

少数動作入力をもとにしたアバターの動作制御とその応用

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/2324/5812>

出版情報：電子情報通信学会技術研究報告. ITS. 103 (639), pp.53-56, 2004-01-26. 電子情報通信学会
バージョン：
権利関係：

少数動作入力をもとにしたアバターの動作制御とその応用

米元 聡[†] 谷口倫一郎^{††}[†]九州産業大学情報科学部知能情報学科 〒813-8503 福岡県福岡市東区松香台2-3-1^{††}九州大学大学院システム情報科学研究院 〒816-8580 福岡県春日市春日公園6-1

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

あらまし 本研究は画像認識からの少数動作入力をもとにアバターの動作制御を行うシステムの開発を目指している。本システムではユーザの動作入力および仮想空間の状況に応じて詳細な動作を自動的に駆動することが可能である。上半身の動作を仮想空間上に提示するアプリケーションを想定し、少数の動作入力からどの程度の自然な動作が提示できるかを検証する。

キーワード 画像認識, Perceptual User Interfaces, 視点制御

Avatar Motion Control Using Human Body Motion

Satoshi YONEMOTO[†] and Rin-ichiro TANIGUCHI^{††}[†] Faculty of Information Science, Kyushu Sangyo University Matsukadai 2-3-1, Higashi-ku, Fukuoka-shi, 813-8503 Japan^{††} Intelligent Systems, Kyushu University Kasuga-koen 6-1, Kasuga-shi, 816-8580 Japan

E-mail: †yonemoto@is.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. We have evaluated the performance of virtual object manipulation by changing view angles and the target positions.

Key words Computer vision, Perceptual User Interfaces, Virtual camera control

1. 概 要

画像認識技術によりユーザの手や顔などの3次元位置をリアルタイムで獲得し、その動作入力をもとに仮想空間内でアバターとして身体動作を表現する方法について述べる。直接的な3次元動作入力をもとにした仮想環境とのインタラクションを実現するための要素技術として画像認識による非接触な3次元身体動作の計測手法を用いることは、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。しかし画像認識を動作入力に用いると、身体動作を精度よく推定することが一般に困難な課題であり[1][2]、比較的安定に得ることが可能な位置は少数に限られるため、少数位置をもとにアバターの身体動

作を表現する必要がある。そこで我々は少数の動作入力を基にして身体動作を再現する方法の開発に着目し、人体モデルを積極的に用いて安定な動作の生成を行う方法の開発に取り組んでいる。身体動作を増強する方法には、物理法則に代表される構造上の制約により補完する方法やシーンの文脈を利用して表現動作を決定する方法が考えられる。本研究では、ユーザは仮想空間において行為を行い、その行為をアフォードする環境とは仮想空間を意味する。仮想環境は完全に把握できるため、仮想シーンの文脈情報を動作の増強に積極的に利用できる。増強された動作を提示することでユーザは効率よく仮想物体操作などのタスクを遂行できる。

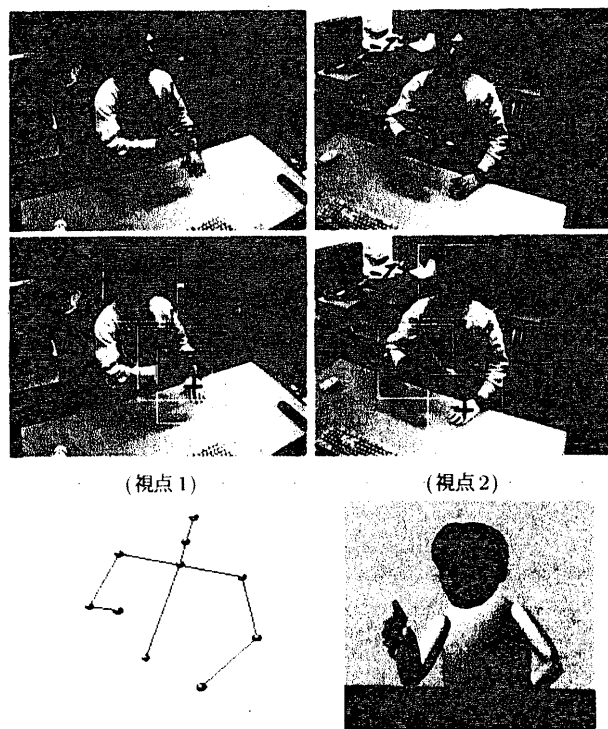


図1 (上) 入力画像と追跡結果 (下) 人体骨格モデルとアバター

2. 画像認識による身体動作の入力と動作生成

本研究では、身体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、肌色領域を身体部位とみなして観測する。2視点以上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次元位置を計算することができる[3]。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。前の時刻での計測位置を初期位置としてこれらの3次元位置の追跡処理を行う。図1(上)に用いる入力画像の例、肌色領域の追跡結果を示す。また、物体操作のための把持状態判定のために、推定した手の位置についてユーザの手の状態(把持状態、開放状態)の推定を行う。得られた少数の位置情報より自然な動作を生成するための姿勢推定法として物理法則に基づいた動作生成を行う方法を用いる。物理法則として身体構造の制約に強力な弾性定数のばねモデルを仮定し、位置入力の変位をもとに当てはめ計算を行うことで姿勢の更新を行う。以上の姿勢推定法により、基本的なアバターの動作の制御が可能となる。図1(下)に顔、両手の重心の3次元位置から再構成された人体骨格モデル、及び人体モデルの当てはめを行った例(アバター)を示す。

3. 仮想空間におけるアバターの動作制御

インタラクションを円滑に行うために、実世界において扱われるアフォーダンスの概念を仮想空間上へ応用する。環境に相当する仮想空間において、オブジェクト(仮想物体)が、人間に相当するアバターに対し意味・価値のある情報、ここでは動作情報をアフォードすると捉える。ユーザの身体動作入力をきっかけに各オブジェクトに定義された動作情報がアフォード

され、システムがそれらより適切なものをピックアップし、実際のアバターの動作やシーンイベントとして反映されることにより遂行される[4]。具体的には、仮想物体の特性に応じた手指の詳細動作や、ドアの開閉に伴うオブジェクトの変化への適応動作、関心のある仮想物体へ顔を向けるための視線制御などが挙げられ、これらは行われているインタラクションについてユーザへの提示効果を高めると考えられる。

4. 実験とシステムの評価

以下では、システムの評価に関する実験について述べる。

4.1 システムの概要

デスクトップ型のインタラクションシステムを構築した。画像認識用にカメラ(SONY VFW-V500)2台を用い、実空間上には提示用の大画面ディスプレイ、机、椅子を配置し、上半身のみが両カメラにより撮像される。1台の汎用PC(CPUクロック2GHz程度)で画像認識から仮想空間の提示処理までを行うことが可能である。本実験では遠隔での提示も考慮して画像認識用に汎用PC1台(Pentium4 2.6GHz)、仮想空間提示用にワークステーション(dual Xeon 3.2GHz)1台を用い、モーションデータ等をリアルタイム送受信させる構成にしている。図2にシステムの概観を示している。基本的には、ユーザが実空間で行う動作に合わせアバターが仮想空間で同様の動作を再現する。一般に全身の精度の高い動作の再現が必要なく、数箇所の位置に関する動作を獲得してアバターを駆動させる程度の目的であれば高価なモーションキャプチャを必ずしも必要としない場合がほとんどである。本システムは、簡易なアバターの身体動作の提示が必要な仮想環境のアプリケーションに有効であり、視点制御により方位角・ズームパラメータなどを制御することも可能である。別の利用方法として、ユーザの身体表現をアバターとして仮想空間上に陽には存在させず、位置情報の入力に応じて視点制御を行うことも可能である。ウォークスルー方式の視点制御による仮想空間ナビゲーションや物体操作を行うことができる。

4.2 アバター動作の増強

仮想空間における物体操作を例にアバターに本アプローチにより自動付加される詳細動作を検証する。物体操作としては仮想物体の把持・移動・開放を行う。ユーザの行う物体操作をリアルタイムに計測した(15fpsで動画像1000フレーム分)。(A)カップ、(B)ティーポット、(C)果物の3つの仮想物体を操作対象とした。図3はフレーム000, 077, 324, 332, 394, 431, 704, 999における仮想物体操作の様子の一部を示している。図4は物体操作における詳細動作が付加された時間帯を示すグラフである。“A: Aff”は仮想物体Aについて自動的に把持動作が駆動されたことを示し、“A: pik”は仮想物体Aについてアバターの状態が完全に把持動作に移行するタイミングを示す。図4のように、フレーム077, 394, 704, 942において把持動作に移行している。図3のフレーム324からフレーム332にかけて把持に関する動作情報がアフォードされ、手指の曲げ動作が自動生成されている。この動作の増強により物体を把持できる位置に手が移動していることを視覚的に確認することが

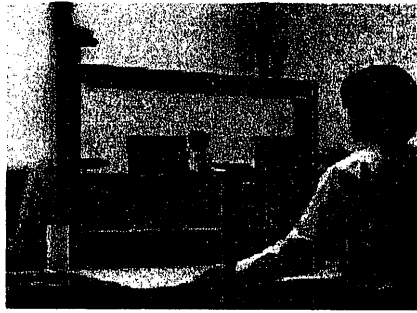


図2 システムの概観

できる。また、フレーム394からアバターの状態がピックアップ処理により把持動作に移行、つまり把持動作を行うモードに切り替わり、フレーム431のように、物体の状態も手に追従するよう変化する。詳細な動作を自動付加することにより、本来計測していないリアリティ情報を増強し、ユーザへの提示効果を高めている。

4.3 物体操作タスクの評価実験

本システムのアプローチでは、物体操作タスクにおいて以下の点が問題となる。

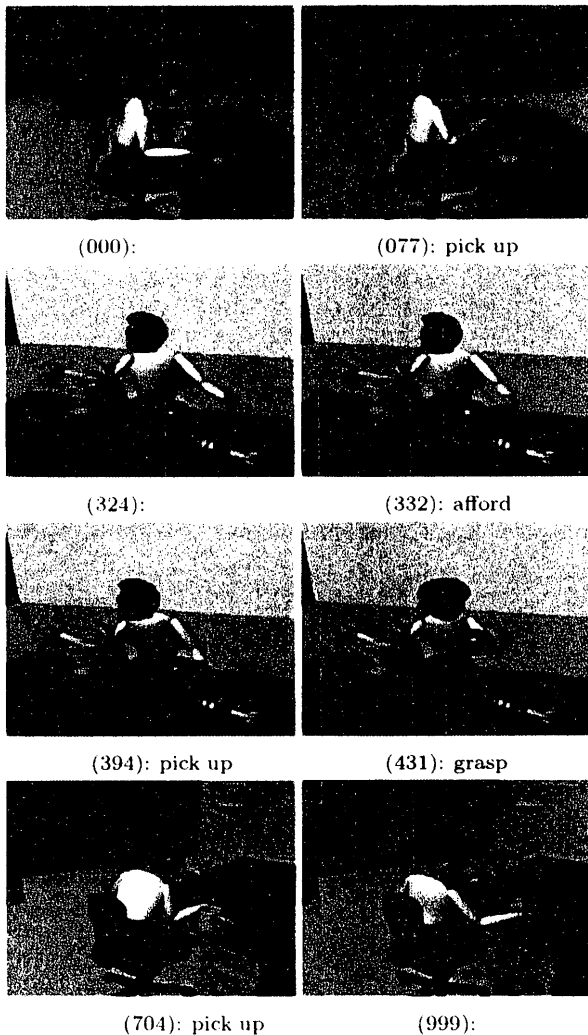


図3 アバターによる物体操作

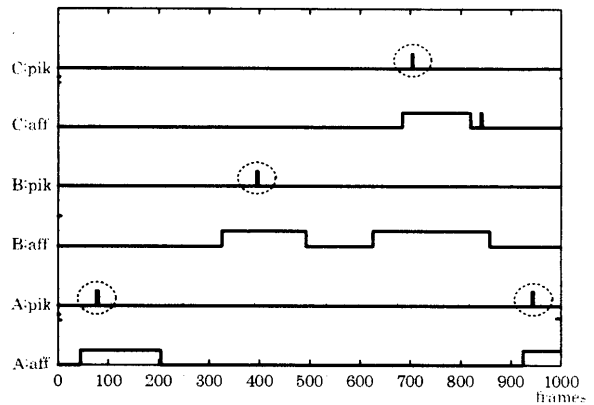


図4 物体操作における付加動作

- 画像認識による動作入力を用いており位置の精度が問題となる。
- ユーザは2次元の提示情報から3次元の仮想空間を把握しなければならない。

肌色領域の重心という推定部位の性質のため精度の向上はあまり期待できない。指の動きを必要とするような細かいタスクも指の姿勢推定を行わない限り実現できない。しかし物体操作タスクにおいては仮想シーンの情報を利用できるため、把持、移動、物体の機能の利用のようなタスクはユーザへの提示効果を高めることである程度改善できると思われる。また、物体操作タスクにおいてユーザが提示情報を効率的に利用するには、その視点の制御の仕方が重要となる。しかしながらタスク遂行に最適な視点を得る方法は仮想物体の様相や物体操作タスク、仮想シーンの状況等に強く依存するため一般化することは困難である。

アバターを表示しながら行う把持・移動タスクについてどの程度の応答性が得られるかを評価するための実験を行った。

- 仮想物体カーソルを1つランダムな位置に提示してそれを制限時間内に把持する
- 仮想物体カーソルを1つランダムな位置に提示してそれを制限時間内に把持・移動する(開始位置、終了位置を提示)

また、タスク遂行は上述のように視点に依存するため、いくつかのタスクが遂行可能な視点を選定し、視点との関係についても計測した。図5にいくつかの視点の例を示す(アバターの手前に表示しているのが仮想物体カーソル)。また、タスク遂行中に視点を自由に制御できる場合(本システムの機能の一部)についても計測した。視点制御の実現方法として、極座標系でのカメラ視点パラメータ記述を用いる。目標点を中心として目標点までの距離、方位角、仰角に関するパラメータで表される。実験では目標点までの距離、仰角に関するパラメータはシーンからタスク遂行に最適なものを経験的に選び固定した。上記(a)(b)のタスクを評価するにあたっては方位角を変化させて行った。タスクの達成率として、成功回数/試行回数を用いた。タスクの成功とは、制限時間内に把持あるいは移動が完了したことを意味する。被験者は1名で十分に物体操作の訓練を行った者である。安定したタスク達成率であったのは自由に

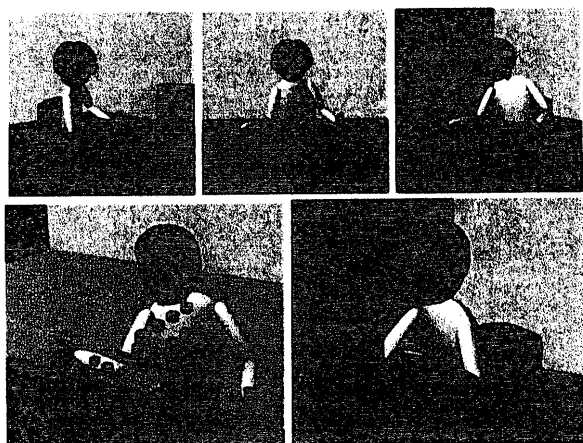


図5 物体操作の評価実験の例 (上) 方位角: -135, -90, -45 の場合。
(下) 用いた仮想物体カーソルの位置, 仮想物体カーソルの移動。

視点を制御した場合であった。固定視点の場合にそれぞれタスク達成率が低下した場合があったがその原因は3人称の視点による操作の難しさにあり、具体的には以下のものと思われる。

- アバター自身により発生するオクルージョン
- 提示視点における奥行き情報の不足

物体操作自身が重要である際はアバターの身体を表示しないかあるいは透過型の表示等の工夫によりアバターとユーザの視線が一致するように視点を定義すればよい。

5. おわりに

画像認識技術によりユーザの手や顔などの3次元位置をリアルタイムで獲得し、その動作入力をもとに仮想空間内でアバターとして身体動作を表現する方法について述べた。画像認識を動作入力に用いると、比較的安定に得ることが可能な位置は少数に限られるため、少数位置をもとにアバターの身体動作を表現する必要がある。身体動作を増強する方法として、物理法則に代表される構造上の制約により補完する方法やシーンの文脈を利用して表現動作を決定する方法を用いた。本研究では、ユーザは仮想空間において行為を行いその行為をアフォードする環境とは仮想空間を意味する。このことから仮想環境を完全に把握できるため、仮想シーンの文脈情報として動作情報をリアリティを高めるために利用できる。上半身の動作を仮想空間上に提示するアプリケーションを想定し、少数の動作入力からどの程度の自然な動作が提示できるかを検証した。評価実験によりアバターの動作が自動的に増強する様子を解析し、ユーザへの提示効果が高まることを確認した。また仮想物体操作タスクの評価実験も行い、視点制御と各物体操作タスクの関係を考慮する必要があることを確認した。今後の課題として、仮想空間操作インタフェースには個人に適応する能力を有することが重要であると考え、検討を行っている。

謝 辞

本研究は科学研究費補助金若手研究(B) 15700172 「少数の3次元計測位置からのヒューマンフィギュア動作制御ライブラ

リの開発とその応用」及び科学研究費補助金特定領域研究(C) 15017270 「直接操作を基にした個人適応能力を有する誘発型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] 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.
- [2] C.Wren, A.Pentland, "Understanding Purposeful Human Motion", in *Fourth IEEE International conference on Automatic Face and Gesture Recognition*, 2000.
- [3] 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.
- [4] Satoshi Yonemoto, Daisaku Arita and Rin-ichiro Taniguchi, Virtual Scene Control Using Human Body Postures, in *Proceedings of 1st IEEE Workshop on Computer Vision and Pattern Recognition for Human Computer Interaction (CVPRHCI)(in conjunction with IEEE CVPR 2003)*, 2003.
- [5] Satoshi Yonemoto and Rin-ichiro Taniguchi, "Vision-based 3D Direct Manipulation Interface for Smart Interaction", in *proc. of International Conference on Pattern Recognition*, pp.655-658, 2002.
- [6] M.Etoh, Y.Shirai, "Segmentation and 2D Motion Estimation by Region Fragments", in *International Conference on Computer Vision*, pp.192-199, 1993.
- [7] Y.Okamoto and R.Cipolla and H.Kazama and Y.Kuno, "Human Interface System Using Qualitative Visual Motion Interpretation", *IEICE*, Vol.J76-D-II, No.8, pp.1813-1821, 1993.
- [8] J.Kuffer Jr., "Autonomous Agents for Real-time Animation", *PhD thesis Stanford University*, 1999.
- [9] Y.Koga, "Planing Motions with Intentions", *Proc. of SIGGRAPH'94*, pp.24-29, 1994.