

# ラケットスポーツ動画の構造解析による映像要約手法の提案

河村俊哉<sup>†1</sup> 福里司<sup>†2</sup> 平井辰典<sup>†1</sup> 森島繁生<sup>†3</sup>

近年、スポーツ動画を手軽に鑑賞できるようになり、効率的な鑑賞方法が必要とされている。その解決策として、従来のラケットスポーツ動画に対する映像要約では、重要なラリーシーンの要約映像を生成したが、ラリーシーンの検出及びその評価方法に問題点が見られ、要約映像の効率的な鑑賞方法についても考慮がなかった。そこで本稿では、ラケットスポーツ動画に対する新たなラリーシーンの検出方法及び各ラリーの重要度をを用いた映像要約手法及びその鑑賞方法を提案する。提案手法では、ショット分割された動画に対し類似ショットのクラスタリング及びラリーを含むクラスタの選定により、精度の高いラリーシーンの検出方法を実現する。その後、各ラリーに対して音響情報を考慮した重要度評価を行い、その結果をユーザが調整することで、任意の時間内での動画の鑑賞を可能とする。さらに、ラケットスポーツに特化した高速再生による動画視聴方法を提案し、さらなる効率的な動画の鑑賞方法を実現する。

## 1. はじめに

動画鑑賞サイトの浸透やスマートフォンの普及により、動画を手軽に鑑賞できる時代に到達した。一方で、動画鑑賞に費やすことができる時間に対し、鑑賞したい動画の総時間は増大し続けている。特にスポーツ動画は、シーズン期間中に1試合あたり数時間にもなる動画が世界中で配信されている。しかし、その内容を把握するためには長時間にわたる試合全てを鑑賞する必要があり、日常生活でまとまった時間が取れない視聴者が試合の流れを含めてスポーツ動画を楽しむことは困難である。

動画を鑑賞する際、特定シーンを検索する一般的な手法として早送りやシークバーの利用が挙げられる。これらの手法は、視聴者の関心が高いシーンやタイミングが予想できる場合に非常に効果的である。しかし、スポーツ動画において、いつ何が起きているのかは予測できないため、未視聴の動画に適用すると試合の過程やその流れを把握できないまま勝敗など最終的な結果に辿り着いてしまい、通常鑑賞したときのような満足感が得られない。以上の理由から、こうした手法では既に鑑賞したことのある動画やCMなどの区間を経験から推測できる箇所のスキップにしか適用できない。

このような背景から、視聴者の興味や関心を保持したまま動画の効率的な鑑賞を促す映像要約に関する研究が盛んに行われている。スポーツ動画を対象とした既存の研究では、ニュース番組のダイジェスト映像のように「得点シーン=重要な場面」であることを前提としているが、劣勢の選手が逆転した得点シーンなど単なる得点シーンと比べ、視聴者の関心が高まるシーンの重要度を捉えるまでには至っていない。特にラケットスポーツの得点シーンは、単純なミスからファインプレーによる得点まで幅広く、視聴者の関心を適切に捉えられなければ満足感が得られない。また、トーナメント形式で行われることの多いラケットスポ

ーツでは、複数の試合を同時進行することが多いため、他のスポーツと比べて配信される試合数も多く試合時間も非常に長い。そのような観点から、ラケットスポーツ動画を対象とした映像要約手法は需要が高いと言える。

そこで本稿では、視聴者が着目するシーンに基づくラケットスポーツ動画の映像要約手法とその鑑賞方法について提案する。前者については、ラケットスポーツを鑑賞する際、最も着目されるラリーシーンを抽出しその重要度評価方法を提案することで、視聴者の満足感が得られる重要度の高いシーンだけを抽出した要約映像の生成を実現する。後者については、試合内容が理解できる程度の高速再生を利用した鑑賞方法について提案することでさらなる効率的な動画鑑賞を可能とする。それによって、通常速度で動画全体を鑑賞する場合と比べ、動画内容の理解と満足度を保持したまま高速で動画を鑑賞することを可能にする。

## 2. 関連研究

現存の映像要約手法には、主に二つのアプローチが挙げられる。ひとつは、システムが判断した重要なシーンだけを抽出し視聴者に提示する手法である。この手法では人手による要約映像と同様、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。そのため、これまでもスポーツ動画を対象とした映像要約に適用されてきた。Liuらは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[1]。この手法では、ラケットスポーツ動画においてラリーシーンが重要であると仮定しており、ラリーシーンを高精度で抽出している。さらに、各ラリーにそれぞれ重要度を付与し、重要度の高いシーンのみを集めた映像要約手法を提案している。しかし、学習に用いた動画に似た音声環境をもつ動画にしか適用できないため、打球音の異なるスポーツや、学習データには含まれない種類の音を含む場合には学習し直さなければならないという問題点があった。また、視聴者は入力動画に適用できる学習データを作成する際、30分程度の音声情報を人手によりラベリングする必要があり、時間的コストがかかるという問題があった。さらに、各ラリーの重要度を評価しているが、

<sup>†1</sup> 早稲田大学

Waseda University

<sup>†2</sup> 早稲田大学/JST

Waseda University/JST

<sup>†3</sup> 早稲田大学理工学術院総合研究所/JST

Waseda Research Institute for Science and Engineering/JST

その評価結果については重要度が高いと判断されたシーンがセットポイントやマッチポイントとなる傾向があると述べるだけに留まっており、具体的な考察がないため妥当性がない。

小林らは視聴者の関心を要約映像に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[2]。つぶやきの内容解析により、応援するチームの同じ人が反応したシーンだけを要約した映像を生成することができるが、リアルタイムに鑑賞している人が多く必要となるなどの制約条件を伴う。両角らは野球中継の映像を対象として、投球シーンや得点時に出現するテロップを利用した映像要約手法を提案した[3]。これはスポーツ動画の要約において「得点シーン=重要なシーン」であるという仮定を用いた研究事例である。しかし、得点シーンの抽出までに留まっており、各得点の試合中での重要度は一切評価されていないため、野球などの得点シーンが比較的少なく重要であるスポーツ以外では要約のメリットが少ない。

一方、動画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。この手法は、動画内容を認識できる限界の速さで高速再生することで、動画全てを鑑賞しながら再生時間を短縮する手法である。この手法では動画全体の情報を汲み取ることができるため、内容を高速で一通り把握するのに効果的である。kuriharaは字幕のあるシーンを字幕が読める程度の速さで再生し、字幕のないシーンを動画内容が把握できる程度で高速再生する手法を提案した[4]。この手法は会話内容を重視するような動画に効果的な手法である。しかし、動画情報とは別に字幕情報が必要なため、スポーツ動画に挙げられるような字幕のない動画に適用することは困難である。

ラケットスポーツに関連した映像要約の研究として、イベント検出に取り組んでいる研究が挙げられる[5][6]。Tienらはテニス動画からサービスやネットプレイといったイベントの検出を行っている[5]。しかし、ラリーシーン検出の精度評価がなされていないため、動画からラリーシーンを高精度で抽出する手法が必要である。また、検出イベントの重要度評価は全く考慮されておらず、視聴者の関心を反映することができない。さらに、検出するイベント内容はスポーツの種類に依存し、それぞれのスポーツに最適な検出手法を設定する必要があるため、煩雑な評価を行うことになり映像要約には適していない。

以上を踏まえ、ラケットスポーツ動画の映像要約において頑健な特徴量に基づいた精度の高いラリーシーン検出手法が必要である。また、要約動画の鑑賞方法として、適切な評価に基づく重要なシーンだけの鑑賞とラケットスポーツに適応した高速再生を利用した鑑賞時間の短縮の二つの手法を併用することで、視聴者は重要なシーンを効率的に

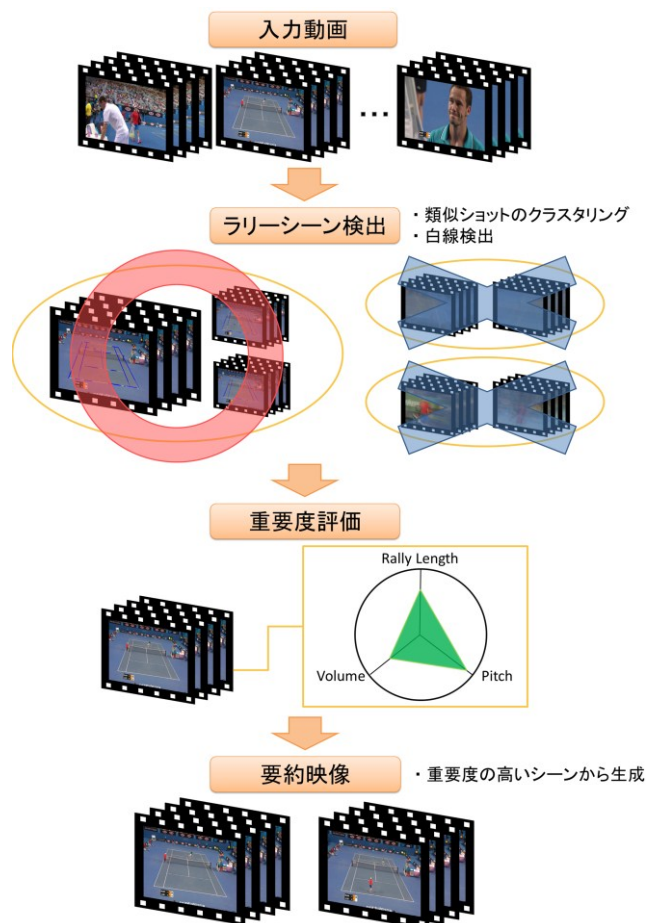


図1 要約映像生成までの流れ

鑑賞できると考えられる。そこで本稿では、ラケットスポーツ動画全般に用いることができるラリー検出とその重要度評価方法及びその鑑賞方法について検討する。

### 3. ラリーシーン検出と重要度評価

本章では、学習なしでのラリーシーン検出及びその重要度評価方法について提案する。初めに、ラケットスポーツにおいて、試合の流れを把握するために有効で視聴者の多くが鑑賞したいラリーシーンの判定方法を提案する。その後、検出したラリーシーンに対して重要度評価を行う。提案手法による要約映像の生成までの流れを図1に示す。

#### 3.1 ラリーシーン検出

事前に手動でショット分割した入力動画に対して、色特徴量に基づいて図2のように類似ショットをクラスタリングする。その後、ラリーシーンのクラスタを識別するために、ラリーシーンが固定カメラで撮影されることを利用し、ショットの平均画像に残る白線の検出を利用する。ラリーシーンのクラスタ判別の流れを図3に示す。

##### (1) 類似ショットのクラスタリング

ここで用いるクラスタリング手法には、Liuらのショットのキーフレームの類似度に基づく手法を用いる[1]。初めに、 $N$ 個のショットに分割された入力映像から各ショット

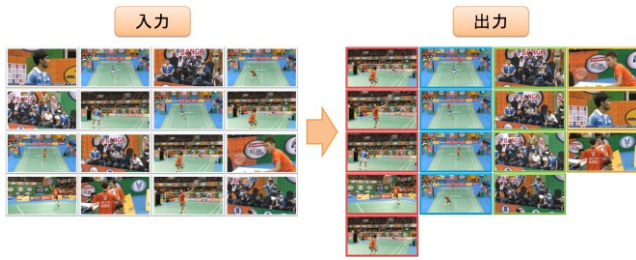


図2 類似ショットのクラスタリング例

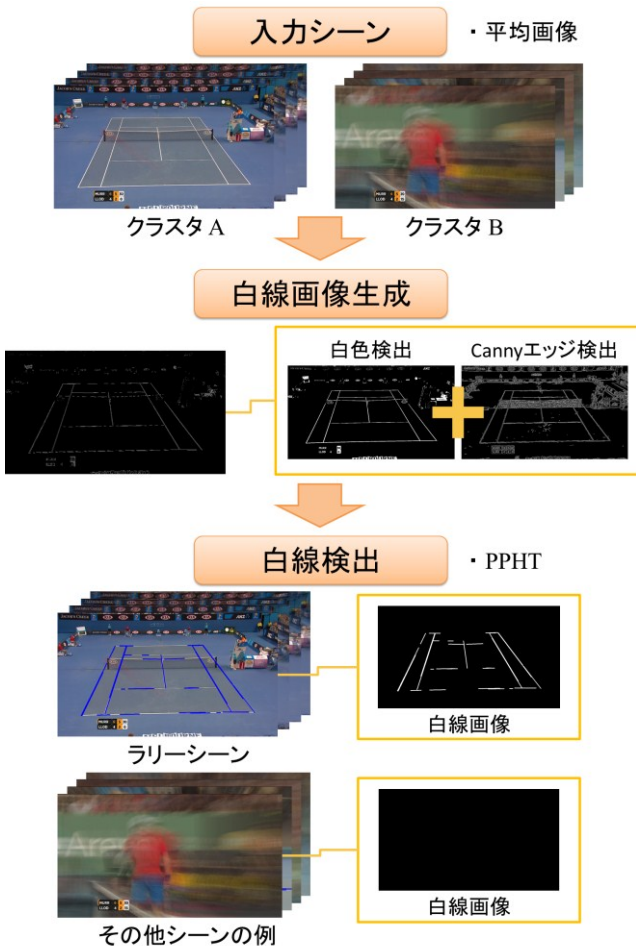


図3 ラリーシーンクラスタ判別の流れ

のキーフレームを抽出する.  $i$  番目のショットから  $m$  枚のキーフレーム  $k_m^i$  を等間隔に抽出する (Liu らと同様に  $m=5$  とした). 次に, 取得したキーフレームに対して HSV 特徴量による 256 次元ヒストグラムを作成する ( $H \times S \times V = 16 \times 4 \times 4$ ). その後,  $i$  番目と  $j$  番目のショットの類似度  $SD(s_i, s_j)$  を以下の式(5)により計算する.

$$SD(s_i, s_j) = \frac{1}{2} [Min\{d(k_m^i, k_n^j)\} + Min\{d(k_m^i, k_n^j)\}] \quad (1)$$

$$d(k_m^i, k_n^j) = \frac{1}{256} \sum_{b=1}^{256} |H_m^i(b) - H_n^j(b)| \quad (2)$$

ここで,  $m, n$  は各ショットのキーフレーム番号であり,

$d(k_m^i, k_n^j)$  はキーフレーム同士の距離である. そして, 式(1)の値が最も小さくなった 2 つのショットを類似ショットとしクラスタリングする ( $s_i = s_i + s_j$ ). 新たに生成したクラスタのもつキーフレームのヒストグラム  $H_m^l$  を更新するため, 以下の式を用いて再構築する.

$$H_m^l = \frac{N_i * H_m^i + N_j * H_m^j}{N_i + N_j} \quad (3)$$

ここで,  $N_i, N_j$  は統合するショット  $s_i, s_j$  がもつフレーム数である. 以上の手順でクラスタリングを行うが, 類似ショットだけのクラスタとするためにクラスタリングを終了する必要がある. その評価式は, クラスタ間・クラスタ内分散をもとにした  $J_l$  とクラスタ数をもとにした  $k_l$  を用いて, 以下の式(4)を最小とする時と定める.

$$E_l = J_l + k_l \quad (4)$$

$$J_l = \frac{\sum_{c=0}^{K_l} J_w^c}{J_l} = \frac{\sum_{c=0}^{K_l} \sum_{i=0}^{N_c} \| \vec{s}_i^c - \vec{s}_{mean}^c \|^2}{\sum_{i=0}^N \| \vec{s}_i - \vec{s}_{mean} \|^2} \quad (5)$$

$$k_l = \frac{K_l}{N} \quad (6)$$

ここで,  $J_l$  はフィッシャーの識別関数をもとに定義した.  $J_l$  はクラスタリングを行う前の各ショット間の分散,  $J_w^c$  はクラスタ  $c$  に含まれるショットの分散,  $N$  はクラスタリング前の総ショット数,  $N_c$  はクラスタ  $c$  に含まれるショット数,  $\vec{s}_i^c, \vec{s}_{mean}^c$  はそれぞれクラスタ  $c$  内 3 番目のキーフレームのヒストグラム (クラスタの中心フレームが最もクラスタの特徴を表すと仮定した) とクラスタに含まれるショットのもつヒストグラムの平均値,  $\vec{s}_i, \vec{s}_{mean}$  はそれぞれクラスタリングを行う前の各ショットのヒストグラムとその平均値である. なお, この計算を行う際, 初期状態ではクラスタ内分散は全てゼロとする. また, 式(6)において  $K_l$  はクラスタリングの過程で存在するクラスタの数であり,  $N$  はクラスタリング前の総ショット数を表している. クラスタリングが進むと  $J_l$  は大きくなるが,  $k_l$  はクラスタ数が少なくなるため減少する. このことを利用し, 式(4)のようなクラスタリングの終了箇所を定める (図4参照). このようにして生成されたクラスタは, ショット内に含まれる色味の似たシーンにより構成される. ラケットスポーツでは, 図2のように選手のアップやラリーシーン, 観客席などのクラスタに分類される.

## (2) 平均画像の生成

ラリーシーン検出の際, Liu らの SVM による手法[1]では音声情報の学習を用いたが, 音声情報は動画ごとに異なり

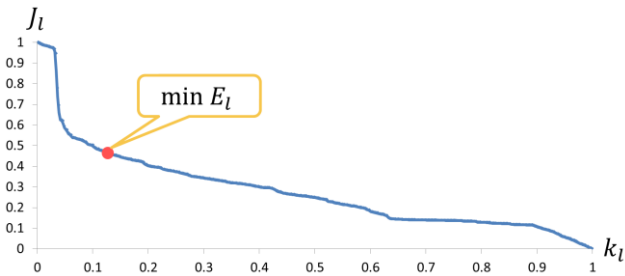


図4 クラスタリングの停止点

学習した音声と似た音声をもつ動画でなければならない、違うラケットスポーツでは打球音が異なるため適用できない、といった問題があった。そこで提案手法では、ラケットスポーツ動画においてロバストな情報である画像特徴量を用いる。具体的には、ラリーシーンは多くの場合固定カメラかつ被写体の動く領域が制限されており、その他のシーンはカメラあるいは被写体が大きく動く、というラケットスポーツ動画における一般的な特徴を利用する。初めに  $i$  番目のショットに対して以下の式により平均画像を生成する。

$$f_{mean}(x, y) = \frac{1}{N} \sum_{i=1}^N f_i(x, y) \quad (7)$$

ここで、 $f_i(x, y)$  は  $i$  番目のフレーム画像、 $N$  は平均画像を生成するために用いたフレーム数である。本稿では、フェードなどカット点周辺の編集効果による影響を取り除くため、平均画像を  $[N/5, 4N/5]$  の範囲で生成した。ラリーシーンは固定カメラで撮影されるため、平均画像はコートの白線が残った状態となる。一方、その他のシーンはカメラや被写体の動きが大きいため、ぼやけた平均画像が生成される。

### (3) 白線検出

平均画像中の白線検出を行い、ラリーシーンとその他のシーンを判別する。白線検出では、平均画像から以下の二つの二値画像を統合したものをを用いる。

一つ目は、白色部分とそれ以外の部分とを二値化することで生成した画像を用いる。イベント検出やボール追跡に関しての先行研究で、ラリーシーンからコートの線を検出するために白色を検出する手法を参考とした [7][8]。白色検出における閾値  $I$  はコートの輝度値によって変化する。本稿では、屋外スポーツであるテニス動画では  $I=150$ 、屋内スポーツであるバドミントン動画では  $I=200$  とした。白色検出のみでは、白線以外にも白い物体がシーン中に映りこんでいた場合に誤検出する可能性がある。そこで、Canny 法を用いて平均画像からエッジ検出を行った。固定カメラ時の平均画像では線が残った二値画像となるが、カメラや被写体が動くショットの平均画像はぼやけているためにエッジが検出されない。

二つの二値画像を重ね合わせて作成した白線画像に対し PPHT(Progressive Perception Hough Transformation)を用い、

白線の有無を判断する処理をクラスタごとに行った。ただし、ラケットスポーツ動画は全ショット数に対するラリーシーンの割合が大きいため、ショット数が全ショット数の 1%未満のクラスタは除外した。そして、クラスタ内に含まれるショット数に対して、白線を含むショットが多い場合にラリーシーンのクラスタとした。以上の手法により、学習をせずに精度が高く様々なラケットスポーツ動画に適用できるラリーシーン検出を実現した。本手法の検出精度については 3.3 節(1)で述べる。

### 3.2 各ラリーシーンの重要度評価

本節では、ラリーシーンの重要度を定量化する手法を検討する。複数のラケットスポーツ動画を検証した結果、ラリーシーンの直後にラリーに対する反応が声援や発話として現れる傾向があった。この事象に着目し、重要度評価の方法としてラリーシーンの後半部分から次のショットの先頭から 5 秒後までの区間で、5 秒間の平均音量とピッチがそれぞれ最大となる箇所を検出し盛り上がり进行评估する。平均音量は次式により定義する。

$$V_{mean}(s) = \frac{1}{5F} \sum_{\tau=0}^{5F} |V(t+\tau)| \quad (8)$$

ここで、 $s$  はラリーシーン番号、 $F$  は音声のサンプリング周波数、時刻  $t$  は音声の開始位置、 $V(t)$  は  $t$  における音量 (振幅) を表す。また、音声のピッチはスペクトラムの重心を利用し、式(9)により定める。

$$P_{mean}(s) = \frac{1}{5} \sum_{t=1}^5 \omega_c(t+\tau) \quad (9)$$

ここで、 $\omega_c$  は毎秒のスペクトラム重心、 $F(\omega)$  は入力波形に FFT(Fast Fourier Transform)をかけたものである。ラリーシーンの長さは以下のように定義する。

$$L(s) = f_{end}(s) - f_{begin}(s) + 1 \quad (10)$$

ここで、 $f_{begin}$ 、 $f_{end}$  はショット  $s$  の終始点に対応するフレーム番号である。

以上、最大平均音量、最大ピッチ、ラリーの長さ三つの特徴量を用いて重要度を評価する。平均音量は、ラリーの重要度が最も顕著に表れるパラメータである。また、声援などが起こらないラリーシーンではピッチは極端に低い値となり、重要でないシーンの識別に特に有効である。さらにラリーシーンの長さ  $L$  は、ラリーの攻防の激しさと相関があると仮定して考慮した。式(8)、(9)、(10)で算出した各値を 0~1 に正規化し、以下の式により各ラリーシーンの重要度を定める。

$$R(s) = V_{mean}(s) + P_{mean}(s) + L(s) \quad (11)$$

### 3.3 実験結果

#### (1) ラリーシーン検出

本手法の有効性を検証するために、実験によりラリーシ

ーン検出の精度を示す。比較手法には、SVM による先行研究[1]を用いた。実験には、全 892 ショットに対し 192 のラリーを含むテニス動画を用いた。評価尺度には Precision, Recall を用いる。正解となるラリーの数を  $T$ 、それぞれの手法により検出したラリーの数を  $D$ 、検出したラリーが実際のラリーと一致した数を  $C$  とすると Precision, Recall は以下の式で表せる。

$$\text{Precision} [\%] = \frac{C}{D} \times 100$$

$$\text{Recall} [\%] = \frac{C}{T} \times 100$$

(12)

ここで使用する正解データは、検証するテニス動画の全ショットに対して人手でラベリングしたデータである。ラリーシーンの検出結果を表 1 に示す。この結果から、学習を用いない本手法によっても従来手法と同等の精度が確認できた。本手法は従来手法と異なり、その他のラケットスポーツ動画への適用が可能である。そこで、その他の動画に対して精度評価を行う。検証動画として上記と異なるテニス動画（全 745 ショット、内ラリーシーン 157 ショット）とバドミントン動画（全 406 ショット、内ラリーシーン 94 ショット）を用いた。その実験結果を表 2 に示す。バドミントン動画に関しては Precision の低下が見られるが、これは試合予定を説明するテロップを囲む白枠を誤検出したためである。しかし、Recall が非常に高く、テニス以外の動画に対しても適切にラリーシーンを取得できることがわかる。

## (2) ラリーシーンの重要度評価

本手法による重要度評価結果の妥当性を検証するため、重要度の算出値とラリーの内容との関係について考察する。テニス①を用いた各ラリーの重要度を図 5 に示す。その結果、重要度の低いラリーシーンとしてフォルト（1 回目のサービスミス）やレット（サーブの打ち直し）と呼ばれるどちらの選手にも得点が入らないシーン（図 5：赤点、59 ショット）が集まった。これらのシーンは試合内容を理解する上で重要でないと判断されているため、生成された要約映像には含まれていない。よって、本稿で定義した重要度は閾値設定で重要でないシーンを除外できうる有効な評価方法であるといえる。

## 4. 映像鑑賞手法

### 4.1 映像鑑賞システムの実装

本稿では、前章までに述べたラリーシーンの重要度ならびに各特徴量を調整することにより、ユーザの嗜好に合った要約映像を鑑賞できる動画鑑賞システムを提案する。具体的には、重要度に基づき鑑賞するラリーシーンの数を選択し、前章で述べた各特徴量の性質をもとにユーザの嗜好を反映させた動画鑑賞を可能とするものである。さらに、ユーザの嗜好を反映させた要約映像を

表 1. ラリーシーン検出精度の比較(テニス①)

	Precision[%]	Recall[%]
先行研究[1]	97.4	97.9
提案手法	98.4	97.4

表 2. 提案手法によるラリーシーン検出精度

	Precision[%]	Recall[%]
テニス②	98.1	100
バドミントン	90.4	100

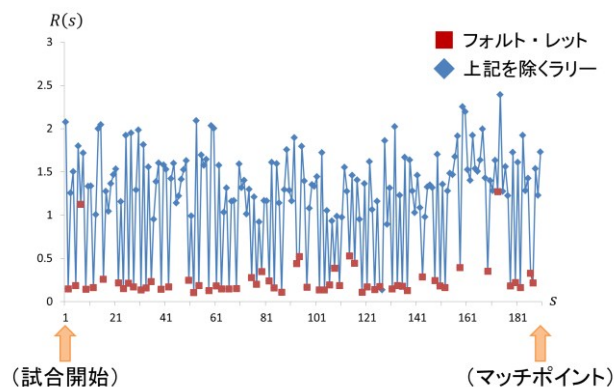


図 5 s 番目のラリーに対する重要度  $R(s)$  (テニス①)

出力する機能も追加した。

これらの機能を使ったとしても、通常再生である限り鑑賞には一定の時間が必要となる。そこで、さらなる効率的な動画鑑賞のために変速再生を利用した鑑賞方法についても検討する。ラケットスポーツ動画には、ラリー内容に視聴者を引き付けるプレイが含まれているとリプレイシーンが付与されている場合がある。著者の鑑賞による調査の結果、リプレイはラリーシーンの全打球ではなく最後数打分だけ再生され、各ラリーシーンにおいて重要な箇所をリプレイする傾向があった。また、ラリーシーンの音声情報は打球音や静寂といったものが多く、ラリーの内容を把握する上で重要ではないため、高速再生を行うことで音声を聞き取れなくても内容理解ができる。Kurihara は動画を高速再生した際、音声、字幕、主映像の順に内容理解可能な再生速度が高くなることを提唱した[4]。この手法は、対象を映画など人物の動作が激しくない動画の鑑賞限界速度をもとに導いた仮定であるため、ラケットスポーツのような動きの激しい動画の鑑賞可能な再生速度の限界値について検討する必要があるが今後の課題とする（現時点では、高速再生による視聴者への負荷を考慮すると 2 倍程度までの高速再生までが適切ではないかと考えている）。

以上に挙げた、ラリーの最後数打の重要度が高いという特徴と人間が鑑賞可能な限界再生速度の二つを考慮し、図 6 に示すようにラリーシーンの最初の部分を高速再生し、

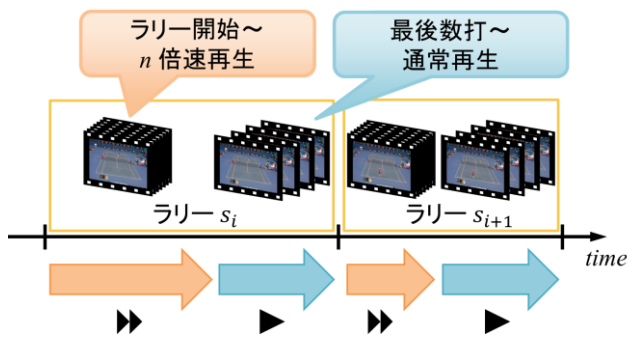


図 6 高速再生を利用した鑑賞方法

最後数打を含む箇所を通常速度で再生する動画鑑賞システムを実装した。これにより、ラケットスポーツ動画の鑑賞において試合の流れを把握しつつ、動画鑑賞時間を大幅に削減することが可能となった。

#### 4.2 本手法による生成動画の総時間

本手法を適用して、どの程度まで動画内容を維持したまま鑑賞時間を削減できるかを検証する。検証は、3章に示した手法で抽出したラリーシーンを4章で示した高速再生によって再生する手法、重要度評価結果が平均値以上のラリーシーンのみを再生する方法の二種類の再生手法を採用した。今回、高速再生法は、各ラリーの最後5秒間を通常再生とし、その他の部分や5秒に満たないラリーシーンは2倍速で再生するものとした。また、検証に用いた動画において重要度の平均値は約1.08であり、3.3節(2)で挙げた重要でないフォルトやレットのシーンをほぼ除外できる値である。テニス①を入力動画とした結果を表3に示す。本手法の適用により重要でないシーンを除外した結果、処理を施さない場合と比べ、鑑賞時間を88.2%削減可能となった。さらに重要度の各特徴量に対して閾値を設定することで、重要な情報を保持したまま効率的に映像を鑑賞できる。

#### 5. まとめと今後の課題

本稿では、ラケットスポーツ動画の映像要約手法として学習なしのラリーシーン検出手法とその重要度評価方法、さらにその効率的な鑑賞のための高速再生手法を提案した。実験の結果、従来のラリー検出と比べ、学習なしで同程度の検出率を得ることができた。今後、条件の異なる動画で検証を行い、手法の有効性をさらに確認していく必要がある。また、重要度評価結果により重要でないシーンの除外に有効性を確認できた。今後の課題として、評価に用いる特徴量の検討や評価の妥当性の検証が挙げられる。特徴量に関しては、選手の総移動量を考慮することで試合内容をより考慮した評価ができると考えられる。評価に関しては、ニュースなどの人手によるハイライト動画との比較実験を行う予定である。動画の変速再生による鑑賞については今後、鑑賞可能な速度の限界値を探る必要がある。今回はラ

表 3. 本手法による映像要約結果

動画内容	総時間[s]	圧縮率[%]
処理なし	6508	—
ラリーシーンのみ	1410	78.3
+高速再生	968	85.1
閾値以上のラリーシーン	1090	83.3
+高速再生	765	88.2

リーシーンに絞った要約手法を提案したが、リプレイシーンなどを分類し、個人の嗜好に応じた要約映像の生成にも着手する予定である。類似ショットのクラスタリング結果を更に有効活用し、ショットのイベント認識ができれば映像編集への応用も可能であると考えている。今後ユーザーのニーズについても調査し、効率的な動画鑑賞を検討していく予定である。

**謝辞** 本研究の一部はJST CREST「OngaCREST プロジェクト」の支援を受けた。

#### 参考文献

- 1) Chunxi, L., Qingming, H., Shuqiang, J., Liyuan, X., Qixiang, Y. and Wen, G.: A framework for flexible summarization of racquet sports video using multiple modalities, *Computer Vision and Image Understanding* 113, pp.415-424, 2009.
- 2) 小林尊志, 野田雅文, 出口大輔, 高橋友和, 井出一郎, 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*, 100(457), pp.165-169, 2011.
- 3) 両角聡, 向井信彦, 小杉信.: 野球中継ダイジェスト版の自動生成システム (高精細画像の処理・表示, 及び一般), *電子情報通信学会技術研究報告 IE, 画像工学*, 103(451), pp.33-36, 2003.
- 4) Kazutaka, K.: CinemaGazer: a System for Watching Videos at Very High Speed, In *Proceedings of AVI'12*, pp.108-115, 2012.
- 5) Tien, M. C., Wang, Y. T., Chou, C. W., Hsieh, K. Y., Chu, W. T. and Wu, J. L.: Event detection in tennis matches based on video data mining, *Multimedia and Expo, 2008 IEEE International Conference on*, pp.1477-1480, 2008.
- 6) Chang, C. K., Fang, M. Y., Kuo, C. M. and Yang, N. C.: Event detection for broadcast tennis videos based on trajectory analysis, *consumer Electronics, Communications and Networks(CECNet), 2012 2nd International Conference on*, pp.1800-1803, 2012.
- 7) Huang, Q., Cox, S., Zhou, X., and Xie, L.: Detection of ball hits in a tennis game using audio and visual information, *Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific IEEE*, pp.1-10, 2012.
- 8) Chen, W. and Zhang, Y. J.: Tracking ball and players with applications to highlight ranking of broadcasting table tennis video, *Computational Engineering in Systems Applications, IMACS Multiconference on*, pp.1896-1903, 2006.
- 9) Lie, L., Hong-Jiang, Z. and Stan, Z. L.: Content-based audio classification and segmentation by using support vector machines, *Multimedia systems*, Vol. 8, Issue 6, pp.482-492, 2003.
- 10) Wang, Y., Zhu, L. and Jin-Cheng, H.: Multimedia content analysis-using both audio and visual clues, *Signal Processing Magazine*, Vol. 17, Issue 6, pp.12-36, 2000.