半自動ビデオアノテーションとそれに基づく意味的ビデオ検索*

山本　大介
名古屋大学 工学部 電気電子・情報工学科
yamamoto@nagao.nuie.nagoya-u.ac.jp

長尾　確
名古屋大学 情報メディア教育センター
nagao@nuie.nagoya-u.ac.jp

1　はじめに

近年 Web ページをはじめ、さまざまな情報検索が頻繁に行われている。しかしながら、ビデオコンテンツに対する Web 検索はいまだ実用化されているとは言い難い。ビデオコンテンツに対する検索にはさまざまな手法 [4] が存在するが、ビデオコンテンツを全自動で解析した結果に基づいて検索する場合、精度の観点からきわめて不十分である。検索の精度を十分に実用的なレベルに引き上げるためにはビデオコンテンツに検索や変換・編集等に有効な意味内容記述をなんらかの方法により付加する必要がある。そこで、コンピュータによるビデオコンテンツの自動解析を行い、人間がその解析結果を効率よく修正・補完できるツールを作成した。さらにそのツールを使用して得られたアノテーションデータに基づいて、高度な意味的ビデオ検索を Web ブラウザを用いて自然言語で行うシステムを試作した。

将来的には MPEG7[1] への対応も考えている。

2　ビデオアノテーションエディタ

本研究で作成したアノテーションツールをビデオアノテーションエディタ（以下 VAE と略す）と呼ぶ。長尾ら [2] が作成したバージョンの VAE をベースに新たに作り直した。VAE は動画像に対してカット検出、オブジェクトトラッキング、シーンおよびオブジェクトへのアノテーション、音声認識を用いたトランスクリプトの作成、XML データ出力等が行えるツールである。

主要な機能として以下のものを備えている。

1. カット検出

   カット検出は、RGB 空間を 4096 分割したカラーヒストグラムを用いて各画素の絶対値差分の合計がある閾値以上になれば新たなカットであると認識している。

2. オブジェクトトラッキング

   オブジェクトトラッキングダイアログ (図 1) を利用しておこなう。アルゴリズムは、矩形範囲をキー画像として、テンプレートマッチングを行っている。トラッキングしたオブジェクトの始めの画像と終わりの画像・さらにその前後 0.1 秒の画像を表示し、トラッキングが成功しているかどうか、一目でわかるように工夫してある。また、手動での修正も可能である。また、MPEG コンテンツはランダムアクセスが遅いために、3 秒ごとに動画をメモリ上に展開し処理を行った。

図 1: ビデオアノテーションエディタの操作画面

3. 複数選択式アノテーション

   あらかじめ、アノテーションに対する 3 つの定義ファイル (オブジェクトの属性を定義する objectDefinitions.xml, オブジェクトの動作を定義する motionDefinitions.xml, シーンの状況を定義する sceneDefinitions.xml) を用意し、それぞれプルダウンメニューを選択することにより意味内容を記

*Semiautomated Video Annotation and its Application to Semantic Video Search by Daisuke Yamamoto (Dept. of Information Engineering, School of Engineering, Nagoya University) and Katashi Nagao (Center for Information Media Studies, Nagoya University)

述する。複数の項目を同時に選択することにより、より複雑な状況も記述可能である。また、ユーザが独自にこれらの XML 定義ファイルを拡張することも可能である。

XML 定義ファイルには、新たな項目を作るだけでなくその項目の説明をする必要がある。RDF スキーマ [3] などグラフ構造を用いた定義の表現方法が存在するが今回はより簡略かつ検索に使いやすいように、新しい項目に関するさまざまな同義語 (日本語、英語を含む) を列挙することにした。この方式ならば、手軽に項目追加が可能であるし、検索時に完全一致、あるいは、部分一致が容易であると考えられるからである。

4. 音声認識

   IBM の音声認識ソフト ViaVoice を使って音声認識を行う。その音声が発話された時間区間の自動抽出も行う。認識結果の修正機能も備えている。

5. 階層構造の表現とシーンの重要度の推定

   文章などと同様に映像にもカットを単位とした階層構造が存在し、半自動的にアノテーションすることが可能である。類似カットのつながりを類似度に応じて自動的にグルーピングすることができる。類似度は、ヒストグラム、シーンの長さ、音声などを考慮して実現可能である。

   また重要度をシーンの長さとアノテーションのデータ量に応じて上げる試みもしている。

6. XML 出力

   記述内容の拡張性と Web ベース検索の容易性を考慮し XML ファイルによる出力を採用した。

3 自然言語による意味的ビデオ検索

VAE によって作られた XML アノテーションデータを、Web ブラウザを用いて検索するシステム (図 2) を Java Servlet と XML データベースを用いて試作した。検索は、自然言語入力によって行っている。

アルゴリズムとしては以下のようになる。

1. 検索キーワードから茶筅を用いて、形容詞・動詞・名詞を取り出す。

2. 形容詞から色にあたる単語（たとえば、赤い・黒い・青い・暗い・明るい等）がある場合はシーンもしくはオブジェクトのヒストグラムも利用して検索結果を絞りこむ。このとき、色にあたる形容詞にかかる名詞が「場面」「光景」「風景」「シーン」「画面」等の場合はシーンについて語っている可能性が高く、それ以外の場合はオブジェクトに関する場合が高いのでそれに応じて点数をつける。

   また、これ以外の名詞・形容詞・動詞は、アノテーションデータに記述されたテキスト情報もしくは、選択式アノテーションによりつけられた記述との部分もしくは完全一致により点数をつける。

3. オブジェクトもしくはシーンを点数順に並び替え、順位づけされた検索結果をユーザーに提示する。

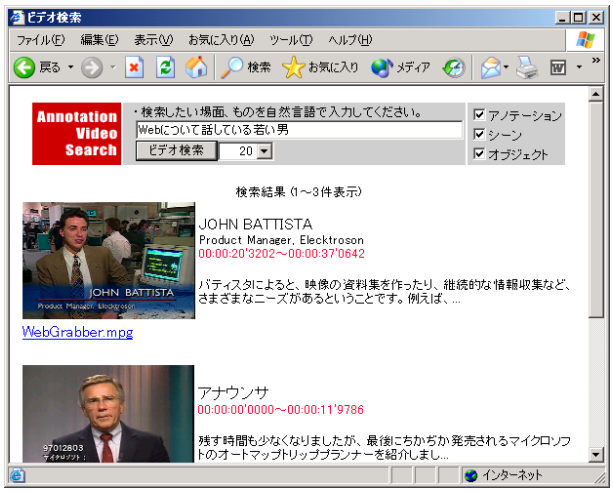

図 2: ビデオの検索画面例

4 おわりに

今回は、ビデオコンテンツに対するアノテーションツールと、自然言語による検索ツールを試作した。従来難しいと思われていた自然言語によるビデオコンテンツ検索が、アノテーションを併用することにより比較的容易になることを示した。これにより、ユーザは Google と同様の感覚で動画像データを意味的に検索できるようになる。

参考文献

[1] MPEG. Mpeg-7, 2002. http://ipsi.fraunhofer.de/delite/Projects/MPEG7/.

[2] Katashi Nagao, Shigeki Ohira, and Mitsuhiro Yonehisa. Annotation-based multimedia summarization and translation. In *Proceedings of the Nineteenth International Conference on Computational Linguistics (COLING-2002)*, 2002.

[3] W3C. Resource description framework (rdf), 2001. http://www.w3.org/RDF/.

[4] 西尾章治朗, 田中克巳, 上原邦昭, 有木康雄, 加藤俊一, 河野浩之. 情報の構造化と検索. 岩波書店, 2000.