

## 音声認識とジェスチャ機能を用いた PC 操作 PC Operation by Speech Recognition and Gesture Function

杉本 拓也†  
Takuya SUGIMOTO

岡村 拓哉†  
Takuya OKAMURA

伊與田 光宏†  
Mituhiko IYODA

### 1. はじめに

近年 PC の操作方法は多々あるがマウス・キーボードによる操作法が最も一般的なものとなっている。しかし、Kinect の多様な機能を用いて、ユーザが行いたい操作をマウスよりも迅速に行えるシステムがあれば便利であると考えた。そこで、開発および実装を行い、ユーザに使用してもらい、評価を頂戴した。

### 2. 概要

本研究では、Microsoft 社から発売されている体感型インタフェース、Kinect を用いる。

Kinect を用いる利点は、PC 操作を行う際に Kinect は動きの認識に映像を用いるため、マウス等のデバイスを持つ必要がないこと、マイクを備えているため音声認識機能を使用できることが挙げられる。したがって、マウス・キーボードによる操作法では不向きな、PC から離れた料理中等の場所でも問題なく PC の操作を行うことができる。さらに、ジェスチャや音声認識を容易に入力することができる。

### 3. ジェスチャ機能

ジェスチャ機能とは、手や腕などの身体の動きや体勢を読みとって何らかの動作を行う機能のことをいう。ジェスチャ入力のメリットは、自然に操作を行うことができる可能性を秘めているという点である。ジェスチャ入力は、手を振る、手を伸ばすなど誰しもが行うことができる、単純な動作を読み取って何らかの他の操作を行うことができるため、自然にジェスチャ入力を行うことができる。本研究では、マウスカーソルの移動、各種クリック操作を PC に入力する等の 4 種類の操作を PC に入力するためにジェスチャ機能を用いる。

### 4. 音声認識機能

音声認識とは、ユーザが発する言葉を認識することによって文字情報として入力したりする機能を指す。音声認識機能の使用例として、予め設定した語とユーザが発した語の比較を行い、一致した場合に OS に実装されているショートカットキーを用いて、PC 操作を行うことが機能として挙げられる。ユーザが容易に PC を操作できる機能を実装できるため、プログラムの調整次第では、プレゼンテーション用などの場面に対応した使用する場面に合ったプログラムを容易に作成できると考えられる。しかし、認識する単語の数を増やし過ぎると、誤検出の可能性が増すため注意が必要である。本研究では WEB ブラウザの呼び出しを行うとき等に音声認識を用いる。

### 5. システム

#### 5.1 システム提案

PC を操作するに当たって Kinect をより効果的に使用するには以下のような機能を付加することが望ましい。また、入力するジェスチャと音声認識の単語に関しては予め設定したものを使用する。

- ・ジェスチャ入力によるマウスカーソルの移動、各種クリック操作を行う機能
- ・音声入力による WEB ブラウザ等の呼び出しを行う機能

ジェスチャ・音声入力に関してはユーザが試し入力を行いデータを取り、そのデータをユーザがジェスチャ・音声入力を使用する際に活かす、といった学習は行わないものとする。

#### 5.2 システム構成

Kinect と PC の接続構成は以下に示す図 1 のようになる。

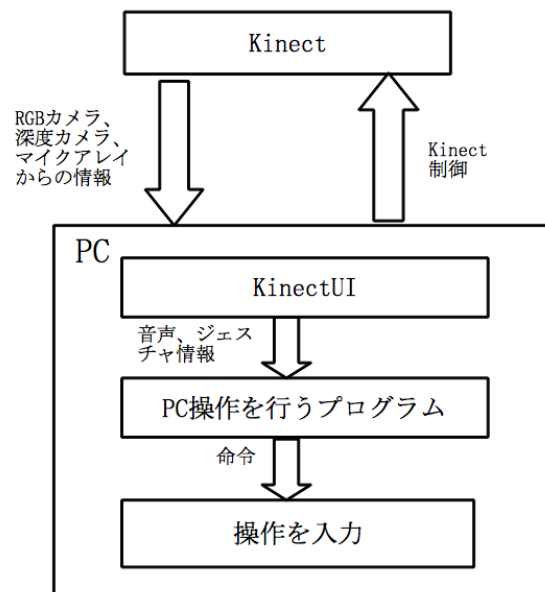


図 1. 全体の構成

主な流れとしては、Kinect によって得られた情報を PC に伝達し、伝達された情報を作成したプログラムによって PC 操作を行うための命令にリンクさせ、リンクさせた命令を使って実際に PC 操作を行うという流れである。

#### 5.3 プログラム構成

本研究で作成したプログラム全体の流れは、図 2 のようになる。

† 千葉工業大学

Chiba Institute of Technology

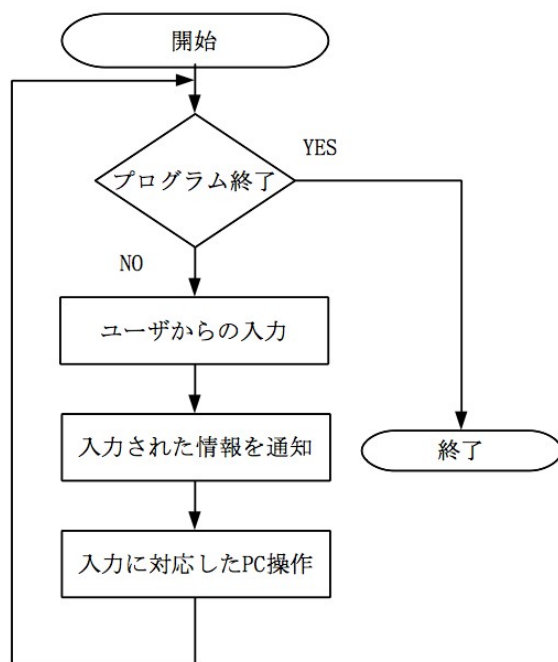


図2. プログラムの流れ

プログラムを起動すると終了操作が行われるか、ユーザからの何らかの入力があるまでは待機状態を維持し続ける。ユーザからの入力を Kinect が受け取ったとき、ユーザからどのような入力を受け取ったかをユーザに通知し、受信した情報を元に PC 内での制御を随時行う。この動作は、ユーザがプログラムを終了するまで、情報の送受信と PC の制御を繰り返し行う。また、ユーザから入力される情報はジェスチャと音声入力に分けられ、さらにジェスチャに関しては、右手と左手では入力できる操作を分けた。具体的に、どのような役割を割り当てたかは後述する。

## 6. プログラムの実行

実際に作成したプログラムを実行すると図3のような画面が表示される。

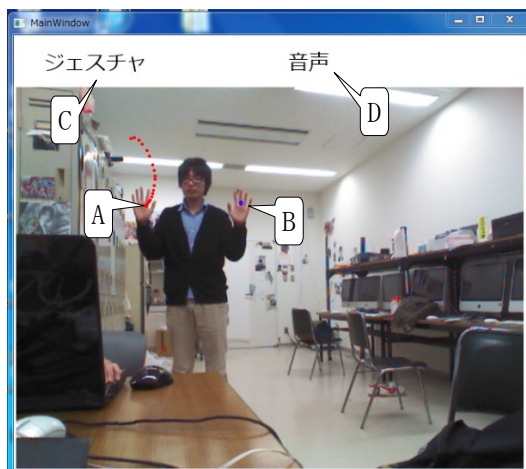


図3. 実行画面

図3に表示されているA～Dは次の通りである。

- A. 右手をトレースしている赤い点
- B. 左手をトレースしている青い点
- C. ユーザが行ったジェスチャを表示するテキストボックス
- D. ユーザが発声した単語を表示するテキストボックス

A、Bは、左右の手をリアルタイムで追跡する青と赤の点を表示している。また、右手にはマウスカーソルの移動とマウスの右クリックを行う役割、左手にはマウスの左クリックとダブルクリックを行う役割をそれぞれ割り当てた。

また、C、Dのユーザが行ったジェスチャとユーザが発声した単語を表示するテキストボックスはユーザがPCに入力を行うたびに入力された事柄に対応したテキストボックスが変化する。

## 7. 検証結果

作成したプログラムを第三者に利用してもらった。アンケート結果を表1として以下に示す。アンケートは5段階評価で表してもらい、1が「とても悪い」、5が「とても良い」とした。また、自由記述欄に意見を頂いた。

表1. アンケート結果

| 質問No. | 質問内容           | 評価  |
|-------|----------------|-----|
| 1     | ジェスチャは使いやすかったか | 3.1 |
| 2     | 音声認識は使いやすかったか  | 4.2 |
| 3     | 誤認識はなかったか      | 2.3 |
| 4     | 素早く操作を入力できたか   | 2.7 |
| 5     | 操作性は良かったか      | 3.2 |

表1をみると、誤認識についての質問の評価が悪いことがわかる。これは、日常的にジェスチャと音声認識機能を使っていないユーザの多くが、誤認識が多いと感じたという評価の1や2といった評価値をつけていた。一方で、ジェスチャと音声認識機能を使っているユーザは、あまり誤認識があるとは思わなかったという評価の4や5の評価値をつけていたことがアンケート用紙から見てとれた。したがって、操作に慣れが必要になってしまったことがわかる。

## 8. おわりに

本研究ではKinectの機能を利用したPC操作の提供を目的とし、システムの開発を行った。さらに、音声認識とジェスチャ機能を用いて、PC操作を行うことができた。

今回の研究において、ジェスチャと音声認識を用いて、PC操作を行うことができたが、作成したシステムが役立つ場面の一つとして料理中のユーザがレシピをPCで見たいような状況で役立つと考えた。今後は、本研究の発展として、料理中のユーザ支援を行うようなシステムの開発に着手したいと考えている。このシステムは、データベース機能を追加することで適宜料理のレシピを閲覧や食材の切り方ができるようなユーザサポートを行う機能を実装したい。

## 参考文献

- [1] 中村薫ほか, Kinect for Windows SDK プログラミング C#編