「棒」入力システムのためのジェスチャ認識の実現

大橋 健[†] 仲程 啓[††]
吉田隆一[†] 江島俊朗[†]

我々は数年前から「棒」が使えるインタフェース環境を構築することを目指して研究を続けてきた．「棒」入力システムでは，青と赤に色分けされた「棒」の動きをカメラでとらえ，動画像処理を施して「棒」の軌跡を抽出し，軌跡パターンを認識・理解をすることで，「棒」の動きにこめられたジェスチャの意味を理解している．しかしながら，「棒」入力システムの構築作業を行う中で，システムの反応の遅さ，「棒」振りの速度に関する制約の強さ，およびジェスチャの認識精度の低さなど種々の改善すべき点が浮き彫りになってきた．これらの問題点を解決するために，本論文では，ボトムアップ的処理（画像処理，パターン認識処理）の改善を図るとともに，トップダウン的な制約，特に，ジェスチャで使用する言語の文法規則に工夫を施すことを行う．すなわち，ジェスチャ言語としては，使用目的に合致し使いやすく，かつ，言語の文法的制約によりジェスチャ認識処理の精度向上が望めるものを設計する．まず，「棒」の動きを止める動作「停止」を始点と終点に置くという言語の設計をすることにより，一文字単位の切り出し精度の向上を図るとともに画像処理や認識処理効率化を図る．次いで，認識に有効な特徴（方向特徴，特異点）と「停止」という区切り情報を統一的に扱うことを可能にする，「棒」の軌跡の（時空間上での）標本化方式を新たに提案する．この方式の導入により，「棒」の移動速度の変動に強く対処できる正確な特徴抽出と区切り記号の抽出が可能になる．最後に，新たな検討のもとに構成した「棒」入力システムは，コマンド入力として有用であることを示す．特に，手が不自由なユーザの入力手段として，その応用が大いに期待できる．

A Gesture Recognition Method for a Stick Input System

TAKESHI OHASHI,[†] KEI NAKAHODO,[††] TAKAICHI YOSHIDA[†]
and TOSHIAKI EJIMA[†]

We are developing a human interface using a stick as the input device. With the stick input, the system captures the painted blue and red parts of the stick by a camera and extracts the stick trajectory by image processing. After that the system recognize the gestures drawn by the stick. We constructed a prototype system using a recognition system that can accept some commands. Some problems such as system response time, the drawing speed limitation, and the gesture recognition accuracy are evaluated. To solve these problems, the system has to improve its bottom up approach and the top down approach. The former includes image processing and pattern recognition. While the latter uses a sign language grammar. The language should be applicable for the task and can support to improve the gesture recognition. In this paper, we designed the sign language combining the pre-pause, sign and post-pause. This restriction increases the accuracy of the sign segmentation. Then we integrate the spatial directional featurs and pause, which are useful in character recognition. Furthermore we proposed a resampling method along the spatial temporal trajectory. This method provides accuracy of the detection featur and stability against temporal distortion. Finally the prototype system showed the stick as a useful input device. This is very beneficial to users with incapacitated hands allowing them to utilize the system using other ports of the body.

1. はじめに

言葉だけで自分の意志を伝えようとするとしばしばもどかしさを感ずることがある．そんなとき，「身振り，手ぶり」を添えることでより自然に自分の意志を伝えることが可能になる．人間とコンピュータのコミュニケーションも然りである．思い浮かべたことを「身振り，手ぶり」で表現することにより，コンピュータとのコミュニケーションが自由に，容易に，そして楽しく行えるようになれば嬉しい．そのような「情報環境」のもとでは想像力が豊かになり，創造性に富んだ多くのものが創出されることが期待される．もし，

† 九州工業大学情報工学部
Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology
†† 大阪大学医学部機能画像診断学
Division of Functional Diagnostic Imaging, Osaka University Medical School

コンピュータが人間のような視覚を備え，オーケストラの指揮者のごとく振られる「棒」の動きを認識・理解できるようになれば，コンピュータが創り出す仮想現実の世界において，「棒」はただの棒ではなく変幻自在の「魔法の棒」として機能する．

このような考えのもとに，我々は数年前から「棒」が使えるインタフェース環境を構築することを目指している[1)~6)]．本システムでは，青と赤に色分けされた「棒」の動きをカメラでとらえ，動画像処理を施して「棒」の軌跡を抽出し，軌跡パターンを認識・理解することで「棒」の動きに託されたジェスチャの意味を理解する．しかしながら，「棒」入力システムの構築作業を行う中で，システムの反応の遅さ，「棒」振りの速度に関する制約の強さ，およびジェスチャの認識精度の低さなど種々の改善すべき点が浮き彫りになっている．

これらの問題点を解決するために，本論文では，ボトムアップ的処理（画像処理，パターン認識処理）の改善を図るとともに，トップダウン的な制約，特に，ジェスチャで使用する言語の文法規則に工夫を施す．すなわち，ジェスチャ言語としては，使用目的に合致し使いやすく，かつ，言語の文法的制約によりジェスチャ認識処理の精度向上が望めるものを設計する．

工夫を施した提案手法の性能を評価するため，「棒」入力システムを構築し，あわせて応用の可能性を調査する．その結果，コマンド入力として「棒」が使えること，特に，手の不自由なユーザの入力手段として大いに期待できることを示す．

2. 関連研究と本研究の位置付け

コンピュータにジェスチャを認識させる手法は，Data Glove などのデバイスを用いる方法[7)~9)]や画像処理を用いる方法[10)~16)]に大別できる．前者は，各指の曲げの状態や位置を検出できるセンサーを用いるので手の姿勢を精度良く計測できる．その反面，装置の設置や調整が必要となりあまり手軽には使えない．後者は，撮影カメラとユーザの間を結ぶケーブルが不要であり自然なインタフェースが実現できる．しかしながら，画像処理を用いているのでオクルージョンが問題となり手の姿勢を同定するのは困難である．ジェスチャを認識するために，手のモデルを単純なスティックモデルで表すアプローチ[17)]や手の輪郭と3次元モデルの対応を推定するアプローチ[18)]などが提案されている．いずれの場合も，曖昧性が含まれてしまい，これを解消するためには複数のカメラからの情報を対応付けるなどが必要となる．文献19)では，各指を色分けして指のオクルージョン問題の解消を図っているが，すべての指が別々に彩色された手袋などを付けなければならない．オクルージョン問題以外でも実時間で画像処理するためには色マーカを付ける手法[20)]が有効であると考えられる．リアルタイムに安定して対象物体を抽出できる手法として，対象物体の位置と色を統計的な blob モデルとして扱う P-finder[21)]なども提案されている．しかし，指の姿勢に至る詳細な形状の測定には向いていない．また，ステレオ法における対応点問題や設置の容易さを考慮すると単眼カメラを用いて処理できることが望ましい[20)]．

これらの研究をふまえて，対象に合わせて適切にインタフェースを設計すれば，ジェスチャ認識が活用できることが示されている[12)~14)]．これらジェスチャ認識を用いたインタフェースでは，一連の動作の中でジェスチャとして意味を持つ部分を特定するスポッティングが重要である[22),23)]．文献13)の指揮者の動きを認識するシステムにおいては，手の位置よりも速度情報の有効性が示されており，スポッティングにおいても速度情報が有益であると考えられる．ジェスチャの切り出しや認識は，オンライン手書き認識[24)~27)]との関連が高いが，棒で空間中に描画するときは物理的な測定面との接触が得られず，描画平面の推定とスポッティングがより重要である．

また，近年人に優しいインタフェースとして何らかの障害を持つユーザが扱えるインタフェース[28)~31)]への関心も高まっている．「棒」は身体の一部に固定してそこを動かすことさえできれば，手に障害を持つユーザでも利用できる．「棒」では3次元空間中のポインティングが容易にできる利点を持っているので，これで描くジェスチャが認識できれば表現の範囲が広がり，これらのユーザが「棒」のみで仮想世界とインタラクションする応用が考えられる．

本論文では，ユーザは二次元平面を想定しその平面上にジェスチャを描くと仮定する．健常者は手でジェスチャを描き，手の不自由なユーザは「棒」を口でくわえてジェスチャを描くと想定する．ユーザは各自の描きやすい平面上にジェスチャを描き，システムがその描画平面を推定し，描かれたジェスチャの認識を行う．

このような状況のもとで画像処理の精度と処理速度を考慮して，ジェスチャ言語の設計を行い，観測データからジェスチャ部分切り出し，描画平面の推定，ジェスチャ認識を行う手法を提示する．認識対象を0から9までの数字10文字とし，数字の組合せを命令として解釈する入力サーバを作成し，既存のアプリケーションのコマンドを呼び出してブラウジングできる試作イ

ンタフェースを示す.

3. 「棒」の3次元軌跡の抽出

3.1 システムの概観

図1 はシステム全体の概念図を示している. 図1のようにユーザはコンピュータあるいは仮想世界と「棒」のジェスチャや音声を用いて対話する. ここで, コンピュータ側は少なくとも1台以上のカメラから画像を取り込み, マイクロフォンから音声が入力できるものとする. 音声と画像はそれぞれ別々に認識処理を経た後に統合され命令等としてシステムに解釈される. ディスプレイについては, 図1では大型ディスプレイの場合を示しているが, 対象となるアプリケーションに合わせてヘッドマウンテッドディスプレイ (HMD) 等に変えてよい. 音声認識関係の説明は省略し, 画像処理の部分のみに関して説明する.

3.2 座標系とカメラモデル

「棒」は, **図2** のように先端から青と赤と白の3色に色分けする. ここで, 後述の3次元座標の推定を容易にするため, 青と赤に塗る長さは同じ長さとする. カメラは固定焦点のピンホールカメラモデルを仮定し, カメラの視線方向を Z 軸方向を合わせることにより, 撮像面上に観測された座標からカメラ座標系の座標を推定する.

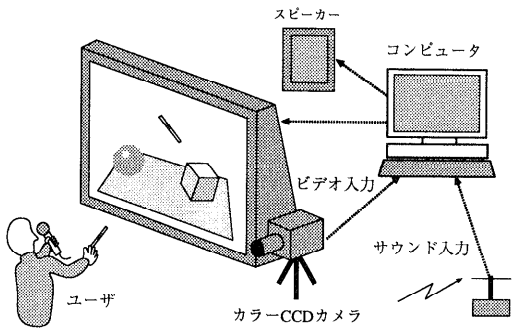


図1 システムの概念図
Fig. 1 The concept of the system.

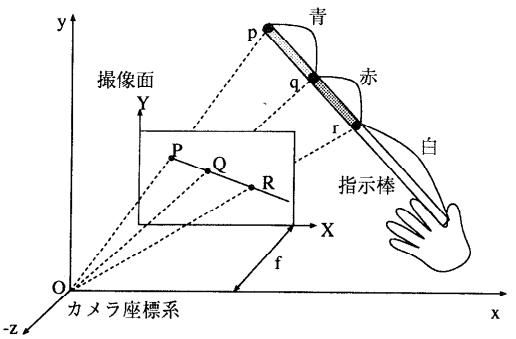


図2 「棒」と座標系の関係
Fig. 2 Coordinate system and the stick.

画像処理は以下の処理から構成される.

(1) 取り込んだ画像中から「棒」領域を抽出する処理
(2) 2次元座標から3次元座標を計算する処理
(3) 時系列のデータからジェスチャを認識する処理

(1), (2) の処理を以下に記し, (3) は次章で詳細を述べる.

3.3 「棒」領域の抽出

まず, 画像中から青と赤の領域を抽出することを考える. 照明の変化の影響を避けるためYUV表色系など輝度と色情報を分離した表現系を利用する. 操作開始に先立ち, あらかじめ「棒」の画像から青と赤の参照色を標本しておき, 抽出処理ではこの参照色との類似度を基準としたしきい値処理を施し2値画像を得る.

次に, 青と赤の参照色との類似度を値とする画像を縦横方向に投影してヒストグラムを作成する. これらから, 「棒」の青領域と赤領域の候補領域を選び出す. 「棒」の青と赤の領域は連続して直線上にあるはずなので, この制約を用いて雑音を除去した後, 正しい領域の組を抽出する.

3.4 3次元座標の推定

「棒」の先端, 青と赤, 赤と白の領域の境目の3点 (図2では点 p, q, r) に着目すると, これらは3次元空間中で直線上に等間隔 k で並んでいる. このとき, これらの点は, カメラの焦点距離などのパラメータにより透視変換され, 図2のように2次元の画像中に写像される. 透視変換により写像された青と赤の領域は, 2次元平面上のやはり直線上に隣接する領域になる.

カメラの座標系を中心投影座標系で考え, 図2のようにカメラの焦点を3次元空間の原点として, 視線の向きに z 軸をとり, カメラの撮像面の水平垂直方向をそれぞれ x 軸, y 軸とする. そして, 撮像面は z 軸と垂直に距離 f の位置にあるとし, その座標軸を X 軸, Y 軸とする. このとき, カメラ座標系における「棒」の先端の点の座標を $p(x_p, y_p, z_p)$, 青と赤の境目の点の座標を $q(x_q, y_q, z_q)$, 赤と白の境目の点の座標を $r(x_r, y_r, z_r)$ とする. そして, これらの点が透視変換され撮像面に作る像の座標をそれぞれ $P(X_P, Y_P)$, $Q(X_Q, Y_Q)$, $R(X_R, Y_R)$ とする. このとき, 点 (x, y, z) と点 (X, Y) の間には $X = fx/z$, $Y = fy/z$ という関係があるので,

$$\begin{cases} p(X_P z_p/f, Y_P z_p/f, z_p) \\ q(X_Q z_q/f, Y_Q z_q/f, z_q) \\ r(X_R z_r/f, Y_R z_r/f, z_r). \end{cases} \quad (1)$$

ここで，$z_p/f = l$，$z_q/f = m$，$z_r/f = n$ とおくと，

$$\begin{cases} p(lX_P, lY_P, lf) \\ q(mX_Q, mY_Q, mf) \\ r(nX_R, nY_R, nf). \end{cases} \quad (2)$$

ここで $\vec{pq} = \vec{qr}$ より，

$$\begin{aligned} &(lX_P - mX_Q, lY_P - mY_Q, lf - mf) \\ &= (mX_Q - nX_R, mY_Q - nY_R, mf - nf). \end{aligned} \quad (3)$$

また，$\bar{pq} = k$ より

$$\begin{aligned} &(lX_P - mX_Q)^2 + (lY_P - mY_Q)^2 + (lf - mf)^2 \\ &= k^2. \end{aligned} \quad (4)$$

これらを解いて，

$$\begin{cases} l = \dfrac{k}{\sqrt{(X_P - X_Q g)^2 + (Y_P - Y_Q g)^2 + (f - fg)^2}} \\ n = hl \\ m = (l + n)/2. \end{cases} \quad (5)$$

ただし，$g = (1+h)/2$，$h = (X_Q - X_P)/(X_R - X_Q)$ である．

このようにして，観測された画像から求められる「棒」の2次元座標 P，Q とカメラの焦点距離 f から式 (2) に代入することにより，カメラ座標系における3次元座標 p，q，r を推定することができる．3次元空間中の同一直線上に順番に並ぶ3点の座標が得られるので，「棒」の方向ベクトルも同時に獲得できる．

3.5 計測が有効な条件

「棒」の3次元座標の測定は，カメラに対する「棒」の角度によって測定精度が大きく変化する．**図3** は「棒」の角度をカメラに対して変えたとき，3次元情報に現れる測定誤差を「棒」とカメラの撮像面との角度が0度のときを基準とした相対値として示したものである．ここでは，画像処理解像度 160×120，「棒」の色の長さを 50 [mm]，カメラとして SONY CCD-MC10 1:2 f = 3.6 [mm] を用いた．

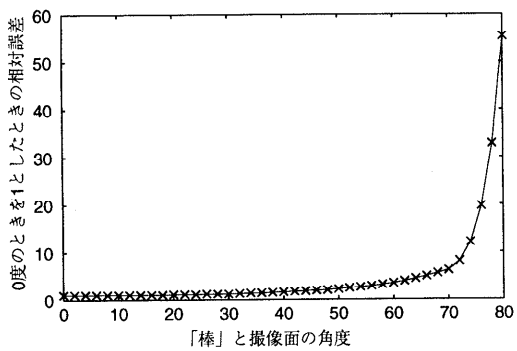


図3 カメラと「棒」のなす角と相対測定誤差

Fig. 3 The relative error ratio of the stick angle with the camera.

図3から，「棒」とカメラの撮像面とのなす角度が45度を超えると誤差が大きくなり，70度を超えると急激に大きくなることが分かる．このことからカメラは「棒」の側面がとらえられるように「棒」の方向とカメラ撮像面のなす角度がなるべく45度以下になるように設置するのが望ましい．また，このレンズは広角であり，画像処理解像度 160×120 とした場合，カメラと「棒」の距離は 20 [cm] から 60 [cm] が適切である．ジェスチャの認識自体はカメラ座標系で観測された値を基に描画平面を推定し，その平面へ写像されるので，これらの条件を満たす限り，任意の位置にカメラを設置してよい．

4. ジェスチャ認識

4.1 ジェスチャ言語の設計

たとえば，ユーザが文字列 "012" を意図して空間に描いた場合を考える．このとき**図4** のような描き方がありうる．このように一連の軌跡には直接文字を示す部分と文字以外の不要な部分とが含まれている．オンライン手書き認識では描画面上に描かれたストロークは文字の一部であるので不要な部分は含まれていない．観測された「棒」の軌跡とユーザの意図した言葉とを対応付けるにはこれらの間を関連付ける記述が必要であり，ここではジェスチャ言語の文法規則と呼ぶことにする．

文字 C，無効文字 N，区切り記号 S を考える．文字の開始と終了に区切り記号があり，文字と文字の間には無効文字が入る場合がある．ユーザが空間中に文字を描くとき，少なくとも1つの文字を描く間はある平面 U を仮定しその平面上に描いていると考えられる．観測された軌跡を描画平面 U へ写像したものが文字 C を表している．

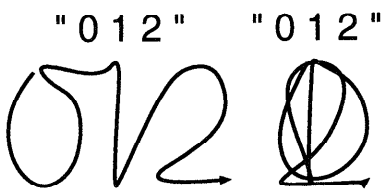


図4 文字列と軌跡の対応例

Fig. 4 Examples of the relation between a string and a trajictory.

連続して文字を描いた場合，文字ごとに平面を変えるのは不自然であり，一連の動作は同じ平面を想定していると仮定しても不具合はない．そこで，一連の軌跡から描画平面 U を規定し，区切記号と文字と無効文字を識別し，文字の部分をユーザが意図した言葉であると解釈する．

次に観測より得られた「棒」の軌跡とジェスチャ言語との対応付けを考える．ジェスチャの文字は描画平面に描かれた記号であるので，一定時間間隔で標本されている．しかし，このままでは速く描いたときと遅く描いたときとのいずれにも対応できる認識手法が必要になる．隠れマルコフモデル（HMM）や動的計画法（DP）を用いれば，ある程度の時間軸の変動は吸収しうる．本来ジェスチャの文字自身は記号の意味しか持たず，それが描かれる時間とは関係ない．よって，描画時間の変動は認識に対する単なる雑音にすぎない．その一方，区切記号は空間的変動が少ない部分であり，時間の経過が大きく意味を持っている．そこで，意味のある特徴を明示的に表現できるように以下の再標本化を行う．

「棒」の軌跡 $\gamma(t)$ を $\gamma(t)=(p(t),t)$ で表す．ここで，t は時刻，$p(t)$ は時刻 t での空間座標である．時刻 t における 3 次元空間中の標本点を $\hat{p}(t)=(x(t),y(t),z(t))$ とし，描画平面 uv への写像を $p(t)=(u(t),v(t))$ とする．つまり，軌跡 $\gamma(t)$ は描画平面上の点 p(t) の移動を表している．

ここで，平面内の単位方向ベクトルと時間方向の単位方向ベクトルをあわせて $e_i=(u_i,v_i,t_i)$ として表現する．このとき平面内の 8 方向を考えると，$e_0=(1,0,0)$，$e_1=(1/\sqrt{2},1/\sqrt{2},0)$，$e_2=(0,1,0),\cdots$，$e_7=(1/\sqrt{2},-1/\sqrt{2},0)$ と表現できる．時間方向には $e_8=(0,0,1)$ と $e_9=(0,0,-1)$ を用意する．このとき，空間方向の基準となる単位 δ と時間方向の基準となる単位 τ は，適当な重み w により $\tau=w\delta$ と対応付けておく．

これらを用いて，軌道上の 2 点 $\gamma(t_1)$ と $\gamma(t_2)$ の距離 D を次のように定義する．

$$D(\gamma(t_1),\gamma(t_2))=\max_{i=0}^{9}\{e_i\cdot[\gamma(t_2)-\gamma(t_1)]\} \quad (6)$$

ここで，"·" は内積である．

式 (6) の距離 D を基準として，平面に写像された軌跡 $\gamma(t)$ の再標本化を行う．具体的には，軌跡 $\gamma(t)$ の始点 $\gamma(t_0)$ から距離 D ごとに標本化操作を行い，ベクトル列 $\{\cdots,d_n(p_n,t_n),\cdots\}$ に変換する．ここで，d_n は動き方向ベクトルで距離計算時に最大となる方向ベクトルを表し，p_n，t_n は標本化した場所と時間

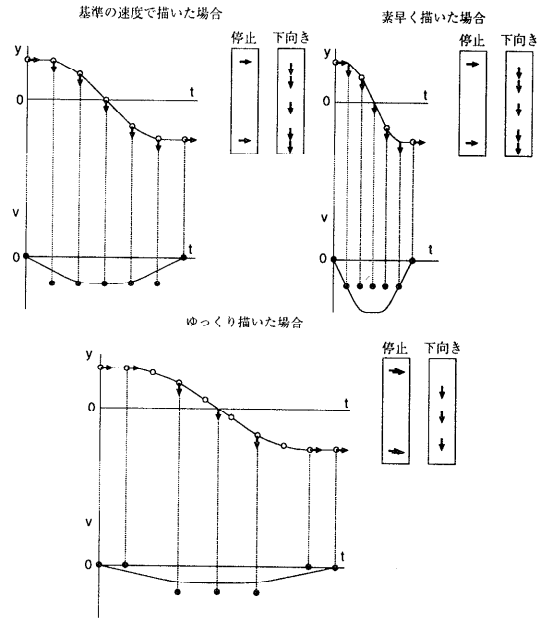


図 5 描画速度変化と抽出される特徴点の対応
Fig. 5 Extracted feature points with variable drawing speed.

を表す．この記述により，空間中で移動が少ない区間は時間方向の速度を反映した距離になり，動きが大きい区間はその動きの方向ベクトルを反映した距離として表現される．この再標本化により，動きが少なく時間のみが経過している部分と空間的な動きが大きい部分とを統一的に扱うことができる．

図 5 に速度を変えて "1" を描いた場合を示す☆．ここで，白丸は単位時間ごとに標本された点を示し，黒丸は上記の距離尺度で等間隔な再標本位置を示し，これらに対応する動きベクトルを矢印に示す．この例が示すように，学習時の描画速度と認識対象の描画速度がずれていても抽出される特徴量の位置や数がほぼ一定している．このためパターン整合法において高い認識率が期待できる．

ただし，実装にあたり観測データは基本的に等時間間隔で標本化したまま利用し，動きが大きい部分のみ線型補間した軌道を等距離で再標本化して利用する．また，時間方向のベクトルが連続する区間は，文字集合 C の区切りの候補，あるいはストロークの始点や終点などの特異点を表している．スポッティングを用いて尤度の高い区間を推定し，特異点を求めることも考えられるが，描画平面を決めずに描かれた文字の認

☆ 図 5 は再標本化の概念を示すための図であり，正確に計算した結果を示してはいない．

識を行うことは困難である．そこで，ユーザに文字の開始と終了時点で一呼吸停止してもらうことにして，システムを簡略化する．この場合，時間方向のベクトルが連続して表れる時点を文字の区切りとして扱い，単独の場合は角の部分などの特徴ベクトルとして扱う．

4.2 描画平面の推定と軌跡の射影

観測された「棒」軌跡の3次元座標の標本点列から描画平面の推定し，その平面上に射影する手法を示す．

描画平面を規定する法線ベクトルは，平面上に描画された任意の連続する3点から得られる外積のベクトルと平行である．そこで，「棒」の軌跡上の時刻 t での空間座標 $\hat{p}(t) = (x(t), y(t), z(t))$ から外積ベクトルを求める．$\vec{p}(t) = (\hat{p}(t+1) - \hat{p}(t))$ とすれば，$\vec{v}(t) = \vec{p}(t-1) \times \vec{p}(t)$ と計算できる．

連続する3つの外積ベクトルの中の2つのベクトル間のなす角の余弦値を計算し，その絶対値が小さい組合せを選ぶ．こうして選ばれた外積ベクトルをクラスタリングし，一番大きなクラスタの平均値が示すベクトルを描画平面の法線として採用する．最後に，この法線ベクトルから規定される描画平面に対して標本点列を射影する．

カメラから得られた軌跡を**図6**左に示し，求められた描画平面へ写像した軌跡を図6右に示す．このようにカメラ方向からとらえた軌跡（図6左）では何を描いているか分からないが，射影後の軌跡（図6右）からは円を描いていたことが判別できる．

4.3 認識処理

4.3.1 標準パターン

今回認識の対象とする動作は‘数字（0〜9）’と‘停止（動いていない部分）’である．基本的には標準パターンと認識対象データとの整合をとる手法を用いる．

認識に用いる標準パターンの概念を**図7**に示す．図7のように，1つの文字に対して8方向の動きベクトルと停止を示すベクトルの合わせて9つの各成分をメッシュ上に加算し，標準パターンを作成する．具体的な手順は以下のとおりである．まず，各々の学習用データの軌跡の外接矩形を求め，縦横比を保持して矩形中心とテンプレートの中心を合わせて正規化する．方向と位置の情報が得られた時点でその場所に値1を加えていき，動きが終了した時点で加えられた値の総和が1になるように正規化して特徴量を決定する．この操作を用意したすべての学習データに対して適応し，標準パターンを作成する．実際に‘数字（0〜9）’と‘停止’の11個のカテゴリを位置と方向の特徴で表現した標準パターンを**図8**に示す．

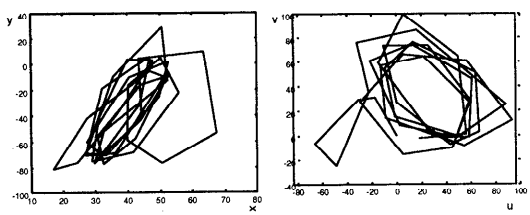


図6 観測された軌跡と平面に射影した軌跡
Fig. 6 The observed trajectory and the projected trajectory on the drawing plane.

4.3.2 類似度評価方法

認識は作成したモデルと与えられた軌跡データから

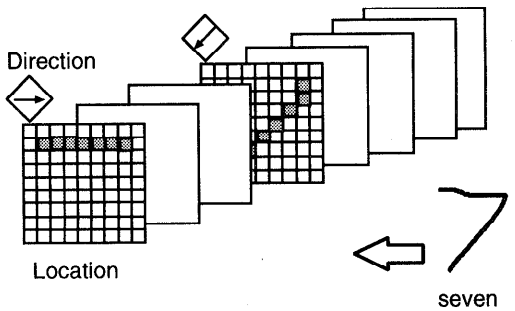


図7 標準パターンの記述
Fig. 7 The description of a template pattern.

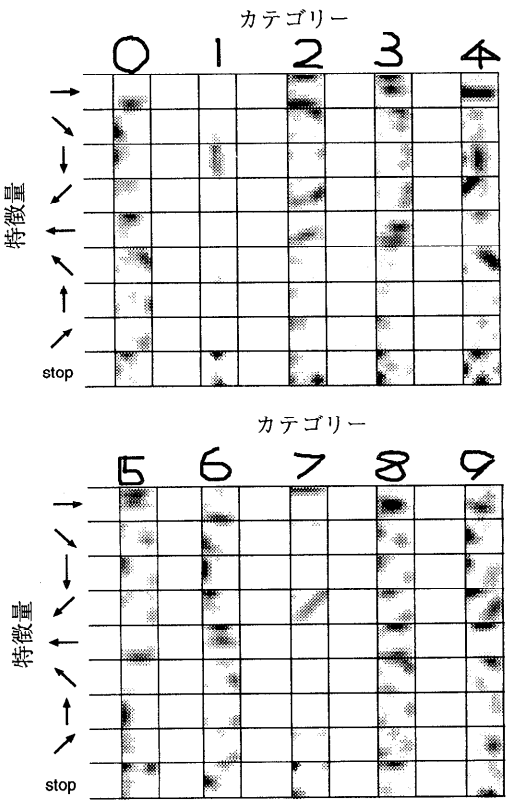


図8 数字の標準パターン
Fig. 8 The template patterns for digits.

同様の処理で変換したものとを比較し，その類似度が最大となったカテゴリを出力する．最も単純な類似度の基準は各カテゴリの標準パターンと対象となるデータのパターンの内積を用いる方法である．しかし，単純な内積では類似パターンの識別が困難な場合がある．そこで，標準パターンと外れているデータが観測された場合に類似度を下げる評価方法と複数の解像度の標準パターンとの類似度の重み付き和を総合的な類似度として用いる方法を導入する．

解像度 (r) のときの認識対象のパターン T とカテゴリ (s) の標準パターン $S_{(s)}$ との類似度 $R^{(r)}(T, S_{(s)})$ を式 (7) のように表す．

$$R^{(r)}(T, S_{(s)})) = \sum_{i,j} T_{ij}^{(r)} \otimes S_{(s)ij}^{(r)} \qquad (7)$$

ただし，$T_{ij}^{(r)}$，$S_{(s)ij}^{(r)}$ は解像度 (r) のとき T と S の ij 要素を表し，i，j はそれぞれ位置と方向を表す指標である．演算 $\otimes$ は式 (8) の処理を行うものとする．

$$T_{ij} \otimes S_{ij} = \begin{cases} T_{ij} \cdot S_{ij} & (S_{ij} \neq 0) \\ -T_{ij} & (S_{ij} = 0) \end{cases} \qquad (8)$$

この演算の効果を**図 9** に示す．類似度は，標準パターンの黒画素部分（出現確率が零でない部分）と認識対象のパターンの黒画素が重なる部分（図 9 の A）については，それぞれの値の積が加算され，認識対象のパターンの黒画素と標準パターンの白画素部分（出現確率が零の部分）が重なる部分（図 9 の B）については負の値がペナルティとして加算される．

そして，総合類似度 $R(T, S_{(s)})$ を式 (9) のように重み付きの和として計算する．

$$R(T, S_{(s)}) = \sum_{r} w_r \times R^{(r)} \qquad (9)$$

ただし，w_r は解像度 (r) の類似度に関する重み係数である．今回の実験では，2×2，3×3，9×9 の 3 つの解像度の類似度を統合して用いた．

4.3.3 認識実験

本手法の有効性を確認するために手と口により「棒」を用いて描いた数字ジェスチャに対する認識実験の結果を示す．いずれの場合も，「棒」により空間に描かれるジェスチャは，紙に描く場合と比べパターンの変動が大きい．この理由としては，手掛かりとなる平面が存在しないことと，描いた軌跡が見えないためである．そこで，被験者と描画方法ごとに個別の標準パターンを作成し，特定の被験者のみに適応した場合を対象として認識実験を行う．

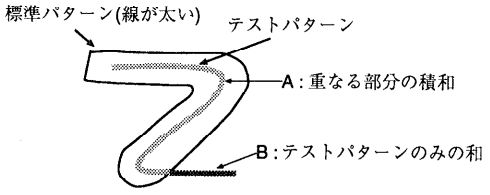


図 9 類似度の算出法
Fig. 9 The calcuration of similarities.

実験の手順は以下のとおりである．まず，学習用に各カテゴリの動作をカメラから入力し，「棒」の 3 次元軌跡データを解析し，ファイルに保存する．次に，このデータから学習に用いる標準パターンを作成する．認識時には，学習時と同様に「棒」の 3 次元データをファイルに保存し，これと標準パターンとの類似度を計算し，最も類似度の高いカテゴリを認識結果として認識率を求める．

計算機は Pentium 150 MHz のパーソナルコンピュータに FreeBSD を載せたものを利用した．他に使用した機材等は，「棒」の色の長さを各 50 [mm]，カメラは SONY CCD-MC10 1:2 f = 3.6 [mm]，画像処理解像度は 160 × 120 で 10[フレーム/秒] で標本化した．

4.3.3.1 手で描いた場合の認識

まず，手で数字を描いて各カテゴリ 20 個（合計 200 個）のデータから標準パターン作成した．次に，同一被験者が評価用に各カテゴリ 50 個合計 500 個のデータを用意した．認識実験の結果，学習に使用したパターンの認識率は正解 199 個/200 個中（99.5%）であり，同一被験者の評価用のデータは正解 488 個/500 個中（97.8%）の認識結果が得られた．標本間隔の違いによる認識率について調べるため，評価用データを 5[フレーム/秒] に間引きして認識を試みたところ，正解 451 個/500 個中（90.2%）の認識結果が得られた．誤認識は “0” と “6” や “1” と “7” など類似文字に表れている．

この結果より，本ジェスチャ認識手法がヒューマンインタフェースに利用できる性能を有していることが確認された☆．

4.3.3.2 口で描いた場合の認識

同様に口で描いた場合について，学習用に各カテゴリ 20 個（合計 200 個）のデータを用いて標準パターンを作成し，評価用に用意した各カテゴリ 20 個（合計 200 個）のデータの認識率は正解 165 個/200 個中（82.5%）であった．

☆ 軌跡の分割が 100% 正確に行われる仮定のもとで，他の特徴量（メッシュ特徴）を用いた認識率は 90% に満たなかった．

手で描いた場合と比べ認識率が低い理由は，カメラの撮影の関係で口で描く場合の方がカメラ座標系の奥行き方向移動が大きな意味を持っている．奥行き方向は3次元計測の測定誤差の影響を受けやすいので，3次元計測や描画平面推定が難しいためと考えられる．実用に向けた安定した認識結果を得るには，さらに測定精度の向上が望まれるが，用途を限れば実用化が可能なレベルであると考える．

5. ジェスチャ認識を用いたインタフェースの試作

「棒」を用いたインタフェースのアプリケーションとして，コマンド入力を行うシステムを紹介する．

5.1 動作とコマンドの対応表の作成

「棒」を用いてコマンドを入力するために，以下に示すような対応表を準備する．

- カテゴリ名 ⇔ カテゴリ番号対応表
  認識システムは，認識結果をこの表で与えられたカテゴリ番号で返す．
- コマンド ⇔ コード列（カテゴリ名）対応表
  ユーザが定義したいコマンドと，それをどの動作の組合せで対応付けるかを記述する．

ユーザはこの対応表を自由に書き換えることで，様々なコマンドを入力することが可能になる（図**10**）．

「棒」の動作を認識するプログラムは，現在の処理状態や認識結果を出力している．これにより利用者は，動作が有効に処理されていることを確認することができる．しかし，これら認識に関する情報とコマンド実行結果の情報を同じ場所に出力してしまうと混乱が起こる．このため，実行結果は別の場所で表示しなければならない．そこで今回はジェスチャ入力を既存のアプリケーションをクライアントとして起動する入力サーバとして実装した．これにより，多くのコマンド入力型のアプリケーションと容易に接続できる．

5.2 電子メールに対する応用例

具体的な実行例として，「棒」で電子メールを操作するアプリケーションとのインタフェースに応用する．メールを送信するためには，文章を入力する必要があり，音声入力やタッチパネルなど他の手段を組み合わせる必要がある．ここでは，「棒」による操作性を確認することが目的であるので，文章を書く以外の表示や削除などを実行の対象とする．ジェスチャと命令の対応表には独立した複数のコマンドで操作するタイプのMail User Agent（MUA）である Massage Handler（MH）のコマンド群（prev, next, show, inc, ..., etc.）と入力サーバの “終了” そして “誤認識された場合の訂正” を記述する．実行の様子を図**11** に示す．同図左上がシステムへの入力画像とその抽出結果を示し，左下が認識の様子を表示している．右のウィンドウは「棒」の操作により表示されたメールを表している．このように，メールボックス内のメールを次々に眺めたり，不要なメールを削除する操作が利用できることが確認できた．

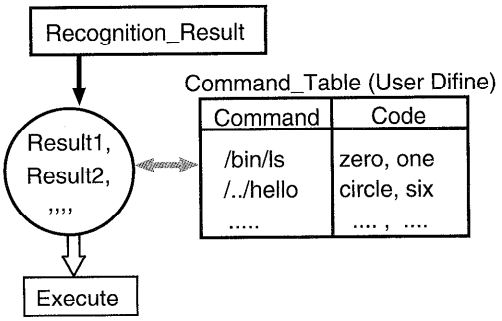


図**10** 対応表を用いたコマンド入力システム
Fig. 10 A command imput system using a translation table.

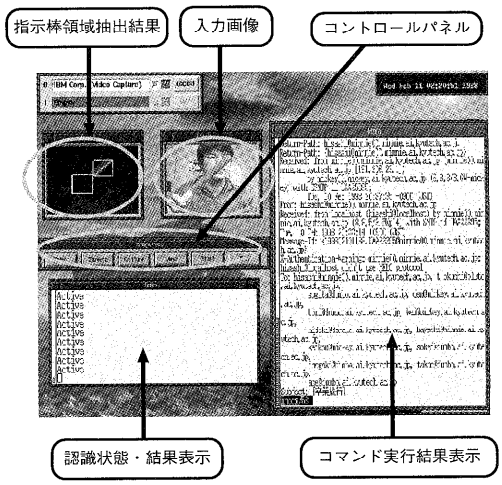


図**11** ジェスチャ入力を用いたメールリーダとインタフェース
Fig. 11 A mail user agent interface using the gesture recognition.

6. ま と め

本論文では，「棒」を用いた非接触で使いやすいインタフェース環境実現のために，ジェスチャ認識システムの改良を行った．

まず，ジェスチャとそれを描く平面のモデル化と「停止」をジェスチャの前後に設けるという簡単な言語の文法的制約を付加することにより，連続する時系列データからジェスチャ部分を安定して切り出せるよう

になった．また，時間方向の特徴である「停止」と空間方向の動きを統一的に扱う表現を導入し，ジェスチャを描く速度変化を吸収する再標本化の手法を示した．

実際の認識実験では，パターン整合法を基本とした認識手法を用いた．標準パターンとのずれをペナルティとして評価する類似度の導入と，複数解像度の類似度を総合することにより，同一被験者の場合に認識率90%以上という十分実用となる認識率が得られた．口で描いたデータに対しても80%以上の認識率が得られ，用途を限れば実用化が可能なレベルであると考えられる．これにより，「棒」入力は手の不自由なユーザの入力手段として有望であることが確認できた．

具体的な応用例として，コマンド入力システムを作成し，既存のアプリケーションのコマンド入力の利用するシステムを試作し，実用化への可能性を示した．

今後は，提案手法の特徴抽出や類似度計算と他の手法とを比較検討し，よりシステム全体の高度化を図る予定である．また，実用化に向けて，照明環境の変化により対応する画像処理技術の導入や他の3次元計測手法との組合せを検討している．

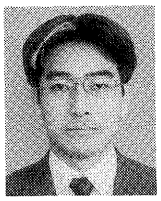
謝辞 有意義なコメント本研究および論文作成にあたり，日頃から貴重なご意見をいただきました，江島・吉田研究室の諸氏に感謝いたします．

参考文献

1) 大橋 健，山之内毅，松永 敦，江島俊朗：指示棒と音声が使えるコミュニケーション環境CoSMoSの提案，インタラクティブシステムとソフトウェアII，竹内彰一（編），pp.29–36，日本ソフトウェア科学会，近代科学社 (1994).
2) Ohashi, T., et al.: Multimodal interface with Speech and Motion of Stick: CoSMoS, *Symbiosis of Human and Artifact*, Anzai, Y., et al. (Eds.), pp.207–212, Elsevier Science Publishers B.V. (1995).
3) 大橋 健，吉田隆一，江島俊朗：指示棒によるジェスチャの認識手法，画像の認識理解シンポジウム MIRU'96 講演論文集分冊 II，pp.49–54 (1996).
4) 大橋 健，吉田隆一，江島俊朗：指示棒を用いた仮想オブジェクトの変形操作，信学技報，Vol.PRMU96, No.95, pp.17–22 (1996).
5) 大橋 健，吉田隆一，江島俊朗：HMMを用いた指示棒の描くジェスチャーの認識，電気関係学会九州支部連合大会論文集，p.49 (1997).
6) 仲程 啓，大橋 健，吉田隆一，江島俊朗：方向特徴を用いた指示棒が描くジェスチャーの認識，信学技報，Vol.PRMU97, No.273, pp.57–64 (1998).
7) 高橋友一，岸野文郎：手振り認識方法とその応用，電子情報通信学会論文誌（D-II），Vol.J73-D-II, No.12, pp.1985–1992 (1990).
8) Seki, Y., et al.: Improvement of Hand-Grasp Measurement System, *Design of Computing Systems: Social and Ergonomic Considerations*, Smith, M.J., et al. (Eds.), pp.459–462, Elsevier Science Publishers B.V. (1997).
9) 澤田秀之，橋本周司，松島俊明：運動特徴と形状特徴に基づいたジェスチャー認識と手話認識への応用，情報処理学会論文誌，Vol.39, No.5, pp.1325–1333 (1998).
10) Pavlovic, V.I., et al.: Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.677–695 (1997).
11) 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.76-D-II, No.8, pp.1813–1821 (1993).
12) 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，電子情報通信学会論文誌（D-II），Vol.J79-D-II, No.7, pp.1218–1229 (1996).
13) 渡辺孝弘，李 七雨，谷内田正彦：インタラクティブシステム構築のための動画像からの実時間ジェスチャ認識手法—仮想指揮システムへの応用，電子情報通信学会論文誌（D-II），Vol.J80-D-II, No.6, pp.1571–1580 (1997).
14) Pavlović, V.I., et al.: Gestural Interface to a Visual Computing Envronment for Melecular Biologists, *Proc. 2nd International Conference on Automatic Face and Gesture Recognition*, pp.30–35 (1996).
15) Bobick, A.F. and Wilson, A.D.: A State-Based Approach to the Representation and Recognition of Gesture, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.12, pp.1325–1337 (1997).
16) Matsuo, H., et al.: The Recognition Algorithm with Non-contact for Japanese Sign Language Using Morphological Analysis, *Gesture and Sign Language in Human-Computer Interaction*, pp.273–284 (1997).
17) クンラポンユーニパン，木下宏揚，酒井善則：スティックモデルを用いた手振りの認識，電子情報通信学会論文誌（D-II），Vol.77-D-II, No.1, pp.51–60 (1994).
18) Shimada, N., et al.: Hand Gesture Recognition Using Computer Vision Based on Model-matching Method, *Symbiosis of Human and Artifact*, Anzai, Y., et al. (Eds.), pp.11–16, Elsevier Science Publishers B.V. (1995).

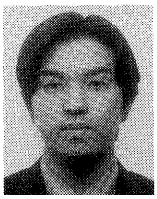
19) 渡辺 賢，岩井儀雄，八木康史，谷内田正彦：カラーブローブを用いた指文字の認識，電子情報通信学会論文誌（D-II），Vol.J80-D-II, No.10, pp.2713–2722 (1997).
20) 中嶋正之，柴 広有：仮想現実世界構築のための指の動きの検出法，電子情報通信学会論文誌（D-II），Vol.J77-D-II, No.8, pp.1562–1570 (1994).
21) Wren, C., et al.: Pfinder: Real-Time Tracking of the Human Body, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.780–785 (1997).
22) 高橋勝彦，関 進，小島 浩，岡 隆一：ジェスチャ動画像のスポッティング認識，電子情報通信学会論文誌（D-II），Vol.J77-D-II, No.8, pp.1552–1561 (1994).
23) 西村拓一，野崎俊輔，向井理朗，岡 隆一：連続DPへの非単調性導入によるジェスチャ動画像からの戸惑い動作のスポッティング認識，電子情報通信学会論文誌（D-II），Vol.J81-D-II, No.1, pp.18–26 (1998).
24) 小高和巳，荒川弘熙，増田 功：ストロークの点近似による手書文字のオンライン認識，電子情報通信学会論文誌（D），Vol.J63-D, No.2, pp.153–160 (1980).
25) 小高和巳，若原 徹，増田 功：筆順に依存しないオンライン手書き文字認識アルゴリズム，電子情報通信学会論文誌（D），Vol.J65-D, No.6, pp.679–686 (1982).
26) 金 長吉，川嶋稔夫，青木由直：部分的構造関係の解析に基づくオンライン手書き漢字認識，電子情報通信学会論文誌（D-II），Vol.J74-D-II, No.12, pp.1706–1714 (1991).
27) 秋山勝彦，中川正樹：オンライン手書き日本語文字認識のための線型処理時間収縮マッチングアルゴリズム，電子情報通信学会論文誌（D-II），Vol.J81-D-II, No.4, pp.651–659 (1998).
28) Shein, F., et al.: Access Considerations Of Human-Computer Interfaces For People With Physical Disabilities, *Symbiosis of Human and Artifact*, Anzai, Y., et al. (Eds.), pp.143–148, Elsevier Science Publishers B.V. (1995).
29) Kiyota, K., et al.: Pen-based Japanese Character Entry System for Visually Disabled Persons, *Design of Computing Systems: Social and Ergonomic Considerations*, Smith, M.J., et al. (Eds.), pp.447–450, Elsevier Science Publishers B.V. (1997).
30) 伊藤英一：身体障害者を支援するコンピュータテクノロジー，*bit*, Vol.25, No.9, pp.14–21 (1993).
31) 伊藤英一，大橋正洋：視線移動を考慮した頚椎損傷者用ペン型ポインティングデバイス，情報処理学会論文誌，Vol.39, No.5, pp.1440–1447 (1998).

(平成10年 6 月1日受付)
(平成10年12月 7 日採録)



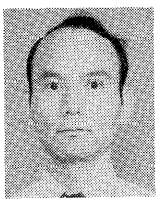
大橋 健（正会員）

1989年長岡技術科学大学工学部電気電子システム工学課程卒業．1991年同大学大学院修士課程修了．同年九州工業大学情報工学部知能情報工学科助手，現在に至る．マルチモーダルインタフェース，画像や音声の認識理解に興味を持つ．電子情報通信学会，日本ソフトウェア科学会，IEEE各会員．



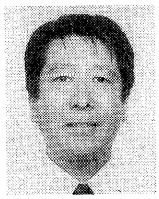
仲程 啓

1998年九州工業大学情報工学部知能情報工学科卒業．同年大阪大学大学院基礎工学研究科情報数理系専攻博士前期課程入学，現在に至る．マルチモーダルインタフェース，手術支援システムの研究に従事．



吉田 隆一（正会員）

1982年慶應義塾大学工学部電気工学科卒業．1987年同大学大学院工学研究科博士後期課程電気工学専攻修了．工学博士．同年九州工業大学情報工学部知能情報工学科助手．1990年同助教授．1993年から翌年にかけてオレゴン科学技術大学院大学において客員研究員．オブジェクト指向計算，分散計算，分散処理システム，オブジェクト指向データベースに興味を持つ．日本ソフトウェア科学会，人口知能学会，IEEE，ACM各会員．



江島 俊朗（正会員）

1973年東北大学工学部通信工学科卒業．1978年同大学大学院博士課程修了．同年東北大学工学部通信工学科助手．1985年同大学情報助教授．同年長岡技術科学大学電気系助教授．1990年九州工業大学情報工学部知能情報工学科教授．1995年から96年までカリフォルニア大学デイビス校客員教授．文字・図形，音声の認識およびヒューマンインタフェースに興味を持つ．工学博士．電子情報通信学会，人工知能学会，IEEE各会員．