

# 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

藤井 崇志<sup>†\*</sup> 長澤 篤<sup>†</sup> 小池 英樹<sup>†</sup> 佐藤 洋一<sup>‡</sup> 中西 泰人<sup>†</sup>

<sup>†</sup> 電気通信大学大学院 情報システム学研究科

〒 182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: {fujii,atsushi,koike,naka}@vogue.is.uec.ac.jp

<sup>‡</sup> 東京大学 生産技術研究所

〒 153-8505 東京都目黒区駒場 4-6-1

E-mail: ysato@iis.u-tokyo.ac.jp

あらまし 現在、情報システムの発展に伴って我々の利用する情報表示環境が大型化されつつある。例として大型のディスプレイのようなものや、また街角では壁に映像を投影したものやディスプレイを埋め込んだものをよく見かける。このような大型の情報表示環境に対する入力手段として頭部姿勢を使うことを考えた。その際にユーザに負担をかけないように頭部にデバイス等を装着しない方法に注目する。しかし従来の頭部位置・姿勢検出手法では検出できる頭部姿勢範囲が狭く、大型の情報表示環境には向かなかった。そこで本報告ではより広範囲でユーザの頭部姿勢を検出する手法について紹介する。また、大型情報表示環境に対しての頭部位置・姿勢を使ったインタラクションについての提案を行う。

キーワード: HCI, Perceptual User Interface, テンプレートマッチング, ポインティング

## Detecting positions and postures of the head for large monitors and its applications

Takashi Fujii<sup>†</sup> Atsushi Nagasawa<sup>†</sup> Hideki Koike<sup>†</sup> Yoichi Sato<sup>‡</sup> Yasuto Nakanishi<sup>†</sup>

<sup>†</sup> Graduate School of Information Systems,

University of Electro-Communications

1-5-1 Chofugaoka Chofu Tokyo 182-8585, Japan

E-mail : {fujii,atsushi,koike,naka}@vogue.is.uec.ac.jp

<sup>‡</sup> Institute of Industrial Science,

The University of Tokyo

4-6-1 Komaba Meguro-ku Tokyo

153-8505, Japan

E-mail : ysato@iis.u-tokyo.ac.jp

**Abstract** Due to recent advances in information systems, large display monitors are developed and in use. For example, there are information displays which use the window or the wall of buildings. In such situation, the traditional input method such as keyboard or mouse is not appropriate. We previously proposed a method for tracking human's head posture without attaching any devices or markers on his/her face. However, our previously proposed method allows users to move his/her head in relatively narrow range of angle. This paper describes an extension of the method to detect wider range of angle, and then proposes to use this method for interaction with such large display.

**Keywords** : HCI, Perceptual User Interface, Template Matching, Pointing

---

\*現在, 株式会社富山富士通所属

## 1 はじめに

近年、コンピュータの発展による情報量の増加・複雑化に伴って、我々の利用するコンピュータの情報表示媒体としてのディスプレイの大画面化が必要とされている。情報の提示方法について注目した場合、情報量の多さや大人数のプレゼンテーションを行うときなどは、広い画面に多くの情報を提示することが有用である。こういった大きな情報表示環境では、より多くの情報提示を行うことができるが、その反面、ユーザの欲しい情報をいかに提示するかといったことも検討される必要がある。提示する情報がいつも一定であったり、決まったパターンであると、ユーザが望む情報を得ることは難しいだけでなく、非効率的である。そこで、ユーザが望む情報をシステム側で提示するようなシステムが期待される。このようなシステムを実現するためには、ユーザの意図をシステム側で推定する必要がある。システム側でユーザの自然な行動を観測して、意図や状況を理解するような状況依存型インターフェース:PUI(Perceptual User Interface)が提案されている。PUIではユーザの行動における表情や見ぶり、音声、視線などの利用があるが、その中でも視線は特にユーザの関心や意図と関連が深い。見ることはユーザにとって自然な動作であり、負荷が少なく疲れにくいといったことその他、大きな画面に対しても高速な移動が可能であり、視線情報を用いることは有効であると考えられる。本報告では視線情報については、ユーザの頭部の姿勢が向けられている方向を視線方向として考える。これは眼球運動も考慮に入れて正確な視線方向を求めようとする場合は固視微動の影響などが大きく、安定に視線方向を検出することができないからである。また、我々は小さな画面でものを見る場合は眼球移動によって見ているものを切替えるが、本報告のような大型情報表示環境を想定した場合は主に頭部を動かしてものを見るが多く、頭部姿勢方向と視線方向がほぼ一致しているとみなす。視線を検出する研究については多く行われているが、あらかじめ何かのデバイスを装着したり、前準備が必要なものでは、拘束感があり誰でも使うことはできない。そこでユーザが何も装着しない非接触型で、受動的に視線を検出するようなシステムがよい。このような非接触でユーザの視線を検出できるようなシステムでは、主に画像処理によって、顔特徴点を追跡する手法 [1] を用いている。更に特徴点をシステム側で時動的に検出する手法 [2] を北島らは用い、ユーザの頭部姿勢を視線情報としてデスクトップ作業への応用を行っている。しかしこの手法 [2] だけで

はユーザの頭部姿勢を検出できる範囲に問題がある。デスクトップ程度であればその端から端までを見渡すこともできるが、より大きな画面に対しては全てを見渡せるまでの検出範囲が得られなかった。そこで本報告では、より広範囲でユーザの視線方向を検出できるようなシステムの構築を目指す。また、広範囲でユーザの頭部姿勢を検出できるようになったことを利用して、大きな画面や空間的な使用を想定し、机型実世界指向インターフェースをモデルとして視線情報を用いた情報提示とインタラクション手法を提案する。

## 2 頭部位置・姿勢検出手法

本章では画像処理を用いたデバイス等の装着のない非接触な頭部位置・姿勢検出手法について説明する。

### 2.1 システム構成

システム構成は図1のようになっており、2台のカメラが取り付けられたカメラユニットと画像処理用PCを用いる。またカメラはあらかじめキャリブレーションしてあり、画像中の3次元位置が計算できるようになっている。

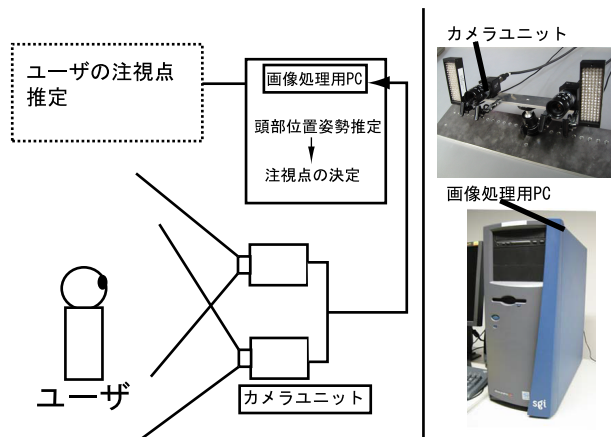


図1: システム構成

### 2.2 処理の流れ

処理の流れは、まずユーザの顔特徴点（ここでは目尻と口元）を検出しテンプレートとして登録したあと、図2のようなテンプレートマッチングによって特徴点を追跡してユーザの頭部位置・姿勢を求めるといったものである。また、2台のカメラの視差を利用したステレオ視の原理を用いて、ユーザの顔特徴点位置から3次元顔平面モデルを構築し、この平面からの法線を

求めてユーザの顔方向ベクトルとしている(図3)。注視点はこの顔方向ベクトルの延長上と対象物との交点で求めている。処理速度は1秒間に30フレーム程度であり、リアルタイム処理が可能である。しかしこの処理だけでは検出可能な頭部姿勢に問題があるので、次節で述べるような拡張のための処理を行った。

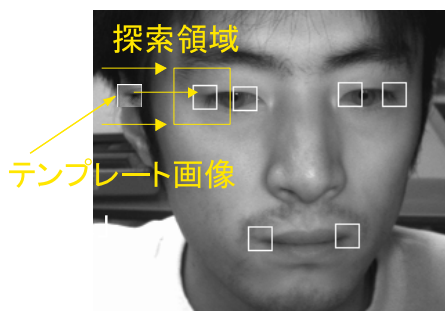


図 2: テンプレートマッチング

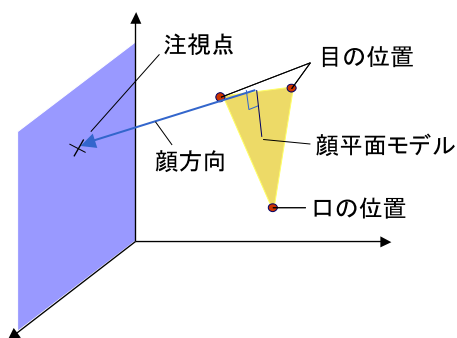


図 3: 顔平面とその法線ベクトル

### 3 検出範囲拡張処理

#### 3.1 更新型テンプレートの使用

本手法ではテンプレートマッチングによって顔特徴点を追跡している。ここで用いるテンプレートとは顔特徴点検出部で求めた特徴点である。この特徴点は図4上のように、正面を向いた状態でテンプレートとして登録される。このため、ユーザの顔の向きが横を向いた状態になると、特徴点の見え方が変化してしまい(図4下)、画像上で特徴点を見つけることが難しくなってしまう。よって、そのまま処理を続けると相関値が低くなってすぐにトラッキングが外れてしまうことになる。

そこでこの問題を解決するために更新型のテンプレートを使用する。これは初期登録されたテンプレート

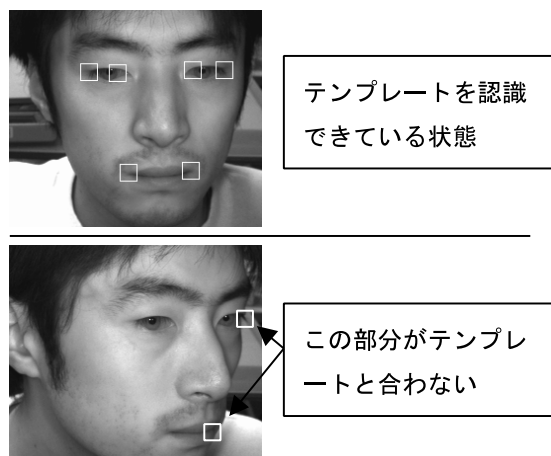


図 4: 顔が横を向いたときの様子

を用いたテンプレートマッチングによって得られた最も相関値の高い場所の画像を新たにテンプレートとして登録し、次のフレームでのテンプレートとして用いるというものである。しかし更新型テンプレートを用いる場合、常にテンプレートを更新し続けると追跡点が徐々にずれていく。この状態が続くと、更新されたテンプレートは初期登録したテンプレートと異なった画像となってしまい、トラッキングが外れるだけでなく、たとえ追跡し続けることができたとしても正確な顔特徴点を追跡していることにはならなくなる。よって更新型テンプレートを用いる場合は、初期登録したテンプレートの情報も保持する必要がある。そこで本手法では、初期登録したテンプレート情報は保持しつつ、状況に応じて新しくテンプレートを登録・更新し、テンプレートマッチングに適用する処理を行う。今まではマッチング処理は2台のカメラ画像をそれぞれ独立で行っていたが、どちらか片方の入力画像に対しては初期登録したテンプレートを用いてテンプレートマッチングを行い、他方の入力画像に対しては更新された横向き状態でのテンプレートを用いてテンプレートマッチングを行う。これは常に初期登録したテンプレート画像をどちらかの処理で用いることによってテンプレート画像がずれていくのを防ぐためである。また、更新するテンプレート画像は初期登録したテンプレートによってマッチングした部分の画像を別カメラの入力画像に対して更新されたテンプレートとして用いる(図5)。

このときに左右どちらの入力画像に対して初期テンプレートを使用するかは、これはユーザの頭部姿勢が右を向いているときは右カメラ画像に対しては初期登録したテンプレートを使用し、得られた追跡点を

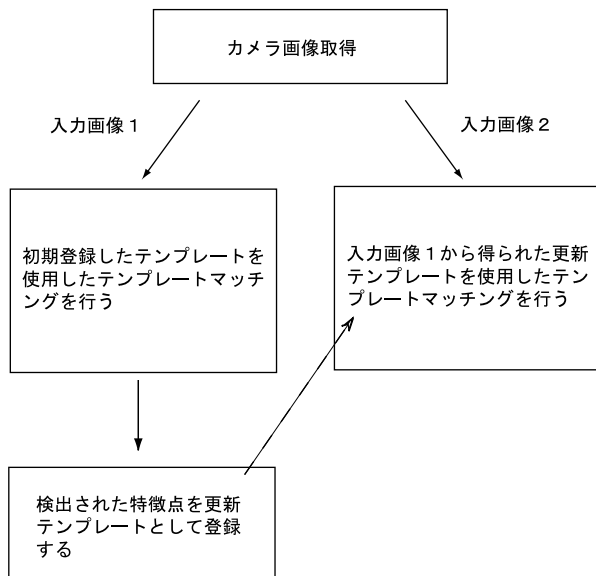


図 5: 更新型テンプレートの使用

左カメラ画像に対してテンプレートとして使用する。同様に左を向いているときは左カメラ画像に対しては初期登録テンプレートを使用し、得られた追跡点を右カメラ画像に対してテンプレートとして使用する。これは図 6 に示すように左右カメラで入力画像の見え方が微妙に異なり、ユーザの顔の向きが徐々に右を向いていくと、右カメラ画像の方が長く右目尻部分をとらえ続ける (特徴点の見え方が初期テンプレートに近い) ことができるからである (図 6)。



図 6: 左右カメラによる見え方の違い

### 3.2 顔方向に応じた使用テンプレートの区別

ユーザがどちらの方向を向いているかは 3 次元顔モデル計算時の顔方向ベクトルによって判断する。よってここでいう顔方向は前フレームでの顔方向のことであり、これをもとに現フレームでの更新テンプレート使用の区別を行う。また、顔方向ベクトルの角度よ

り図 7 のようにユーザの頭部の向きをそれぞれ右向き状態、左向き状態、正面状態として分ける。各状態においてテンプレートマッチング時に使用するテンプレートを区別する。また、正面を向いた状態ではテンプレートの更新を行わず、左右両方のカメラ画像に対しても初期登録したテンプレートを用いることにしている。頭部姿勢の向きによる使用テンプレートは表 1 のようになっている。

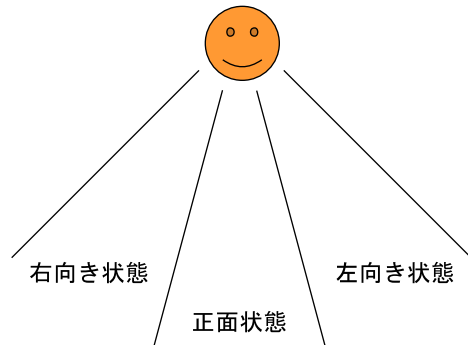


図 7: ユーザの顔の向きへの区別

表 1: 顔の向きに応じたテンプレートの使用

方向	右画像	左画像
右	初期テンプレート	更新テンプレート
左	更新テンプレート	初期テンプレート
正面	初期テンプレート	初期テンプレート

### 3.3 顔方向を用いたフレーム間補正

更新型テンプレートを顔方向に応じて使用することによって、頭部の姿勢検出範囲を拡張することを試みた。しかしユーザの 3 次元顔モデル計算時のフィッティング処理の性質上、計算誤差やテンプレートマッチング時の誤差などが起こる場合がある。更新型テンプレートを使用する場合、顔の向きに応じて更新テンプレートを使用しているため、ユーザが右を向いているのに左を向いた状態だと判断して逆方向の更新型テンプレートを適用してしまうと、マッチング処理の信頼度が急激に下がり、次フレーム以降でトラッキング処理が大きく外れてしまう危険性があり、またトラッキング精度にも多少のばらつきが見られることがある。そこでこのような計算誤差による顔方向ベクトルの誤検出を補正する。テンプレートマッチングで使用するテンプレートは、図 7 のように前フレームで



得られた顔方向をもとに、現フレームで決定される。ここでは一旦、一通りの処理を行い、現フレームでの顔方向までも求める。このとき、顔方向が左向き状態から右向き状態になるには、顔方向は左 正面 右、同様に右向き状態から左向き状態になるには右 正面 左と連続的に変化しているとし、顔方向が左向き状態から正面の状態に変化することはあっても、左向き状態から右の状態に変化することはないとして現フレームにおける顔方向と前フレームにおける顔方向を比較する。その結果、フレーム間で顔方向が異なる場合（顔方向が正面を向いた状態は除く）はエラーとして検出結果を省く。この処理によって、現在の顔の向きに適した更新型テンプレートを使用し続けることができる。ただし、急な顔の旋回もあると思われるので、エラー発生時の次フレームでは初期登録テンプレートを使用することにする。この処理によって3次元顔モデル計算時のエラーを省くことができ、テンプレート更新もスムーズに行うことができると考えられる。

### 3.4 検出範囲拡張処理の評価

頭部位置・姿勢検出システムにおいて、更新型テンプレートの使用と、顔方向ベクトルを利用したフレーム間補正による検出範囲拡張を行った。そこで検出範囲拡張前と拡張後において、それぞれの頭部姿勢検出可能範囲がどの程度までなのかを計測し、評価する。

#### 3.4.1 評価方法

計測方法は以下のとおりである。

- トラッキングシステムにおいて頭部をゆっくりとトラッキングが外れるまで傾けていく。
- トラッキングが外れたフレームの前フレームにおける頭部姿勢の横方向の角度を計測する。
- この動作を右向きと左向きについてそれぞれ10回ずつ行い、平均を求める。

このようにしてトラッキング可能な頭部姿勢の横方向の角度を求めた。

#### 3.4.2 実行結果

結果を表2に示す。

この結果より、ユーザの頭部姿勢は拡張処理前に比べて検出範囲が広範囲になったことが分かる(図8)。また、表2において、拡張処理前後ともに左方向への

表 2: 横方向における検出可能角度

	検出範囲拡張前	検出範囲拡張後
右向き角度	17.7 度	34.2 度
左向き角度	18.5 度	36.8 度
合計視野角	36.2 度	71.0 度

角度が若干広がっているが、これはカメラキャリブレーション時の座標形がわずかに左側に偏ったためではないかと思われる。この角度の拡張は、大きな情報表示環境において視線情報を用いる際には重要である。例えば、1メートル離れた場所において視野角が36度から71度に広がった場合、左右方向を見渡せる範囲は65cmから140cmになる。従来のようなデスクトップ作業を想定した場合は、デスクトップから1メートル離れていたとしても、普通のモニタであればそのほとんどは全域を視線でカバーできると思われる。しかし、同じ1メートル離れた状態でも、ディスプレイが大きなものであった場合、例えば50インチプラズマディスプレイの場合は左右幅が110cm程度あり、その全域を視線で見渡すことは不可能である。しかし、検出範囲を拡張して71度程度の視野角が得られたので、左右幅は140cm程度まで見渡すことができるように、なり、大型情報表示環境の例として50インチプラズマディスプレイを対象とした場合でもその全域をとらえることができるようになった。これにより、次章で提案するような大型情報表示環境においても、有用であると期待される。

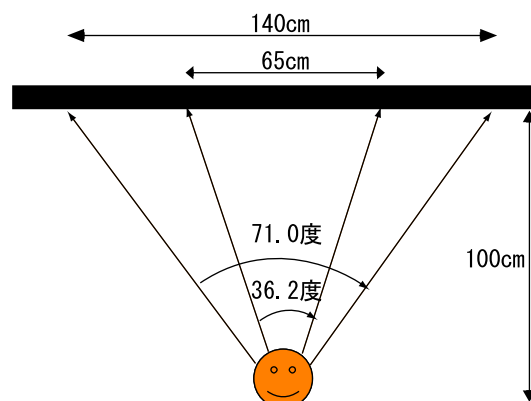


図 8: 頭部姿勢検出範囲の大型環境への拡張

## 4 応用例

本章では、検出可能な頭部姿勢の視野角が広がったことを利用して大型情報表示環境おいての視線情報利用例を提案する。

### 4.1 ユーザの視線による情報提示ディスプレイ

大型情報表示環境に対し、ユーザが見ている部分の情報をピックアップして提示するようなシステムのプロトタイプとして、図9のように50インチプラズマディスプレイ上にテクスチャオブジェクトを映し出した。これはユーザが視線を向けると、そのオブジェクトがピックアップされ拡大表示される。

拡大されたオブジェクト



図9: ユーザがオブジェクトを見ている様子

#### 4.1.1 受動型の情報提示

このシステムでは、ユーザが画面に映し出された絵を見ているだけで、見ている絵がピックアップされて拡大表示されるといったものになっている。そのため、見たあとでユーザが能動的に行動を起こすといった操作は必要なく、何気なく見ているだけで情報が提示されるといった受動的なシステムとなっている。

#### 4.1.2 注視によるユーザの意図判断

また、ユーザが1つのオブジェクトを見続けている時間を計測し、一定時間以上見続けているものにはユーザの興味が大きく反映されていると考え、更に拡大表示される。今回構築したものは更に拡大されるだけであるが、表示するコンテンツによっては詳細情報を表示するといったものや、コンピュータにおけるファ

イル構造のような階層構造をたどることができるようなものでもよいと思われる。

#### 4.1.3 想定する利用環境について

これらはユーザが見ることによって何気なく興味を示したものに關する詳細情報や関連情報を表示するといったことにより、室内の壁に映し出された情報を提示するだけでなく、公共の場などに設置することにより、街を歩く人が足を止めて壁に映し出された情報に何気なく目を向けると、更に細かい情報を提示するような案内システムや広告システムなどの使い方を想定している。最近では駅構内などにも大型ディスプレイが設置されたり、店頭においてプロジェクタでショーウィンドウに映像を投影している店舗もよく見かけ、街角でも大型の情報表示環境が出現しつつある。このシステムでは実際に壁に投影しても図10のように同様のことが行えるので、将来的にも大型情報表示環境への応用が期待される。



図10: 壁に投影した場合の様子

## 4.2 視線と手の動きを併用したポインティング操作

視線方向によってユーザのしている方向を求めることはできるが、受動的ではなく能動的にユーザが操作を行う際のポインティング操作について、視線と手を併用することの提案について述べる。一般的に、PUI環境においてはマウスなどのデバイスを常に持ち歩くことや設置の必要なく、操作ができるような環境が望ましいと考えられる。これに対し、本稿の頭部位置・姿勢検出システムは非接触で、デバイス装着の必要がないことから、PUI環境に対して有効であると考えられる。そこで、大画面におけるポインティング操作

について、手による直接操作を実現している机型実世界指向インターフェースをモデルとして考える。

#### 4.2.1 手による直接操作に伴う問題点

手による操作は直感的であるというメリットはあるが、手の届かないところや手を認識できないところが選択できないといった問題がある。また、手を机上の端から端まで動かす操作の場合は疲労が伴うことや、操作に対して常に手を拘束されてしまうといった問題などが挙げられる。これに対し、視線情報には素早い操作が可能で、ユーザの「見る」行為による入力のため受動的であり、能動的な操作をあまり必要とせず疲労も少ない。その反面、おおまかな範囲でしかポインティングを行うことが難しい。また、固定することが難しいといった特徴がある。これらの特徴をふまえて、ユーザの手の届かない領域の選択を視線情報と手による操作を併用する手法を提案する。

#### 4.2.2 システム概観

手による直接操作が可能な環境として机型実世界指向インターフェース EnhancedDesk において操作を行う。EnhancedDesk の概観は図 11 のようになっており、上部に取り付けられた赤外線カメラによってユーザの手を認識することができる。また、上部（または下部）に取り付けられたプロジェクタによって机上に情報を投影することができ、ユーザは手による直接操作によって直感的にかつインタラクティブな操作を行うことができる。この EnhancedDesk においてユーザの顔領域を撮影できるようにカメラユニットを設置し、頭部の位置・姿勢を計測してユーザの注視点を検出できるようにしてある。

ユーザの手の形状や指先位置の認識は EnhancedDesk 上部の赤外線カメラによる画像を処理することによって行う。図 12 の (a) は赤外線カメラより得られた画像である。これを 2 値化処理したものが (b) で、(c) のように処理範囲を決めて手の認識処理を行う。詳しい処理方法はここでは省く [3]。

#### 4.2.3 視線との併用について

手による操作の場合は、直感的な操作が可能なことと、細かい操作も可能である。そこで、おおまかな領域の指定の操作を視線で行い、細かい操作、微調整を手で行うものとする。視線操作によるポインティン

赤外線カメラ



カメラユニット

図 11: EnhancedDesk の概観

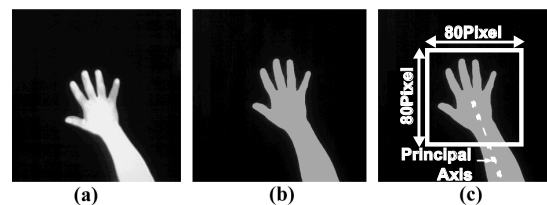


図 12: 赤外線カメラ画像による手認識の様子

グでも十分に選択できる場合は、そのまま視線による選択を行う。この選択操作におけるモードをポインティングモードとしてモード 1, モード 2 として次のように定義する。

モード 1: ポインタ位置は指先位置と等しい。これは手が届く範囲で、また視線によるポインティングではターゲットが小さすぎて選択できないときに使用する。

モード 2: ポインタ位置は視線方向による位置となる。その後、指先を机に出すことによって、指先の位置から相対的にポインタ位置が動くように切り替わる。

このポインティングモードを切り替えながら、ターゲットを 10 個選択するというタスクを行った (図 13,14)。また、これはポインティングについてのタスクであり、決定動作にあたるセレクションについては、ポインティングカーソルが 2 秒間ターゲット上に滞在し



たら選択されるということで行った。また、このタスクをターゲットの大きさを3種類に変更して行った結果以下のような知見を得た。

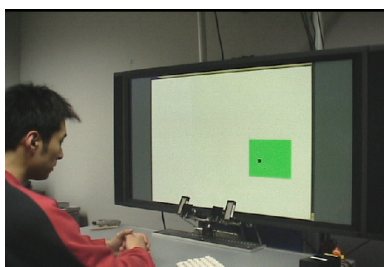


図 13: ポインティングの様子 1

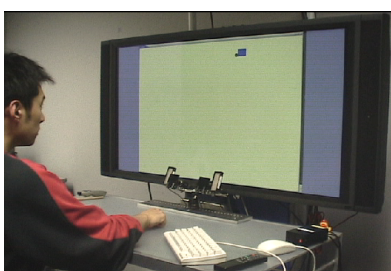


図 14: ポインティングの様子 2

- ターゲットが十分に大きいときは、視線によるポインティングの方が早い
- ターゲットがある程度小さくなると、視線のみでポインティングを行うことは難しい
- 手のみによる選択操作ではターゲットの大きさに係わらず、コンスタントな結果になる(ただし、ターゲットが小さくなると若干時間がかかる場合があった)
- 手と視線を切替えることができるような場合、ターゲットが大きいときは視線のみで選択を行う傾向がある
- モード2においては切替え操作(指先を机に出す)に時間がかかる

これらの操作に関しては筆者が行った場合、視線と手を併用したときが一番素早く操作ができた。しかし、他の人に同様の操作を行ってもらったところ、一番素早い操作ができるわけではなかった。これはモード2におけるポインタ位置の切替え操作に対する慣れの問題だと思われる。よって今後は得られた知見をもとにモードの切替えや、ターゲットの大きさと視線精度の関係などについても調べる必要がある。このこと

はPUI環境においてユーザがデバイスを利用しないで操作を行なうための1つの方法として期待される。

## 5 おわりに

本報告では将来的な大型情報表示環境を想定したときに、視線情報を用いることが有効であると考えた。そこで非接触でユーザの頭部位置・姿勢を検出するシステムにおいて検出可能な範囲を拡張しこれに対応できるような処理を行った。その結果、検出可能な頭部姿勢の視野角が広がったことを確認した。また、大型環境を想定した視線情報の利用例を、ユーザの視線による情報提示ディスプレイと、視線と手の併用による操作方法として提案した。しかし頭部位置・姿勢検出部においてはまだ精度や検出可能な範囲が十分というわけではなく、今後の課題としては更に大きな情報表示環境における検出精度・範囲、ロバスト性の向上などが挙げられる。また、具体的なアプリケーションの構築や評価なども挙げられる。更に、今回のシステムでは同時に1人のユーザの頭部位置・姿勢しか検出できないので、協調作業時における複数ユーザへの対応なども検討していきたい。

## 参考文献

- [1] Y.Matsumoto, Alexander Zelinsky "Real-time Face Tracking System for Human-Robot Interaction," Proceedings of 1999 IEEE International Conference on Systems, Man, and Cybernetics (SMC'99), pp.2-830-2-835, (1999).
- [2] Kotaro Kitajima, Yoichi Sato, and Hideki Koike, "Vision-based face tracking system for window interface: prototype application and empirical studies," Extended Abstracts of 2001 ACM Human Factors in Computing Systems (SIGCHI 2001), pp. 359-360, April 2001.
- [3] 小林貴訓, 佐藤洋一, 小池英樹, "Enhanced Desk のための赤外線画像を用いた実時間指先認識インターフェース," 日本ソフトウェア科学会 Workshop on Interactive Systems and Software (WISS'99), pp. 49-54, インタラクティブシステムとソフトウェア VII 近代科学社, 1999.
- [4] 大野健彦 "視線を用いた高速な目標選択方法," 信学技法 HIP98-33, pp.9-16, (1998).
- [5] A.Zelinsky and J.Heinzmann. "Real-time Visual Recognition of Facial Gestures for Human Computer Interaction," In Proc. of the Int. Conf. on Automatic Face and Gesture Recognition, pp.351-356, (1996).
- [6] 青山晃治, 山村毅, 大西昇, 杉江昇 "一台のカメラによる顔と視線方向の推定," 信学技法 PRU95-233 (1996-03), pp.131-136, (1996).
- [7] Masatake Yamato. "Button Selection for General GUIs Using Eye and Hand Together," Proc. AVI'2000, pp.270-273, (2000).
- [8] Just, M.A. and Carpenter, P.A. "Eye fixations and cognitive processes." *Cognitive Psychology* 8, pp.441-480, (1976).