**2013 年度（平成 25 年度）**

博士論文

# 生体的特徴を用いた
# 人物の属性・行動の分類

立命館大学大学院

理工学研究科総合理工学専攻

東 篤司

# 目次

# 第1章. 序論

## 1.1. 研究背景

近年ではテロ，犯罪の増加により，日常は危険との隣り合わせであると認識する機会が増え，人々のセキュリティに対する関心が非常に高まっている．このような状況を受け，「バイオメトリクス（認証)」を導入した製品の需要は今後さらに高まると予想される．

バイオメトリクスとは「行動的あるいは身体的な特徴を用いて個人を自動的に同定する技術」として定義できる**[1]**．行動的特徴の例としてキーストロークや動的署名，声紋，歩行が挙げられ，また身体的特徴では顔や指紋，静脈，虹彩，網膜，顔の赤外画像，匂い，DNA，耳などが挙げられる．

バイオメトリクスに関して特にコンピュータを用いた画像（信号）処理技術の市場は拡大を続けている．デジタル画像処理技術が一般的になる 1980 年代初期，犯罪捜査にて計算機による指紋照合アルゴリズムが初めて導入された．そして 1985 年頃には 1980 年代と比較してシステム開発コストが低減し，原子力発電施設などの重要施設関連の入退室管理システムとして利用されるようになった．1995 年以降ネットワークの発達により，システムはネットワークに接続された PC や IC カードで構築され，装置コストは更に低下した．これにより市場は装置市場からシステムインテグレーション市場にシフトすることとなる．2003 年以降，モバイル端末認識サービスの市場が立ち上がり，更なる低コスト化が実現されている．近年では銀行 ATM や PC のログインなど身近なサービスに対しても利用されている．そして店舗や公共施設等の監視カメラにおいてもバイオメトリクスは導入されるようになってきている．

監視カメラは，セキュリティへの関心の高まりによるマンションや店舗など設置場所の増加と，従来のアナログカメラから IP カメラへの置き換えにより市場規模の拡大が今後も予想されている．**図 1.1** にアナログカメラを除いた監視カメラ市場規模の遷移のグラフを示す**[2]**．**図 1.1** から市場規模は 2015 年には約 600 万台に到達すると予測されている．このような市場規模の拡大とその普及に伴い，従来のアナログカメラより大幅に高解像度化する IP カメラは，顔認証等のバイオメトリクス技術の認証性能の向上を招き，今後，バイオメトリクス技術の導入が加速すると考えられる．

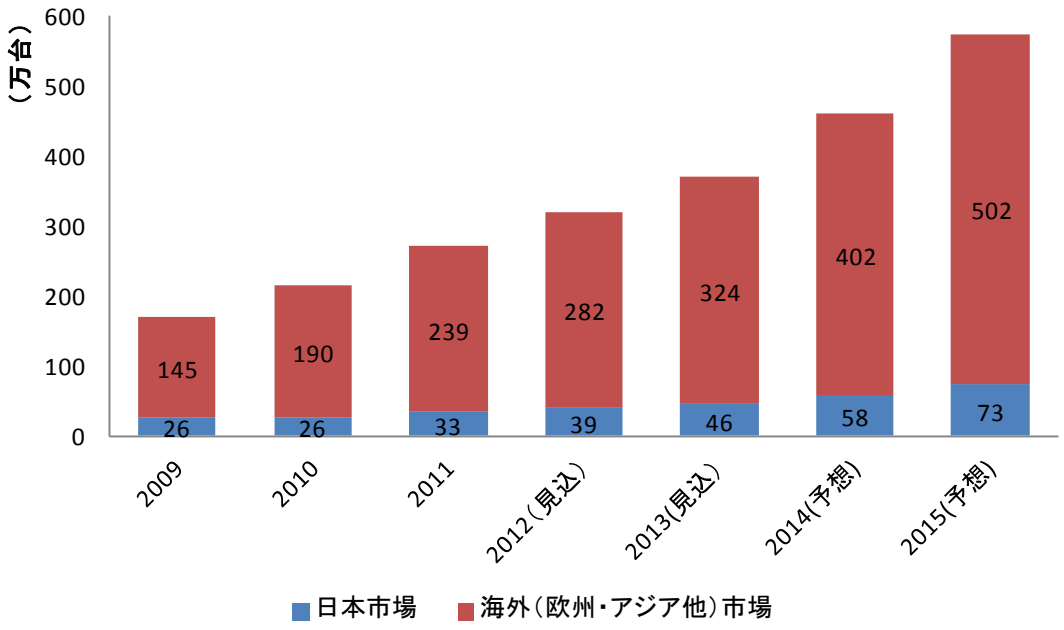



**図 1.1**：世界の監視カメラ市場規模の遷移
（アナログ監視カメラは除く）

バイオメトリクス技術を導入した監視カメラの例として，犯罪捜査への利用が挙げられる．それは登録されている顔画像集合内の犯罪者の顔と映像内の人物が同一人物であると認識した場合，顔画像を拡大映像で保存する機能などである．更に監視カメラはバイオメトリクス技術による個人認証に留まらず，その技術を応用して人物の生体的特徴を基にした異常検知やマーケティングへの利用に発展しつつある．例えば異常検知においては，店舗での「万引き」，一般家庭への「不審者の侵入」，駅や空港における「置き引き」等の犯罪を検知して警備員などに知らせるシステムへの活用が考えられている．またマーケティングにおいては，コンビニ等の店舗内での顧客の行動を解析することで商品の陳列方法の改善や，顧客の顔から年齢層や性別等を分析し，店舗の特性毎に品揃えを最適化する等の活用が考案されている．

現状では監視カメラ映像への画像（信号）処理技術の適用は，外乱の影響が少ない公共施設や店舗内などの屋内環境といった限られた条件下であれば，一部の画像処理機能において実用化されている．しかし実環境を想定した場合，照明や天候の変化などによる外乱や隠れ等のオクルージョンへの対応，人物の見える角度や姿勢の変化に対する汎化性の向上など，現在でも多くの課題が存在する．ゆえにそれらの課題解決への取り組みは多くの研究機関でなされており，今後の更なる技術の高度化は必須であると言える．

## 1.2. 研究目的と論文構成

本研究の目的は，1.1 節の後半で述べた今後の展開が期待される監視カメラの異常検知技術の高度化に取り組み，人間警備員と同等の能力を持たせることである．具体的には身体的特徴である「顔」を基にした人物の属性分類，行動的特徴である「体の動き」を基にした不審者検知のための行動分類についての手法を検討することである．

不審者を高精度に検出して知らせる機能を充実させ，人間警備員と同等の能力を持たせることで監視カメラの有用性は非常に高まる．そのためには 3 つの技術の高度化が不可欠である．それはⅠ．人物の検出，Ⅱ．検出した人物の行動，表情を基にした異常検知，Ⅲ．検知した不審人物の認識技術である．その中でⅠ．の人物の検出技術については研究室単位で既に取り組んでおり，高いレベルの性能を発揮するところまで至っている．Ⅲの認識技術では不審者の情報として年齢，性別や身長といったデータを登録しておけば迅速な不審者の特定が期待できる．

本研究ではⅡ，Ⅲに用いられる顔の属性分類（年齢・性別・表情）と不審者検知のための行動分類についての独自アルゴリズムの提案，実装，そしてその性能評価についての研究を行った．

本稿は全 11 章と付録 A から構成されており，構成のブロック図を**図 1.2** に示す．第 2, 3, 5 章は準備という位置付けで, 顔画像の正規化などに用いられる Active Appearance Model (AAM)や AAM の発展形である Generic AAM，そして後の章に関連する特徴量抽出法について述べる．また付録 A では 6，7，8，9 章の実験の識別器として用いた Support Vector Machine (SVM)について述べる．

第 4 章では Generic AAM を用いた性別分類手法を提案し，性能検証を行う．

第 6 章では，5 章で紹介した特徴量抽出手法に関連する Local Gabor Directiona Patten Histogram Sequence (LGDPHS)と称した新たな特徴量を提案する．それを顔画像の年齢・性別分類に適用し，その性能検証を行う．

第 7 章では 6 章で提案した LGDPHS と 4 章で述べた Generic AAM を用いた年齢・性別推定アルゴリズムを提案し，性能の検証を行う．実験では従来法との性能の比較，更に年齢推定において大学生 20 名の主観評価による見かけ年齢との性能の比較を行う．

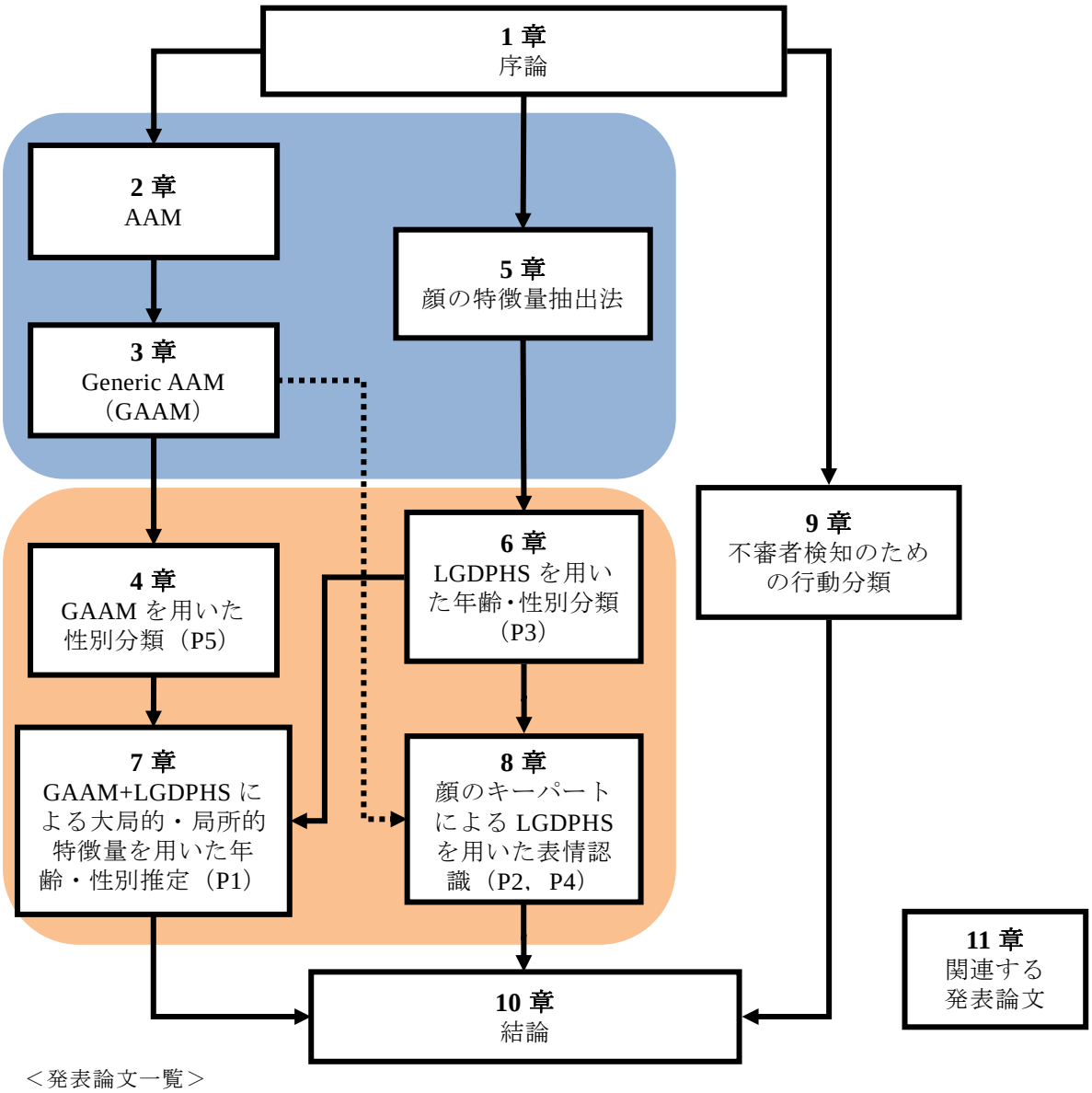
第 8 章では Generic AAM と LGDPHS を顔画像からの表情認識手法に応用する．顔のキーパートに対して LGDPHS を適用する特徴量を提案し，性能を評価する．提案手法は正規化されたキーパートのみから特徴抽出を行うことで認証対象者や表情の変化に対して，位置やスケール，傾きの不変性を保持した特徴量抽出が期待できる．

第 9 章では異常検知のための行動分類手法について検討し，その性能検証を行う．監視カメラの設置環境は多く想定されるが，本稿では寺社・仏閣における不審者検知のための行動分類に焦点を当てる．提案手法は時空間のスケール変動に頑強な局所特徴量を用いて

単純で短い行動である行動素の組み合わせと順序から行動を分類する．

10 章では本稿全体を総括し，得られた知見及び課題をまとめる．

最後に，11 章にて本研究に関係する発表論文の一覧を示す．




＜発表論文一覧＞

P1. “Age and Gender Estimation Using Global and Local Feature with AAM and LGDPHS”（IIEEJ 2012）

P2. “Expression Recognition Using LGDPHS Based Facial Key Part”(IIEEJ 2012)

P3. “Local Gabor Directional Pattern Histogram Sequence (LGDPHS) for Age and Gender Classification” (IEEE SSP 2011)

P4. “Expression Recognition using Local Gabor Directional Pattern Histogram Sequence (LGDPHS)” (NCSP 2012)

P5. “Active Appearance Model による顔特徴量を用いた男女識別”(PRMU 2009)

**図 1.2**：本稿の構成

# 参考文献

[1] バイオメトリクスセキュリティコンソーシアム編：“バイオメトリックセキュリティ・ハンドブック”（2006），（オーム社）

[2] 株式会社 矢野経済研究所：“世界のネットワークカメラ市場に関する調査結果 2013 －アジア・中東圏で高成長、2015 年 575 万台のネットワークカメラ世界市場を予測 －”

# 第2章. Active Appearance Model

## 2.1. まえがき

本章では，本研究で提案する年齢・性別・表情といった顔の属性分類における顔画像の正規化及び特徴抽出のために用いる Active Appearance Model (AAM)の概要について述べる．それは T.F. Cootes らによって提案された手法であり，顔等の予め用意した形状とアペアランスから構成されるモデルと入力物体の二乗和誤差を最小化することでその形状と形状内部のテクスチャの輝度値を同時に低次元で表現できる統計モデルである**[1]**．I. Matthews, S. Baker らは効率的な AAM のフィッティングアルゴリズムである Inverse Compositional アルゴリズムを提案している．この最適化法より Lucas-Kanade アルゴリズムの反復処理の計算量を大幅に削減することに成功し，動画像への適用を可能にしている **[2,3,4]**．近年では AAM は 2.5 次元モデルへ拡張され，顔のトラッキング**[5]**や加齢による個人の顔の変化に対しての人物同定など犯罪捜査へも応用されている．本章は 2.2，2.3 節で形状とアペアランスモデルについて説明し，2.4 節においてそれらモデルの作成事例を示す．そして 2.5 節においてモデルの最適化法について述べ，2.6 節で AAM の特長と課題を挙げる．最後に 2.7 節で本章をまとめる．

## 2.2. 形状モデル

AAM の形状モデルは 3 つのステップより生成される．始めに，学習画像の目や口，鼻，眉などの顔器官や輪郭に対して手動で複数の頂点を打ち，2 次元の座標情報を採取する．次に，採取した頂点の座標群に対して，一般化プロクラステス分析（Procrustes Analysis）**[6]**を施し，形状の正規化を行う．最後に，正規化後の座標に対して主成分分析（Principal Component Analysis; PCA）(多変量データの持つ情報を, 少数個の総合特性に要約する手法) **[7]**を施す．これにより平均形状$\mathbf{s}_0$, 固有値の値を大きい順に並べた$n$個の固有ベクトル $\mathbf{s}_i$が求まり，あらゆる形状$\mathbf{s}$は平均形状$\mathbf{s}_0$と$n$個の基底ベクトル $\mathbf{s}_i$の線形結合で近似的に表現できる：

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i \tag{2.1}$$

この式において，係数$p_i$は形状パラメータであり，$p_i$を変化させることであらゆる形状$\mathbf{s}$を表現することができる．そして$\mathbf{s}_i$は正規直交ベクトルである．

また，形状モデルはメッシュ状に定義され，ある特定の頂点の集合で定義される．数学的に，形状ベクトル$\mathbf{s}$は，頂点$v$の座標を用いて以下のように定義される：

$$\mathbf{s} = (x_1, y_1, x_2, y_2, \dots, x_v, y_v)^T \tag{2.2}$$

## 2.3. アペアランスモデル

アペアランスモデルは平均形状$\mathbf{s}_0$内に含まれるテクスチャの輝度値として定義され，2つのステップより生成できる．手動で打った複数の頂点を基に学習画像を線形補間法により平均形状$\mathbf{s_0}$内にアフィン変換する．最後に形状モデルと同様に PCA を施す．

平均形状$\mathbf{s}_0$内部にある座標$(x, y)$のピクセルを$\mathbf{x} = (x, y)^{\mathrm{T}}$とすると，アペアランスモデルは$\mathbf{x} \epsilon \mathbf{s}_0$の条件の下で$\mathbf{A(x)}$として定義でき，あらゆるアペアランス$\mathbf{A(x)}$は平均アペアランス$\mathbf{A}_0\mathbf{(x)}$と$m$個の基底ベクトル $\mathbf{A}_i\mathbf{(x)}$の線形結合で表現できる：

$$\mathbf{A(x)} = \mathbf{A}_0\mathbf{(x)} + \sum_{i=1}^{m} \lambda_i \, \mathbf{A}_i\mathbf{(x)} \qquad \forall \mathbf{x} \epsilon \mathbf{s}_0 \tag{2.3}$$

ここで，係数$\lambda_i$はアペアランスパラメータであり，この$\lambda_i$の値であらゆるアペアランス$\mathbf{A(x)}$を近似的に表現することができる．そして$\mathbf{A}_i$は正規直交ベクトルである．

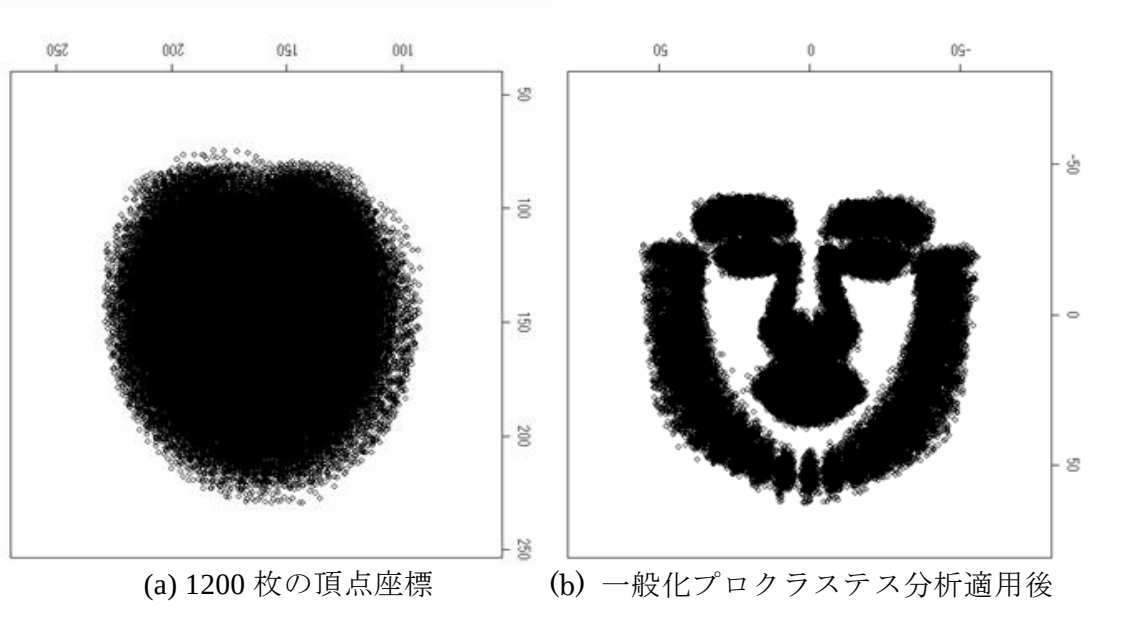
## 2.4. モデル生成事例

本節では上記のモデルの概説を基に実際にモデル生成の事例を示す．HOIP 顔画像データベース**[8]**を用い，**図 2.1** には HOIP 顔画像データベースの男女 20 代から 70 代までのサンプル画像を示す．1200 枚の顔画像に対して頂点を手動で打ち，テキストファイルにその座標を保存する．ここでは 1 つの顔画像に対して 120 点の座標を採取する．また，それぞれの顔画像は上下，左右約 30 度までの顔向きの変動を含んでいる．

**図 2.1**：HOIP 顔画像データベースの 20 代から 70 代の男女のサンプル画像

### 2.4.1. 形状モデルの生成

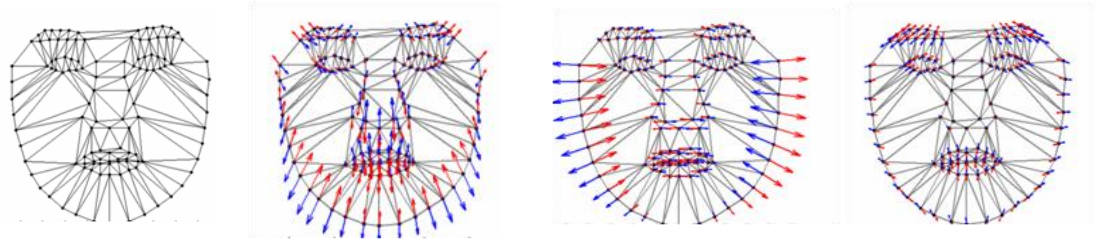
本項では 2.2 節で述べた形状モデル作成法に従い，一般化プロクラステス分析と PCA の適用例を示す．始めに生成される 1200 枚の学習画像の頂点座標群に対して，一般化プロクラステス分析を施した結果を**図 2.2** に示す．




(a) 1200 枚の頂点座標　　(b) 一般化プロクラステス分析適用後の頂点座標

**図 2.2**：プロクラステス分析による正規化の結果

次にプロクラステス分析適用後の頂点座標に対して，PCA を施すことでメッシュ状の形状モデルを生成できる．**図 2.3** に頂点数が 120 点の場合における形状メッシュを示す．



$\mathbf{s}_0$　　$\mathbf{s}_1$　　$\mathbf{s}_2$　　$\mathbf{s}_3$

**図 2.3**：AAM 形状メッシュ

**図 2.3** の左から 1 番目が平均形状$\mathbf{s}_0$である．他は結合係数を$3\sigma_i = \pm 3\sqrt{p_i}$とし，第 1 から第 3 主成分をそれぞれ別々に平均形状$\mathbf{s}_0$と線形結合した結果であり，その変化量を矢印で表現している．$\mathbf{s}_1$は顔の上下方向の動き，$\mathbf{s}_2$は顔の横方向の動きが抽出されていることが確認できる．

### 2.4.2. アペアランスモデルの生成

本項では，2.3 節で述べたアペアランスモデル作成法に従い，PCA の適用例を示す．アペアランスモデルはカラー画像をグレイスケール画像に変換し，メッシュ内部のテクスチャを平均形状内に納まるように正規化する．正規化したメッシュ内のテクスチャに対して PCA を施した結果を**図 2.4** に示す．左から 1 番目は平均テクスチャ$\mathbf{A}_0(\mathbf{x})$である．他は結合係数を$3\sigma_i = \pm 3\sqrt{\lambda_i}$とし，第 1 から第 3 主成分をそれぞれ別々に平均テクスチャ$\mathbf{A}_0(\mathbf{x})$と線形結合した結果であり，第一主成分$\mathbf{A}_1(\mathbf{x})$では結合係数の変化は眉の濃さの変化に関係することが確認でき，また第 2 主成分$\mathbf{A}_2(\mathbf{x})$は口周辺の皺の深さの変化に関係していると言える．このように学習画像に含まれる濃淡値の代表的な特性をアペアランスモデルとして扱う．

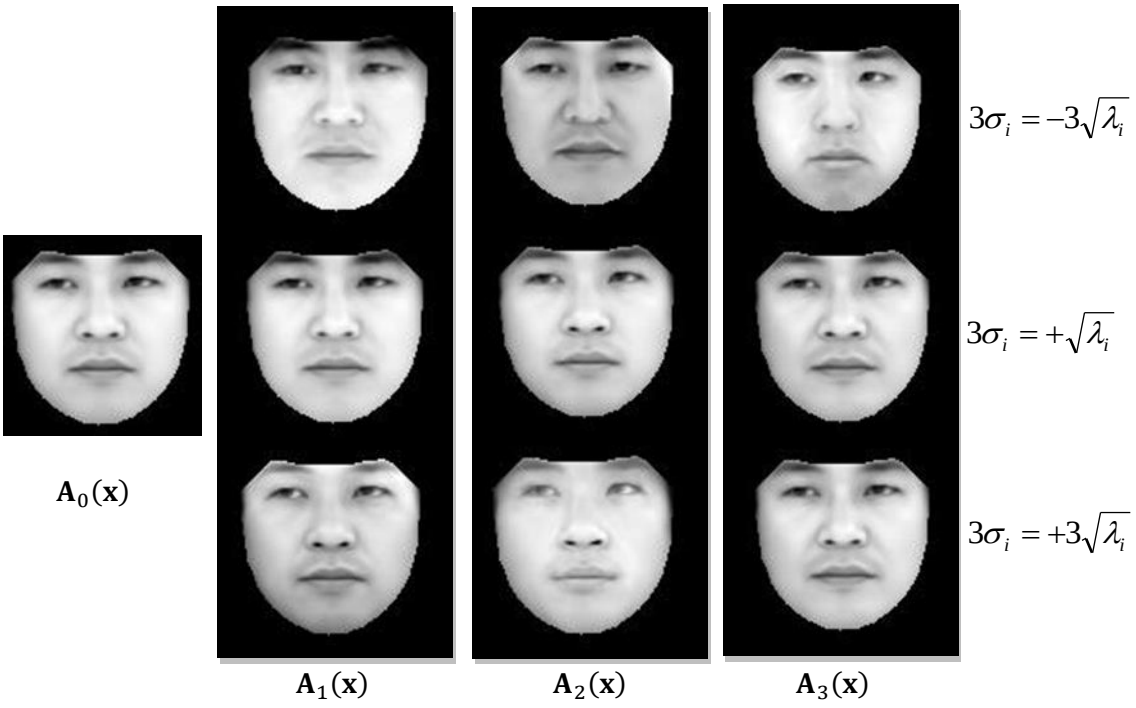



**図 2.4**：結合係数を変化させたときのアペアランスモデル

## 2.5. フィッティング

AAM のフィッティングは画像$I(\mathbf{x})$が与えられたとき，形状パラメータ$\mathbf{p}$とアペアランスパラメータ$\boldsymbol{\lambda}$の最適化として定義できる．$\mathbf{x}$が$\mathbf{s}_0$内のピクセルとする場合，入力画像$I(\mathbf{x})$におけるピクセルは$\mathbf{W}(\mathbf{x};\mathbf{p})$と一致する．ここで$\mathbf{W}$は形状メッシュ内の 3 点から構成されたポリゴンのピース毎のアフィンワープである．フィッティングの誤差関数は式（2.4）の二乗和誤差の最小化として与えられる：

$$\sum_{\mathbf{x}\in\mathbf{s_0}}\left[I\big(\mathbf{W}(\mathbf{x};\mathbf{p})\big)-\mathbf{A}_0(\mathbf{x})+\sum_{i=1}^{m}\lambda_i\,\mathbf{A}_i(\mathbf{x})\right]^2 \tag{2.4}$$

### 2.5.1. Lucas-Kanade アルゴリズム

最適化法として勾配降下法による画像内の位置合わせ法である Lucas-Kanade アルゴリズムについて述べる**[4]**．不変のテンプレート$\mathbf{A}_0(\mathbf{x})$を用いて最適化を効率的に行うとすると，式（2.4）は以下のように再定義される：

$$\sum_{\mathbf{x}}\big[I\big(\mathbf{W}(\mathbf{x};\mathbf{p})\big)-\mathbf{A}_0(\mathbf{x})\big]^2 \tag{2.5}$$

Lucas-Kanade アルゴリズムは反復的にパラメータ$\mathbf{p}$に$\Delta\mathbf{p}$を加算していくことで，誤差関数の最小化を行う．この時，誤差関数は以下の式で与えられる：

$$\sum_{\mathbf{x}}\big[I\big(\mathbf{W}(\mathbf{x};\mathbf{p}+\Delta\mathbf{p})\big)-\mathbf{A}_0(\mathbf{x})\big]^2 \tag{2.6}$$

また，パラメータ$\mathbf{p}$は$\mathbf{p}\leftarrow\mathbf{p}+\Delta\mathbf{p}$より更新される．式（2.6）の 1 次のテイラー展開は：

$$\sum_{\mathbf{x}}\left[I\big(\mathbf{W}(\mathbf{x};\mathbf{p})\big)+\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\Delta\mathbf{p}-\mathbf{A}_0(\mathbf{x})\right]^2 \tag{2.7}$$

ここで, $\partial\mathbf{W}/\partial\mathbf{p}$は$\mathbf{W}$についてのヤコビアン, $\nabla I$は画像$I$の勾配である. 更に式(2.7)を$\Delta\mathbf{p}$について偏微分すると：

$$\Delta\mathbf{p}=\mathbf{H}^{-1}\sum_{\mathbf{x}}\left[\nabla I\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\right]^{\mathrm{T}}\big[\mathbf{A}_0(\mathbf{x})-I\big(\mathbf{W}(\mathbf{x};\mathbf{p})\big)\big] \tag{2.8}$$

$\mathbf{H}$はヘッセ行列であり，式（2.8）より$\Delta\mathbf{p}$が算出でき，これを用いてパラメータを反復的に更新する.

### 2.5.2. Compositional アルゴリズム

Compositional アルゴリズムの概要について述べる. Lucas-Kanade アルゴリズムでは$\Delta\mathbf{p}$を求めることで AAM のパラメータを更新していたが，Compositional アルゴリズムでは既知のワープ$\mathbf{W}(\mathbf{x};\mathbf{p})$を用い，未知の増加パラメータ$\Delta\mathbf{p}$におけるワープ$\mathbf{W}(\mathbf{x};\Delta\mathbf{p})$を計算することにより誤差関数を最小化することが目的となる．このとき誤差関数は以下の式で与えられる：

$$\sum_{\mathbf{x}}\left[I\left(\mathbf{W}(\mathbf{W}(\mathbf{x};\Delta\mathbf{p});\mathbf{p})\right)-\mathbf{A}_0(\mathbf{x})\right]^2 \tag{2.9}$$

また，既知のワープと増加分のワープを用いて更新は以下の式で与えられる：

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p}) \tag{2.10}$$

### 2.5.3. Inverse Compositional アルゴリズム

Inverse Compositional アルゴリズムは画像とテンプレートの役割を置き換えることで, テンプレート$\mathbf{A}_0(\mathbf{x})$を基準にして入力画像$I\left(\mathbf{W}(\mathbf{x};\mathbf{p})\right)$を逆変換したワープの度合いから, $\mathbf{W}(\mathbf{x};\mathbf{p})$を反復的に更新する．誤差関数は，以下の式で与えられる：

$$\sum_{\mathbf{x}}\left[\mathbf{A}_0\left(\mathbf{W}(\mathbf{x};\Delta\mathbf{p})\right)-I\left(\mathbf{W}(\mathbf{x};\mathbf{p})\right)\right]^2 \tag{2.11}$$

また，更新は以下の式で与えられる：

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})^{-1} \tag{2.12}$$

Inverse Compositional アルゴリズムは画像とテンプレートの役割を置き換えることで勾配画像$\nabla\mathbf{A}_0$は不変となり，勾配の計算を反復処理の前に計算することができる．これより計算コストを削減でき，Compositional アルゴリズムと比較し，計算コストを抑えた効率的なフィッティングが可能である．

## 2.6. 特長と課題

上で述べたように AAM は予め作成したモデルと対象物体の二乗和誤差の最小化であり, その最適化法は複数存在する．特に Inverse Compositional アルゴリズムは非常に効率的な手法であり, 動画像への適用も可能である. ここで AAM の特長と課題を以下にまとめる．

- **特長**
  - 対象物の形状とアペアランス情報をパラメータ化でき，高次元の情報を少数次元のベクトルとして特徴抽出できる．
  - $\mathbf{s}_0$内へ変換したテンプレート画像を正規化画像として扱うことで，他の特徴抽出手法の前処理として応用できる．
  - 学習画像内の人物であれば顔の復元が可能であり，犯罪捜査等への応用が期待できる．

- **課題**
  - 極端に初期座標に依存する. モデルの中心座標と対象物体の初期の距離に敏感であり，フィッティングに失敗しやすい．経験的に AAM メッシュと対象物体が半分以上被る程度の初期位置に設定する必要がある．

- 誤差関数内のテンプレート$\mathbf{A_0(x)}$は更新されないので，形状モデルのみの更新による最適化となり，学習画像以外の入力画像が与えられた際，濃淡値の違いによりメッシュが対象に収束しにくい．

## 2.7. まとめ

本章では，本稿における顔画像の正規化や特徴量抽出の基盤となる技術の AAM についての概要を述べた．具体的には AAM の形状・アペアランスモデル，フィッティングのための最適化法について説明し，AAM の特長と課題を整理した．またアルゴリズムの欠点を補った手法であり，実際に本稿内の顔の属性分類に適用した手法である Generic AAM については次の章で述べる．

## 参考文献

[1] T.F. Cootes，J.E. Gareth， and J.T. Christopher :“Active Appearance Models”，IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.6, pp.681-685（2001）.

[2] I. Matthews，and S. Baker :“Active Appearance Models Revisited”，International Journal of Computer Vision，vol.60，no.2，pp.135-164（2004）.

[3] S. Baker，R. Gross，and I. Matthews :“Lucas-Kanade 20 Years On: A Unifying Framework: Part 3” Tech. Report CMU-RI-TR-03-35，Robotics Institute, Carnegie Mellon University，November（2003）.

[4] S. Baker，and I. Matthews :“Lucas-Kanade 20 Years on: A Unifying Framework”，International Journal of Computer Vision，vol.56，no.3，pp.221-255（2004）.

[5] I. Matthews, J. Xiao, and S. Baker :“2d vs. 3d Deformable Face Models: Representational Power, Construction, and Real-Time Fitting”，International Journal of Computer Vision，vol.75，no.1，pp. 93-113（2007）.

[6] T.F. Cootes :“Statistical Models of Appearance for Computer Vision”，Online technical report available from http://www.isbe.man.ac.uk/˜bim/refs.html, Sept. （2001）.

[7] S. Wold，K. Esbensen， P. Geladi :“Principal Component Analysis”，Chemometrics and Intelligent Laboratory Systems，vol.2，no 1，pp.37-52 (1987).

[8] 財団法人ソフトピアジャパン HOIP 顔画像データベース http://www.softopia.or.jp/

# 第3章. Generic AAM (GAAM)

## 3.1. まえがき

本章では，第 2 章で説明した 2 つの AAM の課題（Ⅰ．極端に初期座標に依存する問題．Ⅱ．学習画像以外の入力画像が与えられた際，メッシュが対象に収束しにくい問題）についての改善法を示す．まず TakumiVision 株式会社製の顔検出ライブラリを導入し，顔の初期位置を補正することで，課題Ⅰの解決に取り組む**[1]**．次に特定人物に依存せず，対象人物へのメッシュの収束が可能な R. Gross により提案された Generic AAM (GAAM)を導入することで，課題Ⅱの解決を図る**[2]**．本章の構成としては，まず GAAM について述べ，次に実験において従来の AAM との性能の比較検証を行う．

## 3.2. Generic AAM 概説

Generic AAM の特長は，形状パラメータ$\mathbf{p}$と同様にアペアランスパラメータ$\boldsymbol{\lambda}$の反復的な更新により，誤差関数内のテンプレート$\mathbf{A}(\mathbf{x})$を入力顔画像$\boldsymbol{I}$と類似したテンプレートへと更新することで，フィッティング性能の向上を期待できる点である．このとき誤差関数は以下の式で与えられる：

$$\sum_{\mathbf{x}}\left[\mathbf{A}_0\big(\mathbf{W}(\mathbf{x};\Delta\mathbf{p})\big)+\sum_{i=1}^{m}(\lambda_i+\Delta\lambda_i)\,\mathbf{A}_i\big(\mathbf{W}(\mathbf{x};\Delta\mathbf{p})\big)-\boldsymbol{I}\big(\mathbf{W}(\mathbf{x};\mathbf{p})\big)\right]^2 \tag{3.1}$$

また，式(3.1)の 1 次のテイラー展開は：

$$\sum_{\mathbf{x}}\left[\begin{array}{c}\mathbf{A}_0(\mathbf{x})+\nabla\mathbf{A}_0\dfrac{\partial\mathbf{W}}{\partial\mathbf{p}}\,\Delta\mathbf{p}+\\ \displaystyle\sum_{i=1}^{m}(\lambda_i+\Delta\lambda_i)\left(\mathbf{A}_i(\mathbf{x})+\nabla\mathbf{A}_i\dfrac{\partial\mathbf{W}}{\partial\mathbf{p}}\,\Delta\mathbf{p}\right)-\boldsymbol{I}\big(\mathbf{W}(\mathbf{x};\mathbf{p})\big)\end{array}\right]^2 \tag{3.2}$$

ここで，勾配方向を示す最急降下画像$\boldsymbol{SD}(\mathbf{x})$は：

$$\boldsymbol{SD}(\mathbf{x}) = \left( \left( \nabla \mathbf{A}_0 + \sum_{i=1}^{m} \lambda_i \nabla \mathbf{A}_i \right) \frac{\partial \mathbf{W}}{\partial p_1}, \ldots, \left( \nabla \mathbf{A}_0 + \sum_{i=1}^{m} \lambda_i \nabla \mathbf{A}_i \right) \frac{\partial \mathbf{W}}{\partial p_n}, \right. \\ \left. \mathbf{A}_1(\mathbf{x}), \ldots, \mathbf{A}_m(\mathbf{x}) \right) \tag{3.3}$$

$\Delta\boldsymbol{\lambda} = (\Delta\lambda_i, \ldots, \Delta\lambda_m)^{\boldsymbol{T}}$として$\mathbf{k} = \begin{pmatrix} \mathbf{p} \\ \boldsymbol{\lambda} \end{pmatrix}$，$\Delta\mathbf{k} = \begin{pmatrix} \Delta\mathbf{p} \\ \Delta\boldsymbol{\lambda} \end{pmatrix}$と定義する．$\mathbf{k}$は$n+m$次元の列ベクトルであり，$\Delta\mathbf{k}$は以下の式より与えられる：

$$\Delta\mathbf{k} = -\mathbf{H}^{-1} \sum_{\mathbf{x}} \boldsymbol{SD}(\mathbf{x})^{\mathbf{T}} \mathbf{E}(\mathbf{x}) \tag{3.4}$$

式(3.4)における$\mathbf{H}$はヘッセ行列であり，$\mathbf{E}(\mathbf{x})$はアペアランスモデル$\mathbf{A}(\mathbf{x})$と入力画像$\boldsymbol{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$の差分画像である．**図 3.1** に顔検出器を導入した Generic AAM の更新アルゴリズムの疑似コードを示す.

**Pre-Computation**：

p1) 勾配画像$\nabla\mathbf{A}_0, \nabla\mathbf{A}_i \ \ for\ i = 1, \ldots, m$の計算

p2) ヤコビアン$\partial\mathbf{W}/\partial\mathbf{p}$の計算

p3) 顔検出により，メッシュの初期座標，スケールを取得

**Iteration**：

i1) $\mathbf{W}(\mathbf{x};\mathbf{p})$を使い，ワープ画像$\boldsymbol{I}(\mathbf{W}(\mathbf{x};\mathbf{p}))$を計算

i2) 差分画像$\mathbf{E}(\mathbf{x})$を計算

i3) 式(3.3)を使い，最急降下画像$\boldsymbol{SD}(\mathbf{x})$を計算

i4) 式(3.4)を使い，$\Delta\mathbf{k}$を計算

i5) パラメータの更新．$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})^{-1}$，$\boldsymbol{\lambda} = \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$

**図 3.1**：Generic AAM の更新アルゴリズムの疑似コード

## 3.3. 実験及び考察

AAM と Generic AAM の非学習顔画像に対するフィッティング性能の比較実験を行う．PCA の主成分抽出より求まる形状・アペアランスモデルは，AAM と Generic AAM それぞれ同じモデルを用いる．このとき 2.4 節における HOIP 顔画像データベース（HOIP DB）**[3]** から生成したモデルを利用する．HOIP DB は，男女の約 20 代から 70 代までの 100 人×12 枚，合計 1,200 枚から構成され，**図 2.1** で示したように各顔画像は上下，左右約 30 度までの顔向きの変動を含む．また AAM の顔メッシュは 120 の頂点から構成される．次にテス

ト画像のデータセットには, 独自で採取した HOIP DB とは別の顔画像データセットを用いる．それは一人当たり 3~5 枚の 16 人分，合計 70 枚から構成され，顔の向きや照明の変動を含む．**図 3.2** にテスト顔画像データセットのサンプルを示す.

**図 3.2**：テスト画像データセットのサンプル

本実験では，画像サイズは学習，テスト共に 320×240pix.であり，次元削減のためカラー画像をグレイスケール画像に変換してモデル作成を行う．また照明変動の影響を考慮し，学習モデルとテスト画像の平均・分散を一定にする正規化処理を施す.

AAM と Generic AAM は共に，形状とアペアランスのパラメータにおける次元数を予め設定する必要がある．そこで良好な次元数設定のため，それを経験的に決定する．**表 3.1** に設定した各パラメータの次元数を示す.

**表 3.1**：AAM と Generic AAM における各パラメータの次元数

| | 形状パラメータの次元数 | アペアランスパラメータの次元数 |
|---|---|---|
| **AAM** | 3 | 40 |
| **Generic AAM** | 6 | 25 |

フィッティング率を以下の評価式より定式化する：

$$\text{フィッティング率}(\%) = \frac{\text{成功画像枚数}}{\text{全テスト画像枚数}} \times 100 \tag{3.5}$$

ここで，本実験ではテスト画像に 70 枚使用する．定量的な評価を行うために，テンプレート$\mathbf{A(x)}$内のピクセル数を$\boldsymbol{N}$とすると，エラー画像$\mathbf{E(x)}$を用いてフィッティング誤差は以下の式で定式化できる：

$$\boldsymbol{Fitting\ Error} = \left\{ \sum_{\mathbf{x}} \left( \mathbf{E(x)} \times \mathbf{E(x)} \right) \right\} / \boldsymbol{N} \tag{3.6}$$

式(3.6)のフィッティング誤差が予め設定した閾値以下であれば，フィッティング成功とする．閾値は実験的に決定し，フィッティング成功の条件として AAM メッシュが発散していない場合，かつ各顔の器官を構成する特徴点がそれぞれ適切な器官にフィットしている場合とする．この条件を基に閾値は 177 と設定する．

ここで**図 3.3** にフィッティング失敗例を示す．**図 3.3** から目の特徴点が眉にフィットしている場合，失敗と判定されることを確認できる．また**表 3.2** にフィッティング率の実験結果を示し，**図 3.4** において実験結果のサンプル画像を示す．左列は AAM，右列は Generic AAM のフィッティング結果である．

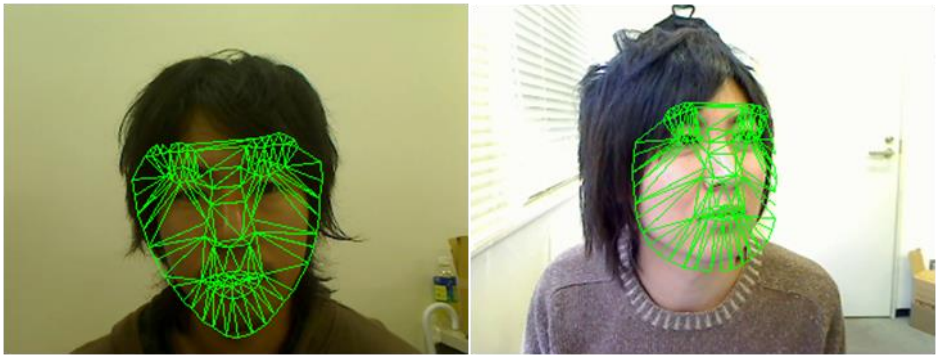


**図 3.3**：Generic AAM のフィッティング失敗例

**表 3.2**：AAM と Generic AAM の各フィッティング率

| 手法 | フィッティング率（%） |
|---|---|
| **AAM** | 18.6 |
| **Generic AAM** | 80.0 |

**図 3.4**：AAM（左列）と Generic AAM（右列）のフィッティング結果画像

**表 3.2** から AAM のフィッティング率は 18.6%と低い性能である．これは顔の傾きによる初期位置のズレ，不変のテンプレートによる照明変動や非学習の入力顔画像に対応できない汎化性の乏しさが原因であると考えられる．

また AAM を顔の向き・傾きの悪条件が存在しない正面顔画像に対してテストした場合，特定人物の画像に対してのみフィッティングが成功する．一方 Generic AAM の場合，対象人物が誰であろうと正面顔画像のフィッティング率は 90%以上となる．この結果から，フィッティングアルゴリズムを改良することで照明変動に対する頑強性，非学習の入力顔画像に対する汎化性の向上を確認できる．また Generic AAM での正面顔画像における失敗例として，**図 3.3** で示すような髪の毛が眉にかかり，眉を目として誤認識する例が挙げられる．このことから隠れ等のオクリュージョンが存在するとフィッティングが難しくなると言える．

次に AAM を顔の向きや傾きの条件を含む画像に対してテストした場合，ほとんどの画像でフィッティングに失敗し，一瞬でメッシュが発散してしまう．一方 Generic AAM の場合，**図 3.4** から目，鼻や口などに対する顔メッシュの正確なフィッティングを確認できる．これはアルゴリズムの改善に加え，顔検出器の導入により，初期位置の影響を受けるリスクを低減できたことが要因だと言える．

実験結果を総括すると，Generic AAM の非学習のテスト画像に対するフィッティング率は 80%であり，従来の AAM と比較し，汎化性や顔の向き・傾きに対する頑強性が向上している．これより Generic AAM の有効性を確認できる．

## 3.4. まとめ

本章では，本稿において実際に顔の属性分類に適用した手法である Generic AAM についての概要を述べた. また独自に採取した顔画像データセットを用い, 従来の AAM と Generic AAM のフィッティング性能の比較実験を行った．実験より Generic AAM のフィッティング率は 80%に到達しており，従来と比較して 60%以上の改善が見られ，Generic AAM の有効性を確認できた．

## 参考文献

[1] Takumi Vision 株式会社 顔検出ライブラリ http://www.takumivision.co.jp/

[2] R. Gross，I. Matthews，and S. Baker : “Generic vs. Person Specific Active Appearance Models”，Image and Vision Computing，vol.23，no.12，pp.1080-1093 (2005).

[3] 財団法人ソフトピアジャパン HOIP 顔画像データベース http://www.softopia.or.jp/

# 第4章. AAM を用いた性別分類

## 4.1. まえがき

本章では，Generic AAM(GAAM)の顔メッシュ座標とアペアランスパラメータを使い，独自の特徴量を提案し，それを用いた性別分類アルゴリズムについて述べる．また HOIP 顔画像データベース**[1]**を用いてその性能を検証する．本章の構成として，まず 4.2 節にて提案手法のフレームワークについて述べる．次に，4.3 節で男女間の差を分析し，顔器官の形状，サイズや比率の情報を含む形状特徴量，唇の色，肌の色や質感などの情報を含むテクスチャ特徴量を示す．また 4.4 節では確率モデルによる分類器であり，設計が容易な単純ベイズ分類器について説明する．そして 4.5 節にて提案手法の性能の評価実験を行い，最後に 4.5 節で本章をまとめる．

## 4.2. 性別分類アルゴリズムの概要

本節では GAAM による顔特徴を利用した性別分類アルゴリズムについて述べる．提案手法のフレームワークを**図 4.1** に示す．

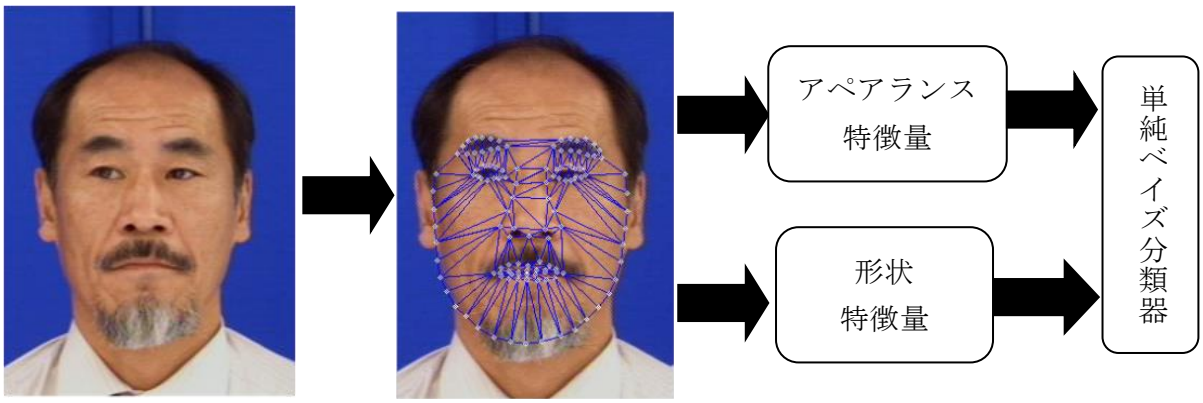



**図 4.1**：提案する性別分類アルゴリズムのフレームワーク

**図 4.1** の提案手法は 3 つのステップから構成される．まず，顔検出器を使い，顔画像内の瞳についての座標を取得する.そして左右の瞳の位置を基準に顔の初期座標を決定する．次に，式(3.1)の誤差関数をパラメータ**λ**，**p** について最適化することで顔メッシュを入力顔画像に収束させる．この時，アペアランスパラメータ**λ**はテクスチャ情報を含む特徴量として扱う．またメッシュを構成する頂点間の形，大きさ，比率を基にして形状特徴量を算出する．詳細については 4.3 節で述べる．最後に，これらの特徴量を用い，単純ベイズにより性別を分類する．

## 4.3. 顔の特徴量

本節では，男女間における顔の特徴差を基に，顔メッシュを構成する複数の頂点座標の関連性より定義できる形状特徴量について述べる.また AAM のパラメータ**λ**を利用したアペアランス特徴量についても示す．ここで，**表 4.1** にて男女間における顔の特徴差の一覧を示す．**表 4.1** の情報を基に性別分類において有効な特徴量の設計を行う．

**表 4.1**：顔特徴の男女による差

| 顔部位 | 男性 | 女性 |
|---|---|---|
| 目 | 細い | 大きい |
| 眉 | 濃く太い | 細い |
| 口 | 顔の横幅に対し大きい | 唇が濃い |
| 鼻 | 大きく幅が広い | 小さい |
| 頬 | 黒め | 面積が大きく，白く明るい |
| ヒゲ | 濃い | なし |

**表 4.1** を基に男女間の差を考慮し，顔の各器官の位置，サイズや比率についての形状特徴量を独自に設計する．**表 4.2** では設計した形状特徴量をまとめている．形状特徴量は顔メッシュが完全に収束した状態でのメッシュの各頂点座標を用いる．またスケールサイズの正規化のため，形状特徴量はメッシュの収束時に算出されるスケールパラメータ**[2]**を使い，平均形状$\mathbf{s}_0$のスケールサイズを基準に正規化する．

**表 4.2**：GAAM による形状特徴量

| パターン | 詳細 |
|---|---|
| **顔の幅** | 口の両端の真横に位置する輪郭上の点を結ぶ直線の距離. |
| **目と顎の距離** | 右目の内端から顎の先端までの距離. |
| **目と鼻との面積** | 小鼻の両端と両目の外端から構成される4点内の面積. |
| **頬の面積** | 右目の下と両端の3点，小鼻の右端の1点，口の右端の真横に位置する輪郭上の1点から構成される5点内の面積. |
| **輪郭と口の距離** | 口の右端に位置する点と，口の右端の真横に位置する輪郭上の点を結ぶ直線の距離. |
| **目と眉の高さの比** | （鼻の先端から両目頭の中心までの距離）/ (鼻の先端から眉頭間の中心の距離) |
| **鼻の幅** | 小鼻の両端の点における距離 |

アペアランス特徴量は，GAAM のアペアランスパラメータ$\boldsymbol{\lambda}$の第 1 から第 6 成分までを用いる．**図 4.2**(a)，(b)はそれぞれ平均テクスチャ$\mathbf{A}_0(\mathbf{x})$に結合係数を$3\sigma_i = \pm 3\sqrt{\lambda_i}$とし，第 1 主成分$\mathbf{A}_1(\mathbf{x})$を加算したモデル, **図 4.2**(c), (d)は第 2 主成分$\mathbf{A}_2(\mathbf{x})$を加算したモデルである．結合係数値の違いにより唇の色や, 頬の色, ヒゲといった性別毎に特徴の違いを確認でき，それらはアペアランス特徴量として扱うことができる．

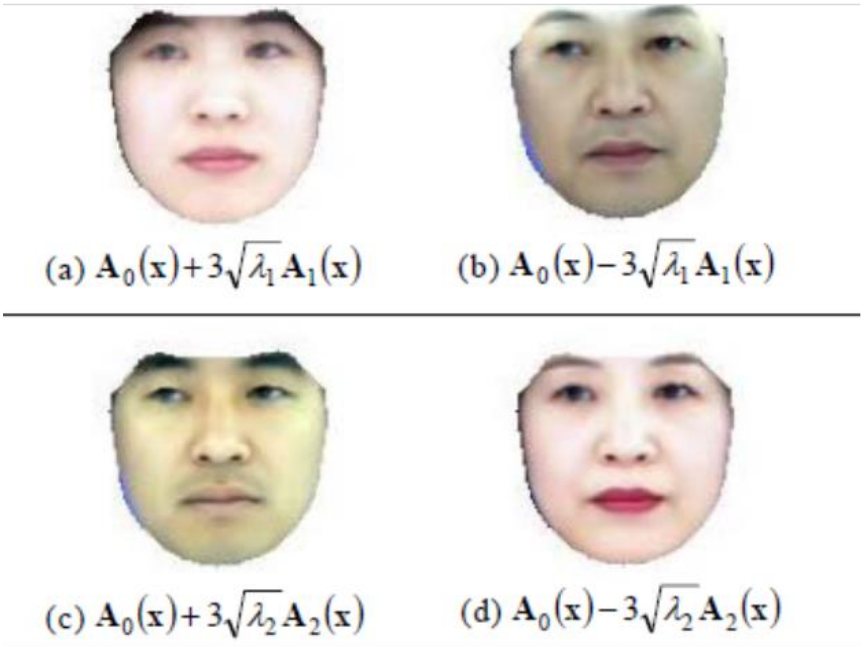



**図 4.2**：アペアランスモデルに含まれる男女の特徴差

## 4.4. 単純ベイズ分類器

生成的手法である単純ベイズ分類器を用い，提案する性別分類アルゴリズムを確率的に性能評価する．それはクラス$C_i$に対する事前確率$P(C_i)$と共に，$P(\mathbf{x}|C_i)$で与えられるクラスで条件付けされた確率密度を生成し，ベイズの定理を用いて事後確率の最大化として定義できる：

$$P(C_i|\mathbf{x}) = \frac{P(\mathbf{x}|C_i)P(C_i)}{P(\mathbf{x})} \tag{4.1}$$

ここで$P(\mathbf{x})$は，$P(C_i|\mathbf{x})$の合計を 1 にするためのスケーリング要素であり，以下の式で定義する：

$$P(\mathbf{x}) = \sum_{\mathrm{i}=1}^{\mathrm{M}} P(\mathbf{x}|C_i)P(C_i) \tag{4.2}$$

## 4.5. 実験及び考察

提案する性別分類アルゴリズムの実験結果について示す．ここで**表 4.3** にテスト画像に用いた HOIP 顔画像データベース(HOIP DB)の年代毎の内訳を示す．それは 17 歳以下から 64 歳までの男性計 89 枚，女性計 95 枚の画像を用いている．また**表 4.4** にて，GAAM モデル構築の際，PCA の主成分抽出に用いた画像における年代毎の内訳を示す．**表 4.3** より，HOIP DB は若い男女の画像枚数は少なく，年代毎に枚数のバラつきが生じることを確認できる．

**表4.3**：テスト画像の年代毎の内訳

| 年齢<br>（歳） | 男性<br>（枚） | 女性<br>（枚） |
|---|---|---|
| **-17** | 8 | 2 |
| **18−25** | 8 | 16 |
| **26−40** | 27 | 30 |
| **41−55** | 27 | 28 |
| **56−64** | 19 | 19 |

**表4.4**：PCAの主成分抽出に用いた画像における年代毎の内訳

| 年齢<br>（歳） | 男性<br>（枚） | 女性<br>（枚） |
|---|---|---|
| **-17** | 5 | 6 |
| **18−25** | 12 | 12 |
| **26−40** | 28 | 25 |
| **41−55** | 18 | 23 |
| **56−64** | 11 | 16 |

### 4.5.1. 実験環境

実験環境を**表 4.5** にまとめる.

**表 4.5**：実験環境

| | |
|---|---|
| OS | Windows XP Professional SP3 |
| CPU | Intel(R) Core(TM)2 Quad CPU Q9450 2.66GHz |
| メモリー | 3.25GB RAM |
| 開発言語 | C 言語 |
| 開発環境 | Visual Studio 2008 |

### 4.5.2. 実験結果・考察

実験では以下の 3 つの評価項目を性能検証に用いる．それは，①「男女それぞれの再現率・適合率」②「顔検出後から結果を出力するまでの処理速度」③「顔は年代毎に違った特徴を見せるので年代毎の再現率」の 3 項目である．再現率・適合率は 2 クラス（正クラス，負クラス）に分類するとき，以下の式より定義できる：

$$\text{再現率}(\%) = \frac{(\text{正クラスに正しく分類された画像枚数})}{(\text{正クラスの全画像枚数})} \tag{4.3}$$

$$\text{適合率}(\%) = \frac{(\text{正クラスに正しく分類された画像枚数})}{(\text{正クラスに分類された画像枚数})} \tag{4.4}$$

再現率は正クラスが男性であるとした場合，男性画像のうちで男性と認識された割合であ

り，適合率とは男性と認識された画像の内で実際に男性である割合を意味する．
ここで GAAM の形状パラメータの次元数を 8，アペアランスパラメータの次元数を 36 に実験的に設定する．そしてアペアランスパラメータの上位 6 次元をアペアランス特徴量とする．**表 4.6** に男女毎の再現率，適合率，処理速度の結果を示し，**図 4.3** では各年代の再現率をグラフとして表している．

**表 4.6**：再現率・適合率・処理速度

| | 男性 | 女性 |
|---|---|---|
| 再現率（%） | 87.6 | 94.7 |
| 適合率（%） | 93.98 | 89.1 |
| 結果出力までの時間 | 244 ms | |

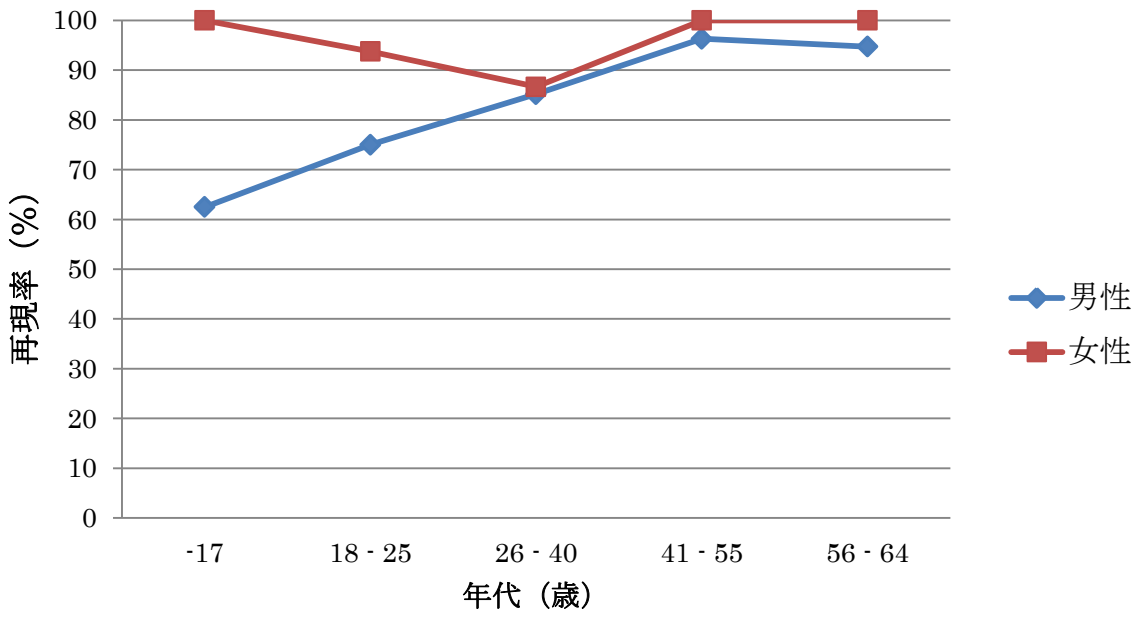



**図 4.3**：年代毎の再現率

**表 4.6** より処理速度は 244ms であり，映像への適用を考えた場合，今後高速化が必要である．また適合率は男女共に優れた数値を示しており，提案アルゴリズムの有効性を確認できる.そして再現率の結果から,女性は男性と比較して再現性が高いことを確認できる．また**図 4.3** の年代毎の再現率のグラフから，本章で示した特徴量では若い男性を十分に識別できないと言える．これは若い男性の顔特徴は女性の特徴に非常に似通っていること，

主成分抽出に用いた画像において，-17 歳から 25 歳までの若い年代が他の年代と比較し，画像枚数が少ないことに起因していると考えられる．また本章では照明や顔の角度変化など含まない良質な環境で撮影された HOIP DB の画像を使い, 提案アルゴリズムの性能を評価した．しかし実環境への適用を考慮した場合，照明や顔の向きに頑強な手法へと発展させることが望ましい．ゆえに以降の第 5，6，7 章において，これら問題の解決策について検討する．

## 4.6. まとめ

本章では GAAM を用いた性別分類アルゴリズムを提案した．提案アルゴリズムは独自に定義した形状特徴量と GAAM のアペアランスパラメータを特徴量として採用した．実験では再現率・適合率，結果出力までの処理速度，年代毎の再現率に焦点を当て，性能の検証を行った．適合率は男性 93.98%，女性 89.1%であり，その有効性を確認できた．また共に年齢が高いほど再現率が高くなり，特に女性は全年代において再現率が男性と比較し，高いことを確認できた．しかし処理速度については 244 ms であり，映像への適用を考えると不十分であると言える．

## 参考文献

[1] 財団法人ソフトピアジャパン HOIP 顔画像データベース http://www.softopia.or.jp/

[2] I. Matthews，and S. Baker：“Active Appearance Models Revisited”，International Journal of Computer Vision，vol.60，no.2，pp.135-164（2004）.

# 第5章. 顔の特徴量抽出法

## 5.1. まえがき

4 章では AAM を用いた性別分類アルゴリズムについて述べた．しかし実験では照明や顔向きの変動などを含まない良質な環境で撮影された HOIP DB の顔画像を使い, アルゴリズムの性能を評価した．仮に照明変動などの条件が付加されると，4 章で提案した手法では分類性能の低下を招くと考えられる．そこで特徴量の改良を行うことでそれら悪条件に対して頑強性を高めることが必要である．本章ではロバスト性の高い 3 つの特徴量抽出法について紹介する．これは 6 章で述べる独自の顔の特徴量に関連する手法である．

## 5.2. 従来の特徴量抽出

本節では，ロバスト性の高い 3 つの特徴量抽出法について紹介する．これらは 6 章における年齢・性別分類などの顔画像解析に関連する代表的な特徴量である．一般的に特徴量は「幾何学的特徴量」と「アペアランス特徴量」の 2 つに大別できる．幾何学的特徴量は目，鼻や口などの顔器官の特徴点の位置座標を計算し，その特徴点間の相関関係などの形状に関する情報を特徴とする．一方でテクスチャ特徴量は特徴抽出フィルタを導入して顔の濃淡情報を特徴とする．本節では照明変動，位置ズレ誤差に頑強なテクスチャ特徴量を 3 つ紹介する．

### 5.2.1. Local Binary Pattern (LBP)

LBP は Ojala らにより提案され，顔の属性分類，認識や検出といった顔画像解析の特徴量として広く利用されている**[1]**．それは局所的なテクスチャの情報を保持しており，抽出されたパターンから構成されるヒストグラムは顔の描写にとって有効な特徴量となる．そして単調なグレイスケールの照明変化に頑強であるが，不規則な照明変化に脆弱であるといった課題がある．適用例として Fang らは PCA を使い，次元数を削減した低次元の LBP 特徴量を構築し，性別分類に応用している**[2]**.

以下では特徴量算出法について説明する．LBP は注目画素とその周辺に配置された画素との輝度差を利用する．これより単調な照明変化に関しては不変のテクスチャパターンと

なる．LBP は 2 つのステップで構成される．ステップ 1 では，注目画素$f_p$の周辺に位置する 8 つの画素$f_p$ $(p = 0, \dots, 7)$を閾値処理より 1 または 0 にラベリングする．それは以下の式で定式化できる：

$$S(f_p - f_c) = \begin{cases} 1, & f_p > f_c \\ 0, & f_p < f_c \end{cases} \tag{5.1}$$

ステップ 2 では，周辺画素を 2 進数から 10 進数に変換した値を，注目画素の値として算出する．それは式(5.2)として定式化でき，**図 5.1** にてその一連の手順を示す．

$$LBP = \sum_{p=0}^{7} S(f_p - f_c)\, 2^p \tag{5.2}$$

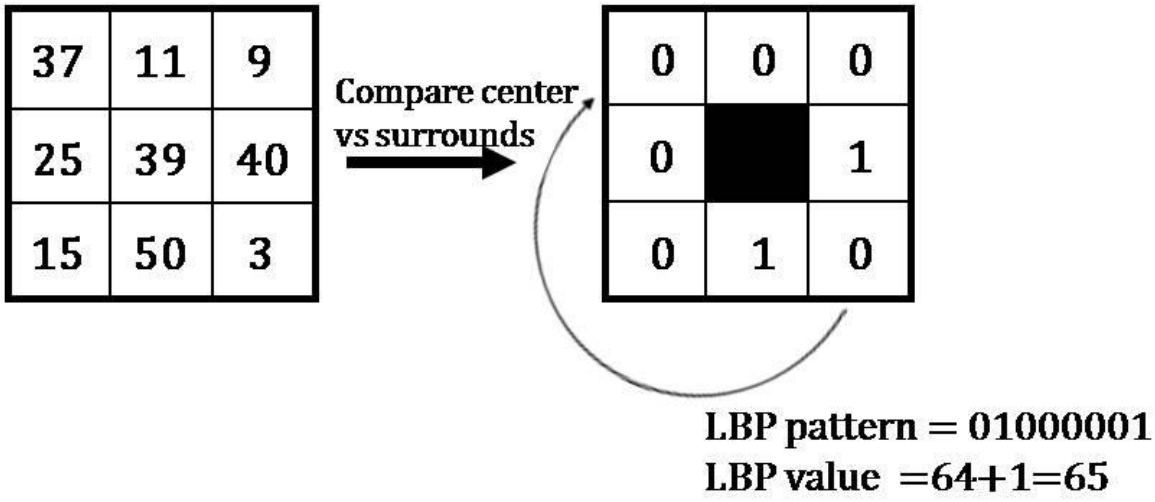



**図 5.1**：LBP の一連の手順

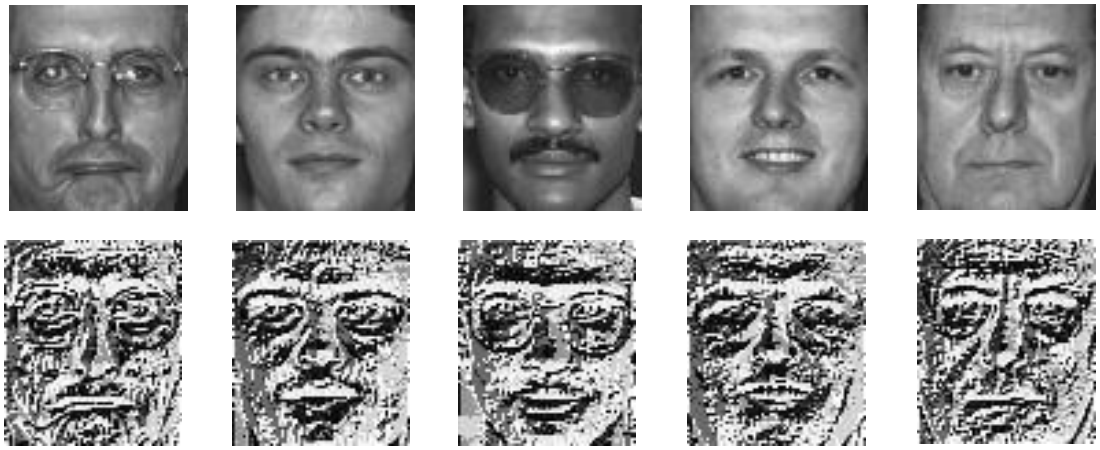


**図 5.2**：LBP を適用した顔画像

**図 5.2** に LBP を適用することで算出された画像を示す．**図 5.2**(上)は LBP 適用前の画像，**図 5.2**(下)は LBP 適用後の画像である．これよりテクスチャの値を符号化し，顔の特徴情報を高める効果が期待できる．

### 5.2.2. Gabor 特徴量

生物，特に脊椎動物の視覚情報処理の仕組みは種によらず基本的には同じ様式であり，眼から入った画像が網膜に投影され,そこから視覚野と呼ばれる大脳の部位に伝達される．視覚野は多層構造をしており，情報の流れとして，網膜に近い方から第一次視覚野，第二次視覚野のように呼ばれる．第一次視覚野の神経細胞は網膜から情報を受け取るが，単一の細胞は網膜上に映った画像のうち比較的狭い領域のみから情報を受け取り，この部分だけを処理して第二次視覚野へ情報を伝える．ここで行われる情報処理は，例えばその細胞が担当している領域の真ん中あたりに明るい部分があるかないかを判定するといった簡単なものである．従って，ある限られた大きさの領域に特定の単純なパターンが含まれているかどうかを判別する特徴抽出の機能を担っている．こうした情報処理は，工学的には特定のパターンにだけ反応する局所的なフィルタと考えることができる．処理される網膜上での領域と，フィルタとして抽出する特徴で細胞の特性は記述されることになり，この特性をその細胞の受容野と呼ぶ．この第一次視覚野の単純細胞の受容野特性は，ガボールフィルタでうまく近似されることが知られている．このフィルタを顔画像に適応し，得られた出力結果は個人の顔に対する特徴量として利用される．Gabor フィルタはガウス関数と正弦・余弦関数からなる関数であり，任意の周波数成分を抽出するフィルタリング機能を持つ．以下で基本的な Gabor フィルタについて述べる．

顔画像に対して Gabor フィルタを適用することで，顔の空間，および周波数領域における局所的な特徴を抽出することができる．一般に顔画像の濃度値情報は照明の変化などによって大きく変わってしまうが，Gabor フィルタを用いることによってその変化を最小限に抑えることができる．以下に Gabor フィルタの定義を示す：

$$\psi_{\mu,\nu}(\mathbf{x}) = \frac{\|k_{\mu,\nu}\|}{\sigma^2} e^{\left(-\|k_{u,v}\|^2 \|\mathbf{x}\|^2 / 2\sigma^2\right)} \left[e^{ik_{\mu,\nu}\mathbf{x}} - e^{-\sigma^2/2}\right] \tag{5.3}$$

ただし，$\mu$と$\nu$はそれぞれ Gabor カーネルの回転角と大きさを表し，$\mathbf{x} = (x, y)$であり，$k_{\mu,\nu}$は以下の式で与えられる：

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{5.4}$$

このとき$k_\nu = k_{max}/f^\nu$，$\phi_\mu = \pi\mu/$（回転角$\mu$の数）である．$k_{max}$は最大周波数，$f$は Gabor カーネルの大きさの間隔を示す係数である．

本稿では 5 スケールで 6 回転角，つまり$\nu \in \{0, \dots, 4\}$，$\mu \in \{0, \dots, 5\}$の条件で Gabor カーネルを作成する．事前に設定するパラメータは$k_{max} = \pi/2$，$f = \sqrt{2}$の条件を与える．

**図 5.3** は 5 スケール 6 回転角，合計 30 の Gabor カーネルの実数部を示している．Gabor フ

ィルタによる特徴（Gabor 特徴量）の抽出は以下の式で定式化できる：

$$\mathbf{G}_{\psi I}(x, y, \mu, \nu) = \mathbf{I}(x, y) * \boldsymbol{\psi}_{\mu,\nu}(\mathbf{x}) \tag{5.5}$$

ここで, $\mathbf{I}(x, y)$はグレイスケールの入力画像, $\boldsymbol{\psi}_{\mu,\nu}(\mathbf{x})$は Gabor フィルタのカーネルであり, $\mathbf{I}(x, y)$と$\boldsymbol{\psi}_{\mu,\nu}(\mathbf{x})$の畳み込み積分より Gabor フィルタの出力$\mathbf{G}_{\psi I}(x, y, \mu, \nu)$を算出できる．ここで，**図 5.4** に Gabor フィルタの適用により算出した Gabor 絶対値成分画像を示す.

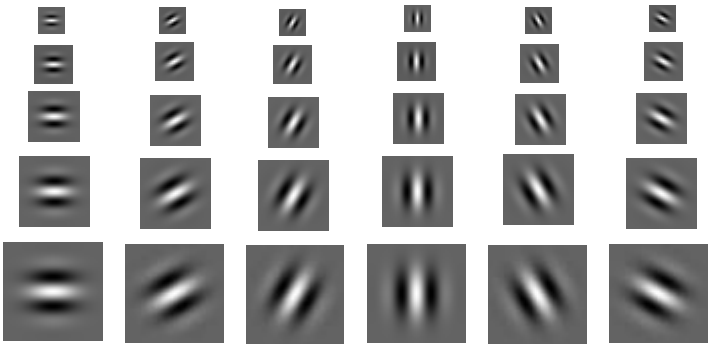
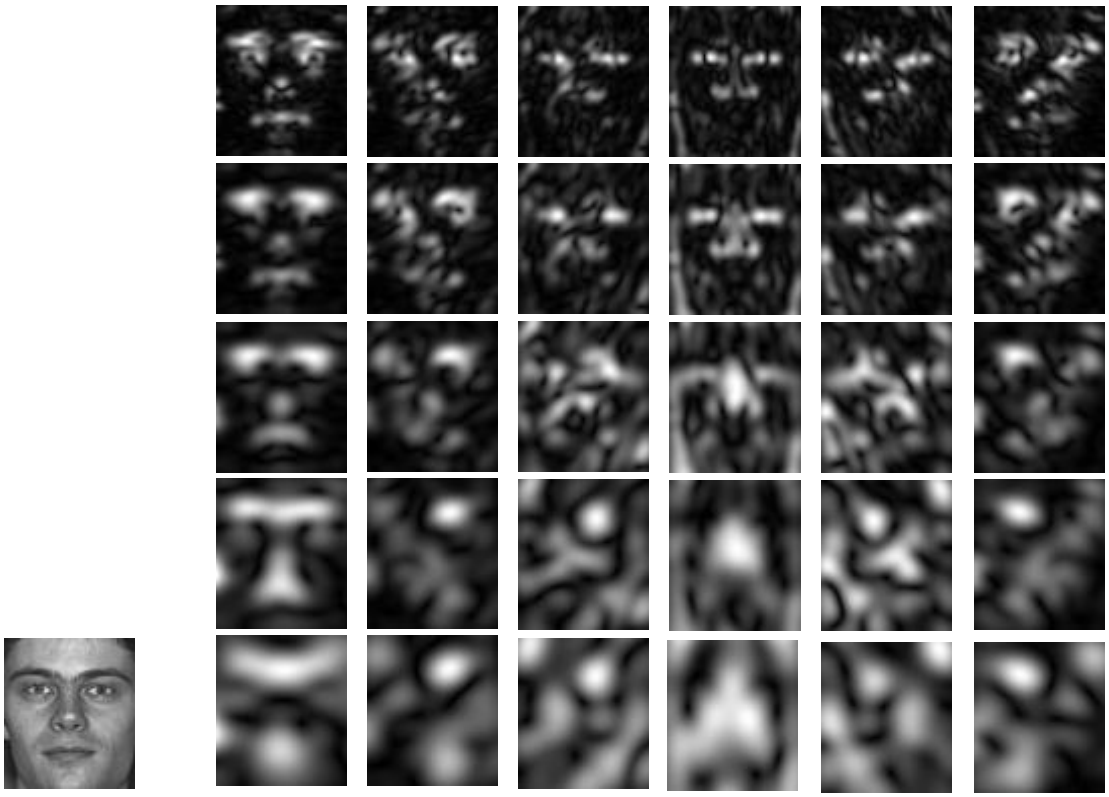


**図 5.3**：Gabor カーネル



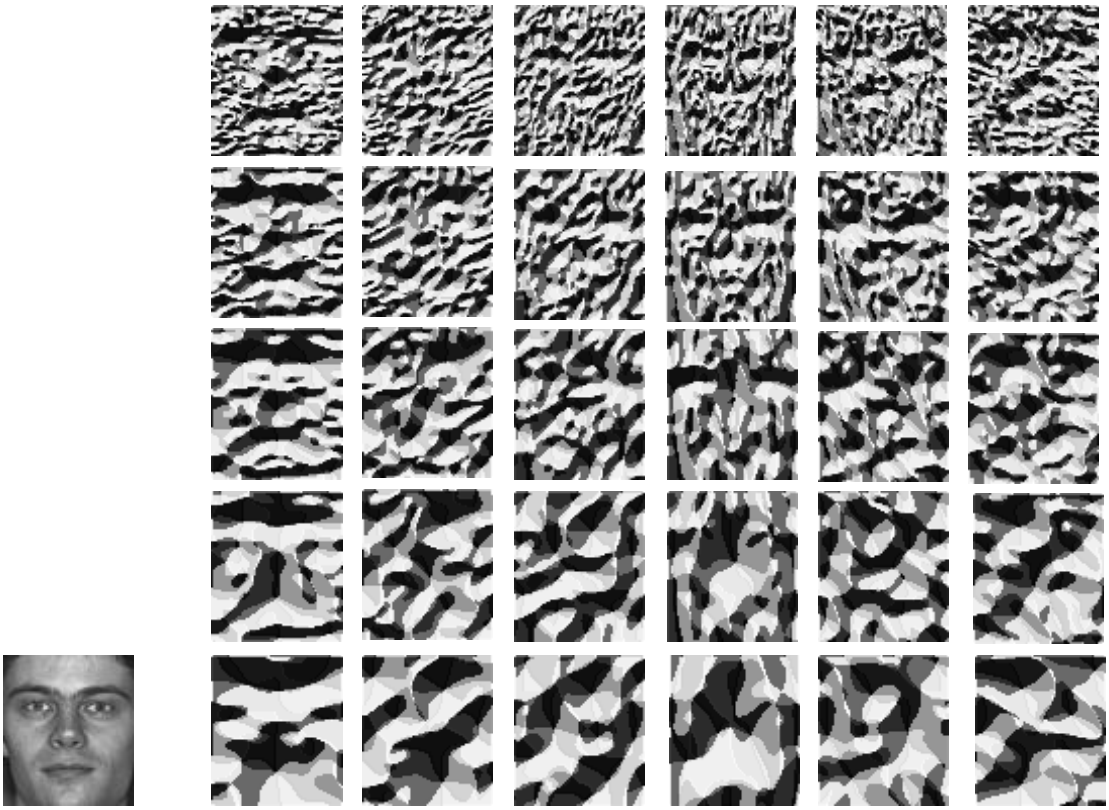
(a)入力画像　　(b) Gabor 絶対値成分画像(GMP)

**図 5.4**：入力画像と Gabor フィルタの適用例

**図 5.4**(a)は Gabor フィルタの適用前画像，**図 5.4**(b)は Gabor フィルタの適用により算出した Gabor 絶対値成分画像(GMP)である．そして**図 5.4** より，GMP は目，鼻や口などの顔器官に対して共起していることが確認できる．また GMP は頬や目元の影に対して共起していない．ゆえに GMP を基にした特徴量は照明変化による影響を最小限に抑えることができると言える．

### 5.2.3. Local Gabor Binary Pattern (LGBP)

Zhang らは濃淡情報の周期性と方向性を含む GMP に対し，LBP を適用することで構成される Local Gabor Binary Pattern (LGBP)を提案している**[3]**． Gabor 特徴量の濃淡変化はゆっくりとした変位であり，それに LBP を適用し，注目画素周辺の濃淡パターンを符号化することで，情報を高める効果が期待できる．近年，LGBP は顔画像解析に広く用いられており，適用例として Xia らは LGBP を性別分類に応用している**[4]**．**図 5.5** に LGBP を顔画像に対して適用した例を示す．



(a)入力画像　　(b) LGBP 適応後の画像

**図 5.5**：LGBP 適用後の画像

## 5.3. まとめ

本章では照明の変化に対して頑強な 3 つの特徴量について紹介した．これらは次章で述べる独自の顔の特徴量抽出法に関連する手法である．まず LBP は単調なグレイスケールの照明変化に頑強である．次に Gabor フィルタは顔の空間，および周波数領域における局所的な特徴量を抽出することができる．そして Gabor フィルタを用いることで照明の影響による顔画像の濃度値情報の変化を最小限に抑えることができる．最後に LGBP は Gabor 特徴量と LBP の 2 つのオペレーターより構成され，Gabor 特徴量のゆっくりとした濃淡変位を LBP の適用により符号化することで，情報を高める効果が期待できる．

## 参考文献

[1] T. Ojala，M. Pietikäineg and T. Mäenpää :“Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns”, IEEE Trans. Pattern Analysis and Machine Intelligence，vol.24，no.7，pp.971–987（2002）

[2] Y. Fang and Z. Wang :“Improving LBP Features for Gender Classification”， Proc International Conference Wavelet Analysis and Pattern Recognition，pp.373–377（2008）

[3] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang :“Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition”，Tenth IEEE International Conference on Computer Vision，vol.1，pp.786–791，（2005）

[4] B. Xia，H. Sun, and Bao-Liang Lu :“Multi-View Gender Classification Based on Local Gabor Binary Mapping Pattern and Support Vector Machines”， In Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. IEEE, pp.3388-3395（2008）．

# 第6章. Local Gabor Directional Pattern Histogram Sequence (LGDPHS)を用いた年齢・性別分類

## 6.1. まえがき

本章では，Local Gabor Directiona Patten Histogram Sequence (LGDPHS)と称した新たな特徴量を提案する．またそれを顔画像の年齢・性別分類に適用し，その性能を検証する．5章で述べたLBPはグレイスケール画像の単調な照明変化に対してロバストであるが，不規則な照明変化などのランダムノイズには敏感な問題がある**[1]**．そこでJabidらはLocal Directional Pattern (LDP)を提案している**[2]**．LBPは隣接する画素の特定方向の輝度値の強度を考慮する一方で，LDPは隣接する画素において全ての異なる方向のエッジ応答を考慮し，その中で重要な方向のエッジ情報のみを符号化する．提案する特徴量は，Gaborの絶対値成分画像(GMP)にLDPを適用することでそれを符号化し，情報を洗練化する効果が期待できる．本章の構成として，まず6.2節にて提案特徴量の詳細について述べる．次に6.3節では提案特徴量を用いた年齢･性別分類アルゴリズムのフレームワークについて説明し，6.4節にて性能検証の実験を行う．最後に6.5節で本章全体をまとめる．

## 6.2. Local Gabor Directional Pattern Histogram Sequence (LGDPHS)

LGDPHSは年齢・性別分類における独自の特徴量であり，**図 6.1**に示す3つの手順に従って算出することができる．始めに5スケール，6回転角のGaborフィルタによって抽出される計30のGMPに対してLDPを適用する．これよりGMPのテクスチャ情報を符号化し，重要性の高い情報を含むLGDPマップへの洗練化の効果が期待できる．次に抽出されたLGDPマップを複数ブロックに分割し，それぞれのブロック毎にヒストグラム列を計算する．最後にそれら全てのヒストグラム列を一つのベクトルとして結合することで，本章で新たに提案する顔特徴量を抽出できる．

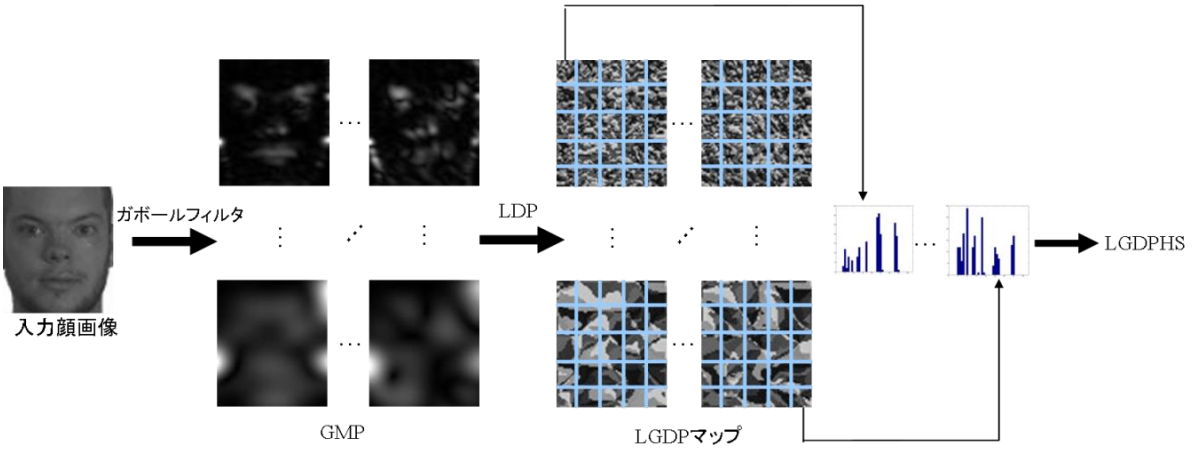



**図 6.1**：LGDPHS のフレームワーク

### 6.2.1. Local Directional Pattern (LDP)

近年，LBP は画像の輝度勾配を符号化する手法として顔画像解析の研究にて広く利用されている．しかし LBP は単調な照明変化にはロバストであるが，複雑な照明変化などのランダムノイズに脆弱な問題がある．原因として，LBP は注目画素の輝度の勾配強度や向きを符号化せずに，その近隣画素に注目し，注目画素との相関的な勾配変化を符号化することで, ある特定方向の勾配のみを符号化してしまっている点である. そこで Jabid らは LDP を提案している．それはあらゆる方向のエッジ応答を考慮し，その中で重要性の高い方向のエッジ情報のみを符号化できる．LDP は 3 つのステップにより算出できる．始めに，8 方向の Kirsch（カーシュ）エッジ応答マスクを適用し，8 つのエッジ応答$(m_0, \dots, m_7)$を求める．ここで 8 方向の Kirsch マスク$(M_0, \dots, M_7)$を**図 6.2** に示す.

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

East $M_0$ North East $M_1$ North $M_2$ North West $M_3$

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

West $M_4$ South West $M_5$ South $M_6$ South East $M_7$

**図 6.2**：8 方向の Kirsch(カーシュ)エッジ応答マスク

次のステップでは 8 つのエッジ応答$m_0, \dots, m_7$をそれぞれ比較し, 上位$t$個の$|m_i|$ $(i = 0, \dots, 7)$を選択する．選ばれた$t$個に 1 の値を割り振り，残りの 8 ビット内の$(8-t)$の値には 0 を割

り振ることで 8 ビットの LDP パターンを算出する. これより情報として重要性の高い方向のエッジのみを符号化できる．最後に，**図 6.3** に示すように 0 と 1 の 8 ビットの 2 進数を 10 進数に変換することで，LDP の符号化された値を算出できる．本実験では$t$ =3 と設定して実験を行う．ここで LDP を適用した顔画像の例を**図 6.4** に示す.

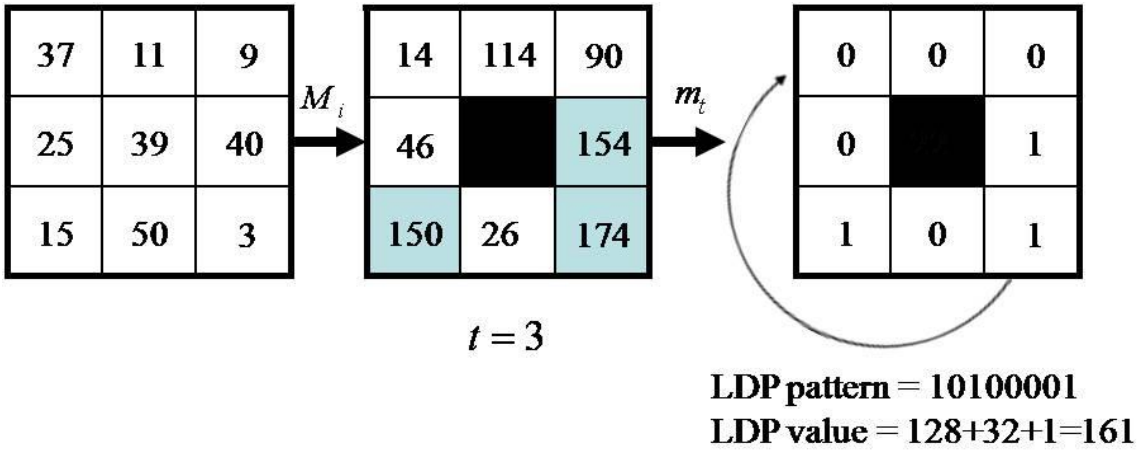
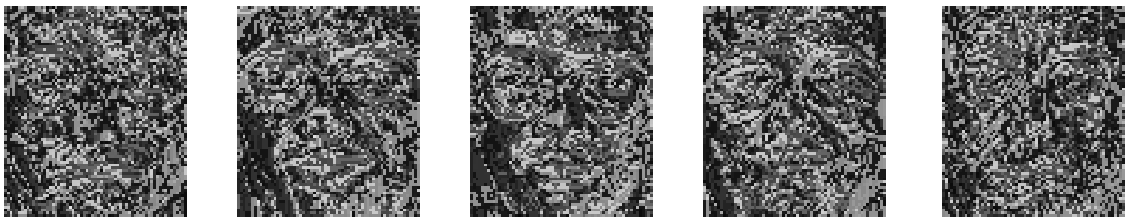



**図 6.3**：LDP の計算法

（a）LDP 適用前の顔画像



（b）LDP 適用後の顔画像

**図 6.4**：LDP を適用した顔画像の例

**図 6.4**(a)は LDP を適用する前の顔画像であり，**図 6.4**(b)は，**図 6.4**(a)の顔画像に対して LDP を適用した結果画像である. また**図 6.5** は LBP と LDP 適用画像をそれぞれヒストグラム化し，特徴量の比較を行っている．**図 6.5** の入力画像は照明変化の影響を受け，頬や額周辺における輝度値の変化が激しい．そこで，それらの領域に対して LBP を適用した場合，変化の激しさを保持した LBP 適用画像が算出されている．また**図 6.5**(a)の頻度ヒストグラムにおいて，頻度は 50 以下または 200 以上の輝度範囲に集中しており，照明変化等のランダムノイズに影響を受け易いことがヒストグラムからも確認できる．一方 LDP の場合，**図 6.5**(b)の頻度ヒストグラムは LBP と比較すると，50 以下または 200 以上の特定輝度範囲への頻度の偏りは見られない．これより LDP は LBP と比較して，照明変化などのランダムノイズの影響を抑える効果が期待できる.

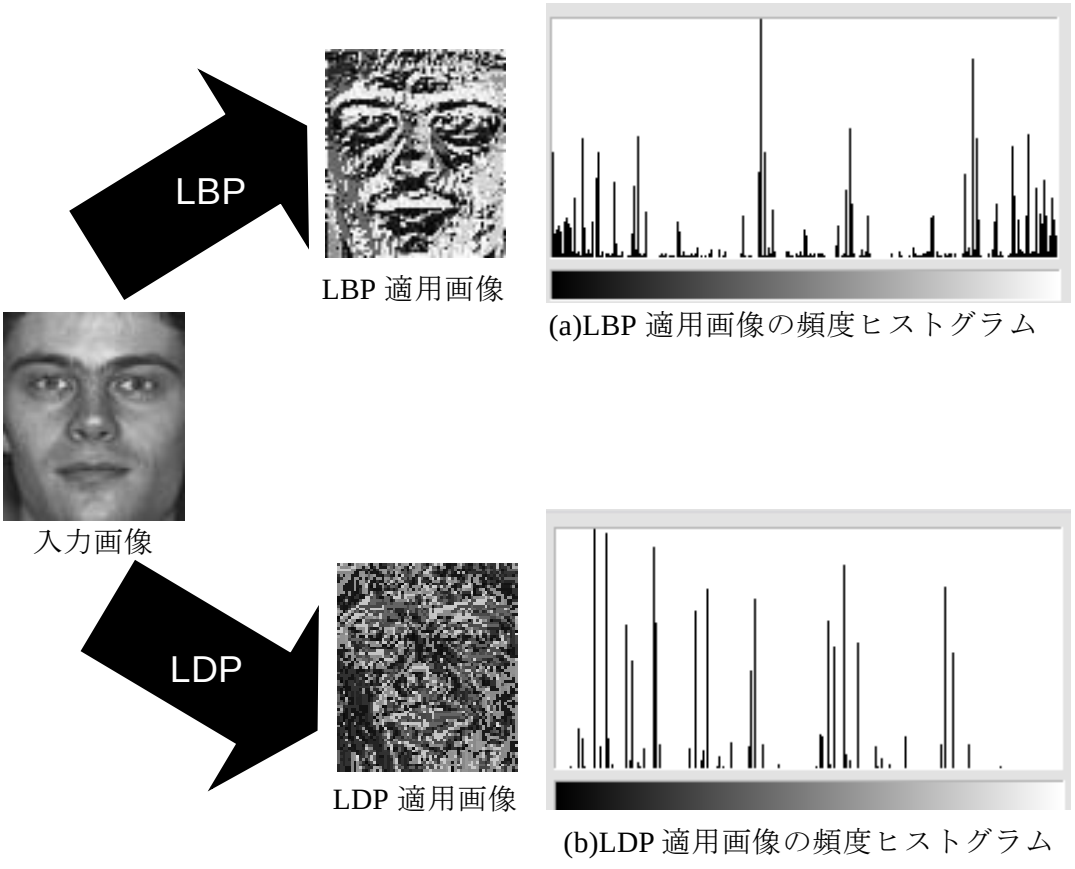
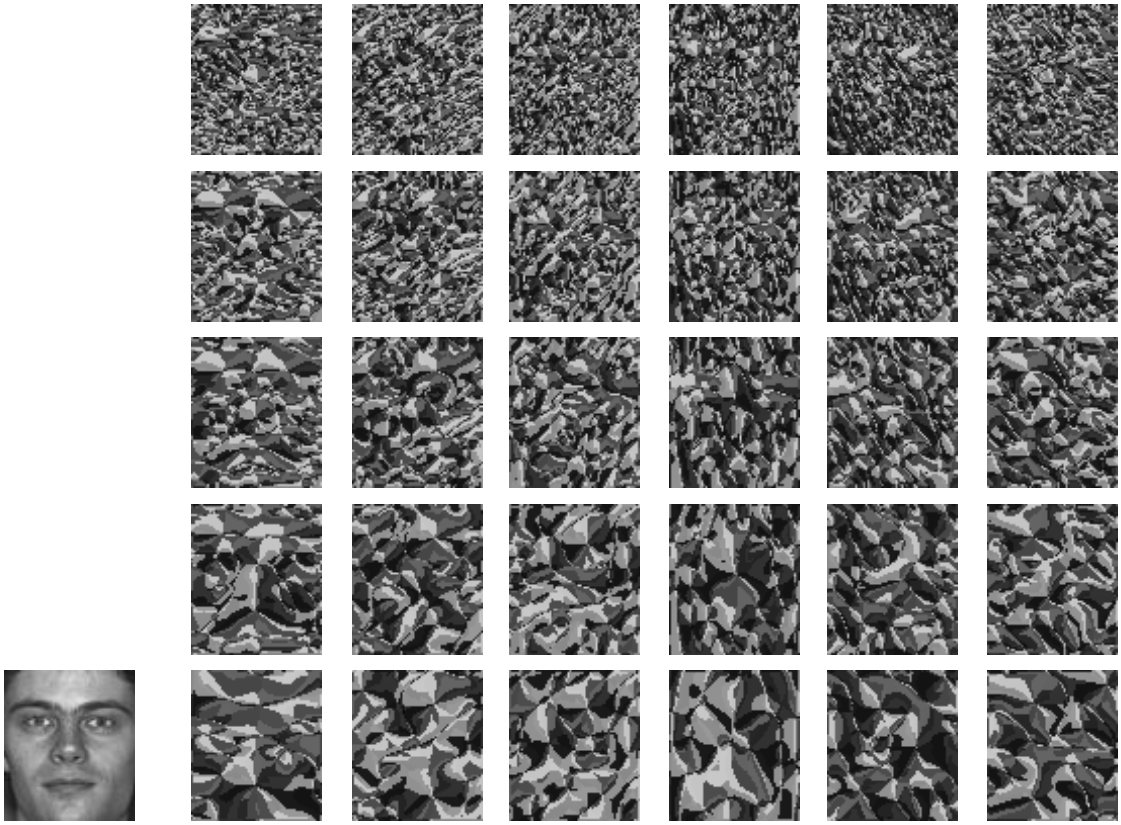



**図 6.5**：LBP と LDP のヒストグラムによる比較

### 6.2.2. Local Gabor Directional Pattern (LGDP)

LGDP の適用画像（LGDP マップ）は，GMP に LDP を施すことで算出できる．手順は 2 つのステップで構成され，まず始めに，5 スケール（$\nu \in \{0,\dots,4\}$），6 回転角（$\mu \in \{0,\dots,5\}$）の Gabor フィルタを顔画像に適用することで計 30 の GMP を導く．次に，GMP に対して

LDP を適用する．これより GMP の濃淡情報を重要性の高い方向のエッジ応答のみを含んだ符号化情報へと変換できる．これより有効性の高い洗練された特徴量を抽出でき，ノイズや不規則な照明変化に対して高い頑強性を期待できる．ここで**図 6.6** にて提案手法である LGDP の顔画像への適用例（LGDP マップ）を示す．



(a)入力画像　　(b) LGDP マップ

**図 6.6**：LGDP 適用例

### 6.2.3.　LGDP のヒストグラム特徴量への変換

本項では，$\nu \times \mu$個の LGDP マップを一つのベクトルへと特徴量化する手順を示す．始めに，各 LGDP マップを$q$個のブロックに分割し，それぞれのブロックからヒストグラムを抽出する．具体的にはグレイスケール画像$f(x,y)$のヒストグラムは 0 から$L-1$ の範囲において，以下のように定義できる：

$$h_i = \sum_{x,y} \mathbf{I}\{f(x,y) = i\}, i = 0,1,\dots,L-1 \tag{6.1}$$

ここで，$i$は$i$番目のグレイスケールの輝度値を示し，$h_i$はそのときのヒストグラムのビンの頻度の値である．ここで，**I**は以下の条件下で成り立つ：

$$\mathbf{I}\{D\} = \begin{cases} 1, & D \text{ is true} \\ 0, & D \text{ is false} \end{cases} \tag{6.2}$$

そして LGDP マップを$q$個のブロックに分割し，それらのブロックは$R_0, R_1, \ldots, R_{q-1}$として示される．$\nu \times \mu$の LGDP マップの中で，$r$番目のブロックのヒストグラムは以下のように定義できる：

$$\mathbf{H}_{\mu,\nu,r} = \left(h_{\mu,\nu,r,0}, h_{\mu,\nu,r,1}, \ldots, h_{\mu,\nu,r,L-1}\right) \tag{6.3}$$

ここで，

$$h_{\mu,\nu,r,i} = \sum_{(x,y)\in R_r} \mathbf{I}\left\{\mathbf{G}_{lgdp}(x,y,\mu,\nu) = i\right\} \tag{6.4}$$

式(6.4)の$\mathbf{G}_{lgdp}$は LGDP マップを表している．最後に，ヒストグラムが全てのブロックにおいて計算され，これらのヒストグラムを一つに集約したヒストグラム列$\mathfrak{R}$は以下の式として与えられる：

$$\mathfrak{R} = \left(\mathbf{H}_{0,0,0}, \ldots, \mathbf{H}_{0,0,q-1}, \mathbf{H}_{0,1,0}, \ldots, \mathbf{H}_{0,1,q-1}, \ldots, \mathbf{H}_{\mu,\nu,q-1}\right) \tag{6.5}$$

この$\mathfrak{R}$を提案する特徴量の LGDP Histogram Sequence(LGDPHS)として扱う．また本章の実験においては，LGDP マップを$q$=5×5=25 個のブロックに分割する．

## 6.3. 年齢・性別分類アルゴリズム

本節では提案特徴量である LGDPHS を用いた年齢・性別分類アルゴリズムについて述べる．**図 6.7** にて LGDPHS を用いた分類アルゴリズムのフレームワークを示す．

学習では，訓練画像から LGDPHS を算出して PCA を施し，累積寄与率 93%における固有ベクトル$\mathbf{u}_1, \ldots, \mathbf{u}_L$を算出する．その時，平均ベクトル$\bar{\mathbf{u}}$も同時に算出される．次に 1 枚の画像から抽出された LGDPHS を$\mathfrak{R}$とすると，式(6.6)に従い固有ベクトル$\mathbf{u}_j$との内積を計算することで特徴スコア$\mathrm{C}_j$を算出できる：

$$\mathrm{C}_j = \mathbf{u}_j^{\mathrm{T}}(\mathfrak{R} - \bar{\mathbf{u}}) \tag{6.6}$$

そして, 特徴ベクトル$\mathbf{C} = (\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_L)^{\mathrm{T}}$を新たな特徴量として扱い, Support Vector Machine (SVM：詳細な説明は付録 A 参照）を用いて識別器を生成する．テスト時は学習と同様に特徴ベクトル$\mathbf{C}$を求め，識別器から顔の属性を分類する．年齢は 4 つのカテゴリーへの分類し，性別は男女の 2 値分類を行う．

ここで**図 6.8** にて，学習画像全てから算出した LGDPHS を行列化し，それに対して PCA を施すことで求まる固有値とその累積寄与率の関係グラフを示す．また学習画像は，本章の実験で使用する**図 6.10** の FERET database に含まれる 590 枚の画像を用いる**[3]**．FERET

database の詳細については 6.4.2 項で後述する.

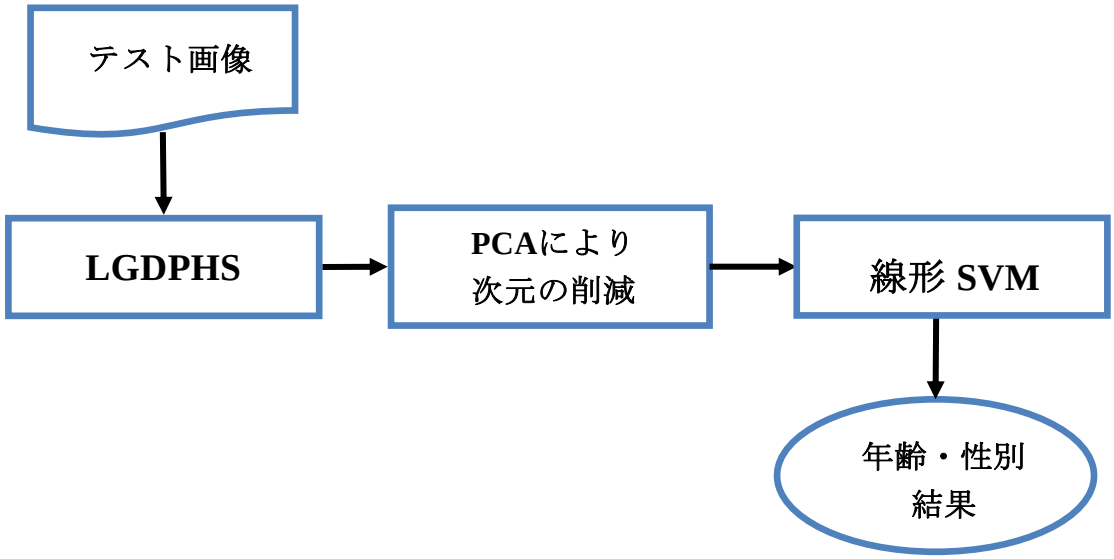



**図 6.7**：提案する年齢・性別分類手法のフレームワーク

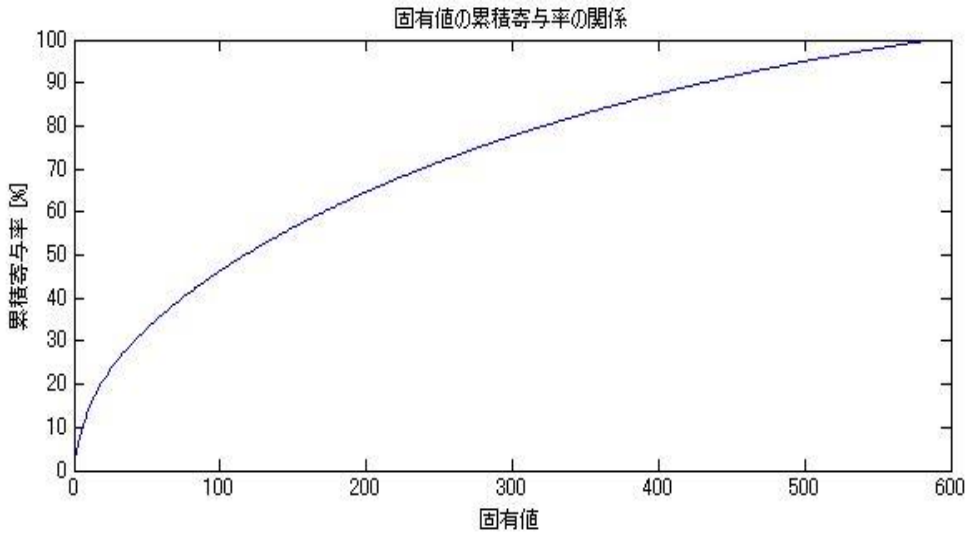



**図 6.8**：LGDPHS に対する固有値数と累積寄与率の関係

**図 6.8** から 25,801 次元の合計 590 の LGDPHS に対して PCA を適用すると，累積寄与率は第 431 成分において 93%に到達する．この時，432 成分以降は顔特徴に対するノイズ情報であり，25,801 次元を 431 次元に削減してもデータの情報量はほとんど保持される．また次元数がサンプル数以下になり，識別器として用いる線形 SVM でのオーバーフィッティングを抑え，汎化性能の優れた識別器の生成が期待できる．

## 6.4. 実験及び考察

本節では，6.3 節で述べた提案特徴量である LGDPHS を用いた年齢・性別分類アルゴリズムと従来手法との性能の比較実験を基に，提案アルゴリズムの照明変動などのノイズに対する頑強性を検証する．またその結果と考察について述べる．本実験ではアルゴリズム自体の性能を評価するため，前処理の顔検出器による正規化はマニュアルで行っている．つまり，実験で使用する全ての顔画像に対して左右の瞳の座標を手動でラベリングし，その座標を基に顔位置の正規化を事前に行う．このとき画像サイズを 65×75pix.に正規化する．また本章で用いる顔画像データベースは，顔の属性分類の評価に広く用いられ，照明変化などの実環境を想定した画像セットである．

### 6.4.1. 実験環境

実験環境を**表 6.1** に示す．年齢・性別分類システムは Windows 上で動作するソフトウェアとして構築している．

**表 6.1**：実験環境

| | |
|---|---|
| OS | Windows XP Professional SP3 |
| CPU | Intel ® Core™2 Quad Q6600 @ 2.40GHz |
| メモリ | 2.00GB |
| 開発言語 | C 言語 |
| 開発環境 | Visual Studio .NET 2008 |
| 使用ライブラリ | CLAPACK |

ここで，システムを構築する際に CLAPACK のライブラリを使用する．CLAPACK は数値計算系のライブラリが数多く含まれており，本章では PCA において，特異値分解での一般化固有値問題の解を求める際に使用する．

### 6.4.2. 実験概要

実験では年齢と性別分類は異なる顔画像データベースを用いる．年齢分類の実験では FG-Net aging database **[4]**を用い，性別分類の実験では FERET database **[3]**を用いる．これらは照明や顔の表情の変化，眼鏡の装着などの条件を含んでいる．データベースの詳細な情報を以下に示す.

#### FG-Net aging database

被写体は 0 歳から 69 歳の 82 名であり，計 1,002 枚の顔画像が含まれる．カラー，グレイスケール画像どちらも含み，一般的なスナップショットやパスポートの写真も一部含んでいる．照明変動，顔向き，表情，口ひげ，帽子や眼鏡の着用など多くの条件を有している．本実験は，照明変動等のノイズに対する頑強性に焦点を当てるため，顔の向きを含んだ画像は用いず，更に男性の画像のみを用いる．そして学習，テスト画像両方含め，252 枚の画像を利用する．**図 6.9** に FG-Net aging database の本実験に使った正規化後のサンプル画像を示す.

(1) 0-9 歳の子供のグループ

(2) 10-19 歳の青年グループ

(3) 20-34 歳の若い大人グループ

(4) 35-64歳の年長の大人グループ

**図 6.9**：4 つのカテゴリーに分類された FG-Net aging database の画像例

**FERET database**

被写体はあらゆる人種の 1,196 名であり，合計で 14,000 枚程の顔画像を含む．また一部の画像は照明や顔の傾きの変化を条件として含んでいる．本実験では，学習，テスト両方含め 1,196 枚の正面顔画像を利用する．ここで，**図 6.10** にて本実験で使用する FERET database の正規化後の画像例を示す．

**図 6.10**：FERET database の画像例（上行は男性，下行は女性）

本実験では年齢分類に FG-Net aging database を用いる．本研究では顔の年齢カテゴリーを 4 つのカテゴリーとして定義する．それは以下に示す 4 つのカテゴリーである．

I. 0-9 歳の子供
II. 10-19 歳の青年
III. 20-34 歳の若い大人
IV. 35-64 歳の年長の大人

ここで，**図 6.9** にて年齢カテゴリーI–IV 毎の FG-Net aging database の画像例を示す．**表 6.2** において本実験に用いる年齢カテゴリー毎の学習とテスト画像の枚数を示す．

**表 6.2**：カテゴリー毎の学習画像の枚数と，テスト画像の枚数

| カテゴリー名 | 学習画像の枚数 | テスト画像の枚数 |
|---|---|---|
| **0-9 歳の子供** | 26 | 40 |
| **10-19 歳の青年** | 32 | 40 |
| **20-34 歳の若い大人** | 32 | 31 |
| **35-64 歳の年長の大人** | 26 | 25 |

次に性別分類の実験では FERET database を用いる．子供はまだ成長過程であるため，男性と女性の顔特徴による違いが少ない．そして子供の性別分類は，他の年齢カテゴリーと比較して極めて難しく,アルゴリズムの性能を比較する意味での実験としては適していない．ゆえに本実験では子供の画像を含まず，他のカテゴリーの画像を多く含むデータベースが望ましい. FERET database は 17 歳以上の人物画像を多く含んでいるので本実験に適している．ここで男女のカテゴリーに分けた FERET database のサンプル画像を**図 6.10** に示し，**表 6.3** において，本実験に用いる学習とテスト画像のそれぞれの枚数を示す.

**表 6.3**：性別分類における学習画像とテスト画像の枚数

| 学習画像の枚数 | | テスト画像の枚数 | |
|---|---|---|---|
| **計**：590 | **男性**：320 | **計**：606 | **男性**：389 |
| | **女性**：270 | | **女性**：217 |

本実験の評価項目として，顔の属性分類に広く用いられている LGBP との性能の比較実験を行う．更に Gabor 特徴量を用いない単独の LBP，LDP も比較対象に加え，それぞれの手法の分類率を算出して性能の比較を行う．分類率は以下の式で定義できる：

$$\text{分類率（\%）} = \frac{\text{（正しいクラスに分類されたテスト画像の枚数）}}{\text{（全テスト画像の枚数）}} \qquad (6.7)$$

### 6.4.3. 実験結果・考察

提案する年齢分類アルゴリズムと従来法との分類率の比較結果を**図 6.11** に示す．また LGDP を用いた際，テスト画像が年齢毎に「どの年齢カテゴリーに分類されているかの割合」を表した結果を**図 6.12** に示す．

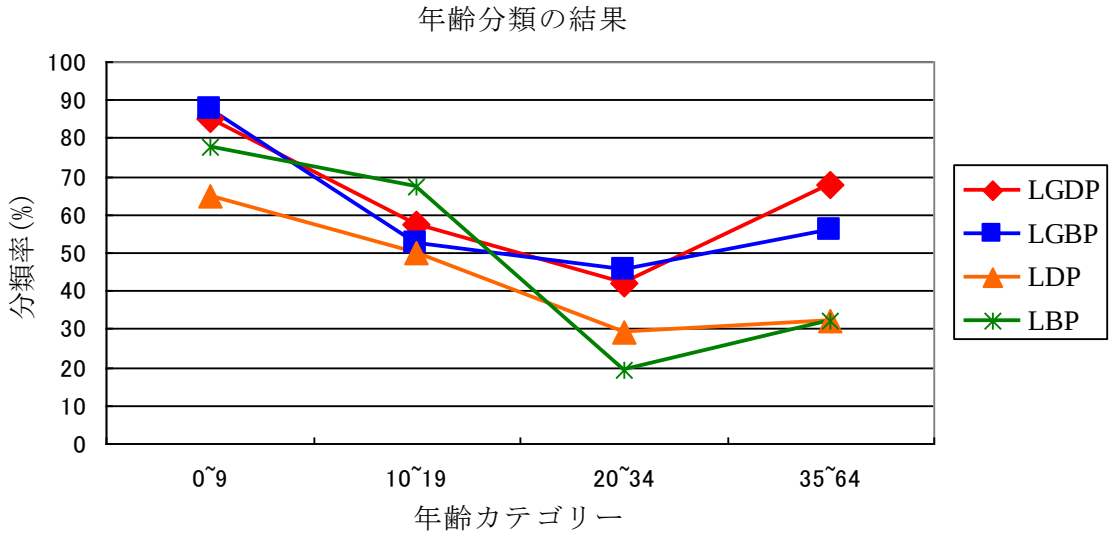



**図 6.11**：年齢分類の結果

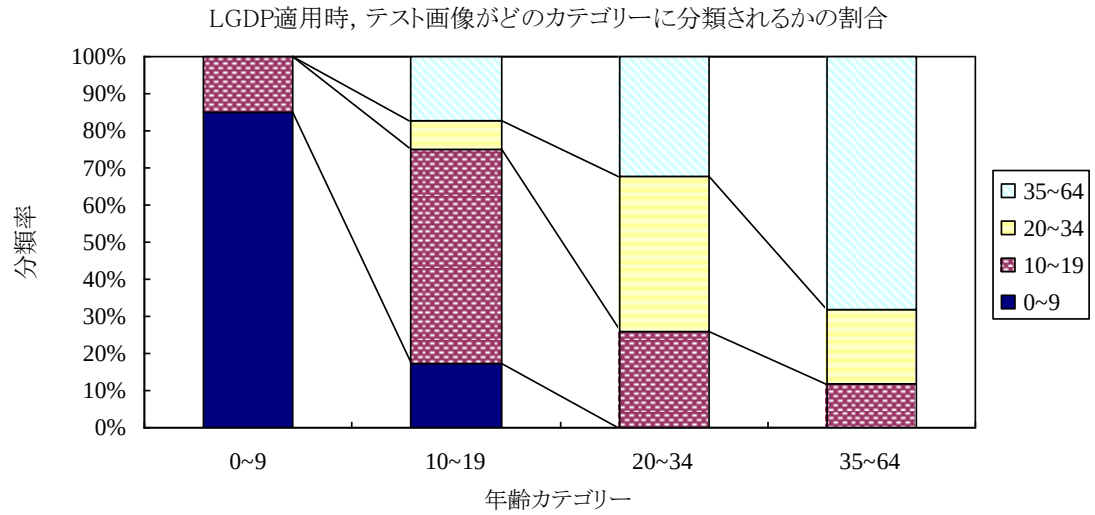



**図 6.12**：LGDP を用いた際，テスト画像が年齢毎でそれぞれ「どの年齢カテゴリーに分類されているか」の検証結果

次に，**表 6.4** にて年齢分類における分類率を従来法と比較した実験結果を示す．また FERET database を使った性別分類における分類率の比較結果を**表 6.5** に示す．

**表 6.4**：年齢分類における分類率の結果

| 手法名 | 分類率（%） |
|---|---|
| **LGDP** | **63.97** |
| **LGBP** | 61.8 |
| **LDP** | 46.3 |
| **LBP** | 52.9 |

**表 6.5**：性別分類における分類率の結果

| 手法名 | 分類率（%） |
|---|---|
| **LGDP** | **91.9** |
| **LGBP** | 88.1 |
| **LDP** | 74.6 |
| **LBP** | 82.7 |

ここでは実験結果を基にした考察について述べる．まず**図 6.11** より，提案手法の年代毎の分類率は，他の手法と比較した場合，各年代において優れた結果を示している．特に 0-9 歳のカテゴリーは 85%と最も優れた分類率であり，提案手法の有効性を確認できる．これは，子供の顔には皺やヒゲがないこと，顔の輪郭が丸みを帯びていること等の子供独特の特徴に起因し，他のカテゴリーよりも優れた結果が得られたと推測できる．また他のカテゴリーの場合，学習に用いた画像の枚数不足が性能低下の一因であると考えられる．本実験で用いた FG-Net aging database における子供と青年の画像はそれぞれ 200 枚以上存在するが，若い大人のカテゴリーでは学習とテストに用いた画像は計 90 枚に満たない．ゆえに十分な画像枚数を有するデータベースの作成は，年齢分類における今後の課題であると言える．

次に**表 6.4** より，提案手法の分類率は，従来法と比較して最大で 18%程向上し，最も優れた性能である．そして**図 6.12** では，提案手法はテスト画像の多くを正解クラスに分類で

きており，仮に誤ったクラスに分類した場合でも，ほとんど正解クラスの隣クラスに分類していることから，その有効性を確認できる．

最後に性別分類の性能を評価した**表 6.5** から， 3 つの従来法において LGBP は最も優れた性能を発揮し，その分類率は 88.1%である．ここで提案手法の分類率は 91.9%なので，それは LGBP よりも優れた性能であることを確認できる．

以上のことから，提案する LGDPHS を用いた年齢・性別分類アルゴリズムは従来法と比較して顔の有効な特徴量を抽出でき，かつ照明変動などのノイズに対する頑強さで優れていると言える．

## 6.5. 課題

**・年齢の推定方法の改善**

本章では 4 つの年齢カテゴリーへの分類を行った．しかし実用性を考えた場合，クラス分類ではなく，最適な年齢を推定する手法が望ましい．改善例として Support Vector Regression (SVR)（7.2.1 項参照）等の回帰手法を適用することが挙げられる．

**・顔領域の正規化**

本章の提案手法では，顔画像は予め左右の瞳の座標を基準にマニュアルで正規化した．しかし瞳の座標を基準にした正規化法では，顔の幅や長さ，向きの変化により位置ズレ誤差の問題を招く．この課題を解決するため，より緻密な正規化法へと発展させる必要がある．具体的には AAM を用いた正規化法の適用や，Gabor フィルタを画像全体にフィルタリングするのではなく，目，鼻や口などに位置する特徴点とその周辺領域に対してフィルタリングするといった改善策が挙げられる．

## 6.6. まとめ

本章では，従来の特徴量抽出法の課題である照明変動などの不規則なノイズに対する脆弱性に焦点を当て，新たな特徴量抽出法を検討した．そしてそれを用いた顔の年齢・性別分類アルゴリズムを述べ，その性能検証を行った．提案特徴量は Gabor フィルタと LDP の二つのオペレーターを用いる．変化が緩やかな GMP の濃淡値に対して LDP を適用することで，濃淡値を重要性の高いエッジ応答の方向を含む符号化情報へと変換できる．これより有効性の高い洗練された特徴量を抽出でき，ノイズや不規則な照明変化に対しての頑強性を高める効果が期待できる．実験では，年齢・性別分類について提案特徴量である LGDPHS と従来法である LGBP，LBP，LDP との性能の比較検証を行った．結果として提案手法は，年齢の分類率が約 64%，特に子供のカテゴリーは分類率が約 85%であり，他のカテゴリーより優れた結果が得られた．そして年齢・性別共に提案手法が従来法より優れた性能であることを確認できた．

# 参考文献

[1] T. Ojala，M. Pietikäineg and T. Mäenpää：“Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns”, IEEE Trans. Pattern Analysis and Machine Intelligence，vol.24，no.7，pp.971–987（2002）

[2] T. Jabid, M.H. Kabir, O. Chae：“Local Directional Pattern (LDP) - A Robust Image Descriptor for Ob-ject Recognition”，IEEE International Conference on Advanced Video and Signal Based Surveillance, pp.482–487（2010）

[3] P.J. Phillips，H.Wechsler，J.Huang and P. Rauss：“The FERET Database and Evaluation Procedure for Face Recogni-tion Algorithms”，Image and Vision Computing, vol. 16, no. 10, pp. 295–306（1998）

[4] The FG-NET Aging database, http://sting.cycollege.ac.cy/~alanitis/ fgnetaging/index.htm.

# 第7章. GAAM による大局的特徴量と LGDPHS による局所的特徴量を用いた年齢・性別推定

## 7.1. まえがき

本章では 6 章で述べた LGDPHS を用いた年齢・性別分類アルゴリズムの課題を踏まえ，それを発展させた新たなアルゴリズムを提案する．実験では提案アルゴリズムの性能を検証するため，従来法との性能比較，更に年齢推定では大学生モニター20 名による主観評価である「見かけ年齢」との比較を行う．提案手法の特徴量は大局的・局所的な 2 つの特徴量から構成される．大局的特徴量として，顔全体の濃淡値を数値化した GAAM のパラメータを用いる．また局所的特徴量として GAAM により正規化した顔領域から抽出する LGDPHS を用いることで，位置ズレ誤差や照明変動に対して頑強となる．本章の構成として，提案手法の詳細は 7.2 節で述べ，実験とその結果を 7.3 節で示す．最後に 7.4 節で本章全体をまとめる．

## 7.2. 提案する年齢・性別推定アルゴリズム

本節では 6 章で述べた課題を踏まえ，新たな年齢・性別推定アルゴリズムについて述べる．具体的には 6.5 節の 2 つの課題である「年齢の推定方法の改善」と「顔領域の正規化」についての課題解決に取り組む．ここで**図 7.1** において提案手法のフレームワークを示す．

アルゴリズムは 4 つのステップから構成される．まず始めに顔画像の正規化を行う．6 章で提案した手法は左右の瞳を基準に顔画像をアフィン変換することで，個人毎の顔器官の位置ズレ誤差を補正した．しかし個人差から生じる顔の幅や長さの違いにより，鼻や口等における位置ズレを十分に補正できず，分類率の低下を招いていた．そこで初期位置を顔検出器より算出する GAAM を使い，顔画像を特定の形状内にワープして正規化し，顔領域を切り出す．本実験ではこのプロセスより 82×92pix.の画像を 70×65pix.へとリサイズする．第 2 ステップでは特徴量抽出を行う．まず GAAM のアペアランスパラメータを

正規化の際に算出する．これは顔全体の濃淡情報を含んでおり大局的特徴量として扱うことができる **[1]**．しかし GAAM のパラメータのみを特徴量として採用することはフィッティング位置の微小なズレや照明変動の影響を受け易くし，頑強性の乏しい特徴量を抽出してしまう恐れがある．そこで大局的特徴量に加え，それら悪条件に影響を受けにくい局所的特徴量を統合した新たな特徴量を提案する．ここで局所的特徴量は GAAM により正規化された顔領域に対し，LGDPHS を適用することで算出する．LGDPHS は照明変動などのランダムノイズに対して頑強である．更に LGDPHS は GAAM の正規化された顔領域から算出されるので，位置ズレ誤差の影響を最小限に抑えることが期待できる．次に局所特徴量に対して PCA を施し，次元数の削減と不必要な情報の除去を行う．ここで特徴ベクトルは，累積寄与率が 93%を満たす固有ベクトルとヒストグラム列の内積から計算できる．

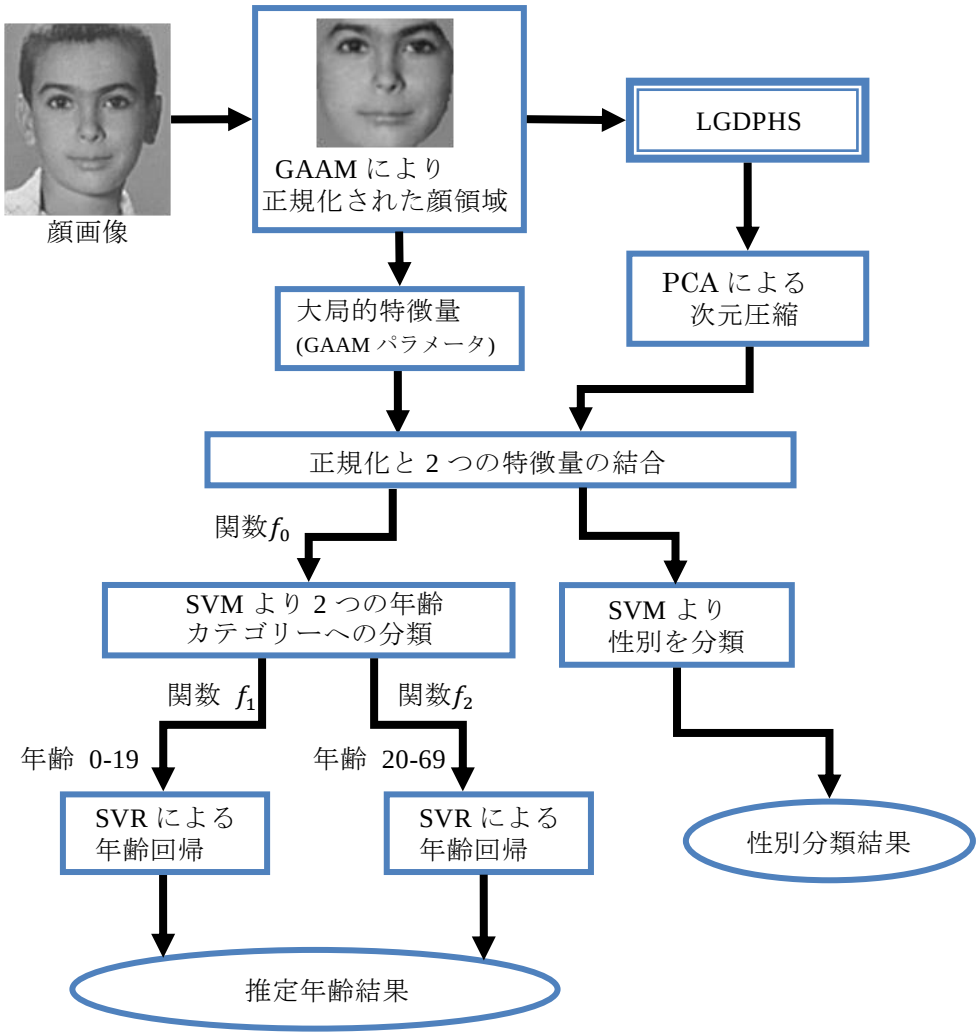



**図 7.1**：提案手法のフレームワーク

第 3 ステップでは 2 つの特徴量（大局的・局所的特徴量）を Min-Max (MM)法により 0 から 1 の範囲に正規化し，それらを一つのベクトルに結合する．この特徴量は，人間が顔から年齢や性別を判断するときの着眼点である顔全体の「見え」と，皺や肌の質感等の細かな情報を疑似的に再現した局所的な特徴量を含んでいると言える．最終ステップでは年齢・性別の推定を行う．学習とテストは異なるプロセスで成り立ち，年齢推定の学習時では, 始めに子供と大人の年齢カテゴリーに分類する. ここで子供は{0, 19}歳, 大人は{20, 69}歳として定義する．そして 2 値分類器$f_0$（**図 7.1** 参照）は SVM を用いて算出する．次に SVM を回帰問題に適用した Support Vector Regression（SVR）を使い回帰関数$f_1$，$f_2$（**図 7.1** 参照）を大人と子供の各カテゴリーにおける学習画像を使い計算する．テスト時では，テスト画像が回帰前の子供と大人のどちらのカテゴリーに属するかを 2 値分類器$f_0$から決定する．もし子供のカテゴリーに分類された場合は回帰関数$f_1$，大人の場合は回帰関数$f_2$を用いて年齢を推定する．また性別分類の識別器としては SVM を用いる．

### 7.2.1. Support Vector Regression（SVR）

本節では SVM（付録 A 参照）を回帰問題に適用した Support Vector Regression (SVR)について述べる**[2]**．SVR は SVM と同様カーネルトリックを用いて非線形モデルへと拡張できる．ここで$\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$，教師信号$y_1, \dots, y_m \in \mathbb{R}$とすると学習データセットは$(\boldsymbol{x}_1, \boldsymbol{y}_1), \dots, (\boldsymbol{x}_m, \boldsymbol{y}_m)$として与えられる．その時回帰関数は以下の式で定義できる：

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b \tag{7.1}$$

この時 SVR は以下の誤差関数の最小化として与えられる：

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}|y_i - f(\boldsymbol{x}_i)|_\epsilon \tag{7.2}$$

ここで$|\cdot|_\epsilon$は$\epsilon$許容誤差関数である．$f(\boldsymbol{x_i})$と$y_i$の差が$\epsilon(>0)$未満のときは，$\epsilon$許容誤差関数の値は 0 であり，以下の式で定義される：

$$|y_i - f(\boldsymbol{x}_i)|_\epsilon = \begin{cases} 0 & if\ |y_i - f(\boldsymbol{x}_i)| \le \epsilon \\ |y_i - f(\boldsymbol{x}_i)| - \epsilon & otherwise \end{cases} \tag{7.3}$$

ここで 2 つのスラック変数$\mathcal{E}_i$，$\widehat{\mathcal{E}_\iota}$を導入することで SVR の誤差関数は以下の式として書ける：

$$\begin{aligned} &\min_{\boldsymbol{w},b,\mathcal{E}_i,\mathcal{E}_\iota} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\left(\mathcal{E}_i + \widehat{\mathcal{E}_\iota}\right) \\ \text{subject to}\quad &y_i - (\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \le \epsilon + \mathcal{E}_i\ with\ \mathcal{E}_i \ge 0 \\ &-y_i + (\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \le \epsilon + \widehat{\mathcal{E}_\iota}\ with\ \widehat{\mathcal{E}_\iota} \ge 0 \end{aligned} \tag{7.4}$$

式(7.4)の$\frac{1}{2}\|\boldsymbol{w}\|^2$は正則化項であり，$C$は正則化項の比重の強さを調整するパラメータである．

## 7.3. 実験及び考察

本節では提案する大局的・局所的特徴量を用いた年齢・性別推定アルゴリズムの性能を検証し，その結果と考察について述べる．年齢推定の実験は大きく分けて 2 つ実施する．それは第 1 に従来法との性能の比較実験である．第 2 にモニターによる主観評価である見かけ年齢と提案手法との性能の比較である．

### 7.3.1. 年齢・性別分類における従来法との比較実験

本項では提案する年齢推定と性別分類アルゴリズムの性能検証のために従来法との性能の比較を行う．

#### 7.3.1.1. 実験概要

年齢推定の実験では 6 章で用いた FG-Net aging database を使用する**[3]**. それは 82 人の被写体についてカラー, グレイススケールどちらの画像も含み, 0 歳から 69 歳までの計 1,002 枚の画像により構成される．ただし 6 章の実験では顔の向きを含んだ画像は用いていなかったが，本実験では-30 度から 30 度までの顔の向きを含む画像も実験対象とする．**表 7.1** において本実験で用いた画像枚数の年代毎の内訳を示す．**表 7.1** から FG-Net aging database は多くの子供の画像を含んでいる一方で，年代が上昇するにつれて画像の枚数が減少することを確認できる．また本実験では学習画像内から 114 枚の画像を選び，それを GAAM の形状・アペアランスモデルの構築に用いる．GAAM のフィッティングにおいては形状パラメータを 9 次元，アペアランスパラメータを 42 次元に設定する．このパラメータ内からアペアランスパラメータ上位の 34 次元を大局的特徴量として用いる．これらの次元数は実験的に決定している．

**表 7.1**：FG-Net aging database の年齢毎の画像枚数の分布

| 年齢範囲（歳） | 画像枚数（枚） | 学習画像枚数（枚） | テスト画像枚数（枚） |
|---|---|---|---|
| **0−9** | 257 | 190 | 67 |
| **10−19** | 267 | 193 | 74 |
| **20−29** | 117 | 82 | 35 |
| **30−39** | 67 | 48 | 19 |
| **40−49** | 42 | 32 | 10 |
| **50−59** | 14 | 10 | 4 |
| **60−69** | 7 | 5 | 2 |
| **0−69** | 771 | 560 | 211 |

**図 7.1** における 2 値分類器$f_0$は線形 SVM を採用し，0-19 歳の子供のカテゴリーでの回帰関数$f_1$の導出には RBF カーネル（式(A.3)）のよる非線形 SVR を用いる．また 20−69 歳の大人のカテゴリーでの回帰関数$f_2$の導出には多項式カーネル（式(A.2)）を用いる．

性別分類の実験では 6 章と同様に FERET database を用いる**[4]**．本実験においては 1,120 枚の顔画像を用いる．ここで**表 7.2** に FERET database の画像枚数の内訳を示す．また子供は成長過程であるので男性と女性の顔の特徴差が少なく，子供の性別分類は他の年齢カテゴリーと比較して極めて難しい．ゆえにアルゴリズムの性能を比較する意味での実験には適していない．そこで本実験においては子供の画像は用いず，他の年齢カテゴリーの画像を使用する. ここで GAAM のモデル構築には学習画像内から 80 枚の画像を選び使用する．そして GAAM のパラメータ数は年齢推定の場合と同値とする．更に SVM の 2 値分類においては RBF カーネルを採用する．

**表 7.2**：FERET database の画像枚数の内訳

| | サンプル数（枚） | 学習画像枚数（枚） | テスト画像枚数（枚） |
|---|---|---|---|
| 女性 | 467 | 280 | 187 |
| 男性 | 653 | 380 | 273 |
| 合計 | 1,120 | 660 | 460 |

ここで年齢推定の実験項目について述べる．他の従来法と性能比較を行うために平均絶対誤差（MAE）と累積スコア（CS）についての検証を行う. MAE は実年齢$y_i$と推定年齢$\hat{y}_i$の絶対誤差の平均であり，以下の式で定式化できる：

$$MAE = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{N} \tag{7.5}$$

ここで，$N$はテスト画像の合計枚数である．また CS は以下の式で与えられる：

$$CS(\theta) = \frac{N_{e \le \theta}}{N} \times 100\% \tag{7.6}$$

ここで，$N_{e \le \theta}$は絶対誤差が$\theta$以下となるテスト画像の枚数を表している．

アルゴリズムの性能を評価するため, MAE の性能検証では比較対象に 4 つの従来法を用いる．第 1 の従来法は LBP+GAAM である．これは顔画像解析に広く用いられている LBP **[5,6,7]**を局所特徴量として用い, 同時に大局的特徴量として GAAM のパラメータを用いる．また GAAM による特徴抽出の前処理として顔画像の正規化が行われる．第 2 の従来法は 6 章で提案した LGDPHS を用いる．第 3 の従来法として GAAM のパラメータを特徴量として用いた手法である．これは特徴抽出のためのフィルタリング処理を行わず，GAAM のメッシュ収束で得られた形状・アペアランスパラメータを特徴量として用いた手法となる．

最後の従来法として，年齢推定において代表的な手法である Geng らによって提案された Aging Pattern Subspace (AGES)を用いる**[8]**．これは年齢パターンをモデル化するために人物の年齢顔画像列として定義されたサブ空間を学習する．テスト時は顔画像をサブ空間に射影し，サブ空間上において最適な顔画像を再構築することでその年齢を推定できる．この手法は本実験においての実験環境とは多少異なるが, Geng らの文献に記されている FG-Net aging database を用いた実験で算出された MAE の結果を本稿における性能検証に用いる．また CS の性能検証に関しては MAE での 4 つの従来手法において AGES 以外を比較対象とする．

次に性別分類の評価項目について述べる．提案手法の性能を検証するために従来法との分類率についての性能の比較を行う. 従来法としては 4 つの手法を比較対象として用いる．第 1，2 の従来法は，年齢推定と同様に LBP+GAAM，LGDPHS を用いる．次に第 3 の従来法は顔認識に広く用いられる LGBPHS **[9]**である．これは 5 章で述べた Gabor フィルタと LBP の 2 つのオペレーターを顔画像に適用することで構成される．その特徴画像を複数のブロックに分割し，ブロック毎に頻度ヒストグラム列を生成する．更にそれらを一つのベクトルとして結合することで LGBP のヒストグラム列が算出できる．最後の従来法として Gabor 特徴量を用いない単独の LBP を比較対象とする．ここで GAAM を用いていない手法は左右の瞳を基準に位置とスケールにおいて正規化の前処理を適用する．そのとき正規化後の画像サイズは 65×75 と設定し, 顔以外の背景領域が画像内に含まれないようにする.

#### 7.3.1.2. 実験結果・考察

ここでは本実験の結果と考察について述べる．**表 7.3** では提案手法を含めた各手法の MAE での性能比較の結果を示している．**表 7.3** から提案手法の LGDPHS+GAAM は従来法の LBP+GAAM と比較して MAE が 0.7 歳程度改善されている. 更にその他の手法と比較しても LGDPHS+GAAM の MAE は最も低い値であることが分かり，提案手法の優位性を確認できる．しかし提案手法の 0－19 歳での MAE は 3.45，20－69 歳では 11.86 であり，提案手法は 20－69 歳において高い誤差を示すことを確認できる．これは学習画像の枚数が 0－19 歳と比較して少ないことが影響していると考えられる.

**表 7.3**：FG-NET aging database での各手法の MAE の結果

| 手法 | MAE（0－69 歳） | MAE（0－19 歳） | MAE（20－69 歳） |
|---|---|---|---|
| **LGDPHS+GAAM（提案手法）** | **6.24** | **3.45** | **11.86** |
| **LBP+GAAM** | 6.92 | 4.27 | 12.26 |
| **LGDPHS** | 7.89 | 5.24 | 13.25 |
| **GAAM** | 10.18 | 4.19 | 22.24 |
| **AGES** | 6.77 | No data | No data |

**図 7.2** では年齢誤差の増加に伴う CS の変化をグラフで表現している. 提案手法の CS は, 誤差の閾値$\theta$＝9 歳のとき 80%以上になり，$\theta$＝15 歳のとき 90%以上の値になることを確認できる．そして提案手法は従来法と比較し，誤差の少ない安定した年齢推定ができると言える.

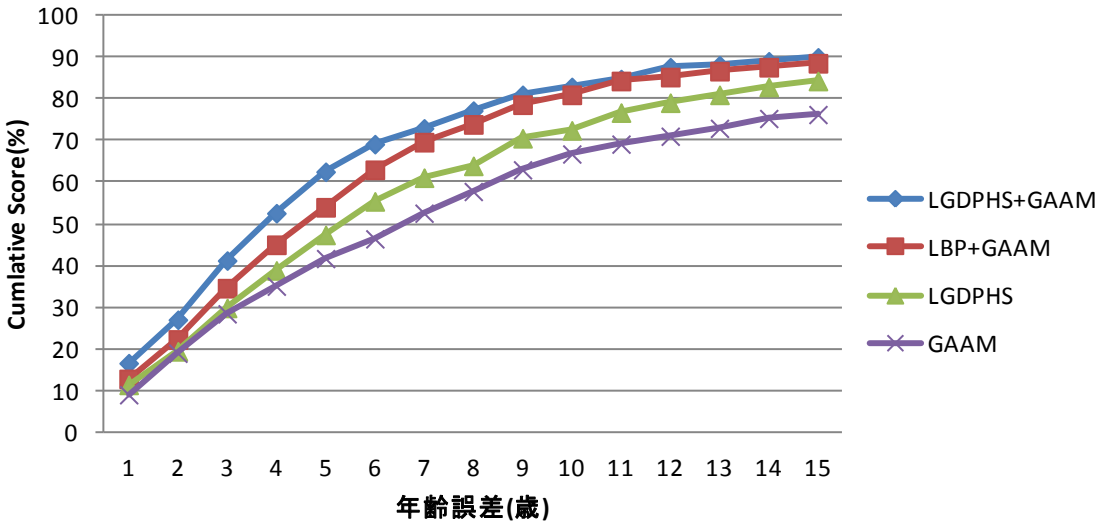



**図 7.2**：FG-NET aging database での各手法の Cumulative Score (CS)

ここで，**表 7.4** において FERET database を用いた性別の分類率の結果を示す．提案手法の分類率は 89.4%に達しており，他の従来法と比較して最も優れていることを確認できる.

以上のことから，提案手法は年齢推定，性別分類において従来手法と比較して優れた性能を示し，その有効性を確認できた.

**表 7.4**：FERET database における性別の分類率の結果

| 手法 | 平均分類率（%） |
|---|---|
| **LGDPHS+GAAM** | **89.4** |
| **LBP+GAAM** | 85.8 |
| **LGDPHS** | 88.9 |
| **LGBPHS** | 84.6 |
| **LBP** | 84.1 |

### 7.3.2. 年齢推定におけるモニターとの比較実験

本項では，FG-Net aging database 内の顔画像を使用し，大学生モニター20 名による主観評価（見かけ年齢）と提案する年齢推定法との性能比較実験を行う．

#### 7.3.2.1. 実験概要

実験項目は，定量的・定性的な観点からモニターによる「見かけ年齢」と提案アルゴリズムの性能の比較を行う．具体的に定量的な評価としては，MAE と CS を評価項目として「見かけ年齢」との性能の比較を行う．定性的な評価としては，提案手法の推定年齢と「見かけ年齢」を散布図で示し，実年齢の直線に対してどのような傾向が表れるかを考察する．ここで散布図は，モニター1 名につき 192 枚の顔画像の「見かけ年齢」を採取し，各画像に対してモニター20 名の平均を取り，その結果を散布している．また「見かけ年齢」のデータ採取のためにモニターに提示した FG-Net aging database における画像枚数の年代毎の内訳を**表 7.5** に示す．モニターの主観評価に用いる画像は各世代できるだけ均等になるような枚数に設定し，データベースからランダムに選択している．

**表 7.5**：FG-Net aging データベースにおけるモニターに提示した画像枚数の年代毎の内訳

| 年齢範囲（歳） | モニターに提示した画像枚数（枚） |
|---|---|
| **0−9** | 40 |
| **10−19** | 40 |
| **20−29** | 40 |
| **30−39** | 39 |
| **40−49** | 33 |
| **50−59** | 0 |
| **60−69** | 0 |
| **0−69** | 192 |

#### 7.3.2.2. 実験結果・考察

ここでは定量的な評価である MAE と CS，定性的な評価である年齢の散布図を用いて提案手法と「見かけ年齢」の性能の比較を行う．**表 7.6** において提案手法，モニターA－T 計 20 名の MAE とその平均の MAE の結果を示す．**表 7.6** から提案手法は MAE＝6.24 歳であり，モニター20 名の MAE の平均は 7.52 歳であることを確認できる．このことから提案手法は，平均して 6.24 歳の誤差で年齢を推測するが，人間が年齢を推測する場合，約 7−8 歳の誤差が生じるので，人間より 1 歳以上優れた性能で年齢を推測でき，その有効性を実証できたと言える．また**表 7.6** の 20 名のモニターにおける最小値は 5.45 歳，最大値は 10.69 歳である．その差は 5 歳以上あり，人間が顔だけから年齢を推測する能力は，個人毎に大

きな開きがあることを確認できる．ゆえに，今後はモニターの人数を更に増やして正確な「見かけ年齢」のデータを採取することが必要であると考えられる．

**表 7.6**：FG-Net aging database における提案手法とモニターの「見かけ年齢」の MAE

| **手法** | | **MAE** |
|---|---|---|
| LGDPHS+GAAM | | 6.24 |
| モニター | A | 6.83 |
| | B | 8.95 |
| | C | 7.84 |
| | D | 10.69 |
| | E | 6.55 |
| | F | 5.82 |
| | G | 7.58 |
| | H | 9.41 |
| | I | 8.85 |
| | J | 5.69 |
| | K | 7.73 |
| | L | 6.50 |
| | M | 6.95 |
| | N | 5.45 |
| | O | 7.19 |
| | P | 7.84 |
| | Q | 5.79 |
| | R | 8.10 |
| | S | 8.61 |
| | T | 8.06 |
| | A－T の平均 | 7.52 |

次に**図 7.3** において提案手法の累積スコア（CS）とモニターによる「見かけ年齢」の CS を比較した図を示す．**図 7.3** から年齢誤差が 0−7 歳以下の場合，提案手法は累積スコアで「見かけ年齢」より 10%以上優れている. しかし誤差が 14 歳以上になると,「見かけ年齢」は提案手法より累積スコアで上回ることを確認できる. またモニターによる「見かけ年齢」

は 94.5%と非常に高い数値を示すことから，人間は誤差 15 歳以内という条件であれば，非常に正確な年齢推定を行えると言える．また提案手法は極端な推定間違いを含んでいることが影響し，年齢誤差を 15 歳から無限大方向に近づけていった場合，CS は 90%程の値を維持することが予測される．以上のことから提案手法は年齢誤差が 14 歳以下という条件下でモニターによる「見かけ年齢」よりも高い確率で正確な年齢の推定ができると言える．

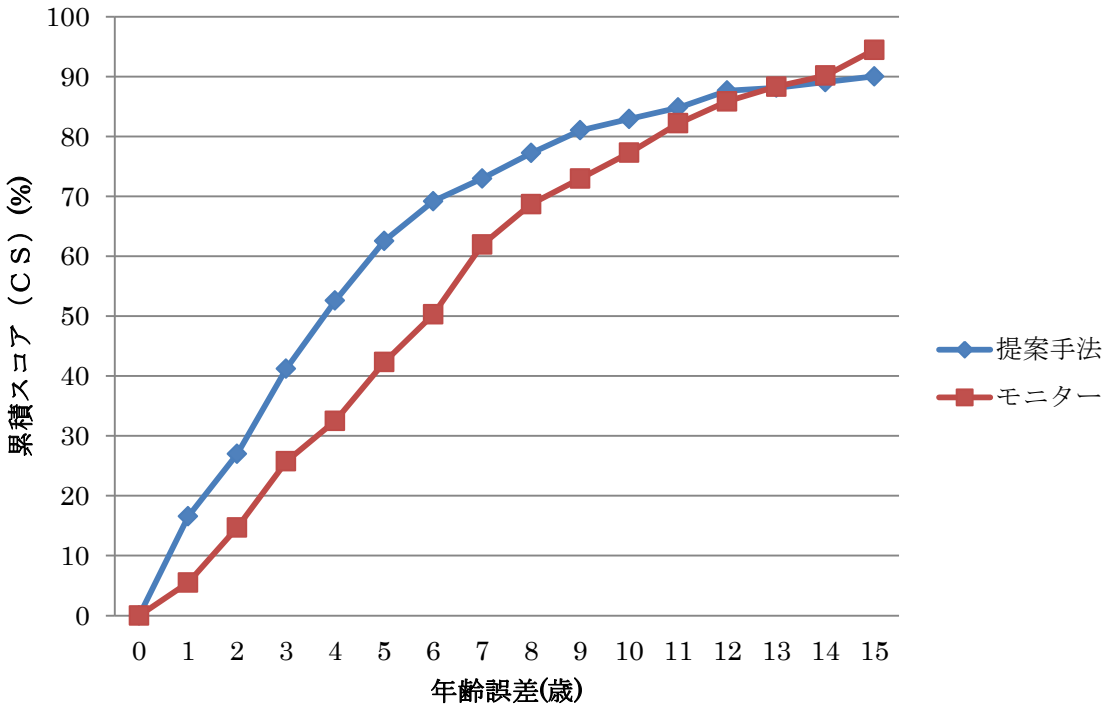



**図 7.3**：提案手法とモニターによる「見かけ年齢」の累積スコア（CS）の比較

最後に定性的な評価の結果を示す．**図 7.4** において，提案手法の推定年齢，モニター20 名による「見かけ年齢」の平均を散布図として示す．また直線は実年齢を表している．**図 7.4** からモニターは常に年齢を高めに推定する傾向にあることを確認できる．また対象年齢が 25 歳以下の場合，提案手法で得られた結果は，データのばらつきが激しいものの，モニターの散布データと比較して実年齢直線の周辺に散布している．しかし対象年齢が 25 歳以上の場合，実年齢より年齢を低く推定する傾向にあり，対象年齢が高く設定されていくにつれ，推定年齢と実年齢の誤差は次第に大きくなることが予測できる．これは学習画像枚数が年齢の増加に伴って徐々に少なくなっていることが原因であると考えられる．

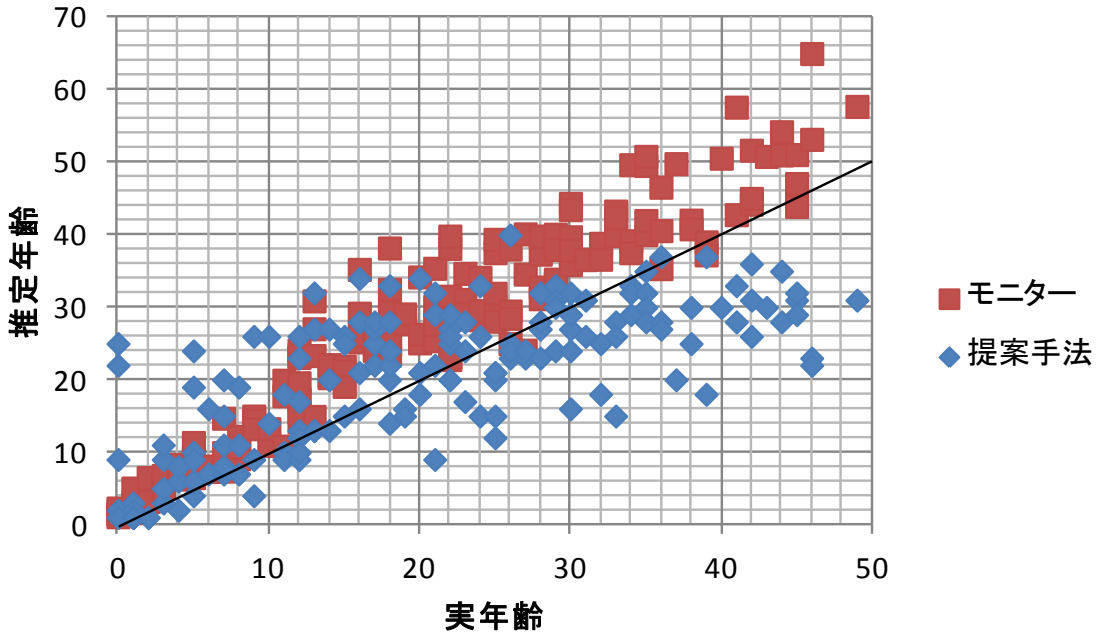



**図 7.4**：実年齢に対するモニターと提案手法の推定年齢の散布図

## 7.4. 課題

**・学習用画像の選別**

学習画像として用いた FG-Net aging database はスナップショットやパスポートの写真を多く含むことから照明，顔の向きや表情の変化を多く含んでいる．提案手法は学習に用いる画像に影響を受けることを予測でき，性能の向上のためには今後，どのような学習画像を用いればよいのかを突き詰めていく必要がある．

**・処理の高速化，メモリ使用量の削減**

提案手法の結果出力までの処理速度は 100ms 程掛かり，映像への適用へと発展させる場合，高速化を検討する必要がある．本研究にて用いた Gabor フィルタは計算負荷が高い畳み込み演算を含む．またスケール数，回転角数が増加すると特徴量の次元数も増え，メモリ使用量が多くなる．今後は Gabor フィルタをダウンサンプリングや特徴点周辺領域に絞って適用するといった改善が必要である．

## 7.5. まとめ

本章では大局的・局所的な 2 つの要素から構成される特徴量を用いた年齢・性別推定ア

ルゴリズムを提案した．大局的特徴量として顔全体の濃淡値，つまり「見え」を数値化した GAAM のパラメータを用い，局所的特徴量として GAAM により正規化された顔領域から抽出した LGDPHS を用いる．これより位置ズレ誤差や照明変動などのランダムノイズに対して頑強な局所特徴が抽出できる．年齢推定の実験では FERET database を用い，提案手法の MAE は 6.24 歳であり，従来法と比較して最も優れていることが確認できた．またモニター20 人の平均の MAE は 7.52 歳であり，提案手法はモニターより 1.2 歳以上優れていることを確認できる．このことから人間が顔から年齢を推測するとおよそ 7－8 歳の誤差があり，提案アルゴリズムは十分に人間と同等の年齢推定能力を有しており，その有効性を確認できた．また性別分類の実験では FERET database を用い，提案手法の分類率は 89.4％となり，従来手法と比較して最も優れた性能であることを確認できた．

# 参考文献


[1] R. Gross，I. Matthews and S. Baker：“Generic vs. Person Specific Active Appearance Models”，Image and Vision Computing，vol.23，no.12，pp.1080-1093（2005）

[2] A.J. Smola and S. Bernhard：“A Tutorial on Support Vector Regression”，Statistics and Computing，vol.14，no.3，pp.199-222（2004）

[3] The FG-NET Aging database, http://sting.cycollege.ac.cy/~alanitis/ fgnetaging/index.htm.

[4] P.J. Phillips，H.Wechsler，J.Huang and P. Rauss：“The FERET Database and Evaluation Procedure for Face Recognition Algorithms”，Image and Vision Computing，vol.16，no.10，pp.295–306（1998）

[5] Y.Fang，Z. Wang：“Improving LBP Features for Gender Classification”，Proc International Conference Wavelet Analysis and Pattern Recognition，pp.373–377（2008）

[6] A. Gunay, V.V. Nabiyev：“Automatic Age Classification with LBP”，23rd International Symposium on Computer and Information Sciences，pp.1–4（2008）

[7] B. Xia，H. Sun, and Bao-Liang Lu：“Multi-View Gender Classification Based on Local Gabor Binary Mapping Pattern and Support Vector Machines”，In Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. IEEE, pp.3388-3395（2008）

[8] X. Geng，Z. Zhou，K. Smith-Miles：“Automatic Age Estimation Based on Facial Aging Patterns”，IEEE Transactions on Pattern Analysis and Machine Intelligence，vol.29，no.12，pp.2234–2240 （2007）

[9] W. Zhang，S. Shan，W. Gao，X. Chen，H. Zhang：“Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition”，Tenth IEEE International Conference on Computer Vision，vol.1，pp.786–791（2005）

# 第8章. 顔のキーパートを用いたLGDPHSによる顔画像からの表情認識

## 8.1. まえがき

本章ではLGDPHSを用いた顔画像からの表情認識アルゴリズムについて述べる．また提案アルゴリズムの性能を検証するため，実験では従来法との性能比較を行う．感情・表情について先駆的な研究を行った心理学者のPaul Ekmanは，人間の感情は6つの普遍的な感情リスト（怒り，嫌悪，恐怖，幸福，悲しみ，驚き）に分類できると提唱した．これより表情認識の研究は，6つの表情(怒り：Ang.，嫌悪：Dis.，恐怖：Fea.，幸福：Hap.，悲しみ：Sad，驚き：Sur.)に無表情(Neu.)を加えた計7つの表情を分類することでアルゴリズムの性能を検証することが一般的となり，さまざまな手法が現在に至るまで提案されている．例えばAAM**[1,2]**やActive Shape Model (ASM)**[3]**等のモデルベースの手法は，顔の頂点座標の相関関係や，連続画像における座標の動き情報を利用する幾何学的アプローチであると言える．また顔画像全体に対してLBPやGaborフィルタを適用し，局所特徴量を抽出するアペアランス的アプローチも広く用いられている．例えばShanらはLBP特徴をAdaBoost**[4]**で学習して最も識別可能な特徴量を選別し，それをSVMによって識別する手法を提案している**[5]**．またHongらはGaborフィルタの周期性を利用した顔の特徴マップを生成し，PCAに加えて線形判別分析(LDA)を適用することで，特徴量の圧縮と選別を行う手法を提案している**[6]**．多くの場合，LBPやGaborフィルタは顔の全体領域に対して適用され，抽出された特徴マップはブロック分割してヒストグラム化される．このヒストグラム化のプロセスにより微小の位置ズレ誤差の問題を回避できる．しかし人物毎の顔構造の違いや表情変化により引き起こされる目・口・鼻などの各顔器官の相対的な位置ズレ誤差は大きな課題である．これは特徴空間における各表情クラスの分布の重なりを招き，識別を難しくする

近年では，このような位置ズレ誤差の問題を防ぐため，顔の主要な領域（キーパート）や重要な特徴点（キーポイント）を利用した局所的なマッチング法が提案されている．Zishengらは一般物体認識に用いられるBag-of-Words (BoW)を表情認識の特徴量に応用している．具体的には複数のキーポイントを検出し，その周辺領域から算出される記述子を

一つの頻度ヒストグラムとして集約することで，位置情報は失われるが，一方で位置ズレ誤差の問題を回避している**[7]**．また Zhang らは顔の動き特徴量に焦点を当て，Gabor 特徴量を使い，最適な局所パッチを抽出し，それを用いた局所的なマッチング手法を提案している**[8]**.

本章では位置ズレ誤差の課題解決のため，顔のキーパートにおける局所特徴量を抽出する独自の表情認識手法を提案する．顔のキーパートは GAAM を使い算出したキーポイントを基準にして抽出する．このキーパートから算出された特徴量によって人物毎の顔構造の違いや，対象者の表情変化に起因する位置ズレ誤差により性能が低下する課題の解決を期待できる．本章の構成として，8.2 節で提案手法について述べ，8.3 節において実験とその結果を示し，8.4 節で課題を挙げる．最後に 8.5 節で本章をまとめる．

## 8.2. 提案する表情認識アルゴリズム

本節では提案する表情認識アルゴリズムについて述べる．そのフレームワークを**図 8.1** に示す．顔全体領域から特徴量を抽出すると人物毎の顔構造の違いや表情の変化による顔のキーパート（目，眉，鼻，口）の位置ズレ誤差が生じる問題がある．そこで前処理として GAAM を用いて顔のキーポイントを抽出し，更に独自に定義した基準に基づき，キーパートを切り出す．これにより位置やスケール，傾きの不変性を保持した特徴量抽出が期待できる．この正規化されたキーパートのみを用いることで，顔の筋肉の微小の動きに起因する顔の局所領域のズレも回避でき，更に認識対象者や表情の変化に対しても位置関係の不変性を保持することが期待できる．

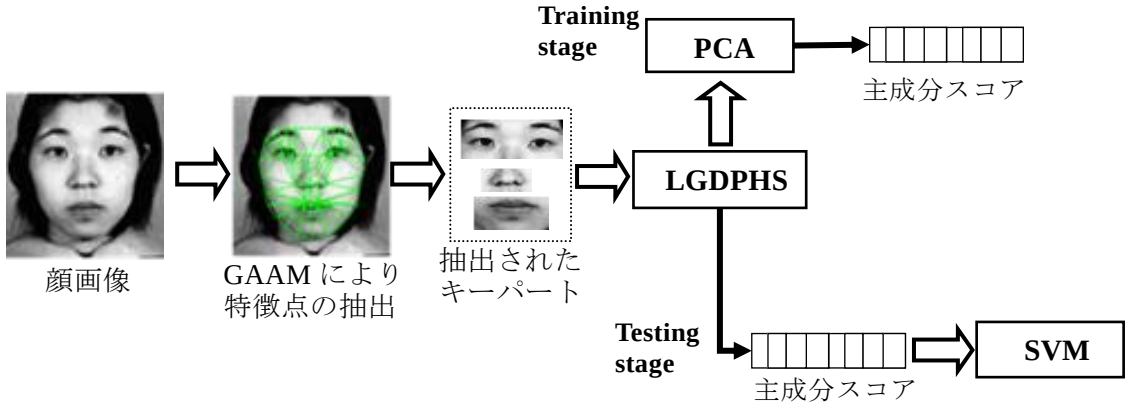



**図 8.1**：提案する表情認識アルゴリズムのフレームワーク

次に，切り出した各キーパートから特徴量を抽出する．**図 8.2** にてパート毎の特徴抽出のフレームワークを示す. 特徴量は 6 章で提案した Local Gabor Directional Pattern Histogram Sequence (LGDPHS)を用いる．これは，まず各パートに対して Gabor フィルタを適用し，顔の表情変化の局所的な特徴を含む Gabor 絶対値成分画像（GMP）を算出する．そして GMP の重要なエッジ応答の方向情報を符号化する LDP を適用することで，顔の細かなパターン情報を洗練化できる．また本章の提案手法が従来の LGDPHS と異なる点として，LGDP マップがパート毎に抽出され，各パートにおいてブロック分割を行う点である．しかし，パートの数だけパラメータ数が従来よりも増えるので，従来と比較してパラメータ値の設定には注意を払う必要がある．ゆえに，特に重要な各パートのブロック数を最適に調整するため，本章の実験ではブロック数を 3 種に設定し，各ブロック数での性能検証を行う. 最後にパート毎にヒストグラム化された特徴量は一つのベクトルとして結合される.

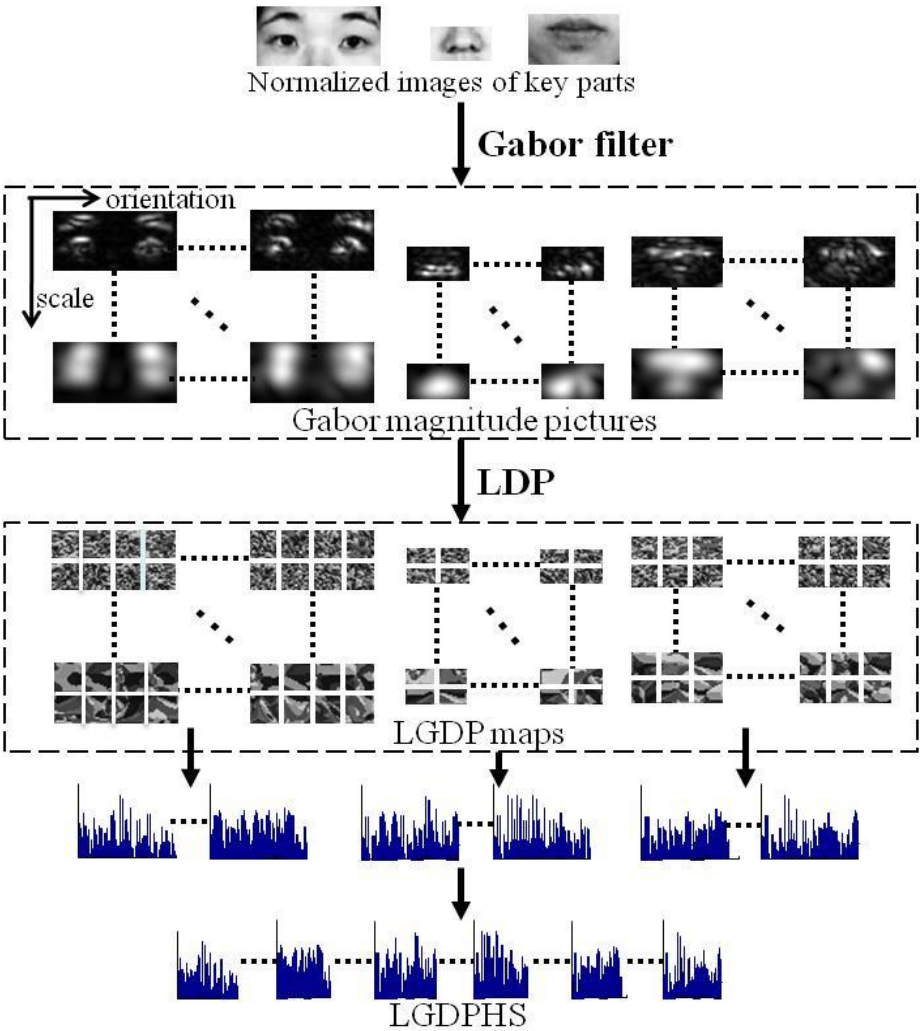
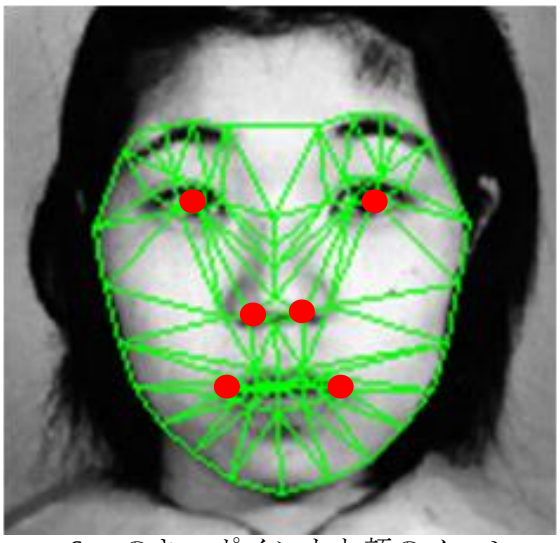



**図 8.2**：顔のキーパートを基にした提案する LGDPHS のフレームワーク

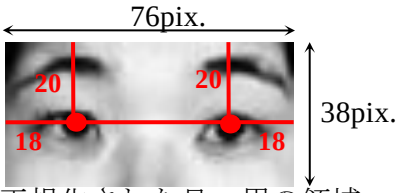
次に，**図 8.1** において学習時では，特徴ベクトルに対して PCA を適用し，LGDPHS の次元数を削減して不用な情報を省く．PCA より算出された固有ベクトルと特徴ベクトルの内積から主成分スコアが計算され，これを新たな特徴量として扱う．テスト時では，未知の画像から主成分スコアが学習時と同様に算出され，最終的に識別器として SVM を用い，7 つの表情の中から最適な表情に分類する．

### 8.2.1. 顔のキーパート抽出

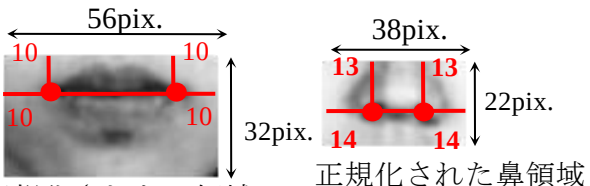
本項では独自に定義した顔のキーパートの切り出し基準について述べる．GAAM を使い顔のキーポイントを検出し，その座標を基準に顔のキーパートを切り出す．本実験では GAAM の顔メッシュの頂点数を 68 点に設定し，**図 8.3(**上)にて各頂点の位置を示す．更に 68 頂点内から顔のキーポイントとして 6 つの頂点（**図 8.3(**上)の太丸）を選出する．



6 つのキーポイントと顔のメッシュ




正規化された目，眉の領域




正規化された口領域　　正規化された鼻領域

**図 8.3**：顔のメッシュと正規化されたキーパート

キーパートは 3 つの領域から構成される．第 1 に目・眉のキーパートである．このキーパートを抽出するため，左右の瞳に位置するキーポイントを基準にする．**図 8.3(**中)にその切り出し規則を示す．それは瞳に位置する 2 点を基準とし，位置・スケール・傾きの調整のため，瞳の位置から画像の端へ向かって水平方向に 18 pix.，垂直方向に 20 pix.のキーパートを切り出す．このとき画像サイズは 76×38 pix.へと正規化される．第 2 のキーパートは鼻領域である．**図 8.3(**下(右))にその切り出し規則を示す．それは鼻下に位置する 2 点のキーポイントを基準にして，水平方向に 14 pix.，垂直方向に 13 pix.の規則に従い，画像を切り出す. 切り出し後の画像サイズは 38×22 pix.となる. 最後のキーパートは口領域である．**図 8.3(**下(左))にその切り出し規則を示す．キーポイントは口の両端に位置し，そのキーポイントを基準に水平方向に 10 pix.，垂直方向に 10 pix.の規則に従い画像を切り出す．切り出し後の画像サイズは 56×32 pix.となる．

## 8.3. 実験及び考察

本節では，提案する表情認識アルゴリズムの性能検証を行い，その結果と考察について述べる．本実験では表情認識の評価に広く用いられているデータベースの Japanese Female Facial Expression (JAFFE)データベースを用いて性能検証を行う．**図 8.4** にて JAFFE データベースのサンプル画像を示す．このデータベースは日本人女性 10 人の 7 つの顔表情｛6 つの表情(怒り：Ang.，嫌悪：Dis.，恐怖：Fea.，幸福：Hap.，悲しみ：Sad，驚き：Sur.)に無表情：Neu.を加えた計 7 つ｝を含み，計 213 枚の画像から構成されている．各人物は各表情に対して 3，4 枚の画像を含み，顔は画像の中心に位置している．本実験では画像の元サイズは 256×256 pix.と十分に大きいので，サイズは左右の瞳の位置を基準に全て 150×150 pix.へとリサイズする．

**図 8.4**： JAFFE データベースにおける 2 人の人物のサンプル画像．左から怒り：Ang.，嫌悪：Dis.，恐怖：Fea.，幸福：Hap.，無表情：Neu.，悲しみ：Sad，驚き：Sur.

ここで実験項目について述べる．Person-independent と Person-dependent な表情認識の 2 つの評価法における性能検証を行う．ここで Person-independent な表情認識とは，学習画像内の人物とテスト画像内の人物が異なる場合であり，学習とテストにおける人物の相違が表情認識の性能に及ぼす影響について評価できる．また Person-dependent な表情認識とは学習画像内の人物とテスト画像内の人物は同一人物が含まれ，学習とテストでは同一人物の異なる画像を用いる．これより Person-independent な表情認識における人物の相違という識別を難しくする条件がなくなり，純粋な表情変化に対してのみの性能評価を行うことができる．

具体的な評価項目として，Person-independent な表情認識においては 5 つの項目（① - ⑤）を設ける．それは，①LGDP マップを分割するブロック数を 3 つ設定したとき，累積寄与率と表情分類率の関係，②3 つのブロック数を設定したとき表情分類率と特徴量の次元数の関係，③4 つの異なるカーネル関数による SVM と 3 つのブロック数を用い，各条件での分類率の性能評価，④7 つの表情についてそれぞれの分類率を Confusion Matrix を使い示す．最後に，⑤従来手法との分類率の性能の比較実験である．ここで学習画像には 7 名の人物における 149 枚の画像を用い，一方テスト画像には学習時と異なる 3 名の人物における 64 枚の画像を用いる．

次に Person-dependent な表情認識における実験では，4 つの実験項目（① - ④）を設定する．それは①線形 SVM を用い，3 つのブロック数を設定したとき累積寄与率と分類率の関係，②3 つのブロック数を設定したとき表情分類率と特徴量の次元数の関係，③7 つの表情についてそれぞれの分類率である Confusion Matrix による評価，④従来手法との分類率の性能の比較実験である．ここで学習とテスト画像は同一人物の画像を用いる．学習画像は各人物の表情について 2，3 枚の画像を用い，テスト画像は学習画像とは異なる 1，2 枚の画像を用いる．そして学習画像の合計としては 137 枚，テスト画像では 76 枚を使用する．

### 8.3.1. 実験環境

本実験における実験環境を**表 8.1** に示す．Windows 上で動作するソフトウェアとして構築している．

**表 8.1**：実験環境

| | |
|---|---|
| OS | Windows XP Professional SP3 |
| CPU | Intel ® Core™2 Quad Q6600 @ 2.40GHz |
| メモリ | 2.00GB |
| 開発言語 | C 言語 |
| 開発環境 | Visual Studio .NET 2008 |
| 使用ライブラリ | CLAPACK |

### 8.3.2. Person-independent な表情認識の実験結果・考察

本項では Person-independent な表情認識における実験結果について示す. 各キーパートから抽出した LGDP マップは複数のブロックに分割する．このとき各パートのブロック数は最適値に設定する必要がある. そこで本実験では 3 種類のブロック数を試す．ここで**表 8.2** にて 3 種類のブロック数の詳細を示す. 次に**表 8.2** に示す 3 つのブロック数（20, 34, 52 blocks）において累積寄与率と表情分類率の関係グラフを**図 8.5** に示す．この時，識別器としては線形 SVM を用いている.

**表 8.2**：3 種類のブロック数における各キーパートの分割方法とそのブロック数

| 合計のブロック数 | 目・眉領域の<br>ブロック数 | 鼻領域の<br>ブロック数 | 口領域の<br>ブロック数 |
|---|---|---|---|
| **20** | 10(5×2) | 4(2×2) | 6(3×2) |
| **34** | 18(6×3) | 4(2×2) | 12(4×3) |
| **52** | 28(7×4) | 9(3×3) | 15(5×3) |

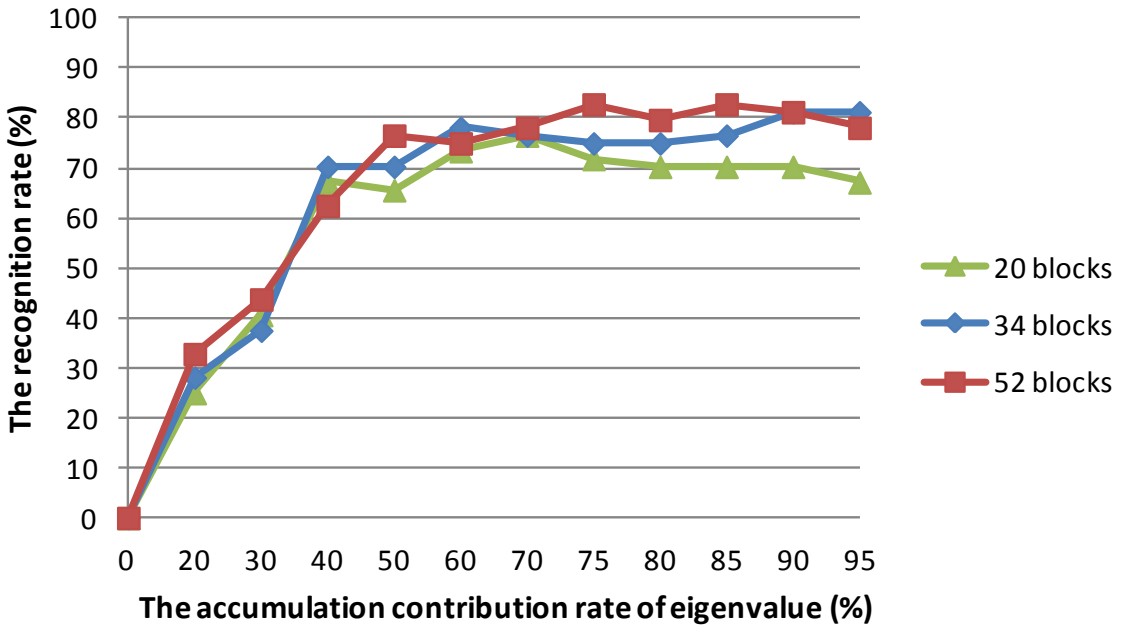



**図 8.5**： Person-independent な表情認識における 3 つのブロック数を設定したときの累積寄与率と表情分類率の関係グラフ

**図 8.5** からブロック数の合計が 52，固有値の累積寄与率が 85%のとき，表情の分類率は 82.8%となり，提案アルゴリズムは最も優れた性能を発揮することを確認できる．また 3 つのブロック数（20，34，52 blocks）はそれぞれ累積寄与率が 85%以上になると，分類率は上昇しない．これより累積寄与率が 85%以上の特徴情報は表情を識別するうえで不必要な情報，即ちノイズである．

**表 8.3**：累積寄与率が 85%での表情の分類率と 3 つのブロック数での次元数

| 合計ブロック数 | 分類率 (%) | 次元数 | PCA 適用後の次元数 |
|---|---|---|---|
| **20** | 70.3 | 25,800 | 87 |
| **34** | 76.6 | 43,860 | 91 |
| **52** | **82.8** | 67,080 | 93 |

**表 8.3** は，3 つの異なるブロック数（20，34，52 blocks）における次元数と各ブロック数での表情の分類率を示している．**表 8.3** の累積寄与率は，**図 8.5** で最も優れた分類率を示した 85%とする．次元数はブロック数に依存しており，例えばブロック数の合計が 34 ブロックのとき Gabor フィルタの Gabor カーネルの数は 5×6＝30，ヒストグラムのビン幅は 43 であり，この時 LGDPHS の次元数は 34×30×43= 43,860 となる．このように次元数が非常に大きくなるので PCA を適用することで，149 枚の学習画像を用いた場合，次元数は 91 次元にまで削減できる．次に**表 8.4** において 4 つの異なるカーネルを使う SVM と 3 つのブロック数（20，34，52 blocks）を用いるとき，各条件における分類率を示す．

**表 8.4**：4 種のカーネルの SVM と 3 つのブロック数を用いた時の各条件での表情分類率

| カーネル名 | 分類率 (%) | | |
|---|---|---|---|
| | 合計ブロック数 = 20 | 合計ブロック数 = 34 | 合計ブロック数 = 52 |
| **線形** | 70.3 | 76.6 | **82.8** |
| **多項式** | 25.0 | 50.0 | 48.4 |
| **RBF** | 35.9 | 51.6 | 39.1 |
| **シグモイド** | 65.6 | 68.8 | 64.1 |

**表 8.4** からブロック数に関わらず線形 SVM は他のカーネル関数を用いた SVM よりも優れた性能を発揮し，特にブロック数が 52 かつ線形 SVM を用いた場合，最も優れた性能になると確認できる．

**表 8.5** では，JAFFE データベースを使った Person-independent な表情認識の Confusion Matrix を示す．これは**図 8.5**，**表 8.4** から最も優れた性能を示すことが確認できた各種パラメータを利用している．つまり SVM のカーネル：線形，ブロック数：52，累積寄与率：85%の条件である．Confusion Matrix は左端の列が対象表情を示しており，表の対角成分は各対象表情が正解の表情であると認識した確率(%)を表現している．また対角成分以外は誤分類した確率であり，対象表情が他の表情に分類した確率は同行へと記される．

**表 8.5**： Person-independent な表情認識における Confusion Matrix

| | **Ang. (%)** | **Dis. (%)** | **Fear (%)** | **Hap. (%)** | **Neu. (%)** | **Sad (%)** | **Sur. (%)** |
|---|---|---|---|---|---|---|---|
| **Ang.** | **100** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Dis.** | 0 | **100** | 0 | 0 | 0 | 0 | 0 |
| **Fear** | 0 | 40 | **40** | 0 | 10 | 10 | 0 |
| **Hap.** | 0 | 0 | 0 | **88.9** | 11.1 | 0 | 0 |
| **Neu.** | 0 | 0 | 0 | 0 | **100** | 0 | 0 |
| **Sad** | 0 | 0 | 0 | 0 | 0 | **100** | 0 |
| **Sur.** | 0 | 0 | 11.1 | 0 | 33.3 | 0 | **55.6** |

**表 8.5** から怒り(Ang.)，嫌悪(Dis.)，無表情(Neu.)，悲しみ(Sad)の表情は 100%の分類率を示しているが，恐怖(Fea.)と驚き(Sur.)は共に 56%以下であり，十分な分類率が得られていない．特に恐怖の表情は嫌悪に間違いやすいことを確認できる．これは**図 8.4** の表情画像を参考にすると，恐怖の表情が嫌悪に非常に似ていることが原因であると考えられる．具体的には左右の眉間の皺や，鼻周辺に皺があること，口の形状は無表情と比較して微小の変化であることなどが挙げられる．また驚きの表情については，多くの顔画像が大きく口を開けていることに起因して GAAM のフィッティングが不正確になり，分類率の性能低下を招いたと考えられる．

次に提案手法の性能と従来手法との性能比較実験の結果を**表 8.6** に示す．本実験では 6 つの従来手法を比較対象とする．それらは Gabor 特徴量と LBP から構成される LGBP，更に Gabor フィルタを適用しない単独の LBP と LDP を比較対象とする．これら 3 つの手法の実験環境は提案手法と同一環境にて実験を行う．その他の 2 つの従来手法については異なる実験環境であるが，JAFFE データベースを用いて Person-independent な表情認識について実験を行った各文献内で記されている結果である．

**表 8.6**：Person-independent な表情認識における提案法と従来法の表情分類率の比較結果

| 参考文献，時期 | 特徴量 | 分類率 (%) |
|---|---|---|
| | **提案手法（GAAM+Parts-based LGDPHS）** | **82.8** |
| | **LGBP(Gabor + LBP)** | 67.2 |
| | **LBP** | 53.1 |
| | **LDP** | 54.7 |
| **[9], 2011** | **LGBP based keyparts** | 77.6 |
| **[4], 2008** | **Boosted-LBP** | 81.0 |

**表 8.6** から，提案手法の分類率は 82.8%に達しており，他の手法より優れていることを確認できる．前処理として左右の瞳の位置を基準にして画像の正規化を行い，特徴量抽出を行う従来法の LGBP, LBP や LDP は背景領域を含まないように 82×92 pix.のサイズに正規化する．しかしそれらは人物の相違や表情の変化により引き起こされる位置ズレ誤差により，性能が低下する問題がある．この問題により，これら 3 つの従来法について分類率は 70%にも到達していない．対照的に顔の局所的なパートのみに対して特徴抽出を行う提案手法を含めたその他の手法（LGBP based keyparts，Boosted-LBP）については，全て 77%以上と高い分類率であることを確認できる．特に提案手法は，LGDPHS を特徴量として用いており, 重要性の高いエッジ方向の情報を特徴量として扱うことで最も高い分類率を示し，その有効性を確認できる．

### 8.3.3. Person-dependent な表情認識の実験結果・考察

本項では Person-dependent な表情認識における実験結果について述べる．ここで本実験では識別器として線形 SVM を用いる．3 つのブロック数（20，34，52 blocks）を設定したときの累積寄与率と表情分類率の関係グラフを**図 8.6** に示す. また 3 つのブロック数（20，34，52 blocks）におけるキーパート毎の詳細なブロック数は，**表 8.2** と同値に設定する．

**図 8.6** から 3 つのブロック数の各グラフは何れも累積寄与率が 85%以上の場合，分類率

が 90%に達している．そして 3 つのブロック数（20，34，52 blocks）の中で最も優れた性能を示すのはブロック数が 34 かつ累積寄与率が 90%の条件であり，このとき表情の分類率は 94.7%となる．

また**表 8.7** において，累積寄与率が 90%の場合における 3 つのブロック数での次元数と PCA 適用後の次元数，そして表情分類率を表にまとめる．

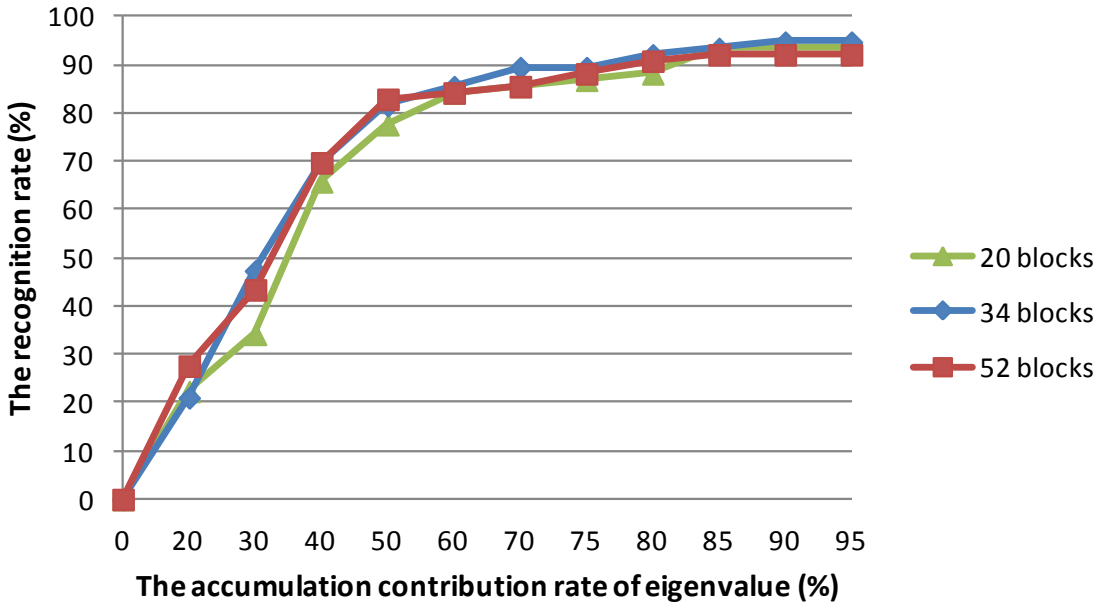



**図 8.6**：Person-dependent な表情認識において 3 つのブロック数（20，34，52 blocks）を設定したときの累積寄与率と表情分類率の関係グラフ

**表 8.7**：累積寄与率が 90%での表情の分類率と 3 つのブロック数における次元数の関係

| ブロック数 | 分類率(%) | 次元数 | PCA 適用後の次元数 |
|---|---|---|---|
| **20** | 93.4 | 25,800 | 98 |
| **34** | **94.7** | 43,860 | 100 |
| **52** | 92.1 | 67,080 | 102 |

**表 8.8** では JAFFE データベースを使った Person-dependent な表情認識の Confusion Matrix を示す. この Confusion Matrix はブロック数が 34 かつ累積寄与率が 90%の条件の下で実験を行っている．**表 8.8** から各表情の分類率は 80%以上に達していることを確認でき，特に怒り(Ang.)，嫌悪(Dis.)，無表情(Neu.)，驚き(Sur.)の表情は 100%の分類率を示している．また恐怖(Fear)の表情は Person-independent な表情認識と同様に最も低い性能であり，恐怖は 7 つの表情の中で最も認識が難しい表情であると言える.

**表 8.8**：Person-dependent な表情認識における Confusion Matrix

| | **Ang. (%)** | **Dis. (%)** | **Fear (%)** | **Hap. (%)** | **Neu. (%)** | **Sad (%)** | **Sur. (%)** |
|---|---|---|---|---|---|---|---|
| **Ang.** | **100** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Dis.** | 0 | **100** | 0 | 0 | 0 | 0 | 0 |
| **Fear** | 0 | 0 | **83.3** | 0 | 0 | 0 | 16.7 |
| **Hap.** | 0 | 0 | 0 | **91.7** | 0 | 0 | 8.3 |
| **Neu.** | 0 | 0 | 0 | 0 | **100** | 0 | 0 |
| **Sad** | 0 | 0 | 0 | 9.1 | 0 | **90.9** | 0 |
| **Sur.** | 0 | 0 | 0 | 0 | 0 | 0 | **100** |

次に提案手法と従来手法との性能比較実験の結果を**表 8.9** に示す．本実験では 7 つの従来手法を比較対象とする．それらは Person-dependent な表情認識の実験と同様に Gabor 特徴量と LBP から構成される LGBP，更に Gabor フィルタを適用しない単独の LBP と LDP を比較対象とする．これら 3 つの手法の実験環境は同一環境において実験を行う．その他 4 つの従来法は，異なる実験環境となるが，各文献内で記されている JAFFE データベースによる Person-dependent な表情認識の実験結果を用いる．ここでその他 4 つの従来手法について概説する．それらは本稿内でそれぞれ Patch-based Gabor，Gabor + FSLP，DCT，KCCA と称する. まず Zhang らが提案した Patch-based Gabor は 8.1 節内で紹介している. 次に Gabor + FSLP は G. Guo らにより提案され，FSLP 法という SVM と同様のマージン最大化法を用い，少数の標本サンプルによる学習を可能にしている．特徴量としては，ラベル付けされた 34 頂点に対して Gabor フィルタを適用する**[10]**. 次に DCT は J. Bin らにより提案され，離散コサイン変換(DCT)を特徴次元数の削減に用いた表情認識法である**[11]**. 最後に KCCA は Z. Wenming らにより提案され，手動で顔画像に 34 個のキーポイント座標をラベル付け

し，Gabor wavelet 変換を用い，それら頂点をラベル化されたグラフ(LG)へと変換する．学習では LG のベクトルと意味論的な表情ベクトルとの相関をカーネル正準相関分析(KCCA)によって学習する**[12]**．**表 8.9** から Gabor 特徴量を用いる手法（提案手法，LGBP，Patch-based Gabor，Gabor+FSLP，KCCA）は LDP や LBP のようなバイナリー化による特徴抽出法と比較し，高い分類率を示すことを確認できる．これより Gabor 特徴量はその周期性から顔の皺等の細かな特徴成分が上手く抽出でき，表情認識の特徴量として有効であると言える. また本章で提案した顔のキーパートを基にした LGDPHS よる表情認識アルゴリズムは分類率 94.74%と従来手法と比較して，最も高い分類率であることを確認できる．本章の実験の結果として，Person-independent，Person-dependent な表情認識の両評価法において，提案手法が有効であることを実証できた.

**表 8.9**：Person-dependent な表情認識における提案法と従来法の表情分類率の比較結果

| **参考文献，時期** | **特徴量** | 分類率(%) |
|---|---|---|
| | **提案手法（GAAM+Parts-based LGDPHS）** | **94.7** |
| | **LGBP(Gabor + LBP)** | 88.1 |
| | **LBP** | 72.4 |
| **[8]，2011** | **Patch-based Gabor** | 92.9 |
| **[10]，2005** | **Gabor + FSLP** | 91.0 |
| | **LDP** | 76.3 |
| **[11]，2008** | **DCT** | 79.3 |
| **[12]，2006** | **KCCA** | 77.1 |

## 8.4. 課題

本章では顔のキーポイント探索に GAAM を用いた．これは学習画像内にテスト画像の人物が含まれない場合でも，安定した AAM のメッシュの収束が可能であることに優位性がある．しかし，GAAM は 6 点のキーポイントを探索するために 68 点の頂点探索が必要であり, 学習時のラベル付けや, テスト時のキーポイント探索において非常に効率が悪い．ゆえにキーポイント探索手法は今後の検討課題であると言える.

また，実験で用いた固有値の数は累積寄与率から決定しており，データに依存した結果が得られている．ゆえに今後は十分な数の画像データを収集し，それを用いた実験を行う必要があると言える.

## 8.5. まとめ

本章では顔のキーパートに対して LGDPHS を適用した特徴量を用い，顔画像からの表情認識手法について述べ，その性能を検証した．提案手法は正規化されたキーパートのみから特徴抽出を行うことで認識対象者や表情の変化によって引き起こされる位置ズレ誤差の問題に頑強な特徴量抽出を期待できる．実験では Person-independent と Person-dependent な表情認識の 2 つの実験を実施した．JAFFE データベースを使い，Person-independent な表情認識の実験では，提案手法は従来法と比較して優れた性能であることを確認できた．また，怒り(Ang.)，嫌悪(Dis.)，無表情(Neu.)，悲しみ(Sad)の表情は 100％の識別率に達している．しかし提案手法は恐怖の表情を嫌悪として分類し易い．これは互いの表情が，左右の眉間の皺や，鼻周辺の皺，口の形状において非常に似通っていることに起因していると考えられる．Person-dependent な表情認識の実験では，提案手法の分類率は 94.74％に達しており，従来手法と比較しても優れた性能であることを確認できた．

# 參考文獻


[1] T.F. Cootes，J.E. Gareth， and J.T. Christopher：“Active Appearance Models”，IEEE Transactions on Pattern Analysis and Machine Intelligence，vol.23，no.6，pp.681-685（2001）.

[2] I. Matthews，and S. Baker：“Active Appearance Models Revisited”，International Journal of Computer Vision，vol.60，no.2，pp.135-164（2004）.

[3] T.F. Cootes，C.J. Taylor，D.H. Cooper and J. Graham：“Active Shape Models-Their Training and Application”，Computer Vision and Image Understanding，vol.61，no.1，pp. 38-59 (1995)

[4] P. Viola and M. Jones：“Rapid Object Detection using a Boosted Cascade of Simple Features”，Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition，vol.1，pp.I-511-I-518（2001）

[5] C. Shan，G. Shaogang and P.W. McOwan：“Facial Expression Recognition based on Local Binary Patterns: A Comprehensive Study”，Image and Vision Computing，vol.27，no.6，pp.803-816 (2009)

[6] H.B. Deng，L.W. Jin，L.X. Zhen and J.C. Huang：“A new Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA”，International Journal of Information Technology，vol.11，no.11，pp.86-96 (2005)

[7] Z. Li，J. Imai， M. Kaneko：“Face and Expression Recognition Based on Bag of Words Method Considering Holistic and Local Image Features”，International Symposium on Communications and Information Technologies，pp. 1-6 (2010)

[8] L. Zhang，D. Tjondronegoro：“Facial Expression Recognition Using Facial Movement Features”，IEEE Trans. Affective Computing，vol.2，no.4，pp.219–229 (2011).

[9] A. Bafandehkar, M. Nazari, M. Rahat：“Pictorial Structure Based Keyparts Localization for Facial Expression Recognition using Gabor Filters and Local Binary Patterns Operator”，International Conference of Soft Computing and Pattern Recognition, pp.429–434 (2011).

[10] G. Guo and C.R. Dyer：“Learning from Examples in the Small Sample Case: Face Expression Recognition”，IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics，vol 35，no. 3， pp. 477-488 (2005).

[11] J. Bin，Y. Guo-Sheng and Z. Huan-Long：“Comparative Study of Dimension Reduction and Recognition Algorithms of DCT and 2DPCA”，Proc. Int. Conf. Machine Learning and Cybernetics，pp. 407-410（2008）

[12] Z. Wenming, Z. Xiaoyan, Z. Cairong, and Z. Li：“Facial Expression Recognition Using Kernel Canonical Correlation Analysis (KCCA)”，IEEE Trans. Neural Networks，vol. 17， no. 1，pp. 233-238 (2006).

# 第9章. 寺社仏閣における不審者検知のための行動分類

## 9.1. まえがき

本章では，異常検知のための行動分類手法について検討し，その性能検証を行う．監視カメラの設置環境は多く想定されるが，我々は寺社・仏閣における不審者検知のための行動分類に焦点を当てる. 文化遺産として指定される寺社仏閣は, 建造物そのものに限らず, 仏像や美術工芸品などの犯罪から守りたい資産を多く保有している. しかしそれらは普段, 参拝者に公開されるので，日中においては参拝客に紛れた窃盗や破損，また夜間においては放火といった被害を受ける場合が散見される．ゆえにこのような犯罪の危険から文化遺産を守る防犯カメラ技術の開発は重要な課題である **[1]**．ここではハードとソフト両面からの解決策が考えられる．ハード面では，ハイビジョン対応の PTZ カメラ（ズームやパンの機能があるカメラ）や, 赤外線高感度カメラなどの改良が成されている. ソフト面では，従来の防犯カメラに対して知性を持たせる，即ち不審者を高精度に検知して知らせる機能を充実させることによって人間警備員と同等の能力を持たせることができると，その有用性は非常に高まることが期待される．ゆえに近年では不審者の異常行動を検知する技術の研究が多く行われている．例えば，鷲見らはエレベータに特化した異常検知に取り組んでいる**[2]**. 平常時の映像からオプティカルフローによる向きや大きさのばらつきを求めて学習モデルを作成する．そして新規の映像から学習時と同様に特徴抽出を行い，閾値処理により暴れ動作の検知を行っている．また村井らはエスカレータでの異常検知に取り組んでいる**[3]**. 具体的には局所的な視覚と動き情報の時間的な変化を捉える Space-Time Patch **[4]** を用い，エスカレータといった動的背景を除去して人物領域を抽出する．その人物領域内より動きベクトルの定常度を閾値から判定し，転倒といった異常行動を検知している．これらの限られた空間での転倒・暴れといった激しい異常行動の検知の研究は実用段階にあると言える．しかし我々が取り組む寺社仏閣での異常検知は多くの問題が存在する．具体的に，異常行動は腕だけの動きで表現される“建造物への傷つけ”や持続的でゆっくりした“賽銭箱を覗く”など正常行動と見分けが付かない程複雑であり，撮影環境も屋外の広い境内であるなど分類を難しくする要因が多く存在する．ゆえに寺社仏閣での行動の複雑

さを考慮すると，従来の正常・異常の 2 クラス分類ではなく，各行動のカテゴリーに分類する手法が望ましい．そこで本研究では不審者検知のための行動分類を目的として寺社仏閣において想定される幾つかの行動を予め定義し，それら行動のカテゴリー分類に取り組む．

その際，我々は動きと視覚情報を同時に表現できる時空間特徴量を用いる．この特徴量は近年の研究で広く用いられているため，次にその関連研究について述べる．代表的な手法としては Cuboid と呼ばれる立方体を映像内から抽出し，それを特徴化する手法が挙げられる．Laptev はハリスオペレーターを 3 次元へと拡張することで時空間のコーナーである Space-Time Interest Point(STIP)を検出し，Histgram of Oriented Gradient(HoG)や Histogram of Oriented Flow(HoF)を Cuboid の記述子として用いる手法を提案している**[5]**．また Dollar らは空間軸に 2 次元ガウシアンフィルタ，時間軸に 1 次元の Gabor フィルタを適用することで Cuboid の中心となる特徴点の検出を行っている **[6]**.

Dollar らの手法は周期的な動きを特徴量化することで行動認識において高い有用性を示し，広く応用されている **[7,8]**．しかし寺社仏閣という環境を想定した場合，広大な領域による人物のスケール変化や個人毎の行動の速さの違いといった課題に Dollar らの手法では対応できない．ゆえに本稿ではそれらの課題を解決する特徴量の提案を行う．

また“賽銭箱を覗く”といった行動は非常に複雑であり，Cuboid を用いる標準的な Bag-of Feater(BoF)**[9]**と SVM（付録 A 参照）による手法での分類は難しいと言える．そこで行動を単純で短い行動（行動素）の列へと分解し，連続する行動素の組み合わせと順序により行動を分類する手法を提案する．例えば，“賽銭箱を覗く”行動は“腰を曲げる”，“賽銭箱に手を添える”，“覗き込む”などの連続する行動素へと分解することが望ましい．

本章では，不審者検知のための行動分類を目的とし，上記で述べた寺社仏閣における 2 つの課題（①撮影場所の広大さに起因し，人物のスケール変化が生じる．また個人毎の行動の速さの違いが分類を難しくする．②“賽銭箱を覗く”といった寺社特有の複雑な行動を分類する必要がある．）の解決に取り組む．9.2 節にて課題①の解決のため，Dollar らの手法を時空間のスケール変動にロバストに発展させた手法について述べる．更に課題②を解決するため，行動を行動素に分解し，その組み合わせと順序から行動分類を行う手法について述べる．9.3 節では実験とその結果について示す．まず既存の映像データセットである KTH データセット**[10]**を用いて，提案する特徴量抽出アルゴリズムについての性能検証を行う．次に独自に採取した寺社仏閣におけるデータセットについて説明し，そのデータセットを用いて寺社仏閣に特化した行動分類システムについての性能評価実験を行う．そして 9.4 節で本章をまとめる．

## 9.2. 提案手法

本節では提案する行動分類システムについて述べる．システムのフレームワークを**図 9.1** に示す．我々は局所特徴量の位置情報を省くことで，それらを一つのヒストグラムとして

簡略化できる Bag-of-Feature (BoF)**[9]**を特徴量として用いる．また BoF は位置情報を持たない局所特徴の集合であるので，人物の見える角度や姿勢の変化の問題に対しても一定の汎化性を期待できる．

訓練時では，始めに複数の行動を含んだ映像から局所特徴量を抽出する．この局所特徴量は提案する時空間でのスケール変動にロバストな特徴量であり，その詳細については 9.2.1, 9.2.2 項にて述べる．次に局所特徴量の集合を k-means 法により量子化し，Visual Word の集合である Codebook を構築する．Codebook からベクトルで表現される局所特徴量とユークリッド距離が最小の Visual Word を求め，それに投票する．この投票結果が出現頻度のヒストグラムである BoF となる．

ここで映像から BoF を抽出する場合について述べる．まず映像内のある一定区間を窓として定義する．このとき窓内には連続する画像列が存在する．本章の実験では窓サイズを 40 と設定している．この時，映像の時間軸方向に窓を 1 フレームずつスライドさせ，複数の画像列を習得する．最後に画像列から複数の局所特徴量を算出し，それを BoF へと集約する．

次に特徴量を行動素へ変換するために BoF を学習データとして用い，教師なしの確率的クラスタリングである Probablistic Latent Semantic Analysis (pLSA）**[11]**を適用する．これより映像内の行動は行動素を意味する連続する記号列へと変換される．そして PrefixSpan 法**[12]**を用いて，出現頻度の高い記号列を抽出し，抽出された記号列を使ってトライ木**[13]**という木構造へと拡張する．エッジには記号，終端ノードには各行動の出現頻度のスコアが格納される．次にテスト時では訓練時と同様に特徴量の抽出を行い，pLSA を用いて行動素を意味する連続する記号列へ変換して，記号列に従いトライ木をルートノードから遷移させていく．遷移の過程で終端記号に格納された行動の出現頻度のスコアが最大の行動をその状態における行動であると推測する．

次に提案する時空間変動に対しロバストな局所特徴量の詳細について述べる．大きく分けて 2 つのプロセス，即ち「時空間の特徴点検出」と「記述子の算出」により構成される．「時空間の特徴点検出」では，人物の手振り，身振りといった動きを捉えることができる Gabor フィルタの周期特性を利用した Dollar らの手法**[6]**を発展させる．Dollar らの従来法は Cuboid のサイズが一定であり，時空間のスケール変動に対して脆弱な課題が存在する．ゆえにカメラからの距離による人物のサイズの変化や体の部位における動作領域のサイズの変化，個体差等による行動の速さの変化などに対応できない．そこでマルチスケール化した時空間特徴点の検出法を提案する．また「記述子の算出」，つまり特徴点を中心とした Cuboid 内部の特徴ベクトルの算出においては 3 次元の勾配特徴を用いる．

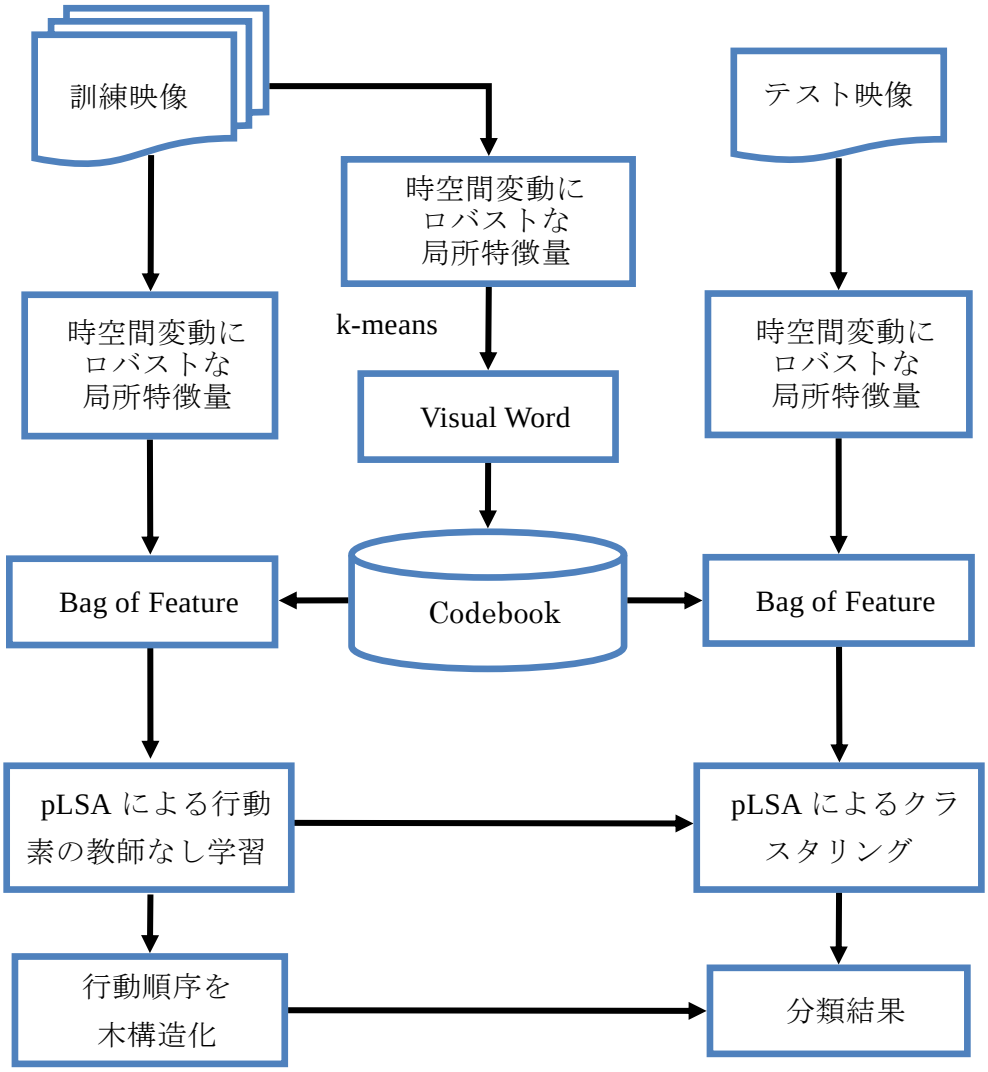



**図 9.1**：行動分類システムのフレームワーク

### 9.2.1. Dollar らによる特徴点検出手法

Dollar らにより提案された従来の特徴点検出手法はレスポンス関数$R$に従い時空間における極大値を特徴点とする．レスポンス関数$R$を以下に示す：

$$R = (I * g * h_{ev})^2 + (I * g * h_{od})^2 \tag{9.1}$$

ここで$I(x, y, t)$は映像中の時間$t$におけるフレーム$I(x, y)$であり，$g(x, y; \sigma)$は 2 次元のガウシアンフィルタで空間軸に適用される．また$h_{ev}$，$h_{od}$は 1 次元の Gabor フィルタで時間軸に適用され，以下の式で定義される：

$$h_{ev} = -cos(2\pi t\omega)\, e^{-t^2/\tau^2}$$
$$h_{od} = -sin(2\pi t\omega)\, e^{-t^2/\tau^2} \quad (9.2)$$

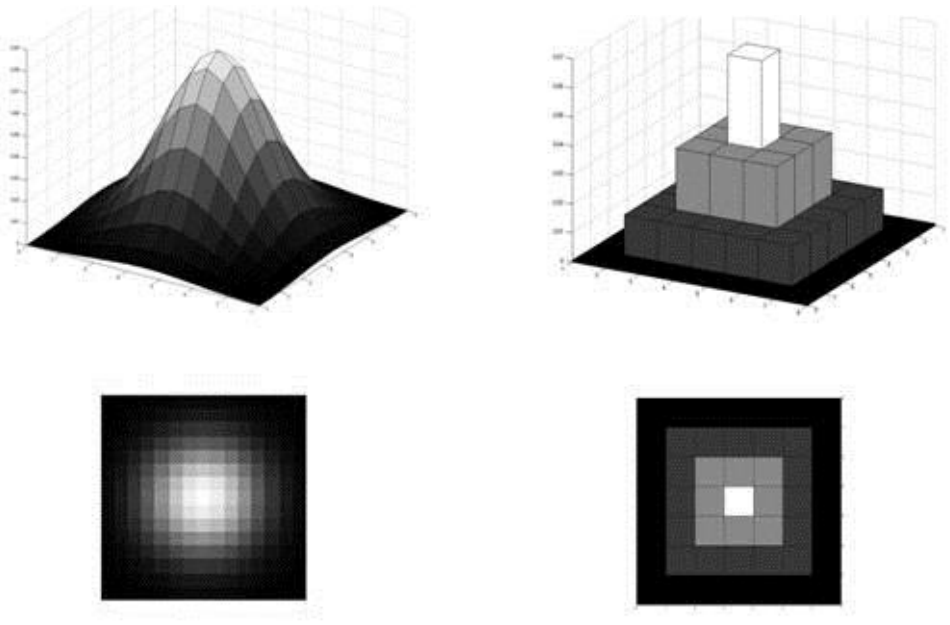
ここで$\sigma$は空間軸のスケールパラメータ，$\tau$は時間軸のスケールパラメータであり予め任意に設定する．また Dollar らの文献**[6]**では$\omega = 4/\tau$と設定されている．

### 9.2.2. 時空間のスケール変動にロバストな特徴点検出

提案手法は複数スケールの空間軸における平滑化，時間軸での 1D Gabor フィルタの適用により，スケール変動にロバストな手法へと発展させた．それは(1)「Box filter による平滑化」，(2)「マルチスケールの平滑化画像作成」，(3)「Gabor フィルタの適用と極大値の検出」の 3 つのプロセスで構成される．ここで空間軸のスケール数を$Octave_\sigma$，時間軸のスケール数を$Octave_\tau$とする．また空間軸でのスケールパラメータの初期値を$\sigma_0$，時間軸での初期値を$\tau_0$と定義する．これら 3 つのプロセスの詳細について述べる．

#### (1) Box filter による平滑化

従来の 2D ガウシアンフィルタによる平滑化は画像の全画素に対して畳み込み演算を実行するため計算コストが高い．そこで計算コスト削減のため，**図 9.2**(b)に示すような Box filter を使いガウシアンフィルタを近似する．



(a) 2D ガウシアンフィルタ　　(b) Box filter による近似

**図 9.2**：Box filter によるガウシアンフィルタの近似

本稿では Box filter を 4 段の Box から構成させるが，この Box の数を減らすことで計算コストを更に削減できる. ここで**図 9.2**(b)上における 4 段の Box の値は**図 9.2**(a)上のガウシアンフィルタの合計値と 4 つの Box の合計値が等しいという条件で求めることができる．このとき 4 つの Box の値を面積が小さい方から$h_{box1}, h_{box2}, h_{box3}, h_{box4}$と定義する．次にインテグラル画像を用いて 4 つの Box 内の積分値を算出する. ここでそれらの積分値を Box の面積が小さい方から$V_{box1}, V_{box2}, V_{box3}, V_{box4}$とすると，最終的な平滑化後の値は$h_{box1} \times V_{box1} + h_{box2} \times V_{box2} + h_{box3} \times V_{box3} + h_{box4} \times V_{box4}$より求めることができる．また本稿の実験においては 4 つの Box のサイズをスケールパラメータが$\sigma_0$のとき, 大きい方から 8×8, 6×6, 4×4，2×2 pix.と設定する.

**(2) マルチスケールの平滑化画像作成**

空間軸におけるスケーリングには画像をダウンサンプリングする方法が多く利用されている**[14, 15]**. しかし$Octave_\sigma$回の画像のダウンサンプリングは計算コストが高い．そこで**図 9.3** に示すように増加率$k_\sigma$ $(k_\sigma = 1, \ldots, Octave_\sigma)$に従い，予め$Octave_\sigma$個の異なるサイズの Box filter を作成しておき，それらを画像に適用する．これより画像のダウンサンプリングを行う必要がなく 1 回のインテグラル画像の計算のみで済み，計算コストを抑えることができる.

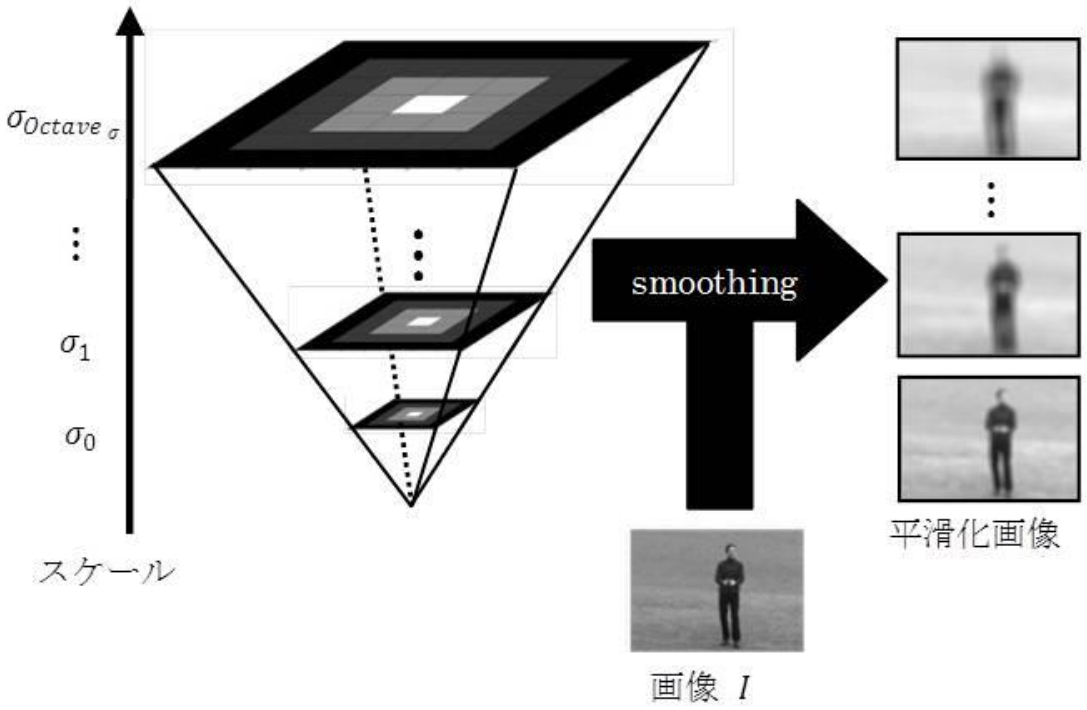
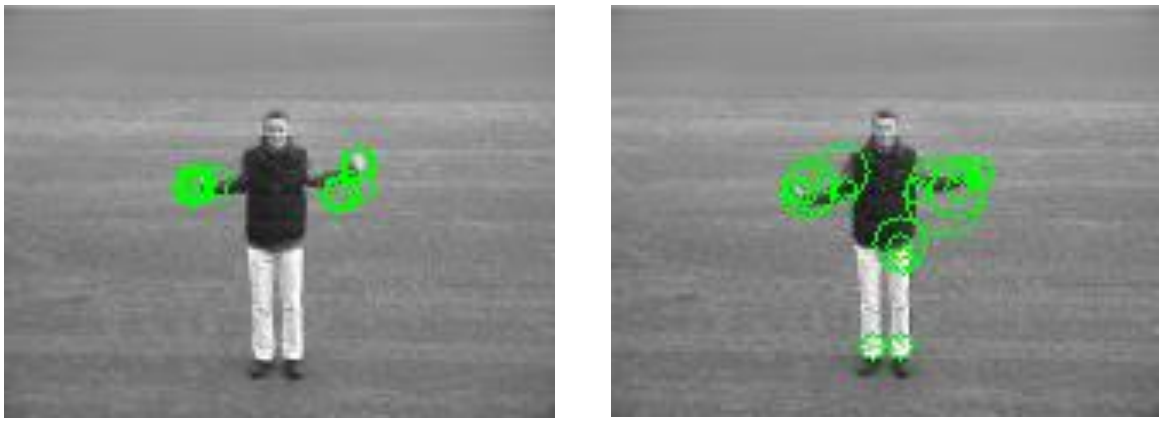



**図 9.3**：マルチスケールの Box filter による画像の平滑化

**(3) Gabor フィルタの適用と極大値の検出**

時間軸に連続する平滑化画像に対して Gabor フィルタを適用し，その周期性から人物の動きを捉える．従来は Gabor フィルタを畳み込む画像の枚数が一定であったため，動きの速さの変化に上手く適応できず，画像枚数に過不足が生じ最適な特徴を抽出できない問題があった．そこで時間軸においてマルチスケールの Gabor フィルタを適用する．つまりスケール数を$Octave_\tau$個に拡張し，スケールパラメータ $\tau$と画像枚数は共に増加率$2^{(k_\tau-1)/2}$ $(k_\tau = 1, \dots, Octave_\tau)$に従い増加させる．ただし増加率は実験的に決定した．これより$Octave_\sigma \times Octave_\tau$の時空間でのスケールパターンが構成され，全てのスケールパターンから極大値の座標を探索することになる．また仮に同座標において異なるスケールパターンから極大値が検知された場合，最大値を有するパターンを採用する．最終的には，極大値における時空間のスケールパラメータをそれぞれ$\sigma_d$，$\tau_d$と定義すると，極大値の座標を中心とした Cuboid が構成され，その Cuboid のサイズをパラメータ$\sigma_d$，$\tau_d$に従わせることで時空間にロバストな特徴点を検出する．ここで**図 9.4** に提案手法により特徴点を検出した結果画像を示す．**図 9.4** における円の中心が検出された特徴点位置であり，円のサイズは Cuboid の空間軸方向のスケールサイズを表現している．提案手法（**図 9.4**(b)）は従来法（図 9.4(a)）と比較し，サイズの異なる Cuboid がより多く検出されていることを確認できる．



(a) Dollar らの従来法　　(b) 提案手法

**図 9.4**：handclapping の動きにおける特徴点検出結果

### 9.2.3. 記述子の算出

Cuboid 内の記述子には 3 次元の勾配特徴を用い，それは 3 つのステップで算出できる．第 1 ステップでは，Cuboid 内の全画素から極座標表現を用いて 3 次元の勾配強度$m_{3D}$，勾配方向$\theta, \phi$を算出する．これらは式(9.3), (9.4)より求めることができる．ここで$\theta$は範囲$(-\pi, \pi)$

での 2 次元勾配方向である．$\phi$は 2 次元勾配方向から時間軸に対しての角度でその範囲は$\left(-\frac{2}{\pi},\frac{2}{\pi}\right)$である．

$$m_{3D} = \sqrt{L_x^2 + L_y^2 + L_t^2} \tag{9.3}$$

$$\phi = tan^{-1}\left(L_t \Big/ \sqrt{L_x^2 + L_y^2}\right) \tag{9.4}$$

$$\theta = tan^{-1}\left(L_y/\sqrt{L_x}\right)$$

ここで, $L_x$, $L_y$, $L_t$は, $\sigma_d$での平滑化画像を$L_{\sigma_d}$とすると, $L_x = L_{\sigma_d}(x+1,y,t) - L_{\sigma_d}(x-1,y,t)$, $L_y = L_{\sigma_d}(x,y+1,t) - L_{\sigma_d}(x,y-1,t)$, $L_t = L_{\sigma_d}(x,y,t+1) - L_{\sigma_d}(x,y,t-1)$として算出できる．第 2 ステップでは，勾配方向ヒストグラムを計算する．それは Cuboid を複数のブロックに分割し，それぞれのブロック毎に算出される．ブロック内の全画素に対して，$\theta$に従い 8 ビンの勾配方向ヒストグラムのいずれかのビンに勾配強度$m_{3D}$の値を割り当てる. また$\phi$に関してはビン数を 4 とし，$\theta$のときと同様に値を割り当てる．これより 2 つの勾配方向ヒストグラムが作成される．最終ステップではそれらを一つのベクトルとして結合する．そしてその特徴ベクトルをその合計値で割り正規化する．ここで提案手法ではブロックの分割法をスケール変化や動きに十分に対応させるため**図 9.5** に示すような 3 つの分割パターンを設ける．それぞれから勾配方向ヒストグラムを計算し，最終的に 3 つの特徴ベクトルを 1 つに結合する. 次元数としてはビン数 : 4+8=12, ブロック数 : 1+9+9=19 より 12×19=228 次元である．

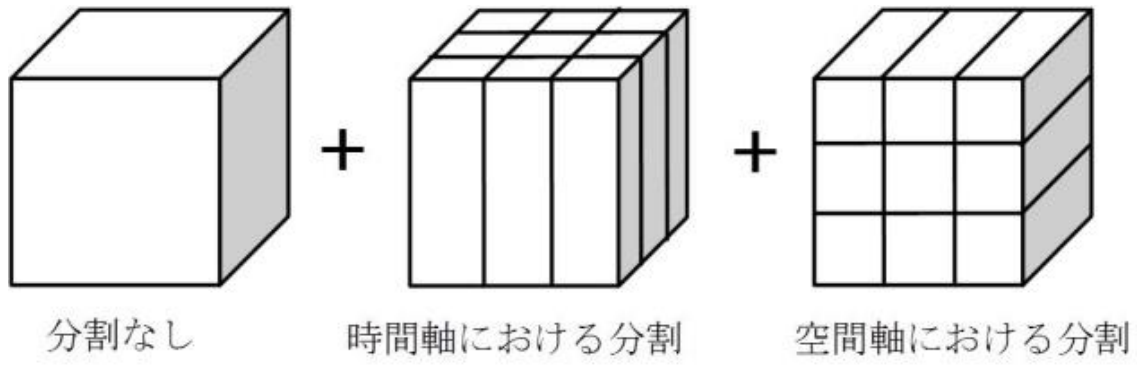



**図 9.5**：3 種類の Cuboid の分割パターン

### 9.2.4. pLSA を用いた行動素の抽出

教師なしの確率的クラスタリングである pLSA**[11]**の適用により映像内の行動を行動素を意味する連続する記号列へと変換する.

ここで Visual Word の集合である BoF の各ビンを Word として定義する．pLSA を適用することで，映像内で発生した Word と潜在トピックに対応する行動素の共起により，映像の各行動素への帰属確率を求めることができる．これより最も高い帰属確率を示す行動素（潜在トピック）を記号として表現する.

潜在トピック数は予め設定する必要があり, 潜在トピックを$z \in Z$とすると, 映像$d \in D$における Word $w \in W$の同時確率は：

$$P(d,w) = \sum_{z \in Z} P(z)P(d|z)\,P(w|z) \tag{9.5}$$

として表現される. この時 pLSA モデルのグラフィカルモデルは**図 9.6** として表現できる.

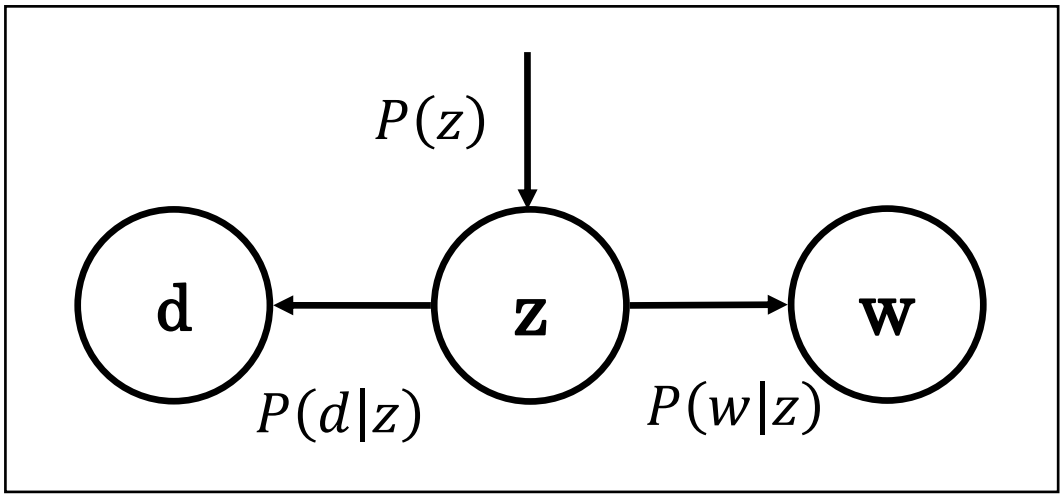



**図 9.6**：pLSA モデルのグラフィカルモデル

ここで，映像$d$における Word $w$の発生確率は，潜在トピック$z$を用いると，以下の式として表現できる：

$$P(w|d) = \sum_{z \in Z} P(w|z)\,P(z|d) \tag{9.6}$$

$P(w|z)$は潜在トピック$z$毎の Word の分布であり, 映像は潜在トピックの混成$P(z|d)$として, モデル化できる．ここで映像$d$における Word $w$の出現回数，即ち BoF を$n(d,w)$とすると全学習映像においての対数尤度は以下の式で与えられる：

$$L = \sum_{d \in D} \sum_{w \in W} n(d,w) \log P(z|d) \tag{9.7}$$

そして式(9.7)の尤度を最大にする$P(z)$，$P(d|z)$，$P(w|z)$を Expectation Maximization (EM)アルゴリズムで計算する．それは Expectation (E)ステップと Maximization (M)ステップから構成され，その反復により同時確率$P(d,w)$を最大化する．E ステップは以下の式より定式化される：

$$P(z|d,w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z' \in Z} P(z')P(d|z')\,P(w|z')} \tag{9.8}$$

これより潜在トピックにおける事後確率$P(z|d,w)$が計算される．次に M ステップよりパラメータの更新を行う．それは以下の式より定式化される：

$$P(w|z) \propto \sum_{d \in D} n(d,w)P(z|d,w) \tag{9.9}$$

$$P(d|z) \propto \sum_{w \in W} n(d,w)P(z|d,w) \tag{9.10}$$

$$P(z) \propto \sum_{d \in D} \sum_{w \in W} n(d,w)P(z|d,w) \tag{9.11}$$

ここで局所最適解に陥ることを防ぐためにアニーリングスケジュールを導入する．E ステップの右辺全体を$\beta$乗$(0 < \beta \leq 1)$し，反復が進むにつれて$\beta$の値を変化させることで，$\beta$が小さければ小さい程，生成される確率モデルの確率分布を平滑化することができる．

### 9.2.5. PrefixSpan による部分記号列の抽出とトライ木への拡張

学習映像の行動は，pLSA により行動素を意味する連続する記号列へと変換できる．次にシーケンシャルデータマイニングを用いて記号列内の集合から頻出する部分記号列を抽出する．ここでシーケンシャルデータマイニングとは，記号列内の出現頻度が最小サポート値以上となる部分記号列を列挙するタスクを指す．また部分記号列とは，順序を保持したまま記号列の任意の要素を抽出可能な記号列のことである．本研究ではシーケンシャルデータマイニングにおいて深さ優先探索で高い出現頻度の部分記号列を抽出する手法である PrefixSpan 法を用いる．PrefixSpan とは，出現する部分記号列に対して特定の記号列に続く接尾辞を取り出す操作である射影を再帰的に行い，出現する部分記号列とその頻度を習得する手法である．ここで**図 9.7** に PrefixSpan の操作手順を示す．

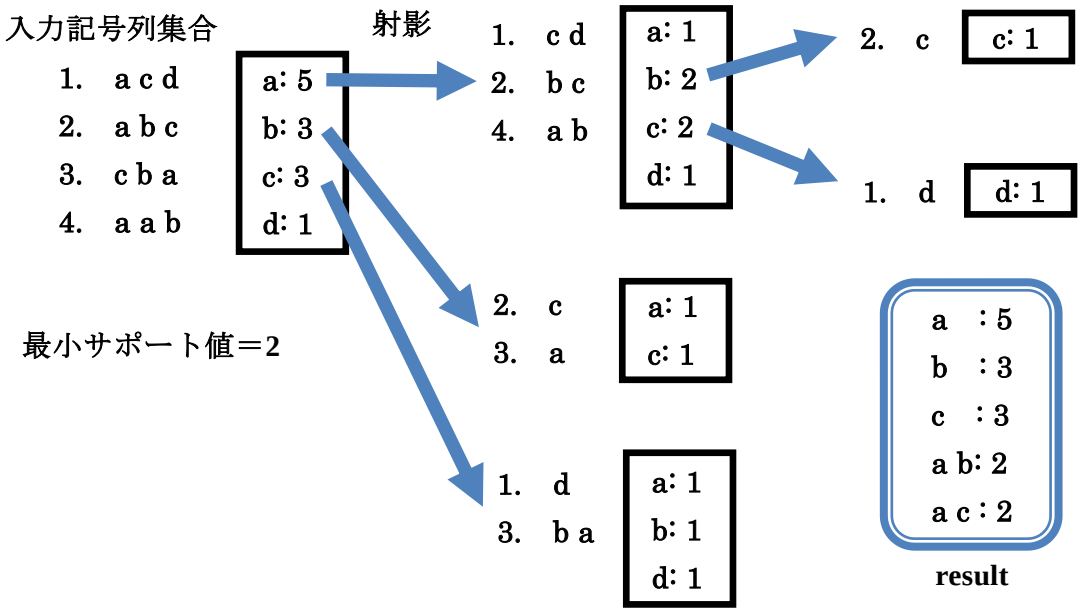



**図 9.7**：PrefixSpan の操作手順

**図 9.7**(左)の 4 つの入力記号列集合内に含まれている記号は a，b，c，d であり，各出現頻度は**図 9.7**(左)の黒枠内に記した数値 5，3，3，1 である．最小サポート値を 2 に設定した場合，2 以上の出現頻度は 5，3，3 であるので，それら出現頻度に対応する a，b，c の記号に続く接尾辞を取り出す．この射影操作を終端記号まで操作する．これより[部分記号列：出現頻度]とするとき，[a：5]，[b：3]，[c：3]，[a b：2]，[a c：2]の 5 つの部分記号列と各出現頻度が抽出できる．

次に抽出された多数の部分記号列をトライ木という順序木構造へと拡張する．トライ木とは記号列や数列などの集合を多分木で表現するデータ構造で，各エッジには記号が付与されている．ルートノードからリーフノードに遷移させていき，対応する経路の存在を確認することで部分記号列を探索できる．終端記号には“#”という記号を付与してやることで，例えば記号列<a b c>内の記号 c の前方の記号列である<a b>の終端も表現できる．そしてリーフノードには出現頻度の値を格納する．**図 9.7** 内の部分記号列とその各出現頻度の結果をトライ木に拡張すると，**図 9.8** のように表現できる．

ここで本章における実験では入力記号列の連続する同記号は 1 つにまとめる．例えば，<c c d>という記号列は<c d>に予め変換される．また行動を分類するためにトライ木のリーフノードの出現頻度を行動毎に算出し，探索の過程で各行動の出現頻度の値が最大となる行動をその状態における行動であると推測する．

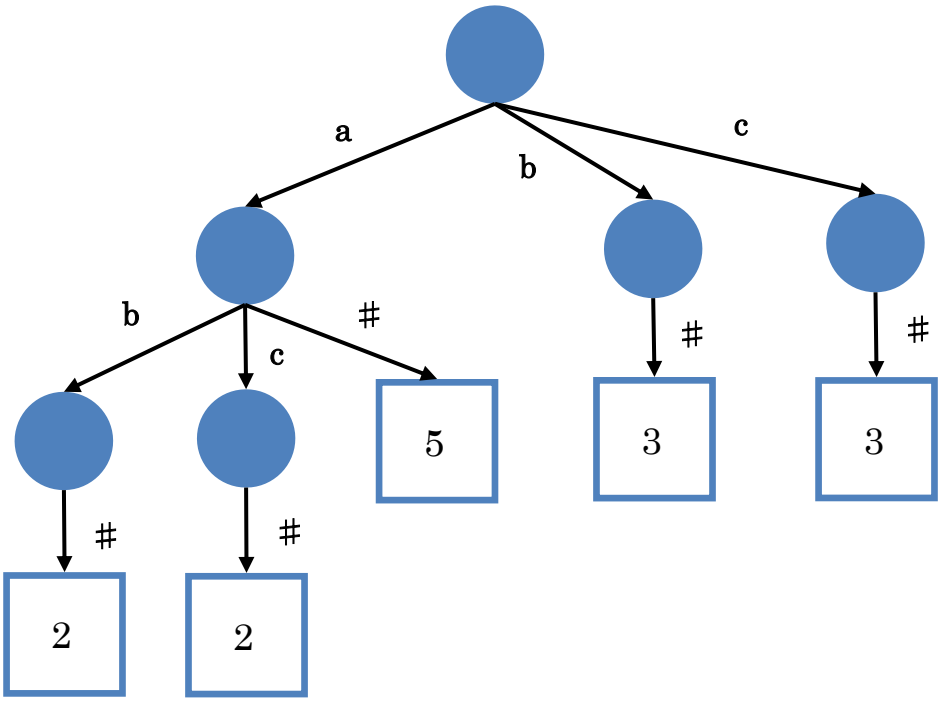



**図 9.8**：部分記号列のトライ木への拡張

## 9.3. 実験と考察

本節では，既存の映像データセットである KTH データセット**[10]**を用いて提案する特徴量抽出アルゴリズムについての性能検証を行う．更に独自に採取した寺社仏閣におけるデータセットについて説明し，そのデータセットを用いて寺社仏閣に特化した行動分類システムについての性能検証を行う．

### 9.3.1. KTH データセットを用いた行動分類

提案する特徴量はガウシアンフィルタに代えて簡易的な Box filter を導入することで従来法である Dollar らの Cuboid と比較し，特徴量自体の性能の低下が考えられる．そこで本項では既存の映像データセットを使い，提案特徴量の性能を検証する．

既存の映像データセットとしては KTH データセットを用いる．**図 9.9** に示すように walking，running，jogging，boxing，hand clapping，hand waving の 6 つの行動で構成され，それぞれの行動は 4 つの異なる背景にて 25 の人物により実演されている．また映像の長さは平均 4 秒であり，画像サイズは 160×120 pix.である．実験では CPU：2.67GHz，実効メモリ：3GBytes のスペックを搭載したプロセッサを使用する．

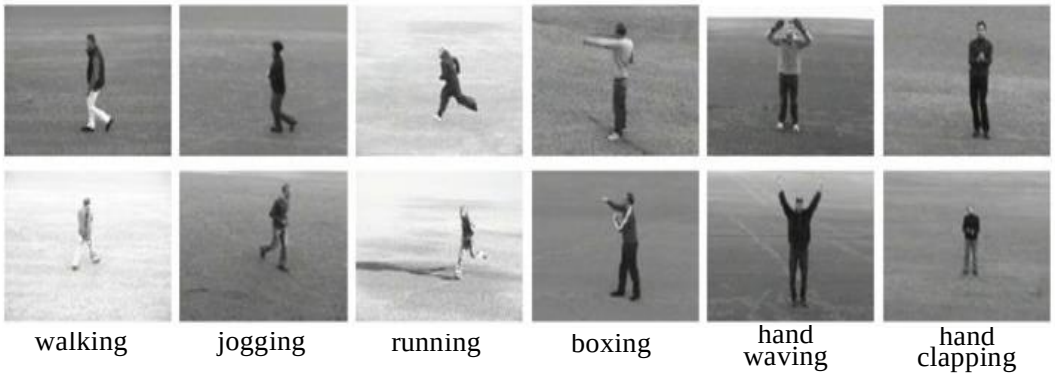



**図 9.9**：KTH データセットのサンプル画像

評価項目として従来手法と提案手法の性能比較実験を行う．比較対象は Dollar らの特徴点検出法**[6]**とする．ただし記述子としては 3 次元勾配特徴を採用する．また従来手法・提案手法互いに Bag-of-Feature を特徴量とし，Codebook の作成が必要であるので Visual Word 数は 1,000 に設定する．また KTH データセットは 5 人単位に区切ることで 5 fold cross validation より学習・テストを行う．識別器としては SVM（付録 A 参照）を採用し，カーネルとしては RBF を用いる. 本実験における特徴点検出での各パラメータは空間軸のスケール数$Octave_{\sigma}$=3，時間軸のスケール数$Octave_{\tau}$=3 とする．また空間軸のスケールパラメータの初期値$\sigma_0$＝1.4，時間軸の初期値$\tau_0$＝1.8 と設定する．各パラメータは実験的に決定している．

ここで**図 9.10** において従来手法との分類率の比較実験の結果を示す．また**表 9.1** は提案手法の各行動の分類率を示した Confusion Matrix である．Confusion Matrix は左端の列が対象行動を示しており，表の対角成分は各対象行動が正解する確率(%)である．また対角成分以外は誤分類する確率であり, 対象行動を他の行動に分類する確率は同行へと記される．**図 9.10** に示すように KTH データセットを使う実験での提案手法の分類率は 80.1%であり従来手法の 77.1%より 3%程の性能向上を確認できる．これより提案手法の特徴量はガウシアンフィルタに代えて簡易的な Box filter を導入しているが, 性能は低下しないことを実証でき，時空間のスケール変動に対しての頑強性を高めたことで若干の性能の向上が見られる．そしてフル HD の映像内から切り出した 56×100pix.のサイズに提案手法を適用する場合，一人物における処理速度は検出された特徴点数に依存するが，60−100fps 程であり，従来手法と比較して処理速度は殆ど同じであることを確認できる．**表 9.1** からは boxing, waving，clapping といった手の動きから構成される行動は提案手法の分類率が 80%以上であり，その有効性を示せている．しかし問題として jogging と running は他と比較しても分類率が低く，人物の動きの速さの違いを上手く捉えられていない．これは Cuboid を時間軸についてスケールの正規化を行ったことが影響していると考えられる．

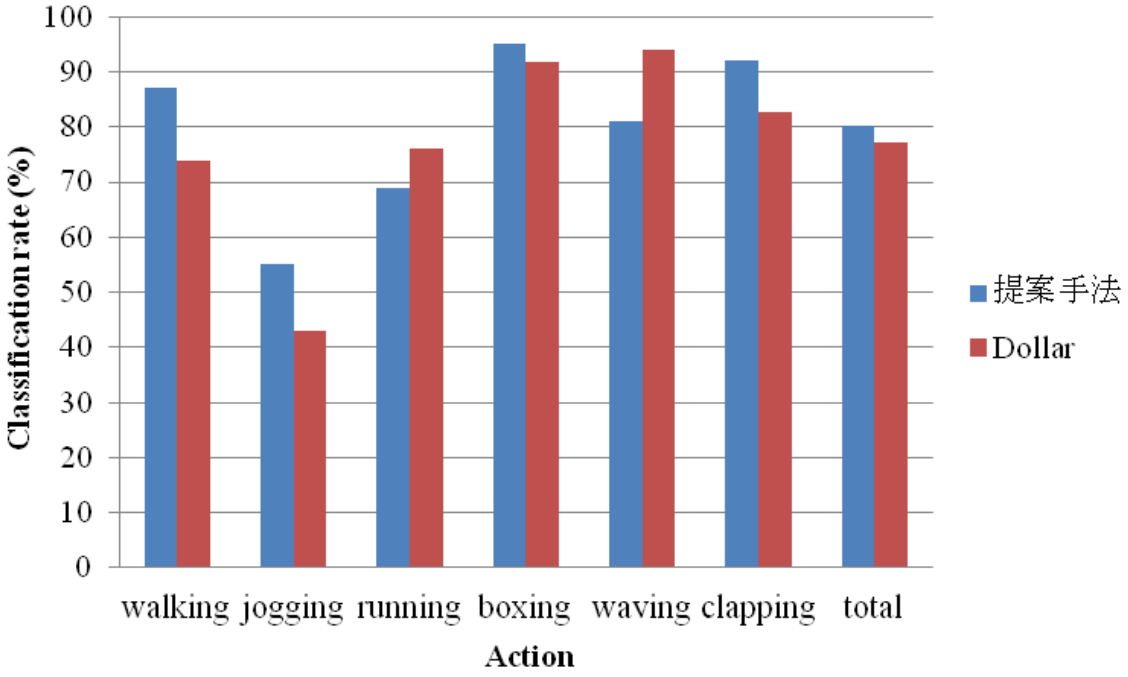



**図 9.10**：KTH データセットでの従来手法との分類率の比較結果

**表 9.1**：KTH データセットを用いた実験における提案手法の各行動の分類率（%）

| | **Walking** | **Jogging** | **Running** | **Boxing** | **Waving** | **Clapping** |
|---|---|---|---|---|---|---|
| **Walking** | **87** | 10 | 1 | 1 | 0 | 1 |
| **Jogging** | 27 | **55** | 18 | 0 | 0 | 0 |
| **Running** | 9 | 22 | **69** | 0 | 0 | 0 |
| **Boxing** | 2 | 0 | 0 | **95.1** | 2 | 1 |
| **Waving** | 2 | 1 | 1 | 8 | **81** | 7 |
| **Clapping** | 0 | 0 | 0 | 5.1 | 3 | **91.9** |

**(%)**

### 9.3.2. 寺社仏閣での独自データセットを用いた行動分類

“賽銭箱を覗く”といった寺社特有の複雑な行動を分類するためには，行動を行動素に分解し，行動素の組み合わせと順序の情報を基に，行動を分類することが望ましい．ゆえに本項では，**図 9.1** に記した提案法である寺社仏閣に特化した行動分類システムを寺社仏閣でのデータセットに対して適用し，その性能検証を行う．

その性能を検証するためには，寺社仏閣に特化した行動を再現したデータセットを作成する必要がある．そこで我々は国宝に指定されている仁和寺の金堂付近に設置された防犯カメラにおいて約 4 カ月間蓄積した映像を観察し，寺社での典型的な行動を正常・異常それぞれ 3 つずつ定義する．正常行動は“歩く”，“拝む”，“写真撮影”であり，異常行動は“しゃがむ”，“賽銭箱を覗く”，“キョロキョロ”である．定義した行動のサンプル画像を**図 9.11** に示す. また本実験における学習とテストに用いる各行動の映像数を**表 9.2** に示す.

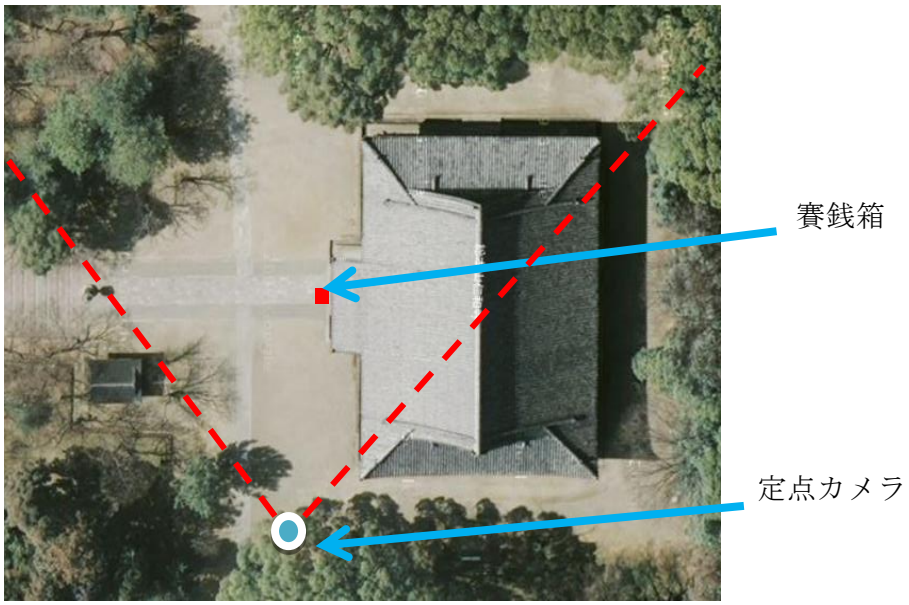
**表 9.2**：学習・テストに用いる各行動の映像数

| 行動種類 | 学習映像数 | テスト映像数 |
|---|---|---|
| **写真撮影** | 23 | 19 |
| **拝む** | 29 | 20 |
| **歩く** | 40 | 30 |
| **しゃがむ** | 27 | 19 |
| **賽銭箱を覗く** | 29 | 18 |
| **キョロキョロ** | 32 | 19 |
| 合計 | 180 | 125 |



**図 9.11**：定義した 6 つの行動のサンプル画像

**図 9.11** より，このデータセットは屋外環境で撮影し，天候変化による照明環境や水溜まりにおける太陽光の反射など分類を難しくする条件を多く含んでいることが確認できる．また各映像は，フル HD 映像内の一部の時間での 1 人物の領域を切り出して作成し，画像サイズと時間の長さは映像毎に異なる．“歩く”，“拝む”といった行動はデータの採取が容易であるが，“賽銭箱を覗く”や“しゃがむ”等の異常行動は採取できる数は限られている．ゆえに防犯カメラから撮影された映像を基にその行動を我々研究チームが実演した映像を一部含んでいる．また防犯カメラの設置位置を**図 9.12**(a)に示す．カメラと賽銭箱の距離は約 30m であり，賽銭箱前における映像内の人物サイズは約 30×60 pix.である．




(a) カメラ位置

(b) 撮影環境

**図 9.12**：カメラ位置と撮影環境

9.3.1 項では Dollar らの手法を時空間に対して頑強性を高めた特徴量抽出アルゴリズムの性能を評価し，従来法と比較して性能の向上を確認できた．そこで本項においては，提案する寺社仏閣に特化した行動分類システムの性能を 9.3.1 項で評価した特徴量抽出アルゴリズム〈以降 Multi-Scale Cuboid（MS-Cuboid）と記述する〉と比較する．この時 MS-Cuboid の識別器として SVM を用いる．2 つ目の評価項目として，提案手法の 6 行動の分類率を Confusion Matrix を使い評価する．ここで本実験では Visual Word 数を 850，pLSA の潜在トピック数は 22，PrefixSpan の最小サポート距離は 2 に設定する．これらパラメータは実験的に決定している．そして特徴点検出における各パラメータは KTH データセットを用いた実験と同様の値とする．**図 9.13** では寺社仏閣でのデータセットを使い，6 行動についての従来法との分類率の比較実験の結果を示す．また提案手法の 6 行動の分類率を示した Confusion Matrix を**表 9.3** に示す．

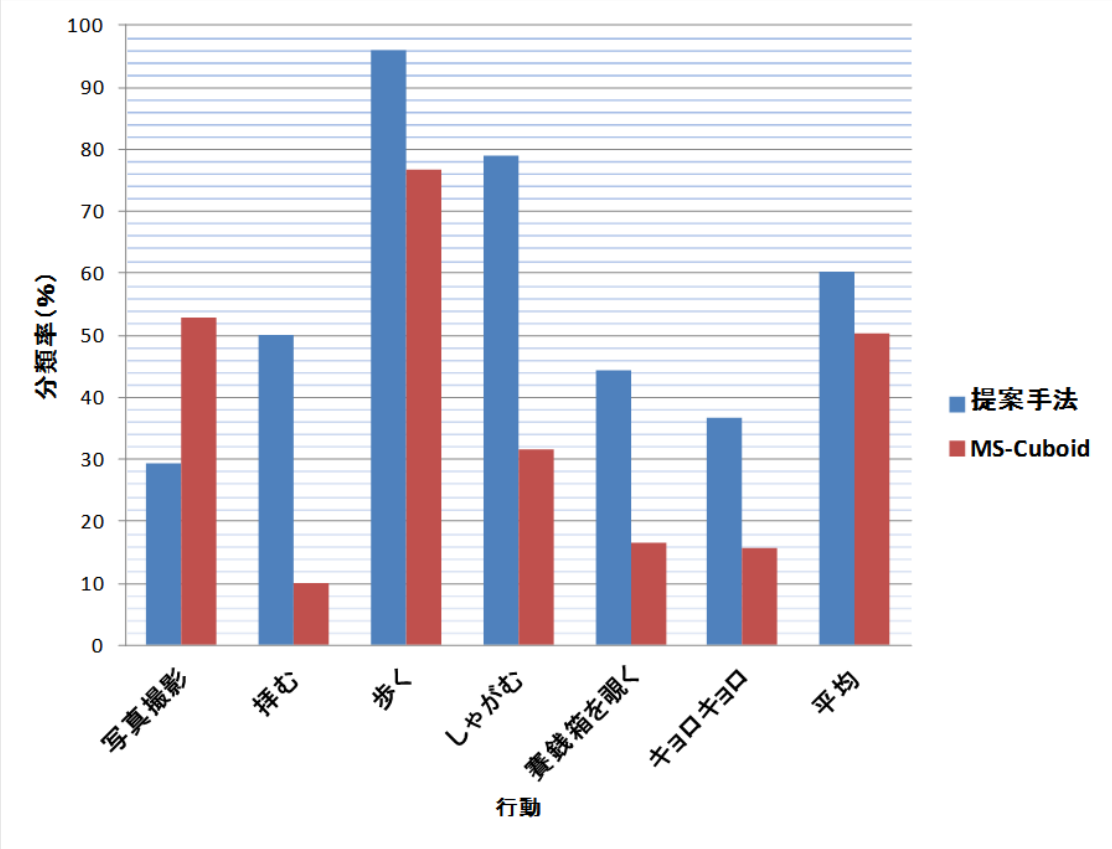



**図 9.13**：寺社仏閣における独自データセットでの MS-Cuboid との分類率の比較結果

**表 9.3**：寺社仏閣におけるデータセットを用いた実験における提案手法の各行動の分類率

| | 写真撮影 | 拝む | 歩く | しゃがむ | 賽銭箱を覗く | キョロキョロ |
|---|---|---|---|---|---|---|
| **写真撮影** | **29.4** | 17.6 | 0 | 0 | 35.2 | 17.6 |
| **拝む** | 40 | **50** | 5 | 0 | 5 | 0 |
| **歩く** | 0 | 0 | **96** | 0 | 4 | 0 |
| **しゃがむ** | 0 | 5 | 0 | **78.9** | 15.7 | 0 |
| **賽銭箱を覗く** | 22.2 | 5 | 0 | 16.6 | **44.4** | 11 |
| **キョロキョロ** | 36.8 | 21 | 0 | 0 | 5.3 | **36.8** |

**(%)**

**図 9.13**，**表 9.3** の実験結果から提案手法の平均分類率は 60.2%であり，MS-Cuboid の 50.4%と比較して 10%程の性能向上を確認できる．また 3 つの異常行動については全て MS-Cuboid よりも分類率が高い．特に最も複雑な“賽銭箱を覗く”行動は 30%程性能の向上を確認できる．“賽銭箱を覗く”行動は体の横移動や頭を下げる動作等の多くの状態変化を含むことから他の行動より分類が難しいが，提案手法は行動を行動素に分解し，行動素の組み合わせと順序に注目することで複雑な行動を従来よりも正確に分類できるようになったと言える．しかし平均分類率は 60.2%であり，未だ実用段階には至っていない．

また正常行動である“写真撮影”は手の動きのみから判断しなければならず，特徴点が十分に検出できずに分類精度が低下している．これは写真を撮り始めると静止状態になってしまい，時間軸方向の変位が失われ，時間軸の Gabor フィルタによる特徴が抽出できないことが原因であると考えられる．同様に“拝む”と“キョロキョロ”は静止状態になることが原因で特徴点を十分に検出できず，誤って“写真撮影”として分類する傾向にある．これらの事例より本章の特徴点検出アルゴリズムの課題は静止した人物の特徴点検出をできないことであり，静止状態においても人物の特徴情報を抽出できるアルゴリズムへの発展は今後の課題であると言える．

また学習映像の数は十分とは言えず，データに依存した結果が得られていると考えられ，今後は実環境映像の更なる採取が要求される．そして異常行動として想定される“建造物

への傷つけ”行為など分類が難しい行動の追加も今後必要である．しかし異常行動の映像については，実際の不審者映像の採取は困難であるので，警察関係者など有識者の見解を参考に映像を作成することで，より迫真性の高い映像を採取できると考えられる．

## 9.4. まとめ

本章では不審者検知のための行動分類を目的とし，時空間のスケール変動にロバストな独自の局所特徴量を用いて行動素の組み合わせと順序から行動を分類する手法である寺社仏閣に特化した行動分類システムを提案した．従来の時空間のスケール変動に脆弱な問題に対して，提案する局所特徴量は空間軸にBox filterで近似したガウシアンフィルタ，時間軸にGaborフィルタをそれぞれマルチスケールで適用することでロバスト性の向上が期待できる．実験では既存のKTHデータセットを用い，独自特徴量が従来手法より3%程高い80%以上の分類率を示したことで，ガウシアンフィルタに代えて簡易的なBox filterを導入しても性能は低下しないことを実証できた．また寺社に特化したシステムとするために，提案手法をMS-Cuboidに加え，行動素の組み合わせと順序から行動を分類する手法へと発展させた．寺社の映像を用いた実験では，3つの異常行動を含んだ6行動を定義し，MS-Cuboidとの分類率の比較検証を行った．提案法の分類率は60.2%であり，従来の50.4%より10%程度の性能の向上を確認できた．しかし未だ実用段階には至っていないと言え，今後は静止状態でも行動の特徴を算出できるアルゴリズムへの発展と実環境映像データの更なる追加が必要である．

## 参考文献


[1] 谷口仁士，山内寛紀：“寺社仏閣を対象とした歴史遺産の無人監視システム（特集『安心・安全』の計測技術）”，O plus E Optics・Electronics，vol. 33，no. 6，pp.589–594，(2011)

[2] 鷲見和彦，関真規人，塩崎秀樹：“安全と安心のための画像処理技術 ：3.画像によるエレベータ内異常検知技術”，情報処理，vol.48，no.1，pp.17–22，(2007)

[3] 村井泰裕，藤吉弘亘，数井誠人：“時空間特徴に基づくエスカレータシーンにおける人の異常行動検知”，信学技報，PRMU 2008，87，pp. 251–258，(2008)

[4] E. Shechtman and M. Irani：“Space-Time Behavior Based Correlation”，Computer Vision and Pattern Recognition，vol. 1，pp. 405–412（2005）

[5] I. Laptev：“On Space-Time Interest Points”，International Journal of Computer Vision，vol. 64，pp. 107–123（2005）

[6] P. Dollar，G. Cottrell, and S. Belongie：“Behavior Recognition via Sparse Spatio-temporal Features”，In Proc. of Visual Surveillance and Performance Evaluation of Tracking and Surveillance，pp. 65–72（2005）

[7] H. Wang，M.M. Ullah，A. Klaser，I. Laptev and C. Schmid：“Evaluation of Local Spatio-Temporal Features for Action Recognition”，In BMVC 2009-British Machine Vision Conference（2009）

[8] J.C. Niebles，H. Wang and L. Fei-Fei：“Unsupervised Learning of Human Action Categories using Spatial-Temporal Words”，International Journal of Computer Vision，vol.79，no.3，pp.299–318（2008）

[9] G. Csurka，C. Dance and L. Fan：“Visual Categorization with Bags of Keypoints”，In Proc. of ECCV Workshop on Statistical Learning in Computer Vision，pp. 1–22（2004）

[10] C. Schuldt，I. Laptev and B. Caputo：“Recognizing Human Actions: a Local SVM Approach”，Pattern Recognition，In ICPR，vol. 3，pp. 32-36（2004）

[11] T. Hofmann：“Unsupervised Learning by Probabilistic Latent Semantic Analysis”，Machine Learning，ｖol. 42，no. 1-2，pp. 177-196（2001）

[12] 工藤拓，山本薫，坪井祐太，松本裕治：“テキストデータベースからの構文構造のマイニング”，情報処理学会研究報告. ICS,[知能と複雑系]，vol.45，pp.139-144（2002）

[13] 青江順一：“トライとその応用 (< 連載講座> キー検索技法 4)”，情報処理，vol. 34，no. 2，pp. 244-251（1993）

[14] S. Leutenegger，M. Chli and R. Siegwart：“BRISK: Binary Robust Invariant Scalable Keypoints”，In IEEE International Conference on Computer Vision，pp. 2548–2555（2011）

[15] D. Lowe：“Distinctive Image Features from Scale-Invariant Keypoints”，International Journal of Computer Vision，pp. 91–110（2004）

# 第10章.結論

本研究では監視カメラに人間警備員と同等の異常検知能力を持たせることを目的に，画像処理技術を用いて身体的特徴の「顔」を基にした人物の属性分類，行動的特徴である「体の動き」を基にした不審者検知のための行動分類手法を検討する．以下に得られた知見を要約する．

第 1 章では，監視カメラの市場規模の拡大に伴い，バイオメトリクス技術やそれを応用した画像処理技術を導入した監視カメラの今後の展開について具体例を交えて述べた．更に本研究の対象としている人物の生体的特徴を基にした異常検知においては，実環境を想定した場合，課題として照明や天候の変化などによる外乱，隠れ等のオクルージョン，人物の見える角度や姿勢の変化など，多くの課題が存在することを示した．

第 2 章では，本研究で提案する顔画像からの年齢・性別・表情など顔の属性分類において，顔画像の正規化や特徴抽出に用いる Active Appearance Model (AAM)の概要について述べた．それは顔などの予め用意した形状とアペアランスから構成されるモデルと入力物体の二乗和誤差を最小化することで，その形状と形状内部のテクスチャの輝度値を同時に低次元で表現できる統計モデルである．

第 3 章では，本稿において顔の属性分類に利用する Generic AAM について述べた．また従来の AAM と Generic AAM のフィッティング率の比較実験を行った．実験より Generic AAM のフィッティング率は 80%に到達しており，従来と比較して 60%以上の改善が見られ，Generic AAM の有効性を確認できた．

第 4 章では，GAAM を用いた性別分類アルゴリズムを提案した．提案アルゴリズムは独自に定義した形状特徴と GAAM のアペアランスパラメータを特徴量として採用した．実験では HOIP 顔画像データベースを用い，再現率，適合率や年代毎の再現率などを評価項目として性能の検証を行った．適合率は男性 94%，女性 89%であり，その有効性を確認できた．また共に年齢が高いほど再現率が高くなり，特に女性は男性と比較し，全年代において再現率が優れていることを確認できた．

第 5 章では，4 章における提案手法の照明変化に対する脆弱性に触れ，照明変化に対して頑強な LBP，Gabor フィルタ，LGBP の 3 つの特徴量抽出法について紹介した．LBP は単調なグレイスケールの照明変化に頑強である．また Gabor フィルタは顔の空間，および周波数領域における局所的な特徴を抽出でき，照明の影響による顔画像の濃度値情報の変

化を最小限に抑えることができる. 最後に LGBP は Gabor 特徴量と LBP の 2 つのオペレーターより構成され，Gabor 特徴量のゆっくりとした濃淡変位を LBP の適用より符号化することで，情報を高める効果が期待できる.

第 6 章では，照明変化などの不規則なノイズに対する脆弱性に焦点を当て，5 章で紹介した従来手法に関連した新たな特徴量抽出法を検討し，それを用いた顔の年齢・性別分類アルゴリズムを述べ，その性能の検証を行った．提案特徴量は変化が緩やかな Gabor 特徴量の濃淡情報に対して LDP を適用することで, 濃淡情報を重要性の高いエッジ応答の方向を含む符号化情報へと変換する．これより有効性の高い洗練された特徴量を抽出でき，ノイズや不規則な照明変化に対しての頑強性を高める効果が期待できる．実験では，提案特徴量である LGDPHS と 3 つの従来法との性能の比較検証を行った. 結果として提案手法は年齢の分類率が約 64%であり，年齢・性別共に提案手法が従来手法と比較して優れた分類率を示すことを確認できた.

第 7 章では，大局的・局所的な 2 つの要素から構成される特徴量を用いた年齢・性別推定アルゴリズムを提案した．大局的特徴量として顔全体の濃淡値，つまり「見え」を数値化した GAAM のパラメータを用い，局所特徴量として GAAM により正規化された顔領域から抽出した LGDPHS を用いる．年齢推定の実験では FERET データベースを用い，提案手法の MAE は 6.2 歳であり，従来法と比較して最も優れた性能であることを確認できた. また主観評価のモニター20 人の平均の MAE は 7.5 歳であり，提案手法はモニターより 1.2 歳以上優れていることを確認できる．このことから人間が顔から年齢を推測するとおよそ 7−8 歳の誤差があり，提案アルゴリズムは十分に人間と同等の年齢推定能力を有しており，その有効性を確認できた．また性別分類の実験では FERET データベースを用い，提案手法の分類率は 89.4%であり従来手法と比較して最も優れた性能であることを確認できた.

第 8 章では，顔のキーパートに対して LGDPHS を適用した特徴量を用い，顔画像からの表情認識手法について述べ，その性能を検証した．提案手法は正規化されたキーパートのみから特徴抽出を行うことで認識対象者や表情の変化より引き起こされる位置ズレ誤差の問題に頑強な特徴量の抽出が期待できる．実験では Person-independent と Person-dependent な表情認識の 2 つの実験を実施した．JAFFE データベースを使う Person-independent な表情認識の実験では，提案手法は従来法と比較して優れた性能であることを確認できた. また, 怒り(Ang.), 嫌悪(Dis.), 無表情(Neu.), 悲しみ(Sad)の表情は 100%の識別率に達している. しかし提案手法は恐怖の表情を嫌悪として誤分類する傾向にある. Person-dependent な表情認識の実験では，提案手法の分類率は 94.7%に到達しており，従来手法と比較しても優れた性能であることを確認できた.

第9章では，不審者検知のための行動分類を目的とし，時空間のスケール変動にロバストな独自の局所特徴量を用い，行動素の組み合わせと順序から行動を分類する手法を提案した．また本研究では，寺社仏閣に特化した異常行動検知に焦点を当て，寺社仏閣の実環境映像を用いて実験を行った．提案する局所特徴量は，従来の時空間のスケール変動に脆弱な問題に対して，空間軸にBox filterで近似したガウシアンフィルタ，時間軸にGaborフィ

ルタをそれぞれマルチスケールで適用することでロバスト性の向上を図った．実験では，既存のKTHデータセットを用い，独自特徴量（MS-Cuboid）が従来手法より3%程高い80%以上の分類率を示したことで，ガウシアンフィルタに代えて簡易的なBox filterの導入により性能は低下しないことを実証できた．また寺社に特化したシステムとするために，提案手法をMS-Cuboidに加え，行動素の組み合わせと順序から行動を分類する手法へと発展させた．寺社の映像を用いた実験では，3つの異常行動を含んだ6行動を定義し，MS-Cuboidとの分類率の比較検証を行った．提案法の分類率は60.2%であり，従来の50.4%より10%程度の性能の向上を確認できた．しかし未だ実用段階には至っていないと言え，今後は静止状態でも行動の特徴量を算出できるアルゴリズムへの発展と実環境映像データの更なる追加が必要である．

以上，本研究では監視カメラに人間警備員と同等の異常検知能力を持たせることを目的に人物の生体的特徴を用いて人物の属性・行動の分類に取り組んだ．今後は，より厳密な生体的特徴の抽出や夜間やオクルージョンなど，より厳しい動画像の条件に対しても適用可能にするため，更なるアルゴリズムの改良が必要である．

# 第11章.本研究に関する発表論文

## 11.1. 論文（学会論文誌）

(1) Atsushi Higashi，Yohei Fukumizu，Tomonori Izumi，Hironori Yamauchi：“Age and Gender Estimation Using Global and Local Features with AAM and LGDPHS” The Journal of the Institute of Image Electronics Engineers of Japan: Visual Computing, Devices & Communications，vol.41，no.3，pp.262-269（2012.5）

(2) Atsushi Higashi，Yohei Fukumizu，Tomonori Izumi，Hironori Yamauchi：“Expression Recognition Using LGDPHS Based Facial Key Parts”，The Journal of the Institute of Image Electronics Engineers of Japan，vol.41，no.5，pp.496-505（2012.9）

## 11.2. 論文（査読付国際会議）

(1) Atsushi Higashi，Toshiyuki Yasui，Yohei Fukumizu，Hironori Yamauchi：“Local Gabor Directional Pattern Histogram Sequence (LGDPHS) for Age and Gender Classification”, IEEE Statistical Signal Processing Workshop，pp.505-508（2011.6）

(2) Atsushi Higashi, Yohei Fukumizu, Hironori Yamauchi：“Expression Recognition using Local Gabor Directional Pattern Histogram Sequence (LGDPHS)”，International Workshop on Nonlinear Circuits and Signal Processing（2012.3）

## 11.3. 論文（研究会等）

(1) 東 篤司，福水 洋平，山内 寛紀：“Active Appearance Model による顔特徴量を用いた男女識別”，電子情報通信学会技術研究報告.パターン認識・メディア理解，vol.109，no.249，pp.103-108（2009.10）

# 謝辞

本研究を行うに当たり，終始変わらぬ御指導並びに御助言を賜った立命館大学理工学部電子情報工学科教授 山内寛紀博士に深い敬意と感謝の意を表します.

また，ご多忙の中，貴重な時間を割いて議論に参加して頂き，きめ細かい御指導・御助言を賜った立命館大学理工学部電子情報工学科准教授 泉知論博士, 及び立命館大学理工学部電気電子工学科准教授 福水洋平博士に心より感謝致します.

また立命館大学情報理工学部メディア情報学科教授 徐剛博士にはAAMの3次元モデル化を行う際の非剛体の因子分解法について，貴重な御助言を頂きました．心より感謝致します.

そしてSVMなどについて有益な御助力を頂いたTakumiVision株式会社 安井俊之氏に心より感謝致します.

第9章の研究を遂行するに当たり，宗教法人真言宗御室派総本山仁和寺様には現地での映像データセットの作成や本稿への写真掲載にご協力して頂き，心から感謝致します．誠にありがとうございました.

また住友電気工業株式会社 梅村充一様には有益な御助言やデータベースの作成への御協力をして頂き，心より感謝致します.

以下に挙げる立命館大学山内研究室卒業生の皆様には，本研究の推進に関して御指導・御協力して頂き, 心より感謝致します. パナソニック システムネットワークス株式会社 高野智章氏には顔画像解析についての有益な御助言を頂きました. ローム株式会社 草尾康隆氏にはプログラムの実装に当たり貴重な御助言を頂きました．そして同期の山内研究室卒業生の皆様には研究活動だけでなく，学生生活においても大変お世話になり，心より感謝致します.

最後に，この研究を遂行するに当たり，数多くの励ましを頂いた山内研究室の皆様をはじめ関係者の皆様には心よりお礼申し上げます.

# 付録A. Support Vector Machine (SVM)

## A.1. はじめに

本付録では，SVM について述べる．現在さまざまな識別器が扱われているが，本稿内の顔の属性分類では 2 値分類器であるサポートベクターマシン(Support Vector Machine, SVM)に注目した．SVM はニューラルネットワークと違い，学習アルゴリズムに与えるパラメータの初期値によって，最終的な解が異なってくるという局所解の問題を持たないという利点がある．以下に SVM の概要を説明する．

## A.2. SVM の特徴

### A.2.1. マージン最大化

SVM では, **図 A.1** のように学習データの中で最も他のクラスと近い位置にあるサンプルを基準として，その距離が最大となる位置に識別境界を設定する．これを「マージン最大化」という．この理論は入力次元数にあまり関係なく高い汎化能力の維持を期待できる．

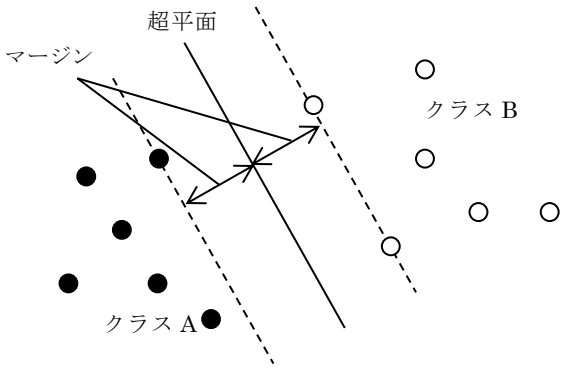



**図 A.1**：マージン最大化

### A.2.2. カーネルトリック

特徴ベクトルを非線形変換し，高次元空間上で線形識別を行う問題に置き換えることで，本質的に非線形な問題に対応させる方法がある．SVM の場合，目的関数や識別関数が入力パターンの内積のみに依存した形になっており，内積が計算できれば最適な識別関数を構成することが可能である．つまり，非線形に写像した空間で二つの要素$\boldsymbol{\phi}(\mathbf{x}_1)$と$\boldsymbol{\phi}(\mathbf{x}_2)$の内積が：

$$\boldsymbol{\phi}(\mathbf{x}_1)^T\boldsymbol{\phi}(\mathbf{x}_2) = K(\mathbf{x}_1, \mathbf{x}_2) \tag{A.1}$$

の様に，入力特徴のみから計算できるなら，非線形写像によって変換された特徴空間での特徴$\boldsymbol{\phi}(\mathbf{x}_1)$や$\boldsymbol{\phi}(\mathbf{x}_2)$を陽に計算する代わりに，カーネル関数$K(\mathbf{x}_1, \mathbf{x}_2)$から最適な非線形写像が構成できる．このように高次元に写像しながら，実際には写像された空間での特徴の計算を避けて，カーネルの計算のみで最適な識別関数を構成するテクニックのことをカーネルトリックという．一般に良く使われるカーネル関数として以下のようなものがある：

- ***d* 次数多項式カーネル**

$$k(\mathbf{x}, \mathbf{x}_\mathrm{i}) = (\mathbf{x} \cdot \mathbf{x}_\mathrm{i})^d \tag{A.2}$$

- **Radial Basis Function (RBF)カーネル**

$$k(\mathbf{x}, \mathbf{x}_\mathrm{i}) = \frac{\exp(-\|\mathbf{x} - \mathbf{x}_\mathrm{i}\|^2)}{c} \tag{A.3}$$

ここで，$c$はスケールパラメータである．

- **シグモイドカーネル**

$$k(\mathbf{x}, \mathbf{x}_\mathrm{i}) = \tanh(\kappa \cdot (\mathbf{x}, \mathbf{x}_\mathrm{i}) + \theta) \tag{A.4}$$

ここで，$\kappa$，$\theta$は任意の実数である．

### A.2.3. 線形 SVM

入力空間$\chi \in \mathbf{R}^n$およびデータ集合$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$が与えられたとすると，線形 SVM の識別関数は次のように定義できる：

$$f(\mathbf{x}) = sign\big(g(\mathbf{x})\big) = sign\big(\mathbf{w}^\mathrm{T}\mathbf{x} + b\big) \tag{A.5}$$

関数$sign(u)$は，$u$ >0 のとき 1，$u$ ≤0 のとき-1 をとる符号関数である．また，自由度として係数$\mathbf{w}$と$b$をパラメータとして与えている．係数$\mathbf{w}$は線形識別器の重みベクトルと呼ばれ，$b$はバイアス項と呼ばれるパラメータである．ここで，$n$個の学習パターン$\mathbf{x}_i(i = 1,2, \ldots, n)$の満たすべき条件を，

$$g(\mathbf{x}) = \mathbf{w}^\mathrm{T}\mathbf{x}_i + b \begin{cases} \geq 1 & \mathbf{x}_i \in \chi_1 \\ \leq -1 & \mathbf{x}_i \in \chi_2 \end{cases} \tag{A.6}$$

とする．点$\mathbf{x}_i$から平面$g(\mathbf{x}) = 0$までの距離は$|g(\mathbf{x}_i)|/\|\mathbf{w}\|$であるから，式(A.6)は，識別関数$g(\mathbf{x}) = 0$から距離$1/\|\mathbf{w}\|$の範囲内，すなわち平面$g(\mathbf{x}) = \pm 1$の間に学習パターンが存在しな

いことを意味する．ここで，$\mathbf{x}_i$の属するクラスを変数$y_i$で表し，

$$y_i = \begin{cases} 1 & \mathbf{x_i} \in \chi_1 \\ -1 & \mathbf{x_i} \in \chi_2 \end{cases} \tag{A.7}$$

と定義し，$\mathbf{x}_i$の教師信号と呼ぶ．また$\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T$とすると，式(A.6)は$y_i$を用いて，

$$y_i \cdot \left(\mathbf{w}^{\mathrm{T}}\mathbf{x_i} + b\right) \geq 1 \qquad i = 1,2, \dots, n \tag{A.8}$$

と書ける．**図 A.2** の平面$H_1, H_2$間の距離(マージン)は$2/\|\mathbf{w}\|$であり，これを最大にする$f(\mathbf{x})$は，扱いやすくするために$\|\mathbf{w}\|^2$を考えて，式(A.8)で表現される制約関数のもと，

$$\tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \tag{A.9}$$

を最小化することで推定できる．

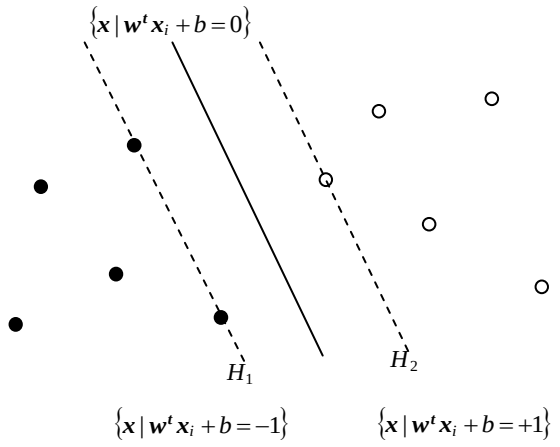



**図 A.2**：制約付き線形識別関数

一般に，制約付きの最適化問題は，その双対問題を考えたほうがより簡単な問題に帰着する場合が多い．そこで，この凸最適化問題を解くため，式(A.9)のラグランジュ関数を計算する．制約条件である式(A.8)は，以下のように書き換えることができる：

$$1 - y_i \cdot \left(\mathbf{w}^{\mathrm{T}}\mathbf{x_i} + b\right) \leq 0 \tag{A.10}$$

この制約条件から，以下のラグランジュ関数が導き出せる：

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(\mathbf{w}^t\mathbf{x}_i + b) - 1) \tag{A.11}$$

ここで，$\alpha_i \geq 0$はラグランジュ乗数である．最適化問題を解くには，このラグランジュ関数を$\alpha_i$について最大化し，$\mathbf{w}$と$b$について最小化する．

最適解においては，パラメータ$\mathbf{w}$と$b$についての$L$の導関数は鞍点において，$L$の勾配が 0 となるので，次式が成立する：

$$\frac{\partial}{\partial b}L(\mathbf{w},b,\boldsymbol{\alpha})=0 \tag{A.12}$$

$$\frac{\partial}{\partial \mathbf{w}}L(\mathbf{w},b,\boldsymbol{\alpha})=0 \tag{A.13}$$

式(A.12)から次式が成立する：

$$\sum_{i=1}^{n}\alpha_i y_i=0 \tag{A.14}$$

また，式(A.13)から次式が成立する：

$$\mathbf{w}=\sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i \tag{A.15}$$

結局, $\mathbf{w}$は学習データの展開式となる. $\mathbf{w}$の解はただ一つに決まるが, ラグランジュ乗数$\alpha_i$はその必要がない.

最適解において，以下の条件が満たされる：

$$\left.\begin{array}{c}\alpha_i[1-y_i(\mathbf{w}^t\mathbf{x}_i+b)]=0 \qquad i=1,2,\dots,n\\ 1-y_i(\mathbf{w}^t\mathbf{x}_i+b)\le 0 \qquad i=1,2,\dots,n\\ \alpha_i\ge 0 \qquad i=1,2,\dots,n\end{array}\right\} \tag{A.16}$$

これはクーン・タッカー(Kuhn-Tucker)条件と呼ばれ，ラグランジュ未定乗数法を用いた際，成り立つ．この条件を満たし，$\alpha_i\ge 0$を有する学習データ$\mathbf{x}_i$をサポートベクターと呼ぶ．$\alpha_i=0$となるサポートベクター以外の学習データは凸最適化問題の解放には関係ないものとなる．つまり，サポートベクター以外の学習データは式(A.8)の制約条件を自動的に満たし，式(A.15)の展開項の部分には現れない.

式(A.11)のラグランジュ関数に式(A.14)，式(A.15)の条件を代入すると，双対問題となる以下の凸最適化問題を得ることができる：

$$\begin{array}{ll}\text{目的関数} & \displaystyle\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j\, y_i y_j\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j\right)\rightarrow \boldsymbol{\alpha}\text{について最小化}\\ \text{制約条件} & \alpha_i\ge 0,\quad i=1,2,\dots,n\\ & \displaystyle\sum_{i=1}^{n}\alpha_i y_i=0\end{array} \tag{A.17}$$

最適な$\boldsymbol{\alpha}$から$\mathbf{w}$を得るには，式(A.15)の関係を用いる．また$b$は

$$b=-\frac{1}{2}(\mathbf{w^t x_{+1}}+\mathbf{w^t x_{-1}}) \tag{A.18}$$

で求められる．ここで，$\mathbf{x_{+1}},\mathbf{x_{-1}}$は，それぞれクラス 1，-1 に属するサポートベクターである.

式(A.15)の展開式を識別関数の式(A.5)に代入することによって，式(A.5)の識別関数を，分類されるパターンとサポートベクターとの内積で評価される次式に書き換えることができる：

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j + b\right) \tag{A.19}$$

以上より，凸二次計画問題を解くことで，識別関数

$$f(\mathbf{x}) = sign(\mathbf{w}^{\mathrm{t}}\mathbf{x} + b) \tag{A.20}$$

を得ることができる．

実際に線形 SVM を実装し，識別関数の決定を行った例を**図 A.3** に示す．

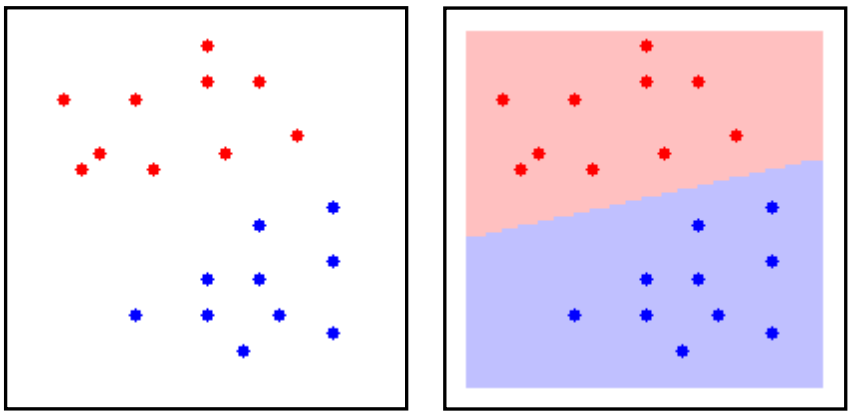


**図 A.3**：線形 SVM による識別結果

### A.2.4. 非線形 SVM

線形 SVM は線形分離可能な場合には高い汎化能力を達成できるが，実際の問題では線形分離可能な場合は多くない．そこで，より一般的な識別関数を推定するため，前処理をして入力ベクトル$\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$を次式のように高次元特徴空間に写像し，その後，その特徴空間で線形 SVM を行うという方法が考えられる．

$$\Phi: \boldsymbol{x_i} \mapsto \boldsymbol{z_i} \tag{A.21}$$

ここで，$\boldsymbol{z_i}$は観測された入力ベクトル$\boldsymbol{x_i}$を高次元特徴空間に写像した結果である．

制約条件を表す式(A.7)と目的関数を表す式(A.8)で表現される高次元特徴空間上での最適化問題は，$\boldsymbol{z_i}$を使って以下のように記述できる．

$$\text{目的関数} \quad \tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \rightarrow \mathbf{w}\text{について最小化} \tag{A.22}$$

$$\text{制約条件} \quad y_i \cdot \left(\mathbf{w}^{\mathrm{T}}\mathbf{z}_{\boldsymbol{i}} + b\right) \geq 1 \qquad i = 1,2,\ldots,n \tag{A.23}$$

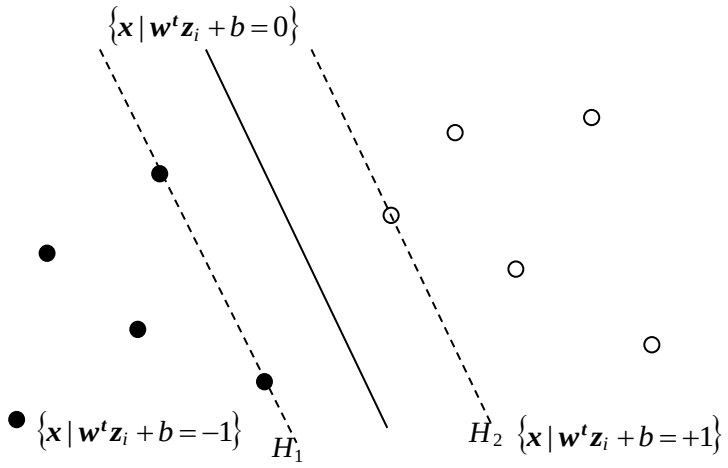



**図 A.4**：非線形 SVM における制約付き線形識別関数

線形 SVM と同様に, この最適化問題を解くため, ラグランジュ関数を計算する. 式(A.23)で表現される最適化問題の制約条件は以下のように書き換えることができる.

$$\text{制約条件} \quad 1 - y_i \cdot \left(\mathbf{w}^{\mathrm{T}}\mathbf{z}_{\boldsymbol{i}} + b\right) \leq 0 \tag{A.24}$$

この制約条件から以下のラグランジュ関数が導き出せる.

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(\mathbf{w}^t\mathbf{z}_i + b) - 1) \tag{A.25}$$

ここで，$\alpha_i \geq 0$はラグランジュ乗数である．最適化問題を解くには，このラグランジュ関数を$\alpha_i$について最大化し，$\mathbf{w}$と$b$について最小化する．

以降は線形 SVM と同様の手順で双対問題となる最適化問題を得ることができる.

$$\text{目的関数} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j \, y_i y_j \left(\mathbf{z}_i^{\mathrm{T}}\mathbf{z}_{\boldsymbol{j}}\right) \rightarrow \boldsymbol{\alpha}\text{について最小化}$$

$$\text{制約条件} \quad \alpha_i \geq 0, \quad i = 1,2,\ldots,n \qquad \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{A.26}$$

式(A.26)で表現される最適化問題を解くためには，高次元特徴空間上で以下の計算をす

る必要がある．

$$(\mathbf{z}^{\mathrm{t}}\mathbf{z}_i) = (\boldsymbol{\phi}(\mathbf{x})\,{}^{\mathrm{t}}\boldsymbol{\phi}(\mathbf{x}_i)) \tag{A.27}$$

式(A.27)で表現される内積の計算には膨大な計算が必要となる．元の観測空間で定義される次式を満たすカーネル関数を用いて，高次元特徴空間上へ写像するという膨大な計算を削減できる：

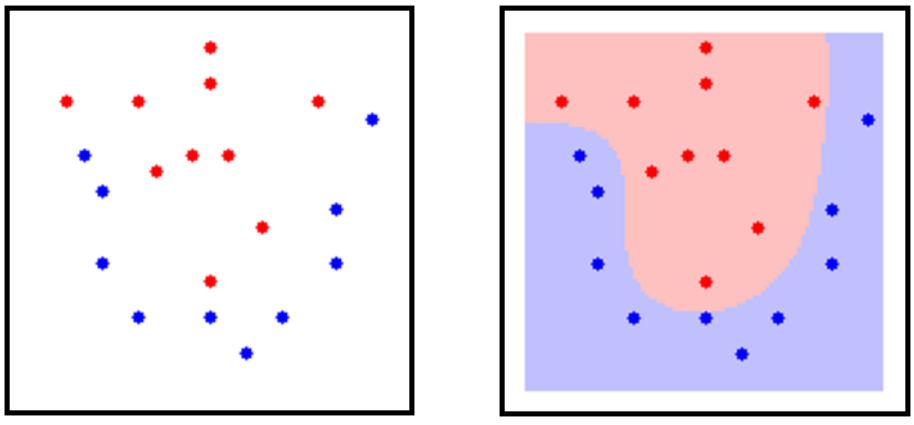
$$\left(\boldsymbol{\phi}(\mathbf{x})^{\mathrm{t}}\boldsymbol{\phi}(\mathbf{x}_i)\right) = k(\mathbf{x}, \mathbf{x}_i) \tag{A.28}$$

このカーネル関数を用いると，高次元特徴空間での式(A.20)に相当する識別関数を導出することができる．式(A.20)の$\mathbf{x}$に$\mathbf{z} = \boldsymbol{\phi}(\mathbf{x})$を代入して，

$$\begin{aligned} f(\mathbf{z}) &= sign\left(\sum_{i=1}^{n} y_i \cdot \alpha_i \cdot \mathbf{z}^{\mathrm{t}}\mathbf{z}_i + b\right) \\ &= sign\left(\sum_{i=1}^{n} y_i \cdot \alpha_i \cdot \boldsymbol{\phi}(\mathbf{x})^{\mathrm{t}}\boldsymbol{\phi}(\mathbf{x}_i) + b\right) \\ &= sign\left(\sum_{i=1}^{n} y_i \cdot \alpha_i \cdot k(\mathbf{x}, \mathbf{x}_i) + b\right) \end{aligned} \tag{A.29}$$

結局，観測空間であるユークリッド空間の内積に代わって，適切なカーネル関数$k$を選択できれば，このカーネル関数$k$に基づく非線形 SVM には，A.2.3 節で述べた線形 SVM の特性がすべて適応できる．

実際に非線形 SVM を実装し，識別関数の決定を行った例を**図 A.5** に示す．



(a)学習データ　　(b)決定境界

**図 A.5**：非線形 SVM による識別結果

**図 A.6** に非線形 SVM の原理を示す．**図 A.6** では，観測空間上のデータ（**図 A.6(**上左**)**）を非線形写像に用い，より高次の特徴空間にマッピングし，特徴空間上で分離可能な超平面を作成することで（**図 A.6(**下左**)**），入力空間では非線形の識別空間になる（**図 A.6(**下右**)**）．**図 A.6(**上右**)**は非線形 SVM が構成できる様子を示している．

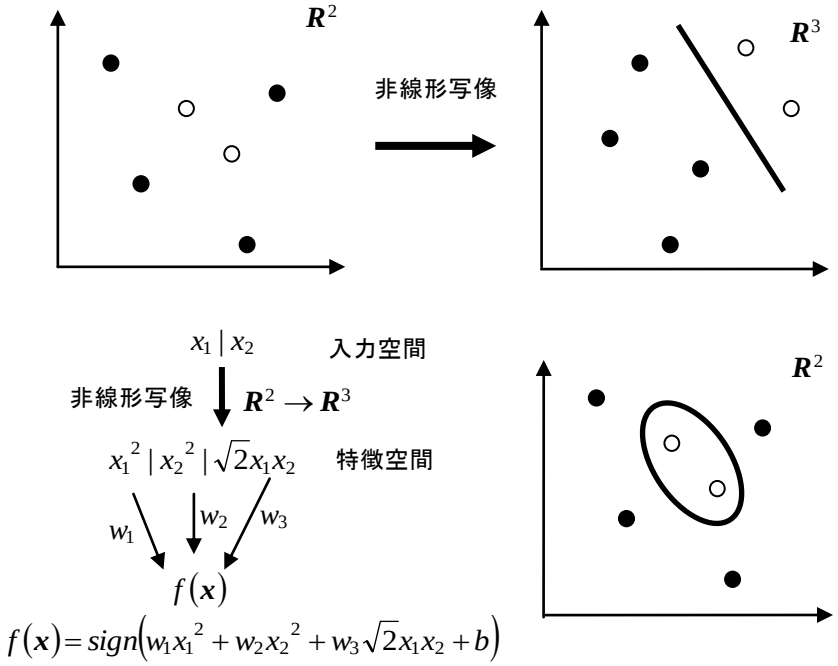



**図 A.6**：非線形 SVM の原理