecvid/

(注2)：http://www.ee.columbia.edu/ln/dvmm/newHome.htm

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により、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 値をひとつなぎにした $\bar{a}_i$ $(i = 1, 2, \ldots, I)$ 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の影響を軽減すると同時に、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 352×240 pixel から 22×15 pixel 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 $\bar{a}_i (i = 1, 2, 3, \ldots, I)$ の正規化値 a_i を算出する。

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

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、

Rüüst‘”›ÃÕ«ÄçwÍiyV›æsO¢$ 4£¤
ˆŠtSMoz¤Ñè”Üw›ÃÕ«Äçq`oj‹rgp^
R`h{ Õ«ÄçwOjz{ ‹wGVM‹w›Í• $M_1;M_2$
x¬œp›Ãí w, qb”{‡czH 1ˆŠwyVxí
M²wyVq`ozéþw¤Ñè”Üw›ÃÕ«Äçt0`o
wÍiyV›æsO{ÍtzH 2ˆŠwyVxÌ M²wyV
q`ozH 1ˆŠwÍiyVUª^•hÑè”Ü›Ì%»t
,z°ùb”à éþwÌ ¢°ùíÌ £ü‡qŠhó:Ñ
è”Üw›ÃÕ«Äç› 1mw›ÃÕ«Äçq`oÍiyV›

誤検出が考えられる。そこで、正解検出数・誤検出数ともに多かった映像グループ 3 に対して以下に示す予備実験を試みた。まず、検出漏れを防ぐために閾値を緩くして正解映像区間をすべて網羅できるように照合を行う。このままだと誤検出数が増加してしまうので、詳細照合により得られた結果に対し、新たな画像特徴としてコリログラム [8] を用いた照合により絞込みを行う。実験の結果、正解検出数・誤検出数はそれぞれ 19, 4 となり、適合率は 83%($\frac{19}{23}$), 再現率は 100%($\frac{19}{19}$) となった。以上より、照合における閾値の調整と新たな画像特徴を用いた絞込みにより精度の高い検出を行えると考えられる。

次に、同一イベント検出において画像情報を利用することの効果について考察する。本実験より、画像情報を用いた同一イベント検出における適合率は 60%($\frac{9}{15}$), 再現率は 50%($\frac{9}{18}$) となり、テキスト情報を用いた検出では適合率 40%($\frac{14}{35}$), 再現率 78%($\frac{14}{18}$) となった。この数字だけ見ると、テキスト情報による検出は適合率こそ低いが再現率は 80% 近くあり、それに比べて画像情報による検出は再現率が 50% と低く、テキスト情報を補完しているようには見えない。しかし、画像情報により検出された同一イベントの中にはテキスト情報では検出することができなかった 4 つの同一イベントが含まれており、両情報による検出結果を併せて用いることで、再現率は 100% となった (表 5)。このことより、画像情報を用いた同一イベント検出は、単体ではあまり高い性能とはいえないが、テキスト情報では得られないものを検出できるため、テキスト情報による検出を補完する情報として有効であり、検索性能の向上に利用可能のであることが確認された。

	適合率	再現率
OR	40%	100%
AND	100%	28%

表 5　画像情報・テキスト情報による検出結果を併せた適合率・再現率

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

6. むすび

本研究では、テキスト情報のみによる検出では十分な結果を得ることが困難である言語横断型の同一ニュースイベント映像検出において、画像情報として同一映像区間の存在を用いることで、テキスト情報による検出を補完し、検出性能の向上に利用できることを示した。実験より、同一映像区間検出の際に若干の未検出・誤検出が生じたり、テキスト情報による検出と比較して再現率が低かったりしたため、画像情報のみを用いた同一イベント検出という点では性能的にいくつか問題があった。しかし、テキスト情報では検出できなかった同一イベントを画像情報を用いることで検出することができ、また、テキスト情報による検出結果と併せることで再現率が 100% となり、テキスト情報を補完する情報としての画像情報の有効性を確認することができた。

今後の課題として、まず、同一映像区間検出の際に発生する未検出・誤検出を削減するため、照合における閾値や新たな画像特徴を用いた絞込みなどを検討し、再現率 100% を目指しつつ適合率も高くすることが挙げられる。また、同一イベント検出における再現率 100% を実現するために画像情報とテキスト情報を併せて用いる検出手法を提案し、重み付け等を検討することで適合率も高くすることを考えている。そして最終的に、世界各国の多種多様なニュース映像を含む、より大量のニュース映像アーカイブに適用して評価をし、高性能な同一イベントの言語横断検索を実現させる。

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, 目加田慶人, 井手一郎, 村瀬洋: “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”, 第 3 回情報科学技術フォーラム講演論文集, vol.3, pp.85–87, Sept. 2004.

[5] Graham Finlayson, Steven Hordley, Gerald Schaefer, Gui Yun Tian: “Illuminant and device invariant colour using histogram equalisation”, Pattern Recognition, vol. 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.

[8] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih: “Image indexing using color correlograms”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition ’97, pp.762–768, June 1997.

[9] 新谷研, 角田達彦, 大石巧, 長尾真: “単語の共起頻度と出現位置による新聞の関連記事の検索手法”, 情報処理学会論文誌, Vol.38, no.4, pp.855–862, Apr. 1997