

パターンの持つ不変的な性質を利用した
物体認識に関する研究

2022年 3月

山口 修

パターンの持つ不変的な性質を利用した
物体認識に関する研究

山口 修

理工情報生命学術院
システム情報工学研究群
筑波大学

2022年 3月

目次

1	はじめに	6
1.1	研究の背景と目的	6
1.2	本論文の構成	11
2	定性的3値表現に基づく画像マッチング	13
2.1	画像のマッチング	13
2.2	定性的3値表現	14
2.2.1	定義	14
2.2.2	定性的3値表現の類似度	16
2.3	定性的3値表現の性質	19
2.3.1	輝度変換に対する性質	19
2.3.2	各種変換、ノイズに対する性能	19
2.3.3	3値画像を用いた従来例との比較	25
2.3.4	増分符号相関との比較	25
2.4	実験	26
2.4.1	テンプレートマッチング	27
2.4.2	トラッキング	30
2.5	ハイブリッドマッチング	30
2.6	深層学習との組みあわせの検討	32
2.6.1	定性的3値表現のカラー化	33
2.6.2	定性的3値表現を用いたCNNの導入	34
2.6.3	QTR画像を入力としたResNetによる実験	35
2.6.4	バイナリニューラルネットワークを用いた高速化	38
2.7	本章のまとめ	39
3	パターンハッシングによる物体認識	40
3.1	アプローチ	40
3.2	パターンハッシング (Pattern Hashing)	42
3.2.1	分散アピアランスモデル	42
3.2.2	モデルの登録	43
3.2.3	ハッシュ関数とハッシングにおける衝突処理	45
3.2.4	部分情報と全体情報の関係記述	45
3.3	パターンハッシングによる認識	47

3.3.1	特徴点検出とモデル選択	47
3.3.2	マッチングのための類似性測度	47
3.3.3	仮説の生成と投票による統合	49
3.3.4	計算量とメモリ量	51
3.4	アルゴリズムの検証実験	51
3.4.1	顔検出	51
3.4.2	複数種類の物体検出	53
3.5	本章のまとめ	54
4	特徴抽出の多重化による線形部分空間法の高精度化	55
4.1	線形相互部分空間法の拡張	55
4.2	顔画像認識	58
4.3	制約部分空間の学習	59
4.4	複数の制約部分空間の生成に対するアンサンブル学習	61
4.4.1	並列的な学習	61
4.4.2	逐次的な学習	62
4.5	多重制約相互部分空間法による識別方法	65
4.5.1	顔画像認識の流れ	65
4.5.2	制約部分空間への射影	65
4.5.3	部分空間同士の類似度算出	65
4.5.4	類似度の結合	67
4.6	多重制約相互部分空間法の実験	67
4.6.1	照明変動に対する識別性能	69
4.6.2	多人数データベースにおける識別性能	77
4.7	白色化変換の導入による拡張	79
4.7.1	白色化相互部分空間法	80
4.7.2	解決すべき課題	81
4.7.3	解決手段の検討	81
4.7.4	固有スペクトル解析による観察	83
4.7.5	疑似白色化変換の導入	85
4.7.6	多重疑似白色化相互部分空間法	87
4.8	多重疑似白色化相互部分空間法による実験	88
4.8.1	多視点物体識別実験	88
4.8.2	顔認識実験	91

4.9	深層学習特徴量との組み合わせに関する検討	92
4.9.1	ハンドジェスチャ認識 (IPN Hand)	92
4.9.2	ビデオベースの顔認識 (YouTube Face dataset)	93
4.10	本章のまとめ	95
5	顔認識への応用例	97
5.1	ポータブルPCにおける顔認識システム	97
5.1.1	PCにおける簡易セキュリティ	97
5.1.2	顔向きや表情の変化に頑健な顔認識アルゴリズム	98
5.1.3	システムの構成	103
5.1.4	Smartface のアプリケーション機能	107
5.1.5	本節のまとめ	112
5.2	公開ベンチマークデータによる顔認識実験	113
5.2.1	FERET DB	113
5.2.2	各手法による認識結果	113
5.3	本章のまとめ	116
6	結論	117

要旨

画像パターン認識においては、画像取得時における撮影条件の変化や撮影方向の違いなどによる見えの変化、識別対象自体の変形など、さまざまな変動により多様に変化する対象を、それらの影響を受けないで、識別することが必要となる。本論文では、パターン認識における不変的な性質に着目し、多様に変化する対象を識別する高度なアルゴリズムを開発するために、次のような三つの視点に着目して、段階的に問題を解決する。

第一の視点は、パターンの存在する環境、発生する過程で起こりうる変化に対する不変性に着目したアルゴリズムの構成を行うことである。第二の視点は、パターンの取得時の変化に対応した不変的な特徴を見出してアルゴリズムに反映させることである。第三の視点は、複雑なパターンの変化を取り扱ってきた伝統的なアルゴリズムを、さらに進化させることである。

この考え方に基づいて、(1) 局所的な輝度変化に対する不変量、(2) 位置や大きさなどの変化に対応する幾何学的な変換に対する不変量の利用法、(3) パターン自体の変形も含む複合的な変化に対応するため不変的な特徴抽出法について考察した。

具体的には、第一の視点に立って、画像マッチングにおける不変量に関する検討を行う。画像マッチングのための測度として、隣接画素の同値関係、大小関係に基づいた定性的3値表現による類似度について述べる。これは、部分的な遮蔽やいくつかの輝度変換に対して耐性をもつ表現となる。その表現を用いたマッチング手法の性質、性能について検証を行う。マッチングに要求される高速性を担保し、不変量としての性質を満たす手法を従来法と比較し、また従来法と組み合わせて利用することで、従来よりも高速かつ柔軟な照合を実現する。

第二の視点からは、対象の位置や大きさなどの変換に対する不変性を導入したアルゴリズムについて検討する。本章では、幾何学的特徴ベースの手法にアピランスペースの考え方を導入した物体認識の枠組みを提案する。部分画像のマッチングを行う際に、各部分画像を作成した際に、付随する不変量を用いて部分画像モデルをハッシングにより高速に選択し、効率的にマッチングを行う”パターンハッシング”について述べる。これは従来法である Geometric Hashing が点や線といった幾何学的特徴をベースとしていたことに対して、濃淡パターンをプリミティブとして利用できるようにしたという観点でも実用上有用な方法である。

第三の視点からは、複合的に起こる変動の中から識別に重要な不変的な特徴を取り出すための特徴抽出法について検討する。これまで顔認識等に応用していた相互部分空間法の高度化を目指し、制約部分空間の生成に対してアンサンブル学習を適用し、複数の制約部分空間を用いて特徴抽出を多重化した多重制約相互部分空間法を提案する。さら

に相互部分空間法の高性能化に関して、線形の特徴抽出系をもつ白色化相互部分空間法について説明し、さらにその改良法を示す。具体的には、人物同士のデータの相関を少なくする変換である白色化変換に対して、個人の識別に有用な成分の選択を行うような疑似白色化変換を導入し、カテゴリの識別能力の改善を図る。

最後に、三つの視点に立って提案した各手法を用いて、多様な変化を行う具体的対象として「顔画像認識」に適用した。提案した画像マッチング法によって、低速な計算資源においても高速な顔検出を実現し、ポータブルPCにおいて相互部分空間法を動作させるための工夫ならびにエンターテインメント向けの機能を提案した。さらに公開ベンチマークデータを用いて、提案手法によって顔画像の検出、顔部品検出といったセグメンテーションを行い、正規化顔画像パターンを対象として、新たに拡張した特徴抽出法によるアルゴリズムの認識実験を行い、従来法に比べて精度向上することを確認した。

本論文では、画像認識の各プロセスにおいて、不変的な性質に着目することで、高速、高精度なアルゴリズムを見出し、実用性を重視した物体認識法を実現できたと考える。

1 はじめに

1.1 研究の背景と目的

社会のありとあらゆる場所に「カメラ」が急速に増加している。街中の商店街や店舗内の「防犯カメラ」、手軽に写真が撮影可能な「デジタルカメラ」、テレビ通話にも利用可能な携帯電話、スマートフォンの「小型内蔵カメラ」、インターホンや家の周りやペットを見守るための「監視用カメラ」など、いまや家庭の中でさえ複数台のカメラが存在している。機器のデジタル化によって、長時間撮影、録画が可能になり、カメラだけでなくその記録装置も普及が進んでいる。しかしながら、撮影されたコンテンツが人間によって確認、視聴されているのは、わずか数%であると考えられる。デジタル画像の普及に伴い、その中身を自動的に内容を把握するための技術である「物体認識技術」の需要は今後さらに伸びると考えられる。

画像データは、デジタル化、テレビ映像の高解像度化と撮像素子の高性能化に伴い、4K、8Kと徐々に高解像度化し、ネットワークを縦横無尽に移動し、巨大なネットワーククラウドに蓄えられる。これまでの「言語」によるシンボル化された情報に基づいた情報処理から、「画像」のような多様な情報を処理する機能を社会に提供することにより、新たな価値が想像され、より豊かな社会の実現が可能となる。そのためには、大量の画像を同時に素早く処理する必要があり、高速かつ高精度な実用性の高いパターン認識技術が重要となる。

人間に備わった機能の一つであるパターン認識は、ものごとの特徴（パターン）を抽出し、さまざまな意思決定、判断のために情報の分類を行う。本論文では、パターン認識における不変的な性質に着目した研究に関して議論する。

画像パターン認識においては、画像取得時における撮影条件の変化や撮影方向の違いなどによる見えの変化、識別対象自体の変形など、さまざまな変動により対象が多様に変化する。認識系は、それらの変動の影響を受けないで、正しく対象の位置を求めたり、カテゴリの識別を行うことが必要となる。

多様に変化するデータの中から不変的な特徴を取り出すことができれば、工学的な視点から有用なアルゴリズムが構成できると考える。しかしながら、不変的な特徴を取り出すには、以下のいくつかの要因の存在が問題となる。

- 対象自体の不安定性
 識別対象となる物体自体の変動が存在する。データの収集環境（光の向き、色、強さ）や、非剛体であれば、物体自体の変形が起こってしまう（図1）。
- カテゴリ化の不完全性
 カテゴリは物理的な測量に基づいた情報からすべてが決定されるものではなく、人間の精神的活動によって定められる。人間が客観的または主観的に決めにくいというカテゴリも存在することや、文化的な背景に依存して変化を伴うものもあり、常にそのカテゴリ境界には変化が起こると考えられる（図2）。
- 観測の不確実性
 同じ対象をセンシングするとしてもそのセンサの性質、性能によって取り入れられる情報が多様である。また、センサの取得過程における計測誤差なども許容しなければならない（図3）。



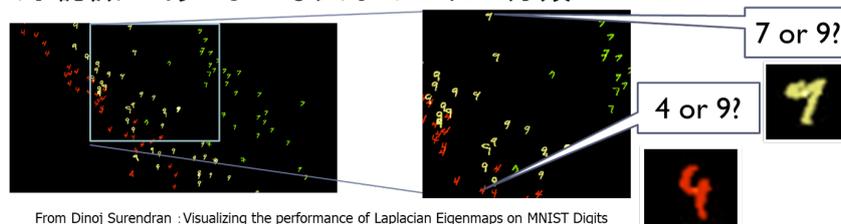
図 1: 対象の変動例（顔の場合）

表情認識:「笑顔」の境界はどこ?



Technique by M. Yuasa, O. Yamaguchi: Real-time face blending by automatic facial feature point detection. IEEE International Conference on Face and Gesture Recognition 2008

文字認識: あいまいな文字はどれに分類?



From Dinoj Surendran : Visualizing the performance of Laplacian Eigenmaps on MNIST Digits

図 2: カテゴリが決めにくい例



図 3: センサによる取得過程で起こる課題

これらは、パターン認識において、あらゆる変化に対応できる不変的な特徴を見つけることが困難であることを示唆している。しかし、工学的なパターン認識では「特徴計測」「特徴抽出」「識別」などのそれぞれのプロセスがあり、それぞれの処理の中に内在する不変性や、変形（変換）を制限することによって不変性を満たすものが存在している。例えば、2次元の画像モーメントを用いた、並進、回転、相似、鏡映変換についての Hu の 7つの不変量 [1] や射影変換に対する複比などの変換を規定した場合の不変量、また自己相関関数などによって導入される時間不変性などが存在する。

パターン認識における不変性は、対象が存在する環境に対して、許容される変換が定まることによって、不変的な特徴が定まる。そこで、それぞれの処理の段階において、うまく不変性を見つけ出して利用することによって、高速かつ高精度な物体認識法を可能としたい。

本論文では、多様に変化する対象を識別する高度なアルゴリズムを開発するために、次のような三つの視点に着目して、段階的に問題を解決する。

第一の視点は、パターンの存在する環境、発生する過程で起こりうる変化に対する不変性に着目したアルゴリズムの構成を行うこと。

第二の視点は、パターンの取得時の変化に対応した不変的な特徴を見出してアルゴリズムに反映させること。

第三の視点は、対象自身の変形なども含む複雑なパターンの変化を取り扱ってきたアルゴリズムを進化させること。



図 4: 各種変化のクラス分けと関連する変換

図 4 は、本論文で対象とするさまざまな変換について、クラス分けを行った図である。光の方向や強さに関連した輝度変換やヒストグラム平坦化といった「フォトメトリックな変換」、次に、平行移動、サイズや回転方向などを変化させる変換を対象とする「幾何学的変換」、そして、顔の向きなどの変化を扱う 3 次元的な変換や部分的な形状変化を伴うような「変形」に区分する。多様な変化に対しては、これらの変換を組み合わせることで近似的に表現する。例えば、ある不変性に着目することで、規定した変換に鈍感な特徴を得ることで、所望の変化にロバストな処理の実現であったり、情報を削減することによって、高速な処理が実現できると考えられる。

このような考え方に基づいて、以下の順序で従来の認識処理を見直していくこととする。図 5 に、本論文における各視点がカバーする範囲の概要図を示す。図のように、横軸にタスクと縦軸に先に説明した各種変換のクラスを割り当てている。タスクに依存して、扱う変化（変換）を規定することとし、あらゆる変化（変換）を取り扱えるわけではなく、変換に対する不変性を扱える枠組みの範囲で、各処理の改善の検討を行っていく。

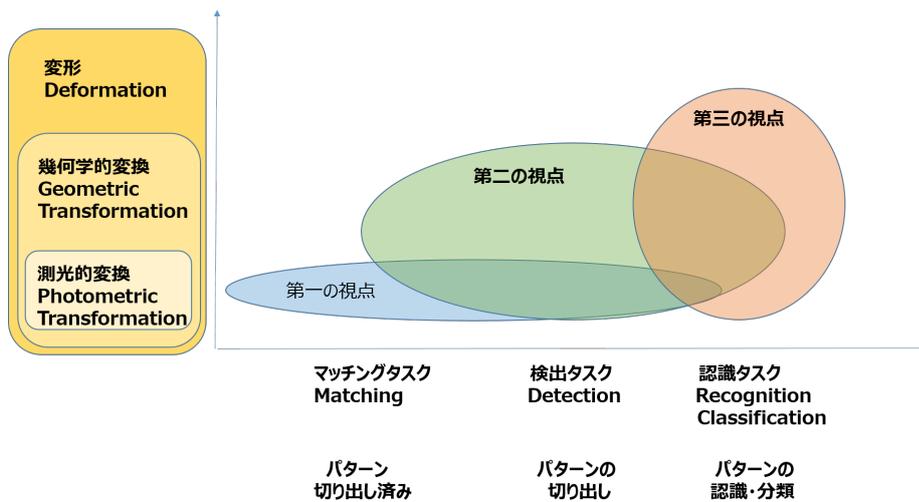


図 5: 本論文における各視点のカバー範囲の概要図

まず、第一の視点では、コンピュータビジョンの中で基本的な技術である画像マッチングを問題の対象とする。識別対象を構成する局所的な領域について、マッチングの精度に関しては、明るさの変化や輝度変換などに影響を受けないことが一つの必要条件となる。そこで、マッチングに要求される高速性を考慮したうえで、変動への対応を行う手法について考察する。さらに、色情報を用いた拡張や、近年の深層学習との組み合わせを検討することで、物体のカテゴリ識別の性能向上方法についても検討を行う。

次に、第二の視点では、複数の画像マッチングに基づいて、物体の検出や物体の構成部品の認識等を行う問題を取り扱う。本研究の対象とする物体認識は、コンピュータビジョンの中心的な問題の一つである。多くの物体認識方法は、幾何学的特徴点ベースの研究 [2] とアピランスベースの研究 [3][4] の2つのタイプに分類される。アピランスベースの研究の火付け役となったのは、Turk and Pentland の顔認識の研究であった [5]。実用的な側面からも顔を計算機で扱おうという取り組みは世界中に広がりを見せた。一方、幾何学的特徴点ベースの研究も、一般物体認識 [6] として研究が盛んである。幾何学的特徴点の周辺の情報を用いて物体認識を行うために、関連して様々な特徴量、記述子の開発 [7]、高速なマッチングに関する研究などが盛んに行われている。この2つの分類されたアプローチを組み合わせることによって、これまでに提案された手法とは異なるスキームを提案することを目指す。また、物体の現れる位置、大きさなどの変化を許容するために、幾何学的変換に対する不変性をモデルに導入し、そのモデルを用いて部分隠れやノイズに対してロバストな認識方法を実現する。

最後の第三の視点では、パターン識別に関するもので、従来から顔認識等に應用されている相互部分空間法 [8][9] の拡張に関する問題に取り組む。顔の認識では、撮影環境の

変動に加え、光学環境の変化、顔の向きの変化、表情の変化など、多種多様な変化が起こり、それらの変化が複合的な変化としてデータに現れる。その変化を許容するための一つの大きな進化として、相互部分空間法を行う前に特徴抽出を行う制約相互部分空間法 [10] が開発された。各個人の識別を行うために、よりよい特徴量を取り出すというアプローチは、さまざまな変動に対する不変性の導入を追究することに対応すると考えられる。本研究では、その特徴抽出法をさらに進展させることを目指す。具体的には、従来の制約空間への射影というプロセスを多重に組み合わせて多角的に特徴抽出を行う手法や、射影変換自体をカテゴリの無相関化を行う変換に替え、それに関連した工夫を加えることを試みる。

本研究の具体的なデータの対象の一つとした「顔」は、様々な捉え方がなされる対象である。一つは、個人を識別するための ID (識別子) としての役割、また、表情という各個人の感情的な内部状態を表すための役割、そして、顔の向き、視線といった情報から、個人の意図をシグナルとして発信する役割がある。これは、人間の精神活動である「知」、「情」、「意」のそれぞれに対して情報を発信できる役割を一つの部位で司る特殊な人間の持ち物である。顔を計算機で扱おうという取り組みのスタートは、約 50 年前に遡る [11]。その後、Turk らによる研究 [5] が発表された頃、顔認識は企業の一研究として応用が期待されることとなった。著者の研究グループにおいても、セキュリティ、エンターテインメントへの応用を目指した「顔」を対象とした研究が開始された。さらに、計算機の性能向上に連れて技術開発が進み、スマートフォンには、フレーム内の顔の位置、数、大きさを自動的に感知し、その人物の配置によって露出時間、フォーカス等を制御する機構が一般的となっている。また、個人照合を行うことで、セキュリティ確保にも応用されている。この多様に变化する対象を、小型のパーソナルコンピュータで個人の識別を行い、アプリケーションに応用するための知見についてもまとめる。実フィールドで動作するアプリケーションで志向した方法や、公開ベンチマークテスト等への適用など、具体的な応用事例について述べる。

1.2 本論文の構成

前節で述べた三つの視点に対応して、本論文は以下のように構成する。

2 章では、第一の視点に立って、画像マッチングにおける不変量に関する検討を行う。画像マッチングのための測度として、隣接画素の同値関係、大小関係に基づいた定性的 3 値表現による類似度について述べる。これは、部分的な遮蔽やいくつかの輝度変換に対して耐性をもつ表現となる。その表現を用いたマッチング手法の性質、性能について検証を行う。マッチングに要求される高速性を担保し、不変量としての性質を満たす手法

を従来法と比較し、また従来法と組み合わせて利用することで、従来よりも高速かつ柔軟な照合を実現する。また、この不変量を深層学習手法と組み合わせることにより、画像の識別タスクへの適応可能性についても検討する。

3章では、第二の視点から、対象の位置や大きさなどの変換に対する不変性を導入したアルゴリズムについて検討する。本章では、幾何学的特徴ベースの手法にアピアランスベースの考え方を導入した物体認識の枠組みを提案する。部分画像のマッチングを行う際に、各部分画像を作成した際に、付随する不変量を用いて部分画像モデルをハッシングにより高速に選択し、効率的にマッチングを行う”パターンハッシング”について述べる。これは従来法である Geometric Hashing[2] が点や線といった幾何学的特徴をベースとしていたことに対して、濃淡パターンをプリミティブとして利用できるようにしたという観点でも実用上有用な方法である。

4章では、第三の視点から、複合的に起こる変動の中から識別に重要な不変的な特徴を取り出すための特徴抽出法について検討する。これまで顔認識等に應用していた相互部分空間法の高度化を目指し、制約部分空間の生成に対してアンサンブル学習を適用し、複数の制約部分空間を用いて特徴抽出を多重化した多重制約相互部分空間法を提案する。さらに相互部分空間法の高性能化に関して、線形の特徴抽出系をもつ白色化相互部分空間法を説明し、さらにその改良法を示す。具体的には、データの相関を少なくする変換である白色化変換に対して、識別に有用な成分の選択を行うような疑似白色化変換を導入し、カテゴリの識別能力の改善を図る。さらに、近年の深層学習特徴量と部分空間ベースの特徴抽出法との組み合わせにより、さらなる高精度化についても検討する。

5章では、顔認識を実例として、2～4章で提案したアルゴリズムを用いて実現したシステム事例について述べる、また、2、3章の方法を用いて、顔の検出、顔の部品検出を行い、顔特徴量パターンを作成し、4章の識別アルゴリズムによって、公開データを用いたベンチマーク評価について述べる。

6章は、本論文のまとめとする。

2 定性的3値表現に基づく画像マッチング

本章では、第1章で述べた第一の視点に立って、画像マッチングにおける不変量に関して検討する。部分的な遮蔽やいくつかのフォトメトリックな変換に対して耐性をもつ表現を目指し、高速かつ柔軟な照合法の実現を目指す。

2.1 画像のマッチング

画像間の類似性の測度を定義し、画像中でその測度が最も高い場所を決定したり、2枚の画像の類似度を判定する方法である画像のマッチングは、物体認識、画像の位置あわせなど、さまざまな場面で用いられる [12]。

あるモデル画像(テンプレート)の画像中の位置を検出するテンプレートマッチングでは、一般に平行移動、拡大、縮小といった幾何学的変換のパラメータの増加に対して探索のコストが大きくなる。よって、実用上はテンプレートとの類似度計算にかかる計算量が少ない方法が求められる。また、明るさの変化を制御するために、ヒストグラム平坦化処理やエッジ強調フィルタ、エッジ抽出処理といった画像処理(前処理)が必要となる場合が多い [4]。しかしながら、これらの前処理は使用する類似性測度やモデル画像のもつ特徴との相性の問題もある。

画素ごとの濃淡情報を用いたマッチング法では、一般に類似性測度として正規化相関(CC: Correlation Coefficient)や残差の2乗和(SSD: Sum of Square Difference)といったものがよく用いられる。濃淡値を直接用いた類似度の定義は、濃淡値の大きさがそのまま類似度の大きさに影響することが特徴である。これらを定量的な類似度と呼ぶこととする。

一方、このような性質をもたない画像のマッチングの研究として、Venotら [13]は、2つの画像を重ねた場合に、それぞれの画素の差の符号が変化した数を特徴量として、マッチングを行っている。差の符号の変化は、差のある位置の「個数」に関連するために、濃淡値の変動に対しても影響が少ない。また、最近では、ロバスト統計の立場から、隣接画素の増分傾向をとらえた増分符号相関が提案されている [14]。

画像検索において、類似画像の検索の索引としてLipsonら [15]は、画像をいくつかのブロックに分け、そのブロック間の画像特徴量の大小関係、同値関係に着目した構造化を行っている。この構造化のベースとなるSinhaら [16]は、画像中からの顔領域検出のために、部分領域の大小関係について、類似している関係を抽出し、その構造をテンプレートとしている。これは、大量の学習データから定性的な関係を抽出してテンプレートとして持つというアイデアである。これらは画像の濃淡値の情報を直接利用するのではなく、相対的な関係や順序といった定性的な側面を捉えている。

一般に、定性的に判断を行うことで、詳細にとらわれずに全体像の特徴を理解することができる。さらに詳細な分析が必要な場合は、その箇所についてのみ定量的に評価を行えば効率がよい。計算を効率良く行い、ノイズが存在する画像に対して変動に強いマッチングを行うには、部分的な変動に対して強いといった特徴量や「おおまかに似ている」といった指標となりうる特徴を持つ定性的な性質を反映した特徴量を利用することが効果的である。

そこで本章では、「隣接画素間の濃淡値の大小関係、同値関係」といった画素間の関係記述を定性的な情報としてとらえた定性的3値表現を提案する。また、それを画像特徴量に利用したマッチング法について述べる。この表現では、部分的な変動やいくつかの濃淡変換に対しては不変となり、定性的な情報のみを用いるため、計算コストを少なくした高速なマッチングが実現できる。

以下、提案する定性的3値表現(QTR)、それをを用いた類似度に定義する。次に、その表現、類似度の性質について述べる。従来法である正規化相関との比較や、コンセプトの類似した、3値輪郭表現、増分符号相関との違いについて述べる。予備実験として、実画像を使ったトラッキング実験、テンプレート検出実験も行い、ならびに定量的、定性的なマッチングを組み合わせたハイブリッドマッチングへの展開について述べる。

2.2 定性的3値表現

提案法を図6を用いて説明する。説明のため、濃度方向の分解能が4段階の画像(4×4画素)を考える。図中央のように左隅の水平方向の2つの画素に注目し、その大小関係、同値関係を調べる。同様に水平方向の2つの画素の大小関係、同値関係を調べると、3×4画素の3値画像が構成できる。同じように垂直方向の関係についても、大小関係、同値関係を調べると4×3画素の3値画像が構成できる。

対象とする画像 I に対し、これらの2つの3値画像 Q_v, Q_h を求める。この表現法を、定性的3値表現(QTR: Qualitative Trinary Representation)と呼ぶ。

2.2.1 定義

画像 I (大きさ $m \times n$ 画素) に対して、位置 (x, y) の画素を画素値を $I(x, y)$ とする。

定性的3値表現は、ある画像 $I_s (0 \leq x \leq m-1, 0 \leq y \leq n-1)$ に対して、次の2つの3値画像 Q_h, Q_v で定義する。

$$QT(I) = \{Q_v, Q_h\} \quad (1)$$

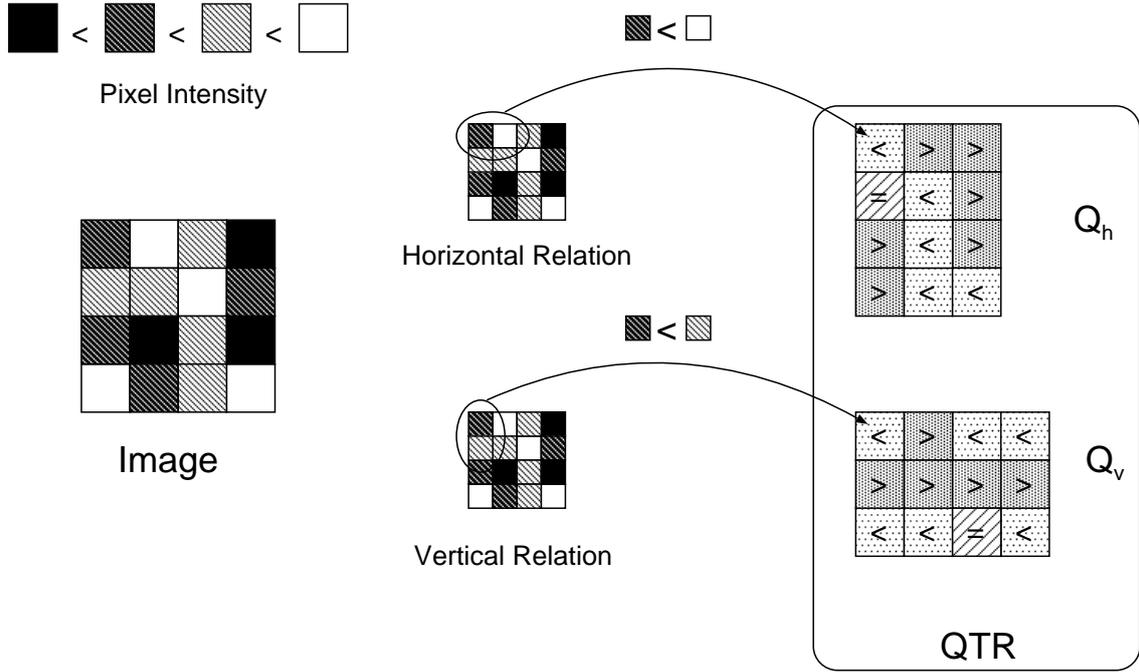


図 6: 定性的 3 値表現

$$\begin{cases} Q_v(x, y) = \text{sgn}(I_s(x, y) - I_s(x + 1, y)) \\ \quad (0 \leq x \leq M - 2, 0 \leq y \leq N - 1) \\ Q_h(x, y) = \text{sgn}(I_s(x, y) - I_s(x, y + 1)) \\ \quad (0 \leq x \leq M - 1, 0 \leq y \leq N - 2) \end{cases} \quad (2)$$

$I_s(x, y)$ は画像の位置 (x, y) での輝度値、 sgn 関数は、

$$\text{sgn}(x) = \begin{cases} 1 : x > 0 \\ 0 : x = 0 \\ -1 : x < 0 \end{cases}$$

である。

このとき、 Q_v は $(m - 1) \times n$ 画素の画像、 Q_h は、 $m \times (n - 1)$ 画素の 3 値画像となる。

ここで、隣接画素が同値と判定する基準を緩和した以下の定義も導入できる。

$$QT^K(I_s) = \{Q_v^K, Q_h^K\}$$

$$Q_v^K(x, y) = \begin{cases} 1 : I_s(x, y) - I_s(x + 1, y) > K \\ 0 : -K \leq I_s(x, y) - I_s(x + 1, y) \leq K \\ -1 : I_s(x, y) - I_s(x + 1, y) < -K \end{cases}$$

$$Q_h^K(x, y) = \begin{cases} 1 : I_s(x, y) - I_s(x, y + 1) > K \\ 0 : -K \leq I_s(x, y) - I_s(x, y + 1) \leq K \\ -1 : I_s(x, y) - I_s(x, y + 1) < -K \end{cases}$$

この定義によれば、 $K = 0$ の場合、

$$QT(I) = \{Q_h, Q_v\} = \{Q_h^0, Q_v^0\} = QT^0(I)$$

となる。

2.2.2 定性的3値表現の類似度

次に、2つのQT表現の間の類似度について定義する。

いま、 D_v, D_h を以下のように定義する。

$$QT(I_1) = \{Q_{v_1}, Q_{h_1}\}$$

$$QT(I_2) = \{Q_{v_2}, Q_{h_2}\}$$

$$D_v(I_1, I_2) = \sum_k \sum_l \begin{pmatrix} 1 : Q_{v_1}(k, l) = Q_{v_2}(k, l) \\ 0 : Q_{v_1}(k, l) \neq Q_{v_2}(k, l) \end{pmatrix}$$

$$D_h(I_1, I_2) = \sum_k \sum_l \begin{pmatrix} 1 : Q_{h_1}(k, l) = Q_{h_2}(k, l) \\ 0 : Q_{h_1}(k, l) \neq Q_{h_2}(k, l) \end{pmatrix}$$

すなわち、それぞれの3値画像の対応する各画素が同じ符号の場合に1を、そうでない場合を0として、すべての画素についての和を取る。よって、最大値は同じ画像を比較したときであるから、

$$D_v(I, I) = n(m - 1)$$

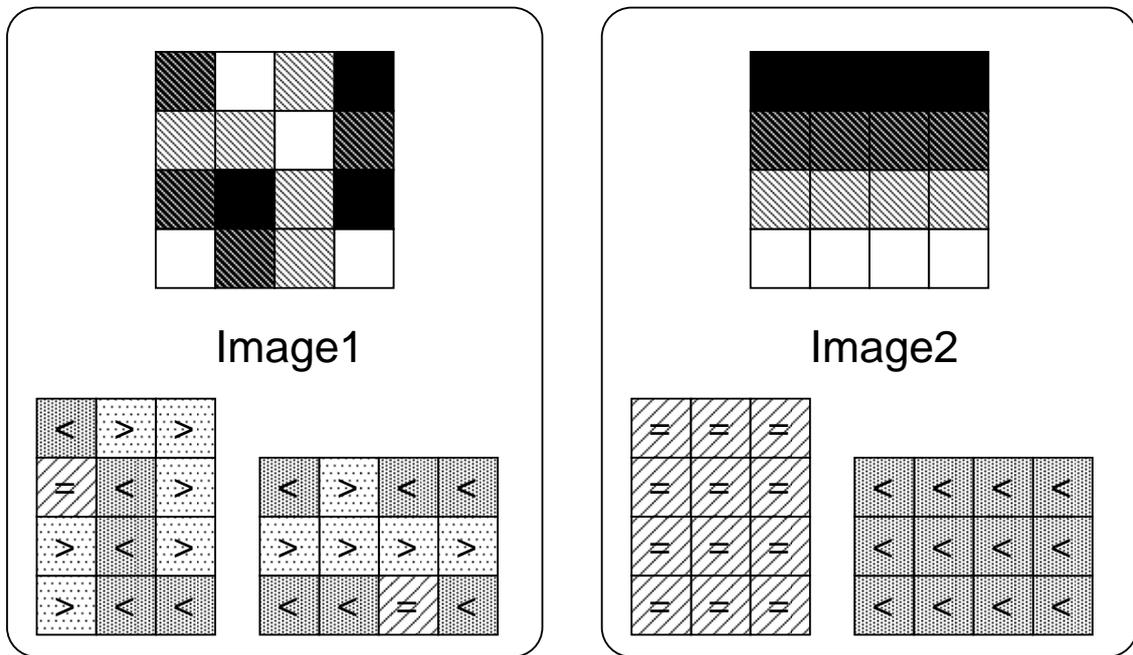
$$D_h(I, I) = m(n - 1)$$

となる。

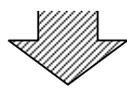
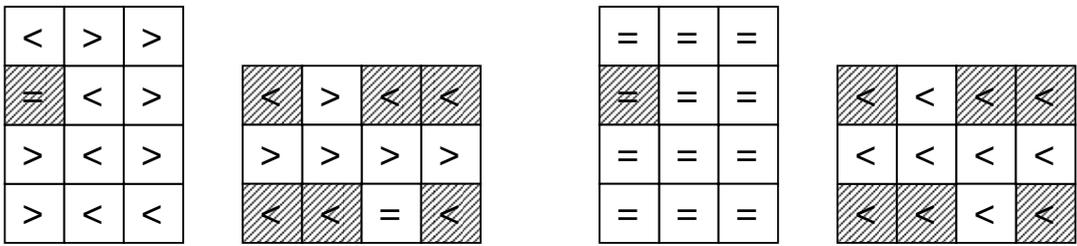
図7のように、2つの画像を比較する場合、そのQT表現を用い、符号が一致した場所の数を利用して類似性を判定する。

本稿では2つの画像 I_1, I_2 の類似度 $QTS(0.0 \sim 1.0)$ として

$$QTS(I_1, I_2) = \frac{D_v(I_1, I_2) + D_h(I_1, I_2)}{m(n - 1) + n(m - 1)} \quad (3)$$



compare



count correspondences
(calculate similarity)

図 7: 定性的 3 値表現の類似度

を用いることにする。この場合、図7の2枚の画像間の類似度は、 $7/24=0.292$ となる。もちろん、テンプレート画像や比較する画像すべてが同じ大きさである場合には、分母の部分は同じであるため、分子部分の大きさのみで比較も可能である。

この類似度の利点としては、各画素に対しての計算に乗算を用いなくて、比較演算、加減演算のみで類似度が求まる点である。また、統計的な平均、分散などを用いていないので、テンプレート内の走査は一回で済む。これは計算コストを考慮する上で有用である。

2.3 定性的3値表現の性質

2.3.1 輝度変換に対する性質

輝度変換に対してのQT表現の性質としては、各画素値 $I(x, y)$ について $I(x, y) \rightarrow aI(x, y) + b$ と変換された画像 $aI + b$ について、

$$QT(I) = QT(aI + b)$$

なる関係が成り立つ ($a > 0$)。これは、各画素間の大小関係が変化しないために同様のQT表現を得る。画像全体で行うノルムの正規化、または平均と分散を用いた正規化によっても特徴量は変化しないことがわかる。

また、コントラスト強調に用いられるヒストグラム平坦化処理についても、ヒストグラム平坦化画像 $HistEq(I)$ について

$$QT(I) = QT(HistEq(I))$$

なる関係がなりたつ。よってQTRは、これらの変換に対する不変量となる。

2.3.2 各種変換、ノイズに対する性能

マッチングのための類似性測度には、正規化相関や残差の2乗和(SSD)がよく使われる。これらの測度はヒストグラム平坦化などの濃淡値フィルタリングにより、類似度の値が影響を受けてしまう。本稿では、定量的な類似度との比較のために、正規化相関(CC)：

$$CC = \frac{\sum_{k,l} (I_1(k,l) - m_1)(I_2(k,l) - m_2)}{\sigma_1 \sigma_2} \quad (4)$$

を用いる。

なお、 $I(k, l)$ は画素 (k, l) の位置の画素値、 m_1, m_2 は I_1, I_2 の画素の平均値。 σ_1, σ_2 は I_1, I_2 の画素の分散値を表す。

実験用のテンプレートとしては、テクスチャの違いによる差異を調べるためにも、図8に示した2種類を用意する。それぞれのテンプレートのサイズは 40×40 画素とし、(a) 正面顔画像と (b) ランダム画像に対して、様々な変換や正規乱数ノイズを加えて、性能評価する。



(a) Face



(b) Random

40x40

図 8: 実験用テンプレート

回転変換に対する変化

回転変換についての変化を調べる。各テンプレートに対して、1度刻みで回転させた画像を生成する。この場合、回転の大きさに応じて、各回転画像の大きさが異なる。テンプレートとの類似度は、その回転画像中をすべて走査し、最大となる類似度をその回転度合いに対する類似度としてグラフを作成した。

図9、図10は、そのグラフを示したもので、いずれのテンプレートにおいても、±5度付近で類似度が急落する。テクスチャの違いによって、若干落ち方に違いがあるものの、傾向は変わらない。これらの類似度は回転には弱い、回転方向のテンプレートをどの程度の刻みで用意すればよいという指標となる。

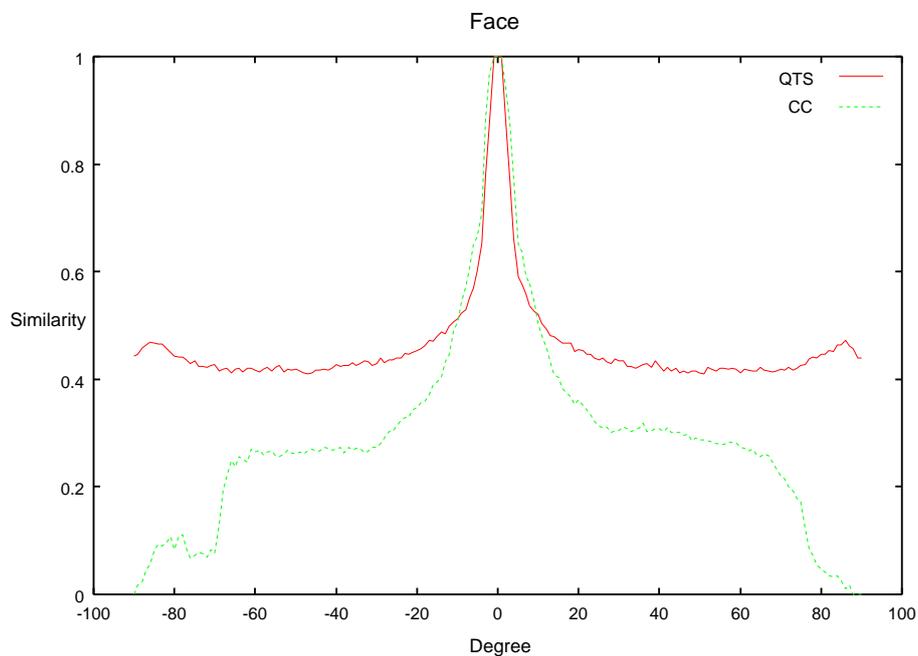


図 9: 回転変換に対する類似度の変化 (Face)

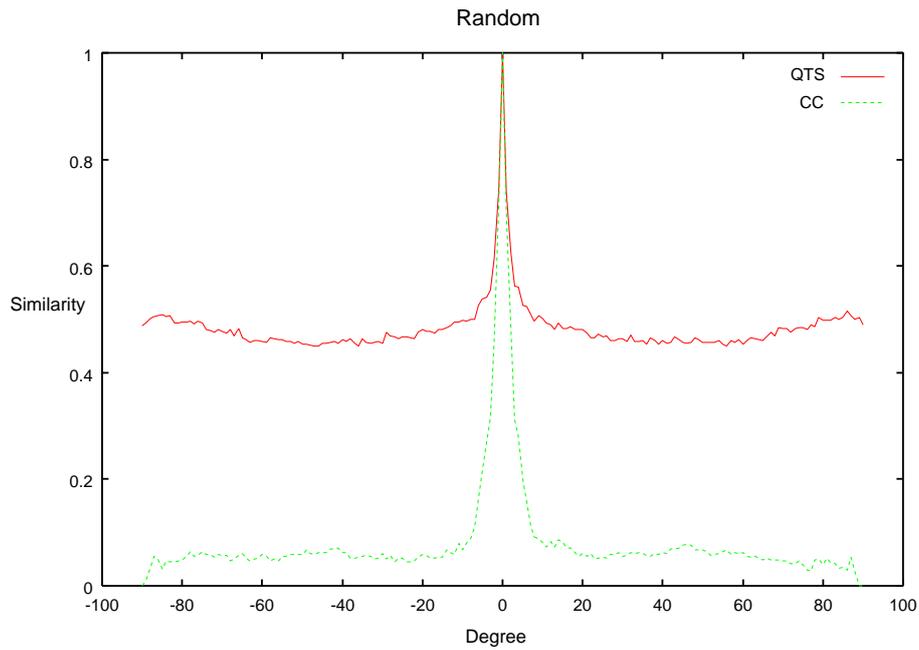


図 10: 回転変換に対する類似度の変化 (Random)

ガウシアンフィルタに対する変化

各テンプレート画像に対して、半径3画素の大きさのガウシアンフィルタを適用する。図11、図12はボケの大きさを決定する分散 σ (図中横軸 Sigma) を変化させ、それぞれの類似度の変化を示したものである。QTSは、CCと比べ、急に類似度が低下する傾向を持つ。このように、ボケについても急峻なピーク値をもつ。

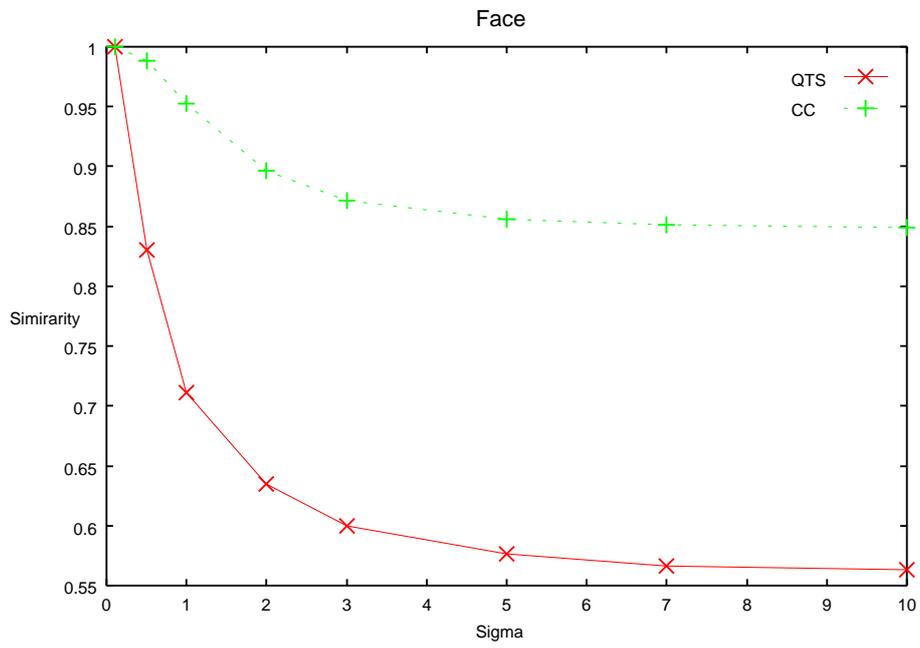


図 11: ガウシアンフィルタに対する類似度の変化 (Face)

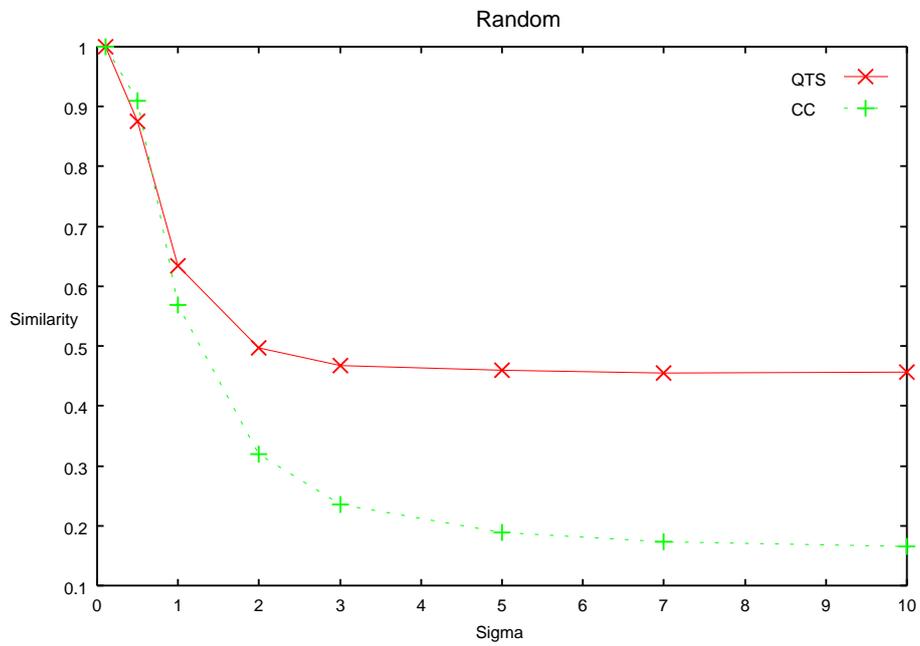


図 12: ガウシアンフィルタに対する類似度の変化 (Random)

正規乱数ノイズに対する変化

次に、各テンプレートに対して、正規乱数ノイズを付加した画像を生成し、それらとの類似度を求める。ノイズ画像は分散 σ を0.5, 1.0, 2.0, 3.0, 5.0, 10.0 と変え、それぞれの分散に対してランダムな画像を各1000枚ずつを求め、これらの画像との類似度のヒストグラムを作成した。なお、類似度はQTSについてのみ求めた。

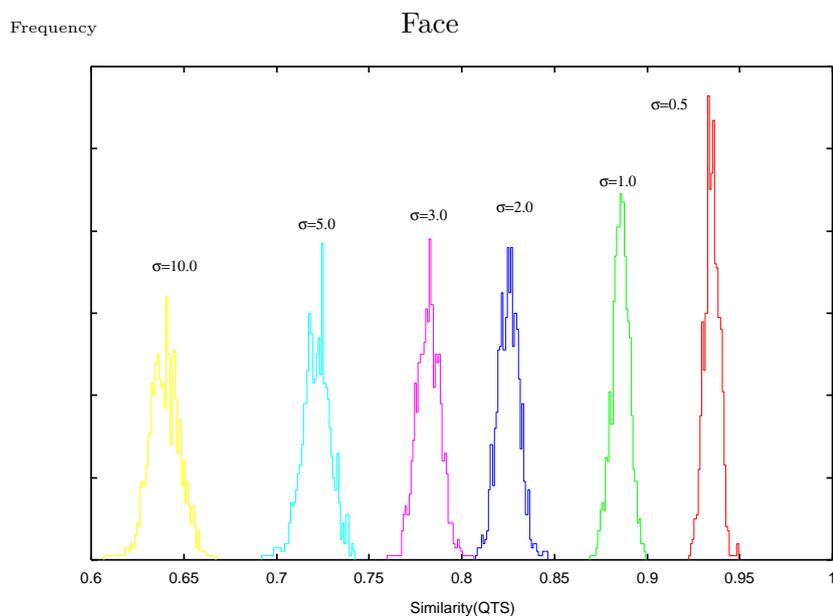


図 13: 正規乱数ノイズに対する類似度の変化 (Face)

図 13、図 14 は、ヒストグラムを示したもので、 σ 値が変化するにつれて、それぞれのヒストグラムの山がなだらかになるのがわかる。また、ノイズに対しても類似度のピークが急峻であることがわかる。顔パターンに対しては、類似度の低下がノイズの大きさに応じて緩やかにみられ、テクスチャによって影響を受けやすいものとそうでないものが存在することが推測される。

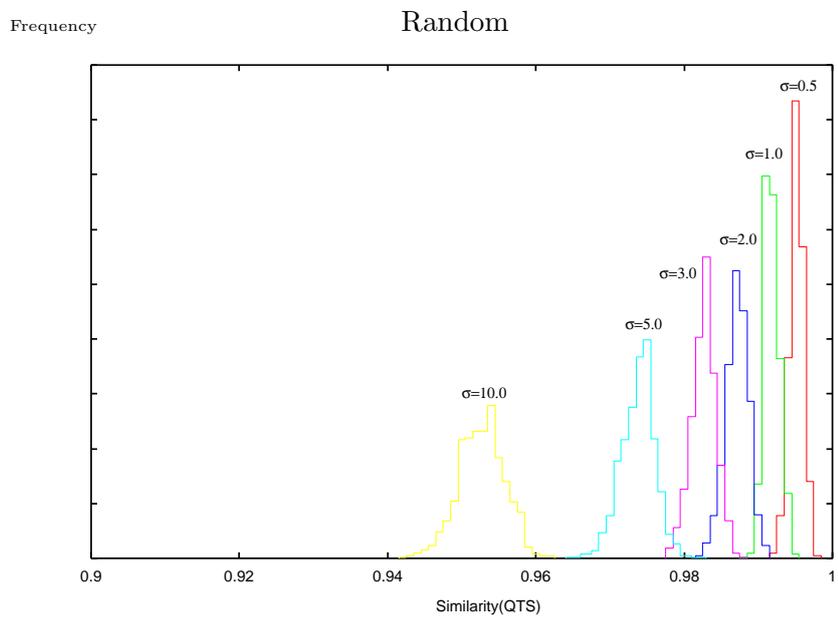


図 14: 正規乱数ノイズに対する類似度の変化 (Random)

2.3.3 3 値画像を用いた従来例との比較

3 値画像を用いてマッチングを行う例として、奥田ら [17] は、3 値輪郭表現を用いて、位置決めを行う手法がある。しかしながら、その 3 値表現を求めるために、ガウシアンラプラシアンフィルタを用いて輪郭付近の符号変化を検出を行っているため、計算コストが大きくなる。この手法に関連する鷺見らの文献 [18] においても、3 値表現の一致度の定義として、exclusive-nor 演算を拡張し、符号が同じ場合に+1、等号と符号が異なる場合は 0、大小異符号の場合に-1 としている。一致度の定義をこのようにした場合、大小異符号のケースのペナルティが大きくなる。すなわち、テンプレート内部での類似性に対する加法性が失われるために、遮蔽ノイズに対する耐性に影響がでる可能性がある。よって、本手法においては、符号の一致度のみを計算することとしている。

2.3.4 増分符号相関との比較

村瀬ら [14] は、隣接画素間の増分傾向をマッチングに利用する、増分符号相関が提案されており、性質が統計的な観点からまとめられている。この増分符号相関の定義では、同値の場合も「増分」とされており、同値関係の扱いが不適切なため、図 15(a)~(e) の画像はすべて同値として扱われる。

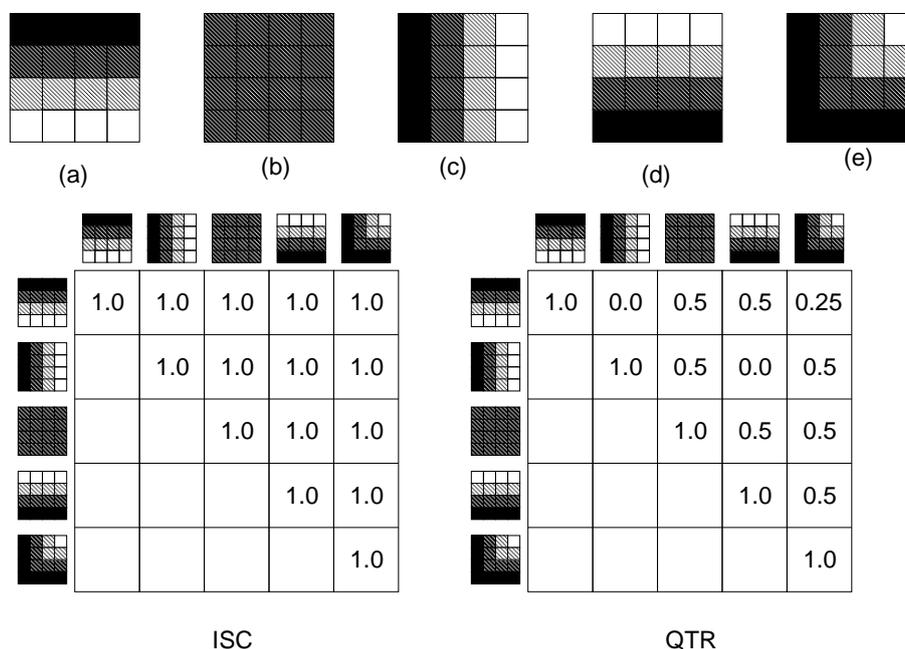


図 15: 増分符号相関 (ISC) と定性的 3 値表現 (QTR) の画像間の類似度の違い

また、佐藤ら [19] は、方向性を考慮した周辺増分符号相関も提案している。これらの類似度は、明度の均一領域において不安定であるという性質を原理的に持っていることが明

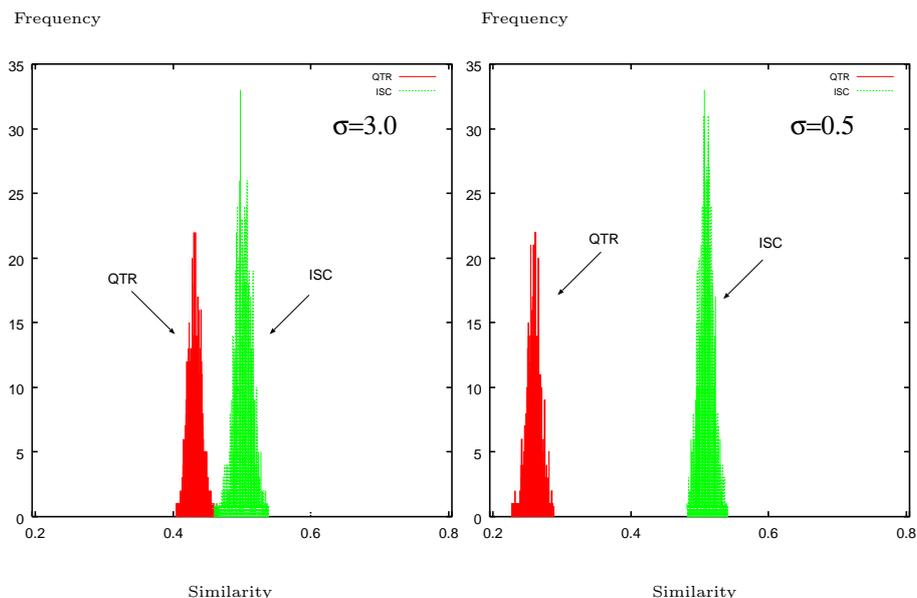


図 16: 増分符号相関 (ISC) と定性的 3 値表現 (QTR) の正規乱数ノイズを付加した場合の影響の比較

らかにされている。提案法の QTR では、方向性に関しては垂直方向を考慮し、同値関係については増分方向とは区別するという立場をとるため、明度の均一領域においての不安定さを幾分改善できると考える。符号の非反転確率を用いて導出された増分符号相関による類似度の期待値は、ノイズの分散を変化させても、0.5 となる。一方、QTR では、符号を 3 つに区別しているため、濃淡変化が少ない部分との判別が容易になる。このことを確認するために、濃淡変化のない画像（すべての画素が同一輝度値）に対して、正規乱数ノイズを加えた 1000 枚の画像を作成し、前章でテンプレートとして用いた顔パターンに対する、QTS、ISC、それぞれの類似度を計算する。ノイズの分散は、 $\sigma = 0.5, 3.0$ の 2 種類の系列を用意し、 σ が小さい程、濃淡変化の少ない画像、すなわち、同値関係 (=) を表す、値 0 の部分が増加する。図 16 は、それぞれの類似度の分布を示しており、 $\sigma = 0.5, 3.0$ を変化させても、ISC の分布が変化していないのがわかる。一方、QTR では、濃淡変化の少ない画像になれば、類似度が低下している。この性質は、均一領域から物体を検出する際の過検出を低減させ、パターンの相違度を適切に与えられること意味する。

2.4 実験

次に、実画像を用いて、領域のトラッキングとテンプレートマッチングによる対象検出の実験を行う。

2.4.1 テンプレートマッチング

次に、検出対象モデルのテンプレートを用いて、画像から対象の位置を検出する実験を行う。本実験では、拡大、縮小を考慮したテンプレートマッチングを行う。テンプレートの大きさは固定とし、様々な大きさの対象を見つけるために、対象画像を数段階に縮小し、それぞれの縮小画像内でテンプレートを検索する。テンプレートの位置を変えながら、それぞれの場所での類似度をもとめ、画像をすべて走査する。すべての拡大率の画像に対して処理を行い、最大値をとる拡大率、位置の情報を対象の場所として出力する。

検出対象としては、各個体の変形が大きいものの、平均的な画像を標準的なモデルとして利用可能な人物の正面顔とした。テンプレート作成のために、文献 [20] の実験のために収集した中から 5000 枚の顔画像データを用いた。この顔画像は PC の前に設置したカメラで撮影し、下側から見上げた画像となっている。実験のテストデータには、財団法人ソフトピアジャパンで開発された顔画像データベース¹を使用し、テンプレート画像とほぼ同じ角度で撮影された画像セットからさまざまな年代、性別の人物データ 300 枚を用いた。画像データは 320×240 画素に縮小し、実験を行った。また、カラー情報は用いないので、白黒濃淡画像に変換して利用する。

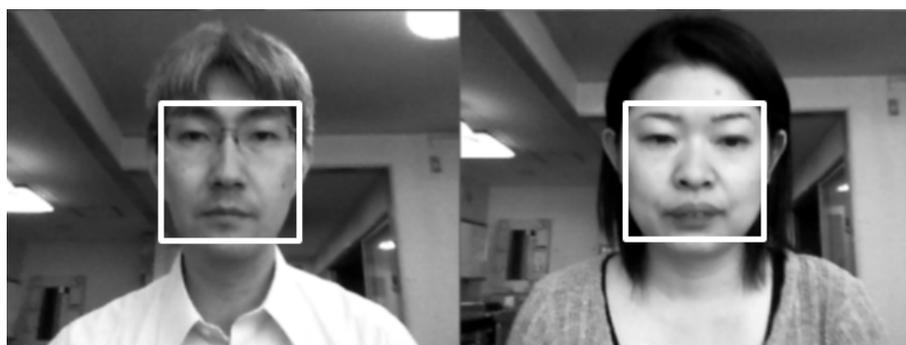


図 17: テンプレートマッチングによる検出の様子

テンプレートは、15×15 画素の大きさとし、それぞれの画像から抽出した顔領域 5000 枚の平均画像を用いた。テンプレートのサイズは、文献 [20] にて顔認識に利用しているテンプレートサイズを基準として、システム設計時に 256 画素であることとしたことと、テンプレートの中心を明らかにするために、各辺が奇数個のピクセル数とし、15×15 画素としている。類似度を正規化相関とした場合と QTR の場合を実験する。各画像には目鼻 4 点の正解点が入力し、検出領域がその 4 点を含む場合に正解と判定した。検出例を図 17 に示す。

¹本データベースは、2022 年現在は利用できない。図 17、図 18、図 20 については、同サイズの別の画像で再構成した画像を使用した。

正解率は、以下の表1のようになり、隠蔽がない場合、ほぼ同等の結果を示した。

表 1: 検出正解率の比較
正規化相関 (CC) 定性的 3 値表現 (QTR)

正解率	97.00%	97.33%
-----	--------	--------

計算量を比較すると、一回の類似度を求める操作において、正規化相関では、225回の減算、乗算が必要であるのに対し、QTRでは420回の加減算と比較演算が必要となる。本実験では、元画像の0.08倍から0.30倍まで0.01倍刻み、24段階の縮小画像を用意し、マッチングを行った。このときの総マッチング回数は、41105回であり、Pentium4 2.0GHzマシンで比較したところ、処理時間は正規化相関の場合67.7msec、QTRの場合118msecであった。当初の見積もりでは、乗算を用いる演算が存在する正規化相関が計算時間がかかると考えていたが、マッチングデータの次元数が約2倍になっていることやデータ構造として各画素を1byte単位で3値画像を表現していること、実装コードの最適化等を行っていないため、QTRに時間がかかる結果となった。そこで、マルチメディア命令による演算の並列化を行った結果、同じQTSの計算は、25.6msecまで減少する。このようにQTRは、演算の効率化を実現しやすいというメリットを持つことが確認できる。さらに、3値画像は各画素を2bitで表現できることから、更なる、省メモリ化、高速化が期待できる。

次に、隠蔽に対する耐性を比較するため、テスト画像に隠蔽ノイズを加え、検出率を求める。隠蔽条件は、顔の半分が影などの影響で隠れることがあることをシミュレーションするため、図18のように画像の左から40%の位置と50%の位置までの部分隠蔽ノイズ画像を作成し、各テスト画像に掛け合わせた画像を作成した。撮影されている顔位置は各個人で微妙に異なっているため、各個人ごとの顔領域の隠蔽状態はそれぞれ異なる。

結果は、同様に表2に示す。正規化相関では、隠蔽ノイズが大きな場合には、検出率の低下が大きい。一方、QTRは、正規化相関の検出率が3%の結果となる隠蔽状態でも60%程度の正解率を得ることができ、部分的な隠蔽ノイズへの耐性を示している。

表 2: 部分隠蔽に対する検出正解率の比較
正規化相関 (CC) 定性的 3 値表現 (QTR)

隠蔽なし	97.00%	97.33%
隠蔽 40 %	72.00%	90.66%
隠蔽 50 %	3.00%	62.66%

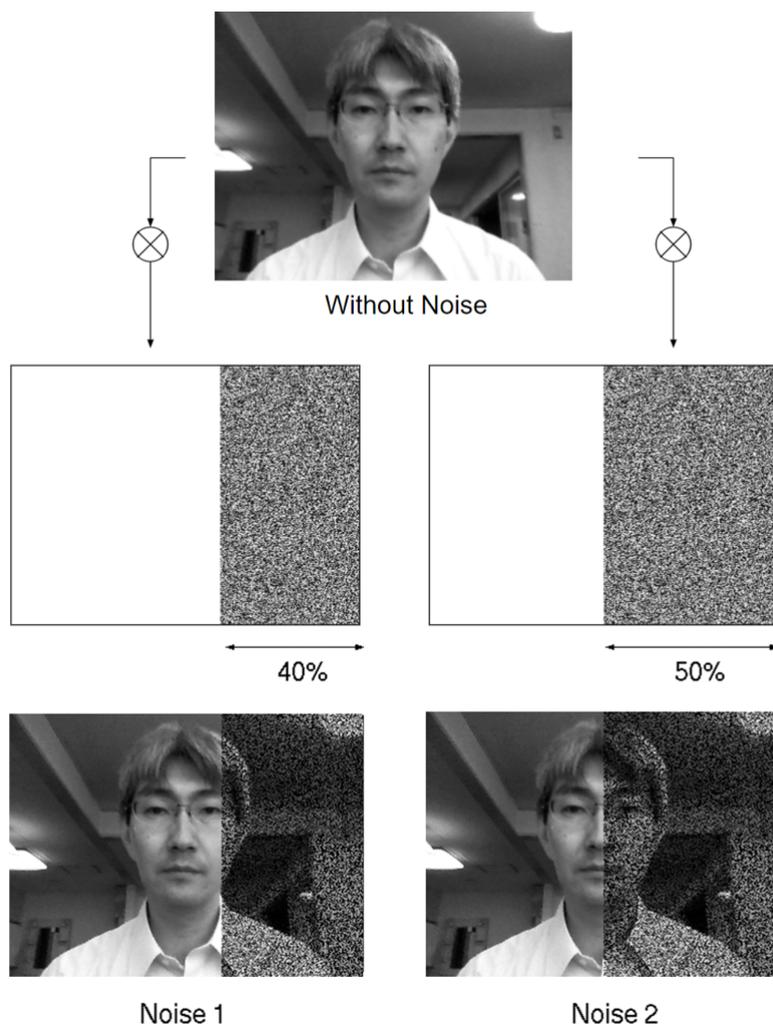


図 18: 遮蔽ノイズの混入

2.4.2 トラッキング

図 19(a) は、手を右から左へ動かした画像系列 (320×240 画素) を表す。テンプレートは、その手領域を囲む画像とし、縮小した 30×30 画素の画像とした。マッチングを行う場合、入力画像を縮小し、各画像から与えたテンプレートと最大の類似度となる位置をもとめる。その場所において抽出した、同じ大きさの画像でテンプレートを更新し、時系列処理を行う。類似度は、正規化相関と QTR の 2 つを用い、トラッキングの性能を比較する。図 19(b) は、正規化相関を類似度として、トラッキングを行った結果である。手領域とその周辺の画像の輝度が高いことが影響しており、最初に与えた位置が常に最大の相関値を示し、検出位置がほぼ変わっていない。

一方、図 19(c) は、QTR によるトラッキング結果である。途中、左端へのずれが生じるものの、複雑な背景下でも手領域を追従しているのがわかる。

2.5 ハイブリッドマッチング

QTR のような定性的な方法を、定量的な方法、もしくはコストのかかる計算と組み合わせる場合を考える。QTR を用いた類似度は計算コストを抑えられるため、概略類似したものを先に求めておき、その後計算コストのかかる手法を必要な部分にだけ類似度計算を適用するといったハイブリッドな手法が考えられる。

具体的には、QT 表現を用いた類似度に対してしきい値を設定し、そのしきい値が越えた場所だけに対してコストの大きな類似度計算を行う。これにより、従来と比較して高速なマッチングが行える。以下の実験では、その効果を示す。

方法は平均顔テンプレートを用いて、QT 表現によるマッチングを行う。その計算の途中結果である、 D_v, D_h に対してしきい値を設定する。 $D_v > Th1, D_h > Th2$ の条件を満たす場所に対して、後処理を行う。この方法の利点は、コストの小さい処理で前処理 (大分類) を行えるだけでなく、後処理 (例えば、定量的なマッチング法) において、誤認識を起こしやすいパターンを性質の異なる前処理で除去できる。

例として、図 18 のノイズのない画像を用いて、しきい値 $Th1, Th2$ を変化させる実験を行った。この実験では、 $Th1 = Th2$ とした。総マッチング回数の 41105 回に対し、削減した結果が表 3 であり、その様子を図 20 に示す。候補として残っている場所に対して、矩形を表示させている。

この画像に対しては、しきい値を 125 にしても、正解が候補に含まれていた。 D_h, D_v の最大値である 210 の 50% に相当するしきい値を 105 に設定すると、2% 程度まで候補が削減される。図 20 より適当なしきい値与えることで、背景の領域等を効率よく削除で



(a)Image Sequence



(b)Tracking by CC



(c)Tracking by QTR

図 19: トラッキングの実験

表 3: QTR による候補削減

しきい値	候補数/総候補数	削減率
90	9944/41105	24.2%
95	5198/41105	12.6%
100	2227/41105	5.41%
105	701/41105	1.71%
110	151/41105	0.37%
115	44/41105	0.11%
120	15/41105	0.03%
125	4/41105	0.01%

きているのがわかる。後処理のコストの大きさや正解が候補に含まれるか否かを判断してしきい値を決定すれば、検出のスピードの改善につながる。

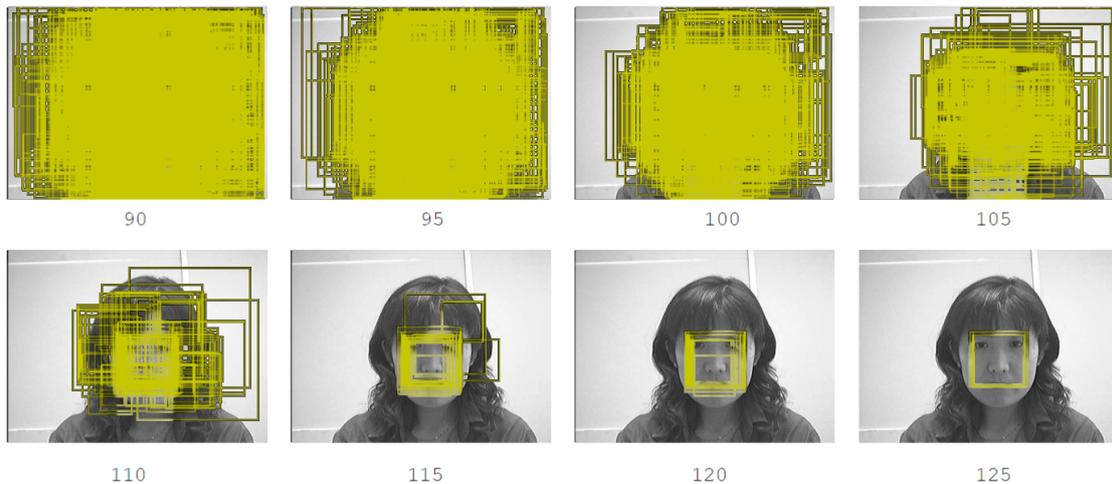


図 20: しきい値を変えた場合の候補削減の様子

2.6 深層学習との組みあわせの検討

QTR は、その不変性により輝度変化に頑健なマッチングを行うことができ、計算量を削減することができた。この表現を用いた多クラスの物体認識である識別タスクへの適用を検討する。近年、物体認識のベンチマークデータとして用いられている CIFAR データ [21] を用いて、クラス識別のタスクの予備実験を行った。10 クラスの画像セット

CIFAR-10は、32x32ピクセルのカラー画像のデータセットであり、5万枚の訓練データ用（画像とラベル）と1万枚のテストデータ用（画像とラベル）の合計6万枚からなる。QTRを求めるため、それぞれの画像を濃淡グレー画像に変換して、3値表現を求める。そして、識別にはk-nearest neighbor法(k=10)を用いた。その結果、分類精度が34.97%となっている。[21]には、初期のCNNモデルによって、82%の正解率であったとの記載があり、この表現を直接用いた多クラスの物体認識では、十分な識別能力がないと考えられる。そこで、本節では、QTRの拡張とDeep Neural Networkとの組み合わせについて検討する。

2.6.1 定性的3値表現のカラー化

物体認識技術における色情報を使用することは、深層学習を用いた各手法においてもRGBの3チャンネルの入力が必須であり、一般的になってきている。しかし、これまでに述べたように、QTRは濃淡グレー画像に対してのみ定義されているため、カラー画像を扱えるように、定義の拡張を検討する。拡張の方法としては、1) RGB各プレーンのQTRを計算して組み合わせる方法

2) 色空間を変換し、色属性ごとにQTRを算出する方法

などのアプローチが考えられる。色空間の変換によっては、HSV表色系のHue（色相）のように周期的な座標系を持つものもある。

1)のアプローチでは、R,G,BのそれぞれのプレーンのQTRを求め、その一致度を求める方法である。

図21は、QTRによるk-nearest neighbor法とCNNネットワークであるResNet-18[22]による識別の結果を示している。QTRによるk-nearest neighbor法では、カラー化したケースでは37.47%と若干向上していることがわかる。

比較として、画像のAugmentationとして水平フリップのみを行い、50エポックの学習を行ったCNNネットワークResNet-18と比較した。こちらについても、色情報を導入する前後で、0.63ポイントしか向上がみられず、CIFAR10を用いた場合の色情報の効果が得られにくいことが見受けられる。

次に、2)のアプローチとして、変換後の色空間を利用する方法として、色相属性に着目する。色相属性は、周期的な構造をもつとともに、明るさなどの変化に対しても不変な属性として認知されている。この色相属性を3値化することで2画素間の色相の類似性を表現する。図22では、色相環の2色間の相対角度に閾値を設定し、類似色、異系色、反対色のそれぞれに1,0,-1を割り当てを行う様子を示している。ここで、2つの画素の選び方にも、さまざまな選び方があり、隣接する2つの画素の間のコードも考えられるが、ここでは、深層学習のAugmentationでもよく用いられている水平フリップに対し

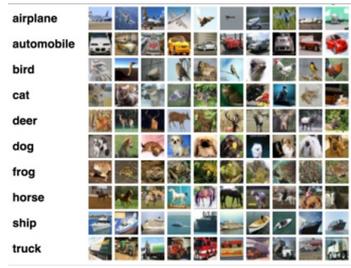
	QTR (10-nearest neighbor) (Intensity (Gray) Only)	CNN※ (Gray Image Only)
	34.97%	83.00%
	ColorQTR (10-nearest neighbor) (Red, Green, Blue)	CNN ※(Color Image)
	37.47%	83.63%
		※ResNet-18, Modest Augmentation, Epoch 50

図 21: CIFAR10 を用いたカラー情報の導入における正解率の変化の様子

て対応する画素の色属性の関係をコード化する。 $H_s(x, y)$ は画素 (x, y) の色相値を表し、レンジを $[0, 179]$ とする。このとき、3 値表現 Q_c は以下のように定義される。

$$Q_c(x, y) = \begin{cases} 1 : d_{ang}(\theta_{xy}) < T_s \\ 0 : T_s \leq d_{ang}(\theta_{xy}) \leq T_d \\ -1 : d_{ang}(\theta_{xy}) > T_d \end{cases} \quad (5)$$

$$\theta_{xy} = H_s(x, y) - H_s(M - 1 - x, y) \quad (6)$$

$$d_{ang}(\theta) = \min(|\theta|, 180 - |\theta|) \quad (7)$$

ここで、 θ_{xy} は、2つの色相のなす角度で、 d_{ang} は値が 180 未満の定義域に入るようにした相対角度を表す。 T_s は類似色を満たす範囲を設定するしきい値、 T_d が異系色と反対色の設定するしきい値となっている。

2.6.2 定性的 3 値表現を用いた CNN の導入

次に、QTR 特徴の識別力を高めるために深層学習ネットワークを用いた改良を検討する。CNN の場合、通常、3 チャンネルのカラー画像を入力とする。この各チャンネルに、

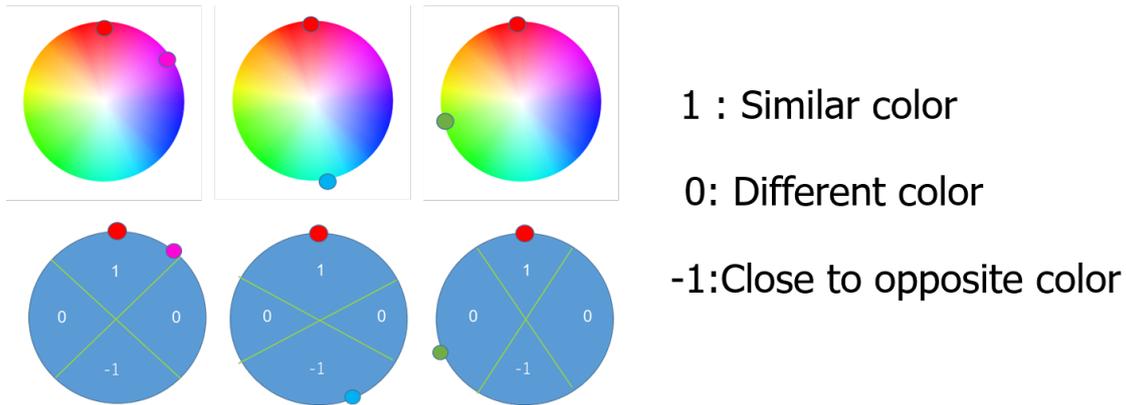


図 22: 色相属性の 3 値化：色相の角度方向に閾値を設定し、相対的な角度に応じて 3 値を割り当てる。

QTR で変換した三値画像を直接入力して学習させることを試みる。これは、通常の輝度画像に対する畳み込み演算と深層化したネットワークを利用して、不変量表現となっている 3 値画像の分類能力を高められる可能性があると考えたためである。QTR の時点で輝度変化にはロバストな表現となっていることから、クラス間の差を深層学習によって引き出すことができれば、識別能力の改善が図られると考えられる。式 2 の 2 枚の 3 値画像 Q_v, Q_h は、構成される画素値は -1, 0, 1 の 3 値である。それぞれのチャンネルに入力されるが、このとき既製のネットワークを用いて学習できるように、1 チャンネルはダミー画像（すべての画素値が 0 の画像）として学習させる。さらに、色情報を考慮した、式 5 で作成される 3 値画像 Q_c を第 3 番目のチャンネルとする構成も考える。

また、2.5 節で行ったように、QTR 表現と正規化相関のハイブリッドによる高精度化という、従来法とのハイブリッドというアプローチも考えられるため、3 値画像 Q_v, Q_h に加え、3 チャンネル目にグレー輝度画像を導入した QTR と多値画像のハイブリッド構成も考える。

2.6.3 QTR 画像を入力とした ResNet による実験

CNN のモデルには ResNet18 を使用する。ここでは、Augmentation として Random-Crop と HorizontalFlip を適用し、各モデルで 100 エポックの学習を行った。モデルの学習時に最も精度の高いモデルを保存した。QTR は、先に述べたように k-nearest neighbor (KNN) 法で評価しており、表 4 に QTR を入力とした ResNet18 を用いた結果と関連する結果を示す。

元の KNN を利用した QTR に基づく分類と比較して、CNN を用いることで 37.47% から 84.96% に性能が向上していることがわかる。これは、[21] で述べられている初期の深

表 4: CIFAR-10 テストセットの識別結果 (ResNet18 に QTR の入力を行ったもの)

			Acc.@1
QTR (10-NN)		Gray	34.97%
		Color	37.47%
ResNet18 with QTR as input	(a)	2ch(Q_v, Q_h)	84.96%
	(b)	3ch(Q_v, Q_h, Q_c)	85.62 %
Hybrid	(c)	2ch(Q_v, Q_h)+Gray	93.61%
ResNet18 (Original)	(d)	Color	94.98 %
	(e)	Gray	93.22 %

表 5: CIFAR-10 の色拡張したテストデータセットの識別結果

			Ave. of Acc.@1
ResNet18 with QTR as input	(a)	2ch(Q_v, Q_h)	82.88%
	(b)	3ch(Q_v, Q_h, Q_c)	83.34 %
Hybrid	(c)	2ch(Q_v, Q_h)+Gray	91.52%
ResNet18 (Original)	(d)	Color	85.71 %
	(e)	Gray	90.95 %

層学習ネットワークでの識別性能と遜色ない結果であり、3値画像のみが入力となっている縮退した情報表現でも識別能力が改善されていることは興味深い。また、QTR に色情報を導入した Q_c を加えることで、グレー画像 (a) 84.96% からカラー画像 (b) 85.62% に向上している。これは色情報の追加に関しては、大きなアドバンテージとはなっていないことは他の例と同じであるが、色に関する情報の追加によって、性能改善することが確認できた。これは、不変性をもつコード画像を追加することで、取り扱える変換のバリエーションが増加し、表現能力が高くなっていることで、性能改善が行われた可能性が存在することを示唆している。

また、QTR と多値画像のハイブリッド構成の方法 (c) では、93.61% まで改善する。これは、通常のカラ画像に対する CNN の結果 (d) 94.98% には及ばないものの、グレー画像のみに対する CNN の結果 (e) 93.22% より良好である。このことから、QTR 不変量の追加は性能改善に効果があると考えられる。

この追加した不変量の効果を検証するために、CIFAR10 のテスト画像に Augmentation

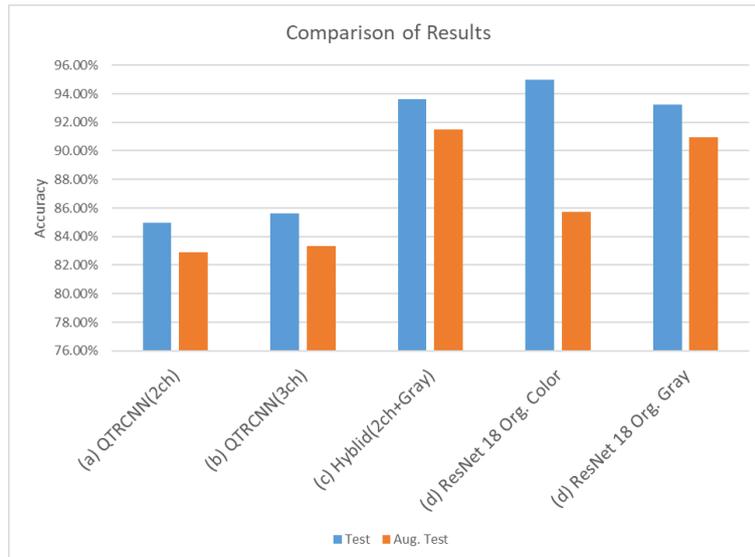


図 23: CIFAR-10 のオリジナルテストデータセットと色拡張したテストデータセットの結果の比較

として ” transforms.ColorJitter ” を適用し、色味や明るさなどをランダムに変化させた CIFAR10 のテスト画像を作成した。各属性の変換確率は、明るさ=0.5、コントラスト=0.5、彩度=0.5、色=0.5 と設定した。それぞれ異なるデータを作成し、5回ずつ実験を行った。

表 5 に各モデルの評価と 5 回のテストの分類精度の平均を示す。また、図 23 は、表 4 と表 5 の結果を比較したグラフである。

まず、QTR を入力した CNN に関連する (a)、(b)、(c) については、オリジナルデータセットと色拡張したデータセットの結果の差はほぼ同じであるが、カラー画像の CNN(d) の性能は大きく低下している。通常の CNN のモデルは、明るさや色の変化が加わったデータに対してロバスト性を失っていることがわかる²。また、色情報が導入されている (b) と (d) を比較しても、色情報に対する不変量 Q_c を導入することで、(a) よりも性能がよく、色の変化に対する耐性も改善がみられることから、不変量の導入が効果があることがわかる。一方、明るさ情報を用いた (c)、(e) の結果は、大きくは下がっていないが、ハイブリッド構成の (c) 91.52% が最も良い結果となっており、不変量による情報の追加は色や画像の明るさに対するロバスト性能を有効に維持できることが確認された。

² もちろん、このネットワークは学習時の Augmentation に Colorjitter が加えられていないことに注意

2.6.4 バイナリニューラルネットワークを用いた高速化

CNNの強力な分類能力を活用することで分類性能は向上するが、マッチングの計算コストは必然的に高くなる。畳み込みニューラルネットワークは、一般的なビジョントラスクに対して高い計算コストとメモリコストを発生させる。

近年、バイナリニューラルネットワークにおいて、計算コストやメモリコストの削減が進んでいる [23][24]。いくつかの手法の中で、3値を入力として2値の重みを用いて計算を行う Ternary-Binary Network (TBN) が提案されている [25]。本手法は標準的なCNNの効率的な近似を提供する。高速化された3値と2値の行列演算に基づき、TBNは標準的なCNNの算術演算を効率的なXOR、AND、ビット数演算に置き換え、メモリ、効率、性能の最適なトレードオフを提供する。

このフレームワークでは、通常、多値画像を閾値処理によって3値画像に変換し、それをネットワークの入力とする。ここで、QTR自体はもともと3値なので、このネットワークとの親和性が高いと考え、適用しました。[25]によると、TBNはオリジナルの浮動小数点演算ネットワークの約40倍高速とされている。

表 6: CIFAR-10 の Ternary-Binary Network による認識結果
(PreAct-ResNet18 の近似)

		Acc@1	Acc@5
TBN with	2ch(Q_v, Q_h)	67.72%	96.78%
QTR	3ch(Q_v, Q_h, Q_c)	68.70%	96.55%
Hybrid	2ch(Q_v, Q_h)+Gray	73.55%	97.61%
Original	Color	78.24%	98.44%
→ 3 levels	Gray	75.06%	98.06%

次に、高速化のために適用したTBNによる認識結果を表6示す。実験に用いたモデルはpreact-ResNet18の近似モデルであり、前節の実験と同様に、異なる入力のケースでモデルを学習し、テストデータで精度を求めた。また、モデルを学習させた際に最も精度が良かったモデルを保存している。TBNの推定結果は、表4の浮動小数点演算ResNetの結果と傾向が同様であることがわかる。もちろん、オリジナルと比較して精度が比較的低く、2値化の影響が大きいためと考えられる。これには、さらに層の深い、規模の大きなネットワークの採用を検討する必要があり、ネットワーク規模と性能、速度のトレードオフについても今後検討が必要である。

2.7 本章のまとめ

本章では、第一の視点に立って、定性的3値表現(QTR)を提案し、マッチングにおけるさまざまな性質について調べた。正規化相関などの従来法と比較し、領域のトラッキング、顔検出実験等を実画像にて評価した。実用上は、定性的3値表現におけるマッチングを前処理として行うハイブリッドマッチングが効果的であることを示した。さらに、QTRを識別タスクへ適用することを検討するために、カラー画像への対応と、畳み込みニューラルネットワークとの組み合わせについても検討を行った。定性的3値表現で得られるコード化画像をダイレクトに畳み込みニューラルネットワークへの入力とすることで、ネットワーク構造等をそのまま利用して学習、識別が可能となる。さらに、多値画像と不変量のコード化画像を同時入力することで性能改善が行われる可能性を確認することができた。

3 パターンハッシングによる物体認識

本章では、第1章で述べた第二の視点に立って、パターンの取得時の変化に対応した不変的な特徴を見出してアルゴリズムに反映させる。主には、第一の視点で扱ったフォトメトリックな変換に対する不変性に対して、対象の撮影時に起こる物体の位置、向き、大きさなど決定する幾何学的変換に対する不変性を導入する。前章で提案した、画像マッチングを複数行うことで、物体の検出や物体の構成部品の認識等を行う問題を取り扱う。幾何学的特徴と部分画像特徴の両者を融合したパターンマッチング法について考察を行い、物体の現れる位置、大きさなどの変化を許容するために、幾何学的変換に対する不変性をモデルに導入し、そのモデルを用いて部分隠れやノイズに対してロバストな認識方法を実現する。

3.1 アプローチ

物体認識は、コンピュータビジョンの中心的な問題の一つである。多くの物体認識方法は、幾何学的特徴点ベースの研究 [2] とアピアランスベースの研究 [3][4] の2つのタイプに分類できる。

本章では、幾何学的特徴点ベースの手法に、アピアランスベースのアプローチを効果的に統合した新しいモデル表現法を構築し、そのモデルを用いて部分隠れやノイズに対してロバストな認識方法について述べる。

2つのアプローチを結合した方法として、Schmid[26] は、Harris のコーナー検出法 [27] によって求めた着目点 (interest points) において、微分ガウシアンフィルタに基づいた濃淡値に不変な局所的な記述を作成し、マッチングに利用した。Matas[28] は、不変点集合記述法 (Invariant Point-Set Signature) を提案し、 n 個の着目点から構成される凸包領域内で、着目点の間の線分上の濃淡情報を用いてモデルを記述した。いずれの方法も、局所的な情報の記述法が中心的な課題であった。

図 24 は提案手法の枠組の概念を示している。人間は複数の断片的な画像だけから顔の位置を推測でき、たとえ一つの部分パターンが欠けたとしても、複数の見え方の可能性によって正しい顔位置を求めることができる。図 24 といくつかの従来例は、モデル表現のための以下のヒントを与える。

- 局所的な見え方の表現

ノイズや隠れを持つ画像から目的の物体の正しい位置を見つけるためには、物体に対する複数の部分画像によるモデルの表現法が必要となる。

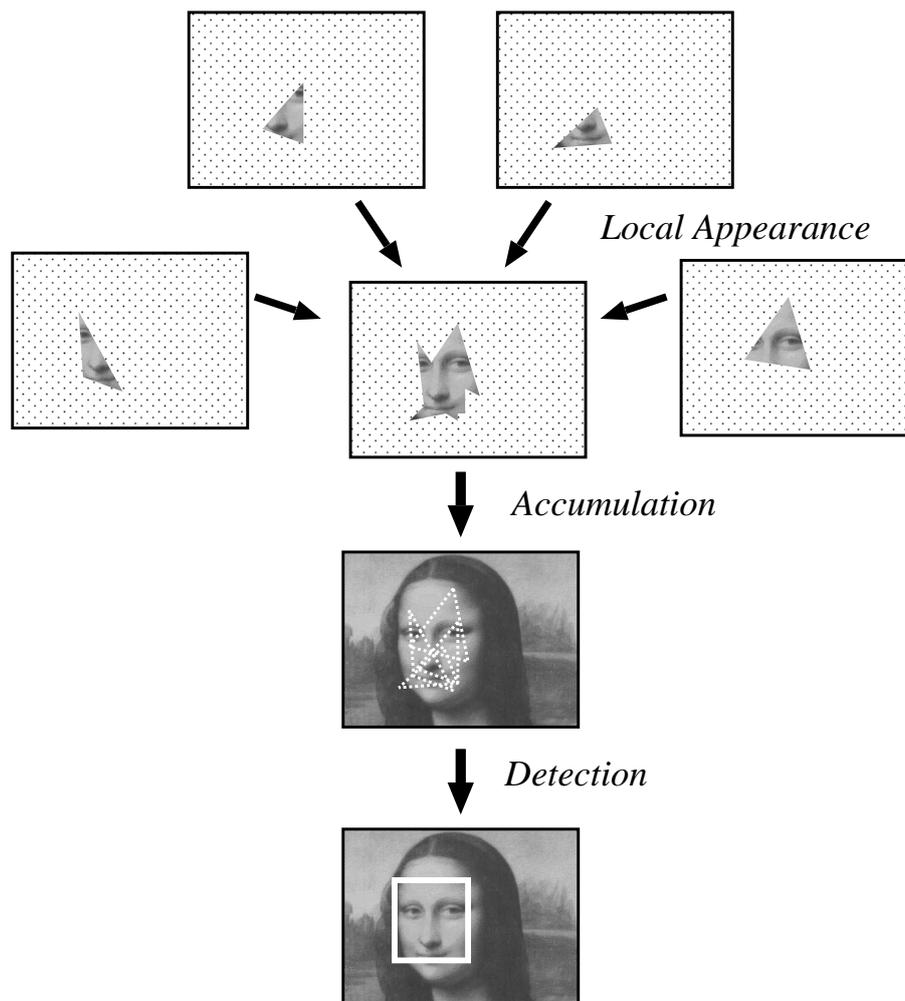


図 24: 提案法のモチーフ: 人間は部分画像を複数用いることで、正しく顔の位置を求められる

- 部分情報と全体情報の関連づけ
各部分画像パターンは、物体全体領域のどの部分であるかという関係を、効率良く関連づけて記憶する必要がある。

本章では、複数の部分画像マッチングを用いた物体認識の新しい方法“パターンハッシング”を提案する。物体のモデル画像は、複数の部分画像パターンに分割してハッシュテーブルに登録する。それぞれの部分画像パターンは、部分画像を切り取ったときの基準として使われた基底ベクトルから生成される幾何学的不変量に基づいた位置に格納される。この不変量は、類似した部分画像モデルをハッシュテーブルから検索するための索引として効果的に使われる。認識時には、テーブルに登録された膨大な部分画像モデルをハッシングにより高速に選択し、入力画像から抽出された部分画像とのマッチングを行う。部分画像同士が類似している場合、対象とする物体が入力画像中に存在すると仮定することができる。モデルの登録時には、部分画像と物体の存在領域との相対的な位置関係を記述しておき、部分画像のマッチングの結果から対象物の位置に関する存在領域が求まる。これらのパターンマッチングを繰り返し、複数の局所的な見え方を集積する。物体の存在領域の仮説を統合するためには投票手続きを用い、その投票結果から大きな投票値を選択することで、物体の位置、方向を検出する。部分画像のパターンマッチングを行うためにモデルパターンを“切り刻み (hash) ”、モデル検索のためにハッシングテクニックを利用することから、本手法をパターンハッシング (Pattern Hashing) と呼ぶ。

3.2 パターンハッシング (Pattern Hashing)

3.2.1 分散アピアランスモデル

提案法は、Geometric Hashing[2]と関連する。Geometric Hashingでは、モデル記述の基本要素は主として点であった。提案法では、基本要素として線や点といった幾何学的特徴の代わりに、新たに部分画像パターンを導入する。モデル画像を部分画像パターンに分割するには、(1) 分割に関する基準、(2) マッチングの際の画像パターンの正規化、を考慮する必要がある。どのように部分画像パターンを作成するかという分割に関する基準は、対象固有の知識に依存しないことが望ましい。よって、モデル画像に対して特徴点を求め、それらの組み合わせによって各部分領域を決定する。また、一般に画像マッチングでは、対象領域のサイズと方向の正規化が必要となる。そこで、正規化のための基底の選択方法も、対象モデル上の特徴点の組み合わせに基づいて決定すれば、上述の(1)(2)を共通化して考えられる。

しかしながら、特徴点の組み合わせを用いた部分パターンへの分割は、特徴点の数に応じて大量のパターンが生成されることが明らかである。これはモデルの選択も複雑になり、さらなるマッチングのコストの増大を招く。よって、モデルの検索のためには、効率のよい索引づけが必要となる。Geometric Hashingでは、2つの特徴点から基底を構成し、ハッシュテーブルを用いたモデルの検索、マッチングを行った。そこで提案法においても、ハッシュテーブルを用いることとし、モデルの部分画像パターンを効率よく選択する方法を与える。具体的には3つの特徴点から定義される2つの基底ベクトルから、同型変換に対する幾何学的不変量を計算し、これをハッシュ関数に適用する。これらの不変量は、部分画像パターンが入力画像の中でさまざまな位置、サイズ、方向で存在しても、不変な性質をもつため、類似したパターンを検索するキーとして都合がよい。

以上の考察から、大量の複数の部分画像パターンをハッシュテーブル上に分散して配置し、不変量を用いることで高速な検索を行うことが可能なモデル表現である、分散アピアランスモデルが構成できる。

3.2.2 モデルの登録

図 25 は分散アピアランスモデルの登録方法を示している。説明のために顔画像 (図 25(a)) を登録する場合を例にする。まず、モデル画像に対して特徴点検出を行う。対象物の全体領域をあらゆる矩形 G_p (図 25(b)) と、その矩形中の特徴点が着目点として選択される。次に、着目点集合の中から任意の3点を選び、ある1点を原点として、2つのベクトルを設定し、これを基底ベクトルとする (図 25(d))。この基底ベクトルは、部分画像パターンの設定と、索引づけのための幾何学的不変量を計算するために使われる。部分画像パターンは図 25(d) のように、基底ベクトルに平行な四辺形の領域を設定し、基底ベクトルを正規直交ベクトルに変換することで正方形パターンに正規化する。これらの正規化部分画像パターンは図 25(c) に示すように、3点の組み合わせの数だけ対象モデルから切り出される。なお、2つの基底ベクトルの成す角度が0度、180度に近い角度の場合は、部分画像がテクスチャとしての意味を持たないため、パターンの登録を行わない。これらの正規化部分画像パターンは、基底ベクトルから計算される幾何学的不変量に基づいたハッシュ関数により、テーブルの所定の位置に格納される。図 25(g) に示すように、テーブル上に登録されるデータは、正規化部分画像パターン $LPat$ に加え、部分画像が検出対象領域 G_p のどの部分に位置するかという相対的な位置関係 $RelG_p$ 、また、対象物全体の正規化画像パターン $GPat$ が共に記憶される。これらの詳細については、後述する。図 25(f) は、モデル画像に対して切り出された顔画像の正規化部分画像パターン集合を示す。これらがハッシュテーブル (図 25(e)) 上に自動的に分散配置される。

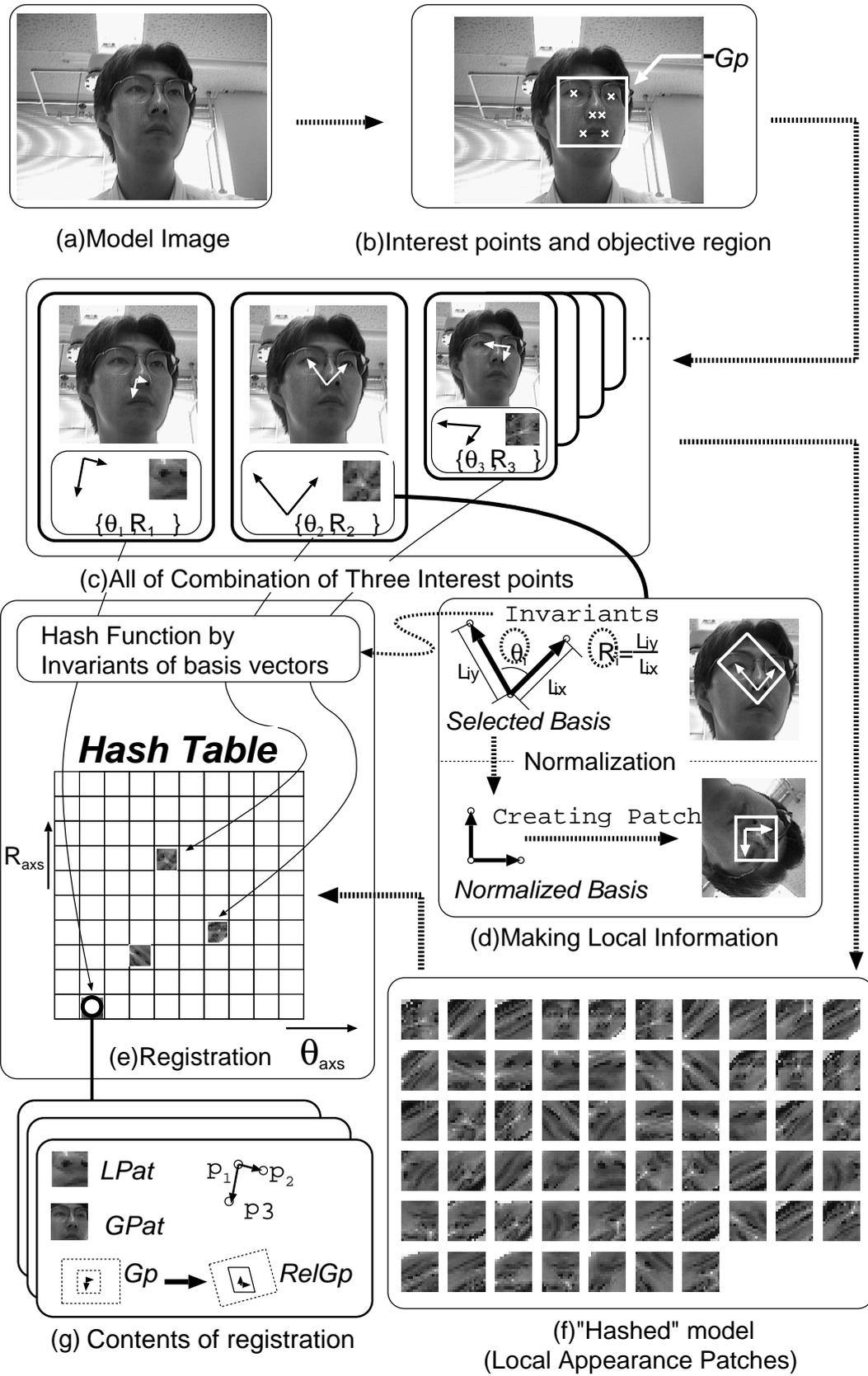


図 25: 分散アピランスモデルの生成

3.2.3 ハッシュ関数とハッシングにおける衝突処理

ハッシュテーブルにおける格納位置を決定するためのハッシュ関数のパラメータには、基底ベクトルを構成するために選択された3点 p_1, p_2, p_3 を用いる。基底ベクトルは、 p_1 を原点、 $\vec{p}_2 - \vec{p}_1, \vec{p}_3 - \vec{p}_1$ を2つのベクトルとする。このときハッシュ関数 HF は以下のように定義する。

$$HF(p_1, p_2, p_3) = \{\theta_{axs}, R_{axs}\} \quad (8)$$

$$\theta_{axs} = \cos^{-1} \frac{(\vec{p}_2 - \vec{p}_1, \vec{p}_3 - \vec{p}_1)}{\|\vec{p}_2 - \vec{p}_1\| \|\vec{p}_3 - \vec{p}_1\|}$$

$$R_{axs} = \frac{\|\vec{p}_3 - \vec{p}_1\|}{\|\vec{p}_2 - \vec{p}_1\|}$$

ここで、2つのベクトルが成す角度 θ_{axs} と2つのベクトルの長さの比 R_{axs} は、平行移動、回転、拡大縮小といった同型変換に対しての不変量となる。図 25(e) に示したハッシュテーブルは2次元であり、一方が θ_{axs} 、他方が R_{axs} の軸を表わしている。 θ_{axs} と R_{axs} は共に量子化され、2つの不変量によって、索引づけが行われる。

複数の部分モデルパターンの登録を逐次行くと、3点の組み合わせによって同じハッシュ関数の値を持つことがあり、ハッシングにおける衝突 (collision) が発生する。ハッシュテーブルの各位置では、チェーン法により、複数のパターンをリスト構造として持つようにする。すなわち、基底ベクトルの成す幾何学的関係が類似した複数の部分モデルは同じ位置に記憶される。

3.2.4 部分情報と全体情報の関係記述

ハッシュテーブルに登録される正規化部分画像 $LPat$ は、対象物の局所的な部分情報であり、部分画像パターンのマッチングにおいて得られる情報は、対象物の部分的な位置情報でしかない。その部分的な位置情報から、対象物の全体の存在領域を計算するためには、部分パターンと検出対象領域との相対的な位置関係の記述が必要となる。

図 26(b) は図 25(g) と同じ登録情報を表し、図 26 の例で部分-全体の関係づけを説明する。モデル画像 (図 26(a)) から得られた図 26(b) の登録情報 M_j は次のように表す。

$$M_j = \{p_1, p_2, p_3, LPat, GPat, RelGp, Label\} \quad (9)$$

p_1, p_2, p_3 は基底ベクトルを構成する3点の座標位置、 $LPat$ は $M \times M$ pixel の正規化部分画像、 $GPat$ は検出対象領域 Gp の部分パターンを表す $M \times M$ pixel の正規化画像である。 $Label$ は物体の種類、カテゴリなどを表す付帯情報で、複数種類の物体を検出を行う際に、どの物体であるかというラベルを与える。 $RelGp$ は、検出対象領域 Gp を表す四

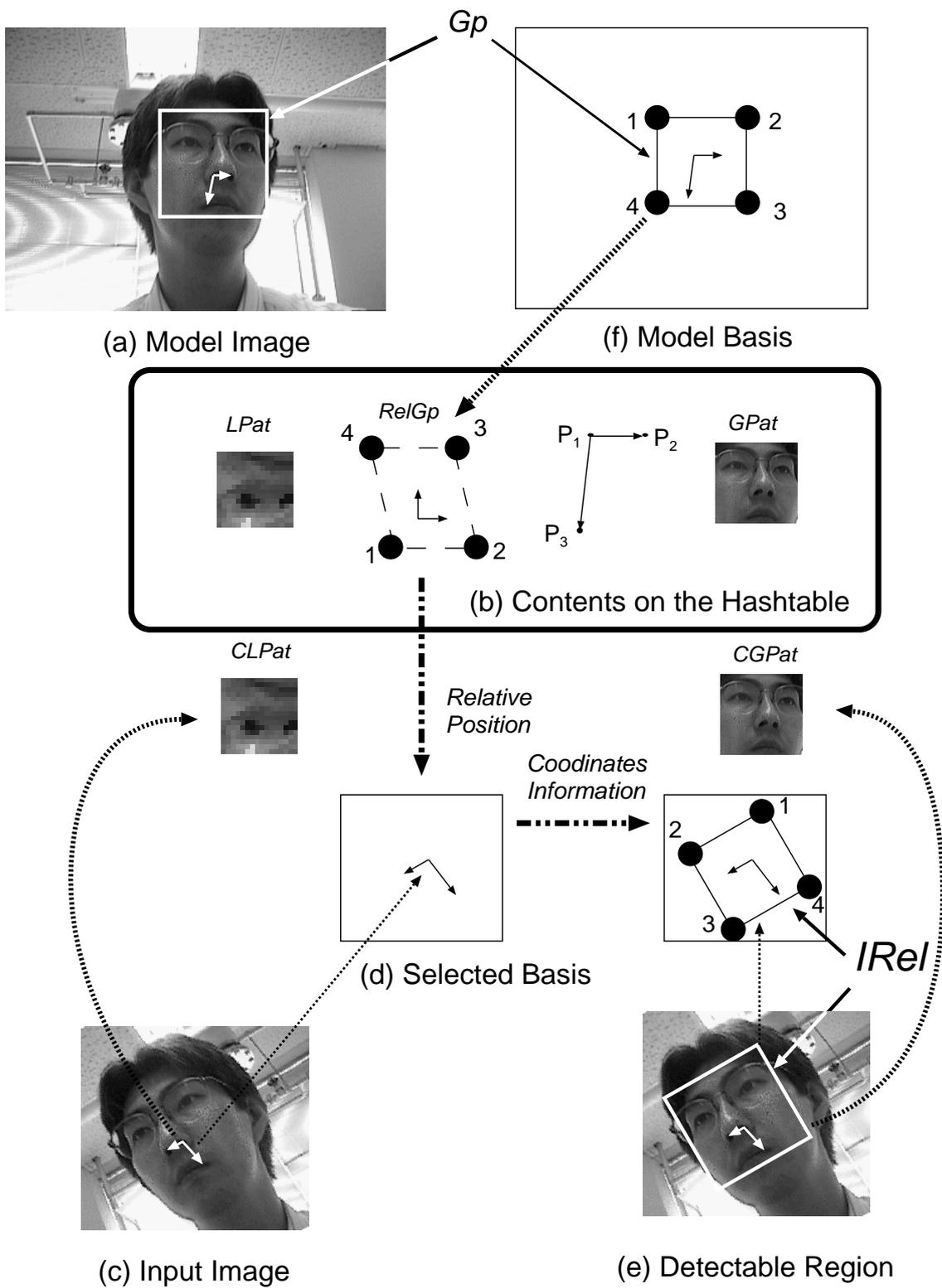


図 26: 部分-全体関係の記述

辺形の4点の座標について、図26(f)の座標系を正規直交系にした場合の相対座標で記述したものであり、対象物の全体の位置を求める際に次のように用いる。図26(c)の入力画像の3点の特徴点を選択し、基底ベクトルを構成したとする。その部分画像パターン $CLPat$ がモデル部分画像パターン $LPat$ とマッチした場合、入力画像上での部分パターンの位置が求まる。そこで、図26(d)の選択された基底ベクトルで張られる座標系を用いて $RelGp$ の表す4点の座標位置を変換する。これにより入力画像(図26(c))における各4点の座標が決定し、図26(e)に示すように、対象物の領域 Gp に対応する物体全体の領域位置 $IRel$ が求まる。この領域 $IRel$ を用いて、入力画像中の検出対象領域のパターン $CGPat$ を抽出でき、 $GPat$ と比較することで、全体的な領域の整合性のパターンマッチングによるチェックが可能となる。

3.3 パターンハッシングによる認識

3.3.1 特徴点検出とモデル選択

次に、分散アピアランスモデルにより表現された物体に対して、複数の部分パターンを用いたパターンハッシングによる認識方法を説明する。図27は、その認識方法の流れを示しており、本図の説明では、部分的に隠れている物体を画像から検出する例を示している。

認識の方法は、最初に入力画像(図27中の(1))に対して特徴点検出を行う(図27中の(2))。これらの特徴点は基底ベクトルを構成するために使われる。一般に隠れやノイズが存在するので、これらの点は、必ずしもモデル登録で与えられた着目点をすべて含んでいる必要はない。モデル登録と同様に、3つの特徴点 p_{1k}, p_{2k}, p_{3k} の組み合わせが選択され(図27中の(3))、各基底ベクトルの不変量が計算される。不変量 θ_k, R_k により、マッチングの対象となるモデルをハッシュテーブルから選択し(図27中の(4))、それぞれのモデルパターンとマッチングを行う。

3.3.2 マッチングのための類似性測度

候補として抽出される部分画像パターンは登録されたモデルの部分画像パターンと比較される。マッチングのための類似性測度には、正規化相関や残差の2乗和(SSD)がよく使われる。これらの測度は、線形濃度変換、ヒストグラム平坦化などの濃度変換処理により、類似度の値が影響を受けてしまう。本研究では、マッチング測度として、定性的3値表現(Qualitative Trinary Representation: 以下QTR)を導入する[29]。この表現

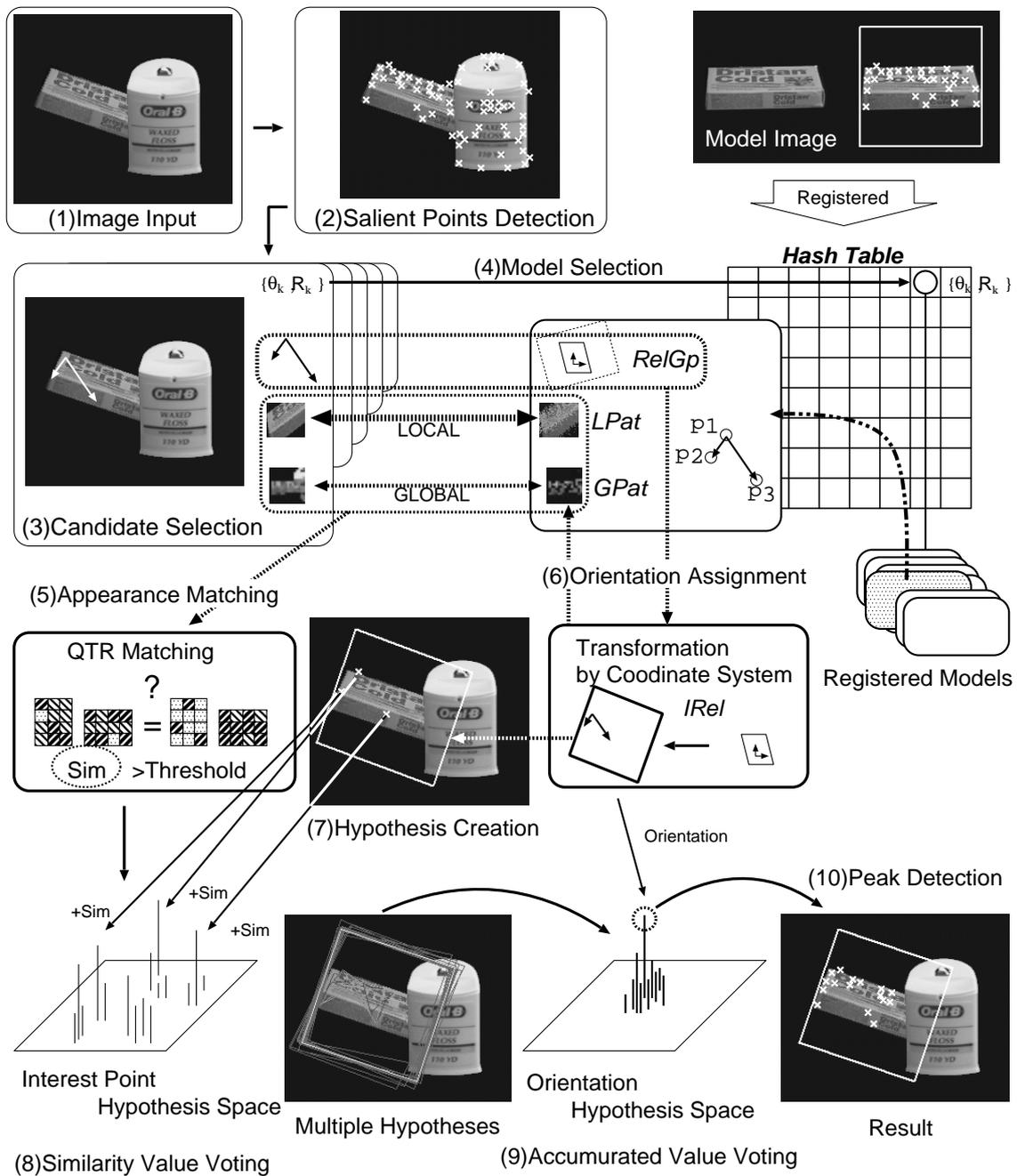


図 27: パターンハッシングによる認識処理の流れ

では、隣接画素値の大小関係、同値関係を利用しており、一般の類似性測度とは異なる性質の類似度が定義される。

QTR は、画像 I_s に対して、2つの3値画像 Q_h, Q_v で定義する。

$$QTR(I_s) = \{Q_h, Q_v\} \quad (10)$$

なお、 Q_h, Q_v は以下の式で定義される。

$$Q_h(x, y) = \text{sgn}(I_s(x, y) - I_s(x + 1, y))$$

$$(0 \leq x \leq M - 2, 0 \leq y \leq M - 1)$$

$$Q_v(x, y) = \text{sgn}(I_s(x, y) - I_s(x, y + 1))$$

$$(0 \leq x \leq M - 1, 0 \leq y \leq M - 2)$$

$I_s(x, y)$ は画像 (x, y) での輝度値、 $M \times M$ pixel の画像が与えられた場合、2つの3値画像 Q_h, Q_v は、 $M - 1 \times M$ pixel 画像と $M \times M - 1$ pixel 画像となる。なお、 sgn 関数は、

$$\text{sgn}(x) = \begin{cases} 1 : x > 0 \\ 0 : x = 0 \\ -1 : x < 0 \end{cases}$$

である。

図 27 中の (5) に3値画像の例を示す。2つの画像 G_1, G_2 の類似度 $QSim(G_1, G_2)$ は、3値画像のそれぞれの画素の一致する回数 D_h, D_v によって定義される。

$$QSim(G_1, G_2) = D_h + D_v \quad (11)$$

$$QTR(G_1) = \{Q_{h_1}, Q_{v_1}\}, QTR(G_2) = \{Q_{h_2}, Q_{v_2}\}$$

$$D_h = \text{Count}(Q_{h_1}, Q_{h_2}), D_v = \text{Count}(Q_{v_1}, Q_{v_2})$$

$$\text{Count}(Q_1, Q_2) = \sum_k \sum_l \begin{cases} 1 : Q_1(k, l) = Q_2(k, l) \\ 0 : Q_1(k, l) \neq Q_2(k, l) \end{cases}$$

この類似度計算には、乗算が必要ではなく、加減算と比較演算のみで構成され、高速な計算が可能である。また、部分的な輝度の変化、隠れに対してもロバストであり、輝度の単調増加変換に対しては、類似度は不変的な性質を持つ。

3.3.3 仮説の生成と投票による統合

登録モデル M_j が選択されると、対象となる3点 p_{1k}, p_{2k}, p_{3k} と $RelGp$ から、相対的な検出対象領域 I_{Rel} が計算され(図 27 中の (6))、部分画像パターン、 $CLPat_k, CGPat_k$ が求まる。この検索対象候補 C_k

$$C_k = \{p_{1k}, p_{2k}, p_{3k}, CLPat_k, CGPat_k, I_{Rel}\} \quad (12)$$

がマッチングの候補となる。

まず、検索対象候補 C_k と登録モデル M_j の部分画像パターン同士の類似度 L_{Sim}, G_{Sim} を求める (図 27 中の (5))。

$$L_{Sim} = QSim(CLPat_k, LPat_j) \quad (13)$$

$$G_{Sim} = QSim(CGPat_k, GPat_j) \quad (14)$$

それぞれの類似度が設定した値を越えた場合に、その場所の部分的な見え方を有する物体が存在することを仮定でき、これを仮説情報として生成する。仮説情報 H_l は、以下のように記述する。

$$H_l = \{p_{1k}, p_{2k}, p_{3k}, IRel, L_{Sim}, G_{Sim}, Label\} \quad (15)$$

ここで、 $IRel$ は物体の存在仮説領域として利用され、入力画像上での物体の位置、サイズ、方向を表す。複数の仮説が画像マッチングにより生成され、部分的な隠れがある場合でも、さまざまな見え方の可能性が仮説情報として生成され、投票による仮説の集積によって対象の位置を検出できる。また、 $Label$ は、物体の種類、カテゴリを表し、式 (9) の M_j から引用する。

図 27 中の (8)(9) は、仮説の投票と統合の様子を模式的に表している。図 27 中の (8) に示す着目点仮説空間は、各特徴点が着目点としてどの程度支持されているかを求めるために用意される。この空間の各セルは物体上の着目点候補と対応し、各仮説情報の 3 点の位置に対応するセルに対して、類似度値 L_{Sim} を投票する (図 27 中の (8))。ここで、ある特徴点が複数の領域の着目点としての可能性がある場合が存在する。例えば、図 28 の例のように、 \triangle で示した特徴点が実線、破線、双方の領域の特徴点となっている。このような場合を考慮して、同じ位置の特徴点であっても、異なる存在領域の場合には、区別して投票を集積しておく。

すべての仮説について投票を行った後、次に、着目点仮説空間において蓄積された類似度値を用いて、姿勢仮説空間への投票を行う。図 27 中の (9) に示す姿勢仮説空間は、画像中のどの位置に物体が存在するかを表現する。仮説情報の存在仮説領域 $IRel$ と対応する姿勢仮説空間のセルに対し、先の着目点仮説空間に蓄積された値を用いて投票を行う。これは、ある領域仮説に関連するすべての特徴点の投票値が集積されることになる。

姿勢仮説空間に集積された投票結果を評価し、物体の存在位置を求める。1 つの対象領域の位置のみが必要な場合は、各セルの投票を数え、もっとも大きな投票をもつ位置を結果とする (図 27 中の (10))。複数の候補を認識する場合は、姿勢仮説空間に投票されたピーク値の大きな順に評価を行い、部分的に重なっている領域の排他処理などの後処理を行う。例えば、図 28 では、複数の実線で示された領域仮説が、同じセルに投票され、

大きな投票値をもつこととなる。この場合、共有されていた点の投票値を、破線部分が示す領域仮説のセルから減じたのち、再評価を繰り返すこととなる。

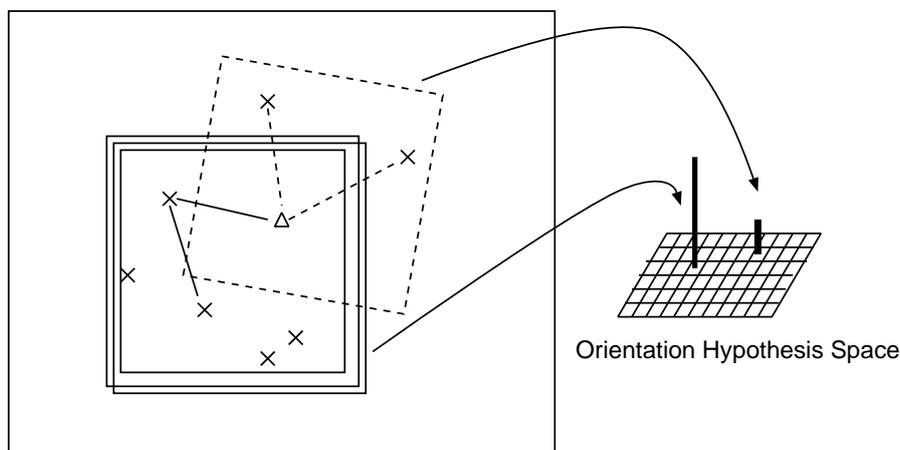


図 28: 複数の領域仮説の例

3.3.4 計算量とメモリ量

3点の組み合わせによって生成される部分モデルのマッチングに関する計算量についてまとめる。登録されているモデルの数を λ 、各モデルについての着目点の数を m 、画像中の特徴点の数を n とした場合、すべての部分モデル同士のマッチングを行う場合の回数は、 $O(\lambda m^3 n^3)$ となる。一方、本手法ではあらかじめモデルの登録がなされているため、認識時のマッチング回数は、ハッシュテーブルの一つの格納場所に登録されている部分モデルの平均数を p とすると、 $O(pn^3)$ となる。

次に、ハッシュテーブルのメモリ量について考察する。ハッシュテーブル内の格納場所の数は、各不変量の量子化サイズによって決定される。しかしながら、メモリ量は部分モデルの登録数によって大きく変化し、ハッシュテーブルにおける各部分モデルの数のオーダーは、 $O(\lambda m^3)$ となる。これに登録情報 M_j で示した構造体を格納するだけのメモリ量を乗じたメモリが必要となる。

3.4 アルゴリズムの検証実験

3.4.1 顔検出

提案法の検証のために、顔画像を利用した実験を行う。収集した顔データベースには 25 名分、950 枚の画像 (320×240 pixel) で、顔の向きをわずかに変えさせながら画像を取

集した。この画像を用いてモデルの登録を行った。各画像に対し特徴点検出を行い、瞳、鼻孔、口角の6点を、着目点として手動で選択して与えた。同時に、顔領域を囲む対象領域 G_p を設定した。特徴点の検出法には、分離度フィルタによる円形領域検出 [30] を利用した。また、部分画像マッチングのためのテンプレートの大きさは、 16×16 pixel とした。

950 枚の顔画像データを登録した場合のハッシュテーブルについて、部分モデルの登録数をグラフにした。図 29 が登録後の状態を示し、各軸は、不変量 θ_{axs} と R_{axs} を表している。この量子化の刻み幅は、角度 θ_{axs} 方向は1度刻み、長さ比 R_{axs} 方向は0.1刻みとした。これらの刻み幅は検出精度にも関連し、今後検討が必要である。登録モデルの冗長性を考えて、長さ比 R_{axs} において、分子 \geq 分母となる組み合わせのみを登録した。登録された部分モデル M_j の数は、100195 個であった。また、一つ以上の登録がある格納場所に関しての平均登録数は約 32 個であった。各部分モデルが不変量による索引によって、テーブル上に分散されている様子がわかる。

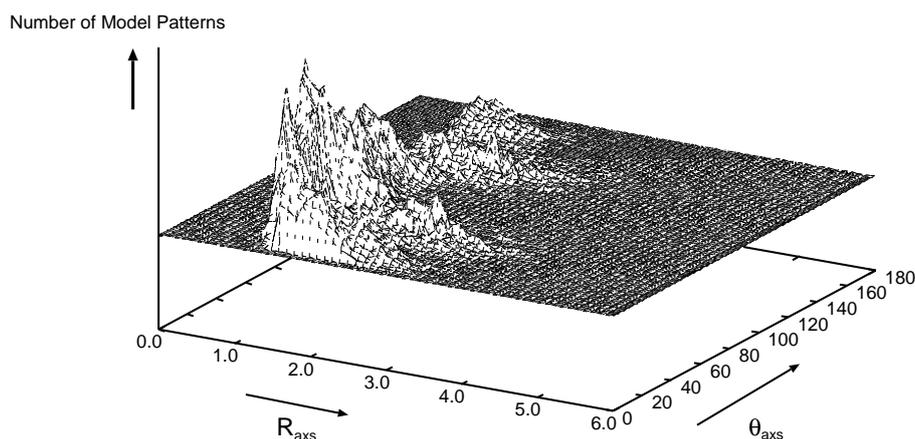


図 29: ハッシュテーブル上の部分モデルの登録数

図 30 は、複数人の正面顔を同時に検出している様子を示したものである。画像平面中での回転したいくつかの顔の領域と、それらの顔のいくつかの着目点を検出している。図 30 左上は、部分的な隠れとなる眼帯をしている人物を含む画像例を示す。図 30 左中の特徴点（円形領域）の候補数は74個であり、3点の組み合わせによる基底ベクトル候補の総数は、388944個となる。図 30 右上の四辺形は、それぞれのマッチングにおいて、類似性が高いと判断された位置候補の $IRel$ を示し、線分はそれらの基底ベクトルを表している。複数の見え方を統合した結果、3つの顔領域を正しく検出している。また、左下のように大きさの異なる対象の場合や、右下のように部分的な隠れをもつ逆さ画像の場合でも検出できている。

試作したプログラムは、Windows XP 上で、Visual C++6.0 で実装され、Pentium4 2.0GHz を CPU とする PC で動作させた。図 30 左上の画像を処理するために、約 4 秒の時間を要した。



図 30: 顔認識の例. 左上: 入力画像、右上: 計算された複数の仮説、左中: 円領域検出の結果、右中: 検出された顔領域と顔特徴点、下段: その他の例

3.4.2 複数種類の物体検出

次に、複数種類のモデルを同時に検出する実験を行った。実験には、Columbia 大学物体画像データベース (COIL)[31] の画像を利用し、あらかじめ 10 個のオブジェクトを選び、ある一方向の見え方の画像のみを使用し、ハッシュテーブルに登録した。特徴点検出には Harris のコーナー検出を用いている。図 31 は、部分的な隠れや回転した 5 つの

物体の合成画像から、登録されている3つの物体を正しく検出することを示している。このとき、式(15)に示した、*Label*を用いて、物体の種類を特定することが可能となる。

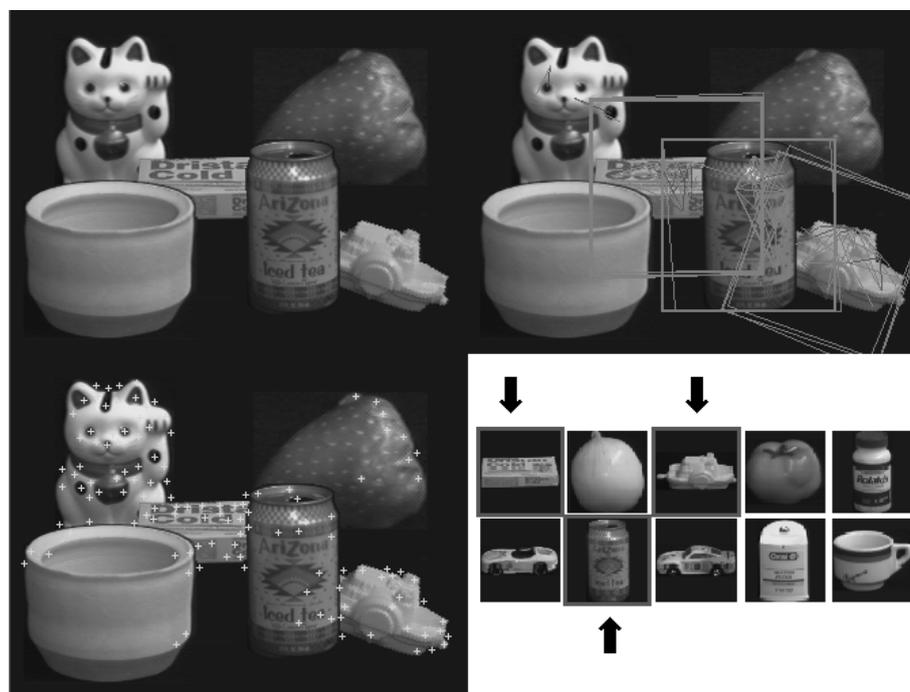


図 31: 複数種類のモデルの検出: 左上: 入力画像、右上: 計算された複数の仮説、左下: 特徴点検出の結果、右下: 登録されている10個の物体と見つかった物体(矢印)

3.5 本章のまとめ

本章では、第二の視点に立って、幾何学的特徴と部分画像特徴の両者を融合したパターンマッチング法について考察を行い、部分パターンのマッチングに基づく新しい物体認識のモデルの表現法と検出方法である“パターンハッシング”を提案した。

本手法は、幾何学的不変量を用いることにより、画像の変形の影響を受けず、また不変量計算の際に求めた変形を適用して部分画像のマッチング計算を行うことにより、画像の部分的隠れに対しても影響を受けにくい手法を実現した。

また、本手法は従来法である Geometric Hashing の拡張となっており、これまで点や線の集合に対してしか使えなかった手法を一般の画像パターンに適用できるように拡張が行ったという観点での位置づけができる。

顔領域の検出や物体データベースを用いた物体検出について、パターンハッシングによる認識実験を行い、任意の位置、サイズ、方向の複数の物体を安定に認識できることで有用性を確認した。

4 特徴抽出の多重化による線形部分空間法の高精度化

本章では、第1章で述べた第三の視点から、複雑なパターンの変化を取り扱ってきた伝統的なアルゴリズムである線形相互部分空間法を、さらに進化させることを目指す。従来から取り組んでいる相互部分空間法について、識別に適した不変的な特徴を取り出すために、特徴抽出を組み合わせることで高精度化を行っている。多種多様な変化が複合的に起こるデータに対して、よりよい特徴量を取り出すための考察を行う。

4.1 線形相互部分空間法の拡張

未知のパターンが入力されたとき、どのクラスに属するかを識別するパターン認識において、認識を精度よく行うための手法の一つとして部分空間法 [32] がある。部分空間法では、1つの入力パターンに対して、クラス毎に登録されているパターンから構成した部分空間（辞書）との類似度を比較して識別を行う。

部分空間法の利点として、分類したいカテゴリごとに辞書が用意されており、その辞書のメンテナンス（更新や削除）を行う際に、そのカテゴリだけの処理を行えばよいということがある。カテゴリ数が非常に多くなった場合には、これらの性質は非常に重要であり、実用性の高い方法として有効に機能する。

この部分空間法を拡張した相互部分空間法が提案されている [8]。相互部分空間法では、獲得した複数の入力パターンと、クラス毎に登録されている参照パターンとの類似度を比較する。複数枚の入力パターンから入力部分空間を生成し、同様に複数枚の参照パターンに対してクラスごとの参照部分空間を求めておき、それぞれの部分空間の間の正準角を利用して類似度を算出する。相互部分空間法は、登録時の画像集合と認識時の画像集合の双方に対して主成分分析によって、変動の少ない部分を抽出することとなり、データ中の不変量を求めることに相当している。

さらに相互部分空間法の前段に特徴抽出を行う方法として、特徴抽出の方法の違いにより、制約相互部分空間法 (CMSM)[10][33]、さらには、非線形主成分分析を用いたカーネル非線形制約相互部分空間法 (KCMSM)[34] やカーネル直交制約相互部分空間法 (KOMSM)[35] も提案され、複数視点からの物体認識においての高い識別性能を得ている。図 32 に系譜を記す。

これらの特徴抽出は、さらにカテゴリの識別精度を向上させるための、各種の変動に不変な特徴を取り出すものとして機能している。よって、特徴抽出を経緯する相互部分空間法は、これらの二つの不変性が導入されていると考えることができる。

しかし、非線形化による拡張については、そのカーネルの再計算に膨大な時間を要することから、頻繁にカテゴリの追加、更新が行われる応用については、十分とはいえない

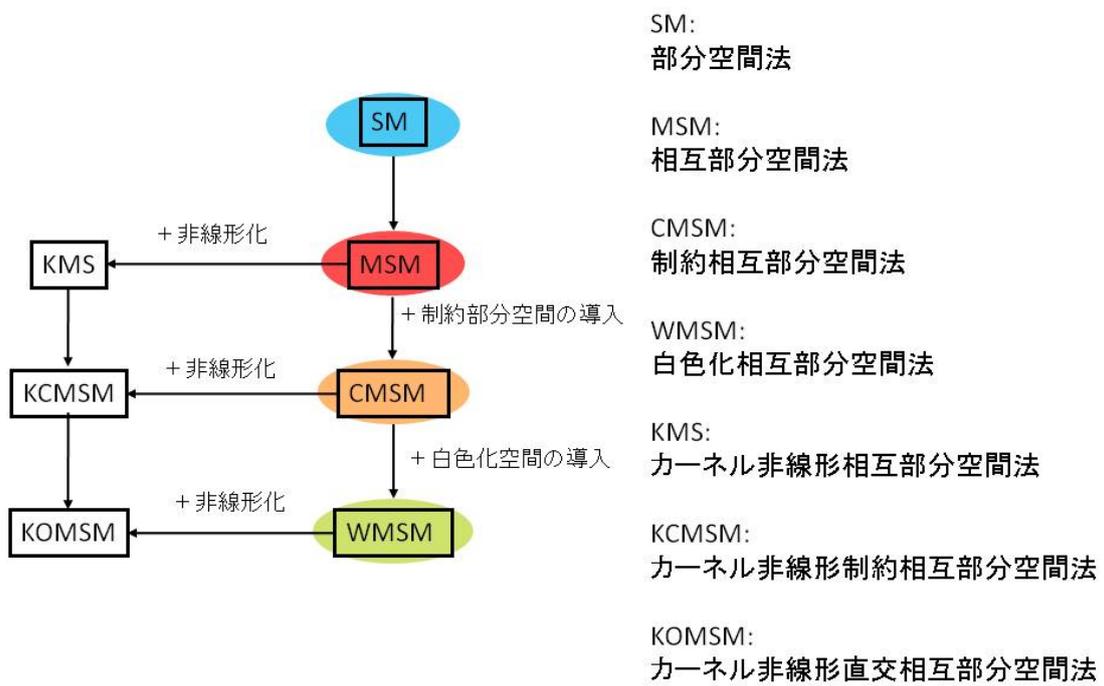


図 32: 相互部分空間法の非線形化による拡張の系譜

い。実用上の観点から計算量の軽減が図られる線形部分空間法を採用し、その拡張方法を本章では検討する。図 33 にあるように、線形線形相互部分空間法の拡張を行い、(1) 従来の制約相互部分空間法の多重化による拡張 (2) 疑似白色化変換の導入、(3) さらに多重化による性能向上、について検討を行う。

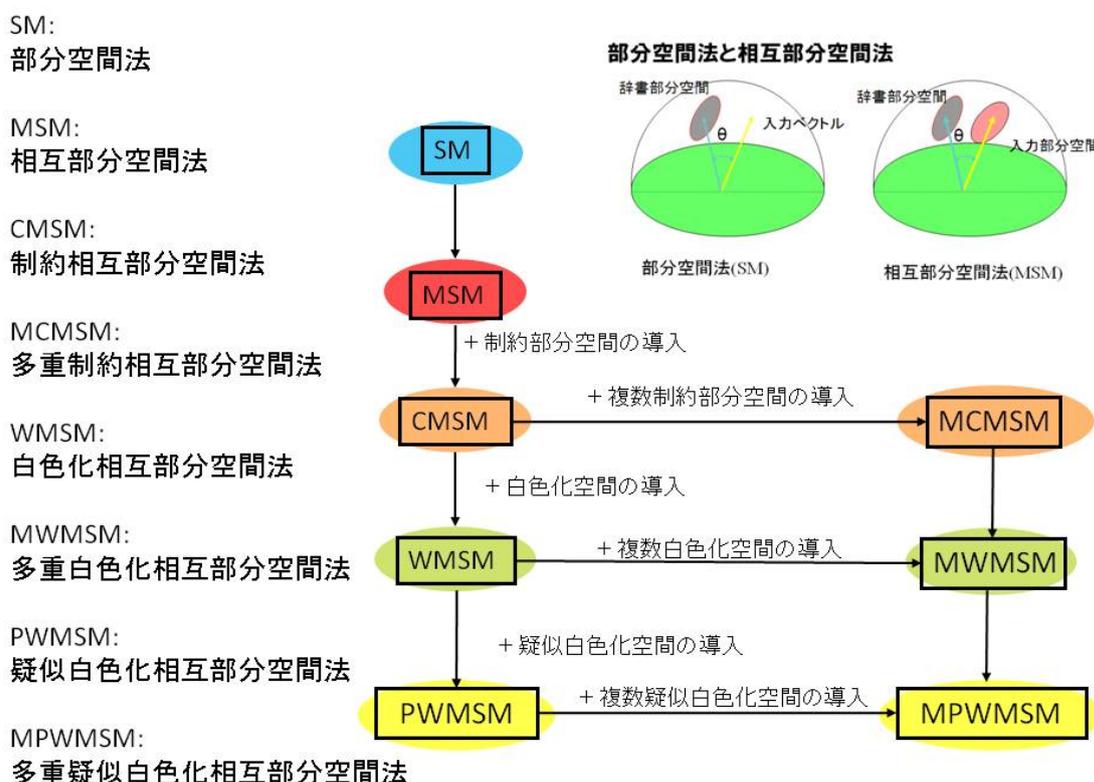


図 33: 線形相互部分空間法の拡張

以下、顔画像認識を具体的なタスクとして、制約相互部分空間法の適用方法について述べ、4.3 で従来の制約部分空間の学習方法について説明する。4.4 で複数の制約部分空間の生成に対してアンサンブル学習を適用する方法について述べる。次に、4.5 で識別時における多重制約相互部分空間法の適用方法について述べる。4.6 で照明変動が生じる環境で撮影されたデータベースと多人数が登録されたデータベースを用いた識別実験で提案手法の有効性を確認する。さらに、4.7 では、カテゴリ間の無相関化を行う変換である白色化変換を導入し、その白色化変換を導入した場合の問題点について述べ、その解決方法として疑似白色化変換の導入法について述べる。そして、4.8 では、識別実験により提案手法の有効性を確認する。

4.2 顔画像認識

顔画像を用いた個人認証は、生体情報をシステムに非接触で入力できるため、ユーザの利便性が高く、ユーザの心理的な負担が少ない [36]。顔画像により個人認証を行う際に、同一人物内における顔の見え方の変化 (パターン変動) は誤認識を引き起こす大きな要因となる。同一人物内のパターン変動は、照明条件、顔向き、表情、眼鏡などの装飾品、髪型などに依存し極めて多様である。このような変動が生じる環境の中で精度よく個人認証を行うために、従来我々は制約相互部分空間法 [10, 37, 33] を提案している。この手法を用いることで実環境でも頑健な個人認証システムを実現することができる [38, 39]。

制約相互部分空間法では、人物毎に参照部分空間を登録しておき、識別対象から獲得した複数枚のパターンから生成した入力部分空間と参照部分空間の類似度を比較する。類似度は、部分空間法 [40] を拡張した相互部分空間法 [8] により部分空間同士のなす角度として求める。識別に有効な特徴を抽出するために、相互部分空間法の前処理として、入力部分空間と参照部分空間を制約部分空間と呼ばれる特徴空間へ射影する (図 34)。制約部分空間へ射影することで、識別に有効と考えられる人物間の差異が強調される。特徴空間への射影による特徴抽出の手法は他にも、主成分分析で求めた固有空間を用いる Eigenfaces [41]、固有空間とそれに直交する補空間を同時に用いる手法 [42]、線形判別分析で求めた判別空間を用いる Fisherfaces [43]、同一人物内の変動を抑えることを目的とした摂動特徴空間を用いる手法 [44] などが提案されており、個人認証の精度向上に有効であることが確認されている。

制約相互部分空間法において、様々な実環境で不特定多数の人物を精度良く識別するためには、制約部分空間の生成に用いる学習パターンに実環境で生じる可能性のある全ての変動を含める必要がある。ところが変動は極めて多様であり全ての変動を獲得するためには、膨大な人数を対象として照明条件などを多様に変化させて撮影しなければならない。しかし、このような大量の学習パターンを収集し制約部分空間を学習することは実用上困難である。

この問題に対応するために、複数の識別器を組み合わせることで少数の学習パターンでも識別精度が向上するアンサンブル学習 [45, 46, 47, 48, 49] に着目する。本章では、制約部分空間の生成に対してアンサンブル学習を適用し、複数の制約部分空間を用いて特徴抽出を多重化する手法を提案する。本章で提案する手法を多重制約相互部分空間法 (Multiple Constrained Mutual Subspace Method : MCMSM) と名付ける。多重制約相互部分空間法では、図 35 のように、入力部分空間と参照部分空間を複数の制約部分空間へ射影し、それぞれの制約部分空間上において入力部分空間と参照部分空間のなす角度を類似度として求める。制約部分空間の個数だけ得られた類似度を結合することで最終的に類似度を決定する。複数の制約部分空間の生成に対してアンサンブル学習を適用す

るために、Bagging[46] に代表される並列的な学習と、AdaBoost[47] に代表される逐次的な学習の考え方を利用する。提案手法により、従来の制約相互部分空間法に比べて高い識別精度が実現できる。

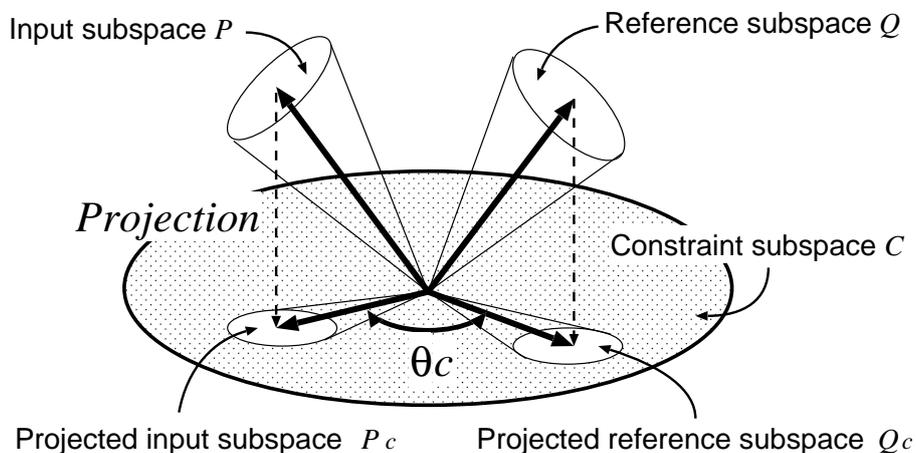


図 34: 制約相互部分空間法 の概念図

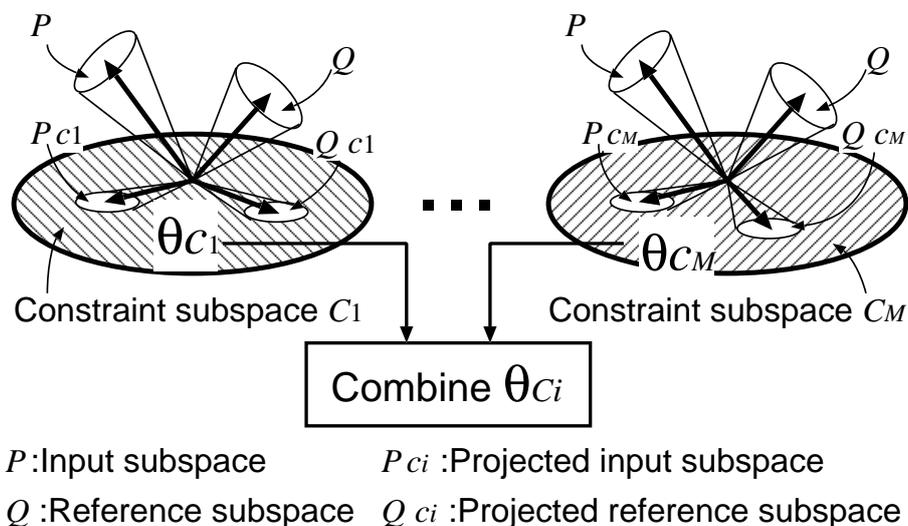


図 35: 多重制約相互部分空間法 の概念図

4.3 制約部分空間の学習

制約相互部分空間法における制約部分空間の学習アルゴリズムについて述べる。まず、同一人物内のパターン変動を学習するために、照明条件や顔向きなどを変化させ撮影し

たパターンを複数人 (L 人) について用意し、人物毎に主成分分析を適用する (図 36)。以下では、それぞれの人物について生成した部分空間を学習用部分空間と呼ぶ。学習用部分空間の固有値が大きいほうの基底ベクトルは、同一人物内の主要なパターン変動を表していると考えられる。

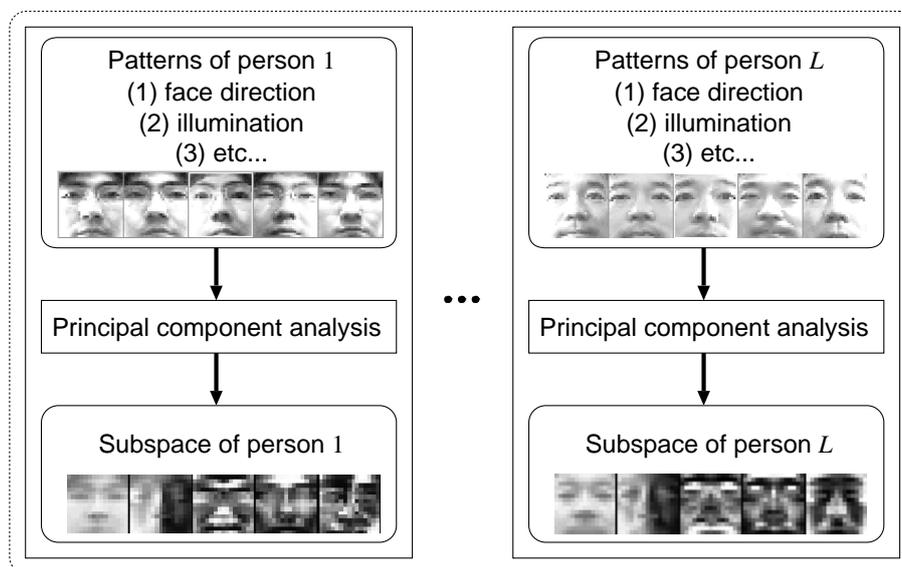


図 36: 学習用部分空間の生成

次に、制約部分空間へ学習用部分空間を射影することで学習用部分空間同士のなす角度が大きくなるように、制約部分空間の基底ベクトルを求める。これにより、人物間の差異が強調され識別に有効な特徴抽出を行うことができる。制約部分空間の生成方法の一つとして、 L 個の学習用部分空間が形成する和空間から一般化差分部分空間を求める手法が提案されている [33]。一般化差分部分空間の生成による制約部分空間の学習の流れを図 37 に示す。一般化差分部分空間は、それぞれの人物についての学習用部分空間の射影行列を $\mathbf{P}_j (j = 1 \dots L)$ とし、人物毎の学習用部分空間の基底ベクトルを $\psi_{jk} (k = 1 \dots N_B)$ とすると、式 (16) の固有値問題より得ることができる。

$$(\mathbf{P}_1 + \mathbf{P}_2 + \dots + \mathbf{P}_L)\mathbf{a} = \lambda\mathbf{a} \quad (16)$$

$$\mathbf{P}_j = \sum_{k=1}^{N_B} \psi_{jk}\psi_{jk}^T \quad (17)$$

固有値が小さい方から選んだ N_C 本の基底ベクトルで張られる空間が一般化差分部分空間となる。なお、 N_B, N_C は実験的に定める。得られた一般化差分部分空間を制約部分空間として用いる。

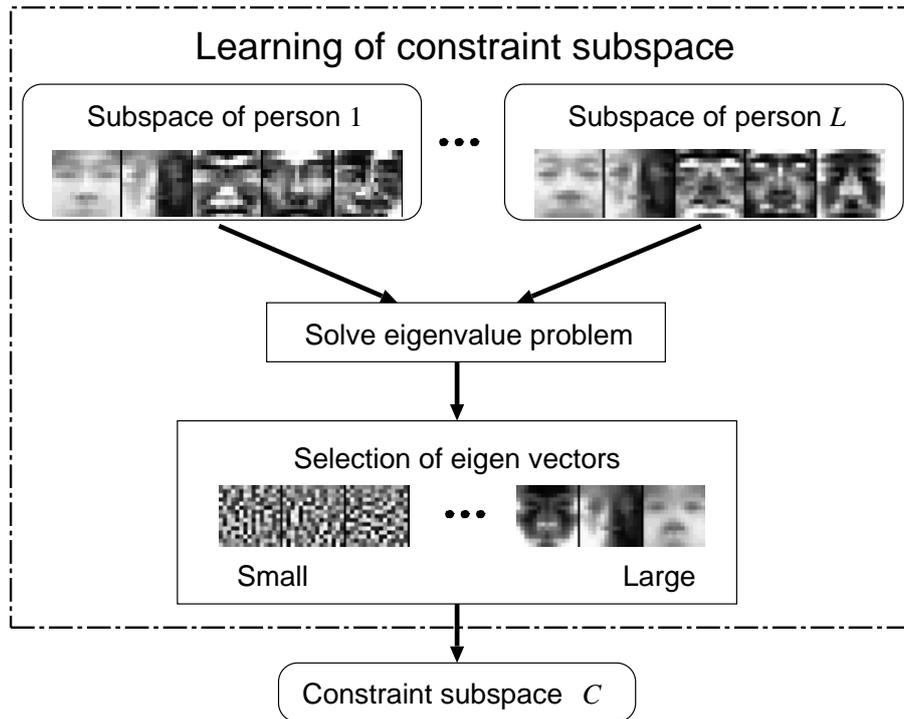


図 37: 制約部分空間の学習の手続き

4.4 複数の制約部分空間の生成に対するアンサンブル学習

4.3 で述べた制約部分空間の学習方法に基づいて複数の制約部分空間を学習する方法を提案する。以下では、並列的な学習と逐次的な学習の二つの方法について説明する。

4.4.1 並列的な学習

複数の制約部分空間を生成するために、アンサンブル学習の一つの手法である Bagging[46] の考え方をを用いる。Bagging は、複数の識別器を用意し、それぞれの識別器から得られた結果を結合することで認識を行う。それぞれの識別器は、全ての学習パターンからランダムにサンプリングされた少数の学習パターンを用いて生成される。学習パターンの選択にランダム性があるため、異なる複数の識別器が生成される。Bagging の考え方を適用するには何をランダムにサンプリングするかを決める必要がある。ここでは、 L 個の学習用部分空間から $L' (< L)$ 個の学習用部分空間をランダムにサンプリングし、複数の制約部分空間を生成する。以下に、 M 個の制約部分空間を生成する手続きを示す。

1. L 個の学習用部分空間から重複しないようランダムに L' 個の学習用部分空間を選択する。

2. L' 個の学習用部分空間の基底ベクトルを用いて 4.3 の手続きで 1 個の制約部分空間を生成する。
3. M 個の制約部分空間が生成されるまで (i) に戻る。

図 38 に制約部分空間の並列的な学習の流れを示す。なお、 L' は実験的に定める。

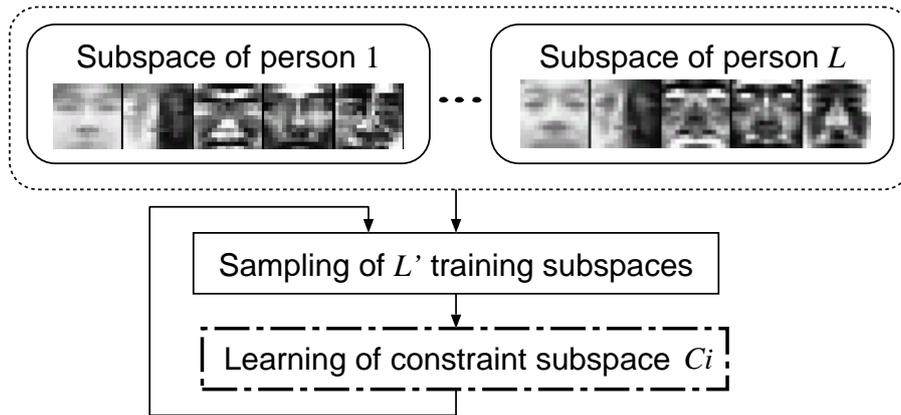


図 38: 並列的な学習の手続き

4.4.2 逐次的な学習

学習に用いる重みの決め方

並列的な学習のように学習パターンを部分的に利用するのではなく、学習パターン全てを利用する逐次的な学習について述べる。逐次的に複数の制約部分空間を生成するために、AdaBoost [47] の考え方をを用いる。AdaBoost は、Bagging と同様に複数の識別器から得られた結果を結合し認識を行う。それぞれの識別器は、Bagging のようにランダムに選択された少数の学習パターンを用いるのではなく、重み付けされた全ての学習パターンを用いて生成される。重みは 1 個の識別器が生成される毎に更新され、新たな識別器で誤識別された学習パターンに対して大きな重みを与える。

逐次的に制約部分空間を学習する場合、何に対してどのように重みを与えるかを考える必要がある。ここで、制約部分空間上での学習用部分空間同士の類似度を考える。図 39(a) のように、制約部分空間 C_i へ射影された学習用部分空間 P_{1C_i} と P_{3C_i} のなす角度が小さい場合、これらは類似しており識別誤りが発生する可能性が増える。そこで、 C_i 上でなす角度の小さかった学習用部分空間に対して大きな重みを与え次の制約部分空間 C_{i+1} を生成することで、図 39(b) のように、学習用部分空間 $P_{1C_{i+1}}$ と $P_{3C_{i+1}}$ のなす角度

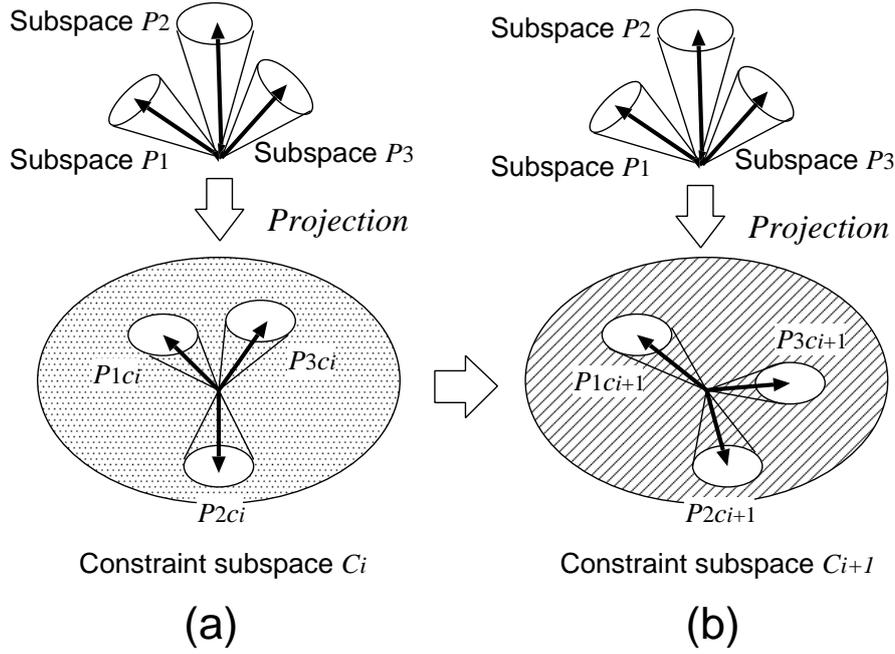


図 39: 逐次的な学習の概念図

を大きくすることを狙う。これにより、制約部分空間 C_i で識別誤りが発生しやすかった部分空間に対しては、制約部分空間 C_{i+1} を用いることで識別精度の向上が期待できる。

逐次的な学習のアルゴリズム

制約部分空間 C_i を学習する時に、 L 個の学習用部分空間 $P_j (j = 1 \dots L)$ に対して与える重みを $W_i(j)$ とすると、以下の手続きで M 個の制約部分空間が生成できる。

1. 重み $W_1(j)$ に初期値を与える。
2. 学習用部分空間の射影行列 $\mathbf{P}_j (j = 1 \dots L)$ に対して重み $W_i(j)$ を与え、式 (18) の固有値問題を解くことで制約部分空間 C_i を生成する。

$$(W_i(1)\mathbf{P}_1 + \dots + W_i(L)\mathbf{P}_L)\mathbf{a} = \lambda\mathbf{a} \quad (18)$$

3. 制約部分空間 C_i を用いてあらたな重み $W_{i+1}(j)$ を決定する。
4. M 個の制約部分空間が生成されるまで (ii) から (iii) を繰り返す。

図 40 に制約部分空間の逐次的な学習の流れを示す。各ステップの重み $W_{i+1}(j)$ は、以下の式で求めることとする。

$$W_{i+1}(j) = \frac{S'_j}{\sum_{j=1}^L S'_j} \quad (19)$$

$$S'_j = \sum_{j'=1, j' \neq j}^L \cos^2 \theta_{C_{ijj'}} \quad (20)$$

ここで、 $\theta_{C_{ijj'}}$ は、学習用部分空間 $P_j, P_{j'}$ を制約部分空間 C_i へ射影した後のなす角度を表す。部分空間がなす角度は後述 (4.5.3 節) する相互部分空間法で求まる。また、式 (20) にしきい値 T 以上の $\cos^2 \theta_{C_{ijj'}}$ から総和をとるという条件を設けることで、類似性の高い学習部分空間のみから制約部分空間を生成することができる。

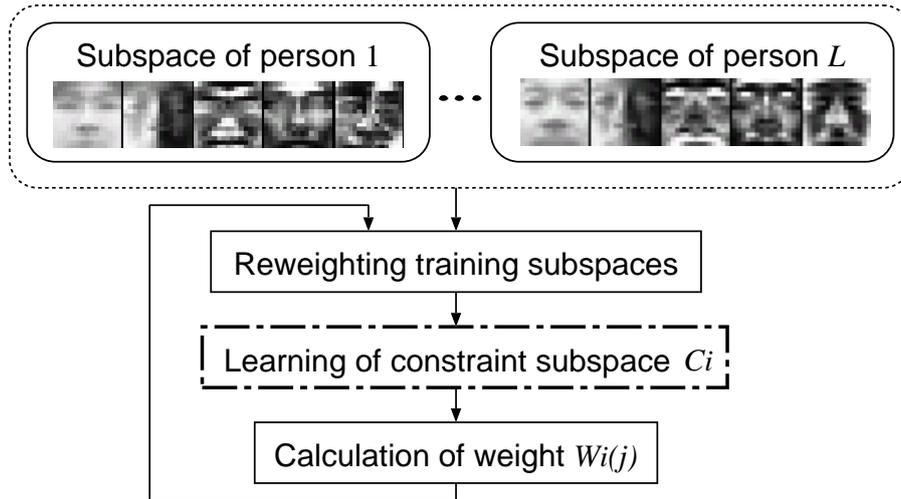


図 40: 逐次的な学習の手続き

4.5 多重制約相互部分空間法による識別方法

4.5.1 顔画像認識の流れ

多重制約相互部分空間法による顔画像認識の流れを図 41 に示す。まず、入力された画像から瞳と鼻孔の位置を基準として顔領域のパターンを切り出し、パターンをラスタースキャンすることで入力ベクトルに変換する。時系列に獲得された複数の入力ベクトルから、主成分分析により入力部分空間を生成する。主成分分析を適用する際は自己相関行列 [50] を用いる。次に、入力部分空間と予め登録されている参照部分空間を複数の制約部分空間に射影し、それぞれの制約部分空間上で類似度を算出する。得られた複数の類似度を結合し入力部分空間と参照部分空間の類似度を決定する。この類似度が最大になる参照部分空間に対応する人物を入力された人物と判定する。また、類似度がしきい値を下まわる場合は、本人と判定せず棄却する。4.5.2 節より、それぞれの処理について詳細を述べる。

4.5.2 制約部分空間への射影

特徴抽出のために、入力部分空間 P と参照部分空間 Q を M 個の制約部分空間 $C_i (i = 1, \dots, M)$ に射影する方法について述べる。

1. 入力部分空間を張る N 本の基底ベクトルを制約部分空間上へ射影し射影ベクトルを求める。
2. 各射影ベクトルの長さを正規化する。
3. N 本の正規化ベクトルに対して Gram-Schmidt の直交化を施す。

直交化された N 本の正規化ベクトルが射影された入力部分空間の基底ベクトルとなる。この手続きを用意した制約部分空間の個数 M だけ繰り返す。参照部分空間も同様の手続きで制約部分空間へ射影することができる。

4.5.3 部分空間同士の類似度算出

制約部分空間 C_i に射影された入力部分空間を P_{C_i} とし、参照部分空間を Q_{C_i} とする。 P_{C_i} と Q_{C_i} 間の類似度 S_{C_i} は、相互部分空間法 [8] により得られる正準角と呼ばれる二つの部分空間がなす角度 θ_{C_i} により式 (21) で決定される。

$$S_{C_i} = \cos^2 \theta_{C_i} \quad (21)$$

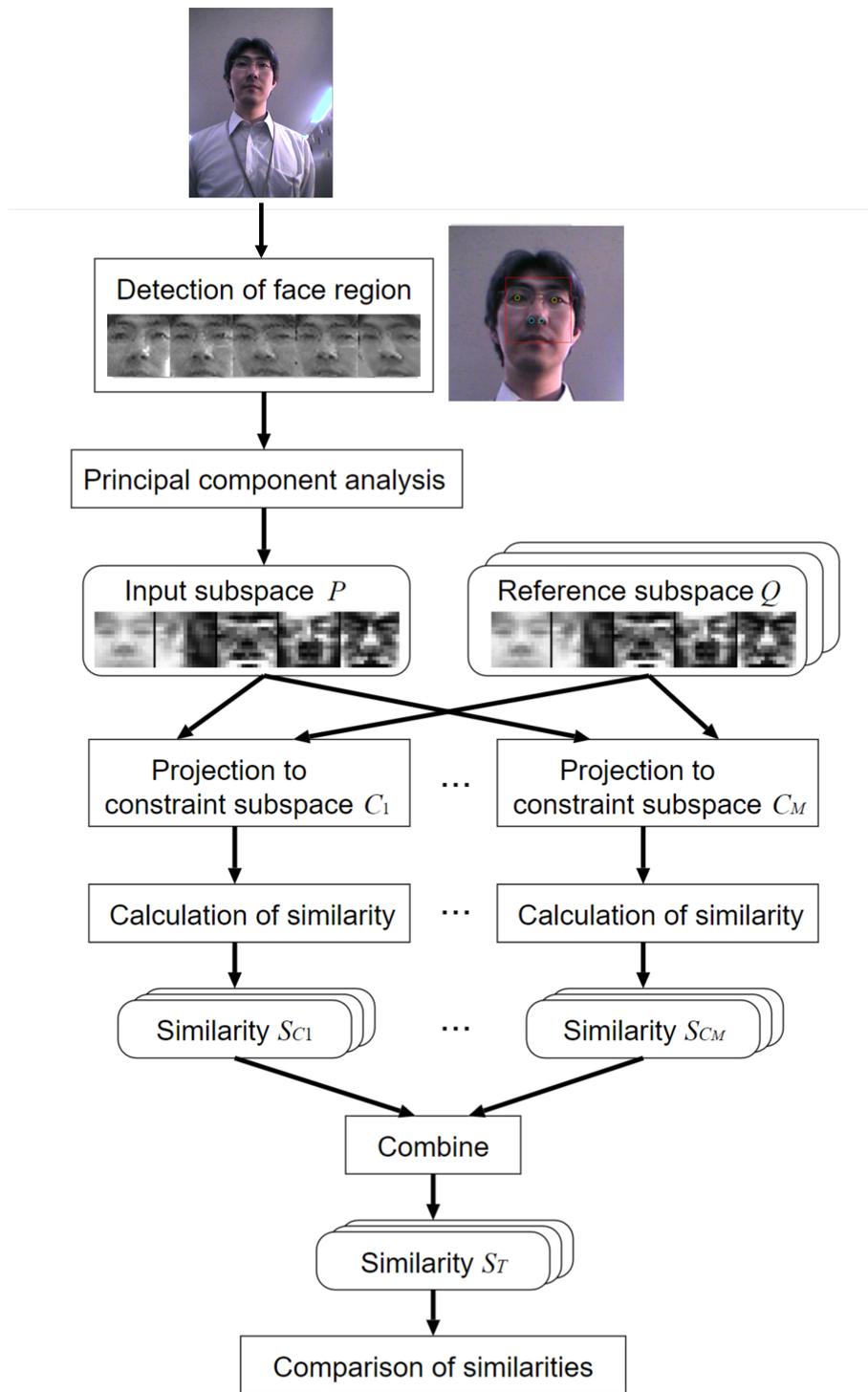


図 41: 多重制約相互部分空間法を用いた顔画像認識の手続き

部分空間同士が完全に一致していれば $\theta_{C_i} = 0$ である。 $\cos^2 \theta_{C_i}$ は、以下の行列 \mathbf{X} の最大固有値となる。

$$\mathbf{X}\mathbf{a} = \lambda\mathbf{a} \quad (22)$$

$$\mathbf{X} = (x_{mn}) \quad (m, n = 1 \dots N) \quad (23)$$

$$x_{mn} = \sum_{l=1}^N (\psi_m, \phi_l)(\phi_l, \psi_n) \quad (24)$$

ここで、 ψ_m, ϕ_l は部分空間 P_{C_i}, Q_{C_i} の m, l 番目の基底ベクトル、 (ψ_m, ϕ_l) は ψ_m と ϕ_l の内積、 N は部分空間の基底ベクトルの本数を表す。

4.5.4 類似度の結合

複数の制約部分空間上で得られた類似度を結合するために、ここでは平均値を用いる。多重制約相互部分空間法の類似度 S_T は式 (25) で決定される。

$$S_T = \frac{1}{M} \sum_{i=1}^M S_{C_i} \quad (25)$$

ここで、 M は射影する制約部分空間の個数、 S_{C_i} は制約部分空間 C_i 上での入力部分空間 P_{C_i} と参照部分空間 Q_{C_i} の類似度を表す。

4.6 多重制約相互部分空間法の実験

表 7: 実験条件

識別率の評価用画像の撮影人数	25 人
制約部分空間の学習用画像の撮影人数	25 人
照明条件 (1 人につき)	10 条件
各照明条件での試行数 (1 人につき)	7 回
1 試行に用いる入力パターン枚数	10 枚
各照明条件での参照パターン枚数	70 枚

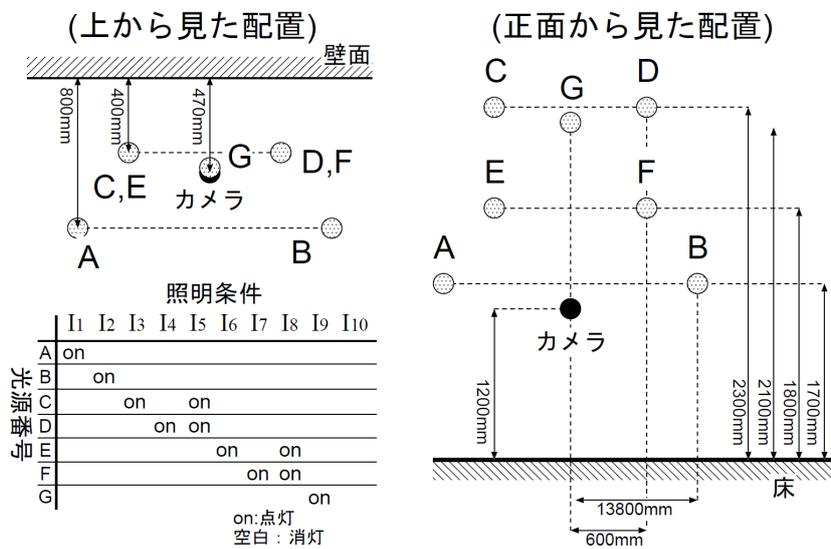


図 42: 光源とカメラの配置

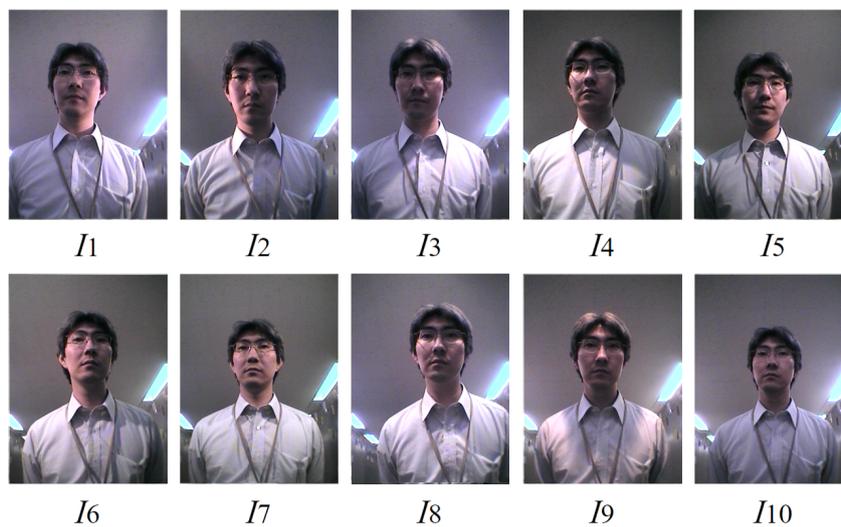


図 43: 各照明条件における撮影画像の例



図 44: 各照明条件におけるパターンの例

4.6.1 照明変動に対する識別性能

実験の仕様

照明変動に対する識別性能を評価するために、登録時と試行時の照明条件を変えて撮影した顔画像による識別実験を行った。撮影人数は50人としている。実験の諸条件については、表7に示している。光源とカメラの配置については、屋内に7個の光源(ハロゲンランプ)を図42のように設定し、光源の点灯と消灯を組み合わせることで10種類の照明条件 I_1 から I_{10} を設定した。なお、天井の蛍光灯は点灯したままで撮影した。図43に、それぞれの照明条件で取得された画像を示す。各照明条件で、1人につき1個の動画像を秒間5フレームで撮影した。動画像は28秒の間に撮影した140枚の画像からなる。被写体の顔向きと立ち位置は、瞳と鼻孔の左右両方がカメラで観測できる範囲で自由に変化させた。顔のパターンは瞳と鼻孔の位置を基準として 30×30 pixels の大きさで各画像から抽出した。各照明条件におけるパターンの例を図44に示す。パターンに対してヒストグラム平坦化を施し、 15×15 pixels にダウンサンプリングし、縦方向の微分オペレータを適用した後、 $15 \times (15 - 1) = 210$ 次元のベクトルに変換しベクトル長の正規化を施した。

識別実験に用いる入力パターンと参照パターンに含まれる人物が、制約部分空間の学習パターンに含まれる人物と異なるように、50人のパターンを25人と25人の二つに分けた。これは、識別対象となる人物のパターン変動(ここでは主に照明変動)を事前に獲得できない状況を想定し、識別対象ではない別人物のパターン変動から識別に有効な制約部分空間を学習できるかを評価することを目的とした。各照明条件の動画像から抽出された140枚のパターンのうち、前半の70枚を参照パターンに、後半の70枚を入力パターンとした。入力部分空間を生成するためには複数のパターンが必要であるため、10

枚の入力パターンで1個の入力部分空間を生成した。以下では、1個の入力部分空間による入力を1回の試行とする。参照部分空間は照明条件毎に70枚の参照パターンから生成した。各照明条件で1人につき7個の入力部分空間を生成し、25人で合計17500回の試行を行った。図45に、照明条件 I_{10} の入力部分空間を生成するための入力パターンと、そのパターンが抽出された画像の一部を示す。撮影画像はカラー画像であるが、処理のために濃淡画像に変換し、両瞳と鼻孔を自動検出し、正規化したパターンを生成している。顔向きを変化させながら撮影したため、鼻の見え方が変化するなどのパターン変動が生じる。被写体の動きがほとんどなく同じパターンが入力された場合、少ない基底ベクトルの本数で累積寄与率は高くなるが、本実験で用いた入力部分空間では8本で99%を超えた。図46に累積寄与率の変化を調べた結果を示す。図中の値は25人の入力部分空間からそれぞれ算出した累積寄与率の平均値を表す。この結果より、顔向きなどを動かすことで、短時間で撮影した10枚の入力パターンに拡がりをもった分布を提供できることを確認した。

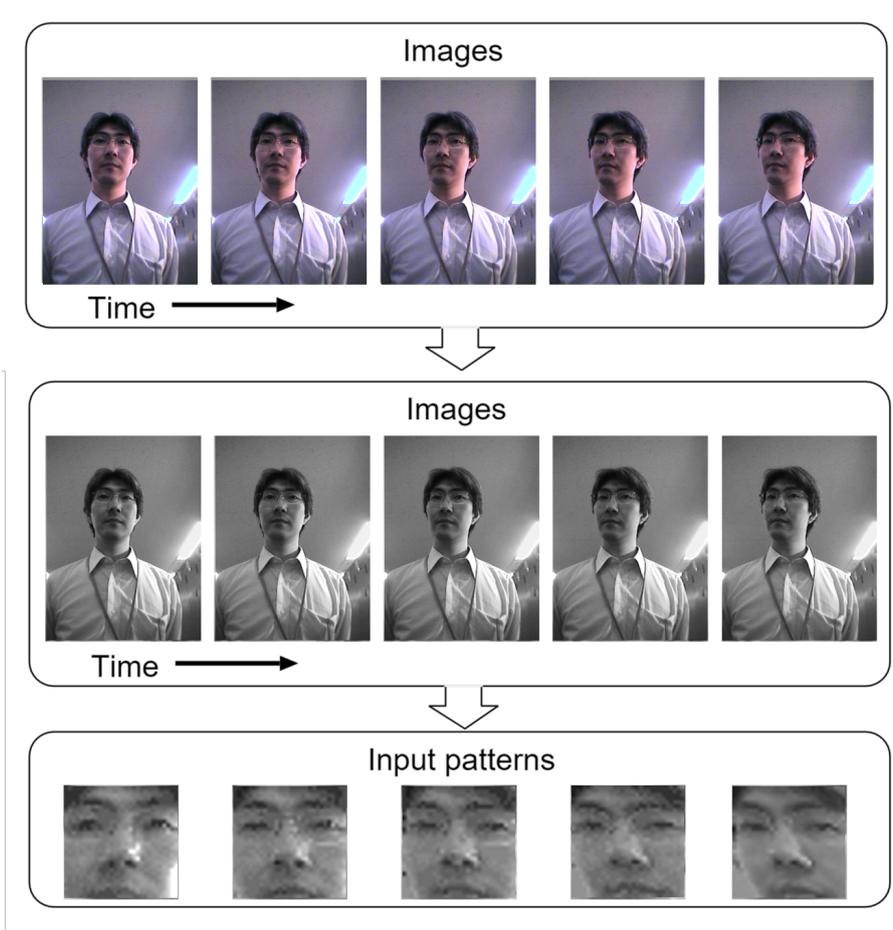


図 45: 入力部分空間を生成するためのパターンの例

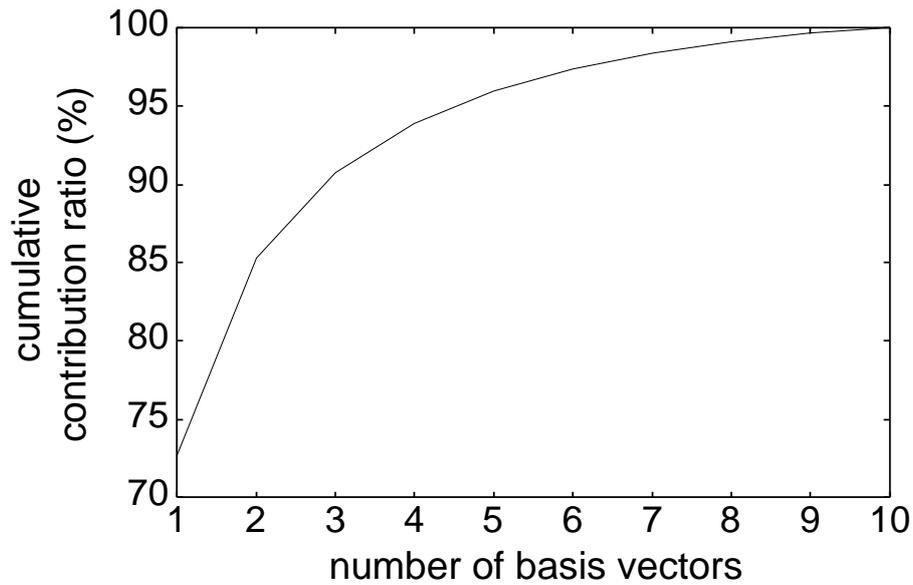


図 46: 入力部分空間の基底ベクトルに対する累積寄与率

制約部分空間の学習パターンは各照明条件で1人につき140枚とした。学習用部分空間は、照明条件によるパターン変動を十分に学習できるように、全照明条件の学習パターンを用いて25個生成した。

従来手法との比較

表 8: 照明変動に対する実験結果

	手法	ER(%)	EER(%)
(a)	最近傍決定則	4.6	23.9
(b)	部分空間法	4.6	12.9
(c)	相互部分空間法	4.6	9.8
(d)	カーネル非線形部分空間法	3.2	10.2
(e)	核非線形相互部分空間法	3.8	8.9
(f)	制約相互部分空間法	4.6	5.0
(g)	多重制約相互部分空間法 (並列)	1.8	4.0
(h)	多重制約相互部分空間法 (逐次)	1.4	3.9

提案手法の有効性を確認するために、識別性能について従来手法と比較実験を行った。それぞれの手法のパラメータは以下のように設定した。

(a) 最近傍決定則 (NN)

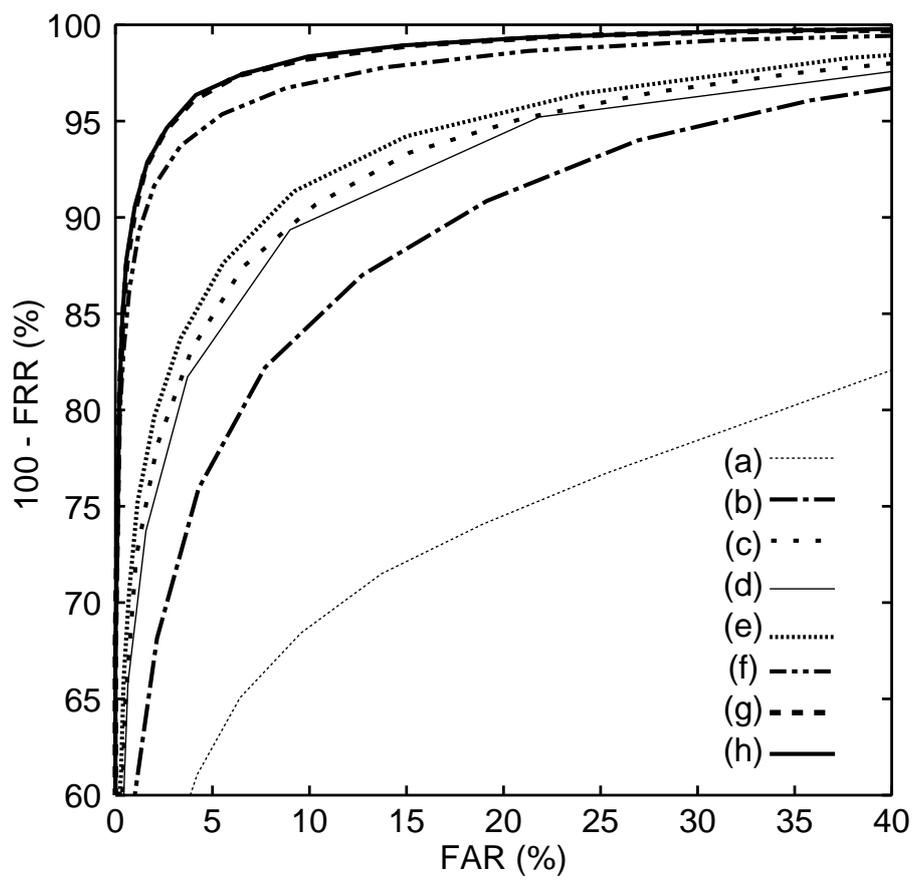


图 47: ROC 曲线

参照パターンを全て記憶しておき、入力パターンと最も類似する参照パターンとの距離を類似度とした。他手法と条件を揃えるため 1 回の試行で 10 枚の入力パターンを用い、10 枚の中で最小となる距離を 1 回の試行における類似度とした。

(b) 部分空間法 (SM)

入力パターンと参照部分空間がなす角度を類似度とした。参照部分空間の基底ベクトルの本数は 40 とした。他手法と条件を揃えるため 1 回の試行で 10 枚の入力パターンを用いた。それぞれの入力パターン毎に類似度を算出し、10 枚の平均値を 1 回の試行における類似度とした。

(c) 相互部分空間法 (MSM)

入力部分空間と参照部分空間とのなす角度 θ の $\cos^2 \theta$ を類似度とした。入力部分空間と参照部分空間の基底ベクトルの本数 N はともに 7 とした。

(d) カーネル非線形部分空間法 (KSM)

特徴空間上で非線形に分布するパターンを取り扱うために、カーネル関数を介して特徴空間を無限次元の空間に拡張した手法 [51] である。カーネル関数は $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\rho^2)$ を用いた。 \mathbf{x} は入力ベクトル、 \mathbf{y} は参照ベクトルを表す。 $\rho = 0.4$ とし、参照部分空間の基底ベクトルの本数は 40 とした。他手法と条件を揃えるため 1 回の試行で 10 枚の入力パターンを用いた。それぞれの入力パターン毎に類似度を算出し、10 枚の平均値を 1 回の試行における類似度とした。

(e) 核非線形相互部分空間法 (KMS)

相互部分空間法をカーネル関数を用いて非線形に拡張した手法 [52] である。(d) と同じカーネル関数を用い $\rho = 0.4$ とした。入力部分空間と参照部分空間の基底ベクトルの本数はともに 7 とした。

(f) 制約相互部分空間法 (CMSM)

単一の制約部分空間で特徴抽出を行い相互部分空間法により類似度を算出した。入力部分空間と参照部分空間の基底ベクトルの本数 N はともに 7 とした。制約部分空間は、 $L = 25$ 人分全ての学習用部分空間から生成し、その基底ベクトルの本数 N_C は 170 とした。学習用部分空間の基底ベクトルの本数 N_B は 30 とした。

(g) 多重制約相互部分空間法 (並列的な学習)(MCMSM-P)

制約部分空間を並列的に学習し多重制約相互部分空間法により類似度を算出した。 $L = 25$ 人分の学習用部分空間からランダムに $L' = 8$ 個を選択し、合計 $M = 10$ 個の制約部分空間を生成した。この学習方法では L' 個の学習用部分空間の選択にランダム性があるために、一連の手続きを 10 回行い、そのつどに後述する ER と EER を算出し平均値を求めた。 N, N_B, N_C は (f) と同じ値を用いた。

(h) 多重制約相互部分空間法 (逐次的な学習)(MCMSM-S)

制約部分空間を逐次的に学習し多重制約相互部分空間法により類似度を算出した。 $L = 25$ 人分の学習用部分空間に対して逐次的に重みを変化させながら、合計 $M = 10$ 個の制約部分空間を生成した。重みの初期値は $W_1(j) = 1/25 (j = 1 \dots 25)$ とし、式 (20) にしきい値 T 以上の類似度から総和をとる条件を加えた。しきい値は $T = 3.5\sigma_i$ とし、 σ_i は制約部分空間 C_i を用いての算出した学習用部分空間同士の類似度の標準偏差とした。 N, N_B, N_C は (f) の CMSM と同じ値を用いた。

識別精度の評価には以下の2つの基準を用いた。

1. エラー率 (ER:Error Rate)

本人類似度よりも他人類似度の方が高くなる割合。本人類似度は入力部分空間と参照部分空間に対応する人物が同じ場合に算出された類似度、他人類似度は異なる場合に算出された類似度である。

2. 等価エラー率 (EER:Equal Error Rate)

FAR(他人受理誤り率) と FRR(本人排除誤り率) が等しい時の割合。FAR は以下の式で求まる。

$$FAR = \frac{\text{他人類似度がしきい値以上の試行数}}{\text{全試行数} - \text{本人の試行数}} \quad (26)$$

一方、FRR は以下の式で求まる。

$$FRR = \frac{\text{本人類似度がしきい値以下の試行数}}{\text{本人の試行数}} \quad (27)$$

顔認識システムの場合、登録されていない未知の人物に対応する必要がある。EER が低いとそのシステムは信頼性が高いといえる。

それぞれの基準についての評価結果を表8に示す。制約部分空間への射影により特徴抽出を行った (f)CMSM,(g)MCMSM-P,(h)MCMSM-S は、行わなかった (a)NN ,(b)SM, (c)MSM, (d)KSM, (e)KMS に比べて、大きく EER が改善されている。複数の制約部分空間により特徴抽出を多重化した (g)MCMSM-P と (h)MCMSM-S は、単一の制約部分空間のみを用いた (f)CMSM と比べて、EER,ER ともに改善されている。図 47 の ROC 曲線でも識別精度が改善されていることが確認できる。これらの結果より提案手法を適用することで、従来の認識手法に比べて識別精度が向上することが確認できた。提案手法は、非線形に拡張された手法 (d)KSM と (e)KMS に比べても性能が優れていた。この理由として、(d)KSM と (e)KMS には制約部分空間のように人物間の差異を強調する特徴抽出が導入されていないことや、評価に用いたパターンには非線形な手法で得意とする顔向き変動の影響が少なかったことが考えられる。線形な手法の組み合わせで、計算

量が爆発的に増加するカーネル関数を用いることなく識別性能の向上が見込めることは、実時間で計算できる点などシステムの観点から意味がある。

次に、制約部分空間の学習パターンに含まれる人物と識別実験に用いるパターンに含まれる人物が同じ場合の従来法と、含まれる人物が異なる場合の提案手法を比較する。含まれる人物を同じにするために制約部分空間の学習パターンとして、入力パターンと参照パターンを用いた。この場合の制約相互部分空間法を (f) CMSM とすると、ER は 4.6% で EER は 4.8% であった。表 8 の (f)CMSM と比べて識別精度は向上したが、学習パターンに含まれない人物から学習した提案手法の (g)MCMSM-P や (h) MCMSM-S の方が識別精度が高い。このことから提案手法は、生じる可能性のある変動成分の全てが学習パターンに含まれていなくとも、従来法と比べて識別に有効な特徴を抽出できるといえる。

制約部分空間の個数についての評価

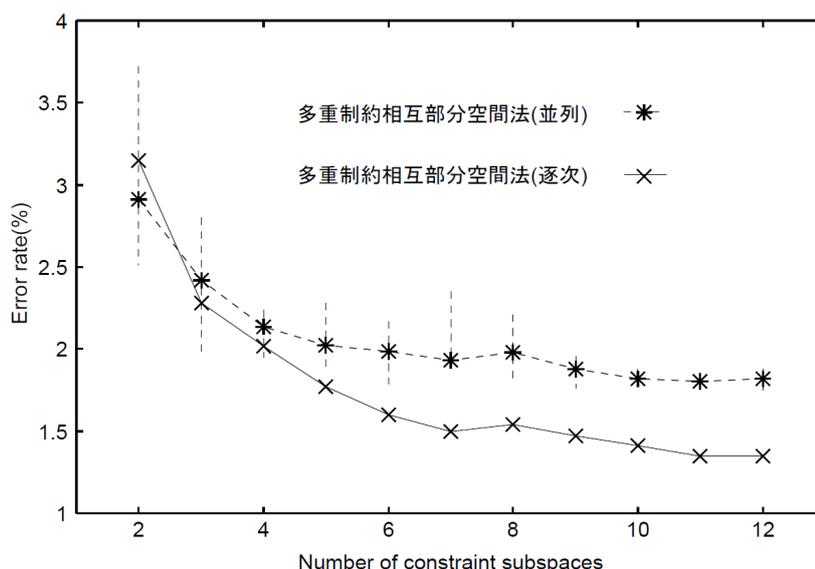


図 48: 制約部分空間の個数を増加させた時の ER

提案手法において制約部分空間の個数と識別性能の関係を調べた。制約部分空間の個数 M を 2 個から 12 個まで増加させたときの ER の変化を図 48 に、EER の変化を図 49 に示す。図中の点線は並列的な学習による評価を 10 回行いその平均値を結んだものであり、点線に付属する誤差棒は結果の最大値と最小値を表している。並列的な学習と逐次的な学習のどちらの学習方法でも、制約部分空間の個数が増加するにつれて ER と EER が改善された。制約部分空間の個数が 8 以下の場合、逐次的な学習では EER を大きく改善できたが、並列的な学習では EER にばらつきが生じた。これは並列的な学習の性能が

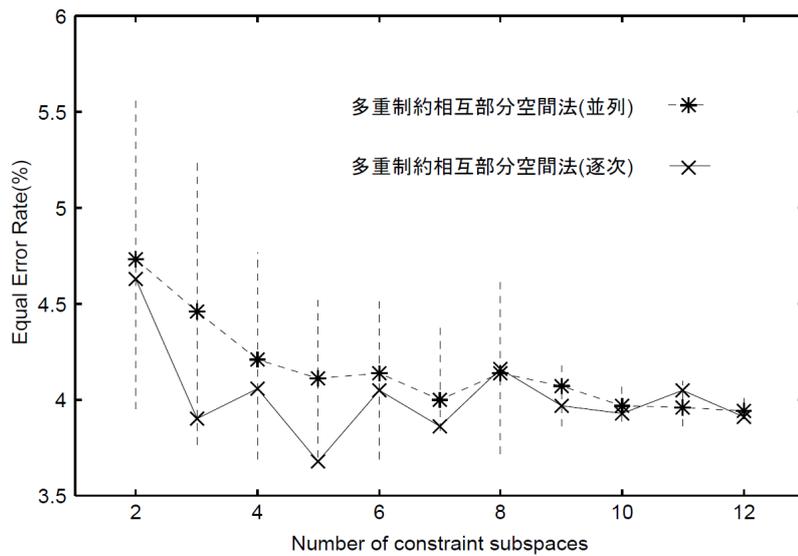


図 49: 制約部分空間の個数を増加させた時の EER

L' の選び方に依存することを示している。並列的な学習は、アルゴリズムが単純であるため実装が容易であり、逐次的な学習のように重み計算が不用であるため処理性能に限りがある場合には有効であると考えられる。メモリの使用量や計算性能に余裕がある場合には逐次的な学習が有効であると考えられる。

学習パターン数についての評価

制約部分空間の学習パターン数と識別性能の関係を実験的に明らかにする。以下では 4.6.1 の設定から学習パターン数を減らした場合について述べる。学習パターンに含まれる人数を半分の 12 人とした時の実験結果を表 9(I) に示す。提案手法の識別性能は、全学習パターンを用いた表 8(f) と比べて ER、EER ともに改善されていた。この結果より、提案手法は学習パターン数が少ない場合でも、学習パターン数が多い時の制約相互部分空間法と比べて高い識別精度を得ることができるといえる。

次に、学習パターンに含まれる照明条件を減らした時の結果を表 9(II) と (III) に示す。(II) では照明条件 I_1 から I_5 を学習パターンとし、(III) では照明条件 I_6 から I_{10} を学習パターンとした。学習パターンには含まれない照明条件で識別性能を評価するために、(I) では照明条件 I_6 から I_{10} で、(II) では照明条件 I_1 から I_5 で識別実験を行った。学習パターンに含まれない照明条件で評価を行っても提案手法が優れていることが確認できた。この結果より、提案手法は従来の制約相互部分空間法と比べて照明変動の影響を受けにくい特徴抽出を行うことができるといえる。

表 9: 学習パターン数を減らした時の評価結果

	Method	ER (%)	EER (%)
(I)	制約相互部分空間法	5.2	5.5
	多重制約相互部分空間法 (並列)	3.0	4.6
	多重制約相互部分空間法 (逐次)	3.1	4.9
(II)	制約相互部分空間法	3.3	2.8
	多重制約相互部分空間法 (並列)	1.3	1.8
	多重制約相互部分空間法 (逐次)	1.4	2.0
(III)	制約相互部分空間法	5.6	4.5
	多重制約相互部分空間法 (並列)	2.9	2.9
	多重制約相互部分空間法 (逐次)	2.2	3.3

4.6.2 多人数データベースにおける識別性能

表 10: 評価結果 (登録人数 500 人).

	Method	ER (%)	EER (%)
(i)	制約相互部分空間法	5.3	2.3
(ii)	多重制約相互部分空間法 (並列)	3.8	1.6
(iii)	多重制約相互部分空間法 (逐次)	3.2	1.6

次に、多人数のデータベースで評価するために 500 人の参照部分空間と 1000 個の入力部分空間を用意した。それぞれの部分空間を生成する際に用いたパターンは、照明条件がほぼ均一な環境で顔向きや立ち位置を不規則に変更しながら撮影した動画像より獲得した。入力ベクトルの次元数は 210 次元とし、入力部分空間と参照部分空間の基底ベクトルの本数 N はともに 7 とした。入力部分空間は 15 枚の入力パターン、参照部分空間は 125 枚の参照パターンから生成した。学習用部分空間は、各人物の参照パターンを用いて $L = 500$ 個生成し、その基底ベクトルの本数 N_B は 10 とした。(i) 制約相互部分空間法、(ii) 並列的な学習を用いた多重制約相互部分空間法、(iii) 逐次的な学習を用いた多重制約相互部分空間法について比較実験を行った。(i) では制約部分空間を $L = 500$ 人全ての学習用部分空間から生成し、その基底ベクトルの本数 N_C は 170 とした。(ii) では $L = 30$ 個から生成した $M = 10$ 個の制約部分空間を用いた。(iii) では重みの初期値を $W_1(j) = 1/500$ とし、式 (20) にしきい値 $T = 5\sigma_i$ 以上の類似度から総和をとる条件を加

えて生成した $M = 10$ 個の制約部分空間を用いた。

表 10 の結果より、登録人数が多人数となるデータベースにおいても従来の制約相互部分空間法より提案手法が優れていることが確認できた。

4.7 白色化変換の導入による拡張

文献 [35] において、3次元モデルを複数視点から撮影した公開データベース (ETH-80[53]) を用いた物体識別実験を行っている。図 4.7 のような類似した物体に対して、複数視点から撮影した物体の部分空間を求め、様々な相互部分空間法によって識別を行っている。表 11 の結果では、直交相互部分空間法 (OMSM)³ が、従来のいくつかの相互部分空間法よりも性能が劣るケースがみられる。白色化変換による特徴抽出を行う白色化相互部分空間法についてもオリジナルの相互部分空間法よりも性能が劣ることとなり、これは特徴抽出の意味を失っていることとなる。



図 50: ETH-80[53] を用いた物体識別実験

表 11: 文献 [54] での評価結果

	Accuracy (%)			
	S[1]	S[2]	S[3]	S[4]
MSM	72.7	73.7	76.3	74.3
CMSM-215	75.7	81.3	76.3	73.7
CMSM-200	73.3	81.0	79.3	77.7
CMSM-190	71.0	73.0	73.0	75.7
WMSM(OMSM)	51.3	54.0	56.0	54.0

MSM: 相互部分空間法

CMSM: 制約相互部分空間法 (ハイフンのあとの数字は制約部分空間の次元数)

OMSM: 直交相互部分空間法 (WMSM: 白色化相互部分空間法)

³この直交相互部分空間法は、白色化変換によるもので、本稿の WMSM と同じである。

そこで、白色化変換にて行われる特徴抽出について、その変換を求めている自己相関行列の固有値の分布の制御に着目して、従来法との比較によって、白色化変換に対する基底の特徴選択法について考察する。

本章では、まず、白色化相互部分空間法 [55] について概説する。

4.7.1 白色化相互部分空間法

各クラスの部分空間の差異を強調する処理として、部分空間集合の白色化変換について述べる。部分空間集合の白色化変換は、部分空間の間の角度が広がるほど小さくなる目的関数の最小値問題の近似解として定式化される。各クラスの d 次元部分空間の集合を V_1, \dots, V_R と V_i と V_j の d 個の正準角を $\theta_{ij}^{(1)}, \dots, \theta_{ij}^{(d)}$ とすると、正準角の余弦の 2 乗和について、以下の式が成り立つ。

$$\sum_{1 \leq i \leq R} \sum_{k=1}^d \cos^2 \theta_{ij}^{(k)} = C_1 \sigma^2 + C_2. \quad (28)$$

ただし、 C_1, C_2 は正の定数、 σ^2 は以下に定義する部分空間の集合の自己相関行列 G の固有値の分散である。

$$G = \frac{1}{R} \sum_{i=1}^R P_i \quad (29)$$

ただし、 P_i は、 V_i の基底 $\psi_{i1}, \dots, \psi_{id}$ で以下のように定義される射影行列 ($1 \leq i \leq R$) である。

$$P_i = \sum_{k=1}^d \psi_{ik} \psi_{ik}^T \quad (30)$$

この式から、部分空間集合の自己相関行列 G の固有値の分散が小さいほど、部分空間集合に属する部分空間の間の角度が広がることが分かる。そのため、以下で定義される、すべての固有値を 1 にする白色化変換 W によって、自己相関行列 G の固有値の分散が最小化される。

$$G = B \Lambda B^T. \quad (31)$$

$$W = \Lambda^{-1/2} B^T. \quad (32)$$

ただし、 Λ は自己相関行列 G の固有値の対角行列、 B はその固有ベクトルを縦に並べた行列とする。この白色化変換によって、部分空間の間の角度が広がるイメージを図 51 に示す。白色化変換によって、角度は均一に広がるため、間の角度が小さい部分空間の組ほど、大きく広がる。このため、白色化変換をあるクラスの部分空間へ適用すると、似ているクラス同士の部分空間ほど角度が広がることになり、似ているクラス間の差に注目し、差を強調する処理となる。

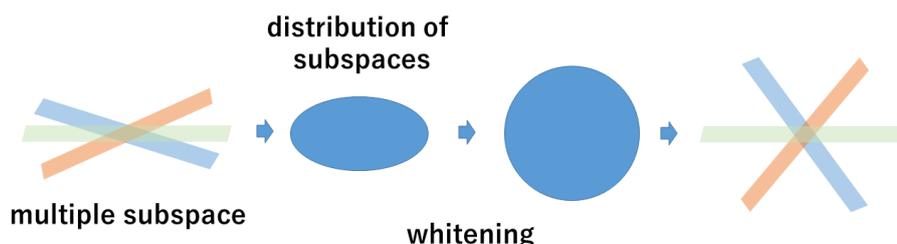


図 51: 白色化変換により、部分空間が広がる様子を示したイメージ図。両端は部分空間の集合を、楕円と円は変換前後の自己相関行列の固有値とその方向を表す。

4.7.2 解決すべき課題

各クラスの参照部分空間同士を類似させないようにするために、白色化相互部分空間法では、参照部分空間同士のなす角度が離れる特徴空間へ線形変換することで、識別精度を高める方法を提供した。しかし、識別したいクラス数が少ない場合や非線形性の強いパターンの場合では、白色化行列による変換によって、識別能力が向上しない場合があった。

4.7.3 解決手段の検討

制約相互部分空間法 (CMSM) ならびに白色化相互部分空間法 (WMSM) は、いずれも参照部分空間から生成した射影行列の自己相関行列 G に基づいて変換行列を求める手法である。射影行列 \mathbf{P}_i は、 ψ_{ij} を第 i カテゴリの参照部分空間の j 番目の正規直交基底ベクトル、 N_C を参照部分空間の基底ベクトルの本数とすると、式 (33) で定義される。

制約相互部分空間法 (CMSM) では、制約部分空間 \mathbf{O}_{CMSM} は、各カテゴリの射影行列を用いて、式 (35) により定義される。

$$\mathbf{P}_i = \sum_{j=1}^{N_C} \psi_{ij} \psi_{ij}^T \quad (33)$$

$$\mathbf{G} = \frac{1}{R}(\mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_R) \quad (34)$$

$$\mathbf{O}_{CMSM} = \sum_{k=1}^{N_B} \phi_k \phi_k^T \quad (35)$$

ここで、 R は参照部分空間の個数、 ϕ_k は行列 G の固有値が小さい方から選択した k 番目の固有ベクトル、 N_B は行列 G の固有ベクトルの本数を表す。

一方、白色化相互部分空間法 (WMSM) では、参照部分空間同士を白色化する変換行列 \mathbf{O}_{WMSM} は、式(36)で定義される。

$$\mathbf{O}_{WMSM} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{B}^T \quad (36)$$

先と同様に、 \mathbf{B} は \mathbf{G} の固有ベクトルを並べた行列、 $\mathbf{\Lambda}$ は \mathbf{G} の固有値による対角行列を表す。

この2つの方法は、同じ G に関する固有値問題を解くが、異なる要素を用いるように見える。しかし、求めた固有値の行列 $\mathbf{\Lambda}$ についての操作を行う方法と考えると次のように考えられる。 \mathbf{O}_{CMSM} は、以下の \mathbf{C}_p を用いて、

$$\mathbf{O}_{CMSM} = \mathbf{C}_p \mathbf{B}^T. \quad (37)$$

ここで、 \mathbf{C}_p は、

$$\mathbf{C}_p = \text{diag}(0, 0, 0, \dots, 0, 1, 1, 1, 1). \quad (38)$$

\mathbf{C}_p はランクが N_B の行列となる。

この固有値分布について例示してみる。図52は、それぞれの方法において固有ベクトルの重みをグラフにしたもので、制約相互部分空間法 (CMSM) では、ある次元以降のベクトルだけを利用することとなり、いずれも重みが1.0となる。白色化相互部分空間法 (WMSM) では、固有値の逆数の平方根となるため、高次元になるほど大きくなる。この図からも、制約相互部分空間法 (CMSM) で削減されている低次元の固有ベクトルは、白色化相互部分空間法 (WMSM) の固有値は小さな値となっており、逆に、高次元の固有ベクトルについては、大きな重みを与えられることになる。白色化相互部分空間法 (WMSM) では、高次元の固有ベクトルで抽出される成分が重要視されることによって、制約相互部分空間法 (CMSM) では起こらないようなクラス間の類似性が高くなる場合が起こり、このことが性能の低下を招いている可能性がある。

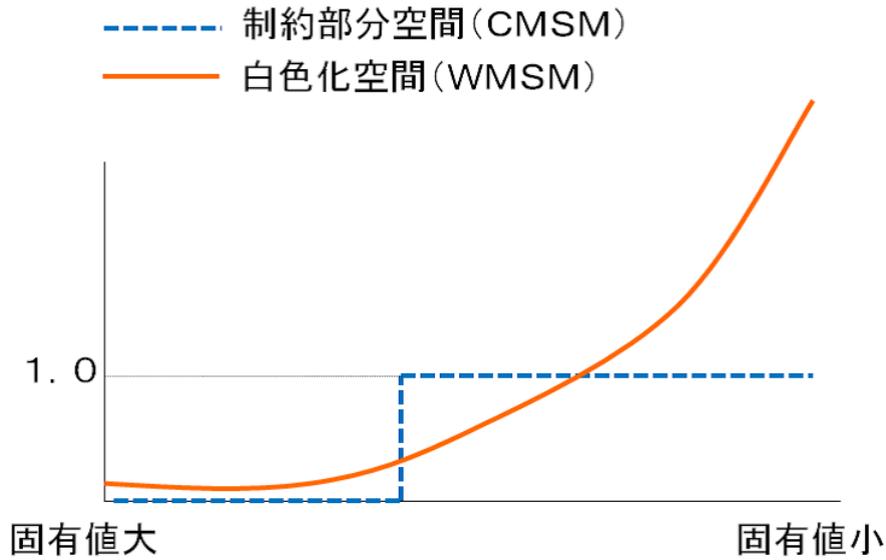


図 52: 制約部分空間と白色化変換の重み

4.7.4 固有スペクトル解析による観察

Fig. 53 は、式 (31) の G の固有値 Λ の降順の分布と、白色化変換の重み $(1/\sqrt{\lambda_k})$ を示している。指数が大きい白色化変換の成分はスケールリングファクターで大きく重み付けされている。

固有スペクトルについては、これまで [56] にて議論がなされている。この論文では、ノイズの乱れと学習サンプル数が有限であることによる小さな固有値の推定値の低さが原因であると指摘している。これは、学習サンプルの数が限られているため、ある次元の固有値が非常に小さくなり、その次元の真の分散をうまく表せなくなることがある。このため、その逆数を白色化変換の重みとして使用した場合、深刻な問題が発生する可能性があるとしている。

そこで、モデル式によって理想的な固有値の推定を行うこととし、本節では、[56] の推定式を用いる。2つのパラメータ α と β 、および推定式は式 (39)(40) と (41) で定義される。

$$\alpha = \frac{\lambda_1 \lambda_m (m-1)}{\lambda_1 - \lambda_m}, \quad (39)$$

$$\beta = \frac{m \lambda_m - \lambda_1}{\lambda_1 - \lambda_m}, \quad (40)$$

$$\hat{\lambda}_k = \frac{\alpha}{k + \beta}, \quad (41)$$

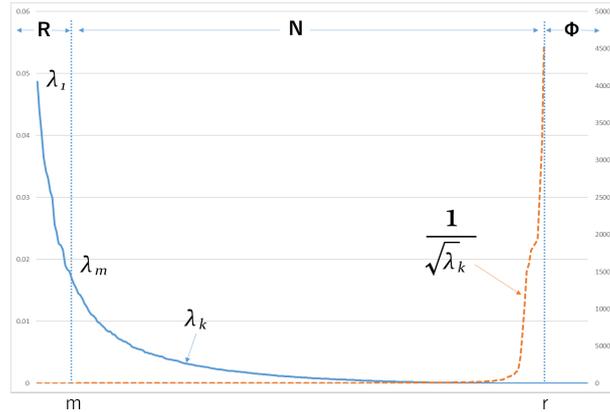


図 53: 固有値の降順の実際の分布（実線）と白色化変換の重み（破線）。 m は Reliable subspace (R) と Noise subspace (N) の分岐点、 r は Noise subspace と Null subspace(Φ) の分岐点である。

ここで、 λ_1 は最大固有値、 λ_m は分割点 m の固有値、 $\hat{\lambda}_k$ は推定固有値である。ここで、部分空間は、Reliable subspace (R), Noise subspace (N), Null subspace (Φ) [56] に区分される。分割点は、図に示すように、信頼性部分空間 (R) と雑音部分空間 (N) を分ける指標 m を示す。図 53 のように、信頼性部分空間 (R) と雑音部分空間 (N) を分離する。

ここで、本稿では、降順に並んだ隣接した固有値に対する固有値比を導入し、推定された分散とどの程度異なるかを観察する。

$$\frac{\hat{\lambda}_{k+1}}{\hat{\lambda}_k} = \frac{\alpha}{k+1+\beta} \times \frac{k+\beta}{\alpha} = \frac{k+\beta}{k+1+\beta}. \quad (42)$$

式 (42) は単調減少する関数で これは、 $\hat{\lambda}_k$ と $\hat{\lambda}_{k+1}$ の比を計算するものである。

固有値比の実測値と推定値をプロットしたものが図 54 である。白色化変換の高次元部分においても、固有値比は常に減衰していることが望ましい。しかし、大きな指数の固有値部分では、その多くが推定値から乖離している。この擾乱は逆固有値スペクトルの大きな振動を引き起こす。インデックス k の大きい成分は学習データによるノイズの影響を強く受けるため、テストデータでの認識性能が悪くなる。

次節では、特徴抽出の計算に用いる基底ベクトルの自己相関行列の固有値の大きさを制御することに焦点をあてて議論する。

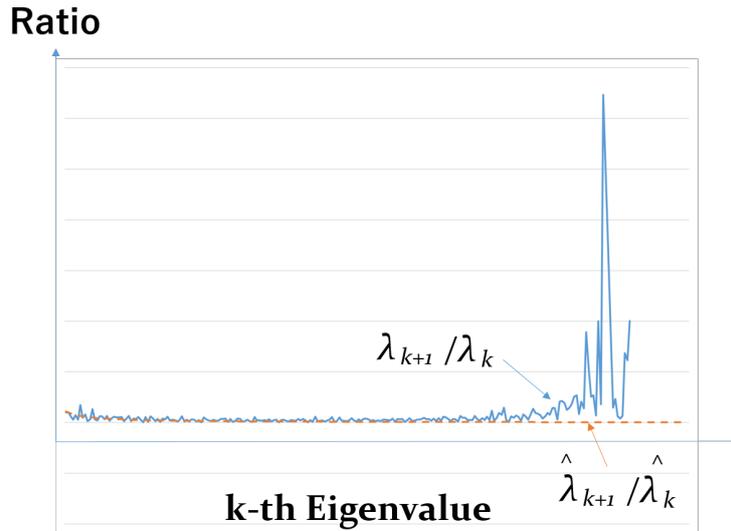


図 54: 実際の固有値比 (実線) と推定した固有値比 (破線) の比較。振動が指数の大きい成分で生じていることがわかる。

4.7.5 疑似白色化変換の導入

そこで、制約相互部分空間法の場合と同様に、一部の基底ベクトルの成分の重みを 0 にすることで、抽出する特徴の選択を行い、白色化相互部分空間法 (WMSM) と制約相互部分空間法 (CMSM) を組み合わせたような重みづけを考える。

図 55 は、それぞれ白色化相互部分空間法 (WMSM) の重みづけに対して、一部分を 0 とした重みづけの様子を示したものである。図 55(a) は、高次元部分を 0 に、図 55(b) は、低次元部分を 0 に、図 55(c) は、高次元部分と低次元部分を同時に 0 に設定したものである。

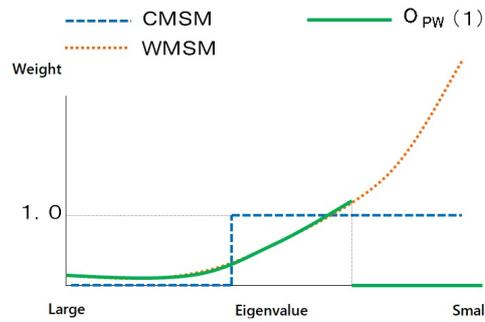
これを実現するためには、白色化行列に対して、 Λ の行列について、対角成分の一部、具体的には、小さい固有値のいくつか、または大きな固有値のいくつかを 0 と置き換えた行列を用意し、変更を加えた疑似白色化行列 \mathbf{O}_{PW} を作成する。

そして、白色化相互部分空間法の白色化行列 \mathbf{O}_{WMSM} を疑似白色化行列 \mathbf{O}_{PW} と置き換えて計算を行うことにより、従来の白色化相互部分空間法で向上しない適用例に対して改善が見込まれる。

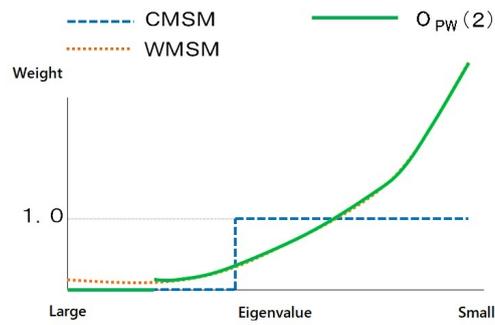
疑似白色化行列 \mathbf{O}_{PW} は式 (43) で定義される。

$$\mathbf{O}_{PW} = \mathbf{C}_s \Lambda^{-\frac{1}{2}} \mathbf{B}_P^T = \Lambda_s^{-\frac{1}{2}} \mathbf{B}^T \quad (43)$$

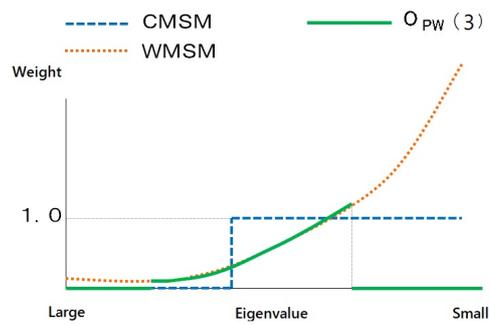
ここで、 \mathbf{B} は固有ベクトルを並べた行列、 Λ は固有値の対角行列、 \mathbf{C}_s は 0 と 1 を要素に持つ対角行列である。この Λ_s を式 (44) とする。(a) では、 Λ の固有値の小さいものから



白色化変換に対する重みの変更 (a)



白色化変換に対する重みの変更 (b)



白色化変換に対する重みの変更 (c)

図 55: 白色化変換に対する重みの変更例

法 (MPWMSM) を導入する。この手法は、4.5 章で導入したように、複数の特徴抽出を行う方法を WMSM/PWMSM にも適用する。セットベースの画像認識に有効な特徴を抽出するために入力部分空間と参照部分空間を多重に特徴変換する。図 56 は、多重擬似白色化相互部分空間法 MPWMSM (Multiple Pseudo-Whitened Mutual Subspace Method) の処理図を示す。複数の擬似白色化変換は式 (43) の射影行列で表現される。また、複数の変換を生成するために、多重制約相互部分空間法で導入したアンサンブル学習が提供するフレームワークを適用してもよい。

類似度を求めるには、それぞれの射影変換を行った部分空間の類似度を複数求め、それらの類似度 S_T は、式 (46) にて以下のように求められる。

$$S_T = \sum_{i=1}^M \alpha_i S_{O_{pw_i}}, \quad (46)$$

ここで、 M は擬似白色化変換の数、 α_i は i 番目の変換 O_{pw_i} の係数、 $S_{O_{pw_i}}$ は O_{pw_i} に射影された $P_{O_{pw_i}}$ と $Q_{O_{pw_i}}$ の間の角度からなる類似度を表している。

4.8 多重擬似白色化相互部分空間法による実験

4.8.1 多視点物体識別実験

3次元モデルを複数視点から撮影した公開データベース (ETH-80 [53]) から、形状が類似している 30 モデルを抜き出して用いる。これらの実験条件は文献 [35] に従っている。各モデル毎に図 57 に示す 41 視点から撮影された画像セットが用意されている。視点は全てのモデルに対して同じである。これらの内、奇数枚目 (21 枚) を各クラス部分空間とし、学習用部分空間用のデータとしても使い、偶数枚目 (20 枚) を評価データとした。すなわち学習視点と評価視点は異なる。評価データについては、20 枚の中から i から $i + 9$ までの 10 枚を取り出して一つのデータセットとし、これを開始フレーム i を 1 から 10 まで変化させて合計 10 個の評価セットを用意した。したがって全試行回数は $300(=10 \times 30)$ 回である。

評価には元画像を 15 ピクセル \times 15 ピクセルのモノクロ画像に変換したものをを用いた。よってデータの次元数 f は $225(=15 \times 15)$ である。次元数は全て 7 である。入力および各クラスの参照部分空間の次元数も 7 とした。また、学習に用いる部分空間の次元数 (N_c) は 16 とした。

図 58 は、 O_{PW} について、高次元部分をカットした場合の認識正解率と等エラー率を示したものである。40 次元目から順次 225 次元目までを 0 になるように変化させた結果



図 57: dog1 の全データ（点線で囲んだ列が学習用画像）

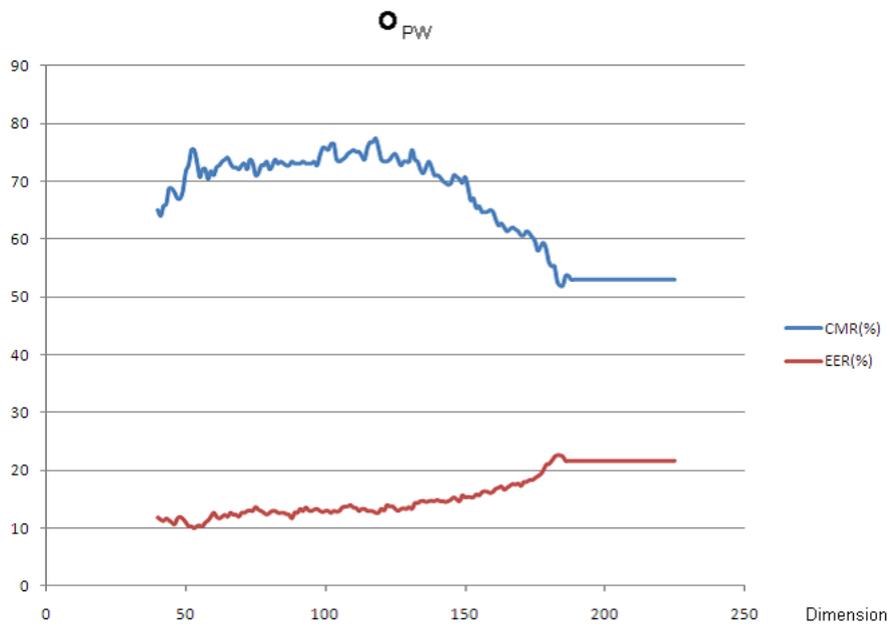


図 58: O_{PW} の変化: 高次元部分のみを削除

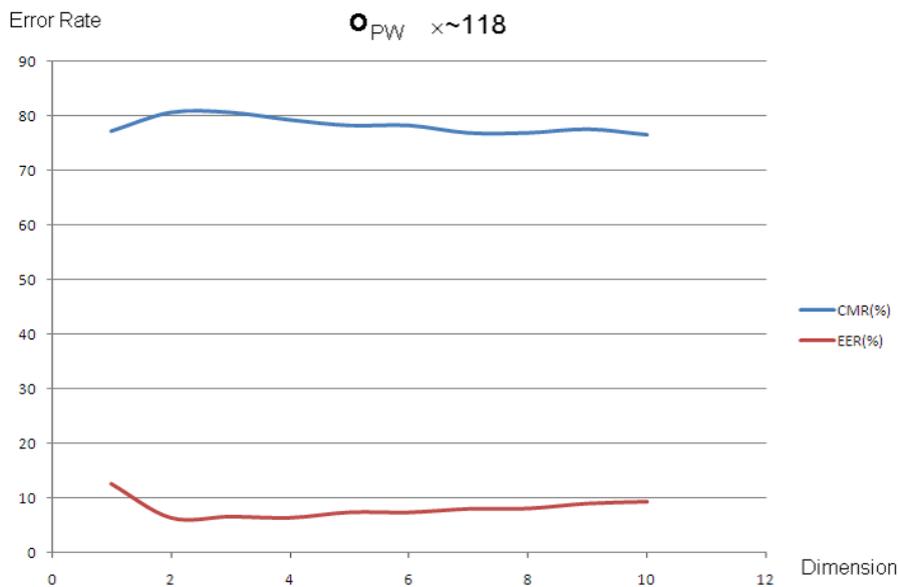


図 59: O_{PW} の変化: 高次元部分 (118 次元以降カット) と低次元部分も削除

である。元の白色化変換が 54.0% に対して、77.3% まで向上する (118~225 次元までを削減)。さらに、この最大値の時の 118 次元を固定し、低次元部分から C_s の要素が 0 になるように変化させた場合の結果が、図 59 である。この場合、最大、80.6% まで改善し、表 11 の S[1] と比較すると、他の線形の特徴抽出系の性能を超えることが分かった。両端の固有値の制御した場合には実験の結果、82.3% (5~104 次元を利用時) となる場合があり、制約相互部分空間法における共通な成分を取り除くといった処理と白色化変換における高次元の不安定な成分の除去を組み合わせることによってさらに高い性能を得ることができていることが確認できる。

さらに、特徴抽出を多重化した MWMSM と MPWMSM を追加で実験を行った。学習データを 3 つに分割し、各パートの 10 サンプルから白色化変換を作成した。その結果、複数の擬似白色化変換 (MPWMSM) (3~104 次元を利用時) を行うことで性能がさらに向上する一方、複数の白色化変換 (MWMSM) を行うことで性能が悪化することが確認された。

表 12: 多重化の実験結果

	Acc. (%) by S[1]
MWMSM	11.0
MPWMSM (3, 104)	84.0

特徴選択により元のデータ次元よりも小さな次元のデータに埋め込むことができるため、参照部分空間のメモリ効率を向上させることができ、計算量も軽減できる点は実用的な意味が大きいと考える。

4.8.2 顔認識実験

次に、顔画像データを用いた認識の実験を行う。データは、表情変化などを含まず、立位の状態で顔を軽く左右上下に変化させて収集したデータセットで、25名 x 8系列、それぞれの系列で照明環境を変動させて取得した。前節の例に比べると非線形性の少ないと考えられる。目鼻の位置情報から正規化顔パターンを作成し、その輝度値を要素を持つ特徴ベクトルを作成した。特徴量の次元数は225次元 (=15x15) である。

学習用部分空間としては、参照部分空間を用い、白色化相互部分空間法 (WMSM) の性能を見る。表 13 は、その結果で、相互部分空間法 (MSM) で 80.0% の性能であるのに対して、白色化相互部分空間法 (WMSM) の性能は学習に用いる部分空間の次元数 (N_c) が 10 次元の場合には、相互部分空間法 (MSM) の性能を下回る。しかし、学習に用いる部分空間の次元数 (N_c) を大きくすると、相互部分空間法 (MSM) よりも性能がよくなり、 $N_c = 20$ のとき、90.8% となる。しかしながら、学習用に必要な次元数 (N_c) を大きくすることは、保存用のメモリが大きくなることから、できるだけ小さい (N_c) での性能向上を図ることを考える。そこで、 $N_c = 10$ の場合に、特徴抽出を行う疑似白色化変換による認識を行うと、90.3% まで性能向上が見られる。このように、サンプル数やクラス数が少ないような場合でも、高性能化が可能となる場合があることが確認できた。

表 13: 顔認識実験

手法	正解率(%)			
MSM	80.0%			
手法	正解率(%)	手法	O_w	正解率(%)
WMSM($N_c=10$)	40.6%	PWMSM($N_c=10$)	0-100	89.7%
WMSM($N_c=15$)	84.6%	PWMSM($N_c=10$)	0-110	89.1%
WMSM($N_c=20$)	90.9%	PWMSM($N_c=10$)	0-120	88.6%
WMSM($N_c=25$)	85.7%	PWMSM($N_c=10$)	0-130	90.3%
		PWMSM($N_c=10$)	0-140	86.3%

4.9 深層学習特徴量との組み合わせに関する検討

近年では、深層畳み込みニューラルネットワーク (DCNN: Deep Convolutional Neural Network) を用いた特徴量が様々な分野の識別性能に高い有効性を示している [57, 58]。また、相互部分空間法と CNN 特徴を組み合わせることで性能が向上することを示した研究もいくつか発表されている [59, 60]。本研究では、対象ドメインに対する学習コストを削減するために、Deep Convolutional Neural Networks (DCNN) の再トレーニングを行わず、既製の CNN 特徴を利用する方針で、疑似白色化相互部分空間法との組み合わせを検討する。すなわち、学習データを活用し、ファインチューニングを加えて CNN 特徴を改善することもできるが、画像セットに対する部分空間ベースの特徴抽出が深層学習特徴量にも、うまく機能することを確認する。

4.9.1 ハンドジェスチャ認識 (IPN Hand)

ハンドジェスチャー認識 (HGR) は、ヒューマンコンピュータインタラクションに不可欠な機能の一つであり、応用範囲が広い。IPN Hand[61] は、リアルタイムハンドジェスチャー認識用のビデオデータセットである。このデータセットに含まれるジェスチャーは、タッチレススクリーンとのインタラクションに焦点を当てたもので、13 のカテゴリを含んでいる。被験者 50 名の内訳は、女性 16 名、男性 34 名となっている。このデータセットは、背景や照明が異なる約 30 の実世界の多様なシーンから収集されている。

[61] にて、分類指標の一つとして評価されている Isolated HGR タスクについて実験を行う。データは被験者ごとにトレーニングセット (74%) とテストセット (26%) に分割し、トレーニング動画 148 本、テスト動画 52 本が作成されている。訓練用とテスト用のジェスチャーインスタンスの数は、それぞれ 3,117 個と 1,101 個である。本実験では、文献で使用されている RGB-seg を使用している。これは、アノテーションデータとしてセマンティックセグメンテーションによる手領域のマスクが提供されており、そのマスクを用いて手領域のみ RGB 画像を領域抽出する。画像は ImageNet training データによって事前学習済みの公開 ResNet100 モデルで処理し、C5 層から 2048 次元の特徴を抽出する。なお、学習データは DCNN の学習には用いていない。学習データは部分空間ベースの特徴抽出にのみ使用している。

映像中の各ジェスチャーインスタンスの開始フレームと終了フレームのインデックスを手動でラベル付けし、部分空間を形成するための訓練に十分な情報を与えている。ビデオシーケンスは、手動で注釈された開始フレームと終了フレームに基づいて、分離されたジェスチャーサンプルにセグメント化される。本手法の評価には分類精度を用いており、これは正しくラベル付けされたサンプルの割合を示す。

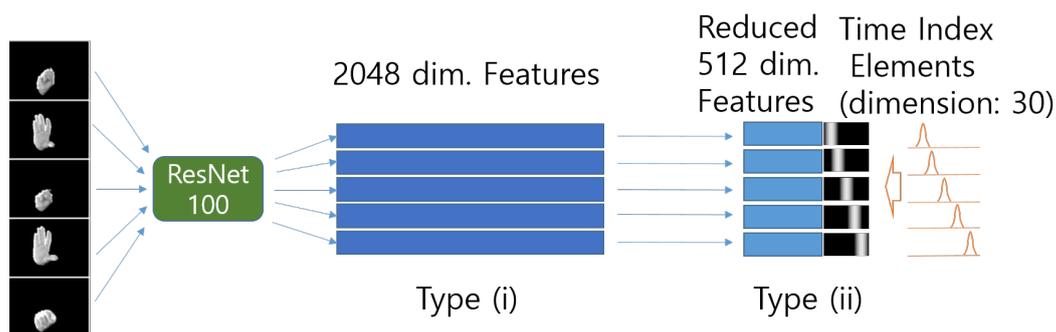


図 60: HGR のフレーム特徴量の説明図

IPN Hand の分類結果を表 14 に示す。図 60 に示すように、2 種類の特徴量を試している。タイプ (i) は DCNNs からフレームごとに抽出した 2048 次元の特徴量、タイプ (ii) は元の DCNNs の特徴量を PCA により 512 次元に削減した特徴量を用いている。さらに、時間推移を表現するために 30 次元の時間インデックス要素を追加する。時間指標には、図 60 のように時間軸に沿って正規分布を移動させる。また、各ジェスチャーの時間長に応じて均等に分割し、30 次元インデックスのピークの異なる位置に割り当てている。

提案手法である MPWMSM は、他の特徴抽出を行った手法より優れていることがわかった。[61] においてテストした 3 次元版 ResNet-50 には及ばないが、提案した MPWMSM は、他の特徴抽出法よりも有効であることが確認された。

4.9.2 ビデオベースの顔認識 (YouTube Face dataset)

次に、動画の顔画像データを用いた認識の実験を行う。YouTube Face dataset (YTF) は、動画上の顔認識のベンチマークとして最も広く利用されているデータセットである。YTF は、1,595 の ID を持つ 3,425 の動画から構成されている。YTF の評価プロトコルでは、5,000 のビデオペアを 10 回に分けてマッチングし、その平均精度を求める。各フォルドは 500 組の映像ペアで構成され、被写体の性質が互いに排他的であることが保証されている。

YTF ベンチマークの 2022/1 現在のトップ精度は、ResNet64 ベースの特徴量を用い、ビデオ内の全顔の単純平均特徴量を適用して最終スコアを算出する SeqFace [62] である。ArcFace [63] では、MS1Mv2 データセットで学習した ResNet 100 ベースの特徴量を使用する。こちらも単純平均特徴量を適用して評価を行っている。そこで、深層学習による state-of-the-arts の CNN 特徴を利用し、提案手法を適用する。

CNN 特徴量としては、ArcFace 損失を用いた ResNet100 モデルを用い、Glint360k データセットで学習させたものを使用する。また、学習済みモデルは InsightFace [64] が公開

表 14: IPN Hand に対する実験結果. (i) 2048 次元の深層学習特徴量、(ii) 時間インデックスをつけた次元削減した 512 次元深層学習特徴量.

		Acc. (%)
(i)	MSM (S[1])	38.23
	MSM (S[5])	43.96
	WMSM (S[5])	50.40
	PWMSM(S[5],(j, k)=(1, 388))	55.40
	MPWMSM(S[5],(j, k)=(5, 388))	57.00
(ii)	MSM (S[10])	35.51
	WMSM (S[10])	47.32
	PWMSM (S[10],(j, k)=(20, 320))	58.94
	MWMSM(S[10])	59.21
	MPWMSM (S[10],(j, k)=(20, 400))	60.13
	3D versions of ResNet-50 [61]	75.11

しているモデルをダウンロードして利用した。特徴ベクトルの次元は 512 次元である。YTF の動画データについては、MTCNN [65] を用いて顔の特徴点を 5 点取得し、顔画像を 112x112 ピクセルに成形する。なお、本実験では、DCNN については YTF データを用いたファインチューニングは行っていない。

表 15: YouTube Face dataset の実験結果

	Acc. (%)
Simple average features	98.22 ± 0.61
MSM (S[4])	98.28 ± 0.69
MWMSM (S[2])	98.30 ± 0.80
MPWMSM (S[2],(j, k)=(3, 277))	98.38 ± 0.60
SeqFace [62]	98.12
ArcFace [63]	98.02

YTF の評価結果を表 15 に示す。正準角度の数 (式)、参照部分空間の次元数 N_p (式)、重みの選択 (式) のパラメータを変えて実験を行った。

Arcface の評価結果と比較すると、学習データの増加 (MS1Mv2 データセットから Glint360k データセット) により、単純な平均特徴量を用いることで精度が向上していることが分かる。さらに、各映像の部分空間表現を用いた MSM を用いることで、精度が

向上していることが確認できた。S[4] と表記した類似度の正準角の数が4である場合に最も良い性能が得られた。

YTF では特定のフォルドをテストする際に10個のフォルドがあるため、1個をテストに使い、残りの9個をMWMSM/MPWMSMの白色化変換の学習に用いた。その後、4.7.6節に、したがって多重化の計算を行う。このとき、各疑似白色化変換の係数はすべて同じとする。

高性能な識別特徴量を用いた場合、WMSMを用いた多重化がMSMよりも優れていることが確認された。なお、今回はYTFの実験プロトコルに整合させるため、バギングにおけるランダム選択は行っていない。各フォルダには個別IDの重複がないため、各フォルダの部分空間を用いて白色化変換／疑似白色化変換を計算する。学習に用いる部分空間の次元数(N_p)は20とした。

結果として、提案した疑似白色化変換が最も良い結果を示し、識別精度は98.38%であることがわかる。また、MPWMSMは、前節で確認したように、MWMSMと同等の改善効果を示した。提案した疑似白色化変換は、白色化変換よりも効果的であることを確認することができた。

4.10 本章のまとめ

本章では、第1章で述べた第三の視点から、相互部分空間法の高精度化を行う方法について議論した。まず、複数の制約部分空間の生成に対してアンサンブル学習を適用し、複数の制約部分空間を用いて特徴抽出を多重化した多重制約相互部分空間法について述べた。提案法では、入力部分空間と参照部分空間を複数の制約部分空間へ射影することで特徴抽出を行う。制約部分空間に射影された入力部分空間と参照部分空間の類似度を算出し、制約部分空間の個数だけ求めた類似度を結合することで最終的に類似度を決定する。複数の制約部分空間を生成するために、アンサンブル学習の代表的な手法である並列的な学習と逐次的な学習の枠組みを導入した。提案手法により従来の認識手法に比べて識別精度が向上することを、照明変動が生じる環境で撮影されたデータベースと大規模人数が登録されたデータベースを用いた識別実験で確認した。

さらなる相互部分空間法の高性能化に関して、線形の特徴抽出系をもつ白色化相互部分空間法を導入し、白色化変換を構成する基底ベクトルに対する特徴選択を行うことで、疑似的な白色化変換を行う方法である疑似白色化相互部分空間法について検討した。非線形性の強いデータや、カテゴリ数、データのサンプル数が少ない場合などに、従来の白色化変換による性能を改善できることを確認した。

また、深層学習特徴量との組み合わせにおいても、既存の深層学習特徴量をファイ

チューニング等のネットワークの再学習を行わず、部分空間表現を用いた複数画像に対する特徴抽出を行うことで、さらなる複雑な変動を考慮して認識精度の向上を図ることができることが確認できた。

5 顔認識への応用例

5.1 ポータブルPCにおける顔認識システム

5.1.1 PCにおける簡易セキュリティ

ヒューマンインタフェース、セキュリティなどといった分野で顔画像を用いた個人識別技術が重要な要素技術として注目されている。顔画像認識には、非接触で、利用者に意識させずに個人の識別が可能といった利点がある。また、個人の識別だけではなく、画像中の顔領域位置、顔特徴点位置の情報を用いた様々なアプリケーションが構築できる [66]。

本章では、可搬型のパーソナルコンピュータ (以下略してPC) 上での動作を前提とした顔認識システムについて述べる。容易に持ち運ぶことができ、サーバのように強力な計算能力を持たない可搬型PCで顔認識システムを構築するために、次の課題について考える必要がある。

- (a) 認識を行う際のユーザの負担軽減
- (b) 計算コストの削減

課題 (a)、すなわちユーザが利用しやすい顔認識システムを構築するには、認識を行う際の人物に対する制約が少ないことが重要である。従来の顔認識システム (例えば [67]) では、画像の撮影時に「真顔で正面を見つめる」といった拘束が必要であった。身構えることなく、自然な体勢で認識できることを目指すには、顔の向き、表情の変化による顔の見え方が変動する場合に対応しなければならない。特に、本論文で対象とする可搬型のシステムでは、両手で持ち上げたりや膝の上に置くなど、システムと人物との位置関係の自由度が大きく、顔の見え方が変化しやすい。

このような状況では、従来の静止画ベースの認識法では誤認識が多くなる [68][9]。そこで、一枚の静止画による認識ではなく、入力画像列全体の分布を利用して判断することで、顔の向きや表情などの変動を吸収する。本システムでは動画像から得られた複数枚の画像を利用した顔認識法 [68][9] を実装する。

また、本システムで想定している状況では、任意の背景から顔画像パターンを検出する必要がある。本システムでは形状情報とパターン情報の組合せによる顔特徴点抽出法 [69] により、この要求を満たす。

一方、課題 (b) については、他のアプリケーションソフトとの共存に配慮して、できるだけ計算コストの少ない実装が必要である。一般に画像処理アプリケーションは、音

声処理等と比較しても、扱う計算量、メモリ量が格段に多いため、実用的なシステム構築のためには、処理を軽減する工夫が必要となる。

これに加えて、本システムでは、ユーザの負担を少なくするアルゴリズムの導入による動画像処理が加わり、計算コストがさらに増大する。対象とする PC では、搭載される計算能力も限られているため、特に計算コストの削減が重要である。本論文では、高速計算のためのアルゴリズムの工夫や、動画像における特徴点のトラッキング、CPU に付加されたマルチメディア命令の積極的な利用、画像入力に対する制御について検討する。

次に、アプリケーションの観点から考えると、顔画像認識の結果をいかに利用するかが問題となる。例えば、セキュリティの側面からは、PC に保存、表示されたデータの盗み見などを防ぐことが挙げられる。また、ヒューマンインタフェースの側面からは、ユーザを識別することにより、個人毎の環境設定など、利用者に応じた情報、サービスの提供が考えられる。

さらに、顔画像認識がもたらす情報には、個人識別だけに限らず、様々な利用法が存在する。本論文で提案するシステム “Smartface” では、アミューズメント的な要素を取り入れて、(1) 個人識別による音声応答、環境設定、(2) 顔認識つきスクリーンセーバ、(3) リアルタイム変装シミュレーション、といった「楽しめるソフトウェア」を実現する。

以下、5.1.2 では、本システムで用いた顔特徴点抽出法と顔の向きや表情の変化にロバストな個人識別アルゴリズムについて述べる。5.1.3 では、システムの構成と計算コスト削減の対応について説明する。5.1.4 では、搭載したアプリケーション機能について述べる。

5.1.2 顔向きや表情の変化に頑健な顔認識アルゴリズム

本システムにおける顔認識処理の流れを図 61 に示す。処理は (1) カメラからの画像入力、(2) 入力された画像中の顔領域の検出、(3) 顔特徴点（目鼻）の位置検出、(4) 位置、大きさなどを正規化したパターン切り出し、(5) 入力部分空間の計算、(6) 登録辞書との類似度計算による識別処理、を順に行って個人識別し、その結果に基づいて (7) アプリケーションの処理を行う。

顔領域の検出

顔領域の検出は、2 章で述べた、高速な定性的マッチング (QTR) と部分空間法 [32] を組み合わせたパターンマッチングを行う。あらかじめ様々な人物、方向の顔パターンデータ (15×15 pixel の画像) を収集し、顔検出用のテンプレート辞書を作成しておく。そのテンプレートを画像全体に走査し、各場所での QTR による類似度を逐次求め、類似度が局所最大かつ、しきい値以上の箇所に部分空間法による検証を組み合わせ、顔領域として検出する。顔領域の大きさの変化に対応するために、入力画像を数段階にスケール

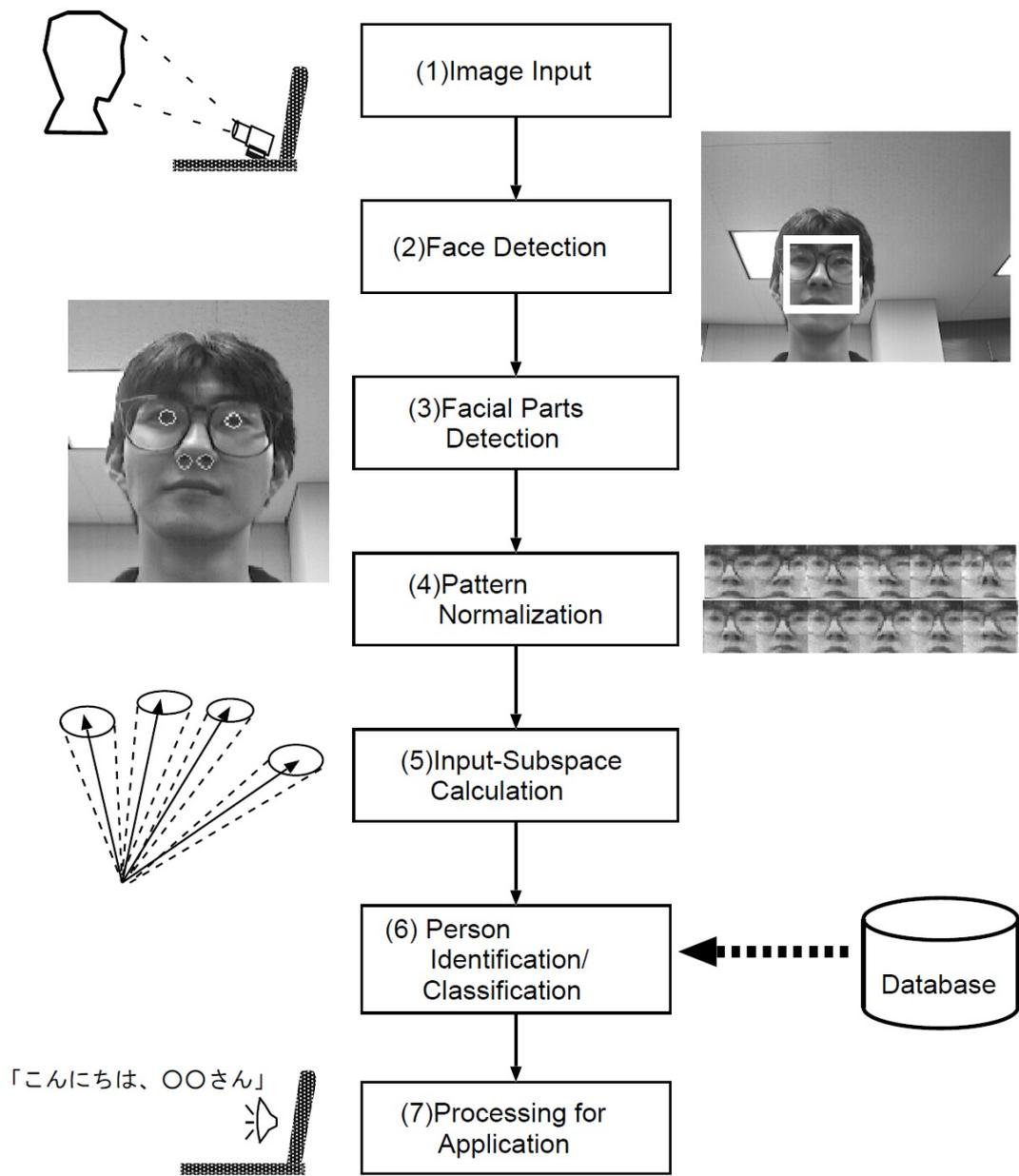


図 61: 処理の流れ図

変化させてマッチングを行う。画像中にしきい値以上の箇所が無い場合は、人が存在しないとして、以後の認識処理を行わない。

複雑な背景の中から顔領域を検出する場合、顔パターンに類似した非顔領域の誤検出が問題となる。この問題に対応するために、文献 [70] で提案されているように、顔の辞書との類似度が高い非顔データを収集し、それらから構成された非顔辞書との類似度を同時に求め、顔、非顔の2つの類似度を用いて検出を行う。

顔特徴点検出とパターンの切り出し

検出された顔領域に対して、過去に提案した特徴点抽出法 [69] を適用し、目鼻（瞳、鼻孔）を検出する。この方法は形状情報とパターン情報を同時に考慮して安定な顔特徴点抽出を行う。

まず、分離度フィルタにより円形領域を特徴点候補として抽出する。分離度フィルタは図 62(a) に示すように2つの円形領域から成り、次式によって求まる分離度 $S(0.0 < S \leq 1.0)$ を出力する。

$$S = \frac{n_1(\bar{P}_1 - \bar{P}_m)^2 + n_2(\bar{P}_2 - \bar{P}_m)^2}{\sum_{k=1}^N (P_k - \bar{P}_m)^2} \quad (47)$$

ここで、 N は領域内の全画素数、 n_1, n_2 は領域 1, 領域 2 の画素数、 P_k は位置 k における輝度、 \bar{P}_m は領域全体での輝度の平均値、 \bar{P}_1, \bar{P}_2 は領域 1, 領域 2 での輝度の平均値を示す。

顔領域内の各画素に対して、フィルタの位置と半径を変えながら分離度を算出する。ここで、分離度の局所最大点が特徴点の候補となる。抽出された候補点の一例を図 62(b) に示す。

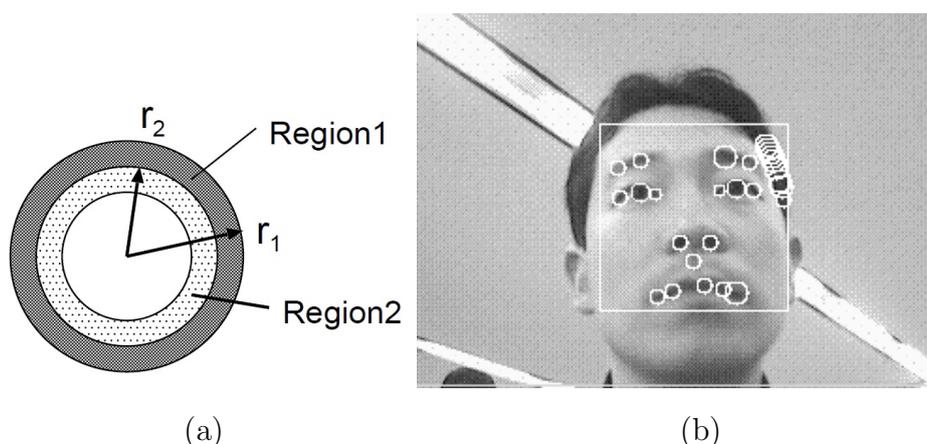


図 62: 円形分離度マスク (a) と特徴点候補 (b)

次に、複数の特徴点候補に対し、部分空間法によるパターン照合を行って、目鼻パターンとの類似性を検証する。最終的には、類似度がしきい値以上の候補点で、顔という幾何学的配置の制約を満たす4点を正しい特徴点とする。なお本処理に関する詳細は [69] に示す。

次に、抽出された目鼻の特徴点を基準にして、顔領域の位置、サイズを正規化し、矩形の濃淡パターンとして切り出す。図 63 は、実際に連続して得られた 20 枚の矩形の濃淡パターン (30×30pixel) である。人間の作業なしで自動的に大量のパターンを生成できる。



図 63: 正規化パターン取得例

相互部分空間法による個人識別

本システムでは、顔の向きや表情変化といった変動を吸収するため、動画像から得られる複数枚の画像を用いた認識法 [68][9] を適用する。

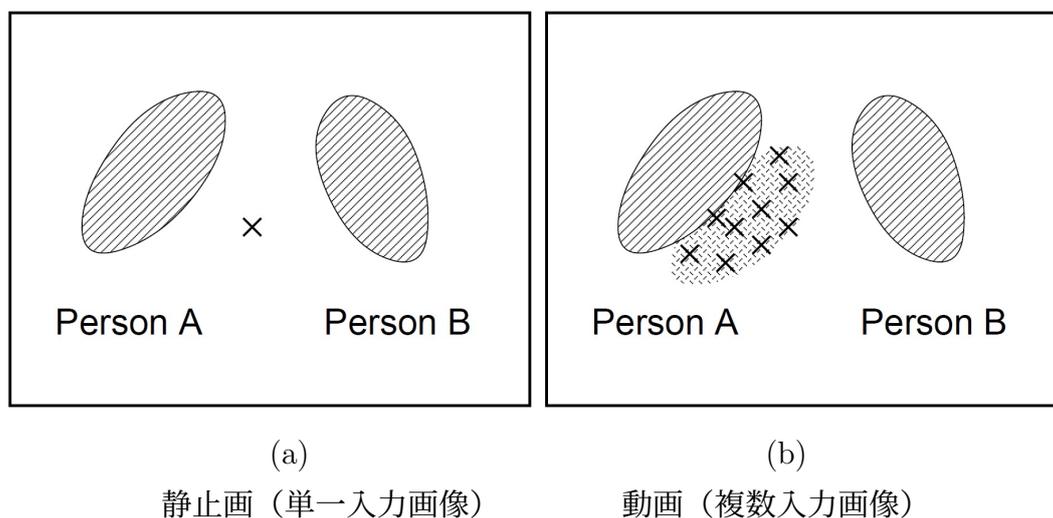


図 64: 静止画と動画による認識の概念図

図 64 は、静止画像による認識と動画像による認識のアプローチの概念図である。この図は特徴空間中の各人物のサンプルデータの分布を網目で表している。図 64(a) は、従来の静止画による認識法を表し、単一の入力 (図中×印) がいずれのサンプル集合に近

いかによって識別することを表している。この例ではどちらの人物に識別すべきかを決定することが困難である。これに対して、時系列で連続して得られる多様な複数の入力データを利用して、図 64(b) に示すように、動画像中の入力顔データの分布 (図中×印の集合) が、どの人物の分布に近いかを求めることにより識別を行えば、変動に強い認識が可能になる。

複数の画像データを用いたパターン認識法として相互部分空間法 [8] を用いる。この認識法では、入力ベクトルも部分空間で表現し、辞書パターンの部分空間との間の角度を類似度として識別を行う。

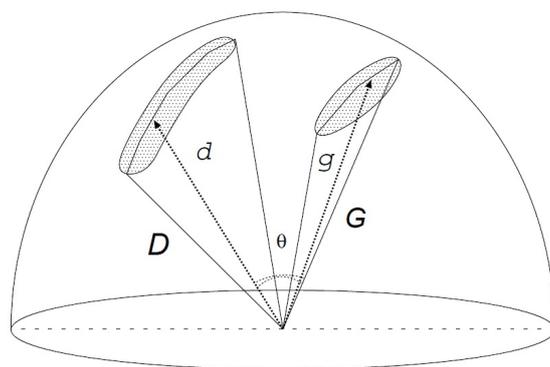


図 65: 相互部分空間法の説明図

図 65 は相互部分空間法の説明図であり、2つの部分空間 D, G のなす角度 θ の余弦は、

$$\cos^2\theta = \sup_{d \in D, g \in G, \|d\| \neq 0, \|g\| \neq 0} \frac{|(d, g)|^2}{\|d\|^2 \|g\|^2} \quad (48)$$

と定義する (d, g は式 (48) が極値をもつためのそれぞれの部分空間上のベクトルとする)。これに関して、2つの部分空間 D, G への正射影作用素を P, Q とする場合、 $\cos^2\theta$ は PQP の最大固有値 λ_{max} となる [8]。

$$\cos^2\theta = \lambda_{max} \quad (49)$$

辞書パターンの部分空間を D 、入力された時系列画像に対する部分空間を G とする。ここで ϕ, ψ を各部分空間 D, G における固有ベクトルとする。実際には PQP という行列の最大固有値を求めるのではなく、式 (50) で表される行列 X の固有値問題を解き、その最大固有値を類似度 (部分空間間類似度) S_{mutual} とすればよい [8]。ここで、辞書部分空間 D の次元を M 、入力部分空間 G の次元を L として、 $L \leq M (1 \leq i, j \leq L)$ とする。

$$X = (x_{ij}) = \sum_{m=1}^M (\psi_i, \phi_m)(\phi_m, \psi_j) \quad (50)$$

$$W^T X W = \Lambda \quad (51)$$

(W: X の対角化行列、 λ_{max} : Λ の対角成分の最大値)

$$S_{mutual}(G, D) = \lambda_{max} \quad (52)$$

本手法の有効性を調べるために、101名分のそれぞれ、登録用データ180枚とテストデータ180枚の2回の画像取得により、36360枚の画像を収集し、オフラインで実験を行った。いずれも、顔の向きの変化や表情の変化を指示した変動のあるデータであり、口の動きがある会話中の顔データなども取得されている。

まず、登録している人物データのうち、最も高い類似度をもつ人物を識別結果とする識別実験を行った。入力/辞書部分空間の次元数を表16のように変化させた場合、次元数10次元のとき、99.0%の識別正解率を得た。これ以上の大きな次元数としても、識別正解率の向上はみられなかった。

また、提示した人物が本人と同一人物であるかどうかを判定する同定実験に関して、従来法との比較を行った[68][9]。本人排除率(False Rejection Rate: FRR)と他人受入率(False Acceptance Rate: FAR)を求め、2つの誤識別率が同じ値をとる等誤識別率(Equal Error Rate: EER)で評価すると、次元数5次元のとき、従来法が8%あるのに対し、本手法では2%にまで抑えることができた。

5.1.3 システムの構成

ターゲットシステム

本システムは、図66の東芝製カメラ付きミニノート型PC Libretto ff 1100(CPU: Intel MMX Pentium 266MHz)に、Windows98上で動作するアプリケーションソフトとして

表 16: 相互部分空間法による識別実験結果

辞書次元数	1	3	5	7	10
入力次元数	1	3	5	7	10
識別正解率	91.1%	96.9%	98.5 %	98.7%	99.0%

実装した。システムには、液晶画面とキーボードの間に CMOS カメラ（図中円内）が装備され、顔を見上げるように配置される。



図 66: 対象 PC システムの外観

図 67 は、ソフトウェアの構造を示したもので、5.1.2 章で述べた顔検出、目鼻検出のエンジンと個人識別エンジンが、それぞれ独立の DLL(Dynamic Linkage Library) として構成されている。これらの顔認識エンジンは、DirectShow のフィルタとして構成されており、そのフィルタを上位の Smartface アプリケーションが呼び出す形で利用される。顔、目鼻検出 DLL、個人識別 DLL は、それぞれが検出用のテンプレート辞書、そして、個人識別用辞書を利用する。

個人識別処理と顔の登録方法

識別処理は画像入力一枚毎に図 61 の処理が行われる。切り出された顔画像パターンに対してヒストグラム平坦化を行い、大きさを 1/4 に縮小し、各画素の濃淡値からなる $15 \times 15 = 225$ 次元のベクトルをデータとする。入力部分空間は、これらのベクトルデータのセットから計算される。識別は、入力部分空間と登録されている各個人辞書の部分空間との類似度の計算を行い、最大の類似度をもつものが、しきい値を越えた場合に、該当人物が存在すると判定する。

顔の登録の際には、個人の名前を入力し、その後、50 枚の画像の取得を行う。登録に要する時間は、画像の取得から辞書計算が完了するまで平均 20 秒程度である。このとき、ユーザが様々な顔の向き、表情変化をさせながら、バラエティのある顔画像を撮影することで、後に精度よい認識が行えるようになる。最大登録人数は 100 名である。

また、登録後の人物の経時変化に対しては、各個人辞書に新しいデータを追加し、辞

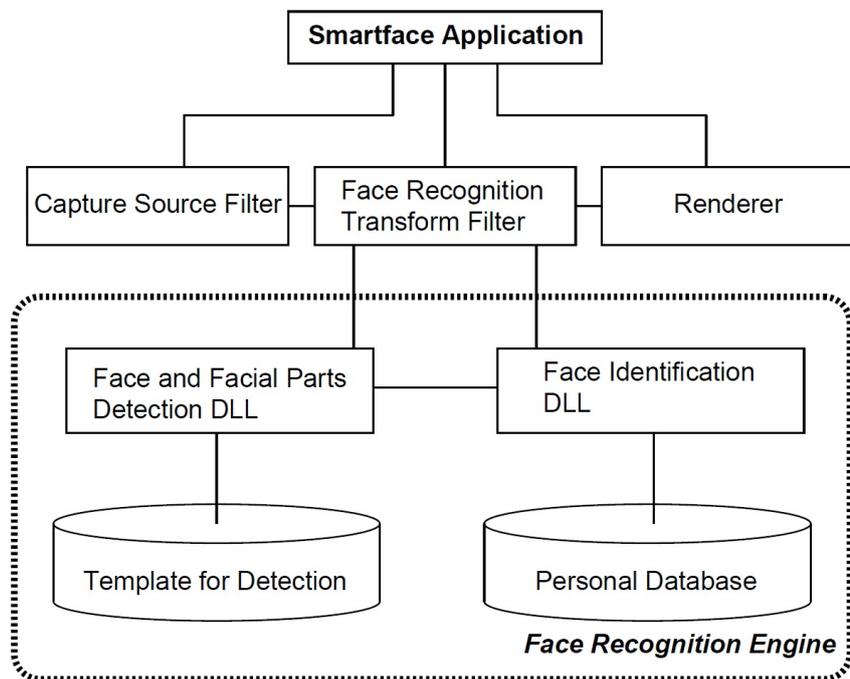


図 67: ソフトウェアの構成図

書を更新することで認識率を維持できることを実験的に確認している [71]。Smartface では、登録ユーティリティにて更新機能を選択し、登録時と同様に複数枚の画像を収集して、それらのデータを辞書に追加できる。

計算コストを考慮した実装 ここでは、実用的な処理速度を実現するために、各認識処理に対する計算コストの削減法について述べる。

[画像入力処理]

入力処理画像の大きさは、 320×240 pixel であり、アプリケーションが起動されると、常時画像のメモリ転送が行われる。顔領域検出の計算コストが大きいため、顔検出が数秒間行われなかった場合には、画像入力の 1 秒あたりの入力枚数（フレームレート）を落とす処理を行う。これにより、1 秒間あたりの処理枚数が低下し、CPU の占有率も低下する。その後、間欠的に顔領域検出が行われ、人物が検出されると、再び通常で検出処理が行われる。

[顔領域検出、目鼻特徴点検出処理]

顔領域検出、目鼻特徴点検出は、毎フレーム同時に実行するのではなく、目鼻特徴点検出が成功したフレームの次フレームでは、顔領域検出処理を行わない。さらに、検出された特徴点付近だけを探索領域として、特徴点のトラッキング処理を行う。目鼻特徴点が検出されない場合、次フレームでは再び顔領域検出の処理から実行される。

[個人識別処理における入力部分空間の計算]

本システムでは、辞書生成時の部分空間の計算は、数枚分のベクトルデータを蓄積した相関行列から、QR アルゴリズムにより行われる。一方、入力部分空間の計算は、毎フレーム行われるため、同様の処理では計算コストが大きい。

相互部分空間法による識別を高速に行うことを実現するために、逐次的に部分空間を計算する部分空間反復法 [32] を採用する。部分空間反復法は、計算に用いるすべての入力ベクトルを保持しないで良いという点と、逐次的に必要な数の固有ベクトルのみが求まるという利点がある。

相関行列の推定値 \hat{R}_k に対して、入力ベクトル g_{k+1} に対し、

$$\hat{R}_{k+1} = (1 - \mu)\hat{R}_k + \mu g_{k+1} g_{k+1}^T$$

により相関行列の推定値を更新する。ここで、 μ は、小さな値を持つ助変数であり、計算の反復回数（データの入力数）を r として、 $\mu = \frac{1}{r+1}$ としてもよい。

次に、固有ベクトルを列にもつ行列 C_k は、以下のように更新される。

$$\hat{C}_k = \hat{R}_k C_{k-1}$$

$$C_k = GS(\hat{C}_k)$$

($GS(\cdot)$ は、Gram-Schmidt の直交化を行う関数)

このアルゴリズムにより、入力部分空間が高速に計算でき、相互部分空間法を実時間で実行できる。

[マルチメディア命令の利用]

さらに、各処理部分で CPU のマルチメディア命令を使用し、高速化を図る。本システムで、最も多用されている処理は、ベクトル内積演算であり、この処理を効率化することが計算コスト削減につながる。

浮動小数点演算を整数化した後、マルチメディア命令を適用し、効果を調べた。表 17 は、前述した入力部分空間の計算と相互部分空間法の計算に要する時間をプロファイラで計測したもので、それぞれ、約 1/2、1/4 にまで削減される。整数化の効果は、部分空間法によるマッチングを行っている顔領域検出、目鼻検出の高速化においても大きな効果が見られた。

なお、整数化に伴う演算精度は、類似度計算において小数第 3 位以降に誤差が認められた。しかしながら、部分空間法における類似度の大小関係は変化しておらず、認識性能には影響しなかった。

表 17: マルチメディア命令による高速化の効果

	マルチメディア 命令未使用 (float)	マルチメディア 命令利用 (short int)
入力部分空間の計算	1700	851
相互部分空間法の計算	2855	645

(単位: μs)

[認識性能を考慮した類似度計算の次元数設定]

識別性能を維持しつつ、計算量を削減するという観点で、類似度計算における入力、辞書部分空間の次元数の設定を行う。表 16 の結果から、7次元と 10次元の場合では、識別性能の低下が少ない。また、式 (50) の計算で発生する内積計算の回数は、10次元の場合の 1/3 程度に減少する。この 2つの理由により、本システムでは、次元数を 7次元に設定する。

実際に可搬型 PC を用いて、50 名による識別実験を行ったところ、表 16 と同様の識別性能を実現し、特徴点検出から個人の識別までを一秒間に 5~6 フレームで処理することが可能となった。

5.1.4 Smartface のアプリケーション機能

キャプチャビューア

システムを起動すると、通常、図 68 に示すキャプチャビューアが開き、入力画像に顔領域、目領域の検出結果が重ねて表示される。認識の状況を知らせるアイコンと、認識の個人類似度のグラフにより、システムの状態がユーザにフィードバックされる。

ユーザ別設定機能

ユーザ別設定機能は、登録人物それぞれに対する環境設定を行う機能で、識別時の音声応答と起動アプリケーションが設定できる。通常は「こんにちは、〇〇さん」と、顔の登録時に入力した名前を音声合成により応答する。任意の音声ファイルを識別時に再生することもできる。また、アプリケーションをあらかじめ選択しておき、ユーザを識別した時点で起動することができる。例えば、メールの到着をチェックするアプリケーションを選択しておけば、当人が PC の前にくると、メールの有無を知らせる、といったことが可能になる。

顔認識つきスクリーンセーバ

従来のスクリーンセーバでは、パスワードが設定されている場合、そのパスワードを入力することで、スクリーンロックが解除される。Smartface スクリーンセーバは、顔認



図 68: キャプチャビューアウィンドウでの認識の様子

識の様々なタイミングでサーバを解除できる。指定可能なタイミングは、(1) 画像中に顔を検出すると解除、(2) データベースに登録されている人の顔を認識すると解除、(3) 特定の一人の人物の顔を認識すると解除、であり、用途やセキュリティの度合いに応じて設定できる。

リアルタイム変装シミュレーション

Smartface では、個人識別を行うだけでなく、顔領域の位置、大きさ、目鼻の位置情報を用いて、自動的かつリアルタイムに変装を楽しめるシミュレーション機能を有する。

この機能は、検出された顔の位置、大きさに応じて、CG 画像などの重ね合わせや画像の部分的な画像処理、変形などが行われる。図 69 は、「ピエロ」への変装の様子を表しており、登録されている様々なコンテンツをキャプチャビューアのボタンで選択できる。

従来、顔画像に装飾を施す場合は、顔領域、顔特徴点の位置等をユーザが手動で与えたり、顔の位置を自分で調整する必要があった。自動検出の例としては、文献 [72] のように、照明条件と背景を無地に固定した環境で静止画を撮影し、顔位置を検出して画像枠などの装飾を行うものがある。Smartface では背景等の制約もない上、動画に対して連続的に装飾できる。図 70 は、複雑な背景の前で、カメラを右から左に動かして取得した画像列であり、顔領域が移動しても、追従して装飾できることを示している。

変装コンテンツは、複数のビットマップを組み合わせたか、画像変形の内容を指定するために、専用のスクリプトで記述する。図 71(a) は図 69 で示した「ピエロ」の変装の全スクリプトである。

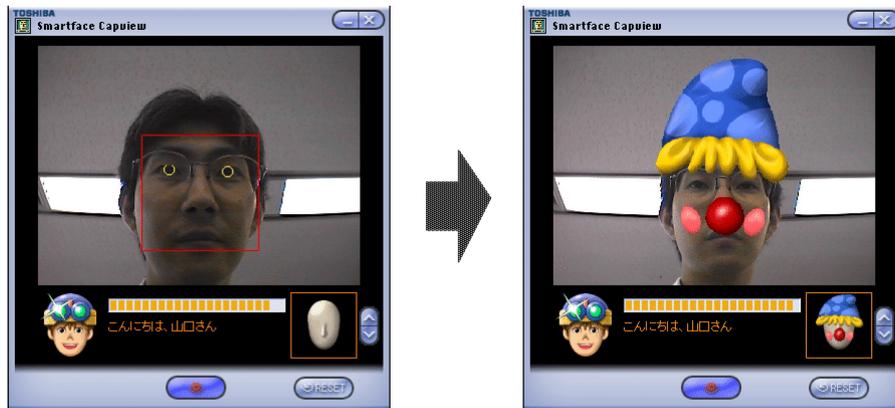


図 69: 変装シミュレーションの様子



図 70: 画像列に対する変装処理

```

# 「ピエロ」の変装スクリプト
FG_EMBEDBMP 2
LEFT_EYE ADD FACE_WIDTH -1.0 ADD FACE_HEIGHT -2.8 ← P1
RIGHT_EYE ADD FACE_WIDTH 1.0 ADD FACE_HEIGHT 0
COLORBMP "data/pierrot1.bmp" 255 255 255

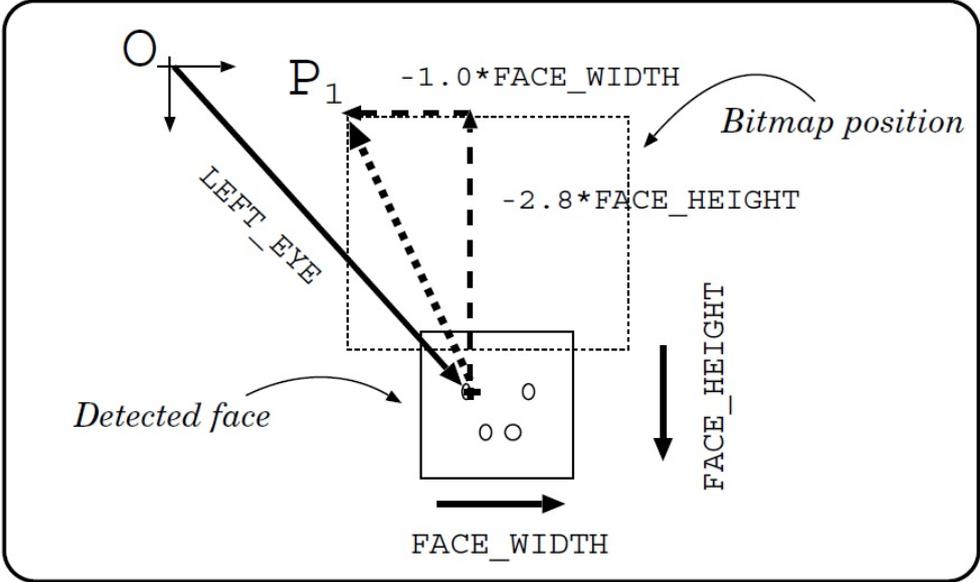
FG_EMBEDBMP 2
CENTER_NOSE ADD FACE_WIDTH -0.5 ADD FACE_HEIGHT -0.5
CENTER_NOSE ADD FACE_WIDTH 0.5 ADD FACE_HEIGHT 0.5
COLORBMP "data/pierrot2.bmp" 255 255 255

FG_EMBEDBMP 2
RIGHT_EYE ADD FACE_WIDTH -0.1 ADD FACE_HEIGHT 0.4
RIGHT_EYE ADD FACE_WIDTH 0.5 ADD FACE_HEIGHT 1.2
COLORBMP "data/pierrot3.bmp" 255 255 255

FG_EMBEDBMP 2
LEFT_EYE ADD FACE_WIDTH -0.5 ADD FACE_HEIGHT 0.4
LEFT_EYE ADD FACE_WIDTH 0.1 ADD FACE_HEIGHT 1.2
COLORBMP "data/pierrot4.bmp" 255 255 255

```

(a)



(b)

図 71: 「ピエロ」の変装スクリプト (a) と相対的位置の指定法 (b)

顔の移動に応じて、任意の位置、大きさで、重ね合わされるビットマップの配置位置や変形対象領域の位置の指定法を図 71(b) に示す。図 71(a) のスクリプトにおけるビットマップを配置する矩形の頂点 P_1 を例に説明する。頂点 P_1 は、左目の検出位置 (LEFT_EYE) を示す「基準点ベクトル」と、相対位置と大きさを規定するための顔領域の水平、垂直軸方向の 2 つの「相対位置ベクトル」 (FACE_WIDTH, FACE_HEIGHT) ならびに、それぞれの倍率によって表現される。

スクリプトでは、図 71(a) で示すように、複数個の頂点を指定し、配置する場所を決定する。この表現法により、顔の大きさ、位置が変化しても、相対的なビットマップの位置、大きさの関係が保たれる。また、ビットマップの重ね合わせだけでなく、画像のモザイク化や拡大縮小などのコマンドなどを用意しており、自由にコンテンツを追加、変更、操作できるよう設計している。

図72は、様々な人物に変装コンテンツを適用した画像例である。これらの機能を MPEG4 などを用いた画像通信のための映像効果前処理として利用することも可能である。



図 72: 様々な変装シミュレーションの様子

5.1.5 本節のまとめ

本節では、可搬型 PC への搭載を前提とした顔画像認識システム “Smartface” について述べた。

可搬型 PC では顔の向きや表情の変化への対応が不可欠となるが、この課題に対して、相互部分空間法を適用した。しかしながら、動画像認識の導入は計算量の増大を引き起こす。これに関しては特徴点のトラッキングや計算アルゴリズムの工夫による解決策を示した。この結果、高い識別性能と少ないユーザ負担という相反する 2 つの条件を同時に満足する顔画像認識システムが実現できた。

このシステムは、4 章で述べた多重制約相互部分空間法が適用された入退室システムである FacePass[39] へ発展している (図 73)。



図 73: 顔照合セキュリティシステム FacePass(東芝のプレスリリースならびにホームページより引用)

5.2 公開ベンチマークデータによる顔認識実験

5.2.1 FERET DB

本節では、公開データベースを用いて顔認識実験を行う。用いる実験データセットは FERET-fafb で、広くベンチマークに用いられているデータセットとして知られている。FERET の種々のプロトコルが存在するが、テストデータの枚数が最も多いものを選び、参照画像 (gallery) 1196 枚と入力画像 (fafb) 1195 枚を用いて評価する。

特徴量は、文献 [73] にて利用されているもので、1つの顔から多数の特徴量を作成し、複数の特徴量の類似度をアンサンブル結合した結果を用いる。

5.2.2 各手法による認識結果

各手法の識別性能について比較実験を行った。それぞれの手法のパラメータは以下のように設定した。

(a) 相互部分空間法 (MSM)

入力部分空間と参照部分空間とのなす最小角度 θ の $\cos^2 \theta$ を類似度とする。入力部分空間と参照部分空間の基底ベクトルの本数 N はともに 7 とした。

(b) 制約相互部分空間法 (CMSM)

単一の制約部分空間で特徴抽出を行い相互部分空間法により類似度を算出した。入力部分空間と参照部分空間の基底ベクトルの本数 N はともに 7 とした。制約部分空間は、

表 18: FERET fafb の各手法の認識結果

	方式	一位正解率 CMR(%)	等エラー率 EER(%)	本人排除率@他人排除率0.1% FRR@FAR0.1%
MSM	相互部分空間法	93.56	3.01	7.53
CMSM	制約相互部分空間法	95.65	1.00	2.59
WMSM	白色化相互部分空間法	96.90	0.67	1.76
PWMSM	疑似白色化相互部分空間法	96.99	0.59	1.67
MWMSM	多重白色化相互部分空間法	97.07	0.59	1.34
MPWMSM	多重疑似白色化相互部分空間法	97.07	0.59	1.51

参照部分空間の全てを学習用部分空間とし、その基底ベクトルの本数 N_C は 180 とした。学習用部分空間の基底ベクトルの本数 N_B は 15 とした。

(c) 白色化相互部分空間法 (WMSM)

(b) の制約相互部分空間法と同一の学習用部分空間を用いて、白色化変換を計算し、すべての入力部分空間と参照部分空間の特徴変換を行い相互部分空間法により類似度を算出した。入力部分空間と参照部分空間の基底ベクトルの本数 N はともに 7 とした。

(d) 疑似白色化相互部分空間法 (PWMSM)

(b)(c) と同一の学習用部分空間を用いて、白色化変換を計算し、白色化変換の特徴次元を 1~253 までを用いる。すべての入力部分空間と参照部分空間の特徴変換を行い相互部分空間法により類似度を算出した。入力部分空間と参照部分空間の基底ベクトルの本数 N はともに 7 とした。

(e) 多重白色化相互部分空間法 (MWMSM)

白色化変換を学習部分空間から並列的に学習し、複数の白色化制約相互部分空間法により類似度を算出し、その平均類似度を結果とする。 $L = 1196$ 人分の学習用部分空間をランダムに選択し、約 240 名程度の学習用部分空間から合計 $M = 5$ 個の白色化変換を生成した。

(f) 多重疑似白色化相互部分空間法 (MPWMSM)

(e) で用いた白色化変換に対して、複数の疑似白色化制約相互部分空間法により類似度を算出し、その平均類似度を結果とする。(e) と同様に学習用部分空間から合計 $M = 5$ 個の白色化変換を生成し、その一部の要素のみを用いる疑似白色化変換を用いる。

表 18 は、それぞれ MSM(相互部分空間法), CMSM(制約相互部分空間法), WMSM(白色化相互部分空間法), PWMSM(疑似白色化相互部分空間法), MWMSM(多重白色化相互部分空間法), MPWMSM(多重疑似白色化相互部分空間法) による認識結果である。

図 74 は、それぞれの ROC カーブを表しており、ベンチマークデータにおいては、MWMSM による多重かによるアンサンブル認識の結果が最も EER が低く性能が改善

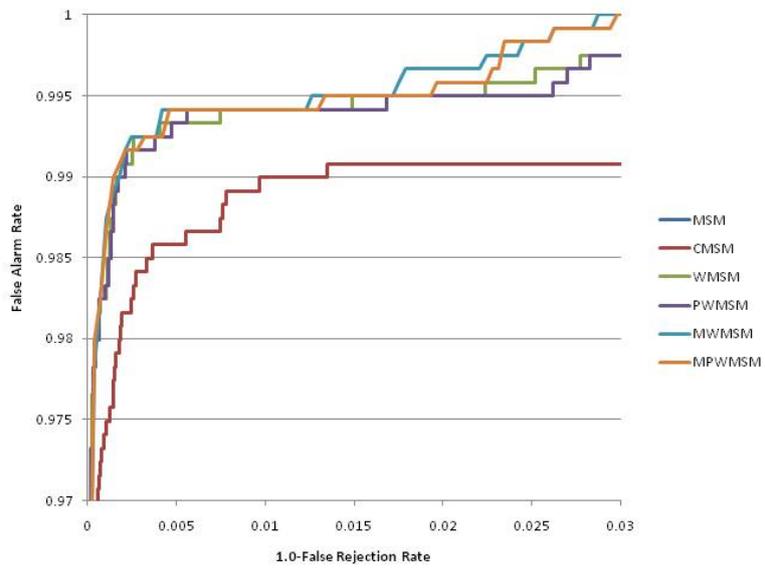
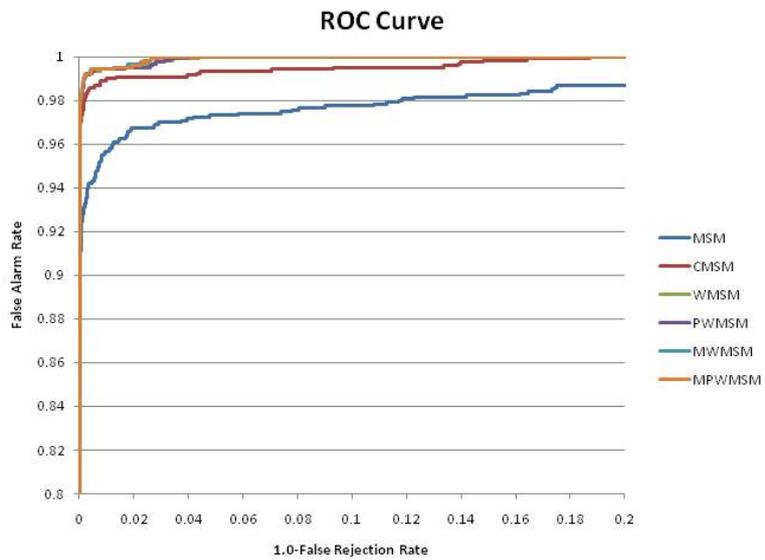


図 74: FERET fafb の ROC Curve(下図は部分拡大図)

していることがわかる。疑似白色化変換については、学習部分空間の十分に次元が高いため、白色化変換による結果とあまり変わらない結果となった。

5.3 本章のまとめ

本章では、三つの視点に立って2～4章で提案した手法を用いて、多様な変化を行う具体的対象として「顔認識」に取り組んだ。2章で提案した画像マッチングにより実装当時低速であった計算資源においても高速な顔検出を実現し、ポータブルPCにおいて4章で述べた相互部分空間法を動作させるための工夫ならびにエンターテインメント向けの機能についても提案を行った。

さらに公開ベンチマークデータを用いて、2、3章で提案した手法を用いて、顔画像のセグメンテーションを行い、正規化顔画像パターンを対象として、4章で述べた新たに拡張した特徴抽出法によるアルゴリズムの認識実験を行い、従来法に比べて精度向上を確認した。この識別アルゴリズムの一部は、米国の顔認識ベンチマークテスト FRVT2006 にエンジンとして提出し、低解像度画像での評価は最も低いエラーレートであった [74]。

6 結論

本論文では、物体認識における各プロセスにおける不変的な性質に着目し、三つの視点に立って、従来の認識処理を見直しを行い、多様に変化する対象を識別する高精度なアルゴリズムを開発し、段階的に問題を解決した。

第一の視点については、コンピュータビジョンの中で基本的な技術である画像マッチングを問題の対象とし、対象を構成する局所的な領域について、マッチングの測度に関して、明るさの変化や輝度変換などに影響を受けない方法を提案した。マッチングに要求される高速性を考慮したうえで、変動への対応を行う手法について考察を行った。

第二の視点については、幾何学的特徴点ベースの手法に、アピアランスベースのアプローチを効果的に統合した新しいモデル表現法を構築し、不変量をインデクシングに利用することにより効率的な部分画像表現を提案した。そのモデルを用いて部分隠れやノイズに対してロバストな認識方法を示した。

第三の視点については、対象を識別するパターン認識技術として、相互部分空間法を取り上げ、複合的に起こる変動に対して不変的な特徴を抽出するための特徴抽出法について、実用的に動作するための高精度化の検討を行った。

2章では、第一の視点に立って、画像マッチングのための不変量について考察し、画像マッチングのための測度として、隣接画素の同値関係、大小関係に基づいた定性的3値表現による類似度を提案した。各画像に対して水平、垂直方向の隣接画素値の差の符号に基づいて3値を割り当てた2つの3値画像を求め、3値画像の一致度合いにより画像の類似性を求める。これは、部分的な遮蔽やいくつかの輝度変換に対して耐性をもつ表現となる。本表現の性質を述べ、いくつかの実験を通して、従来法と比較を行った。本手法は、比較、加減演算のみで計算できるため、計算コストが低くなる。また、本手法を前処理として定量的なマッチングと組み合わせることで、高速かつ精度のよいハイブリッドマッチングが実現できる可能性を示した。

3章では、第二の視点から、対象の位置や大きさなどの変換に対する不変性を導入することを検討した。幾何学的特徴ベースの手法にアピアランスベースの考え方を導入した物体認識の枠組みであるパターンハッシング法を提案した。モデル画像を特徴点を基準にして、複数の部分画像に分割する。次に大きさと向きを正規化した各部分画像を、基準とした特徴点に付随する幾何学的不変量をインデックスとしてハッシュテーブルに登録する。認識時には、テーブルに登録された膨大な部分画像モデルをハッシングにより高速に選択し、抽出された部分画像とのマッチングを効率的に行う。そして、この部分的なマッチング結果を投票手続きにより集積し、入力画像から対象の位置を検出する。検証実験により、隠れのある複数物体を安定に検出できることを確認した。

4章では、第三の視点から、複合的に起こる変動の中から識別に重要な不変的な特徴を取り出すための特徴抽出法に関する研究を行った。複数の制約部分空間の生成に対してアンサンブル学習を適用し、複数の制約部分空間を用いて特徴抽出を多重化した多重制約相互部分空間法を提案した。提案法では、入力部分空間と参照部分空間を複数の制約部分空間へ射影することで特徴抽出を行う。制約部分空間に射影された入力部分空間と参照部分空間の類似度を算出し、制約部分空間の個数だけ求めた類似度を結合することで最終的に類似度を決定する。複数の制約部分空間を生成するために、アンサンブル学習の代表的な手法である並列的な学習と逐次的な学習の枠組みを導入した。提案手法により従来の認識手法に比べて識別精度が向上することを、照明変動が生じる環境で撮影されたデータベースと大規模人数が登録されたデータベースを用いた識別実験で確認した。さらに、相互部分空間法の高性能化に関して、線形の特徴抽出系をもつ白色化相互部分空間法の改良法を示した。非線形性の強いデータやカテゴリ数やデータのサンプル数が少ない場合に、白色化相互部分空間法は、従来のいくつかの相互部分空間法よりも性能が劣るケースが存在する。特徴抽出の計算に用いる基底ベクトルの自己相関行列の固有値の大きさの制御に着目し、白色化変換によって得られる基底に対する特徴選択によって疑似白色化変換を適用する疑似白色化相互部分空間法を提案した。特殊なケースであった性能劣化する場合についても従来法と比較して改善することを確認し、さらに多重化を行った多重疑似白色化相互部分空間法を提案した。

5章では、上述した三つの視点に立って2～4章で提案した手法を用いて、多様な変化を行う具体的対象として「顔画像認識」に取り組んだ。2章で提案した定性的3値表現による画像マッチングによって、実装当時低速であった計算資源においても高速な顔検出を実現し、ポータブルPCにおいて4章で述べた相互部分空間法を動作させるための工夫ならびにエンターテインメント向けの機能を提案した。さらに公開ベンチマークデータを用いて、2、3章で提案した手法を用いて、顔画像の検出、顔部品検出といったセグメンテーションを行い、正規化顔画像パターンを対象として、4章で新たに拡張した特徴抽出法によるアルゴリズムの認識実験を行い、従来法と比較して精度向上することを確認した。これらの技術はいくつかの製品において実用化、製品化されており、社会の中で活用されている。

今後の課題として、

- 2章（特に2.6章）で試行した不変量と深層学習との関連についてのさらなる考察が必要である。今般の深層学習においては、さまざまな画像変換を学習時に適用することによって、汎化性能を向上させる方策がとられているが、不変量を導入することによって、さらなる学習の効率化と高性能化が期待できると推察する。

- 3章の Pattern Hashing については、2章で導入した不変表現を用いた深層学習特徴によるマッチング手法などを導入することで、高性能化が考えられる。今後、実装、評価を行い、実応用への課題解決に利用されることを期待する。
- 4章の相互部分空間法の拡張についても、適切な部分空間の次元数、多重化の射影変換の個数など、いわゆるハイパーパラメータに関する議論を深め、より一般化させたアルゴリズムへの発展を期待する。

今後、様々なメディアを対象としたパターン認識技術を実用化し、より現実世界で動作する画像認識技術を実現していくため、さらなる発展を目指していく。

謝辞

本研究および論文執筆にあたり、ご指導ご鞭撻を賜りました、筑波大学 福井和広教授に心から感謝致します。同教授には、著者の株式会社東芝入社当時の先輩として、ご指導ならびに多数のご助言をいただき、社会人としての研究生活のスタートをバックアップいただくところから、博士論文を仕上げるところにまでお付き合いをいただくことになろうとは考えておりませんでした。感謝の念に堪えません。

また、論文審査に際し、有益なご意見、ご議論をいただきました、筑波大学 亀山啓輔教授、滝沢穂高教授、北原格教授、佐藤雄隆教授（産業技術総合研究所）に深く感謝致します。

奈良先端科学技術大学院大学 木戸出正継名誉教授、大阪芸術大学芸術学部 萩田紀博教授、豊田工業大学 浮田宗伯教授には、過日、奈良先端大での博士課程取得に際し、多大なるご支援をいただきながらも、完遂することができず、その不甲斐なさには弁解の余地がございません。自責の念に駆られ、再度、博士取得にチャレンジしましたことをご報告するとともに、あらためて感謝致します。特に、木戸出正継名誉教授には、東芝採用時からお世話になり、博士論文執筆を強く勧めてくださるとともに、叱咤激励を欠かすことなく頂戴いたしましたこと、感謝致します。

パターン認識の世界に入るきっかけをつくってくださったのは、岡山大学在学時の「松山研」であり、京都大学名誉教授 松山隆司教授、奈良県立大学 浅田尚紀教授、和歌山大学 和田俊和教授には、学生時代だけでなく、会社生活を始めてからも、多大なるご指導をいただいたこと、深く感謝いたします。3章で論じた Pattern Hashing のアイデアに辿り着いたのは、松山研での得た多くの知見がベースになっていると感じています。また、松山研に関係する諸先生方、先輩の方々、同期のみなさま、後輩の皆様にも感謝の意を表します。

本論文の研究の多くは、株式会社東芝研究開発センターにおいて実施したものです。東芝にてご指導いただきました、前田賢一博士、浮田輝彦博士、渡辺睦博士（鹿児島大学教授）、小野口一則博士（弘前大学教授）、岡本恭一氏、中井浩章氏、倉立尚明博士、風間久氏に深く感謝致します。特に、前田賢一博士には、4、5章に関連する相互部分空間法の顔画像認識への適用の道筋を与えていただき、研究上の議論、初めての英語予稿の作成時のご助言など、さまざまな場面での会社における研究の進行に関するご指導、ご支援をいただきました。深く感謝致します。

在籍したラボラトリーの歴代の室長である、麻田治男博士、尾高敏則氏、赤嶺正巳博士（東北大学教授）、江草俊博士、堀修博士、金澤博史氏、伊藤剛氏、西浦正英氏、平和

樹氏、岡田隆三博士、小坂谷達夫氏には、研究を進める上で、素晴らしい環境を作っていただいたことに感謝の意を表します。

顔認識と一緒に取り組んだメンバとして、鈴木薫氏、湯浅真由美氏、窪田進氏、牧淳人博士（スウェーデン KTH 教授）、田中英治氏、若杉智和氏、山田貢己博士、中島朗子博士、小坂谷達夫氏、西山正志博士（鳥取大学教授）、河原智一氏、武口智行博士、柴田智行氏には、数々の研究に関する有益な議論、実システムの実装、評価などだけではなく、楽しい思い出の数々をつくっていただき感謝致します。また、現在も一緒に取り組んでくれている伊藤聡氏、齊藤廣大博士、山地雄士氏をはじめとする現ラボのメンバにも感謝致しております。さらには、顔認識システムの実用化に関しては、岡崎彰夫博士、榎本暢芳氏、佐藤俊雄博士、助川寛氏には、多大なるご指導やご支援をいただき、感謝致します。

両親である山口弘、山口智恵子、また兄である、剛、哲男、和男、そして義姉の佳子には、小さな頃からの多大なる支援と最大限の我儘を認めてくださり、この謝意は言葉では言い尽くせません。

最後に、娘、息子である、蒼乃、絢弘、璃子、そして妻の和美には、家族としていつもながらの温かいサポートに感謝しています。ありがとう。

参考文献

- [1] M. K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory*, Vol. IT-8, pp. 179–187, 1962.
- [2] Y. Lamdan and H. J. Wolfson. Geometric hashing: a general and efficient model-based recognition scheme. *Proc. of ICCV*, pp. 238–249, 1988.
- [3] H. Murase and S. K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, Vol. 14, pp. 5–24, January 1995.
- [4] H. A. Rowley, S. Baluja, and T. Kanade. Rotational invariant neural network-based face detection. *Proc. of CVPR*, pp. 38–44, 1998.
- [5] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86, 1991.
- [6] 柳井啓司. 一般物体認識の現状と今後. 情報処理学会論文誌：コンピュータビジョンとイメージメディア, Vol. 48, No. SIG 16(CVIM 19), 2007.
- [7] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [8] 前田賢一, 渡辺貞一. 局所的構造を導入したパターン・マッチング法. 信学論 (D), Vol. J68-D, No. 3, pp. 345–352, 1985.
- [9] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. *IEEE Proceedings of the 3rd. International Conference on Face and Gesture Recognition*, pp. 318–323, 1998.
- [10] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. *11th International Symposium of Robotics Research (ISRR)*, pp. 192–201, 2003.
- [11] 坂井利之, 長尾真, 木戸出正継. 電子計算機による濃淡図形の処理 (顔写真の場合). 電子通信学会論文誌 C, Vol. 54C, No. 6, pp. 445–452, 1971.
- [12] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, Vol. 24, No. 4, pp. 325–376. (白井 (訳): 「画像の位置合わせ手法の概観」, コンピュータサイエンス acm computing surveys '92 bit 別冊, pp. 77–120).

- [13] A. Venot, J. F. Lebruchec, and J. C. Roucayrol. A new class of similarity measures for robust image registration. *Computer Vision, Graphics, Image Processing*, Vol. 28, pp. 176–184, 1984.
- [14] 村瀬一朗, 金子俊一, 五十嵐悟. 増分符号相関によるロバスト画像照合. *信学論 D-II*, Vol. 83-D-II, No. 5, pp. 1323–1331, 2000.
- [15] P. Lipson, E. Grimson, and P. Sinha. Configuration based scene classification and image indexing. *Proc. of CVPR*.
- [16] P. Sinha. Object recognition via image invariants: A case study. *Investigative Ophthalmology Visual Science*, Vol. 35, pp. 1735–40, May 1994.
- [17] 奥田晴久, 広岡美和子, 橋本学, 鷺見和彦, 吉倉冬彦. 3値・濃淡ハイブリッドマッチングによるパターン同一性判定手法. 第10回「外観検査の自動化」ワークショップ, pp. 91–95, 1998.
- [18] 鷺見和彦, 橋本学. 3値化 blob を中間表現に持つシーン認識手法. *情報処理学会論文誌*, Vol. 34, No. 10, pp. 2085–2095, 1993.
- [19] 佐藤雄隆, 金子俊一, 五十嵐悟. 周辺増分符号相関画像に基づくロバスト物体検出及び分離. *信学論 D-II*, Vol. J84-D-II, No. 12, pp. 2585–2594, 2001.
- [20] 山口修, 福井和広. 顔向きや表情の変化にロバストな顔認識システム”smartface”. *信学論 D-II*, Vol. J84-D-II, No. 6, pp. 1045–1051, 2001.
- [21] A. Krizhevsky. Learning multiple layers of features from tiny images. *Master’s thesis, University of Tront*, 2009.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [23] Haotong Qin, Ruihao Gong, Xianglong Liu, Xiao Bai, Jingkuan Song, and Nicu Sebe. Binary neural networks: A survey. *Pattern Recognition*, Vol. 105, p. 107281, Sep 2020.
- [24] Amir Gholami, Sehoon Kim, Zhen Dong, Zhewei Yao, Michael W. Mahoney, and Kurt Keutzer. A survey of quantization methods for efficient neural network inference, 2021.

- [25] Diwen Wan, Fumin Shen, Li Liu, Fan Zhu, Jie Qin, Ling Shao, and Heng Tao Shen. Tern: Convolutional neural network with ternary inputs and binary weights. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pp. 322–339, Cham, 2018. Springer International Publishing.
- [26] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. PAMI*, Vol. 19, No. 5.
- [27] C. J. Harris and M. Stephens. A combined corner and edge detector. *In Proc. 4th Alvey Vision Conference, Manchester*, pp. 147–151, 1988.
- [28] J. Matas, J. Buriánek, and J. Kittler. ‘object recognition using the invariant pixel-set signature. *Proc. of BMVC*.
- [29] 山口修, 福井和広. 定性的 3 値表現による画像マッチング. 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 102, No. 156, pp. 23–30, 2002.
- [30] K. Fukui and O. Yamaguchi. Facial feature point extraction method based on combination of shape extraction and pattern matching. *Systems and Computers in Japan*, Vol. 29, No. 6.
- [31] S. A. Nene, S. K. Nayar, and H. Murase. Columbia object image library (coil-100).
- [32] E. Oja. *Subspace Methods of Pattern Recognition*. Research Studies Press, England, 1983.
- [33] 福井和広, 山口修. 一般化差分部分空間に基づく制約相互部分空間法. 信学論 D-II, Vol. J87-D-II, No. 8, pp. 1622–1631, 2004.
- [34] K. Fukui, B. Stenger, and O. Yamaguchi. A framework for 3d object recognition using the kernel constrained mutual subspace method. *Proc. of ACCV06*, p. 315–324, 2006.
- [35] K. Fukui and O. Yamaguchi. The kernel orthogonal mutual subspace method and its application to 3d object recognition. *Proc. of ACCV07*, pp. 467–476, 11 2007.
- [36] 赤松茂. コンピュータによる顔の認識 –サーベイ–. 信学論 D-II, Vol. J80-D-II, No. 8, pp. 2031–2046, 1997.

- [37] 福井和広, 山口修, 鈴木薫, 前田賢一. 制約相互部分空間法を用いた環境変動にロバストな顔画像認識- 照明変動を抑える制約 部分空間の学習-. 信学論 (D-II), Vol. J82-D-II, No. 4, pp. 613–620, 1999.
- [38] 小坂谷達夫, 山口修, 福井和広. 制約相互部分空間法を用いた顔認識システムの開発と評価. 情報処理学会論文誌, Vol. 45, No. 3, pp. 951–959, 2004.
- [39] 佐藤俊雄, 助川寛, 横井謙太朗, 土橋浩慶, 緒方淳, 岡崎彰夫. ‘立ち位置変動を考慮した顔照合セキュリティシステム「facepass」の開発. 映像情報メディア学会誌, Vol. 56, No. 7, pp. 1111–1117, 2002.
- [40] S. Watanabe and N. Pakvasa. Subspace method of pattern recognition. *Proc. 1st Int. J. Conf. on Pattern Recognition*, 1973.
- [41] M. Turk and A. Pentland. Face recognition using eigenfaces. *IEEE Proc. Conf. on Computer Vision and Pattern Recognition*, pp. 586–591, 1991.
- [42] T. Shakunaga and K. Shigenari. Decomposed eigenface for face recognition under various lighting conditions. *IEEE Proc. Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 864–871, 2001.
- [43] J.P. Hespanha P.N. Belhumeur and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711–720, 1997.
- [44] 相馬正宣, 長尾健司. 取得過程の極端に異なる顔画像の照合. 信学論 D-II, Vol. J80-D-II, No. 8, pp. 2225–2231, 1997.
- [45] 麻生英樹, 津田宏治, 村田昇. パターン認識と学習の統計学 新しい概念と手法. 岩波書店, 2003.
- [46] L. Breiman. Bagging predictors. *Machine Learning*, Vol. 24, No. 2, pp. 123–140, 1996.
- [47] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, Vol. 55, .
- [48] X. Wang and X. Tang. Random sampling lda for face recognition. *IEEE Proc. Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 259–265, 2004.

- [49] G.-D. Guo and H.-J. Zhang. Boosting for fast face recognition. *Second International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems*, 2001.
- [50] 石井健一郎, 上田修功, 前田英作, 村瀬洋. わかりやすいパターン認識. オーム社, 1998.
- [51] 前田英作, 村瀬洋. カーネル非線形部分空間法によるパターン認識. 信学論 D-II, Vol. J82-D-II, No. 4, pp. 600–612, 1999.
- [52] H. Sakano, N. Mukawa, and T. Nakamura. Kernel mutual subspace method and its application for object recognition. *Electronics and Communications in Japan*, Vol. 88, pp. 45–53, 2005.
- [53] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 409–415, 2003.
- [54] K. Fukui and O. Yamaguchi. Comparison between constrained mutual subspace method and orthogonal mutual subspace method. *Proc. of Subspace 2006*, pp. 63–71, 2006. (In Japanese).
- [55] T. Kawahara, M. Nishiyama, T. Kozakaya, and O. Yamaguchi. Face recognition based on whitening transformation of distribution of subspaces. *Subspace2007(8th ACCV Workshop)*, pp. 97–103, 2007.
- [56] Xudong Jiang, Bappaditya Mandal, and Alex Kot. Eigenfeature regularization and extraction in face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 3, pp. 383–394, 2008.
- [57] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. *In IEEE Conference on Computer Vision and Pattern Recognition workshops*, pp. 806–813, 2014.
- [58] Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson. Factors of transferability for a generic convnet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 9, pp. 1790–1802, 2016.

- [59] Naoya Sogi, Taku Nakayama, and Kazuhiro Fukui. A method based on convex cone model for image-set classification with cnn features. *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2018.
- [60] Akinari Sakai, Naoya Sogi, and Kazuhiro Fukui. Gait recognition based on constrained mutual subspace method with cnn features. In *2019 16th International Conference on Machine Vision Applications (MVA)*, pp. 1–6, 2019.
- [61] Gibran Benitez-Garcia, Jesus Olivares Mercado, Gabriel Sanchez-Perez, and Keiji Yanai. Ipn hand: A video dataset and benchmark for real-time continuous hand gesture recognition. *Proc. of IAPR International Conference on Pattern Recognition (ICPR)*, pp. 4340–4347, 2021.
- [62] W. Hu, Y. Huang, F. Zhang, R. Li, W. Li, and G. Yuan. Seqface: Make full use of sequence information for face recognition. *CoRR*, Vol. abs/1803.06524, , 2018.
- [63] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4690–4699, 2019.
- [64] InsightFace. Insightface model zoo. *GitHub repository*, 2021.
- [65] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, Vol. 23, No. 10, pp. 1499–1503, 2016.
- [66] 長谷川修, 森島繁生, 金子正秀. 「顔」の情報処理. 信学論 D-II, Vol. J80-D-II, No. 8, pp. 2047–2065, 1997.
- [67] 土居元紀, 陳謙, 眞溪歩, 大城理, 佐藤宏介, 千原國宏. 顔画像照合による解錠制御システム. 信学論 D-II, Vol. J80-D-II, No. 8, pp. 2203–2208, 1997.
- [68] 山口修, 福井和広, 前田賢一. 動画像を用いた顔認識システム. 信学技報, Vol. PRMU97-50, pp. 17–24, 1997.
- [69] 福井和広, 山口修. 形状情報とパターン照合の組合せによる顔特徴点抽出. 信学論 D-II, Vol. J80-D-II, No. 8, pp. 2170–2177.
- [70] K.Sung and T.Poggio. Example-based learning for view-based human face detection. *Technical Report A.I.Memo No.1521, A.I. Lab., MIT*, 1994.

- [71] 山口修, 福井和広. 経年変化を抑制する顔辞書更新法. 信学技報, No. PRMU99-25, pp. 43-50, 1999.
- [72] 田畑尚弘. アミューズメントにおける顔研究. 信学会ヒューマンコミュニケーション技術セミナー, HC のための「顔」大研究予稿集, Mar. 1999.
- [73] 小坂谷達夫, 湯浅真由美, 西山正志, 河原智一, 山口修. 1枚の顔画像による特定人物検索システム ~国際顔認識ベンダーテスト (frvt2006) に提出した顔認識エンジン. 画像の認識・理解シンポジウム MIRU2007, DS-09, 2007.
- [74] P. Jonathon Phillips, W. Todd Scruggs, Alice J.O ' Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, Matthew Sharpe. "frvt 2006 and ice 2006 large-scale results march 2007, nistir 7408, national institute of standards and technology. (<http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf>).

発表論文リスト

2章に関連する発表・論文

山口 修, 福井 和広, ” 定性的 3 値表現に基づく画像マッチング,” 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 102(156),pp. 23-30 (2002).

Osamu Yamaguchi, Kazuhiro Fukui, “Pattern hashing: Object recognition based on a distributed local appearance model,” Proceedings of IEEE International Conference on Image Processing (ICIP02) , Vol . III, pp. 329-332, (2002)

3章に関連する発表・論文

山口 修, 福井 和広, ” パターンハッシング – 部分画像と不変量索引を用いた分散アピランスモデル – ,” 画像の認識・理解シンポジウム MIRU 2002, E2-4 (2002).

(MIRU 2002 Excellent papers)

■情報処理学会 山下記念研究賞受賞 (本論文が対象論文)

山口修, 福井和広: “パターンハッシング: 部分画像と不変量索引を用いた分散アピランスモデル,” 情報処理学会論文誌. コンピュータビジョンとイメージメディア, Vol. 44(SIG5(CVIM6)), pp.64-73 (2003)

4章に関連する発表・論文

Osamu Yamaguchi, Kazuhiro Fukui, Ken-ichi Maeda, “Face Recognition Using Temporal Image Sequence,” Proceeding of IEEE Conference of Face and Gesture Recognition(FG’98), pp. 318-323 (1998).

Masashi Nishiyama, Osamu Yamaguchi, Kazuhiro Fukui, ”Face Recognition with the Multiple Constrained Mutual Subspace Method,” Audio- and Video-Based Biometric Person Authentication, 5th International Conference, AVBPA 2005, Hilton Rye Town, NY, USA, July 20-22, 2005, Proceedings. Lecture Notes in Computer Science 3546 Springer , pp. 71–80 (2005)

西山 正志, 山口 修, 福井 和広, “多重制約相互部分空間法を用いた顔画像認識,” 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J88-D-II(8),pp. 1339-1348(2005)

山口 修, 河原 智一: 白色化相互部分空間法における特徴選択に関する考察, 信学技報 PRMU, パターン認識・メディア理解研究会 109(470), pp. 211-216 (2010)

Osamu Yamaguchi, Kazuhiro Fukui, Image-set based Classification using Multiple Pseudo-whitened Mutual Subspace Method. In Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods, ISBN 978-989-758-549-4, ISSN 2184-4313, pages 296-306. DOI: 10.5220/0010836500003122 (2022).

5章に関連する発表・論文

Osamu Yamaguchi, Face Recognition Technology and Its Real-World Application, Perception and Machine Intelligence -First Indo-Japan Conference, PerMin2012, Kolkata, India, January 12-13, 2012. p p.28-34,Proceedings. Lecture Notes in Computer Science 7143, Springer 2012, ISBN 978-3-642-27386-5

その他の発表文献リスト

【修士課程での研究発表】

<口頭発表>

山口修, 松山隆司: 代数的制約記述に基づく配色デザインモデル (II), 1992 年電子情報通信学会春季大会, A-303, 1992.3.25

松山隆司, 山口修: 代数的制約記述に基づく配色デザインシステム, 情報処理学会グラフィックスと CAD シンポジウム, pp.31-40, 1992.9.24

<論文>

松山隆司, 山口修: 代数的制約記述に基づく配色デザインシステム, 情報処理学会論文誌, Vol.34, No.5, pp.941-953, 1993.5

【東芝入社後の研究発表】

<主な口頭発表>

山口修, 福井和広: 分離度特徴を用いた顔画像解析: 目、瞳の検出, 全国大会講演論文集 第 52 回平成 8 年前期 (2),187-188,19960306

■情報処理学会 第 52 回全国大会 (平成 8 年前期 於: 電通大) 大会優秀賞受賞

山口修, 福井和広, 前田賢一: 動画像を用いた顔認識システム, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 97(113),17-24,19970620

知野 哲朗, 福井 和広, 山口 修, 鈴木 薫, 田中 克己: "GAZE TO TALK" メタコミュニケーション能力を持つ非言語 メッセージ利用インタフェース, 情報処理学会 インタラクション '98, 1998

■情報処理学会 インタラクション '98 ベストペーパー賞

山口修, 福井和広: 経時変化を抑制する顔辞書更新法, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 99(118),43-50,19990617

山口修, 福井和広: 動画像を用いた PC 顔認識システム "Smartface", 情報処理学会研究報告. CVIM,[コンピュータビジョンとイメージメディア] 2000(50),99-106,20000531

岩井儀雄, Lao Shihong, 山口修, 平山高嗣: 画像処理による顔検出と顔認識 (サーベイ (2)), 情報処理学会研究報告. CVIM, [コンピュータビジョンとイメージメディア] 2005(38),343-368,20050512

西山正志, 山口修: 物体表面の見え属性分類に基づく Self Quotient Image を用いた顔認識: 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 105(374),33-38,20051020

■PRMU 研究奨励賞受賞論文 (筆頭西山氏表彰)

武口智行, 湯浅真由美, 山口修: 角点を持つ顔特徴点の検出, 第 6 回 計測自動制御学会

(SICE) システムインテグレーション部門講演会 SI2005, 3I4-4, 2005

■ SI2005 ベストセッション講演賞受賞

西山正志, 湯浅真由美, 若杉智和, 柴田智行, 山口修: 断片的な動画像の対応付けを利用した歩行者認識, 第12回 画像センシングシンポジウム SSII06, 2006.

■第12回 画像センシングシンポジウム 優秀論文賞

小坂谷達夫, 山口修: 基準3次元モデルを用いた姿勢と表情の正規化による顔認識, 画像の認識・理解シンポジウム MIRU2006, IS2-49, 2006.

小坂谷達夫, 湯浅真由美, 西山正志, 河原智一, 山口修: 1枚の顔画像による特定人物検索システム ~ 国際顔認識ベンダーテスト (FRVT2006) に提出した顔認識エンジン, 画像の認識・理解シンポジウム MIRU2007, DS-09, 2007

■ MIRU2007 デモセッション賞受賞

山口 修, "局所領域特徴を用いた顔方向認識" 第8回 計測自動制御学会 (SICE) システムインテグレーション部門講演会, 2H3-5 (2007).

山口 修, "Delaunay 三角形分割を導入したパターンハッシング", MIRU2009, IS1-41, (2009)

河原 智一, 山口 修, "最小誤差量子化による特徴ベクトル圧縮の有効性の解明とパターン認識への応用", MIRU2009, IS1-1, (2009)

柴田 智行, 山口 修, "パターン選択を用いた近似k近傍識別手法", MIRU2009, IS2-1, (2009)

山口 修, 河原 智一: 白色化相互部分空間法における特徴選択に関する考察, 信学技報 PRMU, パターン認識・メディア理解研究会 109(470), pp. 211-216, 2010-03-08.

齊藤 廣大, 助川 寛, 山口 修・佐藤 俊雄, "組合せ最適化アプローチによる複数人物顔追跡", 信学技報 PRMU, パターン認識・メディア理解研究会 109(470), pp. 217-225, 2010-03-08.

齊藤 廣大, 助川 寛, 山口 修, 佐藤 俊雄, 榎本 暢芳, "組合せ最適化アプローチによる複数人物顔追跡システム", MIRU2010, IS3-46 (2010)

<解説記事、サーベイなど>

山口修: 動画像を用いた顔認識システム, 画像ラボ 1998年3月号

山口修, 小坂谷達夫: 顔画像による個人認証, ヒューマンインタフェース学会誌 7(2), 101-106, 20050525

山口修: 顔画像認識のセキュリティ応用とその展開, 映像情報インダストリアル, 2007/01/25 発売号 (2月号)

橋本 周司, 北川 克一, 山口 修, 他: パネルディスカッション 膨張するビジョントラゲット-工場から町へ (特集2 ViEW2006 ビジョン技術の実利用ワークショップ), 映像情

報 industrial 39(2), 74-83, 2007-02

山口 修, 西山 正志, 河原 智一: 歩行者を見分ける顔認識技術, (特集 デジタルメディア処理先端技術) 東芝レビュー 62(12), 38-41, 2007-12

Lao Shihong, 山口 修: 実利用が進む顔画像処理とその応用事例: (前編) 顔画像処理技術の動向, 情報処理 50(4), 319-326, 2009-04-15

Lao Shihong, 山口 修: 実利用が進む顔画像処理とその応用事例: (後編) 顔画像処理の応用事例, 情報処理 50(5), 436-443, 2009-05-15

山口 修: 近年の顔画像認識技術の概要, (特集 ここまで来た!顔情報処理技術の最先端) O plus E 31(12), 1403-1407, 2009-12

山口修: 顔画像認識の過去、現在、未来, 応用物理 88(12), 816-819, 2019

<論文>

福井和広, 山口修: 形状抽出とパターン照合の組合せによる顔特徴点抽出, 電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理 J80-D-2(8),2170-2177,19970825

福井和広, 山口修, 鈴木薫, 前田賢一: 制約相互部分空間法を用いた環境変動にロバストな顔画像認識: 照明変動の影響を抑える制約部分空間の学習, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J82-D-II(4),613-620,19990425

山口修, 福井和広: 顔向きや表情の変化にロバストな顔認識システム”Smartface” 電子情報通信学会論文誌. D-II, D-II-84(6), 1045-1052, 2001

■電子情報通信学会論文賞受賞

小坂谷達夫, 山口修, 福井和広: 制約相互部分空間法を用いた顔認識システムの開発と評価, 情報処理学会論文誌 45(3),951-959,20040315

福井和広, 山口修: 一般化差分部分空間に基づく制約相互部分空間法, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J87-D-II(8),1622-1631,20040801

福井和広, 山口修: カーネル非線形制約相互部分空間法による物体認識, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J88-D-II(8),1349-1356,20050801

若杉智和, 山口修, 福井和広: パターンエネルギーを導入した動的輪郭モデルによる唇輪郭抽出, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J88-D-II(9),1857-1865,20050901

福井和広, 山口修: 部分空間法の理論拡張と物体認識への応用, 情報処理学会論文誌. コンピュータビジョンとイメージメディア 46(SIG.15(CVIM.12)),21-34,20051015

前田 賢一, 山口修, 福井和広: 部分空間の正準角を利用した三次元パターンマッチングの一基礎検討: 顔と顔写真との区別を目的として, 電子情報通信学会論文誌. D, 情報・システム. J89-D(6),1288-1296,20060601

西山 正志, 湯浅 真由美, 柴田智行, 若杉 智和, 山口 修: 顔画像の階層的な対応付けを

用いた複数歩行者の認識, 電子情報通信学会論文誌. D, 情報・システム, J90-D(8),2191-2201,2007

<国際会議発表>

Hiroaki Nakai, Yasuhiro Taniguchi, Michihiro Uenohara, Takashi Yoshimi, Hideki Ogawa, Fumio Ozaki, Junji Oaki, Hirokazu Sato, Yukio Asari, Katsuhiko Maeda, Hiroyuki Banba, Toshi Okada, Kyoichi Tatsuno, Eiji Tanaka, Osamu Yamaguchi, Mitsuyoshi Tachimori,: A Volleyball Playing Robot., IEEE ICRA 1998: 1083-1089

YUASA Mayumi, YAMAGUCHI Osamu, FUKUI Kazuhiro: Precise pupil contour detection based on minimizing the energy of pattern and edge, IAPR MVA2002.

Kazuhiro Fukui, Osamu Yamaguchi: Face Recognition using multi-viewpoint patterns for robot vision, 11th International Symposium of Robotics Research (ISRR'2003), pp.192-201, 2003.

Ken-ichi Maeda, Osamu Yamaguchi, Kazuhiro Fukui: "Toward 3-Dimensional Pattern Recognition", SSPR2004 & SPR2004, p.1061-1068, 2004.

Miki Yamada, Osamu Yamaguchi, Akiko Nakashima, Kazuhiro Fukui: Head pose estimation using adaptive scaling template matching, IAPR MVA2005.

Masashi Nishiyama, Osamu Yamaguchi: Face Recognition Using the Classified Appearance-based Quotient Image, IEEE FGR06(49-54).

Tatsuo Kozakaya, Osamu Yamaguchi: Face Recognition by Projection-based 3D Normalization and Shading Subspace Orthogonalization, IEEE FGR06(163-168).

Kazuhiro Fukui, Bjorn Stenger, Osamu Yamaguchi: A Framework for 3D Object Recognition Using the Kernel Constrained Mutual Subspace Method, ACCV06(II:315-324).

Masashi Nishiyama, Mayumi Yuasa, Tomoyuki Shibata, Tomokazu Wakasugi, Tomokazu Kawahara, Osamu Yamaguchi, Recognizing Faces of Moving People by Hierarchical Image-Set Matching, Biometrics07(1-8).

■ Biometrics '07 Best paper award

Mayumi Yuasa, Tatsuo Kozakaya and Osamu Yamaguchi: An Efficient 3D Geometrical Consistency Criterion for Detection of a Set of Facial Feature Points, IAPR MVA2007

Kazuhiro Fukui, Osamu Yamaguchi: The Kernel Orthogonal Mutual Subspace Method and Its Application to 3D Object Recognition. ACCV (2) 2007: 467-476

Mayumi Yuasa, Osamu Yamaguchi: Real-time face blending by automatic facial feature point detection. IEEE FG 2008: 1-2

Tatsuo Kozakaya, Tomoyuki Shibata, Mayumi Yuasa, Osamu Yamaguchi: Facial fea-

ture localization using weighted vector concentration approach. IEEE FG 2008: 1-6

Mayumi Yuasa, Osamu Yamaguchi: Real-time face blending by automatic facial feature point detection. IEEE FG 2008: 1-6

Tomoyuki Shibata, Osamu Yamaguchi: Local Fisher Discriminant Component Hashing for Fast Nearest Neighbor Classification. SSPR/SPR 2008: 339-349

Masashi Nishiyama, Hidenori Takeshima, Jamie Shotton, Tatsuo Kozakaya, Osamu Yamaguchi: Facial deblur inference to improve recognition of blurred faces. CVPR 2009: 1115-1122

Tatsuo Kozakaya, Satoshi Ito, Susumu Kubota, Osamu Yamaguchi: Cat face detection with two heterogeneous features. ICIP 2009: 1213-1216

Hitoshi Sakano, Osamu Yamaguchi, Tomokazu Kawahara, Seiji Hotta: On the Behavior of Kernel Mutual Subspace Method, ACCV Workshops (2) 2010: 364-373 [-] 2000 – 2009

Yuka Kohno, Osamu Yamaguchi, Toshio Sato, Bunpei Irie: Alignment of 3D Shape Data by Hashing Sets of Feature Points MVA2011, 4-22, 120-123 (2011)

Viet-Quoc Pham, Tatsuo Kozakaya, Osamu Yamaguchi, Ryuzo Okada: COUNT Forest: CO-Voting Uncertain Number of Targets Using Random Forest for Crowd Density Estimation, ICCV 2015: 3253-3261

Tomokazu Kawahara, Osamu Yamaguchi: Feature Vector Compression Based on Least Error Quantization, CVPR Workshops 2016: 84-92

<論文 (英文) >

Kazuhiro Fukui, Osamu Yamaguchi: Facial feature point extraction method based on combination of shape extraction and pattern matching. Systems and Computers in Japan 29(6): 49-58 (1998)

Osamu Yamaguchi, Kazuhiro Fukui: "Smartface" -A Robust Face Recognition System under Varying Facial Pose and Expression(Special Issue on the 2001 IEICE Excellent Paper Award) IEICE transactions on information and systems, Vol.E86-D, No.1(20030101) pp. 37-44, 2003

■電子情報通信学会論文賞受賞

M. Yuasa, O. Yamaguchi, K. Fukui,: "Precise pupil contour detection based on minimizing the energy of pattern and edge," IEICE Transactions on Information and Systems, vol.E87-D, no.1, pp.105-112, 2004.

Mayumi Yuasa, Tatsuo Kozakaya, Osamu Yamaguchi: An Efficient 3D Geometrical Consistency Criterion for Detection of a Set of Facial Feature Points. IEICE Transactions

91-D(7): 1871-1877 (2008)

Tatsuo Kozakaya, Tomoyuki Shibata, Mayumi Yuasa, Osamu Yamaguchi: Facial feature localization using weighted vector concentration approach. *Image Vision Comput.* 28(5): 772-780 (2010)

Masashi Nishiyama, Abdenour Hadid, Hidenori Takeshima, Jamie Shotton, Tatsuo Kozakaya, Osamu Yamaguchi: Facial Deblur Inference Using Subspace Analysis for Recognition of Blurred Faces. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(4): 838-845 (2011)