オンラインアノテーションに基づくビデオシーン検索

増田 智樹 † 山本 大介 †† 大平 茂輝 ‡ 長尾 確 ‡‡
† 名古屋大学 工学部 電気電子・情報工学科 †† 名古屋大学 大学院情報科学研究科
‡ 名古屋大学 エコトピア科学研究所 ‡‡ 名古屋大学 情報メディア教育センター

1 はじめに

近年、インターネット技術の発達により Web 上で膨大な数のビデオコンテンツが配信されるようになった。それに伴いビデオコンテンツの検索、要約などに対する要求が高まっている。

筆者らは近年 Web 上で行われているビデオコンテンツを中心とした自然なコミュニケーション活動から得られる情報が、ビデオの検索や要約に利用可能であると考え、それらの情報をオンラインビデオアノテーションとして利用するシステム Synvie を開発した [1]。

本研究では、オンラインアノテーションの有用性を確認するためにそれをビデオシーン検索に利用する仕組みを提案する。オンラインビデオアノテーションには、Synvie の一般公開実験 [2] から得られた情報を利用した。

まず、ビデオシーンに対するタグ（以後シーンタグと呼ぶ）の作成を行った。さらに、シーンタグを利用した新しい発想のビデオシーン検索システムを開発し、シーン検索の被験者実験を行った。

それによって、オンラインビデオアノテーションの有用性を検証した。

2 シーンタグの作成

Synvie に登録されたビデオコンテンツの内 27 個のビデオに対して 3 種類の手法でシーンタグを作成した。利用したビデオのビデオ時間は平均で約 349 秒、最長で 768 秒、最短で 76 秒であり、ビデオの種類は教育、物語、エンターテインメントなど様々なものがある。

2.1 専用ツールを用いたタグ付け

ビデオを視聴しながら、任意の開始時間、終了時間を指定して、そこにタグを付与することができるツールを利用し、1 人のアノテータによってタグ付けを行った。シーンのイベント情報や、人や物体とその様子、テロップ、音声などのオブジェクト情報をシーンタグとして付与した。

ここで、シーンタグの付与に費やした時間を、シーンタグ作成のためのコストとした。その時間は、1 コンテンツあたり平均で約 1480 秒、最長で 3692 秒、最短で 582 秒であった。

2.2 オンラインアノテーションからの自動抽出

Synvie は、Web 上でビデオの任意のシーンに対してコメントの投稿、ブログへの引用を行うことができるビデオ共有システムである。2006 年 7 月 1 日から 11 月 30 日までに一般公開実験によって得られたデータを利用した。この期間に登録されたビデオコンテンツ数は 94 個、ユーザ数は 97 人である。

Synvie のアノテーションからは、テキスト情報とそれに対応する時間情報を得ることができる。

まず、Cabocha を利用してテキスト情報の形態素解析を行い、また形態素解析の際に生成されてしまうカナ 1 文字の語句や「する」「なる」などの一般的な語を不要語辞書を通すことで削除し、名詞、動詞、形容詞を抽出した。未知語は名詞として扱った。それらを時間情報に対応させてデータベースに保存することで、ビデオに対するシーンタグとした。

この処理はすべて機械処理によって行うことができ、またアノテーションデータは人間の自然なコミュニケーション活動から得られるものであるため、シーンタグ作成のために個人が負担するコストは無視できるレベルである。

この処理によって、27 個のビデオコンテンツに対して合計 4136 個、平均で約 153 個、最多で 516 個、最少で 12 個のシーンタグが作成された。

2.3 タグ選択システムを用いた抽出

Synvie によって得られるアノテーションテキストにはビデオに関連のない情報も含まれるため、2.2 で作成されたシーンタグにはノイズが含まれている可能性が非常に高い。

そこで、できるだけ低コストで質の高いシーンタグを作成するために、2.2 で作成されたシーンタグがそのシーンに対するタグとして適切であるかを人手によって Web 上で選択することのできるシステムを開発した（図 1）。具体的には、ビデオの視聴中に 2.2 で作成されたタグが付与されたタイムポイントにくるとビデオが一時停止し、タグが適切であるかを Web ブラウザ上でユーザが選択するシステムである。ビデオウインドウが 2 つ用意されており、タグ選択時には、左のウインドウにシーンの先頭のサムネイルが表示され、右のウインドウでは、シーンの再生を自由に行うことができる。このシステムを利用して被験者実験を行うことで、シーンタグの選別を行った。各ビデオコンテンツに対して 2 人以上の被験者によって実験を行った。

このタグ選択システムを利用したシーンタグ作成のためのコストを、タグの選択に各個人が費やした時

Video Scene Retrieval Based on Online Video Annotation

† MASUDA,Tomoki(masuda@nagao.nuie.nagoya-u.ac.jp)
†† YAMAMOTO, Daisuke(yamamoto@nagao.nuie.nagoya-u.ac.jp)
‡ OHIRA, Shigeki(ohira@nagoya-u.ac.jp)
‡‡ NAGAO, Katashi(nagao@nuie.nagoya-u.ac.jp)
Dept. of Information Engineering, Nagoya University (†)
Graduate School of Information Science, Nagoya University (††)
EcoTopia Science Institute, Nagoya University (‡)
Center for Information Media Studies,Nagoya University (‡‡)
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

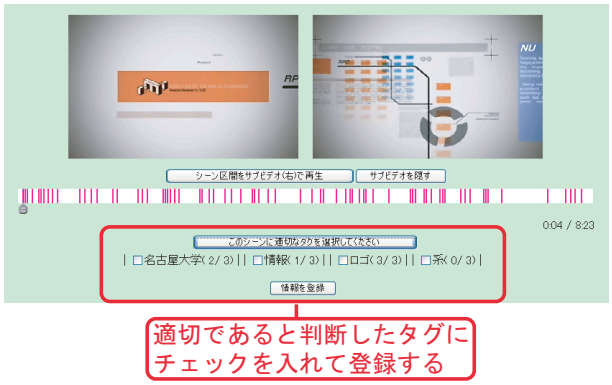

図 1: タグ選択システムの画面例

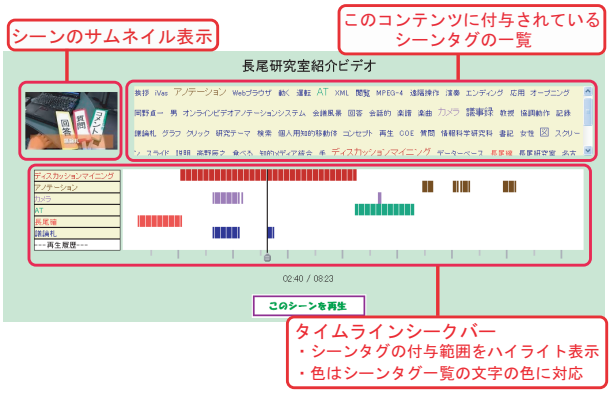

図 2: 検索結果ページの画面例

間とした。その時間は、1 コンテンツあたり平均で約 314 秒で、2.1 でタグ作成に費やした平均時間の約 5 分の 1 であった。また、最長で 1121 秒、最短で 33 秒であった。

この実験によって、27 個のビデオコンテンツに対して合計 1493 個、平均で約 55 個、最多で 277 個、最少で 7 個のシーンタグが作成された。2.2 で自動処理によって作成されたシーンタグのうち 36.2 %が人間の目によってそのシーンを検索するためのタグとして適切だと判断された。

3 ビデオシーン検索

本研究では、タグを利用した新しい発想のビデオシーン検索システムを開発した。そして、そのシーン検索システムに前章の各手法で作成された 3 種類のシーンタグを利用した Web ページをそれぞれ作成し、被験者実験を行った。

3.1 タグを利用したビデオシーン検索システム

この検索システムでは、検索クエリに応じてビデオコンテンツがランキング付けされて表示され、同時に、ヒットしたコンテンツに対して、シーン情報を表示するためのタイムラインシークバーや、付与されているすべてのシーンタグなどが表示される（図 2）。

タイムラインシークバーには、検索クエリとして利用されたタグが付与されているタイムポイントがハイライト表示される。また、シークバーを動かすことでそのタイムポイントのサムネイル画像を見ることができる。さらに、シークバーに対応したビデオの任意の時間からのビデオの再生が可能であるので、シークバーやサムネイル情報をもとに、任意のビデオシーンの視聴ができる。

3.2 シーン検索実験

検索対象となる 9 シーンを設定し、3 種類の手法で作成されたタグを利用して各シーンを手法ごとに 3 人づつ計 9 人が検索を行った。シーンの出題例は「ある動物が親子で映っているシーン」など、必ずしもシーンタグとして付与されている語句が出題文に含まれるのではなく、付与されているシーンタグをヒントにしてシーンを推定できるような出題文とした。また、出題に対する答えが唯一の時間区間であることを明確にするために、文章だけでなくそのシーンのサムネイルをぼかした画像も提示した。

被験者は各出題に対する解答シーンの検索を行い、解答までに費やした時間を自動的に計測した。今回の実験では、全被験者がすべての出題に対して正しいシーンを発見できたため、その観点ではタグ作成の各手法を比較することはできなかった。そこで、シーン検索に費やした時間を比較したのが表 1 である。平均すると 2.1、2.3、2.2 の順で、短時間でのシーン検索が可能であった。2.3 で作成したタグを利用した場合が最も短時間であった出題も存在した。

これらの実験から、オンラインアノテーションを基にして 2.3 のタグ選択システムを利用することで、有用なシーンタグが作成され、人手で詳細に付与したタグを利用した場合と同程度のシーン検索のパフォーマンスが得られると予想される。一方、タグ作成にかかるコストには大きく差があると思われる。したがって、コストパフォーマンスに関して 2.3 の手法が最も優れていると考えられる。

表 1: 各シーンタグと平均検索時間

タグ作成手法	平均検索時間 (秒)
2.1	118.1
2.2	169.6
2.3	145.4

4 おわりに

本研究では、オンラインアノテーションを利用することで低コストでシーン検索に有用なシーンタグを作成することが可能であることを実証し、オンラインアノテーションの有用性を示すことができた。

今後の課題としては、シーンタグ作成法の改良や、タグの付与されていないシーンの検索法の実現、オンラインビデオアノテーションをより大量に獲得するためのシステムの開発などが挙げられる。

参考文献

[1] 山本大介, 清水敏之, 大平茂輝, 長尾確, “Synvie:ブログの仕組みを利用したマルチメディアコンテンツ配信システム”, 情報処理学会第 58 回グループウェアとネットワーク研究会,p13-18,2006.

[2] Synvie Public Beta Service, http://video.nagao.nuie.nagoya-u.ac.jp/