原著論文

Original Paper

タグクラウド共有に基づく協調的映像アノテーション

Collaborative Video Annotation by Sharing Tag Clouds

山本 大介
Daisuke Yamamoto
名古屋工業大学大学院工学研究科情報工学専攻
Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology
daisuke@nitech.ac.jp, http://tk-www.elcom.nitech.ac.jp/~daisuke/

増田 智樹
Tomoki Masuda
名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya Univeristy

大平 茂輝
Shigeki Ohira
名古屋大学情報基盤センター
Information Technology Center, Nagoya Univeristy
ohira@nagoya-u.jp, http://www.nagao.nuie.nagoya-u.ac.jp/members/ohira.xml

長尾 確
Katashi Nagao
名古屋大学大学院情報科学研究科メディア科学専攻
Department of Media Science, Graduate School of Information Science, Nagoya Univeristy
nagao@is.nagoya-u.ac.jp, http://www.nagao.nuie.nagoya-u.ac.jp/members/nagao.xml

keywords: video annotation, video sharing service, video retrieval, Web service

Summary

In this paper, we propose a video scene annotation method based on tag clouds. First, user comments associated with a video are collected from existing video sharing services. Next, a tag cloud is generated from these user comments. The tag cloud is displayed on the video window of the Web browser. When users click on a tag included in the tag cloud while watching the video, the tag gets associated with the time point of the video. Users can share the information on the tags that have already been clicked. We confirmed that the coverage of annotations generated by this method is higher than that of the existing methods, and users are motivated to add tags by sharing tag clouds. This method will contribute to advanced video applications.

1. は じ め に

近年，YouTube[*1]などの映像共有サービスの普及に伴い，膨大な映像コンテンツが Web 上で公開されている．これらの映像コンテンツには大変面白いものや価値があるものも含まれているため，効率よくこれらのコンテンツを検索をしたい要求がある．既存の映像共有サービスにおける映像検索手法は，投稿者によって付与されたコンテンツのタイトルやコメント・タグなどのメタ情報を，テキストキーワードを用いて検索することによって実現されている．このような手法では，映像シーンの内容に基づく検索は実現できない．

本研究における映像シーンとは，映像の切れ目（カット）で区切られた明示的な映像区間（ショット）とは限らず，内容に応じて細かく分割された部分映像として定義する．なぜならば，同一ショットでも映像や内容は変化するため，ショット単位では長すぎる場合があるためである．

映像シーン (つまり映像内部の特定のシーン) を検索するためには，映像シーンの内容を記述したアノテーション [Nagao 01, 井手 09] の付与が有効である．従来のアノテーション手法には，自動アノテーション方式 [Wactlar 96] と半自動アノテーション方式 [Davis 93, Smith 00, Nagao 02] の 2 種類がある．自動アノテーション方式は，音声認識や映像認識技術に基づく方式であり，映像や音声の品質が高いニュース映像などのコンテンツには有効であるが，Web に投稿されている，手ぶれやノイズが多い映像コンテンツには必ずしも有効とはいえない．他方，半自動アノテーション方式とは，MPEG-7[Int 01] のような，専用ツールを用いて，専門家が映像の内容を表すメタ情報を記述する方式である．人が内容を判断するので，映像の画質に依らず，より精度の高い詳細な情報を記述することができるが，人的コストが高いため，膨大な Web 映像コンテンツの全てには適用できない．

これらの問題に対処するために，一般の閲覧者がアノテーションを作成する手法が提案 [Uehara 06, Miyamori 05] されている．我々も，Web ユーザが自発的に映像の内容に対してコメントを関連付ける仕組みとして Synvie[山本 07, Yamamoto 08b] や iVAS[山本 05] を提案してきた．

*1 YouTube. http://www.youtube.com/

本研究の目的は，映像シーン検索に適したアノテーションを効率よく獲得できるシステムを提案することである．ここで，我々が想定する映像シーン検索とは，以下の 3 つの要求を満たすものである．

- **要求 1** ユーザが，HTML を対象とした Web 検索と同様に，固有名詞を含むさまざまなキーワードで映像を検索できること．
- **要求 2** ユーザが，映像コンテンツ単位だけではなく，映像シーン単位で映像を検索できること．
- **要求 3** ユーザが，通常の HTML を対象とした Web 検索と同様に，入力したキーワードと映像シーンとの関連性に応じて，ランキングされた結果として映像シーン検索結果を獲得できること．

これらの要求を満たしたアノテーションが獲得できるシステムを実現するための必要条件として，少なくとも以下の 3 つの要件を満たす必要がある．

- **要件 1** 映像コンテンツ毎に，関連した多様なタグが獲得できる仕組みであること．(要求 1 に対応)
- **要件 2** 要件 1 を満たすことによって獲得された多様なタグを，同一コンテンツ内の映像シーンに満遍なく関連付けることができる仕組みであること (網羅性があること)．映像 A のシーン B をキーワード C で検索したい場合，たとえ映像 A の他のシーン D にキーワード C が付与されていても，シーン B に何もタグが関連付けられていなければ，シーン B を検索結果とすることは困難である．(要求 2 に対応)
- **要件 3** シーンとタグの関連性の評価や選別ができる仕組みであること．単に多くの種類のタグが多くのシーンに付与されているだけでは，検索の再現率は向上するが適合率が向上しない．シーン毎に，それぞれのタグと映像シーンの関連度が，一定の基準に基づいて評価できる仕組みが必要．(要求 3 に対応)

つまり，映像シーン検索を実現するためには，検索キーワードと関連するタグを含むシーンを検索結果の候補として列挙し (要件 1 と要件 2)，シーンとタグの関連度を考慮して (要件 3) 候補シーンをランク付けする処理が必須である．このような情報があれば，例えば，タグと映像シーンとの関連性を用いた既存の映像シーン検索手法 [是津 98, 吹野 02] 等を適用可能になる．この検索手法は，全ての映像シーンに必要な種類のタグが全て付与されていなくても，タグ間やシーンとの関連性を利用することによって，少ないタグで効率よく映像シーン検索が可能になる手法であるが，しかしながら，何もタグが付与されていないシーンを検索できるわけではない．つまり，要件 2 の網羅性が満たされている必要がある．従来手法は，これらの 3 つの要件を同時に支援する仕組みが無いことが，閲覧者のアノテーションに基づく効果的な映像シーン検索の実現が困難な理由の一つである．

本論文では，これらの 3 つの要求を満たすアノテーション手法として，以下の特徴を有するタグクラウドに基づく協調的アノテーション手法 [Yamamoto 08a] を提案する．

- **特徴 1** 映像共有サービスに投稿されている自由投稿コメントから，コンテンツ毎に関連するタグを動的に自動抽出することによって，その映像に関連した多様なタグが自動生成可能である（要件 1 を満たす）．
- **特徴 2** ユーザが映像を閲覧しながら特徴 1 で生成されたタグを，同一コンテンツ内の任意の映像シーンと関連付けることができる簡便なタグクラウドアノテーション手法を提供する．特徴 1 で生成されたタグは，その映像を特徴付けるキーワードが含まれている可能性が高いが，そのタグが，必要とされる，同一コンテンツ内の全てのシーンと必ずしも関連付けられているとは限らない．そのため，提案手法によって，多数の閲覧者がこれらのタグを同じコンテンツの任意の映像シーンに関連付けることにより，より多くのシーンに必要なタグを関連付けることが可能になる．(要件 2 を満たす)
- **特徴 3** 複数のユーザによって映像シーン毎にタグを選択（支持）することができ，また，映像シーン毎のタグの押下状況を共有することができる．これにより，映像シーン毎にタグを選択（支持）したユーザ数に応じて，映像シーンとタグの関連性を評価することが可能である（シーン毎にタグを支持することができる点において，要件 3 を満たす．また，タグの押下状況を共有することによって，副次的にユーザのアノテーションに対するモチベーションを向上させる点において，要件 1 と 2 を満たす）．

本研究では，質の高いアノテーションを追求するのではなく，映像との関連性が少しでもある可能な限り多くのタグを映像シーンと関連付けることを目的としている．

なお，本研究におけるアノテーションとは，映像の内容理解に役立つメタデータ全般のこと，あるいは，それらのメタデータを関連付ける行為と定義する．本研究におけるタグとは，アノテーションの一形態であり，検索などに役立つ映像やその内部シーンの内容を表す短いキーワードのことである．

2. 従来手法の問題点

従来の映像シーンに対する閲覧者によるアノテーション手法は，任意のコメントやタグを関連付ける**コメントアノテーション**手法と，予め用意されたタグをユーザが選択する**ボタンアノテーション**手法の 2 つに分けられる．

コメントアノテーション手法の例として，映像と同期してコメントを入力する手法[*2]や，現在放送中の TV 映像を話題としたチャット文章から映像シーンに対応するキーワードを関連付ける手法 [Uehara 06] などが挙げられる．山本らも，図 1 に示す，映像シーンに対するコメントアノテーション手法を提案してきた．これらの手法

∗2 ニコニコ動画. http://www.nicovideo.jp/

表 1 アノテーション方式の比較.

	要件 1	要件 2	要件 3
コメントアノテーション	○	△	×
ボタンアノテーション	×	○	○
提案手法	○	○	○

はキーボードを用いて入力する必要があるため，映像の閲覧を中止して入力する必要があるなど敷居が若干高い．そのため，任意のタグを関連付けることができる点において要件 1 を満たすが，入力に対する敷居の高さから全ての映像シーンに満遍なくコメントが投稿されることは困難である（ユーザ数が非常に多い場合以外は，要件 2 を満たさない）．事実，図 2 で示すように，コメントアノテーションのメディア時間に対する網羅性は低い．図 2 は横軸が映像の時間軸であり，縦軸は，時間軸を 2 秒毎に区切った時の，その時間区間毎に付与されたユーザコメントから生成されたタグの異なり数である．多いシーンは 20 個のタグがあるが，0 個のシーンも多く，ユーザがコメントを付けるに至ったシーンには偏りがある．また，これらのコメントやタグの映像との関連性を評価・選別するための仕組みは無い（要件 3 を満たさない）．

ボタンアノテーション手法の例として，映像の画面の横に，図 1 の左下のように,「かわいい」「おいしい」「nice」などの専用ボタンを設置し，それを不特定多数の閲覧者が押すことによって映像シーン毎に投票可能な手法が挙げられる．同様な効果をもたらす手法としては，ライブチャットから盛り上がりを抽出して，シーン毎に盛り上がり度を統計的に求める手法 [Miyamori 05] もある．また，アノテーションではないが，地デジの双方向通信機能を用いて番組内でリアルタイム投票を行うことによって，番組を盛り上げる試みも頻繁に行われている．ユーザはタグ付きボタンを押す（支持する）だけなので，映像を視聴しながらアノテーション作成に参加できるなど敷居が低い．事実，Synvie の公開実験において，コメントアノテーションの付与回数が 3534 個に対して，ボタンアノテーションの付与回数は 18854 回と 5 倍以上も多い [山本 07, Yamamoto 08b]．そのため，前者に比べて要件 2 を満たす可能性は高いが，あらかじめ用意したタグ付きボタンの内容しか付与できないため，要件 1 は満たさない．しかしながら，提示されたタグは，そのタグに共感したユーザが押すことを想定しているため，タグの押下回数が，そのまま，そのタグが支持された数と捉えることが可能であり，ボタンの押下回数に応じてタグの評価・選別が可能である (要件 3 を満たす).

つまり，表 1 に示すように，コメントアノテーション手法のように任意の種類のタグが付与でき，ボタンアノテーション手法のように手軽に多く付与され，また，評価・選別できる手法が必要である．

また，増田らは，Synvie で獲得したコメントアノテーションに基づく映像検索システム Divie [Masuda 08] を

図 1 Synvie のアノテーションインタフェース

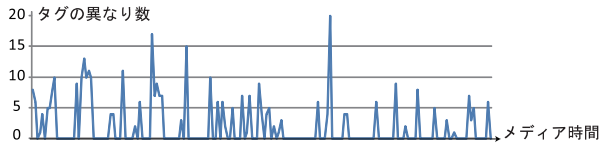


図 2 ユーザコメントの分布（タグ数)

開発した．Divie では，全てのコンテンツに含まれるキーワード（タグ）をタグクラウドの形式でユーザに提示し，ユーザはタグを選択して検索クエリーを入力可能である．キーワードは名詞だけでなく，動詞や形容詞も含まれる．検索キーワード上位 40 件のうち，名詞は 29 件，動詞は 0 件，形容詞は 11 件であるなど，名詞だけでなく形容詞を検索キーワードとして利用することも多かった．これは，映像コンテンツは娯楽作品でもあるため，主観的な検索もよく行われていることを示唆している．これらの主観的なアノテーションは，機械的に認識することが困難であるため，人間によって付与される必要がある．

3. アノテーションシステム

本論文では，一般的な Web ブラウザを用いて，ユーザが映像を閲覧しながら手軽にアノテーション作成に参加できるシステムを提案する．また，簡単な意思表示や閲覧者同士の一体感を生むコミュニケーション手段の一つとしての利用も可能である．

3·1 対象とするコンテンツ

YouTube などの一般的な映像共有システムで使われることを想定しているため，想定するコンテンツは，閲覧者数が数百人〜数万人程度の，一般的な映像コンテンツである．これらの映像コンテンツでは，多くのユーザは映像を見るだけで満足しているが，いくつかのユーザはさまざまなコメントを投稿している．しかしながら，新しいコメントを投稿しないが，何らかの意思表示はしたい，あるいは，動画をただ見るだけではなくて他人と一体感や共感を得たいという人もいるだろう．このようなユー

ザが（全体の 1 割弱を想定），自発的にアノテーション作成に参加することを想定している．また，本アノテーション手法によって取得された結果は統計的に処理することを前提としているため，個人を識別する必要がなく，ユーザにログイン作業を強制することを想定していない．なお，対象とするコンテンツは，映像の内容が時間とともに切り替わる，複数の映像シーンからなる映像コンテンツとする．

また，前提として，Web ユーザは Synvie や YouTube などの映像共有サービスを用いて映像を閲覧し，その映像に興味を持ったユーザは映像に関する記事やコメントを投稿している．Synvie の公開実験*3では，324 人の登録ユーザと，126 個の投稿コンテンツ，映像やシーンと関連付けられた 3534 個の有効なユーザコメントを獲得した．さらに，YouTube などの既存の映像共有サービスにおいても，個々の映像を題材としたコメントやブログエントリーが多数投稿されている．

3·2 提案システムの概要

初めに，図 3 に示すように，ユーザコメントを，4 章 1 節で述べる手法を用いて自動解析することによってタグの集合を獲得する．さらに，それぞれのタグの重要度を 4 章 2 節で述べる手法を用いて推定し，その重要度に応じてそれぞれのタグの大きさを調整することによって，映像コンテンツ毎にタグクラウドを生成する．

次に，提案システムは不特定多数のユーザが任意のタグをクリックすることによって，タグ形式のアノテーションを関連付ける．

提案システムを用いて，タグのクリックにより付与された，タグと映像の関連付け情報は，それぞれの映像シーン (コンテンツ ID, 映像シーン ID) 毎に，対応付けられたタグ (タグ ID) と，それが押下された回数がデータベースに蓄積される．より多くのユーザによって支持されたタグほど，映像シーン検索において関連性の高いタグであると考えている．

ブログなどのユーザコメントからタグを生成している点において，1 章の要件 1 を満たし，タグを選択することによって，映像シーンとタグを関連付けることができる仕組みである点において，1 章の要件 2 を満たし，多くのユーザによってタグが支持することができる仕組みである点において，1 章の要件 3 を満たす．具体的な検証については，5 章で述べる．

3·3 アノテーション手法

具体的なインタフェースを図 4 に示す．画面左側に映像が表示され，右側にタグクラウドが表示される．表示されるタグクラウドは，ユーザが閲覧する毎に次章の手法を用いて動的に生成されるため，ユーザ毎に異なる．

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

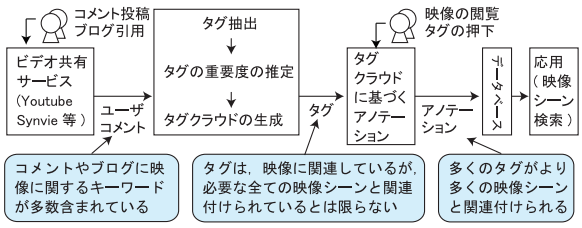


図 3 提案システムの構成

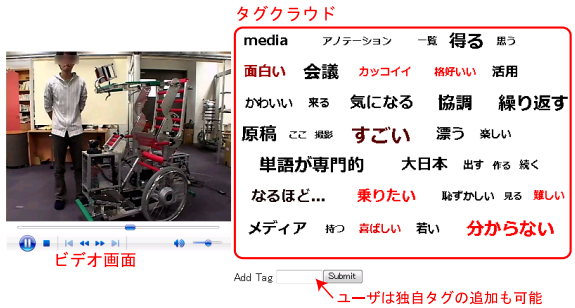


図 4 タグクラウドに基づくアノテーションインタフェース．誰かに支持された（ボタンが押された）タグは，そのシーンにおいてハイライトされる．

ユーザはタグクラウドの中の任意のタグをクリックすることによって，現在閲覧中の映像のメディア時間とタグを関連付けることができる．タグをクリックするとそのタグは赤くハイライトされ，時間とともに徐々にハイライト効果が消える．現在のシステムでは 2 秒で消えるようになっている．ユーザは，何度でも同じタグを押すことが可能である．また，シーン毎に，それぞれのタグは，そのタグをクリックしたユーザ数に応じて赤くハイライト表示される．現在のシステムでは，ユーザ数に応じて黒色から段階的に赤くなり，5 人以上によって支持されると完全に赤くなる．つまり，映像シーン毎にそのタグを支持したユーザがどれくらいいるのかをタグの色で判別することができる．

これらのタグの共有状況の視覚化は，映像のシーンとタグのクリック数を関連付け，これらの情報をデータベースに蓄積することによって，例え，異なる時刻に別々のユーザが閲覧したとしても，あたかも，他のユーザと同時刻に映像を見ているような感覚を実現できる．具体的には，Synvie[Yamamoto 08b] において映像とコメントを同期させる技術と同様に，映像シーン ID とコンテンツ ID をキーとして，AJAX(Asynchronous JavaScript + XML) 技術を用いた非同期通信によりサーバと通信することによって実現している．

また，必要に応じて新しいタグをユーザが追加することも可能である．具体的には，図 4 の Add Tag の項目に任意のキーワードを入力し，Submit ボタンを押すことによって，タグクラウドを構成するタグの一つとして追加される．同時に，タグを追加したときに再生している映像シーンと，タグを関連付けることができる．

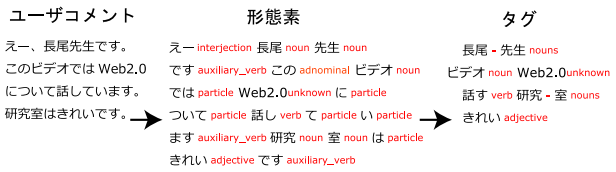


図 5 ユーザコメントからタグの抽出

本インタフェースの利点は，映像の閲覧を中断せずにアノテーション作成に参加できるなど，ボタンアノテーション手法と同様に簡便な手法である．また，タグクラウド形式で提示することによって，従来のボタンアノテーション手法よりも多くのタグを提示している．さらに，タグのクリック状況を視覚的に共有することによって，他のユーザが映像に対してどのような印象や感想をもっているかが分かるため，他のユーザと同じタグを支持することによって共感を呼び，ユーザのアノテーション活動のモチベーションを向上させる効果があると期待している．

4. タグクラウドの生成

4·1 タグ抽出手法

はじめに，Synvie や映像を話題としたブログから収集したユーザコメントからタグを自動生成する．これらのユーザコメントは，映像コンテンツを引用したブログの本文や，映像を話題とした掲示板のコメントから収集したものであり，主に，映像に関する評価や感想などの文章である．これらのユーザコメントは映像全体や，映像の任意の時間範囲と関連付けられている．

想定するタグは,「かっこいい」や「かわいい」などといった閲覧者の主観的な印象を表現する形容詞や,「男性」「パンダ」などといった登場人物やモノを表現する名詞,「走る」「寝る」などといった状況を表す動詞などである．

タグは，図 5 のように，以下の手法によって抽出される．まず，それぞれのコメントを形態素に分解する．形態素のうち，名詞，自立形容詞，自立動詞，未知語をタグとして抽出した．未知語は多くの場合，固有名詞などの名詞である場合が多いので，形式上名詞として扱った．また,「ある」「いる」などのそれ自身で意味を持たない形態素は除外した．それ以外にも不要と考える単語は不要語辞書をつくり除外した．さらに，連続する名詞は複合名詞として結合した．これらの映像コンテンツと関連付けられている形態素がタグである．本論文におけるタグは形態素の品詞情報も保持している．

4·2 タグの重要度の推定

次に，抽出されたタグの重要度を推定する．一般的なタグクラウドはタグの投稿数に応じてタグの文字の大きさを決定している．すなわち，よく利用されるタグほど重要であるという仮定の上に成り立っている．しかしながら，本システムの場合，自動生成されたタグを用いているので，単純に同様のアルゴリズムを用いると，より一般的な単語，たとえば「人」「情報」「モノ」など重要なタグとして解釈されてしまう．特に名詞の場合は，出現頻度の高い単語が重要であるとは限らない．そこで，自然言語処理の分野でよく利用されている tfidf のアルゴリズムを応用した．tfidf とは文章中で重要だと思われる語句を抽出するアルゴリズムであり，tf(単語の出現頻度) と idf(逆出現頻度）によって表現される．ここで，idf とは，一般的な語句を除去するフィルタとして働き，多くのドキュメントに出現する語（一般的な語）の重要度を下げ，特定のドキュメントにしか出現しない単語の重要度を上げる役割を果たす．

映像 c に属するタグ t の出現頻度 $tf_{t,c}$ は，そのタグの映像 c での出現回数 $n_{t,c}$ を用いて，

$$tf_{t,c} = \frac{n_{t,c}}{\sum_k n_{k,c}} \tag{1}$$

と表す．タグ t の逆出現頻度 idf_t は，すべての映像の数 $|D|$ とそのタグが関連付けられている映像の数 $|d : d \ni t|$ を用いて，

$$idf_t = \log \frac{|D|}{|\{d : d \ni t\}|} \tag{2}$$

と表す．これを用いて，ある映像 c におけるタグ t の重要度 $tfidf_{t,c}$ は

$$tfidf_{t,c} = tf_{t,c} idf_t \tag{3}$$

として表現される．タグの重要度 $I_{t,c}$ は，タグが形容詞以外の場合は $I_{t,c} = tfidf_{t,c}$ とし，タグが形容詞の場合は，$I_{t,c} = tf_{t,c}$ とした．形容詞の場合，よく利用される単語は限られており，かつ，そのような単語の重要性は変わらないとの判断からである．

なお，タグの重要度と，要件 3 で述べたタグと映像シーンの関連度は全く違う指標である．タグの重要度は，Synvie などで獲得されたユーザコメントのみから評価したタグの持つ重要性の値である．それに対して，関連度は，シーン毎にタグを支持したユーザ数に応じて決定される，タグとそれぞれのシーンの関連性を示した値である．なお，シーンとタグの関連度の具体的な計算手法は今後の課題である．

4·3 タグクラウドの生成

最後に，タグクラウドを生成する．本研究では，タグクラウドを映像の横に表示し，映像を閲覧しながら Web ユーザが操作することを想定している．そのため非常に多くのタグを列挙すると見にくいため，必要に応じて表示するタグの数を制限する必要がある．タグの数は任意であるが，見やすさを考えると 30 個前後を想定している．重要度が高いタグ 30 個を表示してもよいが，重要度を考慮せずにランダムに 30 個表示することにした．なぜならば，重要度が高い 30 個を限定的に提示するよりも，ランダムな 30 個を多くのユーザに提示する方が，より

表 2 実験用コンテンツ．ショット数とは，カットで区切られた映像区間数．分割シーン数とは，2 秒毎に分割した映像区間数．

	ジャンル	長さ	ショット数	分割シーン数
A	研究紹介	8 分 23 秒	60	251
B	娯楽	12 分 47 秒	4	383
C	動物	2 分 32 秒	16	76
D	CG アニメ	5 分 39 秒	41	339
E	芸術	1 分 41 秒	1	50
F	料理	1 分 16 秒	18	38

多くのタグを選択してもらえるメリットが期待できるからである．つまり，複数ユーザに対してランダムでタグを表示することによって，統計的に平準化されることを期待している．つまり，ここでも複数ユーザによる協調的な効果を期待した．

タグの大きさ $size_{t,c}$ は，タグの重要度 $I_{t,c}$ に基づき，以下のように定義した．ただし，n,m は定数である．

$$size_{t,c} = n \cdot I_{t,c} + m \tag{4}$$

また，単語の視認性を高めるために，あいうえお順にタグを並び替えている．

5. 実験結果と考察

提案手法の有効性を検証するために評価実験に基づく評価を行う．映像シーン検索としての利用を考慮した場合，1 章で述べた 3 つの要件を満たす必要がある．要件 1 は，既に多量に存在する映像に関連したブログやユーザコメントからタグクラウドを生成するため，多様なタグと映像を関連付け可能である点により満たす．しかしながら，提案手法は，前述した通り，機能としては要件 2 や要件 3 は満たすが，その前提として，従来手法であるコメントアノテーション手法よりも多くの映像シーンに対して満遍なくタグが関連付け易い仕組みであることを示す必要がある．そこで，映像シーンに対してタグが関連付けられている割合（網羅性）に着目し，従来のコメントアノテーション手法と提案手法の比較検討を行う．その後，タグの視覚的効果による協調効果を検証する．

被験者は大学生 16 人である．被験者は，あらかじめシステムの使い方を理解しているものとする．対象となるコンテンツは，表 2 に示すように，Synvie に投稿された 1 分〜10 分程度の A から F の 6 コンテンツとする．映像 A はプロが作成した研究室紹介を目的とした映像コンテンツであり，他は素人が作成した映像コンテンツである．これらのコンテンツにはあらかじめ Synvie の公開実験によりユーザコメントが付けられている．また，異なるジャンルの映像を複数用いることによって，コンテンツの特性による影響を平準化する．

5·1 アノテーションの評価

はじめに，従来手法であるユーザコメントアノテーション手法と提案手法を，関連付けられたタグの特性に基づき比較評価を行う．

§1 評価手法

まず，2 つの観点からアノテーションとしての効果を評価する手法について提案する．1 つは，1 つあたりの映像シーンに対してどれだけ多くの種類のタグが関連付けられるか (平均タグ異なり数) という観点である．もう 1 つは，全ての映像シーンのうち，どれだけのシーンに対してタグが付与できているか (網羅性) という観点である．特に，後者は要件 2 を満たすために重要な指標である．つまり，要件 1 と要件 2 を満たすためには，より多くの映像シーンに対して満遍なくタグが付与されることが必要であり，また，それらのシーンに対して多様なタグが付与されていることが望ましい．つまり，平均タグ異なり数と網羅性の両方が高い方がより良いといえる．なお，本実験におけるシーンとは映像を 2 秒毎に機械的に分割した部分映像であるとする．本来ならば，映像シーンは映像の意味内容に応じて分割することが望ましい．しかしながら，実際の利用を想定した場合，これらのシーンを自動で高い認識率で獲得することは極めて困難である．そのため，シーンを経験的に十分に短い時間であろうと想定される 2 秒で機械的に分割した．

まず，映像 x のタグが関連付けられている各シーンに対して，平均いくつのタグが付与されているかを表す関数 $Taverage_x$ を以下の式で表現する．

$$Taverage_x = \frac{\sum_{S'_x \ni s} T(s)}{|S'_x|} \tag{5}$$

ただし，S'_x は映像 x に含まれ，一つ以上のタグが付与されているシーンの集合であり，また，$T(s)$ はシーン s に付与されたタグの異なり数，つまり，シーン毎に含まれるタグの種類の数である．

なお，一つ以上のタグが付与されたシーン集合を対象とする理由は，もしも，すべてのシーンを対象として計算すると，多くのシーンにタグを関連付けやすい提案方式に対して，シーンに対してまばらにしか付与されていないコメントアノテーション手法の $Taverage_x$ の値が極端に小さくなってしまい，コメントアノテーション手法と提案方式をフェアに比較できないためである．

次に，ある映像 x に対して，どのくらいのシーンにタグが付与されているか（網羅性）を表す関数 $Tcover_x$ を以下の式で表現する．

$$Tcover_x = \frac{\sum_{S_x \ni s} K(T(s))}{|S_x|} \tag{6}$$

ただし，S_x は映像 x に含まれるすべてのシーンの集合である．また，$K(u)$ は以下の式で表す．

$$K(u) = \begin{cases} 1 & u > 0 \\ 0 & u \leq 0 \end{cases} \tag{7}$$

なお，本実験では，コンテンツによる違いを吸収するために，映像AからFの評価値の平均を，提案手法の評価値として用いた．

§2 実験結果と考察

図 6 は $Taverage_x$ による評価値を，図 7 は $Tcover_x$ による評価値を表す．本システムでは，アノテーション作成に参加するユーザ数に応じて，シーンとタグの関連付け数が変化するため，表の横軸にアノテーション作成への参加数，縦軸に評価値としたグラフで示している．なお，今回の実験で用いたコンテンツの中で，ユーザコメントを付与した人数のうち，もっとも少ない人数が 11 人であったため，ユーザコメントのデータはユーザ数が 11 人までしか提示していない．

図 6 で示す，シーンに付与される平均タグ異なり数について考察する．コメントアノテーション手法よりも提案手法の方が参加するユーザあたりの取得可能なタグ数が 3 から 5 倍程度少ない．この理由は，コメントアノテーション手法では一つのシーンに対して長いコメントを書くと，より多くのタグが一度に関連付けることができるためである．しかしながら，提案手法はボタンを押すだけといった簡便なインタフェースであるため，より多くのユーザがアノテーション作成に参加することが期待できれば，平均タグ異なり数は増加すると考えられる．事実，今回の実験においては，提案手法ではユーザ数に応じて単調にタグ数が増加しているため，ユーザが増えればより多くのタグが獲得できることが期待できる．

次に，図 7 において，タグの網羅性について考察する．コメントアノテーション手法の網羅性は 40%程度で頭打ちであるのに対して，提案手法では，15 人程度の参加で 100%に近くなる．これは，コメントアノテーションは特定の興味深いシーンにコメントが集中し易いのに対して，提案手法は容易なアノテーション手法であるため，ボタンアノテーション手法と同様に，あまり重要ではないシーンに対してもタグを関連付けていることを示す．また，1 コンテンツあたり，15 人程度のユーザが提案手法に基づくアノテーションに参加することができれば，図 6 と図 7 より，ほぼ全てのシーンに対してシーン当たり平均 4 つのタグが関連付けることができる．本研究で対象とする映像コンテンツは，少なくとも数百人程度の閲覧者がいるコンテンツを対象としているため，全体の 1 割未満のユーザが提案手法を利用することができれば，網羅性の高いアノテーションの作成が期待できる．

これらの結果より，提案手法はコメントアノテーション手法よりも広く浅くタグが関連付けられており，要件 2 を満たす．しかしながら，コメントアノテーション手法にも利点があるため，両者を補完・併用する形で提案手法を用いることが望ましい．具体的な併用手法や，検索アルゴリズムに関する検討は今後の課題である．

5·2 協調的アノテーション手法に関する評価

要件 1 と 2 を効率よく満たすためには，より多くのタグを，より容易に関連付けることができるインタフェースである必要がある．提案手法は，タグの大きさを重要

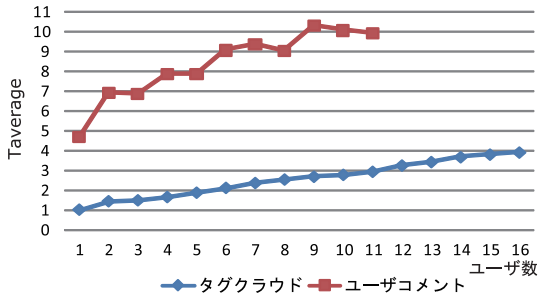


図 6 ユーザ数とシーンあたりの平均タグ異なり数の関係

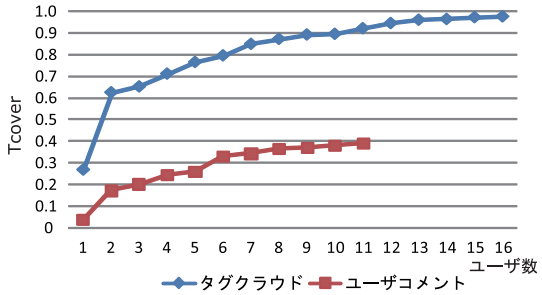


図 7 ユーザ数と網羅性の関係

度に応じて変化させることによって一覧性を向上させたタグクラウドを採用する．また，他のユーザがクリックしたタグの状況を視覚的に共有できるなどの工夫をしている．これらの工夫により，より視認性の高いインタフェースを実現するだけではなく，ユーザのアノテーションに対するモチベーションを向上させる効果を期待している．

そこで，これらのインタフェースの違いによる，タグの関連付け易さを評価するために，我々が提案した仕組みのサブセットを用いて比較実験を行い，これらの協調的アノテーション手法に関する工夫の有効性を検証する．そこで，表 3 で示す 3 つの評価実験を行った．

- **実験 1** タグの大きさを一定にしたタグクラウドを利用し，タグのクリック状況の共有を行わない．
- **実験 2** タグの大きさを重要度に応じて変更したタグクラウドを利用し，タグのクリック状況の共有を行わない．
- **実験 3** タグの大きさを重要度に応じて変更したタグクラウドを利用し，タグのクリック状況の共有を行う．(提案手法)

協調的アノテーション手法に関する評価に用いた映像コンテンツは映像 A と映像 B の二つである．それぞれの実験において被験者を 2 つのグループに分け，同一人物が同じ映像コンテンツを対象とした実験を 2 回以上行わないようにした．これにより，ユーザがはじめて閲覧する映像に対してアノテーションを付与する状況を想定する．それぞれの実験の後，アンケートによる評価を行った．アンケートの項目は，以下の 4 つである．

- 使いやすさ インタフェースの使いやすさ．
- 面白さ システムを利用した映像の閲覧が面白いか．

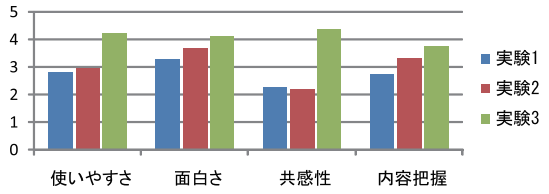


図 8 アンケートによる評価.

- 共有感 他ユーザとの連携感やコミュニティ感があるか.
- 内容把握 映像の内容を理解できたか.

ユーザを増やすためには,「使いやすさ」だけでなく提案手法を用いて映像閲覧したくなる「面白さ」も重要である．さらに，ニコニコ動画などのコメント共有型の映像共有サービスの面白さの一つに，映像を皆で見ているという「共有感」の高さが一因としてあり，それも重要である．さらに，提案方式が映像の閲覧を阻害するインタフェースになっていてはいけないため，映像の「内容把握」が適切になされているかという観点で調査した．

アンケートの結果（リッカードの 5 段階評価）を図 8 に示す．また，これらの結果に対して検定を行った．

まず，実験 1 と実験 2 を比較する．実験 2 では，タグの大きさを重要度に応じて大きさを変化させることによって，ユーザに対して何らかの良い結果がでることを期待していたが，5%水準での有意差はすべての項目において見受けられなかった．また，表 3 に示すように，実験 1 の平均タグクリック回数が 71.3 回であるのにたいして，実験 2 の平均タグクリック回数が 55.3 回と実験 1 と実験 2 においてアノテーション作成のモチベーションを向上させる効果は無かった．つまり，重要度に応じてタグの大きさを変化させることによる効果はあまり無かった．

次に，実験 1, 実験 2 と提案手法である実験 3 を比較する．実験 1 と実験 3 の間には全ての項目において，実験 2 と実験 3 の間では，**使いやすさ**, **共有感**の項目で 5%水準で有意差があった．**共有感**の項目において，実験 2 は 2.1 であるのに対して実験 3 は 4.4 である．実験 2 でも，ユーザのタグを収集している点や，映像は誰でも閲覧できるという点において実際には映像を Web 上で共有しているといえるが，アンケート結果は好ましくない．つまり，共有している状況をリアルタイムに提示することによって，より共有感を拡大させていることが分かる．さらに，他のユーザによってタグがハイライトされている状況を視覚的に共有できることによって，どのような意図をもってタグをクリックするのかが明確になり，**使いやすさ**も実験 2 の 2.9 から実験 3 の 4.2 と大きく向上している．また，実験 3 において，他人のボタンを押した状況が確認できるのがよいというコメントも多数寄せられた．さらに，付与されたタグの数をみると，表 3 に示すように，実験 3 は実験 1 と実験 2 に比べ 2.3 倍から 3.0 倍もの付与回数があった．これにより，タグを共有してハイライト表示するという点が，ユーザを刺激し，ユーザのモチベーションを向上させることが分かる．

表 3 実験条件とタグ押下数．タグ押下数とは，1 コンテンツあたりにタグがクリックされた回数の平均である．

実験	タグの大きさ	タグ共有	タグ数	タグ押下数
1	一定	なし	30	71.3
2	重要度依存	なし	30	55.3
3	重要度依存	あり	30	168.7

6. 関 連 研 究

閲覧者がアノテーションを付与する仕組みはいくつか存在する．ニコニコ動画や YouTube では，Synvie と同じように映像やその時間軸に対してコメントを付与することが可能であるが，これらは，コミュニケーションが目的であり，アノテーション利用を考慮していない．また，WebEVA[Volkmer 05] では，TRECVID[Smeaton 09] のための基礎的情報を付与するために，膨大な画像と映像に対して協調的にいくつかの概念を関連付けることができる．多数のユーザが同じ関係について同時に評価することによって安定した評価が可能になるが，関連付ける概念の数が限られる．一方で，ESP ゲーム [Ahn 04] は，ランダムに表示される画像に対して，互いに連想するであろう言葉（タグ）を提示し，一致するタグの数を競う対戦ゲームである．これにより，画像に関連したタグが多数収集可能になるが，動画像を対象とはしていない．

一方で，映像シーン検索に関する研究の例として，ニュース映像をタグ付け・検索する手法として，Informedia Project[Wactlar 96] が有名である．これらは，音声認識結果やクローズドキャプションから得られる音声書き下しテキストを利用しているため，品質の悪い Web 映像コンテンツへの適用は難しい．また，タグに基づく映像シーン検索として，映像に対する断片的な内容記述と記述間のグラフ的関連性に基づいて映像シーン検索を実現する手法 [是津 98] や，タグと映像の関連性やタグ間の関連度に基づいて映像シーン検索を実現する手法 [吹野 02] が提案されている．これらは，タグに基づいて効率よく検索に利用する手法を提案しており，我々の手法によって取得されたアノテーションにも応用可能であろう．しかしながら，閲覧者が効率よくタグを入力する手法については言及していない．

7. お わ り に

本論文において，タグクラウドを用いた映像シーンアノテーション手法の仕組みを提案した．提案手法は，映像に関連するユーザコメントからタグを生成し，そのタグを閲覧者が映像シーンと関連付けることによって，映像に関連する多種多様なタグを用いたアノテーションが可能になる．さらに，タグ集合をタグクラウドの形式で表示し，またタグのクリック状況を共有することによって，タグ押

下回数が単純な手法の 2.4 倍に増えるなどユーザのアノテーション意欲を向上させることを確認した．また，従来のコメントアノテーション手法のタグの網羅性が 40%前後であったのに対して，提案手法の網羅性が 100%近くになるなど，従来手法のタグの網羅性に関する問題を解決する．また，提案方式は，多くのシーンに満遍なくタグが付与されやすいという特徴だけではなく，そのタグがどれくらいのユーザに支持されているかという割合から，映像シーンとタグの関連度合いが取得できるであろう．これにより，タグと映像シーンの関連性を容易に付与可能になることによって，タグと映像シーン間の関連性を利用した既存の映像シーン検索手法を利用可能になる．

ただし，今回の研究では，従来手法よりも映像シーン検索に適したアノテーションを容易に取得できることを示したが，映像シーン検索を実現するための詳細な手順までは言及していない．また，従来手法によって取得されたアノテーションと提案手法によって獲得されたアノテーションにはそれぞれ利点や欠点があり，どちらかの手法のみを単独で利用するのではなく，これらを併用する手法も検討する必要がある．今後の課題としては，Synvie を含めた実際の映像共有サービスに提案手法を適用する．さらに，映像シーン検索を実現するための，前述した問題を解決した新しい検索手法について検討する．なお，本研究は文部科学省の科研費（20509003）の助成を得た．

◇ 参 考 文 献 ◇

[Ahn 04] Ahn, L. V. and Dabbish, L.: Labeling Images with a Computer Game, in *Proceedings of ACM CHI 2004*, pp. 24–29 (2004)

[Davis 93] Davis, M.: An Iconic Visual Language for Video Annotation., in *Proceedings of the IEEE Symposium on Visual Language*, pp. 196–202 (1993)

[Int 01] International Organization for Standardization (ISO): *Information Technology - Multimedia Content Description Interface (MPEG-7), ISO/IEC 15938:2001* (2001)

[Masuda 08] Masuda, T., Yamamoto, D., Ohira, S., and Nagao, K.: Video Scene Retrieval Using Online Video Annotation, in *Lecture Notes on Artificial Intelligence (LNAI 4914: JSAI 2007 (K. Satoh, et al. Ed.))*, pp. 255–268 (2008)

[Miyamori 05] Miyamori, H., Nakamura, S., and Tanaka, K.: Generation of Views of TV Content Using TV Viewers' Perspectives Expressed in Live Chats on the Web, in *Proceedings of ACM Multimedia 2005*, pp. 853–861 (2005)

[Nagao 01] 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)

[Nagao 02] Nagao, K., Ohira, S., and Yoneoka, M.: Annotation-Based Multimedia Summarization and Translation, in *Proceedings of COLING 2002*, pp. 702–708 (2002)

[Smeaton 09] Smeaton, A. and Kraaij, W.: TREC Video Retrieval Evaluation (2009)

[Smith 00] Smith, J. R. and Lugeon, B.: A Visual Annotation Tool for Multimedia Content Description, in *Proceedings of the SPIE Photonics East, Internet Multimedia Management Systems*, pp. 49–59 (2000)

[Uehara 06] Uehara, H. and Yoshida, K.: Automating Viewers' side Annotations on TV Drama from Internet Bulletin Boards, 情報処理学会論文誌, Vol. 47, No. 3, pp. 765–774 (2006)

[Volkmer 05] Volkmer, T., Smith, J. R., and Natsev, A. P.: A Web-based System for Collaborative Annotation of Large Image and Video Collections: An Evaluation and User Study, in *Proceedings of ACM Multimedia 2005*, pp. 892–901 (2005)

[Wactlar 96] 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)

[Yamamoto 08a] Yamamoto, D., Masuda, T., Ohira, S., and Nagao, K.: Collaborative Video Scene Annotation Based on Tag Cloud, in *Proceedings of PCM 2008*, pp. 397–406 (2008)

[Yamamoto 08b] Yamamoto, D., Masuda, T., Ohira, S., and Nagao, K.: Video Scene Annotation based on Web Social Activities, *IEEE MultiMedia*, Vol. 15, No. 3, pp. 22–32 (2008)

[井手 09] 井手 一郎, 柳井 啓司：セマンティックギャップを越えて〜画像・映像の内容理解に向けて〜, 人工知能学会誌, Vol. 24, No. 5, pp. 691–699 (2009)

[山本 05] 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)

[山本 07] 山本 大介, 増田 智樹, 大平 茂輝, 長尾 確：映像を話題としたコミュニティ活動支援に基づくアノテーションシステム, 情報処理学会論文誌, Vol. 48, No. 12, pp. 3624–3636 (2007)

[吹野 02] 吹野 直紀, 角谷 和俊, 田中 克己：キーワード毎のショット長分布を用いたビデオ映像シーン検索, 情報処理学会研究報告. データベース・システム研究会報告, 第 41 巻, pp. 49–56 (2002)

[是津 98] 是津 耕治, 上原 邦昭, 田中 克己：時刻印付オーサリンググラフによるビデオ映像のシーン検索, 情報処理学会論文誌, Vol. 39, No. 4, pp. 923–932 (1998)

〔担当委員：高間 康史〕

2009 年 10 月 1 日　受理

著 者 紹 介

山本　大介（正会員）

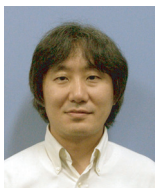
2003 年名古屋大学工学部電気電子・情報工学科卒業.2008 年同大学大学院情報科学研究科博士課程修了.2006 年-2008 年日本学術振興会特別研究員.2008 年-現在名古屋工業大学大学院工学研究科助教. 博士 (情報科学).

増田　智樹

2007 年名古屋大学工学部電気電子・情報工学科卒業. 2009 年名古屋大学大学院情報科学研究科修士課程修了.2009 年-現在株式会社 NTT データ.

大平　茂輝

2000 年早稲田大学理工学研究科情報科学専攻修士課程修了. 2003 年早稲田大学理工学研究科情報科学専攻博士課程単位取得退学. 2001 年-2003 年早稲田大学理工学部情報学科助手. 2003 年-2006 年名古屋大学情報メディア教育センター助手. 2006 年-2009 年名古屋大学エコトピア科学研究所助手 (2007 年-助教). 2009 年-現在名古屋大学情報基盤センター助教.



長尾　　確（正会員）

1987 年東京工業大学総合理工学研究科システム科学専攻修士課程修了. 1987 年-1991 年日本アイ・ビー・エム株式会社東京基礎研究所. 1991 年-1999 年株式会社ソニーコンピュータサイエンス研究所. 1996 年-1997 年米国イリノイ大学アーバナ・シャンペーン校客員研究員. 1999 年-2001 年日本アイ・ビー・エム株式会社東京基礎研究所. 2001 年-2002 年名古屋大学工学研究科助教授. 2002 年-2009 年名古屋大学情報メディア教育センター教授. 2009 年-現在名古屋大学大学院情報科学研究科メディア科学専攻教授.