論　文

３次元形状計測における不完全性のモデル化に基づいた複雑な人物動作の推定

延原　章平[†a)]　　宮本　　新[†]　　松山　隆司[†]

Complex 3D Human Motion Estimation by Modeling Incompleteness in 3D Shape Observation

Shohei NOBUHARA[†a)], Arata MIYAMOTO[†], and Takashi MATSUYAMA[†]

あらまし　重度の接触を伴う複雑な姿勢をとっている人物の３次元表面形状においては，接触面がどのような配置のカメラからも観測不可能となる．そのため観測多視点画像から推定される３次元表面形状にはこの接触面に対応する形状は含まれ得ない．つまり対象本来の３次元表面形状に対して観測から推定された３次元形状には欠損が生じ得る．また接触面以外にも，観測から推定された３次元形状中にはカメラ配置上観測不可能であった部分や，観測された対象表面であっても形状推定手法の限界として残ったアーティファクトのように，本来の対象形状とそもそも形状が一致し得ない箇所が存在する．本研究では観測３次元形状においてこのような欠損，不一致が生じる表面領域をモデル化し，このような領域を除外して本来の形状と観測形状の間の対応付けを行うことで，重度の接触を含む動作も推定できる頑健な姿勢推定手法を提案する．

キーワード　人物動作推定，３次元ビデオ，多視点映像，ICP

1. はじめに

本論文では多視点映像を用いた重度の接触を伴う複雑な人物動作の推定という問題を，姿勢パラメータ p によってその形状が決まる３次元モデル $M(p)$ と，３次元観測形状 M_t の間のマッチング問題であるとして扱う．このような問題に対して従来の手法はモデルと観測が共に同一の表面形状を持っていると仮定し，両者を ICP [1] によってマッチングしていた [2]．つまり図 1(a) のようにモデルと観測の表面形状を構成する頂点集合をそれぞれ $M(p)$ と M_t （図中○と●の点群）とした二部グラフで考えると，$M(p)$ と M_t の要素間の対応関係をユークリッド距離に基づく最近傍探索によって定義し，この距離が最小化するような姿勢パラメータ p をもって姿勢推定としていた．

しかし重度の接触を伴う姿勢を対象がとっている状況下ではモデルと観測の両者が同一の表面形状を持つという仮定は成り立たない．これは以下の２つの理由による．

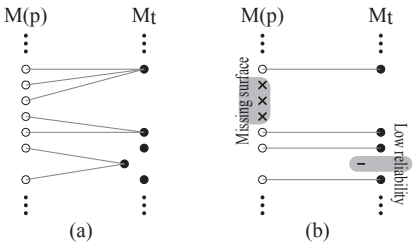

図 1　表面形状同士の場合の対応付け
Fig. 1　Matching between surfaces

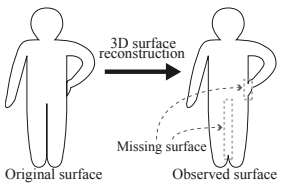

図 2　接触によって観測不可能となった領域における表面形状の消失
Fig. 2　Unrecoverable hidden 3D surface by body contacts

† 京都大学大学院情報学研究科　〒 606–8501 京都市左京区吉田本町
Graduate School of Informatics, Kyoto University　Yoshida-honmachi, Sakyo, Kyoto, 606-8501, Japan
a) E-mail: nob@i.kyoto-u.ac.jp

(a) モデル側： 対象の体節同士が接触した場合，接触面はいかなるカメラからも観測不能なので，観測形状中には対応する表面領域が存在し得ない．つまりモデル側から見ると観測側には適切な対応領域が存在しないことになるが，従来法ではモデル中のこのような表面領域においても対応点を推定形状中で探索するため，誤差の不適切な増大を招く（図 2）．そのためモデル側から接触領域を除外してマッチングを行うべきである．

(b) 観測側： たとえ体節同士が接触していなくても，あるカメラ配置の下で対象を観測する以上は，自己遮蔽などによって対象が観測不可能な領域を伴う姿勢をとることは完全には避けられない．またたとえ観測可能であったとしても，形状復元プロセス自身の不完全性から，対象本来の形状ではない部分がアーティファクトとして観測形状中に含まれることも避けがたい．このように観測形状のうち，カメラから観測不可能な領域や，形状推定の信頼性が低い部分は，モデル側に対応する領域が存在するとは限らないため，やはり誤差の不適切な増大を招く要因となる．そのため頂点集合 M_t から除外してマッチングを行うべきである．

3 次元形状計測における不完全性に関する考察に基づき，図 1(b) のように，モデル側，観測側それぞれの頂点集合 $M(p)$ および M_t から，該当する頂点（図中×および−）をそれぞれ除去した上で対応関係を求め，その上でこの頂点間の距離を最小化する手法を提案する．本手法では (a) モデル側の観測不可能領域を，人体モデルを構成する体節間の距離によってモデル化し，(b) 観測側は推定された形状表面における photo-consistency （観測可能なカメラ間でのテクスチャの一致度．詳細は後述）によって，モデル化する．

以下第 2 節において関連研究に対する本研究の位置付けについて議論を行い，第 3 節で提案する 3 次元形状計測における不完全性モデルの定義を述べる．第 4 節でそれを用いた動作推定アルゴリズムについて述べた後，第 5 節 CG データおよび実観測データに対する姿勢推定を行うことで提案手法の有効性を定量的・定性的に示す．最後に第 6 節で結論と今後の課題を述べる．

2. 関連研究に対する位置付け

姿勢推定に対するこれまでの研究は (1)3 次元の人物モデルを 3 次元の観測データに当てはめるもの [2]〜[6] と，(2)2 次元画像に対してモデルを当てはめるもの [7]〜[15] とに大別することができる [16]．後者には 3 次元モデルを 2 次元画像から得られるエッジ特徴などに当てはめるものや，学習済みの事例との比較を行うもの [7], [8] が含まれるが，重度の接触を伴う複雑な姿勢を 2 次元の投影像から推定することは容易ではない．そのため本研究では (1) の人物モデルを 3 次元データに当てはめるアプローチを採用する． 次に人物モデルを用いた姿勢推定は，(1a) 対象の真の形状を持ったモデル用意して，姿勢のみを推定するアプローチ [2] [3] と，(1b) 汎用的な人物モデルを用意して，姿勢と対象の詳細な形状を同時に推定するアプローチ [4] の 2 つに分類される．たとえば [4] では “soft object” と呼ばれる柔軟な表面形状とそれの位置をコントロールする骨からなるモデルを用いて姿勢（骨の位置パラメータ）と形状（表面形状の変形パラメータ）の同時推定を行っている． ここで (1a) で必要となる対象形状の取得が煩雑である場合，あるいは対象形状が複雑な形状変化を行う場合は (1b) のアプローチがより適していると言えるが，本論文では人物の動作を多関節モデルで表すことができると仮定し，また単純な姿勢では多視点映像から必要十分な精度で形状復元できること [17]〜[19] を踏まえて，(1a) のアプローチを採る．

対象の真の形状を持った多関節モデルを観測形状に当てはめる問題は，これまで ICP に代表される手法によって取り組まれてきた．たとえば [2] では，法線情報をモデルと観測の間のマッチングに使用することで，誤対応を防ぎ，姿勢推定の精度向上を図っている．また [6] ではテクスチャの利用と確率的な最適化プロセスによって誤対応への対応が図られていた．

しかし対象が接触を伴う姿勢をとった場合，図 2 に示したように観測形状とモデル形状の間に差異が原理的に生じ，このような領域ではそもそもモデルと観測の間には対応関係は存在しない．つまり従来の手法が暗黙に仮定していた “モデルと観測は同じ形状を持つ” という前提は成り立たない．このような状況で生じる誤対応に対しては，色情報や法線情報の利用によるアウトライアの除去といったアプローチでは本質的には解決できず，重度の接触を伴う動作の推定には適していないといえる．また 3 次元形状復元によって得られる観測形状に含まれる誤りにも対処できていない． これに対して本論文では，観測プロセスによって生じるモデル形状と観測形状との間の差異を明示的にモデル化することでこの問題の解決を図っている．

3. 3次元形状計測の不完全性のモデル化

本論文では，キャリブレーション済の多視点カメラ群によって対象を撮影し，対象の3次元形状が三角形メッシュデータとして各時刻で観測によって得られていると仮定する．以下，時刻 t における対象の観測データを M_t と表す．また対象の骨格構造は既知であるとし，姿勢パラメータ p によってその表面形状が決定される対象の多関節 skin-bone モデル $M(p)$ が与えられているとする．ここで $M(p)$ の表面形状は対象の真の表面形状と一致しているとする．M_t および $M(p)$ の獲得方法については後述するものとして，本節では我々が提案する3次元形状計測の不完全性のモデルについて述べる．

3.1 人物モデル側におけるモデル化

対象が接触を伴う姿勢をとっている場合，図2に示したように観測される対象形状 M_t は $M(p)$ に比べて接触部分で表面形状が欠落する．そのため単純な ICP のように互いのメッシュ頂点における最近傍点までの距離を誤差量として使用するだけでは，欠落部分において本来存在しないはずの最近傍点を探索してしまい，結果として誤差の不適切な増大を招く．そのため真の姿勢において誤差が最小となり得ず，正しく骨格の姿勢パラメータ p を推定することが困難となる．

そこで我々は，次に述べるように観測側における表面形状の欠落を $M(p)$ の各部位間の距離を用いて，つまりモデル $M(p)$ 自身から近似的に推定し，モデル中からこれらの点を排除した上でモデルと観測の間の誤差量を計算することでこの問題を解決する．つまり部位同士が近接している場合は観測上は図2のように表面形状が欠落する可能性が高い領域であると見なす．具体的には $M(p)$ を構成する各頂点は，それが属する骨，つまり部位ごとに領域分割されている（後述）ため，$M(p)$ を構成する各頂点について，その近傍に他の部位に属する頂点が存在するかどうかを調べ，もし存在するならば観測上は図2のように表面形状が欠落すると見なす．

まず $M(p)$ を構成するある頂点を v とする．v から最も近くかつ v とは他の部位に属する頂点を $v' \in M(p)$ として，v から v' までの符号付き距離 $d(v)$ を

$$d(v) = \begin{cases} \|v - v'\| & \text{if}(v - v') \cdot n(v) \geqq 0, \\ -\|v - v'\| & \text{otherwise}, \end{cases} \quad (1)$$

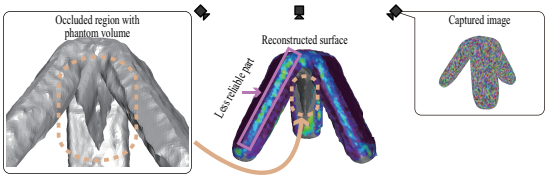

図3 観測不可能領域と観測形状の信頼性
Fig. 3 Unobservable area and reliability of the estimated 3D shape

として定義する．ここで $n(v)$ は頂点 v における法線ベクトルを表し，$d(v)$ は v' が v の属する部位と交差する場合に負の値をとる．このように定義した符号付き距離 $d(v)$ を用いて，本論文ではある頂点 v に関する可観測性 $\rho_c(v)$ を

$$\rho_c(v) = \frac{1}{1 + \exp(-\alpha_c(d(v) - \tau_c))} \quad (2)$$

と定義する．この $\rho_c(v)$ は符号付き距離 $d(v)$ が大きい，つまり他の部位から離れているときは1に近い値をとり，逆に $d(v)$ が小さいか，負のとき，つまり他の部位に接触するか交差しているときは0に近い値をとる．α_c および τ_c はどの程度他の部位に近接したときに v が観測されなくなると見なすかをコントロールしており，観測データに依存するパラメータである．本論文では人手によって与えることとする．ただしこの定式化は計算が簡便である一方で，不適切に大きなパラメータの値を用いると，十分離れた部位をも接触しているものとみなしてしまい，推定精度を損ねることが予想される．したがってより正確には，この定義ではなく頂点の位置，モデルの現在の姿勢，カメラ配置を考慮して計算することがより望ましいと考えられるが，本論文では多視点カメラによる形状復元を入力として用いるため，カメラが対象に対して等方的に分布しているものと仮定し，この式のように頂点位置などによらず近接部位との距離のみでモデル化する．

3.2 観測側におけるモデル化

たとえ体節間に接触が無かったとしても，一定のカメラ配置の下で撮影を行う以上は，自己遮蔽などによって観測不可能な領域を対象が持つことは特に姿勢が複雑であるほど避けがたい．また例え観測可能な領域であったとしても，観測ノイズや形状復元手法の特性によるアーティファクトとして，観測形状中に形状として信頼性の低い部分が含まれうる．このようにどのカメラからも観測されていない領域や，形状として

の信頼性の低い部分は，モデル側に対応する領域が存在するとは限らないため，誤差計算に含めるべきではない．

本研究では人物の 3 次元形状復元手法として現在主流である photo-consistency，形状の滑らかさ，シルエット制約などに基づくコスト関数を graph-cut によって最適化することで，正確さと頑健さを両立した手法 [18], [19] を使用する．得られる形状は visual hull のようにシルエットのみから復元したものと比べるとより正確に対象の表面形状を表しているが，やはり (1) カメラから観測できない領域（図 3 中央，茶色の点線で囲まれた領域と同図左側の拡大図）や，(2)photo-consistency が局所的に悪くともコスト関数を構成する他の項の影響によって対象表面領域とされた部分（図 3 中，紫色の線で囲まれた黄色〜赤色の領域．赤色に近づくほど photo-consistent ではないことを示している．本論文での photo-consistency 関数については後述） など，必ずしもその表面形状が全て photo-consistent とはならず，photo-consistency に基づいてその位置が決定されているわけではない．特に図 3 左の拡大図のように自己遮蔽された領域で生じる phantom volume は，これを photo-consistency に基づいて削ることが不可能であるために，モデル形状との間で重大な差異を生む要因となる（同図の場合，対象は本来 3 本の “足” を持つオブジェクトだが，4 本目の足が phantom volume として中央下部に生じている）．

そこで本研究では，観測形状における photo-consistency の値をその部分における形状の信頼度を表していると見なし，これを後に述べる姿勢推定で使用することで形状として信頼度の低い部分を評価から除外する．本論文では観測形状上の頂点 u における photo-consistency として u を観測可能な全てのカメラ対における Zero-mean Normalized Cross Correlation の平均値を使用し，$[-1:1]$ の値が得られているとする．Zero-mean Normalized Cross Correlation を用いる理由は，撮影画像間での線形な輝度変化に対して不変であるという特徴があり，本研究で想定する基線長が長く，疎なカメラ配置でのステレオに適するためである [20]． ただし図 3 中の灰色領域のように自己遮蔽によって複数台のカメラから観測できない領域については，もっとも低い -1 の値とする．以上を $\mathrm{ZNCC}(u)$ で表すとし，先の式 (2) と同様に

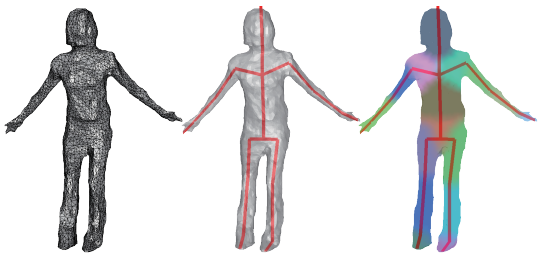

図 4　人物モデルの作成．(a) 撮影画像から復元された 3 次元形状，(b)(a) の形状に骨格モデルを埋め込んだ状態，(c) 骨格モデルと表面形状の対応付けを行った状態．表面の色は各骨への割り当てを示しており，境界部分で色が混ざっている領域は隣接する 2 つの骨の両方に従属していることを示している．

Fig. 4　3D skin-bone model. (a) Estimated 3D shape from multi-viewpoint images. (b) Embedded bone structure into the 3D shape shown by (a). (c) Skin-bone assignment. Colors on the surface indicate the assigned bones. Regions in blended colors indicate that vertices in such regions belong to two bones in their vicinity.

$$\rho_p(u) = \frac{1}{1+\exp(-\alpha_p(\mathrm{ZNCC}(u)-\tau_p))} \qquad (3)$$

によってよりテクスチャが一致している場合に 1 に近い値をとり，一致していないか観測不可能である場合に 0 に近い値をとるように形状の信頼度 $\rho_p(u)$ を定義する．ここで α_p および τ_p はどの程度の一致度をもって形状が正確であると見なすかを決める値であり，観測データとカメラのノイズレベルに依存するパラメータである．本論文の実験では人手によって与えることとする．

4. 提案手法

ここまでで，(1) モデル表面上の頂点 v に対応する点が観測表面上に存在する尤度 $\rho_c(v)$ と，(2) 観測表面上の頂点 u に対応する点がモデル表面上に存在する尤度 $\rho_p(u)$ を定義した．本節ではこれらを用いた複雑な人体の動作推定を可能にする姿勢推定アルゴリズムについて述べる．

以降，まず前節で述べた人物の多関節モデル $M(p)$ を作成する方法について説明し，次に人物の姿勢推定問題の定式化について述べる．

4.1 人物モデルの作成

人物の多関節モデル $M(p)$ はいわゆる skin-bone モデル [21] として表され，関節角に相当する 骨格の姿

勢パラメータ p によってその表面形状である三角形メッシュサーフェースが変形する．本論文では観測した3次元表面形状に既知の骨格構造を対応付けることで，対象の多関節モデル $M(p)$ を作成する．まず観測形状が対象の真の表面形状に十分近いと見なすことのできる時刻 t_0 を観測系列から選び，そのときの観測形状 M_{t_0} に骨格構造を当てはめる（図 4(a) および (b). (b) の太線が当てはめられた骨格構造を示している）．そして M_{t_0} を構成する各頂点は，それぞれ最も近い骨の運動に従属するものとして対応づける（同図 (c). 各色で塗り分けられた領域が，各骨に割り当てられた対象表面領域を表しており，境界部分の色が混ざった領域は，隣接する 2 つの骨に同時に割り当てられていることを示している）．こうして対象の表面形状を表す三角形メッシュが，骨の姿勢パラメータによって変形する多関節人体モデル $M(p)$ が得られる．

なお M_{t_0} の選択と，骨格構造の当てはめを自動かつ適切に行うことはそれ自身困難な課題であるが，本論文の主目的からは逸れるため我々は人手によって行った．また姿勢パラメータ p は，骨格を構成する骨の数を N_p としたとき，人体の中心にあたる位置の骨の位置および姿勢を表す 6 自由度と，それに接続された計 N_p-1 本骨の姿勢（関節角）を表す $3(N_p-1)$ 自由度を合わせて，合計 $3N_p+3$ 自由度を持つものとし，これを最適化の対象とした．具体的な骨の数 N_p については後の評価実験の節において述べる．

4.2 定 式 化

本論文では，人物の姿勢推定問題をモデル $M(p)$ と観測 M_t の形状の差を最小化する骨格の姿勢パラメータ p を求める問題と捉え，パラメータ p によって決まる $M(p)$ と M_t の間の形状誤差関数 $E(M(p), M_t)$ の非線形最小化問題として定式化する．

モデル $M(p)$ と観測 M_t の形状の差 $E(M(p), M_t)$ はモデル表面を構成する各頂点 $v \in M(p)$ とそれに最も近い観測表面形状中の点 u_v との間の二乗距離と，観測表面形状中の各頂点 $u \in M_t$ とそれに最も近いモデル中の点 v_u との間の二乗距離を $\rho_c(v)$ および $\rho_p(u)$ によって重み付けしたものの和とする [2].

$$E(M(p), M_t) = \sum_{v \in M(p)} \frac{\rho_c(v)}{R_c}\frac{\rho_p(u_v)}{R_p}\|v-u_v\|^2 + \sum_{u \in M_t} \frac{\rho_c(v)}{R_c}\frac{\rho_p(u_v)}{R_p}\|u-v_u\|^2 \quad (4)$$

ここで $\|v-u_v\|^2$ および $\|u-v_u\|^2$ はそれぞれ v と u_v，u と v_u の間の二乗距離であり，R_c および R_p はそれぞれ全ての $\rho_c(v)$ および $\rho_p(u)$ の和で，$\rho_c(v)$ と $\rho_p(u)$ を正規化する項である．この式 (4) を最小化する時刻 t の姿勢パラメータ p_t は修正 Levenberg-Marquardt 法によって求める．ただしパラメータ p_t の初期値は前時刻 p_{t-1} の値を用いる．

4.3 3 次元形状計測の不完全性モデルを用いた人物動作の推定

こうして定義した式 (4) を用いて，以下の手順によって対象人物の動作を推定する．

Step 1. 対象モデル $M(p)$ を観測の開始時刻 t_0 における観測形状 M_{t_0} から作る．このとき時刻 t_0 における姿勢パラメータ p_{t_0} は前述のように既知であるとする．

Step 2. 時刻 $t = t_0 + 1$ とする．

Step 3. 式 (4) を最小化する姿勢パラメータ p を求める（4.2 節）．ただし p の初期値は前時刻のパラメータ p_{t-1} とする．得られた姿勢パラメータを時刻 t における推定姿勢パラメータ p_t とし，t が観測終了時刻 t_e であれば Step 4. へ，そうでなければ $t := t+1$ として Step 3. を繰り返す．

Step 4. 得られた $p_{t_0}, \ldots p_{t_e}$ をもって姿勢推定結果として終了．

5. 評 価 実 験

5.1 CG モデルを用いた定量的評価

まず本手法の有効性を定量的に評価するために，図 3 に示したオブジェクトを図 5 のような環境で仮想的に撮影し，得られた 3 次元形状に対して姿勢推定を行う．図 5 に示すように実験ではカメラを 15 台使用し，CG モデルは 100cm の腕が 3 本連結した形状である．またオブジェクトは図 3 の撮影画像の例のようなテクスチャを持っている．この CG モデルに対して $N_p = 6$，つまり 6 本の骨を持ち，$3N_p + 3 = 21$ 自由度を持つ骨格モデルを当てはめて姿勢推定を行った．また式 (2) の α_c および τ_c はそれぞれ 2.0, 2.5 の値を用い，式 (3) の α_p および τ_p はそれぞれ 5.0, 0.95 と

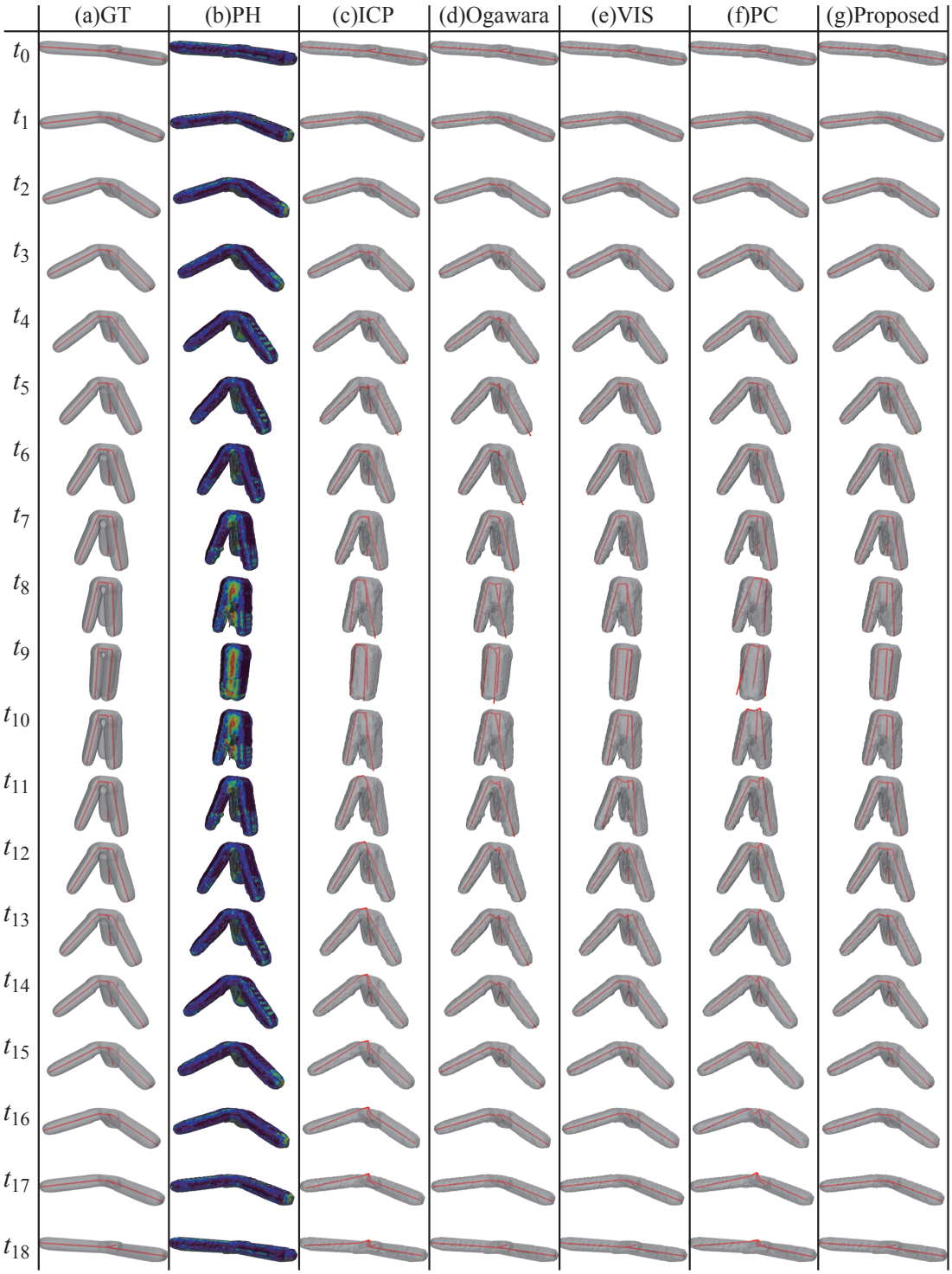

図 6　CG データを用いた定量的評価
Fig. 6　Quantitative evaluation using synthesized data

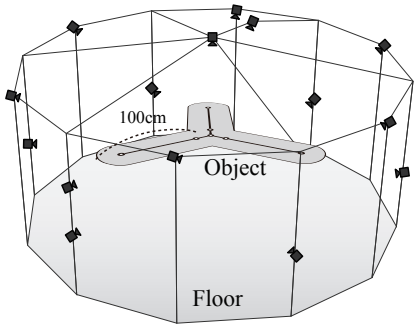

図 5　CG モデルとカメラ配置
Fig. 5　Camera arrangement

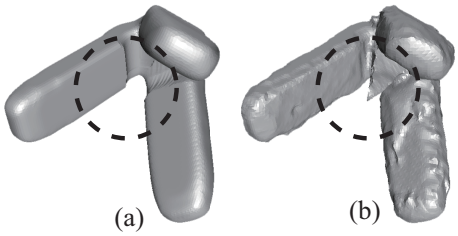

図 7　Photo hull に含まれる phantom volume の例．(a) 図 6 の GT 列，(b) 同図 PH 列 t_6 に示した形状復元結果をそれぞれ斜め下方向からレンダリングした結果．点線内を比較すると，本来の形状 (a) には存在しなかった phantom 形状が (b) に存在していることが確認できる．

Fig. 7　Example of phantom volume in photo hull. (a) and (b) shows 3D renderings of the shapes in the column “GT” (ground truth) and “PH” (estimated 3D shape) of Figure6. The dotted circles indicate that (b) has a phantom volume which is not in the ground truth.

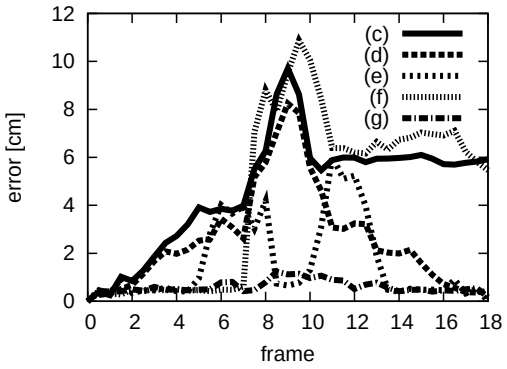

図 8　関節位置の推定誤差
Fig. 8　Estimation error of the node positions

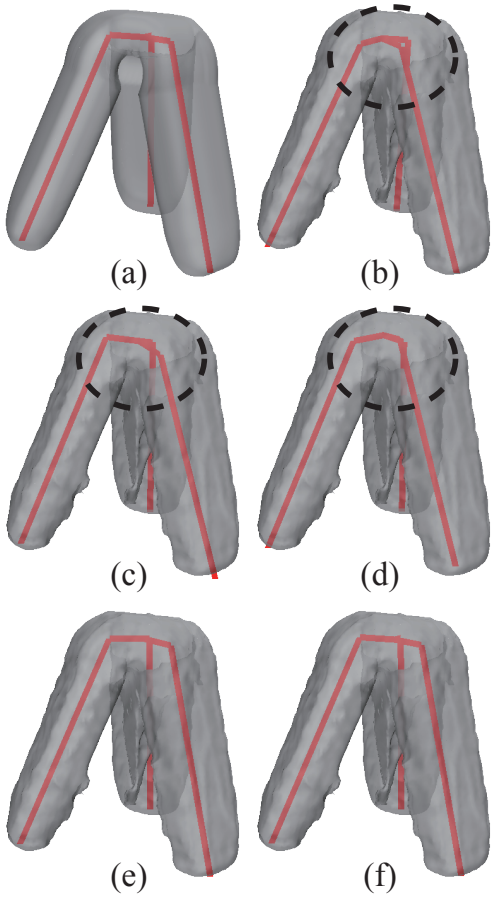

図 9　時刻 t_7 の推定結果（(a) 真値，(b)ICP，(c) 法線を考慮した手法 [2]，(d)ρ_c のみ，(e)ρ_p のみ，(f) 提案手法による推定結果）．推定された骨格姿勢が太線で表示されている．

Fig. 9　Estimation result of t_7. (a) the ground truth, (b) ICP, (c) Ogawara [2], (d) ρ_c only, (e) ρ_p only, (f) the proposed method. Bold lines of each figure illustrate the estimate bone posture.

した．

図 6 および図 8 が姿勢推定結果を表しており，図 6 の各列は左から順に (a) 真の形状とその骨格，(b)photo-hull の形状，(c)ICP によって推定した姿勢，(d) 法線方向による外れ値除去を導入した小川原らの手法 [2] によって推定した姿勢，(e) 本手法で ρ_c のみを用いて推定した姿勢，(f) 本手法で ρ_p のみを用いて推定した姿勢，(g) 本手法で推定した姿勢，を上から下に時系列に沿って表している．図中 (b) の色は photo-consistency の値を表し，青はテクスチャが一致し，赤は一致していないことを，また灰色はカメラから観測されない領

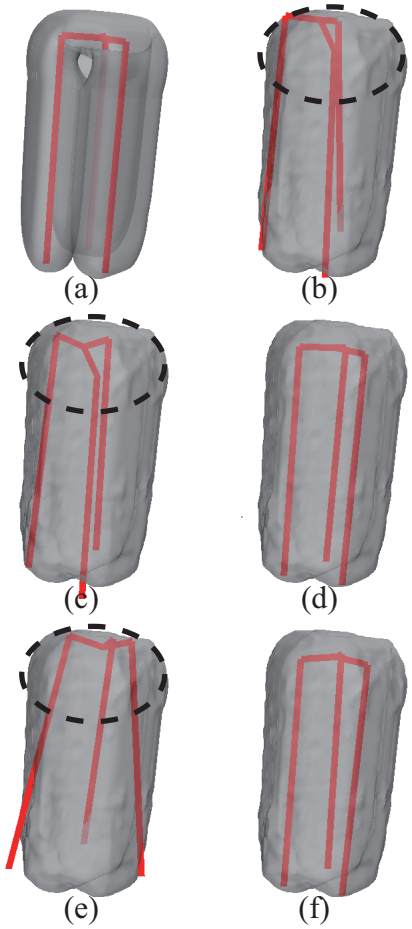

図 10　時刻 t_9 の推定結果（(a) 真値，(b)ICP，(c) 法線を考慮した手法 [2]，(d)ρ_c のみ，(e)ρ_p のみ，(f) 提案手法による推定結果）．推定された骨格姿勢が太線で表示されている．

Fig. 10　Estimation result of t_9. (a) the ground truth, (b) ICP, (c) Ogawara [2], (d) ρ_c only, (e) ρ_p only, (f) the proposed method. Bold lines of each figure illustrate the estimate bone posture.

域を表している．図 11 に示された図 6(b)t_6 の拡大図を見ると，(1) 復元形状の内側が複数台のカメラからは観測不可能で灰色となっていること，(2) 復元形状の内側中央部に元のモデル（図 5）には存在しなかった第 4 の腕のような突起が現れていることが確認できる．この突起は phantom volume によるものであり，どのカメラからも観測不可能であることから原理的に除去はできない． また (d)〜(g) では推定された骨格を太線で (b) の形状に重畳表示している．ここで (d)〜(g) の形状は推定された姿勢に沿ってモデル

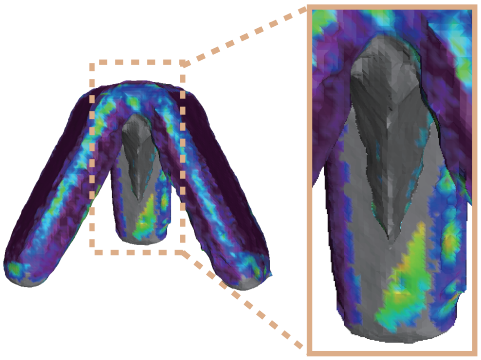

図 11　図 6(b)t_6 の図（左）とその一部の拡大図（右）．色は赤に近いほど photo consistency が低く ρ_p の値が小さいことを，青に近いほど photo consistency が高く ρ_p の値が大きいことを示しており，灰色はカメラから観測できなかった領域を表している．また右の拡大図では，復元形状の内側中央部に元のモデル（図 5）には存在しなかった “第 4 の腕” のような突起が phantom volume として存在していることが確認できる．

Fig. 11　Enlarged image of Figure 6 (b) t_6. Red or yellow regions have low ρ_p values and blue regions have high ρ_p. Gray regions correspond to the surface where no or only one camera can observe. The gray protuberance like “fourth arm” on the center of the body is a “phantom volume” due to self-occlusions.

を変形させたものではないのは，推定された姿勢が本来の形状とずれていることを目視で確認するためである．形状復元には [18] の手法を使用し，メッシュを構成する各 3 角形パッチの大きさは 1cm とした．この手法ではまず visual hull によって対象の大まかな形状を得た後に，カメラから観測可能な領域については photo-consistency に基づいて最適化し，一方でカメラから観測できない領域の形状は visual hull の形状をそのまま踏襲する．このため t_6 などのように対象の 3 本の腕によって自己遮蔽が生じると，図 7 に示すように phantom-volume がそのまま観測形状にも現れ，本来のモデル形状との間で明確な差異となる．図 8 は図 6 の (c)〜(g) に示した推定結果それぞれの関節位置の推定誤差（図 6(a) の骨格との差）を表している．グラフの横軸は時刻 t_0 から t_{18} までを表し，縦軸は推定された関節位置とその真の位置の間の距離の平均値である．これらの結果から，

（1） 時刻 t_6〜t_8，t_{11}〜t_{13} のように観測結果に phantom 領域が含まれる場合，単純な ICP による推定（図 6(c)）や [2] の手法（図 6(d)），ρ_c のみを用いた場合（同図 (e)）では phantom 領域とモ

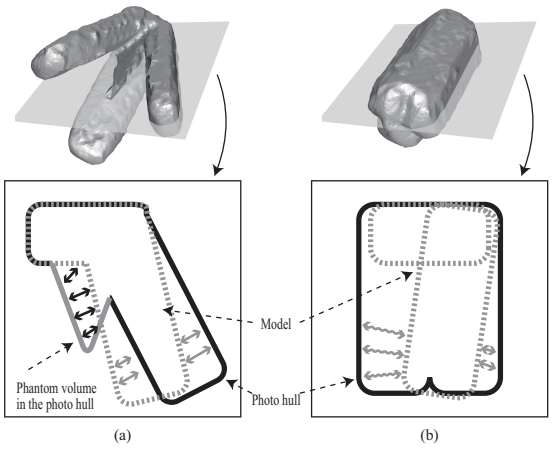

図 12　法線方向による外れ値除去では部位同士の重度の接触に対処できない例．(a), (b) はそれぞれ図 6(d) の t_7 および t_9 の場合について，図上段に示した切断面におけるモデルと観測形状間の対応関係を模式的に図示している．図中の黒太線は観測形状（photo hull）を示しており，灰色太点線はモデル形状ををを示している．また (a) 中の灰色太線は観測形状中に含まれる phantom 領域を示している．(a) のように観測側に phantom 領域が存在した場合，モデル-観測側間には灰色太矢印で示したような正しい対応関係とは別に，黒色太矢印で示したようにモデルと phantom 領域との間でも法線方向が近いために対応関係が定義される．また (b) のように観測側で接触した面に相当する表面形状に相当する部分が欠落した場合，灰色太波矢印で示したように反対側の観測表面形状との間でも法線方向が近いために対応関係が定義されてしまう．

Fig. 12　Limitations of outlier elimination by normal direction. (a) and (b) show the vertical sections of Figure 6 (d) t_7 and t_9 respectively. The top row shows the cutting planes. The black bold lines illustrate the observed 3D shapes (photo hulls) and the gray bold dotted lines illustrate the 3D models. The gray bold lines in (a) illustrate the phantom volume part in the observed 3D shape. In the case that the observation includes phantom volume parts as shown in (a), the matching process produces not only right correspondences as shown by bold grey arrows but also false correspondences as shown by bold black arrows between the phantom volume because the directions of the surface normals can be similar with those of the nearest part of the model. On the other hand, in the case that the collided surfaces are missed in the observation, the matching process produces false correspondences as shown by bold wavy grey arrows because the directions of the surface normals can be similar.

デルの間で誤対応が生じるために姿勢推定精度が悪化する（太線で示した骨格が (a) に示した真の姿勢から大きくずれる）．特に時刻 t_7 の場合について図 9 に図 6 の拡大図を示す．図中 (a)〜(f) はそれぞれ真値，ICP, [2] の手法，ρ_c のみ, ρ_p のみ，提案手法による推定結果に対応しており，点線で囲まれた領域を比較すると (b)ICP の場合，(c) [2] の手法の場合，(d)ρ_c のみの場合に推定位置がずれていることが定性的に確認できる．また図 8 の t_6〜t_8，t_{11}〜t_{13} の区間において (c)ICP, (d) [2] の手法，(e)ρ_c のみの場合の誤差が増大している点から定量的に確認できる．これに対して ρ_p を導入した (f) および提案手法 (g) では，t_6〜t_8 で誤差を抑えることができていることが確認できる．ただし (f) の誤差が t_{11} 以降で現象していない理由は，前時刻の姿勢推定結果を次の姿勢推定の初期値として使用することに起因する誤差の伝搬によるためである．

（2） 時刻 t_9〜t_{10} のようにモデルの体節同士が接触した場合，単純な ICP による推定（図 6(c)）や [2] の手法（図 6(d)），ρ_p のみを用いた場合（同図 (f)）では接触によって消失した表面領域とモデルとの間で誤対応が生じるために姿勢推定精度が悪化する．特に時刻 t_9 の場合について図 10 に図 6 の拡大図を示す．図中 (a)〜(f) はそれぞれ真値，ICP, [2] の手法，ρ_c のみ, ρ_p のみ，提案手法による推定結果に対応しており，点線で囲まれた領域を比較すると (b)ICP の場合と (c) [2] の手法，(e)ρ_p のみの場合に推定位置がずれていることが定性的に確認できる．また図 8 の t_9〜t_{10} の区間において (c)ICP，(d) [2] の手法，(f)ρ_p のみの場合の誤差が増大している点から定量的に確認できる．

（3） 提案手法のように ρ_c と ρ_p を共に用いた場合，図 8 の (g) から，誤差が観測形状の解像度にあたる 3 角形パッチの大きさ 1cm 程度に収まっており，すべての時刻において妥当な精度で姿勢が推定できていると定量的に確認できる．これに対して法線方向による外れ値除去を導入した小川原らの手法 [2] による姿勢推定は，ICP による手法と同程度の性能となっており，本論文が想定する重度の接触が生じる状況においては有効に機能していない．これは phantom 領域との間でも法線

方向が一致してしまったり，そもそも接触した面が消失した場合は，法線の近い別の面との間で偽の対応関係を定義してしまうためである．この状況は図 12 によって説明できる．まず同図 (a) は図 6(d) の t_7 を，(b) は図 6(d) の t_9 をそれぞれ縦に切断して得られる断面を模式的に表している．図中の黒太線は観測形状（photo hull）を示しており，灰色太点線はモデル形状ををを示している．また (a) 中の灰色太線は観測形状中に含まれる phantom 領域を示している．(a) のように観測側に phantom 領域が存在した場合，モデル-観測形状間には灰色太矢印で示したような正しい対応関係とは別に，黒色太矢印で示したようにモデルと phantom 領域との間でも法線方向が近いために対応関係が定義される．また (b) のように観測側で接触した面に相当する表面形状に相当する部分が欠落した場合，灰色太波矢印で示したように反対側の観測表面形状との間でも法線方向が近いために対応関係が定義されてしまう．つまりたとえ法線方向を考慮したとしても “観測側からモデル側” への対応関係を求めてその距離を最小化しようとすると同図 (a) のように phantom 領域に対処できず，“モデル側から観測側” への対応関係を求めてその距離を最小化しようとすると同図 (b) のように接触によって観測側で生じた欠損に対処できない．これは撮影環境によって生じる観測形状側の形状変化を考慮していないことによる限界であるが，これに対して提案手法はこれを明示的にモデル化することによってより頑健な姿勢推定手法を実現できた．

ということができ，実験によって本論文で提案した ρ_c と ρ_p の有効性が定性的，定量的に示された．なお 1 フレーム分の姿勢推定に要した時間は Intel Core2 Duo 3.0GHz において平均約 5 分であった．またこの実験では入力が t_9 を中心として前後半で概ね同じ形状であるが，グラフにおいては (c)ICP の場合と (f)ρ_p の場合に t_{10} 以降で推定誤差が対応する前半時刻と対称になっていない．これはアルゴリズムの設計として前の時刻における推定結果を，次の推定の初期値として使用するために誤差が伝搬することに起因しており，これら 2 つが誤差を含む初期値からの復帰に失敗したことを示している．

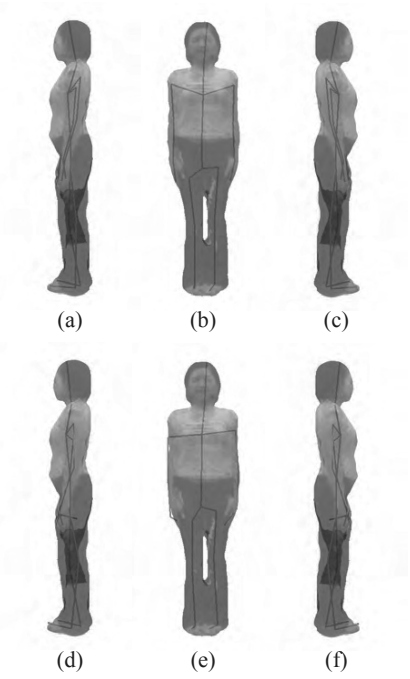

図 14　実画像による本手法と ICP との比較．(a)〜(c)：本手法による推定姿勢をそれぞれ右，正面，左から表示した図．(d)〜(f)：ICP による推定姿勢をそれぞれ右，正面，左から表示した図．各図において，推定姿勢は太線で示されており，これと 3 次元形状復元結果にテクスチャを貼ったものが重畳表示されている．

Fig. 14　Comparision between the proposed method and ICP. (a),(b) and (c) show the results of the proposed method by rendering from left, front and right side respectively. (a),(b) and (c) show the results of ICP by rendering from left, front and right side respectively. Bold lines of each figure illustrate the estimate 3D bone posture.

5.2　実画像による評価

次に本手法を実観測データに適用した結果を図 13 に示す．本実験においても前節と同様のカメラ配置を使用し，各カメラは Sony XCD-X710CR（解像度 XGA, フレームレート 25fps，シャッタースピード 1ms で同期撮影）を使用した．この実験では $N_p = 23$，つまり 23 本の骨を持ち，$3N_p + 3 = 72$ 自由度を持つ骨格モデルを当てはめて姿勢推定を行った． また式 (2) の α_c および τ_c，式 (3) の α_p および τ_p の各パラメータについては，前述の CG モデルを用いた実験と同様に順に 2.0, 2.5, 5.0, 0.95 とした．これらのパラメータが CG でのものと共通である理由は，CG モデルのスケールを実画像に合わせ，またカメラ配置も同じも

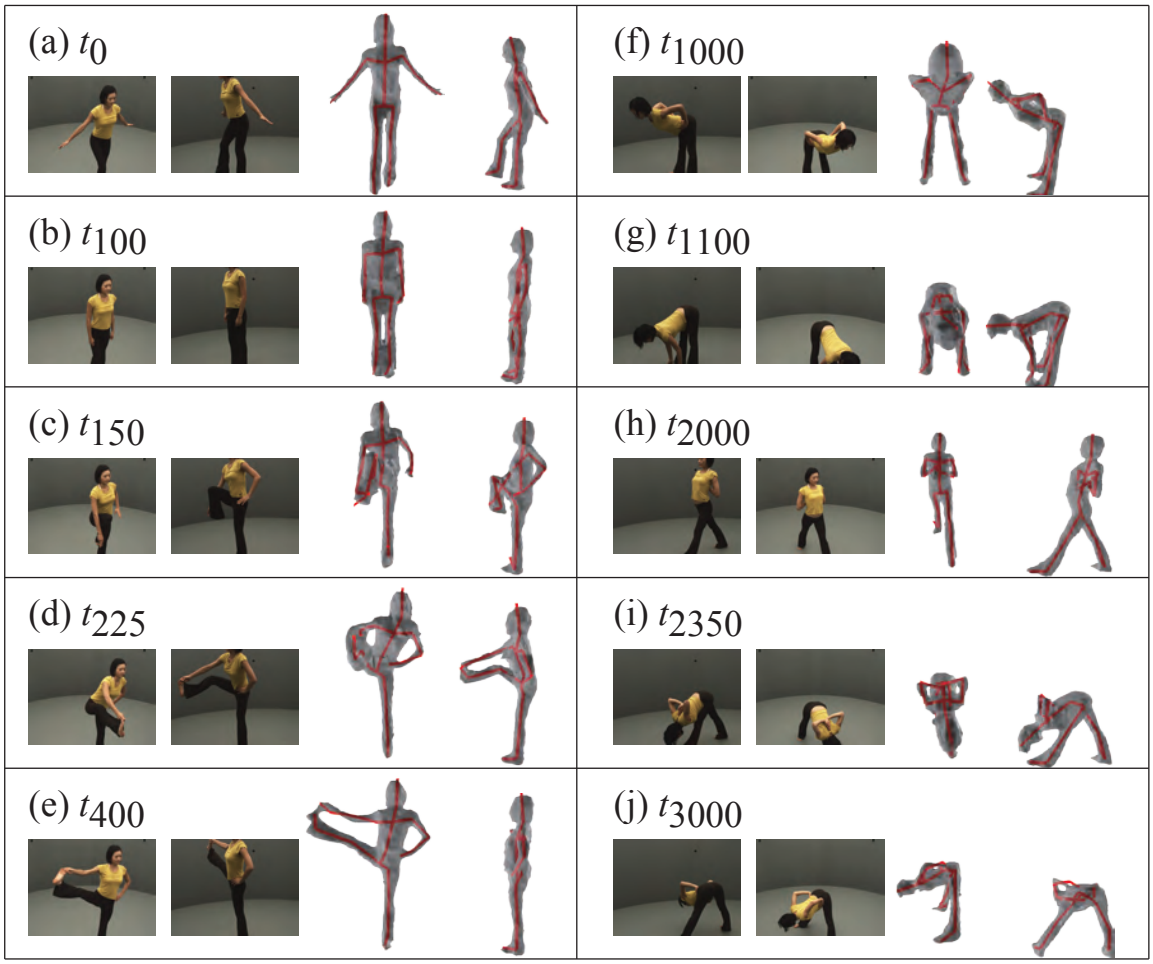

図 13　実画像による評価
Fig. 13　Estimation result using real data

のを用いているためである．

図 13(a)〜(j) は全体として 3000 フレーム，2 分間のヨガを行う人物のシーケンスに対して姿勢推定を行った結果のうち，特徴的なフレームについて入力画像と，(a) の t_0 を初期姿勢として姿勢推定を行った結果を表している．なお姿勢推定結果は灰色で示された観測形状の上に赤色の 太線で重畳表示されている．この結果から，観測形状中では体節同士の接触が様々な箇所で起きているような形状に対しても，本手法によって姿勢推定が可能となっていることがわかる．この実験においても前節の CG の場合と同様に 1 フレームあたり約 5 分の計算時間を要した．

また図 14 に，ICP と比較した図を示す．図 (a)〜(c) は本手法による推定姿勢をそれぞれ右，正面，左からレンダリングした結果であり，図 (d)〜(f) は ICP による推定姿勢をそれぞれ右，正面，左から表示した図である．各図において，推定姿勢は太線で示されており，これと 3 次元形状復元結果にテクスチャを貼ったものが重畳表示されている．なおここで表示している 3 次元形状は多視点映像からこのフレームの形状を復元したものであり，モデルを推定姿勢によって変形させたものではない．これは本来の形状と骨格のずれを目視によって確認するためである．また図 14 中で特に ICP と提案手法の差が顕著である部分の拡大図を図 15 に示す．図中 (a), (c) はそれぞれ図 14(b) および (e) の肩周辺の拡大図であり，(b), (d) はそれぞれ図 14(c) および (f) の腰周辺の拡大図である．

これらの図から，腰や脇のように体節間の接触が起

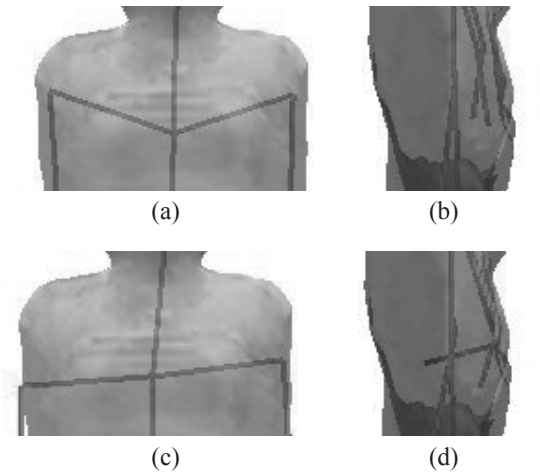

図 15　実画像による本手法と ICP との比較（拡大図）．(a), (c) はそれぞれ図 14(b) および (e) の肩周辺の拡大図であり，(b), (d) はそれぞれ図 14(c) および (f) の掌周辺の拡大図である．各図において，推定姿勢は太線で示されており，これと 3 次元形状復元結果にテクスチャを貼ったものが重畳表示されている．

Fig. 15　Comparision between the proposed method and ICP. (a) and (c) show the “shoulder” area of Figure 14(b) and (e) respectively. (b) and (d) show the “hand” area of Figure 14(b) and (e) respectively. Bold lines of each figure illustrate the estimate 3D bone posture.

きている箇所において，ICP では正しく姿勢が求まっていないことが確認できる．例えば，図 15(a) と (c) を比較すると，肩の位置が ICP で求められた結果では本来の位置から大きくずれており，また図 15(b) と (d) を比較すると，腰に添えられた手の位置が，ICP による結果では大きくずれていることが確認できる．

この結果から，体節間の接触を伴う複雑な運動に関して，本手法が ICP と比較してより頑健に姿勢推定を行うことができることを定性的に確認することができた．

6. 結　　論

以上，本論文では 3 次元形状を用いた姿勢推定問題に対して，(1) 接触によって原理上観測不可能であり，モデル形状中でマッチングに用いるべきではない領域と，(2) カメラ配置と多視点画像からの形状復元の特性として形状の信頼性が低く，観測形状中でマッチングに用いるべきではない領域の，２つのモデル化を提案した．また実際にこのモデルを用いて CG データに対する姿勢推定を行い，2 つのモデルがそれぞれ姿勢推定精度の向上に寄与していることを定量的に示すとともに，実画像に対しても本手法が ICP と比較してより頑健に姿勢推定を行うことができることを定性的に示した．

この中で式 (2) で定義された可観測性 ρ_c のパラメータ α_c および τ_c は，式 (2) の後段で述べたようにカメラが対象に対して等方的に分布していることを仮定している．これに対して，頂点の位置，モデルの現在の姿勢，カメラの配置を考慮すれば，人手によって迫っていされたパラメータ α_c および τ_c を用いずとも自動的に可観測性を決定できると考えられる．そこで 今後はカメラ配置を反映しつつ計算コストを現実的な範囲に抑えた新たなモデル化の検討を行う予定である．

また今回我々は 3 次元形状復元の信頼度として ρ_p を導入したが，形状復元に用いるアルゴリズムによってはカメラから観測可能であるにも関わらず 3 次元形状として復元がなされず部位の欠損が起こりうる．このような形状復元プロセスの特性をより深く踏まえた信頼度モデルについても，今後検討を行う予定である．

謝　　辞

本研究の一部は，文部科学省グローバル COE プログラム「知識循環社会のための情報学教育研究拠点」，文部科学省「大型有形・無形文化財の高精度デジタル化ソフトウェアの開発」プロジェクト，JST CREST「映画制作を支援する複合現実型可視化技術」プロジェクトの補助を受けて行った．

文　　献

[1] M.D. Wheeler, Y. Sato, and K. Ikeuchi, “Consensus surfaces for modeling 3d objects from multiple range images,” Proc. of ICCV, p.917, IEEE Computer Society, 1998.

[2] 小川原光一，李　暁路，池内克史，“関節構造を持つ柔軟変形モデルを用いた人体運動の推定,” Proc. of MIRU2006, pp.994–999，2006.

[3] L. Mundermann, S. Corazza, and T.P. Andriacchi, “Accurately measuring human movement using articulated icp with soft-joint constraints and a repository of articulated models,” Proc. of CVPR, pp.1–6, June 2007.

[4] R. Plänkers and P. Fua, “Articulated soft objects for multiview shape and motion capture,” PAMI, vol.25, no.9, pp.1182–1187, 2003.

[5] J. Starck and A. Hilton, “Spherical matching for temporal correspondence of non-rigid surfaces,” Proc. of ICCV, pp.1387–1394Vol.2, Oct. 2005.

[6] R. Kehl, M. Bray, and L. Van Gool, “Full body tracking from multiple views using stochastic sampling,”

Proc. of CVPR, vol.2, pp.129–136vol.2, June 2005.

[7] G. Shakhnarovich, P. Viola, and T. Darrell, “Fast pose estimation with parameter-sensitive hashing,” Proc. of ICCV, pp.750–757vol.2, Oct. 2003.

[8] K. Grauman, G. Shakhnarovich, and T. Darrell, “Inferring 3d structure with a statistical image-based shape model,” Proc. of ICCV, pp.641–647vol.1, Oct. 2003.

[9] B. Rosenhahn and T. Brox, “Scaled motion dynamics for markerless motion capture,” Proc. of CVPR, pp.1–8, June 2007.

[10] B. Rosenhahn, C. Schmaltz, T. Brox, J. Weickert, D. Cremers, and H.-P. Seidel, “Markerless motion capture of man-machine interaction,” Proc. of CVPR, pp.1–8, June 2008.

[11] A.O. Balan, L. Sigal, M.J. Black, J.E. Davis, and H.W. Haussecker, “Detailed human shape and pose from images,” Proc. of CVPR, pp.1–8, June 2007.

[12] A.O. Balan, M.J. Black, H. Haussecker, and L. Sigal, “Shining a light on human pose: On shadows, shading and the estimation of pose and shape,” Proc. of ICCV, pp.1–8, Oct. 2007.

[13] M.A. Brubaker and D.J. Fleet, “The kneed walker for human pose tracking,” Proc. of CVPR, pp.1–8, June 2008.

[14] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, “Progressive search space reduction for human pose estimation,” Proc. of CVPR, pp.1–8, June 2008.

[15] A. Gupta, T. Chen, F. Chen, D. Kimber, and L.S. Davis, “Context and observation driven latent variable model for human pose estimation,” Proc. of CVPR, pp.1–8, June 2008.

[16] T.B. Moeslund, A. Hilton, and V. Krüger, “A survey of advances in vision-based human motion capture and analysis,” CVIU, vol.104, no.2, pp.90–126, 2006.

[17] T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara, “Real-time 3d shape reconstruction, dynamic 3d mesh deformation and high fidelity visualization for 3d video,” CVIU, vol.96, pp.393–434, Dec. 2004.

[18] G. Vogiatzis, P.H.S. Torr, and R. Cipolla, “Multi-view stereo via volumetric graph-cuts,” CVPR, pp.391–398, 2005.

[19] J. Starck and A. Hilton, “Surface capture for performance based animation,” IEEE Computer Graphics and Applications, vol.27(3), pp.21–31, 2007.

[20] J. Starck, A. Hilton, and G. Miller, “Volumetric stereo with silhouette and feature constraints,” Proc. of BMVC, p.III:1189, 2006.

[21] R. Woodland, “隙間を埋める – ステッチングとスキニングを用いた高度なアニメーション,” Game Programming Gems, M. DeLoura, 川西裕幸, 狩野智英（編）, 第 4.15 章, pp.458–465, ボーンデジタル, 2001.

Abstract While many methods have been proposed to estimate 3D human motion, most of them do not work well for complex body actions such as Yoga dance. Major reasons for this are (1) contacts between body parts which produce colided and unobservable regions where any 3D shape estimation algorithms cannot obtain 3D shapes by definition, and (2) unreliableness of 3D shape estimated using multi-viewpoint images due to self-occlusions or limitations on texture-matching between different viewpoints. These incompletenesses break conventional methods such as ICP-based 3D motion estimation methods since they expect observed surface is equal to the model surface. We solve this problem by introducing a novel modelling of these incompletenesses into our 3D surface-to-surface fitting algorithm. Some experiments demonstrate the advantage of our robust 3D motion estimation method.

Key words 3D motion estimation, 3D video, multi-viewpoint images, ICP