

動作インデックスを用いた映像の自動注釈付けと その柔軟な内容検索への応用

宮森 恒†

† 通信総合研究所けいはんな情報通信融合研究センター
〒619-0289 京都府相楽郡精華町光台 2-2-2

E-mail: †miya@crl.go.jp

あらまし 本稿では、動作インデックスによる映像の自動注釈付けと、それを利用した柔軟な内容検索システムについて報告する。応用例としてテニスを考える。目的は、スマッシュなど一選手の単独動作だけでなく、ネットダッシュするといった複数選手・複数動作からなる、より一般的な特定動作を検索することである。まず、入力映像・音響データから、コート・ネットライン、選手・ボール位置、選手によるボール打撃時刻を抽出し、これらを総合的に評価してフォアハンドスイングなど選手基本動作インデックスを出力する。次に、予め定義済みの動作成立条件とすでに出力された基本動作・一般動作インデックスを評価し、新たに一般動作インデックスを出力する。実験の結果、動作成立条件の再定義と更新により、さまざまな一般動作を柔軟に追加・検索できることを確認した。

キーワード 内容検索，インデキシング，自動注釈付け，動作識別，行動解析，認識

Automatic video annotation using action indices and its application to flexible content-based retrieval

Hisashi MIYAMORI†

† Keihanna Human Info-Communication Research Center, Communications Research Laboratory
2-2-2, Hikari-dai, Seika-cho, Souraku-gun, Kyoto, 619-0289 Japan

E-mail: †miya@crl.go.jp

Abstract This paper presents automatic video annotation method based on action indices and its application to flexible content-based retrieval system. Our test domain is tennis. The goal here is to retrieve not only simple actions by single player, such as smashing, but also more general actions composed of several players and actions, such as "net-dashing". First, input video and audio data are analyzed to extract court-net lines, positions of players and the ball, and the moments a player hits the ball. These data are totally evaluated to output basic action indices by each player, such as forehand swings. Then, the predefined conditions to satisfy a certain general action, and the already generated basic and general action indices are evaluated to output new indices of the corresponding general action. Experimental results show that redefinition and updates of conditions for a specific general action enable the system to append and retrieve various kinds of actions flexibly.

Key words content-based retrieval, indexing, automatic annotation, action identification, behavior analysis, recognition

1. はじめに

近年さまざまな分野においてユーザが利用できるデジタル映像情報が増加の一途をたどっている。インターネット社会の発展とともに、コンピュータ機器・通信環境・インタフェースが高速化・広帯域化し、様々な映像データが大量に随所に蓄積されるようになると、これら膨大な情報にアクセスし映像中の自分の見たい箇所を素早く探し出す内容検索技術は今後ますます重要になると予想される。

内容検索のための注釈付けに関する従来研究は、大きく2つのアプローチに分類できる。

一つは、映像から、色やテクスチャ、カメラ操作[1]-[4]、人物の顔、キャプション、音の種別や大きさ、書き起こしデータに対するTF-IDF[5]-[8]、等々、さまざまなメディアから得られる低レベルの特徴量を自動解析によって抽出し、これらをインデックスとして特定場面に対応づけ内容検索を実現するアプローチである。ドラマやニュース映像など「シナリオ」があるものや例示画像による類似画像検索への適用例が多く報告されている。これらは基本的に低レベル特徴に基づいているため、汎用映像に広く適用できる利点がある一方、識別できる場面の種類がその物理特徴が表現できる範囲に限定され、映像中で実際に何が起こっているかといった高レベルでの内容表現が困難という問題点がある。

他方は、インデックス生成に必要なデータ入力を手作業で行うことを大幅に許容し、内容に関連したインデックスを目的に応じて扱いやすいように設計し、これを生成・利用するものである[9]-[11]。一度手入力できれば、文脈の流れや個々の特徴的な場面を表現するインデックスが利用可能なため、目的に応じた特定場面の検索が実現できる。しかし、多くの場合、適用範囲が限定される点、人手によるインデックス入力の均一性の維持、人手によるコストや負荷軽減のための工夫、インデックス抽出の自動化が課題となる。

一方、ドメイン知識等を用いて特定コンテンツのインデキシングをできるだけ自動化し、結果として実現されるシーン検索もより実用的なレベルに近づけようとする研究が近年進められている[12]-[14]。これらの研究は、いずれもサッカーやバスケットボールといった特定分野に依存したドメイン知識を利用することによって、映像中の対象物の軌跡、相対

位置、および、その時間変化等を抽出し、これらを分析することによって、検索イベントとの対応づけを行おうとする方法である。しかし、従来は、いずれも「位置情報」に基づいたインデックスを用いており、対象物の「動作」をインデックスとして明確に表現しているわけではない。従って、高レベルのイベント内容を表現し、関連づけることが容易でなく、検索可能なイベント種類も限られるという問題点があった。

本稿では、映像中の人物などが行う基本的な「動作」を明示的に表現するインデックスを自動生成し、これに基づき一般的な特定場面を柔軟に内容検索するシステムについて報告する。本方式は、基本動作インデックスを計算機で自動識別・生成する点と、一般動作インデックスを、すでに出力された基本動作・一般動作インデックスと、予め定義済みの動作成立条件から評価することで、随時出力する点が特徴である。

本稿は以下のように構成される。2節では、システム概要について述べる。3節では、映像・音響データの解析による基本動作インデックスの生成について説明する。4節では、動作成立条件の定義と更新による一般動作インデックス生成について述べる。5節では、いくつかの実験結果と考察を示し、6節で結論を述べる。

2. システム概要

図1にシステムの処理概要を示す。本稿におけるテスト対象はテニスとする。試合全体のテニス映像から、スマッシュなどの一選手の単独動作だけでなく、ネットダッシュするといった複数選手・複数動作からなる、より一般的な特定動作を検索することが我々の目的である。

一般に、テニス映像の場合、選手や審判、観客のアップとなるショットがしばしば含まれているが、最も典型的な画面構成はコート斜め上から縦方向にコートが写るようなアングルで撮影された映像である。本稿では、テニスコートなどの背景色を利用した選択手法[14]などにより、選手や審判のアップとなるショットとコート全体を撮影した最も典型的なショットが予め分割されており、後者のショット映像が入力として利用可能であるものと仮定する。

さて、本システムにおける動作インデックス生成手順を図1に示す。

まず、入力映像からコートおよびネットラインを

コートモデルとハフ変換を利用して抽出する．選手位置はコート・ネットライン除去後の2値画像において重なりが最大となる領域を特定することで追跡する．ボール位置は，各選手位置との距離に応じて検出モードと追跡モードを切り替えることで特定する．選手によるボール打撃時刻は，映像データに同期して得られる音響データの周波数領域におけるテンプレートマッチングで抽出する．最後に，ボール打撃時刻における選手とボールの位置関係から，フォアハンドスイング等の選手基本動作を識別し，選手idや基本動作クラスを含む動作インデックスとして出力する．

一般動作については，各動作名とそれが成り立つための諸条件を定義したルール集合が，データベースとは別に予め人手によって用意されている．一般動作のインデックスは，すでに出力されている基本動作や他の一般動作インデックスの間で成り立つ特定の時空間関係を評価することで出力される．

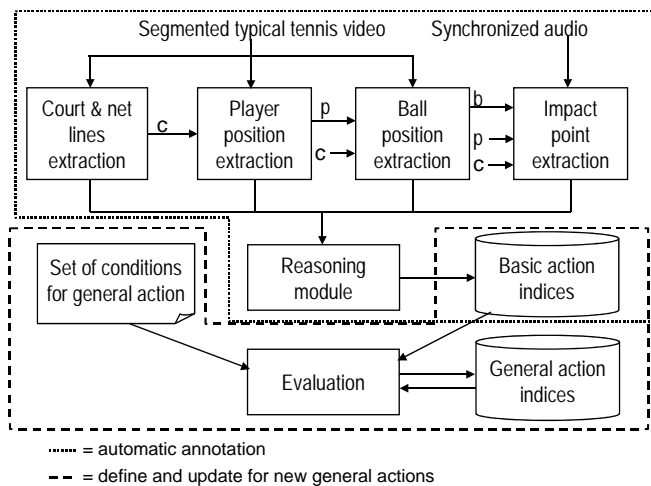


図1. 動作インデックスの生成手順

3. 基本動作インデックスの生成

3.1 コート・ネットラインの抽出

コートおよびネットラインの抽出では，ドメイン知識としてテニスコート仕様を利用する．図2に示すように，コート特徴点 Pc_1, \dots, Pc_{14} ，コートライン Lc_1, \dots, Lc_9 および，ネット特徴点 Pn_1, \dots, Pn_3 ，ネットライン Ln_1, Ln_2 を定め，コートおよびネットモデルとして参照する．

抽出は，コートライン，ネットラインの順に行う．

まず，コートラインの抽出について．図2の各コートラインはその両端のコート特徴点を検出することによって決定する．

1. 時刻 $t = 0$ においては，初期特徴点 $Pc(0)$ を入力として与える．次に， $Pc(0)$ で決まるコートライン $Lc(0)$ について，各ラインをハフ平面に変換する．ハフ平面上における各ピーク点を中心に大きさ w_{th} ， w_{ro} の検出用窓 $Wc(0)$ を用意する．
2. 時刻 $t = t$ においては，まず，原画像の2値化画像 $B(t)$ とコートライン $Lc(t-1)$ の近傍領域との AND をとり，コート近傍のみからなる2値画像（コートライン2値画像と呼ぶ） $Bc(t)$ を生成する．
3. これを各ライン毎にハフ変換し，各検出用窓 $Wc(t-1)$ で制限される範囲でピーク検出を行い，特徴点 $Pc(t)$ を更新する．
4. 再び，コートライン $Lc(t)$ をハフ変換し，検出窓 $Wc(t)$ も更新する．ステップ2へ戻る．

なお，パニングなどによってコート特徴点が画面から外れた場合も，コート中央の特徴点 $Pc_i(t)$ ($i=9,10,12,13$ or $10,11,13,14$) は画面内に常に写っていることを仮定し，連結知識を用いて画面外の点を推定して更新する．同様の理由で，いくつかの初期特徴点は省略してもよい．

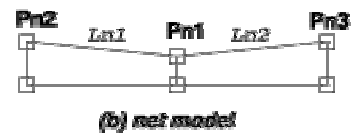
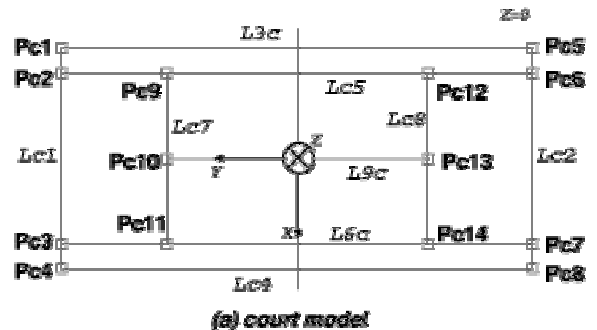


図2. コート・ネットモデル

次に，ネットラインの抽出について．

1. 時刻 $t = 0$ での初期特徴点 $Pn(0)$ を入力として与え，コートラインと同様に，各ライン毎にネットライン $Ln(0)$ ，検出用窓 $Wn(0)$ を用意する．
2. 時刻 $t = t$ においては，原画像の2値化画像からコートライン2値画像を除いた画像 $Bn(t) = B(t) - Bc(t)$ をネットライン2値画像として生成し，これを用いてハフ変換，検

出窓範囲内でピーク検出を行い，特徴点 $Pn(t)$ を更新する．

図3にコート・ネットライン c の抽出結果を示す．特徴点が画面外に出た場合も連結知識を用いた位置推定により，安定した検出が可能である．

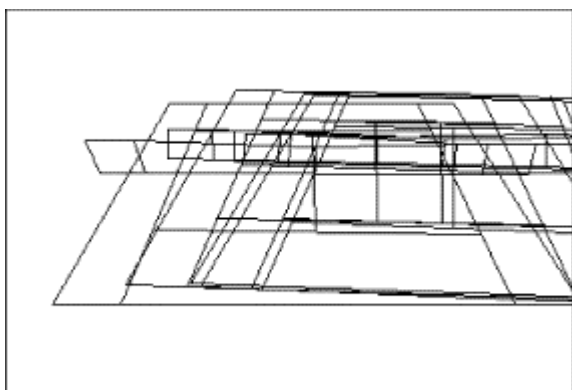


図3．コート・ネットラインの抽出

3.2 選手位置の抽出

選手領域の抽出は次の手順で行う．

1. 時刻 $t = t$ において前後 s フレーム離れた画像との差分を求め適当な閾値で 2 値画像 $B_1(t), B_2(t)$ を作る．

$$B_1(t) = \text{BIN}(I(t) - I(t - s))$$

$$B_2(t) = \text{BIN}(I(t + s) - I(t))$$

2. これら 2 つの差分画像の AND 演算を行った結果得られる 2 値画像を $B_{diff}(t)$ とする．
3. 選手ユニフォーム等の代表色に対応した色クラスタを予め用意し，時刻 $t = t$ における画像 $I(t)$ の点で，その色クラスタ内に含まれるものを 1 として，2 値画像 $B_{label}(t)$ とする．
4. $B_{diff}(t), B_{label}(t)$ からコート・ネットラインを消去する．選手領域との重なり部分を除去したと考えられる領域については拡大縮小処理によって領域を補完する．得られた 2 つの画像の OR 演算を行い 2 値画像 $B(t)$ とする．

$$B'_{diff}(t) = B_{diff}(t) - Bc(t) - Bn(t)$$

$$B'_{label}(t) = B_{label}(t) - Bc(t) - Bn(t)$$

$$B(t) = B'_{diff}(t) \cup B'_{label}(t)$$

5. 2 値画像 $B(t)$ 内で連結領域のラベリングを行い，ノイズの影響を避けるため，これを数フレーム観測し，コート近傍内にある一定以上の面積をもつ領域を選手初期位置とする．

6. 時刻 $t = t$ における一定以上の面積をもつ領域のうち，時刻 $t = t - 1$ の選手領域の近傍に存在し，かつ，面積差が最も小さい領域を時刻 $t = t$ における選手領域 p と判定する．

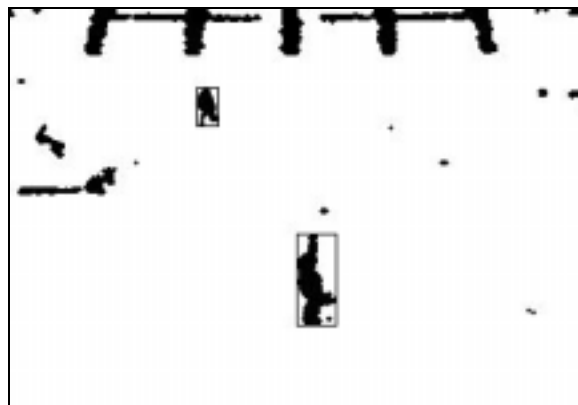


図4．選手領域の検出

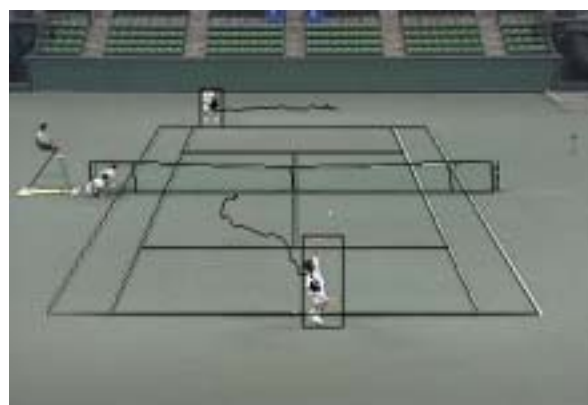


図5．選手位置の追跡

3.3 ボール位置の抽出

ボール追跡を行う際，以下の点が問題となる[15]．

- ボールの大きさが小さい．
- 遠近感や残像のため画像中のボールの大きさがフレームにより変化することがある．
- ボールがライン上やネットの上端にかかる場合にボールの位置確認が全くできないことがある．
- バウンドや投打によってボール軌道が急激に変化する．
- ボールが選手と重なったり隠されたりする場合がある．

よって，3.2 節で述べた選手位置の追跡手法をそのままボール追跡に用いるのは適当とはいえない．そこで，選手とボールの位置関係に応じて処理内容を切り替えることにする．ボール位置 b は以下のように抽出される．

1. 大きさ $b_x \times b_y$ のボールを含むテンプレート

$T_b(x, y)$ を予め作成する。

- 時刻 t において選手領域 p を消去した画像 I'_B 内の選手近傍領域において、テンプレート $T_b(x, y)$ に一致するボール候補位置を全て検出する(検出モード)。時刻 $t = t+1, t+2, \dots$ についても同様にボール候補を検出し、選手位置を中心に放射状に連続して検出されるボール候補を絞り込んでいく。最終的に残った1つのボール候補の組をその時区間のボール軌道と特定する。ステップ3へ。
- テンプレート $T_b(x, y)$ を用いてテンプレートマッチングを行う(追跡モード)。ただし、ボール軌道は微小時間ではほぼ直線とみなせると仮定し、前回検出された移動量をそのまま現フレームに加えた位置を予測中心として探索を行う。選手領域とボール位置との距離がある閾値より小さくなったら、ステップ2へ。そうでなければステップ3を繰り返す。

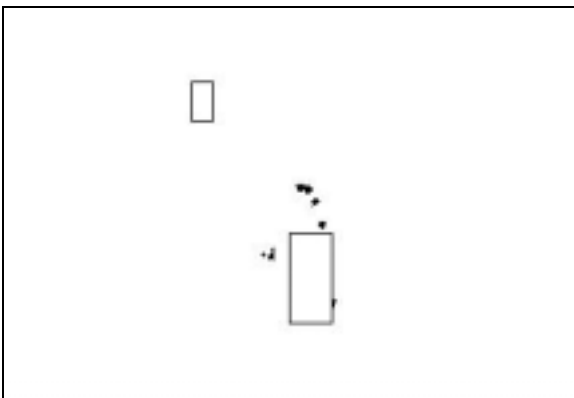


図6. ボール領域の検出

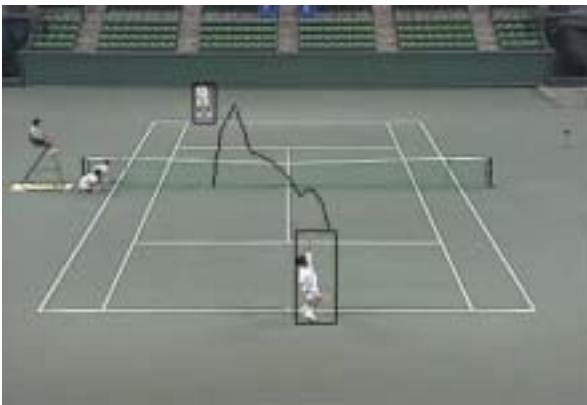


図7. ボール位置の追跡

追跡モードにおいてボール予測を用いない場合、ボールがネットあるいはラインの一部に重なり、そ

こから誤検出を回復できず、追跡が失敗することが多かった。ボール予測により、大きさの比較的小さいボールも追跡を続行することが可能となった。図6, 7に、ボール検出、および、ボール追跡の結果例をそれぞれ示す。

3.4 ボール打撃時刻の抽出

ボール打撃時刻の抽出は、映像に同期して得られる音響データの周波数領域におけるテンプレートマッチングにより行う。

1. テンプレートの選択

映像データよりテンプレートに適していると思われるボール打撃時の映像を手動で任意に数点選び出し、これと同期したボール打撃音の音響データを切り出す。音響データは分解能 16bit, サンプルレート 44.1kHz にてサンプリングされている。本稿で用いるテンプレートのデータサイズは、次の2つの理由より 2048 点 (0.046 秒) とした。

- 観測されるボール打撃音は全て 2048 点というデータサイズ内に収まる。
- FFT 処理を行うため、2 のべき乗のデータ点数が望ましい。

2. マッチングを行う周波数範囲の選択

テンプレートに対して FFT を行い、検出に用いる周波数範囲を設定する。

予備実験の結果、今回選択したテンプレートでは、100Hz ~ 1500Hz の周波数帯域において、ボール打撃による特徴と考えられるスペクトルが現れ、より高周波の帯域ではノイズ成分が多く含まれていることが確かめられた。よって、本稿では 100Hz ~ 1500Hz の周波数範囲をボール打撃時刻抽出に用いる周波数範囲として設定した。

3. 抽出方法

音響データに対し 2048 点単位で切り出しながら FFT 処理を行う。この処理を 128 点 (0.0029 秒) 間隔で順次開始位置をずらしながら全ての音響データに対して行い、テンプレートとの照合を行った。先に設定した周波数範囲内において音響データとテンプレートの相関係数が予め設定した閾値より大きな値を示す音響データの存在時刻をボール打撃時刻とした。

3.5 選手基本動作の識別

選手基本動作の識別は、コート・ネットライン、

選手位置，ボール位置，ボール打撃時刻を総合的に判断することにより行う．選手基本動作のインデックスは次のような構成をもつ．

(obj_id, ts, te, tr, act_class, locus)

ここで，obj_id は選手 ID で，コート内の選手を表す整数値である．ts, te はそれぞれ基本動作の開始・終了時刻を表す．tr は，選手がボールを打った瞬間等の基本動作のキーとなる代表時刻を表す．act_class は，基本動作クラスを表す値で，ここでは，“stay”，“move”，“forehand_swing”，“backhand_swing”，“overhead_swing”の 5 種類の値をとる．locus は，選手軌跡を表し，開始・終了時刻間の任意時刻における選手位置を近似可能ないくつかの代表点を選択的に記録している．

特に，forehand swing, backhand swing, overhead swing の 3 つの基本スイング動作は，ボール打撃時刻における，選手とボールの相対的位置関係を評価することで識別される．本稿では，ボール打撃時刻においてボールが選手を囲む外接四角形上部の識別ラインより上方にあれば overhead_swing, 選手重心に対してフォアサイドあるいはバックサイドにあれば，それぞれ forehand_swing, backhand_swing と判定した (図 8)．なお，識別ラインは，選手外接四角形の y 方向の長さに応じ一定の比率で定まる選手領域上部に設定した．

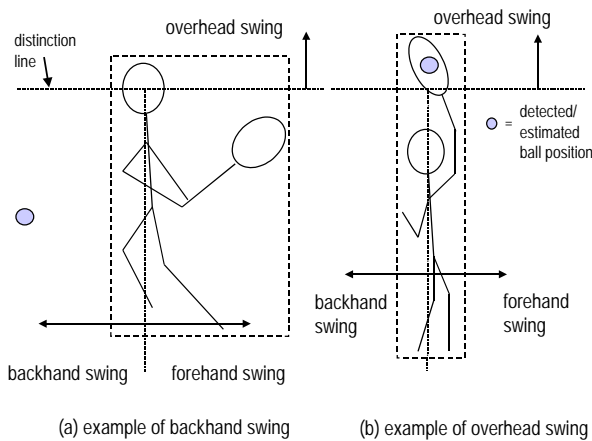


図 8．選手基本動作の識別

4. 一般動作インデックスの生成

本システムにおいては，ある一般動作は，すでに自動注釈付けなどで出力済みの基本動作インデックスや他の出力済み一般動作インデックスが，予め定義されたある特定の時空間関係を満たした時に成立

する，と考える．

例えば，図 9 はテニスにおいて抽出された基本動作インデックスの一例を示している．横軸は時間を表し，縦軸は選手やボールなどオブジェクトの種類を示す．Player1 の“overhead_swing”に相当する矢印は，その時区間の間，Player1 にあたる選手は基本動作“overhead_swing”をしており，コート上の j -i 個の代表点 (p_{i+1}, \dots, p_j) からなる軌跡をたどったことを表している．

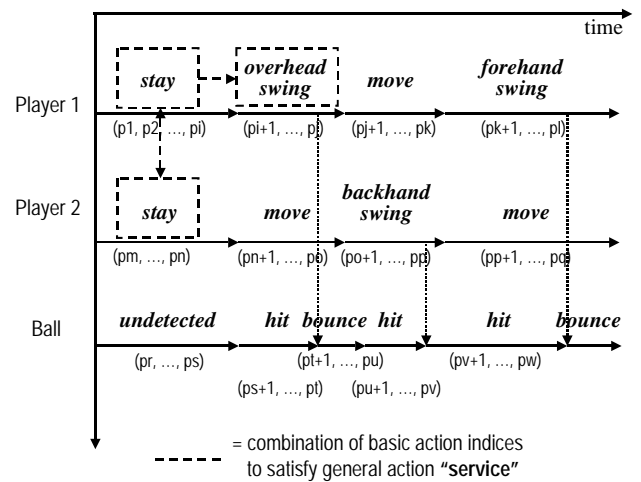


図 9．テニスにおける基本動作インデックス

ここで，テニスの「サービス」という一般動作について考えてみよう．「サービス」動作は，次のような条件で成立すると考えられる：

両選手がある同時刻においてコート外側区画にてともに“stay”し，次にいずれかの選手がコート外側区画において“overhead_swing”を行う．

このような一般動作が成立するための条件は，本システムにおいては，ルール集合としてデータベースとは別に予め人手によって用意されている (図 1)．

そして，一般動作「サービス」のインデックスは，図 9 で示される基本動作インデックス (あるいはすでに出力された他の種類の一般動作インデックス) の中から，ルール集合に記述された「サービス」動作の成立条件を満たすインデックスの組み合わせを特定することで出力される．図 9 の粗い点線部分は，「サービス」動作の条件を満足する基本動作インデックスの組が特定されたことを表している．

以上，一般動作インデックスを，出力済みのインデックスに基づき生成することにより，本システムは次のような利点をもつ．

- 自動注釈付けなどで用いられる画像・音響解析などの一般に負荷の高い種々のメディア情報解

析は、基本動作インデックスを生成する部分まで達成できればよい。

- 一般動作の追加・修正作業は、ルール内容の追加・修正、および、すでに出力済みのインデックスの中から新しいルール内容を満たす部分を再探索することにより、新たなインデックスを生成することで達成可能である。

一般動作は、閲覧する際に要求される種類も多く、内容検索では適宜新しい項目の追加や修正の可能性が考えられる。新しい種類の一般動作の閲覧要求が発生するたびに、新たな画像処理・音響解析を行わなくても、対応するインデックスが生成できるという点で、本システムはインデックス管理の負荷を軽減することができると考えられる。

5. 実験および考察

5.1 基本動作インデックスの生成

男子・女子シングルスそれぞれ1試合ずつの実映像から選択したテストシーケンスに対し、コートライン、選手位置、ボール位置を抽出した。使用したテストシーケンスに対しては、ほぼ正確に追跡が行えたが、一部急激なカメラ動作のために選手・ボール追跡が失敗した箇所があり、これらは今回の以後の実験からは除外することとした。

まず、音響データ解析によるボール打撃時刻の抽出結果を表1に示す。

表1. ボール打撃時刻の抽出結果

	Precision	Recall
total	0.913	0.975

比較的簡単な方法であるにもかかわらず、良好な検出結果が得られた。検出誤りおよびもれの要因として、ボールとラケットの当たり具合により打撃音が変化し、今回実験に用いた検出パターンと周波数特徴が一致しなかったことが考えられる。

次に、抽出したボール打撃時刻に基づく選手基本動作の識別結果を表2に示す。

表2. 選手基本動作の識別結果

	Precision	Recall
total	0.907	0.907

動作識別としては比較的良好な結果が得られているといえる。ここで観測された誤りは、音響解析におけるボール打撃時刻の抽出誤りと、ボール打撃時のボール位置推定に必要となる、ボール追跡処理結果に含まれる誤りであったと考えられる。ボール位置、ボール打撃時刻、選手基本動作の抽出処理で用いら

れるパラメータと、その雑音に対するロバスト性について今後検討する必要がある。

また、選手基本動作の他の識別方法として、シルエットを用いたジェスチャ認識による方法が考えられる。しかし、シルエットによるジェスチャ認識では、必要な学習データ量が多いこと、人物・カメラの向き、解像度の不足、個人性などによって、シルエットの見え方にバリエーションがあり、識別結果がそれらに大きく依存する、という問題点がある。

本稿で用いた方法は、選手とボールの位置関係という比較的単純な要素から判断されるため、特に個人性によるバリエーションの影響はジェスチャ認識による方法よりも少ないと考えられる。

今後より詳しい検証のために、両者の方法を比較する必要があると考えられる。

なお、本システムの処理時間については、現在、1試合（1時間程度）のデータ計算に原データの入力から選手基本動作の出力を得るまでに半日以上かかっている。フレーム間引きや最適化を行うことで、処理速度の向上を図る必要があると考えられる。

5.2 一般動作インデックスの生成

種々のルール定義を行うことで、選手基本動作インデックスから一般動作インデックスを生成した。

例えば、「ネットダッシュする」という動作は、次のように定義できると考えられる：

あるボール打撃時刻における打者選手がネットコート以外の区画にあり、次のボール打撃時刻においてネットコート内に位置する。

一般動作インデックスをこのような定義に基づいて生成し、データベース登録することで、一般動作の内容検索が行えるクライアントを実装した。

図10は、検索条件として、「第1セット第3ゲームでのサービス」を設定し、その検索結果が得られた様子を示す。サービスを含む映像シーンのキー画面（代表時刻に相当する）が表示されており、用意した映像シーケンスからサービスが14件抽出でき、そのうちの10件が表示されていることを示す。

図11は、図10で表示されている10件の検索結果を選択し、条件に一致した部分だけを連続して再生している様子を示す。第1セット第3ゲームのサービス部分だけが選択的に再生され、早送りや巻き戻しといった作業をすることなく、目的とする場面を短時間で確認することができる。

実際、サービス、ネットダッシュするといった複雑な一般動作についても柔軟に定義、検索が可能であることを確認した。



図 10 . 設定した検索条件と検索結果



図 11 . 検索条件に一致した部分の連続再生

6. まとめ

本稿では、動作インデックスによる映像の自動注釈付けを行い、それを適宜参照することにより柔軟な内容検索が可能となるシステムについて報告した。テニス映像を例として取り上げ、自動注釈付けを以下の手順で行った。入力映像・音響データからコート・ネットライン、選手・ボール位置、選手によるボール打撃時刻を抽出した。そして、ボール打撃時刻における選手とボールの位置関係から、フォアハンドスイング等の選手基本動作を識別し、選手 id や基本動作クラスを含む動作インデックスを出力した。一般動作については、各動作名とそれが成り立つための諸条件を定義したルール集合をデータベース

とは別に予め用意し、新たな一般動作のインデックスは、すでに出力されている基本動作や他の一般動作インデックスを参照し、それらの間である特定の時空間関係が成り立つかどうかを評価することで出力される。実際、サービスやネットダッシュするといった複雑な一般動作についても柔軟に定義、検索が可能であることを確認した。

今後は、自動索引付けの精度と処理速度の向上、および、種々のルール定義による一般動作の識別精度について分析を進める必要がある。

文 献

- [1] S.Ravela, et. al.: "Retrieving images by similarity of visual appearance", In the Proc. of the IEEE Workshop on Content Based Access of Images and Video Databases, CAIVD'97, pp.67-74, 1997.
- [2] M.Flickner, et. al.: "Query by image and video content: the QBIC system", IEEE Computer Magazine, pp.23-32, 1995.
- [3] A.Nagasaka, Y.Tanaka: "Automatic video indexing and full-video search for object appearances", IPSJ Trans. Vol.33, No.4, pp.543-550, 1992.
- [4] A.Akutsu et. al.: "Video indexing using motion vectors", In SPIE Proc. Visual Communication and Image Processing '92, pp.522-530, 1992.
- [5] H.Zhang, et al.: "Automatic Parsing and Indexing of News Video", Multimedia Systems, vol.2, no.6, pp.256-266, 1995.
- [6] 有木, 杉山, 石川, 寺西, 桜井: "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, pp.31-38, 1996.
- [7] M.Smith, T.Kanade: "Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques", IEEE Computer Vision and Pattern Recognition (CVPR), 1997.
- [8] Y.Nakamura, T.Kanade: "Semantic analysis for video contents extraction - spotting by association in news video", ACM Multimedia, pp.393-401, 1997.
- [9] 牛尼剛聡, 渡邊豊英: "イベントアクティビティ・モデルに基づくシーン検索における可変的な役割の利用", 情処論文誌, Vol.40, No.SIG3(TOD1), pp.114-123, 1999.
- [10] 植田和憲, 鎌原淳三, 下條真司, 宮原秀夫: "シナリオテンプレートによるストーリー性を重視したダイジェスト生成機構", 情処 DBS-119-24, pp.139-144, 1999.
- [11] 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", 情処論文誌, Vol.41, No.SIG3(TOD6), pp.71-84, May 2000.
- [12] Y.Gong, L.T.Sin, C.H.Chuan, H.Zhang, M.Sakauchi: "Automatic parsing of TV soccer programs", Proc. Int'l Conf. on Multimedia Computing and Systems, pp.167-174, 1995.
- [13] D.D.Saur, Y-P.Tan, S.R.Kulkarni, P.J.Ramadge: "Automated analysis and annotation of basketball video", Storage and Retrieval for Image and Video Databases V, SPIE-3022, pp.167-187, 1997.
- [14] G.Sudhir, J.C.M.Lee, A.K.Jain: "Automatic classification of tennis video for high-level content-based retrieval", Proc. of IEEE Workshop on Content-Based Access of Image and Video Databases, CAIVD'98, 1998.
- [15] Y.Ohno, J.Miura, Y.Shirai: "Tracking players and a ball in soccer games", CVIM, Tech. Report of IPSJ, 114-7, pp.49-56, 1999.