多視点動画像解析による人間の実時間姿勢推定

星野, 竜也
九州大学システム情報科学研究院知能システム学部門

米元, 聡
九州産業大学情報科学部設置準備室

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

出版情報：電子情報通信学会技術研究報告. IE, 画像工学. 101 (455), pp.1-8, 2001-11-14. 電子情報通信学会
バージョン：
権利関係：

社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
IE2001-109 (2001-11)

多視点動画像解析による人間の実時間姿勢推定

星野 竜也† 米元 聡†† 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知能システム学専攻, 福岡県
†† 九州産業大学情報科学部設置準備室, 福岡市

E-mail: †{hoshino,arita,rin}@limu.is.kyushu-u.ac.jp, ††yonemoto@ip.kyusan-u.ac.jp

あらまし 鉛直上方及び水平方向の計6台のカメラを用いて人間の3次元位置・姿勢を実時間で求める手法について述べる。本研究では、人物領域のシルエット輪郭情報を利用して手、足、頭に相当する人体の特徴点を追跡し、複数視点の観測結果を統合し、各部位の3次元特徴点位置を計算する。更に、これらの求まった知覚データから肘や膝の3次元位置も実時間で推定し、人間の姿勢推定としている。

キーワード 実時間モーションキャプチャ, 多視点画像処理, 視点選択, 関節位置推定

Real-time human pose estimation using multi-view image analysis

Ryuya HOSHINO†, Satoshi YONEMOTO††, Daisaku ARITA†, and Rinichiro TANIGUCHI†

† Department of Intelligent Systems, Kyushu University 6–1, Kasuga-koen, Kasuga, Fukuoka
†† Kyushu Sangyo University 2–6–1, Matsukadai, Higashi-ku, Fukuoka

E-mail: †{hoshino,arita,rin}@limu.is.kyushu-u.ac.jp, ††yonemoto@ip.kyusan-u.ac.jp

Abstract This paper presents real-time human motion analysis based on silhouette contour analysis. Our purpose is to develop a computer-vision-based human motion analysis system, which can be applied to human-machine interaction via human gestures. Since vision-based human motion capturing systems are essentially unstable and can only acquire partial information because of self-occlusion, we have introduced a robust pose estimation strategy, which can estimate human postures from limited perceptual cues such as positions of a head, hands and feet. In this paper, we outline a real-time and on-line human motion capture system and demonstrate a simple interaction system based on the motion capture system.

Key words Real-time motion capture, Multi-view image analysis, View selection, Joint position estimation

1. はじめに

人間の動作や行動を実時間で認識し、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、更には人間型ロボットの遠隔操作などの多くのアプリケーションへの適用が期待できる。人間の動作を実時間で観測する技術として、光学式、磁気式のセンサ接触型方式が用いられているが、これらは推定精度としては高いものが得られる反面、一般にシステムが高価になりすぎることや、対象の動きに制限を課してしまうといった問題が生じる。一方、ビデオカメラを利用したコンピュータビジョンの応用による非接触型方式としては、*Pfinder*[1]が既に提案されており、色情報を用いて2次元blob特徴を抽出して実時間で人間の姿勢推定を行っている。しかしながら色情報を用いるため、適用できる人間の服の色の制限や精度の点で問題がある。より詳細に動きを推定する方法としてモデルベース解析法もあるが[3][4]、計算量が膨大なため、オンラインでの動作には課題が残っている。対象の色情報ではなく輪郭による形状情報を用いて人間の全身動作を推定する方法としては、W^4[2]が提案されており、シルエット領域のテンプレートマッチングと人物の頭の位置の推定から、複雑な姿勢にも対応している。また、*Shall We Dance?*[5]では、ヒューリスティックなルールと三眼視を用いて、従来輪郭情報を元にした処理では推定不可能だった姿勢にも対応した3次元位置推定法を提案している。

本論文では、実時間、オンラインでの動作の実現を重要視して、多視点からの輪郭線解析により実時間で全身の動作を計測し、再生するシステムの構築を目指す。また、鉛直カメラと冗長な視点数を利用して、従来のシルエット解析では推定不可能だった姿勢への対応、及び人物の鉛直軸周りの回転にも対応したシステムを目指す。2章では、まずシステムの概要について述べ、3章で画像から人物特徴点を推定する方法を述べる。4章では多視点による冗長性を考慮した3次元姿勢推定法について述べ、5章で提案手法の実験結果と評価を行う。最後に本研究のまとめと今後の課題を示す。

2. システムの概要

2.1 ハードウェア

一般に画像に含まれるデータ量は多いため、画像を処理するには膨大な計算量を必要とする。特に本手法のように多視点からの画像を実時間で処理するためには、極めて高速な計算システムが必要となる。この問題を解決するため、我々は複数のPCを高速ネットワークにより結合したPCクラスタを用い、システムを構築した。本論文で用いたPCクラスタはPentiumIII 700MHzプロセッサを二つ持った8台のノードPCから成り、各PCは高速ネットワークMyrinetにより相互結合されている。また、カメラにはIEEE1394デジタルカメラを用いている[10]。

2.2 処理の流れ

図1にPCクラスタにおける処理モジュールの配置を示す。システムはカメラの台数に応じた並列処理を行なう構成になっている。本手法では、対象人物を水平方向から見るカメラを複数台と、対象人物を鉛直上向きから見るカメラ1台を配置する。各カメラは同時に撮影を行う必要があるため、カメラを接続したPC全てへ撮影開始のメッセージをブロードキャストし、メッセージの到着後に撮影を開始することでカメラの同期をとっている。

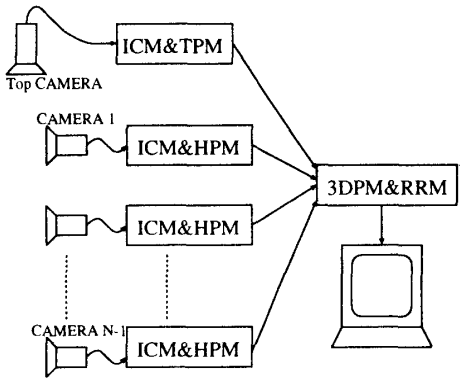

図1 PCクラスタにおける処理モジュールの配置

- **2次元画像特徴抽出モジュール (HPM)**

同期信号に応じて水平カメラから画像を取り込み、獲得した画像を処理して人間の特徴点を抽出するモジュールであり、水平カメラの台数分存在する。処理の結果は3次元復元モジュールへ送られる。

- **鉛直カメラ画像特徴抽出モジュール (TPM)**

鉛直カメラで獲得された画像を処理して人間の特徴を抽出するモジュールであり、HPM同様に、処理の結果は3次元復元モジュールへ送られる。

- **3次元復元モジュール (3DPM)**

各HPM及びTPMから送られてきた2次元画像処理の結果の全てを統合して人物の3次元位置の推

定を行なう。また推定結果を用いて、仮想空間に実時間でCGキャラクタを再生する。

3. 輪郭線解析

各カメラで獲得した画像から対象(人間)の特徴点を推定する手法の概要について述べる。一般に人間の特徴点(頭頂・左右手先・左右足先など)は輪郭線上に現れていると言えるので、本手法では人間の身体領域の輪郭線情報を用いる。

まず、シーンに対象が存在しない画像を背景画像としてあらかじめ獲得しておき、背景差分処理により画像中から人物領域を抽出する。次に得られた人物領域の重心座標を計算する。そして求まった重心位置より真上に向かった時の境界線との交点を開始点として、時計周りに人物領域の輪郭線の検出を行う。検出された輪郭点は開始点から順に2次元座標値をリストとして保持しておく。ここまでの処理はHPM・TPM共通である。

3.1 水平カメラ画像の処理

水平カメラで獲得された画像から求める特徴点は頭頂・左右の手先・左右の足先である。一般にこれらの特徴点は人間の取り得る様々な姿勢において、輪郭線上の尖った点(以後凸点と呼ぶ)として現れる。そこで、輪郭線上の凸点を求めるために、以下の特徴量を計算する。

$$Lv(i) = |\vec{s(i)}^2| + |\vec{g(i)}^2| \qquad (1)$$

ここで、$\vec{s(i)}$は輪郭点リストのi番目の点から輪郭の開始点へのベクトル、$\vec{g(i)}$は輪郭リストのi番目の点から人物領域の重心へのベクトルである(図2参照)。人間の取りうる様々な姿勢においてこの$Lv(i)$の極大点と輪郭線上の凸点が良く一致することが知られている[7]。

一般に、この(1)式から極大点を求めるだけでは、各極大点が人間のどの特徴点に対応しているかまでは知ることができない。そこで本手法では、「前フレームと現在フレームではシルエットの形状はあまり変化しない」と仮定し、前フレームの推定結果を利用することにする。すなわち前フレームで各特徴点を与えたインデックス付近で(1)式の極大点を探索することにする。なお、初期フレームにおける各極大点と各特徴点の対応付けについては、初期姿勢(両手を広げてやや足を広げた直立姿勢)を仮定し、輪郭線上に特徴点の現れる順序をシステムにあらかじめ与えることで解決する。

ただし、ここで述べた特徴点が輪郭線上に現れないことも考えられる。例えば手先が胴体の部分に重なっている姿勢では、手先はカメラから見えているにも関わらず原理的にシルエットから手先を求めることは不可能であり、本システムの問題点である。

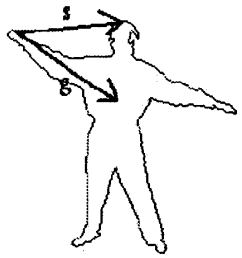

図2 $Lv(i)$の定義

3.2 鉛直カメラ画像の処理

TPMで求めるのは、左右の手先の位置、体の回転角、そして両肩の座標である。まず、両手先については、計算を簡単にするために輪郭線の曲率を用いて手先の推定を行う。前フレームで求めた各手先の付近に存在する輪郭線上で曲率が最大になる点を現在フレームの特徴点(手先)とする。

次に体の回転角については、人間のシルエットの重心周りの主軸を計算する。

$$grad = \frac{m + \sqrt{m^2 + 4M^2(1,1)}}{2M(1,1)} \qquad (2)$$

$$M(p,q) = \sum_x \sum_y (x - G_x)^p (y - G_y)^q I_{xy} \qquad (3)$$

ただし$m = M(0,2) - M(2,0)$である。$M(2,0)$と$M(0,2)$は2次モーメント、$M(1,1)$は慣性相乗モーメントと呼ばれている。ここで、(G_x, G_y)は重心の座標、I_{xy}は座標(x,y)の画素値で、0または1を取る。

ところで、慣性主軸による体の回転角は手の広げ方に影響を受けやすい。例えば両手を前に突き出した姿勢では求まる体の回転角は前に突き出した両手に平行なものになってしまう可能性がある。そこで本手法では、モーメントを計算する前に、対象画像に対してOpeningの処理を行い、腕の部分の画素の除去を行っている。Openingに用いるStructure Elementは$\{(0,0),(1,0),(-1,0),(0,1),(0,-1)\}$である。また、Openingの繰り返し回数は経験的に7回とした(注1)。Openingにより腕の画素を除去した画像につ

(注1):繰り返し回数については、常に一定回数で良いのか、画像により回数を変えていくべきか等、現在調査中である。

いて再度人物の重心を計算し、それから主軸を求める。こうすることにより、手の広げ方に影響を受けにくい、安定した人物の向きの推定が可能となった。

なお、肩の座標については、画像上での人物の肩幅を定数としてあらかじめ与えておくことにより、体の回転角と重心の位置から容易に求めることができる。

4. 3次元姿勢推定法

4.1 視点選択法

本手法では様々な人物の姿勢への対応のため、鉛直カメラを含む、冗長な視点を持たせたが、一般には2視点の視線ベクトルが求まれば、ステレオ視の原理に基づき、特徴点の3次元位置を求めることが出来る。3DPM 部では、TPM の解析結果を用いてステレオ視するのに有効な視点二つを以下のようにして選択している。

まず、TPM で求めた体の回転角を元に、複数の水平カメラの中から人物が向いている方向に最も近いカメラ (Front Camera) を二つと、逆に人物の側面に最も位置するカメラ (Side Camera) を一つ選択する (図 3)。具体的には、人物領域の重心から各カメラへのベクトルと、人物の視線ベクトルの内積を計算する。内積が最大になるもの二つが正面カメラとして選択され、内積が最も小さくなるものが側面カメラとして選択される。

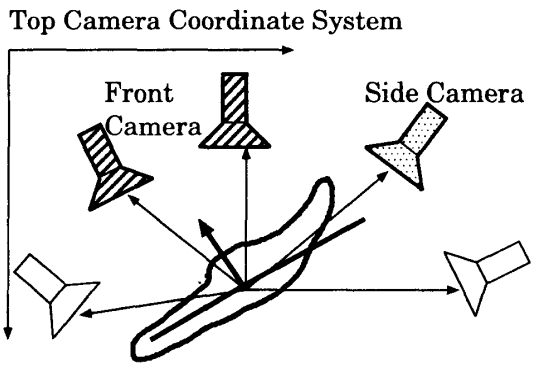

図 3　視点選択法

次に、ステレオ視する2視点を選択する。頭・胴・両足先に関しては、前述した二つの正面カメラでステレオ視が行われる。ただしここで、胴体の3次元位置の復元は人物領域の重心のステレオ視によるものである。一方、両手先に関してだが、まず、鉛直カメラ座標系において、人物領域の重心を通り人物の視線方向の直線と鉛直カメラで求めた手先の位置との距離を鉛直カメラ座標系で算出する。この距離と、鉛直カメラから見た人間の肩幅の長さ (ほぼ一定と仮定する) との比較により、更に視点選択を行う。直線と手先との距離の方が肩幅よりも大きい場合は、正面カメラ二つでステレオ視を行い、逆に小さい場合は、正面カメラにおいて手先が輪郭線上に特徴点として現れなかったものと判断し、正面カメラではなく鉛直カメラと、側面カメラで求めた手先でステレオ視を行う。

4.2 肘と膝の推定について

4.1章で述べた手法で求まる3次元位置は頭・胴体・左右の手先・左右の足先のみであり、人間の全身推定のためには、まだ肘と膝の位置の推定が必要となる。これまで我々は逆運動学を解析的に解く手法を提案してきた [8]。この手法は手先や足先の位置だけの少ない知覚データのみから肘・膝の位置を実時間で求めることが可能であるが、逆運動学は原理的に解が一意に定まらず、場合によっては人間の姿勢としてふわさしくない解が求まることがあった。

そこで本論文では、逆運動学に代る手法として、シルエット形状と肩 (股)・手先 (足先) の位置から肘 (膝) の位置を推定する方法を提案する。ここでは肘の位置の推定について述べるが、膝の位置も同様に推定することができる。

まず前提条件として、上腕及び下腕の長さは既知で定数であるものとする。この時、もしも手先と肩の3次元座標が分かったならば、肘の位置として取りうる値は図 4に示す円上のいずれかに限定されることになる。そこで、円上のある点のワールド座標を求め、これを各カメラの画像平面に投影し、人物シルエットとの相関を調べれば良いことになる。

ここで、$\vec{A} = (A_x, A_y, A_z)$、肩の座標を (x_1, y_1, z_1) とすると、円の半径 r は、

$$r = \frac{\sqrt{4|\vec{A}|^2(l_1^2 + l_2^2) + 4l_1^2 l_2^2 - (l_1^2 + l_2^2 + |\vec{A}|^2)^2}}{2|\vec{A}|} \quad (4)$$

$$T_C = (x_1 + A_x t, y_1 + A_y t, z_1 + A_z t) \quad (5)$$

となる。ただし、$t = \sqrt{\frac{l_1^2 - r^2}{|\vec{A}|^2}}$ である。

肘位置推定の具体的な手順は以下の通りである。

（1） **肘位置推定のためのローカル座標系の設定**

T_C を原点とし、$\vec{A}$ を X 軸、ワールド座標系の Z 軸を円の乗る平面へ投影した直線を Z 軸とするローカル座標系を定義する。これにより肘推定の円がローカル座標系の Y-Z 平面となるため、肘の推定処理が

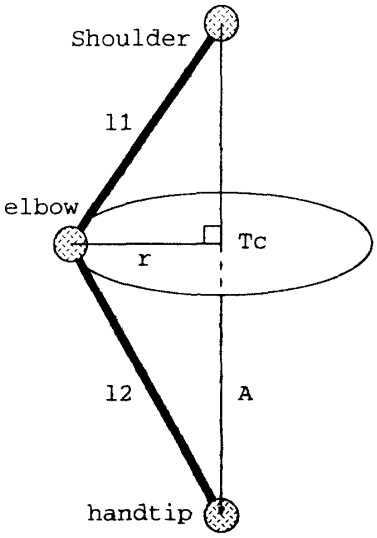

図4　肩・手先、及び肘の位置関係

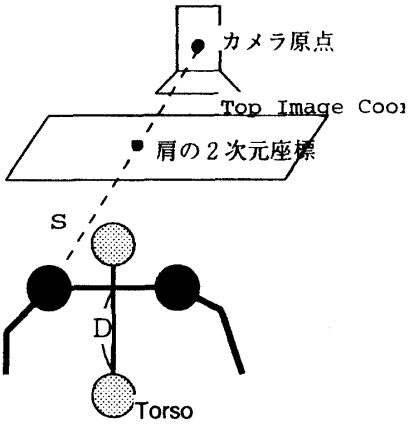

図5　肩のワールド座標の求め方

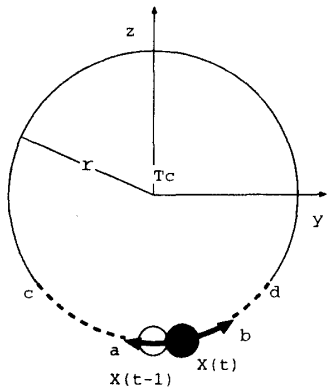

図6　肘の位置の特定法
○：t-1フレーム目での関節位置
●：tフレーム目で推定される関節位置

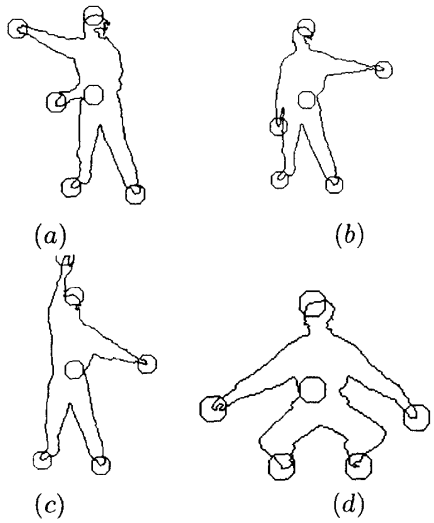

図7　人体特徴点の推定結果の一例

簡単になる。またこのローカル座標系からワールド座標系への変換も容易である。

（2）**肩位置の推定**

TPMにおいて求めた画像中の肩の2次元座標とカメラキャリブレーション情報より、カメラ原点と鉛直画像上の肩を通る直線の式が求まる(図5の破線s)。そして、胴のz座標(注2)に一定量Dを加算した値を肩のz座標とすれば、肩の3次元座標を求めることができる。今回はD=400mmとした。

（3）**肘位置の推定**

tフレーム目での関節位置$X(t)$を推定する様子を図6に示す。(t-1)フレームでの推定結果$X(t-1)$を元に、関節位置の探索範囲を弧cdに限定し、探索範囲上にあり、かつ各画像平面へ投影した時に、その投影位置が全てのシルエットに含まれるような関節位置候補を求める（図6の弧ab)。関節位置は弧abの中点として近似できると考え、$X(t)$を弧abの中点として求める。尚、初期フレームについては従来の逆運動学を用いて仮の肘位置を求め、同様にして肘位置を探索する。

(注2)：シルエット領域の重心位置を胴の位置としている。

5. 実　　験

本章の実験において利用する装置は焦点距離4mmレンズを装着したCCDカメラ6台(約30°間隔に配置した5台の水平カメラ、及び鉛直方向におよそ平行に配置した鉛直カメラ1台)である。なお、これらのカメラはお互いの位置関係がわかるように予めキャリブレーションしておくものとする。

まず、輪郭線解析による人間の2次元特徴点推定の結果画像の一部を図7に示す。輪郭、及び頭頂点、胴体重心、左右手先、左右足先の推定位置を円で示している。これによると、各特徴点が十分精度良く追跡できていると言える。本手法では、基本的に人

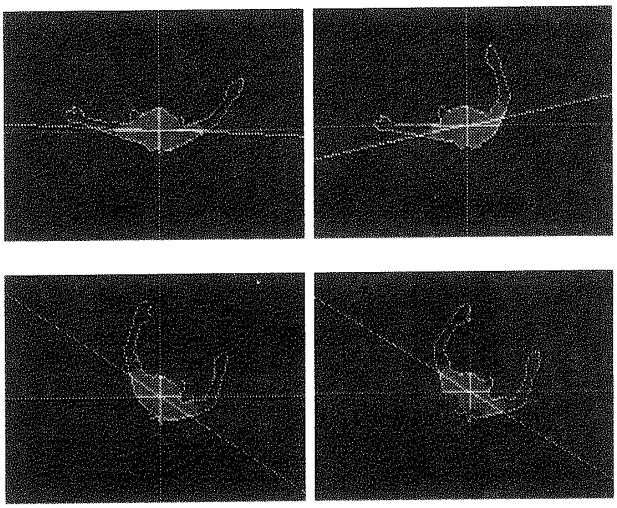

図 8　TPM における画像処理結果の一例

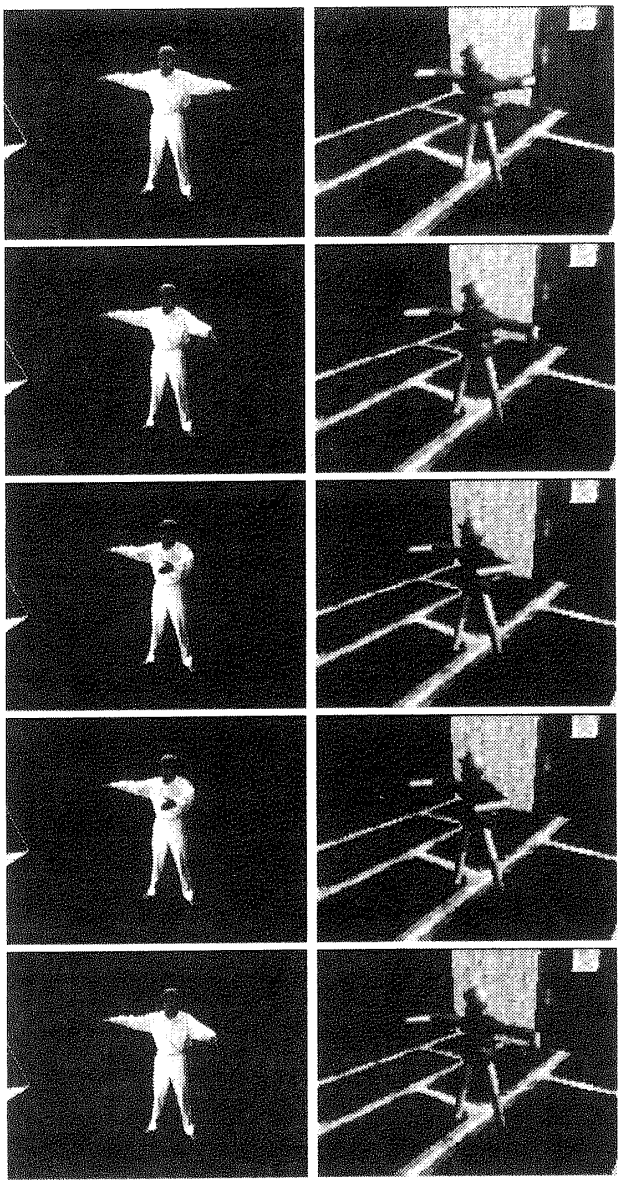

図 9　手が胴に重なる姿勢を含む動作の 3 次元推定結果

間の直立に近い姿勢のみ推定可能であることを想定していたが、(d) のように、しゃがんだ姿勢でも各特徴点追跡が可能であった。又、(a) のように輪郭線上に各特徴点が現れる順番が変わった場合でも追跡可能なことを確認した。もちろん手先などが胴に重なって輪郭線上に現れなくなった場合には追跡は不可能であるが、その後再び輪郭線上に特徴点として現れてから遅くとも 5 フレーム後までには再び検出できることが確認できた。これは単眼の推定結果としては十分なものと言えるであろう。

図 8に Opening 処理を行った画像と主軸を計算した結果を示す。腕の部分の画素がほぼ除去され、胴体部分の画素はほとんど残っており、両手を前へ突き出した姿勢であってもほぼ正確に体の向きが求まっていることがわかる。

本論文では、手先が胴の前に重なっている場合のように、従来シルエット解析手法では推定不可能であった姿勢を冗長な視点数を持たせ、視点選択を行うことで推定可能にすることを目標としている。そこで、左手を真横から胴の前へ手先が弧を描くように動かし、また元の位置へ戻すという動作を行った時の推定結果を検証した。図 9に左手を含めた 3 次元姿勢推定結果の一部を示す。各画像は視点 2 の方向から再現したＣＧで、上からそれぞれ 0, 23, 42, 68, 92 フレーム目の推定結果である。図 9に示されているように、本論文で提案した、多視点を用いてステレオ視に使う視点を動的に選択することで、従来シルエット解析法では推定不可能だった姿勢でも本手法では推定可能であることが分かった。しかしながら、推定結果が不安定な場合もあり、これは主に鉛直カメラにおける体の回転角の推定誤りによるものと思われ、今後改良する予定である。

次に全く同じ入力画像に対して、左肘の推定に逆運動学を用いた場合と、本論文で提案した手法を用いて推定した場合の結果を比較した (図 10)。図の上段は入力画像、中段が逆運動学による推定結果、下段がシルエットを利用した提案手法による結果である。また、左側が上方視点で、右側が水平方向視点である。上方視点、水平方向視点のいずれの場合においても、再構成された CG 人形の左肘の位置は、逆運動学による方法より本論文による提案手法の方が入力画像に近い推定結果となっている。よって、本論文で提案した手法が逆運動学による手法に比べて、有効であると言える。

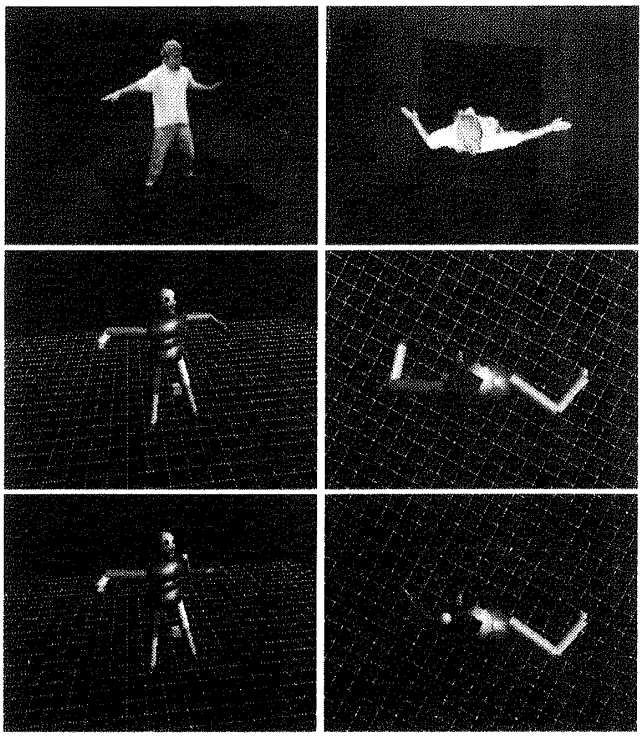

図10　逆運動学とシルエットを利用した提案手法の比較

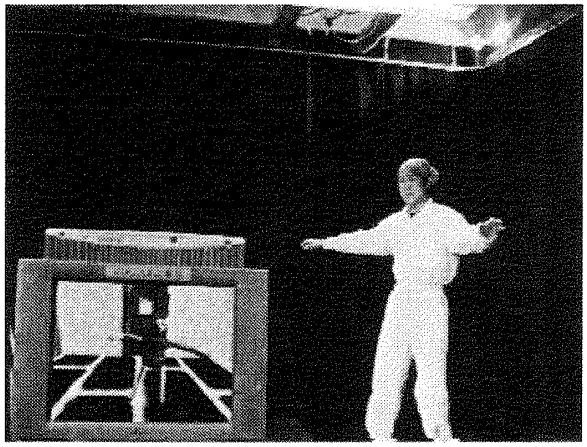

図11　6視点オンラインデモの様子

最後に本システムを6視点でPCクラスタ上に実装して、オンライン、実時間での実行を試みた。図11にオンライン実行中の様子を示す。各処理モジュールでかかった時間は、水平カメラ処理が約20msec、鉛直カメラ処理が約35～45msec、3次元復元モジュールが約26msecであった。したがって、本システムが十分に実時間(注3)かつオンラインで動作することが確認できた。

(注3)：現在使用しているIEEE1394カメラは外部同期を用いると、15fpsの速度でしか画像を取り込めないため、ここでは15fpsを実時間としている。

6. 結　　論

本論文では、多視点動画像の輪郭線解析することにより、人物の3次元姿勢を推定する方法を提案した。本論文で述べた手法をPCクラスタに実装したところ、以下のことが確認できた。

- 鉛直カメラを利用した多視点融合の際の動的視点選択を行うことで様々な姿勢や体の向きに対応が可能

- 輪郭線上に手先が特徴点として現れない場合の姿勢でも推定が可能

- 人物の体を回転させる動作に対しても、復元に使う視点を切替えて推定が可能

- シルエットを元に肘と膝の3次元位置を推定する方法が有効

- PCクラスタ上に実装して、オンライン実時間で十分動作すること

今後の課題としては、まず肘や膝の推定法の改善が挙げられる。今回は単純に関節位置のみを各画像に投影し、その投影位置が全てのシルエットに含まれる関節範囲の中央を解としたが、その他の推定方法との比較検討が必要だと思われる。現在検討している方法は、関節位置だけでなく手先や肩なども画像へ投影し、肩から肘、肘から手先への投影直線がシルエットに最も良く含まれる位置を推定位置とする方法等である。

次に人物の2次元特徴点を求めるのに、色情報も併用することが挙げられる。本論文で述べた手法では、色情報を全く使用することなく人間の3次元の姿勢推定を行うことが可能であるが、動作が不安定な場合もある。そこで、画像処理から特徴点を求める際に、本論文で提案した輪郭を用いると同時に色情報も用いることで、従来よりも高い推定精度を実現し、かつ着用する衣服の制限も緩和できるようなシステムの完成を目指したい。また、提案手法では求めることのできない姿勢、例えばしゃがみ込んだ姿勢などに対する推定方法についても検討する必要がある。

現在我々は、CG人形を実時間で仮想空間に描画し、仮想のサッカーボールを蹴るという簡単なアプ

リケーションを作成しているが、より複雑な人間とコンピュータのインタラクションについても検討して行きたい。

謝辞 本研究は，日本学術振興会科学研究費補助金 (学術創成研究「人間同士の自然なコミュニケーションを支援する知能メディア」課題番号 13GS0003) の補助を受けて行った。

文　献

[1] 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.

[2] I.Haritaoglu, D.Harwood, L.S.Davis, "W4: Who? when? where? what? a real-time system for detecting and tracking people", in *Proc. of the third Int. Conf. Automatic Face and Gesture Recognition*, pp.222–227, 1998.

[3] S.Yonemoto, N.Tsuruta, R.Taniguchi, "Tracking of 3D Multi-Part Objects Using Multiple Viewpoint Time-Varying Sequence" ,Proc. ICPR98, pp.490–494, 1998.

[4] C.Bregler, J.Malik, "Tracking People with Twists and Exponential Maps" ,Proc. CVPR98, pp.8–15, 1998.

[5] 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.

[6] D.Arita, N.Tsuruta, R.Taniguchi, "Real-time parallel video processing on PC-cluster", in *Proc. of SPIE-3452, Parallel and Distributed Methods for Image Processing II*, pp.23–32, 1998.

[7] 高橋, 坂口, 大谷, "三眼視による実時間非接触非装着型三次元人物姿勢推定法", 信学技報 PRMU99-94, pp.47–54, 1999.

[8] 米元, 有田, 谷口, "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論文誌, Vol.54, No.3, Vol.54, No.3, pp.409–416, 2000.

[9] Thomas, Moeslund, Erik, "POSE ESTIMATION OF A HUMAN ARM USING KINEMATIC CONSTRAINTS", 信学技報 PRMU98-265, pp.47–54, 1999.

[10] 吉本, 有田, 谷口, "1394 カメラを利用した多視点動画像獲得環境" , 第 6 回画像センシングシンポジウム SSII2000, pp.285–290, 2000.