九州大学学術情報リポジトリ
Kyushu University Institutional Repository

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

米元, 聡
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/2324/5815

出版情報：映像情報メディア学会誌 : 映像情報メディア. 54 (3), pp.409-416, 2000-03-20. 映像情報メディア学会
バージョン：
権利関係：

KYUSHU UNIVERSITY

The Institute of Image Information and Television Engineers

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 ―

Visually Guided 3-D Animation System Using Real-time Marker-less Motion-Capture with Multiple-Camera Fusion

米元　聡†，　有田大作†，正会員　谷口倫一郎†

Satoshi Yonemoto†, Daisaku Arita† and Rin-ichiro Taniguchi†

Abstract A realtime marker-less motion-capture system is described that can easily and seamlessly map objects in the real world into a virtual environment. In general, virtual environment applications, such as man-machine seamless interaction, require the system to estimate the motion parameters for natural objects such as human bodies in realtime. To achieve this, the developed motion-capture system achieves multiple-camera fusion and reconstructs the parameters for complete human-body motion using real-time inverse kinematics. Implementation of this system demonstrated its ability to work in realtime and online on a pc-cluster.

キーワード：非接触式モーションキャプチャ，多視点融合，実時間動画像処理，仮想空間と実空間の融合

1. まえがき

人間の動作を実時間で計測し，認識・再生することができれば，3次元アニメーションやビデオゲームにおけるキャラクタの動作生成，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作など，多くのアプリケーションでの適用が期待できる．そのような目的を達成するため，我々は多視点カメラを用いた実時間モーションキャプチャシステムの開発を行っている．従来より，人間の動作を実時間でセンシングする技術として，光学式，磁気式などセンサ接触型の方法が用いられてきた．一般に，それらの手法によるものは，システムが大規模になり高価になること，それら装置の性質により測定対象の動きに制限が生じるなど，必要以上に身体的な拘束を課してしまうことなどの問題がある．一方，近年になって，非接触型の方法としてビデオカメラで撮影した映像から3次元動作を計測する方法が，コンピュータビジョンの応用により試みられるようになってきたが[1]，以下のような点で実用レベルにまで達していないのが実情である．

・実時間*での計測やオンラインでの再生ができない．
・センサの配置などによる物理的な制約の他，計測アルゴリズムの性質により動きの種類が限定されるなど多くのシーンに関する制約が必要である．
・一般に必要とされるようなモーションキャプチャのデータとしては，精度の点で課題がある．

しかし，非接触あるいは接触型の方法で，ビデオカメラによりセンシングする技術は，対象の表面情報や部位の形状あるいは対象以外のシーン情報といった，他の接触型の装置で計測不可能な情報を同時に得ることができるという可能性を有しているため，我々は実時間でビデオカメラを用いてセンシング・および認識・再生する技術の実用化を進めている．

ビデオカメラで撮影した映像より人間の全身動作を推定し，仮想空間上でインタラクションする方法としては，Pfinder システム[1] がすでに提案されている．このシステムを用いたアプリケーションは，一般的な領域クラスタリング手法を用いて2次元の *blob* 特徴を抽出し，完全非接触の方法で3次元推定を行う．比較的安価なシステム構成でありながら実時間に近い速さで動作可能であり，非接触方式による精度の低い抽出結果でも多くのアプリケーションが充分に機能することを示した．しかしながら，仮想空間上のキャラクタに全身動作をさせるだけの情報を抽出し，

1999年9月1日受付，1999年10月26日再受付，1999年12月6日採録
† 九州大学 大学院 システム情報科学研究科
（〒816-8580 春日市春日公園 6-1, TEL 092-583-7618）
† Graduate School of Information Science and Electrical Engineering, Kyushu University
(6-1, Kasuga-koen, Kasuga-shi, Fukuoka 816-8580, Japan)
* 我々の目指す「実時間」とは，システムのスループットとして，動画像の入力によく用いられる NTSC ベースのカメラの画像取り込み速度である 30fps を達成し，レイテンシー（遅延）が実時間インタラクションを行うアプリケーションとして支障のない程度であることを指す．

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

実時間動作させるには至っていない．また，より詳細な形状・動きパラメータを推定する方法としては，オフラインで獲得された画像系列に対し，モデルベースで解析する方法[6)7)]などもある．しかし処理時間が比較的膨大であることや，モデル像と画像のマッチングを行うのに必要な精度である初期情報はオフラインでしか得ることができないという制約のため，実時間，オンラインでの適用には課題が残っている．

本論文では，実時間，オンライン動作の実現を最も重要視し，多視点カメラセンサ系により実時間で全身動作の計測および再生を行うシステムについて提案する．以下では，システムの概要および非接触方式で全身を計測するためのアルゴリズムについて説明し，最後に実験結果について述べ全身動作の場合でも仮想空間での再現系まで含めて実時間 (30fps) で動作することを示す．

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

実世界において知能視覚システムが有効に機能するために必要な基本特性として以下が挙げられる[5)]．

・実時間性，オンライン性，ロバスト性，柔軟性

マルチセンサ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最も有効な実現方式のひとつであり，複数のカメラにより獲得された冗長な証拠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うるオクルージョンなど様々なエラー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基本特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，入力信号全体のバンド幅や処理量の点で，高性能な分散システムが必要なってくる．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を開発し，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．システム内の各処理モジュールは，PC 内のプロセスに割り当てられ，同期機構により並列に動作する．PC クラスタの詳細および同期機構の詳細については文献[10)11)]参照のこと．

2.2 システムの流れ

実時間モーションキャプチャのアルゴリズムの概要は以下のようになっている．

（1） 知覚モジュールによる 3-D blob 追跡
　（a） 各視点における画像獲得 (ICM)
　（b） 各視点における 2 次元画像特徴 (2-D blob) 抽出 (2DPM)
　（c） 多視点融合による 3-D blob 復元 (3DPM)

（2） 人体モデルの動き生成，仮想空間のレンダリングおよびリアクション計算 (RRM)

（3） (1)(2) を実時間で繰り返し処理する．

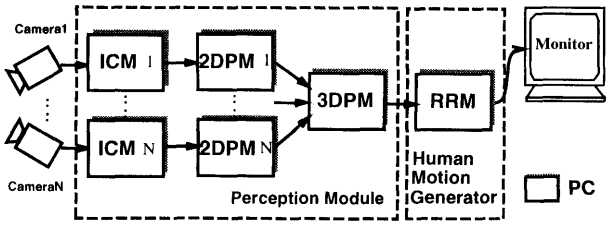

図 1 PC クラスタにおける処理モジュールの配置
An arrangement of the processing modules on PC cluster.

2.3 処理モジュール

図 1 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．それら各処理モジュールを以下のように設計している．

画像獲得モジュール (ICM)

これらの処理モジュール ICM は，画像の取り込み (原画像は 640 × 480 の大きさの 24bit カラー画像) とリサイズ (320 × 240) を行う．各 $\mathrm{ICM}_v(v = 1, \cdots, N; N$ は視点数) はそれぞれ，2 次元画像特徴抽出モジュール ($\mathrm{2DPM}_v$) に画像を送信する．

2 次元画像特徴抽出モジュール (2DPM)

これらの処理モジュール 2DPM は，2 次元画像特徴の抽出処理 (2-D blob 追跡) を行う．各 2DPM はそれぞれ ICM より画像を受信し，処理結果 (2-D blob 重心) を 3 次元復元モジュール (3DPM) へ送信する．

3 次元復元モジュール (3DPM)

処理モジュール 3DPM は，$\mathrm{2DPM}_v(v = 1, \cdots, N)$ で得られた処理結果を融合し，3 次元情報 (3-D blob 位置) の復元を行った後，結果を実時間レンダリングモジュール (RRM) へ送信する．

実時間レンダリングモジュール (RRM)

処理モジュール RRM は，仮想空間の実時間フルスクリーンレンダリングおよびリアクション計算を行う．なお受信した 3 次元情報を用いて人体モデルの動き生成もこの処理モジュールにおいて行う．

3. 知覚モジュール

画像上に観測の容易なマーカ領域などが仮定できない非接触の方式では，人体領域に相当する画像特徴として輪郭特徴，肌色領域などの領域特徴が利用される．本システムでも前述の関連研究[1)]と同様に，特に手先や足先の追跡が比較的安定で抽出精度の人体姿勢依存性が低い色領域 (blob) を考える．

3.1 事前準備

本システムの動作条件としては以下を想定する．

・室内の静的環境下 (固定光源，静止背景，カメラ固定)
・非測定者に関し，洋服 (長袖シャツ，ズボン，靴下) の色は既知であり，肌色となるべく異なる色の服装

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

したがって事前に獲得しておく情報としては，多視点カメラに関する幾何学的キャリブレーションデータ，背景画像，背景差分のための閾値，および洋服，肌色に関する色情報である．色の識別については以下の方法で行う．

（a） 色パラメータの学習

人体領域に相当するblobの色情報としてパラメータ表現を考える．実画像よりシャツ，ズボン，靴下の色および肌色の領域の部分画像を切り出し，それをもとに色パラメータを学習する．基本的な色パラメータ学習の原理は，文献4)で提案された方法による．すなわち，画素のr,g,bを以下のように明度iを媒介変数とした2次関数で表現し，その係数パラメータを学習する．

$$\hat{r}(i) = R_2 i^2 + R_1 i \qquad \hat{g}(i) = G_2 i^2 + G_1 i$$
$$\hat{b}(i) = B_2 i^2 + B_1 i \qquad i = (r+g+b)/3 \qquad (1)$$

各色の学習用画像におけるm個の画素(r_j, g_j, b_j)を用いて，6つのパラメータ，つまり，係数$R_1, R_2, G_1, G_2, B_1, B_2$を各色$r,g,b$について，以下の式により線形最小二乗推定する．

$$\begin{bmatrix} r_1, & g_1, & b_1 \\ & \vdots & \\ r_m, & g_m, & b_m \end{bmatrix} = \begin{bmatrix} {i_1}^2, & i_1 \\ & \vdots \\ {i_m}^2, & i_m \end{bmatrix} \begin{bmatrix} R_2, & G_2, & B_2 \\ R_1, & G_1, & B_1 \end{bmatrix} \qquad (2)$$

（b） 色の識別

ある(x,y)位置の画素の明度iおよび，各色ごとに学習した色パラメータ$R_1, R_2, G_1, G_2, B_1, B_2$を用いて，モデルとなる色$\hat{r}(i), \hat{g}(i), \hat{b}(i)$を学習時と同様に計算する．対応色であるかの識別基準としては以下の式を用いる．

$$error = (r - \hat{r}(i))^2 + (g - \hat{g}(i))^2 + (b - \hat{b}(i))^2 \qquad (3)$$

以下の条件を満たす時，(x,y)の画素がその対応色であると判定する．

$$(i)\ low < i < high, \quad (ii)\ error < error_th \qquad (4)$$

ただし，$high, low$はそれぞれ最高明度の閾値，最低明度の閾値である．また，$error_th$は色識別の閾値である．後述の実験では$high$を230，lowを10，$error_th$を500としている．(i)の基準であらかじめ不安定なデータを取り除き，(ii)の基準で色の類似度を判定する．

3.2 2-D blob追跡

ここでは画像上に観測される肌色領域を顔・両手のblob，靴下や靴などの色領域を足のblobと仮定する．またシャツに相当する上半身領域を胴体のblobと仮定する．blob追跡のアルゴリズムは以下のようになる．まず，背景差分結果に基づき外接矩形を抽出し，同時に低解像度での色識別，胴体の重心推定を行う．次に，その矩形内における識別画素についてk-mean法によるクラスタリングを行い，顔，両手および両足の色領域クラスタを追跡する．ここで用いているクラスタリングの規準は，1) 色の類似度，2) 前フレームでの推定重心位置との距離，である．領域がどの2-D blobであるかの対応は初期設定時，あるいは追跡失敗時に決定する．特にエラー処理はオンラインのアルゴリズムとしては最も重要な部分のひとつであるが，ここでは簡単に，カメラに対し正面を向いており直立の自然な姿勢状態にあるとして，以下のようにヒューリスティックに対応を決定している．

・頭部は外接矩形の上部に近い肌色領域である
・左手(足)は外接矩形の右側に近い色領域である
・右手(足)は外接矩形の左側に近い色領域である

3.3 3-D blob復元

少なくとも2視点から2つの観測点が求められれば，3次元位置を三角測量の原理によって一意に復元できる．しかし，隠蔽により2視点のみでは必ずしも観測できない場合が起こるため，安定に3次元位置を推定するには，より多くの視点についての情報を融合することが必要となる．そこで本手法では多視点情報の融合を以下のように実現する．

多視点情報の選択 信頼性に基づいて統合する視点情報を各blobごとに選択する(信頼性については後述)．

視線ベクトルの計算 各視点ごとのキャリブレーション情報を用いて，カメラ座標原点および2-D blob重心を通る視線ベクトルを求める．

多視点情報の統合 選択した視線を用いて各blobの3次元位置復元計算を行う．

（1） 多視点情報の選択

多視点情報を融合するためには，どの視点の2-D blobを用いるかが重要である．一般に，人体部位に相当するblobが正しく追跡されているかを決定することは容易ではないので，本論文では2-D blobの信頼性を評価するものとして，以下のような画素数による可視性の評価値Rを用いる．

$$R = \begin{cases} N_s & (N_s \geq N_{min}\text{の時}) \\ 0 & (\text{それ以外}) \end{cases} \qquad (5)$$

ここでN_sは探索範囲内の2-D blobと判定された画素の数，N_{min}は2-D blobの存在を決定する最小画素数である．

（2） 多視点情報の統合

後に述べる選択規準により得られた視線$T_j (j = 1, \cdots, J)$を用い，それら視線の交点を求める．実際はノイズなどの影響でこれらの直線はねじれの位置にあることが多い．そこで，これらの直線からその最小距離の点を求めることで近似する．実際この方法では3次元空間での誤差は比較的大きいとされるが，仮想空間にマッピングする対象物情報の精度としては充分であり，また，2-D blobの抽出精度が高くないことからもこのような近似計算で充分である．

最も信頼性の高い視線が，$\mathbf{T}_1 = \mathbf{o}_1 + t_1 \mathbf{d}_1$と表現さ

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

れる時 ($\mathbf{o}_1$は最も信頼性の高い視線を選択した視点のカメラ座標原点，$\mathbf{d}_1$はその単位視線ベクトル)，また，その他の選択された信頼性の高い視線 ($j = 2, \cdots, J$) が，$\mathbf{T}_j = \mathbf{o}_j + t_j\mathbf{d}_j$($t_j$は媒介変数，$\mathbf{o}_j$はカメラ原点，$\mathbf{d}_j$は単位視線ベクトル) と表現される時，最も信頼性の高い視線上にあり，後者の視線群から最小距離である点 $\mathbf{T}$ は，t_1 をパラメータとして求めると以下のようになる．

$$\mathbf{T} = \mathbf{o}_1 - \frac{\sum_{j=2}^{J}(\mathbf{d}_1 \times \mathbf{m}_j, \mathbf{o}_1 \times \mathbf{m}_j - \mathbf{n}_j)}{\sum_{j=2}^{J} ||\mathbf{d}_1 \times \mathbf{m}_j||^2}\mathbf{d}_1 \qquad (6)$$

ただし

$$\mathbf{m}_j = \frac{\mathbf{d}_j}{\sqrt{1+||\mathbf{o}_j \times \mathbf{d}_j||^2}} \quad , \quad \mathbf{n}_j = \frac{\mathbf{o}_j \times \mathbf{d}_j}{\sqrt{1+||\mathbf{o}_j \times \mathbf{d}_j||^2}}$$

この点 $\mathbf{T}$ が 3-D blob 位置 $(T_x, T_y, T_z)^T$である．なお，選択された視線数が 2 よりも小さい場合は，前フレームまでに推定した 3-D blob 位置を保持する．

視点選択規準としては，2 次元の信頼性だけでなく復元される 3 次元情報に関しての妥当性についても考慮されたものである方が望ましい．したがって，各 3-D blob について視線を選択するアルゴリズムは以下のようになる．

（1） 最も信頼性の高い視線，すなわち信頼性 Rの値が最大の視線をまず決める．そして信頼性 Rが 0 でないものを選択候補とする．

（2） 次に，より精度良く 3 次元復元計算を行なうため，以下の規準により，(1) で選択された視線のうち最も信頼性の高い視線との間に成立するエピポーラ制約を満たすもののみをさらに選択する．

- 最も信頼性の高い視線と各視線との間で，ねじれの関係にある 2 視線間の距離 (点 $\mathbf{T}_1$とその視線間距離)D_jを計算し，以下の基準を満たすものを最終的に選択する．

$$D_j < vl_{3D}, \qquad \text{ただし } vl_{3D} \text{ は閾値} \quad (7)$$

- 視線ベクトル $\mathbf{d}_j$と $\mathbf{T}_{1,j} - \mathbf{o}_j$ の内積が以下の条件を満たす．

$$(\mathbf{d}_j \cdot (\mathbf{T}_{1,j} - \mathbf{o}_j)) > 0 \qquad (8)$$

ここで，$\mathbf{T}_{1,j}$は 1，2 の 2 視線を用いて (6) 式より求まる 3 次元位置である．

4. 人体モデルの動き生成

4.1 推定 blob の人体モデルへのあてはめ

知覚モジュールにより得られるのは胴体重心・頭部・両手・両足の計 6 つの 3D blob 位置 (以後，これを知覚データと呼ぶ) のみである．輪郭解析などを行い肘・膝などの関節位置を推定する方法[12)13)]もあるが，不安定な結果しか得られず推定可能な条件も厳しくなる．したがって，限られた少数の知覚データを用いて人体モデルの姿勢を生成するため，以下の逆運動学 (インバースキネマティクス) を用いてこの問題を解決する．なお人体モデルの表現としては，知覚データより求まる位置に両膝・両肘を加えた部位数 14，自由度 23(詳細は後述) を考えた多関節構造モデルとする (図 2)．

4.2 実時間逆運動学

（1） 従来の手法

逆運動学問題は一般に，ロボティクス，コンピュータグラフィクス[14)] の分野で用いられる．解法のアプローチとしては，解析的に解く方法[15)]，数値解析で解く方法[16)]があり，また，ヤコビ法，最適化法などが提案されている．また，腕のみであるが，実際の人体の動きデータを収集しそれらから導出した線形関係式を用いるユニークな方法などもある[17)]．ロボティクスおよびその関連分野では，生体に近い腕の到達運動など研究されているが，実際は動力学と併用されることが多く，質量，弾性など対象に関する多くの事前情報が必要とされる．一方，コンピュータグラフィクスの分野でも近年，実時間での逆運動学問題解法が開発されつつある．しかしながら，利用用途がモデリングツールであり，ユーザとのインタラクションを前提としているため，時系列データを必要としない最適化手法を用い，局所解に陥った場合はユーザが修正するようなアプローチが用いられている．

（2） 提案手法

我々の目的においては逆運動学に関し，上述の一般的手法と異なる以下の性質が望まれる．

- 実時間で両腕両脚の計 4 つの接続リンクに関する逆運動学について同時に解が得られる
- 知覚モジュールにより入力されるゴール (3-D blob) 位置は精度が悪いということを前提とする
- 連続的かつ自然な人体の動きを与える

そこで本論文では，上記の性質を考慮したアプローチとして解析的に解く方法を提案する．まず逆運動学におけるゴールとして様々なものが考えられるが，我々のような目的の場合，知覚モジュールから入力される 3-D blob 位置が正確でないことが起こりうる．特に，ゴール位置が解の存在範囲を越えて設定された場合，連続性が満たされなくなる．よって，実際のゴールを方向およびその距離 (絶対値) と定義することで対処する．つまり，ゴールの方向は必ず満たし，その距離は設定ゴール位置が解の存在範囲内である時のみ 3-D blob 位置に一致させる (図 2の左上参照)．一般に位置姿勢表現はロール・ピッチ・ヨーなどの回転・並進自由度が用いられ，それらが多関節構造の相互依存関係にあるため，このままでは解析的に解くことは困難である．そこで解析的に解けるよう自由度を適切に分ける方法が必要である．

（3） 解　法

2 つの可変リンクに関する運動学方程式を以下のように定義する．

$$\begin{aligned}\mathbf{g}_i = \ & \mathbf{T}^b\mathbf{R}^b(0, {R_y}^b, {R_x}^b)\mathbf{R}^{b'}({R_z}^{b'}, 0, 0) \\ & \mathbf{T}^{l_1}\mathbf{R}^{l_1}({R_z}^{l_1}, {R_y}^{l_1}, 0)\mathbf{R}^{l'_1}({R_z}^{l'_1}, 0, {R_x}^{l'_1}) \qquad (9) \\ & \mathbf{T}^{l_2}\mathbf{R}^{l_2}(0, 0, {R_x}^{l_2})\ \mathbf{t}^e\end{aligned}$$

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

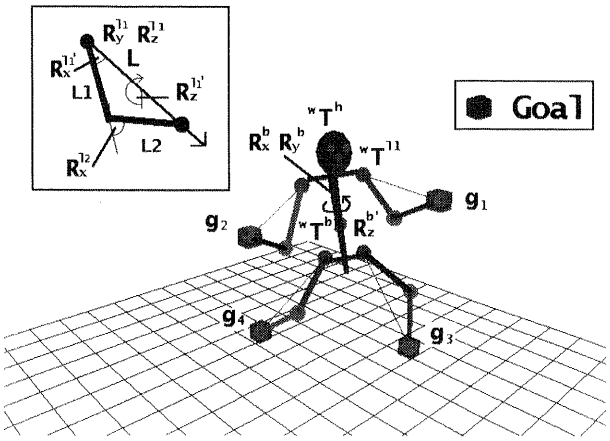

図 2　人体モデルの構造およびゴール方向の定義
Our human model geometry and goal definition.

図 3　知覚データに対する逆運動学適用姿勢の例
Our inverse kinematics responses for several perceptual data.

ここで，$\mathbf{g}_i = (g_x, g_y, g_z, 1)^T$ は，ゴール $i(1,\cdots,4)$ を表す位置ベクトル，$\mathbf{T}^b$ および $\mathbf{R}^b$，$\mathbf{R}^{b'}$ は胴体に関する相対並進，回転行列，$\mathbf{T}^{l_1}$ および $\mathbf{R}^{l_1}$，$\mathbf{R}^{l'_1}$，$\mathbf{T}^{l_2}$ および $\mathbf{R}^{l_2}$ はリンク 1，2 に関する相対並進，回転行列である ($'$のものは解析的に解くために分けられた回転行列を表す)．$\mathbf{t}^e$ はリンク 2 の末端のゴール位置への局所座標系での並進ベクトルである．また $\mathbf{R}(R_z, R_y, R_x)$ の回転自由度 R_z, R_y, R_x はそれぞれロール・ピッチ・ヨーを表す．

すると，逆運動学方程式の解析解としては，その幾何学的な性質より

$$
\begin{aligned}
{R_y}^{l_1} &= -\arccos\left(\frac{g_z - {}^wT_z{}^{l_1}}{||\mathbf{g} - {}^w\mathbf{T}^{l_1}||}\right) \\
{R_z}^{l_1} &= -\arctan\left(\frac{g_y - {}^wT_y{}^{l_1}}{g_x - {}^wT_x{}^{l_1}}\right) \\
{R_x}^{l'_1} &= \arccos\left(\frac{{L_1}^2 + L^2 - {L_2}^2}{2L_1L}\right) \\
{R_x}^{l_2} &= \arccos\left(\frac{{L_1}^2 + {L_2}^2 - L^2}{2L_1L_2}\right) - \pi \\
&\quad (|L_1 - L_2| \leq L \leq L_1 + L_2)
\end{aligned}
$$

が得られる．ここで，L_1, L_2, L はそれぞれリンク 1，リンク 2 の長さ，リンク 1 の原点からゴールまでの距離を表し，${}^wT^*$ はワールド座標を表す (図 2参照)．

また，胴体の方向に関しては，胴体位置 ${}^w\mathbf{T}^b$ および頭部位置 ${}^w\mathbf{T}^h$ により ${}^w\mathbf{T}^h - {}^w\mathbf{T}^b$ の方向に体軸が平行になるように決定する．

$$
\begin{aligned}
{R_x}^b &= -\arcsin\left(\frac{{}^wT_y{}^h - {}^wT_y{}^b}{||{}^w\mathbf{T}^h - {}^w\mathbf{T}^b||}\right) \\
{R_y}^b &= -\arctan\left(\frac{{}^wT_x{}^h - {}^wT_x{}^b}{{}^wT_z{}^h - {}^wT_z{}^b}\right)
\end{aligned}
$$

なお，${R_z}^{l'_1}$ は (9) 式により，両腕両脚のそれぞれに設定されるパラメータであり，腕に関しては，肘の上げ下げ，脚に関しては膝の開き具合に対応する，より人間らしい動きを与えるための回転角であり，これらは逆運動学の解には直接影響を与えないが，人体の姿勢表現には必要なパラメータである (以後，特性角と呼ぶことにする)．また，${R_z}^{b'}$ は人体の向き (パン) を定める回転角であり，これも逆運動学の解には直接影響を与えない，人体の姿勢表現には必要なパラメータであるが，人体の向き (パン) を 6 つの blob 位置のみから推定することは難しいため固定値にしており，その自動推定法の開発は今後の課題である．また，肘・膝の ${R_z}^{l'_1}$ の決定法については実験 5.1 で考察する．

本手法の特徴をまとめると以下のようになる．

- 解析的な方法であるため実時間計算可能
- 少数の知覚データから人体モデルの姿勢を定める目的に特化しているため 2 つの可変リンクのみ適用可能
- 逆運動学の解には直接現れないが，人体の姿勢には影響を与えるパラメータを設定できるため多くの動作表現が実現可能

5. 実 験 結 果

5.1 逆運動学の適用実験

(1) 知覚データへの適用

いくつかの姿勢の知覚データに対し，本論文で提案する逆運動学を適用した結果が**図 3** である．解析解であるため，知覚データに対応する手先・足先の位置はゴールに完全に一致する．ただし，この図の例では，より人間らしい動作を生成するための角 ${R_z}^{l'_1}$ については，両肘・両膝の 2 次元位置を人間が指定し、それをもとに 3 次元位置を自動計算した上で，推定したモデルの肘・膝の 3 次元位置がそれらと最も近くなる値を 1 次元探索により推定している．

(2) 特性角 ${R_z}^{l'_1}$ の実データによる計測

肩・肘・手の位置に合計 3 個のマーカをつけた右腕の動きをモーションキャプチャシステム[8]により 1000 フレーム計測し，その肘の位置から実際の人間の多様な動きに対して特性角 ${R_z}^{l'_1}$ がどのような値をとるのかを推定する実験を行った．まず，右肩の位置に対する右手の相対位置ベク

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

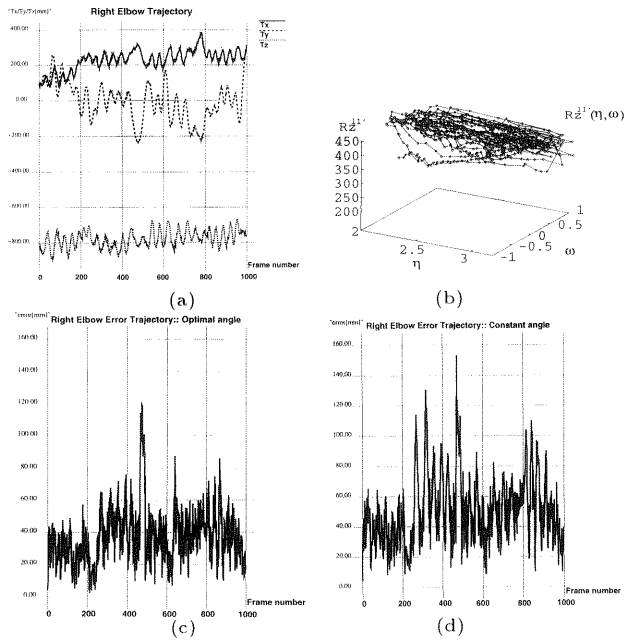

図 4 (a) 実際の肘の位置の軌道 (b) 右腕の動きデータに対する特性角の推移．ゴール方向を緯度，経度 (単位はラジアン) とし，対応する特性角をプロットしたもの．(c) 特性角を推定した場合の肘の位置の誤差 (d) 特性角を定数 (0) にした場合の肘の位置の誤差

Characteristic angle trajectory which was estimated by employing real right arm motion capture data.

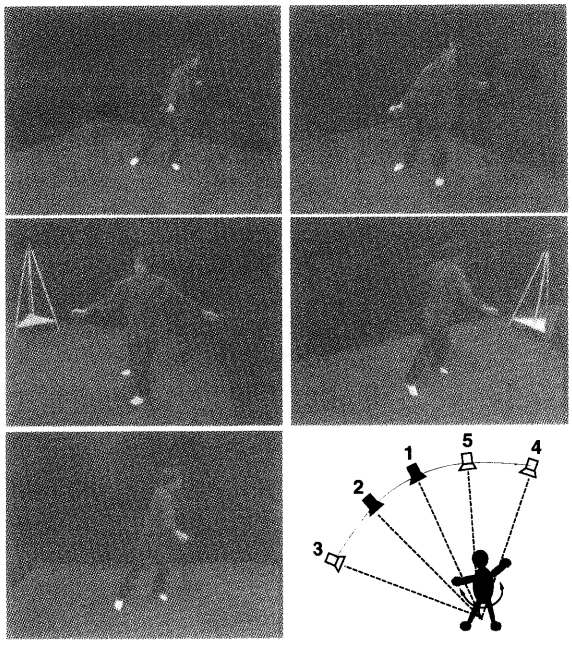

図 5 5 視点入力画像およびそのカメラ配置

Partial 5 input multiviews and camera allocation.

トルをゴール方向として逆運動学を適用する．そして，角 $R_z{}^{l'_1}$ が逆運動学の解に依存しないことを利用して，モデルの肘位置が実際の肘位置に最も近くなる角 $R_z{}^{l'_1}$ を 1 次元探索することで求めた．

図 4(b) は，図 4(a) に示すような腕の動き (手先をらせん状に回転させる動作) を，モーションキャプチャデータより得られる様々なゴール方向に対し推定した $R_z{}^{l'_1}$ を時系列データに沿って連続的にプロットしたものである．ここで，ゴールはその方向を緯度，経度の 2 パラメータで表現している．図 4(b) のグラフにおいて，様々なゴール方向に対し $R_z{}^{l'_1}$ は連続的でほぼ一定の値をとっており，定数にしても充分人間らしい動きを再現することが可能であることが予想される．入力に用いた実際の肘の位置と $R_z{}^{l'_1}$ の探索により推定した肘の位置との間の誤差のグラフを図 4(c) に示す．この場合の誤差平均は 4cm 程度であるが，これは用いたモーションキャプチャデータの精度によるものであると思われる．一方，$R_z{}^{l'_1}$ を定数とした場合の肘の位置の誤差のグラフを図 4(d) に示す．図 4(c) には及ばないにしても，誤差がある程度抑えられているといえる．また，さらに多くのゴール方向に対応する $R_z{}^{l'_1}$ を実際の人体動きデータより収集し，方向と $R_z{}^{l'_1}$ の関係式を導くことで，より人間らしい動きを再現するための値を自動的に決定することが可能であると考えている．

5.2 多視点情報の融合実験と考察

ここでは，多視点情報の融合の妥当性を検証するために，人体のおおよその向きが既知である場合を想定し，共通の視野に対し互いにオクルージョンによる影響を補い合う効果を検証する．**図 5** に示すオフラインで獲得した 5 視点画像系列 (1000 × 5 フレーム) に対し，多視点融合法を適用した．カメラは人体正面を捉えるカメラを規準にほぼ均等に円弧上に配置し，対象の動作は，両手を交互に左右に伸ばす体操であり，規準カメラ (1, 2) でオクルージョンが起こるような内容を仮定した．**図 6**(a)-(c) が左手の肌色 3-D blob 位置に関する追跡結果のグラフである．図中の実線が 5 視点融合時を示し，破線が特定の 2 視点 (1, 2) のみ用いた場合を示している．グラフより多視点融合した場合の方が 2 視点の場合に比べ，オクルージョン時の推定誤りに相当するスパイクが小さいことがわかる．これは冗長性の利用による比較的安定な追跡が達成できたことを示している．一方，図 6(d) は，5 視点融合時の軌跡の一部を拡大したものに融合視点数を加えたグラフである．このグラフから融合する視点数が変化する際に推定結果がやや不安定になっていることがわかる (スパイク状の小さなノイズ)．これは最小二乗推定時の融合情報の少なさに起因すると思われ，対策は今後の課題である．

5.3 オンライン実装およびアプリケーション例

ここでは，以上で述べたモーションキャプチャシステムを実際に実時間・オンラインシステムとして構成し，仮想現実アプリケーションに適用した結果を示す．

本手法を，視覚に基づく仮想空間と実空間の実時間インタラクションシステム (Visually Guided 3D Animation) として実装した (**図 8**)．具体的には，avatar として仮想空間上にモーションキャプチャにより再現されたユーザ (被計測者) が，仮想空間上のサッカーボールを実時間で蹴る，

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

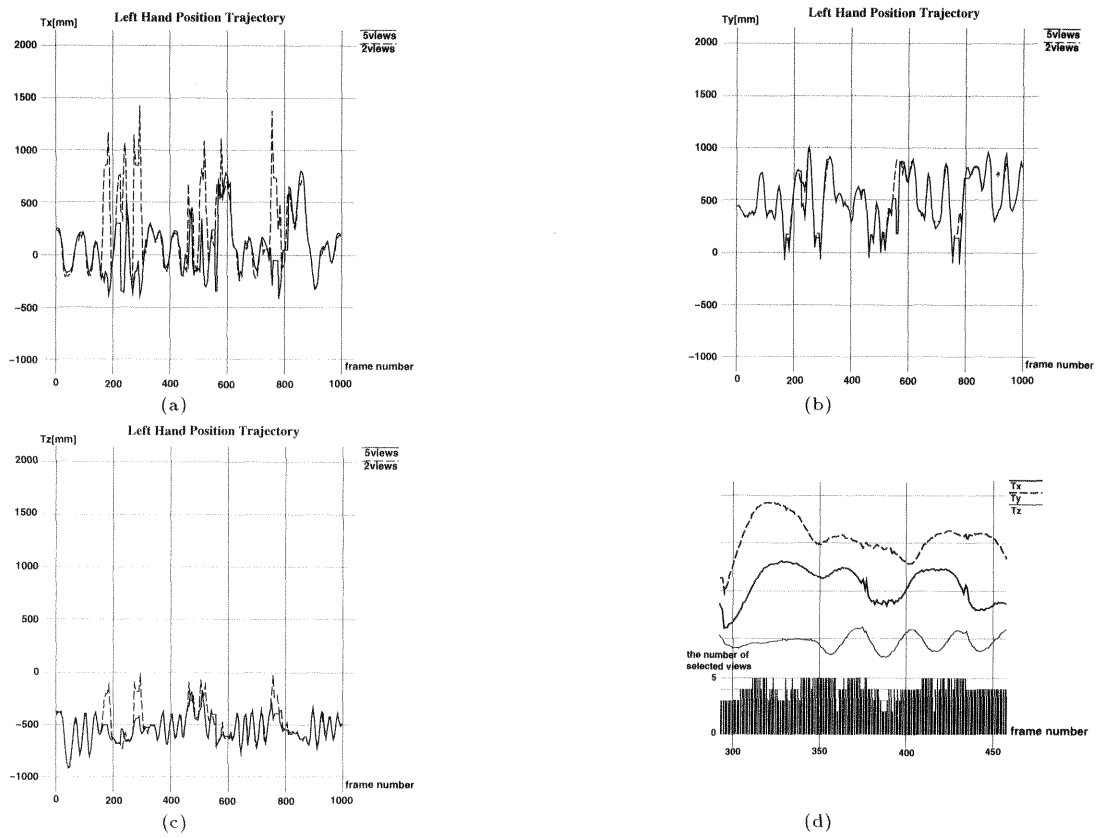

図 6 多視点情報の利用による知覚データの追跡結果 ((a)T_x, (b)T_y, (c)T_z) の比較および (d) 視点数の変化とその融合時の影響.

3-D left hand tracking results: 5 Multiview fusion.

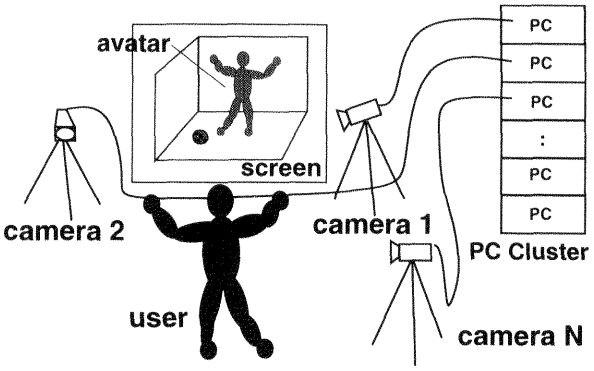

図 7 オンライン・実時間モーションキャプチャシステムの構成
Our online/real-time motion capture system setup.

という例題にした．実空間上のユーザの蹴った足先の位置と仮想空間上でのボール位置との距離により衝突判定を行い，衝突と判定された際には仮想空間上のボールが跳ね返るというリアクションを実現している．このシステムは，現在，PC クラスタの PC 数の制限により，カメラは 2 台，PC は 6 台 (図 1参照) を用いて実装し，実時間 (30fps) かつオンラインで動作している．なお，パイプライン処理による遅延時間は 0.2 秒程度であり，我々の目指す実時間多視点動画像処理システム性能としてはまずまずの結果である．これは処理モジュールの高速化によりパイプラインの段数を縮めることや，パイプライン 1 段あたりの処理時間を小さくすることで対処する予定である．

6. むすび

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を提案し，知覚データに対する実時間逆運動学の導入，および多視点融合の適用について述べた．本システムはオンラインかつ実時間で動作しするため，仮想空間とのインタラクションなど様々なアプリケーションへの適用が可能である．

今後の課題としては，知覚データに対する力学的あるいは感性的な動作フィルタリング[2)]によるより人間らしい動作の生成および人体モデルの高自由度化，多視点情報を利用した知覚モジュールの高精度化などが挙げられる．

〔文　献〕

1) C.Wren, A.Azarbayejani, T.Darrell, A.Pentland: "Pfinder: Real-Time Tracking of the Human Body", IEEE Transaction on Pattern Analysis and Machine Intelligence, **19**, 7, pp.780-785 (1997)
2) 鵜沼, 武内: "コンピュータアニメーションにおける感情を伴った人間の歩行動作の生成方法", 信学論誌, **J76-D-II**, **8**, pp.1822–1831 (1993)
3) 石井, 望月, 岸野: "人物合成のためのステレオ画像からの動作認識法", 信学論誌, **J76-D-II**, **8**, pp.1805–1802 (1993)
4) 岡本, R.Cipolla, 風間, 久野: "定性的運動認識を用いたヒューマンインターフェイス", 信学論誌, **J76-D-II**, **8**, pp.1813-1821 (1993)
5) 松山隆司: "分散協調視覚 – 視覚・行動・コミュニケーション機能の統合による知能の創発–", MIRU'98, TP1-1 (1998)
6) S.Yonemoto, N.Tsuruta, and R.Taniguchi: "Tracking of 3D

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

図 8 入力画像およびオンラインデモの画面．仮想空間上のボールを蹴る様子．
Online demo: real-virtua soccer.

Multi-Part Objects Using Multiple Viewpoint Time-Varying Sequences", Proc. ICPR98, pp.490-494 (1998)

7) C.Bregler and J.Malik: "Tracking People with Twists and Exponential Maps", Proc. CVPR98, pp.8-15 (1998)

8) S.Yonemoto, N.Tsuruta, and R.Taniguchi: "A Real-time Motion Capture System with Multiple Camera Fusion", Proc. ICIAP'99, pp.600-605 (1999)

9) R.Y.Tsai: "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Transaction on Robotics and Automation, **3**, No.4, pp.323-344 (1987)

10) D.Arita, N.Tsuruta and R.Taniguchi: "Real-time Parallel Video Processing on PC-cluster", SPIE-3452, Parallel and Distributed Methods for Image Processing II, pp.23-32 (1998)

11) R.Taniguchi and D.Arita: "A Basic Framework of Real-Time Image Processing on PC-cluster", 2nd International Workshop on Cooperative Distributed Vision, pp.119-132 (1998)

12) Y.Azoz, L.Devi, R.Sharma: "Reliable Tracking of Human Arm Dynamics by Multiple Cue Integration and Constraint Fusion", Proc. CVPR, pp.905-910 (1998)

13) 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)

14) Welman,C.: "Inverse Kinematics and Geometric Constraints for Articulated Figure Manipulation", MSc Thesis, CS, Simon Frazer University (1993)

15) D.Tolani: "An Inverse Kinematics Toolkit for human modeling and simulation", PhD Thesis, University of Pennylvania (1998)

16) J.Zhao, N.Badler, "Inverse Kinematics positioning using nonlinear programming for highly articulated figures", Transactions, on Computer Graphics, **13**, 4, pp.313-336 (1994)

17) Y.Koga: "Planning Motions with Intentions", Proc. of SIGGRAPH'94, pp.24-29 (1994)

米元 聡 1995年，九州大学工学部情報工学科卒業．現在，九州大学大学院システム情報科学研究科博士後期課程在籍中．主として コンピュータビジョン応用に関する研究に従事．

有田 大作 1992年，京都大学工学部情報工学科卒業．1997年，九州大学大学院総合理工学研究科情報システム学専攻博士課程修了．1998年，同助手．主として 並列動画像処理システムに関する研究に従事．

谷口倫一郎 1978年，九州大学工学部情報工学科卒業．1980年，同大学院工学研究科修士課程修了．同年，同大学院総合理工学研究科助手．1988年，同助教授．1996年，同大学院システム情報科学研究科教授．画像処理，コンピュータビジョン，並列処理に関する研究に従事．工学博士．正会員．

NII-Electronic Library Service