

部分空間情報の適応的解析を用いた
画像の認識と強調に関する研究

井上 晃

システム情報工学研究科

筑波大学

2009年3月

概要

本論文では、画像の認識及び強調を目的とし、画像の部分空間情報の適応的解析による、環境変動への適応性実現について述べる。人間が画像を見る場合、全体を一様に眺めているわけではない。その時点でのタスクに応じて着目する部分を適応的に変化させている。画像認識では、対象物体に対して適切な部分領域を選択し、領域毎に照合を行うことにより、遮蔽や環境変化が起きても正しく認識することが可能となる。画像強調では、例えば物体の輪郭部分などのエッジ領域だけに着目することで、鮮鋭度の強調量を適確に判断することができる。

本論文では、このような視覚の部分領域への注目処理は、画像空間中から適切な部分空間を選択する操作であると考え、そしてこれらの部分空間情報を、適応的に解析・統合することによって、変動する環境下でも様々なタスクを正しく実行できると考える。そこで適応性実現に向けて、部分空間の選択の高度化、部分空間の比較・統合、部分空間に基づく適応動作、の3つのアプローチに基づいた、新しい画像認識と画像強調の処理方式を提案する。

画像認識では、様々な対象の中でも環境変動の影響が非常に大きいとされる、人間の顔照合・顔追跡を対象として適応性の改善を試みる。静止画顔照合では、適切な部分領域を選択して領域毎に照合を行う適応的部分領域照合法と、画像から予測される変動画像を生成してそれらの画像群の張る部分空間を用いて照合する摂動空間法を提案する。これらの手法により、遮蔽や照明・姿勢変動に対する頑健性を向上させる。動画顔照合では、部分空間同士の新しい比較方法として、相互投影距離法を提案する。また顔追跡では、特徴点の動きベクトルを剛体運動の部分空間に射影することにより、6つの運動パラメータを高速に求める手法について述べる。

画像強調では、カラー画像の自動高画質化を目的とし、様々な環境下で撮影されたカラー画像に対して、劣化に関する事前知識なしに最適な補正パラメータを算出して強調処理を行う手法について述べる。従来この種の強調処理は、熟練技術者の多大な労力をかけたマニュアル調整が必要であった。そこで本研究では、画質を構成する複数の要因について考察し、各画質要因に影響を及ぼす部分領域を選択・解析することによって、画質特性値を算出する。さらにその特性値が最適値となるように、強調パラメータを自動制御する方法を提案する。画質要因として、鮮鋭度、彩度、コントラスト、明度、ホワイトバランス、記憶色の6種類を用いる。自然画像による画像処理結果を用いた主観評価実験を実施し、様々なシーンの画像に対して自動高画質化が達成されることを示す。

以上、部分空間情報の適応的解析を用いることにより、変動に頑健な画像認識処理と画像強調処理を実現する。

第1章	序論.....	1
第2章	画像の部分空間情報の解析.....	4
2.1.	まえがき.....	4
2.2.	部分空間情報.....	4
2.2.1.	画像座標空間の部分領域選択.....	4
2.2.2.	特徴ベクトル空間の部分空間選択.....	5
2.2.3.	部分空間への射影による認識.....	5
2.3.	適応性実現に向けたアプローチ.....	6
2.4.	画像認識における研究内容.....	7
2.5.	画像強調における研究内容.....	8
2.6.	まとめ.....	9
第3章	部分空間の適応的解析を用いた画像認識.....	10
3.1.	まえがき.....	10
3.2.	適応的部分領域照合と摂動空間法による静止画顔照合.....	10
3.2.1.	静止画顔照合の関連研究.....	10
3.2.2.	適応的部分領域照合法.....	12
3.2.3.	摂動空間法.....	15
3.2.4.	部分領域における摂動空間法.....	21
3.2.5.	顔照合実験.....	22
3.2.6.	静止画顔照合のまとめ.....	25
3.3.	部分空間への相互投影による動画顔照合.....	25
3.3.1.	動画顔照合の関連研究.....	25
3.3.2.	相対する分布との距離.....	27
3.3.3.	擬似マハラノビス距離.....	28
3.3.4.	相互投影距離法.....	28
3.3.5.	動画顔照合への適用.....	30
3.3.6.	照合実験.....	30
3.3.7.	演算量の比較.....	36
3.3.8.	動画顔照合のまとめ.....	37
3.4.	動きベクトルの部分空間射影による高速顔追跡.....	38
3.4.1.	高速顔追跡の関連研究.....	38
3.4.2.	動き推定方法.....	39
3.4.3.	顔追跡システムの実装.....	44
3.4.4.	実験結果.....	48
3.4.5.	高速画像追跡のまとめ.....	49
3.5.	画像認識のまとめ.....	50

第4章	部分領域からの画質特性解析による画像強調	52
4.1.	まえがき	52
4.2.	エッジ領域の解析による画像の最適鮮鋭化	53
4.2.1.	鮮鋭度の最適化	53
4.2.2.	エッジシャープネスの定義	53
4.2.3.	主観的鮮鋭さとエッジシャープネス	55
4.2.4.	画質とエッジシャープネス	57
4.2.5.	鮮鋭度変換アルゴリズム	58
4.2.6.	実験結果	60
4.2.7.	最適鮮鋭化のまとめ	62
4.3.	HSV座標系を用いた彩度・コントラスト・鮮鋭度による自動高画質化	62
4.3.1.	HSV座標系を用いた自動高画質化の概要	62
4.3.2.	HSV座標系	63
4.3.3.	彩度の調整	64
4.3.4.	コントラストの調整	65
4.3.5.	鮮鋭度の調整	66
4.3.6.	3種類の画質改善処理の統合	67
4.3.7.	実験結果	67
4.3.8.	HSV座標系を用いた自動高画質化まとめ	69
4.4.	任意色に対する特色分離補正	71
4.4.1.	選択的色補正の関連研究	71
4.4.2.	6色相別色補正	72
4.4.3.	任意色に対する特色分離補正・HSV方式	73
4.4.4.	任意色に対する特色分離補正・RGB差分方式	75
4.4.5.	特色度の分布	77
4.4.6.	色補正結果	78
4.4.7.	色補正インターフェース	79
4.4.8.	任意色に対する特色分離補正のまとめ	82
4.5.	6つの画質特性解析に基づく自動高画質化	82
4.5.1.	自然画像の自動強調における課題	82
4.5.2.	画質特性解析と補正方法	83
4.5.3.	ホワイトバランス補正	84
4.5.4.	コントラスト補正	85
4.5.5.	明るさの補正	86
4.5.6.	彩度補正	86
4.5.7.	記憶色の補正	87

4.5.8.	鮮鋭度補正	89
4.5.9.	強調パラメータのリミッタ	90
4.5.10.	画像処理結果	91
4.5.11.	並列解析を用いた高速アーキテクチャ	92
4.5.12.	処理速度の比較	95
4.5.13.	実験結果	96
4.5.14.	6つの画質要因解析による自動高画質化まとめ	101
4.6.	画像強調のまとめ	101
第5章	結論と今後の展望	102
謝辞	104
参考文献	105
公表論文リスト	108

第1章 序論

近年、デジタルカメラの普及や演算装置の低コスト化などにより、監視装置や検査装置として画像認識システムの高度化が進行している。人間の視覚は、周辺の照明環境が変動しても、対象物が移動・変形しても、一部が隠れていても、正しく情景を理解することができる。しかしながら、機械による画像認識システムは、人間と比較して周辺環境や認識対象の変動に対する頑健性が低く、利用可能な環境に制約がある。すなわちマシンビジョンシステムにとって、様々な環境変動に対する適応能力の向上が、大きな課題となっている。

人間の視覚システムは、情景を観察する際に全体を一様に眺めているわけではなく、特定の部分領域に逐次注視点を動かすことによって効率的に情報を獲得している。空間的に注視点を動かすだけでなく、注目する部分領域の大きさを変更しながら、顕著な特徴領域を探索している。そして視覚システムの注視動作は、見ている情景とその時点で行っている目的・タスクに依存して適応的に決定されると考えられる。

人間の形状知覚は選択的である。その一例として図と地の分化という現象がある[1]。人間の視野の中に異なる構造が存在すると、視野内にいくつかの部分領域が生じる。こうして生じた視野の領域は2種類に分かれて知覚される。一つは物として浮き出る領域で図(**figure**)と呼ばれ、もう一つは背景として見える領域で地(**ground**)と呼ばれる。この現象を図と地の分化と呼び、図となる領域は形状を知覚できるが、地の領域は形状を知覚できないことが知られている。また観察する人間の意識の中で図と地が反転すると、別の形状が知覚され、新たな物体が現れたように感じられる。このように、部分領域の選択と統合が、人間の視覚システムにおいて重要な働きをしていることは明らかである。

また人間の網膜には、光エネルギーを膜電位に変換する視細胞が、空間的に稠密に配置されている。そして、視覚神経系の大脳皮質 V1 野においては、輪郭の方位、波長(色)、空間周波数成分に対する選択性を持つ細胞の存在が確認されている[2]。すなわち視覚神経系において、実際に、位置情報やエッジ情報、色情報を用いて部分領域を選択し、解析を行っている可能性を示唆している。

以上述べた人間の視覚システムに対する考察より、人間は、部分領域に対する適応的な注視動作によって、様々な認識処理を環境によらずに正確に実行できると考える。従って、マシンビジョンシステムが様々な変動要因の下でタスクを実行するには、適切な部分領域を選択し、解析することが重要な鍵になる。

そこで本論文では、画像の認識処理及び強調処理において、部分空間情報の適応的解析による環境適応の実現方法について述べる。カメラで撮影された画像データを用いて様々なタスクを行うマシンビジョンシステムにおいて、このような視覚の適応処理を実現することにより、実用性の高いシステムを構築できることが期待される。なお本論文では、マシンビジョンシステムが取り扱う画像データ及び特徴量全体を画像空間と呼び、画像空間中の一部を部分空間と呼ぶことにする。

図 1-1 に、本研究の、部分空間情報の適応的解析の概念図を示す。本概念図は、マシンビジョンシステムにおいて、カメラ撮影画像から得られる部分空間を、状況に応じて選択・統合しながらタスクを実行している様子を表している。本研究では、この部分空間情報の選択・統合による適応的解析処理を用いることで、様々な変動に対して頑健なシステムの実現を目指す。

マシンビジョンにおけるタスク例として、物体認識、人物検出、顔認識、画像強調、色補正、画像合成、などがある。本研究ではこの中で、画像認識処理と画像強調処理とを対象として環境適応性の改善を図る。ここで画像認識とは、画像中に写った物体を、決められたカテゴリに識別する作業である。また画像強調とは、画像中に写った物体の視認性を、向上させる作業である。

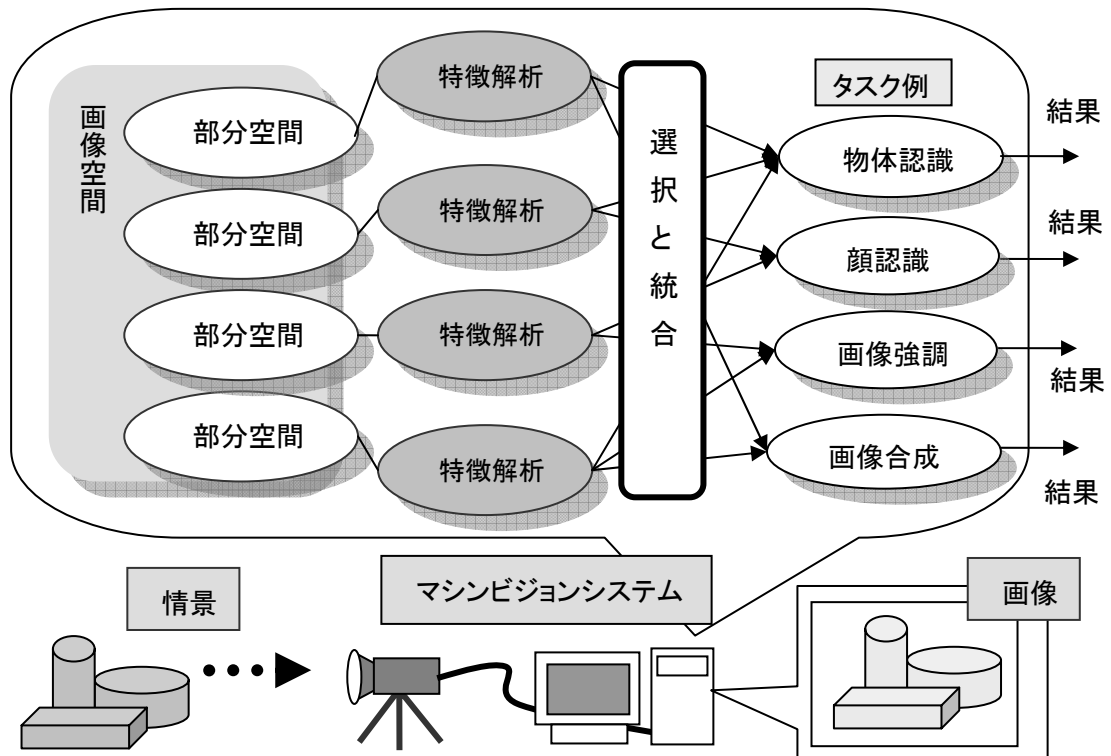


図 1-1: 部分空間情報の適応的解析の概念図

画像中に写る物体の変動要因として、撮影環境変動（照明とカメラ）、姿勢変動、変形・遮蔽がある。画像認識処理においては、これら全ての変動要因に対処することが必要である。そこで本研究の画像認識処理では、対象物体の適切な部分領域を選択し、領域毎に正規化した上で照合を行うことにより変形・遮蔽に対応する。さらに照明と姿勢に対する変動画像群を予測して部分空間を構築することにより、撮影環境変動と姿勢変動に対応する。

画像強調処理においては、撮影画像内容に応じた強調処理を行う必要がある。強調パラメータに対しては、変動要因の中で撮影環境変動の影響が最も大きく、姿勢変動と変形・遮蔽の影響は小さい。但し、記憶色の強調処理のように対象物体に依存する強調処理も存在するため、姿勢や遮蔽の影響も皆無ではない。本研究の画像強調処理では、複数の画質要因を、それぞれ画質特性値として評価するステップを持つことが特徴である。この画質特性値の評価ステップにおいて、部分領域情報を用いる。例えば鮮鋭度強調を行う場合には、物体の輪郭部分等のエッジ領域だけに注目することによって、画像の鮮鋭度を自動判断する。この画質特性値により、撮影環境変動に対応する。さらに、色空間における部分空間情報を適切に用いることで、対象物体にのみ作用する強調処理を実現する。

注目すべき部分領域のみを選択して処理することで、演算量も削減され、瞬時に適切な判断を行うことができる。タスクによっては、複数の部分領域に同時に注目する必要があるため、最終的な判断は、これらの複数の部分領域を統合して全体としての判断が行われる。

本論文では、このように部分領域情報の適応的解析によって、視覚の適応性に近づけるものとする。具体的には、どの部分領域を選択するのか、どのように部分領域を統合するのか、どのように部分領域情報を解析するのが問題である。本研究は、視覚の適応性を実現する一般的なモデルを構築するのが目的ではない。想定する実際の画像処理タスクにおいて、部分空間情報を用いて頑健な適応性を実現することが目標である。

以下、第2章では、部分空間情報を用いた適応性の実現について検討し、部分空間の選択、部分空間の比較・統合、部分空間に基づく動作、の3種類のアプローチについて考察する。次に第3章では画像認識における成果について述べ、第4章では画像強調における成果について述べる。

第3章では、画像認識における具体例として、様々な対象の中でも環境変動の影響が非常に大きいとされる人間の顔照合・顔追跡を課題に適応性の改善を試みる。3.2節では、静止画顔照合の改善手法として、適応的部分領域照合法と摂動空間法を提案する。適応的部分領域照合法は、適切な部分領域を選択して領域毎に照合を行う方法であり、部分遮蔽や表情変化等への頑健性を向上させる。摂動空間法は1枚の画像を元に変動要因を反映した摂動画像群を生成し、それらの画像群の張る部分空間（摂動空間）を用いて照合を行う方法である。3.3節では、動画像照合における部分空間同士の比較方法として、相互投影距離法を提案する。3.4節では、物体追跡における新たな部分空間解析手法を提案する。

第4章では画像強調における具体例として、様々な環境下で撮影されたカラー画像を、劣化に関する事前知識なしに最適な強調パラメータを算出し、強調処理を行う手法について述べる。4.2節では、高画質化のための最適鮮鋭化を提案する。4.3節では、HSV色空間を用いた彩度、コントラスト、鮮鋭度によるカラー画像の自動高画質化手法について述べる。4.4節では、画像中の部分領域への色補正を目的とした、任意色の特色補正法を提案する。4.5節では、鮮鋭度、彩度、コントラスト、明度、ホワイトバランス、記憶色の6種類の画質要因を改善するカラー画像の自動強調手法の実現について述べる。ここでは、画像中の特定の部分領域の解析によってそれぞれの要因に対応した画質特性値を定義し、その画質特性値が最適値となるような強調パラメータの算出方法を提案する。そしてさらに、高速アーキテクチャ導入による、PC上でのリアルタイム処理の実現方法について述べる。

第5章では、本研究の成果を総括し、今後の課題と展望について述べる。

第2章 画像の部分空間情報の解析

2.1. まえがき

顔画像、文字、物体など、様々な対象物に対する画像認識システムの研究が、これまでに数多くなされている。これらの画像認識システムでは認識精度を向上させるため、対象や環境を制限した上で、アルゴリズム最適化が図られてきた。具体的には、(1)認識対象の特定(数字、文字、指紋、検査部品、正面顔、マーカなど)、(2)撮影環境の限定(照明環境の限定、カメラ位置の固定、物体位置の限定、一様背景の利用(前景分離容易化))、などである。

従来の画像認識システムにおいても、一般環境に適応させるため、変動に対する頑健性を向上させる試みが行われてきた。例えば、画像全体の画素値の正規化や、エッジ特徴などの環境変動に強い特徴量の抽出等である[3]。これらの試みは一定の効果があり多くの実用システムに適用されているが、変動に対する部分的な対策に留まっているため、一般環境への適応には不十分であった。

一方、特徴抽出やパターン照合法の研究成果として、局所領域情報の利用や、学習サンプルを用いた部分空間法など[4]、部分空間の解析手法が提案された。しかしながら、従来の部分空間の解析手法は、あらかじめ決められたパラメータによって固定的に解析するものであり、一般環境への適応を意図したものとは言えなかった。そこで本研究では、入力画像に応じた適応的な解析方法を採用することにより、一般環境の変動に適応させることを試みる。

本章では、まず部分空間情報解析の従来手法について概観した後、画像認識と画像強調における、適応性改善のアプローチについて述べる。

2.2. 部分空間情報

本論文における部分空間は、画像空間において、所定のタスク実行に必要な情報を含む部分集合と定義する。部分空間は、空間座標の取り方によって画像座標空間と、特徴ベクトル空間とに分類することができる。画像座標空間における部分空間の一例として、画像中の特定のパターンを含む部分領域や、特定の統計量を有する部分領域がある。また特徴ベクトル空間における部分空間の一例として、少数の主成分ベクトルで張られる部分空間がある。本節では、それぞれの部分空間情報の選択方法と解析手法について概観する。

2.2.1. 画像座標空間の部分領域選択

画像座標空間は、画素単位の要素ベクトルで構成される空間である。画像座標空間における部分空間は、部分領域と呼ばれる。部分領域を選択する方法として、特定のパターンとの相関値の高い領域を抽出する、テンプレートマッチングがある。また、エッジ検出やコーナ検出など、特定の性質を持つ領域を検出する画像処理手法も、部分領域の選択方法と考える。また、特定の画像統計量(画素値の最大値、最小値、中間値など)を有する領域を、部分領域として選択する方法がある。

部分空間を選択するためのパラメータ(テンプレート、統計量など)は通常固定であり、あらかじめ、認識対象に合わせて適切な値に調整しておく必要がある。そのため、環境変動が大きい場合や、画像中のノイズ成分が大きい場合は、しばしば部分領域の選択に失敗し、後段における誤認識の原因となる。

2.2.2. 特徴ベクトル空間の部分空間選択

特徴ベクトル空間は、画素値を変換して得られる特徴ベクトルによって張られる空間である。画像座標空間と異なり、撮影画像中の物体の位置関係は反映されない。従って、部分空間の選択には通常、統計解析が用いられる。

特徴ベクトル空間における部分空間の選択方法として、特徴ベクトルの主成分分析や、判別分析を用いる方法がある[4][5]。撮影画像から得られた認識対象の特徴ベクトル群を主成分分析すると、ある少数の固有ベクトルによって寄与率が 90%を超える場合が多い。そこで、寄与率の大きな少数の固有ベクトルによって張られる部分空間を選択して用いることが、しばしば行われる。判別分析は、クラス内分散に対してクラス間分散を最大化するような、特徴空間を求める処理である。判別分析で得られた特徴空間を、そのまま部分空間として選択することができる。

主成分分析を行うには、ある単一クラス（通常は認識対象）に属する複数の特徴ベクトルを収集し、統計解析する必要がある。また判別分析では、複数のクラス毎に特徴ベクトルを収集し、同様に統計解析する必要がある。複数のサンプルから主成分分析や判別分析を用いて部分空間を求める操作を、学習と呼ぶことがある。本論文では、学習も部分空間の選択方法の一つであると考える。

2.2.3. 部分空間への射影による認識

特徴ベクトルの部分空間を用いた認識手法として、いわゆる部分空間法がある[4]。部分空間法には射影距離法と角度法とがある。いずれも、認識対象のカテゴリが属する部分空間と、テストパターンとの類似度を比較することで、テストパターンがそのカテゴリに属するかどうかを判断する。

部分空間法の射影距離法は、テストデータのパターン部分空間への射影距離によって照合を行う。この射影距離は、図 2-1(a)に示すように、テストデータ A の部分空間への射影距離 D_k である。認識処理では射影距離 D_k が最も小さいクラスを結果として出力する。部分空間の次元数は、認識率が最良となるようにあらかじめ求めておく。

部分空間法の角度法は、テストデータと部分空間とが張る角度によって類似性を計算する。類似度 S は図 2-1(b)に示すように、特徴ベクトル空間における、部分空間とテストデータ A との角度 $\cos^2 \theta$ と等価である。類似度 S は 0 から 1 までの値を返却する。この値が最も大きいクラスを照合結果として出力する。

部分空間法は、認識対象カテゴリに属する学習サンプルが多数得られる場合に用いられ、変動が大きい環境でも比較的良い認識性能を得ることができる。また、行例の線形演算のみで実装できるため、演算量が小さいという利点がある。

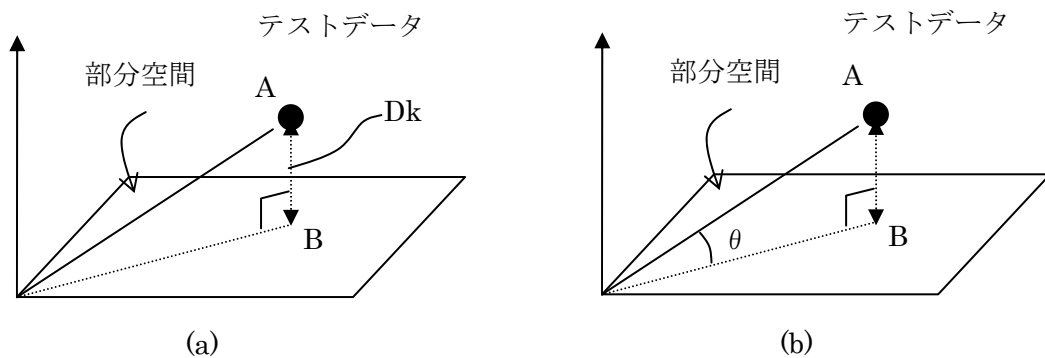


図 2-1: 部分空間法 の概念

2.3. 適応性実現に向けたアプローチ

本論文では、部分空間情報の適応的解析を用いて、画像認識と画像強調における視覚の適応性を再現する方法について述べる。前節で述べた従来の部分空間解析は、あらかじめ決められた固定パラメータや、学習サンプルを用いて処理していたため、実環境における様々な変動（照明や変形、遮蔽など）を反映させることができなかつた。このような変動要因への対応を実現するため、部分空間の適応的解析を次のようなアプローチに基づいて検討した。

1. 部分空間選択の高度化
2. 部分空間の比較・統合
3. 部分空間に基づく適応動作

本論文では、画像認識と画像強調という 2 つのタスクについて、上記のアプローチに基づく画像処理手法を提案する。表 2-1 に、各アプローチに対応する提案手法を分類したものを示す。本節では、表 2-1 に基づいて、タスク別に提案手法の概要を述べる。

まず、部分空間選択の高度化について説明する。従来、画像処理タスクで行われる部分空間の選択は、あらかじめ固定されたパラメータによって実行されていた。例えば、決められた特徴量を用いた部分領域選択や、あらかじめ登録された学習サンプルの統計解析によって選択する部分空間などである。そこで本研究では、タスクに用いる部分空間の選択を、入力画像に応じて適応的に決定することを考える。

画像認識における部分空間選択の高度化として、部分領域照合法と、摂動空間法を提案する。部分領域照合法は、画像を複数の部分領域に分割しておき、部分領域毎の照合スコアを元に、選択する部分領域を動的に変化させる手法である。摂動空間法は、登録された画像に対して予想される変動を推定し、その変動によって張られる部分空間を適応的に求める方法である。

また、画像強調における部分空間選択の高度化として、6 つの画質要因に対応した画質特性値の算出手法を提案する。すなわちこれらの画質特性値算出のため、適切な部分領域選択を検討する。

次に、複数の部分空間の比較・統合について説明する。このアプローチは、異なる複数の部分空間情報の、比較・統合によって適応性能を改善する。画像認識においては、部分空間同士の新しい距離尺度を用いた、相互投影距離法を提案する。本手法は、従来の部分空間法のように単一ベクトルと部分空間との比較ではなく、部分空間同士を直接比較することができる。また、前述の適応的部分領域照合法は、複数の部分空間情報を統合する機能も備えている。

画像強調における部分空間の比較・統合アプローチとして、自動高画質化に向けた 6 つの画質特性解析の統合処理について述べる。

最後に、部分空間に基づく適応動作について説明する。このアプローチは、部分空間情報を選択した後の、動作制御の改善に焦点を当てた研究である。画像認識では、剛体運動する物体の

動きを定義した部分空間を利用することによって、運動パラメータを高速に求める画像追跡手法を提案する。

画像強調における適応動作の研究として、画像の最適鮮鋭化手法、任意色の特色補正法、自動高画質化における強調パラメータ制御法について述べる。画像の最適鮮鋭化は、エッジ領域という部分領域の高周波解析によって、画像の最適鮮鋭化を実現する。任意色の特色分離補正は、色空間における部分領域内での重み制御を行うことにより、自然な色補正を実現する。そして自動高画質化は、部分領域情報から、鮮鋭度、彩度、コントラスト、明度、ホワイトバランス、記憶色の6種類の画質要因を数値化する画質特性値を算出し、それらの値に基づいて、複数の強調パラメータを最適に制御する。

このように本研究では、実環境における様々な変動を反映させるため、3つのアプローチに基づいた部分空間の適応的解析を行う。2.4節では本研究の画像認識処理の研究内容について概観し、2.5節では画像強調処理の研究内容について概観する。

表 2-1: 適応性実現に向けたアプローチ

アプローチ	画像認識	画像強調
部分空間選択の高度化	適応的部分領域照合法 摂動空間法	自動高画質化での画質解析
部分空間の比較・統合	相互投影距離法 適応的部分領域照合法	画質解析の統合
部分空間に基づく適応動作	部分空間射影による高速顔追跡	最適鮮鋭化 任意色の特色分離補正 強調パラメータ自動制御

2.4. 画像認識における研究内容

マシンビジョンにおける画像認識は、入力画像データがどの登録カテゴリに属するのか、を判断するタスクである。画像認識の対象物が何であるかは、あらかじめ既知の情報としてシステムに与えられなければならない。そこで本研究では、様々な認識対象の中でも環境変動の影響が非常に大きいとされる人物顔を、画像認識の対象とする。そして、顔照合と顔追跡を課題として、変動要因への適応性改善を試みる。顔照合の研究は、その歴史は長いですが、近年の監視カメラによるセキュリティ需要拡大により、現在注目を集めている研究分野である。

画像認識では、具体的には次のテーマについて研究を行った。(1)適応的部分領域照合法と摂動空間法を用いた静止画顔照合(3.2節)、(2)部分空間の相互射影による動画顔照合(3.3節)、(3)動きベクトルの部分空間射影による高速顔追跡(3.4節)、である。

(1)適応的部分領域照合法と摂動空間法を用いた静止画顔照合は、1枚の静止画を用いた顔照合性能向上を目的とする。画像空間における主な変動要因は、撮影環境変動(照明とカメラ)、姿勢変動、変形・遮蔽である。対象が人物顔の場合には、これら全ての変動要因が認識性能に影響を及ぼす。本研究では、適応的部分領域照合法と摂動空間法とによって、撮影環境変動(照明変動)、姿勢変動、変形・遮蔽に対応する。本研究は、部分空間選択の適応化と、部分空間の比較・統合のアプローチに基づいている。

(2)部分空間の相互射影による動画顔照合は、部分空間同士の新しい距離尺度を用いた相互投影距離法により、動画像を用いた顔照合性能の向上を目的とする。本研究は、動画に含まれる変動要因を吸収することが可能であり、撮影環境変動(照明変動)と微小な姿勢変動に対して効果がある。本研究は、部分空間の比較・統合のアプローチに基づいている。

(3)動きベクトルの部分空間射影による高速顔追跡は、高速顔追跡システムの頑健性向上を目的とする。特徴点の動きベクトルを剛体運動の部分空間に射影することで、頑健な姿勢推定が可能であり、特に姿勢変動への適応化を改善する。本研究は、部分空間に基づく適応動作のアプローチに基づいている。

以上の研究内容については、第3章において詳細に述べる。

2.5. 画像強調における研究内容

マシンビジョンにおける画像強調は、入力画像を人間が視認し易くなるように変換するタスクである。画像強調は、認識の前処理として利用される場合もあるが、本研究では、画像処理の出力を人間が観察することを前提とする。具体的には、画像の高画質化処理を自動的に行うシステムを目指した研究を行う。画像の高画質化処理は、カラープリンタやデジタルカメラ、デジタルTV等のデジタル画像機器の普及に伴い、利用場面が拡大している技術分野である。

画像強調において、処理対象となる画像は一般の自然画像であり、画像中に写る対象物体を限定することは難しい。従ってこれらの画像群を、自動的に最適に強調するためには、高度な適応性が必要となる。そのためにはまず、複数の画質要因それぞれに応じて画像システムが考慮すべき部分領域を正しく選択することが重要である。そして、選択された部分領域の情報を用いて、どのように強調パラメータを算出するのが課題となる。

画像強調に関しては、次のテーマにおける研究を行った。(1)エッジ領域の解析による画像の最適鮮鋭化(4.2節)、(2)HSV座標系を用いた彩度・コントラスト・鮮鋭度による自動高画質化(4.3節)、(3)任意色に対する特色分離補正(4.4節)、(4)6つの画質特性解析に基づく自動高画質化(4.5節)、である。

(1)エッジ領域の解析による画像の最適鮮鋭化は、自然画像に対し、劣化の事前知識なしに最適な鮮鋭度強調処理の実現を目的とする。本研究は、エッジ領域という部分領域を選択し、解析することによって、画像の最適鮮鋭化を実現する。鮮鋭度強調処理は、撮影環境変動の一部である撮像機器の変動(フォーカス等)に対する適応性を改善する手法である。本手法は、部分空間選択の高度化と、部分空間に基づく適応動作の両方のアプローチに関連する。

(2)HSV座標系を用いた彩度・コントラスト・鮮鋭度による自動高画質化では、HSV色空間を用いて、自然画像の3つの画質要因を、自動的に改善することを目的とする。(1)の鮮鋭度強調に加え、コントラストと彩度の特性値を評価するための部分領域選択・解析と、これらの特性値を元にした強調パラメータ制御を行う。本手法は、画像強調における変動要因の中で、撮影環境に関する変動要因(照明・カメラ)に対する適応性を改善する。本手法は、部分空間選択の高度化と、部分空間に基づく適応動作の両方のアプローチに関連する。

(3)任意色に対する特色分離補正は、部分色空間における重み付けを利用して、特定の物体色を、少ない操作で高品質に補正することを目的とする。本手法では、指定した色を中心として部分空間が選択され、その空間内の色情報を用いて色補正の重みを制御する。画像強調の変動要因の中で、対象物体領域における撮影環境変動を、補正することができる。また本手法は、次に述べる自動高画質化において、記憶色の補正に用いられている。すなわち、記憶色を基準として、撮影環境変動の適応性を改善していると考えられる。なお本手法は、部分空間に基づく適応動作のアプローチに関連している。

(4)6つの画質特性解析に基づく自動高画質化は、自然画像に対して鮮鋭度、彩度、コントラスト、明度、ホワイトバランス、記憶色の6種類の画質要因を、自動的に改善することを目的とする。まずこれらの画質要因に影響を及ぼす部分領域を選択し、画質特性解析を行って、画

質特性値を算出する。次にそれらの特性値を最適値に補正するような、強調パラメータ制御を行う。そして最終的に、6つの画質特性解析を統合した自動強調手法を提案する。本手法は、画像強調における変動要因の中で、撮影環境変動への適応性を改善することができる。また本手法は、部分空間選択の高度化、部分空間の比較・統合、部分空間に基づく適応動作の全てのアプローチに関する手法を含んでいる。

以上、画像強調における研究については、第4章において詳細に述べる。

2.6. まとめ

人間は、情景内における部分領域の情報を適切に選択し、解析することで、物体認識時における状況変化に柔軟に対処している。しかしながら、従来のマシンビジョンシステムでは、認識精度を向上させるため、対象や環境を制限した上で、アルゴリズム最適化が図られていた。変動に対する頑健性を向上させる試みも存在したが、一般環境への適応を積極的に意図したものではなかった。

本章では、このような環境適応性を実現するため、次の3つのアプローチ、(1)部分空間選択の高度化、(2)部分空間の比較・統合、(3)部分空間に基づく適応動作、に基づいた部分空間の適応的解析について検討した。そしてこれらのアプローチに従い、画像認識と画像強調の2つのタスクにおける提案手法の概要を示した。また本提案手法によって、適応性改善が期待される変動要因について考察した。次章以降では、画像認識と画像強調における具体的な画像処理手法について、詳細に述べる。

第3章 部分空間の適応的解析を用いた画像認識

3.1. まえがき

本章では、画像認識における具体的な適応性改善の研究について述べる。画像認識における様々な対象の中でも、環境変動の影響が非常に大きいとされる、人間の顔照合・顔追跡を課題として、適応性の改善を試みる。

はじめに静止画顔照合の改善手法として、適応的部分領域照合法と摂動空間法を提案する。適応的部分領域照合法は、適切な部分領域を選択して領域毎に照合を行う方法であり、遮蔽や表情変化等への頑健性を向上させる効果がある。摂動空間法は1枚の画像を元に、予想される画像変動を適用した摂動画像群を生成し、それらの画像群の張る部分空間（摂動空間）を用いて照合を行う方法である。摂動空間法により、照明変動や姿勢変動に対する頑健性を向上させることができる。

次に、動画像照合における部分空間同士の比較方法として、相互投影距離法を提案する。本手法は部分空間同士を直接比較するため、動画像特有のフレーム間相関を、照合スコアに反映させることができる。照合実験により、従来法に比べて照合精度を維持したまま、大幅に演算量を削減可能であることを示す。

最後に、物体追跡における、新たな部分空間解析手法について述べる。特徴点の動きベクトルを剛体運動の部分空間に射影することで、剛体運動における6自由度の運動パラメータを高速に求めることができる。本手法を人物顔の追跡に応用し、実時間での顔追跡システムを構築する。

3.2. 適応的部分領域照合と摂動空間法による静止画顔照合

3.2.1. 静止画顔照合の関連研究

近年、非接触で人物同定可能な顔照合システムの需要が、急速に高まっている。画像を用いた顔照合については、従来から数多くの研究[6][7]が報告されているが、監視カメラ等を用いた実環境における顔照合では、未だに十分な照合率が得られていないのが現状である。その理由として、以下に示す顔画像特有な見えの違いの存在が挙げられる。すなわち、1) 照明と撮影環境による見えの違い、2) 顔向きの変化による見えの違い、3) 眼鏡装着や表情による見えの違いである。

図3-1に実環境で撮影された同一人物の顔画像の例を示す。図3-1の上段からは、明らかに照明環境による著しい画像パターンの変化が見られる。また下段は、顔向き変化や部分的な遮蔽、眼鏡装着の一例である。



図 3-1: 実環境での顔画像例

実環境での顔照合を実現するため、このような照明、顔向き、表情変化などの、様々な変動要因に左右されない照合方式が必要とされており、顔照合研究の主要な課題となっている。

3.2.1.1. 照明変動

照明変動に対する対策には、以下の方法がある。

(1) 照明変動に強い画像特徴量の利用

照明に有効とされているエッジ特徴(2D-Gabor 特徴など)を利用する方法である。しかし文献[8]では様々な照明条件下でエッジ特徴量を詳細に比較した結果、照明変動に対する効果は限定的であると報告されている。

(2) 照明部分空間の利用

顔表面を完全散乱面と仮定すると、顔表面の照明変動は 3 次元線形部分空間で記述できる、という性質を利用し、登録人物の照明部分空間を生成して照合を行う[9][10][11]。この場合、通常 3 枚以上の異なる照明下で撮影された登録画像が必要である。また文献[12]では逆に照明成分に依存しない顔部分空間を利用する方法が提案されている。

(3) 顔の 3 次元形状の利用

文献[13][14]では、本人の 3 次元形状を用いて照明変動部分空間を生成し、照合画像と照合する方法が提案されている。本人の形状を利用することで、大きな照明変動下においても、高い照合率を達成している。また文献[15]では、汎用的な顔の 3 次元情報を利用して顔の表面反射率を推定することで、照明の影響を排除できることを示した。2 次元画像から顔の 3 次元形状を推定したこと照明成分を補正する方法[16]も提案されているが、画像からの形状推定は一般に誤差が大きく、照合精度への影響が懸念される。

3.2.1.2. 顔向き変動

顔向き変動に対する対策には以下の方法がある。

(1) 事前に人物毎に複数の顔向き画像を登録[17]

比較的高精度な照合が期待できるが、あらかじめ複数の顔向き画像を登録する必要がある。

(2) 顔の 3 次元形状を利用し登録画像と照合画像の顔向きを一致させた上で照合

文献[14]では、本人の 3 次元顔形状を用いることにより照明変動と同時に任意の姿勢における高精度な照合を実現した。また、汎用的な 3 次元顔モデルを用いた研究[18][19][20]も存在する。文献[19]では、数十点の顔の特徴点座標からなる 3 次元顔モデルを使い、特徴点の Gabor Wavelet グラフマッチングを用いて顔向き変化に対応している。

(3) 顔向き変動のモデル化

文献[21]では、SVM を用いた顔向き推定により顔向き変動を吸収した。また文献[22]では、顔向き変動を表現するパラメトリックな部分空間を学習し、顔照合に適用した。

3.2.1.3. 表情、部分的な遮蔽

表情、部分遮蔽に対する研究は、他の変動要因に対する研究に比べて少ない。文献[23]では目、鼻、口の部分テンプレートを利用して、髪形、眼鏡などの影響を排除している。文献[19]の方法も、弾性マッチングを行うため、微小な変動には対応できると思われる。また文献[24]では、文献[19]と同様な Gabor Wavelet グラフマッチングに、ロバスト推定を組み合わせた。遮蔽に対応したとする最近の文献[25]では、顔領域を小領域に分割した後に複数枚の登録画像を用いて主成分分析し、小領域単位で動的計画法と同様な探索を行いながら照合する方法を提案している。文献[25]の方法では、登録する人物毎に複数枚の画像が必要であり、登録画像に大きく依存する点と、原理的に、遮蔽物による影響を完全には排除できない点が問題である。

本節で取り扱う顔照合は、登録画像が 1 枚であることを想定している(静止画面顔照合)。従って複数枚の登録画像を必要とする従来手法は適用できない。すなわち、登録画像 1 枚のみから上記の様々な変動要因に対応するという困難な課題を解決しなければならない。

本節ではまず、表情変化、眼鏡装着などの影響を低減するための、適応的部分領域照合法について述べる。次に、顔向き変動、照明変動に対応するための照合法として、摂動空間法を提案

する。そして、適応的部分領域照合法と摂動空間法とを組み合わせることにより、複数の変動要因に対して高精度な顔照合が可能となることを示す。

3.2.2. 適応的部分領域照合法

実環境における顔画像は、顔画像の一部が撮影環境によって黒く潰れる、顔の一部を手で遮蔽する、または眼鏡を着用する、といった状況が存在する。このため、同じ人物であっても、部分的に登録された顔パターンと、全く異なるパターンとを比較することになる。このような、同一人物における部分的不整合の問題に関しては、上述のようにその原因が多岐にわたるため、モデル化は簡単ではない。複数の顔画像サンプルから、部分的不整合の状態を学習する方法も考えられるが、現実には十分なサンプルを収集することは困難である。従って、顔中心部の画像領域全体を1つの特徴ベクトルとして用いる方法（例えば[26][27][28]）では、このような部分的不整合に対応することは難しい。

このような部分的不整合の問題を解決するため、まず局所部分領域単位の照合について検討した。文献[29]において、目、鼻、口の部分テンプレートと、顔全体のテンプレートとの比較照合実験が報告され、目、鼻、口の部分テンプレートの有効性が報告されている。しかしながら、目、鼻、口といった特徴点が重要なのか、それとも部分テンプレートを利用すること自体が重要なのかは不明確であった。もしも前者であれば、複数の特徴点を正確に検出する必要があるため演算量の増大を招くだけでなく、口元などの不安定な特徴点の検出に対して注力しなければならない。一方、後者であれば、比較的演算量の低い(単純な方式の)検出器と共に利用することができ、広範囲の環境で使用することができる。

3.2.2.1. 部分領域照合

部分領域単位の照合効果を確認するため、画素値を特徴として、顔全体を1つの特徴として照合する手法と、顔全体をN個の局所領域に均等に分割し領域毎に照合する手法とを比較した。

部分領域照合法の概念図を図3-2に示す。本手法では、はじめに顔画像をN個の部分領域 R_i ($i=1\dots N$)に分割する。部分領域の画素値を平均と分散によって正規化し、ノルムを1としたものを特徴ベクトルとする。登録画像のi番目の部分領域の特徴ベクトルを \mathbf{x}_i とし、照合する画像のi番目の部分領域の特徴ベクトルを \mathbf{y}_i とすると、登録画像と照合画像の距離Dは次式で算出される。

$$D = \sum_{i=1}^N f(\|\mathbf{x}_i - \mathbf{y}_i\|), \quad (f(x): \text{単調増加関数}) \quad (3-1)$$

すなわち部分領域毎に独立に距離を算出し、これらの総和を全体の距離とする。本節では $f(x)$ として以下の非線形関数を用いる(θ は実験的に決定)。これは極端に大きい距離値の影響を排除し距離値を一定の上限値に抑えるものである。

$$f(x) = \begin{cases} \theta & \text{if } x > \theta \\ x & \text{if } x \leq \theta \end{cases} \quad (3-2)$$

部分領域照合法は、部分領域毎の輝度正規化によって、領域内の線形な輝度変化を吸収できるという利点がある。さらに関数 $f(x)$ を作用させることにより、極端に大きな距離値の悪影響を排除するため、眼鏡や表情変化による距離値の変動を低減することができる。

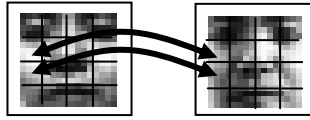


図 3-2: 部分領域照合の概念図

眼鏡、表情、及び照明変化を含む環境で撮影された正面顔画像を用いて顔照合実験を行った。顔画像は手動で与えた両目位置を基準として 30×30 画素の大きさに正規化した。また部分領域は、縦横均等大きさに分割した。登録数は 150 枚 (150 名)、照合数は 1200 枚 (150 名) である。

図 3-3 に実験結果である ROC 曲線を示す。N=1 は領域分割なしで照合した場合、N=9,49 はそれぞれ分割数を変化させた場合の結果である。なお N=9 の場合は縦横各 3 分割し、N=49 の場合は縦横各 7 分割した。比較のため、文字認識分野で利用される加重方向指数ヒストグラム法[30]による結果(Fex N=1)を併記する(空間 8×8 方向 8、全 512 次元)。

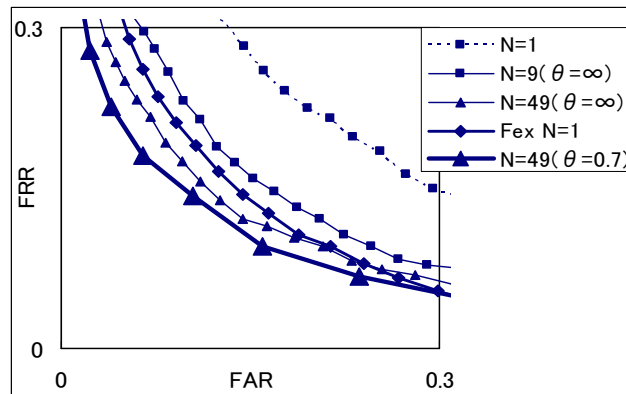


図 3-3: 部分領域照合における ROC 曲線

画像を部分領域に分割し、領域毎に独立に照合した結果を統合する部分領域照合法について考察した。比較照合実験の結果、顔画像を領域分割して照合することによって、照合性能が大きく向上した。すなわち、局所領域単位で照合するほうが、全体領域で照合するよりも明らかに精度が良いという結果が得られた。また各領域の距離値に非線形関数を作用させることにより、眼鏡、表情変化に対して頑健になった。

本実験より、目、口、鼻といった顔特有の部品を用いなくとも、部分領域を用いて照合を行うことで、照合性能を向上させられることが明らかとなった。

3.2.2.2. ロバスト推定を利用した適応的部分領域照合

部分領域照合において、同一人物における部分的不整合を検出し、スコアに反映させる方法について述べる。部分的不整合の検出は、各部分領域のスコアを直接利用して行う。すなわち照合スコアの低い領域を不整合領域とみなす。しかしながら、スコアが低い理由が、前述の不整合なのか、他人画像によるものかは照合時点では不明である。

そこで、スコア値に応じた対策としてロバスト推定の考え方を適用し、全 N 個の領域の中で相対的にスコアの大きい上位 $\alpha\%$ の領域スコアを出力する。本手法による顔パターン間のスコア D は式(3-3)で表される。 S_i は各部分領域における非線形照合関数である。本稿では S_i として式(3-4)の関数を用いた。 $\mathbf{x}(i)$ は登録画像の i 番目の部分領域の特徴ベクトルであり、 $\mathbf{y}(i)$ は照合画像の i 番目の部分領域の特徴ベクトルである。また A はスコア上位 $\alpha\%$ の領域 i の集合であり、

$d(x,y)$ は部分領域単位の照合スコア算出関数である。本手法をここでは、適応的部分領域照合法 (Adaptive Regional Matching) と呼ぶ。

$$D = \sum_{i=1}^N S_i \quad (3-3)$$

$$S_i = \begin{cases} 0 & \text{if } i \notin A \\ d(x(i), y(i)) & \text{if } i \in A \end{cases} \quad (3-4)$$

照合画像が同一人物であって不整合領域が存在する場合には、不整合領域はスコア下位の領域に含まれると考えられるので、全体スコア算出前にその影響が排除される確率が高くなる。また同一人物で不整合領域が存在しない場合は、全ての部分領域で平均的に高いスコアが得られると考えられるので、スコア下位を排除しても全体スコアに影響はない。他人であれば、不整合の有無に関わらず、全ての部分領域のスコア値が低くなる可能性が高いため、全体スコアは低くなる。

適応的部分領域照合法による照合の概念を図 3-4 に示す。図 3-4 において、黒い領域は照合スコアが低く、全体スコア算出には関与しないことを示している。

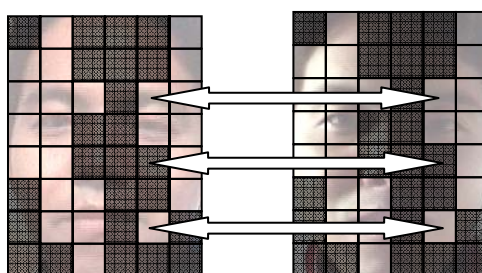


図 3-4: 適応的部分領域照合法の概念図

3.2.2.1 節と同様の条件にて、正面顔画像を用いた顔照合実験を行った。顔画像は手動で与えた両目位置を基準として 30×30 画素の大きさに正規化し、部分領域は縦横均等な大きさに分割した。登録数は 150 枚 (150 名)、照合数は 1200 枚 (150 名) である。

表 3-1 に、照合結果を示す。N=1 の時に一位照合率が 62.7%であったのが、分割数を増やして N=49 とすると 81.8%にまで改善した。また、特徴量として実績のある、加重方向指数ヒストグラム法を用いた照合結果は 79.0%であった。すなわち、特徴量を工夫しなくとも、適応的部分領域照合を適用するだけで、従来の性能を上回ることが可能であることを示した。

なお、分割数 N については、それぞれの実験条件において最適値が存在することが実験的に分かっている。今回の実験では N=49 より大きくしても、性能向上は見られなかった。

表 3-1: 適応的部分領域照合による照合結果

手法	EER	一位照合率
N=1 分割なし	0.214	62.7%
N=9 ($\alpha=50\%$)	0.147	74.1%
N=25 ($\alpha=50\%$)	0.124	80.8%
N=49 ($\alpha=100\%$)	0.131	75.2%
N=49 ($\alpha=50\%$)	0.124	81.8%
加重方向指数ヒストグラム法(従来法)	0.140	79.0%

適応的部分領域照合法の特長は次の通りである。

- ・多数の特徴点座標を必要としない。両目座標等、顔の大きさを正規化可能な最低限の特徴点だけで利用可能。
- ・顔全体を覆う均等分割された N 個の部分領域を用いることができるので実装が簡単。大きさと配置は必ずしも均等である必要はないが、均等分割のほうが処理実装を簡素化できる。
- ・部分的な不整合領域を自動的に排除するため、従来ほとんど顔照合に利用されていなかった、頭髮、顔輪郭、などの情報を含めることも可能。
- ・登録画像と照合画像の組み合わせに応じて自動的に局所領域が選択、重み付けされる。

3.2.3. 摂動空間法

1 枚の登録顔画像から、顔の向きや照明による見えの変化を予測できれば、それらの変動要因に対して頑健な照合システムを構築することが可能となる。そこで登録画像から、予測される見えの変動を施した画像群（変動画像群）を生成し、変動画像群から算出される部分空間（摂動部分空間）と、照合画像との距離値をパターン間距離として利用する手法を提案する。本研究ではこの手法を摂動空間法(Perturbed Subspace Method: PSM)と呼ぶ。

本節で述べる摂動空間法は、単一または複数の汎用的な 3 次元顔モデルを用いて変動画像を生成する。摂動空間法による変動生成は、次のステップで行うことを想定している。

(1)登録画像の姿勢を推定し正規化する。

(2)複数または単一の汎用 3 次元顔モデルと、顔向きが正規化された登録画像を用いて、変動画像を生成する。

顔画像の変動には、照明、姿勢、表情、その他様々な変動を加えることが可能である。実際に本節で生成する摂動部分空間は、姿勢変動と照明変動によって張られる部分空間である。また登録顔画像は正面顔を仮定しており、登録画像の顔姿勢推定は行っていない。

人物顔の照明変動は、少ない線形部分空間で近似できることが分かっており [9][10]、本手法のような部分空間による表現は有効である。また文献[11]より、3次元剛体の姿勢変動に関しては、部分空間上の多様体で表現することが可能であるとされている。従って、冗長性は否定できないものの、十分な次元数を確保すれば線形部分空間による顔の姿勢変動の記述が可能と考える。

図 3-5 に PSM の概念図を示す。摂動部分空間は、変動画像群を主成分分析して得られた固有値の大きな主成分によって張られる、線形部分空間として算出される。本節では、姿勢及び照明変動画像の生成方法を述べた後、PSM を用いたパターン照合法について述べる。

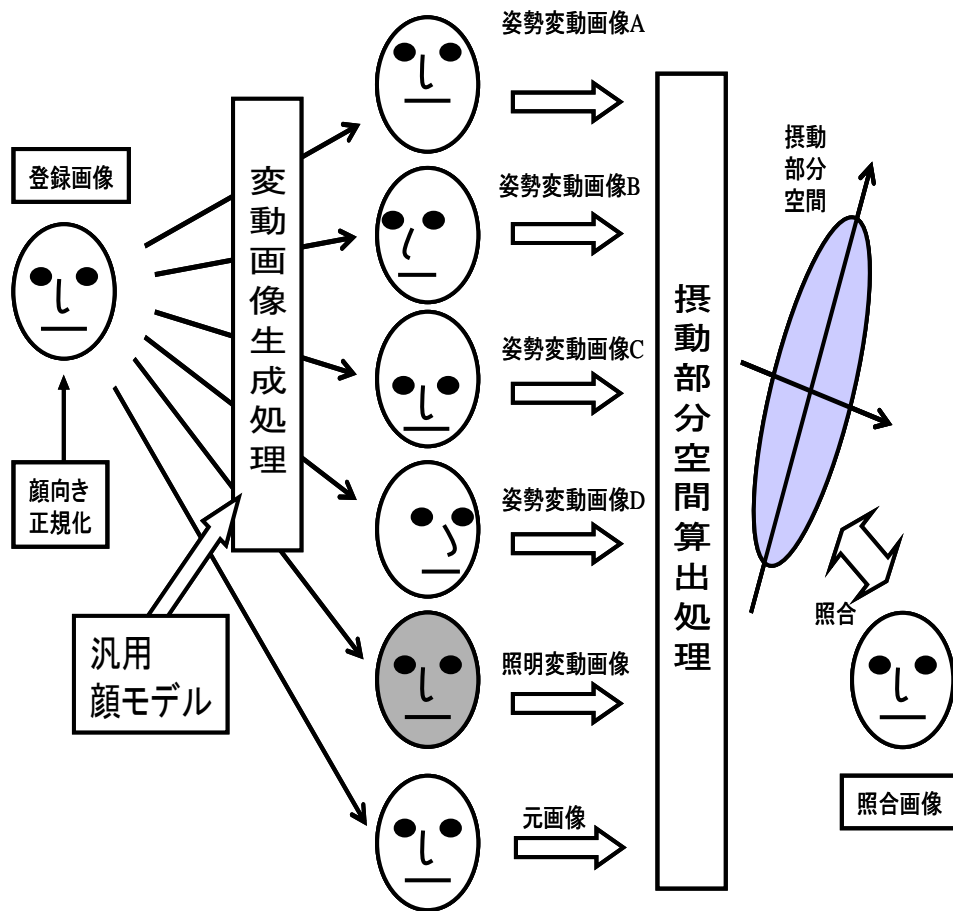


図 3-5: 摂動空間法の概念図

3.2.3.1. 平均顔モデルを用いた姿勢変動画像の生成

顔の姿勢変動に対応する摂動空間を求めるためには、登録画像から姿勢変動画像を生成しなければならない。本節の摂動空間法は、単一または複数の汎用的な 3 次元顔モデルを用いて姿勢変動画像を生成する。姿勢変動生成は次のステップで行う。

(1) 両目座標を用いて、顔画像の位置を補正。

(2) 顔画像の画素値を平均と分散を用いて正規化。

(2) 複数または単一の汎用 3 次元顔モデルと、顔向きが正規化された登録画像を用いて、姿勢変動画像を生成。

両目座標を用いた顔位置補正について述べる。顔画像は、図 3-6 に示すように両目位置を基準として幾何変換される。 L は両目間距離を示す。顔領域は、 L とパラメータ α, β, γ によって定義する。顔領域は $w \times h$ のサイズに回転・縮小される。

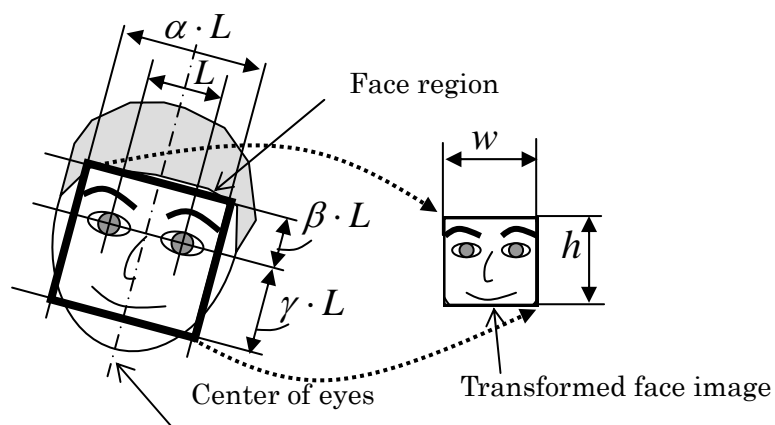


図 3-6: 顔領域の位置補正

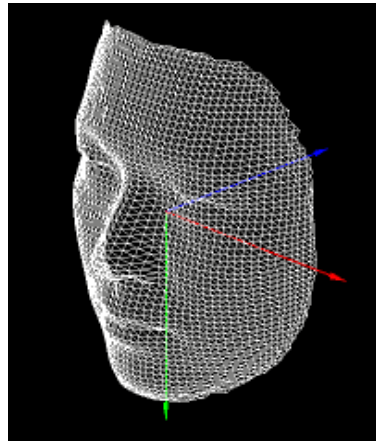
次に、顔画像の画素値を平均と分散を用いて正規化する。正規化処理は、式(3-5)と式(3-6)によって行われる。ここで、 $I(x, y)$ は、幾何変換を行った $w \times h$ サイズの画像であり、 $I'(x, y)$ は正規化後の画像である。 \bar{I} は、画素値の平均値、 σ は標準偏差を表す。式中の a, b は、画素値の範囲を調整するための定数である。正規化後の顔画像は、そのまま特徴ベクトルとして照合実験に用いる。

$$I'(x, y) = \frac{I(x, y) - \bar{I}}{3\sigma} \cdot a + b \quad (3-5)$$

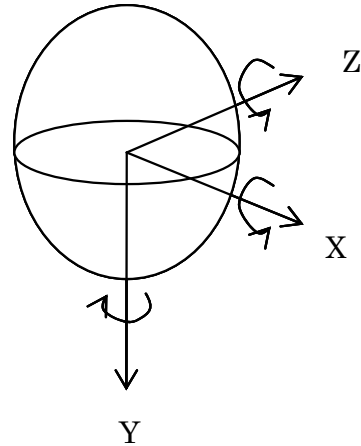
$$\bar{I} = \frac{1}{w \cdot h} \sum_{x=1}^w \sum_{y=1}^h I(x, y) \quad (3-6)$$

$$\sigma = \sqrt{\frac{1}{w \cdot h} \sum_{x=1}^w \sum_{y=1}^h (I(x, y) - m)^2}$$

本研究で用いる汎用 3 次元顔モデルは、実際の人間の顔の測定データより直接求められたものである。具体的には、レンジファインダーによって測定された数百名分の 3 次元顔データに対し適切な正規化を施した後、座標値を平均することで求められた、単一の平均 3 次元モデルである。その他、数学的に簡単に記述できる、楕円体モデル等を用いることも可能である。図 3-7(a) に実験で使用した平均 3 次元顔モデルを示す。



(a) 平均3次元顔モデル



(b) 3次元顔モデル座標系

図 3-7: 平均 3 次元顔モデル

変動画像生成手順は次の通りである。まず登録画像を汎用 3 次元モデルにマッピングし、3 次元座標値を得る。3 次元モデルの座標値を指定されたパラメータによって変換する(登録画像のテクスチャも同時に変換される)。変換後の 3 次元座標を画像平面に再投影することにより、姿勢変動画像が得られる。

顔座標値の変換方法として様々な方法が考えられるが、本研究では、人間の頭部を 3 次元の剛体とみなし次のように座標変換を行った。3 次元の剛体運動は 6 つの自由度(XYZ 軸方向の移動と XYZ 軸回りの回転)を持つ。各自由度に対応する変換は、3 次元座標値の線形変換で記述することができる。それぞれの変換行列は式(3-7)で表される。また元座標を(x,y,z)とすると、変換後の座標値(x', y', z')は、式(3-8)によって表すことができる。

$$\begin{aligned}
 R_x &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta & 0 \\ 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & R_y &= \begin{pmatrix} \cos \theta_y & 0 & \sin \theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & R_z &= \begin{pmatrix} \cos \theta_z & -\sin \theta_z & 0 & 0 \\ \sin \theta_z & \cos \theta_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
 T_x &= \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & T_y &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & T_z &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}
 \end{aligned} \tag{3-7}$$

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = T_z \cdot T_y \cdot T_x \cdot R_z \cdot R_y \cdot R_x \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3-8)$$

なお、本節における実験では、登録顔画像の姿勢はほぼ正面であり、手動によって両目位置を基準とした位置合わせがなされているものとする。姿勢変動画像の生成例を図 3-8 に示す。

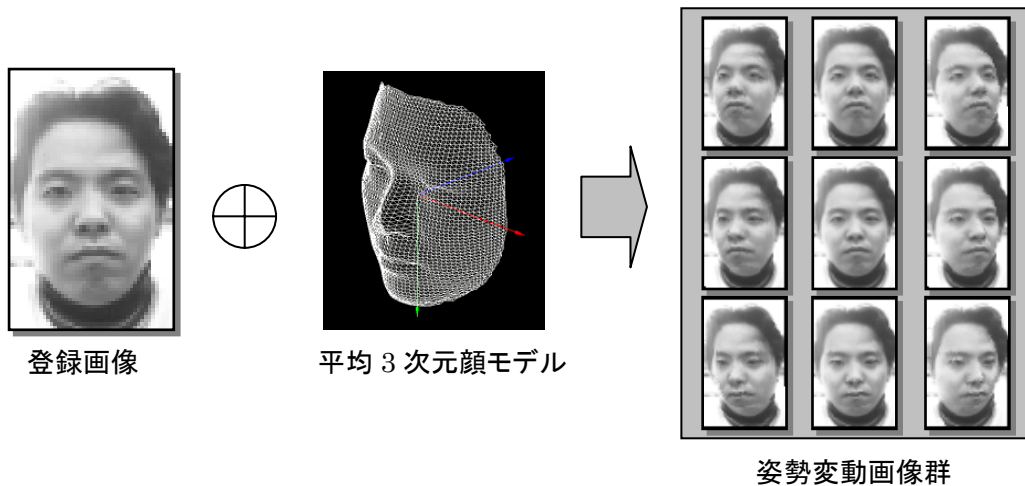


図 3-8: 姿勢変動画像の生成

3.2.3.2. 照明変動画像の生成

照明変動画像を作成する方法について述べる。人物毎の照明変動は、少ない線形部分空間で近似できることが分かっている[9][10]。しかしながら、登録画像のみからでは、登録人物の照明変動を事前に求めることはできない。そここ論文では、人物の照明変動空間を張る基底（照明変動基底）は人物によらず一定であるとみなし、次の方法によって近似を行う。

- ・ 多人数のレンジデータから得られた平均 3 次元顔モデル+平均顔テクスチャを用いこと光源位置を様々に変化させながらレンダリングした画像群を生成する。

- ・ レンダリング画像群を輝度正規化後、主成分分析して得られる基底を照明変動基底とする。以上のようにして得られた照明変動基底を図 3-9 に示す。

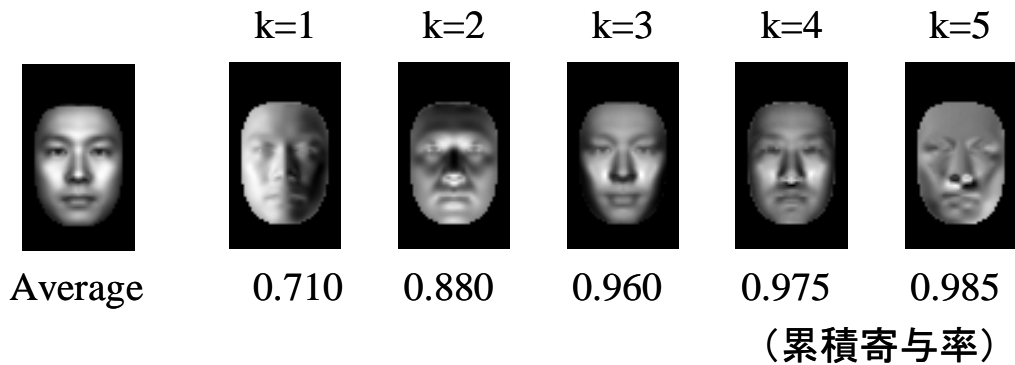
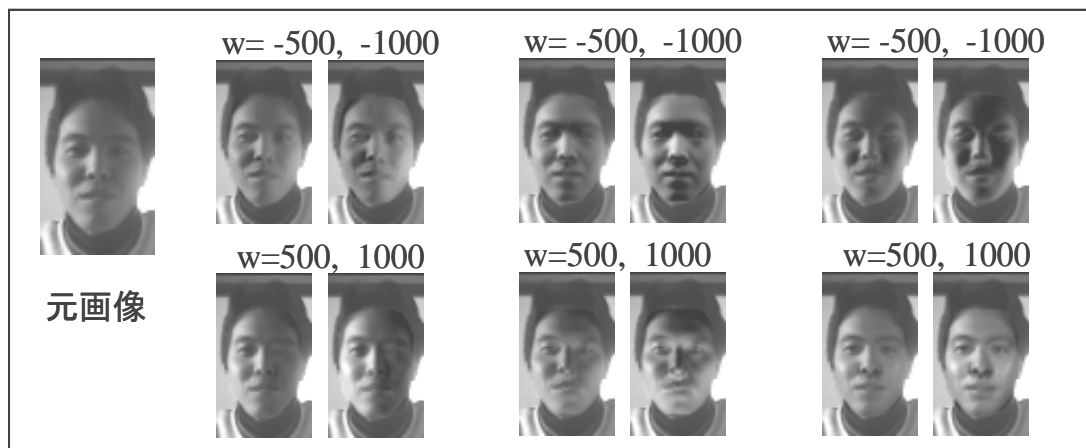


図 3-9: 照明変動基底

照明変動画像は、登録画像に対して照明変動基底成分を付加することによって求められる。
 K 個の照明変動基底ベクトル $\mathbf{v}_k : k = 1 \dots K$ とする。登録画像の特徴ベクトルを \mathbf{x} とすると、
 \mathbf{x} に照明変動基底成分を付加することによって、別の照明画像へ変換することができる。この
 とき、照明変換後の特徴ベクトル \mathbf{y} は式(3-9)で表すことができる(\mathbf{w} は重み係数)。以上のように
 して生成した照明変動画像の一例を図 3-10 に示す。

$$\mathbf{y}^t = \mathbf{x}^t + \sum_{k=1}^K w_k \cdot \mathbf{v}_k^t \quad (3-9)$$



第1 主成分(k=1)付加 第2 主成分(k=2)付加 第3 主成分(k=3)付加

図 3-10: 照明変動画像の例

3.2.4. 部分領域における摂動空間法

摂動空間法と部分領域照合とを統合した手法について述べる。図 3-11 に統合の概念図を示す。本統合手法は、変動画像群を生成後、部分領域毎に摂動部分空間を生成する。これにより、照合を部分領域単位で行うことが可能となる。

部分領域毎の照合処理は、特徴ベクトルと摂動部分空間との距離値によって行われる。部分空間との距離値の計算方法として、式(3-10)の擬似マハラノビス距離を用いる。これは部分空間投影距離の拡張形である。N 個の部分領域を($n=1\dots N$)で表す。X(n)は照合対象における部分領域 n の特徴ベクトル、M(n)は部分領域 n の平均特徴ベクトルを表す。特徴ベクトルは部分領域内の画素値に対し平均と分散によって輝度正規化し、ノルムを 1 としたものである。固有ベクトル $\Phi_i(n)$ と固有値 $\lambda_i(n)$ は、登録された画像群の部分領域 n における固有値の大きい m 個の固有ベクトルと固有値である ($i=1,\dots,m$)。また σ^2 はノイズを表す定数項であり、 $\sigma^2 = 0$ の時は、部分空間への投影距離と一致する。なお式(3-10)は距離値のため、「スコア上位」とは、距離値の小さいものを意味する。

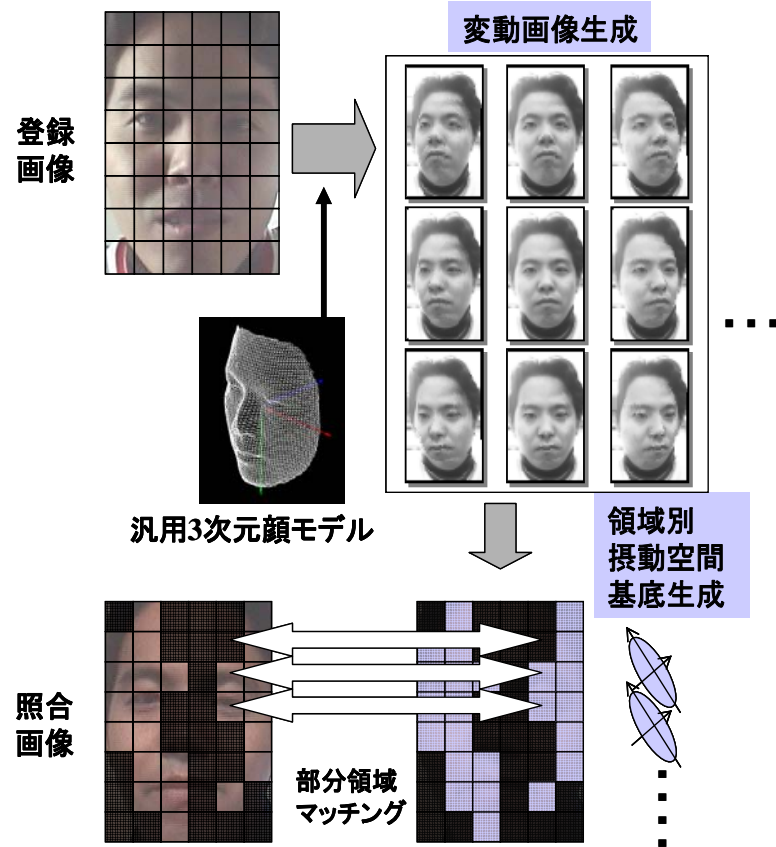


図 3-11: 部分領域毎の摂動空間法

$$d(n) = \|\mathbf{X}(n) - \mathbf{M}(n)\|^2 - \sum_{i=1}^k \frac{\lambda_i(n)}{\lambda_i(n) + \sigma^2} \{\Phi_i(n)' (\mathbf{X}(n) - \mathbf{M}(n))\}^2 \quad (3-10)$$

3.2.5. 顔照合実験

本手法の有効性を確認するため、眼鏡、表情、顔向き、及び照明変化を含む環境で撮影された正面顔画像を用いて顔照合実験を行った。顔画像は手動で与えた両目位置を基準として 30×30 画素の大きさに正規化した。また部分領域は、大きさを同じくして縦横均等に配置した。基本的に縦横 6 分割とし、さらに領域の半分が重なるように重複する部分領域を設定する。そのため分割総数 N は 121 (縦 11 \times 横 11) とした。各部分領域の大きさは 25 画素 (5 \times 5 画素) となった。また摂動空間生成のため、姿勢変動 16 種類、照明変動 6 種類の組み合わせによる変動画像 (合計 96 枚) を生成して照合に用いた。

実験に用いた顔画像データベースは、室内で撮影されたものであり、人数は 150 名で、それぞれ照明環境 3 種類、姿勢 5 種類、表情眼鏡 3 種類からなっている。この中から、(無表情、照明 1、正面顔) の 150 枚を登録画像とし、残りをテスト画像として用いた。

なお本実験では部分領域照合法及び摂動空間法の効果を確認するため、以下の条件について、顔照合実験を行った。

- 1) 未適用(部分領域照合法、摂動空間法共に未適用)
- 2) ARM(部分領域照合法のみ適用)
- 3) PSM(摂動空間法のみ適用)
- 4) ARM+PSM(部分領域照合法と摂動空間法を両方適用)

実験結果を、図 3-12 から図 3-15 にかけて示す。図中では、手法別、各種変動要因別に、一位照合率と EER 値の比較を示す。なお照合精度は、一位照合率が高いほうが良く、EER 値は低いほうが良い結果である。

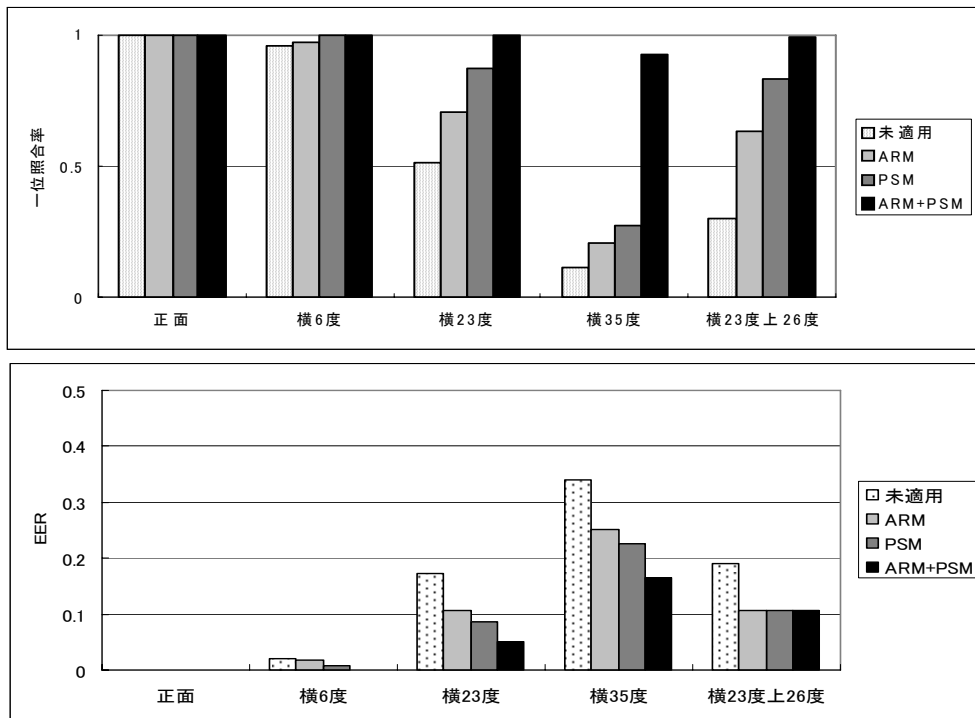


図 3-12: 姿勢別の一位照合率と EER (照明 1、無表情)

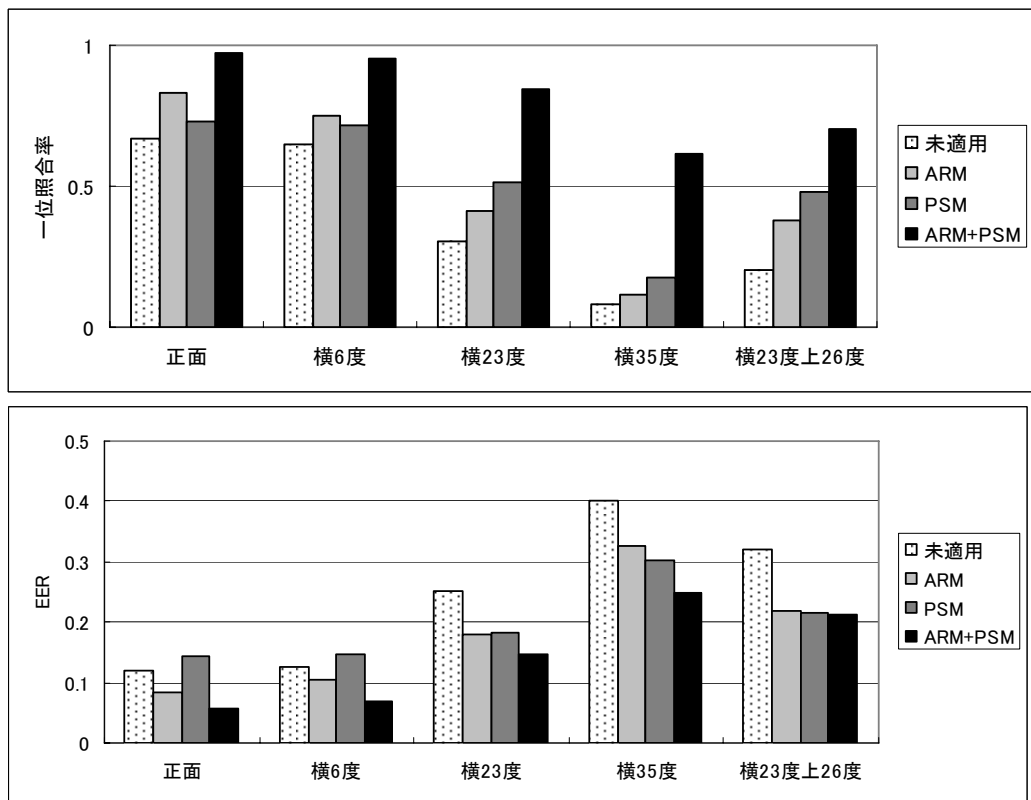


図 3-13: 姿勢別の平均一位照合率と EER

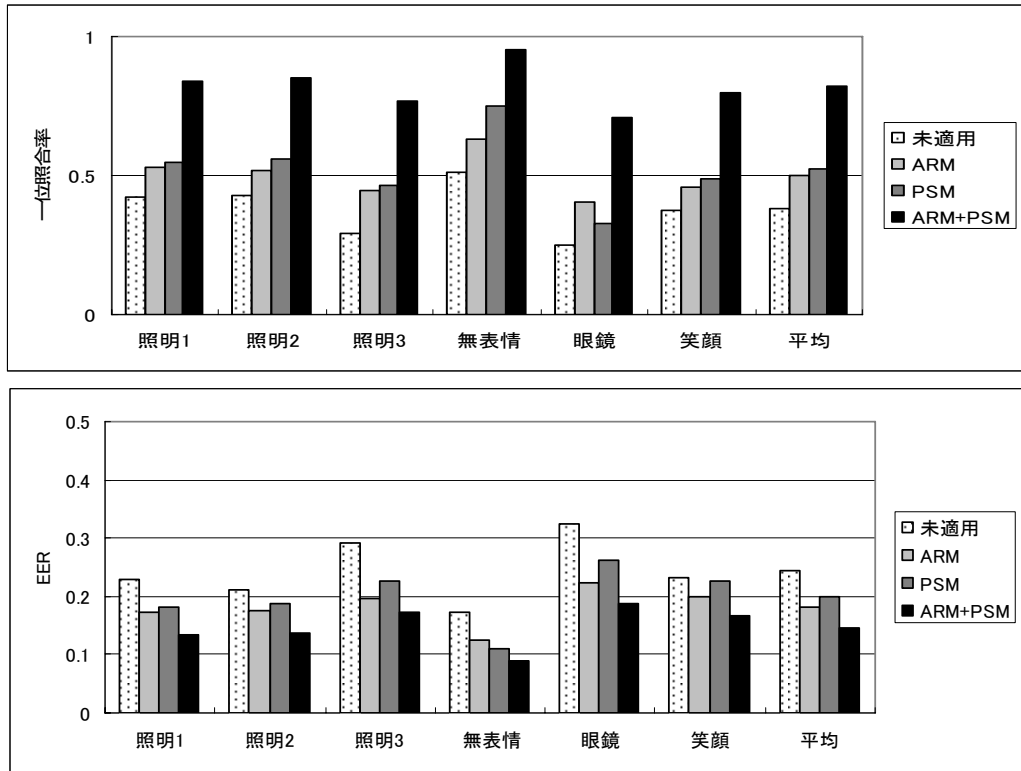


図 3-14: 表情、眼鏡、照明別の平均一位照合率と EER (全姿勢)

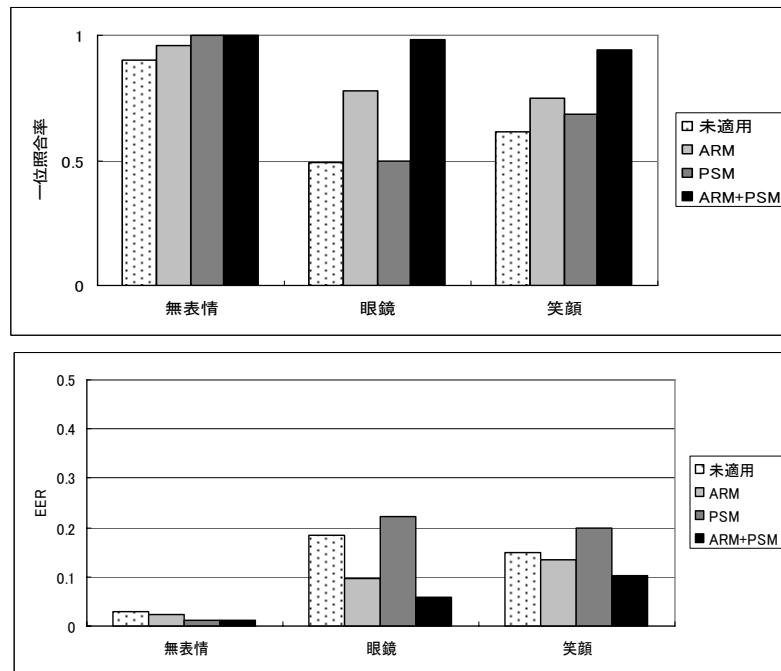


図 3-15: 無表情、眼鏡、笑顔の平均一位照合率と EER (正面姿勢のみ)

姿勢変動に対する性能を評価するため、図 3-12 に照明と表情が登録画像と同一の場合（照明 1、無表情）の、姿勢別一位照合率・EER を示す。図 3-12 より PSM を適用した場合には、適用しない場合に比べて大幅に性能が向上し、横 30 度程度でも高い照合率が得られることが分かった。これは摂動空間法の意図した通りの結果である。

図 3-13 に、姿勢だけでなく、表情や照明の影響を含んだ姿勢別一位照合率・EER を示す。この場合には正面顔のみの場合に比べ、全体的に照合率が低下する傾向が見られるが、それでも ARM+PSM は、未適用や ARM のみに比べて著しく照合率が改善している。

図 3-14 に照明環境、表情、眼鏡装着毎の平均一位照合率、及び全平均一位照合率・EER を示す。図 3-14 より、PSM を適用した場合は、照明別の結果において照合率が改善されており、PSM の照明摂動の効果を確認することができた。ARM のみ適用の場合も、多少改善があるが、これは局所領域における輝度正規化の作用と考えられる。

図 3-14 では、表情、眼鏡別の結果に関して ARM+PSM の照合率が大きく向上しているが、これは姿勢対応の寄与が考えられる。そこで、図 3-15 に姿勢が正面顔の時の、表情、眼鏡装着に対する一位照合率・EER を比較した。その結果、ARM において改善結果が現れており、特に眼鏡装着における上昇が大きいことが分かった。

本実験では全ての場合において、未適用に比較して ARM+PSM の照合率が改善しており、本提案手法は様々な要因が混在する場合でも、効果があることが分かった。

3.2.6. 静止画顔照合のまとめ

本節では、適応的部分領域照合法と、汎用 3 次元顔モデルを用いた摂動空間法を提案し、さらにこれらを統合した静止画顔照合方式について述べた。摂動空間法は、顔向きと照明変動に対応し、適応的部分領域照合法は部分的な遮蔽や表情変化に対応する。本方式により、複数の要因に対して対応可能な顔照合方式を実現した。顔向き、照明、表情、眼鏡装着による見えの変化を含むデータベースを用いて顔照合実験を実施した結果、従来手法に比べて大きく性能が向上することが明らかとなった。

なお、適応的部分領域照合における、最適な分割数と、部分領域の大きさについては、実験的に求めた。摂動空間法における生成画像数、摂動角度、変動量等についても同様である。これらのパラメータは、実験環境や認識対象の分布と関連があると考えており、その理論的解明は今後の課題である。

また、本手法における領域別照合には、基本的に部分空間法の投影距離（式(3-10)の擬似マハラノビス距離）を用いているが、原理的には他の照合方法(Harr 特徴による照合等)を適用することも可能である。

3.3. 部分空間への相互投影による動画顔照合

3.3.1. 動画顔照合の関連研究

顔画像を用いた個人識別技術は、非接触であるため他のバイオメトリクス技術に比べてこと者への負担が少ないという利点がある。また監視システム、ロボット、入門管理システムなど、幅広い応用が考えられるため、その研究は近年活発化している。

3.2 節に述べたように、登録照合共に 1 枚の顔画像を用いる照合技術（静止画顔照合）については、数多くの研究報告がなされている[6][7]。しかしながら、認証端末などの実際の応用場面では、ビデオカメラからの動画像（連続した複数フレーム画像）がシステムに入力されることが多い。このように入力が動画像である顔照合（動画顔照合）は、照合(Query)側データとして静止画ではなく、連続した入力画像シーケンスを用いて照合を行うことが可能である。このような動画顔照合では従来、1 フレーム毎に静止画照合を行い、数フレームの結果を平均して

結果を判定するというアプローチを取ることが多かった。

本研究で対象とする動画像照合は、認証端末やホームロボット等での利用を想定している。従って、顔向きは基本的に正面であるが、上下左右 20 度程度の動きがある。また照合時の照明環境は一定ではないが、登録時に複数の照明環境で登録できることとする。次に画像入力は汎用のビデオカメラを想定し、顔照合システムのユーザに対する反応速度を考慮すると、入力画像シーケンスは最大でも 100 枚程度であり、応答を考慮すれば実際には 30 枚程度の入力であっても動作することが必要となる。同様に応答速度を重視すれば、照合処理自体にも十分な高速性が要求される。

このような目的に適用可能な従来の動画像顔照合法として次のような研究がある。山口ら[30]は、入力画像シーケンスの部分空間と、登録画像群が張る部分空間との角度を利用した相互部分空間法[32]を動画像顔照合に適用し、静止画照合に比べて高精度であることを示した。また坂野ら[33]は、相互部分空間法にカーネル関数を適用した非線形化を行い、その有効性を示した。また文献[34]では、部分空間同士の最小距離値を用いた動画像照合が提案された。しかしながら動画像顔照合に関し、これら以外の研究はほとんどなく、入力画像シーケンスの時系列情報を有効に利用した例は少ないと言える。

相互部分空間法は、部分空間同士のなす角度 θ の余弦の二乗を、部分空間同士の類似度として利用する。入力画像シーケンスによって張られる部分空間の固有ベクトルを列ベクトルとする行列を U 、登録画像によって張られる部分空間の固有ベクトルを列ベクトルとする行列を V とすると、類似度 S は式(3-11)の行列 X の最大固有値として算出することができる。相互部分空間法は、少ない数の固有ベクトルを用いて算出可能なため、少ないサンプル数でも共分散行列の推定誤差の影響といった問題がない。しかしながら、行列 X の最大固有値を照合時に毎回算出する必要があり、演算量が大きいという問題があった。また文献[34]の方法でも、同じく固有値解析を行う必要があり、相互部分空間法と同等の演算量を必要とする。

$$\mathbf{X} = \mathbf{U}^T \mathbf{V} \quad (3-11)$$

本研究では、動画像照合は、入力画像シーケンスによって形成される入力特徴分布と、登録された画像群の形成する辞書特徴分布との距離値を評価することであると考える。このような分布間距離としては、カルバック情報量、ダイバージェンス、バタチャリア距離などがある[35][36]。双方の分布が正規分布であると仮定すると、例えばダイバージェンスは、式(3-12)によって表すことができる。

$$\begin{aligned} \text{div} = & \frac{(M_1 - M_2)^t (\Sigma_1^{-1} + \Sigma_2^{-1}) (M_1 - M_2)}{2} \\ & + \text{tr} [\Sigma_1^{-1} \Sigma_2 + \Sigma_2^{-1} \Sigma_1 + 2 \cdot I] \end{aligned} \quad (3-12)$$

しかしながら、式(3-12)によれば、照合時に毎回入力と登録の共分散行列をマージする処理が必要であり、演算量の増大を招く恐れがある。さらに、本研究の適用シーンのように、サンプルデータが少ない場合、計算に必要となる分散共分散行列の正確な推定は一般に困難であり、有効な距離値を計算することができないという問題があった。

そこで本研究では、動画像顔照合に用いるための新たな分布間距離尺度として、相互投影距離を定義し、本尺度を用いた照合方式「相互投影距離法」を提案する。以下本節では、まず相互投影距離法の考え方と、その算出方法について述べる。次に動画像顔照合への適用について述べる。最後に、従来手法との比較を含めて、本手法を実際の動画像に適用した照合実験結果について述べる。

3.3.2. 相対する分布との距離

入力画像シーケンスによって形成される入力特徴分布 C_1 と、登録された画像群によって形成される辞書特徴分布 C_2 との分布間距離について考える。

特徴ベクトル x と、クラス分布 C_1, C_2 との距離値が、あらかじめ関数 $d_1(x), d_2(x)$ として定義されているものとする。双方の分布の中心ベクトル（平均ベクトル）を m_1, m_2 とすると、 m_2 からクラス C_1 への距離は $d_1(m_2)$ となり、 m_1 からクラス C_2 への距離は $d_2(m_1)$ となる。図 3-16 に $d_1(x)$ と $d_2(x)$ の分布の一例を示す。求める分布間距離は、 m_1 と m_2 とを結ぶ直線上の関数 $d_1(x) + d_2(x)$ 上に存在すると考える。

$d_1(x), d_2(x)$ は $x = m_1, m_2$ の時、最小値 0 を取り、そこから離れるにつれて単調増加する関数とする。この時 m_1, m_2 を結ぶ直線上において $d_1(a) = d_2(a)$ を満たす点 a が一つ存在する。点 a は 2 つの分布から等距離にある点と考えられる。すなわちこの 2 つの分布間の距離は、 m_1 または m_2 から点 a までの距離値の 2 倍であると考えられる。言い換えると、分布間距離値は、点 a における $d_1(x)$ と $d_2(x)$ の和として求められる。本論文ではこれを相互投影距離 D として定義する(式(3-13))。

$$D = d_1(a) + d_2(a) \tag{3-13}$$

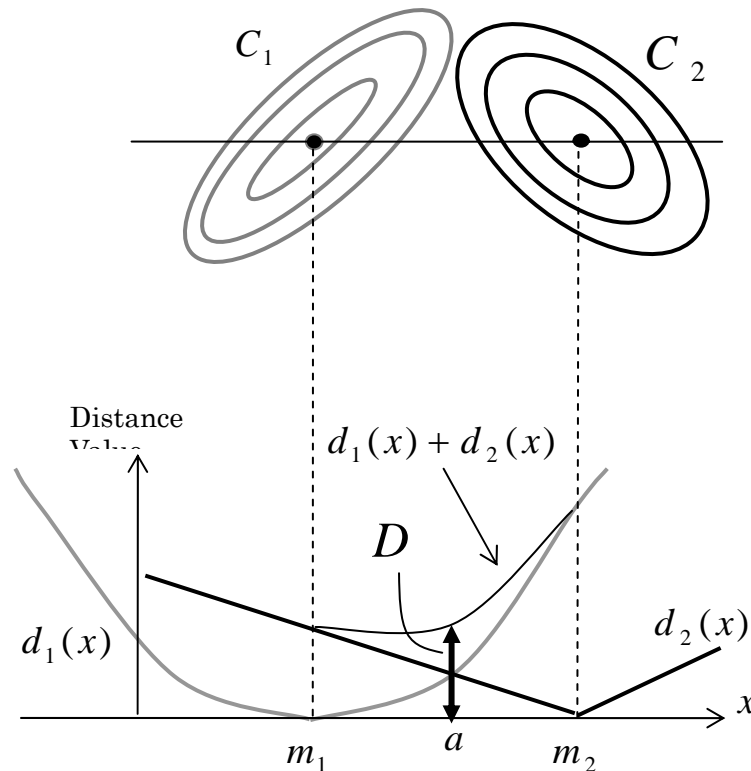


図 3-16: 相対する 2 つの分布間の距離

3.3.3. 擬似マハラノビス距離

次にマハラノビス距離を基にして距離関数 $d_1(x), d_2(x)$ を定義する。マハラノビス距離は、正規分布を仮定した分布と特徴ベクトルとの間の標準化距離であり、広範囲なパターン認識分野に利用されている。次元数 n の特徴ベクトルを x 、分布の平均ベクトルを m 、分布の分散共分散行列を Σ とすると、マハラノビス距離は式(3-14)によって定義される。

$$d_m^2 = (x - m)^T \Sigma^{-1} (x - m) \quad (3-14)$$

しかしながら、認識対象の画像サンプル数が少ない場合には、サンプルから計算した分散共分散行列が特異になり、算出される距離値が不安定になる問題がある。この不安定要因を回避するため、いくつかの距離算出法が提案され、様々な認識応用分野で適用されている[36][38]。ここでは、次のように分散共分散行列を推定する[36]。

$$\Sigma = \hat{\Sigma} + \sigma^2 I \quad (3-15)$$

$\hat{\Sigma}$ は標本サンプルから算出される分散共分散行列で、第二項は分散共分散行列の初期推定値を表す(σ^2 は定数)。分布の標本サンプルから求められる固有値を λ_i 、固有ベクトルを Φ_i とするとマハラノビス距離は次式で得られる。

$$d_M^2 = \sum_{i=1}^n \frac{1}{\lambda_i + \sigma^2} \{\Phi_i^T (x - m)\}^2 \quad (3-16)$$

固有値の小さい成分はノイズが支配的となる。そこで次元 $i > k$ の時 $\lambda_i \ll \sigma^2$ と仮定して変形し式(3-17)を得る。本研究では式(3-17)の平方根を擬似マハラノビス距離と呼び、距離関数 $d_1(x), d_2(x)$ として用いる。なお $d_p^2(x)$ は、 $\|x - m\|^2$ に比例するため、擬似マハラノビス距離 $d_p(x)$ は、 $\|x - m\|$ に比例する。

また式(3-17)より、 $\sigma^2 = 0$ の時は部分空間法の射影距離法と一致することは明らかである。

$$d_p^2(x) = \|x - m\|^2 - \sum_{i=1}^k \frac{\lambda_i}{\lambda_i + \sigma^2} \{\Phi_i^T (x - m)\}^2 \quad (3-17)$$

3.3.4. 相互投影距離法

図 3-17 に、擬似マハラノビス距離を用いた時の距離分布を示す。 $d_1(x), d_2(x)$ は、 m_1 と m_2 とを結ぶ直線上において次の式で表される。(但し $L = |m_2 - m_1|$)

$$d_1(x) = \frac{d_1(m_2)}{L}(x - m_1), \quad (3-18)$$

$$d_2(x) = \frac{d_2(m_1)}{L}(m_2 - x)$$

$d_1(x) = d_2(x)$ を解くことで、等距離点 a が得られる。

$$a = \frac{d_1(m_2)m_1 + d_2(m_1)m_2}{d_1(m_2) + d_2(m_1)} \quad (3-19)$$

式(3-19)を式(3-13)に代入することにより、距離 D は次のように得られる。

$$D = 2 \cdot \frac{d_1(m_2) \cdot d_2(m_1)}{d_1(m_2) + d_2(m_1)} \quad (3-20)$$

式(3-20)を相互投影距離として定義する。すなわち本距離値は、入力と登録の各分布平均を結ぶベクトルを、それぞれの相対する分布に投影して 2 つの擬似マハラノビス距離を算出し、それらの調和平均として定義される。

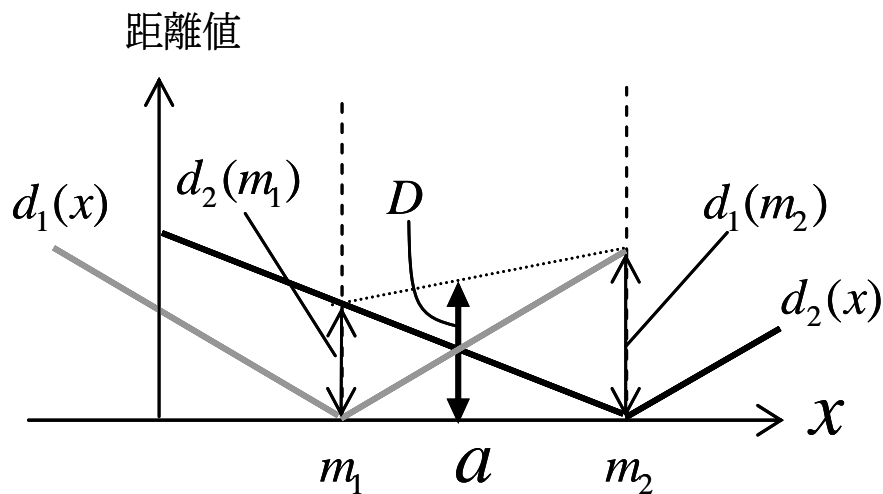


図 3-17: 相互投影距離

提案する相互投影距離法は、相互投影距離 D をパターン間の距離尺度として利用した照合方式である。

ところで式(3-12)に示す正規分布を仮定したダイバージェンスにおいて、第二項を無視するとマハラノビス距離の相加平均となることが分かる。そこで本手法と同様、擬似マハラノビス距離を用いた、擬似ダイバージェンス(式(3-21))を定義し、照合性能の比較を行う。

$$D_D = \frac{1}{2}(d_1(m_2) + d_2(m_1)) \quad (3-21)$$

3.3.5. 動画像照合への適用

相互投影距離法を動画像顔照合に適用するため、次のような手順を用いる。なお動画像顔照合は、登録側とテスト側(Query 側)共に、連続フレーム画像からなる画像シーケンスを用いるものとする。

- 1) あらかじめ、人物毎に登録画像シーケンスを用いて平均値、固有値、固有ベクトルを求めておく(辞書統計量)。
- 2) テスト用の画像シーケンスから、N 枚の連続画像(フレーム f=1 から N まで)を取得。
- 3) 入力 N 枚のテスト画像データから、平均値、固有値、固有ベクトルを求める(入力統計量)。
- 4) 人物毎に、辞書統計量と入力統計量を用いて、相互射影距離 D を計算し、距離値に基づいて照合。
- 5) 次の N 枚(フレーム f=2 から N+1 まで)のシーケンスを用い、3) 4) の処理を繰り返す。以降、フレーム番号を増加し、テスト画像シーケンスが終了するまで繰り返す。

3.3.6. 照合実験

3.3.6.1. 複数の照明環境下での照合(実験 I)

家庭用の入門管理システム、ロボットインターフェース等の応用を考慮すると、登録時に複数の照明環境で登録しておき、照合時の照明が登録と異なる場合のスコア変動を吸収したいという要求がある。また、ロボットインターフェースでは、入力シーケンスの途ことカメラが移ることによって照明方向が変化することが考えられる。よって、このような照明変動のある状況下における本手法の有効性について検証する実験を行った。

実験に用いた画像は、DV カメラを用いて一般家庭内の異なる 12 方向の光源下で撮影された動画シーケンスから、市販のキャプチャカードを使ってフレーム毎に取得した正面顔画像群である。そして撮影画像シーケンスには微小な顔の動きも含まれている。顔画像は輝度値の正規化後、手動で与えた両目位置を基準として 12×18 画素の大きさに正規化する。登録辞書は 12 方向の光源シーンを含んでいる。図 3-18 に正規化後の画像例を示す。また表 3-2 に実験 I に用いたデータセットを示す。



図 3-18: 実験 I における顔画像例

表 3-2: 実験 I に用いたデータセット

撮影環境	登録枚数	テスト枚数
室内 光源 12 方向	200 枚/人 8 人分登録	4420 枚 16 人

顔画像の正規化処理は、次の通りに行う。

- (1) 顔画像はあらかじめグレースケールに変換しておく。顔画像の両目位置(瞳中心)を手動

によって与える。

(2) 顔画像は、図 3-19 に示すように両目位置を基準として幾何変換される。 L は両目間距離を示す。顔領域は、 L とパラメータ α, β, γ によって定義する。顔領域は $w \times h$ のサイズに回転・縮小される。

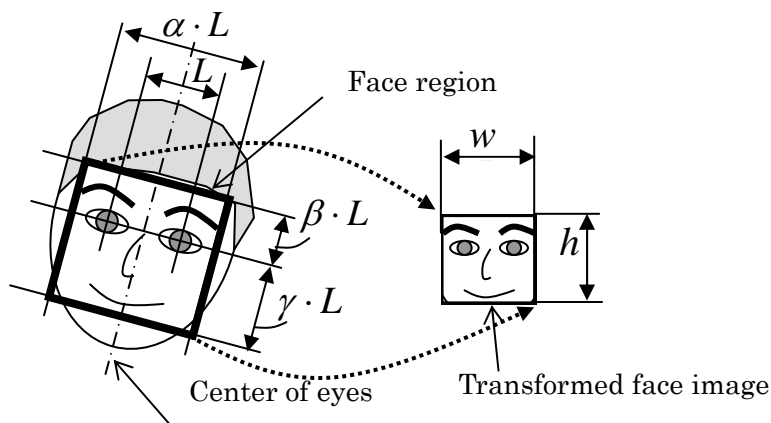


図 3-19: 顔領域の正規化

(3) 顔画像の画素値を平均と分散を用いて正規化する。正規化処理は、式(3-22)と式(3-33)によって行われる。ここで、 $I(x, y)$ は、幾何変換を行った $w \times h$ サイズの画像であり、 $I'(x, y)$ は正規化後の画像である。 \bar{I} は、画素値の平均値、 σ は標準偏差を表す。式中の a, b は、画素値の範囲を調整するための定数である。正規化後の顔画像は、そのまま特徴ベクトルとして照合実験に用いる。

$$I'(x, y) = \frac{I(x, y) - \bar{I}}{3\sigma} \cdot a + b \quad (3-22)$$

$$\bar{I} = \frac{1}{w \cdot h} \sum_{x=1}^w \sum_{y=1}^h I(x, y) \quad (3-23)$$

$$\sigma = \sqrt{\frac{1}{w \cdot h} \sum_{x=1}^w \sum_{y=1}^h (I(x, y) - m)^2}$$

本実験では、入力される画像が登録された本人かどうかを検証する本人同定性能の評価を行った。すなわち、本人棄却率 **FRR** と、他人受け入れ率 **FAR** を算出し、閾値を変化させながらグラフにプロットして評価した。図 3-20 に、実験 I の顔照合実験結果を本人同定性能である **ROC** 曲線によって示す。投影次元数 k は入力 20、登録 20 とし、入力フレーム数は $N=30$ 、そして式(3-17)の σ^2 は 0.001 とした。また、本手法の有効性を検証するため、同じデータセットを用いて部分空間法(投影距離法)、相互部分空間法、擬似ダイバージェンスによる顔照合性能を比較した。部分空間法を用いる場合は、入力 N フレームのスコア平均を用いて照合を行った。なお、図中の表記として、部分空間法を **SM** (Subspace Method)、相互部分空間法を **MSM** (Mutual Subspace Method)、擬似ダイバージェンスを **PD** (Pseudo Divergence)、提案手法である相互投影距離法を **MPM** (Mutual Projection Method) を使用している。

実験 I の顔照合実験結果より、本手法による顔照合は、部分空間法に比べ、大幅に性能が改善されることが分かった。また、擬似ダイバージェンスを用いた場合よりも良く、相互部分空間法による照合とほぼ同等な性能を持つことが分かった。

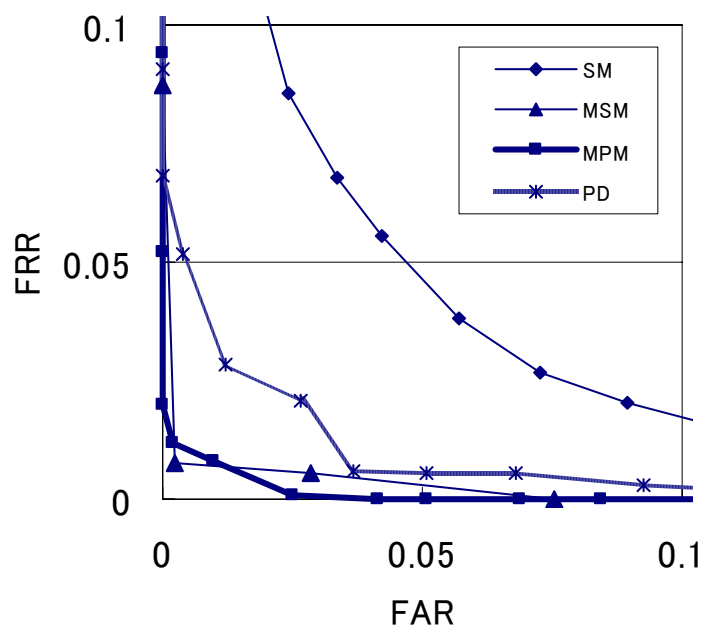


図 3-20: 実験 I の照合実験結果(ROC 曲線)

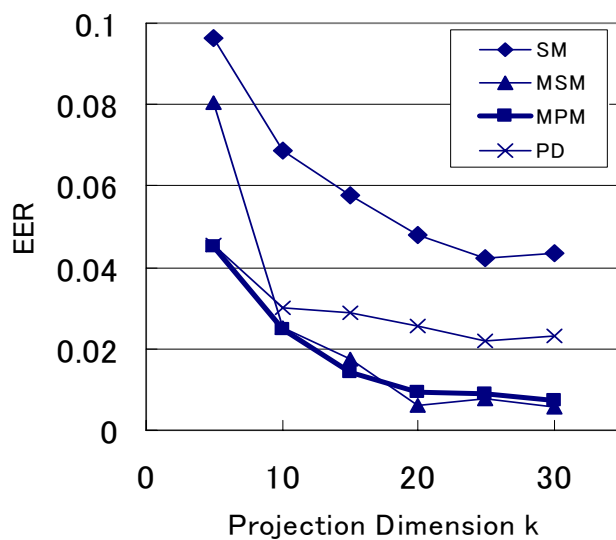


図 3-21: 投影次元数による EER 値推移

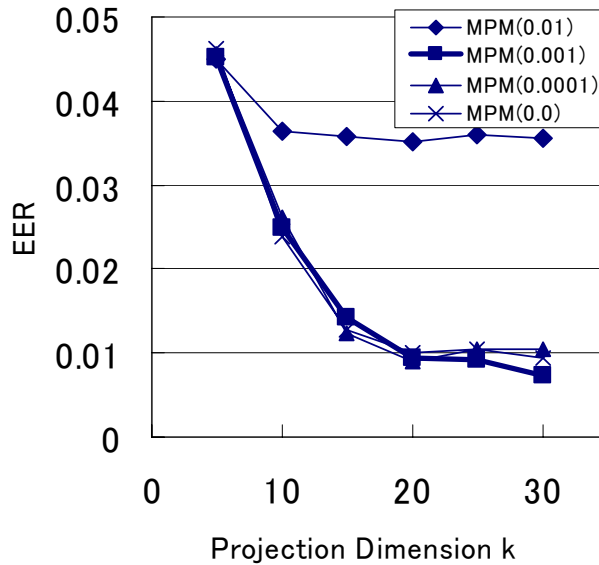


図 3-22: σ^2 及び投影次元数と EER との関係

図 3-21 に、射影次元数 k を変化させた時の EER(FAR と FRR が等しくなる時の値)の推移を示す。なお、EER 値が小さいほど本人同定性能が良いと考える。図 3-21 より、相互投影距離法(MPM) ($\sigma^2=0.001$)による照合は、部分空間法、擬似ダイバージェンスを用いた照合に比べ、全ての次元数 k において性能が良いことが分かった。そして、次元数 k が小さい時は、相互部分空間法(MSM)を用いた場合よりも性能が良く、 k が大きくなるに従ってその差がほとんど無くなることが分かった。但し $k=20$ 以上では相互部分空間法(MSM)による照合のほうが、若干 EER が低くなっている。図 3-22 に σ^2 値と次元数 k とを変化させた時の EER 値の推移を示す。その結果、 σ^2 値を 0.001 から 0.01 に変更すると、照合性能が低下することが分かった。

3.3.6.2. 微小な顔向き変動時の照合(実験 II)

次に照明環境は一定で且つ、 ± 20 度程度の顔向き変動と表情変化のあるデータセットによって、顔照合実験を行った。図 3-23 に本実験で用いる動画像撮影の様子を示す。登録画像シーケンスとテスト画像シーケンスは数日の間隔をおいて撮影された。被験者には、顔を上下左右に動かし、表情の変化をつけてもらうよう指示した。実験 II で用いた顔画像例を図 3-24 に示す。

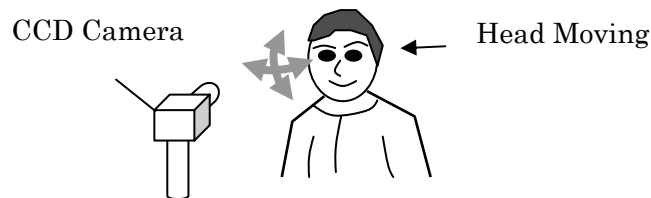


図 3-23: 実験 II における撮影状況



図 3-24: 実験 II で用いた顔画像例

表 3-3 に、実験 II に用いたデータセットを示す。なお、本実験における、顔画像の両目位置による正規化方法は、実験 I と同じである。

表 3-3: 実験 II に用いたデータセット

撮影環境	登録枚数	テスト枚数
実験室環境 固定照明 数日間隔で撮影	800 枚/人 20 人分登録	800 枚/人 24 人

図 3-25 に、実験 II の顔照合実験結果を本人同定性能である ROC 曲線によって示す。実験 I と同様、投影次元数 k は入力側を 20、登録側を 20 とし、入力フレーム数は $N=30$ 、 $\sigma^2=0.001$ とした。図 3-25 より、部分空間法(SM)に比べて、相互投影距離法(MPM)の照合性能は明らかに優位である。部分空間法(SM)を除く 3 手法の差は少ないが、それらの中では相互投影距離法(MPM)による結果が最も良かった。

図 3-26 に、投影次元数 k を変化させた時の EER 値の推移を示す。EER 値比較より、次元数 k を変化させても、部分空間法(SM)に比べて相互投影距離法(MPM)の優位は変わらない。また、MPM,MSM,PD の順序にも影響はなかった。

図 3-27 に、MPM における σ^2 の影響を調べるため、 σ^2 と次元数 k とを変化させた場合の EER 値推移を示す。次元数 $k=20$ あたりまではほとんど影響はないが、 $k=25$ 以上では、 σ^2 が小さいほうが EER 値が改善される結果となった。

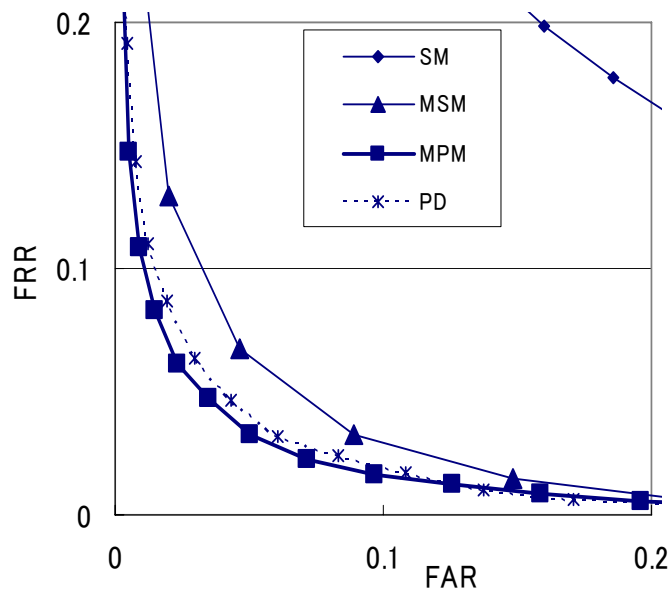


図 3-25: 実験 II の照合実験結果(ROC 曲線)

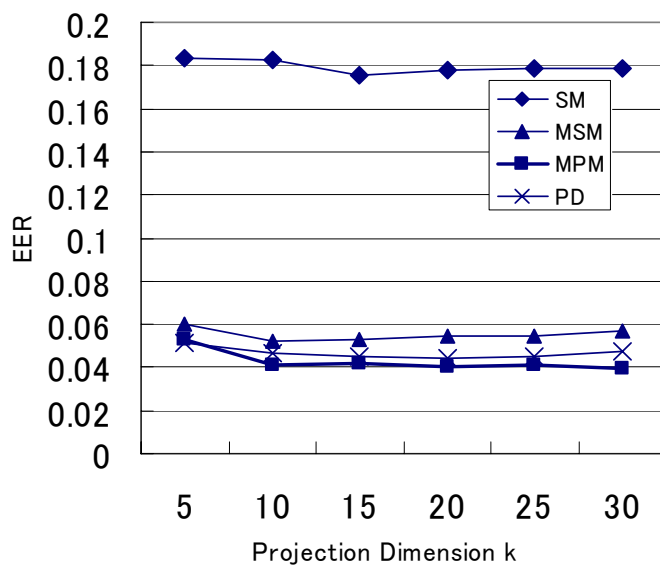


図 3-26: 実験 II の EER 推移

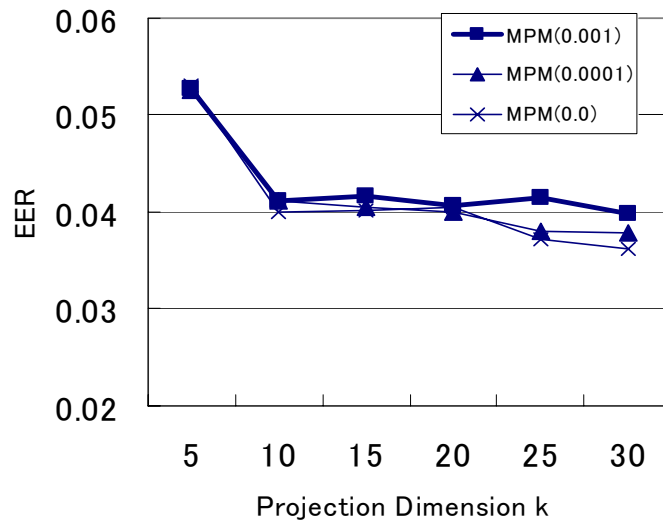


図 3-27: MPM における σ^2 と EER との関係

図 3-28 に参考として、複数人物の識別性能である一位照合率の結果を示す。照合率比較においても、部分空間法の劣位は明らかである。また相互部分空間法(MSM)を用いた場合に、次元数 k が低下するにつれて照合率が低下するという結果が得られた。しかしながら、本実験の人物数が少ないこともあり、部分空間法(SM)を除く 3 手法間の照合率の差は小さいと思われる。

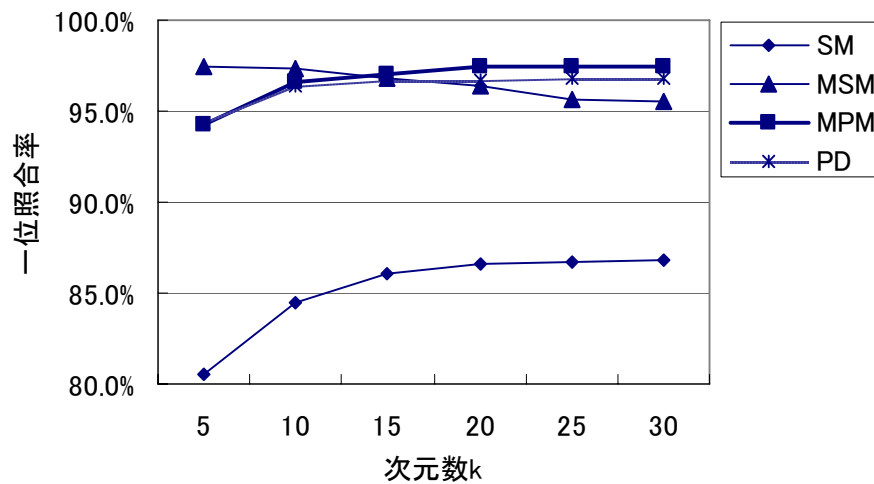


図 3-28: 実験 II の一位照合率

3.3.7. 演算量の比較

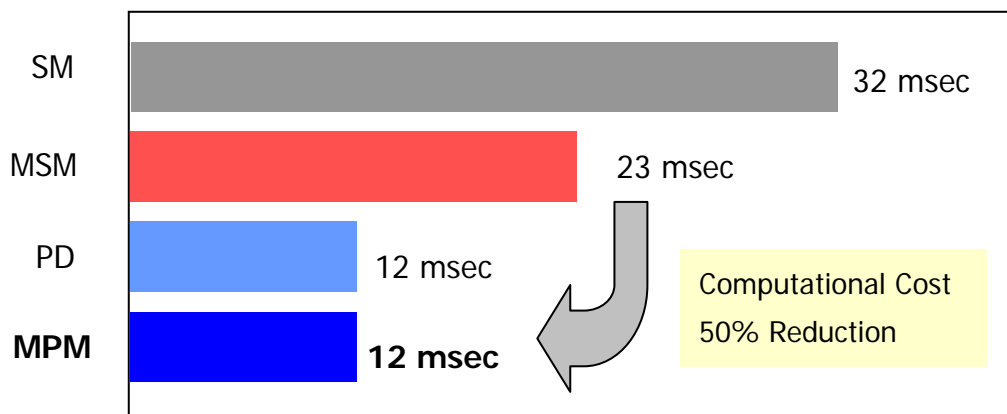
図 3-29 に、実験 I における 1 クエリあたりの平均処理時間を示す。用いた次元数 k は、一連の照合結果から処理速度と性能の最もバランスのとれた値として $k=20$ を採用した。この条件下において、MPM は MSM の処理速度の約 50% で実行可能であることが分かった。図 3-30 に、

処理速度の内訳を示す。これより次元数 k が小さいほど、速度改善効果が大きいことが分かる。
本結果より、MPM は MSM と比較して演算量の面で、大きな利点があることが明らかである。
その理由は、MSM は行列 $\mathbf{X} = \mathbf{U}^t \mathbf{V}$ の最大固有値を算出する必要があり、MPM ではその必要がないからである。(PD についても同じ理由から、演算量は少ない。)

3.3.8. 動画顔照合のまとめ

本節では、入力として複数の顔画像シーケンスを用いる顔照合技術として、相互投影距離法を提案した。相互投影距離法は、入力と登録の各分布平均を結ぶベクトルを、それぞれの相対する分布に投影して 2 つの擬似マハラノビス距離を算出し、それらの調和平均により定義される距離尺度を用いたパターン照合法である。相互投影距離法は、分布の平均と固有値の大きな k 個の固有ベクトルのみを用いて算出するため演算量が小さいのが特徴である。

実際に撮影された画像シーケンスを用いて動画顔照合実験を行った結果、部分空間法による照合に比べ大幅に性能が改善されることが分かった。また、擬似ダイバージェンスを上回り、相互部分空間法に匹敵する照合性能を持つことを示した。また計算時間に関しては、相互部分空間法に比べて大幅に削減することが可能であり、実用性も極めて高いことを示した。



(Frames for each query $N=30$, $K=20$, on PIII 866MHz)

図 3-29: 処理時間比較

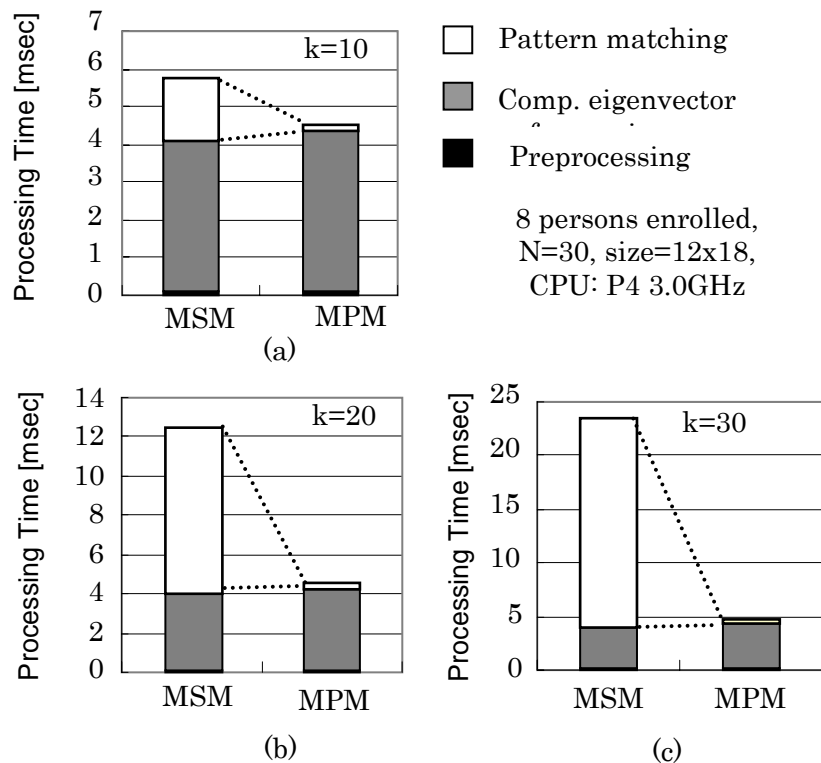


図 3-30: 投影次元数別の処理速度内訳

3.4. 動きベクトルの部分空間射影による高速顔追跡

3.4.1. 高速顔追跡の関連研究

本節では、動きベクトルの部分空間射影を用いた画像追跡法について述べる。物体の運動パラメータを効率的に推定可能であることが特徴である。本研究では、人物顔を対象としてリアルタイム追跡システムを構築するが、追跡アルゴリズム自体は、一般的な剛体運動の追跡処理に利用することができる。

はじめに、画像を用いた人物顔のリアルタイム追跡手法の研究について概観し、提案手法の位置付けについて述べる。人物顔の追跡技術は、ヒューマンコンピュータインタラクション(HCI)、監視、認証システムなど様々な応用が期待されている。しかしながら、照明環境や表情変化によって人物顔の見えが大きく変化するため、人物顔の姿勢を映像から自動的に推定、追跡することは非常に困難とされている。

Black と Yacoob[39]は、2次元の変形可能なテンプレートを用いて顔の動きと表情を追跡した。また Cootes と Taylor[40]は、変形可能なエッジベースの顔モデルを使った追跡技術を提案した。さらに詳細な3次元顔モデルを用いた顔追跡の研究[41][42]においては、コンピュータで生成したテクスチャを含む3次元顔モデルを対象となる顔にモデルをフィッティングさせることで、姿勢推定や顔アニメーションの生成などに適用した。しかしながら、以上述べた研究では、リアルタイムでの実現方法について考慮されていなかった。

リアルタイム顔追跡の関連研究は、大きく2つのカテゴリに分類することができる。第一のカテゴリは、演算性能や精度を高めるために、画像処理ユニット、ステレオカメラ、レンジセン

サ等の特殊なハードウェアを用いる方法である。例えば Harville ら[43]はステレオ画像と特殊ハードを用いて、より高精度なリアルタイム追跡システムを開発した。また松本ら[44]は、ハードウェア画像相関アクセラレータを用いたステレオビジョンによる顔追跡システムを構築した。

リアルタイム顔追跡の第二のカテゴリは、特殊なハードウェアを必要とせずに典型的なパーソナルコンピュータ上で、ビデオレートで動作させる追跡方式の研究である。本研究は、この第二のカテゴリに属する。Gee らは、Temporal consensus tracker と呼ぶ、高速な特徴ベースの追跡方式を提案した[45][46]。Gee らのアルゴリズムはカメラモデルとして Weak Persp こと tive モデルを採用し、Alignment アルゴリズム[47]による 3 点の特徴点からの顔向き推定と、ランダムサンプリング推定とを組み合わせることで、リアルタイム顔追跡を実現した。Jebara ら[48]は運動の表現方法として quaternion[49]を用いた特徴点ベースの顔追跡アルゴリズムを提案した。また Heinzmann ら[50][51]は、アフィン射影を用いた特徴点ベースの追跡方式を提案した。現フレームの特徴点のみから導出される顔姿勢情報は、検出される特徴点の位置が不正確であるためノイズが大きい。よってこのカテゴリの大部分の追跡方式[49][50][51]は、カルマンフィルタによる動きの予測を利用している。すなわち一つ前の状態（位置と速度）から現在の状態を統計的に予測しながら追跡を行っている。しかしながら、物体が突然方向を変えるような場合、カルマンフィルタはうまく適応できない。また実運用においては、環境に応じてフィルタのパラメータ調整が必要である。

近年、リー代数[52]がコンピュータビジョン研究に適用されるようになった。画像中での特定の変換群の動きが、基底ベクトルを用いてどのように記述されるのかについて研究がなされている[53][54]。Drummond らは、物体追跡の研究において、リアルタイムな視覚サーボシステムのための動き推定に、リー代数を導入した[55]。この方式はエッジ領域を追跡するものであり、さらに複雑な 3 次元物体の追跡にも適用され、その有効性が示された[56]。

提案する顔追跡手法は、人物顔の姿勢推定にリー代数を導入する。すなわち顔の特徴点の動きは、特定の変換群 SE(3) (Special Euclidean 3)に含まれると仮定する。SE(3)は 3 次元の剛体運動を表す変換群であり、XYZ 軸周りの回転と並進を表す。SE(3)の自由度は 6 であり、特徴点の動きベクトルの軌跡を 6 つの部分空間に射影することによって、各運動パラメータを直接求めることができる。

本節では、まず提案する顔追跡手法の基本原理について述べ、動きベクトルからの最小二乗法によって剛体運動の 6 つの動きパラメータを簡単に推定できることを示す。そして実際のシステムでの実装方法と、顔追跡結果について述べる。

3.4.2. 動き推定方法

動き推定方法が、本顔追跡システムの心臓部である。動き推定は、透視投影カメラモデルにおける、3 次元特徴点に対する投影行列 \mathbf{P} の推定をベースとする。特徴点 $(X, Y, Z)^t$ の投影座標 $(u, v, w)^t$ は投影行列 \mathbf{P} を用いて次式のように表現できる。

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \mathbf{P} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3-24)$$

画像座標は以下のように与えられる。

$$\begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = \begin{pmatrix} u/w \\ v/w \end{pmatrix} \quad (3-25)$$

投影行列 \mathbf{P} は、カメラ内部パラメータ \mathbf{K} と移動回転行列(Euclidean Matrix) \mathbf{E} との積によってあらわすことができる。移動回転行列 \mathbf{E} は対象物体に対するカメラの相対位置と方向を表す[49]。

$$\mathbf{P} = \mathbf{K}\mathbf{E} \quad (3-26)$$

$$\mathbf{K} = \begin{pmatrix} a_u & 0 & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3-27)$$

$$\mathbf{E} = [\mathbf{R} | \mathbf{t}] \quad \text{with } \mathbf{R}\mathbf{R}^t = \mathbf{I} \quad \text{and } |\mathbf{R}| = 1 \quad (3-28)$$

ここで、 a_u, a_v は水平及び垂直方向の、画像座標系に対するスケーリングを表し、 u_0, v_0 はカメラの光軸中心の座標を表す。 \mathbf{R} は 3×3 の回転行列であり、 \mathbf{t} は移動ベクトルである。

もしも複数の点 $(\tilde{u}_k, \tilde{v}_k)$ において、 (X_k, Y_k, Z_k) との組み合わせが既知であれば、式(3-24)に代入して最小二乗法によって透視行列 \mathbf{P} を求めることができる(自由度 11 のため、最低 6 点の位置情報が必要)。しかしながらこの方法による推定結果は検出誤差などのノイズに弱く、実用的ではなかった。そのため従来手法では、アフィンカメラモデルを利用して自由度を減らす工夫が行われていた[45]。

提案する追跡手法は、透視投影カメラモデルをそのまま用いて行われ、カメラ内部パラメータ \mathbf{K} は事前に与えられるものとする。移動回転行列 \mathbf{E} を推定・更新するアルゴリズムがキーである。すなわち本手法は、カメラの剛体運動パラメータを推定していることを意味する。人物顔は、表情変化等の影響によって正確には剛体ではない。しかしながら表情変化は、姿勢変化と比較してその影響は十分小さいと考え、ここでは剛体運動を仮定する。

移動回転行列 \mathbf{E} は、 3×4 行列であるが、剛体運動の場合は自由度が 6 しかない。この冗長性が、行列の推定が不安定になる原因の一つであった。その推定を行うため、カメラ運動を記述する 6 つのパラメータによって、移動回転行列の推定を行う。

ビデオフレーム間のカメラ運動(カメラが静止しているならば、対象物体の運動を表す)は、以下の \mathbf{M} で表される 4×4 行列の右からの乗算によって表すことができる。

$$\mathbf{M} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{000} & \mathbf{1} \end{bmatrix} \quad (3-29)$$

移動回転行列 \mathbf{E} は、フレーム毎に次式によって更新されるものとする。

$$\mathbf{E}_{t+1} = \mathbf{E}_t \cdot \mathbf{M} \quad (3-30)$$

行列 \mathbf{M} は、SE(3)(Special Euclidean 3)なる群を形成する(\mathbf{R} は正規直交基底)。通常 SE(3)は、

X,Y,Z の移動と、X,Y,Z 軸に関することる回転の 6 つの運動成分に分解することができる。 $\mathbf{M}=\mathbf{I}$ において、6 つの運動成分に対する微分を取ると、式(3-19)に示す 6 つの行列 \mathbf{G}_i が得られる。 \mathbf{G}_i は $\text{SE}(3)$ のリー代数であり、 $\mathbf{M}=\mathbf{I}$ 付近において、剛体運動によって生ずる行列 \mathbf{M} の増分を表す線形ベクトル空間の基底となっている [52][56]。

$$\begin{aligned}
 \mathbf{G}_1 &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
 \mathbf{G}_4 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_5 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_6 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
 \end{aligned} \tag{3-31}$$

図 3-31 は各 \mathbf{G}_i が表す剛体運動成分を示したものであり、矢印はそれぞれの運動によって生じる動きベクトルを示している。

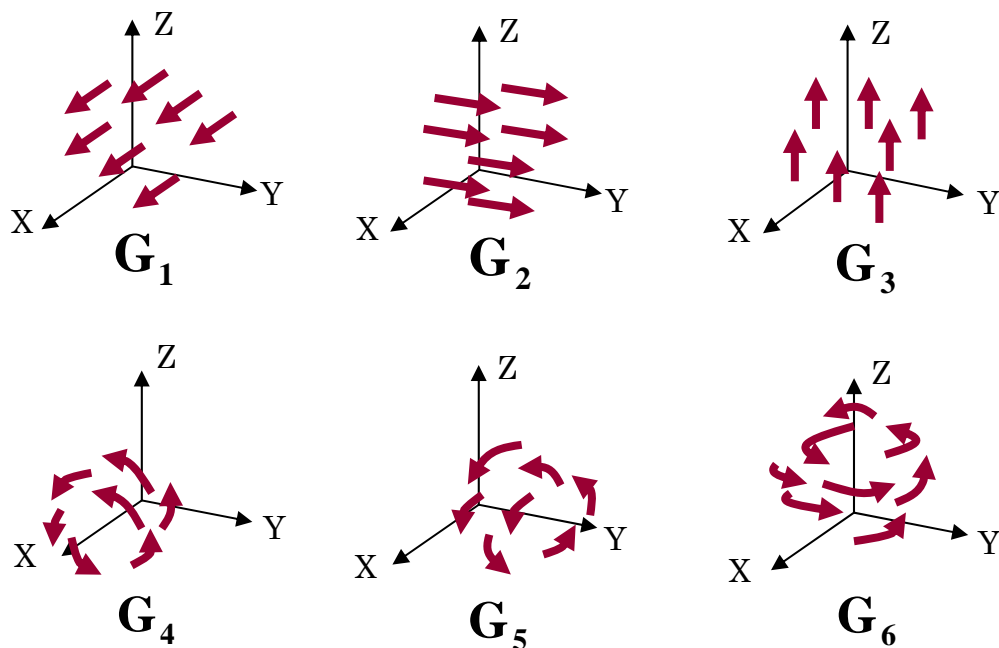


図 3-31: 移動回転行列の構成要素

$\text{SE}(3)$ はリー群である。リー群の性質よりカメラの運動が微小な時、 \mathbf{M} はリー代数 $\mathbf{G}_i : (i=1, \dots, m)$ の線形結合を用いて式(3-32)で表すことができる。($\text{SE}(3)$ の場合は自由度 6 なので $m=6$ である。)

$$\mathbf{M} = \exp\left(\sum_{i=1}^m \alpha_i \cdot \mathbf{G}_i\right) \quad (3-32)$$

式(3-32)をテイラー展開して一次近似を取ると、式(3-33)で表すことができる。

$$\mathbf{M} = \mathbf{I} + \sum_{i=1}^m \alpha_i \cdot \mathbf{G}_i \quad (3-33)$$

すなわち、動画像情報から何らかの方法によって係数 α_i を求めることができれば、行列 \mathbf{M} を算出することができる。各運動要素に対応した係数 α_i を求めるため、 i 番目の運動モード \mathbf{G}_i によってもたらされる運動のもとでの、画像平面(同次座標系)での微分を考える。これは式(3-34)のように表される。

$$\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \mathbf{P} \cdot \mathbf{G}_i \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3-34)$$

式(3-34)を変換することにより、 i 番目の運動モード \mathbf{G}_i によって生じる画像座標系での微分値 \mathbf{L}_i は、式(3-35)のように表される。

$$\mathbf{L}_i = \begin{pmatrix} \tilde{u}' \\ \tilde{v}' \end{pmatrix} = \begin{pmatrix} \frac{u'}{w} - \frac{uw'}{w^2} \\ \frac{v'}{w} - \frac{vw'}{w^2} \end{pmatrix} \quad (3-35)$$

3次元の特徴点座標(X,Y,Z)を式(3-34)に入力することで $(u', v', w)'$ を求め、式(3-35)によってその特徴点における \mathbf{L}_i を算出することができる。ここでは \mathbf{L}_i を運動成分ベクトルと呼ぶ。

図 3-32 を参照し、剛体運動する物体上の特徴点の動きベクトルから α_i を算出する方法について述べる。特徴点の動きが十分小さい時、動きベクトルは運動成分ベクトル \mathbf{L}_i の線形結合で表すことができる。

動きベクトルの長さ d は、係数 α_i と動きベクトルの単位ベクトル \mathbf{n} とを用いて式(3-36)のように表すことができる。また、動きベクトルに垂直な方向の単位ベクトルを $\hat{\mathbf{n}}$ とすると、この方向の成分は0となるので式(3-37)が成り立つ。単位ベクトル \mathbf{n} は式(3-38)によって求められる。

$$d = \sum_{i=0}^m \alpha_i (\mathbf{L}_i \cdot \mathbf{n}) \quad (3-36)$$

$$0 = \sum_{i=0}^m \alpha_i (\mathbf{L}_i \cdot \hat{\mathbf{n}}) \quad (3-37)$$

$$\mathbf{n} = \frac{(\tilde{u}, \tilde{v})_{t+1} - (\tilde{u}, \tilde{v})_t}{|(\tilde{u}, \tilde{v})_{t+1} - (\tilde{u}, \tilde{v})_t|}, \quad (\mathbf{n}, \hat{\mathbf{n}}) = 0 \quad (3-38)$$

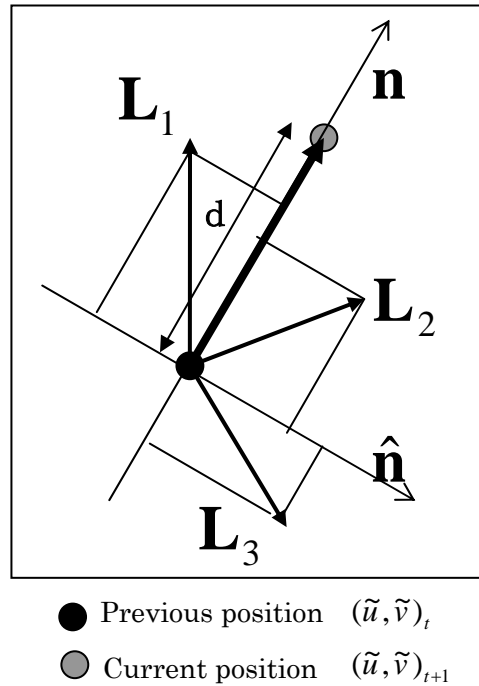


図 3-32: 特徴点の動きベクトルと運動成分ベクトル \mathbf{L}_i との関係

式(3-36)と式(3-37)より、複数の特徴点の画像座標系における動きベクトルを観測すれば、式(3-39)が成り立つ。最小二乗法によって式(3-39)を解くことにより、各運動要素に対応する係数 α_i を推定することができる。

係数 α_i を求める操作は、複数の特徴点の動きベクトルを 6 つの運動モード \mathbf{G}_i によって張られる部分空間に射影し、その射影成分を求めることと等価である。但し \mathbf{G}_i に直接射影するのではなく、画像座標系に変換された運動成分ベクトル \mathbf{L}_i を用いて推定するのが特徴である。

$$\begin{pmatrix} \sum_i a_i(L_i \bullet n_1) \\ \vdots \\ \sum_i a_i(L_i \bullet n_j) \\ \sum_i a_i(L_i \bullet \hat{n}_1) \\ \vdots \\ \sum_i a_i(L_i \bullet \hat{n}_j) \end{pmatrix} = \begin{pmatrix} L_1 n_1 & L_2 n_1 & L_3 n_1 & L_4 n_1 & L_5 n_1 & L_6 n_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ L_1 n_j & L_2 n_j & L_3 n_j & L_4 n_j & L_5 n_j & L_6 n_j \\ L_1 \hat{n}_1 & L_2 \hat{n}_1 & L_3 \hat{n}_1 & L_4 \hat{n}_1 & L_5 \hat{n}_1 & L_6 \hat{n}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ L_1 \hat{n}_j & L_2 \hat{n}_j & L_3 \hat{n}_j & L_4 \hat{n}_j & L_5 \hat{n}_j & L_6 \hat{n}_j \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{pmatrix} = \begin{pmatrix} d_1 \\ \vdots \\ d_j \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3-39)$$

係数 α_i が求めれば、移動回転行列 \mathbf{E} は式(3-30)と式(3-33)とを用いて以下のように算出することができる。

$$\mathbf{E}_{t+1} = \text{Euclideanise} \left(\mathbf{E}_t \cdot \left(\mathbf{I} + \sum_i \alpha_i \cdot \mathbf{G}_i \right) \right) \quad (3-40)$$

式(3-40)で、 $\text{Euclideanise}()$ は $\mathbf{E} = [\mathbf{R} | \mathbf{t}]$ 中の 3×3 部分行列 \mathbf{R} を正規直交基底に強制的に近づける処理である。これは例えば、特異値分解を用いて次のように実現できる。 \mathbf{R} を特異値分解すると、中間の対角行列の要素は本来 1 となる(単位行列)はずである。そこで次のように、 \mathbf{R} を特異値分解し、分解結果のうち \mathbf{U} と \mathbf{V} だけを用いて再構成することにより新しい \mathbf{R}' を求める。式に $\text{Euclideanise}()$ の実現例を示す。

$$\begin{aligned} \text{Euclideanise}(\mathbf{E}) &= [\mathbf{R}' | \mathbf{t}] \\ \mathbf{R} &= \mathbf{U}\mathbf{W}\mathbf{V}^t, \quad \mathbf{R}' = \mathbf{U}\mathbf{V}^t \end{aligned} \quad (3-41)$$

3.4.3. 顔追跡システムの実装

本顔追跡システムは、主にコンピュータのモニタ付近に備え付けられたカメラを用い、コンピュータを操作しているユーザの顔、頭部をトラッキングし、顔の姿勢、向きをリアルタイムに算出、表示するものである。追跡はあらかじめ定めた特徴点(目、口、鼻、眉毛)等を探索しながら行われ、それらの位置関係から顔の向きを求める。なお顔の特徴点の概略 3 次元位置は既知であるとする。またカメラの内部パラメータと外部パラメータは既知とする。図 3-33 に顔追跡処理の流れについて概要を示す。

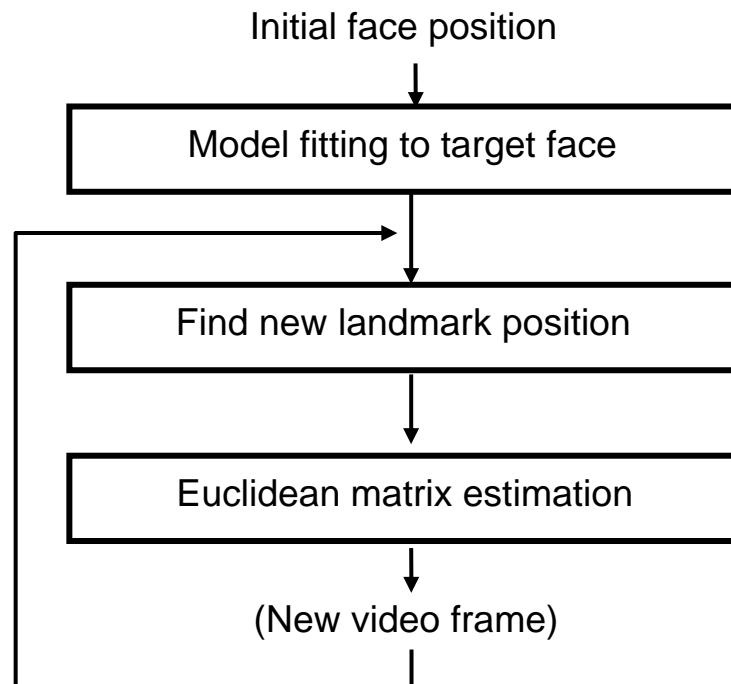
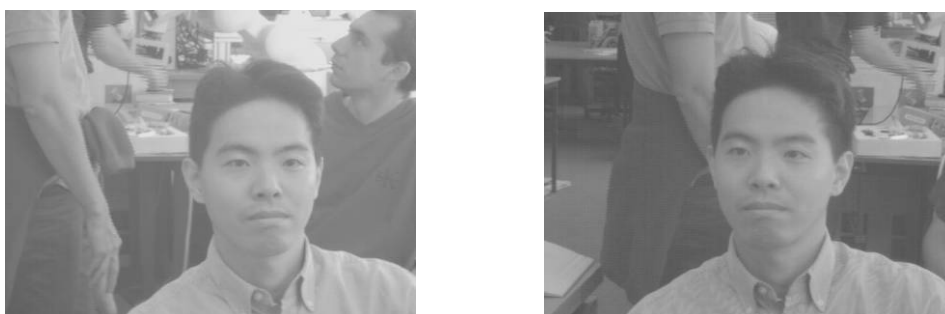


図 3-33: 顔追跡処理の流れ

本顔追跡システムは、事前に顔の特徴点の3次元位置が既知であることが必要である。そこで、ステレオ視のアルゴリズムを用いて特徴点の3次元位置を求めた。図 3-34 に左右カメラの画像を示す。それぞれの画像から対応点を手動で求め世界座標 X を求める。図 3-35 に得られた3次元特徴点の例を示す。本実験では、図 3-35 の特徴点から瞳中心、眉毛中心、鼻孔、口端の合計8点の特徴量を用いる。



(a) 左画像

(b) 右画像

図 3-34: 特徴点位置取得に用いたステレオ画像例

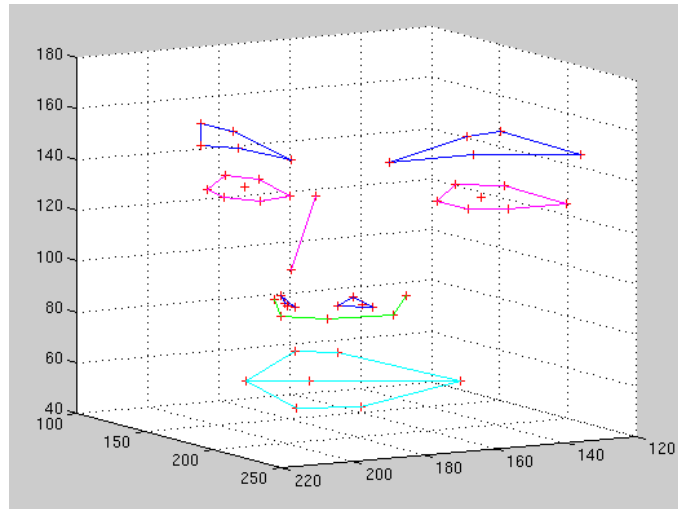


図 3-35: ステレオ画像からの 3 次元位置取得例

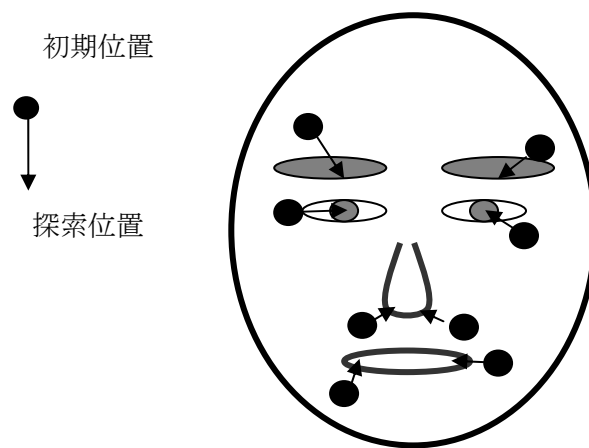


図 3-36: モデルフィッティング(特徴点の初期位置と探索後位置の例)

まず、モニタ上に顔の特徴点モデルを投影し、手動で追跡したい顔の概略位置に合わせる。これが初期特徴点位置である。しかしながら、用いる特徴点モデルは標準的な顔データから作成されるので、個人の固有の特徴を反映していない。従って、手動であっても正確には合わせることはできない。そこで、各特徴点が追跡対象の顔に概略重なったところでモデルフィッティング処理を行う。

図 3-36 にモデルフィッティング処理の概要を示す。図中の黒点が初期特徴点位置である。各特徴点の近傍を探索し、最も特徴が類似している点を求める。図 3-36 ではこれを矢印の先で示している。初期位置と探索後位置をそれぞれ (u, v) , (u', v') とする。モデルフィッティング処理は、画像上の特徴点近傍で、より正確な特徴点位置を探索し、カメラの視線方向の深さを変更することなく画像平面の xy 座標のみを変更することで、新しい顔の 3 次元特徴点座標を得るものである。

初期特徴点位置に関して式(3-42)が成り立つものとする。 z はカメラ座標系における Z 値であり、特徴点までの Z 方向の距離を表す。カメラパラメータは既知とするので、 z は特徴点毎に直ちに求められる。探索後の新しい特徴点についても z は変わらない(視線方向の距離は変わら

ない)と仮定すれば、式(3-43)が成り立つ。これを変形した式(3-44)によって、新しい特徴点の3次元位置(X', Y', Z')が求められる。

$$\begin{pmatrix} zu \\ zv \\ z \end{pmatrix} = \mathbf{K} \cdot [\mathbf{R} | \mathbf{t}] \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \mathbf{K} \cdot \mathbf{R} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \mathbf{K} \cdot \mathbf{t} \quad (3-42)$$

$$\begin{pmatrix} zu' \\ zv' \\ z \end{pmatrix} = \mathbf{K} \cdot \mathbf{R} \cdot \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} + \mathbf{K} \cdot \mathbf{t} \quad (3-43)$$

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = (\mathbf{K} \cdot \mathbf{R})^{-1} \left(\begin{pmatrix} zu' \\ zv' \\ z \end{pmatrix} - \mathbf{K} \cdot \mathbf{t} \right) \quad (3-44)$$

モデルフィッティングが完了すると、ユーザの指示によって顔追跡処理が開始される。顔追跡処理は次のステップによってフレーム毎に繰り返し実行される。

1. 新しいフレームの画像を取得し、現在の特徴点の位置を投影する。
2. 画像平面上で各特徴点の近傍を探索して新しい特徴点位置を求め、動きベクトルを算出する。
3. 特徴点の動きベクトルから、3.4.2 節の手法を用いて移動回転行列(Euclidean Matrix) \mathbf{E} を求める。求められた \mathbf{E} を用いて新しい3次元特徴点座標を得る。
4. 新たな3次元特徴点を画像平面に投影し、特徴点の画像座標を更新し再描画を行う。

実際に構築したシステムでは、追跡する特徴点として局所的な黒点を探索する方式を採用した。局所黒点探索法は、単純に近傍画素で最も輝度値の低い画素を出力するものである。特徴点探索としては、他にも正規化相関によるテンプレートマッチング、コーナ特徴による探索が使用できるが、リアルタイム処理するためには使用する探索アルゴリズムに応じて特徴点の数を制限する必要がある。今回採用した局所黒点は、処理時間も少なく、リアルタイム処理には最も適している。

画像座標の更新・描画ステップにおいて、もしも各特徴点における表面の法線ベクトルが既知であれば、カメラの向きに応じて、可視点、不可視点の判定が可能である。もしも不可視点が存在すると、システム上はどこかに点が見えているものとして探索するため、実際とは異なる座標値を求めてしまう。、そして最終的に得られる運動パラメータ推定の誤差が増大してしまうという問題がある。可視点の判定が可能の場合、もしも不可視と判断されればそのフレームに関して不可視点を描画せず、また探索処理も行わないように制御することができる。これによって顔が横を向いた時などに、特徴点の誤検出を防いで追跡精度の向上を図ることができる。

本システムでは、顔の形状はほぼ凸であると仮定して、顔の表面の法線ベクトルを用いた可視判定を行っている。これは図 3-37 に示すように、特徴点における法線ベクトル \mathbf{n}_c と、カメラの光軸ベクトル \mathbf{v}_c との内積 $\cos \theta$ を評価することで行われる。内積 $\cos \theta$ は式(3-45)から求められる。原理的には内積が 0 以下であれば不可視となる。

$$\mathbf{v}_c = (p_{31}, p_{32}, p_{33})^t$$

$$\cos \theta = \frac{(\mathbf{v}_c, \mathbf{n}_s)}{|\mathbf{v}_c| \cdot |\mathbf{n}_s|} \quad (3-45)$$

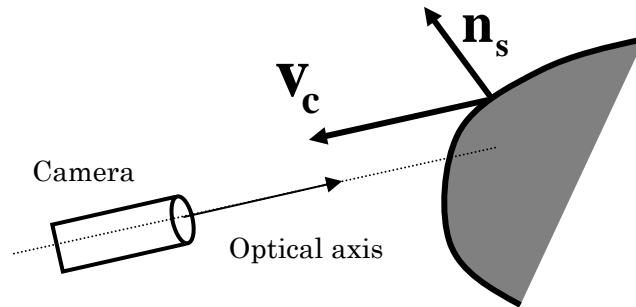


図 3-37: 表面法線ベクトルによる隠れ処理

3.4.4. 実験結果

顔追跡システムによる処理結果を以下に示す。特徴点の位置として、眉毛、目、鼻、口の左右各 2 点ずつ、合計 8 点の特徴点を用いた。なお特徴点の数としては、最低 3 点が必要である。特徴点数が多いほどシステムは頑強になる。

図 3-38 に、顔のモデルフィッティング結果を示す。(a)はフィッティング前、(b)はフィッティング後の画像である。初期特徴点位置による位置合わせは手動であるが、初期位置が多少ずれていても、モデルフィッティングによって正確に位置決めできることが分かる。



(a) 標準顔モデルの初期位置

(b) モデルフィッティング後の位置

図 3-38:モデルフィッティングの例

図 3-39 に、ビデオカメラ映像を用いた顔追跡実験結果を示す。(a)から(f)は一連の追跡映像からキャプチャした画像である。これらはそれぞれ異なる顔姿勢であるが、特徴点が正確に追従

していることが分かる。本手法は、全ての特徴点の動きベクトルを用いて剛体運動を推定するため、一部特徴点の誤検出によって極端に不連続な結果になることが少ない。また、(f)では法線ベクトルによる不可視処理の一例が示されている。右の鼻が不可視と判断されて、画面に表示されていないことが分かる。

本実験は、SGI社のワークステーション O2(CPU: MIPS-R5000, 225MHz)の上で行われ、ビデオフレームサイズは768×576であった。また局所黒点探索は、近傍の21×21画素を探索した。そして、画像の取り込み、追跡処理、表示までのサイクルを、全てビデオフレーム内でリアルタイムに行うことができた。

3.4.5. 高速画像追跡のまとめ

本顔追跡アルゴリズムは高速で、剛体の物体追跡にも利用できる汎用的な手法である。それでも顔の動きが早すぎる場合には、追跡が間に合わない場合がある。追跡する特徴点の探索範囲を増やす、あるいは特徴点数を増やすことで改善が期待できるが、人間の顔表面の安定な特徴点を見つけ出すことが課題である。

また表情変化への追従に対しては現状未対応である。基本的に剛性物体モデルなので、表情が変わった時などに、姿勢推定に失敗する場合がある。

本手法の優位な点は、自由度が6であるにもかかわらず3×4の要素を持つ移動回転行列を、6つの線形結合されたパラメータで表現可能なため、従来手法に比べて頑健な推定が可能であることである。またこれらはx、y、zの移動及びx、y、z軸周りの回転であり、実際の運動方向と対応がある。よって移動パラメータに対してカルマンフィルタを作用させたり、運動の方向を制限したりすることも可能であり、応用性・拡張性の高いアルゴリズムである。

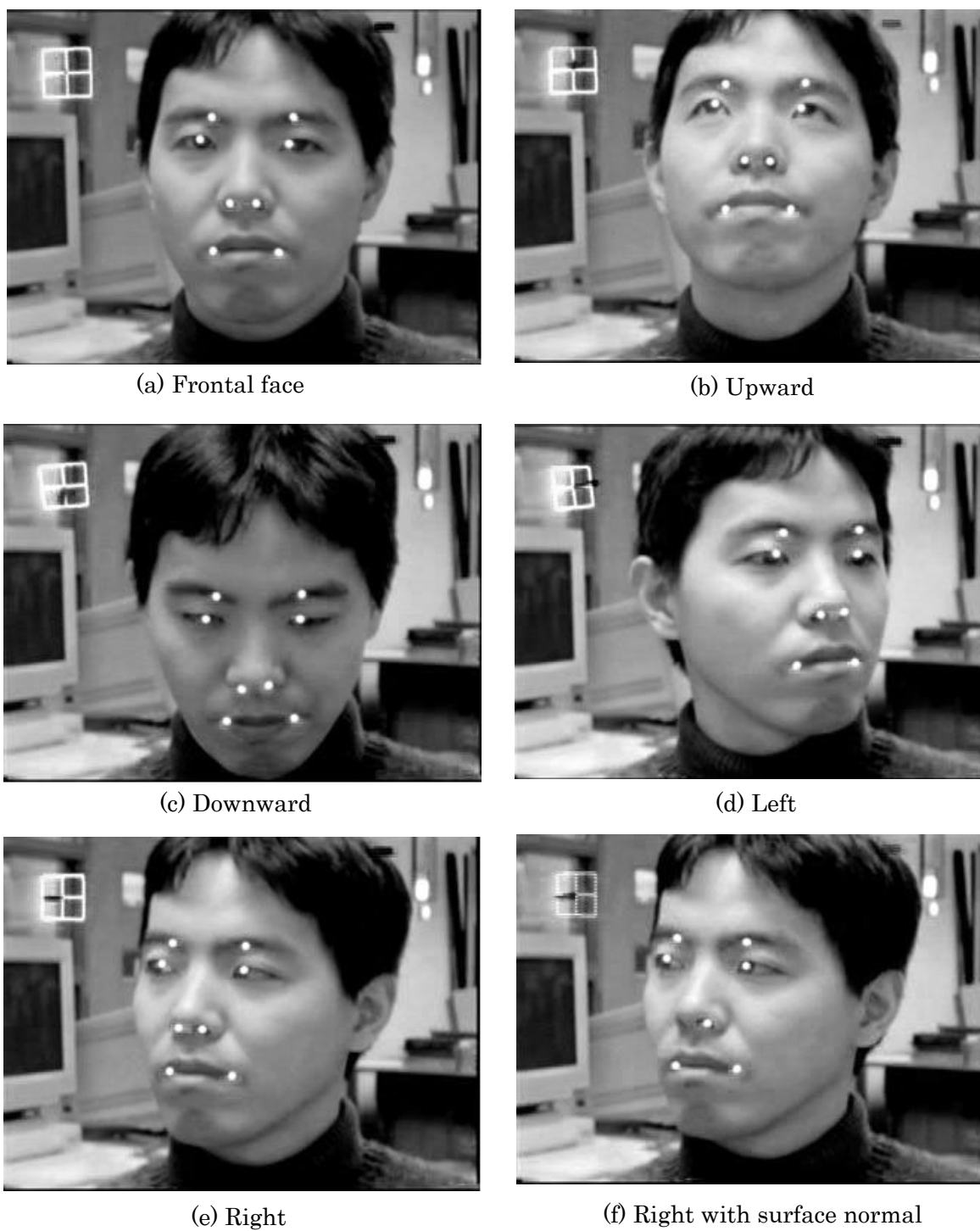


図 3-39: 顔追跡実験結果

3.5. 画像認識のまとめ

本章では、部分空間情報を利用した画像認識処理の研究成果について述べた。画像認識におけ

る具体例として、様々な対象の中でも環境変動の影響が非常に大きいとされる人間の顔照合・顔追跡を課題に適応性の改善を試みた。

はじめに静止画顔照合の改善手法として、適応的部分領域照合法と摂動空間法を提案した。本研究は、部分空間選択の適応化と、部分空間の比較・統合のアプローチに基づいており、入力画像に応じて、照合に用いる部分空間を自動的に変化させながら照合を行うことを示した。適応的部分領域照合法は、適切な部分領域を選択して領域毎に照合を行う方法であり、蔽や表情変化等への頑健性を向上させた。摂動空間法は 1 枚の画像を元に摂動画像群を生成し、それらの画像群の張る部分空間（摂動空間）を用いて照合を行う方法である。摂動空間法により照明変動や姿勢変動に対する頑健性を向上させた。

次に動画像を用いた顔照合方式の検討を行った。本研究は、部分空間の比較・統合のアプローチに基づいており、画像シーケンスによって形成される部分空間同士の比較方法として相互投影距離法を提案した。本手法は部分空間同士を直接比較するため、動画像特有のフレーム間の相関を照合スコアに反映させることができる。照合実験により、従来法に比べて照合精度を維持したまま、大幅に演算量を削減可能であることを示した。

最後に、顔追跡における新たな部分空間解析手法を提案した。本手法は部分空間に基づく適応動作のアプローチに基づいている。特徴点の動きベクトルを剛体運動の部分空間に射影することで、剛体運動における 6 自由度の運動パラメータを直接に求めることが特徴である。これにより、物体の姿勢変動を高速に且つ安定して推定することが可能となった。本手法を人物顔追跡システムに適用し、ソフトウェア処理での実時間顔追跡を実現した。

以上の研究成果より、部分空間情報の適応的解析による画像認識処理の改善を実現した。

第4章 部分領域からの画質特性解析による画像強調

4.1. まえがき

本章では、画像の部分空間情報を利用した画像強調処理について報告する。画像強調における具体例として、人間が観察して最適な画質となるよう、自動的に強調処理を行う研究について述べる。

本研究では、スキャナ入力された写真やデジタルカメラ、ビデオカメラ等から得られたデジタルカラー画像を対象とし、劣化に関する事前の知識なしに、入力画像に応じて自動的に画質を改善する手法を提案する。本手法は入力画像信号を、ディスプレイの色空間内で最適化するものである。

これらの画像をそのままディスプレイに出力すると、多くの場合画質が低下しているのが分かる。その原因は、写真、ビデオ映像などの素材によるもの、センサの特性によるもの、ディスプレイ等出力装置の特性によるものなど、様々である。従来、これらの個々の劣化特性を明らかにし、それらに応じた画像の復元、補正によって画質改善を図ることが行われてきた。しかしながら、一般に流通するデジタルカラー画像を取り扱う場合、上記の劣化のメカニズムは未知であるため、実際には手動で対話的に画質調整を行う必要があった。本研究では入力画像を解析して画質特性値を算出し、最適な画像強調パラメータを推定する方法について述べる。

人間の視覚システムは、周辺環境に適応するため複雑な特性を有している。例えば同じ物体を異なる光源環境で物体を観察しても、人間は色の違いを感じることなく同じ色だと知覚する場合がある。これはカラーコンスタンス[57]として知られる現象である。また、入射光の強さが大きく変化しても、物体の明るさ知覚は変化しないという現象がある。これはライトネスコンスタンス[58]と呼ばれる。このような現象は視覚システムのグローバルな環境適応と考えられる。これらの適応化メカニズムは von Kries[59]と MacAdam[60]によって単純にモデル化された。このモデルでは、環境適応は、網膜にある 3 種類の受容体の応答が等価となるように、それぞれ独立に信号を制御するモデルによって説明されている。この von Kries らのモデルは、デジタルカメラのホワイトバランス補正機能などに幅広く利用されているが、被写体の物体色の影響を考慮していないので、撮影画像の色かぶりや色あせなどの画質劣化を生じる場合がある。

また、人間は特定の自然物に対し、生まれた時からその物体色を記憶していると言われている。これらの色は記憶色と呼ばれる[61]。記憶色は、対象となる自然物の実際の色と測色的に一致している訳ではない。しかし、その自然物の色を再現する際には、記憶色を用いて再現するほうが、人間にとって好ましい色と感じられると言われている。この記憶色を再現する色再現は、自然物の色を人間にとってより望ましい色に変更する処理であり、好ましい色再現と呼ばれる[62]。特定の自然物とは、例えば草木の緑、空の青、人肌、りんごの赤などがある。本研究では、記憶色の補正を、特定の対象物のための局所的な(ローカルな)色補正として考える。すなわち、環境光の変化に対するグローバルな補正処理と、対象物が占める部分領域にのみ作用するローカルな補正処理とを同時に行うことで、人間の視覚システムの持つ適応性に迫ることができる。

画質の主観評価の観点から、人間が知覚する画質の良し悪しについて、いくつかの要因が存在する。例えば、コントラスト、明度、彩度、鮮鋭度、色バランス、光沢感、ノイズ感、記憶色などである。これらの画質要因は、我々の心理的な画質を分析する軸を表している。実際、これらの軸は、画像機器のための手動パラメータ調整におけるパラメータとして利用されている。

本研究は、これらの複数の画質要因を入力画像から解析することにより、視覚の動的な適応性を実現する最適な画像強調処理について述べる。一般に、画像機器における画像強調処理は、固定パラメータによるトーンカーブ補正やマトリクス変換によって行われている。そしてそのパラメータ値は、様々な画像シーケンスから得られた平均的な値を用いるのが一般的である。この方法では 2 つの大きな問題がある。第 1 の問題は、固定パラメータがそれぞれのシーンに

において最適ではないという点である。例えば、コントラストを強調するトーンカーブパラメータは、田園風景シーンの画質を改善する一方で、人間の顔の滑らかな階調を損なってしまう。第2の問題は、特定の物体だけに補正処理をすることができないという点である。

そこで近年、いくつかの適応的色補正の研究が発表されている。Rizzi ら[75]は、グローバルとローカルな影響を両方持つような強調手法を提案した。この方法は色バランス補正と、グローバルとローカルの明度補正を含んでいる。しかしながら記憶色の補正は含まれておらず、演算量も大きいという問題があった。Spitzer ら[76]は、網膜の応答モデルを用いた適応的色変換手法を提案した。この方法は主に局所的な色適応性について焦点を当てており、明度補正や記憶色については対象外となっている。記憶色の補正については、文献[77]において画像の彩度成分が記憶色にとって重要だが、色の自然さを再現するのは困難との報告がある。このように従来はそれぞれ個別領域における強調処理が研究されており、それらを統合した画像強調の研究は少ない。また、自然画像全般を対象とし、演算量も画像機器に適用可能なレベルの研究成果はほとんど存在しない。

本章では、まず鮮鋭度の最適な強調方法について述べる。次に HSV 座標系を用いてカラー画像に適用し、彩度、コントラスト、鮮鋭度の最適化について述べる。そして部分領域に対する新たな色補正方法について述べた後に、6つの画質要因を統合した自動高画質化技術について述べる。

4.2. エッジ領域の解析による画像の最適鮮鋭化

4.2.1. 鮮鋭度の最適化

本節では、デジタルカメラ等から入力された自然画像を対象とし、劣化に関する事前の知識なしに入力画像に応じて鮮鋭さを最適に調節する手法を提案する。画像の鮮鋭さは高すぎても低すぎても画質は低下すると言われている。よって最適鮮鋭化を実現するためには、入力画像の鮮鋭さを評価し、その値が最適になるように画像を変換しなければならない。

画像の鮮鋭さを評価するためには、一般に振幅伝達関数(Modulation Transfer Function: MTF)が用いられる[57][64][65]。しかし写真のスキャナ入力画像などを取り扱う場合は、ふつう MTF などの劣化のメカニズムは未知である。よって従来、このような劣化過程が未知の画像の鮮鋭さを評価することはできなかった。

この問題を解決するため、本研究では MTF を用いずに、エッジシャープネス(ES value)を定義し、この値を用いて画像の鮮鋭さを評価する。エッジシャープネスと主観的な鮮鋭さとの関係を調査し、これらに強い相関関係があることを確認する。さらに、エッジシャープネスと主観的な画質との関係を調査し、エッジシャープネスの最適値を求める。そして画像を任意のエッジシャープネスに変換する手法を考案し、本手法を用いた画像の最適鮮鋭化の有効性を確認する。

4.2.2. エッジシャープネスの定義

人間の視覚システムは、画像の鮮鋭さを判断する際に、主にエッジ領域の特徴に注目している。そこで筆者らは、エッジ領域の高空間周波数成分の平均強度によって鮮鋭度を定義することを提案し、この値をエッジシャープネスと呼ぶことにする。

エッジシャープネスは式(4-1)によって定義される。この中で、 $f(x, y)$ は入力画像、 $E(f)$ は入力画像のエッジ領域を表し、 $A_E(f)$ は $E(f)$ の面積を示す。 $SS(u, v)$ は空間周波数成分の主観的な鮮鋭さに対する重みを表す高周波帯域フィルタであり、 $ss(x, y)$ はその逆フーリエ変換である。式(4-1)の分子は、エッジ領域 $E(f)$ における、高周波成分の積分を表す。 ES_f はエッジ領

域における高周波帯域成分の平均値として定義されており、この値を特性値として画像の鮮鋭さを評価する。

$$ES_f = \frac{\iint_{E(f)} |f(x, y) \otimes ss(x, y)| dx dy}{A_E(f)} \quad (4-1)$$

鮮鋭さは人間の視覚によって感じられる尺度であるため、 $SS(u, v)$ は視覚の空間周波数特性を考慮しなければならない。人間の視覚の空間周波数特性は、帯域濾過特性を示す[66]。よって非常に高い周波数成分は、もはや鮮鋭さに影響を与えなくなる。従って、 $SS(u, v)$ には高周波側に、視覚特性と一致するカットオフを設定する。

文献[67]において、鮮鋭さの重み関数 $Gr(u)$ は、視覚の空間周波数特性 $R(u)$ と、システムのMTF $H(u)$ によって、式(4-2)として定義されている。

$$Gr(u) = 2uR(u)H(u) \quad (4-2)$$

$SS(u, v)$ のコンセプトは $Gr(u)$ と基本的に同様であり、実験では式(4-3)に示すガウシアン差分型のフィルタ $ss(x, y)$ を用いた。図 4-1 に $SS(u, v)$ の空間周波数特性を示す。

$$ss(x, y) = A \exp\left(-\frac{x^2 + y^2}{2\sigma_1^2}\right) - B \exp\left(-\frac{x^2 + y^2}{2\sigma_2^2}\right) \quad (4-3)$$

$$A = \frac{1}{2\pi\sigma_1^2}, \quad B = \frac{1}{2\pi\sigma_2^2}$$

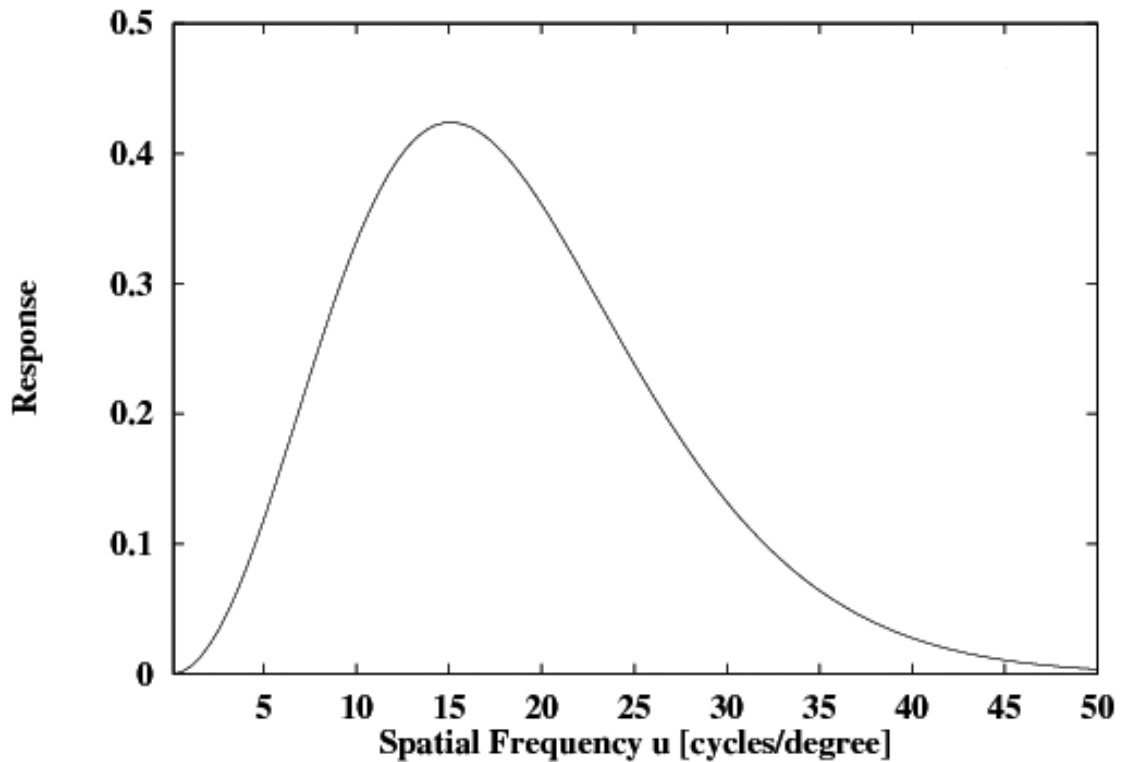


図 4-1: SS(u)の特性

式(4-1)はエッジ領域に関する量($A_E(f)$ と $E(f)$)を含んでいるため、エッジ検出フィルタが必要である。本実験では Sobel フィルタをエッジ検出に用いた。

4.2.3. 主観的鮮鋭さとエッジシャープネス

エッジシャープネスと主観的鮮鋭さとの関係を調査するため、自然画に対して主観評価実験を行った。表 4-1 に、実験の観視条件を示す。8 種類の画像について、それぞれ 6 段階の鮮鋭さを持つ画像を用意し、主観的鮮鋭さとエッジシャープネスを比較した。なお本実験におけるエッジシャープネスは、画像の明度値を L^* (CIE 1976- $L^*a^*b^*$) に変換した上で計算した。評価には 5 段階の評定尺度法を用いた(表 4-2)。

実験結果を図 4-2 と図 4-3 に示す。図 4-2 は被験者 A のデータを示し、図 4-3 は全実験データを示す。全実験データから得られた相関係数は、 $R = 0.793$ であった。本実験結果より、エッジシャープネスと主観的鮮鋭さとは線形な相関関係にあることが分かった。

表 4-1: 観視条件

Display	N7832-23(NEC) 20inch
Display size	1280×1024 pixels
Viewing distance	0.5 m
Umax (spatial frequency)	15.96 cycles / degree
Picture size	512×512 pixels
Number of pictures	8×6 grades
Gray level	8 bits

表 4-2: 主観評価実験における鮮鋭さの段階

Grade	Sharpness
5	Sharp
4	Slightly Sharp
3	Normal
2	Slightly Blurred
1	Blurred

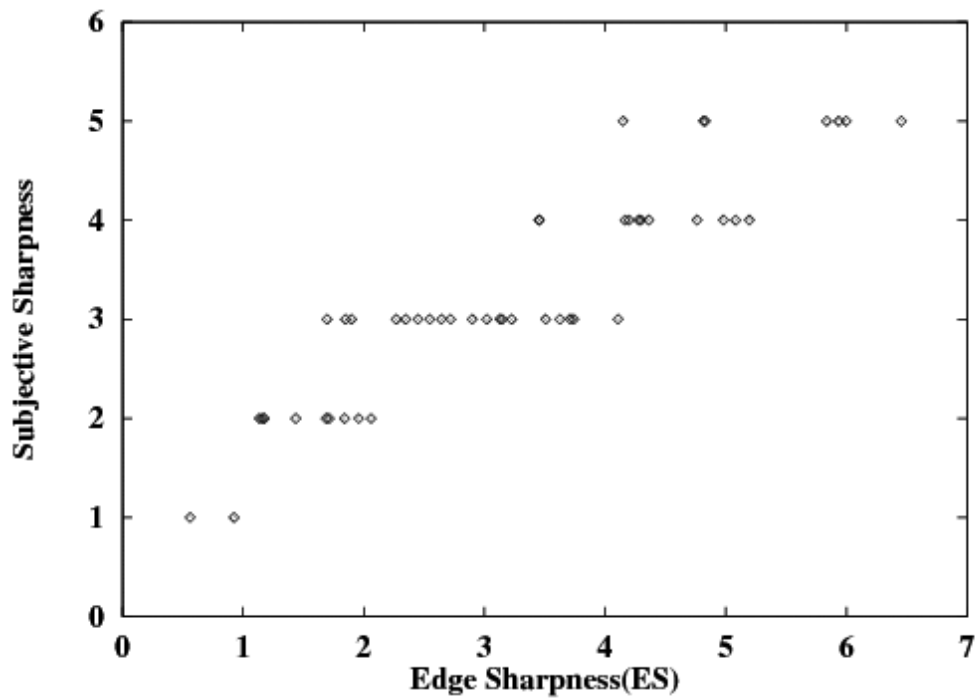


図 4-2: エッジシャープネス(ES 値)と主観的鮮鋭さとの関係(観察者 A)

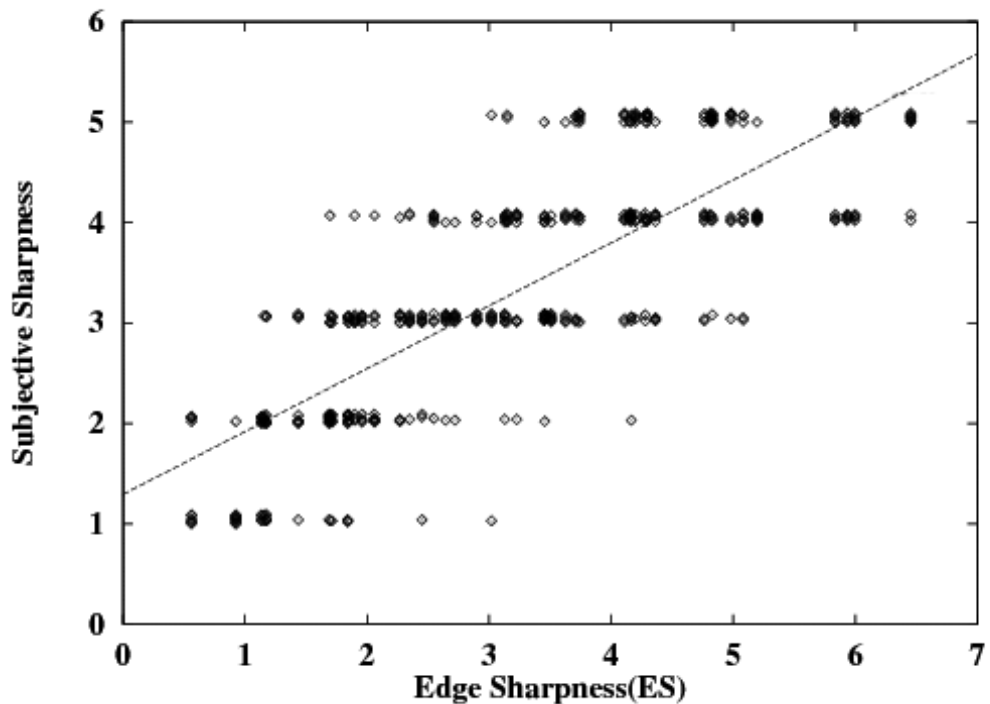


図 4-3: エッジシャープネス(ES 値)と主観的鮮鋭さとの関係(全データ)

4.2.4. 画質とエッジシャープネス

次に画質とエッジシャープネスとの関係を調査するための主観評価実験を行った。観視条件は、先の実験と同じである(表 4-1)。評価尺度を表 4-3 に示す。

表 4-3: 主観評価実験における画質評定

Grade	Quality
+1	The best
3	Good
2	Fair
1	Poor

図 4-4 に、画質とエッジシャープネスとの関係を調査した実験結果を示す。これより、エッジシャープネスがある一定の範囲にあれば、画質が良くなることが分かる。さらに、画質を最適にするエッジシャープネス値(最適エッジシャープネス)が存在することが予想される。実験結果より、最適エッジシャープネス値として $ES_{opt} = 2.63$ が得られた。

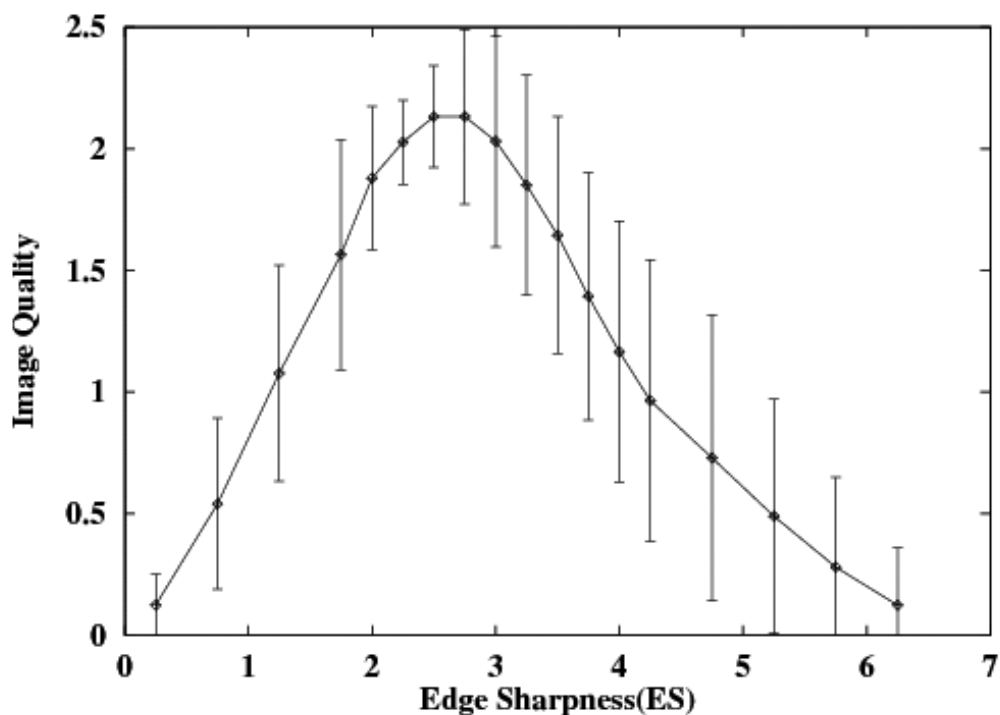


図 4-4: エッジシャープネス(ES 値)と主観画質との関係(エラーバーは σ)

4.2.5. 鮮鋭度変換アルゴリズム

画像のエッジシャープネスを調節するためには、画像の高周波成分を制御しなければならない。基本的な変換式としては、通常の鮮鋭化手法である式(4-4)を用いる。

$$fs = f + k(f \otimes ss) \quad (4-4)$$

f は入力画像 $f(x, y)$ を示し、 fs は鮮鋭度変換後の画像である。 ss は高域濾過フィルタであり、 k は強調係数である。従って、 $k > 0$ の時は鮮鋭化画像が得られ、 $k < 0$ の時には暈けた画像が得られる。ここで ss は、式(4-3)の $ss(x, y)$ を使用することにする。

ここで求めたいのは、所望のエッジシャープネスに変換するための k の値である。式(4-1)と式(4-4)より、変換後のエッジシャープネス値 ES_{fs} は式(4-5)で表すことができる。この式(4-5)を解くことによって所望の k を求める。

$$\begin{aligned} ES_{fs} &= \frac{\iint_{E(fs)} |(f + k(f \otimes ss)) \otimes ss| dx dy}{A_E(fs)} \\ &= \frac{\iint_{E(fs)} |f \otimes ss + k(f \otimes ss) \otimes ss| dx dy}{A_E(fs)} \end{aligned} \quad (4-5)$$

式(4-5)は解析的に解くことはできないので、式(4-6)の近似を用いて分子を単純化する。

$$\begin{aligned} &\iint_{E(fs)} |f \otimes ss + k(f \otimes ss) \otimes ss| dx dy \\ &\approx \iint_{E(fs)} |f \otimes ss| dx dy + k_a \iint_{E(fs)} |f \otimes ss \otimes ss| dx dy \end{aligned} \quad (4-6)$$

式(4-6)を式(4-5)に代入し、 k_a を k の近似として求める。 k_a は式(4-7)で算出することができる。

$$k \approx k_a = \frac{ES_{fs} \cdot A_E(fs) - \iint_{E(fs)} |f \otimes ss| dx dy}{\iint_{E(fs)} |f \otimes ss \otimes ss| dx dy} \quad (4-7)$$

式(4-7)より、 $k > 0$ の時は $k > k_a$ となり、 $k < 0$ の時は $k < k_a$ となることが分かる。 k の代わりに k_a を採用することによって強調が過大になることはないため、 k_a による実用上の問題はない。実際の画像による実験から、 k と k_a の違いは小さく、主観的鮮鋭さへの影響はほとんどなかった。

式(4-7)は以下のように反復法で解く。図 4-5 に処理手順を示す。初期値である k_0 は、 $E(fs) = E(f)$ 及び $A_{E(fs)} = A_{E(f)}$ を仮定することで、式(4-8)によって求めることができる。そしてまず、 k_0 を用いて鮮鋭度変換し、画像 $f_1(x, y)$ を得る。次に f_1 のエッジ領域 $E(f_1)$ を求めた後、新たな $k = k_1$ を式(4-7)から計算する。そして数回の反復の後、所望の k が得られる。

これを式(4-4)に適用することによって、入力画像のエッジシャープネス ES_f は ES_{fs} に変換される。

$$k_0 = \frac{(ES_{fs} - ES_f) \cdot A_E(f)}{\iint_{E(f)} |f \otimes ss \otimes ss| dx dy} \quad (4-8)$$

本アルゴリズムを用い、画像を最適エッジシャープネス値 ES_{opt} に変換することによって、自動的な最適鮮鋭化を行うことが可能となる。

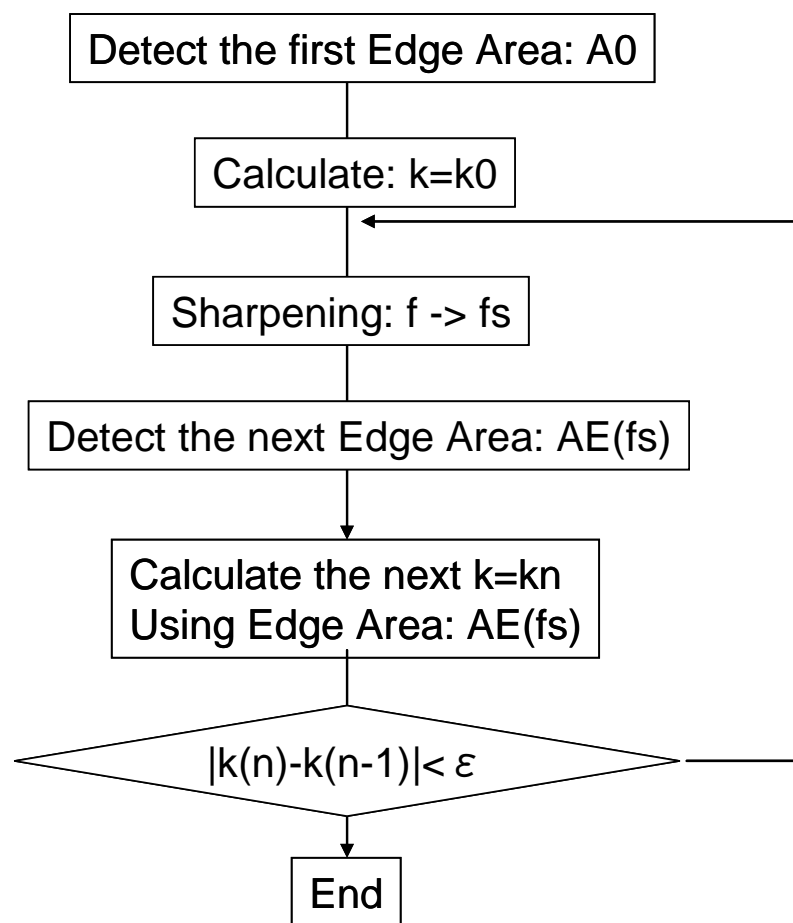


図 4-5: 最適鮮鋭化の処理手順

4.2.6. 実験結果

この最適鮮鋭化手法を、いくつかの写真のスキヤナ入力画像に適用した。表 4-4 はその時の実験結果である。 ES_0 は入力画像のエッジシャープネス、 Q_0 は入力画像の主観画質評価値である。 ES_n は出力画像のエッジシャープネスで、 Q_n は出力画像の主観画質評価値である。画質評価は

3 段階評価で行われた。 k は最適鮮鋭化アルゴリズムによって算出された強調係数である。表 4-4 より、出力画像のエッジシャープネスは、ほぼ目標とする最適値に調整されている事が分かる。また本実験では全ての画像について、主観画質評価値が上昇しており、画質が改善されたことが分かる。また強調係数 k は、入力画像の鮮鋭さに応じて適応的に変化していることが分かる。

表 4-4 自然画に対する最適鮮鋭化の実験結果
(ES=2.63 に調節した時のエッジシャープネスと主観画質の変化)

Image	ES_0	Q_0	k	ES_n	Q_n
A	5.94	1.0	-0.98	2.87	1.9
B	1.16	1.6	4.04	2.57	1.8
C	1.70	2.1	1.18	2.58	2.2
D	1.71	1.7	1.73	2.58	2.6
E	1.13	0.9	3.57	2.57	2.3
F	0.56	0.9	6.88	2.60	2.0
G	1.18	0.5	3.23	2.58	2.2
H	0.93	0.2	6.47	2.54	2.0

図 4-6 に最適鮮鋭化の処理結果画像の例を示す。明らかに異なる鮮鋭さを持った、2 つの入力画像 A と B を本システムに入力したところ、共に最適な鮮鋭さを持つ画像が得られたのが分かる。入力画像 A (図 4-6(a1)) は、暈けた画像のため鮮鋭度が低い。そのため出力では鮮鋭度強調が施されている。一方入力画像 B (図 4-6(b1)) は、もともと鮮鋭度の非常に高い画像である。この場合は、逆に最適値に合わせるために k がマイナスとなり、高周波成分を低減させていることが分かる。



(a1) 入力画像 A(低鮮鋭度画像)



(a2) 画像 A の出力画像 ($k = 9.3$)



(b1) 入力画像 B(高鮮鋭度画像)



(b2) 画像 B の出力画像 ($k = -0.62$)

図 4-6: 画像処理結果

4.2.7. 最適鮮鋭化のまとめ

入力画像の鮮鋭さを評価手法として、エッジシャープネスを定義し、その値を最適な値に変換することによって、最適な鮮鋭化を行う手法を提案した。本手法の特徴は、入力画像の量けに関する事前の知識を必要とせず、自動的に画質改善ができる点にある。

自然画像に対する実験結果は良好であり、入力画像の鮮鋭さに応じた鮮鋭化が行われることが確認された。本手法は、強調制御が効果的であり、実装も容易なため、様々な画像入出力装置への応用が可能である。

4.3. HSV 座標系を用いた彩度・コントラスト・鮮鋭度による自動高画質化

4.3.1. HSV 座標系を用いた自動高画質化の概要

本節では、スキャナ入力された写真や、デジタルカメラ、ビデオ入力からのデジタル画像を対

象とし、劣化に関する事前の知識なしに入力画像に応じて、カラー画像の画質を改善する手法を提案する。本手法は HSV 座標系を用いて、入力画像信号の彩度・コントラスト・鮮鋭度を、ディスプレイの色空間内で最適化するものである。

画像をそのままディスプレイに出力すると、多くの場合画質が低下しているのが分かる。従来、これの劣化特性を明らかにし、それらに応じた画像の復元、補正によって画質改善を図ることが行われてきた。しかしながら、一般に写真のスキナ入力画像などを取り扱う時は、劣化のメカニズムは未知であり、実際には手動で対話的に画質調整を行う必要があった。

本研究では入力画像の劣化過程を考慮せずに、人間の主観的な画質評価要因を基準とした画質改善のアプローチを図る。本手法にて用いる情報は、処理画像である入力 RGB デジタル画像データと、画像の観測状態(出力装置を含む)に関する情報のみである。

本手法の概要を図 4-7 に示す。ここではまず改善すべき画質要因を、色の彩度、明度のコントラスト、鮮鋭度の 3 要因に絞った検討を行う。まず入力画像の部分領域情報から、これらの画質要因に対応した 3 種類の特性値を計算する。次に、画像の観測状態に依存する 3 種類の特性値の最適値を参照して、各要因に対応した画像強調パラメータを計算する。最後に強調パラメータを用いて画像処理を施すことにより、自動高画質化が完了する。

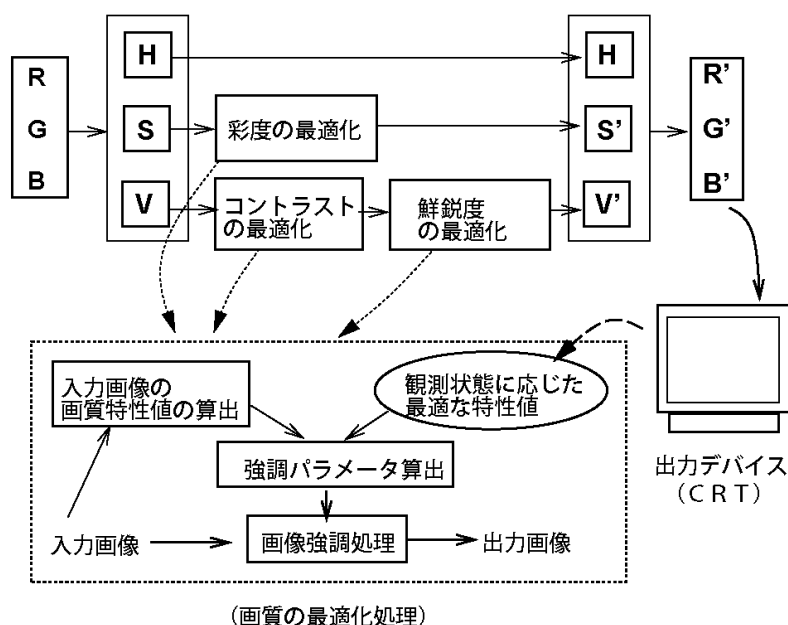


図 4-7: 彩度・コントラスト・鮮鋭度による自動高画質化概要

4.3.2. HSV 座標系

入力画像 f から、彩度、コントラスト、鮮鋭度の 3 つの画質特性値を計算するため、まず f の RGB 値を HSV (色相、彩度、明度) 座標系に変換する。

色相、彩度、明度に分解する座標系はいくつか存在するが、ここでは一般によく使われている Smith の HSV モデル[68]を採用した。図 4-8 に HSV モデルの概要を示す。RGB 値の定義域を $[0, 1]$ で定義すると、H は $[0, 2\pi]$ 、S、V は $[0, 1]$ の値を取る。

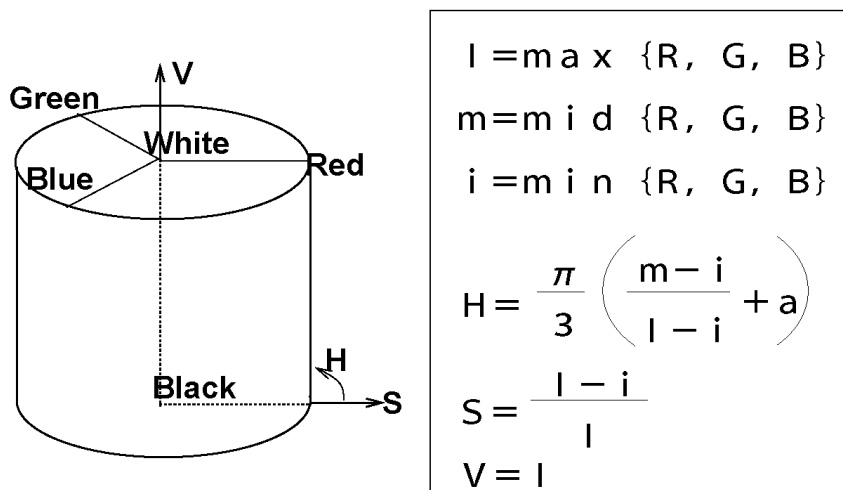


図 4-8: HSV 座標系

4.3.3. 彩度の調整

彩度とは色の鮮やかさを表す。本手法における彩度特性値は、自然画像の色鮮やかさを評価するものある。入力画像における彩度の特性値として、我々は画像内で最も彩度の高い領域のみに注目した。また人間の視覚特性に関して、色の知覚は比較的高周波成分に対する感度が低いことが分かっている[69]。以上より我々は、入力画像 f の彩度特性値 SAF_f を式(4-9)で定義した。

$$SAF_f = \frac{\max(S \otimes G)}{s_H} \quad (4-9)$$

ここで、 \otimes はコンボリューション演算、 G は低周波濾過フィルタ、 S は画像 f の彩度画像を表す。また $\max()$ は画像中の最大値を、 s_H は入力画像の彩度画像の取り得る最大値を表す。彩度強調のための変換は、強調パラメータを c として、式(4-10)の線形変換を用いる。

$$S' = c \times S \quad (4-10)$$

彩度特性値 SAF_f から、 c を計算するためには、 SAF_f をどのレベルに調節すれば良いのかという情報が必要である。これを彩度の最適特性値 SAF_{opt} とすると、この値を持つように画像を変換するための強調パラメータ c_{opt} は、式(4-11)で得られる(高周波成分の影響は無視する)。彩度強調処理の概要を図 4-9(a)に示す。

図 4-9 の c_{opt} は、 $SAF_{opt} = 1.0$ の時の強調パラメータであり、この時は彩度 S のレンジをちょうどいっぱい広げる形の彩度強調処理となる。

$$c_{opt} = \frac{SAF_{opt}}{SAF_f} \quad (4-11)$$

4.3.4. コントラストの調整

コントラストは一般に画像中の明度の高い部分と低い部分の比として定義されている。しかしながらここでは、コントラストは大局的な明度の変化を示すものと考え、入力画像 f のコントラスト特性値 COF_f を式(4-12)で定義した。

$$COF_f = \frac{\max(V \otimes G) - \min(V \otimes G)}{v_H - v_L} \quad (4-12)$$

V は、画像 f の明度 V 画像を表し、 $\max()$ と $\min()$ はフレーム内の最大値と最小値を示し、 v_H と v_L は明度の取り得る最大値と最小値を表す。

コントラスト強調も明度線形変換を用い、変換式は式(4-13)で表される。

$$V' = a \times V + b \quad (4-13)$$

コントラスト特性値 COF_f から強調パラメータ a を計算するためには、コントラスト最適特性値 COF_{opt} が必要である。画像を COF_{opt} を持つように変換する強調パラメータ a_{opt} は、式(4-14)で得られる(高周波成分の影響は無視する)。

$$a_{opt} = \frac{COF_{opt}}{COF_f} \quad (4-14)$$

コントラスト強調処理の概要を図 4-9(b)に示す。図 4-9 の $a0, b0$ は、 $COF_{opt} = 1.0$ の時に明度 V のレンジをちょうどいっばいに広げる場合の強調係数である。係数 a が指定された時に、係数 b の計算が問題となる。我々は図 4-9(b)に示すように、係数 $a0, b0$ の直線において入力明度画像 V の平均値との交点を P とし、明度変換直線が必ず点 P を通るものとして係数 b を計算することにした。

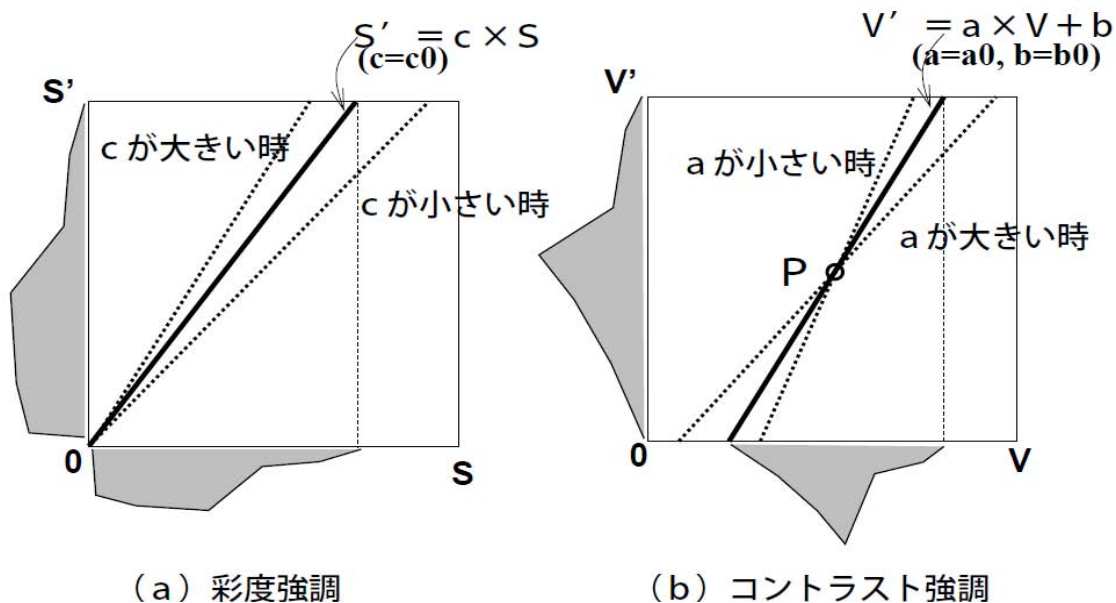


図 4-9: 彩度及びコントラスト強調方法

4.3.5. 鮮鋭度の調整

鮮鋭度は画像中の細かな構造がどれだけ明瞭に表現できるかを示す尺度である。ここでは 4.2 節に述べた鮮鋭度評価値”エッジシャープネス”、を鮮鋭度特性値として採用する。エッジシャープネスは、画像のエッジ領域における高周波成分の平均強度で定義される。

鮮鋭度は画像の高周波成分の特性であるため、視覚特性を考慮すると色情報からの影響はほとんどないと考えられる[69][70]。実際、明度一定下において、色成分にのみ高周波を付加しても鮮鋭さに対する影響はほとんどなく、強く付加すると逆に不自然な画像となってしまふ。

以上より、カラー画像のエッジシャープネスは、明度成分 V のみを用いて式(4-15)で定義する。この中で f は入力画像、 ES_f は f のエッジシャープネスを示す。 $E(V)$ は明度 V 画像のエッジ領域を表し、 $A_{E(V)}$ は $E(V)$ の面積を示す。 ss は各空間周波数の主観的な鮮鋭さに対する重みを表す高周波帯域濾過フィルタである。ここでは、フィルタ ss としてガウシアン差分型フィルタを用いた。式(4-15)の分子は、エッジ領域 $E(V)$ における高周波帯域成分の積分を表す。

$$ES_f = \frac{\iint_{E(V)} |V \otimes ss| dx dy}{A_{E(V)}} \quad (4-15)$$

式(4-15)はエッジ領域に関する量($A_{E(V)}$ と $E(V)$)を含んでいるため、エッジ検出フィルタが必要である。ここでは Sobel のエッジ検出フィルタを用いた。画像のエッジシャープネスを調節するためには、画像の高周波成分を制御しなければならない。基本的な変換式としては、通常鮮鋭化手法である式(4-16)を用いる。強調パラメータは係数 k であり、 $k > 0$ の時は鮮鋭化画像が得られ、 $k < 0$ の時には暈けた画像が得られる。なお、 ss はエッジシャープネスの定義に用いた高周波帯域濾過フィルタと同一である。最適エッジシャープネス値 ES_{opt} に画像を変換す

るための k の値は、式(4-15)と式(4-16)から導出された式(4-17)を用いて近似的に求めることができる。強調係数算出の詳細は 4.2 節を参照のこと。

$$V' = V + k(V \otimes ss) \quad (4-16)$$

$$k \approx k_a = \frac{ES_{opt} \cdot A_{E(V')} - \iint_{E(V')} |V \otimes ss| dx dy}{\iint_{E(V')} |V \otimes ss \otimes ss| dx dy} \quad (4-17)$$

4.3.6. 3 種類の画質改善処理の統合

以上より、3つの画質特性値と、それらに対応する最適特性値から、各強調パラメータを決定することができる。各最適特性値は画像の観測状態や、ユーザの好みに応じて決定する。最適特性値の調整は、従来のように強調量を直接制御するものではなく、人間の知覚する画質要因の特性を制御することになるので、例えばややシャープに、コントラストを強く、といった調整を行うことが可能である。さらに個々の画像に対して適応的に強調パラメータが決定されるので、もともと画質の良いものはそのままの状態でも出力され、画質の悪いものについては、その画質に応じて強調量は自動制御される。

3つの画質改善処理の順序であるが、彩度は他と独立に行うことができるのに対し、コントラストと、鮮鋭度は共に明度 V 成分への処理となるので、互いに依存している。しかしコントラストを変化させると鮮鋭さの主観評価値も大きく変わるのに対し、鮮鋭さを変化させてもコントラストの主観評価値には余り影響がない。従って我々は処理順序として、コントラスト調整を行った後に鮮鋭度調整を行うことにする。

4.3.7. 実験結果

本画質改善手法を、いくつかのカラー写真のスキャナ入力画像に適用した。適用結果はディスプレイ表示による主観評価実験によって評価を行った。評価実験は、暗室と評価用蛍光灯下での2種類(実験 A と実験 B)において行われ、主観評価は、表 4-5 に示す 7 段階の評価尺度によって行われた。この時の観視条件を表 4-6 に示す。

表 4-5: 主観評価実験における画質評価値

Grade	Image Quality
7	Excellent
6	Good
5	Slightly good
4	Fair
3	Slightly poor
2	Poor
1	Bad

表 4-6: 観視条件

Display	N7832-23(NEC) 20inch	
Display size	1280×1024 pixels	
Viewing distance	0.5 m	
White	D65	
Picture size	512×512 pixels	
Gray level	RGB 8 bits×3	
	実験 A	実験 B
環境	暗室	評価用蛍光灯下
画像の種類	5	20
人数	10	2

本実験では総合的な画質改善評価基準として、式(4-18)に示す画質改善率 QI を定義した。 q_0 は入力原画像の画質評価値、 q_{manual} は手動で最適に調節した時の画質評価値であり、 q は自動で画質改善処理した出力画像の画質評価値である。 $QI = 1.0$ の時は、手動による最適な強調処理と同等の画質が得られたことになる。

$$QI = \frac{q - q_0}{q_{manual} - q_0} \quad (4-18)$$

表 4-7 は、実験 A における、画像に応じた適応的な強調パラメータ変化と、それに伴う主観画質評価値の変化を表す。表 4-7 より、画像に応じて強調パラメータが適応的に変化する様子が分かる。

表 4-8 は、各実験において設定した最適パラメータと、得られた平均画質改善率を示す。実験

結果より、平均画質改善率 QI は約 0.7 であった。すなわち、本実験画像は、手動による画質改善と比べて約 70% 程度の品質で出力されたことが分かった。

図 4-10 と図 4-11 に低品質な入力画像に対する処理例を示す。図 4-10 は建物画像の例であり、図 4-11 は人物画像の例である。これらの入力画像に対する出力画像を見ると、適切な強調処理が施されていることが分かる。一方、図 4-12 はもともと高品質な入力画像に対する処理例である。高品質な入力画像に対する出力画像を見ると、ほとんど強調処理が施されず、品質の良い状態のままで出力されている。

4.3.8. HSV 座標系を用いた自動高画質化まとめ

本節では、HSV 座標系を利用して、入力画像に応じて 3 つの画質要因を自動的に改善する手法を提案した。本手法の特徴は、部分領域情報を用いて、人間の主観的な 3 つの画質要因(彩度、コントラスト、鮮鋭度)に関する特性値を算出し、それらを元に強調パラメータを制御する点である。これにより、入力画像の劣化に関する事前の知識を必要とせず、画像データ自身から最適な強調制御を実現する。自然画像に対する強調処理実験の結果は良好であり、入力画像の品質に応じて、強調パラメータが適応的に制御されることを確認した。また主観評価実験によって、画質が改善されることを確認した。実験によって得られた画質改善率より、手動による画質改善と比べてその性能比は、約 70% 程度になることが分かった。本手法は、様々なデジタル画像機器への応用が可能である。

表 4-7: 強調パラメータの変化と主観画質評価値

Image	c	a	k	q_0	q_{manual}	q	QI
No.1	1.26	1.06	7.07	2.7	5.8	5.1	0.77
No.2	1.23	1.12	6.09	1.9	4.9	4.1	0.73
No.3	1.05	1.14	1.48	4.2	6.5	6.3	0.91
No.4	1.00	1.00	3.48	3.7	6.2	4.9	0.48
No.5	1.14	1.11	4.05	3.3	6.2	5.7	0.83

表 4-8: 平均画質改善率の結果

実験	SAF_{opt}	COF_{opt}	ES_{opt}	QI
A	1.0	1.0	5.6	0.71
B	1.0	1.0	5.6	0.66



(a) 入力画像(建物)



(b) 出力画像(建物)

図 4-10: 低品質画像(建物)に対する処理例



(a) 入力画像(人物)



(b) 出力画像(人物)

図 4-11: 低品質画像(人物)に対する処理例



(a) 入力画像



(b) 出力画像

図 4-12: 高品質画像に対する処理例

4.4. 任意色に対する特色分離補正

4.4.1. 選択的色補正の関連研究

カラー自然画像の色補正においては、一般的にトーンカーブによる補正や、マスキングマトリクスによる補正[71]が行われる。しかしながら、画像中のある特定の色相の画素のみを選択的に補正したい場合には、トーンカーブ補正等の手法では正しく補正することができなかった。このような一部の色領域に対する色補正を、選択的色補正と呼ぶ。選択的色補正の従来手法として、特色分離色補正[72]がある。これは、ユーザが R,G,B,C,M,Y の 6 つの色相から補正したい色相を選択し、指定した色相の画素のみを補正する機能であり、印刷原稿の色分解工程等では必須の機能として使用されている。本節ではこの従来手法を、6 色相別色補正、と呼ぶことにする。6 色相別色補正は、指定可能な色相が R,G,B,C,M,Y の 6 つに固定されており、中間の色相や彩度の低い色に対する補正はできないという問題があった。

その一方で従来、 $L^*a^*b^*$ 座標系を基準とする円柱座標系を用いた、選択的色調整機能が提案されている[73]。しかしながら、 $L^*a^*b^*$ 座標系での色補正処理は、6 色相別色補正に比べて多くの計算コストを要するため、通常の PC で対話処理を行うといった用途への適用は困難であった。また、 $L^*a^*b^*$ 座標系で処理されるため、6 色相別色補正との互換性がなく、6 色相別色補正に慣れたユーザにとっては使いづらいものであった。

そこで本節では、6 色相別色補正を包含する、任意色に対する特色分離補正を提案する。本手法には、HSV 座標系[68]を用いた方式(HSV 方式)と RGB の差分を用いる方式(RGB 差分方式)の 2 種類がある。HSV 方式では、任意の色相範囲を設定することが可能であり、RGB 差分方式では、彩度の低い中間色に重点をおいた補正が可能である。

本節では、指定色に対する特色度分布を用いてそれぞれの提案方式の特性を明らかにし、本手法が 6 色相別色補正の完全な拡張であり、指定色として先の RGBCMY を選択した時は、従来と同じ処理結果が得られることを示す。さらに提案手法を用いることで、6 色相別色補正では実

現できなかった、中間色相や中間彩度の色補正が実現できることを示す。

4.4.2. 6色相別色補正

図 4-13 に、RGB 色空間を white と black を結ぶ軸に垂直な平面に投影した図を示す。この投影図は 6 角形で表され、6 つの頂点はそれぞれ RGBCMY である。そしてその頂点を基準に、RGB 空間が色相別に(I)から(VI)までの領域に分割されることが分かる。

6色相別色補正[72]では、まず補正したい色相を、6角形の頂点 R,G,B,C,M,Y の中から選択する。次に指定した色相の特色度を画素毎に抽出する。この特色度は、表 4-9 に従って計算することができる。図 4-13 には、R 色相の特色度 dR の分布も示されている(RGB 画素値は 0 から 1 に正規化されているものとする)。これより、R 色相の特色度を持つ画素は、R の頂点を中心に色領域(I)と(VI)に分布することが分かる。

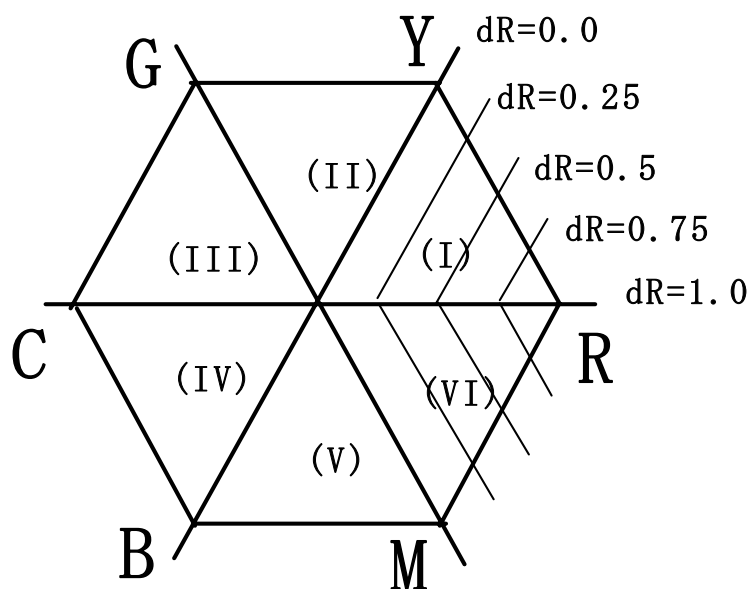


図 4-13:RGB 空間の投影図と R 色相の特色度

表 4-9: 6 色相別色補正の特色度の算出

Area	Condition	dR	dG	dB	dC	dM	dY
(I)	$R > G > B$	$R - G$	0	0	0	0	$G - B$
(II)	$G > R > B$	0	$G - R$	0	0	0	$R - B$
(III)	$G > B > R$	0	$G - B$	0	$B - R$	0	0
(IV)	$B > G > R$	0	0	$B - G$	$G - R$	0	0
(V)	$B > R > G$	0	0	$B - R$	0	$R - G$	0
(VI)	$R > B > G$	$R - B$	0	0	0	$B - G$	0

得られた特色度に対し、式(4-19)を用いて各画素の色を補正する。(a1, a2, a3)は任意の RGB 補正係数である(式(4-19)は R 色相への補正例)。

$$(R', G', B') = (R, G, B) + dR \times (a1, a2, a3) \quad (4-19)$$

すなわち R 色相への補正は、領域(I)と(VI)の画素値にのみ施されることになる。また画素値が R の頂点から離れるにつれて補正量が低減する仕組みになっており、領域(I)と(VI)以外の画素と、連続性を保ったまま補正される。従って、マスクを作成し、マスク領域内の色のみを変更する色補正と比べて、1)対象の色が 6 色相のいずれかであれば、マスクを切りぬく手間がいらぬ、2)指定色相から色空間上で距離が離れるほど変化量が減少するので、領域境界における偽輪郭発生の問題がない、というメリットがある。しかしながら従来の色相別色補正は、補正対象として 6 つの色相しか指定できないという問題があった。

そこで本研究は、任意の色相を指定可能となるよう特色度の概念を拡張する。すなわち特色度とは、補正対象色 X と注目する画素値との色空間における距離であると考え。特色度の概念を拡張することにより、任意色に対する特色分離補正を提案する。

4.4.3. 任意色に対する特色分離補正・HSV 方式

4.4.3.1. 特色度定義と補正方法

図 4-13 の RGB 空間の投影図が、HSV 色座標系[68]の HS 平面と等価であることに着目し、HSV 座標系を用いて新たな特色度 κ を定義する。補正対象となる色 X の HSV 値を (Hue, Sat, Val) とし、色相範囲を m とする。対象色 X として図 4-13 の外周部の色を指定することができるものとする(すなわち $Sat = 1, Val = 1$)、任意の画素の HSV 値を (h, s, v) とした時、その画素の特色度 κ は、式(4-20) で表すことができる。また色補正式は、式(4-21)で表される。

図 4-14 に、HSV 方式による特色度分布の例を示す。

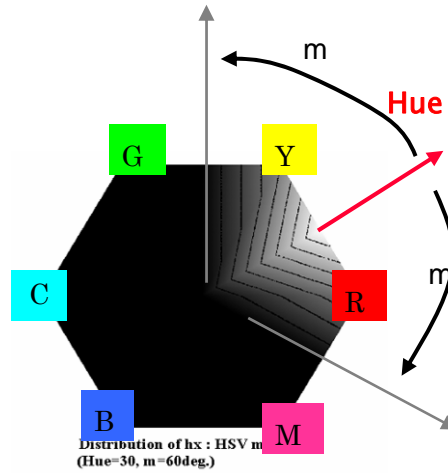


図 4-14: HSV 方式の特色度分布例

$$\kappa = \frac{\phi(m - |Hue - h|)}{m} \times s \times v \quad (4-20)$$

$$\phi(x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } x \geq 0 \end{cases}$$

$$(R', G', B') = (R, G, B) + \kappa \times (a1, a2, a3) \quad (4-21)$$

HSV 方式の特徴は、色相の中心と補正範囲を任意に設定することができるという点である。これにより、6 色相の中間色相を指定したり、非常に狭い範囲の色相を指定したりすることができる。

4.4.3.2. 6 色相別色補正と HSV 方式との互換性

補正対象の色 X が、R,G,B,C,M,Y の頂点であり、なおかつ色相範囲 $m=60$ 度の時、HSV 方式による特色度 κ は、6 色相別色補正における特色度と等価であることを証明する。

式(4-20)を用い、色相 Hue を R 色相(Hue=0)、 $m=60$ 度の場合の特色度 κ_r を計算する。特色度 κ_r が 0.0 より大きいのは、色相 h が $0 \sim 60$ 度の間と、 300 度から 0 度の間にある画素である。今、ある画素の色相 h が 0 から 60 度の間にあるとすると、この画素の RGB 値の中での最大が R、最小は B、中間は G である。従ってこの画素の HSV 値 (h, s, v) は、HSV の定義から式(4-22)のように算出することができる。

$$\begin{aligned}
s &= \frac{R-B}{R} \\
v &= R \\
h &= \frac{G-B}{R-B} \times 60(\text{deg})
\end{aligned}
\tag{4-22}$$

この時、特色度 κ_r は、次のように算出される。

$$\kappa_r = \frac{\phi(60-|0-h|)}{60} \times s \times v = \left(1 - \frac{G-B}{R-B}\right) \times \frac{R-B}{R} \times R = R-G
\tag{4-23}$$

本結果は、表 4-9 の色領域(I)における特色度 dR の値と一致することが分かる。同様に、他の 5 つの色相についても、色相範囲 m が 60 度であれば、特色度の値は一致する。以上より、本 HSV 方式は、6 色相別色補正の拡張方式となっていることが分かる。

4.4.4. 任意色に対する特色分離補正・RGB 差分方式

4.4.4.1. 特色度定義と補正方式

RGB 値の差分を元にして特色度を定義する。本方式による特色度 μ は、指定色 X を (r_0, g_0, b_0) 、特色度を算出する画素の RGB 値を (r, g, b) 、特色範囲の重みづけを W とする時、以下のステップによって算出される。

(1) RGB 差分 (dr, dg, db) の算出

$$\begin{aligned}
(r_0', g_0', b_0') &= (r_0, g_0, b_0) - \min(r_0, g_0, b_0) \\
(r', g', b') &= (r, g, b) - \min(r, g, b) \\
(dr, dg, db) &= (r_0' - r', g_0' - g', b_0' - b')
\end{aligned}
\tag{4-24}$$

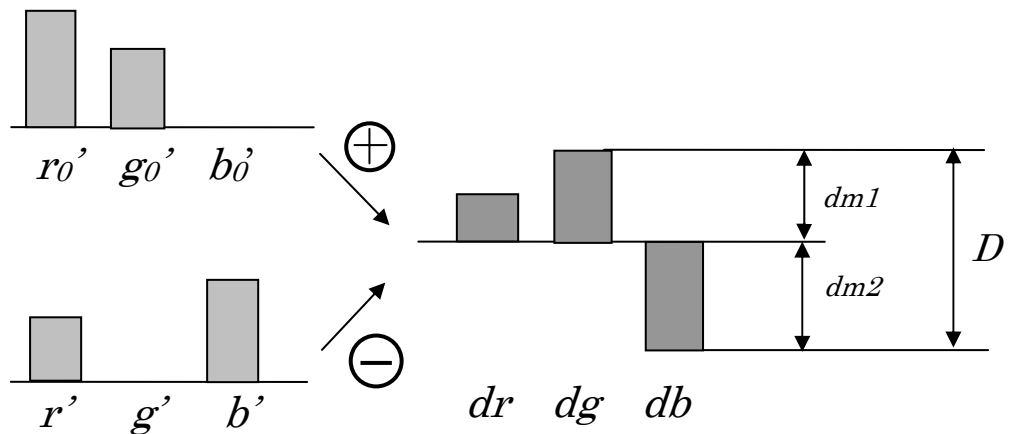


図 4-15: 色間距離の算出

(2) 色間距離 D の算出

次に (dr, dg, db) 中の正值の絶対値の最大 $dm1$ と、負値の絶対値の最大 $dm2$ とを加算して、色間距離 D を求める。 D が 0 の時 2 色は一致し、大きくなるにつれて 2 色は異なる色となる。色間距離の概念を、図 4-15 に示す。

(3) 特色度 μ の算出

特色度 μ は、2 色が一致する時には 1.0、十分離れている時には 0.0 となるよう定義する。但し μ が負の時は $\mu = 0.0$ とする。重み W は特色度が及ぶ範囲を制御する。

$$D = dm1 + dm2 \tag{4-25}$$

$$\mu = 1.0 - W \times D$$

色補正式は次式で表される。

$$(R', G', B') = (R, G, B) + \mu \times (a1, a2, a3) \tag{4-26}$$

RGB 差分方式の特徴は、補正対象色を RGB 値で指定可能な点である。重み W を調整することにより彩度の低い色、あるいは彩度の高い色のみを補正可能である。また HSV 方式に比べると、HSV 座標系への色変換コストが不要なため、高速処理が可能である。

4.4.4.2. 6 色相別色補正と RGB 差分方式との互換性

補正対象の色 X が、R,G,B,C,M,Y の頂点の色であって、重み係数 $W = 1$ の時、特色度 μ は従来の色相別色補正における特色度と等価であること示す。補正対象の色 X を (r_0, g_0, b_0) とし、その値を $r_0 = 1.0$ (最大), $g_0 = b_0 = 1.0$ であるとする。特色度を求める画素の RGB 値が、 $R > G > B$ である時、この画素の特色度 μ_r は次のように計算することができる。

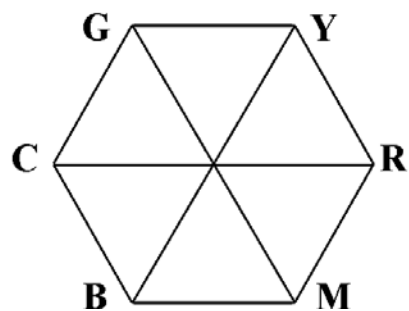
$$\begin{aligned}
 (r', g', b') &= (R - B, G - B, 0.0) \\
 D &= 1.0 - R + G \\
 \mu_r &= R - G
 \end{aligned}
 \tag{4-27}$$

すなわち、RGB 差分方式による特色度 μ_r は、表 4-9 における色領域(I)における特色度 dR の値と一致することが分かる。他の 5 色相についても同様の結果が得られる。本方式は、RGBCMY の 6 色が補正対象色であって、重みが $W=1$ の時には、HSV 方式と同様、6 色相別色補正と等価な色補正となる。

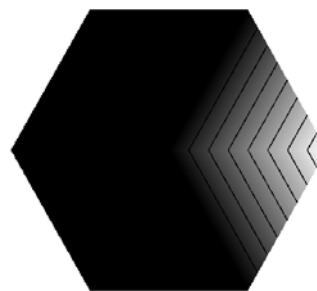
4.4.5. 特色度の分布

提案手法、及び 6 色相別色補正における特色度分布を図 4-16 に示す。本図は RGB 空間を白色方向から見た投影面に対する特色度分布を示しており、特色度 0 から 1 までを 16 等分し、1/16 ごとに表示画素値を 0 とすることで、特色度の等値線を描画している。図 4-16 の(a)は RGBCMY の 6 色相の位置を示し、(b)は従来の 6 色相別色補正による R 色相指定時の特色度分布を示す。また(b),(c),(d)は、HSV 方式において条件を変えた場合の特色度分布を示し、(e),(f),(g)は、RGB 差分方式における特色度分布を示す。図 4-16(b),(c)の結果より、あるパラメータにおいて、HSV 方式が 6 色相別色補正と等価となることが分かる。また(b),(e)の結果より、RGB 差分方式も同様に 6 色相別色補正と等価となる。

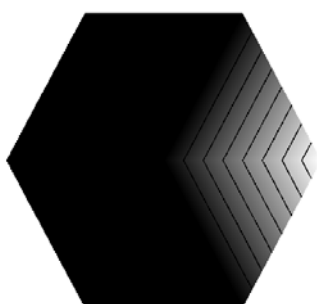
また(c),(d),(f),(g)の分布は、従来の 6 色相別色補正では指定できなかった特色度である。HSV 方式の場合は 6 色相の中間色相に対しても、補正指定が可能であることが分かる。また、RGB 差分方式において重み W を調整することで、一部の狭い色空間を指定できることを確認した。



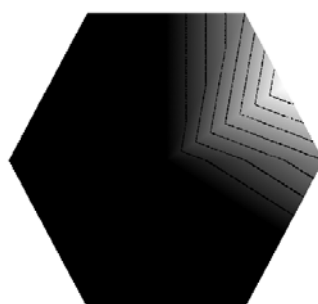
(a) 6色相(RGBCMY)の位置



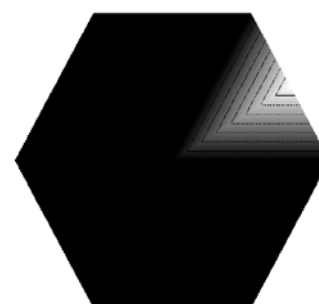
(b) 6色相別色補正の R 指定時の特色度分布



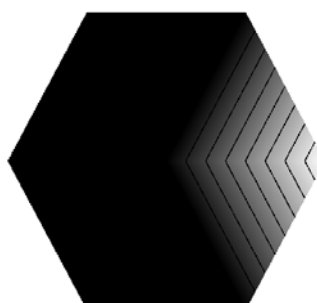
(c) HSV 方式の特色度分布
(Hue=0.0, m=60deg.)



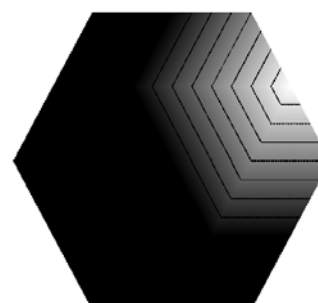
(d) HSV 方式の特色度分布
(Hue=30, m=60deg.)



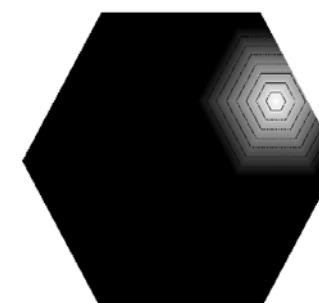
(e) HSV 方式の特色度分布
(Hue=30, m=30deg.)



(f) RGB 方式の特色度分布
(r_0, g_0, b_0)=(1,0,0), W=1
m=60deg.)



(g) RGB 方式の特色度分布
(r_0, g_0, b_0)=(1,0.5,0), W=1
m=60deg.)



(h) RGB 方式の特色度分布
(r_0, g_0, b_0)=(0.8,0.4,0), W=2

図 4-16: 特色度分布例

4.4.6. 色補正結果

図 4-17 に HSV 方式による処理結果を示す。共に(a)が原画像、(b)が従来の 6 色相別色補正に

よる結果、(c)が本手法による結果である。(b)は、6色相別色補正を用い、黄色を指定してゴムボートの色を赤く変更した例である。この場合は背景の川の色も赤く変色していることが分かる。これは川の色が黄色相に含まれるために、6色相別色補正の影響を受けて生じる結果である。一方、HSV方式を用いた例(c)では、ボートの色を直接補正対象色相に指定し、さらに色相範囲 m を 30 度と狭くすることにより、背景に影響を与えることなくゴムボートの色のみが変更されている。

図 4-18 に RGB 差分方式による処理結果を示す。(b)は、6色相別色補正を用いて、赤色の帽子の色を緑に変更した例であるが、赤色相に含まれる肌色が、補正の影響を受けて変色してしまうことが分かる。一方、RGB 差分方式を用いた例(c)では、帽子の色である彩度の高い赤色を直接補正対象色に指定し、さらに重み W を付加することにより、肌色に影響を及ぼすことなく帽子の色を変更可能であることを示している。

4.4.7. 色補正インターフェース

本手法の重要なメリットは、単に任意の色相を補正可能という点だけでなく、ユーザの色補正インターフェースを大幅に改善するという点にある。例えば、表示画像中の補正したい領域をマウスで指定して補正対象の色 A を選択し、別メニューで補正後の色 Z をパレットから指定するという操作で所望の特色補正が実現できる。色相範囲や重みパラメータは選択可能にしておき、強調係数(a_1, a_2, a_3)は色 A と色 Z の差分を元に決定すればよい。図 4-19 に色補正インターフェースの例を示す。(a)は元画像中の黄緑色のヘルメットを指定し、それを青色に変更することを示す。(b1)が HSV 方式による補正結果であり、(b2)が RGB 差分方式による補正結果である。

本節の色補正実験結果は、実際にこの指定方法によって得られたものである。従来の 6色相別色補正では、補正対象の色相が限定されていたため、強調係数はオペレータの経験に基づいて与えなければならなかった。つまり上記のような、簡潔なインターフェースによる色補正は、本方式を用いて任意色相にアルゴリズムを拡張することによって実現可能となる。



(a) 原画像



(b) 6色相別色補正(Y指定)



(c) 提案手法(HSV方式)

図 4-17: HSV方式による処理結果



(a) 原画像

(b) 6色相別色補正(R指定)

(c) 提案手法(RGB差分方式)

図 4-18: RGB 差分方式による処理結果

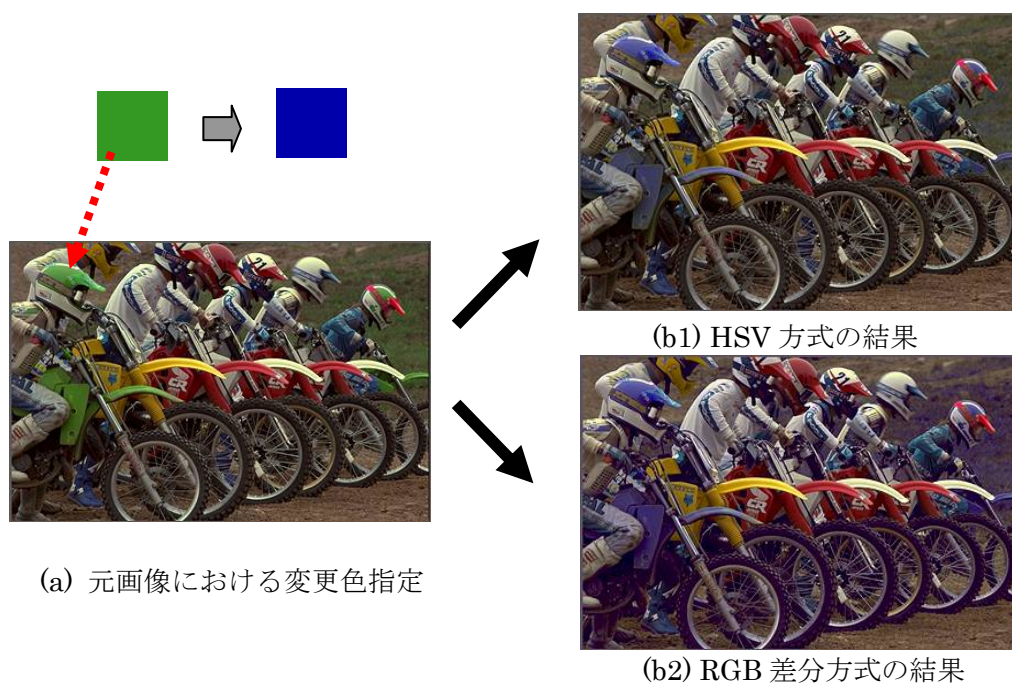


図 4-19: 色補正インターフェースの例

4.4.8. 任意色に対する特色分離補正のまとめ

本節では、特色度の概念を再定義することにより、任意の色成分に対して選択的に補正可能な、2つの色補正方式(HSV方式、RGB差分方式)を考案した。さらにこれらの方式が、従来の6色相別色補正方式を包含する形で構成されていることを示した。

本手法は、マスクを用いた色補正と比べても、以下の利点がある。

- 1) 対象領域の色相が同じであればマスクを切りぬく必要がない
- 2) 部分領域の色補正時の領域境界における偽輪郭が起こりにくい

本手法の適用は幅広く、PC上での簡単なフォトタッチ機能として有効であるだけでなく、非線形な色補正が必要な印刷向けの微妙な色修正にも使用することができる。

本手法は、色の部分空間情報を利用した色補正手法である。本手法を用いて色が均一な特定の対象物の強調処理が可能となる。様々な撮影画像においては、特定の物体色が照明環境の変動によって影響を受けたり、人間の記憶色の影響を受ける場合がある。そのような時、本手法を用いることにより、特定の色領域にのみ適切な補正を施すことができる。

4.5. 6つの画質特性解析に基づく自動高画質化

4.5.1. 自然画像の自動強調における課題

様々なシーンで撮影された一般の自然画像に対し、画像強調技術によって画像品質を適応的に補正する手法について報告する。近年デジタルカメラや、フラットパネルディスプレイを用いた薄型TVの普及により、消費者の購買意欲を刺激するための画質の追求が盛んになっている。それに伴い画像処理の研究分野においても、強調処理による主観画質の向上が重要なテーマとなっている。一般的なデジタルカメラの撮影画像は、特定の劣悪な環境下で品質が劣化するため、適切な強調処理が必要である。通常のデジタルカメラでは、vonKriesのモデルに基づく参照白色を利用した自動カラーバランス補正[59][60]が用いられている。さらに後処理として、固定係数のマトリクス補正やトーンカーブ補正などにより、コントラストや色の補正を行っている。しかしながら、カメラで取得された画像(静止画・動画)は、撮影環境によって画質が大きく異なるため、一律な固定パラメータを用いて全ての画像を適切に補正することは困難であった。

4.3節において、彩度、コントラスト、鮮鋭度による自動高画質化技術を提案した。しかしながら、一般的なカメラで撮影された画像において、これら3つの処理だけでは画質補正できない場合がある。例えば、画像全体の露光不足や露光過多、ホワイトバランスの失敗による色かぶり、あるいは人肌など記憶色再現の不良である。

そこで本節では、彩度、コントラスト、鮮鋭度に加えて、ホワイトバランス、明度、記憶色を含む、6つの画質要因に対応した、新たな自動高画質化手法を提案する。本手法はデジタルカメラで撮影され、RGB値またはYUV値に変換されたデジタルカラー画像を入力対象とし、自動的に最適な強調を施すことによって画質を改善することを目的としている。本手法は6つの各画質要因に対して、画質特性値(QFV)を用いた強調パラメータ制御を行う。画質特性値(QFV)は、4.3節における特性値を再定義したものであり、鮮鋭度特性値等は重複している。入力画像から算出するQFVを基に、適切な強調パラメータを推定することで、これら6つの画質要因に応じて適応的に画像補正することが可能となる。提案する強調手法の特徴は以下の通りである。

- 1) 全体領域への補正(色、階調、鮮鋭度)と、部分領域補正(記憶色)とを有する
- 2) 6つの画質特性値を用いた強調制御
- 3) リアルタイム処理可能な低演算量

本節では、QFVを用いた適応的強調のフレームワークについて述べた後、それぞれのQFVの算出と、強調パラメータ推定方法を述べる。その後、高速アーキテクチャ導入による、PC上で

のリアルタイム処理の実現方法について述べる。最後に本手法を様々なシーンで撮影された自然画像に適用し、主観評価実験によってその有効性を示す。

4.5.2. 画質特性解析と補正方法

本研究の目的は、画像品質におけるいくつかの重要な要因について最適に補正することである。このために、それぞれの画質要因を数値化する新たな測定法を導入する。この測定値を画質特性値(QFV)と呼ぶ。本節では、QFVを用いた自動高画質化手法を提案する。図 4-20 に本手法の概要を示す。

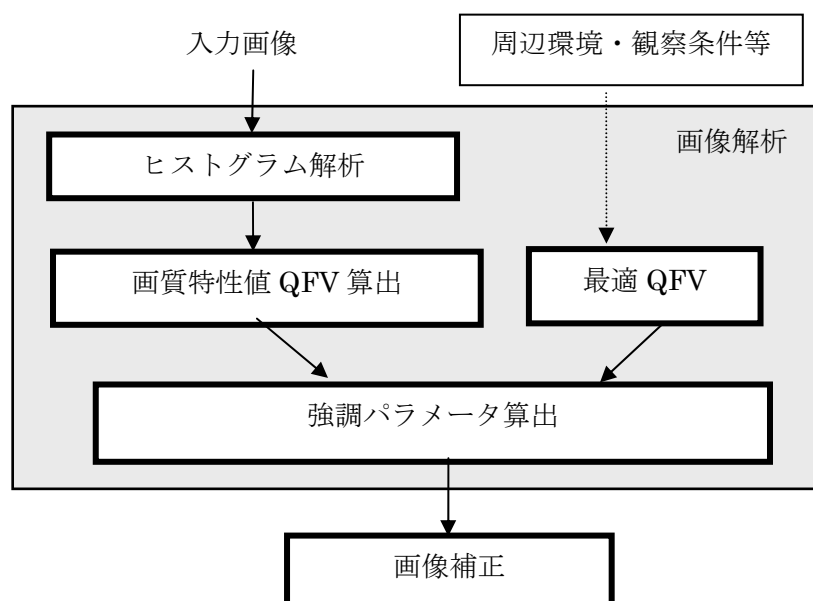


図 4-20: 自動高画質化処理の概要

本処理の流れは次の通りである。

- 1) 入力画像から 6 つの画質要因に対応する画質特性値(QFV)を算出する。
- 2) 入力画像の QFV を、あらかじめ求めておいた最適 QFV と比較する。
- 3) 入力画像の QFV が最適値となるような、強調パラメータを算出する。
- 4) 算出された強調パラメータを用いて画像補正を行う。

表 4-10 に、6 つの画像補正と、それぞれに対応する QFV、及び強調パラメータの一覧を示す。表 4-10 の”Type”は、画像全体に影響が及ぶ補正(Global)か、局所的な補正(Local)かを示している。以下、それぞれの画質特性値(QFV)の算出方法と、画像強調方法について述べる。

表 4-10: 画像補正と QFV

補正処理	Quality Feature Values (QFV)	Enhancement Parameters	Type
ホワイトバランス	$\mathbf{T}_W = (t_{wr}, t_{wg}, t_{wb})$	(w_1, w_2, w_3)	Global
コントラスト	$\mathbf{T}_C = (HD, SD)$	(hl, sd)	Global
彩度	$\mathbf{T}_{SA} = S_{\max}$	c	Global
明度	$\mathbf{T}_L = I_m$	γ	Global
記憶色(3種類)	$\mathbf{T}_P = (H_{skin}, H_{grn}, H_{sky})$	(a_r, a_g, a_b)	Local
鮮鋭度	$\mathbf{T}_S = ES$	k	Global

4.5.3. ホワイトバランス補正

ホワイトバランス補正は、画像の白色点を移動させて色かぶりを補正し、画像全体の色バランスを適正化する処理である。ホワイトバランス特性値 \mathbf{T}_W は、入力画像中で人間が白と感じる、最も明るい部分領域の RGB 平均値である。 \mathbf{T}_W は入力画像の色かぶり度合いを表し、式(4-28)によって定義される。

$$\mathbf{T}_W = (t_{wr}, t_{wg}, t_{wb}) = E_{RGB}(\max(I(x, y) \otimes LPF)) \quad (4-28)$$

$I(x, y)$ は、画像空間座標 (x, y) における明度の分布を示す。 LPF は低周波濾過フィルタであり、 $\max()$ は画像中の最大値を示す。また $E_{RGB}(x)$ は、明度 x 以上を持つ画素領域の平均 RGB 値を表す。 \mathbf{T}_W の定義では微小領域の影響を排除するため、あらかじめ低周波濾過フィルタを作用させている。その理由は、人間が白と知覚できる領域は、ある程度の大きさを有する必要があるからである。低周波濾過フィルタのサイズは画像の解像度と観察環境に依存する。

なお、低周波濾過フィルタを作用させる代わりに、画像中で一定の面積率以上の明るい領域を評価する方法もある。この場合、ホワイトバランス特性値 \mathbf{T}_W は式(4-29)で定義する。

$h(I(x, y), \alpha)$ は、明度画像 $I(x, y)$ において、明度ヒストグラムの明るい側から α %の明度値を表す。式(4-29)の算出方法は、空間フィルタリングが不要であるため、演算量が少ないという利点がある。なお、明度画像 $I(x, y)$ としては、XYZ 座標系の Y や、HSV 座標系の V を用いることができる。

$$\mathbf{T}_W = (t_{wr}, t_{wg}, t_{wb}) = E_{RGB}(h(I(x, y), \alpha)) \quad (4-29)$$

ホワイトバランス補正は、式(4-30)に示される補正式によって行われる。強調パラメータ (w_1, w_2, w_3) は、ターゲットの白色を表す最適ホワイトバランス特性値 $(t0_{wr}, t0_{wg}, t0_{wb})$ を用いて、式(4-31)によって算出される。

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \mathbf{W} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} w_1 + 1.0 & 0 & 0 \\ 0 & w_2 + 1.0 & 0 \\ 0 & 0 & w_3 + 1.0 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (4-30)$$

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} t0_{wr} / t_{wr} - 1.0 \\ t0_{wg} / t_{wg} - 1.0 \\ t0_{wb} / t_{wb} - 1.0 \end{pmatrix} \quad (4-31)$$

4.5.4. コントラスト補正

画質要因としてのコントラストは、画像の大域的な明度変化を表すものとする。ここでは人間の視覚システムは、画像中の最も暗い領域と、最も明るい領域の成分によって、コントラストを判断しているとする。入力画像のコントラストを評価するため、ハイライト特徴 HL と、シャドウ特徴 SD を定義する。コントラスト特性値 \mathbf{T}_c は、この HL と SD とから構成され、式(4-32)で定義される。

$$\begin{aligned} \mathbf{T}_c &= (HL, SD) \\ HL &= \max(I(x, y) \otimes LPF) \\ SD &= \min(I(x, y) \otimes LPF) \end{aligned} \quad (4-32)$$

$I(x, y)$ は、画像空間座標 (x, y) における明度の分布を示す。 LPF は低周波濾過フィルタであり、 $\max()$ は画像中の最大値を示し、 $\min()$ は画像中の最小値を示す。コントラスト強調パラメータは、式(4-33)の hl と sd である。

$$\begin{aligned} hl &= HL_o - HL \\ sd &= SD - SD_o \end{aligned} \quad (4-33)$$

ここで、 (HL_o, SD_o) は最適コントラスト特性値である。コントラスト強調処理は明度ヒストグラムを (HL_o, SD_o) となるように拡大することで行われる。入力画素値を Y とすると、強調後の画素値 Y' は、式(4-34)によって求められる。

$$\begin{aligned} Y' &= a_1 \times Y + a_2 \\ a_1 &= \frac{HL_o - SD_o}{(HL_o - hl) - (SD_o + sd)} \\ a_2 &= HL_o - a_1 \times HL \end{aligned} \quad (4-34)$$

コントラスト強調処理は、様々な色空間を用いて実現可能である。使用する色空間の例として、

RGB空間のRGB各成分、HSV空間のV成分、YUV空間のY成分などがある。しかしながら、用いる色空間によって強調後の彩度成分に違いが生じる。従って、彩度補正処理との併用により、最終画質を調整することが重要である。なお本節の実験では、RGB成分を独立に強調処理を行うものとする。

4.5.5. 明るさの補正

画像全体の明るさ補正は、可変のトーンカーブ補正によって実行される。すなわち、明るさ特性値 I_m に応じてトーンカーブが変化する。明るさ特性値 T_L は、式(4-35)に示すように入力画像の明るさ最大値と最小値の中間値として定義される。

$$T_L = I_m = \frac{(HL + SD)}{2} \quad (4-35)$$

明るさ補正処理は、式(4-36)のいわゆるガンマ補正式によって算出される、トーンカーブを用いて行われる。

$$Y' = \frac{m}{m^\gamma} Y^\gamma \quad (4-36)$$

式(4-36)の m は、画素値 Y の取りうる最大値であり、 Y が 8bit データの場合は 255 である。明るさ特性値 I_m の画像を、最適明るさ特性値 I_{m0} になるように変換するための強調パラメータ γ は、式(4-37)で与えられる。すなわち明るさ補正は、画像中の輝度の中間点を画質特性値とし、この値が最適値となるようなパラメータ γ を用いてガンマ補正を行う。

$$\gamma = \frac{\log(m/I_{m0})}{\log(m/I_m)} \quad (4-37)$$

4.5.6. 彩度補正

画像の彩度は、カラー画像における色鮮やかさを表す指標である。著者らは、人間の視覚システムは画像中の最も鮮やかな部分領域から画像全体の彩度を判断しているものと考えた。そこで、彩度特性値 T_{SA} は式(4-38)によって定義する。

$$T_{SA} = S_{\max} = \max(S(x, y) \otimes LPF) \quad (4-38)$$

$S(x, y)$ は、入力画像中の彩度分布を示す。ここでの彩度とは、HSV座標系における S 値であるものとする。すなわち画素値が (r, g, b) である画素の彩度 S は式で与えられる。

$$S = \frac{\min(r, g, b)}{\max(r, g, b)} \quad (4-39)$$

LPF は低域濾過フィルタであり、人間の視覚システムが高周波の色成分に対して鈍感であるという知見[66]を反映している。強調パラメータ c は次の式(4-40)で与えられる。

$$c = S_o - S_{\max} \quad (4-40)$$

ここで S_o は最適彩度特性値である。彩度補正処理は式(4-41)によって行われる。

$$S' = S \times \frac{S_o}{S_o - c} \quad (4-41)$$

彩度強調処理には、 $CIEL^*a^*b^*$ や YUV などの HSV 以外の色空間が用いられる可能性がある。そこで同じ強調パラメータ c を用いた YUV 色空間での彩度補正処理を式(4-42)に示す。式(4-42)は、 Y 成分はそのままで、 UV 成分にのみ同じ係数で線形強調が施されることを示している。

$$\begin{pmatrix} Y' \\ U' \\ V' \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & \frac{S_o}{S_o - c} & 0.0 \\ 0.0 & 0.0 & \frac{S_o}{S_o - c} \end{pmatrix} \begin{pmatrix} Y \\ U \\ V \end{pmatrix} \quad (4-42)$$

4.5.7. 記憶色の補正

記憶色の補正は、特定の自然物の色を人間の脳に記憶された色に近づける色補正である。記憶色の補正は、特定の物体にのみ補正効果を作用させる必要があるため、部分領域に対する色補正方式が必要である。塚田らは、4.4 節に述べた任意色の特色分離補正を利用して、肌色、青空の色、草木の緑色の 3 種類の重要な記憶色を、適応的に補正する手法を提案した[74]。記憶色の補正は、 $Hunt$ による色再現の分類[62]における、好ましい色再現の実現を目的とする機能であるため、単独で好ましい色再現(**Preferred Color Reproduction**)、と呼ぶこともある。本研究では、塚田らの適応化アルゴリズムを記憶色補正に採用する。

記憶色の画質特性値 \mathbf{T}_p は、3 種類のカテゴリ(肌色、草木の緑色、青空)に対応した、それぞれの代表色である。記憶色補正は、以下のステップで実行される。

- 1) 入力画像から 3 種類のカテゴリに対する代表色を抽出する。
- 2) 代表色に対応したパラメータセットをルックアップテーブルから取得する。
- 3) 取得したパラメータを使って任意色相の特色補正を行う。

代表色は、入力画像中の対象物体領域において最も支配的な色を表す。本手法では、代表色は対象物体領域のカラーヒストグラムから抽出する。ここでは色相ヒストグラムを用いる方法を説明する。対象となる記憶色カテゴリに対して、色相を i 個の領域に分割する。そして、入力画像の対象物体領域から色相ヒストグラム $hist_H(i)$ を生成する。そして代表色インデックス i_{RH} を次式によって求める。

$$i_{RH} = \arg \max_i \{ hist_H(i) \times F_H(i) \} \quad (4-43)$$

ここで、 $F_H(i)$ はあらかじめ求めておいた対象物体の色分布確率である。求める代表色は i_{RH} に関連付けられた色成分 H_x であり、通常は平均 RGB 値、あるいは平均色相値を用いる。すなわち、本手法の記憶色特性値 \mathbf{T}_p は、3 種類のカテゴリに対応した代表色 ($H_{skin}, H_{gm}, H_{sky}$) から成る。

記憶色補正においては、対象物体領域において選択的に色補正することが要請される。そこで 4.4 節に述べた任意色相の特色補正法を利用する。色補正処理は、式(4-44)で表される。

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \kappa \times \begin{pmatrix} a_r \\ a_g \\ a_b \end{pmatrix} \quad (4-44)$$

ここで、 (R, G, B) は入力画像中の任意の画素の RGB 値であり、 (R', G', B') は色補正後の値である。そして (a_r, a_g, a_b) は強調パラメータであり、 κ は特色度と呼ぶ重み値であり、画像中の画素値と対象色との間の類似度を表す。 κ の一般形式を式(4-46)に示す。

$$\kappa = g_H(|c_H - h|) \cdot g_S(|c_S - s|) \cdot g_V(|c_V - v|) \quad (4-45)$$

(c_H, c_S, c_V) は記憶色対象カテゴリの中心 HSV 値である。今、 (c_R, c_G, c_B) を、記憶色対象カテゴリ(肌色、草木色、空色)の中心色と考えると、中心 HSV 値 (c_H, c_S, c_V) は (c_R, c_G, c_B) を色座標変換することで算出される。そして、この中心 HSV 値 (c_H, c_S, c_V) が、代表色 \mathbf{T}_p (記憶色画質特性値)から、ルックアップテーブルを参照することによって取得される。

関数 $g_H(x), g_S(x), g_V(x)$ は重み関数であり、その出力値は $x=0$ の時 1.0 であり、 x が大きくなるに従って 0.0 に近づく。実際に用いた κ の形式を、式(4-46)に示す。

$$\kappa = \frac{\phi(w_H - |c_H - h|)}{w_H} \times s \times v \quad (4-46)$$

$$\phi(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

ここで、 w_H は色相許容角、 c_H は補正対象カテゴリの中心色相である。 (h, s, v) は入力画像の任意画素の HSV 値である。複数の記憶色を同時に補正するためには、式(4-47)のように連続して補正を行えばよい。図 4-21 に記憶色補正結果の一例を示す。

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \kappa_1 \times \begin{pmatrix} a_{r1} \\ a_{g1} \\ a_{b1} \end{pmatrix} + \kappa_2 \times \begin{pmatrix} a_{r2} \\ a_{g2} \\ a_{b2} \end{pmatrix} + \dots \quad (4-47)$$



(a) 元画像 (b) 肌色抽出結果 (c) 補正後の画像

図 4-21: 日本人の肌色に対する記憶色補正結果

4.5.8. 鮮鋭度補正

鮮鋭度（シャープネス）は、画像中の細かな構造をどれだけ明瞭に表現できるかを示す尺度である。画像の鮮鋭度を補正する手法として、4.2 節の最適鮮鋭化手法を用いる。鮮鋭度特性値 \mathbf{T}_s は、エッジシャープネス値(ES 値)を採用する。エッジシャープネス値(ES 値)は画像のエッジ領域における高周波成分の平均強度で定義される。本節の手法における ES 値は、カラー画像の輝度成分として XYZ 表色系の輝度値 $Y(x, y)$ を用いて算出する。式(4-48)に鮮鋭度特性値 \mathbf{T}_s の定義を示す。なお \otimes はコンボリューション演算を表す。

$$\mathbf{T}_s = ES = \frac{\iint_{E(Y)} |Y(x, y) \otimes ss(x, y)| dx dy}{A_E(Y)} \quad (4-48)$$

式(4-48)において、 $Y(x, y)$ は入力画像の輝度成分を示し、 $E(Y)$ はエッジ領域を示す。また $A_E(Y)$ はエッジ領域の面積を示す。 $ss(x, y)$ は主観的鮮鋭さに重みを持つ高周波帯域フィルタである。エッジ領域 $E(Y)$ を算出するためにはエッジ検出フィルタが必要である。ここでは、Sobel フィルタをエッジ検出に用いた後、適切なしきい値処理によってエッジ領域を求める。

鮮鋭度強調パラメータは、入力画像の高周波強調ゲインである。鮮鋭度補正は、式(4-49)に示されるように明度成分の高周波強調によって行われる。なお、 $k > 0$ の時は鮮鋭化された画像が得られ、 $k < 0$ の時は暈けた画像が得られる。

$$Y'(x, y) = Y(x, y) + k \cdot (Y(x, y) \otimes ss(x, y)) \quad (4-49)$$

高周波成分の係数 k は、鮮鋭度特性値の ES 値が最適値となるように適応的に求められる。ES 値の定義式(式(4-48))と、鮮鋭度補正式(式(4-49))とから、強調係数 k の値は式(4-50)によ

て算出される。なお ES_o は鮮鋭度特性値の最適値である。

$$k \approx \frac{(ES_o - ES) \cdot A_E(V)}{\iint_{E(V)} |V(x, y) \otimes H(x, y) \otimes H(x, y)| dx dy} \quad (4-50)$$

k を求める際の演算量を削減するため、 ES 値からテーブル変換によって直接、値を参照することも可能である。この場合、あらかじめ複数のサンプル画像を評価し、 ES 値と対応する k の平均値との、対応テーブルを作成しておく。そして、入力画像から ES 値を算出すると、直ちにテーブル変換によって k 値を推定することができる。実応用では有効な実装方法である。

4.5.9. 強調パラメータのリミッタ

本節で提案する画像の自動高画質化は、多くの画像において適切に動作する。しかしながら、カメラの露光やホワイトバランス設定の不良などの原因により、画像の色分布が極端に偏っている場合は、適切な強調パラメータ推定ができない。このような推定不良は物体のクローズアップや、芸術写真等でも起こる可能性がある。そこで後段のパラメータ制御方法として、モノトーン画像適応処理と、強調パラメータの制限処理とを導入した。

モノトーン画像とは、白黒画像やセピア調、あるいは単一色の壁面など、単一のトーンを有する画像を示す。モノトーン画像適応処理は、コントラスト特性値 HL と SD を用い、式(4-51)に示す方法でパラメータを変更する。 a は元の強調パラメータであり、 a' が変換後のパラメータである。 HL と SD の差分 d を $\beta(d)$ 関数に作用させた値を、元の強調係数 a に乗算することにより、 a' を求める。図 4-22 に $\beta(d)$ 関数の一例を示す。図 4-22 の Dh はモノトーン評価のためにあらかじめ設定する定数である。 HL と SD の差分 d が Dh よりも小さい時は、変換後の強調パラメータ a' は、元の値 a よりも小さくなる。

$$\begin{aligned} d &= HL - SD \\ a' &= a \times \beta(d) \end{aligned} \quad (4-51)$$

モノトーン画像適応処理後の強調パラメータ a' は a_{limit} によって制限される(制限処理)。そして最終的な強調パラメータ a_{fin} が得られる。この制限処理は、実応用において過剰強調を防ぐために必要である。

$$a_{fin} = \min(a', a_{limit}) \quad (4-52)$$

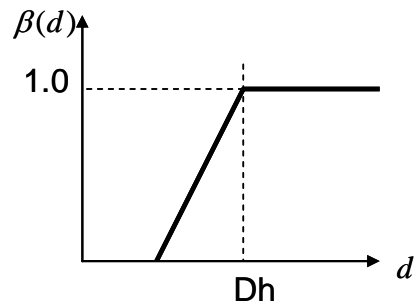


図 4-22: モノトーン画像に対する強調パラメータ適応処理

4.5.10. 画像処理結果

図 4-23 に、本手法の 6 種類の補正処理を全て適用した処理結果の一例を示す。なお、それぞれの画質要因に対する最適 QFV は、観察環境によって調整が必要であるが、標準値でも大抵はうまく動作する。例えば、ホワイトバランスの最適特性値は、メディアの白色値(24bit カラー画像では通常(255,255,255))、コントラストの最適特性値は、画素値の取りうる範囲の最大値と最小値、明度の最適 QFV では、明度の取りうる範囲の中間値(128)付近、などである。但し、記憶色補正については、微調整が必要となる。



(a) 入力画像



(b) 出力画像

図 4-23: 自動高画質化の処理例

4.5.11. 並列解析を用いた高速アーキテクチャ

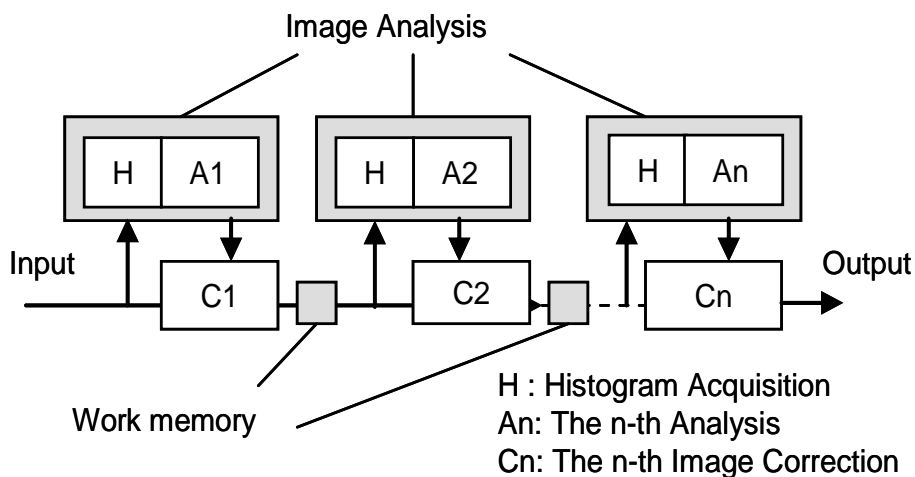
自動高画質化手法を様々な組み込み機器等に適用するためには、演算コストの削減が必要である。そのため、パラメータ抽出と画像補正とを一括して行うことで演算量を削減する高速アーキテクチャについて述べる。

自動高画質化手法は、図 4-24(a)に示すようにそれぞれの画質要因に対応した補正処理を、逐次処理して行く方式を用いていた(逐次方式)。すなわち一つの補正毎にヒストグラムを取得して解析し、算出されたパラメータを用いて補正した出力画像をバッファに出力していた。そのため、一つの補正のたびに出力画像データを書き込む必要があり、速度低下の原因となっていた。

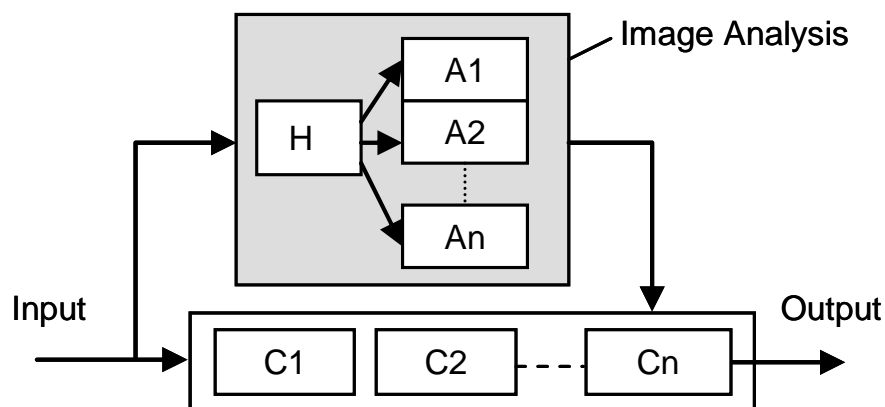
そこで画素単位の色階調補正については、ヒストグラム解析とパラメータ抽出を、複数の画質要因に関して並行して行う方式を検討した(並列方式)。図 4-24(b)に並列方式のブロック図を示す。並列方式では、ヒストグラム取得が一度だけで済み、出力画像の書き込み回数を減らすことができる。そこで高速化版の自動高画質化においては、並列方式による高速アーキテクチャを採用する。

しかしながら並列方式では、本来は前段の出力画像に対して画質解析すべきところを、全てのパラメータを補正前の入力画像を元に算出するために、本来の結果とは異なるパラメータ値が得られる可能性がある。この解析誤差による画質への影響については、後の主観評価実験によ

って検証する。



(a) Serial analyses architecture



(b) Parallel analyses architecture

図 4-24: 自動高画質化の逐次方式と並列方式

一般消費者向け電子機器において、カラー画像伝送には YUV 画像データが最も多く利用されている。従って、本高速アーキテクチャにおいては、YUV 形式を入出力画像の色空間とする。そしてここでは、高速な強調処理を実現するために 6 つの画像補正処理を一部統合した上で、次のステップで実行する。

(1)マトリクス統合処理

入力 YUV から RGB への色変換とホワイトバランス補正と彩度強調を統合する。YUV 画像に

対するホワイトバランス補正と彩度強調を、式(4-53)に示すように単一の3×3行列Ψに統合する。ここで、行列UはYUVからRGBへの変換行列である。

(4-53)

$$\begin{pmatrix} R_1 \\ G_1 \\ B_1 \end{pmatrix} = \Psi \cdot \begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \mathbf{W} \cdot \mathbf{U} \cdot \mathbf{S} \cdot \begin{pmatrix} Y \\ U \\ V \end{pmatrix}$$

(2) トーンカーブ統合処理

コントラスト強調と明度補正処理とを統合する。どちらもトーンカーブ補正であるため、2つのトーンカーブを1つに合成・統合する。

(3) 記憶色補正処理

(4) RGBから出力YUVへの変換

(5) 輝度成分Yに対する鮮鋭度強調

なお、鮮鋭度強調を除く5つの補正処理(彩度、コントラスト、明度、ホワイトバランス、記憶色)は画素単位の処理である。従って、これらの5つの補正処理は1パス(1重のforループ)によって実装することができる。

最終的な高速アーキテクチャによるブロック図を、図4-25に示す。高速アーキテクチャは、鮮鋭度解析(Sharpness Analyses: SA)、鮮鋭度補正(Sharpness Correction)、統合解析(Integrated Tone Analyses: ITA)、統合補正(Integrated Tone Correction: ITC)の4つのブロックから構成されている。統合解析と統合補正のブロックは、画素単位の5種類の処理(彩度、コントラスト、明度、ホワイトバランス、記憶色)の解析及び、補正を行うブロックである。

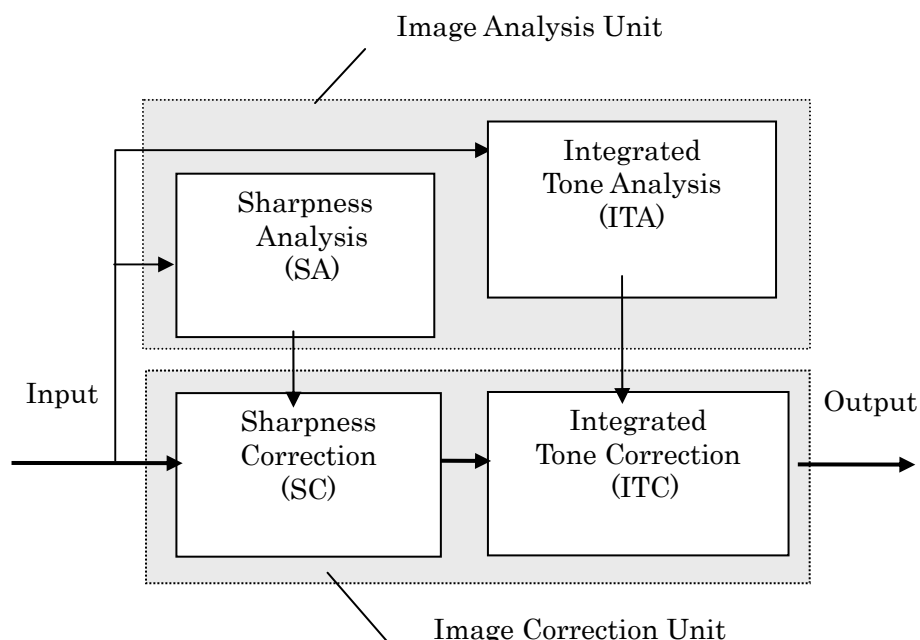


図 4-25: 高速アーキテクチャ

4.5.12. 処理速度の比較

図 4-26 に、逐次方式と並列方式による処理速度の比較を示す。速度値は 7 種類の画像に対する 10 回分の処理速度の平均を示している。横軸の QVGA(s) は QVGA サイズでの逐次方式を示し、他は並列方式（高速版）の結果である。2.4GHz の一般的な PC において、QVGA サイズでは逐次方式が 80ms を要するのに比べ、並列方式では 10ms 以下で実行できた。すなわち、負荷率 30% としても、QVGA サイズでリアルタイム処理可能であることが分かった。

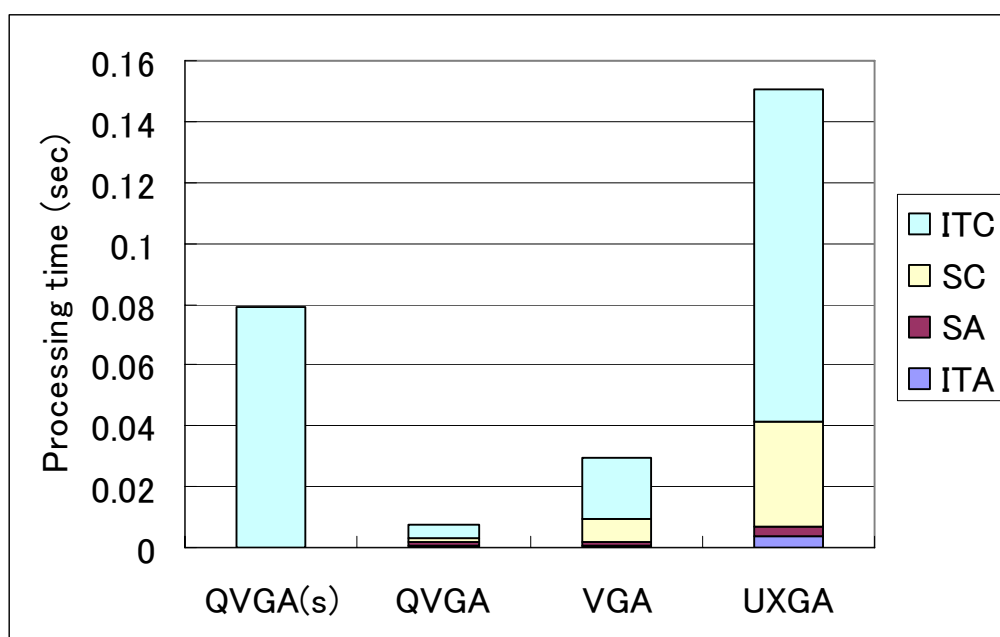


図 4-26: 処理速度の比較(Pentium4 2.4GHz, WindowsXP)

表 4-11 は、組み込み用 CPU である ARM9 のシミュレータによる比較結果である。表内の処理速度は 200MHz にて CPU 使用率 100% 時の推定値である。

表 4-11: CPU サイクル数の比較(ARM simulator: ARM926EJ)

Image size	QVGA	UXGA
Cycle count	19 (M cycle)	247(M cycle)
Processing time (sec) (estimated for 200MHz)	0.11	1.14

4.5.13. 実験結果

図 4-27、図 4-28、図 4-29、図 4-30 は、それぞれ A,B,C,D の 4 種類のカラー画像(ポートレート、風景、オブジェクト、テストチャート[78])に対する自動高画質化処理の出力結果を示す。図中の左側が原画像、右側が処理後の画像を示す。いずれの画像も、高速アーキテクチャによって処理されている。なお本実験では入力 RGB 画像は sRGB 色空間[79]で表現されているものとした。最初の 3 種類(A,B,C)は品質が劣化した入力画像であり、どの画像も高画質に強調処理されていることが分かる。一方で、4 番目の画像 D はもともと高品質な入力画像であり、出力画像も見た目にほとんど変化していない。すなわち、入力画像品質に応じて、低品質の画像のみを強調し、高品質な画像はそのまま出力されていることが分かる。



原画像

処理画像

図 4-27: 6 つの画質要因に対応した高画質化処理結果(A)



原画像

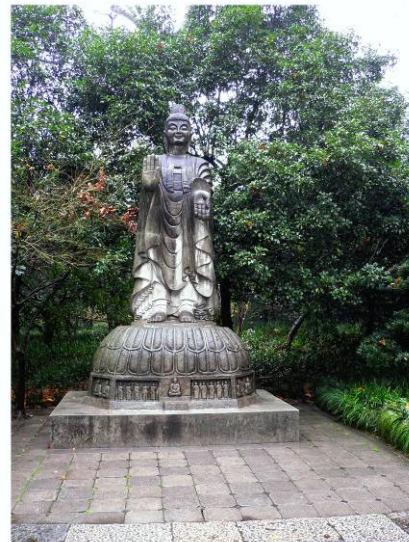


処理画像

図 4-28: 6 つの画質要因に対応した高画質化処理結果(B)



原画像



処理画像

図 4-29: 6 つの画質要因に対応した高画質化処理結果(C)

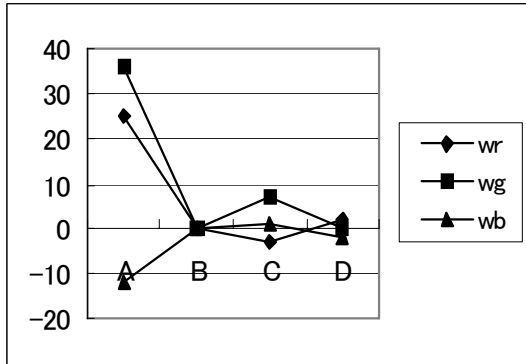


図 4-30: 6つの画質要因に対応した高画質化処理結果(D)

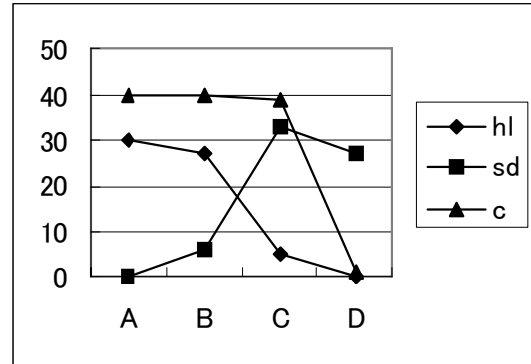
これらの4種類の画像に対する、実際に算出された強調パラメータを表 4-12 に示す。また、強調パラメータの変化をグラフ化したものを図 4-31 に示す。本結果より、画像 D については、いずれの強調パラメータ値も小さな値となっている。すなわち、画像 D は見た目の印象通り、ほとんど強調処理が施されていない。これは、画質解析によって、各画質特性が最適に近いと判断されたことになる。また、画像 A,B,C については、元画像の画質特性に応じて強調パラメータが様々に変化しているのが分かる。

表 4-12: 画像(A,B,C,D)の強調パラメータ

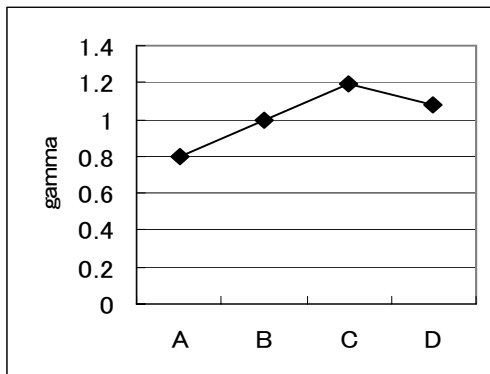
画像	wr	wg	wb	hl	sd	c	gamma
A	25	36	-12	30	0	40	205
B	0	0	0	27	6	40	256
C	-3	7	1	5	33	39	306
D	2	0	-2	0	27	1	277



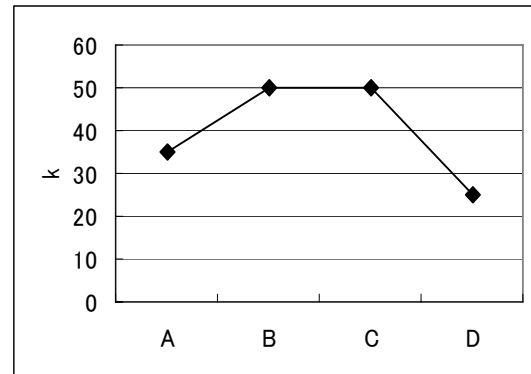
(a) (wr,wg,wb)



(b) (hl, sd) and c



(c) gamma

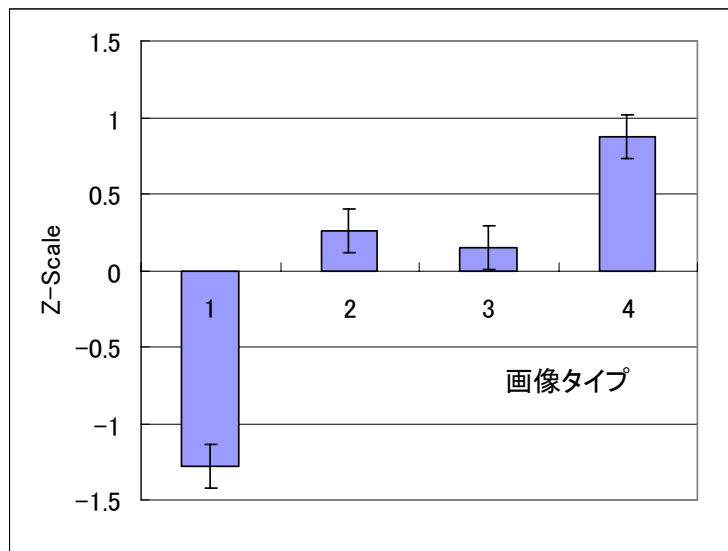


(d) k for Sharpness

図 4-31: 画像(A,B,C,D)における強調パラメータ変化

表 4-13: 主観評価実験の条件

Display	EIZO ColorEdge CG21
Display Resolution	1600x1200
Color Property	sRGB
Picture Size	1280 x 960
Environment	D65 Fluorescent tube
Num. of pictures	30
Num. of subjects	12



画像タイプ: 1.原画像, 2.ITA+ITC(逐次), 3.ITA+ ITC(並列), 4.ITA+SA+ ITC+SC(並列)

図 4-32: 主観評価実験結果

本手法の有効性を検証するため、複数の自然画像を用いて主観画質実験を実施した。評価方法はサーストンの一対比較法に従った[80][81]。実験条件を表 4-13 に示す。図 4-32 は、全評価画像に対する実験結果をサーストンの比較判断の法則を用いて間隔尺度(Z スケール)で表した比較結果である。横軸は画像タイプを示し、1 が原画像、2 が逐次方式で ITC+ITA を処理したもの、3 が並列方式で ITC+ITA を処理したもの、4 が並列方式で全処理(ITC+ITA+SC+SA)を行ったものである。エラーバーは正規分布を仮定した時の 95%信頼区間を示す。

4 種類の画像タイプに対する刺激に対し、分散分析を行った結果、各画像タイプにおける平均刺激値は等しくないことが有意水準 1%で確認された($F(3,44):262.7, P<0.01$)。次にライアン法による多重比較検定(有意水準 1%)を行った結果、画像タイプ 2 と 3 については平均値に有意な差はなく、それ以外の組み合わせについては有意な差が認められた。以上より、本手法による結果 2, 3, 4 は、共に原画像の画質を上回ることが分かった。また 2 (逐次方式) と 3 (並列方式) の間に有意な差はなく、並列方式への移行による画質の低下は見られないことが分かった。また、鮮鋭度補正を含む最終画質結果 4 は、最も良い評価結果となった。

4.5.14. 6つの画質要因解析による自動高画質化まとめ

本節では、デジタルカメラやビデオ画像に対し、6つの画質要因を補正可能な映像の自動高画質化手法について述べた。本手法は、画質特性値 QFV を用いた強調パラメータの適応制御が特徴である。本手法を様々なシーンで撮影された自然画像に適用し、主観評価実験によってその有効性を示した。また高速アーキテクチャ導入により、PC 上でのリアルタイム処理を実現した。

本手法の実装では強調パラメータの制限処理を行っているため、厳密には入力画像の QFV を常に最適値に収束させている訳ではない。また、最適 QFV 値は、特定のシーン（夜景、ポートレート、花など）において若干変化することも分かっている。例えば、ポートレートであれば、鮮鋭度特性値は低めが好まれる。以上より、QFV 値の算出方法についてはさらに改善の余地が残されている。解決方法として例えば、特定シーンへの分類機能を QFV 算出に統合し、特定のシーンに応じて QFV 値を制御する方法が考えられる。分類機能との統合は、今後の課題である。

4.6. 画像強調のまとめ

本章では、画像強調における部分空間情報の適応的解析と、その適用事例について述べた。具体的には、カラー画像の自動高画質化を目的とし、様々な環境下で撮影されたカラー画像を最適に強調する手法について提案した。従来この種の強調処理は、熟練技術者の多大な労力をかけたマニュアル調整が必要であった。そこで本研究では、画質を構成する複数の要因について考察し、各画質要因に影響を及ぼす部分領域の解析による画質特性値の算出方法を提案した。さらに、これらの画質特性値が最適値となるように、強調パラメータを自動的に制御する方法を開発した。

はじめに、鮮鋭化の最適調整を目的とし、エッジ領域における平均高周波成分を解析することによって、鮮鋭度特性値であるエッジシャープネスを提案した。またエッジシャープネス値を用いた鮮鋭度強調パラメータの制御方法を考案し、実画像においてその効果を示した。

次に最適鮮鋭化をカラー画像処理に拡張し、さらに2つの画質特性値を追加することにより、HSV 色空間を用いた彩度、コントラスト、鮮鋭度によるカラー画像の自動高画質化手法を提案した。主観評価実験の結果、本提案手法によって、マニュアル調整による出力画質に比べて約70%の出力画質が得られることが分かった。さらに、画像中の部分領域への色補正を目的とした、任意色の特色分離補正法を提案した。

そして最後に上記の成果を踏まえ、鮮鋭度、彩度、コントラスト、明度、ホワイトバランス、記憶色の6つの画質要因を改善する、カラー画像の自動高画質化の実現について述べた。画像中の特定の部分領域の解析によって、それぞれの要因に対応した画質特性値を定義し、さらにその画質特性値が最適値となるような、強調パラメータの算出方法を考案した。さらに組み込み機器等への実装に向けた、高速化手法について述べた。自然画像を用いた主観評価実験を実施し、様々なシーンの画像に対して有効に動作することを示した。

本章で述べた自動高画質化のための画像強調処理の研究は、処理対象となる画像が一般の自然画像であるため、画像中に写る対象物体を限定することは難しい。従って、非常に高度な適応性を要求されるタスクと言える。本研究の中には、適応性向上に向けた3つのアプローチ、部分空間選択の高度化、部分空間の比較・統合、部分空間情報に基づく適応動作、の考えを全て含んでいる。すなわち、6つの画質特性値算出のための適切な部分領域の選択(部分空間選択の高度化)、そして6つの画質特性解析の統合(部分空間の比較・統合)、最後に画質特性値に応じた強調パラメータ制御法(部分空間に基づく適応動作)、である。

本研究では、このように部分空間情報の適応的解析を広範囲に利用することによって、画像強調処理において高度な適応性を実現することができた。

第5章 結論と今後の展望

本論文では、部分空間情報の適応的解析を用いた画像認識処理及び画像強調処理の提案と、その検証結果について述べた。人間が情景を観察する場合、実行する具体的なタスクに応じて、注目する部分を適応的に変化させている。このような視覚における部分領域への注目処理は、画像空間から適切な部分空間を選択する操作であると考えられる。これらの部分空間情報を、適応的に選択・統合することによって、画像認識と強調処理における様々な変動要因に対する適応性実現に取り組んだ。

第2章では、部分空間情報を用いた適応性実現のためのアプローチについて検討し、部分空間選択の高度化、部分空間の比較・統合、部分空間に基づく適応動作、の3種類のアプローチが存在することを考察した。そして画像認識と画像強調において、これらのアプローチに基づく複数の改善手法を提案した。

第3章では、画像認識における具体的な研究内容について述べた。画像認識における様々な対象の中でも、環境変動の影響が非常に大きいとされる人間の顔照合・顔追跡を課題として、適応性の改善を試みた。3.2節では、静止画顔照合の改善手法として、適応的部分領域照合法と摂動空間法を提案した。適応的部分領域照合法は、適切な部分領域を選択して領域毎に照合を行う方法であり、蔽や表情変化等への頑健性を向上させた。摂動空間法は1枚の画像を元に3次元モデルを用いて摂動画像群を生成し、それらの画像群の張る部分空間(摂動空間)を用いて照合を行う方法である。摂動空間法により、照明変動や姿勢変動に対する頑健性を向上させることができた。3.3節では、動画像照合における部分空間同士の比較方法として、相互投影距離法を提案した。本手法は部分空間同士を直接比較するため、動画像特有のフレーム間の相関を照合スコアに反映させることができる。照合実験により、従来法に比べて照合精度を維持したまま、大幅に演算量を削減可能であることを示した。3.4節では、物体追跡における、新たな部分空間解析手法について述べた。特徴点の動きベクトルを剛体運動の部分空間に射影することで、剛体運動における6自由度の運動パラメータを高速に求めることができる。本手法を人物顔の追跡に応用し、ソフトウェア処理での実時間顔追跡システムを実現した。

このように本論文では、画像認識における環境適応性改善手法として、部分空間解析を用いた新しい顔照合、顔追跡手法を提案し、その有効性を示した。これらの研究では、認識対象となる物体を人物顔として進めてきたが、提案手法の中心にある部分空間解析方式は、顔特有の制約を含んでいない。静止画照合における摂動空間法は、顔の3次元モデルを使用しているが、認識する対象物体に応じて3次元モデルを置き換えればよい。また本来、人物顔は、人工物や文字などに比べて、変動要因の大きな認識対象である。従って提案手法は、顔以外の物体認識・追跡にも十分適用可能であると考えられる。

第4章では、画像強調における具体的な研究内容について述べた。本研究では、カラー画像の自動高画質化手法の実現を目的として、様々な環境下で撮影されたカラー画像を、劣化に関する事前知識なしに最適な強調を行う。従来この種の強調処理は、熟練技術者の多大な労力をかけたマニュアル調整が必要であった。そこで本研究では、画質を構成する複数の要因について考察し、各画質要因に影響を及ぼす部分領域の選択と解析によって、画質特性値を算出する方法を考案した。さらにこれらの画質特性値が最適値となるように強調パラメータを自動的に制御する方法を考案した。4.2節では、高画質化のための最適鮮鋭化を目的とし、エッジ領域における平均高周波成分を解析することによって、鮮鋭度特性値、エッジシャープネスを提案した。またエッジシャープネス値を用いた鮮鋭度強調パラメータの制御方法を考案し、実画像においてその効果を示した。4.3節では、最適鮮鋭化をカラー画像処理に拡張し、さらに2つの画質特性値を追加することにより、HSV色空間を用いた彩度、コントラスト、鮮鋭度によるカラー画像の自動高画質化手法を提案した。主観評価実験の結果、本提案手法はマニュアル補正による出力画質に比べて約70%の画質が得られることが分かった。4.4節では、画像中の部分領域への

色補正を目的として、任意色の特色分離補正を提案した。本補正処理は、記憶色の強調処理に利用されている。

4.5 節では、前節までの成果を踏まえ、鮮鋭度、彩度、コントラスト、明度、ホワイトバランス、記憶色の 6 種類の画質要因を改善するカラー画像の自動強調手法の実現について述べた。画像中の特定の部分領域の解析によってそれぞれの要因に対応した画質特性値を定義し、さらにその画質特性値が最適値となるような強調パラメータの算出方法を考案した。さらに組み込み機器等への実装に向けた、高速化手法について述べた。そして自然画像を用いた主観評価実験を実施し、様々なシーンの画像に対して有効に動作することを示した。

画像強調に対する一連の研究の結果、画像中の注目する部分領域を変化させることによって様々な画質要因を推定可能であり、これらの推定値を利用することで適応的に画像機器の画質を制御可能であることを示した。画像強調においては、処理対象があらゆる自然画像のため、人間の視覚システムに近い環境適応性が必要となる。本研究では、従来から画質評価で培われていくつかの画質要因に基づいて部分空間情報の解析を行うことにより、画像強調処理における適応性を改善した。本研究による画像強調技術は、カラー画像の自動高画質化機能の名称で、サーバ型画像処理システムやカメラ付携帯電話など、様々な画像機器に展開されている。

本論文では、画像認識処理と画像強調処理において、部分空間情報の解析を用いて環境適応性を改善する画像処理方式を提案した。しかしながら、新たな課題もいくつか顕在化している。以下に、本研究における今後の展望について述べる。

画像認識においては、提案手法の人物顔以外の物体に対する適用について、検証・評価が必要である。そして、対象物による違いが認識性能にどのように影響するかを考察し、適応化を改善することにより、一般的な物体認識の実現を果たすことができると考える。適応化の改善方法としては、例えば、カーネル空間等の非線形な部分空間解析を利用する方法が考えられる。また提案手法では、あらかじめ求めておいた 3 次元モデルを用いて摂動空間を生成したが、3 次元モデルの生成自体を動的に行うことができれば、一般的な物体認識における実用性を向上させることができる。

画像強調においては、まず特定の画質要因に対応する部分空間の選択を、さらに高度化することが課題である。現状の画質特性値(QFV)の算出方法は、多数の画像を観察しながら試行錯誤の上で決定したものである。それらの QFV 算出を画像分類技術と融合し、特定のシーンに応じた QFV を算出することによって、さらに適応性を高めることが可能である。また、認識辞書の学習と同様に、複数のサンプル画像を統計解析することによって、対象の部分空間領域を自動選択する枠組みも考えられる。これが実現できれば、人間の労力を軽減するだけでなく、マシンパワーを利用して大量のデータを分析することで、人間には予測不可能な最適な部分空間を選択できる可能性がある。また画像強調に関しては、現状、対象物の姿勢変動について何も対応していない。姿勢変動への対応とは、例えば撮影された対象物を見やすい姿勢・大きさに変換する、といった操作である。実際に姿勢を補正するには、多視点画像を用いて任意視点画像を合成する手法等が必要となる。

さらに、人間の視覚システムに近い適応性を目指す上で、観察者が意識的に注目する部分領域情報の獲得と反映が有用と考えられる。そのためには、注目領域を自動的に判断するメカニズムと、画像認識・強調処理への新たな反映手法の研究が必要となる。

以上、結論と今後の展望を述べた。本研究活動を推進することによって、人間の視覚システムの持つ適応性を、さらに高度に実現できるものと考えている。

謝辞

本論文の作成にあたり、筑波大学大学院博士後期課程在籍時に、懇切丁寧なご指導を頂きました、筑波大学大学院システム情報工学研究科知能機能システム専攻の大田友一教授に、感謝の意を表します。

本論文の副査をお引き受け頂き、また適切なご意見を頂きました、筑波大学大学院システム情報工学研究科知能機能システム専攻の鬼沢武久教授、葛岡英明教授、亀田能成准教授、北原格講師に、感謝の意を表します。また、良い研究環境を提供して頂き、分け隔てなく接して頂いた、画像情報研究室のメンバの皆様に感謝致します。

本論文は、筆者が日本電気株式会社(NEC)において入社以来行ってきた、様々な研究成果を基に執筆致しました。その中で、名古屋市立大学大学院システム自然科学研究科生体情報専攻の田島譲二教授(元 NEC メディア情報研究所主席研究員)には、NEC において熱心にご指導頂き、心より御礼申し上げます。また、在外研究時にご指導頂いた、ケンブリッジ大学工学部ロベルト・チッポラ教授に感謝致します。

本研究の遂行にあたり、多大なご協力とご助言を頂いた NEC 共通基盤ソフトウェア研究所の塚田正人主任研究員、佐藤敦主幹研究員、坂本静生主任研究員(元)に、深く感謝致します。また、有益な討論、評価実験やシステム実装等を通じて、多くのご支援を頂きました、NEC 共通基盤ソフトウェア研究所の鈴木哲明氏、舟山知里氏、戸田真人氏、丸亀敦氏、石山墨氏、細井利憲氏、今岡仁氏、他研究所メンバ、そして NEC 情報システムズの新井雅年氏に、深く感謝申し上げます。

また、日頃の会社業務でのご支援に加え、社会人としての博士号取得に向けて励ましを頂いた、NEC 共通基盤ソフトウェア研究所の芹沢昌宏部長、杉山昭彦主席研究員、チームメンバ、そして側面からご支援頂いた全ての方々に感謝の意を表します。

最後に、本論文執筆を支えてくれた、私の家族に感謝致します。

参考文献

- [1] 大山,今井,和気編,“新編 感覚・知覚心理学ハンドブック,” pp.608-622, 誠信書房, 1994
- [2] 大山,今井,和気編,“新編 感覚・知覚心理学ハンドブック,” pp.287-317, 誠信書房, 1994
- [3] 高木,下田監修,“新編 画像解析ハンドブック,” 東京大学出版会, 2004
- [4] E.Oya, (小川,佐藤訳),“パターン認識と部分空間法,” 産業図書,1996
- [5] 石井,上田,前田,村瀬,“パターン認識,” オーム社,1998
- [6] 赤松,“コンピュータによる顔の認識—サーベイ—,” 信学論(D-II), J80-D-II, No.8, pp.2031-2046, 1997
- [7] W.Zhao, R.Chellappa, P.J.Phillips and A.Rosenfeld, “Face Recognition: A literature survey,” ACM Computing Surveys, vol.35, no.4, pp.399-458, 2003.
- [8] Y.Adini, Y.Moses and S.Ullman, “Face Recognition: The Problem of Compensating for Changes in Illumination Direction,” IEEE Trans. on PAMI, vol.19, no.7, pp.721-732, 1997
- [9] S.Hallinan, “A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions,” Proc. of CVPR, pp.995-999, 1994
- [10] P.N.Belhumeur and D.J.Kriegman, “What is the set of images of an object under all possible illumination conditions?,” Int. J. of Computer Vision, vol.28, no.3, pp.245-260, 1998
- [11] H.Murase and S.K.Nayer, “Visual Learning and Recognition of 3D Objects from Appearances,” Int. J. of Computer Vision, vol.14, no.1, pp.5-25, 1995
- [12] 福井,山口,鈴木,前田,“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,” 信学論(D-II), Vol.J82-D-II, No.4, pp.613-620, 1999
- [13] 石山,坂本,“照明条件と姿勢の両者の自動補正による顔照合,” 信学技報, PRMU2001-162, pp.59-64, 2001
- [14] 石山,坂本,“3次元形状と姿勢非依存な照明基底を用いた顔照合,” 信学技報, PRMU2001-13, pp.33-40, 2001
- [15] 今岡,坂本,“照明変動に強い顔照合,” 信学技報, PRMU2000-80, pp.35-42, 2000
- [16] W. Y. Zhao and R.Chellappa, “SFS Based View Synthesis for Robust Face Recognition,” Proc. of Int. Conf. on AFGR, 2000
- [17] D.J.Beymer, “Face Recognition under Varying Pose,” Proc. of CVPR, pp.756-761, 1994
- [18] 今岡,坂本,“顔の向きによらない顔画像照合方法,” 信学技報, PRMU1999
- [19] L.Wiskott, J.M.Fellous, N.Kruger and C.von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” IEEE Trans. PAMI, vol.17, no.7, pp.775-779, 1997
- [20] V.Blanz, S.Romdhani and T.Vetter, “Face Identification across Different Poses and Illuminations with a 3D Morphable Model,” Proc. of Int. Conf. on AFGR, 2002
- [21] Y. Li, S.Gong and H.Liddell, “Support Vector Regression and Classification base Multi-View Face Detection and Recognition,” Proc. of Int. Conf. on AFGR, 2000
- [22] K.Okada and C. von der Malsburg, “Pose Invariant Face Recognition with Parametric Linear Subspace,” Proc. of Int. Conf. on AFGR, 2002
- [23] 土居,陳,眞溪,大城,佐藤,千原,“顔照合による解錠制御システム,” 信学論, vol.J80-DII, no.8, pp.2203-2208, 1997
- [24] クリッピングデル,伊藤,“動画像の顔検出・追跡・認識への統合されたアプローチ,” PRMU98-200, pp.61-68, 1999
- [25] R.Gross, Y.Jie and A.Waibel, “Face Recognition in a Meeting Room,” Proc. of Int. Conf. on AFGR, 2000
- [26] M.Turk and A.Pentland, “Eigenfaces for recognition,” J. of Cognitive Neuroscience, vol.3, no.1, 1991
- [27] A.Pentland, B.Moghaddam and T.Starner, “View-based and modular eigenfaces for face recognition,” Proc. of CVPR, 1994
- [28] K.Etemad and R.Chellappa, “Discriminant Analysis for Recognition of Human Face Images,” J. Opt. Soc. Am. A, vol.14, no.8, pp.1724-1733, 1997
- [29] R.Brunelli and T.Poggio, “Face recognition: Features versus Templates,” IEEE Trans. on PAMI, Vol.15, No.10, 1993
- [30] 若林,鶴岡,木村,三宅,“特徴量の次元数増加による手書き数字認識の高精度化,” 信学論(D-II),vol.J77-D-II, no.10, pp.2046-2053, 1994
- [31] 山口,福井,前田,“動画像を用いた顔認識システム,” PRMU97-50, pp.17-24, 1997

- [32] 前田,渡辺, “局所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
- [33] 坂野,武川,中村, “核非線形相互部分空間法による物体認識,” 信学論(D-II), Vol.J84-D-II, No.8, pp.1549-1556, 2001
- [34] J.Chen, S.Yeh and C.Chen, “Inter-subspace distance: a new method for face recognition with multiple samples,” Proc. of the ICPR, Vol.3, pp.140-143, 2004
- [35] 高木,下田, “画像解析ハンドブック,” 東大出版会, pp.656-658,1991
- [36] K. Fukunaga, “Statistical Pattern Recognition 2nd ed.,” Academic Press, 1990
- [37] F.Kimura and M.Shridhar, “Handwritten Numerical Recognition based on Multiple Algorithms,” Pattern Recognition, vol.24, no.10, pp.969-983, 1991
- [38] B.Moghaddam and A.Pentland, “Probabilistic Visual Learning for Object Detection,” Proceedings of the ICCV, 1995
- [39] M.Black and Y.Yacoob, “Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motions,” In Proceedings of International Conference on Computer Vision, pp.374-381. 1995
- [40] T.F.Cootes and C.J.Taylor, “Active shape models - their training and application,” Computer Vision and Image Understanding, 61(1): pp.38-59, 1995
- [41] F.Pighin, R.Szeliski and D.H.Salesin, “Resynthesizing Facial Animation through 3D Model-Based Tracking,” In Proceedings of International Conference on Computer Vision, 1999, pp.143-150, 1999
- [42] P.Fua, “Regularized Bundle-Adjustment to Model Heads from Image Sequences without Calibration Data,” International Journal of Computer Vision, 38(2), pp.153-171, 2000
- [43] M.Harville, A.Rahimi, et.al., “3D Pose Tracking with Linear Depth and Brightness Constrains,” In Proceedings of International Conference on Computer Vision, pp.206-213, 1999
- [44] Y.Matsumoto, “An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement,” In Proceedings of International Conference on Automatic Face and Gesture Recognition, pp.499-505, 2000
- [45] A.H.Gee and R.Cipolla, “Fast Visual Tracking by Temporal Consensus,” Image and Vision Computing, Vol.14, No.2, 1996
- [46] R.Cipolla and A.Pentland, “Computer Vision and Human-Computer Interaction,” Cambridge University Press, 1999
- [47] D.P.Huttenlocher and S.Ullman, “Recognizing solid objects by alignment with an image,” International Journal of Computer Vision, 5(2), pp.195-212, 1990
- [48] T.S.Jebara and A.Pentland, “Parameterized Structure from Motion for 3D Adaptive Feedback Tracking of Faces,” In Proceedings of Computer Vision and Pattern Recognition, pp.144-150, 1997
- [49] O. Faugeras, “Three-dimensional computer vision: a geometric viewpoint,” MIT Press, 1993
- [50] J.Heinzmann and A.Zelinsky, “Robust real-time face tracking and gesture recognition,” In Proceedings of the International Joint Conference on Artificial Intelligence, pp.1525-1530, 1997
- [51] J.Heinzmann and A.Zelinsky, “3-D Facial Pose and Gaze Point Estimation Using a Robust Real-Time Tracking Paradigm,” In Proceedings of International Conference on Automatic Face and Gesture Recognition, 1998
- [52] D. H. Sattinger and O. L. Weaver, “Lie groups and algebras with applications to physics, geometry, and mechanics,” Number 61 in Applied Mathematical Sciences. Springer-Verlag, 1986
- [53] K.Kanatani, “Group-Theoretical Methods in Image Understanding,” Springer-Verlag,1990
- [54] 佐藤(淳), “コンピュータビジョン,” コロナ社, 1999
- [55] T.Drummond and R.Cipolla, “Application of Lie Algebras to Visual Servoing,” International Journal of Computer Vision, 37(1), pp.21-41, 2000
- [56] T.Drummond and R. Cipolla, “Real-time tracking of complex structures with on-line camera calibration,” Proceedings of British Machine Vision Conference, pp.574-583, 1999.
- [57] 大山,今井,和気編, “新編 感覚・知覚心理学ハンドブック,” pp.480-484, 誠信書房, 1994
- [58] 大山,今井,和気編, “新編 感覚・知覚心理学ハンドブック,” pp.349-356, 誠信書房, 1994
- [59] J.von Kries, “Chromatic Adaptation,” Source of color science. In: MacAdam, D.L.(Ed), MIT Press, Cambridge, MA, pp.109-119, 1970
- [60] D.L.MacAdam, “Chromatic Adaptation,” J. of Opt. Soc. America, vol.46, pp.500-513, 1956
- [61] 大山,今井,和気編, “新編 感覚・知覚心理学ハンドブック,” pp.484-486, 誠信書房, 1994

- [62] R.W.G. Hunt, "Objectives in Colour Reproduction," J. of Photographic Science, vol.18, pp.205-215, 1970
- [63] E.M.Crane, "An Objective Method for Rating Picture Sharpness: SMT Acutance," J.of the SMPTE vol.73 pp.643-647, 1964
- [64] R.G.Gendron, "An Improved Objective Method for Rating Picture Sharpness: CMT Acutance," J.of the SMPTE vol.82 pp.1009-1012, 1973
- [65] P.G.J.Barten, "Evaluation of subjective image quality with the square-root integral method," J.Opt.Soc.Am. A/vol.7, no.10 pp.2024-2031, 1990
- [66] R.L.De Valois and K.K. De Valois, "Spatial Vision," Oxford Univ. Press pp.147-175, 1987
- [67] 磯野, "テレビ信号の空間周波数レスポンス補正法-鮮鋭さから見た最適化-, " テレビジョン学会全国大会 1-7, 1982
- [68] A.R.Smith, "Color Gamut Transformation Pairs," Computer Graphics vol.12 pp.12-19, 1978
- [69] 坂田, 磯野, "視覚における色度の空間周波数特性(色差弁別いき)," テレビジョン学会誌, Vol.31, No.1, 1977
- [70] G.J.C.van der Horst and M.A.Bouman, "Spatiotemporal Chromaticity Discrimination," J.Opt. Soc. America, vol.59, no.11, pp.1482-1488, 1969
- [71] J.A.C.Yule, "Principles of Color Reproduction," John Wiley & Sons, 1967
- [72] 田島, "カラー画像複製論," 丸善, pp.71-74, 1996
- [73] 金森, 小寺, "色知覚空間におけるカラー画像の選択的色調整," 画像電子学会誌, vol.18, No.5, pp.302-312, 1989
- [74] M.Tsukada, C.Funayama and J.Tajima, "Automatic Color Preference Correction for Color Reproduction," Proc.SPIE, 4300, pp.216-223, 2001
- [75] A.Rizzi, C.Gatta and D.Mariani, "A new algorithm for unsupervised global and local color correction," Pattern Recognition Letters, vol.24, no.11, pp.1663-1677, 2003
- [76] H.Spitzer and S.Semo, "Color Constancy: a biological model and its application for still and video images," Pattern Recognition, vol.35, no.8, pp.1645-1659, 2002
- [77] S.N.Yendrikhovskij, F.J.J.Blommaert and H.de Ridder, "Optimizing Color Reproduction of Natural Images," 6th Color Imaging Conference, pp.140-146, 1998
- [78] ISO/JIS-SCID, "Graphic technology – Prepress digital data exchange – Standard image data (SCID)," ISO/JIS-SCID, JIS X9201-1995, Japanese Industrial Standard, 1995
- [79] IEC 61966-2-1, "Multimedia Systems and Equipment – Colour measurement and management – Part 2-1: Colour Measurement – Default RGB colour space –sRGB"
- [80] 佐藤, "統計的官能検査法," pp.271-281, 日科技連, 1985
- [81] 大串, 福田, 中山, "画質と音質の評価技術," 昭晃堂, 1991

公表論文リスト

【査読付き学術雑誌論文】

- [AI1] A.Inoue and J.Tajima, "Adaptive Image Sharpening Method Using Edge Sharpness," IEICE Trans. Vol.E76-D, No.10, pp.1174-1180, 1992
- [AI2] A.Inoue, T.Drummond and R.Cipolla, "Real Time Feature-based Facial Tracking Using Lie Algebras," IEICE Trans. Vol.E84-D, No.12, pp.1733-1738, 2001
- [AI3] A.Inoue and A.Sato, "Face Recognition Based on Mutual Projection of Feature Distributions," IEICE Trans. Vol.E91-D, No.7, pp.1878-1884, 2008

【査読付き講演論文】

- [AI4] A.Inoue and J.Tajima, "Adaptive Sharpening Method Using Edge Sharpness," Proc of MVA, pp.227-330, 1992
- [AI5] A.Inoue and J.Tajima, "Selective Color Correction for Arbitrary Hues," Proc. of ICIP," Vol.III, pp.38-41, 1997
- [AI6] A.Inoue, T.Drummond and R.Cipolla, "Feature-based Real-time Human Face Tracking Using Lie Algebras," Proc. of MVA, 2000
- [AI7] A.Inoue, T.Suzuki and M.Tsukada, "Adaptive Video Enhancement Using Preferred Color Reproduction," Proc. of VPQM, 2005
- [AI8] A.Inoue and A.Sato, "Face Recognition Using Mutual Projection of Feature Distributions," Proc. of Subspace2007, 2007
- [AI9] T.Suzuki, A.Inoue, M.Tsukada and J.Tajima, "Adaptive Image Quality Improvement Method for Moving Pictures using Smoothing Parameter Transition and Scene Cut Detection," Proc. of the European Conf. on CGIV, 2002
- [AI10] 井上, 田島, "エッジシャープネスを用いた最適鮮鋭化," 第23回画像工学コンファレンス論文集, 1-5, pp.23-36, 1992
- [AI11] 井上, 田島, "彩度、コントラスト、鮮鋭度の調整によるカラー画像の自動画質改善," 画像工学コンファレンス論文集, 1993
- [AI12] 井上, 塚田, 佐藤(洋), 北野, 新井, 田島, "インテリジェント画像編集システム CREST の開発," 第25回画像工学コンファレンス, pp.225-226, 1994 (優秀ポスター賞受賞)
- [AI13] 井上, 田島, "任意色相に対する特色補正機能," 画像センシングシンポジウム 1996
- [AI14] 井上, 佐藤(敦), 坂本, "部分領域マッチングと摂動空間法を用いた顔照合," 画像センシングシンポジウム 2003
- [AI15] 鈴木(哲), 井上, 塚田, 田島, "シーン適応型動画像自動高画質化," 画像センシングシンポジウム 2001
- [AI16] 斯波, 井上, 田島, "顔認識に基づく TV 番組人物検索システム," 画像センシングシンポジウム 2002
- [AI17] 櫻井, 井上, 佐藤(敦), "顔検出のための濃淡特徴を用いた高速な両目ペアの検出," 画像センシングシンポジウム 2001

【学会発表】

- [AI18] A.Inoue and J.Tajima, “Adaptive Quality Improvement Method for Color Images,” Proc of SPIE, Int. Symp.on Electronic Imaging, Vol.2179, pp.429-439, 1994
- [AI19] 井上, 田島, “エッジシャープネスを用いた HSV カラー画像の最適鮮鋭化,” 電子情報通信学会 春季全国大会, pp.736, 1993
- [AI20] 井上, 佐藤(洋), 田島, “Windows95 向け画像編集ツールにおける画像処理機能,” 電子情報通信学会 春季全国大会 1996
- [AI21] 井上, 坂本, 田島, 佐藤(洋), “パーソナルユーザのためのインテリジェント画像修正処理,” 電子情報通信学会研究報告 PRMU, Vol.76, pp.63-70, 1996
- [AI22] 井上, 田島, “任意色相に対する特色補正機能,” 電子情報通信学会 春季全国大会 1996
- [AI23] 井上, 佐藤(敦), “擬似マハラノビス相互距離法を用いた動画像顔照合,” FIT2002
- [AI24] 井上, 佐藤(敦), “相互投影距離法を用いた動画像顔照合,” 電子情報通信学会 研究報告 PRMU, 2002
- [AI25] 井上, 佐藤(敦), “部分領域マッチングを用いた顔照合,” 電子情報通信学会 総合大会, D-12-57, 2003
- [AI26] 坂本, 楠, 井上, 田島, “新聞紙面向け顔画像処理 —精細な背景-人物間輪郭生成—,” 電子情報通信学会 総合大会 1998
- [AI27] 坂本, 楠, 井上, 田島, “新聞紙面向け顔画像処理,” 画像の認識・理解シンポジウム (MIRU), 1998
- [AI28] 鈴木(哲), 井上, 塚田, 田島, “シーンチェンジを考慮した適応的動画像自動高画質化,” 電子情報通信学会 ソサイアティ大会 2001
- [AI29] 斯波, 井上, 田島, “類似顔による誤検出抑制を用いた人物検索システム,” FIT2002
- [AI30] 鈴木(哲), 井上, “周辺輝度適応型階調補正,” 電子情報通信学会 総合大会 2005

【解説】

- [AI31] A.Inoue and J.Tajima, “Adaptive Quality Improvement Method for Color Images,” NEC R&D, Vol.35, No.2, pp.180-187, 1994
- [AI32] 井上, “入出力機器に適用可能なカラー画像の自動高画質化技術,” 画像ラボ, Vol.6, No.9, pp.41-45, 1994
- [AI33] 井上, 坂本静生, “インテリジェント画像処理機能,” NEC 技報 Vol.49, No.9, pp.18-23, 1996
- [AI34] 井上, 鈴木(哲), “動画像の自動画質改善,” NEC 技報, Vol.56, No.3, 2003
- [AI35] 佐藤(敦), 井上, 鈴木(哲), 細井, “顔検出・顔照合エンジンの開発,” NEC 技報, Vol.56, No.3, 2003
- [AI36] 船田, 山下, 井上, 大中, “パーソナルロボット PaPeRo における画像認識,” 精密工学会 サマーセミナー「感性と画像応用」, 2002
- [AI37] A.Sato, A.Inoue, T.Suzuki and T.Hosoi, “NeoFace - Development of Face Detection and Recognition Engine,” Vol.44, No.3, 2003
- [AI38] 塚田, 鈴木(哲), 井上, “デジタル画像処理による自動高画質化技術,” 日本画像学会誌, Vol.43, No.2, pp.90-97, 2004