

アルファベット指文字の肌色抽出と認識に関する基礎的研究

島田 貢明

(2011年2月10日受理)

Experimental Analysis of Skin Color Detection and Automated Recognition for Finger Spelling

Mitsuaki SHIMADA

Abstract : This research is concerned with automated recognition of sign language. In this paper, new algorithm using Neural Network which recognizes finger spelled word of American Sign Language (ASL) is proposed. Experiment is also conducted using actual video of finger spelled words. This is a challenging task since the ASL alphabets are expressed by the various figure of one hand. The captured input image in RGB is converted to HSV color space, because it is more related to human color perception. Automated recognition experiment using proposed algorithm is done to 26 characters of ASL finger spelling. According to the result of experimental analysis, it is proved that the algorithm proposed in this paper is effective in finger spelling recognition.

キーワード key words

指文字 finger spelling, 手話 sign language, 肌色抽出 skin color detection,

画像処理 image processing, 画像認識 image recognition, ニューラルネットワーク neural network

1. 緒言

手話は聴覚障がい者とのコミュニケーションを行う手段のひとつであり¹⁾、講演会などの集会においては手話通訳者が求められているが、その人数はまだ少ないのが現状である。このため聴覚障がい者のコミュニケーション支援を目的として手話認識の研究がこれまで数多く行われてきた^{2), 3), 12)}。しかし、手話の認識をコンピュータで行うためには、まだ解決すべき困難な問題も多く、様々な制約を設けて研究がおこなわれている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る。指文字は、ひらがなやカタカナと同じように音節を表すために使用されるもので、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などに用いられている^{5), 6)}。これ

までの研究では、「あ」から「ん」までのひらがな46文字の指文字を対象として、画像認識のためのアルゴリズム研究及び認識に有効な特徴量の検討を行ってきた^{7), 8)}。しかし、近年コンピュータ用語に限らず様々な分野でアルファベットの用語が用いられるようになっている。そこで今回は、

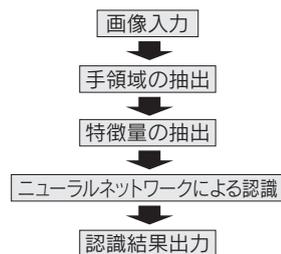


図1. 指文字認識システムの処理の流れ

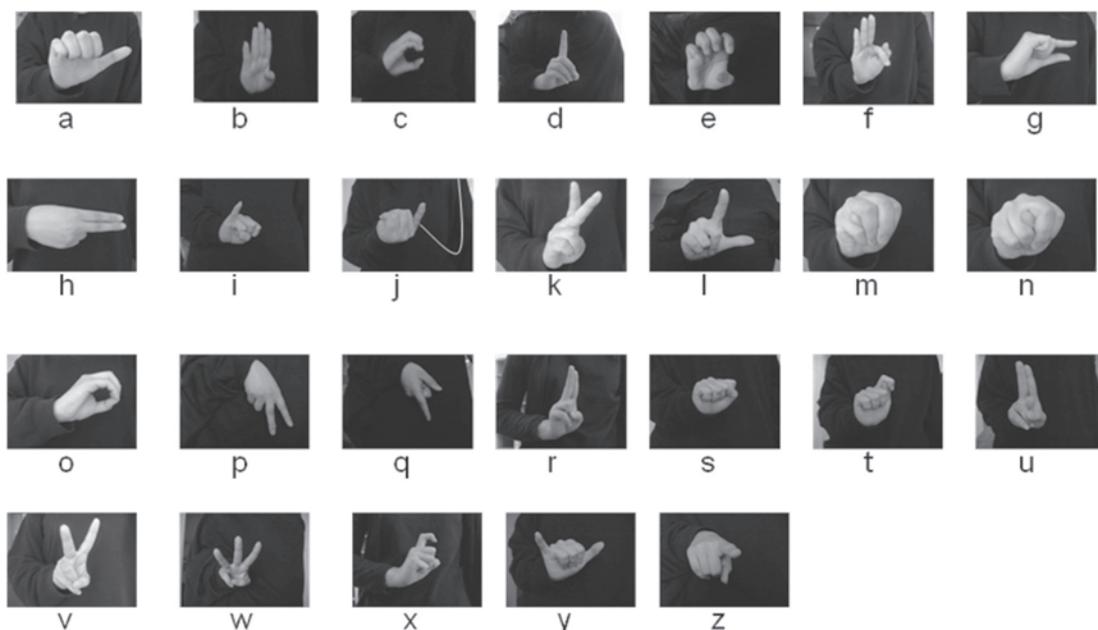


図2. アルファベット指文字 (ASL) の例

新たにアルファベットの指文字としてAmerican Sign Language (ASL)¹³⁾の指文字を対象とした認識実験を行うこととした。今回使用する指文字認識システムは、ひらがな指文字の研究で使用した基本部分を改良している。

本稿では、特に特徴量を正しく抽出し認識を安定的におこなうために不可欠となる手領域の抽出部分に着目し、入力画像から手領域を抽出するための肌色情報について、基礎的なデータを得るための実験的な検討とアルファベットの指文字について認識実験を行ったので、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。処理の概要については後述するが、今回画像の入力は、320×240のカラー動画画像とし、その画像をキャプチャすることで入力画像としている。実際にはUSBカメラにより被験者の画像を直接取り込んで認識実験を行うが、今回は再現性を考慮してビデオ画像を用いた。

アルファベットを表現する指文字には



図3. 画像処理プロセスの例

American Sign Language (ASL) 及びBritish Sign Language (BSL)があるが、BSLは両手で各文字を表現するものであり、日本語の手話で使用される指文字はASLを参考としていることから、本研究においてはASLを用いることとした。入力する画像の背景画像については特別の制約を設けてはいないが出来るだけ手領域を大きく写すこととした。入力画像として利用したアルファベットの指文字 (ASL) 画像の例を図2に示す。「j」、「z」については動きをとまうため今回は動作終了後の画像を学習用画像とした。

その後、入力画像は収縮・膨張処理、ラベリング処理を行って手領域を抽出する⁸⁾。そのプロセス画像の例を図3に示す。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

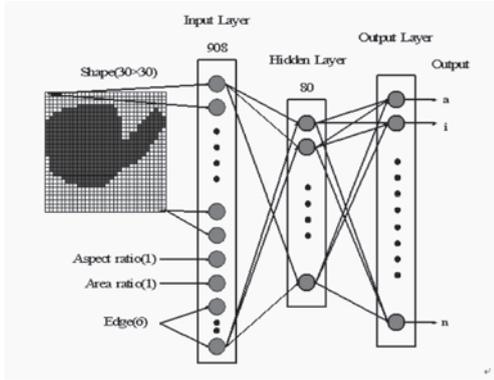


図4. 階層型ニューラルネットワークの構造例

汎化能力をもつ図4に示す階層型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⁾。ニューラルネットワークに入力する情報は、アルファベット各文字の指形状の輪郭画像情報、縦横比、面積比、エッジ比とし、出力層のユニット数は「a」から「z」に対応した26個とした。

3. 手領域抽出のための色情報の検討

これまでの、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表色系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。

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

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

$$L = 0.299R + 0.587G + 0.114B$$

ただし、 r' と g' は次式で表される。

$$r' = R/(R + G + B) - 1/3$$

$$g' = G/(R + G + B) - 1/3$$

表1. 肌色抽出のための条件

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

TSL表色系においては、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。

今回は、肌色抽出のための表色系をHSVとする検討実験を行った。画像は上述の通りRGB値で得られるが、人間の感覚に近いのは、色相 (Hue)、彩度 (Saturation)、明度 (Value) で表現されるHSV表色系であるといわれている。RGB表色系からの変換式を以下に示す。

$$MAX = \max(R, G, B)$$

$$MIN = \min(R, G, B)$$

$$V = MAX$$

$$V = 0 \text{ の場合}$$

$$S = 0$$

$$H = 0$$

$$V = 0 \text{ 以外の場合}$$

$$S = (MAX - MIN) / MAX$$

$$Cr = (MAX - R) / (MAX - MIN)$$

$$Cg = (MAX - G) / (MAX - MIN)$$

$$Cb = (MAX - B) / (MAX - MIN)$$

$$H = \begin{cases} Cb - Cg & R = MAX \\ 2 + Cr - Cb & G = MAX \\ 4 + Cg - Cr & B = MAX \end{cases}$$

$$H = 60 \times H$$

$$H = H + 360 \quad H < 0$$

肌色抽出ではH、S値を色相環にプロットした場合において、Hの範囲は6～38であることが広く知られていることから、今回は入力画像がこの範囲にあるかを実験的に調べてみた。この実験では、画面に表示されている画像の任意の位置のRGBの値を、表色系TSL、YIQ、HSVの各値に変換して表示するソフトウェアを用いて、認識に使用した入力画像の背景と手領域の色情報を実験的に検討した。その色情報の実験における表示画面の様子を図5に示す。このHSV表色系を用い



図5. 色情報の表示画面

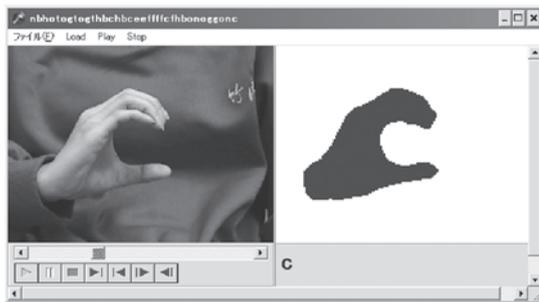


図6. HSV表色系による手領域の検出例

実際の手領域の検出状況を図6に示す。この例は、左側のプレビュー画面において入力された動画像を表示しておりその下部には再生に関するボタンがある。右側の画面では動画からキャプチャした画像をもとに検出された手領域の画像が表示されている。その下部には認識システムの結果が表示されており、図は「C」の手形状に対し正しく「C」の文字を認識している例である。

本実験においてもHは前述した値の範囲と同様な結果を確認することができた。この結果により、手領域の分離にHSV表色系を用いることで、USBカメラによるリアルタイムの画像入力時により安定的な手領域検出が期待できる。

4. アルファベット指文字の認識実験

次に、本システムを用いてアルファベットの指文字を認識する実験を行った。入力は、前述したように、a～zまでの指文字を動画像で各10文字ずつ入力した結果、認識率は表2のようであった。

表2. 認識実験の結果

認識率	アルファベット
100%	f, g, h, i, l, o, p, q, s, t, v, w, x, y, z
90%	a, b, k, r, u
80%	c, e,
70%	n
60%	d
50%	j
20%	m

表3. 認識実験における誤認識の傾向

認識率	アルファベット			
90%	a	g1		
	b	w1		
	k	f1		
	r	k1		
80%	c	o1	t1	
	e	z2		
70%	n	s2	w1	
60%	d	u2	k1	l1
50%	j	a2	z3	
20%	m	n2	s4	e2

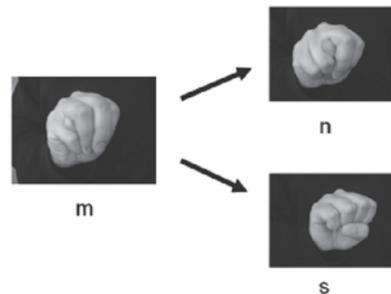


図7. 「m」の誤認識しやすい文字

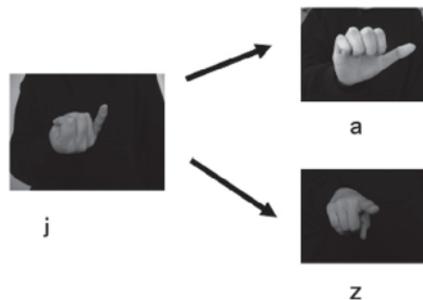


図8. 「j」の誤認識しやすい文字

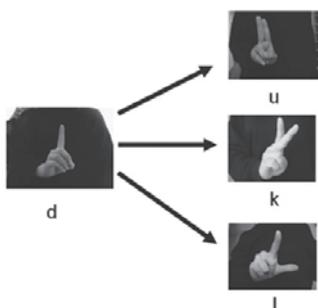


図9. 「d」の誤認識しやすい文字

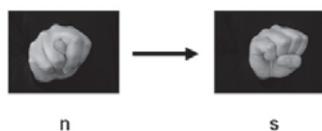


図10. 「n」の誤認識しやすい文字

表に示すように26文字中20文字が90%の認識率を得ることができ良好な認識結果が得られたが、「m」については認識が困難であることがわかった。表3に実験の結果どのような文字として誤認識したのかを示す。表中s4とあるのは「m」を「s」と誤認識した数が4であったことを表している。今回最も誤認識の数が多かった「m」を「n」、「s」と比較した図を図7に示す。図から明らかであるように3文字とも今回の認識システムにおいて扱っている特徴量のうち外接方形から得られる縦横比、面積比からは区別しにくい形状であることがわかる、唯一エッジ比のみがこの3文字を区別するための特徴量であるため認識率を上げるためには入力画像の解像度を上げることで認識率の向上が期待できる。

次に認識率が50%と低い「j」の誤認識しやすい指文字を図8に示す。今回は「j」について動作の終りの指形状を使用したために誤認識したことがあげられるので、今後は、動作のある指文字に対しては、他の文字とは異なる認識プロセスを設けることにより認識率の改善が期待できる。認識率60%の「d」については、図9に示すように指の形状が「u」、「k」、「l」いずれも人差し指が立っている縦長のシルエットであるので、特徴量のうち外接方形から得られる縦横比、面積比か

らは区別しにくい形状であることがうかがえる。「m」同様入力画像の解像度を上げることでエッジ比が文字の区別に有効となり認識率の向上が期待できる。認識率70%の「n」については、図10に示すように「s」の指形状との違いは親指の位置のみでありこれも前述したとおり「d」と同様の改善により認識率の向上が期待できる。なお、今回実施した認識実験の認識率は全体で約88.8%であった。

5. 結 言

本論文では、ニューラルネットワークを用いたアルファベットを表現する指文字画像の認識方法について概説し、認識に用いる特徴量である縦横比、面積比、エッジ比を安定的に得るために不可欠となる手領域の抽出部分に着目し、入力画像から手領域を抽出するための肌色情報について実験的な検討を行った。入力画像より手領域を抽出するためにこれまではTSL表色系を用いていたが、今回はより人間の感覚に近いといわれているHSV表色系の値を用いることとし、肌色検出のための確認実験を行った。実験の結果、HSV表色系を肌色抽出に用いることで、安定的な手領域の抽出に有効であることが確認できた。

アルファベットの指文字認識実験の結果については認識率の低い指文字の形状について考察を行った結果、形状が酷似している特定の文字については区別することが困難であることが分かった。特に「m」については20%と26文字中最も低い結果となった。

今後の課題としては、入力画像の解像度を上げることで認識率を改善すること、動作をともなう指文字の認識手法の導入、CCDを用いたりリアルタイムの画像入力の際に画像における個人差や環境の影響を受けない動画像認識アルゴリズムの研究などがあげられる。

謝辞

本研究における実験のデータ収集及び実施にあたり本学学生の多大なる協力を得た。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：カラーグロー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) 島田貢明：指文字の認識に関する基礎的研究、仁愛女子短期大学紀要、第36号、pp9-14、2003
- 9) 安居院猛、長尾智晴：画像の処理と認識、昭晃堂、2002
- 10) 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
- 11) 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
- 12) P. Buehler, M. Everingham, A. Zisserman : Learning sign language by watching TV (using weakly aligned subtitles) , Proceedings of IEEE Conference on CVPR 2009
- 13) 大杉豊 (全日本ろうあ連盟):国際手話のハンドブック、三省堂、2002