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

林 健太郎† 久野義徳† 白井良明†

本論文では，画像から抽出した上半身の3つの特徴点を基準とし，手の3次元位置，方向を計測する手法を提案する．著者らが先に発表した手法では，3次元情報を計算する際に基準となるアフィン座標系を作るための4つの特徴点をユーザの体から抽出していた．このとき，4点は同一平面上にあってはならないので，ユーザの姿勢は大きな拘束を受けていた．本論文では，アフィン座標系を作る特徴点のうちの3点を体上にとり，その3点が作る平面の法線上の1点を仮想的な4点目とする．このようにすれば3つの特徴点でアフィン座標系が作れるので，ユーザの姿勢を拘束しないヒューマンインタフェースを構築できる．実験として，シミュレーションによって仮想基準点を求める手法を検証する．また，応用例として本手法を用いたプレゼンテーションシステムを構築し，ヒューマンインタフェースとして有効であることを示す．

Position-free Human Interface by Pointing Gestures

KENTARO HAYASHI,† YOSHINORI KUNO† and YOSHIAKI SHIRAI†

This paper describes a method to measure the user's 3D hand position and orientation using 3 features on the upper body of the user. The method proposed by the authors before used 4 feature points attached on the user's body. The user's pose was restricted because these 4 features must not be coplanar. Thus, we propose to use a virtual feature as the 4th feature, which is calculated by using the normal direction of the plane made by the 3 features. This method allows the user to move freely. The experimental results show that the proposed method is useful for the interface of presentation systems.

1. はじめに

今日，仮想現実の技術が飛躍的に向上し，また，現実世界での機械操作が一段と複雑化しつつある．このような中，人間が機械に対して3次元空間上の手ぶりで指示をするためのヒューマンインタフェースの技術が注目されている．このヒューマンインタフェースの技術が確立されれば，一般に操作が容易になるだけでなく，身体の不自由な人のためのインタフェースの可能性も切り開くことができる．

仮想現実などでは従来からデータグローブなど手に特殊なセンサを取り付ける方法がよく用いられている[1]．しかし，これらの方法はユーザ（使用者）にセンサの装着を強制させるもので，ユーザの自由度を制限する．

そこで画像処理によりユーザの手ぶりを解釈するヒューマンインタフェースシステムの検討が行われている．コンピュータビジョンでは3次元位置の計測に関して，エピポーラ拘束を利用したステレオ視の歴史が古く，ステレオ視を用いた手の位置の計測システムが提案されている[2]．しかしエピポーラ拘束を利用したステレオ視では，事前に正確なカメラキャリブレーションを行う必要があり，このことが手軽にシステムを使用する際の妨げとなる．

これを受け著者らは先に，アフィン不変量[3),4)]を用い，カメラキャリブレーションを行う必要のない計測方法を提案した[5),6)]．岡本ら[7)]は1つのカメラから得られる不変量から，定性的な運動認識を行っているが，我々は視点の異なる複数のカメラを用いて，ユーザの体を基準とするアフィン不変量を計算し，ユーザの腕の3次元位置姿勢を定量的に復元している．そしてこのようなシステムを用いたインタフェースとして，手の指示によりCG像を操作するシステムを実現した[5)]．

3次元のアフィン座標系をとるためには，同一平面上にない4点を画像から抽出する必要がある．先のシステムでは，ユーザに座ってもらい，肩，腰の左右両端，膝の位置に特徴点をとった．しかし，スクリー

† 大阪大学大学院工学研究科
Graduate School of Engineering, Osaka University

ンに向かってCGをジェスチャで操作するような，プレゼンテーションシステムのためのヒューマンインタフェースなどを考える場合には，自由に動きまわって自由な姿勢で使用できるのが望ましい．

この問題を解決するために，アフィン座標系を作るための4点のうち1点を仮想点（以下仮想基準点）として画像中の3点から求め，計4点を基準とする座標系から3次元位置を計測する手法を提案する．こうすれば，ステレオ視のようなカメラキャリブレーションが不要で，かつユーザの姿勢に対する制約もないヒューマンインタフェースが構築できる．

ユーザをパンチルトカメラで追跡しながらこの方法を用いれば，ユーザがどこに移動しても，自分の体を中心にして前後左右の3次元位置を解釈するインタフェースが実現できる．本手法ではアフィン不変量を用いているので，パンチルトの具体的な値を知らなくても位置姿勢の推定ができる．したがって，位置姿勢の推定処理とユーザを追跡する処理とが独立に構成できる．

本論文では，シミュレーション実験で仮想基準点を求めるアルゴリズムを検証し，また，実際のプレゼンテーションシステムを作成して，その有効性を確かめる．

2. 3次元位置の計測

本手法では各時点でのユーザの手の位置と方向を，基準点の作る座標系上の相対量として定量的に求める．求めるべき量は位置に対して3自由度と，方向に対して2自由度の5自由度である（腕まわりの回転は含まない）．

2.1 アフィン不変量の計測

提案手法を説明する前に，そのもとになる複数視点の画像からのアフィン不変量[3),4)]について述べておく．

図1 に示すように，3次元空間上に5点 $\boldsymbol{X}_i, i \in \{0,\ldots,4\}$ 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，$\boldsymbol{X}_0$ を原点とする基底ベクトル

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

を設ける．この基底ベクトルを用いると，第5点 $\boldsymbol{X}_4$ 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．

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

ここで，カメラの投影を weak perspective と仮定する．$\boldsymbol{X}_0,\ldots,\boldsymbol{X}_4$，$\boldsymbol{E}_1,\ldots,\boldsymbol{E}_3$ を画像上に投影し，投影された座標をそれぞれ $\boldsymbol{x}_0,\ldots,\boldsymbol{x}_4$，$\boldsymbol{e}_1,\ldots,\boldsymbol{e}_3$ 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．

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

ここで，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，

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

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{\alpha} \tag{4}$$

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 $\boldsymbol{\alpha} = [\alpha\,\beta\,\gamma]^{\mathrm{T}}$ 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[5)]．

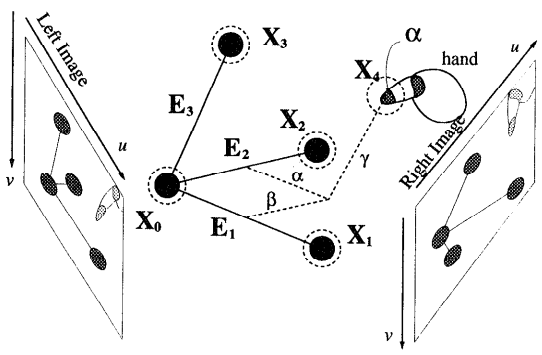


図1 アフィン座標系
Fig. 1 Affine basis.

2.2 3次元上の仮想基準点の計算

文献5)，6)では，ユーザが座った状態で，肩，腰の左右両端，膝の計4つの基準点をとり，手先の点をアフィン不変量として求め，手の3次元位置を得ている．また，それに加え腕の画像上での方向を求めて腕の方向を得ている．

しかし，同一平面上にない人体上の4点をとるためにユーザの姿勢を制限しなければならないという問題点がある．一方，同一直線上にない3点ならば，どんな姿勢においても体上にとりやすいので，本論文では

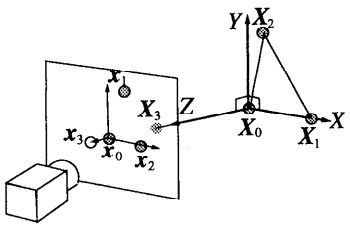


図2 3点が作る平面上にない4つ目の点 X_3
Fig. 2 Virtual point X_3 above the plane made by the existing 3 points.

3点の観測で済む方法を考える．

そこで図2のような物体座標系 XYZ を考える．ただし，体上の3点 $\boldsymbol{X}_0$，$\boldsymbol{X}_1$，$\boldsymbol{X}_2$ は，XY 平面上にあるとする．$\boldsymbol{X}_0$ を原点とし，$\boldsymbol{X}_0$ から $\boldsymbol{X}_1$ に向かう方向を X 軸，$\boldsymbol{X}_0$ から $\boldsymbol{X}_2$ に向かう側で，X 軸に垂直な方向を Y 軸とする．残りの Z 軸は右手座標系を作るようにとる．このとき，Z 軸上に第4の点として仮想基準点 $\boldsymbol{X}_3$ をとることを考える．

仮想基準点 $\boldsymbol{X}_3$ を求めるには，2時点における1台のカメラの画像を用いる．この2時点の間に対象となる3点が動くとする．ただし，カメラが固定で空間上の点が剛体として回転，並進することは，空間上の点が固定で，カメラが回転，並進することと等価である．そこで以下の説明では後者を用い，カメラが運動するときの画像上の特徴点の変位を考える．

2.2.1 カメラの動きと法線方向の幾何学的関係

まずはじめにいくつかの定義を与えよう．最初に図3に示すカメラ座標系 $X_cY_cZ_c$ を与える．カメラの投影モデルとして，カメラ座標系の原点を中心とする球面への投影を考える．ただし，物体座標系の原点 $\boldsymbol{X}_0$ を球面に投影した点を $\boldsymbol{x}_0$ とし，$\boldsymbol{x}_0$ における接平面を画像面と近似する．この画像面上に直交軸 u，v をとる．以後3次元上の点の投影はすべてこの画像面への投影とする．3次元上のカメラの移動量を表すベクトルを $\boldsymbol{V}=[V_1\ V_2\ V_3]^{\mathrm{T}}$ とし☆，$\boldsymbol{V}$ の画像上への投影を $\boldsymbol{A}$ とおく．また，投影中心から $\boldsymbol{X}_0$ までの距離（深さ）を λ とする．λ の勾配をとったもの，$\mathbf{grad}\,\lambda$ が平面の法線方向である．$\mathbf{grad}\,\lambda$ を画像上に投影したものを $\boldsymbol{F}$ とおき，以後これを投影法線と呼ぶ．

さて，3次元空間上の剛体3点 $\boldsymbol{X}_0$，$\boldsymbol{X}_1$，$\boldsymbol{X}_2$ を2つの異なるカメラ位置から撮像した画像があるとする．空間上の3点 $\boldsymbol{X}_0$，$\boldsymbol{X}_1$，$\boldsymbol{X}_2$ を移動前，移動後の画像上に投影したものをそれぞれ $\boldsymbol{x}_0$，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，$\boldsymbol{x}_0'$，

☆ 球面投影カメラの回転は画像上の特徴点を平行移動させるだけで，平面の法線を求めることに寄与しない．したがって，カメラの並進だけを考えればよい．

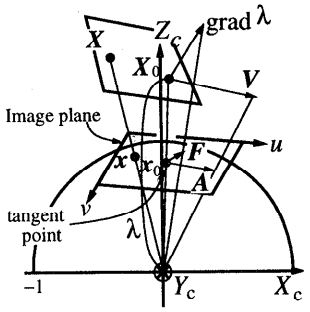


図3 カメラの移動量と平面の法線の画像への投影
Fig. 3 Projection of translational velocity vector and gradient vector.

$\boldsymbol{x}_1'$，$\boldsymbol{x}_2'$ とし，この画像上の3点の集合をそれぞれ I，I' とする．各点の画像上での変位をそれぞれ

$$\boldsymbol{v}_i=\boldsymbol{x}_i'-\boldsymbol{x}_i \quad (i=0,1,2) \tag{5}$$

とする．画像上の3点が，アフィン変換によって I から I' へ移されたと考えれば，その変換は並進成分を除いて以下の行列

$$\mathbf{U}=\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} =([\boldsymbol{v}_1-\boldsymbol{v}_0,\boldsymbol{v}_2-\boldsymbol{v}_0])[\boldsymbol{x}_1-\boldsymbol{x}_0,\boldsymbol{x}_2-\boldsymbol{x}_0]^{-1} \tag{6}$$

で表される．ここで，$[\cdot,\cdot]$ は縦ベクトルを横に並べて作った行列である．また，u_x，u_y，v_x，v_y は変位成分 u，v をそれぞれ x，y 方向に偏微分したものである．さてここで，$\boldsymbol{x}_0$ の近傍点 $\boldsymbol{x}$ でのアフィン変換 $\mathbf{U}$ によって引き起こされる移動ベクトルを $\boldsymbol{v}$ とする．$\boldsymbol{v}$ の拡大量を表す量 $\mathrm{div}\,(\boldsymbol{v})$ と，変位要素を表す量 $\mathrm{def}\,(\boldsymbol{v})$ はアフィン変換行列の要素を用いてそれぞれ

$$\mathrm{div}\,(\boldsymbol{v})=u_x+v_y \tag{7}$$

$$\mathrm{def}\,(\boldsymbol{v})\cos 2\mu=u_x-v_y\approx\frac{V_1\lambda_x-V_2\lambda_y}{\lambda^2} \tag{8}$$

$$\mathrm{def}\,(\boldsymbol{v})\sin 2\mu=u_y+v_x\approx\frac{V_1\lambda_y+V_2\lambda_x}{\lambda^2} \tag{9}$$

と表される．上記で μ は変位の方向を代表する角度である．また，式(8)，(9)の第3項の近似は $\boldsymbol{V}$ のノルム $\|\boldsymbol{V}\|$ が微小な場合に成立する[8]．

次に Z_c 軸が物体座標系の原点 $\boldsymbol{X}_0$ を通る場合を考える．カメラの回転は行列 $\mathbf{U}$ を変化させないので，カメラを適当に回転させることでカメラ座標系をつねにこのようにとってよい．この座標系上での $\boldsymbol{V}$ と λ を用いて $\boldsymbol{A}$，$\boldsymbol{F}$ を具体的に書けば

$$\boldsymbol{A}=\left[\frac{V_1}{\lambda}\ \ \frac{V_2}{\lambda}\right]^{\mathrm{T}} \tag{10}$$

$$\boldsymbol{F}=\left[\frac{\lambda_x}{\lambda}\ \ \frac{\lambda_y}{\lambda}\right]^{\mathrm{T}} \tag{11}$$

となる．ただし λ_x，λ_y はそれぞれ λ の x，y 方向の偏微分である．上式を用いて $\mathrm{def}(\boldsymbol{v})$ を書き直せば

$$\mathrm{def}(\boldsymbol{v}) = \|\boldsymbol{F}\|\|\boldsymbol{A}\| \tag{12}$$

$$\mu = \frac{\angle\boldsymbol{A} + \angle\boldsymbol{F}}{2} \tag{13}$$

となることが簡単な計算から分かる[9]．以上より μ と $\angle\boldsymbol{A}$ が求まれば，投影法線の方向 $\angle\boldsymbol{F}$ が分かる．

2.2.2 正面画像を用いた法線方向の計算

$\angle\boldsymbol{A}$ はカメラの移動量 $\boldsymbol{V}$ から求まるが，カメラの動きは，3次元上の特徴点（つまりユーザ）の動きの逆であり，それを事前に知ることはできない．

そこで，空間上の3点が作る平面に対して，カメラの光軸が直角，つまり $\mathbf{grad}\,\lambda$ が Z_c に平行になる位置から3点を観測し，画像上の特徴点の位置 $\boldsymbol{x}_0^0$，$\boldsymbol{x}_1^0$，$\boldsymbol{x}_2^0$ を登録する．ただしカメラは物体座標系の Z 軸の正の方向から負の方向へ見るものとする．これを正面画像上の3点と呼ぶ．正面画像を登録することは，焦点距離を除くカメラの内部パラメータをキャリブレーションすることともいえる．正面画像が得られれば，それぞれのカメラの，正面画像からの位置の変位より仮想基準点の位置が求まり，アフィン不変量が求まる．アフィン不変量はキャリブレーションが不要であるので結果的に外部パラメータを較正する必要はない．さらに，正面画像を登録することは直接的に内部パラメータを求めているわけではなく，あくまでも画像を登録するだけでよい．この正面画像を得るためには，ユーザにカメラに対してほぼ正面を向いてもらうだけでよく，正確に正面を向いている画像は不要である．この理由は後に2.3節で述べる．

正面画像の3点から，ユーザが動いた後の入力画像から得られる新たな3点への変位を用いて投影法線 $\boldsymbol{F}$ を求めることとする．このようにすれば，カメラの移動方向 $\boldsymbol{A}$ と，平面の投影法線 $\boldsymbol{F}$ は，図4に示すように必ず π の角度を持って現れる．図中の破線はすべて同一平面上にある．

したがって，$\angle\boldsymbol{A} = \angle\boldsymbol{F} \pm \pi$ とすれば，

$$\angle\boldsymbol{F} = \mu \pm \pi/2 \tag{14}$$

とできるので，$\angle\boldsymbol{A}$ を陽に求めずに $\angle\boldsymbol{F}$ が求まる．しかし，画像上で π の角度を持つ2つのベクトルの方向は，どちらを $\angle\boldsymbol{F}$ としてもつじつまがあうので，どちらか一方の正しい解を選択する必要がある．

2.2.3 正しい法線方向の選択

そこで次に，2台のカメラの大まかな位置関係から，正しい $\angle\boldsymbol{F}$ を選択する方法について述べる．ここで，正面画像から入力画像への3点の変位を使った投影法線の計算は，左画像（カメラ1の画像）から右画像

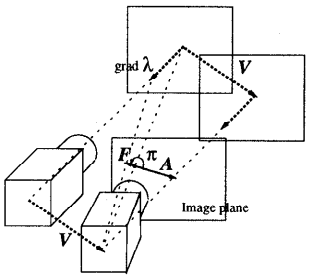


図4 正面画像を撮像した位置からのカメラの移動量 $\boldsymbol{V}$ と，法線 $\mathbf{grad}\,\lambda$ の画像上の投影 $\boldsymbol{A}$，$\boldsymbol{F}$ の関係
Fig. 4 Relation between projections $\boldsymbol{A}$ and $\boldsymbol{F}$ of camera translation $\boldsymbol{V}$ and plane normal $\mathbf{grad}\,\lambda$.

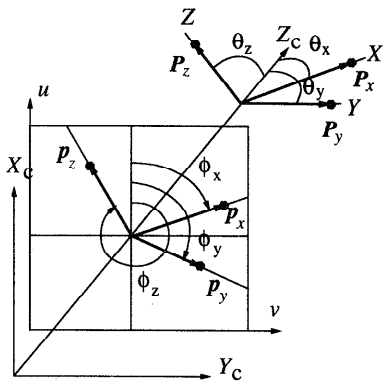


図5 物体上の座標系と画像上の角度の関係
Fig. 5 Relation between the object coordinate system and its angles on the image.

（カメラ2の画像）への3点の変位にも適用できる．2台のカメラの位置関係について正確なキャリブレーションは不要なことが望ましいが，概略の位置関係は既知としても実用上構わない．ここではカメラ1がカメラ2よりも左にあることが分かっているとする．この場合，真のカメラ移動方向 $\angle\boldsymbol{A}$ は $\pm\pi/2$ の範囲に必ずある．今，カメラの概略の移動方向を $\angle\boldsymbol{A}_{ql} = 0$ とし，投影法線方向を求める．これを $\angle\boldsymbol{F}_{ql}$ とすると，真値 $\angle\boldsymbol{F}$ は $\angle\boldsymbol{F}_{ql}$ から $\pm\pi/2$ の範囲に必ずあることが式(13)より確かめられる．したがって，式(14)の2つの解のうち

$$|\angle\boldsymbol{F} - \angle\boldsymbol{F}_{ql}| < \pi/2 \tag{15}$$

を満たす $\angle\boldsymbol{F}$ を選択する．

2.2.4 投影法線上の仮想基準点の位置

ここまでで投影法線の方向 $\angle\boldsymbol{F}$ が求まる．次に仮想基準点 $\boldsymbol{x}_3$ が投影法線上のどの位置にあるかを計算する．

ここで図5のように，各座標軸上に点 $\boldsymbol{P}_x = [1,0,0]^{\mathrm{T}}$，$\boldsymbol{P}_y = [0,1,0]^{\mathrm{T}}$，$\boldsymbol{P}_z = [0,0,1]^{\mathrm{T}}$ をおく．$\boldsymbol{P}_x, \boldsymbol{P}_y$ を入力画像に投影した点 $\boldsymbol{p}_x, \boldsymbol{p}_y$ は，$\boldsymbol{P}_x, \boldsymbol{P}_y$

を正面画像に投影した点 $\boldsymbol{p}_x^0, \boldsymbol{p}_y^0$ を式 (6) で変換することで求められる．ここで，$\boldsymbol{p}_x, \boldsymbol{p}_y$ の u 軸からの角度を ϕ_x, ϕ_y とおく．また，$\boldsymbol{P}_z$ を入力画像に投影した点 $\boldsymbol{p}_z$ の角度 ϕ_z は，式 (15) で求めた $\angle \boldsymbol{F}$ と同じである．

空間上の Z 軸が画像面から見てすべて向こう側を向いている，すなわち $0 < \theta_z < \pi/2$ が成立しているとき，ϕ と θ の間に次のような関係がある[10]．

$$\theta_z = \tan^{-1} \sqrt{\frac{-\cos(\phi_x - \phi_y)}{\cos(\phi_z - \phi_x)\cos(\phi_y - \phi_z)}} \quad (16)$$

もし $\pi/2 < \theta_z < \pi$ ならば，上式の θ_z を $\pi - \theta_z$ に置き換えればよい．

一般には，物体座標系が画像上で拡大縮小されているので，この拡大率を計算する．拡大によって移される点の位置は，任意の画像上の点の位置 $\boldsymbol{x}$ に対して

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{1}{2}(\mathrm{div}\,(\boldsymbol{v}))\boldsymbol{x} \quad (17)$$

と表すことができる[9]．式 (7) より $\mathrm{div}\,(\boldsymbol{v})$ はアフィン変換行列から簡単に求まる．これより拡大率は $1 + 1/2\mathrm{div}\,(\boldsymbol{v})$ である．

したがって，仮想基準点の位置は，$\angle \boldsymbol{F}$ の角度を持つ原点から始まる半直線上にあり，原点からの距離が $(1 + 1/2\mathrm{div}\,(\boldsymbol{v}))\sin\theta_z$ なる点として計算できる．

まとめに，左右画像における 3 点の対応から，アフィン不変量を求めるまでのアルゴリズムを示す．

(1) 最初に 1 度だけ，空間上の 3 点を正面から見たときの画像上の点の位置を初期位置 $\boldsymbol{x}_0^0$, $\boldsymbol{x}_1^0$, $\boldsymbol{x}_2^0$ として記憶する．

(2) 左画像上の 3 つの特徴点 $\boldsymbol{x}_0'$, $\boldsymbol{x}_1'$, $\boldsymbol{x}_2'$ を抽出する．事前に記憶した点の位置 $\boldsymbol{x}_0^0$, $\boldsymbol{x}_1^0$, $\boldsymbol{x}_2^0$ と $\boldsymbol{x}_0'$, $\boldsymbol{x}_1'$, $\boldsymbol{x}_2'$ より式 (6), (7), (8), (9) から div, def を計算する．式 (14), (15), (16), (17) により仮想基準点の位置 $\boldsymbol{x}_3'$ を計算する．

(3) (2) を右画像上の特徴点について行う．

(4) 左右画像上の特徴点と仮想基準点の計 4 点を左右それぞれ $(\boldsymbol{x}_0^l, \boldsymbol{x}_1^l, \boldsymbol{x}_2^l, \boldsymbol{x}_3^l)$, $(\boldsymbol{x}_0^r, \boldsymbol{x}_1^r, \boldsymbol{x}_2^r, \boldsymbol{x}_3^r)$ とおく．式 (3), (4) から，計測対象点 $\boldsymbol{X}_4$ の位置をアフィン座標系上の位置として求める．

2.3 シミュレーション実験

ここでは，式 (8), (9) で述べた近似誤差と，観測位置のノイズの影響をシミュレーションによって確かめる．

計算機上に作る 3 次元空間上の 3 点と仮想点，カメラの関係を**図 6** に示す．空間上の原点付近に 3 点をおき，離れた位置から 2 つのカメラでこの 3 点を撮影する．2 つのカメラの光軸はつねに 3 次元座標系の原点を通るように設定されており，2 つのカメラのベースラインは 400 [mm] 固定にする．原点からベースラインの中心に向かうベクトルは，Z 軸に対して図に示す方向から s の角度をなし，XY 平面に投影した後の，X 軸からの角度が t であるとする．

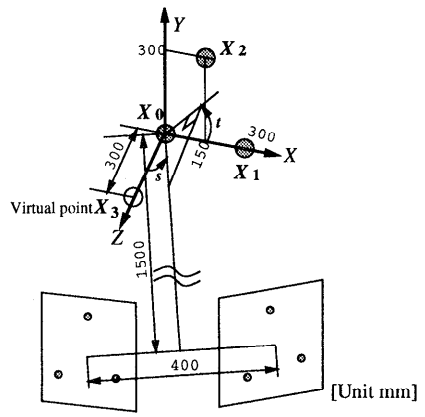
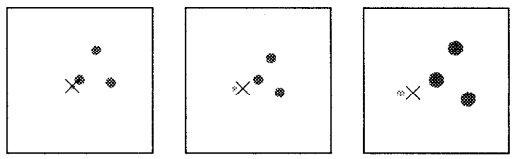


図 6 画像を合成する際のカメラ，特徴点の構成
Fig. 6 Configuration of the cameras and feature points to make the synthesis images.



(a) $s = 25$, $t = 35$　(b) $s = 60$, $t = 35$　(c) $s = 60$, $t = 35$

図 7 投影法線の計算結果
Fig. 7 Experimental results of calculating the normal of a plane.

図 7 に，画像上の 3 点から仮想基準点を求めた結果を示す．×印が仮想基準点，その他の黒丸印が 3 次元空間上の 3 点，白丸印が仮想基準点の真値である．同図 (a) は，$s = 25$ [度]，$t = 35$ [度]，同 (b) は $s = 60$, $t = 35$，同 (c) は $s = 60$, $t = 35$ で，カメラの距離を原点から 1000 [mm] に近付けた場合である．真値とずれがあるのは，式 (8), (9) の近似誤差の影響である．この誤差は s が大きいほど大きくなる傾向がある．このことと，観測位置のノイズの影響を以下の実験で詳しく調べる．

位置の不変量の真値を $\boldsymbol{\alpha}_{true}$，方向の不変量の真値を $\boldsymbol{\beta}_{true}$ とする．2 つの画像上の 3 つの観測点 $\boldsymbol{x}_0^l$, $\boldsymbol{x}_1^l$, $\boldsymbol{x}_2^l$, $\boldsymbol{x}_0^r$, $\boldsymbol{x}_1^r$, $\boldsymbol{x}_2^r$ と，手の位置の観測点 $\boldsymbol{x}_4^l$, $\boldsymbol{x}_4^r$ の各要素にそれぞれ標準偏差 σ_a [pixel] のガウスノイズを加える．さらに，画像上で観測された手の方向と u 軸からの角度 ψ [rad] に標準偏差 σ_b [rad] のガウス

ノイズを加える．2章の手順によって計算した手の位置と方向の不変量をそれぞれ $\boldsymbol{\alpha}_{noise}$，$\boldsymbol{\beta}_{noise}$ とする．$S_N(\cdot)$ を N 個の標本の標本標準偏差，$E_N(\cdot)$ を N 個の標本の標本平均[11]とする．このとき，以下の量

$$\mathrm{IT} = S_N(\|\boldsymbol{\alpha}_{noise} - \boldsymbol{\alpha}_{true}\|) \tag{18}$$

$$\mathrm{IA} = S_N(\|\boldsymbol{\alpha}_{noise} - E_N(\boldsymbol{\alpha}_{noise})\|) \tag{19}$$

$$\mathrm{DT} = S_N\left(\cos^{-1}\frac{\boldsymbol{\beta}_{noise}\cdot\boldsymbol{\beta}_{true}}{\|\boldsymbol{\beta}_{noise}\|\|\boldsymbol{\beta}_{true}\|}\right) \tag{20}$$

$$\mathrm{DA} = S_N\left(\cos^{-1}\frac{\boldsymbol{\beta}_{noise}\cdot E_N(\boldsymbol{\beta}_{nose})}{\|\boldsymbol{\beta}_{noise}\|\|E_N(\boldsymbol{\beta}_{noise})\|}\right) \tag{21}$$

を定義する．ここで，$\boldsymbol{x}\cdot\boldsymbol{y}$ は，$\boldsymbol{x}$ と $\boldsymbol{y}$ との内積である．ITが小さければ $\boldsymbol{\alpha}_{noise}$ が真の位置に近いことを意味する．IAが小さければ $\boldsymbol{\alpha}_{noise}$ が $\boldsymbol{\alpha}_{noise}$ の平均値 $E_N(\boldsymbol{\alpha}_{noise})$ に近いことを意味する．DTが小さければ $\boldsymbol{\beta}_{noise}$ が真の方向に近いことを意味する．DAが小さければ，$\boldsymbol{\beta}_{noise}$ が $\boldsymbol{\beta}_{noise}$ の平均値 $E_N(\boldsymbol{\beta}_{noise})$ の方向に近いことを意味する．

2.3.1 正面画像上の観測点のノイズの影響

ここでは正面画像の観測点 $\boldsymbol{x}_0^0$，$\boldsymbol{x}_1^0$，$\boldsymbol{x}_2^0$ 上にノイズを加えた場合と，ノイズを加えない場合のITの差およびIAの差を調べる．$N = 10^4$ [回]，$t = 35$ [度] に固定し，不変量の真値を $\boldsymbol{\alpha}_{true} = [1,1,1]^{\mathrm{T}}$ に固定する．これは，アフィン座標系を体上にとるとき，実際の手の位置がすべての軸上でほぼ1のオーダで計算されるからである．このとき，アフィン座標軸の実際の長さを 300 [mm] とすると，不変量の1が 300 [mm] に相当する．正面画像に $\sigma = 10/3$ [pixel] のノイズを加えた場合のITを IT_I とする．$\sigma = 10/3$ [pixel] のノイズは，512 × 512 の画像上で ±10 [pixel] 以下の誤差がガウス分布の $\pm 3\sigma \approx 0.997$ の確率で発生する程度のものである．後に示す実際のシステムでは，観測点として肩と腕の稜線の交点の位置など観測ごとの偏差が大きいものを利用するので，この程度のノイズを考えておく必要がある．また，ノイズを加えない場合のITを IT_0 とする．これより $30 \le s \le 70$ の区間で，ノイズを加えた場合と加えない場合の差の絶対値の平均は，$\int_{s=30}^{70} |\mathrm{IT}_I - \mathrm{IT}_0|/(70-30) \approx 0.0252$ である．これは，実際のアフィン座標軸が 300 [mm] の場合，7.6 [mm] に相当する．同様にIAについて，$\int_{s=30}^{70} |\mathrm{IA}_I - \mathrm{IA}_0|/(70-30) \approx 0.0106$ であり，この値は 3.2 [mm] に相当する．当然，より小さな σ のノイズを加えた場合では上記の値はもっと小さくなる．そこで，以下の実験では正面画像にノイズを加えない．また，この事実は2.2.2項で述べたように，システムをヒューマンインタフェースに用いる場合であれば，正面画像を入力するときに正確に正面を向かなくてもシステムの要求精度を満たすことができることを示す．

2.3.2 位置の不変量の誤差

以下では位置の不変量の誤差を代表する値IT，IAを計算する．**図8** (a) は，上記と同様に $N = 10^4$ [回]，$t = 35$ [度]，$\boldsymbol{\alpha}_{true} = [1,1,1]^{\mathrm{T}}$ に固定し，s を0から90 [度] まで変化させたときのITを計算したものであり，(b) は同様にしてIAを計算したものである．また，(a)，(b) ともに破線はノイズを加えない場合，一点鎖線は $\sigma = 1/3$ [pixel] のノイズ，実線は $\sigma = 10/3$ [pixel] のノイズを加えた場合である．$\sigma = 1/3$ [pixel] のノイズは，特徴点抽出のためのマークをつけるなどしてほぼ正確に特徴点の位置が決定できる場合を想定している．また，$\sigma = 10/3$ [pixel] のノイズは前述のとおりである．

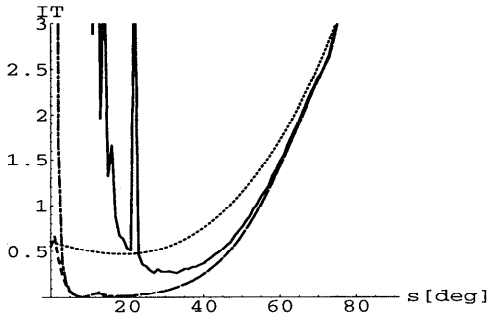
まず (a) は，式 (8)，(9) での近似誤差を裏付ける結果である．つまり観測にノイズがあってもなくても，s が大きくなるに従ってITが大きくなっていく．また，s が小さい領域でノイズの影響が顕著になっている．これは，空間中の基準点が正面に近いときには s を変化させたときの観測点の位置変化が小さく，相対的にノイズの影響が大きくなるためである．ここで，図の点線は基準点の観測つまり $\boldsymbol{x}_0^l$，$\boldsymbol{x}_1^l$，$\boldsymbol{x}_2^l$，$\boldsymbol{x}_0^r$，$\boldsymbol{x}_1^r$，$\boldsymbol{x}_2^r$ を $s = 0$ での観測値に固定し，手の位置 $\boldsymbol{x}_4^l, \boldsymbol{x}_4^r$ のみを s とともに変化させたときのITである．この場合も s が大きくなるに従ってITが大きくなるが，ノイズの影響を受けにくいため，$0 \le s \le 33$ で $\mathrm{IT} < 0.61$ と，s が小さい領域で基準点の観測にノイズを加えた場合のITより値が小さい．したがって，ある適当な値 s_{split} について

(1) $s \le s_{split}$ のとき：
  (i) $s > s_{split}$ から $s \le s_{split}$ に遷移したとき：現在の基準点の観測を捨て，直前の基準点の観測位置を用いて不変量を計算する．
  (ii) (i) 以外では，基準点の観測位置を固定したままにする．

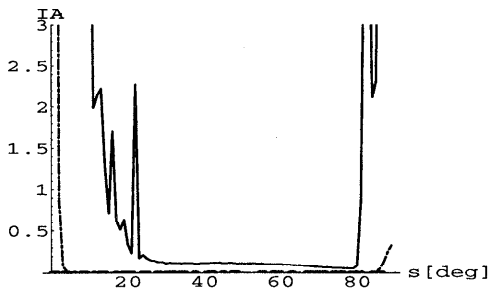
(2) $s > s_{split}$ のとき：通常の処理

のように処理を分けると全体の性能が向上する．観測からは直接 s の値が計算できないので，ある適当な値 s_{split} のときの $d_{split} = \mathrm{def}(v)$ をシミュレーションであらかじめ計算しておき，$s \le s_{split}$ を $\mathrm{def}(v) \le d_{split}$，$s > s_{split}$ を $\mathrm{def}(v) > d_{split}$ と置き換える．

次に (b) では，$\sigma = 1/3$ [pixel] のとき，$5 < s \le 83$ の範囲で $\mathrm{IA} < 0.010$ である．また，$\sigma = 10/3$ [pixel]

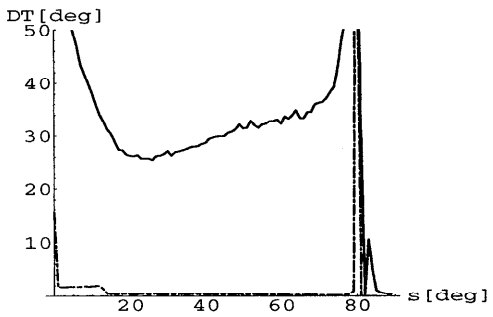


(a) s と IT との関係
IT against s.

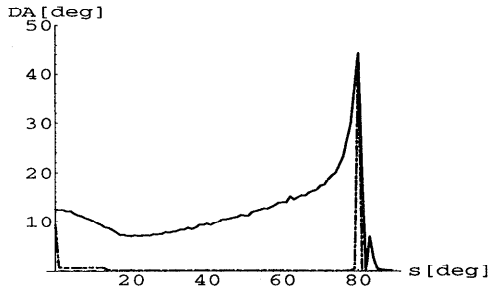


(b) s と IA との関係
IA against s.

図 8 カメラの視点と位置誤差（IT, IA）との関係
Fig. 8 Relation between positional errors and camera viewpoints.



(a) s と DT との関係
DT against s.



(b) s と DA との関係
DA against s.

図 9 カメラの視点と方向誤差（DT, DA）との関係
Fig. 9 Relation between orientation errors and camera viewpoints.

のとき，$29 < s \leq 80$ の範囲で IA<0.13 である．このように，s の広い領域で IA の値が小さくなっているが，やはり s の小さな領域でノイズの影響が顕著になっている．ここで (a) で考えたのと同じように，s の小さな領域で基準点の観測を正面に固定すれば IA ≈ 0.030 となる．したがって，上記と同様に処理を分けた方が全体の性能が向上する．

2.3.3 方向の不変量の誤差

以下では方向の不変量の誤差を代表する値 DT, DA を計算する．**図 9** (a) は，$N = 10^4$ [回]，$t = 35$ [度]，$\boldsymbol{\alpha}_{true} = [1,1,1]^{\mathrm{T}}$，$\boldsymbol{\beta}_{true} = [1,0,1]^{\mathrm{T}}$ に固定し，s を 0 から 90 [度] まで変化させたときの DT を計算したものであり，(b) は同様にして DA を計算したものである．ここで画像上の手の方向を表す直線が，腕の両端点を結ぶ直線であるとする．このとき，腕の長さが画像上で 100 [pixel] 程度の長さであれば，手の先が ± 1 [pixel] 程度の誤差を持つ場合，角度の誤差は $\sigma_b = 0.009/3$ [rad] (0.5 [度]) 程度となる．同様に，手の先が ± 10 [pixel] 程度の誤差を持つとき，$\sigma_b = 0.09/3$ [rad] (5 [度]) 程度となる．以下では，(a), (b) ともにノイズを加えない場合を破線に示し，$\sigma_a = 1/3$ [pixel]，$\sigma_b = 0.009/3$ [rad] (0.5 [度]) のノイズを加えた場合を一点鎖線に示し，$\sigma_a = 10/3$ [pixel]，$\sigma_b = 0.09/3$ [rad] (5 [度]) のノイズを加えた場合を実線に示す．

(a) では，ノイズを加えない場合 $0 \leq s \leq 90$ の区間で DT $< 1.0 \times 10^{-13}$ [度]，$\sigma_b = 0.5$ [度] の場合 $1 \leq s \leq 79$ の区間で DT < 2.0 [度]，$\sigma_b = 5$ [度] の場合 $0 \leq s \leq 77$ の区間で DT < 60 [度] の誤差となっている．また，$s = 80$ [度] 付近で非常に大きな値（ノイズのある場合，100 [度] 程度）をとるのは，式 (8), (9) の近似誤差が大きくなり過ぎて，計算が不安定になるためと考えられる．また，s が小さい領域で大きくなるが，図 8 に示すように，位置の不変量の誤差ほど極端ではない．ただし，もし 3 次元空間上の方向ベクトルがエピ極面上にあるときには，そのエピ極面上のすべての方向ベクトルが解となり，一意に定まらない問題がある．この状況は $\boldsymbol{\beta}_{true}$ のとりうる値の一部で発生し，頻繁に発生するものではない．したがって，これを無視したとしても十分現実的である．もしこれが問題となる場合には，複数カメラを用いて，安定に方向が求まるカメラ対を使うなどの問題の回避が必要である．

(b) では，同様にノイズを加えない場合 $0 \leq s \leq 90$ の区間で DA $< 1.0 \times 10^{-26}$ [度]，$\sigma_b = 0.5$ [度] の場合 $1 \leq s \leq 79$ の区間で DA < 0.70 [度]，$\sigma_b = 5$ [度] の場合 $0 \leq s \leq 54$ の区間で DA < 13 [度] の誤差となっている．(a) と同様に 80 [度] 付近で大きな値をとる．また，DT よりも DA の値を小さくすることが重要であるが，全体として DT より小さな値におさえられている．

以上より，$\sigma_a = 1/3$ [pixel]，$\sigma_b = 0.009/3$ [rad] の場合，$s_{split} = 5$ [度] とすれば，$0 \leq s \leq 48$ の領域で IT < 0.61，IA < 0.010（不変量の 1 が実世界の 300 [mm] に相当する変位の場合，3 [mm] 未満），DT < 2.0 [度]，DA < 0.70 [度]，IT，DT の値を問題にしないのであれば $0 \leq s \leq 79$ で IA < 0.010，DA < 0.70 [度] を実現できる．また，$\sigma_a = 10/3$ [pixel]，$\sigma_b = 0.09/3$ [rad] の場合でも $s_{split} = 29$ [度] とすれば，$0 \leq s \leq 47$ の領域で IT < 0.61，IA < 0.13（前述括弧内の条件と同じ場合，39 [mm] 未満），DT $<$ 60 [度]，DA < 13 [度]，IT，DT の値を問題にしなければ $0 \leq s \leq 54$ で IA < 0.13，DA < 13 [度] が可能である．あるいは，最低 2 つのカメラでつねに $s_{split} \leq s \leq 47$ となるように複数のカメラを設置し，それらのカメラの観測から不変量を計算しても，同じ精度を実現できる．また，複数カメラを使えば，方向の不変量が不安定となる問題を回避できる．

本システムでは，視覚フィードバックによってユーザが考えた位置に正確に指示することが目標である．したがって，不変量の真値（IT）は重要でなく，どれだけ正確に相対的な指示ができるか（IA）が重要である．このことと，$\sigma = 10/3$ [pixel]，$\sigma_b = 0.09/3$ [rad] のように比較的ノイズの大きな場合でさえ IA < 0.13，DA < 13 [度] であることから，本システムが十分有効であるといえる．

3. プレゼンテーションシステム

本手法のヒューマンインタフェースへの応用はいろいろ考えられる．たとえば，仮想世界に入り込んだユーザの様々な指示の解釈や，商品プレゼンテーション時の商品モデルを，様々な角度から見せるためのユーザの指示の解釈などである．これらの指示は多くの場合ユーザを中心とする指示であるので，本手法が有効であることが期待できる．たとえば，プレゼンテーションシステムの場合，ユーザが様々な場所，様々な方向からスクリーンに向かって指示できることが望ましいので，システムの構成として複数（2 つ以上）のカメラで随時ユーザを追跡し，特徴点が正確に抽出されているカメラ対を使ってアフィン不変量を求めるのがよい．

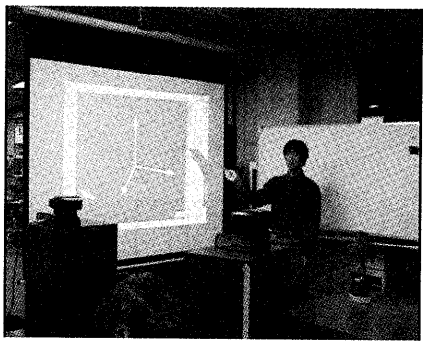


図 10 プレゼンテーションシステム概観
Fig. 10 Presentation system overview.

今回，ヒューマンインタフェースへの応用例として，簡単なプレゼンテーションシステムを作成した．このプレゼンテーションシステムでは，商品に見立てた仮想物体（飛行機モデル）を 3 次元 CG として描き，それをユーザの手指示で様々な方向から見ることができるものである．実際には多くのカメラを使うのが望ましいが，本実験では，2 つのパンチルトカメラ（SONY EVI-D30）を使って実験した．このカメラは，カメラ自身が色情報を用いて人物を追跡する機能を持つ．アフィン座標系は明示的なカメラキャリブレーションを必要としないので，パンチルトの具体的な値を知る必要がない．したがって，カメラのこのような機能を簡単に利用することができる．

システムの全体像を図 **10** に示す．手前の 2 つのカメラで，ユーザを追跡し，手の位置と方向のアフィン不変量から CG の表示位置，方向を決定してスクリーンに映し出す．

本実験では，基準となる特徴点 3 点を，顔の上端点，首の下端点，左肩の点とした．これらの点は，正面画像上で直角に近く，かつ特別なマークをつけなくても抽出できる．

まずオペレータによって与えられた肌色領域内の，すべての画素について色相 (H)，明度 (L)，彩度 (S) の標本平均 (H_m, L_m, S_m) と標本標準偏差 (H_d, L_d, S_d) を最初に 1 度だけ計算して記憶する．以後以下の評価式

$$H_m - H_t H_d \leq H \leq H_m + H_t H_d \quad (22)$$

$$L_m - L_t L_d \leq L \leq L_m + L_t L_d \quad (23)$$

$$S_m - S_t S_d \leq S \leq S_m + S_t S_d \quad (24)$$

を満たす領域を肌色領域として抽出する．ただし (H_t, L_t, S_t) は閾値であり，今回は $(1.2, 2.0, 3.0)$ とした．抽出された領域のうち最大面積のものを顔領域とし，顔領域の上端と首の下端を顔上の 2 つの特徴点とする．

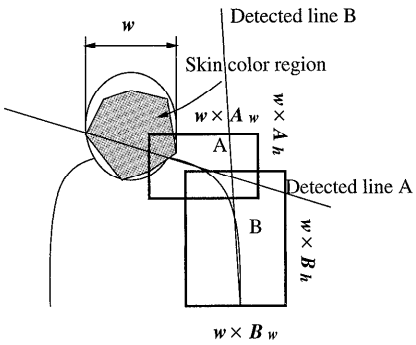
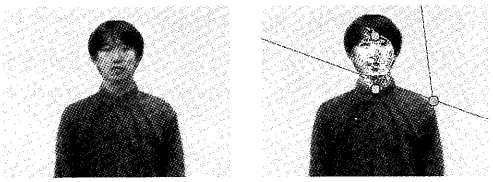


図 11 直線検出のための領域
Fig. 11 Regions for the line detection.



(a) 正面画像
(Frontal image)

(b) 正面画像上の特徴点
(Features)

図 12 正面画像
Fig. 12 Frontal images.

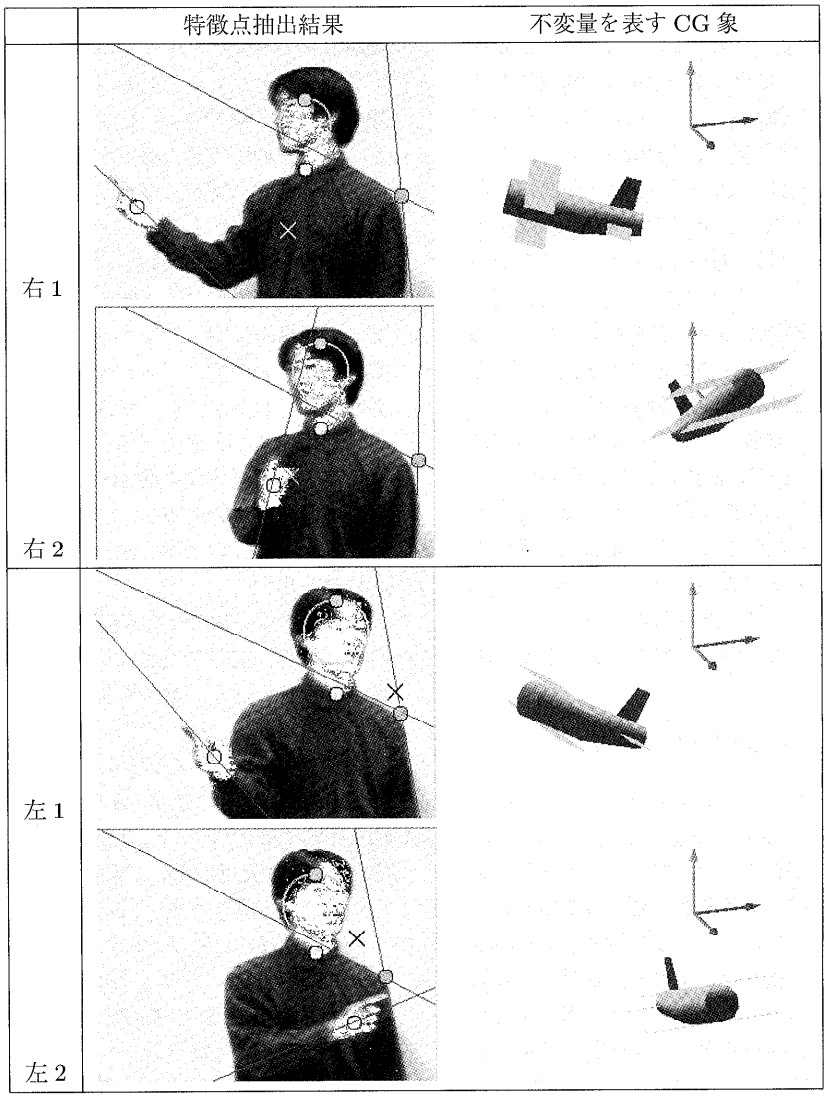


図 13 実験結果
Fig. 13 Experimental results.

続いてエッジ画像からハフ変換[12]を用いて肩と腕の直線検出を行う．**図11** に示すようなA，B 2つの領域をそれぞれ肩の領域，腕の領域として直線検出する．これらの領域はすでに得られた顔領域の幅 w から相対的な位置，大きさに設定する．さらに，テクスチャや背景のエッジの影響を取り除くために，Aの領域では顎の近くを通る直線に，Bの領域では垂直に近い直線に限定する．得られた2直線の交点を肩の特徴点とする．ここまでの画像処理例を**図12** に示す．図では各特徴点の位置を丸印で示してある．

このようにして抽出された体上の3特徴点の位置から2.2節の手順で仮想基準点の位置を求める．

次に手の位置を検出する．顔の領域を抽出したのと同様の方法で，手の肌色領域を抽出し，その重心を手の位置とする．また，領域の2次モーメントを計算して，画像上の指さし方向を検出する．

2.1節の手順で手の位置，方向の不変量を求める．これを首の下端点を原点とし，左肩の点を X 軸，顔の上端点を Y 軸，仮想基準点を Z 軸とする座標系での不変量と見なし，これに応じたCG像を表示する．

特徴点抽出からCG表示までの実験結果を**図13** に示す．左列の上2行はカメラに対して体が右向きの状態，下2行は左向きの状態の画像である．ただし，1，3行めと2，4行めでは人間の方はそれぞれほぼ同じポーズをとっている．図の特徴点抽出画像で，丸印が特徴点の位置，×印が仮想点の位置を表す．図の右列の飛行機の位置と方向が手の位置と方向を表す．カメラに対する体の向きにかかわらず，ユーザの体を中心にした座標系で手の位置，方向がほぼ正しく推定されている様子が分かる．このように表示結果を見ながら使うものでは，表示が思ったようになるまでユーザが手を動かせばよい．実験から，このような使用目的に対しては問題のない結果が得られた．

4. おわりに

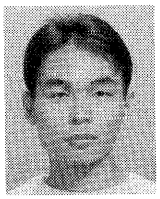
実際のヒューマンインタフェースにおいては，ユーザの自由度を高めることが非常に重要である．先の論文では画像から4つの特徴点を抽出するため，椅子に座った状態でしか操作できなかった．本論文では上半身から得られる3つの特徴点から仮想基準点を求め，これらをアフィン座標系とすることでユーザの姿勢の制約を取り除き，自由なヒューマンインタフェースを実現することができた．

今後は多くのカメラが同時にユーザを追跡し，ユーザがどこにいても，どの方向を向いていても構わないシステムを作成していきたい．

参考文献

1) Takemura, H. and Kishino, F.: Cooperative work environment using virtual workspace, *CSCW 92*, pp.226–232 (1992).
2) Torige, A. and Kono, T.: Human-interface by recognition of human gestures with image processing recognition of gesture to specify moving directions, *IEEE International Workshop on Robot and Human Communication*, pp.105–110 (1992).
3) Vinther, S. and Cipolla, R.: Towards 3D Object Model Acquisition and Recognition using 3D Affine Invariants, Technical Report CUED/F-INFENG TR136, Cambridge University Engineering Department, England (1993).
4) Mundy, J.L. and Zisserman, A. (Eds.): *Geometric Invariance in Computer Vision*, MIT Press (1992).
5) Jo, K.-H., Hayashi, K., Kuno, Y. and Shirai, Y.: Vision-based human interface system with world-fixed and human-centerd Frames using multiple view invariance, *IEICE Trans. Information and Systems*, Vol.E79-D, No.6, pp.219–228 (1996).
6) 林健太郎，久野義徳，白井良明：ユーザ中心と世界固定の視点による空間指示動作の解釈とその応用，信学技報，Vol.95, No.165, pp.85–90 (1995).
7) 岡本恭一，Cipolla, R.，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，Vol.J76-D-II, No.8, pp.1813–1821 (1993).
8) Waxman, A. and Ullman, S.: Surface structure and three-dimensional motion from image flow kinematics, *International Journal of Robotics Research*, Vol.4, No.3, pp.72–94 (1985).
9) Cipolla, R.: *Active Visual Inference of Surface Shape*, Springer (1992).
10) 金谷健一：画像理解—3次元認識の数理，森北出版 (1990).
11) 日本数学会：数学辞典第3版，岩波書店 (1985).
12) Shirai, Y. (Ed.): *ThreeDimensional Computer Vision*, Springer-Verlag (1987).

(平成10年 4 月16日受付)
(平成10年12月 7 日採録)

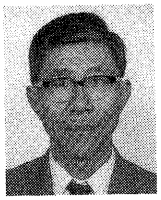


林 健太郎

1972年12月18日生．1995年，大阪大学工学部電子制御機械工学科卒業．1996年，同大学大学院修士課程修了．同年，同大学大学院博士課程入学．コンピュータビジョンの研究に従事．ヒューマンインタフェース，3次元認識等に関心．

久野 義徳（正会員）

1954年4月13日生．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4月より大阪大学工学部電子制御機械工学科助教授．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各会員．



白井 良明（正会員）

1941年8月3日生．1964年名古屋大学工学部機械工学科卒業．1969年東京大学大学院工学系博士課程修了，工学博士．同年，電子技術総合研究所入所，コンピュータビジョン，ロボティクスの研究に従事．1971～1972年，MIT AIラボ客員研究員．1988年大阪大学工学部電子制御機械工学科教授．人工知能学会，電子情報通信学会，日本機械学会，ロボット学会，計測自動制御学会各会員．