

視覚認識による3次元人体動作の解析・生成とインタラクション

米元, 聡
九州産業大学情報科学部知能情報学科

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/2324/5810>

出版情報：映像情報メディア学会技術報告. 26 (9), pp.233-238, 2002-01-30. 映像情報メディア学会
バージョン：
権利関係：

H I R 2 0 0 2 - 5 8 , M I P 2 0 0 2 - 7 7 ,
A I T 2 0 0 2 - 7 7 , N I M 2 0 0 2 - 7 7 ,
V I S 2 0 0 2 - 7 7

視覚認識による3次元人体動作の解析・生成とインタラクション

米元 聡† 谷口倫一郎††

†九州産業大学

〒813-8503 福岡市東区松香台2丁目3番1号

††九州大学大学院システム情報科学研究院

〒816-8580 春日市春日講演6-1

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

あらまし 本研究は実空間と仮想空間との境界を意識させない、高度なインタフェースの実現を目指している。具体的には、視覚センサを用いて人体の3次元動作を非接触に計測・解析し、その動作を入力として、実空間上の人間と仮想空間上のイベントや物体との円滑なインタラクションを実現するシステムを開発している。視覚認識の結果を利用して仮想空間において人体動作を再現する場合、少数の部位の位置情報をもとにリアルな動作を生成することが必要となる。また、人にやさしいシステムであるために、システム側が人間すなわちユーザの意図した動作を把握できることが望ましい。これらの二点を考慮したシステムについて報告する。

キーワード 動作生成, アフォーダンス, インタラクション

Vision-based 3D Human Motion Analysis and Synthesis for Interaction

Satoshi YONEMOTO† and Rin-ichiro TANIGUCHI††

† Kyushu Sangyo University

Matsukadai 2-3-1, Higashi-ku, Fukuoka, 813-8503 Japan

†† Department of Intelligent Systems, Kyushu University

6-1 Kasuga-koen Kasuga Fukuoka, 816-8580 Japan

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

Abstract This paper describes a real-time interaction system which enables 3D direct manipulation. Our purpose is to do seamless mapping of human action in the real world into virtual environments. With the aim of making computing systems suited for users, we have developed a vision based 3D direct manipulation interface as smart pointing devices. Our system realizes human motion analysis by 3D blob tracking, and human figure motion synthesis to generate realistic motion from a limit number of blobs. For the sake of realization of smart interaction, we assume that virtual objects in virtual environments can afford human figure action, that is, the virtual environments provide action information for a human figure model, or an avatar. Extending the affordance based approach, this system can employ scene constraints in the virtual environments in order to generate more realistic motion.

Key words motion synthesis, affordance, interaction

1. はじめに

実空間と仮想空間とをシームレスに融合する高度なインタフェースを実現するために、実世界におけるユーザから仮想空間への情報入力を効率的に行うための装置や技法の開発が進められている。また、近年の高度情報化社会の振興、発展に伴い、個人レベルの生活環境においてもインターネットを介した情報収集など、高度な情報操作を要求されることが多くなってきている。人間中心の情報操作インタフェース機器としては、デスクトップ型、モバイル型の計算機などが普及しているが、操作性に関して見てみると、人間すなわちユーザがシステムに合わせる形で行われている。しかしながら、人間の活動は主として3次元操作であり、人間がシステムに合わせるのではなく、システム側が人間の直接的な操作に合わせるスマートなインタフェースの実現が望まれている。ユーザからの情報入力、なんらかのスイッチや、人間の動作入力装置を用いて間接的、あるいは直接的に行われる。特に、人間の3次元動作の直接的な入力を可能にするものには、モーションキャプチャに代表される3次元動作計測機器があり、装着するセンサ数を増やすことで詳細な動作を計測できるものまで開発され、実用レベルに達しているといえる。しかし、上述のスマートなインタフェースの実現という観点からも、なるべく装置の存在を意識させず[1]、装着、脱着の手間をとらない効率性を有し、しかもユーザの意図した動作をシステムが自動的に理解し仮想空間上で適切に再現できることが望ましいといえる。

本研究では、そのような直接的な3次元動作入力およびユーザの動作意図を考慮することができるインタラクションシステムの要素技術として、コンピュータビジョンによる非接触な動作計測、及び動作理解手法の開発を行っている。非接触での人体動作の計測は、効率やそのスマートさから仮想空間と実空間とのシームレス化に重要である[2][3]。視覚センサを用いた動作計測を行う場合、人体動作を精度よく推定することが一般に困難であるとされているが[4][5][6]、本研究ではこの問題に対処するために、比較的安定に推定可能な手や顔の動作情報から、リアルな人体モデルの動作を仮想空間上に生成することができる方法を提案する。

また、仮想空間と実空間での円滑なインタラクションを遂行するために、計測した3次元動作入力をもとにユーザの本来意図した動作をリアルに再現するための枠組みについても提案する。我々は、3次元動作

入力による3Dインタラクションを円滑に行うためにアフォーダンスの概念を利用する。仮想空間と実空間を融合し高度なインタラクションを実現するインタフェース構築のためにアフォーダンスの概念を利用することは、実世界で利用するのと同様、有用であると思われる。実世界において扱われるアフォーダンスの概念を応用し、仮想空間中のオブジェクト（仮想オブジェクト）がアバターすなわち3次元人体モデルに対し意味・価値のある情報、ここでは動作情報をアフォードするという捉え方である。つまり、仮想空間に存在する様々なオブジェクトにユーザの動作に関連する情報を事前情報として定義しておき、実空間でのユーザの動作に応じてその動作情報をアフォードすることで、リアルな動作の生成することができ、さらに、入力されるユーザの動作系列のみを単に処理するだけでは把握するのが難しいシーンコンテキストや、ユーザの意図を読みとることが可能であると考えられる。

2. 3次元人体動作の解析と生成

2.1 3次元位置の追跡

本研究では、人間の動作を画像より推定するための画像特徴として肌色領域を用いる。肌色領域としては、上半身では顔、両手部分が相当し、肌色という均一色の領域を人体部位とみなして観測する。2視点、あるいはそれ以上の複数視点の画像について同一部位の肌色領域が観測できる場合、ステレオ視の原理によってその領域重心の3次元位置を容易に計算することができる。領域重心位置を追跡するという問題においては、画像特徴の性質上、正確な3次元位置を推定することはあまり重要でないため、本研究では簡易な多視点ステレオ計算法[2]を用いる。対応のとれたそれぞれの視点における重心位置、及びカメラキャリブレーション情報より得られるカメラ座標原点を通過する直線の交点として重心の3次元位置を計算する。したがって上半身の動作においては、顔、両手の重心の3次元位置を求めることになる(図1(左))。

2.2 3次元人体動作の生成

仮想空間において、アバターすなわち3次元人体モデルの動作をユーザの動作に合わせてリアルに再現することは、インタラクションを遂行する際に正しく情報伝達を行うために必要不可欠である。そこで本研究では、画像解析により得られた少数の位置情報よりリアルな動作を生成するための姿勢推定法を導入する。リアルな動作を再現するには、身体構造系の厳密な物理特性に基づいて姿勢を計算することが望まれるが、

すべてのユーザに共通する物理特性を予め獲得しておくことも困難であり、一般に画像解析により求まる位置、速度などの物理特性の精度も悪い上、少数部位の位置情報しか用いることができないことから、厳密なダイナミクス系として3次元人体モデルの動作生成を行うことは難しいと思われる。そこで本研究では、画像解析により得られる位置入力だけをもとに、近似的ではあるが物理法則に基づいた動作生成を行う方法を導入する。

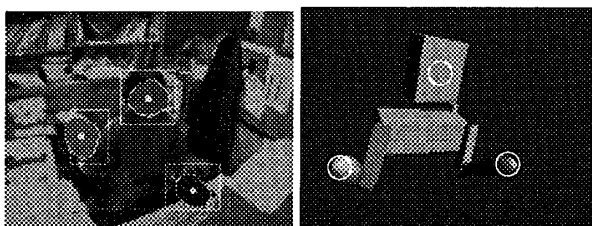


図1 上半身動作：(左)解析。(右)生成。

まず人体の骨格モデルに対し、物理的な制約として重力及びリンクの長さは一定で互いに接続するという多リンク系の制約を条件に、画像解析により得られる位置変位をもとに全リンクの端点、すなわち関節点位置の均衡する位置を推定する。その後、各関節点のペアについて各々の形状を表す部位モデルの当てはめ計算を行い、並進、回転パラメータを求める。この手法により、画像解析では求めることが困難な肩や肘の位置などを物理法則に従うという範囲でリアルに再現することが可能となる。

人体の骨格モデルという多リンク系について、隣接する関節点を結ぶリンクを強力なバネモデルであると仮定し、入力される位置変位が生じてもリンク長を一定に保つよう均衡することを考える。また、計算に必要な速度成分は位置変位から近似的に求め、加速度については陽に考慮しないものとする。

具体的には、各リンクの端点である関節点 i, j のペアについて、以下の計算を収束するよう繰り返す。

(1) $\mathbf{d} = \mathbf{p}_i - \mathbf{p}_j$ 及び $\mathbf{f} = \frac{\mathbf{d}}{\|\mathbf{d}\|} (\|\mathbf{d}\| - r_{ij})K$ を計算する。

ここで、 \mathbf{p}_i と \mathbf{p}_j は関節点の3次元位置であり、 r_{ij} は予め定義する $\mathbf{p}_i, \mathbf{p}_j$ 間のリンク長を表す。 K はバネモデルの弾性定数を表す。

(2) $\mathbf{p}_i, \mathbf{p}_j$ を(1)の \mathbf{f} に基づき、刻み幅 Δ_f だけ更新する。

$$\begin{aligned} \mathbf{p}_i &= \mathbf{p}_i + \mathbf{f}\Delta_f \\ \mathbf{p}_j &= \mathbf{p}_j - \mathbf{f}\Delta_f \end{aligned} \quad (1)$$

さらに、すべての関節点位置 $\{\mathbf{p}_k\}$ ($k = 1, \dots, N$) について、重力に従って刻み幅 Δ_g だけ更新する (\mathbf{g} は重力方向のベクトル)。

$$\mathbf{p}_k = \mathbf{p}_k + \mathbf{g}\Delta_g \quad (2)$$

この繰り返し計算により、全関節点位置 $\{\mathbf{p}_k\}$ ($k = 1, \dots, N$) が入力される位置変位に応じて均衡を保つ。ここではリンク長の制約、重力のみを考慮したが、衝突判定のための制約などの他の制約を式(1)、(2)と同様に考慮することが可能である。

図1(右)に本手法により生成した上半身姿勢の例を示す。図中の3つの丸印が動作入力として与える位置であり、この位置とモデルの状態によりすべての上半身の姿勢が決定される。

3. 視覚認識によるインタラクション

実空間からの動作入力による3Dインタラクションを円滑に行うために、アフォーダンスの概念を利用する。実世界におけるアフォーダンスの概念を応用し、仮想空間中のオブジェクトが人体モデルに対し意味・価値のある情報として動作情報をアフォードするという捉え方である。仮想空間中のオブジェクトとユーザとの間のインタラクションは、ユーザの動作に応じて、各オブジェクトに予め定義された動作情報がアフォードされることにより遂行される。例えば、物体オブジェクトは仮想空間において把持され、操作され、そして手放されることをアフォードする。ユーザの手の位置と物体オブジェクトの距離に応じて把持動作をアフォードすると同時に、手指の詳細な把持動作を自動的に補完生成するよう定義することも可能である。

まとめると、アフォーダンスの概念を利用することは、実空間からの人体動作入力としてユーザの本来意図したリアルな動作を生成するために以下の点で有用であると考えられる。

- 計測できない詳細動作をアフォードされる付加的な動作として追加できる
- アフォードされる動作情報として仮想空間上にシーン制約を課すことで、実空間ではありえない動作を抑制し、生成する動作をリアルにすることができる

前者の例としては、上述のように物体オブジェクトを把持する、すなわち把持されるとアフォードした際に手指の詳細な動作を自動的に生成することが挙げられる。後者の例としては、ドアの開閉のように物体の運動が一定の規則に拘束される場合、アフォードされ

るドアの開閉動作もその運動規則に従う、などが挙げられる。

本研究では、上述のアフォーダンスの概念、及びユーザからの動作入力に基づいて仮想空間と実空間を融合するインタラクションの様式として以下を実現している。

- 仮想空間上に置かれた静的な物体の把持、移動などの基本操作
- 動的オブジェクトとのインタラクション
- シーン制約を考慮したインタラクション
- 実物体を利用したインタラクション

3.1 静的・動的仮想オブジェクトの操作

ユーザは、仮想空間中に定義された、棚上に置かれたカップ、ティーポットのような静的な物体オブジェクトをアフォードする動作に応じて把持、移動などの操作を行うことができる。また、ユーザは擬人化エージェントのような動的なフィギュアオブジェクトとインタラクションすることができる。その場合、ユーザからの動作入力に応じたリアクションが予め定義されるものとする。

3.2 シーン制約を考慮したインタラクション

ユーザからの動作入力に厳密に従うよりは、仮想空間中のオブジェクトの状態及びアフォードする動作情報に応じて現在のシーンに適切な動作を生成する方がより自然な場合がある。オブジェクトの状態が、シーン制約として姿勢や形状変形のパラメータ \mathbf{q} で定義され、 $\mathbf{p}_o(\mathbf{q})$ がそのパラメータによって制限されるオブジェクトの位置、 \mathbf{p}_v がユーザの実際の入力位置であるとしたとき、 $\mathbf{p}_o(\mathbf{q})$ と \mathbf{p}_v の位置が十分近く、パラメータが定義された制限範囲内にあるという条件下では、 $\|\mathbf{p}_o(\mathbf{q}) - \mathbf{p}_v\|$ を最小にする \mathbf{q} が、リアルな動作を生成するのに適したパラメータということになる。このとき生成されるアバターの動作は、 \mathbf{p}_v ではなく、 $\mathbf{p}_o(\mathbf{q})$ に正確に従うことになる。

3.3 実物体情報を利用したインタラクション

物体オブジェクトの回転を行いたい場合など、実物体を利用することで人体動作のみでは実現しにくい操作を代行できると考えられる。例えば回転操作の場合は、実物体と仮想空間中の相当する回転用物体を対応させておくことで、実物体を回転させることにより仮想空間中に回転操作を入力することができる。また、視覚認識の容易な物体を実空間と仮想空間とを結ぶランドマークとして利用することで、計測範囲を制限し、他の実世界情報をシステム内に獲得することを容易に

するようなことも考えられる。

4. 実験

プロトタイプシステムとして、ノート PC(SONY Vaio PCG-R505R/GK) に IEEE1394 規格のデジタルカメラモジュール (SONY DFW-V500) 2 台を接続した、デスクトップ型の実時間インタラクションシステムを構築した。図 2 にシステムの概要及び 2 台のカメラからの入力画像の例を示す。このシステムはデスクトップ型であるため、上半身動作のみを計測、生成の対象とし、システムを動作させるスペースとしては $4m^2$ 程度の空間があればよい。

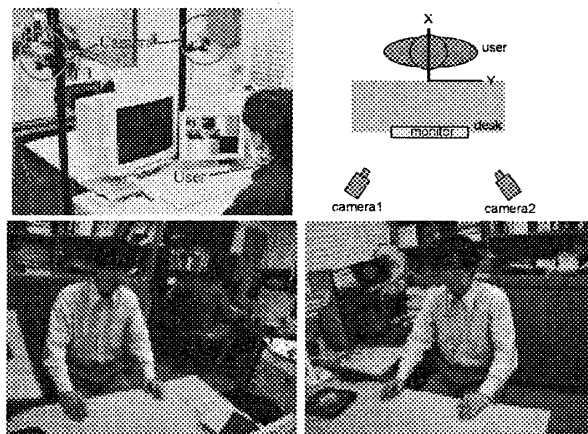


図 2 システムの概要: (上) 実験環境。
(下) 2 台のカメラからの入力画像。

4.1 静的・動的仮想オブジェクトの操作

図 3 に示すように仮想空間上の棚に置いたオブジェクトを、ユーザの動作に応じてアバターが把持、移動を行うことが可能である。実験は、アバターがティーポットからカップに水を入れ、飲む、そして元の場所に戻すという一連の動作を行ったものである。このシーンの例のようにインタラクションに関するインタラクティブなシナリオ、すなわち仮想空間上の物体の配置及びアフォードする動作情報の定義などをシステムに事前に定義しておく。シーンの変更や選択は、ユーザの動作入力をもとに容易に行うことができる。この例では、オブジェクトがアフォードする動作としては、把持される (2,3)、移動される、水を入れられる (4)、水を除かれる (7)、などであり、図 3 の番号順に動作が行われる。実際に水を入れる動作や飲むことによる付加的なリアクション動作の生成は省略した。また、この例では、各オブジェクトがシーンコンテキストによらず定義された動作をアフォードするだけであ

るが、ユーザの行った動作系列をたどることで、より複雑な動作理解が可能であると考えている。

また、動的オブジェクトとしてフィギュアオブジェクトである仮想ロボットを考え、簡単なインタラクションの例としてアバターとの握手動作を実現した(図4)。これは、仮想ロボット側に握手をアフォードするよう実装することで実現しており、ユーザの差し出す手の位置及び仮想ロボットの手の位置に応じて握手動作が行われる。

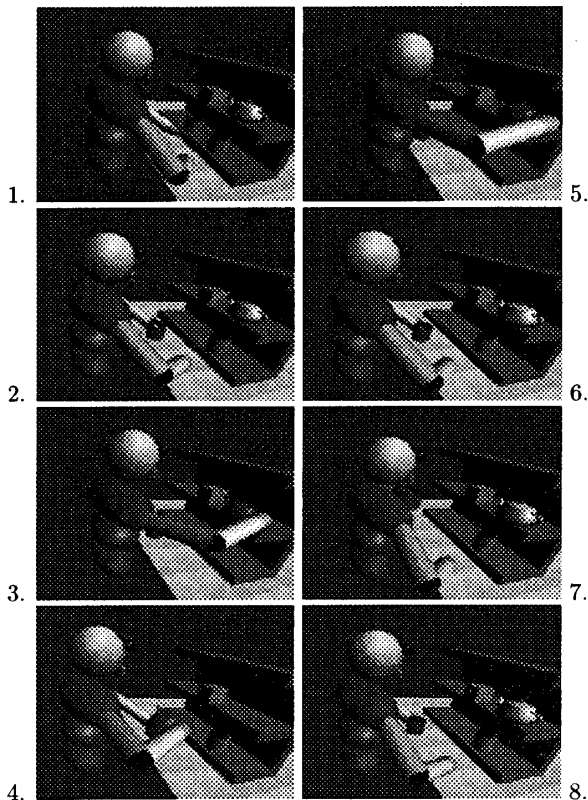


図3 シーン#1: アバターがティーポットとカップを操作する。

4.2 シーン制約を考慮したインタラクション

シーン制約を利用する例として、仮想空間上に設置したドアの開閉動作を考え、ドアの開閉動作はその運動軌道に正しく従うという制約を課した。図5に示すように、ドアの開閉に関する動作がユーザの手の位置の入力とドアノブの距離に応じてアフォードされ、アバターがドアの開閉を行い、棚の中の物体オブジェクトを取り出すシーンを想定した。

この実験では、ドアの状態はその回転軸まわりの回転角パラメータ R によってのみ変更されると仮定する。 R の範囲を $0^\circ \sim 160^\circ$ とし、ドアの円運動による軌道を $p_o(R)$ と表す。ドアの開閉動作は円運動に拘束さ

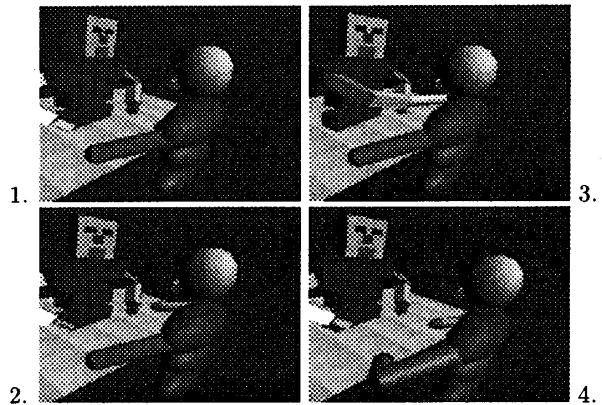


図4 シーン#2: アバターが仮想ロボットと握手する。

れることより、アバターの手の位置もその軌道上に再現されることになる。すなわちユーザからの手の位置の入力 p_v をきっかけに、想定するシーンにおいてより正確な動作が生成されたことになる。

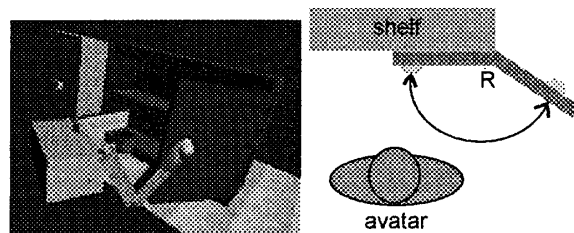


図5 シーン制約を考慮したインタラクション: ドアの開閉動作。

4.3 実物体情報を利用したインタラクション

実物体を利用することで、人間が直接入力することが難しい仮想空間中のオブジェクトの回転操作を実現した。実物体として、2つの識別用マーカを付与した平面パネルを用いた(図6(右))。ユーザが実物体を回転させることで、対応する仮想空間上のディスク及びその上に置いた仮想オブジェクト(dino)を回転することができ(図6(左))、同時に自らの動作も再生することが可能であった。

5. おわりに

システムが人間に合わせる事が可能な高度なインタフェースの構築及び円滑なインタラクションを実現することを目指し、視覚センサを用いて人間の3次元動作を非接触、直接的に入力することが可能で、しかも少数の入力情報からリアルな動作の生成が可能で3次元人体モデルの動作制御技術を開発した。また本研究では、アフォードンスの概念を利用することにより

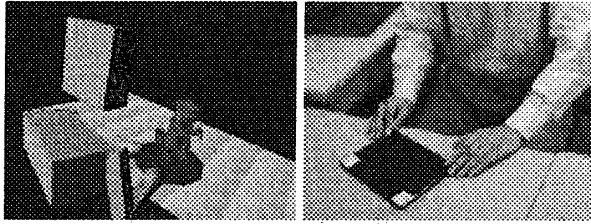


図6 実物体情報を利用したインタラクション: (左) 仮想オブジェクトの回転操作.
(右) 実物体の例.

シーン制約を考慮し、よりリアルなアバターの動作表現を行うことができる。今後、この枠組みを拡張し、ユーザの意図する高次な動作の理解とより現実感のある動作生成を実現する予定である。

謝辞

本研究の一部は科学研究費補助金特定領域研究(C)13224074「個人適応能力を有する高次マルチモーダルユーザインタフェースに関する研究」の補助を受けた。

文献

- [1] M.Weiser, "Some Computer Science Issues in Ubiquitous Computing", *Communications of the ACM*, Vol. 36, No.7, pp75-84, 1993.
- [2] Satoshi Yonemoto, Daisaku Arita and Rin-ichiro Taniguchi, Real-Time Human Motion Analysis and IK-based Human Figure Control, in *Proceedings of Workshop on Human Motion (HUMO2000)*, pp.149-154, 2000.
- [3] S.Yonemoto, R.Taniguchi, "High-level Human Figure Action Control for Vision-based Real-time Interaction", in *Asian Conference on Computer Vision*, 2002 (発表予定).
- [4] C.Wren, A.Azarbayejani, T.Darrell, A.Pentland, "Pfinder: Real-Time Tracking of the Human Body", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.780-785, 1997.
- [5] C.Bregler, "Learning and Recognizing Human Dynamics in Video Sequences", in *Computer Vision and Pattern Recognition*, pp.568-574, 1997.
- [6] D.Metaxas, "Physics-based Deformable Models Applications to Computer Vision, Graphics and Medical Imaging", Kluwer Academic Publishers, 1997.
- [7] C.Wren, A.Pentland, "Understanding Purposeful Human Motion", in *Fourth IEEE International conference on Automatic Face and Gesture Recognition*, 2000.
- [8] J.Kuffer Jr., "Autonomous Agents for Real-time Animation", *PhD thesis Stanford University*, 1999.