

## ニューラルネットによる静止画像からの指文字認識

平山 亮<sup>†</sup> 舟川 政博<sup>‡</sup>

金沢工業大学<sup>† ‡</sup>

### 1. はじめに

聴覚障害者のコミュニケーション補助を目標として、指文字認識の方法を検討した。ウェブカメラでキャプチャした 2 方向からの画像を用い、日本語の指文字の認識、及び、実験を行った。画像から肌色情報を使って手領域を抽出し、2 値化してシルエット情報をした。この画像情報を入力とし、指文字を出力とする、3 層パーセプトロンのニューラルネットを構成し、学習をさせた後、それを用いて認識を行った。動きのある指文字については、連続する複数の静止画像を用いて、同様の方法での認識方法を検討した。また、この認識アルゴリズムを用いた、指文字認識のシステム化について提案する。

### 2. 指文字画像認識方法

#### 2.1 指文字画像

認識対象とする日本語指文字は、動きのない指文字 41 種類、動きのある指文字 4 種類（「の」「も」「り」「ん」）、手指のシルエット形状は変化せず手指全体が左右、上下、前後に動く濁音・半濁音・拗音 34 種類、及び、手指のシルエット形状が変化しつつ手指全体が動く長音「一」1 種類である。ただし、本報告では今までのところ「の」「も」「り」「ん」「一」については試行錯誤の実験中であり結果は出ていない。

指文字画像は 2 台の Web カメラを用いてキャプチャする。既発表のカメラ配置の検討<sup>[1]</sup>において最良の結果であった配置、すなわち、右利きの場合で、指文字者から見て正面やや左寄り、及び左真横の 2 か所から撮影する。肌色成分によって手領域を抽出する際の誤認識を減らすため、実験では、青い長袖の服を着用した。図 1 に入力画像例を示す。

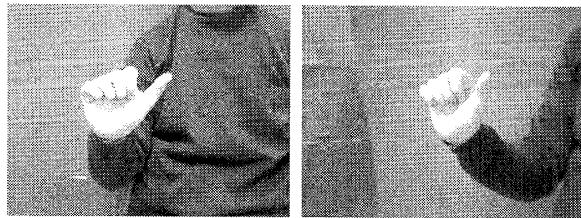


図 1 Web カメラからの 2 眼入力画像

#### 2.2 特徴抽出

入力画像は、肌色の領域情報を使って、手領域の切り出しを行う。これまで行ってきた方法<sup>[2]</sup>と同様に、画素の輝度が  $R > G > B$  の領域を使う。切り出した領域の  $R$  値を使ってグレースケール画像を作成し、続いて、その明暗のしきい値を設定して 2 値化を行い、これをシルエット画像とし、シルエット画像を含む矩形領域である外接長方形（フェレ径）を抽出する。次に、各静止画像について、以下の 6 つの種別で計 638 個の特徴量<sup>[1]</sup>を算出する。4)～6) は外接長方形を  $25 \times 25$  画素に縮小したシルエット画像（図 2）から算出する。

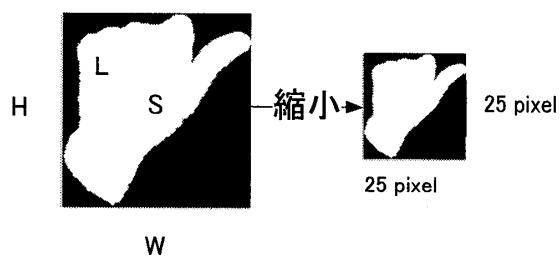


図 2 シルエット画像からの特徴抽出

##### 1) 円形度 (1 種類)

$$C = \frac{4\pi S}{L^2} \quad S: \text{シルエット面積} \quad L: \text{周囲長}$$

##### 2) 縦横比 (1 種類)

$$A = \frac{H}{H+W} \quad H: \text{外接長方形高さ} \quad W: \text{幅}$$

##### 3) シルエットと外接長方形との面積比 (1 種類)

Finger Alphabet Recognition from still images using neural network

<sup>†</sup>Makoto J. Hirayama, Kanazawa Institute of Technology

<sup>‡</sup>Masahiro Funakawa, Kanazawa Institute of Technology

$$D = \frac{S}{H \times W}$$

- 4) 縦を 5 分割し、その矩形内にあるシルエットと矩形との面積比 (5 種類)  
 5) 横を 5 分割し、その矩形内にあるシルエットと矩形との面積比 (5 種類)  
 6) 縮小シルエット画像の画素値 (25x25=625 種類)

### 2.3 ニューラルネットワークによる画像認識

特徴抽出したデータはニューラルネットワークモデルの 3 層パーセプトロン（図 3）により学習させ、認識を行う。入力は前記の抽出した 2 方向からの画像の特徴量（合計 1276），出力は指文字の手指形状の種類である。中間層は 300 ユニットとした。入力は静止画像を対象とし、濁音・半濁音・拗音・長音の手全体の動きについては、カメラ画像に対する手領域（肌色領域）の重心座標の移動により検出する。動きのある指文字である「の」「も」「り」「ん」及び長音「一」については、動き出す前の形状、動いた後の形状、その中間地点での形状をそれぞれ静止画像とし、それぞれ独立に別の形状としてニューラルネットワークで認識したあと、動き出す前、中間地点、動いた後の順に認識された一連の区間を条件分岐により判別して動きのある指文字の認識結果とする。

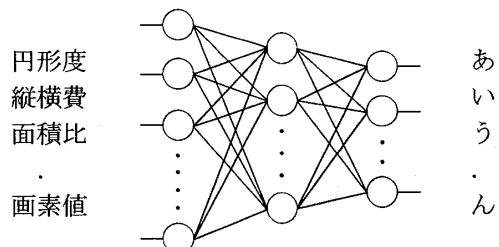


図 3. 指文字静止画像認識ニューラルネットワーク

10 名の被験者から撮影したデータを用いて学習させ、認識実験を行ったところ、静止画像から指文字形状の認識率おおよそ 85% であった。また、濁音・半濁音・拗音・長音の手全体の動きの動作判定<sup>[1]</sup>は、それぞれ 88%，95%，90%，95% の認識率である。

### 3. 実時間指文字認識システムの開発

指文字認識をパソコンのユーザインタフェースとして用いること、手話－日本語翻訳システムの一部として用いること、チャットや拡張現

実感通信システムでのコミュニケーション方法として用いることなどを想定して、提案したアルゴリズムのパソコンへの実装を進めている。又、災害時や外出時のコミュニケーション機器として、携帯情報端末への実装を検討している。アルゴリズム開発及びニューラルネットワークモデルの学習は数値処理ソフト (MATLAB) の環境を使用して行ったが、実時間指文字認識システムの実装は、Windows 環境で、C++ (Visual Studio 2008)，ウェブカメラからのデータ取り込み及び基本的な画像処理には画像処理ライブラリ OpenCV を利用している。

アルゴリズム開発においてはあらかじめ切り出された静止画像を用いたが、稼働システムでは、Web カメラの動画像から認識対象の静止画を実時間で切り出してから認識する必要がある。複数の指文字を連續して表現するとき、一つの指文字から別の指文字に移るあたり部ではシルエット形状が変化するが、移ったあとは短時間その指文字形状を保持する、つまり、シルエット形状の変化がごく小さいことに着目して、抽出する。肌色情報を使って抽出して 2 値化した手のシルエット領域のフレーム間の差分を算出することによって行う方法を実装中である。

### 4. おわりに

今後の課題としては、実時間指文字認識システムの実装を完成し稼働させること、及び、認識率の向上である。特に、形状が類似のため誤認識がおこりやすい文字（「い」と「ち」など）に関する対応、認識対象の時間領域の切り出し（セグメンテーション）をよい精度でできるようにすること、着衣の色の制限や手指形状のばらつきを吸収できるように耐性を高くすることを、検討していきたい。

### 謝辞

本研究の一部は、文部科学省のハイテク事業による私学助成を得て行われた。

### 参考文献

- [1] 木元文章, 舟川政博, 平山亮, “2 方向のシルエット画像による指文字認識精度の向上と有用なカメラ配置の検討,” 第 49 回情報処理学会全国大会, 4T-8, pp. 2-422-423, 2009.
- [2] 舟川政博, 平山亮, “指文字画像からの手指形状特徴量抽出方法の検討,” 第 5 回情報科学フォーラム 2006, pp. 561-562, 2006.