

サッカー選手配置取得のための選手追跡と選手認識の確率的統合 Probabilistic Integration of Tracking and Recognition of Players for Acquisition of Soccer Formation

三須 俊枝[†] 松井 淳[†] サイモン クリッピングデル[†] 藤井 真人[†] 八木 伸行[†]
Toshie Misu Atsushi Matsui Simon Clippingdale Mahito Fujii Nobuyuki Yagi

1. まえがき

膨大な映像コンテンツから所望の番組やシーンを効率的に検索したり、番組ダイジェストの自動生成を実現するためには、番組を意味的な内容を記述したメタデータの付与が必須となる。スポーツ番組においては、対戦カードや出場者、試合結果のような静的な情報に加え、いつどこで何が起きたか(イベント)、時々刻々の戦況といった動的な情報を利用しやすいかたちで付与することが重要である。たとえば、色特徴やテクスチャ特徴、動き特徴などの低レベル特徴はシーンの見た目を記述するには有効であるが、試合の意味を記述する用途には不向きである。このため、各種プレイ等に特化した高レベル特徴を抽出し、メタデータとして付与する試みが盛んである。そこで我々は、サッカー等のチーム競技に関し、高レベル特徴抽出の素材として、あるいは高レベル特徴より汎用的かつ使いやすい中間レベルの特徴として、選手配置情報を利用することを提案している [1]。

選手配置情報を自動的に取得する研究として、選手追跡の頑健化が大きな課題となっており、各種低レベル特徴を融合する手法、多視点情報を利用するもの、時間方向の対応付けにおけるオクルージョンの考慮などさまざまな研究が行われている。しかし、どの移動軌跡が誰のものであるのかは手動で与える必要があった。また、別人物と軌跡が入れ替わってしまうような追跡誤りが致命的になるといった問題があった。そこで、我々は選手追跡と選手認識と組み合わせ、各軌跡が誰のものであるのかを確率的かつ動的に表現することで、名前付きフォーメーション情報を取得する手法を提案した [2]。

本稿では、上記 [2] において直感的な記述にとどまっていた定式化を Bayes の枠組みで展開し、追跡された軌跡群への選手 ID 付与の確率的かつ逐次的な過程と、選手間の位置関係に応じた確率分布の過程とにより表現する。本定式化により、選手認識、時間経過に伴う信頼度低下、選手接近による追跡誤り、およびフォーメーションに関する事前知識を統合的に取り扱うことが可能となった。また、このモデルを Jリーグのサッカー映像に適用し、予備的な実験を行った。本稿では、これらについて述べる。

2. システム構成

図 1 に本システムの構成を示す。選手追跡部は、固定広角カメラ画像から人物をすべて検出・追跡し、そのワールド座標群を出力する。追跡された複数の軌跡を区別するための識別子を軌跡 ID l とし、各軌跡 ID の時刻 t における選手位置を $x_t^{(l)}$ のように表記する。

一方、人物を区別する識別子を選手 ID と呼ぶ。選手追跡において生ずる誤りのため、選手軌跡と選手 ID とは必ずしも全時刻にわたって一対一に対応づけられるとは限らない。すなわち、選手軌跡は、時間方向に関し区分的には同一選手の移動履歴を表すが、長期的には一貫性がない。

そこで、本手法では、各軌跡 ID l が、どの選手 ID であるかを確率分布により表現し、これを時々刻々更新するものとする。

表 1 に変数の一覧を、表 2 に入出力を示す。本システムにおける最終出力は、表 2 の最下行に示す統合結果 $P(n_t^{(l)} | x_{1:t}^{(1:L)}, v_{1:t}, y_{1:t})$ である。すなわち、本システムは、時刻 t までの全軌跡 (1:L) に関する選手位置 $x_{1:t}^{(1:L)}$ 、能動カメラの注視位置 $v_{1:t}$ 、および能動カメラ画像 $y_{1:t}$ が与えられたとき、時刻 t における軌跡 ID l の選手名 $n_t^{(l)}$ を出力する。

軌跡 ID l の軌跡が、選手 ID $n_t^{(l)}$ である確率 $P(n_t^{(l)} | x_{1:t}^{(1:L)}, v_{1:t}, y_{1:t})$ の更新の概念図を図 2 に示す。追跡開始時点 A において、軌跡が誰のものであるのか不明であるとする。このとき、例えば、一様分布を確率分布の初期値として与える。時点 B において、画像認識の結果、その軌跡が鈴木選手のものであることが判明した場合、確率分布は鈴木選手の確率が高く、他の選手が低くなるように更新される。時間経過とともに選手追跡を誤る可能性が高くなる。すなわち、その軌跡が誰のものであるのか次第にあいまいになっていくことから、同図時点 B から時点 C の間の変化のように、

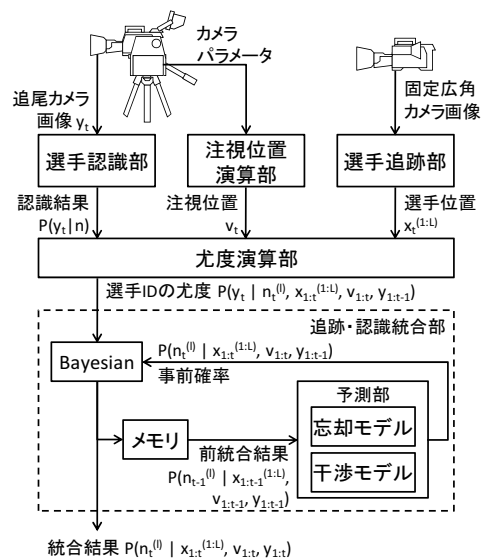


図 1: システム構成

[†]日本放送協会, NHK

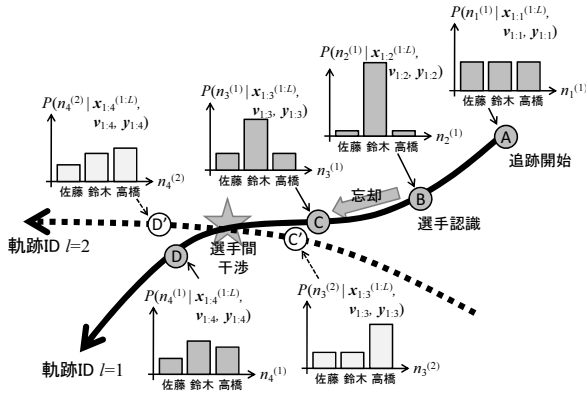


図 2: 確率更新の概念図

例えば次第に確率分布を一様分布に近づける．選手同士が近接した場合には，誤った選手を追跡してしまう可能性がある．そこで，時点 C および C' の確率分布を接近距離に応じた割合で混合して，時点 D および D' の確率分布を生成する．

3. 定式化

能動カメラ画像に基づく選手認識の結果に基づき，広角固定カメラから検出された選手位置に対して選手 ID を確率的に付与する手法を定式化する．

3.1 確率分布の更新式の導出

時刻 t までの全選手位置の履歴 $\mathbf{x}_{1:t}^{(1:L)}$ ，能動カメラの注視位置 $\mathbf{v}_{1:t}$ ，および能動カメラ画像 $\mathbf{y}_{1:t}$ が与えられたとき，各軌跡 ℓ の選手 ID が $n_t^{(\ell)}$ である確率 $P(n_t^{(\ell)} | \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t})$ は，Bayes の公式より次式のように変形できる．

$$\begin{aligned} & P(n_t^{(\ell)} | \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t}) \\ & \propto P(\mathbf{y}_t | n_t^{(\ell)}, \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t-1}) \\ & \quad \times P(n_t^{(\ell)} | \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t-1}) \\ & \stackrel{\text{def}}{=} \text{Likelihood} \times \text{Prior} \end{aligned} \quad (1)$$

表 1: 変数および記法

内容	文字	単位
時点	$t \in \mathbb{N}$	段階
軌跡 ID	$\ell \in \{1, 2, \dots, L\}$	番
選手 ID	$n_t^{(\ell)} \in \{1, 2, \dots, N\}$	番
選手位置	$\mathbf{x}_t^{(\ell)} \in \mathbb{R}^3$	[m]
注視位置	$\mathbf{v}_t \in \mathbb{R}^3$	[m]
能動カメラ画像	\mathbf{y}_t	
空間減衰定数	$R \in \mathbb{R}$	[m]
時間減衰定数	$T \in \mathbb{R}$	[s]
処理時間間隔	$\Delta T_t \in \mathbb{R}$ (2 時点 $t-1$ と t の間)	[s]
時刻範囲	$X_{1:t} \stackrel{\text{def}}{=} (X_1, \dots, X_t)$	
軌跡 ID 範囲	$X^{(1:L)} \stackrel{\text{def}}{=} (X^{(1)}, \dots, X^{(L)})$	

このとき，式 (1) の Likelihood は画像認識により得られる選手 ID $n_t^{(\ell)}$ の尤度である．画像認識の対象は能動カメラ画像により撮影された特定の人物である．

今回は，能動カメラの注視位置 \mathbf{v}_t の最も近傍にいる選手の軌跡 ID $\hat{\ell}$ に対しては認識結果の尤度を，それ以外の選手の軌跡に対しては一様分布を尤度 Likelihood として与えた．なお，注視位置は，水平面と並行な高さ 90cm (人物のおよその重心高) の平面と光軸とが交わる点とした．

$$\begin{aligned} \text{Likelihood} & \stackrel{\text{def}}{=} P(\mathbf{y}_t | n_t^{(\ell)}, \mathbf{x}_t^{(1:L)}, \mathbf{v}_t) \\ & \stackrel{\text{def}}{\propto} \begin{cases} P(\mathbf{y}_t | n_t) & (\text{if } \ell = \hat{\ell}) \\ 1 & (\text{otherwise}) \end{cases} \end{aligned} \quad (2)$$

$$\hat{\ell} = \underset{\ell \in \{1, \dots, L\}}{\text{argmin}} \|\mathbf{x}_t^{(\ell)} - \mathbf{v}_t\| \quad (3)$$

式 (1) の事前確率分布 Prior は，時刻 $t-1$ における軌跡 ID ℓ' の選手 ID $n_{t-1}^{(\ell')}$ を媒介して，次式のように確率伝播することができる．

$$\begin{aligned} \text{Prior} & = \sum_{\ell'=1}^L \sum_{n_{t-1}^{(\ell')}=1}^N \{P(n_t^{(\ell)} | n_{t-1}^{(\ell')}, \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t-1}) \\ & \quad \times P(n_{t-1}^{(\ell')} | \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t-1})\} \\ & \stackrel{\text{def}}{=} \sum_{\ell'=1}^L \sum_{n_{t-1}^{(\ell')}=1}^N \{\text{Predict} \times \text{Previous}\} \end{aligned} \quad (4)$$

式 (4) における Predict および Previous は，それぞれ時刻 $t-1$ から時刻 t への予測と，時刻 $t-1$ における統合結果に対応している．

選手追跡において生じ得る誤りには，追跡対象選手の変形や照明変動による画像特徴の変化に起因する追跡外れや，選手間の干渉時に正しくない選手を捕捉してしまうケースが代表的である．前者の追跡外れは，追跡開始時点からの時間経過とともにリスクが大きくなる．一方，選手間の干渉については，選手同士が近接すればするほど誤りやすくなる．予測 (Predict) を確率伝播により次式のように展開する．

$$\begin{aligned} \text{Predict} & \stackrel{\text{def}}{=} P(n_t^{(\ell)} | n_{t-1}^{(\ell')}, \mathbf{x}_t^{(1:L)}) \\ & = \sum_{n_{t-1}^{(\ell')}=1}^L \left\{ P(n_t^{(\ell)} | n_{t-1}^{(\ell')}, \mathbf{x}_t^{(1:L)}) \right. \\ & \quad \left. \times P(n_{t-1}^{(\ell')} | n_{t-1}^{(\ell')}, \mathbf{x}_t^{(1:L)}) \right\} \\ & \stackrel{\text{def}}{=} \sum_{n_{t-1}^{(\ell')}=1}^L \{\text{Forget} \times \text{Switch}\} \end{aligned} \quad (5)$$

表 2: 入出力

	信号	数式
入力	選手位置	$\mathbf{x}_t^{(1:L)}$
	注視位置	\mathbf{v}_t
	認識結果	$P(\mathbf{y}_t n)$
出力	統合結果	$P(n_t^{(\ell)} \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t})$

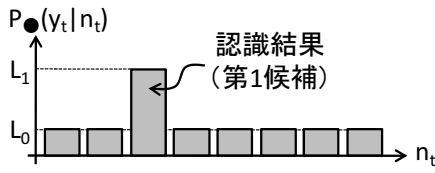


図 3: 尤度関数

このとき、式 (5) の Forget は、軌跡 ID ℓ の選手 ID に関する時刻 $t-1$ から時刻 t への遷移を表す形になっており、時々刻々あいまい性を増す「忘却」のモデル化に用いることができる。一方、式 (5) の Switch は、同一時刻 $t-1$ における軌跡 ID ℓ' から軌跡 ID ℓ への遷移を表しており、接近した軌跡間での追跡の「乗り移り」のモデル化に用いることができる。

今回は、選手 ID の確率分布を、経過時間に対し指数関数的に事前分布 $P(n_t | \mathbf{x}_t^{(1:L)})$ に漸近させることで忘却をモデル化する。

$$\text{Forget} \stackrel{\text{def}}{=} \left\{ 1 - \exp\left(-\frac{\Delta T_t}{T}\right) \right\} \delta_{n_t^{(\ell)}, n_{t-1}^{(\ell)}} + \exp\left(-\frac{\Delta T_t}{T}\right) P(n_t^{(\ell)} | \mathbf{x}_t^{(1:L)}) \quad (6)$$

事前分布 $P(n_t | \mathbf{x}_t^{(1:L)})$ には、全選手配置情報 $\mathbf{x}_t^{(1:L)}$ に応じて、自己の絶対的位置や他選手との相対的位置関係に基づき、その選手が誰なのかを予めモデリングする。

一方、乗り移りは、軌跡間の距離に応じて干渉相手の確率分布を混合させることでモデル化した。今回は、軌跡間の距離の絶対値に対し指数関数的に減衰する重みづけを用いた。

$$\text{Switch} \propto \delta_{n_{t-1}^{(\ell)}, n_{t-1}^{(\ell')}} \exp\left(-\frac{\|\mathbf{x}_{t-1}^{(\ell')} - \mathbf{x}_{t-1}^{(\ell)}\|}{R}\right) \quad (7)$$

式 (4) の Previous は、1 時点前までの認識結果 $\mathbf{y}_{1:t-1}$ から推定した 1 時点前の選手 ID $n_{t-1}^{(\ell)}$ の確率分布である。ここで、条件に含まれる現時点 t における選手配置 $\mathbf{x}_t^{(1:L)}$ および注視位置 \mathbf{v}_t に対する依存性を無視すると、

$$\text{Previous} \stackrel{\text{def}}{=} P(n_{t-1}^{(\ell)} | \mathbf{x}_{1:t-1}^{(1:L)}, \mathbf{v}_{1:t-1}, \mathbf{y}_{1:t-1}) \quad (8)$$

が得られる。以上の式 (1) ~ 式 (8) からなる漸化式により、選手追跡による位置情報に対し、選手認識の結果に基づいて選手 ID を確率的に付与することができる。

3.2 実装

選手認識の手法として、顔画像認識と、背番号画像認識の 2 種類を実装した。顔画像認識および背番号画像認識の結果を尤度 $P_{\text{Face}}(\mathbf{y}_t | n_t)$ および尤度 $P_{\text{Number}}(\mathbf{y}_t | n_t)$ によりそれぞれ表わし、これらの積として、両手法のデータフュージョンを行う。

$$P(\mathbf{y}_t | n_t) \stackrel{\text{def}}{=} P_{\text{Face}}(\mathbf{y}_t | n_t) \times P_{\text{Number}}(\mathbf{y}_t | n_t) \quad (9)$$

なお、認識結果が得られなかったタイミングにおいては、該当する尤度を一様分布とすることで、確率更新に影響を与えないようにする。

認識手法によっては、得られた認識結果の信頼度を定量的に得ることができる。このような場合、その信頼度に応じて尤度を増減することで、より精度の高い追跡・認識統合が可能になると考えられる。しかし、今回は簡単のため認識手法で得られる信頼度は無視して、図 3 に示すように認識結果の第 1 候補のみに対して高い値を有する尤度関数を用いることとした。

また、忘却のモデル (式 (6)) における事前分布 $P(n_t | \mathbf{x}_t^{(1:L)})$ も、簡単のため位置情報を無視した一様分布を用いることとした。

$$P(n_t^{(\ell)} | \mathbf{x}_t^{(1:L)}) \stackrel{\text{def}}{=} \frac{1}{N} \quad (10)$$

4. 構成要素

追跡・認識統合部以外の構成要素について簡単に述べる。

4.1 選手追跡部

実時間オフサイドライン可視化システム [3] の選手検出部を用いた。まず、固定広角カメラ映像から背景差分法により選手シルエットを抽出する。続いて、各シルエットの画像重心を観測量、サッカーコートに固定したワールド座標系における位置、速度を状態量とする拡張カルマンフィルタにより選手位置 (ワールド座標) を推定する。

また、各シルエット領域の上半分の領域 (シャツ領域) の平均色または色ヒストグラムと、予め登録したユニフォーム色情報 (色の平均および共分散、またはヒストグラム) との間の Mahalanobis 距離または Bhattacharyya 距離を評価することにより、該シルエットの選手の所属チームを識別する。

4.2 選手追尾撮影部

追跡の誤るおそれのある選手を重点的に撮影するため、選手追跡部で得られるフォーメーションに基づき他選手と干渉しそうな選手を 1 人選択し、該選手を追尾撮影するよう能動カメラのパン、チルトおよびズームを制御する [4]。

4.3 顔画像認識部

逐次予測型顔検出法 [5] で、追尾映像中の顔領域を高速特定する。可変テンプレートマッチングによる顔認識 [6] を実行し、選手 ID の確率分布を推定する。

4.4 背番号画像認識部

追尾映像中の選手の背番号領域を抽出し、輝度の 2 値化と形状フィルタにより番号候補を抽出する。これに OCR (認識エンジンには、エプソン販売株式会社製「読ん de!! ココ™ 活字文書 OCR ライブラリ」を使用した) を適用することで、背番号を認識する [7]。

5. 実験

J リーグサッカー試合において撮影された映像を使用し、実験を行った。競技場への機材持ち込みの制約から、今回は撮影・収録と、追跡・認識とを分けて実

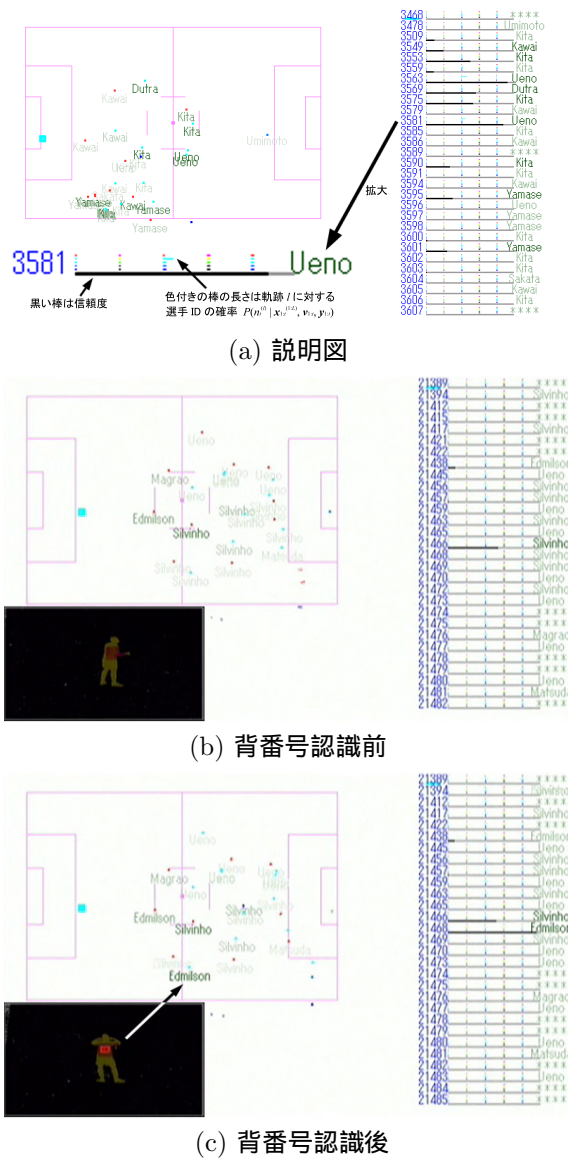


図 4: 実験結果

験を行った。まず、選手追跡部と選手追尾撮影部からなるサブシステムを競技場に設置し、能動カメラの自動撮影を行うとともに、カメラパラメータ（パン、チルト、およびズーム）を記録した。VTR に収録された固定広角カメラ映像と能動カメラ映像を実験室内において再生し、選手追尾撮影部を除くサブシステムにおいて実時間処理を行った。なお、顔画像認識部に関しては、オフラインで処理した結果を用いている。

図 4 に実験結果を示す。各図左上のサッカーコートの中の各点が追跡された選手の位置を表している。右上の 4 桁の数字が軌跡 ID ℓ である。図 4(a) 左下の拡大図の細かい $5 \times 5 = 25$ 本の棒グラフ群で示すように、各軌跡 ID に対し、確率分布 $P(n_t^{(\ell)} | \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t})$ が計算される。これらの中で、最も確からしかった選手 ID

$$\hat{n} = \operatorname{argmax}_{n \in \{1, \dots, N\}} P(n_t^{(\ell)} | \mathbf{x}_{1:t}^{(1:L)}, \mathbf{v}_{1:t}, \mathbf{y}_{1:t}) \quad (11)$$

が各図右端の選手名および各図左上の選手配置図内の選手名として表示されている。各 25 本の棒グラフ群の下に示す黒い棒グラフは、

$$(\text{信頼度}) = \frac{(\text{第 1 候補の確率}) - (\text{第 2 候補の確率})}{(\text{第 1 候補の確率})} \quad (12)$$

により定義する追跡・認識統合結果の信頼度である。なお、選手名の濃淡は信頼度を表している（濃いほど信頼度が高い）。図 4(b) および (c) は、背番号認識の前後における選手 ID の更新をようすを示している。各図左下の画像は、背番号抽出の結果である。(c) において追尾カメラ映像中央の選手の背番号が抽出・認識され、該選手に Edmilson の ID が付与されている。

6. まとめ

選手追跡と選手認識とを確率的に統合し、選手名付きのフォーメーションを取得する手法について定式化を行った。この際、選手 ID の確率分布更新において、選手認識による ID 付与の過程、並びに時間経過及び選手間干渉に起因する追跡誤差が明示的に表れるよう式展開を行いつつ、各要因のモデル化を行った。

本手法に基づくシステムを構築し、Jリーグサッカー映像を用いて動作の予備検証を行った。選手認識結果、時間経過、および選手間干渉によって確率分布が更新され、選手名が随時付与されていく動作を確認することができた。しかし、今回の実験では、テスト画像の不足からシステム全体の動作確認にとどまっており、付与された選手名の精度などの定量的評価は行っていない。今後、実験素材の充実を図り、精度および頑健性の評価を行う。また、番組制作や映像アーカイブス利用における映像検索への応用のため、得られた選手フォーメーションのメタデータ化と、選手名や戦況をクエリとする特定戦況シーンの特定法を確立したい。

参考文献

- [1] 三須, 苗村, 藤井, 八木: 選手フォーメーション解析に基づくサッカーイベント判別法, 映像学誌, vol. 61, no. 9, pp. 1367-1375, 2007.
- [2] 三須, 松井, クリピングデル, 藤井, 八木: サッカー選手配置推定のための顔・背番号認識と選手追跡の統合法, 2008 信学総大, D-12-9, p. 140, 2008.
- [3] 三須, 高橋, 合志, 蓼沼, 藤田, 八木: 実時間画像処理に基づくオフサイドライン可視化システム, 信学論, D-II, vol. J88-D-II, no. 8, pp. 1681-1692, 2005.
- [4] 三須, 高橋, 藤井, 八木: スポーツ番組におけるメタデータ自動生成 ~ サッカー選手追跡・同定のためのデータフュージョン ~, 信学技報, PRMU2005-121, pp. 39-44, 2005.
- [5] 松井, クリピングデル, 松本: ベイズ的動画像顔検出における顔候補領域の逐次予測, FIT2007, no. 3, H-025, pp. 59-60, 2007.
- [6] S. Clippingdale and T. Ito: A Unified Approach to Video Face Detection, Tracking and Recognition, Proc. International Conference on Image Processing ICIP 99, vol.1, pp. 662-666, 1999.
- [7] 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, 情報科学技術レターズ, vol. 2, LI-012, pp. 187-189, 2003.