

ブースティングとモーションディスクリプタの組合せによる人の動作検出

緒 方 健 人[†] ウィリアム クリスマス^{††} ジョセフ キットラー^{††}
 ジュークイタン[†] 石 川 聖 二[†]

本論文では、Efros ら (2003) のモーションディスクリプタおよび Ke ら (2005) のイベント検出器の組合せに基づく動作検出手法を提案する。両手法の直接の組合せは、AdaBoost の学習アルゴリズムと Ke らの弱識別器の特性によって動作検出器の学習の計算時間が増大するという問題がある。よって、本研究ではこの問題を解決するため弱識別器の次元を拡張した改良型組合せ手法を提案する。また、さらなる高速化のため、学習に Look Up Table を導入する。実験では、実際の放送に用いられたテニスの映像に提案手法を適用し各種ストロークの検出を試みた。性能を評価するため、Efros らの手法、Ke らの手法および直接の組合せ手法も適用し結果を比較した。実験結果より、提案する組合せ手法は認識率および計算時間の両面で元となった手法の良い特性をあわせ持つこと、また改良型組合せ手法は直接の組合せによる検出器とほぼ同じ性能でありながら、LUT との併用により学習に要する計算時間を大幅に削減できることが確認できた。

Human Activity Detection by Combining Motion Descriptors with Boosting

TAKEHITO OGATA,[†] WILLIAM CHRISTMAS,^{††} JOSEF KITTLER,^{††}
 JOO KOOI TAN[†] and SEIJI ISHIKAWA[†]

A new, combined human activity detection method is proposed. Our method is based on Efros, et al.'s (2003) motion descriptors and Ke, et al.'s (2005) event detectors. Since both methods use optical flow, it is easy to combine them. However, the computational cost of the training increases considerably because of the increased number of weak classifiers. We reduce this computational cost by extending Ke, et al.'s weak classifiers to incorporate multi-dimensional features. We also introduce a Look Up Table for further high-speed computation. The proposed method is applied to off-air tennis video data, and its performance is evaluated by comparison with the original two methods. Experimental results show that the performance of the proposed method is a good compromise in terms of detection rate and computation time of testing and training.

1. 序 論

動画像処理の研究が、近年のコンピュータ技術の急速な発展とともに注目を集めている。特に人の動作は人とコンピュータがコミュニケーションする際にコンピュータが人を理解するための重要な情報の1つであり、またカメラ映像は非接触で人の動作情報を得られるという利点があるため、人の動作認識の研究がさ

かんに行われている³⁾。人の動作は人を含む映像において重要な情報を持つことが多いため、マンマシンコミュニケーションに限らず様々な分野で必要とされている。スポーツ映像解析もその1つである。スポーツ映像解析の研究には様々な研究課題があるが、人の動作認識は自動解説システム^{1),7)}や戦略解析⁹⁾などの高度なアプリケーションの実現において重要な役割を果たすと考えられる。これらのアプリケーションが実現すれば、映像情報に付加する形で様々な情報を視聴者(ユーザ)に提供でき、ユーザの多面的な理解を補助することができる。

スポーツの動作は明確に定義されているため、人の動作認識の研究で実験サンプルとして頻繁に用いられる。しかし、実際の放送に用いられるスポーツ映像は、

[†] 九州工業大学工学部機械知能工学科
 Department of Mechanical and Control Engineering,
 Kyushu Institute of Technology

^{††} 連合王国サリー大学視覚音声信号処理センター
 Centre for Vision, Speech and Signal Processing, Uni-
 versity of Surrey, U.K.

注目するプレイヤーとその周囲を同時にとらえるように撮影されるため、一般的にプレイヤーの解像度が低い。またスポーツによってはカメラの視点が移動する。このためこれらの研究の多くは実験に解像度の高い独自の映像データを用いる¹⁰⁾か、もしくはスポーツ映像中の解像度の高いプレイヤーのみに注目している¹¹⁾。したがって、解像度の低いスポーツ映像においてこれらの手法が適用可能かどうかは確認されていない。

Efrosらは解像度の低いスポーツ映像に有効な動作認識手法を提案した²⁾。カメラの動きおよびプレイヤーの動きに対応するために、まずプレイヤーを追跡しプレイヤーを中央にとらえた画像を生成する。その画像からオプティカルフローを求め、それを4つのチャンネルに分解して得られるモーションディスクリプタを動作認識の特徴量として用いる。実験ではサッカーの映像でプレイヤーの走る動作を用い、解像度の低い状態での認識に成功した。しかし、この手法の問題点として、動作認識手法の計算コストがあげられる。彼らは k -近傍法を用いているため、認識用データベース内のデータの量が多いほど認識率が高くなるが、その分計算コストも比例的に大きくなってしまふ。

一方、近年提案された手法として、Keらのイベント検出器がある⁵⁾。これはViolaらの提案したAdaBoostに基づく顔などの2-Dパターン検出器⁸⁾を拡張し、時間情報を含む動画からの3-Dパターン検出器としたものである。この手法は学習のための計算時間は長いですが、検出のための計算時間は短い。

本論文では、上述のEfrosらの手法とKeらの手法を組み合わせ、解像度の低い人の動作を高速に検出する手法を提案する。Keらのイベント検出器は2つの成分を持つオプティカルフローに基づいているため、Efrosらの4つの成分を持つモーションディスクリプタと直接組み合わせることができる。しかし、AdaBoostの学習アルゴリズムとKeらの弱識別器の特性によって、オプティカルフローの2倍の成分を持つモーションディスクリプタと組み合わせると学習のための計算時間が増大してしまう。そこで本論文では、これを軽減するため、Keらの弱識別器の次元を拡張した改良型組合せ手法を提案する。また、学習のさらなる高速化のため、Look Up Table (LUT)を導入する。実験では実際の放送に用いられたテニス映像に提案手法を適用し、テニスの各種ストロークの検出を試みた。また、提案手法の性能を評価するため、Efrosらの手法、Keらの手法および直接の組合せによる手法も適用し結果を比較した。実験結果より、組合せ手法は認識率および計算時間の両面において元となった手法の良い

特性をあわせ持つことが確認できた。また、改良型組合せ手法は直接の組合せによる手法とほぼ同じ性能でありながら、LUTとの併用により学習に要する計算時間を大幅に削減できることが確認できた。

提案手法は追跡部と動作検出部から構成される。よって、本論文の構成も同様に、まず2章で本手法で用いた追跡手法について説明し、続いて3章でEfrosらの手法とKeらの手法を組み合わせた動作検出法、および学習の計算時間短縮のために導入した次元拡張弱識別器とLUTについて説明する。4章では提案手法をテニス映像に適用するための詳細について述べる。5章で実験結果を示し、最後に6章で結論を述べる。

2. 追跡手法

Efrosらのモーションディスクリプタは解像度の低い対象に有効な手法であるが、まず対象の追跡が必要である。しかし、一般に解像度の低い、形状の変化する対象の追跡は簡単ではない。Efrosらは単純な正規化相関に基づくテンプレートマッチング手法を用いているが、対象の形状変化が大きい場合、もしくはテンプレートに類似した物体が対象の近くにある場合(局所解問題)追跡に失敗することがある。

本研究では追跡にNummiaroらの提案した色情報に基づくパーティクルフィルタの追跡手法⁶⁾を適用する。パーティクルフィルタ(以下PFと略す)は状態推定手法の1つである。画像追跡の場合、追跡対象の位置や速度などを状態とし、状態空間内における状態の確率分布を多数の重み付き粒子で離散的に近似し、運動モデルに基づき新しい状態分布を予測、そして観測データから得られる尤度により各粒子を評価することで新しい分布を近似する、という処理を繰り返し対象を追跡する。Nummiaroらはこの尤度をヒストグラムに基づく色情報から求めているため、対象の大きさや形の変化に頑強となる。本研究では、さらに類似した色情報を持つ物体が背景に含まれる場合に発生する局所解問題を解決するため、PFの平滑化アルゴリズム⁴⁾を組み合わせる。したがって、追跡のアルゴリズムは以下の2段階からなる。

- (1) 色情報に基づくPFを用いて対象の追跡を行う。このとき、各フレームにおいてサンプリングされ生き残ったパーティクルの状態の履歴をすべて保持しておく。
- (2) 最終フレームの時点で生き残っているパーティクルの履歴を用いて追跡結果を平滑化する。

追跡の結果より、対象を中央にとらえた動画データが動作検出器へと送られる。

3. 組合せに基づく動作検出器

提案する動作検出器は、Efros らのモーションディスクリプタと Ke らのイベント検出器の組合せに基づく．ここではまず 2 つの手法の組合せ，組合せにより動作検出器の学習に必要な時間が増大する原因，およびそれを軽減するために導入した改良型組合せ手法および LUT の学習への導入について説明する．

3.1 Efros らのモーションディスクリプタ (MD)

Efros らの提案したモーションディスクリプタ²⁾ (以下 MD と略す) は以下の処理より求められる．対象の追跡により，対象を中央にとらえた動画データが与えられているものとする．まず，連続する 2 フレームからオプティカルフローを計算し，水平成分 F_x および垂直成分 F_y に分解する．さらに， F_x および F_y をそれぞれ正の成分 F_x^+ , F_y^+ および負の成分 F_x^- , F_y^- に分解する．ここで，もし正の成分に値が入る場合，対応する画素の負の成分は 0 値となり，逆の場合も同様である．そして各成分を 2-D ガウシアンフィルタでぼかし，4 つの成分を持つ MD Fb_x^+ , Fb_x^- , Fb_y^+ および Fb_y^- を求める．この処理の流れを図 1 に示す．

3.2 Ke らのイベント検出器との組合せとその問題点

Ke らのイベント検出器⁵⁾ は Viola らの 2-D パターン検出器の拡張である．この 2-D パターン検出器は AdaBoost に基づいており，その弱識別器の特徴量には Haar-like 特徴が用いられる．Haar-like 特徴は図 2 上部に示されるような形状をしており，2-D パターン検出器の適用されるサブウィンドウ内のある領域の画素値を基に計算される．Ke らはこの Haar-like 特徴を図 2 下部に示される 3 次元の Volumetric 特徴に拡張し，動画像データから 3-D パターンを検出するイベント検出器を提案した．よって，Ke らのイベント検出器は動画像を構成する連続画像のある期間内を見る 3-D サブウィンドウであり，Volumetric 特徴はそのサブウィンドウ内のある 3 次元領域の画素値から計算される．

Ke らは 3-D パターン検出器の入力としてオプティカルフローを用いているが，実際はオプティカルフローを水平成分 F_x および垂直成分 F_y に分解して用いている．そのため，4 つの成分を持つ MD と組み合わせることができる．しかし，この組合せは Adaboost の学習のアルゴリズムと Ke らの用いた弱識別器の特性のため，検出器の学習の計算コストが増大するという問題がある．なお，以降ではこの MD とイベント検出器を直接組み合わせたものを単純組合せ手法と呼

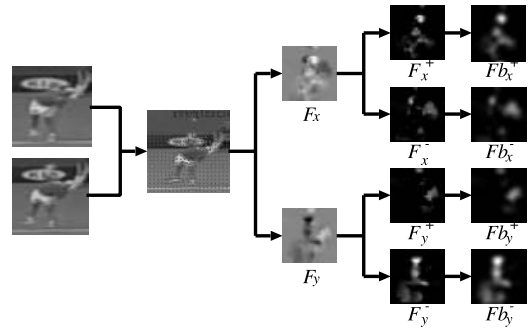


図 1 モーションディスクリプタの計算課程
Fig. 1 Computation of motion descriptors.

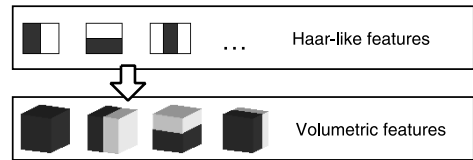


図 2 Haar-like 特徴と Volumetric 特徴
Fig. 2 Haar-like features and Volumetric features.

ぶこととする．

AdaBoost は機械学習法の 1 つで，単純な識別法では識別が難しいデータを，複数の単純な識別器 (弱識別器) の組合せにより識別する手法である．この弱識別器の組合せは学習により自動的に決定される．AdaBoost 検出器の学習には 3 つのセットが必要となる．AdaBoost が検出したいデータを集めた正のサンプルセット，AdaBoost が検出たくないデータを集めた負のサンプルセット，そして弱識別器のセットである．学習の大まかな流れを以下に示す．

- (1) すべての弱識別器の閾値を正と負のサンプルセットから決定する．
- (2) 正と負のサンプルデータ 1 つ 1 つに重みを用意する．最初は同じ値で初期化する．
- (3) 全弱識別器を正と負のサンプルセットに適用し，弱識別器ごとに識別に失敗したサンプルの重みを合計しそれを評価値とする．
- (4) 評価値が最も小さい弱識別器を AdaBoost 識別器に登録し，登録された弱識別器が識別に失敗したサンプルの重みを大きくする．また，このとき評価値からこの弱識別器の重みを求め，登録する．
- (5) (3), (4) の繰返しにより弱識別器を順番に登録していく．

よって AdaBoost 検出器の学習時間は学習に用いるサンプル数，AdaBoost 識別器に登録する弱識別器の数，そして学習時に生成した弱識別器の総数に依存

する。

Ke らの弱識別器は Viola らの弱識別器⁸⁾と同じく、1次元の特徴量から正負の判定を行う。すなわち、弱識別器は計算された Volumetric 特徴 (Viola らの場合は Haar-like 特徴) と弱識別器内のある閾値とを比較し、正か負かの判定を出力する。この閾値は学習時に学習サンプルセットから決定される。ここで、Ke らの弱識別器はオプティカルフロー成分のうち F_x か F_y の一方から Volumetric 特徴を計算する。したがって、Ke らの検出器の学習時に N_v 個の Volumetric 特徴を用いる場合、 F_x と F_y それぞれに対応する弱識別器が必要なため、 $2N_v$ 個の弱識別器セットが必要である。そして単純組合せ手法の場合は成分数 4 の MD を用いるため、オプティカルフローを用いる場合のさらに 2 倍の弱識別器セットが必要となり、学習の計算時間が増大する。

本研究ではこの組合せ検出器の学習時間を軽減するため、Ke らの弱識別器の次元を拡張した次元拡張型弱識別器を用いた改良型組合せ動作検出器を提案する。また、学習の高速化のため LUT を導入する。

3.3 次元拡張型弱識別器

前述したように、Ke らの弱識別器は Volumetric 特徴をオプティカルフローのどちらか一方の成分から計算する。しかし、オプティカルフローは 2 次元の速度ベクトルであるので、それぞれの成分で求められた Volumetric 特徴は 2 次元の特徴ベクトルとして用いることが可能だと考えられる。したがって、本研究では Ke らの弱識別器を拡張し、1 つの Volumetric 特徴をすべての成分から計算し、それを特徴ベクトルとして用いて判定を行う次元拡張型の弱識別器を構成する。これにより、学習時入力成分の数によらず弱識別器の数が Volumetric 特徴の数と同じとなるため学習の計算時間が短縮できる。なお、この拡張は弱識別器を Ke らのものより複雑にするが、次元数が小さければ計算が極端に複雑にはならない。

MD を用いる場合、弱識別器内の特徴ベクトルの次元数は 4 となる。したがって、弱識別器 w の生成する k フレーム時点での特徴ベクトル $\alpha_w(k)$ は以下の式より求められる。

$$\alpha_w(k) = \begin{bmatrix} f_w(Fb_x^+(k), Fb_x^+(k-1), \dots) \\ f_w(Fb_x^-(k), Fb_x^-(k-1), \dots) \\ f_w(Fb_y^+(k), Fb_y^+(k-1), \dots) \\ f_w(Fb_y^-(k), Fb_y^-(k-1), \dots) \end{bmatrix}$$

ここで $Fb_x^+(k)$, $Fb_x^-(k)$, $Fb_y^+(k)$ および $Fb_y^-(k)$ は k フレーム時点での MD, $f_w()$ は弱識別器 w の

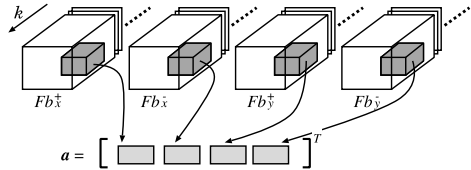


図 3 次元拡張型弱識別器の生成する特徴ベクトル
Fig.3 Computation of the proposed weak classifier.

Volumetric 特徴を計算する関数である。この計算を視覚的に表したものを図 3 に示す。

特徴ベクトルを求めた後、弱識別器は正か負かの判定を行う。提案する弱識別器は、学習時に与えられた正のサンプルセットおよび負のサンプルセットからそれぞれの特徴ベクトルの平均 μ_w^p および μ_w^n を求め、この平均ベクトルとの距離を用いて判定を行う。したがって、弱識別器 w の k フレーム時点での出力 $h_w(k)$ は以下の式で表される。

$$h_w(k) = \begin{cases} 1 & \text{if } d_w^p(k) < d_w^n(k) \\ 0 & \text{otherwise} \end{cases}$$

ここで $d_w^p(k) = \|\mu_w^p - \alpha_w(k)\|$, $d_w^n(k) = \|\mu_w^n - \alpha_w(k)\|$ である。

提案する弱識別器を用いた動作検出器の学習過程、および出力は従来の AdaBoost 検出器と同様である。学習時は、正のサンプルセットとして検出したい動作の MD を、また負のサンプルセットとして検出したい動作を含まない MD を用いる。弱識別器セットには 3-D サブウィンドウ内に形状や位置、大きさの異なる Volumetric 特徴をランダムに生成したものをを用いる。Volumetric 特徴の形状は、図 2 下部に示される Ke らと同じ 4 種類を用いる⁵⁾。そして 3.2 節に示す AdaBoost の学習アルゴリズムにより、最終的な検出器として学習時に与えた弱識別器セットの部分集合と対応する重みが登録される。

動作検出器の出力は、AdaBoost 検出器に登録された弱識別器と対応する重みによる、重み付き多数決によって決定される。

3.4 LUT による計算時間の短縮

3.2 節に示される AdaBoost 学習アルゴリズムの (3) のとき、すなわち AdaBoost 識別器へ登録する弱識別器を選択する際、全弱識別器が正と負のサンプルセットに適用されるが、正と負のサンプルセットは不変なため、サンプルに対する弱識別器の出力はつねに同じである。そこで本研究では、弱識別器の閾値決定後その弱識別器のサンプルセットに対する 2 値の出力を LUT に記録し、その後の弱識別器の選択時にその

結果を参照するようにした．これにより計算時間を大幅に削減できる．しかし、全サンプルの数と全弱識別器の数を掛け合わせた数の2値データを記録する必要があるため、使用するメモリスペースは大きくなる．

4. テニスストローク検出システム

本研究では提案手法をテニスの各種ストロークを検出するシステムとして実装した．テニスのゲーム形式はシングルス、カメラはパンチルトカメラを想定している．システム構成を図4に示し、それぞれの詳細を以下に示す．ここでは1人のプレイヤーに対する処理について述べているが、これを映像内の2人のプレイヤーそれぞれに適用する．

4.1 プレイヤ追跡部

テニス映像内のプレイヤーを、2章で述べたアルゴリズムを用いて追跡する．テニスの映像内ではプレイヤーの体の動きが大きいため手や足を含んだ体全体を追跡することは難しい．そこでプレイヤーの胴体のみを追跡する．なお、現在はプレイヤーの位置と色モデルは最初フレームにおいて手動で初期化している．

追跡の結果より、プレイヤーを中央にとらえた動画データが生成される．

4.2 動作検出部

4.2.1 モーションディスクリプタ生成部

プレイヤーを中央にとらえたデータから3.1節で述べたアルゴリズムでMDを生成する．

4.2.2 動作検出器

最後に、生成されたMDに動作検出器を適用しストロークを検出する．提案する動作検出器は2値識別器であるので、検出器はストロークごとに用意する必要がある．検出器学習用の正のサンプルセットに用いるMDはプレイヤーがボールを打つ瞬間で同期をとっている．本ストローク検出システムでは、動作検出器の3-Dウィンドウのサイズを32[pixel] × 32[pixel] × 20[frame]

とする．

5. 実験

実装したテニスストローク検出システムを、実際に放映に用いられた映像に適用し実験を行った．実験データには2003 Australian tennis tournamentの男子シングルス決勝から、コート全体が写し出されプレイヤーが試合を行っている部分を複数個手動で切り出したものを用いた．提案手法の性能を評価するため、Efrosらの手法およびKeらの手法も同じデータに適用し結果を比較した．

なお、簡単のため映像内でカメラから近い位置でゲームをしているプレイヤーを“手前のプレイヤー”、またカメラから遠い位置でゲームをしているプレイヤーを“奥のプレイヤー”と呼ぶこととする．

5.1 実験データ

テニスはチェンジエンドのルールがあるため、プレイヤーはコートを交代しながら試合を行う．したがって、実験動画データはプレイヤーの配置によって2種類のデータセットに分類できる．すなわち、片方のプレイヤーに注目したとき、そのプレイヤーが手前でプレイしている動画と、奥でプレイしている動画に分けられる．本実験では一方を学習用データセット、もう一方を実験用データセットとして用いた．学習用データセットは18個、実験用データセットは16個の動画を含む．

学習のため、また実験結果の評価のため、すべての動画データから両プレイヤーのボールを打つ瞬間のすべてのフレームを手動で検出し、その時間とストロークの名前を真値として登録した．対象としたストロークは、サーブ(SS)、フォアハンドストローク(FS)、そしてバックハンドストローク(BS)である．このとき、定義するのが難しいストローク、もしくはその他のストロークがいくつか含まれていたため、これらはDon't care(DC)として定義した．DCは学習にも用いず、また実験結果の評価にも用いていない．学習用データセットには奥のプレイヤーのSSが9、FSが19、BSが24個、手前のプレイヤーのSSが9、FSが27、BSが24個含まれている．また実験用データセットには奥のプレイヤーのSSが8、FSが21、BSが21個、手前のプレイヤーのSSが8、FSが18、BSが19個含まれている．

5.2 各手法の設定

実装したEfrosらの*k*-近傍法²⁾の詳細を以下に示す．MDの生成時に用いるガウシアンフィルタの標準偏差は1.225とした．本実験では1-、3-、5-近傍法(1-NN、3-NN、5-NN)それぞれの結果を求めた．ま

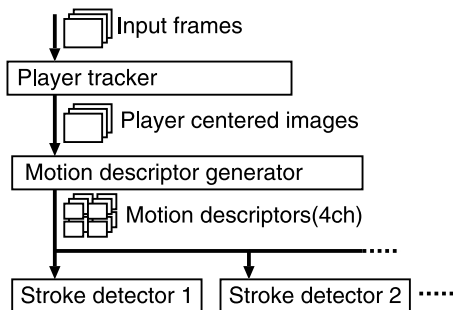


図4 ストローク検出システムの構成

Fig. 4 The construction of the tennis stroke detection system.

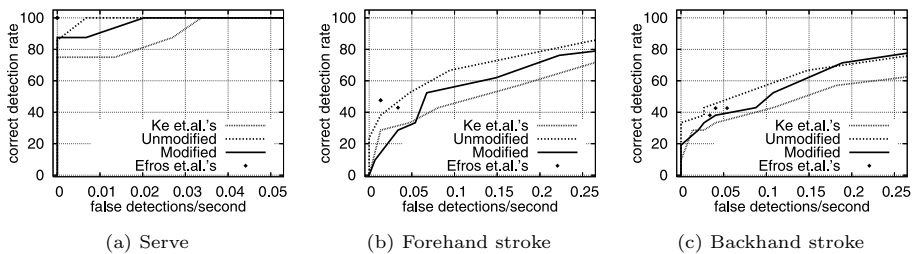


図 5 奥のプレイヤーのストローク検出結果
Fig. 5 Detection results of the far player.

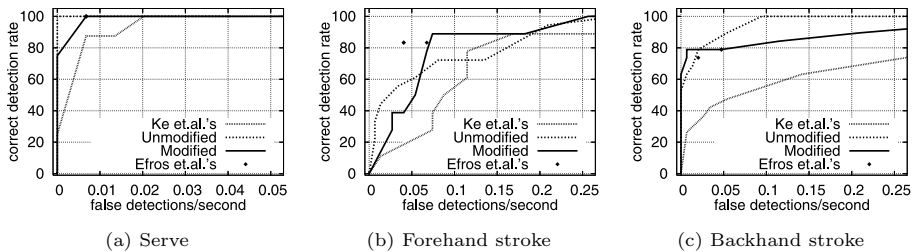


図 6 手前のプレイヤーのストローク検出結果
Fig. 6 Detection results of the near player.

ず、Efros らの k -近傍法を適用することで実験用データの各フレームにおいて学習データセットの中から k 番目までの類似したフレームが選択される。もしこれらのフレームが実験用データセットの真値に登録されたいずれかのストロークに近い (± 5 フレーム) 場合、そのフレームにそのストロークのラベルを持たせ、どのストロークにも近くない場合は“ストロークなし”というラベルを持たせる。そして各フレームにおいて多数決を行い、最も投票数が多いストロークをそのフレームにおける結果とした。以降、この手法を Efros らの手法と呼ぶ。

一方、AdaBoost に基づく手法は、まず 50,000 個の Volumetric 特徴を $32 [\text{pixel}] \times 32 [\text{pixel}] \times 20 [\text{frame}]$ の 3-D ウィンドウ内にランダムに生成し、これらを用いて以下の 3 種類の検出器を構成した。

- オプティカルフローを入力とする Ke らの検出器を 100,000 個の弱識別器から構成 (F_x と F_y それぞれに 50,000 個)
 - MD を直接み組合せた Ke らの検出器、すなわち単純組合せ検出器を 200,000 個の弱識別器から構成 (Fb_x^+ , Fb_x^- , Fb_y^+ と Fb_y^- それぞれに 50,000 個)
 - MD と次元拡張型弱識別器の組合せ検出器、すなわち改良型組合せ検出器を 50,000 個の弱識別器から構成
- これらの 3 つの動作検出器はすべて 1 段階の Ada-

Boost 識別器から構成され、同じ学習サンプルセットを用いて学習した。各検出器内に含まれる弱識別器の数は、Ke らの検出器は 100、単純組合せ検出器は 200、そして改良型組合せ検出器は 50 とした。これは 1 つの検出器が扱う特徴量を同じにすることで、これらの検出器を正当に評価するためである。以降、これら 3 つの検出器に基づく手法をそれぞれ Ke らの手法、単純組合せ手法、そして改良型組合せ手法と呼ぶ。

評価のため、それぞれの検出結果は実験用データセットの真値と比較された。AdaBoost 検出器の結果は時間軸を持つ 1 次元の 2 値データとなるため、まず連続した同じ検出結果を 1 つの検出として統合した。そしてこの検出結果がストロークの真値もしくはその近く (± 5 フレーム) と重なっていれば True Detection、そうでなければ False Detection とした。なお、Efros らの手法の出力は多値であるが、これをストロークごとに分解し上記と同じ処理を施した。

5.3 実験結果

各手法の検出結果の ROC 曲線を図 5 および図 6 に示す。AdaBoost に基づく手法の ROC 曲線は検出器の閾値を変化させて求めた。図 5 は奥のプレイヤーの結果、図 6 は手前のプレイヤーの結果を示す。それぞれ (a) はサーブ、(b) はフォアハンドストローク、(c) はバックハンドストロークの検出結果である。各グラフにおいて、細かい破線が Ke らの手法、粗い破線 ('unmodified' の線) が単純組合せ手法、そして実線

表 1 ストローク検出の計算時間の比較

Table 1 Computation time of the stroke detection.

フレーム数: 1,059

手法	計算時間 [秒]
Efros らの手法	652.19
Ke らの手法	0.2129
単純組合せ手法	0.2826
改良型組合せ手法	0.4219

表 2 組合せ手法の学習計算時間の比較

Table 2 Computation time of the training with LUT and without LUT.

(a) LUT なし

(a) Training without LUT

手法	計算時間 [秒]
単純組合せ手法	74,098
改良型組合せ手法	23,621

(b) LUT あり

(b) Training with LUT

手法	計算時間 [秒]
単純組合せ手法	9,504
改良型組合せ手法	1,295.1

(‘modified’ の線) が改良型組合せ手法の結果を示す。Efros らの手法の結果は黒い点で示される。図 5 (a) ではすべてのサーブがすべての k -近傍法で検出された。(b) では 1-NN と 5-NN が 47.6%, 3-NN が 42.9% であり, (c) は 1-NN と 3-NN が 42.6%, 5-NN が 38.1% の検出率であった。一方, 図 6 (a) でもすべての k -近傍法がサーブを 100% 検出した。(b) ではすべての k -近傍法の検出率は同じ 83.3% であるが 1-NN の過検出数が他より多い。(c) では 1-NN および 3-NN が 78.9%, 5-NN が 73.7% の検出率を示した。

計算時間の点で各手法を比較するため, 実験用データセット内の 1 つの動画に対する奥のプレイヤーのストローク検出に費やした計算時間を測定した。その結果を表 1 に示す。動画に含まれるフレーム数は 1,059 枚であり, この値には追跡および MD の計算時間は含まれていない。この時間は 5 回同じ計算を行って測定された計算時間の平均である。実験に使用したコンピュータの仕様は Pentium4 クロック 3 GHz, メモリは 32 GB, OS は SUSE Linux 10 である。

同様に, LUT の有効性を検討するため, LUT を用いた場合と用いない場合における奥のプレイヤーのバックハンドストローク検出器の学習の計算時間を測定した。その結果を表 2 に示す。これらの値も同様に 5 回測定した計算時間の平均値である。

5.4 考 察

得られた ROC 曲線より, 単純組合せおよび改良型組合せ手法は Ke らの手法をほとんどの部分で上回り,

Efros らの手法の結果に近い, また一部では上回る性能を示したことが分かる。このことより, MD を使用することで Ke らのイベント検出器の性能を改善できることが確認できた。また, どちらも MD を使用しているにもかかわらず単純組合せ手法は改良型組合せ手法より性能が良いことが分かる。この理由として, 強検出器内の弱識別器数の違い, 弱識別器内の判定方法の違いが考えられる。弱識別器数は, 単純組合せ手法の検出器が 200 個に対し改良型組合せ手法は 50 個であり, これは両手法で用いる情報量を同じにするためであった。しかし, 情報量が同じであっても弱識別器の数が異なるため AdaBoost 強検出器内で最終判断を決定する際の重み付き多数決の投票数が 4 倍になる単純組合せ手法が有利になったと考えられる。また, 弱識別器の識別方法は従来の弱識別器がサンプルデータから決定された最適閾値による閾値処理であるのに対し, 4 次元特徴を扱う改良型弱識別器では計算コスト軽減のため正負サンプルの平均とのユークリッド距離の比較により判定を行っている。しかしこれは, 改良型弱識別器にマハラノビス距離などの, より適した識別法を導入することで改善されるものと考えられる。

誤検出について検討を行うと, 今回の実験は 1 人のプレイヤーの動作データで学習を行っているため正のサンプル数が少ない点があげられる。よって様々なゲームから様々なプレイヤーの動作データを収集して学習に用いることで, 性能の改善が見込まれる。なお, Viola らの顔検出器の欠点として, 学習サンプル数が少ないと性能が悪くなるということが知られている。これは Ke らの手法にもいえることであるが, この原因は画像全体に検出器を適用するため, 学習時に正と負のパターンを幅広く用意しておかなければならないためである。しかし, 提案するシステムはプレイヤーを追跡し検出器はプレイヤーの動きのみを見ているため, パターンがある程度限定され少ない学習データでも良い性能を示したものと考えられる。また, 他の原因として, 追跡誤差の影響があげられる。提案する検出器は前処理としてプレイヤーの追跡が必要なため, 追跡の誤差やノイズに影響されることがある。よって, よりロバストな追跡手法を導入することで性能が改善されると考えられる。

計算時間について比較を行うと, 表 1 より Efros らの手法の計算時間が他の AdaBoost に基づく手法と比較して非常に長いことが確認できる。また, 次元拡張型の弱識別器は構造が Ke らの弱識別器より複雑なため, 改良型組合せ手法の検出器は計算時間が他の 2 つの AdaBoost 検出器より長くなっていることも確認で

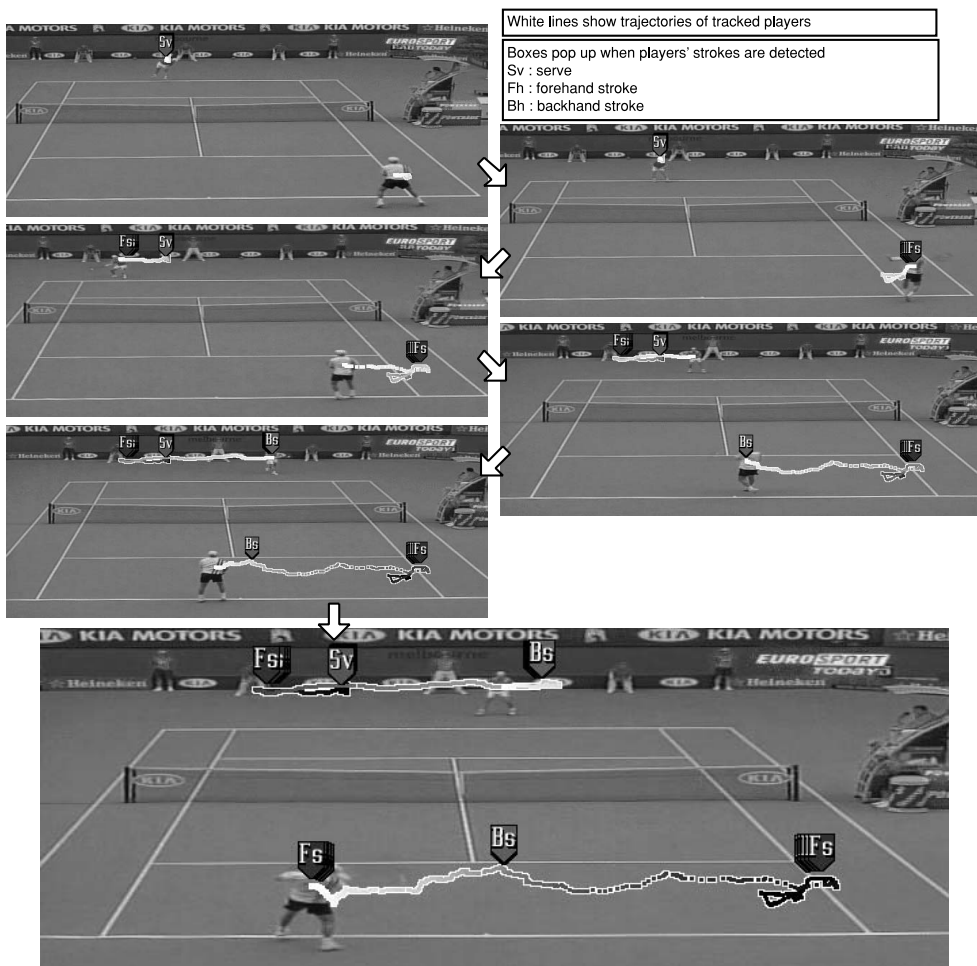


図 7 ストローク検出結果の例．白の軌跡はプレイヤーの移動軌跡，軌跡上の箱はストローク検出結果を表す．Sv の箱はサーブ，Fh の箱はフォアハンドストローク，Bh の箱はバックハンドストロークを表す

Fig. 7 An example of the stroke detection results. White lines show tracking trajectories of the players, and coloured boxes show the detected strokes. A box with 'Sv', a box with 'Fh' and a box with 'Bh' indicate serve, forehand stroke and backhand stroke, respectively.

きる．なお，単純組合せ手法の計算時間が Ke らの手法より若干長いのは，AdaBoost 検出器に含まれる弱識別器の数が異なるためと考えられる．

単純組合せ手法と改良型組合せ手法の検出結果および計算時間を比較すると，単純組合せ手法が若干良い結果を示している．しかし，表 2 に示される学習に必要な計算時間を比較すると，改良型組合せ手法は単純組合せ手法より LUT を使用しない場合で約 3 倍，LUT を使用すると約 7 倍早い．また，LUT の有効性について検討してみると，単純組合せ手法は LUT の使用によって約 8 倍，改良型組合せ手法では LUT の使用によって約 19 倍も学習に要する時間が早くなっている．改良型組合せ手法の学習時間が LUT の使用

によって大幅に軽減された理由として以下が考えられる．3.2 節に示すとおり，AdaBoost の学習では弱識別器を選択する際にすべての弱識別器のすべてのサンプルセットに対する演算が繰り返し何度も行われる．また前述のとおり，次元拡張型の弱識別器は Ke らの弱識別器より構造が複雑なため弱識別器個々の演算時間が長い．しかし，LUT を用いると弱識別器を選択する際の全弱識別器の繰返し演算処理が省略できる．したがって，提案する LUT は弱識別器個々の演算時間が長い改良型組合せ手法に特に有効となる．

学習で使用するメモリスペースについて検討すると，表 2 に示される奥のプレイヤーのバックハンドストローク検出器の学習に使用した学習サンプルの総数

は3,327, メモリ使用量は単純組合せ手法が792 MB, 改良型組合せ手法が306 MBであった。改良型組合せ手法のメモリ使用量が単純組合せ手法より少なかった理由は, (booleanのサイズ) × (正負の学習サンプル数の合計) × (学習に用いる弱識別器数) で決定されるLUTのメモリサイズのうち, 学習に用いる弱識別器数が単純組合せ手法の4分の1でよいためである。なお, 同じ学習プログラムをPentium4クロック3 GHz, メモリ2 GBのPCで実行したが, 単純組合せ手法の学習は多くのメモリを必要とするため学習時間が平均11,897 [秒] となり若干長くなったのに対し, 改良型組合せ手法の学習時間は平均1,257 [秒] とほぼ同じであった。

提案する動作検出システムはプレイヤー追跡の初期化が手動であるので, 完全自動化するためには自動初期化手法が必要となる。しかし, テニスはゲームの初めに決められた位置からサーブするためプレイヤーの初期位置は推定しやすく, それほど困難な課題ではないといえる。初期化の問題が解決されれば, ビデオ映像からプレイショットを自動で切り出すアルゴリズム¹⁾と組み合わせ, テニスの自動解説や戦略解析に応用できると考えられる。検出器の学習に様々なデータのサンプルを用いて性能を高め, 今回の実験で用いた3種類のストローク以外にスマッシュやボレーなどの検出器を構築すれば実用的なシステムとなると考えられる。

実験結果の例を図7に示す。図の白い軌跡はプレイヤーの追跡の履歴を, 軌跡上の記号と色のついた箱はプレイヤーのストロークが検出された場所を表す。‘Sv’ と書かれた箱はサーブ, ‘Fh’ と書かれた箱はフォアハンドストローク, そして‘Bh’ と書かれた箱はバックハンドストロークを表す。この動画データは奥のプレイヤーのサーブから, フォアハンドストローク, フォアハンドストローク, バックハンドストローク, バックハンドストローク, そしてフォアハンドストロークというストロークが含まれ, これらが正しく検出されていることが分かる。

6. 結 論

本論文では, Efros らのモーションディスクリプタとKe らのイベント検出器の組合せに基づく動作認識手法を提案した。単純組合せ手法が持つ学習時間の増大という問題を解決するため, 次元拡張型の弱識別器を導入した改良型組合せ手法を提案し, またさらなる高速化のためLUTを導入した。実験では実際の放送に用いられたテニスの動画像に提案手法を適用し, Efros らの手法, Ke らの手法および単純組合せ手法の

結果と比較することで提案手法の有効性を検討した。実験結果より, 組合せ手法は認識率と計算時間の両面でお互いの良い性能をあわせ持つことが確認できた。また, 改良型組合せ手法は単純組合せ手法とほぼ同じ性能を示しながら, LUTとの併用により学習に要する計算時間を大幅に削減できることが確認できた。

提案手法を用いれば, ユーザーがテニスの試合映像を多面的に理解する手助けとなる様々なアプリケーションが可能となる。たとえば, テレビ放映中に映像とともにどのプレイヤーがどの時点でどのストロークをしたかという情報を提供する, 大量にあるテニス映像の中から特定のプレイヤーの動作をすべて検出しそのプレイヤーの癖やストロークの傾向を調べる, また逆に特定のストロークを含むすべての動画を自動的に取り出しプレイヤーの教育に用いる, などがあげられる。また, 提案手法はスポーツ映像以外の人の動作検出にも適用可能である。特に提案手法は低解像度環境での動作認識が可能であり検出のアルゴリズムも複雑ではないため, たとえばセキュリティ分野の監視カメラ映像などに適用でき, 人の動作を含む映像を多面的に理解する手助けとなると考えられる。

改善すべき提案手法の問題点として, 各検出器が独立して動作するため, 異なる動作の検出結果が重複してしまう点があげられる。この重複を防ぐためには検出結果を統合するための後処理が必要となると考えられる。

参 考 文 献

- 1) Christmas, W., Kostin, A., Yan, F., Kolonias, I. and Kittler, J.: A system for the automatic annotation of tennis matches, *Proc. 4th international workshop on content-based multimedia indexing* (2005).
- 2) Efros, A., Berg, A., Mori, G. and Malik, J.: Recognizing action at a distance, *Proc. ICCV*, pp.726–733 (2003).
- 3) Gavrilu, D.: The visual analysis of human movement: A survey, *CVIU*, Vol.73, No.1, pp.82–98 (1999).
- 4) Isard, M. and Blake, A.: A smoothing filter for condensation, *Proc. ECCV*, pp.767–781 (1998).
- 5) Ke, Y., Sukthankar, R. and Hebert, M.: Efficient visual event detection using volumetric features, *Proc. ICCV*, pp.166–173 (2005).
- 6) Nummiaro, K., Koller-Meier, E. and Gool, L.V.: A color-based particle filter, *Proc. 1st International Workshop on Generative-Model-Based Vision (GMBV'02)*, pp.53–60 (2002).
- 7) Rea, N., Dahyot, R. and Kokaram, A.: Classi-

fication and representation of semantic content in broadcast tennis videos, *Proc. ICIP*, pp.III: 1204-1207 (2005).

- 8) Viola, P. and Jones, M.: Robust real-time object detection, *Proc. IEEE Workshop on Statistical and Computational Theories of Vision* (2001).
- 9) Wang, J. and Parameswaran, N.: Analyzing tennis tactics from broadcasting tennis, *Proc. Intl' Multimedia Modelling Conf.*, pp.102-106 (2005).
- 10) Yamato, J., Ohya, J. and Ishii, K.: Recognizing human action in time-sequential images using hidden markov model, *Proc. IEEE Conf. on CVPR*, pp.379-385 (1992).
- 11) Zivkovic, Z., vd Heijden, F., Petkovic, M. and Jonker, W.: Image processing and feature extraction for recognizing strokes in tennis game videos, *Proc. 7th Annual Conference of the Advanced School for Computing and Imaging (ASCI2001)*, pp.262-267 (2001).

(平成 18 年 6 月 13 日受付)

(平成 18 年 12 月 7 日採録)



緒方 健人 (学生会員)

2002 年九州工業大学工学部卒業。2003 年同大学大学院工学研究科博士前期課程を修了し、現在、博士後期課程に在学中。2005 年 4 月から 1 年間連合王国サリー大学客員学生研究員。2000 年度電気学会九州支部長賞、2004 年度計測自動制御学会九州支部奨励賞、2004 年度 IEEE 福岡支部学生研究奨励賞受賞。ヒューマン・コンピュータインタラクション、パターン認識、コンピュータビジョンに関する研究に興味を持つ。計測自動制御学会、IEEE 各会員。



ウィリアム クリスマス

連合王国オックスフォード大学卒業後、BBC および BP Research International にて映像放送技術、並列処理、実時間画像処理の研究に従事。現在は連合王国サリー大学にて博士号を取得し幾何学的特徴のマッチングに関する研究、マシンビジョンを用いたスポーツ映像の自動解説の研究等に従事。



ジョセフ キットラー

連合王国サリー大学教授。同大学視覚音声信号処理センター長。個人認証、自動検査、ターゲット抽出、ビデオ符号化および探索、視覚認識等を含むパターン認識、画像解析の研究に従事。Prentice-Hall 社“Pattern Recognition: A statistical approach”の共著者。Image and Vision Computing, Pattern Recognition Letters, Pattern Recognition and Artificial Intelligence, Pattern Analysis and Applications の編集委員。



ジュークイ タン (正会員)

九州工業大学大学院設計生産工学専攻博士後期課程修了(工学博士)。同大学工学部 SVBL ポストドク研究員を経て、現在、同大学工学部機械知能工学科制御工学教室助手。98 年度計測自動制御学会九州支部奨励賞、2004 年度情報処理学会九州支部第 13 回奨励賞、同年度知能生命・ロボット国際シンポジウム (AROB2005) 若手奨励賞受賞。主に画像計測および解析、パターン認識、コンピュータビジョン等に関する研究に従事。IEEE、計測自動制御学会各会員。



石川 聖二

東京大学工学部計数工学科卒業。同大学大学院工学系研究科博士課程修了。工学博士。九州工業大学情報工学科講師を経て、現在、九州工業大学工学部教授(機械知能工学科)。この間、連合王国シェフィールド大学客員研究員。オランダ王国ユトレヒト大学客員教授。3次元画像計測、医学画像処理の研究に従事。計測自動制御学会、電子情報通信学会、画像電子学会、IEEE 各会員。