

# 画像を用いた新しい情報機器インターフェースの研究

富長 博\*

## 1. はじめに

日本は今後高齢化社会を迎え多くの介護者が必要になるとされている。高齢人口(65歳以上)は2000年には17%を占め3.7人で一人を支えるようになる。益々少子化の傾向にある人口比率では介護者の物理的、精神的負担を益々増加させるであろう。それを物理的にサポートする技術は多く研究されている。しかし精神面のサポートを肩代わりするような技術の研究は多くない。研究例としてはVirtual Realityを用いて入院患者を擬似的に外出させる、などがそれに当たる。情報機器を用いた精神的なサポートのための技術開発は今後重要な研究課題になるとと思われる。

情報機器、コンピュータは今や多くの分野に応用されている。家庭内応用という視点では24.3%の家庭にコンピュータが導入されているがその中の27.3%の家庭では使用されておらず眠っている。その原因はキーボードなどのインターフェースの不自由さ、すなわち直接操作感の欠如が一因であることは明らかであろう。家庭用ゲーム機がインターフェースに工夫を凝らしパソコンを大きく上回って購入されている(任天堂のファミリーコンピュータ1992年累計出荷台数6000万台、PC98シリーズ1992年累計出荷台数500万台)こと、がその根拠であろう。

このような状況に加え近年コンピュータユーザーは単に情報の収集、編集、出力だけを望んでいるのみではなく、できれば潤いのある生活を支援してほしいと願っていると推測される。それは年齢を問わずパソコン通信で交信相手などと有意義な会話を楽しんだり、音のするコンピュータなどいわゆるマルチメディアパソコンが急増していることから知られる。

このような状況をふまえ本研究においては新しい情報機器のコンセプトとして、ユーザーに潤いのある生活をもたらすような機能を持つことを提案する。潤いのある生活とは提案する情報機器によって気分転換や精神の安定を促進することが可能になることを意味する。そのためにマインドメタファと呼ぶ基本機能を提案する。マインドメタファの目標とする機能は

- ① 情報機器の操作が未学習でも使えるインターフェース
  - ② 見てすぐに理解できる情報機器内の仮想空間との情報交換
  - ③ 音声を用いた相互の情報伝達
- などである。

この報告書では主に①について述べる。以下においてはまず2.において現状の手振りで必要とされる機能の内どの程度表現できるかを検討しその優位性を述べ、3.において手振りインターフェース実現のための手振り認識の方法の一手法を述べ、4.においてその実験結果を示す。

## 2. 手振りの表現できる単語

現在もっとも多くの単語表現できる手振りは手話であろう。手話は聴覚障害者が健常者と同等に会話できるように設計されているため種類、量ともに多い。手話は比較的覚えやすいように連想が簡単になるように設計されている。しかしながら操作になんら知識を持たないユーザーに全ての手話を覚えさせるのは一般的なパソコンの使用法習得より困難な作業となろう。

一方、人間は手話を使わない手振りによる意思伝達方法を身につけている。また外国語を習わずに海外旅行できたという体験談は手振り、身ぶりの有効性、世界での共通性を示す好例であろう。世界における手振りを参考文献<sup>2)</sup>にて調査したところ、入院患者等に必要と思われる感情的な表現については11個の手振りが掲載されていた。その世界における共通度を探ってみた。










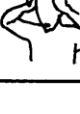

| 手振り   | 一般的な解釈 | 感情的な意味  | 共通度   | 手振り   | 一般的な解釈 | 感情的な意味  | 共通度   |
|---|--------|---------|-------|---|--------|---------|-------|
|    | 静かに    | 静かにしてくれ | 1.0   |    | よくない   | だめ、よくない | 0.425 |
|    | 強い     | 元気      | 0.95  |    | 勝利     | やったー    | 0.3   |
|    | 考える    | 困った     | 0.95  |    | さようなら  | さようなら   | 0.2   |
|   | 食べる    | 食べたい    | 0.95  |   | いやだ    | いやだ拒絶   | 0.075 |
|  | 臭い     | 臭い      | 0.65  |  | 怒っている  | 怒っている   | 0.05  |
|  | OK     | OK      | 0.475 |   |        |         |       |

図1 手振りと共通度

参考文献<sup>2)</sup>では、ある手振りについて国別に20カ国におけるその意味解釈を載せている。ある手振り(手の形)がある意味に解釈されているが国数をA、その手の形を大まかにいくつに解釈できるか、その数をBとすると、その手の形がその意味に解釈されるであろう共通性の高さ(共通度、C)を

$$C = \left(\frac{A}{20}\right)\left(\frac{1}{B}\right)$$

として計算した。その結果を図1に示す。この結果から全てにおいて共通する手振りは少ないということがわかる。しかし意味は違って使われている手振りの形には共通性があり、多くの国において手振りをインターフェースに採り入れれば操作習得に時間がかからない情報機器が生まれるのではないかと考えられる。ここで、ある手振りに対して今まで使ったことのない意味を割り付ける場合とキーボードを使ったことがない人がある意味を情報機器に伝えようとした場合を考える。キーボードで行う場合相当の試行錯誤が必要となる。しかし通常使っている手振り動作であればキーボードの操作方法習得に当たる作業は必要ないので、少ない時間

で手振りによる意思の伝達が可能になるものと思われる。

このように考えて行くと、パーソナルな情報機器における意思伝達では通常使っている手振りに対してはそれに相当する意味を割り付けて良いのではないかと、ということになる。特に緊急性を要するような意思の伝達については伝達のしやすさ(=インターフェースの使いやすさ)が優先されるべきである。

### 3. 手振りインターフェースについて

#### 3.1 手振りインターフェースの研究例

前節の結果よりマインドメタファにおいては手振りインターフェースを第一の入力デバイスとして位置づける。本研究で使用される手振りインターフェースは情報機器に不慣れなユーザーが機器と向かい合う形で利用することを想定している。

例として病院内において使われる情報機器について述べる。このような環境において使用される本研究の手振りインターフェースは

- ・慣習的な動作による意思の伝達ができる。
- ・感情を伝達できる。
- ・病室で利用できるようにするため可能な限り非音声の形態とする。

などの点でマインドメタファ用のインターフェースとして適している。

手振りインターフェースは学習の必要がなく、非音声であり非接触入力のため非常に使い易いインターフェースである。そのためこのようなインターフェースについては手話の認識を含め多くの研究事例がある。動きに基づいた認識と静止画と用いた場合のモデルとの照合によるパターン認識が代表的な例である。動きに基づいた研究としては高橋らのスポッティング認識による動くに認識がある<sup>4)</sup>。この場合は人の動作の時系列を一固まりのパターンとして認識している。静止画を用いた研究としてはユニパンらの隠れマルコフモデルを用いた動作の認識がある<sup>5)</sup>。この研究では手の振りなど体の形が画像上でどうなっているか人体のスティックモデルを用いて検出している。その他手話の認識が多く研究されている<sup>6)</sup>。これらの多くは背景を特殊なものとして、2値画像にした手の形の計測できるパラメータから手話を判別したり、手にマーカーを付けて手の形状を再現している研究が多くある。

本研究では上述のように①病室の環境下で手振りを認識する。②また子供から高齢者まで様々な人が利用する。この場合動きによる認識を行おうとするといつも決まった動きをせねばならないので不便である。「静かに」という動作を認識する際に手を頭から口に動かしてきても、下から動かしてきても認識するようにならなければならない。また人体のモデル照合だけでは本研究では不十分で、意味を解釈することが必要である。

#### 3.2 提案する手振り認識手法

まず考えられる手振り認識手法としては濃淡画像でのパターンマッチングがある。これは既存の画像処理手法で主に位置決め、文字認識などに用いられている。動作機能は単純で

登録された濃淡画像を対象画像の中から実時間で見つけ出すというものである。機能的には単純であるが実時間で動作させるためにハード的に工夫されている。このような装置を手振りの認識に使用できるかどうか検討した。その結果

- ① 違った人の違った手振りを登録せねばならない。
- ② 照明光の影響、影の変化の影響が考えられる。などの理由で適していないと思われる。

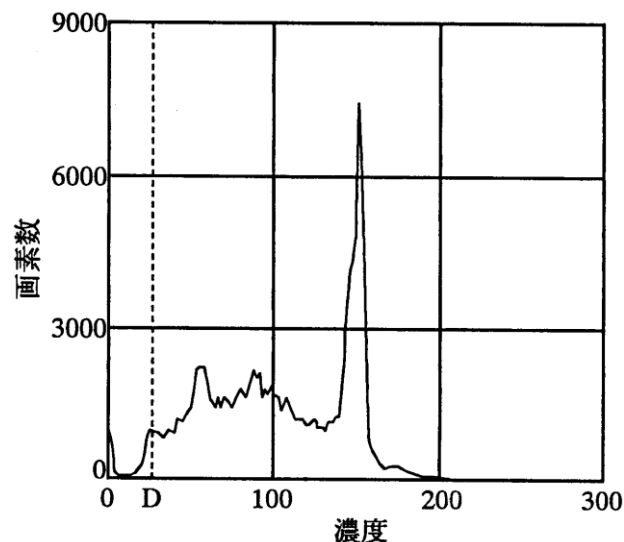


図2 目抽出のためのヒストグラム

次に考えられる方法としては、まず顔を認識してそれから各人体パーツを認識していく手法である。顔はまず目から認識し次に鼻、口などをその明暗から順次検出していく。標準的な顔であると仮定して前述の顔のパーツ位置から顔全体の大まかな位置が判明する。その後それらの位置から肩の位置が判明し画像上で腕の動かせる範囲が決定される。その範囲内での腕、手と思われる人体パーツの候補領域を抽出しあらかじめそれらを認識するパラメータを格納した手振りデータベースとの照合を候補領域との間で行い手振りの認識を行う。といった手法が考えられる。以下本稿ではこの手法を採用する。

この手法の中の顔領域の抽出法方について述べる。

(「顔パーツ抽出アルゴリズム」)

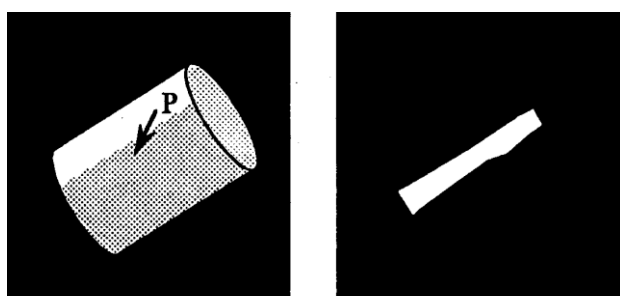
- ① まずユーザーの上半身の背景画像(ユーザーを含む)について濃淡のヒストグラムを作成する。
- ② ヒストグラムのグラフで暗い方から二番目のピークはこの画像の中で信号のない部分以外でもっとも暗い部分と判断する。
- ③ この明るさをしきい値にして、それ以下を2値化して抽出すると目、髪の毛などが抽出される。
- ④ さらに目の部分のみを抽出するために大領域削除、領域の外接長方形の対角線の角度による抽出、対称となる領域を有する領域抽出処理を行う。

このような処理によって抽出された二つの目を結ぶ線の垂直二等分線上に鼻、口はある。鼻、口はそこに手を当てて表現する手振りについて検出するときに位置を知る必要がある。その垂直二等分線上の明るさのプロファイルを検討して、口はもっともその中で暗い、次に暗い部分で目と口の間

にあるのは鼻の下というようにして鼻、口の位置を検出する。このようにして顔のパーツ、ひいては顔の位置を検出できる。実際に計測したヒストグラムの例を図2に示す。図ではDが二番目のピークを持つ濃度である。

次にこの手法の中の腕、手と思われる候補領域の抽出方法について述べる。ここでは腕、指は円筒形状と仮定する。ユーザーは上述の環境下にあるためその下にある腕、指はだまかには一方向からの光で照らされる。具体的には上方からの直接光を仮定し、その反射光は室内各所から腕、指に到達するがその光は直接光よりは弱いものとなるということの意味する。一方向から腕、指は照らされるため円筒の側面に相当する腕、指のある一面が強く照らされ照明光の反対側の側面は暗くなり、その間は明るさが徐々に変化してゆく。その明るさは円筒の円周方向に直角の方向には一定になる。これはその方向の面の法線方向が同一のためである。そのような性質を利用すると円筒の傾き（円筒傾斜角と呼ぶことにする）、すなわち腕、指のだまかな傾きが計測できる。このため計測アルゴリズムは以下のようにする（「腕、手領域抽出アルゴリズム」）。

- ① まず一つの画素Pの明るさとある許容範囲で同じ明るさの画素領域を抽出する（図3）。
- ② その後その画素を中心として仮想的に画素領域の長い方向を見つけるための探索スティックを置き、それを回転させながら探索スティックが抽出された画素領域に最も重なる方向を見つける。その方向を円筒傾斜角候補とする（円筒傾斜角が円筒の角度と等しくなるとは限らないため）
- ③ そのように抽出された円筒傾斜角候補の値を画素Pの位置に明るさの値としてプロットする。角度 $0^{\circ}$ を濃度75にその後の値は $1^{\circ}$ あがるごとに濃度を1プラスした値とした。濃度0はどの方向も同じに検出され方向性がうまく検出されない場合とした。円筒傾斜角候補 $\theta$ は $0^{\circ} \leq \theta < 180^{\circ}$ について求めた。



a.原画像                      b.2値化画像

図3 円筒表面2値化による円筒傾斜角検出用領域の抽出の概念図

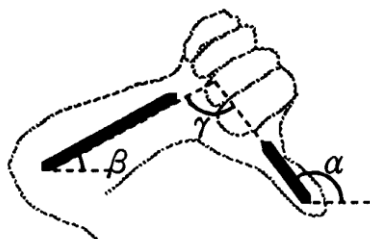


図4 腕、手の円筒傾斜角（ $\alpha$ 、 $\beta$ ）と相対角度（ $\gamma$ ）

このようにすると腕、指の部分が一つの同じ濃度領域となる。さらに腕、指の領域を絞るためにデータベースの中で重要である腕、手の間の相対的な円筒傾斜角度候補を元にして領域を抽出する。そうすることで単に一つの同じ濃度領域を抽出するより、画素間の関連付けができるのでデータベースとの照合が早くなる。ここでは例えば「良くない」という意味に使われる親指を下に向ける手振りを例に説明する（図4）。ここで大事な関係は親指とその他の部分（腕、指）の垂直な関係である（図4では $\gamma$ ）。そこである画素にたいしてその画素と垂直な関係の部分の画素領域を求める。この後意味のないと思われる小領域を膨張→小領域削除という処理で取り除く。ここまで腕、手の候補が絞られれば後はデータベースとの照合作業のみとなる。

本手法の特徴としては明るさを何段階かにわけてそこで抽出された領域の結合、分離作業を知識ベースに基づいてせずに、腕、手領域関係記述データベース照合まで持ち込める点があげられる。それは直接腕、指の領域まで絞り込めるためである。このようにすることで体全体の形状がわからなくても腕、指（手）の存在位置は顔の位置関係、面積などから判明し意味をそこから判断できる場合もあり得るという特徴がある。

#### 4. 手振り認識のための実験と結果

目、鼻、口の検出実験を図5.aと図6.aに対して行った。目の検出結果を図5.bに鼻、口の検出結果を図6.bに示す。鼻と口は点で示されておりその回りにある白い枠は鼻、口検出のための明るさプロファイルを作成するとき横方向の値を平均して作成した時の平均する範囲を示している。

円筒傾斜角による腕、手の検出実験では親指に対して垂直な部分を抽出した。図7が元の濃淡白黒画像、図8が円筒傾斜角を濃度に表した画像である。背景はテキストなどの方向性を持っていなかったため黒く（濃度0）になっている。しかし頭の周辺など背景にも関わらず円筒傾斜角が存在すると抽出された。腕、手の指についてほぼ一つの領域として抽出されている。一方で親指の部分が背景（服のしわ）と同じ円筒傾斜角と抽出されたので区別がつかなかった。図9は図8の親指の部分に対して $90^{\circ}$ の円筒傾斜角を持つ画素を抽出した結果である。このように腕と指の部分が抽出された。

#### 5. おわりに

本研究ではまずマインドメタファという心の安定をもたらすサービスをする情報機器の機能を提案した。またその機能の一部として手振りによるインターフェースをその認識手法とともに提案した。

これらの提案された概念、手法が実用的なものであるかどうかはこの情報機器のプロトタイプを用いた第三者による評価が最もよい方法でありその評価方法についても検討する予定である。

情報機器に慣れた人がこれから多くを占めるようになる一方であらゆる面で（例えばいじめ、テクノストレス、高齢者の一人暮らしなど）心理面におけるサービスが必要になるが

人口が減りそれに関わる人を多くさけないこれからの社会ではマインドメタファのような機能を持つ情報機器が必要になると思われる。

この報告ではプロトタイプ製作までは言及できなかったがプロトタイプ製作が非常に重要になるので今後はその製作に向けての研究を続ける予定である。

### 参考文献

- 1) 福祉士養成講座編集委員会：“老人・障害者の心理”、中央法規出版(1991)
- 2) 金山宣夫：“ノンバーバル事典”、研究社出版(1983)
- 3) 廣瀬通孝：“バーチャルリアリティ”、オーム社(1995)
- 4) 高橋勝彦、関進、小島浩、岡隆一：  
“ジェスチャー画像のスポットティング認識”、  
信学論(D-II)、J77-D-II、8、1552-1561 (1994-08)
- 5) クンラボン ユーニパン、木下宏揚、酒井善則：  
“スティックモデルを用いた手振りの認識”、  
信学論(D-II)、J77-D-II、1、51-60 (1994-01)
- 6) 長嶋祐二、藤井昌紀、長島秀世：  
“カラー画像による指文字に関する基礎検討”、  
テレビジョン学会技術報告、Vol117No58、pp.19~24 (Oct.1993)



図7 「良くない」という手振り



a. 目抽出のための原画像      b. 目の抽出  
図5 目の抽出



図8 円筒傾斜角の明るさ表現画像



a. 鼻、口抽出のための原画像      b. 鼻、口の抽出  
図6 鼻、口の抽出

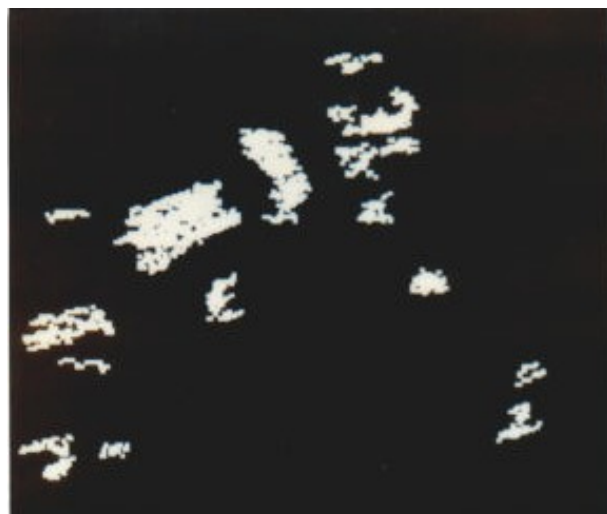


図9 絞り込まれた領域