

音声特徴と画像特徴を利用したニュース映像からの同一場面検出

熊谷はるか[†] 道満 恵介^{††} 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

^{††} 中京大学情報理工学部 〒470-0393 愛知県豊田市海津町床立 101

^{†††} 名古屋大学情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町

E-mail: [†]kumagaih@murase.m.is.nagoya-u.ac.jp, ^{††}kdomain@sist.chukyo-u.ac.jp, [†]ide@is.nagoya-u.ac.jp,
^{†††}ddeguchi@nagoya-u.jp, [†]murase@is.nagoya-u.ac.jp

あらまし 本報告では、同一のイベントを異なる視点で撮影した同一場面検出手法を提案する。近年、大量に蓄積されたニュース映像を内容に基づいて扱う技術が必要になっている。我々はその中でも、同一場面の検出に注目しており、これまで、映像を人物領域と背景領域に分割し、それぞれの領域同士を独立に照合し、その結果を統合する領域別照合手法を提案してきた。しかし、この手法は、ある程度撮影視点が異なる映像には頑健ではあるが、あまりにも異なる映像の検出は困難であった。そこで本報告では、従来の画像特徴に加え、音声特徴を追加した同一場面検出手法を提案する。音声特徴により検出できる映像は限られるが、撮影視点に関わらず同一場面を検出することができるため、精度の向上が期待できる。実際に放送されたニュース映像で同一場面検出実験を行い、提案手法の有効性を確認した。
キーワード 同一場面検出, 撮影視点の違い, 音声特徴, 画像特徴, ニュース映像

Scene Duplicate Detection from News Videos Using Visual Features and Audio Features

Haruka KUMAGAI[†], Keisuke DOMAN^{††},

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 Information Science & Technology, Chukyo University
Tokodachi 101, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan

^{†††} Information and Communications Headquarters, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

E-mail: [†]kumagaih@murase.m.is.nagoya-u.ac.jp, ^{††}kdomain@sist.chukyo-u.ac.jp, [†]ide@is.nagoya-u.ac.jp,
^{†††}ddeguchi@nagoya-u.jp, [†]murase@is.nagoya-u.ac.jp

Abstract In this report, we propose a method to detect scene duplicates from news videos. Scene duplicates are videos taking the same event from different perspectives. Recently, technology to deal with the large amount of news videos based on their contents is needed. Among them, we are focusing on the scene duplicate detection problem. To detect scene duplicates, we have previously proposed a region-based matching method. This previous method is robust to minor difference of perspectives. However, it was difficult to detect shots in the case of a major difference of perspectives. In order to deal with this problem, the proposed method detects scene duplicates by audio features in addition to image features. Although a detectable shot is restricted, audio features can detect scene duplicates regardless to perspectives. Therefore, we expect to accuracy to improve. Scene duplicate detection experiments were conducted using actual news videos. From the results, we confirmed the effectiveness of the proposed method.
Key words Scene duplicate detection, difference of perspective, audio feature, image feature, news video

1. はじめに

近年、記憶装置や通信技術の発達により、放送映像を大量に蓄積できるようになった。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ており、資料的観点から価値が高い。ニュース映像を資料として活用するためには、大量に蓄えたニュース映像を内容に基づいて扱う技術が必要となる。本研究では、そのような技術の中でも、ニュース映像から Near-duplicate を検出する技術に注目する。Near-duplicate とは、ある同一の時点・場所で起こった実世界の出来事（イベント）を撮影した映像、または、異なる時点に同じ対象を撮影した映像のことである。Katayama らは、Near-duplicate が多い映像ほど重要であるという考えに基づき、News Shot Cloud という大規模なニュース映像の一覧表示方法を提案した [1]。Ogawa らは、言語が異なる国におけるニュース中のイベントを同定するために、Near-duplicate を利用した [2]。このように、Near-duplicate はニュース映像を内容に基づいて扱うために、重要な手がかりとなる。

表 1 に示すように、Near-duplicate は 3 種類に分けられる [3]。

(1) Strict Near-duplicate: 同一のイベントを同じ視点で撮影した映像のこと。図 1(a) に示すように、編集、字幕制作などによる違いがあり得る。

(2) Object Duplicate: 図 1(b) に示すように、異なるイベントにおいて、同じ撮影対象を同じ場所で撮影した映像のこと。

(3) Scene Duplicate: 図 1(c) に示すように、同じイベントを異なる視点から撮影した映像のこと。しばしば複数の放送局による撮影映像に見られ、編集の違いから、放送される区間が異なることが多い。

本研究では、(1) と (3) を合わせて「同一場面」と呼び、これをニュース映像から検出する手法を提案する。ニュース映像では、ある 1 つのイベントに関するニュースでも、放送局ごとに使われる映像や報道の仕方が異なる場合がある。このように、ニュース映像は放送局ごとに多様であるが、同一場面を検出できれば、ある 1 つのイベントを様々な視点のニュース映像から理解することが可能になる。

同一場面の検出を目的とした研究として、Takimoto らの研究 [4] や Wu らの研究 [3] がある。Takimoto らは、映像中のフラッシュを利用した検出手法を提案している。しかし、この手法は、フラッシュが映っていない映像は検出できない。Wu らは、映像中の人物の特徴点の不均一性パターンに基づいて同一場面を検出する手法を提案している。この特徴点の不均一性パターンは、人物の動きの緩急を表す特徴である。人物の動きの緩急は、撮影視点の違いに頑健であり、また、同じ撮影対象でも撮影時刻によって変化する。したがって、撮影時刻が異なる (2) と (3) を区別して検出することができる。しかし、この手法は、図 2 に示すような、背景に文字や絵があるような場合（記者会見映像など）では検出が困難となる。このような映像では、背景に特徴点が集まり、人物に特徴点が少なくなるため、人物の動きの緩急が正しく抽出できないためである。そのため、実際のニュース映像へそのまま適用できない。

表 1 Near-duplicate の分類

	イベント	撮影対象	撮影視点
(1)	同一	同一	同一
(2)	異なる	同一	同一でも異なってもよい
(3)	同一	同一	異なる



(a) Strict Near-duplicate: 同一のイベントを同じ視点で撮影



(b) Object Duplicate: 異なるイベントで同じ対象を撮影



(c) Scene Duplicate: 同一のイベントを異なる視点で撮影

図 1 Near-duplicate の分類

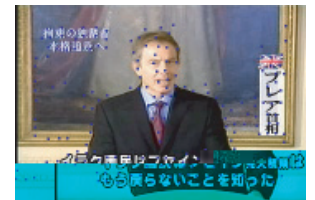


図 2 人物の特徴点が少ない映像 [3]

これに対して我々は、映像の各フレームを人物領域と背景領域に分割し、それぞれの領域同士を独立に照合し、その結果を統合する領域別照合手法を提案した [5]。また、撮影視点がある程度異なる映像に対する有効性を確認した。しかし、あまりにも撮影視点が異なる場合には検出が困難であった。

そこで本報告では、従来の画像特徴に加え、映像に付属する文字放送字幕（Closed Caption; CC）による前処理と、音声特徴を追加した同一場面検出手法を提案する。CC は、ニュース中の音声を書き下したものであり、映像中に音声に基づくテロップが存在する場合は省略されることがある。CC と音声特徴は、それぞれ検出できる映像が限られるが、撮影視点に関わらず同一場面を検出することができる。そのため、CC による前処理を行い、音声特徴と画像特徴を利用することで、精度の向上が期待できる。

以降、2. では提案手法について述べ、3. で評価実験について述べ、その結果について考察する。最後に 4. で本報告をまとめる。

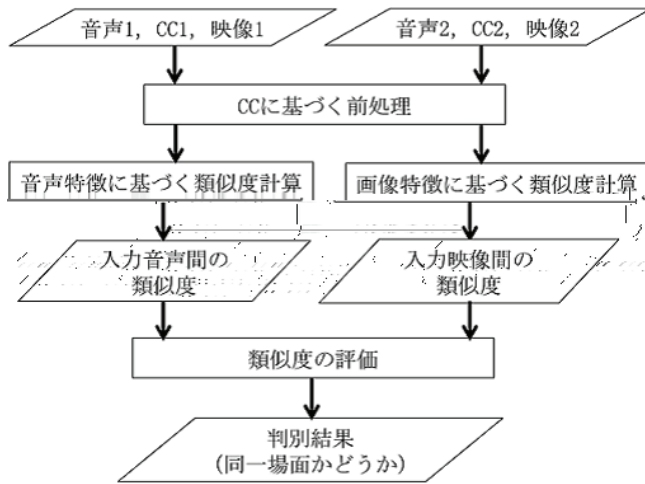


図 3 提案手法の処理の流れ

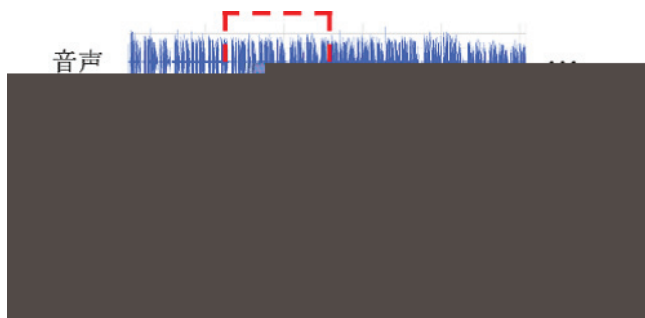


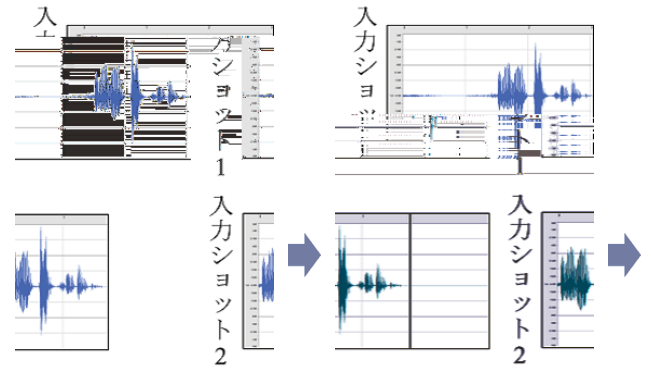
図 4 入力データ：各メディアの破線部分

2. 提案手法

提案手法の処理の流れを図 3 に示す。初めに、入力データについて説明する。図 4 に示されるように、入力データは、音声、映像、そして CC からなる。それぞれ破線で囲まれた部分が入力データである。横軸は時刻を表し、映像の各サムネイルがショットを表す。また、あるイベントに関する意味的なまとまりであるストーリーは、ショットの集まりからなる。映像は、ショットを入力とする。音声は、ショットに対応する区間を用いる。また、CC は、ショットが含まれるストーリーに対応する部分を用いる。まず、入力 CC による前処理を行い、CC が類似する入力データ組を以降の処理対象とする。次に、音声特徴による照合では、音声波形の正規化相互相関を計算して類似度とする。映像中の話者と音声と一致する同一場面では、入力音声組はほぼ一致するため、単純な特徴で照合することができる。画像特徴による照合では、入力ショット組を人物領域と背景領域に分割し、それぞれの領域を撮影視点の違いに頑健な特徴を用いて照合し、その結果を統合して類似度を計算する。最後に、計算された類似度を評価し、入力データ組が同一場面かどうかを判別する。以降、各処理について順に詳述する。

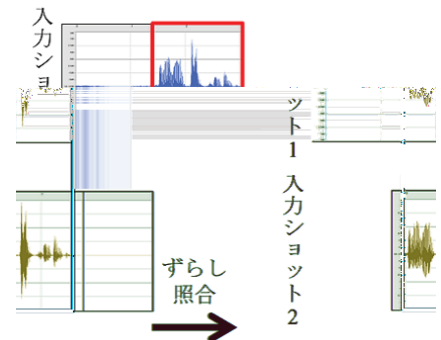
2.1 CC による前処理

同一場面は同一時刻の映像であるため、映像に対応する CC も類似することが期待される。そこで、入力 CC 間の類似度を計算し、類似度がしきい値未満の入力データ組は同一場面の



(a) ダウンサンプリング・正規化
処理後の入力ショット組の音声

(b) 無音部分の挿入



(c) 正規化相互相関の計算

図 5 音声特徴による類似度計算の流れ

候補から外す。類似度の計算では、まず、CC を形態素解析し、品詞判別を行う。そして、得られた形態素の品詞情報を用いてキーワードとなる語を抽出する。このとき、キーワードは名詞と未定語とする。そして、得られたキーワードとその出現頻度をキーワードベクトルとし、その余弦距離を計算し、類似度とする。

入力データの説明で述べたように、入力 CC は、入力ショットが含まれるストーリーに対応する CC とする。入力ショット組が同一場面であるならば、そのショットが含まれるストーリー組も類似し、それに対応する CC も類似することが期待される。ここで、ショットに直接対応する CC を用いないのは、類似度を安定して計算できないほど長さが短いショットが存在すること、テロップが存在する時は CC が存在しない場合があること等の問題があるためである。これらの原因により、正しく CC 組の類似度を測れない恐れがある。一方、ストーリーはショットの集合であるため、ある程度の長さになる。また、アナウンサーが発言する際にはテロップではなく、CC が存在する。また、ストーリー中にアナウンサーによる発言が存在しないことは稀である。そのため、ストーリーの CC を用いると、先述の問題を回避できる。

また、ストーリーの検出手法としては、井手らの CC を用いた手法 [6] などが挙げられる。本論文では、このような手法を用いてストーリーを精度良く検出できるものと仮定する。

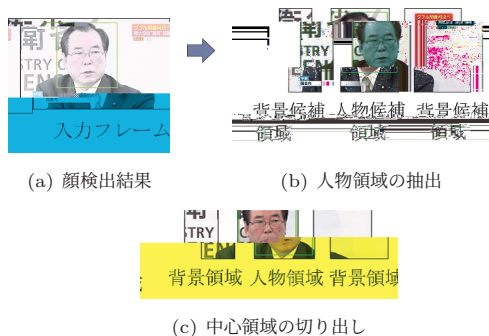


図 6 領域分割

2.2 音声特徴による照合

ニュース映像において、映像中の話者と音声と一致する同一場面では、入力ショット組の音声とほぼ一致する。しかし、そうはならない同一場面、例えばアナウンサーによるナレーションが重畳するような同一場面では、音声は一致しない。そこで、入力ショット組の音声の類似度を計算し、類似度がしきい値以上のショット組を同一場面であると判別する。

具体的には、事前処理として、まず、入力ショットの音声のサンプリングレートを 4kHz に揃える。これは、後の処理の計算時間を短縮するためである。音声特徴では、ほぼ一致する音声を検出することが目的であるので、サンプリングレートの低下はほぼ影響がないと考えられる。次に音声の正規化を行い、ボリュームを揃える。最後に、入力ショット組の音声がある時間よりも短い場合、その音声の類似度は 0 であるとする。これは、極端に短いショットの音声には、音声で判別できる同一場面が存在しないからである。また、そのようなショットの音声を照合に加えると、誤検出の原因となるため、類似度を 0 とすることであらかじめ照合から除外する。

次に、類似度の計算方法について説明する。処理の流れを図 5 に示す。図 5(a) が前処理を行った入力ショット組の音声である。まず、図 5(b) に示すように、入力ショット組の時間長が短い方のショットの音声の末尾に無音部分を加え、入力ショット組の音声の長さを揃える。このようにすることで、短い音声に無音部分が多く含まれることになり、短いショットに長いショットが誤って対応づきやすい問題を回避する。最後に、図 5(c) に示すように、片方をサンプリングレートでずらしながら音声波形の正規化相互相関を計算し、その中で最も大きい相関を入力ショット間の音声の類似度とする。

2.3 画像特徴による照合

画像特徴による照合は、領域分割、領域別の個別照合、個別照合結果の統合からなる。以降、各処理について説明する。

2.3.1 領域分割

まず、図 6(a) に示すように、Viola らのオブジェクト検出手法 [7] を用いて、入力ショットの各フレームから顔を検出する。次に、図 6(b) に示すように、検出された顔の横幅に基づいて、入力ショットを横に 3 分割する。中央の領域を人物候補領域とし、左右の領域を背景候補領域とする。

ニュース映像には、各番組が独自のテロップなどを挿入する



図 7 映像の明るさが異なる同一場面例

場合が多々ある。このようなテロップは、照合の妨げになる。そこで、図 6(c) に示すように、人物候補領域と背景候補領域に分割したショットから、テロップが挿入される可能性が低い中心領域を切り出し、照合に用いる。切り出された中心領域の中央部分を人物領域、左右を背景領域とする。また、背景領域は入力ショットの最初のフレームのみを用いる。これは、背景領域に関しては、ショット中で大きく変化することが少ないためである。

2.3.2 領域別の個別照合

領域別の個別照合は、人物領域の照合と、背景領域の照合からなる。以降、各処理について説明する。

a) 人物領域の照合

人物領域の照合は、映像中の人物の特徴点の不均一性パターンに基づく同一場面の検出手法 [3] により行う。まず、入力ショットから特徴点を検出・追跡し、複数の特徴点軌跡を得る。そして、以降の処理の計算時間を短縮するために、全ての特徴点軌跡を平均する。本研究では、特徴点軌跡の検出・追跡を人物領域のみに限定しているため、外れ値の影響は少ないと考えられる。次に、平均した特徴点軌跡から、Inconsistency をフレームごとに算出する。Inconsistency は、特徴点の動きの緩急を表し、値が大きいフレームでは、特徴点の動きの向きや速度が大きく変化する。最後に、ショット間の類似度を求める。あるウィンドウ内で Inconsistency が極大値をとるフレームを求め、その前後 w フレームの Inconsistency の正規化相互相関の平均を類似度とする。ここで、同一場面は放送される部分が番組によって異なる場合があるため、オフセットごとに類似度を計算し、最も大きい類似度をショット間の類似度 R_{face} とする。

b) 背景領域の照合

本研究では、背景領域の類似度として、色ヒストグラム間の距離を用いる。ここで図 7 に示すように、ニュース映像は、録画状況や各放送局の規定によって同一場面でも映像の明るさが異なる場合がある。この明るさの違いによる検出漏れや誤検出を避けるために、色ヒストグラムには HSV 値の色相 (H) と彩度 (S) のみを用いる。まず、領域分割されたショットの最初のフレームから色相と彩度を求め、2 次元ヒストグラムを計算する。次に、ヒストグラム間の Bhattacharyya 距離を求め、1 から減じることで背景領域の類似度 R_{bg} を計算する。

2.3.3 個別照合結果の統合

人物領域の類似度 R_{face} と背景領域の類似度 R_{bg} から、入力ショット S_i, S_j 間の類似度 R_{fusion} を次式で計算する。

$$R_{\text{fusion}}(S_i, S_j) = \alpha R_{\text{face}}(S_i, S_j) + (1 - \alpha) R_{\text{bg}}(S_i, S_j) \quad (1)$$

ここで、 α は 0 から 1 の値をとる重み係数を表す。

表 2 実験に用いた映像 (2012 年 6 月 7 日～2012 年 6 月 13 日)

放送局	番組名	放送時刻
NHK 総合	ニュース	12:00 ～ 12:20
東海テレビ	FNN スピーク	11:30 ～ 11:55
中京テレビ	NNN ストレイトニュース	11:30 ～ 11:55
CBC テレビ	JNN ニュース	11:30 ～ 11:55
メ～テレ	ANN ニュース	11:45 ～ 12:00



図 8 提案手法による検出成功例

2.4 音声と画像を組み合わせた類似度の評価

映像中の話者と音声とが一致する同一場面では、音声とがほぼ一致するため、音声特徴によって撮影視点に関わらず同一場面を検出できる。そこで、音声特徴の類似度がしきい値以上の入力データ組は、画像特徴の類似度に関わらず同一場面であるとする。入力データが複数の場合、全てのデータ間で類似度を計算する。そして、データ間の類似度が最も大きいものを選択する。また、入力データに同一場面が存在しない可能性があるため、類似度がしきい値以上のものを同一場面と判別する。しきい値は、線形判別分析法によって決定する。

3. 実験・考察

実際に放送されたニュース映像に本手法を適用した実験の結果について述べる。

3.1 実験用映像

本実験では、入力ニュース映像として、2012 年 6 月 7 日から 2012 年 6 月 13 日に放送されたものをを用いた。用いた放送局と番組、放送時刻を表 2 に示す。また、入力ニュース映像のフレームレートは 29.97 [frames/sec]、解像度は 1,440 × 1,080 [pixels] であった。これらの映像から、人物を中心に撮影されたショットを人手で計 678 ショット切り出した。この 678 ショットのうち、200 ショットは同一場面が存在し、478 ショットは同一場面が存在しなかった。なお、同一場面の検出精度への影響を排除するため、ストーリー 262 個は全て人手で検出し、人物領域は人手で切り出した。

3.2 実験方法

提案手法において、CC による前処理の有無と照合に用いる特徴の組み合わせを変えた手法を比較手法として、同一場面の検出精度を比較した。ここで、画像特徴のみを利用する比較手法は従来手法 [5] に相当する。評価基準として、適合率、再現率、F 値を用いた。

3.3 実験結果

各手法による実験結果の適合率、再現率、F 値を表 3 に示す。提案手法が F 値に関して最も高くなり、提案手法の有効性を確認できた。また、図 8 に示すように、撮影視点の異なる同一場面を検出できた。

表 3 実験結果:同一場面検出の検出精度

手法	利用する特徴			再現率	適合率	F 値
	CC	音声	画像			
提案手法	✓	✓	✓	0.645	0.326	0.433
比較手法 1	✓	✓		0.255	0.879	0.395
比較手法 2	✓		✓	0.525	0.243	0.332
比較手法 3		✓	✓	0.480	0.390	0.430
比較手法 4		✓		0.245	0.583	0.345
比較手法 5 [5]			✓	0.500	0.227	0.313

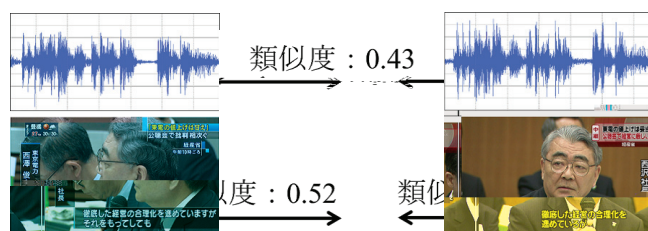


図 9 音声特徴による検出成功例

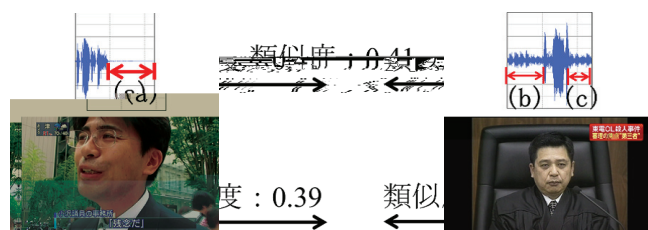


図 10 音声特徴による誤検出例

3.4 考察

音声特徴による照合について考察する。音声特徴による検出成功例の音声波形を図 9 に示す。中心あたりの振幅が小さい部分など、波形の特徴的な部分が類似していることがわかる。そのような特徴的な部分が類似すると、正規化相互相関は大きくなるため、検出することができた。また、データセット中に話者と音声と一致する同一場面は 53 ショット存在し、そのうち、比較手法 4 では 49 ショットが検出された。このことから、音声波形の正規化相互相関による照合の効果を確認できる。また、誤検出例を図 10 に示す。どちらも振幅が大きくなる区間の前後に、振幅の小さい区間 (a), (b), (c) が見られる。音声を確認すると、(a) は微小な雑音、(b), (c) は音楽が背景に流れており、その部分の相関が大きくなり、誤検出された。このように、音声波形の先頭、または末尾における雑音などは誤検出の原因となる。そこで、事前に除去する、または類似度の計算に用いないなどの処理が必要だと考えられる。

音声特徴のみを用いた比較手法 4 と、画像特徴のみを用いた比較手法 5 を比べる。比較手法 4 では、図 11 に示すような、画像特徴では検出が難しい大きく撮影視点の異なるショット組が検出できる。しかし、図 12 に示すように、映像中の話者と音声と一致しないショットと映像中の話者と音声と一致するショットの組み合わせの場合、音声特徴のみでは検出できず、画像特徴のみで検出できる。そこで、これらの特徴を組み合わせた提案手法では、両方のショット組が検出できる。また、表 3 に示

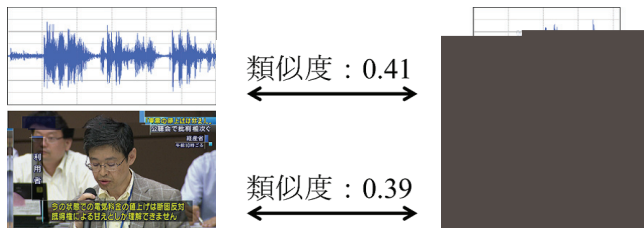


図 11 音声特徴で検出でき、画像特徴で検出できなかった例

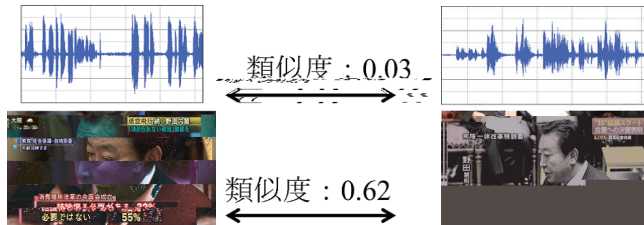


図 12 画像特徴で検出でき、音声特徴で検出できなかった例



図 13 提案手法で検出できなかった例

ように、比較手法 4 は適合率が高いが再現率が低く、比較手法 5 は適合率は低いが再現率は高くなった。これは、音声特徴は映像中の話者と音声一致する同一場面のみを撮影視点に関わらず検出でき、画像特徴は大きく撮影視点の異なる同一場面は検出が難しいが、検出対象を限定しないためであると考えられる。これらの手法と比較し、両方の特徴を用いた比較手法 3 は、適合率と再現率のバランスがとれ、F 値が高くなっており、全体の精度が向上している。よって、性質の異なる 2 つの特徴を用いることは効果的であるといえる。また、提案手法で検出できなかった例を図 13 に示す。一方のショットが話者と音声一致しないショットであり、撮影視点が大きく異なったため、音声特徴と画像特徴で検出できなかった。このようなショットを検出するために、性質の異なる特徴、例えば服の色やネクタイの色などの特徴をさらに組み合わせれば、さらなる精度の向上を実現できると考えられる。

CC による前処理を用いた手法と、用いなかった手法を比較すると、精度が向上している。これは、CC による前処理が有効に働いているためであるといえる。CC による前処理は、類似しないストーリーの絞込みであるため、前処理を行うと適合率が上昇すると予想される。比較手法 1 と 4、比較手法 2 と 5 は適合率が上昇しているが、提案手法と比較手法 3 では再現率が上昇し、適合率が減少している。これは、判別分析法によって決定したしきい値のためであると考えられる。前処理でストーリーが類似する入力データ組のみに絞り込んだために、しきい値が下がり、再現率が上昇したのと考えられる。今後は、最適

なしきい値の決定方法を検討する必要がある。

4. む す び

本報告では、ニュース映像における音声特徴と画像特徴による同一場面検出手法を提案した。また、提案手法の有効性を確認するために、実際に放送されたニュース映像からの同一場面検出を行った。提案手法では、性質が異なる 2 つの特徴により、同一場面を検出した。音声特徴による検出では、音声波形の正規化相互相関を用いて類似度を計算した。画像特徴による検出では、映像の各フレームを人物領域と背景領域に分割し、それぞれの領域同士を独立に照合し、その結果を統合して類似度を計算した。

提案手法の有効性を確認するために、実際のニュース映像を用いた同一場面検出実験を行った。CC による前処理の有無と照合に用いる特徴の組み合わせを変えた手法を比較手法とした。その結果、提案手法の F 値が最も高く、提案手法の有効性が確認できた。今後は、音声処理の方法や最適なしきい値の決定方法、さらなる特徴を組み合わせることを検討していく。また、同一場面検出技術により検出された同一場面には、意味的な関連があるため、同一場面を検出することによって、ニュース映像間を意味的に関連づけることができる。このような関連づけを利用した応用についても、今後検討していく。

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] N. Katayama, H. Mo and S. Satoh, "News Shot Cloud: Ranking TV news shots by cross TV-channel filtering for efficient browsing of large-scale news video archives," Proc. 17th Int. MultiMedia Modeling Conf., LNCS, vol.6523, pp.284-295, Jan. 2011
- [2] A. Ogawa, T. Takahashi, I. Ide and H. Murase, "Cross-lingual retrieval of identical news events by near-duplicate video segment detection," Proc. 14th Int. Multimedia Modeling Conf., LNCS, vol.4903, pp.287-296, Jan. 2008
- [3] X. Wu, M. Takimoto, S. Satoh and J. Adachi, "Scene duplicate detection based on the pattern of discontinuities in feature point trajectories," Proc. 16th ACM Int. Conf. on Multimedia, pp.51-60, Oct. 2008.
- [4] M. Takimoto, S. Satoh and M. Sakauchi, "Identification and detection of the same scene based on flash light patterns," Proc. IEEE Int. Conf. on Multimedia and Expo, pp.9-12, Jul, 2006.
- [5] 熊谷はるか, 道満恵介, 出口大輔, 井手一郎, 村瀬 洋, "ニュース映像中の同一シーン検出のための領域別照合手法の検討," 電子情報通信学会技術研究報告, MVE2011-108, May 2012.
- [6] 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造解析," 電子情報通信学会技術研究報告, PRMU2003-97, Sept. 2003.
- [7] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.511-518, Dec. 2001.
- [8] E. Shechtman and M. Irani, "Space-time behavior based correlation," Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.405-412, June. 2005.