Vol.2017-CG-168 No.30
Vol.2017-DCC-17 No.30
Vol.2017-CVIM-209 No.30
2017/11/9

人体関節を用いた 多視点カメラの外部キャリブレーション法に関する検討

高橋 康輔[1,a)]　三上 弾[1,b)]　五十川 麻理子[1,c)]　木全 英明[1,d)]

概要：本研究では人体関節を対応点として利用する多視点カメラの外部キャリブレーション法を提案する．従来外部キャリブレーションではチェスボードなどの構造が既知な参照物体を利用する方法が主に用いられてきた．また，カメラの共通視野に参照物体を置けない場合であっても，共有視野に存在する物体の投影像に対して得られる自然特徴点を利用する手法が提案されている．これらの手法はカメラの視差が大きい場合には物体の見えが大きく変わり，正しく対応点が取れなくなり適用が困難であるという課題があった．この課題に対し，本研究では映像中の人体関節は様々な視点からでも安定して検出できるという特性に着目し，人体関節を対応点として用いる外部キャリブレーション法を提案する．提案手法では，人体関節の検出点はチェスコーナーに比べて検出誤差を多く含むことを考慮し，検出誤差を許容するように再投影誤差を再定義する．さらに，対応点が人体の関節から構成されるという前提に基づいた制約を導入することで，検出誤差にロバストな外部キャリブレーション法を実現する．実験ではシミュレーションデータおよび実データを用いてその有効性を確認した．

Extrinsic Camera Calibration from Human Joints

KOSUKE TAKAHASHI[1,a)]　DAN MIKAMI[1,b)]　MARIKO ISOGAWA[1,c)]　HIDEAKI KIMATA[1,d)]

1. はじめに

多視点カメラシステムは被写体の三次元形状の復元をはじめ，コンピュータビジョンの様々な研究において広く利用されている．この多視点カメラシステムで撮影した映像に対して三次元画像処理を施すためには，各カメラ間の位置および姿勢を求める外部キャリブレーションを実施する必要がある．

一般に，多視点カメラの外部キャリブレーションを行うためには，チェスボードのような三次元構造が既知な参照物体を利用する方法が用いられる [1,2]．これらの手法は多視点カメラ間での対応点が頑健に得られ，安定して外部キャリブレーションが実施できる一方，スポーツフィールドや舞台のように参照物体を事前に持ち込むことが困難な場合に適用が困難である．このように構造が既知な参照物体を利用できない場合，カメラの共有視野内に存在する物体の投影像に対して自然特徴点を求め，それらを対応点として利用することで外部パラメータを求める手法が提案されている [3]．これらの手法は外部キャリブレーションに利用する対応点として共有視野に存在する直接観測可能な三次元点を参照点として利用しているため，その三次元点が見えない，あるいはカメラの視差が大きくなるに従い見えが大きく変わる場合には，多視点カメラ間で正しく対応が取れなくなり適用が困難であるという課題がある．

これらの課題に対し，本研究では直接観測できる三次元点を用いるのではなく，人体の関節モデルにおける各関節点のように物体の内部に存在する三次元モデルの各点を各カメラにとって共通の参照点として利用することで外部キャリブレーションを実施する．人体の関節点は直接観測することは出来ないが，図 1 に示すように人体の投影像に対して [4,5] などの関節位置推定手法を適用することで各

[1] 日本電信電話株式会社　NTT メディアインテリジェンス研究所
NIPPON TELEGRAPH AND TELEPHONE CORPORATION, NTT Media Intelligence Laboratories
[a)] takahashi.kosuke@lab.ntt.co.jp
[b)] mikami.dan@lab.ntt.co.jp
[c)] isogawa.mariko@lab.ntt.co.jp
[d)] kimata.hideaki@lab.ntt.co.jp

Vol.2017-CG-168 No.30
Vol.2017-DCC-17 No.30
Vol.2017-CVIM-209 No.30
2017/11/9

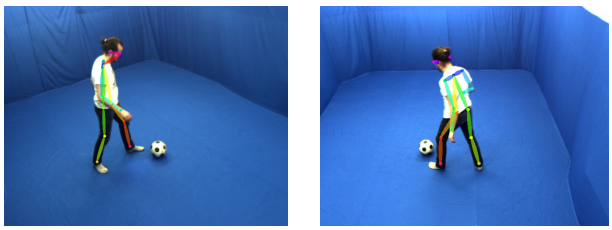

図 1 各関節の検出例．検出した関節を重畳している．

関節の投影点の座標値を得ることができる．これらの関節の投影点は様々な視点から撮影された映像においても安定して共通の対応点として検出できるため，視差の大きな多視点カメラシステムの外部キャリブレーションに利用できると考えられる．

一方，検出された関節点の検出精度は関節位置推定手法の精度に大きく依存し，一般にチェスコーナーの検出点などと比較して数ピクセル以上の大きな誤差を含むため，従来のように再投影誤差を利用したバンドルアジャストメントを実施すると精度が不安定になる．これに対し，提案手法では以下の２点を特徴とする最適化関数を定義する．1点目はこの検出誤差を許容する形で再投影誤差を再定義する．この時，誤差を許容することによって解が一意に定まりにくくなることが懸念されるが，2点目の特徴として参照点として利用している点が人体の関節から構成されるという前提に基づいた制約項を導入することでそれを回避することを狙う．これらの制約項は一本の骨の両端に定めた関節間の距離は全てのフレームにおいて同一であるという長さに関する制約項と，各関節の三次元位置は時間方向に滑らかに変化するという動きに関する制約項から成る．提案手法ではこの最適化関数を非線形最適化することで外部パラメータを求める．

本論文の構成は以下のとおりである．まず２節で関連研究について述べ，３節では本研究で提案する外部キャリブレーション法について述べる．続く４節では実データ及びシミュレーションデータを用いて提案手法の性能を評価し，５節で結論を述べる．

2. 関連研究

本節では多視点カメラのカメラキャリブレーション手法および人体の関節位置推定手法に関して関連研究を述べる．

外部キャリブレーション：外部キャリブレーションに関する研究はコンピュータビジョンの分野において主要なトピックの一つであり，これまでにも様々な研究が提案されてきた．特に，監視カメラ群やスタジアムに設置されたカメラ群のように，共有視野にチェスボードのような構造が既知の参照物体を持ち込むことが困難な場合やカメラ間の視差が大きい場合では，従来のチェスボードを用いた手法や自然特徴点を利用した SfM に基づく手法では正しく対応点が取得できないため適用が困難であるという課題がある．

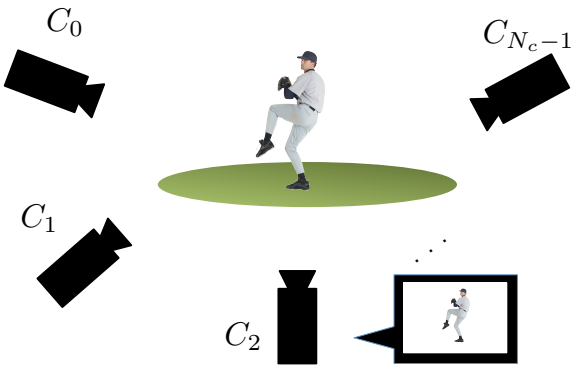

図 2 想定する多視点カメラシステム．N_c 台のカメラを用いて人物を撮影することを想定する．

このような場合，カメラが映すシーンの知識を用いることで外部キャリブレーションを行う手法が提案されている．Huang ら [6] は定点カメラで多人数の人物の移動が撮影できることを想定し，人物の移動軌跡を用いることで外部キャリブレーションを行った．また，Namdar ら [7] らはスポーツフィールドを撮影していることを想定し，フィールド上のラインから計算できる消失点を利用することで各カメラの自己位置を推定している．

これらに対し，図 2 に示すように多視点カメラで人物を撮影していることを想定してキャリブレーションを行う手法も多く提案されている [8–10]．[8, 9] では，人物のシルエットを利用することで外部キャリブレーションを行った．また，Puwein ら [10] は人体の関節位置を対応点として用いる外部キャリブレーションを提案した．[10] では再投影誤差に加え，関節間の長さや動き，オプティカルフローなどからなる最適化関数を最小化することで外部パラメータを推定している．

人体の関節位置推定手法：従来，人体の二次元関節位置推定手法に関する研究では，人物の各部位の空間的関係を木構造のグラフィカルモデルで表した pictoral structures を用いたアプローチ [11] や，スケールの異なる人体の部位間の関係を木構造で記述し，Corase-to-fine に各部位を検出する Hierarchical models に基づくアプローチ [12] が提案されてきた．

これに対し，近年ではニューラルネットワークの発展に伴い，[4,5,13] などの DNN を利用した手法が多く提案されている．[4] は物体認識で使われている AlexNet アーキテクチャを利用して回帰問題として姿勢を推定した．[13] は sequential prediction アプローチである Pose Machine [14] に対して CNN を導入することで高精度な姿勢推定を実現した．Cao ら [5] は [13] をさらに発展させ，Part affinitiy field を導入して各関節間の結合性も考慮することで多人数が映る映像においても頑健かつリアルタイムに検出可能な手法を実現した．

Vol.2017-CG-168 No.30
Vol.2017-DCC-17 No.30
Vol.2017-CVIM-209 No.30
2017/11/9

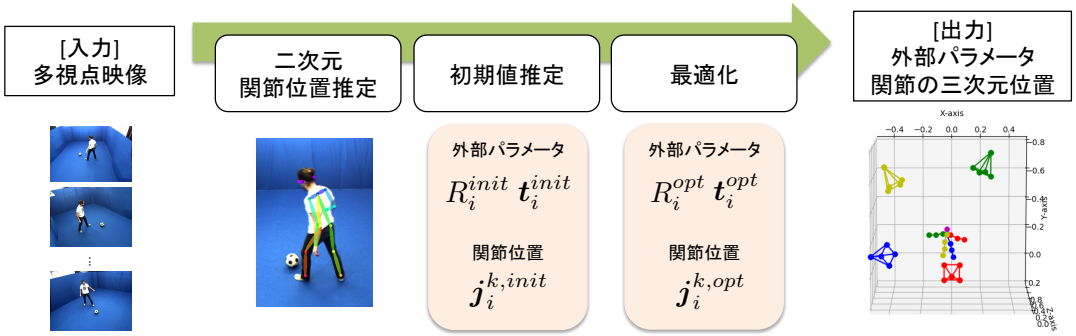

図 3 提案手法の流れ．

3. 提案手法

提案手法も [10] と同様に，人体関節の投影点は視差のあるカメラ間においても安定して検出可能な共通の点であることに注目し，関節の投影点を対応点として利用することで外部パラメータを推定する．提案手法では従来の再投影誤差を検出誤差を許容する形式に再定義することで，検出した関節点が大きな誤差を含む場合でも頑健に外部パラメータの推定が可能になることを目指す．

図 2 に示すように，本研究では $N_c(>2)$ 台のカメラから成る多視点カメラシステムを利用する．各カメラ $C_i(i=0,\cdots,N_c-1)$ の外部パラメータである回転行列および並進ベクトルを R_i および $\boldsymbol{t}_i$ としたとき，これらは

$$\boldsymbol{p}^{C_i} = R_i \boldsymbol{p}^W + \boldsymbol{t}_i \tag{1}$$

を満たす．なお，この時 $\boldsymbol{p}^{C_i}$ は三次元点 $\boldsymbol{p}$ のカメラ C_i 座標系における座標値であり，$\boldsymbol{p}^W$ は世界座標系における座標値を表す．この時，本研究の目的は図 2 に示すように撮影した人物の映像を用いて各カメラの外部パラメータ R_i および $\boldsymbol{t}_i$ を求めることである．本稿では C_0 の座標系を世界座標系とし，ある座標値に関して特に座標系の指定がない場合は世界座標系で表されているものとする．撮影された映像は全て N_t フレームから成る同期済みの映像であり，撮影に用いたカメラの内部パラメータは [1] の手法を用いて事前に既知で有ることとする．

また，本稿では人体を表すモデルとして図 4 に示す $N_j=14$ 点の三次元関節点から成るモデルを使用する．この時，時刻 t における各関節点を $\boldsymbol{j}_t^k(k=0,\cdots,N_j-1)$ と表すこととする．なお，以下では映像中に人物が一人であると想定するが，複数人存在する場合にも容易に拡張が可能である．

提案手法の処理の流れを図 3 に示す．まず，入力として得られた多視点映像に対して二次元関節位置推定手法を適用する．次に，得られた関節位置を対応点として各カメラの外部パラメータを初期値とし，さらにその外部パラメータを用いて各時刻の各関節の三次元位置を求める (3.1 節)．次に，3.2.1 節において定めた再投影誤差，および 3.2.2 節において定めた関節に関する制約に基づいた誤差関数を最適化することで外部パラメータを求める．また，3.3 では実装時における追加の処理に関して述べる．以下では各処理に関して詳しく述べる．

3.1 初期値の推定

図 4 に示すように，カメラ C_i で撮影した映像の時刻 $t(t=0,\cdots,N_t-1)$ において検出された $\boldsymbol{j}_t^k$ の二次元位置を ${}^i\boldsymbol{j}_t^k$ とする．これらの二次元関節点を用い，基準カメラ C_0 とあるカメラ $C_i(i>0)$ 間で 8-point algorithm で求めた基礎行列を分解することで外部パラメータの初期値 $R_i^{init}, \boldsymbol{t}_i^{init}$ を得る．次にこの外部パラメータと投影点の座標値 ${}^i\boldsymbol{j}_t^k$ を利用して三次元関節位置 $\boldsymbol{j}_t^{k,init}$ を求める．最後に，求めた $\boldsymbol{j}_t^{k,init}$ と他のカメラでの二次元関節位置 ${}^i\boldsymbol{j}_t^k$ の間で PnP 問題 [2] を解くことで全てのカメラの外部パラメータの初期値 $R_i^{init}, \boldsymbol{t}_i^{init}, (i=1,\cdots,N_c-1)$ を求める．

3.2 最適化関数

提案手法では，以下のように再投影誤差に関する項 $E_{re}(P,J)$ および関節モデルに関する項 $E_{joint}(P,J,L)$ の 2 つの項から成る最適化関数を定める．

$$E(P,J,L) = \lambda_{rep}E_{rep}(P,J) + \lambda_{joint}E_{joint}(J,L) \tag{2}$$

なお，P は全てのカメラの外部パラメータ，J は全ての関節の三次元位置を表し，L は 3.2.2 節で述べる各関節間の距離を表す．また，λ_{rep} および λ_{joint} は各項の重みを表す係数である．提案手法では 3.1 節で求めた各カメラの外部パラメータおよび三次元関節位置を初期値とし，この式 (2) を Levenburg-Marquardt 法で非線形最適化することで各パラメータの最適値 $R_i^{opt}, \boldsymbol{t}_i^{opt}, \boldsymbol{j}_t^{k,opt}$ を求める．以下では各項について詳しく述べる．

3.2.1 検出誤差を許容する再投影誤差

チェスコーナーの検出精度がサブピクセル精度であることに対し，提案手法で用いる各関節位置は数ピクセル以上の検出誤差を含む．そのため，従来のバンドルアジャストメントのように再投影した点の座標値と検出した点の座標

Vol.2017-CG-168 No.30
Vol.2017-DCC-17 No.30
Vol.2017-CVIM-209 No.30
2017/11/9

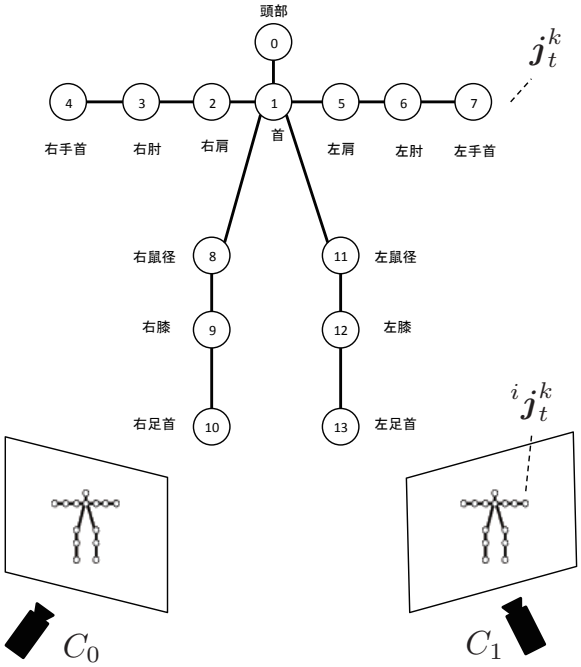

図 4　利用する人体モデル

値の誤差を最小化する場合には，誤差の大きな検出点に最適化されてしまい，外部パラメータの推定精度が悪化する場合がある．そこで，提案手法では再投影誤差における誤差の与え方を緩和することでこの問題を回避する．

提案手法では入力として [5] などの二次元関節位置推定手法で得られた推定値を用いているが，これらは各関節について図 5 のように連続的に変化する confidence map におけるピークの位置を関節位置として出力している．そこで提案手法では，ある関節点の二次元検出点の座標値を $\boldsymbol{j}$ とし，対応する再投影点の座標値を $\tilde{\boldsymbol{j}}$ としたとき，$\tilde{\boldsymbol{j}}$ が confidence map における confidence の高い領域にある場合は再投影誤差の値が小さくなるように誤差の重みを変化させることで検出誤差の影響を緩和する．本稿では confidence map において各関節の confidence の高い領域が正規分布に従うと仮定し，以下の誤差関数を定める．

$$E_{rep}(P,J) = \Sigma_{t=0}^{N_t-1}\Sigma_{i=0}^{N_c-1}\Sigma_{k=0}^{N_j-1} g(\boldsymbol{j},\tilde{\boldsymbol{j}}), \tag{3}$$

$$g(\boldsymbol{j},\tilde{\boldsymbol{j}}) = (f(0) - f(d(\boldsymbol{j},\tilde{\boldsymbol{j}})))d(\boldsymbol{j},\tilde{\boldsymbol{j}}), \tag{4}$$

$$d(\boldsymbol{j},\tilde{\boldsymbol{j}}) = ||\boldsymbol{j} - \tilde{\boldsymbol{j}}||. \tag{5}$$

ただし，$f(x)$ は平均 0，分散 σ^2 の正規分布の確率密度関数とする．

この緩和により，検出誤差による精度悪化を回避することが期待できる一方，解が一意に定まらなくなることが懸念される．これに対し，観測対象が人体であるという前提に基づいて各関節の三次元位置に制約を与えることで，この問題を回避することを試みる．

図 5　[5] を用いて得られた右肩の confidence map の例．

3.2.2　関節に関する制約

観測している対象が人体であるということから，本研究では以下の２つの制約を導入する．

(1)　特定の各関節間の長さは各時刻において一定である．

(2)　各関節の三次元位置は時間方向に連続的に変化する．

これらに関する制約項をそれぞれ E_{length} および E_{motion} とし，E_{joint} を以下のように定義する．

$$\begin{aligned} E_{joint}(J,L) = \lambda_{length}E_{length}(J,L) \\ + \lambda_{motion}E_{motion}(J) \end{aligned} \tag{6}$$

ただし，λ_{length} および λ_{motion} はそれぞれ E_{length} および E_{motion} に関する係数である．以下ではそれぞれの誤差項について述べる．

(1) 各関節間の長さに関する制約：図 4 に示す関節モデルにおいて，k 番目の関節と k' 番目の関節のペアを $\langle k,k'\rangle$ と表すと，$\langle 2,3\rangle$ や $\langle 8,9\rangle$ などはそれぞれ上腕骨や大腿骨といった一本の骨の両端を表していることから，これらの関節間の距離は時間方向に一定であることがわかる．提案手法では $N_p = 9$ 個の関節ペア $P = \{\langle 2,3\rangle, \langle 3,4\rangle, \langle 5,6\rangle, \langle 6,7\rangle, \langle 8,9\rangle, \langle 8,11\rangle, \langle 9,10\rangle, \langle 11,12\rangle, \langle 12,13\rangle\}$ の関節間の距離が一定であるとし，関節間の長さに関する制約項 $E_{length}(J,L)$ を以下のように定める．

$$E_{length}(J,L) = \Sigma_{t=0}^{N_t-1}\Sigma_P |||\boldsymbol{j}_t^k - \boldsymbol{j}_t^{k'}|| - l(\langle k,k'\rangle)| \tag{7}$$

なお，$l(\langle k,k'\rangle)$ は関節ペア $\langle k,k'\rangle$ 間の距離であり，$L = \{l(\langle 2,3\rangle), \cdots, l(\langle 12,13\rangle)\}$ である．なお，$l(\langle k,k'\rangle)$ の初期値は $l(\langle k,k'\rangle) = 1/N_t\Sigma_{t=0}^{N_t-1}||\boldsymbol{j}_t^{k,init} - \boldsymbol{j}_t^{k',init}||$ として与える．

(2) 各関節の三次元位置に関する制約：各関節の三次元位置は時間変化に伴い急に消失したり移動することなく，時間方向に連続的に滑らかに変化すると仮定する．しかしながら，初期値として推定した三次元関節位置 $\boldsymbol{j}_t^{k,init}$ は検出誤差の影響からこの仮定を満たさない場合がある．これに対し，提案手法では各関節の三次元位置の初期値 $\boldsymbol{j}_t^{k,initi}$ に対してカルマンフィルタを適用することで平滑化した三

次元位置を $KF(\boldsymbol{j}_t^{k,init})$ とした時，各 $\boldsymbol{j}_t^k$ が滑らかに変化するよう以下のように $E_{motion}(J)$ を定めた．

$$E_{motion}(J) = \Sigma_{t=0}^{N_t-1}\Sigma_{k=0}^{N_j-1}||KF(\boldsymbol{j}_t^{k,init}) - \boldsymbol{j}_t^k|| \quad (8)$$

なお，本稿ではカルマンフィルタにおいてランダムウォークモデルを用いた．

3.3 実装

提案手法において利用している二次元関節位置推定手法では，オクルージョンなどの理由で関節位置の検出に失敗することがある．そこで，3.1 節で初期値の推定をする際，ある関節点 $\boldsymbol{j}_t^k$ に関して全てのカメラにおいて検出が可能な場合のみ，その検出点を利用することとする．また，初期値の推定において，検出されなかった関節点の三次元位置に関しては，対応する関節点の前後のフレームににおける三次元位置を利用して 3 次スプライン補間を行うことで求めた．

4. 実験

本節では実データおよび CG データを利用して，提案手法の性能を調査する．

4.1 実データを用いた評価

4.1.1 実験環境

実データとして，[15] の Soccer Juggling データセットを利用する．これらは図 6 に示すように 1 名の被写体を 4 視点から撮影した映像から成る．映像は同期済みであり，解像度は 1032 × 778 ピクセル，fps は 30, フレーム数は 530 である．また，関節位置推定手法として [5] を利用し，最適化には Ceres Solver [16] を用いた．

また，評価関数として，回転行列 R に関しては Riemannian 距離 E_R [17] と各軸に分解した際の誤差 E_r を用い，並進ベクトルに関してはスケールを実データに合わせた後に二点間の距離 $E_{\boldsymbol{t}}$ を用いた．それぞれの定義は以下のとおりである．なお，X_g のように g が添付されたパラメータはそのパラメータの真値を表すこととする．

$$E_R = \frac{1}{\sqrt{2}}||\mathrm{Log}(R^\top R_g)||_F, \quad (9)$$

$$\mathrm{Log}R' = \begin{cases} 0 & (\phi = 0), \\ \frac{\phi}{2\sin\phi}(R' - R'^\top) & (\phi \neq 0). \end{cases} \quad (10)$$

ただし，$\phi = \cos^{-1}(\frac{\mathrm{tr}R'-1}{2})$ である．また，E_r は

$$E_r = \frac{1}{3}\Sigma_l|\theta_l - \theta_{lg}| \quad (11)$$

と定める．なお，$\theta_l(l = x, y, z)$ は R が各軸ごとの回転の合成で表される時の角度を表す．さらに，E_t は以下のように表す．

$$E_t = ||s\boldsymbol{t} - \boldsymbol{t}_g|| \quad (12)$$

ここで，s は実データのスケールを表すパラメータであり，$s = ||\boldsymbol{t}||/||\boldsymbol{t}_g||$ として求める．

次に，比較する手法を表 1 に示す．(a) の初期値は 3.1 節で求めた値を用いる．(b) は従来の再投影誤差を用いた手法である．(c) から (f) は式 (2) および式 (6) で導入した誤差項を組み合わせたものであり，(f) が今回の提案手法である．なお，E_{rep} を用いない組み合わせに関しては，解が安定して収束しなかったため比較を省略した．

表 1 比較手法 (各係数の表記は省略).

手法	用いる最適化関数
(a)	- (初期値)
(b)	再投影誤差項 (Bundle adjustment)
(c)	E_{rep}
(d)	$E_{rep} + E_{length}$
(e)	$E_{rep} + E_{motion}$
(f)	$E_{rep} + E_{length} + E_{motion}$

表 2 評価結果（実データ利用）.

手法	E_R	E_r(deg)	$E_{\boldsymbol{t}}$(mm)
(a)	0.1333	4.337	516.4
(b)	0.2057	7.460	921.7
(c)	0.1521	5.451	760.6
(d)	0.1533	5.496	905.8
(e)	0.3634	13.643	1439.5
(f)	0.0558	2.308	363.9

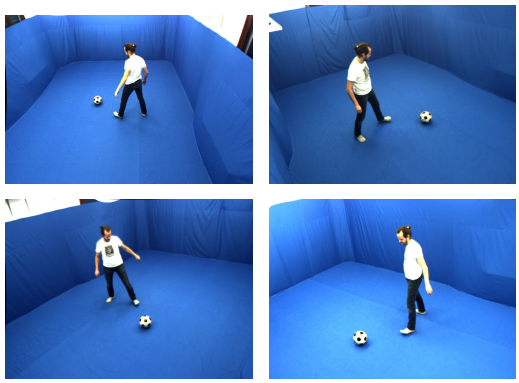

図 6 実データとして用いた入力映像 [15].

4.1.2 結果と考察

表 2 に各手法で求めた各評価関数の平均値を示す．結果から，(b) の結果が初期値より悪化していることが見て取れる．これは従来の再投影誤差を用いた手法では関節位置の検出誤差に最適化されてしまったためであると考えられる．また，この再投影誤差を緩和する形で導入した誤差項 E_{rep} を用いた (c) では，(b) のように大きく悪化することがなくなった一方で，精度の向上は見られなかった．E_{rep} の定義では誤差ごとに重みを変えており，これは例外値に小さな重みを与えるロバスト推定手法の一つである M-estimator

Vol.2017-CG-168 No.30
Vol.2017-DCC-17 No.30
Vol.2017-CVIM-209 No.30
2017/11/9

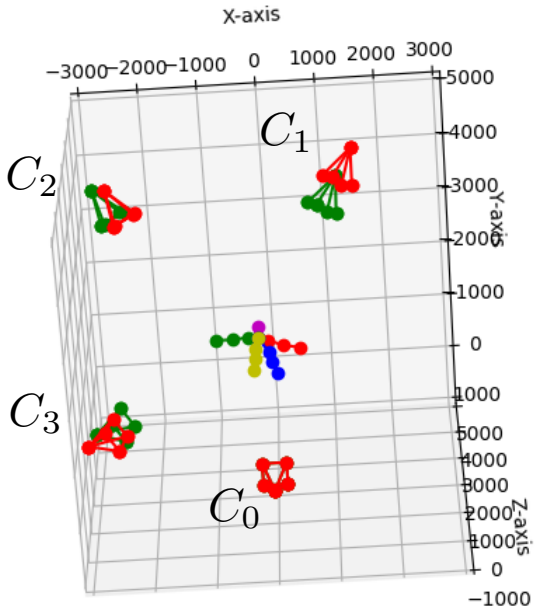

図 7 推定したカメラ位置（赤：提案手法，緑：真値）と，提案手法で推定された各関節の三次元位置．

と同様の振る舞いを行うと考えられる．M-estimator では良い初期値を与えなければ最適解に収束することは保証されないことが知られているが，(c) に関しても再投影誤差を緩和することによって解の候補が多くなってしまい，最適解に収束しなかったと考えられる．一方，(c) に対して観測対象である人体に関する制約項 E_{length}, E_{motion} を導入した (d),(e) ではそれぞれ精度の改善が見られ，さらにこれらを統合した (f) で最も良い精度で推定していることが確認できる．これらの結果から，人体に関する制約項 E_{length}, E_{motion} を導入することで，より良い解に収束したと考えられる．

また，図 7 に提案手法で推定したカメラ位置とその真値，および提案手法で推定した各関節の三次元位置を描画する．図 7 から，提案手法で推定したカメラ位置と真値が概ね同じ位置にプロットされていることが確認できる．

以上の結果から，実データにおいて提案手法は関節位置を利用したカメラ位置推定が可能であることが確認できた．

4.2 CG データを用いた評価

提案手法はチェスボードを持ち込むことが困難なスタジアムのような広域な場所において，カメラ間の視差が大きな状況での利用を想定しているが，そのような条件での外部キャリブレーションは現在も主要な研究課題の一つであり，厳密な真値を用意することが難しい．これに対し，本節では CG データを利用した実験を実施する．CG データでは広域な環境においても外部パラメータの値の真値が得られる．

また，提案手法では式 (2) を最適化することで外部パラメータと同時に人体関節の三次元位置も求まる．提案手法で想定している各関節は人体に内包されているため，実データでの評価が難しい．これらを評価する際にも CG データが利用できる．

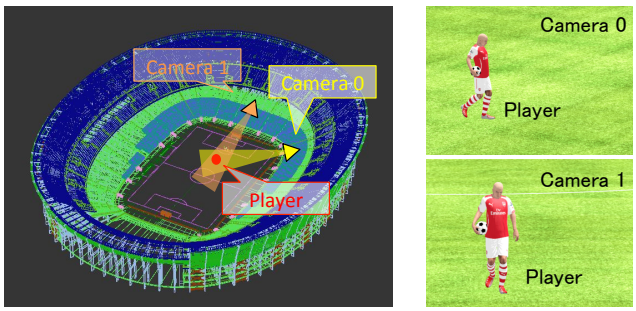

図 8 スタジアム環境を想定した CG データ．左：カメラと被写体の位置関係．右：カメラ 0 およびカメラ 1 で撮影された画像．

表 3 評価結果（CG データ利用，外部パラメータ）．

手法	E_R	E_r(deg)	E_t(mm)
(a)	0.0785	2.009	1571.8
(b)	0.0772	1.606	5137.7
(c)	0.0762	1.741	3984.1
(d)	0.0769	1.623	9162.4
(e)	0.0771	1.635	10365.3
(f)	0.0777	1.711	75.8

4.2.1 結果と考察

4.2.2 実験環境

本実験では図 8(左) に示すように大規模なスタジアムに設置されたカメラで撮影した映像データを用いてキャリブレーションを行う．用いる映像には図 8 のように 1 名の人物が写り込んでいることを想定する．カメラは 2 台利用し，焦点距離は 5333 ピクセル（ミリ換算では 300mm 相当），解像度は 640 × 480 ピクセルとし，レンズ歪みは無いものとした．また，各映像は同期が取れているものとする．映像は 60fps で撮影し，フレーム長は 500 フレームとした．カメラ C_0 を基準とした時，カメラ C_1 の外部パラメータの真値は以下とした．

$$R_g = \begin{bmatrix} 0.5386 & -0.1659 & 0.8260 \\ 0.1209 & 0.9854 & 0.1191 \\ -0.8338 & 0.0357 & 0.5509 \end{bmatrix}, \boldsymbol{t}_g = \begin{bmatrix} -37644.1 \\ -5415.7 \\ 13145.9 \end{bmatrix} \quad (13)$$

外部パラメータの評価に関しては 4.1 節と同様の評価関数を使用する．また，関節位置 $\boldsymbol{j}$ の評価に関しては，開始フレームにおいて関節位置を真値の位置に合わせた後，各時刻における関節位置の誤差を以下の評価関数を用いて評価した．

$$E_j = ||s_j \boldsymbol{j} - \boldsymbol{j}_g|| \quad (14)$$

ここで，s_j はスケールを表すパラメータであり，3.2.2 節で用いた関節ペアに関して $s_j = 1/N_p \Sigma_P ||l(\langle k, k' \rangle) - l_g(\langle k, k' \rangle)||$ として求めた．

Vol.2017-CG-168 No.30
Vol.2017-DCC-17 No.30
Vol.2017-CVIM-209 No.30
2017/11/9

表 4 評価結果（CG データ利用，関節の三次元位置）．

関節	(a)	(b)	(c)	(d)	(e)	(f)
関節 1 (首)	93.1 ± 66.7	80.6 ± 61.0	80.4 ± 58.2	82.1 ± 73.0	80.8 ± 58.1	81.2 ± 44.0
関節 2 (右肩)	115.1 ± 75.8	102.3 ± 60.3	102.3 ± 53.7	109.7 ± 68.6	104.6 ± 57.9	110.1 ± 37.4
関節 3 (右肘)	128.1 ± 58.3	124.3 ± 47.8	121.8 ± 44.3	117.6 ± 52.6	119.8 ± 44.2	134.9 ± 41.2
関節 4 (右手首)	131.6 ± 68.2	124.4 ± 53.6	122.0 ± 47.5	125.0 ± 60.8	118.1 ± 48.5	123.7 ± 32.8
関節 5 (左肩)	114.2 ± 72.1	101.5 ± 68.6	101.9 ± 66.2	115.0 ± 76.1	103.4 ± 66.9	91.6 ± 47.6
関節 6 (左肘)	100.5 ± 56.1	88.0 ± 54.8	91.5 ± 48.6	95.0 ± 60.4	91.8 ± 46.7	93.3 ± 31.3
関節 7 (左手首)	110.2 ± 72.8	101.3 ± 67.5	98.1 ± 59.4	109.1 ± 69.4	95.1 ± 57.7	91.2 ± 31.9
関節 8 (右鼠径)	112.6 ± 45.4	103.4 ± 31.3	100.6 ± 27.9	116.5 ± 36.9	96.8 ± 29.7	94.8 ± 22.5
関節 9 (右膝)	173.7 ± 142.3	168.3 ± 123.1	169.1 ± 116.1	174.4 ± 127.7	164.8 ± 122.9	173.5 ± 105.5
関節 10 (右足首)	137.8 ± 38.7	146.0 ± 33.0	153.2 ± 32.0	148.9 ± 33.8	143.5 ± 33.2	168.0 ± 38.2
関節 11 (左鼠径)	117.0 ± 44.8	105.1 ± 33.9	101.2 ± 27.2	115.9 ± 36.5	99.2 ± 28.1	95.7 ± 23.8
関節 12 (左膝)	127.6 ± 43.7	123.0 ± 35.1	124.7 ± 27.1	132.8 ± 37.2	118.1 ± 27.1	122.2 ± 18.4
関節 13 (左足首)	139.8 ± 55.2	142.9 ± 52.7	149.5 ± 46.2	148.8 ± 49.7	141.5 ± 45.3	157.7 ± 46.5
平均	123.2 ± 64.6	116.2 ± 55.6	116.6 ± 50.3	122.4 ± 60.2	113.6 ± 51.2	118.3 ± 40.1

表 3 に外部パラメータの評価結果を示す．結果から，回転行列に関しては各手法とも大きな差は無いが，並進ベクトルに関しては提案手法 (f) が最も精度良く求められていることが確認できる．(b) に関しては実データと同様に精度が悪化しているが，これは同様に誤差に最適化されてしまったためであると考えられる．一方，(d), (e) に関しては (c) と比較して精度が悪化している．これは (d), (e) で導入した制約項に関し，それらの制約を満たす解が複数存在していたためと考えられる．

表 4 に各関節ごとの三次元位置の評価結果を示す．なお，頭部に該当する関節 0 に関しては，[5] で推定する頭部の位置と CG データで表現する頭部の位置が異なっていたため，評価を省略した．表 4 から，各手法とも概ね 120mm 前後の誤差で推定できていることがわかる．また，各間接ごとの推定誤差の大小に関しては各手法とも共通の傾向を示していることから，これらは [5] の推定誤差が反映されたものであると考えられる．ただ，比較手法は標準偏差が非常に大きく，精度が安定していないことが見て取れる．これに対し，提案手法では比較手法に比べて標準偏差が小さく安定して推定出来ていることが確認できる．

さらに，図 9 に提案手法で推定したカメラ位置およびその真値と，推定された各関節の三次元位置を示す．図 9 から，カメラ位置に関して多少位置ズレはあるものの，概ね同じ位置に推定できていることが確認できる．

以上の結果より，スタジアムのような大規模な環境において，視差の大きなカメラ間でも人体関節を利用することで外部キャリブレーションが実施できることを示した．

なお，本実験ではスタジアムのような広域環境における外部パラメータおよび関節位置の真値取得の困難さの観点から CG データを用いて評価を行ったが，CG データを用いた評価は実際の環境において同等の精度を保証するものではなく，実験結果の信頼性の観点において十分ではない．

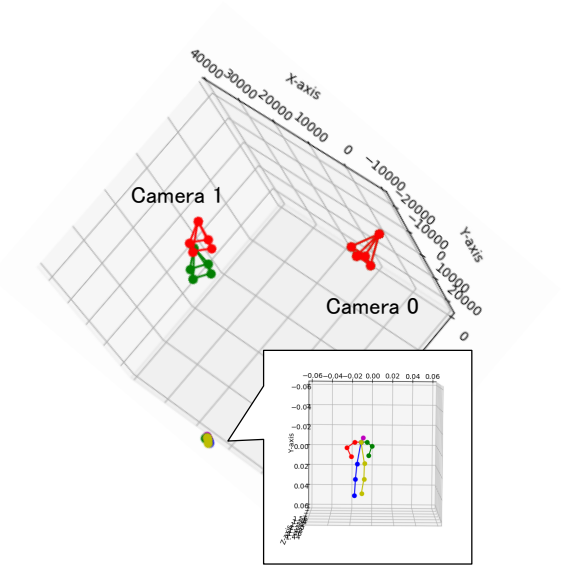

図 9 推定したカメラ位置（赤：提案手法，緑：真値）と，提案手法で推定された各関節の三次元位置．

今後はこの CG データの利便性を利用しつつ，その実験結果の信頼性を保証するためにはどのような追加実験，比較が必要かを検討する必要がある．

5. 結論

本研究では人体関節の投影点を対応点として利用する新たな外部キャリブレーション法を提案した．提案手法では関節点の検出誤差を許容する再投影誤差項と，観測対象が人体であるという前提を利用した関節間距離に関する制約項および関節位置は滑らかに変化するという制約項からなる最適化関数を定義し，非線形最適化することで外部キャリブレーションを実施した．実データおよび CG データを用いた実験では提案手法が概ね正しい外部パラメータを推定しており，その適用可能性を示した．今後は本アプローチによる精度の限界の調査および高精度化に取り組む．

参考文献

[1] Zhang, Z.: A flexible new technique for camera calibration, *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 1330–1334 (2000).
[2] Lepetit, V., Moreno-Noguer, F. and Fua, P.: EPnP: An Accurate O(n) Solution to the PnP Problem, *International Journal of Computer Vision*, Vol. 81, No. 2 (2008).
[3] Hartley, R. I. and Zisserman, A.: *Multiple View Geometry in Computer Vision*, Cambridge University Press (2000).
[4] Toshev, A. and Szegedy, C.: Deeppose: Human pose estimation via deep neural networks, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1653–1660 (2014).
[5] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (2017).
[6] Huang, S., Ying, X., Rong, J., Shang, Z. and Zha, H.: Camera Calibration From Periodic Motion of a Pedestrian, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (2016).
[7] Homayounfar, N., Fidler, S. and Urtasun, R.: Sports Field Localization via Deep Structured Models, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (2017).
[8] Sinha, S. N. and Pollefeys, M.: Camera Network Calibration and Synchronization fromSilhouettes in Archived Video, *International Journal of Computer Vision*, Vol. 87, No. 3, pp. 266–283 (2010).
[9] Boyer, E.: On using silhouettes for camera calibration, *Proc. Asian Conf. on Computer Vision*, pp. 1–10 (2006).
[10] Puwein, J., Ballan, L., Ziegler, R. and Pollefeys, M.: Joint camera pose estimation and 3d human pose estimation in a multi-camera setup, *Proc. Asian Conf. on Computer Vision*, Springer, pp. 473–487 (2014).
[11] Pishchulin, L., Andriluka, M., Gehler, P. and Schiele, B.: Poselet conditioned pictorial structures, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 588–595 (2013).
[12] Sun, M. and Savarese, S.: Articulated part-based model for joint object detection and pose estimation, *Proc. International Conf. on Computer Vision*, IEEE, pp. 723–730 (2011).
[13] Wei, S.-E., Ramakrishna, V., Kanade, T. and Sheikh, Y.: Convolutional pose machines, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 4724–4732 (2016).
[14] Ramakrishna, V., Munoz, D., Hebert, M., Bagnell, J. A. and Sheikh, Y.: Pose machines: Articulated pose estimation via inference machines, *Proc. European Conf. on Computer Vision*, Springer, pp. 33–47 (2014).
[15] Ballan, L. and Cortelazzo, G. M.: Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes, *Proc. International Symposium on 3D Data Processing, Visualization and Transmission* (2008).
[16] Agarwal, S., Mierle, K. and Others: Ceres Solver, `http://ceres-solver.org`.
[17] Moakher, M.: Means and averaging in the group of rotations, *SIAM J. Matrix Anal. Appl.*, Vol. 24, pp. 1–16 (2002).