

LI-003

サッカー映像のフォーメーション解析に基づく実時間イベント検出 Real-time Event Detection Based on Formation Analysis of Soccer Scenes

三須 俊彦[†] 高橋 正樹[†] 蓼沼 眞[†] 八木 伸行[†]
Toshihiko Misu Masaki Takahashi Makoto Tadenuma Nobuyuki Yagi

1. まえがき

近い将来に運用開始されるサーバ型放送 [1] では、番組を意味的なまとまりに区分し、その内容を記述したメタデータの付与 [2] が必須となる。多様な状況やプレイ (以降、イベント) が発生するスポーツ番組用メタデータは、サーバ型放送のほか、映像データベース (NHK が保有する放送アーカイブ等) における映像検索、ダイジェスト番組制作、番組・オブジェクト連動データ放送 [3] のコンテンツ制作など幅広い応用も期待される。しかし、シナリオのないスポーツ映像に対しメタデータを全て手動付与するには莫大な時間、労力、コストを要する。各種スポーツコンテンツにおけるイベント検出を自動化するため、音響解析や言語解析 [4] によりイベントを検出する手法、画像解析による手法 [5][6][7]、それらを複合させた手法 [8] などが提案されている。サッカーなどのフィールド競技では、特に選手のフォーメーション (コート内における選手分布) とその動きが各種の状況の特徴づけることから、選手位置 [9][10][11] や撮影アングル [12][13] に注目したイベント検出手法も多い。

本稿では、固定ハイビジョンカメラで撮影された映像から選手の位置・速度、所属チームを解析し、コーナーキックなどのイベントを実時間検出する手法を提案し、その実験結果を示す。

2. 処理の流れ

今回開発したイベント検出手法は、図 1 に示すように、低レベル特徴抽出、高レベル特徴抽出、イベント判定、ポストフィルタ、およびメタデータ記述の 5 段階で構成される。低レベル特徴抽出部では、各選手の座標などの物理量を抽出する。高レベル特徴抽出部では、抽出された物理量に対し統計処理を施し、チーム全体での重心などの統計量を算出する。続いて、IF~THEN ルールに基づくイベント生起判定と時間フィルタ処理を行い、その結果を XML により記述して出力する。

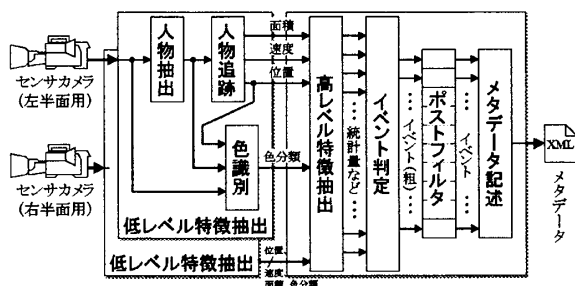


図 1: 処理の流れ

[†]日本放送協会, NHK

表 1: 高レベル特徴量

統計量	変数名
選手人数	n
選手の密集度	ρ
全選手の重心位置	g
{ 左側, 右側 } チームの重心位置	$\{g^{(L)}, g^{(R)}\}$
全選手の平均速度ベクトル	v
全選手の平均速さ (速度の絶対値の平均)	β
タッチライン上選手の速さ	u
タッチライン上選手と g との距離	d
自陣内にいる選手の割合	r
特定領域内の選手有無 (フラグ)	
{ 左奥, 左手前, 右奥, 右手前 } コーナーに攻撃側フィールド選手あり	$\{c_{lf}, c_{ln}, c_{rf}, c_{rn}\}$
{ 奥, 手前 } タッチラインに { 左側, 右側 }	$\{t_f^{(L)}, t_n^{(L)}\}$
チームのフィールド選手あり	$\{t_f^{(R)}, t_n^{(R)}\}$

2.1 低レベル特徴抽出

低レベル特徴抽出部では、各人物 (厳密には各シルエット) i の位置 $x^{(i)}$ 、速度 $\dot{x}^{(i)}$ 、色分類 $c^{(i)}$ 、シルエット面積 $a^{(i)}$ を抽出する。色分類 $c^{(i)}$ としては、左側チームのフィールド選手、左ゴールキーパー、右フィールド選手、右ゴールキーパー、審判、その他人物の 6 カテゴリを設け、それぞれ 0~5 の値を割り当てた。また、オクルージョン時に各シルエットに含まれる人数を推定すべく、シルエット面積 $a^{(i)}$ は距離に応じて正規化した。

なお、屋外環境で撮影される映像を頑健かつ高速に処理するため、筆者らがサッカー中継の映像効果用に実用化したオフサイドライン仮想表示装置 [14] における選手追跡、照明変動対応、実時間処理の各技術を用いている。

2.2 高レベル特徴抽出

高レベル特徴抽出部では、低レベル特徴量に対して統計処理を施すことで、表 1 上の各種統計量を算出する。たとえば、密集度 ρ および平均速さ β は、フリーキックやコーナーキックの際に、選手が密集して「壁」をつくり、各選手がほぼ静止するという特徴に関連した統計量である。以降では、下付き添え字により、イベントの生起した場所または方向性 (f : 奥, n : 手前, l : 左, r : 右) を表すこととする。また、上付き添え字により、動作主の所属チーム (L : 左側チーム, R : 右側チーム) を区別する。

選手人数 n および密集度 ρ は、下式により定義する。

$$n = \sum_{i \in \{j | c^{(j)} \in \{0, 2\}\}} a^{(i)} \quad (1)$$

\leftarrow フィールド選手のための総和

$$\rho = \frac{n}{\sqrt{\det[V + \sigma^2 I]}} \quad (2)$$

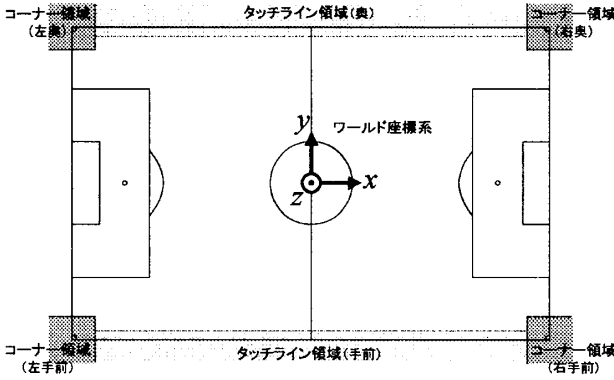


図2: コーナー領域とタッチライン領域

$$V = \sum_{i \in \{j|c^{(j)} \in \{0,2\}\}} [x^{(i)} - g][x^{(i)} - g]^T / \sum_{i \in \{j|c^{(j)} \in \{0,2\}\}} 1 \quad (3)$$

ここに σ は正の定数、 I は2行2列の単位行列である。式(1)に示す選手人数 n の算出においては、オクルージョンの影響で1シルエットに複数選手が含まれる場合に対応するため、シルエット数を単純に数えるのではなく、正規化された面積 $a^{(i)}$ に基づいて n を算出している。一方、式(2)の分母は選手の散らばり具合を面積の次元で定量化したものとなるため、密集度 ρ は人口密度の次元をもつ。

n および ρ 以外の統計量は、全選手あるいは該当チームの選手に関して位置や速度の平均をとるなど、単純に計算することができる。

また、上記の統計量のほか、特定領域(図2)に特定チーム選手が入ったときに TRUE となるフラグも、表1下に示すように高レベル特徴量として定義する。これらのフラグは、コーナーキックやスローイン時に特有のフォーメーションを捉えるのに有効である。

2.3 イベント判定

イベント判定部では、高レベル特徴抽出部で得られた統計量に対して閾値処理を行い、その結果に対して論理演算を施すことで、表2に示す各種イベントの生起を判定する。以下に判定ルールを示す。

$$K = (n \geq N_K) \wedge (\beta \leq B_K) \wedge (r \geq R_K) \wedge (g_x^{(L)} \leq -\Gamma_K) \wedge (g_x^{(R)} \geq +\Gamma_K) \quad (4)$$

$$F = (n \geq N_F) \wedge (\rho \geq P_F) \wedge (\beta \leq B_F) \quad (5)$$

$$C_{\bullet}^{(o)} = (n \geq N_C) \wedge (\rho \geq P_C) \wedge (\beta \leq B_C) \wedge (|g_x| \geq G_C) \wedge c_{\bullet} \quad (6)$$

$$T_{\bullet}^{(o)} = (n \geq N_T) \wedge (\rho \geq P_T) \wedge (\beta \leq B_T) \wedge (u \leq U_T) \wedge (d \leq D_T) \wedge t_{\bullet}^{(o)} \quad (7)$$

$$M_{\ell}^{(R)} = (n \geq N_M) \wedge (v_x \leq -V_M) \quad (8)$$

$$M_r^{(L)} = (n \geq N_M) \wedge (v_x \geq +V_M) \quad (9)$$

$$A_{\ell}^{(R)} = (n \geq N_A) \wedge (v_x \leq -V_A) \wedge (g_x \leq -G_A) \quad (10)$$

$$A_r^{(L)} = (n \geq N_A) \wedge (v_x \geq +V_A) \wedge (g_x \geq +G_A) \quad (11)$$

なお、各右辺の大文字 (N_K など) は閾値を、下付きの x はワールド座標のタッチライン方向成分を表す。ま

表2: イベントの種類

イベント	攻撃チーム	場所・方向	変数名
キックオフ		区別なし	K
フリーキック		ボールを検出しないため区別できず	F
コーナーキック (4種類)	右 左	左奥, 左手前 右奥, 右手前	$C_{\ell f}^{(R)}, C_{\ell n}^{(R)}$ $C_{r f}^{(L)}, C_{r n}^{(L)}$
スローイン (4種類)	右 左	奥, 手前 奥, 手前	$T_f^{(R)}, T_n^{(R)}$ $T_f^{(L)}, T_n^{(L)}$
全体移動 (2種類)	右 左	左方向 右方向	$M_{\ell}^{(R)}$ $M_r^{(L)}$
攻撃 (2種類)	右 左	左方向 右方向	$A_{\ell}^{(R)}$ $A_r^{(L)}$

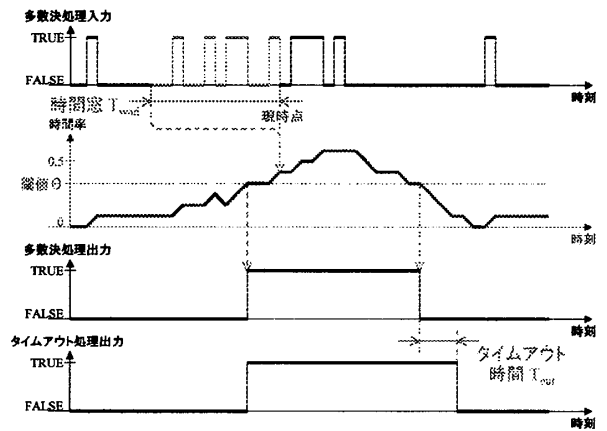


図3: ポストフィルタ

た、 \bullet には場所を表す添え字 (ℓf 等) が、 \circ にはチームを表す添え字 ((L) 等) が入る。

一例として、式(6)で定義される「右側チーム選手が左奥コーナーでコーナーキックを行う」イベント $C_{\ell f}^{(R)}$ の判定ルールは、以下のように言い換えられる。

- 【試合進行中】十分な人数の選手がおり ($n \geq N_C$)、
- 【選手の壁】ゴールポスト付近 ($|g_x| \geq G_C$) に選手が密集 ($\rho \geq P_C$) していて、ほとんど動かず ($\beta \leq B_C$)、
- 【キッカーの存在】左奥コーナーに右側チームの選手がいる ($c_{\ell f}$)。

2.4 ポストフィルタ処理

イベント判定部で得られる判定結果には、散発的な見誤り、見落としが含まれる。これらの誤りは時間的な頻度が小さく、継続時間も短いことから、イベント判定結果それぞれに対し、時間多数決とタイムアウト処理からなるフィルタ処理を適用することとした。フィルタ動作例を図3に示す。

まず、時間多数決処理では、イベント判定部から入力される時系列に対し、幅 T_{wnd} の窓を設定する。窓内において判定が TRUE となった割合を求め、その割合が閾値 θ 以上であった場合に TRUE を出力する。この処理により、イベント判定部出力のばたつきを抑えることができる。

続くタイムアウト処理では、時間多数決の結果が立

```

<?xml version="1.0" encoding="UTF-8" ?>
<SegmentMetadata>
  <AudioVisualSegment>
    <id>Posi_002</id>
    <event>攻撃</event>
    <state>右</state>
    <team>浦和</team>
    <SPoint>924</SPoint>
    <Duration>291</Duration>
  </AudioVisualSegment>
</SegmentMetadata>

```

図4: メタデータ記述例

ち下がった後もなお時間 T_{out} の間は TRUE の状態を維持するよう動作する。たとえば、式 (6) のコーナーキック判定では、主にキック寸前までの動きの小さい区間を検出するが、本タイムアウト処理を導入することでキック後までをイベント区間に含めることが可能となる。

2.5 メタデータ記述

最後にメタデータ記述部では、ポストフィルタ処理出力に基づいてイベントに関する情報を、XML で記述し出力する。図4に示すように、イベントの連番<id>、イベントの種類<event>、場所・方向<state>、動作主の所属チーム名<team>、生起時刻<SPoint>、ならびに継続時間<Duration>が記述される。なお、ポストフィルタ処理出力の各イベントには優先順位を設け、たとえばフリーキックとコーナーキックが同時生起するということがないよう、矛盾の解決も行う。

3. 実験

平成16年12月11日開催のJリーグチャンピオンシップ戦において、2台のハイビジョンカメラで収録した映像を用い、イベント検出の実験を行った。2台のカメラは観客席最上段部(地上高約40m)に互いに隣接して設置した。処理対象は、後半開始の30秒前から始まる約40分である。

図5に各種イベントが検出された際のシーン(左半面用センサカメラ画像)を示す。各図左下には、低レベル特徴抽出、高レベル特徴抽出、ポストフィルタの動作状況の一部を図示している。この動作状況は、図6にも示すように左側の棒グラフ群が高レベル特徴量のパターンを表している。当該パターンがIF~THENルールに合致すると、右側の棒グラフ(時間多数決フィルタ内での時間率)が右に伸びる。この時間率が閾値を超えたとき、最終的にイベント生起と判定される。

図5(a)に示すキックオフの場合、各チーム選手がそれぞれ自陣内に分布していることが特徴的である。フリーキック(図5(b))とコーナーキック(図5(c))では、いずれも選手が密集するという特徴があるが、コーナー(この場合は左手前)に攻撃側選手のキッカーがいるか否かが決め手となっている。図5(d)の攻撃シーンは、ゴール近くでゴールに向かって集団が移動しているときに検出されるが、シュートを試みる場合など

表3: イベント検出精度

イベント	適合率	再現率
キックオフ	100%	100%
フリーキック	100%	40%
コーナーキック	80%	100%
スローイン	47%	58%
全体移動	(100%)	(100%)
攻撃	(100%)	(100%)

重要なシーンが含まれる傾向がある。

表3に各イベントの検出精度を示す。ここで、適合率及び再現率の計算にあたっては、本システムにより検出されたイベント区間と、実試合でのイベント区間に重なりがあれば正解とした。一イベントに対し、同判定が反復された場合(途中「イベントなし」の状態が短時間挟まった場合)については、誤りと見なしていない。全体移動と攻撃の各イベントは、競技ルールとは無関係な物理的な指標であるため、その結果自体を正解と定義した。

キックオフおよびコーナーキックに関しては高い検出精度が達成できた。これらのイベントの検出精度が良好なのは、フォーメーションが非常に特徴的である上、その継続時間も長いからである。

一方、フリーキックの再現率が低くなったのは、特に重要でないフリーキックの場合には必ずしも定型的なフォーメーションをとらずに試合が進んでしまうことがあり、見逃しが生じたためである。

また、スローインの適合率が低いのは、選手の退場時や監督がタッチライン付近に現れたときなどを見誤ることに起因する。スローインの再現率低下の原因は、スローアーが間髪を置かずボールを投げってしまう場合を見逃したためである。時間多数決の時間窓サイズを小さくすることで再現率を改善できるが、逆に適合率の低下を招く。適合率と再現率のいずれを重視すべきか、目的に応じて設定する必要がある。

4. まとめ

サッカーの画像フォーメーション解析に基づく、イベントの実時間検出法を提案し、その実験結果を示した。今回、実験が1例に限られたため、判定閾値等のパラメータ最適化が十分ではなく、また定量評価も十分とは言えないが、サッカー映像に含まれる各種イベントを映像のみから検出できることが示された。

イベント判定のIF~THENルールは主観に大きく依存することから、今後、統計的な手法や状態遷移を導入し、より体系的な学習・判定が行えるシステムの構築を目指す。また、動作主の個人名や、ボールの位置もメタデータとして重要である上、イベント検出精度向上にも有効であると考えられる。そこで背番号認識手法[15]やボール検出手法[9]との組み合わせを検討している。さらに、さまざまな映像特徴量[16]や多視点映像[17]の利用により、低レベル特徴の高精度化・頑健化を図る。

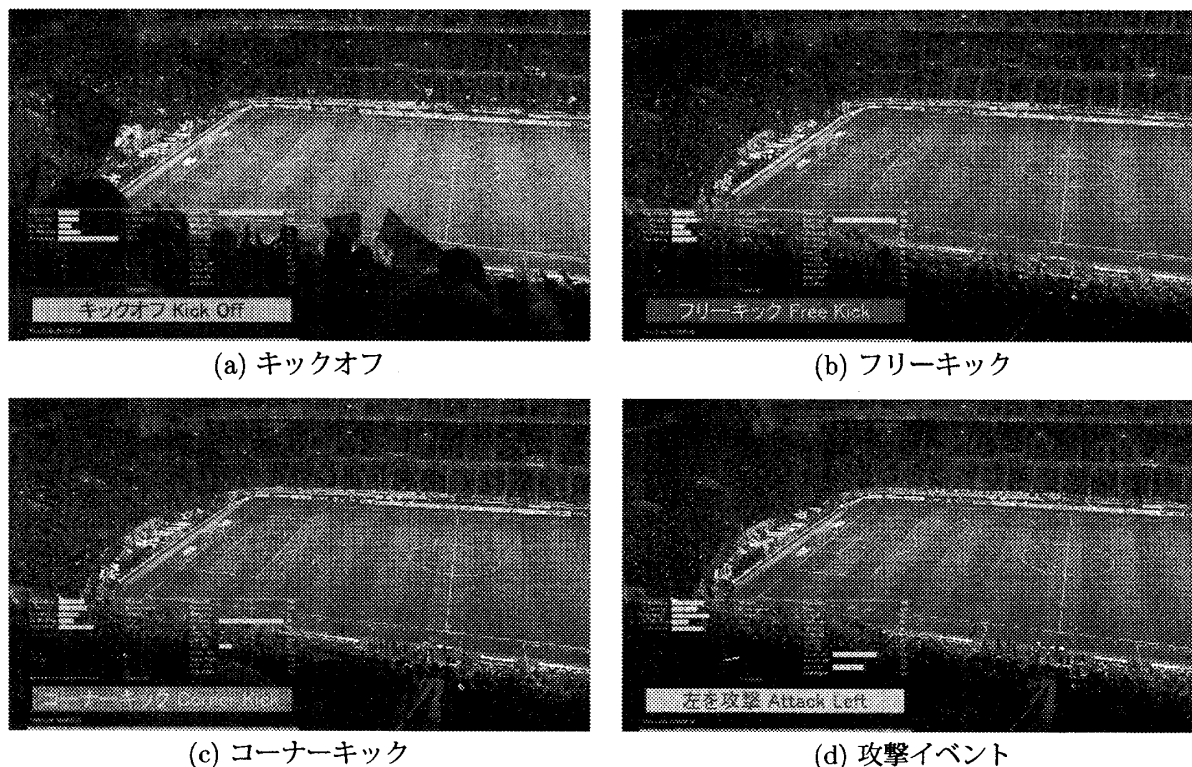


図 5: 各種イベントが検出された画像



図 6: 動作状況のモニタ

参考文献

[1] 映像情報メディア学会編, デジタル放送ハンドブック 第10編サーバー型放送, オーム社, 2003.
 [2] 馬場, 西本, 石川, 中村, 吉村, 栗岡: サーバー型放送におけるメタデータ利用技術の一検討, 信学技報 MoMuC2004-57, pp.11-16, 2003.
 [3] 金次, 三須, 高橋, 合志, 苗村: オブジェクト連動データ放送システムの開発とその記述方式, 情報学研報 2003-AVM-41, 2003.
 [4] 山田, 佐野, 住吉, 柴田: アナウンスコメントを利用したサッカー番組のメタデータ自動生成の検討, FIT2004, E-030, 2004.
 [5] 瀧, 長谷川: チームスポーツにおける集団行動解析のための特徴量とその応用, 信学論 D-II, vol. J81-D-II, no. 8, pp. 1802-1811.
 [6] M. Petkovic, Z. Zivkovic, W. Jonker: Recognizing Strokes in Tennis Videos Using Hidden Markov Models, IASTED International Conference Visualization, Imaging and Image Processing, 2001.
 [7] L. Xie, P. Xu, S.-F. Chang, A. Divakaran, H. Sun, Knowledge and Hidden Markov Models, Pattern Recognition Letters, vol. 25, pp. 767-775, 2004.

[8] 新田, 馬場口, 北橋: 言語と画像の情報統合によるスポーツ映像へのアノテーション付け, MIRU2000, pp. I-319-I-324, 2000.
 [9] Y. Ohno, J. Miura, Y. Shirai: Tracking Players and a Ball in Soccer Games, Proc. Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems, pp. 145-148, 2000.
 [10] Y. Gong, T.S. Lim, H.C. Chua: Automatic Parsing of TV Soccer Programs, ICMCS'95, pp. 167-174, 1995.
 [11] 今, 長谷山, 北島: サッカー映像からのチーム優勢度の判定に関する一考察, 信学技報 ITS2004-46, pp. 35-40.
 [12] A. Ekin, A.M. Tekalp, and R. Mehrotra: Automatic Soccer Video Analysis and Summarization, IEEE Trans. on Image Processing, vol. 12, no. 7, pp. 796-807, 2003.
 [13] P. Xu, L. Xie, S.-F. Chang, A. Divakaran, A. Vetro, H. Sun: Algorithms and Systems for Segmentation and Structure Analysis in Soccer Video, ICME'01, pp.928-931, 2001.
 [14] 三須, 小田原, 高橋, 合志, 宮越, 藤田: 画像オブジェクト追跡に基づくサッカーオフサイドラインの可視化, 情報技術レターズ (FIT2004), vol. 3, pp. 187-190, 2004.
 [15] 三須, 苗村, 高橋, 和泉: オブジェクト追跡と背番号認識の連携による動画像用スポーツ選手同定手法, 情報技術レターズ (FIT2003), vol. 2, pp. 187-189, 2003.
 [16] T. Misu, M. Naemura, W. Zheng, Y. Izumi, K. Fukui: Robust Tracking of Soccer Players Based on Data Fusion, ICPR 2002, vol. 1, pp. 556-561, 2002.
 [17] T. Misu, S. Gohshi, Y. Izumi, Y. Fujita, M. Naemura: Robust Tracking of Athletes Using Multiple Features of Multiple Views, Journal of WSCG, vol. 12, no. 2, pp. 285-292, 2004.