

多視点動画画像解析による身体の実時間姿勢推定

伊達, 直人
九州大学システム情報科学研究院知能システム学部門

吉松, 寿人
九州大学システム情報科学研究院知能システム学部門

有田, 大作
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

<https://hdl.handle.net/2324/5907>

出版情報 : 画像の認識・理解シンポジウム, pp.678-683, 2004-07
バージョン :
権利関係 :

多視点動画画像解析による身体の実時間姿勢推定

伊達 直人[†] 吉松 寿人[†] 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 6-1

E-mail: †{naoto,hisato,arita,rin}@limu.is.kyushu-u.ac.jp

あらまし 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キャラクターの行動生成、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、沢山の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あるが、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、発表者は、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。さらに、本システムを利用したアプリケーションを作成することで本システムの有効性を示す。

キーワード コンピュータビジョン、モーションキャプチャ、多視点画像解析、PC クラスタ

Vision-based Real-time Human Motion Sensing and Its Applications

Naoto DATE[†], Hisato YOSHIMATSU[†], Daisaku ARITA[†], and Rin-ichiro TANIGUCHI[†]

[†] Department of Intelligent system, Kyushu University Kasuga 6-1, Kasuga-shi, Fukuoka, 1816-8580 Japan

E-mail: †{naoto,hisato,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 and we have constructed a “virtual reality walk through” system using this system.

Key words Computer vision, Motion capture, Multi-view image analysis, PC cluster

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の操作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、実時間性を担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のような画像特徴を用いるかが重要である。一般には、プロップ (密な領域) [3] やシルエットの輪郭線 [6] 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。本稿では、これらの画像から検出される限られた情報から人体の三次元姿勢を推定する手法について述べる。特に、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、我々は多視点画像解析というアプローチを行っている。また、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、我々は PC クラスタを利用して、今回新たに、顔の向きを推定、人体モデルの複雑化、および姿勢推定アルゴリズムの改善を行い、姿勢推定の精度向上を図った。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

本手法では、1 台の顔用カメラと約 45 度間隔で円周上に配置した 8 台の水平カメラを用いて処理を行なう (図 1)。これらは固定カメラで、計測空間中に設置したワールド座標系とカメラ座標系の間であらかじめキャリブレーションをしておくものとする [11]。

一般に多眼視によって 3 次元位置の推定を行なうには 2 視点あれば可能であるが、本手法では多視点の情報を統合処理することで、より多くの姿勢へ対応することを目標としている。また、実時間処理を可能にするためにシステムは並列のパイプライン処理を行なう構成にしている。

各モジュールの機能は以下の通りである。

- 画像獲得モジュール (ICM)

画像の取り込みを行なうモジュール。このモジュールは視点の数だけあり、各 ICM は同期信号によって同時に撮影されたカメラ画像を獲得し、顔画像処理モジュール、水平画像処理モジュールに画像データを転送する。

- 顔画像処理モジュール (FPM)

顔用カメラによって撮影した画像から、顔の向き情報を抽出する。処理結果は、3 次元復元モジュールへ転送する。

- 水平画像処理モジュール (HPM)

人物の全身を撮影できる水平カメラで獲得した画像から特徴量の抽出を行なう。処理の結果は 3 次元復元モジュールへ転送する。

- 3 次元復元モジュール (3DPM)

FPM, HPM から受信する画像から抽出した特徴量を統合処理して人物の 3 次元姿勢推定を行なう。推定結果は実時間レンダリングモジュールへ渡される。

- 実時間レンダリングモジュール (RRM)

3 次元復元モジュールから渡されたデータをもとに、仮想空間に実時間で CG キャラクターを描画する。

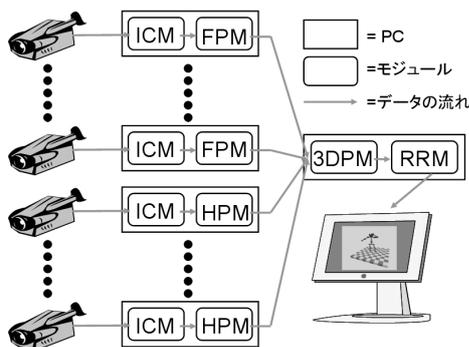


図 1 PC クラスタにおける処理モジュールの配置

3. 特徴抽出

3.1 色プロップの 3 次元位置推定

3.1.1 HPM における処理

本システムでは、性質が異なる 2 つの姿勢推定アルゴリズムを利用している。一方は、人物の肌色領域を計測することによって姿勢推定を行う計測ベースの手法で、他方は、人物の輪郭線の特徴量としたパターン認識によって姿勢推定を行う認識ベースの手法である。計測ベースの手法では、過去の姿勢を参考にして色プロップの追跡を行うために、時間の経過に伴い推定精度が低下する。これに対して、認識ベースの手法の推定精度は時間に依存しない。そこで、本システムでは、認識ベースの手法をリセット処理として利用している。これらの処理に必要な情報を抽出する HPM における処理を以下に示す。

(1) 人物領域抽出

入力画像から人物領域を抽出することで、画像処理対象を限定する。

(2) 肌色領域検出

人物領域から特徴量として肌色領域重心を求める。

(3) パターン認識処理

人物領域の輪郭線情報を利用して、パターン認識処理を行い登録姿勢かどうかを判定する。登録姿勢は両手を開いた直立姿勢。

これらの処理によって抽出した人物領域の画像、肌色領域、パターン認識結果は 3DPM に転送される。

3.1.2 色プロップの 3 次元位置計測アルゴリズム

ここでは、3DPM における色プロップの 3 次元位置計測アルゴリズムについて述べる。

一般的なステレオ計算によって肌色領域を計測するが、視点間でどの点とどの点に対応しているかを決定することは容易ではない。そこで次のアルゴリズムによって肌色領域の対応関係とつつ 3 次元位置を計測する。まずは、全ての視点における肌色領域重心に対応した直線の中から 2 本の直線のペアを選択し交差点判定を行う。2 直線間の距離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。一方、2 直線間の距離が閾値より大きければ選択した点のペアは同一の肌色領域重心ではないとし肌色領域は存在しないものとする。この処理を同一点である可能性のある全ての点のペアに対して適用する。求まった各点は、水平カメラで得られた人物のシルエット画像に逆投影することで、不適切な点を削除する。3 次元空間中の肌色領域を示す各点は、それぞれ最も近い前回のフレームの特徴点 (顔、左手、右手、左足、右足のいずれか) に分類する。前提として、前回のフレームの特徴点の 3 次元位置は正しく推定できているものとし^(注1)、この操作によって、顔のクラスタ、左手のクラスタ、

(注1)：認識ベースの姿勢推定が行われた際に、必ず正しく特徴点が決定されることを利用して初期位置設定を行う

というように各特徴点のラベルのついたクラスタができる。この処理結果例を示す。図2の姿勢に対して得られた肌色領域の3次元分布が図3である。各クラスタの密な点の平均値を現在のフレームの特徴点の3次元位置として決定する。これにより、各特徴点の3次元位置が計測できる。



図2 入力姿勢

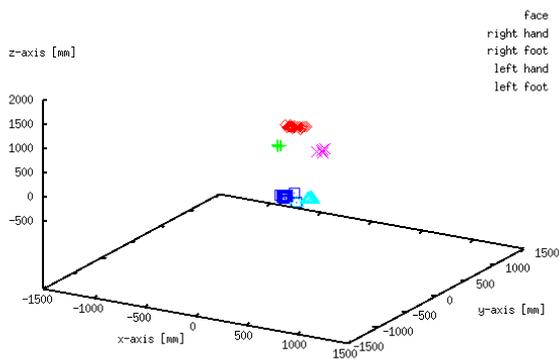


図3 肌色領域分布

3.2 顔の向き推定

3.2.1 処理概要

FPMでは、人物の顔の向きを推定する。顔の向きや視線を推定するためには、顔画像の目や口などの特徴点を求めて特徴点のステレオ計算を行なう手法[10]が存在する。しかし、本システムの場合は、画像から特徴点を抽出できるほどの解像度は無い。そこで、あらかじめ顔の向き情報と対応するテンプレート顔画像を撮影しておき、入力画像の顔領域と最も似ているテンプレート顔画像を探索する手法を採用した。本システムでは、実時間処理を行うため、固有空間法[13]を用いることで高速化を行なった。顔用カメラ画像処理の概要は次の通りである。

(1) 前処理

システムを利用する前に、テンプレート顔画像を撮影しながら、顔の向き情報を超音波センサで獲得する。顔画像は一定の大きさに伸縮することで正規化しておく。テンプレート顔画像のなす固有空間を求めておき、各テンプレート顔画像の固有係数も求めておく。

(2) 顔検出

入力画像に対して、人物領域の切り出しと肌色領域の分割を行ない、最も大きな肌色領域を入力画像の顔領域とする。

(3) テンプレートマッチング

入力画像の顔領域を多次元ベクトルとみなし、最も近い距離にあるテンプレート顔画像を探索する。テンプレート顔画像の探索は、固有空間で行う。

3.2.2 顔テンプレート画像の生成

顔向き検出に利用するテンプレート顔画像は、カラー画像で撮影を行い、顔領域は 80×130 の大きさに伸縮して正規化しておく。このとき、超音波センサで獲得した顔の向き情報も同時に保存しておく。顔の向きの自由度は二自由度とし、視線方向軸についての回転角は考えないものとする。顔テンプレート画像は、テンプレート顔画像群の張る固有空間に射影しておくことで、次元を下げておく。今回は、 $80 \times 130 \times 3$ 次元で300枚のテンプレート顔画像を用意して、これらを60次元(認識精度の保持と処理時間の短縮を両立するために経験的に決定した次元)に下げて利用した。

3.2.3 パターン認識による顔向き検出

入力画像において、人物領域の切り出しと肌色領域分割を行ない、顔領域を決定した後、顔領域をテンプレート顔画像と同じ大きさ(80×130)に正規化する。正規化した顔画像を固有空間に射影し次元を下げた固有係数を求め、最も距離の近いテンプレート顔画像の固有係数を探索する。テンプレート顔画像の探索は、全探索によって行なうが、本システムで用意したテンプレート顔画像の枚数(300枚)の場合は、顔画像処理時間は15msec/frame程度で充分実時間動作可能であった。顔用カメラ画像処理の結果を図4に示す。左から入力画像、選択したテンプレート顔画像、顔の向きを反映させた顔アバタ(分身)である。図に示すように、様々な顔の向きに対して姿勢推定が可能であった。

4. 3次元姿勢推定

本システムでは図5に示すような身体モデルを用いている。このモデルは自由度が23あり、首回りや肩関節等の自由度が高く、身体姿勢の表現能力が高くなっている。しかしその一方で、姿勢推定には工夫が必要であり、ここでは一般的な逆運動学の数値解法と画像認識を併用するアルゴリズムを開発した。

4.1 上半身の姿勢推定

人間の全身姿勢を推定するためには、肩や腰等の位置が必要である。しかし、肩や腰等を画像から精度よく抽出することは容易でない。そこで、限られた情報から人間の全身姿勢を推定するために、肌色領域の3次元位置情報と顔用カメラからわかる顔の向きをもとにして全身の姿勢推定を行う手法を適用した。上半身の姿勢推定は、次の2ステップの処理によって行う。

(1) 逆運動学解法

人体モデルに基づき大まかな人間の姿勢を推定する。推定結果の肩の位置はそのまま利用するが、肘に関し

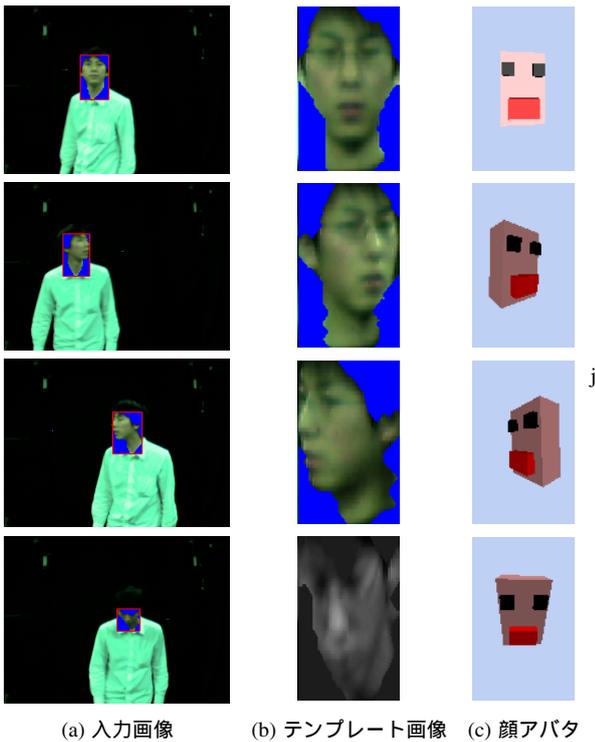


図4 顔向き検出の結果

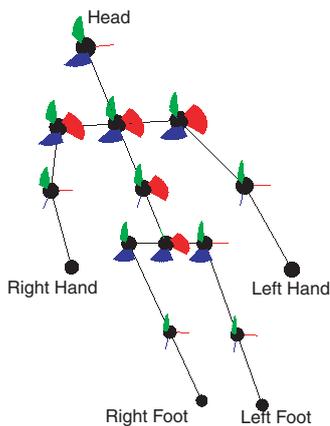


図5 身体モデル

ては、次のステップで探索を行って推定する。

(2) 肘の探索

肩の位置と手先の位置から得られる肘の存在領域と人間のシルエット画像の相関をとることで肘の位置を推定する。

4.1.1 逆運動学解法による一次推定

およその全身姿勢を推定するために、計測した肌色プロップの位置条件を満たす姿勢を逐次計算によって見つける逆運動学解法を行う。現在のフレームで、ワールド座標系における手先の位置 $(x(t), y(t), z(t))$ が人体モデルのパラメータ $T_x(t), T_y(t), T_z(t), \theta_1(t), \theta_2(t), \dots, \theta_i(t)$ によって $\mathbf{P}(T_x(t), T_y(t), T_z(t), \theta_1(t), \theta_2(t), \dots, \theta_i(t))$ と表せるとする。ここで、 $T_x(t), T_y(t), T_z(t), \theta_1(t), \theta_2(t), \theta_3(t)$ は時刻 t における頭部の位置およびワールド座標系と頭部座標系間の回転角、各 $\theta_j(t), 4 \leq$

$j \leq i$ は時刻 t における連結した角パーツの座標系間の回転角とする。次のフレームで手先の位置が $(x(t+1), y(t+1), z(t+1))$ へ、頭部の位置が $(T_x(t+1), T_y(t+1), T_z(t+1))$ に変位し、ワールド座標系と頭部座標系間の回転角が $\theta_1(t+1), \theta_2(t+1), \theta_3(t+1)$ であるとする。このとき、手先の位置 $\mathbf{P}(T_x(t+1), T_y(t+1), T_z(t+1), \theta_1(t+1), \dots, \theta_i(t+1))$ が $(x(t+1), y(t+1), z(t+1))$ に近づくように、各 $\theta_j(t+1), (4 \leq j \leq i)$ を微量ずつ変化させる。この操作を繰り返して得た人体モデルの肩の位置を一次推定の推定結果として用いる。手先の位置変位に連動して肩が動くイメージとなる。正確な位置を計測できるわけではないが滑らかな動作を生成する。なお、人体モデルの角自由度は、頭が2、首元が2、肩が3、肘が1の計8で、 $j=8$ とした。

4.1.2 逆投影探索による2次推定

肌色 blob によって計測した手先の位置と逆運動学解法によって推定した肩の位置から、肘の位置は三次元空間内の円周状に制限されこれを探索円とする。しかし、肘が胴体にめり込むような姿勢がありえないように、探索円上には明らかに解の存在しない領域がある。このような解の存在区間を求めるために次のアプローチを行った。左肘を推定する場合を例に挙げると、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、 y_{local} 軸方向を $x_{local} \times z_{local}$ 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、肘の位置候補は $\mathbf{P} = r \sin(\theta) \mathbf{i} + r \cos(\theta) \mathbf{j}$ と表すことができる。今回、 $0 \leq \theta \leq 120$ という制限をつけることで肘の存在領域を設定した。左肩の位置と左手の位置、 \mathbf{P} によって成される平面上存在する図6のような矩形領域を各視点の人物のシルエット画像に投影して相関 $C(\theta)$ を求め、 C の最大値をあたえる θ に対応する肘の位置を推定結果とする。

4.2 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、足先が正しい位置に変位していくように人体モデルを変化させていく。ただし、腰の中間点位置 (w_x, w_y, w_z) は、頭部の位置 (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)

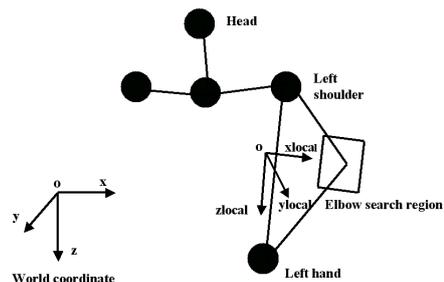


図6 肘の探索区間

によって次式のようにあらわすものとした．

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

ただし， z 方向は実空間の上下方向とする．この様に腰の中間点位置を決定した場合，上半身をある程度起こした姿勢でない正しい姿勢推定はできないという問題がある．今後，腰の位置を正しく推定することが課題である．逆運動学解法によって大まかな姿勢推定を行った後は，逆投影探索により膝の位置を推定する．

5. 計測と認識の融合によるエラー回復

実時間のシステムでは，システムがエラーを起こしたときにシステムをリセットすることなくオンラインでエラー回復を行えることが重要である．肌色領域計測をもとにした身体姿勢の推定手法は，プロップの探索範囲や分類基準に過去のデータを用いるため，ある時点で姿勢推定を誤るとそれ以降正しい姿勢推定が行えなくなることがある．そこで，本システムでは，ある特定の姿勢に限られるもののそれらを安定に認識できるパターン認識アルゴリズムを並行に動作させておき，その姿勢が認識された時点で身体姿勢の計測結果を認識結果に合致するように修正するという手法を導入している．

つまり，パターン認識による手法は登録データに近い入力に対してのみ姿勢推定が可能であるが，過去のデータを手がかりにしないため，精度は時間に依存しない．したがって，従来の手法と融合し，パターン認識がうまくいった場合はパターン認識による姿勢推定結果を適用し，それ以外の場合は肌色領域を計測する手法を適用することでエラー回復が行われ，システムが安定に動作することが出来る．現在，認識手法として水平カメラ画像中の人間の輪郭線と，テンプレートの輪郭線の類似度を求め，類似度が十分大きければ，想定する姿勢になったと判断し，身体パラメータを強制的に修正するようにしている．

6. 実験と評価

6.1 姿勢推定の評価

本システムでは，逆運動学解法によって大まかに人物の姿勢推定を行った後，シルエット画像を逆投影探索することで正しく姿勢推定を行う．ここでは，逆投影探索を行うことでより正確な人物の姿勢推定が可能になるという有効性を示す．図7の一段目に関して，逆運動学解法のみを利用した場合は，右膝位置推定が誤っているが，逆投影探索を行うことで正しい位置に補正できている．2段目に関しては，逆運動学解法のみの場合には，左肘の位置が不自然だが，逆投影探索によって自然な位置に補正されている．その他の画像に関しても，逆投影探索を行うことで，より正確な肘や膝の推定が可能になっており，逆運動学解法に加えて逆投影探索処理を行う手法の有効性を示した．

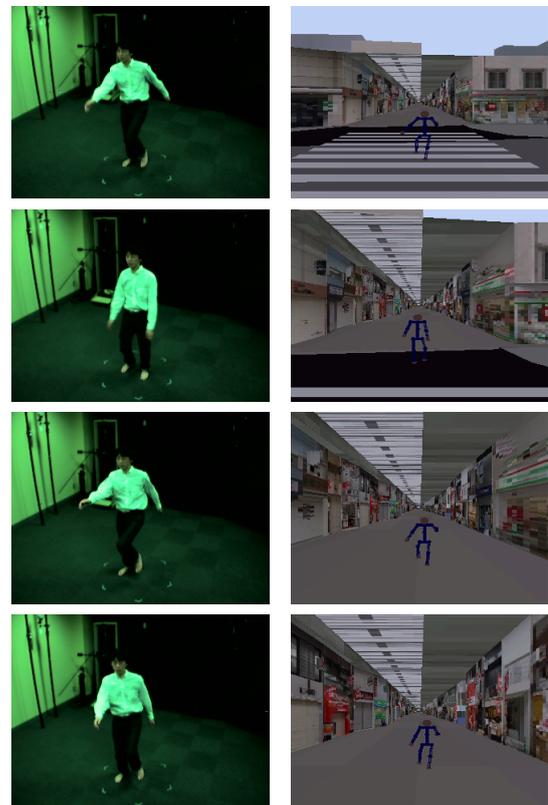
6.2 仮想空間歩行アプリケーション

本システムの応用例として，仮想空間歩行アプリケーションを作成した．仮想空間を歩行する様子を，図8に示す．この

システムでは，モーションキャプチャによってユーザのアバタ(分身)を仮想空間内に表示する．仮想空間は，プロジェクターによってスクリーンに投影することで，ユーザに提示する．また，限られた範囲の計測空間と広大な仮想空間との距離的対応をとるために，ユーザの姿勢情報をコマンドとして認識し，仮想空間の歩行や視界の回転に反映される．これらのコマンドを表1に示す．

表1 仮想空間歩行アプリケーションの身振りコマンド

身振り	実行されるコマンド
右を向く	視界が右に旋回
左を向く	視界が左に旋回
前方に立ち歩く	前進
後方に立ち歩く	後進



(a) 入力画像 (b) 仮想空間歩行の様子

図8 仮想空間歩行

7. おわりに

本稿では，特別なマーカー，センサーを必要としない画像解析による実時間モーションキャプチャシステムについて示した．全身動作を解析するために複数視点のカメラからの情報を融合して人間の姿勢を復元するアルゴリズムについて示した．今回，新たに固有空間法を用いた顔向き検出を行うことで，推定可能な姿勢の自由度を増やすことができた．PC クラスタに実装したシステムは，実時間で動作可能で，現実世界と仮想空間をシームレスにつなぐ様々なインタラクティブシステムに適用できると考える．その一例として，仮想空間歩行アプリケーションを

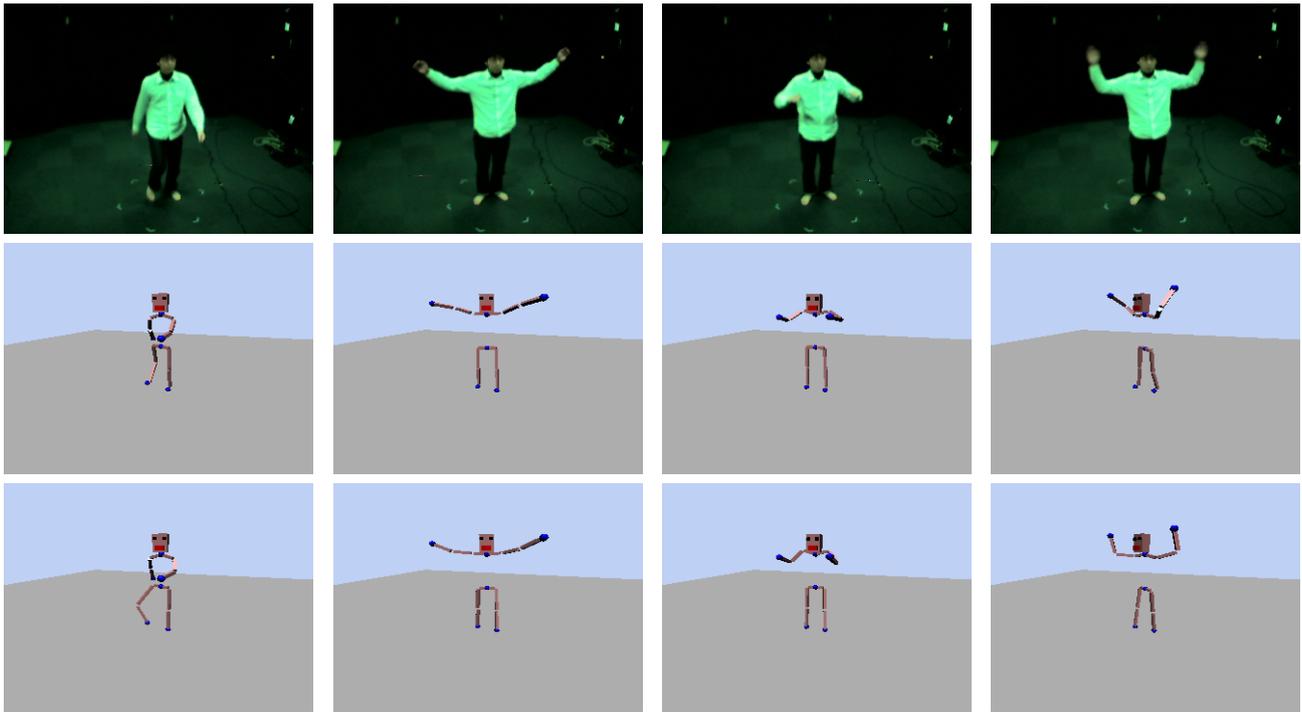


図7 逆運動学解法と逆投影探索の比較．上段:入力画像．中段:逆運動学解法．下段:逆投影探索

作成した．今後は，よりロバストな人間の姿勢推定を可能にするために画像解析アルゴリズムを改善をめざす．それに伴い，適用可能なアプリケーションの幅が広がると考える．

文 献

- [1] J.Carranza, C.Theobalt, M.A.Magnor, H.P.Seidel, "Free-Viewpoint Video of Human Actors", *Proc. of ACM SIGGRAPH 03*, pp.569-577, 2003.
- [2] P.Sand, L.McMillan, J. Popovic, "Continuous Capture of Skin Deformation", *Proc. of ACM SIGGRAPH 03*, pp.578-586, 2003.
- [3] C.Wren, A.Azarbayejani, T.Darrell, A.Pentland, "Pfinder: Real-Time Tracking of the Human Body", *IEEE Trans. on PAMI*, Vol.19, No.7, pp.780-785, 1997.
- [4] J.Deutscher, A.Davison and I.Reid, "Automatic Partitioning of High Dimensional Search Spaces associated with Articulated Body Motion Capture", *2001 IEEE CS Conf. on Computer Vision and Pattern Recognition*, Vol.2, pp.669-676, 2001.
- [5] I.Haritaoglu, D.Harwood, L.S.Davis, "Ghost: A Human Body Part Labeling System Using Silhouettes", *Proc. of International Conference on Pattern Recognition*, pp.77-82, 1998.
- [6] 岩澤, 海老原, 竹松, 坂口, 大谷, "「Shall We Dance?」の構築", 信学技報 PRMU98-114, pp.15-22, 1998.
- [7] R.Taniguchi, D.Arita, "A Basic Framework of Real-Time Image Processing on PC-cluster", *Proc. of 2nd International Workshop on Cooperative Distributed Vision*, pp.119-132, 1998.
- [8] H. Tezuka, A. Hori, Y. Ishikawa, M. Sato, "PM: An Operating System Coordinated High Performance Communication Library", *Proc. of High-Performance Computing and Networking*, pp.708-717, 1997.
- [9] Jordao, L., Perrone, M., Costeira, J.P., Santos-Victor, J. (Inst. de Sistemas e Robotica, Inst. Superior Tecnico, Lisbon, Portugal) "Active face and feature tracking", *Proc. of 10th International Conference on Image Analysis and Processing*, 572-6, xxii+1232, 1999.
- [10] Shigenori Ishii, Daisaku Arita, Rin-ichiro Taniguchi "Real-time Head Pose Estimation with Stereo Vision", *Proc. of the 9th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pp.79-97, 2003.
- [11] Roger Y.Tsai, "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.
- [12] Sung-Eun Kim, Ran-Hee Lee, Chang-Joon Park, and In-Ho Lee, "MIMIC: Real-time Marker-free Motion Capture System to Create an Agent in the Virtual Space", *Proc. of the International Conference on Computers in Education*, 2002.
- [13] H.Murase, S.K.Nayar. "Visual learning and recognition of 3-d objects from appearance", *International Journal of Computer Vision*, Vol.14, No.1, pp.5-24, 1995.
- [14] M.Turk, A.Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, 1991.