コンピュータビジョンと
イメージメディア 106－16
（1997. 7. 25）

動作スポッティングによるシーン検索

舘山公一　　川嶋稔夫　　青木由直

tateyan@huie.hokudai.ac.jp

北海道大学 工学部
〒060　　札幌市北区北13条西8丁目

あらまし　本研究では、野球番組からのユーザ要求シーン検索システムの構築を試みている。本稿ではそのために、野球ビデオデータから、検索時のキーとなるような選手の動き情報やテロップ文字情報を抽出し、不必要な場面を省いてダイジェスト化された各シーンにインデクス情報として付加して、検索のためのデータベースを構築する手法について提案する。選手の動作情報は、連続ＤＰマッチングによる認識・時間区間の同定を行なう。また、実際に野球ビデオデータに対してインデクシングを行う実験を行ない、動作認識に関しては 84 %の認識率で行なえ、本稿で提案したインデクス情報抽出手法の有効性についても示した。

キーワード　画像検索、動作スポッティング、インデクシング

Scene Retrieval by Action Spotting

Kouichi Tateyama,Toshio Kawashima and Yoshinao Aoki

tateyan@huie.hokudai.ac.jp

Hokkaido University,Nishi 8,Kita 13,Kita-ku,Sapporo

Abstract　We propose a method to retrieve scenes from a baseball video on user's demand. In the system, a video-stream is, first, divided into basic segments which contain a single pitching action. The location of player's action is spotted with continuous-DP matching from the segment, and the result of the play is classified. Captions seen in images are also extracted as keys for scene retrieval. A long sequence of game is skimmed based on the information, and stored as a video digest to be retrieved. The method is applied to tens of at bats, and the 84 % of segments are correctly classified. Experimental results show the indexing method is an efficient way to retrieve relevant scenes.

key words　image retrieval, action spotting, indexing

1 はじめに

インターネットなどの新たな画像通信サービスの普及に伴い、個人レベルで、大量の動画像を、しかもオンラインで得られるようになった現状を考えると、個人で所有する動画像はますます広範囲化・肥大化するであろうと考えられる。また、テレビ放送に関してもケーブルテレビ等の普及に伴ってチャンネル数が増加し、受信側においても大量の映像を効率良く扱うことが要求される。そのような大量の動画像ハンドリングのためには、必要な時に必要な画像のみを利用できるような動画像データベースの構築が必要であると考えられる。このようなデータベースには、検索時のキーとなるようなインデクス情報を付加しておくことが必要となるが、その労力負担は軽視できないものがある。

映像の内容をコンピュータによって理解するために解析することを一般的に“映像ブラウジング”という。現在の所、一般的なすべての画像について映像ブラウジングを行なうことができる手法は実現していない。しかし、テレビのいろいろな番組の各々に対して、行なうための研究は進められている。具体的にはドラマシーンから人物とその動きなどを抽出する研究 [3] や、ニュース映像を文字・音声・画像情報を基に索引付けする研究 [4]、スポーツでは大相撲 [5] やサッカー [6]、バレーボール [7] などのシーン解析に関する研究がある。

ＴＶのスポーツ番組は放送の長さの割に、視聴者にとっての重要な箇所は限られていることが多い。非プレー時の“つなぎ”である、ベンチや選手の情報や、あまり試合を左右しないプレーなどが存在するためである。これを要約したものとしてダイジェスト番組があるが、点数が入った場面を中心に重要シーンをまとめている。もしスポーツのビデオ記録にインデクスを付加することができれば、ダイジェスト版と同様の短時間のブラウジングが可能になる。

スポーツ番組では一般的に、得点の変化に関与するシーンが重要シーンである。よって、シーン解析もそのシーンの同定を中心に行なわれる。得点の変化の要因はスポーツによって異なる。例えばサッカーやバレーボールなどの球技では、ボールがある場所に入る、到達することなどが得点に繋がる。よって、フィールド内のボールの位置を検出することによって意味推定を行なうことができる。特にバレーボールやテニスといった、フィールドがネットなどで区切られている場合にはボールがどちらのサイドにあるかを検出するだけで大まかな解析が可能である [7]。また、球技以外のスポーツでは、人のある特定の状態によって得点が変化する。相撲の土俵から外に出された状態や、ボクシングのダウンなどがこれに当たる。よって、解析のためには人間の位置追跡や、状態認識が必要である。

本研究では、野球を題材に選んでシーン解析、インデクシングを行い、スポーツ中継のビデオデータからユーザのリクエストシーンを検索できるシステムの構築を目指す。野球は、球技であるため、ボールの状態を追跡することが重要であるが、得点の変化はランナーのホームインによるため、ボールと選手の両方の状態認識が必要であるという、上記の２パターンの中間に位置するスポーツであると考えられる。しかし、シーン構造的に扱い易い利点もあり (このことに関しては２章で述べる)、比較的簡単な手法で実現することが可能であると考える。実際にはボールが小さいため、困難と思われるボールの追跡に関しても、ボールの位置変化に関係する画像内容の特徴 (打者のバッティング動作など) や、ボールを追跡したカメラの動きなどの単純な特徴によってボールの位置変化を検出することとする。また、ランナーの位置変化の認識を、テロップ情報より実現する。

本稿では、システムの方針とインデクシング時にアンカーとなる画像を検出するための動作スポッティング手法について述べる。２章では、インデクシングのためのアプローチとして、野球のシーン構造、ユーザの要求とそのために必要なインデクス情報について検討する。３章ではシステム全体の方針、４章では実際にインデクシングを行なう手法について説明し、実際にそれに基づいて行なった実検の結果と考察を５章で提示する。最後に６章では今後の課題について検討する。

2 インデクシングのためのアプローチ

2.1 野球におけるシーンの構造と特徴

野球映像では、以下で示すような、意味推定において有利なシーンの構造と特徴がある。

- シーンサイクルの存在

  ある程度決まったシーンのサイクルが存在する。最も基本となるものは、投球1球ごとのサイクルであり、投球シーンに始まり、打者動作・守備シーンを経て、走者の安定状態で終了する。この1球ごとの投球や、バッターの切り替わり、イニングの切り替わりを単位としてシーン構成を考えることによって、時間方向のセグメンテーションが行なえる。また、バッターの切り替わりを単位とした分割で、選手(打者)の特定を行なうことができる。

- カメラ制約

  基本的に、数台の固定カメラからの映像を切り替えることで番組が構成される。主だったものとしては、バッターのアップ(横方向)、ピッチャーのアップ(正面)、投球シーン(バックスタンドカメラからホームを捉えたもの)、守備シーン(フェンス側から内野・外野を捉えたもの)がある。守備シーン以外の固定カメラの映像はパン・チルトの影響があるものの、大局的性質は安定しており、被写体の特徴抽出が容易となり、カットの識別、意味推定を行ないやすい。後の投球シーンにおける動作スポッティングが、比較的単純な手法で行なえるのも、この制約があるからである。

しかし、フィールドの狭いバレーやテニスとは異なり、扱い難い点もまた存在する。それは、一連のプレーが数カットに跨っているということである。投球シーン、走塁のアップ、守備シーンなどを含めて一つのヒットシーンが形成されるわけで、複数のカットの意味推定が全てうまく行なわれなければならない難しさがある。

2.2 ユーザ要求

ダイジェストデータベースを構築するためには、ユーザの要求シーンを検索できるようなインデクス情報を抽出することが必要である。そのために、ユーザが一般的に要求するシーンとはどういうものなのか、またそのシーンを特定するためにはどのような解析を行なえば良いのか検討する。

- 打席結果の特定

  最も主たる要求は打席結果によるものである。「ヒットを打ったシーン」や「三振したシーン」といった要求がこれに当てはまる。本稿では上記の2つに「ホームランを打ったシーン」を加えた3種の要求を考えることとする。これらのシーンを特定するためには、(1) その打者の打席が終了したことの認識、(2) バットスィング有無の認識、(3) ランナーや得点の変化の認識、が必要であると考える。ここで(1)に関しては、バッターの切り替わりによってボールカウントがリセットされることを利用して、ボールカウントのテロップ情報から認識が可能である。(3)についても同様にテロップから認識できる。(2)については簡単なジェスチャ認識が必要となる。

- 選手の特定

  「○○チームの××選手の打席」という要求もまた、考えられ、そのために、チーム・選手の認識が必要である。しかし、チームはテロップ情報より簡単に認識できるが、選手の特定には顔や背番号の認証の問題、もしくはバッターのアップ時に表示されるテロップの選手名認識問題となるが、今回は行なわないこととする。

2.3 ダイジェスト化

また、保存の観点から言えば、不必要な動画像データは極力省かねばならない。検索結果として必要な動画のみをフレームシーケンスから切り出し、インデクシング情報を付加して保存することが望ましい。同じ投球シーンの中でもセットポジションに入る前や、投球するまでは必要なく、データベースには投球し始めるフレーム以降のみで十分と考える。そのために、インデクシング時に並行して抽出しておく情報について検討しておく。

- シーン切り出し始点

  全てのプレーの起点が投球であることを考えると、ダイジェストシーンの始点も投球動作

時が望ましいと考える。よって、その切り出し始点の同定には、投球動作の認識・時間区間同定が必要である。

- シーン切り出し終点

  プレーの終点は基本的には「ランナーやボールの状態が安定した時点」と考えられるが、行なわれたプレーによってそのシーンは異なる。それら全てに対応できる手法は目下検討中であるため、今回は、多少不必要なシーンが含まれてしまうが、基本シーンより後の3カットを残すこととする(カット検出が必要)。

3 方針

システムをインデクシング部分と検索部分に分ける(図1)。インデクシング部分では、ユーザの一般的な検索要求に関連すると思われるシーンを同定しておく。これらのシーンにはヒットや空振りをしたシーンなどが含まれる。これらのシーンに、試合のイニング、カウント、スコアなどが付加され、野球ダイジェストのためのデータベースとして蓄積される。検索部分では、入力されたユーザの検索要求に対して、インデクシング結果をもとにデータベースより検索し、対応する動画像データを出力する。

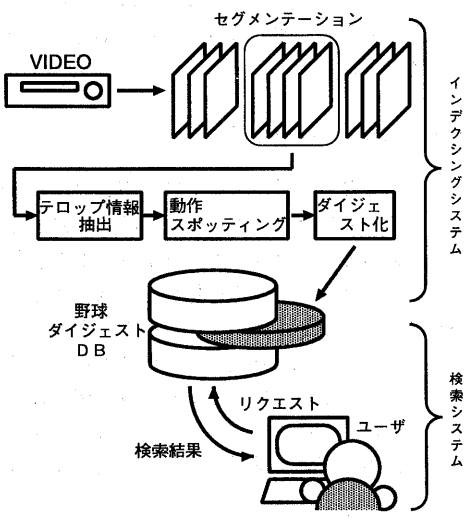

図 1: システム概要図

3.1 インデクシング部分

映像のインデクシングを以下のようなステップで行う(図2)。

1. **時間方向のセグメンテーションにより映像内容の時間的な記述単位を設定する。**

   2章で述べたシーンサイクルを利用して、1球ごとのセグメンテーションを行なう。

2. **各セグメントに対してインデクス情報を抽出する。**

   2章の検討結果より、以下の情報を抽出することとする。

   - テロップ情報(ボールカウント、アウトカウント、ランナー位置、得点)
   - 投球動作・スィング動作の有無、その時間区間

3. **抽出されたインデクスを画像列に付加する。**

   動画像データベースの構築のため、ダイジェスト化された画像列にインデクス情報を付加させたものを蓄積することとする。

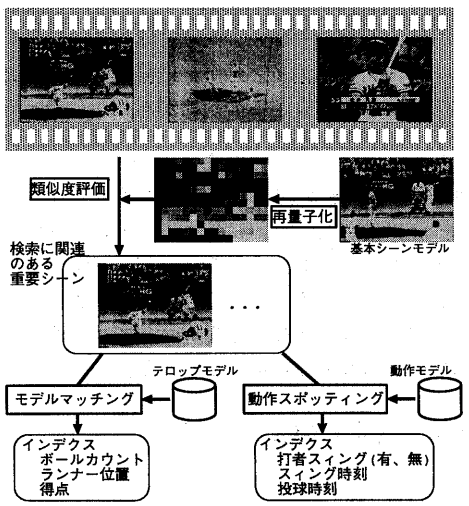

図 2: インデクシング

3.2 検索部分

オフライン部分で保存、インデクシングされた動画像列より、入力されたリクエストに対応したシーンを検索する。ユーザの要求とインデクスとの対応付けは表1のようなものとする。

	ヒット	ホームラン	三振
Swing	あり	あり	あり
Ball	リセット	リセット	リセット
Out	変化なし	変化なし	一つ増加
Runner	変化	リセット	変化なし
Start	投球動作	投球動作	投球動作
End	走塁終了	ホームイン	スィング後

※ Swing:打者スィング動作、Ball:ボールカウント、Out:アウトカウント、Runner:ランナー位置、Start:切り出し始点、End:切り出し終点

表 1: 検索時のキーインデクス

4 インデクシング手法

4.1 セグメンテーション

まず、入力動画像列を投球1球ごとにセグメンテーションする。そこで、投球シーンを基本シーンとして、基本シーンから次の基本シーンまでをひとまとまりとして分割する (図3)。

実際には動画像列に対し、基本シーンのサンプル画像との類似度を計算して基本シーンを検出、その次に現われる基本シーンまでの区間をセグメントとする。類似度は近傍平均により 16x12 画素に再量子化した画素値を要素とするベクトルの内積として求める。一般に基本シーンの映像はバックスタンドに固定したカメラで撮影されるので、セグメントはこのような単純な手法でも安定に検出される。

4.2 テロップ情報抽出

投球シーンに出現するテロップの情報を、以下のステップで抽出する。

1. 輝度値の時空間変化を特徴としてフレームシーケンスからテロップ出現フレームを検出する。

   テロップは瞬間に出現し、出現後は一定時間静止して表示されるという特徴を持つ。そこで、前フレームとの輝度変化が大きく、また後フレームとの輝度変化が小さいピクセルが多いフレームをテロップ出現フレームとする。また、投球シーンのテロップの場合、出現するのは画面の左上と右下の領域であるため、対象領域を狭めることができる。

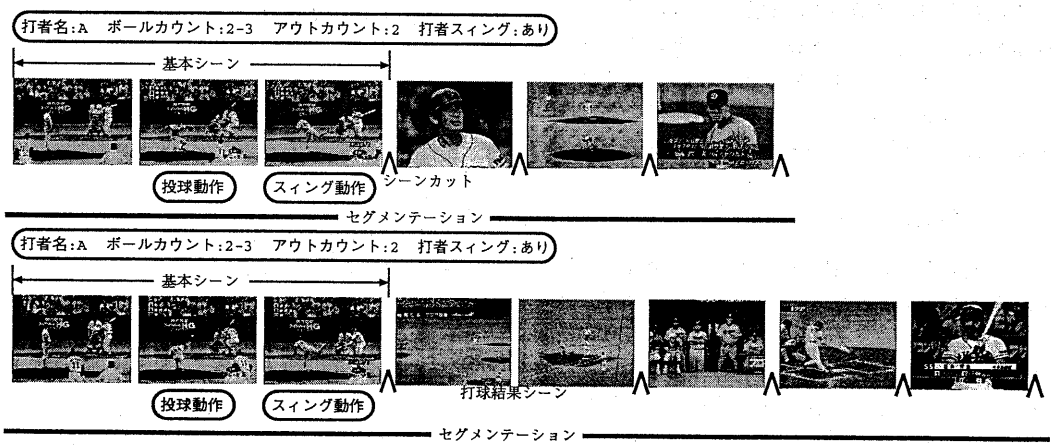

図 3: セグメンテーション

2. テロップ出現フレームからテロップ領域を抽出する。

   テロップは周辺との輝度コントラストが高く、輝度による2値化を行なうことで連結したストロークを分割することなく、文字と背景の大部分を分離した形で領域を形成することができる[2]。しかし、本研究で使用したデータの場合、画面右下の領域に関しては背景と文字との輝度にあまり差がないため、1.で利用した、前との時間差分が大きく、後との時間差分が小さいピクセルをテロップ領域とする。

3. テロップモデルとのマッチングを行ない、テロップを認識する。

   テロップは各局やシーズンによって異なるものが使われる。残念ながら、それらのバリエーションに対応した認識は困難であると考えられる。そこで、どのパターンのテロップなのかを事前に与えることにする。テロップパターンが限定されれば、配置・位置などを既知とすることができ、テロップの認識は簡単なモデルマッチングによって行なうことができる。ボールカウントのボール、得点の数字、ランナーの位置を示すダイヤモンドのモデルを用いてそれぞれの認識を行なう。

4.3 動作スポッティング

投球および打球動作の認識及びその時間区間の同定を行なうために、連続ＤＰマッチングを用いたスポッティング認識を行なう。この方法ではモデルとなる動作時系列画像と入力時系列画像で特徴量の推移を比較し、距離を計算して認識する。

4.3.1 特徴抽出

特徴量としては、指定した領域での濃度値の時間差分量を要素としたベクトルを用いる(図4)。動作スポッティングを行なうバックスタンドカメラからのシーンでは、投手や打者の位置は大きくは変わらないため、特徴量を求める領域は固定することができる。各領域は、それぞれ選手の動きによる時間差分が大きく現れるような位置(バットの起動上、投球時の腕の位置など)を指定しておく。時間差分は衣服や明るさの変化にもロバストになるように、閾値によって1と0に二値化する。領域内の時間差分の総量を面積で割った値をベクトル要素とする。

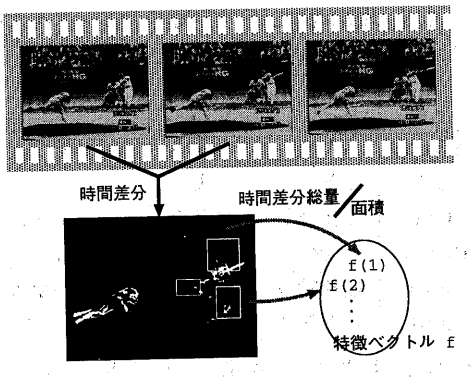

図 4: 特徴ベクトル抽出

4.3.2 スポッティング認識

特徴ベクトルの距離を連続ＤＰの入力とし、累積距離の低下検出により、モデル認識・時間区間同定を行なう[1]。連続ＤＰは入力フレームとモデルパターンの距離を $d(t,\tau)$ として、累積距離 $S(t,\tau)$ を以下のように定義する。
初期条件：

$$S(-1,\tau) = S(0,\tau) = \infty. \ (1 \leq \tau \leq T)$$

漸化式 $(1 \leq t)$：

$$S(t,1) = 3d(t,1).$$

$$S(t,2) = min \begin{cases} S(t-2,1) + 2d(t-1,2) + d(t,2) \\ S(t-1,1) + 3d(t,2) \\ S(t,1) + 3d(t,2) \end{cases}$$

$$S(t,\tau) = min \begin{cases} S(t-2,\tau-1) + 2d(t-1,\tau) + d(t,\tau) \\ S(t-1,\tau-1) + 3d(t,\tau) \\ S(t-1,\tau-2) + 3d(t,\tau-1) + 3d(t,\tau) \end{cases}$$

※ t: 時刻 、τ: モデルパターンの長さに対応するパラメータ

5 実験

同一の試合中の異なる幾人かの選手の打席の基本シーンに対して投球動作とスィング動作のスポッティングを行なった。

図?? はある打席の一連の3球に対して動作スポッティングを行なった結果である。上2枚がそれぞれ投球動作・スィング動作のDPマッチングの出力値の推移を示している。3枚目が基本シーンとの類似度を示したもので、これら3枚のグラフ内の縦線は基本シーンとその他のシーンとの境界線を求めた結果を表したもので、横線は認識時の閾値である。4枚目はシーンカットの検出のために求められた、前フレームとの輝度差分の大きさを示している。

この打席では、1球目は見送り、2・3球目は空振りであったが、認識結果も同じものとなった。また、グラフをみれば、スィング動作の有無によってDPの累積距離に差が大きく現れることがわかり、本稿の特徴抽出による認識手法が妥当であると考えられる。

また、それぞれ動作認識されたシーンにおいて、DPの値が最も小さくなったフレーム画像を右に示す。それぞれ、投球動作・スィング動作の終了時フレームとして妥当であると見られ、時間区間の同定を正しく行えることが確かめられた。

また、図?? はスィング動作とバント動作、投球動作と牽制球動作とのDPマッチングの比較を行なったものである。それぞれ、同選手の打席を扱ったものだが、スィングとバントの差、投球と牽制球の差が見られ、良好な認識結果が得られた。

表??は無作為に抽出された千から2千フレーム程度のシーン(ただし、打席が必ず含まれるもの、CMなども含む)に対して投球・バットスィング動作の認識実験した結果を選手ごとに集計し、一覧にしたものである。認識率を計算すると、スィング動作認識では84.6%、投球動作認識では95.1%であった。スィングがあった打席を「スィングなし」と誤認識される結果が目立つが、A2やB3は、その打者のバットの色が黒っぽく、背景との区別がほとんどつかないためである。この問題に関しては、現在の手法では解決できないため、別のアプローチを考える必要がある。

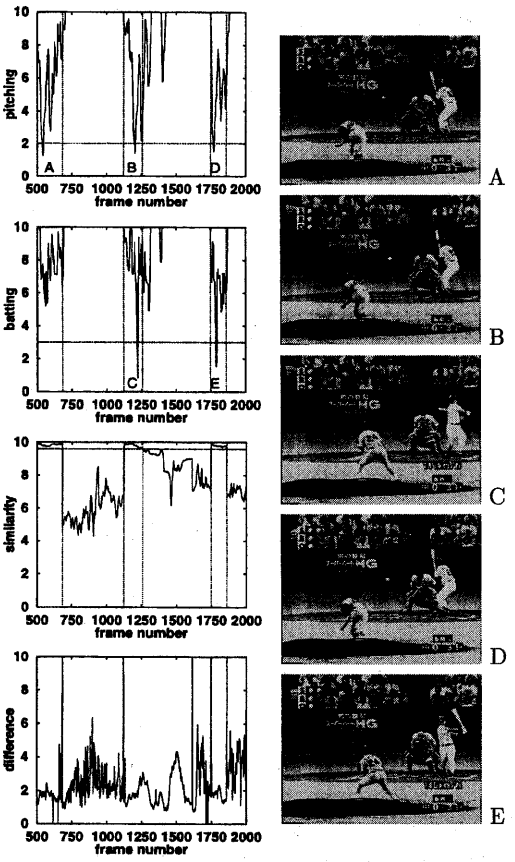

図 5: 実験結果1

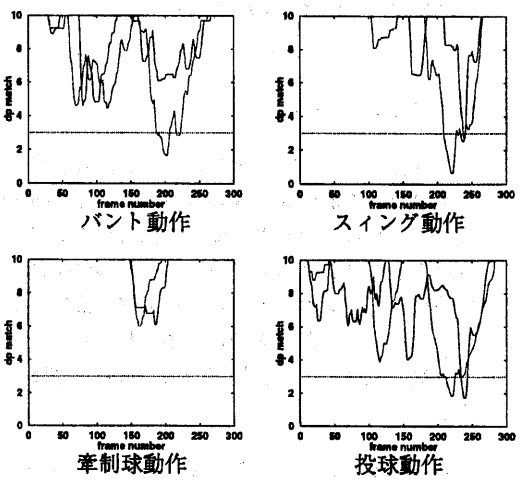

図 6: 実験結果2

選手	スィング認識数	見送り認識数	スィング誤認識数	見送り誤認識数
A 1	1	1	2	0
A 2	0	1	1	0
A 3	1	2	0	0
A 4	2	2	0	0
A 5	2	1	0	0
A 6	2	3	0	0
A 7	3	3	0	0
B 1	1	2	1	0
B 2	1	0	0	0
B 3	0	1	2	0
B 4	1	1	0	0
B 5	1	1	0	0

選手	投球認識数	投球誤認識数
A 1	12	0
A 2	1	0
B 1	11	2
B 2	4	0
B 3	11	0

表 2: 動作認識結果一覧

動作スポッティング以外の面では、テロップ情報の抽出において、得点の数字認識などで誤認識が少し見られるなどの問題点が生じたが、その他のテロップ情報の抽出やセグメンテーションなどの処理課程では概ね良好な結果が得られた。

6 終わりに

本論文では、野球のビデオデータから重要シーンを検索できるシステムの構築を目指し、検索時にキーとなるインデクス情報の抽出法を提案、それによるダイジェストデータベースの構築について検討した。また、実験によって本手法による動作スポッティングの有効性が示された。

今後の課題として、以下のような点が挙げられる。

- ダイジェストシーンの切り出しの終点の同定

  ランナーやボールが安定状態であることを、単純な手法で認識できないか。

- 選手の同定

  打者のアップで表示されるテロップより選手名を認識できないか。もしくは背番号を認識できないだろうか。

- 動作認識・テロップ文字認識の安定性向上

  色のうす暗いバットを背景より分離する手法。打者の様々なスィングのバリエーションにも対応できるような新しい認識法との併用。

- 実際のデータベース、システムの検索部分の構築

参考文献

[1] 高橋 勝彦, 関 進, 小島 浩, 岡 隆一, “ジェスチャー動画像のスポッティング認識”, 信学論(D-2),Vol.J77-D-2,No.8,pp.1552-1561,1994

[2] 桑野 秀豪, 倉掛 正治, 小高 和己, “映像データ検索のためのテロップ文字抽出法”, 信学技報 PRMU96-98,1996

[3] 柳沼 良知, 影山 誠, 坂内 正夫, “2 段階のモデルを用いたビデオ画像からの人物とその動きの抽出方式”, テレビジョン学会技術報告 Vol.17,No.8,pp.1-6,1993

[4] 有木 康雄, 杉山 善明, 石川 則之, 寺西 俊裕, 櫻井 光康, “ニュース映像中の記事に対する音声・文字・映像を用いた索引付けと分類”, 信学技報 PRMU96-97,1996

[5] 田淵 仁浩, 村岡 洋一, “テレビ放送の知的録画-不完全質問処理に基づく個人用大相撲ダイジェスト作成システム-”, 情報メディア 4-5,1991

[6] 皆川 信司, 川嶋 稔夫, 青木 由直, “サッカー中継のシーン解析”, 電気通信学会春季全国大会 D-531,1992

[7] 北 健志, 小沢 慎治, “バレーボール中継におけるシーン解析”, 信学技報 IE95-155,PRU95-242,1996