

Kinect を用いた大画面情報共有・提示のためのマルチモーダルインタフェースの試作

担当者：上松 巧弥 指導教員：長田 茂美 教授

1. まえがき

大画面モニターに PC を接続し、複数人で大画面に表示した情報を共有しながら、プレゼンテーションやグループウェアなどを実施するという場面では、モニターと PC の位置がある程度離れていると、PC 操作の都度、発表者はモニターと PC との間を行き来しなければならず、プレゼンテーションや議論の円滑な進行を妨げるという問題が生じる場合がある。PowerPoint のスライドショーの「進む」、「戻る」操作のような簡単な操作であれば、既存の遠隔操作ができる機器を使えば問題はないが、多くのマウス操作やキーボード操作が要求されるプログラムを実行するような場合には、PC から離れた場所からでも操作できる新しいインタフェースが必要となる。

本研究では、この問題を解決する方法の一つとして、視覚と聴覚を利用したマルチモーダルインタフェースを試作し、マウスとキーボードの代わりに、人間のジェスチャーと音声を用いることの有効性を検討する。なお、人間の骨格情報および音声情報の取得には、Kinect を使用した。

2. Kinect とは

Kinect とは、Microsoft が販売しているセンサーデバイスである。RGB カメラや深度センサー、マイクロフォンアレイなどを搭載しており、カラー画像や距離画像、人間の骨格情報、音声の取得などができる。Kinect には Windows 用と Xbox360 用があるが、Windows 用では、従来からの使用距離 80cm~4m に加え、40cm~3m にも対応しており、本研究では Windows 用を使用している。

3. システム概要

試作したマルチモーダルインターフェースは、PC のマウスとキーボードの代わりに、人間がジェスチャーと音声を Kinect に認識させ、その認識結果に基づいて PC を操作するというものである。具体的なジェスチャーとして、人間の手の位置を認識し、手の動きに応じたマウスカーソルの移動やクリックなどがある。図 1 に、ジェスチャーでのプログラム操作の流れを示す。

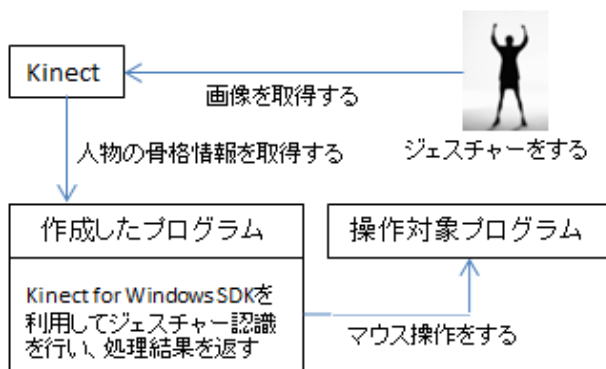


図 1 ジェスチャーでのプログラム操作の流れ

音声の認識については、Kinect から得られた人間の音声情報を基に、Microsoft Speech Platform SDK を用いて音声認識を行っている。音声認識の結果、その中で一定以上の信頼度が得られた単語があれば、その単語を認識したという結果を返す。図 2 に、音声でのプログラム操作の流れを示す。

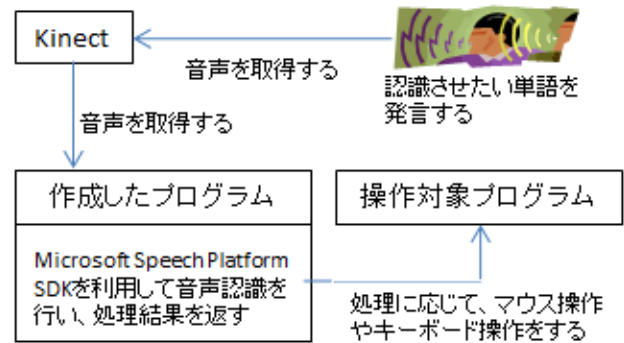


図 2 音声でのプログラム操作の流れ

4. システム評価

評価実験では、Kinect のみを使用し、操作対象プログラムとして、クリックやドラッグなどのマウス操作と文字入力に対応したプログラムを使用した。Kinect のみを使用してプログラム操作を行った結果、ジェスチャーと音声による直感的な操作で、従来のマウス操作やキーボード操作の一部の機能を実現できた。しかし、マウスカーソルを動かす際の感度が良くない点や、ジェスチャーや音声の誤認識が生じた起きた場合に間違った操作をするといった問題がある。このような問題を解決できれば、試作したマルチモーダルインターフェースによる操作対象プログラムの操作性を大幅に向上できるものと思われる。

5. むすび

本研究では、マウスとキーボードの代わりに、人間のジェスチャーと音声を用いたマルチモーダルインタフェースによって、プログラムを操作することの有効性を検討した。現在の段階ではジェスチャー認識や音声認識における、認識するジェスチャーや単語の数が少ないため、今後これらを追加したり、変更したりしていけばより使いやすいシステムを構築できる。

参考文献

- [1] 中村薫 田中和希 宮城英人 著
『KINECT for Windows SDK プログラミング C#編』
日経印刷株式会社(2012年4月10日)
- [2] 川俣晶 著
『新プログラミング環境 C#がわかる+使える』
図書印刷株式会社(2002年6月27日)