コンピュータビジョンと
イメージメディア 108－6
（1997. 11. 27）

人間の身振りと手の形状の統合による手話オンライン認識システムの試作

辻本剛佳、平山健一郎、西堀義仁、山口亨、鎌田一雄
宇都宮大学 工学部 情報工学科

あらまし：手話を用いる人々にとって、インタラクティブな手話インタフェースの構築が望まれている。今まで検討した手話認識システムは、動作の空間的な特徴だけを用いるものであったが、本稿では、簡単な指の特徴を認識する新たな手法を提案し、従来の手法と組み合わせて用いる。これらの特徴について、注意機能を用いた概念ファジィ集合によって連想推論を行い、認識を行なっている。

なお、このシステムでは、特定の人によらずに高い認識率を得ることができることを実験により示す。また、人間からの音声や文字入力に対する手話アニメーションを表示することにより、インタラクティブな会話も行うことができる。

A prototype of Japanese sign language recognition system based on motion feature and shape feature integration

Takayoshi TSUJIMOTO, Kenichiro HIRAYAMA, Yoshihito NISHIBORI, Toru YAMAGUCHI and Kazuo KAMATA
Department of Information Science, Faculty of Engineering, Utsunomiya University

Abstract: It is helpful to develop an interactive Japanese sign language interface for the people with hearing disabilities who uses Japanese sign language. Our previous method used only spatial motion features of sign words. In this paper, we propose another method which processes and analyzes simple shape of hands, and combine it with our previous method. The proposed system switches there two features using attention function on Conceptual Fuzzy Sets(CFS). We also use Fuzzy Associative Memory Organizing Unit System(FAMOUS) to infer the final result. FAMOUS has a good robustness property. Experimental results, using a small set of words, show that our system has person independent feature and high recognition rate. In addition, we realize an interactive interface by generating animations for Japanese sign words.

1. はじめに

現在、手話を話すことができる健聴者の数は少なく、健聴者とコミュニケーションの手段として手話を用いるろう者がスムーズな会話をすることは困難である。このことから、一般的な医療施設等において、患者であるろう者と健聴者である医師等が病状の伝達などのコミュニケーションをはかるとき第三者を介する場合が多く、患者のプライバシー侵害の問題や、詳しい内容を正確に伝えられないという問題などが生じている。このため、これらの問題、環境を改善するために、手話インタフェースの実現は重要であり、インタラクティブな手話インタフェースの実現が望まれている。

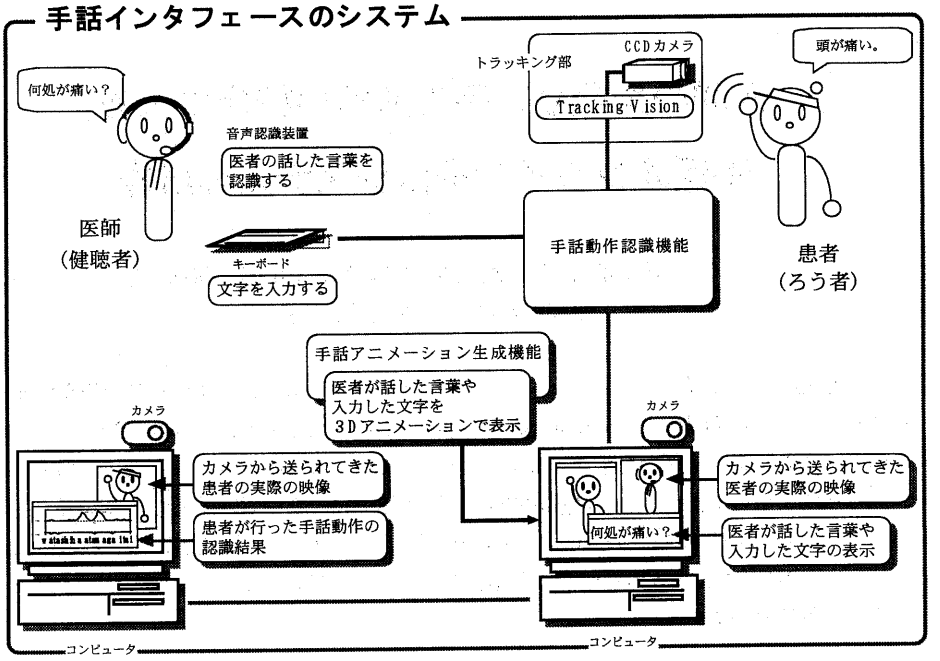

図1 手話インタフェースのシステム構成

2. 手話インタフェース

手話動作認識機能、手話動作生成機能を実現し、それらを統合することによりインタラクティブな手話インタフェースを構築する。

手話インタフェースでは、状況依存とともに、(i)単語分離、(ii)手の動きによる単語認識、(iii)手の形状による単語特定、を統合した認識処理を行う必要がある。(i),(ii)については、山口、吉原らがその手法を提案し実現した[1](以後は「現システム」と呼ぶ)。しかし、これは動作の空間的な特徴だけを用いるもので、手の形状の特徴については考慮していない。本稿では(iii)の機能を実現するために、画像情報処理を行うことによって手の形状を認識する新たな手法を提案し、現システムと組み合わせて用いる(以後は「新システム」と呼ぶ)。そのため本手法では、動作の空間的な特徴と手の形状という2種の特徴について考える。ここでは、大森のモデルによる注意機能[2]と同様の手法を用い、どちらの特徴に注目するかを切換えることにより動作の認識を行う。現システムと新システムにおいて、それぞれの実験結果の比較を行うことによって、本手法の有効性を示す。

2.1 手話動作認識部のシステムの構成

図2に手話動作認識部のシステムの構成を示す。

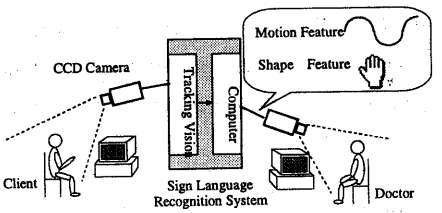

図2 手話動作認識部のシステム構成

ビデオ情報は CCD camera から得て、Tracking Vision に送る。Tracking Vision は人間の皮膚の

gray-level に基づいて、複数のブロックを検出することができるハードウェアであり、人間の動作を２次元座標データに変換し、1/30 秒毎に計算機に送る。それから画像情報データを得る。計算機は入力データから人間の動作の特徴点を引き出して、これらの特徴点に対して連想記憶に格納されたファジィラベルを割り当てる。このあと、連想推論と画像情報処理を行い認識が行われる。本稿ではこの処理をマッチングと呼ぶ。

もし、動作の空間的な特徴だけで明確に手話単語を認識(特定)できない場合は、注意機能をもつＣＦＳを用いることによって、注意している対象を動作の空間的な特徴から手の形状の特徴に切換えることによって,その手話単語を特定することができる。このようにして、医師は手話通訳を介さずに患者であるろう者の意見を理解することができる。

反対に、医師が患者に話しかけると、注意が音声の特徴に切り換わり、その入力に対応した手話アニメーションを生成・表示する。

2.2 注意機能をもつ概念ファジィ集合（ＣＦＳ）

パターンとシンボルを結び付ける手法として Conceptual Fuzzy Sets (CFS) [3] がある。これはパターン-シンボル対の知識を属性別に作成し、それぞれの属性のマトリクスを加算することによって属性を統合して連想メモリで全体を実現する。これは、以下のような式で示される。

$$M = M_1 + M_2 + \ldots + M_n \qquad (1)$$

ただし、$M_1, M_2, \ldots, M_n$ はそれぞれ個別の CFS の連想マトリクスであり、統合された CFS のマトリクスは M である。

本稿では,更に属性を切換える機能を付け加えることにより、どの属性に注目するかを切換える。この切換え機能は海馬モデル [2] に類似した機能であり、注意機能と呼ぶ。提案する CFS を式 (2) のように表現する。

$$M = attn_1 M_1 + attn_2 M_2 + \ldots + attn_n M_n \qquad (2)$$

$$attn_i \in [01] \ , \quad i = 1 \sim n$$

ただし、$attn_1, attn_2, \ldots, attn_n$ は属性別に注意を制御する変数である。図３に注意機能をもつCFSの例を示す。

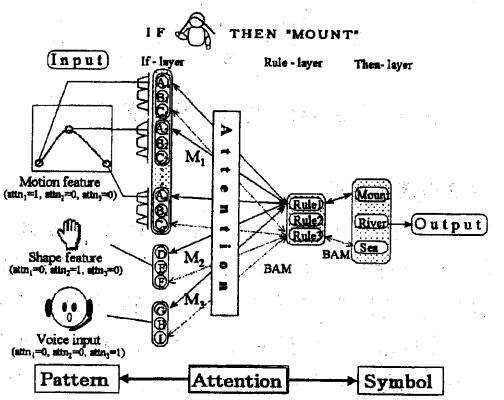

図３ 注意機能をもつＣＦＳ

図３において、もし医師が患者に話しかけたら、このシステムは、その声を認識して、それに対応した手話アニメーションを生成し、表示する。反対に、患者が手話で話すときは、もし、その手話動作を空間的な特徴だけで、判断することができないなら、動作の空間的な特徴から、手の形状の特徴に注意を切換えることによって、その手話単語を特定する。このように、注意機能をもつ CFS は、注意を切換えるための機能である。

3. 動作の空間的な特徴のみのシステムの説明

(ｉ)単語分離、(ii)手の動きによる単語特定で構成しているシステムは、トラッキング部、動作の特徴抽出部、連想推論部から構成される。

3.1 トラッキング部

2.1 でトラッキング部について説明した。トラッキング部は手の位置を２次元座標データに変換する。

3.2 動作の特徴抽出部

この部分が、2次元座標データを受け取り、動作の特徴点（「大きい山」「大きい谷」など、図4参照）を抽出する。

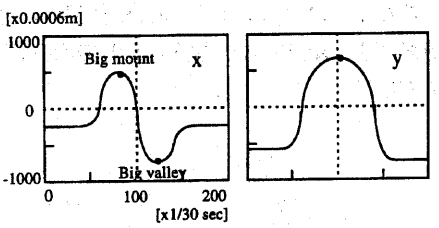

図4　手話「mount」の連続時間データ

3.3 連想推論部

この部分は、Fuzzy Associative Memory Organizing Unit System (FAMOUS) によって連想推論 [4] [5] を行う。連想の入力は動作の特徴点である。すなわち「大きい山」、「大きい谷」などである。FAMOUS は if-layer と rule-layer、then-layer の3層から成り立つ。第1層が入力層であり、第2層が学習によってファジィルールのメンバーシップ関数をつくり、第3層が各々のルールの真値を評価する。双方向連想記憶 (BAM) は if-layer と rule-layer、rule-layer と then-layer の間で用いる。手話は主語、目的語、述語の3つの部分に分けられる。実験では、主語には「私」「あなた」「私たち」、目的語には「山」「魚」「花」「種」「川」「海」「ビール」「猿」、述語には「飲む」「摘む」「登る」「泳ぐ」「焼く」「植える」を例として用いた。手話は主語、目的語、述語の順に計算機に入力する。

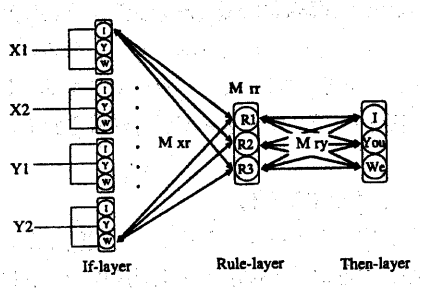

図5 連想記憶システム

主語「私」「あなた」「私たち」を識別する連想記憶システムを図5に示す。

現システムではマッチングを行う際に、動作の空間的な特徴だけを用いるため、手話動作の特徴が互いに近いときには手話の識別が難しくなる。このような場合、動作の特徴と同時に手の形状の特徴についても考慮することが有効であると考える。この際、手の形状の画像情報は、動作の特徴点を検出した瞬間にとる。なぜなら、この瞬間の画像情報が最も特徴的だからである。その手法については次章で説明する。

3.4 手の単純な特徴を得るための画像情報処理

本手法では、手の形状や重心などのような単純な手の特徴についてだけ考える。その主な考えは、2次元の手の画像情報を1次元曲線に変換することである。アルゴリズムは次に述べる通りである。

(a) しきい値「$thres_0$」によって，元の gray-level 画像情報 {img [i] [j]} (i, j = 1, 2, ..., N) を2値画像情報 {bin [i] [j]} (i, j=1, 2, ..., N) に変換する。

$$bin[i][j]=\begin{cases}1, img[i][j] > thres_0\\ 0, otherwise\end{cases} \quad (3)$$

すなわち、gray-level のしきい値には、ここでは経験定数「$thres_0$」を用いる。光の条件が固定的でない場合には、自己適応性のあるしきい値などが好ましい [6]。

(b) 2値画像情報を4方向チェーン-コードに符号化する。

(i_0, j_0) を2値画像情報 {bin [i] [j]} の任意のエッジポイント（端点）と考える。本実験では、あまり詳細な情報は必要としないので、8方向チェーン-コードは使用しない。{bin [i] [j]} のエッジ画像情報を、4方向チェーン-コード（図 6-9）に符号化する。

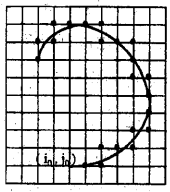

図6 2値画像情報の境界曲線

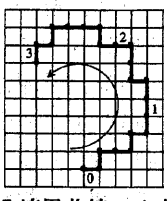

図7 境界曲線の4方向チェーン-コード

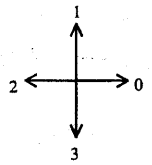

図8 4方向チェーン-コード

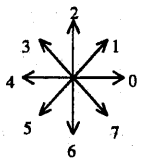

図9 8方向チェーン-コード

$$A_n = a_1 a_2 \cdots a_n,\ a_i \in \{0, 1, 2, 3\},\ (i = 1, 2, \cdots, n) \qquad (4)$$

$a_i a_{i+1}$ $(\forall i \in \{1, 2, \ldots, n\})$ をチェーン-コードの「リング」と呼ぶ。

符号化されたエッジ画像情報を、チェーン-コード曲線と呼ぶ。a_i の水平方向、垂直方向のそれぞれの座標値は coo[i].x , coo[i].y のように示される。

チェーン-コードによって、次のような公式を用いて2値画像情報の重心 (x_c, y_c) を計算する。

$$x_c = M_1^y / S \qquad (5)$$

$$y_c = M_1^x / S \qquad (6)$$

$$S = \sum_{i=1}^{n} a_{i0} y_{i-1} \qquad (7)$$

$$M_1^y = \sum_{i=1}^{n} a_{i0} y_{i-1}^2 \qquad (8)$$

$$M_1^x = \sum_{i=1}^{n} a_{i0} x_{i-1}^2 \qquad (9)$$

a_{i0} の値を表1で示す。

表1 a_{i0} の値

a_i	a_{i0}
0	1
1	0
2	-1
3	0

チェーン-コードの符号と各々の a_i の座標値を得た。

(c) チェーン-コードの4つのコーナーポイントを見つける（図10）。

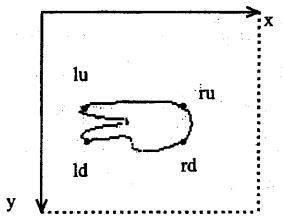

図10 チェーン-コードの4つのコーナーポイト

4つのコーナーポイントはそれぞれ
「lu」:left-up (左-上)、
「ld」:left-down (左-下)、
「ru」:right-up (右-上)、
「rd」:right-down (右-下)
のように定義する。これは次の公式によって得ることができる：

"lu", (coo[i_1].x, coo[i_1].y), which satisfies:
$\text{coo}[i_1].x + \text{coo}[i_1].y = \min_i \{\text{coo}[i].x + \text{coo}[i].y\}$ (10)

"ld", (coo[i_2].x, coo[i_2].y), which satisfies:
$\text{coo}[i_2].x - \text{coo}[i_2].y = \min_i \{\text{coo}[i].x - \text{coo}[i].y\}$ (11)

"ru", (coo[i_3].x, coo[i_3].y), which satisfies:
$-\text{coo}[i_3].x + \text{coo}[i_3].y = \min_i \{-\text{coo}[i].x + \text{coo}[i].y\}$ (12)

"rd", (coo[i_4].x, coo[i_4].y), which satisfies:
$-\text{coo}[i_4].x - \text{coo}[i_4].y = \min_i \{-\text{coo}[i].x - \text{coo}[i].y\}$ (13)

(d) 4つのコーナーポイントにおいて、スタートポイント(手のひらのエッジポイント)を見つける。

例えば図 10 において、「rd」と「ru」は手のひらのエッジポイント(スタートポイント)、「lu」と「ld」は指のエッジポイントである。スタートポイントが複数あるときには任意に1つ選ぶ。

指の幅が手のひらの幅よりも狭いことから、それが指か手のひらかを決めることができる。ここで、ライン走査アルゴリズムを用いることによってその幅を得ることができる。

P_1 を「lu」付近の右下の点とする(図 11 を参照)。水平、垂直方向にライン走査アルゴリズムを活用する。チェーン-コード曲線と水平、垂直方向の交差ポイントの数をそれぞれに数え、どちらか1つが奇数なら、P_1 の代わりに P_1 に近い任意の P_2 を選ぶ。そして再度、同アルゴリズムを活用する。交差ポイントの数が両方偶数になるまでこの手続きをくり返す。水平、垂直方向における全ての交差ポイントの間の距離を計算し、水平、垂直方向のそれぞれの最大距離を得る。小さい方を選び、それを定数 $thres_1$ (指の最大幅を表す)と比較し、もしそれが $thres_1$ より小さければ、「lu」は指のエッジポイントであり、手のひらのエッジポイントは、その他のポイントである。

このようにしてスタートポイントを見付けることができる。

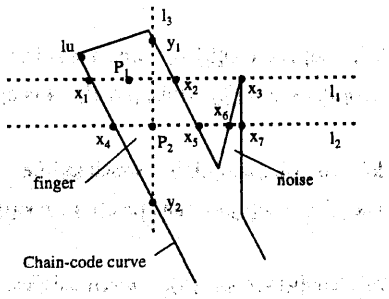

図11 スタートポイントを求める例

P_1 において、水平方向の走査ライン l_1 とチェーン-コード曲線の交差ポイント (x_1、x_2、x_3) の数は3である。そこで P_1 の代わりに P_2 を選ぶ。同様に l_2 との交差ポイント (x_4、x_5、x_6、x_7) の数は4である。x_4-x_5 間の距離、x_6-x_7 間の距離を計算して、前者が水平方向の最大の距離である。同時に、垂直方向の走査ライン l_3 とチェーン-コード曲線の交差ポイント (y_1、y_2) の数は2であり、垂直方向の最大の距離は y_1-y_2 間の距離である。この距離は x_4-x_5 間の距離より大きく、そこで「$thres_1$」と後者を比較する。

(e) スタートポイントから、チェーン-コード曲線上の他の点までのユークリッド距離を計算する(図 12-13)。

局部的に多くの極大、極小を持った距離曲線が得るが、その「小さい山」を削り、「小さい谷」を埋めた「大きい山」の数が指の数となる。

図12 2値画像情報

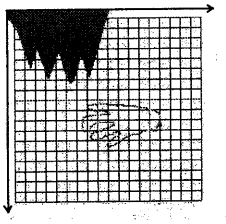

図13 チェーン-コード曲線とユークリッド距離

3.5 上記手法の利点と欠点

(a) 利点

1) 高認識率:手の基本的な6つの形状(0, 1, 2, 3, 4, 5本の指が伸びた状態)について90%以上の高い認識率を示す。2) リアルタイムで処理を行う。3) 雑音の影響を受けにくい。4) 人に依存しない。5) データベースを必要としない。

(b) 欠点

1) 6種類の基本的な形状に限られる:しかし他の特徴と組み合わせて使用することによって、アルゴリズムは有用となる。2) 手は他の皮膚と接触してはいけない:2本の指が接触すると、それは1本の指だと認識してしまう。

4. 医師と患者の手話対話

医療施設などにおいて、インタラクティブに健聴者（医師）とろう者（患者）の会話が行えるようにしなければならない、という問題がある。

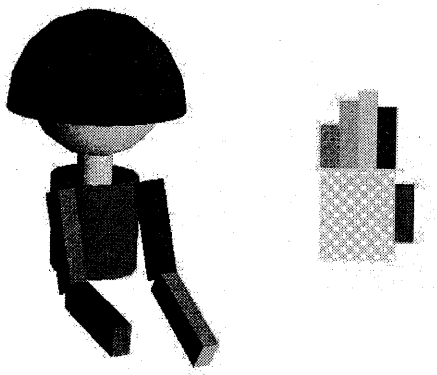

図 14 手話アニメーション（上半身と指）

そこで、手話動作生成機能により、健聴者からの音声や文字入力に対して、図 14 のような手話アニメーションを表示し、反対に、ろう者の手話動作に対して、手話動作を認識した結果を健聴者側に表示することにより、インタラクティブに会話ができるようにした。

5. 実験結果

2 つの手話動作を認識する実験を 5 人について行った。1つの実験は動作の特徴のみ、もう1つの実験は動作の空間的な特徴と手の形状の特徴を組み合わせて実験を行った。現システムでの実験結果の平均認識率は 78.4%、新システムでは 83.9% と認識率は上昇した (図 15 参照)。

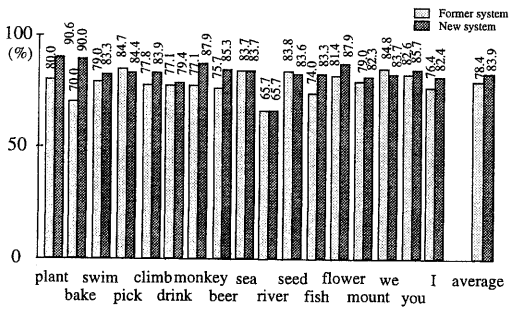

図 15 手話動作の認識率

6. おわりに

本稿では、医療施設等において不可欠な手話インタラクティブインタフェースを構築した結果を述べた。人間の動作の空間的な特徴に加え、簡単な手の形状の特徴にも注目した。この際、注意機能によりどちらの特徴に注目するかを切り替える機能も実現した。さらに、手話アニメーションを表示することにより、健聴者とろう者との間でインタラクティブに会話ができるようにした。

ここで、提案した手話システムとこれまでのシステムでの実験結果の比較によって、単語数は少ないが本手法の有効性を示した。

参考文献

[1] 吉原、山口：動作認識による知的ヒューマンインタフェース、日本ファジィ学会誌 Vol. 7, No. 4, pp. 871-882, 1995

[2] 大森：記号とパターンの結合、日本神経回路学会誌 Vol. 3, No. 2, 65-67, 1996

[3] T. Takagi, A. Imura, H. Ushida and T. Yamaguchi: Laboratory for International Fuzzy Engineering Research

[4] T. Yamaguchi, T. Takagi and T. Mita: Self-organizing control using fuzzy neural-networks, INT J. Control, Vol. 56, No. 2, 1992.

[5] A. Imura, T. Yamaguchi, H. Ushida and T. Takagi: Features of fuzzy associative inference, ELITE-Foundation, EUFIT '94 Promenade 9, D-52076 Aachen.

[6] J. Xu: Image processing and analysis, ISBN 7-03-002502-4/TP. 186, Science Publisher of China, 1992 (in Chinese).