

野球中継のハイライトシーン実時間配信を目的とした特徴のマイニングによるPCシーンの自動検出

Automatic Extraction of PC Scenes Based on Feature Mining for Real Time Delivery of Baseball Highlight Scenes

正会員 熊野雅仁[†], 正会員 有木康雄^{††}, 正会員 塚田清志^{†††}

Masahito Kumano[†], Yasuo Arika^{††} and Kiyoshi Tsukada^{†††}

Abstract To replay baseball highlight scenes in live broadcasts to baseball fans outside, image processing, such as analysis, meta information extraction, and automatic editing, has to be performed in real time. This paper proposes high-speed image processing that automatically extracts PC (Pitcher and Catcher) scenes from live broadcasts of a baseball game in real time using a feature mining technique as a part of baseball highlight scene delivery. This method achieves an F-measure of 97.2% and a processing speed 30 times faster than actual time.

キーワード: 野球中継, ハイライトシーン, リアルタイム, 特徴選択, マイニング

1. はじめに

近年, WWW などの大規模マルチメディア空間や, 携帯電話などのモバイルマルチメディア空間が拡大, 普及している. さらに, 映像や音声などのマルチメディアコンテンツを, 家庭のコンピュータや携帯電話などに実時間配信することが可能となっており, 不足するコンテンツと制作者の人材不足を補うデジタル技術や, 必ずしもプロの制作者を必要としない, もしくは補助するコンテンツ生成技術, またその生成支援技術などが注目されている. コンテンツ不足を補い制作コストを抑え, 必ずしもプロの制作者を必要としない対象として, スポーツ映像が注目されており, 特にプロの試合だけでなく, 民間のスポーツを対象として, 地域性の高いコンテンツを生成するためのデジタル技術も注目されている.

本研究が対象とする野球についても, 高校野球を対象とした, スコアボード付中継映像のインターネット配信実験が

行われている¹⁾. すでに EPG などの番組情報付コンテンツの配信が始まっており, プロ野球の試合内容に関するインデクス情報が将来付与されることを前提として, ユーザの好みに合わせたダイジェストを見ることのできる, インタラクティブ TV システムも提案されている²⁾. また, 人手で更新される WWW 上で公開されているスコアボード情報を利用した, ダイジェストシステムも提案されている³⁾.

このように, 多くの人が場所を問わず, 自分の好みに合った情報を入手できるサービスが提供されようとしている. しかし, 以上のシステムは, いずれも人手, もしくは既に何等かの方法で, インデクス情報が付与されたことを前提としている. 近年の多チャンネル化や, サービス形態の多様化によるインデクス情報の増大を考えると, 人手によりインデクス情報を付与することは現実的でない.

また, 映像コンテンツの場合, 特に商品価値が高いものは, アーカイブ化された過去の映像コンテンツではなく, 同時性を有する, 現在もしくはある出来事が起こった直後の映像コンテンツであり, 実況中継などのライブコンテンツもその一つである. このような, 実況中継を前提とした映像コンテンツの生成では, 実況中継中, あるいは実況中継を完了した直後に配信を始める必要があるため, 予め時間をかけてインデクス情報を付与し, 構造化しておくことができないという問題を解決する必要がある.

以上の点から, 本研究では, スポーツ実況中継, 特に野球の実況中継映像を対象として, 自動的にインデクス情報を付与して, ハイライトシーンを検出することを目的としている. 特に, 本研究が着目するサービスの形態は, 外出

2004年5月31日受付, 2004年9月24日再受付, 2004年10月14日採録

[†] 龍谷大学 理工学部

(〒520-2194 大津市瀬田大江町横谷 1-5, TEL 077-543-5111)

^{††} 神戸大学 工学部

(〒657-8501 神戸市灘区六甲台町 1-1, TEL 078-803-6226)

^{†††} 株式会社毎日放送 メディア開発局

(〒530-0013 大阪市北区茶屋町 17 番 1 号, TEL 06-6359-1123)

[†] Faculty of Science and Technology, Ryukoku University
(Otsu 520-2194 Japan)

^{††} Faculty of Engineering, Kobe University
(Nada, Kobe 657-8501 Japan)

^{†††} Mainichi Broadcasting System, Inc.

(17-1 Chayamachi, Kita-ku, Osaka, 530-0013 Japan)

中の野球ファンに対し、野球中継中のハイライトシーンを速報として、リアルタイムに配信するシステムであり、人手によるインデクス情報を用いないシステムである。このようなシステムを実現するためには、インデクス情報を自動付与するだけでなく、映像をデジタル化しながら並行して、コンテンツ解析を行うオンライン処理、かつすべての処理がリアルタイム内に収まるよう、高精度で高速に処理する手法が必要となる。

そこで、野球の実況中継映像を対象に、オンライン処理を前提としたシステムの実現を主眼に置き、実時間音声認識技術と実時間画像解析技術を統合した、野球映像中のハイライトシーン配信システムの部分システムとして、PCシーンを高速で高精度に自動検出する手法を提案する。

2. 関連研究

映像を構造化するには、音声、言語、画像など、マルチモーダル情報ストリームの協調的処理が有効であることが示されている⁴⁾。本研究は、実時間向けの画像処理とアナウンサーの実時間音声認識を協調させるアプローチを用いているため、この部類に入る⁵⁾。画像処理のみを基本として、デジタル化された後の野球映像にインデクスを付与したり、それをもとに構造化する研究としては、カメラワークを用いてシーンを特定する研究⁶⁾や、その情報をもとに、ホームランシーンのカメラワークをテンプレートとして、DP マッチングにより他のホームランシーンを検出する手法⁷⁾、差分画像から得られる動作情報を用いて、投球やスイングを DP マッチングで検出し、インデクス情報を付与する手法⁸⁾⁹⁾、映像中のテロップを解析する手法¹⁰⁾¹¹⁾がある。また、音声認識のみによりヒットなどのインデクスを付与する手法¹²⁾などが提案されている。この他にも、人手によるメタ情報の付与作業を支援する技術を背景として、検索を目的としたデータマイニングによるシーン検索システムも提案されている¹³⁾。しかし、いずれもオフライン処理を前提としているため、ライブ映像に対応できない。また、クローズドキャプションなどの言語情報を用いる手法¹⁴⁾もあるが、訓練された人手による作業が前提となる。

本研究の対象である、投球ショット(本研究では放送用語として PC[Pitcher and Catcher] ショットと呼ぶ)に対し、そのテンプレートとカメラワークを抽出して映像を構造化する手法¹⁵⁾も提案されているが、PC ショットの検出を高精度化するために、マスキングする領域は手動で指定しているため、最適性に疑問が生じる。

この他にも、ショットの分類を目的として、野球の意味モデルをベイジアンネットで表現する手法¹⁶⁾、最大エントロピー法を用いる手法¹⁷⁾また、ハイライト抽出やダイジェスト作成を目的として、最大エントロピー法を用いる手法¹⁸⁾や、HMM を用いる手法¹⁹⁾などがある。このように、野球に限定しても、2 次的なコンテンツ生成や生成支援・応用システムに対する関心の高さが伺える。



図 1 ショットの種類
Types of shots.

3. ハイライトシーン配信システムの概要

3.1 PC ショットと撮影環境

本研究で取り扱うハイライトシーンは、投手が投げたボールをバッターが打つことにより発生するイベントに限定する。野球の試合進行は、このイベントの繰り返しであると考えられるため、本研究での映像の意味的な最小単位は、投手の投球に始まり、次の投球が行われるまでの区間である。プロ野球中継の場合、球場に配置されるカメラの位置はほぼ決められており、我々は、それらのカメラを用いて撮影されるショットを図 1 のように分類している。図 1 中の記号は、それぞれ PCS(Pitcher and Catcher Shot:ピッチャーの投球場面でバックスクリーンから撮影されているショット)、LS(Long Shot:グラウンドの様子を撮影している高い位置からのショット)、MS1(Middle Shot 1:グラウンドに配置され、選手の動きを追うショット)、MS2(Middle Shot 2:LS より低い位置から撮影されるショット)、TS(Tight Shot:選手の顔などを撮影するショット)、FS(Full Shot:球場全体を撮影するショット)、AS(Audience Shot:観客を撮影するショット)、OS(Other shot:その他のショット)を表している。このような分類法は、ショットサイズを背景として定義している²⁰⁾。本研究では、これらのショット集合から PCSのみを検出することに主眼を置くため、PCS 以外のショットは NPCS(Non-PCS) と呼ぶ。

また図 1 のショットは、複数台のカメラとカメラのスイッチングを行う中継車によって撮影が行われる。しかし、カメラの台数はコスト等に関連して一定とは限らず、1 台のカメラで撮影可能な被写体は、すべて撮影対象とするため、カメラとショットの種類は基本的に対応しておらず、カメラの数が減るほどその傾向は強くなる。このため、ショットの種類は、カメラが特定されてもショットを特定できないため、メタ情報の付与は基本的に人手を必要とする。また、この撮影環境を構築する業者と放送局は、基本的に別会社となる場合が多く、放送局は、中継車から送られてくる映像の放映権について契約を行っている。このため、メタ情報を用いたシステムを構築する場合、メタ情報は新たな取引の対象となり、恒常的なコストを必要とすることになる。

我々は、コストの削減や新たなビジネスの対象となるメタ情報の自動取得や、人手によるメタ情報付与作業の支援

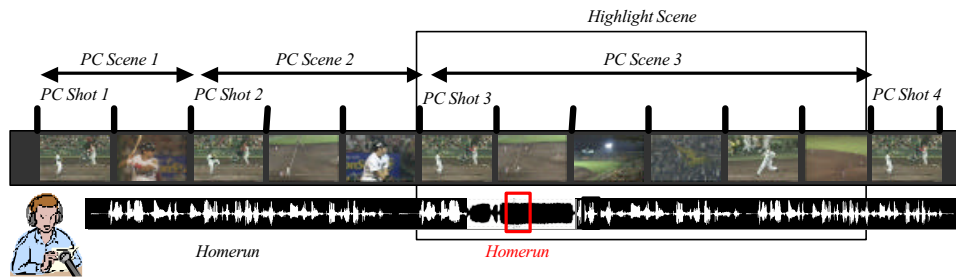


図 2 PC ショットとハイライトシーンの関係
Relation between PC shot and highlight scene.

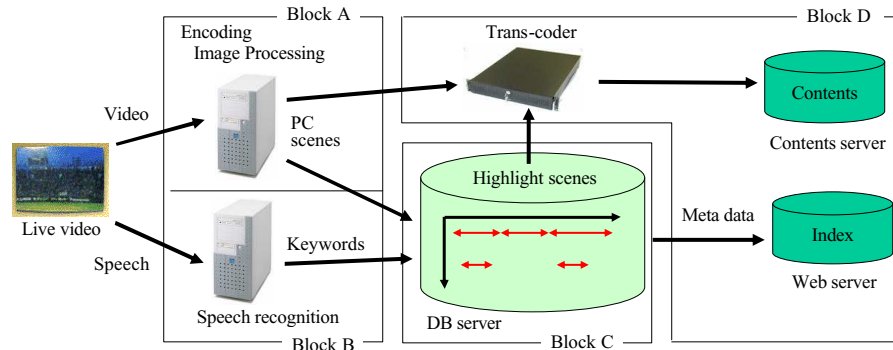


図 3 ハイライトシーン検出システムの概略
Overview of a highlight scene delivery system.

技術に着目しており、音声や画像の解析による問題解決手法に着目している。本研究では、中継車から送られてくる映像を対象とした PC ショットの自動検出に焦点を当てる。

3.2 PC ショットと PC シーン

プロ野球中継映像では、投球が行われるショットとして、図 1 の PCS が必ず撮影される。これは、野球の試合進行という情報の意図を、どのようにして視聴者へ正確に伝えるのか、また大衆に広く受け入れられる表現はどのような提示法が良いのかなどが検討された結果、導かれた経験則としてのショットの一つであり、今では一般化している。このような経験則の別の例として、バッターがボールを打った場合は、基本的に LS が接続される。LS は、バッターが打ったボールがカメラのフレーム内に入るよう、グラウンドの様子を上から撮影し、カメラワークによってボールを追いかけるショットである。このように接続することによって、バッターが打ったという意図をショットの接続という映像表現を用いて伝えているのである。もし、PCS でバッターがボールを打ったにもかかわらず、ここで TS が接続されると、ボールの行方がどうなったかを知ることができず視聴者は混乱する。このように、視聴者に野球の進行をわかりやすく伝えるために守らなければならない規則がある。このような映像制作者の意図を視聴者へ正確に伝えるための規則は、映像文法²¹⁾と呼ばれており、本研究では、この野球の映像文法に従って撮影された映像を基盤とする。

すでに述べたとおり、ハイライトシーンを定義する上での意味的な最小単位は、野球の映像文法を基盤とすることにより、投手の投球に始まり、次の投球が始まるまでの区間として定義できる。つまり、図 2 のように、ある PC ショット

から次の PC ショットまでの区間を PC シーンと定義する。

3.3 PC ショットの判定

PC ショットを抽出するためには、まずショットの切り替えを検出する必要がある。本研究で扱う野球映像では、ショットの切り替えにカットだけでなく、ディゾルブも含まれるため、これらのショットの切り替えを高精度に効率良く検出する方法が望まれる。本研究では、MPEG 1 の映像に対し、高精度かつ実時間の 1/30~1/50 でカットやディゾルブを検出できる手法²²⁾を用いてショットの切り替え点を検出した²³⁾。検出されたショット切り替え点で挟まれた区間を、本報告ではショット区間と呼ぶ。PC ショットは、カメラを完全に固定してはいないが、構図は安定している。そこで、PC ショットの判定を行う場合、ショット区間から代表フレーム(先頭の 1 フレーム)を取り出し、1 枚のフレーム画像が PC ショットに属するものであるかどうかを判定すればよい。この手法を用いることで、大量の画像データを処理する必要がなくなり、計算量を大幅に減らすことが可能となる。ショットの切り替え判定が実時間の 1/30~1/50 であるため、PC ショットの判定が高速であれば、デジタル化の処理を含めてもリアルタイム処理が可能となる。

3.4 ハイライトシーン検出システム

本研究で開発しているハイライトシーン検出システムの全体像を図 3 に示す。図 3 において配信用映像生成部 (Block D) では、入力されたライブ映像をモバイル配信用の映像フォーマットに変換する。メタデータ生成部 (Block A,B) では、ハイライトシーンの映像検索に用いるメタデータを生成する。メタデータ生成の流れは、まず、ライブ映像と中継音声の同期をとり、それぞれ別々の計算機に取り込む。



図 4 PCS の変動要素
Variation in PCSs.

次に、映像解析用計算機 (Block A) では、部分的にデジタル化が終了した入力映像を順次的に処理し、ショットの切り替えが起こった直後の画像フレームを対象として PC ショットの判定を行う。PC ショットと判定された場合は、次に現れる PC ショットを待って PC シーンを切り出し、各シーンにおける始末端の時刻情報をメタデータとして XML 形式でデータベース (Block C) に出力する。

一方、音声解析用 PC (Block B) では、入力された中継音声から無音区間を抽出して音声を分割する。分割された音声区間に対して、音声認識を実行しキーワードを検出する。検出されたキーワードを、その始末端の時間情報とともに XML 形式でデータベースに出力する。これらの XML ファイルにより、ハイライトと判定された映像の区間情報に基づいて、モバイル配信用に変換された映像の区間を携帯端末に配信する。

つまり、ハイライトシーンとは、PC シーンのうち、音声認識と音声のパワーを用いてイベントと判定されたキーワードを含む区間となる。画像解析により PC シーンの区間を決定し、音声認識により PC シーンの中からハイライトシーンを特定するため、画像・音声の協調システムとなっている。このためシステムの性能は、PC ショットをいかに精度よく検出できるかに依存する。本研究は、図 3 中、Block A を実現する。

4. PC ショットの判定法

4.1 PC ショットの変動要素

PC ショットは、安定した画像であると思われるが、実際にはデーゲーム、ナイトゲームの要因以外にも、様々な変動要因がある。図 4 の (a) は典型的な PC ショットであり、(b)~(f) は、PC ショットの画像上、変動する要素を示したものである。PC ショットは、固定されたカメラで撮影されているが、実際の野球映像では、(b) のように、カメラの撮影方向が多少左右にずれる。これは、右バッターや左バッターが PC ショットの中へ適切に入るよう、カメラを左右に動かすことができるようになってきているためである。この点においては、画像のずれに頑健な特徴量が望まれる。また、図 4(c) のように、ピッチャーの位置やフォームが変化したり、(d) のように、バッターの位置が異なる場合もある。この他にも、PC ショットの判定においては、色情報を

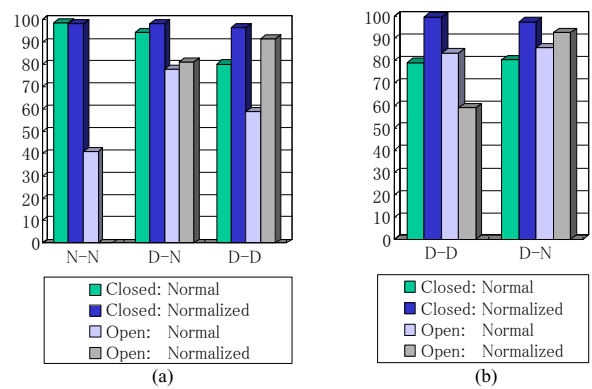


図 5 ヒストグラム法による PCS の検出結果
Result of extraction of PCS by histogram method.

用いる方法が考えられる。しかし、(e) のように、選手のユニホームの色が変化したり、(f) のように、天候や照明の影響によって、全体の色が変化するため、必ずしも安定した特徴量を得ることができない。

4.2 ヒストグラム法による実験

このような構図がある程度定まった画像のシフトや選手の位置関係の変動、また正規化することで色や照明の変化に強い特徴としてヒストグラムが考えられる。図 5(a) は、ユニホーム色を 2 球団に限定し、デーゲーム (D) とナイトゲーム (N) の組合せで実験した結果、また図 5(b) は、ユニホーム色を 3 球団に限定して PC ショットの学習を行い、判定実験を行った結果である。いずれも学習データに用いた画像を対象とした Closed な実験に対しては、非正規化で 80% 以上の精度を示し、正規化すると精度が上がる。しかし、Open な実験にすると基本的に精度が悪くなり、正規化した方が良い場合もあれば、正規化をすることで、精度が 0% になる例もあった。これらの不安定要因を吟味し、新たな処理を単純に追加して PC ショットを精度良く判定しようとするれば、計算量を増大させてしまうことになる。

このように、構図が安定している割に安定して高い精度が得られないことがわかる。これは、不安定要因をすべて特徴量に反映させていることに問題があると考えられる。そこで本研究では、精度を保ちつつ計算量が少なく、また若干の平行移動に強い上に、色の変化に左右されない特徴量を画像中から自動的に検出し、判定を行う手法を提案する。

4.3 特徴量のマイニング

図 4 に示すような PC ショットの代表フレーム画像 (以後、PC 画像と呼ぶ) の構図は、ピッチャーとバッターで特徴付けられるが、ポーズの変動やユニホームの変化などにより、特徴量としては必ずしも安定していない。より安定した特徴としては、グラウンドや、観客席とグラウンドの間の壁などが想定できる。しかし、人手による安定領域の決定は信頼性がない。また、同一種類より多種類の特徴量により PC を決定づける方法が考えられる。そこで、安定した領域を抽出するために、学習データから特徴のマイニングを行い、PC ショット判定の識別率の高い安定した特徴の

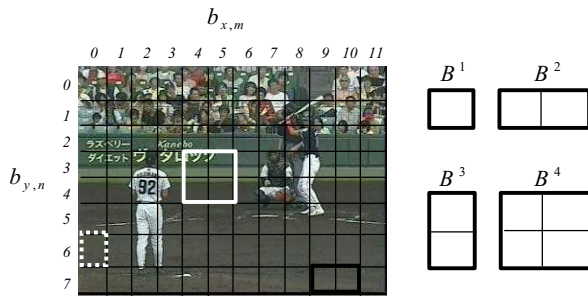


図 6 ブロックと領域の種類
Block and area type.

種類とその特徴に対応した領域を選択し、組合せることで判定処理を行うアプローチを提案する。

まず、PC 画像を図 6 のようにブロックに分割し、図 6 右に示す B^1, B^2, B^3, B^4 といった四つのブロックの組合せブロック内で輝度値の平均値、分散、さらに分散の対数の特徴量として着目する。これは次の理由からである。

- (1) グラウンドや壁の領域は平坦な輝度値を持っているため分散値が小さく、
- (2) 色の大きく異なる部分が含まれていても図柄が安定しており、
- (3) PC 画像のみで構成したオンセット (PC 画像集合) 内では、分散が安定している

また、輝度値の分散の対数を用いる理由は、例えば領域内の輝度値が大きく 2 極化する場合、構図は安定しているにも関わらず、その分散は大きくなる。つまり、その場合はある大きな値で安定している可能性がある。そこで、対数化された特徴が持つ安定性を抽出するためである。

- (4) ブロック内の輝度に大きな差がある場合でも、対数を取ることによって、分散値の桁が安定している

例えば、図 6 の白い枠内には、フェンスの下に位置する壁があり、文字がある。このブロック内では、壁と文字、それぞれの輝度値は安定していると考えられるものの、壁の輝度値と文字の輝度値には大きな差があり、ブロック内の分散は大きくなってしまふ。分散の対数を取ることで、文字の変化や文字位置の変化などによる分散値のばらつきをある程度抑制でき、桁の安定性を得ることができる。この意味においては、観客席の特徴を表現できる可能性もある。

次に、安定度という指標に注目すると、ブロックから得られる 3 種類の特徴量は、PC 画像集合中、どの PC 画像においても、それぞれよく似た値であることが望ましい。したがって、それぞれの特徴量の分散値が PC 画像集合内で小さいブロックを、PC ショットとして判定する際の良い特徴領域と考えることができる。そこで、学習用の野球映像から PC ショットを取り出し、それぞれの代表フレームとして取り出した PC 画像の集合を学習データとしたとき、この学習データから得られるすべてのブロックの特徴量に対して、その PC 画像集合内で分散値を求め、この分散値を昇順に並べて上位に位置するものが良い特徴量を示すブ

ロックであると判定する。このような方法で特徴量を選択することにより、PC 画像判定処理で用いる比較演算を大幅に減らすことが可能となる。

4.4 PC ショット判定の学習アルゴリズム

学習は、二つのステップに分かれる。1 つは、今回用いた三つの特徴量に基づき、特徴量の分散が安定した領域を選択することである。しかし、この段階では、分散の安定した特徴領域を用いることが PC ショットの判定精度にどう影響するかわからない。そこで、学習セットを対象とした Closed の実験により、異なる特徴量の最適な組合せを決定する。

ステップ 1 では、ショット区間の代表フレーム画像を図 6 に示すように、 x 軸を 12 に、 y 軸を 8 の領域に分割する。また、このブロックを図 6 の右のように組合せた四つの領域タイプ $B^1 \sim B^4$ を設定する。これら四つの領域タイプは、フレーム内でブロック単位に移動する。

そこで、これら四つの領域タイプに対して、フレーム内で存在する位置を領域番号 i として記述する。いま、あるフレーム f において、領域タイプ t 、領域番号 i の領域 $B_{f,t,i}^t$ から得られる平均 $M_{f,t,i}$ 、分散 $V_{f,t,i}$ は、次の式 (1),(2) によって求められ、対数分散 $LV_{f,t,i}$ は $\log V_{f,t,i}$ として求められる。ここで、 $Gray(x,y)$ は、画像の位置 (x,y) の濃淡値、また $|B|$ は、ブロック内の画素数を表している。

$$M_{f,t,i} = \frac{1}{|B_{f,t,i}^t|} \sum_{x,y \in B_{f,t,i}^t} Gray(x,y) \quad (1)$$

$$V_{f,t,i} = \frac{1}{|B_{f,t,i}^t|} \sum_{x,y \in B_{f,t,i}^t} (Gray(x,y) - M_{f,t,i})^2 \quad (2)$$

最後に、特徴の安定した領域を求めるために、各特徴量について安定している順に並べた領域リストを生成する。ただし、ここでは、特徴量として平均値をとりあげ領域安定性を計算する方法について述べる。学習に用いる PC 画像 f の枚数を N としたとき、各 PC 画像上で、同じ領域 $B_{f,t,i}^t$ において平均値の分散を、 $V_{M,t,i} = \text{mean}_f(M_{f,t,i} - \text{mean}_f(M_{f,t,i}))^2$ として求める。

次に、 $V_{M,t,i}$ をすべての t と i に対して昇順に並べたとき、順位 n に位置する分散を $V_{M,t,i}^n$ とする。すると、1 番目に位置する分散は、 $V_{M,t,i}^1$ となり、これに対応する領域 $B_{i,M}^{tM}$ をブロック平均値における最安定領域とする。このとき、フレーム f 上の領域 $B_{i,M}^{tM}$ における平均値を M_{f,t,M,i_M} とすると、 f に関してこの値の最大値を式 (3)、最小値を式 (4) として保存する。この最大値と最小値は、自動決定された閾値となる。

$$M_{t_M,i_M}^{max} = \max_f(M_{f,t_M,i_M}) \quad (3)$$

$$M_{t_M,i_M}^{min} = \min_f(M_{f,t_M,i_M}) \quad (4)$$

同様に、分散の分散 $V_{V,t,i}$ 、対数分散の分散 $V_{LV,t,i}$ を計算し、これを基準として昇順に並べ、それぞれ 1 番目に位

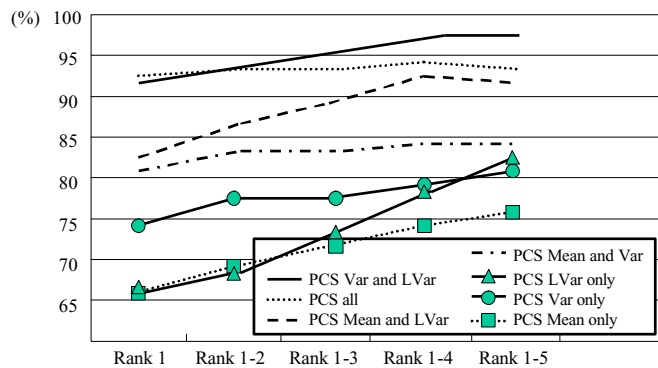


図 7 ステップ 2 の学習実験結果
Result of experiment in learning step 2.

置する分散の分散 $V_{V,tv,iv}^1$ と対数の分散 $V_{LV,tLV,iLV}^1$, また最大値と最小値として, $V_{tv,iv}^{max}$, $V_{tv,iv}^{min}$, $LV_{tLV,iLV}^{max}$, $LV_{tLV,iLV}^{min}$ が得られる.

PC ショットの判定は, 野球映像の各ショットから代表フレームとして取り出した画像を f としたとき,

$$\begin{aligned} (M_{tM,iM}^{min} \leq M_{f,tM,iM} \leq M_{tM,iM}^{max}) \text{ and} \quad (5) \\ (V_{tv,iv}^{min} \leq V_{f,tv,iv} \leq V_{tv,iv}^{max}) \text{ and} \\ (LV_{tLV,iLV}^{min} \leq LV_{f,tLV,iLV} \leq LV_{tLV,iLV}^{max}) \end{aligned}$$

を満たす代表フレーム画像を PC 画像として判定し, その PC 画像が含まれるショットを PC ショットと判定する.

ステップ 2 では, ステップ 1 で得られた特徴量の昇順リストに基づき, リストの上位に位置する特徴量の組合せを生成し, Closed な実験を行って PC 判定の精度を求める. この精度とは, 5 章で定義する $Fmeasure$ を示す. この判定精度を降順に並べたとき, 最も上位にある組合せが判定に最適な特徴の組合せとなる.

図 7 に, ステップ 2 での実験結果を示す. 図中, 横軸は組合せ方を示している. 例えば, Rank1 は, 各特徴の Rank1 のみを組合せた方法, また Rank1-2 は, 各特徴の Rank1 と 2 を組合せた方法である. この結果より, ブロックの分散と対数分散を Rank1-5 まで用いた特徴が, ステップ 2 の Closed 実験の中では最も良い値を示した. 三つの特徴量を用いた場合も比較的高い値を示しているが, ブロックの輝度平均値を取り除いた分散と対数分散の組合せの方が良い結果を示している. これは, 輝度平均もある程度は判定に貢献するが, 逆に輝度情報に左右されるためであると考えられる. 分散と分散の対数はブロック内の輝度の拡散度に依存するため, 色や照明などの変化に大きな影響を受けなかったものと考えられる. ここで, 参考として, 上位 4 位までの選択領域を各特徴量ごとに図 8 に示す. ブロック領域は, y 軸の左上, 左下, x 軸の左上, 右上が属しているブロック番号をそれぞれ y_1, y_2, x_1, x_2 とする.

図 6 の 3 種類の領域 (白点線枠, 黒枠, 白枠) は, 1 の実験において $V_{M,tM,iM}^1$, $V_{V,tv,iv}^1$, $V_{LV,tLV,iLV}^1$ に対応する領域である. 白の破線で囲まれた B^1 タイプのものが $V_{M,tM,iM}^1$,

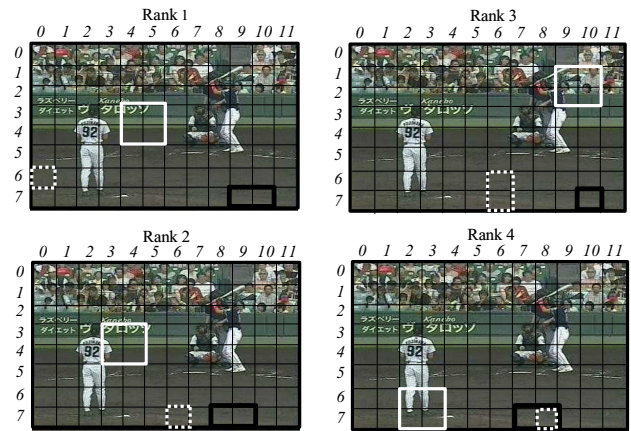


図 8 選択された領域
Selected area(PC).

黒の枠で囲まれた B^2 タイプのものが $V_{V,tv,iv}^1$, 白の枠で囲まれた B^4 タイプのものが $V_{LV,tLV,iLV}^1$ にそれぞれ対応する領域である.

ブロックの平均値と分散に対応する領域は, 画像全体のずれが起こっても輝度値の変化が起こりにくい領域が選択され, また輝度値は一様である部分が選択されていることがわかる. 対数分散に対応する領域は, 壁の輝度値と白字の輝度値の差が大きく, 単純にブロックの分散を計算しただけでは分散値が大きくなると考えられる. しかし, 壁の輝度値と白字の輝度値それぞれが安定しているので, PC 画像集合内の同じ領域を比較すれば, いずれも分散値が近い値を示し, 分散の対数が安定している領域が選ばれていると考えられる.

5. PC ショット判定実験

5.1 実験条件

実験に用いた映像は, 甲子園球場を対象として, 中継車から送信された映像を取り込んだ 320x240, 29.97fps の MPEG1 映像 10 本であり, 人手の情報が付与されていない, それぞれ 2 時間程度の素材映像である. 各映像には, 100~250 程度の PC ショットが含まれている. また, さらなる高速化を目的として, MPEG1 から得られた画像 320x240 を 72x48 ヘダウンサンプリングを行い, カラー画像を濃淡画像へ変換している. このとき, 最小ブロックの縦横の画素数は, それぞれ 6 である.

学習データの選択法としては, 球団ごとの色の違いや昼夜の違いがどのように影響するかを判別するため, 1:阪神 vs 広島のみ, 2:阪神 vs 広島と阪神 vs ヤクルトの組合せというグループに分け, それぞれデーゲーム (D) とナイトゲーム (N) の組合せによる実験を行った. ただし, 提案手法の特徴として, 少ない学習データで高い精度が得られることを示すために, 学習データには, 二つの映像を用いて残り八つの映像を評価する. 例えば, 2:N-N では, 阪神 vs 広島戦のナイトゲームと阪神 vs ヤクルト戦のナイトゲームの 2 本を用いて実験を行う.

表 1 特徴の組合せと Fmeasure のランキング
Fmeasure ranking by combination of features.

教師データ Features	1:N-N			1:D-N			1:D-D			2:D-N			2:D-D		
	M	V	LV	M	V	LV	M	V	LV	M	V	LV	M	V	LV
Rank 1	0-0	1-4	1-4	0-0	1-4	1-4	0-0	1-6	1-6	0-0	1-4	1-4	0-0	1-4	1-4
Fmeasure	96.8%			95.8%			97.5%			96.2%			96.5%		
Rank 2	0-0	1-5	1-5	0-0	1-5	1-5	0-0	1-5	1-5	0-0	1-5	1-5	0-0	1-5	1-5
Fmeasure	95.9%			95.0%			97.3%			95.1%			95.2%		
Rank 3	0-0	1-6	1-6	0-0	1-6	1-6	0-0	1-4	1-4	0-0	1-6	1-6	0-0	1-6	1-6
Fmeasure	95.3%			94.9%			97.2%			95.0%			95.1%		
Rank 4	0-0	1-3	1-3	0-0	1-3	1-3	0-0	1-7	1-7	0-0	1-3	1-3	0-0	1-3	1-3
Fmeasure	93.7%			94.9%			95.5%			94.5%			94.6%		

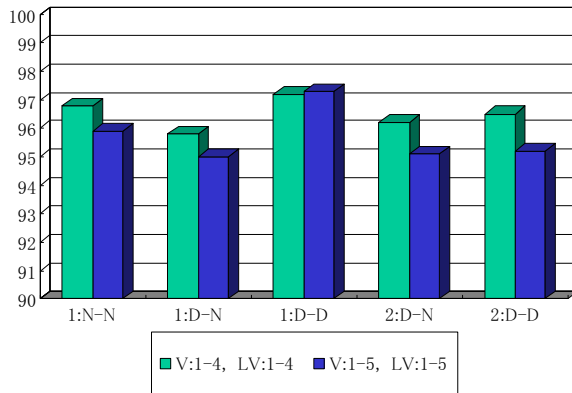


図 9 PC ショット判定結果
Result of PC shot discrimination.

表 2 ステップ 1 における各実験ごとの PC 画像教師データ数
Number of PC images in learning step1 on each experiment.

	1:N-N	1:D-N	1:D-D	2:D-N	2:D-D
PC	277	362	365	402	458

ここで、C を検出対象の正解数、D を正解を検出できなかった数として「未検出数」、E を正解でないものを過剰に検出した数として「過剰検出数」とするとき、再現率 = $C/(C+D)$ 、適合率 = $C/(C+E)$ と定義される。再現率は、検出対象を漏れなく検出できたかという完全性を表現し、適合率は、検出結果の中にどれだけ必要な対象が存在するかという正確性を表現する指標である。システムの性能を評価する場合、漏れがなく、必要な対象だけを抽出することが目的となるため、この再現率と適合率ともに高い値を示すことが求められる。また、Fmeasure は、二つの指標を統合する評価法であり、この場合、再現率と適合率が 100% に達したとき、100% となる。本研究では、未検出がなく、過剰検出のないシステムが理想的であるため、Fmeasure の精度が 100% に近いことが望ましい。

$$Fmeasure = \frac{2 \cdot \text{再現率} \cdot \text{適合率}}{\text{再現率} + \text{適合率}} \quad (6)$$

5.2 実験結果

表 1 に Open な PC ショット判定実験について、上位 5 つの実験結果を示す。縦軸は横軸の特徴量の組合せ方に対する Fmeasure のランキングを示している。また、表 2 に各実験ごとの学習ステップ 1 で教師データとした PC 画像の数を示す。表 1 より、最も高い値を示したのは、学習デー

表 3 実験結果
Result of experiment.

	1:N-N	1:D-N	1:D-D	2:D-N	2:D-D
V:1-4, LV:1-4	96.8%	95.8%	97.2%	96.2%	96.5%
V:1-5, LV:1-5	95.9%	95.0%	97.3%	95.1%	95.2%

タにデーゲームのデータのみを用いた、V と LV の上位 1~6 を組合せた場合である。しかし、この特徴量の組合せは、他の学習データの組合せではいずれも 3 位に位置するのに対し、V と LV の 1~4 や 1~5 は、最高値の 97.5% に近い精度でどの学習データの組合せにおいても高い精度を保持している。

表 1 で 1 位に多く出現する V と LV の上位 1~4 位と、2 位に出現する V と LV の上位 1~5 位の特徴の組合せを抜粋した Fmeasure の棒グラフを図 9、またその数値を表 3 に示す。本研究では、学習データの違いに影響されにくい特徴を良い特徴と位置づけるため、ここではランキングの 1 位に安定して位置する V と LV の上位 1~4 の特徴の組合せを最も優れた特徴と定め、その中で最も高い精度 97.2% をシステム上の最も高い精度と位置づける。

以上の結果より、特徴に分散と分散の対数を用いることが PC ショットの判別に有効であるという結果が得られた。また、いずれも学習データに依存せず、映像 2 本分という少ない学習データで高い精度が得られることも示しており、提案手法の有効性を示すことができた。

ただし、本研究において甲子園球場の PC 画像から得られた特徴量は、他球場でも有効であるとはかぎらない。しかし、本研究で提案した手法は、比較的構図が安定していると思われる画像を対象として、安定要素を探索することに着目した手法である。また、表 3 より、学習データの選定について、昼夜、チームカラー等の画像特徴に注意を払わなくても、いずれも高い精度を示すため、システムの初期調整が容易になると思われる。つまり、表 2 に示した程度の比較的少ない教師データを用いて、検出精度を高くする特徴量をマイニングするため、一つの球場で得られた特徴を他球場に流用するよりも、むしろ各球場ごとに特徴量をマイニングすることに適した手法と考えられる。

実験のグループ 1 における処理時間を計測した。Xeon3.0MHz, 1CPU のパソコン上で、8 特徴 (Rank1-4:L1,L2,L3,L4, LV1, LV2, LV3, LV4) を用いた場合の PC シ

ショット判定の平均処理時間は、最大で8msであった。このように、ブロックの輝度平均、分散、対数分散を求める計算量は非常に小さく、判定に用いる特徴数を8個までに限定することができるため、本手法は精度を高く保ちつつ、計算量を最小限に抑える手法であることがわかる。

6. ま と め

本報告では、野球実況中継映像のハイライトシーンを実時間でネットワーク配信することを目的とし、PCショットを高速で高精度に検出する手法を提案した。PCショット検出では、まず学習ステップ1でいくつかのブロック形状について輝度平均、ブロック内の輝度分散と、その分散の対数値を用い、それらの値が安定なブロックを求めた。次に、ブロックと特徴量の組合せをマイニングし、分散と対数分散によるブロックの上位1~4位を用いたものが学習データに左右されにくく、高い精度を示す特徴量の組合せであることを示した。その特徴量の組合せにより、最大で97.2%の検出精度が得られた。以上より、映像に対するメタデータ付与に関しては、十分に高い精度が得られていると考えられる。提案手法は、PCショットの抽出において、どのような特徴が有効であるかをブロックの位置、形状を考えて発見する手法となっている。今後、この発見的手法を進展させて、PCショット以外のショットについても研究を進めていく予定である。

ハイライトシーン検出を高精度に行うためには、より詳細な構造化を行う手法が必要となる。現在、試合進行の状態遷移を考慮したアナウンサーの音声認識による詳細なメタ情報の取得に関する研究を進めている²⁴⁾。また、ショットの接続に関する映像文法やカメラワークから得られる情報を含め、様々な特徴を用いてハイライトシーンの精度向上を検討する予定である。

【文 献】

- 1) 知念賢一, 吉田豊一, 山口英, 香取啓志: “全国高等学校野球選手権大会 Internet ライブ中継実験報告”, 映像学誌, 51, 6, pp.925-930 (1997)
- 2) 橋本隆子, 白田由香利, 真野博子, 飯沢篤志: “TV 受信端末におけるダイジェスト視聴システム”, 情報処理学論, 41, SIG3, pp.71-84 (2000)
- 3) 田中清, 阿久津明人, 外村桂伸, 秦泉寺浩史: “見たいシーンを見逃さないライブ中継 LiveWatch”, 信学技報, パターン認識・メディア理解, PRMU2002-30, pp.51-56 (2002).
- 4) 馬場口登: “メディア理解による映像メディアの構造化”, 信学技報, パターン認識・メディア理解, PRMU1999-42, pp.39-46 (1999)
- 5) 重森 猛, 金子 剛志, 緒方 淳, 藤本 雅清, 有木 康雄, 塚田 清志, 濱口 伸, 清瀬 基: “音響・言語適応処理を用いたスポーツ実況中継音声の認識へハイライトシーン検出への応用”, 信学技報, 音声研究会, SP2003-166, pp.3-40 (2003)
- 6) 遠藤 齊, 片岡 良治: “カメラモーションに基づく類似動画像検索”, 信学技報, データ工学, DE, 99, 202, pp.147-152 (1999)
- 7) 片岡 良治, 遠藤 齊: “サンプルマッチングによる動画インデクシング支援方式”, 信学技報, 画像工学, IE, 99, 179, pp.59-66 (1999)
- 8) 館山 公一, 川嶋 稔夫, 青木 由直: “野球中継におけるシーン検索”, 第3回知能情報メディアシンポジウム論文集: pp.195-202 (1997)
- 9) 館山 公一, 川嶋 稔夫, 青木 由直: “動作スポッティングによるシーン検索, 情報処理学会研究報告”, CVIM, コンピュータビジョンとイメージメディア, 97, 70, pp.115-122 (1997)
- 10) Zhang, Dongqing and Shih-Fu Chang: “Event Detection in Baseball Video Using Superimposed Caption Recognition”, Proceedings of the 8th International ACM Conference on Multimedia(ACM MM '02), pp.1-6 (2002)

- 11) 広部一弥, 牛尾剛聡, 酒井宏治, 孫魯英, 渡邊豊英: “イベントと状況変化の依存関係に基づいた野球中継のインデキシング支援”, 情報処理学会研究報告. データベース・システム研究会報告, 98, 34, pp.87-94 (1998)
- 12) Yong Rui, Anoop Gupta, Alex Acero: “Automatically Extracting Highlights for TV Baseball Programs”, Proceedings of the 8th International ACM Conference on Multimedia(ACM MM '00), pp.105-115 (2000)
- 13) 金山智一, 瀧本裕一, 小西修: “データマイニング法による映像情報の内容検索”, 情報処理学会, 研究報告, 人文科学とコンピュータ, CH, 99, 59, pp.35-42 (1999)
- 14) 新田直子, 馬場口登, 北橋忠宏: “言語と画像の情報統合によるスポーツ映像からの人物・アクション・イベント抽出”, 電子情報通信学会技術研究報告, パターン認識・メディア理解, PRMU1999-256, 99, 709, pp75-82 (2000)
- 15) 山本拓, 佐藤宏介, 千原國宏: “野球中継映像における各種プレイシーンの自動検索/編集システム”, 2000年電子情報通信学会総大会講演論文集情報・システム 2, D12-77, pp.247 (2000)
- 16) Huang-Chia, Chung-Lin Huang: “A Semantic Network Modeling For Understanding Baseball Video”, Proc. of IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing(ICASSP '03), V, pp.820-823 (2003)
- 17) Wei Hua, Mei Han and Yihong Gong: “Baseball Scene Classification Using Multimedia Features”, Proceedings of the 2002 IEEE International Conference on Multimedia and Expo(ICME '02), I, pp.821-824 (2002)
- 18) Mei Han, Wei Hua, W. Xu and Yihong Gong: “An Integrated Baseball Digest System Using Maximum Entropy Method”, Proceedings of the 8th International ACM Conference on Multimedia(ACM MM '02), pp.347-350 (2002)
- 19) Peng Cheng, Mei Han, and Yihong Gong: “Extract Highlights From Baseball Game Video With Hidden Markov Models”, Proceedings of IEEE International Conference on Image Processing (ICIP '02) I, pp.609-612 (2002)
- 20) 熊野, 有木, 上原, 下條, 春藤, 塚田: “映像編集支援システムのためのショットサイズ自動付与”, 電子情報通信学会論文誌, Vol.J85-D-I, 7, pp.592-602 (2002)
- 21) 天野, 上原, 熊野, 有木, 下條, 春藤, 塚田: “映像文法に基づく映像編集支援システム”, 情報処理学論, 44, 3, pp.915-924 (2003)
- 22) 中島康之, 氏原清乃, 米山暁夫: “部分複合を用いた MPEG データからのカット点検出”, 信学論, J81, D-II, 7, pp.1564-1575 (1998)
- 23) <http://w3-mcgav.lab.kdd.co.jp/mpeg/mpfs/indexe.html>
- 24) 佐古淳, 有木康雄: “タスク依存の知識を用いた実時間音声認識による野球実況中継の構造化”, 日本音響学会, 平成 16 年度春季研究発表会, 1-8-15, pp.35-36 (2004)

くまの まさひと
熊野 雅仁 1990年, 立命館大・理工学部卒。同年, 龍谷大学理工学部実験助手, 現在に至る。映像文法を基盤とした映像編集やデジタルコンテンツの生成, 生成支援技術に関する研究に従事。正会員。

ありき やすお
有木 康雄 1974年, 京大・工・情報卒。1976年, 同大学院修士課程了。1979年, 同大学院博士課程了。1980年, 京大・工・情報助手。1990年, 龍谷大学・理工・電子情報助教授, 1992年, 同教授。2003年, 神戸大学・工・教授, 現在に至る。工博。1987~1990年, エディンバラ大学客員研究員。画像処理, 音声情報処理に従事。正会員。

つかだ きよし
塚田 清志 1979年, 阪大・工・通信卒。同年(株)毎日放送入社。放送技術に従事。2001年, デジタル計画部長。データ放送・インターネット等デジタルメディアに関する研究。放送のデジタル化全般の開発業務に従事。映像情報メディア学会, サイバー関西プロジェクト委員。

Automatic Extraction of PC Scenes Based on Feature Mining for Real Time Delivery of Baseball Highlight Scenes

Masahito Kumano Yasuo ARIKI
Kiyoshi TSUKADA

Faculty of Science and Technology, Ryukoku University
Mainichi Broadcasting System, Inc. 17-1 Chayamachi, Kita-ku, Osaka, 530-0013 Japan

Abstract

To replay baseball highlight scenes in live broadcasts to baseball fans outside, image processing, such as analysis, meta information extraction, and automatic editing, has to be performed in real time. This paper proposes high-speed image processing that automatically extracts PC (Pitcher and Catcher) scenes from live broadcasts of a baseball game in real time using a feature mining technique as a part of baseball highlight scene delivery. This method achieves an F-measure of 97.2% and a processing speed 30 times faster than actual time.

Keyword: baseball live broadcasting, highlight scene, real-time, feature selection, mining

野球中継中のハイライトシーン映像を外出中の野球ファンへ自動配信することを目的とした場合、リアルタイム内で処理可能なハイライトシーン検出法が必要となる。この問題を解決するためには、映像のデジタル化の処理を始めとして、映像解析・メタ情報の付与、ハイライトシーンの自動編集を自動的にリアルタイム内でオンライン処理することが必要となる。本研究では、音声認識技術と高速な映像解析技術を統合することにより、野球中継映像中のハイライトシーンをリアルタイム内で自動的に抽出するシステムの部分システムとして、PC (Pitcher and Catcher) シーンを高速・高精度に抽出する手法について報告する。実験を行った結果、実時間の 1/30 内で処理が可能であり、Fmeasureにより、97.2%の精度を達成した。