ショット分類に基づく映像への自動的索引付け手法

井手　一郎[†]　　山本　晃司[†*]　　浜田　玲子[†]　　田中　英彦[†]

An Automatic Video Indexing Method Based on Shot Classification

Ichiro IDE[†], Koji YAMAMOTO[†*], Reiko HAMADA[†], and Hidehiko TANAKA[†]

あらまし　増大する量への対応及び利用価値の点から，ニュース映像への自動的索引付けの実現が期待されている．これを受けて様々な手法が提案され，なかでも映像に付随する言語情報を利用したものが活発に研究され，既に実用化の域に達しているものも存在する．しかし，それらの多くは言語情報主導であり，映像データベースにとって重要であるはずの，索引（キーワード）と画像内容の一致を考慮したものは少ない．そこで本論文では，このような問題点を踏まえ，画像的に類似した映像は類似した内容を含む，というニュース映像特有の性質を利用し，画像的に典型的な映像に対し，内容を反映した語義をもつ字幕を選択的に付与する自動的索引付け手法を提案し，その有効性を評価する．提案手法を実際のニュース映像に適用したところ，いずれかの典型的映像に分類されたもののうち，全体の 25〜93%，必要な情報が字幕に存在して索引付け可能であったもののうちの 75〜100%に対して索引付けに成功し，一定の有効性を示した．

キーワード　ニュース映像，映像データベース，自動索引付け，ショット分類，字幕解析

1. まえがき

増大する映像資源，殊に速報性と利用価値の点からニュース映像への索引付けを行う必要が高まっている．しかし，現在この作業は主に人手で行われており，増大する量への対応及び，高度な検索を可能にするためのきめ細かな索引付けを行うには不十分であり，自動化が期待されている．

そこで筆者らは，映像中の画像情報及び言語情報を統合的に利用して，このような自動的索引付けの実現を目指している．既に同様の手法による研究が活発に行われているものの，それらの多くは言語情報主導であり，映像データベースにとって重要であるはずの，索引（キーワード）と画像内容の一致を考慮したものは少ない．

本論文ではこのような問題点を踏まえ，画像的に典型的な映像に対し，内容を反映した語義をもつ索引を選択的に付与する自動的索引付け手法を提案し，その有効性を評価する．提案手法は，画像的に類似した映像は類似した内容を含む，というニュース映像特有の性質を利用している．具体的には，典型的なショットに分類した映像に対し，付随する言語情報の中から適切な語義をもつものを選択的に付与することにより，映像内容を反映した索引付けの実現を目指す．

まず，**2.** でテレビニュース映像の特徴と関連研究についてまとめ，**3.** で提案手法の全体像を紹介する．続く **4.**，**5.**，**6.** で各処理の詳細の紹介及び性能の評価を行い，**7.** で結論と今後の課題について述べる．

2. ニュース映像への索引付け

まずニュース映像に特有の性質についてまとめた後に，関連研究の紹介を行う．

2.1 ニュース映像の構成

以下で述べるように，ニュース映像は画像的及び内容的な点から構成を考えることができる．

2.1.1 画像的構成

ニュース映像に限らず，一般に映像（video）は図 1 に示すような階層的構成からなる．用語を以下のように定義する．

- フレーム（frame）：映像を構成する静止画像．
- ショット（shot）：画像的に連続なフレーム群．
- シーン（scene）：画像的に類似したショット群．
- カット（cut）：隣接ショット間の不連続点．

[†] 東京大学大学院工学系研究科，東京都
Graduate School of Engineering, The University of Tokyo, Tokyo 113–8656 Japan
[*] 現在，（株）東芝研究開発センター

2.1.2 内容的構成

図 2 に示すように，ニュース映像の内容的構成は特有の性質をもつ．図中の各ブロックはショットを表し，一つの話題（topic）は「キャスタショット（anchor shot）」で始まり，次のキャスタショットまでの間に「報告シーン（report scene）」などをいくつか挟む構成になっている．

索引付けの際には，話題の境界を判定し，内容的構成を把握することが重要であるため，このような性質は重要な手掛りになる．また，ニュース映像は類似した状況で撮影されることが多いため，画像的特徴量から多くのショットを数通りの典型的なショット分類に分類することができると思われる．

2.2 映像中の言語情報

2.2.1 言語情報源

映像に付随する言語情報として，主音声，副音声，クローズドキャプション，字幕（オープンキャプション）のように，様々なものがある．なかでも字幕は，重要な情報を端的に表すのに用いられることが多く，索引付けに用いる言語情報として有望である．筆者らの統計によると，1 分間に 4 件程度出現することが判明しており，キーワード候補として用いるのに適した頻度であると考えられる．一方で，主音声やクローズドキャプションを利用すると，キーワード候補を得るために冗長な情報の中から重要文や重要語の抽出を行う必要があり，提案手法ではこのような煩雑な前処理を行う必要のない字幕を言語情報源として採用する．

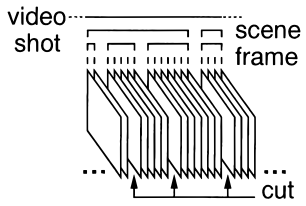

図 1 映像の画像的構成
Fig. 1 Graphical structure of video.

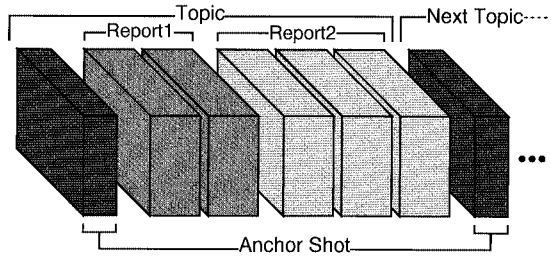

図 2 映像の内容的構成
Fig. 2 Semantic structure of video.

2.2.2 字幕の特徴

字幕は特定の助詞や体言が末尾に存在したり，文中からは省略されたりし，通常の文とは文法的に若干異なる性質をもち [7], [13], [16]，解析にあたり既存の言語処理手法をそのまま適用するのは困難である．このため，**5.** で紹介する手法により，提案手法で必要となる語義属性の解析を行う．

一方，字幕を意味的に分類すると，表 1 に示すようになる．このうち，全字幕の約半数を占める（a），（b），（g）は映像内容を具体的に表すものであり，直ちにキーワード候補になり得る．また，（c）は映像内容を具体的に表すとは限らないものの，話題内容を特定するための重要な手掛りになる．これらを合わせ，全体の 6 割程度の字幕が，映像内容に則した言語情報源として何らかの形で利用可能である．提案手法では，解析しやすさの点から，これらのうち（a），（b），（c）を索引付けに用いる．

2.3 関連研究紹介

一般的な映像データベース構築・検索システムとしては，Informedia プロジェクト [14], [17], [24] が著明である．一連の研究の中でも，CNN（Cable News Network）社のニュース映像を対象にした自動蓄積・検索システムである News-on-Demand [8] は，主音声の認識から得られたテキストやクローズドキャプションから，TF-IDF（Term Frequency Inverse Document Frequency）法により語の珍しさを統計的に判断してキーワードの抽出を行う．このような統計的手法は比較的簡便であり，その点において実用的だが，映像内容と索引との対応が必ずしも保証されないという本質的な問題点がある．

本論文で提案する手法を含め，このような本質的な

表 1 字幕の意味的分類
Table 1 Types of television news captions.

	分類	割合	具体的内容
（a）	場所・組織	30%	○
（b）	人物	15%	○
（c）	タイトル	14%	△
（d）	発言の要約・翻訳	10%	×
（e）	時相	7%	×
（f）	放送技術的	3%	×
（g）	描写	2%	○
（h）	その他	19%	—

問題点を解決するために，映像内容を考慮した自動的索引付け手法がいくつか試みられている．Nakamura と Kanade [10] は，ニュース映像のショットを会見・報告，集合，屋外の 3 通りに分類し，一方でクローズドキャプション中の文を文法的・意味的解析により会見，会議，集合，訪問，場所の 5 通りに分類し，それぞれ適切な対応関係にあるものを対応づけている(注 1)．この手法は，そもそもクローズドキャプションを用いる点，その中の重要文（キーセンテンス）を索引付けに用いる点で提案手法と手法的に異なる．なぜならば，我が国でもテレビジョン放送に対するクローズドキャプションの付与が進められているものの，日本語の場合は漢字変換を伴うことから，ニュース映像のような生放送の映像への実時間での付与は困難であり，その利用を前提にできないためである．また，提案手法では字幕を用いることにより，語句単位での索引付けを行う点で，困難とされる一般的な構文解析や文中の重要語抽出を避けることができ，文単位での索引付けよりもきめ細かな検索要求にこたえられるものと考えられる．

また，これに類似した手法として，Satoh ら [2] は，ニュース映像からの顔画像と人物名の自動的対応付け手法 Name-It を提案した．この手法もクローズドキャプションの表記法の性質を利用して人物名の抽出を行っているほか，顔画像と人物名の対応のみをとる点は，応用例によっては許容可能な制約であるものの，一般的なニュース映像データベースへの索引付け手法としては不十分である．

3. ショット分類に基づく索引付け手法

2. で述べた点を考慮し，ショット分類に基づく索引付け手法を提案する．この手法では，画像的に類似したショットは類似した内容を含む，というニュース映像特有の性質を利用している．このような性質を仮定して，画像的に典型的なショットに対し，特定の語義属性をもつ字幕を選択的に付与する．

提案手法の機構の全体像を図 3 に示す．現時点では，字幕認識及び音声とクローズドキャプションの利用に関する処理は実装されておらず，前者は人手で行った結果を利用し，後者は重要文抽出や要約手法 [1], [15] の利用による将来の拡張を検討している．

本章では，各処理部の簡単な説明を行い，ショット分類，字幕解析，索引付けに関する詳細と評価については，各々章を改めて述べる．

3.1 画像処理部

（1） 映像のディジタル化

まず，PC 上の画像取込みボードを用いて，アナログ信号で録画された映像をディジタル化する．ここでは容量の関係から，NTSC 放送方式の 30 フレーム毎秒から 15 フレーム毎秒にフレームレートを落としてディジタル化を行った．また，フレームの大きさは 320×240 ピクセルとした．

（2） カット検出

カット検出には，大辻ら [20] により比較検討された様々な手法に始まり，MPEG 圧縮のデータ形式を利用した手法 [9] などが存在するが，提案手法では非圧縮映像に対して高精度でカット検出が行える離散余弦変換（DCT）特徴のクラスタリングによる手法 [11] を採用する．この手法は，連続するフレームの DCT 成分によりクラスタを形成し，クラスタからある程度逸脱したフレームが一定数連続した場合にカットを検出する．図 4 に，30 分間のニュース映像（計 208 カット）に対して，クラスタの大きさを 15 フレーム，逸脱フレーム数を 5 フレームとし，逸脱と判断するしきい値を 0.5σ から 4.0σ の間で変えたときの再現率と適合率との関係を求めた結果を示す．この結果，逸脱と判断するしきい値を適切に設定することにより，95%程度の適合率と 80%程度の再現率でカットを検出できることがわかった．

（3） 字幕認識

画像処理の本流と並行して，字幕領域の検出と文字認識を行う必要がある．高精度の OCR（光学的文字認識）技術は存在するものの，テレビ映像中の字幕は，走査線の本数（NTSC 方式の場合 525 本）が少ないために低解像度であることと，文字の背景が均一でないことから，既存手法での認識は困難である．これらの問題は，特に漢字など複雑な文字を扱う際に顕著であるが，背景からの文字の分離と認識に関する優れた手法も提案されつつある [4]〜[6]．このような技術の進歩と合わせて，将来の放送のディジタル化に伴い，字幕に相当する情報が電子的なテキストとして映像に付随して放送される可能性も考慮に入れ，本論文では，字幕認識は実装せずに人手で書き下したものを利用した実験結果を示す．

(注 1)：これらの分類名称は本論文で提案する手法による分類と対応するように訳した．

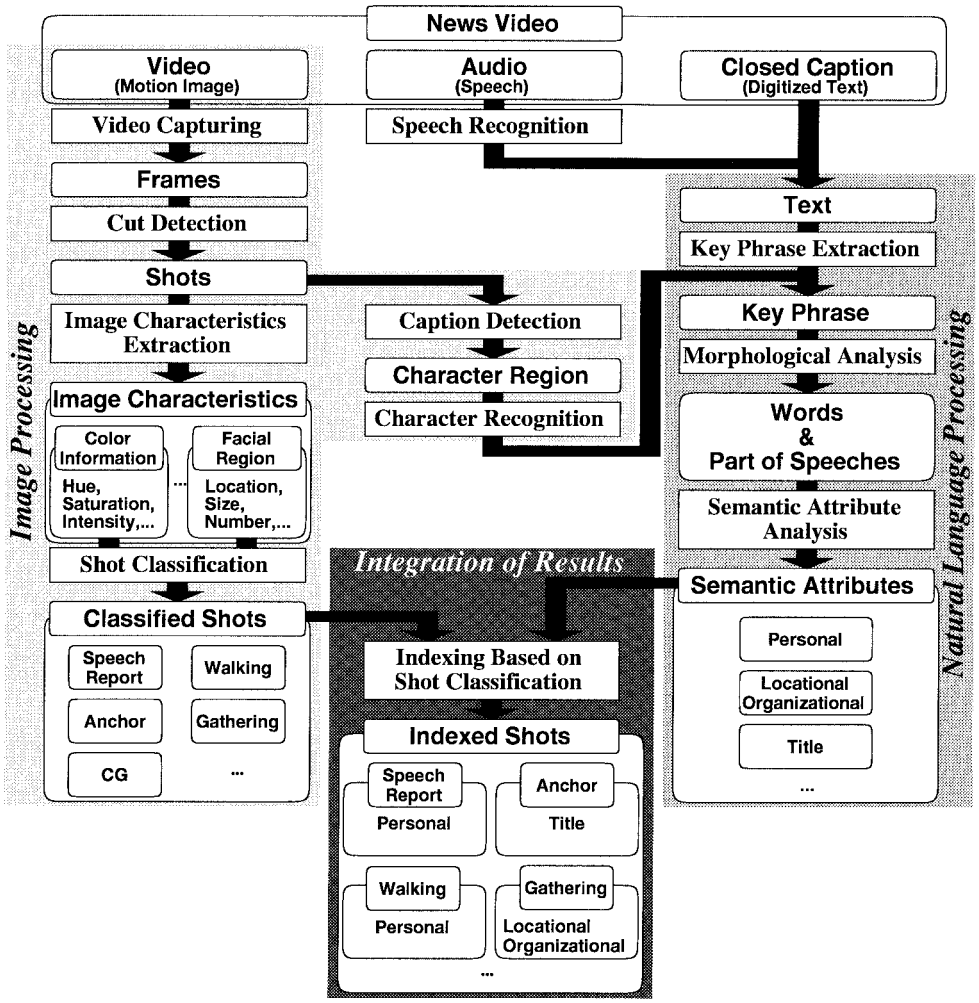

図 3　索引付け機構の全体像
Fig. 3　Overall indexing scheme.

3.2　言語処理部

まず下準備として，字幕から得られた電子化されたテキストに対し，日本語形態素解析システム JUMAN [21] を用いて形態素解析を行う．その後，末尾に名詞が存在するものに対して，その名詞に着目した語義属性の解析を行う．

3.3　統合処理部

ショット分類と字幕解析の結果を受けて，統合処理部で実際の索引付けを行う．ここでは，各々のショット分類に応じて適切な語義の字幕が選択的に付与される．具体的には，人物が会見しているショットに対しては人物に関する字幕を，人々が集合しているショットに対しては集合に関する字幕を付与する，といった索引付けを行う．これにより，映像内容を反映した索引付けが可能になる．

4.　ショット分類

3.1 で述べたような前処理を経た後，各ショットを画像的特徴に基づき，以下の 5 通りに分類する．

- 会見・報告（report/speech）ショット
- キャスタ（anchor）ショット
- 歩行（walking）ショット

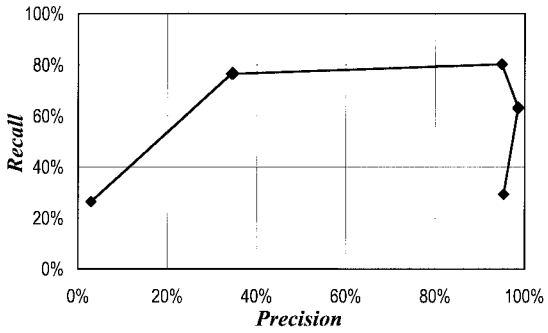

図 4　カット検出性能：再現率と適合率の関係
Fig. 4　Cut detection performance: Relation between precision and recall.

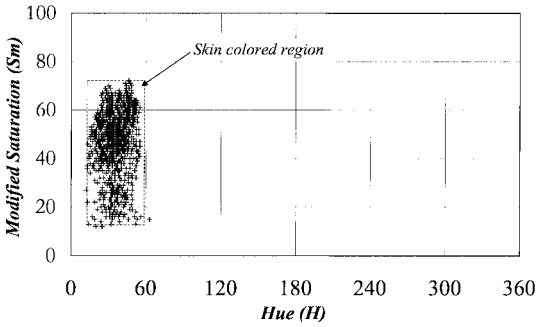

図 5　H-S_m 空間における顔領域の分布
Fig. 5　Skin colored regions on the H-S_m plane.

- 集合（gathering）ショット
- コンピュータグラフィックス（CG）ショット

これらの分類は，実験に用いた全ショットの 57%を占め，残りはいずれにも分類されなかった．

分類規則の詳細を以下で述べるが，大量のデータを処理する際の計算量を極力減らすことを考慮し，複雑な処理により得られる高度な認識結果や特徴量ではなく，比較的簡素な抽出過程により得られる特徴量を主に用いている．

4.1　会見・報告ショット

通常，ニュース映像中で，人物が会見を行ったり，レポータが中継地から報告を行っているショットには，画面の中央部に一つの顔が比較的大きめに映っている．そのようなショットを「会見・報告ショット」と名づけ，（1）顔領域の存在，（2）唇の動きの検出，により検出する．条件（2）は資料映像の肖像写真や単に画面の中央に人物が存在するだけのショットを除くために必要である．

（1）　顔領域の検出

顔領域抽出手法については非常に多くの手法が提案されているが，提案手法では以下の手順に示すような比較的簡単な手法で十分であると判断した．

（a）　肌色領域抽出

肌色領域の抽出には，修正 HSI 表色系 [19] を用いた．このうち，I（明度）は暗い領域を排除するためにのみ用いる．実際の顔領域の色を参考に，H（色相）-S_m（修正彩度）空間中の方形領域を肌色領域と設定した（図 5 参照）．H と S_m の値がこの領域に含まれるピクセルは肌色であると判断した後，数ピクセルのクラスタ成長を経て，一定のピクセル数以上のクラスタを顔領域と判断した．

（b）　テンプレートマッチング

テンプレートマッチングは，肌色領域として抽出された手，壁，机の表面などを除外するために行う．いくつかの顔領域の I から複数の解像度の平均顔を作成しておき，抽出された肌色領域の大きさに応じて比較を行う．なお，撮影条件の違いによる影響を避けるために，肌色領域の I をフレーム全体の I の値によって正規化しておく．

（2）　唇の動きの検出

会見・報告ショットやキャスタショットの場合，映っている顔は良好な照明条件下で撮影された正面顔であることが多く，顔領域が検出されれば，唇の位置を推定することは比較的容易である．推定された唇の位置の周囲での動きが顔領域の他の位置よりも大きければ，唇の動きがあるものと判断する．現時点では唇の動きの検出は，この基準に基づき目視により行っているが，将来的には自動化できるものと考えている．

4.2　キャスタショット

「キャスタショット」は，スタジオでキャスタが放送原稿を読み上げているショットであり，ニュース映像の内容的構成を把握する上で重要な手掛りになる．

（1）　会見・報告ショットからの分離

キャスタショットの検出は基本的に会見・報告ショットと同様に行うため，これらから分離する必要がある．キャスタショットの特色として，スタジオ内で同じような構図で撮影されるため，相互に画像的に非常に類似した特徴を示し，かつ出現頻度が高いことが挙げられる．これらの特色を利用して，会見・報告ショットの検出後，それら相互間の画像的類似度を求め，その結果得られた最大かつ最密なクラスタに含まれるショッ

トをキャスタショットとして分離する．なお，画像的類似度は，色ヒストグラムの差異を χ^2 検定して求めた．

（2） 話題境界の検出

キャスタショットを検出する最大の目的は，話題の境界を検出することにある．しかし，話題の途中に出現することもあり，単純に出現と境界を同一視することはできない．そこで，話題の境界のキャスタショットの特色である，タイトル字幕の出現を利用する．

図 6 に示すように，タイトル字幕の出現前後でフレーム全体のエッジ強度が大きく変化する．そこで，エッジ強度に大きな変化があるキャスタショットを話題の先頭と判断する．

4.3 歩行ショット

「歩行ショット」は，人物が歩行しているショットで，会見や報告以外の人物に関する話題に登場する．歩行の際，顔を含む人物の上半身は上下に振動することを利用して検出する．通常テレビカメラは三脚などに据え付けられているため，上下に動かないものと仮定する．そこで，図 7 に示すように，顔領域の最下部の上下動を追跡することで歩行ショットを検出する．

4.4 集合ショット

「集合ショット」は，会議や集会など，映像中に複数の人物が存在するショットである．一定の大きさ以上の複数の顔領域が存在する場合に集合ショットを検出する．

4.5 CG ショット

「CG ショット」は，CG（コンピュータグラフィックス）で作成されたショットであり，参考資料的な肖像写真や解説のための説明図など，他のショット分類や字幕解析に悪影響を及ぼしかねない内容を含む．そこで，索引付け機構から除去するために分類を行う．CG ショットの特色としては，比較的動きが少ない点が挙げられ，ショット中の静止時間の合計に基づき検出する．以下の実験では，この時間を 1 秒に設定した．

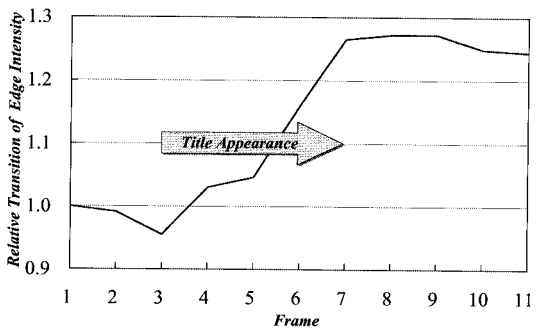

図 6 タイトル字幕の出現に伴うエッジ強度の相対的変化
Fig. 6 Relative transition of edge intensity following the appearance of title caption.

4.6 ショット分類実験

以上の分類規則に基づき，75 分間（計 468 ショット）のニュース映像に対して分類実験を行ったところ，表 2 に示すような結果が得られた．評価に用いる正解は人間が判断した．また，再現率と適合率を以下のように定める．

$$\text{再現率} = \frac{\text{正検出数}}{\text{正検出数} + \text{検出漏れ数}} \quad (1)$$

$$\text{適合率} = \frac{\text{正検出数}}{\text{正検出数} + \text{誤検出数}} \quad (2)$$

誤検出と検出漏れの主な原因は以下のとおりであった．

- 会見・報告ショットで，顔領域が正面顔でなかったため唇の動きが検出できなかった．
- 非常に短い話題の先頭のキャスタショットにタイトルが現れなかった．
- 集合ショットで，背後から撮影されていたために顔領域が検出できなかった．
- 歩行ショットで，遠方から撮影されていたため顔領域の上下動を検出できなかった．

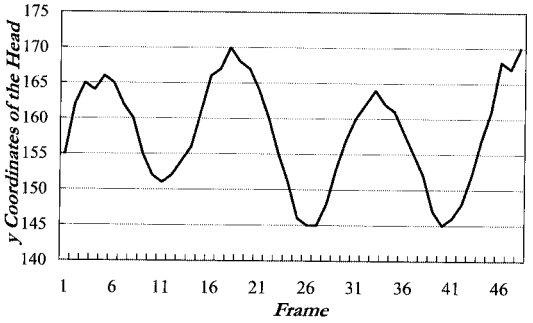

図 7 歩行に伴う顔領域の上下動
Fig. 7 Oscillation of facial region following a person's step.

表 2 ショット分類実験の結果
Table 2 Result of shot classification experiment.

ショット分類	正検出	誤検出	検出漏れ	再現率	適合率
会見・報告	55	3	12	83%	95%
キャスタ	53	0	0	100%	100%
〃（先頭）	37	0	6	86%	100%
歩行	16	0	15	52%	100%
集合	70	13	62	53%	84%
CG	14	1	0	100%	93%

- 静止物体を撮影した映像を CG ショットと誤った．

5. 字幕の語義属性解析

字幕の語義属性解析は，各ショット分類に適切な語義をもつキーワードを付与するために必要な処理である．表 1 に示したように,（1）人物,（2）場所・組織に関する字幕は量的にも内容的にもキーワード候補として適している．これらのほとんどは名詞句であり，日本語において名詞句の語義は末尾の名詞によって定まる傾向があることを利用して語義解析を行う(注 2)．

名詞の語義属性解析に関する関連研究はいくつか存在する．那須川 [12] は，前後の文脈からの固有名詞の語義決定，つまり地名か人名かの判断を行う手法を提案している．一方，渡辺ら [16] は，フレーム中の字幕の出現位置や文法的特徴を参考にしたニュース映像中の字幕の解析手法を提案している．

これらの手法は優れた語義解析性能を有するものの，前後の文脈を把握するに足る言語情報が存在しない字幕に対して前者を適用することは困難である．また固有名詞に限定している点でも提案手法での使用には適さない．一方，後者に関しては，提案手法と類似した目的で使用することを想定した手法であるものの，番組によって異なるデザインなどの編集方針の影響を受けるおそれがある．

更に，情報抽出の分野では，MUC（Message Understanding Conference）の課題として定義された，Named Entity Task [18] が存在する．この課題は，テキスト中の人物，組織，場所，時相，数値にタグ付けを行うものである．しかし，この課題もまた，タグ付け対象のうち，人物，組織，場所については固有名詞に限定している点で提案手法における解析目的と異なる．

提案手法ではこれらの点を踏まえ，末尾の名詞に着目して，各字幕単体での語義解析を目指す．

5.1 「人物名詞」と「場所・組織名詞」の収集

末尾の名詞に着目した名詞句（字幕）の語義解析を行うためには，どのような名詞が（1）人物を,（2）場所・組織を示すかを知る必要がある．このような名詞を「人物名詞」と「場所・組織名詞」と呼ぶことにし，人手による形態素解析済みの新聞記事からなる二つのテキストコーパス [22], [23] から抽出した．

このような名詞の収集の詳細については他稿 [7] に譲るが,「人物名詞」として 3,793 語が,「場所・組織名詞」として 11,166 語が収集された．収集された名詞は固有名詞を含まず,「ボランティア」,「氏」や「台所」,「駅」といった普通名詞であることに注意されたい．

表 3　字幕解析実験の結果
Table 3　Result of caption analysis experiment.

語義属性	正検出	誤検出	検出漏れ	再現率	適合率
人物	79	19	4	95%	81%
場所・組織	140	97	20	88%	59%

5.2 字幕解析実験

ショット分類実験に用いたのと同じ 75 分間のニュース映像中の字幕を解析した結果を表 3 に示す．解析の手順は以下のとおりである．

（1） 日本語形態素解析システム JUMAN [21] を用いて字幕に形態素解析を施す．

（2） 末尾の形態素が固有名詞（人名，地名，組織名）と判定された場合は，その分類に従い，普通名詞と判断された場合は，**5.1** で収集した各辞書中の語との一致を見て分類を行う．

なお，評価のための正解は第三者が判断して決定した．

誤検出と検出漏れの主な原因は以下のとおりであった．

- 両者に属し得る名詞が存在（語義の多様性）．
- 不適格な名詞を収集（雑音）．
- 収集した名詞では不十分（語彙(い)不足）．

後二者は収集手法の改良による改善が見込まれるものの，語義の多様性に関しては，提案手法では対応することができない．場所・組織名詞の適合率が特に低い原因は，収集基準が緩いために相当数の不適格な名詞が混入したためである．これを改善するためには，より詳細な文法的情報が事前にテキストコーパスに付与されている必要がある．

6. 分類されたショットへの索引付け

ショット分類と字幕解析の結果を受けて，各ショット分類ごとに適した語義属性をもつ字幕を索引付けする．

6.1 索引付け手法

各ショット分類に索引付けする字幕の語義属性の対応を表 4 に示す．索引付けすべき字幕の候補を探す範囲は，タイトルが存在するキャスタショットにより検出される話題の境界内である．タイトルが存在するキャスタショットは，表 2 に示すように，比較的高精度で検出できる．

(注 2)：この性質は他のいくつかの東アジアの言語には該当すると思われるが，欧米の諸語には必ずしも該当しない．

表 4　ショット分類と索引付けする字幕の語義属性
Table 4　Shot class and semantics of indexes.

ショット分類	索引付けする字幕の属性
会見・報告	人物
キャスタ（先頭）	タイトル
歩行	人物
集合	場所・組織

表 5　索引付け結果
Table 5　Result of indexing.

ショット分類	全数	可能数	正解数	全数中	可能数中
会見・報告	38	30	25	66%	83%
キャスタ（先頭）	15	15	14	93%	93%
歩行	4	1	1	25%	100%
集合	59	44	33	56%	75%

一方，CG ショットは，その内容の特殊性を考え，索引付け対象から除外する．このような方針のもとに，以下の手順で索引付けを行った．

（1） ショット内で該当する属性をもつ字幕を探す．

（2） 存在しなければ，話題内の画像的に類似したショット内から探す．

6.2 索引付け実験

ショット分類と字幕解析に用いたのと同じニュース映像のうち，30 分の映像に対して索引付けを行った結果を表 5 に示す．ショット分類による索引付け手法単体での評価を行うために，字幕の語義属性解析の結果としては正解のものを用いた．評価は全数中と可能数中に分けて行った．全数とは，各々のショット分類に分類された全ショットの数であり，可能数とは，そのうち正しく索引付けし得る字幕が存在したものの数である．後者のような評価を行ったのは，索引付けに必要な情報が字幕に含まれないことがあるため，そのようなショットを除いた評価を行いたかったためである．

この結果を見ると，全数中では必ずしも実用的な性能を示していないが，可能数中ではすべて 75%以上の索引付け成功率を示しており，他の言語情報の利用により全数中の性能の向上も期待できることがわかる．

7. むすび

本論文では，ニュース映像に対する映像内容と索引との対応を考慮した索引付け手法の提案と評価を行った．全体的な結果は必ずしも実用的な性能を示していないが，手法単体での性能からは，更なる改良により実用的な性能が得られることが期待される．また，ショット分類や字幕解析のための基礎的な技術は既存の手法を利用したが，それらを組み合わせて利用し，画像情報と言語情報とを統合的に処理することの有効性も示した．

今後，映像データベースの検索に対する要求がよりきめ細かなものになることを考慮すると，提案手法のように，映像内容と索引との一致を考慮した索引付け手法の果たす役割は大きいものと考えられる．

問題点としては，ショット分類の種類が少ないため，手法単体としての性能が向上しても，ショット分類が被覆するショットの割合が全体の半数強程度であるため，総合的な索引付け性能の向上には限界がある点が挙げられる．これは明示的に分類規則を記述する以上はやむを得ないため，現在分類規則の統計的な自動獲得手法の検討を進めているところである [3].

謝辞　日本語形態素解析システム JUMAN [21] は，京都大学長尾研究室により開発・配布されているフリーソフトウェアである．京都大学テキストコーパス [22] は京都大学長尾研究室の成果物である．また，RWC テキストデータベース [23] は技術研究組合新情報処理開発機構（RWCP）の成果物であり，同機構の許可の下に利用した．

文　献

[1] 瀬戸喜巳，井手一郎，坂井修一，田中英彦，“見出しからの制約による新聞記事からの重要文抽出，” 平 11 情処学前期全大，vol.2, pp.73–74, March 1999.

[2] S. Satoh, Y. Nakamura, and T. Kanade, “Name-It: Naming and detecting faces in news videos,” IEEE MultiMedia, vol.6, no.1, pp.22–35, March 1999.

[3] 井手一郎，浜田玲子，坂井修一，田中英彦，“言語情報を伴う画像の画像的特徴量と語義の統計的対応付け，” 情処学コンピュータビジョンとイメージメディア研報 99-CVIM-114, pp.137–143, Jan. 1999.

[4] 堀　修，“テロップ認識のための映像からの文字部抽出法，” 情処学コンピュータビジョンとイメージメディア研報 99-CVIM-114，pp.129–136, Jan. 1999.

[5] 松浦克海，鷹尾誠一，杉山善明，有木康雄，“ニュース映像中のテロップ・フリップフレームの検出と文字抽出，” 信学技報，PRMU-98-188, Jan. 1999.

[6] M. Sawaki and N. Hagita, “Text-line extraction and character recognition of document headlines with graphical designs using complementary similarity measure,” IEEE Trans. Pattern Anal. & Mach. Intell., vol.20, no.10, pp.1103–1109, Oct. 1998.

[7] 井手一郎，田中英彦，“末尾の名詞に着目したテレビニュース字幕の語義解析，” 情処学論，vol.39, no.8, pp.2543–2546, Aug. 1998.

[8] H.D. Wactlar, A.G. Hauptmann, and M.J. Witbrock, “Informedia News-on-Demand: Using speech recognition to create a digital video library,” CMU Tech. Rep., CMU-CS-98-109, March 1998.

[9] 金子敏充，堀　修，“尤度比検定による MPEG 圧縮動画像

からのカット検出," 情処学コンピュータビジョンとイメージメディア研報 98-CVIM-109, pp.105–112, Jan. 1998.
[10] Y. Nakamura and T. Kanade, "Semantic analysis for video contents extraction — Spotting by association in news video," Proc. 5th ACM Intl. Multimedia Conf., Seattle WA, USA, pp.393–402, Nov. 1997.
[11] 有木康雄, "DCT 特徴のクラスタリングに基づくニュース映像のカット検出と記事切出し," 信学論（D-II）, vol.J80-D-II, no.9, pp.2421–2427, Sept. 1997.
[12] 那須川哲哉, "文脈情報を利用したキーワード語義決定," 1997 人工知能学全大, no.17-1, pp.348–349, June 1997.
[13] 若尾孝博, 江原暉将, 白井克彦, "テレビニュース番組の字幕に見られる要約の手法," 情処学自然言語処理研報 97-NL-122, pp.83–89, Nov. 1997.
[14] M.A. Smith and T. Kanade, "Video skimming and characterization through the combination of image and language understanding techniques," CMU Tech. Rep., CMU-CS-97-111, Feb. 1997.
[15] 仲尾由雄, "見出しを利用した新聞・レポートからのダイジェスト情報の抽出," 情処学自然言語処理研報 97-NL-117, pp.121–128, Jan. 1997.
[16] 渡辺靖彦, 岡田至弘, 長尾 眞, "TV ニュースで用いられるテロップの意味解析," 情処学自然言語処理研報 96-NL-116, pp.107–114, Nov. 1996.
[17] 金出武雄, 佐藤真一, "Informedia：CMU ディジタルビデオライブラリプロジェクト," 情報処理, vol.37, no.9, pp.841–847, Sept. 1996.
[18] United States Defense Advanced Research Projects Agency (DARPA), Information Technology Office, "Named entity task definition, version 2.1," Proc. 6th Message Understanding Conf., Columbia MD, USA, pp.317–332, Nov. 1995.
[19] 松橋 聡, 藤本研司, 中村 納, 南 敏, "顔領域抽出に有効な修正 HSV 表示系の提案," テレビ誌, vol.49, no.6, pp.787–797, June 1995.
[20] 大辻清太, 外村佳伸, "映像カット自動検出方式の検討," テレビ学技報, vol.16, no.43, pp.7–12, July 1992.
[21] 京都大学大学院情報学研究科知能情報学専攻長尾研究室, "日本語形態素解析システム JUMAN version 3.6," Nov. 1998.
[22] 京都大学大学院情報学研究科知能情報学専攻長尾研究室, "京都大学テキストコーパス version 2.0," June 1998.
[23] 技術研究組合新情報処理開発機構（RWCP）, "RWC テキストデータベース version 1.0," March 1996.
[24] "The Informedia project"; http://www.informedia.cs.cmu.edu/.

（平成 11 年 2 月 25 日受付, 6 月 11 日再受付）

井手 一郎 （学生員）

平 6 東大・工・電子卒. 平 8 同大大学院工学系研究科情報工学専攻修士課程了. 修士（工学）. 現在同研究科電気工学専攻博士課程在学中. 自然言語処理, 情報抽出, マルチメディア統合処理に興味をもっている. 平 7 情報処理学会第 51 回全国大会奨励賞受賞. 人工知能学会, 情報処理学会各学生会員.

山本 晃司 （正員）

平 8 東大・工・電子情報卒. 平 10 同大大学院工学系研究科電気工学専攻修士課程了. 修士（工学）. 同年（株）東芝入社. 現在同社研究開発センター勤務.

浜田 玲子 （学生員）

平 10 東大・工・電子情報卒. 現在同大大学院工学系研究科電気工学専攻修士課程在学中. 情報処理学会学生会員.

田中 英彦 （正員）

昭 40 東大・工・電子卒. 昭 45 同大大学院博士課程了. 工博. 同年東京大学工学部講師. 昭 46 同助教授. 昭 62 より同教授, 現在同大学院工学系研究科教授. この間昭 53〜54 ニューヨーク市立大学客員教授. 計算機アーキテクチャ, 並列処理, 自然言語処理, メディア処理, 分散処理, CAD 等の研究に興味をもっている. 著書「非ノイマンコンピュータ」,「情報通信システム」, 共著書「計算機アーキテクチャ」,「VLSI コンピュータ I, II」,「ソフトウェア指向アーキテクチャ」. 情報処理学会, 人工知能学会, 日本ソフトウェア科学会, IEEE, ACM 各会員.