

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

Proposal of Indicative Video Summary Generation and Presentation Method using Video Contents and Comments

于 多¹ 高間 康史^{1*}

Duo Yu¹ and Yasufumi Takama¹

¹ 首都大学東京大学院システムデザイン研究科

¹Graduate School of System Design, Tokyo Metropolitan University

Abstract: 本稿では、ニコニコ動画などのコメントが付与された動画から指示的要約を生成する手法を提案する。動画は Web 上の主要コンテンツの一つであるが、テキストコンテンツと異なり、見るべき動画の効率的発見は困難である。提案手法では、映像の特徴とコメントの関連性を考慮し、多数のユーザが共感している区間を複数抽出する。抽出した各区間から GIF アニメーションを生成し、それらを並列提示することで効率的な閲覧を支援する。本稿では並列提示に関する予備実験の結果について報告する。

1 はじめに

本稿では、ニコニコ動画¹などのコメントが付与された動画から指示的要約を生成する手法を提案する。2005年頃から、動画共有サービスが急速に発展してきている。動画共有サイトも次々と登場し、特に、YouTube²やニコニコ動画に代表される動画共有サイトの利用者が急増している。2011年のマイボイスコムインターネット調査結果[1]によれば、直近1年間に利用したことがある動画共有サイトは12,477名の回答者のうち「YouTube」が93.9%、「ニコニコ動画」が46.5%となっている。しかし、公開コストの低下により膨大な量の動画が投稿されるようになった結果、短時間で興味ある動画を探すことは困難となっている。

ユーザの動画検索や選択を支援する有効な手段の一つに動画要約が挙げられる。動画要約とは、動画内の各シーンがどのような内容を表現しているかを静止画あるいは短い動画により提示する手法である。利用目的に応じ、指示的動画要約と報知的動画要約の二つのタイプに分けられる[2]。指示的動画要約は、元動画を観るべきかどうか、自分の関心に合うかどうかを判断する際の手がかりとして用いられる要約であり、動画の重要な部分や視聴者の関心を引く部

分を抽出して提示する。これに対し、報知的動画要約は元動画の代替として用いる要約であり、動画の主要な内容を、全体構成もわかるように提示する必要がある。本稿では、指示的動画要約を対象とする。

一方、ニコニコ動画をはじめとする、動画にリアルタイムのコメントを付加できるサービスが注目されている。このサービスにより、ある特定のシーンに対して投稿されたコメントは右から左へ流れるように動画の画面上に表示される。このようなコメントは動画コンテンツと関連するため、指示的動画要約生成の際に、重要な手がかりとして利用可能と考える。

本稿では、ニコニコ動画を対象として指示的動画要約を生成する事を目的として、映像コンテンツとコメントの両方を考慮した重要区間抽出手法を提案する。また、抽出した各区間を並列に提示する手法も提案する。提案手法では、映像の特徴とコメントの関連性を考慮し、元動画から多数のユーザが共感している区間を複数抽出する。抽出した各区間から GIF アニメーションを生成し、それらを並列提示することで、効率的な閲覧を支援する。本稿では、複数区間の並列提示手法に関する予備実験の結果について報告する。

2 関連研究

動画要約に関する研究には、これまで様々な種類

¹ <http://www.nicovideo.jp/>

² <http://www.youtube.com/>

*連絡先：首都大学東京大学院システムデザイン研究科

〒191-0065 東京都日野市旭が丘 6-6

E-mail: ytakama@sd.tmu.ac.jp

や手法がある。本節では、それらを三つの部分に分けてまとめる。

2.1 映像コンテンツを用いた映像要約

映像要約に関する従来研究では、映像コンテンツ内の特徴を用いたものが多い。料理映像の特徴を利用した要約手法[3]では、調理の全体的な流れを視覚的・直感的に理解することを目的としている。調理動作および料理や材料を示す部分を特に重要部分とし、動き検出により重要部分を決定し、要約映像を作成する。

映像中の情報を用いたニュース映像要約手法[4]では、ニュース映像の全体を把握することを目的として「冒頭部」を重要部分と定義し、カット点検出手法、人物領域認識手法とカメラワーク・テロップ検出手法を用いてこれを抽出することで、映像要約を実現する。これらの研究の多くは、特定ジャンルを対象を絞り、その対象に固有の特徴を有効活用しているが、本稿では対象に依存しない映像要約手法を対象とする。

また、映像要約の要素技術として用いられる代表的な技術にカット検出があり、さまざまな手法が提案されている。多次元特徴空間解析に基づく映像のカット検出手法[5]では、映像フレームから抽出した特徴量が特徴空間内で連続的に移動する特性を利用し、多次元特徴空間での軌跡を解析する。カット点検出手法に関する研究では、映像のジャンルにより用いられる特徴が異なるため、ジャンルに応じてカット点の基準を決めることが検出精度向上のために重要である。

2.2 コメントを用いた動画要約

YouTube やニコニコ動画など利用者が動画を投稿できるサービスの出現とともに、動画タイトルやタグ・コメントといった、投稿動画に付与されるテキストを利用した動画要約に関する研究が増加してきている。特に、ニコニコ動画は他の動画共有サイトと異なり、動画にリアルタイムのコメントを付加できるサービスを提供しているため、ニコニコ動画を対象とする研究が多く存在する。

青木らは、映像に伴う音楽のサビ部分の検出や映像要約の生成を、コメント頻度に基づき行う手法を提案し、その可能性について検証している[6]。この研究では、様々なジャンルの映像に対して評価した結果、うまくいくケースとそうでないケースがあることを示している。映像要約は一部失敗しているが、要約の正しさではなく面白さを評価するユーザにな

ら、受け入れられる可能性があるとしている[6]。多胡らは、動画に投稿されたコメントを意味的、量的に分析し、動画の各シーンに対するソーシャルノテーションとみなしてシーン検索・動画要約を行うシステムを提案している[7]。動画を等間隔のユニットに分割し、コメントに基づくキーワードオントロジーを利用してそれらにスコア付けする手法を用いる。キーワードオントロジーは、人手でキーワード群を構築したものであるため、精度が高いものの構築コストが問題になると考える。

2.3 動画要約の可視化手法

これまで、動画要約に関する研究の多くは元動画より短い動画を作成して提示する手法を採用している。代表的なものに、動画内容の把握や特定シーンの再視聴を支援するため、動画要約をいくつかのサブネイルで並べて表示する手法[8]や、動画要約画像を対応する時間軸に配置して可視化する手法[9]などがある。また、小関らは、動画から抽出した画像を漫画的技法を用いてアニメーション化する「ばらばらアニメ」を提案している[10]。

3 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手法

本稿では、効率的な閲覧を支援する指示的動画要約の生成を目的として、映像コンテンツとコメントを用いた重要区間抽出手法と並列提示手法を提案する。指示的動画要約を生成するため、動画の重要な部分や視聴者の関心を引く部分を抽出する必要がある。動画の重要な部分を抽出する際には、画素値など映像内の情報だけではなくコメントなど映像以外の情報も同時に分析することで、動画作成側と視聴者側の両方の視点を考慮することができると考える。

指示的動画要約生成手法では、図1に示すように、まず、ニコニコ動画から動画、基本情報（動画のID、タイトル、再生時間とコメント数など）とコメントを取得して記録する。次に、画素値を利用して元動画を分割する。分割した動画の各区間に対し、コメントを利用して重要度を付与し、重要度の高い順に四つの区間を選択する。さらに、選択した区間それぞれについて、コメントやキーワードが集中する部分空間を抽出することで区間の短縮を行う。

並列提示手法では、単一の動画を生成する従来手法とは異なり、短縮した10秒間の区間に対してそれぞれGIFアニメーションを作成し、それらを時間順に並べて表示する。

3 節では、区間分割、区間選択・短縮、並列提示のそれぞれについて具体的に説明する。

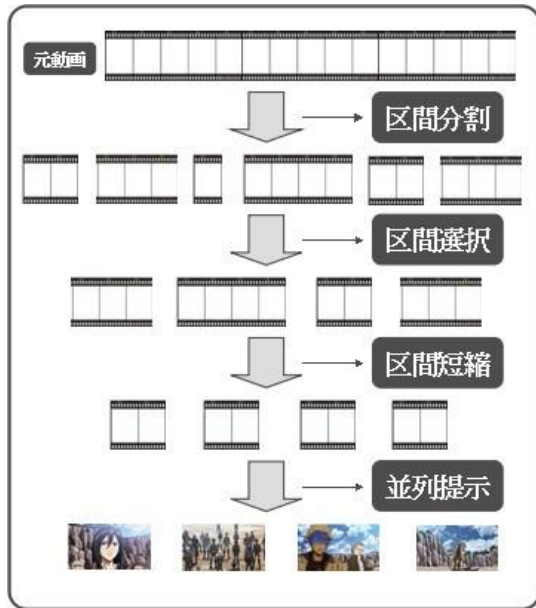


図1 指示的動画要約の生成手順

3.1 区間分割

区間分割では、映像の瞬間的な変わり目（カット点）を検出することで元動画を分割する。まず、元動画から連続しているフレームを全部取得する。時間的に隣接する二枚のグレースケール化したフレーム $f(t-1)$ と $f(t)$ の間で、対応する画素 (x,y) の濃淡値の差 I_t を式(1)により求める。

$$I_t = |I(x, y, t) - I(x, y, t-1)| \quad (1)$$

I_t がある閾値 d より大きい場合、 (x,y) を変化点としてカウントし、変化点の総数 S_t を計算する。図2は、「アニメ」、「ゲーム」、「音楽」、「エンタメ」、「生活」と「スポーツ」の6ジャンルからそれぞれ一般的な動画を取得し、 S_t を計算した結果である。横軸は、動画のフレーム数を示し、縦軸は S_t である。 S_t が大きい場合、フレーム間の変化が大きいとみなせる。図2を見ると、「ゲーム」ジャンルの動画は他のジャンルよりもフレーム間の大きな変化が多く発生していることがわかる。また、「アニメ」と「スポーツ」の動画は変化の大きいところが全区間にわたって比較的均一に分布しているのに対し、「音楽」の動画は特定区間に集中していることもわかる。さらに、「エンタメ」と「生活」の動画の多くは動画投稿者が撮影・編集したものであるため、他のジャンルよりも変化が少ないと考える。

連続する2枚のフレーム間で S_t が大きければカット点と判断するが、エフェクトなどの影響により、

実際にはカット点でない箇所も検出してしまうことがある。そこで、本稿では以降のプロセスで必要とするよりも多めにカット点を検出し、誤検出区間の統合などの後処理を行う。動画の長さに基づき、抽出する区間数 N を定め、カット点数を $2(N-1)$ と設定する。

後処理として、誤検出を除去するため短い区間を合併する。隣接する2つのカット点に挟まれた区間に対し、その長さが10秒あるいは元動画の4%の長さのいずれか長い方よりも短い場合に、前後に隣接する区間の短い方と合併する。

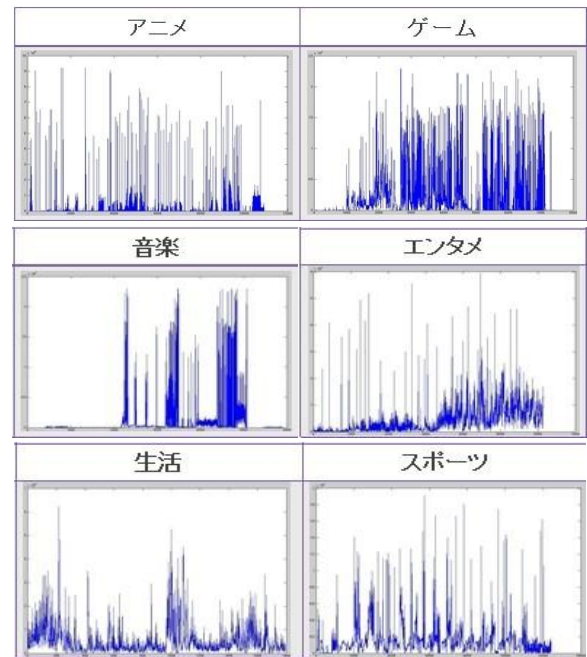


図2 フレーム間画素濃淡値の差

3.2 区間選択・短縮

区間の選択では、ユーザーの関心ある部分を選択することを目的として、3.1 節で得られた各区間にコメント頻度などを用いて重要度を付与し、重要度の高い区間を選択する。

具体的には、区間 k の1秒あたりのコメント数を C_k/T_k と設定する。ここで、 C_k を区間 k のコメント数、 T_k を区間 k の長さとする。また、区間 k に含まれるコメントの割合を C_k/C と設定する。 C は対象動画のコメント数である。以上に基づき、重要度 $IF(k)$ を式(2)より求め、この値が上位の4区間を選択する。

$$IF(k) = \frac{C_k}{T_k} \times \frac{C_k}{C} \quad (2)$$

区間短縮では、コメント数、およびキーワードを含むコメント数を利用することで区間を10秒に短縮

する．ここで、キーワードとは、コメント中に頻出する文字列とする．本稿では対象ジャンルを限定しない要約生成手法を目指すため、キーワードを辞書として用意しておく方法は適さないと考える．そこで、文字Nグラムを採用してキーワードを抽出する．Nは3とする．Nグラムを採用することにより、笑いを表現する「www」や「あああ」などの感嘆表現をキーワードとして扱うことが可能となる．

長さが10秒以上の区間 k から部分空間を抽出する処理について、図3を用いて説明する．部分空間の長さは10秒に固定し、区間 k の先頭から5秒ずつずらしながら、部分空間を抽出する．従って、隣接する部分空間 i と $i+1$ は5秒のオーバーラップを持つ．

部分区間 i のコメント数を C_i 、感嘆表現とキーワードを含むコメント数を K_i とする． $C_i \times K_i$ が最大となる部分空間を抽出し、3.3 節での GIF アニメーション生成対象とする．

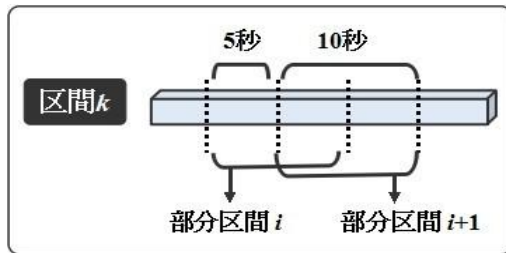


図3 部分区間の定義

3.3 区間の並列提示

並列提示手法では、動画の要約結果を一目で把握可能とするため、3.2 節で抽出した4つの区間を時間順に並べて表示する．各区間は10秒のGIFアニメーションとして提示され、同時に再生される．各動画は10秒と短いため、繰り返し再生する．

4 並列提示に関する予備実験

4.1 実験概要

3.3 節で述べた複数区間の並列提示手法は、短時間で効率的に要約を確認できることが期待される一方、ユーザにとって認知的負荷が高い可能性もある．並列提示する区間数が増えるほど効率性は高くなるが、負荷は高くなる事が想定される．そこで、単一の動画として生成した動画要約を提示した場合と、並列提示手法の比較を行う．

本実験では、20代の工学系学部生・大学院生に、動画要約を見ながら対象動画に関する質問に解答してもらった．ニコニコ動画の「アニメ」から6動画を

を取得して、それぞれ単一のGIFアニメーション(1分割)、4つのGIFアニメーションの並列提示(4分割)、6つのGIFアニメーションの並列提示(6分割)の3種類の要約を生成した．6種類の動画に対し、作成した問題を図4に示す．出題数は4問であり、質問1と4は、内容の類似した動画や同じアニメーションの異なる話を組として、当てはまる動画を選択してもらう形式となっている．質問は、各動画の特徴または内容に関するものであり、実験協力者には各質問に対してそれぞれの動画要約を見てから解答してもらった．

◆質問<1>: 同じシーン(短い)が2回出てくるのはAとBのどちらだと思いますか.

質問<2>: 下の動画要約を見て、この動画と合っている台詞を選んでください.

A「うわああああ! すごい美人!」
B「うわああああ! すごいブス!」
C「すごい! 後ちょっとで溢れそう!」

◆質問<3>: 下の動画要約を見て、この動画と合っているサブタイトルを選んでください.

A「モテないし、ちょっとイメチェンするわ」
B「モテないし、昔の友達に会う」

◆質問<4>: 家族は危険な目に遭ったシーンが出てくるのは、AとBのどちらだと思いますか.

図4 実験に用いた質問

本実験では、同じ動画を複数回見ないように配慮し、実験協力者を3人ずつの6グループに分類した．そして、各実験協力者が6動画と3種類の要約を全部見るように、各要約と質問を組み合わせた．表1は3種類の要約提示方法と質問の組み合わせを示しており、これに基づき各実験協力者グループが解答した質問パターンを表2に示す．

表1 要約提示方法と質問の組み合わせ

	動画①②	動画③④	動画⑤⑥
	質問1	質問2、3	質問4
1分割	N1	N2	N3
4分割	N4	N5	N6
6分割	N7	N8	N9

表2 実験協力者グループと質問パターンの組み合わせ

グループ	実験協力者			1分割	4分割	6分割
A	A1	A2	A3	N1	N5	N9
B	B1	B2	B3	N4	N8	N3
C	C1	C2	C3	N7	N2	N6
D	D1	D2	D3	N1	N8	N6
E	E1	E2	E3	N4	N2	N9
F	F1	F2	F3	N7	N5	N3

図 5 に、グループ B の実験協力者に対する質問と要約の例を示す。実験協力者が各質問に対して「Play」ボタンをクリックすると、解答時間の記録が始まり、解答の選択肢の一つをクリックすると、解答時間の記録が終わる。各 GIF アニメーションの再生時間は 10 秒だが、最後まで行くと先頭に戻って繰り返し再生される。



図 5 グループ B の実験ページの例

4.2 実験結果

表 3, 4, 5 に、各実験協力者の解答の正誤および解答時間について示す。ここで、実験協力者 F3 は他の協力者と比較して極端に長い解答時間となっている。また、実験協力者 B2, C1, C3 と F1 とは、対象動画について実験前に知っていた。

表 3 質問 1 の解答時間と正誤

質問 1			
	実験協力者	解答時間(秒)	正誤
1分割	A1	54	O
	D1	223	O
	A2	138	O
	D2	82	O
	A3	282	O
	D3	106	X
4分割	B1	51	O
	E1	82	O
	B2	73	O
	E2	31	O
	B3	26	O
	E3	35	O
6分割	C1	77	O
	F1	100	O
	C2	70	O
	F2	54	O
	C3	28	X
	F3	471	O

実験結果より、質問1と2についてはどの要約提示方法においてもほとんどの実験協力者が正解しているのに対して、質問3と4では不正解の人が多くなっていることがわかる。これらの結果は、要約提示方法よりも質問の難易度に依存すると考える。質問3は動画タイトルに関する質問であるため、動画タイトルの指す内容と視聴者の受け取り方が異なる可能

性があると考えられる。質問4は特定のシーンの解釈に関する問題であり、実験協力者ごとに解釈が異なったため正解者が少なくなったと考える。

表 4 質問 2 と 3 の解答時間と正誤

		質問 2		質問 3	
	実験協力者	解答時間(秒)	正誤	解答時間(秒)	正誤
1分割	C1	66	O	67	O
	E1	101	O	128	O
	C2	70	O	71	O
	E2	66	O	80	O
	C3	73	O	83	O
	E3	55	O	76	X
4分割	A1	26	O	31	X
	F1	56	O	81	O
	A2	37	O	75	X
	F2	244	O	31	O
	A3	14	O	30	X
	F3	168	O	107	X
6分割	B1	64	O	78	O
	D1	51	O	60	X
	B2	32	O	26	O
	D2	55	O	96	X
	B3	27	O	35	X
	D3	110	O	53	X

表 5 質問 4 の解答時間と正誤

質問 4			
	実験協力者	解答時間(秒)	正誤
1分割	B1	348	X
	F1	121	O
	B2	31	O
	F2	34	X
	B3	122	X
	F3	186	O
4分割	C1	20	O
	D1	52	X
	C2	68	O
	D2	63	O
	C3	41	X
	D3	94	O
6分割	A1	23	O
	E1	136	O
	A2	212	O
	E2	64	X
	A3	163	O
	E3	32	X

図6, 7は、要約提示方法間で解答時間の平均と分散をそれぞれ比較したものである。これらの結果を見ると、質問2と3は解答時間の平均、分散共に小さいのに対し、質問1と4では解答時間の平均・分散共に大きくなっている。これは、質問1と4が比較問題であることが一因と考える。すなわち、基本的には二つの動画を視聴しなくてはならないため、単一の動画を対象とした場合よりも倍程度の時間がかかってしまうと言える。しかし、1分割よりも4, 6分割の方が解答時間が減少していることから、複数動画を並列提示する提案手法の有効性が確認できる。特に4分割が質問によらず解答時間が短く、分散も小さい傾向にあることから、認知的負荷の点でバランスがとれていると考える。更に、質問1のように内容ではなく映像の特徴のみで解答できる場合は、4分割によ

り、短時間で正確な解答が可能と考える。

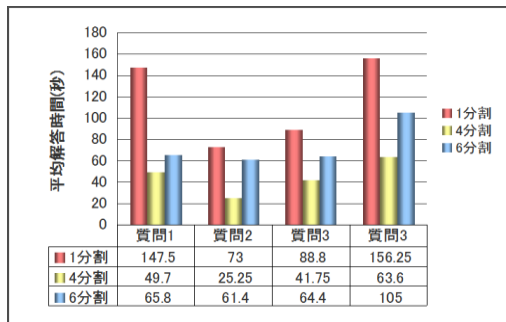


図 6 平均解答時間

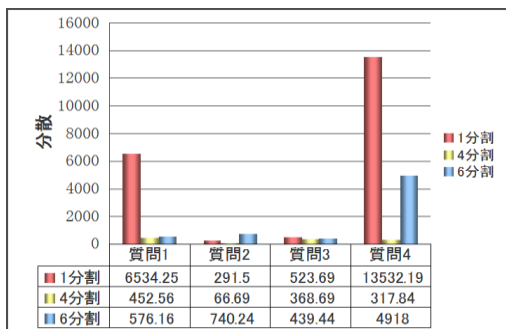


図 7 分散

実験後、3 種類の要約提示方法の見やすさについて実験協力者に感想を聞いたところ、質問 1 と 4 のように内容の近い動画の比較問題では、分割されていると集中して見るのが困難だと思う実験協力者が多数いた。また、6 分割が見やすいと思う実験協力者は一人もいなかった。これらの原因と考えられることとして、複数の動画が提示された場合、それらの時間的順序がわかりにくいという意見があった。そのため、時間的な順序がわかりやすいように提示する工夫が必要と考える。また、実験協力者の多くは 1 分割が最も見やすいと回答しているが、これは慣れの問題も大きいと考える。実験結果では、複数分割でも正解率は下がらず、解答時間は短縮されることが示されていることから、提示方法を更に改善すれば、多数の動画から見たい動画を選択する場合などに役立つと考える。

5 おわりに

本稿では、見るべき動画の効率的発見を支援する指示的動画要約を生成するため、ニコニコ動画を対象として映像コンテンツとコメントの両方を考慮した重要区間抽出手法を提案した。また、抽出した区間を並列提示する手法も提案した。

提案する重要区間抽出手法では、画素値など映像

内の情報だけではなく、コメントなど映像以外の情報も同時に分析することで、動画作成側と視聴者側の両方の視点を考慮している。また、並列提示手法の利用可能性について予備実験を行った結果、4 分割程度であれば正解率を維持しつつ解答時間を短縮可能であることを示した。

今後は、提案する要約生成手法の評価実験を行う予定である。また、時間軸の追加などにより並列提示手法を改善することで、より見やすい提示手法を検討する予定である。

参考文献

- [1] 自主企画アンケート結果: 動画共有サイト (第 2 回), <http://www.myvoice.co.jp/biz/surveys/15713/index.html>, (2013 年 9 月 8 日現在)
- [2] 奥村学, 難波英嗣: テキスト自動要約 (知の科学), オーム社, p.6, 2005.
- [3] 三浦宏一, 浜田玲子, 井手一郎: 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. PRMU, パターン認識・メディア理解, Vol. 102, No. 155, pp. 15-20, 2002.
- [4] 永橋功丞, 宮岡伸一郎: 映像中の情報を用いたニュース映像要約手法の研究, 第 70 回全国大会講演論文集, pp. 2-357-2-358, 2008.
- [5] 岩元浩太, 山田昭雄: 多次元特徴空間解析に基づく映像のカット検出手法, 第 4 回情報科学技術フォーラム, Vol. 4, No. 3, pp. 65-66, 2005.
- [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] 多胡厚津史, 中川博之, 田原康之, 大須賀昭彦: ニコニコ探検くらぶ: ソーシャルアノテーションとキーワード群に基づく動画要約, 情報処理学会シンポジウム論文集, Vol.2010, No.4, pp.47-50, 2010.
- [8] 中村貴洋, 青木秀憲, 宮下芳明: マングの手法を用いたニコニコ動画ナビゲーション, 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, pp. 103-110, 2008.
- [9] 笠松沙紀, 伊藤貴之: 動画像データの要約可視化インタフェースの一手法, 第 1 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), E9-5, 2009.
- [10] 小関悠, 角康之: ばらばらマトリクス: 漫画技法を用いた映像を要約するシステム, 情報処理学会シンポジウム論文集, Vol. 2005, No. 4, pp. 177-178, 2005.