

## 動きの変化に着目した動作パターンの 1次元符号化に基づく身振り認識

小 荒 健 吾<sup>†1</sup> 西 川 敦<sup>†2</sup> 進 戸 健 太 郎<sup>†3</sup>  
石 井 薫<sup>†4</sup> 安 井 陽 一<sup>†5</sup> 宮 崎 文 夫<sup>†2</sup>

身振りは一般に、手話の指文字のようにその形状で表現する「静的」なものと、手まねき動作のように動作パターンで表現する「動的」なものに大別できる。本論文では、画像を用いた動的な身振り認識を対象とした動作パターンの記述法と認識法を提案する。まず、画像中の動領域の重心を求めることで、手の動きを抽出する。次に、身振り動作の特徴は動きが変化する部位に強く現れると考え、動領域の重心位置、速度が極大/極小となる点を特徴としてとらえ、符号を割り当てる。この符号を、事象が生じた時刻順に、同時刻に複数の事象が生起している場合には、割り当てた符号の値が小さいものから順に並べることで1次元の符号列を生成する。さらに、文字列比較のアルゴリズムを応用したパターン識別法を提案する。また、認識結果の信憑性の評価法、および識別可能なパターンの組合せの評価法についても検討する。

## Gesture Recognition Based on 1-Dimensional Encoding of Hand Motion Changes

KENGO KOARA,<sup>†1</sup> ATSUSHI NISHIKAWA,<sup>†2</sup> KENTARO SHINTO,<sup>†3</sup>  
KAORU ISHII,<sup>†4</sup> YOICHI YASUI<sup>†5</sup> and FUMIO MIYAZAKI<sup>†2</sup>

We propose a new method for describing and recognizing hand gestures. At first, we calculate the position and velocity of the center of gravity of moving region in video image using an interframe block subtraction method. We then consider that the maximum/minimum points of such motion information represent motion pattern well, we encode them into a 1-dimensional code sequence by a very simple rule. The encoded pattern is recognized by comparing with template patterns using a method based on string comparison algorithms. We also discuss a method for evaluating the reliability of the recognized results and define a criterion for evaluation of pattern similarity.

### 1. はじめに

人がコミュニケーションを行うとき、自然に身振りを交えて意志を表現することから示唆されるように身振り動作は人にとって直観的なものであり、こうした動作を認識する手法の開発は、人にやさしいヒューマ

ンインタフェースを構築するうえで有用である。とくに、ユーザの動きに制限を与えない非接触の身振り認識手法の開発が望まれる。非接触の手法として、カメラで撮影した動画像から得られる情報の時系列を解析し動作(身振り)認識を実現しようとする研究がさかんに行われている<sup>1)~10)</sup>。

牛田ら<sup>1)</sup>は、手の座標値やある点を基準とした角度などの量が山や谷となる点と定常的な点を特徴量としてとらえ、ファジー連想記憶を用いて身振りの認識を行っている。林ら<sup>2)</sup>はステレオ画像から求まる人の顔の上端点、首の下端点、左肩の点の3つの点を基準としてユーザ位置に拘束されることなく手の3次元的な位置、方向を認識するシステムを開発している。この手法では手の抽出は肌色領域を求めることで行われている。しかしながら、画像処理の手法上、背景環境が整備されている必要があると考えられる。大和ら<sup>3)</sup>は、

†1 三菱電機株式会社名古屋製作所

Nagoya Works, Mitsubishi Electric Corporation

†2 大阪大学大学院基礎工学研究科システム人間系専攻

Department of Systems and Human Science, Graduate School of Engineering Science, Osaka University

†3 日本 IBM 株式会社

IBM Japan Ltd.

†4 株式会社本田技術研究所栃木研究所

Tochigi R & D Center, Honda R & D Co. Ltd.

†5 大阪大学大学院工学研究科知能・機能創成工学専攻

Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University

背景差分により得られた身体領域を含む画像を格子領域に区切り、ベクトル量子化することでコードブック (code book) を作成して動画像をシンボル列に変換、さらに HMM を適用してパターンの認識を行っている。渡辺ら<sup>4)</sup>は、Maskable Template Model により身体の特定位を動画像から抽出し、切り出された身体部位に KL 展開 (Karhunen-Loève Expansion) を適用することで低次元のジェスチャ空間を構成し、その空間上で表現されたジェスチャ曲線から特徴的な点を抽出した。また、複数のカメラからの入力画像を1枚に統合し、ひとまとめに処理を行う手法を提案している。しかしながら、これらの手法は身体部位を抜き出すのに2値化処理を行っているため、背景環境や照明条件に影響を受けやすいと考えられる。

これらのような比較的環境の影響を受けやすい画像特徴量を抽出する手法に対し、動きの場を入力として利用する手法も提案されている。Davisら<sup>5)</sup>は、画素が時間の関数となる2次元画像モデル MHI (Motion History Image) を提唱し、人間の動作認識に適用した。このモデルは、フレーム間差分の2値化により得られる動領域を時間履歴を考慮して重ね合わせることで生成される。桐島ら<sup>6)</sup>はフレーム間差分情報の2値化処理により検出された運動領域の重心まわりの領域分布を形状特徴量とし、これに身振りの名前などの情報を加えたラベルづけを行い動作パターンを記述している。さらに、差分画像を足し合わせた軌跡画像を用いることでその動作が別の身振りの一部か独立した身振りかを判定する手法を提案している。しかし、これらの手法では画像全体に対するフレーム間差分を利用しているため、動作の種類や状況によってはノイズの影響を受けやすいと考えられる。西村ら<sup>7)</sup>も動領域の抽出にフレーム間差分を用いているが、入力画像を低解像度化し各点に関して3フレームにわたる平均をとることでノイズの影響を低減している。低解像度化された画像はそのまま特徴ベクトルとして連続 DP に引き渡され身振り動作の検出 (spotting) と認識処理が行われる。この手法で用いている特徴ベクトルは位置依存のため、カメラに対する動作者の位置が異なる場合にはうまく機能しないと考えられる。

画像中の動領域をとらえる方法としてオプティカルフローを利用した研究もいくつか提案されている。畠ら<sup>8)</sup>はオプティカルフローにより動きの場を検出し、得られた動きの場に KL 展開を適用することで低次元のジェスチャ空間に動き情報を投影し、さらにモデルデータをクラスタリングすることで1次元の符号列を生成している。身振り認識は生成された符号列に離散

HMM を適用することで行われる。画像やフロー場などの2次元分布に KL 展開を用いる場合、対象の平行移動に弱い欠点があるが、まず顔を検出し、それに合わせて処理領域を移動させることで動作者の平行移動にロバストな手法を構築している。西川ら<sup>9)</sup>は、オプティカルフローから計算される手の移動方向の変化率 (曲率) を基に、あらかじめ用意された基本動作の列として符号化を行い、動作の個人差の影響を受けにくい手振りの認識を実現している。また、移動ロボットに身振り認識用のカメラを搭載するなど動作者とカメラとの距離に変化が生じる場合を考慮したパラメータ調整法を検討している。しかしながら急激な方向転換の際にサンプリング時間やノイズの問題などから曲率が適切に得られないこともある。この問題に対処するために符号化にあるルールを設けて処理を行っている。以上の2つの方法では、オプティカルフローを実時間計算するために専用のハードウェアを利用していたのに対し、Cutlerら<sup>10)</sup>は、フレーム間差分とオプティカルフローを併用することにより、一般的な PC のみで両手の身振りを実時間認識できるシステムを開発した。彼らのシステムでは、フレーム間差分値が大きな画素についてのみオプティカルフローを計算することでトータルの計算コストの低減を図っている。得られたフロー場をフローの方向の類似性に基づいてたかだか2個のクラスタに分割することで両手の検出を行い、得られたクラスタの大きさや方向情報を用いたルールベース的な手法により身振り認識を達成している。これらの研究をはじめ、一般に、身振り認識の問題は、(1) 認識対象となる手先などの部位を抽出し、(2) その運動から得られる時系列情報を符号化し、(3) 符号化された情報を識別する、という3つの問題に帰着できる<sup>11)</sup>。

本研究では、ヒューマンマシンインタフェース (たとえば、コンピュータ画面の操作や移動ロボットの誘導など) に用いるコマンドとしてあらかじめデザインされた「手全体の動きで表現する」身振り動作を認識対象とし、(1) 照明などの環境条件の影響を受けやすい状況下でもロバストに動作する、(2) 専用の特殊なハードウェアを必要としない (計算量の少ない) 画像ベースの身振り認識手法を開発することを目的とする。この目的を達成するために、本論文で提案する手法では、まず、画像からの対象領域抽出にはフレーム間差分を用い、運動する物体を抜き出す手法をとる。ただし、画素ごとの差分ではなく、画像を格子状の小領域 (ブロック) に区切り各小領域ごとのフレーム間差分を行うことで、ノイズによる影響の低減を図る。

次に、得られた動きの場の重心の位置ならびに速度を計算し、それらの時系列における極値を特徴量としてとらえる。またこれらの特徴量の時系列を単純なルールに従って1次元のパターンに符号化する。この符号化手法により動作のパターン表現を大幅に圧縮することができる。さらに、符号化された1次元パターンと参照情報(テンプレート)とのマッチングを行うことで身振り動作の種類を識別する手法を提案する。この方法は文字列比較アルゴリズム<sup>12)</sup>をベースにしているという点に特徴があり、これにより、動作パターンを表す符号列を明瞭な指標で比較することができる。

本研究と同様の目的、範囲を持つ研究として、上述したCutlerらの研究があるが、我々の手法はフローの演算やフロー場の解析をいっさい必要とせず、単に動きの場の重心を求めているという点で、Cutlerらの方法より大幅に計算量が少ない。指の細かな動きや形状変化ではなく手振りを行う「手全体」(あるいは体全体)の動きをとらえるには、このアプローチで十分と考えている。すなわち、従来の身振り動作記述、認識手法に対する提案手法の最大の特長は、「動きの場の重心の位置・速度変化」という計算コストの小さい大雑把な情報のみを用い、さらにその情報を単純な1次元の符号列として表現することにある。加えて、本研究では、この1次元符号列という単純なデータ構造を利用して、認識結果の信憑性の評価法および識別可能なパターンの組合せの評価法についても明示的に検討を行っている。これらは、身振り認識手法をヒューマンマシンインタフェースへ実際に応用する際には考慮しなければならない重要なポイントであるにもかかわらず、従来の研究においてほとんど言及されなかった点である。

本論文の構成は次のとおりである。まず2章において、小領域フレーム間差分を利用した、運動領域の抽出法および身振り動作の切り出し手法を概説する。3章では、抽出された動作のパターンを1次元的な情報に符号化する手法について説明する。4章においては、提案手法により得られる符号列を利用し、入力パターンを識別する手法を提案する。5章では、ここまで議論してきた符号化・認識手法を、実際に身振り認識に適用し、その検証を行う。6章では、提案手法の応用例として身振りによる移動ロボットの誘導を取りあげる。最後に7章でまとめと課題を述べる。

## 2. 身振り動作の検出

### 2.1 前提条件

本研究では、文献9)と同様、環境および認識対象

に対して次の仮定をおく。

- (1) 認識対象は片手全体で行う動きの大きな身振り動作とし、手指の形状の変化などの細かな動きは認識対象としない。
- (2) 一連の身振り動作はカメラ視野内で行われる。
- (3) カメラ視野内には他の動物体は存在しない。さらに、後述する「身振り動作の切り出し」問題を簡単化するために、次のような前提条件も設ける。
- (4) ある身振り動作と別の身振り動作の間にはいったん静止する(少なくとも静止することを意識する)。

第4の仮定は複数の身振り動作が連続的に行われることを対象外にしているが、身振りによる指示を受けて反応を示すアプリケーションとの対話的なインタフェースを考えた場合には、ユーザはアプリケーション側の応答を見てから次のコマンド(身振り)を行うという形態が考えられるため、妥当な仮定であると考ええる。

### 2.2 画像からの動き検出

カメラによる非接触な身振り認識を考えたとき、初めに必要なのは画像からの身振り動作の検出である。本研究では、上述した前提条件をふまえ、画像中での運動領域(動きの場)に着目する。動きの場を求める手法として最も単純なものは、連続する画像フレームどうしの差分処理(フレーム間差分)である。しかしながら単純にフレーム間差分をとるだけでは、画素ごとの細かなノイズの影響を受けやすい。そこで、図1に示すように、まず画像を複数の小領域に切り分ける。そしてそれぞれの小領域ごとに、連続するフレーム間の差分値を求めその和が一定値よりも大きい領域を「動きあり」と判定する。ここではとくに、この動きの場の重心位置とその速度を動き情報として利用する。

### 2.3 身振り動作の切り出し

本研究では2.1節で述べたとおり、1つ1つの身振りの間で動作がいったん停止することを仮定している。しかしながら、実際には、動作者自身は静止したつもりでも停止後の反動や無意識的な揺れなどにより、画像から動き検出を行う低レベルな処理にとっては「動きがある」と見なされることがある。また、意図的な

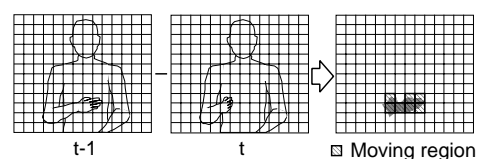


図1 小領域フレーム間差分

Fig. 1 Interframe block subtraction.

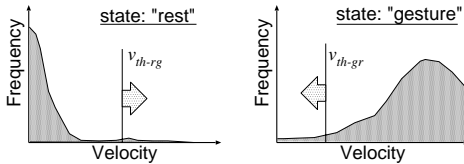


図2 「待機」「身振り」各状態時の動領域重心速度分布（模式図）  
 Fig.2 The distribution of velocity of the centroid of moving region in the “rest”/“gesture” states.

身振り動作の途中で動きが反転する場所などにおいては速度が一時的にゼロに近くなることもある。こうした場合、単に「動きの有無」を判定するだけでは適切に動画像列から意図された身振り動作を切り出すことはできない。

以上をふまえて、本論文では、次の手法で意図的な身振り動作を切り出す。まず、「待機」と「身振り」という2つの状態を考える。図2に、待機状態、身振り状態における動領域重心の速度分布の模式図を示す。この図のように、一般に、待機状態においては、動領域重心の移動速度は低いところに集中し、高い速度が生じる頻度はきわめて低く、逆に、身振り状態では、比較的速度の大きい部分に分布が多く現れ、低い部分の分布は少なくなると考えられる。ここで、待機状態のとき、ある閾値  $v_{th-rg}$  を超える速度ならば身振り状態への遷移を考慮し、身振り状態のとき閾値  $v_{th-gr}$  よりも低い速度が現れたならば待機状態への遷移を考慮する。身振り動作が開始されるまでは閾値は大きめに、いったん身振り動作が開始されれば折り返し動作などで身振りの途切れ以外に速度が低下すると考えられるため、閾値を低めに設定する（すなわち  $v_{th-rg} > v_{th-gr}$ ）。しかしながら、速度の分布は待機状態と身振り状態とで重なり合う部分があるため、速度閾値からの状態遷移の条件を満たした状態が一定時間（待機 → 身振り： $n_{th-rg}$ 、身振り → 待機： $n_{th-gr}$ ）継続した場合に、完全に状態が遷移したものと見なし、継続時間が不十分な場合には状態遷移はなかったものと判定する。この処理の流れをまとめたフローチャートを図3に示す。なお、「身振り動作」として切り出す情報は、状態遷移中のフレームを考慮し、待機状態から身振り状態に遷移する際には、遷移と判定された時点からの  $n_{th-rg}$  フレーム分過去に遡って情報を抜き出し、身振り状態から待機状態に遷移する際には、 $n_{th-gr}$  フレーム分の情報を無視するものとする。

3. 動作パターンの1次元符号化

3.1 身振り動作パターンの特徴の抽出

本研究では、身振り動作パターンの特徴は、動きが

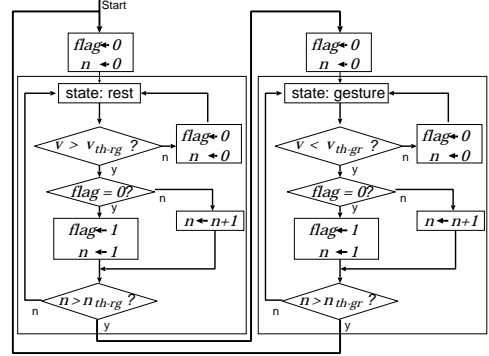


図3 「待機」「身振り」状態遷移のフローチャート  
 Fig.3 Flowchart of the transition between the states “rest” and “gesture.”

変化する箇所（タイミング）に強く現れるという考えに基づき、前節までの方法によって入力画像列より切り出された「動領域重心の位置  $(p_x, p_y)$  および速度  $(v_x, v_y)$  の時系列」が各軸においてそれぞれ極大/極小となる点をパターンの“特徴”としてとらえる。

極大/極小点は次のようにして求める（ここでは、便宜上、 $p_x, p_y, v_x, v_y$  を代表して  $Val$  という記号で表現する）。まず、時系列から細かな変動の影響を取り除くため適当なパラメータのガウシアンフィルタを施す。次に、サンプリング時のフレーム番号を  $t_s$ 、その時刻における入力値を  $Val(t_s)$ 、時間的に近傍と見なすフレーム数を  $N_n$  とするとき、次の2つの条件  $\{Val(t_s) - Val(t_s - 1)\}\{Val(t_s + 1) - Val(t_s)\} < 0$ 、ならびに、 $\{Val(t_s) - Val(t_s - N_n)\}\{Val(t_s + N_n) - Val(t_s)\} < 0$  を同時に満たすならば、 $Val(t_s) - Val(t_s - 1) > 0$  のとき「極大点」、 $Val(t_s) - Val(t_s - 1) < 0$  のとき「極小点」と見なす。これは、隣接フレームのみではノイズによる局所的な極大/極小点を検出しやすくなる一方、近傍フレームだけでは条件を満たす部位が、ある時点ではなく区間として現れる可能性があるからである。

3.2 1次元事象符号列の生成

以下では、前節の方法によって抽出された動作パターンの“特徴”を「事象（event）」と表現するものとする。今、入力  $d$  における特徴量  $f$  の値が  $p$  となる時点を考えてとき、事象が発生するのは、 $d = \{\text{カメラ1, カメラ2, カメラ3, ...}\}$ 、 $f = \{\text{動領域重心の位置 } x \text{ 座標, 同 } y \text{ 座標, 動領域重心の速度 } x \text{ 座標, 同 } y \text{ 座標}\}$ 、 $p = \{\text{極大, 極小}\}$  の場合であり、身振り動作をとらえるカメラの台数を  $N$  台としたとき、事象の種類としては全部で  $N \times 4 \times 2 = 8N$  通りの可能性がある。これらの事象を表現するためにそれぞれ1

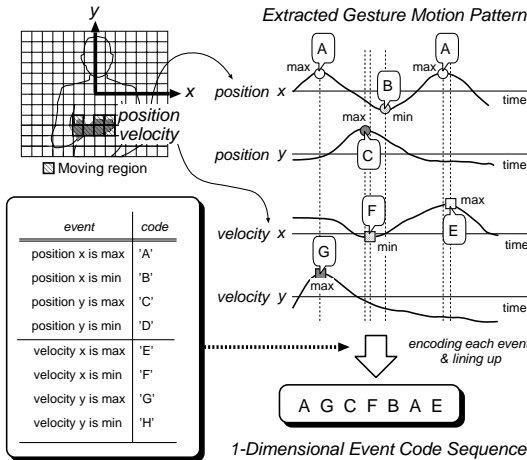


図 4 1次元事象符号列の生成例

Fig. 4 Example of generating the 1 dimensional event code sequence.

つの符号 (code) を用意しておく。たとえば、用いるカメラが 1 台 ( $N = 1$ ) の場合には、可能な 8 通りの事象のために、8 つの符号 (たとえば、アルファベット文字 A ~ H) を用意しておけばよい (図 4 参照)。

実際に発生した事象ごとの符号を、生じた時刻順に並べ、同時刻に複数の事象が生起する場合は、割り当てた符号の値 (文字コード) が小さいものから順に並べることで 1 次元の符号列を生成する (図 4 参照)。これを「1 次元事象符号列」(1-Dimensional Event Code Sequence) と呼ぶ。

#### 4. 動作パターンの認識

一般に、パターン認識のためには、認識対象となる動作パターンの知識 (モデル) を何らかの形で用意する必要がある。本研究では、あらかじめ対象動作を撮影した画像列より生成した「1 次元事象符号列」をモデルテンプレートとして保持しておくものとする。ただし、動作ごとのバラツキやノイズの影響で、動作に対応した符号列は一定しないため、1 つの動作パターンあたり複数個の動作から得られた符号列をテンプレートとして保持するものとする。

このとき問題は、入力された符号列と最も一致する動作パターンをテンプレート群から発見すること (パターンマッチング問題) に置き換えられる。3.2 節で例にあげたように、1 次元事象符号列は文字列にきわめて類似したものであるため、ここでは、文字列比較のアルゴリズムである SED (Shortest Edit Distance: 最小エディット距離) と LCS (Longest Common Subsequence: 最長共通符号列) を求める手法<sup>12)</sup>を応用し

たパターン識別法を提案する。

##### 4.1 最小エディット距離の計算

文字列の類似度を評価する「最小エディット距離 (SED)」を求める手法を以下に示す。

SED とは、2 つの文字列 A, B があるとき、文字列 A に編集操作 (文字の挿入, 削除) を加えて文字列 B を生成するのに必要な最小編集作業量を意味する。

SED を求めるために、エディットグラフというもの考える。エディットグラフとは、文字列 A, B の各文字 ( $A_1, A_2, \dots, A_N$ ), ( $B_1, B_2, \dots, B_M$ ) を  $x$  軸,  $y$  軸にそれぞれ並べて格子を形成し,  $A_i = B_j$  となる場合には, ( $i-1, j-1$ ) から ( $i, j$ ) の対角を結合したものである。

SED を求める問題は、原点 (0, 0) よりスタートし、グラフの格子を形成している縦横線、対角線を通り、右下 (ゴール) に到達するまでの最短経路を求める問題に置き換えられる。

エディットグラフ上の点 ( $i, j$ ) から、右方向への移動により  $x = i + 1$  となれば文字  $A_{i+1}$  の削除, 下方向への移動により  $y = j + 1$  となれば,  $A_i$  に  $B_{j+1}$  を挿入, 対角線への移動が行われれば編集操作はなしとなるので、右または下への移動を距離 1, 対角線上の移動を距離 0 とし、スタートからゴールまでの距離の総和を求めれば編集距離が求まる。

エディットグラフを用いて SED を求めるにはいくつかのアルゴリズムが提案されているが、本研究では、処理する符号列長がたかだか数十と短く、また後に説明する最長一致文字列の決定の容易さから、最も単純な  $O(N^2)$  アルゴリズム<sup>12)</sup>を用いる。

##### アルゴリズム 1 最小エディット距離の計算

文字列 A, B の文字列長をそれぞれ  $N, M$ , エディットグラフ上の点 ( $i, j$ ) での編集量を  $ED(i, j)$  とする。

- (1)  $ED(i, 0) \leftarrow i, ED(0, j) \leftarrow j, ED(0, 0) = 0$  と設定する。
- (2) 開始点 ( $i, j$ )  $\leftarrow (1, 1)$  とする。
- (3) 次の処理を行う。
  - (a)  $A_i = B_j$  のとき
 
$$ED(i, j) \leftarrow ED(i-1, j-1)$$
  - (b)  $A_i \neq B_j$  のとき
 
$$ED(i, j) \leftarrow \min(ED(i-1, j), ED(i, j-1)) + 1$$
- (4)  $i < N$  なら  $i \leftarrow i + 1$  として (3) の処理を繰り返す。
- (5)  $i \leftarrow 1$
- (6)  $j < M$  なら  $j \leftarrow j + 1$  として (3) の処理を繰り返す。

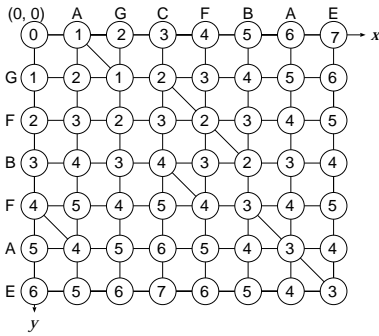
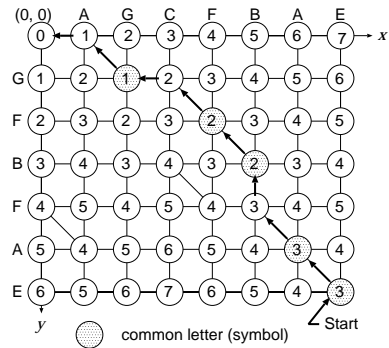


図5  $O(N^2)$  アルゴリズムの適用時のエディットグラフ  
Fig. 5 Edit graph after applying the  $O(N^2)$  algorithm.



**LCS = GFB AE**

図6 LCSの検出例

Fig. 6 Example of finding LCS.

(7)  $ED(N, M)$  の値が SED として求まる。

アルゴリズム 1 を適用した結果求まったエディットグラフ ( $ED$  のテーブル) の一例を, 図 5 に示す。

4.2 最長共通符号列の発見

アルゴリズム 1 で求めた  $ED$  のテーブル (編集量テーブル) をゴールから逆にたどることで最長共通符号列 (LCS) を求めることができる。具体的には次のような処理を行う。

アルゴリズム 2 最長共通符号列の検出

共通文字列を  $CS$ , 文字列  $CS$  の  $p$  番目の文字を  $CS(p)$  で表すものとする。

(1) 開始点  $(i, j) \leftarrow (N, M)$ ,  $p \leftarrow 0$  とする。

(2) 次の処理を行う。

- (a)  $A_i = B_j$  のとき  
 $CS(p) \leftarrow A_i, p \leftarrow p + 1$   
 $(i, j) \leftarrow (i - 1, j - 1)$
- (b)  $A_i \neq B_j$  かつ  
 $ED(i - 1, j) < ED(i, j - 1)$  のとき  
 $(i, j) \leftarrow (i - 1, j)$
- (c)  $A_i \neq B_j$  かつ  
 $ED(i - 1, j) \geq ED(i, j - 1)$  のとき  
 $(i, j) \leftarrow (i, j - 1)$

(3)  $j > 0$  かつ  $i > 0$  ならば (2) の処理を繰り返す。

(4)  $CS$  の文字列の順序を逆転したものが LCS。

例として図 5 から LCS を求めたものを図 6 に示す。

4.3 一致パターンの検出

比較する 2 つの事象符号列を  $S$  と  $T$  とする。ここで,  $S$  を比較対象 (入力された身振り動作から得られた符号列),  $T$  をテンプレートとする。まず, アルゴリズム 1 を用いて  $S$  と  $T$  の SED を計算する。次に, SED となるときの編集量テーブルにアルゴリズム 2 を適用して  $S$  と  $T$  の LCS ( $LCS(S, T)$ ) を求める。

2 つの事象符号列の一致度  $P_{sim}(S, T)$  は次の式で計算される:

$$P_{sim}(S, T) = \frac{\text{len}(LCS(S, T))}{\max(\text{len}(S), \text{len}(T))} \quad (1)$$

ここで,  $\text{len}(A)$  は符号列  $A$  中の符号列長を表す。

入力から得られた事象符号列  $S$  とテンプレートの符号列とを比較し, 最も一致する動作種を検出する手順をアルゴリズム 3 に示す。

アルゴリズム 3 一致パターンの検出

1 動作あたりに保持する符号列の数をテンプレート保持数と呼び,  $N_t$  と表記する。また,  $\mathcal{M} = \{\text{すべての動作種}\}$ ,  $\mathcal{N} = \{1, 2, \dots, N_t\}$  と定義する。さらに, ある動作種  $m$  のテンプレートを  $T(m)$ , そのテンプレートに含まれる特定要素  $j$  の符号列を  $T_j(m)$  と表記する。

- (1) すべての  $m \in \mathcal{M}$  について,  
 $P_{sim}(S, T(m)) = \max_{j \in \mathcal{N}}(P_{sim}(S, T_j(m)))$  を計算する。
- (2) 最も一致度が高い動作種として,  
 $\max_{m \in \mathcal{M}}(P_{sim}(S, T(m)))$  となる  $m$  を求める。

4.4 認識結果の信憑性の評価

この手順では, 入力された符号列は, 「必ず」テンプレートにある動作のいずれかとして認識される。しかしながら, 実際には意図的ではない動作や, システムにとって未知な身振りが入力されることもあり, 認識結果の信憑性の評価法について検討する必要がある。これに関して, 実際に前節までに述べた方法を適用して予備的な身振り認識実験を行った結果, 「正しい認識が行われた場合には, 一致度が最も高いものと第 2 位の  $P_{sim}$  の差が比較的大きく, 誤った認識が行われた際には小さく現れる」傾向が見られた<sup>13)</sup>。そこで, 本論文では, 「第 1 位と第 2 位の一致度の差」を評価値とし認識結果が適切か否かを判定する手法を導入する。

ある入力  $S$  に対する一致度が第 1 位, 第 2 位と判定された動作種をそれぞれ  $m_1, m_2$  とすると, 第 1 位と第 2 位の一致度の差は

$\Delta P_{sim12} = |P_{sim}(S, T(m_1)) - P_{sim}(S, T(m_2))|$  となる. ここで, ある閾値  $Th_{ambiguous}$  を設け次の式により認識結果の信憑性を判定する.

$$\begin{cases} \Delta P_{sim12} \geq Th_{ambiguous} : \text{信憑性あり} \\ \Delta P_{sim12} < Th_{ambiguous} : \text{不明瞭} \end{cases} \quad (2)$$

「不明瞭」と判定された身振りに関しては, 認識結果は信頼できないものとし, その結果を棄却する.

## 5. 検証実験と考察

提案手法を検証するために, 動作者の正面および上方(動作者の手を見下ろす位置)に配置した 2 台のカメラからの入力を符号化, 認識する実験を行った.

### 5.1 検証実験システムの概要

まず, 正面と上方から動作者の手をとらえる位置にカメラを配置するため, 図 7 に示すようなカメラ台を製作した. 台にはキャストが取り付けられており被験者(動作者)に対する位置決めを行うことができる. とくに, 上部カメラの高さはフリーストップ機構を持つ昇降アームにより調整可能である. 本実験では, 正面カメラ(camera 1)と上部カメラ(camera 2)は被験者が自然な姿勢で構えたときの手の位置から, それぞれ約 600 mm と 1,000 mm に設定した. また, このとき camera 1 は手を画面の中心にとらえるように傾けて設置した.

身振り動作の処理には, ビデオキャプチャカード(アイ・オー・データ製 GV-VCP2/PCI)を搭載した 1 台のパーソナルコンピュータ(CPU: Intel Pentium III 700 MHz, メモリ: 128 MB, OS: Linux)を利用した. 2 台のカメラからの映像は, 4 画面分割器(ビデオデバイス製 TV-550)によりそれぞれが画面の 4 分の 1 の領域に割り当てられ, 1 つの映像信号としてビデオキャプチャに入力される. キャプチャ解像度は  $320 \times 240$  画素, 各点 8 ビット, すなわち, カメラ 1 台あたりの画像解像度は  $160 \times 120$  画素となる. 本実験では, これを  $20 \times 15$  個のブロック(つまり小領域のサイズは  $8 \times 8$  画素)に分割した. フレームレートは毎秒 10 フレーム

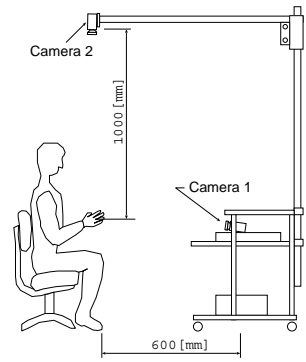


図 7 身振り認識のためのカメラ配置

Fig. 7 Placement of cameras for capturing gestures.

ム, 「動きあり」と見なすフレーム間差分の閾値(輝度差)は 30 とした. また, 身振りを切り出す際の動作速度および継続時間の閾値(2.3 節参照)は, それぞれ次のように設定した:  $v_{th-rg} = 15(\text{pixel/frame})$ ,  $v_{th-gr} = 1(\text{pixel/frame})$ ,  $n_{th-rg} = 2(\text{frame})$ ,  $n_{th-gr} = 3(\text{frame})$ . さらに, 時系列データのノイズ処理用ガウシアンフィルタの標準偏差(3.1 節参照)は 1.5 に設定した. 身振りの認識処理は, 1 つの動作が終了するたびに行われる. 認識に要する時間は保持しているテンプレートの数と符号長, ならびにシステムの性能自体にも依存するが, 本実験の場合, 動作の区切れ検出(身振り動作の切り出し)から認識完了までの時間遅れは約 0.1~0.3 秒であった. なお, 本実験では, 背景環境や照明条件については特別な整備は行っていない. ただし, 各カメラの絞りについては, あらかじめマニュアルで良好な状態(民生用ビデオカメラの自動露光調整機能により撮影される画像とほぼ合致する程度)になるよう調整しておいた.

### 5.2 入力動作パターンとテンプレート

本研究では応用例としてロボットの誘導を考えている(6 章参照). 誘導に用いる身振りとして, 一般に, 手の前後あるいは左右方向の往復運動や, 円(弧)を描く運動, およびこれらの組合せ動作が考えられる. こうした点をふまえ, 本実験では, 図 8 に示す「アーチ(arch)型」「サークル(circle)型」「オーム(ohm)型」「前振り(mae)型」, および「横振り(yoko)型」の 5 形式, 前振り型を除きそれぞれ左右両方向(同じ手で動作方向が逆)の計 9 種類の動作パターンを認識対象として選択した.

実験条件を変えても被験者の入力動作を一定にできるように, 直接認識処理系に画像入力を渡すのではなく, いったん DV(デジタルビデオ)に録画しておき, それを再生することで処理系に入力した. 被験者

ここでは紙面の都合上, 2 台のカメラを入力とするシステムのみ扱っているが, 提案手法自体には“使用するカメラの台数に何ら制約はない”ことに留意されたい. すなわち, カメラが 1 台あるいは逆に 3 台以上でもすべてのアルゴリズムがそのまま動作可能である. たとえば, 文献 13) に, カメラ 1 台のシステムによる提案手法の動作例が掲載されているので参照されたい. 最適なカメラの台数と配置は, 認識すべき動作の種類にも依存する興味深い研究課題であるが, 本論文では取り扱わない.

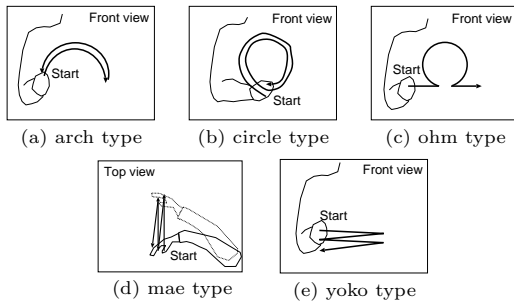


図8 認識実験に用いる身振り

Fig. 8 Gestures for the experiment.

の数は4名、それぞれの動作パターンについて30回の動作を行った。このうち前半10回の動作はテンプレート生成用とし、この10回分の動作から各被験者ごとに無作為に5個を選び、その被験者のテンプレートと定めた。すなわち、本実験のテンプレート保持数は、 $N_t = 5$ である。認識実験には後半の20回の動作を入力とした。

### 5.3 結果と議論

本実験では、とくに次の3点を検証・評価した。

- (1) 提案手法の身振り認識性能
- (2) 認識結果の信憑性判定法の有効性
- (3) テンプレート内のパターン間の差異

以下で、それぞれの検証結果と考察を述べる。

#### [ 議論1 ] 提案手法の身振り認識性能

各人のテンプレートに対し、

- テンプレート生成者の身振りのみを認識対象とする場合
  - すべての被験者の全身振りを認識対象とする場合
- についてそれぞれの認識実験を行い、その認識率の全被験者分の平均値を計算した。

その結果、図9に示すように、テンプレート生成者と認識対象の身振りを行った被験者が同一の場合(図中で Template Generator と記載されている棒グラフ)には、全体で90%以上の認識率が得られた。一般に、同一人物が生成する動作パターンの場合、その動作の速さや空間中に描くパターンの大きさは、各試行ごとに微妙に異なる(パターンが時空間的に伸縮する)が、その人物が持つ手の振り方の“くせ”(ある部分の動作を行っているときに手を振る速度が速くなる傾向がある、全体にひしゃげた形で手を振るなど)を含むパターンの本質的な特徴は変わらない。すなわち、この結果は、提案した動作パターン符号化手法が、同一人物が生成した動作パターンの変動(時空間的な伸縮)に柔軟に対応している(個人の動作パターンの不変的な特徴をうまく抽出している)ことを示すものであり

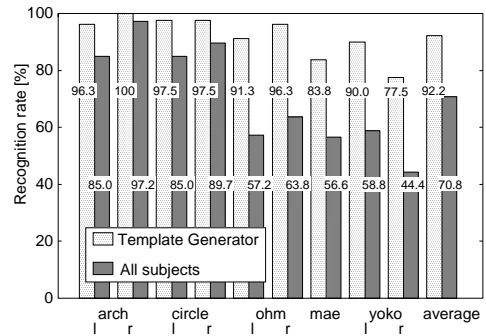


図9 テンプレート生成者の動作入力と全被験者の動作入力時の認識率

Fig. 9 Recognition rate using input from the template generator and from the all subjects.

意義深い。これは、1次元事象符号列の本質的な特徴(事象の絶対的な生起時刻に重点を置かず、事象間の相対的な関係に着目している点)と深く関連している。従来法の多くが、このようなパターンの伸縮に対処するために、DP(動的計画法)やHMM(隠れマルコフモデル)などを駆使して複雑な認識系を構築していたのに対し、ここでは、認識系に関しては、文字列比較アルゴリズムベースのきわめてシンプルな構造のものを用いているにすぎないことに留意されたい。

一方、すべての被験者の身振りを認識対象とした場合(図中で All subjects と記載)には全体で70%程度の認識率にとどまっている。とくに、ohm型、yoko型、mae型の身振りにおいて、認識率に大きな低下がみられた。これは、主に手の直線的な折り返し動作の部分での個人差が影響しているためと考えられる(図8参照)。いいかえれば、本手法(動作パターンの符号化)は、時系列パターンの時間的な伸縮に柔軟に対応する一方で、同時に、動きの個人差も表現(場合によっては強調)してしまうと考えられる。このような動作パターン自体の変動にも対処するためには、単一人物のテンプレートのみでは本質的に不十分であり、従来法のように、複数人物のテンプレートからのモデル学習プロセスが必要になるとと思われる。

次に、認識率を身振りの種類別に見てみると、arch型やcircle型の認識率が95%以上(全被験者を対象とした場合でも85%以上)と高く、ohm型も比較的高い認識率が得られているのに対して、mae型とyoko型(とくにyoko\_rの方)の認識率が低い。これは、提案手法では動きの極値の出現順にのみ着目しているため、右横方向に手を振るyoko\_r型とmae型の場合、その動作形状の性質上(右手でmae動作を行う場合、左から右へ振る動きが自然に生じてしまう)、類似した



表1  $Th_{ambiguous}$  と誤認率 (M) および棄却率 (R) [単位: %]  
 Table 1 Relationship between  $Th_{ambiguous}$  and miss-judge rate(M) and rejection rate(R) [unit: %].

	$Th_{ambiguous}$					
	0.00		0.02		0.05	
	M	R	M	R	M	R
arch_l	3.8	0	2.7	3.8	1.4	10.0
arch_r	0	0	0	1.3	0	2.5
circle_l	2.5	0	1.3	1.3	0	6.3
circle_r	2.5	0	1.3	1.3	1.5	6.3
ohm_l	8.8	0	5.6	10.0	4.7	21.3
ohm_r	3.8	0	1.4	7.5	1.6	15.0
mae	16.3	0	12.9	10.0	2.5	31.3
yoko_l	10.0	0	5.9	11.3	5.7	30.0
yoko_r	22.5	0	11.2	13.8	6.9	30.0
average	7.8	0	4.7	9.0	2.7	16.9

符号列が割り当てられることがあり、パターン間に差が現れにくくなっていることが一因と考えられる。この点に関しては、議論3で改めて考察を行う。

#### [ 議論2 ] 認識結果の信憑性判定法の有効性

4.4 節で提案した認識結果の信憑性判定の閾値  $Th_{ambiguous} = \{0.0, 0.02, 0.05\}$  としたときの誤認率 M (正しい認識が行われなかった確率)、棄却率 R (不適切な入力として棄却された確率) を計算した。

その結果、表1に示すように、 $Th_{ambiguous}$  の値が大きくなるほど、誤認率 M は低下するが、棄却率 R は増加する傾向が見られた。この結果より、認識結果の信憑性判定法自体は有効に動作しているといえる。ただし、 $Th_{ambiguous}$  の値の優劣については一概にどれが良いとはいえない。これは、アプリケーションに応じて設計者が入念に調整すべきものであると考える。たとえば、ペットロボットとのコミュニケーションなど、認識精度そのものよりも何らかの反応が得られるということに重点を置くようなアプリケーションでは、棄却される入力が低減されるように、 $Th_{ambiguous}$  を低めに設定する方が良い。一方、誤認識の増大が致命的な結果を招くようなタスクにおいては、認識精度 (誤認率の低減) を最優先し、 $Th_{ambiguous}$  を高めに設定すべきである。アプリケーション構築の一具体例として、6章において「身振りによる移動ロボットの誘導」を取りあげ、 $Th_{ambiguous}$  の導入・設定例を示す。

#### [ 議論3 ] テンプレート内のパターン間の分離度

議論1より、認識率は認識対象とする動作パターンの組合せに依存すると考えられる。ここでは、これを判定する基準として、パターンどうしの差異を評価する方法を考える。

本研究の提案手法では、パターンは1次元の符号列として表記され、テンプレートも同様のデータを保持

している。そこで、テンプレートに保持している符号列どうしを比較し、パターンどうしの分離度 (類似性が低い度合い) を示す指標を提案する。

テンプレートには、ある動作  $m$  について複数個 ( $Nt$  個) のデータが保持されているものとし、その  $i$  番目の符号列を  $T_i(m)$  と表記する。さらに、テンプレートに含まれるある動作  $m$  と  $n$  のそれぞれ  $i, j$  番目の要素どうしのパターン一致度を  $P_{sim}(T_i(m), T_j(n))$  と表現するものとする。このとき、テンプレート内の動作  $m$  と自身以外の動作  $n$  (つまり  $n \neq m$ ) を比較した場合におけるパターン一致度の平均  $E_{po}(m, n)$  は、次の式で与えられる。

$$E_{po}(m, n) = \frac{\sum_{i=1}^{Nt} \sum_{j=1}^{Nt} P_{sim}(T_i(m), T_j(n))}{Nt \cdot Nt}$$

一方、テンプレートに含まれる同じ動作種  $m$  どうしを比較したときのパターン一致度の平均  $E_{ps}(m)$  は、次のように書ける。

$$E_{ps}(m) = \frac{\sum_{i=1}^{Nt} \sum_{j=1, j \neq i}^{Nt} P_{sim}(T_i(m), T_j(m))}{Nt(Nt-1)}$$

これらの量を用いて、動作  $m$  に対する  $n$  の分離度を次のように定義する。

$$Sep(m, n) = E_{ps}(m) - E_{po}(m, n) \quad (3)$$

この定義から分かるように、一般に  $Sep(m, n) \neq Sep(n, m)$  である。 $Sep(m, n)$  が低いほど、動作  $m$  を  $n$  と混同する可能性が高くなる。たとえば、 $Sep(m, n)$  の値が低く、 $Sep(n, m)$  の値が高い場合には、動作  $m$  を  $n$  と混同する可能性は高いが、動作  $n$  を  $m$  と混同する可能性は低いといえる。

実験に使用した各被験者のテンプレートごとに  $Sep(m, n)$  を計算したところ、比較的誤認が多かった yoko 型や mae 型では、4名の被験者の値を平均して 0.17 程度、その他のパターンでは約 0.21 ~ 0.34 であった。また被験者によっては認識率が低かった yoko 型に対する circle 型の評価などで 0.08 程度の低い値となっていることが分かった。これらより、本指標がテンプレートに保持しているパターン間の差異を評価する基準として利用できると考えられる。たとえば、あるアプリケーションのためにあらかじめデザインされた身振り動作パターンの組合せが適当か否かを事前に検討するのに利用できる。次章でこの一例を示す。

#### 6. 応用実験：身振りによるロボットの誘導

提案手法の応用例として、身振りにより移動ロボットを誘導 (遠隔操作) する実験を行った。

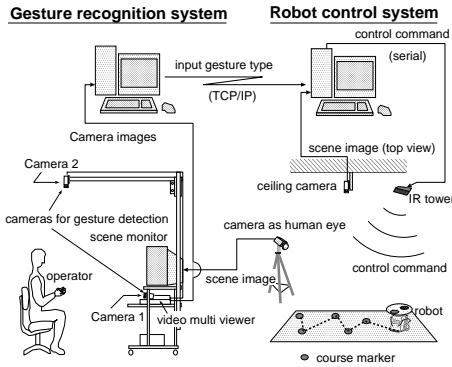


図 10 身振りによる移動ロボット誘導システム  
Fig. 10 Gesture-based robot navigation system.

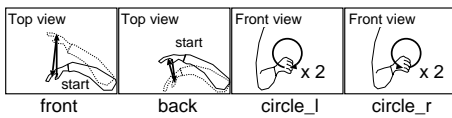


図 11 ロボットの誘導に用いる身振り動作  
Fig. 11 Gestures for robot navigation.

6.1 実験システムの概要

本実験では、ロボットを所定のスタート位置からマークによって示されたコースに沿って走行させ、ゴール位置まで誘導することをタスクとする。図 10 に示すように、被験者（操作者）はロボットを動かすフィールドを斜めに見下ろす形で配置されたカメラで撮影した画像をモニターで確認しながら、身振りによりロボットに指示を行う。身振りは 2 台のカメラにより入力・処理され、身振り認識結果が制御コマンド（前進、後退、左旋回、右旋回）に変換される。ロボットへの制御コマンドの送信は赤外線通信ポートを通じて無線で行われる。ロボットは連続的に移動するのではなく、身振りに対する行動を一定量行った後に停止して次の指示を待つ。本移動ロボット誘導システムの詳細は文献 13) を参照されたい。

ロボットの誘導を行うために用いる身振りは、図 11 に示すように、手を前、後にそれぞれ 1 往復させる動作（front, back）と、左回り、右回りにそれぞれ 2 回手を回す動作（circle\_l, circle\_r）の計 4 種類とし、front を前進、back を後退、circle\_l を左旋回、circle\_r を右旋回に割り当てた。

6.2 身振り認識予備実験とロボット誘導実験

実際にロボットの誘導実験を行う前に、行動指示に用いる身振りの認識精度について検証した。前章の実験と同様、4 名の被験者に対象動作（front, back, circle\_l, circle\_r）を各 30 回行ってもらい、これらを DV に録画し、初めの 10 回分から  $Nt = 5$  としてテ

表 2 ロボット誘導用身振りの誤認率 (M)/棄却率 (R) [ 単位 : % ]  
Table 2 Mis-judge rate(M)/rejection rate(R) of the gestures for robot navigation [unit:%].

	$Th_{ambiguous}$					
	0.00		0.02		0.05	
	M	R	M	R	M	R
front	1.3	0	1.3	0	0	1.3
back	2.5	0	2.5	0	2.5	1.3
circle_l	0	0	0	0	0	0
circle_r	0	0	0	0	0	0
average	0.9	0	0.9	0	0.6	0.6

ンプレートを作成し、残りの 20 回を検証用の入力としたときの誤認率と棄却率を表 2 に示す。

表 2 より、 $Th_{ambiguous}$  が 0.02 の場合においても棄却率がゼロとなったが、ロボットを誘導するという意図がともなう場合には、動作のパラッキが予備実験時より大きくなったり、動作の区切れで十分に手が停止しなかったりすることも考えられるため、実際には  $Th_{ambiguous} = 0.02$  でも棄却が生じると考えられる。また、同様の理由から、 $Th_{ambiguous} = 0.05$  とした場合、棄却率が必要以上に増大し、操作者に不快感を与えることも考えられるため、以降のロボット誘導実験では  $Th_{ambiguous} = 0.02$  を採用する。

ここで、表 2 の誤認率、棄却率とも、前章の 9 種類の身振りに対する検証実験結果（表 1）と比べるとかなり低くなっていることに注目されたい。たとえば、表 1 の身振り mae と表 2 の身振り front は、図 8 と図 11 を比較すれば分かるように、手を振る繰返し回数のみが異なる動作パターンであるが、誤認率ならびに棄却率は、mae に比べて front の方がはるかに低くなっていることが分かる。これは、mae と front の動作パターンの違い（手を振る繰返し回数の差）が原因ではなく、5.3 節の議論 1 や議論 3 でも述べたように、認識対象となる他の身振り動作との組合せに起因するものである。前章の実験では、mae と類似の符号列が割り当てられる可能性が高い yoko 型が認識対象として存在しており、このことが両者の認識率を悪化させていた。一方、本実験の場合、表 2 の結果より、front と類似の符号列が割り当てられる可能性の高い（すなわち、front のパターン分離度を大幅に低下させる）認識対象が含まれていないと推察される。このことを確認するために、5.3 節で定義した身振りのパターン分離度（被験者全員の平均値）を計算した。その結果、表 3 に示すように、本実験における 4 種類の身振りの組合せに関するパターン分離度の最低値は 0.22（back に対する front の分離度）であり、前章の検証実験における身振りパターンの組合せ（5.3 節の議論 3）に

表3 ロボット誘導用身振りのパターン分離度の評価  
Table 3 Pattern classification rates of the gestures for robot navigation.

	front	back	circle_l	circle_r
front	—	0.32	0.33	0.31
back	0.22	—	0.28	0.27
circle_l	0.39	0.44	—	0.24
circle_r	0.35	0.42	0.23	—

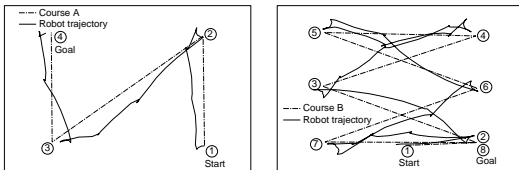


図12 ロボットの移動軌跡の一例(被験者1)  
Fig.12 Trajectories of the robot (Subject 1).

比べると,良好なパターン分離度が得られていることが確認できる。

以上の予備実験をふまえて,2つのコースを設定し,4名の被験者に身振りによりロボットを誘導してもらったところ,全被験者とも,おおむね指定されたルートに沿ったロボットの誘導に成功した(図12)。

## 7. おわりに

本論文では,身振り認識の手法について,「身振り動作の検出」,「運動パターンの符号化」,「動作パターンの認識」の3つの段階にわけて,それぞれの問題ごとに考察を進めてきた。

まず「身振り動作の検出」では,環境の影響を受けにくい動きの場の重心位置/速度を利用した。次に「運動パターンの符号化」では,動作の特徴は運動に変化が生じる時点で強く現れるという考えから,運動の重心位置/速度の極大点,極小点を動きの特徴量としてとらえた。こうして抽出された運動の特徴量に符号を割り当て,それを出現時刻順に並べ1次元の符号列(1次元事象符号列)を生成した。この1次元事象符号列は,時間の同期がとれば他の時系列情報も同じ枠組みで取り扱うことができるため,センサフュージョンやマルチモーダルインタフェースへの利用は容易であると考えられる。「動作パターンの認識」には,文字列比較のアルゴリズムを応用し,パターンどうしの類似度を評価する手法を検討した。

提案手法を検証するため,いくつかの基本動作の認識率を測定した。その結果,認識に用いたテンプレート生成者と動作入力者が同一の場合には平均して90%以上の高い認識率が得られた。その一方で,テンプレ

ト生成者と動作入力者が異なる場合には認識率は低くなる傾向が見られた。この傾向は,動作者の癖が出やすいと思われる動きの急激な変化をとともなう形状にとくに強く現れた。これらのことから,提案した1次元事象符号列は,個人ごとの動作の差異を記述している(個人ごとの動作の特徴を的確にとらえている)と考えられる。

また,正解時と誤認時との一致度の傾向に着目し,不明瞭な身振り入力を判定する手法を提案した。

さらに,認識しうるパターンの組合せを評価する方法を提案した。これはより良いテンプレートを構築する指標になると考えられる。これらを利用し,より良いテンプレートを構築する手法を検討することは,今後の課題の1つである。

## 参考文献

- 1) 牛田博英,山口 亨,高木友博:ファジー連想記憶システムを用いた動作認識,電子情報通信学会論文誌,Vol.J77-D-II, No.8, pp.1571-1581 (1994).
- 2) 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).
- 3) 大和淳司,倉掛正治,伴野 明,石井健一郎:カテゴリ別VQを用いたHMMによる動作認識法,電子情報通信学会論文誌,Vol.J77-D-II, No.7, pp.1311-1318 (1994).
- 4) 渡辺孝弘,谷内田正彦:複数入力画像の固有空間法による実時間ジェスチャ認識,電子情報通信学会論文誌,Vol.J81-D-II, No.5, pp.810-821 (1998).
- 5) Davis, J.W. and Bobick, A.F.: The Representation and Recognition of Human Movement Using Temporal Templates, *Proc. 1997 IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp.928-934 (1997).
- 6) 桐島俊之,佐藤宏介,千原國宏:身振りの動画認識によるユーザインタフェース,パターン認識理解研究会資料PRU95-190,電子情報通信学会技術研究報告(1995).
- 7) 西村拓一,向井理朗,野崎俊輔,岡 隆一:低解像度特徴を用いた複数人物によるジェスチャの単一動画からのスポッティング認識,電子情報通信学会論文誌,Vol.J80-D-II, No.6, pp.1563-1570 (1997).
- 8) 畠 直志,岩井儀雄,谷内田正彦:動き情報と情報圧縮を用いたロバストなジェスチャ認識手法,電子情報通信学会論文誌,Vol.J81-D-II, No.9, pp.1983-1992 (1998).
- 9) 西川 敦,大西映生,西村正典,平野敦士,小荒

健吾, 宮崎文夫: 局所相関演算に基づくオプティカルフローを用いた身振り動作の認識手法, 情報処理学会論文誌, Vol.40, No.8, pp.3118-3133 (1999).

- 10) Cutler, R. and Turk, M.: View-based Interpretation of Real-time Optical Flow for Gesture Recognition, *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, pp.416-421 (1998).
- 11) 間瀬健二: 顔とジェスチャの検出および認識, 日本ロボット学会誌, Vol.16, No.6, pp.745-748 (1998).
- 12) 津田伸秀: テキストエディタ(特集 プログラミング tips 大全第1章), *C MAGAZINE*, Vol.11, No.10, pp.45-55, ソフトバンクパブリッシング (1999).
- 13) Koara, K., Nishikawa, A., Shinto, K., Ishii, K., Yasui, Y. and Miyazaki, F.: Gesture Recognition based on 1-Dimensional Encoding of Motion Changes, *Proc. 10th Int. Conf. on Advanced Robotics*, Budapest, Hungary, pp.639-644 (2001).

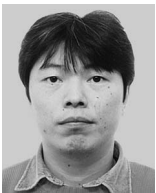
(平成 14 年 1 月 21 日受付)

(平成 14 年 12 月 3 日採録)



小荒 健吾

昭和 49 年生。平成 9 年大阪大学基礎工学部機械工学科卒業。平成 13 年同大学大学院基礎工学研究科システム人間系専攻博士後期課程修了。現在、三菱電機(株)に勤務。在学中は、画像を用いた身体運動の解析と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。博士(工学)。日本ロボット学会, 電子情報通信学会各会員。



西川 敦(正会員)

昭和 42 年生。平成 7 年大阪大学大学院基礎工学研究科物理系専攻博士後期課程修了。同年米国南カリフォルニア大学客員研究員。平成 8 年大阪大学基礎工学部助手。平成 9 年同大学大学院基礎工学研究科助手となり現在に至る。この間平成 6 年~8 年日本学術振興会特別研究員。動画像を用いた人間の動作認識とその応用に関する研究に従事。博士(工学)。日本ロボット学会, 電子情報通信学会, 日本コンピュータ外科学会, IEEE 各会員。



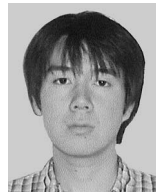
進戸健太郎

昭和 51 年生。平成 12 年大阪大学基礎工学部機械工学科卒業。平成 14 年同大学大学院基礎工学研究科システム人間系専攻博士前期課程修了。現在、日本 IBM(株)に勤務。在学中は、画像を用いたヒューマンインタフェースの研究開発に従事。



石井 薫

昭和 53 年生。平成 13 年大阪大学基礎工学部システム科学科卒業。現在(株)本田技術研究所に勤務。在学中は、身振りをを用いたヒューマンインタフェース応用に関する研究に従事。



安井 陽一

昭和 53 年生。平成 13 年大阪大学基礎工学部システム科学科卒業。現在、同大学大学院工学研究科知能・機能創成工学専攻博士前期課程在学中。身振り動作パターンの認識手法に関する研究に従事。



宮崎 文夫

昭和 27 年生。昭和 54 年大阪大学大学院基礎工学研究科物理系専攻博士課程中途退学。同年大阪大学基礎工学部助手。昭和 61 年同助教授。平成 3 年同教授。平成 9 年同大学大学院基礎工学研究科教授となり現在に至る。この間昭和 62 年~63 年米国カリフォルニア大学客員準教授。ロボットの知能化に関する研究に従事。工学博士。日本ロボット学会, システム制御情報学会, 計測自動制御学会, IEEE 各会員。