

2D-8

コンピュータによる手話単語の認識*

細川 昌也 浅見 賢 国井 利泰†
 東京大学 理学部‡

概要

手話は耳の不自由な人や話すことができない人にとっては有用なコミュニケーション手段である。手話においては、どの動作パターンが何を表しているかあらかじめ決められており、手話の受け手は相手の動きをそれらと照らし合わせて意味を判断している。こうした手話の認識をコンピュータで行うことができれば、テレビ電話における手話での通信をとっていても、わざわざ大量の画像データを送らなくても認識した結果をパラメータ化して送れば通信スピードがより速くなるし、キーボードからの入力や音声による入力以外に新たな入力手段としても有効である。現在はまだ地域差のある手話も、テレビ電話で手話が使えとなれば、テレビ電話で使える手話をいわゆる共通語として普及させることも可能となる。これまで、指文字の認識や手話のパラメータの検討さらには右手だけをつかう簡単な手話単語の認識がされてきている。この論文では、より複雑な手話単語の認識を目的とする。ただし、指文字の認識は今後の課題とする。

1 方法

手話を構成する要素としては、次の三つが挙げられる。

- (1) 手の形
- (2) 手の位置
- (3) 手の動き

手話動画像を認識するには、それから一定の時間間隔ごとの静止画像を取り出して上記の三つの要素がどうなっているか調べるのが一般的な方法だと考えられる。しかし、手話のように動作パターンが決まっているならば、それぞれの静止画像について要素全部を調べなくてもかなりのところまで手話単語を識別できると考えられる。そこで、この論文では、一定の時間間隔ごとの静止画像について、顔と両手の占める部分に注目してパターン化し、その時間的変化を追うことによってどの程度まで手話単語を識別できるかを調べる。

2 システムの概要

2.1 パターン単純化

この論文において使用した手話単語の動画像には、(財)全日本ろうあ連盟が企画、編集し、京都市聴覚障害センターが製作したビデオ「新しい手話1」(カラー、VHS)を使用する。システムにおいて使用するパターンは、一定の時間間隔ごとに取り出された静止画像を、顔と両手の占める部分を1にそれ以外の部分を0に変換したものである。本来ならば領域検出法を使うべきところであるが、手話者が黒色の長袖の服を着ていて、背景が無地で顔や手とは違った色(青色)のこのビデオを使用することで、画像のパターン化を簡略化することができる。ここでは、背景の色を黒色に変換し、残りの色を全て白色に変換することでパターンを一層単純化することに成功した。手話者の服装や背景の色などを工夫することは、パターン化の効率を上げるのに重要である。

2.2 小システム

小システムは、あらかじめ幾つかのパターンを持っていて、外部から入力としてパターンを与えられると、自分の持っているパターンのなかでそれに一番近いものを出力する働きを持っている。あらかじめ持っているパターンで学習済みのパターン情報ネットに与えられたパターンを入力して一番近いパターンを割出す方法をとる。システムはこれらの小システムの集合体とこれを管理するコントローラーからなる。

2.3 コントローラー

2.3.1 手話単語とパターンの管理

手話単語は、時間とともに変化する一連のパターンと対応している。コントローラーは、それらのパターンを手話単語の動画像のスタートからの経過時間で分類して木構造で管理している。例えば、手話単語の動画像から静止画像を取り出す一定の時間間隔を h とし、そうした場合に静止画像が三つ取り出せるとするとそのパターンは図1のようにコントローラーによって管理される。更にコントローラーは図2のようにパターン群を小システ

*Computer Recognition of Words in a Sign Language

†Masaya Hosokawa, Satoshi Asami, Toshiyasu L. Kunii

‡The University of Tokyo

ムに割り当てている。手話単語を新規登録したいときは、時間とともに変化するその手話単語の一連のパターンを木構造に追加すればよい。ただし、その際は、小システムは学習し直したり、新たな小システムを追加する必要性が生じる。

2.3.2 手話単語の認識過程

具体的に図2を使って説明する。認識したい手話単語の動画像から一定の時間間隔 h で取り出した静止画像をパターン化したものを、手話単語の動画像のスタートからの経過時間順に、パターン1、パターン2、パターン3とする。コントローラーは、パターン1を小システム1に渡し、その結果得られたパターンとつながっている小システム3にパターン2を渡す。次に、小システム3が出力するパターンとつながっている小システム8にパターン3を渡す。パターン3で入力するパターンは終りなので、小システム8からパターンが出力されるとコントローラーは、それまで木構造を辿ってきた一連のパターンから対応する手話単語を認識結果として出力する。

3 まとめと今後の課題

このシステムを使えば、手話動画像から一定の時間間隔で取り出される静止画像について、手の形、手の位置、手の動きをいちいち全部調べる必要がないので手話単語の認識の効率が向上する。さらにこれはニューラルネットワークに直接実現できるので、それにより高速化が行える。このインプリメンテーション進行中である。

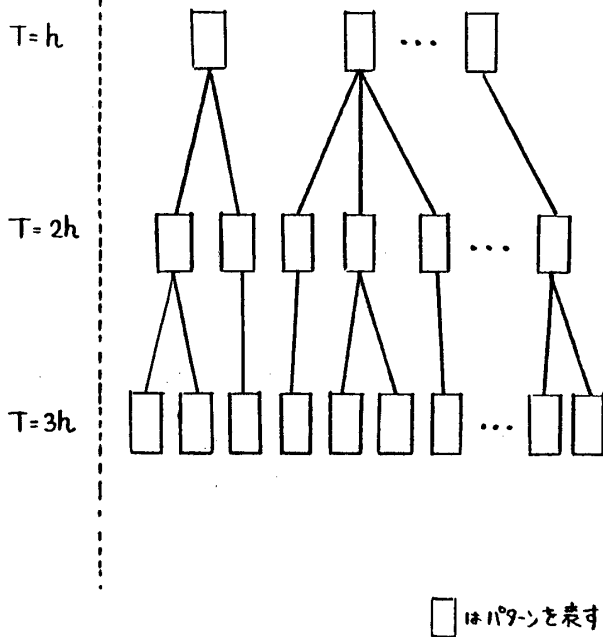
4 参考文献

- [1] 落合、浅井、鎌田：「輪郭特徴を用いた指文字認識方法」,PRU88-44, 電子通信情報学会技術研究報告 (1988)
- [2] 鎌田、他：「テレビ電話を用いた手話会話」 HC90-11, 電子通信情報学会技術研究報告 (1990)
- [3] 藤野、他：「機械処理のための手話構成パラメータの検討」 HC90-10, 電子通信情報学会技術研究報告 (1990)
- [4] 田村、川崎：「手話動画像認識システム」 コンピュータビジョン 44-1(1988)
- [5] N.Ohbo, K.Shimizu, and T.L.Kunii, "A graph-theoretical approach to region detection," *Proc. IEEE COMPSAC '79*, pp.751-756,1979; TECHNICAL REPORT 79-10, DEPARTMENT OF INFORMATION SCIENCE, Faculty of Science, University of Tokyo (1979)

< 図 1 >

手話動画像のスタートからの経過時間(T)

Tに対応するパターン群



< 図 2 >

Tに対応するパターン群

