

修士論文

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

350703264 増田 智樹

名古屋大学大学院 情報科学研究科  
メディア科学専攻

2009年1月



# 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

350703264

増田 智樹

本研究では、映像シーンのブログ記事への引用と映像シーンプレイリストの作成を支援する仕組みをユーザに提供し、その編集履歴から映像アノテーションを抽出することによって、効率的な映像シーン検索を実現する仕組みを提案する。

まず、映像シーンを引用する従来の仕組みを改良し、複数の映像コンテンツを対象とした新たな引用の仕組みを提供するシステムを開発した。具体的には、映像シーンを効率的に選択するためのユーザインタフェースを提供し、さらに、共引用された複数のシーンをブログ上で同期的に再生する仕組みを提供する。これらの仕組みによって、引用されたシーンがセグメント情報として蓄積され、シーンに対して記述されたテキストが関連付けられるだけでなく、シーン間の関連を抽出することができる。引用に加え、シーン間の関連をより手軽に抽出するための仕組みとして、引用されたシーンを集めたプレイリストの作成を支援するシステムを開発した。プレイリストは、作成の手がかりとした情報によって分類され、その種類に基づいて抽出された代表キーワードをシーンに対して関連付ける仕組みとなっている。さらに、それぞれの仕組みに基づいて計算されたシーン間の関連度を組み合わせることで、シーン間のリンク構造の解析を行い、その結果を利用することで、シーンとキーワード間の関連度を計算した。

そして、以上の仕組みから収集される映像アノテーションを利用することで効率的に映像シーンを検索するシステムの開発を行った。本システムの検索結果には、該当する映像コンテンツに関するタグクラウド、引用されたシーン区間を表す時間軸シークバー、シーンに関連付けられたコメント文などが提示される。ユーザは、提示される情報によって映像コンテンツを俯瞰し、映像シーンやその文脈を理解することで、最適なシーンを発見することができる。また、シーン間のリンク構造を活用することによって、検索の再現率や適合率を上げるための検索アルゴリズムを考案した。

提案手法の有効性を検証するために、被験者実験によってシーン引用とプレイリスト作成のデータを収集し、映像シーン検索を行った。従来手法の引用の仕組みによって抽出可能なアノテーションと、提案手法によって抽出したアノテーションをそれぞれ利用して検索を行った結果、提案手法を適用することで、シーン検索の再現率、F 値、検索結果上位の適合率が改善された。特に、再現率には 35 %以上の改善が見られ、提案手法の映像シーン検索に対する有効性を確認することができた。



# A Study on Video Scene Quotation and its Application to Video Scene Retrieval

350703264

Tomoki MASUDA

In this paper, we propose an efficient mechanism for retrieving video scenes. The mechanism uses video annotations extracted from user activities to quote video scenes and to create video scene playlists.

In first, we have developed a user interface that users can easily quote an arbitrary time segment of video contents. Our quotation system allows the users to pick up multiple video scenes from multiple video contents. In addition, the system offers mechanisms for browsing multiple video scenes synchronously on Weblogs by quoting multiple video scenes in the same paragraph. When the users quote video scenes, their segments are created and texts written in paragraphs are associated to the video scenes, and video scenes are associated to the other video scenes.

To extract other relations among video scenes, we have also developed a system that enable users to create video scene playlists by using quoted scenes. The playlists are classified based on their creation method, and their representative keywords extracted based on the annotation contents are associated with the video scenes. The relationships among video scenes are analyzed to integrate several kinds of relations extracted two different methods for video scene annotation. Based on this, our system calculates relevance values between video scenes and keywords.

Furthermore, we have developed a tag-based video scene retrieval system. The output of our video scene retrieval is a list of videos each having a collection of keywords, called a tag-cloud, associated with the video and a timeline seek bar on which quoted time segments are highlighted and associated text comments are displayed. These information let users overview and understand video content, and, as a result, video scenes that users want to watch are found. Our developed retrieval algorithm makes precision and recall be increased based on the relationships among video scenes.

Finally, we have performed the subjective experiment to evaluate our proposal mechanism by collecting video annotations and retrieving video scenes. We have retrieved video scenes by using both video annotations extracted from the conventional mechanism and extracted from the proposal mechanism. As a result, recall, F-measure, and top-ranked precision are improved by applying our proposal mechanism. Especially, recall is improved from 53.7 % to 88.8 %. From the result, the usefulness of our proposed mechanism for retrieving video scenes was confirmed.



# 目次

|            |                                         |           |
|------------|-----------------------------------------|-----------|
| <b>第1章</b> | <b>はじめに</b>                             | <b>1</b>  |
| <b>第2章</b> | <b>Web コミュニティ活動に基づく映像アノテーション</b>        | <b>7</b>  |
| 2.1        | 映像コンテンツを話題の中心とした Web コミュニティ活動 . . . . . | 8         |
| 2.2        | 映像全体に対するアノテーション . . . . .               | 11        |
| 2.3        | 映像中の部分に対するアノテーション . . . . .             | 12        |
| 2.3.1      | コメント投稿に基づくアノテーション . . . . .             | 12        |
| 2.3.2      | ボタンクリックに基づくアノテーション . . . . .            | 15        |
| 2.3.3      | 引用に基づくアノテーション . . . . .                 | 15        |
| 2.3.4      | 映像編集に基づくアノテーション . . . . .               | 16        |
| 2.4        | 映像シーンの引用とは . . . . .                    | 16        |
| 2.5        | 従来の引用手法の問題点 . . . . .                   | 17        |
| <b>第3章</b> | <b>映像シーンの引用とプレイリストの作成に基づくアノテーション</b>    | <b>21</b> |
| 3.1        | 映像シーンの引用に基づくアノテーション . . . . .           | 21        |
| 3.1.1      | 引用区間の選択によるセグメンテーション . . . . .           | 22        |
| 3.1.2      | 共引用による複数シーン間の関連付け . . . . .             | 26        |
| 3.1.3      | ブログ記事の編集によるテキストアノテーション . . . . .        | 26        |
| 3.1.4      | 映像シーン引用ブログ . . . . .                    | 27        |
| 3.1.5      | 映像シーンへの自動トラックバック . . . . .              | 28        |
| 3.2        | 映像シーンプレイリストに基づくアノテーション . . . . .        | 29        |
| 3.2.1      | プレイリストに利用するシーンの選択 . . . . .             | 29        |
| 3.2.2      | プレイリストの編集 . . . . .                     | 31        |
| 3.2.3      | プレイリストの共有と再利用 . . . . .                 | 32        |
| 3.3        | アノテーションの解析 . . . . .                    | 32        |
| 3.3.1      | シーンに対するタグの生成 . . . . .                  | 33        |
| 3.3.2      | プレイリストを代表するキーワードの抽出 . . . . .           | 33        |
| 3.3.3      | シーン間のリンク構造 . . . . .                    | 35        |
| <b>第4章</b> | <b>映像シーン検索</b>                          | <b>37</b> |
| 4.1        | タグクラウド . . . . .                        | 37        |
| 4.1.1      | タグのインクリメンタル検索 . . . . .                 | 38        |

|            |                               |           |
|------------|-------------------------------|-----------|
| 4.2        | 検索アルゴリズム                      | 39        |
| 4.2.1      | シーンのスコアリング                    | 40        |
| 4.2.2      | リンク構造に基づくシーンスコアの再計算           | 40        |
| 4.2.3      | コンテンツのスコアリング                  | 41        |
| 4.3        | 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     | 42        |
| 4.4        | リンク構造の利用に関する考察                | 45        |
| <b>第5章</b> | <b>実験と考察</b>                  | <b>49</b> |
| 5.1        | アノテーションの収集                    | 49        |
| 5.1.1      | 実験方法                          | 49        |
| 5.1.2      | 実験結果と考察                       | 50        |
| 5.2        | 映像シーン検索                       | 54        |
| 5.2.1      | 実験方法                          | 54        |
| 5.2.2      | 実験結果と考察                       | 55        |
| 5.2.3      | まとめ                           | 60        |
| <b>第6章</b> | <b>関連研究</b>                   | <b>61</b> |
| 6.1        | ツールを用いた半自動アノテーションに関する研究       | 61        |
| 6.2        | マルチメディアコンテンツに対するアノテーションに関する研究 | 62        |
| <b>第7章</b> | <b>おわりに</b>                   | <b>65</b> |
| 7.1        | まとめ                           | 65        |
| 7.2        | 今後の課題と展望                      | 67        |
| 7.2.1      | 大量のデータに基づく提案手法の検証             | 67        |
| 7.2.2      | 映像シーンを検索する仕組みに関する評価           | 67        |
| 7.2.3      | プレイリスト作成に対する動機付けの向上           | 68        |
| 7.2.4      | 高度な言語解析等による意味的関係性の抽出          | 68        |
| 7.2.5      | 学習によるシーン関連度計算のための最適なパラメータ抽出   | 69        |
| 7.2.6      | 全シーンを対象としたシーンの重要度の計算          | 69        |
|            | <b>謝辞</b>                     | <b>71</b> |
|            | <b>参考文献</b>                   | <b>73</b> |



# 第1章 はじめに

近年、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コンテンツが存在するようになった。また、YouTube<sup>1</sup>などの動画共有サービスの出現によって、誰でも手軽にWeb上に映像コンテンツを公開することができるようになり、商用の映像コンテンツだけでなくアマチュアが作成した非商用の映像コンテンツも大量に公開されるようになった。これらの多種多様な映像コンテンツの増加とともに、検索や要約といった応用に対する要求が非常に高まっており、今後もさらに高まっていくと考えられる。

映像要約や映像シーン検索などの高度な応用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[11]。アノテーションとは、映像コンテンツに限らず、テキスト、画像、音声、音楽などのコンテンツ一般に対して適用可能な技術であり、これまで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[23]。

これまでの研究において、画像認識や音声認識技術を利用することで映像アノテーションを抽出する自動アノテーション手法[17][15][4]や、画像認識や音声認識の結果を利用しながら、さらに専任の人間が専用のツールを用いて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[12][22][9]が提案されている。

自動アノテーション手法には、人手を掛けることなく全自動で映像アノテーションを作成することができるという大きな利点がある。さらに、ニュース映像などの、音響や照明などの設備が整った環境で撮影され、ある程度型のきまった映像コンテンツに対しては、高精度に情報を抽出することが可能である。しかし、本研究で対象とするWeb上の映像コンテンツには、そのようなプロフェッショナルによって作成されたコンテンツだけでなく、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。そのようなコンテンツには、ノイズや手ぶれ、ピンボケなどが多く含まれるため、自動認識精度は一般にあまり高くない[3]。また、プロフェッショナルによって作成されたコンテンツでも、Web上にはスポーツ映像やドキュメンタリ映像などの多様なものが存在するため、抽出される情報に統一的な基準を持たないことから汎用性が低く、逆に単一の手法の導入は解析精度が低くなるという問題点がある。そのため、自動アノテーション手法をWeb上の映像コ

---

<sup>1</sup><http://youtube.com/>

コンテンツへ適用することは必ずしも有効ではないと考えられる。一方で、半自動アノテーション手法は、自動認識のミスを手によって修正し、詳細なアノテーションの作成を行うことができるが、膨大な数の映像コンテンツ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や、専門家の深い知識がなければ必ずしも質の高い映像アノテーションを作成することができないという問題点が存在する。

そこで、筆者らは Web 上で行われている自然なコミュニティ活動（以下では、Web コミュニティ活動と表記する）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ための手法を提案する。

近年、ブログや SNS（ソーシャルネットワーキングサービス）、SBM（ソーシャルブックマーク）、電子掲示板などのサービスが普及し、それらのサービスの利用者たちは共通の話題を持ち、Web 上で日常的にコミュニケーションを行っている。これらの活動全般のことを Web コミュニティ活動と呼んでいる。これらのサービスを利用する一般ユーザが発信する情報の中には、非常に意味のある情報も含まれ、時には社会に対して大きな影響を与えることもある。最近では国内最大の電子掲示板である 2ちゃんねる<sup>2</sup>や、国内最大の SNS である mixi<sup>3</sup>などに書き込まれた情報による社会への影響が最も典型的な例であろう。これらの Web サービス利用者の増加傾向は日本国内にとどまらず、世界最大の SNS である MySpace<sup>4</sup>の登録アカウント数は 2 億を超えている（2008 年 5 月現在）。また、Web 上で誰でも編集可能な百科事典である Wikipedia<sup>5</sup>は、情報の信頼性が比較的高く、また非常に広く深い情報を有する百科事典であると言え、世界中で利用されている。辞典を共同執筆するという活動も一種の Web コミュニティ活動であり、一般ユーザが発信する情報を大量に収集することで集合知と呼ばれる大きな利用価値のある情報となるという典型的な例である。

そして、このような Web コミュニティ活動が、映像コンテンツを話題の中心として行われるようになった。例えば、映像に対してコメントを投稿したり、映像をブログや SNS の日記に引用してその紹介文を書いたりする行為が頻繁に行われている。これらの活動によって蓄積される情報を映像アノテーションとして利用した場合、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の負担すべきコストが極めて小さいと考えることができるという利点や、内容や撮影環境の異なる映像コンテンツに対しても、映像の種類に依存しない統一的な仕組みを提供することにより汎用性の高いアノテーションを収集できるという利点がある。さらに、自由な情報記述が可能であるため、機械処理では解析することが困難な意味情報を含んでいる可能性がある。しかし、多くのサービスでは、映像コ

---

<sup>2</sup><http://www.2ch.net/>

<sup>3</sup><http://mixi.jp/>

<sup>4</sup><http://www.myspace.com/>

<sup>5</sup><http://www.wikipedia.org/>

ンテンツ全体に対するコメントの投稿や、ブログへの引用のみの機能しか実装されておらず、そこからは映像コンテンツの時間軸に対する情報を収集することができないという問題点がある。そのため、映像アノテーションとして利用する場合に映像シーン検索などの高度な応用を実現することは不可能である。

しかし近年、映像コンテンツの任意のタイムコードに対しての情報記述を行うことができるサービスも提供されるようになった。ニコニコ動画<sup>6</sup>では、動画の任意のタイムコードにコメントを投稿することが可能であり、そのコメントは動画上に表示される。ニコニコ動画には、1000 万人以上の会員登録者が存在（2008 年 11 月現在）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ったユーザたちが Web 上で互いにコミュニケーションをしたいという要求が十分にあるということを実証しているとも言える。しかし、ニコニコ動画で投稿されているコメントには、映像に直接もしくは間接的にも関係のない情報が大量に含まれるため、有用な映像アノテーションのみを選別して抽出することが非常に困難である。その原因として、コメント投稿の匿名性やユーザ層のコミュニティの偏りが考えられる。さらに、各コメントは、任意のタイムコードに対するものであり、より高度な応用を実現するためには、時間区間（以下では、シーンと表記する）に対する情報が必要である。

筆者の所属する研究室で開発された Synvie<sup>7</sup>は、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（以下では、シーン引用と表記する）などの情報から映像アノテーションを収集することが可能な動画共有サービスである [18]。Synvie は 2007 年 6 月から一般公開実験によるデータ収集を行っており、各機能から収集される映像アノテーションの解析や分析を行ってきた。筆者も公開当初から現在までサイトの運営に携わっている。Synvie で収集されるアノテーションには、映像に関係のない情報も含まれるが、それらを選別することによって映像シーン検索が可能であることを示してきた [10]。

しかし、従来手法では、映像コンテンツに対するアノテーションの網羅性に偏りがあるという問題や、アノテーションを自動的に選別する必要があるという問題が残されていた。具体的には、投稿されたテキストとその対象となるタイムコードやシーンが一对一对応で関連付けられるため、映像コンテンツに対して網羅的にテキスト情報を関連付けることが困難であった。その結果、アノテーションの網羅性が低くなり、検索の再現率が低いという問題があった。さらに、ユーザに記述される玉石混淆なテキストの中から検索に有用なキーワードのみを抽出する仕組みが十分でなかったため、検索の適合率を上げることも困難であった。

そこで、これらの問題を解決するために映像シーンの引用に関してより深く考察した。これまでに、Synvie で収集されたアノテーションの分析から、任意のタイムコードに対するコメントに含まれるテキストに比べ、シーン引用によって記

---

<sup>6</sup><http://www.nicovideo.jp/>

<sup>7</sup><http://video.nagao.nuie.nagoya-u.ac.jp/>

述されるテキストの方が良質なものが多く含まれるということも実証されている [18]。引用とは、紹介、参照、論評やその他の目的で自己の著作物中に他者の著作物の原則として一部を参照あるいは掲載することである [25]。これを映像シーンに適用し、シーン引用を定義すると、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、自身のブログなどの Web サイトで参照し、そのシーンに対する紹介や論評を記述することである。それによって、引用されたシーン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。

シーン引用を含む映像コンテンツの引用には、映像に対するハイパーリンクのみを掲載する形式と、ハイパーリンクに加え引用先のコンテンツ上で映像を再生可能にする形式が存在する。前者の場合、引用先のコンテンツから引用元のコンテンツの設置場所へのユーザの流れを誘発するため、引用元のコンテンツを保持している人、つまりサービス提供者にとって都合の良い形式である。一方、後者の場合、引用先のコンテンツ上で映像視聴が行われることによって、前者と比べると引用元のコンテンツの設置場所へのユーザの流れは減衰すると考えられる。しかし、前者よりも後者の形式の方が、引用先が明確である点や引用先の情報をコンテンツ上で閲覧可能になっている点で、引用者であるユーザにとって魅力的である。このように両者にはそれぞれ特徴があるが、アノテーションを収集するという観点においては、ユーザにとって魅力的なサービスを提供する後者の方が適切であると考えられる。また、映像コンテンツをコンテンツ単位で引用するか、シーン単位で引用するかという違いにも着目してみると、コンテンツ単位での引用は、記述されたテキストがコンテンツ中のどこを指しているのかわからない。一方、シーン単位で引用を行った場合は、引用元の該当箇所、つまり引用文の指し先（指示対象）を明確にすることが可能であるため、そのテキストをアノテーションとして再利用するためにはシーン単位での引用の方がより有効である。

Web 上では、テキストを引用したり、映像コンテンツをコンテンツ単位で引用するという行為は頻繁に行われるようになったが、映像シーンを効率良く引用する仕組みは未だ存在しない。これまでの Synvie に実装されていた映像シーンを引用する仕組みも、ユーザにとって非常に負担の大きいものであった。そこで、本研究では映像シーンをより手軽に引用できる仕組みを考案し、新しいユーザインタフェースを備えた映像シーン引用システムの開発を行った。さらに、シーン引用によって得られる映像アノテーションに加え、映像シーン間の関連度を計算し、さらにキーワードをより多くのシーンに関連付ける仕組みを組み合わせることによって、より効率良く映像シーンを検索するための仕組みを提案する。

本研究で開発した映像シーン引用システムでは、映像と同期して時系列順に並ぶサムネイル画像を参照することによって直感的に映像シーンを選択できる仕組みになっている。ユーザは、任意のシーンを選択し、そのシーンに対してテキストを記述し、それを HTML 形式のコンテンツ（主にブログ）に引用することがで



きる。引用されたシーンはその引用先のコンテンツ上で Adobe Flash Player<sup>8</sup>によって再生が可能である。また、複数の映像シーンを並べて引用することによって、それらのシーンを同期再生する機能を実装し、映像シーン間の関連付けを可能にした。

さらに、より手軽な仕組みによって映像シーン間の関連付けを行うために、引用されたシーンを利用した映像シーンプレイリストを作成するシステムを開発した。映像シーンプレイリストとは、引用されたシーンから任意の数のシーンを選択し、それらを連続的に再生する仕組みである。プレイリストによってコンテンツ間の関係を抽出するという研究はこれまでも行われている[5]。プレイリストの作成手順を複数用意することによって、その作成手順からプレイリストの作成意図を自動的に判別し、その属性によってアノテーションの利用方法を変化させる仕組みについても述べる。

そして本研究では、このような仕組みを提供することによって映像アノテーションを収集し、それらを解析して利用することで効率的に映像シーンを検索する仕組みを提案する。まず、シーンの引用時に記述されたテキストからタグを生成し、そのタグを利用して検索を行う。タグとはマルチメディアコンテンツの検索や分類を行うために付与されたキーワードのことである。タグを利用することで、複数のコンテンツをキーワードによって関連付けることが可能になる。タグを利用したサービスとしては写真共有サービス Flickr<sup>9</sup>やソーシャルブックマークサービス del.icio.us<sup>10</sup>などが世界的に利用されている。多くの動画共有サービスでも、映像コンテンツに対してタグが付与されるようになったが、それらは映像全体に対して付与されるものであるため、映像検索には非常に有効な手法であるが、映像中の任意のシーンを検索したいという要求には応えることができない。しかし、シーン引用によって収集されるアノテーションテキストからタグを生成することで、シーンに対してタグを付与することが可能である。そこで、本システムでは、シーンに関連付けられたタグ（以下では、シーンタグと表記する）によって検索を行う。検索結果としては、該当する映像コンテンツに関するタグクラウド、引用されたシーン区間を表す時間軸シークバー、シーンに関連付けられたコメント文などが表示される。これらの情報によって、シーンタグが関連付けられた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を支援することができる。

この映像シーン検索システムを使用して、これまでの映像シーン引用によるアノテーション手法と、提案手法である映像シーン引用とプレイリスト作成の組み合わせによるアノテーション手法の比較実験を行った。まず、一般公開実験と被験者実験によってシーン引用とプレイリスト作成のデータを収集した。そして、シー

---

<sup>8</sup><http://www.adobe.com/products/flashplayer/>

<sup>9</sup><http://www.flickr.com/>

<sup>10</sup><http://delicious.com/>

ン引用によって収集されたアノテーションから、シーンとキーワードの関連付けのみを行ったデータと、シーン引用とプレイリスト作成によって収集されるアノテーションから、シーンとキーワードの関連付けに加えシーンのリンク構造の解析を行ったデータの2つのアノテーションデータを用意した。そして、それぞれのアノテーションデータを利用してシーン検索を行い、適合率と再現率に基づく比較を行うことで、提案手法の有効性の検証を行った。

以下に本論文の構成を示す。第2章では、Webコミュニティ活動に基づく映像アノテーションについて詳細に述べる。第3章では、映像シーンを効率的に引用する仕組みと映像シーンプレイリストを作成する仕組みについて、またそれぞれの仕組みが利用されることによって抽出可能な映像アノテーションについて述べる。第4章では、収集される映像アノテーションを活用したシーン検索のアルゴリズムとユーザインタフェースについて述べる。第5章では、本研究における提案手法の有用性を検証した実験について述べる。第6章では、関連研究について述べる。そして、最後の第7章で、本論文の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

映像コンテンツを検索可能にしたり、その他の応用を実現するために、映像内容に関するメタ情報、またメタ情報を作成することを映像アノテーションと呼ぶ。映像アノテーションには、記述規格も提案されており、例えば、XML ベースの MPEG-7[20] が存在する。

アノテーションはコンテンツの検索や要約などの応用のために作成されるものであるが、対象となる映像コンテンツの種類や、応用などによって必要となるアノテーションが異なる。例えば、検索のためのアノテーションを作成する場合、ニュース映像とスポーツ映像では必要となる情報が異なる。ニュース映像はアナウンサーやナレーションの音声情報が最も重要であり、スポーツ映像では、実況や解説者の音声などに加え、選手の位置情報などの映像情報も非常に重要となる。Web 上には多種多様な映像コンテンツが存在し、最近では、アマチュアである一般人が自身の作成した映像コンテンツを手軽に公開できるようになったため、今後もそのような個人的なコンテンツが Web 上に氾濫していくと考えられる。それらのコンテンツはプロフェッショナルによって作成されるコンテンツとは撮影環境や機材などが異なることに加え、コンテンツの内容や構造にも制約がないため、それに対応したアノテーションの作成手法も必要である。

応用に関して述べると、映像シーン検索を実現する場合、検索の適合率や再現率はアノテーションの質と量に依存する。ニュース映像では、アナウンサーの発話文や字幕、画面中のオブジェクトなどに関するアノテーションが必要であり、スポーツ映像では、選手の名前やチーム名、選手などの位置情報やプレイ内容、実況内容などが必要である。また、映像要約を実現するには、検索のための情報に加え、さらに深い意味情報が必要であり、映像の任意の要素に関する意味的なアノテーションを必要とする。このほかの応用としては、コンテンツの推薦や統合などが考えられ、未だに考案されていないような映像コンテンツの利用方法も存在するであろう。このように、応用によっても必要とされるアノテーションが異なる。

また、アノテーションに関する重要な要素に、作成に要する人的コストがある。ある映像コンテンツに対して、十分に人手をかけて詳細なアノテーションを作成すれば、高度な検索や要約などの応用が比較的容易に実現できるであろう。しか

し、応用が実現できればそれで良いのではなく、そのために費やされる人的コストも考慮しなければいけない。

つまり、アノテーションを作成する際には、コンテンツの種類、コンテンツの利用目的、その精度、人的コストといった複数の要素を考慮する必要がある。それらのバランスをうまくとり、応用を十分な精度で実現できるようなアノテーション作成手法が求められている。これまでも、様々な手法によって映像アノテーションを作成する研究が行われてきた。しかし現在に至るまで、上記のバランスを十分に実用的なレベルで達成した手法は発見されていない。例えば、ニュース映像の検索に利用するアノテーションを作成するといったような、限られたコンテンツの特定の目的に特化した場合には、これらの複数の要素を考える必要はないように思われるが、近年の Web 上の映像コンテンツの増加傾向を考えると、コンテンツの属性や応用に幅広く対応できるアノテーション手法が必要である。

以上のような背景のもと、本研究では、Web コミュニティ活動によって蓄積される情報をアノテーションとして利用する手法に着目した。本章では、Web コミュニティ活動に基づいて収集することのできるアノテーションの特徴や、そのようなアノテーションを収集することのできる既存の研究、サービスとその特徴について述べる。さらに、本研究で特に注目した映像シーンの引用について詳細に述べ、従来の引用手法の問題点と本研究で解決すべき点について述べる。

## 2.1 映像コンテンツを話題の中心とした Web コミュニティ活動

近年、ブログや SNS、SBM、電子掲示板などのサービスが普及し、そこには共通の話題を持った人たちが集まり、コミュニケーションがなされている。これらの活動全般のことを Web コミュニティ活動と呼んでいる。そして、映像コンテンツを話題の中心とした Web コミュニティ活動とは、例えば、コンテンツに対してコメントを投稿したり、コンテンツを引用してブログや SNS の日記を執筆するといったユーザ活動の事である。具体的な研究やサービスの事例については次節で詳細に述べる。

本研究は、そのような Web コミュニティ活動における人間の行為から映像コンテンツに関する情報を抽出し、それをアノテーションとして利用する手法に着目した。例えば、映像コンテンツに対して投稿されたコメントに含まれるテキストを解析して関連付けることで、その映像コンテンツを検索可能にするといったことが最も簡易な例である。

この手法は、不特定多数の人間の知識を少しずつ顕在化し、それを大量に収集することで、大きな利用価値のあるアノテーションとして利用しようというアプローチである。この考え方は、映像コンテンツ以外のものに関してはすでに認知



され、実用化されている。例として、Web 上で誰でも編集可能な百科事典である Wikipedia が挙げられる。Wikipedia に記述されている情報は、非常に広範囲で、信頼性も比較的高い。全世界で 15 万人以上の人間が編集に携わっていることや、日本国内での利用者が 700 万人を超えていることからその信頼性の高さが表れている。このように不特定多数の一般ユーザの知識を少しずつ集め、その結果大きな知識を含む集合となった情報は集合知と呼ばれている。またそのようなユーザのことを、ユーザ一人一人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という場合において、大量の一般ユーザが行う分類が、少数の専門家が行うものよりも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。

Wikipedia のようなテキストコンテンツに加え、Google Image Labeler<sup>1</sup>では画像コンテンツに対するキーワードの関連付けの精度を、不特定多数のユーザの力を利用して高めている。Google Image Labeler は、ユーザが表示された画像に対してキーワード入力を行い、他のユーザの入力したものと一致すればそのユーザたちに得点が入り、その得点を競わせるといったサービスである。

このように、Web 上では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るようになり、その実用性が評価されているが、映像コンテンツに関しては、このような研究やサービスの事例が未だに多く存在せず、アノテーションの詳細な分析を行った例も少ない。しかし、画像やテキストに比べて高精度の意味情報の解析が困難なメディアである映像こそ Web コミュニティの力を利用してアノテーションを生成すべきであり、そのために効率的に良質なアノテーションを収集するための手法を考案する必要がある。

次に、Web コミュニティ活動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、インターネット上で行われる自然なユーザ活動から情報を獲得するため、その情報を発信している各個人は、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。このことは、アノテーションを作成するために個人が負担するコストが非常に小さいということである。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が増加するが、各個人が負担するコストは増加しない。また、映像の種類に依存しない統一的な仕組みを提供することにより汎用性の高いアノテーションを作成できるという利点もある。この手法では人手によってアノテーションが付与されるため、動画像認識ではその精度に大きく影響するような、画質や音質、撮影環境などにも影響を受けない。

さらに、もう 1 つの利点としては、コンテンツの解析からは決して抽出できないような情報をアノテーションとして収集できる可能性があるという点である。例えば、映像を引用したブログには、その映像そのものの内容だけではなく、それ

<sup>1</sup><http://images.google.com/imagelabeler/>

に関連した情報も記述される可能性が高い。ドラマを引用したブログの例を挙げると、そのドラマの出演者が以前に出演したドラマや、テレビ番組などについて述べられる可能性がある。そのことで、引用されたドラマとその他のドラマやテレビ番組との間に関連付けを行うことができる。このことはブログだけでなくコメントに関しても同様の可能性がある。そのような情報は専用ツールによって人間が付与することも可能であるかもしれないが、そのためには、映像の内容やそれに関連する情報について非常に詳細な知識が必要となり、1人の人間で付与することは困難であると考えられる。幅広く深い情報を映像アノテーションとして収集することができれば、検索や応用といった特定の応用だけでなく、その他の幅広い応用の実現可能性が広がる。

しかし、問題点もいくつか存在する。まず、アノテーションの質に関する問題である。個の手法によって収集されるアノテーションの中には応用の実現に適さないものも存在する。例えば、ブログやコメントに含まれる全てのテキストが、映像について、または映像に関連する情報について語られたものであるとは限らず、全く関係のない内容も含まれるということが十分に考えられる。そのためこの手法では、玉石混淆の情報の中から本当に利用価値のある情報とそうでない情報を選別する、もしくは応用によって利用する情報を変化させるような手法が必要になると考えられる。また、自然言語によって記述されたテキストを収集する場合には、それらのテキストを解析して、応用システムが利用可能な情報を抽出する必要もある。

次に、アノテーションの量に関する問題である。人手をかけずに質の高いアノテーションを収集できたとしてもその量が少なければ、映像シーン検索や要約などの高度な応用を実現することはできない。この手法では、アノテーションを作成していることを意識しながら行う活動からではなく、それを意識せずに行うコメントの投稿やブログの執筆などの自然な活動から映像アノテーションを収集する。そのため、情報の発信源である一人一人の人間が付与する情報量には制約がなく、その情報量は、多い場合もあれば極めて少ない場合もある。しかし、長時間の映像コンテンツに対する詳細な情報を1人で記述することを期待するのは非現実的なことであるため、複数人の人間によって各映像コンテンツに対する情報記述がなされる必要があると考えられる。また、限られた人間による情報のみからアノテーションを収集しても、この手法の利点があまり発揮されず、利用価値が低いものになってしまうと考えられる。こ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。多くの人間を集めるには、システムを自然に利用させる動機付けを与えなければならない。そのためには、まず、システム自体の使いやすさや、おもしろさ、話題性などが必要である。映像コンテンツを閲覧するためのページの見やすさやデザイン、ページ自体の軽さや検索のしやすさなどといったアノテーション行為と直接関係のない部分から、効率的にコメントを投稿できたり、手軽に映像を引用できたりするためのユーザインタフェース

などのアノテーション行為そのものに関わる部分まで、考慮すべき点が多い。また、映像を共有するシステムである場合は、映像コンテンツ自体のおもしろさも非常に重要な要素であるが、著作権や肖像権の問題が絡んでくるため運用には注意が必要である。

以上のことをまとめると、Web コミュニティ活動に基づく映像アノテーション手法の利点は次の通りである。

- 集団の力を利用してアノテーションを作成するため、個人の負担するコストが非常に小さい
- 動画像認識では解析できないような意味情報を含む可能性がある

また、問題点は次の通りである。

- 映像に関係のない情報をアノテーションとして付与してしまう可能性がある
- 十分な量のアノテーションを収集するためにユーザへの動機付けが必要である

これらの利点を活かしつつ問題点を解決するようなアノテーション手法を考案する必要がある。

## 2.2 映像全体に対するアノテーション

映像コンテンツに対して評価の入力やコメントの投稿をしたり、コンテンツをブログに引用したりといった機能が多くの動画共有サービスで提供されている。また、コンテンツ登録時に入力されるタグもコンテンツ全体を対象としたアノテーションである。これらの機能では、手軽に情報入力を行うことができるためユーザにとっての負担が少なく、比較的収集が容易なアノテーションであり、映像コンテンツそのものを扱う際には有用な情報になりうる。

しかし、これらの情報をそのままアノテーションとして利用する際、入力された情報の指し先は映像コンテンツ全体としてしか扱うことができない。そのため、映像コンテンツそのものを検索したり、推薦したりといった応用は実現可能であるが、映像シーン検索や要約のような高度な応用を実現することは不可能である。

だが、このようなアノテーションの中にもシーンに関する情報が含まれている可能性がある。もし映像中の一部分に対して言及している記述がされた場合、記述された情報が本当に映像全体を指しているのか、それともシーンを指しているのか、その場合どのシーンを指しているのかといったことを抽出することが可能になれば、その情報はシーンを扱った高度な応用にも有用なアノテーションに変

化しうる。しかし、コンテンツ全体を対象としたアノテーションのみでは、テキスト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、またほんの一瞬しかでてこないような情報が記述されるとそれは映像コンテンツそのものを検索するという用途に関して言えばノイズになる可能性もある。そのため、本研究では、その他のアノテーション手法と組み合わせることでそれらの情報をシーンに対して関連付けるということを行い、その情報を映像シーン検索にも利用し実験を行った。詳細は第3章以降で述べる。

## 2.3 映像中の部分に対するアノテーション

前節でも述べたとおり、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の種類と特徴について、複数のアノテーション機能を提供している筆者の所属する研究室で開発された Synvie を中心に述べる。

Synvie は、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（以下では、シーン引用と表記する）などの情報から映像アノテーションを収集することが可能な動画共有サービスである。Synvie は 2007 年 6 月から一般公開実験によるデータ収集を行っており、各機能から収集される映像アノテーションの解析や分析を行ってきた。筆者も公開当初から現在までサイトの運営に携わっている。本研究では、より効率的で良質なアノテーションの収集を目的として、Synvie をベースに開発を行った。

### 2.3.1 コメント投稿に基づくアノテーション

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として、コメント投稿に基づくアノテーションが挙げられる。Synvie では、映像コンテンツの閲覧中に任意のタイムコードに対するコメント投稿の仕組みを提供することによってアノテーションを収集している。例を図 2.1 に示す。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。コメントを共有することで、他ユーザとのコミュニケーションを図ることができ、それがユーザにとってのコメント投稿への動機付けとなる。

類似した機能を提供しているサービスにニコニコ動画がある。ニコニコ動画では、動画の任意のタイムコードにコメントを投稿することが可能であり、そのコメントは動画上にオーバーレイ表示される。ニコニコ動画には、1000 万人以上の会員登録者が存在（2008 年 11 月現在）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 Web





図 2.1: 映像中の任意のタイムコードに対するコメントの投稿

上で互いに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ということを実証しているとも言える。

このようなアノテーションの収集手法には2つの問題点がある。まず、アノテーションの対象が時間区間を持っていないということである。つまりアノテーションの指示対象は、シーンとは呼べず、あるシーン中のあるタイムコードでしかない。そのため、映像中のシーンを検索したい場合、その手がかりとしてタイムコードの情報を利用することができるが、シーンという単位で検索することができない。また、各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、統計的に複数のアノテーションを扱うことも難しい。この問題に対する対処を行っている研究に SceneNavi [24] があり、サービスの一般公開も行っている<sup>2</sup>。SceneNavi は、映像の視聴に同期した掲示板型のコミュニケーションや、非同期での掲示板型のコミュニケーションを行うことができる。特徴は、映像アーカイビングシステム SceneCabinet [27] を用いることで、映像をまとまった時間区間であるシーンへ分割しておき、各シーンに対してコミュニケーション空間を作り出すという点である。SceneCabinet は、映像からカット、テロップ、カメラの動き、音楽、人の声といった豊富なインデックスを、映像処理技術によって検出する。そのため、タイムコードではなく、シーンに対して閲覧者のコメントを関連付けることができる。もしこのシステムをプロフェッショナルの作成した商用コンテンツに対してのみ適用するのであれば、非常に有用な情報を収集できるかもしれないが、それを Web 上の多種多様なコンテンツに適用した場合には、シーンという区切りの信頼性に疑問が残る。意味のない区切り方がされたシーンに対し

<sup>2</sup><http://asp.super-space.jp/snv/Home.html>

でテキストアノテーションが関連付けられたとしても、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、シーンのセグメンテーションも不特定多数のユーザに頼るというアノテーション手法が、本研究の主要なテーマの一つである映像シーン引用である。映像シーン引用に関する詳細は次節で、本研究で提案する映像シーン引用の仕組みについては次章で詳細に述べる。

この形式のアノテーション手法に対する2つ目の問題点としては、前節でも述べたアノテーションの質の問題である。コメントの投稿は、映像を視聴しながら思ったことを入力して投稿するというスタイルで行われるため、「かっこいい」「すごい」などの印象語が多く含まれ、映像の内容そのものについて記述されることがあまり多くないと予想され、この傾向は他の動画サービスにも表れている。また、コメントの投稿は、他人のコンテンツに対して情報を追加するという行為であるため、自身には責任のない形で情報記述が行われることが多いと考えられる。そのため、不完全な文章であったり、コンテンツとは全く関係のない文章が入力されることが多くなる。

アノテーションの手軽さは、Web コミュニティ活動に基づくアノテーション収集手法における考慮すべき大きな要素の1つであるため、その点においてコメント投稿によるアノテーションは有効である。だが手軽さを優先するだけで結果として多くのノイズを含んでしまうのでは、有用なアノテーション手法であると言えない。そのため、玉石混淆なアノテーションの中から良質なものを選別するための手法を組み合わせることによって有用なアノテーションとして利用可能な形にする必要がある。

コメント投稿の機能を拡張したものに、任意のタイムコードの任意の画像領域に対するコメント投稿からアノテーションを収集するというものがある。具体的に、Synvie では対象となるタイムコードの静止画像に対して、マウスで矩形範囲を選択した後にコメントを記述する。北山らは時間区間と画像領域を指定したコメント投稿を行うシステムを提案し、コメントを利用した視聴支援方法についても提案を行っている [7]。この手法では、アノテーションテキストの指し先をより明確にすることができるという利点がある。しかし、画像領域を選択する際には、映像を一時停止しなければならないため、コメント投稿によるアノテーションに比べ操作が煩雑であることに加え、収集されるテキストはコメント投稿によるアノテーションと同様の特徴を持つと考えられる。北山らの研究でも、テキストを解析して検索に利用するといったことは行っていない。

コメント投稿と同等の手軽さで、より映像内容に近い情報を収集するための手法として、ユーザによって協調的にタグ付けをさせるという研究も行われている。検索や分類のためのキーワードを投稿させることによって、直接的に映像に関係のあるものを集めようという手法である。筆者らはソーシャルブックマークの機能を映像シーンに適用することでアノテーションを収集する手法の提案もこれまでにやってきた [26]。しかし、ユーザに対してキーワードでの入力を強制するも

のではないため、コメントと同様の特徴を持つテキストが記述されることも多い。また、タグ付けという活動は、コメント投稿に比べユーザに対する動機付けが弱いと推測される。2006年11月から Ask ビデオ<sup>3</sup>ではシーンに対するタグの投稿機能が提供されていたが、数万人のユーザを抱えているにも関わらず現在はその機能の提供を中止している。

### 2.3.2 ボタンクリックに基づくアノテーション

次に、ボタンクリックに基づくアノテーションについて述べる。Synvie では任意のタイムコードに対する印象を「Nice」か「Boo」で評価するボタンと、次回の視聴やブログ引用の際の目印をつけておくボタンがある。このような映像に対する印象や目印をクリックで入力するという機能は多くの動画共有サービスでも提供されている。YouTube では、コンテンツに対する 5 段階評価をマウスクリックで行う。YouTube の機能はコンテンツ全体に対する情報であるという点で Synvie とは異なるが、非常に多くのユーザに利用されている機能である。ボタンクリックは、コメント投稿よりもさらに手軽であり、映像の人気度などの評価情報を収集するという点では非常に優れた手法であり、アノテーションを集計してユーザにフィードバックすることでユーザが閲覧するコンテンツを選ぶ手がかりとなる。また、ボタンクリックによるアノテーションは、それだけではコンテンツの内容情報をアノテーションとして収集する手段にならないが、他の手法によって収集されたテキスト情報を利用することで内容情報をコンテンツに関連付ける手法にもなりうる。筆者らは、コメントなどのアノテーションテキストから生成したタグクラウドをユーザに提示し、そのタグをクリックさせることによってタグをタイムコードに関連付ける手法の提案も行っている [19]。

### 2.3.3 引用に基づくアノテーション

引用に基づくアノテーションとは、映像コンテンツ全体もしくは一部がブログなどの Web ページで参照され、そのシーンに対してコメント文が記述されることによって蓄積されるアノテーションである。Synvie では、映像の任意のシーンをブログへ引用することができる。映像シーンが引用されると、セグメンテーションとテキスト記述の複数のアノテーション収集が可能である。映像シーンの引用については次節で詳細に述べる。

---

<sup>3</sup><http://video.ask.jp/>

### 2.3.4 映像編集に基づくアノテーション

セグメンテーションの情報を収集するために優れていると考えられるアノテーションの収集方法として、オンラインでの映像編集がある。motionbox<sup>4</sup>やjumpcut<sup>5</sup>、sprasia<sup>6</sup>などでは、動画共有サービスに加えて、Web ブラウザ上で動画を編集する機能を提供している。しかし、映像編集は、ユーザにとって負担が大きいという問題点や、映像編集によって生成されるシーンに対してテキスト情報を収集することが困難であるという問題点がある。

## 2.4 映像シーンの引用とは

Web コミュニティ活動に基づくアノテーション手法には、前節で述べたように多様なものが存在する。その中でも本研究では、映像シーンの引用に着目し、従来の引用手法を改良すると共に新たなアノテーション手法を組み合わせることで、効率的に良質なアノテーションを収集する仕組みを提案する。本節では、映像シーンの引用によって収集可能なアノテーションの特徴について述べる。

映像シーンの引用を定義すると、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。

映像シーン引用のアノテーション手法としての利点は、テキストによるアノテーションを収集することができるだけでなく、セグメンテーションとしてのアノテーションも蓄積することができるという点である。コメント投稿やボタンクリックによるアノテーションでは、特定のタイムコードもしくはあらかじめ定義されたセグメントに対して情報が追加されていくだけであるが、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。これによって映像コンテンツをシーン単位で扱うことが可能になり、高度な応用の実現の可能性が生まれる。また、引用の際に定義されるシーンは、人手によって定義されるため、何らかの意味をもったまとまりである可能性が高いと言える。

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あると考えられる。映像シーン引用は、ユーザにとって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。そのため、コメントのような他人のコンテンツに対して行われる活動に比べて、質の高い情報記述や意味のあるセグメンテーションが

---

<sup>4</sup><http://www.motionbox.com/>

<sup>5</sup><http://jumpcut.com/>

<sup>6</sup><http://www.sprasia.com/>



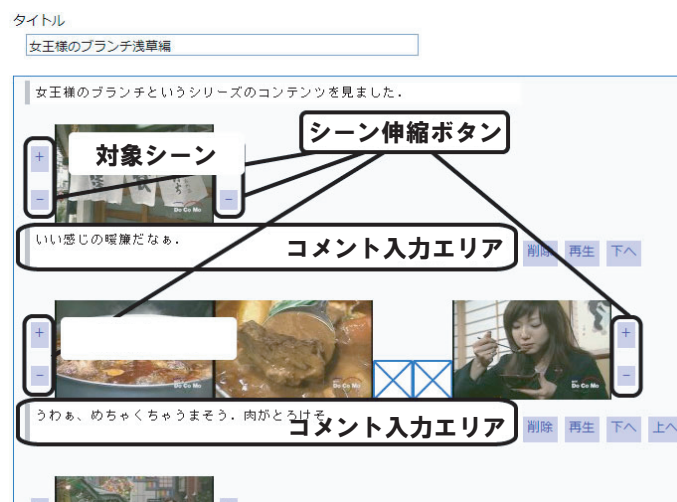


図 2.2: 従来手法における映像シーン引用のユーザインタフェース

なされるといった期待ができる。Synvie におけるこれまでのアノテーションの分析からも、コメント投稿によるアノテーションよりも引用によるアノテーションのほうが良質なテキストが多く含まれるということも実証されている。

このように、シーン引用はコメント投稿にセグメンテーションの要素を加え、さらに良質な情報を収集することが期待できるアノテーション収集手法であると考えられる。Synvie では、映像シーンを引用す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。しかし、従来の引用手法にはいくつかの問題点があり、映像シーン検索などの応用を高精度で実現することが困難であった。そのため、本研究ではその問題点を明らかにし、それを解決する手法の提案を行う。

## 2.5 従来の引用手法の問題点

従来の Synvie における映像シーン引用のユーザインタフェースが図 2.2 である。ユーザが映像コンテンツを選択して引用を開始すると、シーンを引用してブログの執筆を行うためのページが開く。そのページには、コンテンツ視聴時にコメント投稿またはボタンクリックを行ったタイムコードのサムネイル画像が並ぶ。そして、シーン伸縮ボタンをクリックして引用するシーンの開始時間と終了時間を選択することで、引用シーンの時間区間を決定する。そしてそのシーンに対するコメントを編集する。シーンとテキストの編集を終えると HTML 形式のテキストが生成され、それをブログサービスに投稿することができる。

この手法の問題点はまず、シーンを引用するためのユーザインタフェースと引用によって作成されるコンテンツ (これをシーン引用ブログと呼ぶ) にある。



図 2.3: 従来手法における映像シーンを引用したブログ

シーンを引用するためのユーザインタフェースにおける問題は、シーンの時間区間をサムネイル画像のみで決定するという点である。サムネイル画像を用いることによってシーンの開始時間と終了時間を明確に決定できる点は有効であるが、それだけでは音声の情報が除去されており、また映像中の細かな動きを閲覧することができない。また、ボタンクリックによってシーンの時間区間の伸縮を行う点にも問題がある。数秒程度のシーンを設定するためには特に不便ではないが、数十秒やそれ以上の時間区間をシーンとして引用したい場合には非常に効率が悪い。例えば、1分のシーンを引用したい場合はボタンを30回クリックする必要がある。

次に、図 2.3 が従来手法によって生成されるシーン引用ブログの例である。引用されたシーンがサムネイル画像で表現され、画像にはハイパーリンクが関連付けられており、クリックするとそのシーンを Synvie のサイト内で閲覧することができる。この引用形式の場合、ブログ閲覧者を Synvie へ誘導できるため、Synvie へのユーザの流れを増やすことができる。しかし、ブログ閲覧者にとってはブログ内でそのシーンを閲覧できることが理想的であると考えられ、アノテーションをより多く収集するためにはより魅力的な引用コンテンツを作成可能にする必要がある。また、Synvie のページに飛ばないとシーンが閲覧できないということによってブログのコンテンツとしての魅力が下がることで、その結果ブログ閲覧者が減り、結局は Synvie へのユーザの流れも作り出せない可能性もある。

このように従来のシーン引用手法は、システムを利用するユーザにとっての利便性と魅力に欠ける。そこで、本研究では、これらの問題を解決するために、シーンをより効率良く引用し、より魅力的なシーン引用ブログを作成できる仕組みを

提供することによって、アノテーションを収集するためのシステムの開発を行った。詳細は次節にて述べる。

以上の問題点に加え、従来手法には、収集可能なアノテーションにも問題がある。これまでに、従来手法で収集されたアノテーションを利用して映像シーン検索を行うという実験を行った [10] が、いくつかの課題が残されていた。それは、シーン検索にとって有用なアノテーションとそうでないものを機械的に選別する必要があるという課題と、コンテンツに対するアノテーションの網羅性を高め、シーンに対するアノテーションの量を効率的に増やすことができないという課題である。

これらの問題を解決するためには、まず、引用に基づくアノテーション間の関連度を計算し、それを検索結果に反映させることが有効である。アノテーション間とは、引用されたシーン間、テキストとシーン間などが挙げられる。例えば、あるシーンとあるシーンはどの程度の関連性があるか、またどのような属性の関係であるかを計算することによって、あるシーンに対して新たに情報が追加された場合に、そのシーンと関連の強いシーンに対してもその情報を関連付けることが可能になる。それによって、各アノテーションの対象範囲を広げることができると考えられる。また、テキストとシーンの間の関連度計算は、シーンに対する適切なキーワードを選択することに貢献する。

本研究では、従来の映像シーン引用の仕組みに加え、引用されたシーンを連続的に再生するプレイリスト作成の仕組みを組み合わせることによって各アノテーション間の関連度計算を可能にし、より効率的な映像シーン検索を実現するための手法の提案を行う。



## 第3章 映像シーンの引用とプレイリストの作成に基づくアノテーション

本研究では、複数の映像コンテンツの映像シーンを手軽に引用することができるシステムを開発し、さらに引用されたシーンを複数集めたプレイリストを作成できるシステムを開発した。本章では、それぞれのシステムの詳細について、また、それぞれのシステムが利用されることによって収集される情報から抽出可能な映像アノテーションについて述べる。

### 3.1 映像シーンの引用に基づくアノテーション

本研究では、映像シーンのブログ記事への引用を支援する仕組みを提供し、その仕組みを利用したユーザの詳細な編集履歴を蓄積することによって、ブログ記事の文章構造と映像のシーン構造とを関連付けたアノテーションの抽出法を提案する。本研究では、映像シーンを引用したブログのことを映像シーン引用ブログと呼ぶ。

映像シーンの引用先としてブログ記事を採用している理由は、ブログは、記事ごとにパーマリンク (Permalink) [1] やトラックバック (Trackback) [21] などの仕組みが実装されているため、引用先の情報を手軽に収集することが可能であるからである。また、映像コンテンツをブログに引用して紹介記事を書くという活動はすでに広く認知されているため、新しいユーザ活動の仕組みである映像シーン引用も、引用先をブログ記事にすることによってユーザにとって受け入れられやすくなると考えられる。

本節では、本研究で開発した映像シーン引用ブログの作成を支援するシステムについて、ユーザインタフェースやブログ執筆手順、また、その編集履歴から収集可能なアノテーションについて述べる。また、本システムで作成される映像シーン引用ブログについても述べる。

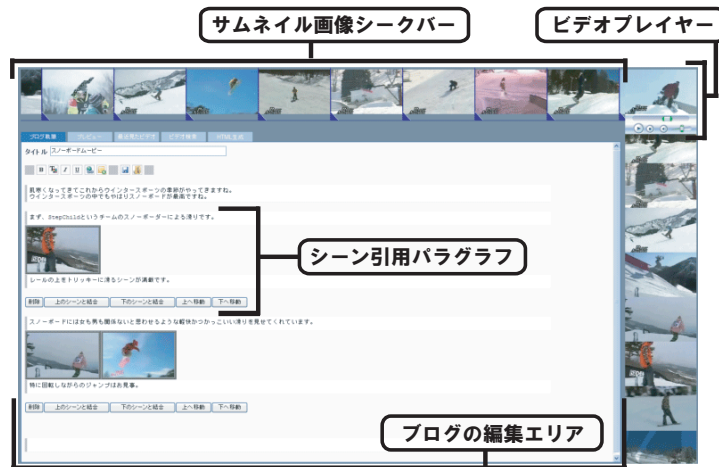


図 3.1: 映像シーンを引用するユーザインタフェース

### 3.1.1 引用区間の選択によるセグメンテーション

#### サムネイル画像シークバーの生成

図 3.1 に映像シーン引用ブログを執筆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 3.1 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、ユーザが視聴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を提供している。また、映像の視聴中に図 3.1 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、上から下へ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を移動し、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引用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、ビデオ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いる。





図 3.2: 引用するシーン区間の選択

### 引用するシーン区間の選択

サムネイル画像の上でマウสด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、タイムラグ無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。しかし、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、効率良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 3.2）。1 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、右クリックによって表示されるポップアップウィンドウ上のプレイヤーでプレビュー再生を行う。そして、シーン区間の調整とプレビューを終えたら、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を決定したら、右クリックメニューから、映像シーンを引用したパラグラフをブログ執筆エリアに追加する。ブログ執筆エリアでは、ユーザが設定した代表フレームに対応するサムネイル画像によって映像シーンが表現される。このサムネイル画像をクリックすると右上のビデオプレイヤーが引用されているシーンの開始時間から再生を始めるため、コンテンツ中のどこを引用したかをいつでも確認することができる。

映像シーンが引用されることによって、シーンというセグメント情報が映像コ



図 3.3: 引用されたシーンの再引用

コンテンツに対して関連付けられる。ユーザが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、セグメンテーションが行われることによって、シーンというセグメントに対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。

### 被引用シーンの再利用

ユーザは引用するシーン区間の詳細を自身で設定できるだけでなく、すでに引用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 3.3 のように表示され、クリックすることでそのシーンをプレビュー再生することができる。さらに、そのシーンに対して記述されたテキスト、引用先のブログへのハイパーリンクを表示することによって、シーンの内容を理解する手掛かりとなるだけでなく、そのシーンを話題の中心とした議論が行われる可能性もある。

引用された区間を再利用可能にすることによって、ユーザのシーン引用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。全く同じ対象を複数のユーザが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というような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



図 3.4: 複数の映像コンテンツからの引用

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、同一の対象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、シーンの再利用を可能にすることによって、他のユーザと同じ個所を引用したいという意図で微妙にずれたシーンを選択してしまうというのを防ぐことができ、各シーンに対する情報量を増やしていくことができる。

### 複数の映像コンテンツからの引用

従来のシーン引用の仕組みでは、単一の映像コンテンツからのシーン引用のみを提供していたが、本システムでは複数の映像コンテンツの中のシーンを同一のブログ記事に引用することができる。

複数の映像コンテンツからシーンを引用するために、引用インターフェースは、コンテンツを新たに読み込むだけでなく、最初に読み込んだコンテンツと入れ替える、もしくは新たに追加して読み込むという機能を有する。新たに追加した場合は、最初に読み込んだものの左下に、同様の形式でコンテンツが読み込まれる（図 3.4）。最大で3つのコンテンツまで同時に読み込むことが可能である。このようにして複数のコンテンツを並べて読み込むことで、映像内容を見比べながら引用するシーンを探すことが可能である。

複数の映像コンテンツに含まれるシーンが同一のブログ記事に引用されることによって、シーン間だけでなくコンテンツ間にも関連付けが行われる。

### 3.1.2 共引用による複数シーン間の関連付け

従来のシーン引用の仕組みでは、引用されたシーンは、そのシーンのみを閲覧する仕組みであったが、本研究では、複数のシーンを同一パラグラフに引用することで、同期的に再生する仕組みを提供している。シーンを比較したいという意図や、シーンを並べて再生することでより魅力的な映像を表示したいという意図によって、共引用が行われると考えられる。これまで、複数のコンテンツやシーンを同時に再生するためには、映像編集ソフトを利用してコンテンツ自体を編集するか、複数のプレイヤーを JavaScript などのプログラムによって操作するという手段しかなかった。しかし、本システムでは1つのプレイヤーの中で複数のシーンを同時に再生することができるため、手軽にシーンを並べた再生を行うことができる。

共引用は、マウスクリックによってシーンを引用したパラグラフを結合するという手軽な操作で行うことができる。また、共引用を行う場合、それぞれのシーンの再生時間を合わせたいという要求が考えられるため、シーンの設定時には、サムネイル画像左隅に設定中のシーンの総再生時間を表示している。

同一のブログ記事に対して引用されたシーンの間には、ブログ全体の大きなテーマに関して関連性があると言えるが、シーンの深い内容に関する関連性は低いと考えられる。しかし、複数のシーンが同一のパラグラフに共引用された場合、それらのシーンには映像内容に関する何らかの強い共通点があるという理由で引用されたと考えられるため、より関連性が高いと言える。

### 3.1.3 ブログ記事の編集によるテキストアノテーション

パラグラフには、サムネイル画像の上下にテキストエリアが用意されており、シーンに対するコメントをそのテキストエリアに記述することによってパラグラフの編集を行う。パラグラフの編集画面を図3.5に示す。パラグラフに記述されるテキストは、そのパラグラフに引用されているシーンに対して関連付けられる。ここで記述されるテキストには、解説や、描写、感想などといった、引用シーンに対して関連のあるテキストが含まれると考えられる。さらに、ブログという新しいコンテンツ制作活動のために記述されるテキストであるということや、制約のない自然言語による自由な記述を許すことから、映像シーンに関する詳細なテキスト情報を収集できる可能性がある。また、情報が映像の品質や種類に依存せず、画像処理や音声処理によっては解析することができないような情報が含まれる可能性がある。また、本システムでは、プレーンテキストを記述するだけでなく、BタグやIタグなどのHTMLタグをボタンクリックで挿入可能なエディタを提供している。それらのタグで囲まれたキーワードは、そのパラグラフ中の代表的なキーワードを抽出するための指標になると考えられる。各パラグラフに対するテキス

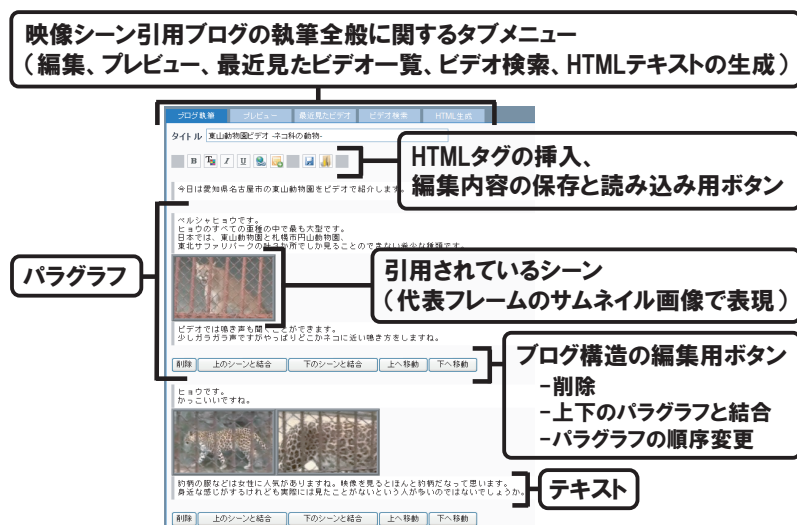


図 3.5: シーン引用パラグラフの編集

トを編集し、パラグラフの順番の入れ替えなどを行うことでブログ記事の編集を完了する。

### 3.1.4 映像シーン引用ブログ

映像シーン引用ブログは、映像コンテンツにおける任意の時間区間である映像シーンと、そのシーンを代表する画像、およびシーンに対して言及したテキストを含むパラグラフの集合によって構成される（図 3.6）。

図 3.1 のような仕組みで映像シーン引用ブログの執筆を終えたら、HTML 形式のテキストが自動生成され、ユーザはそれを既存のブログサービスに貼り付けて投稿することができる。ブログ中では、各パラグラフに対して Adobe Flash Player オブジェクトが埋め込まれる<sup>1</sup>。ブログページの読み込み時には、各プレイヤーには、代表フレームに対応するサムネイル画像が表示されている。そして、プレイヤーにマウスカーソルを置くと再生アイコンが表示され、クリックすると映像シーンが再生される。

従来のシーン引用の仕組みでは、引用されたシーンはサムネイル画像のみで表現され、画像にはシーンを Synvie 内で再生するためのハイパーリンクが関連付けられていた。しかし、この映像シーン引用ブログでは、ブログ上でシーンの再生ができるため、ユーザは効率的にシーンを閲覧することができる。シーンの再生

<sup>1</sup>一部のブログサービスでは、Adobe Flash Player の埋め込みに非対応である。MovableType、Seesaa ブログ、FC2 ブログ、livedoor ブログ、LOVELOG では埋め込みに対応していることを調査済みである（2009 年 1 月現在）。



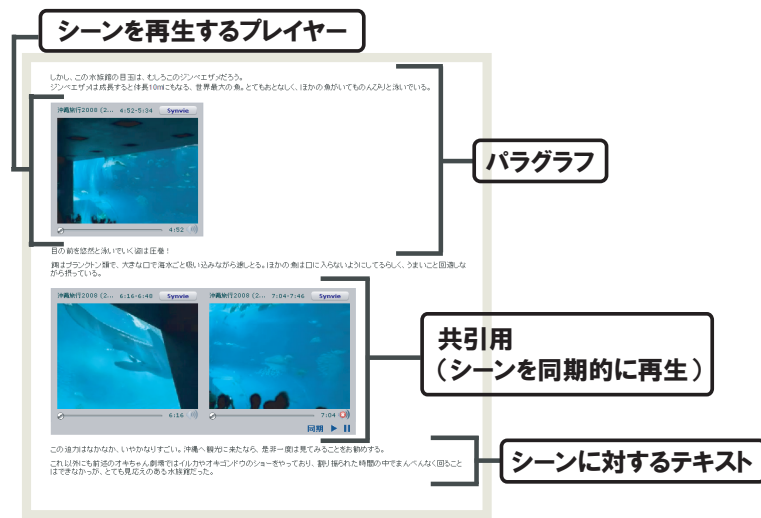


図 3.6: 映像シーン引用ブログ

がブログ記事内で完結してしまうことで、コンテンツ元である Synvie にユーザを誘導しにくくなるという問題点があるが、プレイヤーに Synvie へのハイパーリンクを設置することで、オリジナルコンテンツへのユーザの興味を喚起する対応を行っている。また、ハイパーリンクではないシーン再生が可能になることで、ブログ記事そのもののコンテンツの価値が上がり、シーン引用の活動そのものに対する動機付けにつながり、結果として Synvie へのユーザの流れも作りだすことができると考えている。また、シーンの再生ログを蓄積することで、そのシーンやブログ記事の重要度を計算するための指標に利用することができる可能性がある。

### 3.1.5 映像シーンへの自動トラックバック

本システムで生成される映像シーン引用ブログの HTML には、公開されたブログが閲覧された際にその情報をサーバに送るためのタグが挿入されている。そのため、映像シーン引用ブログが公開され、その記事が閲覧されると、その引用元である映像シーンに対して引用先であるブログパラグラフが自動でトラックバックされる。トラックバックのリンクが自動的に追加されることで、映像シーン引用ブログが公開されたことをシステムが確認することができる。

トラックバックがなされると、その情報は共有され、Synvie 上で映像シーンが視聴された際に、映像下に引用先のパラグラフへのハイパーリンクとそのパラグラフに書かれているテキストの一部が表示される。この仕組みによって、映像コンテンツを視聴したユーザが、映像シーン引用ブログを容易に発見できる。また、トラックバックを自動で行うことで、引用された映像コンテンツの投稿者が、自



分のコンテンツのシーンが引用されたことを知ることができる。

## 3.2 映像シーンプレイリストに基づくアノテーション

映像シーンが引用されることによって、シーンという単位の意味のあるセグメント情報が蓄積される。また、シーンに対して言及されたテキストが関連付けられることによって、それらのシーンが検索可能になる。しかし、引用に基づく情報のみでは、テキスト中に含まれるキーワードに対する重みが一律であり、それらを機械的に選別するのは困難である。そのため、シーンを検索するために適切なキーワードとそうでないキーワードの両方が同一の重みで混在することになる。また、ユーザ活動から得られる情報量には偏りが見られるため、アノテーションの網羅性が低く、各シーンに対して関連付けられるテキストアノテーションの量にも偏りが存在すると考えられる。

そこで、シーンを連続的に再生するシーンプレイリストを作成する仕組みをユーザに提供し、その仕組みを利用したユーザの詳細な編集履歴を蓄積することによって問題の解決を図る。この仕組みによって、コンテンツに対するアノテーションの網羅性を高め、さらにそれぞれのシーンに対するキーワードの重みづけを行うことで、シーン検索の適合率と再現率を向上させることができると考えられる。本研究では、映像シーンを連続的に再生する仕組みを映像シーンプレイリストと呼ぶ。

本節では、本研究で開発した映像シーンプレイリストの作成を支援するシステムについて、ユーザインタフェースや作成手順、また、その編集履歴から収集可能なアノテーションについて述べる。

### 3.2.1 プレイリストに利用するシーンの選択

本システムでは、プレイリストを作成するためだけにシーンを作成し、それを利用するという仕組みではなく、過去にユーザによって引用されたシーンのみを再利用してプレイリストを作成するという仕組みを提供する。映像シーンを引用する仕組みと映像シーンプレイリストを作成する仕組みをそれぞれ独立に提供しても、従来手法によるアノテーションと同等のものを複数の方法で収集するだけになってしまうと考えられる。しかし、本研究で提案するのは、引用とプレイリストのそれぞれの編集履歴を密に組み合わせることによって、アノテーションの量と質を共に向上させていくアプローチである。つまり、引用によって蓄積された情報がプレイリスト作成のために再利用されることで、それらの情報の選別と拡充が行われ、結果として良質で網羅性の高い映像アノテーションを蓄積していく仕組みを提案する。

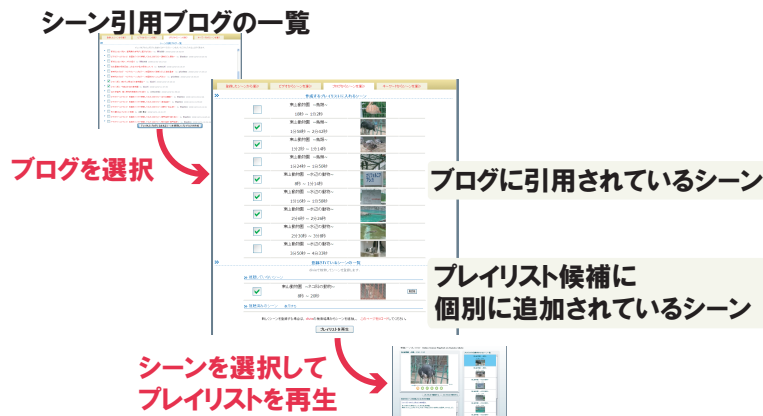


図 3.7: プレイリストに利用するシーンの選択

プレイリストを作成するために、まずプレイリストに利用するシーンの候補を選択する。シーンをプレイリスト候補に追加するためには、1シーンずつ候補に追加する方法と複数シーンを同時に候補に追加する方法がある。

1シーンずつ候補に追加するためには、キーワード検索によってシーンを探し、その検索結果上でシーンの追加を行う。この検索には本研究で開発した映像シーン検索システム [10] を用いる。検索のインターフェース等に関する詳細は次章で述べる。

また、以下の3種類の方法で複数のシーンを同時にプレイリスト候補に追加することができる。

- 映像コンテンツを基にシーンを選択

任意の数の映像コンテンツを選択することで、それらコンテンツから引用されている全てのシーンがプレイリスト候補に追加される。

- 映像シーン引用ブログを基にシーンを選択

任意の数のブログを選択することで、それらのブログ中に引用されている全てのシーンがプレイリスト候補に追加される。

- キーワードを基にシーンを選択

任意の数のキーワードを選択することで、そのキーワードが関連付けられている全てのシーンがプレイリスト候補に追加される。

これらの方法でプレイリスト候補に追加されたシーンと、1シーンずつ追加されたものは併用することが可能である。このようにして複数のシーンを同時にプレイリスト候補に追加することで、効率的にプレイリストを作成できる。映像シー



図 3.8: プレイリストの再生と編集

ン引用ブログを手がかりとしてプレイリストに利用するシーンを選択する例を図 3.7 に示す。

また、どの情報を手がかりにプレイリスト候補を決定したかという履歴を蓄積することによってプレイリストに対して属性を付与することができる。これらの属性はアノテーションを解析する際に利用可能である。詳細は次節にて述べる。

以上のようにして、プレイリスト候補を追加したら、さらにその中からプレイリストに利用するシーンを選択し、プレイリストの再生画面を開きプレビュー再生を行いながら編集を行う。

### 3.2.2 プレイリストの編集

プレイリストの再生画面例が図 3.8 である。再生画面を開くとプレイリストの再生が自動的に開始され、プレイリストに含まれるシーンが連続的に再生される。図 3.8 のように、左上にプレイヤーが設置され、それぞれのシーンが再生される。左下には、そのシーンの引用先であるブログのタイトルとシーンに対して記述されたテキストが表示され、そのブログへのハイパーリンクが関連付けられている。また、右にはプレイリストに含まれているシーンが上から順番にサムネイル画像とタイトルによってリスト表示されている。再生中のシーンはハイライト表示され、それぞれのシーンをクリックするとそのシーンから再生することができる。

この画面では、プレイリストの再生だけでなく、編集も行うことができる。プレイヤーの下に配置されている編集ボタンをクリックすると、プレイリスト編集用のポップアップウィンドウが表示され、各シーンの順番の入れ替えは、ドラッグ&ドロップによる直感的な操作によって行うことができる(図 3.8 右)。また、ボタンクリックによってシーンを削除することができる。ポップアップを閉じると

編集内容が反映され、再びプレイリストが再生される。このようにして編集と再生を繰り返すことでプレイリストの編集を行う。

これまでの一連の編集履歴を保存し、プレイリスト候補のシーン集合と最終的に残ったシーン集合の差分を見ることによって、シーン間の関連度を測る指標にもなると考えられる。

### 3.2.3 プレイリストの共有と再利用

シーンの順番の入れ替えや削除を終えたら、プレイヤーの下に配置されている保存ボタンをクリックして、プレイリスト保存用のポップアップを表示させる。ポップアップ上で、タイトルとタグ、コメントの入力を行い、プレイリストを共有するかどうかを選択し、保存する。プレイリストを共有するという選択がされた場合、そのプレイリストは一覧ページに追加され、他のユーザも閲覧することができる。もし共有しない場合は、自身のページにのみ表示される。

プレイリストを保存するという行為は、そのプレイリストを再び閲覧したい、もしくは他のユーザと共有したいという意図で行われると考えられる。もし共有しないという選択がされた場合は、そのプレイリストは自身のためだけに作成されたものであり、それは単純にそのユーザの嗜好に適したシーンの集合である可能性が高い。しかし、プレイリストを共有するという選択がなされた場合、自身で閲覧するだけでなく他ユーザにも閲覧されることを意図しているため、そのプレイリストに含まれるシーンは何らかのテーマや意図に沿ったシーンの集合である可能性が高い。そのため、共有されたプレイリストに含まれているシーン間に対して関連付けを行うことができる。

さらに、ユーザは共有されているプレイリストを閲覧できるだけでなく、プレイリストに含まれるシーンを利用して派生のプレイリストを作成することや、シーンをブログへ引用することができる。このようにして、プレイリストからもシーンの再利用を可能にすることによって、引用やプレイリスト作成に対する手間を減らし、映像シーンを話題としたユーザ活動の活性化を図ることができると考えられる。

## 3.3 アノテーションの解析

本システムでは、それぞれのユーザ活動から収集される情報をなるべく情報劣化がない形式でアノテーションとして蓄積する。

映像シーンが引用されることで、シーンというセグメント情報、さらにシーンに対して記述されたテキストがアノテーションとして蓄積され、また、共引用によって複数シーンの間に対する関連付けが行われる。また、映像シーンプレイリ

ストが作成されることで、プレイリストに含まれるシーンの間に関連付けが行われることに加え、プレイリストに対するタイトルやタグなどのテキスト情報などがアノテーションとして蓄積される。しかし、これらのアノテーションは、そのままの形で応用に利用しようとしても、アノテーションに内在する深い情報を機械が理解することができず、引用とプレイリストの履歴を十分に活かすことができない。そのため、アノテーションを解析することによって、より利用価値の高い情報に変換する必要がある。

まず、アノテーションのテキスト情報からコンテンツの意味内容を表す情報の抽出を行う。具体的には、映像シーンの内容を表現するキーワード（一般にタグと呼ばれる）の抽出を行う。次に、プレイリストに対する代表的なキーワードを抽出し、プレイリスト中のシーンに関連付ける。また、引用やプレイリスト作成に基づいて関連付けが行われたシーンの集合からリンク構造を推測する。

### 3.3.1 シーンに対するタグの生成

映像シーンの引用によってそれぞれのシーンに対して記述されたテキストから、キーワードの抽出を行う。この解析によって生成される、シーンに対して関連付けられたキーワードのことをシーンタグと呼ぶ。まず、日本語係り受け解析器 Cabocha[8] を利用してテキストを形態素に分割する。次に、それぞれの形態素から、名詞、動詞、形容詞、固有名詞、未知語を抽出する。さらに、一般的に不要語と判断可能な形態素（例えば、する、ある、なる、できる等）を事前に作成した不要語辞書を利用して除外する。そして、本研究で作成した連結ルールによって連続する名詞等の連結を行う。この処理によって出力されたキーワードをシーンに関連付けてデータベースに保存することでシーンタグが生成される。本システムでは、主に日本語によるテキスト記述を仮定しているため、形態素解析には日本語形態素解析システムである Cabocha を利用しているが、連続するアルファベット列を連結してキーワードとして抽出しているため、英文にも対応可能である。

### 3.3.2 プレイリストを代表するキーワードの抽出

プレイリストが作成されることによって、プレイリストに含まれるシーン間に関連付けを行うことができるだけでなく、タイトルやプレイリストの作成履歴を解析することによって、プレイリストに含まれるシーンに対してテキストを関連付けることができる。

本研究では、2つの方法によってプレイリストを代表するキーワードを抽出し、プレイリストに含まれるシーンに対する関連付けを行った。

まず1つ目は、プレイリストのタイトルからキーワードを抽出するというアプローチである。プレイリストのタイトルは、そこに含まれるシーンに共通したキー



表 3.1: プレイリストの代表的キーワードの候補

|              |                        |
|--------------|------------------------|
| 作成の手がかり      | プレイリストを代表すると推測されるキーワード |
| コンテンツ（単数）    | 各シーンに共通するキーワード         |
| コンテンツ（複数）    | 各コンテンツに共通するキーワード       |
| シーン引用ブログ（単数） | 各シーンに共通するキーワード         |
| シーン引用ブログ（複数） | 各ブログに共通するキーワード         |
| キーワード（単数・複数） | そのキーワード                |

ワードが含まれる可能性が高いということは容易に想像できる。一方、プレイリストに対して付与されたタグやコメントなどは、プレイリスト中のそれぞれのシーンに対する記述がなされる可能性があるため、それらの情報をプレイリストに含まれるシーンすべてに関連付けることは不適切な場合が多いと思われる。そのため、タイトルのみを利用し、そこからキーワードの抽出を行う。

2つ目のアプローチとして、プレイリストを作成する際に手がかりとした情報からキーワードの抽出を行う。各プレイリストには、作成の際に、「コンテンツ」「ブログ」「キーワード」のどれを基にシーンを選択したかによる属性の付与がなされており、その属性情報を基に、プレイリストに対する代表キーワードを抽出する。それぞれの属性を持つプレイリストの代表的なキーワードは、表 3.1 に示すキーワード集合に含まれると考えられる。

表 3.2: プレイリストに対する代表キーワードの抽出

| 作成の手がかり       | 複数コンテンツ                          | 複数ブログ               | ブログ・コンテンツ           |
|---------------|----------------------------------|---------------------|---------------------|
| 推測に<br>利用する情報 | 映像タイトル、<br>タグ、コメント※ <sup>1</sup> | ブログタイトル、<br>ヘッダ、フッタ | 含まれるシーン<br>に対するタグ   |
| 判定条件          | 過半数より多い<br>コンテンツに含まれる            | 過半数より多い<br>ブログに含まれる | 過半数より多い<br>シーンに含まれる |

※<sup>1</sup> タグ、コメントは、コンテンツ登録時に投稿者によって付与されたもの

また、コンテンツやブログを手がかりとしたプレイリストに対する代表キーワードを抽出するために具体的に利用する情報と、代表キーワードとして判定する基準を表 3.2 に示す。例えば、3つの映像シーン引用ブログを手がかりにプレイリストが作成された場合、2つ以上のブログのタイトル、ヘッダ、フッタに共通して含まれるキーワードが代表キーワードとして抽出される。この判定基準に関しては、



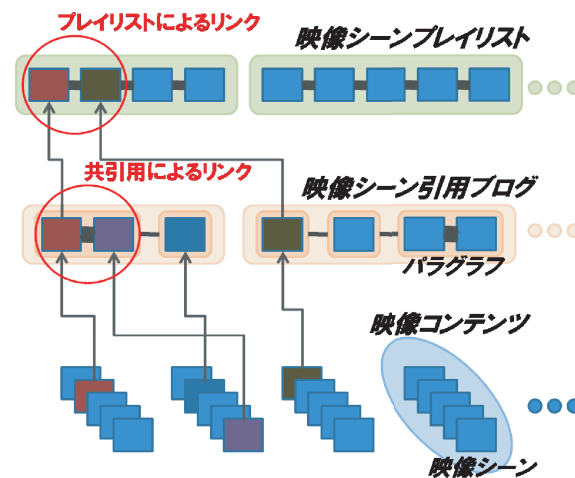


図 3.9: 映像シーンのリンク構造

理論的な根拠によって決定することが困難であるため、経験的に決定していく必要があると考えられる。

このアプローチは、プレイリストによって収集されたテキストからではなく、すでにシーン引用などによって蓄積されたテキストアノテーションからキーワードを抽出している。そして、そのキーワードをプレイリスト中のシーンに関連付けることによって、蓄積されていたテキストアノテーションの対象シーンを増やすことにつながり、その結果、それぞれのキーワードによる検索対象範囲を広げることに貢献する。

### 3.3.3 シーン間のリンク構造

映像シーンの引用と映像シーンプレイリストの作成が行われることで、以下の4種類の属性によってシーン間に対して関連付けが行われる。

- A) 同一の映像コンテンツに含まれるシーン間
- B) 同一のブログに引用されているシーン間
- C) 同一のパラグラフに引用（共引用）されているシーン間
- D) 同一のプレイリストに含まれるシーン間

そして、それぞれの属性に対するシーン間の関連の強さについて考察し、リンク構造の推測を行う。

まず、Aのように同一の映像コンテンツに含まれるシーン間には、それぞれのシーンが引用された時点で、ブログやプレイリストの構造に関係なく関連付けが行われる。映像コンテンツ全体に関する関連性があるという推測によって関連付けが行われるが、それぞれのシーンの深い内容に関して関連性があるかどうかを推測することができない。例えば、スノーボードをテーマとしたコンテンツに含まれるシーンで、スノーボードをしているシーンと雪山を映したシーンでは、そのシーンに写っている意味内容に関しては深い関連性がない。そのため、Aの属性による関連はそれほど強くない。

次に、映像シーン引用ブログの構造に係のあるリンク属性であるBとCについて考察する。映像シーン引用ブログは、複数のパラグラフによって構成され、また、パラグラフは複数のシーンによって構成される。ブログを執筆する際には、何らかの一貫したテーマで映像シーンが引用されると考えられるため、同一ブログの中に引用された映像シーン間には、そのテーマに関して関連性があると推測される。また、Aはコンテンツごとに閉じたリンクであるが、BやCは、複数の映像コンテンツから同一のブログに対してシーンが引用されることによって、映像コンテンツ間を越えたリンク構造を推測することが可能である。Bのリンクは、Aによるものと同様に大きなテーマによってシーン間が関連しているという可能性が高いため、それぞれのシーン間の関連はそれほど強くないと考えるのが妥当である。一方、同一パラグラフへの引用（共引用）は、それぞれのシーンを同時に見ることを意図して行われるため、シーン間には内容情報に近い関連性があると考えられる。そのため、Cの属性に対するシーン間には、AやBの属性よりも強い関連があると考えられる。さらに、映像シーンプレイリストに含まれるシーンは何らかのテーマや意図に沿ったシーンの集合である可能性が高いと考えられるため、含まれるシーン間にはそのテーマによる関連性があると考えられる。しかし、プレイリストは大きなテーマで作成される場合と、よりニッチなテーマで作成される場合があるため、含まれるシーンの関連を一律に強いと判断するのは危険である。例えば、スノーボードというテーマでプレイリストが作成される場合と、ジャンプというテーマでプレイリストが作成される場合がある。これらを推測する手がかりとして、前節で述べたプレイリストの代表キーワードや、プレイリストの属性が利用可能であると考えられる。しかし、これを考慮せずに平均的なシーン間の関連を考えた場合、Cよりも関連が小さく、AやBよりも関連が大きいと考えられる。

以上のアノテーション情報から推測されるリンク構造の概念図を図3.9に示す。第5章で述べる実験では、このリンク間の強さを示すパラメータを  $C > D > B > A$  の順で0.1～0.5の範囲で経験的に決定した。しかし、データ量が変化すると共に、検索の適合率や適合率を最適にするパラメータを自動的に再計算する仕組みが必要であり、詳細は第7章の今後の課題にて述べる。

## 第4章 映像シーン検索

映像シーンの引用とプレイリストの作成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タグ（シーンタグ）、シーン間のリンク構造を活用することによって、効率的に映像シーンを検索する仕組みを提案する。具体的には、近年の情報検索に頻繁に利用されているタグクラウドの仕組みをシーンタグに適用するという手法に加え、検索結果の提示方法に特徴を持たせた。さらに、リンク構造を活用することで検索キーワードに対するシーンの関連付けを拡張した。

また本章では、シーン引用とプレイリスト作成によって収集されるアノテーションを利用した場合のシーン検索について、従来のシーン引用の仕組みによって収集されるアノテーションと比較して考察する。

本システムは、2007年4月からWebによる一般公開実験を行っている<sup>1</sup>。

### 4.1 タグクラウド

本システムは、シーンタグによるタグクラウド [14] を利用して映像シーン検索を行う。タグクラウドには主に2つの利点がある。まず、検索に利用可能なキーワード（シーンに対して関連付けられているキーワード）が提示されることによって、探したいシーンに対する適切な検索キーワードや、興味のあるキーワードを発見する手がかりとなるという利点がある。さらに、タグの提示方法（大きさや色）によってキーワードの人気度や注目度を推測する手がかりにもなるという利点がある。

検索システムのトップページには、アノテーションを解析することによって生成されたシーンタグが名詞・動詞・形容詞にカテゴリ分けされ、それぞれのカテゴリの中では50音順にソートされたタグクラウドが表示される。品詞によってカテゴリ分けを行うことは、検索に利用したいタグを効率的に発見するために効果的であると考えられる。例えば、「おもしろい」や「楽しい」などの類似した意味の形容詞で検索を行いたいと思った場合に、形容詞のみを参照してタグを選択したいという要求が考えられる。もし、分類がなされていない場合は、全てのタグを参照してタグを探さなければならない。

---

<sup>1</sup>divie, <http://video.nagao.nuie.nagoya-u.ac.jp/search/top>

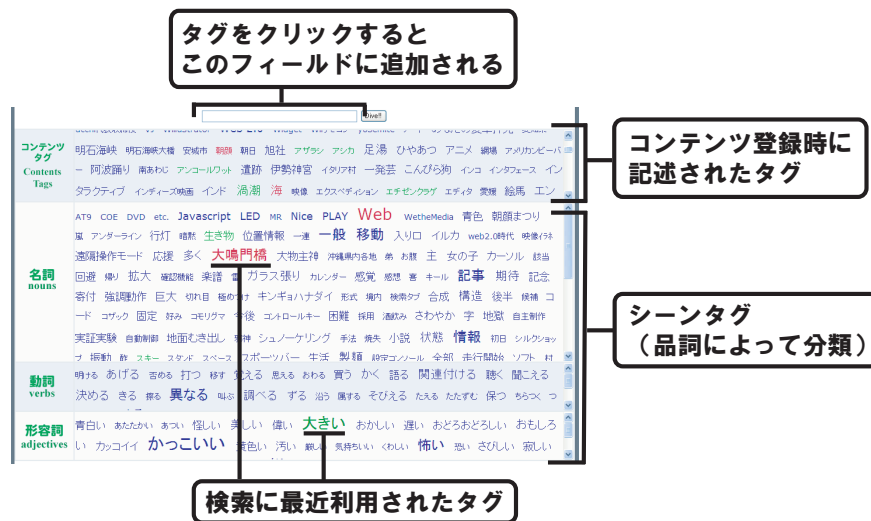


図 4.1: タグクラウド

また、それぞれのタグの大きさ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、アクセス時の前1週間以内に検索に利用されたシーンタグは赤色に、1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、多くのユーザにとってタグを発見するための有用な情報となる。

そして、表示されているタグをクリックすることによって、検索キーワードを入力するテキストエリアにそのキーワードが追加される。そのため、キーボードによるテキスト入力なしでも検索を行うことができる。また、複数のタグを利用して検索を行うことも可能である。

#### 4.1.1 タグのインクリメンタル検索

タグクラウドは、小スペースに大量のタグを表示することができるが、表示されるタグが増加するほど検索に利用したいタグを発見することが困難になり、またページを読み込む処理に対するブラウザの負荷が大きくなる。そのため、ページを読み込む時点では、表示されるタグの数を制限し、インクリメンタル検索によって全てのタグにアクセスできる仕組みを実装した。インクリメンタル検索とは、絞り込み検索とも言われ、検索キーワードを1文字入力するたびに徐々に検索結果を絞り込んでいく検索方法のことである。具体的には、テキストエリアに文字を入力することで、入力文字から始まる、もしくは読みの始まりが入力文字であるシーンタグのタグクラウドがテキストエリア下にポップアップ表示される(図 4.2)。例えば、「か」と入力された場合、「会議」、「かっこいい」のいずれも検

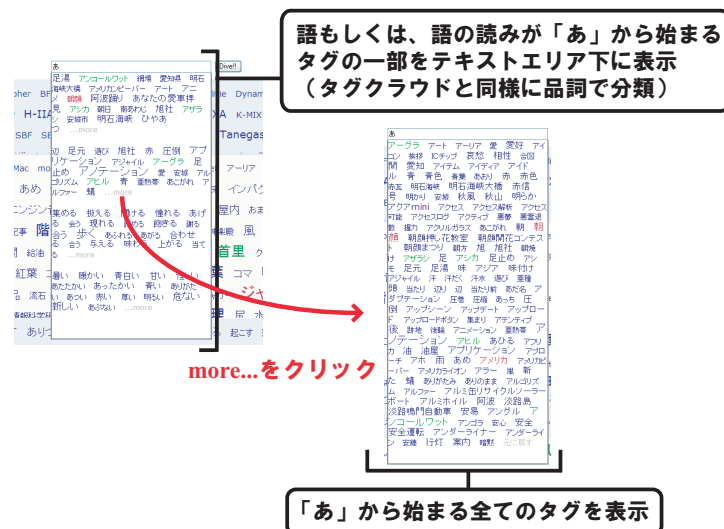


図 4.2: タグのインクリメンタル検索

検索結果として返される。アルファベット文字が入力された場合は、大文字小文字の区別はされない仕様とした。また、ポップアップ表示上でもデフォルトで表示されるタグ数は制限され、「more...」という文字をクリックすることで、検索にマッチした全てのタグを表示する。

タグの読み込みには Ajax (Asynchronous JavaScript and XML) の技術を利用しており、テキストエリアの内容が変わるごとにデータベースへのアクセスがページの読み込みと非同期で行われる。この機能によって、シーンテキストから抽出される膨大な数のタグの中から効率良く検索キーワードを探すことが可能になり、さらに、データベースに存在しないキーワード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。

## 4.2 検索アルゴリズム

本システムでは、検索が行われると、タグが関連付けられているシーンのスコアリングがなされ、次に、シーン間のリンク構造によってシーンスコアの再計算を行う。そして最終的に、含まれる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る。その結果、ランキングされたコンテンツが検索結果として返される。

### 4.2.1 シーンのスコアリング

検索キーワードを  $k$  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) 引用に基づいて生成されたシーンタグ  $k$  が関連付けられている
- B) 代表キーワードが  $k$  であるプレイリストに含まれる
- C) タイトルに  $k$  が含まれるプレイリストに含まれる（ただし B との重複は除く）
- D) タイトルに  $k$  が含まれるコンテンツ中のシーンである

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キーワード  $k$  の数によって決定され、該当する全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。

上記の条件と一致するシーンの、それぞれの条件に対するスコアについて述べる。シーンタグは引用のみの情報、プレイリストタイトルはプレイリスト作成のみの情報から抽出されるが、プレイリストの代表キーワードは、引用とプレイリストを組み合わせた情報から抽出される（3.2 節）。そのため、代表キーワードは、シーンに対して 2 つの異なる観点によって関連付けが行われるため、より関連度が高いと言える。一方、それぞれのシーンに対して個別に記述されるテキストであるシーンタグと、シーンの集合に対して記述されるテキストであるプレイリストタイトルでは、シーンタグの方が関連度が高いと考えられる。また、D による一致条件に関して述べると、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強いが、シーン内容そのものにはそれほど関連がないことが多いと推測される。それに加え、本システムは最終的にコンテンツのスコアリングを行うため、このスコアを高くしてしまうと、引用によって記述されたテキストにかかわらず、引用されている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非常に低いスコアとした。以上から、上記の条件に関して、 $B > A > C > D$  という重みでスコアを決定した。現在は、スコアの具体的な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自動的に算出する仕組みが理想的であり、それは今後の課題の 1 つである。

### 4.2.2 リンク構造に基づくシーンスコアの再計算

映像シーン引用の際に記述されるテキストをシーンに関連付ける仕組みには、アノテーションの網羅性が高くなりにくいという問題点がある。映像に詳細にメタ情報を付与するという動機でアノテーションが作成されるわけではないため、全てのシーンに対して、シーンに関連付けられるべきキーワードが漏れなく記述さ



れるということは期待できない。そのため、あるキーワードで検索を行った際に、本来スコアリングされるべきシーンであるにも関わらず、スコアが0となるシーンが存在しやすく、検索の再現率が低くなりやすいと考えられる。

この問題を解決するために、キーワードが直接関連付けられているシーンだけでなく、そのシーンと関連の強いシーンに対してもスコアリングを行う仕組みを導入した。この仕組みには、前章で述べた、引用とプレイリスト作成によって生成されるシーンのリンク構造を利用する。

検索キーワード  $k$  が関連付けられているシーン  $s$  と関連の強いシーン  $s_0$  のスコアリングを以下のように表現する。

$$scenescore(s_0, K) += relation(s, s_0) \times scenescore(s, k) \quad (4.1)$$

ここで  $scenescore(s, k)$  はキーワード  $k$  に対するシーン  $s$  のスコアである。また、 $relation(s, s_0)$  は、 $s$  と  $s_0$  の関連の強さであり、リンク構造から計算を行う。具体的には、3.2節で述べたシーン間に対する4種類のリンク属性によって0.1~0.5の値が与えられる。ここで、 $s_0$  は、検索キーワード  $k$  が関連付けられていないシーンだけでなく、検索キーワード  $k$  が引用やプレイリスト作成によって関連付けられているシーンである場合も考えられる。このスコアリングの再計算では、スコアの高いシーンと関連の強いシーンには高いスコアが与えられる仕組みとなっている。ただし、検索キーワード  $k$  が引用やプレイリスト作成によって直接関連付けられていないシーンに対しても同様の処理を行うと、キーワードとは関連の低いシーンに対してもスコアリングを行ってしまうと予想されるため、この計算は再帰的に繰り返すわけではなく、検索キーワード  $k$  が引用やプレイリスト作成によって関連付けられている全てのシーンに関してこの処理を行うことで終了する。

### 4.2.3 コンテンツのスコアリング

以上の仕組みによって算出されたシーンのスコアを基にして、検索キーワードに対する映像コンテンツのスコアリングを行う。まず、1つのキーワードで検索した場合、映像コンテンツ  $x$  に対するキーワード  $k$  に対する重み  $score(x, k)$  を以下の式で表現する。

$$score(x, k) = \frac{\sum_{s \in S_x} scenescore(s, k)}{|S_x|} \quad (4.2)$$

ここで、 $S_x$  は映像コンテンツ  $x$  に含まれる引用によってセグメントされた映像シーンの集合であり、 $scenescore(s, k)$  は、キーワード  $k$  に対するシーン  $s$  の重みを意味している。各コンテンツに対して、含まれるシーンのスコアを足し合わせるだけでなく、そのコンテンツに含まれるシーン数で割ることによって、コンテンツ中の検索キーワードと関連するシーンの密度を考慮することができる。

タグクラウドの中から  
検索キーワードを選択  
(図4.1, 図4.2)

検索結果として  
映像コンテンツがランキング表示される



映像コンテンツの俯瞰と  
シーンの映像内容の閲覧  
(図4.4)



図 4.3: 映像シーン検索の流れ

次に、複数キーワードで検索を行った場合に、映像コンテンツ  $x$  に対する検索キーワードの集合  $K$  に対する重み  $score'(x, K)$  を以下の式で表現する。

$$score'(x, K) = \left(1 + \frac{\sum_{k_i, k_j \in K} D(x, k_i, k_j)}{|K|}\right) \times \sum_{k \in K} score(x, k) \quad (4.3)$$

ここで、 $D(x, k_i, k_j)$  は映像コンテンツ  $x$  中に、キーワード  $k_i$  と  $k_j$  が共に関連付けられている映像シーンの数である。この式では、検索キーワードの集合のうちの2つのキーワードがコンテンツ  $x$  中の同一シーンに関連付けられている場合に、その映像シーンの重みを上げている。これによって、シーンタグの生成元であるテキストの文脈を検索結果に反映させている。

### 4.3 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、含まれるシーンを効率的に探すことのできる仕組みを提供することによって、最終的に最適な映像シーンが発見される。

具体的には、検索結果に含まれる全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全てのシーンタグによるタグクラウド（以降では、シーンタグクラウドと表記する）を表示する。本システムにおける映像シーン検索全

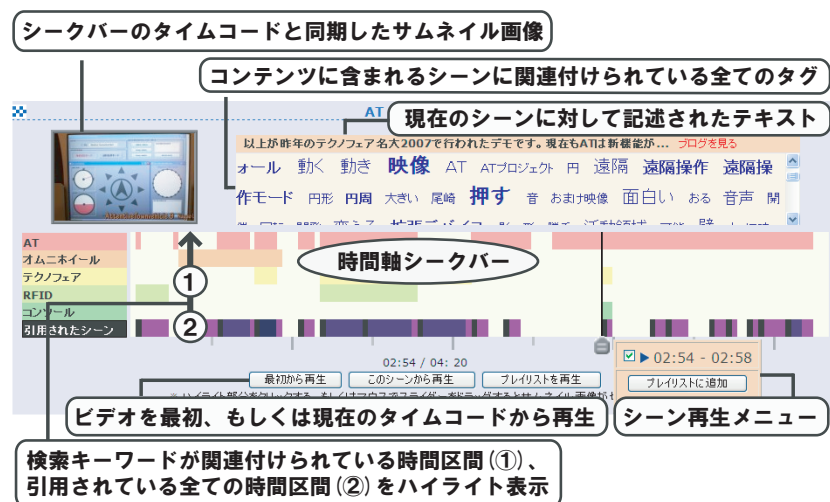


図 4.4: 映像シーン検索のユーザインタフェース

体の流れを図 4.3 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図 4.4 に示す。

このインタフェースでは、時間軸シークバーは、マウสด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、コンテンツ内で引用されている全ての時間区間がハイライト表示され、引用されている頻度が高いほど濃い色でハイライト表示される。シークバーをマウสด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が表示され、サムネイル画像にマウスカーソルを置くとコンテンツのビデオサイズに拡大表示される。さらに、そのタイムコードを含むシーンに関して、引用先であるブログコンテンツに対するハイパーリンクと、執筆されているテキストが表示される。ブログへのハイパーリンクを表示することで、映像コンテンツからブログへ、ブログから映像コンテンツへといった映像シーンを話題としたコンテンツに関するユーザ活動を促進することができる。また、ブログの執筆者にとっては、自身のコンテンツへの入口が増えるという利点があるため、シーン引用活動の動機付けにもなる。映像シーンを探す手がかりとして時間軸シークバーによるシーク操作の他にシーンタグクラウドを利用することができる。タグクラウドが表示されることによって、そのコンテンツ全体の情報をキーワードによって俯瞰でき、さらに、タグをクリックすることによってそのキーワードを時間軸シークバーに反映させることができる（つまりキーワードが関連付けられている時間区間がハイライト表示される）。

それらの操作を繰り返してシーンの情報を閲覧し、サムネイル画像やテキストを基に閲覧したいシーンを発見したら、Synvie 上でその開始時間からのビデオ再

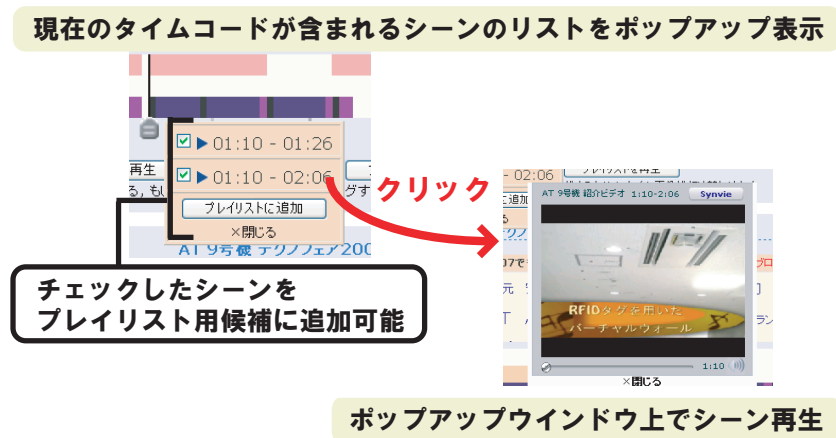


図 4.5: ポップアップウィンドウによるシーンの再生

生を行う、もしくは、ポップアップウィンドウ上でそのシーン区間の再生を行う（図 4.5）ことで映像シーン検索を実現する。

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、コンテンツの手軽な内容理解を支援する。さらに、シーンタグやテキスト、サムネイル画像を用いることで、ユーザは映像コンテンツそのものにアクセスせずに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、映像シーン検索を効率的に行うことができる。

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ると思われる。また、そのような仕組みが理想的であるが、それを高い精度で実現するためには、全ての映像コンテンツに対して非常に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 上のユーザ活動から収集することは非現実的であり、また人手によってアノテーションを作成することは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吸収して効率的に映像シーンを検索するために、このような検索の仕組みを採用した。この仕組みには、最終的には探している映像シーンにたどり着くことを目的としているものの、検索の過程でシーンの前後の文脈や全体におけるシーンの位置付けを確認することができるという利点もある。また、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って、求めているものをより正確に発見できると考えられる。

### プレイリストとの連携

3.2節でも述べたように、本検索システムにおける検索過程で、シーンを個別にプレイリスト用候補に追加することができる。具体的には、図4.5のポップアップウィンドウ上で、シーンに対応するチェックボックスにチェックを入れ、「プレイリストに追加」ボタンをクリックすることによってプレイリスト用候補に追加される。また、この検索結果上からプレイリスト作成画面を開くことも可能である。

あるキーワード $k$ を手がかりにしてプレイリストを作成したいと考えた時に、 $k$ が引用によって関連付いているシーン全てをまとめてプレイリスト用候補に追加できる(3.2節)。しかし、引用によるキーワードの関連付けの網羅性が低い場合、つまり $k$ が関連付けられているシーンの数が十分でない場合、それだけではユーザにとって満足のできるプレイリスト作成を支援することができない。そのような場合に、本検索システムを利用してプレイリスト用候補とするシーンの検索を行う。例えば、 $k$ を検索キーワードとしてシーン検索を行うことによって、 $k$ が関連付けられているシーンの周辺のシーンや、同じコンテンツに含まれるシーンも発見することができ、それをプレイリストに利用することができる。また、検索過程において新たに興味を持ったシーンをプレイリストに利用することもできる。この仕組みによって、ユーザにとってより自由なプレイリスト作成を支援することができることに加え、引用では関連付けることのできなかったキーワードをプレイリスト作成履歴からシーンに対して関連付けることができるという利点がある。

検索過程におけるプレイリストの再生は、プレイリストを作成するという動機によるものだけでなく、検索結果に含まれるシーンを連続的に再生することで自分の求めるシーンを発見したいという動機によっても行われると考えられる。検索キーワードが関連付けられたシーンが大量に存在した場合に、その1つ1つを順に見ていくのではなく、それらのシーンをプレイリスト再生によって閲覧した方が効率良くシーンを探すことができると考えられる。

## 4.4 リンク構造の利用に関する考察

映像シーン検索に利用するデータセットとして、従来の映像シーン引用の仕組みから蓄積されるアノテーションを利用した場合と、本研究の提案手法である映像シーン引用と映像シーンプレイリスト作成の仕組みを組み合わせることによって蓄積されるアノテーションを利用した場合の相違点について考察する。

それぞれのアノテーションを検索のデータセットとして利用した場合、検索対象となるシーンは、全て引用によって生成されるシーンであるという点は、両方のデータセットにおいて共通している。

しかし、それらのシーンに対して関連付けられるテキストが異なる(表4.1)。従来の仕組みでは、シーンに対して、引用された時に記述されたテキストのみが関



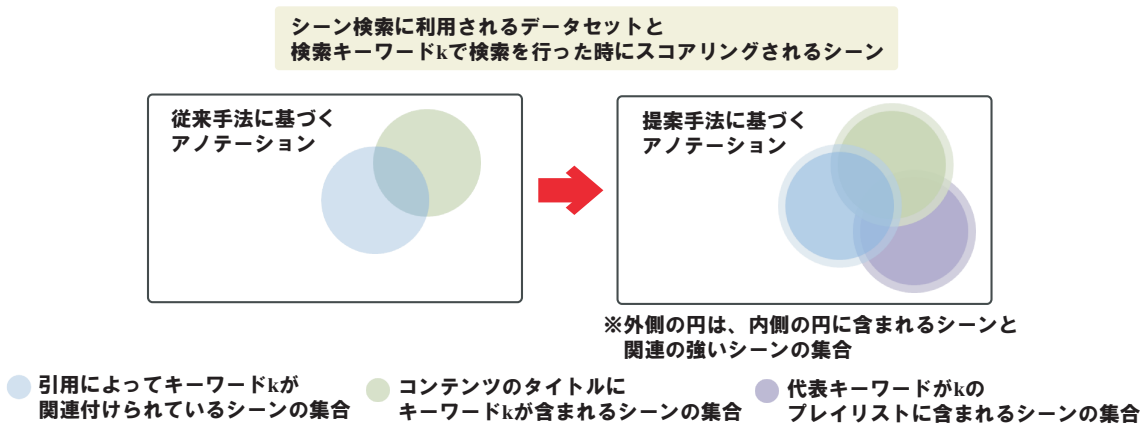


図 4.6: 検索されるシーンの比較

連付けられるが、提案手法では、引用された時に記述されたテキストに加え、そのシーンが利用されたプレイリストの代表キーワードが関連付けられる。さらに、提案手法では、共引用やプレイリスト作成履歴に基づいて生成されるシーン間のリンク構造を利用することができる。あるシーンが検索キーワードとマッチした場合に、関連の強いシーンもスコアリングされる仕組みによって、実質的にそれらのシーンにもその検索キーワードが関連付けられていることと同等であると言える。

検索キーワード $k$ で検索が行われた場合を例に挙げる。従来手法によるデータセットでは、 $k$ が引用によって関連付けられたシーンと $k$ がコンテンツのタイトルに含まれるシーンが検索される。一方、提案手法では、それらのシーンに加え、 $k$ を代表キーワードとするプレイリストに含まれるシーンも検索される。さらに、検索されたシーンと関連の強いシーンに対してもスコアリングがなされるため、それらのシーンも検索することができる。この関係を図 4.6 に示す。

表 4.1: シーンに関連付けられるテキストによるアノテーションの比較

| アノテーション | シーンに関連付けられるテキスト                                                        |
|---------|------------------------------------------------------------------------|
| 従来手法    | 引用されたパラグラフに記述されたテキスト                                                   |
| 提案手法    | 引用されたパラグラフに記述されたテキスト<br>利用されたプレイリストの代表キーワード<br>関連の強いシーンに対して関連付けられたテキスト |

図 4.6 からわかるように、提案手法では、検索キーワードに対するシーンの関連付けの拡張がなされることによって、検索されるシーンが増加する。そのため、



Web コミュニティ活動に基づくアノテーションを利用した映像シーン検索の課題の一つであった、検索の再現率を改善できると考えられる。また、図 4.6 中で、円の重なる部分に対応するシーンは、検索キーワードが引用とプレイリスト作成の両方の仕組みによって関連付けられているシーンの集合であり、それらのシーンとキーワード間には強い関連があると考えられる。この考察に基づいて、複数の仕組みによってキーワードが関連付けられているシーンのスコアを高くする検索アルゴリズムを適用することによって、再現率だけでなく検索結果上位の適合率も改善できると考えられ、提案手法ではそのようなアルゴリズムで検索を行う。



## 第5章 実験と考察

本研究における提案手法の有効性を検証するために、映像シーン引用とプレイリスト作成に基づくアノテーションを利用して、映像シーン検索を実現し、その評価を行った。

まず、Web による一般公開実験と被験者実験によってシーン引用とプレイリスト作成のデータを収集した。次に、シーン引用によって収集されたアノテーションからシーンとキーワードの関連付けのみを行ったデータセット（従来手法）と、シーン引用とプレイリスト作成によって収集されたアノテーションからシーンとキーワードの関連付けに加えて、シーンのリンク構造の解析を行ったデータセット（提案手法）を用意した。そして、それぞれのアノテーションデータを利用してシーン検索を行い、適合率と再現率によって比較を行うことで、提案手法の有効性の検証を行った。

前章で述べたように、映像シーン検索の仕組みでは、検索クエリに対する結果として映像コンテンツがランキングされる。しかし、今回の実験では、コンテンツのランキングによって手法の検証を行うためには十分なデータ数を収集することができず、検索対象が非常に少なくなってしまったため、コンテンツのランキングに加え、主に検索キーワードに対するシーンそのもののランキングによって比較・評価を行った。

### 5.1 アノテーションの収集

提案手法に基づく映像アノテーションの映像シーン検索に対する有効性を検証するために、まず引用とプレイリスト作成のデータを被験者実験を通して収集した。

#### 5.1.1 実験方法

本実験に参加した被験者は11人であり、インターネットサービスの利用に慣れており、またYouTubeやSynvieなどの動画共有サービスの閲覧を行ったことのある人によって構成されている。

まず、被験者はSynvieに登録されている142個のコンテンツの中から任意の映像コンテンツを3個以上選択し、それらのコンテンツの映像シーンを引用したブ

ログの作成を行った。ここで、選択する映像コンテンツの上限や、単一のブログ中に引用するシーンの数、各コンテンツに対して引用するシーンの数などの指定は行っていない。

そして、全ての被験者の映像シーン引用のデータ収集後に、被験者は任意のシーンを利用した映像シーンプレイリストを5個以上作成した。ここでは、プレイリストに入れる映像シーンの数などの指定は行っていないが、プレイリストを作成する3種類の手がかり（キーワード、コンテンツ、ブログ）をそれぞれ1回以上は使用してもらうように指示を行った。

### 5.1.2 実験結果と考察

Webによる公開実験と被験者実験を通して、計11人の人間によって、48個のブログ、53個のプレイリストが作成された。そして53個のコンテンツに対して合計403個のシーンが生成され、計5589語（重複を除くと2021語）のキーワードが抽出された。

#### 従来の引用インタフェースとの比較

本研究で開発した映像シーンを引用するユーザインタフェースの評価を行うために、過去のユーザインタフェース（図2.2を参照）を利用したシーン引用のデータ（従来手法）と、本研究で開発したユーザインタフェース（図3.1を参照）を利用したシーン引用のデータ（提案手法）の比較を行った。従来手法のデータは、2006年7月～2008年4月までにSynvieの公開実験と被験者実験によって収集されたデータである。それぞれのユーザインタフェースを利用したときの、ブログに含まれるシーンの数と、シーンの長さを比較したものが表5.1である。それぞれのユーザインタフェースを利用して作成されたシーン引用ブログの作成者は異なるが、作成者の大半を本研究室の人間が占め、情報リテラシの差はほとんどないと考えられるため、ユーザインタフェースを評価する上では信頼できる比較を行うことができると考えられる。

表 5.1: 映像シーンを引用するユーザインタフェースの比較

| UI   | ブログ数 | シーン数/ブログ数 | シーンの平均時間長（秒） |
|------|------|-----------|--------------|
| 従来手法 | 45   | 7.62      | 10.2         |
| 提案手法 | 48   | 8.40      | 19.0         |

表5.1から、1ブログあたりに含まれるシーンの数、それぞれのシーンの平均時間長が共に今回の実験で作成されたシーン引用ブログの方が上回っていることが

わかる。特に、引用されたシーンの平均時間長に関しては、2倍近くの値になっており、より時間区間の長いシーンが多く引用されたことがわかる。

従来のユーザインタフェースには、長い時間区間のシーンを引用するための効率が非常に悪いという問題点があった（2.5節）。そのため、本研究では、より直感的に映像シーンを選択できることによって効率良く引用を行うことができることを目的としてユーザインタフェースの改良を行った。その結果が、引用された平均時間長に表れたと考えられ、ユーザインタフェースの改良の有効性を確認することができたと考えられる。

ここでは、シーンの平均時間長からユーザインタフェースの比較を行ったが、プロダクトの作成やシーンの引用に要した時間の比較を行っていないため、引用の効果だけの評価を行い、費やされたコストに関しては考慮されていない。今後は、引用に費やされた時間の計測を行うことによって、費用対効果の観点からユーザインタフェースを比較する必要があると考えられる。

### 共引用に関する考察

次に、共引用（同一パラグラフへの複数シーンの引用）について考察する。本実験では、105個のパラグラフで共引用が行われた。共引用が行われたパラグラフを全て閲覧し、それぞれの共引用の行われた目的を主観的に評価した結果、3種類の引用目的が観測された。

まず、意味的にほぼ等しいシーンを並べて引用することによって、単一の内容に対する映像効果を高める目的で共引用がなされているケースであり、共引用のうち約70%を占めた。全く同じ風景を撮影した別のシーンなどが引用されており、引用されているシーンの内容がほとんど変わらないため、全てのシーンに対するコメントをそれぞれ記述しているのではなく、引用されているシーンを1まとまりにして、それに対するコメントの記述がなされていた。そのため、引用されているシーンのうちいくつかは欠けても、そこに記述されたコメントとは矛盾しない。例として、水族館のコンテンツから4つのシーンを並べて引用し、「水族館気分を味わってみましょう。」というコメントが記述されていたケース（図5.1左）や、異なるコンテンツに含まれる同じ場所を撮影したシーンを並べて引用し、「金刀比羅宮の鳥居です。」という記述がされていたケースが挙げられる。以上のことから、この目的によって共引用されたシーンの間には、意味的に非常に近い関連があると考えられる。

次に、複数のシーンを比較する目的で共引用がなされているケースであり、約25%がこのケースであった。ある共通点でシーンを集め、それぞれのシーンの異なる点を比較するということとなされていた。また、記述されるコメントの中には、それぞれのシーンに違いがあるということを明示的に表現しているものも多く見られた。例として、4つの日本の城のシーンが引用され「左上が松山城、右



図 5.1: 共引用の例

上が彦根城、左下が大阪城、右下が首里城です。」という記述がされていたケース（図 5.1 右）や、ある 1 つの研究に関するシーンが引用され「去年の 9 月に出版した時の様子（左）と今年の 9 月に出版した時の様子（右）」と記述されていたケースが挙げられる。並列に並べる目的による共引用に比べると、この目的によって共引用されたシーンの間には関連性が低い、ある共通点によって引用されていることは明らかであり、シーン間には類似関係があると考えられる。

そして、最も少なかったのが複数のシーンを連続的に閲覧させる目的で共引用がなされるケースであり、約 5 % の共引用がこの目的で行われていた。同じコンテンツの中から、いくつかのシーンを抜き出し、ダイジェストのように再生させることを意図したのが見られた。例として、ある研究のコンテンツの、8 秒～26 秒、36 秒～54 秒、62 秒～80 秒のシーンが引用され「下のシーンを順番に見てもらえれば大まかな流れが分かると思います。」という記述がされていたケースが挙げられる。この目的で共引用されているシーンの間には、同じコンテンツに含まれるシーンであるということ以外に意味的に深い関連があまりないことが多いと考えられる。

今回の実験では、共引用を行う際に、その目的や意図を入力する仕組みを実装していなかったため、これらの分類を機械的に行うことができなかった。また、言語処理によって分類を行おうと試みたが、高い精度での分類を行う手法を実現することができなかった。そのため、シーン間の関連の強さを計算する際に、共引用の目的を反映させることができなかった。しかし、上で述べたように共引用の目的によって、含まれるシーン間の関連の強さが異なることが明らかであるため、今後は、その目的を入力もしくは高い精度で推測する仕組みを実現する必要



がある。

### プレイリストの代表キーワードの抽出

複数のブログを手がかりに作成されたプレイリストと複数のコンテンツを手がかりに作成されたプレイリストに対して、3.3 節で述べた手法によって代表キーワードの抽出を行った。推測の条件は表 3.2 と同様である。複数のブログを手がかりとして作成されたプレイリストに対しては、手がかりとしたブログのタイトル、ヘッダ、フッタに含まれるキーワードの中で、過半数より多いブログに含まれるキーワードが存在した場合、それを代表キーワードとして抽出する。ヘッダ、フッタとは、ブログエントリの最初と最後に書かれるテキストのことであり、ブログ全体に対して記述されるコメントである。また、複数のコンテンツを手がかりとして作成されたプレイリストに対しては、コンテンツのタグ（投稿時に投稿者によって付与されるタグ）、タイトル、コメントに含まれるキーワードの中で、過半数より多いコンテンツに含まれるキーワードが存在した場合、それを代表キーワードとして抽出する。

実験によって収集したプレイリスト作成のデータから代表キーワードの抽出を行った結果を表 5.2 に示す。プレイリストに含まれる全てのシーンの中にキーワードで表現されるオブジェクトや場所などが映っているかどうかという判断基準によって、抽出されたキーワードがプレイリストに対して適切であるかどうかを評価した。

表 5.2: プレイリストに対する代表キーワードの抽出結果

| 手がかり  | プレイリスト数 | キーワードを抽出できた数 | 適切なキーワード |
|-------|---------|--------------|----------|
| ブログ   | 11      | 7            | 5        |
| コンテンツ | 12      | 10           | 8        |

表 5.2 から、ブログを手がかりに作成されたプレイリスト、コンテンツを手がかりに作成されたプレイリストのどちらに対しても、60 %以上の割合で代表キーワードを抽出することができ、そのうち 70 %以上の割合で適切なキーワードを抽出できたことがわかる。3.3 節で述べたように、プレイリストのタイトル情報からキーワードの抽出を行い、そのキーワードをプレイリストに対して関連付けるが、この手法によって抽出された代表キーワードの関連付けも併用することが有効であるということを実験結果から確認できたと考えられる。

また、抽出されたキーワードは具体的に、「四国」、「動物」、「ネコ」、「渦潮」などであり、比較的上位概念のキーワードが抽出された。この結果は、抽出に利用する情報そのものに上位概念のキーワードが含まれやすいということや、上位概

念のテーマによってシーンを集めてプレイリストを作成するというケースが多いことが原因であると考えられる。

より下位概念のキーワードの抽出を試みるために、代表キーワードの抽出に利用する情報を、プレイリストに含まれているシーンに対して関連付けられているキーワードとし、過半数以上のシーンに対して関連付けられているものを代表キーワードとして抽出を行った。その結果、代表キーワードを抽出できたプレイリストは2割にも満たず、抽出されたキーワードの傾向に関しても変化がほとんど見られなかった。

以上の結果から、代表キーワードを抽出するために利用する情報は適切であったと考えられる。また、プレイリストに含まれるシーンに対して、引用によって関連付けられているキーワードに共起関係があまり見られなかったことは、引用のみの情報からのキーワードの関連付けの網羅性の低さを表していると言えるため、プレイリスト作成の仕組みによってシーンとキーワードの関連付けを拡張することは有効な手法であると考えられる。

## 5.2 映像シーン検索

### 5.2.1 実験方法

収集されたアノテーションデータを利用して映像シーン検索を行った。また、提案手法であるシーン引用とプレイリスト作成の仕組みから収集されるアノテーションの映像シーン検索に対する有効性を、従来手法による引用の仕組みから収集されるアノテーションと比較することによって検証を行った。

まず、シーン引用によって収集されたアノテーションから、シーンとキーワードの関連付けのみを行ったデータと、シーン引用とプレイリスト作成によって収集されるアノテーションから、シーンとキーワードの関連付けやシーンのリンク構造の解析を行ったデータの2つのアノテーションデータを用意した。

次に、検索の評価に利用するキーワードの選出を行った。実験によって収集されたテキストから抽出された2021語のうち、プレイリスト作成に影響を受けたキーワードから、形容詞や「ビデオ」・「映像」・「動画」などのシーンとの適合性を客観的に評価できないキーワードを除き、その中から、シーン検索システムの公開実験で検索に利用された頻度の高い上位20語を選出した。具体例として、「動物」、「魚」、「スノーボード」、「ジャンプ」、「四国」などのキーワードを選出した。

そして、選出したキーワードを検索クエリとし、それぞれのアノテーションデータを利用して検索を行った時の再現率と適合率を計算することによってデータセットの比較を行った。

## 5.2.2 実験結果と考察

### 映像シーンのランキング

検索キーワードに対する映像シーンのランキングによる評価を行う。

まず、選出した全てのキーワード（20 語）と全てのシーン（403 シーン）との適合性の判断を人手によって行った。この適合性の判断は、検索キーワードで表現されるオブジェクトや場所などが、シーン中に、シーン長の 3 分の 1 以上の時間映っているかどうかという客観的な判断基準によって行った。

そして、用意した 2 つのデータセットを利用して検索を行い、再現率と適合率、F 値によって評価を行った。F 値とは、再現率と適合率の調和平均であり、再現率と適合率の両方の観点を考慮した指標によって評価を行うことができる。また、適合率は、検索結果全件に対する適合率と、検索結果上位に対する適合率によって評価を行った。ただし、検索結果上位 15 件までの適合率の比較を行おうとしたが、従来手法によるデータセットを利用した場合、検索結果が 15 件以上存在するものは 20 語中 10 語しか存在しなかったため、検索結果の上位に対する適合率は、検索結果が 15 件以上存在する全ての検索キーワード 10 語による結果を評価した。

各データセットを利用して検索を行った時の再現率、適合率、F 値、検索結果上位 15 件に対する適合率を表 5.3 に示す。また、検索結果上位 15 件までの検索結果件数に対する適合率の平均を比較したグラフを図 5.2 に示す。

表 5.3: 映像シーンのランキングに関する結果

| データセット | 再現率    | 適合率    | F 値    | 15 位適合率 <sup>※</sup> |
|--------|--------|--------|--------|----------------------|
| 従来手法   | 53.7 % | 62.5 % | 57.2 % | 74.0 %               |
| 提案手法   | 88.8 % | 52.1 % | 65.7 % | 82.7 %               |

※検索結果上位 15 件に対する適合率

表 5.3 から、F 値に関して提案手法の方が 8.5 % 改善された結果となり、検索の総合的な観点から提案手法が有効であったということがわかる。より詳細な評価を行うために再現率と適合率について比較と考察を行う。

まず、再現率について述べる。表 5.3 からわかるように、提案手法のデータセットを検索に利用した場合、従来手法のデータセットに比べて再現率が約 35 % 改善されたことがわかり、再現率に対する提案手法の有効性が確認された。この結果は、リンク構造の解析によってシーンに対するキーワードの関連付けが拡張されたことが大きく影響していると考えられる。また、従来手法のデータセットに対する再現率が 50 % 台という低い値であったことから、シーン引用のみのアノテーションの網羅性の低さを再確認することができ、従来手法に対する問題点も確認

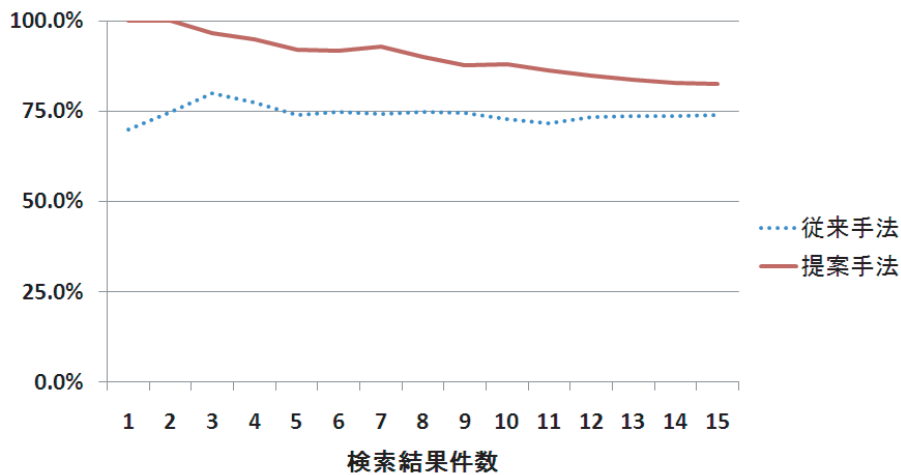


図 5.2: 映像シーンのランキングにおける検索結果件数に対する適合率

できた。従来手法による再現率が非常に低いことは、提案手法を適用することの有効性だけでなく、その必要性も示していると考えられる。

次に、適合率について述べる。表から、検索結果全件に対する適合率は、従来手法に優位性が見られた。一方、検索結果上位に対する適合率に関しては、提案手法に優位性があることがわかる。検索結果上位のそれぞれの適合率を詳細に見てみると、上位 5 件、10 件、15 件に対する適合率の平均は、提案手法によるデータセットに対して 92.0 %、88.0 %、82.7 %であり、従来手法によるデータセットに対して 74.0 %、73.0 %、74.0 %であった。この結果から、検索結果 10 位以内の適合率は 15~20 %の改善が見られる。参考として、両方のデータセットに対して検索結果が 10 件以上あるキーワード 14 語に対して同様の比較を行ったが、検索結果上位 10 件における適合率の平均が、提案手法によるデータセットに対して 82.9 %、従来手法に対して 72.1 %となり、検索結果上位 15 件による結果とほぼ同様の結果となった。

提案手法では、検索結果全件に対する適合率は低下したが、上位の適合率は改善された結果となった。この結果から、シーン引用とプレイリスト作成といった複数の仕組みによって関連付けられたキーワードとシーンには強い関連性があるという推論が正しかったと考えられる。

さらに、検索アルゴリズムを変更して検索を行うことによってこの推論に関する検証を行う。提案手法では、複数の仕組み（シーン引用とプレイリスト作成）によって検索キーワードが関連付けられているシーンほど高いスコアが与えられる。そして、このアルゴリズムを変更し、複数の仕組みによってキーワードが関連付けられていることの効果を弱めて検索を行った場合の検索結果上位に対する適合率を計算し、提案手法との比較を行った。具体的には、まず、シーン引用によっ

てキーワードが関連付けられているシーンのスコアを 20 倍にして検索を行った。このアルゴリズムを適用して検索を行うと、プレイリスト作成によってキーワードが関連付けられたシーンに対するスコアリングは相対的に非常に低い値となり、プレイリスト作成によるアノテーションは、キーワードとシーンの関連付けを拡張することによりのみ貢献することになる。次に、上とは逆に、プレイリスト作成によってキーワードが関連付けられているシーンのスコアを 20 倍にして検索を行った。これらのアルゴリズムを採用して検索を行った場合と、提案手法によって検索を行った場合の、検索結果上位に対する適合率を表 5.4 に示す。

表 5.4: 複数の仕組みによってキーワードが関連付けられることに関する評価

| 検索アルゴリズム            | 10 位適合率 | 15 位適合率 |
|---------------------|---------|---------|
| シーン引用による関連付けの重みを増大  | 79.0 %  | 72.7 %  |
| プレイリストによる関連付けの重みを増大 | 74.0 %  | 69.0 %  |
| 提案手法                | 88.8 %  | 82.7 %  |

表から、検索結果上位 10 件、15 件に対する適合率ともに提案手法が約 10 % 高い結果となったことがわかる。また、提案手法以外の場合は、従来手法のデータセットを用いた結果（表 5.3）よりも低い適合率となってしまった。この結果から、複数のアノテーションの仕組みによって関連付けられたキーワードとシーンの関連度が高いという推測を適用したことによって、検索結果上位に対する適合率を改善できたということを確認することができた。そのため、複数の仕組みを通してキーワードの関連付けが行われたシーンは、その関連度が高いという推測が可能であることを確認することができた。

以上の結果から、今回の実験では、提案手法によって映像シーン検索の再現率を大幅に改善できたことに加え、検索結果上位に対する適合率の向上にも有効であったと言える。検索結果全件に対する適合率は下がってしまったが、コンテンツの量が増えていくにつれて、検索結果全件にどれだけ適切なものが含まれるかどうかよりも、検索結果の上位に適切なものがより多く含まれることの方が重要であると考えられる。また、再現率と適合率を共に改善できた結果となり、キーワードとシーンとの関連付けの網羅性を高める効果と、キーワードとシーンの適合性を適切に推測する精度を高める効果の、本提案手法の効果として予想する 2 つの効果を共に確認することができたと言える。

## 映像コンテンツのランキング

次に、検索キーワードに対する映像コンテンツのランキングによる評価を行う。ただし、今回の実験では、検索対象となる映像コンテンツやアノテーションのデー



タが不足していたため参考程度の結果として述べる。

まず、映像シーンのランキングの評価と同様に、選出した全てのキーワード（20語）と全てのコンテンツ（53コンテンツ）との適合性の判断を人手によって行った。この適合性の判断は、検索キーワードで表現されるオブジェクトや場所などが、コンテンツ中に、10秒以上もしくはシーン長の3分の1以上の時間映っているかどうかという客観的な判断基準によって行った。10秒という値は、引用されたシーンの平均時間長の約2分の1の値である。

そして、用意した2つのデータセットを利用して検索を行い、再現率と適合率、F値によって評価を行った。また、適合率は、検索結果全件に対する適合率と、検索結果上位に対する適合率によって評価を行った。ただし、従来手法によるデータセットを利用した場合の適合率が非常に低かったため、検索結果上位に対する適合率は、上位2件までの参考程度の結果を示す。各データセットを利用して検索を行った時の再現率、適合率、F値、検索結果上位2件に対する適合率を表5.5に示す。

表 5.5: 映像コンテンツのランキングに関する結果

| データセット | 再現率    | 適合率    | F 値    | 2 位適合率 <sup>※</sup> |
|--------|--------|--------|--------|---------------------|
| 従来手法   | 43.3 % | 63.2 % | 51.4 % | 80.0 %              |
| 提案手法   | 83.9 % | 53.2 % | 65.1 % | 83.3 %              |

※検索結果上位2件に対する適合率

表5.5からもわかるように、シーンのランキングに対する結果とほぼ同様の結果となった。F値に関して提案手法の方が14.7%と大きく改善された結果となった。

次に、再現率について述べる。表5.5から、提案手法のデータセットを検索に利用した場合、従来手法のデータセットに比べて再現率が50%以上改善されたことがわかる。また、従来手法のデータセットに対する再現率が40%台という非常に低い値であったことから、シーン引用のみのアノテーションの網羅性の低さをこの結果からも再確認することができた。アノテーションの網羅性の違いは、検索結果件数にも大きく表れた。具体的には、従来手法によるデータセットを利用した場合、検索結果が0件であるキーワードが3語存在し、さらに、検索結果が3件以上存在するキーワードは、20語中6語、2件以上存在するキーワードは20語中15語しか存在しなかった。一方、提案手法によるデータセットを利用した場合は、全てのキーワードに対して検索結果が3件以上存在し、20語中19語は検索結果件数が6件以上存在した。このように、各手法のデータセットに対して、検索結果件数にも非常に大きな差が見られた。これらの結果から、映像コンテンツのランキングに関しても、再現率に対して提案手法が非常に有効であることを確認することができた。この結果も、シーンのランキングと同様に、リンク構造の解析に



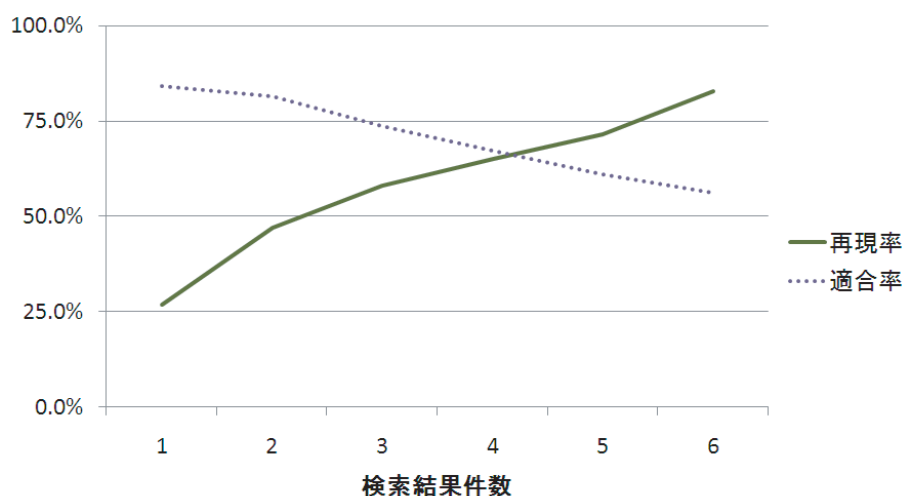


図 5.3: 提案手法のデータセットを利用した場合の検索結果件数に対する再現率と適合率の関係（映像コンテンツのランキング）

よってシーンに対するキーワードの関連付けが拡張されたことが大きく影響していると考えられる。

適合率に関しても、シーンのランキング同様の結果が得られ、検索結果全件に対する適合率は従来手法が上回る結果となった。また、検索結果上位に関する評価は、上位2件に対する適合率の比較となり、どちらのデータセットに対しても約80%と、ほとんど差が見られなかった。この原因は、検索対象となるコンテンツが不足していたことや、上位2件のみしか見ていないことによって必然的に差が出にくいからであると考えられ、この結果は参考程度の比較評価である。このように、検索結果上位に関する比較評価が困難であったため、提案手法のデータセットを利用した場合の検索結果件数と適合率、再現率の関係について観察を行った。提案手法のデータセットを利用した場合の、検索結果が6件以上存在した全ての検索キーワード19語に対する、再現率と適合率の検索結果件数に対する関係を示したグラフを図5.3に示す。グラフから、再現率が50%を超える検索結果件数2件～3件の時点における適合率が約75%、再現率が75%を超える検索結果件数5件～6件の時点における適合率が約60%程度であることがわかる。この結果からわかるように、再現率が75%の時点における適合率は低い値になってしまった。この原因として、検索対象が少なすぎるため、検索結果に含まれるノイズ（誤った検索結果）1つあたりの適合率に対する影響が非常に大きくなってしまい、少数の誤った結果によって平均の適合率が大きく下がってしまったからであると考えられる。検索結果件数1件の時点で再現率が25%を超えていることから、検索対象が非常に不足していたということが見て取れる。

### 5.2.3 まとめ

本実験では、提案手法を適用することで、シーン検索の再現率、F 値、検索結果上位に対する適合率を改善することができ、実験を通して、提案手法の映像シーン検索に対する有効性を確認することができた。この結果から、提案手法によるキーワードとシーン間の関連付けの拡張は再現率の向上に大きく貢献すると考えられる。また、検索結果上位に対する適合率が改善されたことから、複数の仕組みを通してキーワードの関連付けが行われたシーンは、その関連度が高いと考えられる。しかし、提案手法を適用することによって検索結果全件に対する適合率が低くなるという結果となり、シーンとキーワードの関連付けの拡張によって、正しい関連付けだけではなく、誤った関連付けを行ってしまうということが実験によって明らかになった。この問題を解決するための今後の課題として、誤った関連付けを減らすために最適な関連付けを行うパラメータを学習によって決定する仕組みや、高度な言語解析や新たな仕組みによるキーワードとシーンとの意味的関係性を抽出し、その属性を検索に反映させる仕組みなどが考えられる。

また、今回の実験では、映像コンテンツや映像シーン引用ブログ、映像シーンプレイリストのデータ数が不足していたことも問題点である。データの不足によって、主に映像シーンのランキングの観点によって評価を行ったが、本研究で提案している映像シーン検索の仕組みに対する直接的な評価、つまり映像シーンのランキングと映像コンテンツのランキングの両方の観点による評価を十分に行うことができなかったと言える。そのため、より大量のデータを収集することによって、両方の観点によって提案手法の有効性を示すことが今後の課題として挙げられる。

## 第6章 関連研究

本研究との関連が特に強い 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や Web サービスなどについては第2章で詳細に述べた。そのため、本章では、専用アノテーションツールを用いた映像アノテーションの研究について触れ、さらに、Web コミュニティ活動によって映像以外のコンテンツに対してアノテーションの付与を行っている研究や実運用されているシステムについて述べる。

### 6.1 ツールを用いた半自動アノテーションに関する研究

専用アノテーションツールを用いて、詳細な映像アノテーションを作成する手法が提案されている。ツールの多くは、XML 形式で記述するアノテーションの仕組みである MPEG-7[20] によってアノテーションの記述を行う。MPEG-7 は、主に映像コンテンツに対して専任の作業者が、信頼性の高い情報を記述する目的で利用される。記述形式を詳細に標準化してあるものの、仕様が細かすぎるため実用化にはなかなか結びついていない。専用ツールを用いて映像アノテーションを作成する手法には、アノテーションの作成にかかるコストが大きく、Web 上の膨大な数の映像コンテンツに対して直接適用することは困難である。しかし、詳細な映像アノテーションを作成することができるため、商用コンテンツなどの一部のコンテンツに対しては非常に有効な手法である。また、これらの手法で提案されているアノテーションの記述方法や属性の付与、ユーザインタフェースなどは、映像アノテーション全般における重要な問題であるため、関連研究として取り上げた。

代表的な例として、長尾らが開発した Video Annotation Editor (ビデオアノテーションエディタ) [12] がある。ビデオアノテーションエディタは、音声認識や、カット検出・オブジェクトトラッキングなどを行うことができ、それらの機械処理から生成されるアノテーションを人間の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。リコーが開発した MovieTool [22] は、MPEG-7 記述用の映像アノテーションツールであり、映像コンテンツに対して階層構造表現などの詳細なアノテーションの記述を行うことができる。また、MPEG-7 スキーマを動的に取り込む仕組みがあり、MPEG-7 スキーマで定義されているすべての

メタデータを記述できる。さらに、MPEG-7 記述の編集において利用可能なタグ候補を表示する機能や、MPEG-7 の文法を定義した MPEG-7 スキーマと MPEG-7 記述との整合性のチェックを行う機能がある。IBM が開発した MPEG-7 対応アノテーションツールである VideoAnnExAnnotation Tool [9] は、映像コンテンツに対して、シーンやイベントやオブジェクトなどの意味属性をアノテーションとして付与することができる。木構造によって、意味属性をカテゴリ分けしてアノテーションを保存しているという特徴がある。さらに、この研究では、ツールを拡張し、Web を用いた協調作業によって映像コンテンツに対してアノテーションを行う仕組みも提案している [16]。この手法では、複数人の人間によって同時にアノテーションを作成することができるため、TRECVID のための大量の映像と画像のコレクションに対して協調的に内容情報を関連付けることができる。複数人の人間によって作業を行うことで、アノテーション作成の効率を上げている。

## 6.2 マルチメディアコンテンツに対するアノテーションに関する研究

任意の Web コンテンツに対するアノテーションを記述し、そのアノテーションを共有する仕組みに Annotea[6] が存在する。Annotea では、アノテーションの記述形式として RDF (Resource Description Framework) を採用している。現状では、専用の Web ブラウザが必要であるという点などの欠点もあるが、Javascript や AJAX の機能を用いて、通常の Web ブラウザで利用可能なシステムもいくつか存在する。この仕組みは、アノテーションそのものを情報共有に使用する目的であるため、コンテンツの中身を検索するために、アノテーションを解析するといったことは行われていない。

また、任意の Web サイトを部分単位でブックマークすることのできるサービスに Clipmarks<sup>1</sup>が存在する。Clipmarks は、Web ブラウザである Mozilla Firefox<sup>2</sup>をプラグインによって機能拡張することで使用可能になる。Clipmarks は、ソーシャルブックマークとしての仕組みも実現しており、ブックマークした Web サイトの一部分を、他のユーザと共有したり、ブックマークに対してコメントを記述することができる。この情報が蓄積されることによって、Web サイトの一部分に対するアノテーションを収集可能であると考えられ、またそれに基づいてブックマークされたコンテンツを検索する仕組みも提供している。しかし、Clipmarks は HTML のタグ構造を解析することによって部分単位でのブックマークを行うため、映像コンテンツの一部分（シーン）をブックマークすることはできず、映像シーンに対するアノテーションを収集することはできない。

---

<sup>1</sup><http://clipmarks.com/>

<sup>2</sup><http://www.mozilla.com/>

画像に対するアノテーションを収集することのできる Web サービスに画像共有サービスである Flickr<sup>3</sup>が存在する。Flickr は、個人が撮影した画像を Web 上で共有できるサービスであり、画像の部分に対するコメント記述、画像全体に対するタグの記述、画像のブログエントリへの引用を行うことのできる仕組みを提供している。また、タグに基づいて画像の共有と分類が可能である。タグを用いた情報の分類手法をタクソノミーと呼び、Web2.0 と呼ばれる Web 技術の草分け的存在である。Flickr においては、タグは検索と分類に利用され、コメントはコミュニケーションのために利用されるという位置付けである。Flickr のタグによる検索の問題点は、他のユーザのことを意識したキーワードだけでなく、タグ付けをした人自身のためだけのキーワードがタグとして付与されることが多いという点である。

アノテーション活動にエンターテインメント性を取り入れることで、画像アノテーション活動の活性化を図っているサービスに、ESP Game [2] や GoogleImage Labeler<sup>4</sup>がある。例えば、Google Image Labeler は、ユーザが表示された画像に対してキーワード入力を行い、他のユーザの入力したものと一致すればそのユーザたちに得点が入り、その得点を競わせるというものである。オンラインでのタグ付けをコンテスト形式で行うことでユーザのモチベーションを高め、結果として、画像アノテーションを蓄積する。この手法は画像に対するものであるが、映像コンテンツに対しても似たような手法を適用することが可能であれば、映像アノテーションの収集手法として非常に有用なものになると考えられる。

音楽に対するアノテーションの仕組みとして、梶らの研究 [23] が存在する。音楽の時間軸や楽譜に対してユーザがアノテーションを作成することができる仕組みである。楽譜の任意の要素に対してユーザの主観的な感想や評価を関連付けることが可能であり、アノテーションに基づいて、検索や推薦などのを行う仕組みについても提案を行っている。音楽は映像に比べて、ある程度の型が決められているコンテンツであるため、この手法を直接映像コンテンツに適用することはできないが、推薦などのアノテーションを利用する仕組みについては Web 上のコンテンツに共通する応用例である。

---

<sup>3</sup><http://www.flickr.com/>

<sup>4</sup><http://images.google.com/imagelabeler/>





## 第7章 おわりに

本章では、本論文をまとめ、今後の課題と展望について述べる。

### 7.1 まとめ

本研究では、映像シーンのブログ記事への引用と映像シーンプレイリストの作成を支援する仕組みをユーザに提供し、その編集履歴から映像アノテーションを抽出することによって、効率的な映像シーン検索を実現する仕組みについて提案を行った。さらに、被験者実験によって収集したアノテーションデータを利用して映像シーン検索を行うことで、提案手法の有効性を確認した。

第2章においては、映像コンテンツを話題の中心とした Web コミュニティ活動とそれに基づく映像アノテーションについて述べた。まず、Web コミュニティ活動に基づいてアノテーションを収集する手法の利点と問題点や、既存の研究、Web サービスについて述べた。そして、本研究で着目した映像シーンの引用に関する特徴や利点を述べた後、従来手法に残されていた課題について論じた。

第3章においては、本研究で開発した、映像シーンを効率的に引用するためのシステム、引用されたシーンを集めたプレイリストを作成できるシステムについて、また、それぞれのシステムが利用されることによって収集される情報から抽出可能な映像アノテーションについて述べた。まず、シーン引用を支援するシステムに関して、映像シーンをより効率的に選択するためのユーザインタフェースの開発を行い、引用先であるブログ上でシーン再生を行う仕組みを実現した。また、共引用された複数のシーンを同期的に再生する仕組みや、複数の映像コンテンツからシーンを引用する仕組みによってシーン間の関係を規定できる仕組みを実現した。次に、引用されたシーンを連続的に再生するシーンプレイリストを作成する仕組みを提供することによって、映像シーン間の関係をより容易に抽出するシステムを開発した。また、プレイリストは、作成の手がかりとした情報によって分類され、その種類に基づいて抽出された代表キーワードをシーンに対して関連付ける仕組みとなっている。さらに、それぞれの仕組みに基づいて計算されたシーン間の関連度を組み合わせることで、シーン間のリンク構造の解析を行い、その結果を利用することで、シーンとキーワードの関連付けの拡張とシーンとキーワードの関連度計算を行った。

第4章においては、映像シーンの引用とプレイリストの作成によって収集される映像アノテーションを利用して、効率的な映像シーン検索を実現する仕組みを提案した。本研究で開発した検索システムでは、検索結果に、該当する映像コンテンツに関するタグクラウド、引用されたシーン区間を表す時間軸シークバー、シーンに関連付けられたコメント文などが表示される。ユーザは、提示された情報によって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良く理解することで、最適なシーンを発見することができる。また、シーン間のリンク構造を活用することによって、検索の再現率や適合率を上げるための検索アルゴリズムを提案した。

第5章においては、提案手法の有効性を検証するために、被験者実験によって収集したシーン引用とプレイリスト作成のデータを用いて映像シーン検索の実験を行った。具体的には、シーン引用によって収集されたアノテーションから、シーンとキーワードの関連付けのみを行ったデータセット（従来の映像シーン引用の仕組みと同等）と、シーン引用とプレイリスト作成によって収集されるアノテーションから、シーンとキーワードの関連付けやシーンのリンク構造の解析を行ったデータセット（提案手法）の2つのアノテーションデータを用意し、比較を行った。その結果、提案手法を適用することで、シーン検索の再現率、F 値、検索結果上位の適合率が改善された。特に、再現率に関しては非常に大きな改善が見られ、提案手法によるキーワードとシーン間の関連付けの拡張の有効性を確認することができた。また、検索結果上位の適合率が改善されたことから、複数の仕組みを通してキーワードの関連付けが行われたシーンは、その関連度が高いであろうという考察についても実証することができた。検索結果全件に対する適合率については従来手法に若干の優位性が見られたが、コンテンツが増えるにつれて検索結果上位の適合率はそれ以外の適合率に比べてより重要になると考えられるため、検索結果全体の適合率はそれほど大きな問題にはならなくなると思われる。

第6章においては、第2章で述べたWebコミュニティ活動に基づく映像アノテーション手法以外の関連研究について、専用アノテーションツールを用いた映像アノテーションの研究や、Webコミュニティ活動によって映像以外のコンテンツに対してアノテーションの付与を行っている研究について述べた。

本研究では、大量のデータを用いた提案手法の検証を行うことができなかったが、今後、映像シーンの引用、プレイリストの作成、映像シーンの検索というユーザ活動サイクルが活発になり、より大量のアノテーションが蓄積されることによって、さらに高い精度の映像シーン検索や、要約といった高度な応用が実現可能になると思われる。

## 7.2 今後の課題と展望

### 7.2.1 大量のデータに基づく提案手法の検証

本研究で行った実験では、映像コンテンツやアノテーション（シーン引用ブログとシーンプレイリスト）のデータが不足していたため、主に映像シーンのランキングの観点によって評価を行った。しかし、本研究で提案している映像シーン検索の仕組みに対する直接的な評価、つまり映像シーンのランキングと映像コンテンツのランキングの両方の観点による評価を十分に行うことができなかった。そのため、より大量のデータを収集することによって、両方の観点によって提案手法の有効性を示す必要がある。大量のデータを収集するためには、本システムのユーザビリティなどの改良が必要であると考えられる。また、現在はSynvieに登録されている映像コンテンツに対してのみ利用できるシステムになっているが、ブラウザのプラグインや、クライアントソフトとして開発することによってWeb上の任意の映像コンテンツに対して利用可能にするという手段も考えられる。大量のアノテーションが蓄積されることによって、より精度の高い映像シーン検索や、要約といった幅広い応用の実現可能性が高まると考えられる。

### 7.2.2 映像シーンを検索する仕組みに関する評価

本研究では、引用とプレイリスト作成の仕組みに基づくアノテーションの有効性に関しては、従来手法との比較評価によって確認することができた。しかし、提案している映像シーン検索の仕組みに関しては、アノテーションを利用してシーンが検索可能であることを確認できたものの、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の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映像シーンそのものを検索結果として返すユーザインタフェースによる検索によってシーン検索を行った場合の検索時間などの比較を行ったり、検索結果上で提示される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することなどが必要である。ただし、本研究で提案している映像シーン検索の仕組みは、最終的には探している映像シーンにたどり着く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前後の文脈や全体におけるそのシーンの位置付けを確認することができるという利点もあるため、単純に検索時間のみの比較によっては評価できない。

### 7.2.3 プレイリスト作成に対する動機付けの向上

提案手法の効果をより高めるためには、ユーザからより多くのアノテーションを収集する必要がある、そのためにはシステムを利用するための動機付けが必要となる。より多くのユーザにプレイリスト作成の活動を動機づけるためにまず、プレイリストに利用するシーンについて考えていく必要がある。本研究で開発したシステムでは、引用されたシーンのみをプレイリストに利用できる仕組みを提供している。しかし、引用されたシーン以外はプレイリストに利用することができないため、プレイリスト作成のみをしたい人にとっては必ずしも都合のよい仕組みではない。この仕組みを採用している理由は、引用とプレイリストのそれぞれの編集履歴を密に組み合わせることによって、アノテーションの量と質を共に向上させていくアプローチをとっているからであるが、シーン引用とプレイリスト作成の活動が活発化し、ユーザが自然に両方の仕組みを連携して利用できるようになれば、プレイリストに利用するシーンを限定する必要はなくなる。そのため、プレイリストに利用するシーンの時間区間を、引用されたシーンに限らず、より柔軟に選択できる仕組みを採用することも今後考えていくべきことである。

また、プレイリストをブログへ引用する仕組みを提供することも今後実現すべき課題の1つである。プレイリスト引用を可能にすることで、プレイリストに含まれる全てのシーンに対するテキストアノテーションを容易に収集できることに加え、そのテキストからプレイリスト内のシーン間の意味的関連性を抽出できる可能性もある。これを実現するためには、ブログ内におけるプレイリストの表示や操作に関するユーザインタフェースを設計する必要がある。

### 7.2.4 高度な言語解析等による意味的関係性の抽出

本研究では、シーン間、シーンとキーワード間に対して、その関連付けが行われた仕組みによって属性を分類している。しかし、第5章の実験でも明らかになったように、引用や共引用、プレイリスト作成には、より深い目的や意図が内在していると考えられる。引用の意図を適切に抽出することができれば、シーン間のリンク属性を詳細に分類できるため、シーン間に対する関連度の計算をより適切に行うことができると考えられる [25]。例えば、シーンを並列に並べる目的で共引用が行われた場合とシーンを比較する目的で共引用が行われた場合で、含まれるシーン間の関連度を変化させることができる。。また、意味的関連性を抽出することができれば、検索や推薦、要約等の応用に応じてシーンの関連度や重要度を適切に変えることができると考えられる。

意味的関連性を抽出するためには、より高度な言語解析によって、シーンとキーワード間の関連性を抽出する方法や、オントロジーの概念を用いることによるキーワード間の関連性を抽出するなどといった方法が考えられる。言語解析などの機



械的な手法に加え、より明確な意味的関連性を抽出するためには、ユーザに引用の意図などを自然に入力させるための仕組みが必要であると考えられる。

### 7.2.5 学習によるシーン関連度計算のための最適なパラメータ抽出

本研究では、シーン間のリンク構造を解析するために、シーン間の関係属性（共引用されている、同一のプレイリストに含まれるなど）を利用し、それぞれの属性に対して関連の強さを表すパラメータを割り当てることで、シーン間の関連度の計算を行っている。しかし、現状では、属性に対する関連の強さの上下関係のみを固定し、パラメータの詳細な調整は経験的に決定している。しかし、データ量が増加した時、現在のパラメータ設定が常に適切であると言えないため、データ量が増加するごとに、検索の適合率や適合率を最適にするパラメータを自動的に再計算する仕組みが必要である。そのためには、パラメータを調整したときの適合率や再現率を計算するための学習データが必要であり、その学習データを人手によって作成する必要がある。学習データに必要とされる量は、検索対象となる映像シーンや映像コンテンツの数にも依存すると考えられる。

### 7.2.6 全シーンを対象としたシーンの重要度の計算

本研究で提案した映像シーン検索の仕組みでは、検索キーワードが関連付けられているシーンの集合内におけるリンク構造を利用することによって各シーンの重要度を決定している。しかし、キーワードに関わらず、全てのシーンに対するリンク構造から各シーンの重要度を計算する手法も考えられる。全シーンを対象とした各シーンの重要度を計算するために、Web ページの重要度を計算する手法であるページランク (PageRank) [13] の仕組みを映像シーンに適用する方法が考えられる。ページランクは、重要なページにリンクされているページは重要であるという仕組みに基づいて、Web ページに対する重みづけを行う。映像シーン引用と映像シーンプレイリストによって生成されるリンク構造を利用することで、ページランクを映像シーンに適用し、重要なシーンと関連の強いシーンは重要という仕組みに基づいてシーンのスコアリングを行う手法を考案した。この手法で計算される値をシーンランクと呼ぶ。あるシーン  $s$  に対するシーンランク  $R(s)$  は以下の式によって表現される。

$$R(s) = c \sum_{u \in B_s} \frac{R(u)W(s, u)}{\sum_{t \in B_u} W(t, u)} \quad (7.1)$$

$B_s$  は、 $s$  と関連のあるシーンの数である。 $W(s, u)$  は、 $s$  と  $u$  の関連度であり、3.3 節で述べたシーン間の関連度を適用する。 $c$  は、値を正規化するための定数である。この式によって、より多くのブログへの引用やプレイリストへの利用がさ

れているシーンほどシーンランクが高くなるため、シーンの重要度を計算するための1つのパラメータとして利用することができると考えられる。

しかし、ページランクと同様に、この式で有効な値を計算するためにはある程度大規模のリンク構造が必要であると考えられるため、実験においてはこの値を検索に利用しなかった。また、ページランクとは異なり、双方向リンクによって計算を行っている。そのため、ユーザが意図的に、シーンランクの高いあるシーンと共引用することで、そのシーンのシーンランクを上げることができてしまうという問題があり、小規模のリンク構造ではこの影響がさらに大きく働いてしまう。

また、シーンランクもページランクと同様に、検索結果の候補が複数得られたときに、どちらを優先するかの手がかりに用いるものである。そのため、シーンの絶対的な重要度を計算するためには、重要なシーンとはどのようなシーンかということをより厳密に定義した上で、例えば、シーンが視聴された回数であったり、シーンを引用したユニークユーザ数であったりといったいくつかの指標を組み合わせることで重要度を計算する必要がある。



## 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、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。

大平茂輝助教には、研究の本質的なことから、技術的なことまで幅広く貴重な御指導、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。

研究室 OB である名古屋工業大学工学研究科山本大介助教には、本研究の基礎的な研究であると言える Synvie に関することに加え、研究の進め方や技術に関する御指導を頂き、大変御世話になりました。ここに御礼申し上げます。

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。

石戸谷顕太郎さんには、ネットワークなどの基礎的な技術に関することや、プログラミングに関することに関して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。

研究室 OB である、友部博教さん、梶克彦さん、成田一生さん、林亮介さん、伊藤周さんには、研究室在籍中に、研究やプログラミングに関する御指導、御意見を頂き、また、研究室での活動の中でも大変御世話になりました。ありがとうございました。

尾崎宏樹さん、安田知加さん、森直史さん、木内啓輔さん、井上泰佑さん、杉浦広和さんには、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、及び元秘書の金子幸子さんには、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、影ながら見守っていただき、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] Esma Aimeur, Gilles Brassard, and Sebastien Paquet. Using personal knowledge publishing to facilitate sharing across communities. In Proceedings of the Twelfth International World Wide Web Conference (WWW2003), 2003.
- [2] Luis von Ahn, Laura Dabbish. Labeling images with a computer game. In Proc. Conf. on Human Factors in Computing Systems, pp.319-326, 2004
- [3] 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.
- [4] Mike Dowman, Valentin Tablan, Hamish Cunningham, Borislav Popov. Web-Assisted Annotation, Semantic Indexing and Search of Television and Radio News. World Wide Web Conference 2005, 2005.
- [5] Yufen Dou, Eisuke Itoh, Sachio Hirokawa, Daisuke Ikeda. An Approach to Analyzing Correlation between Songs/Artists Using iTunes Playlists. Proc. International Conference on Intelligent Agents, Web Technology and Internet Commerce (IAWTIC'2005), pp.28-30, 2005.
- [6] J. Kahan, M.-R. Koivunen. Annotea: an open RDF infrastructure for shared web annotations. In World Wide Web, pp.623-632, 2001.
- [7] Daisuke Kitayama, Natsumi Oda, Kazutoshi Sumiya. A Social Video Sharing System using User Comments based on Temporal Duration and Pointing Region. Proc. of International Workshop on Information-explosion and Next Generation Search (INGS2008), pp.55-58, 2008.
- [8] Taku Kudo, and Yuji Matsumoto. Fast Methods for Kernel-Based Text Analysis. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-2003), pp.24-31, 2003.
- [9] C.-Y. Lin, B. L. Tseng, J. R. Smith. ideoAnnEx: IBM MPEG-7 Annotation Tool for Multimedia Indexing and Concept Learning. Proc. IEEE Intl. Conf. on Multimedia and Expo (ICME), 2003.

- [10] Tomoki Masuda, Daisuke Yamamoto, Shigeki Ohira, and Katashi Nagao. Video Scene Retrieval Using Online Video Annotation. Lecture Notes on Artificial Intelligence (LNAI 4914: JSAI 2007 (K. Satoh, et al. Ed.), Springer-Verlag, pp.54-62, 2008.
- [11] 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.
- [12] Katashi Nagao, Shigeki Ohira, and Mitsuhiro Yoneoka. Annotation-based multimedia summarization and translation. In Proceedings of the Nineteenth International Conference on Computational Linguistics, pp.702-708, 2002.
- [13] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd. The pagerank citation ranking: Bringing order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998.
- [14] A. W. Rivadeneira, Daniel M. Gruen, Michael J. Muller, and David R. Millen. Getting our head in the clouds: toward evaluation studies of tagclouds. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.995-998, 2007.
- [15] Michael A. Smith, Takeo Kanade. Video Skimming and Characterization through the Combination of Image and Language Understanding. IEEE International Workshop, ICCV98, 1998.
- [16] Timo Volkmer, John R. Smith, and Apostol Paul Natsev. A web-based system for collaborative annotation of large image and video collections: An evaluation and user study. In Proceedings of the 13th ACM International Conference on Multimedia, pp.892-901, 2005.
- [17] Howard D. Wactlar, Takeo Kanade, Michael A. Smith, and Scott M. Stevens. Intelligent access to digital video: Informedia project. IEEE Computer, Vol.29, No.5, pp.140-151, 1996.
- [18] Daisuke Yamamoto, Tomoki Masuda, Shigeki Ohira, and Katashi Nagao. Video Scene Annotation Based on Web Social Activities. IEEE MultiMedia, Vol.15, No.3, pp.22-32, 2008.
- [19] Daisuke Yamamoto, Tomoki Masuda, Shigeki Ohira, Katashi Nagao. Collaborative Video Scene Annotation Based on Tag Cloud. PCM 2008, 397-406, 2008.

- [20] ISO. Information Technology - Multimedia Content Description Interface(MPEG-7), ISO/IEC 15938:2001. International Organization for Standardization(ISO), 2001.
- [21] Benjamin and Mena Trott. mttrackback - TrackBack Technical Specification. movabletype.org, <http://www.movabletype.org/docs/mttrackback.html>, 2002.
- [22] Ricoh. Movie Tool, <http://www.ricoh.co.jp/src/multimedia/MovieTool/>, 2002.
- [23] 梶克彦, 長尾確. 楽曲に対する多様な解釈を扱う音楽アノテーションシステム. 情報処理学会論文誌, 特集「便利で身近な音楽情報処理」, Vol.48, No.1, pp.258-273, 2007.
- [24] 寫田聡, 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.
- [25] 林 亮介, 土田 貴裕, 大平 茂輝, 長尾 確. 閲覧時アノテーションを利用した Web ドキュメントの引用とその応用. 情報処理学会第 70 回全国大会講演論文集, 2008.
- [26] 増田 智樹, 山本 大介, 大平 茂輝, 長尾 確. ソーシャルブックマークの仕組みに基づく映像シーンアノテーション. 情報処理学会第 70 回全国大会講演論文集, 2008.
- [27] 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.