Vol.2014-DBS-160 No.18
Vol.2014-OS-131 No.18
Vol.2014-EMB-35 No.18
2014/11/18

画像特徴量とコメントを用いたニコニコ動画の指示的要約サムネイルの生成手法

松原 宏和[1,a)]　新妻 弘崇[1,b)]　太田 学[1,c)]

概要：ニコニコ動画や YouTube といった有名な動画共有サイトには膨大な動画が投稿されており，動画のあらすじは興味のある動画を効率よく探すために視聴者にとって大変有益である．そこで本稿では，動画の画像特徴量とコメントを用いて動画の重要場面を検出し，要約サムネイルを生成する手法を提案する．提案する動画要約手法の特徴は，画像特徴量の変化に基づいて検出した各シーンの動画をコメントに基づいて分類し，さらに各シーン中で最も盛り上がる場面のサムネイルを抽出する点にある．これにより，動画内容を網羅しつつ，視聴者が盛り上がる場面を抽出する．評価実験により提案手法の有効性を確認する．

A method of generating indicative summary thumbnails of Nicovideo using image features and comments

1. はじめに

近年，ニコニコ動画 [*1] や YouTube[*2] などの動画共有サイトが盛んに利用されている．ニコニコ動画では，2013 年 2 月時点で 2000 万件以上の動画が投稿されている [*3]．また YouTube では，2012 年 1 月時点で 1 分当たりおよそ 60 時間分の動画が投稿されており，1 日の動画視聴回数は 40 億回に及ぶ [*4]．この膨大な動画の中からユーザが自分に興味のある動画を探し出すのは困難である．そのためユーザが動画のあらすじを簡単に理解できたり，ユーザが自分に興味のあるシーンのみを選択できたりすることが重要となる．

ニコニコ動画と YouTube にはいくつか異なる点がある．最も顕著な違いは，ニコニコ動画と YouTube のコメントの投稿システムである．YouTube では，動画全体に対するコメントが投稿される．しかしニコニコ動画では，動画の再生中にコメントを投稿することができる．そのコメントは投稿された再生時刻に表示されるようになる．そのためニコニコ動画のコメントは，動画中の特定の再生時刻におけるシーンに対する視聴者の感想が反映されていると考えられる．また，ニコニコ動画のコメントに書き込める文字数は最大 75 文字である．このため視聴者は動画を見ている途中で短い感想を投稿する傾向がある．このことから，動画の各シーンはその再生時刻に投稿されたコメントにより様々なシーンに分類可能であると考える．

視聴者が動画を検索する際に参考となるメタデータとして，動画のサムネイル，タイトル，説明文がある．本研究では，動画内容の把握や検索を支援するためにサムネイルに注目し，これに多くのコメントから得られる情報を関連付ける動画要約を提案する．動画要約は，指示的動画要約と報知的動画要約の 2 タイプに分類できる [1]．指示的動画要約は，要約の元になった動画を見るかどうかの判断材料として用いられる．一方，報知的動画要約は，元の動画の代替として用いられ，元動画に含まれる情報を可能な限り含んだ要約となる．本研究では，ユーザが自分に興味のある動画を探す際に参考となる要約を目的としているため，指示的動画要約を対象とする．また，現在のニコニコ動画のサムネイルは，画像 1 枚が表示されているだけである．そこで例えば，動画中から重要な場面を抽出しそれを GIF アニメーションにして，サムネイルの代用にすれば，動画の選別に有効である．

[1] 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University
[a)] matsubara@de.cs.okayama-u.ac.jp
[b)] niitsuma@suri.cs.okayama-u.ac.jp
[c)] ohta@de.cs.okayama-u.ac.jp
[*1] http://www.nicovideo.jp/
[*2] http://www.youtube.com
[*3] http://ja.wikipedia.org/wiki/ニコニコ動画
[*4] http://youtubejpblog.blogspot.jp/2012/01/60-40.html

本研究では，ニコニコ動画のコメントに着目し，画像特徴量の変化を利用して分割した動画の各シーンを，動画の再生時刻でのコメントに基づいて分類し，各シーンで最も盛り上がる場面を抽出する．具体的にはまず画像の画素値を求め，その値が大きく変化する動画の再生時刻をシーンの切り替わりとして検出し，動画をシーンに分割する．次に，各シーンの動画の再生時刻に投稿されたコメントに基づいて，各シーンをいくつかのクラスに分類し，その分類結果に基づいて，各シーン中で最も重要と判断される再生時刻の画像を抽出する．さらに，抽出した画像に重要なコメントを関連付けることで要約サムネルを生成する．

本稿の構成は次の通りである．2 節で関連研究について述べ，3 節で提案する要約サムネイルの生成手法について説明する．4 節で実際に生成した要約サムネイルの実験について述べ，5 節でまとめと今後の課題について述べる．

2. 関連研究

2.1 シーン抽出

ニコニコ動画のコメントを用いて，視聴者の観点が入れ替わるシーンを抽出する手法が提案されている [2]．この研究では，視聴者のコメントを評価することで，動画の各シーンに対する視聴者の観点や反応を反映した特徴付けが行えると主張している．そのため，視聴者が盛り上がっている場面だけでなく，視聴者の観点が入れ替わるような重要なシーンの抽出が行える．コメントは肯定と否定で評価し，評価表現辞書 [3] を用いている．本研究では，このような肯定・否定でコメントを評価せず，コメントをいくつかのクラスに分類するのでその点が異なる．

動画のコメントを用いて，映像に登場する人物が注目されているシーンを抽出する研究もある [4]．あらかじめ用意した人物名辞書とコメントのパターンマッチにより，動画に登場する人物名の抽出を行う．そしてその人物名が含まれるコメントの前後 k 秒間を登場人物の登場シーンとして抽出する．登場シーンに投稿された登場人物を含むコメント数と，感情を表す表現を含むコメント数の和により登場シーンの注目度を求め，最も注目度の大きいシーンを推定する．ここでは，コメント数と感情を表す表現を含むコメントの数によりシーンの抽出を行っているため，批判で注目されているのか賞賛されているため注目されているのかなどは判別できない．しかし本研究では，コメントを「laugh」や「cry」などのクラスに分類するため，各シーンがユーザにとってどういったシーンであるかを提示することが可能である．

コメント数の変化を利用して動画の重要場面を抽出する研究もある [5]．この研究では，重要場面をキーフレームと呼ぶ．6 人のユーザが 5 本の動画からキーフレームを 1 秒単位で選ぶという実験の結果から，コメント数が多い時刻から 3 秒後が最もキーフレームとして適切であるとしている．そこで，コメント数の極大値を求め，その極大値の 3 秒後のフレームをキーフレームとして抽出した．この研究では，コメント数の変化のみを用いており，コメントの内容を考慮していない点が本研究とは異なる．

2.2 映像要約

ニコニコ動画のコンテンツを 30 秒の映像に要約する研究がある [6]．この研究ではコメント数が多い箇所を動画中で重要な箇所であると仮定し，要約映像を生成する．まず，動画の開始と終了からそれぞれ 3 秒間を取得し，これを要約映像の開始，終了とする．次に，コメント数が最も多い時刻の前 2 秒間，後 3 秒間を取得し，要約映像に追加していく．さらに，次にコメント数が多い時刻から同様に映像を取得し，要約映像に追加する．これを，要約映像が 30 秒になるまで繰り返す．この研究でも，コメントはその頻度のみを用いており，その内容は考慮していないため，この点が本研究とは異なる．

また楽曲動画限定で，視聴者の反応と音響特徴量を用いて 15 秒の動画を生成する研究も報告されている [7]．この研究では，視聴者の盛り上がり検出技術と，楽曲動画のサビ検出技術，その両方を組み合わせた手法によりサムネイル動画を自動生成する仕組みを実現している．サムネイル動画を 15 秒という短時間に設定したため，「オリジナル楽曲動画を見たいと思わせる，ユーザにとって質が高く感じるシーンを抽出すること」に注目している．この研究では，要約動画を生成するが，本研究ではサムネイル画像を生成する．そのため，ユーザは，動画タイトルや説明文などと同時に要約サムネイルを見ながら動画の一覧から，気に入る動画を探すことが出来る．

映像コンテンツとコメントを利用して指示的動画要約をする研究もある [8]．この研究では，時間的に隣接する二枚のフレーム間の画素値が，ある閾値より大きい箇所を動画のシーンの変わり目として検出し動画を分割している．そしてコメントの頻度を用いることで重要なシーンを 4 つ選択し，それぞれのシーンをコメントの頻度を用いて 10 秒に短縮することで動画を要約している．この研究も，コメントの内容は考慮していない．

3. 要約サムネイルの生成手法

本研究は，指示的動画要約により，短時間での動画の内容把握を支援することを目的とする．そのため動画中の類似した場面を削減する．

本節ではまず 3.1 節で動画のシーンの切り替わりを検出する方法，3.2 節で各シーンをコメントに基づいて分類する方法を説明する．3.3 節では各シーンでの重要場面の抽出法，3.4 節で抽出した重要場面とコメントに基づいて指示的要約サムネイルを生成する方法を説明する．

Vol.2014-DBS-160 No.18
Vol.2014-OS-131 No.18
Vol.2014-EMB-35 No.18
2014/11/18

3.1 画素値の変化を利用したシーンの切り替わり検出

本研究では，ある再生時刻 $t1$ から $t2(t1 < t2)$ までを「シーン」と呼び，ある再生時刻 $t3$ の画像を「フレーム」と呼ぶ．動画の映像には 1 秒間に数十フレームの画像が含まれている．本研究では，簡単のため，その毎秒数十フレームの画像を全ては利用しない．毎秒ごとに最初の 1 フレームの画像のみを代表フレームとして選択し，他のフレームは無視することで処理を簡単化する．以下全てこの毎秒 1 フレームを処理の対象とする．

画像特徴量を利用したシーン検出に関する研究は多くある [9][10] が，これらは映像のジャンルを絞り，そのジャンルに特徴的な映像の動きなどを利用している．本研究では，映像のジャンルを絞らずにシーンの切り替わりを検出するために，1 フレーム中の画素値の平均値のみを用いる．

一般的に画像は複数の画素で構成されており，それぞれの画素が RGB（赤，緑，青）の 3 色の情報を保持している．この RGB にそれぞれ 0 から 255 までの値を割り当てることで，様々な色を表現している．本研究では，RGB の値を平均したものを画素値，1 枚の画像の全ての画素値を平均したものを平均画素値と呼ぶ．この平均画素値の変化によりシーンの切り替わりを検出する．具体的にはまず，要約する動画の全フレームの平均画素値を計算する．次に，全フレームの平均画素値から，その標準偏差を計算する．最後に，動画の再生時刻順に隣り合う二つのフレームの平均画素値の差が標準偏差以上であれば，後ろのフレームの再生時刻をシーンの切り替わりとして検出する．動画のシーンの切り替わりを検出することで，動画を各シーンに分割する．

3.2 コメントによるシーン分類とそれに基く動画の要約

ここでは，3.1 節の方法で分割した動画の各シーンをコメントを用いてクラスに分類し，それに基づいて動画を要約する方法を説明する．ニコニコ動画のコメント投稿時刻は，ミリ秒単位であるが，本研究では小数点以下を切り捨て，秒単位で扱う．またあらかじめ，全てのコメントのカタカナはひらがなに，「ぁ」「ぃ」「ぅ」「ぇ」「ぉ」の小文字は大文字に正規化する．以下でまず，各シーンの再生時刻に投稿されたコメントを用いたシーンの分類について説明する．

シーン分類ではまず，コメントを各クラスに分類する．分類するコメントのクラスを表 1 にまとめる．コメントと表 1 の各クラスの要素のパターンマッチによりコメントを分類する．ただし，マッチする要素が複数ある場合はクラスの重複を認める．また表 1 の要素はこれが全てではなく一部のみを示している．表 1 のクラス「scream」の要素は，浅井ら [11] が提案した叫喚フレーズ抽出を参考に，我々 [12] が提案した叫喚表現の抽出方法で抽出したものである．例えば，「きたあああああ」のように母音が連続するものは，「きたあ」のように母音を 1 文字に変換し，この「きたあ」を叫喚表現とする．分類のクラスとその要素は，「ニコニコ動画コメント@ばっちりサーチ.net」[*5] を参考にした．「ニコニコ動画コメント@ばっちりサーチ.net」ではコメントを 20 クラスに分類している．ニコニコ動画の「エンタメ・音楽」「生活・一般・スポ」「政治」「科学・技術」「アニメ・ゲーム・絵」の各ジャンルの総合ランキング Top100 の上位 10 件と下位 10 件の合計 100 件の動画のコメントを「ニコニコ動画コメント@ばっちりサーチ.net」で分類した．その結果，分類数上位 5 クラスは「笑い」「絶叫」「賞賛」「興奮」「疑問」であった．そこで本研究では「笑い」「賞賛」を「laugh」に「絶叫」「興奮」を「scream」，「疑問」を「question」とし，ニコニコ動画によく見られる「cry」と「greeting」を追加して 5 クラスとした．これにより，シーンの動画のコメントを全て分類する．

表 1　分類するコメントのクラスとその要素の例

クラス	要素
laugh	w，笑，わろた，わらた，吹いた，吹く， 腹筋，噴いた，噴く
cry	泣く，感動，涙，切ない，胸にこみ上げる，ぐっとくる，うるっとくる
greeting	う p，うぽつ，きょんにちは，ちょりーす，どうも，にゃんぱすー，おかえり，おかえりなさい，ばい，から来ました，のし，gj，ばいにー，さよなら，乙，じゃあの，あばよ，お疲れ，otu，otsu，nosi，またな，またね
question	?，なぜ，何故，どうして，なんで，何で，何を，どうやって，何処
scream	きたあ，うお※ 3.2 節で説明する

次にコメントの分類結果を用いてシーンを分類する．分類するシーンのクラスは，表 1 のコメントを分類するクラスと同じである．以下の式で，シーン毎にクラス c のスコア $Score_c$ を計算し，この値が最大となるクラスにそのシーンを分類する．

$$Score_c = \sum_{vpos=s}^{e} \frac{Comment_c(vpos)}{N_c}$$

ここでシーン k の開始時刻を s，終了時刻を e としている．$vpos$ は動画の再生時刻である．$Comment_c(vpos)$ は，ある再生時刻 $vpos$ でのクラス c に分類されたコメント数で，N_c は動画全体でクラス c に分類された全てのコメント数である．

次にシーンの分類結果を用いて動画を要約する．本研究では，ユーザの反応をシーンごと定める．しかし，各シーンをクラスに分類後，隣り合うシーンのクラスが同じ場合，ユーザの反応に変化がないとみなして，それらのシーンを結合する．例として，図 1 に再生時間が 60 秒の動画のシーンの切り替わりと，シーン分類の例を示す．横軸が

[*5] http://nicomment.batch-re-search.net

Vol.2014-DBS-160 No.18
Vol.2014-OS-131 No.18
Vol.2014-EMB-35 No.18
2014/11/18

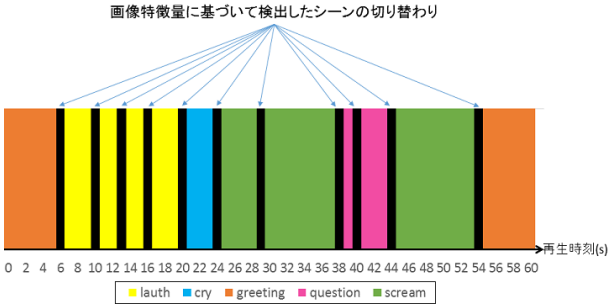

図 1 シーンの切り替わりとシーンの分類の例

再生時刻で，黒の縦縞がシーンの切り替わり時刻，各色が各クラスを表している．画素値に基づいて検出したシーンの切り替わり時刻は，6，10，13，16，20，24，29，38，40，44，54 秒である．よって画素値に基づけば，この動画は 12 シーンに分割できる．一方 6 秒〜19 秒の間にシーンの切り替わりが 3 つ存在するが，これらのシーンはコメントにより全て「laugh」に分類されているため，本研究ではこれらのシーンは結合する．そのためこの例では，0〜5 秒，6〜19 秒，20〜23 秒，24〜37 秒，38〜43 秒，44〜53 秒，54〜60 秒の 7 シーンとなる．画素値に基づいて検出したシーンは，12 シーンであったため，シーン数が削減されている．

3.3 コメントによる重要フレームの抽出

3.2 節の方法で要約した各シーンから重要フレームを抽出する方法を説明する．各シーンにおける重要フレームとは，各シーンの特徴をよく反映している箇所であると本研究では仮定する．動画の重要場面として，そのシーンで最もコメント数が多い箇所の 3 秒後とする方法 [5] やコメント数が多い箇所とする方法 [6] がある．しかしこの方法は，コメント数しか利用しておらず，コメントの内容を考慮していない．そこで我々は，3.2 節で説明した各シーンの分類結果を利用する方法を提案する．シーンの分類によって，各シーンにおけるユーザの反応がわかる．そこで各シーンの特徴を反映している箇所は，ある再生時刻における全てのコメント数に対して，そのシーンの分類クラスに分類されたコメント数が最も多い箇所であるとする．3.2 節で説明したように各シーンは 5 つのクラスのいずれかに分類される．よって以下の式でシーン中の再生時刻 $vpos$ におけるフレームの重要度を計算し，シーン中でこの値が最も高い時刻フレームを重要フレームとする．

$$Frame_{vpos} = \frac{Comment_c(vpos)}{Comment(vpos)}$$

$Comment_c(vpos)$ は，クラス c に分類されたシーンの再生時刻 $vpos$ でクラス c に分類されたコメント数であり，$Comment(vpos)$ はある再生時刻 $vpos$ での全てのコメント数である．

3.4 要約サムネイルの生成

3.3 節の方法で重要フレームを抽出できるが，そのフレームを見ただけでは，それがどうのような場面であるのかわからない場合がある．そこで 3.3 節の方法で抽出した重要フレームに，その場面での重要なコメントを付与して，要約サムネイルを生成する．具体的には抽出したフレームの再生時刻で，最も重要と判断するコメントを抽出し，そのフレームと関連付けるて要約サムネイルを生成する．重要コメントの抽出には，tfidf 法を用いる．コメント中の各単語 t_i の重要度 $tfidf_i$ を以下のように定義する．

$$tfidf_i = tf_i \times \log \frac{|D|}{df_i}$$

ここでは 1 動画の同再生時刻に投稿されたコメントの集合を 1 文書とみなす．D は総文書数であるから，60 秒の動画があり，その全ての再生時刻にコメントが投稿されていれば，60 となる．tf_i は，各文書中における単語 t_i の出現頻度である．出現文書頻度 df_i は，単語 t_i が出現する文書数である．ニコニコ動画のコメントは 75 文字の制限があり，文章としてはそれほど長くない．そのため重要コメントは，コメント中の単語の tfidf 値の総和が最も大きいものとする方法もある．しかし本研究の目的は，短時間での動画内容の把握を支援することにあるので，長いコメントを提示することは目的に反する．そのため以下の式で，抽出したフレームの再生時刻 $vpos$ のコメント $Comment_j$ の重要度 $Score_j$ を定め，これが最大のコメントを重要コメントとして抽出する．

$$Score_j = \frac{\sum_i tfidf_i}{L_j}$$

$\sum_i tfidf_i$ は，コメント $Comment_j$ 中の全ての単語の tfidf 値の和で，L_j はコメント $Comment_j$ を形態素解析することで得られる形態素数である．つまりこの重要度はコメント中の形態素数で正規化している．そして最も $Score_j$ が大きい重要コメントを重要フレームと関連付けることで要約サムネイルを生成する．

4. 評価実験

重要フレーム抽出法の有効性を確かめるために，3 節で提案した方法で生成したサムネイルと，重要フレームをそのシーンでコメント数が最大となる再生時刻のフレームとした場合の要約サムネイルを比較した．また 3 節の方法で得られるサムネイルのフレーム数が，画素値により検出した各シーンから得られたフレーム数と比べてどの程度少な

Vol.2014-DBS-160 No.18
Vol.2014-OS-131 No.18
Vol.2014-EMB-35 No.18
2014/11/18

くなるかについても実験した．さらに，要約サムネイルを被験者が評価した実験について説明する．これらの実験には，ニコニコ動画の動画を収集して利用した．また動画のコメントには，国立情報学研究所のダウンロードサービスにより株式会社ドワンゴが提供する「ニコニコ動画コメント等データ」を利用した [13]．

4.1 重要フレーム抽出に関する実験

3 節の方法で重要フレームを抽出して得られる要約サムネイルと，各シーンでコメント数が最も多い場面を重要フレームとする場合の要約サムネイルを比較した．表 2 に実験で使用する動画 *6 のメタデータを示す．またそれぞれの方法で抽出した重要フレームの再生時刻を図 2 に示す．図 2 の横軸は動画の再生時刻であり，縦軸は各再生時刻におけるコメント数である．各再生時刻中の黒の縦縞は画素値の変化により検出したシーンの切り替わりであり，そのシーンの間の各色がシーンの分類クラスである．灰色の丸印がコメント数最多の方法で得られるフレームの再生時刻であり，青色の三角印が 3 節の方法で得られる重要フレームの再生時刻である．図 2 より，この動画は，「笑い」で始まり，次にユーザが「疑問」に思う場面があり，「泣ける」場面があり，また「疑問」に思うような場面があり，「泣ける」場面があり，「叫ぶ」ような場面があり，最後に「挨拶」で終わることがわかる．コメント数が最多の場面を重要フレームとする場合の要約サムネイルを図 3 に示す．また 3 節で説明した提案手法によって生成される要約サムネイルを図 4 に示す．図 3，4 ともに要約サムネイル中のフレームを再生時刻順に並べている．左上が最も再生開始時刻に近く，右下が終了時刻に近いフレームなっている．各フレームの左上の番号は説明のためで，実際に生成する要約サムネイルにはない．

これら二つのサムネイルで大きく異なるフレームは，1，2，10 フレーム目である．図 3 の 1 フレーム目は料理の材料を示しており，図 4 の 1 フレーム目は料理名とその料理のイメージ図を示している．材料は字が細かく一見しただけでは，理解が難しい．そのため最初のフレームとしては図 4 の方が良いといえる．2 フレーム目は，かぼちゃを調理している場面から調理の説明に切り替わる場面と，かぼちゃを調理する場面である．図 2 を見ると，図 3 の 2 フレーム目は画素値により検出したシーンの切り替わりにあたる．そのためこれを見ただけでは，どのような映像であるか判断が難しい．このようにシーンの切り替わり上にあるフレームは，画像が見づらい可能性があるため，抽出しないなどの処理が必要である．またこのフレームを含むシーンは「question」に分類されている．実際にこのシーンの映像とコメントを見ると，映像に調理とは関係のない

*6 http://www.nicovideo.jp/watch/sm3190026

表 2 重要フレーム抽出実験に使用する動画のメタデータ

タイトル	ジャンル	再生時間 (s)	コメント数
魔女の宅急便に登場の，かぼちゃとニシンのパイを作ってみた	料理	413	4,0877

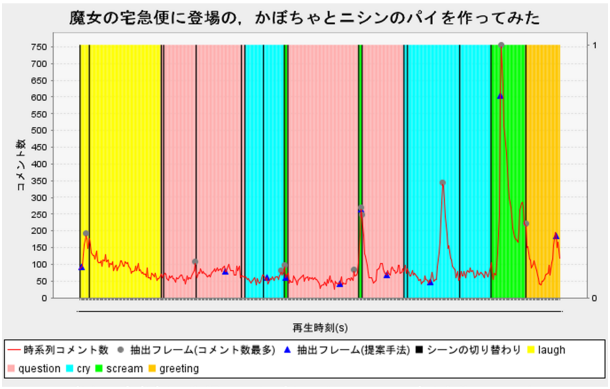

図 2 重要フレームの再生時刻

「ヤクルト」という飲料が映っており，それに対してユーザが疑問を感じているコメントが多く投稿されていた．図 4 の 2 フレーム目の右端に「ヤクルト」が映っており，このシーンのユーザの反応の理由がわかる．

図 3 の 10 フレーム目も 2 フレーム目と同様に，シーンの切り替わり上のフレームであるため，画像がわかりにくい．このシーンは動画の締めくくりの場面であり，シーンの分類は「greeting」である．多くのユーザがネット用語で「さよなら」や「またね」という意味で使われる「ノシ」というコメントを投稿しており，図 4 の 10 フレーム目ではシーンとコメントの同期が取れている．このことから，提案手法は，そのシーンの特徴をとらえることができているといえる．しかし，図 4 の 5，7 フレーム目はシーンの切り替わり上ではないが，少しわかりにくい画像を選択している．

図 3 の 10 フレーム目は完成した料理のシーンから動画の締めくくりに移り変わる場面であるが，これに関連付けられているコメントは完成した料理に対するコメントである．このコメントは一つ前の 9 フレーム目に関連付けられている方が望ましい．ユーザは動画を見てからコメントを書き込むため，映像に対して少し遅れてコメントが投稿されることがこの原因である．そのため得られたフレームの少し後のコメントを関連付けるなどの工夫が必要である．また図 3 の 3，7 フレーム目に関連付けられたコメントには，ニコニコ動画のコメントにおいて特徴的な「←」や「↑」などの「矢印」が含まれている．この「矢印」を含むコメントは他のユーザが投稿したコメントに対するコメントであり，動画に対するコメントでないことが多い．実際に図 3 の 3，7 フレーム目に関連付けられたコメントは，これら

Vol.2014-DBS-160 No.18
Vol.2014-OS-131 No.18
Vol.2014-EMB-35 No.18
2014/11/18

図 3　コメント数最多のフレームを選択した要約サムネイルの例

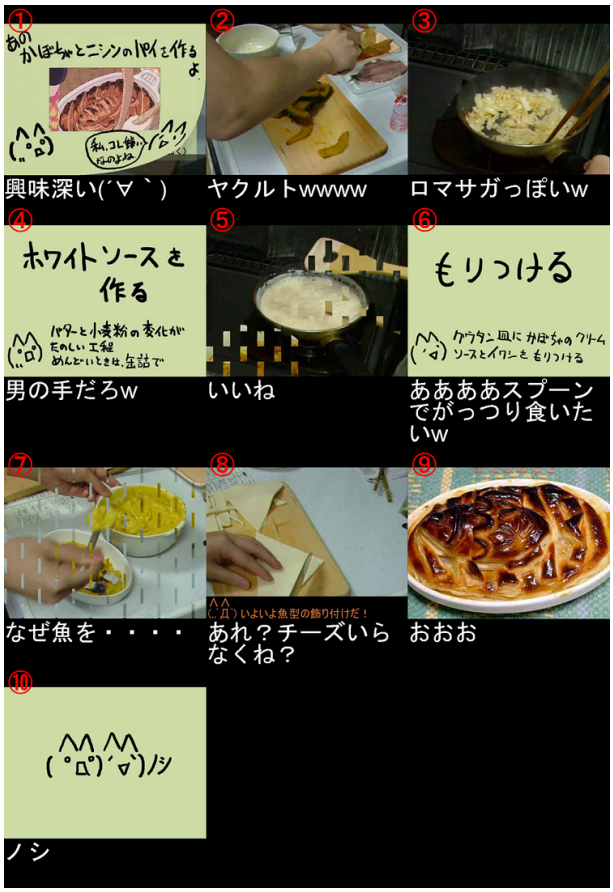

図 4　提案手法で生成される要約 (有) サムネイルの例

のフレーム画像とは直接は関係のないコメントである．フレームと関連付けるコメントも，シーンの分類結果を用いてそのフレームと関連のあるものとすべきと考える．

4.2　要約サムネイルに関する実験

生成した指示的要約サムネイルが動画をどの程度要約できているかを調べた．これを確認するために，下記の二つの方法で要約サムネイルを生成した．一つ目は，3.2 節で説明した提案手法で生成したサムネイルである．これを要約有サムネイルと呼ぶ．二つ目は，3.2 節で説明した方法でシーンを分類するが，隣り合う同じクラスのシーンを 1 シーンにまとめずに，画素値により検出した各シーンから重要フレームを抽出する方法で生成したサムネイルである．これを要約無サムネイルと呼ぶ．図 5 に，要約有と無のサムネイルの各フレームの再生時刻と，各シーンの分類結果を示す．図 5 で，灰色の丸印が要約無サムネイルのフレームの再生時刻であり，青色の三角印が要約有サムネイルのフレームの再生時刻となっている．得られるフレーム数は，要約有サムネイルが 10，要約無サムネイルが 17 であるため，これらには 7 フレームの差がある．ここで，要約有サムネイルが要約無サムネイルに対してどれだけフレーム数を削減したか確認するため，削減率を以下のように定義する．

$$\text{削減率} = \frac{\text{要約有サムネイルのフレーム数}}{\text{要約無サムネイルのフレーム数}} \times 100(\%)$$

そうすると表 2 の動画の削減率は約 59% である．

要約されたフレームを具体的に考察する．要約有サムネイルはすでに図 4 に示した．同じ動画の要約無サムネイルを図 6 に示す．まず図 4 と図 6 の大きな違いは，図 4 の 2 フレーム目と図 6 の 3，4，5 フレーム目に見られる．これらのフレームを含むシーンでは，かぼちゃを調理している．図 6 では，かぼちゃを調理するフレームが 3 フレームあるが，図 4 ではこれを 1 フレームに要約している．次に，図 4 の 3 フレーム目と図 6 の 7，8 フレーム目を比較してみる．これらのフレームはたまねぎを調理しているシーンである．先程と同様に，図 6 ではたまねぎの調理場面が 2 フレームであるが，図 4 では 1 フレームに要約している．しかし料理の手順は図 6 の 3 フレーム目や 6 フレーム目を見ると良く分かる．本研究では，汎用的な動画要約を目指したが，例えばこのような料理動画では，その要約映像を見るだけで料理手順を確認できれば有用と考えられる．このような動画のジャンルに対応した動画要約は今後の課題としたい．

その他の動画に対しても実験を行い削減率を確認した．

Vol.2014-DBS-160 No.18
Vol.2014-OS-131 No.18
Vol.2014-EMB-35 No.18
2014/11/18

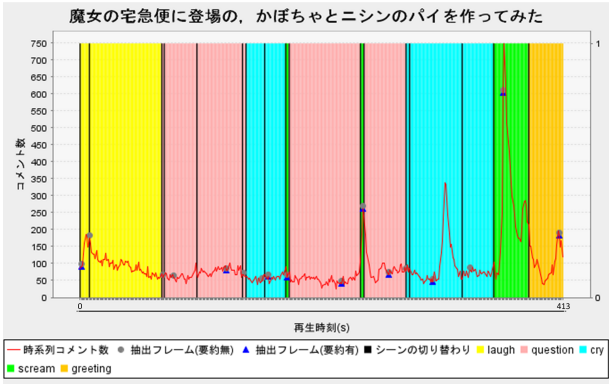

図 5　要約有と無のサムネイルで抽出するフレームの再生時刻

実験に使用した動画 *7*8*9 とその削減率を表 3 にまとめる．表 3 でジャンルが「料理」の動画は，表 2 の動画のことである．また「無」は要約無サムネイルのフレーム数，「有」は要約有サムネイルのフレーム数である．いずれの動画も，要約有サムネイルの方が要約無サムネイルよりフレーム数は少ないが，その削減率には幅がある．また，各動画の全フレームの平均画素値の標準偏差を計算すると表 4 のようになった．それぞれの動画を視聴すると，ジャンル「動物」の動画は，映像にほとんど動きや変化がなかった．つまりこの動画は，平均画素値の時系列の標準偏差が相対的に小さく，実際に映像の変化も少ない．しかし本研究ではコメントを利用することで，映像にほとんど変化がなくてもシーンの切り替わりを検出でき，シーン数の削減が可能である．またジャンル「スポーツ」の動画は，再生時間はジャンル「料理」の動画より短いが，フレーム数は数倍多い．さらに平均画素値の標準偏差は「料理」の動画より小さく，映像の変化が少ないことがわかる．実際に「スポーツ」の動画を視聴すると，この動画はサッカーの試合に関するもので，主に選手やサッカーグラウンドが画面に現れた．登場物の変化が少ないために標準偏差が小さくなっている．しかしゴールシーンなどでのエフェクトや，選手やグラウンドだけでなく観客などに切り替わる場面があるため，これらのシーンで切り替わりが検出されやすくなり，フレーム数が多くなっている．

4.3　被験者実験

本研究で生成した要約サムネイルは，ユーザが動画を選別する際に利用することを想定している．そのため，ユーザが要約サムネイルを見て予想されるその動画に対する期

*7　【実況 デンマーク戦全ゴール】FIFA ワールドカップ 日本代表ハイライト：http://www.nicovideo.jp/watch/sm11177576
*8　勝手に入るゴミ箱作った：http://www.nicovideo.jp/watch/sm18391671
*9　怒っていた猫が急に話しかけて来たけど、ネコ語だからわからない：http://www.nicovideo.jp/watch/sm11126185

図 6　要約無サムネイルの例

表 3　削減率を求めた動画のメタデータ

動画	ジャンル	再生時間 (s)	フレーム数		削減率
			無	有	
ニシンのパイ	料理	413	17	10	59%
W 杯日本代表	スポーツ	373	50	39	78%
動くゴミ箱	ニコニコ技術部	183	12	6	50%
怒る猫	動物	55	6	4	67%

待値と実際の動画の評価値は一致することが望ましい．被験者に要約サムネイルと動画を評価させ，要約サムネイル

Vol.2014-DBS-160 No.18
Vol.2014-OS-131 No.18
Vol.2014-EMB-35 No.18
2014/11/18

表 4　動画の平均画素値の標準偏差

ジャンル	標準偏差
料理	41.2
スポーツ	10.5
ニコニコ技術部	55.4
動物	3.3

を見ることで実際の動画の評価を予想できるか検証した．実験には，ニコニコ動画の「生活」と「科学・技術」の動画からそれぞれ 5 動画を選び使用した．これらの動画を岡山大学工学部の情報系の学部生・大学院生合わせて 9 名からなる被験者が評価した．

まず，被験者に上記 10 本の動画の要約サムネイルを評価してもらい，その数日後に実際の動画を見て評価してもらった．要約サムネイルの評価は，要約サムネイルを見て予想されるその動画に対する期待値を 5 段階で評価してもらう．また実際の動画の評価は，その動画の良し悪しを 5 段階で評価してもらう．要約サムネイルの評価ではタイトルと生成した要約サムネイル，実際の動画の評価ではタイトルと動画のみを提示した．

実験のために作成した要約サムネイルの評価システムの画面例を図 7 に示す．中央に生成した要約サムネイルがあり，これは GIF アニメーションとなっている．この GIF アニメーションは無限にループ再生される．要約サムネイルの下部に評価値の記入欄がある．被験者による評価結果を表 5 に示す．表 5 の T はユーザが要約サムネイルを閲覧したときの評価で，V は実際の動画を閲覧したときの評価である．この結果から，要約サムネイルと動画の評価の差は 0 から 2 が多い．しかし，User g のみ動画 2 と 8 で，評価の差がそれぞれ 4，3 となった．特に動画 2 では，他の User は要約サムネイルと動画の評価の差はあまり見られないが，User g のみ大きく差を付けている．被験者の中で User g のみがニコニコ動画を利用したことが無かったため，User g はタイトルや要約サムネイルを見るだけではどのような動画なのかを想像することが困難だった可能性がある．動画 2 は，猫に関する映像で，複数匹の猫の画像や動画を繋ぎ合わせて作成したものであった．User g は，動画 2 の要約サムネイルは猫の画像が多く出てくるのみで，猫以外の要素が伝わらなかったため要約サムネイルの評価を低くしていた．一方他の User は，タイトルと要約サムネイルを見ることである程度動画を想像できたと考えられる．

また，要約サムネイルの期待値 T に対して，実際の動画の評価値 V を真値とすると平均 2 乗誤差（Root Mean Squered Error）は以下のようになる．

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - V_i)^2}$$

表 6 に，全 User と User g を除いた場合の T の平均値，V の平均値，平均 2 乗誤差をまとめる．User g を除いた場合の，平均 2 乗誤差は 1 未満でありため，普段ニコニコ動画を利用する User は，要約サムネイルを見ることでその動画への評価を予想することがある程度できていると言える．

図 7　実験システムの例

表 5　被験者による要約サムネイルと動画の評価結果

User / 動画＃	a		b		c		d		e	
	T	V	T	V	T	V	T	V	T	V
1	3	4	4	4	4	4	4	3	3	2
2	3	4	4	5	3	3	5	5	3	3
3	4	5	3	2	5	5	3	3	3	2
4	5	3	4	3	5	4	4	3	4	4
5	4	5	4	4	4	5	4	5	3	2
6	5	4	4	4	4	5	5	5	5	5
7	4	4	4	4	5	5	5	5	3	3
8	3	5	5	4	4	5	3	4	4	5
9	4	4	4	3	4	5	4	4	3	4
10	2	3	3	3	5	5	4	3	3	3

User / 動画＃	f		g		h		i	
	T	V	T	V	T	V	T	V
1	3	4	3	5	3	2	3	4
2	5	5	1	5	5	5	4	5
3	5	3	4	5	4	3	4	4
4	5	4	2	4	5	4	5	5
5	4	5	5	5	5	5	5	4
6	4	4	2	5	5	5	4	5
7	4	3	5	5	4	5	5	4
8	3	5	1	4	3	5	2	3
9	3	4	4	5	5	5	4	5
10	2	2	3	5	4	3	5	4

表 7 に全被験者の要約サムネイルによる期待値と動画の評価値の対応をまとめた．表 7 より，要約サムネイルの期待値と実際の動画の評価値が一致しているものは 33，その差が 2 以上のものは 11 ある．この 11 のうち 9 は要約サムネイルの期待値が低く，実際の動画の評価値は高いものである．よって動画の魅力が要約サムネイルに十分反映されていない可能性がある．この 9 のうち 6 は User g が与えた評価のため，ユーザの分析も合わせて必要であると考えている．

4.4　考察

本研究では，シーン分類を利用して動画を線形に要約す

表 6　T，V の平均値とその平均 2 乗誤差

	T の平均値	V の平均値	RMSE
全 User	3.843	4.129	1.121
User g を除いた場合	3.938	4.001	0.929

表 7　期待値と評価値毎の評価数

		要約サムネイル					
		5	4	3	2	1	計
動画	5	15	16	5	1	1	38
	4	8	11	7	1	1	28
	3	2	8	6	2	0	18
	2	0	0	5	1	0	6
	1	0	0	0	0	0	0
	計	25	35	23	5	2	90

る手法を提案し，比較的短時間の動画を実験に用いてその有効性を評価した．[6] や [7] の動画要約に関する研究では，動画を 15 秒や 30 秒の映像に要約している．コメント数に基づいて動画を要約する研究 [6] では，コメント数が多い箇所が重要であるとし，コメント数が多い箇所から順にシーンを選択して要約動画を生成する．これらの研究では長時間の動画に対しても必ず一定長の映像に要約できる．一方，本研究の提案手法は線形要約のため長時間の動画を要約すると，大量の要約サムネイルが生成されることが予想される．そこで今後は，要約サムネイルの枚数に上限を定め，制限枚数内のサムネイルに要約方法を検討したい．

5. おわりに

本研究では，動画共有サイトにおけるユーザの動画選択を支援するために，ニコニコ動画の要約サムネイルの生成法を提案した．本研究では，画像の画素値を用いることで動画のシーンの切り替わりを検出し，動画のコメントに基づいてシーンを分類することで，シーン数を削減した．さらに，コメントを用いて重要な場面を抽出し，さらに重要コメントをその画像に結び付けて要約サムネイルを生成した．

重要フレーム抽出に関する実験では，コメントによるシーンの分類結果を用いることで，そのシーンの特徴的なフレームを抽出できることがわかった．また要約サムネイルに関する実験では，画素値のみによるシーン検出で得られるサムネイルに比べて，22%〜50% 削減できることがわかった．被験者実験では，普段ニコニコ動画を利用しているユーザに対しては，要約サムネイルは有効であることがわかった．しかし，要約サムネイルを見て抱く動画の期待値が実際の動画の評価より低くなるものもあるため，これらを詳しく分析する必要がある．またシーン分類のクラス数を本研究では 5 としたが，参考にした「ニコニコ動画コメント@ばっちりサーチ.net」では，コメントを 20 クラスに分類しており，ニコニコ動画のコメントに特徴的なクラスもについても検討する必要がある．画像とコメントの関連付けの方法についてもさらに洗練したい．また要約サムネイルの枚数に上限を定め，重要シーン抽出などを行いながら，制限枚数内のサムネイルに要約する方法についても検討したい．

参考文献

[1] 奥村学，“テキスト自動要約（＜特集＞自然言語による情報アクセス技術）”，情報処理，Vol.45，No.6，pp.574-579，2004.
[2] 山内嶺，北山大輔，“ダイジェスト映像自動生成のための観点の入れ替わりに基づいた特徴的シーン抽出”，第 6 回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2014），F4-2，2014.
[3] 高村大也，乾孝司，奥村学，“スピンモデルによる単語の感情極性抽出”，情報処理学会論文誌ジャーナル，Vol.47，No.02，pp.627-637，2006.
[4] 佃洸摂，中村聡史，山本岳洋，田中克己，“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”，情報処理学会論文誌，Vol.52，No.12，pp.3271-3482，2011.
[5] 谷直紀，山崎俊彦，相澤清晴，“コメント数の動的な変化を利用した CGM 動画要約”，電子情報通信学会総合大会講演論文集，D-12-5，2009.
[6] 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 2008-HCI-128/2008-MUS-75，Vol.2008，No.50，pp.37-42，2008.
[7] 中村聡史，山本岳洋，後藤真孝，濱崎雅弘，“視聴者反応と音響特徴量に基づくサムネイル動画の生成手法”，情報処理学会論文誌，Vol.6，No.3，pp.148-158，2013.
[8] 于多，高間康史，“映像コンテンツとコメントを利用した指示的動画要約生成・提示手法の提案”，人工知能学会 インタラクティブ情報アクセスと可視化マイニング研究会 (第 5 回)　SIG-AM-05-05.
[9] 三浦宏一，浜田玲子，井手一郎，坂井修一，田中英彦，“料理映像の特徴を利用した要約手法の検討”，電子情報通信学会技術研究報告，PRMU2002-22，pp.15-20，2002.
[10] 出口嘉紀，吉高淳夫，“映画の文法に基づく要約映像の生成”，情報処理学会研究報告，データベース・システム研究報告，Vol.2004，No.3，pp.33-40，2004.
[11] 浅井洋樹，秋岡明香，山名早人，“きたあああああああああああああああああ！！！！！ 11: マイクロブログを用いた教師なし叫喚フレー ズ抽出”，第 5 回データ工学と方法マネジメントに関するフォーラム（DEIM Forum 2013），A4-4，2014.
[12] 松原宏和，新妻弘崇，太田学，“動画共有サイトにおける動画ナビゲーションのためのコメント要約手法”，第 6 回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2014），F4-3，2014.
[13] 大学共同利用機関法人情報・システム研究機構国立情報学研究所 (NII)：情報学研究データリポジトリ
http://www.nii.ac.jp/cscenter/idr/nico/nico.html