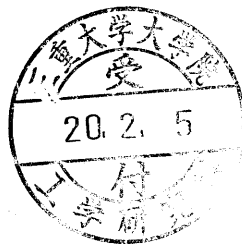


修 士 論 文

正面顔画像における
正中線の自動検出と
その顔部品抽出への応用



平成 19 年度修了
三重大学大学院工学研究科
博士前期課程 情報工学専攻

仲尾 望

はじめに

顔画像の認識を高精度に行うためには、入力された画像から対象人物の顔のみを含む領域を正しく切り出す必要がある。切り出す顔領域の決定には、両目や鼻、口などの顔部品の位置情報の利用が有用であるが、これらの顔部品の正確な抽出には高い計算コストが要求される。本研究では、この顔部品抽出の高精度化、高速化に利用できる顔正中線（中心線）を検出する新しい手法を開発した。顔部品の抽出処理の前に、顔の回転と中心位置を提案手法で検出した正中線を利用して補正すれば、顔部品抽出に必要な画像の探索を削減できるため、計算コストを大幅に削減できる。

提案手法は、1 個の正面顔が含まれるグレイスケール画像から顔の正中線を検出する。検出対象である正中線は、目と目を結ぶ線分（眼間線分）の垂直二等分線とする。このような正中線を検出することは、画像中の顔の傾きと中心位置とを検出することと等価である。提案手法は一般化 Hough 変換（Generalized Hough Transformation : GHT）によって正中線を検出する。GHT は、テンプレートとして与えられたパラメータ記述されない自由な形状の図形と同じ、あるいはよく似た図形を対象画像から検出するマッチング手法である。本手法では人物画像の 2 値エッジ画像に対して、その鏡像画像をテンプレートとして GHT によるマッチング処理を行い、マッチング結果より顔正中線を決定する。本研究ではさらに、顔の左右対称性の検出のための新しい GHT の高速化アルゴリズムを提案する。正中線自動検出手法の有効性を、FERET データベースから抽出した 2409 画像を用いた実験により評価した。実験の結果、提案手法は、従来手法よりも高速かつ安定に正中線を検出できることがわかった。

さらに、本研究では検出された正中線の顔部品抽出（目の抽出）に対する寄与について検討した。自動検出した正中線を利用して、濃度勾配特徴を用いた両目抽出実験を行った結果、両目抽出に必要な計算時間が約 40 分の 1 に短縮され、90.24%の抽出成功率で両目を抽出することができた。

本研究は、GHT の顔正中線検出への新しい応用と、対称軸の検出を行う新しい高速化アルゴリズムを提案した。提案手法の有効性と信頼性は、大規模な画像データベース上での実験により実証された。

今後の課題として、(1) 正中線検出性能，顔部品抽出性能のさらなる改善，(2) 顔部品抽出処理の高速化，(3) 抽出した顔部品を用いた顔検出が挙げられる．正中線の検出性能がさらに向上すれば，顔部品抽出の性能も向上が期待でき，処理の高速化にも寄与する．提案手法で抽出した目を用いれば，顔領域の検出，顔サイズの正規化，顔の傾きや向きの補正などを行うことができ，画像を用いた顔認識に多大に寄与すると期待できる．

目次

はじめに	i
第 1 章 緒言	1
1.1 研究の背景	1
1.2 本研究の目的	2
1.3 関連する先行研究	2
1.4 論文の構成	3
第 2 章 顔画像処理に関する従来研究	4
2.1 顔検出と顔認識, 顔部品抽出	4
2.2 顔検出の代表的な手法	5
2.2.1 事前知識に基づく顔検出	5
2.2.2 特徴に基づく顔検出	6
2.2.3 テンプレートマッチングに基づく顔検出	6
2.2.4 見え方 (appearance) に基づく顔検出	7
第 3 章 正中線自動検出	8
3.1 顔正中線とは	8
3.2 提案手法の概要	9
3.3 前処理	10
3.4 一般化 Hough 変換 (Generalized Hough Transformation : GHT)	12
3.5 GHT の高速化アルゴリズム	13
3.6 正中線の決定	13
3.7 まとめ	14
第 4 章 正中線を利用した顔部品抽出	16
4.1 顔部品抽出とは	16
4.2 テンプレートマッチングによる両目の抽出	18

4.2.1	問題の定義	18
4.2.2	特徴抽出と特徴変換	21
4.2.3	距離関数	23
4.2.4	テンプレートマッチング法の問題点	23
4.3	顔正中線を利用した高速化	24
4.3.1	正中線による探索自由度の削減	24
4.3.2	エッジ強度の2値化による探索範囲の削減	25
第5章	評価実験	27
5.1	FERET データベース	27
5.2	定量的評価のための指標	28
5.2.1	正中線検出の正しさの指標	28
5.2.2	顔部品抽出の正しさの指標	29
5.3	パラメータ決定のための予備実験	29
5.3.1	正中線検出におけるパラメータ決定	30
5.3.2	顔部品抽出におけるパラメータ決定	30
5.4	性能評価実験	32
5.4.1	顔正中線	32
5.4.2	顔部品抽出評価実験	37
第6章	結言	42
6.1	考察	42
6.2	今後の課題	45
付録A	目領域の大きさモデルの決定	46
付録B	校正用目座標の修正	50
謝辞		51
参考文献		51

第 1 章

緒言

1.1 研究の背景

バイオメトリクス認証に応用できる全自動の顔認識・認証の実現には、入力画像から顔領域を切り出す顔検出と切り出された領域の顔を認識する顔認識 [1] との両方が必要である。顔検出では、人物の顔が 1 個、あるいは複数含まれている（あるいは含まれていない）画像が入力として与えられる。このとき、顔の大きさは未知であり、例えば、 512×768 画素の大きさの入力画像において、 64×64 画素の領域に顔が存在する場合や、画像全体の 512×768 画素を顔が占める場合があり得る。顔検出の課題は、このような入力画像から顔領域を切り出し、非顔領域と分割することである。特に、顔として切り出される領域は、衣服や毛髪などを含まずに、額とあご、左右の頬までを含む、いわゆる顔面を正しく捉えていることが求められる。これは、顔認識 [2] の精度が顔検出の精度に強く依存するためである。例えば、Chen ら [3] は、検出された枠が顔面を正しく捉えていないとき、枠内に含まれた非顔物体が顔認識に必要な特徴ベクトルの抽出処理に悪影響を及ぼすことを示した。

人物の顔には、目や鼻、口など目印となる顔部品が存在し、それらの位置情報は顔検出に利用できる。もし、これらの顔部品の位置を知ることができれば、顔の検出と顔領域の抽出は、容易に、かつ正確に行うことができる。

しかし、顔部品の抽出は、高い計算コストを要求する。顔認識において、最も単純な対象である正面顔の画像でさえ、各顔部品の位置、顔の大きさ、回転など多くのパラメータを推定する必要がある。もし、何らかの手法によって顔部品抽出のための手がかりを得ることができれば、顔部品抽出はより単純化でき、総合的な計算コストを削減できると考える。

顔の正中線、言い換えれば顔の線対称軸は、顔部品抽出の計算コストを削減するための手がかりとして有望な候補の一つとして考えられる。顔の正中線を検出することは、顔の

傾きと眼間の中心位置とを検出することと等価である。よって、検出された正中線は、顔の傾きと位置を正規化するのに利用できる。これは、顔部品抽出の複雑性を削減することになる。また、顔正中線は顔認識のためにさらなる寄与を果たすと考えられる。例えば、Quintiliano らは、顔画像の対称化が顔認識の精度を向上させると報告した [4]。ここでの対称化とは、照明環境が明るい側の顔から暗い側の顔を再現することを意味し、これには顔正中線が用いられている。

1.2 本研究の目的

本研究では、一般化 Hough 変換 (Generalized Hough Transformation : GHT) に基づいた、顔正中線検出手法を提案する。本手法は、1 個の正面顔が含まれるグレイスケール画像から顔の正中線を検出する。画像中の顔は、しばしば傾いているため、検出手法は顔の傾きに頑健でなければならない。本論文では、顔正中線の自動検出技術と、その検出精度を、FERET データベース [5] に含まれる顔画像を用いた実験によって示す。

提案手法は、入力画像の 2 値エッジ画像とその鏡像画像のマッチングを行うことで、顔正中線を検出する。GHT はマッチングのために用いられる。対象が 2 値画像の場合、GHT は相関法と同じアルゴリズムであると知られている。しかし、GHT は計算コストが低くノイズに対し頑健である点で優位である。本論文ではさらに、顔の左右対称性の検出のための GHT の高速化アルゴリズムを提案する。

次に、検出された正中線を顔部品抽出 (Facial Feature Extraction : FFE) へと応用する。見え方に基づく顔認識において、顔部品の位置情報は顔面領域を正確に切り出すためなどに重要な役割を果たすが、その抽出には高い計算コストを要求する。提案手法では、検出された正中線を利用して、この計算コストを削減する。本論文では、濃度勾配特徴を用いた両目抽出の手法と、正中線が顔部品抽出にどのように貢献するのかについて述べる。また、同データベースを用いて両目の抽出実験を行い、その抽出精度を定量的に評価する。

1.3 関連する先行研究

顔の正中線を自動検出する手法として、X.Chen らが提案した顔正中線自動検出手法 [6] がある。彼らの手法では、任意の直線の両側の輝度値の差 (Gray Level Differences : GLD) に基づいて、対称性の尺度として定義された Y 値を最大にするような直線を顔の線対称軸として検出する。彼らのアプローチには次の 2 つの欠点がある。(1) Y 値は照明環境の変化に影響されやすい。例えば、顔が左右のいずれかから照明に照らされている場合、GLD は容易に悪影響を受ける。(2) Y 値が最大となる軸を求めるには膨大な計算コストがかかる。なぜならば、軸を求めるためには、回転や位置など推定すべきパラメータ

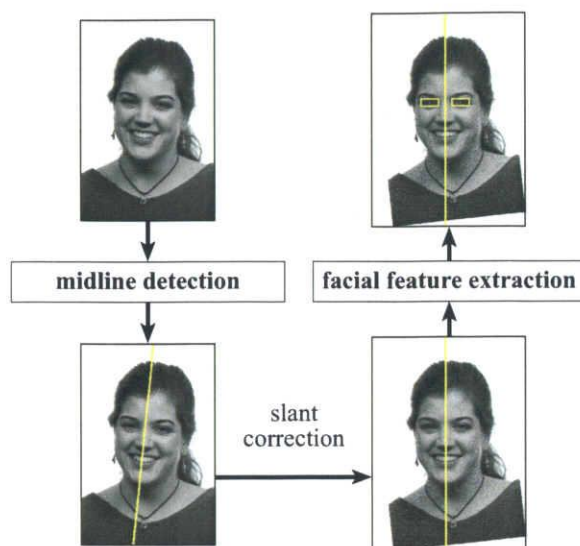


図 1.1: 提案手法の概要

Overview of the proposed eyes extraction method

の組み合わせの全てで Y 値の計算処理が必要となるためである。

これ以外に, Hiremath ら [7] によって提案された手法がある。この手法では, 顔の線対称軸を含めた *Lines-of-Separability (LS)* 顔モデルによって, 顔が記述される。この *LS* 顔モデルを得るためには, 顔の線対称軸を得る前に, 正面顔画像から両目を抽出する必要がある。この観点からいうと, この手法は提案手法とは反対に先に顔部品を抽出してから対称軸を求めるのでボトムアップな手法であると言える。

他の手法のグループとして, *Active Appearance Model* [8] のような, 顔の形や見え方のモデルを適用する手法がある。これらの手法は, 入力顔画像に対して, 事前に学習させたモデルやテンプレートを適用するものである。提案手法では, 正中線の検出にはそのような事前の学習は必要としない。

1.4 論文の構成

本研究では, まず顔の正中線を検出し, 続いて検出された正中線を利用して両目の抽出を行う。研究全体の流れ図を図 1.1 に示す。

本論文ではまず, 第 2 章にて顔画像処理に関する従来研究について紹介し, この分野の研究分野のこれまでの研究動向について概説する。続いて, 顔の正中線を検出するための提案手法について第 3 章で述べ, 検出した顔正中線の顔部品抽出への応用について第 4 章で述べる。さらに, 第 5 章では, 提案手法の正中線検出と顔部品抽出性能を定量的に評価した実験結果について述べ, 第 6 章にて考察と今後の課題について述べる。

第 2 章

顔画像処理に関する従来研究

顔画像の自動認識は、建物の入退出管理、銀行 ATM などでの個人識別、ロボットと人間との間のインターフェース、ビデオ映像検索クエリとしての利用など、様々な応用分野があり、盛んに研究されている。この研究分野の最終的な目標は、実環境下における安定した顔画像の自動認識であるが、これは計算機にとっていまだ難しい課題であると言わざるを得ない。

ここでは、従来の顔画像処理に関する研究について本研究に関連する顔検出を取り扱うものを中心に紹介し、この分野のこれまでの研究動向について概説する。

本章の内容よりも詳しいサーベイ論文 [1, 2] が発表されているので、それらも参照されたい。

2.1 顔検出と顔認識、顔部品抽出

顔を含んだ画像、または画像列（シーケンス）は、例えば視覚に基づいた知的マンマシンインターフェースなどにおいて基本となる情報である。顔認識や顔の追跡、表情の認識など多数の顔画像処理に関する研究が報告されているが、これらのほとんどは入力画像中から適切に顔領域が切り出されていることを仮定している。

ここではまず、顔画像処理の分野で特に重要な 3 テーマである顔検出、顔認識、顔部品抽出のそれぞれについて概略と相互の関係について説明する。

入力画像に複数の顔が含まれていた場合、それらの顔を発見し適切な大きさで画像から切り出す処理を顔検出と呼ぶ。顔検出の最終的な目標は、画像中の顔のサイズや向き、形状の変動にかかわらず正しく顔を検出することである。

顔認識とは、入力画像（probe と呼ばれる）中の顔が、事前に取得されているデータベース（gallery と呼ばれる）に含まれるどの人物の顔かを判定する処理である。この処理では probe は gallery 内のすべての顔と比較される。多くの顔認識手法では、額からあご、左右

の頬までのいわゆる顔面画像を認識対象としている。これは、頭髪や衣服の変化による画像変動が認識結果に与える悪影響を低減するためである。つまり、顔認識を目的とした顔検出では、余分な領域を含まない顔面領域のみを正しく切り出す必要がある。

顔検出手法の多くは、顔の検出に重点を置いたため、顔の境界線を正確かつ厳密に決定することが困難である。後述する顔特徴に基づく顔検出で利用されているように、顔部品（両目、鼻、口）などを抽出し、それらの位置情報をもとに統計的モデルなどによって顔の境界線を決定するアプローチが提案されている。顔部品抽出はそのための基本的な処理であり、表情認識や動作認識にも応用されている。

2.2 顔検出の代表的な手法

顔が含まれた画像全体から顔のみを含む領域を検出し切り出す処理を、ここでは顔検出と呼ぶ。この処理を顔の同定（Face Localization）として顔検出（Face Detection）とは分けて議論する場合もあるが、本研究ではこれらを区別しない。

顔の検出は顔認識における重要な前処理である。ここでは顔検出技術に関するサーベイ論文である文献 [1] で紹介されている研究のうち、本研究に関連の深いものを紹介する。

自由な環境条件で撮影された画像から顔を検出する問題は、以下に示すような困難さを含んでいる。

1. **画像中の顔の姿勢**。被写体である顔とカメラとの相対的な位置関係の変化により、画像中の顔の見え方は大きく変動する。位置関係によっては顔の特徴の大部分が隠される場合もある。
2. **顔構成の変動**。例えば、メガネや髭などの存在や変化により、顔の見え方は容易に変化する。
3. **表情の変動**。表情が変動することにより、顔の見え方が変化する。
4. **顔の隠れ**。顔が他の物体で隠される場合がある。
5. **画像の向き（カメラの回転）**。カメラが光学軸上で回転することにより顔が回転しているように見える場合がある。
6. **画質**。撮影場所の光源環境やカメラの特性により画像の画質が異なる。

これらの問題は、撮影環境を制御することにより軽減できる場合もある。

2.2.1 事前知識に基づく顔検出

このアプローチでは、顔に関する事前知識に基づく規則（ルール）によって顔検出手法が定義される。例えば、画像中にそれぞれ対称な位置関係で両目がありその中心線上に鼻

と口が存在する，のように顔部品とそれらの位置関係を説明することの容易性が，この方法が採用された理由である．主な手順としては，何らかの処理により顔の候補を複数抽出し，それぞれの候補を顔ルールに従って真に顔であるかどうか判定する．これにより擬陽性（False Positive）検出を減少できる．

これらの手法の問題は，多くの場合に対して一般的に有効なルールの作成が極めて困難であることである．具体的には，特定の撮影環境で撮影された画像に対して有効なルールが開発されたとしても，そのルールが想定しない範囲の変動が画像中の顔に存在した場合，顔の検出は失敗しやすくなる．

このアプローチの代表的な手法として Yang らによる Multiresolution rule-based method[9] がある．

2.2.2 特徴に基づく顔検出

事前知識に基づく手法に対して，顔検出に利用できるような顔の不変な特徴を見つけ出す研究が行われてきた．例えば，画像から顔部品や顔に特有な肌色領域を取り出すことで顔を検出する研究が多く行われてきた．これらの特徴が得られれば，統計的モデルによりそれらの位置関係などを検証することで，対象領域の顔らしさを評価することができる．

これらの手法の問題点は，このような顔特徴の画像での見え方が光源環境に容易に悪影響を受けること，顔特徴の境界が顔そのもののそれよりも弱く表現されることなどである．また，顔特徴の検出が顔そのものを検出する問題と同じ困難さを持っていることに注意する必要がある．

このアプローチの代表的な手法としては，顔部品を用いる手法 [10, 11, 12]，顔テクスチャを解析する手法 [13]，肌の色を利用する手法 [7, 6] などが存在する．

2.2.3 テンプレートマッチングに基づく顔検出

事前に手動で定義またはパラメータ表現された標準的な顔テンプレートと入力画像とのマッチングにより顔を検出する手法である．入力画像が与えられたときには，例えば，顔輪郭，両目，鼻，口などでそれぞれ独立に標準テンプレートとの相関を計算し，相関の値に基づいて顔の有無を決定する．この手法の利点は実装が容易であることである．しかし，この手法は顔の大きさ，姿勢，形状の変動に対して十分に対応できない場合がある．これらの変動に対処するために多重分解能，多重スケールテンプレートの導入や，可変形状テンプレートなどが提案されている．

このアプローチの代表的な手法としては，形状テンプレート [14]，Active Shape Model[15] などがある．

2.2.4 見え方 (appearance) に基づく顔検出

見え方に基づく顔検出手法の，テンプレートマッチング手法との大きな違いは，テンプレートが事前に定義されるのではなく，画像から収集された実例の学習によって得られる点である．基本的にこのアプローチは，顔画像と非顔画像との特徴を的確に記述するために，統計的解析と機械学習の手法を用いている．学習された特徴は分布モデルや判別関数の形式で記述され，顔検出に利用される．計算効率と検出性能の向上のために次元削減が行われる．

このアプローチの代表的な例は，主成分分析を利用して顔の見かけ特徴を抽出する Eigenface[16]，ニューラルネットワークを用いる手法 [17] 等である．

第 3 章

正中線自動検出

この章では、本研究で提案する、正面顔画像から顔正中線を自動検出する手法について述べる。本研究では、人物の顔は左右対称であると仮定しており、顔の線対称軸を正中線として検出する。安定した線対称軸を検出するために、一般化 Hough 変換 (Generalized Hough Transformation : GHT [18][19]) を用いる。GHT は、テンプレートとして与えられた、パラメータ記述されない任意の形状の図形を画像から検出する方法である。

3.1 顔正中線とは

本研究においては、顔正中線を眼間線分 (目と目を結ぶ線分) の垂直二等分線と定義する。図 3.1 は入力画像の一例で、角度 θ だけ傾いた頭部を含んでいる。このときの顔正中線は頭部と同じ傾き θ を持った直線として検出される。画像上から直線を検出する問題

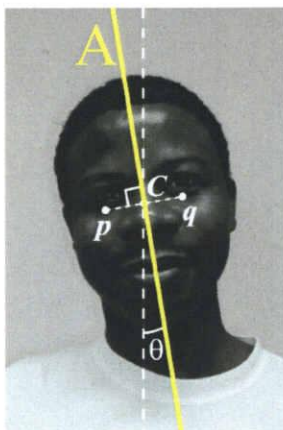


図 3.1: 顔正中線 (線対称軸) の例

Example of facial midline as the symmetry axis

は、その直線が通る一個の点とその直線の傾きとを求めることに等しい。正中線の場合、眼間線分上の一点と顔の傾きとを検出する問題とみなすことができる。図 3.1 の直線 A は、点 $\mathbf{c} = (c_x, c_y)$ と角度 θ とを用いて、以下の式で表される。

$$\frac{x - c_x}{\sin \theta} = \frac{y - c_y}{\cos \theta}. \quad (3.1)$$

提案手法は、正中線を軸として互いに線対称となる 2 点から、2 つのパラメータ \mathbf{c} と θ を決定する。点 $\mathbf{p} = (p_x, p_y)$ と点 $\mathbf{q} = (q_x, q_y)$ とを、互いに線対称とする線対称軸上の点 \mathbf{c} は $\mathbf{c} = \frac{\mathbf{p} + \mathbf{q}}{2}$ と表すことができる。そして、角度 θ は、線分 $(\mathbf{q} - \mathbf{p})$ に直交する角度として得られる。従って、式 (3.1) を、 \mathbf{p} 、 \mathbf{q} を用いて、以下の式で表すことができる。

$$\frac{x - \left(\frac{p_x + q_x}{2}\right)}{\sin \theta} = \frac{y - \left(\frac{p_y + q_y}{2}\right)}{\cos \theta}, \quad (3.2)$$

$$\theta = \tan^{-1} \left(\frac{q_y - p_y}{q_x - p_x} \right). \quad (3.3)$$

正中線検出における課題は、図 3.1 において点 \mathbf{p} 、点 \mathbf{q} として示されている、互いに左右対称となる 2 点を検出することである*¹。

3.2 提案手法の概要

提案手法は、主に図 3.2 に示される 3 つの処理で構成されている。

提案手法の入力画像は、図 3.2(a) に示すような、512 × 768 画素のグレイスケール画像で、1 個の正面顔を含む。画像中の顔の大きさ、回転、傾きは任意とする。

まず第一に、エッジ抽出、2 値化、ノイズ除去を含む前処理を行う。本手法は、正面顔は大局的に左右対称であるという仮定に基づいている。しかし、顔に左右のいずれか片方から照明が当たる場合には、顔の見え方の対称性が損なわれてしまう。このような照明による影響を軽減するために、前処理を行う必要がある。前処理を終えた結果画像は、GHT を適用するために十分な強いエッジ成分を含んでいる。前処理の結果画像の例を図 3.2(b) に示す。

続いて、鏡像テンプレートとのマッチングを GHT によって行う。GHT では、適切な実行結果を得るために、妥当な参照点を決める必要がある。参照点の例を、図 3.2(b) の点 \mathbf{p} に示す。GHT は、参照点と左右対称関係にある点を検出する。この検出された点を、本研究では抽出点と呼ぶ。抽出点の例を、図 3.2(c) の点 \mathbf{q} に示す。

*¹ 実装上は、角度 θ は後述する GHT により検出可能である。対象とする顔画像は離散画像であるので、式 (3.3) を用いて算出した θ より、GHT により検出した θ の方が離散化誤差は小さいと考えられる。

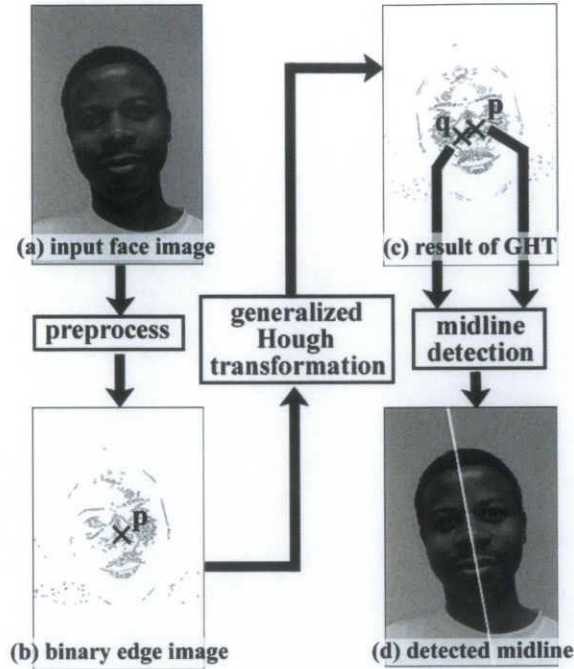


図 3.2: 提案手法における主な3つの処理

Three main stages of the proposed facial midline detection

第3の処理では、互いに線対称関係にある点 p 、点 q の座標を用いて、式 (3.2) より、顔正中線を決定する。

それぞれの処理について、続く節にて述べる。

3.3 前処理

提案手法における前処理部では、入力画像から2値エッジ画像を作成する。第2ステージで実行する GHT のアルゴリズムは2値画像のみを対象とするため、適切な処理結果を得るためにも2値画像の作成は極めて重要な処理である。例えば、2値画像が多くのノイズを含む場合は、GHT の計算コストが増大するだけでなく、結果にも悪影響を与える。一方で、エッジ成分が小さすぎる場合は、GHT の信頼度は低下してしまう。

まず最初に、原画像 $I = I(p)$ の画素 p に対して以下の Sobel オペレータ [20] を適用してエッジ強調を行う。

$$I'(p) = \sqrt{G_x(p)^2 + G_y(p)^2}, \quad (3.4)$$

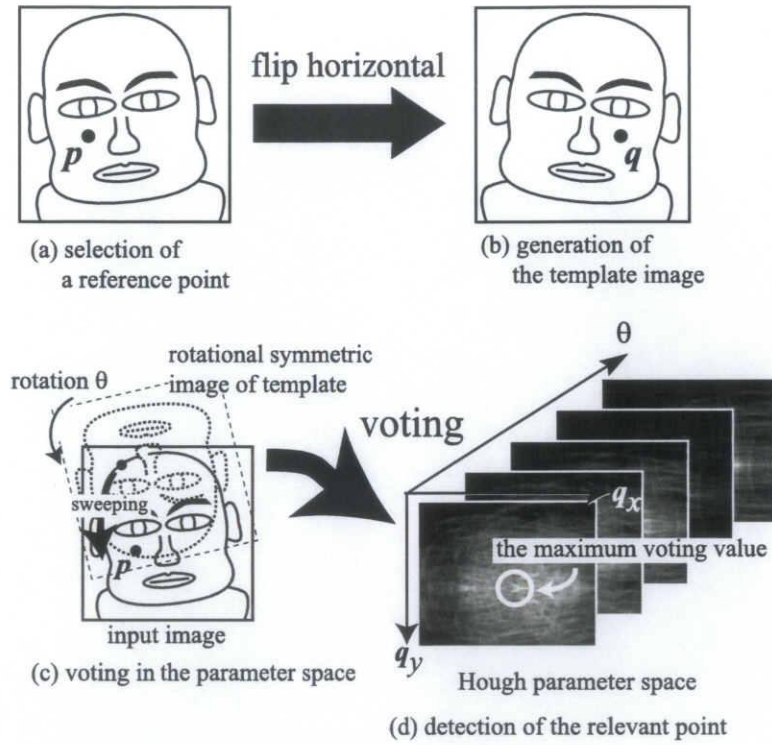


図 3.3: 提案手法における一般化 Hough 変換

The generalized Hough transformation in the proposed method

$$\begin{aligned}
 G_x(p) &= I(p_x + 1, p_y - 1) + 2 \times I(p_x + 1, p_y) + I(p_x + 1, p_y + 1) \\
 &\quad - \{I(p_x - 1, p_y - 1) + 2 \times I(p_x - 1, p_y) + I(p_x - 1, p_y + 1)\}, \\
 G_y(p) &= I(p_x - 1, p_y + 1) + 2 \times I(p_x, p_y + 1) + I(p_x + 1, p_y + 1) \\
 &\quad - \{I(p_x - 1, p_y - 1) + 2 \times I(p_x, p_y - 1) + I(p_x + 1, p_y - 1)\}.
 \end{aligned}$$

次に、エッジ画像を p -タイル法 [21] で 2 値化する。 p -タイル法では、しきい値 T より大きな画素値を持つ画素の数が、全画素数の $p\%$ となるように T を設定する。本手法では、 T より小さい輝度値を持つ画素は白、輝度値が T より大きければ黒として 2 値化する。 Sobel オペレータは、原画像のノイズに基づくエッジも強調してしまうため、適切な 2 値化を行ったとしても、結果の 2 値画像にはいくつかのノイズが含まれる。このノイズを削減するため、その連結成分の周囲長が L_l 画素以下、 L_u 画素以上の成分を削除する。このとき、連結成分の長さは、8-近傍境界追跡 [22] によって取得する。ここでのしきい値 L_l 、 L_u は、実験によって決定する。この実験については、5.3 節にて述べる。境界追跡により、各エッジ成分は輪郭線として表される。

3.4 一般化 Hough 変換 (Generalized Hough Transformation : GHT)

一般化 Hough 変換 (GHT) は図形検出手法で, 2 値の入力画像から, テンプレートとして与えられた図形と同じ, あるいはよく似た図形を検出する. GHT は, 画像中のノイズや図形の欠落に対して頑健であると知られている. 2 値画像を対象とした図形検出という観点から, GHT はテンプレートマッチングの高速化アルゴリズムとして用いられている. 一般に入力画像中の図形は, スケール, 位置, 回転などの幅広い自由度を伴っており, その図形に対して, テンプレートをマッチングさせなければならない. つまり, テンプレートに対して様々な変換を行いながら, 入力画像との適応度を調べる必要がある. この計算コストは膨大である. 計算時間を削減するために, GHT では図形の持つ変動の自由度と等しい次元数のパラメータ空間への投票を行う.

本手法における GHT の目的は, 参照点と左右対称関係にある抽出点を検出することである. 人物の顔は左右対称であるという仮定から, エッジ画像も左右対称であると仮定する. 前処理で得られた 2 値エッジ画像の鏡像画像をテンプレートとして用いる. つまり GHT を用いて, 2 値エッジ画像から, それとよく似た図形であると仮定した鏡像画像を検出する. GHT が図形を検出すれば, 左右対称関係にある 2 点は容易に決定できる.

提案手法における GHT の流れを以下に示す.

- (1) 参照点の決定: GHT では, 参照点を決定する必要がある. 参照点の決め方には任意性があるが, この参照点は後に続く処理である GHT の処理に大きな影響を及ぼすため, 非常に重要であると言える. 佐藤ら [23] は, 参照点としてエッジ画素の重心を用いれば, 信頼性の高い GHT の処理結果が得られることを明らかにした. 本手法ではこのことより, 参照点として 2 値エッジ画像の黒画素 (エッジ画素) の重心を用いる. 重心は以下の式によって得られる.

$$\mathbf{p} = \frac{\sum_{j=0}^N \mathbf{e}_j}{N} = \left(\frac{\sum_{j=0}^N e_{jx}}{N}, \frac{\sum_{j=0}^N e_{jy}}{N} \right). \quad (3.5)$$

ここで, $\mathbf{e}_j = (e_{jx}, e_{jy})$, N はそれぞれエッジ画素の位置とエッジ画素の総数を表す. 参照点の例を, 図 3.3 (a) の点 \mathbf{p} に示す.

- (2) テンプレート画像の生成: 上述のように, 2 値エッジ画像を垂直軸に対して反転させた鏡像画像をテンプレート画像として用いる. 2 値エッジ画像が垂直軸に対して左右対称であれば, 原画像とテンプレートは抽出点 (図 3.3(b)) において, 適切に重なり合うと考えられる.

- (3) **パラメータ空間への投票：** 提案手法における GHT のパラメータ空間は, q_x , q_y , 回転角度 θ の 3 次元からなる. これらは, 図形変化の自由度に相当する. 図 3.3(c), (d) は, 提案手法における投票処理を図示したものである. テンプレートの点対称図形 (図 3.3(b)) は, 2 値エッジ画像上の全てのエッジ画素の上を走査する. 走査処理と同時に, 現在のテンプレートの位置と回転角度に対応するパラメータ空間において, 投票を累積していく.
- (4) **抽出点の検出：** パラメータ空間の中から最大の投票度数を持つ点が, テンプレートの位置と回転角度として検出される.

3.5 GHT の高速化アルゴリズム

上述の処理は, テンプレートを 2 値エッジ画像に適合させるための適切な情報を提供する. しかし, この処理に所要する計算時間は小さくはない. 特に, 3 次元のパラメータ空間への投票処理の所要時間が大きい. この処理時間を削減するために, 本手法では高速化アルゴリズムを考案した. このアルゴリズムは, 高速化のためにパラメータ空間の次元数を削減する. 人物の顔が垂直軸に対して左右対称に表示されていれば, 顔は画像中において直立していると言える. これより, 参照点と抽出点それぞれの垂直方向の位置 p_y, q_y は, 入力画像とテンプレート (鏡像画像) の間で等しい. 入力画像とテンプレートが常に鏡像関係を保つように, お互いを逆方向に回転させると, 入力画像とテンプレート間の q_y の違いを除去することができる. これにより, パラメータ空間の次元数を q_x と θ の 2 次元に削減できる. 図 3.3, 3.4 は, この手法の基本的な概念を図示している. なお, 本手法では, 人物の頭部の傾きの範囲を $\theta \in [-10^\circ, 10^\circ]$ と仮定している.

この高速化のアルゴリズムによって, GHT の計算時間を大幅に削減できた. 予備実験の結果, GHT の所要時間は 10[s] から 0.15[s] へと短縮された (2.6 GHz Intel Core2 processor).

3.6 正中線の決定

GHT により得られた θ, p, q を式 (3.2), (3.3) に代入して正中線を決定する.

$$\frac{x - \left(\frac{p_x + q_x}{2} \right)}{\sin \theta} = \frac{y - \left(\frac{p_y + q_y}{2} \right)}{\cos \theta}. \quad (3.6)$$

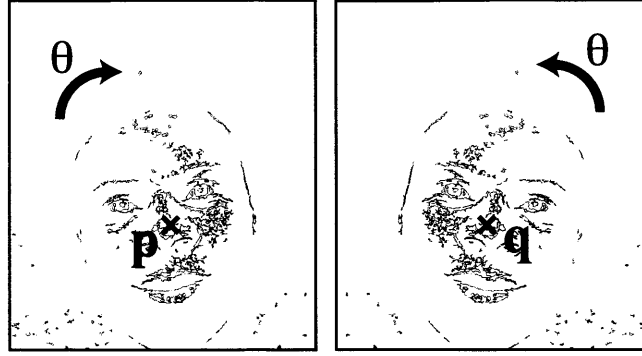


図 3.4: 高速化アルゴリズムの基本的考え方
Basic idea of the fast midline detection

ここで、表記を簡単化するため、以下の変換を行う。

式 (3.6) に $y = 0$ を代入する。

$$\frac{x - \left(\frac{p_x + q_x}{2}\right)}{\sin \theta} = -\frac{\left(\frac{p_y + q_y}{2}\right)}{\cos \theta}. \quad (3.7)$$

式 (3.7) を x について解き

$$x = \left(\frac{p_x + q_x}{2}\right) - \left(\frac{p_y + q_y}{2}\right) \tan \theta, \quad (3.8)$$

を得る。

式 (3.8) は、検出された正中線の $y = 0$ における値 (x 座標) である。式 (3.8) の右辺を x_0 とし、正中線 $L(x_0, \theta)$ を

$$L(x_0, \theta) : \frac{x - x_0}{\sin \theta} = \frac{y}{\cos \theta}, \quad (3.9)$$

と再定義する。

3.7 まとめ

提案手法の流れを図 3.5 にまとめる。提案手法は、グレースケール画像 (a) を入力とし、前処理 (3.3 節) によって 2 値エッジ画像 (b) を得る。次に、2 値エッジ画像 (b) とその鏡像画像であるテンプレート (c) とのマッチングをとるために GHT (3.4, 3.5 節) を行う。GHT の結果として (d) が得られる。この図において、点 p と点 q が線対称であることが GHT により導かれたので、これら 2 点を使って正中線を決定する (e)。

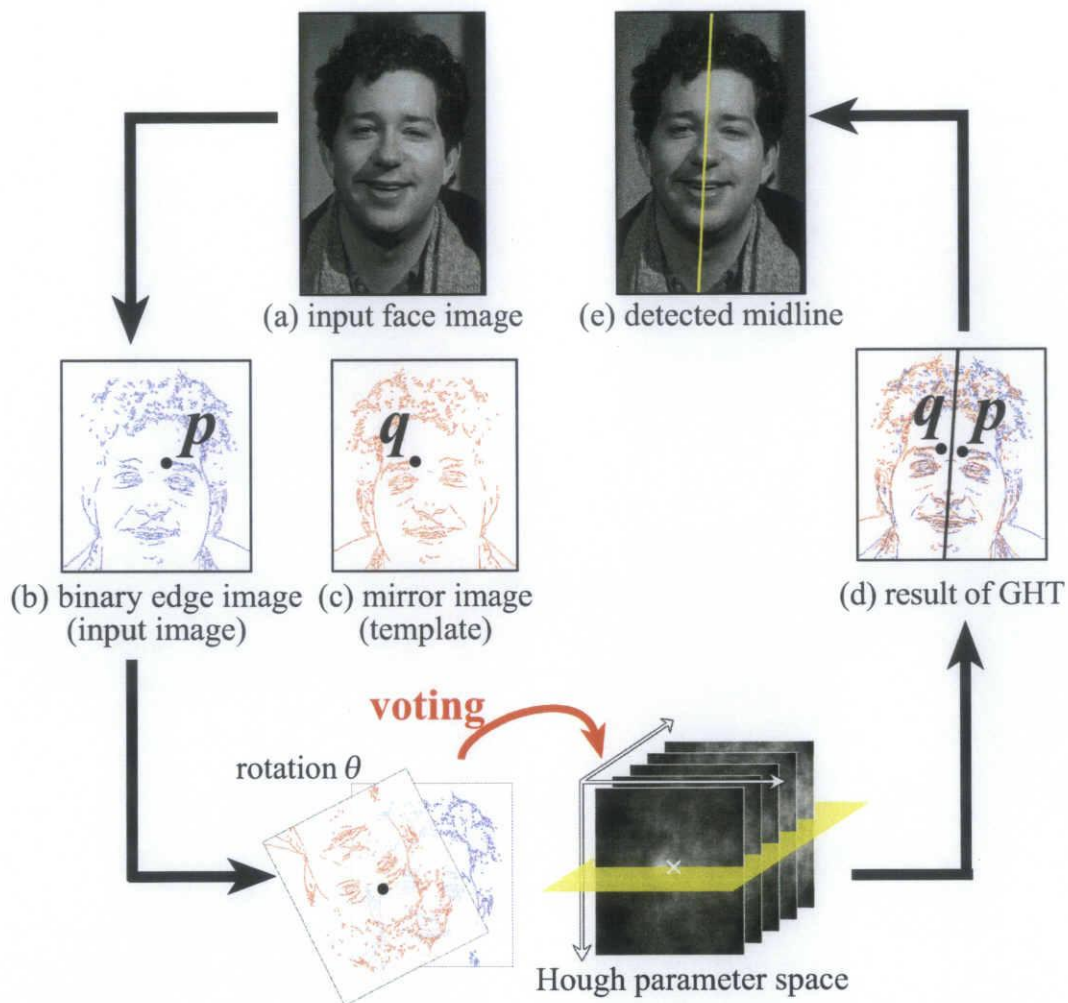


図 3.5: 提案手法における正中線抽出過程の一例

Example of actual facial midline detection in the proposed method

第 4 章

正中線を利用した顔部品抽出

この章では、検出した正中線の顔部品抽出への応用について述べる。第 3 章で述べた手法で検出された正中線を用いて、顔部品抽出に必要な探索を削減する。一般的に顔部品抽出では、画像中の顔の大きさ、傾き、位置などの変化を考慮する必要がある。これに対して、正中線を利用すれば、正中線から顔の傾きと顔の中心軸の位置を得ることができるので、問題をより単純化することができ、顔部品抽出の計算コストを削減できる。

ここでは、一般的に用いられているテンプレートマッチング手法によって、正面顔から両目を探索する場合を考える。図 4.1 は、顔正中線によって顔部品抽出がどのように単純化されるかを示す。正中線を用いずに、目を探索する場合には、図 4.1 (a) のように、主に以下に挙げる 4 つの自由度を考慮する必要がある。

1. x : 水平位置
2. y : 垂直位置
3. θ : 顔の傾き
4. d : 目と目の間の距離

しかし、顔正中線を用いると、 x, θ を事前情報として得ることができるため、図 4.1 (b) のように 2 つの自由度を削減する事ができる。

なお、本研究では顔部品の中でも目を抽出することを目的とする。適切に両目の位置を求めることができれば、その他の顔部品や顔全体の抽出も可能となる。

4.1 顔部品抽出とは

見え方に基づいて顔を認識するには、図 4.2 に示すように、顔検出と顔認識の両方の処理が必要になる。入力画像の人物を正しく認識するには、髪型や衣服、背景の変化が、認識処理に与える影響を除去するため、まず顔領域を検出し、その顔領域のみを顔認識へと

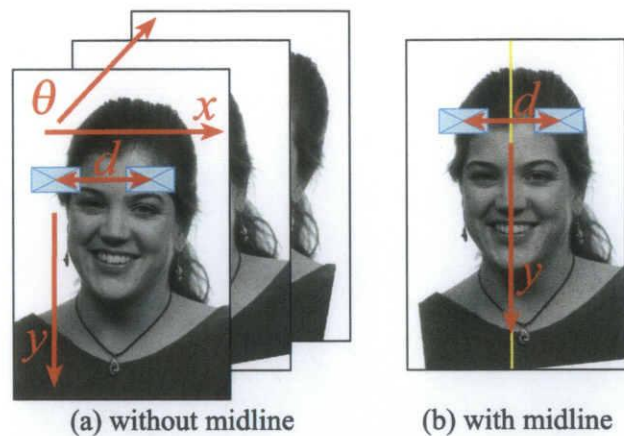


図 4.1: 正中線による顔部品抽出の簡単化

contribution of detected midline to simplification of facial feature extraction

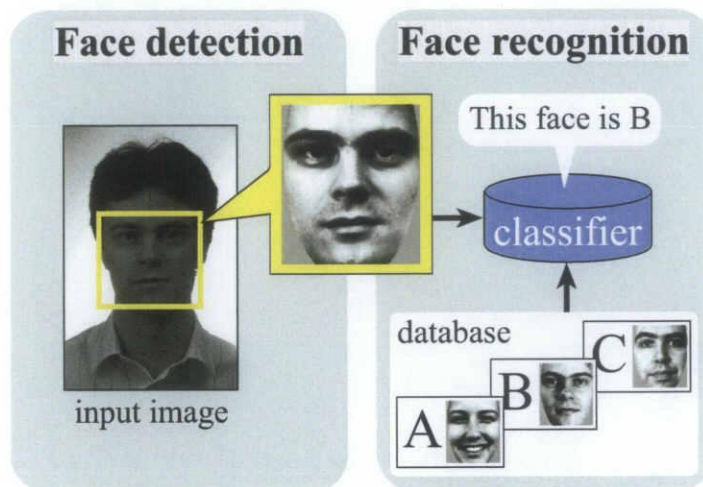


図 4.2: 見え方に基づく自動顔認識システム

Appearance-based automatic face recognition

入力する。顔認識 [2] の精度は顔検出の精度に強く依存するため、安定に顔領域を検出できる手法が求められる。

顔検出手法の従来の方法について、図 4.3 に示す。従来の顔検出手法は、顔全体探索型手法と、顔部品探索型手法の 2 手法に大別できる。前者は、画像中から顔全体を探索する手法で、任意の大きさの顔ウィンドウを用いて、画像の大きさを変えながら画像中を探索する。SVM[24] やニューラルネットワーク [17] などの分類器や距離関数を用いる手法は、この手法に分類される。一方、後者は、顔部品の位置情報を顔検出の手がかりとして利用する手法である。目、鼻、口などの顔部品を抽出することができれば、それらの位置情報を用いて顔全体の領域を求めることができる。また、顔の傾きや向きも知ることができるため、より強力な手法である。

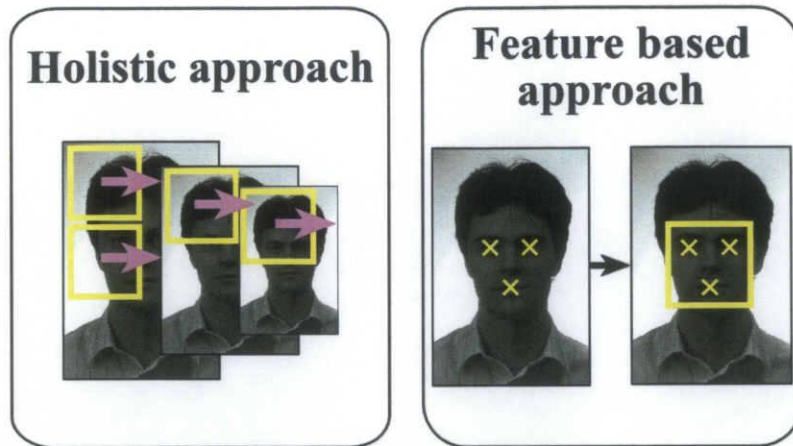


図 4.3: 顔検出手法の従来手法

Conventional approaches for face detection

一般的に、目の位置と両目の距離はあまり変化しないため、目の検出は顔検出の非常に重要な前処理となる。目の位置に基づいて、顔サイズの正規化、顔位置の決定、顔画像平面内の回転の補正などが行われる。本研究では、安定した両目の抽出を行うことを目的とする。しかし、目の抽出は多次元の探索を必要とするため、高い計算コストが必要となる。提案手法では、この計算コストを削減するために顔正中線を応用する。

4.2 テンプレートマッチングによる両目の抽出

対象物のパターンがテンプレートとして表現されるとき、これを探索範囲内で移動させて、最もよく合った領域を探す操作をテンプレートマッチングと呼ぶ。テンプレートマッチングは、画像処理におけるマッチング手法で最も単純で実装が容易である。本研究では、目に外接する領域をテンプレートとして用い、テンプレートマッチングによって、最も目らしい位置を抽出する。

4.2.1 問題の定義

本研究における両目の抽出では、テンプレートマッチング法を用いて、図 4.4 中の $w \times h$ で表される目領域の目らしさを評価し、両目の位置 p_r , p_l を抽出する。

テンプレートマッチング法では、事前に何らかの方法で生成されたテンプレートと、画像中の任意の領域との類似度を、何らかの方法で評価し、その類似度が最大となる領域を目的の領域として抽出する。これを本研究で取り扱う両目の抽出に当てはめると、テンプレートとして目を切り出した領域を用い、画像中からそのテンプレートとの類似度が最も最大となる領域を目領域として抽出する。

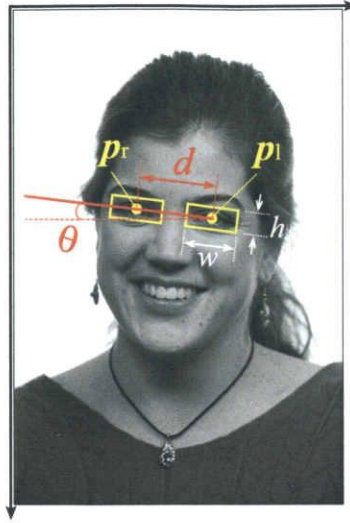


図 4.4: 本研究における両目の抽出

Definition of eyes extraction in the proposed method

両目の画像中の大きさを抽出する必要がある場合は, w, h もそれぞれ自由なパラメータとする必要がある. 本研究では, 両目の位置 p_r, p_l のみを求めるとし, 両目の大きさは p_r, p_l から以下の線形モデルで算出できるとする.

$$d = |p_r - p_l| = \sqrt{(p_{rx} - p_{lx})^2 + (p_{ry} - p_{ly})^2}, \quad (4.1)$$

$$\begin{pmatrix} w \\ h \end{pmatrix} = \begin{pmatrix} 0.51 \\ 0.22 \end{pmatrix} d + \begin{pmatrix} 6.04 \\ 3.15 \end{pmatrix}. \quad (4.2)$$

このモデルの導出と妥当性の検証については付録 A を参照されたい.

このようにして求められた w と h を用いて目領域を以下のように定義する. まず右目領域 S_r は,

$$S_r = \left\{ p_r + iu + jv \mid i = \left[-\frac{w}{2}, \frac{w}{2} \right], j = \left[-\frac{h}{2}, \frac{h}{2} \right] \right\}, \quad (4.3)$$

$$u = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}, v = \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}, \quad (4.4)$$

$$\theta = \tan^{-1} \frac{(p_{ry} - p_{ly})}{(p_{rx} - p_{lx})}. \quad (4.5)$$

続いて左目領域 S_l は,

$$S_l = \left\{ p_l + iu + jv \mid i = \left[-\frac{w}{2}, \frac{w}{2} \right], j = \left[-\frac{h}{2}, \frac{h}{2} \right] \right\}, \quad (4.6)$$

となる.

ここで取り扱う両目の抽出においては, 左右の目を独立に抽出すると, 先に述べた線形モデルによる目領域の決定が困難になるため, p_r, p_l を組として抽出する. また, 正中線

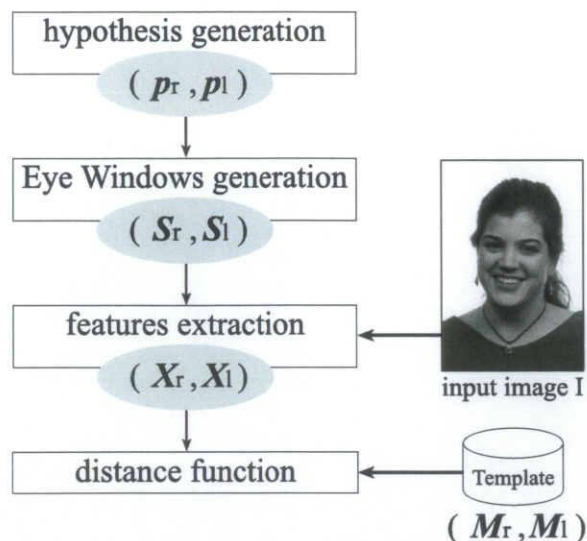


図 4.5: テンプレートマッチングにおける 1 サイクル

Proposed facial feature extraction

に対して両目の位置が線対称であるという仮定を用い、右目と左目を対で探索することで目の抽出精度向上を目指す。さらに検出された正中線が誤っていた場合には両目の対称性が崩れるため、棄却処理を導くことができる。取り扱う問題は、画像からテンプレートとの類似性が最大となる p_r, p_l の組を求める問題となる。

テンプレートマッチング法では、この最大の類似性を持つ領域を発見するために、図 4.5 に示すような処理を 1 サイクルとした繰り返し処理を行う。1 サイクルの処理は、以下の通りである。

- (1) 仮定の生成： 任意の p_r, p_l の組を生成する。
- (2) 領域の生成： p_r, p_l より、領域 S_r, S_l を線形モデル (4.2) で生成する。
- (3) 特徴ベクトルの抽出と変換： (2) で生成された領域内の画素の見え方を記述する特徴ベクトルを抽出する。

$$X_r = \text{GFV}(I, S_r), \quad (4.7)$$

$$X_l = \text{GFV}(I, S_l). \quad (4.8)$$

- (4) 距離関数による類似性の評価： 特徴ベクトルとテンプレートの類似性を距離関数により評価する。

$$S = D(X_r, M_r) + D(X_l, M_l). \quad (4.9)$$

4.2.2 特徴抽出と特徴変換

本研究では、目領域内の画像の見え方を濃度勾配特徴を用いて表現する。濃度勾配特徴は、手書き文字認識において高い性能を有する特徴量である [25]。本手法では、この濃度勾配特徴をグレイスケールの顔画像から抽出する。

まず始めに、次の流れで濃度勾配の強度と向きを取得する。

1. 平滑化：入力画像 $I = I(\mathbf{p})$ の画素 \mathbf{p} に対して以下のガウシアンフィルタを施す。

$$I'(\mathbf{p}) = \frac{\sum_{i=-N/2}^{N/2} \sum_{j=-N/2}^{N/2} f(i, j) \bullet I(p_x + i, p_y + j)}{\sum_{i=-N/2}^{N/2} \sum_{j=-N/2}^{N/2} f(i, j)}, \quad (4.10)$$

$$f(i, j) = \exp\left(-\frac{i^2 + j^2}{2\sigma^2}\right). \quad (4.11)$$

ここでは、カーネルサイズ N は 5, $\sigma = 1.5$ とした。

2. 濃度勾配の抽出：カーネルサイズが 3×3 の Sobel オペレータを用いて、勾配の強度と向きを求める。ある画素 \mathbf{p} における、勾配の強度 $|G(\mathbf{p})|$ 、勾配の向き $\theta(\mathbf{p})$ は以下で得る。

$$|G(\mathbf{p})| = \sqrt{G_x(\mathbf{p})^2 + G_y(\mathbf{p})^2}, \quad (4.12)$$

$$\theta(\mathbf{p}) = \tan^{-1}\left(\frac{G_y(\mathbf{p})}{G_x(\mathbf{p})}\right). \quad (4.13)$$

なお、 $G_x(\mathbf{p}), G_y(\mathbf{p})$ はそれぞれ勾配の水平、垂直成分である。また、勾配の向きの値域は $[-\pi, \pi]$ である。

以上により抽出した濃度勾配画像の例を図 4.6 に示す。図は、勾配の強度を明度で表し、勾配の向きを色相によって表している。目、鼻、口などの顔部品に強く特徴的な勾配分布があることが見てとれる。

図 4.6 に示した各画素ごとの勾配の強度、向きを、そのまま特徴ベクトルとして用いることは、その次元数が大きすぎるため現実的ではない。本手法では以下の処理によって、次元数の削減を行い濃度勾配特徴ベクトルを抽出している。対象領域に対して、図 4.7(a) のように $n_x \times n_y$ の小領域に分割する。さらに、図 4.7(b) のように量子化数 n_d により向きごとの量子化を行う。続いて、 $n_x \times n_y$ 個の小領域それぞれの内部で、 n_d 次元の向きごとに、濃度勾配強度を積算していく。この処理によって、 $n_x \times n_y \times n_d$ 次元の原特徴ベクトルを得る。

次に原特徴ベクトルに対して、以下の特徴変換を行う。

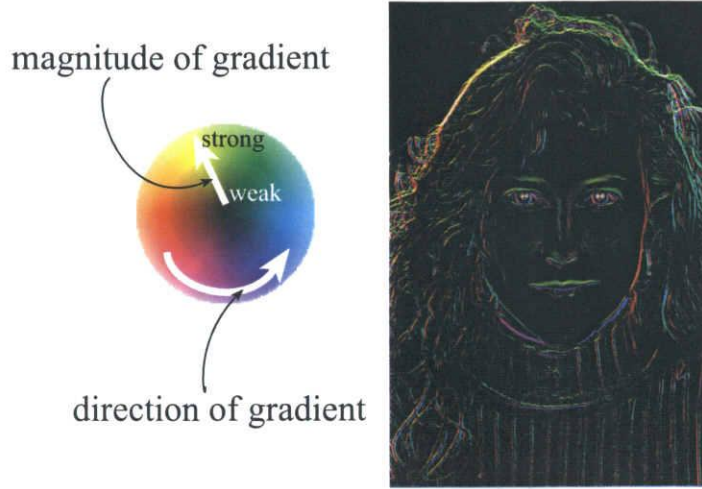


図 4.6: 濃度勾配の画像例
Example of gradient features

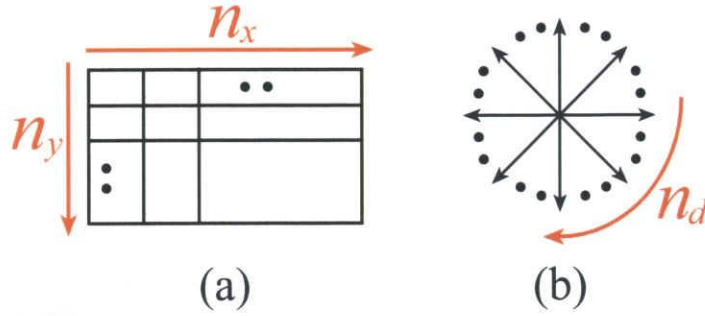


図 4.7: 濃度勾配特徴抽出における小領域分割と方向量子化
sub-block segmentation and direction quantization in gradient feature vector extraction

1. 大きさの正規化：以下の式によりベクトルの大きさの正規化を行う。なお、 v_i は原特徴ベクトルの第 i 要素、 a は重み付けの係数を表している。また、 w, h はそれぞれ目領域の幅と高さを表している。

$$v'_i = \frac{a \times v_i}{w \times h}. \quad (4.14)$$

2. 次元削減：ガウシアンフィルタを用いて、空間と向きに対する平滑化を行い、次元を削減する。
3. べき変換：指数 γ を用いて、以下の式でべき変換を行う。

$$x_i = (v'_i)^\gamma. \quad (4.15)$$

なお、特徴抽出には様々なパラメータが存在するが、これらはパラメータ決定実験によって決定した。この実験については 5.3.2 節にて述べる。以上の処理で得られる勾配特

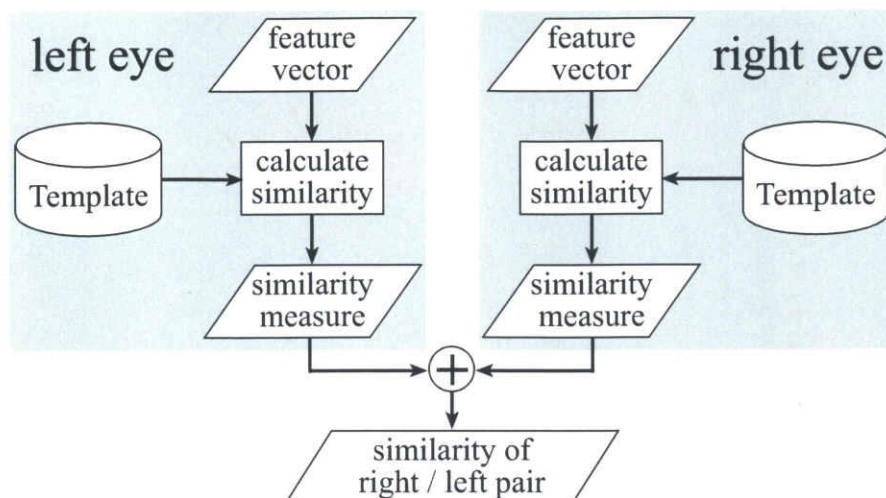


図 4.8: 類似性の評価
evaluation of similarity

特徴ベクトルは 400 次元のベクトルであり，原画像の画素値をそのまま用いる場合に比較して大幅に次元数が削減される．

4.2.3 距離関数

抽出された特徴ベクトルとテンプレートとの類似性を，距離関数により評価する．処理の流れを図 4.8 に示す．

抽出された特徴ベクトルを評価対象のデータとして，テンプレートの特徴ベクトルとの類似性を計算する．提案手法では，類似性の尺度としてコサイン係数（cosine coefficient）を用いている．なお，コサイン係数は以下で求められる．

$$D(\mathbf{X}, \mathbf{M}) = \frac{\mathbf{X}^T \bullet \mathbf{M}}{\|\mathbf{X}\| \bullet \|\mathbf{M}\|}, \quad (4.16)$$

$$\mathbf{X} = (x_1, x_2, \dots, x_{400})^T. \quad (4.17)$$

ここで， \mathbf{X} は 4.2.2 節の方法で取得したテストデータ， \mathbf{M} はテンプレートで，本研究では学習データの標本平均ベクトルである．

図 4.8 に示されているように，左目と右目それぞれ独立して，コサイン係数により類似性を評価する．得られた左右の類似性を足しあわせ，その領域の目らしさとする．

4.2.4 テンプレートマッチング法の問題点

テンプレートマッチング法を用いて画像中から \mathbf{p}_r , \mathbf{p}_l の 2 点を抽出することの問題点は，取りうる \mathbf{p}_r , \mathbf{p}_l の全ての組合せにおいて類似性の評価を行う必要があることである．

これは、画像のサイズを N とすると、 $O(N^2)$ の計算量であり、膨大な計算時間を要求する。提案手法は、 $\mathbf{p}_r, \mathbf{p}_l$ の抽出の高速化のために、顔正中線を応用する。

4.3 顔正中線を利用した高速化

テンプレートマッチング法を用いて両目を抽出する場合、 $\mathbf{p}_r, \mathbf{p}_l$ の組合せ数は膨大になり、高い計算コストを要求する。提案手法は、検出された正中線を用いて、探索すべき $\mathbf{p}_r, \mathbf{p}_l$ の組の数を削減する。

4.3.1 正中線による探索自由度の削減

本研究における両目の抽出では、画像中の顔の位置、大きさ、傾きは任意である。テンプレートマッチングでは $\mathbf{p}_r, \mathbf{p}_l$ の組合せを画像全体に対して生成して評価する必要がある。従って、 $\mathbf{p}_r = (p_{rx}, p_{ry})$, $\mathbf{p}_l = (p_{lx}, p_{ly})$ となり 4 自由度が存在する。

正中線 $L(x_0, \theta_0)$ が垂直になるように画像を回転させると、式 (4.5) の θ は 0 になる。すなわち、 $p_{ry} = p_{ly}$ として探索することができる。また、両目は正中線に対し線対称な位置に存在すると仮定すると、

$$p_{rx} = x_0 - d/2, \quad (4.18)$$

$$p_{lx} = x_0 + d/2, \quad (4.19)$$

となり、最終的に探索する変数の数は 2 となる。

この条件

$$\mathbf{p}_r = (p_{rx}, p_{ry}), \mathbf{p}_l = (p_{lx}, p_{ly}), \quad (4.20)$$

$$p_{rx} = x_0 - d/2, p_{lx} = x_0 + d/2, \quad (4.21)$$

$$p_{ry} = p_{ly} = p_y, \quad (4.22)$$

$$\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (4.23)$$

のもとで、目領域 (4.3), (4.6) を定義し直すと以下の通りになる。

$$\mathbf{S}_r = \left\{ \left(x_0 - \frac{d}{2} + i, p_y + j \right) \middle| i = \left[-\frac{w}{2}, \frac{w}{2} \right], j = \left[-\frac{h}{2}, \frac{h}{2} \right] \right\}, \quad (4.24)$$

$$\mathbf{S}_l = \left\{ \left(x_0 + \frac{d}{2} + i, p_y + j \right) \middle| i = \left[-\frac{w}{2}, \frac{w}{2} \right], j = \left[-\frac{h}{2}, \frac{h}{2} \right] \right\}. \quad (4.25)$$

目領域の例を図 4.9 に示す。

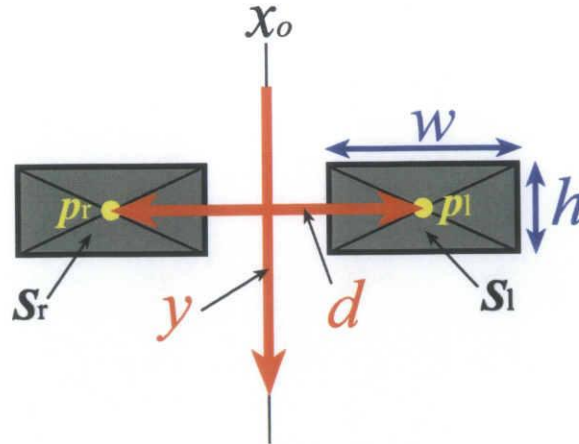


図 4.9: 目領域の例

Example of Eye Windows

4.3.2 エッジ強度の2値化による探索範囲の削減

画像中で、顔部品を含む領域は、額や頬、背景などの領域に比べ、特徴的な（一般的に強い）勾配強度を持つと考えられる．これを利用して、目領域内の濃度勾配強度の平均がしきい値 T 以下であれば、その領域は顔部品を含まないとみなし探索の対象から外す．なお、しきい値 T は以下のように求める．ここで、 R は探索範囲を表し、 $A(R)$ 、 $A(S)$ はそれぞれ領域 R 、 S の面積である．

$$T = \frac{\sum_{p \in R_r} G_{mr}(p) + \sum_{q \in R_l} G_{ml}(q)}{A(R_r) + A(R_l)}, \quad (4.26)$$

$$G_{mr}(p) = \frac{\sum_{p \in S_r(d,y)} |G(p)|}{A(S_r(d,y))}, \quad G_{ml}(q) = \frac{\sum_{q \in S_l(d,y)} |G(q)|}{A(S_l(d,y))}, \quad (4.27)$$

$$R_r = \left\{ \left(x_0 - \frac{d}{2}, y \right) \mid d = [d_{\min}, d_{\max}], y = [y_{\min}, y_{\max}] \right\}, \quad (4.28)$$

$$R_l = \left\{ \left(x_0 + \frac{d}{2}, y \right) \mid d = [d_{\min}, d_{\max}], y = [y_{\min}, y_{\max}] \right\}. \quad (4.29)$$

しきい値処理によって探索範囲を絞り込んだ例を図 4.10 に示す．図 4.10 (a), (b) それぞれの画像中に赤色で示されている領域は、しきい値処理前後の探索範囲を表している．図 4.10 (a) は、4.3.1 節の処理において、2つのパラメータ d, y を設定すべき領域で、顔全体を対象とする．これに対して、しきい値処理を行った結果が図 4.10 (b) である．勾配が弱い額や頬の領域が取り除かれている．図 4.10 (c) は、濃度勾配強度平均のヒストグラムである．このヒストグラムに対して平均値 T によるしきい値処理をした結果を、青と赤の2クラスで示している．赤で示されている濃度勾配強度平均を持つ領域だけが探索の

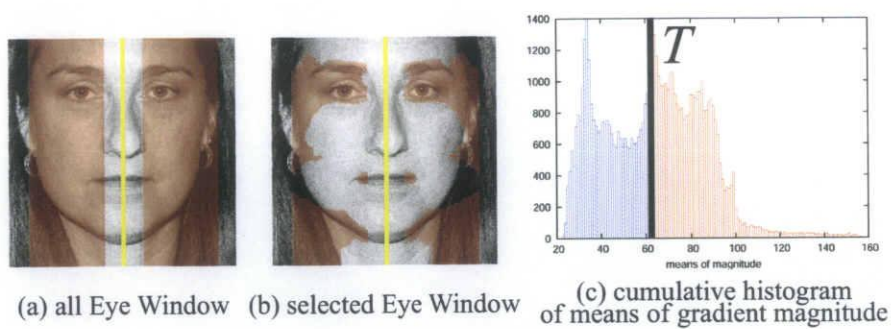


図 4.10: 探索対象となる目領域の例

Results of search area selection

対象となり，図 4.10 (b) のように選択される．

第 5 章

評価実験

5.1 FERET データベース

The Facial Recognition Technology (FERET) データベース [5] は、顔認識アルゴリズムの評価を目的に収集されたデータベースである。FERET データベースは、1000 人を超える個人の顔を収集した大規模なデータベースである点が特徴で、顔認識手法の評価では標準的に用いられている。本来、FERET データベースは顔認識アルゴリズムの評価用に収集されたデータベースであるので、顔検出アルゴリズムの評価には適さないと言われている [1]。これは、データベース内の全ての画像が、比較的良好な光源環境下で撮影され、無地または簡素な背景と 1 名の人物の顔のみを含んでいるため、顔検出として容易に取り扱えらると思われているためである。しかし、顔部品抽出の観点からは、FERET データベース中の画像は、決して簡単な対象とはいえない。その理由は、データベースに含まれる 1 人物の画像数が 2 から 5 画像程度と少ないこと、表情の変化や眼鏡の着脱を伴うため同一人物でも顔部品の見かけが大きく変化することなどである。

本研究で行う評価実験は、FERET データベースに含まれる、正面顔を撮影した **fa**、**fb** グループを対象とする。**fa** グループは顔認識において主に学習用、参照用に用いられるグループである。**fb** グループは **fa** グループと同日に撮影された画像からなるが、**fa** グループの画像とは表情が異なる。**fa**、**fb** グループはそれぞれ 1207 画像、1202 画像から構成される。すなわち、本研究では合計で 2409 画像、867 人の人物から構成されるデータベースを用いる。なお、各画像には事前に目視により抽出された顔部品の座標が校正用（正解）情報として与えられている。

本研究では、4.2.3 節で述べたテンプレート生成のための学習データと、付録 A に述べる目領域線形モデルの設計のために、**fa** グループから無作為に取り出された 200 枚の学習データを用いる。さらに、各パラメータを決定する（5.3 節）ための予備実験用に、400 枚の画像を **fa**、**fb** グループから無作為に取り出した設計用データを用いた。

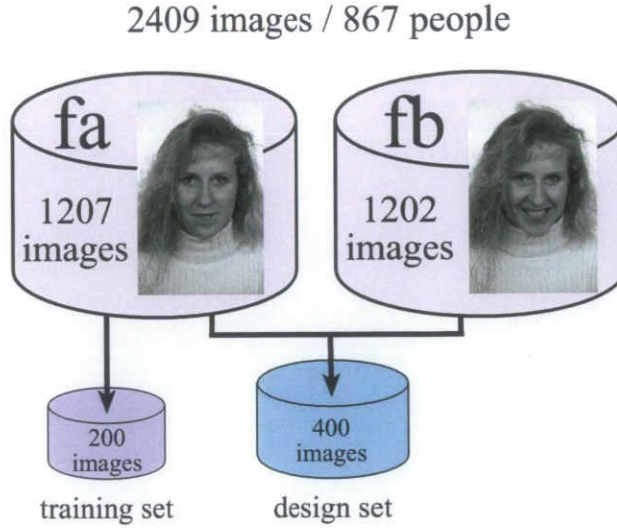


図 5.1: FERET データベース
FERET database

各データの構成を図 5.1 に示す。

5.2 定量的評価のための指標

5.2.1 正中線検出の正しさの指標

正中線の検出精度を定量的に評価する。精度評価実験では、検出した正中線と、真の正中線を比較する。ここで、真の正中線とは、事前に目視抽出されている両目の座標を用いて定めた正中線である。目視抽出された両目の座標 (c_{rx}, c_{ry}) , (c_{lx}, c_{ly}) を式 (3.2), (3.3) に代入し、以下を真の正中線 L_t とする。

$$\frac{x - \left(\frac{c_{rx} + c_{lx}}{2} \right)}{\sin \theta} = \frac{y - \left(\frac{c_{ry} + c_{ly}}{2} \right)}{\cos \theta}, \quad (5.1)$$

$$\theta = \tan^{-1} \left(\frac{c_{ry} - c_{ly}}{c_{rx} - c_{lx}} \right). \quad (5.2)$$

文献 [6] では、正中線の検出精度を評価するために、2つの指標、角度誤差 $\Delta\theta$ と位置誤差 s を用いている。角度誤差 $\Delta\theta$ は、検出した正中線と真の正中線との間の角度の誤差である。また、位置誤差 s は、眼間線分上における2つの正中線間の距離を表したものである。図 5.2 は、この2つの指標を図示したものである。

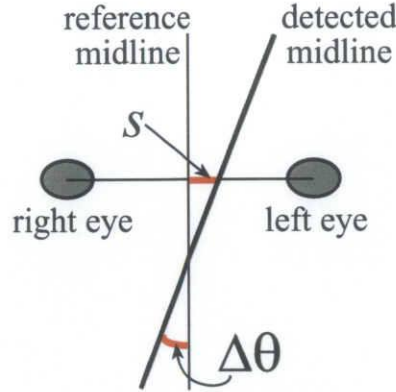


図 5.2: 精度評価のための角度誤差と位置誤差
angle and distance errors for evaluation of detected midline

5.2.2 顔部品抽出の正しさの指標

目の抽出精度を定量的に評価する．この精度評価方法は，正解データとして与えられている手動で取得された真の目の座標 C_r, C_l と，検出した目の座標 p_r, p_l を比較する．評価方法として，文献 [11][12] で用いられている評価指標 d_{eye} を用いる．評価指標の式を以下に表す．次式の C_l, C_r はそれぞれ左目，右目の座標を表しており， $\|C_l - C_r\|$ は眼間の距離である．

$$d_{eye} = \frac{\max(d_l, d_r)}{\|C_l - C_r\|}, \quad (5.3)$$

$$d_l = \|C_l - p_l\|, d_r = \|C_r - p_r\|. \quad (5.4)$$

式 (5.3) は，検出した目の座標と真の目の座標との距離の，左右における最大値を検出誤差とし，その誤差を，眼間の距離で正規化したものになる．文献 [11] では， $d_{eye} < 0.05$ のときを抽出成功として定義している．これを図示すると図 5.3 のようになる．図中の左右の目の上にある黄色の円の内部が正解となり，左目，右目の両方が，この円内の精度で検出されなければならない．

5.3 パラメータ決定のための予備実験

提案手法にはあらかじめ決定しておくべきパラメータがある．これらのパラメータを設定するために，400 枚の設計用データ（5.1 節を参照）を用いた，パラメータ決定予備実験を行った．

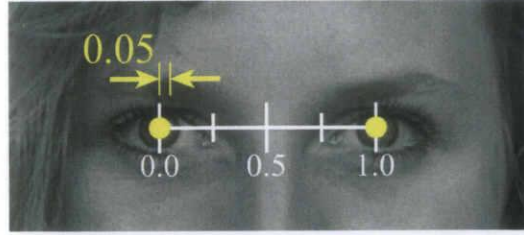


図 5.3: 目抽出の正しさの尺度
criterion of correctly extracted eyes

5.3.1 正中線検出におけるパラメータ決定

第 3 章で述べた正中線検出で必要となるパラメータは，前処理（3.3 節）における p -タイル法で用いるエッジ画素の割合 p とノイズ除去で用いる周囲長のしきい値 L_l ， L_u の 3 つである．GHT に基づく正中線検出手法を，設計用データを入力として，次のパラメータ，

$$p \in \{5, 6, 7, 8, \dots, 20\} [\%], \quad (5.5)$$

$$L_l \in \{5, 10, 15, \dots, 30\} [\text{pixels}], \quad (5.6)$$

$$L_u \in \{500, 600, 700, \dots, 1500\} [\text{pixels}], \quad (5.7)$$

の組み合わせで実行した．それぞれのパラメータの組み合わせにおいて，角度誤差と位置誤差（5.2.1 節を参照）に基づいて検出精度の評価を行った．評価実験の結果， $(p, L_l, L_u) = (7, 15, 1400)$ の組み合わせが最も高い検出成功率 84.5% を与えたので，この組み合わせをパラメータとして決定する．

5.3.2 顔部品抽出におけるパラメータ決定

ここでは，第 4 章で述べた顔部品抽出で用いるパラメータについて述べる．ここで決定すべきパラメータは，探索範囲を定義するために必要なパラメータ $d_{\min}, d_{\max}, y_{\min}, y_{\max}$ と，濃度勾配特徴を抽出するときに用いるパラメータである．

まず， $d_{\min}, d_{\max}, y_{\min}, y_{\max}$ について議論する．探索すべき領域の大きさは $(d_{\max} - d_{\min}) \times (y_{\max} - y_{\min})$ となり，さらに，右目と左目のそれぞれでテンプレートとの類似性を評価するので，評価の対象となる領域の大きさは $2 \times (d_{\max} - d_{\min}) \times (y_{\max} - y_{\min})$ となる．この数が増えれば，目の抽出に要する処理時間も増加するので，探索範囲はできるだけ小さく，かつ必ず目を含むように選択される必要がある．図 5.4 は，設計用データにおける眼間線分の長さをヒストグラム表示したグラフである．この結果から，眼間距離

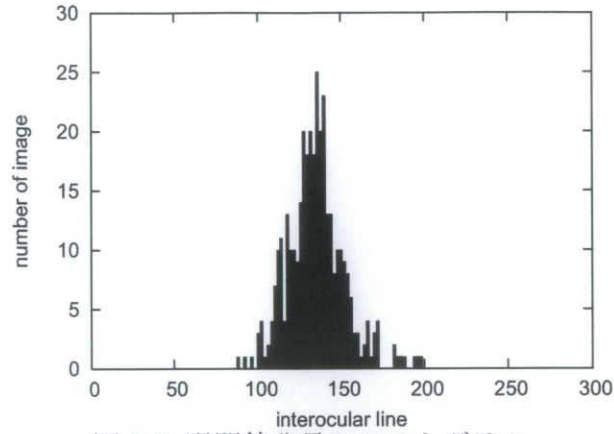


図 5.4: 眼間線分長のヒストグラム

Histogram of interocular line length

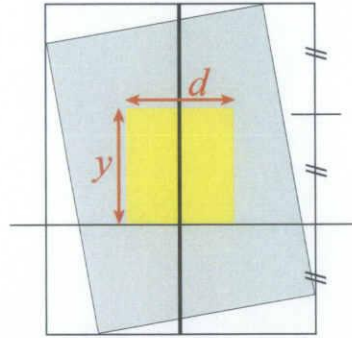


図 5.5: 探索領域

initial searching region

d の範囲 $[d_{\min}, d_{\max}]$ は $[50, 250]$ とした. なお, 実装時には, 正中線を軸として左右方向に 1 画素ずつ変化するので, 以下に示すように d は 2 画素刻みになる.

$$d = \{2 \times d' \mid 25 \leq d' \leq 125\}. \quad (5.8)$$

また, 垂直方向の探索範囲 $(y_{\max} - y_{\min})$ は 300 画素とした. 図 5.5 に探索領域の例を示す. 探索領域は, その下端 y_{\max} を回転補正後の画像縦サイズの $2/3$ に設定した.

続いて, 特徴量抽出に関するパラメータについて述べる. まず, 原特徴ベクトルを抽出するときの $n_x \times n_y$ と n_d と, 特徴変換を行うときの, 空間, 方向の次元削減の有無, ベキ変換の指数 γ を決定する.

これらを決定するために, 目の抽出手法を設計用データに対して実行し, その抽出精度を比較した. なお, ここでは正中線検出は理想的に行われたと仮定し, 式 (5.1) で表される真の正中線 L_t を用いた.

特徴ベクトルとして, 表 5.1 に示されている 4 種類の特徴ベクトルを用いた. なお, 表中の次元削減とは, ガウシアンフィルタによって特徴量の間引きをする処理で, $\frac{n_x+1}{2} \times \frac{n_y+1}{2}$,

表 5.1: 評価対象とした特徴ベクトル抽出の条件

evaluated conditions for feature vector extraction

	特徴ベクトル		原特徴ベクトル			
	$n_x \times n_y \times n_d$	次元数	$n_x \times n_y$	次元削減	n_d	次元削減
パターン 1	$3 \times 3 \times 16$	144	5×5	○	32	○
パターン 2	$5 \times 5 \times 8$	200	9×9	○	16	○
パターン 3	$5 \times 5 \times 16$	400	9×9	○	32	○
パターン 4	$5 \times 5 \times 16$	400	5×5	×	32	○

あるいは $\frac{n_d+1}{2}$ へと次元削減される。また、4つのパターンそれぞれにおいて、べき変換の指数 γ を以下のパターンで組み合わせ実行した。

$$\gamma = \{0.3, 0.5, 0.8, 1.0\}. \quad (5.9)$$

その結果、設計用データにおいて最も高い抽出成功率 93.5% を与えたのはパターン 4 の特徴ベクトルで、べき変換の指数は $\gamma = 1.0$ のときであった。よって、小領域分割数 $n_x \times n_y$ は 5×5 、方向量子化数 n_d は 32 で、次元削減は方向のみ行い、 $5 \times 5 \times 16$ の 400 次元の特徴量を採用する。また、べき変換の指数は $\gamma = 1.0$ なので、べき変換を行わないことと等価になり、べき変換の処理を取り除く。

最後に、式 (4.14) で述べた、重み付け係数 a であるが、本手法では、べき変換を行わずに、評価関数としてコサイン係数を用いているため、評価される類似性は理論上 a に依存しない。しかし、ベクトルの大きさが小さすぎると計算誤差を生じる可能性があるため、 $a = 4000$ とする。これは、目領域の平均面積が約 4000 [画素] であるためである。

ちなみに、評価関数としてユークリッド距離を用いた場合でも同様の予備実験を行ったが、抽出成功率は最高でも 72.75% であり、全ての条件においてコサイン係数の抽出成功率より低かった。

5.4 性能評価実験

5.4.1 顔正中線

提案手法の有用性を実証するため、FERET データベースの画像を用いて評価実験を行った。入力画像は、5.1 節で述べた 2409 枚の正面顔画像を用いている。

3 節で述べた顔正中線検出手法を実験した。検出した正中線の例を図 5.6 に示す。各画像中の黄色の線は、検出された正中線である。提案手法は、様々なサイズや頭部の回転角度においても、正しく正中線を検出することができる。



図 5.6: 提案手法による正中線検出結果

Result of the midline detection

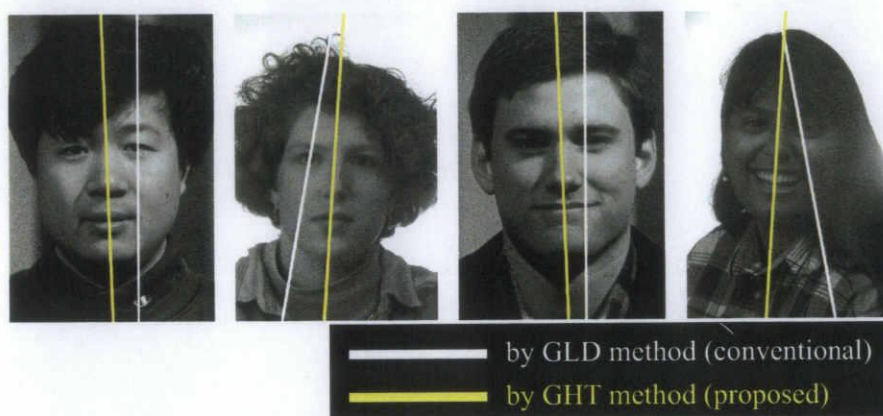


図 5.7: 正中線検出の提案手法と従来手法の比較結果例

visual comparison of extracted midlines

次に、図 5.7 は、従来手法である GLD に基づいて正中線を検出した結果と、提案手法の GHT に基づいて正中線を検出した結果との比較である。図中の白色と黄色の線はそれぞれ、GLD と GHT による結果である。GLD の結果では、顔に対して非対称な照明環境によって正しく正中線を検出できていない。一方で GHT による結果では、このような条件においても正しく検出されている。これらの例では、顔の左右で明るさが異なっており、GLD による手法は、このような照明環境の非対称性に影響を受けやすいことが示された。このため、大域的な線対称性ではなく、局所的な線対称性を検出している。これに

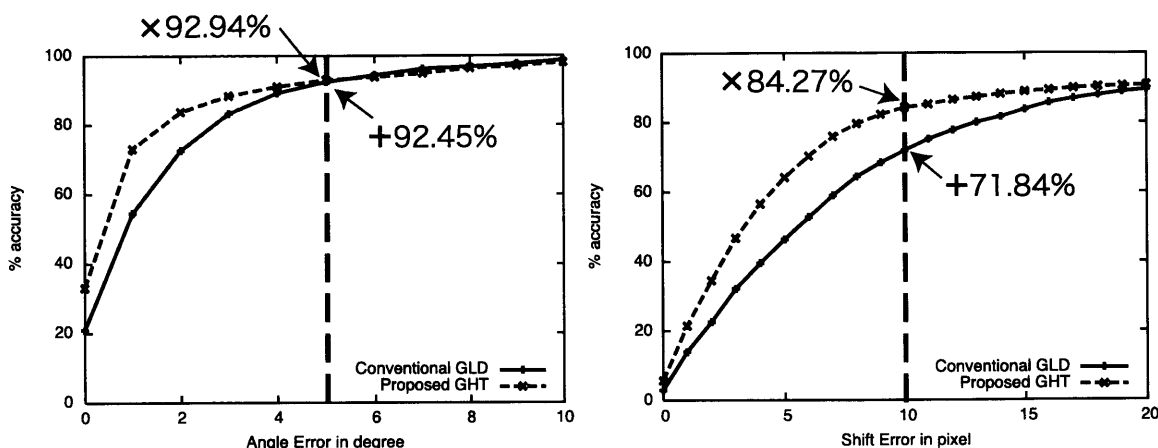


図 5.8: 検出精度の累積ヒストグラム

Performance evaluation by cumulative histograms

対して、GHT に基づく提案手法では、照明が暗い側の顔でもエッジを検出できるので、大域的な左右対称性に基づいて、理想的な正中線を検出している。

提案手法の優位性を示すために、提案手法と Chen ら [6] の従来手法のそれぞれについて、角度誤差と位置誤差 (5.2 節を参照) を調査した。図 5.8 は、提案手法と従来手法の検出精度を角度誤差と位置誤差の累積ヒストグラムで評価したものである。なお、GLD に基づく従来手法は、GHT に基づく提案手法と同じ環境で実行したものである。

図 5.8 より、検出した正中線の 92.94% の正中線は角度の誤差が 5° 以下であることがわかる。これは、提案手法によって入力画像の 92.94% の顔の回転角度が正しく推定できたことを意味する。また、検出した正中線の 84.27% の正中線は位置の誤差が 10 画素以下であることから、入力画像のうち 84.27% の正中線の位置を正しく検出できたと言える。これらの結果より、提案手法は FERET データベース上で Chen ら [6] の従来手法よりも優れていると言える。提案手法は正中線を検出するための手法として良好な結果を得られることがわかった。また、2.66 GHz Intel Core2 CPU 上で提案手法を実行したところ、所要時間は 369.12 [s] となり、フレームレートに換算すると 6.53 [frames/s] となった。

次に、角度誤差と位置誤差について詳細な調査を行った。入力となる顔画像は、頭部がいくらか傾いていると考えられる。正中線を検出する手法は、これらの頭部の傾きによらず正しく正中線を検出することが望ましい。様々な頭部の傾きに対して、正中線検出手法における傾き検出の傾向を調査した。FERET データベースは、様々な頭部の傾きを持つ顔画像が多く含まれている。FERET データベースの画像の頭部の傾きごとにグループ化し、それぞれのグループで実際に正中線を検出し、そのときの角度誤差と位置誤差の平均を調査した。図 5.9 は、GLD に基づく従来手法と、GHT に基づく提案手法のそれぞれについて、平均誤差と真の頭部の傾きとの関係をプロットしたものである。この結果よ

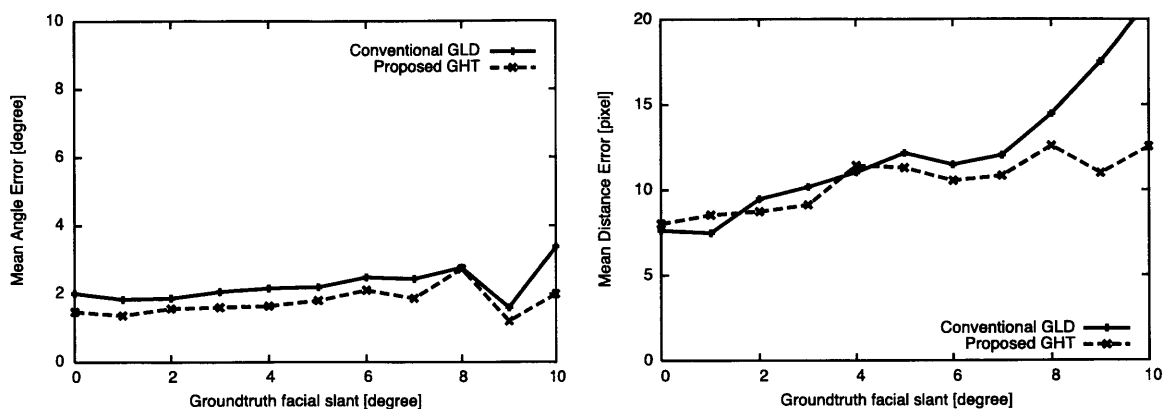


図 5.9: 角度誤差と位置誤差の頭部の傾きによる傾向

The tendency of mean angle error and mean distance error with the facial slant

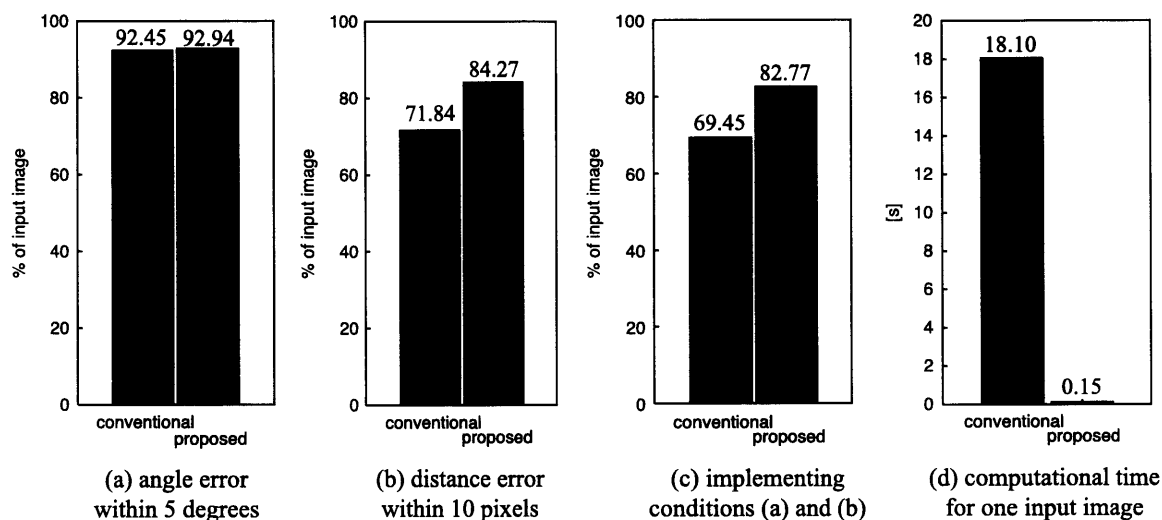


図 5.10: 提案手法と従来手法の検出成功率と実行時間

Performance comparison between the proposed and the conventional methods.

り，次のことが観測できる。

1. ほとんどの頭部の傾きにおいて，提案手法の方が従来手法よりも，平均誤差が小さい。
2. 頭部の傾きが大きくなるにつれて，従来手法では位置誤差が大きくなる．一方，提案手法では頭部の傾きによらずほぼ一定である。

図 5.10 (a) から (c) までは，これらの誤差をグラフにしたものである．(a) と (b) は，それぞれ角度誤差が $\pm 5^\circ$ 以下，位置誤差が ± 10 画素以下の画像の割合である．つまり，正中線検出の成功率である．(c) は，角度と位置の誤差のどちら条件も満たすときの割合で

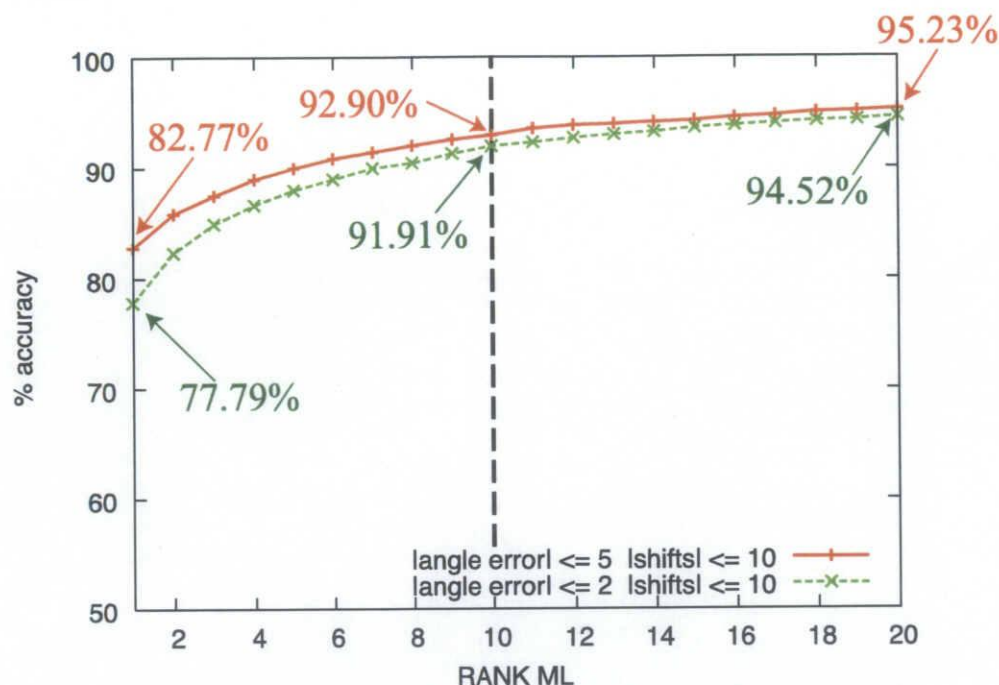


図 5.11: 提案手法による顔正中線検出成功率

Performance accuracy of the detected midlines

ある。これらの結果は、正中線検出の精度が提案手法によって改善されたことを示している。さらに、両手法における計算時間を比較した。図 5.10 (d) は、1 枚の画像を処理するのに所要時間が 18.1[s] から 0.15[s] に短縮されたことを示している。

ここまでの結果は、1 枚の画像から 1 本の正中線のみを検出し、その成功率を評価したものである。もちろん 1 個の顔には 1 本の正中線しか存在しないが、検出された正中線を顔部品抽出などに応用する場合、複数の候補を検出して後の処理に利用の方が都合が良い場合がある。このような場合には複数候補に対する評価が必要である。そこで、図 3.3 にある Hough パラメータ空間から、極大点を複数個取り出し、正中線の候補とした。この極大点の数を以降は R_{ml} と表記する。図 5.11 は、横軸に R_{ml} 、縦軸には図 5.10 (c) と同様に、角度と位置の誤差のどちらの条件も満たすときの数をプロットしたものである。例えば、 $R_{ml} = 5$ のときは、正中線の候補を 5 本取り出したときに、その中に検出成功となる正中線が含まれているかどうかを表している。ここで、文献 [6] において角度の誤差の許容範囲は $\pm 5^\circ$ 以下としていた。しかし、その正中線を顔部品抽出へと用いる場合には、角度の誤差が大きく影響するため、許容誤差範囲をさらに厳しくした $\pm 2^\circ$ 以下で成功率を調査した。グラフには 2 つの評価尺度での成功率をプロットしている。赤線のプロットは、図 5.10 (c) と同じく角度の許容誤差範囲は $\pm 5^\circ$ 以下、緑線のプロットは $\pm 2^\circ$ 以下を表している。

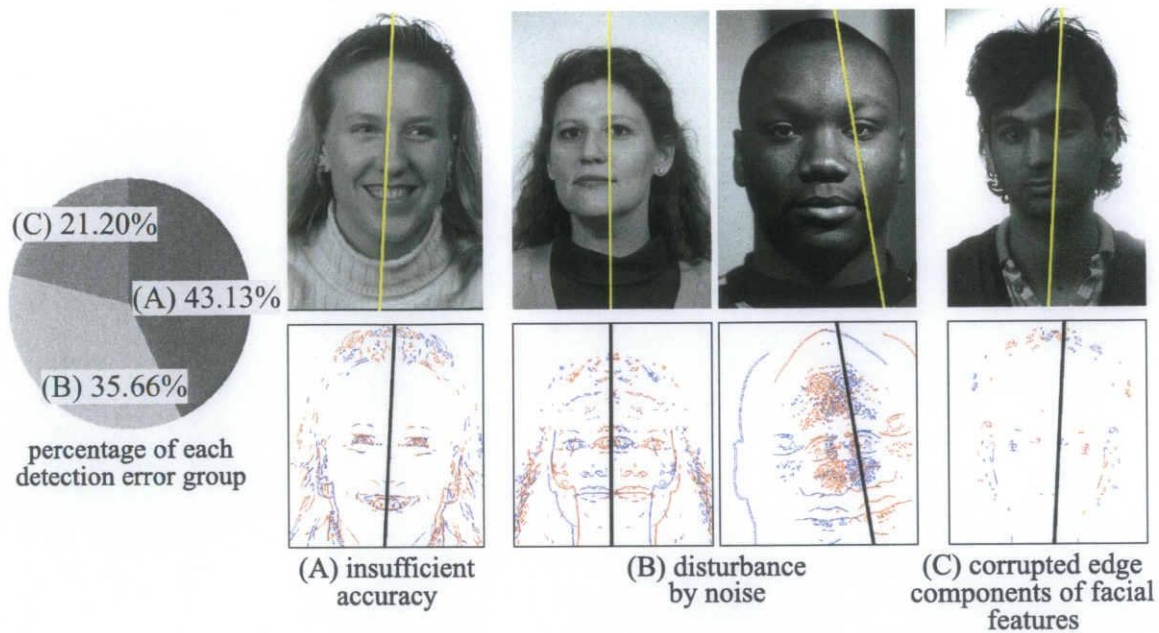


図 5.12: 正中線検出の失敗例

Examples of midline detection error

続いて、提案手法の失敗例について考察する。図 5.12 は、正中線の誤検出例とその内訳を表している。画像の上段は正中線検出結果画像で、下段は一般化 Hough 変換のマッチング結果画像である。図中左の円グラフは、 $R_{ml} = 1$ の場合（図 5.11 における赤線の条件）の検出エラー 17.23%（415 画像）における各グループの割合を示す。誤検出のうち、最も多い失敗例は（A）のグループで、検出の精度が不十分であるため失敗となる。マッチング結果は適切に見えるが、検出の精度が前述した許容誤差範囲を超えてしまう。次に多いグループ（B）は、前処理で生成した 2 値エッジ画像（3.3 節を参照）に含まれるノイズが、一般化 Hough 変換に悪影響を及ぼした例である。これに対してグループ（C）は、顔部品のエッジ成分を消失したグループである。提案手法では、顔の左右対称性を利用しているため、顔部品のエッジ成分を積極的に利用し、毛髪や衣服、照明による鏡面ハイライトに相当するエッジはノイズとして除去すべきである。しかし、提案手法は固定パラメータ（5.3.1 節を参照）を用いて 2 値エッジ画像を生成しているため、グループ（B）（C）のように GHT にとって不適切な 2 値エッジ画像が生成されてしまう場合がある。

5.4.2 顔部品抽出評価実験

次に、第 4 章で述べた提案手法を用いて、FERET データベースを用いた目の抽出実験を行った。図 5.13 は、検出した正中線を用いて目を抽出した結果画像である。これらの

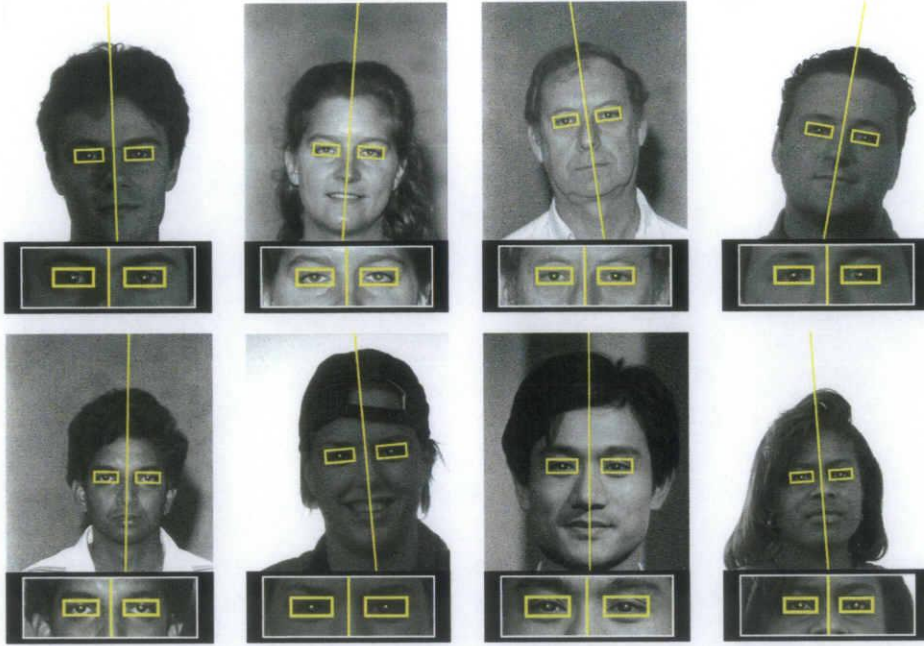


図 5.13: 検出した正中線を用いた両目抽出結果

Results of the facial feature extraction where the midlines are employed as a guide for restriction to the vertical scan-line number of one.

図では、全ての目が正しく抽出されている。テンプレートマッチングを行う前に、正中線から顔の傾きと、位置を得ることができたので、マッチング処理をより単純化できた。正中線を用いて顔の傾きと中心位置を補正して目を検出する場合の両目抽出に必要な計算実行時間は、正中線を用いない場合に比べて約 1/40 に短縮された。

目の抽出の正しさの指標として式 (5.3) を用いて $d_{\text{eye}} < 0.05$ のときを抽出成功とする [11] 場合の成功率を図 5.14 に示す。図は、真の正中線 L_t (式 (5.1)) を用いた場合と、検出した正中線を用いた場合のそれぞれの結果をプロットしたもので、横軸に目の候補数 R_{eye} 、縦軸に成功率を表している。また、検出した正中線は、 $R_{\text{ml}} = 1$ の正中線 $L(x_0^{(1)}, \theta^{(1)})$ と、 $R_{\text{ml}} = 10$ の正中線の集合 $\{L_1(x_0^{(i)}, \theta_0^{(i)})\} (i = 1, 2, 3, \dots, 10)$ 、 $R_{\text{ml}} = 20$ の正中線の集合 $\{L_2(x_0^{(j)}, \theta_0^{(j)})\} (j = 1, 2, 3, \dots, 20)$ を用いた。図 5.11 によると、角度の誤差が $\pm 2^\circ$ 以下、位置の誤差が ± 10 画素以下の正中線は、 $R_{\text{ml}} = 1$ では 77.79%, $R_{\text{ml}} = 10$ では 91.91%, $R_{\text{ml}} = 20$ では 94.52% である。このことから、 $L(x_0^{(1)}, \theta^{(1)})$, $\{L_1(x_0^{(i)}, \theta_0^{(i)})\}$, $\{L_2(x_0^{(j)}, \theta_0^{(j)})\}$ のそれぞれとを中心に、この誤差範囲内で複数の正中線を生成した。 $L(x_0^{(1)}, \theta^{(1)})$, $\{L_1(x_0^{(i)}, \theta_0^{(i)})\}$, $\{L_2(x_0^{(j)}, \theta_0^{(j)})\}$ そ

れぞれに対して、生成した正中線の集合 \mathcal{L}_1 , \mathcal{L}_{10} , \mathcal{L}_{20} を以下で定義する.

$$\mathcal{L}_1 = \{L(x_0, \theta)\}, \quad (5.10)$$

$$x_0 = x_0^{(1)} - 10, x_0^{(1)} - 8, \dots, x_0^{(1)}, \dots, x_0^{(1)} + 10,$$

$$\theta = \theta^{(1)} - 2, \theta^{(1)}, \theta^{(1)} + 2.$$

$$\mathcal{L}_{10} = \left\{L_1\left(x_0^{(i)}, \theta^{(i)}\right)\right\}, \quad (5.11)$$

$$i = 1, 2, \dots, 10,$$

$$x_0^{(i)} = x_0^{(i)} - 10, x_0^{(i)} - 8, \dots, x_0^{(i)}, x_0^{(i)} + 10.$$

$$\mathcal{L}_{20} = \left\{L_2\left(x_0^{(j)}, \theta^{(j)}\right)\right\}, \quad (5.12)$$

$$j = 1, 2, \dots, 20,$$

$$x_0^{(j)} = x_0^{(j)} - 10, x_0^{(j)} - 8, \dots, x_0^{(j)}, x_0^{(j)} + 10.$$

ここで、 $x_0^{(j)}$, $\theta^{(j)}$ はそれぞれ $R_{ml} = j$ のときの式 (3.9) における x_0 , θ である. グラフの黒, 緑, 青, 赤の線はそれぞれ L_t , \mathcal{L}_1 , \mathcal{L}_{10} , \mathcal{L}_{20} を用いて目を抽出したときの成功率である. 提案手法で検出した正中線集合 \mathcal{L}_{20} を用いて両目を抽出した結果, $R_{eye} = 5$ の成功率は 90.24% になった. 正中線集合 \mathcal{L}_{10} を用いて両目を抽出した場合は, 目の候補を 20 個 ($R_{eye} = 20$) 取り出しても, 抽出成功率は 90% に達しない. 一方で, 正中線を 20 本 ($R_{ml} = 20$) 取り出した正中線集合 \mathcal{L}_{20} を用いて両目を抽出した場合, 目の候補を 5 個 ($R_{eye} = 5$) 取り出せば, 抽出成功率は 90.24% となる.

次に, $R_{eye} = 1$ のときの目の抽出精度を, 図 5.15 の累積ヒストグラムに示す. グラフは横軸に式 (5.3) から得られる抽出誤差 d_{eye} を, 縦軸に画像の割合を表したものである. 図 5.14 と同様に, 黒, 緑, 青, 赤の線はそれぞれ, L_t , \mathcal{L}_1 , \mathcal{L}_{10} , \mathcal{L}_{20} を表す.

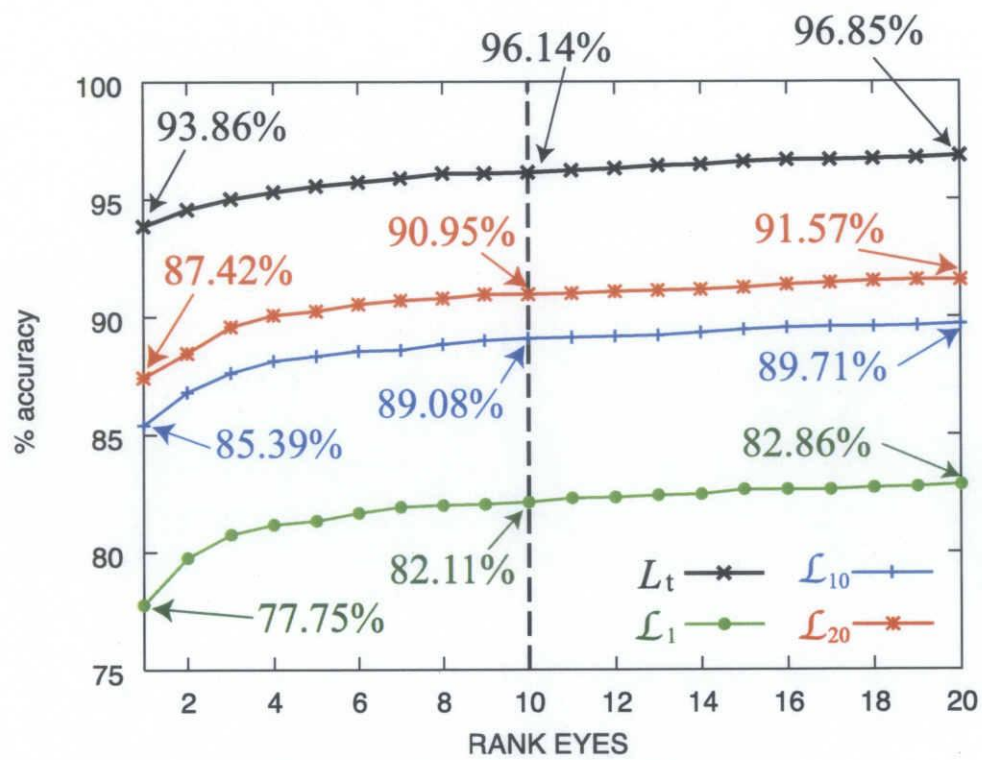


図 5.14: 提案手法による両目抽出精度

Performance accuracy of the extracted eyes

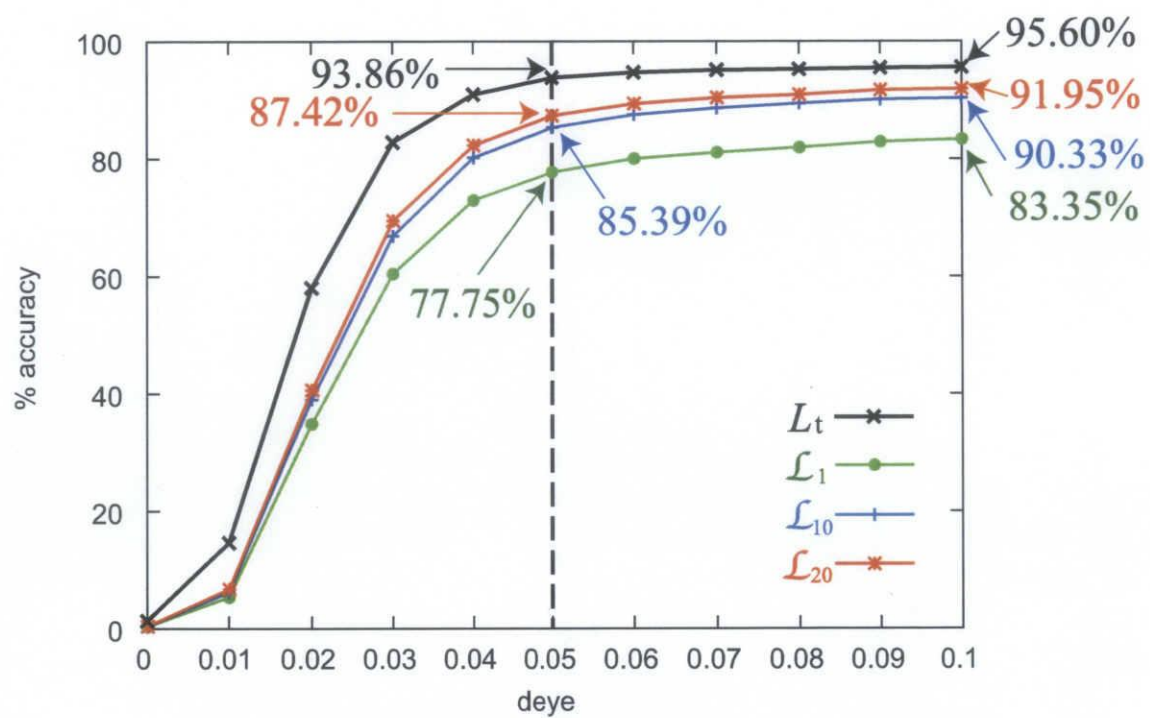


図 5.15: 両目抽出精度 d_{eye} (累積ヒストグラム)

Cumulative histograms of d_{eye}

第 6 章

結言

6.1 考察

一般的には，テンプレートマッチング法に基づく物体抽出手法では，計算コストと位置の推定の解像度とにトレードオフの関係がある．もちろん，これは，2 値画像に対するテンプレートマッチング法の高速アルゴリズムと見なせる GHT においても同様である．提案手法による正中線検出では，顔の傾きと水平位置は，1.0 度および 1.0 画素の解像度で点 q によって推定される．もしここで，それぞれの推定の解像度をさらに細かくすると，計算時間は増加する．しかし，同じ解像度の条件下で，GHT に基づく提案手法の方が GLD に基づく従来手法よりも，計算時間が小さいことが実験結果より得られた．

図 5.10 の結果から，提案手法の方が従来手法よりも検出精度が高いことがわかった．また，計算コストも，提案手法の方が小さいことが示された．その理由として，提案手法では左右対称性検出のために，2 値エッジ画像と GHT を利用しているためである．提案手法では Sobel オペレータと p -タイル法によって，原画像であるグレースケール画像よりも光源変化に対して頑健な 2 値エッジ画像を得ることができた．

上述した提案手法の有能性を証明する実験は，無地または単調な背景の画像を多く含む FERET データベースを用いて行われた．もし，複雑で自由な背景を持つ画像に提案手法を適用すると，複雑な背景のエッジ成分は，GHT による左右対称性検出に悪影響を及ぼす．そのような場合には，既存の顔検出手法 [1, 12, 11] を利用して，背景を取り除くことが必要になる．これは，提案手法が正しく正中線を検出するための一助となる．図 6.1 (a) は，複雑な背景の画像を用いて正中線検出を試みて失敗した例である．また，図 6.1 (b) は，背景の領域を考慮せずに顔領域のみを用いて提案手法を適用し，正中線が正しく検出された例である．

次に，顔部品抽出について考察する．本研究では，両目を抽出するために両目の中心座標 p_r, p_l を求める手法を提案した．テンプレートマッチング法により画像中から 2 点を求

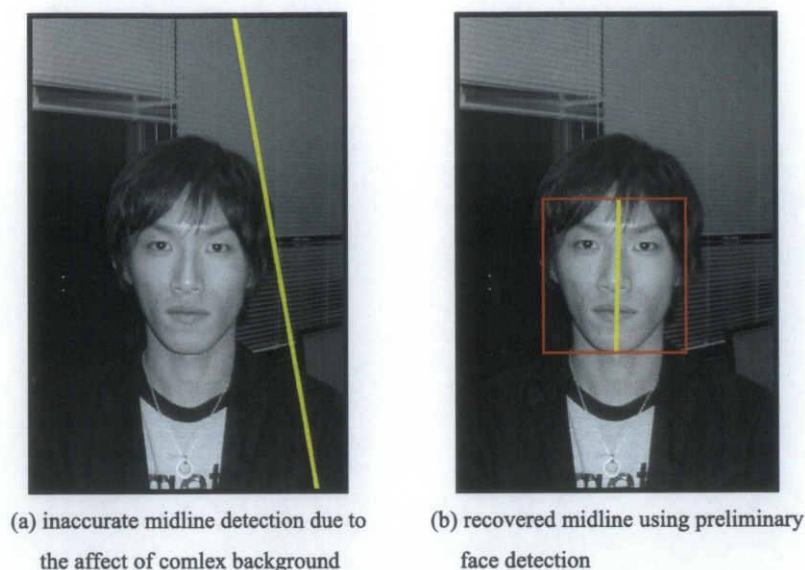


図 6.1: 複雑な背景を持つ画像に対する正中線検出

An example of midline detection that the proposed GHT-based method was applied on an image containing complex background

める問題は、高い計算コストを要求するが、提案手法で検出された顔正中線を用いることで、大幅なコスト削減が実現できた。

提案手法は、両目が正中線に対して線対称であるという仮定の下で探索を行うが、図 5.14 の結果から、真の正中線 L_t を用いたときの目の抽出成功率は 93.86% であり、約 6% は抽出に失敗する。誤抽出の内訳を図 6.2 にグラフと画像で示す。画像中の黄色の線が L_t 、黄色の枠が検出された目領域と、その中点が目の位置 p_r, p_l である。誤抽出を以下の 5 グループに分類した。

- (A) 失敗例が最も多いグループで、画像中の人物が眼鏡をかけている。眼鏡レンズ上の鏡面反射や着色レンズの眼鏡によって、目の見かけが変化するため、誤抽出したと考えられる。画像 (A) は、目領域で眼鏡が反射して正しく目領域を捉えることができず、眉を目と誤抽出している。
- (B) 画像中の人物が目を閉じている例である。顔部品抽出に用いたテンプレートは、目を閉じている画像を含んでいないため、マッチングできない。テンプレートの検討が求められる。
- (C) 4.3.2 節の勾配強度の 2 値化処理に失敗した例である。本来は探索の対象となるべき領域が、取り除かれてしまい、目の抽出に失敗している。計算コストの面から、探索の対象領域をできるだけ少なくすることが必要だが、このような失敗を改良するためにも、2 値化処理に改善が必要である。

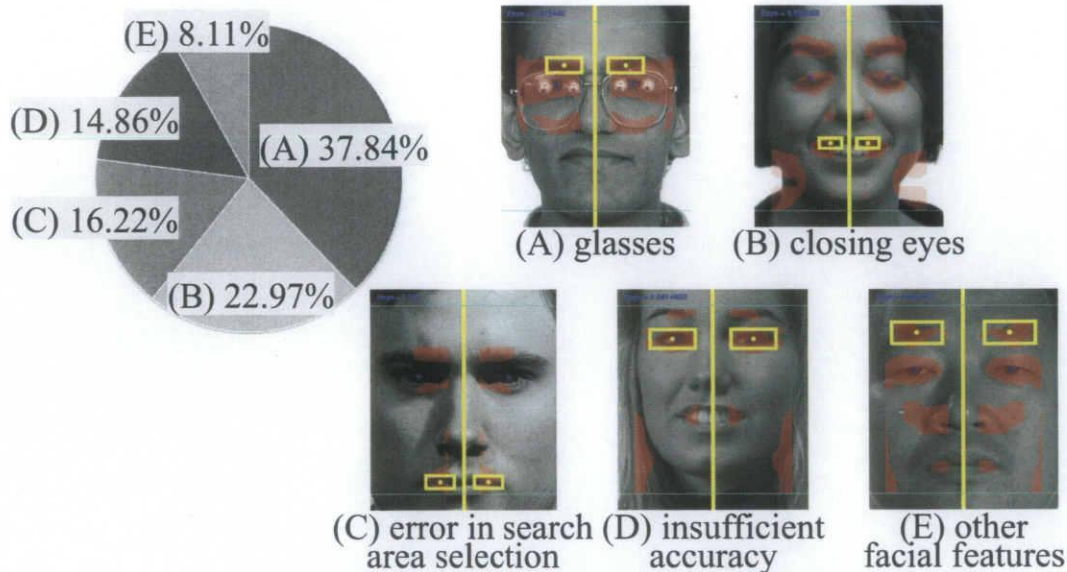


図 6.2: 両目抽出に失敗した例と抽出失敗の原因
examples and reasons of incorrect extraction

- (D) 目領域の付近に検出されているが、式 (5.3) の評価尺度でその誤差 d_{eye} が $d_{\text{eye}} \geq 0.05$ となっているグループである。仮に許容誤差を $d_{\text{eye}} < 0.1$ とすれば成功となるグループで、この範囲の誤差を有していても、目付近を検出できることは有意義である。この抽出された目を用いて、顔を検出する次の行程に、十分に寄与できると考える。
- (E) 他の顔部品（眉、鼻腔、口）を誤抽出している例である。マッチングに用いるテンプレートの検討、他の評価関数での実験など、検討する余地がある。

以上の失敗例を改善すれば、提案手法で自動検出した正中線を用いた場合にも、精度の向上が期待できる。

6.2 今後の課題

本研究では、画像から顔の正中線を検出する手法を提案した。GHT に基づく本手法は、高速で平易で、良好な性能を持つ。さらに、検出された正中線を顔部品抽出へと応用させ、濃度勾配特徴を用いて目の抽出を行った。正中線を利用する事で、計算コストを大幅に削減でき、良好な結果を得た。

今後の課題として、(1) 顔正中線検出性能、顔部品抽出性能のさらなる改善、(2) 顔部品抽出処理の高速化、(3) 抽出した顔部品を用いた顔検出、が挙げられる。

正中線検出の主な失敗原因としては、2値エッジ画像に含まれるノイズの影響で、GHT

によるマッチングに失敗したと考える．これらを改善するために，前処理（3.3 節）のさらなる改良を行う必要がある．この改良によって，複雑な背景を持つ画像や，複数の顔が含まれる画像からも正中線を自動検出したい．また，顔部品抽出においては，眼鏡が反射している顔画像や，目を閉じている顔画像に，多くの失敗が見られる．これらの見え方が複雑な目も抽出できるように改善する必要がある．本研究では，目らしさの類似性を評価するために，コサイン係数しか用いていないので，他の評価関数でも実験を行う必要がある．また，テンプレート生成のための学習データセットとして，FERET データベースの fa グループを用いており，目を閉じている例はデータセットには含まれていない．データセットのデータ数の増加により改善できるか，実験する必要がある．

また，顔部品抽出のさらなる高速化が求められる．安易に探索範囲を削減すると，抽出精度の低下が懸念される．本手法では，濃度勾配強度の平均により探索範囲を削減したが，正解となる領域を探索範囲から除外してしまったり，あるいは探索範囲を全く削減できなかったりと，適切な結果を得られない失敗例もある．探索範囲を適切に削減し，高速化を実現できる手法を検討する必要がある．

最後に，抽出された目を用いた顔検出を目標として挙げる．本手法では，入力画像から正中線を検出し，正中線から目を抽出した．工程をさらに進め，目を用いて顔領域を切り出し，顔検出の精度を評価したい．また，切り出された顔領域が顔認識の精度にどのような影響を与えるかも検討したい．

付録 A

目領域の大きさモデルの決定

提案手法で考案した目領域の大きさを決定する線形モデルについて述べる．このモデルは，人物の顔の大きさが画像ごとに異なることから考案したものである．

図 A.1 は，ある人物画像の例である．図中の黄色の長方形は目の外接長方形である．赤線は左目と右目の中心を結んだ線分で，眼間線分と呼ばれる．二つの画像における人物の顔の大きさと同様に，目の大きさも異なる．このように画像ごとに目の大きさ（黄色の長方形の大きさ）は異なる．入力画像に対して，眼間線分が決まれば目の大きさと位置が一意に決まるモデルを提案した．

モデルの例を図 A.2 に示す．図中の長方形を目の領域とし，その長方形の中心間を結ぶ直線を長さ d の眼間線分とする．このモデルは以下の式によって構成される．

$$\begin{pmatrix} w \\ h \end{pmatrix} = \begin{pmatrix} 0.51 \\ 0.22 \end{pmatrix} d + \begin{pmatrix} 6.04 \\ 3.15 \end{pmatrix}. \quad (\text{A.1})$$

モデルの式 (A.1) は，回帰分析により取得したものである．変数は d のみで， d が決まれば，長方形の幅と高さが一意に決まり，眼間線分の両端に長方形の中心をあわせることで，長方形の位置も決まる．



図 A.1: 画像ごとに異なる目領域の大きさ
Example of different size of Eye Window

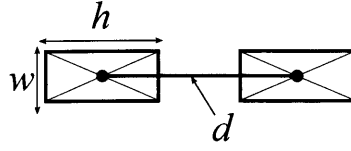


図 A.2: 目領域の線形モデル

Proposed Eye Window Model

次に、このモデルを決定した回帰分析実験 [26] について述べる。サンプルとして、200 枚の画像を学習データとして用いた。学習データについては 5.1 節にて述べる。これらのサンプルにおいて、事前に目視抽出されている両目の座標から眼間線分長 d を求めた。また、目視で両目の外接長方形を求め、その長方形の幅と高さを左右で平均し、これを各画像の w, h として求めた。これら d, w, h を実測値として用いて回帰分析を行った。なお、ここでの回帰分析は、 $y = a + bx$ なる回帰直線を求めることで、 $x = d$ としたときに、 $y = w, h$ が決まる直線を決定する。以下の式を用いて、回帰直線 $y = a + bx$ を求める。なお、サンプル数を n 、実測値を $x_i, y_i (i = 1, 2, \dots, n)$ 、それぞれの実測値の平均を \bar{x}, \bar{y} とする。また、 S_{xx} は x の偏差平方和、 S_{xy} は x, y の偏差積和である。

$$b = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - (\sum x_i \cdot \sum y_i) / n}{\sum x_i^2 - (\sum x_i)^2 / n}, \quad (\text{A.2})$$

$$a = \bar{y} - b\bar{x} = \frac{1}{n} (\sum y_i - b \sum x_i), \quad (\text{A.3})$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}, \quad (\text{A.4})$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}. \quad (\text{A.5})$$

式 (A.2), (A.3) より得られた結果が、式 (A.1) の線形モデルである。

続いて、式 (A.1) の回帰直線の有意性を検証するために、 t 検定を行った。

t 検定とは、行われた回帰分析が意味のある回帰か検定するもので、回帰直線が $y = a + bx$ で与えられているときに、その傾き b が有意に 0 から偏っているかどうかを調べる。検定のために、 b の標準誤差を求め、 b 値の有意性を判定する。ここで、 b の標準誤差 S_b は以下の式で求める。なお、推定値を $Y_i (= a + bx_i)$ とする。また、 s は回帰直線からの標準偏差を表す。

$$S_b = \frac{s}{\sqrt{S_{xx}}}, \quad (\text{A.6})$$

$$s = \sqrt{\frac{\sum (y_i - Y_i)^2}{n - 2}} = \sqrt{\frac{b \cdot S_{xx}}{n - 2}}. \quad (\text{A.7})$$

次に、 b 値の偏り度合を求めるために、式 (A.6) によって得られた標準誤差 S_b を用い

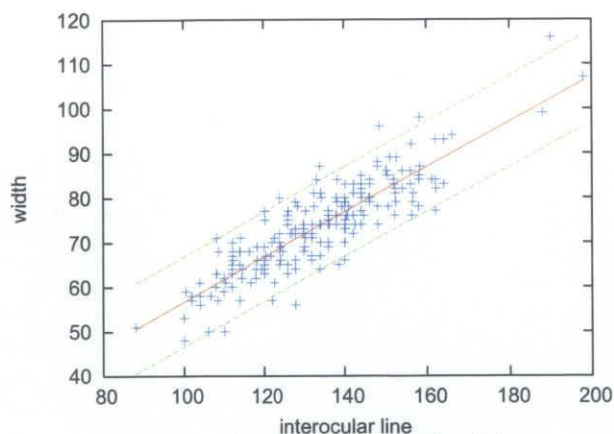


図 A.3: 眼間線分と目領域の幅

width of window v.s. length of interocular line segment

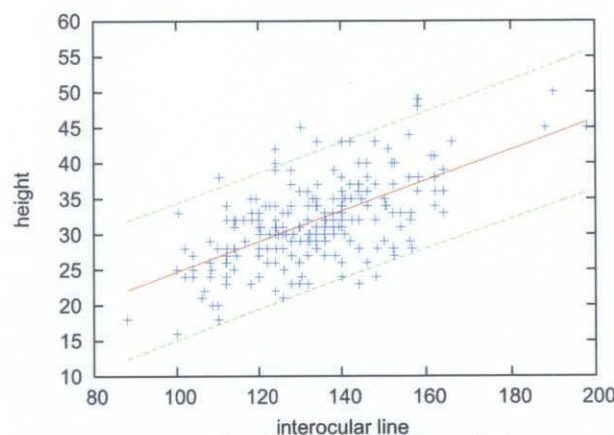


図 A.4: 眼間線分と目領域の高さ

height of window v.s. length of interocular line segment

て、以下の式で標準化する。

$$t = \frac{b}{S_b}. \quad (\text{A.8})$$

得られた統計量 t と、 t 分布表で調べた t 値の生じる確率により有意性の判定を行う。

本実験で得られた 2 つの回帰式 (A.1) に対して検定を行ったところ、それぞれの統計量 t_w, t_h は $t_w = 24.885, t_h = 11.092$ となった。本実験で用いた画像は 200 枚であり、 $n = 200$ のときの両側確率 0.05 の t 値は、 t 分布表 [27] によると $t_{0.05} = 1.972$ であり、 $t_w > t_{0.05}, t_h > t_{0.05}$ を満たしている。よって、回帰式 (A.1) は 5% の有意水準で有意であるといえる。

以上より、提案手法では式 (A.1) を目の大きさを決めるモデルとして採用する。

実験データのプロットを図 A.3, A.4 に示す。グラフの横軸は眼間線分 d 、縦軸はそれぞれ w, h を表している。青の点は、あるサンプルの d と、左右の目で平均した w, h をプ

ロットしたものである。また、赤の直線は、青の点から求めた回帰直線 $y = a + bx$ で、式 (A.1) に相当する。さらに、緑の波線は、回帰直線の 95% 信頼区間を表している。95% 信頼区間の領域の中に、ほとんどのサンプルが含まれていることが見てとれる。なお、95% 信頼区間は式 (A.9) によって求めた。

$$y = a + bx \pm t_{0.05} \cdot s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}. \quad (\text{A.9})$$

付録 B

校正用目座標の修正

本研究で用いた FERET データベースは、各画像に目視抽出された顔部品の座標が校正用（正解）情報として与えられている。本研究におけるそれぞれの実験では、両目の座標 C_r, C_l を用いた。しかし、これらの目の正解座標を目視で確認したところ、正しく目の中心位置を捉えていない例が存在した。本研究では、このような精度の校正用情報を用いて目の抽出用のテンプレート生成や自動抽出の精度評価を行うことは望ましくないと判断し、より精度の高い校正用目座標を目視抽出を行った。本論文で用いた両目の座標 C_r, C_l は、これらの新たに目視抽出した校正用の目座標である。FERET により与えられた目座標と今回修正を行った目座標との例を図 B.1 に図示する。上段の画像中の赤色の円と下段の画像中の黄色の円はそれぞれ、FERET により与えられた目座標と新たに目視抽出した目座標である。

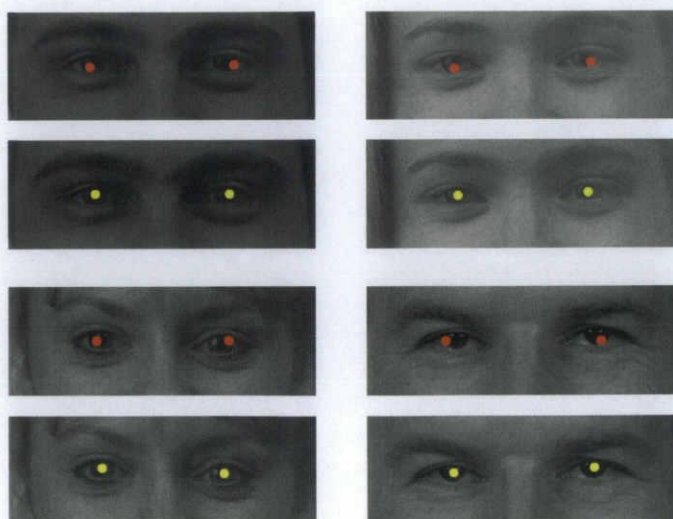


図 B.1: 校正用目座標の修正

Correction of ground truth eye coordinates

謝辞

本研究に関し、研究に対する基本姿勢を御教授下さった木村文隆教授、いつでも丁寧に指導をして下さった優しい若林哲史准教授、至る所まで面倒を見て下さった大山航助教に深く感謝いたします。また、充実した学生生活を支えて下さり、日々おしゃべりに付き合っ下さったヒューマンインタフェース研究室事務員田中みゆきさんに感謝いたします。学会前には親身にご指導して下さいった **Pal** 先生、日頃英語の添削をして下さった **Busagala** さん、楽しい時間をたくさんくれた優しい劉さん、癒しの雰囲気を与える謝さん、いろんな勉強をさせてくれる陳さん、国境を越えた皆さんと出会えてとっても嬉しいです。そして、自由奔放で個性的な先輩方、若さたっぷりの後輩たち、楽しく充実した時間をたっぷりくれた同級生の皆さんに感謝します。また、ポルトガルで出会った世界各国の優しい方々、素敵な思い出をありがとうございました。ヒューマンインタフェース研究室から授かった全ての出会いは私の一生の宝物です。

三重大大学での生活を有意義で刺激的にしてくれた、ヒューマンインタフェース研究室の人々、雰囲気、環境、その全てに深く感謝いたします。

最後になりましたが、長きにわたる私の学生生活を支えてくれた両親、私の大切な姉、私の宝物のおいこたち、めいこたち、いつまでも心の支えの沖縄のどうしぐわあーに今一度の感謝を表して、本論文の結びといたします。

Acknowledgements

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

参考文献

- [1] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp. 34–58, 2002.
- [2] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, Vol. 35, No. 4, pp. 399–458, 2003.
- [3] L.-F. Chen, H.-Y. M. Liao, J.-C. Lin, and C.-C. Han. Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof. *Pattern Recognition*, Vol. 34, No. 7, pp. 1393–1403, 2001.
- [4] P. Quintiliano, R. Guadagnin, and A. Santa-Rosa. Practical procedures to improve face recognition based on eigenfaces and principal component analysis. In *The 5th International Conference on Pattern Recognition and Image Analysis: New information Technologies*, pp. 372–375, 2001.
- [5] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, pp. 1090–1104, 2000.
- [6] X. Chen, P. J. Flynn, and K. W. Bowyer. Fully automated facial symmetry axis detection in frontal color images. In *4-th IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 106–111, 2005.
- [7] P. S. Hiremath and Ajit Danti. Detection of multiple faces in an image using skin color information and lines-of-separability face model. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 20, No. 1, pp. 39–61, 2006.
- [8] G. Dedeoğlu, T. Kanade, and S. Baker. The asymmetry of image registration and its application to face tracking. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 29, No. 5, pp. 807–823, 2007.
- [9] Guangzheng Yang and Thomas S. Huang. Human face detection in a complex background. *Pattern Recognition*, Vol. 27, No. 1, pp. 53–63, 1994.

- [10] K.C. You and R. Cipolla. Feature-based human face detection. *Image and Vision Computing*, Vol. 15, No. 9, pp. 713–735, 1997.
- [11] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, H. Kälviäinen, and J. Mates. Feature-based affine-invariant localization of faces. *IEEE trans. On Pattern Analysis and Machine Intelligence*, Vol. 27, No. 9, pp. 1490–1495, 2005.
- [12] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz. Robust face detection using the hausdorff distance. In *Proc Int'l Conf. Audio- and Video-Based Biometric Person Authentication*, pp. 90–95, 2001.
- [13] Ying Dai and Yasuaki Nakano. Face-texture model based on sgld and its application in face detection in a color scene. *Pattern Recognition*, Vol. 29, No. 6, pp. 1007–1017, 1996.
- [14] I. Craw, D. Tock, and A. Bennett. Finding face features. In *Proceedings of Second European Conf. Computer Vision*, pp. 92–96, 1992.
- [15] A. Lanitis, C.J. Taylor, and T.F. Cootes. An automatic face identification system using flexible appearance models. *Image and Vision Computing*, Vol. 13, No. 5, pp. 393–401, 1995.
- [16] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86, 1991.
- [17] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 23–38, 1998.
- [18] D.H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, Vol. 13, No. 2, pp. 111–122, 1981.
- [19] E.R. Davis. New framework for analysing the properties of the generalized hough transform. *Pattern Recognition Letters*, Vol. 6, pp. 1–7, 1987.
- [20] 田村秀行（編）. コンピュータ画像処理, pp. 182–199. オーム社, 2002.
- [21] 田村秀行（編）. コンピュータ画像処理, pp. 138–142. オーム社, 2002.
- [22] 田村秀行（編）. コンピュータ画像処理, pp. 151–152. オーム社, 2002.
- [23] M. Sato and H. Ogawa. Optimization of the generalized hough transform. In *Proceedings on 4-th International Conference on Image Analysis and Processing*, 1987.
- [24] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *Proc. of Conference on Computer Vision Pattern Recognition*, pp. 130–136, 1997.
- [25] M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura. Handwritten numeral recognition using gradient and curvature of gray scale image. *Pattern Recognition*, Vol. 35, No. 10,

pp. 2051–2059, 2002.

- [26] 市原清志. バイオサイエンスの統計学／ Statistics for Bioscience, pp. 204–217. 南江堂, 1990.
- [27] 市原清志. バイオサイエンスの統計学／ Statistics for Bioscience, p. 346. 南江堂, 1990.