

拡張機型インタフェースのための複数指先の追跡とその応用

岡 兼 司[†] 陳 欣 蕾^{††}, 中 西 泰 人^{††}
佐 藤 洋 一[†] 小 池 英 樹^{††}

本論文では、拡張機型インタフェース上で複数の指先軌跡を安定に計測するための手法を提案するとともに、計測された指先軌跡を用いて様々なインタラクションを行うための手法について説明する。まず、赤外線カメラを用いて検出された複数の指先位置について、隣り合う画像フレーム間における指先どうしの対応を考慮することによって、複数の指先軌跡を実時間で安定に計測する。さらに、手の形状に基づいて、ダイレクトマニピュレーションとシンボリックジェスチャという2種類のインタラクションモードを連携して利用するための機構を実現する。本手法に対していくつかの評価実験を行ったところ、良好な結果を得ることができた。

Real-time Tracking of Multiple Fingertips and Its Applications for Augmented Desk Interface Systems

KENJI OKA,[†] XINLEI CHEN,^{††} YASUTO NAKANISHI,^{††} YOICHI SATO[†]
and HIDEKI KOIKE^{††}

In this paper, we propose a fast and robust method for tracking multiple fingertips; we then demonstrate gesture recognition and an application based on measured fingertip trajectories for augmented desk interface systems. Our tracking method can track multiple fingertips reliably even in a complex background under a dynamically changing lighting condition without any markers. After multiple fingertips are detected in each input infrared image frame, correspondences of those fingertips between successive image frames are determined based on a prediction technique. Our gesture recognition system is particularly advantageous for Human-Computer Interaction (HCI) in that users can achieve interactions based on both symbolic gestures and direct manipulation with their own hands.

1. はじめに

本などの実物体とそれに関連する電子情報を統合的に扱うことを可能とする作業環境の実現に向けて、近年、拡張機型インタフェースについての研究が進められている^{1),4),6),10),14),15),18)}。そこで我々は、図1に示すような拡張機型インタフェース EnhancedDesk を開発してきた⁵⁾。EnhancedDesk では、先駆的な研究である DigitalDesk¹⁸⁾ と同様に、机上に投影されたアプリケーションをユーザ自身の指先で操作することが可能である。さらに、机上に投影された仮想物体と実物体の両方を統合的に操作することが可能な環境の

実現を目指している。

この拡張機型インタフェースでは、ユーザの手指の追跡技術が重要な役割を果たす。従来の手指追跡手法では、手領域を抽出するために肌色抽出や背景差分などの手法が多く用いられてきたが⁸⁾、拡張機型インタフェース上では、プロジェクタによる投影の影響によりユーザの手の色や背景が動的に変化するため、これらの従来手法では安定に手指を追跡することが困難であった。このような問題に対して、我々の先行研究において、中赤外領域光を撮影する赤外線カメラを利用することにより、動的に光源環境が変化する場合や背景が複雑な場合にも、マーカなどのデバイスを用いることなく、複数の指先位置を計測するための手法を開発してきた¹¹⁾。また、この計測結果を利用することにより、机上に投影されたアプリケーションをユーザ自身の手によって直接的に操作すること、すなわち、ダイレクトマニピュレーションに基づくインタラクションを可能としてきた。

[†] 東京大学生産技術研究所

Institute of Industrial Science, The University of Tokyo

^{††} 電気通信大学院情報システム学研究科

Graduate School of Information Systems, University of Electro-Communications

現在、NTT アドバンステクノロジー株式会社

Presently with NTT Advanced Technology Corporation

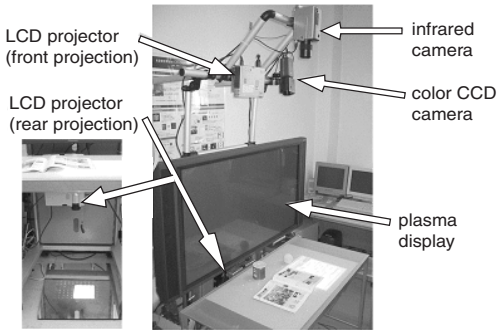


図 1 EnhancedDesk の概観
Fig. 1 Overview of EnhancedDesk.

しかしながら、この先行研究にはいくつかの問題点が残されていた。最初の問題点として、先行研究の指先位置計測手法では個々の入力画像フレームで独立に指先位置を特定していただいただけであり、個々の指先についての追跡を行っていなかったことがあげられる。このために、ユーザの時系列の動作に関する情報が欠如していることが問題となってきた。また、過去の計測結果を考慮せずに個々の画像フレームで独立に指先位置計測を行うことは、指先位置計測の安定性の欠如にも大きく影響してきた。

さらに先行研究では、認識可能なインタラクションの種類が大幅に制限されてしまうという問題も存在した。一般に、手指の追跡結果をインタラクションに利用する際には、大きく分けて 2 種類のインタラクションモードが存在する。その 1 つがシンボリックジェスチャである^{7),13),19)}。ここでは、ある一定時間における一連の動作により定義される操作をシンボリックジェスチャとして考える。一方で、もう 1 つのインタラクションモードがダイレクトマニピュレーションであり、個々の入力画像フレームで計測された情報を直接的にインタラクションに利用するモードとして定義される^{3),12),16)}。先行研究では、個々の入力画像フレームから得られる情報をもとにダイレクトマニピュレーションによるインタラクションを可能としていた一方で、一連の手指動作に関する情報を獲得できなかったためにシンボリックジェスチャを認識することが困難であった。手の動作に基づく多様なインタラクションを実現するためには、シンボリックジェスチャも含めた 2 種類のインタラクションモードを連携して利用することが不可欠である。

そこで本論文では、拡張机型インタフェース上で複数の指先動作を安定に計測するための手法を提案する。さらに、指先動作の追跡結果を利用して、2 種類のインタラクションモードを連携して利用するための機構

を提案する。

まず、指先動作の計測のために、個々の入力画像フレーム中で検出された指先位置に対して予測フィルタを適用することにより、次の画像フレームにおける指先位置を予測する。次いで、これらの予測位置と次フレームで実際に検出された指先位置との最適な組合せを決定することにより、隣り合う画像フレーム間で指先どうしの対応付けを行う。本手法により、指の曲げ伸ばしなどにより指先の検出数が変化する場合や指先の検出に失敗した場合にも、複数の指先動作を実時間で安定に計測することが可能となる。

また、ダイレクトマニピュレーションとシンボリックジェスチャという 2 種類のモードの連携を可能とするために、計測された指先動作を利用してこれら 2 種類のモードを判別し、それぞれのモードに適した処理を適用できるような機構を実現する。本論文では、机上での細かなマニピュレーション作業では親指と人差し指が重要な役割を果たすという観察に基づき、ユーザの親指に着目したモードの判別を行うこととする。その後、判別されたダイレクトマニピュレーションに対しては指先位置を直接的に利用した処理を適用するのに対し、シンボリックジェスチャに対しては Hidden Markov Model (HMM)⁹⁾ に基づく認識処理を行い、その結果をインタラクションに利用する。

さらに、これらの手法を利用した拡張机型インタフェースのアプリケーションを紹介する。

以下、本論文は次のような構成となっている。2 章において複数の指先動作を安定に計測するための手法を説明し、3 章で指先動作に基づく多様なインタラクションを実現するための機構について述べる。そして 4 章で本手法を利用して開発されたアプリケーションを紹介し、最後に 5 章で結論を述べる。

2. 複数の指先動作の実時間計測

2.1 個々の画像フレームにおける指先、手のひら中心および指の付け根の検出

本章では、我々の先行研究において開発された個々の入力画像フレームから実時間で複数の指先を検出するための手法について簡単に説明する¹¹⁾。

従来手法で利用されてきた肌色抽出や背景差分に基づく手領域の抽出手法には光源環境や背景の状況に影響されやすいといった問題がある。そこで本手法では、このような問題点を回避するために、中赤外領域光を撮影する赤外線カメラによってユーザの手の熱放射を計測する。これにより、得られる赤外線カメラ画像中の手に対応する領域では輝度が非常に高くなる。その

ため、この画像に対して適当な閾値の二値化処理を適用することにより、背景が複雑な場合や光源環境の変化が起こる環境下においても、ユーザの手に対応する領域を画像中で安定に抽出することが可能となる。

手領域を抽出した後、複数の指先検出を高速に行うために、手のひらと指全体を含むような一定の大きさの探索領域を設定する。ユーザが机上で作業を行う場合、カメラと手の間の距離は比較的一定であるため、同一ユーザの場合には探索領域の大きさが一定であっても本システムは安定に稼働する。

次いで、決定された探索領域内において指先を検出する作業を行う。人間の指全体の形状は半球状のキャップを持つ円柱として近似されるため、本手法では適当な大きさの円形テンプレートを用いて正規化相関に基づくパターンマッチングを行い、複数の指先候補点を検出する。このようにして得られた指先候補点について、多重マッチングや指関節周辺での誤検出を除去することにより、複数の指先位置を特定する。

その後、手のひらの中心位置を計測するための処理を行う。本手法においては、手のひらの中心を手領域の境界からの距離が最大となる点として定義することにより、手の開閉などの形状変化に対しても安定に手のひらの中心位置を決定することを可能としている。このような位置を得るために、抽出された手領域に対して、領域があらかじめ定義された面積より小さくなるまで、収縮処理を繰り返し適用し、その後、残った領域の重心を計測するという処理を行う。

以上の処理によってユーザの指先および手のひらの中心を検出することが可能となるが、3.1節における親指判別を行うためには指の付け根の位置を検出することが必要となる。まず、抽出された手領域に対して、指領域が消失するために十分な回数の収縮処理を適用することにより、指領域を含まない手のひらの領域を抽出する。この収縮処理の回数としては、指先検出で用いた円形テンプレートの大きさ、すなわち指の太さに比例する回数を設定する。次に、抽出された手のひら領域と指を含む手領域との差分をとることにより指領域を抽出する。さらに、得られた指領域に対して、先に検出された指先位置を考慮することにより、指先に対応する端点とは反対側の端点を特定する処理を行う。この処理は抽出された指領域が長方形の領域となることを利用しており、たとえば指領域に対して指先に対応する点が左上にある場合、最も右下にある画素を検出することにより実現される。このようにして抽出された点を指の付け根の位置として決定する。

なお、以上で述べたような二値化処理の閾値や探索

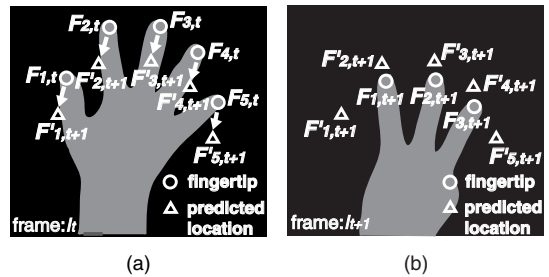


図2 指先対応付け手法

Fig. 2 Method for taking correspondences of fingertips.

領域の大きさ、円形テンプレートの大きさはユーザによってある程度異なる。そのため、これらのパラメータは初期化時に取り込んだ赤外線カメラ画像に基づいてユーザごとに設定される。これにより、ユーザに依存しない安定したシステムの稼働が可能となる。

2.2 指先動作の計測

本計測手法では、2.1節の手法を用いて検出された複数の指先について、それらを隣り合う画像フレーム間で対応付けることにより、背景や光源環境の動的な変化が起こる場合にも、複数の指先の動作を実時間で安定に計測することを可能とする。以下、個々の入力画像フレームで検出された指先どうしを画像フレーム間で対応付けるための手法について説明する。

2.2.1 指先動作計測のための指先対応付け手法の概要

複数の指先動作の計測を目指し、隣り合う画像フレーム間で指先どうしの対応付けを行うための手法について、その概要を説明する。ここでは、 t 番目の画像フレーム I_t で n_t 個の指先が検出された場合を想定し、これらの指先位置を $F_{i,t}$ ($i = 1, 2, \dots, n_t$) とする(図2(a))。まず、次の画像フレーム I_{t+1} での指先位置 $F'_{i,t+1}$ を予測フィルタを用いて予測する。次いで、画像フレーム I_{t+1} で実際に検出された n_{t+1} 個の指先位置 $F_{j,t+1}$ ($j = 1, 2, \dots, n_{t+1}$) と予測位置 $F'_{i,t+1}$ との比較を行う(図2(b))。これらの検出位置と予測位置との間で最適な組合せを求めることにより、隣り合う画像フレーム間での指先どうしの対応付けが実現され、図3のように複数の指先の軌跡を実時間で計測することが可能となる。

2.2.2 指先位置の予測

本項では、ある画像フレームで検出された指先位置を基に、予測フィルタを用いて次フレームにおける指先位置を予測する手法について説明する。本手法では、予測フィルタとして特徴点追跡などにしばしば利用されるカルマンフィルタ¹⁷⁾を適用する。なお、本項で

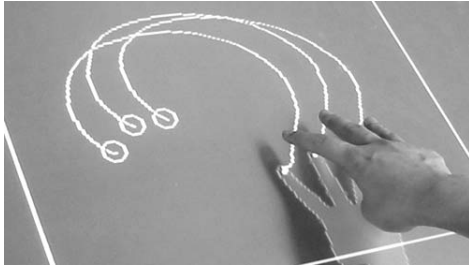


図3 指先軌跡の計測

Fig. 3 Measurement of fingertip trajectories.

説明する処理を各指先に対して独立に適用する。

まず、本手法では各画像フレームに対して各指先の位置と速度についての推定を行うため、状態ベクトル x_t を次のように定義する。

$$x_t = (x(t), y(t), v_x(t), v_y(t))^T. \quad (1)$$

ここで、 $x(t)$ 、 $y(t)$ 、 $v_x(t)$ 、 $v_y(t)$ は、 t 番目の画像フレームにおける指先位置 $(x(t), y(t))$ 、および指先の速度 $(v_x(t), v_y(t))$ を表す。

次に、観測ベクトル y_t を t 番目のフレームで検出された指先位置として定義する。このとき、状態ベクトル x_t と観測ベクトル y_t を次の基本システム方程式により関連付ける。

$$x_{t+1} = Fx_t + Gw_t, \quad (2)$$

$$y_t = Hx_t + v_t. \quad (3)$$

この式における F は状態遷移行列、 G は駆動行列、 H は観測行列を表し、また、 w_t は状態ベクトル x_t の速度成分に加わるシステム雑音、 v_t は実際の位置と検出された位置との誤差に相当する観測雑音を示す。

さらに、本手法では画像フレーム間隔 ΔT が非常に短いため、連続する画像フレーム間での各指先の移動を等速直線運動として近似する。そこで、 F 、 G 、および H を次のように定義する。

$$F = \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

$$G = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T, \quad (5)$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \quad (6)$$

ここで観測行列 H は作業空間である机上で定義されるワールド座標系と画像座標系との変換を表現する行列であり、カメラモデルとして透視投影を考えた場合、一般には射影変換で記述される。一方で、本論文では

議論を簡潔にするために、一般性を失うことなく、状態ベクトル x_t の (x, y) 座標と観測ベクトル y_t の (x, y) 座標が一致するとしている。

また、システム雑音 w_t と観測雑音 v_t については、二次元零ベクトルを平均ベクトルとして持つ一定のガウス雑音として仮定する。それゆえ、 w_t および v_t の共分散行列はそれぞれ $\sigma_w^2 I_{2 \times 2}$ 、 $\sigma_v^2 I_{2 \times 2}$ となる。ここで、 $I_{2 \times 2}$ は 2×2 単位行列である。

以上の定義を用いて、カルマンフィルタを次式のように定式化する。

$$K_t = \tilde{P}_t H^T (I_{2 \times 2} + H \tilde{P}_t H^T)^{-1}, \quad (7)$$

$$\tilde{x}_{t+1} = F \{\tilde{x}_t + K_t (y_t - H \tilde{x}_t)\}, \quad (8)$$

$$\tilde{P}_{t+1} = F (\tilde{P}_t - K_t H \tilde{P}_t) F^T + \frac{\sigma_w^2}{\sigma_v^2} \Lambda. \quad (9)$$

ここで、 $\tilde{x}_t = \hat{x}_{t|t-1}$ であり、この $\hat{x}_{t|t-1}$ は y_0, \dots, y_{t-1} から推定される x_t の推定値を表す。また、 $\tilde{P}_t = \hat{\Sigma}_{t|t-1} / \sigma_v^2$ であり、 $\hat{\Sigma}_{t|t-1}$ は $\hat{x}_{t|t-1}$ の推定誤差の共分散行列を表す。さらに、 K_t はカルマンゲインであり、 $\Lambda = GG^T$ である。

以上の式を計算することにより、次の画像フレーム、すなわち $t+1$ 番目の画像フレームにおける予測位置を \tilde{x}_{t+1} の中の成分 $(x(t+1), y(t+1))$ として得る。

なお、2 フレーム以上後における指先位置、すなわち m ($m > 1$) フレーム後での予測が必要な場合には次式により計算を行う。

$$\hat{x}_{t+m|t} = F^m \{\tilde{x}_t + K_t (y_t - H \tilde{x}_t)\}, \quad (10)$$

$$\hat{P}_{t+m|t} = F^m (\tilde{P}_t - K_t H \tilde{P}_t) (F^T)^m + \frac{\sigma_w^2}{\sigma_v^2} \sum_{l=0}^{m-1} F^l \Lambda (F^T)^l. \quad (11)$$

ここで、 $\hat{x}_{t+m|t}$ は y_0, \dots, y_t から推定される x_{t+m} の推定値である。また、 $\hat{P}_{t+m|t} = \hat{\Sigma}_{t+m|t} / \sigma_v^2$ であり、 $\hat{\Sigma}_{t+m|t}$ は $\hat{x}_{t+m|t}$ の推定誤差の共分散行列を表す。

2.2.3 予測位置と実際の検出位置との対応付け

本項では、実際に検出された指先位置と式 (8) や式 (10) によって予測された指先位置とを対応付けるための手法について説明する。

実際には、検出された指先位置と予測位置との間の複数の組合せについて距離の二乗和を計算し、その和が最小となる組合せを最適な組合せとして決定する。

ここで計算コストの削減を図るために、手のひらの中心のまわりにおける指先の順番を考慮することによ

これはごく簡単な近似であり、本来であれば、正規化相関のマッチングスコアなどに基づいて各画像フレームごとに推定されるべきであるが、これに関しては今後の課題とする。

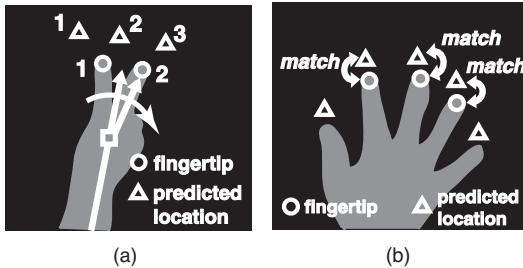


図 4 指先の検出位置と予測位置との対応付け

Fig. 4 Correspondences of detected fingertips and predicted fingertips.

り、計算すべき組合せの数を減らす(図 4(a)). すなわち、指の交差などによる入力画像中の指先の順番変化は起こらないと仮定し、図 4(a) の場合では、3 通りの組合せ (i) $\bigcirc 1-\Delta 1$ & $\bigcirc 2-\Delta 2$, (ii) $\bigcirc 1-\Delta 1$ & $\bigcirc 2-\Delta 3$, (iii) $\bigcirc 1-\Delta 2$ & $\bigcirc 2-\Delta 3$) についてのみ計算を行うこととする. これにより、考慮すべき組合せの数は最大で ${}_5P_5$ 通りから ${}_5C_5$ 通りに削減される.

次に、入力画像中でいくつかの指先を検出できなかった場合の対応について説明する. このような状況の一例を図 4(b) に示す. この図では、手の高速移動時における手領域のぶれなどの影響により、親指と小指についての検出に失敗している. 本手法では、検出に失敗した指先の予測位置に関する情報を次のようにして利用することにより、指先追跡を続行するかどうかについて判断する.

ある予測位置に対応する指先位置が見つからなかった場合、その予測位置に対応する式 (9) 中の共分散行列 \hat{P}_{t+1} の (1, 1) 成分を調べる. この成分は指先位置予測についての不確定性を表していることから、この成分がある閾値よりも小さい場合にはこの予測位置を信頼し、この予測位置に対応する指先は何らかの原因により偶然検出されなかったにすぎないと見なす. そこでこの場合には、予測位置を実際の指先位置と見なして指先追跡を続行することとする. 一方で、この (1, 1) 成分が閾値よりも大きい場合には、予測された位置は信用できないと判断し、その指先に関する追跡を終了する. 現在の実装では、不確定性に関する閾値として経験的に定められた一定の値を使用している.

また、予測位置の数よりも検出された指先位置の数の方が多い場合、どの予測位置にも対応付けられなかった指先位置について追跡を開始する. このとき、その指先位置から計算される予測位置の不確定性が閾値よりも小さくなるまでは新たな指先の軌跡として見なさないこととする.

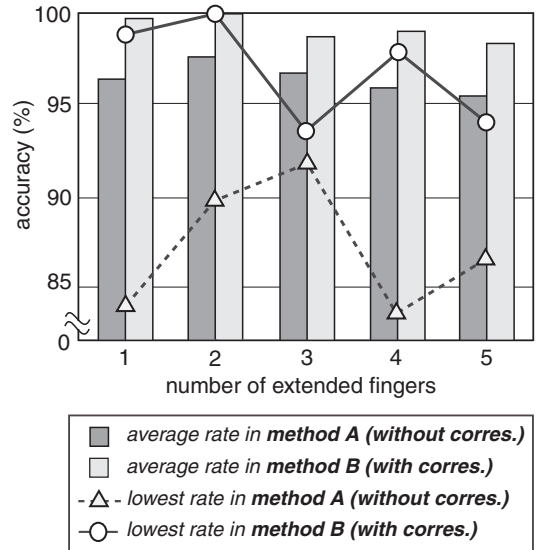


図 5 指先追跡手法の評価実験

Fig. 5 Evaluation of finger tracking.

2.3 追跡手法の評価実験

前節で紹介した複数指先の追跡手法の信頼性について評価実験を行った. 特に、隣り合うフレーム間での指先どうしの対応を考慮することにより、指先追跡の信頼性がどの程度向上するかについて比較実験を行った.

本実験には 7 名の被験者が参加した. 実験環境としては、Intel Pentium III 500 MHz を搭載した Linux PC を使用し、画像処理ボードには日立 IP5005 を、赤外線カメラにはニコン LAIRD-S270 を使用した.

各試行において、被験者は図 4(a) のように伸ばした指の本数を一定に保ったまま机上で自由に手を動かした. 最初に、被験者は 1 本だけ指を伸ばして 30 秒間手を動かした. 次いで、伸ばした指の本数を 2 本、3 本、4 本、5 本として同様に動かした. 各試行はそれぞれ 30 秒間行ったため、各試行につき約 900 フレームのデータが得られた.

本実験では、隣り合う画像フレーム間で検出された指先どうしの対応関係を考慮する場合としない場合の指先追跡性能について比較実験を行った. このとき、公平な比較を行うために、カメラ出力をいったんビデオに録画し、そのビデオに対して各追跡手法を適用した.

実験結果を図 5 に示す. 手法 A は指先どうしの対応を考慮しない手法を表し、手法 B は対応関係を考慮する手法を表している. また、図中の棒グラフは追跡された指先の数についての正解率における被験者 7 名の平均値を示し、折れ線グラフは正解率についての

被験者 7 名中の最低値を示す。

図 5 を見ると、フレーム間で指先どうしの対応を考慮することによって、指先追跡性能が大きく向上していることが分かる。さらに、折れ線グラフにあるように正解率の最低値についても大幅な向上が見られる。このことから、本追跡手法はユーザにかかわらず非常に信頼性が高く、実用的なアプリケーションへの応用に適していることが分かる。

3. 指先動作に基づくインタラクション

本章では、計測された指先動作に基づいて、ダイレクトマニピュレーションとシンボリックジェスチャを連携して利用するための機構について述べる。

本研究で提案する機構は拡張机型インタフェースに

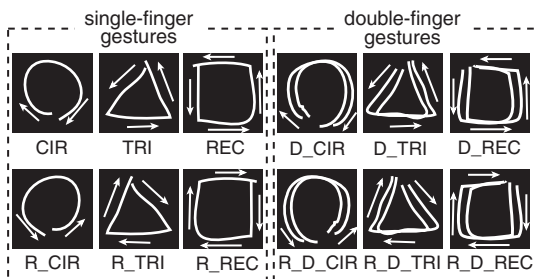


図 6 シンボリックジェスチャの例
Fig. 6 Examples of symbolic gestures.

おける様々なアプリケーションでの利用が可能であることを前提としているが、ここでは具体例として、4章で示すような描画システムで利用されるダイレクトマニピュレーションとシンボリックジェスチャ、さらにはそれらを組み合わせるための機構について述べる。

まず、ダイレクトマニピュレーションとしては、物体をつまんで移動させるような動作を行うことによって、仮想物体をマニピュレーションするインタラクションが中心となる。このとき、個々の入力画像フレームにおいて計測された指先位置を利用して、マニピュレーションのモードや移動先の位置などの決定を行う。一方、シンボリックジェスチャとしては、人差し指 1 本で図 6 に示すような軌跡を描くという一連の動作を行うと、その軌跡の意味を認識器により認識したうえで、描いた位置に認識結果を表示するようなインタラクションを行う。また、2 本指で同様に軌跡を描くと選択範囲を指定できるような操作なども考えられる。

このような手指による様々なインタラクションを実現する機構として、本研究では図 7 に示す機構を利用する。まず、指先動作の情報から手の動作がダイレクトマニピュレーションであるか、シンボリックジェスチャであるかを判別する。次いで、ダイレクトマニピュレーションとして判別された動作については、指先どうしの間隔や伸ばした指の本数に基づいて「回転」

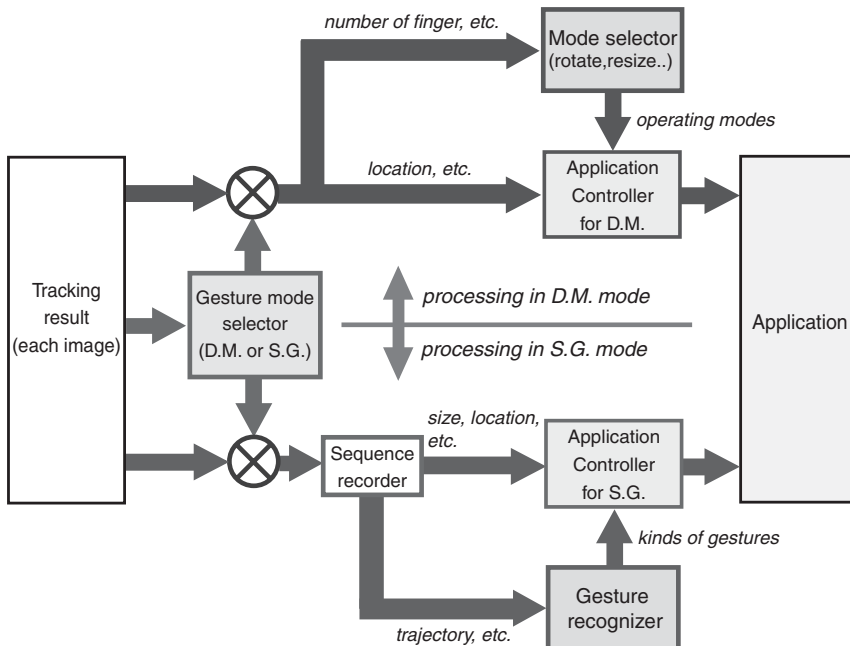


図 7 手指によるインタラクションのための処理の流れ
Fig. 7 Interaction based on both direct manipulation and symbolic gestures.

や「移動」などの操作モードを選択してから、指先の位置情報をもとに操作モードの变量を制御する。一方、シンボリックジェスチャとして判別された動作については、HMMに基づいて構築された認識器に追跡結果のシーケンスを入力してジェスチャの種類を認識し、さらに軌跡情報からジェスチャの位置や大きさなどの情報を付加するといった処理を行う。

このとき、ダイレクトマニピュレーションとシンボリックジェスチャとを判別については、机上での細かなダイレクトマニピュレーションでは人差し指に加えて親指が非常に多く用いられるという観察に基づいて、親指の動作に注目してこれら 2 種類のモードを切り替えることとする。すなわち、ユーザが親指を折り曲げることにより親指が追跡されていないときのジェスチャをシンボリックジェスチャとして、親指を伸ばして親指が追跡されているときのジェスチャをダイレクトマニピュレーションとしてインタラクションモードの切替を行う。なお、本研究における指の判別（ある追跡中の指が親指であるか人差し指であるか、など）においては、五指をすべて判別するのではなく、親指かそれ以外の指かという判別を行うこととする。これは、親指かそれ以外の指かの判別だけでもインタラクションモードの切替に関して大きな支障はないという判断に加えて、親指はその他の指と比較して手の構造上判別が容易であるという理由による。

以下、この機構を実現するために必要な要素技術として、3.1 節では追跡中の複数の指先の中から親指を判別するための手法について、3.2 節ではシンボリックジェスチャの種類認識手法について説明する。

3.1 親指の判別

本節では親指とそれ以外の指先を区別するための手法について説明する。本手法では、図 8 に示すように、指の向き（手のひらの中心から指の付け根へと向かう向き）と腕の向き（腕の慣性主軸の向き）との間の角度 θ を求め、この角度 θ をもとに親指であるかどうかを判別する。ここで指の付け根を利用する理由は、指を動かしたときでも、指先と比較して指の付け根はほとんど動かないためである。

まず初期化の段階で、親指の標準角度 θ_T と人差し指の標準角度 θ_F ($\theta_T > \theta_F$) を求めておく。ここで標準角度とは、親指らしさや人差し指らしさを表すための基準となる角度である。これらは、初期化時に取り込んだ手の画像に対して 2.1 節で述べた手法を適用することにより親指と人差し指の付け根位置を検出した後で、これらの付け根に対する角度 θ を計算することによって求められる。

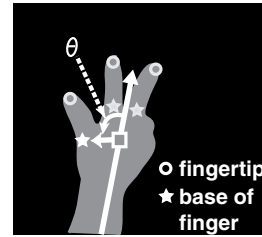


図 8 角度 θ の定義
Fig. 8 Definition of θ .

その後、個々の入力画像フレームにおいて、2.1 節で述べた手法により検出された指の付け根に対する角度 θ を計算する。ここで、ある指の軌跡の開始点から k 番目のフレームにおける角度 θ を θ_k とする。また、現在のフレームがその指の軌跡の開始点から N 番目のフレームにあたるとする。このとき、親指らしさを表すスコア s_T は次式により計算される。

$$s'_T(k) = \begin{cases} 1.0 & \text{if } \theta_k > \theta_T, \\ \frac{\theta_k - \theta_F}{\theta_T - \theta_F} & \text{if } \theta_F \leq \theta_k \leq \theta_T, \\ 0.0 & \text{if } \theta_k < \theta_F, \end{cases} \quad (12)$$

$$s_T = \frac{\sum_{k=1}^N s'_T(k)}{N}. \quad (13)$$

こうして得られた s_T が 0.5 以上のとき、その指を親指として見なす。

本研究の親指判別手法では、手領域を含めた腕の慣性主軸方向を考慮した角度を利用していることや親指の付け根と人差し指の付け根の位置が大きく離れていることに加えて、過去の履歴を考慮したスコアの計算を行っているため、不自然な手の姿勢を長時間持続させない限り、安定に親指の判別を行うことが可能である。

そこで、本節で提案する親指の判別性能についての評価実験を行った。実験環境は 2.3 節の実験と同様である。本実験では、机上の作業を想定した 3 種類のタスクについて実験を行った。まず、タスク (a) が人差し指だけを利用した描画作業、そして、タスク (b) が親指と人差し指によって物体をつまむ作業である。ここで、タスク (a) はシンボリックジェスチャによるインタラクションを想定した作業であり、一方、タスク (b) ではダイレクトマニピュレーションによる操作を想定している。また、タスク (a) に対する実験の補足として、親指だけを利用した描画作業であるタスク (c) に対する実験も行った。なお、本研究の指先追跡手法では指先どうしを完全に密着させると追跡が不安定になるため、タスク (b) におけるつまむ作業では完全には密着させない程度に指先どうしを近

表 1 親指判別の評価実験

Table 1 Evaluation of distinguishing a thumb.

タスク	(a)	(b)	(c)
平均値 (%)	98.3	99.4	98.2
標準偏差 (%)	4.6	0.8	3.6

表 2 シンボリックジェスチャ認識の評価実験

Table 2 Evaluation of symbolic gesture recognition.

シンボリックジェスチャの種類	single-finger	double-finger
平均値 (%)	99.2	97.5
標準偏差 (%)	0.5	1.8

づけることにより「つまむ」状態を表すこととする。各タスクにおける親指判別の正解率を表 1 に示す。この表によると、本実験ではいずれのタスクにおいても非常に高い正解率で親指の判別が行われた。このことから、本手法では信頼性の高い親指判別が実現されていることが分かる。

3.2 シンボリックジェスチャの認識

他のシンボリックジェスチャ認識^{7),13),19)}と同様に、本手法においても HMM を利用してシンボリックジェスチャの種類を認識を行う。認識器への入力については、検出された指先の本数と指先の動きの向きという 2 種類のデータ系列を使用する。これにより、指先の動作と伸ばした指の本数の両方に依存したシンボリックジェスチャ認識が可能となる。

ここで、指先の動きの向きについては次のように求める。まず、追跡中の各指先についての動きのベクトルを求め、それらの平均ベクトルを計算する。次いで、角度の一周を 16 分割したときに、この平均ベクトルの向きに対応する角度がどこに含まれるかに応じて、コード 1 からコード 16 までのコードを割り当てる。さらに、16 種類のコード以外に指先がほとんど静止した状態を表すコードを用意し、計 17 種類のコードを指先の向きを表す入力として使用する。このときに指先の動きの平均ベクトルを用いた理由は、親指以外の指でシンボリックジェスチャを行う場合、伸ばした指の指先が 1 つのシンボリックジェスチャの中で同様の動きをとることが多いためである。

そこで、本節で紹介したシンボリックジェスチャ認識の性能について評価実験を行った。認識するシンボリックジェスチャとして、図 6 に示すような指先軌跡による 12 種類のシンボリックジェスチャを用意した。

1 名によって行われたシンボリックジェスチャを各シンボリックジェスチャに対して 80 試行分用意し、それらを用いて HMM の学習を行った。認識実験には学習データを採取した 1 名を含まない 6 名が参加し、各シンボリックジェスチャに対して 20 回の試行

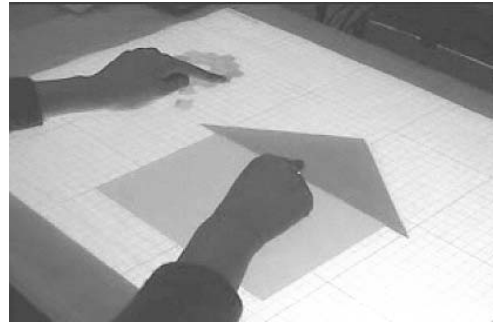


図 9 両手描画システム

Fig. 9 Two-handed drawing system.

を行った。このとき、シンボリックジェスチャの順番や大きさ、描く位置については各被験者の任意で決定した。得られた実験結果は表 2 のとおりである。表には 1 本指のシンボリックジェスチャ (single-finger symbolic gestures) と 2 本指のシンボリックジェスチャ (double-finger symbolic gestures) についての認識の正解率を示している。この結果から、1 本指のシンボリックジェスチャの認識率は 100% に近く、非常に頑健であることが分かる。また、2 本指のシンボリックジェスチャ認識についても高い認識率を示している。

4. 拡張機型インタフェースへの応用

本研究では、拡張機型インタフェース Enhanced-Desk 上で両手を有効に利用できるアプリケーションを開発している。その一例が図 9 の両手描画システムである²⁾。

この描画システムでは、本論文のシンボリックジェスチャ認識を利用して、ユーザの右手の人差し指で描いた図形と同じ位置で同じ形と大きさを持つ電子図形を表示することができる。たとえば、ユーザが指先で机の上に四角形に似た軌跡を描くと、システムはこれが四角形であると認識し、さらにユーザの指先の軌跡から四角形の位置と大きさを計算し、四角形が電子図形として机上に表示される。

また、ユーザが右手の親指と人差し指でつまむ動作を行った場合、親指と人差し指の間隔に基づいて「つまむ」と「はなす」のモードを判別し、電子図形のダイレクトマニピュレーションを行うことが可能である。

さらに、ユーザは自身の左手の位置に表示されているメニューの 1 つを指定してから、そのメニューの操作モードに従って右手で図形を描くことや操作対象の図形を指定するといった操作も可能となっている。このメニューはパイメニューと呼ばれる階層構造を持つ

扇形のメニューであり、左手の追跡結果に基づいてつねにユーザの左手の位置に表示される。このメニュー操作によって、コピーやペースト、色の変更といった操作を実現している。

以上の機能を組み合わせて描画作業を行うことにより、GUIでの描画ツールと比較して高速で直感的に描画することが可能となる。

5. おわりに

本論文では、拡張机型インタフェース上で複数の指先動作を安定に計測するための手法を提案した。さらに、様々なインタラクションを実現するために、ダイレクトマニピュレーションとシンボリックジェスチャという2種類のモードを連携して利用するための手法について説明した。まず、複数の指先動作を安定に計測するために、個々の入力画像フレーム中で検出された指先について、予測フィルタを利用して隣り合うフレーム間で指先どうしの対応付けを行う。次に、親指を検出することによってこれら2つのモードを明確に区別し、さらにシンボリックジェスチャに関してはHMMを利用した認識を行う。本手法に関して様々な評価実験を行ったところ、良好な結果が示された。また、本手法を利用したアプリケーションとして両手描画ツールを紹介した。

今後の課題の1つとして、手領域をより安定に抽出することがあげられる。現在の手法では赤外線カメラを利用することにより光源や背景の変化に対しては安定に手領域を抽出することが可能となっているが、その一方で、手の温度が非常に低いユーザの場合には手領域を安定に抽出できないという問題がある。このため、赤外線カメラ画像だけでなくカラーカメラ画像も併用することにより、より安定した手領域抽出を実現し、手指追跡の信頼性の向上を目指す。

また、三次元的に手指を追跡することも課題の1つであると考えられる。現在のところ、本追跡手法は机上での二次元的な追跡にしか利用できない。拡張机型インタフェースで利用する場合には二次元的な追跡で十分であるが、机以外での利用を行う場合には三次元的な追跡を利用したインタラクションが必要になることが考えられる。今後は、様々なアプリケーションへの実用に耐えうる三次元手指追跡手法について検討を重ねていく。

謝辞 本研究の一部は、文部科学省科学研究費補助金特定領域(C)(2)「人間の意図・行動理解に基づく柔軟なヒューマン・マシン・インタラクションの実現」(課題番号13224051)の助成により行われた。

参考文献

- 1) Arai, T., Machii, K. and Kuzunuki, S.: Retrieving electronic documents with real-world objects on InteractiveDesk, *Proc. ACM Symposium on User Interface Software and Technology (UIST '95)*, pp.37-38 (1995).
- 2) Chen, X., Koike, H., Nakanishi, Y., Oka, K. and Sato, Y.: Two-handed drawing on augmented desk system, *Proc. Working Conf. Advanced Visual Interfaces (AVI 2002)*, pp.219-222, ACM Press (2002).
- 3) Crowley, J., Berard, F. and Coutaz, J.: Finger tracking as an input device for augmented reality, *Proc. IEEE Int'l Workshop on Automatic Face and Gesture Recognition (FG '95)*, pp.195-200 (1995).
- 4) Fjeld, M., Ironmonger, N., Guttormsen-Schar, S. and Krueger, H.: Design and evaluation of four AR navigation tools using scene and viewpoint handling, *Proc. IFIP Int'l Conf. Human-Computer Interaction (INTERACT '01)*, pp.214-223 (2001).
- 5) Koike, H., Sato, Y., Kobayashi, Y., Tobita, H. and Kobayashi, M.: Interactive textbook and interactive Venn diagram: Natural and intuitive interface on augmented desk system, *Proc. ACM Conf. Human Factors in Computing Systems (SIGCHI 2000)*, pp.121-128 (2000).
- 6) MacKay, W.: Augmenting reality: Adding computational dimensions to paper, *Comm. ACM*, Vol.36, No.7, pp.96-97 (1993).
- 7) Martin, J. and Durand, J.: Automatic handwriting gestures recognition using hidden Markov models, *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG 2000)*, pp.403-409 (2000).
- 8) Pavlovic, V., Sharma, R. and Huang, T.: Visual interpretation of hand gestures for human-computer interaction: A review, *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, Vol.19, No.7, pp.677-695 (1997).
- 9) Rabiner, L. and Juang, B.: An introduction to hidden Markov models, *IEEE ASSP Magazine*, pp.4-16 (1986).
- 10) Rekimoto, J. and Saito, M.: Augmented surfaces: A spatially continuous work space for hybrid computing environments, *Proc. ACM Conf. Human Factors in Computing Systems (SIGCHI '99)*, pp.378-385 (1999).
- 11) Sato, Y., Kobayashi, Y. and Koike, H.: Fast tracking of hands and fingertips in infrared images for augmented desk interface, *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recog-*

inition (FG 2000), pp.462–467 (2000).

- 12) Segan, J. and Kumar, S.: Shadow gestures: 3D hand pose estimation using a single camera, *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '99)*, pp.479–485 (1999).
- 13) Starner, T. and Pentland, A.: Visual recognition of American sign language using hidden Markov models, *Proc. IEEE Int'l Workshop on Automatic Face and Gesture Recognition (FG '95)*, pp.189–194 (1995).
- 14) Underkoffler, J. and Ishii, H.: Illuminating light: An optical design tool with a luminous-tangible interface, *Proc. ACM Conf. Human Factors in Computing Systems (SIGCHI '98)*, pp.542–549 (1998).
- 15) Underkoffler, J. and Ishii, H.: Urp: A luminous-tangible workbench for urban planning and design, *Proc. ACM Conf. Human Factors in Computing Systems (SIGCHI '99)*, pp.386–393 (1999).
- 16) Utsumi, A. and Ohya, J.: Multiple-hand-gesture tracking using multiple cameras, *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '99)*, pp.473–478 (1999).
- 17) Welch, G. and Bishop, G.: An introduction to the Kalman filter, *Course Notes 8 of ACM SIGGRAPH 2001* (2001).
- 18) Wellner, P.: Interacting with paper on the DigitalDesk, *Comm. ACM*, Vol.36, No.7, pp.87–96 (1993).
- 19) Yamato, J., Ohya, J. and Ishii, K.: Recognizing human action in time-sequential images using hidden Markov model, *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '92)*, pp.379–385 (1992).

(平成 14 年 7 月 10 日受付)

(平成 15 年 1 月 16 日採録)

(担当編集委員 佐藤 淳)



岡 兼司

2002 年東京大学大学院工学系研究科修士課程修了。同年同大学院情報理工学系研究科博士課程入学。現在に至る。ヒューマン・コンピュータ・インタラクションおよびコンピュータビジョンに関する研究に従事。電子情報通信学会会員。



陳 欣蕾

2002 年電気通信大学大学院情報システム学研究科博士前期課程修了。同年 NTT アドバンステクノロジー株式会社入社。現在、Human Interface Assessment & Design Center に所属し、インタフェースの研究開発・評価に従事。



中西 泰人(正会員)

1998 年東京大学大学院工学系研究科博士課程修了。同年より電気通信大学大学院情報システム学研究科助手。博士(工学)。感性情報処理、ヒューマンインタフェース、モバイルコミュニケーション等の研究に従事。ヒューマンインタフェース学会、人工知能学会、ACM 各会員。



佐藤 洋一(正会員)

1990 年東京大学工学部機械工学科卒業。同大学院情報工学専攻を経て、1991 年よりカーネギーメロン大学計算機科学部ロボティクス学科博士課程に在籍。1997 年 Ph.D. in Robotics 修了。同年より東京大学生産技術研究所研究機関研究員、講師を経て、現在同研究所助教授。コンピュータビジョン、ヒューマン・コンピュータ・インタラクション、およびコンピュータグラフィックスに関する研究に従事。1999 年山下記念研究賞、Int'l Conf. Shape Modeling and Applications '97 最優秀論文賞、MIRU2000 最優秀論文賞、1999 年度日本バーチャルリアリティ学会論文誌論文賞、IEEE VR2001 Honorable Mention for the Outstanding Paper Award 受賞。電子情報通信学会、日本バーチャルリアリティ学会、ACM、IEEE 各会員。



小池 英樹（正会員）

1991 年東京大学大学院工学系研究科情報工学専攻博士課程修了．工学博士．同年電気通信大学電子情報学科助手．1994 年同大学院情報システム学研究科助教授．現在に至る．1994～1996 年，1997 年 U.C. Berkeley 客員研究員．情報視覚化の研究に従事．特に視覚化へのフラクタルの応用，Perceptual User Interface，情報セキュリティへの視覚化の応用に興味を持つ．1991 年日本ソフトウェア科学会高橋奨励賞，2000 年情報処理学会 DICOMO2000 最優秀論文賞，2001 年 IEEE VR2001 Honorable Mention for the Outstanding Paper Award 受賞．ACM，IEEE/CS，日本ソフトウェア科学会各会員．
