

バーチャルキーボード: ビデオ画像からの頑健な実時間指先検出の実現

Virtual Keyboard: Realization of Robust Real-Time Fingertip Detection from Video-Image

松井 望[†]

Nozomu MATSUI

山本 喜一[†]

Yoshikazu YAMAMOTO

[†]慶應義塾大学大学院 理工学研究科
Graduate School of Science and Technology,
Keio University

概要

我々の研究であるバーチャルキーボードは、特殊なハードウェアを使わずにユーザがキーボードを作成し利用できるシステムである。バーチャルキーボードシステムは、CCDカメラで撮影したユーザの指先の位置を実時間で検出し、システム内で仮想的に作成されているキーボードをタイプするためのシステムである。これまでに我々の研究で得られた結論の1つが、“ビデオ画像から指先の位置を実時間で正確に検出する必要性”である。2次元ビデオ画像から毎フレーム各指先座標を正確に検出する事は容易でなく、これに関わる処理コストは非常に高い。従って、このアプローチを実用的なバーチャルキーボードシステムの実現に利用するには困難であると考えて来た。しかし昨今の計算機の発展により、エンドユーザが利用している計算機でも実用に近いレベルでバーチャルキーボードシステムを利用可能になって来た。本稿ではバーチャルキーボードシステム実現に必要な、頑健な実時間指先位置検出の実現方法について述べる。

キーワード

モダリティンタフェース, バーチャルキーボード, 実時間指先検出, 単眼視物体認識.

1 導 入

昨今の計算機の進化は非常に早く、ラップトップPCも小さくて処理能力の高いものが多く存在する。一般に、PCにはキーボードが必要である。当然ラップトップPCが小さくなれば、キーボードも小さくなる。しかしキーボードがあまりにも小さくなると、能率を下げる原因になり得る。結果、ラップトップPCの大きさはキーボードの大きさに制限

され、ある程度のサイズよりも小さくできない状態にある。

もともとPCのキーボードは、一部の人にしか受け入れられていないインタフェースであった。現在は電子メールサービスの普及もあって、一般のユーザでもPCのキーボードに抵抗を感じる人は少なくなってきた。しかし、少し前まではPCのキーボードが苦手なためにPCが利用できずにいるユーザが多く存在した。これは、インタフェースを利用する際に共通して起こり得る問題の1つである。

例えば、ユーザがあるインタフェースを初めて利用する場合を想定する。そこには既に用意されて

いる既定のインタフェースが存在する．PCのキーボードも同様である．まずユーザは，PCを利用するためにPCキーボードの利用に慣れる必要がある．ただPCを使いたいだけであるのに，インタフェースの使用訓練を強いられて来たのである．

このように，インタフェースはユーザではなく第三者によって与えられる場合がほとんどである．しかもこれらインタフェースの大半は，目で見てキーの位置などを覚えるなどの訓練が必要である．また目の不自由な人にとっては，インタフェースの使用訓練は非常に苛酷なものとなる．

我々は，この問題を解決するためにバーチャルキーボードシステム [1]を提案した．バーチャルキーボードシステムではCCDカメラが接続されたPCさえあれば，使用訓練を必要としない自由な形状のキーボードを利用することができる．また手の大きな人でも，自分に適した大きさのキーボードを自ら作成し利用することができる．我々のバーチャルキーボードシステムが実用システムになるためには，2次元ビデオ画像から指先の位置を実時間で正確に検出できる事が必要となる．

本稿は，以下のように構成されている．第2節では，研究背景について述べる．第3節では関連研究との比較，第4節では実現方法，第5節では評価，第6節では応用例，そして第7節では本稿のアプローチについてまとめる．最後に第8節において今後の課題点について述べる．

2 背景

バーチャルキーボードは，ユーザ自身がキーボードを仮想的に作成できるシステムである．本システムは特殊なデバイスを必要としない．必要となるデバイスは，1台のCCDカメラとビデオキャプチャーカードが接続されたPCだけである．バーチャルキーボードシステムは，CCDカメラで撮影したユーザの指先の動きからキー入力を推定する．本システムの特徴の1つは，物体を追跡するのではなく正確な指先の位置を追跡する点である．このような実時間で指先の位置を正確に追跡しようという研究は，他に例を見ない．

これまで人間の指先の正確な位置を実時間で検出する方法についてあまり研究がなされてこなかった理由には，

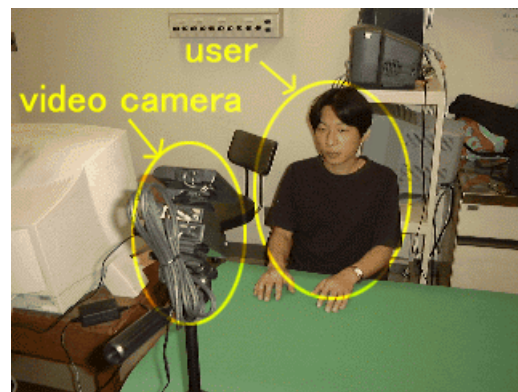


図1 バーチャルキーボードシステム環境

1. 指先を検出する必要があるアプリケーションが考えられてこなかった
2. 画像処理コストの面から従来の計算機環境では実用に見合わなかった

などの理由が考えられる．しかし理由1に関しては，我々の研究によるバーチャルキーボードシステムや第6節で挙げる応用例など，多くの可能性を持ったアプリケーションが考えられる．また理由2については，昨今の計算機処理速度の向上によって回避できると考える．

これらの点から，2次元ビデオ画像を基にした指先位置検出方法を確立することには意義があると考え，本稿では我々のバーチャルキーボードシステムにおいて実装した指先検出方法について述べる．

3 関連研究との比較

新しいインタフェースを提供するための研究は，他にも多く行なわれている．

FingerRing [2] [3]は，指先で机などをタイプすると入力として認識するシステムである．この研究では，机をタイプする際に生じるショックを指輪のような特殊センサーで検知することによって実現されている．

自動手話認識生成システム [4]は，手話を知らない人でもろう者とコミュニケーションを取ることができるシステムである．この研究では手の形状認識を正確に行なうために，2台のCCDカメラが使用されている．

いずれも画期的なシステムだが，これらのシステムの利用には特別な環境が必要となる．バーチャルキーボードシステムは，特殊なセンサーも2台の

CCD カメラも必要としない．1 台の PC に 1 台の CCD カメラが標準となりつつある現在では，一般家庭のエンドユーザでも簡単に利用できる，自由で手軽な新しいインタフェースである．

4 実現手法

本稿で述べる指先位置検出法は，以下の手順で実現されている．

1. 背景の除去
2. 手と手の分離
3. 手以外の物体の除去
4. 手の骨格化
5. 指先の検出

理想的なビデオ撮影環境では，上記の処理群だけで十分頑健な指先検出が実現できる．しかし一般の利用環境下では，理想的なビデオ環境が提供されているとは限らない．より頑健な指先検出を実現するには，上記の処理群以外にもいくつかの処理が必要であることが，これまでの我々の研究から分かっている．

以下に上記の処理群およびこれら以外の必要な処理について，処理の順序に従って具体的に説明する．

4.1 背景以外の物体を抽出する

本稿で述べる指先位置検出法は，まずビデオカメラから送られてくる RGB 画像情報から背景以外の部分を抽出する事から始まる．以下にその具体的な実現手順を述べる．

まず，背景だけが撮影されている画像をあらかじめ 1 枚撮影しておく．背景は単一色であると仮定する．この背景画像から，背景の RGB 色情報それぞれの範囲を取得しておく．実際に抽出する物体が撮影されている画像には，背景色の部分と手などの物体の部分撮影されている．この画像の中から，あらかじめ取得しておいた背景色の RGB 色情報値の範囲内におさまっている部分を背景として除去する．こうすることで，背景以外の物体の部分だけが残される．

この抽出方法はある程度の成績を納めており，十分に実用に耐え得るものとする．図 2 は，黒い学習機の除去結果である．

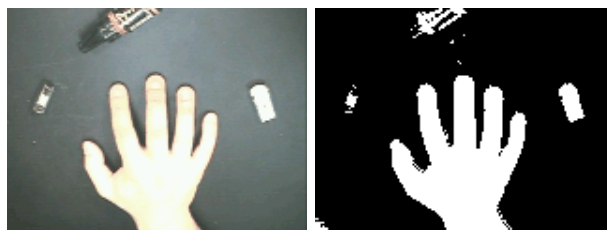


図 2 背景の除去

抽出された物体は，物体が存在する領域を白とする白黒画像に変換され後の処理に渡される．2 値画像は，ラベリングによる物体特徴抽出が容易に行なえるなどの利点がある．

4.2 手と影の分離

前の処理で背景以外の部分を抽出した 2 値画像には，ノイズや手の影，机の汚れなどが含まれている場合が多い．また実際の使用環境においては，手以外の物体が撮影されている場合が大いにあり得る．



図 3 分離可能な手と影の接続状態

これら手以外の物体は，後の処理でラベリングを利用した特徴抽出法で除くことが可能である．しかし，手の影は物体特徴抽出法でも取り除けない場合が多い．これは，手と手の影はお互いが接触しており，抽出結果もそれらがつながった状態である事が多いからである．

これまでの我々の研究結果から，手と手の影はただ 1 つの画素によって接続している場合が多いことが分かっている．この状態の画素をなんらかの方法で取り除くことができれば，影と手が分離できる確立が高くなる．

この実現は，画素の接続性を調べることで可能である．画素の接続性とは，その画素がいくつの領域の接続に関与しているかを表すものである．図 4 は

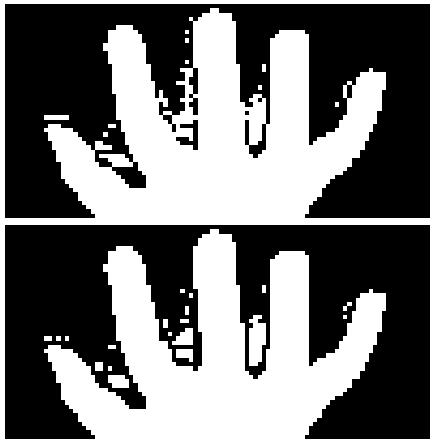


図4 手と手の影の分離

影の分離前後を表している。

4.3 無効物体の除去

実際の使用環境では、机の上に手以外の物体が置かれている場合が少なくない。この場合でも、ラベリングを用いれば手以外の物体の除去が可能である。

ラベリングは、2値画像での物体特徴抽出によく用いられる手法である。物体特徴抽出法では、物体の重心、周囲長、面積などさまざまな特徴を知ることができる。

ここでは、物体の面積情報を用いて手以外の部分を除去している。図5は、無効物体除去前の2値画像と除去後の2値画像を表している。



図5 無効物体の除去

4.4 手の骨格化

バーチャルキーボードシステムの実現に必要な情報は手の領域ではなく、あくまで正確な指先位置である。本稿で述べる指先位置検出方法では、簡単に抽出するための処理として“骨格化”を用いている。



図6 骨格化処理

この骨格化には、エッジ処理でよく用いられる細線化を応用している。この処理によって、抽出された手は図6の様にちょうどレントゲン写真によって撮影された手の骨格のような状態になる。

4.5 指先の検出および無効な指先の除去

前の処理までで正しく骨格化されていると仮定すれば、骨格の先端を検出すればそれらが指先となる。しかし骨格化で得られる結果は、理想のものでない場合が少なくない。



図7 無効な指先

図7は、指先でない部分が指先として検出されてしまう例である。これは影の分離に失敗したことによって、手の輪郭が整っていない場合に起こり得る症状である。

これら無効な指先は、ラベリングで得られた物体特徴を利用することで除去が可能である。ここでは、物体特徴のうちの重心座標情報を利用する。手の重心から一定の距離以内に存在する指先を無効とすることで、これらの指先の除去が可能になる。

骨格化を用いた指先位置検出では、人間の爪の中心あたりが指先として検出される(図8)。これは人間が感覚的に指先と認識している位置と一致している。

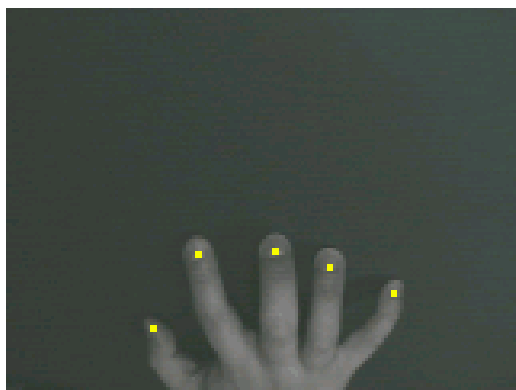


図 8 検出された指先

5 評 価

本節では、本稿で述べた指先位置検出法が実用可能なものか各処理毎に速度面、頑健性面、および実際にバーチャルキーボードシステムに利用した時の入力認識率について評価した結果を述べる。

5.1 速度評価

評価環境は以下の通りである。

- ・画像サイズ 160 × 120 画素
- ・計算機 PentiumIII 500MHz, DualCPU
- ・OS FreeBSD-3.3R

表 1 は、100 フレーム画像を処理した時のそれぞれの処理工程の平均処理速度である。

処理名	処理速度 ($\mu s/frame$)
背景の除去	1961
手と影の分離	4318
手以外物体除去	26471
骨格化	5462
指先選択	237
計	38449

表 1 速度評価

この結果を評価する 1 つの指針として、PC キーボードを 1 分間にタイプする速度を基準としてみる。

PC キーボードでは、タイピングが得意な人ならば 1 分間に 500 字をタイプすることが可能である。一般のタイピングデバイスでこのタイプ速度に勝る

ものは、他に存在しないと考えるよくだらう。

我々のバーチャルキーボードシステム入力検出部は、1 つのキーの入力推定に連続するビデオ画像フレームを最低で 3 フレーム必要とする。よって 1 字を認識させるための画像処理にかかる時間は、およそ $115500 \mu s$ となる。1 分間に換算すると、およそ 519 字のタイピング動作認識が可能である。この数字は、十分に実用速度に耐えるものと言えるだろう。

5.2 頑健性評価

今回の開発は、実用環境で正確に指先の位置を検出することが目的である。図 2 は、一般の学習機での背景の除去の結果を示している。一般の学習機では、光源の位置によって机面が反射する。しかし、本稿で述べた背景除去法では机の部分の正しく除去することに成功してゐる。また、白色の背景でもラベリングによる無効物体領域の除去によって手以外の部分を除去することに成功している。光源の状態や CCD カメラの能力にも左右されるが、本稿で述べた指先検出法は一般に人間が無理なく本を読める明るさであれば正しく指先の位置を検出することに成功している。

5.3 入力認識率評価

本稿で述べた指先検出法の評価例として、実際に約 300 字の文字入力を行なった時の入力認識率 (表 2) を挙げる。評価時のキーボードパターンは図 9 である。

サンプル入力文字数	296
認識された文字数	448
正常認識率	66%
重複認識率	15%
誤認識率	19%
過剰認識率	40%

表 2 入力認識率

ユーザの意図に反して認識されるキー入力は、重複認識と誤認識の 2 つに分類できる。重複認識とはユーザの意図に反して同じキーを続けて入力認識することであり、オートリPEAT機能を使用している時と同じような状態になる。誤認識とはユーザが意

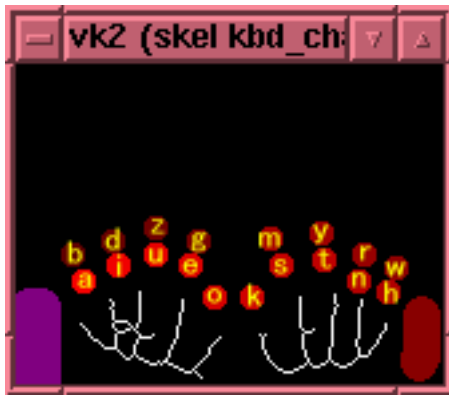


図9 キーボードパターン

図しないキーが入力認識されてしまうことである。過剰認識は重複認識と誤認識の総和である。

上記の結果を見る限り、本稿で述べた指先検出法を用いたバーチャルキーボードシステムは、過剰な認識が多くユーザが意図する文字列よりもかなり多くの文字が認識されている。一方、ユーザが意図する入力文字が認識されない場合はなかったことが伺える。

この過剰認識の原因の一つには、指先検出部が指先を正しく検出できない事が挙げられる。指先検出部が指の間の影などを指先として検出してしまうと、ユーザが指を動かさなくても影が動けば入力認識されてしまう恐れがある。この無効な指先の除去には前節で述べた手法を用いているが、上記の結果を見る限り十分ではないことが分かる。

過剰認識はキーボードの配置パターンによって減少させることが可能であり、図9はそのキーボードパターンの一例である。このキーボードパターンはユーザの手がホームポジションに位置する時、手の下にキーが存在しないように考慮されている。しかし、キーボードパターンに制限を与えることで入力認識率を向上させる事は本末転倒であり、望ましくない。

これらの結果には入力認識部の能力も関係してくるが、指先の検出が常に安定して正しく行なわれていれば入力認識率も向上すると考える。今後、指先検出部のさらなる向上とともに、インテリジェントな入力検出部の開発が必要である。

6 応用例

本稿で提案した指先位置検出法は、以下の様な仮

想インタフェースに応用可能である。

6.1 キーボード

図10と図11は、我々の研究であるバーチャルキーボードで現在実現されている例である。

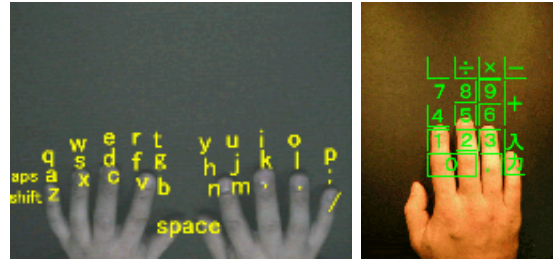


図10 キーボード例



図11 オリジナルキーボード例

指先位置検出法を用いたバーチャルキーボードシステムならば、図10のようなキー同士が近接したキーボードでも実現が可能である。

6.2 ピアノ

ピアノもキーボードの1つである。PCのキーボードと違う部分は、キーを叩く速度が違えば発せられる音の大きさが違ってくる点である。

この点も指先位置検出法を利用した入力認識法ならば、実現が可能である。

6.3 障害者支援

目が不自由な人にとって、既存のキーボードでは必要だった使用訓練によってかかる負担は計り知れない。キーボードのキーの位置を目で見えて覚える事が可能な人でも、使用訓練は非常に苦であり訓練期

間中の作業効率は非常に悪い。

指先位置検出方法を用いたバーチャルキーボードシステムならば、目が不自由な人でも自分の好きな様にキーボードを作ることが可能である。これにより、今まで不自由だったキーボードの使用訓練の必要がなくなり、障害者がキーボードを利用する際の負担が非常に軽減される事が期待できる。その上、高価で特殊な障害者用のインタフェースを購入する必要もなくなり、より多くの人を手軽にバーチャルキーボードを利用することが可能であると考えられる。

6.4 教育用途

指先位置検出方法は、キーボード以外にも応用が可能である。例えば、幼児教育でよく利用されるかるたゲーム(ひらがなカード遊び)などにも応用が可能である。CCDカメラでかるたを撮影し、カードの位置をバーチャルキーボードでのキーにあたる部分に割り当てる。ユーザ(幼児)がかるたゲームシステムに指定されたカードの位置を指先で差し示し、一致すれば正解を告げるようなシステムも実現可能と考える。

6.5 タッチパネル

指先位置検出方法を用いれば、CCDカメラによるタッチパネルも実現が可能となる。同様に、タッチパネル機能を持たないプロジェクターのような表示デバイスでもタッチパネル機能を実現することができると思われる。

7 結論

結論として、本稿で述べてきた指先位置検出方法は速度、応用性それぞれの面において十分実用可能であると言える。

頑健性については、CCDカメラが正しく物体を撮影できる環境であれば正確に指先を検出できている。一般のユーザが利用している机の上では、十分実用に耐えることが確認できた。

8 今後の課題

本稿では、ビデオから送られて来る画像以外はずべて2値画像を使用してきた。画像からの物体抽出

に2値化する方法は伝統的でありその取り扱い方法も多く存在するが、画像情報量の損失は明らかである。また我々の研究であるバーチャルキーボードに必要なものは物体抽出ではなく、あくまで正確な指先の位置である。

今回の開発ではまず背景以外の物体の抽出から作業が始まるが、この手法がバーチャルキーボードシステムにとって最適であるかどうかはまだ結論に達していない。また、RGB色系よりも肌色抽出に適した色系が存在することも既に知られている。

今後はこれらの色系を利用するなどしてできるだけ情報量を損失しないまま、より頑健で正確な指先の位置を検出する方法を開発し、多くの評価をする必要がある。

謝辞

本稿を仕上げるにあたり協力して下さった鈴木純一氏、辻将悟氏に感謝致します。

参考文献

- [1] 松井 望他. バーチャルキーボードの提案. 日本ソフトウェア科学会第15回大会論文集, pp. 213-216, 1998.
- [2] 福本 雅朗他. Keyboard for wearable computer Fingering. 電子情報通信学会論文誌 A, Vol. J79-A, 1996.
- [3] Masaaki FUKUMOTO, et al. Body coupled fingering: Wireless wearable keyboard. *SIGCHI97*, 1997.
- [4] 郵政省 通信総合研究所 総合通信部ユニバーサル端末研究室. 自動手話認識生成システム. <http://www.crl.go.jp/st/st821/research/sl-j.htm>, 1998.
- [5] 岩澤 昭一郎他. 熱画像からの実時間全身像の姿勢推定法の検討. 電子情報通信学会 信学技報, Vol. OFS96-30, pp. 25-30, 1996.
- [6] 有福 義範他. キー入力判定のための判別式の検討. 電子情報通信学会技術研究報告, HCS97-21, 1997.
- [7] 渡部 富男他. ヒューマンインタラクション解析のための身体的バーチャルコミュニケーションシステム. 電子情報通信学会技術研究報告, HCS98-1, 1998.
- [8] Christopher Wren, et al. Pfinder: Real-Time Tracking of the Human Body. *SPIE*, Vol. 2615, pp. 89-98, 1996. MIT Media Laboratory.
- [9] 中司 修二他. 動画における部位画像の追跡に関する一検討. 電子情報通信学会 信学技報, pp. 119-123, 1998.
- [10] 山田 正紀他. 実時間人物全身像動き検出. 映像情報メディア学会技術報告, Vol. 21, pp. 61-66, July 1997.
- [11] 田中 昭二他. 画面領域の物理的特徴に基づく注目領域抽出法の評価. 映像情報メディア学会技術報告, Vol. 22, pp. 7-12, March 1998.