サッカー映像におけるホイッスル音声情報を利用したイベント検出

伊藤 拓紀[1,a)]　滝口 哲也[2,b)]　有木 康雄[2,c)]

1. はじめに

近年，個人撮影機器の高画質化やメモリの大容量化など，アマチュアのための撮影環境が整い，映像が誰でも簡単に撮影可能となった．これに伴って，映像編集のニーズが高まり，映像を個人の嗜好に基づいて編集することを支援する自動映像生成技術が注目を集めている．

この技術の一つにデジタルカメラワークがある．これは，スポーツの試合において，フィールド全体を高解像度カメラで撮影し，その映像を切り取り，連続的につなぎ合わせることによって新しい映像を作成するという技術である．この技術によって，視聴者の嗜好に即した自由な編集が可能となる．デジタルカメラワークを実現するには，そのスポーツ映像におけるイベントを検出し，その種類に基づいて映像を編集する必要がある．本研究では，世界的に人気のあるスポーツであるサッカーにおけるイベント検出手法を提案する．ここでのイベントとは，サッカーの試合における，ゴールキックやコーナーキック，ゴールといったアウトオブプレーを指す．

2. 関連研究

我々は，従来，ルールベースの投票式手法によるイベント検出システムを提案してきた [1]．この手法は，サッカーの試合におけるイベントが，ボールと選手の位置関係から一意に定まることを利用している．具体的には，選手やボールの位置情報と各イベントを対応付けたルール集合を事前に用意しておき，フレーム毎に選手やボールの追跡結果をそのルール集合と比較して各イベントの発生度を計算することで，イベントを検出する．しかしながら，この手法では，発生時間が短く，ルール集合で定義することが困難なフリーキックやスローインといったイベントの検出精度が低いという問題があった．また，主観に大きく依存し，ルール集合など多数のパラメータを手動で与えなければならなかった．

一方，矢崎ら [2] は，ベイズ隠れマルコフモデルを用いてスポーツイベントの検出を行った．この手法は，スポーツなどの時系列データの時間的依存関係を表現することができる一般化隠れマルコフモデルによって各イベントのモデリングを行い，イベントを検出している．特徴量としては，選手の位置情報を用いている．従来のルールベースの投票方式と比較し，ルール集合のパラメータを設定する必要がなく，他のスポーツ映像にも応用できる手法となっている．しかしながら，フリーキックやスローインといったイベントの検出精度においては改善されていない．そこで，本研究では，従来検出精度が低いとされていたフリーキックなどの検出精度の改善が期待できる手法を検討する．

サッカーでは，多くのイベントの発生時，審判によってホイッスルが鳴らされる．例えば，キックオフやファール，フリーキックの発生時である．本研究では，各時間でホイッスル音声の認識を行い，ホイッスル音声が認識された時には必ず何らかのイベントが発生したものと考える．そして，特徴量として選手の位置，速度情報だけでなく，ボールの情報も用いて，各イベントを隠れマルコフモデルによってモデリングを行い，イベント検出する手法を提案する．

3. イベント検出

サッカーの試合におけるイベントは，ボールと選手の位置，速度から一意に定まると考えられる．例えば，コーナーキックというイベントであれば，ボールはフィールドの四角のいずれかに速度 0 で存在し，ゴール前には両チームの選手が密集していることが考えられる．つまり，選手やボールを頑健に追跡することが，イベント検出をする上で重要となる．また，イベント発生時に鳴らされるホイッスル音声を認識することで，より確実に検出することが可能になると考えられる．

[1] 神戸大学大学院システム情報学研究科
〒 657–8501 兵庫県神戸市灘区六甲台町 1–1
[2] 神戸大学自然科学系先端融合研究環
〒 657–8501 兵庫県神戸市灘区六甲台町 1–1
[a)] itoh@me.cs.scitec.kobe-u.ac.jp
[b)] takigu@kobe-u.ac.jp
[c)] ariki@kobe-u.ac.jp

3.1 選手追跡システム

人追跡において高い性能が期待できるパーティクルフィルタを用いて，サッカー映像中の選手追跡を行う手法は数多く提案されている [3]．しかしながら，映像のフレーム間で複数選手の位置情報を用いていないため，オクルージョンなどにより一度追跡対象を見失うと，再度発見するのが困難であり，長時間の追跡が実現できないという問題があった．そこで，本研究では，複数選手の位置情報を時間状況グラフとして表現しておき，これにガイドされる形でパーティクルフィルタを実行することにより，オクルージョンが起こっても選手の誤検出，誤追跡を大幅に減らすことが可能な追跡システムを用いる [4]．ここで，時間状況グラフとは，図 1 のように，映像において毎フレーム，背景差分法により抽出した選手領域の集合を基に構築され，その選手領域内に存在する選手数（図 1 中のコンポーネント数）を遷移させていくグラフである．

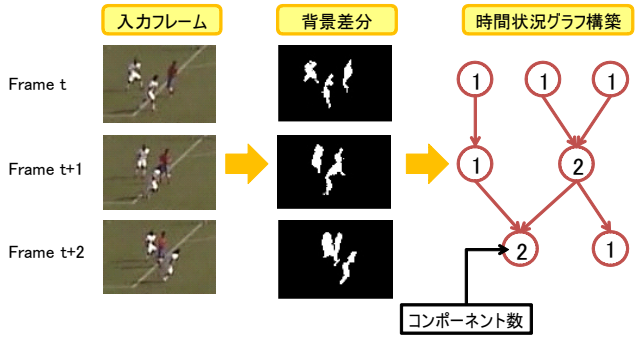

図 1 時間状況グラフ構築の流れ

3.2 イベント検出の流れ

イベント検出の流れを図 2 に示す．まず，動画データから選手及びボールの追跡を行う．ボール追跡に関しては，我々が提案している全面探索と局所探索を適応的に切り替える手法 [1] を用いる．そして，得られた選手情報を基に，フィールドを格子状に分割したグリッドベースの密度や速度平均などの情報に加工する．その後，ホイッスル音声情報をフリーキックなどのホイッスルが鳴らされるイベントのフラグ判定として用い，隠れマルコフモデルによってイベントを検出する．

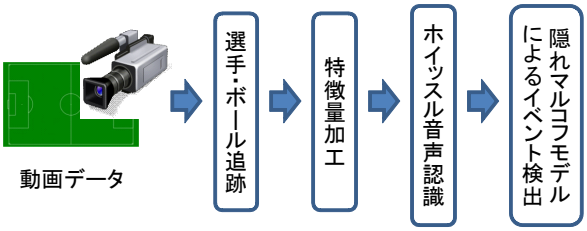

図 2 イベント検出の流れ

4. 予備実験

実際の視点固定の単眼サッカー映像に対して，本研究の特徴量の一つとして挙げられる選手位置を獲得する選手追跡実験を行った．時間状況グラフを用いないパーティクルフィルタによる手法と，時間状況グラフを組み合わせたパーティクルフィルタによる提案手法を比較する．そして，追跡精度評価は，式 (1) で定義される選手の平均追跡精度 A_p で行った．

$$A_p = \frac{1}{SP}\sum_{i=1}^{S}\sum_{j=1}^{P}\frac{W^{i,j}}{N^{i,j}} \tag{1}$$

$N^{i,j}$ はサンプル映像 i において，選手 j が映っているフレーム数，$W^{i,j}$ はサンプル映像 i において，選手 j の追跡成功フレーム数を表している．あるフレームにおいて，手動で与えた 2 次元画像座標正解データと，ワールド座標空間で追跡した結果を 2 次元画像座標に変換した座標データとの距離が，10 画素以内であれば追跡成功とした．ただし，S はサンプル映像数，P は全選手数である．サンプル平均の追跡精度は，表 1 より従来手法の 72.15 %から 79.50 %へ 7.35 ポイント向上した．

表 1 平均追跡精度

手法	A_p(%)
パーティクルフィルタのみ	72.15
時間状況グラフ+パーティクルフィルタ	79.50

5. まとめ

本研究では，サッカー映像において検出精度が低いとされるフリーキックやスローインといったイベントの検出精度を高めると考えられる手法を提案した．予備実験として，選手追跡の精度実験を行い，イベント検出の特徴量となる各フレームにおける選手位置が獲得できることを確認した．今後は，イベント検出実験を行って，性能を評価し，自動映像生成技術の一つである実況，解説音声の生成を行っていく．

参考文献

[1] 西濃 拓郎, 滝口 哲也, 有木 康雄, “単眼サッカー映像におけるボールの 3 次元位置情報を用いた状況認識” 画像の認識・理解シンポジウム (MIRU2009), IS2-61, pp. 1269–1276.

[2] 矢崎 智浩, 三須 俊枝, 中田 陽平, 他, “ベイズ隠れマルコフモデルを用いたスポーツイベント検出の高精度化” 電子情報通信学会技術研究報告, ヒューマン情報処理 109(471), pp. 401–406, 2010 年 3 月.

[3] M. Breitenstein, F. Reichin, B. Leibe, E. Koller-Meier and L. V. Gool, “Robust tracking-by-detection Using a Detector Confidence Particle Filter”, ICCV2009, pp. 1515-1522, 2009-09.

[4] H.Itoh, T.Takiguchi, Y.Ariki, “3D Tracking of Soccer Players Using Time-Situation Graph in Monocular Image Sequence” ICPR2012, WePSBT2.25, 2012-11.