Synvie:映像シーンの引用に基づくアノテーションシステムの構築とその評価

山本大介[†1] 増田智樹[†2]
大平茂輝[†3] 長尾確[†4]

本論文では，映像コンテンツに関連した Weblog コミュニティから映像コンテンツに関する意味情報をアノテーションとして取得する仕組みを提案する．具体的には，複数の映像シーンを引用した Weblog エントリーをユーザが手軽に編集可能な Web ベースのツールを提供し，その編集履歴から映像の任意の複数のシーンと Weblog エントリーの文書構造とを関連付け，アノテーションとして蓄積する仕組みを実装した．これにより，従来からある映像シーンに対する掲示板型コミュニケーションからの情報抽出を実現する仕組みより，質の高いアノテーションを獲得することが可能になる．これらを実現するための具体的なプラットフォームとして Synvie と呼ばれるシステムを開発した．さらに，本システムを一般に公開した実証実験を行い，それによって取得されたアノテーション情報の特性を解析することによって，本システムの有用性を示した．

Synvie: An Annotation System Based on Quotation of Video Scenes

Daisuke Yamamoto [†1], Tomoki Masuda [†2], Shigeki Ohira [†3]
and Katashi Nagao [†4]

In this paper, we propose a mechanism which acquires semantics of video contents as annotations from related Weblog communities. In particular, we have implemented a Web-based tool which user can generate a Weblog entry quoting video scenes easily. This tool can acquire relationships which associate multiple video scenes with a document structure of a Weblog entry from editing histories. In the result, we can acquire higher quality annotation than previous works which acquire information from communication with online message board systems. We have developed an online video quotation system "Synvie." Moreover, we have analyzed real annotation data which were accumulated using the public beta service which we are providing, and confirmed the utility of our system.

1. はじめに

近年，インターネットの発達と共に，映像・音楽などのマルチメディアコンテンツが Web 上で頻繁に配信・共有されている．それらのコンテンツはプロが作成したコンテンツだけではなく，個人が撮影・作成したコンテンツも爆発的に増加しており，それらのコンテンツをいかに効率よく配信・管理・検索するかといった問題が顕在化してきている．その一方で，Weblog や SNS，Wiki などの登場により個人や Web コミュニティからの情報発信が一般化し，影響力も増している．

従来，映像コンテンツの検索や要約などの応用を行う場合，映像認識や音声認識等の自動解析技術を利用し映像に関するメタ情報を取得する自動アノテーション方式[12] や，専任の作業者が映像のシーンに対するアノテーション情報を専用のツールで付与する半自動アノテーション方式[4),8),10),11)] によって付与されたアノテーション情報を利用する必要があった．しかしながら，とりわけ個人が作成したコンテンツの場合，手ぶれ・ピンぼけ・雑音・不明瞭な声などといった撮影者の技能の問題や，カメラ付き携帯電話やデジカメといった撮影機器の性能問題等から映像や音声の品質のばらつきが大きく自動認識・解析は極めて限定的にしか利用できない．また，専任の作業者による半自動アノテーションを行うためには，視聴者が限定され費用対効果が見合わない等の理由から，全ての映像コンテンツに対するアノテーションを施すことは困難である．

そこで本研究では，マルチメディアコンテンツとそれらを取り巻く Web コミュニティとを効果的に融合させる仕組みを提案し，それらのコミュニティにおけるユーザの自然な知的活動からコンテンツに関する知識をアノテーション[9] として獲得・蓄積・解析することを目的としている．具体的には，二つのコミュニケー

†1 名古屋大学情報科学研究科
Graduate School of Information Science, Nagoya University
†2 名古屋大学工学部
Department of Technology, Nagoya University
†3 名古屋大学エコトピア科学研究所
EcoTopia Science Institute, Nagoya University
†4 名古屋大学情報メディア教育センター
Center for Information Media Studies, Nagoya University

ション手段を提供する．一つ目は，映像コンテンツの任意のシーンに対して，コンテンツの内容に対する感想や評価などの情報の関連付けを支援する掲示板型コミュニケーションの仕組みであり，二つ目は，任意の映像シーンを引用した Weblog エントリーを生成し，映像シーンとそれらの記事の文書構造との関連付けを支援する Weblog 型コミュニケーションの仕組みである．これらの仕組みを作成することによって，ユーザらによる映像を題材としたコミュニケーションを支援する．さらには，コンテンツの内容とこれらのコミュニケーションとを詳細に結びつけることによって，コンテンツに付随する様々な情報をアノテーションとして獲得する．このような方式ならば，映像の画質や音質に左右されず，また，アノテーションにかかるコストも発生しない利点があり，上述した自動・半自動アノテーションの問題を回避できる．これらのコミュニケーションは単体のコンテンツのみに閉じているのではなく，コンテンツを部分引用する仕組みによって，Web 全体を対象とするより広がりをもつコミュニティの構築を支援する．また，人気のある映像コンテンツほどより多くのアノテーションの取得が可能になる利点がある．

そこで，本論文では，映像シーンへのアノテーションの仕組み，映像シーン単位でのコンテンツの引用に基づく Weblog エントリーからのアノテーション取得方法の提案，コミュニケーションに特化した具体的なインタフェースの提案，及び，システムの公開実験に基づく分析・評価を行い，コミュニケーションから得られるアノテーションを用いたアプリケーション作成のための指針を提示する．

2. 関 連 研 究

映像に対するコメント付与や Weblog への引用といったサービスは YouTube☆や Google Video☆☆など既にいくつか限定的ながら提供されている．これらのサービスでは，個人所有の映像コンテンツを Web 上に公開し，コンテンツ閲覧者がその映像に対してのコメント付与による掲示板型コミュニケーションや個人の Weblog への埋め込みなどが日常的に行われている．これらのコミュニケーションを映像に対するアノテーションとして捉えることは可能であるが，アノテーションの対象がコンテンツ単位であるなど粒度が荒く，映像のシーン検索などの応用に利用することは困難であり，限定的な応用にしか利用できない．

映像のシーン単位に対するアノテーションの例としては，iVAS[17]，SceneNavi[13),16)]，MPEG-7[7] などが存在するが，映像コンテンツと Weblog などの外部の Web サイトとを詳細に関連付け，そこからアノテーションを抽出しようとする試みはない．コンテンツに関連するコミュニティは，Weblog 等の他の関連するコミュニケーションシステムにも分散する可能性が高く，そこに重要な知識が存在している可能性も高い．

また，映像と外部の Web サイトを関連付けてアノテーションを抽出する研究の例としては，Dowman ら[5] による，ニュース映像の音声認識結果と CNN の Web ニューステキストの内容を比較することによって自動的に該当するニュース記事を特定し，その記事から映像コンテンツに関連した情報の取得を試みる仕組みがあるが，ニューストピック単位での関連付けであるため粒度が荒く，映像コンテンツはニュース記事に限定され，また，音声認識や言語解析結果に極めて依存したリンクであるため，そのリンク自体の精度や再現性も高くない．

3. コンテンツの形式化

一般に映像コンテンツはバイナリデータ列であるため，意味内容を考慮した上で柔軟に扱うことは困難である．コンテンツコミュニティからアノテーションを効率よく取得するためには，機械や人間にとって扱いやすい枠組みを提供することが望ましい．そこで，HTML コンテンツの管理・配信・機械的処理等で一定の成果をあげている Weblog の仕組みを参考にした．

3.1 Weblog に学ぶ

Weblog では，エントリー毎に，Permalink[1)]，Trackback[3)] 等の仕組みを実装することによって異なるサイトにまたがるエントリー間のリンク付けを可能にしている．また，XML Feed[2),6)] の仕組みを利用することによって，コンテンツの情報を機械が理解可能な形で積極的に配信している．このような仕組みを実装することによって，Weblog コミュニティは急激な発展を遂げることが可能になり，RSS リーダや Weblog 検索などの様々な応用を生み出してきた．つまり，これらの仕組みを映像コンテンツのシーンに対して適用することによって，既存の Web や Weblog エントリーなどと親和性が高く，映像シーン単位でのリンクや処理が可能になる．

3.2 映像シーンとショットの定義

本研究では，図 1 のように，映像は複数のショットからなるリストであると定義する．ショットは，一般に映像のカット（切れ目）から次のカットまでの時間範囲を示すが，必ずしもカットが意味的な内容の切れ目であるとは限らないので，長いショットは一定時間間隔に分割しても良いこととする．本システムでは間隔を 2 秒とした．また映像を Web 上でより扱い易くするために，それぞれのショットの内容を表すサムネイル画像をあらかじめ用意する．シーンとは，複数の連続するショットからなり，意味的につながりを持っているものと定義する．ひとつのショットが複数のシーンに属することも許す．

3.3 映像シーンに対する Permalink

映像の任意のシーンに対してアノテーションなどの処理を施すためには，それらのシーンに対して固有の Permalink を記述できる必要がある．そこで，本研究

☆ http://www.youtube.com/
☆☆ http://video.google.com

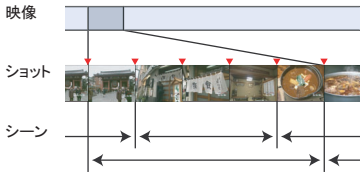

図 1 本論文における，映像のシーンとショットの定義
Fig. 1 Definition of Video Scene and Video Shot

では梶ら[14] によって提案されている Element Pointer の仕組みを採用した．Element Pointer は任意のコンテンツの部分要素に対して URI を関連付ける仕組みであり，それぞれのコンテンツの URI が一意であることが保証されている．

映像コンテンツ全体に対する Permalink は以下のように，固有の ID を用いた URI を記述する．

```
http://[server]/[content ID]
```

また，任意のシーンに対する Permalink は，以下のように固有の ID とその時間区間を記述する．複数の時間区間に対する Permalink を記述する場合は，コンマで区切って複数記述する．

```
http://server/[content_id]#epointer(
urn:aps:timeline(begin,end),
urn:aps:timeline(begin,end), ...)
```

これらの仕組みにより，映像の任意の時間区間に対して，固有の Permalink を記述することができる．

4. 映像シーンへのアノテーション

アノテーションには，直接的なアノテーションと，間接的なアノテーションの二つのタイプが存在する．直接的なアノテーションとは，従来からあるコンテンツの属性情報や構造情報・意味情報など，検索や要約等の応用を目的としたアノテーションである．間接的なアノテーションとは，コンテンツの意味情報を，コンテンツに付随するコミュニケーションやコンテンツに対するレビュー記事の執筆などの編集履歴から間接的にアノテーションとして捉える仕組みである．本論文では，主にオンライン上での間接的アノテーションについて扱う．

4.1 映像シーンへのテキストアノテーション

ユーザがコンテンツの任意のシーンに対して容易にコメントの付与などのアノテーションを可能にする仕組みが必要である．そのために，筆者らが以前の研究で作成したオンラインビデオアノテーションシステム iVAS[17] の仕組みを発展させて利用する．

ユーザは，ネットワークからアクセス可能な任意の映像コンテンツに対して，Web ブラウザを用いてアノテーション及び閲覧を行う．本研究では，シーンに対してコメントを記述することをシーンテキストアノテーションと呼ぶ．図 2 に示すように，映像の任意のショットに対してコメントを付与できる簡便なインタフェースであり，映像の閲覧を継続したままアノテーションを付与可能である．

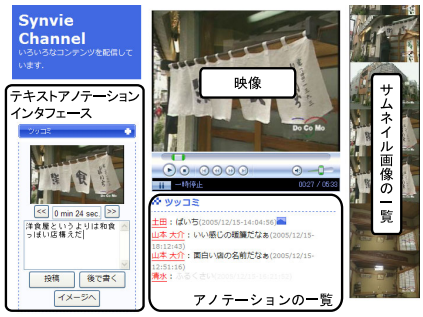

図 2 シーンテキストアノテーション．ユーザは映像の任意のショットに対してコメントを付与可能である．また，現在の映像に同期したアノテーションを表示可能である．
Fig. 2 Scene Text Annotation

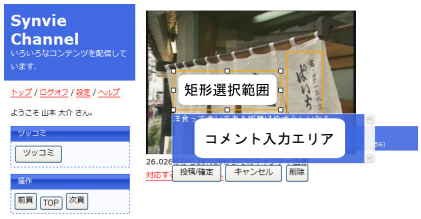

図 3 シーン領域テキストアノテーション
Fig. 3 Scene Region Text Annotation

これにより，ユーザは映像コンテンツに対して，電子掲示板感覚で他のユーザとコミュニケーションを図ることが可能になる．

4.2 映像シーン領域へのテキストアノテーション

シーン領域テキストアノテーションとは，図 3 のように，任意の映像シーンの任意の矩形範囲に対してコメントを付与するためのインタフェースである．対象となるシーン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付与する．これにより，映像の任意のショットの矩形領域を対象としたアノテーションの付与が可能になる．このインタフェースでは，映像の閲覧を一時的に停止する代わりに，より詳細で対象が明確なアノテーションを付与可能である．

4.3 映像シーンへのボタンアノテーション

次に，映像シーンに対するより簡易なアノテーションとして，二種類のボタンによるアノテーションを提案する．一つは，映像に対するマーキングとしての機能であり，任意のシーンに対して “チェック” を行う仕組みである．これは，次章で述べる映像シーンの引用の手がかりとして用いられ，他のユーザとの共有は行わない．二つ目は，iVAS[17] において提案されたシーンボタンアノテーションである．シーンボタンアノテーションでは，映像の任意の時間に対してマウスクリックであらかじめ用意された閲覧者の主観的な印象を表すボタンを押すことによって統計的に評価する仕組みである．本システムでは，nice と boo の二種類のボタンを用意した．インタフェースを図 4 に示す．

図 4 シーンボタンアノテーション.
Fig. 4 Scene Button Annotation.

4.4 コンテンツへのアノテーション

映像シーンに対するアノテーションだけでなく，YouTube などの従来の動画共有サイトで一般的に行われている，コンテンツ全体に対するコメント投稿の機能も実装した．これによって取得されるアノテーションを，コンテンツテキストアノテーションと呼ぶ．また，タイトル情報などあらかじめコンテンツに埋め込まれているメタデータも，コンテンツの内容を示す重要な情報でありアノテーションとして扱う．

5. 映像シーンの引用に基づくアノテーション

次に，映像シーンを引用する仕組みを提案する．Web 上でテキストを引用することはしばしば見受けられるが，任意の映像シーンを効率よく引用する仕組みはあまりない．そこで，アノテーションの取得を前提とした映像シーンの引用の仕組みを提案する．ここで取得したいアノテーション情報は，映像の任意のシーンと関連する Weblog エントリーの文書構造とを結びつけたリンク情報とそのコメントである．そこで，ユーザが映像コンテンツの任意のシーンを容易に引用し，そのコンテンツに関する記事（コンテンツ引用記事）の作成を支援する仕組みを提供する．特に，ビデオコンテンツを引用した記事をビデオブログと呼ぶことにする．映像シーンの引用は，シーンの内容を表すサムネイル画像とそのシーンに対するリンク及びコメントの組からなる．映像シーンを引用して，それに対応するコメントを記述することは，間接的に，映像シーンとコメントを関連付けるアノテーションとして捕らえることが可能であり，シーン引用アノテーションと呼ぶ．

5.1 引用シーンの選択

ユーザはコンテンツを閲覧する際，自身にとって興味のあるシーンに対してシーンテキストアノテーションやシーンボタンアノテーションなどの何らかのアノテーションを施すことによって手がかりを残す．システムは，これらの手がかりをコンテンツ引用記事の執筆のための引用シーン候補としてユーザに提示する．

5.2 コンテンツ引用記事の編集

ユーザが，Weblog などで通常のエントリーを書くのと同様に，一般的な Web ブラウザを用いてコンテンツ引用記事の編集が可能になる仕組みを提案する．

本研究では，二つの編集インタフェースを提案する．一つ目は，連続する映像シーンを引用するのに適した編集インタフェース (図 5) である．これは，引用シーンをショット単位で時間的に展開させることで引用シーンの時間範囲を伴う修正・変更が可能であり，より正確に選択することが可能なインタフェースである．このインタフェースは，シーンの流れやストーリを対象とした記事を記述するのに適したインタフェースである．連続する映像シーンとブログエントリー上の対応するパラグラフ上のコメントとを関連付けることを連続シーン引用アノテーションと呼ぶ．

図 5 連続シーン引用アノテーションインタフェース
Fig. 5 Continuous Scene Quotation Interface

二つ目は，複数の非連続な映像シーンを引用するのに適した編集インタフェース (図 6) である．過去にユーザが施したシーンテキストアノテーションやシーンボタンアノテーションに対応するショットが右側のストックに保持されており，その中から任意のショットをドラッグアンドドロップ形式で複数選択し，その複数のショットに対してコメントを付与することが可能なインタフェースである．これは，複数の連続しないショットに対してコメントを記述することに適したインタフェースであり，シーンやストーリよりも特定のオブジェクト（たとえば特定の人物など）を対象とした記事を記述するのに適したインタフェースである．また，映像シーン検索機能と併用することで，他のコンテンツのシーンの引用も可能である．これによって取得されるアノテーションを非連続シーン引用アノテーションと呼ぶ．

ユーザはこの二つのインタフェースを使い分けながら記事を作成可能である．

このようにして編集された記事は HTML 文書として生成され，任意の Weblog サイトに投稿されると同時に，アノテーションデータベースに蓄積される．

6. 実験と評価

アノテーションシステムの評価及び映像コンテンツに関するアノテーションの収集のために，本論文で提案した Synvie☆の公開実験を行った．2006 年 7 月 1 日から公開を開始し，2006 年 10 月 22 日までの期間において収集されたデータに基づき評価を行う．この期間に，登録ユーザ数 97 人，投稿コンテンツ 94 個，1 コンテンツあたりの平均メディア時間は 321.5 秒，総閲覧数は 7318 回に達した．収集されたアノテーショ

☆ http://video.nagao.nuie.nagoya-u.ac.jp/

図 6 非連続シーン引用アノテーションインタフェース
Fig. 6 Discontiunuous Scene Quotation Interface

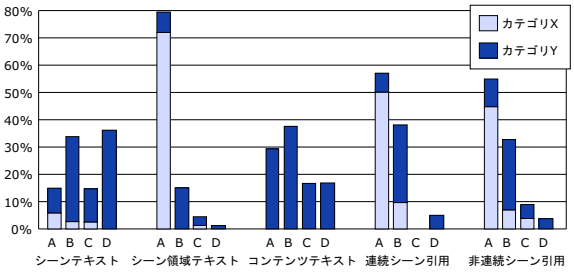

図 7 アノテーションタイプ毎のアノテーションの質の比較
Fig. 7 Quality of Annotation in Each Method

ンは，表 1 で示すように計 4768 個であった．なお平均キーワード数とは，1 アノテーションあたりの茶筌[15]によって取得されたキーワード（名詞と未知語）の数から算出した目安である．

コンテンツテキストアノテーションが YouTube 等の従来システムで実用化されているアノテーション，シーンテキストアノテーションが iVAS 等の従来システムによって取得されるアノテーションと捉え，本論文ではこれらに加えてシーン引用アノテーションを提案している．これらのアノテーションタイプの違いによるアノテーションの質と量を比較することによって，シーン引用アノテーションの有用性を示す．

6.1 アノテーションの質的な分類

収集されたアノテーションを評価するために，それぞれのアノテーションのコメント内容に対して，以下のとおり，アノテーションの意味に基づく分類を行った．

A 主にシーンの内容を説明・解説するコメント．
B 主にシーンに対する直接的な感想や意見などからなるコメントで，シーンに関連するキーワードが含まれるもの．
C 主に，シーンの内容から派生した話題に関するコメント．
D 感嘆符のみ，形容詞のみなど単独では内容を理解できないもの．あるいは，撮影手法，映像の品質に対する感想など，シーンの内容とは関係のない話題からなるコメント．

さらに A，B，C のカテゴリに関して，コメントの文章としての正しさに基づき，

X コメントに主語・述語・目的語が存在するなど，十分に内容を表現している．
Y 十分に内容を表現しているとはいえない．

のサブカテゴリに分類した．なお，分類は二人の評価者によって同時に行い，異なる意見が出た場合には話し合いによる調整を行った．

$A-X$ のアノテーションの例としては，朝顔の展示に対して映像撮影者が自身の Weblog で「名古屋式盆養切込みづくりの朝顔です。 蔓を伸ばさずに盆栽仕立てにしていてとてもユニークです。100 年の歴史があるそうです。」と記述したコメントのように，シーンの内容を的確に表現しており言語解析などを行うことによって，より多くの知識の抽出が期待できる．$A-Y$ のアノテーションの例としては，Web アプリケーションのデモ映像で画像のアップロードを行っているシーンに対する「画像のアップロード」というコメントのように，シーンの内容を表現しているキーワードを含んでいるが，十分に内容を表現しきれていないものである．$B-X$ は「私にとっての朝顔は、こういう蔓を上へ上へと伸ばしていくタイプです。」のようにシーンに対しての感想や意見を述べているものであり，$B-Y$ は「どれだけお菓子使うんだよ！笑」のような表現であり，共にシーンの内容に関するキーワードの抽出が期待できる．$C-X$ の例としては映像中に表示される URL のキャプションに対して「リサイクルトナー専門店のようです。 著作権フリーの CG, 音楽を製作されているみたいです。」．$C-Y$ の例としては「長尾先生といえば，アノテーションの研究」などである．これらは，関連する話題について記述しており，必ずしも映像シーンの内容を直接的に表現していないが，シーンに関する補助的な情報としての利用が期待できる．D の例としては「すごっ！」や「キター」など，単独では意味を成さないコメントや，「なんでこの回だけ映像がぶれてるのでしょう？ウィンドウズメディアエンコーダーという無料ソフトで、ノンインターレス化できるので是非。」など映像の品質に関する話題などが含まれる．

アノテーションタイプごとにカテゴリ分けし，集計したものを図 7 に示す．

6.2 考　　察

まず，アノテーションの量の観点から考察する．アノテーションの数は，シーンボタンアノテーション，シーンテキストアノテーション，シーン引用アノテーション，シーン領域テキストアノテーション，コンテンツテキストアノテーションの順に多かった．ここで，アノテーションの量は，そのアノテーションを付与する手軽さや扱いやすさに関係していると仮定する．従来型のコンテンツ全体に対するコンテンツテキストアノテーションよりもシーンテキストアノテーションの方が多いため，シーンテキストアノテーションは手軽なアノテーションであったと推察できる．一見すると，コンテンツテキストアノテーションの方が，シーンを選択する手間が無い分手軽であるように感じられるが，

表 1 公開実験によって取得されたアノテーション
Table 1 Result of Open Experiment

対象単位	行為	データ型	アノテーションタイプ	取得数	平均キーワード数
コンテンツ	投稿	文	コンテンツテキスト	40	0.78
シーン	投稿	ボタン情報	シーンボタン	3412	0
		文	シーン領域テキスト	187	2.5
			シーンテキスト	795	1.6
	引用		連続シーン引用	283	7.9
			非連続シーン引用	51	4.1

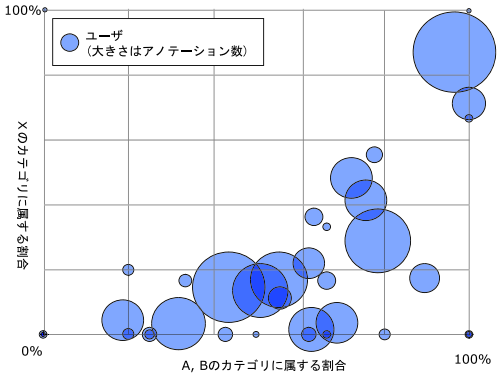

図 8 アノテーションを施した数及び品質に基づくユーザの分布．一つの円が一人のアノテータにあたり，円の大きさが投稿したアノテーションの数にあたる．右上に行くほど質の高いアノテーションを施したユーザである．
Fig. 8 Quality of Annotation in Each User

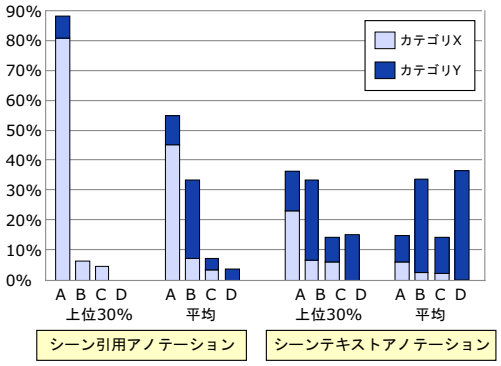

図 9 アノテーションタイプ及びユーザ毎のアノテーションの質の比較．上位 30%とは，サブカテゴリ X に属するアノテーションを施した割合が多い人，上位 30%を示す．
Fig. 9 Quality of Annotation in Each Method and User

シーンテキストアノテーションの方が，アノテーション対象を現在再生中のシーンに限定しているため，他の閲覧者と話題を共有し易く比較的短いコメントで内容を記述できる，些細な問題や話題でもコメントを投稿し易いなどの理由から，よりアノテーションを付与しやすいためだと推察できる。

次に，アノテーションの質の観点から考察する．厳密な質の定義は応用に依存するが，ここでは，コメント内容の品質が高くシーンの内容を的確に表現し，引用シーンに関連するキーワード等を含んでいるものとする．具体的には，$A > B > C > D$ の順で質の高いアノテーションであるものとし，また，サブカテゴリ X の方が Y よりもアノテーションの質が高いものとする．この観点からみると，図 7 で示すように，シーンテキストアノテーション・コンテンツテキストアノテーションに比べて，連続 / 非連続シーン引用アノテーションの質の方が高い．特に，シーンテキストアノテーションにおいてサブカテゴリ X に属する割合は 11%なのに対し，シーン引用アノテーションは 59%になり，より正確な文章が記述されていること，また，シーンテキストアノテーションにおいてカテゴリ D に属する割合が 36%も存在しているのに対して，シーン引用アノテーションの場合は 4.8%であるなど，無関係なコメントや “荒し” と呼ばれるコメントが少ないなどの点で，シーン引用アノテーションの方がより質が高い傾向があるといえる．

つまり，アノテーションの質や量はアノテーションタイプに依存する．これは，閲覧者が映像を見ているという前提が成り立ち，その場限りのコミュニケーションを目的としたシーンテキストアノテーションよりも，映像コンテンツを閲覧しているとは限らない不特定多数に向けた Weblog エントリーの執筆を目的としたシーン引用アノテーションの方がより丁寧な文章を記述する傾向があり，より質の高い情報を記述していると捉えることができる．また，掲示板よりもブログの方が一般的により良い文章が書かれている現状を反映した結果ともいえる．一見面倒で操作が多いアノテーションも，ブログを書く等といった人間の自然な日常活動の一部として取り込む事ができれば，十分な質と量を伴うアノテーションの取得が可能になることが分かる．

次に，アノテーションと人との関連性を考察する．図 8 に示すように，良いアノテーションを施す人もいれば，そうでない人もいる．つまり，人に応じてアノテーションの質や量は異なり，ばらつきがある．そこで，アノテーションタイプと人との関係を考えてみる．図 9 に示すように，質の高いアノテーションを施す人（ここでは，サブカテゴリ X のアノテーションを付与した数の割合が多い人，上位 30%）がシーン引用アノテーション方式を用いて施したアノテーションのうち 80%が一番質の高いカテゴリである $A - X$ に分類される．これは，シーン引用アノテーション全体の平均の 45%や，質の高いアノテーションを施した上位 30%の人がシーンテキストアノテーションを用いて付

与した平均 22%よりも圧倒的に多い．

これにより，多くのアノテーションが集まった場合は，評価が高いユーザが施したシーン引用アノテーションを重視し，あまりアノテーションが集まらなかった場合は，シーンテキストアノテーションの情報も活用するなど，場合によって使い分ける事が可能になる．

つまり，アノテーションの量と質は人にもアノテーションタイプにも依存する．逆に言えば，人やアノテーションタイプが，アノテーションの質の推定パラメータの一つとして利用することが可能になる．具体的なアルゴリズムは現状ではデータが十分揃っていないため今後の課題としたいが，学習アルゴリズムを用いてパラメータを動的に決定することを検討している．

まとめると，YouTube 等で実用化されているコンテンツテキストアノテーションよりも，筆者らが iVAS で提案してきたシーンテキストアノテーションの方がより多くのアノテーションを収集することが可能であり，また，シーンテキストアノテーションよりも，本論文で提案したシーン引用アノテーションの方がより質の高いアノテーションの収集が期待できる．シーンテキストアノテーションとシーン引用アノテーションを併用することによって，質・量とも高いアノテーションの収集が可能になる．また，シーン引用アノテーションの方がより対象コンテンツを詳細に選択できる仕組みである観点から見てもより質が高いといえる．

7. アノテーションの解釈とその応用に向けて

本システムでは，テキストアノテーションやシーン引用アノテーションを，なるべく情報劣化がない形式で蓄積する．そのため，本研究で意味するところのアノテーションは事実の列挙にすぎず，それ自身が機械によって理解可能な情報とは限らない．つまり，本研究によって取得されたアノテーションを用いたアプリケーションを構築するためには，アノテーションに何らかの解釈を与え，機械が理解可能な情報に変換した上で，演算を行う必要がある．具体的な応用については今後の課題とするが，アノテーションの解釈の一例について本章で議論する．

7.1 テキストアノテーションからの意味情報の獲得

映像シーンに対するテキストアノテーションは，図 10 のように，対応する映像シーンとコメントとを「シーンテキストアノテーション」というラベルのついたグラフで表現される．コメントには映像シーンに関する情報を含んでいる場合が多く．間接的に映像シーンに対するアノテーションとして利用可能である．

7.2 コンテンツ引用記事からの意味情報の獲得

コンテンツ引用記事は，図 11 のように，引用した映像シーンと Weblog エントリーのパラグラフとを「シーン引用」というラベルのついたグラフで表現される．これにより，二つの観点からアノテーションとして知識の獲得が可能である．

一つ目は，テキストアノテーションとしての解釈である．引用シーンに対する Weblog エントリー上で

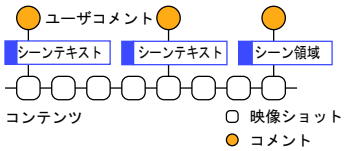

図 10 映像シーンへのアノテーションのモデル
Fig. 10 Video Scene Annotation Model

のコメントは，映像シーンに関連する情報を含んでいると考えられ，引用シーンに対するテキストアノテーションとして捉えることが可能になる．また，映像を閲覧しつつリアルタイムに投稿されたテキストアノテーションには，誤字脱字や，言葉足らずなコメントが少なからず存在する可能性が高いが，Weblog 上で書かれたコメントはより質の高いコメントである可能性が高い．

二つ目は，複数の映像シーンを意味的に関連付けるという解釈である．連続シーン引用アノテーションによって選択された連続するショットからなる引用シーンでは，それに対応するコメント内容という観点に基づきシーンの連続性があるとみなすことができる．また，非連続シーン引用アノテーションを用いて選択されたショットの集合は，対応するコメントの意味内容という観点に基づいて，シーンの関連性があると考えられる．また，一つの Weblog エントリーで複数のコンテンツを引用した場合，その Weblog エントリーの内容に基づいて，これらのコンテンツの意味的な関連性があると捉えることが可能になる．複数のコンテンツを引用した Weblog エントリーの例としては，CG アニメーション「ノラネコピッピ 1 話」と実写映像である「ノラネコピッピのモデルになった猫♪」を同時に引用し比較する記事などである．

7.3 Weblog・マルチメディアコンテンツネットワーク

本システムにより，Web と映像コンテンツの垣根を越えた引用に基づく詳細なネットワークを形成する．これにより Weblog ネットワークにコンテンツを組み込むことが可能になる．Weblog・マルチメディアコンテンツネットワークでは，コンテンツを扱う粒度をコンテンツ/エントリー単位から映像シーン/パラグラフ単位へとより詳細に，コンテンツに関連するコミュニティをサイト内から Web 全体に広げ，コンテンツ間のリンクをナビゲーションのための Hyperlink から引用に基づく意味的なリンクへと，より詳細なコンテンツネットワークを構築する．これにより，コンテンツに付随する様々な知識を抽出するためのフレームワークとして機能し，検索やコンテンツ推薦などの様々な応用のための基礎的データとして利用されることが期待できる．

8. おわりに

本論文では，映像シーンへのアノテーション，映像シーン単位でのコンテンツの引用に基づく Weblog エ

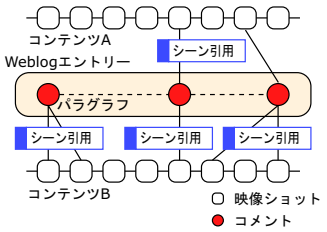

図 11 映像シーン引用に基づくアノテーションのモデル
Fig. 11 Annotation Model Based on Quotation of Video Scenes

ントリーからのアノテーション取得方法の提案，コミュニケーションに特化した具体的なインタフェースの提案と公開実験に基づく評価を行った．これにより，それぞれのアノテーションタイプによって得られるアノテーションの傾向をアノテーションの量と質の観点から分析を行い，それぞれのアノテーションに特有の傾向が見られることが分かった．特に，関連するWeblog エントリーから情報を抽出することが質の高いアノテーションを抽出する手助けになることが示せたことが有用であると考えている．また，これらのアノテーションは，二つの観点により映像の構造的情報も抽出可能である．一つは，コンテンツを引用することによってそれぞれのショット間の意味的な関係の抽出が期待できる．もう一つは，引用によって複数のコンテンツ間の意味的な関係の抽出が期待できる．

なお，本論文では主に映像コンテンツを対象として述べているが，同様な考え方は音楽や画像など他の様々なコンテンツに対しても有効である．今後は，収集されたアノテーション情報を用いて，ビデオシーン検索等のより具体的なアプリケーションの構築とその評価を行う予定である．

謝　　辞

本研究は独立行政法人情報処理推進機構 (IPA）による 2005 年度上期未踏ソフトウェア創造事業の支援を受けた．

参 考 文 献

1) Aimeur, E., Brassard, G. and Paquet, S.: Using Personal Knowledge Publishing to Facilitate Sharing Across Communities, *Proceedings of the Twelfth International World Wide Web Conference (WWW2003)* (2003).
2) Beged-Dov, G., Brickley, D., Dornfest, R., Davis, I., Dodds, L., Eisenzopf, J., Galbraith, D., Guha, R., MacLeod, K., Miller, E., Swartz, A. and van der Vlist, E.: *RDF Site Summary (RSS) 1.0*, RSS-DEV Working Group, http://web.resource.org/ rss/1.0/spec (2001).
3) Benjamin and Trott, M.: *mttrackback - TrackBack Technical Specification*, movabletype.org, http://www.movabletype.org/docs/ mttrackback.html (2002).
4) Davis, M.: An Iconic Visual Language for Video Annotation., *Proceedings of the IEEE Symposium on Visual Language*, pp. 196–202 (1993).
5) Dowman, M., Tablan, V., Cunningham, H. and Popov, B.: Web-Assisted Annotation, Semantic Indexing and Search of Television and Radio News, *Proceedings of the The 14th International World Wide Web Conference 2005 (WWW 2005)*, pp.225–234 (2005).
6) Hoffman, P. and Bray, T.: *Atom Publishing Format and Protocol (atompub)*, http://www.ietf.org/html.charters/atompub-charter. html (2005).
7) MPEG: *MPEG-7*, MPEG-7 Consortium, http://www.mp7c.org/ (2002).
8) Nagao, K., Ohira, S. and Yoneoka, M.: Annotation-Based Multimedia Summarization and Translation, *Proceedings of the Nineteenth International Conference on Computational Linguistics (COLING-02)*, pp.702–708 (2002).
9) Nagao, K., Shirai, Y. and Squire, K.: Semantic Annotation and Transcoding: Making Web Content More Accessible, *IEEE MultiMedia*, Vol.8, No.2, pp.69–81 (2001).
10) Ricoh: *Movie Tool*, http://www.ricoh.co.jp/src/multimedia/MovieTool/ (2002).
11) Smith, J.R. and Lugeon, B.: A Visual Annotation Tool for Multimedia Content Description, *Proceedings of the SPIE Photonics East, Internet Multimedia Management Systems*, pp. 49–59 (2000).
12) Wactlar, H.D., Kanade, T., Smith, M.A. and Stevens, S. M.: Intelligent Access to Digital Video: Informedia Project, *IEEE Computer*, Vol.29, No.5, pp.140–151 (1996).
13) 山田一穂, 宮川　和, 森本正志, 児島治彦：映像の構造情報を活用した視聴者間コミュニケーション方法の提案, 情報処理学会研究報告 2001-GN-43, Vol.43, No.7, pp.37–42 (2001).
14) 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, Vol.48, No.1, pp.258–273 (2007).
15) 奈良先端科学技術大学院大学自然言語処理学講座：形態素解析システム茶筌, http://chasen.aist-nara.ac.jp/ (2003).
16) 鷲田　聡, 宮川　和, 森本正志：ファンコミュニティサイトにおける映像シーン連動型掲示板コミュニケーションの分析, 電子情報処理学会技術研究報告 (HCS2005-50), pp.69–74 (2005).
17) 山本大介, 長尾　確：閲覧者によるオンラインビデオコンテンツへのアノテーションとその応用, 人工知能学会論文誌, Vol.20, No.1, pp.67–75 (2005).