情報処理学会論文誌　Vol. 52　No. 12　3471–3482 (Dec. 2011)

映像に付与されたコメントを用いた登場人物が注目されるシーンの推定

佃　　洸　摂[†1]　中　村　聡　史[†1]
山　本　岳　洋[†1,†2]　田　中　克　己[†1]

映像の編集や検索を行う際に，ある人物が視聴者の注目を集めているシーンだけを切りだしたり，ある人物がより視聴者の注目を集めている映像を検索したいということはよくある．また，現在視聴している映像の関連映像として，登場人物の活躍パターンが似ている映像を推薦してほしいということもよくある．しかし，映像に登場する人物がどの程度視聴者の注目を集めているかは，映像のテキストデータや画像解析だけでは判断できないことが多く，視聴者の反応に基づいて注目されている人物を推定する必要がある．そこで本稿では視聴者の反応として映像の再生時刻に沿って付与されたコメントを用いて，映像に登場する人物が視聴者の注目を集めているシーンの推定と各シーンにおける各登場人物の活躍の度合いの推定を行う手法の提案をする．ただし，本稿ではニコニコ動画に投稿された映像の中で，コメントが一定数以上付与された映像に登場する，視聴者に名前が広く知れわたっている人物を対象とする．

Estimating the Spotlighted Scene of Characters Based on the Comments Posted to a Video

Kosetsu Tsukuda,[†1] Satoshi Nakamura,[†1]
Takehiro Yamamoto[†1,†2] and Katsumi Tanaka[†1]

When a user edits or searches a video clip, he/she often hopes to extract a scene or search video clips in which a character is active. Moreover, he/she also often hopes to be recommended a video clip that has a similar activity pattern of characters as a relevant video clip that he/she is watching. However, it is difficult in many cases to judge the active characters from only text data or image analysis. We estimate the degree of attention based on viewers' reaction. In this paper we use comments posted to a video clip as the viewers' reaction. We propose a method to estimate the spotlighted scenes of each character in a video clip and the degree of it. We especially target video clips that is uploaded to Nico Nico Douga and a character whose name is known widely.

1. はじめに

近年，YouTube[⋆1]やニコニコ動画[⋆2] などの映像共有サービスが広く利用されている．YouTube には 2010 年 1 月の時点で 10 億本以上，ニコニコ動画には 2011 年 4 月の時点で 570 万本以上の映像が投稿されており，非常に多様な映像の視聴が可能になっている．その一方で，映像の数が多くなりすぎているため，視聴したいと考える映像を検索するのは困難である．また，ユーザには限られた時間しかないため，発見したすべての映像を視聴することは難しい．そのため，ユーザは限られた時間の中で自分の視聴したい映像を効率良く検索したり視聴したりすることを望むことが多い．具体的には以下のような例があげられる．

- 映像を最初から最後まですべて視聴する時間はないので，ある登場人物が注目を集めているシーンを集めたダイジェスト映像を視聴したい．
- キーワードで映像を検索したときに，検索結果に含まれる映像をすべて視聴する時間はないので，ある登場人物がより映像全体にわたって視聴者の注目を集めている映像や，映像の中で雰囲気が大きく変わるきっかけとなる活躍をしている映像を優先的に視聴したい．
- ユーザが気に入ったある映像の関連映像として，タイトルや説明文などのテキストが類似している映像ではなく登場人物の活躍パターンが類似している映像を推薦してほしい．

映像のダイジェスト生成や映像の盛り上がり度を計算するため，映像を分析してシーンの切り換わりを検出したり[1)]，映像中の音声を分析することでシーンの盛り上がりを検出する[2)] 研究は行われているが，映像に対する視聴者の反応を利用しておらず，コンテンツ作成者の意図を反映しているにすぎず，映像の作成者が注目させようと意図している人物と，視聴者が注目している人物は必ずしも一致するとは限らない．一方，映像に対して付与されたコメントを視聴者反応と見なし，盛り上がりなどを検出することで映像のダイジェストやランキングを行う研究はなされている．しかし既存の研究では，映像の各シーンに対する，

†1 京都大学
Kyoto University
†2 日本学術振興会特別研究員
JSPS Research Fellow
⋆1 http://www.youtube.com/
⋆2 http://www.nicovideo.jp/

視聴者の感想を集約すること[3]や注目の大きさを求めること[4]を目的としている．そのため，視聴者の注目が集まっている登場人物までは求めることができないという問題がある．また，膨大な映像に対する適用事例がない．

ある人物がある映像内で注目されているシーンは，映像のタイトルやタグ，説明文といったテキスト情報からは判定不可能である．また，画像解析によってその人物の画面への登場の有無を判定することはある程度可能であるが，注目の度合いは判定できない．そこで本研究では，映像の中で各登場人物が視聴者の注目を集める言動をとっているシーンとそのシーンにおける注目の大きさを視聴者が映像に対して付与したコメントを分析することで求めることを目的とする．ここで，視聴者が注目を集める登場人物の言動には，その人物が活躍している場合や批判されている場合など様々な状況が考えられるが，本稿では簡単のためこれらの状況の違いは考えないものとする．

実際にユーザが，ある人物が注目されているシーンの抽出や映像の検索を行う際は，視聴者から見た登場人物の注目の大きさに基づいたシーンや映像を返す方がユーザには望ましい．そこで本稿では視聴者にとって登場人物が注目を集める言動をとっているシーンを「注目シーン」と定義し，視聴者が感じる登場人物の注目シーンとその大きさに着目する．

ある人物が注目されているシーンでは，視聴者はその人物に対して何らかの反応をすると考えられる．たとえば，その人物に視聴者の視線が集まったり，その人物の言動に対して笑い声を出したりする反応があげられる．これらに加え，ニコニコ動画のように映像の再生時刻に沿ったコメントの付与が可能なサービスの登場により，視聴者によって映像に付与されたコメントも映像に対する反応の 1 つと考えられるようになった．そこで本稿ではニコニコ動画に付与されたコメントの観察に基づき，ある人物が注目されているシーンでは，その人物に対して映像の視聴者がコメントを付与するという仮説を立て，映像に付与されたコメントの内容を基に，登場人物が注目されているシーンおよび注目の大きさを推定し，仮説の検証を行う．ただし本研究では，映像に付与されたコメントを用いるため，コメントが一定数以上付与されている映像のみを対象とする．また，コメント中に含まれる人物名を用いて注目シーンの推定を行うため，視聴者に名前が広く知れわたっている人物を対象とする．

ニコニコ動画は，日本で最も有名な映像共有サービスの 1 つである．2010 年 11 月の時点で，1,600 万人以上の利用者がおり，500 万本以上の映像に対して 28 億件以上のコメントが付与されている．ニコニコ動画のインタフェースを**図 1** に示す．ユーザは映像を視聴しながら，任意の場面に対してコメントの付与が行える．また，他のユーザは，再生時刻に沿って，画面の右から左へと流れて表示されるコメントを見ることができる．

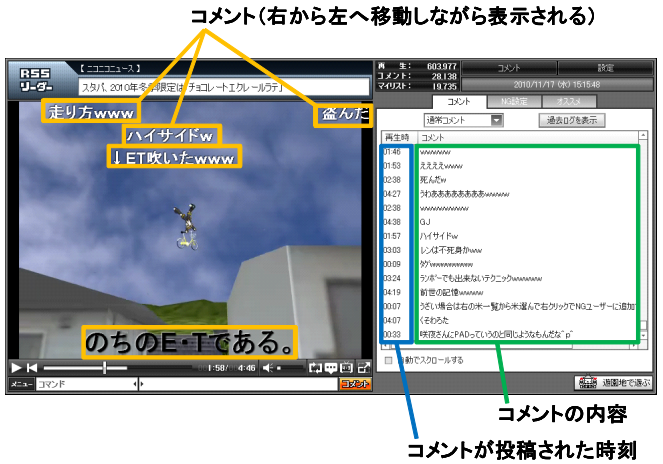

図 1　ニコニコ動画のインタフェース
Fig. 1　Interface of Nico Nico Douga.

2. 関 連 研 究

映像の再生時刻に応じてユーザがコメントを付与できるサービスを対象とした研究は近年さかんに行われている．

青木ら[5]はニコニコ動画の映像に対して付与されるコメントの出現頻度を用いて，映像内で最も重要な箇所の判別や映像の要約を試みている．また，Nakamura ら[6]はニコニコ動画の映像に対して付与されたコメントから喜び，悲しみなどの印象情報を抽出し，インデックスを作成することで，印象に基づく映像検索およびランキングを可能としている．いずれの研究においても，映像に登場する人物には着目していない点で本研究とは異なる．

Diakopoulos ら[7]は Shamma ら[8]の行った研究を基に，テレビで放送された大統領討論会に対する Twitter[⋆1]上のコメントを用いて，映像に登場する人物に対してポジティブな感情を持ったユーザ数とネガティブな感情を持ったユーザ数の比較を行っている．しかし，この研究では映像内の各時刻における話者や話される内容などの情報をあらかじめ用意したうえでコメントを分析するため，こうした発信情報がない映像に対しては手法を適用できな

⋆1 http://twitter.com/

い．我々が対象とする映像共有サイトに日々アップロードされる膨大な映像はそうした情報を持たないため，Diakopoulos らの手法は利用できないが，我々の手法は視聴者の反応を利用するものであるため，一般のユーザが作成した映像に対してもその映像に登場する各人物に対する視聴者の反応の大きさを求めることが可能である．

テレビ番組と同期してコメントの書き込みが行われる Web 上の実況チャットを利用した研究では，Uehara ら[9] が実況チャットのコメントを分析し，番組内である人物やテーマが話題となるシーンを求めている．また，宮森らは実況チャット上に現れる特徴的な表現を処理することで，番組の盛り上がり場面や視聴者の嗜好・趣味に沿ったリアクションなど，視聴者視点に関連するメタデータを抽出したり[10]，それに基づいて視聴者の盛り上がりや自分と類似した嗜好を持つ他人が興味を示す部分など，様々なビューを作成したり[11] している．これらの研究では，ある登場人物がより注目されている映像を視聴したいというユーザの欲求を満たすことはできないが，本研究では映像間での人物の注目の大きさを考慮し比較することで，そのような意図を持ったユーザに映像をランキングして提示することが可能である．

また，独自のアノテーションシステムを開発し，それを利用した研究も行われている．山本ら[12] は任意のビデオシーンに対するコメントの付与や，ブログへの引用ができるオンラインビデオアノテーションシステム Synvie を開発し，Masuda ら[13] がそのシステムを用いて，映像のシーンに対して付与されたコメントや引用元のブログ記事の文章からタグを自動生成し，映像のシーン検索を可能としている．しかし，彼らのシステムを利用することで，ある人物が登場するシーンを検索することは可能になるが，登場人物に対する注目の大きさに基づいて検索することはできない．

3. 注目シーンおよび注目の大きさの推定

本章では，映像に対して付与されたコメントを用いて，映像に登場する人物が視聴者からの注目を集める言動をとっているシーンを抽出する方法と，注目の大きさを数値化する方法について述べる．ここで，コメントの情報にはコメントの内容と，映像中で各コメントが付与された時刻を示す時刻情報が含まれる．

映像に付与されるコメントの中には，映像の登場人物に対して付与されるコメントが存在する．ニコニコ動画では，映像の再生時刻に沿ってコメントが映像上に表示されるため，「すごい！」や「この動き面白い！」のようなコメントであっても，そのコメントが付与された対象となる登場人物が他の視聴者にも明確である．一方で，コメントで言及している対象となる登場人物が明らかな場合であっても，「レンすごい！」や「ミクかっこいい！」のように，「レン」や「ミク」といった，コメントの対象となる人物の名前を含むコメントも存在する．そこで本研究では，名前を含むコメントを用いて，映像の各登場人物の注目シーンを推定し，名前を含むコメントおよびその周辺コメントを用いて，各注目シーンの注目の大きさを推定する手法を提案する．

表 1　人物名辞書に登録した人物の例
Table 1　Examples of characters who are registered to the dictionary of names of characters.

正式な表記	別の表記（ニックネームなど）
初音ミク	初音，ミク
KAITO	カイト，兄さん
天海春香	天海，春香，はるるん，はるかっか，閣下
秋月律子	秋月，律子，律，りっちゃん

3.1 コメントに含まれる人物名の抽出

ニコニコ動画の映像に付与されるコメントは口語的な記述が多いため，自然言語処理によりコメントから人物名を抽出することは難しい．そこで，本研究では簡単のためあらかじめ用意した人物名辞書とのパターンマッチを行うことによりコメントからの人物名抽出を行う．ここでは，ニコニコ動画上の各映像でよく登場する各人物に対して，その人物を表す別の表記（ニックネームなど）を考えられるだけ手作業で辞書に登録した．辞書に登録された人物とその人物を表す別の表記の例を**表 1** に示す．映像に付与されたすべてのコメントに対して辞書を用いたパターンマッチを行うことで，その映像の中で言及された人物と各人物への言及回数，および映像中で各人物の名前を含むコメントが付与された時刻の情報を得ることができる．

3.2 注目シーンの抽出

コメントの中にある人物名が含まれていれば，そのコメントが付与された時刻には，その人物が視聴者の注目を集める言動をとっている可能性が高いと考えられる．ある映像 v における，すべてのコメントの付与時刻の集合を $T = \{t_1, t_2, \cdots t_n\}$ とし，T のうち，人物 x の名前を含む全コメントの付与時刻の集合を T_x とする．このとき，各 $t \in T_x$ の前後 k 秒間は人物 x が注目を集める言動をとっていると仮定する．この仮定のもとで，時刻 $t_i \in T_x$ 以降に付与されたコメントで x の名前を含むもののうち，最も近い付与時刻を t_{i+1} とすると，$t_i + k \geq t_{i+1} - k$ を満たしていれば，人物 x は $t_i - k$ 秒から $t_{i+1} + k$ 秒にわたって視聴者の注目を集めていると考える．このように，$2k$ 秒以内に連続して同一人物の名前を含

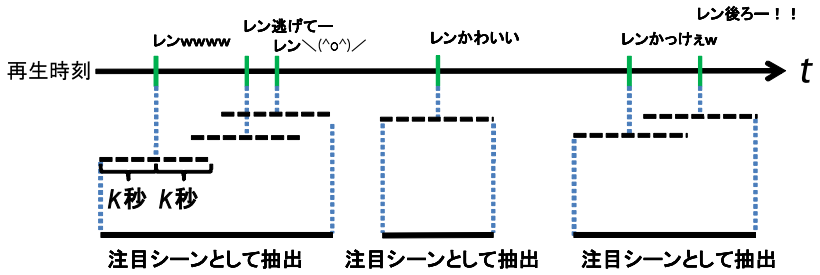

図 2　名前を含むコメントを用いた注目シーンの抽出
Fig. 2　Extracting attention scenes based on comments including a character's name.

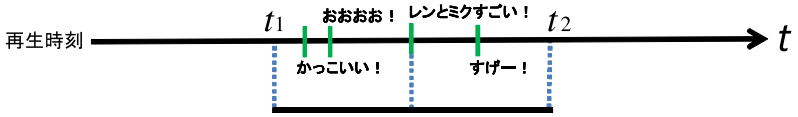

図 3　1 つのコメントに複数の登場人物の名前が含まれる例
Fig. 3　An example of a comment that includes names of some characters.

むコメントが付与されている期間をその人物の注目シーンと定義する．たとえばある映像において，「鏡音レン」という人物の名前を含むコメントが映像の再生時刻軸上に**図 2** のようにマッピングされたとすると，図中の実線で示された時刻帯が「鏡音レン」の注目シーンとして抽出される．ただし，映像のある場面において，視聴者の注目を集めるような活躍をしている人物は 1 人とは限らず，複数人の場合もある．そのため本稿では，1 つのコメント内に複数の人物の名前が含まれる場合，そのコメントは複数人物の言動に対して視聴者が反応したものであると考える．そして，そのコメントが付与された前後 k 秒間は，名前が記述されている人物全員の注目シーンであるとする．たとえば，**図 3** において，「初音ミク」と「鏡音レン」の注目シーンはいずれも時刻 t_1 から t_2 の間で，注目の大きさはいずれも 4 となる．つまり，ある登場人物 x と y 両方の名前を含むコメントが時刻 t_i に付与された場合，$t_i \in T_x$ かつ $t_i \in T_y$ となる．

3.3　注目の大きさの数値化

ある人物が視聴者の注目を集める言動をとっているときに，視聴者は必ずしもその人物の名前を含むコメントを付与するとは限らない．本研究では，名前を含むコメントが付与された時刻と近い時刻に付与されたコメントは，その人物の言動に対して付与されたコメントであると仮定し，その数が多いほど，視聴者のその人物に対する注目の大きさは大きいと考える．ただし，映像に付与されるコメントは，すべてが映像に登場する人物に対するものであるとは限らない．コメントの中には，映像の内容を説明するコメントや視聴者同士で会話のやりとりをするように付与されるコメント，映像の内容とはまったく関係のないコメントなどがある．ここで，映像の登場人物の言動に対して付与されたコメントには，何らかの感情を表す語が含まれていると仮定する．たとえば，ある人物が視聴者を笑わせる言動をとるシーンでは，その人物の言動に対して喜びの感情を含むコメントが付与され，ある人物が視聴者の批判を買う言動をとるシーンでは，その人物の言動に対して否定的な感情を含むコメントが付与されると考えられる．そこで本研究では Nakamura ら[6] が作成した喜び，悲しみ，驚き，肯定，否定のそれぞれに対応する語の集合からなる辞書を用いて，辞書に登録されている語を含むコメントのみを解析の対象とする．

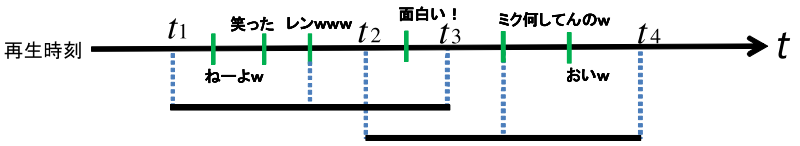

図 4　ある人物の名前を含むコメントと隣接したコメントに他の登場人物の名前が含まれる例
Fig. 4　An example of a comment that is adjacent to a comment including a name of a character and includes a name of another character.

映像に付与されたコメントのうち，人物名辞書に登録されているいずれの人物の名前も含まず，感情を表す語を 1 つ以上含むコメントの付与時刻の集合を T_{imp}，登場人物 x の注目シーンの集合を $R_x = \{r_1, r_2, \cdots, r_n\}$ とする．また，注目シーン r_i の開始時刻を s_i，終了時刻を e_i と表すと，登場人物 x のある活躍シーン r_i における注目の大きさ $f_{\mathrm{attention}}(s_i, e_i)$ は次式により表される．

$$f_{\mathrm{attention}}(s_i, e_i) = |\{t | s_i \leq t \leq e_i, t \in T_x \cup T_{\mathrm{imp}}\}|$$

つまり，注目シーン r_i の間に付与されたコメントのうち，人物 x の名前を含むコメントと感情を表す語を含むコメントの数の和により，人物 x のある活躍シーン r_i における注目の大きさ $f_{\mathrm{attention}}(s_i, e_i)$ を表す．ただし，名前を含むコメントの中にはその人物が注目を集める言動をとっていないときに付与されるものもあるため，$f_{\mathrm{attention}}(s_i, e_i)$ の値が閾値 α 以下のシーンはノイズとして除去する．また，T_{imp} に含まれるコメントが複数の登場人物の注目シーンが重複した時刻に付与されたコメントである場合，その場面では複数の登場人物が活躍していると考え，注目シーンが重複しているすべての登場人物の言動に対して

付与されたものであるとする．つまり，**図 4** において，時刻 t_2 から t_3 の間は「初音ミク」と「鏡音レン」両者の注目シーンであり，この時刻帯に付与されたコメントは両者の注目の大きさを計算する際に同じように 1 ずつ加算される．したがって，この例において「初音ミク」の注目の大きさは 3,「鏡音レン」の注目の大きさは 4 となる．

4. 実　　験

本章では，登場人物が視聴者の注目を集める言動をとるシーンにおけるその人物に対するコメントの付与の有無と，注目シーンを推定するために感情を含むコメントのみを対象とすることの有用性を調べるために，被験者を用いた評価実験を行った．今回評価の対象とした映像は，「アイドルマスター」，「MMD（MikuMikuDance）」，「政治」のカテゴリに属する映像である．アイドルマスターは，アイドルを育成するゲームシリーズの名称であり，ニコニコ動画ではゲームの登場人物を用いて作成したアニメや，ゲーム中の映像に別の音楽を重ねた映像などが投稿されている．MMD は無償で利用できる，映像作品を作成するためのソフトウェアであり，キャラクタのモデルを読み込むことで様々なキャラクタを用いて映像を容易に作成することを可能としている．MMD のカテゴリにはこのツールを用いて作成された映像が含まれる．政治のカテゴリには，政治家を登場人物として作成したアニメや，国会での討論などの映像が含まれる．各カテゴリに属する映像集合は，タグに「アイドルマスター」，「MMD」，「政治」のそれぞれの単語を含む映像とした．これらのカテゴリを選択した理由は，以下の 3 つである．

- 1 つの映像に多くの人物が登場する．
- 様々な人物の組合せからなる映像が存在する．
- 特定の人物が注目される映像が極端に多いということはなく，映像によって登場人物の役割が様々である．

4.1 データ収集

実験を行うにあたり，我々はニコニコ動画に投稿されている映像の中でコメントが 1,000 件以上付与されている約 17 万件の映像について，それぞれのタイトル，説明文，タグ，コメント，再生数，コメント数，お気に入り数を収集し，データベースに格納した．なお，コメントはクロール時の最新の 1,000 件を収集した．収集したコメントのうち，人物の名前を含むコメントの例を**表 2** に，感情を表す語を含むコメントの例を**表 3** に示す．表 2 では人物の名前を表す語を，また表 3 では感情を表す語をそれぞれ太字で表している．

表 2　登場人物の名前を含むコメントの例
Table 2　Examples of comments that include a character name.

ネル逃げてー
毎回**レン**逃げ足速え ww
春香さん怖すぎるだろ ww
ここの**やよい**すっごくかわいい
麻生がんばれ
亀井かっけぇw

表 3　感情を表す語を含むコメントの例
Table 3　Examples of impression comments.

喜び
ちょ**wwwwwwwwwww**
これは**笑える**
悲しみ
ここが一番**泣ける**…
切ないよ〜
肯定
やっぱ**かっこいい**！
かわいいよー
否定
うげ，**きもい**
この態度**むかつく**なー

表 4　4.2 節の実験に用いた映像の情報
Table 4　Information of videos used in the experiment in section 4.2.

カテゴリ	本数	再生時間		
		最短	最長	平均
アイドルマスター	6,483	2 秒	271 分 36 秒	12 分 17 秒
MMD	534	30 秒	46 分 17 秒	5 分 40 秒
政治	2,831	19 秒	181 分 43 秒	15 分 36 秒

4.2 名前を含むコメントの割合

まず，提案手法の適用可能性を示すために，すべてのコメントに対する名前を含むコメントの割合を調べた．そのために，「アイドルマスター」，「MMD」，「政治」の 3 つのカテゴリについて，各カテゴリに含まれるすべての映像に付与されたすべてのコメントからランダムに 3,000 件ずつ抽出した．コメントを抽出する際に対象とした映像の情報を**表 4** に示す．その後，抽出されたすべてのコメントを著者が目視で確認し，名前を含むコメントの割合を求めた．その結果，名前を含むコメントの割合は「アイドルマスター」カテゴリで 9.6%，「MMD」カテゴリで 10.1%，「政治」カテゴリで 7.9%であった．

4.3 正解セットの作成

次に，被験者実験により，映像の各登場人物の注目シーンの正解セットを作成した．そのために，「アイドルマスター」，「MMD」，「政治」の各カテゴリから，評価の対象とする映像として，再生時間が 1 分 15 秒以上 6 分未満の映像をランダムに 12 本ずつ選択した．本

表 5　4.3 節の実験に用いた映像の情報
Table 5　Information of videos used in the experiment in section 4.3.

カテゴリ	本数	再生時間		
		最短	最長	平均
アイドルマスター	12	2 分 32 秒	4 分 49 秒	3 分 45 秒
MMD	12	3 分	5 分	3 分 59 秒
政治	12	1 分 26 秒	5 分 47 秒	3 分 54 秒

表 6　1 人の被験者の 1 つの映像に対する評価結果の例
Table 6　Example of evaluation result of a subject for a video clip.

	0:00–0:15	0:15–0:30	0:30–0:45	···	4:45–5:00	5:00–5:09
初音ミク	0	1	0	···	0	0
弱音ハク	1	2	2	···	1	0
巡音ルカ	0	0	0	···	2	1
KAITO	0	0	2	···	0	2

実験に用いた映像の情報を**表 5** に示す．被験者数は 5 名で，内訳は 20 代の男性 4 名と女性 1 名である．そして，1 つの映像につき 3 名の被験者が評価を行った．評価方法は，映像ごとに，映像に登場する人物の一覧と映像を 15 秒ごとに区切った表を用意し，各登場人物が各区間において注目を集める言動をとっている度合いを 3 段階で評価してもらった．3 段階の内訳は，0 が注目を集める言動をまったくとっていない，1 がやや注目を集める言動をとっている，2 がかなり注目を集める言動をとっている，とした．なお，本評価を行う際に用いた人物名辞書は 44 件からなる．被験者は映像を視聴する際，コメントを表示しないようにし，映像のみを閲覧して評価を行った．評価の結果の例を**表 6** に示す．

その後，映像ごとに 3 名の被験者の評価結果を集約し，正解セットを作成した．正解セットの作成方法は次のとおりである．

ある映像 v で人物 x の各時刻区間において 3 名の被験者のうち 2 名以上が 1 または 2 をつけている区間をすべて求め，その区間を映像 v において人物 x がやや注目を集める言動をとっている，またはかなり注目を集める言動をとっているシーンとする．これをすべての映像のすべての登場人物について求めたものを正解セット 1 とする．

さらに，ある映像 v で人物 x の各時刻区分において 3 名の被験者のうち 2 名以上が 2 をつけている区間をすべて求め，その区間を映像 v において人物 x がかなり注目を集める言動をとっているシーンとする．これをすべての映像のすべての登場人物について求めたものを正解セット 2 とする．

Algorithm 1 $15i$ 秒以上 $15(i+1)$ 秒未満の区間における人物 x の注目の大きさの計算

Require: $i(i \in \mathbb{Z} \wedge i \geq 0), R_x$
  $S_i = 0$
  for r_j in R_x **do**
    if $s_j < 15i$ **then**
      if $e_j > 15i$ and $e_j \leq 15(i+1)$ **then**
        $S_i += f_{\text{attention}}(15i, e_j)$
      else if $e_j > 15(i+1)$ **then**
        $S_i += f_{\text{attention}}(15i, 15(i+1))$
      end if
    else if $s_j \geq 15i$ **then**
      if $e_j \leq 15(i+1)$ **then**
        $S_i += f_{\text{attention}}(s_j, e_j)$
      else
        $S_i += f_{\text{attention}}(s_j, 15(i+1))$
      end if
    end if
  end for
  return S_i

4.4　注目シーンの抽出精度の検証

映像の登場人物が視聴者の注目を集める言動をとっているときはその人物に対するコメントが付与されるという仮説を検証するために，3.2 節のパラメータ k および 3.3 節の閾値 α の値を変化させたときの適合率，再現率，F 値を求めた．適合率，再現率，F 値は各映像の各登場人物ごとに次のようにして求められる．

まず，正解セット 1 において 2 名以上が 1 または 2 をつけた区間を 1 とし，それ以外の区間を 0 とすると，映像 v の登場人物 x の注目シーンは 0 と 1 の配列として表される．これを l_1 とする．l_1 と，提案手法により注目シーンとして抽出された区間を比較して適合率および再現率を求めるために，提案手法についても映像を 15 秒ごとに区切った各区間の各登場人物の注目の大きさを求める．$15i$ 秒以上，$15(i+1)$ 秒未満の区間における登場人物 x の注目の大きさの求め方を Algorithm 1 に示す．ただし，i は 0 以上の整数であり，ある映像 v における登場人物 x の注目シーンの集合を $R_x = \{r_1, r_2, \cdots, r_n\}$ とする．たとえば，人物 x の注目シーンが 20 秒から 43 秒までの 1 カ所のみであった場合，20 秒から 30 秒ま

表 7　正解セット 1 における各カテゴリの F 値の最大値とそのときの k，α，再現率，適合率の値
Table 7　Maximal value of F-measure and the value of k, α, recall, and precision in answer set 1.

	MMD	アイドルマスター	政治	全カテゴリ
F 値	0.626	0.693	0.747	0.629
再現率	0.728	0.756	0.722	0.785
適合率	0.605	0.723	0.834	0.605
k（秒）	4.0	2.5	0.50	5.0
α（個）	3	5	8	7

でと 30 秒から 43 秒までに分割し，20 秒から 30 秒までの間の注目の大きさが α 以上であれば 15 秒以上，30 秒未満の区間は x の注目シーンとなり，30 秒から 43 秒までの間の注目の大きさが α 未満であれば 30 秒以上，45 秒未満の区間は x の注目シーンとはならない．つまり，S_i の値が α 以上であればその区間の値を 1 に，α 未満であればその区間の値を 0 とした．これにより提案手法により求められる注目シーンも 0 と 1 の配列として表される．これを l_2 とする．

このとき，映像 v の登場人物 x の適合率 $P(v,x)$，再現率 $R(v,x)$，F 値 $F(v,x)$ はそれぞれ次の式により求められる．

$$P(v,x) = \frac{f_{\mathrm{common}}(l_1,l_2)}{f_{\mathrm{notzero}}(l_2)}$$

$$R(v,x) = \frac{f_{\mathrm{common}}(l_1,l_2)}{f_{\mathrm{notzero}}(l_1)}$$

$$F(v,x) = \frac{2\cdot P(v,x)\cdot R(v,x)}{P(v,x)+R(v,x)}$$

ここで，$f_{\mathrm{notzero}}(l_i)$ は l_i の要素の 1 の個数を表し，$f_{\mathrm{common}}(l_1,l_2)$ は l_1 と l_2 の同じ時刻区間の値がともに 1 となっている区間の数を表す．これをすべての映像のすべての登場人物について求め，その平均値を k，α におけるスコアとする．本実験では k の値を 0.5 から 10 まで 0.5 ずつ変化させ，α の値を 1 から 100 まで 1 ずつ変化させた．α の値が大きくなると，注目シーンとして抽出されることが少なくなるため，適合率は上昇したが再現率が低下した．一方で k の値が大きくなると，様々な区間が注目シーンとして抽出されるため，再現率は上昇したが適合率は低下した．正解セット 1 において全カテゴリの F 値の最大値は $k=5.0$，$\alpha=7$ のときの 0.629 であった．また，このときの適合率は 0.605，再現率は 0.785 であった．さらに，各カテゴリの F 値の最大値とそのときの k，α，再現率および適合率の値を**表 7** に示す．

表 8　正解セット 2 における各カテゴリの F 値の最大値とそのときの k，α，再現率，適合率の値
Table 8　Maximal value of F-measure and the value of k, α, recall, and precision in answer set 2.

	MMD	アイドルマスター	政治	全カテゴリ
F 値	0.513	0.641	0.762	0.560
再現率	0.375	0.869	0.625	0.741
適合率	0.875	0.562	1.0	0.523
k（秒）	3.5	1.0	5.5	5.0
α（個）	72	4	83	23

表 9　テストセット 1 における各手法の F 値の最大値の比較
Table 9　Comparison of maximal value of F-measure for each method in test set 1.

	MMD			アイドルマスター			政治			全カテゴリ		
	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率
提案手法	0.626	0.728	0.605	0.693	0.756	0.723	0.747	0.722	0.834	0.629	0.786	0.605
比較手法 1	0.621	0.731	0.596	0.687	0.756	0.702	0.733	0.679	0.857	0.628	0.743	0.636
比較手法 2	0.588	0.645	0.606	0.678	0.731	0.710	0.719	0.823	0.685	0.621	0.614	0.760

次に，正解セット 2 において 2 名以上が 2 をつけた区間を 1，それ以外の区間を 0 として，先ほどと同じ範囲で k と α を変化させ，全カテゴリの F 値の最大値を求めたところ，$k=5.0$，$\alpha=23$ のときに最大値 0.560 となった．このときの適合率は 0.523，再現率は 0.741 であった．各カテゴリの F 値の最大値とそのときの k，α，再現率および適合率の値を**表 8** に示す．

4.5　感情を含む周辺コメントを使用する有用性の検証

3 章では，名前を含むコメントの周辺コメントのうち，感情を含むコメントのみを使用する手法の提案を行った．そこで本節では，次にあげる 2 つの手法との精度の比較を行い，提案手法の有用性を検証する．

- 比較手法 1：3.3 節において，注目の大きさを求める際に映像に付与されたすべてのコメントを対象とする．そして，各区間の注目の大きさが閾値 β 以下であればノイズとして注目シーンから除く．
- 比較手法 2：映像を 15 秒ごとに区切り，人物 x の注目シーンを，各区間に含まれる x の名前を含むコメントの数のみから求める．この場合も，ある区間の名前を含むコメントの数が閾値 γ 以下であればその区間をノイズとして注目シーンから除く．

正解セット 1 および正解セット 2 に対して提案手法と 2 つの比較手法を適用したときの各カテゴリの F 値の最大値およびそのときの再現率と適合率の値を**表 9** および**表 10** に示

表 10　テストセット 2 における各手法の F 値の最大値の比較
Table 10　Comparison of maximal value of F-measure for each method in test set 2.

	MMD			アイドルマスター			政治			全カテゴリ		
	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率
提案手法	0.513	0.375	0.875	0.641	0.869	0.562	0.762	0.625	1.00	0.560	0.741	0.523
比較手法 1	0.567	0.458	0.854	0.631	0.830	0.577	0.667	0.500	1.00	0.539	0.696	0.507
比較手法 2	0.500	0.389	0.778	0.633	0.583	1.00	0.667	0.500	1.00	0.547	0.513	0.743

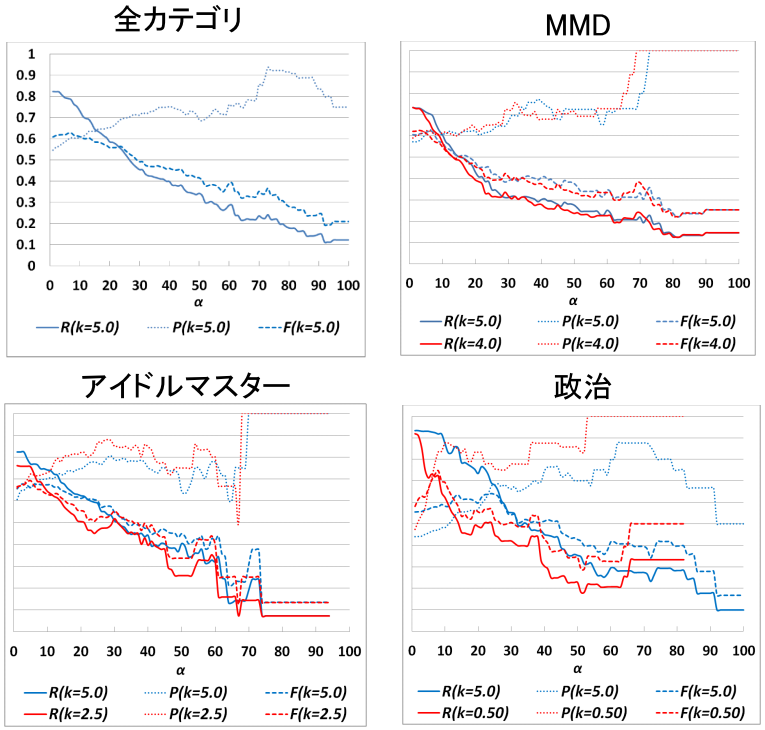

図 5　正解セット 1 において k を固定し，α を変化させたときの適合率，再現率，F 値の遷移
Fig. 5　Transition of precision, recall, and F-measure by fixing k and changing α for test set 1.

す．なお，β と γ の値はいずれも 1 から 100 まで 1 ずつ変化させ，最適な値を求めた．この結果から，いずれのテストセットにおいても全カテゴリの F 値の最大値は提案手法が両

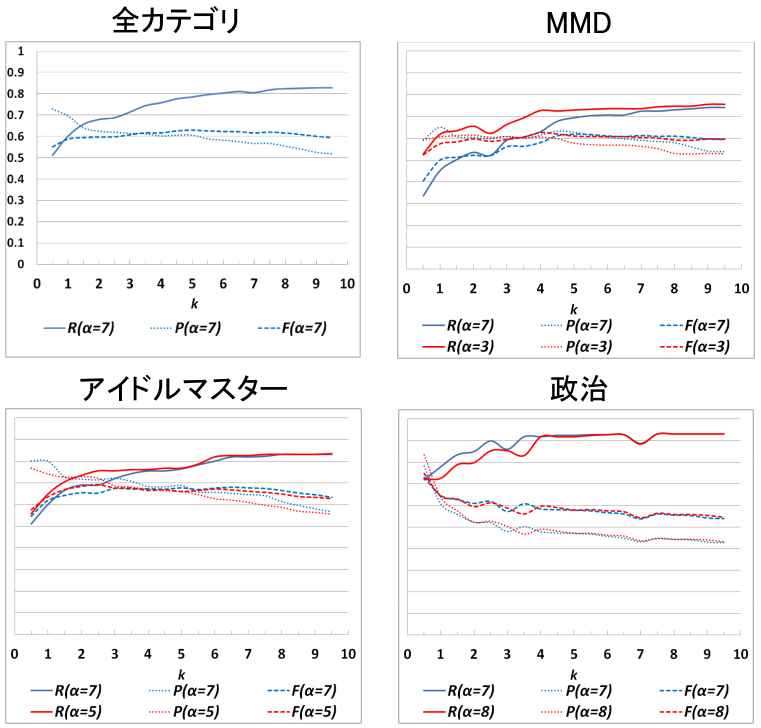

図 6　正解セット 1 において α を固定し，k を変化させたときの適合率，再現率，F 値の遷移
Fig. 6　Transition of precision, recall, and F-measure by fixing α and changing k for test set 1.

比較手法を上回っていた．

また，正解セット 1 において各カテゴリで F 値が最大となったときの k の値を固定して α の値を変化させたときの F 値，再現率，適合率の値の変化を**図 5** に示す．図中の赤の実線が再現率，赤の点線が適合率，赤の波線が F 値を表している．さらに，図中の青の実線，点線，波線はそれぞれ，全カテゴリにおいて F 値が最大となったときの k の値を固定して α の値を変化させたときの再現率，適合率，F 値を示している．同様にして，正解セット 1 において各カテゴリで F 値が最大となったときの α の値を固定して k の値を変化させたときの再現率，適合率，F 値の変化を**図 6** に示す．さらに，それぞれの場合の正解セット 2 における再現率，適合率，F 値の変化を表したグラフを**図 7** および**図 8** に示す．

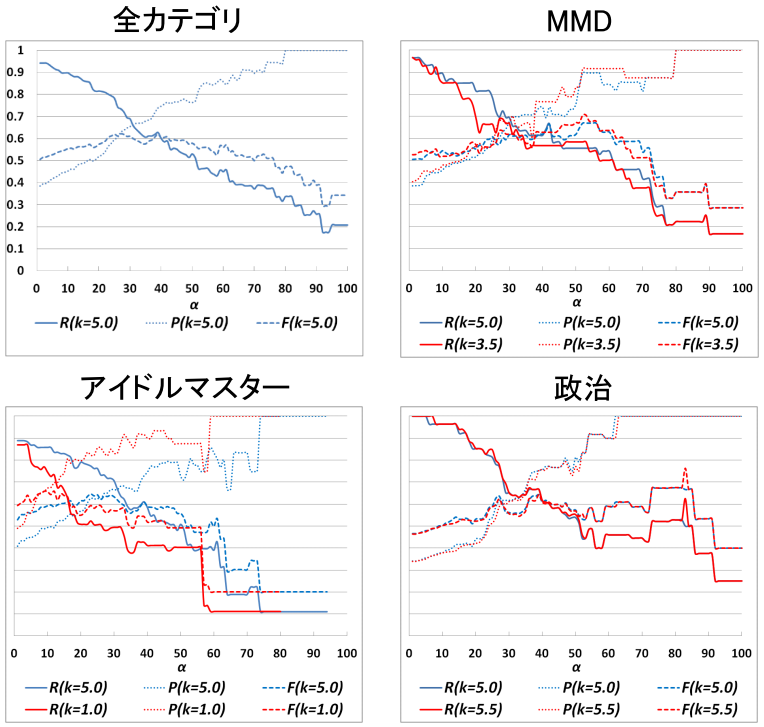

図 7　正解セット 2 において k を固定し，α を変化させたときの適合率，再現率，F 値の遷移
Fig. 7　Transition of precision, recall, and F-measure by fixing α and changing k for test set 1.

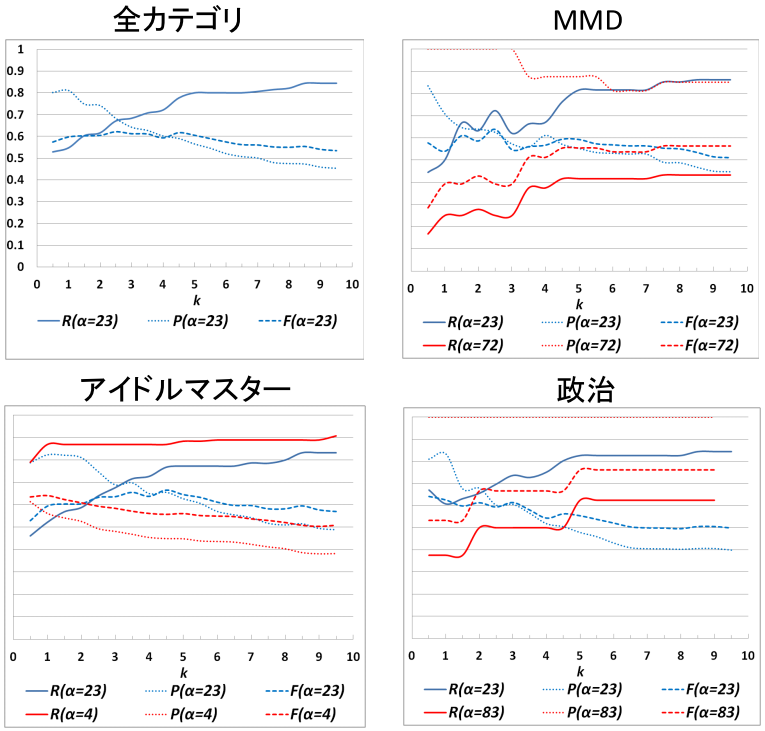

図 8　正解セット 2 において α を固定し，k を変化させたときの適合率，再現率，F 値の遷移
Fig. 8　Transition of precision, recall, and F-measure by fixing α and changing k for test set 1.

5. 考　　察

4.4 節の結果より，正解セット 1 における全カテゴリの F 値の最大値は 0.629 であり，比較的高い精度で注目シーンの特定が行えているといえる．被験者がスコアとして 1 をつけたシーンとしては，画面に少数の人物がある程度の大きさで映っていたり，会話をしたりしているものが多かった．よって，映像に付与されたコメントを解析することで，映像のタイトルやタグ，説明文には書かれていないが映像に登場している人物の判定などを行うことができると考えられる．カテゴリ別に見ると，特に政治のカテゴリにおいては F 値の最大値が 0.747 と高い値になった．また，図 5 および図 6 より，k の値を固定して α の値を増加させると適合率は上昇し，再現率は低下することが分かる．これは，α の値を増加させることで，注目シーンとして抽出されるシーンの数が減少するためである．逆に，α の値を固定して k の値を増加させると，適合率は低下し，再現率は上昇することが分かる．これは，k の値を増加させることで，閾値 α を超えやすくなったためである．また，図 6 を見ると，α の値を固定して k の値を変化させても，F 値はそれほど変化していないことが分かる．つまり，k の値を変化させることで，適合率と再現率のどちらを重視するかを選択できるといえる．図 5 を見ると，「政治」カテゴリにおいて，$k = 5.0$ で α の値が小さいとき，他のカテゴリよりも再現率が高くなっている．これはつまり，k と α が他のカテゴリと同じ値でも，閾値 α を超えやすいということであり，すべての映像に用いたコメントの数が 1,000 件であ

るにもかかわらず，このような差が生じたということは，「政治」カテゴリにおいては，他のカテゴリに比べて，登場人物が活躍しているシーンにコメントが集中的に付与され，登場人物が活躍していないシーンではコメントはあまり付与されないということを表している．

正解セット 2 の全カテゴリの F 値の最大値は 0.560 であり，こちらも比較的高い精度で注目シーンの特定を行えていた．被験者がスコアとして 2 をつけたシーンとしては，視聴者を笑わせるような言動や感動させる言動をとっているシーンが多かった．

図 7 と図 8 を比較すると，特にアイドルマスターのカテゴリにおいては，k を変化させたときよりも α を変化させたときのほうが各値におおきな影響が出ていることが分かる．これは，他のカテゴリに比べて登場人物に対するコメントが映像の再生時刻によらず万遍なく分散していることを表している．このことは，図 7 において，α の値を増加させていくと，ある α の値を境にして，再現率が大きく落ちることからも正しいと考えられる．つまり，アイドルマスターのカテゴリのように，コメントが万遍なく分散しているカテゴリにおいては，α の値をうまく決めることが重要であるといえる．

また，登場人物の名前が言及されていない場合であっても，映像中のコメントには登場人物の名前が含まれていなくてもその登場人物に特有の定型語を利用することで登場人物を特定することは可能であると考えられる．このような定型語の例として，「弱音ハク」という人物に対する「ツマンネ」という定型語や，「ドナルド」という人物に対する「ランランルー」という定型語があげられる．これらの語はいずれも，ある映像がきっかけとなって使用され始めた語であり，各人物の名前とはまったく無関係である．しかし，時間の経過とともに，使用され始めるきっかけとなった映像とは関係なく，これらの人物に向けて付与されるという現象が見られる．このような定型語を特定する手法の 1 つとして，以下のような方法があげられる．まず，ある登場人物の名前を含むコメントが付与された時刻と近い時刻に付与された，いずれの登場人物の名前も含まないコメントを収集する．これを複数の動画の様々な登場人物に対して行う．収集された各登場人物の周辺コメントの形態素解析を行い，各登場人物間で共通に得られた語やある登場人物から多く得られる語を求める．ある登場人物のみから多く得られる語の中で人物名辞書に登録されていない語は，その人物に対して付与される定型語である可能性が高いといえる．つまり，登場人物の名前を含むコメントが付与されていない映像であっても，そのような定型語の有無を検証することで，視聴者の注目を集める言動をとっている人物の特定ができると考えられる．

6. ま と め

本稿では，ある登場人物が注目されているシーンではその登場人物に対して映像の視聴者がコメントを付与するという仮説を立て，その検証を行った．仮説を検証するために，映像に付与された時刻同期コメントを用いて映像に登場する人物が視聴者の注目を集める言動をとっているシーンを推定し，それぞれのシーンにおいてその人物に対する注目の大きさを推定する手法の提案を行った．本稿で提案した手法では，登場人物の名前を含むコメントの分布から各登場人物が視聴者の注目を集めているシーンを，登場人物の名前を含むコメントの数とその周辺の感情を含むコメントの数からその登場人物に対する注目の大きさを求めた．

実験では，被験者が注目を集める言動をとっていると判定した登場人物について，そのシーンの特定の精度を検証した．その結果，いずれのカテゴリにおいても比較的高い精度で各登場人物が注目されているシーンを特定することができていた．一方で，最適なパラメータの値はカテゴリにより大きく変わるため，今後はコメントの傾向とパラメータの値の関係をより詳細に考察することで，適切な値を推定する手法を考える必要がある．さらに，現在は名前を含まないコメントが向けられる対象となる人物を十分に考慮できていないが，たとえばある人物とともに出現しやすい語を求めることで，名前が明記されていなくてもコメントの対象となる人物を判定できるようにする予定である．

本稿ではニコニコ動画に投稿された映像に付与されたコメントの観察に基づいて仮説を立て，その検証を行ったが，映像とその再生時刻に沿ったコメントの取得はニコニコ動画以外でも可能である．Youtube では映像のタイムコードに結び付けたコメントをつける機能があるが，これは他のユーザに対して映像の特定の部分を紹介するためのインデックスとして用いられることが多い．Youtube でタイムコードと結び付けられたコメントの中には，映像の内容を説明するときに「誰が何をしている」のように登場人物の名前を含むコメントが付与されることもあるが，Youtube でのタイムコードと結び付けられたコメント数が十分ではないため，本手法は Youtube に投稿された映像には適用不可能であると考えられる．他の例として，ネット掲示板「2 ちゃんねる[⋆1]」や短文投稿サイト「Twitter」があげられる．ネット掲示板「2 ちゃんねる」ではあるテレビ番組に対して専用のスレッドが作成され，その番組の内容に対して掲示板にコメントが投稿される．2 ちゃんねるにおいては，ニコニコ動画と同様に，ある登場人物が活躍しているシーンではその人物の名前を含む

⋆1 http://www.2ch.net/

コメントが掲示板に投稿される傾向にある．また，短文投稿サイト「Twitter」でも，テレビ番組に加えて，Ustream で配信される映像に対して Twitter でコメントを述べるという現象が見られる．Twitter でも，映像の内容に対するコメントでは登場人物の名前を含むコメントが投稿されることもある．そのため，本稿の提案手法を用いることでテレビ番組や Ustream で配信される映像についても，ある登場人物が活躍しているシーンを抽出できる可能性は高いと考えられる．今後は，これらのサービスについてもニコニコ動画とのコメントの傾向の違いを調査したうえで，提案手法の適用を試みる予定である．

謝辞　本研究の一部は，グローバル COE 拠点形成プログラム “知識循環社会のための情報学教育研究拠点”，特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」計画研究 “情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究”（研究代表者：田中克己，A01-00-02，課題番号 18049041），文部省科学研究費補助金若手研究（A）“インタラクティブな再ランキング・再サーチを可能とする次世代検索に関する研究”（研究代表者：中村聡史，課題番号 23680006）によるものです．ここに記して謝意を表します．

参 考 文 献

1) 橋本隆子，白田由香利，真野博子，飯沢篤志：TV 受信端末におけるダイジェスト視聴システム，情報処理学会論文誌：データベース，Vol.41, No.3, pp.71–84 (2000).
2) 宮森　恒：映像と音響情報の協調による内容検索のためのテニス動作自動注釈付け，電子情報通信学会論文誌 D-II，情報・システム，II–パターン処理，Vol.86, No.4, pp.511–524 (2003).
3) 大黒泰平，加藤友規，土居清之，亀山　渉：番組に対する視聴者入力情報からの時系列キーワード抽出の改善に関する検討，情報科学技術フォーラム一般講演論文集，Vol.3, No.2, pp.81–82 (2004).
4) 上原　宏，吉田健一：インターネット上の対話文に基づくドラマ番組の構造化：注目状態グラフによる視聴者コミュニティの嗜好パターン認識（マルチメディアとパターン認識理解，一般），電子情報通信学会技術研究報告 PRMU，パターン認識・メディア理解，Vol.104, No.369, pp.25–30 (2004).
5) 青木秀憲，宮下芳明：ニコニコ動画における映像要約とサビ検出の試み，情報処理学会研究報告 HCI，ヒューマンコンピュータインタラクション研究会報告，Vol.2008, No.50, pp.37–42 (2008).
6) Nakamura, S. and Tanaka, K.: Video Search by Impression Extracted from Social Annotation, *Web Information Systems Engineering–WISE 2009*, pp.401–414 (2009).
7) Diakopoulos, N. and Shamma, D.: Characterizing debate performance via aggregated twitter sentiment, *Proc. 28th International Conference on Human Factors in Computing Systems*, pp.1195–1198, ACM (2010).
8) Shamma, A., Kennedy, L. and Churchill, E.: Tweet the debates, *ACM Multimedia Workshop on Social Media* (*WSM*) (2009).
9) Uehara, H. and Yoshida, K.: Automating Viewers' Side Annotations on TV Drama from Internet Bulletin Boards, *IPSJ Digital Courier*, Vol.2, pp.145–154 (2006).
10) 宮森　恒，中村聡史，田中克己：番組実況チャットを利用したテレビ番組のメタデータ自動抽出方式，情報処理学会論文誌：データベース，Vol.46, No.18, pp.59–71 (2005).
11) Miyamori, H., Nakamura, S. and Tanaka, K.: Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web, *Proc. 13th Annual ACM International Conference on Multimedia*, pp.853–861, ACM (2005).
12) 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 = Transactions of the Japanese Society for Artificial Intelligence: AI, Vol.20, pp.67–75 (2005).
13) Masuda, T., Yamamoto, D., Ohira, S. and Nagao, K.: Video scene retrieval using online video annotation, *New Frontiers in Artificial Intelligence*, pp.54–62 (2008).

(平成 23 年 4 月 11 日受付)
(平成 23 年 9 月 12 日採録)

佃　　洸摂（学生会員）
京都大学大学院情報学研究科社会情報学専攻修士課程在学中．電子情報通信学会学生会員．

中村　聡史（正会員）
京都大学大学院情報学研究科社会情報学専攻特定准教授．2004 年大阪大学大学院情報学研究科博士後期課程修了．博士（工学）．主にヒューマンコンピュータインタラクション，ウェブ検索の研究に従事．日本データベース学会会員．

山本　岳洋（学生会員）

京都大学大学院情報学研究科博士後期課程在学中．2009 年京都大学大学院情報学研究科修了．ウェブ検索の研究・開発に従事．日本データベース学会学生会員．

田中　克己（正会員）

京都大学大学院情報学研究科社会情報学専攻教授．1976 年京都大学大学院博士前期課程修了．博士（工学）．主にデータベース，マルチメディアコンテンツ処理，ウェブ検索の研究に従事．IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，日本データベース学会各会員．