2006－CVIM－152（5）
2006／1／19

実時間ビジョンベースモーションキャプチャシステムのパラメータ自動推定と精度評価

荒木 裕　有田 大作　谷口 倫一郎

九州大学大学院システム情報科学府・研究院
〒816-8580 福岡県春日市春日公園 6-1
E-mail:{araki, arita, rin}@limu.is.kyushu-u.ac.jp

概要　人間の動作や行動を実時間で計測する技術は，3 次元アニメーションやビデオゲームにおける CG キャラクタの行動生成，人間と機械との仮想空間上でのインタラクション，さらには人間型ロボットの遠隔操縦など，多くのアプリケーションへの利用が期待できる．現在，人間の動作情報を獲得するために，特別なマーカやセンサといったモーションキャプチャデバイスを装着する手法が実用化されているが，それらはユーザに身体的な制約を課し不快感を与えることもある．そこで，我々は，身体的制約の無い，画像解析によるモーションキャプチャシステムを開発した．本稿では，このシステムで必要となる被写体の身体的事前知識をビジョンベースで獲得する手法を提案し，姿勢推定の精度評価を行った．

Automatic Parameter Estimation and Accuracy Evaluation of Vision-based Real-time Human Motion Capture System

Yutaka Araki, Daisaku Arita, Rin-ichiro Taniguchi

Department of Intelligent system, University of Kyushu
6-1, Kasuga-koen, Kasuga, Fukuoka 816-8580 Japan
E-mail:{araki, arita, rin}@limu.is.kyushu-u.ac.jp

abstract　In this paper, we discuss a vision-based real-time human motion capture system. Vision-based motion capture has a merit that it does not impose any physical restrictions on humans, which provides a natural way of measuring human motion. This system is useful for action generation of CG character in 3-dimensional animation or video games, man-machine interaction in a virtual space, remote control of humanoid robots, etc. Then, we have developed a vision-based real-time human motion capture system. This paper proposes a automatic mesurement of human body parameter and show the result of accuracy evaluation experiment for the system.

1. はじめに

バーチャルリアリティシステムやビデオゲーム等の様々なインタラクティブシステムにとって，人間と機械のシームレスな三次元インタラクションは重要な技術である．そのようなインタラクションシステムの実現のためには，実時間で人体の動作を計測するシステムが必要とされる．現在，人間の動作情報を獲得するために，沢山の特別なマーカやセンサを人体に取り付ける手法があるが，これらのデバイスは計測対象に身体的な制約を課し，ユーザに不快感を与えることもある．近年では，そのような身体的制約の無い画像特徴を利用したビジョンベースのモーションキャプチャシステムが研究されている．ビジョンベースのアプローチにはまだ解決すべき問題もあるが，人間と機械のシームレスなインタラクションを実現するための非常にスマートなアプローチといえる．ただし，実用的なインタラクティブシステム開発のためには，システムの実時間性が極めて重要である．従って，モデルの精度を重視し，オフラインでの適用を前提とした計算量の多い手法[1),2)]を適用するのは難しい．モデルの自由度をある程度犠牲にした上で，実時間性を確保するのが現状では妥当であるというのが，筆者らの立場である．

ビジョンベースのアプローチでは，人間の動作解析のための手掛りとしてどうのような画像特徴を用いるかが重要である．一般には，ブロッブ (密な領域)[3)] やシルエットの輪郭線[4)] といった画像特徴がよく利用されるが，安定に得られる画像特徴は限られている．

これまで我々は，身体モデルを用いた多視点画像解析によるモーションキャプチャシステムを構築してきた[5)]．このシステムではブロッブとシルエット輪郭から人体の

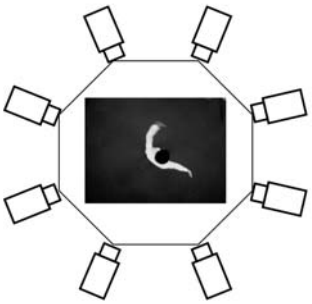


図 1 カメラ配置

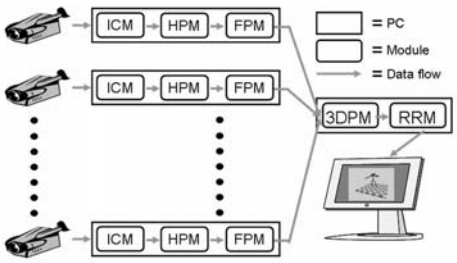


図 2 PC クラスタにおける処理モジュールの配置

三次元姿勢を推定している．また，人間が様々なポーズをとる際に起こる自己隠蔽に対処するために，多視点画像解析というアプローチをとっている．さらに，莫大な量の情報を獲得し処理する多視点画像処理の実時間処理を可能にするため，PC クラスタ[6] を利用している．

今回新たに，姿勢推定の際に用いられるユーザの身体パラメータを撮影画像から取得するための方法を考案し，システムの利便性向上を図った．また，システムの姿勢推定方法の精度評価実験を行った．本稿では，これらの点について報告する．

2. システムの概要

2.1 PC クラスタによる実時間多視点動画像処理

実世界において認識機構などの知的機能を備えた撮像系である知能視覚システムが有効に機能するためには実時間性・オンライン性・ロバスト性・柔軟性といった基本特性が必要とされる．マルチセンサーフージョンのように多様なセンサ情報融合を実現するためのフレームワークを用いることで，ロバストで柔軟なシステムが実現できると期待されている．特に，多視点動画像処理はもっとも有効な実現方法のひとつであり，複数のカメラにより獲得された冗長な情報を互いに参照することで，単一のカメラのみによる観測で起こり得る隠蔽などの様々な問題に対処することが可能となる．しかし，この場合，実時間性やオンライン性の実現が困難になっていることが多い．これらの問題点を解決し，実時間多視点動画像処理を実現するには，高性能な分散システムが必要となる．我々は，そのような高性能分散システムとして，高速ネットワークにより接続された複数の PC から構成される PC クラスタを用い，実時間処理のための同期機構やエラー処理機構提供するプログラミング環境 RPV を開発し[6]，その上に実時間モーションキャプチャシステムを実装した．RPV によって，システム内の各処理モジュールは PC 内のスレッドに割り当てられ，同期機構により並列に動作する．

2.2 カメラの配置

本システムでは，人物の全身が撮影可能なカメラを人物を囲むように円周上に配置して利用する．なお，今回の実験では 8 台のカメラを利用した (図 1)．これらのカメラは固定されており，計測空間中に定義したワールド座標系とカメラ座標系間であらかじめキャリブレーションが行われているものとする[7]．

一般に多眼視によって 3 次元位置の推定を行うには，少なくとも 2 視点から同じ観測点が求められれば，三角測量の原理により一意に推定可能である．しかし，2 視点しか用いない場合，どちらか一方の視点で隠蔽が起こっただけでその観測点の 3 次元位置推定が失敗してしまうため，本手法では多くの視点から得られる情報を統合処理することにより，隠蔽などの影響を受けにくい 3 次元位置推定を行う．

2.3 処理モジュール

図 2 は PC クラスタの各 PC 上に本システムの処理モジュールを並列パイプライン配置したものである．図 2 から分かるように，本システムではカメラ台数より 1 台多い PC から成る PC クラスタを利用する．そして，各 PC では以下に挙げるモジュールが動作している．

画像獲得モジュール (ICM)
人物全身の画像の取り込みを行なう．このモジュールは視点と同じ数だけ存在し，各 ICM は同期信号よって同時刻に撮影されたカメラ画像を獲得する．獲得された画像は直ちに HPM に送信される．

全身画像処理モジュール (HPM)
ICM から送られてきた画像から特徴点の抽出を行なう．処理結果は FPM へ送信される．また，人物シルエット画像が 3DPM に送信される．

顔画像処理モジュール (FPM)
抽出された特徴点の中から顔領域を選び出し，テンプレートマッチングにより顔の向きを推定する．処理結果は他の特徴点の情報とともに 3DPM へ送信される．

3 次元復元モジュール (3DPM)
HPM と FPM から受け取った情報を統合処理することにより，人物の 3 次元姿勢推定を行なう．推定結果は RRM へ送信される．

実時間レンダリングモジュール (RRM)
3DPM から受け取ったデータをもとに，仮想空間上に実時間で CG キャラクターを描画する．

以下，システムの性能評価実験について理解するため

に必要な HPM と 3DPM の処理内容について，それぞれ 3 と 4 で詳しく述べる．そののち，5 で身体パラメータの取得方法について，6 で性能評価実験について述べる．

3. 全身画像からの特徴点抽出

センサーなどを用いない非接触での人物動作解析には，人体領域に相当する画像特徴として人物の肌色領域や輪郭線などの領域特徴が一般的に利用される．本システムでは，手先や足先の追跡が比較的安定に行える肌色領域と，人体の輪郭線情報の二つの異なる特徴量を同時に利用する．肌色領域追跡による計測ベースと輪郭線を用いた認識ベースの二つの手法を用いる（具体的には 4 で述べる）ことで，エラー処理を含む安定した人体の姿勢推定が実現される．

以下，HPM における画像特徴点の抽出の詳細について述べる．

3.1 肌色領域によるブロッブの 2 次元位置推定

画像上に観測される肌色領域を顔・両手・両足のブロッブとする．肌色領域の追跡により，これら 5 点の 2 次元カメラ座標位置を推定する．

肌色ブロッブ追跡の前処理として，カメラから獲得した画像に対し，

背景差分およびオープニング処理をおこなうことにより人物領域を抽出する．次に，抽出された人物領域に対して肌色画素の検出を行う．肌色画素の検出には輝度変化の影響を比較的受けにくい色相を利用する．また，予め RGB すべての画素の組み合わせにおいて肌色尤度を計算した表を用意しておき，実行時にはその表を参照することにより高速に肌色画素を識別する．

次に，検出した肌色画素に対して，処理が高速な射影探索法[8] を用いてクラスタリングを行い肌色ブロッブを検出し，各肌色ブロッブの重心を求める．具体的には，画像の縦軸横軸に対して肌色画素のヒストグラムを計算し，縦軸横軸ともにその値が閾値を超えた領域を肌色領域候補とする．肌色領域候補例を図 3 に示す．この場合，実際の肌色ブロッブは 5 つであるが，候補領域として 25 個の領域が検出されている．これらの候補領域に対し，内部にある肌色画素の総数を求め，その値が閾値以上の領域の肌色画素を肌色ブロッブとし，その重心座標を画像上の特徴点とする．特徴点として検出された肌色ブロッブの重心を図 4 に示す．検出された肌色領域は FPM と 3DPM に送信される．

3.2 輪郭線を用いたブロッブの 2 次元位置推定

本システムでは，3.1 節の処理と平行して輪郭線を用いた特徴点の抽出処理を行う．画像上の人物の輪郭線情報を検出し，予め登録しておいた輪郭線データとのマッチングを行うことにより，顔・両手・両足のカメラ座標位置を検出する．

本システムでは画像中の人物は 1 人であると仮定しているため，シルエットの境界線は，人物のシルエット像

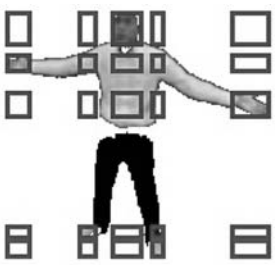


図 3 肌色領域候補

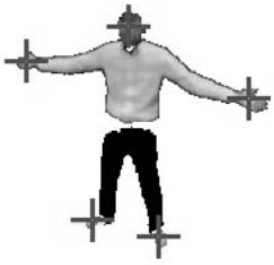


図 4 肌色領域重心

図 5 両手を真横に広げた直立姿勢

の重心から真上に向かって探索することによって検出される．その境界線の一点を開始点とし，時計回りに輪郭線を検出する．

シルエット重心の座標を (x_g, y_g)，n 点検出した輪郭線上の点のうち i 番目にあたる点の座標を (x_i, y_i) $(i = 1, \cdots, n)$ とする．このとき，輪郭線情報 $dx[j], dy[j]$ $(j = 1, \cdots, k)$ を

$$\begin{cases} dx_{tmp}[i] = x_i - x_g \\ dy_{tmp}[i] = y_i - y_g \\ dx[j] = dx_{tmp}[int((n \times j)/k)] \\ dy[j] = dy_{tmp}[int((n \times j)/k)] \end{cases} \quad (1)$$

と定義する[4]．データ列 $dx_{tmp}[i], dy_{tmp}[i]$ における要素数 n はフレーム毎に変化する．そこで，要素数が k 個になるようにサンプリングを施した $dx[j], dy[j]$ を輪郭線特徴量として用いる．本システムでは，カメラから遠ざかったときに平均的に検出される輪郭点数である $k = 200$ とした．また，認識対象の登録姿勢として，“両手を真横に広げた直立姿勢” を用いた（図 5）．登録姿勢の輪郭を表すデータ列 $dx_{reg}[j], dy_{reg}[j]$ $(j = 1, \cdots, k)$ と，各フレームで検出される輪郭線 $dx[j], dy[j]$ に対して，以下の式が同時に成立する場合のみ，入力画像が登録姿勢と同じ姿勢であると認識する．

$$\begin{cases} \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} \mid dx[j] - dx_{reg}[j] \mid < dx_{th} \\ \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} \mid dy[j] - dy_{reg}[j] \mid < dy_{th} \end{cases} \quad (2)$$

ただし，dx_{th}，dy_{th} は経験的に決定した値であり，後述

の実験ではそれぞれ 0.04，0.08 としている．

入力画像が登録姿勢であると判定された場合，画像から顔・両手・両足の座標を検出する．検出には，重心からの距離が局所的に最大となる輪郭部を顔・両手・両足に対応させる．図 6 に登録姿勢における輪郭線 $dx_{reg}[j], dy_{reg}[j]$ の値，及び推定された各部位の位置を示す．

4. 人体の 3 次元姿勢推定

肌色追跡，及び輪郭線情報によって計測された顔・両手・両足の各カメラでの座標位置とテンプレートマッチングにより検出された顔の向きのデータが，それぞれ HPM，FPM から 3DPM へと送信される．3DPM では，これらの情報を統合処理することで，人物の様々な姿勢を推定する．3DPM で行われる処理の概要を以下に示す．

1. 各カメラで顔向き推定の結果を統合し，全カメラによる顔向き推定を行う．
2. 肌色追跡によって計測された顔・両手・両足の 3 次元位置を推定する（計測ベースの推定）
3. 輪郭線により計測された顔・両手・両足の 3 次元位置を推定する（認識ベースの推定）
4. 2.3. の結果からどちらの値を使用するか選択を行う（計測と認識の融合によるエラー回復処理）
5. 顔・両手・両足の 3 次元位置と顔向き情報をもとに，全身の姿勢を推定する

4.1 各カメラからの顔向き情報の統合

カメラから撮影した画像中の肌色領域と身体の肌色部位との対応付けを行うことは困難である．そこで FPM では、画像中の肌色領域全てにおいて、テンプレートマッチングを適用し、各領域においての肌色重心と選択されたテンプレート画像との距離を求める．そして 3DPM で、前フレームでの頭部の 3 次元位置をカメラ画像に投影した座標に最も近い肌色領域を顔領域として選択する．さらに、このようにして求められた各カメラからの顔の向きがのうち最もテンプレート画像との距離が近いものを全カメラによる顔の向きとして採用する．

4.2 計測ベースによる肌色領域の 3 次元位置推定

各カメラ画像から肌色追跡によって得られた肌色領域重心より，肌色領域の 3 次元位置を推定する．2 次元画像中の 1 点は撮影中心を通る 3 次元空間中の直線に対応付けられる．2 視点以上のカメラで同一の点が見えている場合，各視点からの拘束線の交点をとることでその点の 3 次元位置を推定することができる．しかし，実際には各視点における肌色重心に対応する直線どうしが交わることはほとんどない．さらに，同じ肌色領域を表す肌色重心の視点間での対応を取ることも容易ではない．そこで視点間での肌色重心の対応とりつつ，安定に 3 次元位置を計測するアルゴリズムを開発した．

1. 全ての視点における肌色領域重心を 3 次元空間中の直線に変換する．直線への変換には Tsai カメラモデル[7] を用いる．
2. 視点の異なる全ての直線の組について交差判定を行う．その際，直線間の距離を交差点判定の評価値として用いる．評価値が閾値以下なら，選択した 2 点は同一の肌色領域重心とみなし，2 直線が最も近づく点を肌色領域の 3 次元位置とする．
3. 求まった各 3 次元位置をすべての視点からの人物のシルエット画像に逆投影する．このとき，1 視点でもシルエット画像中に存在しない点は肌色領域ではないとみなし削除する．
4. 残った 3 次元空間中の肌色領域を示す各点を，前フレームでの特徴点 (顔，左手，右手，左足，右足のいずれか) にクラスタリングし，各クラスタの密な点を現フレームにおける特徴点の 3 次元位置として決定する．

図 7(a) の入力姿勢に対して得られた肌色領域の 3 次元分布を図 7(b) に示す．最終的に，各クラスタの中心部が特徴点として抽出される．

4.3 認識ベースによる肌色領域の 3 次元位置推定

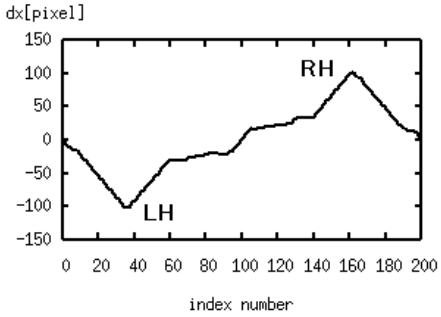
各カメラ画像から輪郭線情報によって得られた特徴点から，肌色領域の 3 次元位置を推定する．まず，顔の向きの情報をもとに，人物の正面に近いカメラを 2 台選択する．選択された 2 台のカメラ画像の両方に対して，そのときの姿勢が登録姿勢であると判定された場合，三角測量の原理により頭・左手・右手・左足・右足の 3 次元位置を推定する．この場合は，各カメラから得られる特徴点と身体の各部位との対応は分かっているので，各部位の 3 次元位置は一意に決定される．

4.4 計測と認識の融合によるエラー復帰処理

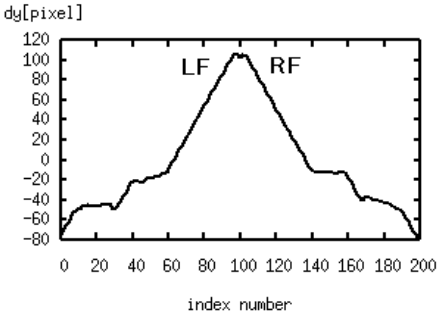
モーションキャプチャをインタラクションのツールとして考えた場合，システムがエラーを起こしたときに，リセットすることなくオンラインでエラーからの復帰を行えることが重要である．肌色追跡をもとにした推定手法は，ブロッブの探索範囲や分類基準に過去のデータを用いるため，ある時点で推定を誤るとそれ以降は正しい推定が行われない．一方，輪郭線情報をもとにした手法は，認識できる姿勢は登録姿勢に近い姿勢に限られるものの，過去のデータを手がかりにしないため精度は常に一定である．本システムでは，これらの 2 つの手法を融合し，入力姿勢と登録姿勢の類似度が大きければ，輪郭線情報をもとにした推定結果を適用し，それ以外の場合は肌色追跡をもとにした手法を適用することでエラーからの復帰を実現し，極端な精度の低下を防いでいる．なお，初期姿勢として登録姿勢をとることによって，次フレーム以降の処理が可能となる．

4.5 全身の姿勢推定

前節までの処理では，人物の頭部・両手・両足の 3 次元位置および顔の向きが得られるが，本システムで用いている人体モデル（図 8）でも分かるように，人体の全身姿勢を推定するためには，さらに肩・肘・腰・膝など



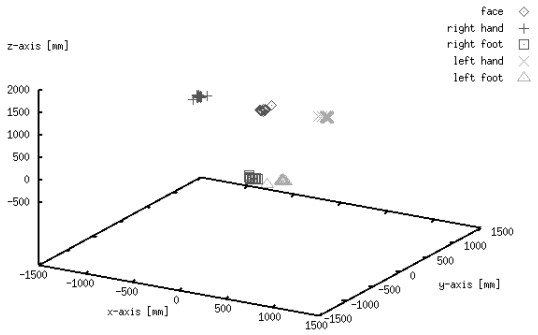
(a)$dx_{reg}[j]$



(b)$dy_{reg}[j]$

図 6　輪郭線情報インデックス

(a)



(b)

図 7　肌色領域の 3 次元位置：(a) 入力画像，(b) 肌色領域分布

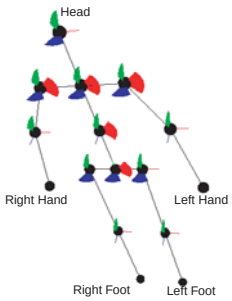


図 8　身体モデル

の 3 次元位置を求める必要がある．

そこで，すでに得られている部位の位置から全身の姿勢情報を得るために，ロボティックスやコンピュータグラフィックスの分野で一般に用いられる逆運動学を用いた数値解と人物シルエットを用いた画像処理を同時に利用するアルゴリズムを開発した．以下では，このアルゴリズムについて上半身と下半身に分けて説明する．

4.5.1　上半身の姿勢推定

上半身の姿勢推定は，下記の 2 段階の処理によって行う．

1.　逆運動学解法による 1 次推定

おおよその全身姿勢を推定するために，計測された肌色ブロッブの位置条件を満たす姿勢を，逆運動学解法を用いて計算する．解法には，計算量の面から実時間性を考慮し CCD 法[9] を用いる．まず，得られた肌色ブロッブの 3 次元位置情報から頭部位置を用いて，首元の位置を求める．このとき，頭部から相対的にみた首元の位置は変化しないとして，頭部位置に固定値を加えることで首元の位置を求める．この首元を固定点として終端点の手先までの関節に対して CCD 法を適用し，この結果を 1 次推定とする．この際，初期姿勢は前フレームの姿勢を用い，手先の肌色ブロッブと人体モデルの手先の位置との距離を最小にするよう繰り返し計算をする．精度のよい計測はできないが，滑らかな動作が生成される．なお，CCD 法での人体モデルの各自由度は，首元が 2，肩が 3，肘が 1 の計 6 である．

2.　逆投影探索による 2 次推定

この手法の基本的なアイディアは，肘の存在範囲は，手先と肩の 3 次元位置によってある円周上（この円を探索円と呼ぶことにする）に限定されることを利用するものである．以下，左肘を推定する場合を例として説明する．

まず，肘推定のための座標系を図 9 のように定義する．探索円の中心から探索円周上の点へのベクトルのうち，首元から左肩へのベクトルとの xy 成分 (ワールド座標系は

(a) (b) (c) (d) (e)

図 10 身体パラメータ取得のための基本姿勢 (a, b, c, d, e)

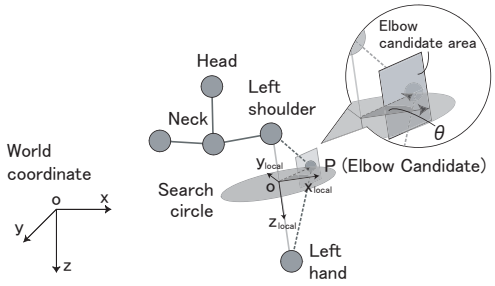


図 9 肘の探索範囲

人間が直立しているとして下が z の正である右手座標系とする) の内積が最も大きくなるものを x_{local} 軸方向，左肩から左手へのベクトルを z_{local} 軸方向，$x_{local} \times z_{local}$ 方向を y_{local} 軸方向と設定する．探索円の半径を r，x_{local} 方向の単位ベクトルを $\mathbf{i}$，y_{local} 方向の単位ベクトルを $\mathbf{j}$ とすると，肘の存在位置候補は $\mathbf{P} = rsin(\theta)\mathbf{i} + rcos(\theta)\mathbf{j}$ と表すことができる．

次に，θ の値を変えながら，そのときの 3 次元位置 $\mathbf{P}$ を各視点のシルエット画像に投影し，図 9 の肘周辺の矩形領域とシルエット画像との重なりを評価値として各視点から投票を行う．投票の際には，処理の高速化のため，矩形領域から 5×7 のサンプリング点を求め，シルエット画像と重なる数を投票数とする．そして投票数が一番大きくなる θ の値より求められる $\mathbf{P}$ を肘の位置推定結果とする．このとき，肘が体にめり込むのを回避するため，経験的に $0 \leq \theta \leq 120$ に肘の存在領域を限定している．

4.5.2 下半身の姿勢推定

上半身と同様，下半身に対しても逆運動学解法と逆投影探索による姿勢推定を行う．左右の腰の中間点位置を固定点とし，終端点の足先が正しい位置に変位するように人体モデルを変化させる．なお腰の中間点位置 (w_x, w_y, w_z) は，図 9 と同じ世界座標系において，頭の位置 (h_x, h_y, h_z) と右足の位置 (rf_x, rf_y, rf_z)，左足の位置 (lf_x, lf_y, lf_z) を用いて式 (3) によって求められるものとしている．

$$\begin{cases} w_x = h_x \\ w_y = h_y \\ w_z = 0.55h_z + 0.225rf_z + 0.225lf_z \end{cases} \tag{3}$$

5. 身体パラメータの取得

本システムにおける姿勢推定では，人体モデルにおける身体パラメータとして，ユーザの体の各関節間の長さをあらかじめ計測しておく必要がある．身体パラメータの取得は，これまで手作業により行ってきたが，対象人物が変わるたびにこの作業を行うことは時間と手間がかかり，またできたとしても正確に身体パラメータを取得できるかどうか分からないという問題があった．そこで，身体パラメータを多視点撮影画像から計測することを考える．身体が左右対称であるとすると，求めるべき身体パラメータは，首，肩幅，上腕，下腕，腰幅，上腿，下腿の 7 種類の関節間の長さとなる．

4.2 節の方法により，肌色を検出することで顔・手先・足先の位置を算出することができる．そこで，これらの情報を用いて身体パラメータの取得を行なう．まず，顔・手先・足先の位置が，身体パラメータの情報を顕著に表すような図 10 の基本姿勢を撮影する．そして撮影画像から求めた顔・手先・足先の 3 次元位置の情報から身体パラメータを算出する．具体的な各関節間の長さの算出方法を以下に示す．ここで，3 次元位置の座標系は，対象人物から向かって左が x 軸，正面が y 軸，鉛直下向きが z 軸とし，z 軸と床の交点位置の z 成分 z_g はカメラキャリブレーションの際に求めているとする．

首

図 10(a) の両手の中心と頭部の 3 次元位置の長さ．

肩幅

図 10(c) の両手の 3 次元位置の長さ l_{chh}．

上腕

図 10(b) の両手の 3 次元位置の長さを l_{bhh} としたときの $(l_{bhh} - l_{chh})/2$．

下腕

図 10(a) の両手の 3 次元位置の長さを l_{ahh} としたときの $(l_{ahh} - l_{bhh})/2$．

腰幅

肩幅と同じ l_{chh}．

上腿

図 10(d) の頭部の 3 次元位置の z 成分 z_{dh} を用いて $l_{dhg} = z_{dh} - z_g$ とする．また，図 10(a) の頭部と両足の中心の 3 次元位置の長さを l_{ahf} としたときの $l_{ahf} - l_{dhg}$．

下腿

図 10(e) の頭部の 3 次元位置の z 成分 z_{eh} を用いて $l_{ehg} = z_{eh} - z_g$ としたときの $l_{dhg} - l_{ehg}$．

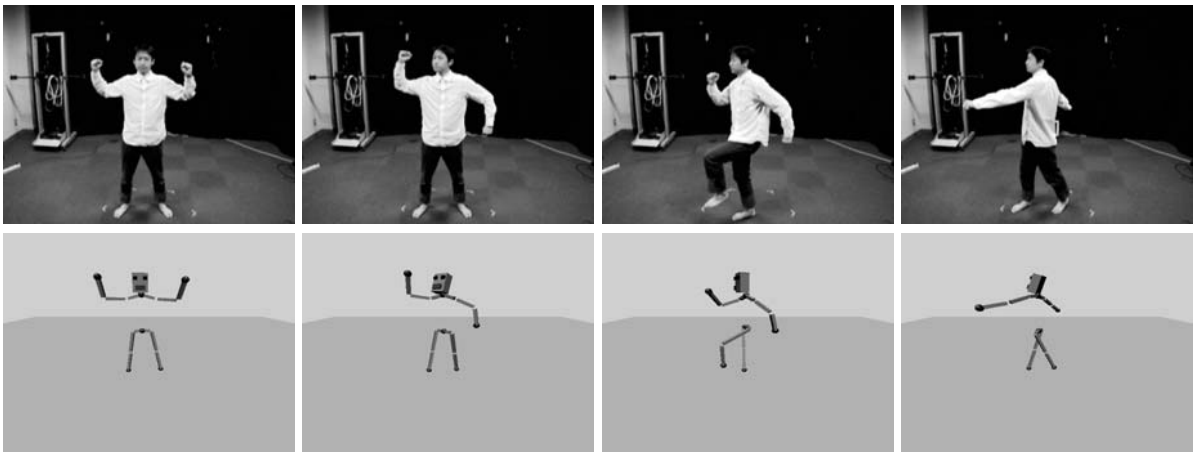
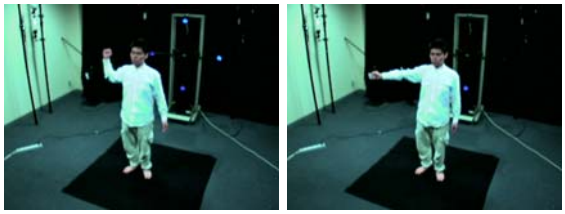


図 12 入力画像とモーションキャプチャの結果．上段:入力画像．下段:CG アバタ



(a) (b)

図 11 実験の基本姿勢 (a, b)

表 1 各処理の計算コスト

処理	時間
ブロップ追跡 (HPM)	9ms
顔の向き推定 (FPM)	8ms
3 次元姿勢推定 (3DPM)	18ms

6. 性能評価実験

本論文で述べたモーションキャプチャシステムに対する性能評価実験を行った．図 12 は，実際にユーザの姿勢推定を行った結果を CG アバタに反映させたものである．実験環境として，8 台のカメラを用いて 12fps で撮影した画像を入力として姿勢推定を行った．評価項目は処理速度と精度で，実時間性は各モジュールの計算時間，精度は本システムの手法上，誤差が最も増大しやすい肘位置の推定誤差をみることで評価する．

6.1 処 理 速 度

本システムにおける各モジュールの処理時間を表 1 に示す．PC クラスタにおける一段目の PC での処理時間の合計は約 17ms、二段目の PC での処理時間は約 18ms であり、パイプライン処理を行なっているため，理論的に 55fps ほどのフレームレートでの動作が可能である☆．これは、十分に実用的な処理速度を確保できているといえる．

☆ 利用しているカメラ性能の限界のため，実際には 12fps で実験を行なっている．

6.2 肘の 3 次元位置推定誤差

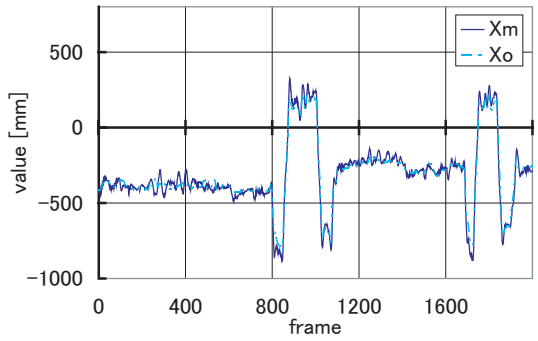
本システムの姿勢推定法の誤差を定量的に評価するために，本システムとは別に 3 次元位置を計測できる装置を用いて肘位置の真値を獲得し，本システムの計測結果と比較する．ここでは，超音波により 3 次元位置を計測する装置を用いる．性能評価実験での計測の際の動作を以下に示す．

1. { 基本姿勢 1（図 11(a)），基本姿勢 2 （図 11(b)） } で { 肘の前後運動，肘の上下運動，肘の前後方向の円運動，膝のみ屈伸，上半身固定で歩き回る }
2. 基本姿勢 1 で { 肘を固定して手先を円運動，肩と手先を固定して肘だけ円運動 }
3. 基本姿勢 1 と 2 の繰り返し

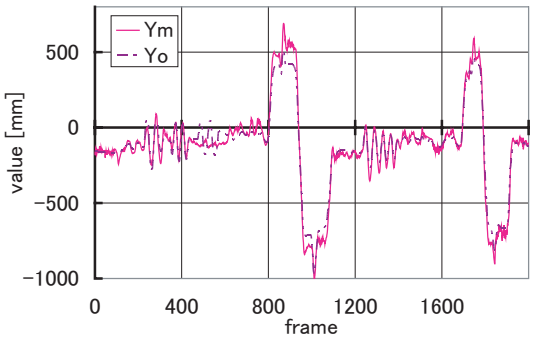
図 13 に 2000 フレーム分の肘の軌跡を，図 14 に真値と推定値のユークリッド距離を示す．ここで，図 13 の座標系は，ユーザから向かって左方向を X 軸正方向，正面方向を Y 軸正方向，鉛直下向きを Z 軸正方向とする右手系である．この実験での平均誤差は 75mm ほどであった（およそ腕の直径と等しい）．

誤差が生じる原因は，大きく二つ考えられる．一つは身体モデルが実際の人間よりも簡単化されたものであるため，その時の姿勢によって精度にばらつきが生じてしまうことによるものである．

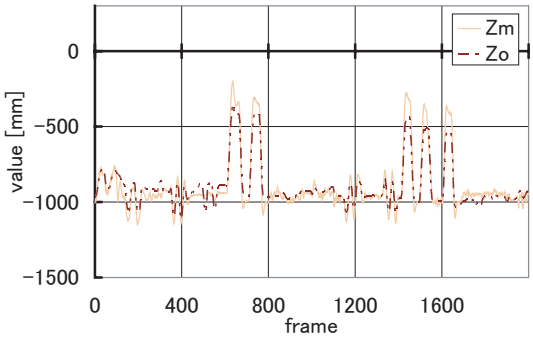
もう一つは，肘位置探索の逆投影探索でシルエット画像との重なりを利用しているが，重なったシルエットの位置が肘であるかどうか分からないため，視点によっては腕以外の部分と探索円が重なったことによる誤った投票が行われる可能性があることによるものである．また，モデルや肘位置探索のアルゴリズム上の問題以外にも，誤差の中には超音波センサの座標系とモーションキャプチャの座標系との位置合わせの際の誤差や、カメラキャリブレーションの精度に関する誤差も含まれていると考えられる．



(a)X 軸



(b)Y 軸



(c)Z 軸

図 13 肘の 3 次元位置の軌跡: （X_m, Y_m, Z_m : 推定値, X_o, Y_o, Z_o : 真値）

7. おわりに

本論文では多視点動画像処理による実時間全身モーションキャプチャシステムで必要となる身体パラメータの自動推定法として，複数のカメラで決められた姿勢の人物を撮影し，その画像から測定できる頭部・両手・両足の 3 次元位置を基準として体の各関節間の長さを自動的に算出する方法を提案した．また，性能評価実験により，システムの処理速度が十分であることを確認し，平均誤差がだいたい腕の直径と同等程度ということで，精細ではないが，およそユーザの姿勢の推定が可能であることが分かった．

本システムは、実時間で動作するため、仮想空間におけるインタラクションなど様々な応用アプリケーションへの適用が期待できる．

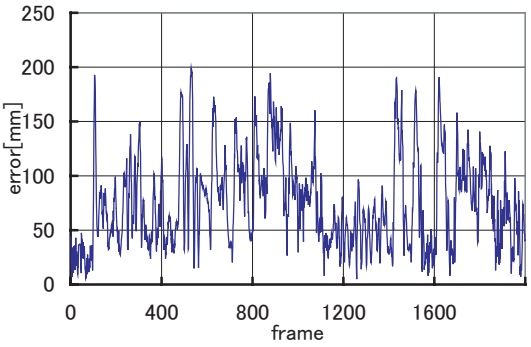


図 14 肘の 3 次元位置推定誤差

今後の課題としては、より複雑な姿勢の推定を実現することが挙げられる．例えば，本システムでは計算の簡単化のため，胴体が鉛直になっていることを前提としているため，顔は常に腰の真上に位置していなければならない．このためには，腰位置を別の方法で算出するなどの工夫が必要となる．

参 考 文 献

1) Carranza, J., Theobalt, C., Magnor, M. and Seidel, H.: Free-Viewpoint Video of Human Actors, *Proc. of ACM SIGGRAPH 03*, pp.569–577 (2003).
2) Sand, P., McMillan, L. and Popovic, J.: Continous Capture of Skin Deformation, *Proc. of ACM SIGGRAPH 03*, pp.578–586 (2003).
3) Wren, C., Azarbayejani, A., Darrell, T. and Pentland, A.: Pfinder: Real-Time Tracking of the Human Body, *IEEE Trans. on PAMI*, Vol.19, No.7, pp.780–785 (1997).
4) 岩澤昭一郎，海老原一之，竹松克浩，坂口竜己，大谷　淳：“Shall We Dance?” の構築，*PRMU98-114*, pp.15–22 (1998).
5) 米元　聡，有田大作，谷口倫一郎：多視点動画像処理による実時間全身モーションキャプチャシステム－視覚に基づく仮想世界とのインタラクション－，映像情報メディア学会誌，Vol.54, No.3 (2000).
6) 有田大作，花田武彦，谷口倫一郎：分散並列計算機による実時間ビジョン，情報処理学会論文，Vol.43, No.11, pp.1–10 (2002).
7) Tsai, R.: A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, *IEEE Trans. on Robotics and Automation*, Vol.3, No.4, pp.323–344 (1987).
8) Jordao, L., Perrone, M., Costeira, J. and Santos-Victor, J.: Active Face and Feature Tracking, *Proc. of the 10th International Conference on Image Analysis and Processing*, pp.572–576 (1999).
9) Welman, C.: Inverse Kinematics and Geometric Constraints for Articulated Figure Manipulation, PhD Thesis, Simon Frasier University (1993).