

Kinect を用いたジェスチャ翻訳の設計と開発

黒木 勇磨

1. はじめに

ジェスチャとは言葉を用いず身振り手振りで行うコミュニケーションの一つである。この手法は言語障害のある方や異なる言語間でのやりとりでは重要度が大きい、しかし必ずしも伝えたいことが相手に伝わるとは限らない。さらに認識の違いや文化の違いによって真逆の意味に解釈されることも少なくはない。そこで、正しくジェスチャの意味を判断し音声で出力すれば、そのような誤解も減少すると考える。本研究ではジェスチャを認識し、翻訳、そして音声で出力するシステムの設計と開発を行う。

2. Kinect センサとは

Kinect センサ (図 1) とは、Microsoft 社が販売しているコントローラーを用いずに身体



図 1 Kinect センサ

の動き、ジェスチャなどによる操作を可能とする周辺機器である。Kinect センサに搭載されている距離カメラ (解像度 640×480, フレームレート 30fps) や RGB カメラ (解像度 640×480, フレームレート 30fps) を用いてユーザーやユーザーの骨格を番号で認識 (図 2, 例えば左肩は③など)、追跡しユーザーの動きを 3 次元の座標データとして取得することによってジェスチャなどの操作を可能としている。

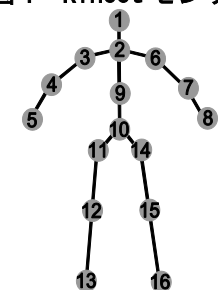


図 2 骨格の認識

3. ジェスチャ翻訳システムの構築

3.1 システムの概要

本システムは PC に接続した Kinect センサとスピーカーを用いる。ユーザーが Kinect センサの前に立ちジェスチャを行うことにより、そのジェスチャに対応した音声

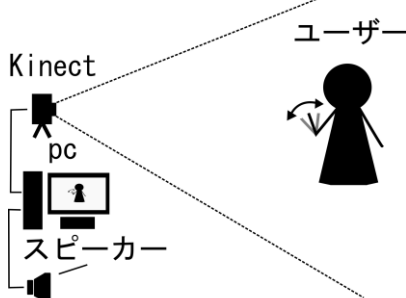


図 3 概要図

3.2 ジェスチャの表現と認識

本システムは、「バイバイ」、「オッケー」、「ダメ」の三種類のジェスチャの翻訳を実装している。それぞれのジェスチャ認識に必要な情報である、左手 (⑤)、左肘 (④)、左肩 (③)、右手 (⑧)、右肘 (⑦)、右肩 (⑥) の座標を Kinect センサを用いて取得する。またそれらの座標を利用し、手と肩の位置関係を肘を基準とした象限で取得する。図 4 (左) は、左肘を基準とした象限を表している。図 4 (右) は、左手の象限が 2、左肩の象限が 1 となる。これらの象限の変化と座標の変化によってどのジェスチャが行われたのかを認識し音声で出力する。またジェスチャは上記のような動作をすると認識するように表 1 のように定義した。

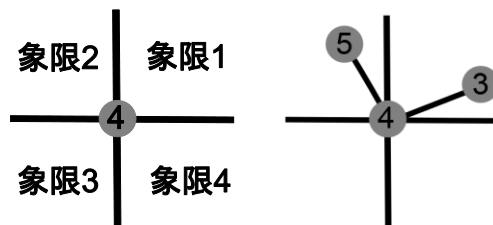


図 4 左肘を基準とした象限と位置関係の具体例

表 1 ジェスチャの動作定義

ジェスチャ	動作
バイバイ	⑤ in 象限 2→1→2→1 と遷移したら出力 (⑧の場合、象限の遷移は逆に定義)
オッケー	⑤ in 象限 1 ③ in 象限 4 かつ ⑤の X 座標 - ⑧ in 象限 2 ⑧の X 座標 ≤ 80 ⑥ in 象限 3
ダメ	⑤ in 象限 1 かつ ⑤の X 座標 - ⑧ in 象限 2 ⑧の X 座標 ≥ 150

また例としてバイバイの実際の動作を図 5 に示す。図 5 (左) はバイバイ動作の全体図、図 5 (中央, 右) は左肘 (④) を基準とした左手 (⑤) の象限の移動を表している。

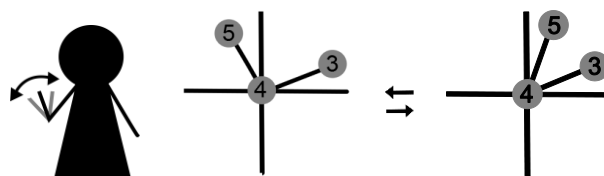


図 5 バイバイ動作の遷移

3.3 システムの構築

本システムは、Visual C++ 2010 Express, Kinect for windows SDK を用いて開発を行う。[1]

4. 実行例

実行例を図 6 に示す。ここでは「ダメ」のジェスチャを行っている。音声出力の結果を明示的に示すため、ペイントの起動を行っている。

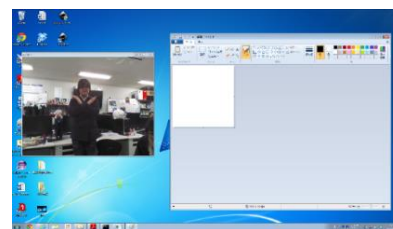


図 6 実行画面

5. まとめと今後の課題

本研究では、Kinect センサを用いて「バイバイ」「オッケー」「ダメ」の三種類のジェスチャを翻訳し音声で出力することができた。ジェスチャの種類を増やせはより多くのジェスチャの翻訳が可能になると考えられる。今後の課題として、個人個人の微妙なジェスチャの違いによる誤動作をなくし様々なパターンに対応したシステムを構築する必要がある。

[参考文献]

- [1] 中村 薫, 斎藤 俊太, 宮城 英人: KINECT for Windows SDK プログラミング C++編, 秀和システム, 2012.