

スポーツ映像解析ソリューション

Sports Video Analysis Solution

田 靡 雅 基
Masamoto Tanabiki

古 山 純 子
Junko Furuyama

齋 藤 浩
Hiroshi Saito

竹 中 慎 治
Shinji Takenaka

関 井 大 気
Taiki Sekii

要 旨

近年、実際の試合映像から選手のさまざまな情報を抽出するスポーツ映像解析が重要視されている。そのなかでも、ポジショニング分析や運動量換算の基礎データとなる選手の位置情報は特に重要な情報である。特にチームスポーツ形式の競技において、試合状況に応じた戦術変更を可能とするリアルタイムおよび取得する時間間隔を高密度化した詳細な解析への期待が大きい。従来のスポーツ映像解析は、手動入力に頼る解析が主流であり、リアルタイムかつ密な時間間隔での解析を同時に実現する場合に膨大なコストを必要とする。これに対し、筆者らは、90%以上の精度で複数選手を同時に追跡し、自動で位置情報を取得可能な画像認識技術を開発した。画像認識技術を搭載するシステムと解析者との相互作用において、解析者の少ない、簡単な操作だけで試合中に詳細な解析が可能な映像解析ソリューションを提案する。

Abstract

Recently, sports video analysis has been attracting attention. It is very important to have geodata on some players because they are primary data to evaluate the players' positioning and amount of motion. In team sports, real-time and detailed analysis is expected.

However, some conventional sports image analyses rely on manual input by staff. With manual input, it is difficult to realize simultaneously both real-time digitization and detailed analysis. We have tackled this problem by developing new image-recognition technology. Through interaction between a staff member and an image-recognition system, just the minimal amount of simple operations by the staff member can realize both real-time and detailed analysis in a match.

1. はじめに

近年、実際の試合映像から選手やボールの位置情報を取得し、位置情報を活用して選手個人の運動量や、選手同士のポジショニングを可視化するスポーツ映像解析が注目されている。しかしながら、従来のスポーツ映像解析は、選手やボールの位置情報を手動で抽出し、人手による膨大なコストを必要とするため、自動で位置情報を取得できるシステムの需要が高まっている。

特に、競技に勝つことが重要であるスポーツチームから、試合状況に応じた戦術変更を可能とするソリューションへの期待が大きい。映像と解析結果をセットにして監督・コーチから選手に指示を出すことで説得力が増し、コミュニケーションが円滑になることが実証されている。

人物を対象とした自動の映像解析には、歩行者を主な対象とし、監視カメラ映像から人物領域を切り出し（検出）、同一人物の位置変化を把握（追跡）する画像認識技術がよく知られている[1]。しかしながら、スポーツ選手がフィールドを四方八方に、速度を変えて移動しつつさまざまな姿勢をとるため、スポーツ映像解析には監視カメラ映像解析と異なる手法が求められる。

また、例えばサッカーの場合、チーム全体での戦術共有に最適なハーフタイムまでに試合前半分の解析を完了

させる即時性が必要となる。

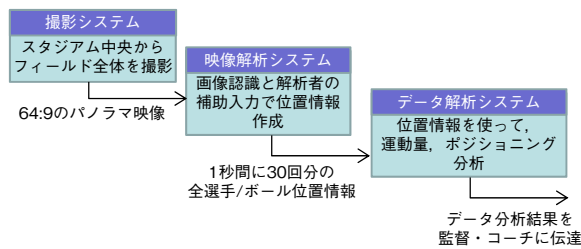
そこで、筆者らは、選手の移動に伴う選手間遮蔽や、プレー中の姿勢変動に頑健な画像認識技術を考案した。さらに、ボールも解析対象に加え、画像認識技術と専門知識不要の簡単な人的補助を組み合わせ、即時処理を可能とするスポーツ映像解析システムを開発した。

次に本稿の構成を示す。第2章では、スポーツ映像解析システムの概要について述べる。続く第3章では、スポーツ映像解析の中核である画像認識技術について、技術課題と解決手法について論じる。第4章では、データ活用要望が世界規模で最大となるサッカーを対象に画像認識技術を適用した評価結果を述べる。最後の第5章で本取り組みをまとめる。

2. スポーツ映像解析システムの概要

提案するスポーツ映像解析システムは3つのサブシステムで構成される。全選手・ボールが散在するフィールド全体を1枚のパノラマ映像として撮影するシステム、パノラマ映像を画像認識技術で解析し、位置情報生成の担当者（以下、「解析者」）の補助入力を経て全選手・ボールの位置情報を1秒間に30回取得する映像解析システム、位置情報から選手個人の運動量や選手同士の連動性を算

出するデータ解析システムである（第1図）。

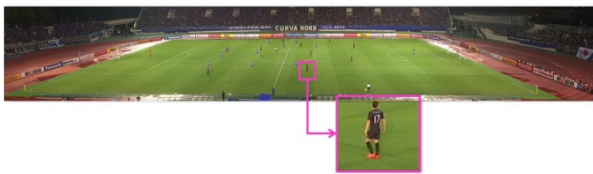


第1図 スポーツ映像解析システムのデータフロー

Fig. 1 Data flow in system to analyze sports video

2.1 撮影

撮影にはUltra Wide Angle Camera (UWAC)を活用する。UWACは、4台のHDカメラがとらえた映像をリアルタイムでステッチング処理し、4倍ワイド（64：9画角）かつ継ぎ目のないパノラマ映像を生成する（第2図）。



第2図 解析対象となるスポーツ映像

Fig. 2 Sports video for analysis

フィールド全体をとらえることで、全選手の動きやフォーメーションを一望できる。さらにHD×4倍の高解像度は、拡大すれば背番号まで確認できる高画質であり、特定の選手の細かな動きの確認にも適している。また、監督・コーチが選手に指示する場合に、両者による映像確認によって、データ分析結果と選手の主観・記憶との差を容易に埋めることができるため、大域的・局所的な指示・理解を促進できるUWAC映像の有効性が実証されている。

2.2 映像解析

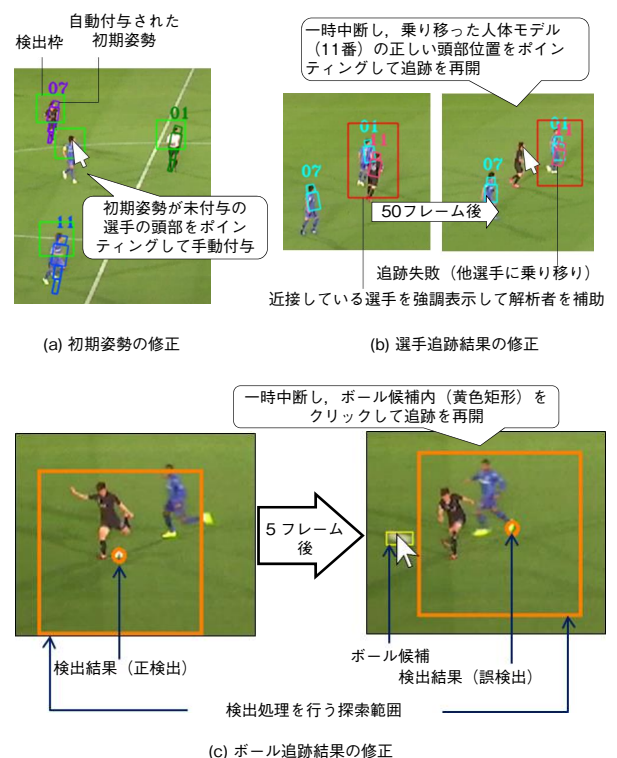
本システムではUWACで撮影した映像を15秒ずつにファイル分割し（以下、「パック」）、1秒あたり30フレームのカラー画像を含むパックごとに解析を行う。これは、多くのチームスポーツにおいて、1回あたりのインプレー持続時間が15秒以下となることに基づいている。

選手位置情報生成において、1パックあたり30秒以内で位置情報を生成できれば、連続する2パックの1パックずつを2人が循環させて担当することで、試合中に30秒ごとのデータ更新が可能となる。また、ボール位置情報生成

において、1パックあたり、15秒以下で位置情報を生成できれば、1人で担当することができる。これにより、サッカーチームが要望する、ハーフタイムまでに試合前半の解析を完了させることができる。

映像解析システムは、解析した選手位置・ボール位置を選手形状・ボール形状に模した図形（以下、「解析結果」）で表現し、カラー画像に重畳して出力する。第3図は、解析者が映像解析システムとインタラクションするための画面および作業例である。同図については4.2節で詳述する。解析者が、選手やボールと解析結果との位置ずれを目視確認すると、映像解析システムに解析の一時中断を指示する。続いて、解析者が正しい位置を入力した後、映像解析システムに再開を指示する。このように、解析者が解析結果に対して簡単な修正を施すだけで、映像解析システムは正確な位置情報を逐次に生成する。

なお、位置情報の自動生成を実現する選手・ボール追跡技術の詳細は第3章で述べる。



第3図 映像解析システムの画面例

Fig. 3 User interface of sports video analysis

2.3 データ解析

映像解析によって取得した一次情報（選手位置情報・ボール位置情報）を単独、もしくは組み合わせることで第1表に示す二次情報を算出し、運動量や連動性の客観的な評価が可能となる。運動量を示す走行距離やスプリ

ント回数を手がかりに、選手・コーチが、期待するパフォーマンスを発揮していない選手の交代を決定したり、パス成功率や選手間距離を参照して、選手同士がポジショニングを再確認したり、競技に勝つための意思決定ツール、コミュニケーションツールとして現場で活用される。

第1表 データ解析の事例

Table 1 Examples of analyzed data

一次情報	二次情報	
選手位置	運動量	走行距離・速度、スプリント回数など
	連動性	プレーエリア、選手間距離など
選手位置+ ボール位置	運動量	ボール支配率、ボール保持時間など
	連動性	パス本数・成功率、パス主要ルートなど

3. パノラマ映像を活用した映像解析技術

本章では、映像解析システムの中核である画像認識技術について説明する。

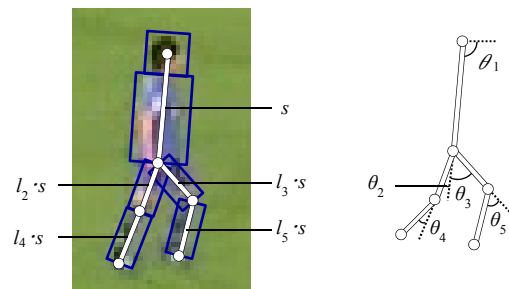
3.1 選手検出

選手検出処理は、入力画像から特徴量と呼ばれるベクトルを抽出し、その特徴量から選手が含まれるかどうかを判定する検出器を用いて、検出（人物が存在する領域の切り出し）を行う。例えば[2]では、監視カメラ映像に対し独自特徴量を使った高精度な人物検出を示している。画像解析システムは、選手検出によって、選手の頭部を中心に上半身を内包する矩形（くけい）領域（以下、「検出枠」）を特定し、この検出枠を後続のチーム識別／初期姿勢付与処理で利用する。

3.2 チーム識別／初期姿勢付与

次節で述べる選手追跡において、バックの先頭画像で選手の初期姿勢を自動付与する必要がある。本節では、前節で得られた検出枠から選手の初期姿勢を推定する方式について述べる。また、追跡結果の位置情報をチームごとに選別するには、初期姿勢を付与する段階でチームを識別しなければならない。そのため、本節では初期姿勢付与と同時にチームを識別する方式についても述べる。

第4図に示すように、初期姿勢は、複数の関節と、関節をつなぐ直線により表現される。これらの関節のうち、まず頭部と腰のみを推定し、その他の関節位置については、頭部と腰位置を通る直線状の一定の位置に固定する。頭部・腰位置の推定にはPictorial Structures[3]と呼ばれる、機械学習に基づく姿勢推定アルゴリズムを用いる。最初に、選手を含む小領域を、検出枠の位置を基準として所定サイズの画像で切り出す。続いて切り出した選手画像



第4図 人体モデル

Fig. 4 Human shape model

からPictorial Structuresにより初期姿勢を推定する。

Pictorial Structuresでは、一般的な学習・識別アルゴリズムであるRandom Forestと動的計画法による2段階の手順により関節位置を推定する。Random Forestは、あらかじめ準備した2000枚の頭部・腰関節を含む画像および頭部・腰関節を含まない非関節画像を用いた学習によって構築され、選手画像の画素ごとに頭部・腰関節尤（ゆう）度（関節らしさ）を出力する。続いて、動的計画法を用いて、自画素の周辺画素の関節尤度が高く、かつ他の関節との距離が一定の範囲に収まる画素群を関節位置として求める。なお、事前に構築するRandom Forestは汎用的で、映像解析システムが稼動する現場での再構築・更新を不要とする。

以上の手順で得られた初期姿勢から、チームを次のように識別する。チーム識別には、Support Vector Machine (SVM) [4]と呼ばれる機械学習アルゴリズムを用いる。まず、頭部位置を基準として画像を切り出す。切り出すサイズは頭部・腰位置間の距離を基準として決定する。切り出した画像内の色がフィールド色と異なる領域（以下、「前景」）を活用して、前景の特徴量（色ヒストグラム）を抽出し、SVMに入力する。SVMは、入力された特徴量からチーム識別番号を出力する。なお、各チーム数人程度の選手画像をあらかじめ準備し、SVMを構築しておく必要がある。例えば、試合開始前に両チームの選手画像をUWAC映像から切り出してSVMを更新し、試合中に利用する。

3.3 選手追跡

選手追跡処理では、直前の画像における選手の位置・関節（部位）情報を基に、現在の画像上の選手の部位位置を推定する。

スポーツ映像では、複数人の選手が独立に複雑な姿勢・動きをするため、人体部位が遮蔽状態になることが多々あり、これに起因して誤追跡が発生する。本追跡方式では、前述の通り複数の部位情報を利用しているため、

一部の部位に遮蔽や姿勢変動が生じて他部位に重みを付けることで頑健な追跡が可能である。また、選手のさまざまな姿勢の部位情報をあらかじめ学習しモデル化することで、人体部位の連結関係に統計的な制約をもたせ、人間らしい動きに絞った効率的で高精度な追跡を実現している。

〔1〕人体モデル

追跡に用いる尤度として部位ごとの色ヒストグラムの類似度を用いているため、追跡には各部位の色情報を取得する際の部位を内包する矩形情報が必要となる。そこで、人体モデルを生成するにあたり、モデルの姿勢パラメータを以下のように定義した（第4図）。

$$\mathbf{x}=(\mathbf{u}, s, \mathbf{l}, \boldsymbol{\theta}), \quad \mathbf{l}=(l_2, l_3, l_4, l_5),$$

$$\boldsymbol{\theta}=(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$$

ここで \mathbf{u} は腰位置座標、 s は頭部中心から腰位置までの長さ、 \mathbf{l} は各部位長の s に対する比率($l_2 \sim l_5$)、 $\boldsymbol{\theta}$ は腰関節の絶対角度 θ_1 と各部位間の相対角度 $\theta_2 \sim \theta_5$ を表す。ただし、 $\boldsymbol{\theta}$ は左回りを正とする値である。これらのパラメータは追跡に必要な矩形情報と各部位間の相対関係を表現する。なお、各部位矩形幅は部位長から一意に求まるものとし固定パラメータとして定義した。以上を主成分分析によって寄与度の高い5主成分まで抽出し姿勢状態空間の基底ベクトルとして定義した。

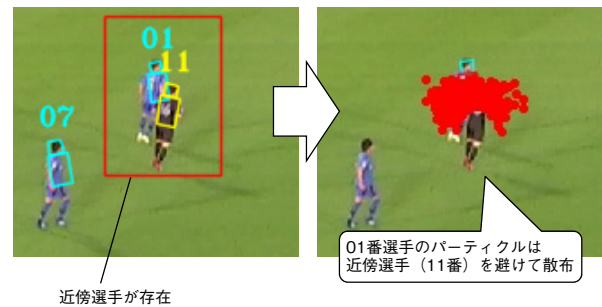
〔2〕追跡方式

学習した人体モデルが定義されている姿勢状態空間内をパーティクルフィルタ(PF)によって探索することで、選手の位置・姿勢を推定する。PFは、推定したい状態空間内に多数のパーティクルと呼ばれるサンプル点を散布し、各パーティクルにおける尤度を算出し、その尤度分布を状態確率として推定する手法である。パーティクル散布、尤度計算、状態推定を繰り返すことで効率的に探索することが可能となる。

パーティクルは現在より過去の画像において推定した状態確率に応じて散布される。すなわち、尤度値が高いパーティクルが密集した領域に多く散布される。そのため、追跡対象選手の近傍に同チーム(ユニフォームが同一)の他選手がいた場合、その選手付近の尤度値も高くなり誤追跡が発生する確率が高くなる。

これに対し筆者らは、追跡対象選手が他の追跡中選手と重なっているか、または、近傍に他選手が存在するかを判定し、次のような3段階の調整を加えた。①近傍選手の周囲にパーティクルを散布しないよう制御することで、他選手領域で尤度が高くなる事象を削減する(第5図に01番と11番選手が近づいた際の01番選手のパーティクル散布位置を右図赤点で示す)。②人体部位の大半が他の追跡選手と重なっており尤度が低い場合は、他チーム選手

の陰に隠れたとみなし、過去移動量を元に線形的にパーティクル散布範囲を絞り込む。③近傍選手が存在する場合は、次状態を推定する際に使用するパーティクルの数を削減し、より確からしい状態確率に基づいてパーティクルを散布する。以上の調整により、人物の遮蔽や姿勢変化に頑健な追跡を実現した。



第5図 近傍選手を避けたパーティクル散布

Fig. 5 Example of scattered particles

3.4 ボール追跡／検出

ボールは小サイズで遮蔽も多く、選手の足・靴と色・形が類似しているため、誤追跡／誤検出しやすいという課題がある。そこで本方式では、前景からボールらしいサイズの前景領域をボール候補として抽出し、移動距離・サイズ・色情報を特徴量として最尤のボール候補を探索する追跡処理を基本とする。

しかしながら、ボールが選手に近接すると、ボールの前景と選手の前景が結合する。結合によってボールらしいサイズより大きな前景領域となることが原因でボール候補が得られず、追跡が困難となる。そこでボールらしいボール候補が得られない場合は、過去の画像のボール位置周辺に探索範囲を限定して検出処理を行う。具体的には、ボール位置周辺のカラー画像を用いて、色ヒストグラムを特徴量としたテンプレートマッチングを行い、色情報が類似する領域をボール領域として検出する。

ボールが完全に遮蔽された場合は検出もできないため、現在の画像に対し、追跡／検出ができた直前の画像におけるボール位置と、追跡／検出ができた直後の画像におけるボール位置とから、補間により位置を算出する。

このように、追跡、検出、補間という3つの処理を組み合わせることで、小さくかつ遮蔽されることの多いボールの位置を算出する。UWAC映像ではフィールドの奥と手前でボールの見え方の変化が大きく、特に奥のボールが極小(直径4画素程度)となって誤検出が増加する。そこで、テンプレートマッチング計算パラメータをボール位置に応じて自動変更する機能や、ボール周辺領域の色情報を用いた誤検出除去機能を追加することで、見え方

の変化が大きい映像においてもボール位置を算出可能とした。

4. 解析者の作業負担評価

本章では、映像解析システムとの相互作用によって位置情報を生成する解析者の作業負担について評価を行う。

4.1 評価方法

サッカープロリーグの公式戦を撮影し、無作為に抽出した40パック（10分間に相当）を評価対象とする。

具体的に、映像解析システムが出力する画像認識結果に対し、解析者が施す修正回数を計測する。さらに、事前に手動入力によって全選手・ボールの正解位置情報と、許容される誤差（＝画素数）を40パック分の全画像にわたってリスト化し、画像認識と解析者の相互作用によって作成した位置情報との差分を基に、生成した位置情報の正確さを確認する。

4.2 目標とする作業量

2.2節で述べたように、1パックに対し、選手位置情報生成に30秒以内、ボール位置情報生成に15秒以内を目標とする。

選手位置情報生成において、30秒の内訳は、画像認識処理時間15秒、修正作業15秒を目安とする。解析者は、画面に表示された画像認識結果を確認しつつ、①追跡処理開始前と、②追跡処理中、に修正作業を行う。画面上、チーム別に色分けされた人体モデルと選手IDを1組とした解析結果が、画像上の全選手に重畳されている。修正作業は、①初期姿勢が未付与の選手を確認し、選手頭部位置のポインティングによって初期姿勢を付与する作業と、②人体モデルが他選手に乗り移った際、乗り移った選手IDを選択し、人体モデルを再付与するべき選手頭部位置をポインティングする作業、に大別される。①②の例を第3図 (a) (b) にそれぞれ示す。いずれも、1選手あたり平均3秒要することを実験で確認しており、①②を合計して5回以下が修正作業15秒以下を実現する目標となる。

またボール位置情報生成における15秒の内訳は、画像認識処理時間10秒、修正作業5秒を目安とする。複数IDを区別する選手追跡と違って、ボールのID選択作業は不要となる。第3図 (c) で示すような、追跡を再開するべきボール位置へのポインティング作業を平均2秒で実施できるため、修正作業2回以下を目標とする。

4.3 評価結果

第3章で述べた画像認識の精度が間接的に影響し、映像解析における解析者の作業コスト（手間）の直接指標となる修正必要回数を計測する。

〔1〕選手追跡開始前の修正必要回数

各パックの先頭画像に対し、映像解析システムは選手検出に続いてチーム識別／初期姿勢付与を自動で実施し、選手ごとにIDと人体モデルで表現した初期姿勢（以下、「付与結果」）を選手に重畳して画面に表示する。

画面出力されたID、初期姿勢の少なくとも1つを間違っている選手が1人でも存在すると、解析者は全選手の付与結果を確認する作業を余儀なくされる。時間制約が厳しい映像解析において、この作業時間を無視できないため、映像解析システムはチーム識別あるいは初期姿勢付与の信頼度が低い選手に対し、付与結果を出力しない。これにより、解析者は付与結果が重畳されていない選手を探索する作業に集中できる。

実験において、修正必要回数は1パックあたり平均2.2回であった。

〔2〕選手追跡中の修正必要回数

各パックの先頭画像に対し、初期姿勢の修正作業を完了後、すなわち全選手に初期姿勢が付与された状態から追跡を開始する。

映像解析システムは、追跡対象とするパックの先頭画像から最終画像に至る全画像において、全選手に対して選手個別のIDと人体モデルを重畳して画面表示する。解析者は、同一選手に同一IDが付与され続けること、選手と人体モデルの腰位置の差が許容誤差以内であることを目視確認し、いずれかに異常が発生した場合に、修正を施す。

実験において、修正必要回数は1パックあたり平均1.1回であった。さらに、全選手の位置が全パックで許容誤差以下となることを確認した。

〔3〕ボール追跡／検出

解析者がパックの先頭画像に手動でボール位置を付与した後、映像解析システムに追跡／検出開始を指示する。

映像解析システムは、全画像に対し、認識したボール位置とボールサイズを基に円形状の図形をボールに重畳して画面表示する。解析者は、画面表示された図形とボールの位置が一致しているかを目視確認し、位置ずれが大きいと主観で判断したときにボール位置を手動で再付与し、追跡／検出を再開させる。

実験において、修正必要回数は1パックあたり平均0.7回であった。さらに、ある画像におけるボール位置と、その画像を中心とした前後数枚の画像におけるボール位

置を連結したボール軌跡の類似度計算において、全パックで許容誤差以下となることを確認した。

以上により、選手位置情報生成における1パックあたりの修正作業回数は合計で平均3.3回、ボール位置情報生成は平均0.7回となった。これにより、解析者の少なく、簡単な操作だけで許容誤差以内の正確な位置情報を試合中、即時に生成できることを確認した。

5. まとめ

本稿では、サッカーの試合中に、高精度な選手やボールの検出／追跡と、解析者の少なく、簡単な操作によって位置情報を取得し、チームが勝つための戦術修正を可能とするスポーツ映像解析ソリューションについて述べた。

しかしながら、解析者が修正箇所の発見に高い集中力を維持しなければならず、解析者の“余裕のなさ”が現状の映像解析ソリューションの課題である。これに対し、例えば、解析者の修正内容から検出／追跡の失敗パターンを映像解析システムが自動獲得し、試合経過に応じて検出／追跡の失敗を極小化するように改善していきたい。これにより、試合経過に応じて解析者の修正回数が減り、作業工数をチームが勝つための別作業に割り当てられる時間的・精神的余裕を生み出していく。

さらに今後は、人体モデルの特徴を活かし、サッカー以外の展開と、獲得する情報の高度化を目指す。

人体モデルはスポーツ種目固有の動き・姿勢を表現できる汎用性を備えており、アメリカンフットボール、バスケットボール、バレーボール、アイスホッケーなどとの親和性が高い。また、スポーツチームは、位置情報を基礎データとした分析を確立した後、プレー中の姿勢・手足に着目した動作分析を目指している。提案する人体モデルは、複数部位で構成されており、部位間の位置関係や、部位内の関節位置を把握することに適している。この特徴を活かし、姿勢・手足の動きをとらえた高度な動作分析を実現していきたい。

参考文献

- [1] 橋本潔 他, “姿勢変動に伴う身体特徴変化の統計的モデリングによる遮蔽に頑健な人物追跡”, ViEW2011, no.B-30, pp.29-30, 2011.
- [2] H. Cher Keng et al., “Shrink boost for selecting multi-LBP histogram features in object detection,” In CVPR, pp.3250-3257, 2012.
- [3] M. Dantone et al., “Human pose estimation using body parts

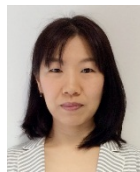
dependent joint regressors,” In CVPR, 2013.

- [4] J. Weston et al. “Feature selection for SVMs,” Advances in Neural Information Processing Systems 13. MIT Press, 2001.

執筆者紹介



田摩 雅基 Masamoto Tanabiki
AVCネットワークス社
イノベーションセンター
Innovation Center, AVC Networks Company



古山 純子 Junko Furuyama
AVCネットワークス社
イノベーションセンター
Innovation Center, AVC Networks Company



齋藤 浩 Hiroshi Saito
AVCネットワークス社
イメージングネットワーク事業部
Imaging Network Business Div.,
AVC Networks Company



竹中 慎治 Shinji Takenaka
(株) パナソニック システムネットワークス開
発研究所
Panasonic System Networks R&D Lab. Co., Ltd.



関井 大気 Taiki Sekii
(株) パナソニック システムネットワークス開
発研究所
Panasonic System Networks R&D Lab. Co., Ltd.