

画像認識による身体動作入力にもとづく誘発型インターフェースの構築

米元, 聰
九州産業大学情報科学部知能情報学科

中野, 裕史
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, 優一郎
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/2324/5811>

出版情報：電子情報通信学会技術研究報告. IE, 画像工学. 102 (632), pp.171-176, 2003-01-28. 電子
情報通信学会
バージョン：
権利関係：

画像認識による身体動作入力にもとづく誘発型インターフェースの構築

米元 聰[†] 中野 裕史^{††} 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, ††{nakano,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自然な3Dキャラクタの動作制御を行うことで仮想空間との円滑なインタラクションを可能にするシステムの開発を目指している。ユーザの動作入力に応じて、システムが誘発を行うことでユーザの本来意図した動作を自然な形で代行表現する誘発型ユーザインターフェースの枠組みについて述べる。具体的には、誘発的なインタラクションを実現するために、仮想空間におけるアフォーダンスの概念によるアバターの適切な動作制御の導入及び身体動作に基づき仮想空間の視点制御を行う方法の検討を行った。

キーワード 画像認識, Perceptual User Interfaces, インタラクション, 視点制御

Affordance-based Perceptual User Interfaces

Satoshi YONEMOTO[†], Hiroshi NAKANO^{††}, 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-kon 6-1, Kasuga-shi, 816-8580 Japan
 E-mail: †yonemoto@is.kyusan-u.ac.jp, ††{nakano,rin}@limu.is.kyushu-u.ac.jp

Abstract This paper describes an 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 also developed a virtual camera control mechanism using physical motions.

Key words Computer Vision, Perceptual User Interfaces, Interaction, Virtual Camera Control

1. 概要

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ー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発展に伴い、個人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のように、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ー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次元世界における操作から成り立っていることから、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ーフェースの実現が望まれる。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。しかし、上述のスマートなインター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、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る非接触な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な課題であり、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、計測した身体動作の入力をもとにユーザの本来意図した動作を適切に再現するための誘発型ユーザインタフェースの枠組みについても検討する。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[7]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することが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る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] 参照）。

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、両手部分が相当し、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[6]。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。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。また、物体操作における把持判定のために、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。

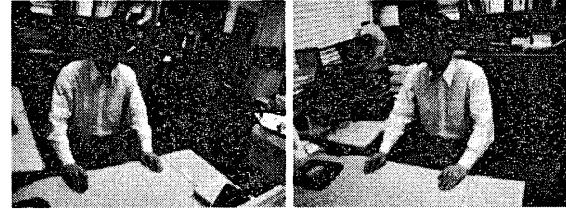


図1 2つの視点からの入力画像の例。

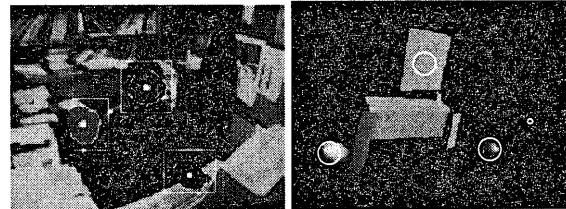


図2 上半身動作推定の例。（左）肌色領域。（右）動作生成の例。

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動作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、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上、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、画像認識により得られる位置入力のみを用いて、物理法則に基づいた動作生成を行いう方法を導入する。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。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。肩や肘は、3点の位置変位に応じて自動的に計算される。

3. 誘発型ユーザインタフェース

本研究は、アフォーダンスの考え方などを基にして、仮想空間中のオブジェクトがインタラクションに関与する情報を積極的にシステムに与えることにより、インタラクションが誘発的に行われることを目指している。そこで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、以

以下の2つの誘発的なインタラクションを実現する機構を導入する。

- 仮想空間におけるアフォーダンスの概念によるアバターの適切な動作制御
- 提示する仮想空間の視点の制御

3.1 物体操作時のアバターの適切な動作制御

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（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く、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、適切な動作を生成することができると考えられる。仮想空間中のオブジェクトとアバターとの間のインタラクションは、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をアフォードされ、システムがそれ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。ここで、アフォードされる、とは、現実世界の場合とは異なり、ユーザの身体動作入力や、他の仮想シーンのイベントによりその動作情報が利用可能な状態になることを指す。例えば、カップやペンのような仮想物体は、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バターに対しアフォードする。実際にそれらの動作情報が利用可能であるかの判断やそれが選択されるのかは、アバターの手と仮想物体との関係（距離や状態）に応じてシステムがピックアップ処理を行い、その動作をアバターや関連するオブジェクトへ適用することにより達成される。カップの把持動作がピックアップされた場合は、カップの状態はアバターの手の動作に同期して移動するよう変化し、アバターには手指の詳細な把持動作を自動的に補完生成する動作が適用される。自動制御される動作としては、具体的には、仮想物体の特性に応じた上記の手指の詳細動作や、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むけるための視線制御などがあげられ、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ていると考えられる。

3.2 提示する仮想空間の見えを制御する

仮想空間の提示の仕方、つまり視点の制御を工夫することでユーザの行為を誘発することができる。

まず本研究で想定する視点制御に関するパラメータの定義を行う。

3.2.1 視点制御のパラメータ

注視する目標点の位置を中心に考え、極座標系でカメラ視点パラメータを記述する（図3）。カメラ視点より目標点までの距離（視点球の半径）を r 、方位角を $azimuth$ 、仰角を $elevation$

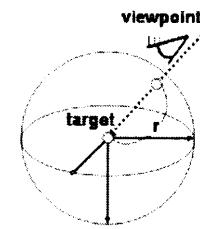


図3 仮想カメラ視点と目標点の関係。

とし、カメラ視点より注視する目標点の位置 (T_x, T_y, T_z) 、とおくとき、カメラ視点の位置 (C_x, C_y, C_z) は以下の式によつて計算される。また view up ベクトル（上方向を決めるベクトル）は上記より導出する。

$$\begin{aligned} C_x &= T_x + r \cdot \sin(azimuth) \cdot \cos(elevation) \\ C_y &= T_y + r \cdot \cos(azimuth) \cdot \cos(elevation) \\ C_z &= T_z + r \cdot \sin(elevation) \end{aligned} \quad (1)$$

これらの視点制御のパラメータ記述はマウスによる視点制御で利用されている方法であり、マウス操作の代わりにユーザの動作にそのまま割り当てることも可能である（後述）。

以下、上記で定義した視点制御パラメータに関連して視点制御を行うための方法をいくつか述べる。

3.2.2 視点制御の方法(1)

最も単純で直接的な視点制御の方法は、ユーザの頭部位置に応じて視点の制御を行うことである。ユーザが左右へ頭部を傾ける際の変位を上記の視点制御のパラメータのうちの方位角 $azimuth$ の変位に対応させ、ユーザが前後へ頭部を傾ける際の変位をカメラ視点より目標点までの距離 r の変位に対応させる。このような直接的な視点制御法はデスクトップ型の仮想空間操作システムにおいて従来より実現されている。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あるが、課題はその他の視点制御パラメータをどうマッピングするかである。

3.2.3 視点制御の方法(2)

物体操作のためのユーザの動作やアフォーダンスの概念の応用による動作補完により、ユーザの関心のある対象を特定することができる。ユーザが円滑にインタラクションを遂行できるようにするためにには関心のある対象を視野中心に捉えることが望ましい。本研究で想定する視点制御のパラメータに関連させるとカメラ視点より注視する目標点の位置 (T_x, T_y, T_z) が対象の位置にそのまま相当する。また上記の視点制御の方法(1)と同時に利用できる。現在想定しているのはカメラ視点の瞬時の切り替えであるため関心対象が頻繁に移動する場合や、複数対象がある場合などには、カメラ視点の軌道が連続的に切り替わる方が望ましいがこれは今後の課題である。

3.2.4 視点制御の方法(3)

通常マウス操作により行われる視点制御をユーザの動作に

マッピングする方法、具体的には、仮想カーソルを画面中央付近に表示させ、そのカーソルを掴んだ状態で行われるユーザーの手の変位に合わせて視点を制御する。カーソルを掴んだ状態がマウスのボタンのクリック時に相当する。仮想カーソルをアバターの真正面の空間に置く場合は、左右の手の変位を方位角 *azimuth* の変位、上下の手の変位を仰角 *elevation* の変位に対応させる。またカーソルをデスクトップ平面に置く場合は通常のマウス操作と同様の感覚になり、左右の手の変位を方位角 *azimuth* の変位、前後の手の変位を仰角 *elevation* の変位に対応させる。仮想カーソルを複数用意することで、前後の手の変位をカメラ視点より目標点までの距離 r の変位に、あるいは手の上下左右の変位を、提示画面に平行に注視点をスライドさせるための変位に対応させることができる。

なお提示画面に平行に注視点をスライドさせるための計算は以下の式によって行う。

$$\begin{aligned}\Delta T_x &= \Delta x \cdot \cos(\text{azimuth}) \\ &\quad + \Delta y \cdot \sin(\text{azimuth}) \cdot \sin(\text{elevation}) \\ \Delta T_y &= -\Delta x \cdot \sin(\text{azimuth}) \\ &\quad + \Delta y \cdot \cos(\text{azimuth}) \cdot \sin(\text{elevation}) \\ \Delta T_z &= -\Delta y \cdot \cos(\text{elevation})\end{aligned}\quad (2)$$

ただし、手の左右の変位及び上下の変位により加えられるスケーリングした変位をそれぞれ Δx , Δy とし、計算される注視点の変位を $(\Delta T_x, \Delta T_y, \Delta T_z)$ とあらわしている。

3.3 視点制御の適用

上記の視点制御の各モードをシーンに応じて切り替えながら利用することを想定している。以下ではその適用例について示す。

3.3.1 アバターとしての身体表現を必要とする場合

ネットワークを通して共有化された仮想空間を想定する場合や他者（他のユーザや、他の擬人化エージェント）がユーザの動作を知覚することを想定する場合など仮想空間でのアバターとしての身体姿勢表現を提示することが必要な場合には、ユーザは実空間に対応した限られた仮想空間においてインタラクションを行う[8]。もちろん実空間と対応させる仮想空間を切り替えることは可能である。アバターの動作制御は前述した方法で行う。ユーザは上記の視点制御の方法(1)(2)(3)の何れも利用することが可能である。

3.3.2 アバターとしての身体表現を必要としない場合

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。この時、ユーザは上記の視点制御の方法(1)(2)(3)の何れも利用することが可能である。空間が限定されないため、連続的に視点制御を行い、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。また、物体操作時には、アバターの代わりにユーザの手や頭の位置を示すカーソルを表示させることでインタラクションの補助を行う。また、

視点制御の方法(2)の応用として、手を示すカーソルと提示像中の関心対象との2次元上での位置あわせ、及び把持状態の判定による関心対象の特定により関心対象を視野中心に捉える視点制御を行うことが可能である。この制御により、2次元の提示像を見ながら3次元空間を移動するための効率をあげることができる。またウォークスルー方式において、視点の前後への移動は注視点の移動に相当する。上述の視点制御の方法を応用し、頭部位置の変位を注視点の変位に対応させることによって直接的な制御方法として実現できる。

3.3.3 モードの選択

身体動作を利用したユーザインターフェースの機能に欠かせないのが、従来キーボードで行われている様々なコマンドの実行を代行させる工夫である。本研究では既存の音声認識装置・手法を用い、音声コマンドによる上記のアバターモード、ウォークスルーモードの選択や起動・終了を想定する。

4. プロトタイプシステムの実装と実験

4.1 システムの概要

プロトタイプシステムとして、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 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像される。画像認識が簡単な方式であるため、1台の汎用PCで画像認識から仮想空間の提示処理までを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次元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 アバターとしての身体動作の再現

図4に示すように、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。画像認識により、顔、両手の3次元位置を推定し、それらを用いて上述の物理法則に基づく身体姿勢推定法によりアバターの身体姿勢を更新する。肘や肩の位置に関しては実空間のユーザの姿勢情報を用いずに自動的に計算されることになる。また、顔の向きや、手首の向き等に関しては、関心を向けている仮想物体へのインタラクションに応じて自動制御される。また、仮想物体操作における把持判定の目的で実空間でユーザが行う手の状態を把持状態、開放状態にあるか推定する。

4.3 視点制御の基礎実験 1

図5に視点制御による目標物体への視点移動の例を示す。入力されるユーザの動作に従い、目標物体へアバターの右手を近づける。この時、方位角、カメラ視点から目標点までの距離に関するパラメータは、視点制御の方法(1)により予めユーザ自身が頭部位置の変化により調整する。右手の接近により目標物体の動作情報がアコードされ、まず視点制御により注視点を目標物体へ移動する。視野の中心に目標物体を捉えることでアバターによる把持・移動を行いやすくする提示の効果を与える。

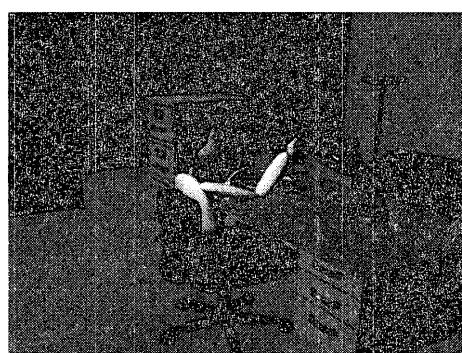
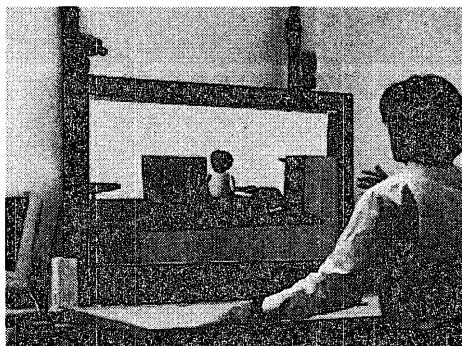


図4 システムの概要：（上）設備および実験環境、（下）アバターを表示した仮想空間の例。

その後、物体の把持に関する手指の動作が制御され、物体の移動が行われる。視野中心に関心対象を捉えることは必ずしもユーザの意図した視点制御とはいえないかもしれないが、その他の視点制御のパラメータは目標物体を中心に定義されることになり、提示効果は高まると思われる。また、提示効果をさらに高めるために方位角などの自動制御を行うことも考えられるが、本研究ではユーザの頭部位置変化による直接的な制御（視点制御の方法(1)）を用いている。ただし仰角については方法(1)では制御できないため、その自動化について検討を行っており今後の課題とする。

4.4 視点制御の基礎実験 2

視点制御の方法(3)について実装を行い、アバターの身体を表示させない場合について、視点制御を伴う仮想空間ナビゲーションへと適用した。仮想空間ナビゲーションを実現する場合、実空間での操作空間に関する座標系を仮想カメラの視点を基準とした座標系に対応づけ、ユーザの身体動作入力を常に仮想カメラ視点位置を基準とした動作入力として扱う。実現する視点制御については本質的にマウスと同等の操作を手による操作に代行しているにすぎないが、ハンズフリーによるユーザの3次元空間の操作性の向上は期待できる。図6に仮想空間ナビゲーションの実現例を示す。ユーザは視点制御(3)の方法で、右手による操作により注目対象（ぬいぐるみ）の位置へとウォークスルーし、任意の視点を選択する。その後、物体操作モードに切り替え、注目対象を把持・移動する。

4.5 考 察

物体操作のように実空間に対応させた仮想空間を想定する場合、ユーザの位置が提示情報を知覚するためにディスプレイの目の前に制限されることや計測範囲の制限により、把握可能な仮想空間が制限されてしまうことになる。そこで、本研究ではこの制限をなくすためアバターとして身体を表示させないモード（仮想空間ナビゲーション）を導入し、ユーザの判断によりモードを切り替えることで関心対象を中心としたシーンへ移動することを想定している。この切り替えによってアバターの位置を新たな関心対象の付近に更新する必要があるが、不自然さをなくすため、アバターの存在可能な位置は地面に立てる位置や、椅子に座れる位置などシーンにより限定するものとする。



図5 関心対象への視点制御の適用例。

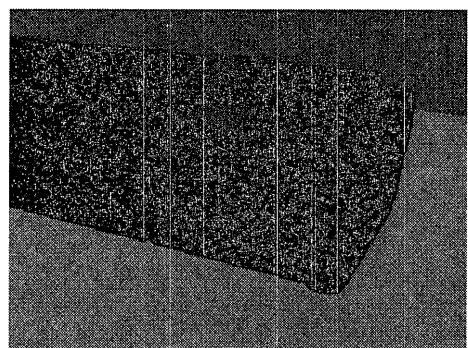
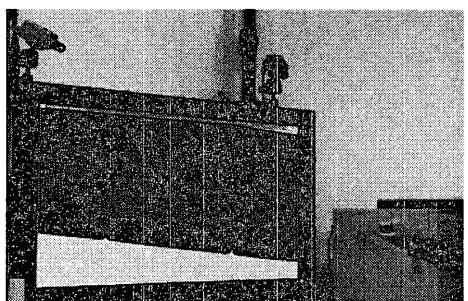


図6 身体動作による仮想空間ナビゲーション。（上）実行例、（下）仮想空間の例。

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の円滑なインタラクションを可能にする誘発型ユーザインタフェースの枠組みについて検討し、ユーザの動作入力に応じてシステムが誘発を行うことでユーザの本来意図した動作を自然な形で代行表現するための動作制御の方法、身体動作にもとづく仮想カメラ視点の制御の方法について述べた。これらの誘発的なインタラクションを行うための工夫により、実空間上のユーザが2次元の提示像を見ながら3次元仮想空間中の仮想物体を操作する際の効率を上げることが可能となる。今後の課題として、身体動作に基づく仮想カメラ視点制御の改良、本研究の個人適応に関する検討、より完成度の高い誘発型ユーザインタフェースの構築を目指し、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(C) 14019071「個人適応能力を有する高次マルチモーダル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] 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.
- [4] Satoshi Yonemoto and Rin-ichiro Taniguchi, "A Direct Manipulation Interface with Vision-based Human Figure Control", HCI International 2003 (発表予定).
- [5] 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", 電子情報通信学会論文誌, Vol.J76-D-II, No.8, pp.1813-1821, 1993.
- [6] 川戸, 鉄谷, "リング周波数フィルタを利用した眉間の実時間検出", 電子情報通信学会論文誌, Vol.J84-D-II, No.12, pp.2577-2584, 2001.
- [7] 佐々木正人, "アフォーダンス - 新しい認知の理論", 岩波書店, 1994.
- [8] J.Kuffner Jr., "Autonomous Agents for Real-time Animation", *PhD thesis Stanford University*, 1999.