

Perceptual Glove : 多視点画像に基づく 手形状・姿勢の実時間入力とその応用

齋藤 真希子[†], 佐藤 洋一^{††} 小池 英樹[†]

ユーザにとってより自然なヒューマン・コンピュータ・インタラクションを実現するためにはキーボードやマウスなどの従来の入力デバイスによる操作に加え、ユーザの手振りによる直接操作が有効であると考えられる。そこで本論文では、複数カメラからの入力画像をもとに手の3次元位置・姿勢推定および手形状判別を実時間でを行うための手法を提案する。特に、手形状判別にニューラルネットワークによるパターン識別を利用することにより、複数のユーザによるさまざまな手形状を信頼性良く認識することを可能としている。また、具体的なインタラクションに対する提案手法の有効性を評価するために、3次元仮想物体の手による直接操作および没入型映像環境内におけるナビゲーションの2つのタスクについて予備的な主観評価実験を行った結果について報告する。

Perceptual Glove: Real-time Input of 3D Pose and Configuration of User's Hand Using Multiple Cameras and Its Applications

MAKIKO SAITO,[†] YOICHI SATO^{††} and HIDEKI KOIKE[†]

In this paper, we introduce a method for tracking a user's hand in 3D and recognizing the hand's gesture in real-time without the use of any invasive devices attached to the hand. Our method uses multiple cameras for determining the position and orientation of a user's hand moving freely in a 3D space. In addition, the method identifies predetermined gestures in a fast and robust manner by using a neural network which has been properly trained beforehand. This paper also describes results of user study of our proposed method and its application for several types of applications, including 3D object handling for a desktop system and 3D walk-through for a large immersive display system.

1. はじめに

日常生活のあらゆる場面において計算機が利用されるようになるにつれ、ヒューマン・コンピュータ・インタラクション(HCI)の重要性がいっそう増している。現在、マウスとキーボードを入力装置としアイコンやウィンドウなどのデスクトップメタファによる操作を基本とするグラフィカル・ユーザ・インタフェース(GUI)が広く一般的に利用されている。テキストを中心とした計算機端末の利用形態からGUIによる利用形態へと発展してきた過程においてさまざまな改良が加えられ、GUIによるインタフェースはワードプロセッシングやウェブブラウジングなどの作業におい

て効率の良い作業環境を実現している。

しかしながらその一方では、GUIはすべての作業に関して必ずしも効率の良いインタラクションを提供するとは限らない。特に、自由度の高い入力が必要不可欠となる作業に関しては、マウスとキーボードを入力手段としたGUIでは自然な操作環境を実現することが困難である。たとえば、マウスにより3次元物体を操作することを考えた場合、入力デバイスであるマウスの動作が2次元平面に拘束されているために、6自由度を持つ物体の3次元位置・姿勢を操作することは必ずしも直感的であるとはいえない。

このようにGUIに代表される現在のHCIの枠組みでは、人間同士の対話のように自然なインタラクションを実現することが難しい。その主たる原因は以下の3つにまとめられる¹⁾。

- (1) 人とコンピュータのインタラクションに使える入力方法(モード)の数が非常に少ない。人からコンピュータへの入力はキーボードやマウスであり、コンピュータから人への表示は主とし

[†] 電気通信大学情報システム学研究所
Graduate School of Information Systems, University of
Electro-Communications

^{††} 東京大学生産技術研究所
Institute of Industrial Science, The University of Tokyo
現在、日本電気株式会社 NEC ネットワークス開発研究所
Presently with NEC Networks

て小さい2次元画面に限られている。

- (2) 各モードを通して享受できる情報量が少ない。特にコンピュータへの入力作業は効率が悪い。
- (3) インタラクションの手段が不自然である。多くの入力作業で身体がキーボードやマウスなど接触型の装置に拘束されている。また、情報の質に適合したモードが存在しないことも不自然さを拡大している。

これらの課題に対して、PUI(Perceptual User Interface)と呼ばれるアプローチが近年注目を集めている²⁾。GUIに代表される従来のHCIではシステムへの情報入力はいくまでもキーボードやマウスなど入力デバイスを介してのみ行われていたのに対し、PUIは計算機側にさまざまなセンシング機能を持たせることにより状況を把握(percept)しながら、行動主体であるユーザにとってより自然なインタラクションを実現しようとするインタフェースとなっている。

本研究では、PUIの考え方に基づく新たな入力方法として人間の手によるジェスチャ、すなわちユーザが自分の手を3次元空間内で動かすことによる直接操作によりインタラクションをとる手法が有望であると考へた。このことは、人間同士の共同作業などにおいてジェスチャによって伝達される情報の多様性を考えれば容易に想像することができる。そこで本研究では、直感的で使いやすいインタフェースの実現を目指し、ユーザの手に接触型センサやマーカなどを付与することなく3次元位置・姿勢の推定およびジェスチャ判別を行うための手法を提案する。本研究におけるこの提案手法をPerceptual Gloveと呼ぶことにする。

本論文の構成は以下のとおりである。まず、2章では本研究における提案手法と関連する研究事例について紹介する。次に、3章において提案手法における手の3次元位置・姿勢推定およびジェスチャ判別処理の各々についてその実現方法を説明する。4章では、提案手法の性能を評価するために行った評価実験について報告し、5章では、提案手法を具体的な作業に利用した場合の有効性に関する評価実験の結果を紹介する。最後に、6章で本研究のまとめを述べる。

2. 関連研究

ユーザの手振りをインタラクションに利用しようとする試みが人工現実感やコンピュータビジョンなどの分野で数多く報告されてきている。

人工現実感では、磁気センサなどによる3次元位置センサを備えたグローブ型のデバイスを用いてユーザの手の位置姿勢と形状を実時間で精度良く計測す

ることが一般的に行われてきた。このような入力デバイスの最も初期のものとしてはVPL Research社のDataGlove³⁾が有名であり、ほかにも多く開発されている⁴⁾。このようなグローブ型の入力デバイスは人工現実感システムにおける重要な入力手段として広く一般的に用いられてきたが、自然なインタラクションという観点からは、着脱の手間やグローブ型デバイスとコントローラを接続するケーブルなどによる自由な動作の制限などの問題により必ずしも適しているとはいえない。

このようなことから、ケーブルなどを必要としない非接触型のアプローチとして画像処理に基づく手法が研究されてきた。その1つのアプローチとして何らかのマーカなどを利用するものがある。たとえば、Cipollaらによる手法⁵⁾やMaggioniによる手法⁶⁾では、手に取りつけたカラーマーカや特別な形状のマーカをもとに画像中における指先位置や手の位置・姿勢を安定に検出している。このようなアプローチはグローブ型デバイスの利用よりはインタラクションへの応用に適しているものの、ユーザにとってより自然なインタラクションを実現するためにはマーカなどをユーザの手に貼付けるということはできる限り避けることが望ましい。

このため、マーカなどを利用しないアプローチも積極的に研究されてきている。このような場合には、まず入力画像中においてユーザの手に対応する画像領域を抽出するという前処理が必要とされる⁷⁾。この目的のために黒い背景などを想定し環境を限定することによる2値化、視点固定かつ静的な背景を仮定することによる背景差分、手の肌の色に関する仮定に基づく色抽出、などが一般的に行われることが多い。これらはどれも一長一短があり、想定される利用法に応じて適当な方法を選択する必要がある。このようにしてユーザの手領域が抽出された後、手領域の形状特徴をもとに手中心や指先の位置が推定される。たとえば、Maggioniらによる手法⁸⁾やSatoらによる手法⁹⁾では手領域の輪郭形状をもとにして画像中における指先の2次元位置を推定している。

一般に、単眼画像のみからでは手の3次元位置・姿勢を求めるには情報が不足してしまううえに、手のように複雑な形状を持つ物体ではセルフオクルージョンが問題となる。このため、複数視点からの入力画像を用いることにより手の3次元位置・姿勢を推定する手法が提案されている。たとえば、Fukumotoらの手法¹⁰⁾では2台のカメラを利用して空間中で指し示されたユーザの指先位置を推定するFinger-pointerと

いう手法が紹介されている。しかしながら、この手法では手の姿勢がそれほど大きく変化しない場合が想定されており、認識される手身ぶりも比較的単純なものであった。また最近では、単眼画像のみから手の3次元位置情報を推定する興味深い手法が Segen らによって紹介された¹¹⁾。この手法では手の影をうまく利用することにより、単眼画像のみから2眼画像を利用したのと同じように手の姿勢を推定している。

また別のアプローチとして、リンク機構で近似される手の3次元モデルに基づく手法も提案されてきている。このアプローチでは入力中の手領域の2次元的な幾何特徴のみを用いるのではなく、観察された手領域に手の3次元モデルがあてはまるように手の3次元位置・姿勢や各関節角度を推定するという処理が行われる。このような研究の例としては、Rehg らの手法¹²⁾、岩井らの手法¹³⁾、島田らの方法¹⁴⁾などがある。これらの手法では手の3次元モデルを利用することにより、各指の関節角度など細かな手形状を推定することが可能であるという面で有利な一方で、パラメータ推定などに必要となる数値計算コストが高くなってしまい実時間処理が難しいという課題や、セルフオクルージョンなどにより安定に手形状を推定するのが難しい場合が存在するという課題などがある。

このようにユーザの手振りのインタラクションへの利用が試みられてきており、ここで紹介した研究例以外にも多くのものが存在する¹⁵⁾。しかしながら、自然なインタラクションを実現しようとする場合に、背景などの環境を限定することなくユーザの手の3次元位置や形状を比較的安定にかつ実時間で推定することが可能であることが必要不可欠となり、これまでの手法では必ずしもすべての条件を満足するものがなかった。そこで本研究では、実際のアプリケーションへの応用をふまえ、手振りによる自然なインタラクションを実現するのに適した手法を紹介する。

本研究で提案する手法は、多視点画像をもとに安定に手の3次元位置・姿勢を求めるという点、複数の手形状を信頼性良く判別可能、リアルタイム処理の実現という点などにおいて、内海らの手法^{16),17)}と関連が深い。内海らの手法では、5台のカメラからの映像をもとに手の3次元位置・姿勢を推定し、入力画像を選択的に用いながらP型フーリエ記述子で近似された手領域の輪郭形状をもとに複数の手形状の判別をリアルタイムで行っている。実際に、3次元仮想物体の操作などへの応用も報告されており、実際のアプリケーションへの応用という面からも有効であると考えられる。しかしながら、不特定ユーザによるさまざまな手

形状への拡張性などの面が明らかにされているとはいえない。また、報告されているシステムでは毎秒約10フレームの処理速度が実現されているが、自然なインタラクションを考えた場合にはビデオフレームレートもしくはそれに近い処理速度が望ましい。本研究では、これらの点についてより有効な手法を提案する。

3. Perceptual Glove の実現手法

本研究では、より自然な HCI のための入力手段として、ユーザの手の3次元位置・姿勢の計測ならびに手形状の判別処理を行うための手法を実現することを目指す。特に、本研究における提案手法では以下にあげるような特徴を重視している。

- (1) 複数カメラからの入力画像をもとに、手の3次元位置および姿勢を安定にかつ正確に計測することが可能であること。
- (2) ビデオフレームレートもしくはそれに近い更新による実時間処理が実現されていること。
- (3) 複数の手形状を安定に判別することが可能であり、不特定のユーザや新たな手形状に対して拡張が容易であること。

本研究における提案手法は図1に示すように大きく分けて入力画像中における手領域の抽出、手の3次元位置ならびに姿勢の推定、手形状の判別、の3つの処理により校正される。

まず、Perceptual Glove の前処理として、2台のカラーカメラからの入力画像それぞれにおいてユーザの手に対応する画像領域を抽出する。このような画像領域の抽出には背景差分や色に基づく領域抽出が一般的に用いられている。本研究では通常のオフィスなどの室内環境(図1)を想定するために、入力画像には複雑な背景や計算機端末上の表示など動的に変化する領

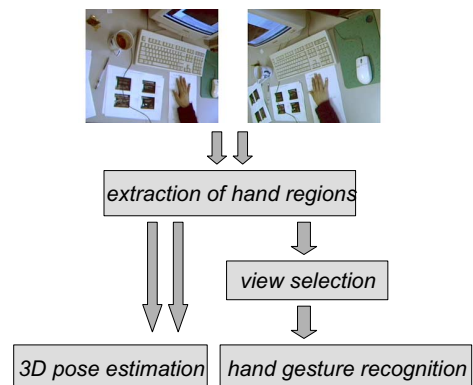


図1 Perceptual Glove の概略

Fig. 1 Overview of Perceptual Glove.

域が含まれることが想定される．一方で，照明条件などはそれほど大きく変化せず比較的安定であることが期待される．

そこで，提案手法では以下で述べるように色に基づくアプローチにより手領域の切り出しを行うこととした．入力画像としては YUV カラー画像 (256 × 220 画素) として取り込まれたものを HSV (色相, 彩度, 明度) カラー画像に変換したものをを用いる．最初にメディアンフィルタを適用することにより画像ノイズによる影響を軽減した後，彩度の十分高く，かつ色相が対象とする肌色の色相に一致する領域を特定する．一般にはさまざまな外乱により複数の領域が抽出されるため，ここでは候補となる複数の連結領域の中で最大のものをもって求める手領域とする．以上の処理により，机上にさまざまなオブジェクトがある場合にも，特定の背景色などを用いることなく手に対応する画像領域を比較的安定に検出することが可能となる．

3.1 多視点画像からの手姿勢推定

2台のカラーカメラからの各入力画像中で手領域を求めた後に，それらの手領域の幾何特徴より手の3次元位置ならびに姿勢を推定する．このために2台のカラーカメラはあらかじめ Tsai のカメラキャリブレーション手法¹⁸⁾を用いて各々のカメラパラメータを求めておく．

このカメラパラメータをもとに，各々の入力画像に対する投影中心 P_1, P_2 (図2) が決定される．次に，入力画像中における手領域の重心位置をもとにして，各入力画像の投影中心 P_1, P_2 から手の3次元位置への方向ベクトル V_1, V_2 が決定される．これらの値から手の3次元位置は以下の式で表される直線 L_1, L_2 の交点として求めることができる．実際には手領域抽出の際の誤差などにより L_1, L_2 は交差しないため，2直線への最近接点として手の3次元位置を求める．

次に，手の3次元姿勢として図2に示す Roll, Pitch, Yaw の3つの回転角度を推定することになるが，ここでは3次元位置を求めるために利用した手領域の重心位置に加え，手領域の上端点・右端点・左端点を利用する．手領域の重心位置から手の3次元位置を求めたのと同様にして，それら3つの点の3次元位置を求める．そして，これらの重心点と上端点より手の主軸方向すなわち Pitch と Yaw を求め，右端点と左端点の3次元位置から Roll を求める．

3.2 ニューラルネットワークによる手形状判別

手の3次元位置・姿勢だけでなく，その手によるジェスチャを判別することができれば，簡単なコマンドの実行やユーザの意図理解に基づくインタラクションを

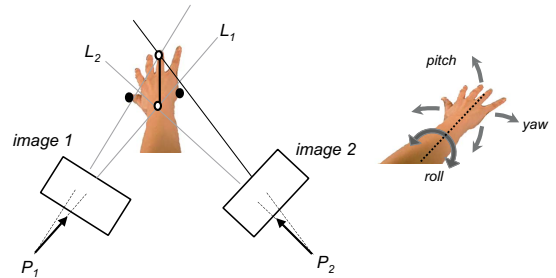


図2 姿勢推定のための3次元位置推定

Fig. 2 Estimation of 3D position and orientation.

実現するのに有用であると考えられる．そこで，提案手法では手によるジェスチャの判別を実現する．なお，ここでは手の動きによる動的なジェスチャではなく，静的なジェスチャの判別を対象としている．具体的には，入力画像中の手の形状があらかじめ定義される手の形状，たとえば closed, open, pointing などの形状のどれに相当するかを実時間で判別することを考える．

手のジェスチャ判別においては，複数利用者間あるいは同一利用者であっても手の形状や動作速度などが必ずしも一定とは限らない．そのため，本研究における提案手法では複数の手形状を安定に判別することに加え，不特定のユーザや新たな手形状に対して拡張が容易であることが重要であると考えられる．

このような性質を実現すべく，ニューラルネットワーク¹⁹⁾を用いて手形状の判別を行うこととした．ニューラルネットワークはさまざまな種類の信号を入力とするパターン識別に非常に有効であることが知られており，音声や画像などの識別に広く用いられている．ニューラルネットワークの持つ優れたパターン識別能力に加え，計算コストの観点からもニューラルネットワークの利用は有効である．一般に，教師信号付の学習データをもとにネットワークの重み係数を逐次修正していく学習段階の処理はその計算コストが非常に高いものとなりうる．しかしながら，いったん学習が完了した後は，新たな入力データに対して，その手形状を判別するための識別処理は非常に高速に実行することが可能となる．このようなことから，実時間処理が前提となるインタラクションへの利用に適しているといえる．今回使用したニューラルネットワークモデルは，図3に示すように3層構造を持つニューラルネットワークである．また学習アルゴリズムにはバックプロパゲーション法を使用した．

手のような多関節物体を観察する際に，1台のカメラにより撮影される画像ではオクルージョンのためにその形状を正しく判別することは困難となる．そこで，

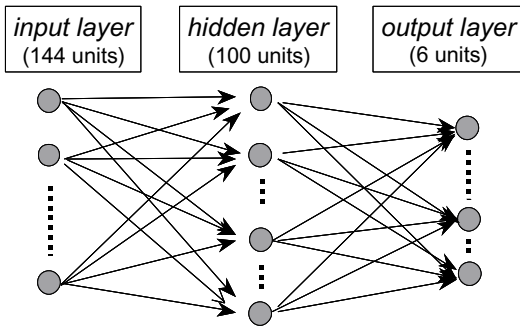


図3 ニューラルネットワークモデル
Fig. 3 Neural network model.

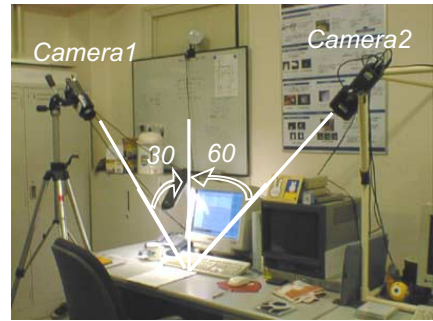


図5 システム概観
Fig. 5 Overview of system.

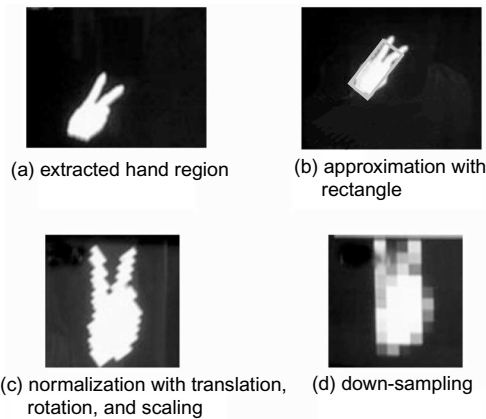


図4 正規化処理
Fig. 4 Normalization of input image.

図1に示すような複数枚の画像から手形状の判別に最も適しているものを選択的に利用する¹⁶⁾。提案手法では、手領域の画像モーメントから計算される手領域の面積ならびにアスペクト比に基づき²⁰⁾、面積がある一定の閾値以上であればアスペクト比が1に近いものがニューラルネットワークへの入力として適した画像として選択される。

このようにして選択された画像は、次に述べるような正規化処理によりニューラルネットワークへの入力データとされる。この正規化処理は、提案手法で利用しているニューラルネットワークでは入力データの位置ずれ、回転、スケールの変化などによって識別結果が影響を受けてしまうために、任意の手の位置、方向においても手形状判別を問題なく行えるようにするために必要となる。

正規化処理は手領域の重心位置、主軸方向、それにアスペクト比(図4(b))に基づき以下のように行う。まず、手領域の重心位置が画像中央にくるように領域を平行移動し、領域の主軸方向が画像座標軸のいずれ

かと平行になるように回転する。次に、手領域の縦横長さが一定の長さとなるように画像軸方向に拡大縮小する。このようにすることにより、手領域が画像中のどこに存在した場合にも、つねに一定の大きさ、一定の回転方向で、アスペクト比が1の入力画像へと正規化される(図4(c))。

このようにして得られた手領域画像は、そのままではニューラルネットワークへの入力データとしてデータ量が多すぎるため、平滑化および間引き処理により画素数を削減したのちに簡略化した画像の各画素の濃淡値をニューラルネットワークの各入力ユニットの値とする(図4(d))。これは計算コストの面以外にも、照明条件などの変動などによる入力データの変動に影響されにくくするというメリットもあわせ持つ。この際に、あまり解像度を低くしてしまうと手形状の識別能力が低下し、あまり解像度を高くしてしまうと学習段階と識別段階の両方において計算コスト的に不利になるため、どの程度まで解像度を落とすのがよいかという点が問題となるが、提案手法では手形状の識別能力と計算コストとのバランスと考慮した上で経験的に 12×12 画素とし、ニューラルネットワークの入力層のユニット数を144としている。

4. 実験

4.1 システム構成

提案手法による手の3次元位置・姿勢推定ならびに手形状判別に関する評価実験を行った。図5に示すように、2台のカラーカメラは手から約1.5mの位置に配置され、手の位置を中心としてそれらのカメラの視線方向がなす角度は約90度とした。この評価実験ではユーザの手の方向(回転主軸)がスクリーンに向けられていることが多いと想定されたため、手の主軸方向と直交する平面上に2台のカメラを配置している。

2台のカラーカメラからの入力画像は2枚の画像処

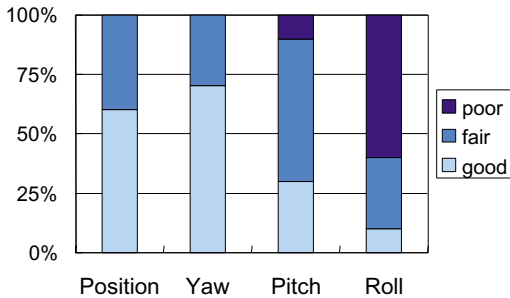


図 6 手の 3 次元位置・姿勢推定結果

Fig. 6 Performance evaluation of estimation of 3D position and orientation.

理ボード (日立: IP5005) を備えた PC (PentiumII 450 MHz, WindowsNT 4.0) によって取り込まれ, およそ毎秒 20 フレームの速度で 3 次元位置・姿勢推定および手形状判別処理が行われる.

4.2 3 次元位置・姿勢推定の評価

本実験では, 提案手法による手の 3 次元位置・姿勢の推定に関して, 日常的に計算機を利用している大学院生 10 名を被験者として主観評価実験を行った. この実験では, 計算機画面上に 3 次元 CG で描画された手を被験者の手の 3 次元位置・姿勢により直接動かしながら, どの程度直感的な操作ができるかということの評価した. まず最初に, 全員の被験者にインタフェースに慣れてもらうために約 2 分間, 自由にシステムを操作してもらった. その後, それぞれの被験者に 1 分間ずつシステムを操作してもらい, 実験が終了後, 直感的な操作が行えたかという点について, 良い, 普通, 悪い, の 3 段階で主観的に評価してもらった.

その結果, 図 6 に示すように, 3 次元位置の推定については被験者の意図に沿った, 微妙な操作が可能であることが分かった. また, 3 次元姿勢推定に関しても Yaw と Pitch 方向の回転についておおむね良好な結果を得ることができた. 一方, Roll 方向の回転については低い評価となっている. これは, 実験において Roll 方向に何度程度回転しているということは検出できるが, その回転方向を誤って推定することが起きていたことが主な原因であると思われる. しかしながら, このような現象は Roll 方向に大きく手を回転させた場合にのみ生じており, そのような操作が必要とされない作業内容に対しては本システムでの手の動作範囲で十分であり, かつ実用的な精度を兼ね備えていると思われる. もしさらに Roll 方向の回転範囲を広げることが必要な場合には, カメラの台数を増やすことで対応できるものと考えられる.

なお, 本論文では予備的な主観的评价実験のみを報

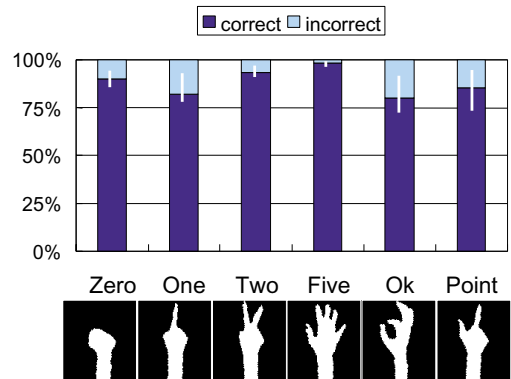


図 7 手形状判別結果

Fig. 7 Recognition of hand shapes.

告する. 定量的な評価に関しては Polhemus などのセンサにより実際の手の 3 次元位置・姿勢を計測する必要があり, 本実験では行っていない. これについては今後の検討課題としたい.

4.3 手形状判別の評価

手形状判別に関する評価実験では, 図 7 に示す 6 種類の手形状を用いて実験を実施した. 学習用のデータとしては, 4 名の被験者それぞれに図に示す 6 種類の手形状を提示してもらい, 各形状に対して 20 枚の画像を記録した. このようにして集めた合計 480 枚の画像データを用いてニューラルネットワークの学習を行った.

その後, 学習データを提供した 4 名を含む合計 10 名の被験者により, 提案手法による手形状判別に関する評価実験を行った. 各被験者は 6 つの手形状を順に 10 秒ずつ提示するように求められ, 提示するように求められた手形状とシステムにより判別された手形状がどの程度一致するかを評価した. 本システムは毎秒 20 フレームの処理速度を持つため, 各被験者に対して 1 分間で 1,200 回程度の識別データを得た. 図 7 に手形状の判別結果を示す. 図中のグラフの中にある白い線は正解率の分散を示している.

この実験の結果, 手形状が正しく判別された率は全体で 85% 程度となっており, 各被験者について約 1200 フレームの全入力画像のうち平均して 85% のフレームにおいて正しく手形状が判別されていることが分かる. 特に, いくつかの手形状に関しては, ほぼ 100% の正解率が得られている. 一方で, 誤認識の主なケースとしては, セルフオクルージョンに起因すると思われる誤認識や, 別のサインに移行する場合の誤認識などが見受けられた. 認識率が低かった Ok と Point については, 手形状が個人によって大きく異なったため, 被験者により提示された手形状が学習データに含まれて

いる手形状と異なっていたものと思われる。これに対しては、学習データとして利用する手形状の画像を増やすことで正解率を向上させることが可能であると考える。

One についても他のものよりも誤認識率が高いが、原因としては Two と形状が類似しているため、手の姿勢によってはこの 2 つの手形状を判別することが難しかったことが考えられる。このようにセルフオクルージョンに起因する誤認識は、追加のカメラを用いることにより視点選択の範囲を広げることで回避できるが、個々の手形状によるセルフオクルージョンの程度と必要な視点数の関係についてはさらに検討する必要がある。

5. 応用例および考察

前章までに述べた手の 3 次元位置・姿勢推定ならびに形状判別手法を利用し、仮想空間内において対話的な操作を可能とする実時間システムを構築しその有効性を評価した。具体的には、ユーザの手による 3 次元仮想物体の直接操作および没入型仮想空間内におけるナビゲーションを例に、提案手法がどの程度直感的な作業環境を提供できるかという点に関して主観的評価実験を行った。

5.1 机上における 3 次元物体操作

さまざまなアプリケーションを考えた場合、モード切替えが頻繁に必要なものや直接値を入力するのではなく曖昧な感覚を利用して操作を行いたいものなどは、提案手法により実現される手振りによるインタラクションが特に有効であると期待される。

ここでは、こうした要素を含んだアプリケーションの例として仮想空間における 3 次元物体の直接操作を考える。利用者は手振りによってコマンドを発し、仮想物体の配置・拡大縮小・回転などの操作を行うことができる。具体的な操作を表 1 に示す。なお、本システムでは手形状の遷移時などに手形状が誤判別されるのを軽減するため、過去 5 フレームのうち 2 フレーム以上で同一形状と判別されないと新しい手形状に遷移しないこととした。

このアプリケーションを例として、提案手法による 3 次元物体の直接操作がどの程度直感的に行えるかということの評価するために、10 名の被験者による主観的評価実験を行った。ここでは 4.2 節で述べた評価実験と同様に、まず操作インタフェースに慣れることを目的にすべての被験者に約 2 分の間自由にシステムを使ってもらった。その後、被験者に操作実験として 1 分間、自由に 3 次元物体を操作するように指示した。

表 1 3 次元物体操作ジェスチャ& モード対応表
Table 1 Control modes used for 3D object handling.

認識ジェスチャ	対応モード
zero	モードなし
one	回転 (Yaw)
two	回転 (Pitch)
five	移動
point	拡大・縮小
ok	把持

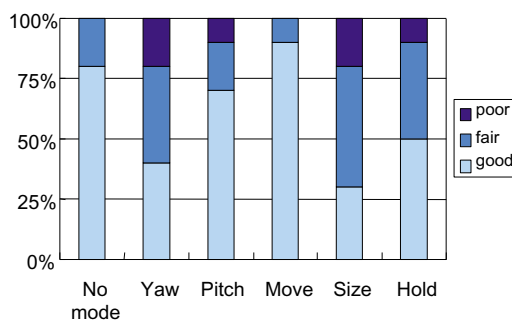


図 8 3 次元物体操作における評価結果

Fig. 8 Performance evaluation of 3D object handling.

その結果、どの程度直感的に操作することができたかということについて、主観的に良い、普通、悪い、の 3 段階で評価してもらった。

この主観評価実験の結果を図 8 に示す。モードなし状態においては手の 3 次元位置だけを描写し、対象となる物体には触っていない状態としている。この状態に関しては前章で述べた手の 3 次元位置推定実験の結果とほぼ同様の良い結果を得ることができた。これに対して、回転方向に関しては Yaw, Pitch とともにあまり良い評価が得られていないことが分かる。これは、前章の手形状判別実験において判別が難しいとされた One, Two を物体の回転操作に割り当てたために、回転方向の判定が信頼性良く行えなかったことに起因していると思われる。一方、把持・移動については非常に良い操作感を得ることができた。拡大、縮小に関してもあまり好ましい結果は得られなかったが、これは提案手法の性能に起因するというよりも、操作感として拡大縮小の動作を考えた場合に片手で伸ばすという操作方法に違和感を覚えた人が多いようだった。

全体としては、直感的な操作方法として非常に使いやすかったという意見も多い一方で、コマンドを増やして機能を増やすということや、実際同じ作業をマウスとキーボードで行うことで比較して使用してみたいとの意見があった。

5.2 仮想空間内のナビゲーション

デスクトップ端末に表示される 3 次元物体を手で直

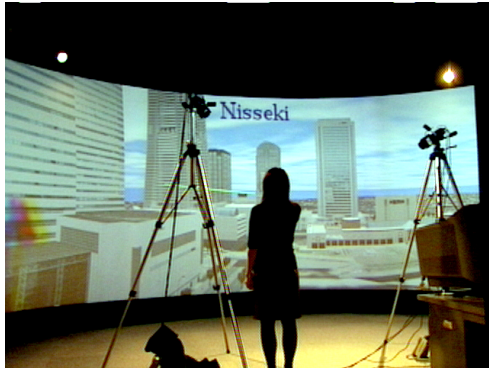


図9 没入型映像空間内ナビゲーションへの応用
Fig.9 Application for indoor navigation.

接操作するという作業に加え、大型立体視映像により表示される没入型映像空間内におけるナビゲーションに対して手による直接操作を応用しその有効性を評価した。

本実験に利用した大型立体視映像表示装置の外観を図9に示す。本表示システムは半径4m、高さ2.7m、水平角度150度の円筒形スクリーンを備え、3台のCRT式プロジェクタによって継ぎ目のない広視野角立体映像の投影を実現している。本実験では、机上のシステムと同様に3次元位置・姿勢推定および手形状の判別を行った後、その結果をソケット通信を用いて、映像生成用グラフィックスワークステーション(SGI Onyx2)に送っている。

本実験では、3次元空間内の移動および対話的なポインティングの2つの作業モードを手による直接操作でコントロールすることが実現されている。具体的には、図7におけるジェスチャのうちFiveがウォークスルーモードに対応しており、このモードでは前進を基本状態とし、手を左右に傾けることにより進行方向を変更できるようになっている。そして、手の3次元位置・姿勢を変えらることにより視点の上下の変更や後退も可能となっている。また、図7に示すPointingのジェスチャを認識した場合にはポインティングモードに変更され視点の移動は停止する。このポインティングモードでは手の位置から指し示している方向に緑色のラインが描画され、そのラインで仮想空間内に存在するビルなどのオブジェクトを指し示すことによりオブジェクトの名称を対話的に表示させることが可能となっている。そして、次にナビゲーションモードに戻すときにはOkのジェスチャをコンピュータに認識させることで、モードの変更を行う。

このシステムを利用して10名の被験者による主観評価実験を行った。前章で述べた3次元物体の直接操

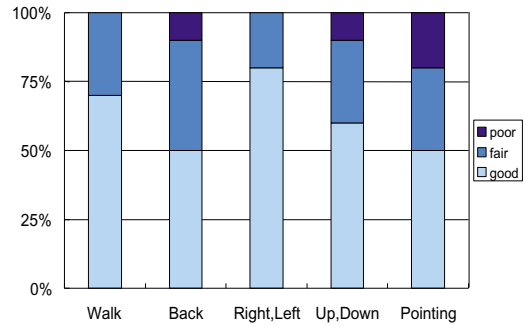


図10 ナビゲーションタスクにおける評価結果
Fig.10 Performance evaluation of indoor navigation.

作の場合と同様に、被験者には最初2分間の練習期間の後に1分間自由にシステムを使用してもらい、実験終了後にどの程度直感的に操作することができたかという点について、主観的に良い、普通、悪い、の3段階で評価してもらった。その結果、図10に示すように手による直接操作により没入型映像空間内のナビゲーションが問題なく行われたことが分かった。本実験では3次元物体操作と比較して良好な結果が得られていることが分かるが、これは手による直接操作を考えた場合に、手近な対象物体を操作するという操作感覚よりも室内空間における移動・ポインティングといった作業に関する操作感覚の方がより自然であると好まれた結果であると考えられる。

6. まとめ

本論文では、2台のカラーカメラからの入力画像をもとに手の3次元位置・姿勢推定ならびに手形状判別を行うための手法を提案した。特に、ニューラルネットワークにより手形状判別処理を行うことにより、新たなジェスチャや不特定のユーザに対しても計算コストの増大を招くことなく容易に拡張可能という面で有利な手法となっている。現在のシステムでは毎秒20フレーム程度の処理速度が実現されており、さまざまなインタラクションへの応用が可能である。また、提案手法による3次元位置・姿勢推定ならびに手形状判別の有効性を評価するために、10名の被験者による基礎的な主観評価実験を行い良好な結果が得られた。

さらに、提案手法による直接操作が具体的なアプリケーションに対してどの程度有効であるかということの評価するために、3次元物体の直接操作ならびに没入型映像空間内におけるナビゲーションという2つの作業を例に、主観的評価実験を行った。この結果、提案手法による3次元位置・姿勢推定の性能は具体的なインタラクション作業に利用する際に必ずしもすべて

の面において十分であるとはいえない一方で、手による直接操作がユーザにとって直感的な作業感覚を実現するのにおおむね有効であることを確かめた。

今後の課題としては、人間の持つ多様な操作モードに対応できるようにするため、手のジェスチャとして静的な形状の識別に加え、動きに基づくジェスチャへと拡張することが考えられる。また、Polhemusなどのセンサを用いることにより提案手法による手の3次元位置・姿勢推定の客観的評価が必要であると考え、さらに、提案手法の高精度化をはかるといふ観点から3台以上のカメラを利用した場合への拡張も重要であると考え、このように提案手法を拡張することにより、人間が直感的に利用できる有用なアプリケーションを実現することを目指す。

謝辞 本研究の一部は、文部省研究プロジェクト：科学研究費・創成的基礎研究「人間主体のマルチメディア環境形成のための情報媒介機構の研究」(09NP1401)の研究助成により行われた。ここに記して謝意を表す。

参 考 文 献

- 1) 黒川隆夫：ノンバーバルインタフェース，オーム社(1994).
- 2) Turk, M.: Perceptual user interfaces, *Comm. ACM*, Vol. 43, No.3, pp.33-34 (2000).
- 3) Zimmermann, T.G., Lanier, J., Blanchard, C., Bryson, S. and Harvill, Y.: A hand gesture interface device, *Proc. ACM Conf. Human Factors in Computing Systems and Graphics Interface*, pp.189-192 (1987).
- 4) Sturman, D.J. and Zeltzer, D.: A survey of glove-based input, *IEEE Computer Graphics and Applications*, Vol.14, pp.30-39 (1994).
- 5) Cipolla, R., Okamoto, Y. and Kuno, Y.: Robust structure from motion using motion parallax, *Proc. 1999 IEEE Intl. Conf. Computer Vision*, pp.374-382 (1993).
- 6) Maggioni, C.: A novel gestural input device for virtual reality, *Proc. 1993 IEEE Annual Virtual Reality International Symposium*, pp.118-124 (1993).
- 7) Cipolla, R. and Pentland, A.: *Computer vision for human-machine interaction*, Cambridge University Press (1998).
- 8) Maggioni, C. and Kammerer, B.: Gesture-Computer — History, design and applications, *Computer Vision for Human-Machine Interaction*, Cipolla, R. and Pentland, A. (Eds.), pp.23-51, Cambridge University Press (1998).
- 9) Sato, Y., Kobayashi, Y. and Koike, H.: Fast tracking of hands and fingertips in infrared images for augmented desk interface, *Proc. 2000 IEEE Intl. Conf. Automatic Face and Gesture Recognition (FG 2000)*, pp.462-467 (2000).
- 10) Fukumoto, M., Suenaga, Y. and Mase, K.: Finger-pointer: Pointing interface by image processing, *Computers and Graphics*, Vol.18, No.5, pp.633-642 (1994).
- 11) Segen, J. and Kumar, S.: Shadow gestures: 3D hand pose estimation using a single camera, *Proc. 1999 IEEE Conf. Computer Vision and Pattern Recognition*, pp.479-485 (1999).
- 12) Rehg, J.M. and Kanade, T.: Visual tracking of high DOF articulated structures: An application to human hand tracking, *Proc. ECCV'94*, pp.35-46 (1994).
- 13) 岩井儀雄, 八木康史, 谷内田正彦: 単眼動画からの手の3次元運動と位置の推定, 信学論(D-II), Vol.J80-D-II, No.1, pp.44-55 (1997).
- 14) 島田伸敬, 白井良明, 久野義徳: 確率に基づく探索と照合を用いた画像からの手指の3次元姿勢推定, 信学論(D-II), Vol.J79-D-II, No.7, pp.1210-1217 (1994)
- 15) Huang, T.S. and Pavlovic, V.I.: Visual interpretation of hand gestures for human-computer interaction: A review, *IEEE Trans. Pattern and Machine Intelligence*, Vol.19, No.7, pp.677-694 (1997).
- 16) 内海 章, 大谷 淳, 中津良平: 多数カメラを用いた手形状認識法とその仮想空間インタフェースへの応用, 情報処理学会論文誌, Vol.40, No.2, pp.585-593 (1999).
- 17) Utsumi, A. and Ohya, J.: Multiple-hand-gesture tracking using multiple cameras, *Proc. 1999 IEEE Conf. Computer Vision and Pattern Recognition (CVPR'99)*, pp.473-478 (1999).
- 18) Tsai, R.Y.: A versatile camera calibration technique for high accuracy machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Journal of Robotics and Automation*, Vol.3, No.4, pp.323-344 (1987).
- 19) 中野 馨, 飯沼一元: ニューロコンピュータ, 技術評論社, pp.12-47 (1989).
- 20) Freeman, W.T. and Anderson, D.B.: Computer vision for interactive computer graphics, *IEEE Computer Graphics and Applications*, Vol.19, No.3, pp.42-53 (1998).

(平成 12 年 4 月 26 日受付)

(平成 13 年 10 月 16 日採録)



齋藤真希子 (正会員)

1998年法政大学工学部システム制御工学科卒業。2000年電気通信大学大学院情報システム学研究科修士課程修了。同年、日本電気株式会社入社。現在、NEC ネットワーク開発研究所に所属。モバイル・ネットワーク、ユビキタス・コンピューティングの研究に従事。ヒューマン・コンピュータ・インタラクションに興味を持つ。IEEE VR2001 Honorable Mention for the Outstanding Paper Award 受賞。



佐藤 洋一 (正会員)

1990年東京大学工学部機械工学科卒業。同大学院情報工学専攻を経て、1997年カーネギーメロン大学計算機科学部ロボティクス学科博士課程修了。Ph.D. in Robotics。同年より東京大学生産技術研究所研究機関研究員、専任講師を経て、現在同研究所助教授。コンピュータビジョン、ヒューマン・コンピュータ・インタラクション、およびコンピュータグラフィックスに関する研究に従事。平成11年山下記念研究賞、MIRU2000 最優秀論文賞、平成11年度日本バーチャルリアリティ学会論文誌論文賞、IEEE VR2001 Honorable Mention for the Outstanding Paper Award, 1997 Int. Conf. Shape Modeling and Applications 最優秀論文賞を受賞。電子情報通信学会、日本バーチャルリアリティ学会、ACM、IEEE 各会員。



小池 英樹 (正会員)

1991年東京大学大学院工学系研究科情報工学専攻博士課程修了。工学博士。同年電気通信大学電子情報学科助手。1994年同大学院情報システム学研究科助教授。現在に至る。1994~1996年、1997年 U.C. Berkeley 客員研究員。情報視覚化の研究に従事。特に視覚化へのフラクタルの応用、Perceptive User Interface、情報セキュリティへの視覚化の応用に興味を持つ。1991年日本ソフトウェア科学会高橋奨励賞、2000年情報処理学会 DICOMO'2000 最優秀論文賞、2001年 IEEE VR2001 Honorable Mention for the Outstanding Paper Award 受賞。ACM、IEEE/CS、日本ソフトウェア科学会各会員。