

指の開閉動作を用いた多人数向けテーブルトップエンタテインメントシステムの実装

Tabletop Entertainment System with a Finger Opening and Closing Gesture

間宮 暖子 佐藤 俊樹 福地 健太郎 小池 英樹*

Summary. 従来のビデオゲームにおけるユーザインタフェースは、押しボタン式のゲームパッド等を用いるのが一般的であったが、近年、より直感的なインタフェースに注目が集まっている。そこで本研究では、手軽な指の開閉動作をトリガとして用いた、フリーハンドの直感的な動作で遊べるユーザインタフェースを提案する。今回実装した認識システムでは、カメラで複数の手を認識することで、指の開閉動作を検出することができ、同時に手の位置や腕の角度も認識することができる。さらに、数百 fps での高速な認識が可能で、多人数で同時に操作を行った場合にも高速性が損なわれないという利点がある。また本研究では、このシステムを応用した多人数向けテーブルトップエンタテインメントアプリケーションを実装し、実際にプレイしてもらうことで意見を聞いた。

1 はじめに

従来のビデオゲームのユーザインタフェースは、押しボタン式のゲームパッド等を用いることが一般的であった。しかし近年、より直感的なエンタテインメント向けのユーザインタフェースに注目が集まっている。入力デバイスを用いて直感的な入力を行う手法もあるが、デバイスを用いた場合に、ゲームの種類によっては、どうしてもデバイスを通した間接的な操作間隔を感覚を得てしまうという問題がある。

一方で、画像認識を用いたビジョンベースの認識手法を用い、このような入力デバイスを一切使わず、ユーザの身体動作を認識することで、直感的で、直接的な入力を実現する手法もある。フリーハンドで入力することにより得られる楽しさもあり、さらに、入力デバイスを用意したり、装着したりする必要が無いため、誰でも参加でき、より多人数で遊ぶことができるゲームを実現できる。しかし、ビジョンベースの認識手法では、ビデオゲームの入力に用いる際に必要な入力の高速度の実現に問題があった。一般的に、アクションゲームのような比較的動きの激しいビデオゲームにおいては、入力が行われた際にシステムがその高速な入力に追従できないと、ユーザはストレスを感じる。特に、多人数で同時に操作を行った場合には、認識速度の低下は顕著であった。

そこで本研究では、手軽な指の開閉動作をトリガとして用いたフリーハンドによる直感的な入力が行え、多人数に対応でき、また高速性も持ったエンタテインメント向けのユーザインタフェースを提案する。

2 指の開閉動作

指の開閉動作は図1のように、親指と人差し指の先をくっつけたり、離したりする動作である。このような指の動作は、物を摘む時の動作に似ている。

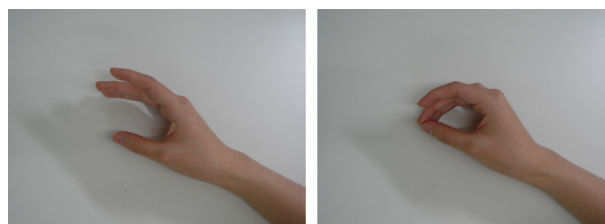


図 1. 指の開閉動作

この動作を用いることは、次のような利点がある。まず、テーブル上で手軽に行うことができ、簡単であり、また疲れにくい点が挙げられる。また、指を開いた状態と閉じた状態の区別が明確に行えることに加え、高速に開閉の切替を行うことも可能であるため、高速な入力にも対応できる。さらに、シンプルな画像処理を用いて、高速で、ロバストな検出が可能である点も挙げられる。テーブル上の手を使ったジェスチャとしては、指差し動作が代表的であるが、指先の検出は処理が複雑になりがちで、画像処理のコストも高くなってしまふ恐れがある。一方で開閉動作を認識するには、指で囲まれた領域の有無を調べるだけの簡単な処理で済み、また手の位置や方向についても同時に取得することが可能である。

3 関連研究

親指と人差し指を使ったジェスチャ入力を行うユーザインタフェースは、これまでもいくつか研究がな

Copyright is held by the author(s).

* Haruko Mamiya, Toshiki Sato, Kentato Fukuchi and Hideki Koike, 電気通信大学大学院 情報システム学研究科 情報メディアシステム学専攻

されてきた。まず Wilson らは、親指と人差し指を使った摘む動作を応用したインタフェースの TAFFI[5] を提案した。TAFFI は親指と人差し指で摘んだ際の、指で囲まれた領域を認識することで、摘む動作、手の位置、手の回転をロバストに検出でき、片手、もしくは両手を使って摘んでひっぱるといったジェスチャ入力を実現した。しかし、TAFFI が手の位置、手の回転に着目している点に対し、本システムでは腕と手の位置から方向を設定することで、計算コストを軽減している。また、TAFFI は 1 人用のシステムで 30fps 程度なのに対し、本システムではより高速な動きの認識に対応しており、多人数での操作も考慮している。

一方、本研究室では、親指や人差し指を使い、指で物を弾く動作を認識することのできるおはじきインタフェース [3] の研究を行っている。おはじきインタフェースが認識対象とする指で物を弾く動作は、指に力を入れて物を弾き飛ばす動作であるが、本システムで用いる指の開閉動作は、指に力を入れずに行うものであり、開閉の動作を連続的に行うことが可能である。おはじきインタフェースにおいては、ユーザの弾いた強さの推定を行っているが、本システムではユーザの手の位置と角度、開閉の検出のみを簡易に認識することで、高速性を損なわずに多人数での利用に対応した。

テーブルトップシステムのエンタテインメント利用を行った研究としては、ボードゲームの再現を得意とする Philips 社の Entertaible[4] 等がある。多くのテーブルトップエンタテインメントシステムでは、手の位置や形を認識することは可能だが、動作を認識することはできない。本システムでは身体動作の認識を行い、高速なジェスチャ入力を用いたゲームの実現が可能である。

デバイスを用い、指の開閉動作を検出する研究として、加速度センサを利用して指をタップする動作を検出し、タッピングの間隔を用いて入力を行うことが可能な福本らの UbiButton[2] 等がある。このようなデバイスを利用した手の動きの認識手法には、複数人にデバイスを着脱する手間や、デバイスが高速な動作の妨げになることがある。対して、本システムにおいては、ユーザは特別な用意無しにすぐにご利用することができるという利点があり、そのため、他のユーザがシステムを利用しているときに、割り込んですぐさま共に利用することもできる。

4 システム構成

本システムでは、テーブル上での指の開閉動作における、指を開いた瞬間、閉じた瞬間の検出を行う。さらに、テーブル上のユーザの手の位置、狙った方向の検出も同時に行う。

開閉動作の検出には、手にセンサ等のデバイスを装着する手法もある [1]。しかし、開閉動作の検出、

手の位置、狙った方向といった複数の情報を取得する必要があるため、複数のセンサ等を手に装着する必要が発生し、装置の装着が煩雑になったり、ユーザの動作の邪魔になるという問題が考えられる。そこで本システムではビジョンベースの認識手法を用いて認識を行った。ビジョンベースの認識手法は、カメラと画像処理を用いてユーザの動作の認識を行う手法であり、ユーザがセンサ等のデバイスを装着する必要がないという利点がある。

また、高速な動きに対応するために、本システムでは高速度カメラを用いて撮影を行った。高速な処理を行うためには、画像処理のコストを軽減させる必要がある。そこで本システムでは複雑で処理コストの高い画像処理は行わず、背景差分や領域の検出といったシンプルで高速に処理が可能な画像処理のみを用いることで認識を高速化した。背景差分を精度良く行うためには、ディスプレイ画像による認識への影響を避けなければならない。本研究では、偏光フィルタと液晶ディスプレイを組み合わせる手法 [6] を用いた。液晶ディスプレイから発せられる光は既に偏光しているため、カメラのレンズに偏光方向に直交させた偏光フィルタを装着することで、カメラに入るディスプレイ映像を遮断し、ディスプレイ画像に影響を受けない認識を可能にした (図 2)。

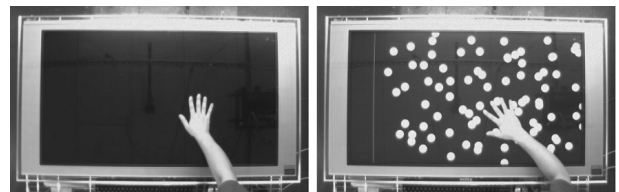


図 2. 偏光フィルタ装着後 (左) と装着前 (右)

テーブルトップシステムを実現するには、液晶ディスプレイをテーブルとして用いる手法のほかに、プロジェクタを用いてテーブル上に投影を行う手法がある。しかしプロジェクタを用いた場合、部屋を暗くする必要があり、高速度カメラによる撮影に支障を来す恐れがある。また、システムの規模が大きくなり設置が困難になったり、投影した画像がゆがむといった問題もある。

本システムのハードウェア構成は図 3 のようになる。カメラはテーブル面から約 170cm 上方の天井に固定し、テーブル全体を含む画像を撮影する。カメラは Point Grey Research 社の DragonflyExpress を使用した。このカメラは 640×480 pixel の解像度で最大 200fps の撮影が可能な 8bit グレイスケールカメラである。また ROI (Region of Interest) を設定することで、 320×240 pixel で 350fps、 160×120 pixel で 450fps といったさらに高速な撮影が可能になる。テーブルとしては、表示部分 $102\text{cm} \times 58\text{cm}$ の大型液晶ディスプレイを用い、高さ約 100cm の台に横

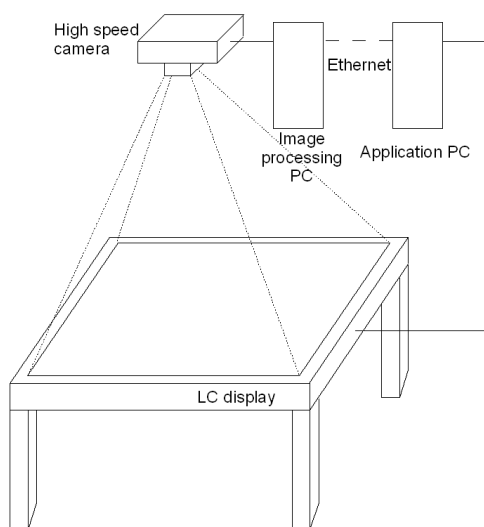


図 3. ハードウェア構成

置きに設置した。この時のカメラ画像におけるディスプレイの表示部分の空間解像度は約 2.3mm/pixel になる。カメラの画像はリアルタイムに画像処理用 PC (Xeon 3.2GHz, 2GB RAM) 上で処理される。結果は Ethernet 経由でアプリケーションが動作するアプリケーション用 PC (PentiumD 3.2GHz, 4GB RAM) に送信され、アプリケーション画像はテーブル型のディスプレイに描画される。

5 認識の流れ

開閉動作の検出の流れは次のようになる。

1. キャリブレーション・背景の取得
カメラ座標とディスプレイ座標との変換を行うため、射影変換を用いたカメラキャリブレーションを行っておく。また背景差分用にカメラ画像内に人が入っていない状態の画像を背景画像として保存しておく。これらの処理はプログラム開始時に 1 度だけ行う。
2. 腕の検出
背景差分を用い、背景画像とカメラからの入力画像の差分を取ることで、テーブル上の全てのオブジェクトの領域の 2 値画像を得る。次に、得られた複数の領域の 2 値画像に対し、領域分割を行うことで連続領域の検出を行い、またそれぞれの領域の面積を計算する。さらにこの領域の面積を基に、約 30cm² 以上の領域のみを抽出することで、図 4(左, 中央) のような腕の 2 値領域を腕の個数分得る。また同時に、腕領域の重心を求め、テーブル上での腕の位置を求める。
3. 手の状態の検出

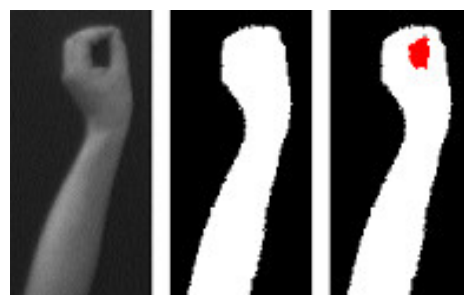


図 4. 入力画像 (左) と腕の抽出結果 (中央), 及び指で囲まれた領域の検出結果 (右)

検出したユーザの手が図 1(左) のような開状態にあるか、もしくは図 1(右) のような閉状態にあるかを検出する。まず 2 の検出処理で得られた全ての腕領域に対し、ピクセルの階調を反転させ、2 と同様に連続領域の検出と領域分割を行い、約 1cm² 以上の領域を検出し、腕領域内に図 4(右) のような、閉状態の手領域のみに存在する指で囲まれた小領域 (白ピクセル内の黒ピクセル) が存在するかを調べる。この領域がある場合、手が閉状態であると判断し、無い場合は開状態と判断する。

4. 位置と角度の検出

閉状態にある手の位置は、3 で検出した腕領域内の指で囲まれた小領域の重心を計算することで求める。また同時に、手の位置と 2 で求めた腕の位置の 2 点を用い、図 5 のように腕から手の方向のベクトルを求め、ユーザが狙う方向を設定する。

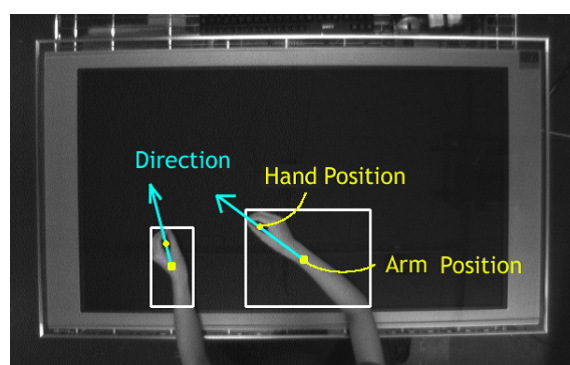


図 5. 腕の位置と手の位置を用いた手の方向の設定

5. 腕の対応付け

前フレームで検出された腕と、現フレームで検出された腕の対応を調べる。まず前フレームで検出された全ての腕の位置と、現フレームで検出された全ての腕の位置の距離をそれぞれ調べ、6cm 以内で最も位置に近い腕を同

じ腕と判断する．また，現フレームで検出された腕について，前フレームにおいて位置の近い腕が発見できなかった場合は，これを新しく検出された腕とみなす．逆に，前フレームで検出された腕で，現フレームで位置の近い腕が見つからなかった場合，この腕は無くなったとみなす．

6 アプリケーション

本研究では，これらのシステムを用い，次のようなテーブルトップアプリケーションを実装した．

6.1 シューティングゲームへの応用

指の開閉動作の認識を応用したテーブルトップアプリケーションの一つとして，シューティングゲーム「AGEBall Shooting」を実装した．



図 6. テーブルトップシューティングゲーム

このゲームは，図 6 のように，手を用いて弾を撃つ位置や方向を指定でき，複数人で遊ぶこともできるゲームである．手を動かすことにより，キャラクターの位置を自由に移動でき，また腕の曲げ伸ばしや手首の角度により弾を撃つ方向を指定することができる．また，両手を使うことにより 1 人で 2 キャラクターまで同時に操作することができる．さらに，図 7 のように，複数人でテーブルを囲み，4 人以上で協力して遊ぶこともできる．

ユーザがテーブル上に手を出して閉状態とすると，手の位置にマーカが表示され，同時に弾の発射方向を示す矢印が表示される．このマーカと矢印は，常にユーザの手の位置に追従し，矢印によってユーザは簡単に狙いを付けることができる．閉状態から指を開き開状態に変化させると，弾が矢印の方向へと発射される．こうしてユーザは指の開閉状態を繰り返して弾を撃ち，画面上に次々と現れる敵を倒して遊ぶことができる．

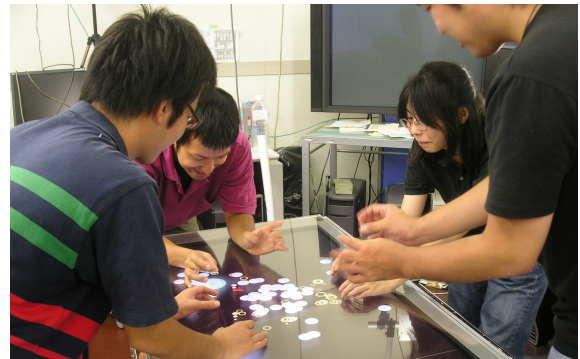


図 7. 多人数でのプレイ

6.2 「手で食べるゲーム」への応用

次に，図 8 のような，テーブル上に表示された食べ物を手で食べて遊ぶゲームを実装した．このゲームは，テーブル上に表示された食べ物を，口でついで指の開閉動作を行うことで，表示された食べ物を食べて遊ぶゲームである．食べ物は様々な大きさがあり，プレイヤーは様々なルールに従い，他のプレイヤーと時には協力し合い，時には競い合いながらテーブル上の食べ物を食べて遊ぶことができる．このゲームの場合，テーブル上の食べ物を早く食べる必要があり，この時に高速な開閉動作の認識の必要性が生まれる．

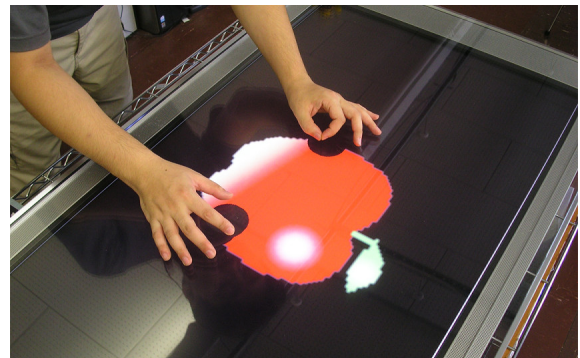


図 8. 「食べる」ゲーム

7 考察

7.1 認識速度について

本システムでは最大 200fps 以上で撮影が可能な高速カメラと，低処理コストの画像処理を用いることで，テーブル上に手が複数本ある場合にも高速な認識が行えるようにしている．そこで実際に，本システムに多数の腕を認識させた場合のフレームレートを調べたところ，テーブル上に 20 本の手を同時に入れた場合でも 180fps での検出が可能であった．ま

た、テーブルの周囲や縁を除いたディスプレイの表示面だけに、約 450×250 pixel の ROI を設定して認識を行ったところ、手が 20 本の状態でも約 300fps での検出が可能であった。従って、今回テーブルの大きさでは同時に遊べる人数は高々 4 人から 6 人程度であると考えられるため、認識の速度は十分であると言え、今後テーブルを大型化し、さらに多人数での認識を行う場合にも十分対応が可能であるといえる。また、より高速な CPU を使用することや、より高速に撮影が可能なカメラの使用により、さらに認識速度を高めることも期待できる。

7.2 ユーザ評価

今回実装したシューティングゲームアプリケーションについて、本研究室の大学院生と、本学オープンキャンパスの参加者約 30 人に対し実際に遊んでもらい、遊んでいる時の被験者の様子や被験者から得られた感想を次のように考察した。

まず、どの被験者も、特別な操作の説明をせずとも、身振りによる簡単な説明ですぐに動作を理解でき、遊ぶことができていたことが挙げられる。前に遊んでいた人の様子から、見よう見まねでゲームに参加する人もいた。また、両手で遊んだ場合にも、操作の混乱等は見られなかった。これらのことから、今回着目した指の開閉動作がシンプルな動作であると共に、両手でも直感的に扱える動作であるといえる。

次に、複数人で遊べる点が良いという意見が多く得られた。複数人で同時に遊ぶことができるビデオゲームについては、これまでゲームセンターや家庭用ゲーム機においても様々存在していた。しかし本システムのように、テーブルを囲んで複数人で遊ぶことができるビデオゲームは少なかった。また、テーブル上には他のプレイヤーの手が存在し、テーブルの周りにはプレイヤーの体が存在するため、他のプレイヤーの手を避けながら移動したり、テーブルの周りを歩き回ることによって、より高い現実感が得られると考えられる。さらに、プレイヤー同士の手や体が接触することで、より対面的なコミュニケーションが促進されることも考えられる。

次に、連続で弾を発射することで、爽快感が得られるという意見もあった。また長時間のプレイでも、手が疲れるという意見は少なかった。加えて、多くのプレイヤーが、連続で弾を発射することで、テーブル上に表示された弾の飛ぶ方向を示す矢印を用いずに、狙った方向に弾を飛ばしていることがわかった。

一方で、複数人で遊ぶには現在のテーブルでは狭すぎるという意見も得られた。今回実装したシステムでは、46inch の液晶テレビを用い、同時に遊べるプレイヤーの数を最大 4 人までに設定していたが、ユーザがテーブルの一方向へ偏るとユーザ同士の体がぶつかり、狭く感じられてしまう場合もあった。また同時に、もっと多人数で遊びたいという意見も

得られた。これらの点については、今後テーブルの大型化を行い、対応していく予定である。

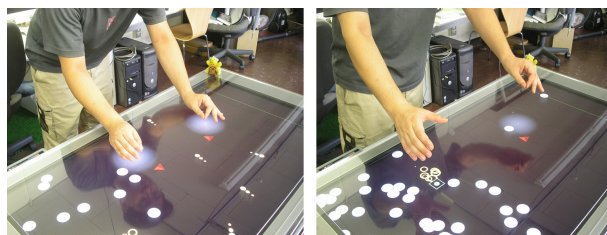


図 9. 手を密着させた場合 (左) と、手を離れた場合 (右)

またユーザによっては、図 9(左) のように、手をテーブルに密着させて動作を行う人と、図 9(右) のようにテーブルから数十 cm ほど離して動作を行う人がいた。今回、ユーザには手の高さについては説明を行わなかったため、ユーザは好みに応じて手を密着させた位置や、離れた位置で遊んでいたと思われるが、ステレオカメラを用いることで、このテーブル上の手の高さも入力に利用することもできると考えられる。例えば、今回のようなシューティングゲームでは、手をテーブルに近づけたり離したりすることで、近づけると攻撃、遠ざけると回避といった攻撃と回避の切替の要素を加えることもできる。

システムのロバスト性については、テーブルを覗き込んだユーザの頭が手を覆い隠してしまい、手がカメラに撮影されなくなる現象も見られた。この現象は、カメラを斜めに配置するか、複数台設置し、隠れる場所を減らすことで軽減できると考えられる。また、テーブルの境界付近にユーザの手がある場合、十分に腕の領域を抽出することができずに、正確な方向が設定できない現象も見られた。この問題については、テーブルを広くすることや、今回の実装では認識範囲外であったテーブルの縁まで認識範囲に入れることで解決できると考える。

7.3 高速な操作

従来の十字ボタンを用いた入力手法では、ボタン等を押すことで、ボタンを押した方向へキャラクターを移動させることができた。この際、ボタンを押す強さで移動速度を変化させることもできるものもあったが、ボタンの数により移動方向が限られている点や、ボタンによる操作の難しさにより、キャラクターをプレイヤーの意図したとおりに高速に動かすことは非常に困難であった。マウスを用いた移動は、この点でボタンベースの移動に比べ簡単かつ高速に操作を行うことが可能であるが、本システムではフリーハンドで、さらに、両手を使って操作を行うことができるため、ジェスチャ入力による操作が高速に、かつ簡単に行うことができる。また、認識のフレームレートも高いため入力の遅延も少ない。

7.4 開いた瞬間と閉じた瞬間の使い分け

今回実装したシューティングゲームにおいては、指を開いた瞬間に弾が発射されるようにしていた。そこで指を開いて弾を発射する操作について、発射の瞬間にユーザがどのような感覚を受けるかを実際に複数のユーザに聞いて調べてみた。一方、閉じた瞬間に弾が発射されるようにした場合のことも考え、実際に弾が閉じた場合に発射されるように設定を変更し、同様に意見を聞いてみた。まず開いた際に発射される設定での意見は、「物を投げている感覚」が強いという意見が多くあり、加えて「弾いている感覚」に近いという意見や、「指を開いた際の隙間から何か魔法のような物が出ている感覚」があるという意見も得られた。次に、閉じた場合については、「指でボタンを押して弾を発射している感覚」に近いという意見が多かった。しかしいずれの場合にも、操作には直感性があり、楽しく遊ぶことができるという好意的な回答が得られた。また、今回のシューティングゲームでは、1発ずつ狙って弾を発射するのではなく、連続で弾を発射することが殆どである。この場合、閉じた瞬間と開いた瞬間がほぼ同時になることから、連射を行ううえではどちらで操作を行っても良いと考える。

一方、もう一つのアプリケーションである「食べるゲーム」においては、「食べる」時の口の動きを指を閉じる動作に見立てているため、指を閉じる動作以外の入力では「食べる」感覚が得られず、ゲームとしての趣旨が変わってしまうという問題があり、相応しい操作であるとは言えないだろう。

以上のことから、閉じる動作と開く動作では動作のイメージが異なるため、ゲームによっては適切に動作の選択を行わないとゲームの直感性を損なう恐れがあるが、連続で入力を行うゲームの場合は、どちらで操作を行っても良いことがわかる。

7.5 例外的な動作の認識について

本システムではロバストに指の開閉動作の検出を行えるようにしているが、腕として認識された領域内に小領域があれば、指の開閉動作でなくても認識された状態になってしまう。例えば両手の指を使って輪を作ったり輪を崩したりする動作や、手を開いたり握ったりすることで、指の開閉と同様の検出が行われることになる。これらの入力手法は本システムにおいては想定外の入力であるが、エンタテインメント向けのシステムであることから、本来の入力の妨げになるような問題ではないため、このような派生的な入力手法を制限する処理を特に加えたりはしなかった。

7.6 アプリケーションについて

今回実装した2種類のアプリケーションの他にも、指の開閉動作を用いた物を摘んで運ぶゲームや、投

げるゲームが考えられる。また、両手を有効に使ったアプリケーションとして、例えば両手間の距離、手の位置関係を用いることで、弓矢を引く動作における、狙った方向と、弦を引く強さの入力を用いたゲームが考えられる。今後、これらについては実装を行い、運用、及び評価を行う予定である。

8 今後の展望

現実装では、テーブル上の手の追跡を行うことはできるが、その手がどのユーザのものなのかを識別することができない。ユーザの識別を行う方法としては、テーブル上だけでなく、テーブルの周囲のユーザの位置を取得し、テーブル上の手と対応付ける処理を行うことが考えられる。ユーザの識別が可能になると、両手を使った高速なジェスチャ入力や、他のユーザと協力した入力も行えるようになる。また、ユーザ毎の得点等のゲーム上必要なパラメータの記録も行うことが可能になる。

さらに今回の実装では、指を閉じている状態のみ、手の位置を検出することができた。しかし、カメラで抽出した手の形状データをアプリケーションで直接用いることで、例えばシューティングゲームで手に接触判定を持たせたりすることも可能となる。

現在、図1のような指の開閉動作だけの認識を行っているが、これに似た動作の、手の甲を上にしたたり、手の平を上にしての指の開閉動作も認識できるようにしたい。このような動作も認識するためには、カメラを傾けて設置する方法等が考えられる。

参考文献

- [1] D. A. Bowman, C. A. Wingrave, J. M. Campbell, V. Q. Ly, and C. J. Rhoton. Novel Uses of Pinch Gloves for Virtual Environment Interaction Techniques. In *Virtual Reality*, pp. 122–129. Springer London, 2004.
- [2] M. Fukumoto and Y. Tonomura. Whisper: A Wristwatch Style Wearable Handset. In *Proceedings of the SIGCHI*, pp. 112–119. ACM Press, 1999.
- [3] T. Sato, K. Fukuchi, and H. Koike. OHAJIKI Interface: Flicking Gesture Recognition with a High-Speed Camera. In *Proceedings of ICEC'06*, Vol. 4161 of *Lecture Notes in Computer Science*, pp. 205–210. Springer, 2006.
- [4] E. van Loenen and et al. Entertaible: A Solution for Social Gaming Experiences. In *Proceedings of IUI'07*, pp. 16–19, 2007.
- [5] A. D. Wilson. Robust Computer Vision-Based Detection of Pinching for One and Two-Handed Gesture Input. In *Proceedings of UIST'06*, pp. 255–258. ACM Press, 2006.
- [6] 西川 渉, 他. 偏光を応用したテーブルトップシステムの提案. WISS2007 論文集, 2007.