

大画面ポインティングシステムのための 回帰モデルによる単眼画像からの指差し位置の推定

新谷 晃 市^{†1} 間下 以大^{†1,†2}
清川 清^{†1,†2} 竹村 治 雄^{†1,†2}

近年、ショッピングモールや駅といった公共の場で、ディスプレイの大型化が進んでいる。そのため、大型ディスプレイに適したポインティングインタフェースの必要性が増している。デバイスを把持する方法では、Wii リモコンやレーザーを用いたインタフェースなどが提案されている。しかし、デバイスを把持しない方法では決定的なインタフェースがない。本研究では、デバイスを持たずに、直感的に大型ディスプレイとのインタラクションを可能にするポインティングインタフェースを提案する。提案手法は単眼画像を元にユーザの指差し動作を学習し、指差し位置を推定する。学習と推定手法には、主成分回帰モデルを用いた。プロジェクタを用いて試作システムを作成し、提案手法の有効性の評価を行った。

Pointing Interface for a Large Screen based on Regression Model of Monocular Images

KOICHI SHINTANI,^{†1} TOMOHIRO MASHITA,^{†1,†2}
KIYOSHI KIYOKAWA^{†1,†2} and HARUO TAKEMURA^{†1,†2}

As large screens become more and more common in areas like shopping malls, designing a suitable pointing interface becomes increasingly important. A class of pointing interfaces involves holding a special device, such as a Wii remote, but there has not been a conclusive interface presented requiring no device at all. This thesis presents a pointing interface that allows intuitive interaction with a large screen. The interface is trained against monocular images of hand pointing and estimates the coordinate being pointed at by a user. A PCA regression model is constructed for learning and estimation. This method was implemented in a prototype system with a large, projected screen, and experiments with the prototype system confirmed the validity of the proposed method.

1. はじめに

ここ数年におけるコンピュータの一般化には目覚ましいものがある。それに伴ってディスプレイの大型化も進んでおり、駅構内等の公共の場所においても大型ディスプレイが普及してきた。一方、ポインティングインタフェースにおいても、これまでのマウスやトラックボールのような機械を操作するものに加え、タッチパネルや位置計測を利用したデータグローブ¹⁾ などユーザの動作をそのまま機械に伝えるインタフェースが一般化してきた。また、これまで指示対象を示すために用いられてきた指示棒やレーザーにディスプレイとのインタラクション機能を持たせる研究も行われている²⁾。

現在、公共の場所におけるディスプレイとのインタラクションには、主にタッチパネルが利用されている。これは、タッチパネルが利用者が何も持たずに、対象に触れるという直感的な動作によって操作できるインタフェースであるからと考えられる。しかし、タッチパネルは手の届く距離での利用に限定されるという欠点がある。天井や壁面など利用者の手の届かない位置に設置されたディスプレイとのインタラクションを考慮すると、レーザー等の指示動作による非接触で利用可能なインタフェースが求められる。

一方、心理学や認知科学の分野では、指示動作は人間の視覚や空間認知特性と深く関係していることが知られており、研究対象とされている^{3),4)}。Brain⁵⁾によると、人間の空間認知は手の届くような距離 (reaching distance) と手の届かない距離 (walking distance) に分類されている。手の届かない距離 (walking distance) において、吉田⁴⁾らにより、空間認知特性によって、実空間と認知空間に誤差が生じることが確認されている。そのため、レーザーは手の届かない距離にあるような大画面とのインタラクションにおいては、認知特性の影響により、ユーザの動作をそのまま機械に伝えることができない。

今後、大型ディスプレイとのインタラクションを取る機会が増えると思われる。そのとき、既存のポインティングインタフェースでは認知特性による誤差や、手の届かない距離にある大画面での利用は考慮されていないため、様々な不便が予想される。例えば、タッチパネルは、机や台の上の小型ディスプレイのポインティングには有効であるが、大型ディスプレイはユーザとディスプレイの距離が遠いため不向きである。また、データグローブは距離

^{†1} 大阪大学大学院 情報科学研究科

Graduate School of Information Science and Technology, Osaka University

^{†2} 大阪大学 サイバーメディアセンター

Cybermedia Center, Osaka University

が遠くても有効であると考えられるが、装着の手間が煩わしいため、ユーザが気軽に扱えるものではない。

このことから、大画面に対する指示動作をそのまま機械に伝えるために、ユーザの動作に基づく新しいインターフェースが必要であり、これから増えると思われる大型ディスプレイとのインタラクションを持った、ユーザが自然に扱える大画面用のポインティングインターフェースが望まれる。そこで本研究では、大画面に対しユーザの自然な動作によるポインティングが可能なインターフェースを提案する。

2. 関連研究

本章では既存のポインティングインターフェースを整理し、大画面での利用に適した、空間認知特性を考慮したインターフェースについて検討する。

2.1 これまでのポインティングインターフェースについて

まず、ポインティングインターフェースについて説明する。マウスやデータグローブなど、様々なインターフェースがあるが、以下の基準により分類した結果を図1に示す。

機械中心、人間中心による基準

機械中心のインターフェースは、ユーザが機械に合わせた入力を行うインターフェースとし、人間中心のインターフェースは、機械がユーザに合わせた出力を行うインターフェースとした。例えばマウスの場合、キーボードのように機械的な入力をユーザに要求せず、タッチパネルのようにユーザが直感的に扱えるものではないため軸の中心付近に位置する。

利用時の画面の大きさによる基準

主に利用されている画面の大きさによって、そのインターフェースの位置を決定した。本研究の目的である、大画面に対しユーザの自然な動作によるポインティングが可能なインターフェースは、図1では右上に位置する。

2.2 大画面で利用可能なポインティングインターフェース

本節では、大画面で利用可能なポインティングインターフェースの例を挙げ、その手法を紹介する。

- Vogelら⁶⁾の手法はモーションキャプチャカメラシステムを用いる。ポインタの座標をユーザ座標系に合わせ、手袋に装着したマーカによって手の3次元情報を得ることで、ユーザの指先方向へのポインティングや、手の相対移動量を基準としたポインティングを実現している。
- 中村らのHands-In⁷⁾は、スクリーン付近に設置した単眼カメラを用いる。カメラ画像



図1 ポインティングインターフェースの分類

に映ったユーザの手を囲む矩形領域を決定し、矩形の相対移動量を基準としたポインティングと、矩形の位置を基準としたポインティングを実現している。

- Wiiリモコン⁸⁾は、赤外線カメラと赤外線LEDランプを用いる。画面付近に赤外線LEDランプを配置し、Wiiリモコンに内蔵されたカメラで撮影されたランプの位置を基準にポインタの座標を計算する。

これらのインターフェースは、大画面に対する指示動作をカメラで捉えることにより、人間中心よりのポインティングを実現している。しかし、ポインティングの基準となるものは、指差し時の指先方向や把持しているものである。このため、空間認知特性を考えた真に人間中心のインターフェースとはいえない。

2.3 カメラを用いたインターフェース

カメラを用いたインターフェースは、コンピュータビジョン、パターン認識を利用してユーザの動作を認識し、入力として扱う。この分野の研究では、EnhancedDesk⁹⁾や、比嘉らの研究¹⁰⁾などがあり、指先検出やジェスチャ認識を用いることで様々な操作を実現している。

カメラを用いたインターフェースの長所はユーザがデバイスを特に意識することなく、いつでも操作が行える点である。そのため直感的なインターフェースの作成が可能である⁷⁾。例えば手のひらの位置を認識し、握るジェスチャをすることで対象物を掴むなど直接的な操作の実現を可能とする。また、デバイスの装着も基本的には必要ないため、ユーザに与える拘束感が少ないといった利点が挙げられる。しかし、環境の影響を受けやすいという欠点があ

る。以下は一般的に挙げられる、カメラを用いたインタフェースの問題点である。

- オクルージョン：物体が重なっているとき、1台のカメラで得られる単眼画像では、必要な情報を得ることができないという問題が発生する。この解決法としてはカメラを複数用いてオクルージョンを軽減する方法やモデルマッチング¹¹⁾を用いて形状を推定する方法などがある。
- 必要領域の抽出：カメラから得られた動画像には、インタフェースの処理には不必要な情報が多く含まれている。そのため必要領域のみを抽出する必要がある。Vogelらの研究⁶⁾では、赤外線センサカメラによって手に装着したグローブの動作の情報のみを得ている。このように、撮影デバイス自体を工夫することで、必要な情報のみを取得するのがハードウェアの解決である。別の手法として、肌色抽出や背景差分を取るなどの画像処理を行い、必要な情報のみを取得するソフトウェア的解決もある。
- 利用環境の動的変化：カメラで取得される動画像は、周りの光源や光源位置などの照明環境に大きく影響される。また、背景の変化が起こる場合は、前述の背景差分を取ることは困難となる。肌色抽出を行う場合も、周囲に肌色に近い物体が多ければ難しくなる。
- 計算量：画像から人の動作等を認識する手法は数多く存在するが、一般に認識精度と計算量にはトレードオフの関係がある。ポインティングインタフェースへの応用を考えると、入力に遅延が発生しない程度の計算量でなければならないため、計算量の少ない単純な手法でありながら、使用に耐えうる精度が得られねばならない。

カメラを用いるポインティングインタフェースのためには、これらの問題を解決し、ユーザの動作を安定して認識するシステムにする必要がある。

人間中心のポインティングインタフェースを考えた場合、把持・装着物の使用は指示動作に影響を与える可能性があり、人間中心とはいえない。カメラを用いた場合、把持・装着物は必要なく、指示動作時の画像を入力として扱えるため、人間中心のポインティングインタフェースの実現が可能である。しかし、利用環境に影響を受けやすく、計算量によってはインタラクション性が低くなってしまふ。

そこで、大画面ポインティングインタフェースとして、以下の項目を解決すべき条件とする。

条件 1 空間認知特性の影響を考慮した指差し位置推定

条件 2 非接触、非把持

条件 3 大画面で利用可能

従来研究におけるカメラを用いたインタフェースが各条件を満たすか否か表 1 にまとめ

た。久米らの手法¹²⁾は手首にカメラを装着し、指に反射材を装着する。そしてカメラ画像内の反射材の座標によってポインティングを行う。佐藤らの IHP¹³⁾では、人差し指でポインティングを行い、指先の延長線を指示方向としている。表 1 に示すいずれの研究においても、ポインティングの基準は指先の延長であったり、カメラ画像内の必要領域の座標などである。また、EnhancedDesk⁹⁾は、条件 1、条件 2 を満たすが、机上 (reaching distance) での利用が想定されているため、大画面 (walking distance) で利用可能という前提からは外れている。上記の条件のうち非接触・非把持で利用可能なポインティングインタフェースはすでに提案されているものの、人間中心のポインティングインタフェースは提案されていない。

3. 提案手法

3.1 アプローチ

前章までに述べたように、人間中心のポインティングインタフェースが求められている。人間中心のポインティングインタフェースとは、空間認知特性を考えたポインティングインタフェースである。

指示動作における空間認知特性は、周りの環境や、指示対象との距離によって影響が異なるとされているが、指示動作にどのような影響があるのかはつきりしていない。よって、空間認知特性をモデル化することは難しいと考えられる。また、個人差も含まれるため、ユーザの指示動作をあらかじめ学習する手法とする。

空間認知特性を考慮するにあたり、指差し動作のどの部分に影響が現れるのかを決めなければならないが、空間認知特性はまだ未知の部分が多いため、特定は難しい。よって、指差し画像をそのまま扱うことができる重回帰モデルを利用する。画像をそのまま用いると回帰係数が 10 万以上となり、計算量が大きく、ポインティングインタフェースとしての操作性を得られない。そこで、本研究では主成分分析によって画像の持つデータ量を圧縮し、主成分回帰モデルとする。

表 1 カメラを用いたインタフェースの例と性質

	空間認知特性	非接触・非把持	大画面
Vogel ⁶⁾	×	△	○
久米らの手法 ¹²⁾	×	△	×
Hands-In ⁷⁾	×	○	○
EnhancedDesk ⁹⁾	○	○	×
Interactive Hand Pointer ¹³⁾	×	○	△



図 2 提案手法の流れ

1	2
3	4

図 3 2×2 の画像の場合、行ベクトルは [1 2 3 4] となる

3.2 提案手法による全体の処理の流れ

提案手法の処理の流れは図 2 のようになっている。処理は学習フェイズと推定フェイズにわかれており、学習フェイズにおける主成分分析および回帰係数推定の結果を推定フェイズで用いる。

3.2.1 教師データの取得

学習に必要な教師データについて説明する。教師データは指差し位置とその位置に対する指差し画像（以下教師画像）の組である。指差し位置を変え、この組を複数得ることで学習データとする。

3.2.2 背景差分、ノイズ除去

背景差分とは、画像から背景を取り除き、移動物体を抽出する方法である。この処理により、指差し動作のシルエットを抽出できる。しかし、背景との部分的な類似などにより、画像にはノイズが乗ってしまうため、ノイズ除去を行う必要がある。学習フェイズ、指差しフェイズのどちらにおいても、この処理を施すことにより環境の影響の軽減が見込まれる。

3.2.3 主成分分析

主成分分析とは、与えられた入力ベクトルのうち、相関関係のあると思われる要因を圧縮して、主成分と呼ばれるいくつかの成分にし、その値を見ることで、入力ベクトルの特性を求めることができる手法である。主成分は最も相関の高い要因から順に第一主成分、第二主成分、などと呼ばれる。提案手法において、主成分は以下の手順によって求められる。

まず、各教師画像を図 3 のようにそれぞれ行ベクトル \mathbf{X} に変換する。これらのベクトルからなる行列 \mathbf{X} の共変動行列 \mathbf{R}_x を $\mathbf{R}_x = (\mathbf{X} - \overline{\mathbf{X}})^T(\mathbf{X} - \overline{\mathbf{X}})$ と定義する。ここで $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \cdots \mathbf{X}_N]^T$ 、であり、 N は教師データの個数である。

次に、 \mathbf{R}_x の固有値、固有ベクトルを求める。 \mathbf{R}_x の固有値を $\lambda_0, \dots, \lambda_N$ とし、それに対応する固有ベクトル $\mathbf{v}_0, \dots, \mathbf{v}_N$ とする。ここで、 $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_N$ とする。このとき \mathbf{v}_1 が第一主成分ベクトル、 \mathbf{v}_2 が第二主成分ベクトル、 \dots 、 \mathbf{v}_n が第 n 主成分ベクトルとなる（ただし n は教師データの個数 N 以下の数）。

主成分ベクトルを用いることにより、入力ベクトルを 1 次元～ N 次元の固有空間に射影することができる。次式 2 によって教師画像 \mathbf{X} から特徴量ベクトル \mathbf{S} が得られる。特徴量ベクトル \mathbf{S} は \mathbf{X} の n 次元固有空間への射影であり、射影行列 \mathbf{A} は式 (1) によって得られる。

$$\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \quad (1)$$

$$\mathbf{S} = \mathbf{X}\mathbf{A} \quad (2)$$

また、以下の式により、 n 次元における主成分の寄与率 P を求めることができる。

$$P = \frac{\sum_{i=1}^n \lambda_i}{\sum_{i=1}^N \lambda_i} \quad (3)$$

3.2.4 回帰係数推定

主成分回帰モデルを作るために、得られた \mathbf{S} に対し以下の式により重回帰分析を行い、偏回帰係数 \mathbf{B} を推定する。

$$\mathbf{Y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}^T \quad (4)$$

$$\mathbf{B} = \begin{bmatrix} b_0 & b_1 & \cdots & b_n \end{bmatrix} \quad (5)$$

$$\mathbf{D} = \begin{bmatrix} 1 & S_{11} & \cdots & S_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & S_{N1} & \cdots & S_{Nn} \end{bmatrix} \quad (6)$$

$$\mathbf{F} = (\mathbf{Y} - \hat{\mathbf{Y}})^2 = (\mathbf{Y} - \mathbf{D}\mathbf{B})^2 \quad (7)$$

式 7 の値が最小となる \mathbf{B} が偏回帰係数である (\mathbf{Y} は正解データ、 $\hat{\mathbf{Y}}$ は推定データ)。この \mathbf{B} を求める式は以下の通りである。

$$\frac{\partial \mathbf{F}}{\partial \mathbf{B}} = -2\mathbf{D}^T[\mathbf{Y} - \mathbf{D}\mathbf{B}] = 0 \quad (8)$$

$$\therefore \mathbf{B} = [\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{Y} \quad (9)$$

以上より、偏回帰係数 \mathbf{B} が得られる。

表 2 FL2-08S2C の仕様

イメージャ	ソニー 1/3" プログレッシブスキャン CCD ICX204
最大解像度	1032×776
最大フレームレート	30FPS
消費電力	2W 以下 (12V)
インターフェース	9pin IEEE-1394b コネクタ, 4pin GPIO
寸法	29×29×30mm
重量	約 58g

3.2.5 指差し位置推定

推定対象となる指差し画像を教師画像同様に背景差分、ノイズ除去の後、図 3 のように行ベクトル I に変換する。この I に対して、学習フェイズにおける式 1 で得られた A と式 9 で得られた B を用いて以下の式により推定結果 \hat{y} を得る。

$$\hat{y} = b_0 + \mathbf{AI} \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}^T \quad (10)$$

4. 試作システム

本章では、前章で述べた提案手法の実装及び試作したシステムの構成について述べる。

4.1 背景差分、ノイズ除去

3 章で述べたように、提案手法はユーザの指差し画像を元にモデル化を行う。本節では、教師データ及び入力となる指差し画像の処理について述べる。

まず、背景となる画像および指差し画像 (640×480 画素) をグレースケール [0~255] に変換する。次に、各画素値の絶対値差分を取り、閾値以上の画素のみを残す。本研究では、背景にあまり変化がない状況を想定しているため、試作システムでは閾値を 20 とした。次に、膨張収縮処理により、反射や影によるノイズの除去を行う。試作システムでは、膨張処理を 2 回実行した後、収縮処理を 1 回実行した。以上の処理を指差し画像に対して行う。

4.2 システムの構成

提案手法では、指差し画像を元にモデル化を行う。教師データとして、指差し位置の情報を得る必要があるため、カメラ以外にプロジェクタを用いた。

以下、システムを試作するにあたり、使用した機器について述べる。

4.2.1 単眼カメラ

試作システムに用いた単眼カメラ Flea2 FL2-08S2C (Point Grey Research) の仕様を表 2 に示す。解像度、出力フォーマット、フレームレートなどを柔軟に変更できる。

表 3 計算機の仕様

OS	Linux Ubuntu8.04
CPU	Intel Core2Duo 3.0GHz
メモリ	DDR2 1.0GB×2

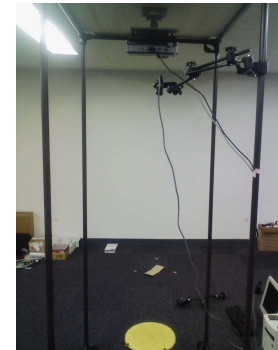


図 4 試作システム

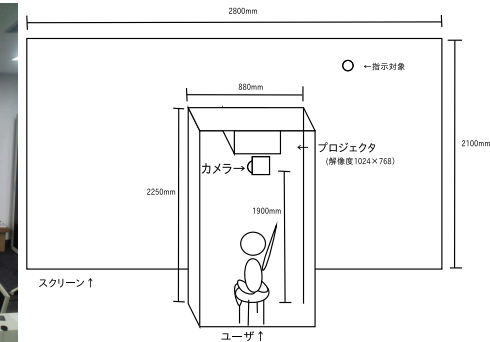


図 5 試作システムの構成

4.2.2 プロジェクタ

指差し位置を座標として得るために、スクリーンに投影した画像に対する指差しを行う。指差し動作の回帰モデル化にあたり、スクリーンを指す動きが大きくなるほど有効であると考えられる。本システムは公共の場での利用を考えているため、実環境を想定し大画面で出力できることが望まれる。試作システムにはプロジェクタ MP522ST (BenQ) を用いた。

4.2.3 計算機

試作システムに用いた計算機のスペック詳細を表 3 に示す。

以上のデバイスを用いて、試作システムを作成した。図 4、図 5 はその概観と構成である。カメラの位置は、ユーザ正面からの撮影では顔表面や頭髮のノイズ除去が難しく、背面からの撮影は画面がチラつき、背景差分が不可能だった。このため、指示動作として肘から指先までが写るようにユーザの頭上にカメラを配置した。

5. 実験

カメラとユーザの位置関係は、ユーザの体格等によって異なるため、各被験者の教師データを取得し実験を行った。今回の実験は、情報科学を専攻する 6 人の大学生および大学院生を対象に行った。



図 6 スクリーンに投影する画像 図 7 図 6 に対する教師画像 図 8 補間データを撮影する際の指差し位置の例

まず、教師データの取得について述べる。図 6 のようにスクリーンを縦横 10 分割する直線と画面縁との各交点、直線同士の各交点に円を表示し、その円に対する指差し画像を撮影する。次に、背景差分、ノイズ除去により指差し動作部分のシルエットを抜きだし、教師画像（図 7）とする。1 つの交点に対して 1 枚の教師画像を得る。また、撮影の際、手は膝上などに戻さずスクリーン左上から連続して順次指差しを行うものとする。得られた教師データを用い、式 (1) の A および、式 (9) の B を得る。

得られた A, B に対し、教師画像を入力とした場合の推定座標と正解座標 100 組による平均推定誤差と主成分によって張られる固有空間の次元数との関係を図 9 に示す。平均推定誤差は推定座標と正解座標のピクセル間距離である。固有空間の次元数を上げていくと、教師画像に対する推定誤差は少なくなっていく、99 次元で最小となることが確認された。また、ある被験者（以下被験者 A）のデータで固有空間を 20 次元とした場合の正解座標の円と推定座標を結んだ画像を図 10、固有空間を 99 次元とした場合の正解座標の円と推定座標を結んだ画像を図 11 に示す。なお、カメラの解像度が 640×480 画素の場合、毎秒 10 フレーム以上で推定可能であり、操作性については問題ないと思われる。

5.1 指差し精度の評価

5.1.1 教師画像と同じ位置を指差した場合の指差し位置推定

本実験では、教師データと同じ位置を 1 回ずつ指差した画像（テストデータ）、教師画像とは異なる位置を指差した場合の画像（中間データ）を用意し、その推定精度を検証した。図 12 はテストデータを入力とした場合の全被験者の平均推定誤差と固有空間の次元数との関係を示したものである。20 次元以降は推定誤差に大きな変化は見られなかった。また、被験者 A のデータで固有空間 20 次元における誤差（図 13）、99 次元における誤差（図 14）を比べたところ、推定精度にほとんど差は見られなかった。しかし、20 次元、99 次元どちらにおいても図 10 に見られるものと同様の誤差が確認された。

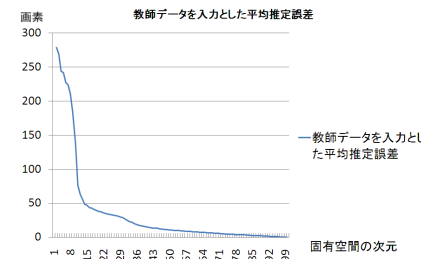


図 9 教師データを入力した場合の平均推定誤差と固有空間の次元数の関係

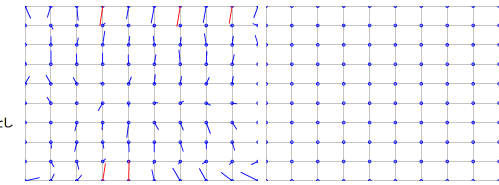


図 10 教師データを入力として 20 次元の固有空間を用いた指差し位置推定結果

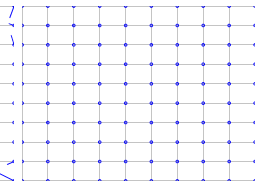


図 11 教師データを入力として 99 次元の固有空間を用いた指差し位置推定結果

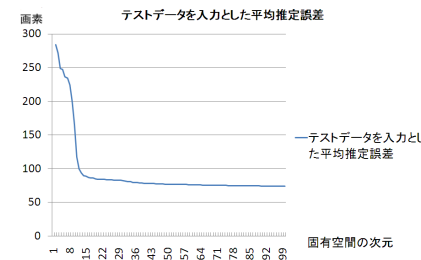


図 12 テストデータ：平均推定誤差と固有空間の次元数の関係

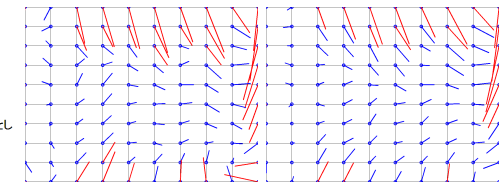


図 13 テストデータ：固有空間を 20 次元とした場合の指差し位置推定結果

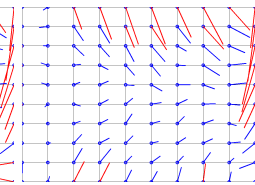


図 14 テストデータ：固有空間を 99 次元とした場合の指差し位置推定結果

5.1.2 教師画像とは異なる位置を指差した場合の指差し位置推定

この実験では、図 8 に示すように、教師データで指差した位置の中間地点として、各格子の中心 81 箇所に対する指示画像を元に推定を行った。全被験者の平均推定誤差と主成分の次元数の関係（図 15）は、テストデータと同様の結果となった。また、被験者 A のデータで固有空間 20 次元（図 16）、99 次元における誤差（図 17）を比べたところテストデータと同様の結果が得られた。この結果はどの被験者についても同様であった。

5.2 他のポインティングデバイスとの比較

既存のインタフェースとの比較として、マウス（ロジクール、MX-A）、ジャイロマウス（ロジクール、MX-A）、提案システムによるポインティングタスクの実行時間を計測し、比較評価を行った。実験内容を以下に示す。

- 中間データを取得した際の円の中心と同様の位置をスタート地点としてランダムに選択

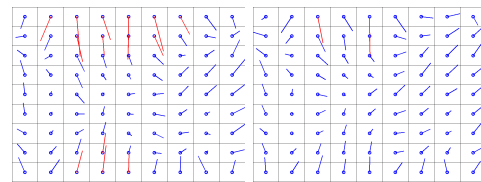
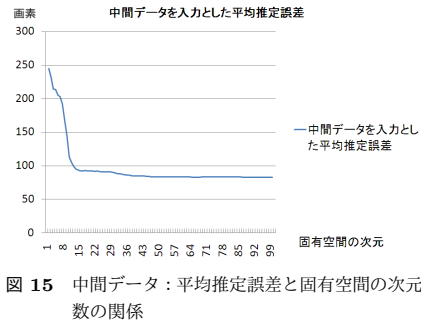


図 16 中間データ：固有空間を 20 次元とした場合の指差し位置推定結果
図 17 中間データ：固有空間を 99 次元とした場合の指差し位置推定結果

ゴール地点のエリアサイズ (半径)	画素
サイズ 1	25
サイズ 2	50
サイズ 3	75

表 4 ゴール地点のエリアサイズ (単位：画素)

する。

- 選択された地点から一定の距離にゴール地点を設定する。
- スタート地点からゴール地点に移動する時間を計測する。

スタート、ゴールの判定は各々の地点から一定の範囲内に 0.4 秒留まることで判定した。スタート地点とゴール地点との距離は、151, 251, 416, 691 1151 画素の 5 段階。スタート地点の大きさを半径 25 画素、ゴールの判定となるエリアの大きさを半径 25, 50, 75 画素の 3 段階とした。距離を 5 段階に分けたが、最長距離はスクリーンの対角線上に位置する円の中心を結んだ 1151 画素であり、次に長い距離をスクリーンの縦方向の端に位置する円の中心同士を結んだ直線距離 691 画素とした。2つの距離の比率に基づき残りの 3つの距離を決定した。以上の操作を各距離、各エリアサイズにつき 4 回行い、各被験者に対して 5×3×4 の 60 回を距離、エリアサイズをランダムに選び実行した。

実験結果を図 18 に示す。実験において、多くの被験者が提案手法ではスクリーンの 4 隅を指すことが出来ず、対角線を結ぶ最長距離に対して十分なデータが取れなかったため、今回は 151, 251, 416, 691 画素の 4 段階の距離を評価対象とした。

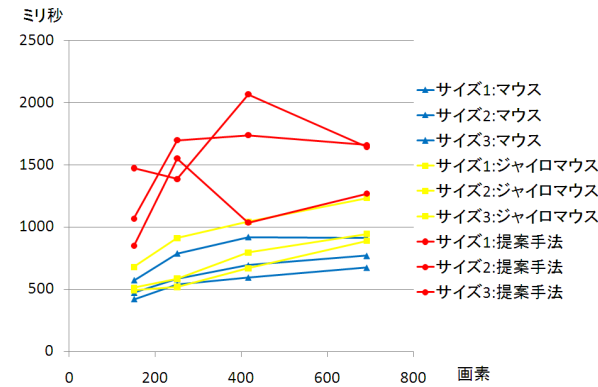


図 18 各距離、各エリアサイズのタスク処理の平均時間

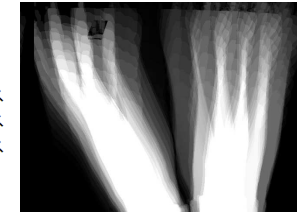


図 19 試作システムによる固有画像の例 (第一主成分)

5.3 考察

5.3.1 推定誤差に関する考察

5.1.2 節の実験の結果、図 13, 図 16 のように、スクリーン端の領域において特に大きい誤差が確認された。また、誤差はスクリーンの端から中心に向かう傾向が見られた。これは回帰モデルを利用したため、スクリーン中心は学習データが十分だが、スクリーンの端に向かうに従い学習データが不足するためと考えられる。

実験中、試作システムによるカーソルの動作を確認したところ、どの被験者においても横方向にカーソルを動かす場合にカーソルが波打つように動いていた。また、縦方向についてはこのような動作は確認されなかった。これは固有画像 (図 19) と重ならない位置に腕がある場合、上下の位置の学習画像の影響を受けていると思われる。

5.3.2 固有空間の次元数に関する考察

教師データを入力とし、固有空間を 20 次元とした場合の図 10 に見られる誤差の発生する位置、誤差の向きが 5.1.1 節の実験、5.1.2 節の実験の結果と一致していることがそれぞれ確認される。誤差のあまり発生していないスクリーン中心付近の推定位置は固有空間を 20 次元とした場合、99 次元とした場合で大きな変化は見られなかった。このことから、教師データが正しく取得できていれば、固有空間は 20 次元で十分な推定結果が得られると思われる。

5.3.3 既存インタフェースとの比較

図 18 から、マウス、ジャイロマウスはエリアサイズが大きくなるにつれ、タスクの処理時間が短くなり、スタートとゴールの距離が離れるにつれ、タスクを処理する時間が長くなっている。これは、ポインティングタスクにおける、ターゲットまでポインタを動かす時間とターゲットオブジェクトのサイズ、ターゲットまでの距離の関係を示した、フィッツの法則¹⁴⁾に従っている。一方、提案手法ではエリアサイズに対してはマウス、ジャイロマウスと同様の変化を見せたが、スタートとゴールの距離とタスクの処理時間の関係はマウス、ジャイロマウスほどではないことが確認された。これはマウス、ジャイロマウスの操作はカーソルを目で確認して動作にフィードバックしながら動かすが、提案手法においてはフィードバックの必要がないためと思われる。

6. おわりに

本研究では、大画面で利用可能な人間中心のポインティングインタフェースを実現するために、単眼カメラと主成分回帰モデルによる指差し位置推定手法を提案した。そして、指差し位置の推定精度の評価、提案手法と既存インタフェースとの比較を行った。その結果、回帰モデルを用いたことによるスクリーン端の大きな推定誤差が認められたものの、スクリーン中心については概ね安定して推定結果が得られることを確認した。また、既存インタフェースはフィッツの法則に従ったが、提案手法はフィッツの法則には従わなかった。今後の課題として、学習領域の検討を行いカーソル動作の改善や、推定誤差の大きくなる領域を無くすことが挙げられる。これによって、より扱いやすく直感的なポインティングが可能になると考えられる。

謝辞 本研究の一部は、文部科学省グローバル COE プログラム（研究拠点形成費）の補助による。

参 考 文 献

- 1) CyberWorld, Inc.: <http://www.cwonline.com/> ((最終アクセス日時:2009 年 5 月 8 日))。
- 2) 福地健太郎:レーザーポインタの軌跡を用いた映像パフォーマンスの試み, インタラクシオン 2005 論文集 (情報処理学会シンポジウムシリーズ), No.4, pp.63-64 (2005)。
- 3) Soechting, J.F. and Flanders, M.: Sensorimotor representations for pointing to targets in three-dimensional space, *Journal of Neurophysiology*, Vol.62, No.2, pp.582-594 (1989)。

- 4) 吉田千里, 乾敏郎:ポインティング動作に基づく空間認知特性の検討, 映像情報メディア学会技術報告, Vol.24, No.51, pp.9-16 (2000)。
- 5) Brain, W.R.: Visual disorientation with spatial reference to lesions of the right cerebral hemisphere, *Brain*, Vol.64, pp.244-272 (1941)。
- 6) Vogel, D. and Balakrishnan, R.: Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays, *Proceedings of UIST 2005 - the ACM Symposium on User Interface Software and Technology.*, pp.33-42. (2005)。
- 7) 中村卓, 高橋伸, 田中二郎:ハンドジェスチャを用いた公共大画面向けインタフェース, *DICOMO2006* (2006)。
- 8) 任天堂株式会社:<http://www.nintendo.co.jp/> ((最終アクセス日時:2009 年 5 月 8 日))。
- 9) 佐藤洋一, 小池英樹:Enhanced Desk: 拡張机型インタフェースによる紙情報と電子情報の統合, 情報処理学会研究報告. CVIM,[コンピュータビジョンとイメージメディア], Vol.2000, No.50, pp.73-77 (2000)。
- 10) 比嘉広樹, 堂上高司:手指障害者のための非接触型入力インタフェースの開発, 生体医工学, Vol.46, No.1, pp.31-36 (2008)。
- 11) 岩井儀雄, 八木康史, 谷内田正彦:単眼動画からの手の 3 次元運動と位置の推定, 電子情報通信学会論文誌 D, Vol.J80-D-2, No.1, pp.44-55 (1997)。
- 12) 久米祐一郎, 島田正文:腕装着型カメラと反射指標を用いた装用型入力デバイスの検討, 映像情報メディア学会誌, Vol.60, No.2, pp.249-253 (2006)。
- 13) 佐藤信, 坂根茂幸:実空間にマークを投影するインタラクティブハンドポインタの構成, 電気学会論文誌 C 電子・情報・システム部門誌, Vol.121-C, No.9, pp.1464-1470 (2001)。
- 14) Card, S.K., Moran, T.P. and Newell, A.: *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Assoc Inc. (1983)。