

ニュース文と画像内容の一致性を考慮した 複数ニュース映像の要約

張 よう[†] 谷繁龍之介^{†*} 道満 恵介^{††,†} 川西 康友[†]
井手 一郎[†] 出口 大輔^{†††,†} 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

^{††} 中京大学 工学部, 〒470-0393 愛知県豊田市貝津町床立 101

^{†††} 名古屋大学 情報戦略室, 〒464-8601 愛知県名古屋市千種区不老町

* 現在, (株) デンソー.

E-mail: †zhangy@murase.m.is.nagoya-u.ac.jp, †{ide,murase}@is.nagoya-u.ac.jp ††ddeguchi@nagoya-u.jp

あらまし 本報告では、ニュース文と画像内容が一致する要約映像の作成手法を提案する。大量のニュース映像を資料として活用する際、あるトピックに関する映像を全て閲覧するには時間を要するという問題を解決するために、ニュース映像の自動要約技術を開発する。従来手法では、ニュース映像の文字放送字幕 (CC: Closed-Caption) を用いた映像要約がなされている。しかし、CCのみを利用するニュース映像の要約手法では、映像の重要な構成要素である画像内容が考慮されないという問題があった。そこで提案手法では、Visual Concept を用いて画像内容を分析し、重要文として抽出したニュース文の内容と一致するシーンを選択して、それに重要文の音声を重ねることで、複数ニュース映像を要約する。提案手法の有効性を確認するために、2種類の被験者実験を行った。その結果、様々なニューストピックに対して、提案手法が有効であることを確認した。

キーワード ニュース映像, 要約, Visual Concept, 文字放送字幕

Summarization of news videos based on the consistency of news text and image contents

Ye ZHANG[†], Ryunosuke TANISHIGE^{†*}, Keisuke DOMAN^{††,†}, Yasutomo KAWANISHI[†],
Ichiro IDE[†], Daisuke DEGUCHI^{†††,†}, and Hiroshi MURASE[†]

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

^{††} School of Engineering, Chukyo University
101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan

^{†††} Information Strategy Office, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

* Currently at DENSO CORPORATION

E-mail: †zhangy@murase.m.is.nagoya-u.ac.jp, †{ide,murase}@is.nagoya-u.ac.jp ††ddeguchi@nagoya-u.jp

Abstract This report proposes a method for summarizing news videos by selecting scenes whose image contents are consistent with the news texts. News videos are valuable resources because daily events are recorded in them. Therefore, there is a demand for efficiently watching a large amount of them. Some text-based methods for summarizing news videos have been proposed. However, they summarize news videos based on only their text contents, which leads to the problem that they do not take into account their image contents. Thus, the proposed method selects scenes whose image contents are consistent with the text contents considering “Visual Concepts” of each scene. Two kinds of subjective experiments were performed in order to evaluate the performance of the proposed method. As a result, we confirmed the effectiveness of the proposed method for various news topics.

Key words News video, summary, Visual Concept, Closed-Caption

1. はじめに

近年、テレビ放送の多チャンネル化および映像共有ウェブサイトの普及などにより、実世界に流通する映像データの量が爆発的に増加している。このような背景をうけ、大量の映像データを効率的に閲覧するための技術が求められている。放送映像の中でもニュース映像は実世界の出来事を記録したものであるため、資料的な価値が高い。しかし、資料として活用する際に、あるトピックに関する映像を全て閲覧するには長い時間を要する。そのため、ニュース映像の検索 [1] や複数のニュース映像の要約 [2] に関する研究がなされている。調査対象の話題に関するニュース映像群の一部を視聴して理解するだけでなく、話題の移り変わりを追跡し、内容を総合的に理解することが重要であるため、1つのトピックに関する複数のニュース映像を要約する技術が求められている。

ニュース映像は、主に動画と音声から構成されている。近年、日本や米国などでは、多くのニュース映像に文字放送字幕 (CC: Closed-Caption) と呼ばれる音声を書き下したテキストデータが付与されて放送されているため、テキスト処理によりニュース映像の音声内容を分析することが可能になった。それに伴い、ニュース映像の文字放送字幕に着目した映像要約の研究がいくつかなされている [3] [4]。しかし、図 1 で示すように、テキストのみを利用するニュース映像の要約手法では、映像の重要な構成要素である画像内容を考慮していないため、ニュース文の内容と関連が少ないシーンが選ばれる場合があるなど、必ずしも視覚的に有用な情報が含まれないことがあった。そこで、ニュース文の内容だけでなく、画像内容も考慮したニュース映像の要約手法が必要であると考えた。

一方、映像の画像特徴に着目した研究もいくつかなされている [5] [6]。これらの手法は、画像の色ヒストグラムや画像の DCT 係数等の画像特徴を利用して映像を要約する。しかし、画像特徴を利用する点で本研究と似ているが、映像要約の際に映像内容を考慮しておらず、本研究とは異なる。

本研究では、画像内容を理解するために Visual Concept [7] を利用する。Visual Concept とは、画像中に含まれる人物や物体、自然など、画像の視覚的な内容を表すものである。その例を図 2 に示す。Visual Concept を用いてニュース映像に含まれる視覚的な内容を把握することで、ニュース文の内容を最も適切に表現するシーンを選択できると考えられる。

2. ニュース文と画像内容の一致性を考慮した複数ニュース映像の要約

本手法では、ニュース文の内容と一致するシーンを選択することで複数のニュース映像を 1つの映像に要約する。提案手法の処理の手順を図 3 に示す。入力とは既存手法で構築したニュース映像のトピックスレッドである。次に、入力映像の文字放送字幕に関するテキスト処理と、動画に関する画像処理を行う。最後に、抽出された重要文とシーンの Visual Concept により適切なシーンを選択して、それに重要文の音声を重畳することで、要約映像を作成する。



(a) 既存手法が選択したシーン (b) ニュース文の内容と一致するシーン (ニュース文の内容と不一致)

図 1 ニュース文:「一連の日程を終えて帰国しました。政府専用機で羽田空港に到着した安倍総理大臣」に対する画像内容

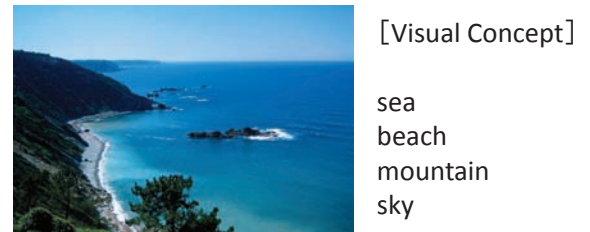


図 2 Visual Concept の例

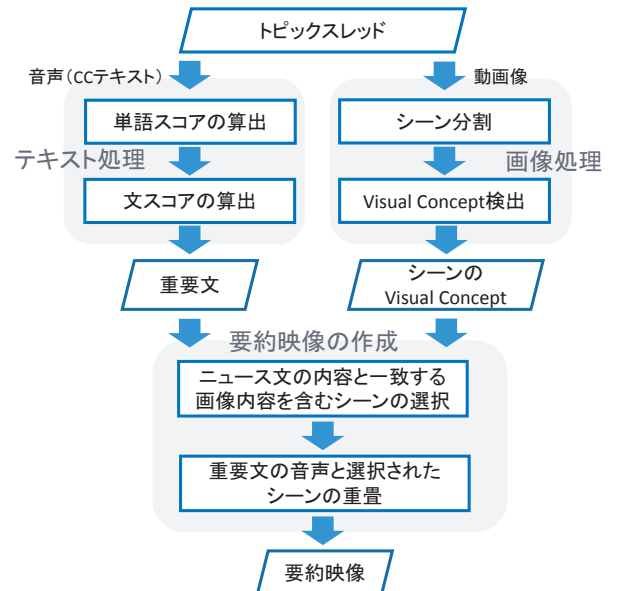


図 3 提案手法の処理の流れ

2.1 事前処理

本手法は、文字放送字幕が付与されて放送されている日本のニュース映像を対象とする。事前処理により、あるトピックに関連するニュースストーリーの時系列意味構造を構築して、その中から最も主要な遷移を抽出する。その例を図 4 に示す。処理の手順としては、まずニュース番組の映像から各ストーリーを分割し、その音声内容である文字放送字幕を抽出しておく。次に、井手らの手法 [8] により、特定のトピックに強く関連するトピックスレッド構造を構築する。そして、加藤らの手法 [9] により、ストーリーの長さ、放送時間、放送順序などの特徴を考慮して、各ストーリーの重要度を算出する。重要度がしきい値以上のストーリーを最も多く含むトピックスレッドを主要なものとして抽出する。

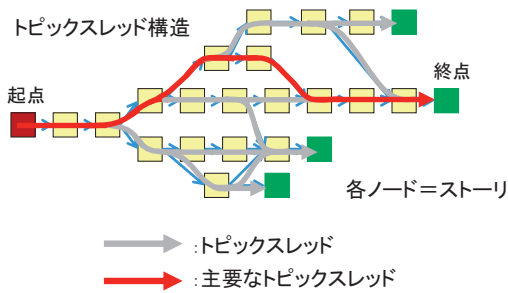


図4 加藤ら[9]の手法により推定した主要なストーリー遷移の例

2.2 テキスト処理

本節では入力映像の文字放送字幕に関するテキスト処理について説明する。

2.2.1 単語スコアの算出

単語スコアの算出方法を説明する。各ストーリーにおいて、tf-idfを用いて単語にスコア付けをする。tf-idfとは、文書中の単語に関する重み付け方法の一種であり、主に情報検索や文章要約などの分野で利用される。具体的には、以下の手順で処理を行う。

(1) トピックスレッドに含まれる全てのストーリーに付随する文字放送字幕を形態素解析する。

(2) 名詞と未知語をキーワードとして抽出した後、各キーワードに対し、あるストーリー中の出現回数を表す特徴量 Term Frequency (キーワードの出現頻度)である $tf_{i,j}$ とそのキーワードを含むストーリー数を表す特徴量 Inverse Document Frequency (逆文書頻度)である idf_i を以下の式により求める、

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$idf_i = \log \frac{|D|}{|\{d|i \in d, d \in D\}|} \quad (1)$$

$n_{i,j}$ はストーリー j 内でキーワード i が出現する回数、 $\sum_k n_{k,j}$ はストーリー j における全てのキーワードの出現回数の和、 $|D|$ は総ストーリー数、 $|\{d|i \in d\}|$ はキーワード i を含むストーリー数である。

(3) 各ストーリーにおいてキーワードの tf-idf ベクトルを作成する。 $\omega_{i,j}$ を j 番目のストーリーにおけるキーワード i の tf-idf 値とすると、

$$\omega_{i,j} = tf_{i,j} \cdot idf_i \quad (2)$$

となる。

このように各ストーリー内でのキーワードの出現回数と、そのキーワードを含むストーリー数を両方考慮することによって、出現頻度が高いキーワード(一般的なキーワード)の重要度を下げ、出現回数が少なく、珍しいキーワードの重要度を上げる作用がある。

2.2.2 文スコアの算出

文スコアの計算方法を説明する。まず、ストーリーに付属する文字放送字幕を句点により文に分割する。次に、単語スコアを基にストーリー内の各文にスコア付けを行う。続いて、各文に含まれるキーワードのスコアの平均に基づき、Visual Concept の

語彙と一致する単語の数により重み付けをし、文スコアを計算する。具体的には、ストーリー中の文 l のスコア S_l を次の式のように定義する。

$$S_l = \frac{1}{|W_l|} \sum_{i \in W_l} \omega_{i,j} (N+1) \quad (3)$$

ここで、 W_l を文 l に含まれるキーワードの集合とする。 N は文中のキーワードのうち、Visual Concept の語彙と一致するものの数を表す。この処理によって最終的に算出される文スコアは、出現頻度及び視覚的に情報量が多いキーワードを多く含む文ほど高くなる。最後に、ストーリー内での文スコアが最も高い文を重要文として抽出する。

ここで、文中の単語と Visual Concept との対応付け方法を説明する。ニュースでは、同義語や類義語を状況に応じて使うことがある。例えば、Visual Concept の「リーダー」のカテゴリに対し、ニュース文中では「首脳」、「指導者」、「首相」、「総理大臣」、「大統領」等、様々に表現される。このような類義語間の1対多の対応関係を判断するため、日本語 WordNet [10] を用いて Visual Concept の語彙を拡張する。日本語 WordNet は、日本語の概念辞書(意味辞書)である。日本語 WordNet では、日本語の単語が同義語のグループに分類され、定義や他の同義語のグループとの関係が記述されている。本研究では、文中のキーワードが Visual Concept の語彙またはその同義語と一致すれば、その Visual Concept を含むとみなす。

2.3 画像処理

入力映像の動画像に関する画像処理について説明する。

2.3.1 シーン分割

シーンとは、画像的に連続するフレーム群である。そのため、一般に同じシーン内において連続する2フレームの画像特徴は極めて類似する。本研究では、フレーム間のHSVヒストグラムの差分を用いて、前後の画像特徴量が急激に変化する箇所を検出し、動画像をシーンに分割する。具体的には、隣接した2枚のフレームのHSVヒストグラム間の距離を計算し、その差がしきい値を超えていれば、この2枚のフレームをシーン境界として分割する。ここで、HSV色空間は色を色相(Hue)、彩度(Saturation)、明度(Brightness)の3要素で表現し、人間の感覚に近い表現方法である。

2.3.2 Visual Concept 検出

同じシーン内の画像内容はあまり変わらないと考え、シーンの先頭フレームを代表フレームとして、Visual Concept を検出する。本研究では、一般的的事象に対応するBVLC GoogLeNet 検出器および独自に構築した「人」に関する検出器の2種類の Visual Concept 検出器を用いる。BVLC GoogLeNet Visual Concept 検出器は、画像中の「物体」、「場所」、「活動」など様々な Visual Concept を検出できるが、ニュース映像に出現頻度が高い「人」について細かく分類できない。そこで、ニュース映像に適用するために、「人」とその属性に関する検出器を独自に構築し、BVLC GoogLeNet 検出器と併用することにした。この2つの検出器の組み合わせにより1,000種類以上の Visual

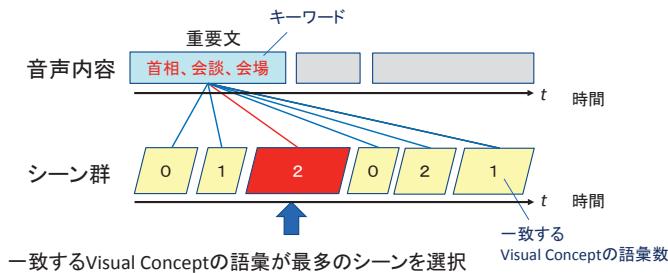


図5 ニュース文の内容と一致する画像内容を含むシーンの選択方法

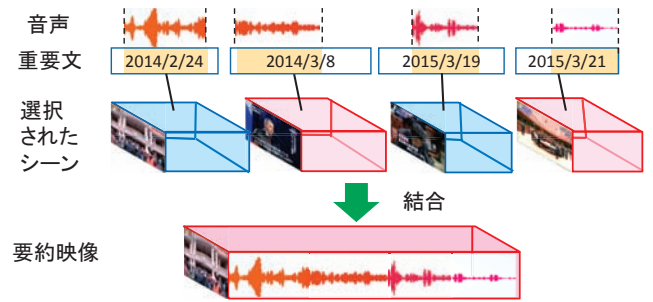


図6 要約映像の作成方法

Concept を検出する。

2.4 要約映像の編集

抽出された重要文と検出された Visual Concept からシーンを選択して、要約映像を作成する方法について説明する、

2.4.1 ニュース文の内容と一致するシーンの選択

図5に示す通り、以下の基準で、ニュース文に適したシーンを選択する。

(1) 2.3.1で説明した処理により、1つのニュースストーリーの動画から分割されたシーン群から、文中のキーワードと一致する Visual Concept の語彙数が最大のシーンを選択する。

(2) 基準(1)のシーンが複数ある場合、時間的に近い方を選択する。

(3) 選択したシーンの長さが文に対応する音声より長い場合は途中で打ち切り、短い場合は既に選択したシーンに加えて、基準(1)に戻って新たなシーンを選択し直す。

基準(1)において、文中のキーワードと画像中の Visual Concept との対応付け方法は2.2.2で説明した対応付け方法と同じである。日本語 WordNet を用いて、キーワードの同義語や同義語のグループを取得することで、Visual Concept の語彙を拡張する。

2.4.2 要約映像の作成

通常、ニュース番組では、アナウンサーが放送局でニュースの概要を説明する映像(アンカシーン)と、ニュースの詳しい内容を説明する映像の2種類で構成される。以下にトピックスレッド中の各ストーリーから、上記の処理で得られた重要文と、それと内容的に一致するシーンを用いて要約映像を作成する。要約映像の作成方法は、加藤らの手法[9]を参考にし、図6に示すように、以下の処理手順で行う。

(1) 各ストーリーから抽出した重要文を時系列に並び替える。

(2) 重要文と同期した音声を切り出す。

(3) 各重要文の画像内容と一致したシーンを切り出し、文の出現順序に従ってその音声を重畳することで、要約映像を作成する。

3. 実験

提案手法の有効性を評価するために、2種類の被験者実験を行った。以降、順に説明する。

3.1 ニュース文の内容と一致するシーンの選択方法に関する評価

本節では、要約映像を作成する際にニュース文と一致する

表1 各データに対する提案手法の選択率

トピック	データ	選択率	トピック	データ	選択率
被災地訪問	1	56%	自然災害	8	69%
	2	47%		9	44%
経済	3	34%	政治	10	78%
	4	72%		6	47%
	14	84%	11	59%	
電車の事故	5	84%	スポーツ	13	84%
	7	93%	教育	12	91%
	15	94%	平均		69.2%

シーンの選択方法に関する評価実験について説明する。

3.1.1 データセット

本実験では、NHK News7の文字放送字幕から抽出したニュース文15文を用いた。実験に用いたニュース文の抽出条件を以下に示す。

- 対象：NHK News7の文字放送字幕
- 期間：2013年1月14日～2013年5月12日
- 条件：Visual Conceptの語彙と一致するキーワードを1つ以上含む

3.1.2 被験者実験

20代の男女32名による被験者実験を行い、提案手法の評価を行った。実験方法の詳細を以下に示す。まず、ニュース文を被験者に提示した。次に、提案手法と比較手法により作成した映像を1本ずつ被験者に視聴してもらい、以下の設問に対する回答を得た。

● 設問：ニュースの内容について、視覚的に良く表現している映像はどちらか

● 回答方法：提案手法により作成した映像/比較手法により作成した映像のいずれかを選択

比較手法として、ニュース文と同期したシーン(元映像)をそのまま切り出す手法を用いた。

評価尺度として提案手法の選択率を用いた。これは以下の式により求められる。

$$\text{提案手法の選択率} = \frac{\text{提案手法を選択した人数}}{\text{被験者人数}} \quad (4)$$

3.2 実験結果と考察

各データに対して、32名の被験者に評価をもらった。被験者実験の結果は、全てのデータにおいて、提案手法で作成した映像の平均選択率が69.2%であった。各データに対する提案

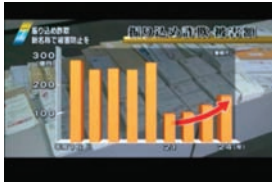


(a) 比較手法



(b) 提案手法

図 7 データ 5 (電車の事故) において各手法で選択した映像



(a) 比較手法



(b) 提案手法

図 8 データ 3 (経済) において各手法で選択した映像

手法の選択率を表 1 に示す。事故やスポーツ、教育といったトピックでは提案手法の選択率が高く、自然災害、政治、経済などのトピックに対しては、選択率が低いものもあった。

提案手法の選択率が高かった例を図 7 に示す。データ 5 は、「きょう午後、兵庫県高砂市の山陽電鉄の踏切で、特急電車がトラックと衝突して、1 両目と 2 両目が脱線し、駅のホームに衝突しました」という内容であった。提案手法では、文中の「電車」、「トラック」などのキーワードを用いて、映像中から「電車」、「トラック」の Visual Concept が検出されたシーンを選択した。一方、比較手法では、アナウンサだけが映っているシーンであったため、視覚的な一致性が低かった。スポーツ、教育などのトピックに対しても、同じ理由で、提案手法の選択率が高かったと考えられる。

提案手法の選択率が低かった例を図 8 に示す。データ 3 は、「警察は、被害防止に向けたキャンペーン強化に乗り出します」という内容であった。提案手法では、文中の「警察」というキーワードに対して、同じストーリー中から「警察」の Visual Concept が検出されたシーンを選択した。一方、比較手法では、「振り込め詐欺の被害額が年々増加している」ことを示す表が映っていた。比較手法を選択した被験者は、「警察」だけが映っているシーンよりも、図表などが映っているシーンの方がより解りやすく、ニュース文の内容をより良く表現していると判断したと考えられる。これにより、今後は、Visual Concept に文字や図に関する検出器を加えて、このような映像を選択できるようにする必要があると考えられる。

3.3 要約映像の作成方法に関する評価

本節では、要約映像の作成方法に関する評価実験について説明する。

3.3.1 データセット

本実験では、入力映像として NHK News7 を使用し、加藤らの手法 [2] で抽出したトピックスレッド 3 本を実験に用いた。以下、表 2 に示す 3 つのストーリーを起点とする主要なトピックスレッドをデータ 1~データ 3 と呼ぶ。

被験者実験 1 とは異なる 20 代の男女 15 名による被験者実験

表 2 実験に用いたトピックスレッド構造とその内容

データ	放送年月日, ストーリ番号	話題
1	2013/11/21, 08	東京電力の原発再開
2	2014/03/25, 13	2014 年クリミア危機
3	2014/09/14, 05	スコットランド独立運動

表 3 各手法によって作成された映像の長さ [秒]

手法	データ 1	データ 2	データ 3
提案手法, 比較手法 3	78	88	43
比較手法 1, 比較手法 2	54	101	73

表 4 各比較手法に対する提案手法の被験者による平均選択率

	提案手法の選択率			平均
	vs. 比較手法 1	vs. 比較手法 2	vs. 比較手法 3	
データ 1	60%	80%	80%	60%
データ 2	60%	80%	100%	80%
データ 3	60%	80%	20%	67%

を行った。実験を開始する前に、トピックに関する事前知識を共有するために、Wikipedia における各ニューストピックの概要を一読してもらった。実験方法の詳細を以下に示す。

被験者に映像を視聴してもらい、以下の設問を回答させた。

- 設問：音声の内容を視覚的により良く表現している映像はどちらか

- 回答方法：提案手法により作成した映像/比較手法により作成した映像のいずれかを選択

また、比較評価については、以下の 3 つの手法で要約映像を作成した。

- 比較手法 1：tf-idf のみを用いた単語スコアに基づいて選択した文と同期したシーン（元映像）をそのまま切り出して結合したもの。

- 比較手法 2：tf-idf のみを用いた単語スコアに基づいて選択した文の内容に沿って、Visual Concept 検出を用いて内容的に対応するシーンを切り出して結合し、音声差し替えたもの。

- 比較手法 3：tf-idf と文中の Visual Concept と対応するキーワード数に基づいて選択した文と同期したシーン（元映像）をそのまま切り出して結合したもの。

- 提案手法：tf-idf と文中の Visual Concept と対応するキーワード数に基づいて選択した文の内容に沿って、Visual Concept 検出を用いて内容的に対応するシーンを切り出して結合し、音声差し替えたもの。

実験では、各ストーリーからスコアが最も高い文を重要文として選択して用いた。評価尺度は被験者実験 1 と同様に提案手法の選択率を用いた。

3.4 実験結果と考察

各手法によって作成された映像の長さを表 3 に、被験者実験の結果を表 4 にそれぞれ示す。全てのデータに対して、比較手法 1 及び比較手法 2 よりも、提案手法の方が平均選択率が高かった。比較手法 3 と比較すると、提案手法の平均選択率はデータによって大きな差があった。すなわち、提案手法では



(a) 比較手法 3



(b) 提案手法

図 9 ニュース文：「住民投票の実施を求めたのに対して、イギリス政府は、地元住民の意思を尊重するとしまして、1回に限って、住民投票の実施に合意したということなんです」に対して、選択されたシーン。



(a) 比較手法 3



(b) 提案手法

図 10 ニュース文：「しかし、住民投票の詳細な条件を決めるときに、サモンド党首は、未来のことを決めるのだから、若い人たちも参加すべきだと主張しました」に対して、選択されたシーン。

データ 1、データ 2 に対する有効性が高かったが、データ 3 に対する有効性が低かった。

比較手法 1 について考察する。比較手法 1 は、重要文の選択方法とシーンの選択方法が提案手法と異なり、tf-idf により重要文を選択して元映像をそのまま切り出す。それに対して、提案手法は、重要文を選択する際に、Visual Concept と対応するキーワードの数を考慮し、重要文と一致したシーンを切り取って要約に用いる。そのため、シーンの切れ目と音声の切れ目の時間的なずれにより不自然さが生じたと考えられる。これが原因で、提案手法の選択率が低かったと考えられる。

比較手法 2 に対して、提案手法の選択率が高かった。これに関して、手法の違いは重要文の選択方法だけで、重要文を選択する際に、Visual Concept と対応するキーワードの数を考慮するか否かである。このことから、重要文を選択する際に Visual Concept を考慮することが重要であると考えられる。

比較手法 3 について考察する。比較手法 3 と提案手法の違いは、シーンの選択方法である。この結果から、シーンを選択するときに、Visual Concept を考慮することが重要であると考えられる。しかし、データ 3 では提案手法の選択率が低かった。その原因としては、データ 3 にはモノログシーン（演説やインタビューなどのシーン）を多数含むためと考えられる。このようなシーンに対して、提案手法を適用すると、話者と音声の不一致により不自然さが生じやすい。データ 3 について、提案手法と比較手法 3 により選択したシーンの一部を図 9 と図 10 に示す。このような映像データに対する対策は、モノログシーンでは、例外的に話者本人の音声を使用することなどが挙げられる。なお、モノログシーンの検出には Kumagai らの手法 [11] などがあるため、それらを利用することが考えられる。

4. むすび

本報告では、あるトピックに関する複数のニュース映像の自動要約を目的とし、音声と画像の内容の一致性を考慮した要約手法を提案した。提案手法では、トピックスレッド構造から選択した主要なトピックスレッドを入力とし、そのトピックスレッド内のニュース映像の文字放送字幕から抽出した重要文と画像内容が一致するシーンを選択することで要約映像を作成した。

今後は、要約映像を作成する際に、モノログシーンは話者本人の音声を使用するなどの例外ルールを追加することや、要約映像の長さをユーザが指定できるようにすることなどが考えられる。

謝 辞

本研究の一部は科学研究費補助金及び国立情報学研究所との共同研究による。

文 献

- [1] 井手 一郎, “ニュース映像の検索”, 映像情報メディア学会誌, Vol.64, No.3, pp.306–311, Mar. 2010.
- [2] K. Kato, I. Ide, D. Deguchi, and H. Murase, “Generation of a video summary on a news topic based on SNS responses to news stories”, Proc. 4th Int. Workshop on Crowdsourcing for Multimedia, pp.21–26, Oct. 2015.
- [3] D. R. Radev, H. Jing, M. Stys, and D. Tam, “Centroid-based summarization of multiple documents”, Information Processing and Management, Vol.40, Issue 6, pp.919–938, Nov. 2004.
- [4] 山田 一郎, 柴田 正啓, “ニュース記事の定型性を利用した話題要約の検討”, 情報処理学会研究報告, 2001-NL-148-7, Mar. 2002.
- [5] F. Wang and B. Merialdo, “Multi-document video summarization”, Proc. 2009 IEEE Int. Conf. on Multimedia and Expo, pp.1326–1329, June 2009.
- [6] J. Almeida, N. J. Leite, and R. S. Torres, “Vison: Video summarization for online applications.” Pattern Recognition Letters, Vol.33, no.4, pp.397–409, Mar. 2012.
- [7] M. J. Huiskes, B. Thomee, and M. S. Lew, “New trends and ideas in Visual Concept detection”, Proc. 11th Int. Conf. on Multimedia Information Retrieval, pp.527–536, Mar. 2010.
- [8] 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, Sept. 2008.
- [9] K. Kato, I. Ide, D. Deguchi, and H. Murase, “Estimation of the representative story transition in a chronological semantic structure of news topics”, Proc. 4th ACM Int. Conf. on Multimedia Retrieval, pp.487–491, Apr. 2014.
- [10] F. Bond, T. Baldwin, R. Fothergill, and K. Uchimoto, “Japanese SemCor: A sense-tagged corpus of Japanese”, Proc. 6th Int. Conf. of the Global WordNet Association, pp.9–16, Jan. 2012.
- [11] S. Kumagai, K. Doman, T. Takahashi, D. Deguchi, I. Ide, H. Murase, “Speech shot extraction from broadcast news videos”, Int. J. of Semantic Computing, Vol.6, No.2, pp.179–204, June 2012.