

複数の深度センサを用いたユーザの向きに依存しない ジェスチャ認識手法

都竹尚紀[†] 梅澤猛[‡] 大澤範高[‡]

千葉大学工学部[†] 千葉大学大学院融合科学研究科[‡]

1. はじめに

ヘッドマウントディスプレイ (HMD) を利用した VR 空間では手元が確認できないためデバイスを用いずに操作可能なジェスチャ入力が望まれる。しかし HMD を装着してジェスチャ入力を行う VR 空間ではユーザが映像に即して向きを変えるため、ユーザとセンサの向きが一致せず、ジェスチャが認識されない問題がある。本研究では HMD 装着時の骨格情報を用いたジェスチャ入力において、1 台の Kinect を用いて骨格情報を座標変換する手法と、複数の Kinect を使用して取得した骨格情報の各部位に信頼度を持たせて各骨格情報を統合する手法を組み合わせる。そしてユーザの角度とジェスチャの認識率の関係を実験で求めることで提案手法の有用性を示す。

2. 関連研究

Kinect を用いた向きに依存しないジェスチャ認識手法として長谷川らはユーザの向きから骨格情報を座標変換する手法 [1] を提案した。しかし 1 台の Kinect から骨格情報を取得しているため、Kinect との角度が大きくなるほどセルフオクルージョンにより認識精度が低下する。また 2 台の Kinect を用いたモーションキャプチャとして 3 次元点群を作成しパーティクルフィルタから部位を推定する手法が提案されている [2]。しかし実時間性に乏しく、また高いスペックの PC を必要とする。本研究では Kinect で利用可能な骨格取得を利用して得られた各骨格情報を利用することにより、ユーザへの負担なく複数の Kinect を用いた骨格取得を可能とする。

3. 条件設定

3.1. 想定環境

今回は 2 台の Kinect 向かい合わせに配置し、角度は座標変換を行うため図 1 のように設定する。今回は一方の Kinect を正面、他方の Kinect を背面と設定し、ユーザの向き θ は正面の Kinect とユーザのなす角度とする。

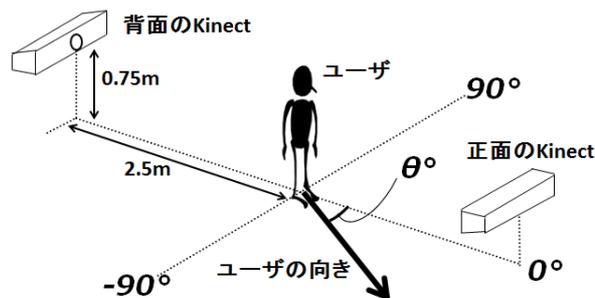


図 1 想定する環境

3.2. ジェスチャ認識

ジェスチャは右手で行い、右手、右手首、右肘の三次元座標を利用する。ジェスチャ認識には DTW 法 (動的時間伸縮法) を用いる。DTW 法とは 2 つの時間的なシーケンスの相違度を計算する手法である。登録ジェスチャには [1] で示された個人差の少ないジェスチャとそれらの反対のジェスチャ、円運動の計 10 種のジェスチャを利用する。



図 2 登録ジェスチャー一覧

4. 1 台の Kinect での骨格取得の改良

[1] では上半身の部位を通る面の法線ベクトルから向きを取得しているが、両肩の部位を利用すると奥の肩のセルフオクルージョンにより誤差が発生するため、本手法では肩の中心、背骨、右肩と左肩のうち手前にある肩の 3 点を通る面の法線ベクトルから向きを取得する。そして得られた角度を基に座標変換を行う。その際 Kinect で得られるユーザの重心の座標をユーザ座標の原点として、相対座標としてから変換を行う。

Gesture recognition independent of the orientation of the user with multiple depth sensors

[†]Naoki Tsutake, School of Engineering, Chiba University

[‡]Takeshi Umezawa, Noritaka Osawa, Graduate School of Advanced Integration Science, Chiba University

5. 複数の Kinect での骨格取得

5.1. 信頼度

2 台の Kinect で得られた骨格情報を統合するために本研究では信頼度を定義する。

信頼度とは Kinect で得られた骨格部位の取得状況を各部位ごとに算出したものである。取得状況の 3 パターンを表 1 に示す。

表 1 骨格部位の信頼度

取得状況	意味
TRACKED	正しく取得されている
INFERRED	前データや他の部位、骨格の幾何学情報により推定
NOT_TRACKED	推定からも取得が不可

5.2. 骨格情報の統合

取得した骨格情報の信頼度の優先順位を高い方から TRACKED, INFERRED, NOT_TRACKED と定める。4 の手法で座標変換された 2 つの骨格情報の各部位において信頼度の優先度が高い部位の座標を統合する骨格部位の座標とする。優先度が同一の場合は 2 つの部位の座標の平均値をとる。

6. 実験

6.1. 概要

ジェスチャの認識精度を次の 3 手法で行う。i) [1] の従来手法(従来手法と記載)
ii) [1] を改良した手法(提案手法(1 台)と記載)
iii) 信頼度を用いて複数の骨格情報を 1 つに統合する手法(提案手法(2 台)と記載)

(ii) と (iii) を組み合わせる提案手法の有用性を確認する。今回の実験で (i) は従来手法を利用してプログラムを作成し、追実験したものである。

6.2. 環境

今回は各角度 (0°, ±30°, ±60°, ±90°) で UpRight, DownLeft の 2 ジェスチャを 3 人×10 回で合計 30 回ずつ行い、認識率を比較する。

6.3. 結果

UpRight, DownLeft のジェスチャ認識率の結果をそれぞれ図 3 (a), (b) に示す。

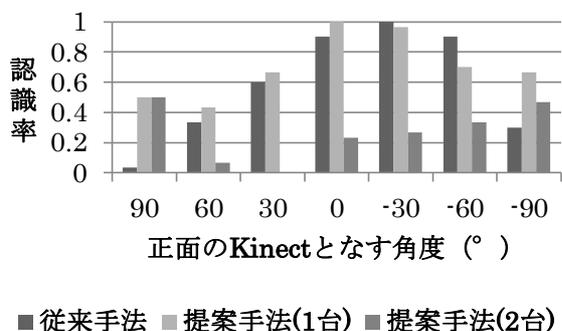
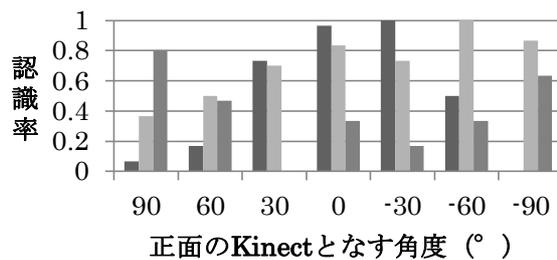


図 3-(a) UpRight のジェスチャ認識率



■ 従来手法 ■ 提案手法(1台) ■ 提案手法(2台)

図 3-(b) DownLeft のジェスチャ認識率

7. 考察

従来手法に比べ提案手法(1 台)は正面から離れていてもジェスチャを認識出来ているが、正面から離れるほど認識率が低下する傾向にある。提案手法(2 台)は正面付近を認識率が極端に低かった。これは背面の Kinect では骨格認識が左右の手を反対に認識してしまうことによるものと思われる。±90° では認識率が上がっているが、提案手法(1 台)の方が認識率が高い場合もあり、今回の 2 つの部位の座標の平均値をとる手法とは別の統合の方法が必要であるといえる。提案手法(2 台)の±60°, ±90° の認識精度が向上すれば、提案手法(1 台)と提案手法(2 台)を組み合わせることで、正面から離れた向きにおいてもジェスチャ認識が可能となると考えられる。

8. おわりに

今回の実験を踏まえ、さらに多くの被験者によるジェスチャのデータや集めることで、各角度における各手法の特徴がより顕著になると考えられる。また、各手法で得られる骨格座標の精度の比較検討を行うことで、2 台の Kinect でのジェスチャ認識の精度の向上が望める。今回は 2 台の Kinect を正面、背面と固定したが、ユーザの向きに合わせ正面、背面を判別することで 360° のジェスチャ認識が可能となるとと思われる。

参考文献

[1] 長谷川秀太, 赤池英夫, 角田博保, “姿勢を考慮したハンドジェスチャを利用する機器操作の提案・評価”, 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2012-HCI-147, No. 24, pp. 1-6, 2012.
[2] L. Zhang, J. Sturm, D. Cremers, D. Lee, “Real-Time Human Motion Tracking using Multiple Depth Cameras”, In Proc of the International Conference on Intelligent Robot Systems (IROS), 2012.

謝辞

本研究は JSPS 科研費 21300318 の助成を受けたものです。