

ステレオカメラによる3次元計測と信号音による誘導を用いた タッチパネル操作支援システム

○大西由純 山下淳 金子透
静岡大学工学部機械工学科

本研究では視覚障害者のためのタッチパネル操作支援システムを提案する。利用者は音声入力によりシステムに指示を与える。ステレオカメラによる情報からシステムはタッチパネルボタンと利用者の指を検出し、両者の位置関係から任意のボタンへ指先を誘導するための信号音を出力する。利用者は信号音を頼りに指先をボタンの位置まで移動させることによってタッチパネルの操作を行う。実験により人の視覚に頼ることなくタッチパネル操作が可能であることを確認した。

1. 序論

近年、銀行のATMや駅の券売機等、タッチパネル操作を取り入れた機器が至る所に設置され、これらは日常生活をおくる上で必要不可欠なものとなりつつある。タッチパネル機器は従来の機器のようなボタンの凹凸が無く、また、点字を併記することもできないため、視覚障害者がこれら进行操作することは困難である。この問題を解決する方法として第三者の補助を受けることもできるが、個人情報漏洩の観点から好ましくない場合もある。また、テンキーと音声案内による補助入力システムを内蔵したタッチパネル機器も存在するが、コスト削減等の観点からその機能が省かれることもあり、利用可能な環境が限定される。

そこで、本研究では視覚障害者が単独でのタッチパネル操作を可能とする支援システムの構築を目的とする。視覚障害者のための研究としては、カメラを利用し文字情報を取得することで視覚障害者が周囲環境を把握するための研究[1][2]や、身に着けた小型カメラにより取得した周囲環境の様子を3次元音響を用いて伝える研究[3]等がある。これらの研究は歩行支援を目的としたものであり、視覚障害者の行動範囲を広げるための汎用的な技術を提供する。一方で、視覚障害者の生活環境における事細かな要求に対しては、これら汎用的な技術のみでは対応することができない。視覚障害者のためのタッチパネル操作を支援する他の研究としては、携帯型カメラを利用し、指先周辺の文字情報を取得し利用者にもその情報を伝える研究[4]がある。この研究の手法では指先周辺の情報を対象としているため、指先を目的の

ボタンへと動かす際に指標となるものが必要となる。本研究では視覚障害者の生活範囲をより広げるため、指先を目的のボタンまで誘導するタッチパネル操作支援システムを構築する。

タッチパネル操作で比較的多くみられる操作は数値入力である。例えば、ATMでは暗証番号や金額の入力の際に数値入力が求められる。そこで、本研究では図1に示すような数値入力を行うタッチパネルの操作を想定しシステムを構築する。また、構築したタッチパネル操作支援システムを評価し、システムの有効性を確認する。



図1 操作を想定するタッチパネル

2. システムの概要

機器の構成を図2に示す。利用者はマイクを介した音声入力によりシステムに指示を与える。システムはステレオカメラから画像を取得し、タッチパネル操作の誘導に必要な情報を取り出した後、この情報を信号音や音声という形でヘッドホンに出力する。利用者はこのヘッドホン出力から得られる情報を基にタッチパネル操作を行う。

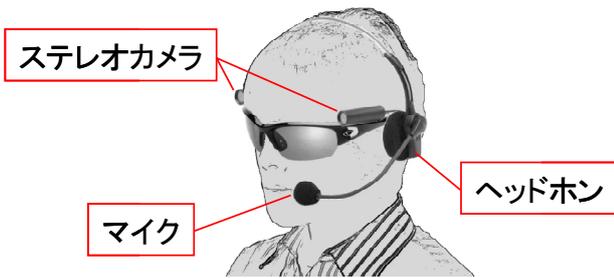


図2 機器の構成 (概念図)

3. タッチパネル操作の支援処理の流れ

システムの処理の流れを図3に示す。まず、タッチパネル上の操作するボタン名を音声により入力する。次にステレオカメラから画像を取得する。取得した画像からボタン及び指先の検出を行い、各々の3次元座標を算出する。求めた3次元座標から指先と操作するボタンの相対位置を計算し、相対位置情報に基づいた信号音を生成してヘッドホンに出力する。以降、画像取得から信号音出力までの処理を繰り返し行う。逐次更新される信号音を頼りに指先をボタンの位置まで移動させることによって利用者はタッチパネル操作を行う。

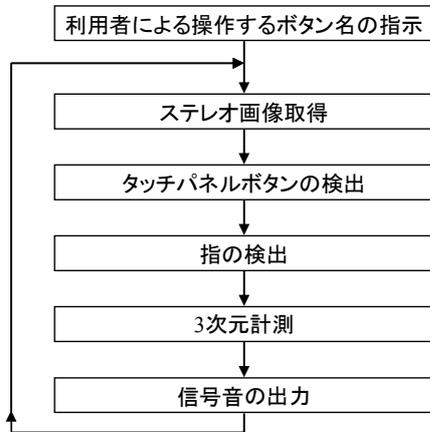


図3 誘導処理の流れ

3.1 操作するボタン名の指示

タッチパネル操作を支援するために、システムは利用者が操作しようとするタッチパネルのボタンを知る必要がある。まず、利用者は音声入力により目的とするボタンの数字、或いは記号をシステムに伝える。これだけではシステムが正しく入力を認識したのかが確認できないため、システムには認識した数字を音声出力により復唱させる。復唱された数字が誤ったものであるなら、利用者は音声入力により再度数字を入力する。尚、誤動作防止のため、音声

入力の際は機械スイッチを一度クリックしてからおこなう。システムはクリック後に発せられる利用者の一単語のみを認識するように設定されている。

3.2 ステレオ画像取得

ステレオカメラを用いてステレオ画像を取得する。前提条件として、ステレオカメラはタッチパネル画面に向けられ、その全景を映していることとする。

3.3 タッチパネルボタンの検出

タッチパネル上の文字を認識することでボタンを検出する。画像から文字情報を抽出するにあたり文献[1]に基づく手法を用いた。文字が持つ一般的な特徴(アスペクト比、背景と黒字部のバランスなど)を照合し、原画像から同様の特徴を持つ領域を抽出する。抽出した領域は文字を内包する可能性が高く、この領域に対して文字認識処理を行う。文字認識で得た文字情報と画像上での認識位置との対応を取ることで各文字の位置を特定し、この文字の位置を各々のボタン位置とみなす。

タッチパネル上のボタンは操作を行う手の動きにより隠れてしまう場合がある。このような状況ではボタンの検出を正しく行うことはできない。本研究では次に示す手法を用いてボタンが手によって隠れるオクルージョン問題の軽減に努めた。まず、最初にボタンが隠れない状態で検出した全ボタンの画像上における位置を記録しておく(図4(a))。次に、タッチパネル上のボタンの一部が隠れた画像(図4(b))を得た際、画像上で確認できるボタンの位置だけを取り出し、最初に記録したボタンとの対応から、ボタン位置を記録した画像→ボタンの一部が隠れた画像への射影変換行列を導出する。最後に、位置を記録した画像のボタン位置を射影変換することで隠れた部分のボタン位置を補完し、オクルージョン問題を軽減する。

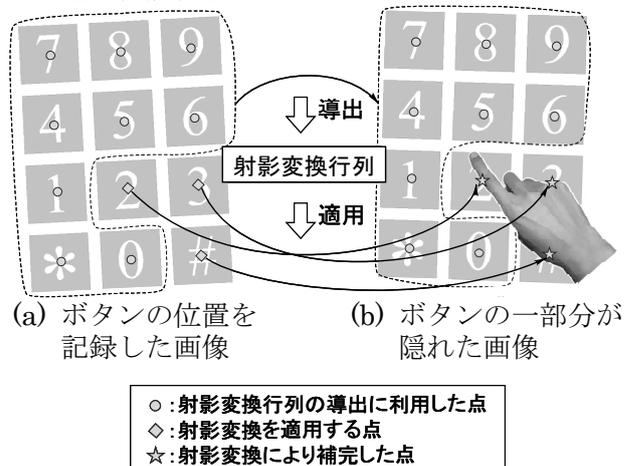


図4 射影変換によるボタン位置の補完

3.4 指の検出

本システムでは、指の位置、姿勢の情報に基づき信号音を生成する。従って、ここでは画像上における指先の位置、及び指付根の位置を導出する。

はじめに原画像（図 5(a)）の肌色領域を手の領域として抽出し、その輪郭線を取得する（図 5(b)）。取得した輪郭線を簡略化し（図 5(c)）、手の領域の重心から最も遠い位置にある輪郭線上の鋭角の頂点を指先とする（図 5(d)）。求めた指先頂点を挟む 2 直線から指さし方向の直線を求め（図 5(e)）、その直線上において手領域の重心に最も近い点を指付根の位置として定める（図 5(f)）。

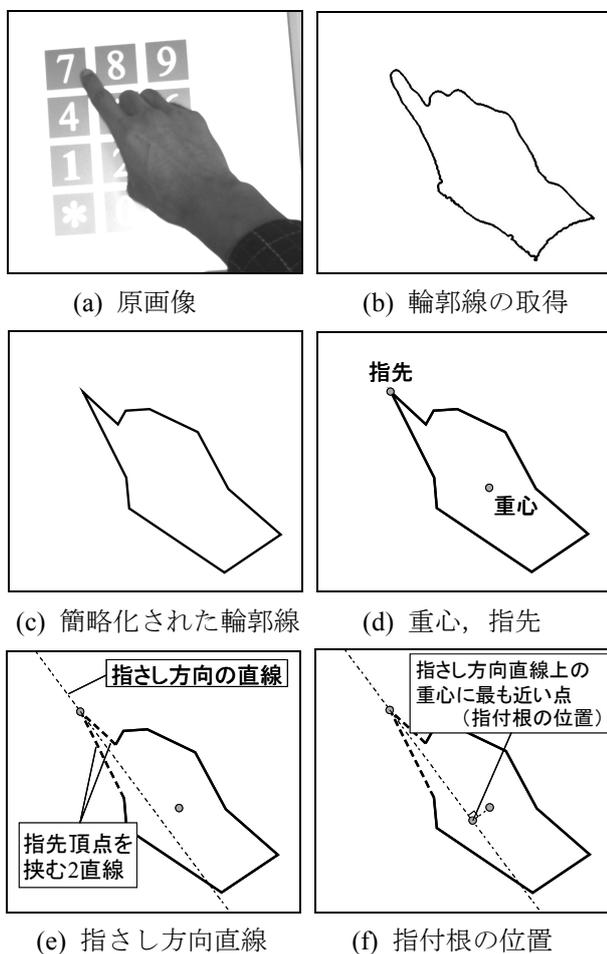


図 5 指の検出

3.5 3次元計測

左右両画像に対してボタン、指を各々検出した後、検出結果を基に左右画像の対応をとり、3次元計測処理を行う。ここでは、平行ステレオカメラを用いて、ボタンと指先、及び指付根の3次元位置座標を算出する。

3.6 信号音の出力

タッチパネル操作の誘導には信号音を利用する。信号音を使用する理由として、音声との対比がよく持ち上げられる。音声によって言葉を伝えるには複数の音節が必要であり、従って、単純な情報の伝達であってもある程度の時間の長さ（スパン）を必要とする。一方、信号音はそのスパンを自由に設定でき、また、スパンの短い信号音を用いたとしても、音程、音量等を利用することである程度の情報を持たせることが可能である。操作の誘導に信号音を用いることにより比較的短いスパンで多くの状況を把握できるため、指先とボタンの位置関係の変化に対して利用者は即座に反応することが可能となる。

信号音の生成には、操作するボタンと指との位置・姿勢関係から得られる 2 つのパラメータを用いる。信号音生成の際のパラメータには、

- タッチパネル面における指さし方向の投影線を基準としたボタンの方向を示すベクトル $\mathbf{b-h}$
- タッチパネル面における指先の投影点とボタン間の距離 $|\mathbf{b-r}|$

を利用する（図 6）。前者は指さし方向の投影線がパネル面の領域を分断するため、分断された左右領域のどちらにボタンが存在するかの指標となる。後者は指先とボタン間の距離的感覚を提供する。これら 2 つのパラメータにより信号音を生成することで直感的なタッチパネル操作が可能となる。

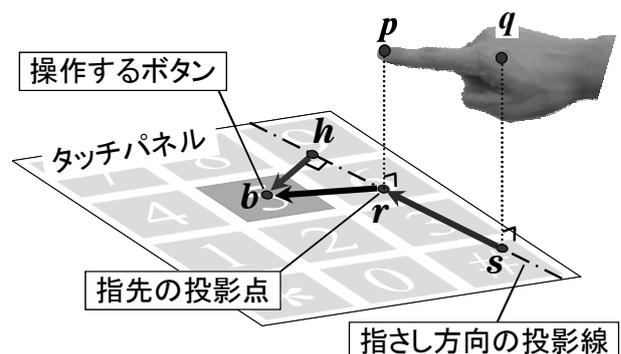


図 6 信号音生成に用いる位置情報

各パラメータ算出の手順を述べる。まず、タッチパネル面における指先の投影点 \mathbf{r} と目標とするボタン \mathbf{b} との距離 $|\mathbf{b-r}|$ を求める。タッチパネル面における指先の投影点 \mathbf{r} は、

$$\mathbf{r} = (x_p - ka, y_p - kb, z_p - kc) \quad (1)$$

$$\text{ただし, } k = \frac{ax_p + by_p + cz_p + d}{a^2 + b^2 + c^2} \quad (2)$$

また, 係数 a , b , c , d はタッチパネルの平面方程式 $ax + by + cz + d = 0$ により与えられるものである. これは予めタッチパネルの全ボタンの位置より求めておく必要がある.

求めた指先の投影点 \mathbf{r} と目標とするボタン位置 \mathbf{b} 間のベクトル $\mathbf{b} - \mathbf{r} = (x_b - x_r, y_b - y_r, z_b - z_r)$ よりその距離 $|\mathbf{b} - \mathbf{r}|$ を求め, これを指先の投影点と目標とするボタンとの距離とする.

$$|\mathbf{b} - \mathbf{r}| = \sqrt{(x_b - x_r)^2 + (y_b - y_r)^2 + (z_b - z_r)^2} \quad (3)$$

次に, タッチパネル面における指さし方向の投影線を基準としたボタンの方向を示すベクトル $\mathbf{b} - \mathbf{h}$ を求める. まず, タッチパネル面における指の投影ベクトル $\mathbf{r} - \mathbf{s} = (x_r - x_s, y_r - y_s, z_r - z_s)$ を求めるため指付根の投影点 \mathbf{s} を求める.

$$\mathbf{s} = (x_q - la, y_q - lb, z_q - lc) \quad (4)$$

$$\text{ただし, } l = \frac{ax_q + by_q + cz_q + d}{a^2 + b^2 + c^2} \quad (5)$$

ボタンの方向を示すベクトル $\mathbf{b} - \mathbf{h}$ は,

$$\mathbf{b} - \mathbf{h} = (\mathbf{b} - \mathbf{r}) - t(\mathbf{r} - \mathbf{s}) \quad (6)$$

$$\text{ただし, } t = \frac{(\mathbf{r} - \mathbf{s}) \cdot (\mathbf{b} - \mathbf{r})}{(\mathbf{r} - \mathbf{s}) \cdot (\mathbf{r} - \mathbf{s})} \quad (7)$$

求めた $\mathbf{b} - \mathbf{h}$ から指さし方向の投影線を基準としたボタンの方向 (左右) を判断するため, 次式 D の値を算出する.

$$D = \frac{\mathbf{p} - \mathbf{r}}{(\mathbf{r} - \mathbf{s}) \times (\mathbf{b} - \mathbf{h})} \quad (8)$$

D が正の値となる場合, 指さし方向の投影線より左にボタンがあると判断し, 負の値をとるときには右にあると判断する. ただし, ボタンが指さし方向の投影線上に存在する場合も想定して, あるしきい値 θ を設け, $|\mathbf{b} - \mathbf{h}|$ の値が次式を満たす場合, ボタンは指さし方向の投影線上にあるものと判断する.

$$|\mathbf{b} - \mathbf{h}| \leq \theta \quad (9)$$

次に, 出力する信号音の仕様について述べる. 信号音はステレオ出力とする. 信号音の左右音量のバランスの変化を「指さし方向の投影線を基準としたボタンの方向 (式(8), 式(9))」に対応させる. ボタンが指先の投影線の右領域に存在する場合はバランスを右に, 左領域に存在する場合は左にシフトさせる. 尚, 指さし方向の投影線上にボタンが位置する

場合には左右同音量の出力とし, 更にその旨を知らせる別の信号音を付加する. 利用者は音が聞こえる方向に指先を向けることでボタンの方向を把握することができる.

また, 信号音には断続的な音を使用する. この断続音の時間的間隔を「指先の投影点とボタン間の距離 (式(3))」に対応させる. 指先の投影点とボタン間の距離が大きい場合は信号音の間隔を長く, 小さい場合は短くする. 更に, 「指先の投影点とボタン間の距離」が 0 に近い場合, 信号音を別のものに切り替え, 指先の直下にボタンが存在することを利用者知らせる.

4. 実験

4.1 実験環境

実験機器の構成を図 7 に示す. 図 7(a) に示すタッチパネルを用いて構築したシステムの検証を行った. 実験に使用したカメラは, PointGreyResearch 社製 Bumblebee2 ステレオカメラ (図 7(b)) で, 解像度 1024×768 , フレームレート 20fps で左右の同期したカラー画像を取得することができる. また, 実験に用いた PC は Intel Core2Duo(3.00GHz) を搭載し, OS には Microsoft 社製 WindowsXP を用いている. マイクとヘッドホンには両者が一対となったヘッドセットを利用した. 尚, ヘッドセットは数多く普及する一般的なものを用いた.

画像処理に関する多くの部分は Intel 社が公開している OpenCV ライブラリを利用し構築を行った. 信号音は MIDI を介して生成し, また, 音声入出力の部分には WindowsXP に付属する音声認識・音声合成エンジンを利用した. この環境で, ボタン検出からボタンの 3 次元座標算出までの平均処理時間は 47ms, 指の検出から信号音生成に必要なパラメータ算出まで平均 63ms であった. 全工程が一巡する際の処理速度は 150ms 前後であった.



(a)

(b)

図 7 実験機器の構成

4.2 実験手順

図 7(b)に示すように、視覚障害者の操作を想定するため被験者には目隠しをした。タッチパネルの全景が映る位置でカメラを体に固定した後、被験者にタッチパネルの操作を開始するよう指示を出した。被験者はタッチパネル上の0~9の数字、及び*、#の記号の中からひとつを選択し、音声入力によりシステムに伝える。システムが数字、或いは記号を正しく認識した瞬間を開始時刻として正しいボタンを押すまでの時間を計測し、これを操作の所要時間とした。誘導の成否の判断は対象となるボタンを押すことができた際に成功とし、他のボタンを押した場合やシステムが通常とは異なる挙動を示し誘導が不可能となった場合に失敗とした。尚、音声入力を行った際に生じた誤認識は失敗にカウントせず、被験者が操作を最初からやり直すことで、指先の誘導処理のみを評価した。

4.3 実験結果

支援システムのモニタ出力を図 8 に示す。モニタ出力に描かれた指先の記号、独立して描かれた9の数字、dist、dir のパラメータはシステムが導出した誘導のための情報を示している。指先に描かれた記号は、指先の位置及び指さし方向を示す。独立して描かれた9の数字は操作するボタンとその位置を示す。左上表記 dist に示されたバーグラフは指先の投影点とボタン間の距離を示し、dir は指さし方向の投影線を基準にしたボタンの方向を“<-”（左），“○”（中央），“->”（右）で示す。尚、モニタ出力はシステムの挙動確認用のみ設置したものであるため、被験者には一切見えていないことを明言しておく。

モニタ出力からは指先とボタンの相対位置情報が正しく取得できていることが判る。図 8(a)~(f)は、時系列順に並んでおり、指先をボタンの位置まで誘導し、最終的にボタンを押下している様子が伺える。

図 8(a)の状態は、指さし方向に対して目標とするボタンが左にあり、この場合、被験者はヘッドホンの左側より信号音が聞こえるので、信号音に従い左に指を動かす(図 8(b))。指を左に動かし続けると、指さし方向の直線上に目標のボタンが合致し、信号音はヘッドホンの両側から聞こえ、更にその旨を知らせる別の信号音が付加される(図 8(c))。この状態で指さし方向に指を前後させると、断続的な信号音の時間的間隔が長短の変化を示すので、この間隔が短くなる方向に指を進める(図 8(d))。指を進めると指先は目標とするボタンの上部に到達し、指先

の直下に目標のボタンがあることを知らせる信号音がヘッドホンから出力される(図 8(e))。被験者はこの信号音を聞き指先を真下に降下させることで目標とするボタンを押下することができる(図 8(f))。

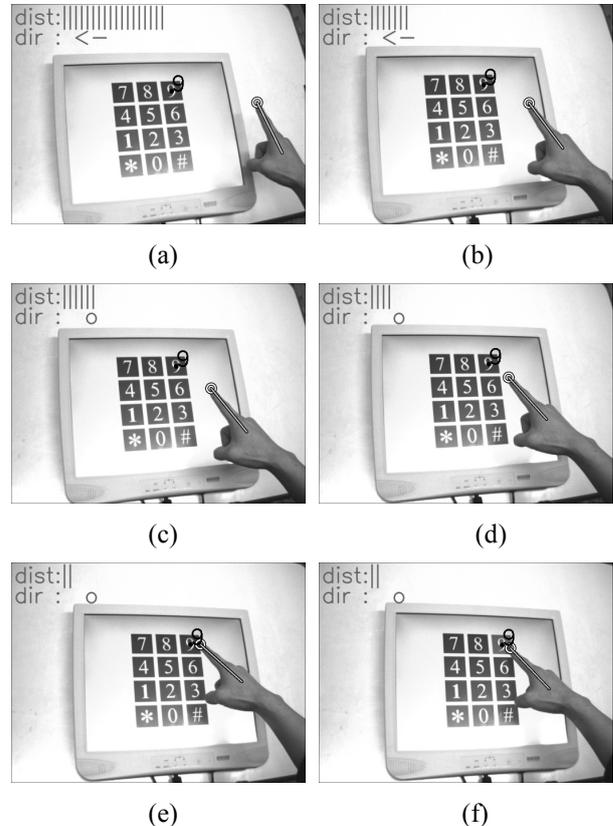


図 8 モニタ出力

9人の被験者が各々10回ずつ操作した結果を表 1 に示す。全試行回数 90 回のうち操作が成功した回数は 70 回、成功率は 78%となった。また、成功時における操作 1 回の平均所要時間は 22 秒であった。

表 1 実験結果

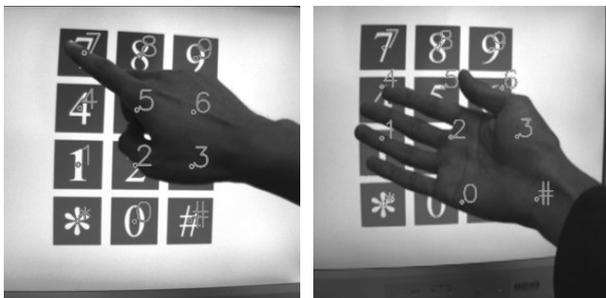
全試行回数	90 回
成功回数	70 回
成功率	78 %
成功時における平均所要時間	22 秒

失敗としては、手によって対象のボタンが隠れるオクルージョンにより、文字の検出ができずに誘導不能となる例が見られた。また、誘導に時間がかかった例として、指とボタンとの位置・姿勢関係を基準とした座標系が被験者の頭の中で描く座標系とは完全に一致せず、被験者が信号音を聞きながら幾度も修正を繰り返すため、誘導に無駄な動きが生じる事象が見られた。

いくつかの課題が見られたものの、視覚に頼ることなく行ったタッチパネル操作が成功率 8 割弱を収める良好な結果を得た。

5. 考察

今回用いたオクルージョンの軽減策では、タッチパネル上の文字が手によって隠れてしまう問題を十分に解決することはできなかった。本研究で用いたオクルージョン対策の成否例を図 9 に示す。タッチパネル上のボタンの数字に重なって表示されている数字はシステムが検出した各ボタンの位置である。成功例では、手によって隠れた部分のボタン位置が正しく推定されている。一方、失敗例では、隠れたボタンの位置が歪な形で広がり、ボタン位置の推定結果が正しくないことがわかる。射影変換行列の導出は四角形を形成する 4 点以上の対応した位置情報がなければ正しく行うことができない。従って、失敗例のような 7, 8, 9 を結ぶ直線の端点と * で形成される 3 点（或いはそれ以下の点数）の構成ではオクルージョン対策とはなり得ない。パネル上の文字の位置は狭い範囲に集約しているため、すべてのボタンがカメラの死角に入ることもあり得る。従って、オクルージョン対策はパネル上のボタンの文字以外の情報も利用し解決しなければならない。



(a) 成功例

(b) 失敗例

図 9 オクルージョン対策の結果

本研究では、指とボタンの位置・姿勢の情報を利用してタッチパネル支援システムを構築した。しかしながら、人が感じる位置・姿勢と実際の位置・姿勢との間には微妙にずれが生じるため、支援システムの利用者はシステムが指示するとおりの指の動きを正確に辿ることはできない。今後の展望として、システムが人の動きを見て、その人の座標系を汲み取り修正を加えることがよりよい方法であるかの確認を行う。また或いは、人の座標系に依存せず、タッチパネルの座標系に従ったほうがよいかの確認も行う。タッチパネルディスプレイには外枠があり、こ

の外枠は指の触感覚により把握することができる。また、ATM 等であれば大きな本体にタッチパネルディスプレイが埋め込まれており、本体の向きは容易に把握できるので、タッチパネルディスプレイの設置方向も容易に想像がつく。これら外枠や本体等の周囲状況を触感覚の知覚により把握すれば、タッチパネルに対して常に正面に構えることが可能で、その正面に構えた際の座標系を基準とした方がより扱いやすい支援システムとなる可能性もある。

6. 結論

本研究では、ステレオカメラによる 3 次元計測と信号音による誘導を用いることで、人の視覚に頼ることなくタッチパネルの操作が可能であることを確認した。

今後の課題として、手によってタッチパネル上のボタンの文字が隠れてしまうオクルージョン問題に対して、タッチパネル周辺の情報を利用することによりその対策の強化を図ることが挙げられる。また、操作の基準となる座標系を利用者の座標系に近づくよう修正を加えつつ誘導を行う、或いはタッチパネルを正面からとらえた座標系を用いるといった試みにより、利用者にとってより扱いやすい支援システムの追求を行う。これら課題を解決することでより強固なタッチパネル操作支援システムの構築が可能となる。

参考文献

- [1] 鈴木悠司, 平岩裕康, 竹内義則, 松本哲也, 工藤博章, 山村毅, 大西昇: “視覚障害者のための環境内の文字情報抽出システム”, 電気学会論文誌 C, Vol.124, No.6, pp.1280-1287, 2004.
- [2] 田中誠, 後藤英昭: “視覚障害者用のウェアラブルな文字認識デバイス”, 情報処理学会研究報告 コンピュータビジョンとイメージメディア, Vol.2008, No.115, pp.125-130, 2008.
- [3] 河井良浩, 小林真, 皆川洋喜, 宮川正弘, 富田文明: “3 次元仮想音響による視覚障害者支援システム”, 電気学会論文誌 C, Vol.120, No.5, pp.648-655, 2000.
- [4] 笹山典江, 山本和彦, 加藤邦人: “視覚障害者のための携帯型文字認識システムの提案”, 第 12 回画像センシングシンポジウム予稿集, pp.478-481, 2006.