ヒューマンインタフェース 75−10
（1997. 11. 14）

知的なインタフェースのための視覚機能の実現

～動画像を用いた顔認識システム～

田中 英治　山口 修　福井 和広　前田 賢一

(株) 東芝 関西研究所
〒658 神戸市東灘区本山南町 8-6-26
E-mail:e-tanaka@krl.toshiba.co.jp

あらまし　本稿では、知的なヒューマンインタフェースに重要な視覚機能を付与する顔認識手法について報告する。顔認識機能を持ったシステムの具体的事例を挙げ、認識方法における問題点を明らかにする。これまでユーザの正面顔を識別する方法として部分空間法を用いていたが、HIの観点からは表情や顔向きの変化の影響の少ない手法が必要であり、動画像を用いた相互部分空間法による顔の識別法を採用した。さらに、システムの小型化のためにPC上での実装を行ない、表情や顔向きの変化による影響と識別性能について評価を行なった。

キーワード　ヒューマンインタフェース、顔画像認識、PCシステム、相互部分空間法

Realization of Visual Function for Intellectual Interface

- Face Recognition System using Temporal Image Sequence -

Eiji Tanaka ,Osamu Yamaguchi, Kazuhiro Fukui and Kennichi Maeda

TOSHIBA Kansai Research Laboratories
8-6-26 Motoyama-Minami-Cho, Higashinada-ku,
Kobe 658 Japan
E-mail:e-tanaka@krl.toshiba.co.jp

Abstract　This paper presents a face recognition method providing important visual function to intellectual interface. We first introduce some concrete examples of face recognition system , and clarify the problems. "Subspace Method" has been used as a method to recognize a front view of a user's face , but from the viewpoint of human interface, it is neccessary to have a method that is less sensitive to the variation of expression or facial posture. We then adopt a face recognition method by "Mutual Subspace Method" usingimage sequence. Further,we implement the method to the system on PC for the purpose of practical use, where the influence of variation of expression or facial posture is estimated and the performance of face recognition is evaluated.

Key Word : Human interface , Face recognition , Personal computer system , Mutual subspace method

1 はじめに

ヒューマンインタフェース (HI) をより高度で知的なものとするためには、ユーザを含む周囲の状況を理解する能力がシステムに求められる。この能力により、システムはユーザの状態やユーザ（システム）周辺の状況に対応し、ユーザはより少ない負担と訓練でより自分に適ったサービスを提供してもらえるようになる。システムが人間のように状況を理解するためには、視覚、聴覚などの複数の感覚チャネルと、その認識機能をシステムが備えている必要がある。特に、視覚はユーザが発する個人性、表情、身振りなどのノンバーバル言語の多くを非接触で比較的広い距離範囲で受容でき、重要かつ有用な感覚であると言える [1][2]。

視覚機能はこれらの情報を動画像を処理することで常時取得可能であり、非接触の状況理解の基本となる機能であると位置付けられる。画像情報を用いて、ユーザの行動、個人性を認識するための負担をかけない方法としては、顔画像認識が挙げられる [3]。顔画像を用いて、人物の検出、追跡、識別を行うことにより、ユーザの存在／接近／離脱や移動、またユーザが誰であるのかを示す情報を得ることができ、システムはユーザに特化した対応が可能になる。

我々はこれまで部分空間法 [4] による正面顔の顔画像認識方式を用いて、いくつかの対面型のシステムの上でのHIにおける視覚機能を実現してきた [5][6][7]。本稿では、それらのシステムにおけるHIの顔認識機能の問題点を整理し、その問題点を克服するための動画像を用いた顔識別方法について述べる。また、近年のパソコンの著しい性能向上と普及により、パソコン上で動作する実用的な顔認識モジュールが今後広く求められるようになると考えられる。そこで、本稿では提案した顔認識方式をパソコン上に実装し、ヒューマンインタフェースへの応用の観点から評価した結果について報告する。

2 HIにおける顔認識

ヒューマンインタフェースにおける顔認識は次の3段階の処理が必要である。

1. 人物の検出段階
2. 人物の追跡段階
3. 人物の識別段階

まず、1. の検出段階は、人間の存在の有無を画像中から顔領域を検出することによって判断する。存在する場合には顔の位置情報も検出することができる。人間が存在する場合に、顔を検出することによって、システムが反応し、人間の手を介さずに自発的な動作をすることができる。

2. の追跡段階では、対象となる人物が適切な大きさで適切な位置に映るようにカメラを制御する。追跡機能によって、人間が動いている場合でも、カメラの向きを人手で操作する必要がなく、また人間がカメラの視野を気にすることなくシステムを利用することができる。

3. の識別段階では、検出された顔領域に対して、その人物が誰であるかを識別する。この識別機能によって、人物に応じた動作が可能となり、人間どうしのコミュニケーションに近いインタフェースが実現される。

これらの各機能の具体的事例を以下に示し、顔認識における問題点を考察する。

2.1 具体的事例における問題点

- 固定カメラによる顔認識

  人間の顔のCGを使ったナレーションエージェント「Rachel」[7] は、状況に応じた動作をさせるべく顔認識の機能を設けている。ここではカメラとして固定カメラを用いている。カメラの視野内に人間が存在するか否かを顔検出の機能で判別し、存在する場合にはその検出された方向にCGの顔を向かせ、ナレーションを開始する。さらに、検出された人物が誰であるかを顔によって識別し、識別結果に応じてナレーションの内容を選択する。このような機能を備えたインタフェースは、人間の代わりとして自発的にユーザに対する動作が出来るという点で優れている。

  このシステムを実験的に運用したところ、カメラが固定されているために、人物を検出できる範囲が非常に狭いという問題点があった。この問題点に対して、カメラを固定せず、一度検出した人物を追跡できるようにすると効果的である。

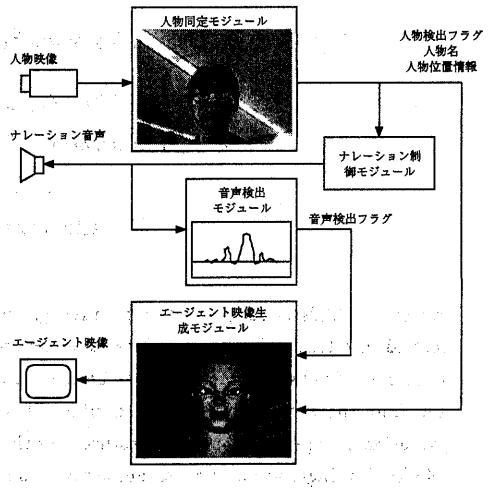

図 1: ナレーションエージェント

- アクティブカメラを用いた顔認識

「人間に優しいロボット」を開発コンセプトとして作成されたビーチバレーロボット「TOMORROW」[6] は、バレーの対戦相手を識別するために顔認識機能を備えていた。ロボットの腕に当たる部分に付けられたズームレンズ付きのカメラは、腕の上下左右の角度を調整することができ、人間の顔を追跡することができる。図2のように、3人の人間がカメラの視野に入った場合、順番に各人の顔が適切な大きさで映るようにズームしながら、画面中央にその顔が収まるようにカメラの向きを変えてフレーミング(位置合わせ)する。この時、顔の位置が動けばカメラによりトラッキングを行う。フレーミングが完了すると人物の識別を行なう。

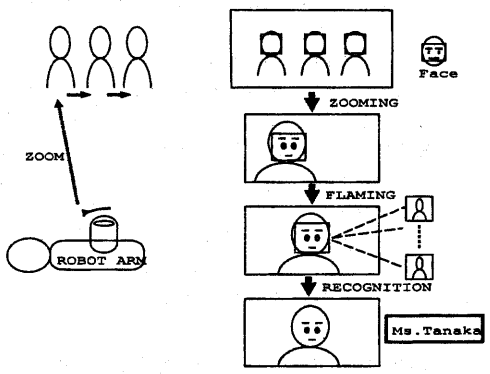

図 2: ロボットによる顔認識

このシステムでは顔を追跡する機能を取り入れることで、人間の立つ位置の制限をなくすことができ、また動きなどに対応することが可能となった。

ここに挙げた2つのシステムでは、いずれも静止画で正面向きの顔を認識することを前提としているため、トラッキングと同時に識別を行なうことは困難であり、さらに顔の向きや表情を変化させた場合には正しく識別が出来ないという問題がある。ユーザが比較的自由に表情や位置を変えるHIシステムでは、人物認識に安定した動作が要求され、人間自身の起こす顔の向きや表情の影響を受けない認識法が必要である。

3 動画像を用いた顔認識

これまでのシステムで問題となっていた識別処理に関して以下に改良法を述べる。顔の向きや表情変化といった変動に対応するため、動画像から得られた時系列画像を用いた顔認識方法を用いる [8]。これは登録時だけでなく、認識の際にも多数枚の多様なデータとして時系列画像を収集して利用し、識別には部分空間法を拡張した方法である相互部分空間法 [9] を用いる。相互部分空間法は、登録パターンだけでなく、入力パターンも部分空間として表現し、登録パターンの部分空間と、入力パターンの部分空間のなす角度を評価して識別を行なう。以下、相互部分空間法について詳しく説明する。

3.1 相互部分空間法

従来の部分空間法 [4] では、g を入力画像、ϕ を登録パターンを主成分分析して得られた各固有ベクトル（辞書と呼ぶ）として、類似度 $S_{normal}(g)$ を以下の式で定義する。

$$S_{normal}(g) = \frac{1}{||g||^2} \sum_{n=1}^{N} (g, \phi_n)^2 \qquad (1)$$

$(\cdot, \cdot)$ は内積を表す。

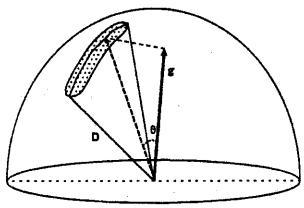

図 3: 部分空間法の説明図

図3は部分空間法の説明図であり、未知入力ベクトル g を部分空間 D に射影したベクトルとの角度、すなわち部分空間との角度から式 (1) のように類似度を定義している。

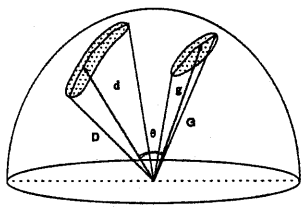

図 4: 相互部分空間法の説明図

一方、相互部分空間法では、入力ベクトルも部分空間で表現し、辞書パターンの部分空間との間の角度を類似度として識別を行なう。図4は相互部分空間法の説明図であり、2つの部分空間 D, G のなす角度 θ の余弦は、

$$cos^2\theta = \sup_{d \in D, g \in G, ||d|| \neq 0, ||g|| \neq 0} \frac{|(d, g)|^2}{||d||^2 ||g||^2} \qquad (2)$$

と定義する（d, g は式 (2) が極値をもつためのそれぞれの部分空間上のベクトルを表す）。これに関して、2つ

の部分空間 D, G への正射影作用素を P, Q とする場合、$cos^2\theta$ は PQP の最大固有値 λ_{max} となる [9]。

$$cos^2\theta = \lambda_{max} \tag{3}$$

辞書パターン部分空間を D、入力された時系列画像に対する部分空間を G とする。ここで ϕ, ψ を各部分空間 D, G における固有ベクトルとする。実際には PQP という行列の最大固有値を求めるのではなく、式 (4) で表される行列 X の固有値問題を解き、その最大固有値を類似度（部分空間間類似度）S_{mutual} とすればよい [9]。ここで、D の部分空間の次元を M、G の部分空間の次元を L として、$L \leq M (1 \leq i, j \leq L)$ とする。

$$X = (x_{ij}) = \sum_{m=1}^{M} (\psi_i, \phi_m)(\phi_m, \psi_j) \tag{4}$$

$$W^T X W = \Lambda \tag{5}$$

(W:X の対角化行列、λ_{max} : Λ の対角成分の最大値)

$$S_{mutual}(G, D) = \lambda_{max} \tag{6}$$

相互部分空間法を顔認識に用いることで、入力の時系列画像の冗長性を利用して、表情や顔向きの変化に対して安定した識別が可能となる。

3.2 顔認識処理の流れ

顔認識処理は、図 5に示すように、(a) カメラから入力された画像中の顔領域の検出、(b) 顔部品（瞳、鼻孔）の特徴点抽出、(c) 位置、大きさなどを正規化したパターンの切りだし、(d) 登録パターン辞書との類似度計算による識別処理、を順に行なって入力画像中の人物が誰であるかを識別する。

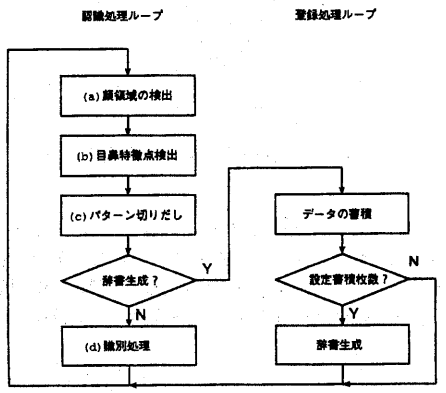

図 5: 処理の流れ図

(a) の顔領域の検出においても、部分空間法を用いる。入力画像から切り出した領域とあらかじめ登録した顔パターンとの類似度を求め、その類似度があるしきい値よりも大きくなる領域を顔領域とする。(b) では画像の分離度とパターン情報を用いて目鼻の特徴点を抽出する (詳細は [10])。(c) の正規化パターンの切り出しでは、目と鼻の位置を基準として顔の正規化濃淡パターンを生成する。(d) の識別処理では、正規化された入力画像の複数のパターンを用いて部分空間を計算し、相互部分空間法により人物の識別を行う。

4 顔認識システムの実装

将来の HI システムの応用場面では、PC による実装が多くなる。今後、画像を用いたアプリケーションは一般的になってくると考え、顔認識システムを PC 上で実装した。

システムは、CPU PentiumPro 200MHz を持つ PC でソフトウエアのみで実現される。画像入力のためのカメラは、図 6下のように、ディスプレイの下側に設置して、顔を撮影する。カメラは、東芝製 IK-C40 CCD カラーカメラ、レンズは焦点距離 6.5mm のものを使用した。入力画像の取り込みには PCI バスを利用した画像入力ボードを用いた。取得画像は、サイズが 320×240 pixel でありカラーであるが、特徴点抽出と識別にはモノクロに変換した濃淡情報のみを利用している。

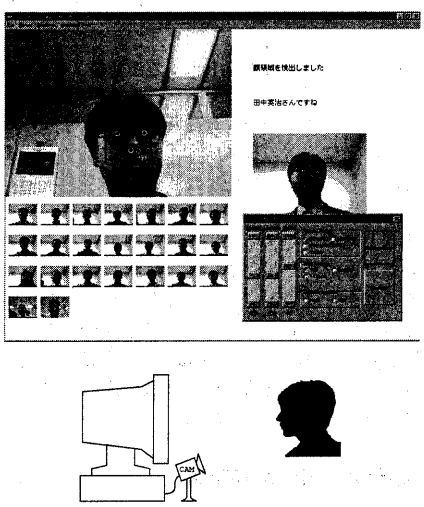

図 6: 顔認識システムの概観 (上：画面表示、下: カメラの配置)

図 6上は、システムの表示画面であり、入力画像の表示と、識別結果の表示、登録されている顔画像の表示を

行なう。

登録については、用意されたコントロールパネルウインドウ(図6上の右下)を操作することにより、辞書生成を指示するだけで図5に示したように、認識処理から登録処理に移行できる。辞書登録では、顔画像パターンを蓄積し、個人を識別するための辞書が生成される。登録処理が終了すると、再び認識処理に移行する。

入力画像の表示部分(図6上の左上)には、顔検出、目鼻検出の様子が表示される。認識が行なわれると、識別結果として登録時に撮影した顔画像を表示する(図6上の右上)と同時に、識別された人物と同じ辞書画像を点滅させる(図6上の左下)。試作システムでは、100人の登録人物に対する照合処理を一秒間に3～4回の速度で実行できる。画面上へ現在の入力画像を逐次表示させない状態では、一秒間に5～6回の処理も可能である。

5 実験

5.1 表情や顔向きの変化による影響

表情や顔向きの変化によって顔の識別性能がどのように変化するかを実験により調べた。

図7,8 は、ある一人の人間が表情を変えたり、顔の向きを変えた時の時系列画像データに対して、その人の登録辞書を用いて部分空間法と相互部分空間法の類似度の時系列変化を比較したものである。破線が従来の部分空間法による類似度、実線が相互部分空間法による類似度である。

図7では正面向きの顔のみ収集して登録を行なった辞書を用いており、入力として、無表情の区間(A)、表情を変化させた区間(B)、顔向きを変化させた区間(C)、の類似度の変化を示したものである。

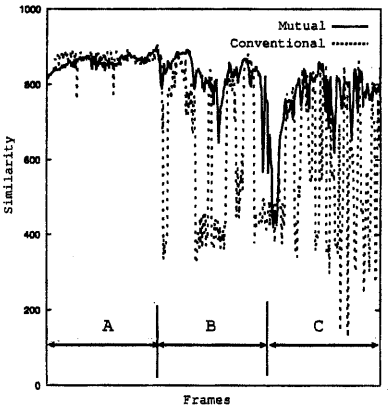

図 7: 無表情の顔の辞書を用いた時の類似度の時系列変化の比較(実線:相互部分空間法、破線:部分空間法)

図7の場合は無表情の顔を登録した辞書を用いているので、部分空間法では、表情のある時や顔向きの変わった時に同じ顔パターンが辞書にないため、区間B,Cでは類似度が低くなっている。

図8は表情を変化させたりや顔向きを変え、様々な種類の顔を登録した辞書を用いて、図7と同じ入力画像系列を与えた場合の類似度の変化を示したものである。

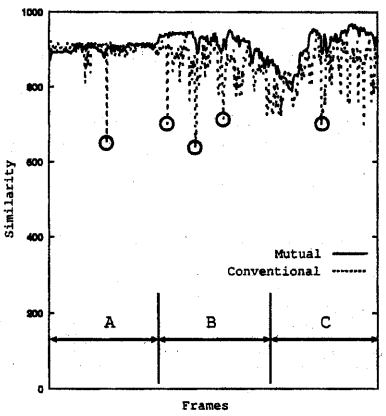

図 8: 表情や顔向きに変化のある顔の辞書を用いた時の類似度の時系列変化の比較(実線:相互部分空間法、破線:部分空間法)

図8の場合は辞書に様々な種類の顔を登録しているため、入力画像に表情のある場合や顔向きに変化のある場合でも、図7と比較して、部分空間法による類似度の変動は少なくなっている。しかし、相互部分空間法に比べ、なお変動が大きく、図8に示した○印のフレームのように類似度が大きく下がってしまうことがあり、誤認識を起こしてしまう。

相互部分空間法を用いた場合には、図7,8 いずれの場合も、表情や顔向きによる類似度の変動を吸収しており、高い類似度を保ったままでいることが分かる。このように人間自身が起こす状態変化に影響を受けにくい認識方法がHIシステムの機能として必要とされる。

5.2 識別正解率の比較

次に、従来の部分空間法と相互部分空間法を識別正解率によって比較した。ここで、正解とは、ある人物のデータを入力した時に、辞書に登録した人間の中から、入力データに相当する人物の類似度が最大となることを指す。また、識別正解率とは、辞書登録している全ての人物のデータを順次入力した場合、識別正解率=(正解となった回数)／(全ての試行回数)で表される数値で、ここでは100名のデータに対して識別正解率を求めた。

従来の部分空間法による識別正解率は、辞書の次元数を変化させた場合、表1のようになった。

表 1: 従来の部分空間法での識別正解率

次元数	15	20	25	30
識別正解率	92.9 %	93.3%	93.5 %	93.7 %

また、相互部分空間法を用いた場合の識別正解率は辞書の部分空間の次元数、入力画像の部分空間の次元数を変化させた時、表2のようになった。相互部分空間法を用いると、従来の部分空間法よりも、識別率が向上していることが分かる。辞書の次元数も小さいため、記憶容量の小型化という点において有利である。

表 2: 相互部分空間法での識別正解率

辞書の次元数	3	3	5	5
入力の次元数	3	5	3	5
識別正解率	96.2%	96.7%	97.0%	97.0%

いずれの方法においても、辞書の次元数が増えるにしたがって識別正解率は上がるが、類似度を求めるための計算時間は、それにつれて増大することになる。PC システムにおいてリアルタイム性が要求される場面ではこの所要時間がネックとなる。応用に合わせた認識精度と処理時間のトレードオフが問題となる。

6 おわりに

本稿では、知的 HI のために必要とされる顔認識方法についての報告を行なった。

従来の静止画像による部分空間法を用いた顔認識では、顔の向きなどによる入力画像の変動に対処するために大量の辞書登録パターンを用いる必要があった。しかし、動的にユーザの状況が変化する HI システムには認識能力が不十分であった。

時系列の入力パターンを用いた相互部分空間法による顔認識を行なうことで、入力画像の変動に対し、少ない辞書データで安定かつ精度の高い識別性能が得られることが分かり、HI システムで顔認識を利用する際にも安定した性能を発揮できることを確認した。

今後は、アクティブカメラシステムと動画像による顔認識システムを組み合わせることにより、より多様な状況で安定した HI システムの視覚機能として発展させていく予定である。

参考文献

[1] 間瀬 健二:" マルチモーダル・インタフェースのための画像処理", 第 2 回画像センシングシンポジウム, pp.123-128(1997)

[2] 黒川 隆夫:" ノンバーバルインタフェース", オーム社 (1994)

[3] 末永 康仁、間瀬 健二、福本 雅朗、渡部保日児:"Human Reader: 人物像と音声による知的インタフェース, 信学論 (D),Vol.J75-D-II,No.2,pp/190-202(1992)

[4] E.Oja:"Subspace Methods of Pattern Recognition", Research Studies Press,England(1983)

[5] 田中 英治、倉立 尚明、福井 和広: " アクティブカメラを使った遠隔教育システムの試作" , 情報処理学会誌,vol.38,No.4,pp.346-347(1996)

[6] 辰野 恭市:" ヒューマンフレンドリロボット", 東芝レビュー 5 月号,pp.104–106(1997)

[7] 鈴木 薫、山口 修、福井 和広、田中 英治、倉立 尚明、松田 夏子: " 人に近いインタフェースを目指して ～擬人化インタフェース Rachel の試作 (1) ～" , 情処学会 HI 研究会研究報告,96-HI-69-7, pp.47-53(1996)

[8] 山口 修、福井 和広、前田 賢一:" 動画像を用いた顔認識システム" , 信学技報,PRMU97-50,pp.17-24(1997)

[9] 前田 賢一、渡辺 貞一: " 局所的構造を導入したパターン・マッチング法" , 信学論 (D),vol.J68-D,No.3, pp345–352(1985)

[10] 福井 和広、山口 修:" 形状情報とパターン照合の組合せによる顔特徴点抽出", 信学論 (D-II),vol.J80-D-II, No.8,pp.2170–2177(1997)