

博 士 論 文

画像処理を用いたユビキタスインターフェースの研究
—性別推定と家電製品制御システムに関する基礎研究—

平成 17 年度

岐阜大学大学院工学研究科

電子情報システム工学専攻

学位論文：博士(工学)甲290

川野 卓也

目次

1 序論.....	1
2 特徴抽出手法と識別手法.....	3
2-1 特徴抽出.....	3
2-2 高次局所自己相関特徴.....	5
2-2-1 特徴パターンの構築.....	6
2-3 四方向面特徴.....	8
2-3-1 エッジ抽出.....	9
2-3-2 正規化.....	11
2-3-2-1 内挿を使用する場合.....	11
2-3-2-2 平均画素法を使用する場合.....	11
2-3-3 局所加重平均フィルタ.....	12
2-3-4 特徴抽出.....	13
2-4 GABOR ウェーブレット特徴.....	16
2-4-1 フーリエ変換.....	16
2-4-2 短時間フーリエ変換(Gabor 変換).....	18
2-4-3 ウェーブレット.....	19
2-5 次元削減.....	21
2-5-1 主成分分析.....	22
2-5-2 判別分析法.....	23
2-6 識別機.....	24
2-7 その他の識別法.....	24
2-7-1 部分空間法(CLAFIC 法).....	24
2-7-2 累積寄与率を用いた部分空間法.....	25
2-7-3 複合類似度法.....	26
2-7-4 混合類似度法.....	28
2-8 特徴空間の評価.....	29
3 家電制御支援システムへの応用.....	31
3-1 システム構成.....	33
3-2 命令実行ポーズ.....	33
3-3 ポーズ識別の流れ.....	35
3-3-1 肌色候補領域の抽出.....	36
3-3-1-1 R 値, B 値を用いる手法.....	36
3-3-1-2 U 値, V 値を用いる手法.....	36

画像処理を用いたユビキタスインターフェースの研究
--性別推定と家電製品制御システムに関する基礎研究--

3-3-2	顔周辺矩形の作成.....	40
3-3-3	エッジ画像の作成.....	41
3-4	識別機の検討.....	43
3-4-1	部分空間法を用いた識別実験.....	44
3-4-2	累積寄与率を用いた識別実験.....	45
3-4-3	複合類似度法を用いた識別実験.....	47
3-4-4	混合類似度法を用いた識別実験.....	49
3-4-4-1	学習データに対する識別実験.....	49
3-4-4-2	未知データに対する識別実験.....	51
3-4-5	線形判別分析.....	54
3-5	仮想学習データを用いた学習.....	55
3-5-1	判別空間が安定する枚数.....	55
3-5-2	仮想学習データ作成.....	56
3-6	高次局所自己相関特徴の傾きに対する対応.....	59
3-6-1	平行移動に対する高次局所自己相関特徴の検証.....	59
3-6-2	傾きに対する高次局所自己相関特徴の検証.....	60
3-6-3	傾きを考慮した辞書作成方法.....	62
3-6-4	可視化による特徴空間の観察.....	65
3-7	相対位置情報を持たせた特徴量抽出の提案.....	73
3-7-1	識別実験.....	73
3-7-2	エッジ情報がある場合の考察.....	75
3-8	特徴空間を考慮した辞書作成法の提案.....	76
3-8-1	識別実験.....	76
3-8-2	クラス内分散, クラス間分散の考察.....	77
3-8-3	クラス間距離に対する考察.....	78
3-9	異なった認識対象の拡大.....	80
3-9-1	命令実行ポーズ.....	80
3-9-2	手領域の抽出方法.....	82
3-9-3	特徴量抽出.....	82
3-9-4	識別実験.....	83
3-10	システム操作方法.....	84
3-10-1	上半身のポーズを用い操作する場合.....	85
3-10-2	手先のポーズを用い操作する場合.....	86
3-11	まとめ.....	87
4	性別推定について.....	88

画像処理を用いたユビキタスインターフェースの研究
--性別推定と家電製品制御システムに関する基礎研究--

4-1 四方向面特徴・線形判別分析を用いた識別.....	88
4-2 実験.....	89
4-2-1 データベース.....	89
4-2-2 最適な特徴次元数.....	90
4-2-3 最適な解像度.....	91
4-2-4 使用する特徴量.....	92
4-2-5 未知データに対する評価方法.....	94
4-2-6 最適な方向面.....	95
4-3 クラスタ判別法による特徴空間の観察.....	98
4-4 結果及び考察.....	100
4-5 各年代に対する詳細な考察.....	102
4-6 まとめ.....	103
5 むすび.....	104

1 序論

生活様式をより便利に変える情報技術の一つとしてユビキタスコンピューティングがある[1]~[7]. またそれを実現するための技術として, パターン認識を用いた画像処理技術がある. 画像処理技術は, 情報技術(画像処理チップ等のハードウェアの向上, デジタルカメラ等の一般家庭への普及), 計算機速度の飛躍的な向上と大容量メモリの低価格化に伴い, 工業, 医療等の専門分野で使用されるだけでなく, 日常生活でも活用されつつある. さらに, 近年急速に研究が進んでいるバイオメトリックス[8][9], ITS[10]~[19]の分野にも画像処理技術は導入されつつある. 例えば, 現在 JAL では人間の顔の形や瞳の紋様から個人を識別しチェックインを行う, e-チェックインの実証実験を行っている. またトヨタは, ドライバーの顔の向きを画像認識して警告するシステムを開発し, 2006年春に発売する新型車に搭載する. このような技術の延長線上には, 一般家庭をターゲットとしたカメラ搭載情報家電の導入が考えられ, 介護機能や防犯, 家事の自動化も実用化されつつある.

そこで, 本研究では, ユビキタスコンピューティング応用技術の基礎研究として, 画像処理を用いた家電製品制御システムと顔部品画像を用いた男女推定に関する研究を行った.

まず家電製品制御システムに関しては, クラス識別方法, データ収集方法, 特徴抽出方法, 辞書作成法について検討した. 本システムは入力画像から肌色領域を抽出して, 特徴量を抽出する領域を求め, 領域内に存在する肌色領域のエッジより高次局所自己相関特徴を抽出した後に識別機により識別を行う.

特徴抽出の問題については, メッシュ特徴を使用すると, 少量のずれにより特徴量が大幅に異なってしまう. しかし, 高次局所自己相関特徴は平行移動に不変なため物体が平行移動した場合でも特徴量は変わることはない. 本研究で扱っているデータは, 対象物が動かない場合でも, 特徴抽出領域が微妙にずれってしまう場合があるため高次局所自己相関特徴が適している.

画像から特徴量を抽出する際に, 高次元の特徴量を抽出し, そのまま識別機を使用し判別してしまうと, 識別を行うのに最適な特徴量を使用しているにもかかわらず, 次元の呪い[20]により正しい認識結果を得られない場合が考えられる. そこで, 本システムに対して最適な識別機を検討した.

学習データから有効かつ安定した判別空間を構築するためには, 1枚の画像から抽出される特徴次元数の数倍以上の枚数が学習データとして必要であるといわれている. しかし, 辞書作成のための学習データを得る際に, 現実問題として短時間では十分な枚数が得られない, 容量が制限されているなど様々な問題がある. そこで, 本研究では実在する学習データから仮想的にデータを増や

すことを提案した。今回は、原画像を幾学的に傾けることによって、仮想学習データを作成した。また、仮想学習データを作成することにより利用者がベッド上で傾いた場合でも対応できる辞書作成についても検討した。具体的には分散比、クラスタ判別法により作成された辞書に対する評価を行った。

平行移動に強いことは高次局所自己相関特徴の利点ではあるが、次のような欠点も含んでいる。高次局所自己相関特徴は空間上の位置情報は保持していないため、ある物体の形や向きを変えずに位置を変えただけならば同じものとして認識されてしまう。そこで、特徴抽出領域を分割することにより、大まかな相対位置の情報を取り込んだ特徴量を抽出し、相対位置情報を持たせた特徴量についての考察を行った。

本研究で扱っているデータは、肌色領域から特徴を抽出しているため、服装の違いによって同じクラスでも、得られる特徴量が異なる場合が出てくる、服装ごとに辞書を作成すれば良いが、服装をどのように判断するかが問題となってくる。大幅に異なる特徴量が存在しても、同じクラスとして取りまとめたい場合があるとす。しかし、大幅に異なる特徴量を1つのクラスとすると、クラス内分散が大きくなりすぎてしまい、クラス間分散が小さくなってしまふことが考えられる。このような特徴空間が作成されてしまうと、判別が困難になってしまう。そこで複数の辞書作成法を提案し、特徴空間の分離度等を考察した。

さらに、認識対象が異なる場合でも同じアルゴリズムで識別できるかを考察した。具体的には体の一部のみが写っている場合でも今まで議論してきた手法で対処できるかどうかを検討した。

次に、顔部品画像を用いた男女識別に関する研究を行った。人間の男女を形態的に判別する場合、女性は男性に比べて目が大きく、鼻が小さく、顎は細い、というように、髪型を除けば、額の広さ、顎の尖り具合等、顔に差があることが知られている。

そこで本研究は、顔部品に注目した場合、顔のどの部分に性別を区別する特徴が存在するかを議論する。男女識別には対象画像から四方向面特徴[21]を抽出した後に、線形判別分析を適用する方法を用いる。またクラスタ判別法により特徴空間を可視化し、どの顔部品が良好な特徴空間を作成しているかを観察することにより評価を行う。さらに、男女識別に成功した各顔部品画像の平均画像を作成することにより、どの部分に特徴の差が表れているかを調べる。最後に年代による、性別の差について考察を行う。

本論文は5章で構成されている。第2章では特徴抽出手法と識別手法について、第3章は家電製品制御システムへについて、第4章で顔部品画像を用いた性別推定について、最後に第5章は本研究のまとめについて記す。

2 特徴抽出手法と識別手法

2-1 特徴抽出

パターン認識において、まず行われることはパターンの観測に基づく特徴の抽出である。しかし、特徴抽出は問題の性質に大きく依存するために、一般論を語ることは難しい。その上、特徴は無限に多くあると考えられるので、適当な連続性を仮定して有限個の特徴を抽出しなくてはならない。例えば、画像理解の問題では、画素数を特徴次元数として入力すると非常に大きな次元数になるが、対象の性質に基づいた特徴を用いれば、非常に少ない次元数の特徴で対象をうまく表現することができる。実際どのような特徴を抽出すれば良いかという問題は重要である。特徴の数が少なければ識別は不完全になるであろうし、特徴を多く抽出しても構造が複雑になってしまう。また、特徴の次元数を多く増やせば増やすほど、学習に必要なデータの数が指数関数的に増大する。従って必要かつ十分な特徴を抽出しなくてはならない。画像理解のために用いられる特徴量として様々な特徴量が存在する。

手書き文字認識の分野で実績のある特徴量「加重方向指数ヒストグラム法」は文字の局所的な輪郭成分を向きごとに足し合わせた量にガウスフィルタを作用させて文字変形に対して頑健になっている[20]。一方、文字ストロークが文字全体の中でどの向きにどれだけあるかを表現する「外郭方向寄与度特徴」[23][24]や、背景から見て文字の直線成分がどこにどれだけ存在するかを表現する「拡張セル特徴」[25][26]は、文字の構造を表現できる特徴量である。また、筆記者を特定するための特徴量として手書き数字パターンの傾きを考慮した「Inclination-DCT 特徴量」がある[27]。

分割されたブロック間の関係を考慮した特徴量として、孫らは「方向線素特徴量」を提案している[28][29]。複数の Gabor フィルタを用いてテクスチャの特徴抽出を行う。Gabor フィルタは出力が複素数と成る帯域フィルタであり、パラメータの異なるフィルタの集合を用いることにより画像パターンの変化の度合と方向性についての情報を得ることができる[30][31]。

3次元物体モデルの類似検索のための様々な形状特徴量が提案されている。形状特徴量は、回転に対して影響を受ける形状特徴量と、回転に対して影響を受けない特徴量(回転不変な特徴量)に分類することができる。前者には、頂点密度分布を量子化した特徴量[32][33]、表面積、体積、重心位置、法線ベクトルなどの特徴量[34]などがある。後者の回転不変な特徴量としては、面情報を重心座標系として物体の位置に依存しない特徴量[35]、レーブグラフに基づいた3次元形状の相似度による特徴量[36]などがある。主成分分析を利用することで、

物体に対して固有の傾きと位相を一意に定めて回転不変な特徴量とする手法としては、3次元物体モデルのテンソル、法線、座標に主成分分析を応用した特徴量[37]、球面調和関数に主成分分析を用いた手法[38][39]、不連続フーリエ変換と主成分分析を用いた特徴量[40]などがある。更に、主成分分析による傾きと位置の固有化が、必ずしも類似性の判断に適さないことから、物体特徴量を層となった球に分割し、調和関数を計算する特徴量などが提案されている[41]。

合成開口レーダ(SAR)の画像データ解析において使われる特徴量として、多偏波画像からのテクスチャ特徴量の算出方法として、一般的に用いられているGLCM法がある。この手法はテクスチャを近傍画素の距離と方位から定義するもので、原画像から部分画を抽出し、その部分画に対してグレーレベル共起行列(GLCM: Gray Level Co-occurrence Matrix)を求め、その行列要素から部分画の特徴量を算出する手法[42]～[45]がある。類似画像検索を行うための特徴量として、色特徴量、微分方向特徴量を使用したもの[46]、カラーヒストグラムを使用したものがある。一方、古くから行われている手書き文字認識の分野では、文字の局所的な輪郭成分を向きごとに足し合わせた量にガウスフィルタを作用させて文字変形に対して頑健な特徴量をもつ四方向面特徴[21]などがある。

また特徴を抽出した際に特徴次元数が大きくなりすぎると、各次元の特徴ベクトルの相関が高くなってしまい、特徴空間が形成できないなど、問題が起きる可能性が高い、そこで特徴空間の次元圧縮手法のアルゴリズムが重要になる。

本章では本研究で使用する高次局所自己相関特徴、四方向面特徴、Gaborウェーブレット特徴について記す。また次元圧縮手法として、主成分分析、線形判別分析について記す。さらに、クラスを判別する際に使用する識別機についても記す。

2-2 高次局所自己相関特徴

自己相関関数は平行移動に対して不変であることが知られている。その高次への拡張が高次自己相関関数である[47]~[49]。画面内の対象領域を $f(r)$ とすると N 次自己相関関数は、変位方向(a_1, a_2, \dots, a_N)に対して

$$x^N(a_1, a_2, \dots, a_N) = \int f(r)f(r+a_1)\cdots f(r+a_N)dr \quad (2-1)$$

で定義される。よって、高次自己相関関数は次数や変位方向 (a_1, a_2, \dots, a_N) の取り方によって無数に考えられる。

本研究では次数 N を $2(N=0, 1, 2)$ に、変位方向の範囲を参照点の周りの局所性な 3×3 に限定した。このようにした理由は、実際的な応用を図るためには、変位方向(a_1, a_2, \dots, a_N)、次数 N を絞り込まなくてはならないためである。変位方向の範囲を 3×3 に限定した理由は、画面に関する加法性を満たすには、変位方向は r の局所領域に限定されなければならない。さらに、画像では離れた転換の相関よりも局所領域内の点間の相関の方が重要であるためである。また、次数を $N=2$ に限定した理由は、 N は3次から8次までのマスクも作成することができるが、これ以上高次の特徴は数が多い割には意味がないことと、2次まで平行移動不変の完全系となり、 180° 傾いた違いも反映できることである。もちろん、変位 a_i を局所的な場合に限定したので、厳密には平行移動に対し不変な完全系にはならないが、厳密な意味での完全系は現実的には不可能である。また、実際上不可欠な要請ではない。ここで傾き等の同値性も考えると、上記の25種はさらに縮退をして、例えば、 90° 傾き同値では、0次1種類、1次2種類、2次6種類、の計9種類となるが、ここでは平行移動同値類のみを考えることにする。

3×3 の局所領域で次数が $N=2$ の場合、単純に考えると、9点の配置(このうち1点は常に 3×3 の中心で良い)は 2^8 通りある。0次から8次までのすべてのマスクは256枚必要となるが、平行移動に関して同値なマスクを重複して数えている。実際、式(2-1)により注目している9点を一様に平行移動して考えても結果は同じである。したがって、マスク上の配置を一様に並行移動して重なり合うマスクは互いに同値であり同値類として一つで簡約化できる。

このように変位方向(a_1, a_2, \dots, a_N)の範囲を局所的な方向に限定し、 N 次自己相関関数によって得られる特徴量の値を高次局所自己相関特徴と呼ぶことにする。

0次のマスクは1種(No.1)で、これは図形の総和を求める特徴である。1次は4種あり(No.2~No.5)、これらは従来のいわゆる自己相関である。2次は20種(No.6~No.25)存在する。2次までで合計25種存在する。図2-1に2次までのマ

スクパターンを示す。

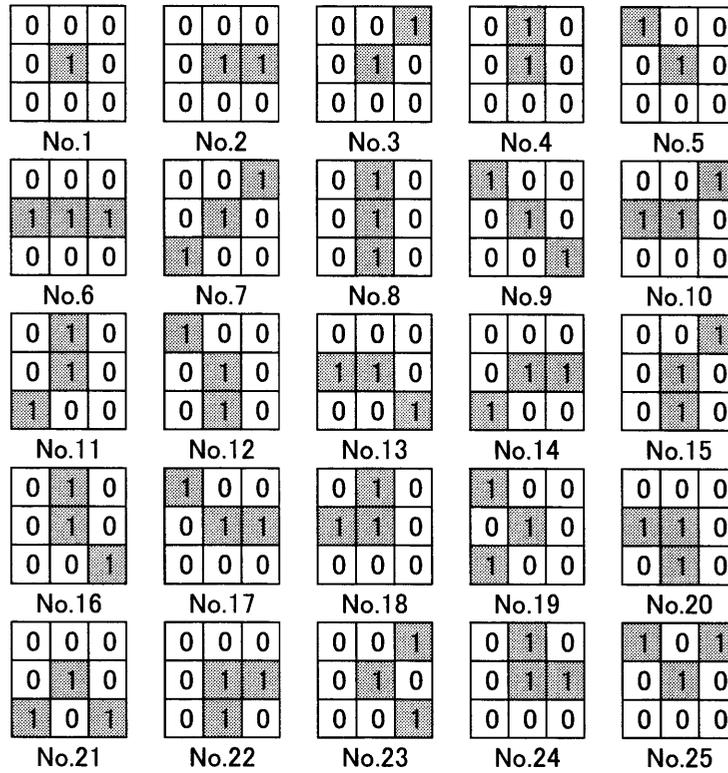


図 2-1 2 値画像へ適用する 25 種のマスクパターン

画像の高次局所自己相関特徴は、画像の各位置に図 2-1のマスクの中心を合わせ、図中“1”と示された個所の濃淡値を掛け合わせるとともに、マスクを移動するたびにその値を畳み込み演算をする(積和する)ことによって求めた。2 値画像の場合は各マスクの数が特徴量となる。

上記 25 種の各マスクに対して得られる高次自己相関関数の値を x_j で表すと 25 次元の特徴ベクトルを得ることができる。各特徴 x_j は図形の局所的性質を大域的に要約したものであり、平行移動に不変な互いに独立な特徴と考えられる。また、これらの局所的マスクによる畳み込み演算は計算による処理も簡単である。

2-2-1 特徴パターンの構築

25 パターンへの振り分け方を図 2-2を使って説明する。3×3 のマスクが図 2-2(a)の点線上にきたとき、図 2-1のマスクパターンにしたがって、次のように 25 パターンに値が積み上げられる。まず、中央に画素があるので No.1 に+1、そして No.2 にも+1、No.3 に+1、No.4 に+1、No.16 に+1、No.23 に+1、No.24 に+1 となる。次に、3×3 のマスクが図 2-2(b)の点線上にきたとき、No.1 に+1、No.4 に+1、No.5 に+1、No.8 に+1、No.9 に+1、No.12 に+1、No.13 に+1、No.16

に+1, No.18 に+1, No.20 に+1 となる. 結果として表 2-1のようになる. これを対象画像上の全画素について行い, No.1 から No.25 までそれぞれを, 特徴量を抽出する対象画像領域の画素数で割った値が最終的な特徴ベクトルとなる. したがって, 25 種のマスクパターンの出現確率を計算していることになる. 表 2-1に図 2-2に対する 25 パターンの積み上げ結果を示す. これは画素の値が 0, 1 の場合であり, 濃淡画像で 25 パターンを構築する場合はそれぞれのパターンの畳み込み演算により構築される. また, 図 2-2の特徴量を抽出する対象画像領域の画素数は 25 なので, 最終的な正規化された特徴ベクトルは 25 で割って表 2-2のようになる.

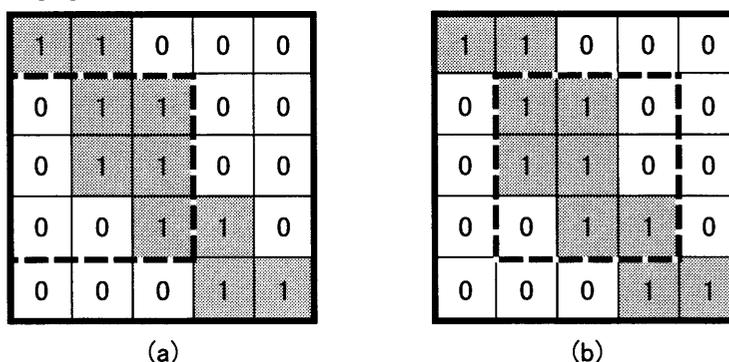


図 2-2 25 パターンへの振り分け方

表 2-1 25 パターンの積み上げ結果

特徴番号	1	2	3	4	5	6	7	8	9
特徴量	2	1	1	2	1	0	0	1	1
特徴番号	10	11	12	13	14	15	16	17	18
特徴量	0	0	1	1	0	0	2	0	1
特徴番号	19	20	21	22	23	24	25		
特徴量	0	1	0	0	1	1	0		

表 2-2 実際の特徴ベクトル

特徴番号	1	2	3	4	5	6	7	8	9
特徴量	0.08	0.04	0.04	0.08	0.04	0	0	0.04	0.04
特徴番号	10	11	12	13	14	15	16	17	18
特徴量	0	0	0.04	0.04	0	0	0.08	0	0.04
特徴番号	19	20	21	22	23	24	25		
特徴量	0	0.04	0	0	0.04	0.04	0		

2-3 四方向面特徴

人間の視覚系を模倣した特徴量として、コントラストフィルタ[50]~[52]、Gabor フィルタ[53]を用いた特徴量がある。これとは別に Gabor 特徴とほぼ同等な特徴量として、文字認識で一般的に使用されている四方向面特徴がある[21]。四方向面特徴は、水平、垂直、左下がり、右下がりの各方向面で構成される。原画像(顔画像)からエッジ検出フィルタを用いて四方向面を検出し、得られた画像にガウシアンフィルタをかけ、低解像度化したものを図 2-3に示す。次節より四方向面特徴の具体的な抽出方法について記す。

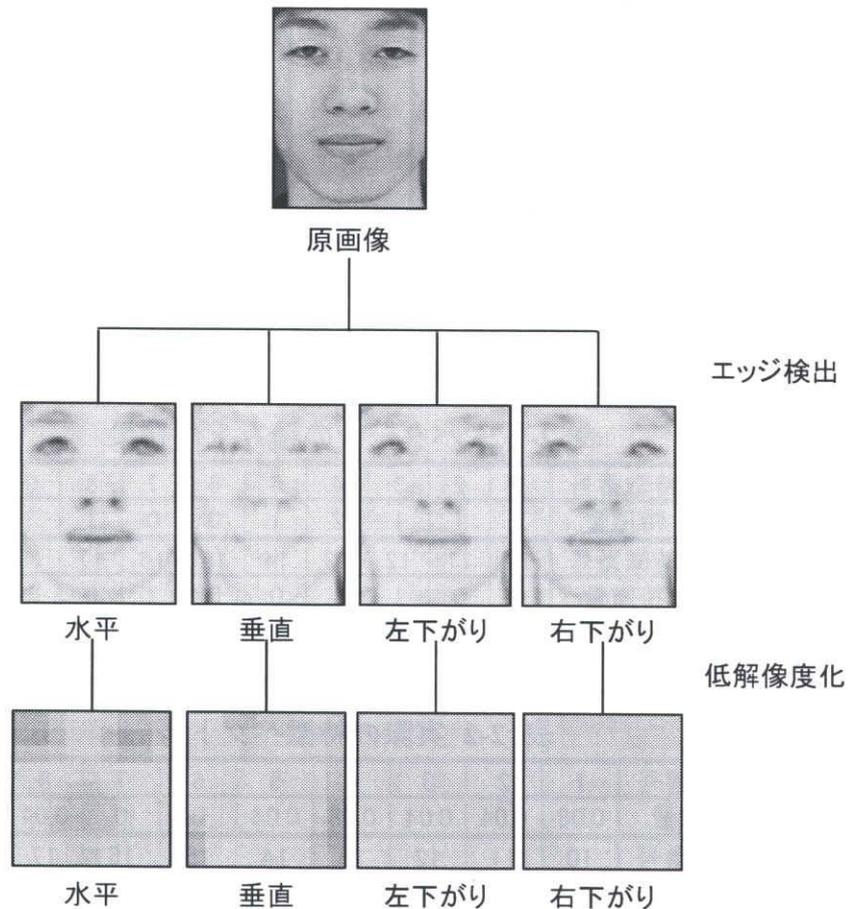


図 2-3 顔画像における四方向面パターンの例

2-3-1 エッジ抽出

エッジを検出する方法と言っても様々な方法がある。四方向面特徴では方向別のエッジを必要とするため、1次微分(グラジエント)に基づく方法の中で、テンプレート型のものを使用する。

これは、画像上でのエッジ付近の濃淡パターンを想定したテンプレートを準備し、画像間との相関を計算することによってエッジ要素を検出する方法である。実際には8方向のエッジに相当する8個の3×3テンプレートを用いることが多い。例えば、Robinsonのエッジ検出オペレータと呼ばれるものでは、図2-4のようなテンプレートが用いられている。

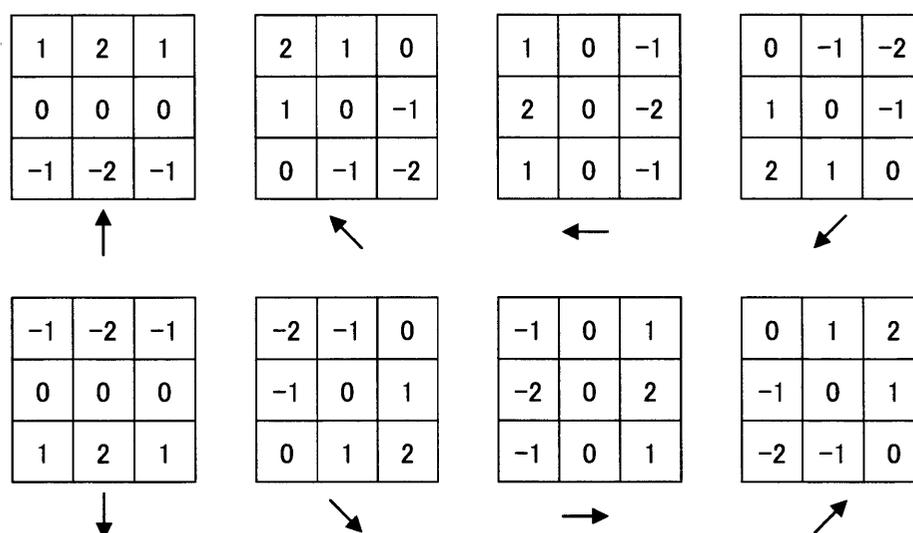


図 2-4 Robinson オペレータ(テンプレート型)

各テンプレートの下に示されている矢印はエッジ上で暗い部分から明るい部分への方向を示している。

各テンプレートについて画像に対し畳み込み演算を行い、最大出力が得られたテンプレートの方向をグラジエントの方向とし、その時の出力値をグラジエントの強度とする。その他8方向テンプレートの例を以下に示す。

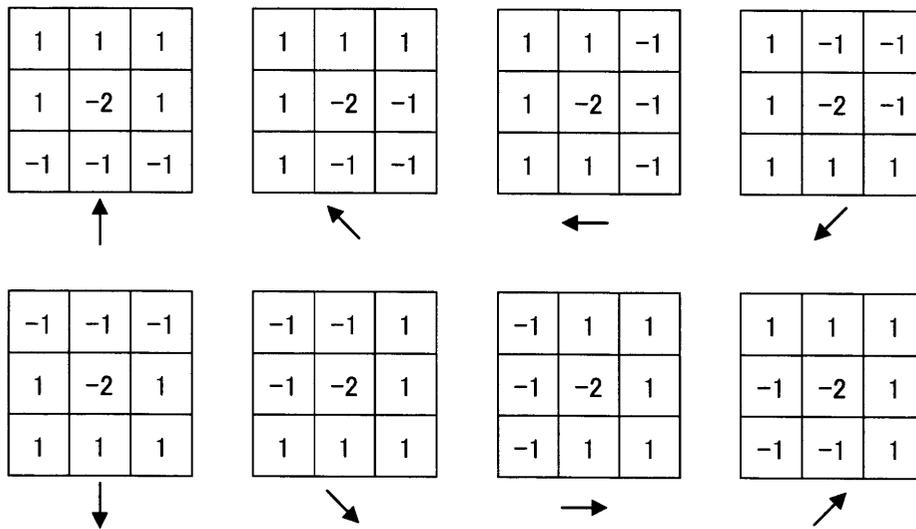


図 2-5 Prewitt オペレータ(テンプレート型)

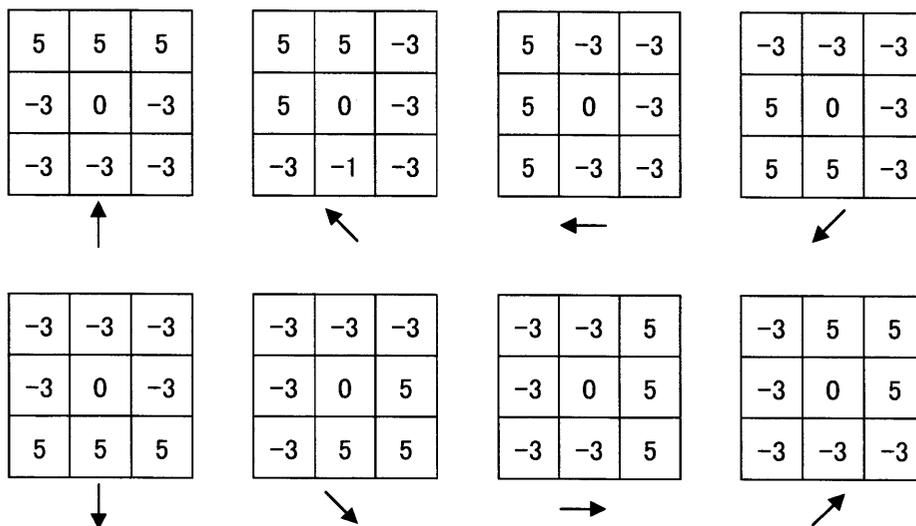


図 2-6 Kirsch オペレータ(テンプレート型)

テンプレート型は、 3×3 のマスクを使用した場合、実際 8 方向のエッジが存在する。8 方向のエッジ面を特徴量として使用することも考えられるが、2-5節で記すように、特徴次元数を無駄に増やしてしまうと、学習に必要な画像枚数が膨大な量となってしまふ。また、Robinson オペレータは、Prewitt, Kirsh と異なり、あるテンプレートに対して 180° 方向が異なるテンプレートのエッジ強度値は符号が反転しているだけである。そこで、四方向面特徴では、Robinson オペレータにより抽出されたエッジ強度値の絶対値を使用する。

また、 3×3 のマスクを使用すると、対象画像の上下左右の端 1 画素分(エッジ抽出を行った対象画像の周辺の画素分)はエッジ情報が抽出されないため、端に存在する画素は特徴量として使用してはいけない。

2-3-2 正規化

対象画像から特徴量を抽出する際、特徴次元数をそろえなくてはならない。このため、各画像から抽出される、各特徴面の解像度を同じ大きさに揃える。この際、内挿を使用して画像の解像度を拡大させてしまうと、実際に存在しない情報が加わってしまうため、内挿を使用して拡大をしてはいけない。よって、解像度の大きさを揃える際には、対象画像よりも小さい解像度に必ず揃えなくてはならない。同じ大きさに揃える方法として2種類の方法が考えられる。

2-3-2-1 内挿を使用する場合

内挿を使用して一度に縮小してしまうと、内挿点(画像データを内挿したい点、拡大縮小後の画像内の点)を求める際に、使用されない観測点(画像データが存在する点、拡大縮小前の画像内の点)が現れる可能性がある。具体例として、 144×152 の対象画像を 8×8 の大きさに縮小する場合を考える。共1次内挿法[54]を使用する場合、周囲の観測点4点を使用することになる。周囲の観測点の決め方にもよるが、各辺のサイズを50%以下の大きさに縮小した場合に、使用されない観測点が現れてしまう。このため、各辺のサイズを50%より小さいサイズに縮小すると、使用しない情報量がでてきてしまう。同様に、共3次内挿法[54]を使用する場合は、各辺のサイズを25%より小さいサイズに縮小すると、使用しない情報量がでてきてしまう。よって共1次内挿法を使用して 144×152 の対象画像を 8×8 に縮小する場合は、 $144 \times 152 \rightarrow 128 \times 128 \rightarrow 64 \times 64 \rightarrow 32 \times 32 \rightarrow 16 \times 16 \rightarrow 8 \times 8$ と段階的に縮小を行わないと情報量が必要以上に失われてしまう。

また、縦または横サイズが奇数だった場合、観測点の決定方法に注意しないと、使用しない情報量が現れてしまう。具体例として、 1×7 の画像を 1×4 に縮小する際に、観測点の決定方法によっては右端の画素を使用しないケースが現れる。これでは右端に重要な情報が存在した場合には確実に情報が失われる。よって、観測点を決定する際に右端の情報量も必ず使用するよう注意しなくてはならない。

2-3-2-2 平均画素法を使用する場合

例として 6×6 の画像を 4×4 の画像に縮小する場合を考える。まず、対象画像のサイズと縮小するサイズとの最小公倍数を求め、対象画像を最小公倍数の大きさに拡大する。拡大する際には内装を使用せず、対象画像の画素値をそのまま入力する。例として、 6×6 の画像を 12×12 に拡大したものを図 2-7に示す。

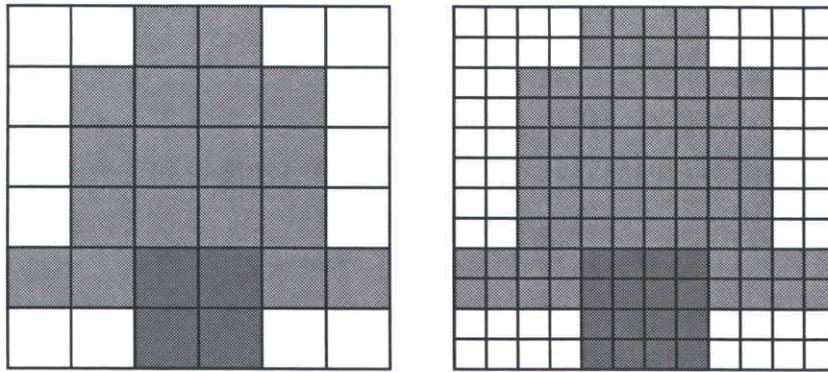


図 2-7 拡大例

拡大した画像を，縮小したい画像のサイズ分のブロックに分けてそのブロック内の要素の平均を求め，縮小後の各要素とする．例として，12×12に拡大したものを4×4に縮小したものを図 2-8に示す．

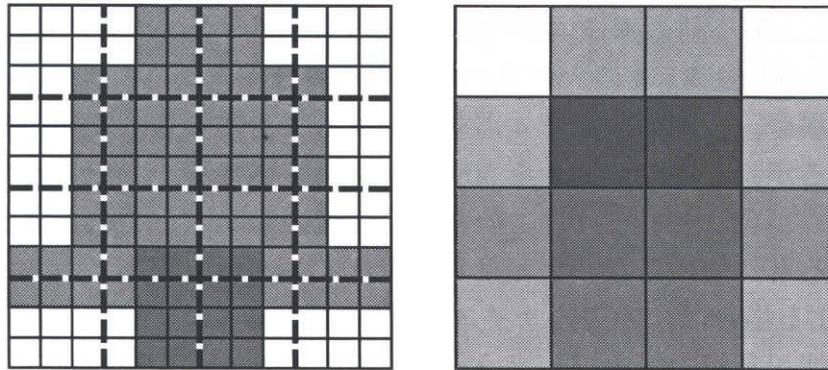


図 2-8 縮小例

2-3-3 局所加重平均フィルタ

局所加重平均フィルタは中心画素に近い画素の寄与が大きいとして各画素の重みを考慮したものである．重み係数行列としては，図 2-9に示される3×3の大きさのものが使われることが多い．

1/10	1/10	1/10
1/10	2/10	1/10
1/10	1/10	1/10

1/16	2/16	1/16
2/16	4/16	2/16
1/16	2/16	1/16

図 2-9 局所加重平均フィルタ

正規分布を用いた3×3を超える重み係数も頻繁に用いられる．これはガウシヤンフィルタ(Gaussian filter)と呼ばれる．原点周りの正規分布 $G(I, j)$ は標準偏差を σ とすれば式(2-2)で与えられる．

$$G(i, j) = \exp\left\{-\frac{i^2 + j^2}{2\sigma^2}\right\} \quad (2-2)$$

式(2-2)に従って、設定した局所領域の全画素について G を重み係数として与え、最後に全重み係数の合計が 1.0 になるように正規化する。平滑化の程度は σ で制御するが局所領域の大きさが $(2N+1) \times (2N+1)$ の場合、 $\sigma=N/2$ とするのが一つの目安である。中心画素の重みを増すことによって、平滑化の程度は局所平均フィルタより自然なぼかしを実現できる。

加重平均フィルタを使用する際も、テンプレート型エッジ検出フィルタと同様に画像の端に存在する画素の扱いに注意をしなければいけない。例えば、 $N \times N$ のマスクを使用すると、対象画像の上下左右の端 $[N/2]$ 画素分(加重平均フィルタを適用した対象画像の周辺の画素分。また $[]$ はガウス記号といわれるもので、 $i=[u]$ とは u の少数部分を切り捨てて u の整数部分を i とすることを意味している)は、画素が存在しない点の計算は行わず、存在する画素に対してのみフィルタがかかるようにしなくてはならない。

2-3-4 特徴抽出

今回、具体的には特徴量(初期特徴ベクトル)を以下のように求めた。まず、濃淡画像の各画素における濃淡値の水平、垂直、左下がり、右下がりの四方向のエッジ勾配の絶対値を **Robinson** フィルタによって求め、それぞれ方向別に分解した四枚の画像を作成する。次に四枚の画像を平均画素法または、内挿をもちいて画像を縮小する。この際必要に応じてガウシアンフィルタを使用する。低解像度化された四枚の画像の各画素値を特徴ベクトルの 1 つの要素として定義した。これにより、低解像度化された四枚の画像の総画素数が 1 枚の画像から得られる特徴次元数となる。内挿を用いて特徴抽出を行うフローチャートを図 2-10 に、平均画素法を用いて特徴抽出を行う方法を図 2-11 に示す。

画像処理を用いたユビキタスインターフェースの研究
 --性別推定と家電製品制御システムに関する基礎研究--

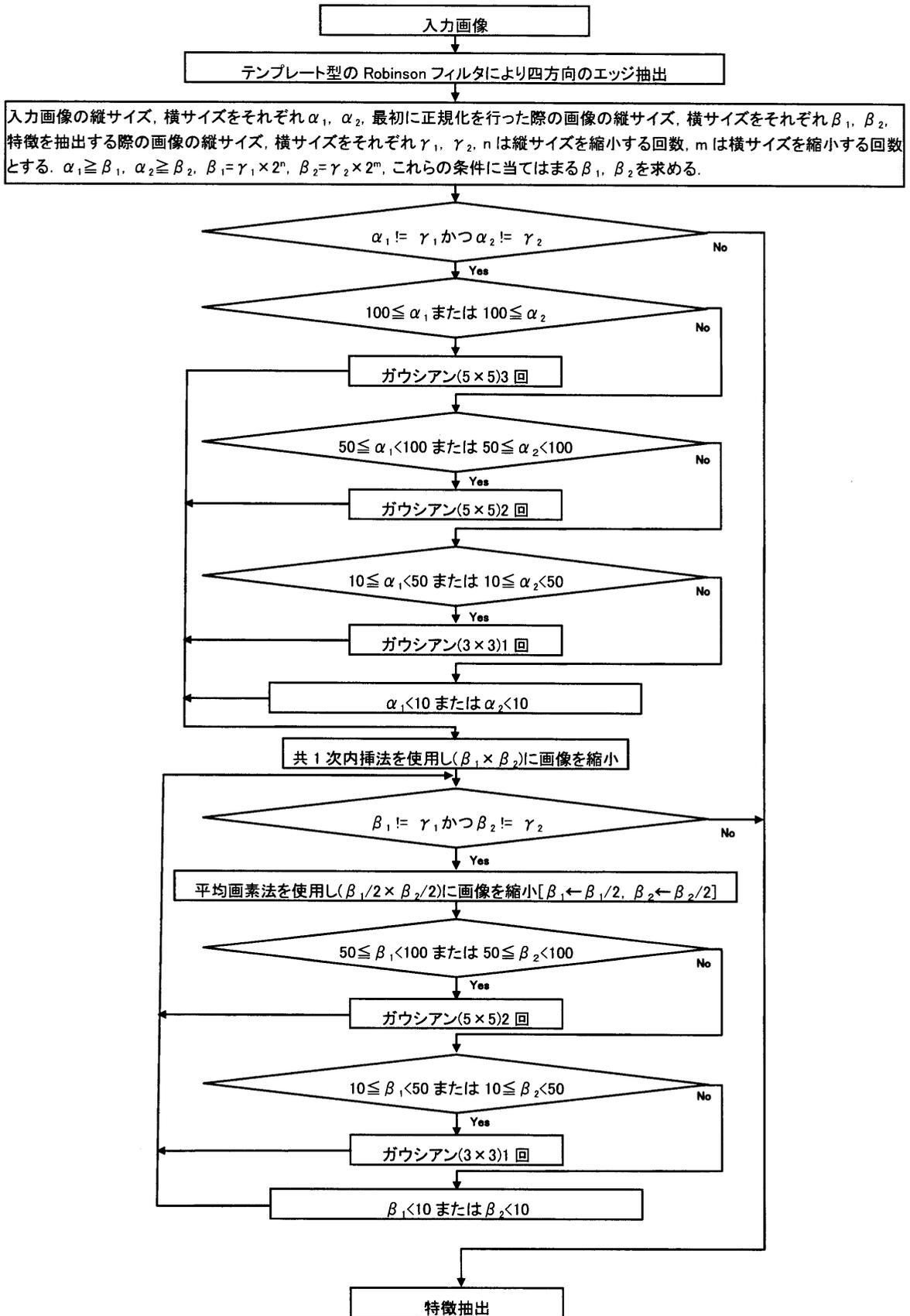


図 2-10 内挿を用いた特徴抽出法

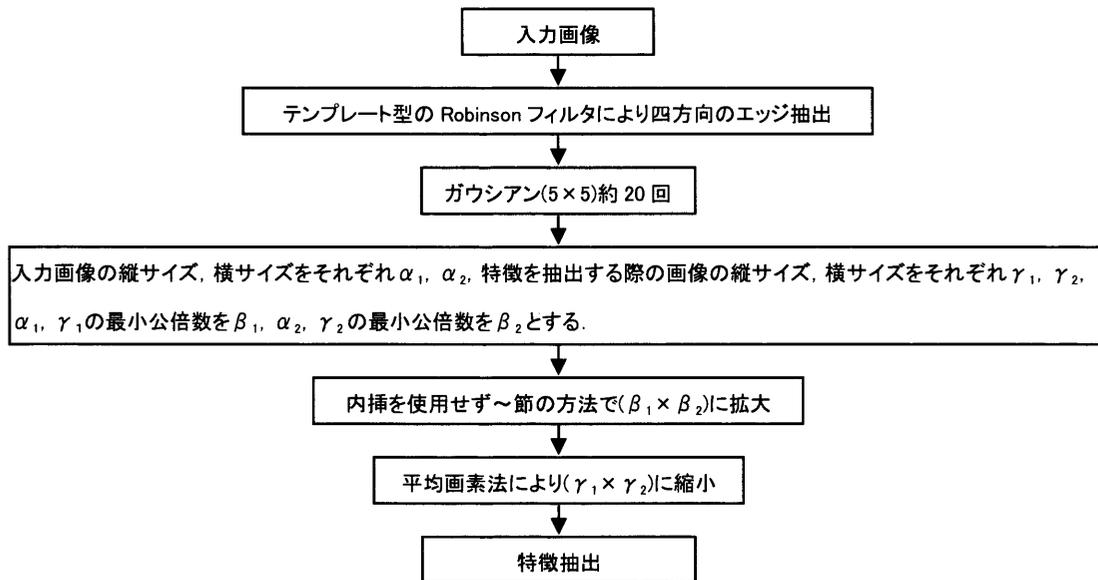


図 2-11 平均画素法を用いた特徴抽出法

2-4 Gabor ウェーブレット特徴

四方向面特徴と同じ性質を持つといわれている，Gabor ウェーブレットについて考察を行った．Gabor ウェーブレットは簡単に説明すると，ガウス関数と正弦関数の積である．これを数式で示すと式(2-3)になる．

$$\phi_{k,\theta}(x,y) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2(x^2+y^2)}{2\sigma^2}\right) \left\{ \exp(ik(x \cos \theta + y \sin \theta)) - \exp\left(-\frac{\sigma^2}{2}\right) \right\} \quad (2-3)$$

原画像に対して式(2-3)をフィルタリングすることにより，特徴を抽出するものである．式(2-3)以外にも Gabor ウェーブレットと呼ばれている式が多くある[55]～[59]．しかし，式(2-3)と比較した場合[55]～[59]は振幅の大きさ，フィルタサイズの大きさが変化しただけであって，結果的にはどれもほぼ同じ結果が得られる．

しかし，簡単な説明だけでは，なぜ Gabor ウェーブレットによってエッジが抽出されるのかを理解することは難しい．そこで，これらを理解するために，フーリエ変換，短時間フーリエ変換，ウェーブレットについての説明を行う．

2-4-1 フーリエ変換

画像処理で使用されるフーリエ変換とは，画像を種々の周波数の正弦波に分割することである．また，フーリエ変換は，画像の見方を画像の座標ベースのものから周波数ベースのものへ変換を行う数学的な手法としても考えることもできる．具体的な例を用いて説明する．まず，原画像(図 2-12(a))から低周波成分のみを抽出したものを図 2-12(b)に，逆に原画像から高周波成分を抽出したものを図 2-12(c)に示す．

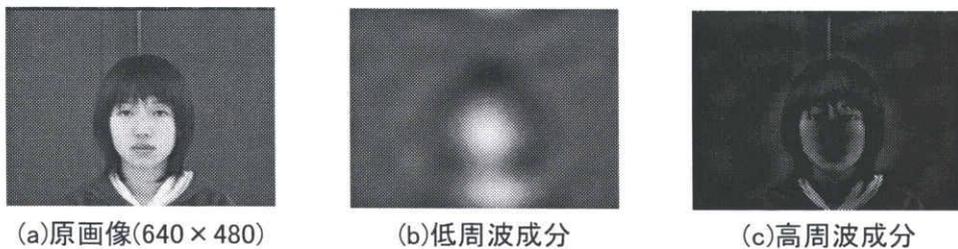


図 2-12 フーリエ変換

$$F(\omega_1, \omega_2) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \underbrace{\exp(-i\omega_1 x)}_{\text{複素正弦波}} \underbrace{\exp(-i\omega_2 y)}_{\text{原画像}} f(x, y) \quad (2-4)$$

図 2-13は，画像を周波数成分で表現したものである．ちなみに，式(2-4)の記

号で示した場合、横軸が ω_1 、縦軸が ω_2 となっている。この画像がどのように得られるか説明する。例えば、原画像(図 2-14(a))に、 $\omega_1=\pi$ 、 $\omega_2=\pi$ の場合の正弦波フィルタ(図 2-14 (b))を掛け(畳み込み演算を行い)、得られた画像(図 2-14 (c))の画素の合計値(スペクトル)が、図 2-13の一つの画素値として代入される。また、原画像(図 2-15 (a))に、 $\omega_1=\pi/36$ 、 $\omega_2=\pi/18$ の場合の正弦波フィルタ(図 2-15(b))を掛け(畳み込み演算を行い)、得られた画像(図 2-15(c))の画素の合計値(スペクトル)が、図 2-13の一つの画素値として代入される。代入される位置を矢印で示す。

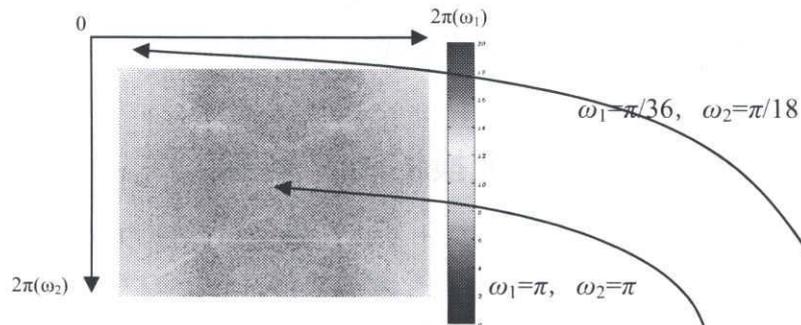


図 2-13 スペクトル成分

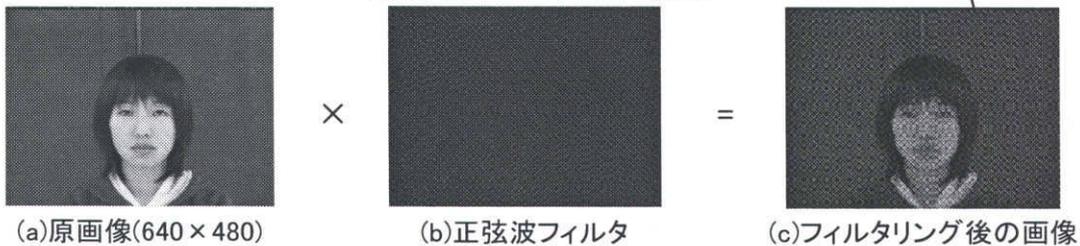


図 2-14 スペクトルの求め方(その 1) [$\omega_1=\pi$, $\omega_2=\pi$]

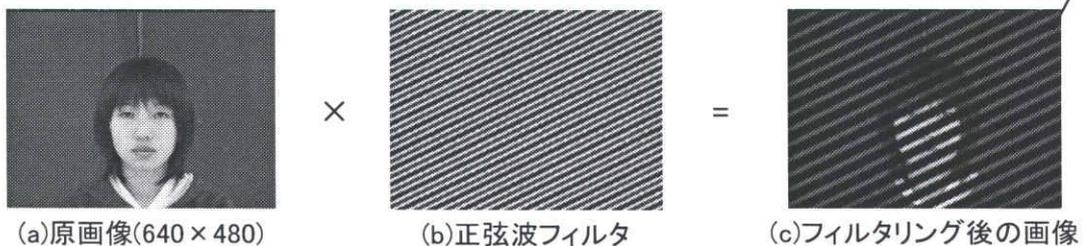


図 2-15 スペクトルの求め方(その 2) [$\omega_1=\pi/36$, $\omega_2=\pi/18$]

このようにして周波数成分に変換されたものが求められる。求められた値 ($F(\omega_1, \omega_2)$) はスペクトルと呼ばれるものであるが、これは、周波数に対する各成分の分布を示す。例えば、ある周波数(ω_1, ω_2)で大きいスペクトル値が求められた場合、周波数(ω_1, ω_2)の成分(正弦波)が多く含まれていることになる。 $F(0, 0)$ は原画像の画素値をすべて加算したものなので、必然的に大きな値となる。(こ

の成分は直流成分とも呼ばれる)

周波数解析はテクスチャ画像に対しては有効であるが、テクスチャ構造が含まれない顔画像、文字画像に対しては有効ではない。その理由として、周波数領域への変換で、画素の位置情報が失われることがある。画像の周波数成分だけを調べると、特別な事象が生じる(エッジが存在する)位置を知ることができない。画像が、ある範囲に渡り、あまり変化しない場合(このことを定常と言う)、この欠点はあまり問題にはならない。しかし、ほとんどの特徴となる現象は、かなりの非定常性か、または急変部等を含んでいる。これらの特徴が画像の最も重要な部分になることが多く、そのためフーリエ解析は、顔画像や文字画像から特徴を抽出するにはあまり適していない。

2-4-2 短時間フーリエ変換(Gabor 変換)

2-4-1節で述べたように、フーリエ変換では画像の位置に対する情報が欠落している。では、「フーリエ変換を画像全体に行うのではなく、画像を分割し、分割された場所に対しフーリエ変換を行うことにより、位置情報も含めれば問題ないのではないか?」と考え出されたのが、短時間フーリエ変換(Gabor 変換)である。「分割されることにより画像の局所的な情報を得ることができ、この中にエッジ情報も含まれるのではないか?」と期待できそうなものであるが、都合よくエッジの情報が抽出できるものではない。周波数分解能と空間分解能は、一方が高くなると他方は低下する。これを「不確定性の原理」と言う。すなわち、短時間フーリエ変換では、周波数分解能と空間分解能を両立させることができないことを示している。つまり、画像からエッジ成分を抽出するにはある程度画像を細分化しなくてはならないが、細分化しすぎた場合、周波数に変換してもエッジ成分が含まれる周波数が手に入らない。式(2-5)は2次元の短時間フーリエ変換の式である。

$$F(\omega_1, \omega_2, a, b) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \underbrace{\frac{1}{\sigma^2} \exp\left(-\frac{(x-a)^2 + (y-b)^2}{2\sigma^2}\right)}_{\text{ガウス関数}} \underbrace{\exp(-i\omega_1 x) \exp(-i\omega_2 y)}_{\text{複素正弦波}} \underbrace{f(x, y)}_{\text{原画像}} \quad (2-5)$$

フーリエ変換と同様、 $F(\omega_1, \omega_2, a, b)$ はスペクトルを表し、求められるものは図2-13のように周波数成分に変換されたものが得られる。ただフーリエ変換と異なるところは、 a, b によりフーリエ変換を行う位置を決定しており、正弦波フィルタにガウス関数が掛けられ(畳み込み演算が行われ)抽出される情報が分割されていることである。よって、分割した数だけ周波数成分に変換されたスペクトル画像を得ることができる。これにより、位置情報を得ることになる。

具体的に説明すると、例えば、原画像(図 2-16(a))に、 $\omega_1=\pi/36$, $\omega_2=\pi/18$, $a=100$, $b=100$, $\sigma=100$ の場合の正弦波フィルタ(図 2-16 (b))を掛け(畳み込み演算を行い)、得られた画像(図 2-16 (c))の画素の合計値がスペクトルとなる。また、原画像(図 2-17(a))に、 $\omega_1=\pi/36$, $\omega_2=\pi/18$, $a=100$, $b=100$, $\sigma=50$ の場合の正弦波フィルタ(図 2-17 (b))を掛け(畳み込み演算を行い)、得られた画像(図 2-17 (c))の画素の合計値がスペクトルとなる。

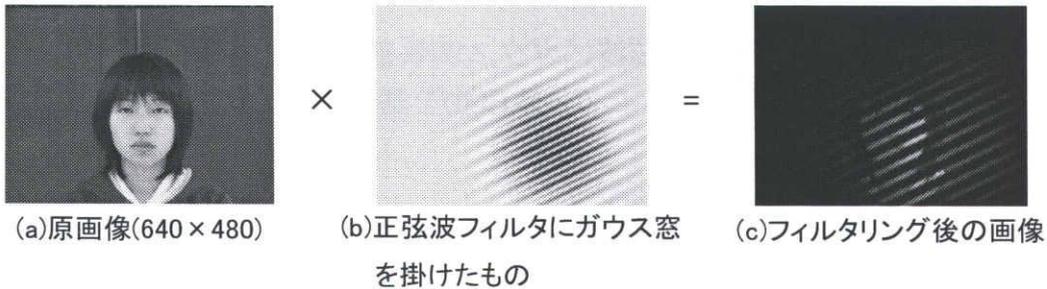


図 2-16 短時間フーリエ変換のスペクトルの求め方(その 1) [$\omega_1=\pi/36$, $\omega_2=\pi/18$, $a=100$, $b=100$, $\sigma=100$]

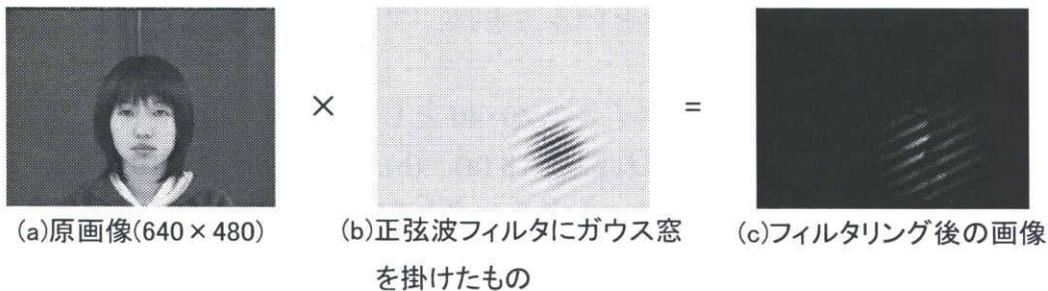


図 2-17 短時間フーリエ変換のスペクトルの求め方(その 2) [$\omega_1=\pi/36$, $\omega_2=\pi/18$, $a=100$, $b=100$, $\sigma=50$]

先に「不確定性の原理」について記したが、周波数分解能が上がるということは図 2-16 (b)の線の数が増えることである。また空間分解能が上がるということは、図 2-17 (b)のガウス窓が小さくなる(円が小さくなる)ということである。図 2-16, 図 2-17より周波数分解能と空間分解能を両立することは不可能であることが解る。

2-4-3 ウェーブレット

ウェーブレット変換は、1つのマザーウェーブレットという基本的な関数を拡大・縮小させることにより、画像の周波数-位置(時間)の解析を行う手法である。

今回使用したマザーウェーブレットは式(2-6)である。パラメータの θ は、マザーウェーブレット $\Phi_{\theta}(x,y)$ をどの程度傾かせるかを決定するものであり、方向成分に相当する。

$$\phi_{\theta}(x,y) = \underbrace{\frac{1}{\sigma^2} \exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)}_{\text{ガウス関数}} \left\{ \underbrace{\exp(i(x\cos\theta + y\sin\theta)) - \exp\left(-\frac{\sigma^2}{2}\right)}_{\text{複素正弦波}} \right\} \quad (2-6)$$

この関数 $\Phi_{\theta}(x,y)$ は、複素正弦波にガウス関数を乗じたものとなっているが、周波数 ω に依存していない。ここで、式(2-6)を基にして、位置方向に拡大・縮小したものが式(2-7)となる。

$$\phi_{k,\theta}(x,y) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2(x^2+y^2)}{2\sigma^2}\right) \left\{ \exp(ik(x\cos\theta + y\sin\theta)) - \exp\left(-\frac{\sigma^2}{2}\right) \right\} \quad (2-7)$$

パラメータの k は、マザーウェーブレット $\Phi_{\theta}(x,y)$ を位置方向に拡大・縮小する比率を決定するものであり、周波数に相当する。また、本来シフトという概念が存在するが、シフトさせるということは、フィルタを移動させることと同様である。

$k=\pi/3$, $\theta=0$, $\sigma=2$, フィルタサイズを 40×40 とした場合の複素正弦波、ガウス関数、ウェーブレットをそれぞれ図 2-18 (a), (b), (c) に示す。また、ウェーブレットを原画像(図 2-18(d)) にフィルタリングしたものを図 2-18 (e) に示す。これより、原画像からエッジが抽出されていることが確認できる。

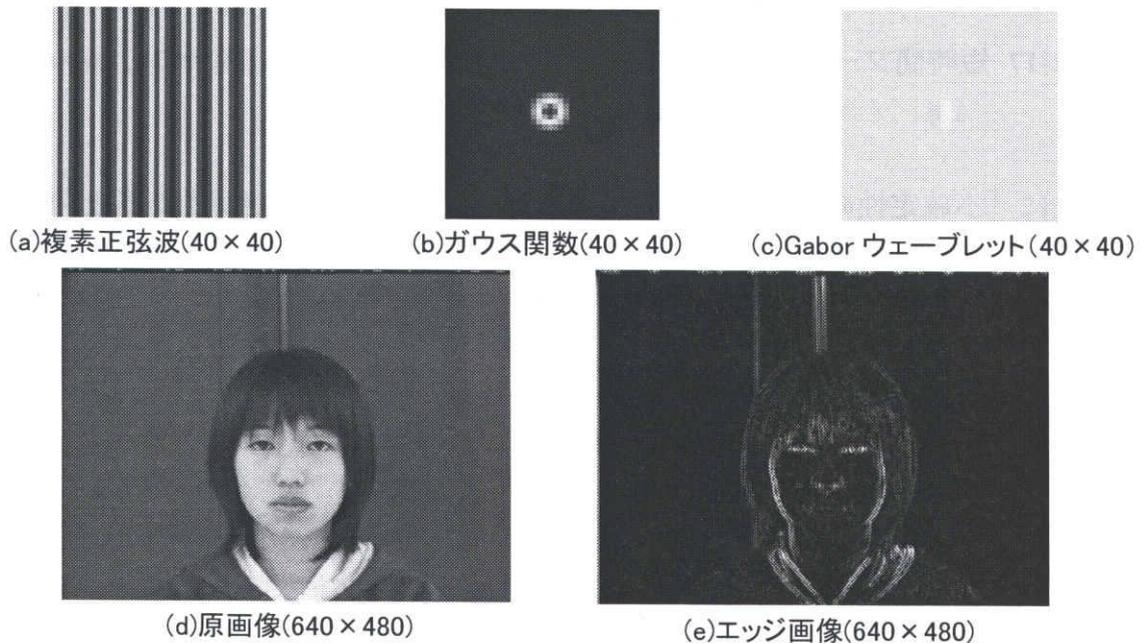


図 2-18 Gabor ウェーブレット変換 [$k=\pi/3$, $\theta=0$, $\sigma=2$, フィルタサイズ= 40×40]

2-5 次元削減

一般に識別機を設計する場合、特徴を増やしすぎるという傾向に陥りやすい。これは、特徴の数を増やせばそれだけ情報量が増え認識率も上昇すると期待することによるものである。これが必ずしも得策ではない理由は次の三点に集約できる。第1に、特徴の数を増やせば増やすほど相関の高い特徴の組が混入する可能性が高まり、期待したほどの効果が得られない。第2に、統計計算に要する計算量は少なくとも次元のべき乗のオーダーになるから、特徴次元の次元の増大は計算量の爆発を引き起こす。第3に、有限個の学習パターンから識別機を設計する際、次元を高くしていくと誤認識率がかえって上昇するという事実がある。このような理由から特徴空間の次元削減を適度に行うことはパターン認識の重要な課題の1つとなっている。

次元を削減する手法としては、テーブルを用いて特徴を選択する特徴選択技術と、行列を用いて特徴ベクトルを変換する次元圧縮技術に分類される。

次元圧縮技術では主成分分析と線形判別分析が重要な技術である。主成分分析は、分散最大基準もしくは原点移動を許した平均二乗誤差最小基準により求まる部分空間、すなわち共分散行列の上位固有値に対応する固有ベクトルを基底とする部分空間を使い、高次元の特徴空間での分布形状を線形の範囲内で最適に近似する方法であり、単純に変数の減少を担うためには最も素直な方法である。この方法は単に分布を近似して低次元に写像する方法であるため、理論的には識別精度の向上はおこりえない。しかし、低次元に写像されたため、学習に必要なデータ数が少なくなり、確率密度関数が再現しやすくなるため現実には識別精度の向上が見られることは珍しくない。

一方、線形判別分析は空間の線形変換によりクラス内分散・クラス間分散比を最大にする部分空間を求める方法であり、識別を考慮した特徴空間の変換法である。このため、積極的に識別に有効な空間に写像する方法であるために、理論的にも実験的にも識別精度は向上する。しかし、学習していないクラスが入力されることが想定される場合には、予測不可能な挙動を起こすことがあるので注意が必要である。

本節では本研究で使用する主成分分析、線形判別分析について簡単に説明する。

2-5-1 主成分分析

KL 展開は、線形空間における特徴ベクトルの分布を最もよく近似する部分空間を求める手法である。具体的には式(2-8)のように特徴ベクトル x を変換行列 A により近似する部分空間上のベクトル y に変換する。

$$y = A^T x \quad (2-8)$$

KL 展開による次元削減法は、分散最大基準と平均二乗誤差最小基準の 2 つの評価基準を用いるが、今回は分散最大基準を用いた KL 展開で次元削減を行う方法を記す。分散最大基準とは射影後の空間でのパターン分布の分散が最大となるものを選択するというものである。射影後の部分空間においてパターンのバラつきが大きければ、その空間は原特徴空間のパターン分布の特徴をより良く保存していると考ええる。

まず与えられたデータの原特徴空間における共分散行列 C を求める。平均ベクトル m 、共分散行列 C は以下の式で求めることができる。

$$m = \frac{1}{n} \sum_{i=0}^n x_i \quad (2-9)$$

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - m)(x_i - m)^t \quad (2-10)$$

n はデータの総数、 x_i は i 番目のデータを示す。次に得られた共分散行列 C の固有値、固有ベクトルを求める。固有値を求めるためには行列の多項式を解かねばならないが、高次の多項式の解を求めるのは非常に困難であるため近似解析手法を用いる。固有値の近次解手法には様々なものがあるが、今回はヤコビ法を用いて近似を行う。固有値を大きいほうから d 個取り出し、これらの固有値に対応する固有ベクトルを d 個取り出す。取り出された d 個の固有ベクトルを変換行列 A とする。処理の流れを図 2-19 に示す。

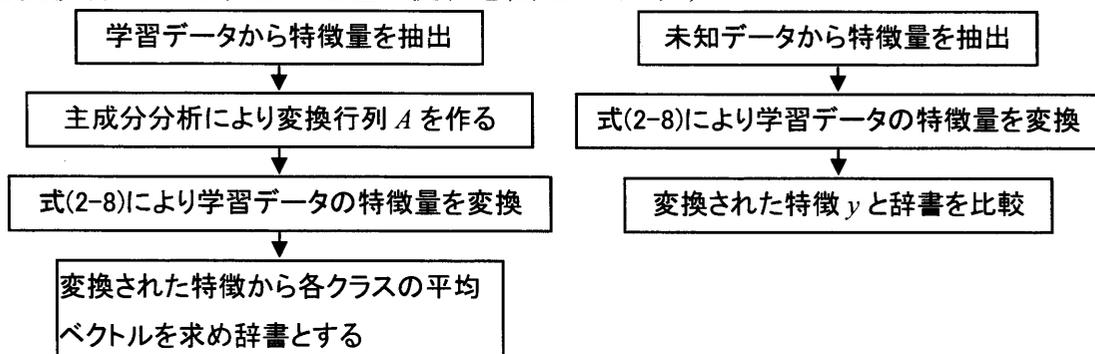


図 2-19 処理の流れ

2-5-2 判別分析法

今、初期特徴ベクトル x の K クラスのサンプル集合 $C_k = \{x\} (k = 1, \dots, K)$ が与えられているものとする。線形判別分析は、より少ない次元でクラス間の分離を最大限強調する線形写像

$$y = A^T x \quad (2-11)$$

を構成する多変量解析手法である。次に記すように、最適な変換行列 A は、固有値問題

$$\sum_B A = \sum_W A \Lambda \quad (2-12)$$

$$A^T \sum_W A = I_N \quad (2-13)$$

の解として求まる。

このとき N は判別空間の次元で、 $N \leq \min(K-1, M)$ である。 M は初期特徴ベクトル x の次元数、 Σ_B および Σ_W は、それぞれクラス間共分散行列およびクラス内共分散行列、 Λ は固有ベクトル、 I_N は $N \times N$ の単位行列である。線形判別分析によって得られた新特徴 y は、与えられたクラス間を最もよく判別する特徴としての意味を持つ。

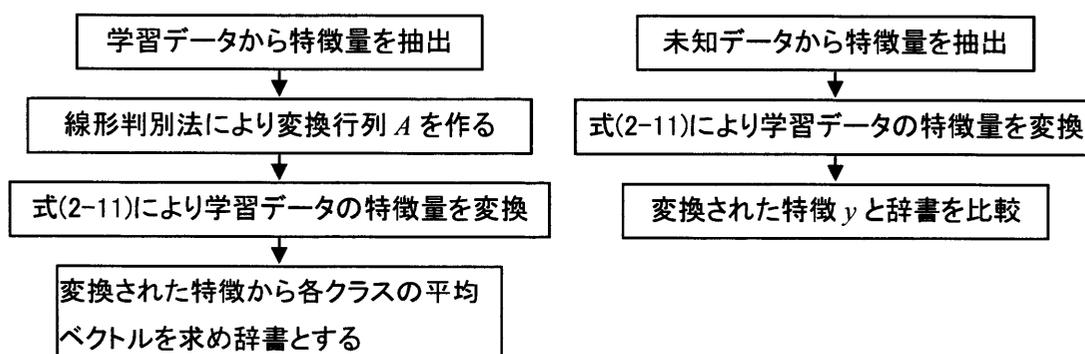


図 2-20 処理の流れ

2-6 識別機

識別機には、平均値からのユークリット距離を用いる1次識別機、マハラビノス距離を用いる2次識別機、k-最近傍法、ニューラルネットワーク[60]、また最近話題になっているサポートベクターマシーン[61]などが挙げられる。

本研究では、特徴量の多様性、正当性の検討が必要となる。このような理由から、処理量が多く、多数のパラメータを含むニューラルネットワーク、サポートベクターマシーンは適していない。また、識別面を作成する際に、マハラビノス距離を用いる2次識別機、k-最近傍法はかなりのサンプル数がないと正確な識別面が得られない可能性がある[20]。現実問題としてサンプルを集める際に、コストがかかるため十分な枚数が得られない、容量が制限されているなど様々な問題がある。そこで、本研究では平均値からのユークリット距離を用いる1次識別機を用いることとする。これは処理が軽く、パラメータが少なく、理解しやすい識別機である。

2-7 その他の識別法

ここまで述べたように、一般的には特徴抽出部、識別部は分かれて認識系が確立されているわけであるが、CLAFIC[62]法、複合類似度法[63]、混合類似度法[63]等の部分空間法は、特徴選択と識別を分離することなく、特徴空間そのものを利用して識別するという手法である。具体的には、各クラスの分布に対し原点移動を許さない平均二乗誤差最小基準によって求まる部分空間、すなわち相関行列の上位固有値に対応する固有ベクトルを基底とする、部分空間が使われる。

2-7-1 部分空間法(CLAFIC 法)

本節で記す部分空間法(CLAFIC 法)は特徴選択と識別を分離することなく、特徴空間そのものを利用して識別するという興味深い手法である。

具体的なクラスの識別法を記す。まず、学習データが1クラス辺り100枚あり、1枚の画像から抽出される特徴次元数は25次元とする。またクラス数は6とする。各画像から特徴量を抽出した後、各画像に対し自己相関行列を作成する。これにより、 25×25 の自己相関行列が各クラス100個作成され、合計600個の自己相関行列が作成される。各クラスに対し、行列の各要素の平均値を算出し、平均自己相関行列 A_i を作成する。これにより 25×25 の平均自己相関行列が6個作成される。6個の平均自己相関行列をそれぞれ固有値展開し、固有

値, 固有ベクトルを求める. 次に未知データに x ついては式(2-14)により各クラスに対する類似度を求める. 求められた類似度の中で最大類似度を示したクラスを結果とする. ここで x は未知データ, i はクラス番号, j は大きい順に固有値を並べたときの固有値の順位, u_{ij} は固有値を λ_{ij} とし, λ_{ij} に対応する固有ベクトル, $S_i(x)$ はクラス i の類似度を示す. 処理の流れを簡単に示したものを図 2-21 に示す.

$$S_i(x) = \sum_{j=1}^d (x' u_{ij})^2 \quad (2-14)$$

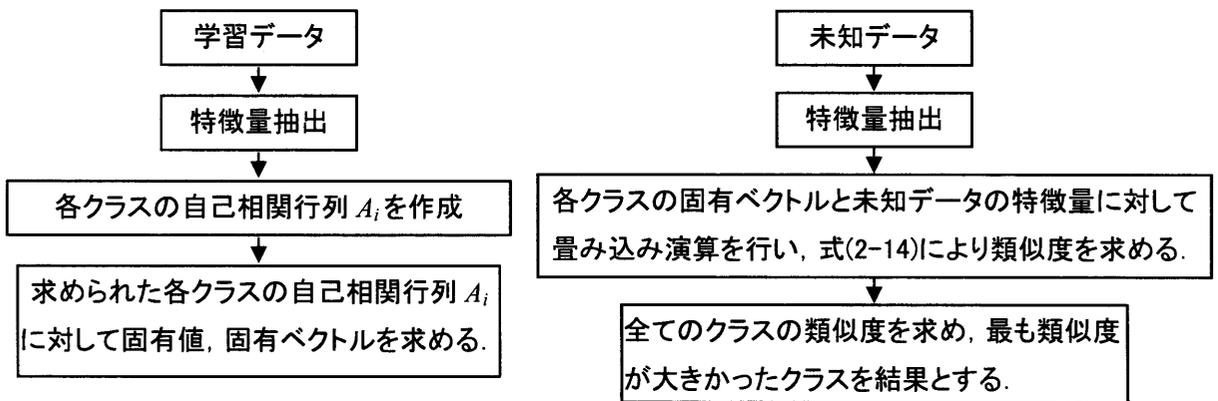


図 2-21 処理の流れ

2-7-2 累積寄与率を用いた部分空間法

固有値展開により得られた固有値から累積寄与率を求め, これにより各クラスの部分空間の次元数を決定する. 累積寄与率は,

$$a(d_i) = \frac{\sum_{j=1}^{d_i} \lambda_{ij}}{\sum_{j=1}^d \lambda_{ij}} \quad (2-15)$$

で求め, 全てのクラスに対して共通なパラメータ(累積寄与率) k を選んで

$$a(d_i) \leq k \leq a(d_i + 1) \quad (2-16)$$

となる次元数 d_i を各クラスが選択するようにした. これにより類似度は式(2-17)によって求められる.

$$S_i(x) = \sum_{j=1}^{d_i} (x' u_{ij})^2 \quad (2-17)$$

具体的なクラスの識別法を記す. まず, 学習データが 1 クラス辺り 100 枚あり, 1 枚の画像から抽出される特徴次元数は 25 次元とする. またクラス数は 6

とする。各画像から特徴量を抽出した後、各画像に対し自己相関行列を作成する。これにより、 25×25 の自己相関行列が各クラス 100 個作成され、合計 600 個の自己相関行列が作成される。各クラスに対し、行列の各要素の平均値を算出し、平均自己相関行列 A_i を作成する。これにより 25×25 の平均自己相関行列が 6 個作成される。6 個の平均自己相関行列をそれぞれ固有値展開し、固有値、固有ベクトルを求める。パラメータ k を決定し、式(2-16)の条件に当てはまる次元数をクラスごとに設定する。次に未知データに x については式(2-17)により各クラスに対する類似度を求める。ここで x は未知データ、 i はクラス番号、 j は大きい順に固有値を並べたときの固有値の順位、 u_{ij} は固有値 λ_{ij} に対応する固有ベクトル、 $S_i(x)$ はクラス i の類似度、 d_i はクラス i の特徴次元数を示す。処理の流れを簡単に示したものを図 2-22 に示す。

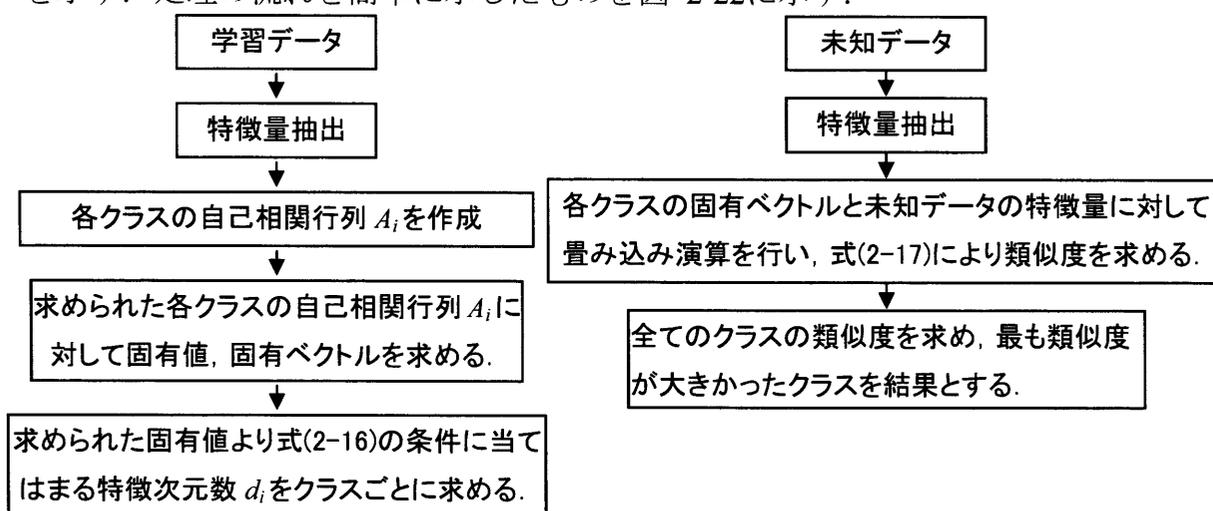


図 2-22 処理の流れ

2-7-3 複合類似度法

文字認識の一手法として複合類似度法がある。この手法は現在では部分空間法の一つの変形として位置付けられている[63]。複合類似度法の定義は式(2-18)で与えられる。

$$S_i(x) = \sum_{j=1}^d \frac{\lambda_{ij}(x'u_{ij})^2}{\lambda_{i1}x'x} \quad (2-18)$$

ここで x は未知データ、 i はクラス番号、 j は大きい順に固有値を並べたときの固有値の順位、 λ_{ij} は固有値、 u_{ij} は固有値 λ_{ij} に対応する固有ベクトル、 $S_i(x)$ はクラス i の類似度を示す。また、この式の分母の $x'x$ はクラスに依存しないので省くことができるが、値を正規化するために導入されている。CLAFIC 法と複

合類似度の違いは、複合類似度では各固有ベクトルに $\lambda_{ij}/\lambda_{i1}$ の係数がかかっていることである。つまり固有値により重みがつけられていることになる。

複合類似度法ではこの類似度を最大とするクラス i を識別結果とする。複合類似度法では基本的には j を d まで変化させるが、一般的に自己相関行列の固有値 λ_{ij} は j が大きくなるに従って急激に小さくなるため、 d の代わりに適当な値 $d_i (< d)$ で打ち切った式

$$S_i(x) = \sum_{j=1}^{d_i} \frac{\lambda_{ij}(x'u_{ij})^2}{\lambda_{i1}x'x} \quad (2-19)$$

で類似度を計算してもその値はほとんど変わらない。実用上は累積寄与率が十分大きくなる次元数 d_i で計算を打ち切っても十分である。実際、累積寄与率が十分に大きくなる d_i を利用すれば結果は次元数 d まで計算した場合とほとんど変わらない。これは d_i の値の選び方は CLAFIC 法ほど識別結果に影響を与えないことを意味している。また適当な次元 d_i で打ち切るにより計算効率は向上する。

具体的なクラスの識別法を記す。まず、学習データが 1 クラス辺り 100 枚あり、1 枚の画像から抽出される特徴次元数は 25 次元とする。またクラス数は 6 とする。各画像から特徴量を抽出した後、各画像に対し自己相関行列を作成する。これにより、 25×25 の自己相関行列が各クラス 100 個作成され、合計 600 個の自己相関行列が作成される。各クラスに対し、行列の各要素の平均値を算出し、平均自己相関行列 A_i を作成する。これにより 25×25 の平均自己相関行列が 6 個作成される。6 個の平均自己相関行列をそれぞれ固有値展開し、固有値、固有ベクトルを求める。パラメータ k を決定し、式(2-16)の条件に当てはまる次元数をクラスごとに設定する。次に未知データに x ついては式(2-19)により各クラスに対する類似度を求める。ここで x は未知データ、 i はクラス番号、 j は大きい順に固有値を並べたときの固有値の順位、 u_{ij} は固有値を λ_{ij} とし、 λ_{ij} に対応する固有ベクトル、 S_i はクラス i の類似度、 d_i はクラス i の特徴次元数を示す。処理の流れを簡単に示したものを図 2-23 に示す。

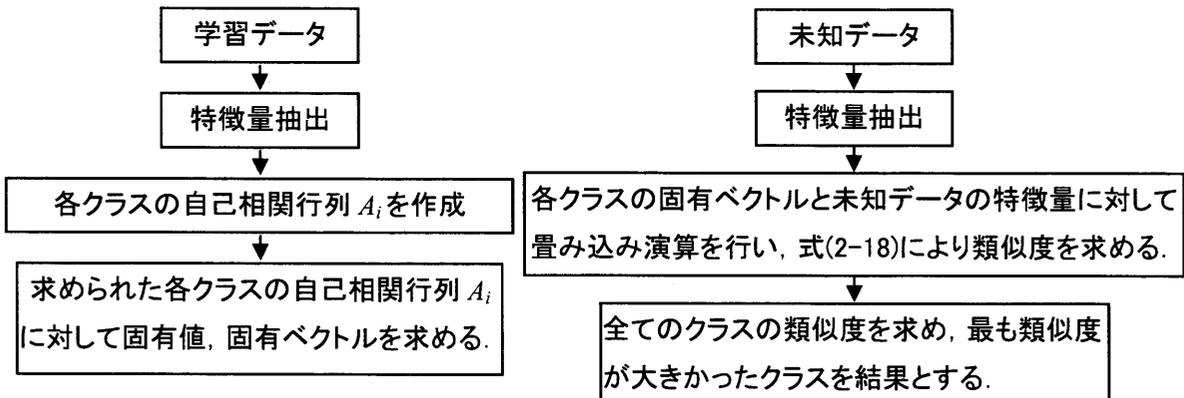


図 2-23 処理の流れ

2-7-4 混合類似度法

複合類似度法や CLAFIC 法では, 存在するクラスのパターンのみからその部分空間を作成する. そのためそのクラスを精度良く表現するには最適な部分空間になっているが, クラス間の判別のためには必ずしも最適な部分空間になっているとは限らない. つまり, 元のデータを分離するのに適した空間とは異なる. 混合類似度法は, 複合類似度法にクラス間の分離機能を導入したものである[63]. 混合類似度法は

$$S_i(x) = \sum_{j=1}^d \frac{\lambda_{ij}}{\lambda_{i1}} \frac{(x^t u_{ij})^2 - u(x^t v_i)^2}{x^t x} \quad (2-20)$$

によって表される. ここで v_i は類似クラス k の平均パターン m_k とクラス i の学習パターン集合との差分を表して

$$v_i = \frac{m_k - \sum_{j=1}^d m_k^t u_{ij} u_{ij}}{\sqrt{m_k^t m_k - \sum_{j=1}^d (m_k^t u_{ij})^2}} \quad (2-21)$$

によって定義される. また x は未知データ, i はクラス番号, j は大きい順に固有値を並べたときの固有値の順位, λ_{ij} は固有値, u_{ij} は λ_{ij} に対応する固有ベクトル, u はパラメータである. 混合類似度の式は複合類似度と同様に, ある適当な次元 d_i で加算を打ち切ることにより計算効率を高めることができる. v_i は, 類似したクラスの平均パターン m_k をクラス i の部分空間に射影したベクトルと, 元のベクトルの差分ベクトルであり, 大きさは 1 に正規化されている. 混合類似度は類似クラス間の差を強調するものであると考えられる.

具体的なクラスの識別法を記す。まず、学習データが1クラス辺り100枚あり、1枚の画像から抽出される特徴次元数は25次元とする。またクラス数は6とする。各画像から特徴量を抽出した後、各画像に対し自己相関行列を作成する。これにより、 25×25 の自己相関行列が各クラス100個作成され、合計600個の自己相関行列が作成される。また各クラスに対し25次元の平均ベクトルも作成しておく。これにより平均ベクトルが6個作成される。各クラスに対し、自己相関行列の各要素の平均値を算出し、平均自己相関行列 A_i を作成する。これにより 25×25 の平均自己相関行列が6個作成される。6個の平均自己相関行列をそれぞれ固有値展開し、固有値、固有ベクトルを求める。次に平均ベクトル、固有ベクトルを用い、式(2-21)より差分ベクトルを求める。パラメータ k を決定し、式(2-16)の条件に当てはまる次元数をクラスごとに設定する。未知データに x については式(2-20)により各クラスに対する類似度を求める。処理の流れを簡単に示したものを図 2-24に示す。

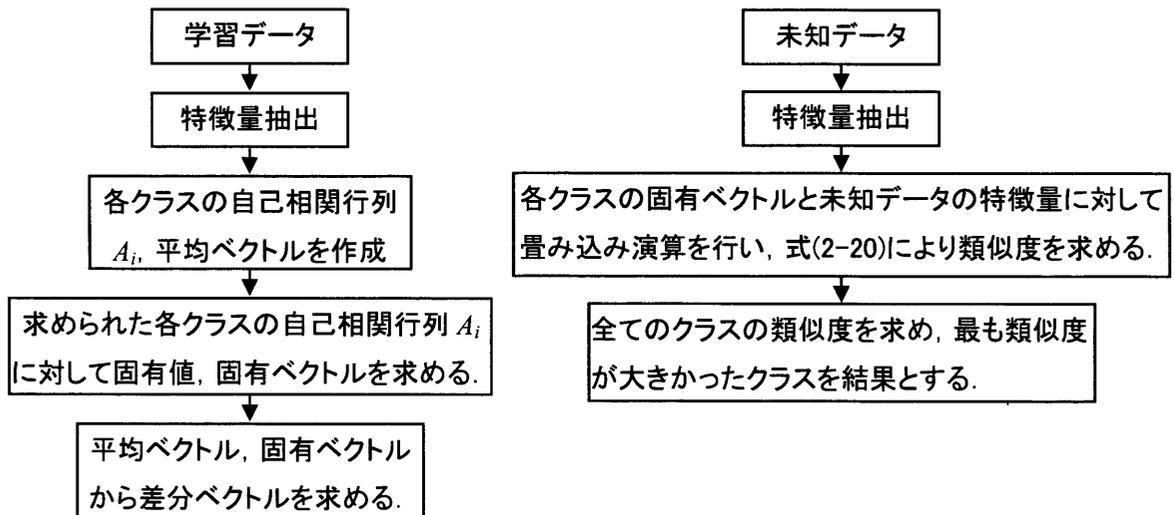


図 2-24 処理の流れ

2-8 特徴空間の評価

一般的なクラス内分散・クラス間分散比の求め方は以下の通りである。

クラス i に属するパターンの集合を X_i とし、 X_i に含まれるパターン数を n_i 、平均ベクトルを m_i とする。また、全パターン数を n 、全パターンの平均ベクトルを m とする。ここで、クラス内分散を σ_w^2 、クラス間分散を σ_b^2 、クラス数を c で表すと

$$\sigma_W^2 = \frac{1}{n} \sum_{i=1}^c \sum_{x \in \chi_i} (x - m_i)' (x - m_i) \quad (2-22)$$

$$\sigma_B^2 = \frac{1}{n} \sum_{i=1}^c n_i (m_i - m)' (m_i - m) \quad (2-23)$$

となる。すなわちクラス内分散はクラスの平均的な広がりを表し、クラス間分散はクラス間の広がりを表している。従ってそれらの比

$$J_\sigma = \frac{\sigma_B^2}{\sigma_W^2} \quad (2-24)$$

を定義すれば、 J_σ が大きいほど優れた特徴であると判別することができる。 J_σ はクラス内分散・クラス間分散比である。これ以降 J_σ を分散比と呼ぶ。これはクラス内距離で正規化したクラス間距離と見ることもできる。しかしこの方法の欠点として、必ずしも実際の分布の分離度を反映していないことが上げられる。これは分布の重なり具合を評価していないためである。

認識率の良さに加えて、さらに特徴空間上でクラスの重なりがなければ特徴は高い信頼を持つことができると考えられる。そこで本研究ではクラスタ判別法[64]を用い特徴空間の可視化を行い、観察を行う。これは原特徴空間で構造解析を行い、このクラスタ構造を保存し低次元空間への写像を求め可視化を行う手法である。原特徴空間でクラスの重なりがあった場合は、根本的にクラスの分離が困難であることが予測される。そこで、各パターンの特徴空間での程度重なりが存在するかを観察し、原特徴量がもっている情報を把握する。

3 家電制御支援システムへの応用

我々は、下半身が不自由な病人や高齢者に対する介護・看護をサポートするシステムの構築と、システムを用いたパターン認識の多様性の研究を目的とし、またその適用例として利用者自身の身体の一部を用いた家電製品制御システムを作成し研究を多数行ってきた[65][66].

利用者自身の身体の一部を用いた家電製品の制御としては、ジェスチャ認識[67]～[71]，ポーズ認識，音声認識[72]～[76]，視線検出[77]～[79]などが考えられる。

ジェスチャ認識，ポーズ認識を行う際に，マーカー等のデバイスを用いることが考えられる。しかし，高齢者はマーカー等のデバイスを携帯させたとしてもはずしてしまう可能性がある。またジェスチャ認識の場合では，認識に失敗すると一連の動作を最初からやり直さなければならない。このため高齢者にとって動きのあるジェスチャより，静止しているポーズの方が身体への負担が少ないと考えられる。

また，音声認識の場合は，TV のように音の出るものを操作するとき，利用者の声が TV の音にかき消されてしまう可能性がある。高齢者や障害者の人が大声を出すことは困難であるため現実的ではない。

視線検出の場合は，高齢者は老眼鏡などの眼鏡をかけている場合が多く，このような場合視線を抽出することは困難である。

家電製品を制御するためには，利用者が伝えたいことを確実にシステムに伝えることが重要である。このため，静止して確実に命令を送ることができるポーズを用いることが適していると考えられる。またポーズを認識する際にマーカー等のデバイスを用いる代わりに，肌色を用いることにした。これにより高齢者に存在を意識されないマーカーが実現できる。

ここでは，本システムで用いた特徴量抽出，識別手法に主眼をおき，識別機の検討，高次局所自己相関特徴の性質，特徴空間の性質について考察してみる。具体的には，識別方法の検討，学習枚数についての検討，高次局所自己相関特徴の傾きに対する検討，特徴量抽出方法についての検討，辞書作成方法の検討を行った。

識別方法の検討であるが，画像から抽出された特徴量が最適であったとしても，抽出された特徴ベクトルの各要素の相関が高い，必要以上の特徴次元数を抽出してしまい特徴空間を形成できないといった問題も起こりえる。そこで，本システムに最適な識別機について検討を行う。

学習枚数の検討についてであるが，学習データから有効かつ安定した判別空間を構築するためには，1 枚の画像から抽出される特徴次元数の数倍以上の枚

数が学習データとして必要であるといわれている。そこで、本研究では実際に存在しない学習データを作り出し、データを増やすことを目的とする。

高次局所自己相関特徴の傾きに対する検討についてであるが、高次局所自己相関特徴は傾きに対して弱いことが知られている。そこで画像に幾学的な変換を施し傾いた画像を作成し、仮想的に傾かせた画像を学習させることにより、傾きにも対応できるような特徴空間の作成を目的とする。

特徴量抽出方法の検討についてであるが、平行移動に強いことは高次局所自己相関特徴の利点ではある。しかし、次のような欠点も含んでいる。例えば、特徴量が少ししか存在しない場合、高次局所自己相関特徴は空間上の位置情報は保持していないため、ある物体の形や向きを変えずに位置を変えただけならば同じものとして認識されてしまう。そこで、大まかな相対位置の情報を取り込んだ特徴量抽出方法についての考察を目的とする。

辞書作成法の検討についてであるが、大幅に異なる特徴量が存在しても、同じクラスとして取りまとめた場合があるとする。しかし、大幅に異なる特徴量を1つのクラスとするため、クラス内分散が大きくなりすぎてしまい、クラス間分散が小さくなってしまふことが考えられる。このような特徴空間が作成されてしまうと、判別が困難になってしまう。そこで複数の辞書作成法を提案し、特徴空間の分離度等を考察してみることを目的とした。

3-1 システム構成

今回制御する家電製品はテレビ、扇風機の2つを対象とした。本システムは人物のポーズを入力するカメラ、ポーズの識別及び家電製品制御コマンドを赤外線学習リモコンに送るコンピュータ、現在制御できる家電製品が何であるかを示すメニューを表示させるためのディスプレイからなる。赤外線学習リモコンには、あらかじめ家電製品の制御信号を学習させておいた。システムの構成を図3-1に示す。

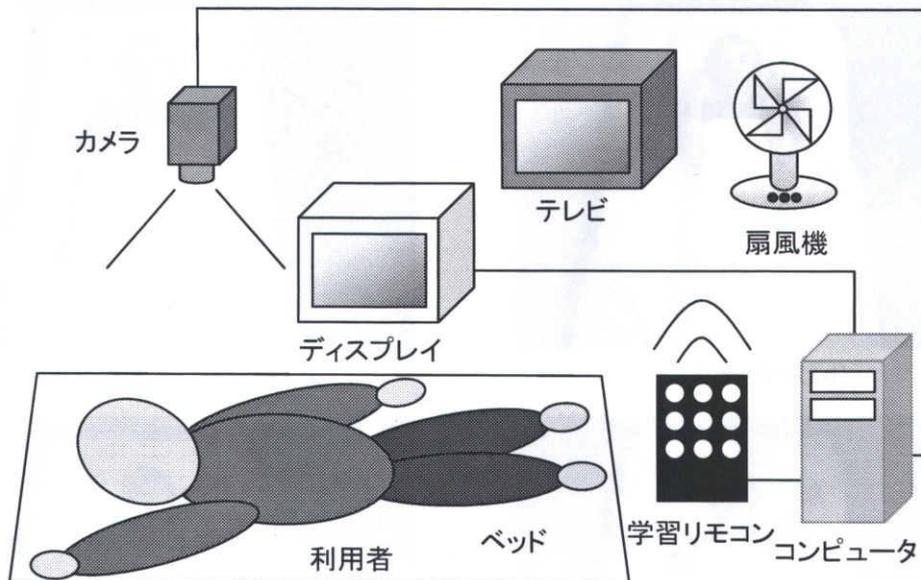
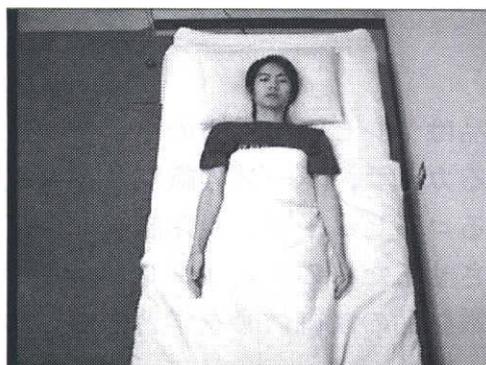


図 3-1 システム構成

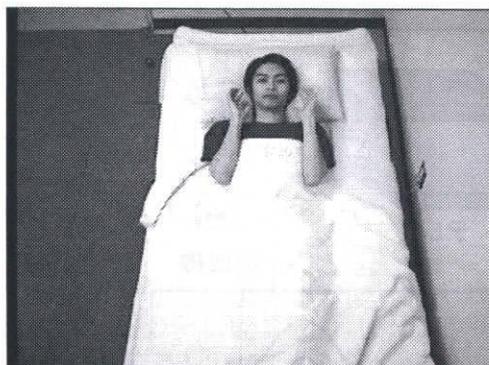
3-2 命令実行ポーズ

各種の命令を実行するために割り当てられたポーズを図3-2に示す。利用者はベッドの上に横になった状態で家電製品を操作することができる。以下に図3-2に示すポーズをまとめる。

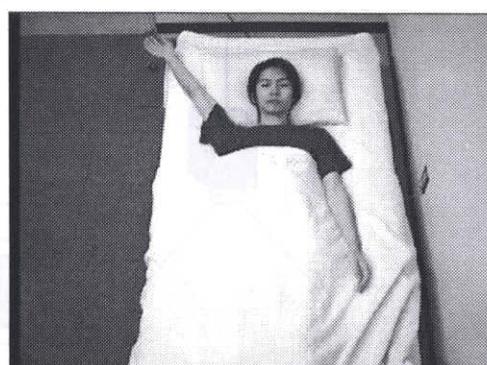
- (a) ベッドにただ横たわっているだけの状態(ノーマルポーズ)
- (b) 身体の正面で両腕を折りたたみ両手を両耳に翳した状態(スイッチオン)
- (c) 右腕を右斜め上にまっすぐ伸ばした状態(チャンネルアップ)
- (d) 右腕を右斜め上にまっすぐ伸ばしながら肘より先を左斜め上方向に曲げた状態(チャンネルダウン)
- (e) 右腕を身体の正面で折りたたむようにして手を耳にかざすような状態(ボリュームアップ)
- (f) 右腕を身体の右側で折り曲げ耳をふさぐような状態(ボリュームダウン)
- (g) 両腕を使って両耳をふさぐような状態(スイッチオフ)



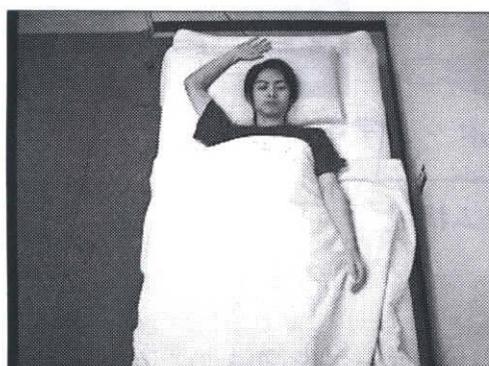
(a) ノーマルポーズ



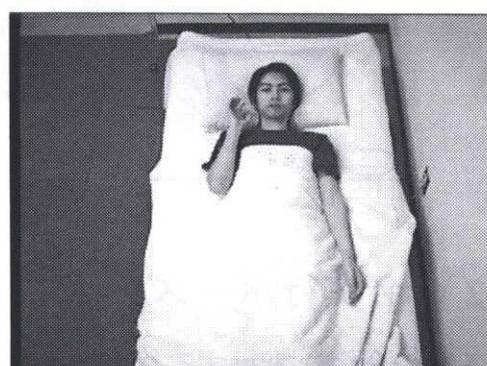
(b) スイッチオン



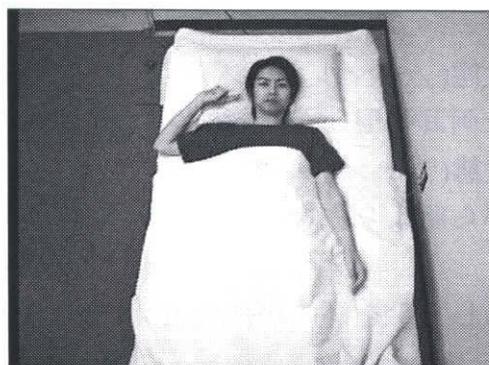
(c) チャンネルアップ



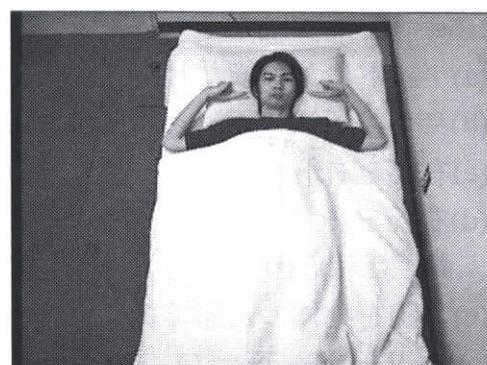
(d) チャンネルダウン



(e) ボリュームアップ



(f) ボリュームダウン



(g) スイッチオフ

図 3-2 命令実行ポーズ

3-3 ポーズ識別の流れ

ポーズ識別の流れを図 3-3に示す。まずベッドの上に横たわった利用者をカメラで捉え、画像を取り込む。得られた画像から肌色領域を抽出し、頭の重心点と利用者の顔、腕または手の部分を抜き出す。頭の重心点からは顔周辺矩形を作成し、利用者の顔、腕または手の部分からはエッジを抽出する。エッジ画像に顔周辺矩形を当てはめ、高次局所自己相関特徴を顔周辺矩形内から抽出する。高次局所自己相関特徴を抽出した後、識別機により、あらかじめ用意されている学習辞書との類似度をを比較することによりポーズ識別を行う。

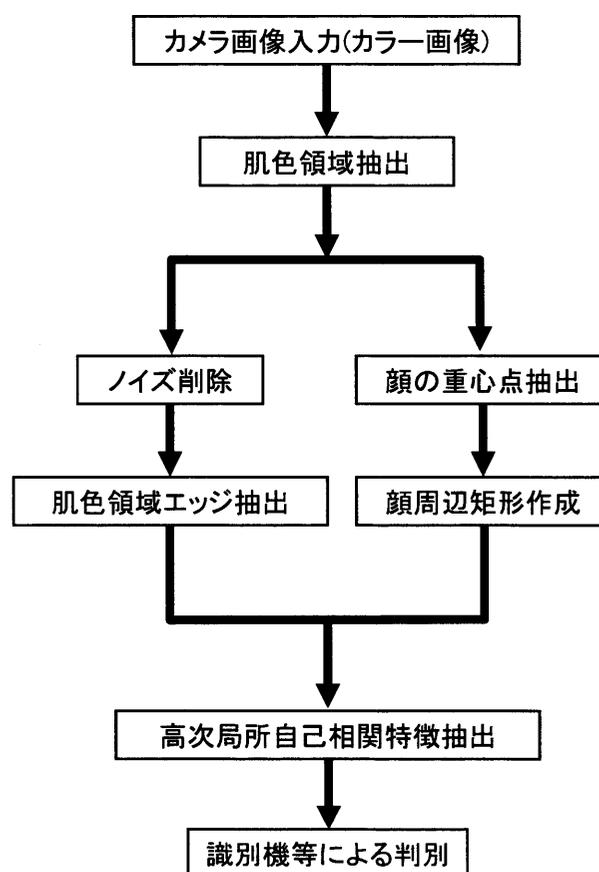


図 3-3 ポーズ識別の流れ

3-3-1 肌色候補領域の抽出

ポーズ認識を行う場合、マーカーを用いて認識させることが考えられるが、マーカーを失ってしまった場合に制御ができなくなってしまう。そこで特殊な機器なしでの認識を目的とするため、ここでは失うことが無いと肌色を用いることにした。肌色領域抽出は個人差や照明条件などに影響されるため、それに対応する方法が必要であり、多くの研究が行われている[81]~[83]。

3-3-1-1 R 値, B 値を用いる手法

リアルタイム処理を行う場合、計算コストが少ないほど実用性がある。そこで、背景が単純である場合は、高速に肌色候補領域を抜き出すことができる RGB 画像の R 値と B 値を用いた処理を行う。R 値-B 値の処理[84]をすることによって肌色らしい領域の画素値が比較的高く、それ以外の領域の画素値が低い画像が得られる。閾値を設定して 2 値化処理を行うことにより肌色領域とその他の領域とに分割することができる。この後で記す 3-4 節で用いられる実験データはこの手法を用い、肌色を抽出している。閾値は固定値 $30(=R-B)$ で設定した。図 3-4 (a)に入力画像を、(a)に対して肌色候補領域抽出を行った画像を (b)に示す。

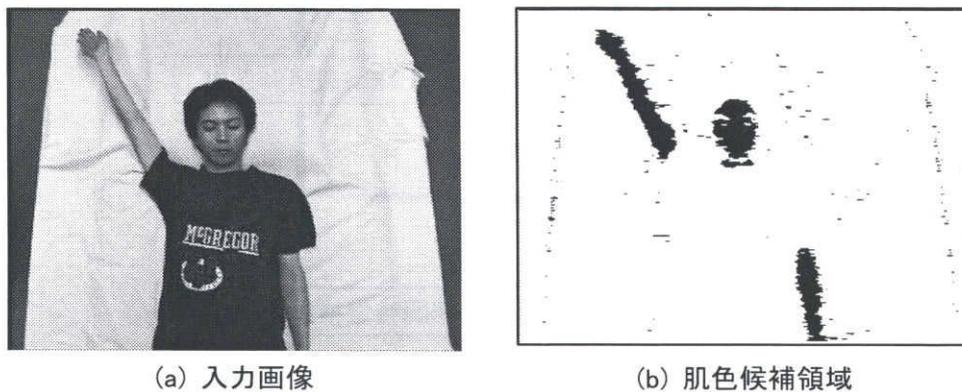


図 3-4 肌色領域抽出結果

3-3-1-2 U 値, V 値を用いる手法

3-3-1-1節の処理は、背景が単純な場合、高速に肌色領域を抜き出すことができるが、固定閾値を設定して 2 値化処理を行っているため、照明条件等が変化した場合、その度に閾値を調整しなくてはならない。そこで、均等知覚色空間の 1 つである LUV 表色系において、基準肌色 UV 値を設定し、ダイナミックに閾値を変化させる方法[85]がある。

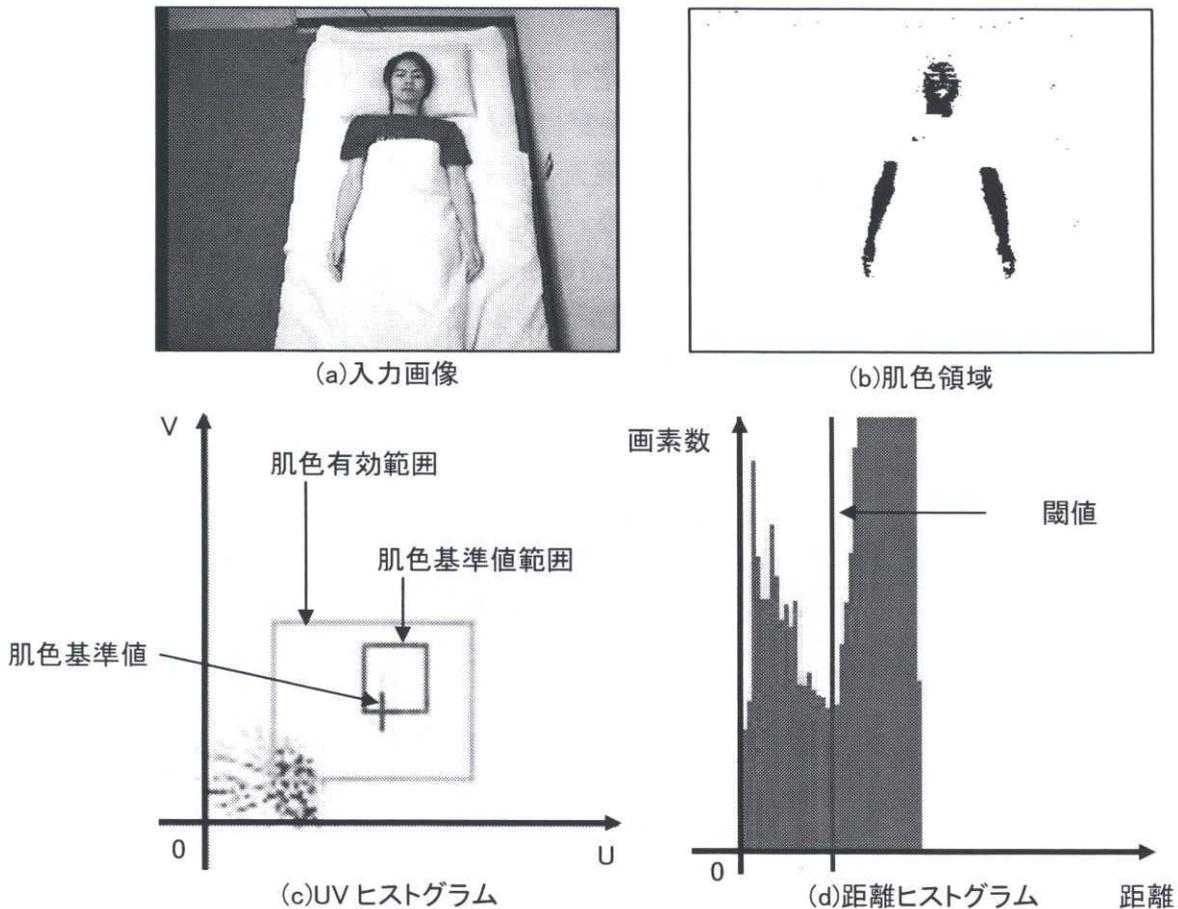


図 3-5 肌色領域抽出結果

LUV 表色系を用いた方法は、入力画像の各画素の RGB 値を LUV 表色系の U, V 値へ変換し、2次元の UV ヒストグラムを作成する。次に予め定めた肌色基準値範囲内で画素数が最大の U, V 値を肌色基準 U, V 値とする。次に、予め定めた肌色領域範囲内の U, V 値と肌色基準 U, V 値との距離を計算する。肌色基準値範囲及び肌色領域範囲の決定方法は後に示す。

次に、肌色基準 U, V 値からの距離の画像のヒストグラムを作成して判別分析法により 2 値化し、肌色領域を抽出する。図 3-5 (a)に入力画像を、図 3-5 (b)に抽出された肌色領域を、図 3-5 (c)に UV ヒストグラムを、図 3-5 (d)に距離ヒストグラムを示す。図 3-5 (c)の横軸は U 値、縦軸は V 値を示し、2次元ヒストグラム中に示されている大きい四角で囲まれている領域が肌色有効範囲、小さい四角で囲まれている領域が肌色基準値範囲、十字の中心は肌色基準値を示す。また、図 3-5 (d)の横軸は肌色基準値からの距離、縦軸は画素数を示し、1次元ヒストグラム中の縦線は閾値を表している。入力画像は夜、ブラインドを閉めた状態で部屋の蛍光灯をつけた状況で撮影されたものであり、環境は一定となっている。

● 肌色有効範囲について

肌色基準値範囲及び肌色領域範囲を決定するために、以下のようなデータを用意した。環境は以下の2種類を用意し、撮影は晴れた日の8時から16時まで、一時間おきに50枚ずつ撮影を行った。環境光としては窓から差し込んでくる日光、天井に設置してある蛍光灯の光がある。

各時間帯で、上記の2つの環境の照度を10回ずつ測定した。照度を測定した位置は、ベッドに人が寝た場合に顔の位置が来る位置に固定し照度を測定した。上記の2つの環境における、各時間帯の照度平均を表3-1に示す。1行目にかかっているものは撮影した環境を、1番左の列にかかっている数値は撮影した時間を、それ以外に書かれている数値の単位はlx(ルクス)である。

表 3-1 照度平均

	ブラインドあり	ブラインドなし
8時	557.1	644.1
9時	577.6	831.4
10時	555.9	735.1
11時	561.6	682.4
12時	552.5	656.8
13時	549.6	668.3
14時	550	704.1
15時	547.8	633.8
16時	541.7	576.6

表3-1より部屋の照度は、541.7～831.4 lx まで変化している。これらの環境上で肌色を抽出できるような肌色基準値範囲を決定することとした。

今回は肌色基準有効範囲として $135 < U < 150$, $125 < V < 140$, 肌色有効範囲として $115 < U < 160$, $110 < V < 145$ とした。このように肌色基準有効範囲と肌色有効範囲を設定した場合の肌色抽出率を表3-2に示す。1行目に書かれているものは、撮影した環境を、1番左の列に記されている数値は照度を測定した時間を、表内の括弧の左に記されている数値は抽出率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。抽出できたかどうかの判断は目視により確認した。

表 3-2 肌色抽出率

	ブラインドあり	ブラインドなし
8時	100 (50/50)	100 (50/50)
9時	100 (50/50)	100 (50/50)
10時	100 (50/50)	100 (50/50)
11時	100 (50/50)	100 (50/50)
12時	100 (50/50)	100 (50/50)
13時	100 (50/50)	100 (50/50)
14時	100 (50/50)	100 (50/50)
15時	100 (50/50)	100 (50/50)
16時	98 (49/50)	100 (50/50)

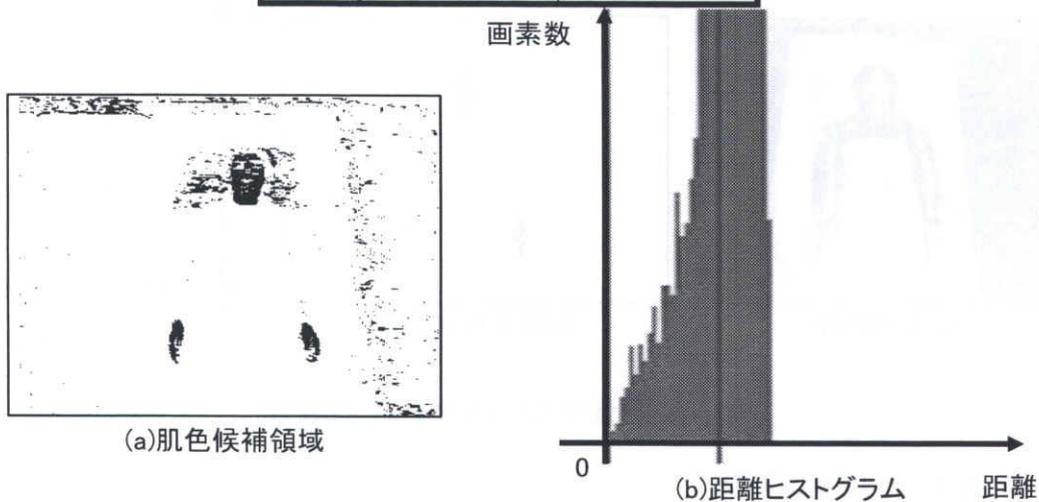


図 3-6 肌色候補領域抽出失敗例

ブラインドありの環境で、肌色抽出に失敗した例を図 3-6に示す。図 3-5 (d)では1次元ヒストグラムの山が2つできているが、図 3-6(b)は1つしかできていない。これにより、判別分析法によりうまく閾値が設定できなかったためである。ブラインドありの場合は環境光が蛍光灯のみとなるため、基本的に環境光が一定になる。このためほぼ 100%肌色候補領域が抽出されている。ブラインドなしの場合は表 3-1より照度平均が、576.6~831.4 lx と大幅に変化している。これは、日光が環境光として加わっているためである。しかし 100%肌色領域が抽出されている。表 3-1の中で最も平均照度の低かった場合を図 3-7に、最も平均照度の高かった場合を図 3-8に示す。図 3-7 (b), 図 3-8 (b)より、安定して肌色候補領域が抽出でき、図 3-7(c), 図 3-8(c)の UV ヒストグラムの形状もさほど変化は見られない。これらの結果より、今回設定した肌色有効範囲と肌色基準値範囲の場合、照度が 540~830 lx 程度は、UV ヒストグラムが安定し、安定して肌色領域が抽出できることが確認できる。



図 3-7 照度平均が 540 程度の場合



図 3-8 照度平均が 830 程度の場合

3-3-2 顔周辺矩形の作成

命令実行ポーズは顔付近において顔と腕によって形成されている。従って、顔からはなれた部分における情報はあまり意味を持たない。そこで必要な情報のみを用いるために、顔付近の領域を切り出して特徴抽出を行うのが有効であると思われる。3-3-1節で求められた肌色領域に対してメディアンフィルタを適応し、ごま塩ノイズを削除する。次に膨張処理を行い、肌色でないために抜け落ちた目や口の部分を埋める。さらに収縮処理を行い、肌色領域を削っていくことにより、比較的小さな肌色領域である手や腕の領域や、その他のノイズを除去する。そこで残された肌色領域の重心を算出する。しかし肌色領域の重心点は髪の毛の部分を考慮していないため、実際の頭の重心点と誤差が生じる。このため求められた重心点の値から縦軸方向に一定の値を引くことにより調整する。これによりここで得られた点は頭部領域の重心点であると考えられる。この点を基に顔周辺矩形を作成する。肌色領域を収縮処理により削った画像に対して顔周辺矩形を当てはめた画像を図 3-9に示す。

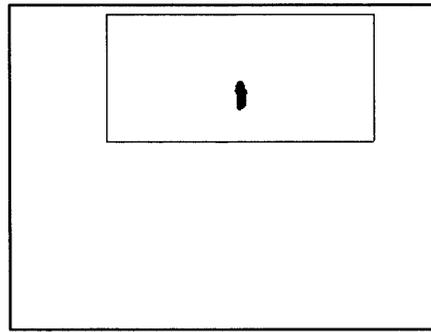


図 3-9 顔周辺矩形作成

3-3-3 エッジ画像の作成

3-3-1節で求められた肌色領域に対してメディアンフィルタを適応しごま塩ノイズを削除する。次に、膨張処理を行い、肌色でないために抜け落ちた目や口といった部分を埋める。さらに収縮処理を行い、ノイズを削っていくことにより利用者の顔や手、腕の領域を抜き出す。この画像を図 3-10(a)に示す。この画像に対して差分型の Prewitt オペレータを用いてエッジ抽出を行う。得られたエッジ画像に3-3-2節で得られた顔周辺矩形をあてはめる。図 3-10(a)よりエッジ抽出を行い、顔周辺矩形を当てはめたものを図 3-10(b)に示す。図 3-10(b)より実際に抽出される高次局所自己相関特徴の値を表 3-3に示す。また、各ポーズのエッジ画像に顔周辺矩形を当てはめたものを図 3-11に示す。顔周辺矩形内から高次局所自己相関特徴を抽出することによって、矩形内の対象の位置の変動に対する、平行移動の不変性の利点が最大限に発揮される。

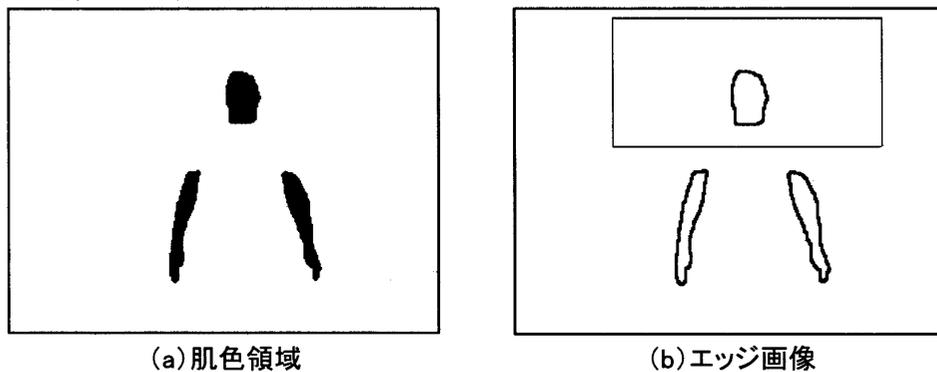


図 3-10 顔と腕の領域

表 3-3 実際の特徴ベクトル

特徴番号	1	2	3	4	5	6	7	8	9
特徴量	0.014	0.010	0.009	0.011	0.009	0.006	0.003	0.009	0.003
特徴番号	10	11	12	13	14	15	16	17	18
特徴量	0.005	0.006	0.006	0.005	0.005	0.006	0.006	0.005	0.008
特徴番号	19	20	21	22	23	24	25		
特徴量	0.006	0.008	0.004	0.008	0.006	0.008	0.004		

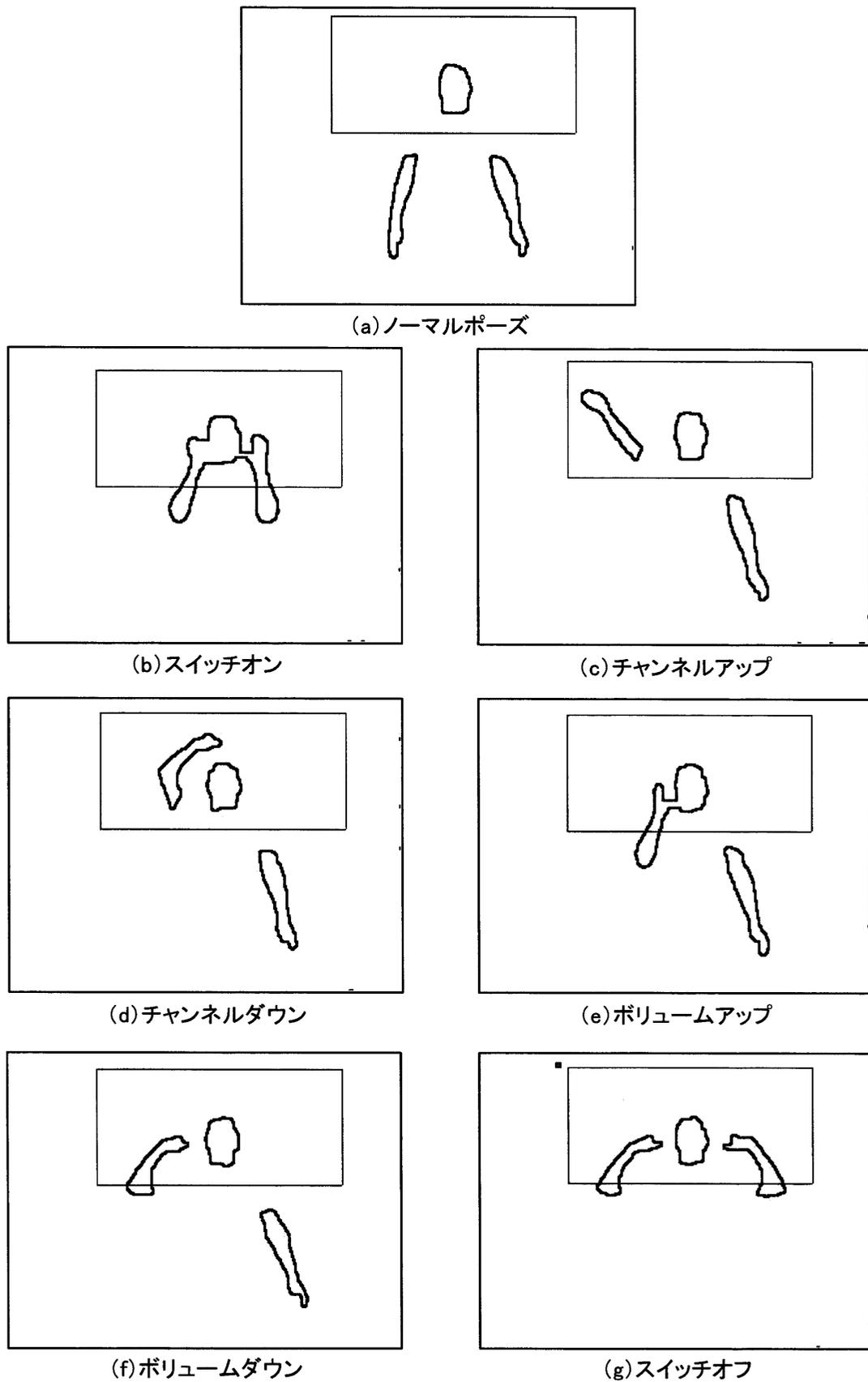


図 3-11 命令実行ポーズの肌色エッジに顔周辺矩形を当てはめたもの

3-4 識別機の検討

どの識別機が最適な判別を行うか検討した。実験に用いる画像データは、研究室に設置したベッドの上に利用者が横になった状態で、天井に固定されたカメラにより撮影した。画像サイズは320×240pixelである。ここでは利用者を男性一人とした。使用したカメラはSONY製 CCD-Z1、キャプチャカードはインタウェア製 PicPon Linux、カメラとキャプチャカードをコンポジットケーブルにより直接接続し、画像をPCにキャプチャした。

撮影環境は室内で撮影を行い、外部からの環境光は天井に常備されている蛍光灯のみである。被験者は学習データ、未知データ共に半袖を着用している。学習データは図 3-2に示されているポーズで(b)～(g)までのポーズ(スイッチオン、チャンネルアップ、チャンネルダウン、ボリュームアップ、ボリュームダウン、スイッチオフ)6種類のデータを各100枚、合計600枚用意した。未知データとして、異なる時間に撮影を行い、異なる服装を着用したものを学習データと同様に合計600枚用意した。これらの画像から図 3-3に示される方法にて、25次元の高次局所自己相関特徴を抽出した。但しこれらのデータを撮影した際に、環境光はほぼ一定であるため、3-3-1-1節の手法を用い肌色抽出を行った。学習データ、未知データの一例としてチャンネルアップのポーズを図 3-12に示す。

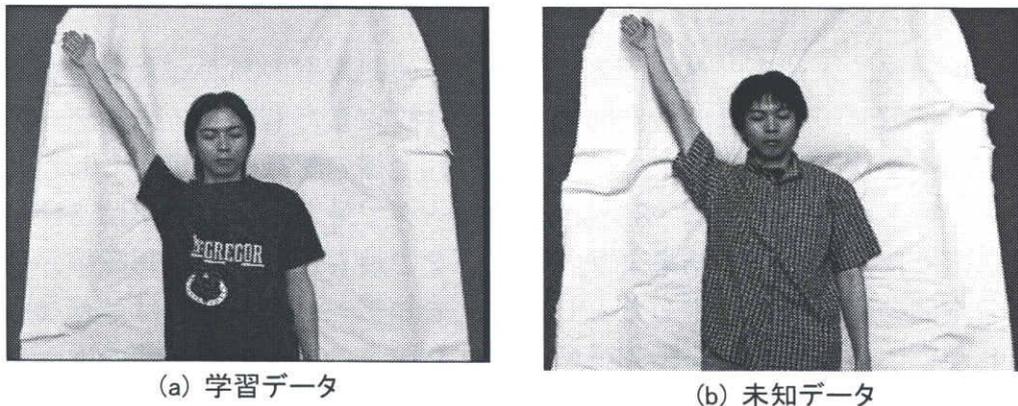


図 3-12 実験データ

3-4-1 部分空間法を用いた識別実験

3-4節で述べられたデータに対して部分空間法を適用し識別能力に関して検討した。学習データに対する認識率を表 3-4、未知データに対する認識率を表 3-5に示す。特徴次元数と記されている行にある数値は、類似度を求める際に利用した u_{ij} の個数(部分空間の次元数)、認識率と記されている行にある数値は認識率(正解枚数/総枚数[=600]、単位は%)を示す。

表 3-4 学習データに対する認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	93.5	91.3	88.5	96.7	97.3	97.5	98.7	96.3	98.2
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	96.2	96.0	96.2	94.7	93.2	90.0	84.5	82.5	79.3
特徴次元数	19	20	21	22	23	24	25		
認識率	75.8	70.5	68.7	55.2	45.7	17.2	16.7		

表 3-5 未知データに対する認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	80.1	81.3	65.5	75.7	79.8	77.3	79.3	79.7	75.0
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	68.5	68.2	63.3	64.5	63.8	68.7	64.0	62.7	65.7
特徴次元数	19	20	21	22	23	24	25		
認識率	63.3	56.3	54.5	43.5	56.2	16.7	16.7		

学習データ、未知データの両方で最も全体の認識率の高かったものを塗りつぶした。学習データでは部分空間を7次元に設定したものが、未知データでは部分空間を2次元に設定したものがそれぞれ全体の認識率が良い。しかし学習データで最も高い認識率は98.7%、未知データで最も高い認識率は81.3%と低い結果だった。またこの結果より学習データで最も良い全体の認識率が得られた部分空間の次元数に設定したとしても、未知データに対しては最適な結果が得られないことがいえる。誤認識例は様々存在するが、部分空間の次元を高くするにつれて、どのポーズもチャンネルダウンのポーズと誤認識する枚数が多くなってしまい、最終的に部分空間を25次元に設定したときには学習データ、未知データともに全てのポーズをチャンネルダウンと誤認識してしまった。

3-4-2 累積寄与率を用いた識別実験

学習データから算出された各ポーズに対する累積寄与率を表 3-6に示す. 表 3-6の中で一番左の列にかかっている数値は, 類似度を求める際に利用した u_{ij} の個数(部分空間の次元数)を示す.

表 3-6 各ポーズの累積寄与率

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
1	0.999773268	0.999730385	0.999831553	0.999721001	0.999742450	0.999789075
2	0.999876616	0.999855174	0.999931047	0.999861908	0.999876316	0.000873656
3	0.999943627	0.999936812	0.999965884	0.999945931	0.999951578	0.999943713
4	0.999968346	0.999985762	0.999990326	0.999980525	0.999984891	0.999988831
5	0.999992091	0.999990382	0.999992849	0.999986333	0.999990099	0.999991707
6	0.999993862	0.999993483	0.999994512	0.999990165	0.999992711	0.999994006
7	0.999995311	0.999995416	0.999995978	0.999992665	0.999994312	0.999995671
8	0.999996508	0.999996500	0.999996905	0.999994479	0.999995547	0.999996609
9	0.999997504	0.999997447	0.999997708	0.999995963	0.999996620	0.999997484
10	0.999998222	0.999998185	0.999998220	0.999996903	0.999997591	0.999998288
11	0.999998741	0.999998591	0.999998664	0.999997730	0.999998187	0.999998891
12	0.999999082	0.999998984	0.999998967	0.999998287	0.999998615	0.999999188
13	0.999999385	0.999999233	0.999999251	0.999998635	0.999998985	0.999999390
14	0.999999613	0.999999398	0.999999459	0.999998923	0.999999264	0.999999536
15	0.999999722	0.999999555	0.999999627	0.999999184	0.999999509	0.999999658
16	0.999999820	0.999999679	0.999999738	0.999999411	0.999999642	0.999999757
17	0.999999881	0.999999785	0.999999829	0.999999589	0.999999757	0.999999828
18	0.999999919	0.999999867	0.999999899	0.999999733	0.999999840	0.999999884
19	0.999999949	0.999999921	0.999999943	0.999999831	0.999999906	0.999999932
20	0.999999972	0.999999956	0.999999959	0.999999892	0.999999937	0.999999960
21	0.999999986	0.999999973	0.999999974	0.999999937	0.999999957	0.999999972
22	0.999999993	0.999999986	0.999999983	0.999999961	0.999999973	0.999999982
23	0.999999997	0.999999993	0.999999990	0.999999976	0.999999985	0.999999990
24	0.999999999	0.999999998	0.999999996	0.999999988	0.999999997	0.999999996
25	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000

3-4-1節の実験で学習データに対して全体の認識率が 100%になったセルは濃く塗りつぶした. また認識率が 90%以上になったセルを薄く塗りつぶした. この結果より各ポーズの部分空間の次元数は, スイッチオンは 1~23 次元, チャンネルアップは 1~21 次元, チャンネルダウンは 3 次元以外, ボリュームアップも 3 次元以外, ボリュームダウンは 4~7 次元と 9 次元, スイッチオフは 1, 2 次元と 5~14 次元のどれかに設定することが望ましい. これらの次元数に対応する累積寄与率を見てみると, 0.999998 より小さい値を持つ次元数を各ポーズの部分空間の次元数に設定することが良いと考えられる. これをもとに部分

空間の次元数を求めたところ、表 3-7 のようになった。表 3-7 の特徴次元数を用いた学習データの識別結果と未知データの識別結果を表 3-8 に示す。表 3-8 一番左の列はクラスの種類を、表内の括弧の左に記されている数値は認識率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。

表 3-7 各クラスにおける部分空間の次元数

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
次元数	9	9	9	11	10	9

表 3-8 認識率

	学習データ	未知データ
スイッチオン	100.0(100/100)	79.0(79/100)
チャンネルアップ	100.0(100/100)	100.0(100/100)
チャンネルダウン	97.0(97/100)	69.0(69/100)
ボリュームアップ	100.0(100/100)	91.0(91/100)
ボリュームダウン	97.0(97/100)	69.0(69/100)
スイッチオフ	99.0(99/100)	49.0(49/100)
全体	98.8(593/600)	76.2(457/600)

学習データの全体の認識率は 98.8%、未知データの全体の認識率は 76.2%であった。ポーズによっては認識率が向上したものも見られたが、全体の認識率が3-4-1節の結果よりも悪くなっており、累積寄与率が必ずしもそのまま認識率に反映しないことが確認できる。

次に累積寄与率の値を実際に変化させてみて累積寄与率が0.999998前後の認識率を調べた。学習データの認識率を表 3-9に、未知データの認識率を表 3-10に示す。表 3-9、表 3-10内の数値は認識率(正解枚数/総枚数[=600]、単位は%)を示す。

表 3-9 累積寄与率を用いた場合の学習データに対する認識率

累積寄与率	0.9997	0.9998	0.9999	0.99991	0.99992	0.99993	0.99994
認識率	0	81.5	95.8	90.5	90.5	90.5	90.5
累積寄与率	0.99995	0.99996	0.99997	0.99998	0.99999	0.999991	0.999992
認識率	85.8	86.7	88.5	84.7	91.2	95.8	96.5
累積寄与率	0.999993	0.999994	0.999995	0.999996	0.999997	0.999998	0.999999
認識率	98.0	98.0	98.0	99.0	97.7	98.8	97.3
累積寄与率	0.9999991	0.9999992	0.9999993	0.9999994	0.9999995		
認識率	96.5	97.3	97.5	97.7	93.8		

表 3-10 累積寄与率を用いた場合の未知データに対する認識率

累積寄与率	0.9997	0.9998	0.9999	0.99991	0.99992	0.99993	0.99994
認識率	0	74.0	73.5	82.0	82.0	82.0	82.0
累積寄与率	0.99995	0.99996	0.99997	0.99998	0.99999	0.999991	0.999992
認識率	68.8	59.5	65.5	54.7	71.0	73.5	77.8
累積寄与率	0.999993	0.999994	0.999995	0.999996	0.999997	0.999998	0.999999
認識率	77.7	77.0	76.7	79.2	79.2	76.2	71.8
累積寄与率	0.9999991	0.9999992	0.9999993	0.9999994	0.9999995		
認識率	69.7	66.1	68.5	66.8	65.1		

学習データ、未知データの両方で最も全体の認識率の高かったものを塗りつぶした。これらの結果を見ると、表 3-9で最も認識率が高くなった時のパラメータの値は 0.999996 であるのに対し、表 3-10で最も認識率が高くなった時のパラメータの値は 0.99991~0.99994 の間であり、学習データで高い認識率が得られたパラメータを用いたからとしても未知データに対して良い結果が得られるとは限らない。

3-4-3 複合類似度法を用いた識別実験

処理の流れは図 2-22に示す方法で行った。全てのクラスについて部分空間の次元を 1~25 次元まで変化させて実験を行ったが、学習データ、未知データともに認識率にまったく変化は見られなかった。結果を表 3-11に示す。一番左の列はクラスの種類を、表内の括弧の左に記されている数値は認識率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。

表 3-11 認識率

	学習データ	未知データ
スイッチオン	97.0(97/100)	60.0(60/100)
チャンネルアップ	100.0(100/100)	100.0(100/100)
チャンネルダウン	90.0(90/100)	71.0(71/100)
ボリュームアップ	95.0(95/100)	84.0(84/100)
ボリュームダウン	87.0(87/100)	100.0(100/100)
スイッチオフ	92.0(92/100)	70.0(70/100)
全体	93.5(561/600)	80.8(485/600)

類似度の変化がほとんどないためにこのような結果になったと考えられる。そこで、学習データとして使用した画像の中から、スイッチオフのポーズをした画像を 1 枚入力データとして選び類似度の検証を行った。上位 1 個の固有値に対応する固有ベクトルのみ使用した場合に、式(2-19)により求められた類似度を表 3-12の 2 行目に、25 個全ての固有値に対応する固有ベクトルを使用した場合に、式(2-19)により求められた類似度を表 3-12の 3 行目に示す。同様に、

未知データとして使用した画像の中から、スイッチオフのポーズをした画像を1枚入力データとして選び類似度の検証を行った。上位1個の固有値に対応する固有ベクトルのみ使用した場合に、式(2-19)により求められた類似度を表3-13の2行目に、25個全ての固有値に対応する固有ベクトルを使用した場合に、式(2-19)により求められた類似度を表3-13の3行目に示す。

表 3-12 学習データの類似度

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
1	0.976936260	0.999578743	0.999578743	0.984628786	0.998872045	0.999900458
25	0.976938366	0.971703727	0.999578756	0.984629985	0.998872133	0.999900463

表 3-13 未知データの類似度

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
1	0.970188188	0.966300597	0.999426894	0.978666203	0.998078481	0.999653825
25	0.970190881	0.966303359	0.999426918	0.978667895	0.998078620	0.999653840

この結果を見てみると類似度は小数点以下第5位辺りからしか変化が見られない。この原因としては、複合類似度法では固有値により重みが増えられているが、この固有値が2番目に大きい値以降が極端に小さくなってしまい限りなく0に近い値になってしまっていることが考えられる。例として表3-14に学習データのスイッチオフの固有値を降順にソートしたものを示す。

表 3-14 学習データのスイッチオフの固有値

次元数	固有値	次元数	固有値	次元数	固有値
1	0.001480545826	10	0.000000001824	19	0.000000000105
2	0.000000297560	11	0.000000001335	20	0.000000000050
3	0.000000144524	12	0.000000001169	21	0.000000000028
4	0.000000064753	13	0.000000000864	22	0.000000000023
5	0.000000007041	14	0.000000000599	23	0.000000000012
6	0.000000005275	15	0.000000000431	24	0.000000000002
7	0.000000004670	16	0.000000000260	25	0.000000000001
8	0.000000003778	17	0.000000000165		
9	0.000000002756	18	0.000000000133		

表3-14から解るように2次元目以降を類似度に加えたとしてもほとんど類似度の値が増えていない。これより固有値に重みをつけたとしても結果がほとんど変わらないため、累積寄与率が十分に大きくなる d_i で計算を打ち切った方が良いと考えられる。

3-4-4 混合類似度法を用いた識別実験

本システムの場合でチャンネルダウンのポーズとボリュームダウンのポーズの全体的な形状はきわめて類似している。そのため画像全体を均等に観測するような複合類似度法を用いた場合次のような問題が起きると予想される。チャンネルダウンのポーズをした画像を入力データとした場合のボリュームダウンクラスに対する類似度と、ボリュームダウンのポーズをした画像を入力データとした場合のチャンネルダウンクラスに対する類似度は、近い値をとることになる。これは複合類似度法では両者の間で誤認識を生ずる可能性が高いことを示している。ここで、チャンネルダウンとボリュームダウンの違いは右腕を折り曲げる角度である。つまりこの画像中で、この部分に着目したような類似度を定義すれば両者の分離が改善されることが予想される。

3-4-4-1 学習データに対する識別実験

本来、類似クラスはクラスごとに異なるものを設定するべきであるが、各クラスの類似クラスがきちんと定義できないため、類似クラスは固定して、学習データに対する識別実験を行った。またパラメータの値は1に設定した。類似クラスをスイッチオンにした場合の結果を表 3-15に、類似クラスをチャンネルアップにした場合の結果を表 3-16に、類似クラスをチャンネルダウンにした場合の結果を表 3-17に、類似クラスをボリュームアップにした場合の結果を表 3-18に、類似クラスをボリュームダウンにした場合の結果を表 3-19に、類似クラスをスイッチオフにした場合の結果を表 3-20に示す。特徴次元数と記されている行にある数値は、類似度を求める際に利用した u_{ij} の個数(部分空間の次元数)、認識率と記されている行にある数値は認識率(正解枚数/総枚数[=600]、単位は%)を示す。

表 3-15 類似クラスをスイッチオンに設定した場合の学習データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	93.8	93.2	94.5	94.8	94.0	93.2	94.3	94.3	94.3
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	94.3	94.3	94.3	94.3	94.3	94.3	94.3	94.3	94.3
特徴次元数	19	20	21	22	23	24	25		
認識率	94.3	94.3	94.0	94.0	93.5	93.5	16.7		

表 3-16 類似クラスをチャンネルアップに設定した場合の学習データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	93.2	93.3	94.5	93.5	93.5	93.8	93.5	93.5	93.5
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.5
特徴次元数	19	20	21	22	23	24	25		
認識率	93.5	93.5	93.5	93.5	93.5	93.5	16.7		

表 3-17 類似クラスをチャンネルダウンに設定した場合の学習データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	93.0	94.3	95.0	94.3	94.8	93.8	94.0	93.5	93.5
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.8
特徴次元数	19	20	21	22	23	24	25		
認識率	93.8	93.8	93.5	93.5	93.5	93.5	99.0		

表 3-18 類似クラスをボリュームアップに設定した場合の学習データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	93.0	93.2	94.5	94.7	94.7	94.7	94.2	93.8	93.8
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	93.8	93.8	93.8	93.8	93.8	93.8	93.8	93.8	93.8
特徴次元数	19	20	21	22	23	24	25		
認識率	93.8	93.8	93.5	93.5	93.5	93.5	89.7		

表 3-19 類似クラスをボリュームダウンに設定した場合の学習データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	92.8	94.3	95.7	94.7	94.7	94.8	93.5	93.5	93.5
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.5
特徴次元数	19	20	21	22	23	24	25		
認識率	93.5	93.5	93.5	93.5	93.5	93.5	0.0		

表 3-20 類似クラスをスイッチオフに設定した場合の学習データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	93.3	93.2	93.7	93.8	93.8	94.3	93.5	93.5	93.5
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.5	93.5
特徴次元数	19	20	21	22	23	24	25		
認識率	93.5	93.5	93.5	93.5	93.5	93.5	16.7		

チャンネルダウン，ボリュームアップ以外のクラスは，獲得する次元数を 25 次元に設定すると急激に認識率が下がる．この理由は式(2-21)の分母の計算をするときに $\sqrt{\quad}$ の中がマイナスになってしまう，もしくは 0 になっていることが原因である．これにより差分ベクトルが計算できない．また，表 3-16，

表 3-19, 表 3-20の 7~24 次元までの認識率に注目すると獲得する次元数が増えていても各ポーズの認識率が一定になっている。この理由としては式(2-20)の中で固有値により重みをつけているが、固有値は獲得する次元数を多くすればするほど対応する固有値の値は限りなく 0 に近づいていくため、類似度の値がほとんど変化しなくなってしまう。この現象は3-4-3節の複合類似度法にも見られた。ただ複合類似度法のように全ての特徴次元数に対して認識率が一定になることはなかった。

また類似クラスを変更したとしても、各次元に対する各ポーズの認識率および全体の認識率にさほど影響が見られなかった。そこで、学習データとして使用した画像の中からスイッチオフのポーズをした画像を 1 枚入力データとして選び、上位 1 個の固有値に対応する固有ベクトル使用した場合の類似度と、上位 24 個の固有値に対応する固有ベクトルを使用した場合の類似度を調べた。25 個全ての固有ベクトルを使用した場合を調べなかった理由は、25 次元の類似度は差分ベクトルが計算できない場合があるため類似度が出てこない場合が存在するためである。結果をそれぞれ表 3-21と表 3-22に示す。表 3-21の中で一番左の列に書かれているものは設定した類似クラスを、1 行目に示されているものは学習されているクラスの種類を示している。表 3-22の 1 行目に示されているものは学習されているクラスの種類を示す。上位 24 個固有値に対応する固有ベクトルを使用した場合は、全ての類似クラスに関して求められる類似度が等しかったため類似度の値のみ示した。

これらの結果より獲得する次元数を高く設定した場合、類似クラスを適当に設定しても良いことになる。

表 3-21 類似クラスを設定し、上位 1 個の固有ベクトルを使用した場合の学習データの類似度

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
スイッチオン	0.976787382	0.966422330	0.999299430	0.973343180	0.998453467	0.999887713
チャンネルアップ	0.976887011	0.961474060	0.999210863	0.983231265	0.998138679	0.999900181
チャンネルダウン	0.940942019	0.933153354	0.999274769	0.957887512	0.996450223	0.999612696
ボリュームアップ	0.957931256	0.957406401	0.999309256	0.984626647	0.998502574	0.999880597
ボリュームダウン	0.954585259	0.943798578	0.999568229	0.970019787	0.998857092	0.999846351
スイッチオフ	0.953959221	0.943501422	0.999236450	0.969337404	0.997789283	0.999883034

表 3-22 類似クラスを設定し、上位 24 個の固有ベクトルを使用した場合の未知データの類似度

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
類似度	0.976936696	0.971696943	0.999578080	0.984621496	0.958869538	0.999900186

3-4-4-2 未知データに対する識別実験

次に未知データに対する識別実験を行った。またパラメータの値は 1 に設定した。類似クラスをスイッチオンにした場合の結果を表 3-23に、類似クラスを

チャンネルアップにした場合の結果を表 3-24に, 類似クラスをチャンネルダウンにした場合の結果を表 3-25に, 類似クラスをボリュームアップにした場合の結果を表 3-26に, 類似クラスをボリュームダウンにした場合の結果を表 3-27に, 類似クラスをスイッチオフにした場合の結果を表 3-28に示す. 特徴次元数と記されている行にある数値は, 類似度を求める際に利用した u_{ij} の個数(部分空間の次元数), 認識率と記されている行にある数値は認識率(正解枚数/総枚数 [=600], 単位は%)を示す.

表 3-23 類似クラスをスイッチオンに設定した場合の未知データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	77.0	75.3	73.3	79.8	80.2	81.0	80.8	80.5	79.5
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	79.0	80.8	81.2	80.8	80.8	81.7	81.7	81.7	80.8
特徴次元数	19	20	21	22	23	24	25		
認識率	80.8	80.8	81.3	81.3	81.3	81.3	16.7		

表 3-24 類似クラスをチャンネルアップに設定した場合の未知データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	76.3	74.3	72.7	80.5	81.8	81.7	81.5	82.0	81.7
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	81.8	83.0	83.0	83.7	83.3	83.2	83.3	83.8	81.7
特徴次元数	19	20	21	22	23	24	25		
認識率	81.7	81.7	81.7	81.7	81.3	81.3	16.7		

表 3-25 類似クラスをチャンネルダウンに設定した場合の未知データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	79.0	82.2	76.7	81.3	81.0	80.7	81.0	79.7	79.7
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	79.6	80.7	81.7	81.5	82.0	82.5	82.0	82.5	81.3
特徴次元数	19	20	21	22	23	24	25		
認識率	80.8	80.8	80.8	80.8	80.8	80.8	57.3		

表 3-26 類似クラスをボリュームアップに設定した場合の未知データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	73.8	73.5	89.0	80.2	80.2	80.3	79.8	79.0	79.3
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	79.3	80.2	80.5	80.8	80.8	80.8	80.8	80.8	80.8
特徴次元数	19	20	21	22	23	24	25		
認識率	80.8	81.3	81.3	81.3	81.3	81.3	74.0		

表 3-27 類似クラスをボリュームダウンに設定した場合の未知データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	79.3	85.6	73.5	80.5	80.0	80.3	80.8	81.2	81.5
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	80.7	81.2	82.0	82.0	82.0	82.0	82.3	82.3	80.8
特徴次元数	19	20	21	22	23	24	25		
認識率	80.8	80.8	80.8	81.3	81.3	81.3	0.0		

表 3-28 類似クラスをスイッチオフに設定した場合の未知データの認識率

特徴次元数	1	2	3	4	5	6	7	8	9
認識率	77.8	80.7	69.5	80.8	81.0	81.0	81.0	81.0	81.3
特徴次元数	10	11	12	13	14	15	16	17	18
認識率	81.7	80.3	80.8	80.8	80.8	81.7	81.7	81.7	81.3
特徴次元数	19	20	21	22	23	24	25		
認識率	81.3	81.3	81.3	81.3	81.3	81.3	16.7		

未知データの場合は学習データのとくと異なり獲得する次元数をかなり多くしないと各ポーズの認識率が一定にならなかった。しかしこのこと以外は、学習データのとくと同じような現象が起きた。次に前節と同様に未知データとして使用した画像の中からスイッチオフのポーズをした画像を1枚入力データとして選び、上位1個の固有値に対応する固有ベクトルを使用した場合の類似度と、上位24個の固有値に対応する固有ベクトルを使用した場合の類似度を調べた。25個全ての固有ベクトルを使用した場合を調べなかった理由は、25次元の類似度は差分ベクトルが計算できない場合が存在するためである。結果をそれぞれ表3-29と表3-30に示す。一番左の列に書かれているものは設定した類似クラスを、1行目に示されているものは学習されているクラスの種類を示している。

表 3-29 類似クラスを設定し、上位1個の固有ベクトルを使用した場合の学習データの類似度

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
スイッチオン	0.969980351	0.962395939	0.999416612	0.964463471	0.998071603	0.999340912
チャンネルアップ	0.970179839	0.951629345	0.999403018	0.977483383	0.997856146	0.999427135
チャンネルダウン	0.954016441	0.943457683	0.999575514	0.969411712	0.998156232	0.999881262
ボリュームアップ	0.947463768	0.952781930	0.999405866	0.978647004	0.998078186	0.999327179
ボリュームダウン	0.942295840	0.934305478	0.999136180	0.959258137	0.998037257	0.999653606
スイッチオフ	0.940409896	0.932722846	0.999159800	0.957352143	0.996505842	0.999509316

表 3-30 類似クラスを設定し、上位 24 個の固有ベクトルを使用した場合の未知データの類似度

	スイッチオン	チャンネルアップ	チャンネルダウン	ボリュームアップ	ボリュームダウン	スイッチオフ
スイッチオン	0.970190311	0.966289146	0.999426918	0.978078280	0.998078280	0.999651265
チャンネルアップ	0.970190326	0.966289141	0.999426918	0.978662916	0.998078280	0.999651265
チャンネルダウン	0.970190329	0.966289146	0.999426918	0.978662916	0.998078280	0.999651265
ボリュームアップ	0.970190329	0.966289146	0.999426918	0.978662929	0.998078280	0.999651265
ボリュームダウン	0.970190329	0.966289146	0.999426918	0.978662916	0.998078283	0.999651256
スイッチオフ	0.970190329	0.966289146	0.999426918	0.978662916	0.998078280	0.999650951

1 個の固有ベクトル使用した場合の類似度と、24 個の固有ベクトルを使用した場合の類似度は、共に前節の結果と異なり、類似クラスが違くと類似度も異なった。しかしほとんど差がないため獲得する次元数を高く設定した場合、類似クラスを適当に設定しても良いことになってしまう。

3-4-5 線形判別分析

まず、学習データより係数行列を作成する。次に学習データから得られた原特徴を、係数行列を用いて写像変換し新特徴を作成する。次に新特徴よりクラスごとの平均ベクトルを求める。これを学習辞書とする。

未知の画像が与えられた場合の識別は、その画像の特徴を写像変換することにより新特徴を算出し、学習辞書とのユークリット距離を比較し最も近いクラス平均ベクトルを持つクラスに決定するという方法を用いて実現できる。

3-4節に示されている学習データ、未知データに対して識別実験を行った。学習データ、未知データに対して認識率はともに 100%であった。

この結果より、これ以降の実験では線形判別分析を使用する。またここから先の実験では実環境に近づけた画像を用いて実験を行うため、肌色抽出は 3-3-1-2節で示した手法を用いる。

3-5 仮想学習データを用いた学習

学習データから有効かつ安定した判別空間を構築するためには、1枚の画像から抽出される特徴次元数の数倍以上の枚数が学習データとして必要であるといわれている。しかし、学習データを集める際に、現実問題としてコストがかかる、容量が制限されている、利用者に負荷がかかる、集められるデータ数が限られているなど、十分な枚数が得られないような様々な問題がある。そこで、本研究では実際に存在する学習データから仮想学習データを作り出し、データを擬似的に増やすことを提案する。今回は、原画像に幾学的な傾きを与えることによって、仮想学習データを作成した。この仮想学習データを実在する学習データに加え、新たに辞書を構築することで認識率が上がることを示す。

3-5-1 判別空間が安定する枚数

ここで用いられている特徴次元数は、25次元なので必要な学習データの枚数は最低でも50枚以上が目安となる。実際に判別空間が安定するにはどの程度学習枚数が必要になるかを検討した。

実験に用いる画像データは、研究室に設置したベッドの上に利用者が横になった状態で、天井に固定されたカメラにより撮影した。画像サイズは320×240ピクセルである。ここでは利用者を男性一人とした。使用したカメラはSONY EVI-G20、キャプチャカードはI/Oデータ製GV-VCP2M/PCI、カメラとキャプチャカードをコンポジットケーブルにより直接接続し、画像をPCにキャプチャした。

学習データは次のようにして用意した。最初は図3-2に示されている各ポーズを2枚で合計14枚、次に、各ポーズ3枚で合計21枚というように、学習データの数を徐々に増やしていき、それぞれ辞書を作成する。未知データとしては、利用者が学習データを撮影した位置とほぼ同じ位置に横たわり、学習データと異なる時間に撮影したものを、各ポーズ100枚、合計700枚用意した。図3-13(a)に学習データの例を、(b)に未知データの例をそれぞれ示す。

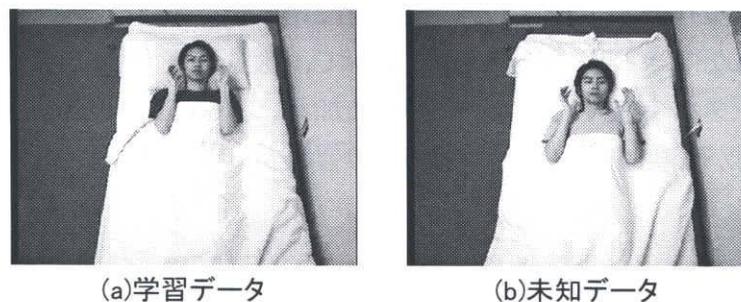


図 3-13 実験に用いるデータ

実験の結果を図 4-2 に示す。グラフの横軸は辞書を作成するときに用いた学習データの枚数を示し、縦軸は認識率(正解枚数/総枚数)を示す。

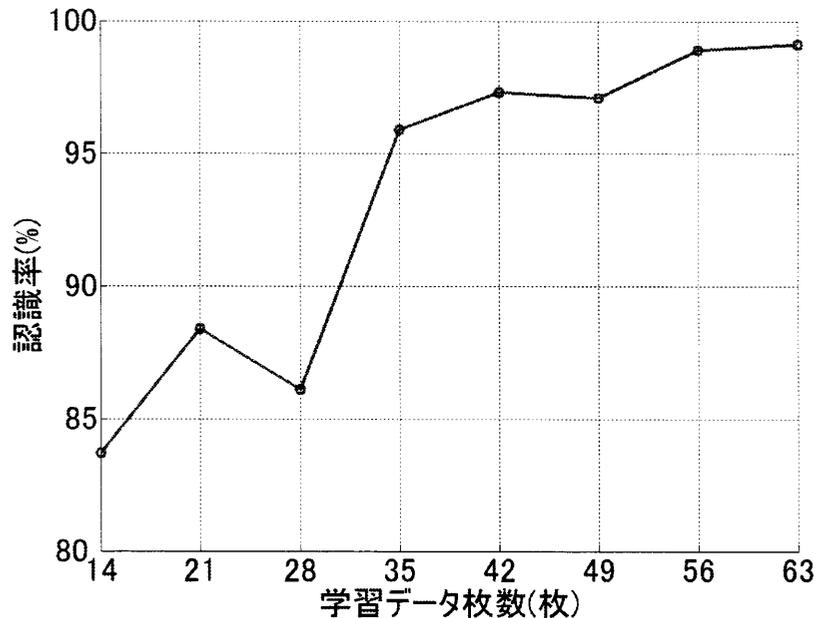


図 3-14 学習データ枚数について

学習データ数が 35 枚になるあたりまで認識率が安定していない。この理由として、学習データ枚数が少なすぎるため判別空間が安定していないためである。

図 3-14より 56 枚以上学習させている場合に、認識率が安定していることが解る。これより、学習データから有効かつ安定した判別空間を構築するためには、1 枚の画像から抽出される特徴次元数の数倍以上の枚数が学習データとして必要であることが確認できる。

3-5-2 仮想学習データ作成

しかし、多種の学習データを収集するのは現実問題として、利用者に負荷がかかり、集められるデータ数が限られているなど、十分な枚数が得られないような場合が多々ある。そこで、本研究では実際に存在する学習データから仮想学習データを作り出し、データを擬似的に増やすことを提案する。今回は、原画像に幾学的な傾きを与えることによって、仮想学習データを作成した。この仮想学習データを実在する学習データに加え、新たに辞書を作成しなおし、実際に判別空間が安定するにはどの程度学習枚数が必要になるかを検討した。

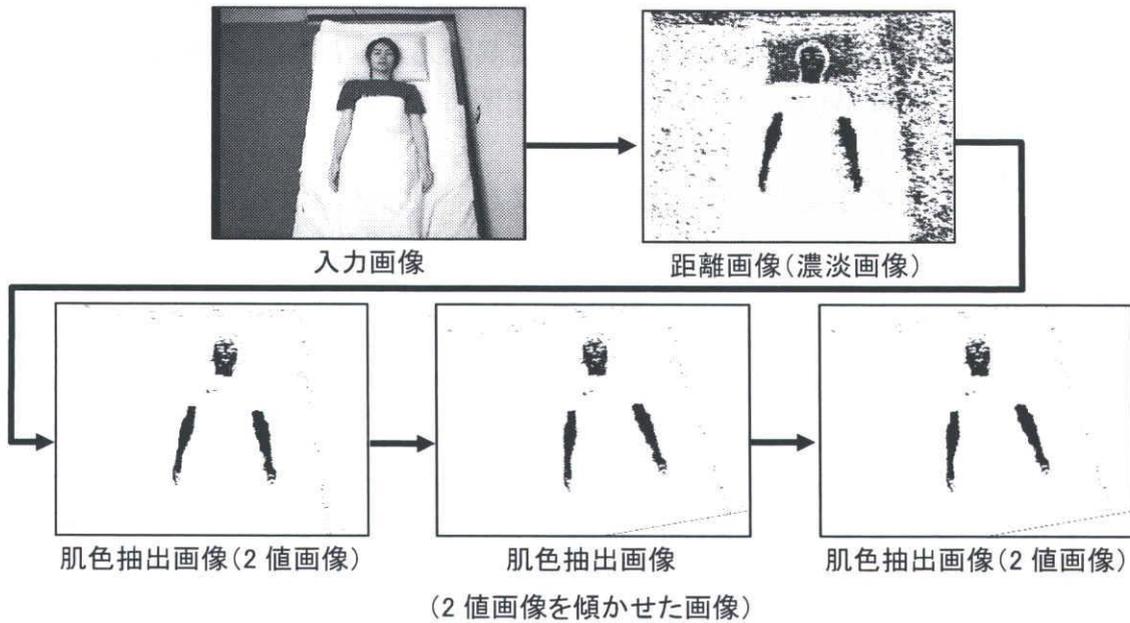


図 3-15 肌色抽出画像(2 値画像)を傾かせた場合

仮想学習データは次のようにして作成した。入力画像から各画素で肌色基準 U , V 値からの距離を求め、距離画像(濃淡画像)を作成する。次に判別分析法により肌色抽出画像(2 値画像)を得て、その画像を傾かせる。この際、中心軸は画像の中心とした。対象となる画像が 2 値画像なので、補間されて 0, 255 以外の画素値が補完されたとしても 0, 255 の値の近い方に振り分けられなくてはならない。よって中間値の 128 で閾値を決めることができる。よって傾かせた画像を 128 の閾値により再び 2 値化させれば、量子化誤差を最小に抑えられる。実際の処理の流れを図 3-15 に示す。また、傾かせない場合に得られる肌色抽出画像を図 3-16(a) に、本節で求められた肌色抽出画像を図 3-16 (b) に示す。図 3-16 (a)(b) を比較するとほとんど量子化誤差が見られないことが解る。

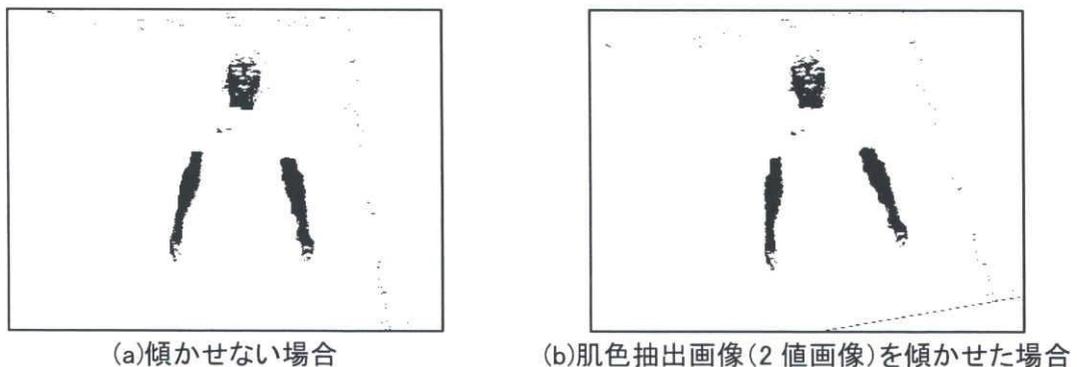


図 3-16 肌色候補領域についての比較

この画像に対し3-3-3節の処理を行い、高次局所自己相関特徴を抽出する。肌色領域と中心軸との相対位置は画像によって異なるが、顔周辺の領域を切り出して、切り出した領域内から高次局所自己相関特徴を抽出しているため、さほど影響は見られない。

今回は、実在する学習データを各ポーズ2枚、合計14枚用意した。最初に、学習データを時計回り、半時計回りに 1° 傾かせることにより、28枚の仮想学習データが作成される。学習データと仮想学習データを合わせた合計42枚から辞書を作成する。次に、学習データを時計回り、半時計回りに 2° 傾かせることにより、先ほどと同様に28枚の仮想学習データが作成される。この仮想学習データを先ほど得られた42枚のデータに加えた合計70枚から辞書を作成する。このように角度を 0° から 1° ごとに 5° まで増やし、学習データ枚数を増やした。未知データとしては、3-5-1節と同様のものを用いた。

実験の結果を図3-17に示す。グラフの横軸は辞書を作成するときに用いた学習データの枚数を示し、縦軸は認識率(正解枚数/総枚数)を示す。

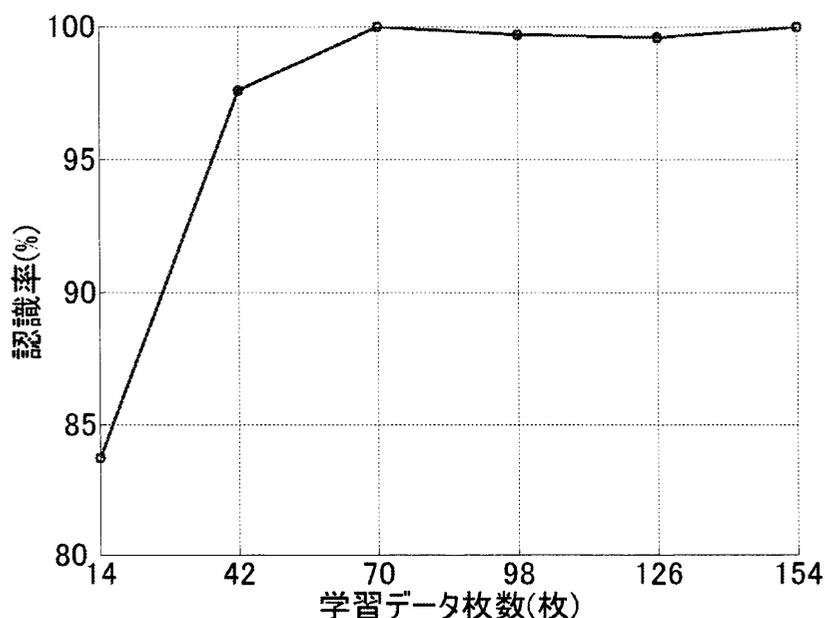


図 3-17 仮想学習を用いての識別結果

仮想学習データを加えることによって、認識率は急激に上昇した。これより仮想学習データを作成することの有効性が示された。また、図3-17において学習枚数が増えるにつれて、認識率がほぼ一定の値となり、飽和していることが確認できる。 $\pm 2^\circ$ 傾かせたデータを加えた70枚の学習データを用いたときの認識率でほぼ上限に達しているため、この程度の学習枚数で判別空間が安定することが確認できる。

3-6 高次局所自己相関特徴の傾きに対する対応

高次局所自己相関特徴は平行移動に強いが、傾きに対しては弱いことが知られている。この場合、認識対象となる物が傾いた画像を学習させれば、傾いた場合にも対応できることが考えられる。しかし、実際に傾いた画像を何枚も獲得することができない場合がある。そこで3-5-2節で用いた方法により、画像に幾学的な変換を施し傾いた画像を作成し、仮想的に傾かせた画像を学習させることにより、傾きにも対応できるような特徴空間の作成を目的とする。

3-6-1 平行移動に対する高次局所自己相関特徴の検証

高次局所自己相関特徴が平行移動に強いかを検証するために以下のようなデータセットを用意し識別実験を行った。

3-5-1節で述べた状況で実験環境を整えた。学習データとして図 3-2に示されている各ポーズを 100 枚ずつ、合計 700 枚の画像を用意した。未知データは学習データを撮影した時間と異なる時間に撮影を行い、服装も異なるものを着用し、学習データと同様に 700 枚用意した。学習データを撮影した位置とほぼ同じ位置に横たわり、ベッドに対して利用者がまっすぐ横たわっているものを固定データセット、利用者が学習データを撮影した位置から左側に平行移動しているものを横移動データセット、利用者が学習データを撮影した位置から下側に平行移動しているものを縦移動データセットとした。図 3-18 (a)に学習データの例を、図 3-18 (b)に固定データセットの例を、図 3-18 (c)に横移動データセットの例を、図 3-18 (d)に縦移動データセットの例を示す。

表 3-31 各データセットに対する認識率

固定データセット	横移動データセット	縦移動データセット
100(700/700)	99.9(699/700)	100(700/700)

表 3-31にそれぞれのデータセットに対する平均認識率を示す。1行目に書かれているものは、未知データの種類を示す。表内の括弧の左に記されている数値は認識率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。表 5-1 より高次局所自己相関特徴は平行移動に強いことが確認された。

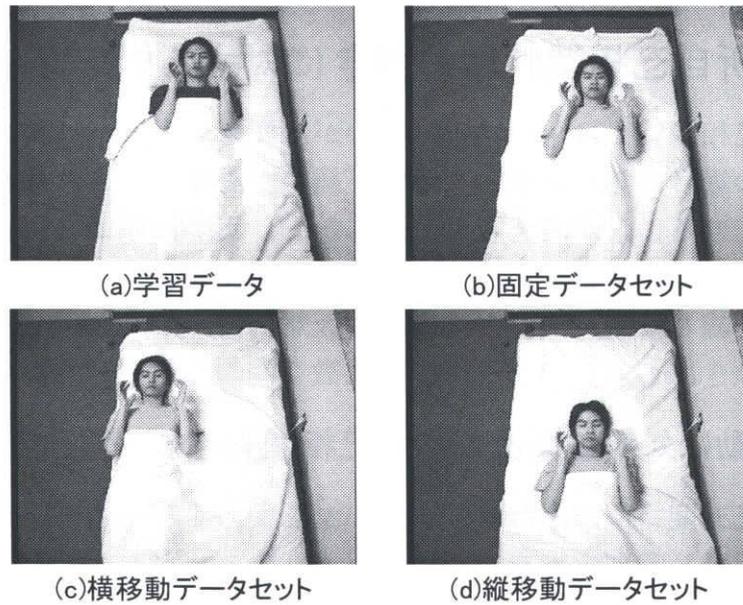


図 3-18 実験データ

3-6-2 傾きに対する高次局所自己相関特徴の検証

次に高次局所自己相関特徴が傾きに対して、実際に弱いかどうかを検証してみるため識別実験を行った。

3-5-1節で述べた状況で実験環境を整えた。学習データは3-6-1節で使用したものをを用いる。未知データは学習データを撮影した時間と異なる時間に撮影を行い、服装も異なるものを着用し、学習データと同様に700枚用意した。利用者が半時計回りに傾いているものを半時計回りデータセット、利用者が時計回りに傾いているものを時計回りデータセットとした。図 3-19(a)に半時計回りデータセットの例を、図 3-19(b)に時計回りデータセットの例を示す。

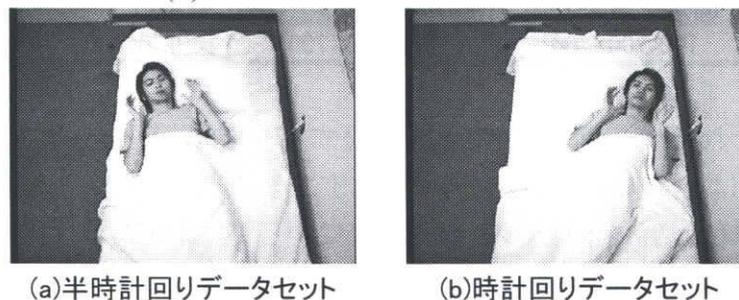


図 3-19 実験データ

また、表 3-32に各ポーズに対する認識率を示す。1行目に書かれているものは、未知データの種類を示し、一番左の列にかかっているものはクラスの種類を示す。表内の括弧の左に記されている数値は認識率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。

表 3-32 各ポーズに対する認識率

	半時計回りデータセット	時計回りデータセット
ノーマルポーズ	100(100/100)	94(94/100)
スイッチオン	0.0(0/100)	0.0(0/100)
スイッチオフ	100(100/100)	100(100/100)
チャンネルアップ	100(100/100)	100(100/100)
チャンネルダウン	97.0(97/100)	96.0(96/100)
ボリュームアップ	13.0(13/100)	4.0(4/100)
ボリュームダウン	98.0(98/100)	21.0(21/100)
全体	72.6(508/700)	59.3(415/700)

表 3-32の全体の認識率より高次局所自己相関特徴は傾きに弱いことが確認できる。また、スイッチオン、ボリュームダウンの認識率が著しく低いことが解る。

スイッチオンの認識率が低い原因として、半時計回りデータセットの場合、利用者が半時計回りに傾いた際に腕のエッジの傾きが、図 3-11(d), (f)に示すように、傾いていない場合のチャンネルダウン、ボリュームダウンの腕のエッジの傾きに類似しているため、チャンネルダウン、ボリュームダウンと誤認識している。一方、時計回りデータセットの場合、利用者が時計回りに傾いた際に腕のエッジの傾きが図 3-11(c)に示すように、チャンネルアップに類似しているためチャンネルアップと誤認識している。

ボリュームアップの認識率が低い原因として、半時計回りデータセットの場合、利用者が半時計回りに傾いた際に腕のエッジの傾きが、図 3-11(f)に示すように、傾いていない場合のボリュームダウンの腕のエッジの傾きに類似しているため、ボリュームダウンと誤認識している。一方、時計回りデータセットの場合、利用者が時計回りに傾いた場合に腕のエッジが、傾いていない場合と比較して、顔周辺矩形内に入らない。このため腕のエッジの特徴量が無くなってしまい、顔領域のエッジのみからなるノーマルポーズとの距離が近くなってしまい誤認識している。

3-6-3 傾きを考慮した辞書作成方法

半時計回り、時計回りデータセットでは利用者は約 15° 左右に傾いている。しかし、利用者は常に一定の角度に傾いているわけではなく、例えば右側に 10° 傾いている場合もあれば、左側に 5° 傾いている場合もある。よって学習を 0° ~ ±15° まで行えば、0° ~ ±15° まで利用者が傾いたとしても対応できると考えられる。このことより、利用者が傾いていない学習データを左右 0° ~ ±15° まで傾かせ、仮想学習データを作成する。これにより学習データ枚数は 21700 枚となる。この学習データを用いて辞書を新たに作成し、識別実験を行った。未知データは固定データセット、半時計回りデータセット、時計回りデータセットを使用する。

表 3-33に学習データ、それぞれのデータセットに対する平均認識率を示す。1 行目に書かれているものは、データの種類を示す。表内の括弧の左に記されている数値は認識率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。

表 3-33 各データセットに対する認識率

学習データ	固定データセット	半時計回りデータセット	時計回りデータセット
99.0(21491/21700)	100(700/700)	94.6(662/700)	87.5(613/700)

仮想学習を行ったことにより認識率が大幅に改善された。しかし、学習データであるにもかかわらず認識率が 100%にならず、多少低くなっている。これは、仮想学習データを加えた際にクラス間の分離が悪くなっている可能性が示唆される。そこで2-8節で述べた手法を用いて仮想学習を行っていない場合と仮想学習を行った場合の辞書を評価する。表 3-34に仮想学習を行っていない場合と仮想学習を行った場合のクラス内分散、クラス間分散、分散比を示す。

表 3-34 特徴空間の評価

	クラス内分散	クラス間分散	分散比
仮想学習なし	0.00060	0.0484	81.2
仮想学習あり	0.00061	0.0068	11.2

表 3-34より、仮想学習を行ったことにより、クラス内分散はさほど変化しなかったものの、クラス間分散が大幅に小さくなっていることが確認できる。これより、傾きには対応できるようになったが、クラス間分散が小さいため誤認識する可能性が高くなってしまったと考えられる。

実際に仮想学習枚数を増やしていった場合に、学習データ、未知データの固定データセットの認識率、クラス内分散、クラス間分散、分散比がどの程度変

化するかを調べた。半時計回り、時計回りデータセットを調べなかった理由としては、半時計回り、時計回りデータセットは左右に約 15° 傾いている。このため左右 15° の仮想学習データを作成し、学習した場合は認識ができるが、学習していない場合は認識できないと考えられる。今回は学習していない辞書が大半を占め、明らかに認識ができない場合が現れるためである。

学習データに左右 1° から 1° おきに 10° まで仮想的に傾かせたデータを加えた辞書を辞書 0-10deg とする。このように辞書 0-1deg から辞書 0-15deg まで作成する。仮想学習を行った辞書を用い識別実験を行った。表 3-35に実験結果を示す。一番左の列に示されているものは辞書の種類を示し、学習データ、固定データセットの列で括弧の上に記されている数値は認識率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。

表 3-35 認識率と特徴空間の評価

	学習データ	固定データセット	クラス内分散	クラス間分散	分散比
0-1deg	100 (2100/2100)	100 (700/700)	0.00060	0.0407	68.2
0-2deg	100 (3500/3500)	100 (700/700)	0.00060	0.0362	60.6
0-3deg	100 (4900/4900)	100 (700/700)	0.00060	0.0302	50.4
0-4deg	100 (6300/6300)	100 (700/700)	0.00060	0.0245	40.9
0-5deg	100 (7700/7700)	100 (700/700)	0.00060	0.0217	36.3
0-6deg	99.97 (9096/9100)	100 (700/700)	0.00060	0.0155	25.9
0-7deg	99.99 (10498/10500)	100 (700/700)	0.00060	0.0164	27.3
0-8deg	99.98 (11897/11900)	100 (700/700)	0.00060	0.0145	24.0
0-9deg	99.96 (13295/13300)	100 (700/700)	0.00061	0.0123	20.3
0-10deg	99.95 (14693/14700)	100 (700/700)	0.00061	0.0108	17.7
0-11deg	99.94 (16090/16100)	100 (700/700)	0.00061	0.0097	15.8
0-12deg	99.80 (17465/17500)	100 (700/700)	0.00061	0.0088	14.4
0-13deg	99.59 (18822/18900)	100 (700/700)	0.00061	0.0081	13.2
0-14deg	99.30 (20157/20300)	100 (700/700)	0.00061	0.0074	12.1
0-15deg	99.04 (21491/21700)	100 (700/700)	0.00061	0.0068	11.2

表 3-35より, 仮想学習枚数を増やすことにクラス間分散が小さくなる傾向がみられ, 学習データの認識率も少しではあるが, 悪くなっていることが解る.

実際に全ての角度を学習させてしまうと学習時間もかかり, またクラス間の分離も悪くなることが考えられる. そこで全ての角度を学習させるのではなく, 2度おき, 3度おきというように間隔をあけて仮想学習データを作成した. 辞書は以下のものを用意した.

表 3-36 辞書の種類

2interval	0°, ±2°, ±4°, ±6°, ±8°, ±10°, ±12°, ±14° と 0° から 2° 間隔で仮想学習データを用意
3interval	0°, ±3°, ±6°, ±9°, ±12°, ±15° と 0° から 3° 間隔で仮想学習データを用意
4interval	0°, ±4°, ±8°, ±12° と 0° から 4° 間隔で仮想学習データを用意
5interval	0°, ±5°, ±10°, ±15° と 0° から 5° 間隔で仮想学習データを用意

実際に仮想学習枚数を増やしていった場合に, 学習データ, 未知データ(固定データセット, 半時計回りデータセット, 時計回りデータセット)の平均認識率, クラス内分散, クラス間分散, 分散比がどの程度変化するかを調べた. 表 3-37に学習データの認識率と未知データの平均認識率を, 表 3-38に各データセットの平均認識率を, 表 3-39にクラス内分散, クラス間分散, 分散比を示す. 表 3-37, 表 3-38, 表 3-39の一番左の列に示されているものは辞書の種類, 表 3-37, 表 3-38内の括弧の左に記されている数値は認識率(単位は%)を, 括弧内に記されている数値は(正解枚数/総枚数)を示す.

表 3-37 学習データ, 未知データに対する認識率

	学習データ	未知データ
2interval	99.20(10415/10500)	93.8(1969/2100)
3interval	98.94(7618/7700)	95.4(2004/2100)
4interval	99.65(4882/4900)	91.9(1930/2100)
5interval	98.67(4834/4900)	97.0(2037/2100)

表 3-38 各データセットに対する認識率

	固定データセット	半時計回りデータセット	時計回りデータセット
2interval	100(700/700)	94.1(659/700)	87.1(610/700)
3interval	100(700/700)	96.1(673/700)	90.1(631/700)
4interval	100(700/700)	90.7(635/700)	85.0(595/700)
5interval	100(700/700)	98.0(686/700)	93.0(651/700)

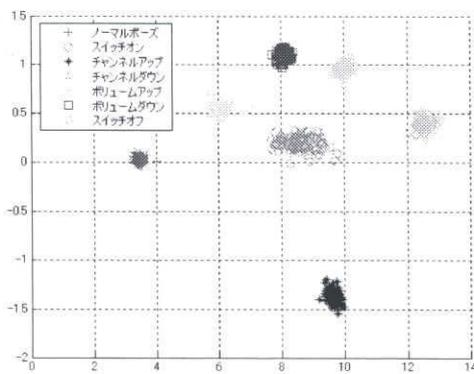
表 3-39 特徴空間の評価

	クラス内分散	クラス間分散	分散比
2interval	0.00061	0.0072	11.7
3interval	0.00061	0.0065	10.6
4interval	0.00061	0.0077	12.6
5interval	0.00061	0.0063	10.3

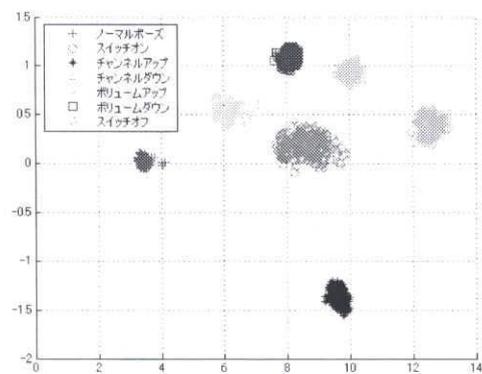
認識率は 5interval が最も良かったが、表 3-39よりさほどクラス内分散、クラス間分散の差はみられない。むしろ分散比は辞書 0-15deg とさほど変わらない。これより、大きく傾かせた仮想学習データを学習させるほどクラス間分散が小さくなるのが解る。また、分散比が変わっていないにもかかわらず、間隔を空けるほど認識率が上昇している理由は、仮想学習する間隔が小さいほど、クラス間の重なりが増えてきていることが考えられる。間隔をあけるごとにクラス間の重なりが小さくなり、判別が行いやすい空間になっていることが考えられる。しかし、間隔を開けすぎると、学習されていない角度が多くなるため学習されていない角度の補間ができなくなる可能性がある。このため今回は 5° おきに学習させることが有効であるとした。

3-6-4 可視化による特徴空間の観察

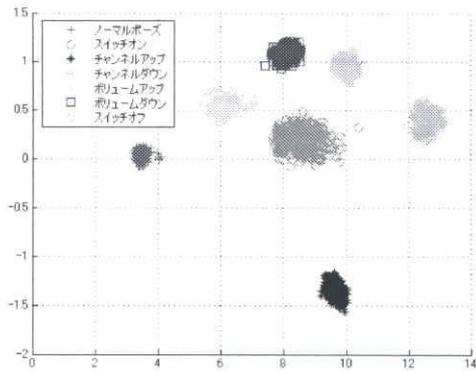
前節ではクラス内分散・クラス間分散比によって特徴空間を評価した。しかし、特徴量の取る値により表価値が一定にならないため、計算された値が正しいかどうかの判断が難しい。また分布の重なり具合を評価していないため、分布の分離度も反映されていない。そこで実際に主成分分析、クラスタ判別法により元の特徴空間を 2次元に写像し、特徴空間がどのようになっているか可視化することにより観察する。実験データは表 3-35に記されている学習データを使用する。主成分分析により特徴空間を 2次元に圧縮したものを図 3-20に示す。



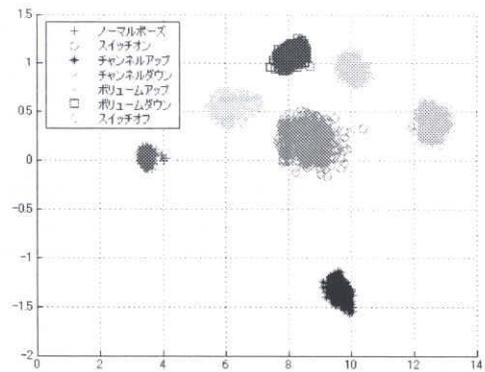
(a)0deg(仮想学習なし)



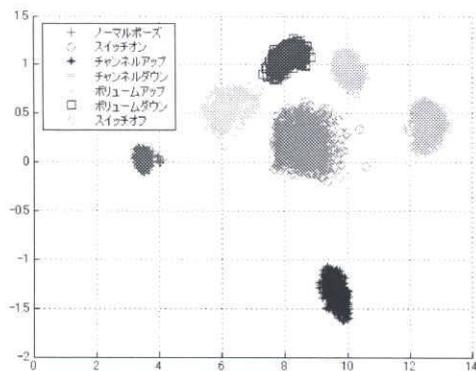
(b)0-1deg(仮想学習あり)



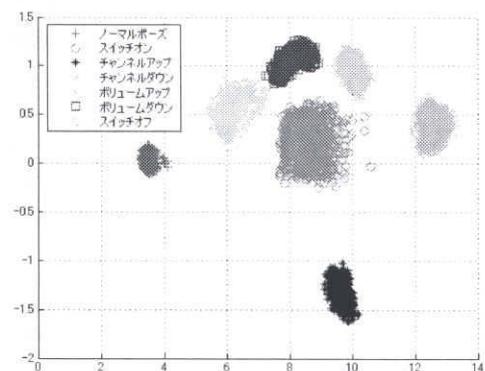
(c)0-2deg(仮想学習あり)



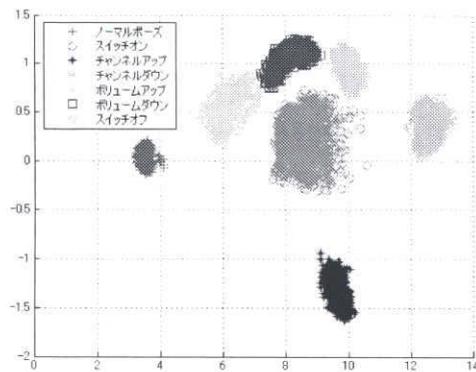
(d)0-3deg(仮想学習あり)



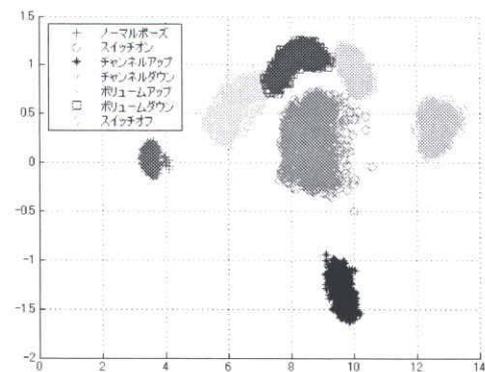
(e)0-4deg(仮想学習あり)



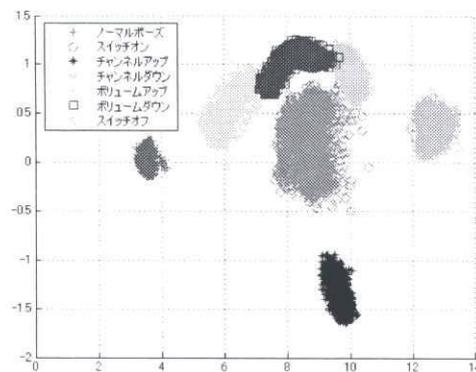
(f)0-5deg(仮想学習あり)



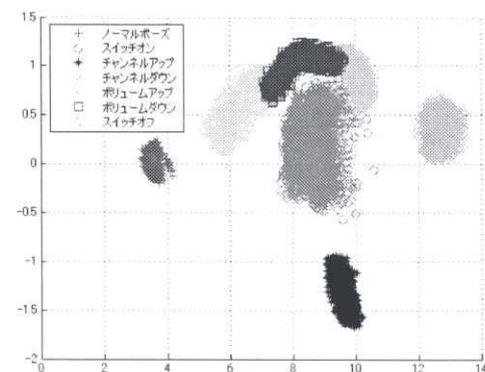
(g)0-6deg(仮想学習あり)



(h)0-7deg(仮想学習あり)



(i)0-8deg(仮想学習あり)



(j)0-9deg(仮想学習あり)

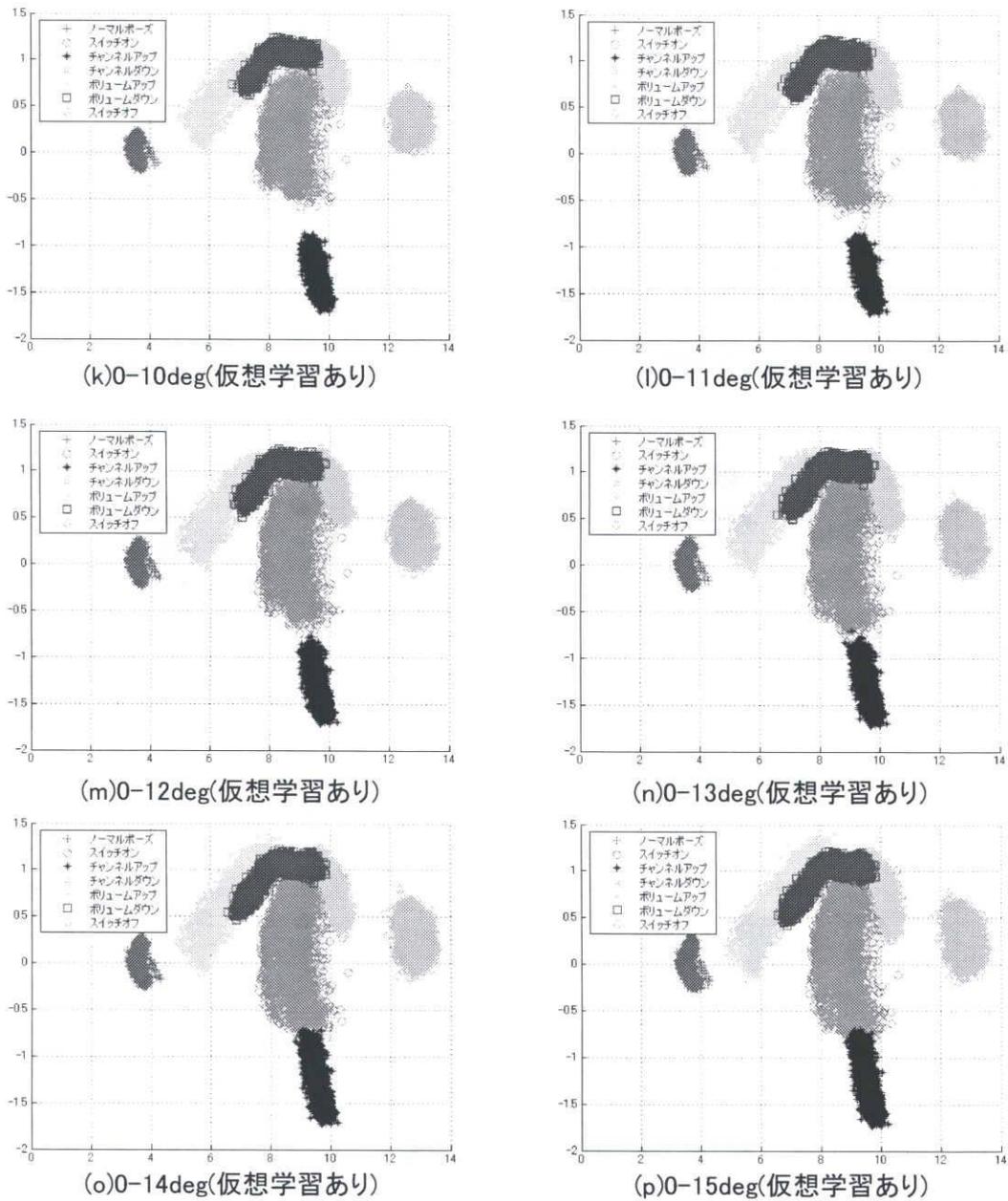
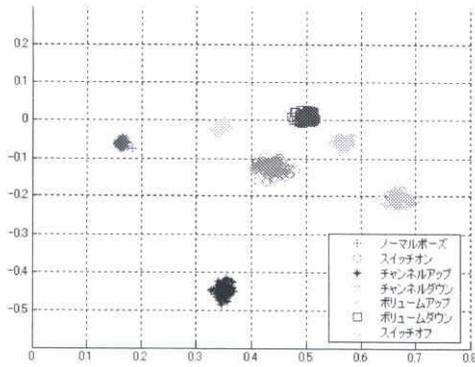
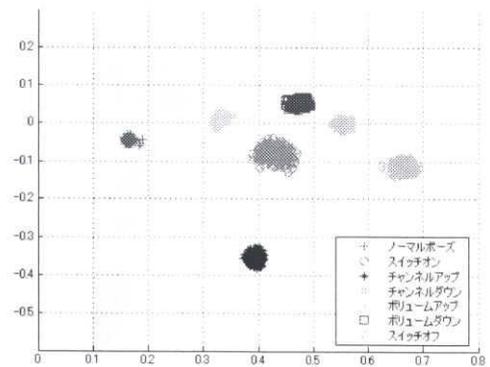


図 3-20 主成分分析により可視化された特徴空間

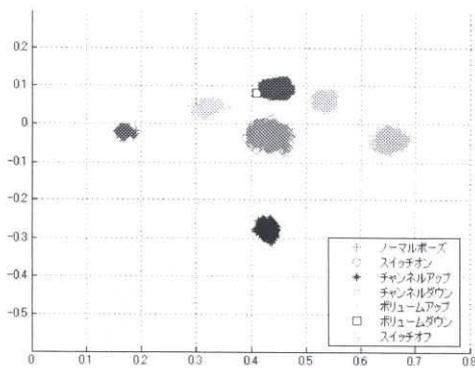
汎用性を持たせたデータが増えるにつれ徐々に各クラスの分散が大きくなり、クラス間分離度が悪くなっている様子が解る。次にクラスタ判別法により特徴空間を2次元に圧縮したものを図 3-21に示す。



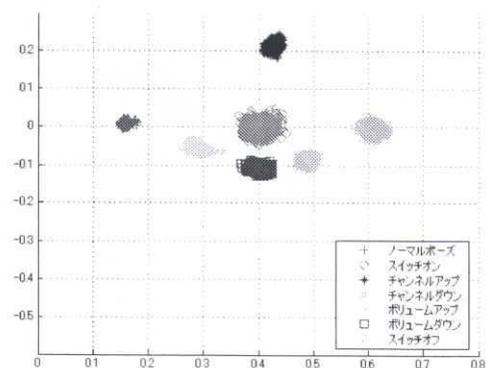
(a)0deg(仮想学習なし)



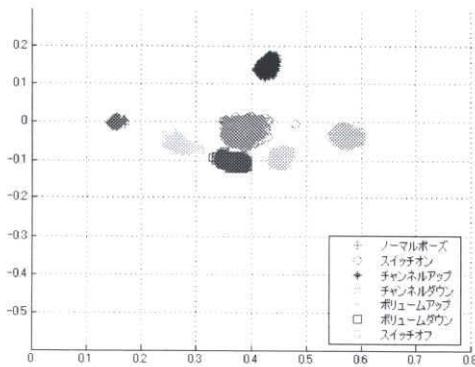
(b)0-1deg(仮想学習あり)



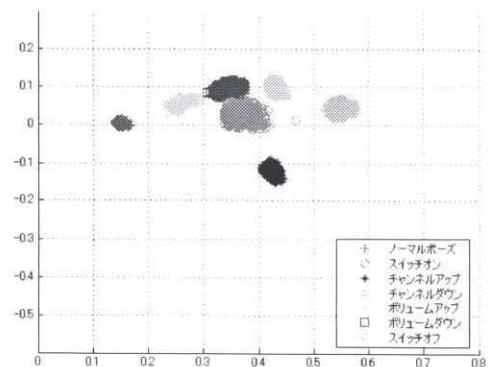
(c)0-2deg(仮想学習あり)



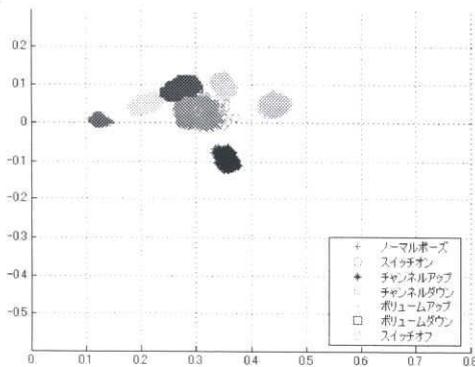
(d)0-3deg(仮想学習あり)



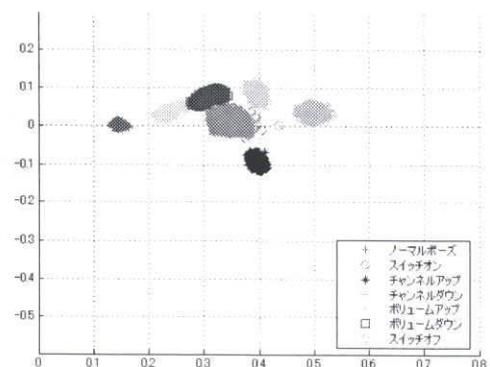
(e)0-4deg(仮想学習あり)



(f)0-5deg(仮想学習あり)



(g)0-6deg(仮想学習あり)



(h)0-7deg(仮想学習あり)

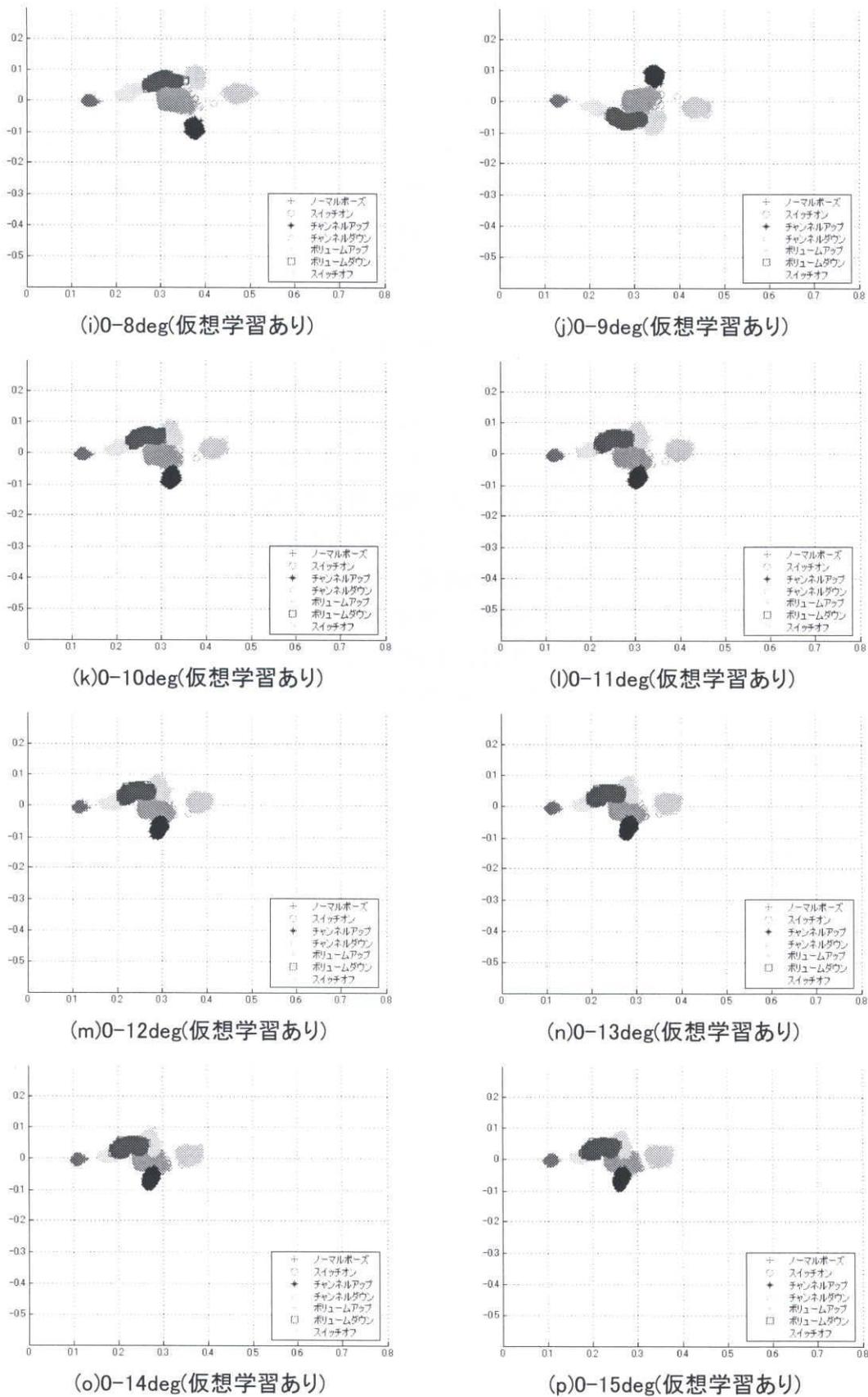


図 3-21 クラスタ判別法により可視化された特徴空間

クラス間分離度が悪くなっている様子に変わりはないが、主成分分析のときと異なり、各クラスの分散はさほど変化がないことが確認できる。

図 3-20(a), 図 3-21(a)より仮想学習を行わない場合は、判別しやすい特徴空間が作成されている。一方、図 3-20(b)~(p), 図 3-21(b)~(p)より多くの仮想学習を行うにつれ、クラス間の重なりができてしまい判別が難しい特徴空間が作成されていることが視覚的に解る。また、図 3-21と表 3-35を比較してみると、表 3-35のクラス内分散、クラス間分散の値が図 3-21に表されていることも確認できる。

図 3-20より主成分分析により次元圧縮を行った場合は、仮想学習がない場合、仮想学習がある場合の両方とも、クラスが分布している位置が同じである。

一方、図 3-21よりクラスタ判別法により次元圧縮を行った場合、仮想学習がない場合、仮想学習がある場合のほとんどのクラスの分布している位置が変わっている。これは、主成分分析が特徴空間全体の分布(共分散)のみしか考慮せずに軸を決定していることに対して、クラスタ判別法はクラスの分布(クラス内分散、クラス間分散)を考慮して軸を決定しているためである。

次に表 3-36に示されている学習データを用いた場合、特徴空間がどのように変化していくかを観察する。図 3-22に主成分分析により特徴空間を 2 次元に圧縮したものを、クラスタ判別法により特徴空間を 2 次元に圧縮したものを図 3-23に示す。

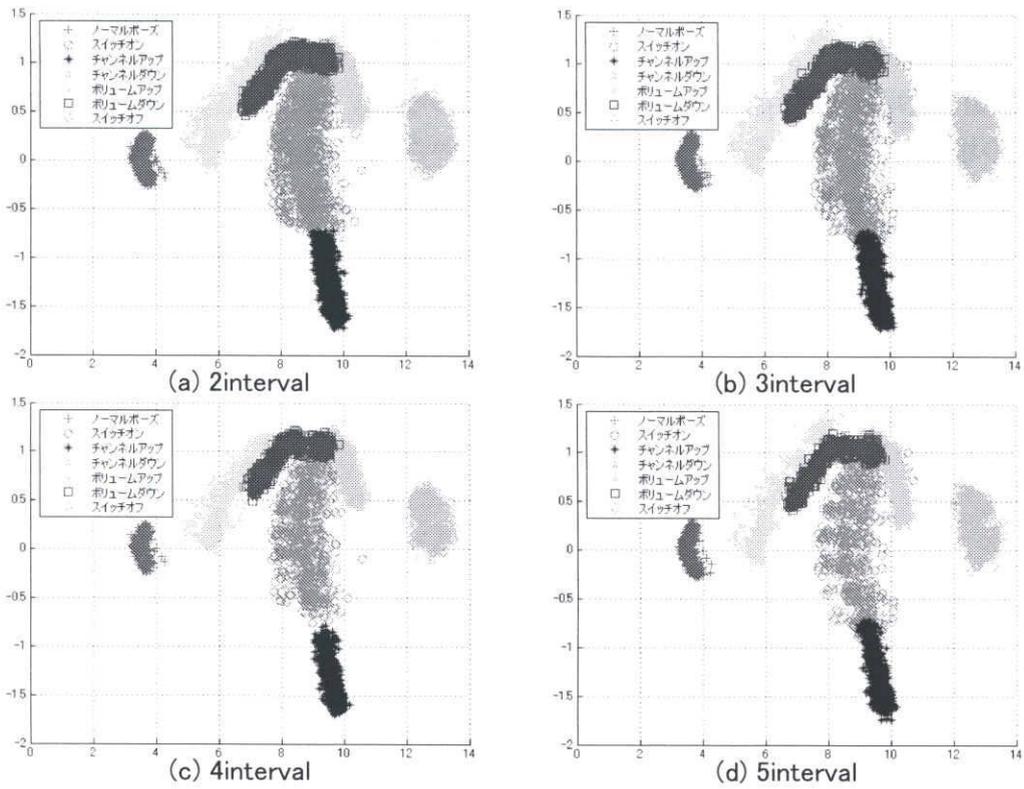


図 3-22 主成分分析により可視化された特徴空間

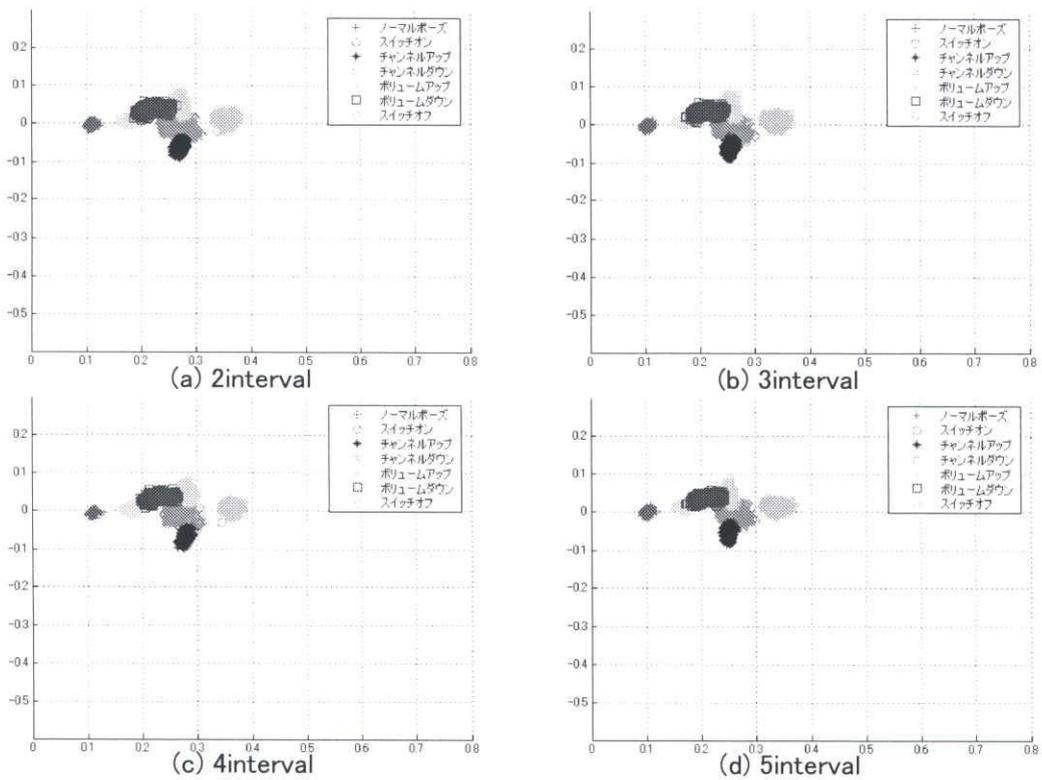


図 3-23 クラスタ判別法により可視化された特徴空間

また、図 3-22, より主成分分析により次元圧縮を行った場合は, 2interval, 5interval の両方とも全体的に分布の形は変わっていないが, 学習されていない角度の場所にデータがプロットされていないことが解る.

一方, 図 3-23よりクラスタ判別法により次元圧縮を行った場合, どちらの分布も差異はあまり見られない.

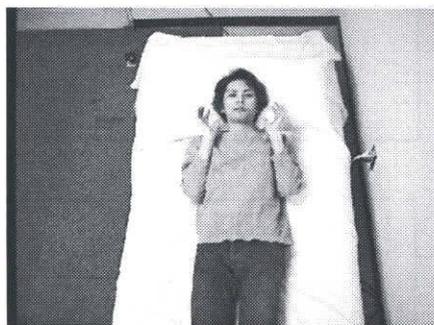
3-7 相対位置情報を持たせた特徴量抽出の提案

平行移動に強いことは高次局所自己相関特徴の利点ではあるが、その特徴量は局所的であり間に広がりをもった対象の区別に適していないため、空間上の相対位置情報が必要な場合が生じる。そこで、特徴抽出領域を分割することにより、大まかな相対位置の情報を取り込んだ特徴量を抽出し、相対位置情報を持たせた特徴量について考察を行う。

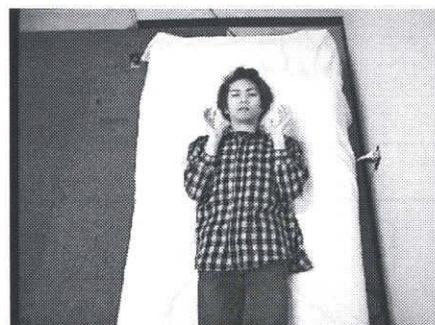
3-7-1 識別実験

今まで利用者が半袖を着用した事例について考察してきたが、利用者が長袖を着用する場合について考察する。長袖の場合肌色領域をあまり露出していない画像となり、これにより得られる特徴量が少なくなることが考えられる。そのためどの程度服装の影響が認識に現れるかを検証する。

3-5-1節で述べた状況で実験環境を整えた。学習データとして図 3-2に示されている各ポーズを 100 枚ずつ、合計 700 枚の画像を用意した。この学習データを、長袖学習データセットとする。未知データは学習データを撮影した時間と異なる時間に撮影を行い、学習データを撮影した位置とほぼ同じ位置に横たわり、服装も異なるものを着用し、学習データと同様に 700 枚用意した。この未知データを、長袖未知データセットとする。図 3-24 (a)に長袖学習データセットの例を、図 3-24 (b)に長袖未知データセットの例を示す。また各ポーズのエッジ画像を図 3-25に示す。



(a)長袖学習データセット



(b)長袖未知データセット

図 3-24 実験データ

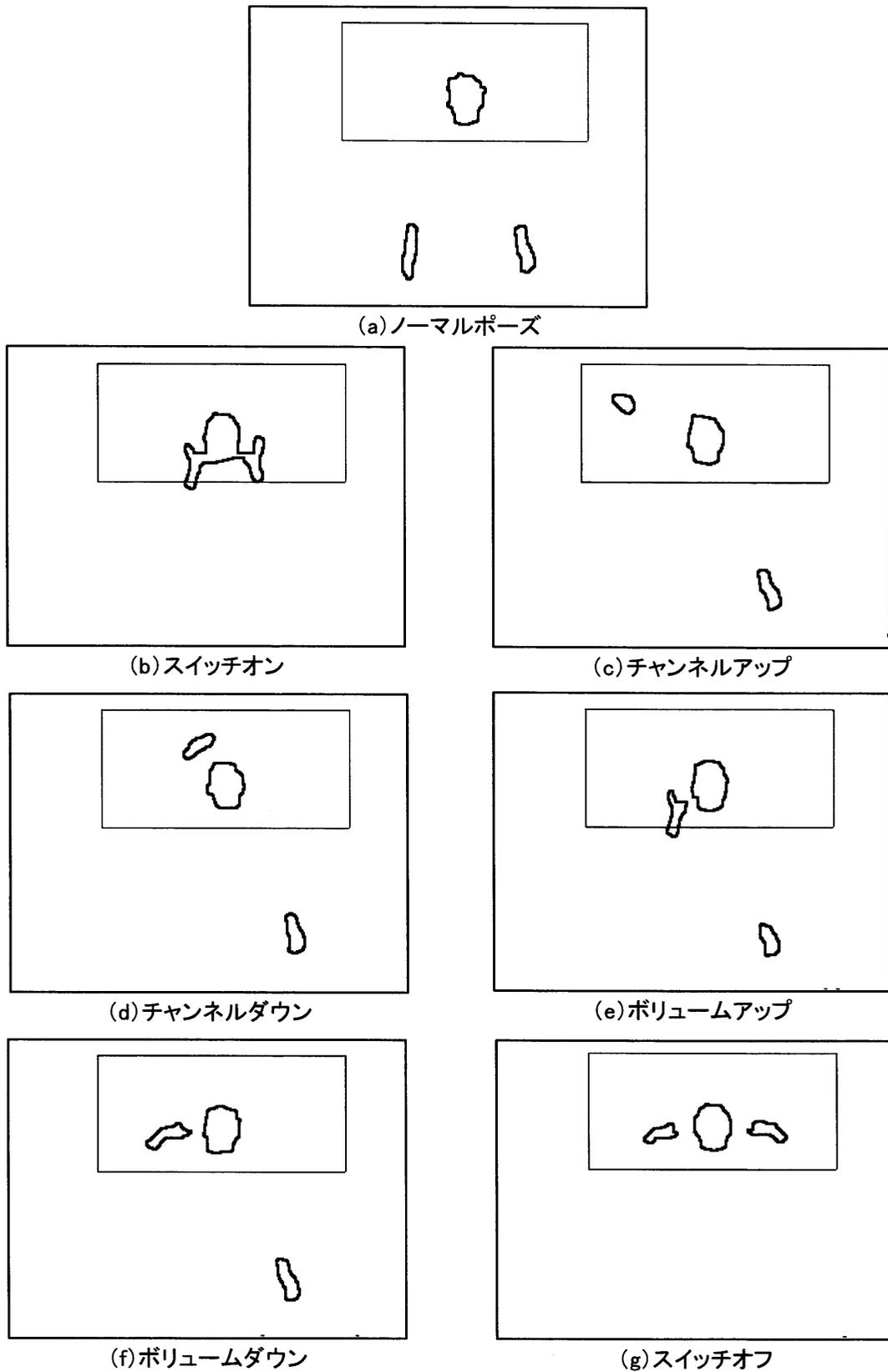


図 3-25 エッジ画像

長袖学習データセットの認識率は 100%(700/700)であった。長袖未知データセットの認識率は 96.1%(673/700)であった。認識率が低下した原因として、図 3-25に示すように、長袖を着用しているため腕の肌色領域のエッジ情報がなくなってしまう、手先と顔の肌色領域のエッジ情報しか得ることができない。このような場合、利用者がポーズを変えた場合でも、得られる手先のエッジの形状があまり変わらないため、同じ特徴として認識されると考えられる。

そこで、相対位置情報を考慮するために顔の重心点を中心とした 3×3 の 9 領域に分割する。分割された各々の領域から特徴量を抽出する。これにより、高次局所自己相関特徴に顔を中心とした空間上の相対位置情報を持たせる。

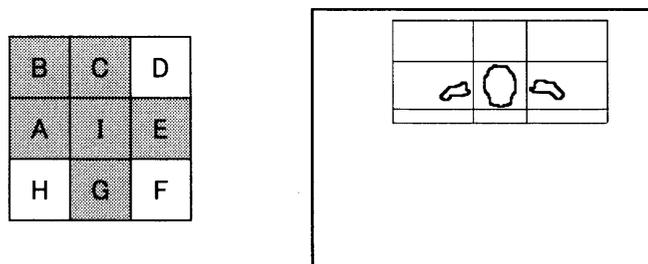


図 3-26 顔周辺矩形の分割

このとき、図 3-26の左側に示すようにそれぞれの領域に番号をつける。全ての領域から 25 次元の高次局所自己相関特徴を抽出すると $25 \times 9 = 225$ 次元となる。しかし、今回使用している命令実行ポーズに対して不必要な領域が存在する。このため、命令実行ポーズに対して不必要な部分を除いた A, B, C, E, G, I の 6 つの領域のみを特徴抽出の対象とした。またこれにより、むやみに次元数を増やしていないため、学習に必要なサンプル数をさほど増やす必要もなく、特徴空間も確率密度関数が比較的作成しやすい空間にすることができる。これら 6 つの領域から 25 次元の高次局所自己相関特徴を抽出する。これにより特徴次元数は、 $25 \times 6 = 150$ 次元になる。

この相対位置情報を持たせた 150 次元の高次局所自己相関特徴を用いて、先ほどと同様の 3-7-1 節のデータセットに対して識別実験を行ったところ、認識率は学習データ、未知データ共に 100%(700/700)になった。

3-7-2 エッジ情報がある場合の考察

エッジ情報が多い場合でも相対位置情報を持たせた特徴を識別できるかどうかを試すために、利用者が半袖を着用している画像を用意した。学習データと未知データを 3-7-1 節と同様の条件で用意し、識別実験を行った。この学習データを半袖学習データセット、未知データを半袖未知データセットとする。図 3-27 (a)に半袖学習データセットの例を、図 3-27 (b)に半袖未知データセットの例を示す。

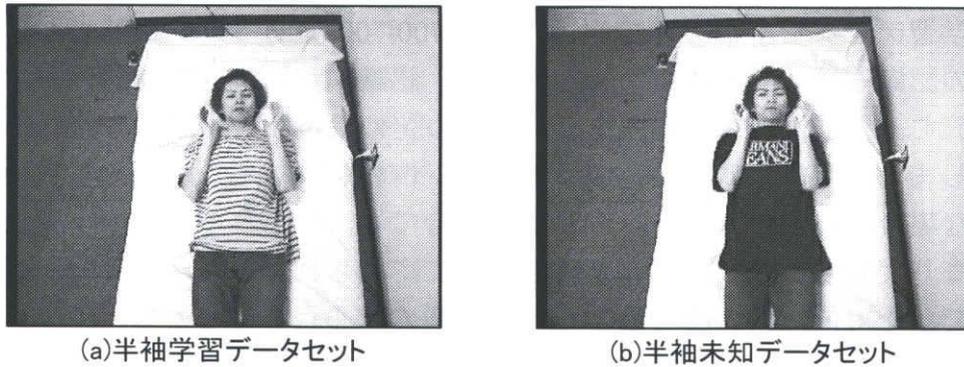


図 3-27 実験データ

こちらでも認識率は学習データ，未知データ共に 100%(700/700)となったため，相対位置情報を持たせた高次局所自己相関特徴はエッジ情報がある場合でも有効であることが確認できた。

3-8 特徴空間を考慮した辞書作成法の提案

大幅に異なる特徴量が存在しても，同じクラスとして取りまとめた場合があるとす。しかし，大幅に異なる特徴量を1つのクラスとすると，クラス内分散が大きくなりすぎてしまい，クラス間分散が小さくなってしまふことが考えられる。このような特徴空間が作成されてしまうと，判別が困難になってしまう。本研究で扱っているデータは，肌色領域から特徴を抽出しているため，第3-7節で述べたように服装の違いによって同じクラスでも，得られる特徴量が異なる場合が出てくる，服装ごとに辞書を作成すれば良いが，本研究で使っているシステムは，服装をどのように判断するかが問題となってくる。そこで複数の辞書作成法を提案し，特徴空間の分離度等を考察する。

3-8-1 識別実験

実験データは，3-6-1節，3-7-1節で用意したものを用いる。半袖学習データセット，長袖学習データセットから，相対位置情報を持つ特徴量(特徴次元数 150次元)を抽出し，表 3-40に示すような4つの辞書を作成した。

表 3-40 辞書の種類

辞書 S7	利用者が半袖を着用した物を用意し，各ポーズで1つクラスを作成し，7つのクラスに分けた辞書。
辞書 L7	利用者が長袖を着用した物を用意し，各ポーズで1つクラスを作成し，7つのクラスに分けた辞書。
辞書 SL7	利用者が半袖・長袖を着用した物を用意し，服装の形状が異なる場合でも同じポーズであれば，それを1つのクラスとし，7つのクラスに分けた辞書。
辞書 SL13	利用者が半袖・長袖を着用した物を用意し，同じポーズでも服装の形状が異なる場合は異なるクラスとし，13個のクラスに分けた辞書。

表 3-40に示される 4 種類の辞書を用いて識別実験を行った。表 3-41に半袖未知データセット，長袖未知データセットの認識率を示す。表 3-41の 1 行目に書かれているものは，未知データの種類，一番左の列に示されているものは辞書の種類を示し，表内の括弧の左に記されている数値は認識率(単位は%)を，括弧内に記されている数値は(正解枚数/総枚数)を示す。

表 3-41 半袖未知データセット，長袖未知データセットに対する認識率

	半袖未知データセット	長袖未知データセット
辞書 S7 (半袖のみ, クラス数 7)	100(700/700)	54.7(383/700)
辞書 L7 (長袖のみ, クラス数 7)	71.4(500/700)	100(700/700)
辞書 SL7 (半袖・長袖, クラス数 7)	100(700/700)	100(700/700)
辞書 SL13 (半袖・長袖, クラス数 13)	100(700/700)	100(700/700)

辞書 S7, 辞書 L7 は，学習された服装の形状に対しては高い認識が得られているが，学習されていない服装の形状に対しては認識率が大きく低下している。一方，両方の服装の形状を学習させてある辞書 SL7, 辞書 SL13 に対しては服装の形状によらず高い認識率が得られている。本節で議論しているクラスタリングの問題としては辞書 SL7, 辞書 SL13 にあたるため，これ以降は辞書 SL7, 辞書 SL13 について議論する。

3-8-2 クラス内分散，クラス間分散の考察

2-8節で述べた手法を用いて前節で述べた辞書 SL7, 辞書 SL13 を評価する。実際にクラス内分散，クラス間分散，分散比がどの程度変化するかを調べた。表 3-42にクラス内分散，クラス間分散，分散比を示す。表 3-42の一番左の列に示されているものは辞書の種類を示す。

表 3-42 特徴空間の評価

	クラス内分散	クラス間分散	分散比
辞書 SL7	0.013	5.4	405.9
辞書 SL13	0.027	12.4	457.2

分散比が大きくなる辞書が得られることが望ましい。また，分散比が大きいということは，クラス内分散が小さく，クラス間分散が大きい辞書のことである。このことを考えるとクラス内分散は辞書 SL7の方が小さく，クラス間分散は辞書 SL13の方が大きい。これは表 3-42に示されているクラス内分散の値は，各クラスのクラス内分散の合計値であり，辞書 SL7 よりクラス数の多い辞書 SL13 のクラス内分散値の値が大きくなるためである。しかし，最終的には分散比は辞書 SL13の方が大きいため辞書 SL13の方が最適な辞書といえる。

3-8-3 クラス間距離に対する考察

実際に分散比が小さいということは、クラス間の距離が取れていないことが予想される。そこで、クラス間の距離を見るために以下のようなデータセットを用意した。未知データとして、半袖・長袖未知データセットと同じ条件でノーマルポーズ→スイッチオン→チャンネルアップ→チャンネルダウン→ボリュームアップ→ボリュームダウン→スイッチオフ→ノーマルポーズという順番にポーズを連続的に変化させ合計 160 枚獲得した。半袖を着用しているものを半袖連続データ、長袖を着用しているものを長袖連続データとする。毎秒 5 フレーム獲得している。第3-7節の実験結果より、相対位置情報を持つ辞書を使用する。

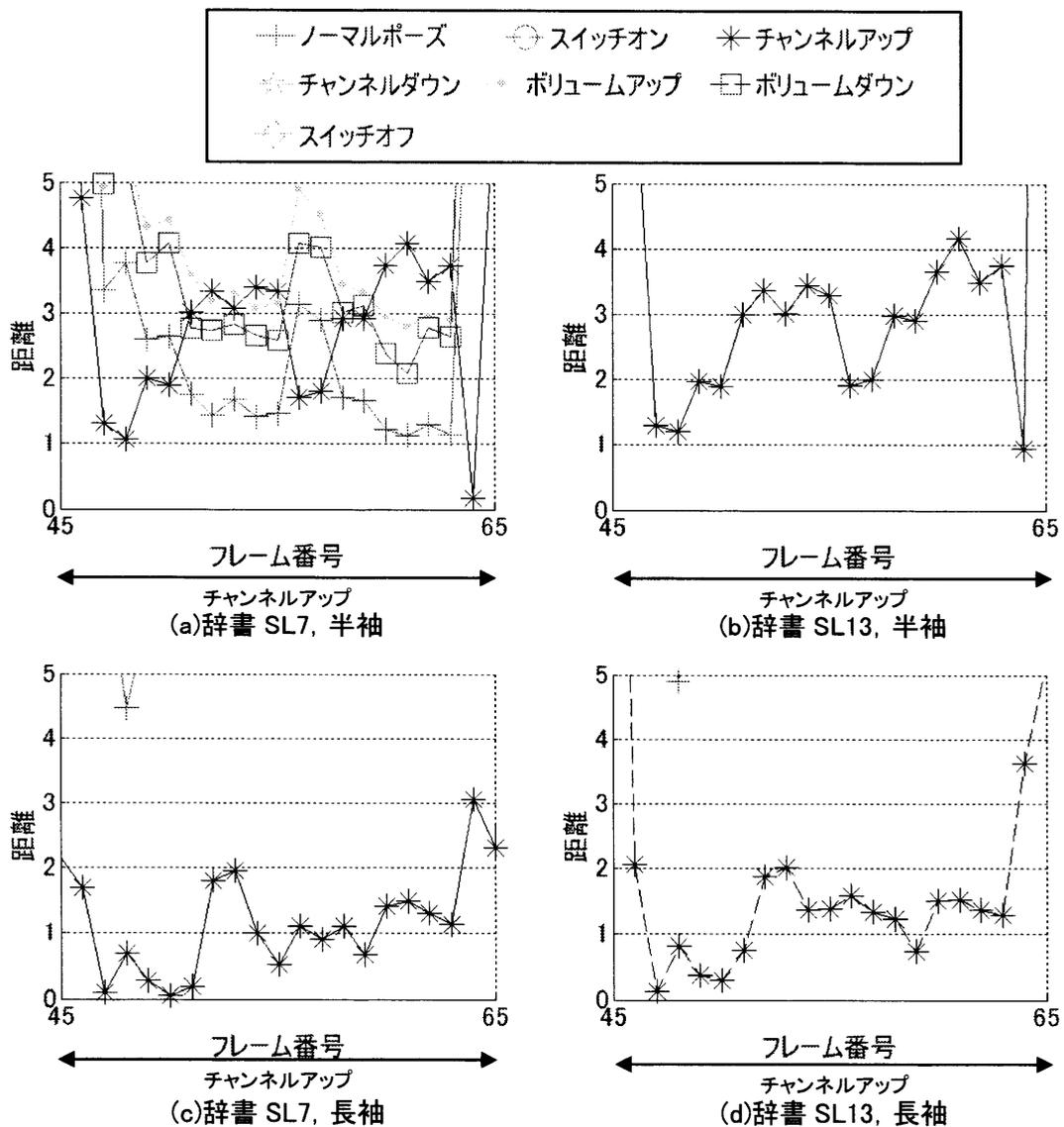


図 3-28 距離値

表 3-43 連続データに対する認識率

	半袖連続データ	長袖連続データ
辞書 SL7 (半袖・長袖, クラス数 7)	88.1(141/160)	93.8(150/160)
辞書 SL13 (半袖・長袖, クラス数 13)	95.0(152/160)	91.9(147/160)

表 3-43に辞書 SL7, 辞書 SL13 を用いた半袖・長袖連続データに対する認識率を示す. 表 3-43の 1 行目に書かれているものは, 未知データの種類, 一番左の列に示されているものは辞書の種類を示し, 表内の括弧の左に記されている数値は認識率(単位は%)を, 括弧内に記されている数値は(正解枚数/総枚数)を示す.

相対位置情報を持った辞書を使ったにもかかわらず全体的に認識率が低下してしまった. これは, ポーズ間動作がシステムに入力されてしまうためである. また, 長袖を着用している場合は辞書 SL7 を用いた方が良く, 半袖を着用している場合は辞書 SL13 を用いた方が良い.

各辞書が, クラス間の距離が取れているかを調べるため, 入力画像と学習辞書に登録されているクラスとの距離値をグラフにしたものを示す.

辞書 SL7 を用いて半袖連続データを入力したものを図 3-28(a), 辞書 SL13 を用いて半袖連続データを入力したものを図 3-28(b), 辞書 SL7 を用いて長袖連続データを入力したものを図 3-28(c), 辞書 SL13 を用いて長袖連続データを入力したものを図 3-28(d)に示す. 図 3-28はチャンネルアップ動作の画像番号 45 から 65 までを示した. グラフの縦軸は距離値を表し, 横軸はフレーム番号を示す. グラフの下側に実際の入力ポーズを示す. また辞書 SL13 を用いているグラフ中の実線は半袖の辞書との距離を示し, 点線は長袖の辞書との距離を示す.

図 3-28(b), (d)より辞書 SL13 は長袖, 半袖連続データについて, クラス間の距離が取れていることが確認できる. これに対し, 辞書 SL7 は表 3-41より, 辞書 S7, 辞書 L7 よりは服装の形状に対応していることが確認できるが, 図 3-28(a)よりクラス間の距離が取れていないことが確認できる. これにより辞書 SL13 が最適な辞書といえる.

3-9 異なった認識対象の拡大

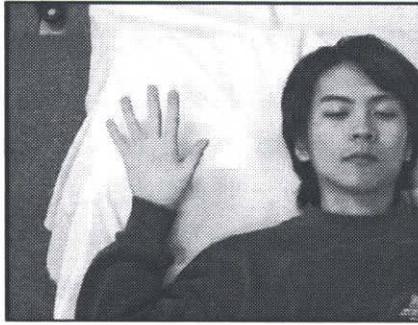
本節では、認識対象が異なる場合でも同じアルゴリズムで識別できるかを考察してみる。本研究で使用してきたシステムではこれまでに利用者の全身が写っている画像を対象としてきた。本節では体の一部のみが写っている場合でも今まで議論してきた手法で対処できるかどうかを検討する。

その一例として手のポーズを取り上げる。まず解像度の問題を解決するために、利用者の全身が写っている画像から、手領域にズームを行い、その後に高次局所自己相関特徴を抽出し識別する方法を提案した。

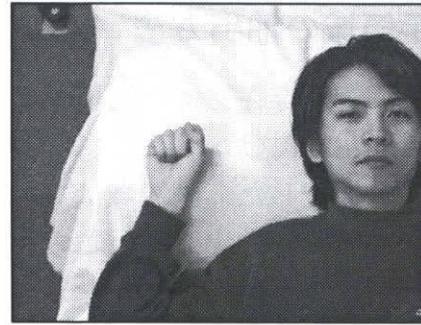
3-9-1 命令実行ポーズ

各種の命令を実行するために割り当てられた手のポーズを図 3-29に示す。利用者はベッドの上に横になった状態で家電製品を操作することができる。以下に図 3-29に示すポーズを提案し、それらに括弧内の命令を割り当てる。

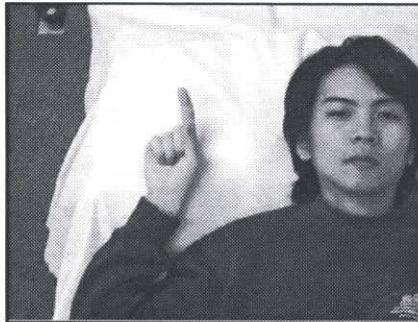
- (a)5 本全ての指を立てる。(ノーマルポーズ)
- (b)5 本全ての指を折る。(リセットポーズ)
- (c)人差し指のみを立てて、それ以外の指を折る。(スイッチオン)
- (d)人差し指と中指を立てて、それ以外の指を折る。(チャンネルアップ)
- (e)人差し指と中指と薬指を立て、それ以外の指を折る。(チャンネルダウン)
- (f)親指のみを折り、それ以外の指を立てる。(ボリュームアップ)
- (g)親指を立てて、それ以外の指を折る。(ボリュームダウン)
- (h)親指と人差し指を立てて、それ以外の指を折る。(スイッチオフ)



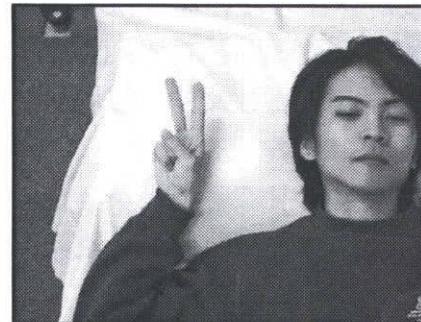
(a) ノーマルポーズ



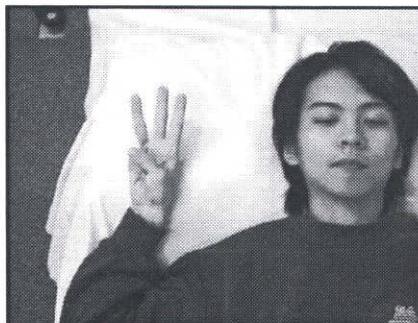
(b) リセットポーズ



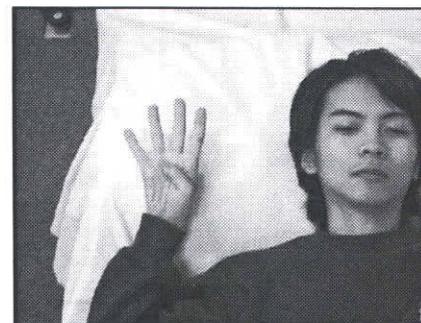
(c) スイッチオン



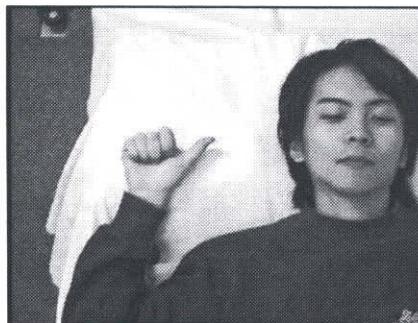
(d) チャンネルアップ



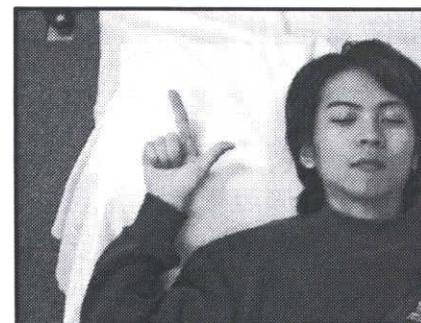
(e) チャンネルダウン



(f) ボリュームアップ



(g) ボリュームダウン



(h) スイッチオフ

図 3-29 命令実行ポーズ

3-9-2 手領域の抽出方法

最初に、手が存在している場所を示すために、利用者がベッド上で右手のみを一定時間左右に動かすようにした。動領域、肌色領域がともに現れた部分を手の位置と決定した。動領域はフレーム間差分により、肌色領域は3-3-1-2節に示す方法を用いて抽出した。この積で決められた領域は、どちらか一方が現れなかったとき、当然検出されない。しかしこの情報だけでは、顔領域、左手領域を誤検出してしまう可能性がある。そこで探索領域内に論理積により求められた領域が存在しない場合は手領域が存在しないとした。

探索領域の決め方として、図 3-30に示される画像で、顔領域は最も大きい肌色領域である。最も大きい領域の左端の x 座標、下端の y 座標を求め、この点を基に図 3-30の左上に示すような探索領域を決定する。

もし探索領域中に手領域が存在しなかったら、処理を行わない。もし一定時間手領域を見つけたなら、一定時間後に領域内の肌色領域の重心点を求め、この重心点を基にカメラ方向を決め、ズームを行い手の拡大画像を得る。カメラからベッドまでの距離は一定のためズームは固定で行った。

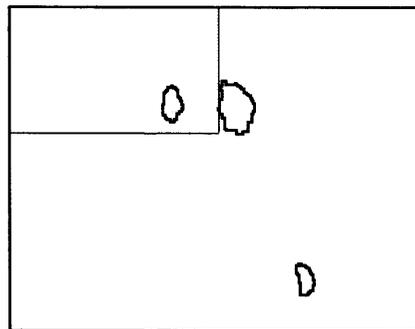
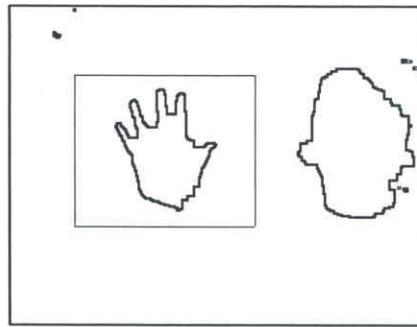


図 3-30 探索領域

3-9-3 特徴量抽出

拡大して得られた画像から3-3-1-2節に示す方法で肌色抽出を行い、膨張・収縮処理によりノイズを消去し、次に手領域の重心位置を求める。この重心位置を元にして、手周辺矩形を作成しエッジ画像に手周辺矩形を当てはめる。図 3-29(a)のノーマルポーズのエッジ画像に手周辺矩形を当てはめたもの図 3-31に示す。次に、手周辺矩形内から高次局所自己相関特徴を抽出する。この際矩形内は分割しないため、抽出される特徴量は25次元となる。得られた特徴量に対して、線形判別分析を行い、入力された画像がどのポーズに当てはまるかを識別する。



(a)ノーマルポーズ

図 3-31 命令実行ポーズのエッジ画像

3-9-4 識別実験

適用性を調べるためにポーズの識別実験を行った。3-5-1節で述べた状況で実験環境を整えた。学習データとして、図 3-29に示される 8 つの命令ポーズを各画像につき 100 フレーム分、計 800 枚の画像を用いた。未知データとしては、学習データで撮影した位置と異なる位置で右手領域を検出し、学習データと同様のポーズをとった画像を合計 800 枚用意した。未知データの一例を図 3-32 に示す。図 3-32で行っているポーズはノーマルポーズである。



図 3-32 未知データ

長袖学習データセットの認識率は 100%(700/700)であった。長袖未知データセットの認識率は 99.1%(694/700)であった。結果より対象が変わった場合でも、今まで議論してきた手法で対処できることを確認し、また高次局所自己相関特徴が平行移動に対しても強いことを確認した。

3-10 システム操作方法

ディスプレイ上にメニューが表示されているかどうかで、家電製品の操作を行うことが可能か否かをユーザーに伝えるようにした。この際、高齢者は機器操作に不慣れな方々が多いと考えられるため、システムのバリアフリー化が重要なポイントとなる。そこでメニューの文字を大きくし、操作命令が書いてある横に絵を載せることにより、操作を容易にした。

操作方法は、最初に制御可能である家電製品の一覧を示すメニューを立ち上げ、ユーザーはこの中から制御したい家電製品を選択する。次に、選択された家電製品の操作一覧がディスプレイ上に示される。ユーザーはこの操作一覧を見ながら家電製品を制御する。図 3-33に制御可能である家電製品の一覧を示すメニューを、図 3-34に各家電製品の操作一覧を示す。

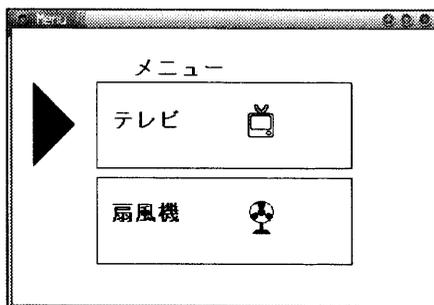


図 3-33 家電製品選択メニュー

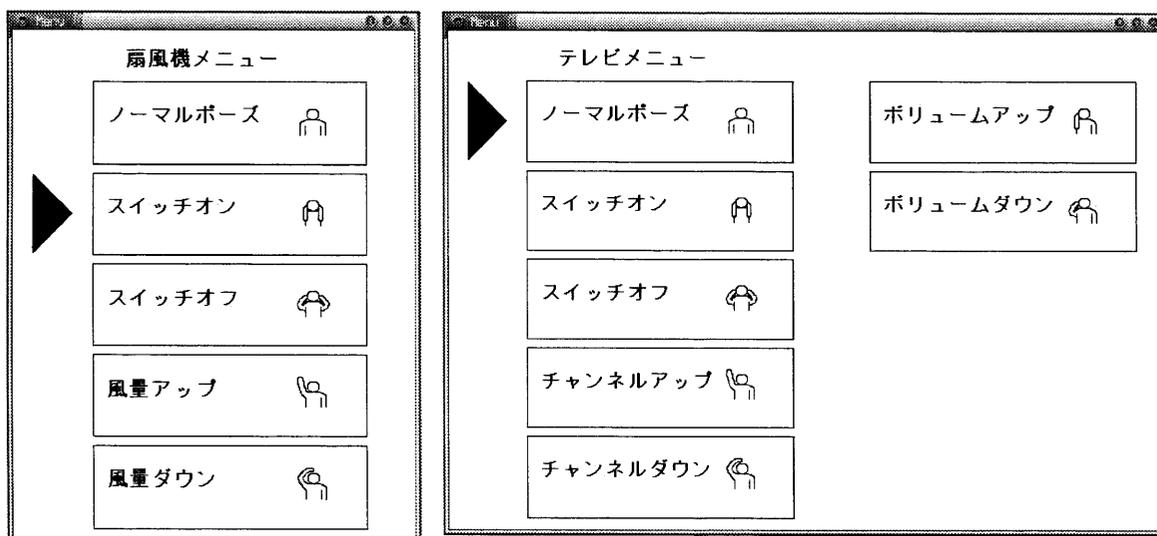


図 3-34 各家電製品の操作一覧メニュー

3-10-1 上半身のポーズを用い操作する場合

実際の操作では、まず図 3-2 (b)のように特殊なポーズをすることにより図 3-33に示すメニューがディスプレイ上に表示される。メニューが表示されている間だけ命令受け付け可能となる。次に図 3-2 (c)(d)のアップダウンのポーズを入力することで操作したい機器にカーソルを合わせ、(a)のポーズをすることによって機器を決定する。TV を制御する場合は、スイッチを図 3-2(b)(g), チャンネルを(d)(e), ボリュームを(f)(g)のポーズを用い制御する。扇風機を制御する場合は、スイッチを図 3-2(b)(g), 風量調整を(c)(d)を用い制御する。このように、2 段階の手順を踏んで家電製品を制御することにより、また図 3-2 (b)のような普段起こりにくいポーズを最初のスイッチにすることにより誤動作を防ぐことができる。一定時間(ここでは 30 秒とした)命令が送信されない場合は、誤動作防止のためメニューを消し、最初から操作を行うようにした。図 3-35に実際の操作の流れをまとめる。

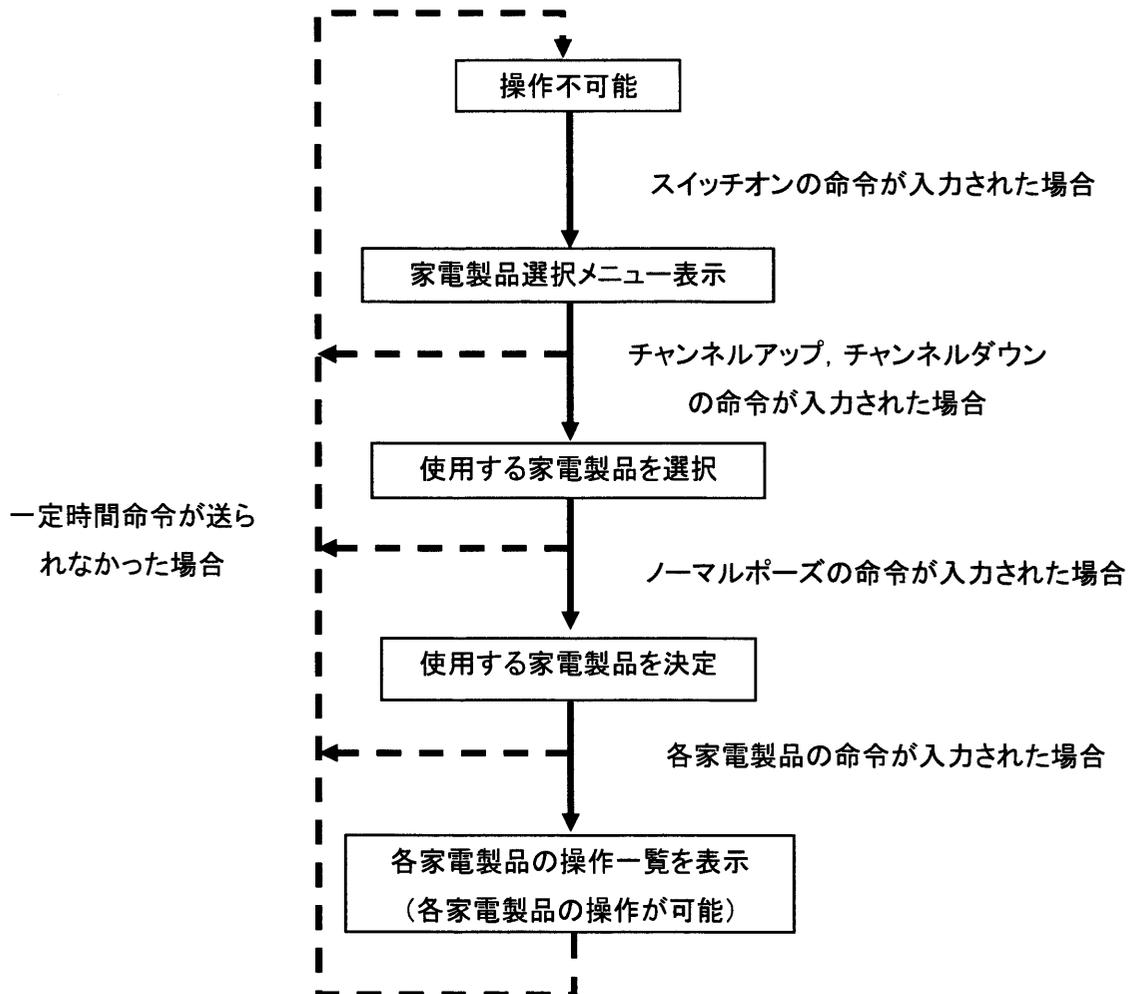


図 3-35 操作の流れ

3-10-2 手先のポーズを用い操作する場合

実際の操作では、利用者がベッド上で手を振ることにより、カメラが手の位置にズームされ、図 3-33に示すメニューがディスプレイ上に表示される。メニューが表示されている間だけ命令受け付け可能となる。次に図 3-29(d)(e)のアップダウンのポーズを入力することで操作したい機器にカーソルを合わせ、(a)のポーズをすることによって機器を決定する。TV を制御する場合は、スイッチを図 3-29(c)(h)、チャンネルを(d)(e)、ボリュームを(f)(g)のポーズを用い制御する。扇風機を制御する場合は、スイッチを図 3-29(c)(h)、風量調整を(d)(e)を用い制御する。このように、2 段階の手順を踏んで家電製品を制御することにより、またズームされない限りメニューが立ち上がらないため、誤動作を防ぐことができる。一定時間(ここでは 30 秒とした)命令が送信されない場合、または図 3-29(b)のようなリセットポーズを行った場合は、誤動作防止のためメニューを消し、最初から操作を行うようにした。図 3-36に実際の操作の流れをまとめる。

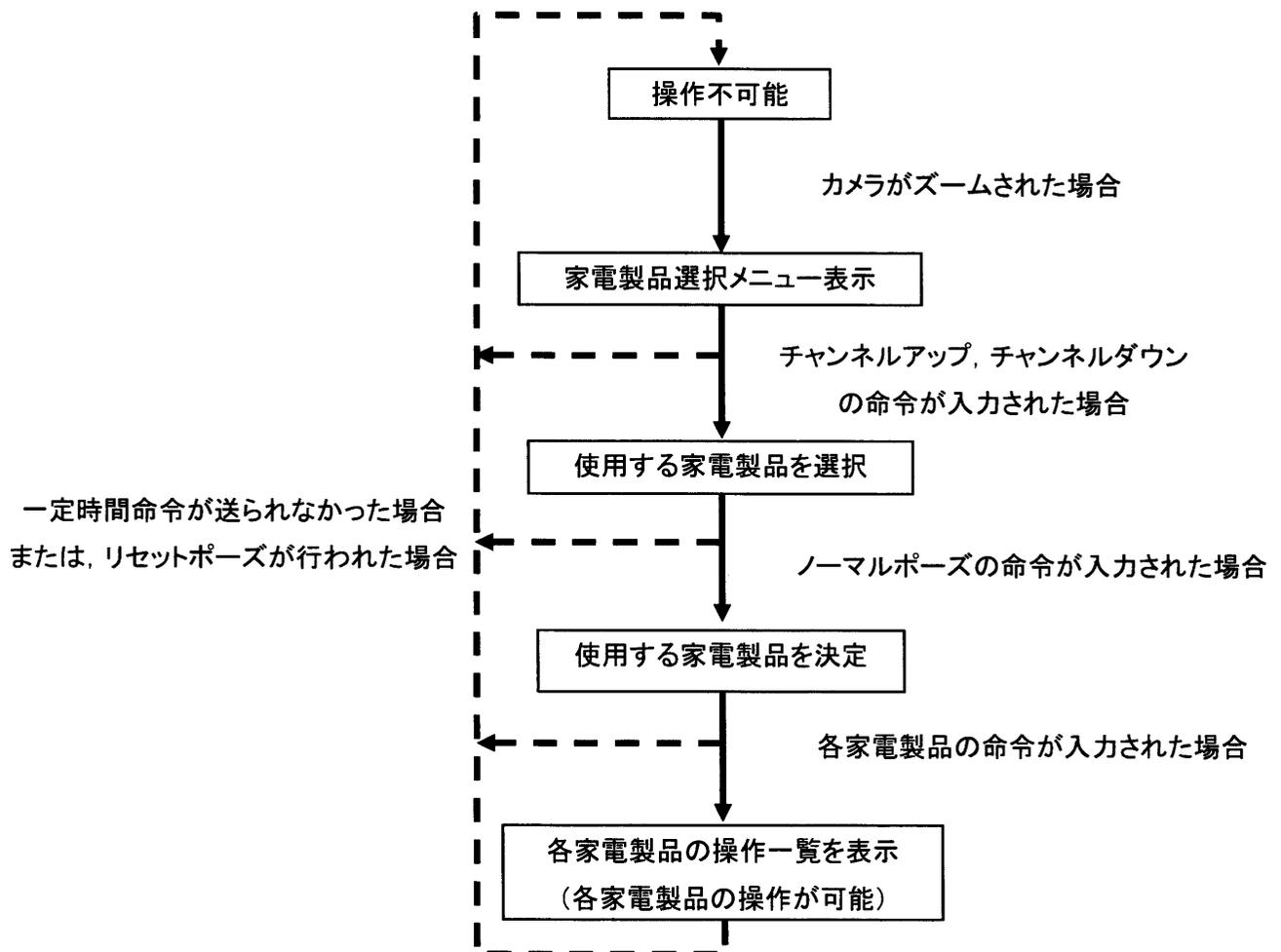


図 3-36 操作の流れ

3-11 まとめ

本研究では、高次元特徴と多変量データ解析を用いた画像認識を、家電製品制御システムへの応用事例を用いてデータ収集、特徴抽出、辞書作成法について検討し、パターン認識の多様性に関する考察を行った。また、高次元特徴として対象物体の平行移動に強い高次局所自己相関特徴を抽出し、次元圧縮は識別に有効な空間に写像する線形判別分析を用い、識別は学習辞書からのユークリッド距離を用いた。これらの認識システムを組み込んだ家電製品制御システムを用いて研究を行った。

辞書作成のための学習データを得る際に、短時間では十分な枚数が得られないなど様々な問題がある。そのような問題に対処するため、実際に存在する学習データに幾学的な変換を加えることにより、仮想学習データを作り出し、データを増やすことを提案し考察を行った。実験より仮想学習データを加えることによって、認識率は大きく改善された。これより仮想学習データを作成することの有効性が示された。

高次局所自己相関特徴は、傾きに対して弱いことが知られている。この場合、画像に幾学的な変換を施し傾かせた画像を作成し、仮想的に傾かせた画像を学習させることにより、傾きにも対応できるような特徴空間の作成を行った。実験より大きく傾かせた仮想学習データを学習させるほどクラス間分散が小さくなることが確認できた。また、適度な間隔を持たせて傾き角度を学習させることが有効であることも確認した。また高次局所自己相関特徴は、局所的であり空間に広がりをもった対象の区別に適しておらず、空間上の相対位置情報が必要な場合が生じる。そこで、特徴抽出領域を分割することにより、大まかな相対位置の情報を取り込んだ特徴量を抽出し、相対位置情報を持たせた特徴量についての考察を行った。実験結果より相対位置情報を持たせた高次局所自己相関特徴が有効であることを確認した。

クラスタリングの問題については、大幅に異なる特徴量を1つのクラスとした場合、クラス内分散が大きくなりすぎてしまい、クラス間分散が小さくなってしまふことが予想される。今回は複数のクラスタリング手法を提案し、それぞれ辞書を作成した。これらの辞書についての分離度等をクラス間距離の面から考察し、最適な辞書作成法を提案した。これらの実験よりポーズ認識の多様性に対する有効性が示された。

また本研究ではパターン認識の適用性として、認識対象が異なる場合でも同じアルゴリズムで識別できるかを考察した。具体的には手のポーズの判別に対しても認識部分のアルゴリズムが有効であるかを調べた。実験システムよりポーズ認識の適用性に対する有効性が示された。

4 性別推定について

人間の男女を形態的に判別する場合、女性は男性に比べて目が大きく、鼻が小さく、顎は細い、というように、髪型を除けば、額の広さ、顎の尖り具合等、顔に差があることが知られている[55].

そこで、顔画像を用い男女識別を行った研究として、画像を大域的正規化し、入力顔と性別平均顔との特徴点間の距離を求め、この距離と重み係数を乗じて統合したものを評価関数として判別する手法[87], 顔画像の形状, テクスチャ, 輝度パターンなどの特徴量を用い, 最近傍法, サポートベクターマシーンなどを複数組合せた識別器により判別する手法[88]がある. また, 正面顔を一定の大きさに正規化し, 主成分分析, GA の順で性差に関する特徴を選択し, 最後にニューラルネットを用い性差を判別する手法[89], 特徴量を Gabor ウェーブレット変換と Retina サンプリングにより抽出し, この特徴を用いてサポートベクターマシーンにより性別を推定する手法[90], 等がある. 一方, [91]は他の顔との差異を強調した似顔絵を用い, 被験者がこの似顔絵から性別認識が行えるか注視個所, 注視回数 of 点により考察を行っている.

しかし, 先に述べた研究は顔全体の特徴を利用したもので, 顎, 唇のみといった顔部品に注目しているものは見られない. そこで我々は, 顔部品に注目した場合, 顔のどの部分に性別を区別する特徴が存在するかを議論する. 男女識別には対象画像から四方向面特徴[21]を抽出した後に, 線形判別分析を適用する方法を用いる. この手法は, 個人を判別する方法として適用されている.

本研究では, どの顔部品が男女識別に有効であることを示す. また, 男女識別に成功した各顔部品画像の平均画像を作成することにより, どの部分に特徴の差が表れているかを調べる. 最後に年代による, 性別の差について検証した.

4-1 四方向面特徴・線形判別分析を用いた識別

始めに, 学習データの初期特徴ベクトルより最適な係数行列を算出し, これらのベクトルを係数行列により写像変換する. 次に, 写像変換後の各クラスの平均ベクトルを求め, これを学習辞書とする. 同様に, 未知データの初期特徴ベクトルも, 係数行列により写像変換する. 最後に, 変換後のベクトルと学習辞書を, ユークリッド距離を用いた 1 次識別器により比較することで識別を行う.

4-2 実験

4-2-1 データベース

今回はデータベースとして、ソフトピアジャパン HOIP 顔画像データベース [92]300 名(男女 150 人・正面・無表情・15~64 歳・日本人)から、1 名につき 4 枚、合計 1200 枚の画像を用意した。画像サイズは 640×480pixel で、RGB 各 8 ビットである。但し、女性は化粧をしている可能性がある。実験では濃淡画像を用いるため、カラー画像を濃淡画像に変換した。どのような顔部品の特徴が性別分類に有効であるかを確認するために、手動で切り出しを行い、以下の 6 つのパターンを用意した。表 4-1 に各パターンの説明を記す。またそれぞれの切り出した画像を図 4-1 に示す。

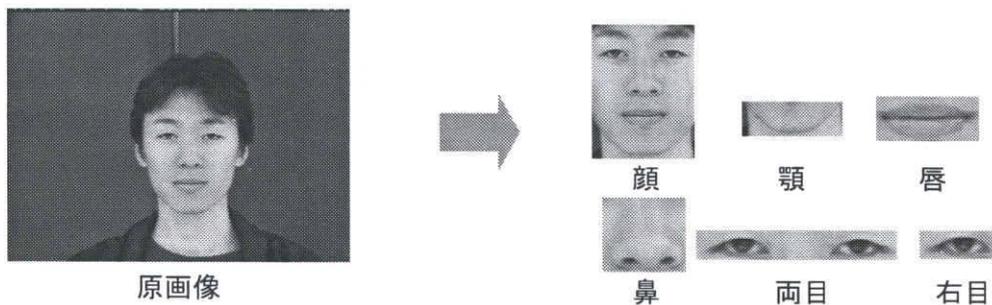


図 4-1 実験データ

表 4-1 各パターンについての説明

顔	眉毛の頂点を上端、顎を下端、両目の外眼角を左右の両端として切り出す。
顎	赤唇縁の最も下の部分を上端、顎の尖っている部分を下端、上端のラインと顎が交わる点を通り、上端のラインに対する垂線を左右の両端として切り出す。
唇	赤唇縁の最も上の部分を上端、赤唇縁の最も下の部分を下端、口角を左右の両端として切り出す。
鼻	下眼瞼縁の最も下の部分を上端、小鼻を左右下の端として切り出す。
両目	上眼瞼縁の最も上のラインを上端、下眼瞼縁の最も下のラインを下端、両目の外眼角を左右の両端として切り出す。
右目	上眼瞼縁の最も上のラインを上端、下眼瞼縁の最も下のラインを下端、右目の内眼角と外眼角を両端として切り出す。

4-2-2 最適な特徴次元数

性別推定を行うために、認識するための有効な特徴空間を調べる必要がある。そこで、より良い判別結果を求めるために、特徴次元数を増やしてしまうと2-5節でも記したように、次元の呪いに陥ってしまう。しかし、特徴量を圧縮することにより、どの程度特徴空間に影響をもたらすかは具体的に分かりにくい。また、特徴が適切でなければ高精度の認識は実現できないため、特徴空間の評価は重要である。そこで、特徴空間に関して、圧縮前と2-5-2節の手法を用いて圧縮した後のクラス内分散・クラス間分散比を調べることにより、圧縮することの利点を調べてみる。

図 4-1に示される各パターンから四方向面特徴を抽出する。表 4-2の 2 行目は低解像度を行った際の各パターンの解像度を、3 行目に各パターンから抽出される特徴次元数を示す。

男性のクラスを 600 枚、女性のクラスを 600 枚としてクラスタリングを行い、2-8節の方法により分散比を調べた。表 4-3の 2 行目は原特徴に対する分散比であり、表 4-3の 3 行目は判別分析により次元圧縮が行われた特徴に対する分散比である。また圧縮前の特徴次元数は 256 次元、圧縮後の特徴次元数は 1 次元となる。

表 4-2 各パターンにおける特徴次元数

パターン	顔	顎	唇	鼻	両目	右目
各面における解像度(pixel)	8×8	8×8	8×8	8×8	8×8	8×8
次元数	256	256	256	256	256	256

表 4-3 特徴空間の評価

パターン	顔	顎	唇	鼻	両目	右目
圧縮前の分散比	0.0967	0.0573	0.0996	0.0391	0.0427	0.0525
圧縮後の分散比	8.201	5.834	2.679	3.035	5.311	1.579

表 4-3より、圧縮前の分散比は全てのパターンに対して圧縮後の分散比より相当低くなっていることが確認できる。

圧縮前の分散比が低い数値を示した原因として、次の 2 点が考えられる。1 つは、特徴次元数が多い場合、相関の高い特徴が存在する可能性が大きい。もう 1 つは、特徴次元数が増えた場合、必要な学習パターンの数は指数関数的に増えてくる。よって今回 256 次元の特徴空間を利用したが、高々 1200 個の学習データでは特徴空間の分布を表すことができないと考えられる。このため期待したほどの分散比が得られなかった。

一方、圧縮後の分散比が高い数値を示した原因として、次のことが考えられ

る。判別分析は、積極的にクラス内分散が小さく、かつクラス間分散が大きい空間に次元圧縮する方法である。また次元圧縮を行ったために特徴空間の分布を表すために必要な学習データが原特徴空間の次元数を扱った場合と比べ、大幅に少なくなり、今回の 1200 枚の学習データで特徴空間の分布が表せたと考えられる。

また、顔部品を全て含む顔パターンは、比較的良好な部分空間が作成されていることが予想される。一方、顔以外の各顔部品のパターンは、顔パターンと比較すると分散比の値が低い。しかし顎、両目のパターンは比較的高い値を保っている。

4-2-3 最適な解像度

特徴量を抽出する際に、特徴量の正規化を行う必要がある。正規化を行う際に気をつけなくてはならないことは、各パターンに対する形状特徴をなるべく失わないように正規化を行う必要がある。例えば、図 4-1に示されている両目パターンは横幅(x 方向)が長く、x 方向の特徴量が重要になる可能性がある。そこで正規化によって特徴量の優位性が変化するかを調べた。表 4-4の 2 行目は低解像度を行った際の各パターンの解像度を、3 行目に各パターンから抽出される特徴次元数を示す。また、形状特徴を考慮しない場合と考慮した場合の相違点を簡単にまとめたものを表 4-5に示す。

表 4-4 形状特徴を考慮した際の各パターンにおける特徴次元数

パターン	顔	顎	唇	鼻	両目	右目
各面における解像度(pixel)	8×8	4×16	4×8	8×8	4×32	4×8
特徴次元数	256	256	128	256	512	128

表 4-5 形状特徴に関する相違点

形状特徴を考慮しない場合	<ul style="list-style-type: none"> ● 低解像度化した際の画像サイズは全てのパターンに対し 8×8 ● 抽出される特徴次元数は全てのパターンに対し 256 次元の特徴量を抽出 ● 各パターンの形状特徴を失っている可能性がある
形状特徴を考慮した場合	<ul style="list-style-type: none"> ● 低解像度化した際の画像サイズは各パターンにより異なる ● 抽出される特徴次元数は各パターンにより異なる ● 各パターンの形状特徴は保たれている

特徴量の優位性を調べるために表 4-2、表 4-4の特徴量を使用し性別推定実験を行った。識別機は平均値を辞書とした最近接識別器を用い、1200 枚のデータに対し、leave one out 法[20]を用いて評価を行った。認識率を表 4-6に示す。形状特徴を考慮しない場合を 2 行目に、形状特徴を考慮した場合を 3 行目に示す。また表内の数値の単位は%(正解枚数/総枚数[=1200])である。

表 4-6 形状特徴を考慮した際の各パターンにおける認識率

パターン	顔	顎	唇	鼻	両目	右目
形状特徴を考慮しない場合	97.7	94.9	92.2	90.6	83.8	84.2
形状特徴を考慮した場合	98.8	96.0	91.8	90.6	87.0	84.9

今回特徴抽出法を変えたことにより、唇以外の認識率が同等以上の結果が得られた。また、表 4-3で観察された分離度による予測とほぼ同等の結果を得た。

4-2-4 使用する特徴量

ここまでは四方向面特徴を用いて実験を行ってきたが、実際に四方向面特徴が最も適しているとは限らない。そこで表 4-7示す 4 種類の特徴を用いることにより、認識率がどのように変化するかを考察する。

表 4-7 特徴量の種類

エッジ特徴(Edge)	濃淡画像に対して Prewitt フィルタを用いることによりエッジ画像を作成し、それに対してガウシアンフィルタをかけ低解像度化したもの
濃淡特徴(Shade)	濃淡画像に対してガウシアンフィルタをかけ低解像度化したもの
四方向面特徴(FDF(all))	画像の各画素における濃淡値の縦、横、右斜め、左斜めの四方向のエッジ勾配をそれぞれ方向別に分解した四枚の画像を作成し、それらに対してガウシアンフィルタをかけ、低解像度化したものである。但しエッジを抽出する際に、全ての方向面に値を入れる
四方向面特徴(FDF(max))	画像の各画素における濃淡値の縦、横、右斜め、左斜めの四方向のエッジ勾配をそれぞれ方向別に分解した四枚の画像を作成し、それらに対してガウシアンフィルタをかけ、低解像度化したものである。但しエッジを抽出する際に、各面のそれぞれのピクセルに対して濃度値を比較し、最大値を取る面に対してのみ値を入れる

評価を均等にするために、全ての特徴量の次元数をそろえた。従って Edge, Shade の場合、各パターンを低解像度化した画像サイズは表 4-8の 2 行目、3 行目に示される解像度となり、FDF1, FDF2 の場合、各パターンを低解像度化した画像サイズは表 4-8の 4 行目、5 行目に示される解像度となる。これらより、各パターンの最終的な特徴次元数は表 4-8の 6 行目に示されるものとなる。識別機は平均値を辞書とした最近接識別器を用い、1200 枚のデータに対し、leave one out 法を用いて評価を行った。認識率を表 4-9に示す。Edge 特徴を使用した場合を 2 行目に、Shade 特徴を使用した場合を 3 行目に、FDF(all)を使用した場合を 4 行目に、FDF(max)を使用した場合を 5 行目に示す。形状特徴を考慮しない場合を 2 行目に、形状特徴を考慮した場合を 3 行目に示す。また表内の数値の単位は%(正解枚数/総枚数[=1200])である。

表 4-8 各パターン、各特徴における特徴次元数

パターン	顔	顎	唇	鼻	両目	右目
解像度(Edge)	16×16	8×32	8×16	16×16	8×64	8×16
解像度(Shade)	16×16	8×32	8×16	16×16	8×64	8×16
解像度(FDF1)	8×8	4×16	4×8	8×8	4×32	4×8
解像度(FDF2)	8×8	4×16	4×8	8×8	4×32	4×8
特徴次元数	256	256	128	256	512	128

表 4-9 各パターン、各特徴における認識率

パターン	顔	顎	唇	鼻	両目	右目
Edge	95.9	91.1	87.4	83.7	75.4	76.5
Shade	99.3	92.4	93.3	89.3	81.9	83.7
FDF(all)	98.8	96.0	91.8	90.6	87.0	84.9
FDF(max)	96.5	91.3	90.0	95.4	82.5	89.8

まず Edge の場合は全てのパターンに対して認識率が最も低いことが確認できる。次に Shade の場合は顔、唇パターンに対して最も高い認識率が得られ、顎パターンに対しては FDF(max)よりは高い認識率が得られた。しかし、その他のパターンに対しては、FDF(all)、FDF(max)より低い認識率が得られている。

顔、唇が高い認識率が得た理由として、女性は男性に比べて肌の色が白いため、女性は口紅を塗っている人が多いためと考えられる。しかし Shade の場合、今回はデータベースを用いて実験を行っており、環境が整いすぎていたためにこのような結果が得られたと考えられる。光による影響が入ってきた場合にこの特徴は使えなくなるであろうことが懸念される。

また、FDF(max)は鼻、右目に関しては最も良い認識率が得られている。この原因を記す前に表 4-10を見てもらいたい。

表 4-10 各パターンの縦、横の大きさの平均

パターン	顔	顎	唇	鼻	両目	右目
高さ平均	145.9958	47.9908	35.9883	52.0233	21.0167	20.0000
幅平均	144.0292	143.9433	75.9750	57.0075	131.0000	44.0108

表 4-10は各パターンの縦・横サイズの平均であり、単位は pixel である。興味深いことに縦サイズ平均、横サイズ平均の大きい方の値に注目した場合、鼻の場合は約 57 pixel、右目の場合は約 44 pixel となっており、画像サイズとしては小さいと考えられる。推測ではあるが、特徴を抽出するための原画像のサイズが小さい場合、FDF(max)を用いた方が認識率が良くなり、特徴を抽出するための原画像のサイズが大きい場合、FDF(all)を用いた方が認識率が良くなると考えられる。

4-2-5 未知データに対する評価方法

ここまでの評価は leave one out 法を用いてきた。しかし leave one out 法は未知データが1枚となってしまうため、学習データの中に同一人物のデータが含まれることになってしまう。このため、性別の情報だけでなく個人の情報も含まれてしまい、認識率が向上していることも考えられる。そこで n fold Cross Validation 法[20]によって学習データの中に未知データの人物が含まれない場合も検討する必要がある。そこで n fold Cross Validation 法を用いてデータの評価を行った。被験者1人につき4枚のデータがあるため、同一人物である4枚を1つの集合した。これより、n=300となる。n fold Cross Validation 法の認識率を表 4-11示す。また表内の数値の単位は%(正解枚数/総枚数[=1200])である。

表 4-11 n cross validation 法による認識率

パターン	顔	顎	唇	鼻	両目	右目
Edge	86.4	86.5	83.6	76.1	70.2	74.3
Shade	94.1	87.0	90.2	83.2	76.7	80.4
FDF(all)	93.7	89.8	87.3	77.7	78.3	79.8
FDF(max)	88.9	84.7	86.1	89.4	73.1	86.1

n cross validation 法により評価したことにより、全体的に認識率が低下した。leave one out 法による評価を用いた際に認識率が高くなる理由としては、未知データとして使用された人物が学習データに含まれてしまい、未知データで使用された人物の特徴を予め学習した為である。

使用された1200枚の画像が異なる人物を使用した場合、leave one out 法は有効な結果が得られると考えられる。しかし、今回のように、個人情報に依存せずに特徴を評価したい場合で、同一人物の画像が複数枚存在する場合は n cross validation 法を用いた方が、データベースに対して正しい評価を行うことができる。ここまでの全体的な評価とからこれから先の実験で使用する特徴量は FDF(all)を使用する。

4-2-6 最適な方向面

前節までの実験で四方向面特徴は性別推定に対して適している特徴量であることがいえる。しかし、特徴抽出を行う際の面数を多くした方が、より良い特徴量となる可能性がある。そこで、多方向面の特徴を抽出するために、Gabor ウェーブレットを使用する。今回は特徴面を1~10 方向、特徴を抽出する解像度を1×1~10×10 とした。また、特徴抽出の対象となる画像は図 4-1に示す顔パターンを対象とし、線形判別分析による性別推定実験を行った。特徴抽出のアルゴリズムは図 2-10に示されているものを使用する。

特徴を抽出する解像度の1 辺の長さをそれぞれ2 倍にし、これを繰り返す、1 辺の長さが128 を超えた場合に計算を終了し、それまでに求められたサイズを用い、低解像度化を行った。具体的な数値を図 4-2に示す。また表内の数値の単位は pixel である。

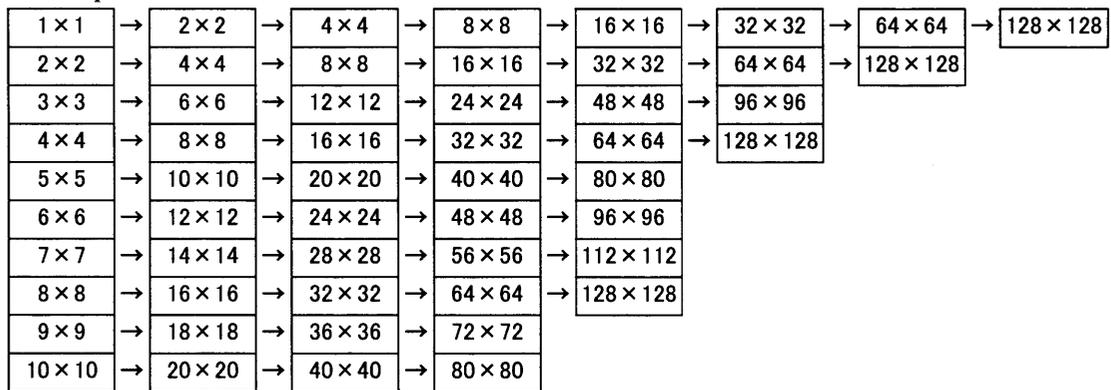


図 4-2 最初の正規化における解像度

1 方向を抽出するといっても、どの角度を抽出するかが分からない。今回は π を方向面の数で割った値の倍数を使用した。具体的な数値を図 4-3に示す。また表内の数値の単位は° である。

1 方向	2 方向	3 方向	4 方向	5 方向	6 方向	7 方向	8 方向	9 方向	10 方
0	0	0	0	0	0	0	0	0	0
	90	60	45	36	30	25.7	22.5	20	10
		120	90	72	60	51.4	45	40	20
			135	108	90	77.1	67.5	60	30
				144	120	102.9	90	80	40
					150	128.6	112.5	100	50
						154.3	135	120	60
							157.5	140	70
								160	80
									90

図 4-3 特徴抽出面について

各解像度と特徴抽出面の数による特徴次元数の変化を表 4-12に示す。表内の数値は特徴次元数を示す。

表 4-12 各解像度，各方向別において抽出される特徴次元数

	1方向	2方向	3方向	4方向	5方向	6方向	7方向	8方向	9方向	10方向
1×1	1	2	3	4	5	6	7	8	9	10
2×2	4	8	12	16	20	24	28	32	36	40
3×3	9	18	27	36	45	54	63	72	81	90
4×4	16	32	48	64	80	96	112	128	144	160
5×5	25	50	75	100	125	150	175	200	225	250
6×6	36	72	108	144	180	216	252	288	324	360
7×7	49	98	147	196	245	294	343	392	441	490
8×8	64	128	192	256	320	384	448	512	576	640
9×9	81	162	243	324	405	486	567	648	729	810
10×10	100	200	300	400	500	600	700	800	900	1000

これまでに示した特徴量を用いて識別実験を行った。識別には線形判別分析を用いた。1200枚のデータに対し、n cross validation法を用いて評価を行った。表 4-13に認識率を示す。今回は1人につき4枚画像が存在するため、この4枚を1セットとした。このためn=300となる。また表内の数値の単位は%(正解枚数/総枚数[=1200])である。

表 4-13 各解像度，各方向別における認識率

	1方向	2方向	3方向	4方向	5方向	6方向	7方向	8方向	9方向	10方向
1×1	64.8	63.8	64.9	66.7	65.1	66.3	65.4	65.8	65.7	65.3
2×2	64.2	68.2	68.2	71.1	70.8	70.6	70.1	70.3	71.6	72.2
3×3	75.2	78.0	81.8	82.5	82.3	82.3	82.3	81.6	81.3	81.5
4×4	78.5	81.8	86.1	85.4	85.7	85.8	85.8	86.2	86.0	86.9
5×5	77.5	81.2	88.2	88.8	88.1	88.0	87.3	87.6	85.2	86.5
6×6	80.4	82.3	87.3	88.5	89.5	87.1	87.2	88.0	87.1	88.2
7×7	82.4	84.0	90.1	90.3	89.0	90.3	88.9	89.0	88.3	86.7
8×8	84.9	88.3	91.6	91.3	89.4	89.3	89.5	88.1	89.4	85.7
9×9	86.7	88.5	89.5	90.9	87.8	89.5	88.7	87.9	84.5	81.5
10×10	87.1	88.6	89.4	90.3	89.7	88.8	86.9	84.9	82.3	74.3

表 4-13では分かりづらいので、グラフにしたものを図 4-4に示す。

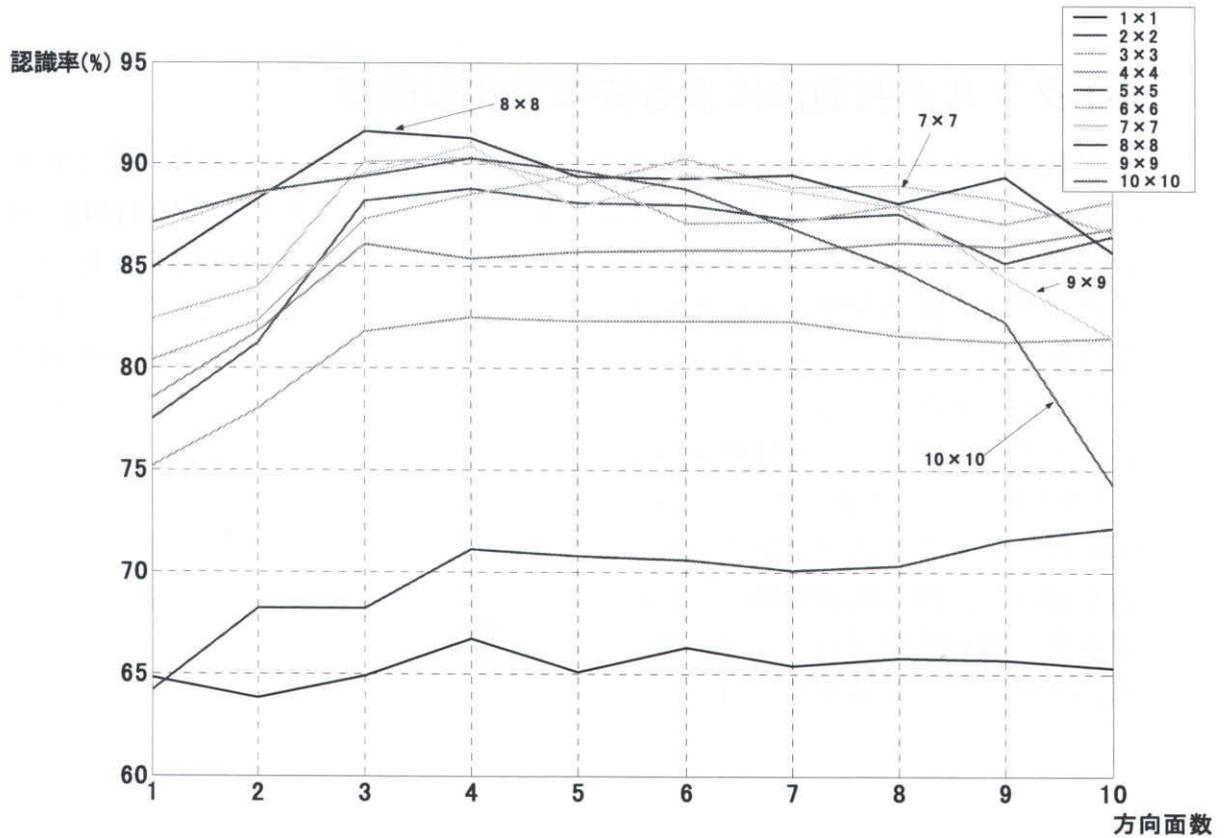


図 4-4 認識率

図 4-4よりほとんどの解像度に対して、方向面が 5 枚の場合は、方向面が 4 枚の場合よりも認識率が下がっている。また、方向面が 7 枚の場合もほとんどの画像に対し、方向面が 6 枚の場合よりも、認識率が僅かではあるが下がっている。

解像度については、7×7あたりから識別面数が多くなるに従い、認識率が落ちてゆく傾向が見られる。これは、特徴次元数が多くなりすぎたため、次元の呪いが起こっていると考えられる。

今回の実験では、識別面数を増やせば特徴次元数が必然的に増える。これにより、特徴を抽出する解像度がかなり小さい場合でも、認識率が多少は上昇すると予想していたが、実際には変化は無く、3×3、もしくは4×4で認識率のピークが表れている。

よって、Gabor ウェーブレットにより、識別面を増やしたからといって、認識率が劇的に変わるわけではなく、むしろ、特徴を抽出する解像度の方が、情報量として重要であることが確認できた。

4-3 クラスタ判別法による特徴空間の観察

認識率の良さに加えて、さらに特徴空間上でクラスの重なりがなければ特徴は高い信頼を持つことができると考えられる。そこでクラスタ判別法[64]を用い特徴空間の可視化を行い、観察を行う。これは原特徴空間で構造解析を行い、このクラスタ構造を保存し低次元空間への写像を求め可視化を行う手法である。原特徴空間でクラスの重なりがあった場合は、根本的にクラスの分離が困難であることが予測される。そこで、各パターンの特徴空間でどの程度重なりが存在するかを観察し、原特徴量がもっている情報を把握する。

3-2-1 節で述べた各パターンに対し、男性のクラスを 600 枚、女性のクラスを 600 枚としてクラスタリングを行った。これらのデータに対し、クラスタ判別法を適用し、四方向面特徴による特徴空間を 2 次元に可視化したものを図 4-5 に示す。濃い色でプロットされた点が男性のデータ、薄い点でプロットされた点が女性のデータを示す。

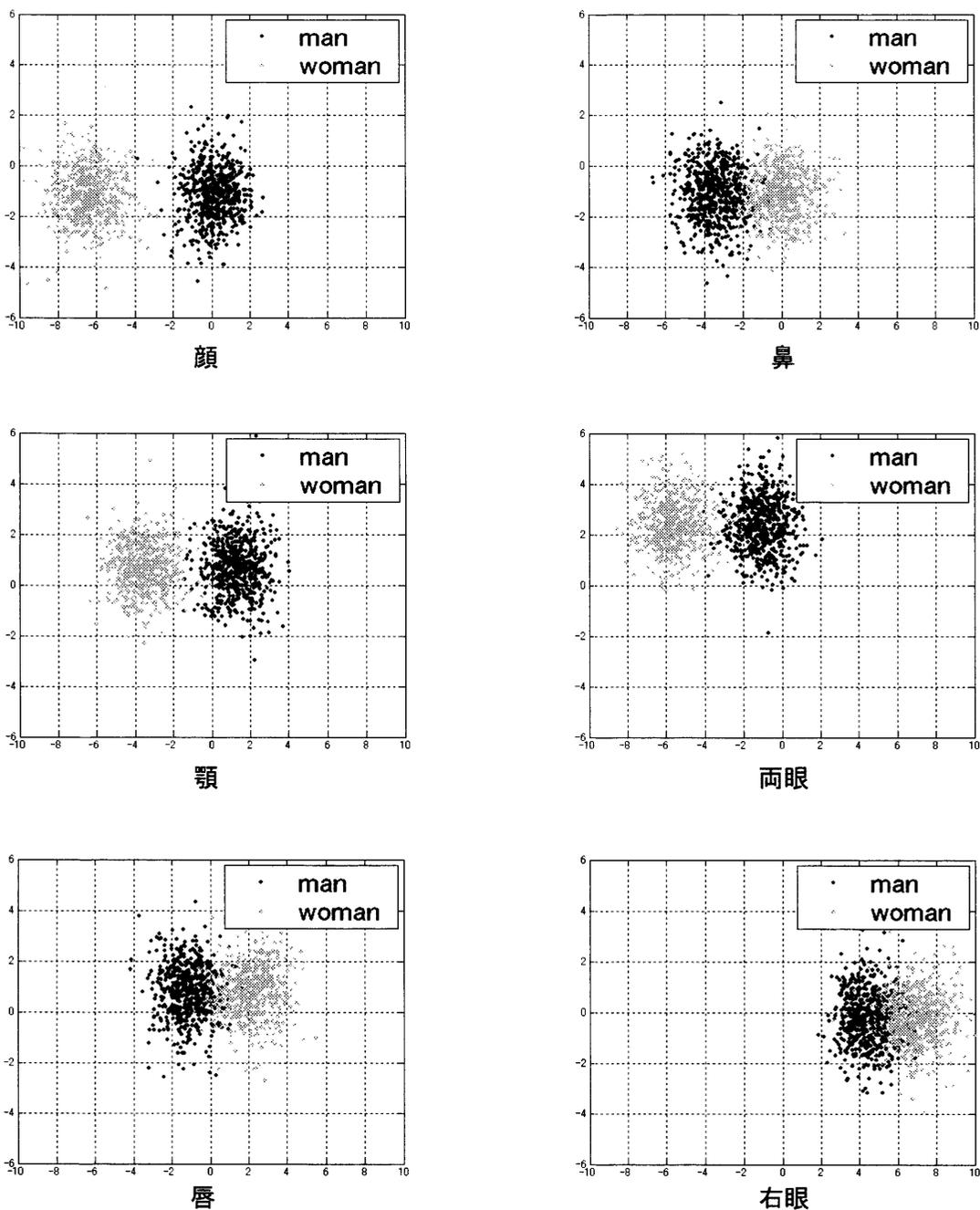


図 4-5 各パターンにおける四方向面特徴の特徴空間

4-4 結果及び考察

表 4-11より、認識率が良好なパターンは、顔、顎、唇の3つのパターンで、それ以外の鼻、両目、右目は先の3つのパターンと比べると10%程認識率が低下している。

10%も認識率に差が表れた原因を調べるために、先の分類実験で性別を正しく分類した各顔部品のパターンより、図 2-3に示すような四方向面パターンを作成し、それぞれの面に対し平均画像を作成した。分類実験により、高い認識率を示した顔、顎、唇の3パターンについて、男性と正しく分類された方向別の平均画像をそれぞれ図 4-6に、同様に女性の場合を図 4-7に示す。

顔のパターンの場合、男性より女性の方が顎、唇、目の辺りに強いエッジが表れている。顔のパターン内で顔部品別に見ると、女性の場合、顎は全ての方向面に対し、上顎の部分に差が見られ、唇は特に唇の水平エッジに差が見られる。目については、男性に比べ女性の方がやや大きい。これらのうち、唇、目は女性の化粧による影響と考えられる。このように明確な差が随所に見られたため、顔のパターンが、最も高い認識率を得ることができたと考えられる。また、顔のパターンの認識率が、顎、唇のパターンと比べて、さらに良かったことから、男女間の判別には各顔部品の位置関係も重要であると考えられる。加えて、目のエッジに明らかに男女差が見られる反面、右目の認識率が低かった理由は、女性の目への化粧による個人差が大きく、男女分類に悪影響を及ぼしたためと考えられる。

顎のパターンの場合、男性より女性の方が全ての方向面に対して上顎の部分に強いエッジが表れている。顔のパターンでも上顎の部分に差が見られたが、顎のパターンだけの方がより明確な差が出ている。

唇のパターンの場合、垂直成分を除いた方向面において、男性より女性の方が肌と唇の境の部分に強いエッジが表れている。これは口紅などによるエッジの出やすさの影響と考えられる。

これまでに述べたように、顔、顎、唇のパターンでは、各面の平均画像を作成した際に、性別を判断するための明確な差が見られたが、その他のパターンでは各面の平均画像に明確な差が表れなかった。

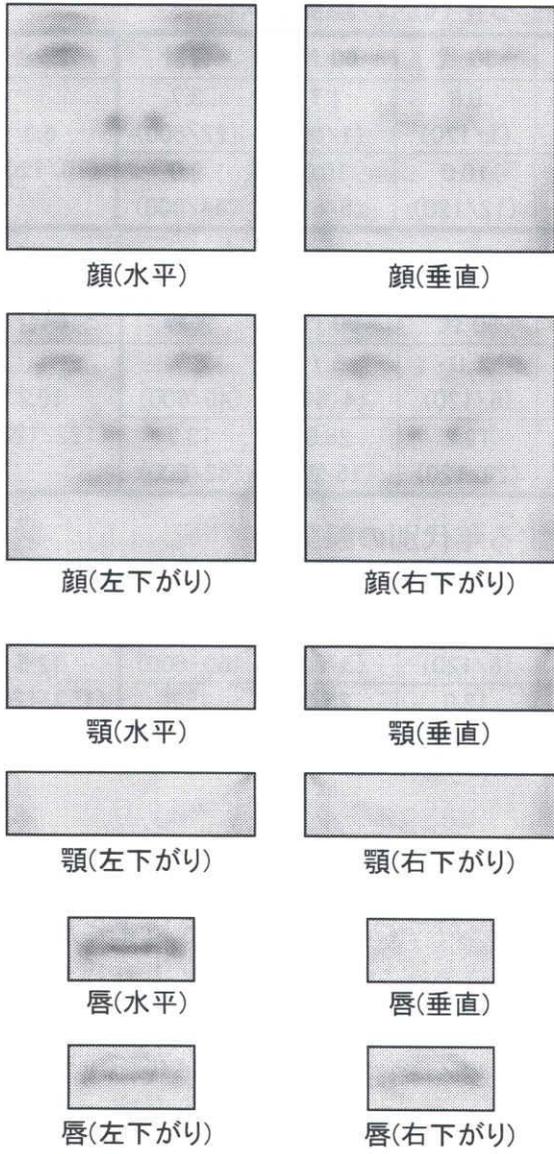


図 4-6 男性の平均画像

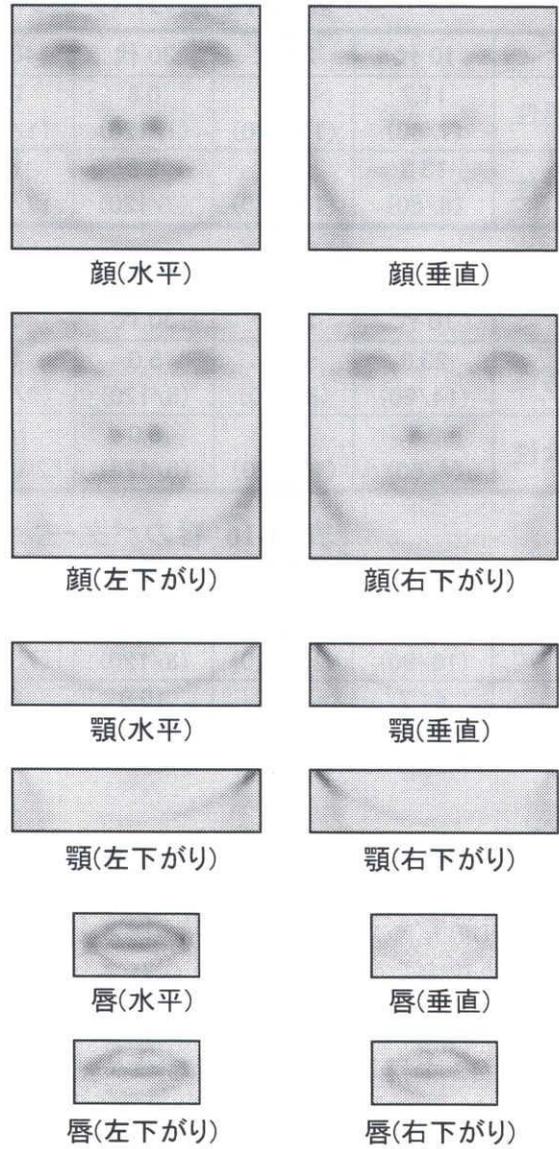


図 4-7 女性の平均画像

表 4-14 顔のパターンに対する年代別の誤認識率

	10代	20代	30代	40代	50代	60代	合計	総合
男性	11.7 (7/60)	9.1 (11/120)	0.8 (1/120)	0.8 (1/120)	0.8 (1/120)	1.7 (1/60)	3.7 (22/600)	6.3 (76/1200)
女性	13.3 (8/60)	11.7 (14/120)	5.8 (7/120)	5.8 (7/120)	10.0 (12/120)	10.0 (6/60)	9.0 (54/600)	

表 4-15 顎のパターンに対する年代別の誤認識率

	10代	20代	30代	40代	50代	60代	合計	総合
男性	23.3 (14/60)	6.7 (8/120)	5.0 (6/120)	1.7 (2/120)	5.0 (6/120)	6.7 (4/60)	6.7 (40/600)	10.2 (122/1200)
女性	6.7 (4/60)	11.7 (14/120)	5.0 (6/120)	16.7 (20/120)	19.2 (23/120)	25.0 (15/60)	13.7 (82/600)	

表 4-16 唇のパターンに対する年代別の誤認識率

	10代	20代	30代	40代	50代	60代	合計	総合
男性	26.7 (16/60)	16.7 (20/120)	6.7 (8/120)	5.8 (7/120)	5.0 (6/120)	5.0 (3/60)	10.0 (60/600)	12.8 (153/1200)
女性	41.7 (25/60)	11.7 (14/120)	10.8 (13/120)	9.2 (11/120)	12.5 (15/120)	25.0 (15/60)	15.5 (93/600)	

4-5 各年代に対する詳細な考察

前節でも述べたように、例えば、唇のパターンでは性差が口紅等による影響により表れていると考えられる。そこで、このような影響が年代により異なる可能性があると考え、先の分類実験で高い認識率を得た、顔、顎、唇の3パターンに注目した。これらのパターンに対し、性別分類を行う際に、年代により、どの程度差が現れるかを調べた。表 4-14、表 4-15、表 4-16に各パターンの誤認識率の詳細な内訳を示す。表内の括弧の上に記されている数値は認識率(単位は%)を、括弧内に記されている数値は(正解枚数/総枚数)を示す。

表 4-14、表 4-15、表 4-16より、顔のパターンはどの年代に対しても高い誤認識率は見られなかった。しかし、顎のパターンの場合 10代の男性、60代の女性で高い誤認識率が見られ、唇のパターンでは 10代の男性、10代の女性、60代の女性で高い誤認識率を得た。この中で、唇のパターンにおいて、10代の女性の誤認識率は 41.7%とかなり高い数値を示している。これは、10代の女性は口紅をつけている人が他の女性の年代と比べて少ないことが原因と考えられる。また、各性別、各年代について表 4、5、6に示すように、それぞれを1つのクラスと見なし、各パターンで誤認識率が高いものから順に選出した場合、上位に必ず 10代の男性、60代の女性を含む現象が見られた。これより、10代の男性、60代の女性は性別が分類しにくい可能性が高い。

4-6 まとめ

複数の顔部品からそれぞれ四方向面特徴を抽出し、線形判別分析を行った場合、顔のどの部分に性差が現れるかを検討した。顔、顎、唇のパターンが高い認識率を示し、この中で最も識別が行いやすいパターンは全ての顔部品の位置情報を含んでいる顔パターンであり、次いで顎、唇のパターンの順に識別が行いやすいことが判明した。また、各性別、各年代でクラス分けを行い、各クラス間の性別分類の差を調べたところ、10代の男性、60代の女性のクラスにおいては、誤認識率が高くなる可能性を示唆した。

5 むすび

本論文では、画像処理を用いたユビキタスインターフェースの研究として家電製品制御システムと顔部品を用いた性別推定に関して考察を行った。

家電製品制御システムに関しては識別機、辞書作成法、特徴抽出法、異なった認識対象への拡大について検討した。識別機に関しては、部分空間法、複合類似度法、混合類似度法、線形判別分析による比較を行ったところ、本システムでは線形判別分析が最も適していた。また、辞書作成方法に関しては、仮想学習データを作成し、学習データとして使用することで利用者が傾いた場合に対応できるようにし、十分な学習データが得られない場合でも安定した辞書が作成できることを確認した。特徴抽出法に関しては顔周辺矩形を分割し、各領域から高次局所自己相関特徴を抽出することにより、利用者が半袖を着用している場合、長袖を着用している場合に対応できることを確認した。さらに、服装ごとに辞書を用意することなく、一つの辞書で複数の服装に対応できる辞書を作成し、その辞書が有用であることも確認した。最後に手先のポーズを用いて本システムを稼働させた。これは、上半身を使用したポーズの識別アルゴリズムと同様のアルゴリズムを用いており、異なった認識対象でも本システムの有用性を示した。家電製品制御支援システムに関する今後の課題は、被験者が変わった場合でもシステムが稼働する方法についての検討、特徴量についての検討、環境が変わった場合への対応が考えられる。顔部品を用いた性別推定に関しては、識別を行う際の特徴次元数、解像度、特徴量、方向面について考察を行い、それぞれ有効な特徴次元数、解像度、特徴量、方向面を確認した。データ評価方法は *n fold Cross Validation* と *leave one out* 法の比較を行い、個人情報に性別推定に与える影響を調べた。また特徴空間に関して数値的、視覚的に考察を行い、顔、顎のパターンが有用であることを確認した。また各年代に関する考察を行った。顔、顎、唇のパターンが高い認識率を示し、この中で最も識別が行いやすいパターンは全ての顔部品の位置情報を含んでいる顔パターンであり、次いで顎、唇のパターンの順に識別が行いやすいことが判明した。また、各性別、各年代でクラス分けを行い、各クラス間の性別分類の差を調べたところ、10代の男性、60代の女性のクラスにおいては、誤認識率が高くなる可能性を示唆した。顔部品を用いた性別推定に関する今後の課題は、各顔部品による特徴、もしくは結果の統合により識別精度を向上させることが課題である。さらに、分散値等の統計学的な観点や認知科学的な視点などから、有効な特徴が存在するかどうかについて検討しなくてはならない。また、実環境におけるリアルタイムデモシステムの構築を行う必要がある。

謝辞

本研究は工学部応用情報学科山本和彦教授のご指導の下に遂行いたしました。山本教授には、指導教授として多大なる御助言をいただきました。記して深く感謝いたします。

博士論文をまとめるにあたりご指導いただきました副査の工学部人間情報システム工学科、谷和男教授、医学研究科再生医科学専攻再生工学講座知能イメージ情報分野工学研究科、藤田廣志教授、工学部応用情報学科、速水悟教授に感謝いたします。

工学部応用情報学科加藤邦人助手をはじめ、元地域結集型共同研究プロジェクト(HOIP)主任研究員、三洋電機(株)ハイパーメディア研究所の本郷仁志氏、山本研究室卒業生の伊藤慎哉氏(日信ソフトエンジニアリング(株))ならびに山本研究室の皆様方には熱心かつ有益な御討論ならびに御協力を頂きました。ここに記して深く感謝申し上げます。また家電製品制御支援システムの成果は実用化に向けて中部コンピューター(株)に技術移転されました。技術移転に当り協力いただいた長井幸志氏、花木晋也氏に感謝いたします。

研究の一部は科学技術庁地域結集型共同研究事業の一環として行われました。

参考文献

- [1] A.Ohsuga, Y.Nagai, Y.Irie, M.Hattori and S.Honiden: "PLANGENT: An approach to making mobile agents intelligent", IEEE Internet Computing, vol.1, no.4, pp.50-57 (1997)
- [2] 川村隆浩, 田原康之, 長谷川哲夫, 大須賀昭彦, 本位田真一: "Bee-gent: 移動型仲介エージェントによる既存システムの柔軟な活用を目的としたマルチエージェントフレームワーク", 信学論(D-I), vol.J82-D-I, no.9, pp.1165-1180 (1999)
- [3] M.Hattori, N.Kase, A.Ohsuga and S.Honiden: "Agent-based Drivers' Information Assistance System", New Generation Computing, vol.17, no.4, pp.359-367 (1999)
- [4] 池谷直紀, 加瀬直樹, 大須賀昭彦: "エージェント技術を適用したヒューマンナビゲーションシステム", 情報処理学会研究報告, vol.1999, no.080, 1999-MBL-010, pp.97-104 (1999)
- [5] M.Weiser: "Some Computer Science Issues In Ubiquitous Computing", Communications of the ACM, vol.36, no.7, pp.74-83 (1993)
- [6] 山田茂樹, 上岡英史: "ユビキタスコンピューティング; ネットワークとアプリケーション", 信学論(B), vol.J86-B, no.6, pp.863-875 (2003)
- [7] M.Weiser: "The Computer of the 21st Century", Scientific American, vol.265, no.3, pp.66-75 (1991)
- [8] 竹内英世, 梅崎太造, 松本憲幸, 平林勝巳: "かすれ指紋画像の判定・復元機能をもつ指紋照合法", 信学論(D-II), vol.J87-D-II, no.2, pp.574-585 (2004)
- [9] 中田康之, 安藤護俊: "色抽出法と固有空間法を用いた読唇処理", 信学論(D-II), vol.J85-D-II, no.12, pp.1813-1822 (2002)
- [10] 佐久間聡, 高橋裕子, 鈴木章, 塩昭夫, 大塚作一: "一對のラインセンサカメラを用いたナンバプレート位置を基準とした車速計測", 信学論(D-II), vol.J87-D-II, no.6, pp.1233-1243 (2004)
- [11] M.-P.D.Jolly, S.Lakshmanan and A.K.Jain: "Vehicle Segmentation and Classification Using Deformable Templates", IEEE Trans. Pattern Anal. & Mach. Intell., vol.18, no.3, pp.293-308 (1996)
- [12] W.Y.Kan, J.V.Krogmerier and P.C.Doerschuk: "Model-based Vehicle Tracking from Image Sequences with an Application to Toad Surveillance", Opt. Eng., vol.35, no.6, pp.1723-1729 (1996)
- [13] T.N.Tan, G.D.Sullivan and K.D.Baker: "Closed Form Algorithms for Object Pose and Scale Recovery in Constrained Scenes", Pattern Recognit., vol.29, no.3,

- pp.449-461 (1996)
- [14] 金平実, 定野温: "運転操作補助に関する車両制御技術", 計測と技術, vol.40, no.3, pp.214-219 (2001)
- [15] D.Pomerleau and T.Jochem: "A Rapidly Adapting Machine Vision for Automated Vehicle Steering", IEEE Expert: Intelligent Systems and Their Applications, vol.11, no.2, pp.19-27 (1996)
- [16] D.J.Kang, J.W.Choi and I.S.Kweon: "Finding and Tracking Road Lanes Using Line-snakes", Proc. of Conf. on Intelligent Vehicle Symp., pp.189-194 (1996)
- [17] Y.Wang, D.Shen and E.K.Teoh: "Lane Detection Using Catmull-Rom Spline", Proc. of Conf. on Intelligent Vehicles Symp., Stuttgart, vol.1, pp.51-57 (1998)
- [18] 小森鉄平, 藤原幸広, 藤田政之, 内田健康: "ビジュアルサーボによる自動車の自動操舵", 電学論(C), vol.120, no.4, pp.501-506 (2000)
- [19] 小平実, 大友照彦, 田中敦, 岩月正見, 大内隆夫: "ニューラルネットを用いた移動ロボット車の障害物回避走行制御", 信学論(D-II), vol.J79-D-II, no.1, pp.91-100 (1996)
- [20] 坂野鋭, 山田敬嗣: "奇怪!!次元ののろいー識別問題, パターン認識, データマイニングの初心者のためにー", 情報処理, vol.43, no.6, pp.658-663 (2002)
- [21] K.Yamamoto: "Present state of Recognition Method on Consideration of Neighbor Points and Its Ability in Common Database", ICICE, Trans.Inf. and Syst., vol.E79-D, no.5, pp.417-42 (1996)
- [22] 松井俊弘, 山下郁生, 若原徹, 吉室誠: "第1回文字認識技術コンテスト実施結果", 「手書き文字認識技術の過去・現在・未来」シンポジウム講演論文集, pp.15-22 (1993)
- [23] 荻田紀博, 内藤誠一郎, 増田功: "外郭方向寄与度特徴による手書き漢字の識別", 信学論(D), vol.J66-D, no.10, pp.1185-1192 (1983)
- [24] 堀桂太郎, 根本孝一, 伊藤彰義: "文字の輪郭線に着目した手書き漢字の特徴抽出法ー外郭局所的輪郭線特徴と外郭局所的モーメント特徴ー", 信学論(D-II), vol.J82-D-II, no.2, pp.188-195 (1999)
- [25] 岡隆一: "セル特徴を用いた手書き漢字の認識", 信学論(D), vol.J66-D, no.10, pp.17-24 (1983)
- [26] 梅田三千雄: "手書き漢字認識手法の個性的なサーベイ", 「手書き文字認識技術の過去・現在・未来」シンポジウム講演論文集, pp.54-63 (1993)
- [27] 川勝英史, 力宗幸男: "カテゴリー間筆記特性を用いた手書き数字による筆記者の特定ー異なるカテゴリー対のパターンを比較する方法ー", 信学論(D-II), vol.J83-D-II, no.7, pp.1647-1657 (2000)

- [28] 孫寧, 田原透, 阿曾弘具, 木村正行: "方向線素特徴量を用いた高精度文字認識", 信学論(D-II), vol.J74-D-II, no.3, pp.330-339 (1991)
- [29] 大町真一郎, 阿曾弘具: "つぶれを考慮した低品質印刷文字の高精度認識", 信学論(D-II), vol.J79-D-II, no.9, pp.1534-1542 (1996)
- [30] D.Gabor: "Theory of Communication", Journal of IEE, vol.93, pp.429-457 (1946)
- [31] J.G.Daugman: "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-dimensional Cortical Filters", Journal of Optical Society of America A, vol.2, pp.1160-1169 (1985)
- [32] M.T.Suzuki, T.Kato and H.Tsukune: "3D Object Retrieval Based on Subjective Measures", Proc. of 9th International Conf. and Workshop on Database and Expert Systems Applications, pp.850-856 (1998)
- [33] 鈴木一史, 加藤俊一, 築根秀男: "主観的類似度に対応した 3 次元多面体の検索", 信学論(D-I), vol.J82-D-I, no.1, pp.184-192 (1999)
- [34] 西村剛, 寺本純司, 長田秀信, 紺谷精一: "3 次元物体データベースにおける類似物体検索の検討", 情処学データベースシステム研報, vol.2000, no.069 2000-DBS-122, pp.479-486 (2000)
- [35] 江本雄治, 小早川倫広, 大西建輔, 星守, 大森匡: "幾何情報を用いた 3 次元物体検索のための特徴量抽出", 人文科学とコンピュータシンポジウム, 情報処理学会, pp.155-162 (2000)
- [36] M.Hiraga, Y.Shinagawa, T.Kohmura and T.Kunii: "Topology Matching for Fully Automatic Similarity Estimation of 3D Shapes", Proc. of the 28th Annual Conf. on Computer Graphics and Interactive Techniques, pp.203-212 (2001)
- [37] E.Paquet and M.Rioux: "Nefertiti:A Query by Content System for Three-dimensional Model and Image Databases Management", Image and Vision Computing, vol.17, pp.157-166 (1999)
- [38] D.V.Vranic, D.Saupe and J.Richter: "Tools for 3D-object Retrieval: Karhunen-Loeve Transform and Spherical Harmonics", Proc. of IEEE 2001 Workshop Multimedia Signal Processing, pp.293-298 (2001)
- [39] D.Saupe and D.V.Vranic: "3D Model Retrieval with Spherical Harmonics and Moments", Proc.of DAGM 2001, pp.392-397 (2001)
- [40] D.V.Vranic and D.Saupe: "3D Shape Descriptor Based on 3D Fourier Transform", Proc. of EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS 2001), pp.271-274 (2001)
- [41] T.Funkhouser, P.Min, M.Kazhdan, J.Chen, A.Halderman, D.Dobkin and D.Jacobs: "A Search Engine for 3D Models", ACM Trans. Graphics, vol.22, no.1,

- pp.83-105 (2003)
- [42] R.M.Haralick, K.Shanmugam and Its'hak Dinstein: "Textural Features for Image Classification", IEEE Transactions on Systems, Man and Cybernetics, vol.3, no.6, pp.610-621 (1973)
- [43] 辻内順平編著: "応用画像解析", 共立出版, pp.109-112 (1981)
- [44] 新井康平: "テクスチャ情報の季節変動を考慮した TM 画像分類", 日本写真測量学会誌, vol.26, no.4, pp.24-31 (1987)
- [45] 橋本岳, 松尾優: "合成開口レーダ画像のテクスチャ解析の一方法", 信学技報, IE86-88, pp.33-40 (1986)
- [46] 武者義則, 広池敦, 杉本晃宏: "類似画像検索における特徴空間の可視化インターフェース", 信学論(D-II), vol.J82-D-II, no.10, pp.1626-1633 (1999)
- [47] N.Otsu and T.Kurita: "A New Scheme for Practical, Flexible and Intelligent Vision Systems", Proc.of IAPR Workshop on Computer Vision, pp.431-435 (1988)
- [48] T.Kurita, N.Otsu and T.Sato: "A Face Recognition Method Using Higher Order Autocorrelation and Multivariate Analysis", Proc. of 11th International Conf. on Pattern Recognition, vol.2, pp.213-216 (1992)
- [49] T.Kurita and N.Otsu: "Texture Classification by Higher Order Local Autocorrelation Features", Proc. of Asian Conf. on Computer Vision, pp.175-178 (1993)
- [50] J.J Atick and A.N. Redlich: "What does the Retina Know about Natural Scenes?", Neural Computation, vol.4, no2, pp.196-210 (1992)
- [51] B.A Olshausen and D.J. Field: "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by v1? ", Vision Research, vol.37, pp.3311-3325 (1997)
- [52] R.P.N Rao and D.H Ballard: "Efficient Encoding of Natural Time Varying Images Produces Oriented Space-time Receptive Fields", Tech. Rep., 97.4, Dept of Corp Sci, Univ of Rochester (1997)
- [53] J.Daugman: "Entropy Reduction and Decorrelation in Visual Coding by Oriented Neural Receptive Fields", IEEE Trans. Biomed. Eng., vol.36, no.1, pp.107-114 (1989)
- [54] 高木幹雄, 下田陽久: "新編 画像解析ハンドブック", 東京大学出版会 (2004)
- [55] 中野宏毅, 吉田靖夫, 藤田和弘: "ガボール関数を用いたカラー液晶ディスプレイのマクロ欠陥検査支援", 信学論(D-II), vol.J80-D-II, no.3, pp.734-744 (1997)
- [56] 三谷重知, 呉海元, 塩山忠義: "ガボール変換による自動車抽出", 信学論

- (D-II), vol.J83-D-II, no.12, pp.2641-2651 (2000)
- [57] 田中秀郎, 吉田靖夫, 深見公彦, 中野宏毅: "ガボールフィルタの振幅および位相情報を用いたテクスチャ画像の領域分割", 信学論(D-II), vol.J84-D-II, no.12, pp.2565-2576 (2001)
- [58] 堀田一弘: "部分的な隠れに頑強な顔検出法", 信学論(D-II), vol.J88-D-II, no.8, pp.1624-1633 (2005)
- [59] 河田耕三, 有本卓: "ガボール展開係数を用いた階層的テクスチャ解析", 信学論(D-II), vol.J78-D-II, no.3, pp.437-444 (1995)
- [60] 山田敬嗣, 佐藤敦: "ニューラルネットによるパターン認識", 信学論(D-II), vol.J82-D-II, pp.852-859, pp.977-984, pp.1046-1053, pp.1248-1255, vol.J83-D-II, pp.50-56 (1999-2000)
- [61] 前田英作: "痛快!サポートベクトルマシン", 情報処理, vol.42, no.7, pp.676-683 (2001)
- [62] S. Watanabe and N. Pakvasa, "Subspace method in pattern recognition, " Proc. of 1st Int. J. Conf. on Pattern Recognition, Washington DC, pp. 2-32 (1973)
- [63] 飯島泰蔵: "パターン認識理論(基礎情報工学シリーズ)", 森北出版 (1989)
- [64] 末永高志, 佐藤新, 坂野鋭, "クラスタ構造に着目した特徴空間の可視化—クラスタ判別法—", 信学論(D-II), vol.J85-D-II, no.5, pp.785-795 (2002)
- [65] 伊藤慎也, 山本和彦, 本郷仁志, 加藤邦人, 丹羽義典: "寝たきりの人を支援するユビキタスインターフェース(システムの提案と基礎実験)", 第5回知能情報メディアシンポジウム, pp.111-116 (1999)
- [66] 島田大輔, 山本和彦, 加藤邦人, 川野卓也, 岩田英樹: "カメラと携帯電話を用いた介護支援システムの提案", 信学技報, PRMU2000-84, pp.67-72 (2000)
- [67] 岩淵耕一, 岩崎圭介, 竹村治雄, 岸野文郎: "画像処理を用いた実時間手振り推定とヒューマンインタフェースへの応用", 信学論(D-II), vol.J79-D-II, no.7, pp.1218-1229 (1996)
- [68] 西村拓一, 向井理朗, 野崎俊輔, 岡隆一: "動作者適応のためのオンライン教示可能なジェスチャ動画像のスポッティング認識システム", 信学論(D-II), vol.J81-D-II, no.8, pp.1822-1830 (1998)
- [69] 渡辺孝弘, 李七雨, 谷内田正彦: "インタラクティブシステム構築のための動画像からの実時間ジェスチャ認識手法—仮想指揮システムへの応用—", 信学論(D-II), vol.J80-D-II, no.6, pp.1571-1580 (1997)
- [70] 高橋勝彦, 関進, 小島浩, 岡隆一: "ジェスチャー動画像のスポッティング認識", 信学論(D-II), vol.J77-D-II, no.8, pp.1552-1561 (1994)
- [71] 大和淳司, 大谷淳, 石井健一郎: "隠れマルコフモデルを用いた動画像から

- の人物の行動認識", 信学論(D-II), vol.J76-D-II, no.12, pp.2556-2563 (1993)
- [72] 馬場朗, 芳澤伸一, 山田実一, 李晃伸, 鹿野清宏: "高齢者音響モデルによる大語彙連続音声認識", 信学論(D-II), vol.J85-D-II, no.3, pp.390-397 (2002)
- [73] B.J.Grosz, D.Appelt, P.Martin and F.Perira: TEAM: "An Experiment in the Design of Transportable Natural Language Interfaces", Artif. Intell., vol.32, no.2, pp.173-244 (1987)
- [74] 伊田政樹, 森弘之, 中村哲, 鹿野清宏: "据置き型情報提供端末向き雑音処理を用いた音声入力インタフェース", 信学論(D-II), vol.J84-D-II, no.6, pp.868-876 (2001)
- [75] 藤本雅清, 有木康雄: "GMMに基づく音声信号推定法と時間領域SVDに基づく音声強調法の併用による雑音下音声認識", 信学論(D-II), vol.J88-D-II, no.2, pp.250-265 (2005)
- [76] 井ノ上直己, 今井裕志, 橋本和夫, 米山正秀: "誤認識訂正のための繰返し音声検出手法", 信学論(D-II), vol.J84-D-II, no.9, pp.1950-1959 (2001)
- [77] 伊藤和幸, 数藤康雄, 伊福部達: "重度肢体不自由者向けの視線入力コミュニケーション装置", 信学論(D-I), vol.J83-D-I, no.5, pp.495-503 (2000)
- [78] Y.Matsumoto and A.Zelinsky: "An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement", Proc. of 4th FG, pp.499-504 (2000)
- [79] 伊藤和幸, 伊福部達: "ビデオキャプチャ画像処理による視線検出及び意思伝達装置への応用", 信学論(D-I), vol.J88-D-I, no.2, pp.527-535 (2005)
- [80] 瀬川能昭, 遠藤利生, 村上和人, 鳥生隆, 輿水大和: "似顔絵生成のための顔と瞳の輪郭抽出方法", 映像情報メディア学会誌, vol.51, no.11, pp.1935-1944 (1997)
- [81] 呉海元, 陳謙, 谷内田正彦: "ファジイパターン照合を用いた色彩画像からの顔検出システム", 信学論(D-II), vol.J80-D-II, no.7, pp.1774-1785 (1997)
- [82] 高橋圭子, 松浦正樹, 杉山岳弘, 阿部圭一: "人間による画像の色分類結果と領域分割結果に基づいた色空間の比較評価", 信学論(D-II), vol.J84-D-II, no.7, pp.1378-1388 (2001)
- [83] 秦清治, 広刈正義: "人間の感覚特性を組み込んだカラーパターン検査", 電気学会研究会資料, 情報処理・産業システム情報化合同研究会, IP-02-10, IIS-02-21, pp.55-58 (2002)
- [84] 島田直幸, 武藤淳一, 森島繁生: "Pan Tilt Zoom Cameraによる目及び唇形状抽出・追跡", 1999年信学総大, D-12-71, p.244 (1999)
- [85] N.Hagai, H.Hongo, K.Kato and K.Yamamoto: "Detection of Facial Parts Using Edge and Color Features", Proc. of the 4th International Conference on Virtual

- System and Multimedia (VSMM'98), pp.648-653 (1998)
- [86] 吉川左紀子, 益谷真, 中村真: "顔と心顔の心理学入門", サイエンス社 (1993)
- [87] 安本護, 林純一郎, 輿水大和, 丹羽義典, 山本和彦: "平均顔との距離を用いた性別・年齢推定手法の提案", 信学技報, PRMU2001-138, pp.1-6 (2001)
- [88] 和泉恭子, 伊賀亮達, 林尚典, 深野元太郎, 大谷哲也: "顔画像における複数の特徴量を用いた性別・年齢推定手法", 第9回画像センシングシンポジウム SSII2003, pp.567-570 (2003)
- [89] Z.Sun, X.Yuan, G.Bebis and S.Louis: "Neural-network-based Gender Classification Using Genetic Eigen-feature Selection", IEEE International Joint Conference on Neural Networks, 2002 (<http://www.cs.unr.edu/~bebis/genderIJCNN.pdf>)
- [90] 細井聖, 瀧川えりな, 川出雅人: "ガボールウェーブレット変換とサポートベクタマシンによる性別・年代推定システム", 第8回画像センシングシンポジウム SSII2002, pp.243-246 (2002)
- [91] 西松豊典, 安井聡, 鳥羽栄治, 村上和人, 輿水大和: "ヒューマンインタフェースへの似顔絵画像の応用 - 似顔絵における性別・年齢の推定 -", 信学論(A), vol.J80-A, no.8, pp.1260-1265 (1997)
- [92] 本郷仁志, 安本護, 渡辺博己, ジャン・クリストフ・テリヨン, 山本和彦: "パーセプトルーム構築のための多方向顔画像データベース開発", 信学技報, PRMU2000-108, pp.7-12 (2000)

論文

- [1] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志: "ポーズ認識に夜家電製品制御支援システム", 電学論(C), vol.124, no.3, pp.769-775 (2004)
- [2] 川野卓也, 加藤邦人, 山本和彦: "各顔部品における四方向面特徴を用いた性別分類", 日本顔学会論文誌四巻一号, pp.81-87 (2004)

海外発表

- [1] Kazuhiko Yamamoto, Takuya Kawano and Kunihito Kato: "Pose Recognition System for a Ubiquitous Interface", Proc. of International Conference on Image and Signal Processing (ICISP2001), pp.155-160 (Morocco) (2001)
- [2] Takuya Kawano, Kazuhiko Yamamoto, Kunihito Kato and Hitoshi Hongo: "Integration of Pose Recognition for a Person Wearing Short or Long Sleeves", Proc. of the International Conference on Pattern Recognition (ICPR2002), vol.3, pp.1023-1026 (Canada) (2002)
- [3] Takuya Kawano, Kazuhiko Yamamoto and Kunihito Kato: "A Comparison of the Gender Differentiation Capability between Facial Parts", Proc. of the International Conference on Pattern Recognition (ICPR2004), vol.3, pp. 350-353 (England) (2002)
- [4] Takuya Kawano, Kunihito Kato and Kazuhiko Yamamoto: " An Analysis of the Gender and Age Differentiation Using Facial Parts", Proc. of International Conference on Systems, Man and Cybernetics (SMC2005), vol.3, pp. 3432-3436 (USA) (2005)

国内発表

- [1] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志, 丹羽義典: "TV 制御支援システムの改良と評価実験", 平成 12 年度電気関係学会東海支部連合大会, 748, p.374 (2000)
- [2] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志, 丹羽義典: "TV 制御支援システムのための顔と手の相対情報を利用したポーズ認識", 2001 年信学総大, D12-41, p.208 (2001)
- [3] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志, 丹羽義典: "服装を考慮したポーズ認識による家電制御支援システムの研究", 情処学 CVIM 研報, 127-22, pp.151-156 (2001)
- [4] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志, 丹羽義典: "高次局所自己相関特徴の相対位置関係を用いた TV コントロールのためのポーズ認識", 第 7 回画像センシングシンポジウム SSII200 1, pp.341-346 (2001)
- [5] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志, 丹羽義典: "ポーズ認識を利用した介護支援のための家電制御支援システムの改良", 信学技報, PRMU2001-114, pp.9-14 (2001)
- [6] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志, 丹羽義典: "多段階にポーズ認識を用いた複数の家電製品の制御システム", 平成 13 年度電気関係学会東海支部連合大会, 765, p.383 (2001)
- [7] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志: "家電製品制御システムにおけるポーズ認識辞書学習法の比較", 電気学会研究会資料, 情報処理・産業システム情報化合同研究会, IP-02-8, IIS-02-19, pp.43-48 (2002)
- [8] 川野卓也, 山本和彦, 加藤邦人, 本郷仁志: "家電制御システムで用いる辞書学習法の評価について", 平成 14 年度電気関係学会東海支部連合大会, 666, p.333 (2002)
- [9] 川野卓也, 加藤邦人, 山本和彦: "顔部品間における性別推定能力性能の評価", 信学技報, PRMU2003-141, pp. 7-12 (2003)
- [10] 川野卓也, 加藤邦人, 山本和彦: "四方向面特徴を用いた顔部品による性別推定性能の考察", 画像の認識・理解シンポジウム(MIRU2004)講演論文集 I, pp.648-653 (2004)