

H-049

校正制御点近傍のSIFT特徴を用いるカメラパラメータ推定法 Camera Parameter Estimation Using Control Points with Adjacent SIFT Features

三須 俊枝[†] 藤井 真人[†] 柴田 正啓[†]
Toshie Misu Mahito Fujii Masahiro Shibata

1. まえがき

実写画像への映像合成や、実写画像からの被写体の3次元情報抽出においては、撮影時のカメラ位置、姿勢、画角などのカメラパラメータを知る必要がある。

従来、レンズや雲台に装着/内蔵したロータリエンコーダにより機械的にカメラパラメータを計測する手法が、スタジオや大規模スポーツ中継 [1] における映像合成に適用されている。映像合成に特化したバーチャルスタジオでは、視覚的なマーカを天井パトロンに設置 [2] したり、幾何学的な濃淡パターン印刷されたクロマキーバックや赤外線マーカを用いて、これらをカメラ(放送用または専用センサ)で撮影することで光学的にカメラパラメータを取得する装置も実用化されている。

ロータリエンコーダなど内界センサを用いる方式は、被写界に対しては非侵襲的であるためスポーツ中継などにも多用されているものの、撮影時に特殊な雲台等を用いる必要がある。また、事前にカメラの設置位置や姿勢の原点を詳細に校正する必要がある。

一方、光学的な計測方式では、画像の見た目から直接カメラパラメータを推定するため、事前校正が不要となる。また、たとえ誤差が生じても見た目上では目立たないというメリットがある。しかし、被写界に侵襲的であるというデメリットが大きい。

被写界にすでに存在する視覚的特徴を利用してカメラ校正を行う手法が考えられる。例えば、サッカーコートの白線と、モデルをカメラパラメータで投影した像との重なり具合から、カメラパラメータを推定する手法 [3] が提案されている。しかし、世界座標の知れた場所の視覚的特徴が顕著であるとは限らないため、これらを校正用の制御点として直接用いることには無理がある。そこで、本稿では、特徴点と制御点を別個のものとして捉える。特徴点と制御点との間の、画像上における位置関係と、両者の視覚的なパターンのスケール及び方位をデータベースに保持することで、特徴点の見た目から制御点の画像座標を推定できるように構成する。制御点群の、推定された画像座標及び、指定された世界座標に基づき、透視 N 点問題(本稿では、透視 4 点問題)を解くことでカメラ校正を行う。

2. 動作原理

本稿では、世界座標が既知の地点を制御点と呼ぶ。制御点の世界座標と画像座標のペアが十分な数(例えば、4 対以上)だけ得られれば、当該画像を撮影した際のカメラパラメータを算出することができる。

しかし、制御点は、必ずしも視覚的に際立った特徴を有するとは限らないことから、本手法では、制御点近傍に存在する視覚的に特徴的な点(特徴点)を抽出

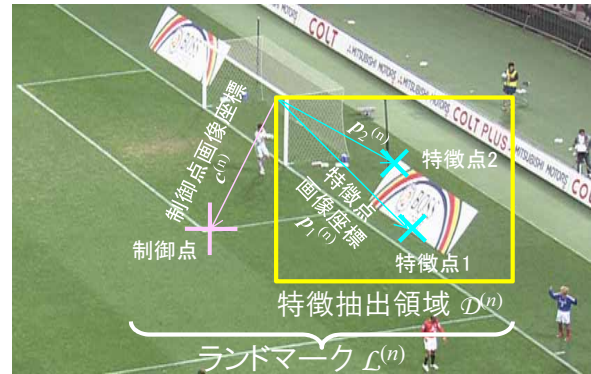


図 1: ランドマーク

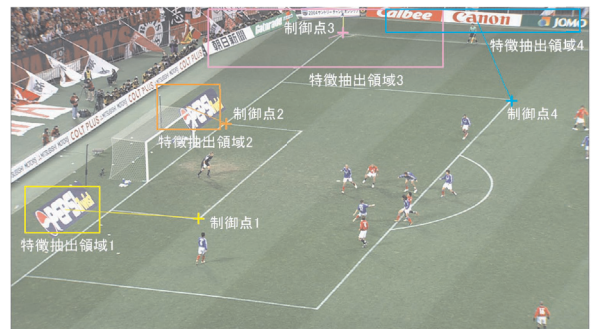


図 2: ランドマーク配置

し、これらの配置と見た目を手掛かりにして間接的に制御点の画像座標を推定する。

各制御点に対して 1 個以上の特徴点を関連付けることとし、これらを総称してランドマークと呼ぶこととする。図 1 にランドマークの概念図を示す。この例では、制御点はゴールエリアの 1 頂点に置かれている。この制御点に対応して、ユーザが矩形(特徴抽出領域)を指定し、その中から特徴点(この例ではゴール手前の看板上に二つの特徴点)を抽出する。こうしたランドマークを、図 2 に示すように複数配置する。

制御点位置を一にする複数のランドマークをおいてもよいし、逆に一つの特徴抽出領域に対して異なる制御点位置のランドマークをおいてもよいが、特徴抽出領域と制御点は近い方が好ましい。

本システムは、ユーザによるランドマーク登録機能と、自動処理によるカメラパラメータ推定機能を有する。図 3 に本システムのブロック図を示す。

ランドマーク登録は、カメラパラメータ推定に先だって事前に行ってもよいし、カメラパラメータ推定の中で追加・削除・修正を行ってもよい。

[†]日本放送協会, NHK

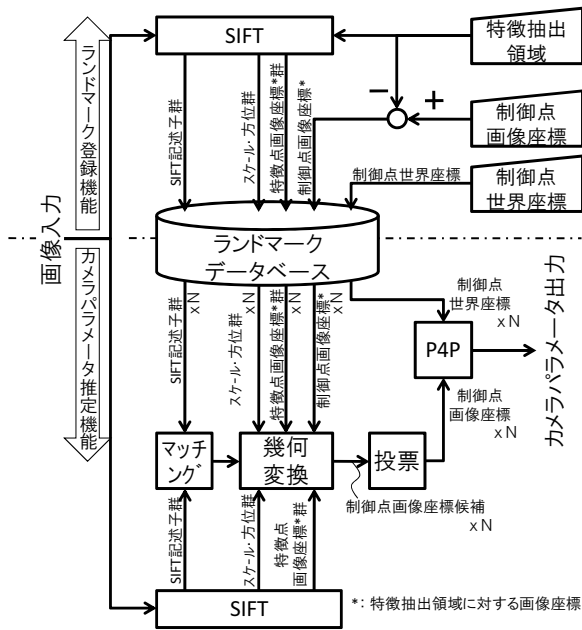


図 3: ブロック図

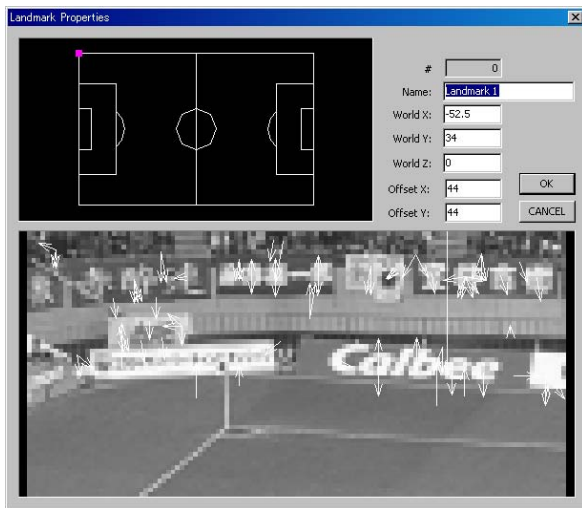


図 4: 抽出特徴点の表示と世界座標の指定

3. ランドマーク登録

以下では、1つのランドマーク (n 番目のランドマークを $\mathcal{L}^{(n)}$ とする) を登録する場合について説明する。以下の操作を繰り返すことで、 N 個のランドマークを登録する。 n 番目のランドマークに関する変数には上付きの (n) を付するものとする。

ランドマーク登録機能では、図 1 の十字線で示した制御点と、黄色の矩形で示した特徴抽出領域 $D^{(n)}$ をユーザがマウス操作等によって指定すると、本システムは特徴抽出領域から $M^{(n)}$ 個の特徴点を自動的に抽出する。また、ユーザに対し、図 4 に示すような制御点の世界座標を入力するためのダイアログを提示する。

表 1: 1 ランドマーク $\mathcal{L}^{(n)}$ あたりの情報
(a) 特徴点共通情報

制御点画像座標	$c^{(n)} = [c_x^{(n)}, c_y^{(n)}]^T$
制御点世界座標	$w^{(n)} = [w_x^{(n)}, w_y^{(n)}, w_z^{(n)}]^T$
特徴点総数	$M^{(n)}$

(b) 特徴点個別情報

	特徴点 1...	特徴点 m	...	特徴点 $M^{(n)}$
画像座標	$p_1^{(n)}$	$p_m^{(n)} = \begin{bmatrix} a_m^{(n)} \\ b_m^{(n)} \end{bmatrix}$...	$p_{M^{(n)}}^{(n)}$
方位	$r_1^{(n)}$	$r_m^{(n)}$...	$r_{M^{(n)}}^{(n)}$
スケール	$s_1^{(n)}$	$s_m^{(n)}$...	$s_{M^{(n)}}^{(n)}$
SIFT 記述子	$f_1^{(n)}$	$f_m^{(n)}$...	$f_{M^{(n)}}^{(n)}$

3.1 手動での指定

特徴点を抽出すべき領域 $D^{(n)}$ と制御点の画像座標 $c^{(n)} = [c_x^{(n)}, c_y^{(n)}]^T$ を画像上で指定し、また、制御点の世界座標 $w^{(n)} = [w_x^{(n)}, w_y^{(n)}, w_z^{(n)}]^T$ を入力する (図 4 の上半分)。

3.2 特徴点の抽出

画像上の特徴抽出領域 $D^{(n)}$ 内から、SIFT (Scale Invariant Feature Transform) [4] による複数のキーポイントを特徴点として抽出する。抽出された特徴点の総数を $M^{(n)}$ 個とする。

m 番目 ($m \in \{1, 2, \dots, M^{(n)}\}$) の特徴点の位置を $p_m^{(n)} = [a_m^{(n)}, b_m^{(n)}]^T$ 、方位を $r_m^{(n)}$ 、スケールを $s_m^{(n)}$ とおく。抽出された特徴点を図 4 の下半分に示す。この図では、特徴点の位置、方位およびスケールを矢印で表現している。

さらに、各特徴点に関して 128 次元の特徴量 (SIFT 記述子) $f_m^{(n)}$ を求める。

3.3 ランドマークデータベースへの登録

以上の手動指定および自動処理により得られたランドマークに関する情報 (表 1) を図 3 中央のランドマークデータベースに登録する。なお、本稿では、制御点の画像座標 $c^{(n)}$ および特徴点の画像座標 $p_m^{(n)}$ は、特徴点抽出領域 $D^{(n)}$ の位置に対する相対座標 (領域 $D^{(n)}$ の左上頂点を原点とする) で表すものとする。

表 2 にランドマークデータベースへのランドマーク情報の登録例を示す。なお、この表では各特徴点の画像座標、方位、スケールおよび SIFT 記述子の値の記載は省略した。また、第 1 列の「制御点画像座標」の各欄には、参考までに特徴抽出領域 $D^{(n)}$ 内の画像を示した。

4. カメラパラメータ推定

ランドマークデータベースに登録されたランドマークを入力画像中から検出し、その画像座標に基づいてカメラパラメータを算出する。

表 2: ランドマークデータベースへの登録例 ($N = 100$)

制御点 画像 座標	制御点 世界 座標	特徴 点数 $M^{(n)}$	ランドマーク名	世界座標 $w_x^{(n)}$
			画像座標 $c_x^{(n)}$	世界座標 $w_y^{(n)}$
			画像座標 $c_y^{(n)}$	世界座標 $w_z^{(n)}$
		# SIFT Descrs. 44	Landmark 1	$w_x = +0.000$
			$C_x = +69.0$	$w_y = +34.000$
			$C_y = +36.0$	$w_z = +0.000$
		# SIFT Descrs. 37	Landmark 2	$w_x = +0.000$
			$C_x = +76.0$	$w_y = +9.150$
			$C_y = +98.0$	$w_z = +0.000$
		# SIFT Descrs. 24	Landmark 3	$w_x = +0.000$
			$C_x = +70.0$	$w_y = +0.000$
			$C_y = +126.0$	$w_z = +0.000$
		# SIFT Descrs. 31	Landmark 4	$w_x = +0.000$
			$C_x = +68.0$	$w_y = -9.150$
			$C_y = +167.0$	$w_z = +0.000$
		# SIFT Descrs. 33	Landmark 100	$w_x = -36.000$
			$C_x = -8.0$	$w_y = -20.160$
			$C_y = +145.0$	$w_z = +0.000$

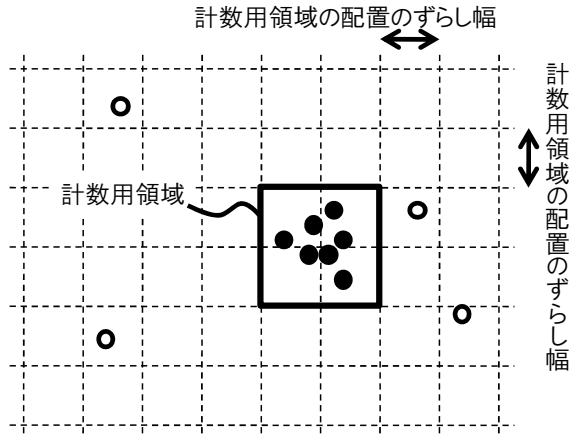


図 5: 投票による制御点画像座標の推定

4.1 入力画像からの特徴点抽出とマッチング

入力画像全体に対して SIFT を適用し, SIFT 特徴 (位置 $[\alpha_k, \beta_k]^T$, 方位 ρ_k , スケール σ_k , 記述子 ϕ_k) ($k \in \{1, 2, \dots, K\}$) を抽出する. ここに K は抽出された特徴点の個数である.

ランドマークデータベース内の各ランドマーク $\mathcal{L}^{(n)}$ に関し, その各 SIFT 記述子 $f_m^{(n)}$ に最も近い SIFT 記述子 ϕ_k を求め, そのインデックス k を $\kappa_m^{(n)}$ とおく.

$$\kappa_m^{(n)} = \operatorname{argmin}_{k \in \{1, 2, \dots, K\}} \|\phi_k - f_m^{(n)}\|_1 \quad (1)$$

ここに, $\|\cdot\|_1$ は L_1 ノルムを表す.

4.2 幾何変換による制御点画像座標の推定

ランドマークデータベース内のランドマーク $\mathcal{L}^{(n)}$ の制御点の画像座標 $c^{(n)}$ と, 第 m の特徴点に関する幾何情報 (画像座標 $p_m^{(n)}$, 方位 $r_m^{(n)}$, スケール $s_m^{(n)}$) と, 入力画像から抽出されたランドマークの幾何情報 (画

像座標 $\pi_k = [\alpha_k, \beta_k]^T$, 方位 ρ_k , スケール σ_k と, 前節のマッチング結果 $\kappa_m^{(n)}$ とに基づき, 位置, 方位, およびスケールの補償を行うことで, 制御点の画像座標 $\chi_m^{(n)}$ を推定する.

$$\chi_m^{(n)} = \pi_{\kappa_m^{(n)}} + \quad (2)$$

$$\frac{\sigma_{\kappa_m^{(n)}}}{s_m^{(n)}} \begin{bmatrix} \cos(\rho_{\kappa_m^{(n)}} - r_m^{(n)}) & -\sin(\rho_{\kappa_m^{(n)}} - r_m^{(n)}) \\ \sin(\rho_{\kappa_m^{(n)}} - r_m^{(n)}) & \cos(\rho_{\kappa_m^{(n)}} - r_m^{(n)}) \end{bmatrix} (c^{(n)} - p_m^{(n)})$$

4.3 投票による制御点画像座標の代表値の決定

続いて, すべての特徴点 m に関して, 推定画像座標 $\chi_m^{(n)}$ を画像平面上にプロットする. 理想的には, これら推定画像座標は 1 点に凝集するはずである. 実際には, 特徴量のマッチングや方位等の幾何補償は誤りや誤差を含むため, ある一点付近に広がりをもって凝集する (図 5 の黒丸) とともに, 若干の外れ値 (図 5 の白丸) を生じると考えられる. そこで, 本稿では制御点の推定画像座標の投票 [5] により, 代表値を決定する.

まず, 図 5 の点線のように画像を正方形格子状に区切る. 最小の正方形 2×2 個分の太い実線内 (以下, 計数用領域と呼ぶ) \mathcal{S}_i ($i \in \mathcal{I}$ は計数用領域を互いに区別するインデックス) に存在する推定画像座標値 (黒丸) の数 $\nu_i^{(n)}$ を数える.

$$\nu_i^{(n)} = \left| \left\{ \chi_m^{(n)} \mid \chi_m^{(n)} \in \mathcal{S}_i, m \in \{1, 2, \dots, M^{(n)}\} \right\} \right| \quad (3)$$

計数用領域を格子の間隔ですらしつつ (計数用領域が面積で $1/2$ ずつオーバーラップするよう移動), この黒丸の数 (得票数) を最大化する計数用領域 (そのインデックスを $\hat{i}^{(n)}$ とおく) を求める.

$$\hat{i}^{(n)} = \operatorname{argmax}_{i \in \mathcal{I}} \{\nu_i^{(n)}\} \quad (4)$$

最後に, 得票数が最大となった計数用領域 $\mathcal{S}_{\hat{i}^{(n)}}^{(n)}$ における推定画像座標 $\chi_m^{(n)}$ の平均値をランドマーク $\mathcal{L}^{(n)}$ の制御点画像座標の代表値 $\bar{\chi}^{(n)}$ とする.

$$\bar{\chi}^{(n)} = \frac{1}{\nu_{\hat{i}^{(n)}}^{(n)}} \sum_{\mu \in \{m \mid \chi_m^{(n)} \in \mathcal{S}_{\hat{i}^{(n)}}^{(n)}, m \in \{1, 2, \dots, M^{(n)}\}\}} \chi_\mu^{(n)} \quad (5)$$

なお, 今回は, 得票数の最大値 $\nu_{\hat{i}^{(n)}}^{(n)}$ が所定の数 θ に満たない場合には, 該ランドマーク $\mathcal{L}^{(n)}$ は無効とし, 以下のカメラパラメータ推定には用いないこととした. 以下, 有効なランドマーク $\mathcal{L}^{(n)}$ のインデックス n の集合を \mathcal{N} と書くこととする.

$$\mathcal{N} = \{n \mid \nu_{\hat{i}^{(n)}}^{(n)} \geq \theta\} \quad (6)$$

また, 有効なランドマークのインデックス n を昇順に並べた数列を $(n_j)_{j=1, 2, \dots, N'}$ とおく.

4.4 カメラパラメータの算出 (透視4点問題)

すべてのランドマーク $\mathcal{L}^{(1)}, \mathcal{L}^{(2)}, \dots, \mathcal{L}^{(N)}$ について、以上の処理を実行することにより、制御点の画像座標 (推定画像座標の代表値) と世界座標との対 $(\bar{x}^{(n)}, w^{(n)})$ ($n \in \mathcal{N}$) を $N' = |\mathcal{N}|$ 対得ることができる。

ここで、対 $(\bar{x}^{(n)}, w^{(n)})$ ($n \in \mathcal{N}$) において、その世界座標 $w^{(n)}$ がすべて同一平面上にあって、かつその3つ以上が同一直線上に乗らないようなものが4対以上存在すれば、透視4点問題 (P4P) の解法 [6] によりカメラパラメータ (ホモグラフィ行列 H) を求めることが可能である。

以下、制御点は平面 $w_z = 0$ 上に拘束されるものとし、P4Pの解法を適用するものとする。画像座標 \bar{x} の齊次座標を \tilde{x} 、世界座標 w の XY 平面内成分 $[w_x, w_y]^T$ に対する齊次座標を \tilde{w} のように、いずれもチルダ (\sim) で表すこととする。

$$\tilde{x} = \begin{bmatrix} \bar{x} \\ 1 \end{bmatrix}, \quad \tilde{w} = \begin{bmatrix} w_x \\ w_y \\ 1 \end{bmatrix} \quad (7)$$

このとき、画像座標と世界座標とは、実係数 λ を用いて次式により相互変換できる。

$$\lambda \tilde{x} = H \tilde{w} \quad (8)$$

ここで、ホモグラフィ行列 H は以下により算出可能である (上付きの $+$ は右側擬似逆行列: $Z^+ = [Z^T Z]^{-1} Z^T$)。

$$H = [W^+ C X]^T \quad (9)$$

$$X = [\tilde{x}^{(n_1)} \quad \tilde{x}^{(n_2)} \quad \dots \quad \tilde{x}^{(n_{N'})}]^T \quad (10)$$

$$W = [\tilde{w}^{(n_1)} \quad \tilde{w}^{(n_2)} \quad \dots \quad \tilde{w}^{(n_{N'})}]^T \quad (11)$$

$$C = \frac{v_{n_{N'-1}}}{u_{n_{N'-1}}} \text{diag} \left(\frac{u_{n_1}}{v_{n_1}}, \frac{u_{n_2}}{v_{n_2}}, \dots, \frac{u_{n_{N'-1}}}{v_{n_{N'-1}}} \right) \quad (12)$$

$$\begin{bmatrix} u_{n_1} & u_{n_2} & \dots & u_{n_{N'-1}} \end{bmatrix} = [\tilde{x}_{n_{N'}}]^T X^+ \quad (13)$$

$$\begin{bmatrix} v_{n_1} & v_{n_2} & \dots & v_{n_{N'-1}} \end{bmatrix} = [\tilde{w}_{n_{N'}}]^T W^+ \quad (14)$$

とする。

4.5 RANSAC による頑健化

今回は、外れ値や誤差の大きなサンプルが含まれる場合に対応するため、前節の P4P 解法に基づく RANSAC [6] を構成した。

5. 実験

パン、チルト、ズーム操作を伴うサッカー映像を用いて実験を行った。この映像は、実際のサッカー中継番組に用いられたカメラのうちの1台 (ほぼハーフウェイラインの延長上に設置) の映像であるため、放送番組制作用のカメラワークである。

試合開始後数分間の映像上においてランドマークを100個 (一部を表2に示す) を設定し、引き続きカメ

表 3: 実験に用いたパラメータ等

入力画像	
解像度	480 × 270 画素
	8ビットグレースケール
	(1920 × 1080 4:2:2 YP _b P _r から変換)
処理時間間隔	
	可変 (コマ落とし)
ランドマーク	
ランドマーク数 N	100 個
制御点決定に必要な得票数 θ	5 対
特徴点数 $M^{(n)}$ ($n = 1, 2, \dots, 100$)	16 ~ 142 個
SIFT	
コントラストに対する閾値	0.005
曲率に対する閾値	10.0
ガウシアン画像解像度下限	8 × 8 画素
RANSAC	
最大繰り返し回数	100 回
インライア閾値	30 画素
P4P に要するランドマーク数	4 個
最低限必要なインライア数	$\max\{0.7N', 5\}$ 対

ラパラメータの推定を行った。表3に実験に用いたパラメータを示す。

実験結果の一部を図6~図9に示す。検出・推定された制御点をX印で示した。明るいX印がRANSACでインライアと判定されたもの、暗いX印がアウトライアと判定されたものである。また、サッカーコートモデルをカメラパラメータ (ホモグラフィ H) により投影した結果を黄線で示した。

図6のペナルティエリア奥のLandmark 21とLandmark 97のように複数のX印が重なって表れているのは (異なるカメラアングル等にて) 同じ制御点を重複登録しているからである。

図6では、Landmark 85 (正しくは右奥コーナー) が観客席付近に誤って検出されているが、RANSACにおいて本ランドマークがアウトライアとして判定され、サッカーコートモデルはほぼ正しく実写のライン付近に投影されている。

図7は、右ペナルティエリア付近の撮影画像に対する結果であるが、ほぼ正確なカメラパラメータ推定が行われている。このケースでは、右ゴールとその周囲の看板に特徴抽出領域 $\mathcal{D}^{(n)}$ を設定したランドマークが用いられている。

図8では、制御点群が奥側に偏在したため若干末広がりの歪みを生じている。

図6~図8の場合、人物の大きさとの対比でわかるように、誤差は1m程度である。この精度が、例えば戦術解析やCG合成に十分なものであるか否かは、その目的に応じて今後詳細に検討する必要がある。

図9では、制御点群が左奥に極端に偏在したため、投影像に大きな歪みを生じている。このような投影状態は通常の撮影条件では生じないため、事後処理で失敗を検出することは可能であると考えられる。

今回の実験では、平均約16.4秒に1回の頻度でカメラパラメータが得られた (推定された結果には誤りを

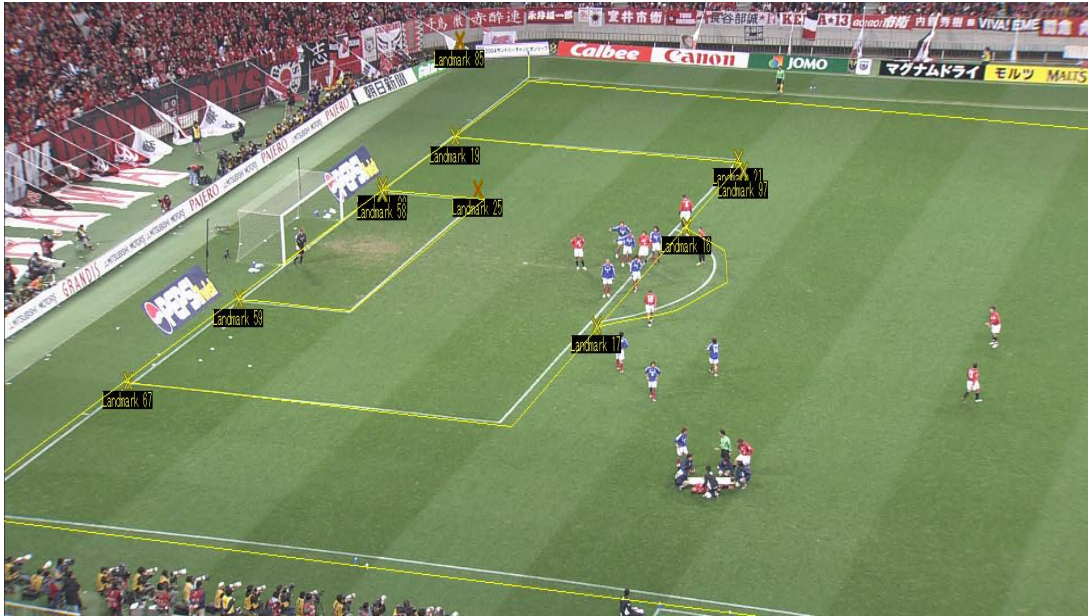


図 6: 実験結果 1 (左半面)

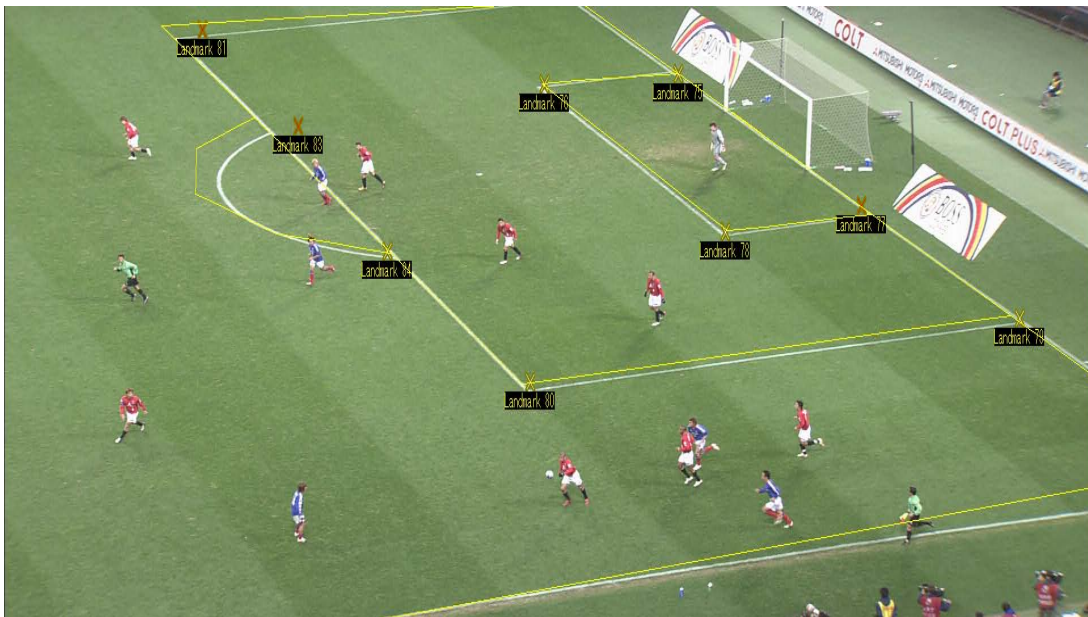


図 7: 実験結果 2 (右半面)

含む)。これは、処理負荷の大きさに加え、十分な個数のランドマークが写っていない期間は推定結果が得られないためである。なお、最小の時間間隔は約 1.4 秒であった。なお、実験に用いた計算機は Xeon-3.2GHz のデュアルプロセッサ搭載機である。

6. まとめ

制御点付近に視覚的に顕著な特徴がない場合にもカメラ校正を可能とすべく、制御点と特徴点を独立に配置可能な手法を提案した。本手法では、パンやズームに伴う画像の回転・拡大を補償しつつ制御点の画像座標

を得るため、特徴点の抽出と記述に SIFT を採用した。

実験の結果、十分な特徴点对応が得られれば精度よくカメラパラメータを推定できることが確認できた。しかし、時間的に動作は間欠的であり、推定結果もぶれる傾向にあった。

今後、時間方向のフィルタリングにより推定結果を滑らかにするとともに、他のセンサ(ジャイロなど)や他の推定手法(動きベクトルなど)を統合して、時間分解能の高い高精度かつ頑健なカメラパラメータ推定を実現したい。

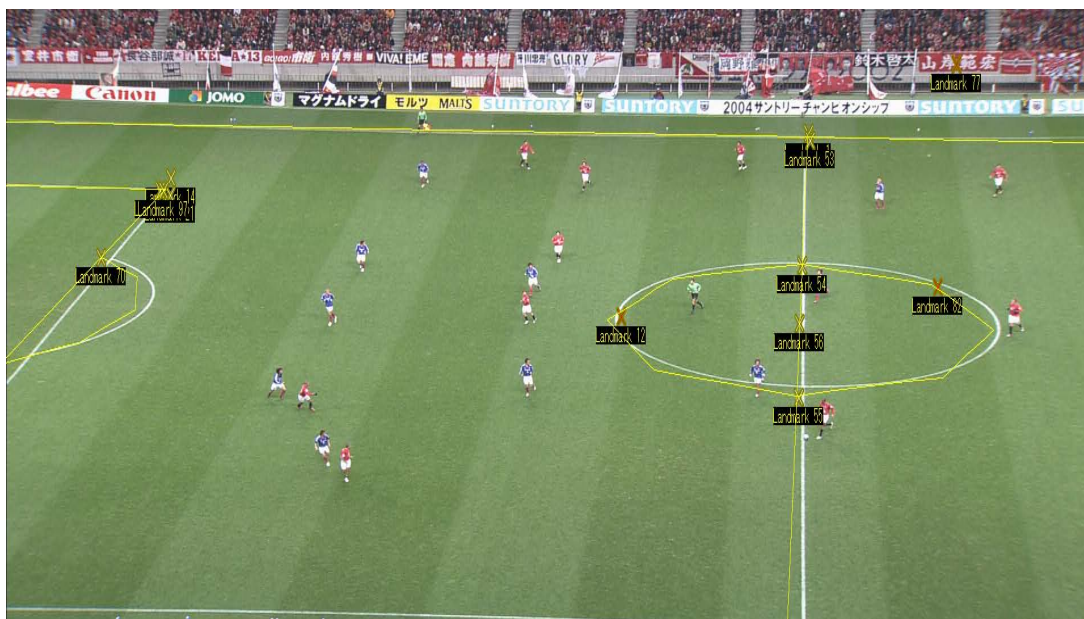


図 8: 実験結果 3 (中央)

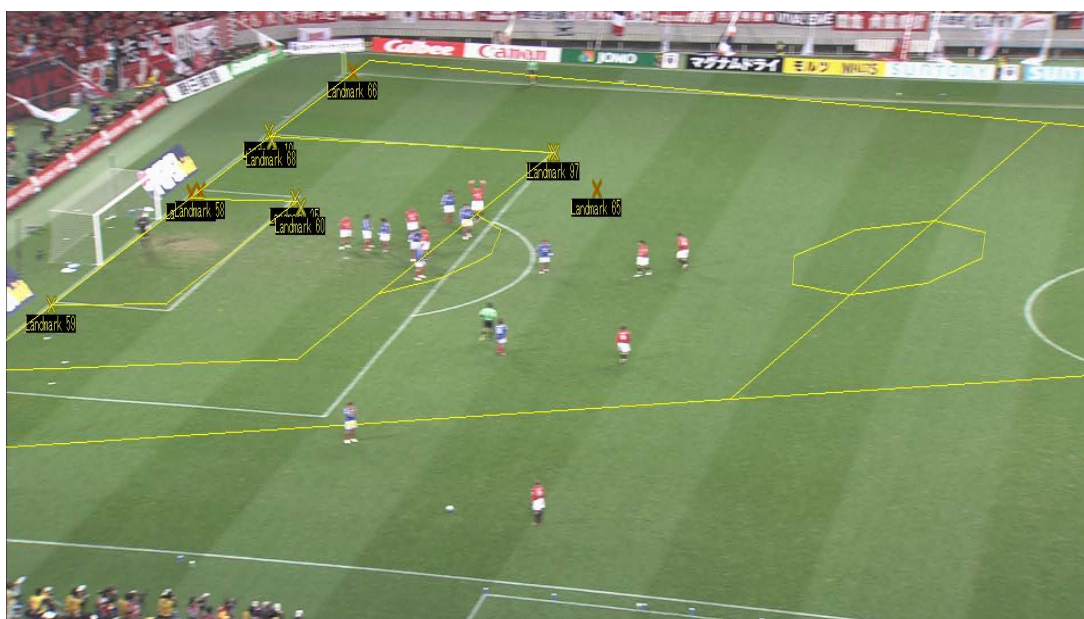


図 9: 実験結果 4 (失敗)

参考文献

- [1] 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.
- [2] Vinten Radamec: Free-D 3D カメラトラッキングシステム, <http://www.vintenradamec.com/free-d-ja.php>.
- [3] 渡辺, 長谷山, 北島: サッカー映像における撮影フィールド領域の検出に関する一考察, FIT2003, J-092, pp. 387-388, 2003.
- [4] D. G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints, Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- [5] 高木, 藤吉: SIFT 特徴量を用いた交通道路標識認識, 第 13 回画像センシングシンポジウム SSII07, LD2-06, 2007.
- [6] M. A. Fischler and R. C. Bolles: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Comm. of the ACM 24, pp. 381-395, 1981.