

姿勢を考慮したハンドジェスチャーを利用する 機器操作の提案・評価

長谷川 秀太^{†1} 赤池 英夫^{†1} 角田 博保^{†1}

徒手にて離れた位置にある機器を操作するための、基礎的な調査およびその一適用例を紹介する。近年、深度カメラによって被写体の三次元位置情報の取得が容易になり、それを利用した新しいインタラクションの発展が予想される。そこで手によるジェスチャーによる機器操作を念頭におき、ジェスチャーの調査を行った。その結果を用い、ユーザの向きによらずジェスチャーを認識する手法を提案・実装し、実験によって有用性を評価した。また、適用例としてジェスチャー操作の動画再生システムを実装し、有用性を評価した。

Bare hand gesture in consideration of posture within remote control

SHUTA HASEGAWA,^{†1} HIDEO AKAIKE^{†1}
and HIROYASU KAKUDA^{†1}

By the advent of off-the-shelf depth cameras, obtaining 3D-positional information has become easy and we think that interaction techniques utilizing such devices will be developed more and more. So, we first investigated a suitable set of hand gesture for controlling remote equipments, especially home appliances. Moreover, with taking into consideration of user's posture, we thought it is not necessary to face to the target. On the basis of the concept, we proposed and implemented a flexible method to operate a remote equipment by gestures without any controller in hands, and evaluated it through some experiments.

^{†1} 電気通信大学大学院情報理工学専攻情報・通信工学専攻
Graduate School of Informatics and Engineering Department of
Communication Engineering and Informatics The University of Electro-Communications

1. はじめに

離れた位置にある機器を徒手にて操作するための基礎的な調査およびその適用例を示す。過去の研究においてジェスチャーを入力とする研究がある。方法としては、装置を体に装着し情報を取得する方法¹⁾や、画像情報から手の位置、動作を認識する方法などがある²⁾。しかし、装置を体に装着する場合着脱に手間がかかる、画像認識による方法では動作認識が難しいという問題がある。ところで、近年安価で入手できる深度カメラの登場によって被写体の三次元位置情報の取得が容易になり、それを利用した新しいインタラクションの発展が予想されている³⁾。そこで、まず、家電機器操作を念頭におき、利用しやすいジェスチャーの調査を行った。次に、深度情報を用い、操作者の姿勢も考慮してそれらのジェスチャーを認識する手法を提案・実装した。また、ジェスチャーにて操作する動画再生システムを作成し、提案手法の有用性を評価した。

2. 提案システム

本研究では、深度カメラを用いて空中での手によるジェスチャーの動作特性を調査し、手により指示可能な範囲と精度の調査の実施、3次元位置情報を利用したジェスチャーによる電子機器操作手法の提案と実装、ユーザの向きによらずジェスチャーを認識可能にするシステムの提案と実装を行った。そして、そのシステムを実装した動画プレイヤーを作成した。

2.1 設計方針

ユーザによって違いが少ないジェスチャーを選別し、そのジェスチャーを利用して電子機器を操作できるシステムを作成する。今回の設計方針として、以下の2点がある。

- ジェスチャーにより簡単に操作ができる。
- ユーザの向きや体勢などの制限を低減させる。

2.2 ジェスチャー認識

ジェスチャー認識は、動的時間伸縮法 (Dynamic Timewarping、以後は DTW と略記) によって実装した。DTW はデータ長やピーク値をとるタイミングが一致するように、2つの波形データの同期を取るアルゴリズムであり、これによりジェスチャーにかける時間や動作距離によらず同一のジェスチャーの認識を可能とした (図1)。

2.2.1 試作

電子機器操作の一例として、DTW を用いてジェスチャーを認識し操作を行う動画プレイヤーを作成し、HI2011にて対話発表を行なった⁴⁾。対話発表において、ジェスチャーの誤

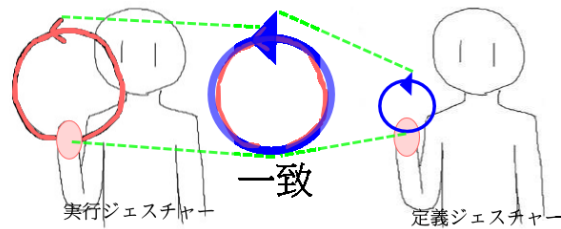


図1 DTWでのジェスチャー認識

認識は少なく、ユーザによらず認識が可能だという意見が得られた。

2.3 姿勢認識

本研究ではユーザの向きを認識する方法として、骨格トラッキングと顔認識を使用する(図2)。まず、ユーザの上半身の骨格を取得しそれを含む平面を作成する。それに垂直な方向を奥行き方向とする。その奥行き方向をユーザの向きとした。ユーザの実行したジェスチャーを向きに合わせて座標変換することにより、ユーザの向きによらずジェスチャー操作が可能になると考えられる。

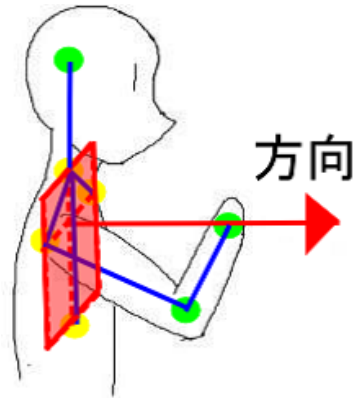


図2 向き認識

3. ジェスチャー調査実験

高い認識精度を見込めるジェスチャーを選定するためにジェスチャー調査実験を行った。

3.1 実験概要

利き手で指定のジェスチャーを深度カメラの前で実行し、その軌跡を取得した。ジェスチャーは、動作開始から5秒間をジェスチャーとした。記録されたジェスチャー数は5セッション×21ジェスチャー×3回×5人=1,575タスクとなった。ジェスチャーには上下左右前後など一方向にスライドするもの(スライドジェスチャーと呼ぶ)、バイバイするものや上下や左右に往復するもの(往復ジェスチャーと呼ぶ)、円を書くといった特徴的な動きがあるものを使用した。使用したジェスチャーは図3、図4となる。

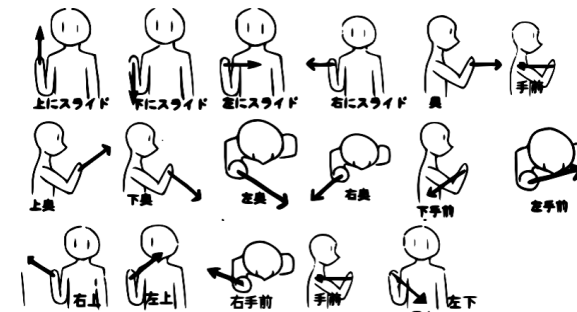


図3 スライドジェスチャー



図4 その他のジェスチャーと往復ジェスチャー

本研究は離れた機器を操作するのが目的のため、いわゆる10フィートUIを採用し、深度カメラから3m離れた位置に椅子を設置し、カメラに正対して座り、行うよう指示した。

実験装置の深度カメラには Microsoft の Kinect *1 を使用した。実験中の被験者をビデオカメラにより録画した。

3.2 結 果

各ジェスチャー動作の安定性を以下の方法で数値化した。各タスクの開始点を原点をしたときの終点座標の平均値を基準座標として設定し、それと各タスクの終点座標との差から標準偏差を計算した。この標準偏差の値をジェスチャーの安定性とした。今回の調査では、各辺に対し 5cm 程度のぶれに収まる標準偏差 100cm^3 を基準とした。

取得したジェスチャーの軌跡を解析し、直線系のジェスチャーにおけるユーザ毎の標準偏差と全被験者の標準偏差 (図 5) を求めた。

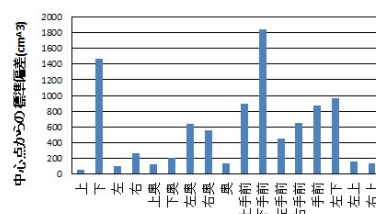


図 5 ジェスチャーごとの標準偏差

標準偏差が 100cm^3 を下回ったジェスチャーは、表 1 のようになった。

表 1 偏差が 100cm^3 未満の人数

	上	下	左	右	上奥	下奥
人	5	0	5	3	4	2
	左奥	右奥	上手前	下手前	左手前	
人	2	3	1	1	2	
	右手前	手前	奥	左下	左上	右上
人	0	1	4	3	3	4

奥行きを使用しないものは上と左の誤差が低く、全員が標準偏差 100cm^3 を下回った。利き手側である右方向は左側よりも誤差が大きく、標準偏差 100cm^3 を下回った人数は 5 人中

3 人となった。手前は、奥に比べ評価が低くなった。よって、単体のジェスチャーとして、標準偏差が 100cm^3 未満の人数が 5 人中 4 人以上存在する上、左、奥、上奥、右上が有用であると言える⁴⁾。

4. ジェスチャー認識実験

ジェスチャー認識の認識精度調査のため、ジェスチャー認識実験を行った。

4.1 実験概要

予備実験よりジェスチャーセットを作成し、ジェスチャー認識アルゴリズムを使用して認識実験を行った。ジェスチャーは標準偏差が小さいジェスチャー (上、左、奥、右上) と、それに対応するジェスチャー (下、右、手前、左下)。認識は難しいと思われるが特徴的なジェスチャー (時計回り、反時計回り) を使用し、全 10 種類とした。

タスクは指定したジェスチャーを深度カメラの前で実行させるものとした。ジェスチャーはカメラに正対して行った。被験者は、本研究室に所属する学生 9 人である。年齢は 21 歳から 25 歳で全員男性であり、全員右利きであった。実験は、10 種類のジェスチャーを 1 セッション中 3 回行い 10 セッション行った。第 1 セッションの前に、正誤フィードバックをつけたシステムを使わせジェスチャーを覚えさせた。実験では正誤フィードバックは行わなかった。

4.2 結 果

実験より得られたログデータを分析し、考察を行う。

ジェスチャー認識実験の全被験者の正答数の平均を図 6 に示す。

平均認識率が 78.5 % となった。

ジェスチャーの平均認識率は、ジェスチャー全体の平均は 78.5 % となった。最も認識率が高いジェスチャーは右にスライドで 99.6 %、二番目が左にスライドで 98.1 %、3 番目が下にスライドで 94.8 % だった。反対に最も認識率が低いジェスチャーは時計回りで 39.3 %、次に手前にスライドで 45.6 %、3 番目が反時計回りで 55.6 % だった。スライドジェスチャーの平均は 86.3 %、時計回りと反時計回りの平均は 47.4 % となった。

実行したジェスチャーとシステムによって認識されたジェスチャーを、表 2 に示す。

反時計回りのジェスチャーは、全誤認識回数が 120 回で、最も多い誤認識は右にスライドするもので、全誤認識の 45 % だった。次は、左にスライドするもので、全誤認識の 29 % だった。3 番目に多かったものは上にスライドするもので、全誤認識の 17 % となった。時計回りのジェスチャーは、全誤認識回数が 150 回で、最も多い誤認識は右にスライドする

*1 <http://www.xbox.com/kinect/>

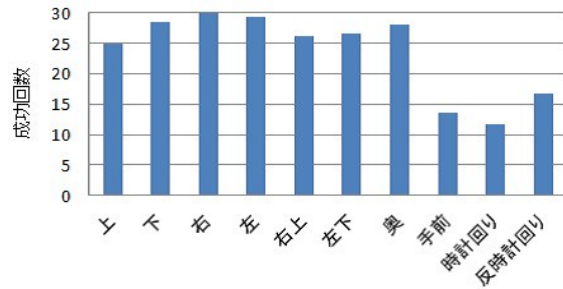


図 6 全被験者の成功回数の平均

表 2 実行ジェスチャーと認識されたジェスチャー

	上	下	右	左	右上	左下	奥	手前	時計回り	反時計回り
上	223	0	3	0	0	0	0	0	42	2
下	1	257	2	0	0	6	4	0	0	0
右	0	4	266	0	0	0	0	0	0	0
左	1	0	15	247	6	1	0	0	0	0
右上	0	0	4	0	250	0	0	0	0	16
左下	3	1	32	1	1	232	0	0	0	0
奥	0	1	30	0	0	0	238	1	0	0
手前	2	27	107	0	0	0	3	131	0	0
時計回り	33	2	76	3	34	0	2	0	120	0
反時計回り	20	7	54	35	1	1	0	0	2	150

もので、全誤認識の 51 % だった。次は、右上にスライドするもので、全誤認識の 23 % だった。3 番目に多かったものは上にスライドするもので、全誤認識の 22 % となった。これらの誤認識は、それぞれ反時計回り、時計回りの動作の一部になるため、その時点でジェスチャー認識が発生し、誤った認識が多く発生してしまったのだと思われる。

手前にスライドするジェスチャーは、全誤認識回数が 139 回で、最も多い誤認識は右にスライドするもので、全誤認識の 77 % だった。これは、右手でジェスチャーを行なっているため、自然に手を引くと手前と同時に右にも動いてしまうのが原因だと考えられる。

5. 向き認識実験

向き認識の認識精度調査のため認識実験を行った。

5.1 実験概要

ジェスチャー認識実験と同様のジェスチャーセットを使用して、向き認識アルゴリズムを使用して認識実験を行った。

タスクは指定したジェスチャーを深度カメラの前で実行させるものとした。カメラに対して 0 度と左右 30、60、90 度に向けて行った。被験者は、本研究室に所属する学生 9 人である。年齢は 21 歳から 25 歳で全員男性であり、全員右利きであった。実験は、10 種類のジェスチャーを 1 セッション中 3 回行い、各角度 3 セッションで計 21 セッション行った。第 1 セッションの前に、正誤フィードバックをつけたシステムを使わせジェスチャーを覚えさせた。実験では正誤フィードバックは行わなかった。

5.2 結果

実験より得られたログデータを分析し、考察を行う。

各角度、各ジェスチャーにおける、0 度の時とのジェスチャー認識率の有意差を表 3 に示す。

左 60 度、左 30 度は 0 度の時と有意差が無かった。しかし、左 90 度では、右、左、右上、左下、奥のジェスチャーに有意差があり、右 90 度では下、奥、手前のジェスチャーに有意差があった。よって、両 90 度のジェスチャーにおいて、0 度と同様に使用できるとはいえない結果となった。これは、90 度横を向いているため、ジェスチャーを行う右手が体に隠れてしまう事、左右 90 度では向き認識に必要な骨格情報が完全に取得する事が出来ず、向き認識にぶれが出てしまった事が原因だと考えられる。

右側に向いている時の下へスライドするジェスチャーの認識率が低いが、これらはほとんど左下と誤認されており、右 30 度が 98.3 %、右 60 度が 85.5 %、右 90 度が 86.0 % となっている。これは、右側を向くため、ジェスチャーを行う右手が体に隠れてしまう事が原因だと考えられる。

6. 実証実験

本研究の応用例として作成した、利用者の姿勢を考慮したジェスチャーで操作する動画プレイヤーを、自由に使用させる実証実験を行った。

6.1 実験概要

本研究で作成したジェスチャー認識システム、向き認識システムを使用して実際の環境で機器を操作した時の有用性調査のため、実験を行った。タスクは本システムを使用して動画プレイヤーを操作するもので、実際に使用する事を想定して自由に使用させた。実験は 10

表 3 0度との有意差

	左 90	左 60	左 30	右 30	右 60	右 90
上	無	無	無	無	無	無
下	無	無	無	有	有	有
右	有	無	無	無	無	無
左	有	無	無	無	無	無
右上	有	無	無	無	無	無
左下	有	無	無	無	無	無
奥	有	無	無	無	無	有
手前	無	無	無	無	無	有
時計	無	無	無	有	無	無
反時計	無	無	無	無	無	無

分程度行い、実験時のジェスチャー認識の成功率と、実験後のアンケートで評価を行った。

6.2 実験内容

6.2.1 実行ジェスチャーと操作

実行ジェスチャーと操作の対応は、HI2011で発表した動画プレイヤー⁴⁾と同様の10種類とした(図7)。タスクとして、自由にジェスチャーで操作させた。操作する時に実行時に操作を声に出してから実行させ、認識ジェスチャーを取得した。実験開始前に、練習として各ジェスチャーを最低1回ずつ実行させた。

被験者は、本学に所属する学生6人である。年齢は21歳から25歳の男性であり、全員右利きであった。4人は前の実験に参加してジェスチャーを知っており、2人は前の実験に参加していないのでジェスチャーを知らない被験者とした。

実験時の様子は図8である。

6.2.2 アンケート

実験終了後に、本システムに対するアンケートを実施した。アンケートでは、実行しやすさ、覚えやすさ、心地良さ、楽しさ、疲れやすさの主観評価を5段階のリッカート尺度で記入させた。そして、良い点・悪い点を自由記述で記入させた。

6.3 結果

実証実験でのジェスチャー認識率を図9に示す。

全ジェスチャーの平均認識率は、被験者1が85.5%、未経験者が78.7%となり、全員の平均認識率は83.3%となった。平均認識率は、ジェスチャー認識実験は78.5%、向き認識実験では正面での認識率が66.0%だったため、本実験の認識率の方がジェスチャー認識実験と比較して4.8%、向き認識実験と比較して17.3%高い結果となった。これは、前2つ



図7 ジェスチャーと操作の対応

の実験がどのジェスチャーが認識されたかフィードバックが無いのに対し、本実験では動画プレイヤーの操作がフィードバックとなっているため、被験者がジェスチャーを修正できるからだと考えられる。

アンケートで回答させた結果を表4に示す。

表4 主観評価

項目\被験者	経験者 1	経験者 2	経験者 3	経験者 4	未経験者 1	未経験者 2	平均
実行しやすさ	4	4	4	5	3	4	4.0
覚えやすさ	3	4	4	4	5	5	4.2
心地よさ	3	3	3	4	4	3	3.3
楽しさ	4	4	5	5	4	5	4.5
疲れにくさ	3	3	2	3	3	3	2.8

主観評価では、平均値が4未満の結果は心地良さが3.3、疲れにくさが2.8の2点となっ

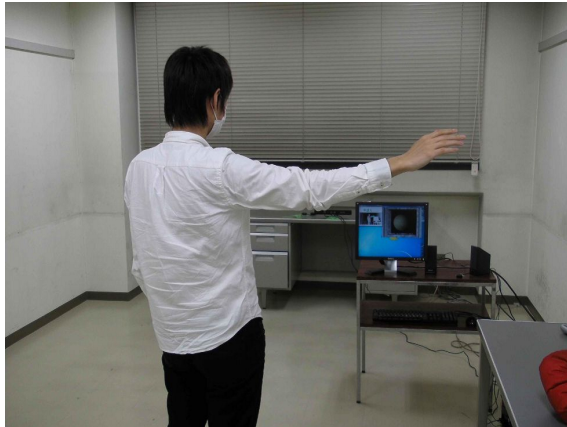


図 8 実験の様子

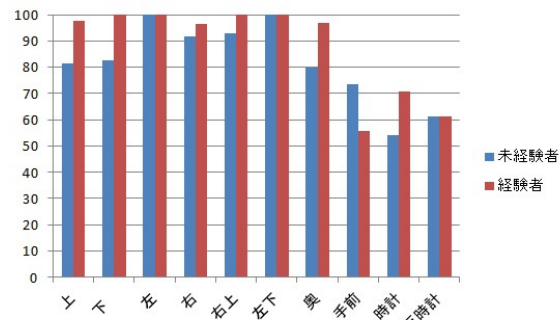


図 9 ジェスチャーごとの認識率

た。これは自由記述より、トリガーである手の静止時間が長い、手の静止がトリガーのため、連続で操作したくてもできない、長期間使用していると肩が疲れる、と言った意見があった。

良い点の自由記述から、ジェスチャー操作が楽しいため操作したくなる、ジェスチャーが直感的である、画面を見ずに操作できるため、顔や視点の移動が必要無いといった点があった。

7. おわりに

本研究では、深度カメラを用いて空中での手によるジェスチャーの動作特性を調査し、手により指示可能な範囲と精度の調査の実施、3次元位置情報を利用したジェスチャーによる電子機器操作手法の提案と実装、ユーザの向きによらずジェスチャーを認識可能にするシステムの提案と実装を行い、それぞれについて実験により有用性を評価した。さらに、実証実験として、上記のシステムを使い、徒手で動画プレイヤーを操作する実験を行い有用性を評価した。

その結果、ジェスチャー認識実験では平均認識率 78.5 % となった。向き認識実験では、左 60 度、左 30 度は 0 度の時と有意差が無く、正面を向いている時と同じように使用できるとわかった。しかし、左 90 度、右 90 度は有意差があり、認識率が大きく減少する結果となった。これは、90 度横を向いているため、ジェスチャーを行う右手が体に隠れてしまう事が原因だと考えられる。そして、右方向に向いた時は、下にスライドするジェスチャーが左下にスライドするジェスチャーと誤認識されることが多く存在した。

動画プレイヤーを操作する実証実験では平均認識率 83.3 % となった。ユーザのアンケートでは、実行しやすさ、覚えやすさ、心地よさ、楽しさについて平均より良い結果が得られた。疲れにくさは評価 2.8 と低い評価となった、これは、ジェスチャー - 開始のトリガーが手を静止させる事にしたため、トリガーの誤動作があった、連続して操作したい時に静止時間のため連続して操作できないという点があげられる。

参考文献

- 1) Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S. : A Hand Gesture Interface Device; CHI+GI '87, pp.189-192 (1987)
- 2) Sato, Y., Saito, M., Koike, H.: Real-Time Input of 3D Pose and Gestures of a User's Hand and Its Applications for HCI; IEEE Virtual Reality Conference 2001, pp.79-86 (2001)
- 3) Andrew D.W., Hrvoje, B.: Combining Multiple Depth Cameras and Projectors for Interactions On, Above, and Between Surfaces; UIST '10, pp.273-282, (2010)
- 4) 長谷川 秀太, 角田 博保, 赤池 英夫: 空中での手によるジェスチャーを利用したインタラクション手法のための調査と検討; ヒューマンインタフェースシンポジウム 2011, 2535D (2011).