

機械学習を用いた空中手書き文字認識の検討*

花房 大佑^{*1}, 西田 好宏^{*2}

Studies on Aerial Handwritten Character Recognition Using Machine Learning

Daiyuu HANAFUSA ^{*1} and Yoshihiro NISHIDA ^{*2}^{*1} Department of Applied Science and Engineering

This paper presents a method to recognize a character handwritten in the air. We had prototyped an aerial handwritten hiragana character recognition system that detect single character period by hovering in the air using DP (dynamic programming) matching. In this time, we investigate an aerial handwritten character recognition system using machine learning. We use two features, vector and picture. Feature of vector is relative direction of handwriting information. It is a time series information so we use LSTM(Long Short Term Memory) to learn data of characters. Feature of picture is absolutely handwriting information. We use CNN(Convolution Neural Network) to learn data of characters, because this model is suitable for picture recognize. We compared correct recognition rate of using LSTM and using CNN. We estimated features of both methods.

Key Words : Character Recognition, Handwriting in the Air, Machine Learning

1. 緒 言

クラウド・コンピューティングやSNSの普及により、いつでも、どこでもふと思いついたアイデアなどをメモとしてクラウドに保存したり、出来事を呟きあったりする機会が増えている。そのためにはテキストデータの入力が必要不可欠であるが、パソコンのキーボードによる文字入力には両手での操作が必要になる。また、音声認識は、公共施設や騒音が激しい場所では利用が困難な場合がある。そのような背景の中、新たな文字入力方法の選択肢の一つとして、空中手書き文字入力システムの開発を行っている⁽¹⁾。

一方、一般的な手書き文字入力の歴史は古く郵便番号読み取りに代表されるオフライン文字認識からタブレット等に利用されたオンライン文字認識に発展して来た。現在においても中国語の手書き文字認識を中心に文字区切り処理⁽²⁾や学習のためのアノテーション⁽³⁾、手書きデータの前処理⁽⁴⁾などの研究が進んでいる。一般的な手書き文字入力は画数(ペンの上下)を文字の特徴として利用できるが、空中手書き文字入力は一筆書きで連続的に筆記された筆跡情報を基に文字の認識を行う点が大きく異なる。さらに、人が正確に空中に書くのは困難で、一般的に筆跡を検出するセンサの精度は良くないという問題も存在する。

空中手書き文字入力の研究事例として、園田らはビデオカメラとヘッドマウントディスプレイを使用して、手の動きを撮影し、連続DPマッチングを用いて文字の認識を行った⁽⁵⁾。また、大坪らは加速度センサを内蔵したWiiリモコンを振って文字の筆記を行い、マルチストリームHMMでモデル化することにより文字を認識した⁽⁶⁾。これらに対し、我々はLeap MotionやKinect, Myoを用いて取得した相対的な方向ベクトルを、端点フリーDPマッチングを用いて文字認識を行っている⁽⁷⁾。

現在、我々は空中手書き文字入力システムを利用して、病気や障害により対話を用いたコミュニケーションが困難な方を支援するシステムの開発を目指している⁽⁸⁾。今回は文字の入力方法と認識方法において、従来の手法とは異なる方法のフィジビリティスタディーを行う。具体的には車椅子の肘掛けの前に設置したカメラに向かって文字を筆記し、利用者が筆記を行う際の負担を減らし、利用者に合わせたカスタマイズを可能にするシステムについて検討する。また認識については、筆跡を画像情報として処理する場合と従来のDPマッチングと同様

* 原稿受付 2019年3月29日

^{*1} 大学院博士前期課程 応用理工学専攻^{*2} 工学部 電気電子工学科

E-mail: nishida@fukui-ut.ac.jp

に方向ベクトルの時系列情報として処理する場合において、それぞれ機械学習を用いて文字認識を行った。その結果、画像を用いた場合の認識率は 96.52[%]、方向ベクトルを用いた場合の認識率は 100[%]であり、先行研究⁽¹⁾の認識率 92.6[%]、先行研究⁽⁷⁾の認識率 76.8[%]を共に上回ることを確認した。

2. 今回の空中手書き文字入力システム

2.1 文字の入力方法

空中に文字を筆記する様々な方法についての検討がなされている⁽⁹⁾。先行研究において、空中に文字を入力する際には加速度や筋電位を取得する装置や、指先や手の形を検出するセンサ付きのカメラが必要であった。本検討では特殊な装置を使用せずに、一般的な web カメラと色情報を使用して筆跡の取得を行う。web カメラを使用する場合は画面を見ながら筆記する必要があるが、座標の変換や重力加速度の除去が不要で、筆記者が装置を身に着ける必要がないという利点がある。またカメラ本体は比較的安価で、容易に入手することができるので誰でも気軽にシステムを使用することが可能である。

2.2 文字の認識方法

先行研究において、文字認識は相対的な移動方向情報を利用して行われていた。入力された一筆書きの文字を筆記方向に応じて 8 種類の方向コード列に変換し、端点フリー DP マッチングを用いて、辞書データと比較することにより文字の認識を行う。理由としては、アルゴリズムが容易であること、文字を 8 種類のコードのみで表現するため、比較的単純な DP マッチングでも十分な認識率が期待できること、端点フリーの為、文字の書き始めのノイズを軽減できることが挙げられる。一方この方法は、辞書データの作成は容易ではなく、利用者ごとの筆記の癖に対応しきれないといった弱点が見られた。そこで、本検討では機械学習を用いて入力された文字の学習・判定を行う。機械学習を使用することにより、入力された文字の特徴を捉えた学習データの作成が可能となり、利用者に応じたカスタマイズが容易になることが予想される。

従来までの文字認識は、相対的な移動方向のみを使用して行われていた。その理由は、文字の大きさや筆記する場所に関わらず文字の判定ができるからである。しかしながら、相対的な移動方向情報のみで文字認識を行うと、異なった書き順で文字を筆記した場合に正しく判定できない。そこで本検討では従来の相対的な移動方向情報に加えて、絶対的な座標情報である画像を用いた文字の判定も行う。双方の特徴を捉えそれぞれの利点を知ることによって、空中手書き文字入力システムの発展を目指す。

今回のシステム概要を Fig.1 に示す。利用者はカメラに向かって文字を空中に筆記し、画像処理を用いて筆跡の座標を記録する。その座標を基に筆跡を、絶対的な情報である画像と、相対的な情報である方向ベクトルに分割する。それぞれの情報はあらかじめ学習された筆跡情報と比較して、入力された文字の推定を行う。本検討では画像を用いた文字認識には CNN を、方向ベクトルを用いた文字認識には LSTM を使用した。

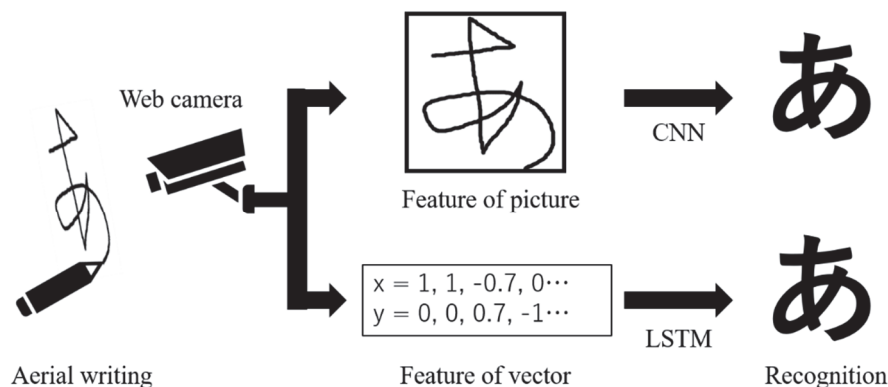


Fig.1 Aerial Handwritten Character Recognition

3. アプリケーションの開発及び筆跡データの収集

3.1 文字入力アプリケーションの開発

始めに空中に一筆書きで筆記した筆跡情報を出力・保存するアプリケーションの開発を行った。利用者はカメラに向かって文字を空中に筆記し、画像処理を用いて筆跡の座標を記録する。画像処理には OpenCV を用いた。

筆跡座標の取得方法を Fig.2 に示す。筆跡の取得には色情報を使用した。初めに特定の色（今回は緑色）を白色に、その他の色を黒色に置き換える二値化処理を行う (①, ②)。この際にノイズの除去を行った。次に画面上の白色の部分から、面積が最大となる部分を選択して、中心座標を記録する。記録した座標を使用して、筆跡を画像と時系列の方向ベクトルの 2 つの情報として出力を行った。

筆跡を画像として出力する際は、現在の筆跡座標と、1 フレーム前の筆跡座標を結ぶ線を描画する。この処理をフレーム毎に繰り返し (③)、最後に筆跡のみを画像として出力する (④)。出力する画像は Fig.3 に示すように、28 ピクセル×28 ピクセルの大きさに統一した。

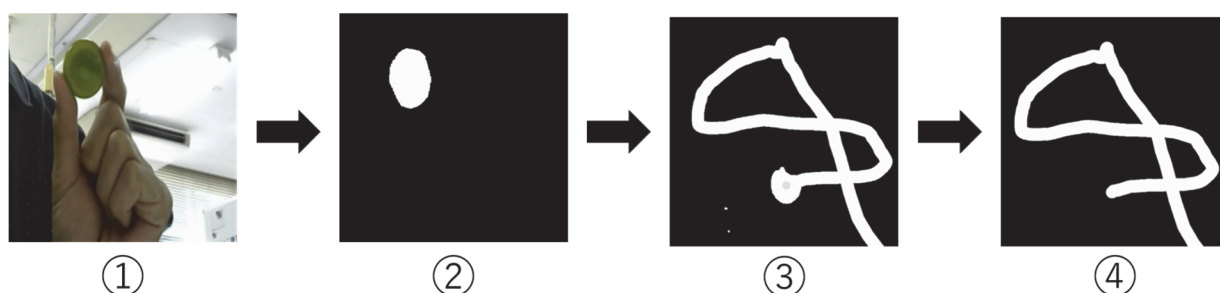


Fig.2 Track a Handwriting

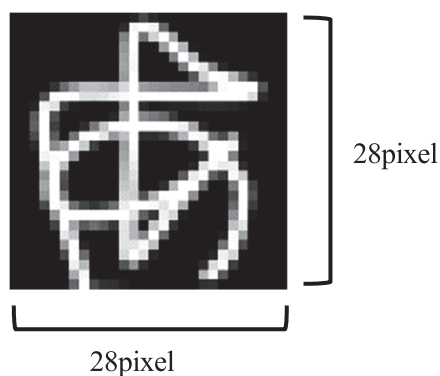


Fig.3 Output Picture

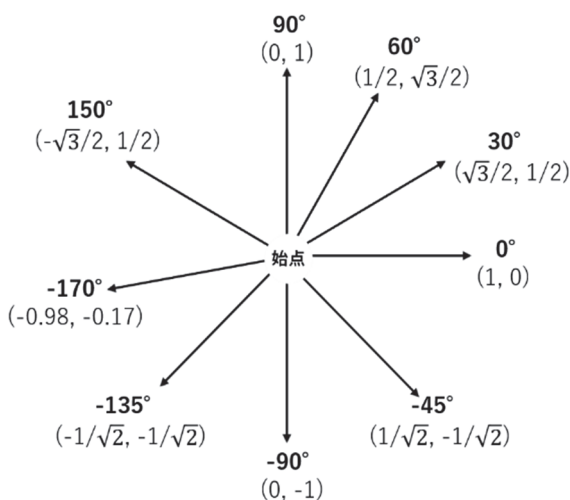


Fig.4 Track a Handwriting

筆跡を方向ベクトルとして出力する際、最初に現在の筆跡座標と、1 フレーム前の筆跡座標の角度(θ)を求める。次に求めた角度を使用して X 軸方向のベクトルデータ($\cos\theta$)と、Y 軸方向のベクトルデータ($\sin\theta$)をそれぞれ csv ファイルとして出力する。座標と方向ベクトルの関係の一例を Fig.4 に示す。例えば現在の座標と 1 フレーム前の座標の角度の差が 60° である場合、X 軸方向のベクトルデータは $1/2 (=0.5)$ 、Y 軸方向のベクトルデータは $\sqrt{3}/2 (=0.87)$ となる。具体例としてひらがなの「い」を 6 分割してベクトルを出力する場合は、初めに fig.5 に示すように現在の座標と 1 つ前の座標の角度の差を算出する。例の場合、角度の差を順に挙げると、 -90° 、-

45° , 60° , 60° , -45° , -90° となる．次に，算出した角度の差に応じた X 軸方向と Y 軸方向それぞれの方向ベクトルを求める．例の場合における方向ベクトルは Table.1 に示すように， $(X, Y) = (0, -1), (0.71, -0.71), (0.5, 0.87), (0.5, 0.87), (0.71, 0.71), (0, -1)$ となる．例では方向ベクトル数はそれぞれ 6 個であるが，実際はすべての文字において，方向ベクトルの数を 100 個とした．

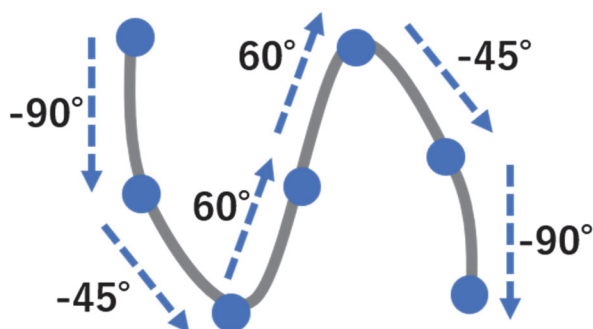


Fig.5 Vector “い”

Table.1 Vector Code “い”

| X | Y |
|------|-------|
| 0 | -1 |
| 0.71 | -0.71 |
| 0.5 | 0.87 |
| 0.5 | 0.87 |
| 0.71 | -0.71 |
| 0 | -1 |

空中で文字を筆記する際には，文字と文字を区切る動作が不可欠である．今回は Fig.6 に示すように，画面下部に境界線を設定し，境界線の下部から筆記を開始する事とした．筆記後は筆記点を，境界線の下部に再び移動させることで文字の区切りを行う．この方法を採用した理由は，この方法は利用者に特別な動作を強いることなく，負担が少ない筆記が可能になると考えたからである．画像およびベクトルデータの出力は，文字区切りと同時にいった．

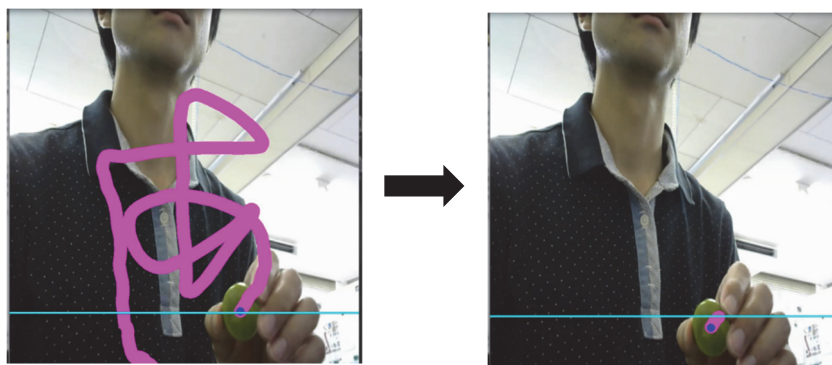


Fig.6 Separate Character

3.2 筆跡データの出力

出力された画像と方向ベクトルデータの双方を用いて機械学習を行い，文字認識の検討を行う為に作成した空中手書きアプリケーションを用いて筆跡情報を収集した．今回の認識対象は，濁音・半濁音を除いた，ひらがな 46 文字とした．被験者 1 名が先端に色の付いたペンを使用して 46 文字をそれぞれ 50 回ずつ，計 2300 回筆記を行い，筆跡情報を画像とベクトルデータとして出力した．

4. 文字認識

4.1 筆跡の分類とデータセットの作成

出力された画像と方向ベクトルデータを用いて、機械学習と文字の判定を行う。学習・認識には Sony が提供しているディープラーニング・ツールの Neural Network Console を使用した。例として画像のデータセット作成について説明する。作成した筆跡情報を、Fig.7 に示すように文字ごとに分類してフォルダに入れる。その後、筆跡データの 8 割を学習用、2 割を評価用に分割して、筆跡情報のファイルパスと正解のラベルを一覧にした csv ファイルを作成する。この際にファイルの順番はランダムに並べ替えた。正解ラベルはフォルダ名と同じものを割り当てる。csv ファイルの一例を Fig.8 に示す。x:image は筆跡情報のファイルの保存場所を、y:label は正解の文字番号を表す。同様に方向ベクトルデータの分類とラベリングを行った。

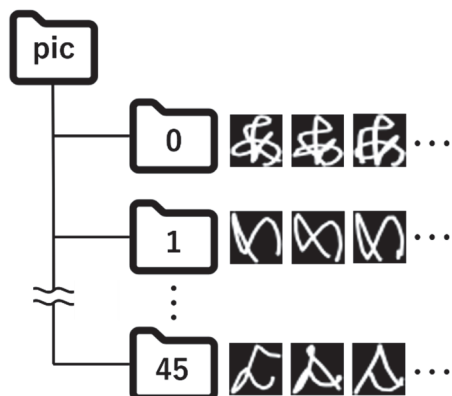


Fig.7 Data Set

| x:image | y:label |
|-------------------------------|---------|
| ¥33¥19_0603_1016_50405694.png | 33 |
| ¥40¥19_0604_0859_04621181.png | 40 |
| ¥20¥19_0603_1048_10983465.png | 20 |
| ¥17¥19_0604_0951_05262838.png | 17 |
| ¥44¥19_0604_0906_13245175.png | 44 |
| ¥24¥19_0604_0919_11863428.png | 24 |
| ¥40¥19_0604_0901_53026680.png | 40 |
| ¥29¥19_0604_0947_05896156.png | 29 |
| ¥0¥19_0604_0932_18076052.png | 0 |
| ¥26¥19_0604_0906_55902856.png | 26 |
| ¥39¥19_0604_0909_33403193.png | 39 |

Fig.8 Data List

4.2 画像を使用した文字認識

一般的に、画像に対しては CNN が用いられることが多い。そのため今回の学習には CNN の基本的なモデルを採用した。具体的には、1998 年に提案され、CNN の先駆けとなった LeNet⁽¹⁰⁾ を基にモデルを作成した。4.1 で作成した学習用と評価用それぞれのデータセットを読み込むと、Fig.9 のように画像と正解ラベルの一覧が表示される。

| Index | x | y |
|-------|--|----|
| 1 | ¥4¥19_0604_0906_55902856.png c, 28, 28 | 4 |
| 2 | ¥42¥19_0604_0906_55902856.png c, 28, 28 | 42 |
| 3 | ¥0¥19_0603_1048_10983465.png c, 28, 28 | 0 |
| 4 | ¥0¥19_0604_0901_53026680.png c, 28, 28 | 0 |
| 5 | ¥38¥19_0604_0947_05896156.png c, 28, 28 | 38 |

Fig.9 Picture List

| Index | x | y |
|-------|---|----|
| 1 | ¥4¥19_0717_1236_43102665.csv 100, 2 | 4 |
| 2 | ¥33¥19_0717_1245_41652589.csv 100, 2 | 33 |
| 3 | ¥4¥19_0717_1337_34951171.csv 100, 2 | 41 |
| 4 | ¥28¥19_0717_1240_1621136511.csv 100, 2 | 22 |
| 5 | ¥41¥19_0717_1236_43102665.csv 100, 2 | 41 |

Fig.10 Vector List

その後、Neural Network Console のネットワーク自動最適化機能を利用して、ネットワーク構成とパラメータが異なるモデルを複数作成して、その中で最も精度の高いモデルを採用した。学習に使用した CNN のモデルを Fig.11 に示す。各層横の数字はノード数を表す。使用する画像は、サイズが縦横 28 ピクセルのグレイスケール画像なので入力層が $28 \times 28 \times 1$ ノードとなる。中間層は、畳み込み層とプーリング層が 2 層ずつ、全結合層が 1 層、出力層が 46 ノード（認識対象のひらがな 46 文字）から構成される。中間層、出力層では活性化関数に ReLU 関数を使用しており、出力層ではさらにドロップアウトを行い、誤差関数に交差エントロピーを使用した。テスト用データを用いて文字の評価を行ったところ、認識率は 96.52[%]であった。

4.3 方向ベクトルを使用した文字認識

方向ベクトルは、時系列情報によって構成されているので、学習モデルは時系列データを扱うことのできる RNN(Recurrent Neural Network)の中の LSTM を使用した。LSTM を用いたモデルにおいても CNN の時と同様に基本的なモデルとして、LSTM 層が 1 層の簡単なモデルで検討を行った。4.1 で作成した学習用と評価用それぞれのデータセットを Neural Network Console に読み込むと、Fig.10 のように方向ベクトルと正解ラベルの一覧が表示される。その後、Neural Network Console のネットワーク自動最適化機能を利用して、ネットワーク構成とパラメータが異なるモデルを複数作成して、その中で最も精度の高いモデルを採用した。学習に使用した LSTM のモデルを Fig.12 に示す。各層横の数字はノード数を表す。使用したベクトルデータは X 軸、Y 軸に対してそれぞれ 100 個なので、入力層が 100×2 ノードとなる。中間層は LSTM 層とプーリング層が 1 層ずつ、出力層が 46 ノード（認識対象のひらがな 46 文字）から構成される。出力層では誤差関数に交差エントロピーを使用した。テスト用データを用いて文字の判定を行ったところ、認識率は 100[%]であった。

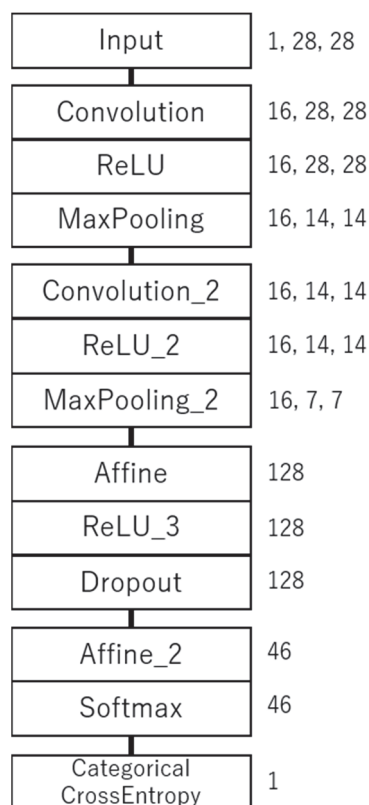


Fig.11 CNN Model

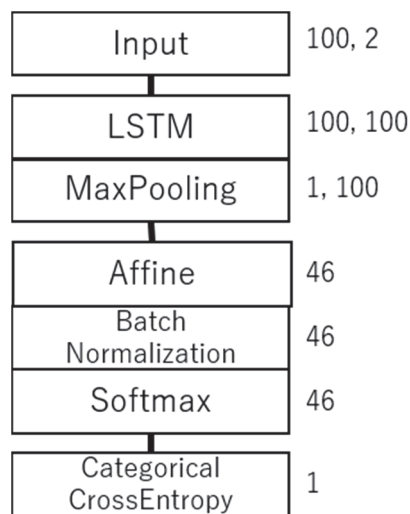


Fig.12 LSTM Model

5. 検討と考察

本検討では従来のセンサや特殊なカメラを使用せずに、一般的な web カメラと先端に色の付いたペンを使用して、空中の筆跡取得を行った。その結果、従来よりも手軽に筆記できることが確認できた。しかし色情報を用いた場合は、原理上光の反射や背景の色によってノイズが生じる問題があるため、色情報を用いない筆跡取得方法についても引き続き検討を行う。

空中に筆記した筆跡を画像データとベクトルデータに分類して、それぞれに対して学習・判定を行った。その結果、いずれの場合においても高い文字認識率を確認できた。しかしながら、本検討では被験者 1 名が丁寧に筆記した場合の筆跡情報を用いて学習と判定を行ったため、筆記した文字の形がほぼ同一であった。よって、理想的な状況においては、十分な認識精度があることが確認できた反面、画像を用いた場合とベクトルデータを用いた場合それぞれの長所・短所の比較はできなかった。今後は多くの人の筆跡データを収集して、それぞれの手法の特徴を理解することを試みる。

本検討では、ディープラーニングを用いて空中に筆記した文字の認識を行った。先行研究における DP マッチングを用いた文字認識と比較すると、あらかじめ辞書データを作成する必要が無く、利用者の筆記方法や筆記の癖に応じたカスタマイズが容易であることが考えられるので、今後も積極的に活用していきたい。

6. 結 言

空中手書き文字入力システムにおいて、文字入力にカメラ画像を使用し、文字認識にディープラーニングを採用した場合のフィジビリティスタディーを行った。その結果、一般的な web カメラを用いて例えば指先に所定の色の指サックを付けることで、手を持ち上げて空中で筆記して手を下すという一連の動作で負担の少ない文字入力が実現できることを確認した。また、認識方法については従来の DP マッチングと同様の方向ベクトルの時系列情報を用いた方式で高い認識率が得られることを確認した。今後は被験者及び認識文字パターンを増やし筆跡のデータセットを増強していくことが必要である。今回の結果を基にシステムの改良を重ね、それぞれの利点を生かした文字入力・認識方法の確立を目指す。その他にも転移学習・アンサンブル学習を用いることで空中手書き文字入力システムの更なる発展を目標とする。

文 献

- (1) 西田好宏, “二次元移動方向に基づく空中手書きひらがな文字認識に関する研究”, 和歌山大学大学院博士後期課程学位論文, 甲第 44 号 (2011).
- (2) R. Messina and J. Louradour, “Segmentation-free handwritten Chinese text recognition with LSTM-RNN”, Proc. Int’l Conf. Document Analysis and Recognition (ICDAR), (2015).
- (3) SUN, Li et al, “Deep LSTM networks for online Chinese handwriting recognition”, In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). IEEE, pp. 271-276 (2016).
- (4) ZHANG Xu-Yao, BENGIO Yoshua, LIU Cheng-Lin, “Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark”, Pattern Recognition, 61. pp. 271-276 (2017).
- (5) 園田智也, 村岡洋一, “空中での手書き文字入力システム”, 電子情報通信学会論文誌, vol.J86-D-II, no.7, pp. 1015-1025 (2003).
- (6) 大坪由香利, 中井満, “ペン先方向の加速度特徴を活用した空中手書き文字認識”, 電子情報通信学会論文誌, vol.J97-D, no.3, pp. 574-583 (2014).
- (7) 吉田大志, 西田好宏, “センサを利用した空中手書き文字入力システム及び文字区切りに関する研究”, 平成 26 年度福井工業大学修士論文 (2019).
- (8) 花房大佑, 西田好宏, “空中手書き文字入力システムを用いたコミュニケーションツールの作成”, 平成 30 年度電気関係学会北陸支部連合大会, F2-13 (2018).
- (9) 中井満, “より自然な空中筆記インタフェースを目指して”, 平成 29 年度電気関係学会北陸支部連合大会, I-1 (2017).
- (10) Y LeCun et al, “Gradient-Based Learning Applied to Document Recognition”, Proc of the IEEE, Vol.86. pp2278-2324 (1998).

(2019 年 4 月 26 日受理)