

ラリーシーンに着目した映像自動要約による ラケットスポーツ動画鑑賞システム

河村 俊哉^{1,a)} 福里 司^{1,3,b)} 平井 辰典^{1,c)} 森島 繁生^{2,3,d)}

受付日 2014年6月30日, 採録日 2014年11月10日

概要: 本論文では, ラケットスポーツ動画を対象とした映像要約手法に基づくシステムを提案する. 既存研究では重要なラリーのみで構成される要約映像を生成したが, 事前に手作業で音響情報をアノテーションする必要があり実用的な手法ではなかった. そこで本研究では, 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とラリーの重要度評価尺度およびその鑑賞方法を提案する. 提案手法では, 画像特徴量を用いたショットクラスタリングおよびラリーシーンを含むクラスタの自動選定により, 高精度なラリーシーン検出を実現する. その後, 各ラリーのショット中の音響情報を考慮したラリーの重要度評価を行うことで, ラケットスポーツ動画の中から重要度の高いシーンのみで構成された要約映像の生成を可能とする. さらに, ラケットスポーツに特化した高速再生機能を用いた動画鑑賞インタフェースを提案し, 試合内容の理解度を保持したまま短時間での動画鑑賞を実現する.

キーワード: 映像要約, ラリーシーン検出, 重要度評価, 高速再生, ユーザインタフェース

A System for Viewing Racquet Sports Video with Automatic Summarization Focusing on Rally Scene

SHUNYA KAWAMURA^{1,a)} TSUKASA FUKUSATO^{1,3,b)} TATSUNORI HIRAI^{1,c)} SHIGEO MORISHIMA^{2,3,d)}

Received: June 30, 2014, Accepted: November 10, 2014

Abstract: In this paper, we propose a video browsing system customized for racquet sports, which reflects interests of users. The method to summarize video that is composed of important rally shots has been discussed in a previous research. However, it is not a practical system because the system needs annotation for auditory events in advance. Therefore we realize automatic rally shot detection based on shot detection and shot clustering using white line detection method. Then, the system calculates the importance of each rally shot on the basis of audio features. As a result, the summarized video facilitates users find and review the information they need. In addition, our method makes it possible for users to create summarized video with minimal effort. Furthermore, we propose a high-speed playback method customized to racquet sports video and realize efficient video browsing experience.

Keywords: video summarization, rally shot detection, shot evaluation, fast-forwarding, user interface

¹ 早稲田大学
Waseda University, Shinjuku, Tokyo 169-8555, Japan
² 早稲田大学理工学術院総合研究所
Waseda Research Institute for Science and Engineering,
Shinjuku, Tokyo 169-8555, Japan
³ JST CREST
a) s_kawamura@ruri.waseda.jp
b) tsukasa@moegi.waseda.jp
c) tatsunori_hirai@asagi.waseda.jp
d) shigeo@waseda.jp

1. はじめに

スポーツ動画鑑賞は動画サイトやテレビを通じて楽しめる娯楽の1つである. その一方で, 試合時間が長いため, 動画を鑑賞するには多くの時間が必要となる. 特にラケットスポーツの試合はトーナメント形式で行われることが多いため, 同時進行される複数の試合を観るためには録画や再放送を利用する必要がある. その結果, 観たい試合の動

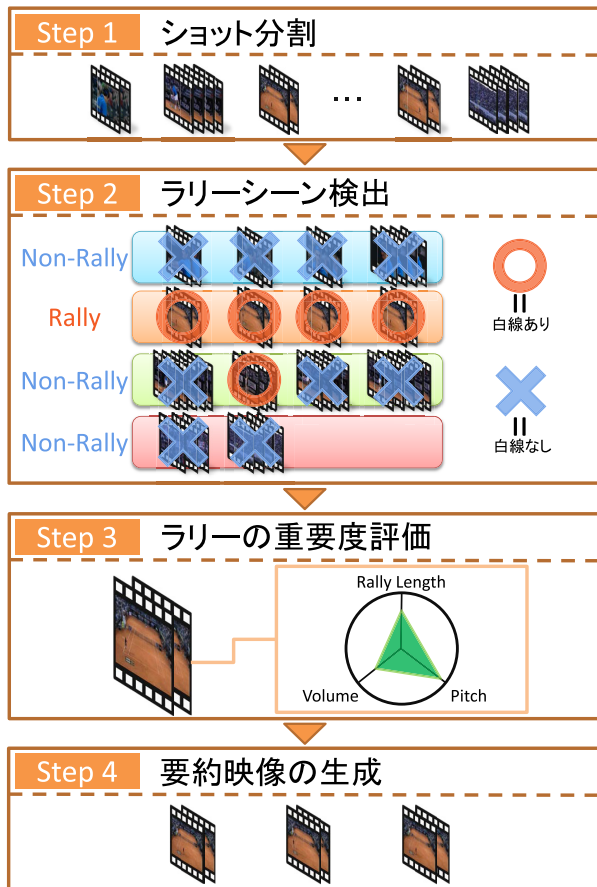


図 1 研究概要
Fig. 1 System overview.

画を鑑賞するために必要な時間はリアルタイムでの試合鑑賞に比べて増加してしまう傾向にある。日常生活において動画鑑賞に費やせる時間は限られており、時間内で試合の流れをすべて理解しながら鑑賞できる試合の数には限りがある。このような観点からラケットスポーツ動画の効率的な鑑賞方法が必要とされている。

短時間で試合内容を把握する方法として、ニュースやウェブ上のハイライト動画の鑑賞があげられる。このような動画の鑑賞は、短時間で試合結果を把握したい場合に有効である。しかし、ハイライト動画は試合結果を理解するための最低限の情報が切り出された内容にすぎないため、1つ1つのプレーなど、試合の詳細を理解することができない。さらに、すべての試合に対してハイライト動画が存在するわけではないため、多くの試合については一試合を通した通常の方法によってしか鑑賞できない。このような背景から、動画中の重要なシーンのみから構成される要約映像を生成する研究が注目されている。特にラケットスポーツ動画では、得点シーン1つをとっても、単純なミスからファインプレーまで幅広く存在するため、視聴者の関心を適切にとらえた要約が必要となる。このような問題に取り組んだラケットスポーツ動画に対する映像要約の既存研究 [1] では、試合内容を把握するうえで重要なシーンを

ラリーシーンと仮定し、ラリーシーンの検出および要約映像の生成手法を提案している。しかし、映像要約をするためには動画の冒頭 30 分程度に手作業でアノテーションする必要がある、そのために必要な時間を考慮と、実用的な手法であるとはいえない。このため、要約映像を容易に生成できる新たな手法が必要である。

本論文では、画像特徴量を用いた事前学習を用いないラリーシーン検出手法および、ラリーの重要度に基づく自動映像要約手法を提案する。本論文における要約映像を生成するまでの流れは図 1 のとおりである。ショットクラスタリング手法とラリーシーンではコートの白線が映り、カメラが静止している時間が長いという特徴を利用し、ラリーシーンを抽出する (図 1, Step2)。検出したラリーに対し、音響情報とラリーの長さに基づき重要度を算出する (図 1, Step3)。最後に、ユーザが動画を鑑賞する時間に応じてラリーの重要度に閾値処理を行い、重要なラリーのみで構成される要約映像を生成する (図 1, Step4)。また、動画の解析結果に基づくインタラクティブな要約と、高速再生を用いることによる試合内容の理解度を保持したままの短時間鑑賞のためのシステムについても提案する。

2. 関連研究

動画を効率的に鑑賞するための手法は、動画中の重要なシーンのみを鑑賞する手法と高速再生を利用して動画を短時間で鑑賞する手法の 2 つのアプローチで研究されている。1 つ目の手法はニュースのハイライト映像のように人手で映像を編集して要約したかのような映像を生成することを目的としており、これまでも様々なスポーツ動画の映像要約に適用されてきた [1], [2], [3], [4], [5]。両角らは野球中継動画において得点時にテロップが表示される特徴に着目し、得点シーンを要約する手法を提案した [2]。この手法はサッカーなど、得点が入るとテロップを表示するようなすべてのスポーツ動画に適用できる点で優れている。しかし、ラケットスポーツ動画のように得点が頻繁に更新され、得点の度にテロップが表示されるわけではないようなスポーツ動画への適用は困難である。Tjondronegoro らはホイッスル音、盛り上がり、テキスト情報の有無からハイライトシーンを検出し、様々なスポーツに適用できる映像要約手法を提案した [3]。また、Zhao らは、放送されるスポーツ動画に含まれるリプレイシーンでは前後にロゴショットが出現する傾向を利用し、様々なスポーツに対応したリプレイ検出を実現した [4]。ハイライトやリプレイはどちらも熟練の編集者が試合の中で特に重要な箇所と感じたシーンであるため、関心度の高いシーンであるといえる。しかし、ハイライトやリプレイは細かなプレー内容に着目したものであるため、試合全体の流れは考慮しない傾向にある。つまり、試合内容を理解するための要約映像の

生成には不向きである。小林らは野球の試合中継動画に対する要約手法として SNS の 1 つ、Twitter のつぶやき数が多い時間と試合の中継時間を照らし合わせることで、関心の高いシーンを要約する手法を提案した [5]。この手法では、視聴者の応援するチームが活躍したシーンのみの要約を可能としたが、中継が行われていない時間がある場合や視聴者の数が極端に少ない場合には適用できないという問題があった。また、ラケットスポーツ動画へ適用することを考える場合、ラケットスポーツのポイントとポイントの時間間隔が短いという特徴から、つぶやきの盛り上がりなどのポイントに対する反応であるかは明確でない。Liu らは試合中の音響情報とそのラベルを SVM により学習することで、ラケットスポーツ動画を要約する手法を提案した [1]。この手法では、ラリーシーンを高精度に抽出したうえで、各ラリーの重要度に応じて順位付けをすることで重要なシーンのみの鑑賞を実現している。しかし、入力動画に対し 30 分程度の音響情報を毎秒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。

2 つ目の手法は、動画をすべて鑑賞しつつ再生時間を短縮することを目的としている [6], [7]。この手法では、動画の内容を認識できる限界の速さで高速再生して動画をすべて鑑賞するため、短時間で動画内容を一通り把握できる。栗原らは、動画中の音響信号を発話区間とその他の区間とに分類し、発話区間は発話を聞き取れる速さで再生し、その他の区間は動画中の動きを把握できる速さで高速再生する手法を提案した [6]。しかし、この手法は発話という観点に注目しているため、映画などの登場人物の発話が重要となる動画が対象となる。そのため、スポーツ動画のように発話区間がほとんどないような動画には適していない。また、Cheng らは動画の再生速度を加速、減速によって調整しながら動画を鑑賞する手法を提案した [7]。この手法によって、ユーザは観たい箇所は重点的に、その他のシーンは重視する度合いに応じて再生速度を変化させながら鑑賞することができる。しかし、この手法は動画の構造についての事前知識がなければうまく鑑賞することができない、毎回同じようなシーンを鑑賞したいとは限らない、観たいシーンとあまり観なくてよいシーンがいつ現れるか分からないため再生速度の制御が難しい、といった問題がある。

ラケットスポーツに関連した研究として、ラリー中に起きたイベントを検出する手法が提案されている [8], [9], [10], [11]。Tien ら、Chang らのテニス動画のラリー中に含まれるサーブやネットプレイといったイベントを検出する手法 [8], [9]、Huang らのテニス動画のラリーシーンからボールの打点を検出する手法 [10]、Chen らの卓球動画のラリー中のボールと選手を追跡する手法 [11] などがあげられる。イベント検出を利用することでラリーの内容まで考慮した各ラリーの重要度を評価できると考えられるが、スポーツごとに最適な検出方法を設定する必要が

あるため、ラケットスポーツ動画全般を対象とした要約手法の実現は難しい。ラケットスポーツ動画中のイベント検出を行うためには事前にラリーシーンを検出することが不可欠である。Zhonga ら、Kijak らによるラリーシーン検出手法では、入力動画に対してショットクラスタリングを行った後、ラリーシーンを選択している [12], [13]。Zhong らは物体（選手）の動きや線の検出を用いた手法を提案した [12]。具体的には、色特徴量に基づきクラスタリングモデルを事前に作成している。しかし、この手法の実現には、入力動画が持つコートの色と同じクラスタリングモデルを用意する必要がある。さらに、ラリーシーンを決定する手法として、各ショットのキーフレーム 1 枚に対する線検出を行っているが、適切なキーフレームを選択できなければラリーシーンに似た別のシーン（リプレイシーンなど）との区別が困難となる。Kijak らはテニス動画のショットを 4 つのクラスタに分類し、キーフレームの色特徴量と HMM によるショットの出現パターンの学習によってラリーシーン（一回目のサービスミス、短いラリー、ラリー）およびリプレイの検出を行った [13]。しかし、前述したイベント検出手法と同様に、テニス動画に特化した手法であり、ラケットスポーツの種目ごとに検出手法を変更する必要がある。

以上をふまえ、ラケットスポーツ動画の映像要約を自動で行うためには、高精度かつ自動でラリーシーンを検出する手法が必要である。また、再生時間の長いラケットスポーツ動画を視聴者が重要なシーンだけに着目してより効率的に鑑賞するためには、適切な評価に基づく重要なシーンの抽出と高速再生による鑑賞時間の短縮という 2 つの手法の長所を統合して利用することができる。本研究では、ラケットスポーツ全般に対して、事前学習を行わない自動ラリーシーン検出手法とその重要度の評価手法および、効率的な鑑賞方法について検討する。

3. ラケットスポーツ動画の構造

本論文で扱うラケットスポーツ動画の構造について記述する。放送されるラケットスポーツ動画は 3 つの区間に分けられる。表 1 に 3 つの区間（放送開始～試合開始前、試合開始～試合終了、試合終了後～放送終了）それぞれに含まれる主なシーンを示す。表 1 に含まれるシーンの中で

表 1 ラケットスポーツ動画中に含まれる主なシーン
Table 1 Representative scenes included in racquet sports video.

区間	主なシーン
放送開始～試合開始前	実況・解説者の会話、選手紹介、練習風景、客席の様子
試合開始～試合終了	ラリー、コートチェンジ、客席の様子、リプレイ、コート整備、選手のアップ
試合終了後～放送終了	実況・解説者の会話、客席の様子、インタビュー、表彰

試合内容の理解に直接つながるシーンはラリーシーンとリプレイシーンであると考えられる。ラリーシーンはつねにスコアボードが表示されており、かつラリー中の全打球が含まれる。一方、リプレイシーンではスコアボードが表示されず、ラリー中の数打しか含まれない。よって、本論文では試合内容を理解するうえで最も重要なシーンはラリーシーンであると仮定し、ラリーシーンによって構成される要約映像の生成を実現する。

4. ラリーシーン検出

Liu らにない、試合内容を把握する上でラリーシーンが最も重要であると仮定し、ラケットスポーツ動画中のラリーシーン検出手法について検討する。図 2 にラリーシーン検出の流れを示す。

4.1 ショット分割

ラリーシーンを検出するために、Liu ら [1] と同様に

ショット分割およびクラスタリングを行う。しかし、Liu らが用いたカラーヒストグラム差分によるショット分割ではフェードによるカットを考慮していないため、隣接するショットが誤って同一ショットであると推定されるという問題があった。そこで、Lian ら [14] のピクセル差分、RGB ヒストグラム差分、ブロックマッチング差分を用いた手法により、カット点およびフェード区間を検出する。

4.2 ショットクラスタリング

前節の手法により N ショットに分割された動画に対し、Liu ら [1] のクラスタリング手法を用いて、ショットを類似した色特徴量を持つショットのクラスタに分類する (図 2, Step1)。クラスタリングは各ショットのキーフレームを用いて行う。 i 番目のショットのキーフレームは、等間隔に抽出した L 枚のフレーム k_l^i ($l = \{1, L\}$) とする (Liu らと同様に $L = 5$ とした)。抽出した各キーフレームからクラスタリングのための特徴量として、HSV 特徴量による 256 次元ヒストグラムを作成する ($H * S * V = 16 * 4 * 4$)。その後、 i 番目と j 番目のキーフレームを用いて、ショット間の距離を式 (1) により算出する。

$$SD(s_i, s_j) = \frac{1}{2} (M + \hat{M}) \quad (1)$$

$$M = \min \left\{ d(k_{l_1}^i, k_{l_2}^j) \right\} \quad (2)$$

$$d(k_{l_1}^i, k_{l_2}^j) = \frac{1}{256} \sum_{b=1}^{256} |H_{l_1}^i(b) - H_{l_2}^j(b)| \quad (3)$$

s_i, s_j はそれぞれ i 番目、 j 番目のショット、 l_1, l_2 はショットのキーフレーム番号、 $d(k_{l_1}^i, k_{l_2}^j)$ はキーフレーム間の距離、 $H_{l_1}^i(b)$ はヒストグラムの b 番目のビンの値である。また、 M, \hat{M} はそれぞれキーフレーム間の距離が最小であることを表す。式 (1) による計算を N ショットのすべてのショットの組合せで行い、式 (1) を最小とするようなショットの組を類似ショットとしてクラスタリングする。つまり、一度の試行につき 2 つのショットがクラスタリングされる。次に、新たに生成されたクラスタ n の持つキーフレームのヒストグラム H_l^n を更新するため、式 (4) を適用する。

$$H_l^n = \frac{N_i * H_l^i + N_j * H_l^j}{N_i + N_j} \quad (l = \{1, L\}) \quad (4)$$

N_i, N_j はそれぞれ統合する前のショット s_i, s_j が持つフレーム数である。式 (1) から式 (4) までの処理を繰り返すことでクラスタリングを行うが、処理の終了条件を定めなければ式 (1) の距離が最小となるショットどうしをクラスタリングし続けてしまう。そこで、最終的なクラスタリング回数を設定するための評価式として、クラスタ内・クラスタ間分散を基にした J_n 、クラスタ数を基にした k_n を用いて式 (5) を定める。式 (5) が最小となる回数を最終的なクラスタリング回数とする。



図 2 ラリーシーン検出の流れ

Fig. 2 Overview of rally shot detection.

$$E_n = J_n + k_n \quad (5)$$

$$J_n = \frac{\sum_{c=0}^{K_n} J_w^c}{J_{inter}} = \frac{\sum_{c=0}^{K_n} \sum_{i=0}^{N_c} \|H_{L/2}^i - H_{mean}^c\|}{\sum_{i=0}^N \|H_{L/2}^{i_{init}} - H_{mean}\|} \quad (6)$$

$$H_{mean} = \frac{1}{N * L} \sum_{i_{init}=0}^N \sum_{l=1}^L H_l^{i_{init}} \quad (7)$$

$$H_{mean}^c = \frac{1}{L} \sum_{l=1}^L H_l^n \quad (8)$$

$$k_n = \frac{K_n}{N} \quad (9)$$

J_{inter} はクラスタリングを行う前の各ショット間の式 (1) の距離の分散, J_w^c はクラスタ c 内の各ショット間の式 (1) の距離の分散, N , N_c はそれぞれクラスタリングを行う前の総ショット数, クラスタ c 内のショット数, K_l はクラスタリングの過程で存在するクラスタ数である. $H_{L/2}^{i_{init}}$, $H_{L/2}^i$ はそれぞれクラスタリングを行う前の各ショットの持つ中心のキーフレームのヒストグラム (ショットの中心フレームが最もショットの特徴を表すと仮定した) とクラスタ c のクラスタの中心フレームが持つヒストグラム, H_{mean} , H_{mean}^c はそれぞれ各ショットが持つヒストグラムの平均値とクラスタ c 内のショットが持つヒストグラムの平均値である. また, クラスタリングを行う前のクラスタ内の分散はすべて 0 とする. 図 3 に評価式の各項とクラスタリングの試行回数との関係を示す. 式 (5) は, クラスタリングが進むにつれて J_l が大きくなるが, k_l はクラスタ数が少なくなるため減少する傾向を持つ. クラスタリングの結果, 生成されたクラスタは色味の類似したショットにより構成される. ラケットスポーツ動画のショットをクラスタリングした場合, 選手のアップやラリー, リプレイ, 観客席などのクラスタが生成される (図 2, Step1).

4.3 ラリークラスタの判別

ラリーシーンの特徴に着目してラリークラスタの自動判別を行う (図 2, Step2). 初めに, 総ショット数に対する

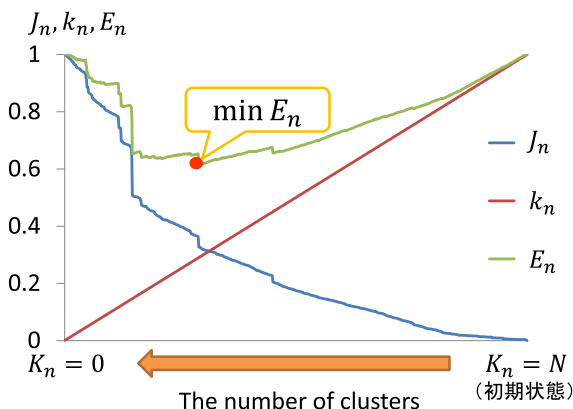


図 3 クラスタ数に対する評価式 E_l , J_l , k_l の変動

Fig. 3 Instances of transition of evaluation function (E_l , J_l and k_l) for the number of clusters.

ラリーシーンのショット数の割合は 10%以上を占める傾向があるため, ショット数の少ないクラスタはラリーシーンのクラスタではないといえるため候補から除外する. 次に, ショット数の多いクラスタの中からラリークラスタを選定するため, ラリーシーンには必ず存在するコートの白線に着目する. しかし, 白線の有無のみに着目してラリークラスタを判定するだけでは白線の存在するリプレイやその他のシーンとの区別ができない. そこで, 各ショットに映る白線の動きに着目する. ラリーシーンでは選手が大きく画面から外れるような動きをしない限りカメラは静止しているため, 一定時間以上白線は同じ位置に映り続ける傾向がある. 一方, ラリーシーン以外ではカメラは被写体を追いかけるような動きをすることが多く, 1つのショットの間に白線が同じ位置に映っていることは少ない. ラリーシーン以外でもカメラが動かず白線が同じ位置に映るようなショットも存在するが, そのようなショットはそのショットを含むクラスタの総ショット数に対して非常に少ない. 以上の特徴をふまえたうえで, 各ショットの平均画像 (図 2, Step2.A) を生成すると, ラリークラスタ内のショットから生成した平均画像からは白線が検出される確率が高くなる. 表 2 に, 約 2 時間のテニス動画に対して白線検出を行った際のラリークラスタとその他のクラスタにおけるクラスタ内の総ショット数と平均画像から白線が検出されたショット数の例を示す. 表 2 のラリー (A) のように, 白線が必ず存在するにもかかわらず平均画像から白線が検出されないショットも見られた. これは, ラリー中にカメラ領域から外れた選手を追跡するようにカメラが動いた結果, 平均画像中の白線がぼやけるためである. そこで, 各ショットの平均画像を基に白線の検出を行い (図 2, Step2.C), クラスタ内のショット数に対する白線が検出されたショットの割合が一定以上のクラスタをラリークラスタと判定する.

ラリークラスタを判別するための白線検出の具体的な手順について述べる. 初めに, 各ショットの平均画像 (図 2, Step2.A) に対し, 白色領域検出およびエッジ検出を用いて生成した 2 枚の二値画像の論理積によって白線画像を生成する (図 2, Step2.B). ここで, 平均画像をグレースケール化した際に輝度値が 170 以上となる領域を白色領

表 2 クラスタの種類とそのクラスタ中の総ショット数および白線が検出されたショット数の例 (テニス動画)

Table 2 Example of white line detection for each clusters in a tennis video.

クラスタの種類	総ショット数	白線が検出されたショット数	割合 [%]
ラリー (A)	160	97	60.6
ラリー (B)	50	38	76.0
選手のアップ	86	3	3.5
リプレイ	29	1	3.4

表 4 本手法によるラリーシーン検出精度
Table 4 Accuracy of our rally shot detection method.

入力動画の種類	大会名	ラリー数	全ショット数	Recall	Precision
テニス (クレー)	Rome Semi-Final 2012	192	962	0.984	0.900
テニス (ハード)	Indian Wells Semi-Final 2012	157	744	1.000	0.981
バドミントン 1	Mumbai Masters 2013	94	408	1.000	0.949
バドミントン 2	BWF Championships Final 2013	109	789	0.972	0.972
卓球 1	世界卓球 2013	70	290	1.000	0.642
卓球 2	London Olympics Semi-Final 2012	88	629	1.000	0.688
Total	-	710	3822	0.992	0.864

表 3 ラリーシーン検出精度の比較

Table 3 Comparison of the accuracy of rally shot detection (Liu's method and our system).

検出手法	Recall	Precision
Liu et al. [1] (半手動)	0.984	0.900
本手法 (全自動)	0.984	0.900

域, エッジは平均画像に対して Canny 法を適用したものとした. このようにして生成した白線画像に対し, 確率的 Hough 変換を用いることで白線の有無を検出する (図 2, Step2.C). ただし, ラリーシーンに限らず表示されるスコアボード領域が白枠である場合, ラリーシーン以外でも白線が検出される可能性がある. そこで, 白線を検出する領域は [(画像の高さ)/5, (画像の高さ)*4/5] の範囲とする. 白線の有無を判定後, クラスタ内のショット数に対して平均画像に白線の残ったショットの数が 60%以上を占めるクラスタをすべてラリークラスタとする. 以上の手順により, ラリーシーンの自動検出が可能となる.

4.4 検出精度の評価

本手法によるラリーシーンの検出精度を評価するための実験を行った. まず, 本手法と Liu ら [1] の検出手法との比較実験を行った. 実験には 962 ショット中 192 ショットがラリーであるテニス (クレーコート) 動画を用いた. また, 評価尺度として Precision, Recall を用いた. また, 評価尺度として Precision, Recall を用いた. 実際のラリーのショット数を T , 各手法によって検出したラリーのショット数を D , 検出したラリーと実際のラリーとが一致したショット数を C とすると, Precision, Recall はそれぞれ以下の式で表される.

$$Precision = \frac{C}{D} \quad (10)$$

$$Recall = \frac{C}{T} \quad (11)$$

比較実験の結果を表 3 に示す. この結果から, 本手法と Liu らの手法は同等の検出精度を得られたことが分かる. この結果はショットクラスタリングで得られたクラスタの中からラリーシーンで構成されるクラスタを適切に選択できていることを示している. よって, 手作業でアノテ

ションを行う必要があった Liu らの手法と同等の検出精度をすべて自動で実現できたといえる. また, 表 4 にはテニス, バドミントン, 卓球動画にそれぞれに対して本手法を適用した結果を示す. この結果, すべての動画で高い Recall が得られたが, 卓球動画においては他動画に比べ Precision が低くなった. これは, 卓球動画ではラリー前やコート整備時にコートを映したラリー以外のシーンを検出してしまったためである. しかし, ほとんどの動画で Recall の値が高く, 要約映像を生成するにはほぼすべてのラリーを検出することができる.

5. ラリーの重要度評価

5.1 重要度の評価手法

重要なラリーのみで構成される要約映像を生成するため, 4 章で検出した各ラリーの重要度を算出する手法について述べる. 本論文の目的を達成するため, ラリーの重要度は, (1) 試合内容を理解するために不可欠なラリー (ゲームポイント, マッチポイント) は重要度が高いとして評価されること, (2) 視聴者を惹きつけるラリーは試合の流れに関係なく重要度が高いとして評価されること, の 2 点を満たす必要がある. 今回はラリーの重要度を定量化するにあたり, (2) のような評価を行うことを目標として重要度を定義する. テニス, バドミントン, 卓球それぞれの動画に対して複数種類のラリーを観察した結果, ラリーの内容が観客を惹きつけるほど, ラリー後に観客から起こる歓声が大きくなるという傾向があった. この歓声を評価するため, ラリー後の音量 V およびピッチ P を重要度算出のための特徴量として考慮する. 音量は歓声の大きさを評価するもので, ラリーの内容が観客を惹きつける内容である場合やそのラリーが試合で重要な得点である場合に特に大きくなる傾向がある. ピッチは歓声の音の高さを評価するもので, 観客が盛り上がっているほど高い音の歓声を出す傾向に基づいている. ピッチは得点が入ったかどうかで値が大きく異なる傾向があり, たとえばテニス動画において一回目のサービスミス (フォルト) 時のピッチに着目すると値は極端に小さくなる. また, ラリーが長いほどラリー中の攻防が激しく, 視聴者はそのラリーに惹きつけられると仮定し, ラリーの長さ L も考慮する. 以上を考慮して, r

番目のラリーに対して、最大音量 V_r , ピッチ P_r , ラリーの長さ L_r の3つの特徴量をそれぞれ抽出した後に標準化を行い、重み係数 $\alpha, \beta, \gamma, \delta$ を用いて、重要度 I_r を式 (12) により算出する。

$$I_r = \alpha V_r + \beta P_r + \gamma L_r + \delta \quad (12)$$

最大音量 V_r , ピッチ P_r の具体的な算出方法について記述する。ラリー後に現れる歓声が最も盛り上がっている箇所をそのラリーの盛り上がりとして考慮するため、最大音量 V_r , ピッチ P_r は、各ラリーのショットの後半 (位置 x_{begin} とする) からそのラリーのショットの次ショットの冒頭5秒間まで (位置 $x_{end} + 5$ とする) の区間において、5秒間の長さの窓を1秒ごとにシフトさせながら分析し、その最大値を用いる。 r 番目のラリーに対する最大音量 V_r は式 (13) により算出される。

$$V_r = \max_{x_i \in [x_{begin}, x_{end}]} \left\{ \frac{1}{5F_s} \sum_{n=1}^{5F_s} V(x_i + n) \right\} \quad (13)$$

x_i は分析窓の探索開始位置、 F_s は音声のサンプリング周波数、 V は音量 (振幅) である。ピッチ P_r は、1秒ごとに窓をシフトさせながらサンプリングしたスペクトラム重心 ω_c を基にした式 (14) により算出される。ここで、スペクトラム重心とは音響信号に対する周波数空間上での重心を指し、音響信号中に含まれる高域の音響信号の割合が高いほど大きな値となる。

$$P_r = \max_{x_i \in [x_{begin}, x_{end}]} \left\{ \frac{1}{5} \sum_{n=1}^5 \omega_c(x_i + n) \right\} \quad (14)$$

また、ラリーの長さ L_r は r 番目のラリーのショット中に存在するフレーム数である。式 (12) の重み係数 $\alpha, \beta, \gamma, \delta$ は事前に行った主観評価実験によって得られた値である。具体的には、ラケットスポーツの種目ごとに各ラリーの面白さを5段階評価してもらい、その値を目標値として V_r, P_r, L_r による重回帰分析を行うことで算出した (実験の詳細については5.2.1項で述べる)。以上の方法によって各ラリーの重要度を算出する。算出されたラリーの重要度に対して閾値処理を行うことで任意の時間内に収まる要約映像の生成が可能となる。

5.2 評価実験と結果

5.2.1 主観評価実験による評価関数の算出

5.1節で定義した重要度の評価関数 (式 (12)) の重み係数を算出するため、主観評価実験を行った。実験では、男女7名 (内、テニス経験者1名、バドミントン経験者1名) に合計120ラリー分の動画を鑑賞してもらい、それぞれの面白さを5段階評価してもらった。被験者には各スポーツのルールを把握してもらうため、実験前にテニス、バドミントン、卓球動画を10分以上鑑賞してもらった。また、実験に用いた動画は、テニス動画 (2試合分)、バドミントン

動画 (2試合分)、卓球動画 (2試合分) の中から、できるだけラリーの内容や長さに偏りがないように各動画から20ラリーずつを試合の流れに関係なくランダムに抽出することで作成した。主観評価結果に基づく式 (12) の重み係数は、各ラリーの評価値の平均値を用いた。評価結果を観察したところ、スポーツの経験者と未経験者の評価結果は似た傾向にあり、評価結果が異なる場合は個人差であると考えられる。よって、本論文で定義した評価関数 (式 (12)) によって、スポーツ経験を問わずに満足のいく要約映像を生成できる。

5.2.2 評価関数の相関係数

本論文で定義した評価関数 (式 (12)) によって算出した重要度と視聴者の関心との間に相関があることを示すため、主観評価実験の評価値に対する、重要度および各特徴量 (最大音量、ピッチ、ラリーの長さ) それぞれの相関係数を算出した。その結果を図4に示す。この結果から、テニス、バドミントン、卓球すべてについて本論文で定義した重要度が主観評価実験の評価値との相関が高く、視聴者の関心を適切に表現できていることが分かる。

5.2.3 評価結果

本論文で定義した評価関数 (式 (12)) を用いて、約2時間のテニス動画の重要度の評価を行った。その結果を図5に示す。この結果から、フォルト (1回目のサービスミス) やレット (サーブのやり直し) と呼ばれるシーンの重要度が

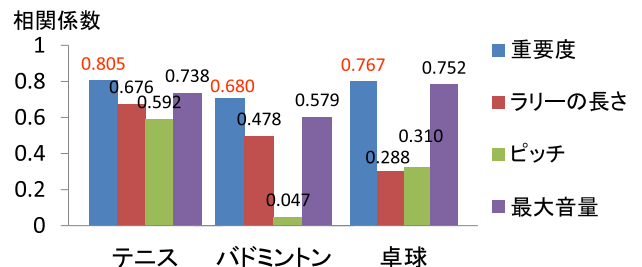


図4 主観評価値に対する評価関数による重要度および各特徴量との相関係数

Fig. 4 Correlation coefficient between subjective values and evaluation function values (an importance of rally and each feature value).

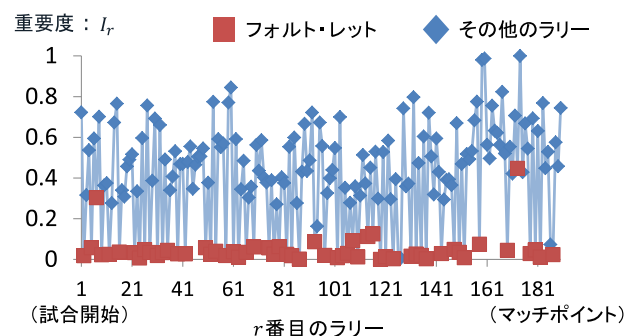


図5 重要度評価の結果例

Fig. 5 Instances of transition of evaluation function.

低いと判定されていることが分かる(図5, 赤プロット: 56ショット). よって, 本論文で定義した評価式は閾値処理により重要でないシーンを除外できる有効な評価手法であるといえる. また, 本論文で定義した評価関数によって重要度が高いと判定されたラリーを観察したところ, 視聴者を惹きつけるラリー以外にも試合の流れを把握するうえで重要なポイント(ゲームポイントやマッチポイント)が多く含まれた. これは評価関数(式(12))において「盛り上がり」を考慮したことで, 視聴者を惹きつけるラリーと試合内容を理解するためのラリーの両方を重要なラリーと判定できたことによる. 一方で, サービスエースという視聴者が関心を惹きつけやすいと考えられるシーンが, ラリーの長さが短いため適切に評価されなかった. このような特定のイベントを適切に評価できるよう, 今後, 各特徴量間の関係性を考慮する, 選手の動きを考慮できるような特徴量を定義する, といった改善を行っていく.

6. 鑑賞インタフェースの実装

6.1 鑑賞インタフェース

本論文では, 前章で定義したラリーの重要度や後述する高速再生機能を用いることで, ユーザのニーズに合わせて要約映像を鑑賞できるシステムを提案する(図6). このシステムによって, ユーザは動画を鑑賞しながら重要度の閾値処理を行い, 試合への興味に応じて鑑賞時間をインタラクティブに調整できる. また, タイムラインではラリー区間を可視化している(図6下部, 青色区間). この可視化は, ラケットスポーツ動画の鑑賞経験があるユーザに対して特に有効であり, ラリーとラリーの間隔が長い箇所の前後はゲームやセットの開始位置あるいは終了位置であると推測できるため, 特定のゲームやセットの探索も容易となる.

6.2 高速再生による鑑賞支援

重要度に対する閾値処理によって鑑賞するラリーの数を



図6 ユーザインタフェースの外観

Fig. 6 User interface.

制限したとしても, 特定の時間内に鑑賞できるラリーの数には限りがある. そこで, 高速再生を用いて再生時間を短縮しつつ, 観たい情報を把握できるような機能を実装した. この機能を用いる目的は純粋に試合を楽しむためではなく, 試合内容の理解度を保持したまま短時間で把握するためである. 具体的には, ラリーの始めの区間を高速再生(n 倍速)し, ラリーの最後数打を含む区間(t 秒間)を通常再生または高速再生で再生する機能である. ラリーの始めの区間の再生速度については, 動画内容の理解度に応じた3段階の速度を設定した. 動画内容の理解度と再生速度の関係を求めるため, 著者によるラケットスポーツ動画の観察の結果, ラリー内容を理解するための要素として次の3つがあると仮定した(得点, 選手の動き, 球の軌跡). 1つ目の「得点」は, どちらの選手が得点したかどうかの理解度を指し, 試合内容の理解, 特に勝敗に関わる重要な要素である. 2つ目の「選手の動き」は, ラリー中に選手がどのように打球したかの理解度を指し, 球に追いつくまでの移動や打ち方からラリー中のそれぞれの選手が優勢または劣勢かを判断する基準となる. 3つ目の「球の軌跡」は, ラリー中に球がどのような軌道を通ったかの理解度のことで, 球速や軌道もラリー内容を把握することができる. つまり, 以上の3つの要素に対する理解度を保持したまま動画を鑑賞できる再生速度を明らかにすることで, 短時間でラリーシーンの内容を理解することができる. そこで, 主観評価実験によりラケットスポーツにおけるラリーの理解度と再生速度の関係を検証した. 具体的には, テニス, バドミントン, 卓球の各6ラリーを10倍速から通常再生速度まで0.5倍速刻みで順に再生していき, それぞれの要素に対する理解度を5段階評価してもらった実験を20代の男女10名(内, ラケットスポーツ経験者4名)に対して行った(栗原ら[6]の実験方法を参考にした). 図7にテニス, バドミントン, 卓球の各動画に対する得点, 選手の動き, 球の軌跡の理解度(全被験者の平均値)と再生速度の関係を示す. この結果, 得点, 選手の動き, 球の軌跡の理解度はスポーツの種目によらず再生速度が上がるにつれてどれも等しく下がることが分かった. よって, 図7の結果からスポーツごとに, 理解度が80%以上, 50%以上となるような再生速度および, 通常再生速度の3段階の再生速度を設定できる機能を実現した. また, ラケットスポーツ動画におけるリプレイシーンではラリー中の最後数打のみ再生される傾向にあること, 再生速度が上がると得点の理解度が低下することの2点に基づき, ラリーの最後数打の区間のみを通常の再生速度で鑑賞できる機能も追加した.

6.3 本システムによる要約結果

本システムを用いて動画時間が約2時間のテニス(クレイコート)動画を鑑賞した際の情報を紹介する. このときの鑑賞時間の例を表5に示す. ここで, 要約映像を生成

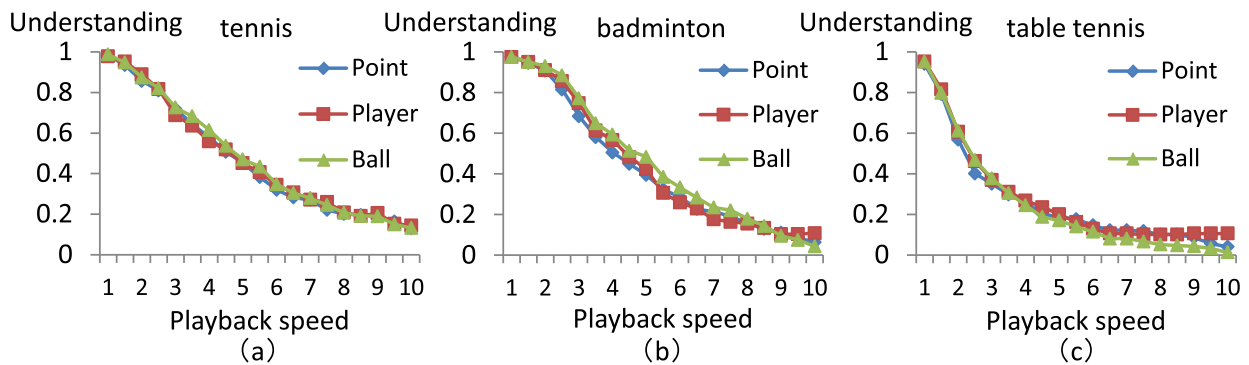


図 7 再生速度と理解度の関係 (a) テニス, (b) バドミントン, (c) 卓球

Fig. 7 Relationship between the playback speed and the understanding (a) tennis, (b) badminton, (c) table tennis.

表 5 本手法による動画の鑑賞時間例

Table 5 Example of time required for browsing a racquet sports video.

	総時間 [s]	圧縮率 [%]
元動画	6,508	-
ラリーのみ	1,410	21.7
重要なラリーのみ	1,181	18.1
重要なラリーのみ + 高速再生 (理解度 80%) [n = 2.49, t = 5]	768	11.8
重要なラリーのみ + 高速再生 (理解度 50%) [n = 4.18, t = 5]	656	10.1

する際の重要度の閾値には、5.2.3 項で述べた重要でないシーン (図 5, 重要度 0.2 未満のラリー) を除くことが可能な値を設定した。より高い閾値を設定することでさらに重要なラリーのみで構成される要約映像を生成することもできる。しかし、閾値を上げると歓声上がるようなラリーが要約映像に残る一方で、試合の流れまで理解しての鑑賞が難しくなる。また、視聴者が満足する要約内容は視聴目的によって大きく異なる (試合内容を分かる程度に鑑賞したい、試合中で惹きつけられる数ラリーのみを鑑賞したいなど) ため、適切な閾値を設定することは難しい。よって、ユーザインタフェースに閾値設定機能を実装し、ユーザが要約映像の長さを鑑賞中に変更できるようにした。また、動画内容の理解度が 80%、50% となる再生速度で鑑賞する際には、どちらの選手が得点したか理解できるようにラリーの最後 5 秒間を通常再生する設定とした (5 秒未満のラリーは高速再生した)。この結果、重要なラリーシーンの情報を保持したまま元動画の 10% 程度の時間で鑑賞可能となった。

6.4 本システムに対するユーザ評価

本インタフェースを用いて、ラケットスポーツ動画を鑑賞してもらうユーザ評価を行った。実験においては、被験

者にシステムの使い方を事前に説明したうえでテニス動画、卓球動画、バドミントン動画を鑑賞してもらう形式で行った。初めに、ラリーシーンのみを再生して動画を鑑賞する方法については、ラリーとラリーの間の時間を省いて鑑賞できるため動画をすべて鑑賞する場合よりも退屈せず良かったという意見が得られた。また、選手がどのように打ったかといった細かい動きまで観たい人にはリプレイシーンが必要な可能性があるが、試合の良い場面のみを楽しみたい人にとってはラリーシーンのみで十分であるという意見も得られた。次に、本論文で定義した重要度に対する閾値を自由に設定してもらうことで適切な要約結果が得られるかアンケート調査を行った。その結果、閾値をインタラクティブに変えることで、よりユーザの好みに応じた映像要約結果が得られることが分かった。特にテニス動画においては、フォルト・レットを取り除く閾値を設定して鑑賞した場合においても違和感なく鑑賞できることが分かった。今後、誤検出の修正機能や、重要度算出方法の変更機能などによって、さらに質の高い要約結果が得られるようなシステムを検討している。最後に、高速再生機能によって動画を鑑賞してもらう実験を行った。高速再生を用いた鑑賞においては、高速再生から通常再生に切り替えることでラリーの終わりが予見できてしまうことが想定されるため、この機能は短時間で試合内容を把握したいユーザ向けの機能である。しかし、被験者からはポイントが決まる瞬間が強調されて良いという肯定的な意見もあった。一方で、再生速度が切り替わる瞬間に違和感があるという意見があった。今後、再生速度を滑らかに変化させるといった改善が必要である。

7. まとめと今後の課題

本論文では、ラケットスポーツ動画を対象とした映像の自動要約手法として、ラリーシーンの特徴を利用したラリーシーン検出手法とその重要度の評価手法を提案した。また、さらなる効率的な鑑賞のためのユーザインタフェースも提案した。ラリー検出では事前のラベリングを必要と

しない自動化を実現しながら、従来手法と同等の高い検出率が得られた。ただし、この手法はカメラが一定時間以上静止することを前提としているため、すべてのスポーツに適用することは難しい。また、放送されるようなラケットスポーツ動画が対象であり、一般のカメラで撮影されたような動画には対応できない。今後は白線への着目方法を変える、白線以外の特徴を考慮する、といった方法により適用できる動画の拡張が見込まれる。ラリーの重要度評価では、ラリーの長さ、ピッチ、最大音量の3つの特徴量を考慮することでラリーの重要度を評価し、重要でないシーンの除外という観点での有効性が見られた。しかし、テニス動画の場合、視聴者の関心を惹きつけやすいサービスエースがラリーの長さの考慮のために低く評価されることがあった。そのため、ラリーの重要度を評価する際には各特徴量間の関係まで考慮する必要がある。今後はより試合内容を反映した要約を行うため、ラリー中の選手の動きやリプレーションについても考慮していく予定である。提案した鑑賞インタフェースでは、興味に応じて要約結果をインタラクティブに変化させながら動画を鑑賞することが可能となった。また、高速再生機能を利用することで試合内容の理解度を保持したままさらに短時間で要約映像を鑑賞する方法を実現した。このインタフェースは、従来のテレビなどにおけるリモコンで実現可能な早送りやトラックのスキップ機能を、より動画内容に即した高度な形で拡張したインタフェースである。動画内容の理解に基づく鑑賞インタフェースにより、さらに効率的な動画の鑑賞が可能となることが期待できる。今後、鑑賞するスポーツの経験者と未経験者との興味の違いやユーザ個人の嗜好を反映できる特徴量についてさらなる検討を重ね、より効率的かつ満足に動画を鑑賞できるシステムの実現を目指したい。

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。

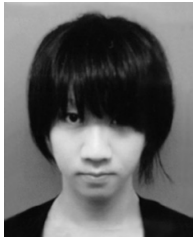
参考文献

- [1] Liu, C., Huang, Q., Jiang, S., Xing, L., Ye, Q. and Gao, W.: A framework for flexible summarization of racquet sports video using multiple modalities, *Computer Vision and Image Understanding*, Vol.113, No.3, pp.415–424 (2009).
- [2] 両角 聡, 向井信彦, 小杉 信: 野球中継ダイジェスト版の自動野球中継ダイジェスト版の自動生成システム, 電子情報通信学会技術研究報告, Vol.103, No.451, pp.33–36 (2003).
- [3] Tjondronegoro, D., Chen, Y.P.P. and Pham, B.: Integrating highlights for more complete sports video summarization, *IEEE Trans. Multimedia*, Vol.11, No.4, pp.22–37 (2004).
- [4] Zhao, F., Dong, Y., Wei, Z. and Wang, H.: Matching logos for slow motion replay detection in broadcast sports video, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.1409–1412 (2012).
- [5] 小林尊志, 野田雅文, 出口大輔, 高橋友和, 井手一郎, 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, Vol.100, No.457, pp.165–169 (2011).
- [6] 栗原一貴, 佐々木洋子, 緒方 淳, 後藤真孝: 音声区間自動検出技術を用いた変速再生方式による映像の高速鑑賞システムの検討, 情報処理学会ヒューマンコンピュータインタラクション研究報告, Vol.2012-HCI-149, No.13, pp.1–5 (2012).
- [7] Cheng, K.Y., Luo, S.J., Chen, B.Y. and Chu, H.H.: SmartPlayer: User-centric video fast-forwarding, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.789–798 (2009).
- [8] 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, *Proc. IEEE International Conference on Multimedia and Expo*, pp.1477–1480 (2008).
- [9] Chang, C.K., Fang, M.Y., Kuo, C.M. and Yang, N.C.: Event detection for broadcast tennis videos based on trajectory analysis, *Proc. 2nd International Conference on Communications and Networks (CECNet)*, pp.1800–1803 (2012).
- [10] Huang, Q., Cox, S., Zhou, X. and Xie, L.: Detection of ball hits in a tennis game using audio and visual information, *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (2012).
- [11] Chen, W. and Zhang, Y.J.: Tracking ball and players with applications to highlight ranking of broadcasting table tennis video, *Proc. IEEE International Conference on Computational Engineering in Systems Applications*, Vol.2, pp.1896–1903 (2006).
- [12] Zhong, D. and Chang, S.F.: Structure analysis of sports video using domain models, *Proc. IEEE International Conference on Multimedia and Expo*, pp.713–716 (2001).
- [13] Kijak, E., Oisel, L. and Gros, P.: Temporal structure analysis of broadcast tennis video using hidden Markov models, *Proc. SPIE Storage and Retrieval for Media Databases*, Vol.5021, pp.289–299 (2003).
- [14] Lian, S., Dong, Y. and Wang, H.: Efficient temporal segmentation for sports programs with special cases, *Proc. Advances in Multimedia Information Processing – PCM 2010 PT I*, Vol.6297, pp.381–391 (2010).



河村 俊哉 (学生会員)

1991年生。2014年早稲田大学先進理工学部物理学科卒業。同年より同大学院修士課程に在学。2013年より動画像処理に関する研究に従事。ACM会員。



福里 司 (学生会員)

1990年生。2012年早稲田大学先進理工学部応用物理学科卒業。2014年同大学院修士課程修了。同年より同大学院博士後期課程に在学。2011年より動画画像処理，コンピュータグラフィックスに関する研究に従事。画像電子学

会，ACM各会員。



平井 辰典 (学生会員)

1988年生。2011年早稲田大学先進理工学部物理学科卒業。2012年同大学院修士課程修了。同年より同大学院博士後期課程に在学。2010年より音楽情報処理，動画画像処理に関する研究に従事。2013年より日本学術振興

会特別研究員(DC1)。2014年米Microsoft ResearchにてResearch Internとして研究に従事。現在に至る。2012年度IPA未踏スーパークリエイター。ACM会員。



森島 繁生 (正会員)

1959年生。1987年東京大学大学院工学系研究科博士課程修了。工学博士。同年成蹊大学工学部専任講師。1988年同助教授。2001年同電子工学科教授。2004年から早稲田大学理工学部応用物理学科教授。現在に至る。新潟

大学非常勤講師を併任。1991年電子情報通信学会業績賞，2010年電気通信財団テレコムシステム技術賞受賞。画像電子学会副会長，芸術科学会理事，日本顔学会理事。日本音響学会，映像情報メディア学会，電子情報通信学会，日本心理学会，IEEE，ACM各会員。