

Imaging Highlight

日本語の指文字を用いたタイピングシステムのプロトタイプ構築

井上勝文*

(2019.4.9 受理)

Prototype Construction of Typing System with Japanese Finger-spelled Sign Language

Katsufumi INOUE*

Japanese Finger-spelled Sign Language is a visual language and this represents Japanese syllabary characters called "Hiragana". To assist the communication with hearing-impaired people, many researchers have already been proposed automatic finger-spelled sign language recognition systems. In this paper, we introduce an example of that recognition system which includes a recognition method with hand shape information that is extracted from a hand region by using a RGB-D camera. In addition, we introduce a spotting technique with SVM and rule-based approaches and show a prototype of typing system with Japanese finger-spelled sign language by converting a Hiragana word into its Kanji representation.

Keywords: Japanese finger-spelled sign language, RGB-D camera, Typing system, Spotting, Gesture recognition

日本語の指文字は、ひらがな 50 音を指の形で表現する視覚言語の一種で、手話単語にない単語を表現するのに使用されている。このため、指文字を自動認識できるようになると、聴覚障がい者とのコミュニケーション促進につながる事が期待されている。本稿では、これの実現を目指している、日本語の指文字の自動認識技術の一事例について紹介する。具体的には、RGB-D カメラを用いて手領域を抽出し、手領域の形状等の情報から表現されている指文字を認識する技術と、連続表現された指文字を一文字一文字に分割するスポッティング技術について紹介する。また、その応用事例として、指文字で表現されたひらがな文字を漢字変換して提示するタイピングシステムについて紹介する。

キーワード：日本語の指文字、RGB-D カメラ、タイピングシステム、スポッティング、ジェスチャ認識

1. はじめに

日本の社会問題の一つとして、聴覚障がい者の問題がある。厚生労働省の調査によると、日本の聴覚障がい者の数は 30 万人以上にのぼると報告されている¹⁾。聴覚障がい者とのコミュニケーション手段の一つとして、指文字がある。日本語の指文字は、ひらがな 50 音を指の形で表す視覚言語の一種で、手話単語にない単語などを表現するために使用されている。このため、指文字を自動認識できるようになれば、健常者と聴覚障がい者とのコミュニケーションが促進されることが期待されている。

日本語の指文字は、Fig. 1 に示す通り大きく分けて静止文字と動作文字に大別できる。前者は、手を動かさずに指の形だけで表現され、後者は、手を動かして表現される。指文字の自動認識を実現するには、これら一文字一文字を正確に認識する必要がある。また、指文字で表現された単語を認識するには、連続表現された指文字を一文字一文字にスポッティングする必要がある。これまでの研究では、指文字一文字毎の認識に焦点を当てたものが多く^{2,3)}、その先に話を進めているものは少ない。そこで、RGB-D カメラと簡単なアルゴリズムを用いて、連続表現された指文字をスポッティングし、漢字変換するタイピングシステムのプロトタイプを作成してみた⁴⁾ので、本稿ではこれについて紹介する。なお本システムは、大阪府立大学大学院工学研究科倫理委員会の承認を得て作成している。

* 大阪府立大学 大学院工学研究科
〒599-8531 大阪府堺市中央区学園町 1-1

* Graduate School of Engineering, Osaka Prefecture University
1-1, Gakuencho, Naka, Sakai, Osaka 599-8531, Japan

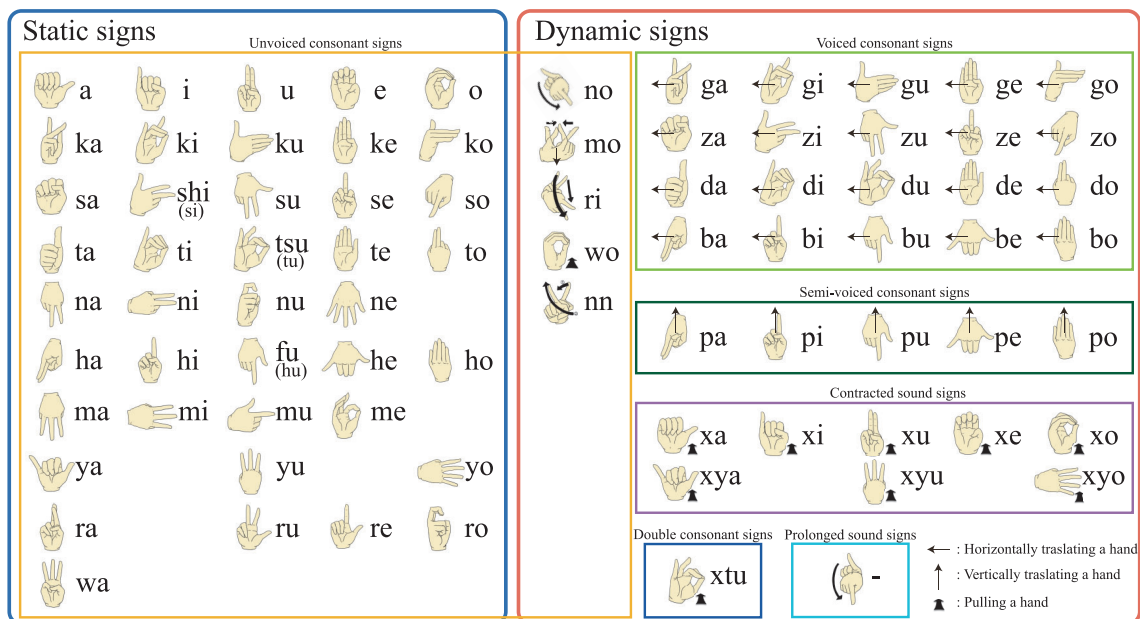


Fig. 1 Hand signs of Japanese finger-spelled sign language.

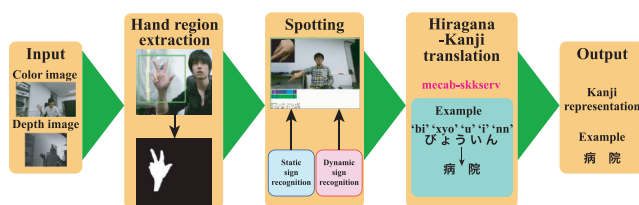


Fig. 2 Overview of the processing of prototype typing system.

2. 指文字認識システム及び タイピングシステムの紹介

本稿で紹介する日本語の指文字認識システム及びタイピングシステムの処理の流れを Fig. 2 に示す。まず、RGB-D カメラを用いて手領域を抽出し、得られた領域から抽出された画像特徴量や手の動作情報を基に静止文字と動作文字を認識する。またこれらの情報を基に連続表現された指文字を一文字一文字にスポットティングする。そして得られたひらがな列を漢字変換し提示する。以下で具体的な処理について概説する。

2.1 手領域抽出手法

日本語の指文字を表現する際に、人物や表現のタイミングによって手や腕の見え方が大きく異なる。指文字認識では、手や指の形が重要であるため、腕の情報は認識処理においてノイズとなる。そこで、まず RGB-D カメラを用いて手領域を正確に自動抽出する⁵⁾。ここで、本稿では、指文字表現者は右手で表現していることを前提に話を進める。

まず、RGB-D カメラから得られる距離情報を基に人物骨格情報を抽出する。この情報から右手付近の画像を得る。そして距離情報を基に背景を除去し、頭と手の距離情報より手と腕を含む領域を切り出す。本稿ではこの領域を手腕領域と呼ぶ。そ

して得られた領域に対して、Time-Series Curve⁶⁾を用いて領域の輪郭情報より手首点や指先点などの情報取得し、これを基に腕の領域を削除した手のみの領域を取得する²⁾。

2.2 静止文字認識とそのスポットティング

前述の処理で得た手領域画像から HOG 特徴量を抽出し、one-v.s.-one ルールの SVM で静止文字の特徴量を学習する。日本語の指文字における静止文字は 41 種存在するので、合計で 820 個の SVM が構築される。認識処理では、820 個の SVM を調査し、最も得票数の多い指文字を認識結果とする。上述の処理をフレーム画像毎に行うことで、表現されている指文字の認識結果列が作成される。

次に、得られた認識結果列を基に、同一文字が認識されている区間を、その文字が表現されていた区間としてスポットティングする⁷⁾。この際に、ノイズの影響で表現されている静止文字を誤認識する可能性があるが、予備実験よりその頻度は少ないことが分かっている。このため、スポットティングされた区間が短い場合、前後の区間の文字が同じ場合はその文字の区間とし、異なる場合は文字の切り替え途中などと判断し認識結果をリジェクトする。

2.3 動作文字認識とそのスポットティング

具体的な処理の前に、動作文字について更に細かく分類する。動作文字には、「の」、「も」、「り」、「を」、「ん」に加え、長音や濁音・半濁音や拗音・促音のような小書き文字が含まれる。本稿ではこの内、「の」、「り」、「ん」、長音のように手首を返して表現されるものを「非平行移動文字」、残りの文字のようにベースとなる静止文字を横や上、手前へ平行移動させて表現されるものを「平行移動文字」と呼ぶ。以下で具体的な処理について概説する。

まず、各フレーム画像間の手の重心の移動量 $D_i(x, y, z)$ を求める。そしてこれが閾値 Φ_1 を超えた場合、手が移動したと判定し、この時刻を T とする。時刻 T 前後において、 $D_i(x, y, z)$

が閾値 Φ_2 を下回る直近の時刻 (T_s, T_e) とし、この時刻のフレーム画像 F_s と F_e を取得する。次に、これら2つのフレーム画像を用いて、非平行移動文字か平行移動文字かを判定する。これを実現するために、非平行移動文字は、手首を返して表現するため、動作前後で指の方向が変わるのに対し、平行移動文字は指の方向が変わらない、という情報を活用する。時刻 (T_s, T_e) における手腕領域を取得し、背景を黒、それ以外を白とした画像をそれぞれ得る。そして、細線化処理を用いて、手腕領域の中心線を取得する。指文字は手を出して表現するため、中心線の端点の内、RGB-D カメラから遠い点を肘点、その反対の端点を指先点とする。これらの端点を用いることで指の方向の変化を判定する。この時変化が大きい場合は表現されている文字は非平行移動文字、小さい場合は平行移動文字と区別する⁸⁾。

もし非平行移動文字と判定された場合、時刻 T_e における指の方向を基に、表現されている文字を認識する。この時ユーザ側から見て右方向の場合「ん」、下方向の場合「長音」と認識する。また、この方向が左方向で、指先が1つの場合「の」、2つの場合「り」と認識する。

一方、表現された文字が平行移動文字の場合、手の重心の移動方向で判定する。具体的には、ユーザ側からみて右方向の場合濁音、上方向の場合半濁音、手前方向の場合小書き文字、下方向の場合「も」と判断する。さらに、前者3つにおいて、静止文字認識と同様に F_s の文字を認識することで具体的な文字を認識する。

動作文字のスポッティングについては、上述の処理において求めた F_s と F_e の区間を動作文字が表現された区間としてスポッティングすれば良い⁷⁾。

2.4 ひらがな—漢字変換処理

上述の処理でスポッティングされて得たひらがな文字列を漢字変換する。本プロトタイプのタイピングシステムでは、MeCab⁹⁾に基づく漢字変換エンジン mecab-skkserv¹⁰⁾を使用する。Fig. 3 に「あした」と指文字で表現した際のシステムの動作例を示す。本タイピングシステムでは、プロトタイプとして mecab-skkserv で漢字予測変換された第一候補のみを提示している。

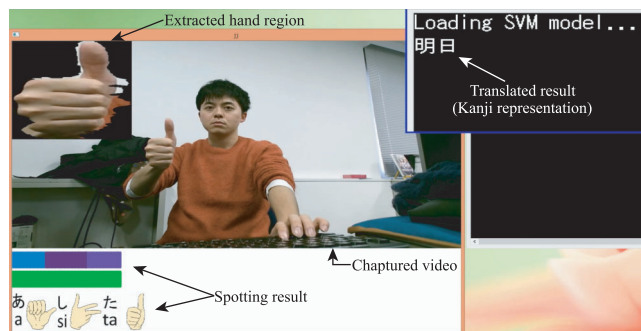


Fig. 3 Example of implementation of the prototype typing system.

3. タイピングシステムの評価

本稿で紹介したプロトタイプシステムを簡易的な実験で精度評価とユーザビリティ評価を行って見たので、その結果について紹介する。

3.1 実験条件

本システムでは、RGB-D カメラとして Microsoft 社が提供している Kinect for Windows v2 を用いた。本実験では、実験参加者に Kinect から約 1 [m] 離れた位置で指文字を表現してもらった。撮影は、学習用データとテスト用データ取得の2つのパターンで行った。学習用データでは、静止文字認識用の SVM を構築するために、各静止文字を 16 名の実験参加者にそれぞれ 10 回表現してもらい撮影した。一方テスト用データ撮影では、Table 1 で示す 8 つの単語を 6 名の実験参加者 (内 4 名は学習用データ撮影にも参加) にそれぞれ 10 回表現してもらった。動作文字認識に用いる閾値 Φ_1 は、水平方向が 150、奥行方向が 80 とした。また閾値 Φ_2 は 20 とした。

3.2 精度評価

本システムの精度評価するために、文字分割成功率・単語認識成功率・単語内文字認識成功率の3つの指標を用いる⁵⁾。文字分割成功率は、表現された単語を構成する文字数と同じ指文字数が検出できた割合、単語認識成功率は、各単語を構成する文字と認識された文字が全て一致する割合である。また単語内文字認識率は、表現された文字毎の認識精度である。Fig. 4 に単語毎の評価結果を示す。8 つの単語に対する平均文字分割成功率、平均単語認識成功率、平均単語内文字認識率は、それぞれ 75.6%, 49.2%, 82.4% だった。

Fig. 4 より、「はっぴょう」などの文字数、動作文字数が多い単語では、一文字毎の精度は悪くないものの手の切り替え時のノイズによりスポッティング精度に課題が残った。これは、動作文字のベースとなる静止文字が、他の静止文字と類似しているため SVM で正しく認識できなかったことが挙げられる。「はっぴょう」の場合では、「ち」と「つ」が類似している影響が大きいと思われる。また、「じゅず」は文字数が少ないにも

Table 1 List of testing words and character class transition (S: static signs, V: voiced consonant signs, SV: semi-voiced consonant signs, C: contracted sound signs, NT: non-translation signs).

Words	Transition of character classes
a shi ta	S->S->S
re shi pi	S->S->SV
zi xyu zu	V->C->V
u do nn	S->V->NT
ri xyo u ri	NT->C->S->NT
ha xtu pi xyo u	S->C->SV->C->S
a nn ze nn pi nn	S->NT->V->NT->SV->NT
shi xyu xtu ti xyo u	S->C->C->S->C->S

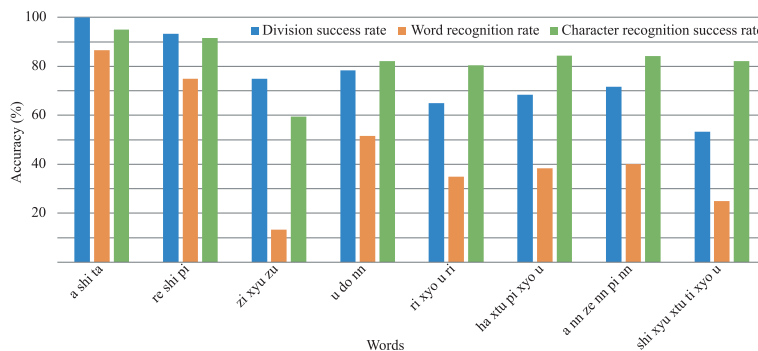


Fig. 4 Experimental results of the performance of the prototype typing system.

Table 2 List of questionnaire for evaluating the usability of the prototype typing system.

No.	Question
Q1	The spotting process is smooth.
Q2	The constraint for user to use this system is not severe. (Distance between camera and user, captured angle, etc.)
Q3	Hearing-impaired people can communicate with healthy people by using this system.
Q4	This system can be an alternative communication tool for hearing-impaired people compared with writing message, keyboard, etc.

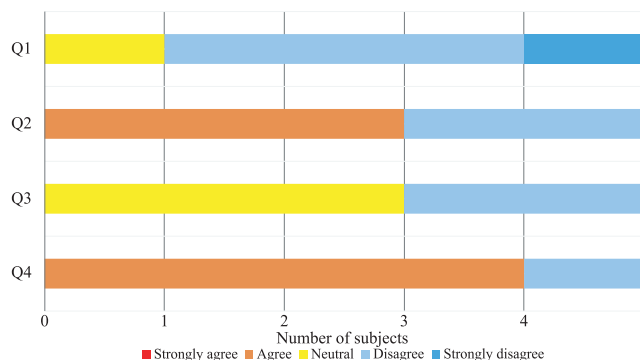


Fig. 5 Result of questionnaire with 5 subjects.

関わらず単語内文字認識率にも課題が残った。構成される動作文字の動作方向が全て異なるため、文字切替時の動作が他の動作文字と誤認識したことが挙げられる。このため、今後文字表現時と文字切替時の動作を区別するアルゴリズムを考える必要がある。

上述の通り、まだ精度に難はあるが、「あした」や「れしぴ」といった文字数や動作文字数が少ない単語では、認識・スポットティングできていることから、プロトタイプとしてある程度のベースラインを示すことができたのではと考える。

3.3 ユーザビリティ評価

本プロトタイプシステムのユーザビリティを評価するために、テスト用実験参加者 5 名に Table 2 に示す計 4 問のアンケートをとった。各アンケートは 5 段階で評価している。Fig. 5 に結果を示す。Fig. 5 より、システムの処理精度や聴覚障がい者との円滑な意思疎通の実現性についてまだ否定的な意見が多いため、システムを改善する必要がある。一方、システムの制約条件や他の手法の代替手段としては、肯定的な意見が多かったため、システムの柔軟性や必要性を確認することができた。このことからプロトタイプとしてある一定のユーザビリティ性を有していることを確認できたと考えられる。

4. おわりに

本稿では、日本語の指文字におけるタイピングシステムに関して、RGB-D カメラを用いたプロトタイプシステムの構築方法について紹介した。簡易実験より、本システムは SVM やルールベースの認識手法といった単純な手法のみでも、文字分割成功率、単語認識成功率、単語内文字認識率それぞれ 75.6%、49.2%、82.4% を達成した。またユーザビリティ評価より、筆談などの代替手段としての必要性を確認した。ただ、精度やユーザビリティ性についてはまだまだ改善の余地が多々残っているので、本稿を通じて聴覚障がい者の問題に少しでも関心を持ってもらえれば幸いである。そして、新たなシステムが誕生することにより、将来的に聴覚障がい者との円滑なコミュニケーションが促進されることに少しでもつながって貰えれば幸いである。

謝 辞

本プロトタイプ作成は、科学研究費補助金（若手研究（B）JP16K21287）の補助による。また、本プロトタイプ作成は、白石孝弥氏（元大阪府立大学）、松岡遼氏（元大阪府立大学）、丸山瑞己氏（現大阪府立大学）の協力によるもので、ここに感謝の意を表す。

参 考 文 献

- 1) <http://www.mhlw.go.jp/toukei/saikin/hw/shintai/06/>, (accessed 2019-03-25).
- 2) J.C. Terrillon, S.A. Pilpre, Y. Niwa, and K. Yamamoto, "Robust Face Detection and Japanese Sign Language Hand Posture Recognition for Human Computer Interaction in an Intelligent Room," Proceedings of 15th International Conference on Vision Interface (2002), pp. 1-8.
- 3) K. Inoue and T. Saitoh, "Real-time Finger Character Recognition using Range Image," Proceedings of IEICE Technical Report, IEICE (2012), pp. 87-92 [in Japanese].
- 4) R. Matsuoka, K. Inoue, and M. Yoshioka, "Investigation of a Typing System with Japanese Finger-spelled Sign Language," Proceedings of 2019 Young Researcher Symposium of Kansai Section of The Society of Instrument and Control Engineers and The Institute of Systems, Control and Information Engineers, The Society of Instrument and Control Engineers Kansai Section and The Institute of Systems, Control and Information Engineers (2019), pp. 121-124.
- 5) K. Inoue, T. Shiraishi, M. Yoshioka, and H. Yanagimoto, "Depth Sensor Based Automatic Hand Region Extraction by Using Time-Series Curve and Its Application to Japanese Finger-spelled Sign Language," Journal of Procedia Computer Science, **60**, pp. 371-380 (2015).
- 6) E.J. Keogh, L. Wei, X. Xi, S.H. Lee, and M. Vlachos, "LB_Keogh Supports Extract Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures," Proceedings of the 32nd International Conference on Very Large Data Bases, the VLDB Endowment (2006), pp. 886-893.
- 7) K. Inoue, T. Shiraishi, R. Matsuoka, and M. Yoshioka, "Spotting Analysis of Sequential Representation of Japanese Finger-spelled Sign Language with RGB-D Camera," Journal of IEEJ Transactions on Electronics, Information and Systems, **138**, pp. 1230-1241 (2018) [in Japanese].
- 8) K. Inoue, T. Shiraishi, R. Matsuoka, and M. Yoshioka, "Investigation of Japanese Dynamic Finger-spelled Sign Language Recognition with RGB-D Camera," Proceedings of 22nd Korea-Japan Joint Workshop on Frontiers of Computer Vision, IEEE Nagoya Section (2016), pp. 95-100.
- 9) <http://taku910.github.io/mecab/>, (accessed 2019-04-05).
- 10) <http://www.chasen.org/~taku/software/mecab-skkserve/>, (accessed 2019-04-05).



井上 勝文

2012年大阪府立大学大学院工学研究科博士後期課程(電気・情報系専攻)を修了。博士(工学)。2010年日本学術振興会特別研究員(DC1)。2012年同特別研究員(PD)。2012年パナソニック株式会社に入社。2014年大阪府立大学大学院工学研究科助教。2019年同研究科准教授。主として、パターン認識、画像センシング、物体認識、信号処理などの研究に従事し、現在に至る。2010年電気関係学会関西連合大会奨励賞、ACIS2016 Best Paper Award 各受賞。電子情報通信学会、電気学会、IEEE 各会員。