

指文字の認識に関する基礎的研究

島 田 貢 明

(2004年 1 月20日受理)

Experimental Analysis of Feature Quantities for Finger Spelling Recognition

Mitsuaki SHIMADA

Abstract: This research is concerned with finger spelling recognition using neural network. The process of finding out a hand region from an input image and the recognition procedure using neural network are described. The feature quantities of hand region are analyzed experimentally, and are discussed. About the aspect ratio and area ratio which are introduced as external form information of hand region, it proved that they are effective in finger spelling recognition. In order to recognize the shape of a finger spelling with the similar silhouette, the edge ratio is also introduced as internal form information of hand region. It proved by experimental analysis that it is also effective in finger spelling recognition as feature quantities

Key words: 指文字 finger spelling, 特微量 feature quantities,
ヒューマンインターフェース human interface, 画像処理 image processing,
ニューラルネットワーク neural network

1. 緒 言

動作を用いてコミュニケーションを行う手段のひとつに聴覚障害者間で用いられている手話がある¹⁾。手話を理解する人はまだ少ないのが現状であるため、あらゆる場面で手話通訳者が求められている。このため聴覚障害者とのコミュニケーション支援を目指して手話をコンピュータによって認識する研究がこれまで数多く行われてきた^{2),3)}。しかし、手話をコンピュータで認識するには困難な問題が数多く存在するため研究を行う際には様々な制約を設けているのが現状である⁴⁾。手話動作の中でも定義が標準化されているものに指文字があり、日本においてはひらがなやカタカナと同じように音節を表すものである。指文字は、手話において固有名詞や手話単語が確立していない場合たとえばコンピュータ用語の表現などに用い

られている^{5),6)}。本研究では、「あ」から「ん」までの46文字の指文字を対象としている。これまでの研究では、画像認識のためのアルゴリズム研究を行ってきた⁷⁾。そこで本稿では、研究に用いる指文字認識システムおよび認識アルゴリズムにおいて用いている特微量について概説し、加えて基礎的なデータを得るために特微量として手領域の縦横比、面積比および指文字画像のエッジ情報の分布が認識に及ぼす影響を詳細に調べる実験を行ったので、その結果について報告する。

2. 指文字認識システムの概要

本研究において使用する指文字認識システムの処理の流れを図1に示す。

入力する画像のサイズは320×240ピクセルのカラー画像とし、撮影条件については、手とカメラの距離は30～60cm、服装・背景の色は肌の色領

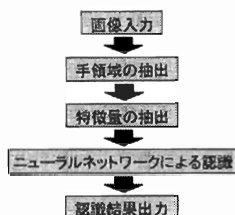


図1. 指文字認識システムの処理の流れ

域でない色とした。入力画像として利用した指文字画像の例を図2に示す。動きを伴う「の」、「も」、「り」、「を」、「ん」の5文字については最後の指形状を入力画像とした。入力された指文字画像のもつ RGB の色情報を色合い (Tint)、鮮やかさ (Saturation)、輝度 (Luminance) で指定する TSL 空間に変換し肌色情報をもつ手領域を抽出した。RGB 表色系のデータ (R, G, B) を TSL 表色系のデータ (T, S, L) に変換する式を次式に示す。

$$T = \begin{cases} \arctan(r'/g')/2\pi + 1/4, & g' > 0 \\ \arctan(r'/g')/2\pi + 3/4, & g' < 0 \\ 0, & g' = 0 \end{cases} \quad (2-1)$$

$$S = [9/5(r'^2 + g'^2)]^{1/2} \quad (2-2)$$

$$L = 0.299R + 0.587G + 0.114B \quad (2-3)$$

表1. 肌色抽出のための条件

Tint	Saturation	Luminance
$0.368 < T < 0.41$	$0.621 < S < 0.79$	$30 < L$

ただし、 r' と g' は次式で表される。

$$r' = R/(R+G+B) - 1/3 \quad (2-4)$$

$$g' = G/(R+G+B) - 1/3 \quad (2-5)$$

肌色の抽出に当たり表1に示す値の範囲を実験的に求めて用いた。

その後、収縮・膨張処理、ラベリング処理を行って手領域の抽出を行った⁹⁾。

抽出した手領域から縦横比、面積比、エッジ比の特徴量を抽出した後、これらの特徴量を用いて、未知パターンに対する汎化能力をもつニューラルネットワークによって指文字画像の認識を行った⁹⁾。現在のところ本システムによる認識実験の認識率は約80%である。認識に用いた階層型ニューラルネットワークの構造例を図3に示す。他に視覚系ニューラルネットワークでの検討もおこなっている¹⁰⁾。ニューラルネットワークに入力する情報は、図に示すような指形状の輪郭情報、縦横比、面積比、エッジ比とし、中間層のユニット数は80、出力層のユニット数は「あ」から「ん」に対応した46とした。

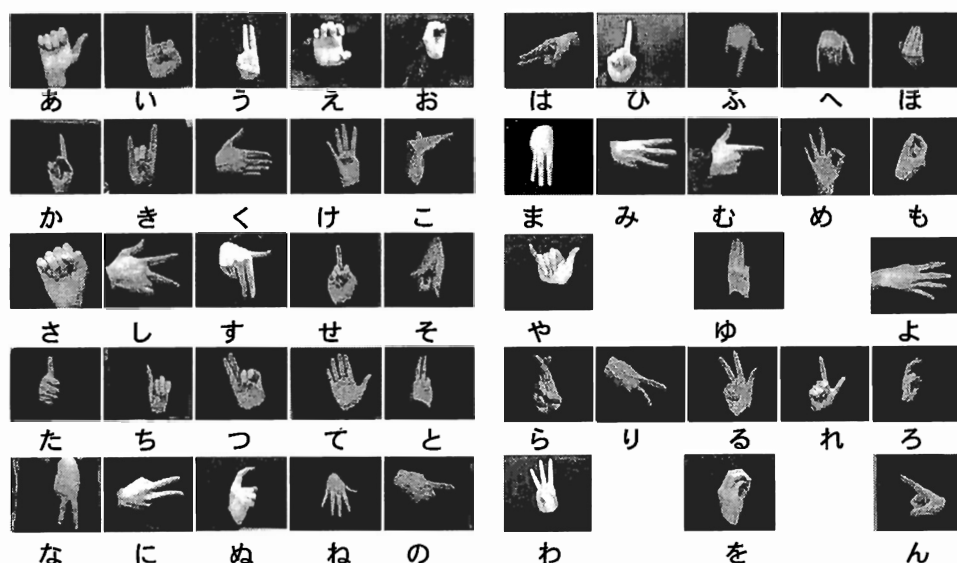


図2. 指文字画像の例

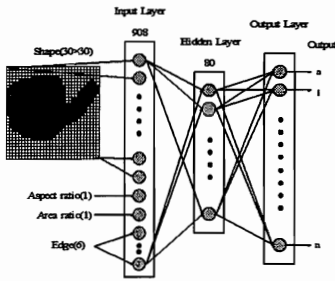


図3. 階層型ニューラルネットワークの構造例

3. 認識に用いる特徴量について

認識には、指の輪郭情報に加えて、外部形状情報の特徴量として図4に示すような縦横比・面積比を用いる。これらの特徴量を認識に用いる理由は指のシルエット画像いわゆる指の輪郭情報だけでは非常に似ているものがあり、それらを区別することが困難であるからである。指の形状としては、手を握った状態のもの、手を開いた状態のものがあり、手の外形情報を特徴づける値として手領域の外接方形の縦横比と外接方形の内部に手領域が占める割合を表す面積比が有効と考えられる。

しかし、手の外形情報だけでは指形状が非常に似ている場合にはその区別が困難であるので、この問題を解決するために加えて内部形状情報の導入をおこなった。内部形状情報の候補としては、手領域の輝度値の分布が考えられるが、撮影場所など影響を受けやすいため適さない。そこで、光環境に影響されにくいエッジ情報を特徴量として用いた。図5にエッジ情報の抽出例を示す。垂直

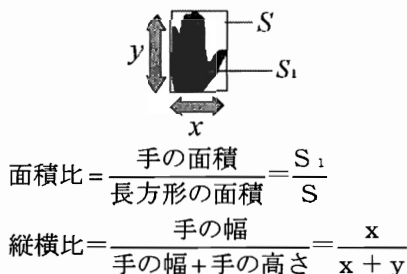
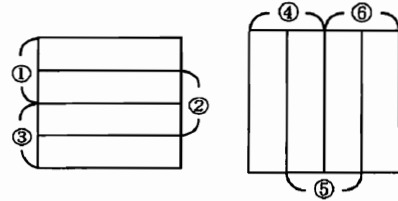


図4. 面積比・縦横比の定義



(a) エッジの表示例



(b) 画像の分割



(c) エッジ比の計算画面例

図5. エッジ情報の抽出例

エッジ成分と水平エッジ成分は抽出された手領域をもとに次式を用いて求められている。

$$\Delta X(i, j) = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad (4-1)$$

$$\Delta Y(i, j) = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (4-2)$$

図5(a)は手領域において求められたエッジの例を表している。図より外部情報では得ることの出来なかった手の内部情報がエッジとして検出されていることがわかる。すなわち、エッジ比を調べることにより指が折り曲げられているのか伸ばされているのかを知ることができる。図5(b)は画像情報を6分割する方法を示している。図5(c)は、画像情報より図5(b)で示した6つの領域のエッジ比を次式

$$\text{エッジ比} = \frac{\text{垂直エッジ成分}}{\text{垂直エッジ成分} + \text{水平エッジ成分}}$$

により算出した画面例を示している。表示されている数値は、上から順に①～⑥の分割領域に対応している。エッジ比は式からわかるように、エッジ全体の中で垂直エッジが占める割合を表す。

4. 実験結果

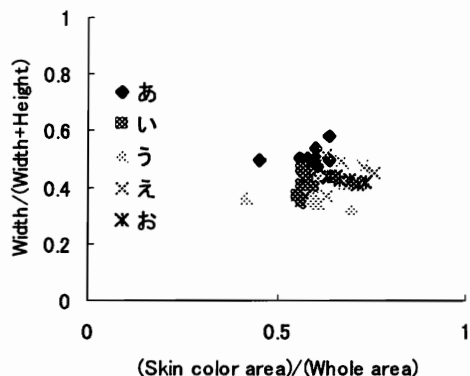
図6に「あ」、「い」、「う」、「え」、「お」の5文字について、各10種類の画像データを、縦軸に縦横比を横軸に面積比をとりグラフ化した結果を示す。手を握ったような手形状の場合には縦横比の値は小さくなり、逆に指を伸ばしたような手形状の場合にはこの値は大きくなる。面積比については、手を握ったような手形状の場合には外接方形の中に占める肌色の割合が大きくなるため面積比の値は大きくなる。逆に大きく指を広げた手形状の場合にはこの値は小さくなる。図からわかるように、これらの値は文字ごとに特徴をもって分布しており認識に用いる特徴量として有効であることがわかる。

次に、認識実験の過程において、お互いに誤認識されやすいグループとして「い」、「く」、「ち」、「む」（以下「いくちむ」と「う」、「け」、「せ」、「た」、「と」、「ひ」、「ら」、「ろ」（以下「うけせたとひらろ」）の2グループがあるので、その特徴量を比較してみた。その結果を図7に示す。図7(a)は「いくちむ」のグループの面積比・縦横比の分布を示している。図からわかるように面積比では「む」は分離が比較的容易であるが縦横比では相互に分離が困難なことがわかる。図7(b)は「うけせたとひらろ」のグループの面積比・縦横比の分布を示している。図より面積比のほうが縦横比より分離が比較的容易ではあるが、いずれもオーバーラップしている部分が多く縦横比・面積比の外部形状情報を表す特徴量では分離が容易ではないことがわかる。

次に指文字画像の内部形状情報をあらわす特徴量として導入したエッジ比について同様な実験をおこなった。その結果を図8に示す。図8(a)は、「いくちむ」の画像よりエッジを検出した結果を画像で示す。この画像より、手の内部の情報を知る

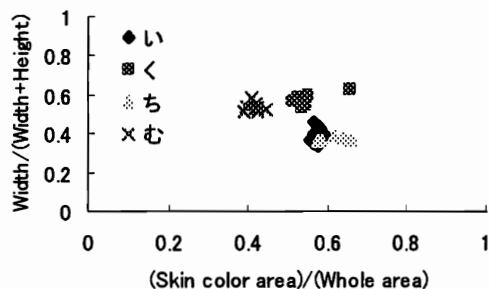


(a) 指文字画像

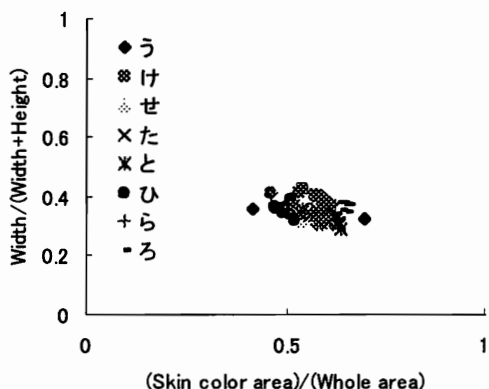


(b) 縦横比・面積比の分布例

図6. 指文字画像と縦横比・面積比の分布例

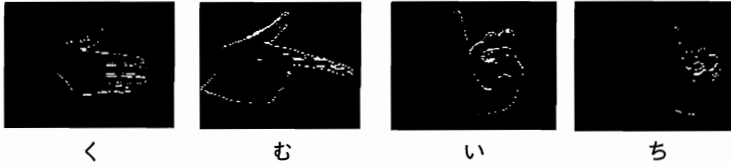


(a) 「いくちむ」の面積比・縦横比の分布

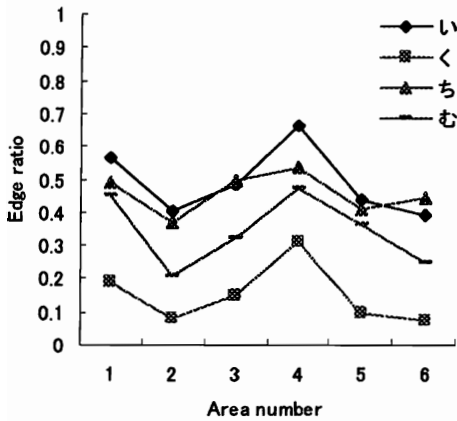


(b) 「うけせたとひらろ」の面積比・縦横比の分布

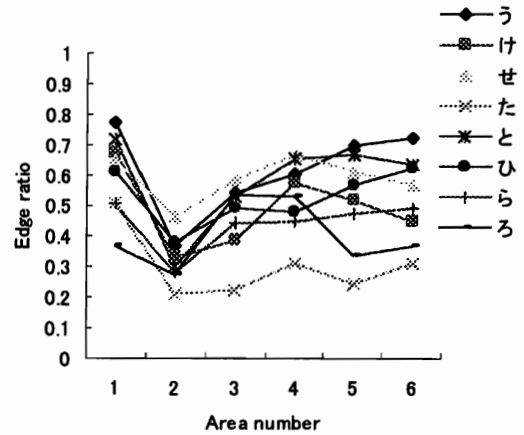
図7. 誤認識されやすいグループの分布例



(a) エッジ抽出画像例



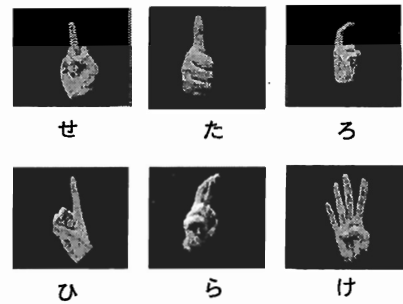
(b) 「いくちむ」におけるエッジ比の分布



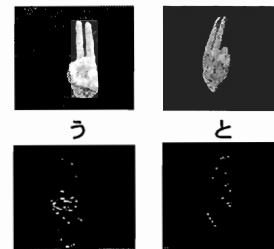
(c) 「うけせたとひらろ」エッジ比の分布

図 8. エッジ抽出画像とエッジ比の分布

ことが出来るが、「くむ」、「いち」のグループが非常に似ていることが画像よりわかる。そこで、前節で述べたように手領域を 6 分割し、それぞれの部分についてエッジ比を計算しグラフ化して比較した。図 8(b)は「いくちむ」のグループの結果を示す。グラフ縦軸はエッジ比であり、横軸は分割した領域の番号である。この値は、10個の画像より求めた値の平均値である。このエッジ比は垂直エッジ成分が垂直および水平エッジ成分の和に占める割合を表しているの、画像に垂直エッジ成分が多く含まれる場合にはその値は大きくなり、逆に水平エッジ成分が多く含まれる場合にはこの値は小さくなる。図より「く」の画像には水平エッジ成分が多く含まれており「い」の画像には垂直エッジ成分が多く含まれていることがわかる。グラフより、先ほど述べたとおり「くむ」はグラフの傾向がよく似ているが値が異なるため比較的容易に分離可能である。しかし、「いち」については、グラフの値が非常に似ているためエッジ比では分離が困難であることが予想される結



(a) 比較的認識されやすいグループ



(b) 比較的認識が困難なグループ

図 9. エッジ情報の特徴によるグループ分け

果を得た。同様の検討を「うけせたとひらろ」のグループにもおこなった。グラフからわかるようにいずれも②の領域の値はほぼ同じ値を示しているが、他の領域の値を比較することで「せたらひらけ」のグループは、「いくちむ」のグループ程ではないが、比較的分離可能であることがわかる。しかし、「うと」については、グラフの形状も非常に似ており分離が困難であることが予想される。以上エッジ比のグラフよりわかることについて述べたが、このことを実際の指文字画像で再確認をおこなった。エッジの特徴によるグループ分けを図9に示す。図9(a)は比較的認識されやすいグループの画像であり図9(b)は比較的認識が困難なグループの画像とエッジ画像を示している。図から指の折り曲げ方など不安定な要素はあるもののエッジ比を導入することで輪郭形状が非常によく似た指文字も認識可能とすることができることが確認された。

5. 結 言

本論文では、ニューラルネットワークを用いた日本語カナを表現する指文字画像の認識方法について概説し、手形状の輪郭情報だけでは認識が困難な指文字画像の認識に内部形状情報と外部形状情報の特徴量を導入する方法について述べた。入力画像より抽出された手領域の外部形状情報として縦横比・面積比を内部形状情報としてエッジ比を定義し、この特徴量の分布が文字ごとに固有に分布していることを実験的に明らかにし、認識に有効であることを示した。また、指文字を輪郭情報が似ているなどの原因から、これまでの研究において相互に誤認識しやすいことがわかっている指文字2グループについて取り上げ、それぞれのグループ内の各指文字の縦横比・面積比およびエッジ比の分布を考察することでこれらの特徴量が認識に有効であることが確認できた。

今後の課題としては、画像における個人差や環境の影響を受けない認識手法の確立および動画認識へ適用するためのアルゴリズムの研究などがあげられる。

謝辞

本研究の実施にあたり福井仁愛学園後援会の特別研究助成より援助を受けた。また、実験のデータ収集にあたり本研究室所属の卒業研究の学生の協力を得た。ここに記して感謝の意を表する。

参考文献

- 1) NPO 手話技能検定協会：実用手話辞典、新星出版社、2002
- 2) 今川和幸、呂 山、猪木誠二、松尾英明：肌色領域により隠れて見える場合を考慮した手話動画像からの手の実時間追跡、信学論、J81-D-「II」、No.8、pp1787-1795、Aug.1998
- 3) Marcus Vinicius LAMAR, Md. Shoaib BHUIYAN, Akira IWATA: Hand Gesture Recognition Using T-CombNET: A New Neural Network Model, IEICE TRNS. INF.&SYST., VOL.E83-D, No.11, Nov 2000
- 4) 渡辺賢、岩井儀雄、八木康史、谷内田正彦：カラーグロブを用いた指文字の認識、信学論、J80-D-「II」、No.10、pp2713-2722、Oct.1997
- 5) 谷千春、井崎哲也：みんなの手話、NHK 出版、2002
- 6) 長谷川洋：コンピュータ用語と手話（手話の世界）、マガジンサポート、pp74-77、2002
- 7) M.Shimada, S.Iwasaki, T.Asakura : Finger Spelling Recognition using Neural Network with Pattern Recognition Model, Proceedings of SICE Annual Conference 2003, pp2176-2181, Aug.4-6, 2003
- 8) 安居院猛、長尾智晴：画像の処理と認識、昭晃堂、2002
- 9) 島田貢明、岩崎聡、朝倉俊行：ニューラルネットワークを利用した指文字認識、第45回自動制御連合講演会概要集、pp393-394、Nov.26-27, 2002
- 10) 計測自動制御学会：ニューロ・ファジィ・AIハンドブック、オーム社、1994