

ヘッドマウントディスプレイにおける 指文字を用いた文字入力手法の提案

久保田千尋^{†1} 湯村翼^{†1}

概要：ヘッドマウントディスプレイの普及において、文字入力手法が課題である。指先などでカーソルを操作し仮想キーボードを操作する手法が主流であるものの、手の認識に深度センサや複数台のRGBカメラが必要である。そこで本研究では、MediaPipe Handsを用いることによりRGBカメラ1台で手の形状を推論し、手話の一種である指文字によって文字を入力するシステムを提案する。本稿ではアメリカ手話(ASL)の指文字に対応するプロトタイプを実装し、問題点や実用性を洗い出し、本システムの有用性について検討した。

1. はじめに

メタバースが盛り上がりを見せ、拡張現実(Augmented Reality: AR)や複合現実(Mixed Reality: MR)といった分野が注目される。それらのコンテンツを利用するデバイスとしてヘッドマウントディスプレイ(HMD)が挙げられる。

コンピュータでの文字入力に最も一般的に使われるデバイスは物理キーボードである。しかし、物理キーボードは机において使用する必要があるなどの制約があるため、HMDの文字入力手法として用いるには親和性が低い。そのため、外部入力デバイスを用いずにHMDを装着した状態での利用が可能な文字入力手法が必要であると考えられる。そこで本研究では、HMDの特性を生かすため、外部入力デバイスを用いず、HMDに内蔵されたRGBカメラのみを用いる文字入力手法を提案する。カメラに向けて指文字を表現することによって、文字の入力を可能とする。

指文字とは手話と同様に聴覚障害者が用いるコミュニケーション手法の一つである。手話は主に手指や腕を用いた動作により単語を表現する。しかし手話が存在しない単語や固有名詞を表す場合には手の形状によって1文字ずつ表現する指文字を用いて表す。

本稿ではHMDにおいて指文字を用いる文字入力システムを実装し課題を洗い出したうえ、実用性を検討する。

2. 関連研究

2.1 HMDにおける文字入力手法

土井ら[1]は深度カメラによって指先の位置を検出し、空中でのタップやフリックジェスチャを認識することにより仮想キーボードを操作する文字入力手法を提案した。この様に深度カメラまたはRGBカメラを用いて撮影した映像から指先または手全体を認識し仮想キーボードを操作する手法は多くみられる。Microsoft社製HoloLens[2]に搭載されるMRTK(Mixed Reality Keyboard)[3]でも同様に内蔵カメラ

を用いて指先の位置を検出し、視界内に表示されるQWERTYキーボードを操作し文字を入力する。

また、村田ら[4]はVRコンテンツ向けの密閉型HMDに視線計測装置を装着し、視線によって文字の入力を可能とした。頭を動かすことでHMD上での操作を可能としたYukangら[5]の手法は、文字入力には対応していないが仮想キーボードの操作に応用することも可能であろう。これらはすべて手や目、頭を用いてカーソルを動かすことによって仮想キーボードを操作し文字を入力するものであり、事節で挙げる事例は手の形状によって直接文字を入力するものである。

2.2 手話を活用した文字入力手法

木村ら[6]はカメラから得た映像から手の形状および動作を検出し、手話を認識することにより手話を用いて文字を入力できる辞書システムを開発した。また、生野ら[7]はスマートフォン上で内蔵カメラによって得た映像から手の形状を検出し指文字を認識することにより指文字を用いて文字を入力することが可能なシステムを開発した。

3. 提案手法

前節で紹介したようにHMDにおいてカメラを使用し手を検出することによって文字を入力する試みは多く存在する。ただし、それらはカーソルを操作し仮想キーボードを使用するものであるため、指先の揺れなどによって正確かつ迅速に文字を入力することが難しいと考えられる。また、手の形状を利用する文字入力システムはいくつかの先行研究があるが、いずれもHMD向けに作られたものではない。そのため本研究では、HMDで文字入力をする際、指文字を用いて文字を入力するシステムを提案する。

提案システムでは、ユーザはHMD前面に内蔵されたRGBカメラからおおよそ30cmほど離れた位置に右手を用いてカメラに向け指文字を示す(図1)。カメラから得た映像を解析して指文字を認識し、対応する文字を入力する。なお、入力したい文章を1文字ずつ全て入力することは非常に不便である。そのため、入力補完機能も実装する。本研究にて使用

^{†1} 北海道情報大学情報メディア学部

する指文字はアメリカ手話(ASL)[8]に対応する。さらに、入力補完機能の実装にあたって入力補完候補の選択や文字の削除、入力の確定といった制御用文字を表す指文字が必要となるため、独自に作成した指文字を使用する。



図1 本システムを利用している様子

4. 実装

4.1 システム構成

本研究の提案手法の概念実装として、システムの開発を行った。本システムはHMD,PCによって構成される。HMDはEPSON MOVERIO BT-30Eを使用した。MOVERIOにはRGBカメラが内蔵されているため、内蔵カメラから映像を取得する。しかしOSが搭載されておらず、外部から映像を入力する必要があるため、カメラから得た映像をPCに送信し処理した後HMDに映像を出力する。

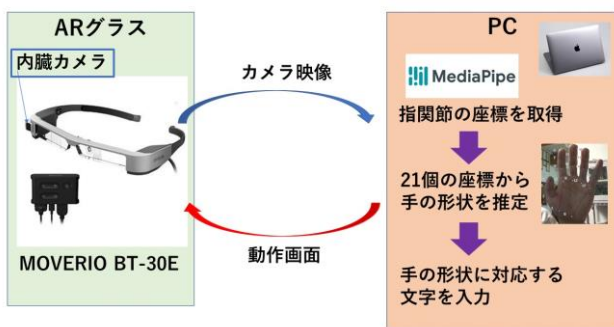


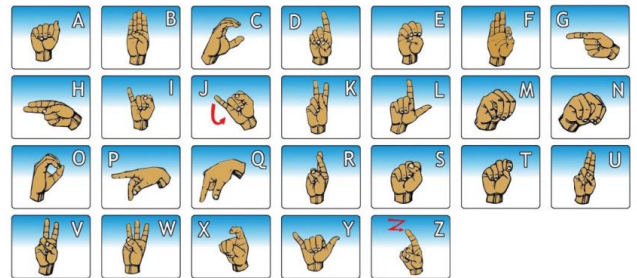
図2 システム概要

4.2 指文字の検出

内蔵カメラから得た映像からMediaPipe Hands[9]によって手の検出及び手の形状の認識を行う。MediaPipe Handsとは機械学習により動画から手を検出し、姿勢を推論するライブラリであり、Googleからオープンソースで提供される。対応言語はC++, Python, JavaScriptとなっており、本システムではPythonを使用した。本システムでは、手の位置・領域を検出するモデルである「Palm Detection Model」によって手を検出した後、「Hand Landmark Model」によって検出された手から21個の指関節の三次元座標を取得する。なお、奥行きを表すz座標は手首を起点とした相対座標である。MediaPipe Handsから得た手の各点の三次元座標をもと

に、相対的な位置関係などからルールベースで手の形状を推定する。推定された手の形状がシステムに登録された手の形状と一致した場合に指文字であると認識する。本システムでは、アメリカ手話(ASL)(図3)に対応しているため[A-Z]の26種類に加え、8種類の制御用文字(図4)と、合わせて34種類の手の形状をシステムに登録した。

ASL - FINGERSPELLING ALPHABET



SHARE & PRACTICE AMERICAN SIGN LANGUAGE ONLINE
WWW.SIGNLANGUAGEFORUM.COM/ASL

図3 アメリカ手話(ASL)の指文字(文献[8]より引用)



入力補完候補から1番目を選択
/空白文字を1文字入力



入力補完候補から2番目を選択
/空白文字を2文字入力



入力補完候補から3番目を選択
/空白文字を3文字入力



1文字削除



2文字削除



3文字削除



全文字削除



文字入力の確定
(ENTER)

図4 制御用指文字

4.3 文字の入力

前節の手法により認識した指文字をシステムに入力する。入力された文字は入力未確定の文字列として変数に格納され、入力補完機能の関数に渡される。入力補完機能にはオープンソースで提供されているライブラリであるFast Autocomplete[10]を使用した。入力補完候補は入力未確定の文字列に最も近いものから3つの単語が提案され、画面左

上に表示される。入力補完候補の中に入力したい単語が存在した場合に、ユーザは入力補完候補選択用指文字を示す。入力システム内では、入力補完候補内の単語を入力確定済みの文字列として変数に格納する。入力未確定の文字列をそのまま入力確定する場合には、文字入力の確定を表す指文字を示す。入力システム内では入力未確定の文字列を入力確定済みの文字列として変数に格納する。

4.4 画面表示

システムの実行画面を図 4 に、実行画面を合成したユーザ視点を図 5 に示す。

画面左上部に入力未確定の文字列及び入力補完候補を表示し、画面左下部に入力確定済みの文字を表示する。画面内には認識した手の各点の座標をもとに手の骨格を表示する。これによりユーザは自らの手がカメラの画角に入るよう調整することができ、手の形状が正しく認識されているかを確認することが可能となる。

本研究で使用した HMD である MOVERIO では、図 5 に示すように視界左手に仮想ディスプレイが表示される。本システムでは右手を用いて指文字を示すことを想定しているため、画面と自らの手が重なり視認性が低下することはないと考えられる。MOVERIO において画面内の黒い部分は透過して表示される。そのため、実際には左上部下部に表示されている文字及び手の骨格のみが視認できる。

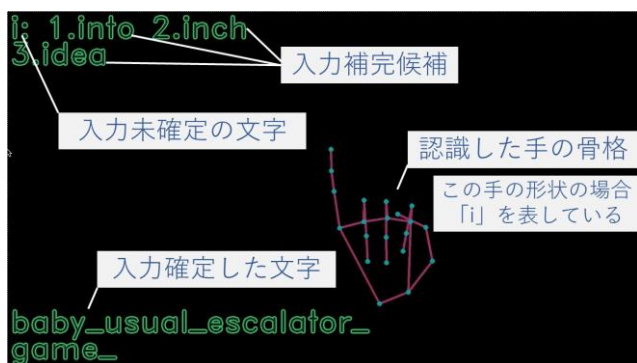


図 5 システム実行画面

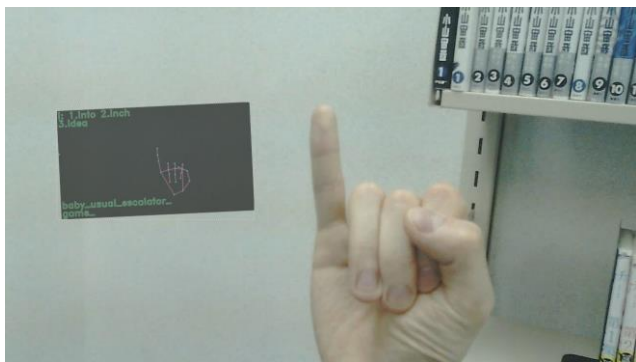


図 6 実行画面を合成したユーザ視点

5. 議論

本研究では、実装しシステムを利用する過程でいくつか

の課題が浮き彫りとなった。本節ではそれらの課題を挙げる。

5.1 指文字認識の正確性

特定の指文字が認識されにくいという課題があるということがわかった。認識されにくい場合の主な理由として、他の指の奥に隠れてしまう指の姿勢推定が不正確である、他の指文字と近似しているといったものが挙げられる。前者の問題が起りやすい例として「O」、「P」、「Q」といった手の側面を見せる指文字を示した際、手前の指に隠れた指の姿勢推定が不安定となり正しく認識されない場合が多くみられた。この問題に関してはハンドトラッキング技術の向上により改善されると考えられる。また「M」、「N」、「T」これら 3 種類の指文字は親指を内側に入れた拳を作り、親指の先を小指と薬指の間から出した場合「M」を表し、薬指と中指の間から出した場合「N」を表し、中指と人差し指の間から出した場合「T」を表す。3 種類それぞれ近似しているうえ、親指の先端以下が常に隠れているため姿勢推定が不正確であるといった 2 つの重大な問題を抱えている。ハンドトラッキング技術の向上により多少改善する見込みはあるものの、正確に認識することは困難であると考えられる。したがって指文字を文字入力に用いる場合これら 3 種類の指文字に関しては代替となる指文字を使用することによって、より正確な文字入力が可能となるであろう。

5.2 入力補完機能

本システムにて入力補完機能として使用した Fast Autocomplete には当初 40,000 語程度の英単語を集めた辞書データを使用していた。しかし処理の負荷が高くなり、フレームレートが低下した。そこで 2,000 語程度の簡単な英単語を集めた辞書データを作成し使用したところ、処理負荷は軽減され問題なく動作した。ただし実用的な文字入力手法を開発する場合 2,000 語程度の辞書データでは足りない可能性がある。その場合、他の入力補完機能を実装することも考えられる。

5.3 実用性

認識精度の低さに関しては改善の余地があり、認識精度が向上するにつれ実用性も向上するであろう。

続いて、ユーザは制御用指文字を含め 34 種類の指文字を習得しなければならず、日常的に指文字を用いないユーザにとって本システムを快適に利用することが困難であるという問題が挙げられる。この問題は、使用可能な指文字の一覧を表示する機能を追加する事により解決できる可能性がある。

腕を持ち上げ続けることにより起こる腕への疲労の蓄積も課題である。これは、既存手法の指先を用いて仮想キーボードを操作する手法と同様の課題である。本研究にて用いた HMD である MOVERIO ではカメラが正面を向いた状態で内蔵されているためユーザは自らの手をカメラの画角に収めるべくわずかに腕を持ち上げなければならない、ユーザ

の腕に疲労が蓄積してしまう。しかしながら手を検出するためのカメラが下向きに内蔵されている場合ユーザは手を持ち上げる必要が無い。そのため腕の疲労は軽減されると考えられる。

また、既存手法である手指などを用いてカーソルを動かして仮想キーボードを操作する手法で文字を入力する場合画面を注視しなければならないものの、本システムではカメラの画角に手が入ってさえいれば良く、画面を注視する必要がない。そのため視覚障害者向けの文字入力手法として応用することも考えられる。

6. まとめ

本研究では、HMD 内蔵カメラによって得た映像から MediaPipe Hands によるハンドトラッキングにて指文字を認識し、文字を入力するシステムを提案した。実装したところ指文字の認識精度の低さが目立ったため、機械学習を用いる等認識手法を改め認識精度を向上させる必要がある。

必要な機材は RGB カメラ 1 台のみであるがゆえ安価に実装できる。そのため導入が手軽であり、使用する環境を選ばないという利点がある。HoloLens に搭載される MRTK のように手の指先を用いて仮想キーボードを操作する手法と、本システムを組み合わせて使用することも考えられる。このように他の手法と組み合わせる応用も期待できる。本研究は HMD における文字入力の快適性を向上させ、HMD の普及に貢献するであろう。

参考文献

- [1] 土井 秀信,ヘッドマウントディスプレイ向けの直感的な文字入力インターフェースの提案,東京大学大学院 修士論文, 2017
- [2] “HoloLens2 製品ページ”,
<https://www.microsoft.com/ja-jp/hololens>, (参照 2021-12-16).
- [3] “MRTK のシステムキーボードの概要”,
<https://docs.microsoft.com/ja-jp/windows/mixed-reality/mrtk-unity/features/ux-building-blocks/system-keyboard?view=mrtkunity-2021-05>, (参照 2021-12-16).
- [4] 村田 朋来,盛川 浩志,小宮山 撰 : VR 空間における視線入力に関する基礎的検討. 研究報告ヒューマンコンピュータインタラクション (HCI) . 2018-HCI-177.22. P1-6
- [5] Yukang Yan, Chun Yu, Xin Yi, Yuanchun Shi : HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Volume 2, Issue 4. 2018, p1-23
- [6] 木村 勉,鈴木 洸輔,石濱 春,吉田 あすか,神田 和幸:手話認識機能を備えた手話辞書システムの開発. 研究報告アクセシビリティ (AAC) 2021, p1-6
- [7] 生野優輝,外村佳伸 : UbiScript:スマートフォンを用いた手指ジェスチャー認識. 情報処理学会インタラクション 2021. 2021
- [8] “ASL Fingerspelling Alphabet”,
<https://www.signlanguageforum.com/asl/fingerspelling/alphabet/>, (参照 2021-12-19)
- [9] “MediaPipe Hands”,
<https://google.github.io/mediapipe/solutions/hands.html>, (参照

- 2021-12-16)
[10] “Fast Autocomplete”,
<https://github.com/seperman/fast-autocomplete>, (参照 2021-12-16)