映像における話題シーン検出と話題性に基づく受動的映像視聴システム

王　元元† 渡部　雅俊† 河合由起子† 角谷　和俊††

† 京都産業大学コンピュータ理工学部　〒 803–8555 京都府京都市北区上賀茂本山
†† 兵庫県立大学環境人間学部　〒 670–0092 兵庫県姫路市新在家本町 1 丁目 1–12
E-mail: †{yuanw,kawai}@cc.kyoto-su.ac.jp, ††g1145497@cse.kyoto-su.ac.jp, †††sumiya@shse.u-hyogo.ac.jp

あらまし　VOD サービスの普及により，ユーザがさまざまな映像コンテンツを視聴することが可能であるが，映像の任意の時点で話題と関連する映像を閲覧したい場合，ユーザが関連映像を検索することが必要である．そこで，本研究では，映像の話題に対してシーンを検出し，映像投稿サイトや画像検索から各シーンの話題に関するコンテンツの追加および削除が可能な受動的映像視聴システム（TV-Binder）の構築を目的とする．本システムは，映像の話題に対して各シーンの話題性と時間制約に基づき，適切なコンテンツの追加ならびにシーンの削除を判定することで，4 種類の映像コンテンツを自動生成する．本論文では，映像からの話題発見に基づくシーンの検出手法を提案し，各シーンの話題性に基づき構築した受動的映像視聴システムを検証する．

キーワード　話題抽出，シーン検出，受動的映像視聴

1. はじめに

近年，ビデオ・オンデマンド（VOD）サービスの普及により，ユーザは観たい時にいつでも映像コンテンツを視聴できるようになった．しかしながら，映像視聴時に任意の時点で興味や疑問が生じた場合，それらに関連する話題のコンテンツ（Web ページや画像，Youtube 映像など）を検索しなければならない．例えば，ユーザがスイスの食べ物について知りたいと思った場合，スイスを話題にした映像を視聴すると考えられる．しかし，その映像のメイントピックがスイスの歴史という話題の場合，ユーザの興味に関連する映像の受動的視聴はできていないため，ユーザの興味を満たすことが困難である．

そこで，本研究では，ユーザの興味や疑問，つまり視聴映像の話題語を抽出する必要がある．また，一方で，ユーザに時間的な制約があり，限られた時間の中で映像の要点だけを知りたい場合もあるため，話題語に基づきシーンを検出し，検出されたシーンの話題性の高低度を判定することで，詳細度に基づいて話題性の高い（あるいは低い）コンテンツを追加し，話題性の高い（あるいは低い）シーンを削除し，多様性のある映像コンテンツを自動生成する手法を提案する．具体的には，まず，映像の字幕データを抽出し，映像全体の話題語を抽出する．この話題語の出現頻度を用いてシーンを検出し，検出されたシーンに対する話題性を判定する．次に，各シーンに出現する話題語の検索ヒット数に基づき，話題性の高低度を算出し，生成する映像コンテンツの種類に応じてランキングする．最後に，各シーンの話題性と映像全体の再生時間の長さに基づき，シーンの削除と新たなシーンの追加を決定する．なお，追加コンテンツは，本論文は投稿映像サイト Youtube(注1) から検索された映像や Google 画像検索(注2) より検索された画像，Google マップ検索(注3) より地名を含む地図を用いて，関連性の高い順と映像の再生時間に基づき選択する．シーンの削除と取得した関連するコンテンツを挿入することで，シーンの話題性と詳細度に合わせた 4 種類の映像コンテンツが生成される．

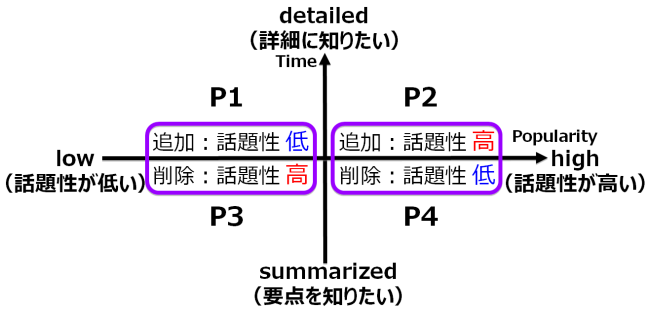

図 1　話題性と詳細度に基づく 4 種類の映像コンテンツ

本論文の構成は以下のとおりである．次章では提案システムの概要と関連研究について述べる．3 章では，映像における話題語の抽出ならびにシーンの検出手法を説明し，4 章では，提案手法に基づき生成される映像コンテンツの編集方法について述べ．5 章では，実装したシステムの有用性を図るための評価実験について述べた後，最後に，6 章でまとめと今後の課題について述べる．

2. システム概要と関連研究

2.1 受動的映像視聴システム（TV-Binder）

本研究では，字幕データを用いることで映像の話題語の抽出からシーンの検出を行い，また，検出したシーンの話題性の高低度を判定することで，映像の再生時間制約に基づき，投稿映像サイトや画像検索，マップ検索からシーンの話題性に関連する新たなコンテンツの追加，話題性の高い・低いオリジナル映

（注1）：https://www.youtube.com/
（注2）：https://www.google.co.jp/imghp?gws_rd=ssl
（注3）：https://www.google.co.jp/maps

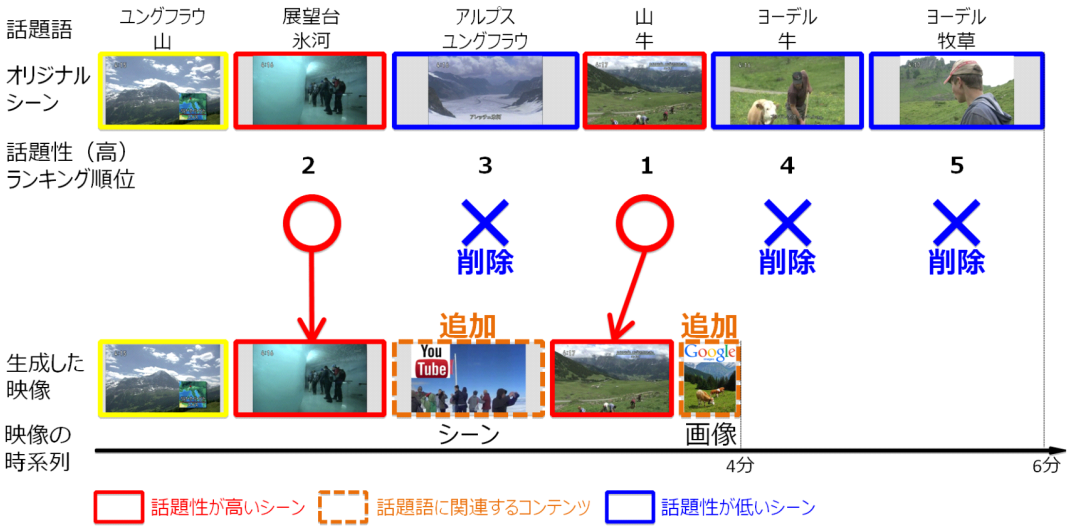

図 2　受動的視聴システム（TV-Binder）の概要図

像のシーンを削除することで，4 種類の映像コンテンツを生成する受動的視聴システム（TV-Binder）の構築を目指す．これにより，本提案システムでは図 1 のように 4 種類の映像コンテンツを生成する．

- P1 の場合：マニア向けの詳細映像
- P2 の場合：一般向けの詳細映像
- P3 の場合：マニア向けのダイジェスト映像
- P4 の場合：一般向けのダイジェスト映像

この 4 種類の映像コンテンツの中で，P1 と P3 に対しては，シーンの話題性が低い場合に新たなコンテンツを追加し，話題性の高いオリジナルシーンの削除を行う．逆に，P2 と P4 に対しては，シーンの話題性が高い場合に新たなコンテンツを追加し，話題性の低いオリジナルシーンの削除を行う．また，P1 と P2 は，映像の話題についてより詳細に知りたい場合とし，オリジナル映像より再生時間が長い映像コンテンツが生成され，逆に，P3 と P4 は，映像の話題の要点だけを知りたい場合とし，オリジナル映像より再生時間が短い映像コンテンツが生成される．追加する新たなコンテンツに関しては，本論文は投稿映像サイトや画像検索，マップ検索より話題語に関連する映像や画像，地図を検索し，関連性の高い順と映像の再生時間の長さに基づき選択し挿入する．

図 2 に TV-Binder で一般向けのダイジェスト映像が生成される P4 の場合の概要を示す．映像全体の字幕データより，話題語抽出に基づき話題性が高い，低いシーンをそれぞれ検出する．検出されたシーンの話題性の高い順によりランキングする．赤い枠は話題性が高いと判定されたシーンであり，破線枠はそのシーンに含まれる話題語に関連する別の映像コンテンツを投稿映像サイトより取得し，追加する．逆に，青い枠は話題性が低いと判定されたシーンであり，該当シーンをオリジナル映像より削除する．黄色の枠は映像の編集を行わないシーンとなる．

以上より，例えば，話題性が低く，詳細を知りたい場合（P1）は，マニア向けの長編の詳細映像が視聴でき，一方で，話題性が高く，要点を知りたい場合（P4）は，一般向けのダイジェスト映像が視聴できる．

2.2　関 連 研 究

これまでに，映像のシーン分割に関する研究は活発に行われている．映像の構造解析を用いたシーン分割手法では，ニュース映像ようにショットの明確な繰り返し構造を有する映像を対象とし，それらの階層クラスタリングに基づくグラフ分割問題を解くことでシーン分割を可能とする [8], [9], [12]．また，近年では，映像が画像，音響などの複数の要素から成り立っていることから，それらを統合して用いるマルチモーダル処理を導入することが一般的となっている [6], [1]．その中でも，帆足ら [13] は色配置，動き，そして音などの特徴量を算出し，シーンが変化するショット境界を SVM などの機械学習に基づいて検出することでジャンルに汎用なシーンを分割する手法が提案した．これらの研究では，映像コンテンツの内容より画像や音声などの映像メディアの特徴に着目し，シーンの検出を可能としている．本研究は，ユーザが映像視聴時の任意の時点で生じる話題に基づいて受動的映像を視聴することを目的とする．そのため，映像の字幕から映像の内容における話題を抽出し，話題に対応するシーンを検出する．

現存の映像要約手法には，主に 2 つのアプローチがあげられる．1 つは，システムが判定した重要なシーンを抽出し視聴者に提示する手法である [3]．この手法では重要なシーンをまとめて鑑賞できるため，満足感が得られるシーンを絞って鑑賞できる．Liu ら [5] は SVM (Support Vector Machine) による音声の学習を用いたラケットスポーツに対する映像要約手法を提案した．小林ら [10] は視聴者の関心を要約映像には反映させた手法として，SNS の 1 つ Twitter のつぶやきを利用し重要なシーンを推定する手法を提案した．我々も視聴者の関心である話題に関するシーンの追加および削除による受動的視聴システムの構築を目的とする．

一方で，映像そのものを要約するのではなく，高速再生により再生時間を短縮するという手法が注目されている．Kurihara [4] は字幕付き映画 DVD を対象として，字幕のない箇所は高速再

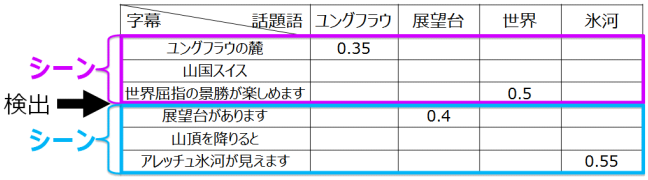

図 3　シーンの検出例

生し，字幕のある箇所については字幕を読むことが可能なように再生することで，内容を把握しつつ全体として通常よりも短時間による鑑賞を可能にする．Foulke ら [2] は，音高を変えない音声の高速再生が理解の上で有効であることを示した．Vemuri ら [7] は音声情報の高速再生時に，その音声の音声認識結果のテキスト情報を提示することによるユーザの情報処理速度の向上の試みを検討した．また，青木ら [11] は聴覚のみを用いた音楽検索インタフェースとして高速再生を導入している．本研究は音声や映像などの再生速度を自動調整することではなく，ユーザが映像視聴時に任意の時点で生じる興味や疑問に関する話題を発見し，映像の話題に対して各シーンの話題性と時間制約に基づき，適切なシーンの追加ならびに削除を判定することで，短時間でユーザが興味のある話題に関する内容を受動的に視聴することができる．

3.　映像における話題語抽出とシーン検出

本研究では，映像全体の話題語を抽出するため，実際の映像コンテンツにおける字幕データを用いて出現する単語および出現時間を抽出する．具体的には，映像コンテンツ全体の字幕データより，単語 $W_{1,\cdots,i}$ を抽出する．抽出した単語 $W_{1,\cdots,i}$ のうち，tf 値が閾値以上の単語を話題語 k として抽出する．tf 値は字幕に出現するそれぞれの単語の出現頻度である．本研究でのシーンとは，ある話題に関する内容のまとまりであると考える．そこで，映像コンテンツの時系列に沿って，抽出された話題語のうち，tf 値の総和が閾値以上の場合，シーンの切れ目として検出を行う．

図 3 に閾値=0.8 の場合のシーン検出例を示す．表の列は話題語であり，左から tf 値が高い順となっている．表の行は字幕の各セクションであり，上から時系列の順になっている．映像の時系列に沿って，話題語の tf 値の総和が 0.8 以上の時点でシーンの切れ目として検出を行う．括弧の部分に含まれる部分が検出されたシーンとなる．

4.　話題性に基づく映像コンテンツの生成

4.1　コンテンツの追加・削除判定

3 章で検出したシーンに対して，オリジナル映像に新たなコンテンツの追加，オリジナル映像からシーンの削除を判定する．まず，検出したシーンの話題語 k を検索クエリとして，投稿映像サイトより各シーンごとに映像の検索を行い，映像の検索ヒット数を取得する．次に，検索ヒット数から追加と削除の判定を行うための閾値を算出する．閾値 α は以下の式を用いて算出する．

$$\alpha = \frac{|Search(k)|}{N} \tag{1}$$

$|Search(k)|$ は映像全体の話題語により検索された結果数であり，N は提案手法より検出したシーンの総数である．

この閾値 α と各シーンの検索ヒット数 $|Search(k_i)|$ の比較により，以下の判定式を満たす新たなコンテンツの追加とオリジナルシーンの削除を決定する．

- P1，P3 の場合

if $|Search(k_i)| < \alpha$ then 追加，else if $|Search(k_i)| \geqq \alpha$ then 削除

- P2，P4 の場合

if $|Search(k_i)| \geqq \alpha$ then 追加，else if $|Search(k_i)| < \alpha$ then 削除

4.2　追加コンテンツの再生時間算出

前節の各シーンの検索ヒット数を用いて，話題性の高い (あるいは低い) シーンの割合により追加コンテンツの再生時間を算出する．削除するオリジナルシーンは，各シーンの検索ヒット数を用いて，シーンの話題性が高いランキング（P1，P3）・低いランキング（P2，P4）順位により削除を行う．追加する新たなコンテンツは，各シーンの検索ヒット数を用いて，シーンの話題性が低いランキング（P1，P3）・高いランキング（P2，P4）順位により追加を行う．以上の削除・追加判定条件に含まれないシーンは映像の編集を行わない．また，各シーンの検索ヒット数 $|Search(k_i)|$ を用いて，それぞれの追加コンテンツの再生時間 t_i は下記の式により算出する．

- P1，P3 の場合：$|Search(k_i)| < \alpha$

$$t_i = \frac{|Search(k_i)|}{N_l} \times T \tag{2}$$

N_l は提案手法の式 (1) より全ての話題性が低いシーンの検索ヒット総数であり，T は追加コンテンツの総再生時間である．

- P2，P4 の場合：$|Search(k_i)| \geqq \alpha$

$$t_i = \frac{|Search(k_i)|}{N_h} \times T \tag{3}$$

N_h は提案手法の式 (1) より全ての話題性が高いシーンの検索ヒット総数であり，T は追加コンテンツの総再生時間である．

例えば，検出されたシーン数が 10 で，5 分 30 秒のオリジナル映像に対して，再生時間は 11 分の詳細映像（P2）を生成する場合とする．まず，オリジナル映像（5 分 30 秒）より，倍になる詳細映像（11 分）のうち，5 分 30 秒の半分となる 2 分 45 秒のオリジナルシーンを削除する．そこで，話題性の低い順により 5 つのオリジナルシーンを削除することと仮定する．次に，増加する 5 分 30 秒と削除されるオリジナルシーンの 2 分 45 秒に合わせて 8 分 15 秒（495 秒）の新たなコンテンツを追加する．そこで，オリジナル映像における残った 5 つのシーンから話題性が高いと判定される 3 つのシーン（A，B，C）に関連する新たなコンテンツの追加を行うことと仮定する場合，話題性が高い順でシーン A の検索ヒット数が 50，シーン B の検索ヒット数が 30，シーン C の検索ヒットが 20 である場合，式 (3) よりシーン A に関連する追加コンテンツの再生時間は (50/(50+30+20))×495 秒 =246 秒，シーン B に関連する追加コンテンツの再生時間は (30/(50+30+20))×495 秒 =149 秒，シーン C に関連する追加コンテンツの再生時間は (20/(50+30+20))×495

秒 =99 秒となる．

なお，本論文は生成コンテンツの時間制約に基づき追加コンテンツの再生時間は閾値 β を用いて，以下の判定式を満たす追加コンテンツの種類を判定する．

- if $t_i \geqq \beta$ then 投稿映像サイト Youtube からの映像や Google マップ検索からの地図（話題語が地名である場合）を追加
- if $t_i < \beta$ then Google 画像検索からの画像や Google マップ検索からの地図（話題語が地名である場合）を追加

また，生成コンテンツの話題性の高低度に基づき追加コンテンツを選択する．

- P1，P3 の場合

投稿映像サイト Youtube からの話題性が低いシーンに含まれる話題語との関連性の高い順に基づき映像，Google 画像検索からの話題性が低いシーンに含まれる話題語との関連性の高い順に基づき画像，Google マップ検索からの話題性が低いシーンに含まれる地名の詳細地図を選択する．

- P2，P4 の場合

投稿映像サイト Youtube からの話題性が高いシーンに含まれる話題語との関連性の高い順に基づき映像，Google 画像検索からの話題性が高いシーンに含まれる話題語との関連性の高い順に基づき画像，Google マップ検索からの話題性が高いシーンに含まれる地名の広範囲地図を選択する．

5. 評 価 実 験

本章では，構築した受動的映像視聴システムに関する評価実験を行った．被験者は 20 代の大学生 10 名，「シリーズ世界遺産 100(注4)」の以下の 3 本の映像から，TV-Binder による P1～P4 の 4 種類の映像コンテンツを生成した．実験 I：追加コンテンツが映像のみを用いて生成した場合を表 1；実験 II：追加コンテンツが映像や画像，地図を用いて生成した場合を表 2 に示す．

- オリジナル映像 O1：スイスアルプス　ユングフラウとアレッチュ～スイス～（再生時間：5 分 31 秒）
- オリジナル映像 O2：ヴァッハウ渓谷の文化的景観～オーストリア～（再生時間：5 分 30 秒）
- オリジナル映像 O3：イエローストン国立公園～アメリカ～（再生時間：5 分 30 秒）

提案手法により n=3 で検出されたシーン数は O1 が 9 個，O2 が 11 個，O3 が 13 個で，シーンの平均再生時間は約 30 秒であった．追加コンテンツの再生時間の閾値 β=10 で追加コンテンツの種類を判定した．また，詳細映像（P1，P2）の平均再生時間は 8 分 22 秒で，ダイジェスト映像（P3，P4）の平均再生時間は 2 分 32 秒であった．

実験 I と II で生成した P1～P4 の 4 種類の映像コンテンツに対して下記の設問項目についてアンケートを実施した．

- $Q1$：映像の内容が理解できた
- $Q2$：シーンの切り替わりの違和感がなかった
- $Q3$：映像の内容に関係ないシーンがなかった
- $Q4$：P1，P2 の場合，映像は長いと感じた

（注4）：http://www.nhk.or.jp/sekaiisan/s100/

表 1　実験 I：生成した P1～P4 の 4 種類の映像コンテンツ

	再生時間	追加シーン長（数）	削除シーン長（数）
O1	5 分 31 秒	–	–
P1 映像 O1	8 分 21 秒	5 分 5 秒（3）	2 分 24 秒（4）
P2 映像 O1	7 分 1 秒	4 分 45 秒（2）	3 分 28 秒（4）
P3 映像 O1	3 分 30 秒	1 分 7 秒（3）	3 分 10 秒（6）
P4 映像 O1	2 分 16 秒	1 分（2）	4 分 18 秒（6）
O2	5 分 30 秒	–	–
P1 映像 O2	7 分 26 秒	5 分 17 秒（5）	3 分 25 秒（6）
P2 映像 O2	6 分 42 秒	4 分 51 秒（3）	3 分 24 秒（6）
P3 映像 O2	2 分 36 秒	1 分 40 秒（3）	4 分 36 秒（8）
P4 映像 O2	2 分 21 秒	1 分（1）	4 分 18 秒（8）
O3	5 分 30 秒	–	–
P1 映像 O3	10 分 43 秒	8 分 27 秒（6）	3 分 14 秒（7）
P2 映像 O3	9 分 53 秒	8 分 11 秒（3）	3 分 49 秒（7）
P3 映像 O3	2 分 39 秒	1 分 34 秒（3）	4 分 27 秒（10）
P4 映像 O3	2 分 25 秒	47 秒（3）	4 分 47 秒（10）

表 2　実験 II：生成した P1～P4 の 4 種類の映像コンテンツ

	再生時間	追加シーン長（数）	追加画像長（数）	追加地図長（数）	削除シーン長（数）
O1	5 分 31 秒	–	–	–	–
P1 映像 O1	8 分 21 秒	5 分 5 秒（3）	4 秒（2）	–	2 分 24 秒（4）
P2 映像 O1	7 分 1 秒	4 分 45 秒（2）	–	–	3 分 28 秒（4）
P3 映像 O1	3 分 38 秒	1 分 7 秒（3）	8 秒（2）	–	3 分 10 秒（6）
P4 映像 O1	2 分 24 秒	1 分（2）	8 秒（2）	–	4 分 18 秒（6）
O2	5 分 30 秒	–	–	–	–
P1 映像 O2	7 分 41 秒	5 分 17 秒（5）	16 秒（4）	–	3 分 25 秒（6）
P2 映像 O2	5 分 37 秒	4 分 51 秒（3）	8 秒（2）	–	3 分 24 秒（6）
P3 映像 O2	2 分 19 秒	1 分 40 秒（3）	8 秒（2）	4 秒（1）	4 分 36 秒（8）
P4 映像 O2	3 分 10 秒	1 分（1）	12 秒（2）	–	4 分 18 秒（8）
O3	5 分 30 秒	–	–	–	–
P1 映像 O3	11 分 7 秒	8 分 27 秒（6）	12 秒（3）	12 秒（3）	3 分 14 秒（7）
P2 映像 O3	10 分 1 秒	8 分 11 秒（3）	4 秒（1）	4 秒（1）	3 分 49 秒（7）
P3 映像 O3	1 分 19 秒	1 分 34 秒（3）	12 秒（3）	4 秒（1）	4 分 27 秒（10）
P4 映像 O3	1 分 42 秒	47 秒（3）	12 秒（3）	–	4 分 47 秒（10）

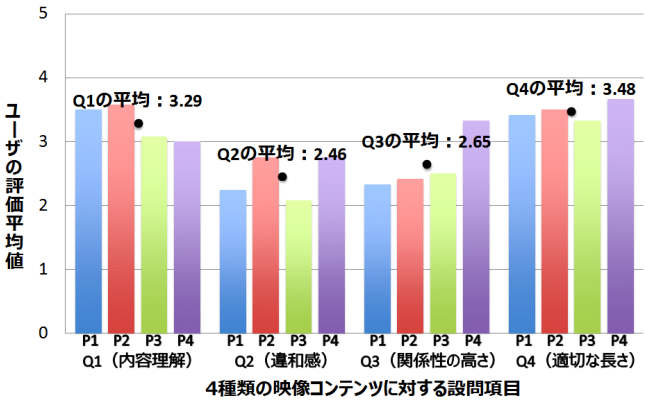

図 4　実験 I：各設問項目の評価結果

P3，P4 の場合，映像は短いと感じた

- $Q5$：興味があるトピックの記述とどの興味程度か
- $Q6$：映像の内容に関係ないトピックの列挙

実験 I の 5 段階評価による $Q1$～$Q4$ の評価平均値を図 4 に示す．

- $Q1$ は詳細映像（P1，P2）の評価が高く，映像の内容を理解するために必要とされるシーンが削除されてしまうため，ダイジェスト映像（P3，P4）の評価が低くなった．
- $Q2$ は生成した映像の全体が低い結果となった．これは，実験で使用したオリジナル映像はナレーションが含まれているのに対して，追加した映像はナレーションがないため，シーンの切り替わりの違和感があったと考えられる．特にマニア向け

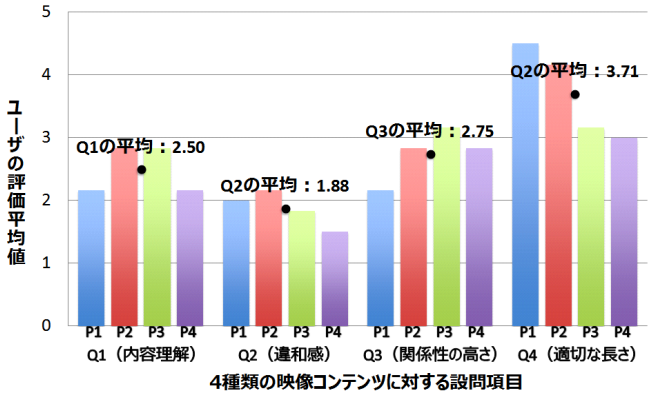

図 5　実験 II：各設問項目の評価結果

の映像（P1，P3）の評価が低くなった．

- *Q*3 はマニア向けのダイジェスト映像 P3 のみの評価が高く，映像に関連ないシーンを追加されてしまうため，特に詳細映像（P1，P2）の評価が低くなった．
- *Q*4 は詳細映像（P1，P2）は長いと感じた評価に偏った．ダイジェスト映像 P3 は平均値より高い評価となり，ダイジェスト映像は短いと感じた評価が多くなった．これにより，生成した映像コンテンツの再生時間の長さは妥当な結果であることが確認できた．
- *Q*5 に対して，被験者が興味を持ったトピックは，約 8 割は P1～P4 の 4 種類の映像コンテンツに追加と判断されたオリジナルシーンから抽出した話題語と同じ語であった．つまり，話題語を用いた受動的視聴が被験者の興味を惹くような結果となった．また，被験者が P1～P4 の 4 種類の映像コンテンツを視聴する前に興味を持ってなかったトピックに対して，興味喚起を図ることができた．
- *Q*6 に対して，ダイジェスト映像（P3，P4）は映像に関連のないトピックが 18 件，詳細映像（P1，P2）に対しては，映像と関連のないトピックが 24 件という結果となった．被験者が記述した関連のないトピックから，話題語には関連しているが適切でない映像が追加されてしまうという傾向があった．

以上より，*Q*1 と *Q*4 は全体的に高い評価を得られた．しかしながら，*Q*2 は全体的に低い評価となった．また，*Q*3 は特に P1，P2，P3 の評価が低くなった．

実験 II の 5 段階評価による *Q*1～*Q*4 の評価平均値を図 5 に示す．実験 I（追加コンテンツ：映像のみ）の結果と実験 II（追加コンテンツ：映像，画像，地図）の結果の比較を行う．

- *Q*1 は特に P1，P4 の評価が低く，実験 I と同じく，映像の内容理解に必要なシーンが削除されていると考えられる．
- *Q*2 は全体が低い結果となった．また，実験 I の結果と比較すると実験 II の評価の方が低くなった．これは画像や地図のような静止画が増えたことにより，映像を視聴する際に違和感を強く感じてしまったと考えられる．
- *Q*3 は特にマニア向けの詳細映像 P1 の評価が低くなった．しかし，実験 I の結果と比較した場合，実験 II の評価の方がよくなった．これは追加するコンテンツの種類が増えたことにより，映像に関連するコンテンツがより詳細になったと考えられる．
- *Q*4 は詳細映像（P1，P2）は長いと感じた評価に偏った．ダイジェスト映像 P3 は平均値より高い評価となり，ダイジェスト映像は短いと感じた評価が多くなった．実験 I と同じく，生成した映像コンテンツの再生時間の長さは妥当な結果であることが確認できた．
- *Q*5 は実験 I と同じく，被験者が興味を持ったトピックがオリジナル映像から抽出した話題語と同じ語が多かったため，興味喚起を図ることができた．
- *Q*6 に対して，ダイジェスト映像（P3，P4）は映像に関連のないトピックが 8 件，詳細映像（P1，P2）に対しては，映像と関連のないトピックが 12 件という結果となった．P4 以外，実験 I の結果と比較すると，追加するコンテンツが増えたことによる映像と関連のないトピックが少なくなった．

以上より，2 つの実験を比較した結果，*Q*1，*Q*2 は，実験 I の評価が良い結果を得られた．しかし，*Q*3，*Q*6 は，実験 II の評価が良い結果を得られた．*Q*4，*Q*5 は，どちらも結果に明確な差が見られないという結果となった．今後の課題として，短い映像を生成するため，シーンの削除の妥当性と追加コンテンツの種類が増えた場合の切り替わりの違和感についての検討が必要である．また，追加するコンテンツとオリジナルコンテンツとの関連性分析の検討を行う必要がある．

6. おわりに

本研究では，映像の字幕データを用いることで，映像の話題発見に基づきシーンを検出し，検出したシーンの話題性に基づく受動的映像視聴システム（TV-Binder）を構築した．TV-Binder でオリジナルシーンの削除と新たなコンテンツの追加により生成した 4 種類の映像コンテンツを用いて評価実験を行った．実験結果により有効な結果であることを確認した．

今後，ユーザが映像に対するインタラクション機能を付与し，新たに挿入するコンテンツのタイプ（Web ページや音声など）選別とその提示方式を検討する．また，音声認識による音声情報と字幕の両方から話題語の抽出を検討する．

謝　辞

本研究の一部は，総務省戦略的情報通信研究開発推進事業（SCOPE）および JSPS 科研費 26280042 の助成を受けたものである．ここに記して謝意を表す．

文　献

[1] M. Del Fabro and L. Böszörmenyi. State-of-the-art and future challenges in video scene detection: a survey. *Multimedia systems*, 19(5):427–454, 2013.

[2] E. Foulke and T. G. Sticht. Review of research on the intelligibility and comprehension of accelerated speech1. *Speech and Hearing Science: Selected Readings*, 72(1):78, 1974.

[3] S. Kawamura, T. Fukusato, T. Hirai, and S. Morishima. Efficient video viewing system for racquet sports with automatic summarization focusing on rally scenes. In *ACM SIGGRAPH 2014*, page 62. ACM, 2014.

[4] K. Kurihara. Cinemagazer: a system for watching videos at very

high speed. In *Proc. of the International Working Conference on Advanced Visual Interfaces (AVI2012)*, pages 108–115. ACM, 2012.

[5] C. Liu, Q. Huang, S. Jiang, L. Xing, Q. Ye, and W. Gao. A framework for flexible summarization of racquet sports video using multiple modalities. *Computer Vision and Image Understanding*, 113(3):415–424, 2009.

[6] P. Sidiropoulos, V. Mezaris, I. Kompatsiaris, H. Meinedo, M. Bugalho, and I. Trancoso. On the use of audio events for improving video scene segmentation. In *Analysis, Retrieval and Delivery of Multimedia Content*, pages 3–19. Springer, 2013.

[7] S. Vemuri, P. DeCamp, W. Bender, and C. Schmandt. Improving speech playback using time-compression and speech recognition. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI2004)*, pages 295–302. ACM, 2004.

[8] M. Yeung, B. Yeo, and B. Liu. Segmentation of video by clustering and graph analysis. *Computer Vision and Image Understanding*, 71(1):94–109, 1998.

[9] 吉田壮, 小川貴弘, and 長谷山美紀. 歌謡番組における映像の構造に注目したシーン分割手法. **電子情報通信学会論文誌** *D*, 97(7):1177–1188, 2014.

[10] 小林尊志, 野田雅文, 出口大輔, 高橋友和, 井手一郎, and 村瀬洋. Twitter の実況書き込みを利用したスポーツ映像の要約. **電子情報通信学会技術研究報告**. *MVE*, **マルチメディア・仮想環境基礎**, 110(457):165–169, 2011.

[11] 青木秀憲 and 宮下芳明. 視覚を用いない状況下での高速楽曲探索インタフェースの設計と検証. **情報処理学会論文誌**, 51(2):356–364, 2010.

[12] 宋妍, 小川貴弘, and 長谷山美紀. 映像の構造に注目した mcmc 法に基づくシーン分割法. **電子情報通信学会論文誌** *D*, 97(3):560–573, 2014.

[13] 帆足啓一郎, 菅野勝, 内藤正樹, 松本一則, and 菅谷史昭. 汎用的特徴量に基づく動画像話題分割手法. **電子情報通信学会論文誌** *D*, 89(10):2305–2314, 2006.