ポインティングデバイスとしての身体動作

久 野 義 徳†

人間は様々な身体動作で情報を発信できる．ここではそれを主としてポインティングデバイスとしてヒューマンインタフェースに用いることについて述べる．まず，ビデオカメラの画像から手の動きを認識することにより，手の動きで CG 像やロボットを動かすことのできるシステムを紹介する．次に，行きたい方向に顔を向けることにより操縦できる知的車椅子について述べる．この車椅子では，周囲の歩行者の顔を見て，その顔の向きにより車椅子に気づいているか判断し，避け方を変える．最後に，この種のシステムに関して今後の課題を議論する．

Body Actions as Pointing Devices

Yoshinori Kuno†

Humans can transmit information through various body actions. This paper describes several systems using such body actions mainly as pointing devices for human interfaces. Firstly, we show systems that can recognize our hand motions from video images. We can thereby move computer graphic images or robots by our hand movements. Then, we introduce an intelligent wheelchair. We can turn it in a desired direction by turning our face in the direction. It changes the collision avoidance method with an approaching pedestrian by judging whether his/her noticing it from his/her face direction. Finally, we discuss issues concerning such human interfaces.

1. はじめに

人間同士のコミュニケーションでは言語のほかに，視線，表情，ジェスチャなど多数の非言語的行動が重要な役割を果たしている．そこで，コンピュータと人間のコミュニケーションといえるヒューマンインタフェースに非言語的行動を利用しようという研究がさかんになっている[1]．人間の非言語的行動には無意識的・非意図的なものが多いが，コンピュータへの意思伝達に使うという点から，現時点では，意識的・意図的な行動が対象になっていることが多い．すなわち，ジェスチャでコマンドを送ったり，手や視線，あるいは顔の向きで対象を指し示したり，対象の動きを操作するものである．これらは，ジェスチャによるコマンドを除けば，マウスなどによるポインティングにあたる．つまり，身体動作がポインティングデバイスとして使われている．これに関する先駆的な研究は MIT のメディア研究所で開発された “Put-That-There” であろう[2]．ここでは磁気センサからの情報をもとに，操作対象や移動先を手で指し示すことができた．このようなことをビデオカメラの画像データから行うことができれば，装着物も不要で人間の行動を拘束しないものが実現できる可能性がある．そこで，多くの研究が進められている．ここでは，それに関して著者が関与してきた研究を紹介し，今後の課題を議論する．

2. ポインティングデバイスとしての手の動作

指でものを指したり，手で対象の動きを示したりするのは，よく用いられる表現手段である．また，手でものの形や大きさを示すことも多い．これらは直接的なポインティングの動作であり，それをビデオカメラの画像から認識できれば，そのまま使いやすいインタフェースになると期待される．特に，マウスの動きが 2 次元平面上に限定されるのに対し，手は 3 次元空間で動かせる．したがって，3 次元空間を対象とした場合に便利なインタフェースが実現できるのではないかと考えて研究を進めてきた．これはコンピュータビジョンによるジェスチャ認識のヒューマンインタフェースへの応用の研究ということになる．ジェスチャ認識に関しては多くの研究があるが，Pavlovic ら[3]はそれらを技術面からモデリング，解析，認識に分けて整理している．また，応用システムもまとめている．また，Quek[4]もジェスチャに関する用語の定義と関連研究を

† 埼玉大学
Saitama University

自身のジェスチャ認識によるヒューマンインタフェースとともに紹介しており，参考になる．

ジェスチャ認識関連の研究を大きく分けると，コンピュータビジョンの基礎技術を中心としたものと，応用システムの実現を中心としたものに分けられる．手は3次元の関節物体で複雑な形状変化が可能で，コンピュータビジョンの認識対象として興味深い．そこで，手の3次元モデルを用いて，手の形や動作を求めようという研究が多く行われている[5)~7)]．ここでは触れないが，著者の関連したグループでも研究を進めている[8),9)]．しかし，応用指向の研究では，手全体[10)]や指先[11)]を追跡して得られる動き情報と簡単な形状特徴による認識[12)~14)]を組み合わせて用いる（文献は中心的に用いられている方であげている）のが主で，基礎技術で検討されていることと隔たりのあることが多い．ヒューマンインタフェースの応用システムでは，実時間で確実に動作する必要があり，応用ごとに利用できる拘束や知識を活用して，簡単な処理で動作するシステムが実現されている．

我々のアプローチはコンピュータビジョンの基礎技術の検討もするが，どちらかといえば先に述べた分類では応用システム実現の方に入る．しかし，たとえばテレビの操作をする[10)]というような特定の応用の実現のためではなく,（コンピュータビジョンを用いた）ヒューマンインタフェースに共通な課題をコンピュータビジョンを用いて解決することを研究の主眼とする．

Norman はユーザ中心のデザイン[15)]を提唱しているが，ヒューマンインタフェースはユーザである人間を中心に考えるべきである．これまでは，機械の方の都合に人間が合わせる側面があったが，これからは人間の方に機械が合わせるべきだと考えて研究を進めている．コンピュータビジョンによるジェスチャ認識を用いたこれまでのヒューマンインタフェースを見ると，人間中心とはなっていない面が多い．人間は所定の場所に座るか立つかして，カメラの視野に手が入るように注意して，機械の方の都合で定められた不自然で大げさなジェスチャをしなければならない．また，論文には明示的には書かれていないが，一般には使用しないときは手をあまり動かさない方がよいと推察される．音声に比べてジェスチャは技術的に開始の検出が困難なのに，音声と違い，人間は手を意思伝達以外の場合にも（むしろそちらが主目的だが）よく動かす．このような，これまでは人間に使う際に制約が課せられていたのを解消する方法について研究を行ってきた．

2.1 運動視差を利用した CG 像の操作

コンピュータビジョンの応用としてヒューマンインタフェースが有望だと考えて，最初に行った研究である[16),17)]．ここでは，手の回転や並進の3次元運動をロバストに求める方法として，以下に述べるような運動視差に基づく方法を検討した．まず，対象上の同一平面上にない4点を画像上で追跡する．ここで，4点のうち1点を除いた3点で作る3角形を考え，4点目がその3角形上にあると仮定する．そして，ある時刻から次の時刻に3角形が動いたとき，4点目がどこに移動するか求める．実際には，4点目は3角形の作る平面上にないから，その画像上の位置は仮定により求めた位置と異なる．この差が運動視差である．これを利用して，安定に3次元運動を求めるアルゴリズムを提案した．

実際には，手の上に4点を定めて追跡するのは困難なので，4色の色球を手袋につけ，それを追跡することによりリアルタイムで動作するシステムを完成した．そして，1992年のデータショウに手の動きにより3次元 CG 像を操作できるシステムとして参考出品するなど，コンピュータビジョンの技術がヒューマンインタフェースに利用できることを示した．

このシステムを使って気づいたのは，回転運動の指示の際の問題である．3次元空間での並進運動は手の動きで簡単にできる．しかし，回転運動は少しならよいが，大きく，たとえば，対象物を何回転もぐるぐる回したいという場合には問題が生じる．このシステムでは，手の動きをそのまま対象物の動きとしたが，人間の手は，そのようにぐるぐるとは回せない．実際に，人間がそのような意図を伝えるときは，手を所定の方向に回しては戻すことを繰り返す．このシステムでは，そのような動作をされると，対象物は所定の方向に回転したり戻ったりを繰り返すことになる．そのときは，この問題を解決するために棒に球をつけたものを作り，それを人間が持って動かすことにした．実際に使ったところでは，棒の方が操作しやすかった．しかし，コンピュータビジョンを使ったインタフェースとしては，こういう補助具を使わなくてすむようにしたい．そのためには，人間の動作を直接対象の動きにするのでなく，間に人間の意図を理解する処理が必要になる．以上については4章で議論する．

2.2 空間の基準

ビジョンを使ったヒューマンインタフェースの利点の1つは，ケーブルなどに拘束されずに自由な位置で使えるということである．しかし，実際に先のシステムを開発して使用してみると，そのような位置の自由が実際には得られていないことが分かった．ビジョンを使ったシステムでは，当然のことであるが，対象（手）

がカメラの視野に入っていなければならない．ところが，現在のカメラの解像力で認識に十分な画像を得るためには，対象が画像中にかなり大きく写っていなければならない．したがって，カメラの視野はあまり広くできない．そこで，先のシステムでは手がカメラの視野からはずれないように，モニタを見ながら注意して使わなければならなかった．これでは良いヒューマンインタフェースとはいえない．通常のカメラで視野を広げるには，アクティブカメラとして対象を追跡すればよい．しかし，3 次元情報を得るためにカメラのパン・チルトの精度を出そうとすると精密な機構が必要になり，コストの点から問題になる．

さらに，使用者が自由に動くことを考えると空間の基準の問題が生じる．先のシステムではディスプレイ中の CG 像が対象なので，使用者はだいたいはディスプレイの方を向いているので問題はない．しかし，ロボットを手の動きで操作する場合などを考えると，自由な位置で使えるシステムだと，使用者の位置や向きが使用している状況によって変わる可能性がある．このような場合，使用者が空間をどのようにとらえて指示を出しているかを考えなければならない．具体的にいうと，「右」とか「左」を何を基準に考えているかということである．たとえば，ロボットが近くにいて見えているときに，指である方向を指し示した場合，実世界でその向きにロボットに行ってもらいたいと考えるのが普通であろう．この場合は，指の向きを世界座標で考える必要がある．しかし，ロボットが遠くにいて，ロボットに積まれたカメラの映像を見ながら指示を送る場合は，ロボット（その上に積まれたカメラ）を自分の身体と重ねあわせて，位置関係を考えると思われる．すなわち，ロボットを右に動かすときは，自分の身体を基準にして手を右に動かすようなことが自然だと思われる．

以上の 2 つの問題を解決するために，複数視点画像からのアフィン不変量[18]を用いた CG 像や移動ロボットを操作するインタフェースを開発した[19),20)]．

はじめに，複数視点画像からのアフィン不変量について簡単に述べておく．**図 1** に示すように，3 次元空間上に 5 点 $\mathbf{X}_i, i \in \{0,..,4\}$ があると仮定する．それらのうち同一平面上にない 4 点を用いて，$\mathbf{X}_0$ を原点とする基底ベクトル

$$\mathbf{E}_i = \mathbf{X}_i - \mathbf{X}_0 \quad (i \in \{1,2,3\}) \tag{1}$$

を考える．この基底ベクトルを用いると，第 5 点 $\mathbf{X}_4$ は α，β，γ を適当に選ぶことにより，次のように表すことができる．

$$\mathbf{X}_4 - \mathbf{X}_0 = \alpha\mathbf{E}_1 + \beta\mathbf{E}_2 + \gamma\mathbf{E}_3 \tag{2}$$

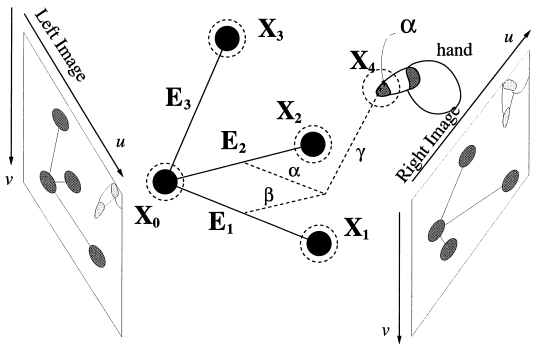


図 1 複数視点アフィン不変量
Fig. 1 Multiple view affine invariance.

ここで，カメラの投影を weak perspective と仮定する．$\mathbf{X}_0, \ldots, \mathbf{X}_4, \mathbf{E}_1, \ldots, \mathbf{E}_3$ を画像上に投影し，投影された座標をそれぞれ $\mathbf{x}_0, \ldots, \mathbf{x}_4$，$\mathbf{e}_1, \ldots, \mathbf{e}_3$ とすると，異なる位置から観測された 2 枚の画像それぞれについて式 (2) と同様に以下の関係が成り立つ．ただし，両画像上での各点の対応は求まっているとする．

$$\left.\begin{aligned} \mathbf{x}_4^l - \mathbf{x}_0^l &= \alpha\mathbf{e}_1^l + \beta\mathbf{e}_2^l + \gamma\mathbf{e}_3^l \\ \mathbf{x}_4^r - \mathbf{x}_0^r &= \alpha\mathbf{e}_1^r + \beta\mathbf{e}_2^r + \gamma\mathbf{e}_3^r \end{aligned}\right\} \tag{3}$$

ここでは，左右（2 台のカメラの配置は任意だが，ここでは便宜上，左右という言葉を使う）の画像上の点それぞれに l，r をつけて区別している．式 (3) では，それぞれが 2 次元ベクトルの方程式であるので，未知数 3 に対して，式の数は 4 である．これを成分で書くと，

$$\begin{bmatrix} x_{4u}^l - x_{0u}^l \\ x_{4v}^l - x_{0v}^l \\ x_{4u}^r - x_{0u}^r \\ x_{4v}^r - x_{0v}^r \end{bmatrix} = \begin{bmatrix} e_{1u}^l & e_{2u}^l & e_{3u}^l \\ e_{1v}^l & e_{2v}^l & e_{3v}^l \\ e_{1u}^r & e_{2u}^r & e_{3u}^r \\ e_{1v}^r & e_{2v}^r & e_{3v}^r \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}$$
$$\mathbf{x} = \mathbf{A}\alpha \tag{4}$$

となる．ただし，u，v はそれぞれ画像上の (x, y) 座標を表すベクトルの要素である．式 (4) を最小二乗法で解くことにより，アフィン不変量 $\alpha = [\alpha\ \beta\ \gamma]^{\mathrm{T}}$ を求めることができる．

また 3 次元方向ベクトルを求める場合には，2 点の 3 次元データを用いてもよいが，2 枚の画像上で対応する 1 点と，3 次元方向を求めたい対象の画像上での方向が分かれば求められる[20)]．

これを利用して CG 像やロボットを操作するインタフェースを開発した．この方法では，基準点が 2 台のカメラに写ってさえいればよい．したがって，機械的に精密なパンチルト機構でなく，簡単なもので特徴点を追跡すれば 3 次元情報が得られる．これで使用者

図 2 ロボットを操作している様子
Fig. 2 Robot operation.

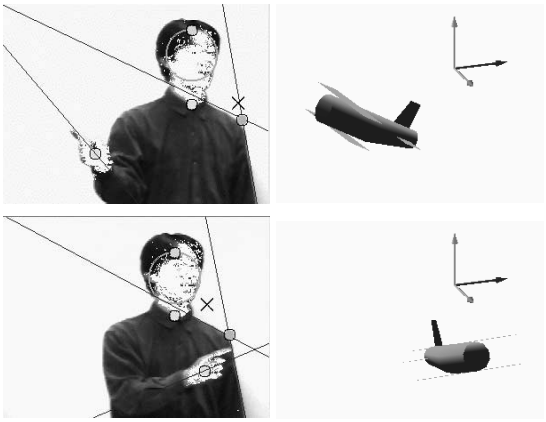


図 3 実験結果
Fig. 3 Experimental results.

の位置の制限の問題が解決できる．そして，基準点をシーンの中の固定物上にとれば，固定した世界座標で，使用者の身体の上にとれば，その人を中心に考えた座標系で空間情報が得られる．

このシステムでも，身体の上の特徴点を追跡するのは難しいので，**図 2** のように身体の上に 4 つの球をつけて基準点とした．また，マーク付きの手袋をはめて手の特徴点とした．

しかし，実際には人体の上に基準点となる同一平面上にない 4 点をとるのは，座った姿勢でないと難しい．すなわち，このままでは本当に立って自由に動くことはできない．そこで，身体上には 3 点だけを基準点としてとり，4 点目は仮想的に 3 点の作る平面の法線上にある点を考える方法を考案した[21]．これにより，使用者は本当に自由に動けるようになった．また，その利点を活かした応用として，液晶プロジェクタの表示画面内の CG 像を動かすシステムを開発した．**図 3** にそのシステムによる実験例を示す．図中の○が画像から求めた基準点で，×が仮想基準点である．図の左側の動作により右側に示されたように CG 像が表示される．

2.3 操作を意図した手の動きの選択

これまでに紹介したものは，手の動きに応じて対象物を動かすのが主な機能であったが，これに加えて，ものをつかんで置いたり，さらに両手で持って伸ばしたり縮めたりをジェスチャで行えるようにしたシステムを開発した[22]．しかし，これらのものには共通の欠点がある．それは，操作を意図しないで手を動かした場合にも，対象物を動かすジェスチャと見なされてしまう可能性があることである．したがって，操作を意図しないときは手を動かしてはいけないという，使いにくいものになってしまっている．そこで，これを解決するために他の非言語的行動の利用を検討した．ここでは，対象物を操作しようというときはそれを見ているはずであるという仮定に基づき，非言語的行動として視線を用いることにした．すなわち，視線が対象物に向いているときに手を動かしたときだけ，操作の意図があると考えることにした．実際には視線の代用として顔の向きを求めて使用した[23),24]．システムを開発して実験の結果，有効性を確認した．ただし，当然ではあるが，対象物の方を向いていても，それを操作する意図のない場合もある．これを解決するために，対象物を操作しようとしてじっと見ている場合と，画面をぼんやりと見ている場合を顔の向きの変化パターンから識別できないか検討している．しかし，現在までのところ，まだ良好な結果は得られていない．

その他，手の動作を利用するものとして，スライドショーの操作を行うシステムについて検討した．ノート PC にカメラを搭載したものが発売されたとき，そのカメラを活用するものとして，手の動きによりスライドを前に進めたり戻したりできるものを開発した[25]．さらに，液晶プロジェクタで投影した画面をカメラでとらえ，指示棒やレーザポインタなどで画面を指したときに，その位置情報を得られるようにした．これを用いてスライドの前後のほかに，スライドの一覧を出して，その上で指示したものを表示することもできるようにした[26]．これは，直接的なポインティングデバイスとしての手の動作の利用といえる．

3. ポインティングデバイスとしての顔の向き

手のジェスチャのほかにポインティングデバイスとして検討されているのは視線，あるいはその概略の情報と考えられる顔の向きである．視線情報のヒュー

マンインタフェースの利用に関しては視線の測定法とともに大野が詳しく論じている[27]．視線の主な利用法は画面上のメニューやアイコンの選択である．その際，見たものをすべて金に変えてしまったフリギア国の Midas 王の話にちなんで “Midas Touch Problem” と呼ばれる問題がある[28]．すなわち，見たものがすべて選択されるのでは，選択の意図のないときは画面上を見ることができなくなってしまう．そこで，見たものを本当に選択したいのか確認するための手段が必要になる．これには一定時間以上の注視を用いること[28),29)]などが提案されているが，使用者に負担の少ない方法とはいえない．そこで，大野は注視の必要がなく，高速に負担もなく選択できる方法として，Qucik Glance Selection Method を提案している[30]．これは画面上で選択領域と情報提示領域を明示的に区別して，選択領域を見たときにはただちに選択が行われるようにしたものである．

このような改善手法も提案されているが，いずれにしても画面上のオブジェクトを選択するためには正確な視線情報が必要であり，近赤外 LED による投光などの補助[31]を使わずにビデオカメラ画像だけからそれを得るのはかなり難しいと考えられる．また，もともと手でものを指すのは自然な動作だが，意識的に顔や目を動かしてものを指すことはあまりない．顔や目は見たいものがある方向に，それを見るために動かすものである．それを他者が見ると，その人が何に注目しているかという情報が得られる．そこで，顔の向き程度ですむ概略の視線情報で，かつ意識的な細かいポインティングとは違う使い方で有効なものはないかという方向で研究を進めている．その 1 つが，前章の最後に述べた，意図的に手を動かしているときの抽出である．なお，この顔の向きについてもジェスチャ認識の場合と同様に，それを得るコンピュータビジョンの基礎技術よりも，応用の観点から一般的なものに広がる技術を検討するというアプローチで研究を進めている．

3.1 使用者の顔の向きの利用

顔の向きのヒューマンインタフェースへの応用として，使用者の顔の向きで操縦できる知的車椅子を開発した[32)~34)]．使用者が行きたい方向を向けばそちらに回転する．これは意識的に行わなければならないが，目的の程度回転すると，使用者はほとんど無意識的に顔を正面に向ける．これが回転を止めることになる．自動車のハンドル操作では，回転の終了のために意識的に回転と逆方向にハンドルを回すという操作が必要になるが，そのような意識的な操作の負担が軽減される．実際に病院でリハビリテーション中の車椅子利用者に乗っていただいたときも，事前の指示は「顔の向きに曲がります」というだけで，5 人の被験者全員が操作でき，この点では有効性が確認できた．

図 4 知的車椅子の外観
Fig. 4 Intelligent wheelchair.

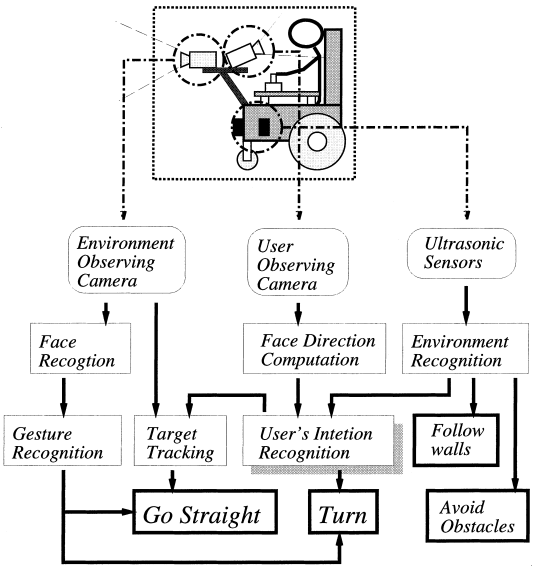
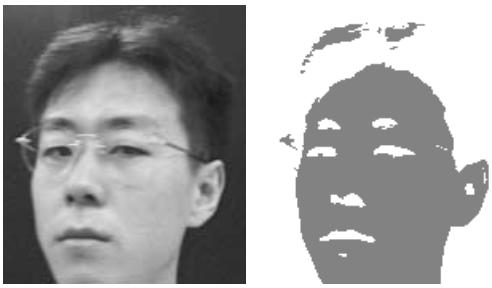


図 5 システム構成
Fig. 5 System configuration.

知的車椅子の外観を**図 4** に，構成を**図 5** に示す．センサとしては，搭乗者を見るカメラと外部を見るカメラ，それに 16 個の超音波センサがある．搭乗者を見るカメラから**図 6** に示すようにして顔の向きを求める．入力画像 (a) から明るい領域を求め (b)，そこから雑音的な部分を除き顔領域を求める (c)．図中の縦線は顔領域の重心を通る直線である．最後に顔領域の中の



(a) 入力画像 (b) 高明度領域

(c) 顔領域 (d) 顔特徴

図 6 顔の向きの計算
Fig. 6 Face direction computation.

目や口などの顔の特徴に対応する暗い部分を抽出し，それら全体の重心を求める (d)．図中の細い縦線はその重心を通る直線である．顔全体の重心と顔特徴の重心のずれが，概略の顔の向きに相当する情報になる．

このシステムでは顔の向きは，ほぼフレームレートで求められる．しかし，顔はつねに少しは動いているので，顔の動きで車椅子を直接制御しては動きが安定しない．そこで，細かな動きには反応せず，意識的に顔を動かしたときにだけ反応するように，あるフレーム数のデータを平均して平滑化して使用する．しかし，どの程度平滑化するかが問題である．回転を意図しない細かな動きに反応しないようにするには平滑に用いるフレーム数を多くしてやればよいが，そうすると回転させようというときにも，なかなか回転を始めず，操作感が悪くなる．実験の結果，1 つの固定値では難しいことが分かった．しかし使用者の感想から，左右に曲がるときは意識して顔を動かすので平滑の程度を大きくしても操作感はそれほど悪くないが，回転を止めるとき，すなわち，顔を正面に戻すときは，ほとんど無意識的な動作のためすばやく動かすので，そのときに反応が遅いと操作感が悪いということが分かった．そこで，左右へ曲がるときは平滑の程度を大きく，中央に戻すときは小さくすることにした．

しかし，顔をゆっくりと動かしても，そちらに曲がるつもりでないこともある．たとえば，壁に貼ってあるポスターなどを走りながら見るような場合は，壁の方を向いていてもそちらに曲がりたいのではない．また，一般に何かが近付いてきたら，そちらを見るのが普通である．この場合も，そちらに曲がりたいわけではない．そこで，超音波センサのデータを用い，近くに物体がある場合は，その方向への顔の向きの平滑の程度を大きくするようにした．すなわち，少しそちらを向いただけでは曲がらないようにした．ただし，近くに物体がある場合にも，そちらに本当に行きたい場合もあるので，じっとそちらを向けば曲がるようになっている．

以上は，2.3 節で検討したのと同様な操作を意図した部分の検出の問題になる．こでは，人間の行動は環境条件により限定されるということを仮定して，この問題を低減している．

3.2 周囲の人の顔の向きの利用

人間は人混みの中でも相手を観察してうまく避けて進んでいくことができる．この場合の観察の対象はおもに視線や顔である．その観察により，相手がこちらを見ていないようなら，こちらから避けるようにする．相手がこちらに気づいているようなら，観察を続け，互いにどう避けるか考える．ときには同じ方向に避けて，まずい場合もあるが，相手を観察することでかなり良い障害物回避を行っているといえる．

このような機能を知的車椅子にも実現しようと研究を進めている[35),36)]．これまでは機械のインタフェースというと，使用者の利便や快適さしか考えていなかった．しかし，車椅子のように使用者以外の人間ともかかわるようなものでは，使用者以外の周りの人間に対しても快適なものである必要があると考えての研究である．

はじめに予備的な検討をするために，車椅子が多く走行する病院の廊下の様子をビデオで撮影し，その映像から車椅子と人間の間の回避パターンを分析した．その結果，以下の 3 通りの場合があることが分かった．医師，看護婦，歩行に支障のない患者・見舞客など，車椅子より速く動く人たちが車椅子に気づいている場合，歩行が困難で車椅子より遅い動きの人の場合，車椅子に気づいていない人の場合である．最初の場合には人間の方が避けるのが普通で，後の 2 者の場合には車椅子の方が避ける．すなわち，車椅子としては相手の速度と車椅子に気づいているかの情報があれば，相手に対して適切な回避行動ができることになる．

そこで，速度については超音波センサにより求め，また，気づいているかについては，顔の向きを観察し，車椅子の方に頻繁に顔が向けられていれば気づいてい

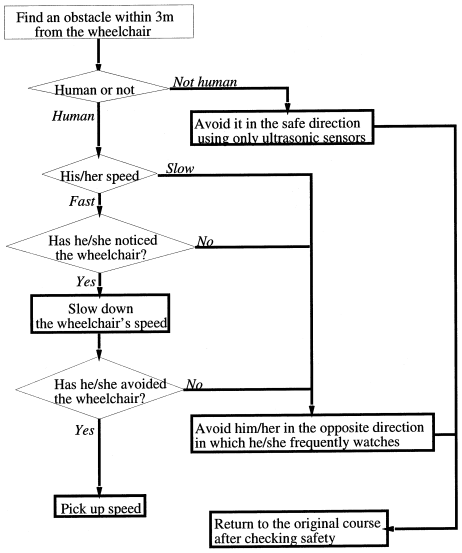


図 7 衝突回避法の決定
Fig. 7 Avoidance method decision.

ると考えることにした．実際に車椅子に種々の回避行動をとらせ，歩行者の快適性を調べた結果，先に述べた第 1 の場合，車椅子の方が気づいていることを相手に示し，相手に避けてもらうことを促すために，少し手前で速度を落とすことが良いことが分かった．車椅子より遅く歩いている人や車椅子に気づいていない人に対しては，車椅子の方から避けるようにする．もちろん，第 1 の場合でも，歩行者の方が避け始めなければ，車椅子の方が避けなければならない．

以上のことを次のようにして実現した．超音波センサで障害物を検出したら，それを人間と仮定して，標準的な人間の大きさと超音波センサの距離データから，カメラのパン・チルト，それにズームレンズの焦点距離を変えて，人間だとしたら顔の周辺あたりになる部分の画像データを入力する．その画像から肌色の検出と目などの顔特徴の検出を行い，顔部分が検出できるなら，その物体は人間だと判定する．そうでない場合は，車椅子は超音波センサのデータを基に障害物を回避する．これには静止障害物のほかに後ろ向きの人間も含まれるが，後ろ向きで車椅子に気づいていないなら，超音波センサのデータに基づいての回避で支障はない．人間と判定した場合は，以後，顔領域の追跡を行う．そして，目の位置から顔の向きを求める．顔が一定の頻度以上で車椅子の方向を示すときは，車椅子に気づいていると判定する．また，超音波センサのデータから対象物の速度も計測する．この 2 つのデータから，**図 7** に示すように衝突回避の方法を決定する．

4. 今後の課題

身体動作のヒューマンインタフェースへの利用についてこの 10 年近くの間，研究してきたことを述べた．これを通じて，この分野で特徴的な検討事項として以下の 3 つの問題があることが分かった．

(1) 意図的部分の検出

マウスやキーボードなら，操作を意図して使用すれば，使用者の意図が伝わるかは別にしても操作した事実はほぼ確実に機械の方に伝わる．しかし，手や顔の動作によるものでは，その点に問題がある．普段あまり現れないような複雑あるいは大げさな動作を用いるなら問題はあまりない．しかし，使いやすいインタフェースということで自然で簡単な動作を用いると，このような動作は機械の操作以外の場合にも現れる可能性があるので，操作を意図したときを識別する必要がある．視線のところで述べた “Midas Touch Problem” もこれに関連するものである．2.3 節では顔の向きにより操作を意図して手を動かした場合を検出するようにした．これはある程度有効だが，顔を対象の方に向けていても，それを操作するつもりでないときもある．また，知的車椅子では顔の動きの速さと周囲環境から操作を意図して顔を動かした部分を検出するようにした．これは有効な方法だが，これで完全なわけではない．身体動作を用いたヒューマンインタフェースでは，このように意図的部分の検出が重要な問題になる．この能力を高めることを考えるとともに，多少の間違いがあっても，システム全体としては十分に有効であるような使い方を考える必要がある．

(2) 比例的動作と記号的動作の切替わり

たとえば手の動作の場合，手で示した形や動きに比例あるいは相似的に対象に指示を与える場合と，手の動作が何らかの記号としてある意図を示す場合がある．2.1 節の最後に述べたように，提案のシステムでは前者の場合しか考えていない．しかし，実際に使用してもらったところを見ると，使用者は両者が存在するということなど意識せずに，対象を少し動かすときは手を操作意図に応じて比例的に動かし，くるくる回したいときは，手の往復動作で表現する．このようなことが，ある程度固定したパターンしかなければ，ジェスチャ認識を行えばよいということになるが，それで十分かどうかは検討の必要がある．結局，第 1 項と同じく，身体動作を単純にヒューマンインタフェースに利用するのではなく，そこから使用者の意図を理解して利用する必要があるということになる．

Quek[4]や Pavlovic[3]はジェスチャの種類をさらに細か

く分類している．現状ではそれらのうちの特定のものしか認識対象になっていないが，今後は，ユーザが切替えを意識せずにそれらを自由に使えるようにする必要がある．

(3) 直接的な操作対象の欠如

ビジョンを用いたシステムの長所の1つは装着物など，使用に際して他の器具が不要なことである．しかし，応用によっては，このことが必ずしも長所にならない場合もある．たとえば2.1節の例では，手でCG像を動かすよりも，実際には棒に球を付けた物体を手に持って，それを動かした方が好まれた．これは手を使用した場合の方が自己隠蔽などのために動きの認識の失敗が多かったのと，前項で述べたように手では大きな回転が指示しにくいためもあるが，対象を動かすのに空中で手を動かすより，対象と見なせる物体を動かした方が使用者にとっては感覚的に扱いやすかったことも理由と考えられる．動かすことをアフォードする物体があった方が，何もなしで手を動かすより良いという，アフォーダンスに関する問題である[15]．

(4) 失敗への対応・回復

これはこの分野に限ったことではないが，ビジョンによるシステムで失敗を完全になくすのは難しい．たとえジェスチャ認識の認識率が高くても，誤ることがある限り，それが問題にならないようにしておかなければならない．まれにではあっても，意思の伝えられない場合があっては，ヒューマンインタフェースとしては問題である．

この問題については，インタラクションによる失敗の回復を検討中である．ヒューマンインタフェースでは，機械の相手として人間が存在する．したがって，その人間を活用して，すなわち人間とのインタラクションにより失敗の回復ができないか検討している．しかし，機械の使用者は一般にビジョンの専門家ではない．そこで，どのようにすれば人間に負担にならない形で，その人が機械がどういう失敗をしているか知り，そしてそれを回復するための有益な情報を与えられるかが問題になる．これについては，機械の方が音声で何が分かって何が分からないのかという現状を伝え，それに対する人間の音声やジェスチャによる反応を認識することにより，問題を解決する方法を検討している[37)~39)]．また，ジェスチャ認識では，未知のジェスチャをされたり，学習したものでも環境の変化により認識が確かでなかったりする場合がある．この場合，ジェスチャの意味を推定し，その結果に基づく行動（推定できないときは可能な行動の中から適当に選択）を少し示し，それに対する人間の反応を見て推定が正しかったかどうかを判定するロボット用のインタフェースも検討している[40)]．

以上，本文で述べた研究を通じて気づいた問題について述べた．そのうち，最初と最後の問題に関しては研究を行ってきたが，今後さらに進展させる必要がある．他の問題については，今後の課題ということで，実際の検討はまだ行っていない．最後の問題の解決法として音声の利用を研究中であると述べたが，このようなマルチモーダルインタフェースが上記の他の問題の解決にも有効な方法であると考えている．この種の研究の発端になった“Put-That-There”[2)]も音声とジェスチャを用いたマルチモーダルインタフェースであったし，ほかにも初期の研究からマルチモーダルインタフェースは検討されていた[41)]．そして，現在も活発に研究されている[42)]．黒川は[1)]マルチモーダルインタフェースからさらに進んだものとして，モードを意識せずに使えるモードフリーインタフェースを目指すことを提案しているが，これも重要な方向であると思われる．

ここでは，コンピュータビジョンのヒューマンインタフェースへの応用の観点からの著者らのグループの研究を述べたが，基礎技術ももちろん重要で，それについても研究を進めている．2章のはじめに述べた3次元モデルを用いた手の形状推定[8),9)]や人体の動作解析[43)]について研究を行っている．ジェスチャ認識については隠れマルコフモデル（HMM）がよく使われている．HMMによる動作認識に関しては大和らが興味深いサーベイを著している[44)]が，形が変化しながら動くような複雑なジェスチャではHMMによるモデル化では不十分であると考え，switching linear modelを利用したジェスチャ認識を提案している[45)]．さらに，両手の動作間の関連を考慮して両手のジェスチャを認識するcoupled switching linear model[46)]を検討している．

以上述べたように，本稿では応用面からの研究を中心に述べたが，基礎技術とともに，今後両者とも検討を進めていく必要がある．さらに，マルチモーダルインタフェースの重要性を考えると，ビジョンだけでなく他分野も考慮しなければならず，総合的な研究の必要な分野であるといえよう．

5. おわりに

ポインティングデバイスとしての身体動作ということで，手のジェスチャと顔の向きをコンピュータビジョンの技術で求め利用する研究について述べた．この種のヒューマンインタフェースでは高速に反応しなければ，実際に使用することはできない．実験システムを

作り，想定したような動作は確認できたが，実装の面では速度や安定性に関して何とか使えるレベルにしたという程度である．したがって，本当に他の手段に比べて有効かどうかの定量的な評価の段階までには至っていない．前章で述べた課題に加え，この点も今後の課題である．

謝辞 本研究の一部は科学研究費補助金（07650492, 09221217, 09555080, 12650249, 13224011）による．

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) Bolt, R.A.: Put-That-There, *Computer Graphics*, Vol.14, No.3, pp.262–270 (1980).
3) Pavlovic, V.I., Sharma, R. and Huang, T.S.: Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE Trans. PAMI*, Vol.19, No.7, pp.677–695 (1997).
4) Quek, F.K.H.: Eyes in the Interface, *Image and Vision Computing*, Vol.13, No.6, pp.511–525 (1995).
5) Mochimaru, M. and Yamazaki, N.: The Three-Dimensional Measurement of Unconstrained Motion Using a Model-Matching Method, *Ergonomics*, Vol.37, No.3, pp.493–510 (1994).
6) Rehg, J.M. and Kanade, T.: Model-Based Tracking of Self-Occluding Articulated Objects, *Proc. 5th ICCV*, pp.612–617 (1995).
7) 亀田能成，美濃導彦，池田克夫：シルエット画像からの関節物体の姿勢推定法，電子情報通信学会論文誌 D-II，Vol.J79-D-II, No.1, pp.26–35 (1996).
8) 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三次元姿勢推定，電子情報通信学会論文誌 D-II，Vol.J79-D-II, No.7, pp.1210–1217 (1996).
9) 島田伸敬，白井良明，久野義徳，三浦 純：緩やかな制約知識を利用した単眼視動画像からの関節物体の形状と姿勢の同時推定，電子情報通信学会論文誌 D-II，Vol.J81-D-II, No.1, pp.45–53, (1998).
10) Freeman, T.F. and Weissman, C.D.: Television Control by Hand Gestures, *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp.179–183 (1995).
11) Quek, F.K.H., Mysliwiec, T. and Zhao, M.: FingerMouse: A Freehand Pointing Interface, *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp.372–377 (1995).
12) Maggioni, C.: GestureComputer — New Ways of Operating a Computer, *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp.166–171 (1995).
13) Kjeldsen, R. and Kender, J.: Visual Hand Gesture Recognition for Window System Control, *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp.184–188 (1995).
14) Hunter, E., Schlenzig, J. and Jain, R.: Posture Estimation in Reduced-Model Gesture Input Systems, *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp.290–295 (1995).
15) Norman, D.A.: *The Psychology of Everyday Things*, Basic Books, New York (1988). 野島久雄（訳）：誰のためのデザイン，新曜社 (1990).
16) Cipolla, R., Okamoto, Y. and Kuno, Y.: Robust Structure from Motion Using Motion Parallax, *Proc. IEEE 4th International Conference on Computer Vision*, pp.374–382 (1993).
17) 岡本恭一，ロベルト チポラ，風間 久，久野義徳：定性的運動認識を用いたヒューマンインタフェースシステム，電子情報通信学会論文誌 D-II, Vol.J76-D-II, No.8, pp.1813–1821 (1993).
18) Mundy, J.L. and Zisserman, A. (Eds.): *Geometric Invariance in Computer Vision*, MIT Press (1992).
19) Kuno, Y., Hayashi, K., Jo, K.H. and Shirai, Y.: Human-Robot Interface Using Uncalibrated Stereo Vision, *Proc. 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.525–530 (1995).
20) Jo, K.H., Hayashi, K., Kuno, Y. and Shirai, Y.: Vision-Based Human Interface System with World-Fixed and Human-Centered Frames Using Multiple View Invariance, *IEICE Trans. Information and Systems*, Vol.E79-D, No.6, pp.799–808 (1996).
21) 林健太郎，久野義徳，白井良明：ユーザの位置の拘束のないジェスチャによるヒューマンインタフェース，情報処理学会論文誌，Vol.40, No.2, pp.556–566 (1999).
22) Jo, K.H., Kuno, Y. and Shirai, Y.: Manipulative Hand Gesture Recognition Using Task Knowledge for Human Computer Interaction, *Proc. 3rd IEEE International Conference on Face and Gesture Recognition*, pp.468–473 (1998).
23) 石山智之，久野義徳，島田伸敬，白井良明：視線情報による選択的ジェスチャ認識に基づくヒューマンインタフェース，第 4 回画像センシングシンポジウム講演論文集，pp.175–178 (1998).
24) Kuno, Y., Ishiyama, T., Nakanishi, S. and Shirai, Y.: Combining Observations of Inten-

tional and Unintentional Behaviors for Human-Computer Interaction, *Proc. CHI 99 Conference*, pp.238–245 (1999).
25) 島田伸敬，村嶋照久，久野義徳，白井良明：プレゼンテーション補助のためのジェスチャインタフェース，第5回画像センシングシンポジウム講演論文集，pp.67–70 (1999).
26) 古川大輔，島田伸敬，久野義徳，白井良明：ジェスチャによるプレゼンテーション支援システム，インタラクション2000論文集，pp.53–54 (2000).
27) 大野健彦：視線インタフェースから視線コミュニケーションへ—視線のある環境を目指して，情報処理学会研究報告，Vol.2001, No.87 (2001-HI-95, 2001-CVIM-129), pp.171–178 (2001).
28) Jacob, R.J.K.: The Use of Eye Movements in Human Computer Interaction Techniques: What You Look at Is What You Get, *ACM Trans. Inf. Syst.*, Vol.9, No.3, pp.152–169 (1991).
29) Hansen, J.P., Anderson, A.W. and Roed, P.: Eye-Gazed Control of Multimedia Systems, *Symbiosis of Human and Artifact*, Anzai, Y., Ogawa, K. and Mori, H. (Eds.), Vol.20A, pp.37–42, Elsevier Science (1995).
30) 大野健彦：視線を用いた高速なメニュー選択作業，情報処理学会論文誌，Vol.40, No.2, pp.602–612 (1999).
31) Morimoto, C.H., Koons, D., Amir, A. and Flickner, M.: Pupil Detection and Tracking Using Multiple Light Sources, *Image and Vision Computing*, Vol.18, No.4, pp.331–335 (2000).
32) Adachi, Y., Kuno, Y., Shimada, N. and Shirai, Y.: Intelligent Wheelchair Using Visual Information on Human Faces, *Proc. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.354–359 (1998).
33) 足立佳久，中西　知，久野義徳，島田伸敬，白井良明：顔の視覚情報処理を用いた知的車椅子，日本ロボット学会誌，Vol.17, No.4, pp.423–431 (1999).
34) Nakanishi, S., Kuno, Y. and Shirai, Y.: Robotic Wheelchair Based on Observations of Both User and Environment, *Proc. 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.912–917 (1999).
35) Murakami, Y., Kuno, Y., Shimada, N. and Shirai, Y.: Collision Avoidance by Observing Pedestrians' Faces for Intelligent Wheelchairs, *Proc. 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, CD-ROM (2001).
36) 村上佳史，久野義徳，島田伸敬，白井良明：知的車椅子のための歩行者の顔の観察に基づく衝突回避，日本ロボット学会誌，Vol.20, No.2, pp.206–213 (2002).
37) Takahashi, T., Nakanishi, S., Kuno, Y. and Shirai, Y.: Human-Robot Interface by Verbal and Nonverbal Communication, *Proc. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.924–929 (1998).
38) Cheng, S., Kuno, Y., Shimada, N. and Shirai, Y.: Human-Robot Interface Based on Speech Understanding Assisted by Vision, *Advances in Multimodal Interfaces — ICMI 2000*, Tan, T., Shi, Y. and Gao, W. (Eds.), Lecture Notes in Computer Science 1948, pp.16–23, Springer (2000).
39) Yoshizaki, M., Kuno, Y. and Nakamura, A.: Human-Robot Interface Based on the Mutual Assistance between Speech and Vision, *Proc. Workshop on Perceptive User Interfaces*, CD-ROM (2001).
40) 村嶋照久，久野義徳，島田伸敬，白井良明：人間と機械のインタラクションを通じたジェスチャの理解と学習，日本ロボット学会誌，Vol.18, No.4, pp.590–599 (2000).
41) Koons, D.B., Sparrell, C.J. and Thorisson, K.R.: Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures, *Multimedia Interfaces*, Maybury, M.T. (Ed.), pp.257–276, AAAI/MIT Press (1993).
42) 長谷川修：マルチモーダル対話における視覚の役割とその応用，情報処理学会研究報告，Vol.2001, No.87 (2001-HI-95, 2001-CVIM-129), pp.165–170 (2001).
43) 林健太郎，久野義徳，島田伸敬，白井良明：動的ロバストキャリブレーションによる人体の姿勢復元，電子情報通信学会論文誌D-II，Vol.J83-D-II, No.3, pp.977–987 (2000).
44) 大和淳司，上田修功，和田俊和：動作認識のための状態遷移モデル—HMMの高度化と非HMM手法の成長，人工知能学会誌，Vol.17, No.1, pp.41–46 (2002).
45) Jeong, M.H., Kuno, Y., Shimada, N. and Shirai, Y.: Recognition of Shape-Changing Hand Gestures Based on Switching Linear Model, *Proc. International Conference on Image Analysis and Processing*, pp.14–19 (2001).
46) Jeong, M.H., Kuno, Y., Shimada, N. and Shirai, Y.: Complex Gesture Recognition Using Coupled Switching Linear Model, *Proc. 5th Asian Conference on Computer Vision*, pp.132–137 (2002).

(平成13年12月25日受付)
(平成14年 3 月 8 日採録)

(担当編集委員　八木 康史)

久野 義徳（正会員）

1954 年生．1977 年東京大学工学部電気工学科卒業．1982 年同大学大学院電子工学専攻博士課程修了．同年（株）東芝入社．1987〜1988 年カーネギーメロン大学計算機科学科客員研究員．1993 年大阪大学工学部電子制御機械工学科助教授．2000 年 4 月より埼玉大学工学部情報システム工学科教授．コンピュータビジョンおよびそのロボットやヒューマンインタフェースへの応用に関する研究に従事．工学博士．電子情報通信学会，日本機械学会，日本ロボット学会，計測自動制御学会，人工知能学会，IEEE，ACM 各会員．