

手話認識のための動き特徴に基づく学習データの自動合成

相本幸治[†] 山田寛[†] 松尾直志[†] 白井良明[†]

[†] 立命館大学

滋賀県草津市野路東 1-1-1

E-mail: †{aimoto,yamada,matsuo,shirai}@i.ci.ritsumei.ac.jp

あらまし 本論文では HMM を用いた手話の学習・認識のための合成データの自動作成について述べる. HMM の学習には多数のデータを用いるのが望ましいが, 手話動作のデータ収集はコストが高い. そこで各単語の中の類似動作の部分の特徴量を交換することで新しいデータを自動合成し, 学習データを増やして学習を改善する. 類似動作の発見には, k -means 法によるクラスタリングを用いる. 適切な類似動作の数は未知であるため, クラスの統合・分割を繰り返して最適な類似動作分類を得る. 手動による類似動作分類をもとに作成した合成データと提案法で自動作成した合成データを HMM のモデル学習に用い, 認識実験結果を示す.

キーワード 手話認識, 動画像処理, HMM, 状態分割, 合成データ

1. はじめに

日本語を手話に翻訳するシステムでは, テキストで入力された日本語を CG アニメーションで表示するものが実用化されている [1] [2]. 一方, 手話を日本語に翻訳システムはデータグローブなどの装着デバイス [3] を用いた手法とカメラなどで撮影された映像を用いる手法があり, 前者は試験的に実用化されている. また [4] では指先 5 点と手首 1 点に異なる色を割り当てたカラーグローブを用いて指の切り出しを行っている. しかし装着デバイスを用いる方法では特徴抽出が早く容易であるが, 装着に伴うわずらわしさや持ち運びの不便さが実用化の問題となる. 一方, 映像を用いる手法ではそのようなわずらわしさから解放されるが, ノイズや肌色領域の重複のため手や顔の取得が難しいという問題がある. 本論文では話者への負担を考え, 映像から手領域の形状や動きを抽出して特徴量を求め, 手話認識を行う方法について考える. 手話動作時の特徴から認識を行う方法として, 動的時間伸縮法 (Dynamic time warping, DTW) [5] や隠れマルコフモデル (Hidden Markov Model, HMM) [6] [7] などがある.

手話動作は同じ単語であって速度や動きが異なる場合がある. そこで, ここでは時間的な伸縮や多少の動きの違いに対応できる HMM を用いる. HMM は, 音声認識 [8], 表情認識 [9] [10] [11], ジェスチャー認識 [12], 行動認識 [13] [14] の分野でよく用いられている.

HMM は複数の「状態」の時間的な遷移関係と各状態での特徴量の分布の組で表され, 手話認識においては各々の状態が簡単な手の動きに対応する. 学習時には各手話単語のそれぞれに対応した HMM を作成し, 認識時は入力特徴量を出力する尤度が最も高い HMM に対応する単語を認識結果とする. 手話認識システムの概要を図

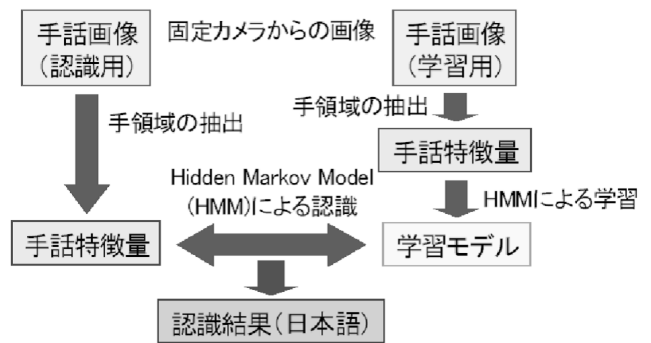


図1 手話認識システムの概要

1に示す.

手話単語動作は, 発話者によって動作の一部が省略されたり動きが多少変形されたりすることがある. HMM による認識精度の向上を図るためには, その手話動作で起こり得る動きや手形状を全て網羅した学習データを用いて学習するのが望ましいため, 多数の学習データが必要となる. しかし手話を行える人が少なく, また正確に顔と両手のパラメータを抽出するための撮影環境を整えなければならず, 手話動画の取得にはコストがかかる.

この問題を解決するため, 新庄ら [19] は, 画像における手の見え方によって手形の分類を行い, 異なる手話で, 同じ手形の特徴量を交換することで合成データを作成している.

本論文では, 手の動きに着目し, 異なる手話の類似の動きの特徴量を交換することによって学習データを合成する.

2. 手話認識に用いる特徴量

手話においては, 手の細かい動きに重要な意味がある場合, 手は顔に近い位置で動かされる傾向がある. 一方,

顔から離れた位置では細かな位置の違いはそれほど重要ではなく、むしろ話者や発話ごとの差が大きいため細かな違いによって区別するのは適切ではない。そこで、手領域の重心位置を表す特徴量として、顔領域の重心からのユークリッド距離の対数と顔重心からの方向を用いる。距離の対数を用いることで顔に近い動作は細かく区別しつつ、顔から離れた位置での動作の細かな違いを無視することができる。ここでは川東ら [16] の方法で以下のような特徴を抽出し、動きに関する特徴として用いる。

- 右手左手のそれぞれについての位置情報
 - (1) 顔領域重心からの距離の対数
 - (2) 顔領域重心からの距離の対数の変化
 - (3) 顔領域重心からの方向
 - (4) 顔領域重心からの方向の変化
- 左手領域重心からの右手領域重心までの距離
 手の形状の特徴には面積・円形度・慣性主軸方向・突起数・周囲長などがある [18]。本研究では手領域を画像系列から抽出しているため手の向きや形状の少しの変化で指同士の領域がくっついたり離れたりする。そこで、これらによる影響の少ない以下の特徴を用いる。

- (1) 面積
- (2) 円形度
- (3) 慣性主軸

3. 手話単語の学習・認識

HMM の各状態がもつ出力の確率分布としては Gauss 分布を仮定し、そのパラメータ及び遷移確率を Baum-Welch アルゴリズムによって学習する。

認識時には対象となる特徴ベクトル時系列の出力尤度を Viterbi アルゴリズムで各単語の HMM について求め、最大の尤度を与える HMM に対応した単語を認識結果とする。

HMM の学習では、最初に画像系列をいくつかの状態に分割して、各状態の特徴の平均と分散を与える必要がある。手の位置・速度に関する特徴量を用いて画像系列を静止区間と動区間に分割する。区間分割の流れを以下に示す [16]。

- (1) 手領域の重心位置の移動速度によって静止区間と移動区間に分類する。
- (2) 移動区間のうち、振動区間を求める。
- (3) 移動区間に運動方向が急変する場合は分割する。
- (4) 静止区間内で手の形状が変化している区間を形状変化区間とする。
- (5) 両手手話の場合は、片手のみの分割結果を統合して両手手話としての区間とする。

4. 手動による合成データの作成

手話単語を状態分割すると、異なる手話単語であっても同じ動き特徴を持つ区間が多く存在する。そこで各単語を区間分割し、手の速度が大きい運動区間の手の速度

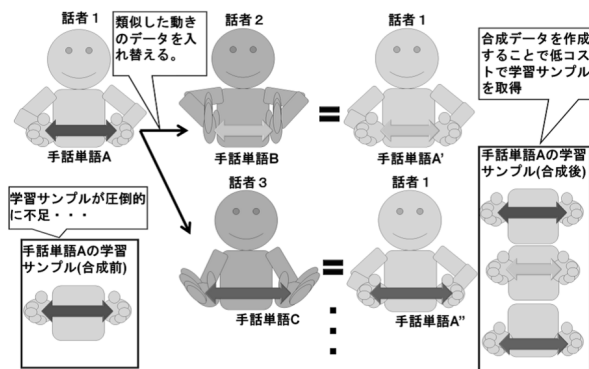


図2 学習データ合成システムの概要

と運動方向に着目することで、類似した動作(基本動作という)を行う区間を決定し、その区間の両手の運動に関する特徴量(運動パラメータ)を交換することで新しい学習データを合成する。

基本動作を決めるため、これまで認識を行っている手話単語の中から、少なくとも2単語以上で類似な動きをする区間があるものを選択した。具体的には、表1に示す22単語から8種類の基本動作を抽出した。

表1 実験で使用した手話単語の基本動作

右手	左手	手話単語例
↑	静止	姉, 兄 (片手手話)
↑	静止	ありがとう, 料理
↓	静止	ありがとう, 料理, 座る, やめる, 最低
↑	静止	合格, 最高
↓	↑	世話, うれしい, 試験, どちら, 姉妹, 兄弟
↑	↓	世話, うれしい, 試験, どちら, 姉妹, 兄弟
→	←	短い, 小さい, 集まる, 休み
←	→	大きい, 長い, 久しぶり

学習データ合成の概要を図2に示す。

学習データの合成の際に、同じ話者による同一単語から得られるデータを交換して合成したデータを用いて学習しても、各状態に割り当てられる学習用特徴ベクトルの集合は共通であり、HMMが持つ運動パラメータの出力分布と状態遷移確率は変化しないため認識率を改善できない。そこで、学習データを合成する際には、異なる話者の異なる単語でデータを合成する。図3に合成で用いる学習データの組み合わせの例を示す。「久しぶり」という単語の手形状データと「長い」という単語の運動パラメータを部分的に交換する場合、話者Aの「久しぶり」、それとは異なる話者Bの「長い」を組み合わせると合成データを作成する。

手話単語を静止区間と運動区間に分割した後、ユーザーが動画を1フレームづつ確認し、類似した基本動作をしている運動区間を1つのクラスタとするクラスターリングを行う。動作の類似性を判定するのはデータの合成で交換する運動パラメータは以下である。

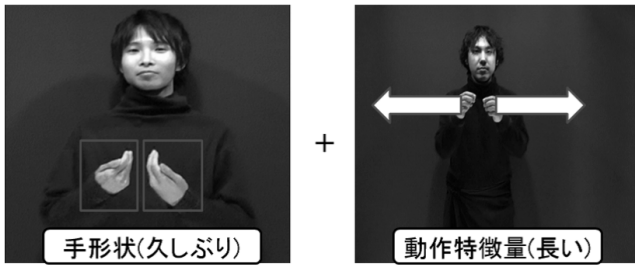


図3 異なる話者による合成データの作成

- (1) 手と手の相対座標の対数
- (2) 手の速度の対数
- (3) 顔と手の相対座標の対数

なお、手と手、あるいは顔と手が近い場合には相対位置の差が重要になるので、対数を用いている。つぎに、同じクラスに属している区間で運動パラメータの入れ替えを行うことで学習データの合成を行う。人が類似した基本動作を選択する場合、図4のように手の運動パラメータが同じであっても、動作が行われる手の重心位置に大きな違いがある区間同士は交換しない。

なお、交換する運動区間のフレーム数が一致しない場合は、パラメータの入れ換えを行う時に運動区間のフレーム数の差を補完し、運動パラメータは線形補完する。



(a) 胸より下で行う動作



(b) 胸より上で行う動作

図4 パラメータの交換は行わない類似した基本動作の例

以上のように、手動による合成データの作成には、合成に使用する手話単語の動きを1フレームずつユーザーが確認して合成データの作成を行うため、類似動作のクラスタリングの手間がかかる。

5. 学習データの自動合成

手話単語サンプルの運動区間から基本動作を自動的に決定し、各運動区間をクラスタリングできれば、学習データの自動合成ができる。自動合成のために交換する運動パラメータの類似性は、手動での類似性に用いた特徴に、両手の位置を加えたものとしている。すなわち、以下に示す10次元の特徴ベクトルを用いる。

- (1) 左右手領域の重心座標 $(x_L, y_L), (x_R, y_R)$
- (2) 左右手領域の重心の速度 $(v_{xL}, v_{yL}), (v_{xR}, v_{yR})$
- (3) 顔領域と左手領域の重心間距離 d_L
- (4) 顔領域と右手領域の重心間距離 d_R

運動区間の類似性をこの特徴ベクトル間のユークリッド距離で評価してクラスタリングを行う。

クラスタリングは、まず k -means 法を用いて初期クラスタリングを行う。具体的には、運動区間データの数を N 、各運動区間の特徴ベクトルを $\mathbf{x}_i (i = \dots, N)$ 、クラス数を K とし、以下に示すステップで行う。

- (1) 各運動区間 \mathbf{x}_i をランダムに1から K のクラスに割り振る。
- (2) 割り振った運動区間をもとに各クラス k について中心 \mathbf{c}_k を計算する。中心の計算は割り当てられたデータの各パラメータの平均が使用される。
- (3) 各 \mathbf{x}_i についてクラス k の中心 \mathbf{c}_k との距離を求め、 \mathbf{x}_i を最も近い中心のクラス k' に割り当て直す。

$$k' = \arg \min_{k=1 \dots K} \|\mathbf{x}_i - \mathbf{c}_k\|$$

- (4) 上記の処理で全ての \mathbf{x}_i のクラスの割り当てに変化がない場合は処理を終了する。それ以外の場合は新しく割り振られたクラスから中心を再計算して(2)~(3)の処理を繰り返す。

なお、ここで用いている画像のサイズは 640×480 ピクセルであり、重心座標、相対座標、相対距離は $0 \sim 400$ ピクセルまで、速度は $-20 \sim 20$ ピクセル/フレームまでの値となる。そこで他のパラメータとスケールを合わせるために速度の値は30倍している。

5.1 終了動作区間、準備動作区間の決定

本研究で使用する手話動画には、図5のように初期フレームの両手を下ろした状態から両手を上げる動作区間、図6のように手話動作が終わり、手を下す動作区間のように手話として意味をなさない動きである動作区間が存在する。本論文ではこれらの動作区間を終了動作区間、準備動作区間と定義する。

これらの区間には手話単語や話者の違いなどによってばらつきが見られるため、クラスタリングを行う際にこれらの区間を用いると、新しいサンプルを使用するたびに新たな準備動作区間と終了動作区間が現れ、クラス数の決定が困難になる。準備動作区間と終了動作区間にはそれぞれ顔と手の距離が大きく、また手の速度の y 成分

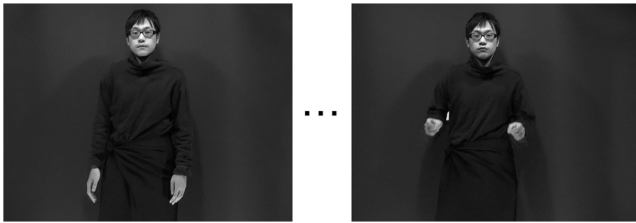


図 5 準備動作区間

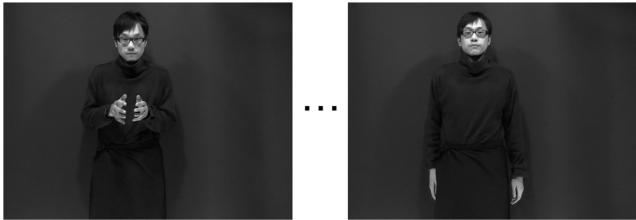


図 6 終了動作区間

が大きい傾向にある．そこで閾値を設定することで準備動作区間と終了動作区間をクラスタリング対象から外す．クラスタリング対象とする運動区間の判定を以下に示す．

(1) 両手手話の場合は顔と右手の相対距離，顔と左手の相対距離の平均が共に 140 ピクセル以上である．片手手話の場合は顔と右手の相対距離の平均が 140 ピクセル以上である

(2) 両手の重心の y 成分の平均が 160 ピクセル以上である (両手手話)，右手の重心の y 成分の平均が 160 ピクセル以上である (片手手話)

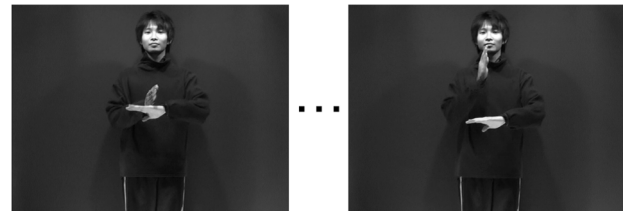
(3) 両手の速度の y 成分の平均が 270 以上である (両手手話)，右手の速度の y 成分の平均が 270 以上である (片手手話)

(4) 手の速度の y 成分が負の値であれば終了動作区間とし，正の値であれば準備動作区間としてクラスタリング対象から外し，1~3 に当てはまらない運動区間はクラスタリング対象とする

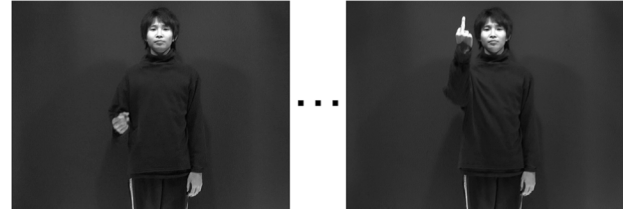
今回の実験で用いた図 5, 5 は環境を固定して撮影しており，話者・カメラ間の距離は約 2.3m である．上記の判定方法で用いるパラメータはこの撮影環境において話者 3 人の動作をもとに決定した．

5.2 クラスの統合・分割

k -means 法は初めにランダムに設定される中心の位置によって，類似した基本動作であっても別のクラスに割り当てられてしまったり，類似していない基本動作が同じクラスに割り当てられる場合がある．例えば，図 7 のように，「ありがとう」や「料理」のように右手を上へ上げて左手が胸の付近で静止している手話単語と兄，姉のような片手手話は右手と左手の速度は類似しているが，顔との相対座標や両手の重心位置などに差がある．ところが，最初に設定された中心の位置によっては，同じクラスに属してしまう．また，図 8 のように「大きい」という手話単語が特徴空間上にほぼ同じ特徴ベクトルを持

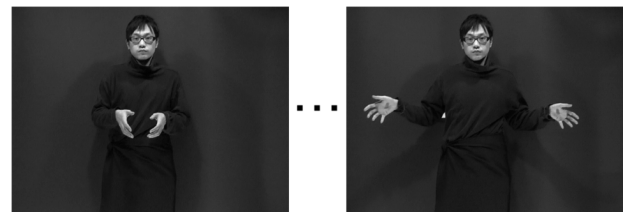


(a) ありがとう

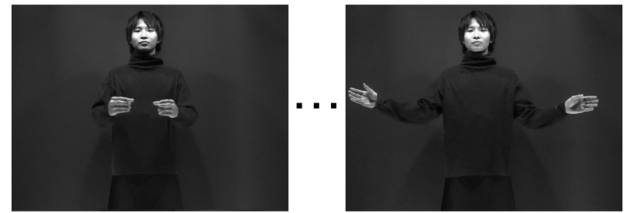


(b) 兄

図 7 同じクラスに属してしまう類似していない運動区間例



(a) 大きい



(b) 大きい

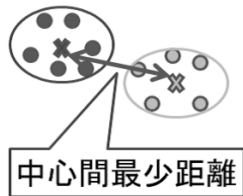
図 8 異なるクラスに属してしまう類似運動区間例

つ中心が複数存在する場合には，話者の違いなどによるパラメータの小さな差で別々のクラスに属してしまう場合がある．

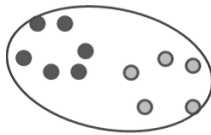
このようなクラスタリングの不安定に対処するため，クラスの分割と統合を行う．クラスの中心間のユークリッド距離が最少となるペアを求め (最小距離を d_{\min} とする)，そのペアを統合するクラスの候補とする．また各クラス内のサンプル間距離を求め，その中で最大のクラス内サンプル間距離 r_{\max} を持つクラスを分割するクラスの候補とする．

クラスの統合処理を行う際には，統合候補となるクラスを統合した場合のサンプル間の最大距離 r が， r_{\max} より大きくなってしまふと，別のクラスに属していた動作が統合される可能性がある．そこで， r が r_{\max} より小さい場合に限り 1 つのクラスに統合する (図 9)．

分割処理を行う際には，分割候補のクラスを分割した場合の中心間の距離 d を求め，それが d_{\min} より小さくなる場合には，類似した動作を 2 つのクラスに分割して

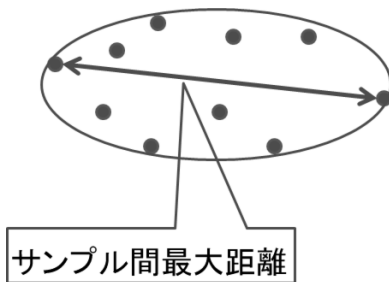


(a) 統合前

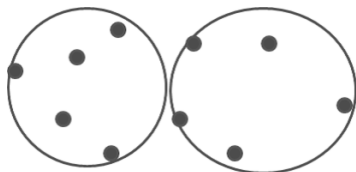


(b) 統合後

図 9 クラス統合処理



(a) 分割前



(b) 分割後

図 10 クラス分割処理

しまう可能性がある。そこで、 d が d_{\min} より大きい場合に限り2つのクラスに分割する(図10)。

以下に統合・分割処理の手順を示す。

(1) クラスの中心間距離最小のペアとその最少距離 d_{\min} を求める。最大のクラス内サンプル間距離をもつクラスとその距離 r_{\max} を求める。

(2) 中心間の距離が最少のペアを1つのクラスに統合する。

(3) 統合したクラスのサンプル間の最大距離 r が r_{\max} より大きくなった場合は、統合を解除し、処理を終了する。

(4) 最大のサンプル間距離を持つクラスを最大距離のサンプルを初期値として k -means法によって2つのクラスに分割する。

(5) 分割後のクラスの中心間距離 d が d_{\min} より小さくなった場合は、分割を解除し、処理を終了する。

(6) (3)もしくは(5)で処理が終了するまで(1)~(5)を繰り返す。

5.3 自動クラスタリングの実験結果

自動合成のためのクラスタリングには準備動作区間、

表 2 統合・分割処理を行わないクラスタリング結果

	右手	左手	手話単語例
1	↑	静止	姉, 兄, ありがとう, 料理
2	↓	↑	世話, 試験, どちら
3	↓	静止	ありがとう, 料理, 座る, やめる, 最低
4	↑	静止	合格, 最高
5	↓	↑	うれしい, 試験, 姉妹, 兄弟
6	→	←	短い, 小さい, 集まる, 休み
7	↑	↓	世話, うれしい, 試験, どちら, 姉妹, 兄弟
8	←	→	大きい, 長い, 久しぶり
9	↑	↑	手話として意味を持たない動作区間
10	↑	静止	手話として意味を持たない動作区間
11	↓	↓	手話として意味を持たない動作区間
12	↓	静止	手話として意味を持たない動作区間
13	↓	↓	手話として意味を持たない動作区間

表 3 統合・分割処理を行うクラスタリング結果

	右手	左手	手話単語例
1	↑	静止	姉, 兄, ありがとう, 料理
2	↑	静止	ありがとう, 料理
3	↓	静止	ありがとう, 料理, 座る, やめる, 最低
4	↑	静止	合格, 最高
5	↓	↑	世話, うれしい, 試験, どちら, 姉妹, 兄弟
6	→	←	短い, 小さい, 集まる, 休み
7	↑	↓	世話, うれしい, 試験, どちら, 姉妹, 兄弟
8	←	→	大きい, 長い, 久しぶり
9	↑	↑	手話として意味を持たない動作区間
10	↑	静止	手話として意味を持たない動作区間
11	↓	↓	手話として意味を持たない動作区間
12	↓	静止	手話として意味を持たない動作区間
13	↓	↓	手話として意味を持たない動作区間

終了動作区間と定義することができなかつた手話として意味を持たない動作区間が必ずあるため、表1に示した基本動作にその分だけクラス数を増やしてクラスタリングを行っている。

クラスタリングに使用した運動区間は準備動作区間、終了動作区間と判断された運動区間を除く142の運動区間を使用し、クラス数は13で実験を行っている。クラスの統合・分割を行わないクラスタリングの実験結果例を表2、統合・分割処理を行ったクラスタリングの実験結果例をの表3に示す。

k -means法はランダムで初期の中心位置が決定されるため、クラスタリングを行う度にクラスタリング結果が変化する。統合・分割処理を行わないクラスタリングではクラスタリングを行う度に類似していない基本動作が同じクラスに割り当てられ、類似基本動作が別々のクラスに割り当てられる。表2のように、類似していない基本動作が同じクラス1に割り当てられる。また、クラス2とクラス5は類似基本動作であるが、別々のクラスに

割り当てられるものや、表2の9~13のクラスに属している手話として意味を持たない動作区間が5つ以上のクラスに分類されてしまうものなどがあり、手動で行ったクラスターリングと同じ結果になることはほとんどなかった。しかし、表2のクラス1, クラス2, クラス5などの誤って分類されたクラスを統合・分割処理を行うことで、表2のように類似動作が同じクラスにクラスターリングされることができた。

6. 認識実験

手話単語認識に合成データを用いることの妥当性を実験によって確かめた。実験には3人の話者による手話画像を用い、単語は手話単語辞典から日常会話に使用される単語の中から22単語を選び撮影した。フレームレートは30フレーム/秒である。

HMMの学習・認識に用いる手話特徴量は以下である。

- (1) 手と手の相対座標の対数
- (2) 手の速度の対数
- (3) 顔と手の相対座標の対数
- (4) 慣性主軸の特徴量

今回の実験に使用した単語は、6種類の類似した動きに分類できる。それぞれの動きに対して実験に使用した単語を以下に示す。

- (1) 姉, 兄
- (2) どちら, 兄弟, 試験, 世話, 姉妹, うれしい
- (3) 短い, 小さい, 集まる, 休み
- (4) 料理, ありがとう, やめる, 最高, 座る
- (5) 久しぶり, 長い, 大きい
- (6) 合格, 最低

認識実験には全てのデータを学習に用いた実験と合成データを学習に用いた実験を行った。合成データを用いた認識実験には合成データの有用性を示すために学習データの中からいくつかを省き、その代わりに合成データを学習データとして使用して認識実験を行った。

6.1 全てのデータを学習に用いた認識結果

各単語について話者ごとに3シーケンスずつ、合計9シーケンスのサンプルがある。そのうち話者毎にサンプルの内の1シーケンスを認識データとし、残り2シーケンスのサンプルを学習データとして、63シーケンスの認識データと214シーケンスの学習データを用いて実験を行った。実験結果を表4に示す。

表4 全てのデータを学習に用いた認識結果

	成功数	認識率
話者 A	22	100%
話者 B	19	90.5%
話者 C	18	90.0%
合計	59	93.6%

6.2 手動合成データを学習に用いた認識結果

表5の実験では、各話者の学習データから1シーケンスずつ省き、その省いた学習データの代わりに43シーケンスの合成データを学習に使用して実験を行う。

()内は手動合成データを学習に使用していない場合の成功数, 認識率を示し, ()がないものは手動合成データを学習に使用している場合の成功数, 認識率を示している。

表5 43シーケンスの手動合成データを学習に用いた認識結果

	成功数	認識率
話者 A	21(20)	95.5%(90.9%)
話者 B	19(16)	90.4%(76.2%)
話者 C	17(17)	85.0%(85.0%)
合計	57(53)	90.1%(84.1%)

表6の実験では、3人の話者の内1人を学習データから省き、その省いたデータの代わりに30シーケンスの合成データを学習に使用して実験を行う。また、手動合成データは合計162シーケンスを作成している。

表6 30シーケンスの手動合成データを学習に用いた認識結果

	成功数	認識率
話者 A	20(17)	90.9%(77.3%)
話者 B	20(20)	95.2%(95.2%)
話者 C	18(17)	90.0%(85.0%)
合計	58(54)	92.1%(85.7%)

6.3 自動合成データを学習に用いた認識結果

表7は表5, 表8は表6と同じ条件で実験を行っている。自動合成データは合計314シーケンスを作成している。

()内は自動合成データを学習に使用していない場合の成功数, 認識率を示し, ()がないものは自動合成データを学習に使用している場合の成功数, 認識率を示している。

表7 43シーケンスの自動合成データを学習に用いた認識結果

	成功数	認識率
話者 A	21(20)	95.5%(90.9%)
話者 B	19(16)	90.4%(76.2%)
話者 C	17(17)	85.0%(85.0%)
合計	57(53)	90.1%(84.1%)

表 8 30 シーケンスの手動合成データを学習に用いた認識結果

	成功数	認識率
話者 A	20(17)	90.9%(77.3%)
話者 B	20(20)	95.2%(95.2%)
話者 C	18(17)	90.0%(85.0%)
合計	58(54)	92.1%(85.7%)

7. おわりに

本研究では、手話の動き特徴に関する特徴量に基づき、既存の手話動画データから新たな学習データを合成し、認識実験を行った。本論文では合成データを使用する2種類の実験を行い、どちらの実験も合成データを使用しないものよりも認識率が向上し、合成データの有用性を示すことができた。また、除外する学習データと代用する合成データの組み合わせを替えて認識実験を行った結果、認識率の上昇は使用した合成データによって2.1~6.4ポイントの上昇し、同じ話者の基本動作を交換して作成した合成データのみを学習に使用するよりも様々な話者で作成した合成データを学習に使用した方が認識率の向上が見られた。また、長いや大きいなどの手話単語は合成データを使用しない認識実験では尤度の差が僅差であるため、合成データを使用する認識実験でどちらか一方が認識に成功すると、もう片方が認識に失敗するという結果になった。これらは手の形状に大きな差がなく、合格、最高や姉、兄のように合成データを作成できる総数が少ない基本動作に見られたため、サンプル数を増やしどの基本動作にも十分な合成データを準備する必要がある。また、表9に示すように合成データの作成を自動で行うことで手動に比べて合成データ作成の作業時間を大幅に短縮することができた。

表 9 手動作成と自動作成の作業時間の比較

作成の種類	合成データの作成数	時間(sec)
手動	162	7560
自動	314	20

8. 今後の課題

今後の課題としては、さらに多くの単語について合成データの作成を行い、学習データに使用していない未知の話者の手話単語であっても認識できるようにすること。新庄らが行った手形状を用いた合成データの作成と本論文での動き特徴に基づいた合成データの作成を統合し、図11のように所有していない手話単語を合成データにより作成することが今後の課題として挙げられる。

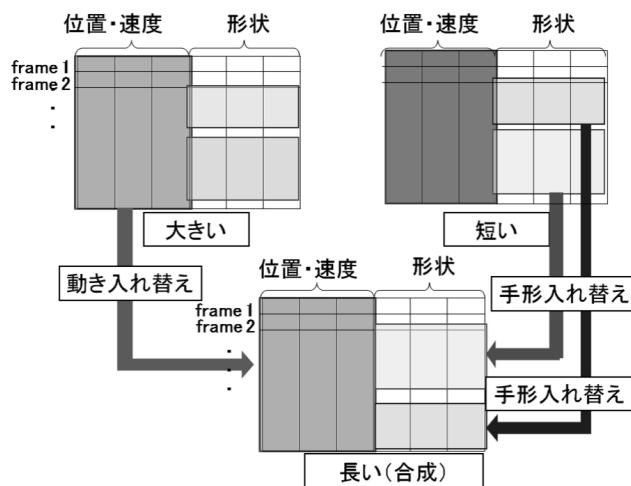


図 11 手の動きと形状に基づく合成データの作成方法

文 献

- [1] H. Sagawa, M. Ohki et al, "Pattern Recognition and Synthesis for a Sign Language Translation System", Journal of Visual Languages and Computing Vol. 7, No. 1, pp. 109-127, 1996.
- [2] 森本, 川村, 黒川, "胃部レントゲン検査の指示に用いる手話アニメーションの作成とその評価", 信学技報, Vol. 105, No. 67, pp. 37-42, 2005.
- [3] H.Sagawa, M.Takeuchi, "A Method for Recognizing a Sequence of Sign language Words Represented in Japanese Sign Language Sentence", Proc. Int. Conf. Automatic Face and Gesture Recognition (FG2000), pp. 434-439, 2000.
- [4] 佐治, セン, 森本, 黒川, "HMM を用いた複雑な手形状を伴う手話単語認識", ヒューマンインタフェースシンポジウム 2003 論文集, pp. 281-284, 2003.
- [5] 大崎 竜太, 上原 邦昭, "Dynamic Time Warping 法を用いた身体運動の動作識別", 情報処理学会研究報告. データベース・システム研究会報告 98(58) pp. 233-240, 1998.
- [6] Grobel, K. and Assan, M., "Isolated sign language recognition using hidden Markov models", Proceedings of the International Conference on System, Man and Cybernetics (1997), pp. 162-167, 1997.
- [7] Starner, T., Weaver, J. and Pentland, A. RealTime American Sign Language Recognition Using Desk and Wearable Computer Based Video, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 12, pp. 1371-1375, 1998.
- [8] 中川聖一, "確率モデルによる音声認識", 電子情報通信学会, コロナ社 (1988).
- [9] 坂口, 大谷, 岸野, "隠れマルコフモデルによる顔画像からの表情認識", テレビジョン学会誌, Vol. 49, No. 8 1995.
- [10] 大塚, 大谷, 中津, "連続出力確率密度分布を用いたHMMによる動画からの複数人物の表情認識", 電子情報通信学会論文誌, vol.J80-D-II NO.8 1997.
- [11] Takahiro Otsuka, Jun Ohya, "Spotting Segments Displaying Facial Expression from Image Sequences Using HMM", 0-8186-8344-9/98, IEEE 1998.
- [12] Tatsuya Ishihara, Nobuyuki Otsu "Gesture Recognition Using Auto-Regressive Coefficients of Higher-Order Local Auto-Correlation Features", FG2004, pp. 583-588, 2004.
- [13] 大和, 大谷, 石井, "隠れマルコフモデルを用いた動画からの人物の行動認識", 電子情報通信学会論文誌, Vol.

- [14] Christopher R. Wren, Brian P. Clarkson, Alex P. Pentland, “Understanding Purposeful Human Motion”, FG2000, pp. 378-383, 2000.
- [15] 岡澤, 西田, 堀内, 市川, “わたり区間を含む単位を用いた手話認識手法の検討”, 信学技報, vol. 103, No. 747, pp. 13-18, 2004.
- [16] K. Kawahigashi, Y. Shirai, J. Miura, and N. Shimada, “Automatic Synthesis of Training Data for Sign Language Recognition Using HMM”, Proc. International Conference on Computers Helping People with Special Needs, pp.623-626, 2006.
- [17] 松尾直志, 白井良明, 島田伸敬, “手話認識のためのHMM構造の自動生成”, ヒューマンインタフェースシンポジウム 2008 論文集, pp.915-922, 2008.
- [18] von Agris, U., Zieren, J., Canzler, U., Bauer, B., Kraiss, K.F., “Recent developments in visual sign language recognition.”, Springer Universal Access in the Information. Society 6(4), pp. 323–362, 2008.
- [19] 新庄智子, 山田寛, 松尾直志, 白井良明, 島田伸敬, “HMMを用いた手話認識のための学習データの自動合成”, ヒューマンインタフェースシンポジウム 2007 論文集, pp.475-480, 2007.