情報工学コース卒業研究報告

映像と論文の部分引用関係に基づく映像シーン検索の高度化に関する研究

2011年2月

棚瀬　達央

目 次

第1章　はじめに

近年のインターネット技術の発達やブロードバンド回線の普及によって、Web上に膨大な数の映像のコンテンツが存在するようになった。またYouTubeなどの動画共有サービスによって、手軽にWeb上に映像コンテンツを公開、閲覧することができるようになり、映像コンテンツはより身近なものとなった。映像コンテンツの中でも学術的な内容を含む映像コンテンツ、例えば、研究成果のデモ映像や、講演風景の映像などは、研究活動などにおいて利用価値が高く、そのような映像コンテンツの増加に対して、視聴者の閲覧する時間は限られているため、シーン検索や要約のような高度な応用に関する技術が求められている。

シーン検索や要約を実現するためには映像の内容に関する詳細なメタ情報が必要不可欠である。それらのメタ情報のことを映像アノテーションと呼んでいる[1]。アノテーションは映像コンテンツに限らず、テキスト、画像、音声、音楽などの多くのコンテンツに対しても適用できる技術であり、これまでにも多様なコンテンツに対するアノテーションの研究が行われている[5]。

これまでの研究において、映像アノテーションを画像認識や音声認識技術を利用することで抽出する自動アノテーション手法[6]や、自動アノテーションで得た結果に基づき、専任の人間が専用ツールを用いて、手動で高品質な映像アノテーションを作成する半自動アノテーション手法[8][9]が提案されている。

自動アノテーション手法の利点は、人手を掛けることなく機械的に映像アノテーションを作成できることにある。しかし、この手法で得られる情報は、音声やテロップ情報、カット検出やオプティカルフローなどの情報であり、映像に含まれる意味的な情報を抽出することは困難である。また、Web上の映像コンテンツには、アマチュアによって作成されたコンテンツが含まれている。そのような映像には手ぶれ、ピンボケなどのノイズが多く含まれるため[11]、機械による処理では、作成されるアノテーションの質が低下するという問題がある。

半自動アノテーションでは機械処理の認識ミスを手動で修正し、詳細なアノテーションを作成することができるが、アノテーションを作成するための、人的コストが非常に高く、膨大な映像コンテンツの数に対して適用するには費用対効果が見合わないという問題点がある。また、アノテーションを作成する人によってはアノテーションの質にばらつきがあるという問題点もある。

筆者が所属する研究室では、Web上で行われている自然なコミュニティ活動に着目し、そこから得られる情報を映像アノテーションとして利用するオンラインア

ノテーション手法が考案された。Web 上のコミュニティ活動とは、ブログや SNS、電子掲示板などのコミュニケーションサービスにおけるコミュニティ活動のことをいう。これらのサービスの利用者は現在も増加傾向にあり、サービスを利用する一般人が発信する情報には、非常に意味のある情報が含まれる場合もあり、時には社会に対して大きな影響を与えることもある。最近では、世界最大の SNS である Facebook や短いテキストメッセージである「つぶやき」のやり取りを行う Twitter などに書き込まれた情報が、エジプトにおいて市民の反政府行動の組織化や広報に大きな影響を与えたことが例として挙げられる。この Web 上のコミュニティ活動において集められるアノテーションは不特定多数の人間による集団の力が大きく影響するため、各個人に対する人的コストが極めて小さいと考えることができる利点や、多種多様なコンテンツに対して、コンテンツの種類に依存しない統一的な仕組みを提供する事により汎用性の高いアノテーションが取得できる利点がある。

このようなコミュニケーション活動が、映像コンテンツに対しても行われるようになり、それらの活動から得られる情報から効率よく利用価値のある情報の抽出を行うという研究が筆者らの研究室で進められた。その結果、開発されたのがオンラインビデオアノテーションシステム Synvie[2] である。Synvie では投稿された映像の任意のタイムコードに対するコメントの投稿、任意のシーンのブログへの引用から映像アノテーションを収集している。増田らが開発した映像シーン検索システム Divie[3] では、実際に Synvie で収集されたアノテーションに基づいた映像シーン検索を実現している。Divie では、収集されたコメントやブログの文章から、タグと呼ばれるマルチメディアコンテンツの検索や分類のために付与されるメタ情報を抽出する。多くの映像共有サービスでも、映像コンテンツに対してタグが付与されるが、それらは映像全体に対して付与されるものであった。Divie では映像シーン単位でタグが付与されるため、映像内部の任意のシーンを検索することができる。また、検索結果として、シーン区間を表す時間軸シークバー、シーンに関係付けられたコメントやブログの文章が表示される。これらの情報によって、シーンタグが存在するシーンだけでなく、その周辺の情報を検索結果上で閲覧することができ、シーンを含む映像コンテンツ全体の俯瞰支援を可能としている。

しかし、Synvie で映像アノテーションとして収集されたコメントやブログの文章から抽出されたタグには、映像シーンと関係のない不適切なタグが含まれやすい。収集されるコメント文やブログ文のテキストには特に制約がないため、シーンとは関係のない内容のテキストが含まれる場合がある。その結果、シーンに関係のない不適切なタグが付与され、Divie の検索結果として検索したタグとは関係のない映像が出てきてしまい検索の精度が低下してしまう問題があった。

本研究では、学術的な内容を含む映像コンテンツに対して、質の高い映像アノテーションを獲得するために、Synvie で利用しているコメントやブログの文章に代えて、論文の文章に着目した。

これまでに様々な論文が学会などで発表されているが、論文を読む際の問題として、文章としての表現方法の難しさや、見慣れない用語の多さから、文章や図だけでは十分に理解できないという問題が挙げられる。特に、読みたい論文の分野に詳しくない者や、論文に読み慣れていない者などにとっては、論文読解は敷居が高いことがある。それゆえに、論文の著者が論文に関連する映像を作成したり、ある研究者が論文を紹介する際に、既に存在するシーンから論文と関係のある映像を引用することが行われている。実際に研究発表では、聴衆に興味を持たせるため、発表者がデモ映像を流すことは盛んに行われている。計算機科学分野の学会である ACM（Association for Computing Machinery）のサイト[1]では論文と一緒にデモ映像が提供されていることもある。

また、研究室のホームページ[2]などでは、その研究室で執筆された論文と共に、研究を知ってもらうきっかけとして、論文の紹介映像を公開している。他にも、研究者のブログ[3]などでは、他者が執筆した論文を紹介するために、YouTube などに投稿された映像を引用して記事を書いていることもある。このようにして、論文を探す際に見つかった論文と関連する映像は、論文の内容を理解する上で非常に有用なものであるため、論文と共に自身の研究資料の一つとして参照されることがあると考えられる。

論文を映像アノテーションとして使用する利点として、コメントやブログの文章をアノテーションとして利用する場合、ユーザによって映像に対して関係のない内容が含まれる可能性があるが、論文の文章は、深い知識を持つ研究者が書いているため、コメントやブログの文章よりも映像と関連する専門的な情報が多く含まれると考えられる。したがって、コメントやブログの文章より論文の文章からは映像シーン検索に有用な特徴的な語が多く抽出されることが期待できる。また、もう 1 つの利点として、コメントやブログの文章は、ユーザによって、主語や目的語が存在しないなど、十分に内容を表していない、文章として正しくない場合があったが、論文は研究者などが時間を掛けて執筆した洗練された文章であるため、文章としての適切さが保障されている。ゆえに、論文の文章からは、信頼性の高い、文章中に現れる語と語の関連性を獲得できることが期待できる。

そこで、本研究では、この論文と映像が共に閲覧されることと、論文の文章の質の高さに着目し、映像シーン検索における検索精度を向上させる仕組みとして、論文部分の文章と映像の部分であるシーンが共引用されたことによる関係（映像と論文の部分引用関係）から映像アノテーションを獲得する仕組みと、それに基づく映像シーン検索を実現する仕組みについて提案する。まず、映像と論文の部分引用関係による映像アノテーション獲得のための仕組みについて説明する。

本研究における引用とは、研究活動において、あるコンテンツの内容を参照し、

[1]ACM,Association for Computing Machinery http://www.acm.org/
[2]明治大学 宮下研究室 http://miyashita.com/publications/
[3]気ままに有機化学：論文 http://chemistry4410.seesaa.net/category/2379659-3.html

その内容を基にユーザが文章を執筆することを指す。この引用を行うためのシステムとして、DRIP システム [4] を利用する。これは土田らが開発した大学研究室の研究活動を対象に個人の知識活動を支援するシステムで、個人の過去に行った会議の情報を部分引用し、次の会議までのユーザの様々な研究活動（論文やプログラムの作成など）を専用のブラウザ上で関係付けをすることができる。DRIP システムを継続的に利用し関連付けの情報を蓄積していくことで、ユーザの研究活動が可視化され、現在までの活動の経緯やテーマにおける活動の位置づけを確認したり、保留状態にしている内容を確認することができ、効果的な知識活動を行うことが期待できる。この DRIP システムにおいて、ユーザが読んだ論文の部分と視聴した映像シーンも扱えるようにDRIP システムを拡張し、DRIP システム上で、引用された論文の部分と映像シーンの関係付けを行えるようにした。

次に、映像と論文をそれぞれ部分的に引用するためのシステムについて説明する。まず、論文部分の文章を引用するために、筆者らの研究室で開発された TDAnnotator を利用する。このシステムでは、研究活動の上で参考にする論文をシステム上に登録することで、Web ブラウザ上での論文の閲覧と、論文の任意の部分に対してのアノテーションを付与することができ、このアノテーションをユーザ同士で共有することができる。この論文の任意の箇所にアノテーションの付与が行える機能を利用し、任意の論文部分の文章をDRIP システムで引用できるようにした。映像内の任意のシーンを定義し、引用するために前述の Synvie の仕組みを利用する。Synvie では、映像からブログへ任意のシーンを引用することができる。この機能を利用して、映像内のシーンを定義し、DRIP システムで引用できるようにした。

そして、DRIP システムを用いて、論文部分と映像シーンを共に引用することで、両者の関係付けを行うことにより、映像と論文の部分引用関係を獲得した。実際に、このシステムを用いることにより、収集した論文と映像のデータから、映像と論文の部分引用関係を作成し、映像アノテーションとして論文のテキストを獲得した。そして、この論文のテキストに基づいた映像シーン検索を実現するために、獲得されたアノテーションのテキストから形態素解析を行うことにより、シーンに対するタグを抽出した。

論文の文章からは、ブログの文章やコメントよりも検索に有用な、特徴語のタグが得られやすいことを示すために、被験者に論文の部分が対応付けられた映像シーンに対してコメントしてもらい、コメントから得られるタグと、論文の文章から得られる名詞・複合名詞のタグを比較した。そして、論文から得られる特徴語のタグの傾向と、コメントから得られる特徴語のタグの傾向について分析し、考察した。

最後に、映像と論文の部分引用関係から得られた映像アノテーションである論文の文章に基づいた映像シーン検索を行うための仕組みを実現した。まず、タグの入力の際に、引用された論文部分の文章中に含まれるタグ間の共起頻度の情報

に基づいて、関連タグを表示するなど表示方法を工夫した。次に、引用されたシーンを検索する仕組みに加え、そのシーンを含む映像コンテンツを検索し、映像コンテンツ全体を俯瞰するインタフェースを作成することにより、引用されたシーンに限定されない映像シーンの閲覧手法を実現した。さらに、映像シーンの閲覧の際に、論文の文章や TDAnnotator へのハイパーリンクを表示することにより、映像に関連する論文に容易にアクセスできるようにした。

以下に本論文の構成を示す。第 2 章ではこれまで我々が行ってきた映像シーン検索を実現するための研究について詳細に述べる。第 3 章では、映像と論文の部分引用関係による映像アノテーションと実験について詳細に述べる。第 4 章では部分引用関係により得られた映像アノテーションに基づく映像シーン検索システムについて詳細に述べる。第 5 章では、関連研究について述べる。最後に第 6 章では、まとめと今後の課題について述べる。

第2章　Webコミュニティ活動に基づく映像アノテーションと映像シーン検索

映像シーン検索を実現するためには、映像の意味内容を記述したメタ情報が必要であり、これを映像アノテーションと呼ぶ。検索に必要となるアノテーションは、検索の対象となる映像コンテンツの種類によって異なる。例えば、ニュース映像はアナウンサーやナレーションの音声や、表示されるテロップの文章の情報などが重要となるが、スポーツ映像は実況や解説者の音声情報に加え、写っている選手の状態や位置情報なども非常に重要となる。本研究で検索の対象とする学術的な内容を含む映像、研究成果のデモ映像や講演風景の映像の場合は、映像の解説者の音声や、テロップなどの映像に写っている情報以外にも、写っているシステムや機材の正式な名称や、映像に関連する専門用語などの特徴的な語が重要となる。

最近では、Web上に手軽に映像を投稿することができるようになり、撮影設備の整った環境で撮影されるプロフェッショナルな映像だけでなく、映像の撮影や編集に慣れていない研究者などが作成した学術的な内容の映像も多く存在するようになり、今後もそのような映像コンテンツが増加していくと考えられる。それらの映像コンテンツは、手ぶれやピンボケなどのノイズが含まれることや、テロップなどの文章がまったく存在しないこともあるほか、作成される映像の内容や構造は何も制約がないため、それに対応した映像アノテーションを獲得する方法が必要となる。

また、映像アノテーションを作成する上で重要なこととして、作成に要する人的コストが挙げられる。ある映像コンテンツに対して、十分な時間と人手をかけ、詳細な映像アノテーションを獲得することができれば、高度な検索は比較的容易に実現できるだろう。しかし、近年の映像コンテンツの量を考慮すると、検索を実現するために必要な人的コストについても考えなければならない。

つまり、アノテーションを獲得する際には、コンテンツの種類や、アノテーションに必要となる人的コスト、アノテーションの詳細度など、複数の要素を考慮する必要がある。これらの要素を考慮し、効率よくアノテーションを獲得し、検索などの応用を十分な精度で実現する手法が求められている。これまでにも映像ア

ノテーションを作成するために、様々な手法が研究されてきた。

従来研究によると、アノテーション手法には、自動アノテーション手法 [6][7] と、半自動アノテーション手法 [8][9] がある。自動アノテーションは、音声認識や画像認識によって作成されるアノテーションである。このアノテーション手法はアノテーションの付与を自動で行うことができるため、人的コストは低いが、この手法が有効な映像は、ニュース映像やスポーツ映像などの決まった環境で撮影された映像のほか、ピンボケなどのノイズがない映像に限られる。自動アノテーションの結果に対して、専任の人間が、専用ツールを用いて修正を加えるやり方が、半自動アノテーションである。このアノテーション手法は、人手によって修正・追加されるため、良質なアノテーションを得られる利点があるが、大量の映像コンテンツのデータを扱う際には、費用対効果があまり高くないことや、映像に対して深い知識を持った人でなければ、質の高いアノテーションが得られない可能性がある。

筆者が所属する研究室では、ニュース映像やスポーツ映像などの限られた撮影環境で作成される映像コンテンツだけでなく、近年増加したアマチュアが作成した一般的な映像などにも幅広く対応した映像アノテーションを獲得する手法が考案された。その手法を実現したのが、オンライン映像アノテーションシステム Synvie[2] である。Synvie では、Web コミュニティ活動によって蓄積される情報に基づいた映像アノテーションの作成を行っている。ここで獲得された映像アノテーションに基づいて応用となる映像シーン検索を実現したのが、増田らが開発した Divie[3] である。Divie では、Synvie で獲得された映像アノテーションから、検索に利用するためのキーワードとなるタグを抽出し、映像シーン単位での検索を可能にしている。

本章では、Synvie を用いて、Web コミュニティ活動の情報から得られたアノテーションの特徴と、そのアノテーションに基づいた映像シーン検索システム Divie の特徴について述べる。さらに、Divie の映像シーン検索における問題点と、本研究で解決すべき点について述べ、新たなアノテーション手法を提案する。

2.1　Web コミュニティ活動に基づく映像アノテーションシステム Synvie

2.1.1　Web コミュニティ活動に基づく映像アノテーション

映像シーン検索や映像要約といった応用を実現するための映像アノテーションを収集する目的で、筆者の研究室で開発されたのが映像アノテーションシステム Synvie である [2]。

Synvie では、Web 上のコミュニティ活動から映像アノテーションを収集している。Web 上のコミュニティ活動とは、近年のブログや SNS、電子掲示板などの Web 上のサービスに集まった、共通の話題を持った人々によって行われるコミュニケーション活動全般のことを言う。また、映像を話題にした Web 上のコミュニティ活動とは、ある映像に対して、コメントを投稿したり、映像コンテンツを引用して、ブログや SNS の日記を執筆するなどといったユーザの活動である。この Web 上の活動で収集される映像アノテーションは、不特定多数の人間により収集されるものであるが、それが蓄積されていくことで、利用価値の高いものとして利用できると考えられる。この考え方は、映像コンテンツ以外のものに関して、すでに実用化されている。例として、Web 上で誰でも作成、編集が可能な電子百科事典である Wikipedia[1]が挙げられる。Wikipedia では、日本語だけでも 70 万 (2010 年現在) を超える非常に幅広い言葉についての記事が存在し、比較的信憑性も高い。全インターネットアクセスのランキング[2]では、10 位以内 (2010 年現在) に入っており、その利用者の多さからも信憑性の高さがうかがわれる。このように、不特定多数のユーザの情報が蓄積され、大きな知識の集合となったものを集合知と呼ぶ。また、そのような一人ひとりのユーザのことをユーザが知識を生み出す貢献者という意味で、User As Contributor と表現することもある。 さらに、Web 上のコンテンツを分類する際に、少数の専門家が分類するよりも、不特定多数の一般ユーザが分類する方が分類の精度が高くなりうるという folksonomy（フォークソノミー）という概念も広まっている。

この folksonomy に基いたサービスとして、お気に入りの Web ページを登録する際に、ユーザが自由にページにタグ付けするソーシャルブックマークがある。ソーシャルブックマークでは、不特定多数のユーザのブックマーク情報を共有することで、タグの関連付けの精度を高めている。

このように、Web 上では、多くのマルチメディアコンテンツに対するアノテーションを Web コミュニティ活動から収集するサービスや研究が行われ、実用性が評価されている。そこで、機械的に高度な解析が困難である映像に対してこそ、Web コミュニティ活動からの集合知の力を利用してアノテーションが有効であると考えられる。このような背景から近年、そのような手法を用いた研究が行われてきた [13][18]。その 1 つが前述の映像アノテーションシステム Synvie である。次節で詳細に説明するが、Synvie では、コメント投稿機能やブログへの映像シーン引用により、映像の部分に対する映像アノテーションを収集している。

この Web コミュニティ活動から得られるアノテーションの、一般的に考えられる特徴について述べる。この手法での大きな利点の 1 つはアノテーション作成のための人的コストが非常に小さいことである。アノテーション作成者は、インターネット上で、自然なコミュニティ活動という形でアノテーションを映像に対して

[1] http://ja.wikipedia.org/wiki/
[2] Alexa Internet http://www.alexa.com/

付与するため、ユーザ自身はアノテーションを施しているという印象を持たない。システムを利用する人数が増えるほどアノテーションは蓄積されるが、ユーザに対する負担が増加することもない。また、映像の種類に関わらず、統一的な仕組みでアノテーションを加えることができるため汎用性の高いアノテーションが得られることも大きな利点である。この手法は人手によりアノテーションが作成されるため、アマチュアが作成したさまざまな種類の映像コンテンツに対しても問題なく適用できる。

もう 1 つの大きな利点は、機械による解析では決して獲得できないアノテーションが得られる可能性がある点である。映像に対するコメントや、映像を引用したブログでは、映像の内容だけでなく、映像に関連した情報が記述されることもある。例えば、ある研究に関する映像がブログに引用された場合、ブログには、映像に出現したシステムについて、写っている内容以外にも、そのシステムの詳細な情報や、他の研究にも利用されていることなど関連する話題が記述されることが期待できる。映像に関連する多様な情報を獲得することで、他の研究映像とその映像を関連付けることもできる。このような情報は、映像の内容に対して深い知識を持つ人でなければ、作成することは困難であるが、不特定多数の人間の中にそのような人物が存在すれば、獲得できる可能性がある。このように、映像に関連する広範囲な情報が得られれば、検索や要約以外の応用にも適用できると考えられる。

しかし、問題点もいくつか存在する。１つはアノテーションの質に関する問題である。この手法で得られるアノテーションは、検索や要約などに適さない場合がある。例えば、コメントやブログの文章では、文章には一切の制約がないため、映像と関係あるものばかりではなく、映像の内容とは間接的にも関係あるとは言えない文章が存在することも考えられる。そのため、この手法では、様々な文章の中から本当に利用価値のあるものを選別する仕組みか、応用の仕方によってアノテーションの利用の仕方を変化させる必要がある。

もう 1 つの問題はアノテーションの量に関する問題である。人的コストを低くして質の高いアノテーションを獲得できたとしても、アノテーションの量自体が少なければ検索や要約などの高度な応用には不十分であることが多い。この手法の、１人１人がアノテーションを付与するという意識を持つことなく、映像に対してコメントの投稿やブログへの引用を行えることは利点だが、アノテーションを意識していないがために、映像に付与されるアノテーションの量には制約がなく、極端に多い場合もあれば、少ない場合もある。１人の人間に長時間の映像にアノテーションを付与させることは非常にコストが高いため、複数人の人間にアノテーションを作成してもらう必要があるが、そのためには、システムの利用者を増やすことが必要不可欠である。多くの人間を集めるには、システムを使うための動機づけが必要となる。そのためには、システム自体の使いやすさ、話題性、面白さなど、アノテーションに関わらない部分も考慮しなければならない。

山本らは、以上の利点と問題点を踏まえ、Synvie を開発し、運用を行ってきた[2]。次節では、Synvie のアノテーションシステムについて詳細に述べる。

2.1.2 映像全体へのアノテーション

現在存在する多くの動画共有サービスでは、映像に対する評価やコメントの投稿、ブログへのコンテンツの引用といった機能が提供されている。また、映像の投稿を行う際にも、コンテンツに対するタグや説明文の入力を要求されることがあるが、これらはコンテンツ全体に対する質の良いアノテーションとして扱うことができる。これらの機能では、手軽に情報を入力できるため、ユーザにとっての負担が少なく、自然に映像に対するアノテーションが収集でき、映像そのものを扱う際に有用な情報になりうる。

しかし、これらの情報をそのまま映像アノテーションとして利用する際、入力された情報は、映像コンテンツ全体に対するアノテーションとしてしか扱うことができない。したがって、映像コンテンツ単位での検索や推薦といった応用を実現することは可能であるが、映像シーン単位での検索や推薦、映像の要約といった高度な応用は実現不可能である。

だが、このような映像全体に対するアノテーションの中にも映像の部分に対する情報も含まれている可能性がある。もし、映像のあるシーンに対して言及したコメントが映像に投稿された場合、記述されたコメントの情報が、映像全体に対しての記述なのか、それともシーンに対してなのか、その場合どのシーンを指しているかといったことを判別することが可能になれば、その情報は映像のシーンを扱う応用に関しても利用できるアノテーションとすることができる。しかし、コンテンツ全体に対するアノテーションを、シーンに対して正しく関連付けることは、非常に困難であり、コンテンツ全体に対するアノテーションに、映像の中で一瞬でしか現れていないような情報が記述されてしまうことがあれば、映像コンテンツ単位で検索する用途に関して言えば、ノイズとなる可能性もある。増田らは、映像全体のアノテーションをその他のアノテーション手法を組み合わせることで、全体に対するアノテーションの情報をシーンと関係付けることを行い、その情報をシーン検索にも利用している。本研究で提案するアノテーション手法による映像シーン検索に関しても、同様にして、その関係付けの情報を利用する。その詳細は次章で述べる。

2.1.3 映像の部分に対するアノテーション

映像コンテンツの高度な応用を実現するためには、映像中の部分に対するアノテーションが必要である。本節では、そのようなアノテーションを収集する Synvie

の 2 つのアノテーション機能について述べ、得られるアノテーションの種類と特徴について述べる。

Synvie では、映像の任意のタイムコードに対するコメントの投稿が行える機能を有している。コメント投稿は Web コミュニティ活動に基づく映像中の部分に対するアノテーションの最も典型的な例である。Synvie では、投稿されたコメントがそのタイムコードに同期してプレイヤー下に表示される。コメントが他ユーザと共有されることで、ユーザ同士のコミュニケーションを促し、それがユーザにとってコメント投稿への動機づけとなる。Synvie のコメント投稿を行うインタフェースを図 2.1 に示す。映像の視聴途中でボタンをクリックすると、左のようにクリックした時刻の映像ショットの画像が表示され、その時刻に関係付けてコメントを投稿することができる。

図 2.1: Synvie のコメント投稿インタフェース

類似した機能を提供しているサービスにニコニコ動画[3]が存在する。ニコニコ動画では映像の任意のタイムコードにコメントすることが可能であり、そのコメントは映像上にオーバーレイ表示される。ニコニコ動画の総動画数は 500 万[4]以上 (2010 年 4 月) に達し、大量のコメントが毎日のように投稿されている。このことは、映像コンテンツに対して共通の興味を持つユーザが、映像に対してコミュニケーションをするという要求を十分に持っていることを実証しているといえる。

しかし、このようなアノテーションの収集方法には問題点が 2 つ存在する。ま

[3]http://www.nicovideo.jp/
[4]http://www.nicovideo.jp/video_top/

ず、前節でも述べたようにアノテーションの質が問題としてあげられる。コメントの投稿は映像の視聴者が感じたことを入力されることで行われるが、投稿されるコメントには特に制約がないため、「かっこいい」「すごい」などの印象語が多く含まれ、映像の内容について記述されることはあまり多くないと予想される。実際この傾向は他の動画サービスでも表れている。また、主語や目的語がないなど、不完全な文章であったり、映像の内容とは全く関係の無い文章が投稿される可能性もある。アノテーション手法として手軽ではあるが、その手軽さのため、多くのノイズを含んでしまうのはアノテーション手法として有効とは言えない。このコメントで収集されるアノテーションをシーン検索や要約などの応用を利用するためには、アノテーションとして適切な物を選別するための手法が必要である。

コメントの投稿によるアノテーション手法に対する 2 つ目の問題点は、アノテーションの対象が任意のタイムコードであり、時間区間を持っていないため、シーンに対して、アノテーションを行っていることにはならない点である。そのため、映像シーン単位で検索したい場合、手掛かりとしてタイムコードの情報を利用することはできるが、シーンそのものは検索できない。また、タイムコードに投稿されたコメント間の関係などを抽出することが困難であるため、統計的に複数のコメントを扱い、高度な応用に用いることも難しい。この問題に対する対処を行っている研究に SceneNavi[13] があり、サービスの一般公開も行っている。SceneNavi では映像アーカイビングシステム SceneCabinet[14] を用いることで、映像をショット切換、テロップ、カメラワーク、音楽区間、音声区間などから、機械的にシーンの分割が予め行われる。その区切られたシーンに対して、コメントを投稿することでコミュニケーションが行えるが、Web 上の多種多様なコンテンツに適用する場合には機械処理によるシーンの分割の信頼性に疑問が残る。意味のないシーンに区切られたシーンに対してアノテーションが付与されたとしても、それは有用な情報とは言えない。このような問題を解決するために、シーンの分割をシステムのユーザに行ってもらう手法が、映像シーン引用である。Synvie では、映像シーン引用に基づきアノテーションを付与する機能が提供されている。

Synvie では、映像の任意のシーンをブログに引用する機能を有しており、引用に基づくアノテーションを収集している。映像シーンに対する引用を定義すると、他者もしくは自分の著作物である映像コンテンツの一部であるシーンを、自分のブログなどの Web ページで参照し、そのシーンに対する紹介や論評を記述することである。これにより、引用されたシーンの時間区間とそのシーンに対して記述されたテキストを映像アノテーションとして収集することができる。

この映像シーン引用のアノテーションの利点は前述した通り、アノテーションを収集することができる点だけでなく、セグメンテーションとしてのシーンの分割もアノテーションとして収集することができる点である。映像シーン引用では、シーンというユーザが定義したセグメントと、そのセグメントに対する情報を同時に収集することが可能である。これにより、映像をシーン単位で扱うことが可

能となり、映像シーン検索や要約などの応用の実現可能性が出てくる。また、定義されたシーンは、引用の際に、人手により定義されるので、何らかの意味を持ったまとまりである可能性が高いと考えられる。このようなシーンには、機械的な処理では生成できないシーンも存在すると考えられる。

また、そのような利点に加え、アノテーションの質という観点でも利点が考えられる。映像シーンの引用はユーザ自身のコンテンツであるブログを執筆するという活動であるため、コメントのような他人のコンテンツに対する活動に比べて、映像シーンに対する良質な情報記述や意味のあるセグメンテーションがなされることが期待できる。これまでの Synvie で収集されたアノテーションのデータの分析からも、コメントによる投稿よりも映像シーン引用によるアノテーションの方が質の高いテキストが多く含まれることが実証されている。

このようにして、コメントの投稿に加え、映像シーン引用から、良質な情報を収集することができる仕組みが実装されており、アノテーションの収集と分析を行ってきた。増田らは、この得られたアノテーションから映像シーン検索を実現するシステムを開発した。次節ではこの開発された映像シーン検索システムについて詳細に説明する。

2.2　映像シーン検索システム Divie

2.2.1　シーンタグの作成

映像検索を実現するためには、映像に対するキーワードであるタグを付与する必要がある。タグを利用した検索は現在でも様々な動画共有サービスにも用いられている。しかし、これまでのサービスで付与されるタグは映像全体に対して付与されるものがほとんどであった。Synvie では、前述した通り、映像シーンに対してアノテーションを収集することを実現した。そこで、増田らが開発した映像シーン検索システム Divie では、Synvie で収集された映像の任意のタイムコード及び、任意のシーンに関連付けられた文章から映像の部分に対するタグを抽出し、映像シーン検索を実現している。Divie における収集されたアノテーションであるコメントやブログの文章からシーンのタグを作成するまでの具体的な処理手順は以下の通りである。

1. テキスト情報の形態素解析を行う。
2. 事前に作成した不要語辞書を用いることで、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除去する。
3. 固有名詞、名詞、動詞、形容詞、未知語を抽出する。

4. 時間情報に対応させてデータベースに保存する。

これらの処理はすべて機械処理によって一括で行われている。この処理で生成されたシーンタグに基づいて、Divie では映像シーン検索を実現している。Divie では、機械処理によりに得られたシーンタグの特徴を考慮し、それらを有効に活かしたシーン検索システムを実現している。

2.2.2 Divie における映像シーン検索

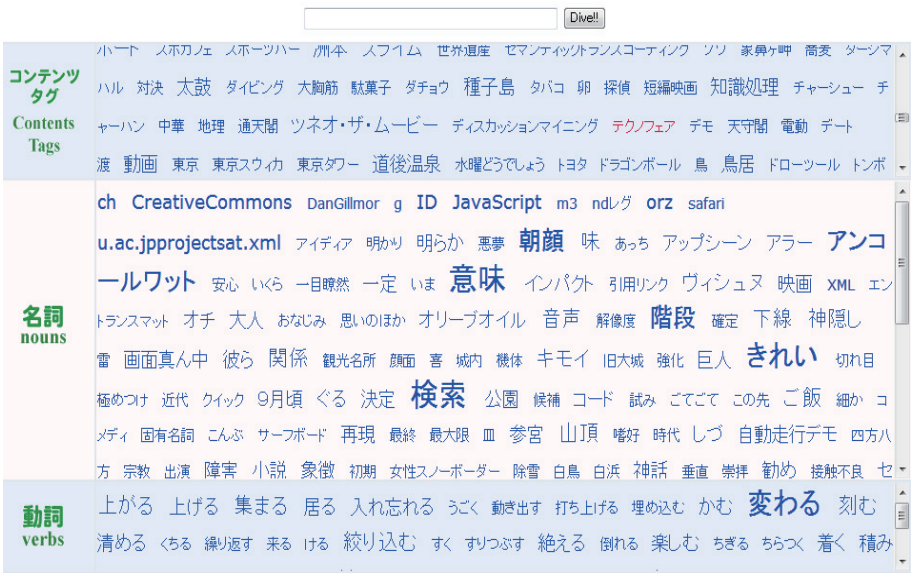

図 2.2: Divie の検索インタフェース

図 2.2 に、Divie のシーン検索インタフェースを示す。本インタフェースには、コンテンツタグ、シーンタグのすべてが表示されている。コンテンツタグとは、映像コンテンツ全体に対するタグであり、コンテンツ投稿時に投稿者が直接入力するタグと、タイトル、サブタイトル、投稿時のコメントを形態素解析して抽出される名詞タグ（未知語を含む）である。また、シーンタグは品詞に分けられている。それぞれのシーンタグをクリックすることで、テキスト入力することなくタグを利用することができる。さらに、最近検索に利用されたタグは色や大きさを変えることによって、検索に利用されやすいタグがどれなのかをユーザに示している。

Divie の検索クエリに対する検索結果は図 2.3 のように表示される。Divie の検索システムでは、検索クエリに基づき、コンテンツのランク付けを行い、その結果を表示する。検索クエリにマッチしたコンテンツに対しては、シーンのサムネイル画像と時間軸シークバー、シーンタグの一覧が同時に表示される。シークバーを

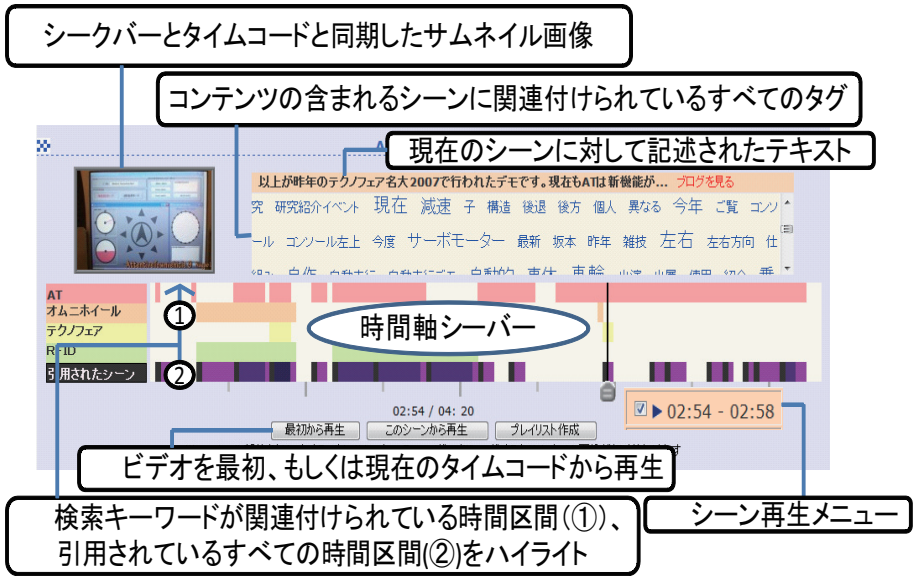

図 2.3: Divie の検索結果表示

動かすと、そのタイムコードに対応したサムネイル画像が左上の画像に表示される。また、中央部分に表示される時間軸シークバーには、検索クエリとして用いられたシーンタグが付与されている箇所をハイライト表示される。シーンタグ一覧に存在するタグをクリックすると、そのタグが検索クエリに追加され、そのコンテンツのみを対象とした再検索が行われる。この操作により、映像コンテンツに付与されたタグについてそれぞれ関連付けられた時間区間が可視化され、映像に直接アクセスすることなく、コンテンツ全体を俯瞰することができる。一般的に、映像シーン検索という言葉からは、検索キーワードに対して適切なシーンをピンポイントに検索結果として表示する仕組みだと認識され、そのような仕組みが理想的であるが、それを高精度で実現するためには、すべての映像コンテンツに対して網羅的かつ信頼性の高いアノテーションが必要となる。そのようなアノテーションを Web コミュニティ活動から収集することは非現実的であり、また人手によって作成することも費用対効果が見合わない。そのため、映像アノテーションの網羅性や信頼性の欠如を補い、効率的な映像シーン検索をするために、Divie はこのような検索の仕組みになっている。この仕組みには、検索の過程でシーンの前後文脈や全体におけるシーンの位置づけなども確認できる利点があり、シーンをダイレクトに視聴するのではなく周辺の情報を確認しながらシーンを探すことにより、求めているシーンをより正確に発見できると考えられている。

増田らは、ユーザによる映像シーンのプレイリストへ引用の情報を用いた検索アルゴリズムにより、映像シーン検索の精度を改善している [3]。

2.3 従来の映像シーン検索の問題点

前節までに述べてきたDivieの映像シーン検索にはいくつかの問題点が存在する。その一つは、映像シーンに対して、無関係なタグが付与され、検索の精度が低下してしまうことである。また、一般ユーザによるコメントからでは、学術的な内容の映像の検索の際に有用な特徴的なタグが得られないことも多いだろう。

無関係なタグが付与される例として、自動的な移動が可能な乗り物の研究映像に対して「マンガを読みながらでも勝手に進むから楽だね」「画質が悪い、いいカメラはなかったのか」というコメントがなされ、「マンガ」「カメラ」「画質」という間接的にも映像の内容とは関係しないタグが付与されることがある。

このように、シーンに対して無関係なタグが付与されるということは、映像シーンに対して内容とは関係性の低いコメントが付与されることに起因する。これは、前述したWebコミュニティ活動から得られるアノテーションに見られる特徴の1つであるアノテーションの質の問題である。ブログへの映像シーン引用により得られたアノテーションテキストは、コメントのテキストよりは、質の高い情報が得られることは実証されているが、ブログで記述される内容はコメントと同様に制限はないため、不完全な文章や、シーンと間接的にも関係のない文章が書かれることがある可能性がなくなったわけではない。ユーザによっては、あまり関係のない文章を多く書くユーザも存在する可能性もあり、実際、コメントを行うユーザや、ブログを引用したユーザによって得られる質にはかなりの差が生じていた[15]。

次に、検索の際に有用な特徴的なタグが得られない問題について述べる。特徴的なタグとは、映像に写っている物体の名称や、使用されている機材やアルゴリズム名などの専門用語を含む、映像と関連する特徴的な語のことである。このようなタグは、映像に対して深い知識を持った専門家が付与するアノテーションからでなければ、抽出されにくいと考えられる。Webコミュニティ活動における不特定多数のユーザからアノテーションを収集する場合には、そのような専門家がアノテーションを付与するとは限らず、そのようなアノテーションが含まれていたとしても、知識のない人が付与したアノテーションが他に多く含まれる中から、機械的に、専門家が行ったものであるかどうかを判別することは非常に困難である。

以上で述べたアノテーションの質を考慮し、特徴的なタグを多くシーンに付与するためには、シーンと無関係な文章が含まれにくく、特徴的な語を多く含んだアノテーションを獲得する仕組みが必要となる。

そこで、本研究では、学術的な内容の映像には、関連する論文が存在すると仮定し、映像に関連付ける良質なアノテーションのテキストとして、論文の文章を用いた新しいアノテーションの手法と、それに基づいた映像シーン検索の手法を提案する。

第3章　映像と論文の部分引用関係に基づく映像アノテーション

近年、Web 上で様々な論文が公開されるようになり、Google Scholar[1]や CiNii[2]などの論文検索サイトによって、手軽に論文を手に入れることが可能となった。研究者などの間では、論文は非常に身近なものであり、研究活動（論文やプログラムの作成など）において、過去の論文を読み、自身の研究の参考にすることは日常的に行われている。自身の研究の参考にできる論文を探す際には、研究者は、論文と共にしばしば論文と関連する映像を発見することがある。例えば、計算機科学分野の学会である ACM(Association for Computing Machinery) の Web サイトでは、論文と一緒にその論文の研究に関するデモ映像が投稿されている場合がある。また、研究室のホームページなどでは、その研究室で執筆された論文と共に、研究を知ってもらうきっかけとして、論文の紹介映像を公開している。他にも、研究者のブログなどでは、他者が執筆した論文を紹介するために、YouTube などに投稿された映像を引用して記事を書いていることもある。このようにして、論文を探す際に見つかった論文と関連する映像は、論文の内容を理解する上で非常に有用なものであるため、論文と共に自身の研究資料の一つとして参照されることがあると考えられる。

本研究では、研究活動において、論文と映像が共に研究資料として扱われることに着目し、論文の文章を映像アノテーションとして収集するために、映像と論文の引用が手軽に行えるシステムを開発した。また、詳細な関係付けが行えるように、映像の部分であるシーンを引用するために作成するシステムと、論文の部分の文章を引用するシステムを開発した。以下では、これらの映像と論文の部分同士の共引用を映像と論文の部分引用関係と呼ぶことにする。本章では、それぞれのシステムの詳細と、部分引用関係が生成されるまでの手順について、また、これらの映像と論文の部分引用関係により収集されるアノテーションの特徴について述べる。さらに、これらのシステムが利用されることによって収集される映像アノテーションからは、検索に有用な特徴語が抽出されやすいことを示すために行った、映像に対するコメントと論文から得られた名詞の比較実験について詳細に述べる。

[1] http://scholar.google.co.jp/schhp?
[2] CiNii http://ci.nii.ac.jp/

3.1　知識活動支援システム DRIP

筆者が所属する研究室で開発されたシステムに DRIP[3] システムが存在する [4]。このシステムでは、ユーザはディスカッションレコーダによって作成されたゼミコンテンツの引用や、タグの付与を行うことで、ゼミコンテンツを要約・分類・整理し、そこから導出される、文献調査やシステム構築といった様々なタスク内容の作成を行うことができる。DRIP システムでは、そこで引用したゼミコンテンツやそこから作成されたタスクを専用ブラウザ上に可視化して表示し、それらを閲覧しながら発表資料を作成する機能を有している。これを行うために、DRIP システムのユーザは、コンテンツを引用して発表資料を作成するために、Web サーバ・クライアント型アプリケーションを利用する。このアプリケーションで記録される様々なコンテンツとそれらのリンク情報を表示したインタフェースを図 3.1 に示す。

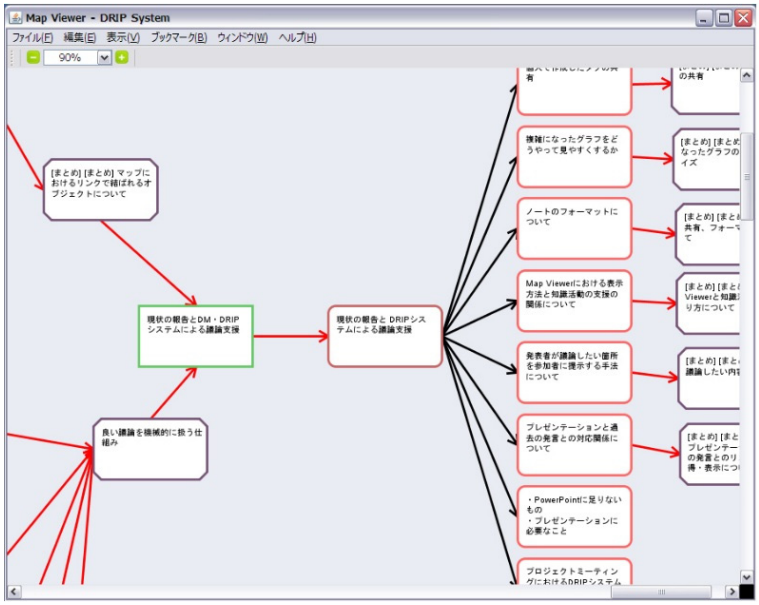

図 3.1: DRIP システムのコンテンツのリンク表示

このインタフェースではコンテンツをノード、コンテンツ間のリンクをエッジとみなしたグラフ構造を表示しており、ユーザは必要に応じてその配置を自由に変更できる。コンテンツにはゼミコンテンツや、そこから生まれた発言、タスクに関するノートがあり、新たに作成したノードを既にあるノードに対して関係付けを行うこともできる。このようなリンク付けの情報により、図 3.1 のようにユーザの研究活動を可視化することで、過去の議論と現在の研究成果の間の文脈情報を俯瞰することができる。

[3] Discussion-Reflection-Investigation-Preparation

本研究ではこのクライアントアプリケーションを拡張し、映像と論文をコンテンツとして引用する仕組みを開発した。具体的には、詳細な研究活動の情報を記録するため、映像と論文をそれぞれ部分的に引用し、それらの関係付けを行う仕組みを、映像や論文の部分を指定する仕組みと連動させることで実現した。映像と論文の部分を指定する仕組みについては、次節で詳細に述べる。

3.2 論文の部分の引用

映像と論文を関連付ける際、論文の文章全体が映像に対して関連性があるとは限らない。論文には構造が存在し、序論、本論、結論に大きく分かれて章立てがなされている。研究の背景に関する文章、研究の具体的なアプローチについての説明、研究の評価実験の文章など、様々な情報が記述されている。これらすべての文章を1つの映像に対応付けることは困難であり、研究活動の際にも、論文の文章すべてを引用して参考にすることはほとんどない。したがって、研究の際に自身の研究に有用な部分だけを引用できるシステムが必要となる。ここで引用される部分は何らかの意味のまとまりを持っていると考えられ、それに対応する映像あるいはそのシーンが存在すれば、引用された論文の文章は、映像アノテーションとして質の高いものとなると考えられる。

本研究では、筆者の研究室で開発された論文アノテーションシステム TDAnnotator を用いて論文の部分を手軽に指定し、論文部分のデータとして扱える仕組みを開発した。このシステムではサービスに投稿された論文の閲覧と、論文の部分に対してアノテーションの付与が行える。

図3.2に論文にアノテーションを行うユーザインタフェースを示す。ユーザは投稿された論文を閲覧している際に、論文の部分を短形選択により決定すると、ポップアップウィンドウ上にアノテーションを記述する画面が表示され、コメントや翻訳などの、アノテーションが行える。ここで論文部分に対して付与されたアノテーションは他者にも公開され、共有することができる。本研究では、論文の部分を研究資料の1つとしてDRIPシステムへの引用を可能にするために、指定した論文の部分のアノテーションをDRIPシステムから参照できる機能を論文アノテーションシステムに追加した。論文の部分を選択した際に表示されるポップアップウィンドウでは論文の部分に付与されたアノテーションの一覧が表示され、引用したいアノテーションを選択することで、それを、DRIPシステム上で引用することが可能になる。ここでは、論文の文章のテキストは論文の部分に対するアノテーションの１つとして扱われ、DRIPシステム上で元の論文コンテンツと論文の部分と共に引用される。ユーザは、論文の部分を自分で設定できるだけでなく、すでに存在する他のユーザが設定した論文の部分を再利用して、その論文の部分に付与されたアノテーションを引用することも可能である。

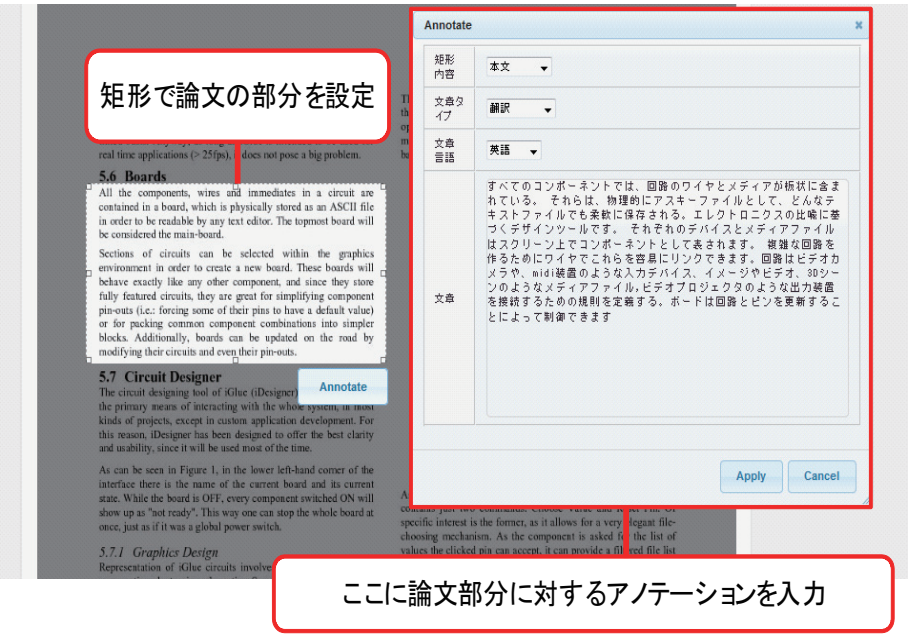

図 3.2: 論文アノテーションシステム TDAnnotator の画面例

3.3 映像の部分の引用

映像の部分とは映像シーンである。研究資料の 1 つとして映像コンテンツを利用する際、映像コンテンツ全体では情報量が多すぎる場合がある。例えば、あるシステムの紹介映像が存在し、システムの詳しいアルゴリズムについてのみ参考にしたい場合、映像全体を引用したという情報しか記録されない場合、映像を見返す際に、アルゴリズムについて説明している映像の箇所をまた探さなければならない。コンテンツの再生時間が長いほどそのような手間が増加するものと考えられる。

そこで、映像コンテンツの任意のシーン区間を、研究活動で利用したコンテンツとして手軽に作成して、引用するために、筆者の所属する研究室で開発された Synvie の映像シーン引用の機能を元に、映像コンテンツから任意のシーン区間を決定する機能を開発した。本節では、本研究で開発した映像シーン引用を支援するシステムのユーザインタフェースと既に存在する映像シーンの再利用の方法について述べる。

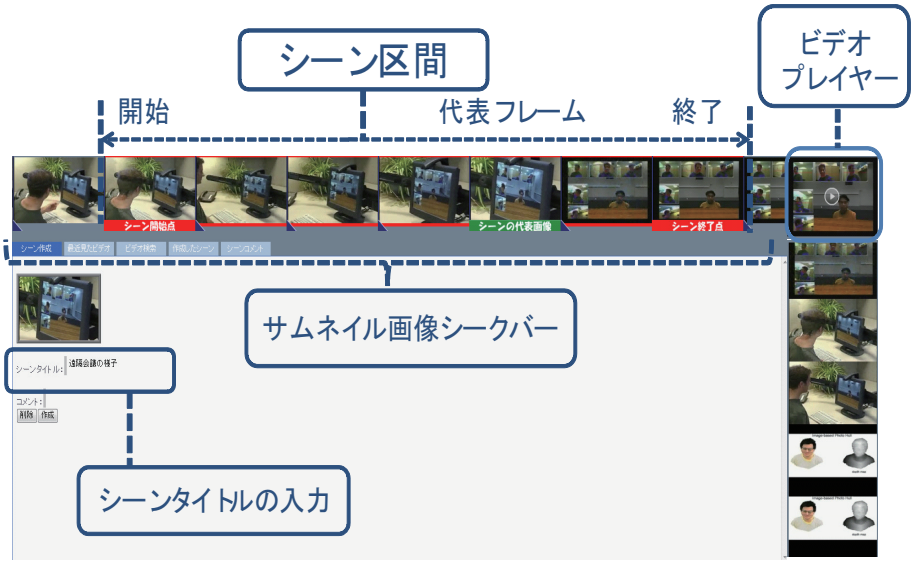

図 3.3: 映像シーンを引用するインタフェース

3.3.1 映像シーンの作成

サムネイルシークバーの生成

図 3.3 に映像シーンを引用するためのユーザインタフェースを示す。まず、引用したいシーンを含む映像コンテンツを図 3.3 のインタフェースに読み込む。コンテンツの読み込みには、ユーザが選択した時期の近いコンテンツから選択する仕組みや、キーワードによって選択する仕組み、さらに、最近作成したシーンを含んだコンテンツから選択する仕組みを提供している。また、映像の視聴中に図 3.3 のインタフェースを開くと自動的にそのコンテンツが読み込まれる。

コンテンツが読み込まれると、右上に映像を再生するためのストリーミングビデオのプレイヤーが設置され、プレイヤーを再生すると上下左右にサムネイル画像が流されていく。このサムネイル画像は、映像コンテンツの登録時にあらかじめ生成され、データベースに保存されている URL の情報から読み込まれる仕組みとなっている。

右から左へ水平に流れるサムネイル画像は 2 秒単位のサムネイル画像であり、上から下に垂直に流れるものは 10 秒から 60 秒までの中でユーザが指定した単位のサムネイル画像であり、どちらもプレイヤーの再生時間と同期して画面上で移動し、マウスドラッグによるシーク操作が可能である。水平に流れるサムネイル画像は作成する映像シーンを決定する目的で利用され、垂直に流れるサムネイル画像は映像の再生時間の長いコンテンツに含まれるシーンを引用したい場合に長い時間単位で映像シーンを飛ばしてシークする目的で利用されることを前提として

図 3.4: 引用するシーン区間の選択

いる。

引用するシーン区間の選択

サムネイル画像の上でマウスドラッグを行うことで、画像がシークされ、それに同期して早送りあるいは巻き戻しされる。ストリーミングビデオを利用してシーンを探す場合、タイムラグなしにシークすることは出来ないため、シーンを細かく参照するためには非常に手間がかかる。逆に細かく参照せずにシーンを調整すると映像内容が見落とされる可能性がある。本システムのようにシーク可能なサムネイル画像を参照する仕組みを提供することによって、効率よく詳細に映像内容を閲覧し、シーンを探すことが可能になる。

シークによって引用したいシーンを発見したら、引用する区間をマウスクリックによって詳細に設定する (図 3.4)。一回目のクリックでシーンの開始フレームを、2 回目のクリックで終了フレームを、3 回目以降のクリックで微調整を行う。選択されているシーンは、サムネイル画像下がハイライト表示されるため、シーン区間を直感的に確認することが可能である。シーンの選択を行ったら、Ctrl キーを押しながらマウスクリックを行うことで、シーンの代表フレームの設定を行う。デフォルトではシーンの開始フレームが代表フレームとなる。

引用するシーン区間と代表フレームが決定したら、右クリックメニューから映像シーンを映像シーン引用エリアに追加する。シーン引用エリアでは、決定したシーンに対して、シーンタイトルを入力する。シーンタイトルを決定すると、決定されたシーンをDRIP システム上で引用することが可能となる。DRIP システムでは映像とその映像から定義された映像シーンの二つの情報が自動的に関連付けられる。

映像シーンが定義されることにより、シーンというセグメント情報が映像コンテンツに関連付けられる。ユーザが研究活動のために何らかの意図である時間区

間を選択したということはそのシーンは何らかのまとまりをもったセグメントであると考えられる。このようにセグメントが作られることによってシーンというセグメントに関するメタ情報の関連付けが可能になり、さらに、シーン単位での応用システムの実現が容易になる。

また、引用されるシーンはユーザの研究資料として利用するために記述されたものであり、ここで入力されるシーンタイトルは、映像シーンに対するアノテーションとして利用価値のあるものと考えられる。

3.3.2 シーンの再利用

図 3.5: シーンの再利用

ユーザは作成したシーン区間の詳細を自身で設定できるだけでなく、すでに作成されているシーンを再利用してそのまま引用することが可能である。再利用可能な区間は図 3.5 のように表示され、クリックすることで、そのシーンのプレビュー再生とシーンに対して記述されたコメントの閲覧が行える。

作成された区間を再利用可能にすることによって、ユーザのシーン作成に対する負担を軽減することができる。さらに、アノテーションを蓄積するという観点から見ても、複数のユーザによって引用されたシーンの同一性を明確にできるという利点がある。まったく同じ対象を複数のユーザが DRIP システムに引用したいと思った場合に、例えば、あるユーザはあるコンテンツの 10 秒から 20 秒のシーン、別のユーザは 12 秒から 22 秒のシーンを選択する微妙なずれが生じる可能性がある。この場合、それぞれの指し先が意味的に同一のシーンかどうかを判別する

図 3.6: シーン一覧

ことができないため、2 つの別のシーンとして扱うことしかできない。この場合、それぞれのシーンに対する情報は、それぞれのシーンにしか関連付けることができず、同一のシーンに関する情報を増やしていくことができなくなってしまう。しかし、シーンを再利用可能にすることによって、他のユーザと同じ意図で、微妙にずれたシーンを選択してしまうことを防ぐことができ、シーンに対する情報量を増やしていくことができる。

また、図 3.5 に示すように、すでに作成されているシーンを一覧で表示し、DRIP システムで参照可能にすることができる。一覧にはシーンタイトル、映像シーンを含む元の映像コンテンツのタイトルを表示し、映像シーンの再生も行える。シーンを一覧で表示することで、シーンタイトルやコンテンツタイトルを手掛かりとして、過去に作成されたシーンから、ユーザの研究に関連する映像を探すことができ、シーンを新たに作成することなく、映像シーンの引用が行える。

3.4 映像と論文の関係付け

第 3.2 節と第 3.3 節で述べた、映像と論文の部分を引用するそれぞれのシステムを用いて、ユーザは、DRIP システムで参照可能にした映像と論文の部分を研究資料として利用することができる。そして、このアプリケーション上で映像シーンと関連する論文の部分の関係付けを行う。図 3.7 に DRIP システムで関係付けが行われた論文の部分と映像シーンのノードを示す。映像コンテンツのノードから伸

びるエッジの先は映像の部分要素を表す映像シーンである。また、映像コンテンツのノードの上にあるノードが、論文アノテーションシステムに登録されている論文コンテンツであり、そこから伸びているノードが論文の部分である。論文の部分のノードが指しているノードは論文の部分に付与されたアノテーションである。ここで表示されているアノテーションは論文の部分に記述されているテキストである。このアノテーションのテキストと映像シーンを関係付けると、図3.7で示す楕円形の関連を表すノードが作成される。この関連ノードの情報は、ユーザがDRIPシステムの情報を保存した際に、サーバに送信され、データベースに保存される。本研究では、データベースに保存されたDRIPシステムのユーザが作成した関連ノードの情報を基に、映像シーンのアノテーションとして論文の部分に対するテキストを獲得する。

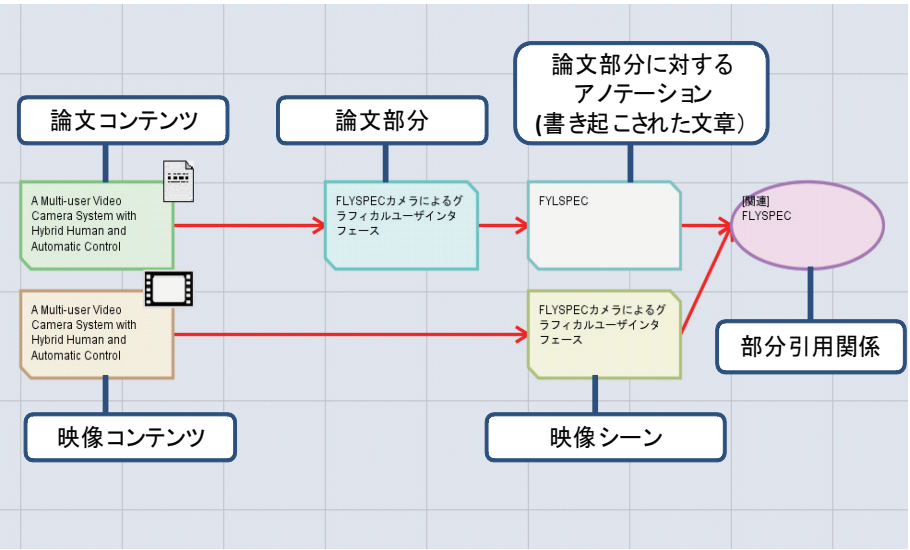

図 3.7: 映像と論文の関係付け

3.5 論文の部分引用関係に基づく映像アノテーション

本節では、前節までに述べてきた研究活動により作成される映像と論文の部分引用関係に基づいてアノテーションを収集する手法に関して、一般的に考えられる特徴を述べる。

この手法の大きな利点は、論文の文章という、その論文の研究分野に関して専門的な知識を持った研究者が記述したテキストを用いる点である。つまり、このアノテーションで映像に関連付けられたテキストには、専門家がオフラインで詳細にアノテーションを行う際に得られるような専門用語などの特徴語が多く含まれ

る可能性がある。学術的な内容を含む映像シーンを検索する際には、このような特徴語は非常に有用なタグになると考えられる。また、このアノテーションで得られる特徴語は、Web コミュニティ活動に基づいて収集するアノテーション手法で記録・蓄積されたアノテーションからは抽出されにくいような特徴語を含んでいることも期待できる。本研究で収集するアノテーションからは Web コミュニティ活動に基づいたアノテーションよりも、特徴語が多く存在すること、また、そのような特徴語は Web コミュニティ活動に基づくアノテーションで集まる特徴語の質と異なるということを実証するために、映像に対するコメントを収集し、論文の文章から得られる特徴語とコメントから得られる特徴語を比較する実験を行った。その実験の詳細は次節で述べる。

もう 1 つの利点としては、Web コミュニティ活動で集めるアノテーション手法と同様に、アノテーションを行う各個人は、アノテーションを作成しているという意識が特にないという点である。そのため、オフラインで専門家が詳細に映像に対してアノテーションを行う際に、専門家個人が負担するコストよりも、ユーザの負担が少ないと考えられる。また、システムを利用する人が増えれば増えるほど、アノテーションの量は増えるが、各個人が負担するコストは増加しない。

さらに、文章の文法的的確さが挙げられる。Web コミュニティ活動に基づくアノテーションには、入力されるテキストには特に制約がないため、文法的にあまり的確でない文章、つまり、主語や目的語がない文章が記述されることや、感動詞など、単体では、意味をなさないものが含まれる。しかし、論文は査読などが行われていて、ある程度洗練された文章であるから、文章として文法的にあまり的確でない文章はほぼないものと考えられる。そのような文章からは、容易に語と語の関連性を統計的に獲得できる可能性がある。本研究では、この考えに基づき、アノテーションとして収集されたテキストの文章に現れる語の共起頻度からタグの関連性を計算し、それを利用した映像シーン検索を実現した。この詳細については次章で述べる。

このアノテーション手法にはアノテーションの量に関する問題がある。このアノテーション手法により、負担をかけずに、質の高いアノテーションが収集できると考えられるが、映像に論文が関連付けられなければ、シーン検索ができない。また、映像全体に対して、網羅的にアノテーションが付与されるとは限らず、関係付けられる論文の数も限られている。これらの問題を解決するためには、システム利用者を増やすことが必要不可欠である。多くの人間にこのシステムを利用してもらうためには、システム自体の使いやすさが必要である。また、論文と映像をより容易に引用するためのユーザインタフェースや、映像と論文を関係付ける動機づけなど考慮すべき点は多い。

3.6 実験

本研究における映像と論文の部分引用関係に基づくアノテーションから得られる特徴語の量と質について検証するために、視聴する映像に対して専門的な知識を持たない被験者からコメントを収集し、そこに含まれる特徴語と、映像に関連付いた論文の部分の文章から得られる特徴語を比較して評価する実験を行った。実験に使用するデータセットの作成のために、まず、すでに Web で公開されている学術的な内容を含む映像と、その映像に関連する論文を 50 セット収集した。次に、論文を TDAnnotator に登録し、対応する映像を視聴しながら、映像の内容と関連する部分が記述されている論文部分を特定し、DRIP システムで参照できるようにした。関連付けられる論文の文章の量は長くとも 1 パラグラフとした。さらに、シーン引用インタフェースに映像を読み込み、論文部分に対応する映像シーンの区間を決定し、DRIP システムで参照できるようにし、DRIP システムでそれらを関連付け、映像シーンと関連する論文の文章をデータセットとして獲得した。収集した映像の内容は、すべて学術的な内容を含む映像で、システムのデモ映像や、研究紹介の映像である。

3.6.1 実験内容

本実験は、インターネットサービスの利用に慣れており、YouTube や Synvie などの動画共有サービスの閲覧を行ったことがある本研究室の学生 10 名を被験者として行った。

まず、提案手法により作成された映像シーンの中から、被験者にコメントしてもらう 22 シーンを選別した。ここでは、シーンの区間が短すぎてシーン単体で見た場合、シーンの内容が理解できない映像シーンは除外した。

被験者には、選別された映像シーンのサムネイルを 5 枚と、映像に関連付いた論文の概要を見てもらい、コメントできそうなシーンを選択してもらった。コメントできるかできないかの判断には、ある程度の数の名詞を抽出できるようにするために、自然に 50 字以上の文を記述できるかを基準にした。上限を設定していないのは、コメントの文字数が増えたからといって、特徴語が増えるとは限らないためである。また、選択する映像シーンの上限と下限は設けていない。

このときに 2 人以上のコメントが集まるシーンのみを実験対象とし、合計 16 シーンに対してコメント収集の実験を行った。被験者は、映像全体を視聴した後に、その映像コンテンツに存在するそれぞれの被験者が選んだ実験対象のシーンにコメントを投稿した。そして、収集されたコメントと論文の部分の文章に存在する名詞・複合名詞が特徴語かどうかの判別を人手により行った。特徴語であるかの基準は、システム名や、人名などの固有名詞のほかに、少なくとも映像中の文脈として特別な意味として用いられているもの、特徴のあるものを選んだ。たとえば、

「ペン」という単語は一般名詞であるが、システムで用いる特別な装置の意味で映像の中で使用されていたため特徴語扱いしている。逆に、「ユーザ」という単語はシステムを利用する人という意味で専門用語として捉えられる場合もあるが、映像シーンを表現する際に、特徴のあるものではないため、特徴語として扱っていない。また、ここで判定される特徴語は、映像と直接関係しているかどうかは考慮されていない。

3.6.2 実験結果

被験者実験を通して、計10人の人間によって、92個のコメントが作成され、論文の文章とコメントを形態素解析した結果から名詞・複合名詞を抽出し、それらから人手により特徴語を選別した。形態素解析にはCabocha[4]を用いた。

論文の文章とコメントから映像シーン毎に抽出される名詞の異なり数と、そこに含まれる特徴語の異なり数を調べた。コメントの場合は、同じシーンに投稿されたコメントすべてを1つのアノテーションとして集計し、1つのシーンに対する複数人の被験者のコメントに含まれる同じ名詞は異なり数1として計算した。抽出されたすべての名詞・複合名詞の異なり数とそこに含まれる特徴語の数、そこから計算される割合は表3.1のようになった。参考として、シーンに関係付けられた論文の文章を読んだ後に、映像に対してコメントした被験者のデータも記載する。表3.2に被験者ごとのシーン当たりの特徴語の数と、それぞれの平均コメント長を示す。

表 3.1: 得られたタグ数と、それに占める特徴語の割合

	特徴語異なり数	全異なり名詞数	特徴語の割合 (%)
論文	158	572	27.6
コメント	107	418	25.5
論文を読んだコメント	156	457	34.1

3.6.3 考察

本実験において、特徴語とは、映像シーンの内容を記述する上で、特徴的な名詞・複合名詞を指し、抽出される単語のうち、特徴語の数が多いほど、抽出する前の文章の中には多くの特徴的な情報が含まれていることになる。

[4]http://chasen.org/~taku/software/cabocha/

表 3.2: シーン当たりの被験者ごとの特徴語の数

	何も読んでいない被験者のコメント	関連する論文を読んだ被験者のコメント	平均コメント長
被験者 1	4.0	4.9	99.0
被験者 2	1.2	3.4	78.8
被験者 3	3.0	4.5	76.2
被験者 4	2.0	2.0	76.2
被験者 5	2.4	3.8	62.8
被験者 6	4.0	4.0	126.7
被験者 7	3.5	5.8	77.7
被験者 8	2.0	4.8	75.3
被験者 9	4.0	3.3	116.5
被験者 10	2.5	5.3	163.2

実験結果から、論文内のシーンに関連付けられた文章から名詞・複合名詞を抽出した場合に、その中に含まれる特徴語の割合は 27.6 %、複数人のコメントの場合は、25.5 %であった。この値は、大きな差があるとは言えないが、論文の文章からは、被験者のコメントよりも文章中に含まれる名詞の中には特徴語が多く含まれることが実証できた。

本実験では、被験者に筆者が所属する研究室の学生を選び、さらに、実験対象には学術的な内容を含む映像のみを用いて、50 字以上という文章量が書ける興味のある映像シーンを被験者に選択してもらっている。これはある程度被験者に、背景知識を持っている映像を選ばせていることになる。もし、学術的な内容を含む映像を Web に公開し、コメントを収集した場合には、映像に対して全く知識のない人間が行うようなコメントも収集されることになる。そのようなコメントには、ただ単に、「すごいなぁ」、「これはどういう仕組みなんだろう」などといった、まったく特徴語が現れない文章が増加することが考えられる。そのため、実際に一般的なコメントを多く収集した場合は、コメントの特徴語の割合はもっと低くなると考えられる。

この実験で収集されたコメントと論文の文章から抽出されるそれぞれの特徴語の質について関して考察する。論文の文章にしか現れなかった特徴語、コメントからのみしか得られなかった特徴語、どちらの文章にも現れた特徴語が存在した。どちらの文章にも多く含まれる特徴語としては、映像に実際に写っている、「リモコン」「ペン」「アイコン」などの特徴語が多く含まれた。このような特徴語は一般語であることが多く、コメントの文章にも記述されやすいためと考えられる。

論文の文章にしか含まれなかった特徴語には、「Coliseum」「OpenGL」「モーショ

ンセンサ」「グラフィカルインタフェース」など、システム名、システムに利用した装置や、システムの概念を表す抽象語などが多く見られた。このような名詞は固有名詞が多く、一般の人々にはなじみのない名詞が多い。したがって、そのような単語を使い慣れない人のコメントには、このような特徴語が含まれる可能性は低く、そのような名詞を含むコメントをするのは、映像に関して知識を持つ人物に限られると考えられる。逆に、コメントにしか含まれない名詞には「これはKinectを使えば簡単に実現できるのではないか」「ネットゲームに応用できたら面白そう」などの、映像とは直接的に関係しない主に映像から被験者が連想したようなコメントに含まれる「Kinect」「ネットゲーム」などの特徴語が含まれていた。

このような特徴語は、映像アノテーションとしてノイズになる場合があるため、今回の実験では特徴語という観点のみで評価したが、映像との関連性を考慮した評価の場合は、コメントのほうが関係性の低い特徴語を多く含むと考えられる。

参考として示した、論文を読んだ人によるコメントの特徴語の割合をみると、論文よりも6.5％大きい値となっている。これは、コメントの文章に論文の情報が記述され、かつ、その文章を読んだ上での映像シーンに対する意見が含まれるため、論文よりも特徴語が増えたと考えられる。論文を読んだ被験者は、映像に関する深い知識を得た被験者であり、論文の著者によるコメントと近いものになるため、この値は妥当であると考えられる。一般に、このような知識のあるコメントが多く収集されることは稀であり、また、コメントという制約のない文章には関係性の低い特徴語も含まれる可能性もあるため、論文を読んだ被験者によるコメントは特徴語の量だけを見て、完全に優位であるとは言えない。

次に被験者ごとのコメントの文章に含まれる特徴語の数について考察する。実験結果では、被験者によってシーンあたりに生成される特徴語の数に差が見られた。参考として、論文を読んだ際のコメントと比較すると明らかに論文を読んで知識を身に付けた際には、ほとんどの被験者の特徴語の数が何も読んでいないときのコメントより増加している。このことから、文章中に含まれる特徴語の数は、コメントを行う人の映像の背景にある知識の量に大きく影響すると考えられる。

本実験により、映像シーンに関連付けられた論文の文章からはコメントより特徴語が多く抽出されることと、コメントに含まれる特徴語の数は映像の内容に関する知識量に関係することが実証された。

今回の実験では、コメントを投稿するまでのユーザのコストと、映像と論文の部分を関連付けるまでのコストの比較を行わなかった。アノテーション手法を比較する上で、アノテーションを獲得するまでの人的コストは重要な要素であるが、映像と論文の部分引用関係の作成までにかかる人的コストを定量的に測定することが困難であると考えられるため、実験では比較しなかった。アノテーションの手法としては、コメントによる人的コストは非常に小さいと考えられるが、提案するアノテーション手法も、コメントと同様に、ユーザはアノテーションを作成することを意識していないため、アノテーションそのものを目的とするオフライ

ンアノテーションなどよりは負担が小さいと考えられる。しかし、システムの部分引用関係を獲得するまでのインタフェースの利便性などを考慮するとコメントの投稿よりは負担が大きいと考えられるため、この人的コストを考慮することは今後の課題の一つである。

第4章　映像シーン検索

本章では、論文と映像の部分引用関係によって蓄積される、シーンというセグメント情報、シーンに対して関連付けられたテキスト及び、それから抽出されたタグを活用することによって、効率的に論文と関連する映像シーンを検索する仕組みを提案する。

本システムにおける映像シーン検索全体の流れを図4.1に示す。この検索システムでは、ユーザが、はじめにタグを入力するか、一覧からタグを選択すると、連動して引用された映像シーンの一覧が表示される。また、検索ボックスの横の検索ボタンを押すと、検索クエリのタグが付与されているシーンを含む映像コンテンツの検索を行い、その検索結果のページに遷移する。このページでは、タイムコードと同期して、論文の文章や映像のサムネイルを同時に表示するなど提示方法を工夫し、コンテンツ全体を俯瞰できるようにした。以下では、各機能について詳細に説明する。

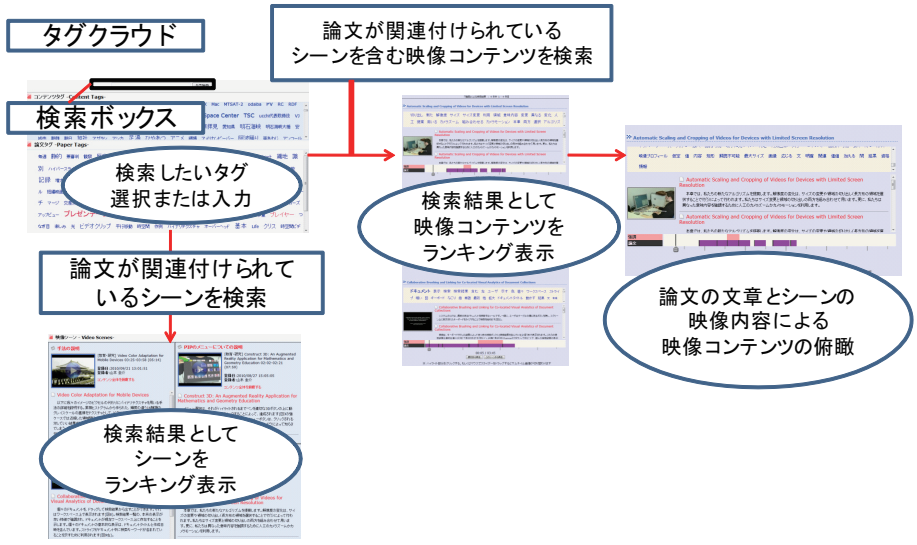

図 4.1: 映像シーン検索の流れ

4.1　タグクラウド

本システムは、論文のテキストから抽出されるシーンタグによるタグクラウドを利用して映像シーン検索を行う。タグクラウドとは、タグが大小さまざまなフォントで表示されるものを指すが、主に 2 つの利点がある。まず、検索に利用可能なタグが提示されることによって、ユーザが探したい映像シーンに対するキーワードや、興味のあるキーワードを発見する手掛かりとなる。さらに、ユーザが、タグの色や大きさの提示方法によって、そのタグの人気度や注目度を推測することができる。

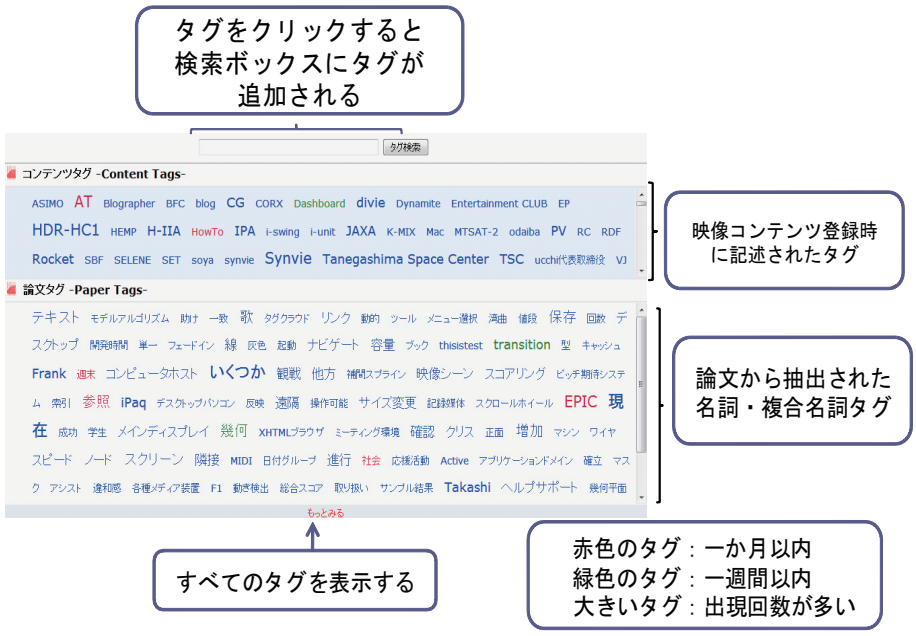

図 4.2: タグクラウド

図 4.2 に本システムのトップページに表示されるタグクラウドを示す。トップページでは、コンテンツ登録時に、登録者によって付与されたコンテンツタグと論文のテキストから生成されたシーンタグが 50 音順にソートされ、タグクラウドとして表示される。シーンタグには、特徴語が多く含まれる名詞・複合名詞のタグを利用した。名詞・複合名詞のシーンタグだけではなく形容詞と動詞といったシーンタグで検索したい場合は、一覧表示ボタンを押すことで選択が可能となる。

それぞれのタグの文字サイズはシーンタグの出現頻度によって決定されている。さらに、アクセス時から一週間以内に検索に利用されたシーンタグは赤色に、一ヶ月以内に利用されたものは緑色でハイライト表示される。これらの仕組みによって、検索に利用されやすいタグほど発見しやすくなり、ユーザにとって有用な情報となる。タグクラウドに存在するタグをクリックすることによって、クエリを

入力する検索ボックスにそのタグが追加される。これにより、ユーザはテキスト入力を行うことなく検索が可能となる。クエリには、複数のタグを利用することも可能である。

4.1.1 関連タグ

本システムでは、映像アノテーションの論文の文章における語の共起頻度からタグ同士の関連度を計算した。語の関連度を測る計算指標として、文章中の語のクラスタリングに用いられている相互情報量を利用した [10]。ここでタグ $w1$ とタグ $w2$ 間の相互情報量は、以下の式で計算される。

$$M(w1, w2) = log\frac{N * freq(w_1, w_2)}{(freq(w_1) * freq(w_2))} \tag{4.1}$$

ここで、N は映像アノテーションとして収集された論文の文章から抽出されたすべてのシーンタグの総数である。また、$freq(w)$ はタグ w のアノテーション中での出現回数、$freq(w_1, w_2)$ は一文中での共起回数である。この相互情報量を計算することにより、検索キーワードとして選択されたタグと関連するタグを、ユーザに提示する仕組みを実装した。具体的には、検索ボックスに入力された、またはタグクラウドからクリックされたタグから、そのタグとの相互情報量の値が大きい上位 5 つのタグを Ajax（Asynchronous JavaScript and XML）の技術を用いて、検索ボックスの下に関連タグとして表示するようにした (図 4.3)。クエリにタグが複数含まれる場合は、それぞれのタグに関して相互情報量を計算して、それぞれの関連タグを並べて表示する。このようにして関連タグを表示することにより、ユーザに、入力したタグと関連するタグを即座に発見させることができる。もし、関連タグの中に、ユーザが利用したいタグが存在すれば、そのタグは目的の映像シーンを検索するための有用な手掛かりとなる。本研究では、提案手法による大量のアノテーションデータを集めることができなかった上に、収集した論文の分野に偏りがあったため、関連タグとして表示されるタグの妥当性について検証は行わなかった。検索に利用するためにより良い関連タグを提示するためには、大量の論文のテキストデータに基づいたタグのクラスタリングを行う必要があるが、それは今後の課題である。

4.1.2 インクリメンタル検索

ユーザは、検索ボックスでキーワードを入力する際に、インクリメンタル検索によって、すべてのタグにアクセスできる。インクリメンタル検索とは、絞り込み検索とも言われ、検索ボックスに文字を入力することで、入力文字から始まる、も

図 4.3: 関連タグ

しくは読みの始まりが入力文字であるシーンタグが検索ボックスの下にポップアップで表示される (図 4.4)。例えば、検索ボックスに「し」と入力された場合、「システム」「自然言語処理」などの検索結果がポップアップ表示される。ポップアップで表示されるタグの数は制限されており、[..more] という文字をクリックすることで、検索にマッチしたすべてのタグを閲覧することができる。Ajax の技術を利用してタグの読み込みを行うため、検索ボックスの内容が変わるごとに、ページを再読み込みすることなく、データベースへアクセスが行われる。この機能により、膨大な数のタグの中から効率よく目的のタグを探すことに加え、データベースに存在しないタグによって検索してしまうということを防ぐ効果が期待できる。また、インクリメンタル検索の結果のうち、アノテーション中での出現頻度の高いタグの関連タグをポップアップの一番下に表示するようにした。インクリメンタル検索によりある程度タグが絞り込まれた際、入力しようとしていたキーワードとは違うが、目的の映像シーンに近いキーワードが関連タグとして表示された場合に、この関連タグが有用な情報となる。

4.2　論文が関連付けられた映像シーンの検索

本システムでは、タグクラウドからクリックにより選択されたタグ、及び検索ボックスに入力されたキーワードと同期して、キーワードの検索結果として論文の文章が関連付けられている映像シーンをランキング表示する仕組みが実現されている。図 4.5 にその検索結果の例を表示する。この検索の詳しいアルゴリズムについては、次節で説明するが、シーンに付与されているタグとテキストに基づいてスコアリングを行っている。この検索結果では、シーンの情報として、シーンのタイトル、シーンのサムネイル、及び、シーンに関連付けられた論文の部分の文章、さらにその論文のタイトルを表示する。ユーザは論文の文章を閲覧するこ

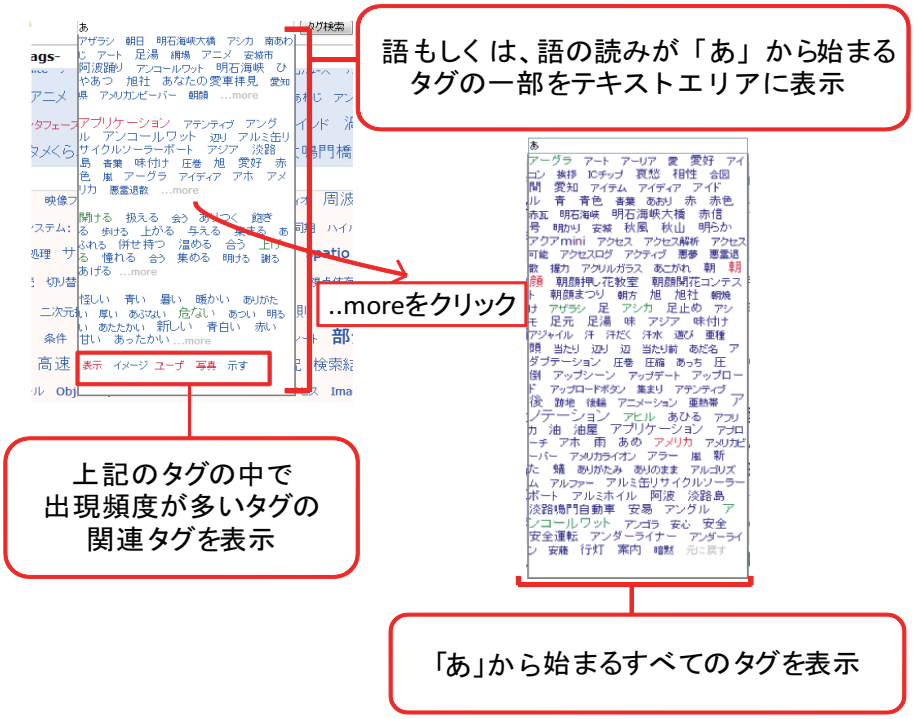

図 4.4: インクリメンタル検索

とで映像シーンの内容を概ね理解することが可能となる。また、映像に関連付けられている論文のタイトルを知ることができ、表示された論文のうち、興味のあるもののタイトルをクリックすることで、論文タイトルをクエリとした映像シーンの検索結果に切り替わる。シーンタイトルをクリックすると、シーンを含む元のコンテンツが、このシーンの開始時間から再生される。2 章でも述べたように、Divie における映像シーン検索では、ピンポイントに適切なシーンを検索結果として表示する仕組みではなく、コメントやブログの文章に基づく映像アノテーションの特徴である網羅性や信頼性の問題を考慮した上で映像シーンを検索するために、含まれる映像シーンのスコアから映像コンテンツのスコアを算出して、映像コンテンツをランキング順に表示していた。本研究で収集される論文のテキストは、信頼性の高いアノテーションとして見ることができるため、この検索インタフェースではピンポイントに映像シーンを検索できるようにした。これにより、アノテーションが行われている映像シーンを容易に検索することができる。一方で、論文の文章はアノテーション目的で記述された文章ではなく、シーンのセグメントもユーザによって行われたものであるため、論文によるアノテーションも、元の映像コンテンツに対して網羅的に行われるわけではない。よって、Divie と同様な映像コンテンツ全体を俯瞰するインタフェースも必要となると考えられるため、この検索結果では Divie と同様な映像コンテンツ全体を俯瞰するインタフェースへ

のハイパーリンクを表示するようにした。これにより、検索結果として表示されたシーンの、前後の文脈の情報や、映像コンテンツ全体に対する位置づけなどの詳細な情報を確認できるようなる。本システムで開発した映像コンテンツ全体を俯瞰するインタフェースについては第 4.5 節で詳細に説明する。

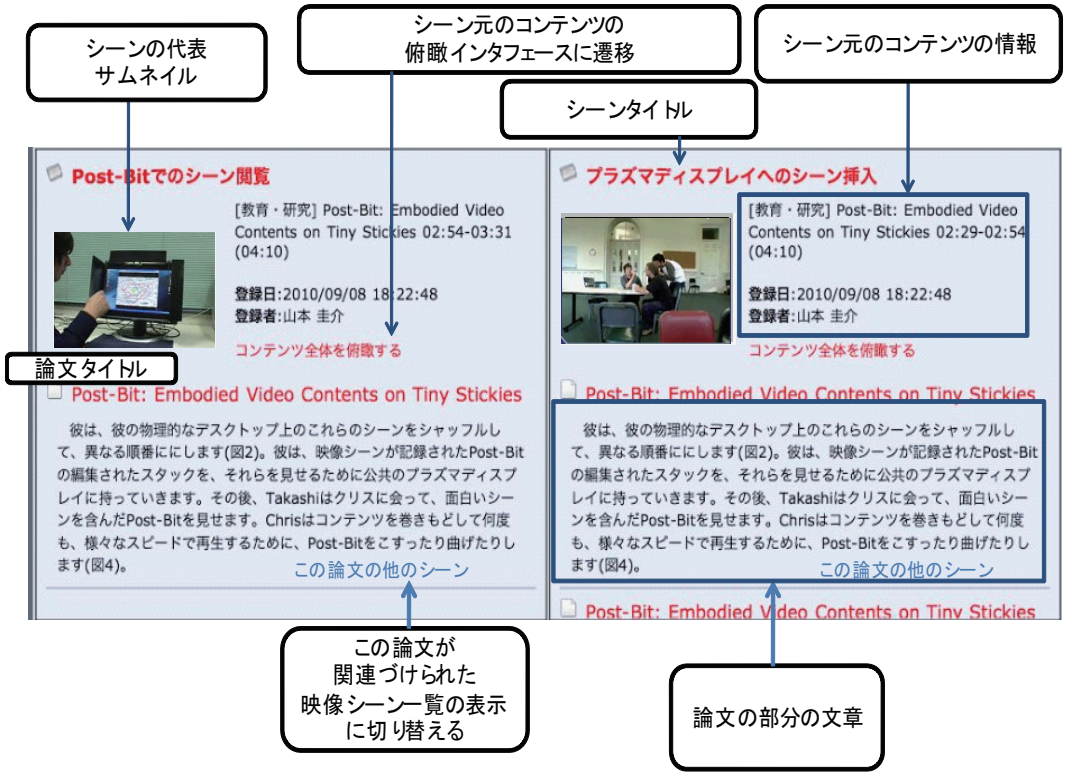

図 4.5: 論文が関連付けられているシーンの検索結果

4.3　検索アルゴリズム

本システムでは、タグの選択によりクエリを決定して検索すると、タグが関連付けられているシーン、または関連付けられているテキストにタグが含まれるシーンのスコアリングが行われる。前節で説明した、引用されたシーンの検索では、このスコアリングの結果を用いる。引用されたシーンを含む映像コンテンツの検索の際は、シーンのスコアに基づいてコンテンツがスコアリングされ、その結果によって、ランキングされたコンテンツが検索結果として表示される。

4.3.1 シーンのスコアリング

検索クエリとしてのシーンタグをkとした場合に、以下の条件に当てはまるシーンに対してスコアリングが行われる。

A 論文の文章から抽出されたシーンタグ k が関連付けられているシーン

B シーンタイトルに k が含まれるシーン

C 関連付けられている論文部分のテキストに k が含まれるシーン (A との重複を除く)

D 関連付けられている論文のタイトルに k が含まれるシーン

E タイトルに k が含まれる映像コンテンツ中のシーン

そして、該当するそれぞれの条件に対するスコアは、該当条件と関連付けられているタグの個数によって決定され、該当するすべての条件に対するスコアを足し合わせることによって各シーンのスコアリングを行う。

検索クエリのタグ k が関連付けられているシーン s のスコアリングを以下のように表現する。

$$scenescore(s,k) = \sum_{c \in C} score(c) \times N(c,k) \tag{4.2}$$

ここで、C は $A-E$ までの該当条件の集合であり、c はその中の該当条件の 1 つの要素である。また、$score(c)$ は該当条件 c の重みで $N(c,k)$ は、該当条件 c におけるタグ k の個数である。タグが複数入力された場合は、検索クエリに含まれるタグの集合 K に対する重み $scenescore'(s,K)$ を以下の式で表現する。

$$scenescore'(s,K) = (1 + \frac{\sum_{k_i,k_j \in K} D(s,k_i,k_j)}{|K|}) \times \sum_{k \in K} scenescore(s,k) \tag{4.3}$$

ここで、$D(s,k_i,k_j)$ は、シーン s にタグ k_i と k_j が共に関連付けられていたかどうかの値である。この式では、検索タグの集合に含まれる異なるタグが、同一のシーンに関連付けられている場合、そのシーンの重みを上げている。例えば、ある 2 つの検索タグで検索された場合に、1 つのタグのみでヒットしたシーンよりも、2 つのタグでヒットしたシーンスコアを優先して高くしている。

上記の条件と一致するシーンの、それぞれの条件に関するスコアについて述べる。本システムは、タグによる検索を主に行うので、テキストとのパターンマッチングによる $B-E$ のスコアよりタグが付与されているシーンのスコアを優先し

て高いスコアにした。また、シーンタイトルは映像の内容を簡潔に表した文章と考えられるため、シーンに対して、論文の文章よりも関連度が高いとして、シーンタイトルに含まれる条件のスコアを、論文のテキストに含まれる条件のスコアより高くした。一方、シーンに関連付けられている論文のタイトルや映像コンテンツのタイトルは、広い意味ではそれぞれのシーンとの関連があると考えられるが、シーンの内容そのものにはそれほど関連がないと推測されるため、シーンタイトルや論文のテキストよりも低いスコアとした。以上から、上記の条件に関して、$A > B > C > D = E$ という重みでスコアを決定した。現在は、スコアの値を経験的に決定しているが、最適なスコアを学習によって決定する仕組みが理想的であり、それは今後の課題の一つである。

4.3.2　コンテンツのスコアリング

前節で説明した検索結果では、このシーンスコアに基づいて、論文が関連付けられているシーンがランキング表示される。一方、検索クエリのタグが論文が付与されているシーンを含む映像コンテンツの検索を行った場合は、式 4.2 のシーンスコアを基にして、検索クエリに対する映像コンテンツのスコアリングを行い、検索結果としてコンテンツをランキング表示する。ただし、コンテンツのスコアリングを行う際は、コンテンツタイトルによるシーンスコアを高くしてしまうと部分引用関係によって関連付けられたテキストに関わらず、作成されたシーンが多いコンテンツほど高いスコアリングが行われてしまう問題があるため、E のスコアは、D よりさらに低いスコアとした。

コンテンツのスコアリングについては、シーンスコアを式 4.2 として、増田らの Divie の検索システムと同様のアルゴリズム [3] を用いた。このアルゴリズムでは、コンテンツスコアはコンテンツに存在するシーンのスコアの平均値であるが、複数のタグが入力された場合は、各タグのコンテンツスコアの合計に加え、複数のタグが同一のシーンに付与されているものが存在するコンテンツを優先してランキングの上位に表示するようにしている。例えば、2 つのタグで検索した場合は、片方のタグしか付与されていないシーンを含んだコンテンツよりも、両方のタグが付与されているシーンを含んでいるコンテンツの方が高いスコアとなる。

4.4　論文の文章と時間軸シークバーを利用した映像コンテンツの俯瞰支援

前節のアルゴリズムによって、映像コンテンツに対するスコアリングが行われ、検索結果として映像コンテンツがスコアの高い順にリスト表示される。そして、それぞれのコンテンツに対して、コンテンツに含まれるシーンとそれに関連付けら

れた論文の情報を表示する仕組みを提供することによって、目的の映像シーンが発見されやすくなる。また、このインタフェースでは、映像シーンを発見するだけではなく、その映像について深く知ることを可能にするために、映像と関連する論文へのリンクを表示することで、ユーザの知的活動を支援することができる。

具体的には、検索結果に含まれるすべてのコンテンツに対して、コンテンツ中の任意のタイムコードに対する情報を閲覧するための時間軸シークバーと、コンテンツに関連付けられているすべてのシーンタグによるタグクラウド、シーンに関連付けられている論文の部分の文章、及び関連付けられた文章の元となった論文の情報を表示する。図 4.6 に検索結果に含まれるコンテンツに対してコンテンツの俯瞰とシーンの映像内容の閲覧を行うユーザインタフェースを示す。

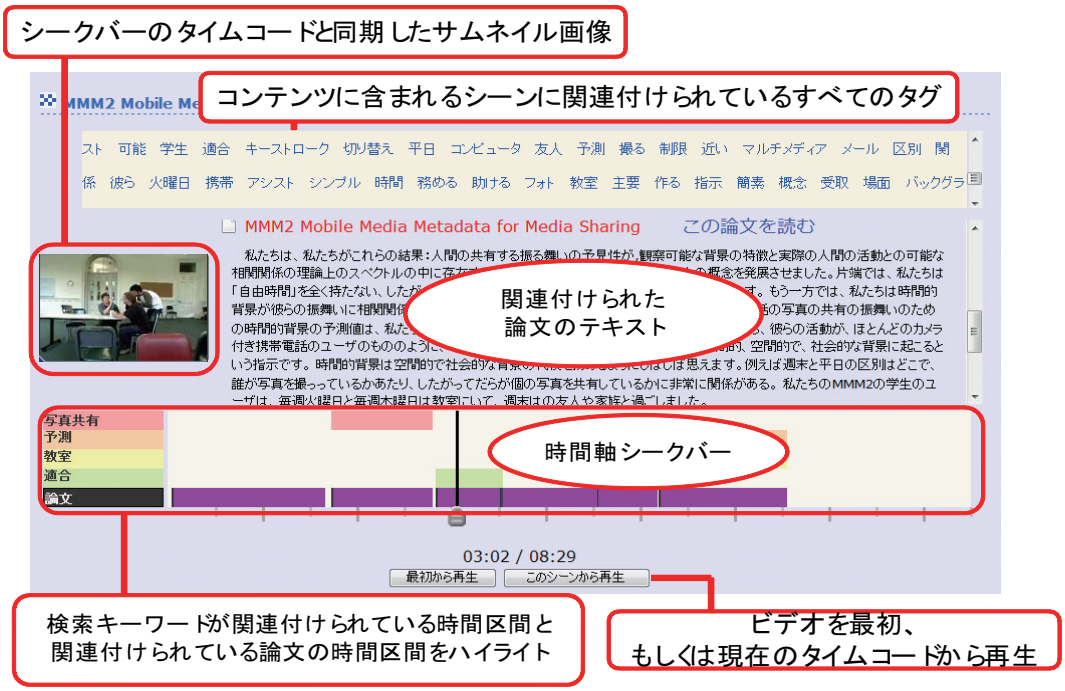

図 4.6: 映像コンテンツの俯瞰インタフェース

このインタフェースでは、時間軸シークバーは、Divie の検索インタフェース [3] と同様に、マウスドラッグによるシーク操作によって任意のタイムコードやシーンに対する情報を閲覧するために利用される。時間軸シークバー上には検索キーワードが関連付けられている時間区間と、関連付けられた論文ごとの、このコンテンツに含まれるすべてのシーンの区間が表示される。シークバーをマウスドラッグすると、そのタイムコードと同期したサムネイル画像を読み込んで表示する。さらに、そのタイムコードを含むシーンが存在する場合は、関連付けられている論文部分のすべての文章をサムネイルの画像の横に表示する。同じタイムコードに複数のシーンが存在した場合は、表示するシーンの情報を切り替えることができる。

このようにして、サムネイルの画像を見ながら、映像に関連付けられた論文の文章を読むことで、そのシーンの内容について理解を深める事ができる。また、その論文の文章の元となるコンテンツが存在する TDAnnotator へのハイパーリンクを表示し、別ウィンドウで論文が閲覧できるようにした (図 4.7)。リンク先では、映像と関連付けられている論文部分が、ハイライトされて表示される。これにより、素早く論文部分の周辺の文脈を確認することができ、映像に対する理解をより深めることができる。

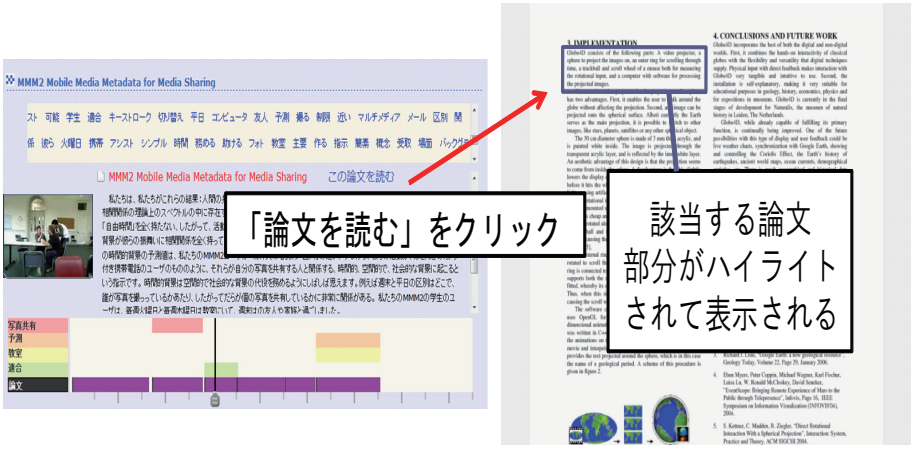

図 4.7: TDAnnotator へのハイパーリンク

コンテンツに存在するすべてのシーンのタグを表示したタグクラウドは、コンテンツ全体の情報を俯瞰するためのキーワードとして利用することができる。また、タグをクリックすることによって、それが関連する時間区間を時間軸シークバーに反映させることができ、それによってシーンの絞り込みや、コンテンツに含まれる別のシーンについて知ることができる。

また、コンテンツに関連付けられている論文の情報を検索結果の下部に表示し、その論文をクエリとした映像シーン検索、及び論文の閲覧を可能にした。論文をクエリとした映像シーン検索では、クエリの論文が関連付けられているシーンを含む映像コンテンツを検索することができる。この検索では、クエリの論文が関連付いているシーンの数でコンテンツのスコアリングを行う。この検索は、1 つの論文が複数の映像コンテンツに対して関連付いていた場合に有効である。論文の閲覧は、論文部分のハイパーリンクと同様に別ウィンドウで表示された TDAnnotator のページで行える。

それらの操作を繰り返してシーンの情報を閲覧し、サムネイル画像や論文の文章を基に閲覧したいシーンを発見したら、映像コンテンツが存在するページでそのシーンの開始時間からビデオ再生を行うことで映像シーン検索を実現する。このインタフェースでは再生するシーンの開始時間をユーザが設定できるため、す

でにあるシーンの時間区間に限定されない映像シーンの柔軟な閲覧が行える。

これらの仕組みによって、映像コンテンツの俯瞰効果を高め、コンテンツの内容理解を支援する。Divie と同様に、シーンタグや映像に関連するテキスト、サムネイル画像を用いる映像コンテンツそのものにアクセスすることなくシーンの情報を閲覧することができるため、一般に時間がかかるビデオへのアクセスを最小限にして、映像シーン検索を効率的に行うことができる。

さらに、このインタフェースでは映像シーン検索に加え、映像シーンと関連する論文へのハイパーリンクを表示することで、ユーザの論文に対する興味・関心を高め、研究などの知識活動を支援する。映像から論文というアクセスのパスが増えることで、新たな映像と論文の部分引用関係の獲得も期待できる。

本章では、映像と論文の部分引用関係から得られた映像アノテーションに基づいた映像シーン検索システムを開発した。しかし、3 章の実験で、検索に有用である特徴語が多く抽出されることはある程度確認できたが、本システムを用いた映像シーン検索の精度については、実験を行えるだけのアノテーションのデータが不足していたため、検証を行うことができなかった。また、論文の文章をシーンと共に表示することによるユーザに対する影響などの、ユーザインタフェースに関する検証も今後の課題の 1 つである。

次章では、関連研究として、映像コンテンツに対するアノテーションに関するその他の研究や、タグ間の関連性を抽出するために、今まで行われている研究について述べる。

第5章　関連研究

本章では、映像アノテーションに関連する研究として、専用ツールを用いた映像アノテーションの研究と、Web コミュニティ活動に基づく映像アノテーション手法に関する研究、さらに、映像の中で使用されたプレゼンテーションのスライドを映像アノテーションとして用いる研究について述べる。また、本研究で関連タグの表示を行ったことに関連して、タグのクラスタリングに関する研究について述べる。

5.1　専用ツールを用いた半自動アノテーションに関する研究

映像アノテーションを専用のツールを用いて作成する手法が提案されている。これらの手法で使用されるツールの多くは、MPEG-7[1]によってアノテーションの記述を行う。MPEG-7 は、XML 形式でマルチメディアコンテンツに関するアノテーションを記述する仕組みで、映像に対する何らかの知識を持った専門家が、その映像に対して詳細な情報を記述する目的で使用される。しかし、再生時間の長い映像コンテンツに対しては、アノテーションを行う専門家にとってコストが高く、多種多様な分野の学術的内容の映像が存在する場合に、このアノテーション手法を適用することはあまり現実的でない。一方で、この手法で作成されるアノテーションは、専門家の手によるもので、視聴者やシステムのコンテンツ理解を目的として作成されたものであるから、論文の文章から抽出したタグよりも信頼性の高いメタ情報と見ることができ、映像全体に対して網羅的に記述されることも期待できる。よって、短いコンテンツや商用などの一部の映像コンテンツに対してアノテーションを行う際には、非常に有効な手段である。

代表的なツールにビデオアノテーションエディタ [8] がある。ビデオアノテーションエディタは、音声処理や、カット検出・オブジェクトトラッキングなどの機械処理から自動的に生成されるアノテーションを人手で編集することで、詳細な映像アノテーションを作成することができる。機械処理による自動アノテーションと、人手によるアノテーションを 1 つの時間軸を元に統合して処理できるのが特徴となっている。

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

IBM が作った MPEG-7 対応アノテーションツールである VideoAnnExAnnotation Tool[9] は映像コンテンツに、シーンや存在するオブジェクトに対して意味属性を映像アノテーションとして付与することができる。特徴として、木構造によって意味属性ごとにカテゴリ分けして、アノテーションを保存している。また、シーンの一部の領域に対してアノテーションを付与することも可能である。さらに、これを拡張したツールでは、Web を用いて、複数人の人間による協調作業によりアノテーションを作成することができるため、大量の映像に対しても比較的効率よくアノテーションを行うことが可能である。

これらの専用ツールを用いたアノテーション手法では、映像の内容に対する詳細な知識を持った専門家が付与するとは限らないため、そのアノテーションから、映像に写っていない専門用語などの特徴語を獲得できるとは限らない。

5.2 Web コミュニティ活動に基づく映像アノテーションに関する研究

2 章で述べた Synvie 以外の Web コミュニティ活動に基づく映像アノテーションに関する研究について述べる。

映像シーン連動型掲示板コミュニケーションシステム SceneNavi[13] では、Web 上に存在している映像を閲覧しているユーザの間でコミュニケーションを行うシステムであり、映像に対して同期、または非同期で掲示板型のコミュニケーションが可能である。特徴として、映像アーカイビングシステム SceneCabinet[14] を用いることにより、映像からまとまった時間区間であるシーンを分割しておき、各シーンに対してアノテーションが付与できるという点が挙げられる。SceneCabinet では、ショット切換、テロップ、カメラワーク、音楽区間、音声区間といった映像処理技術により高い精度でシーンを検出している。この手法では、あらかじめ、映像を分割してシーンをユーザに対して提示して、コミュニケーションにおけるコメントによりアノテーションを獲得しているが、多種多様な映像コンテンツに適用する場合は、機械によるシーン分割の際に、意味的な情報が考慮されているかどうかなど、シーン分割の信頼性に関して問題点がある。

次に、日本最大の電子掲示板である 2 ちゃんねるの書き込みから得られる情報から、番組コンテンツのインデキシングをしたり、盛り上がりを検出したりすることで、様々な視点による視聴可能なビューを生成する研究が宮森らによって行われている [18]。具体的には、テレビ番組の放送中に、リアルタイムで番組について 2 ちゃんねるに頻繁に書きこまれるアスキーアートや、盛り上がりや落胆を表すような単語をパターンマッチングにより検出する。開発されたシステムでは、映像に対する視聴者の反応の大きさや盛り上がり、落胆の度合いなどを利用した様々な視点による番組の視聴を行うことができる。掲示板に書き込みを行った視

聴者の ID により、特定の視聴者のみの書き込みを見ることも可能である。この映像アノテーションでは、大量のアノテーションが集まりやすく、盛り上がるシーンなどを検索する際には有用であるが、掲示板に書かれるコメントの質は高いとは言えず、専門用語などの特徴的な語をうまく抽出する際にはあまり有効ではないと思われる。

5.3 講義スライドを用いた映像アノテーションに関する研究

講義映像に対して用いられたプレゼンテーションのスライドから映像アノテーションを作成する手法が提案されている。本研究で使用した映像の中にも、スライドを利用した発表映像などが含まれていたため、関連研究として取り上げた。

山本らが開発した講義コンテンツ共有システム [16] では、講演者のスライド操作情報を記録するソフトウェアを用いて、スライドの切り替え時刻や、スライド内のアニメーションの操作時の時刻によって、シーンを決定し、シーンと対応するスライドに対してアノテーションの付与を可能にしている。具体的には、ユーザによるスライドの部分の引用や質問提示版に対するコメントに基づいてアノテーションの獲得を行っている。これらのアノテーションには、Web コミュニティ活動における映像に対するコメントと同様に、口語的な文章が多く含まれ、必ずしも映像の内容を的確に表現しているとは限らず、映像と関係性の薄い文章が付与された場合、あまり検索に有用とは言えない。ゆえに、獲得したアノテーションを取捨選択し、検索により有用なアノテーションを作成する仕組みが必要とされる。

小林らが試作したプレゼンテーションコンテンツ蓄積検索システム [17] では、文字認識や画像認識の技術を利用して映像中のシーン情報抽出とスライド同定による同期情報の抽出を自動的に行うツールを用いている。このツールは、動画ファイルと発表資料のスライドを入力として、映像の 1 秒毎に対応するスライドを認識し、そのシーン情報と同期情報を XML ファイルとして出力する。さらに、このファイルから、スライドに存在する文字列とその位置の情報をメタデータとして獲得し、検索の際には、スライドのキーワードを含む文字列のインデント情報により重みを計算している。このスライドの構造に基づく映像アノテーションは、プレゼンテーション資料と映像さえあれば、自動でアノテーションを作成できるため、アノテーションを作成するコストが少ない利点がある。また、スライドに記述されている内容と、講義映像で説明されている内容は、関係性が強いと考えられるため、アノテーションとして比較的信憑性が高いと考えられる。ただし、この手法が適応できる映像は、プレゼンテーションスライドを用いた講義映像や発表映像に限られる。また、スライドの情報は、講義映像中の発話内容を要約したテキストである可能性があるため、本手法のアノテーションである論文より情報

量が少ないことも考えられる。

5.4　タグのクラスタリングに関する研究

高度な検索を行うためには、大量のアノテーションデータから抽出されるタグから、さらに、ユーザにとって重要なタグを選別する必要がある。本研究では、アノテーション中のタグの共起頻度を用いた相互情報量により、タグ間の関係性を計算したが、これまでにも、何らかの形でタグの関連度を計算し、それに基づいてタグをクラスタリングする研究が行われてきた。Li らはソーシャルブックマークのタグについて、ISR (Inter-Section Ratio) 手法を用いた階層化を行い，その結果を用いたタグ閲覧システム ELSABer を提案している [12]。ISR 手法では、タグ間の上下関係を、タグの特定の Web ページの出現頻度と、共起確率により求めている。

また、榊らは、Web 全体の単語の出現確率や共起頻度を検索エンジンにより求め、出現確率のばらつきを考慮するために、ISR 法の代わりに χ^2 値を用いた単語間の関連度の指標を使って、単語の親子関係を導出する手法を考案している [19]。χ^2 値は、あるデータ集合内での統計的な偏りを表す指標であり、機械翻訳の手法でも用いられている。本研究で用いた相互情報量は、関連度を計算する際に、各語の出現確率に数千倍、数万倍といった開きがある場合、値の信頼性は低くなるという問題があるため、大量のアノテーションデータを扱う際にはこの研究の手法の方が有効であると考えられる。

第6章　まとめと今後の課題

6.1　まとめ

本研究では、映像と論文の部分引用を行うための仕組みを実現し、そのユーザによって作成された部分引用関係から、映像アノテーションとして論文に含まれるテキストを獲得する仕組みについて提案を行った。さらに、従来手法であるオンラインアノテーションのコメントの文章から得られるタグと、論文の文章から得られるタグを比較することにより、提案手法の有効性について確認した。そして、獲得した映像アノテーションに基づいて、学術的な内容の映像シーンを検索するシステムを開発した。

本論文の第 2 章においては、オンラインアノテーションに基づく映像シーン検索について述べた。まず、オンラインアノテーションの特徴について筆者の研究室で開発された Synvie を参照しながら述べた。次に、オンラインアノテーションに基づいたシーン検索システムである Divie の特徴や利点についてを述べ、最後に、Divie の検索によるシーン検索の問題について論じた。

第 3 章においては、本研究で開発した、映像と論文を部分引用するための複数のアノテーションシステムと、そのシステムで得られる映像アノテーションの特徴について、さらに、その特徴の、検索における有効性について検証するための実験について述べた。まず、研究活動において引用したコンテンツを記録することができる DRIP システムに関して述べ、映像あるいは論文コンテンツを引用して、それらのコンテンツを管理する仕組みについて説明した。次に、映像の部分であるシーンを引用するために、登録された映像コンテンツからシーンの開始時間と終了時間を決定して、シーンを定義し、その情報を DRIP システムに取り込むための仕組みについて説明した。また、登録された論文部分の文章を引用するために、論文部分を、矩形領域として選択し、その部分に対して翻訳やコメントなどのアノテーションの付与が行える TDAnnotator について述べ、そのアノテーション情報を同じく DRIP に取り込む仕組みを説明した。さらに、これらのシステムから得られた情報に基づいて、DRIP システムにおいて引用された映像シーンと論文部分を関連付ける仕組みについて述べた。

この仕組みを用いて作成された、論文と映像の部分引用関係から得られる論文の文章は、映像アノテーションとして質が高いことを、文章の正しさや、専門用語などの特徴語の得られやすさなどから述べた。次に、論文から得られる特徴語の

量や質について検証するために、上記のシステムを用いて作成した部分引用関係から抽出された名詞タグを、論文と共に引用された映像シーンに対するコメント(従来手法により獲得)から抽出された名詞タグと比較した。その結果、論文には、論文からは特徴語が得られやすいことが分かった。また、論文の特徴語には、コメントから得られるタグよりも固有名詞や専門用語が多く含まれることが分かった。これにより、論文から抽出されるタグはコメントよりも特徴的なタグが多いことが確認でき、従来手法であるコメントからのタグ抽出よりも検索に有効なタグであることが分かった。

第 4 章においては、提案手法により得られた論文の文章を映像アノテーションを用いて、学術的な内容の映像シーンを効率的に検索する仕組みを提案した。本研究で開発した検索システムは、最初に、論文の文章から抽出されたタグをタグクラウドとして表示し、ユーザーが検索クエリとして利用するタグを発見するための手掛かりを提供した。また、クエリを入力する際にも、インクリメンタル検索や、関連タグを用いて、ユーザが利用したいタグを選択しやすくした。関連タグは、論文の文章の中に存在する共起頻度によりタグ同士の関連度に基づいて表示される。また、タグの入力あるいは選択と同期して、タグクラウドの下に、論文が関連付けられているシーンを一覧表示するようにした。これにより、いままで、研究活動で引用された映像シーンをすぐに発見することができる。また、検索クエリが送信されると、タグが付与されているシーンを基に、映像コンテンツがランキング表示される。表示されたコンテンツ内では、映像コンテンツに対するタグクラウド、そのコンテンツに含まれる引用されたシーンに関連付けられた論文の文章、引用されたシーン区間を表す時間軸シークバーなどが表示される。このように提示された情報によって、引用されたシーンを含む映像コンテンツ全体を俯瞰し、シーンやその周辺の映像内容を効率よく理解することができ、目的の映像シーンを容易に発見することできる。また、論文の文章には TDAnnotator へのハイパーリンクが付与されており、オリジナル論文の閲覧が簡単にできるようにした。論文を読むことにより映像を深く理解することができる上に、映像から論文というアクセスのパスが増えたことで、新たな映像と論文の部分引用関係が獲得できるようにしている。

第 5 章においては、アノテーション手法に関して、専門ツールを用いた映像アノテーションに関する研究、Web コミュニティ活動に基づく映像アノテーションに関する研究、スライドを利用した映像アノテーションに関する研究、さらに、関連タグに関連して、タグのクラスタリングに関する研究を関連研究として説明した。

6.2 今後の課題

6.2.1 論文と映像の関係付けに対するモチベーションの向上

本研究で用いた映像と論文の部分引用関係は、実験のために、筆者らの研究室で作成したものであり、集められた論文や映像は特定の分野に偏っている。高度な映像シーン検索を実現するためには、大量のアノテーションデータが必要不可欠であり、アノテーションを大規模に収集するためには、多くのユーザに、本システムを使用して映像と論文の関連付けを行ってもらわなければならない。そのためには、互いに関連する映像と論文を容易に発見することができる手段が必要となる。

本システムを使用するユーザのモチベーションについて考察すると、ユーザ自身が論文を執筆し映像を撮影した場合や、ユーザが参考にしたい論文を探す際に、論文と映像が一緒に公開されていれば、映像と論文を関連付けるモチベーションは高いと考えられる。しかし、論文を探している際に、論文のみが公開されていた場合は、論文と関連している映像を探すというモチベーションは一般には高くないだろう。論文に関連する映像があると記述されている場合を除き、存在するかどうか分からない、論文と関連する映像を積極的に探すという状況は考えにくい。逆にユーザが、学術的な内容の映像を発見した際に、関連する論文があるかどうか探すモチベーションは比較的高いと思われる。ユーザにとって、興味のある映像ならば、より深く映像の内容について知りたい、という欲求が生まれると考えられるためである。したがって、論文と映像の関連付けに対するモチベーションの向上のためには、学術的な内容を含む映像コンテンツが充実している環境を構築することが有効であると考えられる。今後、学術的な内容の映像コンテンツのみを対象とした映像公開サービスの展開や、そのための検索システムなどが開発されれば、映像から論文を探すモチベーションの向上が期待できる。

また、映像と論文を探すこと以外にも、我々が開発したシステム自体の使いやすさ、すなわち、手軽に論文や映像を引用するためのユーザインタフェースも モチベーションの向上における重要な要素であるため、システムを運用する上で考慮する必要がある。

6.2.2 映像シーン検索に関する評価

本研究では、映像と論文の部分引用関係に基づく映像アノテーションからは、従来手法であるコメントより、検索に有用な特徴語が得られやすいことを実験により示した。しかし、検索の精度を検証するだけのアノテーションデータが不足していたために、開発した検索システムを用いた映像シーン検索についての評価実験は行うことができなかった。また、検索の精度だけでなく、検索のユーザインタ

フェースや、検索過程におけるコンテンツの俯瞰支援効果についての検証を行うことができなかった。これらの評価を行うためには、検索クエリに対して、引用された映像シーンを表示するインタフェースと、映像コンテンツ全体を俯瞰するインタフェースとを検索時間で比較したり、検索結果上で提示される論文の文章の情報によって映像内容をどれだけ理解できるかどうかを被験者実験によって検証したりすることが必要である。ただし、映像全体を俯瞰するインタフェースは、最終的には探している映像シーンを発見することを目的としているものの、検索の過程で、そのシーンの全体における位置付けや、シーンの前後の文脈を確認できるという利点もあるため、必ずしも検索時間のみでは、評価することはできない。そのような利点を評価する実験も今後の課題である。

6.2.3　関連タグに関する評価

本研究では、タグの論文中での共起頻度により相互情報量を計算し、タグ間の関連度を計算し、ユーザーが検索クエリとして選択したタグのとの相互情報量が高いタグを関連タグとして表示した。しかし、関連タグの検索における有効性について、データが不足していたため検証できなかった。本来ならば、大量のアノテーションデータに基づいたタグのクラスタリングを行い、関連タグを用いた場合とそうでない場合とでの検索精度の比較を行う必要がある。

また、今回は、論文部分のテキスト（あるいは論文部分で付与された翻訳テキスト) での 1 文中の共起頻度を用いたが、単一の論文の文章全体での共起頻度、あるシーンに関係付いたすべての映像アノテーションに含まれるテキスト中での共起頻度、及びそれらを組み合わせた共起頻度など、様々な共起頻度の求め方が考えられる。また、共起頻度だけでなく、同時にクエリに用いられる複数のタグなど、ユーザの検索履歴の情報も組み合わせることも考えられる。

さらに、「共起する」というだけの漠然と広い意味での関連ではなく、高度な言語解析等により、タグ間の親子関係、類義語、反意語、などといった意味的関連性を抽出することができれば、タグの関連度を検索の目的によって変化させることができ、タグ間の関連によるシーンの推薦などの応用にも利用できる可能性がある。

6.2.4　論文部分の文章と映像シーン間の意味的関係の抽出

本研究では、論文部分と映像の間に、ユーザが自由に関連付けを行っているが、どのような意味で関連付けているかという深い意味的関係については考慮されていない。映像シーンに対して論文の部分を関連付ける理由としては、「アルゴリズムが似ている」、「同じ実験機材を使っている」、「正反対の研究である」など様々

な理由が考えられる。このような関係が、高度な言語解析技術や、ユーザに引用の意図を自然なやり方で入力させることなどによって抽出できれば、ユーザの検索の目的によって検索結果の内容を変えるなどといった高度な検索が実現できると考えられる。これらについて対処することも今後の課題の一つである。

謝辞

本研究を遂行するにあたり、指導教員である長尾確教授をはじめ、数多くの方々に御支援、御協力を頂きました。この場で、感謝の言葉を申し上げたいと思います。

長尾確教授には、ゼミ等で、研究に対する姿勢や心構えといった基礎的な考え方から、研究に関する貴重な御意見、論文執筆に関する御指導を頂くなど、大変御世話になりました。心より御礼申し上げます。

松原茂樹准教授並びに、大平茂輝助教には、ゼミ等で、研究の本質的なことに関する貴重な御意見を頂き、大変御世話になりました。心より御礼申し上げます。

土田貴裕さんには、プログラミングや研究に関する様々なアドバイスや御指導を頂き、大変御世話になりました。ありがとうございました。

石戸谷顕太朗さんには、プロジェクトのミーティング等で、ネットワークなどの基礎的な技術に関することや、研究に対する基礎的な考え方まで数多くの御指導を頂き、大変御世話になりました。ありがとうございました。

山本圭介さんには、研究に使用するデータの収集や、実験の分析の際に、数多くの御指導、協力を頂き、大変お世話になりました。ありがとうございました。

森直史さん、木内啓輔さん、井上泰佑さん、高橋勲さん、磯貝邦明さん、渡邉賢さんには、ゼミ等で貴重なご意見を頂いたことに加え、研究室における様々な活動の中で御世話になりました。ありがとうございました。

長尾研究室秘書である鈴木美苗さんには研究室における生活全般に関する様々な面で御世話になりました。ありがとうございました。

最後に、日々の生活を支えていただいた両親にも最大限の感謝の気持ちをここに表します。ありがとうございました。

参考文献

[1] Katashi Nagao, Yoshinari Shirai, and Kevin Squire. Semantic annotation and transcoding: Making Web content more accessible. IEEE MultiMedia, Vol.8, No.2, pp.69-81, 2001.

[2] Daisuke Yamamoto, Tomoki Masuda, Shigeki Ohira, Katashi Nagao. Video Scene Annotation Based on Web Social Activities. IEEE MultiMedia, Vol.15, No.3, pp.22-32, 2008.

[3] 増田 智樹. 映像コンテンツの高度な引用とシーン検索への応用に関する研究. 修士論文, 情報科学研究科 メディア科学専攻, 名古屋大学, 2009.

[4] 土田 貴裕, 大平 茂輝, 長尾 確. ゼミコンテンツの再利用に基づく研究活動支援. 情報処理学会論文誌, Vol.51, No.6, pp.1357-1370, 2010.

[5] 梶 克彦. デジタルコンテンツのアノテーション基盤技術とそれに基づく音楽情報処理に関する研究. 博士論文, メディア科学専攻, 名古屋大学大学院情報科学研究科, 2007.

[6] Mike Dowman, Valentin Tablan, Hamish Cunningham, Borislav Popov. Web-Assisted Annotation, Semantic Indexing and Search of Television and Radio News. Proceedings of the 14th International Conference on World Wide Web, 2005.

[7] Michael A. Smith, Takeo Kanade. Video Skimming and Characterization through the Combination of Image and Language Understanding. Proceedings of IEEE International Workshop on Content-Based Access of Image and Database, p.61-71, 1998

[8] Katashi Nagao, Shigeki Ohira, and Mitsuhiro Yoneoka. Annotation-based multimedia summarization and translation. Proceedings of the Nineteenth International Conference on Computational Linguistics, pp.702-708, 2002.

[9] Ching-Yung Lin, Belle L. Tseng and John R. Smith. Video collaborative annotation forum: Establishing ground-truth labels on large multimedia datasets. Proceedings of the NIST TREC 2003 Text Retrieval Conference, 2003.

[10] 松尾 豊, 石塚 満. 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム. 人工知能学会論文誌, Vol.17, pp.217-223, 2002.

[11] Moshe Ben-Ezra, Shree K. Nayar. Motion-based motion deblurring. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.26, No.6, pp.689-698, 2004.

[12] Rui Li, Shenghua Bao, Ben Fei, Zhong Su, Yong Yu. Towards Effective Browsing of Large Scale Social Annotations. Proceedings of the Sixteenth International World Wide Web Conference, pp.943-952, 2007.

[13] 嶌田聡, 宮川和, 東正造, 森本正志, 奥雅博, 映像シーン連動型掲示板コミュニケーションを用いたコミュニティ協働型メタデータ抽出方法. 電子情報通信学会論文誌, Vol.J91-D, No.5, pp.1231-1242, 2008.

[14] 谷口 行信, 南 憲一, 佐藤 隆, 桑野 秀豪, 児島 治彦, 外村 佳伸. SceneCabinet: 映像解析技術を統合した映像インデクシングシステム. 電子情報通信学会論文誌, Vol.J84-D-II, No.6, pp.1112-1121, 2001.

[15] 山本 大介, Web コミュニティ活動に基づく映像アノテーションとその応用に関する研究. 博士論文, 大学院情報科学研究科メディア科学専攻, 名古屋大学, 2008.

[16] 山本 大介, 増田 智樹, 大平 茂輝, 長尾 確. 映像アノテーションを獲得・管理する講義コンテンツ共有システム. 情報処理学会第 70 回全国大会, 2008.

[17] 小林 隆志, 村木 太一, 直井 聡, 横田 治夫. 統合プレゼンテーションコンテンツ蓄積検索システムの試作. 電子情報通信学会論文誌, Vol.J88-D-I No.3 pp.715-726, 2005.

[18] 宮森 恒, 中村 聡史, 田中 克己. 番組放送チャットに基づく視聴者視点を利用した放送番組のビュー生成. 日本データベース学会 Letters, Vol.4, No.1, pp.93-96, 2005.

[19] 榊 剛史，松尾 豊，石塚 満. Web 上の情報を用いた関連語のシソーラス構築について. 自然言語処理，Vol.14, No.2, pp.3-31, 2007.