

画像認識による仮想環境直接操作インタフェースの 構築とその評価

中野, 裕史
九州大学システム情報科学研究所知能システム学部門

米元, 聡
九州大学システム情報科学研究所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所知能システム学部門

<http://hdl.handle.net/2324/5937>

出版情報 : 火の国情報シンポジウム, 2004-03
バージョン : accepted
権利関係 :



身体動作認識による 仮想環境直接操作インタフェースの構築とその評価

中野裕史*, 米元聡**, 谷口倫一郎***

九州大学大学院システム情報科学府*

九州大学大学院システム情報科学研究所***

〒816-8580 春日市春日公園 6-1

九州産業大学情報科学部知能情報学科**

〒813-8503 福岡市東区松香台 2-3-1

E-mail: *{nakano,rin}@limu.is.kyushu-u.ac.jp, **yonemoto@is.kyusan-u.ac.jp

本研究では、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次元身体操作の計測動作を開発し、利用する。非接触での身体動作の計測は、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。しかしカメラを用いて動作計測を行う場合、身体動作を精度よく推定することが一般に困難であるため、比較的安定に推定可能な手や顔の動作情報のみから身体動作の推定を自動的に行うアプローチにより、リアルな身体動作を仮想空間上に再現することを目指している。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、構築したインタフェースについて評価を行う。

キーワード 画像認識, Perceptual User Interface, インタラクション

3D direct manipulation interface by vision-based human action sensing

Hiroshi Nakano, Satoshi Yonemoto, Rin-ichiro Taniguchi

Department of intelligent systems*

Kyushu University

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

Department of Intelligent Informatics**

Kyushu Sangyo University

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

E-mail: *{nakano,rin}@limu.is.kyushu-u.ac.jp, **yonemoto@is.kyusan-u.ac.jp

This paper presents a real-time interaction system which consists of human motion tracking using skin-color blobs and human motion synthesis. Our purpose is to do seamless mapping of human action in the real world into virtual environments. For the sake of realization of smart interaction, we assume that virtual environments can afford human figure actions, that is, the virtual environments provide action information for a human figure model or an avatar. In this paper, we present a real-time, on-line and smart desktop interaction.

Key words Computer Vision, Perceptual User Interface, 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事が多くなってきている。その一方では、高度化・複雑化の一途を辿る機械や情報機器類と人間の乖離も進んでいる。人間中心の操作を行なうことができる情報機器として、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になっている。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ものではなく、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求められている。

ユーザからシステムへの情報入力は、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ーに代表される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、実用レベルまで達していると言える。しかし、現在利用されているモーションキャプチャシステムは高価かつ大規模であり、容易に扱うことのできるユーザインタフェースとしては、適しているとは言えない。上で述べている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、脱着の手間を取らない効率さを有し、しかもユーザの意図した動作をシステムが自動で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言える。

本研究では、直接的な3次元動作入力を基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り非接触な3次元身体動作の計測手法を開発し利用する。非接触での身体動作の計測は、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。しかしカメラを用いて動作入力を行う場合、身体動作を精度よく推定することが一般に困難であり、詳細な動作を長時間安定に推定するための手法の開発がまず必要となる。そこで、本研究では、比較的安定に推定可能な手や顔の動作情報のみを計測し、身体動作を再現するための姿勢推定を自動的に行うアプローチにより、リアルな身体動作を仮想空間上に生成することを目指している。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、計測した身体操作の入力をもとにユーザーの本来意図した動作を適切に再現するためのインタフェースを構築し、それについて検討する。

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^[2]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（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捕らえ方である。つまり、仮想空間に存在する様々なオブジェクトにアバターの動作に関連する情報を事前情報として定義しておき、アバターの動作や他のオブジェクトとの関係に応じてアフォードされた動作候補の中から適切なものをピックアップすることで、リアルな動作を生成できると考えられる。

2. 画像認識による身体動作の入力

以下では画像認識により身体動作の計測手法^[2]及びそれらの少数動作入力をもとにした自然な人体モデルの姿勢推定手法について概要を述べる^[3]。

2.1 身体動作の計測

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

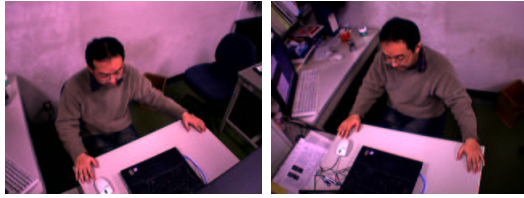


図 1: 2視点からの入力画像の例

担当し，肌色という均一色の領域を身体部位とみなして観測する^[4]．実際の肌色領域を含む画像より予めモデルパラメータとして色情報 (R,G,B) を学習しておき，そのモデル色と各画素の色情報 (R,G,B) とのマッチングを行うことで肌色画素を抽出する．抽出された画素がどの肌色領域に属するかは，最も近い位置にある肌色領域へクラスタリングすることによって決定する．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．図1に2視点からの入力画像の例を示す．図2(左)に肌色領域の推定例を示す．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，厳密な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を用いる．対応の取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．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．

2.2 身体姿勢の推定

インタラクションを遂行する際，他者に正しく情報伝達を行うための手段としてアバターの動作をユーザの位置に合わせて適切に再現することが必要な場合がある．そこで，本研究では，画像認識により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を導入する．人間らしい自然な動作

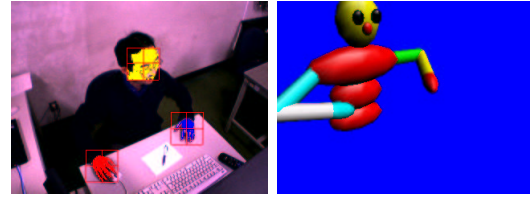


図 2: 上半身動作推定の例．(左) 肌色領域 (右) 動作生成の例

を再現するためには，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，一般に画像認識により求まる位置，速度などの物理パラメータの推定精度も悪い。うえ，少数部位の位置情報しか用いることができないなどの理由から，厳密な動力学系として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，画像認識により得られる位置情報のみを用いて，物理法則に基づいた動作生成を行う方法を導入する．物理法則として身体構造の制約としてバネモデルを仮定し，位置入力の変異をもとに当てはめ計算を行うことで姿勢の更新を行う．図2(右)に示す上半身動作の例のように，3点の位置から上半身の姿勢の推定が可能である．

3. 仮想環境直接操作インタフェース

本システムでは，カメラから得られた実空間情報をもとに実装されるユーザインタフェースを適用するために，自分の分身となるアバターを用いたシステムを構築する．アバターの表示を確認することで，3次元位置の認識結果を視覚的に理解することが容易にできる．実空間からの人体動情報より具現化したアバターを操作することで，仮想空間中のオブジェクトを把持したり，移動させたり，放したり，回転させたりすることができる．また，実空間物体を用いてインタラクションを行なうことも可能である．本研究では，ユーザからの動作入力に基づいて仮想空間と実空間が融合するインタフェースシステムの種類として以下のものを実現している．

- 積み木を利用したインタラクション
- ネットワークを利用したインタラクション

3.1 積み木を利用したインタラクション

ユーザは、仮想空間中に定義された箱やティーポットのような静的なオブジェクトを把持、移動、回転、放す、などの操作を行うことができる。仮想空間内のオブジェクトの配置や動作情報などは、あらかじめ定義されている。これらの動作の定義は、仮想空間内に表現されるアバターの手とオブジェクトとの距離や時間などによって定義されており、シーンの選択や変更を行なう時も、ユーザの動作入力をもとに容易に行なうことができる。この例では、単に仮想空間のオブジェクトの把持・移動・回転・放すを行なっているが、「把持」は、手とオブジェクトの距離が一定の距離に近付き、そしてある程度その距離が保たれた時に、初めてそのオブジェクトを掴むと言う動作を行なう。「移動」は、オブジェクトを把持している時のみ行なう事ができ、「放す」ことでその動作を終了する。「回転」は、「把持」している状態のとき把持していない手を近づけることで、回転動作を行うことができる。「放す」は、オブジェクトが置きたい場所に置かれる、もしくは手が余り動かない状態になった時、オブジェクトを放したいと判定させることで放すことができる。フィードバックとしてオブジェクトの色を変える事にした。これらの動作を関連づけることで複雑な操作を行なうことができると考えている。

そこで仮想空間中に表示された大きさの違う積み木を単に積むだけのタスク、模様がついた箱を複数表示しておきパズルのようにうまく並べることである模様を完成させるタスクを構築した。今回このような単純なモデルを選んだ理由は、与えられた積み木からイメージしやすい、つまりわかりやすいメンタルモデル（実際の機器の振る舞いと整合性に関らず、ユーザが心的に持っている機器の動作に関するモデル）⁵⁾であるからである。

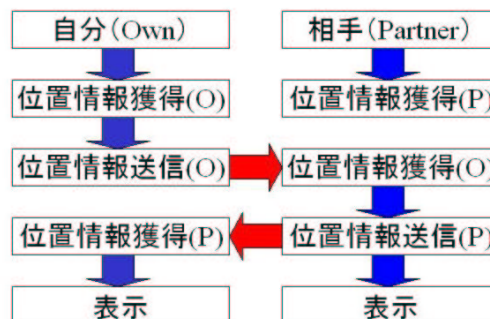


図 3: 情報の送受信の流れ

また、仮想空間と実空間のユーザとの間の円滑なインタラクションを遂行するために、提示する仮想空間の視点の制御とフィードバックを導入する。直接的で理解しやすい視点制御の方法は、ユーザの頭部位置に応じて視点の制御を行なう事である。ユーザが左右へ頭部を傾けることで目標点を中心に左右に回転することができ、前後へ頭部を傾ける事で目標点に対してズームイン、ズームアウトを行なう事ができる。ユーザの動作に直接的にマッピングする視点制御の方法は有効ではあるが、課題はその他の視点の制御パラメータをどのようにマッピングするかである。フィードバックの導入により、現在オブジェクトを「把持」しているのかどうか知る事ができる。

3.2 ネットワークを利用したインタラクション

一般に仮想世界の利用には、

- 単独
- 協調作業(遠隔)

の二つがある。特に遠隔での協調作業には仮想化が不可欠でありこのようなシステムの応用が考えられる。そこで遠隔地にいる利用者とネットワークを通じてインタラクションを行うシステムを構築した。これは、同じ仮想環境内のオブジェクトの把持、移動、回転放すなどを動作を行なう事ができ、また、アバ

ター同士のインタラクションを行なう事も可能である。

今回は、仮想環境に自分と遠隔地の使用者のアバターの2人を表示するようなシステムを構築した。システムの流れを説明すると、仮想環境内の自分のアバターを Own, 相手のアバターを Partner とし、使用者側も自分のアバターを Own, 相手のアバターを Partner する。ここで必要なパラメータは自分自身と相手の位置情報であり、システムの流れは、図3で示す。まず Own, Partner の位置情報を求めたあとで、Partner にその情報を送信して、Own に Partner の位置情報を獲得し表示を行なっている。

4. 実験と評価

4.1 システムの概要

上記の実験として、デスクトップ型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 DFW-V500) 2台を用い、実空間上には、提示用の大画面モニタ (50型プラズマディスプレイ), 及び机, 椅子を配置し、上半身のみがカメラにより撮影される。画像認識が簡易な方法であるため、1台の汎用 PC で画像認識から仮想空間の提示処理まで行うことが可能である図4にシステムの概要を示している。ユーザは2次的に提示される仮想空間を見ながら、3次元仮想物体とインタラクションを行う。

4.2 実験手順

被験者にこのインタラクションシステムの概念及び操作方法について約10分間説明し、実装したインタフェースの使用例を見学させた。その後、システムに慣れてもらうために約5分間操作練習をしてもらった。その後、実際に実験を行ってもらった。

まずは、積み木を利用したインタラクションの実験として、積み木を大きい物を下から順番に置いていくタスクを実行してもらった(図5(上))。このとき、視点制御がある場合とない場合の2通り実験してもらった。次に模様をついた4個の積み木をパズルのように組み合わせるタスクを実行してもらった(図



図 4: システム概要

5(下))。最後にネットワークを利用したインタラクションの実験として、仮想空間上に4個の模様のあるオブジェクトが存在し、それらのオブジェクトを自分と遠隔地にいる相手と共同で組み合わせるというタスクを実行してもらった。場所は研究室間で一人は被験者、もう一人はシステムに慣れている人で実験を行なった。図6にネットワークを利用したインタラクションの例を示す。

また、評価のために、作業開始から終了までの時間を計測し、実験終了後にアンケートを5段階評価で行なわせ、最後にシステムに対する感想などを記述させた。被験者は、仮想空間とのインタラクションの経験がないグループ(グループ A), 経験のあるグループ(グループ B)を対象に視点制御を行なわない積み木(タスク A), 模様をついた積み木(タスク B), 1タスクにつき3回ずつ実験を行なった。

4.2.1 予備実験

ネットワークを利用するインタラクションを行なう際、どれほどの遅延が発生するか実験した。実験を行なった場所は研究室の離れた部屋間、九州大学内の離れたキャンパス間、そして九州大学と FTTH 間で実験を行なった。

遅延は平均しては、研究室内が 33.5msec, 九州大学キャンパス間が 33.6msec, 九州大

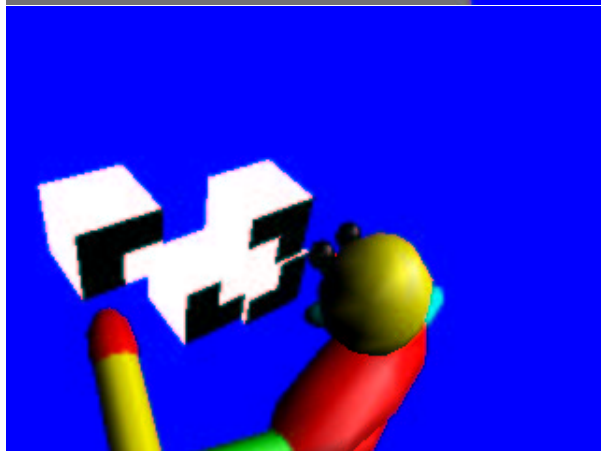


図 5: 積木を利用したインタラクションの例
(上)単純な積み木(下)模様をついたパズル

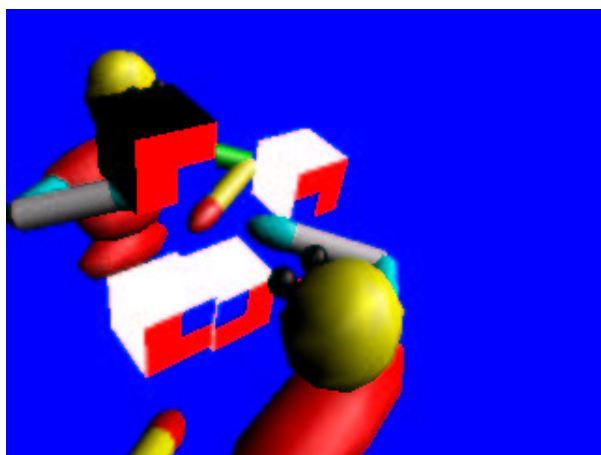


図 6: ネットワークを利用したインタラクションの例

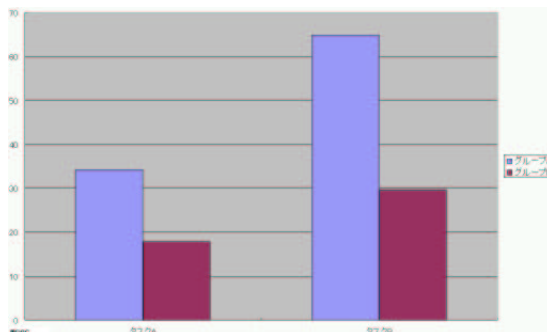


図 7: 実験結果

学と FTTH 間が 52.8msec だった。また、故意に遅延を発生させ遅れて感じるか実験したところ、100msec 以上は明らかに遅く感じた。故意の遅れで 50msec までは遅れを感じなかった。現在のインターネット環境では、ほぼ問題なく実行できると思われる。

4.3 実験結果及び考察

図 7 に 4 名ずつ計 8 人の被験者の平均タスク時間を示す。グループ A とグループ B では、作業時間に差が生じる結果結果となった。特にグループ B はタスクを完了するまでの時間にバラツキが感じられた。これは自分の思った動作と違う動作を行なった時に手間取っていたこと、グループ A に比べて慣れていないのが原因だと考えられる。また実験後に行なったアンケートの結果、直観的でわかりやすい、面白いなどの好意的な意見を得る事ができた。また、視点制御は円滑なインタラクションを行なうために必要であるという意見を得る事ができた。

一方で、ユーザが仮想空間中のオブジェクトを把持、移動などをする際に手間取る場面も多く見られた。これはユーザの手が他の肌色領域と重なる事によるトラッキングミスや、自分が思っているところに置けない、オブジェクトが接近している時うまく把持、移動などをうまく行なう事ができないなどが原因である。これらの問題が操作の違和感や疲れ、ストレスにつながっているためユーザによりよく操作してもらうために改善する必要

がある。

5. おわりに

画像認識からの少数動作入力をもとにアバターの自然な動作制御を行う技術及び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インタフェースについて検討し、構築したインタフェースについて評価を行った。今後の課題として、ネットワークを介した協調作業の改良、専門知識を持たない人でも仮想空間に違和感なく溶け込めるような理解しやすいインタフェースを目指し、多くのインタラクションシーンへと応用することでその効果を実証することが挙げられる。

謝辞

本研究の一部は科学研究費補助金特定領域研究「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。

参考文献

- [1] M. Weiser, “Some Computer Science Issues in Ubiquitous Computing,” *Communications of the ACM*, Vol. 36, No. 7, pp75-84, 1993.
- [2] S. Yonemoto, D. Arita and R. Taniguchi, “Real-Time Human Motion Analysis and IK-based Human Figure Control,” *Proc. Workshop on Human Motion (HUMO2000)*, pp. 149-154, 2000.
- [3] S. Yonemoto and R. Taniguchi, “Vision-based 3D Direct Manipulation Interface for Smart Interaction,” in *Proc. of International Conference on Pattern Recognition*, pp. 655-658, 2002.
- [4] 川戸慎二郎, 大谷 淳, “眉間の検出・追跡による顔きと首振りのリアルタイム検出,” *信学技法*, PRMU99-108, Nov. 1999.
- [5] D. A. ノーマン “誰のためのデザイン?”, 新曜社, 1990.
- [6] 佐々木正人, “アフォーダンスの構想 - 知覚研究の生態心理的デザイン,” 東京大学出版会, 2001.