

# 顔画像認識システムの開発

堀 武司, 波 通隆, 大崎 恵一

## Development of a Face Recognition System

Takeshi HORI, Michitaka NAMI, Keiichi OHSAKI

### 抄 録

一般家庭への各種情報機器の普及に伴い、誰にとっても親しみやすいユーザインターフェース技術が求められている。顔画像を用いた個人識別は、ユーザによる操作が必要ないため、対話的ユーザインターフェースにおけるユーザ認識手法として有効である。そこで我々は、様々なアプリケーションへ組み込んで利用できる汎用の顔画像認識システムの開発を行った。本システムは、視野内の人物頭部の追跡、学習用顔画像の登録、および実時間での顔画像の識別を行う事が出来る。13人の成人男性を対象に本システムの評価実験を行った結果、正面顔で98.5%、顔の姿勢変化を含む場合で93.5%の識別率が得られた。

**キーワード：**顔検出、顔識別、対話型インターフェース、高次局所自己相関特徴、多変量解析

### Abstract

As various home information appliances become popular, human-computer interface technologies familiar to everybody is needed. Personal identification by using human face is suitable to the user recognition method for interactive systems since it requires no user's operation. So we developed a general-purpose face recognition system, which can be incorporated into various applications. The system is able to track a human face in the field of view, register face images for pattern learning, and classify faces in real time. In experiments with 13 adult male persons, 98.5% and 93.5% of recognition rates were achieved under frontal and varying facial poses respectively.

**KEY-WORDS :** Face detection, face classification, interactive interface, higher order local autocorrelation feature, multivariate analysis

## 1. はじめに

高度情報化社会の到来に伴い、様々な情報機器やサービスが日常生活の中に普及、浸透してきている。そのため、だれもが使いこなす事ができる、親しみやすいユーザインターフェース技術が求められている。その一つの方向性として、視覚・音声情報処理などの技術を組み合わせ、言葉やジェスチ

ャ、表情などを複合的に用いて人間的なコミュニケーションを目指す「マルチモーダル対話技術」<sup>1)</sup>が提案されている。

コンピュータが人間と対話するために最初に必要となる機能は、対話の相手となるユーザが誰であるかを正しく識別する能力である。現在のコンピュータシステムでは、パスワードやICカードキー、指紋などのバイオメトリクスによるユーザ識別が用いられている。これらの手法は信頼性は高いものの、計測の際にはユーザがキーボードを打鍵したり指をセンサに押し当てるなど何らかの操作が必要であり、自然な対話の流れが阻害される。

一方、ユーザに負担をかけない自然な認証技術として、人

事業名：創造的研究推進事業

課題名：視覚情報に基づくマルチメディアコンピュータ対話システムに関する研究

の顔画像情報を用いた個人識別（以下、顔画像認識）が注目されている。顔の計測はビデオカメラ等によって受動的に行う事が出来るため、ユーザは計測作業を意識する必要がない。コンピュータに顔画像認識能力を持たせる事で、ユーザが端末の前に座っただけで相手が誰かを認識し、相手や状況に応じた適切な応答を即座に行う事が可能となる。

そこで我々は、対話的インターフェースやその他様々な情報システムに組み込んで利用する事が出来る、汎用の顔画像認識システムの開発を行った。

以下、開発した顔画像認識システムの機能と処理手法の概要、識別性能を評価するための実験とその結果について順に述べる。

## 2. 顔画像認識システム

### 2.1 顔画像認識システムの概要

我々が開発した顔画像認識システムは、以下の三つの機能を提供する。

- 1) 画像中の人物の頭部を自動的に追跡し、顔画像の切り出しを行う。
- 2) パターン辞書への顔画像の登録と管理を行う。
- 3) 入力された顔画像を辞書に登録された顔画像パターンと比較し、個人の識別を行う。

顔画像認識システムのハードウェアは一般的なコンピュータ(PCもしくはワークステーション)と画像入力用カメラから構成される(図1)。処理はソフトウェアにより実装され、特殊なハードウェア、センサ等は特に必要としない。

顔画像認識システムの実行画面を図2に示す。画面左上にはカメラから入力された画像、頭部検出の結果を示す矩形枠が表示される。下部には、入力画像から切り出した顔画像が表示される。システムは学習、識別の二つのモードを持ち、それぞれ次の様な動作を行う。

**学習モード：**これから学習させるユーザの名前をダイアログから入力する。その後、パターン学習用として所定の枚数



図1 顔画像認識システムの構成



図2 顔画像認識システムの実行画面

(数十枚程度)の顔画像を自動的に撮影する。撮影が終わると、取得した顔画像データを用いて、パターン辞書の更新を行う。

**識別モード：**カメラの視野内に入ってきた人物を自動撮影し、パターン辞書と比較して個人の識別を行う。識別結果として、一致度が高かった順に数人分の候補のユーザ情報(名前、顔アイコン、ユーザIDなど)とパターンの一致度を画面に表示する。

### 2.2 顔画像認識処理の流れ

本システムにおける顔画像認識処理フローの概略を図3に示す。まず、ビデオカメラ等で撮影された画像の中から人物の顔の位置を検出し、顔を含む部分画像の切り出しを行う。次に、切り出した顔画像から画像特徴量の抽出を行い、パターン認識部へ送る。パターン認識部は、学習モードでは予めクラス分けされた学習用データからパターン辞書を構成し、識

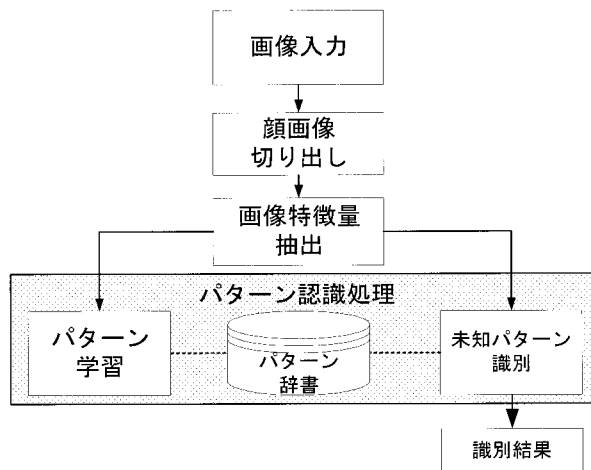


図3 顔画像認識処理フロー

別モードでは入力されたデータをパターン辞書と照合してどのクラスに所属するかを推定する。

各段階の処理にどのような手法を用いるかによって、様々な顔画像認識システムの構成が考えられる。本研究では、背景差分で得られた人物シルエット情報に基づく頭部の追跡処理、高次局所自己相関特徴を用いた特徴量抽出、線形判別分析によるパターン学習・識別手法を用いて顔画像認識システムの構築を行った。

### 2.3 頭部の追跡と顔の切り出し

最初に、入力された画像から頭部の位置を検出し、顔画像領域のみを切り出す。

多くの顔画像認識手法、特に固有顔法に代表される画素値を特徴量として直接使用する手法では、精密な画像の位置合わせが必要であり、切り出し精度の良否が認識性能に大きく影響する<sup>2)</sup>。そのため、これらの手法を用いる場合には、最初におおまかな頭部位置を推定して切り出しを行った後、目鼻口などの顔部品の位置を検出し、アフィン変換などで位置の正規化処理を行う場合が多い<sup>3)</sup>。

しかし、実環境下で顔部品の位置を高速かつ安定に検出する事は困難な問題である。本研究では、後段のパターン認識処理として画像の位置ずれに対して頑健な手法を用いる事により顔の精密な位置合わせ処理を省略し、頭部に外接する正方形領域による大まかな切り出しを行った。

#### 2.3.1 人物シルエットの抽出

本システムでは、背景差分処理で得られる人物シルエット情報を用いて頭部の位置推定を行う。背景差分による物体検出は最も基本的な画像処理の一つであるが、実環境で使用するには、照明の変化やカメラの自動露出の影響による輝度変動、背景自体の長期的な変化などに対する頑健性が要求される。

本システムでは、輝度変動の影響を受けにくくするため、単純に画素単位で差分を取るのではなく、背景の輝度パターンの変化を検出する手法を利用した。まず背景画像および各時刻の入力画像を、 $16 \times 16$ 画素程度の小ブロックに分割する。照明条件の変動などで画像全体の輝度レベルが変動した場合は、個々の画素値は変化するがブロック内の相対的な輝度分布パターンはある程度保存されるものと考えられる。一方、人物などの物体によって背景の遮蔽が起きた場合は、ブロック内の輝度パターンは背景と全く異なるものに変化する。そこで、背景画像と入力画像の間における輝度パターンの変化量を評価する事で、遮蔽の有無の検出を行った。

シルエット抽出結果の例を図4に示す。パターン変化の検出には、輝度変動に対して頑健な尺度である正規化相関を使用した。左の画像は背景画像取得時と同じ条件、右の画像はカメラのゲインを調整して輝度を変化させた場合であるが、ど

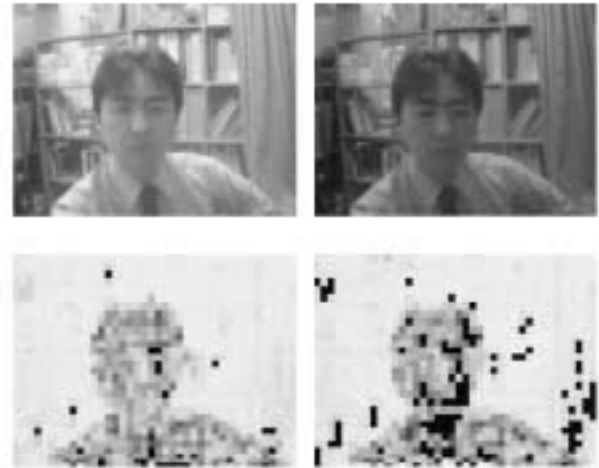


図4 背景差分によるシルエット抽出結果(左:輝度変化なし,右:輝度変化あり)

ちらもほぼ同様な人物シルエットを得る事が出来た。

#### 2.3.2 頭部位置の推定

次に、得られた人物シルエット情報を用いて頭部位置の検出を行う。任意の条件下でシルエットから頭部位置を正確に識別するのは困難であるため、本システムでは撮像時の条件として、

- ・シーン中には、背景を除き人物以外の物体は存在しない。
- ・人物はほぼ直立に近い姿勢である。

という制約を加える。そして画面内から

- ・シルエットの最上部に位置し、左右と上方が開けている。
- ・一定サイズの円形(正方形)領域である。

という条件を満たす領域を人物の頭部領域として検出する。ここで用いた前提条件は常に満たされるとは限らないが、顔認識システムが実際に利用される状況の大部分はカバーしており実用上さほど問題にはならない。

具体的な処理方法としては、図5に示す2重の領域1, 2を持ったテンプレートを用い、それぞれの領域内の画素の輝度分布の差を、分離度  $\eta$

$$\eta = \frac{\sigma_b^2}{\sigma_T^2} \quad (1)$$

$$\sigma_b^2 = n_1(\bar{P}_1 - \bar{P}_m)^2 + n_2(\bar{P}_2 - \bar{P}_m)^2 \quad (2)$$

$$\sigma_T^2 = \sum_{i=1}^N (P_i - \bar{P}_m)^2 \quad (3)$$

によって評価する。

ただし、 $\sigma_T^2$ は全領域の分散、 $\bar{P}_1, \bar{P}_2, \bar{P}_m$ は領域1, 2および全体の輝度平均、 $n_1, n_2$ は領域1, 2の画素数である。

この尺度は分離度フィルタ<sup>2,3)</sup>が用いているものと同じであるが、使用するテンプレートの形状が異なる。背景差分で得られたシルエット画像をテンプレートで走査しながら分離

度を評価し、分離度が最大となる点を頭部位置とする。また、分離度の最大値が閾値を下回る場合は、シーン中には人物頭部は存在しないものと判断する。

本手法によって頭部検出を行った時のシルエット画像、分離度画像、および元画像上での頭部検出結果の例を図6に示す。入力画像のサイズは640×480画素で、そこから128×128画素の顔画像領域を抽出するものとした。テンプレートは領域1が128×128画素、領域2が256×256画素のものを使用した。画像中の矩形の枠線は検出された頭部位置を示す。

本システムではシーン中には人物は一名しか存在しないものと仮定して処理を行っているが、分離度の極大点を個別に検出する事により複数人物の同時検出への拡張も可能である。

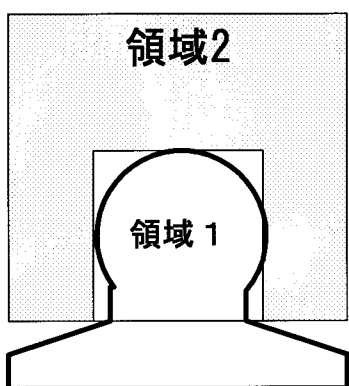


図5 頭部位置検出のためのテンプレート



図6 シルエット画像からの頭部位置推定

## 2.4 画像特徴量の抽出

次に、切り出された顔画像から、画像特徴量の抽出を行う。

2.3節で述べた通り、本システムにおける顔切り出し精度は高くないため、画像パターンの位置ずれに対する頑健性が求められる。そこで、画像特徴量の抽出にはパターンの平行移動に対する不変性を有する高次局所自己相関特徴<sup>1,7)</sup>を使用した。

自己相関関数を高次に拡張した高次自己相関関数は、 $N$

個の変位 ( $a_1, \dots, a_N$ ) を用いて、

$$x_N(a_1, a_2, \dots, a_N) = \int f(r) f(r+a_1) \dots f(r+a_N) dr \quad (4)$$

と表される。自然画像を処理対象とする場合、近接する画素同士の相関が重要であると考えられるので、変位を参照画素の8近傍に制限し、また自己相関関数の次数  $N$  を高々2次までとする。この時、変位の取り方は平行移動で等価となるものを除くと図7に示す25通りとなる。ただし、No.1~10までの変位パターンに関しては、重複する参照を許して0~2次までの自己相関関数を全て求める事で、パターンの種類は35通りとなる。それぞれの変位パターンで画像を走査しながら対応する画素値の積を求め、その値を全画面について足し合わせる事で、一枚の画像から35次元の特徴量ベクトルが得られる。

高次局所自己相関特徴は、 $3 \times 3$ 画素内の局所的な特徴を抽出するため、対象とする画像の性質によっては観測のスケールが適切でない可能性がある。本システムでは、元画像を段階的に縮小し得られた多重解像度画像からそれぞれ特徴量抽出を行い、それらを統合する事で大まかな特徴から細かな特徴までを同時に抽出した。実際には、 $1/2$ づつ4段階に縮小された画像を用い、 $35 \times 4 = 140$ 次元の特徴量ベクトルを得た。

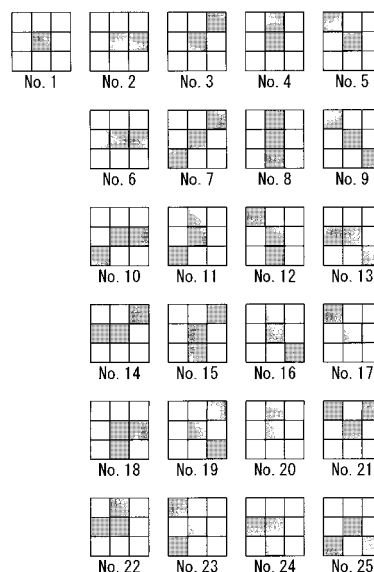


図7 高次局所自己相関特徴のための変位パターン

## 2.5 パターンの学習・識別処理

最後に、得られた特徴ベクトルを用いて、学習用パターンからの辞書の構成、および未知パターンの識別処理を行う。

特徴量抽出によって得られた特徴量ベクトルは元の画像の持つ特徴を反映しているが、これをパターンの比較に直接用いるのは適切ではない。

第一に、原特徴量ベクトルは、目的とする識別に有効な特徴(本システムにおいては個人間の差異を表す特徴)とそれ

以外の特徴（例えば照明や顔向きの変異）が混在しているため、原特徴量同士の比較結果がクラスの類似度に直接対応するとは限らない。また、特徴ベクトルの各変量は正規化されていないため、パターン間の位置関係はスケールの取り方によって大きく変化してしまう。

第二に、原特徴量ベクトルは次元数が大きく、互いに相関の大きい変数の組を多数含んでおり冗長である。冗長な特徴量は計算コストの面で不利である事に加え、与えられた学習パターン数に対して特徴ベクトルの次元数が大きすぎる場合には、「次元の呪い」<sup>6)</sup>と呼ばれる現象によって識別性能が低下する事が知られている。

これらの問題を解決するためには、原特徴量に対して適切な特徴選択と次元削減を行う事が必要である。本システムでは、多変量解析の一種である線形判別分析を用いてこれを行った。

線形判別分析<sup>5,6)</sup>は、予めクラス分け情報が与えられた学習用データ集合を用い、元の特徴ベクトルをクラス判別に適した特徴（判別スコア）へ変換する線形写像を求める手法である。判別スコア空間では、同一クラスに属するデータは互いに近くに、また異なるクラスのデータは互いに遠くなる様に分布するため、判別スコア空間上での距離を用いてパターンの類似度を評価する事ができる。

線形判別分析では、学習データ集合のクラス間共分散行列、およびクラス内共分散行列から定義される判別基準を最大化する（クラス内分散に対するクラス間分散の比を最大化することによって、与えられた学習クラスを判別するのに最適な写像行列を計算する。このとき、得られる判別スコア空間の次元数の上限は、(クラス数 - 1)に制限される。

パターンの学習時には、学習データから判別スコア空間を求め、その上での各クラスの学習データの重心を代表データとしてパターン辞書を構成する。パターンの識別時には、入力されたパターンとこれらの代表データとの距離を求め、最も距離が近いクラスに所属するものとして分類する。

### 3. 識別性能の評価試験

本システムによる個人識別の性能を定量的に評価するための評価試験を行った。

#### 3.1 実験データ

13人の被験者を対象に、実験用の顔画像データの収集を行った。撮影は、蛍光灯照明下の室内でテクスチャのない白色の壁を背景として実施した。被験者が頭部をゆっくりと運動させている所をビデオカメラで録画し、正面および上下左右に向きが変化する顔画像の時系列データを得た（図8）。



図8 実験に使用した顔画像データの例

#### 3.2 正面顔の識別

最初に、正面顔のみで構成されたデータを用いて学習、識別実験を行った。正面顔データは、撮影された顔画像データからほぼ正面向きと判断されるものを目視により選別して作成した。各クラスのデータを時系列順に交互に取り出して二群に分け、一群を学習用データ、もう一群を評価用データとした。また、特徴量には元画像を4段階に縮小した多重解像度画像から得られた特徴ベクトル140次元を全て利用した。

線形判別分析によって得られた判別スコア空間上におけるデータの分布を図9に示す。判別スコア空間上の各クラスのデータが互いに分離して分布している事が確認できる。クラス数は13群であるため、有効な判別スコア空間の次元数は12次元となる。

得られた判別スコア空間を用いて、評価用データの識別を行った。識別には判別スコアの上位6次元までの値を用い、各クラスの学習用データの重心と評価用データとのユークリッド距離をそれぞれ比較して最も小さい距離のクラスに分類した。リジェクト条件は考慮せず、それぞれの評価用データは必ずいずれかのクラスに分類されるものとした。各クラスのデータ数、識別正答率を表1に示す。クラスによってややばらつきが見られるものの、識別正答率95~100%、平均して98.5%の良好な識別結果が得られた。よって、正面顔が安

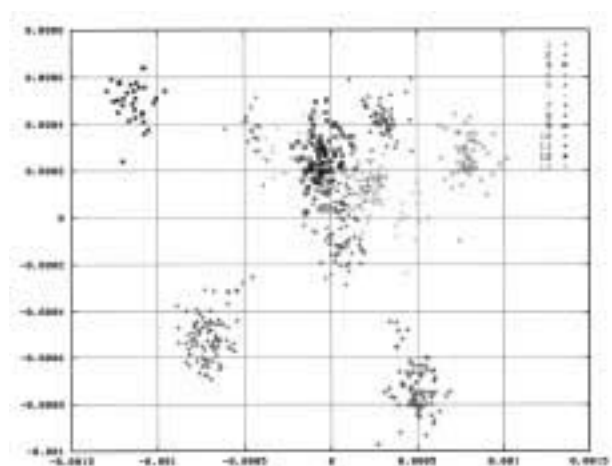


図9 正面顔データの判別スコア空間

表1 正面顔データの識別率

クラス	データ数	識別正答率
No 1	18	0.944
No 2	40	0.975
No 3	29	1.000
No 4	14	1.000
No 5	69	1.000
No 6	34	0.941
No 7	77	1.000
No 8	88	0.977
No 9	60	0.983
No 10	51	1.000
No 11	53	1.000
No 12	31	1.000
No 13	75	0.986
Total	639	0.985

定して取得できる環境では、本手法によって十分な精度の識別が可能であるといえる。

### 3.3 多重解像度画像の効果の検証

特徴量抽出時における画像解像度選択と多重解像度画像の利用の効果を検証するために、解像度の組み合わせを様々に変化させながら学習・識別実験を行った。4段階に縮小された画像（以下、解像度の高い順にP1～P4とする）を単独で用いた場合、および多重解像度処理の段数を増減させた場合について、正面顔の時と同様の手順により実験を行い識別能力の優劣を評価した。それぞれの条件における識別正答率を表2に示す。

単一解像度の画像を用いた場合、P2（64×64画素）が識別率最大となり、それより低い解像度では識別率が低下する結果が得られた。より解像度の高いP1がP2よりも識別率が低いのは、P1を用いた時の観測スケールが顔画像の特徴抽出には細かすぎたためと考えられる。

また、多重解像度画像を用いた場合は、段数が増えるにつれて識別率が改善され、P1、P2、P3の組み合わせで識別率

表2 多重解像度のレベルを変化させた場合の識別率

画像				識別正答率
P1	P2	P3	P4	
○	—	—	—	0.91080
—	○	—	—	0.92175
—	—	○	—	0.88419
—	—	—	○	0.88106
○	○	—	—	0.93114
○	○	○	—	0.98122
○	○	○	○	0.97340

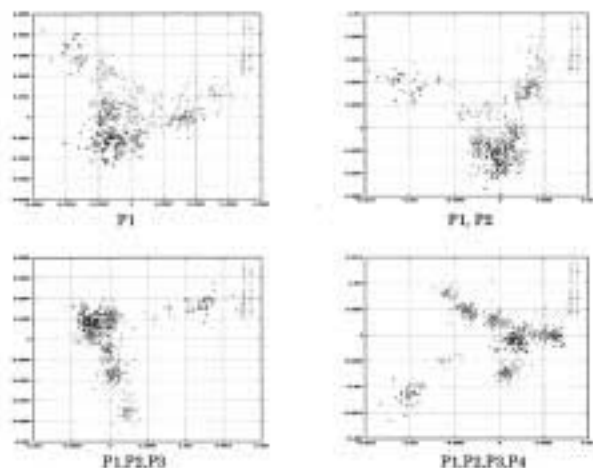


図10 多重解像度のレベルを変化させた場合の判別スコア空間

が最大となり、判別スコア空間上のデータ分布形状（図10）も明らかに改善された。しかし、さらに段数を増やしたP1、P2、P3、P4の組み合わせでは、識別率は横ばいかもしくは若干低下する傾向が見られた。

いずれの場合においても、多重解像度処理を用いた方が単一解像度による結果よりも良い識別性能が得られており、特徴量抽出時における多重解像度画像の利用が識別性能向上に有効である事が確認できた。

### 3.4 顔向き変化を含む場合の識別

顔画像認識の利点の一つは、ユーザを拘束せず自然な状態で計測出来る事である。そのため、カメラを意識して完全な正面顔を呈示した場合だけではなく、ある程度顔向きが変化した場合においても正しく認識が出来る事が望ましい。

本システムでは、入力された顔の姿勢を検知し補正するといった処理は特に行わず、顔向きの変化もパターン変動の一部とみなし様々な顔向きの画像を一括して学習させる事で顔向き変化への対応を行った。

顔向き変化を含む全ての顔画像を用いて学習用、評価用データを構成し、正面顔の場合と同様の手順によって実験を行った。判別スコア空間上のデータ分布（図11）は正面顔の場合よりも各群の分離が不明瞭となり、識別率（表3）も本人受入率が平均87.8%と正面顔の時より低下した。特にクラス9、13など一部のクラスの識別率に大きな落ち込みが目立った。しかし、判別スコアを上位6次元までで打ち切らず、12次元全てのデータを用いて識別を行った場合には、一部クラスの大幅な識別率低下は見られなくなり、平均の識別率も93%台まで改善された。これは、比較的離れたクラスの情報が上位軸に集まり、近接したクラスを分離する情報が下位軸に追いやられているためと考えられる。しかし、クラス数がさらに大きくなった場合には、判別スコアの次元数を無制限に増やす事は難しいため、被験者数を100人程度まで増やし

た大規模実験を行うなどの検証作業が必要である。

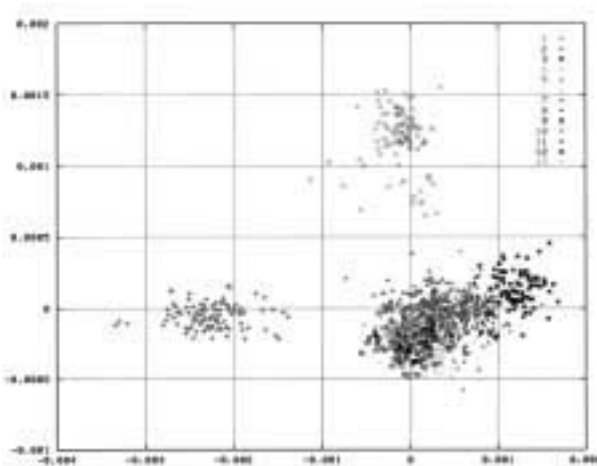


図11 顔向き変化を含む場合の判別スコア空間

表3 顔向き変化を含む場合の識別率

クラス	データ数	識別正答率	
		判別スコア6次元	12次元
No 1	47	0.936	0.936
No 2	97	0.835	0.866
No 3	86	0.942	0.988
No 4	28	0.929	0.964
No 5	125	0.856	0.880
No 6	137	0.876	0.912
No 7	171	0.994	0.994
No 8	189	0.963	0.963
No 9	123	0.756	0.878
No 10	113	0.973	0.965
No 11	104	0.846	0.981
No 12	102	0.912	0.912
No 13	103	0.553	0.913
Total	1425	0.878	0.935

### 3.5 考 察

ここで行った識別実験は、それぞれの画像を単独で評価している。しかし、実際の顔画像認識の場面では頭部追跡の結果から同一人物のものと判っている画像が多数得られるため、これらの情報を統合して評価する事によりシステム全体の認識性能をさらに改善できる可能性がある。

本実験で用いたデータは、学習・評価データ共にほぼ一定の条件下で撮影されているが、学習時と識別時の照明条件等が大きく異なる場合には、識別性能が低下する。しかし、顔向き変化と同様に、各種条件の変動の影響も一緒に学習させる事によって、識別性能の改善が可能である事が確認されている。一般に、顔画像認識システムを使うユーザは同一システムを長期間継続して使う場合が多いと予想されるが、このような場合には環境変動の影響を含んだ新しい学習データを逐

次収集していく事が可能である。

## 4. おわりに

ユーザインターフェース技術などへの応用を目的とする顔画像認識システムの開発を行った。本システムは、人物頭部の自動検出と個人識別処理を実時間で実行できる。また、学習用顔画像の新規登録と更新作業が簡単な操作によりオンラインで可能である。

成人男性13人を被験者とした評価実験では、正面顔に対して平均98%、顔向き変化を含む場合でも93%程度の識別正答率が得られた。これは入退室管理や電子決済など厳密な認証が必要な用途には必ずしも十分ではないが、対話的ユーザインターフェース構築のための簡便な個人識別手段としては十分実用的な性能といえる。

今後は、本システムをベースとして、顔検出技術、パターン認識技術など個々の要素技術に関してさらに改良を進め、システムとしての性能向上を図る。

本システムの応用としては、CG 擬人化エージェントや音声認識、合成技術などと組み合わせにより、マルチモーダル対話システムの構築を行う予定である。

また、インターフェース分野以外にも、人を検知するためのインテリジェントなセンサ技術の一つとして、在宅高齢者モニタシステムや、人と協調する自律型ロボットにおける視覚認識技術など、様々な分野において顔画像認識技術の適用を検討していく。

## 謝 辞

本研究を進めるにあたり、独立行政法人産業技術総合研究所(旧 工業技術院電子技術総合研究所)フェローの大津展之様、同知能システム研究部門ヒューマンセンタードビジョン研究グループリーダーの坂上勝彦様に御指導、御助言を頂きました。記して感謝申し上げます。

## 引用文献

- 1) 長谷川修・坂上勝彦・速水悟：実世界視覚情報を対話的に学習・管理する人間型ソフトウェアロボット，信学論 Vol.J82-D-II No.10, pp.1666-1674(1999)
- 2) 福井和広：ユーザインターフェースにおける顔画像認識，第6回画像センシングシンポジウムチュートリアル講演会応用コーステキスト，pp18-32(2000)
- 3) 福井和広・山口修：形状抽出とパターン照合の組み合わせによる顔特徴点抽出，信学論(D-II)，vol.J80-D-II No8, pp.2170-2177(1997)
- 4) 金子俊一：実世界マシンビジョンのためのロバスト画像照合技術，電学論C, 121巻5号, pp.830-834(2001)
- 5) 大津展之・栗田多喜夫・関田蔵：パターン認識 理論と

応用, pp.214, 朝倉書店(1996)

- 6) 石井健一郎・上田修功・前田英作・村瀬洋: パターン認識, オーム社, pp204(1998)
- 7) 栗田 多喜夫・堀田 一弘・三島 健: Log-Polar 画像の高次局所自己相関特徴を用いた大きさに不変な顔画像の認識, 信学論, Vol.J80-D2 No.8, pp.2209-2217 (1997)