

NewsWho'sWho: ニュース映像アーカイブからの人物情報ポータル構築

井手 一郎^{†‡} 關岡 直城[†] 小笠原 崇[†] 木下 智義[‡] 孟 洋[‡]
片山 紀生[‡] 佐藤 真一[‡] 高橋 友和^{†*} 村瀬 洋[†]

[†] 名古屋大学大学院情報科学研究科 〒464-8603 名古屋市千種区不老町 1

[‡] 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

[‡] (有)ネットコンパス 〒104-0033 東京都中央区新川 2-14-4

* 日本学術振興会

E-mail: [†] {ide, nsekioka, toga, ttakahashi, murase}@murase.m.is.nagoya-u.ac.jp,

[‡] {ide, mo, katayama, satoh}@nii.ac.jp, [‡] kino@netcompass.co.jp

あらまし 記憶装置の大容量化の恩恵をうけ、大量に放送映像を蓄積し、検索や内容理解をする試みが盛んになっている。我々は特定のニュース番組の映像を過去5年にわたり蓄積し、そこから高次の意味情報を抽出することを目指している。本講演では、その一環として取り組んでいる、ニュース番組に登場する人物に関する情報抽出手法について述べ、登場人物間の人間関係やモノログ映像を総合的に提供するポータルのプロトタイプを紹介する。

キーワード 放送映像アーカイブ, 社会ネットワーク, 映像内容理解

NewsWho'sWho: Creation of a biography portal from a news video archive

Ichiro Ide^{†‡} Naoki Sekioka[†] Takashi Ogasawara[†] Tomoyoshi Kinoshita[‡] Hiroshi Mo[‡]
Norio Katayama[‡] Shin'ichi Satoh[‡] Tomokazu Takahashi^{†*} and Hiroshi Murase[†]

[†] Nagoya University, Graduate School of Information Science 1 Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan

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

[‡] NetCOMPASS, Ltd. 2-14-4 Shinkawa, Chuo-ku, Tokyo, 104-0033, Japan

* Japan Society for the Promotion of Science

E-mail: [†] {ide, nsekioka, toga, ttakahashi, murase}@murase.m.is.nagoya-u.ac.jp,

[‡] {ide, mo, katayama, satoh}@nii.ac.jp, [‡] kino@netcompass.co.jp

Abstract Retrieval and understanding contents in broadcast video archives has become a hot-topic in proportion to recent advance in data storage technologies. We have been attempting to extract high-level semantic information from a television news archive which consists of daily news program recorded over the past five years. In this talk, we will introduce information extraction methods that extract information related to people that appear in the news programs, together with a biography portal that provides the users with information on human relationship, monologue scenes and so on.

Keywords Broadcast video archive, Social network, Video contents understanding

1. はじめに

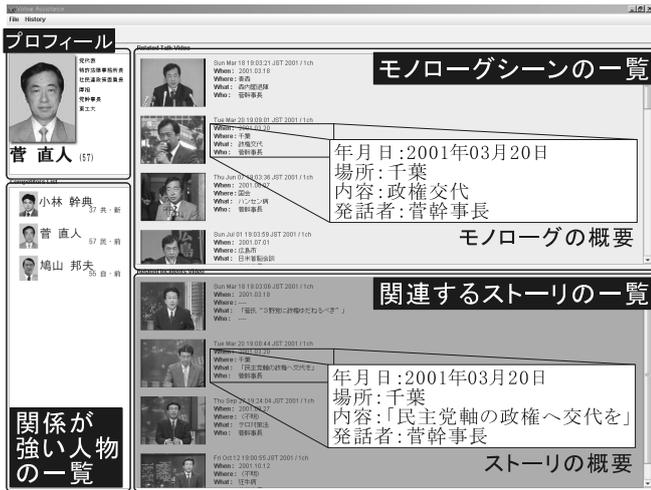
記憶装置の大容量化の恩恵をうけ、評価型映像検索ワークショップ TREC Video[1, 2]をはじめ、大量に蓄積された放送映像の検索や内容理解への取り組みが盛んである。このような背景のもと、我々は特定のニュース番組の映像を過去5年にわたり蓄積してきた[3]。ニュース映像アーカイブは、規模が大きくなるにつれ、単なる断片的な映像の集積ではなく、高次の意味情報を豊富に含むマルチメディア知識ベースと見なせるようになる。

本講演では、ニュース番組に登場する様々な人物に関する情報をニュース映像アーカイブから抽出して提供するポータルのプロトタイプを紹介する。

このポータルは、注目する人物に関する生の映像のみならず、映像内容やさらにはアーカイブ全体から抽出された高次の知識も併せて提供する。具体的には、注目する人物に関して、次のような情報を提供する。

- 他の登場人物との人間関係
- モノログ(発言・演説・インタビューなど)とそれに対する他の人物のコメントや反響
- 関係するニュースストーリー

一般に、新聞記事やニュース原稿では 5W1H(When, Where, Who, What, Why, How)が内容の伝達に重要とされている。ニュース映像アーカイブを対象とした検索に関して、従来の研究は一般的なキーワードと放送日



(a) 人物情報ビュー



(b) 映像ビュー

図 1: 人物情報ポータル NewsWho'sWho のプロトタイプインタフェース

(When)を指定するものが多かった。5W1H の他の属性に注目した検索手法としては、Christel et al.による地図上の領域 (Where) を指定したニュース映像検索インタフェース[4]がある。著者らもこれまで 5W1H のうち、特に When, Where, Who, What に注目したニュース映像索引付け手法を提案してきた[5]。しかし、これらはいずれも生の映像から抽出した 1 次的な情報に基づく索引付け・検索手法に過ぎない。本研究では 1 次的な情報にとどまらず、それらの相互関係から得られる 2 次的な情報も併せて提供することで、利用者による映像内容の理解を深めたいと考えている。

本稿は以下のように構成されている:第 2 節ではポータルのプロトタイプインタフェースを紹介する。それを支える技術として、まず第 3 節でニュースの意味構造解析について、続く第 4,5 節で得られた意味構造に基づく人間関係の解析とモノログシーンの検出手法について紹介する。最後に第 6 節でまとめる。

2. 人物情報ポータル:NewsWho'sWho

提案する人物情報ポータルのプロトタイプインタフェースを図 1 に示す。このインタフェースは以下の 2 つのビューからなる。

・ 人物情報ビュー(図 1(a))

特定の人物に関する情報を提供する。その人物のモノログシーンや関連するニュースストーリー、関係が深い人物の一覧などからなる。

・ 映像ビュー(図 1(b))

特定のストーリーやモノログに関する情報を提供する。指定したストーリーやモノログの映像を実際に視聴できるほか、そのストーリーやモノログを承けた他の人物によるコメントの一覧なども表示される。

このインタフェースの構築は現時点では全自動化されていないが、このような情報を提供するために、下記の機能の自動的な実現に取り組んできた。

- ・ ニュースの意味構造解析
- ・ 人間関係の抽出
- ・ モノログシーンの検出

以下の各節では、これらの技術について紹介する。

3. ニュースの意味構造解析

ニュース映像アーカイブに含まれる知識を抽出するためには、まずその意味構造を明らかにする必要がある。そのために、映像に付随して文字放送字幕 (Closed-Caption) として提供される音声書き下しテキスト (以下 CC テキスト) を用いて、(1) ニュース映像内のストーリー分割、(2) ニュース映像をまたいだストーリー間の関係を表すトピックスレッド構造の 2 つを解析する[5]。本節では簡単にそれらの手法を紹介する。なお、CC テキストは事前処理により、音声と同期してある。

3.1. ストーリ分割

番組単位の CC テキスト中の各文へ次の処理を適用することにより、ストーリーの境界を検出する。

1. 各文に形態素解析を施し (JUMAN[6] 3.61 を使用)、名詞列を抽出する。更に各名詞列に対して語義属性解析[7]を施し、「一般」、「人物」、「場所/組織」、「時相」のいずれかに分類する。
2. 手順 1. で抽出した名詞列の出現頻度を要素とするキーワードベクトルを属性毎に作成する。
3. 各文の境界において、前後 m (以下の実験では 1~10) 文のキーワードベクトルを結合し、それらの間の類似度を求め、さらにその最大値を採る。類似度は、ベクトル間のコサイン尺度を用いる。

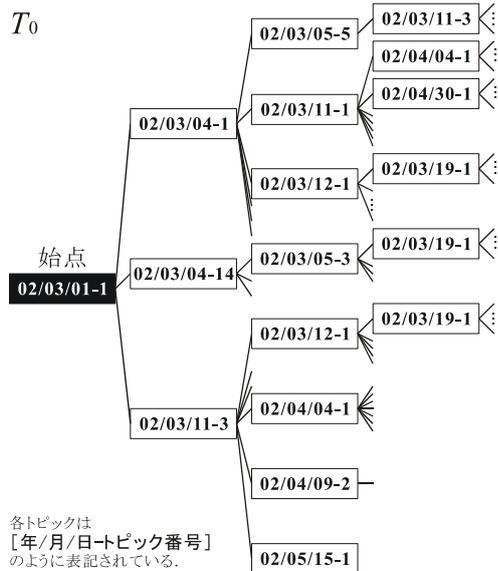


図 2: 単純な関連ストーリーの時系列木構造の例

- 属性毎の類似度の加重和を求め、閾値 θ_{seg} 以下なら、そこにストーリー境界を検出する。加重和の重みは、予備実験において人手で与えた 384 の境界から学習した値; (一般:人物:場所/組織:時相) = (0.23 : 0.21 : 0.48 : 0.08), $\theta_{seg} = 0.17$ とした。

手順 4. で用いたのとは別に用意した、人手で与えた 130 のストーリー境界による評価により、1 文以内の誤差を認めた場合に、適合率 90.5%, 再現率 95.4% の検出性能が得られた。また、2006 年 2 月 26 日の時点で、2001 年 3 月 16 日から 1,720 日分のニュース映像に対して上記の処理を適用したところ、21,164 件のストーリーが検出された。

3.2. トピックスレッド構造の解析

時間とともに変遷するニューストピックの内容を理解するためには、3.1 節で述べた手法により分割された一つ一つのストーリー間の関連を求める必要がある。このような関連を表すために、我々はストーリー間の時間的・意味的関連に基づく「トピックスレッド構造」という有向グラフ構造を提案している [8]。以下にこの構造の特徴と解析方法について述べる。

トピックスレッド構造とは、時系列的な順序関係を維持したまま、意味的に関連が強いストーリーのノードを接続した有向グラフ構造である。グラフ中に同一ストーリー(ノード)の重複を許さないため、意味構造を簡潔に表現する構造になる。図 3 にスレッド構造の例、図 2 にスレッド解析を施す前の状態、つまり重複を許した単純な関連ストーリーの時系列木構造の例を示す。

具体的には、以下の手順に従って指定されたストーリーを始点とするトピックスレッド構造を解析する。

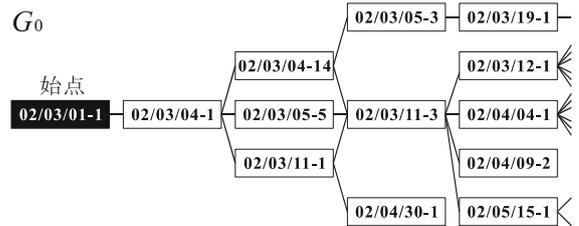


図 3: トピックスレッド構造の例

- 関連するストーリーの単純な時系列ツリー構造を作成する。下記の条件を満たしつつ、再帰的に関連ストーリーの木構造を展開する。
 - リンクが張られたノード間で、親は子よりも古いストーリーであるように時系列性を維持。
 - 兄弟は左から右へ長幼の順にソート。
 ここで、ストーリー間の類似度は、ストーリー中の名詞列出現頻度ベクトル間のコサイン尺度と定義した。ストーリー間の類似度が閾値 θ_{trk} 以上なら、それらは類似していると判断する。
- 手順 1. に従って構築されたストーリー木 T_0 (例: 図 2) 中の全ての部分木 $T_i(j)$ につき、自分より左に同一の部分木 $T_s(j)$ があるとき、以下の処理を施す。
 - $T_s(j)$ が $T_i(j)$ の先祖ならば $T_s(j)$ を除去。
 - $T_i(j)$ が $T_s(j)$ の(親を除く)先祖の子孫ならば $T_i(j)$ と $T_s(j)$ を統合。
 a) における部分木の除去は、間に一つもストーリーを経由しない短絡パスの生成を防ぐためである。

以上の処理により、図 3 に示したようなスレッド構造 G_0 が得られる。以下では、上記のストーリー及びスレッドというニュース固有の意味構造に基づき、人物に関する情報を抽出する手法について述べる。

4. 人間関係の抽出

ある集団において、構成員をノード、人間関係をエッジで表現すると、人間関係のネットワークを表現できる。Kautz らによる取り組み [9] に端を発して、このような社会ネットワークの解明や分析に関する研究が盛んになされている。しかしこれらの研究の多くは、ウェブ空間のリンク構造、電子メールヘッダ中の共起、文献や電子掲示板における引用関係など、テキスト情報のみから関係を抽出している。

これに対し、我々はテキスト情報の他に画像中の人物の顔の出現パターンを利用する手法を提案してきた [11]。まず 4.1 節で従来式の CC テキストに出現する人物を表す名詞列の共起を手がかりとした人間関係の抽出手法について、続く 4.2 節で画像中に出現する顔領域の出現パターンを手がかりとした抽出手法について述べる。なお、各手法の詳細については、文献 [10, 11] を参照されたい。

4.1. 文字放送字幕テキスト中の人物名の出現パターンによる抽出

3.1 節で述べた手法に基づき抽出されたストーリー単位で、人物名詞の共起頻度を計数する。ここで共起頻度が高い人物名詞間に強い関係があると考えられる。これをアーカイブ中の全ストーリーにわたって集計することで、一般的な人間関係が抽出される。また、3.2 節で述べた手法で得られるトピックスレッド構造や一定の時期の範囲で集計すれば、あるトピックの登場人物間の人間関係や、経時的な人間関係の変化を明らかにすることもできる。以下に具体的な処理の手順を示す。

人物名詞の抽出

3.1 節で述べたストーリー分割手順 1. で用いたのと同じ手法 [4] により CC テキスト中の人物名詞を抽出する。この手法は末尾の名詞に基づいて名詞列全体の語義属性を分類するため、固有名詞(人名)を含む名詞列(例:小泉-首相)だけでなく、普通名詞のみからなる名詞列(無名の人物, 例:会社-幹部)をも検出できる特徴がある。

なお、この手法を用いてニュース映像に重畳して表示される字幕(Open Caption)テキスト 2,549 件(ほぼ全て名詞列または体言止め)を分析したところ、再現率 72.47%、適合率 82.13%の結果が得られている。

共起頻度の計数

ストーリー中に出現する人物名詞の共起頻度を計数する。ストーリー s 中の人物名詞 p の出現頻度を $f(p, s)$ とすると、集計対象のストーリー集合 S における人物 p_i と p_j の関係の強さ $R_s(p_i, p_j)$ を次のように定義する。

$$R_s(p_i, p_j) = \sum_{s \in S} f(p_i, s) f(p_j, s)$$

なお、現時点では、同一人物の異表記や役職などの経時的変化には対応していない。

実験

2001 年 3 月から 2002 年 1 月に放送された内政関係ストーリーの映像 18 区間(約 70 分)の CC テキストに以上の処理を適用したところ、32 ノード(人物名詞)間に 312 エッジ(人間関係)が抽出された。得られた人間関係グラフの一部を図 5(a)に示す。

4.2. 画像中の顔の出現パターンによる抽出

4.1 節で述べた手法によりテキスト中の人物名詞の共起頻度に基づく人間関係を抽出した。しかし、本研究で対象とする放送映像の性質を考えると、テキスト情報により明示的に描写されない人間関係が画像情報として提供される可能性が考えられる。たとえば、著名な人物について、画像中にその人物の顔が出現することにより、

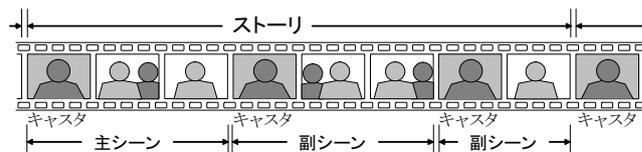


図 4: ストーリー内のサブストーリー構造

その話題に関与していることが記号的に示されることがある。逆に、テキスト中で言及されない無名の人物については、顔の出現を手がかりにしない限りは、人間関係を抽出し得ない。そこで、以下に述べる手順により、映像中に顔が出現するパターンに応じて人間関係を抽出することを目指している。

顔の検出

従来、パターン認識の分野で様々な顔検出手法が提案されている。以下の実験では、Intel 社が公開している OpenCV ライブラリ [12] 中の Haar-like 特徴による高速顔領域検出プログラムを用いた。

顔の出現パターンに応じた関係の強さの計算

以下の 2 つに分類して顔の出現パターンを考え、それぞれに関係の強さを表現する重みを付与する。

1) 同一画像内の共起

同一画像内における人物の物理的な共起頻度を数える。重みは W_f とする。

2) 同一ストーリー中の異なる画像間の共起

3.1 節で述べた手法に基づき抽出されたストーリー単位で、人物の時間的な共起を数える。ここで、ストーリー内のサブストーリー構造を考慮して重みを与える。そのために、サブストーリーの冒頭に出現することが多いキャストショットを検出し、次のキャストショットまたはストーリーの終端までの映像区間を「シーン」と呼ぶ。このなかで特に冒頭のものを「主シーン」、それ以外のものを「副シーン」と名付け、以下のように人物の顔が出現するシーン毎に異なる重みを付ける。図 4 にサブストーリー構造の例を示す。

- ・ 主シーン → 主シーン: $W_s(m, m)$
- ・ 主シーン → 副シーン: $W_s(m, s)$
- ・ 副シーン → 主シーン: $W_s(s, m)$
- ・ 副シーン → 副シーン: $W_s(s, s)$

人物毎の重みの集計

以上の処理により得られるのは、正確には顔領域間の関係である。人間関係を得るためには、異なる画像に出現する顔を照合し、同一人物の「顔寄せ」をしなければならない。しかし、放送映像から得られる顔画像の品質は必ずしも高くないため、現時点では人手で照合した結果を利用している。

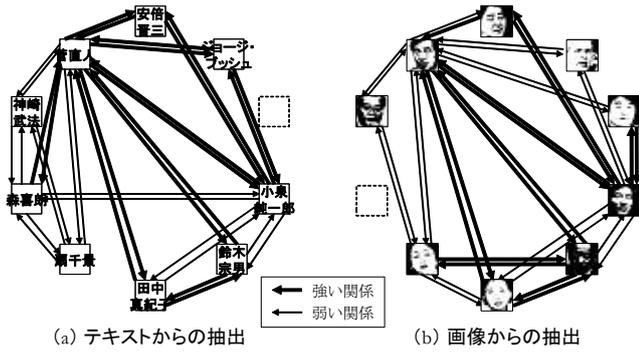


図 5: 抽出された人間関係グラフの例

実験

予備実験による検討の結果, 重みを $(W_s(m, m) : W_s(m, s) : W_s(s, m) : W_s(s, s) : W_j) = (5 : 4 : 3 : 2 : 1)$ と設定し, 登場する顔領域毎に人間関係の強さを評価した.

図 5(a)を得たのと同じ映像に対して本手法を適用し, 人手で顔寄せをしたところ, 68 ノード間に 1,414 エッジが抽出された. 図 5(a)に対応する画像情報を用いて得られた人間関係グラフを図 5(b)に示す.

4.3. テキストからの抽出と画像からの抽出の比較

4.1 節及び 4.2 節で得られた人間関係を比較したところ, 次のような結果が得られた.

- ・ テキスト情報のみから抽出された人物: 5 名
- ・ 画像情報のみから抽出された人物: 41 名
- ・ 両者から共通して抽出された人物: 27 名

図 6 に例示したグラフを見ても, 実際に重要と思われる人物や関係が欠落している箇所があることが分かる. 以上より, テキスト情報・画像情報の双方を利用した抽出を行うことの重要性が示唆された.

5. モノログシーンの検出

モノログシーンとは, TREC Video[1, 2]の高次特徴抽出検索タスクにおいて, “News subject monologue: segment contains an event in which a single person, a news subject not a news person, speaks for a long time without interruption by another speaker. Pauses are ok if short.”とあるように, 番組関係者以外の人物が発言・演説・インタビューに答えるシーンのことである. 本研究では, テキストにない映像ならではの特徴とは, 発言内容にとどまらず発話者の容貌や雰囲気を知ることができることと考え, このようなモノログシーンを抽出することとした.

従来 TREC Video の参加者らを中心として, モノログシーン検出手法[13, 14]が提案されてきた. しかし, これらの手法では, 番組関係者の名前や話者モデルなどの外部情報を事前に与える必要があった. 本研究ではそのような外部情報が入手できないニュース映像を想定し, 入力映像中の音声から話者モデルを動的に作成して, 番組関係者の発話を除去する手法を提案する.

以下の各節で本手法の処理の流れについて述べる. なお, 詳細については文献[15]を参照されたい.

5.1. モノログシーン候補の検出

モノログシーンが備えるべき要件として, 顔の存在がある. 通常は正面向きの顔がフレーム中央部に大きく映っていることが想定されるため, 以下の条件を満たす顔が検出された場合, モノログシーンの候補とする.

- ・ 向き: 4.2 節で用いたのと同じ OpenCV 付属の顔検出プログラム[12]が対応できる範囲の正面顔.
- ・ 位置: フレームを 4×4 ブロックに分割したとき, 重心が中央部の 2×2 ブロック内に存在.
- ・ 大きさ: フレーム全体の 8%以上の面積.

5.2. 番組関係者による発話の除去

5.1 節で検出されたモノログシーン候補のなかには, 番組関係者のモノログや, 画像は一般の人物であるものの音声は番組関係者, というシーンが含まれる. そこで, 以下の手順に従って話者モデルを動的に作成して, 番組関係者の発話区間を除去する.

番組関係者の発話モデルの動的生成

番組関係者はスタジオで原稿を読み上げるキャストとその他のレポーターなどからなる. 本手法ではキャストは 1 名しか存在しないという仮定をおき, 個別に発話箇所を推定する.

- ・ キャスタの発話箇所
CC テキスト冒頭の発話はキャストであると推定.
- ・ その他の番組関係者の発話箇所
CC テキスト中で以下のいずれかの条件を満たす文の直後の発話はその他の番組関係者であると推定.
 - ・ 条件 1: 呼びかけ.
末尾が“[固有名詞]+さん”.
 - ・ 条件 2: “記者”, “取材”, “中継”のいずれかの語を含み, 末尾が過去形ではない.

CC テキストと音声は事前に同期してあるので, このようにして検出された発話箇所を開始時刻とし, 次の無音区間(0.5 秒以上 FFT のパワースペクトラムが一定値以下の区間)の直前までの区間の音声特徴を抽出する. 話者モデルはベクトル量子化法により作成する.

番組関係者による発話の検出と除去

以上のようにして作成したモデルを用い, モノログショット候補の音声部と照合し, 番組関係者による音声と判断した場合は候補から棄却する. その結果, 一般のモノログショットが残ることになる.

5.3. モノログへの話者名の付与

次に, 検出されたモノログの話者を推定する. 本研究では事前に外部情報を与えないことを想定しているため, 下記の手順により, CC テキストから推定する. 同様の機能の実現を目指した先行研究として, NameIt システムにおける顔画像への人名付与手法[16]がある. 本研究でもその条件を準用し, 下記の条件を同時に満たす人物名詞を人物名候補として索引付けする.

- ・ **条件 1:** モノログシーンと同じストーリー中で言及される.
- ・ **条件 2:** モノログシーンの直前または直後の文中で言及される.

さらに, その人物名詞が文の主格をとる場合に $w(>1)$ 倍加重する.

5.4. 実験

2004 年 1 月のニュース映像 31 本(約 15 時間)に提案手法を適用した. 形態素解析には JUMAN[6] 5.1 を, 構文解析や時制の判定には KNP[17] 2.0 を用いた.

その結果, 5.1 節で述べた手順により検出されたモノログシーンは, 再現率 78.5%, 適合率 30.4%の精度で検出された. さらに 5.2 節で述べた手順により番組関係者を除去したところ, 再現率 76.6%, 適合率 55.0%の精度が得られるようになった.

5.3 節で述べた手順による発話者名の索引付け実験において, 主格をとる場合の重みを経験的に $w = 3.5$ とした. また, 本手法は CC テキスト中に存在しない名前は索引付けできないため, 人手で CC テキスト中に正解があることを確認したモノログショットに限って実験を行った. その結果, 上位 3 位以内に正解話者名を含んだモノログショットは 62%であった.

以上の実験により, 事前に外部情報を与えることなく, おおむね良好にモノログシーンを検出できることが示された.

6. おわりに

本講演では, 大規模ニュース映像アーカイブから登場人物に関する様々な情報を抽出する手法及びそれらを統合的に提供するインタフェースのプロトタイプを紹介した. 今後は, 各処理の性能向上及び自動化に取り組むほか, 5W1H の他の情報との連携や, 抽出した情報を利用したより高度の意味内容理解について検討する.

謝辞

本研究の一部は文部科学省科学研究費補助金(課題番号 15700116, 16016289), 名古屋大学 21 世紀 COE プロジェクト「社会情報基盤のための音声・映像の知的統合」, 国立情報学研究所との共同研究による.

- [1] United States National Institute of Standard and Technology: “TREC Video Retrieval Evaluation”, <http://www.nist.gov/projects/trecvid/>
- [2] 帆足啓一郎, 菅野 勝, 松本一則: “映像情報検索とその評価技術の最前線”, 情報処理, 46, 9, pp.1016-1023 (Sept. 2005)
- [3] 片山紀生, 孟 洋, 佐藤真一: “映像インデクシング研究のための大規模放送映像アーカイブシステムの試作”, 情処学研報 (DBS), 2002, 41, pp.17-23 (May 2002)
- [4] M.G. Christel, M. Olligschlaeger, and C. Huang: “Interactive maps for a digital video library”, IEEE Multimedia Mag., 7, 1, pp.60-67 (Jan.-Feb. 2000)
- [5] I. Ide, R. Hamada, S. Sakai, and H. Tanaka: “An attribute based news video indexing”, Proc. ACM Multimedia 2001 Workshops –Multimedia Information Retrieval–, pp.70-73 (Oct. 2001)
- [6] 京都大学長尾研究室, 東京大学黒橋研究室: “日本語形態素解析システム JUMAN”, <http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html>
- [7] 井手一郎, 浜田玲子, 坂井修一, 田中英彦: “テレビニュース字幕の語義属性解析のための辞書作成”, 信学論 (D-II), J85-D-II, 7, pp.1201-1210 (July 2002)
- [8] 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”, 信学技報, PRMU2003-97 (Sept. 2003)
- [9] H. Kautz, B. Selman, and M. Shah: “The hidden web”, AI Mag., 18, 2, pp.27-36 (Summer 1997)
- [10] I. Ide, T. Kinoshita, H. Mo, N. Katayama, and S. Satoh: “trackThem: Exploring a large-scale news video archive by tracking human relations”, Proc. 2nd Asia Information Retrieval Symp., Lecture Notes in Computer Science, 3689, pp.510-515 (Oct. 2005)
- [11] 小笠原崇, 井手一郎, 高橋友和, 村瀬 洋: “放送映像からの人物相関グラフの構築”, 第 19 回人工知能学会全大, 1F4-02 (June 2005)
- [12] Intel Corp.: “Open source computer vision library”, <http://www.intel.com/technology/computing/opencv>
- [13] A. Hauptmann, R.V. Baron, M.-Y. Chen, M. Christel, P. Duygulu, C. Huang, R. Jin, W.-H. Lin, T. Ng, N. Moraveji, N. Papernick, C.G.M. Snoek, G. Tzanetakis, J. Yang, R. Yan, and H. Wactlar: “Informedia at TRECVID2003: Analyzing and searching broadcast news video”, TREC Video Retrieval Evaluation 2003 Online Proc. (Nov. 2003)
- [14] A. Amir, M. Berg, S. Chang, W. Hsu, G. Iyengar, C. Lin, M. Naphade, A. Natsev, C. Neti, H. Nock, J.R. Smith, B. Tseng, Y. Wu, and D. Zhang, “IBM Research TRECVID-2003 video retrieval system”, TREC Video Retrieval Evaluation 2003 Online Proc. (Nov. 2003)
- [15] 關岡直城, 高橋友和, 井手一郎, 村瀬 洋: “ニュース映像中のモノログシーン検出による発言集の自動作成”, 信学技報 PRMU (Mar. 2005 発表予定)
- [16] 佐藤真一, 中村裕一, 金出武雄: “Name-It: 動画像処理と自然言語処理の統合による映像内容アクセス手法”, 第 3 回知能情報メディアシンポジウム論文集, pp.187-194, (Dec. 1997)
- [17] 京都大学長尾研究室, 東京大学黒橋研究室: “日本語構文解析システム KNP”, <http://www.kc.t.u-tokyo.ac.jp/nl-resource/knp.html>