論　文

放送番組を素材としたマルチメディア百科事典の自動構築

Automatic Generation of a Multimedia Encyclopedia from TV Programs

正会員　三浦菊佳[†1]，正会員　山田一郎[†1]，正会員　住吉英樹[†1]，
正会員　八木伸行[†1]，　奥村　学[†2]，　徳永健伸[†3]
Kikuka Miura[†1], Ichiro Yamada[†1], Hideki Sumiyoshi[†1],
Nobuyuki Yagi[†1], Manabu Okumura[†2] and Takenobu Tokunaga[†3]

Abstract　This paper proposes a method for automatically generating a multimedia encyclopedia composed of video clips using closed-caption text information. The goal is to automatically index each video segment of the television program by the principal video object. We focus on several features of the closed-caption text style in order to identify the principal video objects. Using Quinlan's C4.5 decision-tree learning algorithm and the predicted accuracies of production rule indicators, we extract one object noun for each video shot. To show the effectiveness of the method, we conducted experiments on the extraction of video segments in which animals appear in twenty television programs on animals and nature. We obtained a precision rate of 74.6 percent and a recall rate of 51.4 percent on the extraction of video segments in which animals appear, and generated a multimedia encyclopedia comprising 322 video clips showing 82 kinds of animals.

キーワード：クローズドキャプション，百科事典，被写体抽出，機械学習

1．ま　え　が　き

放送局では多種多様な番組が大量に制作されており，効率的な番組の蓄積と有効利用も求められている．NHKでは，現在，NHKアーカイブス[1)]として約60万本もの番組を蓄積しており，そのうち約6,000本を公開ライブラリーとして一般向けに提供している．しかし，大部分の番組は新たな番組制作のために局内で参照される程度で，充分活用されているとは言いがたい．放送番組には，良質で信頼性の高い情報が豊富に含まれるため，重要な情報を取出せればそれらを新たなコンテンツサービスの資源として利用できる．われわれはそのサービスの一つに，マルチメディア百科事典[2)]を考えている．放送された番組から特定のシーンや情報を抽出し，それらを集めて映像データベースを構築することで，映像付きの百科事典として活用する．例えば，自然や動物を扱う番組からは動物が映っている映像区間や，肉食・夜光性といった習性など，紀行番組からは歴史建造物やその土地の行事など，料理番組からは料理レシピなど，あらゆる番組からさまざまな知識を収集することを想定している．

番組をマルチメディア百科事典のような用途に活用するためには，番組のどの映像区間に何（被写体）が映っているのかという情報（メタデータ）が重要な役割を果たす．しかし，番組のほとんどはメタデータを付与されていない状態にあり，大量の番組に人手で付与していては多大な時間とコストがかかる．そこでわれわれは，自動で番組を解析してメタデータを生成する研究を進めている[3)]．特に，映像の被写体が何であるかは，映像を2次利用する上で重要な情報であるが，画像認識による解析は番組映像のようにあらゆる条件の被写体を特定するには困難であり，映像とともに扱われる文字，音声や言語情報を材料とするのが現実的だと思われる．

総務省では，聴覚障害者のために2007年までに付与可能なすべてのテレビ番組で字幕放送を行うことを目標に掲げており，字幕放送番組が近年急激に増加している[4)]．この字幕情報（以後，「クローズドキャプション」）は，番組中の出演者の発話内容やナレーションをもとに作成されている．そのため，映像内容を説明した文章を多く含んでおり，映

2007年5月2日受付，2007年9月6日再受付，2007年10月12日採録
†1 NHK 放送技術研究所
（〒157-8510 世田谷区砧1-10-11，TEL 03-5494-3383）
†2 東京工業大学 精密工学研究所
（〒226-8503 横浜市緑区長津田町4259，TEL 045-924-5067）
†3 東京工業大学 大学院 情報理工学研究科
（〒152-8552 目黒区大岡山2-12-1，TEL 03-5734-2105）
†1 NHK Science and Technical Research Laboratories
(1-10-11, Kinuta, Setagaya-ku, Tokyo, 157-8510, Japan)
†2 Precision and Intelligence Laboratory, Tokyo Institute of Technology
(4259, Nagatsuta-cho, Midori-ku, Yokohama city, Kanagawa, 226-8503, Japan)
†3 Department of Computer Science, Tokyo Institute of Technology
(2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552, Japan)

像中の被写体を推定するために有益な情報と考えられる．

そこで，本論文ではクローズドキャプションを利用して，映像中の主要な被写体（以後，「主被写体」）を推定する手法を提案する．テレビ番組で使われる言葉には，その言い回しに一定の特徴が存在する．例えば，「場所紹介」や「人物紹介」など特定の事柄を表現するために同じような言い回しが多用されることが報告されている[5]．映像中の主被写体を説明するときの表現にも同様に一定の特徴があると仮定し，提案手法ではこの特徴を決定木アルゴリズムにより学習する．このとき，映像の切替わり点により区切られる映像区間を映像の単位（以後，「映像カット」）と考え，一つの映像カットから一つの主被写体を推定する．そして，主被写体が推定された映像区間をビデオクリップとして収集し，マルチメディア百科事典を自動生成する．

以下，2章で関連研究についてまとめ，3章で映像とクローズドキャプションの関係についての調査結果を報告する．4章で主被写体推定処理の詳細を述べ，動物や自然をテーマとした20番組を対象とした実験を5章で説明し，従来手法との比較を行う．6章で処理結果を利用したアプリケーションとしてマルチメディア百科事典について言及し，7章でまとめと今後の課題について述べる．

2．関　連　研　究

これまでに，クローズドキャプションを利用した映像中の被写体を推定する手法として，SatohらによるName-It[6]がある．この手法では，顔画像解析とオープンキャプション（専用の機器を通さずに見ることのできる通常の字幕）の認識処理を組合せて高精度に映像中の人物を特定している．しかし，ニュース映像の人物のみを対象としているため，あらゆる被写体に適用するのは難しい．

また，Agnihotriらはトーク番組や報道番組の要約手法[7]を提案しており，キーワードやキーフレーズを手がかりとしている．柴田らは料理番組において，材料などの被写体特定を行っている[8]が，談話構造解析のほか映像処理を利用してクローズアップショットにおける被写体の特定を行っており，あらゆる大きさ，位置での被写体を特定することは難しい．

またGoogle社は，クローズドキャプションを利用して番組を検索し，代表画像とともに提示するシステムGoogle Video[9]を公開している．このシステムは，検索語をクローズドキャプションに含む番組の提示を行っている．しかし，必ずしも検索語が被写体となっている映像が抽出されているとは言えない．

3．予　備　調　査

表1にクローズドキャプションの例を示す．本論文で扱うクローズドキャプションには，番組開始時刻からの「時：分：秒.フレーム」で表される画面提示開始時刻と，そこで表示されるテキストのデータが含まれる．この時刻により映像と対応付けることができる．

表1　クローズドキャプションの例
Examples of closed captions.

画面提示時刻	テキスト
00:22:53.22	スイギュウの母親が気づきました。
00:23:04.25	ライオンの位置を確認するといちもくさんに走り出します。
00:23:18.26	スイギュウは深みに向かって逃げ込みます。
00:23:23.11	泳いで後を追うライオン。
00:23:28.28	スイギュウの巨体は水をはね飛ばし群れで駆け抜けます。
00:23:37.02	ライオンはあっという間に離されてしまいました。
00:23:41.22	水の深いところに逃げ込まれてはライオンも手が出ません。
00:23:47.05	湿地の深みを上手に利用してスイギュウが逃げきりました。

クローズドキャプションには映像内容を具体的に説明する文章が存在する．われわれは，このような文章を見つけることができれば，映像に現れる主被写体を推定できると考えた．そこで，NHKで放送された『地球！ふしぎ大自然』20番組のクローズドキャプションを対象として予備調査を行った．

例えば，新たな被写体が映像に登場したとき，クローズドキャプション中にそれに言及する文が現れる可能性は高いと考えられる．時間対応するクローズドキャプション中に映像中の主被写体を表す名詞がある映像カットのうち，その名詞が映像カット点直後の一文中にあった割合は84%であった．これにより，映像の切替わりとクローズドキャプションには相関があるといえる．

また，主被写体を推定するために，目的の言葉を抜き出すキーワードマッチングを行うだけでは，映像に現れる主被写体を精度良く取出すことができない．例えば，表1のクローズドキャプションの例において，実際の映像で主被写体がライオンであるのは4行目の文に対応するカットのみである．他の文に対応するカットでは，ライオンという単語が出現しているにもかかわらずスイギュウが主被写体であり，ライオンは主被写体ではない．つまり，ライオンという単語がクローズドキャプションに存在する映像を単純に抜き出すだけでは目的外の映像を抽出してしまう可能性がある．そこで，映像中の主被写体を説明している文を抽出し，解析して主被写体名詞を抽出しなくてはならない．

例えば，「泳いで後を追うライオン」といった体言止め（省略の形）の文や，「これはライオンです」といった文末が断定の助動詞の文は，映像内容を説明していると考えられる．そこで，前述の二つの特徴を持つ具象物名詞が，対応する映像カットの主被写体となっていたかを調査した．結

表2　“体言止め”，“名詞＋「です」”が映像の主被写体を表現する割合
Results regarding whether the concrete nouns in sentences ending in the target noun or in a phrase composed of the target noun and the predicative auxiliary verb "desu" refer to principal video objects.

適合率	再現率	F 値
479 / 748 (64.0%)	479 / 2776 (17.3%)	0.272

果を表2に示す．ここで具象物名詞とは，目で見て手で触れられるものを表現した名詞を指し，国立国語研究所の分類語彙表[10]の上位5桁を判断基準とした．適合率の結果から，体言止めの具象物名詞や断定の助動詞「です」が後続する具象物名詞は，映像カットの主被写体となる傾向があることがわかる．しかし再現率が低いことから，ほかの文体表現による主被写体の説明が行われていると考えられる．再現率を改善するためには，このほかの特徴も抽出する必要がある．

4．主被写体推定処理

前章で，映像内容を説明する記述には体言止めや断定の助動詞の後続など一定の特徴があることを示した．本章では，機械学習を用いて上記以外の特徴も抽出し，映像カットの主被写体をクローズドキャプションから推定する手法について述べる．処理手順を図1に示す．まず，映像カットごとに主被写体を表す名詞の正解値が与えられた学習データから特徴を抽出する．そして，それをQuinlanのC4.5決定木学習アルゴリズム[11]に入力する．この結果，クローズドキャプションに含まれる各名詞について主被写体であるか否かを判定する決定木が生成される．次に，生成された決定木を，優先順位付けされたプロダクションルールに変換する．このプロダクションルールを利用して，テストデータの名詞について主被写体であるか否かを判定し，主被写体候補を抽出する．この際，一つの映像カットに対応する文から複数の主被写体候補名詞が抽出された場合，予測精度を指標に主被写体名詞を一つに絞る．以下に，特徴抽出，決定木生成，プロダクションルール生成，主被写体推定について説明する．

4.1　特徴抽出

映像内容を説明する文の特徴を抽出するために，クローズドキャプションの各文に含まれるすべての自立語名詞を対象として表3に示す属性に対する属性値を付与する．表中の①は，主格や所有格などの文の格構造の特徴を抽出する属性である．②は，助動詞の省略など文末の文体を判定する属性である．③は，固有名詞や具象物名詞などの名詞の判定をする属性である．⑤は修飾句の数を獲得する属性であり，主被写体となる重要な名詞には補足説明が多用されると考えられるため設けている．⑦は有無を判定する属性であり，「これ」などの指示詞は映像中の被写体を参照することが多いと報告されている[12]ため設けている．⑧は，「ある」「いる」など，存在を表す動詞に係るか否かを判定す

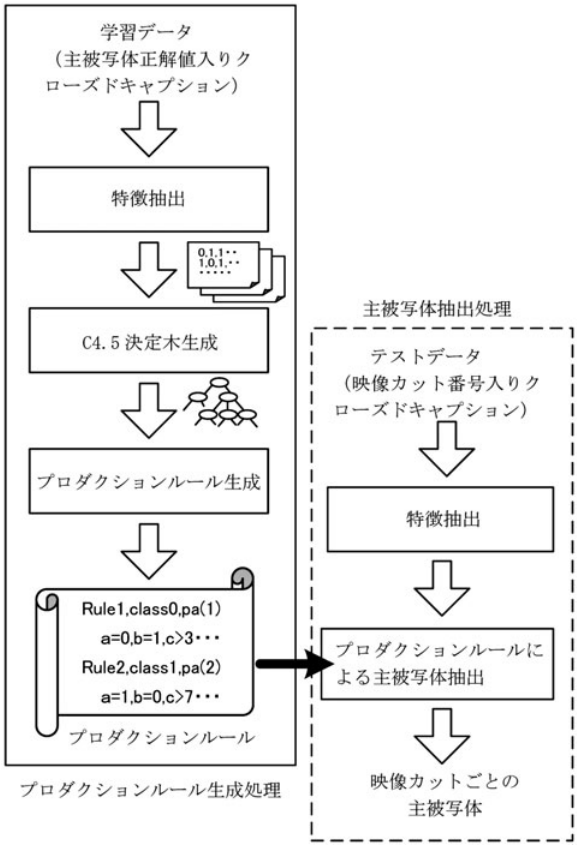


図1　主被写体抽出処理手順
Procedures used for extracting principal video objects.

表3　クローズドキャプションから抽出した特徴
Features extracted from closed captions.

属性	属性値
①対象名詞が含まれる名詞句の付属語の種類	は格 or NOT，が格 or NOT，の格 or NOT，に格 or NOT，を格 or NOT，と格 or NOT，へ格 or NOT，で格 or NOT，副助詞 or NOT
②文末の文体，句の係り先	体言止め or NOT，断定の助動詞「です」が後続 or NOT，対象名詞句の係り先が最終文節 or NOT
③対象名詞の種類	固有名詞 or NOT，サ変名詞 or NOT，形容動詞語幹 or NOT，数量名詞 or NOT，番組中の新語 or NOT，具象物名詞 or NOT
④対象名詞の重要度	正の数値（TFIDF 値）
⑤対象名詞句を修飾する文節数	自然数（0，1，2，…）
⑥主語の有無	主語有 or 主語無
⑦指示語の有無	指示語有 or 指示語無 or 指示語自体
⑧存在を表す動詞の有無	存在を表す動詞有 or 存在を表す動詞無
⑨カット点との関係	カット点後 1 文目 or NOT

る属性である．⑨は映像カットの切替わり直後の一文目か否かを判定する属性であり，映像が切替わった直後に被写体の紹介をすることが多いと考えられるため設けている．

4.2 決定木生成

映像の主被写体となる名詞が，上記の特徴のどのような組合せになるかを学習する．抽出した特徴を用いて，「主被写体である」，「主被写体でない」の2クラスに分類する決定木を生成する．まず，映像カットごとに主被写体が特定された正解値を持つクローズドキャプションを学習データとする．そして，学習データのクローズドキャプション中の全自立語名詞に4.1の特徴抽出処理を行う．そして，その結果をC4.5決定木学習の入力とする．決定木学習は，まずすべての属性を分岐条件の候補とし，分岐前後の情報利得比が最大になる属性を分岐条件とする．この処理を繰り返すことにより，中間ノードに分岐条件が配置され，葉ノードに学習データの事例が配置された決定木を生成する．

4.3 プロダクションルール生成

4.2で得られた決定木を，属性の分岐条件と分類クラスからなるルール記述の集合であるプロダクションルールに変換する．初期値としてすべての葉に対応するルールを生成する．この結果，葉の数だけルールが生成される．次に，各ルールから予測精度が低下しない範囲で分岐条件を除外する．ルールiに対する予測精度paは以下の式で算出される．

$$pa(i) = 100\left(1 - pe(i)\right) \tag{1}$$

$$pe(i) = \left\{ p \;\middle|\; \sum_{j=0}^{E_i} \binom{N_i}{j} p^j (1-p)^{N_i - j} = CF,\quad 0 \le p \le 1 \right\} \tag{2}$$

E_i ：ルールiにおける誤り事例数
N_i ：ルールiに適合する総事例数
CF：枝刈り度(0.25)

paは誤り率をはかる指標であり，CFとは枝刈りの度合いを示す定数で，本実験では標準値の0.25[11]を用いた．この値は，C4.5決定木学習アルゴリズムを用いたさまざまな手法で多用されている値である．分岐条件の除去により生じた重複ルールを除去し，残ったルールに対し予測精度の降順に優先順位付けを行う．この優先順位付けされたルール集合をプロダクションルールとする．プロダクションルールは，決定木の枝刈りと違い中間ノードに存在する分岐条件も除去することができ，また人による可読性に優れている．さらに，予測精度の値により各ルールの優先順位を付けることができる．これらの利点から，本手法では決定木により直接判定を行わず，プロダクションルールを採用している．

4.4 主被写体推定

学習データから生成されたプロダクションルールをテストデータに適用し，主被写体を推定する．ここでは，クローズドキャプションに対応する映像カット番号が付与されたテストデータを用い，映像カットごとに主被写体を推定する．まず，テストデータの各名詞に対してプロダクションルールに含まれるルールを優先順位順に適用する．最初に合致したルールが持つ分類クラスをその名詞の判定値とする．一つの映像カットに主被写体と判定された名詞が一つであった場合，その名詞を主被写体として出力する．一つの映像カット中に主被写体と判定された名詞が複数あった場合，最も予測精度の高いルールが適用された名詞を出力する．複数の主被写体と判定された名詞のうち，予測精度が同値であった場合，複数の選択を許し同値のものをすべて出力する．反対に，一つの映像カット内に主被写体と判定された名詞が一つも存在しない場合，「主被写体なし」と出力する．

5．主被写体推定実験

提案手法の有効性を検証するために，NHKで放送された番組『地球！ふしぎ大自然』を対象とした主被写体推定実験を行った．この番組は，世界の動物や植物などの紹介を趣旨に制作されている．本論文の目的である主被写体を説明する記述が多く，実験に適していると考えたため，この番組を使用した．以下，詳細を述べる．

5.1 プロダクションルールによる主被写体推定実験と考察

実験用データ作成のため，『地球！ふしぎ大自然』20番組分のクローズドキャプションの各文にカット番号を人手で付与した．なお，カット点は映像解析技術を利用すれば高い精度で自動抽出することもできる[13]．このクローズドキャプションに含まれるすべての自立語名詞に対して，人手により映像カットごとに一つの主被写体を特定し，これらを学習データとした．この結果，学習データに含まれる全自立語名詞19,520個の正解値の内訳は，2,776個（14.2%）が「主被写体である」，16,744個が「主被写体でない」であった．映像カットに対応するクローズドキャプションに主被写体となる名詞が存在しない場合も多くみられ，4,351カット中1,575カット(36.2%)に主被写体となる名詞が存在しなかった．

主被写体であるか否かの正解値が付けられた19番組を学習データ，残り1番組分のテストデータとし，クロスバリデーションにより合計20回の主被写体推定の評価実験を行った．この際，特徴抽出処理における係り受け解析には南瓜[14]を使用した．評価結果を**表4**に示す．提案手法による推定結果の F値は0.502であり，表2に示す体言止めと断定の助動詞を手掛かりとした手法のF値0.272と比べて有効性が確認できた．特に，再現率では26.4% 向上しており，プロダクションルールを用いた手法により表2の二つの文末表現以外の特徴も抽出できたといえる．

次に比較手法として，単語の重要度を利用して主被写体を推定する実験を行った．映像カットごとにTFIDF値が最も高い名詞を「主被写体である」としたときの実験結果を**表5**に示す．TFIDF値は，情報抽出において単語の重要度を示す代表的な指標の一つである[15]．再現率ではTFIDF値を用

表4　プロダクションルールによる主被写体抽出実験結果
Results of the experiment using the proposed method.

適合率	再現率	F値
1213/ 2075 (58.5%)	1213 / 2776 (43.7%)	0.502

表5　TFIDF値のみを利用した主被写体抽出実験結果
Experimental results when using the TFIDF value.

適合率	再現率	F値
1385 / 4354 (31.8%)	1385 / 2776 (49.9%)	0.389

いた手法が提案手法より良い結果が得られているが，適合率は低く，その調和平均を示すF値で比較すると，単語のTFIDF値のみで主被写体を推定するより，他の属性も考慮した提案手法が有効であるといえる．

本研究は，クローズドキャプションというテキスト情報から，映像カットの主被写体を推定することを目的としている．しかし番組では，映像内容についてまったく言及しない場合や，クローズドキャプションのみからでは推測できないような映像，例えば「ナマケモノはもう目の前です」の映像がナマケモノではなくオウギワシであるなど，を扱う場合もある．よって，クローズドキャプションのみからすべての主被写体を推定することは難しいと考えられる．そこで，どの程度までクローズドキャプションのみで主被写体が推定可能か，人手による実験を行った．正解データの作成者とは異なる被験者が，クローズドキャプションのみから主被写体を推定した結果と，映像を見て作成した正解データと比較した．実験と用いたものと同じ20番組を対象とした結果を表6に示す．この値が，自然言語処理によるアプローチの限界値と考える．提案手法による実験結果は，適合率は上限値65.3%のところ58.5%，再現率は77.5%のところ43.7%であり，一定の精度が得られているといえる．

5.2　プロダクションルールの検証

生成されるプロダクションルールは，対象とする学習データにより異なる．そこで，一つの番組に対して生成されたものを対象として，実験で得られたプロダクションルールの検証を行った．表7に主被写体推定に成功した事例数の上位4ルールと判定例を示す．表中のルール番号は，ルールが持つ優先順位を示し，小さい番号ほど優先度が高い．Rule23は，判定対象とする名詞のTFIDF値が一定値以上かつ断定の助動詞「です」を伴うという特徴を持つルールである．Rule25は，判定対象とする名詞が体言止めとなるという特徴を含むルールである．Rule42は，TFIDF値が一定値以上の重要度を持ち，かつカット点後の一文目となり，「は格」により文の主語となるという特徴を持つルールである．Rule27もRule42と同様，判定対象とする名詞のTFIDF値が一定値の重要度を持ち，かつカット点後の一文目となり，「が格」により文の主語となるという特徴を持つルールである．これらは主被写体の特徴的な表現であることがわかる．

表6　人による主被写体抽出結果
Experimental results for human identification.

適合率	再現率	F値
2151 / 3294 (65.3%)	2151 / 2776 (77.5%)	0.709

表7　主被写体抽出に成功した事例数上位4ルールと判定例
Top four rules in terms of the number of cases of successful principal video object extraction.

ルール番号［該当数］	ルール	判定例（下線部が主被写体と判定）
Rule 23［10事例］→主被写体と判定	②断定助動詞/③サ変名詞でない/④TFIDF値>10.7743	水の中のラッコです。
Rule 25［9事例］→主被写体と判定	②体言止め/③サ変名詞でない，具象物名詞/⑤修飾する文節数>0/⑥主語無/⑦指示語無	哺乳類の中で最後に住み着いたラッコ。
Rule 42［5事例］→主被写体と判定	①は格/③具象物名詞/④TFIDF 値>9.13727⑤修飾する文節数<=0/⑨カット点後1文目	ラッコは起きている時間のほとんどを毛繕いに費やします。
Rule 27［5事例］→主被写体と判定	①が格/②係り先最終文節/③サ変名詞でない、数量名詞でない/④TFIDF 値>10.7743/⑦指示語無⑨カット点後1文目	子供が2匹います。

提案手法により誤判定した事例のうち，人手による主被写体推定実験でも同じ誤判定であったものの割合を表8に示す．主被写体でない名詞を主被写体名詞と誤抽出してしまった37名詞のうち15名詞が，「葉が大好物」のように，葉も映っているが主被写体はサルであるケースや，「ジャイアントロベリアはサワギキョウの仲間」のように，ジャイアントロベリアと推定したが映像はサワギキョウであるケースなど，人間も誤って抽出していた．このケースは言語処理のアプローチのみからでは困難だと考えられる．また，提案手法で主被写体名詞を抽出できなかった81名詞のうち，36名詞が人手でも抽出できていなかった．

主被写体推定に失敗した原因として，「AのB」という表現の解析の難しさが挙げられる．例えば「ニホンジカの群れです」という文に対して，提案手法では被写体である「ニ

表8　提案手法で失敗した事例のうち人手でも同じ失敗をした割合
Ratio of the number of manual extraction errors to the number of errors when the proposed method was used.

主被写体でない名詞を主被写体と誤抽出	主被写体名詞を未抽出
15 / 37	36 / 81

表9　被写体を動物に限定したときの抽出結果
Evaluation results of experiment for the objects of animals using the proposed method.

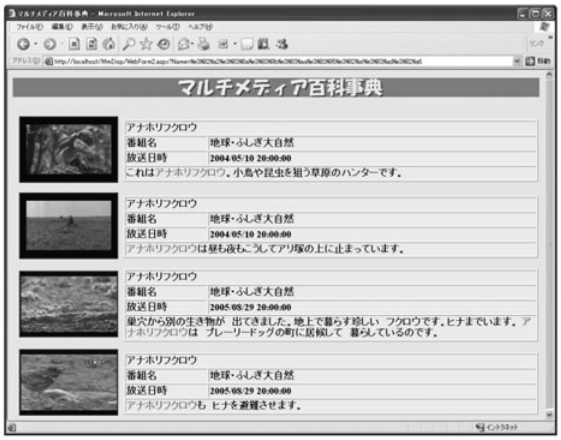
適合率	再現率	F 値
332 /445 (74.6%)	332 / 646 (51.4%)	0.609

ホンジカ」は「の格」を伴っているため被写体として抽出され難く，「群れ」は具象物でないため被写体なしと推定される．人間であれば，「群れ」は「ニホンジカ」の様態を表し，「ニホンジカ」が被写体となっていると想像できる．提案手法では，構文特徴を重視して単語が持つ語彙的な特徴を充分考慮していなかったため，人間による推定結果との再現率に差が生じた．単語が持つ意味概念などを決定木学習の素性として与えることにより，提案手法における再現率の改善が見込まれる．また，同じ被写体でも「ラッコ」(動物名)，「コジロウ」(愛称)，「生き物」(分類名) など番組中で複数の異なる表現がなされており，言い換え表現の抽出処理が必要と考えられる．談話構造解析などを利用して文脈を理解することで，精度の向上が図れると考えられる．

6．マルチメディア百科事典への応用例

実験で用いた『地球！ふしぎ大自然』20番組を対象として，提案手法により動物が主被写体となっている映像区間を抽出した．この結果，82種類の動物の映像区間322個が得られた．その結果を表9に示す．処理対象を動物に限定したため精度は向上している．この抽出した映像区間をビデオクリップとしてデータベースに登録し，Web上からアクセスすることにより，興味のある動物の動画を容易に視聴できるマルチメディア百科事典を試作した．図2にマルチメディア百科事典の画面を示す．図2 (a) は，トップ画面で，番組から主被写体として抽出された動物名を表している．各項目のサムネイル画像は，ビデオクリップの初めの1フレームから作成している．サムネイルをクリックすると2ページ目が開かれ (図2 (b))，対象番組から集められたビデオクリップの一覧が示される．さらに，そのサムネイルをクリックすると，目的の動物のビデオクリップとその映像区間のクローズドキャプションを見ることができる．また，トップメニューの検索ボックスに目的の動物の名前を直接入れて，目的の動物のビデオクリップを検索することもできる．

(a) トップメニュー



(b) 動物別メニュー

(c) 動画再生画面

図2　マルチメディア百科事典画面
Generated multimedia encyclopedia.

このように，アーカイブスのさまざまな番組から必要なシーンを項目ごとにまとめた百科事典を構築すれば，貴重な映像資産を有効利用できる．このような用途において，提案した手法は効果的で充分な精度を持つと考える．現状では，大量の映像の中からユーザが希望する被写体の映像を見つけるために相当な時間を必要とするが，提案手法により構築したマルチメディア百科事典では，あらかじめ映像に対する被写体が推定され，整理されているため，映像を検索する時間を大幅に短縮できる．解析結果には誤りも含まれているが，映像とナレーションを見ればユーザは誤

りか否かを容易に識別できる．また，予測精度の高いルールのみを採用することにより適合率を上げることができ，低下する再現率は処理対象とする番組数を増やすことによりカバーすることができると考える．今回の実験では20番組を処理対象としたが，本手法を他の番組に適用していくことにより，さらに大量のビデオクリップを自動的に獲得できる．このように，過去の良質な番組を対象として自動的に百科事典を構築すれば，教育用途などで放送された番組を有効に活用することが期待できる．

7．むすび

本論文では，放送番組中で映像の主被写体を説明するときの言葉の特徴を機械学習し，主被写体を映像カットごとに推定する手法を提案した．動物や自然をテーマに扱う20番組を対象とした実験を行った結果，単語の重要度を利用した従来手法と比較して，提案手法の有効性を確認した．また，推定された映像区間を集めてマルチメディア百科事典の試作機を自動構築した．このように，映像の特定の区間を収集してアプリケーションとして活用することにより，教育用途などへの応用が期待できる．今回は自然番組に適用したが，同じように画面に映っている被写体を説明することを意図した，紀行・教養・ドキュメンタリなどの番組についても応用可能と考える．一方，ドラマやニュースなど被写体を説明することを意図しない番組については，被写体を説明する表現自体が少ないため本手法の応用は難しい．

今後，言い換え表現の抽出，「AのB」の表現の解析により，今回の実験において誤判定されたものについて改善をはかる予定である．また，ゼロ主語補完，代名詞の照応解析などを考慮していくことにより，クローズドキャプション中に主被写体を表す名詞が直接出現していない場合の推定処理を進める．

さらに，動物の習性や行動などを表現する映像の自動抽出に取組み，より充実した内容を提供できるマルチメディア百科事典の構築を目指す．

〔文　献〕

1) NHKアーカイブス，http://www.nhk.or.jp/nhk-archives/
2) “マルチメディア百科事典 膨大な映像資産の有効活用に向けて”，NHK技研R&D，99，p.56（2006）
3) 住吉，佐野，山田，松井，Clippingdale，望月，三須，佐藤，小林，今井，松村，八木：“メタデータ制作・活用システムの試作”，映情学技報，29，70，BCT2005-159，ME2005-205，AIT2005-120，pp.13-18（2005）
4) 総務省：“平成16年度の字幕放送の実績”，http://www.soumu.go.jp/s-news/2005/050811_6.html
5) 山田，三浦，住吉，八木，奥村，徳永：“AdaBoostを利用した字幕テキストからの定型表現文章区間抽出”，情処学自然言語処理研資，2006，82，2006-NL-174（5），pp.25-30（2006）
6) S. Satoh and Y. Nakamura and T. Kanade: "Name-It; Naming and Detecting Faces in Video by the Integration of Image and Natural Language Processing", IJCAI-97, pp.1488-1493（1997）
7) L. Agnihotri, K. Devara, T. McGee and N. Dimitrova: "Summarization of Video Programs based on Closed Captions", Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Media Databases, Proc. SPIE, 4315, San Jose, pp.599-607（2001）
8) 柴田，黒橋：“言語情報と映像情報を統合した隠れマルコフモデルに基づくトピック推定”，第12回言語処理学会年次大会発表論文集，pp.476-479（2006）
9) Google Video，http://video.google.com/
10) 国立国語研究所：“分類語彙表”，増補改定版（2004）
11) J.R. Quinlan: "C4.5 Programs for Machine Learning", Morgan Kaufmann（1993）
12) 西田，徳永，山田：“テレビの情報番組における指示詞の照応先同定”，第12回言語処理学会年次大会発表論文集，pp.472-475（2005）
13) B.L. Yeo and B. Liu: "Rapid Scene Analysis on Compressed Video", IEEE Trans. Circuits & Syst. Video Technol., 5, pp.533-544（1995）
14) 工藤，松本：“チャンキングの段階適用による係り受け解析”，情処学論，43，6，pp.1834-1842（2002）
15) G. Salton and C.S. Yang: "On the Specification of Term Values in Automatic Indexing", Journal of Documentation, 29, 4, pp.351-372（1973）

三浦　菊佳　2002年，慶應義塾大学理工学部物理情報工学科卒業．同年，NHK入局．名古屋放送局を経て，2004年より，放送技術研究所にて，自然言語処理，情報抽出の研究に従事．正会員．

山田　一郎　1991年，名古屋大学工学部情報工学科卒業．1993年，同大大学院修士課程修了．同年，NHK入局．名古屋放送局を経て，1996年より，放送技術研究所にて，自然言語処理を利用した情報抽出，メタデータ生成，知識獲得の研究に従事．2003年～2004年，スタンフォード大客員研究員．現在，放送技術研究所専任研究員．正会員．

住吉　英樹　1980年，広島県立広島工業高校電気科卒業．同年，NHK入局．広島放送局を経て，1984年より，放送技術研究所にて，コンピュータを応用した番組制作システムの研究に従事．現在，放送技術研究所専任研究員．工学博士．正会員．

八木　伸行　1978年，京都大学工学部電気工学科卒業．1980年，同大学大学院電気工学専攻修士課程修了．同年，NHK入局．甲府放送局，放送技術研究所，技術局，編成局を経て，現在，放送技術研究所知能処理グループリーダー．2005年より，東京工業大学特任教授（兼任）．画像・映像・メディア情報処理，コンピュータアーキテクチャ，コンテンツ制作技術，デジタル放送などの研究開発に従事．工学博士．正会員．

奥村　学　1984年，東京工業大学工学部情報工学科卒業．1989年，同大学院博士課程修了．同年，東京工業大学工学部情報工学科助手．1992年，北陸先端科学技術大学院大学情報科学研究科助教授．2000年，東京工業大学精密工学研究所助教授．現在，同大学精密工学研究所准教授．自然言語処理，知的情報提示技術，語学学習支援，テキストマイニングに関する研究に従事．工学博士．

徳永　健伸　1983年，東京工業大学工学部情報工学科卒業．1985年，同大学院理工学研究科修士課程修了．同年，（株）三菱総合研究所入社．1986年，東京工業大学大学院博士課程入学．現在，同大学大学院情報理工学研究科准教授．自然言語処理，計算言語学，情報検索などの研究に従事．工学博士．