Vol.2016-AAC-1 No.10
2016/7/30

Kinect を用いた手話の数字認識における識別手法の検討

佐藤貴弘[†1]　堀内靖雄[†1]　川本一彦[†1]　下元正義[†2]　眞崎浩一[†2]　黒岩眞吾[†1]
鈴木広一[†2]

概要：本論文では Kinect を用いた手話の数字認識手法について検討する．Kinect によって得られた画像から手領域の輪郭線を抽出し，これを特徴量に用いてパターンマッチングによる識別を行った．認識対象は 1～9，10～90 の 18 種類の数字である．手話者クローズ実験を行ったところ 99.9%の認識率が得られた．また，手話者オープン実験では 1 人評価 3 人テンプレートでクロスバリデーションを行ったところ 89.6%の認識率が得られた．

キーワード：手話数字認識，Kinect，パターンマッチング，最近傍探索

1. はじめに

現在，ろう者の社会進出に伴い，ろう者と聴者がコミュニケーションをとる機会が増加している．手話を用いるろう者と手話を知らない聴者とのコミュニケーションは筆談や手話通訳者を介して行われる．筆談を用いた場合，手話を母語とするろう者にとっては，日本語を書くことは負担となり，コミュニケーション速度が遅くなってしまう．また，手話通訳者を介する場合は，手話通訳者の数が限られている上に，守秘義務を伴う通訳者であっても，プライバシーに関わることであれば話しづらいことがある.そこで，手話対話システムの開発が望まれている．手話対話システムとは手話を入力し，その手話に対する適切な応答を手話で出力するシステムであり，病院の受付や道案内などへの利用が期待されている．本研究では手話対話システムの入力部である手話認識について検討を行う．また，手話表現の中でも特に数字の認識に関する検討を行う．

2. 先行研究と目的

手話の数字認識の研究は少ないが，関連した研究として指文字認識の研究が活発に行われている．指文字認識を行うには手領域の抽出が必要となるが，従来のビデオカメラによる手話の認識手法[1]では背景の色の制約を受けるという問題があった．それに対して距離計測カメラを用いる手話認識手法[2][3]ではカメラから物体までの距離が取得可能であり，カメラから一番近い物体を手として認識し，手領域を正確に抽出することができるため，近年では距離計測カメラである Kinect 等を用いた指文字を認識する研究が盛んに行われている[4][5]．識別に用いる特徴量は，HOG や HLAC などの画像局所特徴[6][7]や，指の本数や手の輪郭線を用いた研究[8][9]がある．

本研究では Kinect を用いた手話の数字認識を行う．手領域の抽出方法として，SDK のスケルトントラッキングと深度情報を組み合わせることでより正確に手領域を検出する[10]．Kinect によって得られた画像から OpenCV のライブラリを用いて手の輪郭線を取得し,実験データを作成する．これを用いて，テンプレートデータとのパターンマッチングにより識別を行う手法を提案する．

3. Kinect による手の輪郭線の取得

Kinect によって撮影された映像から手の輪郭線を取得する手順について述べる．まず，図 1 のように時系列順に画像が得られる中から，数字を表している瞬間の画像フレームを探す．これは，数字を表す際に手を前に押し出すスタンピングと呼ばれる動作を検出することで実現する．Kinect の深度画像を用いることで，各画素におけるカメラから被写体までの距離を測ることが出来る．スタンピングを行う瞬間の手の位置は，深度画像に写っている範囲の中で Kinect までの距離が局所的に一番近くなると考えられる．そこで，各フレームの最小画素値（最小距離に相当）を記録していき，極値を求めることでスタンピングの瞬間のフレーム検出を行う．ここで注意が必要なのは，スタンピングによって数字を提示する瞬間の他に，手を基本姿勢に戻す際にも極値が検出されてしまうことである．これを防ぐために，深度画像全体の中から数字が提示される範囲をあらかじめ指定する．そして，そのブロックの中で最小の画素値を記録していき，極値を検出したフレームを抜き出すようにした．

図 1　Kinect で撮影された赤外線映像の画像フレームの例

†1 千葉大学
Chiba University
†2 みずほ情報総研(株)
Mizuho Information ＆ Research Institute, Inc.

続いて，数字を表している瞬間の画像フレームから，手の輪郭線を取得する．手の輪郭線を取得する工程では，まず始めに手のマスク画像を作成する．Kinect のスケルトントラッキング機能から得られる手の位置座標から，その地点の深度情報を抽出し手の位置から奥行方向 8cm を切り出すことで，手のマスク画像が作られる（図 2(a)）．続いて，このマスク画像を用いて赤外線画像から手領域のみを抽出する （図 2(b)）．ここで一つ問題がある．マスク画像を奥行方向 8cm まで切り出したときに，手領域の範囲に腕まで含んでしまうことがある．このため，実験参加者には腕に図 3 に示すようなゴムバンドを巻いて頂き撮影を行った．ゴムが赤外線を吸収するため，ゴムバンドの部分は赤外線画像中では映らなくなる．この画像に 2 値化処理を行うと，手領域とその他の領域が分けられた 2 値画像が生成される（図 2(c)）．さらにラベリング処理を行うことで手領域のみの画像を取得した（図 2(d)）． そして，OpenCV の輪郭線追跡機能によって輪郭線の位置座標を取得した．

(a)マスク画像　(b)赤外線画像を重ねた画像　(c)2 値化処理をした画像　(d)ラベリング処理をした画像

図 2　輪郭線取得までの流れ

図 3　手首に巻いて頂いたゴムベルト

4. 実験データ作成

4.1 標準化

実験参加者によって手の大きさが異なると考えられる．また，収録を行う際の Kinect とカメラの距離の違いによって，手の大きさが変わると考えられる．提案手法では，輪郭線同士の距離を類似度として識別を行うため，同じ数字を表す手の大きさは同一になることが望まれる．したがって，手の大きさの違いを吸収するために標準化を行う必要がある．ここでは標準化の手法について述べる．

まず，標準化を行うために手の輪郭線データを極座標系に変換する．図 2 (d)のラベリングした画像から手領域の重心を求め，重心から手の輪郭線の各位置座標までのユークリッド距離と角度を計算し，極座標変換を行う．続いて，各実験参加者ごとの全収録データに対して，式（1）によって標準化を行う．さらに，類似度計算を行うためにユークリッド座標系に戻す必要があるが，標準化を行ったデータをそのままユークリッド座標系に戻すことは出来ない．そこで，実験参加者の手の大きさを表す平均値を基準として逆標準化を行う．逆標準化に用いる式を式（2）に示す．全実験参加者の平均値のうち，中央値となる参加者を探し，その平均値（μ'）と標準偏差（σ'）の値を用いて逆標準化を行う．最後に，逆標準化を行ったデータをユークリッド座標系に変換する．これによって手の大きさの違いが吸収される．

$$x' = \frac{x - \mu}{\sigma} \qquad (1)$$

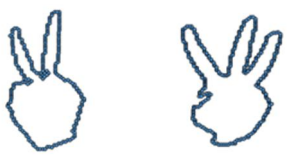
$$x'' = \sigma' x' + \mu' \qquad (2)$$

4.2 回転

手を提示する際の角度は呈示ごとにわずかに異なる．この角度の違いを吸収するために，輪郭線データを回転させる操作を行う．パターンマッチングを行う 2 つの輪郭線データのうち，片方の輪郭線データを－20° ～＋20° の範囲で 1° ずつ回転させながら，もう片方の輪郭線データとの 2 乗誤差を計算していき，相互相関が最大となる角度を算出する．ただし，このときに用いるデータの形式は極座標系であり，2 乗誤差を計算する前処理として線形補間を行っている．同じ角度に 2 つ以上のデータが表れる場合は，大きい方のデータを取るようにした．相互相関が最大となる角度に回転させたデータでパターンマッチングを行った．

5. 最近傍探索による数字識別

本研究では距離尺度として，最近接点までの距離の和を用いる．評価データの輪郭線の各点から，テンプレートデータの輪郭線の各点までの距離を計算していき，その最小値の和で評価を行う．ただしこの手法には問題がある．例えば，図 4(a)に示すような 2 という数字が図 4(b)に示す 3 という数字に誤認識されてしまうようなケースである．つまり，2 つのデータを重ね合わせた際に 2 つのデータの距離が近い部分だけを評価に用いるので，評価に用いられない指の部分がマッチングに考慮されないのである．この問題を解決するために，同様の計算を逆方向（テンプレート→評価データ）でも行い，2 つの評価値のうち，大きい方を採用する方法で認識を行った．



(a)2 のデータ　(b)3 のデータ

図 4　実験データの例

6. 評価実験

提案手法の評価実験として，2 つの実験を行った．ひとつは評価者のデータをテンプレートとする手話者クローズ実験である（評価データはテンプレートから除外）．もうひとつは，評価者以外の 3 人のデータをテンプレートとする手話者オープン実験である．手話者オープン実験では，ベースライン，標準化，回転，標準化＋回転の 4 つの条件で実験を行った．

6.1 実験条件

実験で使用した数字は，1～9, 10～90 の 18 単語であり，各単語ごとに 10 回の収録を行った．手話サークルに所属する聴者 4 名に協力していただき，18 単語×10 回×4 名の 720 個のデータを収録した．慣れや疲れによる影響を小さくするため，収録は一定時間ごとに十分な休憩を挟んで行った．

手話者クローズ実験では，1 名分の収録データ 180 個のうち，評価データを除く全てのデータ（179 個）をテンプレートデータとして，180 個全てを評価データとして 4 名分の評価実験を行った．

手話者オープン実験では，1 名分の収録データ 180 個を評価データ，他の 3 名分の収録データ 540 個をテンプレートデータとして 4 名分の評価実験を行った．

6.2 実験結果

手話者クローズ実験の結果を表 2 に示す．実験参加者 A の実験で 1 つだけ誤認識が発生した．評価データの 80 を 70 に誤認識したものである．

手話者オープン実験の結果を表 3 に示す．さらに，標準化+回転の条件での混同行列を表 1 に示す．

表 2　手話者クローズ実験の結果

実験参加者	認識率(%)
A	99.4
B	100
C	100
D	100
平均	99.9

表 3　手話者オープン実験の結果

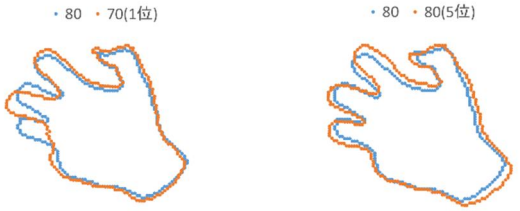
実験参加者	認識率(%)			
	ベースライン	標準化	回転	標準化+回転
A	62.2	91.7	59.4	95.6
B	48.3	81.1	58.3	97.2
C	72.2	78.9	70.6	78.9
D	60.6	70.6	73.3	86.7
平均	60.8	80.6	65.4	89.6

表 1　Confusion Matrix(標準化+回転)

	1	2	3	4	5	6	7	8	9	10	20	30	40	50	60	70	80	90
1	**40**																	
2		**40**																
3			**40**															
4			4	**36**														
5					**40**													
6						**36**									4			
7							**40**											
8							9	**31**										
9									**37**									3
10	1									**39**								
20											**40**							
30												**40**						
40											1	6	**33**					
50														**40**				
60															**40**			
70																**25**	15	
80															1	18	**21**	
90									7					5			1	**27**

6.3 考察

まず，手話者クローズ実験で発生した誤認識について分析する．誤認識が発生したデータは，実験参加者 A の評価データ 80 である．テンプレートデータ第 1 位から第 4 位まで 70 が選ばれており，第 5 位で初めて正解である 80 が選ばれた．評価データ 80 とテンプレートデータ第 1 位の 70 のデータをプロットした図と，評価データ 80 とテンプレートデータ第 5 位の 80 のデータをプロットした図をそれぞれ図 5（a），図 5（b）に示す．2 つのデータを比較すると，図 5（a）のデータは手首側の違いは小さく，指側の違いが大きい．一方で図 5（b）のデータは，手首側も指側も全体的に違いがある．そこで，誤認識の原因となっている輪郭線の部位を探すために，指側と手首側で輪郭線を区切って 2 つのデータを作成し，改めて実験を行った．その結果を表 4 に示す．表中の指側のみの結果の 5 位が 263.3 となっており，手話者クローズ実験で 5 位だったデータが指側のデータのみでは 1 位となっている．したがって，手首側の輪郭線の違いが誤認識に影響していたと考えられる．また，指側の輪郭線において指の逆側の輪郭との距離が短くなる（指の内側と外側の誤認識）ことで輪郭線同士の距離が縮まり，誤認識の原因になったとも考えられる．今後考えられる手法として，Kinect から得られる depth 値（Kinect から被写体までの距離）を使って，手首側の輪郭線を削除して指側の輪郭線のみを用いる手法が考えられる．また，指の内側と外側の誤認識を防ぐために，指部分を塗りつぶしてマッチングを行う手法が考えられる．



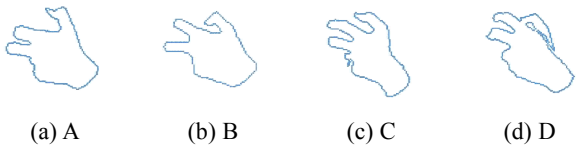
(a)評価データ 80 とテンプレートデータ第 1 位の 70　(b)評価データ 80 とテンプレートデータ第 5 位の 80

図 5　評価データとテンプレートデータ

表 4　実験結果の比較

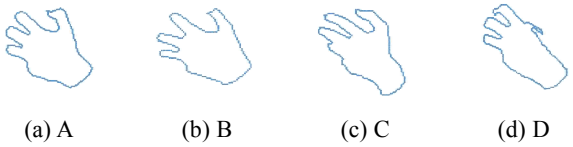
手話者クローズ実験での順位	実験結果の値		
	全ての輪郭線	指側のみの輪郭線	手首側のみの輪郭線
1 位	398.3	283.4	124.2
2 位	411.7	283.0	138.8
3 位	428.8	271.5	173.3
4 位	435.8	325.5	112.0
5 位	439.2	263.3	176.3

続いて，手話者オープン実験の結果の中で特に誤認識が多い数字について原因を分析する．まず，70 を 80 に誤認識する原因及び 80 を 70 に誤認識する原因について分析する．これらの誤認識は，実験参加者 C と D の間で相互的に発生している．表現の個人差を確認するために，4 名の実験参加者の 70 のデータと 80 のデータの例をそれぞれプロットした図を，図 6 と図 7 に示す．70 について，実験参加者 D の親指部分の輪郭線が他の実験参加者とは異なって見える．これは，親指の指先を掌の後ろ側に隠れるように呈示しているためである．また，実験参加者 C と D は薬指のわずかな突出が確認できる．さらに，人差指と中指の関節角度も実験参加者によって異なっている．次に 80 について，70 と同様に実験参加者 D は親指の指先を掌の後ろ側に隠れるように呈示している．また，実験参加者 C の薬指が中指と接しているため，中指と薬指が 1 本の太い指のように見える．このように，同じ数字であっても表現には個人差があり，誤認識の原因のひとつになっていると考えられる．ここで，ひとつひとつのマッチングについて誤認識の原因を分析する．最初に実験参加者 C の 70 を 80 に誤認識している原因について考察する．評価データである実験参加者 C の 70 とテンプレートデータ第 1 位である実験参加者 D の 80 のデータをプロットした図を図 8 に示す．実験参加者 C の人差指と中指が，実験参加者 D の中指と薬指と重なっていることが確認できる．実験参加者 D は 70 と 80 を提示するときに親指を突き出さずに指先を掌の後ろ側に折り曲げてしまう癖があるため，親指が突出しない．このため評価データとテンプレートの間で指の重なりがひとつ隣の指にずれてしまい誤認識になったと考えられる．続いて，実験参加者 D の 70 を 80 に誤認識している原因について考察する．評価データである実験参加者 D の 70 とテンプレートデータ第 1 位である実験参加者 C の 80 のデータをプロットした図を図 9 に示す．実験参加者 D の 70 は，薬指がわずかに突出している．また．実験参加者 C の 80 は薬指を曲げる角度が大きいため，人差指や中指と比べると突出が小さい．他の実験参加者の 70 の手形ならば突出しない薬指と，他の実験参加者と比較して突出が短い薬指のデータがマッチしてしまったことで誤認識が生じたと考えられる．実験参加者 C と D の 80 を，参加者 D と C の 70 にそれぞれ誤認識している原因については同様の理由によると考えられる．



(a) A　(b) B　(c) C　(d) D

図 6　実験参加者 4 名の 70 のデータ



(a) A　(b) B　(c) C　(d) D

図 7　実験参加者 4 名の 80 のデータ

図 8　評価データ C70 とテンプレートデータ第 1 位の D80

図 9　評価データ D70 とテンプレートデータ第 1 位の C80

次に，8 を 7 に誤認識する原因について考察する．この誤認識は，評価データである実験参加者 C の 8 に対して，実験参加者 A の 7 のテンプレートデータが第 1 位に選ばれたものである．評価データである実験参加者 C の 8 と，テンプレートデータ第 1 位である実験参加者 A の 7 と，第 3 位で初めて正解として選ばれた実験参加者 B の 8 のデータをプロットした図を図 10 に示す．誤認識の原因となっている輪郭線の部位を探すために，指側と手首側に区分して再度実験を行った．その結果を表 5 に示す．全ての輪郭線で実験を行った場合は，実験参加者 A の 7 のデータの方が評価データに対して類似度が大きいと判定されたが，指側のみの輪郭線で実験を行った場合は，実験参加者 B の 8 のデータの方が類似度が大きくなった．この原因として，手話者クローズ実験の誤認識と同様に手首側の輪郭線の影響が考えられる．つまり，指側の輪郭線のズレの影響を相殺してしまう程に手首側の輪郭線が評価データに対して近い形であったと考えられる．手話の数字表現において数字の情報を持つのは指の形であるため，手首など情報を持たない部位は削除してマッチングを行う手法が考えられる．

図 10　評価データ C8 とテンプレートデータ第 1 位の A7 と第 3 位の B8

表 5　実験結果の比較

	実験結果の値		
	全ての輪郭線	指側のみの輪郭線	手首側のみの輪郭線
実験参加者 A の 7(テンプレートデータ第 1 位)	845.7	648.0	198.1
実験参加者 B の 8(テンプレートデータ第 3 位)	849.3	376.8	490.2

最後に，4 を 3 に誤認識する原因について考察する．この誤認識は，評価データである実験参加者 C の 4 に対して，実験参加者 D の 3 のテンプレートデータが第 1 位に選ばれたものである．手首部分の輪郭線の違いの影響を取り除くために，指側のみのデータを作成し改めて実験を行ったところ，表 6 に示すように評価データ 4 に対しても 3 が最も近いという結果が出た．評価データである実験参加者 C の 4 と，テンプレートデータ第 1 位である実験参加者 D の 3 と，第 16 位で初めて正解として選ばれた実験参加者 D の 4 のデータをプロットした図を図 11 に示す．図の中で，指の輪郭線の外側と内側が一致している箇所が確認できる．このように，指の本数が異なる場合でも指の外側と内側が一致することで，指の本数が異なる輪郭線データに誤認識してしまうことがあることが確認された．

表 6　実験結果の比較

	実験結果の値		
	全ての輪郭線	指側のみの輪郭線	手首側のみの輪郭線
実験参加者 D の 3(テンプレートデータ第 1 位)	601.3	495.6	198.3
実験参加者 D の 4(テンプレートデータ第 16 位)	848.8	632.5	217.9

図 11　評価データ C4 とテンプレートデータ第 1 位の D3 と第 16 位の D4

7. おわりに

本研究では手話の数字認識を行うため，Kinect によって取得した手領域の画像から輪郭線を抽出し，評価データとテンプレートデータの両方から相手の輪郭線に対して最近傍探索を行い，計算したユークリッド距離の合計を用いて識別を行う手法を提案した．その結果，手話者クローズ実験では 99.9%，手話者オープン実験(条件：標準化+回転)では 89.6%の認識率を得た．

今後の課題として，認識率を向上させるために特徴量に改良を加えることが挙げられる．具体的には，指の外側と内側が一致することで発生する誤認識を解消するために指の内側を塗りつぶす特徴量を導入することや，輪郭線の部位の中でも数字の情報を持たない手首側の部分を特徴量から除外することで，指側の輪郭線だけを特徴量として用いることが考えられる．

また, 本研究で用いた手法はパターンマッチングであり，評価データに対して近いデータが存在しない場合は認識に失敗してしまうという問題があった．今後の実験では，データ収録者数を増やすことで個人差に対応することが考えられる．

謝辞　実験に協力して頂いた手話サークルの方々に深く感謝いたします．また，本研究は文部科学省科学研究費補助金基盤研究(C)15K00223 の補助を受けています．

参考文献

[1] 谷端信彦，島田伸敬，“手話認識のための手指抽出と単語認識”，信学技報 WIT2001-22, pp.37-42，2001
[2] 佐藤新，篠田浩一，古井貞熙，“ToF カメラによる 3D 手話認識”，画像の認識・理解シンポジウム(MIRU2010)，IS3-44，pp.1861-1868，2010
[3] 森昭太，松尾直志，白井良明，島田伸敬，“手話認識のための距離情報を用いた隠蔽を含んだ顔・手領域抽出”，情報処理学会研究報告 IPSJ SIG Technical Report，Vol.2013-CVIM-186 No.20，pp.1-6，2013
[4] 井上快，齊藤剛史，“Kinect を利用した指文字認識に関する検討”，信学技報 MBE2012-81，pp.45-50，2013
[5] 織茂裕介，玉國裕司，高橋大介，岡本教佳，“Kinect を用いた指文字認識の検討”，映像情報メディア学会技術報告，ITE Technical Report Vol.38,No.9 ME2014-36，pp.31-32，2014
[6] Lukas Prasuhn，Yuji Oyamada，Yoshihiko Mochizuki，Hiroshi Ishikawa，“A HOG-based hand gesture recognition system on a mobile device”，2014 IEEE International Conference on Image Processing (ICIP)，pp.3973-3977，2014
[7] T. Kurita，S. Hayamizu，“Gesture recognition using HLAC features of PARCOR images and HMM based recognizer”，Proceeding of Third IEEE International Conference on Automatic Face and Gesture Recognition，pp.422-427，1998
[8] 青木俊和，野村昌輝，高塚崇文，田村仁，“RGB-D カメラを用いた指文字認識”情報処理学会第 75 回全国大会，4ZB-3，pp.4-187-188，2013
[9] 藤本光一，松尾直志，島田伸敬，白井良明，“輪郭部分特徴の階層構造学習による三次元手指姿勢推定の高速化”，IS3-64，pp.2007-2014，2010
[10] Zhou Ren，Junsong Yuan，Jingjing Meng，Zhengyou Zhang，“Robust Part-Based Hand Gesture Recognition Using Kinect Sensor”，IEEE TRANSACTIONS ON MULTIMEDIA,，VOL. 15, NO. 5，pp.1110-1120，2013