多数カメラを用いた両手手振りの検出

内海 章[†] 大谷 淳[†] 中津良平[†]

画像処理により両手の3次元位置・姿勢・形状を検出，追跡するシステムを提案する．画像処理により手の動きを検出する際には，カメラに対する手の姿勢の変化により観測される形状が変化する自己オクルージョンと，一方の手が他方の手を隠してしまう相互オクルージョンが大きな問題となる．我々のシステムは，多視点画像を利用し観測に最適な視点を動的に選択することで，これらのオクルージョンの問題を低減する．左右の手の動きはそれぞれカルマンフィルタで追跡され，追跡結果に基づいて次フレームで利用する視点を選択する．選択された視点のうち相互オクルージョンを含まない画像から得られる特徴量によりフィルタの更新が行われる．手の姿勢は各カメラで得られる距離変換特徴から，手形状は姿勢推定結果に基づいて選ばれる最適視点画像で抽出される輪郭情報から，それぞれ推定される．本システムは仮想空間における物体操作等のユーザインタフェースとして広く応用できると考えられる．

Multiple-camera-based Multiple-hand-gesture-tracking

AKIRA UTSUMI,[†] JUN OHYA[†] and RYOHEI NAKATSU[†]

We propose a method of tracking 3D position, posture, and shapes of human hands from multiple-viewpoint images. Self-occlusion and hand-hand occlusion are serious problems in the vision-based hand tracking. Our system employs multiple-viewpoint and viewpoint selection mechanism to reduce these problems. Each hand position is tracked with a Kalman filter and the motion vectors are updated with image features in selected images that do not include hand-hand occlusion. 3D hand postures are estimated with a small number of reliable image features. These features are extracted based on distance transformation, and they are robust against changes in hand shape and self-occlusion. Finally, a "best view" image is selected for each hand for shape recognition. The shape recognition process is based on a Fourier descriptor. Our system can be used as a user interface device in a virtual environment, replacing glove-type devices and overcoming most of the disadvantages of contact-type devices.

1. はじめに

コンピュータビジョンの技術によって人間と機械を結ぶ新しいインタフェースを実現しようする研究がさかんである．その理由の1つとして，カメラという受動的なセンサを用いることで従来のインタフェースにはない非接触性という特徴をシステムに持たせられることがあげられる．表情や身ぶり手振りといった人間の動作をカメラで撮影することによって計算機に理解させ，人間の意図を計算機に伝達することがこれらの研究の課題である．

人間の動作の中でも，手の動きは日常でも物を運ぶ，指さす，といった場面で自然に使われており，直感的に使うことのできるインタフェースとして有望である．我々はこの点に着目し，片手の動きを画像処理で検出する研究を進めてきた[1),2)]．手は体の他の部位と比べて自由度が大きく，計算機へのコマンド指示に限ってもその表現力が高い．本論文では，我々の開発した両手手振り認識システムについて述べる．両手の動きを検出することでさらに自由度が高まり，応用分野も広がると考えられる．

手の動きを用いたインタフェースについては従来より多くの研究者によって研究が行われてきた．初期の研究の多くは，センサ付きの手袋を用いたものであり，仮想空間操作等のタスクで多くの報告がある[3),4)]．しかしながら，これら接触型の装置は取扱いが容易ではなく，長時間の使用にはそぐわない．

これら接触型の装置の欠点を克服するために，コンピュータビジョンの研究者がめざしたのが，テレビカメラを用いた手の動き検出である[5)~9)]．しかしながら，従来提案されたシステムの多くは画像処理装置に

† ATR 知能映像通信研究所
ATR Media Integration & Communications Research Laboratories

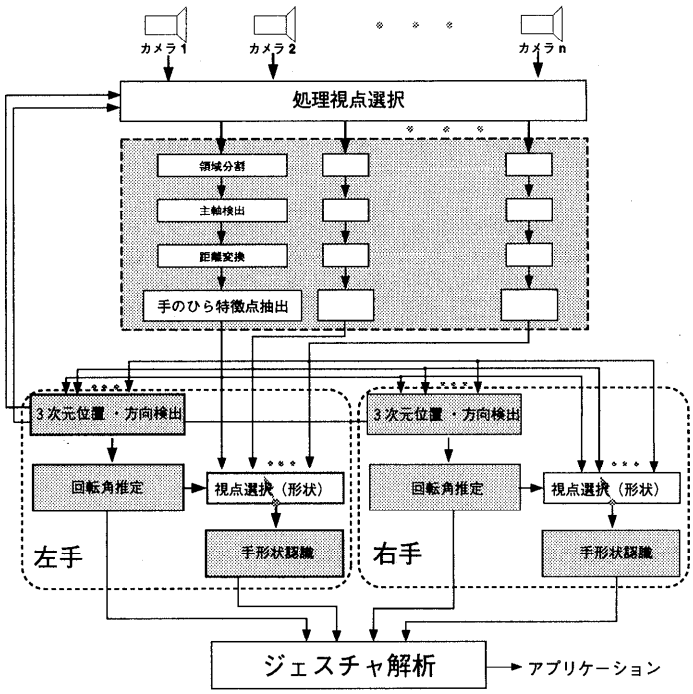


図1 システム構成
Fig. 1 System diagram.

特有の自己オクルージョンの問題に対処していなかった．自己オクルージョンの問題とは，対象物体の動きにともなう観測方向の変化により，得られる画像特徴が大きく変動し，連続的に安定な検出を行うことが困難になることである．我々のシステムではこの問題に対処するために，多視点画像から得られる，手の動き・変形による見え方の変化に対してロバストな画像特徴を使って片手の位置・姿勢を安定に推定する手法を考案した[1]．位置・姿勢を検出できれば，手のひらを正面から観測する最適な視点の画像を用いて手形状認識を行うことができる[2]．

本論文では，この手法を両手の追跡に拡張する．両手の追跡を考えた場合，先ほどの自己オクルージョンに加え，一方の手が他方の手を隠すオクルージョン（相互オクルージョン）が問題となる．我々のシステムではこの問題に対処するため，多数のカメラのうち相互オクルージョンを起こす確率の低いものを動的に選択しながら追跡処理を行う．追跡にはカルマンフィルタを用いる．カルマンフィルタにより両手の追跡処理を行うシステムとしてはすでに Azarbayejani らによる提案があるが[10]，彼らのシステムが固定カメラによって位置のみを追跡するのに対し，我々の手法は追跡結果を利用してカメラを動的に選択し，姿勢・形状をも検出する点が特徴である．

我々は，手振り認識のアプリケーション例として両手の動きによるコマンドベースの仮想シーン操作システムを構築している．本システムでは，ユーザが仮想シーン内に自由にオブジェクトを生成し，それらを操作するという一連のタスクをすべて手の動きによるコマンドで実現している．我々の手振り認識システムの可能性はこれらのアプリケーションに限るわけではなく，人間と計算機のインタラクションを必要とする様々な分野に応用できると考えている．

本論文では，2 章において提案システムの概要と仮定について述べた後，3 章でシステムで用いる観測モデルについて説明する．続いて，4 章で処理画像の選択・特徴抽出処理について述べ，5 章では位置追跡，姿勢・形状推定の手法を実験結果とともに示す．6 章では，本システムを用いたアプリケーションの一例として，手振りを用いた仮想空間操作システムについて述べる．7 章で本論文をまとめる．

2. システムの概要と仮定

本論文で提案するシステムの構成を図 1 に示す．本システムは多視点で得られる時系列画像から両手の動きを追跡する．計算コストを低減するため，全カメラから得られた画像のうち限られた数の画像のみを適応的に選択し処理対象とする．この視点選択は，両手の

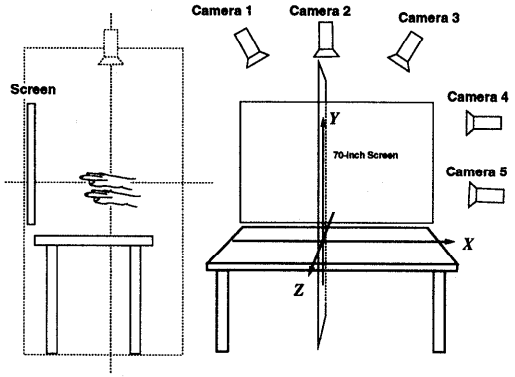


図2 実験環境
Fig. 2 System configuration.

ひらの現フレームでの予測位置に基づいて行う．

選択された各画像から2次元の画像特徴を抽出する．特徴抽出処理は2値化，エッジ検出，距離変換からなる．本システムでは，特徴抽出を安定に行うため，単色からなる一様な背景を仮定した．なお，選択された画像内においても依然としてオクルージョンが発生する可能性は残っている．本システムでは，各画像がオクルージョンを起こしているか否かを特徴抽出時に判別し，オクルージョンが発生していると判断された画像は以後の追跡処理に用いないという方法をとる．これにより，オクルージョンが発生している画像から誤った特徴を抽出して処理が不安定になるのを回避できる．

続いて，両手のひらの予測位置に基づいて画像特徴と追跡モデルの対応づけを行い，位置 (x,y,z)・指先方向 (a,e) の追跡モデルを更新する．追跡モデルは等速運動を仮定している．さらに指先方向まわりの回転角 r を推定し，位置・姿勢の推定結果から形状特徴の抽出に最も適した画像を左右それぞれの手について選択し，形状認識を行う．なお，両手の追跡モデルはそれぞれ検出空間内で左寄り，右寄りに初期位置を持ち，検出空間左側から入ってきた手を左手，右側から入ってきた手を右手として追跡を開始する．

位置・姿勢追跡および形状認識の結果はジェスチャ解析処理に渡され，解析結果がアプリケーションプログラムに送られる．なお，両手間の距離が極端に小さくなりすべての視点でオクルージョンが発生する場合には，本システムによる位置・姿勢追跡および形状認識は不可能となるため，処理を中止し初期状態に戻る．

今回実装したシステムでは，**図2**に示すように観測用に5台のカメラを設置し，これらのうち3台分の画像を選択し同時に処理する．実装システムの処理速度は約10 Hzである．

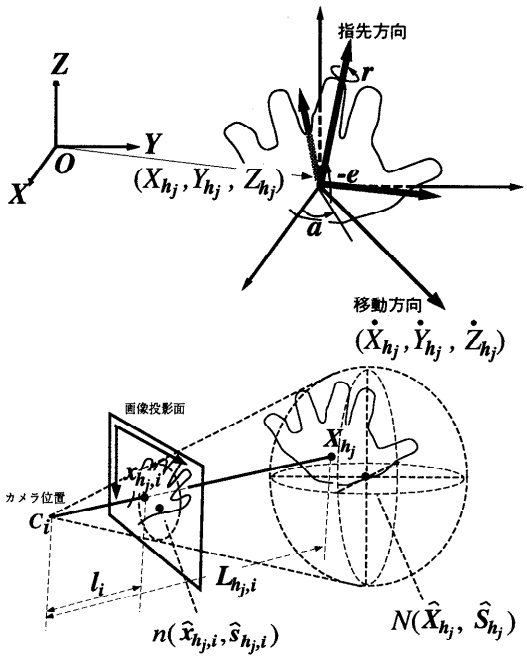


図3 観測座標系
Fig. 3 Observation coodinates.

3. 観測モデル

本章では，提案システムで用いる観測モデルについて述べる．まず，位置の観測について考える．提案システムでは，左右の手の動きを等速運動を仮定してそれぞれ独立に追跡する．**図3**上に示すように，世界座標系 (X,Y,Z) 上での手 h_j の位置を $(X_{h_j}, Y_{h_j}, Z_{h_j})$，移動速度を $(\dot{X}_{h_j}, \dot{Y}_{h_j}, \dot{Z}_{h_j})$ とすると，手の状態 $\boldsymbol{X}_{h_j}$ は，次の6次元ベクトルで表される．

$$\boldsymbol{X}_{h_j} = \left[X_{h_j} \;\; Y_{h_j} \;\; Z_{h_j} \;\; \dot{X}_{h_j} \;\; \dot{Y}_{h_j} \;\; \dot{Z}_{h_j} \right]'. \tag{1}$$

ここで添字 j は左右いずれかの手を示す $(j = l, r)$．

図3下のように，手 h_j の位置をカメラ C_i で観測するとき，観測結果は C_i の画像面上の投影点 $\boldsymbol{x}_{h_j,i}$ として得られる．ここで，i はカメラの番号である．観測を繰り返すことにより状態 $\boldsymbol{X}_{h_j}$ を逐次的に推定していく．状態推定の結果は6次元ガウス分布 $N(\hat{\boldsymbol{X}}_{h_j}, \hat{\boldsymbol{S}}_{h_j})$ で表される．$\hat{\boldsymbol{X}}_{h_j}$，$\hat{\boldsymbol{S}}_{h_j}$ が得られたとき，新たな観測により画像上で得られる投影点は，上記ガウス分布の位置成分（3次元）を画像面上に投影した分布をとると考えられ，弱透視変換を仮定すれば，この分布は2次元ガウス分布となる．ここでは，その平均を $\hat{x}_{h_j,i}$，分散を $\hat{s}_{h_j,i}$ とする．

次に，手の幅の観測について述べる．ここでは，カ

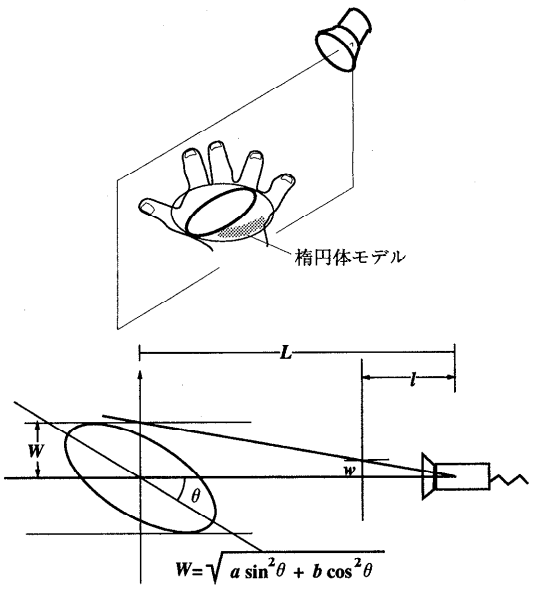


図4 楕円体モデル
Fig. 4 Ellipsoidal palm model.

メラ C_i の光軸と垂直方向の手の幅の2分の1を W_i で表す（図4下）．先の状態分布の投影と同様に弱透視変換を仮定すると，W_i は，画像面上で w_i として観測される．手 h_j とカメラ C_i の距離を $L_{h_j,i}$，カメラ C_i の焦点距離を l_i とすると，

$$w_{h_j,i} = \frac{l_i}{L_{h_j,i}} W_i \tag{2}$$

となる．

本システムでは，簡単のため手のひらを図4に示す楕円体モデルで近似する．このとき，図中 $\theta = 0$ に位置するカメラについて W は $\sqrt{a\sin^2\theta + b\cos^2\theta}$ と表すことができ，画像面上での観測値 $w_{h_j,i}$ は，

$$w_{h_j,i} = \frac{l_i}{L_{h_j,i}} \sqrt{a\sin^2\theta + b\cos^2\theta} \tag{3}$$

となる．ここで，a，b は楕円体モデルの形状を定める定数である．

次章以降では，これらの観測モデルを用いた処理の詳細について述べる．

4. 画像特徴の抽出

4.1 処理視点の選択

本節では，両手の予測位置に基づいた処理画像の選択方法について述べる．

手 h_j について時刻 $t-1$ までの観測による時刻 t の状態分布の平均を $\hat{\boldsymbol{X}}_{h_j,t}$，分散を $\hat{\boldsymbol{S}}_{h_j,t}$ とすると，時刻 t に画像上で得られる手 h_j の投影位置 $x_{h_j,i,t}$ の分布は，前章で述べたように2次元ガウス分布 $N(\hat{x}_{h_j,i,t}, \hat{s}_{h_j,i,t})$ で表される．

このとき，カメラ C_i 上でのオクルージョン発生確率を考える．オクルージョンが発生するのは左右の手が画像面上で互いの手の幅以内に近づいた場合である．したがって，前節の観測モデルから，オクルージョンが発生する条件は，

$$\begin{aligned} |x_{h_l,i,t} - x_{h_r,i,t}| &\le w_{h_l,i} + w_{h_r,i} \\ &= \left(\frac{l_i}{L_{h_l,i}} + \frac{l_i}{L_{h_r,i}}\right) W \end{aligned} \tag{4}$$

となる．

ここで，左辺に現れる $x_{h_l,i,t} - x_{h_r,i,t}$ の分布は2次元ガウス分布 $N(\hat{x}_{h_l,i,t} - \hat{x}_{h_r,i,t}, \hat{s}_{h_l,i,t} + \hat{s}_{h_r,i,t})$ となるから，その確率密度関数を $f_i(x)$ とすると上記条件を満たす確率 p_i は，

$$p_i = \int_S f_i(x)dx \tag{5}$$

となる．ただし，$\int_S$ は式(4)を満たす $x = x_{h_l,i,t} - x_{h_r,i,t}$（中心 $O(0,0)$，半径 $\left(\frac{l_i}{L_{h_l,i}} + \frac{l_i}{L_{h_r,i}}\right) W$ の円周内部）についての積分を示す．したがって，オクルージョン発生確率を小さくするため，すべての画像に対して p_i を計算し，その値が小さいものから順に処理画像を選択する．

なお，式(5)の計算は煩雑であるので，実際には確率密度を式(4)の円の中心での値 $f_i(O)$ で代表させた次の近似式 $\tilde{p}_i$ を用いた．

$$\tilde{p}_i = 2\pi \left(\frac{l_i}{L_{h_l,i}} + \frac{l_i}{L_{h_r,i}}\right)^2 W^2 \times f_i(0) \tag{6}$$

4.2 特徴抽出処理

前節で選択された画像に対して，特徴抽出処理を行う．まず，選択された各画像中の肌色領域を色情報および輝度情報を用いて手領域のシルエットを抽出する．本システムでは両手の動きを対象としているので，得られる領域の個数は0，1，2のいずれかである（領域の持つ画素数が所定の閾値を超えない場合はノイズと判断して除外し，また3個以上の領域が得られる場合には，画素数の多いものから2個を選択する）．

得られた各シルエットに対し距離変換を行い（ここで距離変換は領域内の各画素が領域境界までの最短距離を持つようにする変換をいう），極大点として手の2次元位置を定める（図5）．極大点の持つ距離変換値を3章で述べた手の幅 w の観測値と考える．極大点の2次元位置（重心点）は追跡モデルの更新に，その距離変換値 w は本節で述べるオクルージョン検出と5.3節で述べる回転角の推定にそれぞれ用いる．

次に同じシルエット画像に対して，Sobelフィルタ

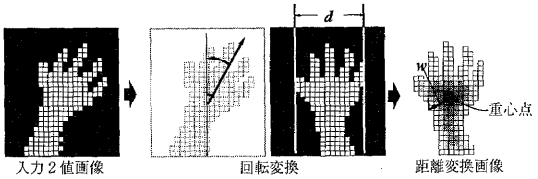


図5 特徴抽出
Fig. 5 Feature extraction.

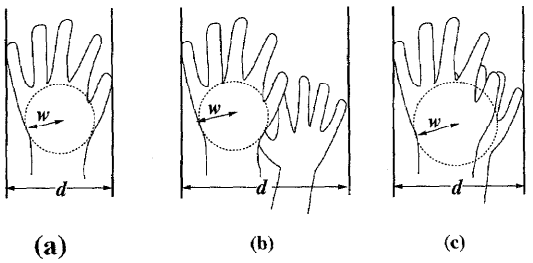


図6 オクルージョン検出
Fig. 6 Occlusion detection.

によるエッジ検出処理を施し平均エッジ方向として画像内の指先方向を検出[11]，各シルエットについて指先が上方を向くように回転変換を行う．このときのシルエットの横幅を d（図5）とする．さらに 5.3 節で述べる手形状認識のために重心点よりも上方部分の輪郭線を抽出する[2]．

ところで，ここで処理した画像は前節で述べたように依然としてオクルージョンを含んでいる可能性がある．そこで，次のような方法でオクルージョン判別を行う．

先に述べた距離変換値 w は画像内での手領域の大まかな大きさ（半径）を与えるから，それぞれの手が単独のシルエットとして観測されている場合には，シルエットの幅 d は距離変換値 w の2倍に近い値をとる（図6(a)）．一方で，左右の手が重なって1つのシルエットとして観測されると，シルエットの幅 d が大きくなる（図6(b)）．したがって，幅 d が w の2倍に比べ十分大きいとき，オクルージョンが発生していると判断する．

$$\frac{d}{w} > threshold. \qquad (7)$$

今回の実装では式(7)の閾値を 2.5 とした．

なお，左右の手の重なりが大きくなると，再び両者の違いが小さくなるため，本手法によるオクルージョン検出は困難となる（図6(c)）．しかしながら，このような極端なオクルージョンは，前節で述べた視点選択により容易に回避できると考えている．

図7に同時刻に観測した3視点の画像に対する特徴抽出結果を示す．画像中の実線は検出された指先方向を，実線の下端点は重心点をそれぞれ表している．式(7)の評価の結果，同図左の画像ではオクルージョンが発生していると判断された．**表1**は，各領域に対して得られた d，w，d/w の値を示す．

図7 位置・方向検出
Fig. 7 Position & orientation detection.

表1 図7の画像に対する変数値
Table 1 Variable values for images in Fig. 7.

	左	中(領域1)	中(領域2)	右(領域1)	右(領域2)
d	48.4	36.3	25.0	31.4	14.1
w	10	18	13	14	7
d/w	4.84	2.02	1.92	2.24	2.02

図8 オクルージョン検出の例
Fig. 8 Examples of occlusion detection.

表2 図8の画像に対する変数値
Table 2 Variable values for images in Fig. 8.

	左	中	右
d	95.1	55.5	36.5
w	8	14	14
d/w	11.9	3.97	2.61

同様にオクルージョンと判断された例とそのときの各変数値を**図8**，**表2**にそれぞれ示す．

なお，前節で述べた視点選択では手のひらどうしのオクルージョンのみを考えており，腕領域と手のひらのオクルージョンは評価していない．発生したオクルージョンについては，特徴抽出時に検出し処理対象から外す対応をとった．また今回の実装では予測位置がカメラの撮影範囲外となる場合も考慮していない．画角評価も含めた視点選択法については今後さらに検討していきたい．

5. 多数カメラを用いた両手手振り検出

5.1 位置追跡

位置追跡について説明するため，3章で述べた観測モデルをあらためて**図9**に示す．

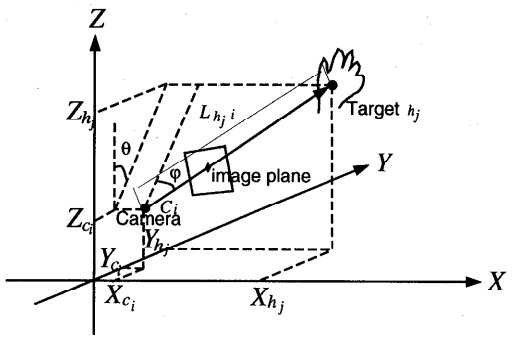


図9 位置観測モデル
Fig. 9 Observation model (hand position).

ここで，$\varphi_{h_j,i}$ はエピポーラ線と $Y-Z$ 面のなす角度，$\theta_{h_j,i}$ はエピポーラ線の $Y-Z$ 面への投影と Z のなす角度をそれぞれ示す．$\boldsymbol{R}_{\varphi_{h_j,i}}\boldsymbol{R}_{\theta_{h_j,i}}$ は，エピポーラ線を Z 軸に平行にする回転変換を示す．

手 h_j が N 台のカメラで観測されている状況を考える．ここに，$\hat{\boldsymbol{X}}_{h_j,t-1}$ は $\boldsymbol{X}_{h_j}$ の時刻 $t-1$ における推定値である（推定値 $\hat{\boldsymbol{X}}_{h_j,t-1}$ の持つ共分散を $\hat{\boldsymbol{S}}_{h_j,t-1}$ とする）．このとき，時刻 t における状態は以下のように計算できる．

$$\overline{\boldsymbol{X}}_{h_j,t}=\boldsymbol{F}\,\hat{\boldsymbol{X}}_{h_j,t-1} \tag{8}$$

$$\overline{\boldsymbol{S}}_{h_j,t}=\boldsymbol{F}\,\hat{\boldsymbol{S}}_{h_j,t-1}\,\boldsymbol{F}'+\boldsymbol{Q}. \tag{9}$$

ここで，$\boldsymbol{F}$ は状態遷移を示す行列で以下のように書ける．$\boldsymbol{Q}$ は状態遷移のゆらぎを示す共分散行列である．

$$\boldsymbol{F}=\begin{bmatrix}1&0&0&1&0&0\\0&1&0&0&1&0\\0&0&1&0&0&1\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix} \tag{10}$$

このとき，状態 $\boldsymbol{C}_i=[\ X_{c_i}\ \ Y_{c_i}\ \ Z_{c_i}\ \ 0\ \ 0\ \ 0\ ]'$ を持つカメラによる1回の観測は，エピポーラ線の方向を示す $\theta_{h_j,i}$，$\varphi_{h_j,i}$ によって以下のように示される．

$$\boldsymbol{H}\boldsymbol{R}_{\varphi_{h_j,i}}\boldsymbol{R}_{\theta_{h_j,i}}\boldsymbol{C}_i=\boldsymbol{H}\boldsymbol{R}_{\varphi_{h_j,i}}\boldsymbol{R}_{\theta_{h_j,i}}\boldsymbol{X}_{h_j}+e_p. \tag{11}$$

e_p は2次元の観測誤差（平均 $[\ 0\ \ 0\ ]'$，共分散 $E_{h_j,i}$），$\boldsymbol{H}$ は以下に示す観測行列である．

$$\boldsymbol{H}=\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\end{bmatrix} \tag{12}$$

観測誤差はカメラからの距離の増加にともない増加する．ここでは，観測誤差を以下のように近似する（E はカメラの観測誤差を表す定数行列）．

$$E_{h_j,i,t}=\frac{\bar{L}_{h_j,i,t}}{l_i}E\simeq\frac{L_{h_j,i,t}}{l_i}E. \tag{13}$$

ここで，手とカメラの実際の距離 $L_{h_j,i,t}$ は知りえないので，代わりにカメラ $\boldsymbol{C}_i$ と予測位置 $\overline{\boldsymbol{X}}_{h_j,t}$ の距離 $\bar{L}_{h_j,i,t}$ で近似する．

観測結果と式(8)，(9)から，手 h_j の状態は次のように推定される．

$$\hat{\boldsymbol{X}}_{h_j,t}=\hat{\boldsymbol{S}}_{h_j,t}\left(\overline{\boldsymbol{S}}^{-1}_{h_j,t}\overline{\boldsymbol{X}}_{h_j,t}+\sum_i\boldsymbol{R}_{\varphi_{h_j,i,t}}\boldsymbol{R}_{\theta_{h_j,i,t}}\boldsymbol{H}'\left(E^{-1}_{h_j,i,t}\right)\cdot\boldsymbol{H}\left(\boldsymbol{R}_{\varphi_{h_j,i,t}}\boldsymbol{R}_{\theta_{h_j,i,t}}\right)'\boldsymbol{C}_i\right). \tag{14}$$

ここで，

$$\hat{\boldsymbol{S}}^{-1}_{h_j,t}=\overline{\boldsymbol{S}}^{-1}_{h_j,t}+\sum_i\boldsymbol{R}_{\varphi_{h_j,i,t}}\boldsymbol{R}_{\theta_{h_j,i,t}}\boldsymbol{H}'\left(E^{-1}_{h_j,i,t}\right)\cdot\boldsymbol{H}\left(\boldsymbol{R}_{\varphi_{h_j,i,t}}\boldsymbol{R}_{\theta_{h_j,i,t}}\right)'. \tag{15}$$

式(14)，(15)において，$\sum_i$ はオクルージョンを起こしていないすべての処理画像についての和を表す．

本システムでは，各画像特徴と左右の追跡モデルを2次元画像面上で対応づける．追跡モデルの3次元ガウス分布 $N(\overline{\boldsymbol{X}}_{h_j,t},\overline{\boldsymbol{S}}_{h_j,t})$ を各画像面に投影し，画像面上での追跡モデルの分布を得る．前述のとおり弱透視変換を仮定すると，この分布は2次元ガウス分布で表される．各追跡モデルにマハラノビス距離の意味で最も近い画像特徴を対応づけ，上述の方法でモデルの更新を行う[12]．$\overline{\boldsymbol{X}}_{h_l,t}$，$\overline{\boldsymbol{X}}_{h_r,t}$ は $Y-Z$ 面について対称な位置で，$\overline{\boldsymbol{X}}_{h_l,t}$ は左側（X 軸負），$\overline{\boldsymbol{X}}_{h_r,t}$ は右側（X 軸正）に初期位置を持つ．初速は0とする．

図10 に，約50フレームの画像に対する追跡結果を示す．ここでは，被験者が前方に伸ばした左右の手を互いに近づく方向に移動させ，水平（X 軸）方向に交差させる操作を行った．

図11 に追跡時の視点選択の様子を示す．同図上段は，処理視点選択のため式(6)により算出したオクルージョン発生確率 $\tilde{p}_i$ の値を，同図中段は参考のためビデオテープに記録した入力画像から目視により確認したオクルージョン発生フレームをそれぞれ全5視点について示している．

同図下段は，視点選択により選ばれた3視点のカメラ番号をそれぞれ示す．なお，本図において，△，□

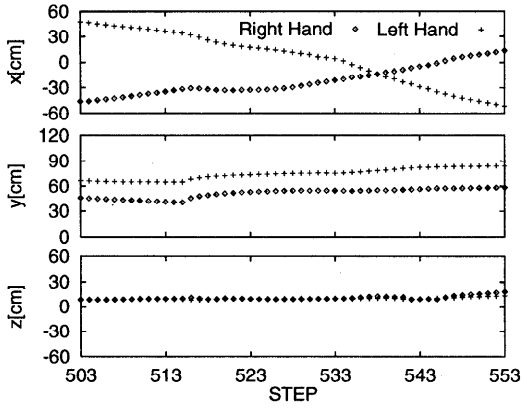


図 10 位置追跡結果（X 軸方向）
Fig. 10 Position tracking results (for X axis).

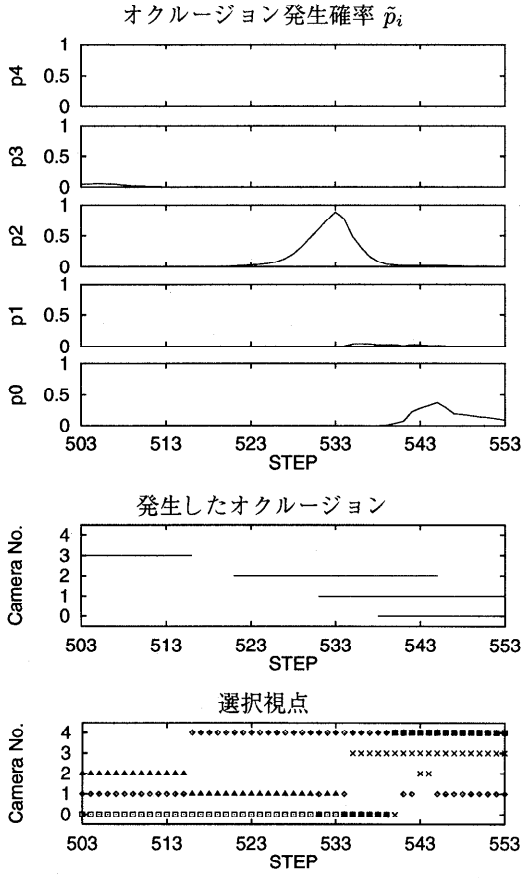


図 11 視点選択結果（X 軸方向）
Fig. 11 Camera selection results (for X axis).

は後述の手形状認識に用いられた視点を，× は選択されたもののオクルージョン判別処理においてオクルージョンが検出され，その後の処理に用いられなかった視点をそれぞれ示している．本図から，両手の位置が，オクルージョン回避のための視点選択をともないながら安定に追跡されていることが分かる．

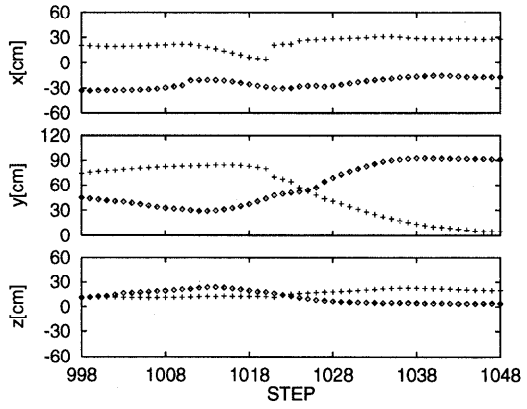


図 12 位置追跡結果（Y 軸方向）
Fig. 12 Position tracking results (for Y axis).

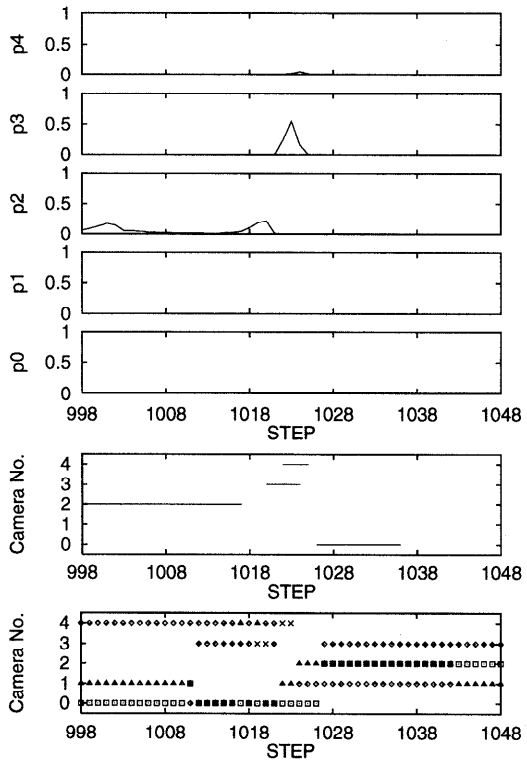


図 13 視点選択結果（Y 軸方向）
Fig. 13 Camera selection results (for Y axis).

図 12～**図 15** は，同様の実験を垂直（Y 軸）および奥行き（Z 軸）方向の交差運動について行った結果を示している．ここにみられるように，本手法によりいずれの方向についても安定した追跡を行うことができる．なお，追跡結果の一部に不連続がみられるが，これは重心検出時のノイズによるものと考えられる．

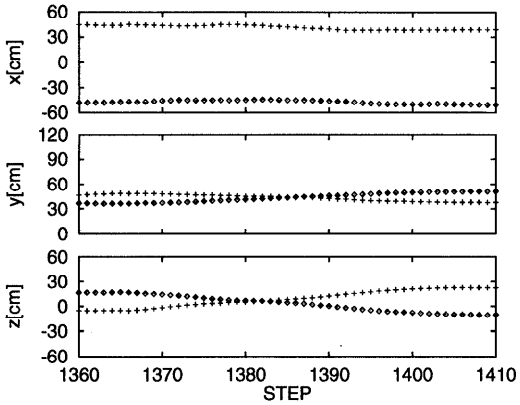


図 14 位置追跡結果（Z 軸方向）
Fig. 14 Position tracking results (for Z axis).

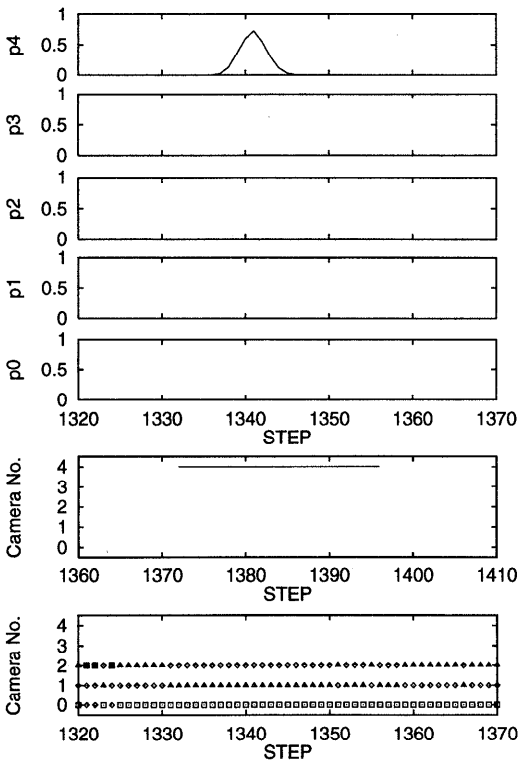


図 15 視点選択結果（Z 軸方向）
Fig. 15 Camera selction results (for Z axis).

ところで，前述の追跡処理では手の動きが等速運動であることを仮定しているが，実際の手の動きはこの仮定から外れることが多い．しかしながら，ほとんどの場合，等速運動からのずれはシステムのゆらぎとして吸収可能である．**図 16** は，一例として後述の仮想空間操作システム（図 22）で四角形を描いた際の位置追跡結果を示している．このように，四角形の軌跡の

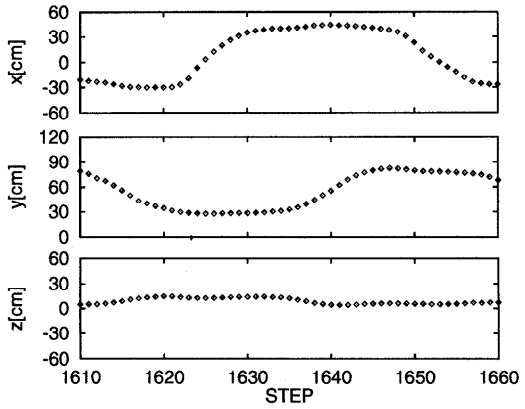


図 16 位置追跡結果（非等速運動）
Fig. 16 Position tracking results (for a hand motion with non-constant velocity).

ような不連続な動きも安定に追跡することができる．

5.2 指先方向推定

指先方向についても位置と同様に時系列情報に基づいて追跡する．追跡にあたっては，指先方向が前後のフレームで大きく変化しないと仮定した．

まず，**図 17** 左のように，指先方向を半径 1 の単位球上の点として表す．ここでは方向のみが重要であるので，簡単のため世界座標を平行移動し，カメラ位置 C_i を単位球の中心に一致させている．ここで，前時刻 $t-1$ の予測方向を V_{t-1} とする．

4.2 節で述べたように，エッジ検出によって各画像内で 2 次元指先方向が定まるが，これは実際の指先方向（3 次元）を画像上に投影したものと考えられるから，指先方向は重心点を通るエピポーラ線，2 次元指先方向の両方を含む平面と単位球の交線上に存在することになる．この交線上で V_{t-1} に最も近い点を P_i とする．

次に，前時刻からの手方向の変化が小さいとの仮定から，V_{t-1} を中心として P_i を含む単位球上の小領域を平面と見なす（図 17 右）．この平面上に V_{t-1} を原点とする直交座標 a_v-e_v をとると，先ほどの平面と単位球の交線は，P_i を通る直線で近似できる．

この直線と a_v のなす角を γ とすると，この観測は位置追跡における式 (11) と同様に以下のように表される．

$$\boldsymbol{H}_v \boldsymbol{R}_{\gamma_i} P_i = \boldsymbol{H}_v \boldsymbol{R}_{\gamma_i} V + e_o. \tag{16}$$

ただし，$\boldsymbol{R}_{\gamma_i}$ は直線を軸 e_v に平行にする回転変換，$\boldsymbol{H}_v$ は以下に示す観測行列である．

$$\boldsymbol{H}_v = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{17}$$

e_o は 1 次元の観測誤差で平均 0，分散 $\sigma_{h_j,i,t}$ とする．

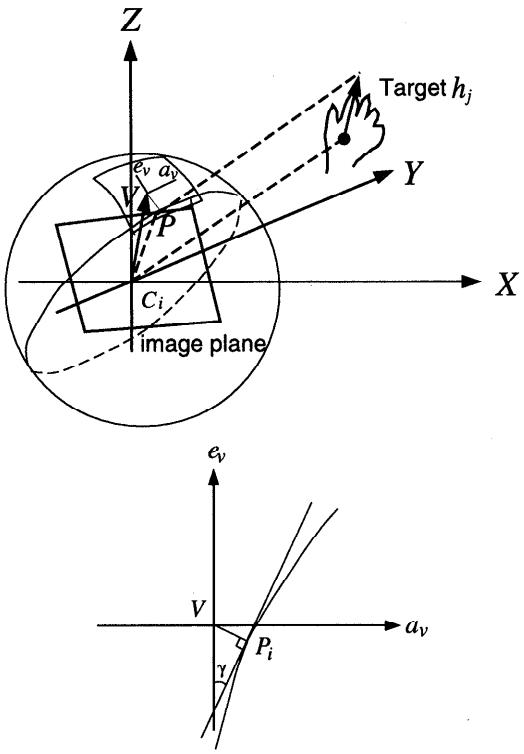


図 17 指先方向観測モデル
Fig. 17 Observation model (hand orientation).

観測誤差には，位置観測の場合と同様にカメラからの距離による重みづけを行う．

$$\sigma_{h_j,i,t} = \simeq \frac{L_{h_j,i,t}}{l_i}\sigma. \quad (\sigma \text{ は定数}) \tag{18}$$

観測結果から，a_v-e_v 上での方向推定は以下のように表される．

$$\hat{P}_{h_j,t} = \hat{T}_{h_j,t}\left(\hat{T}_{h_j,t-1}\hat{P}_{h_j,t-1} + \sum_i \boldsymbol{R}_{\gamma_{h_j,i,t}}\boldsymbol{H'}_v\sigma^{-1}_{h_j,i,t}\boldsymbol{H}_v\boldsymbol{R'}_{\gamma_{h_j,i,t}}\boldsymbol{V}_i\right). \tag{19}$$

ここで，$\hat{T}_{h_j,t}$ は推定値 $\hat{P}_{h_j,t}$ の持つ誤差の共分散を示す．$\hat{P}_{h_j,t}$ を，オイラー角 a，e に変換し，指先方向を得る．

前節の実験と同じ画像に対する指先方向推定の結果を**図 18** に示す．本実験では被験者は常時両手をほぼ前方（Z 軸の負方向）に向けていた．実験結果は，e がつねに 90 度付近の値を示しており，推定が良好に行われていることを示している．

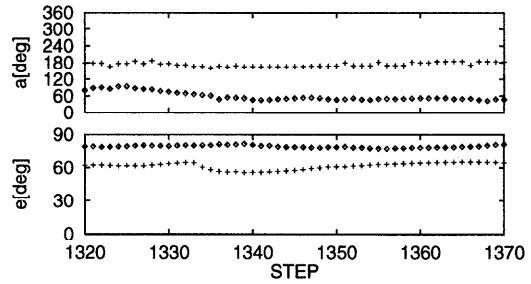


図 18 指先方向追跡結果
Fig. 18 Orientation tracking results.

5.3 手姿勢（回転角）推定および形状判別

手の指先方向回りの回転角は多視点の観測画像内で手の特徴点の持つ距離変換値から推定される[1]．これらの値は，手の形状変化・自己オクルージョンの影響を受けにくく，安定な姿勢推定が可能である．3 章で述べたように，楕円体モデルの仮定により，手のひらの法線方向から角度 θ の位置にあるカメラで観測される距離変換値 w は $\frac{l_i}{L_{h_j,i}}\sqrt{a\sin^2\theta + b\cos^2\theta}$ となる．

観測にガウス誤差を仮定すると，m 台のカメラから観測値 $w_1,\cdots,w_m$ が得られる確率 $P(w_1,\cdots,w_m|\theta)$ を最大にする角度 θ として手姿勢を推定することができる．

$$P(w_1,\cdots,w_m|\theta) = \prod_{i=1}^{m} P(w_i|\theta-\theta_{c_i}) \tag{20}$$

$$P(w|\theta) = \frac{1}{\sqrt{2\pi}\sigma_w}\exp\left(-\frac{1}{2\sigma_w^2}\cdot\left(w-\sqrt{a\sin^2\theta+b\cos^2\theta}\right)^2\right). \tag{21}$$

手の位置・姿勢推定の後，利用可能な視点のうち最も手のひらを正面から捉える画像を選んで，手形状の識別を行う（**図 19**）．前述の実験結果（図 11）で，△は右手用に選ばれた画像を，□は左手用に選ばれた画像をそれぞれ示している．

選択されたカメラのシルエット画像から抽出した輪郭線をフーリエ記述子で記述し[13]，低次成分を用いて手形状の判別を行う[2]．

図 20 は，今回の実装での判別対象である 7 種類の形状を示している．**表 3** に，オクルージョンが起こらない条件での各形状およそ 300 フレームについての認識結果を示す．ここにみられるようにほとんどの形状について，90%以上の良好な結果が得られた．これは，次章で述べる仮想空間操作を含む多くのアプリ

表3 形状検出結果
Table 3 Shape recognition accuracy.

	Input Hand Shape						
	shape1	shape2	shape3	shape4	shape5	shape6	shape7
Input Frames	299	298	300	297	299	297	298
Correct Answer	299	295	271	275	261	297	295
Recognition Rate (%)	100	99.0	90.3	92.6	87.3	100	99.0

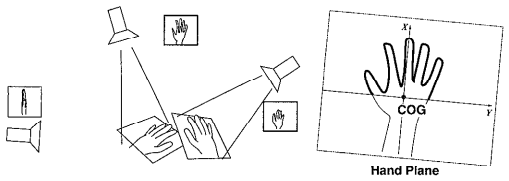


図19 手形状認識
Fig. 19 Hand shape recognition.

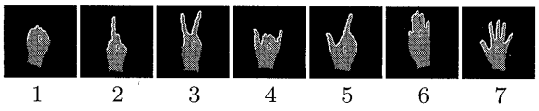


図20 実験用手形状（7種類）に関する輪郭検出結果
Fig. 20 Seven hand shapes and extracted contours.

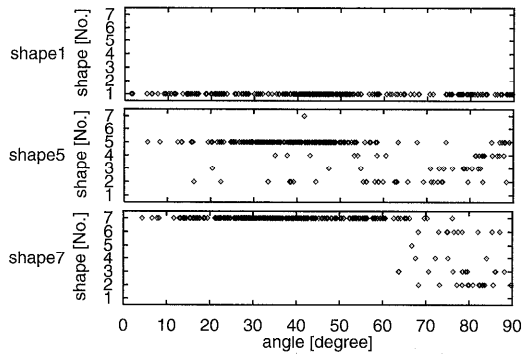


図21 選択視点の変化に対する形状認識の安定性
Fig. 21 Stability of shape recognition versus difference of a selected viewpoint.

ケーションについて十分な認識精度と考えられる．

一方で，手どうしの相互オクルージョン等により，利用可能な視点数が限られる場合には，正面に近い像が得られないケースが発生し，認識精度が悪化する．**図21**に，形状認識に使われたカメラの光軸と手のひら法線間の角度 θ の変化と認識結果の関係を示している．形状認識に利用する視点が手のひらの正面から外れるに従い，認識誤りが増加する様子が分かる．なお，認識精度の悪化の度合いは形状ごとに異なるため，両手の接近する可能性の高いジェスチャに安定した認識が可能な手形状を当てることで，オクルージョンによる認識誤りの影響を低減できる．

表4 コマンド一覧
Table 4 Command list.

コマンド	手の位置	形状遷移
形状生成	外部	7 → 1
把持・移動	内部	7 → 1
拡大・縮小	内部	7 → 2
色・テクスチャ変更	内部	7 → 4
削除	内部	7 → 5
結合	内部	7 → 6
分割	内部	7 → 3
伸縮	内部（両手）	7 → 1

6. 両手を用いた仮想空間操作

本手法を用いて，両手による仮想空間インタラクションを行うシステムを構築した．本システムは，文献2)で紹介した仮想シーン操作システムを拡張したものであり，ユーザは手振りジェスチャによって，仮想的な形状を生成し，位置，大きさ，色，形状を変更することができる．**表4**に現在利用可能なコマンドの一覧を示す．表中,「手の位置」はコマンド生成にユーザの手の位置が仮想オブジェクトの内部・外部いずれにあるかを示している．数字は，図20の形状番号を示す．伸縮コマンドを除くすべてのコマンドは左右の手で独立に発行することができる．以下では各コマンドの動作について説明する．

形状生成

形状生成コマンドでは，ユーザは手の運動軌跡により生成される特定の形状を指定することができる．ユーザは，手のひらを閉じたあと（7 → 1），手を動かすことにより軌跡を描き，手のひらを開くことでコマンドを終了する．あらかじめ用意した形状のうち，描いた軌跡に近い外形を持つ形状が生成される．

指示操作

形状生成を除くすべてのコマンドは，手を仮想物体の内部に入れることで対象となる形状を選択する．コマンドの発行は，内部で手の形状を変化させることで行う．**図22**は，左手で円錐形の形状を把持しながら，右手で新しい形状を生成している様子を示している．

図 22 両手を使った操作の例（把持+生成）
Fig. 22 Independent manipulation with both hands.

図 23 両手を使った操作の例（伸縮）
Fig. 23 Stretch command.

スライド操作

把持・移動，拡大・縮小，色・テクスチャ変更，伸縮の各コマンドについては，ユーザは形状内部での手形状変形によるコマンド発行後に手の位置を変化させることで，仮想物体の属性（位置，大きさ，色，形状等）を変更する．手形状を形状7に戻すことでコマンドは終了する．**図 23** は，伸縮コマンドにより仮想物体の引き延ばしを行っている様子を示している．

以上のコマンドにより，ユーザは仮想空間内に自由に形状を生成し，それらを操作することができる．その際，手の動作の検出に最適な視点が動的に選択されるため，ユーザはカメラの位置を特に意識する必要がない．

動的な視点選択を行わず利用できる視点数が限られたシステムでは，両手の運動により容易に相互オクルージョンが発生するため，検出可能な手の動作はカメラの位置に依存した限定的なものとなる．図 22，図 23 に示したように，本システムでは両手を独立に運動させて仮想空間操作を行うことが可能であり，このことは多視点情報を利用した本手法の有効性を示すものと考えられる．

7. ま と め

多数カメラを用いて両手の動きを検出するシステムを提案した．提案手法では，複数視点で得られる画像特徴から両手の3次元位置・姿勢を追跡する．位置・姿勢追跡にはカルマンフィルタを利用しており，フィルタから得られる次フレームの予測位置に従い，利用するカメラを選択している．これにより，多視点の情報を少ない計算コストで効率的に利用することができる．手形状認識には，位置・姿勢の推定値に基づいて左右それぞれの手形状の認識に最適な視点を選択して利用する．このように，本システムでは追跡と形状認識について2段階の視点選択を行っており，これにより手どうしのオクルージョンと自己オクルージョンを低減している．本論文では，実画像を用いた実験によって本手法の有効性を確認した．

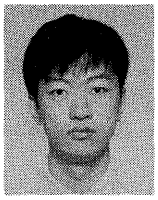
謝辞 システムの実装にご尽力くださった裏隠居宏，井村茂雄の両氏に深く感謝します．

参 考 文 献

1) 内海 章，宮里 勉，岸野文郎，大谷 淳，中津良平：距離変換処理を用いた多視点画像による手姿勢推定法，映像情報メディア学会誌，Vol.51, No.12, pp.2116–2125 (1997).
2) 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).
3) 吉田美寸夫，ジュリA.ティヘリノ，宮里 勉，岸野文郎：手振りと言語による仮想物体形状生成インタフェース，テレビ誌，Vol.50, No.10, pp.1482–1488 (1996).
4) Hanqiu, S.: Hand Interface in Traditional Modeling and Animation Tasks, *J. of Comput. Sci. & Technol.*, Vol.11, No.3, pp.286–295 (1996).
5) Moghaddam, B. and Pentland, A.: Maximum Likelihood Detection of Faces and Hands, *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp.122–128 (1995).
6) Cipolla, R., Hadfield, P.A. and Hollinghurst, N.J.: Uncalibrated Stereo Vision with Pointing for a Man-Machine Interface, *Proc. IAPR Workshop on Machine Vision Applications*, pp.163–166 (1994).
7) Rehg, J.M. and Kanade, T.: Visual Tracking of High DOF Articulated Structures: An Application to Human Hand Tracking, *Computer Vision-ECCV '94*, LNCS, Vol.801, pp.35–46 (1994).

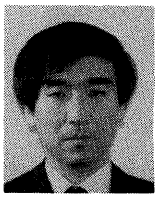
8) Iwai, Y., Yagi, Y. and Yachida, M.: Estimation of Hand Motion and Position from Monocular Image Sequence, *Proc. ACCV'95*, Vol.II, pp.230–234 (1995).
9) Davis, J. and Shah, M.: Determining 3-D Hand Motion, *Asilomar Conference in Signals, Systems and Computers*, pp.1262–1266 (1994).
10) Azarbayejani, A. and Pentland, A.: Real-Time Self-Calibrating Stereo Person Tracking Using 3-D Shape Estimation from Blob Features, *13th International Conference on Pattern Recognition*, pp.627–632 (1996).
11) 石淵耕一，岩崎圭介，竹村治雄，岸野文郎：画像処理を用いた実時間手振り推定とヒューマンインタフェースへの応用，信学論（D-II），Vol.J79-D-II, No.7, pp.1218–1229 (1996).
12) Cox, I.J.: A Review of Statical Data Association Techniques for Motion Correspondence, *International Journal of Computer Vision*, Vol.10:1, pp.53–66 (1993).
13) 上坂吉則：開曲線にも適用できる新しいフーリエ記述子，信学論（A），Vol.J67-A, No.3, pp.166–173 (1984).

(平成11年1月5日受付)
(平成11年6月3日採録)



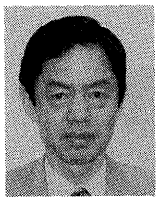
内海　章（正会員）

1991年大阪府立大学工学部金属工学科卒業．1993年大阪大学大学院基礎工学研究科情報工学修士課程修了．同年ATR通信システム研究所に入社．画像処理，ヒューマンインタフェースの研究に従事．現在，ATR知能映像通信研究所研究員．



大谷　淳（正会員）

1979年，東京大学大学院精密機械工学専攻修士課程修了．同年，電電公社（現NTT）電気通信研究所入所．以来，画像処理，カラー記録等の研究に従事．1988年より1年間米国Maryland大学客員研究員．1992年（株）ATR通信システム研究所に出向．現在（株）ATR知能映像通信研究所第一研究室長．仮想空間を介したコミュニケーション方式の研究に従事．工学博士．



中津　良平（正会員）

1969年京都大学工学部電子工学科卒業．1971年同大学大学院修士課程修了．同年日本電信電話公社（現NTT）武蔵野電気通信研究所入所．1980年横須賀電気通信研究所．主として音声認識の基礎研究，応用研究に従事．1990年NTT基礎研究所研究企画部長，1991年NTT基礎研究所情報科学研究部長．1994年よりATRに移り，現在（株）ATR知能映像通信研究所代表取締役社長．マルチメディア要素技術の研究およびマルチメディア技術を応用した通信方式の研究等に従事．工学博士（京大）．