

人間の行動計測・理解にもとづく実世界作業支援

佐藤洋一
生産技術研究所

1 はじめに

本プロジェクトでは、画像処理による人間の行動の実時間計測・認識技術の開発、およびその実作業支援への応用に関する研究を進めている。特に、紙書類などの実在メディアとマルチメディアコンテンツなどの電子メディアの双方を利用した作業に着目し、ユーザにとってより直感的かつ自然なインタラクションを実現するために必要とされる行動計測技術およびジェスチャ認識技術を開発した。図1に本研究で利用する拡張机型インタフェースシステムの概観を示す。本報告では、具体的に以下の各点に関して研究成果の概略を報告する。

- ユーザ行動の理解：複数指先軌跡の実時間計測
- ユーザ行動の理解：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の統合
- インタラクションにおける視線の利用
- 両手による直接操作の評価

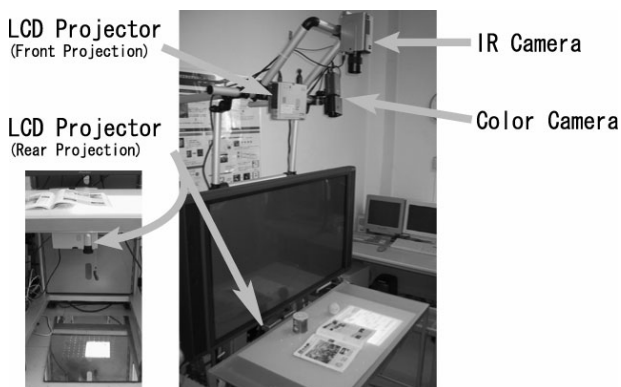


図1 拡張机型インタフェース

2 複数指先軌跡の実時間計測にもとづくジェスチャ認識

実在世界におけるユーザのインタラクションを考慮した実世界指向インタフェースを実現す

るためには、ユーザが行っている行動およびその意図を理解することが必要不可欠となる。そのためには様々なモダリティによる入力を利用することが考えられるが、その1つに指先の動きの追跡が挙げられる。人間は実物体を操作するときには手を使うことが多いが、特に細かい作業を行うときには指による操作を行うため、指先位置を正確に認識することは人間の動作を理解する上で重要な役割を果たすことになる。

以上の見地からこれまでも指先位置を認識する研究は行われてきたが、従来の指先位置認識には様々な問題があった。具体的には、手の向きが常に特定の向きにあることを仮定したり、指の本数は常に1本にしておくという制約を設けていたりしたため、人間が自然な振る舞いを行う場合には十分な認識性能を得られないことが問題であった。

これに対し本研究では、ユーザの複数の指先位置を安定に実時間で追跡するための手法を開発している。提案手法の特徴としては、赤外線画像の利用と、指先形状に注目したテンプレートマッチングによる高速かつ安定な指先位置検出が挙げられる。通常の可視領域光を検出するカメラではなく中赤外領域（ $5\mu\text{m}$ 程度の波長領域）に感度を持つカメラを利用すると、特定の温度領域からの放射光を検出することができる。その結果、物体表面における反射光による影響を受けずにユーザの肌領域を安定に抽出することが可能となる。これは、机上に様々な電子情報を投影しながらインタラクションを行う拡張机型インタフェースにおいてユーザ行動の理解する目的には特に有効である。このようにして得られた赤外線画像に対して正規化相関にもとづくテンプレートマッチング等の処理により指先位置を検出すると同時に収縮処理を適用することにより手のひらの中心位置を検出する。

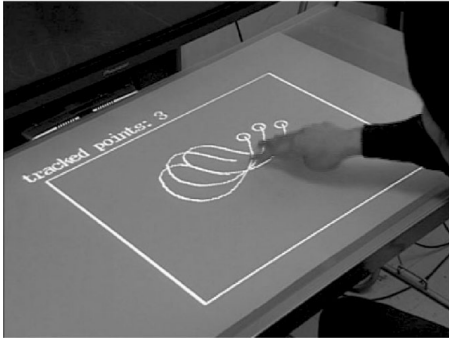


図2 指先軌跡の実時間計測

さらに、各フレームで検出された指先位置について、フレーム間の指先位置の対応関係を求める。具体的には、各画像フレームの指先位置を検出する際に直後のフレームでの各指先の位置をカルマンフィルタにより予測する。ここでは、各指先位置の運動として等速直線運動を仮定し基本システムを定式化している。さらに、各フレームにおいて検出された指先位置を観測値として入力する。また、正規化相関の際のマッチングスコアをもとにして観測誤差を推定し、カルマンゲインの計算に用いている。次に、このようにして予測される各指先位置と直後のフレームで実際に検出された指先位置との対応関係を求めることにより、安定かつ高速に複数の指先位置の軌跡を計測することが可能となる。図2に複数の指先軌跡を実時間で計測している様子を示す。

現在のシステムでは、PC (CPU: PentiumIII 1GHz, OS: Linux kernel 2.4.9) に画像処理ボード (HITACHI IP5005) を使用して複数指先軌跡の追跡処理を行っている。プログラムを十分に最適化した状態ではないが、毎秒 25 フレームから 30 フレームのビデオフレームレートでの実時間処理が実現されている。また、両手指先位置の検出も可能となっており、その場合の処理も毎秒 15 フレーム程度のスピードで実行することが可能である。

身振りなどの身体動作によるインタラクションを考えた場合、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の2つのモードに大別される。例えば、追跡された手や指先の位置情報はダイレクトマニピュレーションのための入力として直接的に使用することが可能であり、このような例として描画ツールや仮想物体のマニピュレーションのために手指の追跡技術を利用する研究などが報告されている。また、シンボリック

クジェスチャの認識には隠れマルコフモデルなどの統計的手法がしばしば用いられる。

これら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つのモードをユーザにとって自然な形で組み合わせることは、幅広いインタラクションを可能とするために重要である。しかしながら、従来の研究例ではこれらのモードの組み合わせという点はあまり注目されてこなかった。

これに対して本研究では、指先の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という2つのモードの自然な組み合わせを提供することを試みている。特に、詳細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場合が多いという観察に基づき、ユーザの親指に基づいたモードの切り替えを提案している。このようにしてダイレクトマニピュレーションと判別された場合には追跡された指先位置を直接的にインタラクションに用いるとともに、シンボリックジェスチャのモードでは HMM に基づき複数指先軌跡からのジェスチャ認識を行っている。複数ユーザによる評価実験の結果、これらのモードの切り替えおよび複数指先軌跡からのジェスチャ認識を信頼性良く行えることが確認された。

3 インタラクションにおける視線の利用

ユーザの机上における手や指の動作から作業内容を理解しようとする場合に、意味のある動作とそうではなく偶然取られた動作との判別が問題となる。例えば意識的に何かを指差しているのか、それとも単に机上に置いた手から偶然に1本の指のみが検出されたのかを正確に区別することは手指先位置の計測のみでは不可能である。

そこで、実世界指向インタフェースを実現するためには、音声や視線などの別のモダリティも利用してユーザの行動意図を理解することが必要不可欠となる。その中でも特に人間の目や顔の動きは、人の注意や意図と深く関係していることや、高速に移動できる、疲れないといった特徴があるため、コンピュータによる状況の認識及びユーザの意図理解への応用が期待できる。

本研究では、ステレオ画像処理にもとづきユーザの顔画像から頭部の位置および姿勢をリアルタイムで検出することが可能なシステムを構築

した。その後、環境光の変化などの外乱に対してより安定な処理を実現するために近赤外領域光に感度を持つ CCD カメラを用いたシステムへと改良した。更に、追跡開始時やシステムが顔の追跡に失敗したときに自動的に顔特徴を再検出し追跡を再開するためのアルゴリズムを開発した。現在のシステムはユーザの顔領域を観察するカメラ 2 台と顔領域を照らす近赤外照明装置、画像処理を行う PC、処理結果を出力するモニタから構成される。本システムでは頭部姿勢に関して最大誤差約 ± 1 度、頭部の 3 次元位置に関して最大誤差約 ± 1 mm の精度での検出が可能となった。処理速度は PC (Pentium III 550MHz, Linux OS) と HITACHI IP5005 画像処理ボードを使用し、毎秒 20 フレームでのリアルタイム処理が実現されている。

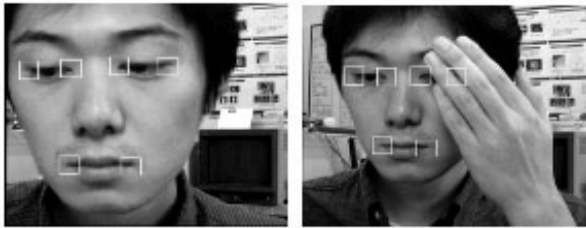


図3 トラッキングの様子

これまでに視線情報をヒューマンインタフェースの分野に応用する研究は盛んに行われてきたが、それらの多くはマウスやタッチパネルなどの入力デバイスを視線に置き換えたものであった。これらのシステムでは、コンピュータの操作のすべてを視線で行おうとしているが、視線を向けたコマンドすべてが選択されてしまうという問題などにより、コンピュータインタフェースとして必ずしも効率的ではなかった。

このような問題は俗に Midas touch problem と呼ばれ、視線情報をインタラクションに利用しようとする際の障害の一つとなっている。これは、視線情報は位置を指定するには適しているのに対し、イベントを発生させるのに適していないことに起因する。このため過去の研究事例では入力のタイミングを決定するために、ユーザの瞬き、ある一定時間以上同じ場所を見つめる、専用メニューのアイテムを見つめる、などを利用していた。しかし、ユーザの日常的な動作では作業しているウィンドウを常に見ているわけではなく、手元にある本などを参照しながらキーボード入力を行うことが多い。そのため、従来のシステムではユーザに常に目的のウィンドウを見ながら作業す

ることを強制してしまい、逆にユーザに負担をかける結果となりがちであった。このため、視線情報を利用した場合でも必ずしも作業時間が短縮できるとは限らなかった。

ここでは、インタフェースにおける視線情報の有効な利用法を検討するために、マウスやキーボードなどの従来型の入力デバイスとの比較において視線がどのような特徴を持つか、ユーザによる作業性評価を行った。特に、ウインドウインタフェースを想定し、(1)フォーカスウィンドウの切り替え、(2)ウィンドウの移動、(3)ウィンドウの重なり制御、(4)ズームやスクロールなど 3 次元の操作、の各基本操作を、頭部の位置・姿勢を用いて実行した場合の作業効率を評価した。特に、ここでは入力の際のあいまいさを解消するために単キー入力と視線方向を併用するアプローチを用い評価実験を行った。

その結果、ウィンドウ操作に関するものはそれぞれ目的のウィンドウを選択するために要した時間をマウスと比較すると、ウィンドウ選択に関して約 19.4%、移動に関して約 5.4%、重なり制御に関しては約 23.4% 高速であるという結果が得られた。3 次元操作の実験はズーム機能を用いて地図上の目的の位置を探すのに要する時間を測定し、キーボードなどを用いた場合と比較して効率が良いことがわかった。

3 両手直接操作にもとづくインタラクション

EnhancedDesk ではユーザが両手による直接操作で実在メディアと電子メディアを同時に取り扱うことができる点が大きな特徴となっている。マウスを両手で一つずつ使うダブルマウスなど、両手による操作の有効性がユーザインタフェースの分野で注目されているが、本システムのようにユーザの両手による直接的操作を対象とした例は見られない。ユーザは投影されたオブジェクトの端を文字通り直接つかむことができ、物体を移動したり、回転させたり、拡大したりといった操作を両手で直接行うことが可能である。

しかしながら、両手による直接操作がマウスなどの従来型のポインティングデバイスと比較してどのような点で優れているのかということについては明確な答えが得られていない。そこで本研究では、机型実世界指向インタフェースにおいてより効率の良いインタラクションの方法を設

計するためには両手による直接操作の特性を調べる必要があると考え、ユーザによる作業評価を進めている。

特に本研究では、2次元図形の描画作業を対象とし、利き手とは反対の手（現在のシステムでは左手）によってパイメニューと呼ぶ円周上に配置された階層メニューしながら、両手もしくは片手で描画することを可能とするツールを実装している。

このツールを利用して、マウス入力+CRT画面表示、マウス入力+机上投影表示、両手による直接入力+机上投影表示のそれぞれについて作業実行時間を比較した。この結果、ほとんどの場合で両手による直接操作が最も短時間で作業を実行できることがわかった。その理由としては、2次元図形の描画作業では両手操作による自由度の高さが有効に活用されること、作業を行う場所と映像が投影される場所が一致していることから直感的な操作が可能となること、などが考えられる。今後は、よりさまざまな作業内容について両手による直接操作の有効性を評価していくことを予定している。

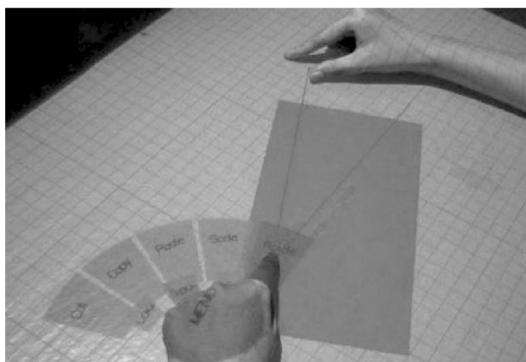


図4 両手による描画作業の評価

発表文献

1. Kenji Oka, Yoichi Sato, and Hideki Koike, "Real-time fingertip tracking and gesture recognition," *IEEE Computer Graphics and Applications*, Vol. 22, No. 6, pp. 64-71, November/December 2002.
2. 中西泰人, 岡兼司, 倉持正之, 松川昌平, 佐藤洋一, 小池英樹, "Narrative Hand: 高速な手指認識システムのメディアアートへの応用," 情報処理学会論文誌, Vol. 44, No. 2, pp. 309-316, February 2003.
3. Yasuto Nakanishi, Kenji Oka, Masayuki Kuramochi, Shohei Matsukawa, Yoichi Sato and Hideki Koike, "Narrative Hand: applying a fast finger-tracking system for media art," *Int'l Symposium on Electronic Art (ISEA 2002)*, October 2002.
4. Yasuto Nakanishi, Takashi Fujii, Kotaro Kitajima, Yoichi Sato, and Hideki Koike, "Vision-based face tracking system for large displays," *Proc. ACM Ubiquitous Computing (Ubicomp 2002)*, pp. 152-159, September 2002.
5. Kenji Oka, Yoichi Sato, and Hideki Koike, "Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems," *Proc. 2002 IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG 2002)*, pp. 429-434, May 2002.
6. Kenji Oka, Imari Sato, Yasuto Nakanishi, Yoichi Sato, and Hideki Koike, "Interaction for entertainment contents based on direct manipulation with bare hands," *Proc. 2002 IFIP Int'l Workshop on Entertainment Computing (IWEC 2002)*, pp. 391-398, May 2002.
7. Xinlei Chen, Hideki Koike, Yasuto Nakanishi, Kenji Oka, and Yoichi Sato, "Two-handed drawing on augmented desk system," *Proc. 2002 Int'l Working Conf. Advanced Visual Interfaces (AVI 2002)*, pp. 219-222, May 2002.
8. 岡兼司, 陳欣蕾, 中西泰人, 佐藤洋一, 小池英樹, "拡張机型インタフェースのための複数指先の追跡とその応用," 情報処理学会 画像の認識・理解シンポジウム (MIRU 2002) 論文集, pp. I-39-48, July 2002.
9. 岡兼司, 佐藤洋一, 小池英樹, "手動作認識の安定化に向けた注視点情報の利用," 電子情報通信学会研究会報告, PRMU2002-65, pp. 53-58, July 2002.
10. 藤井崇志, 長澤篤, 小池英樹, 佐藤洋一, 中西泰人, "大型情報表示環境のための頭部位置・姿勢検出手法とその応用," 情報処理学会研究会報告, 2002-HI-98, pp. 33-40, May 2002.