

音声対話システムのための Kinect センサーによる人の接近検出 Detection of Approaching a Person by a Kinect Sensor for Speech Dialog System

早川 元貴[†] 實廣 貴敏[†]
Genki Hayakawa Takatoshi Jitsuhiro

1. はじめに

これまで研究機関で試作されてきた音声対話システムや近年話題になっているスマートスピーカは、基本的にユーザが話しかけると答える。対話性を向上させるには、時にはシステム側からの問いかけがあるとよいが、むやみに話しかけても意味がない。ユーザの状況を理解した上で話しかけると効果的と考えられる。据え置いた対話システムに興味があれば、近づこうとするだろうし、そうでなければ、通り過ぎる。これらの動作を検出できれば、話しかけるタイミングを図る参考になると考えられる。

そこで、本研究では、音声対話システムにおいて、目の前にいる人たちに呼びかけ、対話できるようにするために、彼らの行動が、こちらに接近しているのか、離れて行っているのか、単に通り過ぎているのかを検出する方法を検討する。具体的には、Microsoft 社の Kinect センサーを利用して人物の動作を推定する研究[1][2]を参考に、Kinect センサーを用い、人物を検出し、その動きを推定する。移動方向や速度を用いて識別を行う。

2. 接近・離脱の推定方法

2.1 動作の定義

ユーザの動作として、(1)「接近」、(2)「離脱」、(3)「停止」、(4)「通過」の 4 種類を定義する。図 2 に本研究で使用する xz 座標系を示す。問題を簡単にするために、それぞれの動作パターンを図 3、図 1、図 4 のように決める。

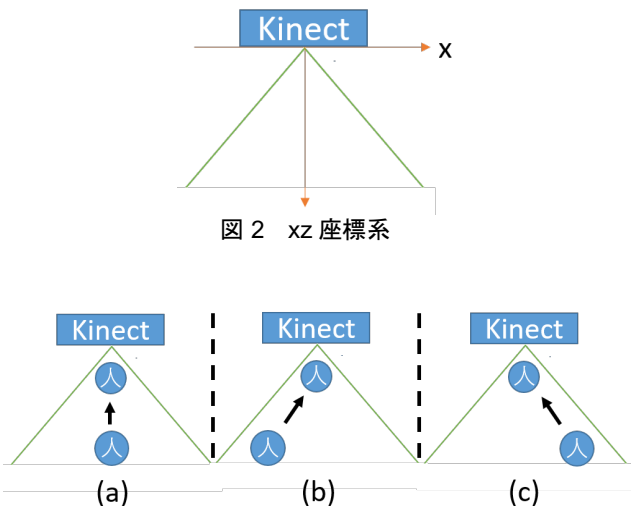


図 3 接近の動作パターン

[†] 愛知工科大学, Aichi University of Technology

Kinect から下へ向かって左右に開いている線分は、Kinect により深度距離を推定できる範囲を示す。

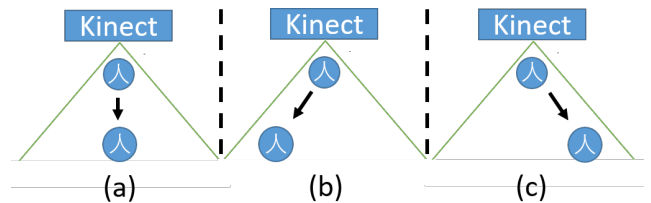


図 1 離脱の動作パターン

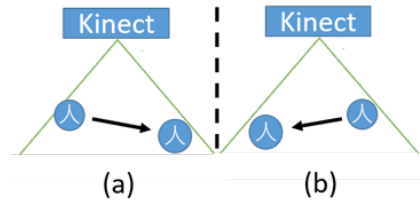


図 4 通過の動作パターン

2.2 動作推定方法

図 5 に今回用いる動作推定方法の流れを示す。接近と離脱の判定をするために、頭部の座標値から距離を求めて速度を算出し、歩行しているかどうかの判定をする。この時に歩いていないと推定された時は、止まっているので「停止」と判定する。歩いている時には、頭部の座標値から傾きを求める。そして、Kinect を通過しているかどうかの判定をする。通り過ぎていると推定された時は、「通過」と判定し、そうでない時は、速度の符号で判定をする。符号がマイナスの時は「接近」、プラスの時は「離脱」と判定する。

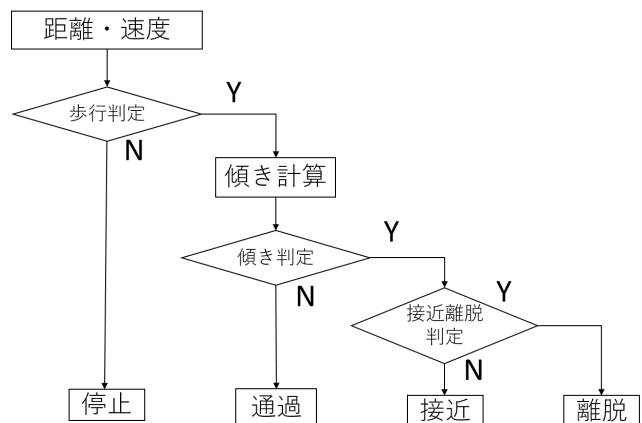


図 5 動作推定方法

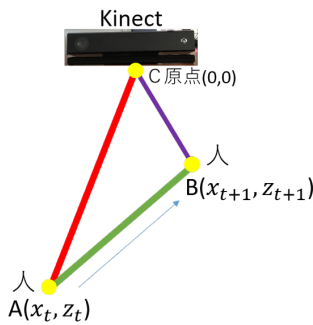


図 6 人物座標の定義

2.3 動作推定方法：距離と速度の計算

距離を推定する方法として 2 種類検討する。深度値から [提案法 A] では時刻 t と $t+1$ での原点からの距離を求め、距離の変動から接近/離脱を推定、速度は 2 点間の距離との大小から求めた符号を付与する。[提案法 B] では Kinect からの距離を使用し、速度を推定する。

[提案法 A]

まず、時刻 t での座標 (x_t, z_t) と時刻 $t+1$ での座標 (x_{t+1}, z_{t+1}) での Kinect (原点) からの距離を求め、

$$d_t = \sqrt{x_t^2 + z_t^2}$$

接近/離脱の判定を行う。

$$d_{t+1} \geq d_t \text{ のとき } f(d_t, d_{t+1}) = 1, \text{ 離脱}$$

$$d_{t+1} < d_t \text{ のとき } f(d_t, d_{t+1}) = -1, \text{ 接近}$$

次に、座標間の距離 D_{t+1} を求める。

$$D_{t+1} = \sqrt{(x_{t+1} - x_t)^2 + (z_{t+1} - z_t)^2}$$

速度を下記のように求める。

$$v_{t+1} = f(d_t, d_{t+1}) \times D_{t+1} / T$$

ここで、Kinect が 1 秒間あたり 6 個の頭部座標を取得できることから、 $T = 1/6$ [s] とする。

[提案法 B]

提案法 A での速度の定義を下記のようにする。

$$v_{t+1} = (d_{t+1} - d_t) / T$$

2.4 動作推定方法：欠損修正、歩行判定、傾き判定

[欠損修正]

Kinect でユーザーの検出を行い、骨格認識後、頭部を追跡し、頭部座標を取得する。しかし、完全ではなく、時折、座標が得られない場合がある。その時は前後時刻の座標の平均を座標として用いる。x 座標では、

$$\bar{x}_t = (x_{t-1} + x_{t+1}) / 2$$

とする。z 座標も同様とする。

[歩行判定]

安定的に判定するために、スムージングをした上で、歩行か停止の判定を行う。速度 v_t が速度の閾値 v_1 を超えているとき 1、そうでないとき 0 とする。前後フレームのこの値を使い、移動平均を計算する。その値が 1 より小さいとき停止、1 のとき歩行と判定する。

[傾き判定]

「通過」か、「接近/離脱」を判定するために、Kinect に対する傾きを推定する。前後フレームの座標を用い、最小 2 乗法から傾きを計算する。閾値との比較で判定を行う。

3. 評価実験

3.1 実験条件

大学研究室内で $3\text{m} \times 3\text{m}$ の範囲、Kinect が検出可能な前方 70 度内で被験者に歩行してもらった。被験者は 20 歳の 20 名で、あらかじめ決めたルートを一人 10 回程度、歩行してもらった。人物検出プログラムは Microsoft 社 Kinect for Windows SDK 2.0, NtKinect[3] を用いて作成した。

3.2 実験結果

図 7 に速度の閾値を変えたときの再現率、適合率を示す。また、表 1 にそのときの最大 F 値を示す。提案法 A, B で、接近、離脱においては大きな差はなく、9 割近く検出できた。さらに、傾きの閾値で得られる「通過」判定の F 値最大値は提案法 A は 70.6%, B は 46.6% であった。B では、横に移動したときの変化量が小さいため、速度から「停止」と推定されることが多かったためである。

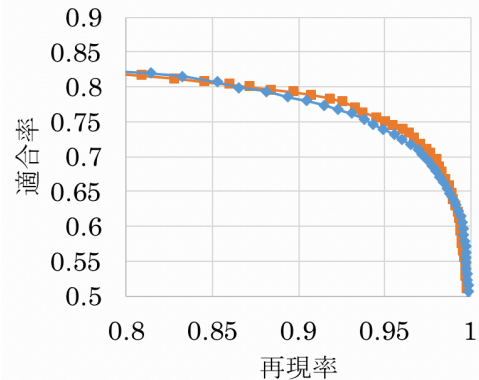


図 7 速度閾値を変えたときの再現率、適合率

表 1 速度閾値を変えたときの F 値の最大値

	F 値[%]		
	停止	接近	離脱
提案法 A	84.6	89.7	88.7
提案法 B	83.8	89.4	88.2

4. まとめ

音声対話システム前の行動を Kinect センサーにより、「接近」「離脱」「停止」「通過」動作を推定する手法を検討した。今後、より自然な動作から必要な行動を検出できるように改善して行きたい。

謝辞

本研究の一部は JSPS 科研費 26330211 の助成を受けた。

参考文献

- [1] 有満大輝, “Kinect を用いた頭部追跡によるリアルタイム音声強調の研究”, 大分大学卒業論文 (2013)
- [2] 神園卓也, 高野茂, 馬場謙介, 村上和彰, “深度情報を含む映像からの行動認識に関する研究”, 情報処理学会研究報告 (2013)
- [3] 新田善久, “NtKinect : C++ Class Library for Kinect V2”, 第 172 回ヒューマンコンピュータインタラクション(HCI)研究会(2017)