

## 映像シーンの引用に基づく映像アノテーションとその応用

増田 智樹<sup>†</sup> 石戸谷 顕太朗<sup>†</sup> 大平 茂輝<sup>††</sup> 長尾 確<sup>‡</sup><sup>†</sup>名古屋大学 大学院情報科学研究科<sup>††</sup>名古屋大学 エコトピア科学研究所 <sup>‡</sup>名古屋大学 情報メディア教育センター

## 1 はじめに

近年、Web 上に膨大な数の多種多様な映像コンテンツが存在するようになり、シーン検索などの応用に対する要求が非常に高まっている。それらを実現するためには映像アノテーションと呼ばれる、映像内容に関するメタ情報が必要不可欠である。我々は、Web 上で不特定多数の人間が行うコミュニティ活動から映像アノテーションを収集する研究を行ってきた [1]。

本研究では、映像シーンの引用に着目し、従来の引用の仕組みを改良し、さらに新たな仕組みを組み合わせることによってより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、映像シーンを効率的に引用するためのユーザインタフェースを備えたシーン引用システムを開発した。次に、引用されたシーンを再利用したプレイリスト作成を支援するシステムの開発を行った。そしてそれらの仕組みによって収集されるアノテーションを利用して映像シーン検索を行うシステムを実現し、評価を行った。そして、Web による一般公開実験<sup>1</sup>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、それらのアノテーション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検証した。

## 2 映像シーンの引用と

プレイリスト作成に基づくアノテーション

## 2.1 映像シーンの引用支援

映像シーンを効率的に引用できる仕組みを提供し、それを利用したユーザによる編集履歴を映像アノテーションとして蓄積するシステムを開発した。

まず、映像コンテンツを最近見た映像もしくは検索によって選択し、図 1 のインタフェースに読み込む。複数のコンテンツを読み込むことが可能であり、シーンを見比べながら引用を行うことができる。

ユーザは読み込まれた映像コンテンツに対するビデオと、ビデオに同期して上下左右に流れるサムネイル画像を参照することによって引用するシーンの時間区間の設定を行う。サムネイル画像をマウスドラッグによってシーク操作することで引用したいシーンを探し、マウスクリックによってシーンの開始、終了、代表フレームを設定する。引用されたシーンは、編集画面では代表フレームに対応するサムネイル画像によって表現され、プレビュー画面やブログ中では、Adobe

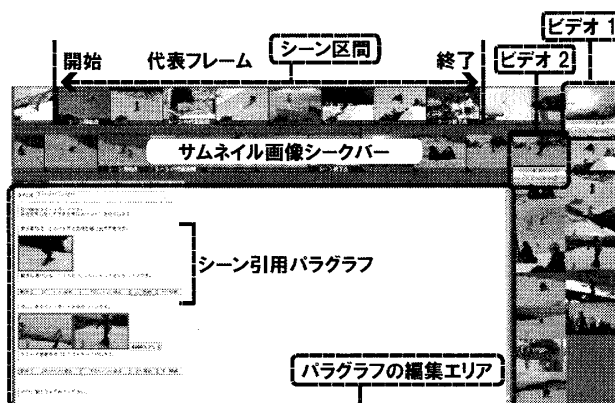


図 1: 映像シーンを引用するユーザインタフェース

Flash Player でシーン再生を行うことができる。さらに、複数のシーンを同一パラグラフ中に並べて共引用することもでき、共引用されたシーンは単一のプレイヤー中で同期再生することができる。そして、シーンに対するテキストを記述することで、シーンを引用したパラグラフの作成を行い、同様の作業を繰り返すことによって映像シーンを引用したブログの執筆を行う。

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蓄積され、さらに、それぞれのシーンに対してテキスト情報が関連付けられる。

## 2.2 映像シーンプレイリストの作成支援

引用されたシーンを利用して手軽にプレイリストを作成できる仕組みを提供し、その作成履歴から映像アノテーションを収集するシステムを開発した。

まず、プレイリストに利用するシーンを、ブログ、コンテンツまたはキーワードを基に検索する。例えば、あるブログを選択することによってそのブログ中に含まれるシーンをプレイリストに利用することができる。任意の数のシーンを選んだら、プレビュー再生を行いながらシーンの削除や並び替えを行う。最後に、タイトル等を記述し保存することでプレイリストが作成され、他ユーザと共有することができる。

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ため、それらのシーン間には関連性があると推測される。

## 2.3 アノテーションの解析

まず、シーンに対して関連付けられたテキストを解析し、映像シーンの内容を表現するキーワード（以降、シーntagと表記する）の抽出を行った。

次に、プレイリストに対する代表的なキーワードを、プレイリストのタイトルやプレイリスト作成の手がかりとしたブログやコンテンツから抽出し、プレイリスト中のシーンに対して関連付けを行った。

そしてさらに、推測されるシーンの関連性からシーン間のリンク構造の解析を行った（図 2）。ある 2 つのシーン間に対して、同一のコンテンツに含まれている

Video Scene Annotation and its Application  
Based on Quoting Video Scenes<sup>†</sup> MASUDA, Tomoki  
(masuda@nagao.nui.e.nagoya-u.ac.jp)<sup>†</sup> ISHITOYA, Kentaro  
(ishitoya@nagao.nui.e.nagoya-u.ac.jp)<sup>††</sup> OHIRA, Shigeki(ohira@nagoya-u.ac.jp)<sup>‡</sup> NAGAO, Katashi(nagao@nui.e.nagoya-u.ac.jp)  
Graduate School of Information Science, Nagoya University  
(†)

EcoTopia Science Institute, Nagoya University (††)

Center for Information Media Studies, Nagoya University (†)

Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

<sup>†</sup> Synvie beta, <http://video.nagao.nui.e.nagoya-u.ac.jp/>

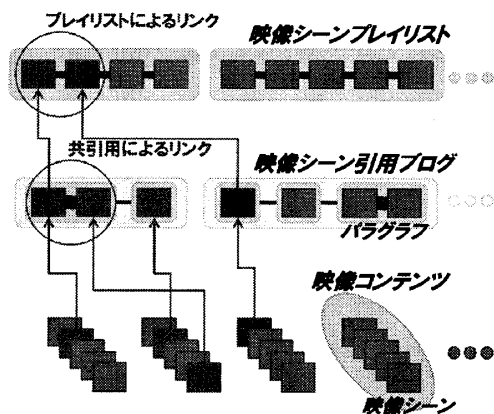


図 2: 映像シーン間のリンク構造

場合 (A)、同一のブログに引用されている場合 (B)、ブログ内で共引用されている場合 (C)、同一のプレイリストに含まれている場合 (D) の 4 種類の属性によって、関連付けを行った。それぞれの属性による関連の強さは、 $C > D > B > A$  とした。

### 3 映像シーン検索

我々が開発してきた映像シーン検索システム [2] を改良し、解析したアノテーションを利用することで、映像シーンを効率的に検索する仕組みを提案する。

#### 3.1 タグに基づく映像シーン検索システム

検索のトップページには、シーンタグの一覧が表示され、その中から検索キーワードを選択し、検索を行う。その結果、キーワードが関連付けられたシーンにスコアリングがなされ、含まれるシーンのスコアの総計によって、コンテンツがランキングされる。そして、それが検索結果として返され、それぞれのコンテンツに対して、シーン情報をブラウザ上で閲覧するための時間軸シークバーとサムネイル画像、シーンタグの一覧が表示される (図 3)。時間軸シークバーには、検索キーワードが関連付けられている時間区間と、そのコンテンツにおいて引用されている全ての時間区間がハイライト表示される。シークバーを動かすことで、そのタイムコードに同期してサムネイル画像が切り替わることに加え、そのタイムコードを含むシーンをポップアップウィンドウによって視聴することができる。また、一覧中のタグをクリックすることでそのタグをシークバーに反映させることができる。任意のタイムコードからの映像視聴も可能である。

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、手軽な内容理解を支援することで、効率的な映像シーン検索を支援することができる。

#### 3.2 実験と考察

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 コンテンツに対して合計 403 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。

そして、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) を用意し、シー

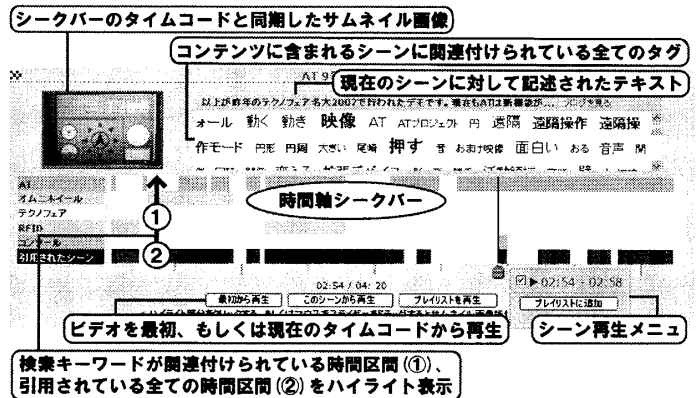


図 3: 映像シーン検索のユーザインタフェース

ン検索に利用することで比較を行った。

本研究で提案する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、コンテンツのランキングで比較を行うには検索対象が少なすぎるため、検索キーワードに対するシーンそのもののランキングを比較する。

評価に利用した検索キーワードには、抽出された 2021 語の中から、形容詞などの適合性を客観的に評価できない語を除き、プレイリスト作成に影響を受けた語の中から、公開実験による検索に利用された頻度の高い上位 20 語を選択した。検索キーワード 20 語に対する再現率の平均と、両方のデータセットに対して検索結果が 15 件以上存在する全ての検索キーワード 10 語に対する適合率の平均は、表 1 のようになった。表 1 から、再現率、適合率ともに提案手法に優位性があることが示された。特に、再現率に大きな差が見られ、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。

表 1: 映像シーン検索実験結果

データセット	再現率	適合率
従来手法	0.537	0.740
提案手法	0.888	0.827

※ 適合率は、検索結果上位 15 件までの適合率

#### 4 まとめと今後の課題

本研究では、映像シーンの引用と映像シーンプレイリストの作成を支援し、それらを利用するユーザ活動から収集される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仕組みを提案した。また、実験によって提案手法の有効性を確認した。今後の課題として、高度な言語解析などによって引用時に記述されるテキストとシーンとの間の意味的關係を推測し、検索に反映させる仕組みの考案などが挙げられる。

##### 参考文献

- [1] 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.
- [2] 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