顔映像情報からの人の状態推定に関する研究

２０１６年　３月

嶋田　敬士

顔映像情報からの人の状態推定に関する研究

嶋田　敬士

システム情報工学研究科

筑波大学

２０１６年　３月

目次

図目次

表目次

第1章　序論

1.1　背景

近年，携帯電話，スマートフォンやタブレット端末などをはじめとした携帯情報端末にはカメラが標準的に搭載されるようになり，また家庭内でも PC，TV や家庭用ゲーム機にカメラを搭載しようとする流れもあり，我々の生活の中でカメラを利用することは特別なことではなくなってきた．そのような人々の生活環境の変化に伴い，この 10 年ほどで人々のコミュニケーションの方法も従来の文章や音声中心によるものから映像を用いたものも一般的になるほど変化してきており，映像情報の活用の場は急速に広がりを見せてきている．事実，従来はデータサイズや通信速度の問題からテキストや音声に比べて記録・配信に課題があった映像情報も記憶媒体の進化，高速な通信網やクライドコンピューティングによる潤沢な IT 資源の整備を背景に，インターネット上には様々な映像が溢れ，一般ユーザがそれらを公開・視聴できる環境が整ってきた．一方，実世界においても主にセキュリティを目的として学校，駅や空港などの公共の場や店舗・商業施設，一般家庭に至るまで監視カメラが設置され屋内外の映像を常時記録し続けており，今や街中でカメラを見つけること自体が特段珍しいことではなく，事実そういった監視映像が重大事件の解決の決め手となる例も少なくない．一方，デバイスとしてのカメラのコストも，1990 年代初頭にデジタルカメラへ搭載され，2000 年代から携帯電話に，2005 年頃からスマートフォンへも標準搭載されるにつれ爆発的に市場に供給されたことで急速に低下してきており，今やカメラ機能は単に搭載されるだけでは特別な商品付加価値を持たない域にまで達した．そこで昨今では，撮像目的のみに留まらずユーザ・インタフェースとしてカメラを活用しようという新たな流れもある．ゲーム機や TV などでは搭載されているカメラを利用したジェスチャ認識機能などにより商品の差別化がなされてきている．このような外的環境の変化に伴い撮像される側，つまり被写体たる人間の意識も次第に変化してきており，自身が撮影されることへの敷居も低くなってきている．つまり現代において映像情報は様々な形で人々の日常生活の中に溶け込み，撮像する側，される側双方にとって容易に利用可能な身近で有益な情報媒体となってきた．

このように非侵襲・非接触・非拘束に対象をセンシング可能なカメラの普及及びそこから得られる映像情報を積極的に利活用しようという社会環境の変化に伴い，映像情報を単なる記録・配信の対象から，解析・理解の対象へと応用しようとする試みもなされてきており，その根幹となる画像認識の分野では基礎から応用まで様々な研究・開発が益々活発に行われてきている．

1.2　研究の目的

カメラ映像を活用した代表的なアプリケーションの一つとしてセキュリティがあげられる．特に近年，個人情報保護法の施行や犯罪の増加や凶悪化などにより人々のセキュリティに対する意識・関心は非常に高まっている．こういった社会情勢を背景に，従来は単なる記録に留まっていた映像情報が，画像認識技術の高まりも相まってより高度な認証技術へと活用されてきている．例えば近年，人物の生体固有情報を利用する認証方式である所謂バイオメトリクス認証として顔，指紋，虹彩，静脈や声紋など用いたものも実用化されてきており，社会的なカメラの普及と画像処理・認識関連研究の活性化と実用化促進にも大きな影響を与えている．これら，バイオメトリクス認証の中でもカメラの非接触･非侵襲･非拘束というメリットを活かした技術として顔認証があげられる．認証対象者は手をかざしたり，声を発したりするなどの特別なアクションを起こす必要がないことがないためシームレスな認証技術として着目されている．しかし，実際の運用においては課題もあり，例えば，カメラを注視する必要があったり，ほぼ正面からの顔画像でしか認識をできなかったりなど，必ずしもその利点を最大限に活かしている状況とはなっていない．つまり，人物の顔を画像認識の対象とした場合，その真のメリットを引き出すには人物の広範囲に変化する顔姿勢に対応した技術が望まれる．

広範囲の顔姿勢に対応した技術が開発できると，被写体に対する拘束条件を緩くできるため実に様々な応用が考えられる．日本が世界に誇る産業の一つとして自動車産業があげられるが，近年の自動車業界全体の動向やユーザの関心を俯瞰してみると，環境・安全に対する注目度が非常に高くなってきていると感じられる．もはや従来の自動車の評価基準であった動力性能と同等かそれ以上に環境・安全性能が問われる時代になってきている．そのような状況を背景に，主に安全を目的として車外あるいは車内へのカメラ搭載が進んできているが，車外の障害物検知や車内での運転手の居眠りや脇見といった状態検知においては，照明環境の問題，カメラの設置位置やドライバの動きの速さや大きさから実用上の課題もまだ多い．しかし例えば，市街地の交差点における交通事故はドライバの安全確認が不十分だったことによるものも多く，もし顔姿勢が検知できれば，ドライバの安全確認行動の有無を判定し注意喚起することも可能となり，道路での安全性の向上に大きく寄与できる．

自動車などの交通事故も社会問題の一つではあるが，日本が抱える大きな社会問題のひとつに超高齢社会の到来があげられる．日本は先進国の中でも特に急速に高齢化が進んでおり，それに付随する医療・介護・福祉などの社会保障に関する課題も多く，解決が急務である．特に近年は，要介護者等の自立を促すためのリハビリテーションなどの

重要性が年々増してきている．リハビリテーションでは患者と療法士がコミュニケーションを取りながらその回復に努めるという構造上，療法士の主観による評価が一般的で療法士は患者の様子を注意深く観察し評価を行う必要がある．しかし，患者の症例によっては評価項目の妥当性や点数の信頼性などの問題があり，異なる症例でも共通に適用可能な客観的評価手法を通じて，その介入効果を検証することが望まれている．一方，リハビリテーションの主体である患者に表出する状態に目を向けると，心身賦活に伴って広く一般的に表情が豊かになることが経験的に分かっている．なかでも笑顔は患者のポジティブな状態をよく反映しており，回復度合いが良好な場合に多く見られるようになる．そこで，人物の顔表情の中でも特に笑顔に着目しその定量化とリハビリテーション効果の客観的評価へと応用することを考える. 本取組は高齢社会に対する IT 技術の応用の一例に過ぎないが，このような研究の積み重ねにより，他国に先んじて得られた知見，技術やサービスは将来，国際競争力のある新たな日本のサービス産業として育つ可能性もあると考える．

表情は人物（広くは動物）の顔に表れる動的な状態変化の一つであるが，一方でその人物の属性やプロフィールといった静的な状態を把握したいという強いニーズがある．人物の属性としては，性別，年齢，血液型，人種など生まれ持った永続的なものから職業，趣味や特技など一時的なものまで様々あるが，見た目から判別可能な基本的属性である性別や年齢が顔画像から自動的に推定できるだけでも多くの応用が考えられる．例えば，デジタルサイネージ用ディスプレイの上部にカメラを設置して視聴者の年齢や性別を推定し，視聴者に応じた広告を動的に表示したり，ディスプレイや店舗の看板の前に留まった時間を計測するなどして広告の効果測定として利用したりすることができる．また商店街やショッピング・モールなどに設置されている監視カメラを利用して来店者の人数，年齢や性別を推定することで，曜日や時間帯による客数・客層の変化を把握したり，タイムセールや販売促進の効果を測定したりすることできる．現代社会における我々の生活は多くの選択肢があるが故に複雑化し，人々の嗜好や生活様式も多様化してきている．このような状況下において，自動的に消費者を把握・分析し，その情報をマーケティングに活用できる仕組みを構築することは，消費者の真のニーズを掴むことが難しい今の世の中における経済活性化のための有効な手段の一つとなり得ると考える．

本論文は，映像情報の中でも特に人の“顔”に着目し，顔映像情報からその人の眼，鼻や口などの顔器官，属性（プロフィール），姿勢や表情などの静的又は動的な状態を推定すること試み，その推定手法と実環境への応用について提案するものである．人々の日常生活への映像情報の普及に伴い，人物の動作，行動や状態を認識し，それを応用して様々なサービスに役立てようという試みへの期待は高まってきている．しかし，被写

体や撮像環境に特別な拘束条件を課さない実環境下の映像は非常に多様であり，いかなる状況においても安定して映像を認識・理解することは困難である場合が多い．本論文では，これら実環境下での人物状態推定を試み，ロバストな技術の確立を目指すと共に提案技術を応用した具体的なアプリケーションを設定して実用化に向けた成果と課題を確認しつつ，顔映像情報の認識技術を広く人々の社会生活へ活用することを目指した．

1.3 本論文の構成

本論文は全 6 章より構成され，第 1 章では研究の背景と目的について述べた．

第 2 章では，画像認識のための特徴量と識別手法について，本論文と関連する手法を中心に説明する．

第 3 章では，人間の動作の中でも特に顔姿勢の変化に着目し，その顔映像情報からの推定手法について検討する．そして，今後整備されていく道路交通網における路車間協調システムを見据え，運転中のドライバの安全確認行動時の顔姿勢変化の追跡を試み，安全性の向上に役立てることを提案する．まず交通死亡事故の発生が多い市街地交差点において求められる顔姿勢変化の検出範囲を明らかした後，自動車運転時という実利用環境を想定し，昼夜問わず撮像可能な近赤外線カメラをドライバの正面となる車内ダッシュボード上に設置すること仮定し，撮像される 256 階調グレイ・スケール映像を処理対象とする．顔及び各種顔器官の検出に最も性能が良い識別器の 1 つとして知られる SVM（Support Vector Machine） [1] [2] [3]を用い，求められる広範囲の顔姿勢変化にも対応可能な各顔器官の探索範囲を調査し，ロバストかつ高速な検出器の構築を試みる．そして，検出された顔器官の撮像画像内での 2 次元座標位置から，予め 3D スキャナを用いて構築した 3 次元顔モデルとパーティクル・フィルタにより実 3 次元空間内でのドライバの顔姿勢の推定を試みる．提案手法を実験室内での基礎実験と実際に車道を運転した時の映像へ適用し，その結果を通じて本提案手法の実用可能性と今後の課題について述べる．

第 4 章では，人間の感性情報の中でも視覚的に表出する部分である表情に着目し，その顔映像情報からの推定と，医療・ヘルスケア分野への応用として従来定量化が困難であったリハビリテーション効果の客観的評価方法について検討する．我々人間は，相手の仕草や顔の表情を視覚的に確認できることで，より深く相手の意図や状態を理解・把握することができる．本章では，そうした表情の中でも音楽療法中の患者の症状の改善過程に伴って頻繁に観察されるようになるポジティブな表情変化である笑顔に着目し，SVM により笑顔度（笑顔の度合い）を推定し定量化する手法と，その療法効果評価への

応用方法について述べる．はじめに研究用に公開されている画像データベースを用いて提案する笑顔度推定手法の性能評価を行った後，実際の音楽療法セッション中の記録映像から患者の笑顔度を定量的に測定することを試みる．そして，最後に療法経過や介入内容による笑顔度の変化を多重比較することによる音楽療法効果の客観的な評価方法ついて従来の主観評価結果と比較しながらその有効性と将来性を示す．

第 5 章では，人間が生まれながらにして備え持っている基本的な属性の中でも性別と年齢に焦点を当て，その顔映像情報からの推定手法と応用システムについて検討する．人物の性別・年齢情報は様々な利用可能性があるが本章では特に，近年急激にそのニーズが増加しているこれらのマーケット調査分野への応用を検討する．例えば，従来店舗や商業施設を利用する顧客の性別・年齢を知るためにはアンケートやカード利用履歴による方法や，目視による確認方法があったが，いずれも調査対象が限定的で高コストであり, また個人情報保護の観点からそれを収集すること自体が難しい時代になってきた．そこで我々は，定点カメラで商業施設等の来場者を撮影し，上半身画像から人物を検出することで来場者数をカウントし，更に顔画像から性別・年齢を推定することで自動的に人物属性構成比率を算出するデモグラフィック調査のための来場者モニタリング・システムを提案する．本章では，基礎実験を通じて性別・年齢の顔画像からの推定可能性について論じた後，SVM により性別を，SVR（Support Vector Regression） [4]により年齢を推定する手法を提案する．そして提案手法を性別・年齢が既知である画像データベースに適用した結果と実際に商業施設に設置した来場者モニタリング・システムから得られたカメラ映像に適用した結果について述べる．

そして最後に，本論文で述べた研究成果と今後の課題・展望について第 6 章で総括し結論とする．

第2章　画像認識のための特徴量と識別手法

画像から目的とする対象物を検出・識別するためには，一般的にはまず入力画像内の任意の大きさの矩形領域に着目する．そして，その領域から対象物の特徴をよく抽象化した画像特徴量を抽出した後に，その抽出された画像特徴量を用いて対象物の存在有無を判別するという過程をたどる．本章では，本論文に関連する画像特徴量及び識別手法について説明する．

2.1　画像特徴量

入力された画像から対象物を識別するためには，通常は予め画像内に存在する対象物の大きさと存在位置はわからないため，注目領域の大きさを変化させつつ画像内を走査しながら識別しようとする対象物の大域的又は局所的な色，模様，形状や構造といった特徴をうまく抽象化した特徴量を抽出する必要がある．識別しようとする対象物によって様々な特徴量が提案・研究されているが，本論文に関連した特徴量について以下に簡潔に説明する．

2.1.1　正規化輝度値

最も単純な画像特徴量の一つは入力画像の輝度値そのものを利用するものである．画像は 2 次元の情報であるが，今 $N \times N$画素の入力画像 Iがあったとき，これを例えば左上画素の輝度値から右下画素の輝度値へと順に並べると $N \times N$次元のベクトル vへと変換できる（図 2-1 参照）．

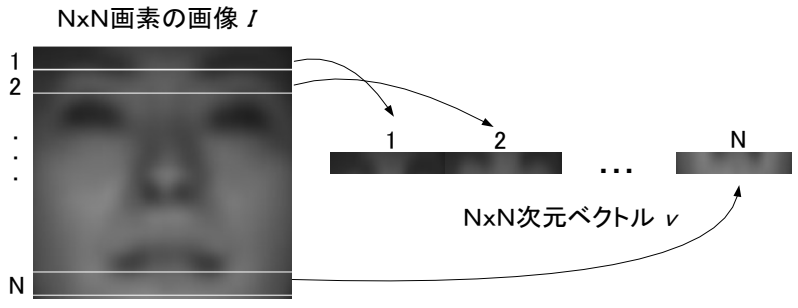


図 2-1　2 次元画像のベクトルへの変換例

ここで本論文では，入力画像の輝度値が環境光や照明等の外乱により変動することを

考慮して，式(2.1)により算出されるベクトル $\mathbf{v}$をその L2 ノルム $\|\mathbf{v}\|_2$で正規化した正規化輝度値ベクトル $\mathbf{v}'$を特徴量とする．

$$\mathbf{v}' = \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \tag{2.1}$$

2.1.2 LIH（Local Intensity Histogram）

LIH [5]は画像の局所領域内の輝度ヒストグラムを算出し，それらを連結した特徴量である．局所領域に分割することで領域間の配置関係を記述でき，またヒストグラムを用い輝度値の分布のみに着目することで局所領域内での物体の形状変化に強い特徴量である．その算出ステップを以下に，そして格子セル数を 8 × 8，ヒストグラムビン数を 8 としたときの笑顔/非笑顔検出時の特徴抽出例を図 2-2 に各々示す．

1. 顔画像をある一定の大きさの格子セル（$M \times N$）に分割する
2. 各格子セル領域内で L 階調輝度ヒストグラムを求める
3. 各格子セルの輝度ヒストグラムの総和が 1 になるように正規化を施す
4. 格子セルごとに求められた正規化された局所輝度ヒストグラムを連結し，$M \times N \times L$ 次元の特徴量を得る

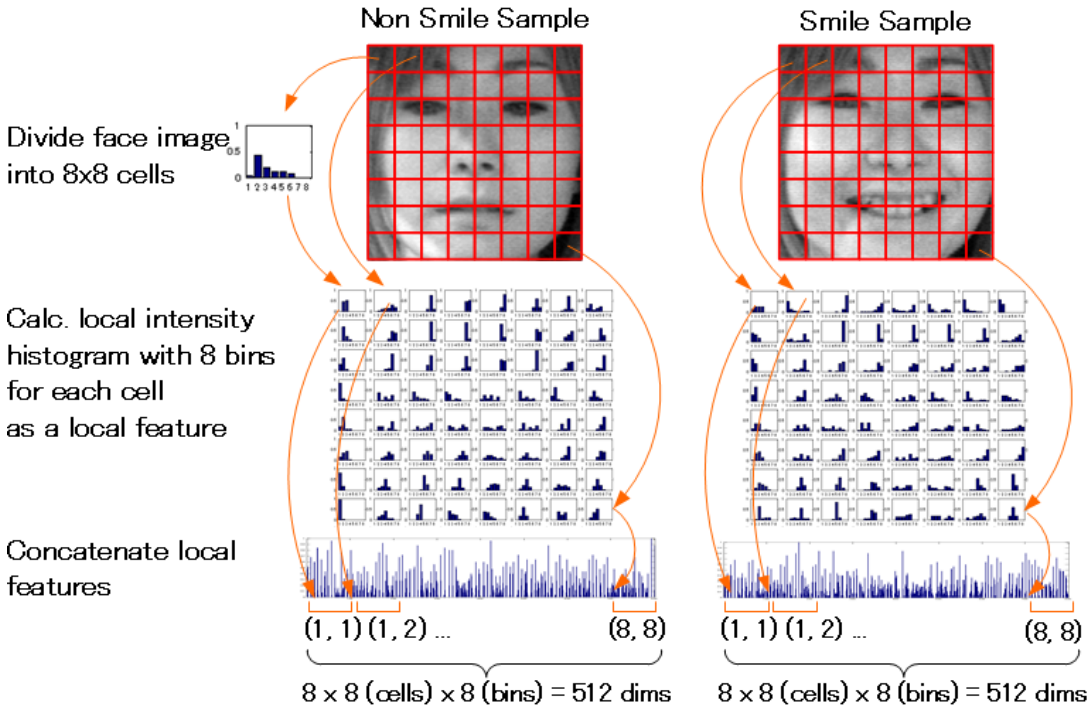


図 2-2　笑顔検出のための LIH 特徴量の抽出例

2.1.3 LBP（Local Binary Patterns）

LBP [6]は輝度変化に強いテクスチャ特徴を抽出できる特徴量で, その改良手法を含め, 顔検出，顔認証や表情認識など画像認識分野 [7] [8] [9]でよく用いられる特徴抽出法の１つである．その計算方法は，注目画素 n_c に対して，半径を R，近傍数を N，i 番目の近傍画素の値を n_i とすると，式(2.2)により求まる．

$$LBP_{R,N}(x,y)=\sum_{i=0}^{N-1}s(n_i-n_c)2^i,\quad s(x)=\begin{cases}1 & x\geq 0\\ 0 & otherwise\end{cases} \tag{2.2}$$

例えば近傍数 $N=8$，半径 $R=1$ とした場合，ある注目領域内での LBP 及び LBP ヒストグラムは図 2-3 のように計算でき, このとき LBP ヒストグラム特徴量は 256 次元となる．

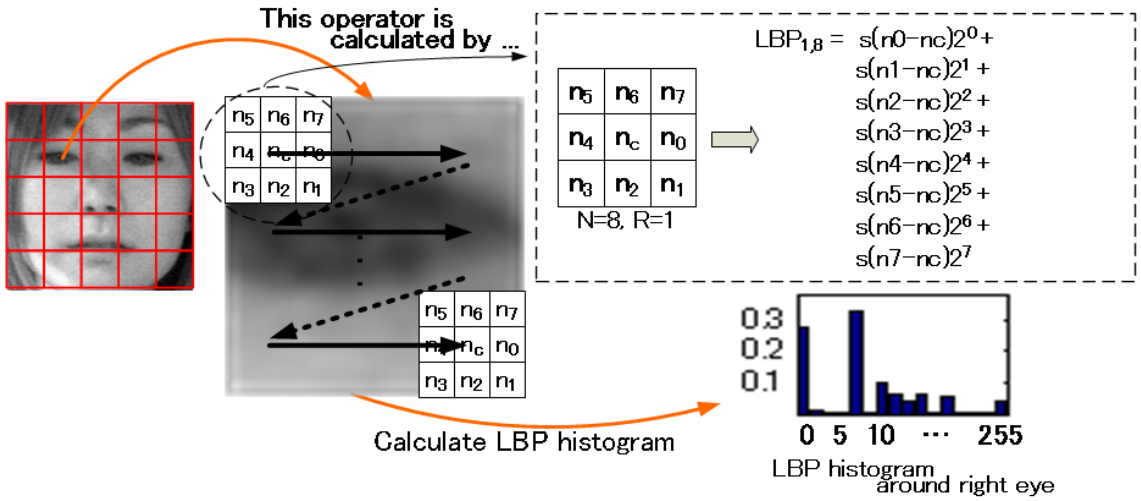


図 2-3　LBP の計算例

また LBP は様々な改良が提案されており，例えば，中心画素値ではなく近傍画素値の平均を求めその値と近傍画素値との大小比較を行う improved LBP [10]や uniform パターンと呼ばれるパターンを用いることで次元数を大幅に削減できる uniform LBP [11]などがある．Uniform LBP では，LBP を算出した結果のビット配列において最大でも 2 回の 0→1 又は 1→0 への遷移のみを許す. 例えば, 01111000, 11111111 や 11111011 は各々uniform パターンであるが 01010100，01011110 や 10101101 は uniform パターンではない．このように uniform ではないビット配列となった LBP をヒストグラムの算出時に考慮しない結果，例えば $N=8$，$R=1$ の場合に uniform LBP ヒストグラムを

特徴量とした場合には次元数を 256 次元から 59 次元へと圧縮できる．

2.1.4 CS-LBP（Center Symmetric – Local Binary Patterns）

CS-LBP [12]は単純な計算方法ながら，輝度変化に強いテクスチャ特徴を LBP よりもコンパクトに表現できることで知られており，近年物体検出などでよく使われる特徴抽出法の 1 つである．その計算方法は，注目画素に対して，半径を R，近傍数を N，i 番目の近傍画素の値を n_i，エンコード閾値を T とすると，式(2.3)により求まる．

$$CS-LBP_{R,N,T}(x,y)=\sum_{i=0}^{(N/2)-1}s(n_i-n_{i+(N/2)})2^i,\quad s(x)=\begin{cases}1 & x\geq T\\ 0 & otherwise\end{cases} \tag{2.3}$$

ここで本論文では，近傍数 $N = 8$，半径 $R = 1$ としたため，注目画素を n_c としたとき，ある注目領域内での CS-LBP は図 2-4 のように計算できる．

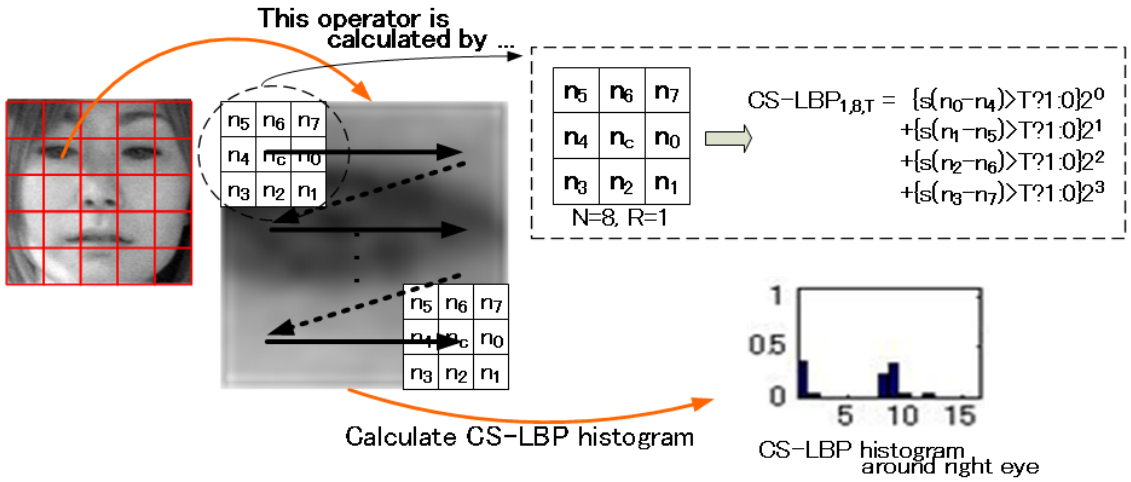


図 2-4　CS-LBP の計算例

このようにして計算される CS-LBP を用いた CS-LBP ヒストグラム特徴量の算出ステップを以下に示す．

1. 顔画像をある一定の大きさの格子セル（$M \times N$）に分割する
2. 各格子セルの CS-LBP 値のヒストグラム（CS-LBP ヒストグラム）を算出する
3. 各格子セルの CS-LBP ヒストグラムの総和が 1 になるように正規化を施す
4. 格子セルごとに求められた CS-LBP ヒストグラムを連結して $M \times N \times 16$ 次元の

特徴量を得る

また, 図 2-5 に格子セル数を 5×5 とした時の笑顔/非笑顔検出時の CS-LBP ヒストグラム特徴抽出例を示す.

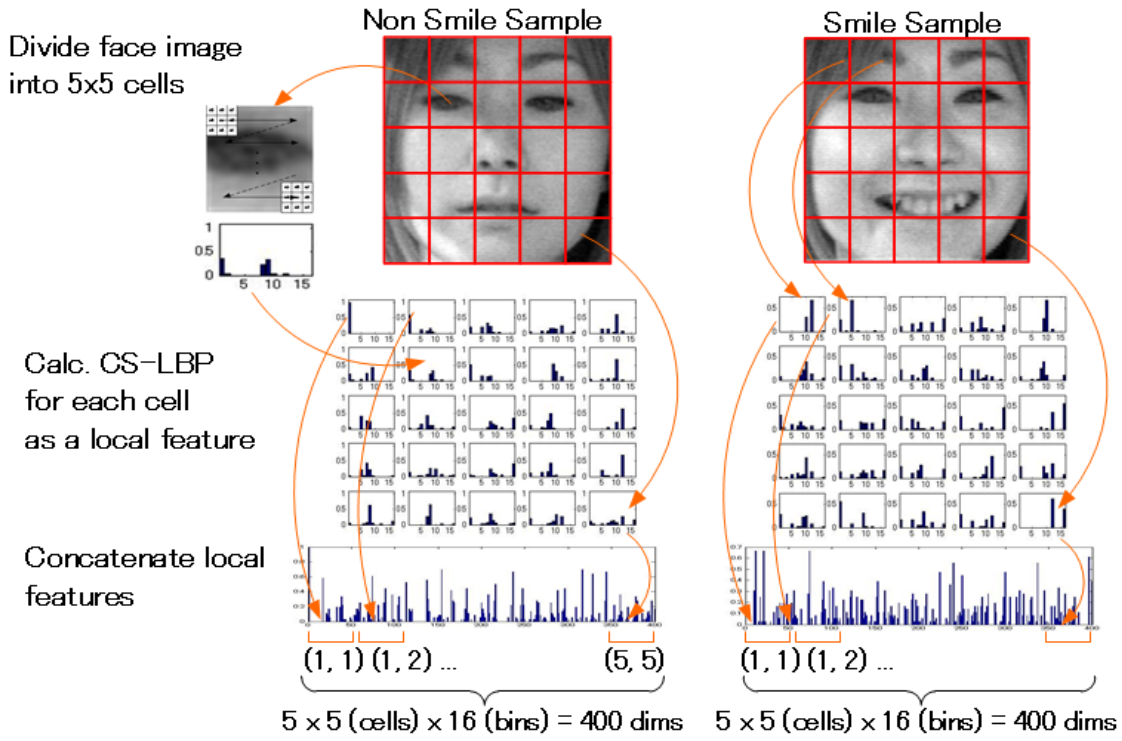


図 2-5　笑顔検出のための CS-LBP 特徴量の抽出例

2.1.5 LIH + CS-LBP

LIH + CS-LBP 特徴量はその名の通り，LIH 特徴量と CS-LBP 特徴量を結合した特徴量で，注目矩形領域を格子状のセルに分割し，各セル内で LIH 特徴量と CS-LBP 特徴量を算出した後に，それらを連結した特徴量である．図 2-6 に LIH 特徴量及び CS-LBP 特徴量共にセル数 5×5 としたときの性別・年齢推定時の特徴抽出例を示す．

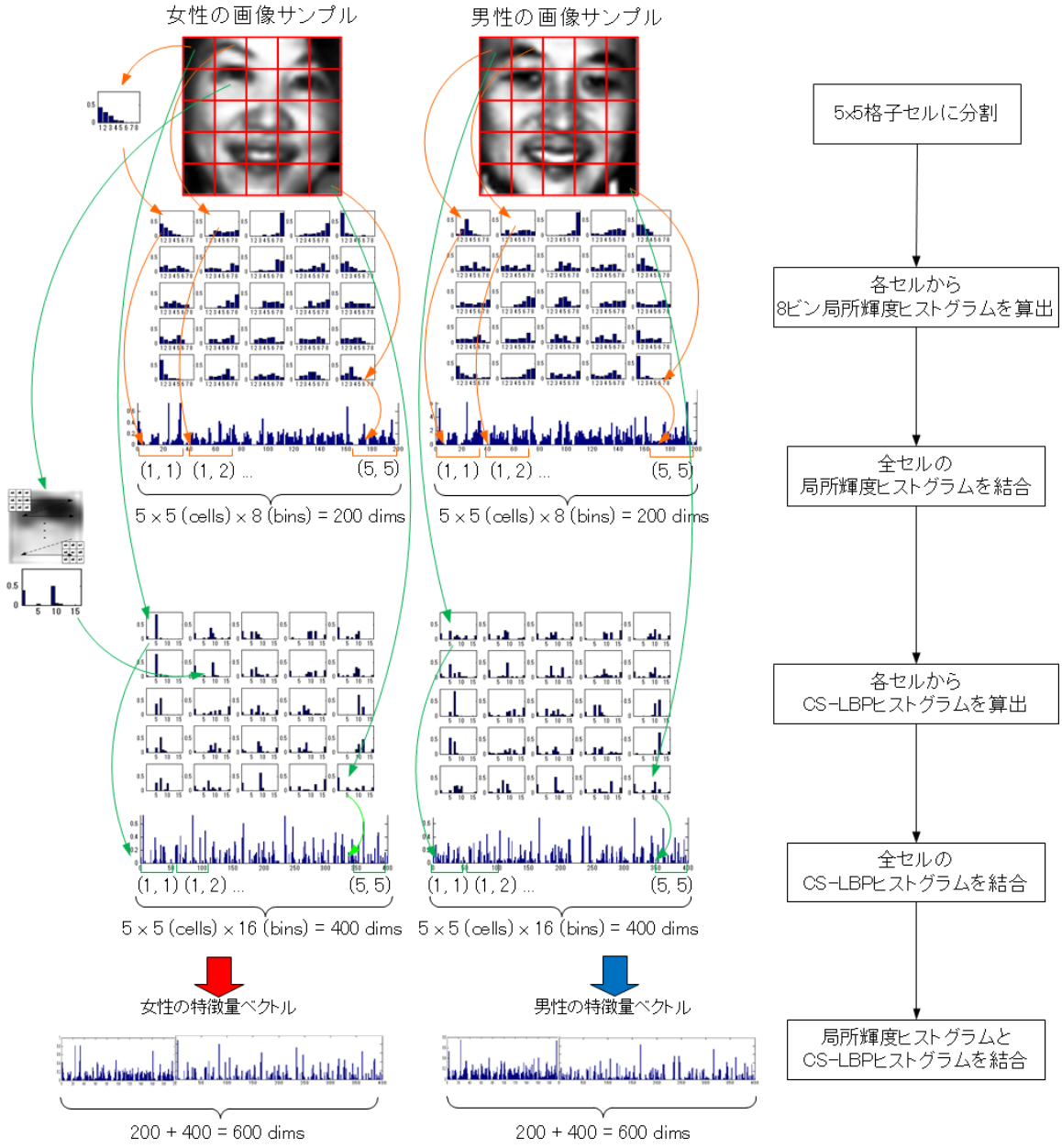


図 2-6　性別推定のための LIH + CS-LBP 特徴量の抽出例

2.1.6 Haar-Like

Haar-Like 特徴量 [13]は画像中に複数定義された局所的な矩形領域間の輝度値の総和の差を求めることで, 画像中の局所領域の明暗情報を抽出することができる. 例えば, 図 2-7 に示したように白い矩形領域の輝度値の総和と黒い矩形領域の輝度値の総和の差を求めたものが各 Haar-Like 特徴の特徴量 H となる.

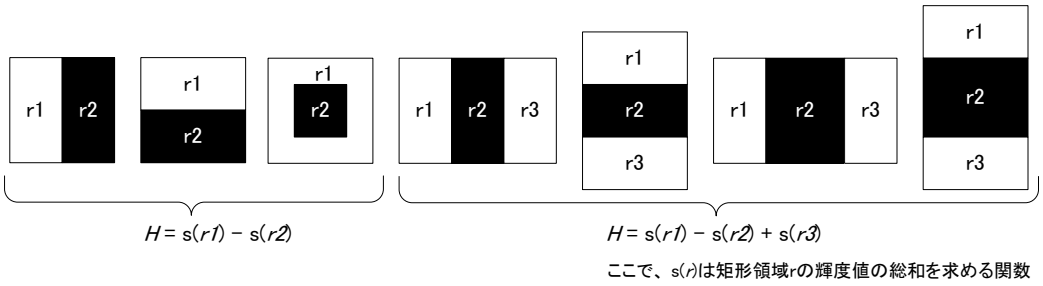


図 2-7 Haar-Like 特徴の例

このように Haar-Like 特徴は画像内の任意の位置に配置された矩形領域同士の相対的な明暗情報しか反映しないため, 例えば入力画像の明るさ変動に対してロバストである. また入力画像から予め左上から右下へと画素値を積算したインテグラル・イメージ（積分画像）を作成しておくことで，矩形領域の位置と大きさに依らず数回の加減算のみで高速に特徴量を算出できる.

2.1.7 HOG（Histogram of Oriented Gradients）

HOG 特徴量 [14] [15]は，画像勾配の強度を方向別ヒストグラムとして表現した特徴量で物体のシルエット/形状情報をうまく抽象化できることから人物検出などによく用いられる特徴量である．その算出フロー概要を図 2-8 に示す.

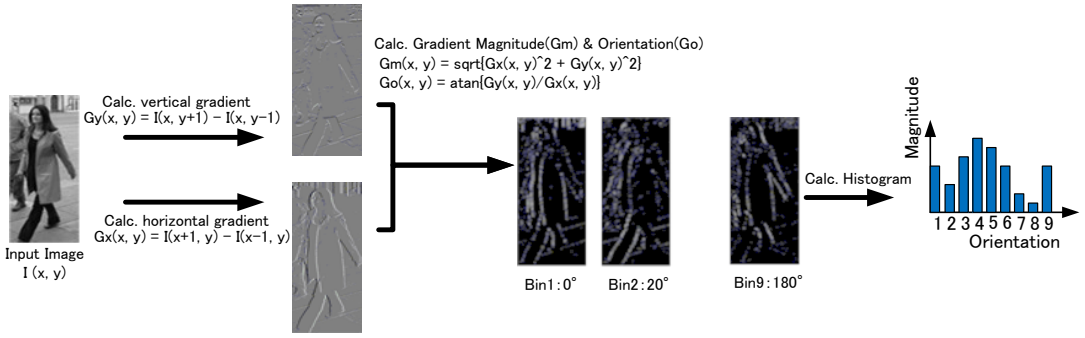


図 2-8 HOG 特徴量の抽出例

はじめに，入力画像 Iから垂直勾配画像 G_y及び水平勾配画像 G_xを各々求め，算出された各々の勾配画像から勾配強度 G_mと勾配方向 G_oを算出した後，勾配方向 0～180 度を 20 度ごとの 9 つの方向に分割し各々の方向での勾配強度ヒストグラムを求める．人物検出などでは入力画像を格子状のセルと呼ばれる小領域に分割し，そのセルごとに HOG 特徴量を算出した後，数個のセルで構成されるブロック単位で HOG 特徴量を連結，正規化しブロックごとの正規化された HOG 特徴量を求め，最終的には全ブロックの正規化された HOG 特徴量を連結して人物検出のための特徴量として用いる場合が多く，この場合の次元数は数千のオーダーとなる．

2.2 識別手法

入力画像内の任意の大きさ，位置に配置された注目領域から抽出された特徴量を用いて対象物を識別するために種々の識別手法が提案・研究されている．以下に本論文に関連する識別手法について簡潔に説明する．

2.2.1 SVM（Support Vector Machine）

SVM [1] [2] [3]はノンパラメトリックな統計的学習手法の一つで, その特徴は識別器の学習時にマージン最大化を基準とし，学習サンプルに対して最適な識別平面が一意に求まることである．マージン最大化とは，学習データの中で他のクラスに最も近いデータ間のユークリッド距離を最大にするように識別平面を設定することで，このとき識別平面を構成するのに選ばれた学習サンプルをサポート・ベクタと呼ぶ.

SVM の識別関数は入力ベクトルを$\mathbf{x}$, 学習により得られた重みベクトルを$\mathbf{w}$, 閾値を b とすると式(2.4)のように表わされる．

$$y = sign(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b), \quad sign(u) = \begin{cases} 1 & u \geq 0 \\ -1 & otherwise \end{cases} \tag{2.4}$$

今，n個の学習サンプル $\mathbf{x}_{\mathrm{i}}\,(\mathrm{i} = 1, \ldots \mathrm{n})$ と、各々のサンプルに対する -1または $+1$なる正解ラベル $t_{\mathrm{i}}\,(\mathrm{i} = 1, \ldots, \mathrm{n})$ が与えられているとする．仮に学習サンプルが全て線形分離可能であるならば，式(2.5)を満たすようなパラメータ $\mathbf{w}$及び bが存在する．

$$t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 \quad (i = 1, \ldots, n) \tag{2.5}$$

これは 2 枚の超平面（$\mathbf{w}^{\mathrm{T}}\mathbf{x}+b=1$及び$\mathbf{w}^{\mathrm{T}}\mathbf{x}+b=-1$）によって学習サンプルが完全に分離されていることを示しており，またこのとき，2 枚の超平面と識別平面との距離（＝マージン）は $\frac{1}{\|\mathbf{w}\|}$となる．従って、このマージンを最大とするパラメータ $\mathbf{w}$と bを求める問題は式(2.6)のように定義できる．

$$\begin{aligned} \min\ \ & L(\mathbf{w})=\frac{\|\mathbf{w}\|^2}{2} \\ \text{subject to}\quad & t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+b)\geq 1 \quad (i=1,...,n) \end{aligned} \tag{2.6}$$

この最適化問題を解くために，ラグランジュ乗数 $\alpha_i \geq 0$ (i = 1, ... , n) を導入すると、目的関数 Lは式(2.7)で置き換えることができる．

$$L(\mathbf{w},b,\alpha)=\frac{\|\mathbf{w}\|^2}{2}-\sum_{i=1}^{n}\alpha_i\left\{t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+b)-1\right\} \tag{2.7}$$

ここで，式(2.7)を $\mathbf{w}$, b 各々について偏微分した後，それらを 0 に等しいとすると式(2.8)に示す 2 つの条件が得られる．

$$\begin{aligned} \mathbf{w}&=\sum_{i=1}^{n}\alpha_i t_i\mathbf{x}_i \\ 0&=\sum_{i=1}^{n}\alpha_i t_i \end{aligned} \tag{2.8}$$

これらの条件式を式(2.7)に代入すると結局式(2.9)に示すような αに関する双対問題として目的関数 $L(a)$ を最大化すればよい．

$$\begin{aligned} \max\ \ & \mathrm{L}(\alpha)=\sum_{i=1}^{n}\alpha_i-\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j t_i t_j\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j \\ & \text{subject to}\ \sum_{i=1}^{n}\alpha_i t_i=0\ \ (\alpha_i\geq 0) \end{aligned} \tag{2.9}$$

結果，得られるサポート・ベクタ集合を S としたとき，サポート・ベクタとなる非 0 のラグランジュ乗数 $\alpha_i > 0$ (i ∈ S) に対応した学習サンプル $\mathbf{s}_i$ (i ∈ S)を用いてパラメータ $\mathbf{w}$は式(2.10)により求まる．

$$\mathbf{w} = \sum_{i \in S} \alpha_i t_i \mathbf{s}_i \tag{2.10}$$

一方，パラメータ b はサポート・ベクタが 2 つの超平面上のどちらかにあるということを利用して任意のサポート・ベクタ $\mathbf{s}_j$ (j ∈ S) から式(2.11)により一意に求まる．

$$b = \mathbf{w}^{\mathrm{T}} \mathbf{s}_j + t_j \tag{2.11}$$

結局式(2.4)の識別関数はサポート・ベクタ $\mathbf{s}$と各サポート・ベクタに対する係数 αと正解ラベル tを用いて式(2.12)のように表現できる．

$$y = sign(\sum_{i \in S} \alpha_i t_i \mathbf{s}_i^{\mathrm{T}} \mathbf{x} + b) \tag{2.12}$$

SVM のもう一つの大きな特徴として，カーネル・トリックを使うことで線形のみならず非線形分離の問題へも対応できるということがあげられる. カーネル・トリックでは，入力ベクトルを非線形写像し，写像された空間での線形分離を考える．このとき，個々のベクトルを非線形写像するのではなく，非線形写像空間での内積をカーネル関数で置き換えることで，高次元ベクトルの計算が不要となり，簡単な計算で非線形分離の問題に対応できる．カーネル関数は問題に応じて様々なものが提案されているが，本論文では式(2.13)で表わされるガウス・カーネルを用いた．

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{\left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2}{2\sigma^2}) \tag{2.13}$$

また，カーネル・トリックを用いた非線形 SVM の識別関数は，学習によって得られるサポート・ベクタ集合をS，サポート・ベクタを $\mathbf{s}_i$ (i ∈ S)，各サポート・ベクタに対応

した係数及び正解ラベルを各々 $\alpha_{i,} t_i$ (i ∈ S)とすると，式(2.14)のように表わされる．

$$y = sign(\sum_{i \in S} \alpha_i t_i K(\mathbf{s}_i, \mathbf{x}) + b) \tag{2.14}$$

ここで本論文では，SVM の学習に研究用途で広く使われている libSVM [16]を用いた．

2.2.2 AdaBoost

AdaBoost [17]は Adaptive Boosting の略で，その名が示すように Boosting と呼ばれる統計的学習手法の一種である．Boosting とは複数の弱識別器から構成される強識別器を学習する集団学習アルゴリズムの一つで，AdaBoost では特に学習の過程で学習サンプルに対する重みを適応的に更新しながら強識別器を構成する弱識別器の選択を行う．

今，n個の学習サンプルxとその正解ラベル y (= 0 又は 1) の組 (x_1, y_1), (x_2, y_2), … , (x_n, y_n) が与えられた場合，その学習ステップは以下の通り．

1. 学習サンプルに対する重み w を初期化する

$$\mathrm{w}_1(\mathrm{i}) = \frac{1}{n} \qquad (i = 1,2, \dots, n)$$

2. For t = 1,2, …, T

2-1.各学習サンプルに対する重みを正規化する

$$\mathrm{w}_\mathrm{t}(\mathrm{i}) \leftarrow \frac{w_t(i)}{\sum_{j=1}^{n} w_t(j)}$$

2-2.t 回目の試行における各弱識別器のエラー率を求める

$$\epsilon_\mathrm{j} = \sum_{i=1}^{n} w_i |h_j(x_i) - y_i|$$

2-3.最もエラー率ϵ_tの低かった弱識別器$\mathrm{h_t}$を選択する

2-4.各学習サンプルに対する重みを更新する

$$\mathrm{w}_{\mathrm{t}+1}(\mathrm{i}) = \mathrm{w}_\mathrm{t}(i)\beta_t^{1-e(i)}$$

$$\text{where} \quad \mathrm{e(i)} = \begin{cases} 0 & h_t(x_i) == y_i \\ 1 & Otherwise \end{cases}$$

$$\text{and} \quad \beta_\mathrm{t} = \frac{\epsilon_\mathrm{t}}{1 - \epsilon_\mathrm{t}}$$

3. 最終的な強識別器 H は以下で定義される

$$H(\mathrm{x}) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & Otherwise \end{cases}$$

$$\text{where} \quad \alpha_t = \log \frac{1}{\beta_t}$$

AdaBoost では，このようにして学習された複数の重み付けされた弱識別器 h から構成される強識別器 H を用いて識別を行う．

2.2.3 SVR（Support Vector Regression）

SVR [4] は SVM を回帰へと拡張したもので，SVM 同様カーネル・トリックにより非線形問題への対応が可能であり，また未知のデータに対する高い汎化能力も期待できる．本論文では SVR としてε-SVR を用いた．ε-SVR では式(2.15)のように，ある入力 $\mathbf{x}$ が与えられたときに線形関数 $f(x)$ と観測値 y との誤差 r が事前に与えられた ε 以下の場合には損失を 0，ε を超えた場合のみ損失とみなした上で，この誤差 r をモデルの複雑度を考慮しつつ最小にする関数 $f(x)$ を推定する．

$$\begin{aligned} r &= \max(0, |f(\mathbf{x}) - y| - \varepsilon) \\ &\text{where} \quad f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b \end{aligned} \tag{2.15}$$

そのためには今，n 個のサンプル $\mathbf{x}_{\mathrm{i}}\ (\mathrm{i} = 1, \ldots \mathrm{n})$ と、各々のサンプルに対する観測値 $y_i\ (\mathrm{i} = 1, \ldots, \mathrm{n})$ が与えられたとすると，式(2.15)により算出される誤差 r を予測値が観測値を上回っている場合に ξ，下回っている場合に $\hat{\xi}$ とおくと，式(2.16)で表される最小化問題を解けばよい．ここで第 2 項はモデルの複雑度を表現し，第 1 項の係数 C はモデルの複雑度と誤差の大きさとのバランスを調整するパラメータである．

$$\begin{aligned} \min \quad & L(\mathbf{w}) = C\sum_{i=1}^{n}(\xi_i + \hat{\xi}_i) + \frac{\|\mathbf{w}\|^2}{2} \\ \text{subject to} \quad & (\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - y_i \leq \varepsilon + \xi_i \\ & y_i - (\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \leq \varepsilon + \hat{\xi}_i \\ & \xi_i, \hat{\xi}_i \geq 0 \ \ (i = 1, \ldots, n) \end{aligned} \tag{2.16}$$

これは SVM 同様にラグランジュ乗数を導入すると結局，式(2.17)で表される最大化問題

に帰着する．ここで，αは予測値が実測値を上回る場合に対するラグランジュ乗数で，$\hat{\alpha}$は予測値が実測値を下回る場合に対するラグランジュ乗数である．

$$\begin{aligned}\max\ L(\alpha,\alpha^*) = &-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat{\alpha}_i-\alpha_i)(\hat{\alpha}_j-\alpha_j)\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j \\ &-\varepsilon\sum_{i=1}^{n}(\hat{\alpha}_i+\alpha_i)+\sum_{i=1}^{n}(\hat{\alpha}_i-\alpha_i)y_i \\ \text{subject to}\quad &0\le\alpha_i\le C\quad(i=1,...,n) \\ &0\le\hat{\alpha}_i\le C\quad(i=1,...,n)\end{aligned} \tag{2.17}$$

結果，関数 $f(x)$ は式(2.17)の問題を解くことによって求まるサポート・ベクタ集合を S とすると，非 0 なるラグランジュ乗数 α_i, $\hat{\alpha}_i > 0$ (i ∈ S) とそれに対応したサンプル $\mathbf{s}_i$ (i ∈ S) を用いて式(2.18)のように定義でき，また SVM 同様にカーネル・トリックを用いた場合の非線形 SVR は式(2.19)となる．

$$f(x)=\sum_{i\in S}(\hat{\alpha}_i-\alpha_i)\mathbf{s}_i^{\mathrm{T}}\mathbf{x}+b \tag{2.18}$$

$$f(x)=\sum_{i\in S}(\hat{\alpha}_i-\alpha_i)K(\mathbf{s}_i,\mathbf{x})+b \tag{2.19}$$

ここで，Kはカーネル関数であり SVM 同様に様々なカーネルを用いることができるが本論文では式(2.13)で表されるガウス・カーネルを用いた．また，本論文では SVM と同様に SVR の学習にも libSVM [16]を利用した．

第3章　顔及び顔器官の検出とそのドライバ姿勢推定への応用

3.1　はじめに

ITS（Intelligent Transport Systems） [18]の研究開発は，2009 年 2 月に官民共同で公道における公開デモンストレーション [19]が実施されるなど実用化に向け着実にその歩みを進めつつある．中でも，UTMS（Universal Traffic Management Systems） [20]を中心に研究開発が進められている DSSS（Driving Safety Support Systems） [21] では，インフラ協調型システムによりドライバが安全に運転できるよう支援することを目指している．DSSS が広く整備されると，例えばドライバから視認困難な自動車，二輪車，歩行者等を各種センサで検出し，その情報を交通情報板などに表示したり，路車間通信により車載装置に提供したりするなどして，注意を促すようなサービスの提供が想定される．これにより，一般道における見通しの悪い交差点等での交通事故の減少，ドライバの判断負荷の軽減などの効果が期待されている．近年の自動車乗車中の死者数を年齢層別にみると，高齢者が 4 割以上（構成率 43.3 %）を占め最も多く，次いで若者（同 12.3 %），50 歳代（同 10.8 %）の順に多い [22]．このような状況を踏まえると，若年層よりも視聴覚能力/情報処理/状況判断/身体反応速度が衰えてきている高齢層に対しても有用な注意喚起を行えるシステムが望まれる．しかし，走行路周辺の様々な情報が収集できてそれを伝達可能となる一方で，ドライバへ伝わる情報が溢れることが予想される．そのような状況下において，例えば収集した全ての情報をドライバへ伝えてしまうと，ドライバがその情報を処理できないだけでなく，逆に本来集中すべき運転行動に対しても好ましくない影響を与えることが懸念されるため，それを解決するには伝達情報の収集と併せて，真に伝えるべき情報か否かの取捨選択手段が必要不可欠であると考えられる．つまり, 実用的なサービスにするためには, 単に取得された情報全てを提示するのではなく，ドライバの運転状態をリアルタイムに把握した上で必要な時に必要な情報だけを提示するなど実運用性・利便性の面での課題を解決する必要があるといわれている．

ドライバの運転状態把握とは，例えば運転中のドライバの顔姿勢，視線の検出やドライバの集中力，注意力や眠気の検知 [23] [24]などである．中でも，顔姿勢や視線が検出できれば，交差点等での左右確認の有無，道路上の障害物や歩行者を確認したか否かなどを把握することができ，DSSS により取得可能となる時々刻々と変化する走行路周辺状況と協調することで，ドライバの運転状態に応じた適応的な情報提示が可能となり，ユーザに快適な道路インフラ・システムの実現に大きな役割を果たすと考えられる．

以上のことから，本研究ではドライバの安全確認行動に応じて適切な支援を行う安全運転支援システムの実現を目指し，運転中のドライバの顔姿勢推定に着目した．ドライバの顔姿勢を推定するためには，姿勢センサを装着する方法や，特殊なマーカーなどを付けてカメラで撮影する方法 [25]，ステレオ・カメラなど複数のカメラを用いる方法 [26]があるが，このように何らかの接触デバイスをドライバに装着させる方法や複数のカメラを用いる方法は運用面，コスト面での障壁がある．そこで，本研究では，ドライバの負担がなく，実用的で安価なシステムの実現を目指し，ドライバの正面ダッシュボード上に設置した単眼の車載カメラからの映像のみで運転中のドライバの顔姿勢を推定することを検討した（図 3-1 参照）．

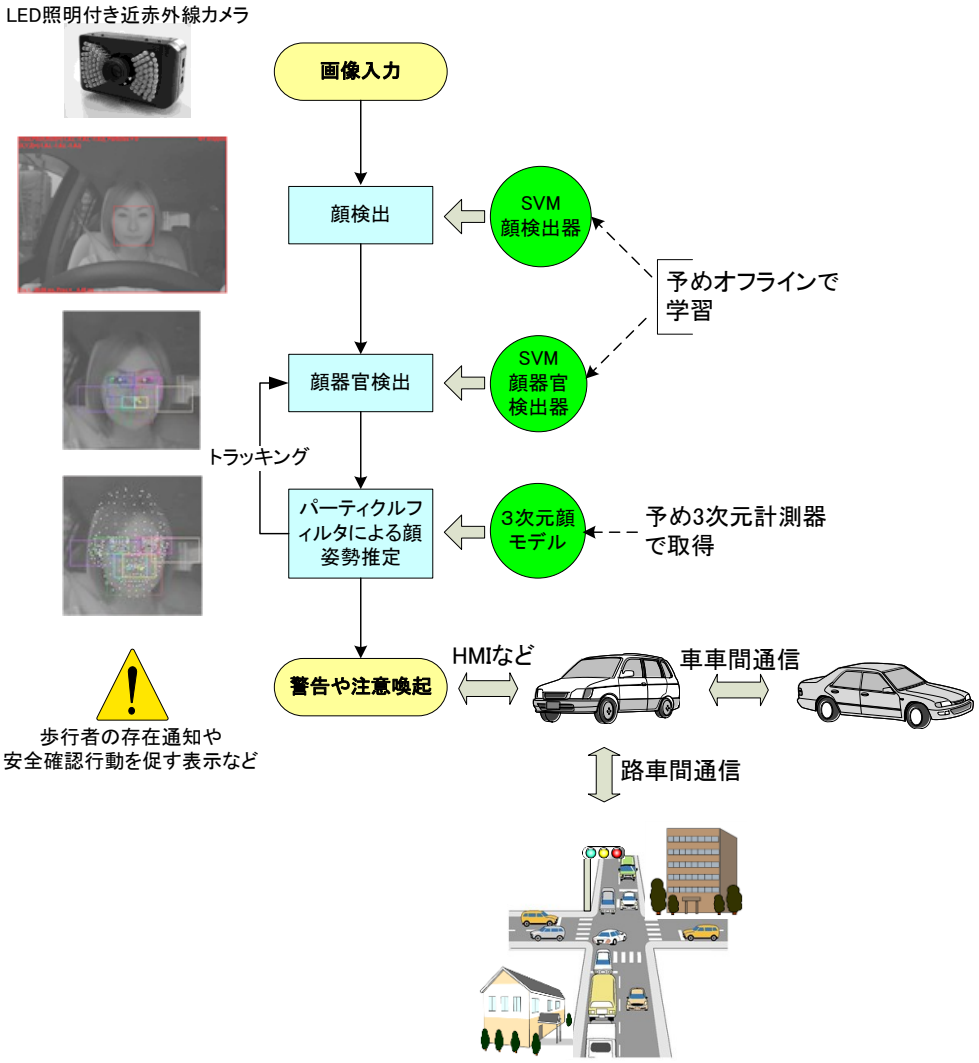


図 3-1　提案手法を利用した DSSS サービス・イメージ

本研究では，まず市販の顔姿勢計測システムを用いて運転中のドライバの安全確認行動に必要な顔姿勢の検出範囲を明らかにした後,顔及び顔器官検出のアルゴリズム及び,3 次元顔モデル，パーティクル・フィルタ [27] [28] [29]を用いた顔姿勢推定のアルゴリズムについて説明し,これら提案手法を各種評価映像に適用した結果について紹介する．なお，検出器の学習や評価実験には，研究所員数名の顔映像データ，財団法人ソフトピアジャパンから使用許諾を受けた顔画像データベース[1], 人材派遣会社を介して謝金雇用された 20～60 歳代の健常成人男性 33 名,同女性 33 名の教習所内での周回運転中の顔映像データなどを用いた．被験者の公募に際しては，日常的に運転をしていることを条件とし，実験への参加は，実験内容の説明を受けての自由意志とした．また，実験当日に再度，実験の内容や危険性，個人情報の扱い，途中棄権の権利などを説明した後，書面による参加意思の確認を行った．

[1] 本章に使用した顔画像データの一部は，財団法人ソフトピアジャパンから使用許諾を受けたものです．権利者に無断で複写，利用，配布等を行うことは禁じられています．

3.2　　安全確認行動

警察庁の統計 [22]では，死亡事故件数を道路形状別にみると，市街地の交差点（構成率 32.0 %）が最も多く，次いで非市街地の単路（同 31.3 %），市街地の単路（同 19.2 %）の順に多く発生している．そこで本研究では，ドライバの安全確認行動の中でも特に，交差点における安全確認行動に着目した．

まず信号のない交差点での安全確認の有無を判定する上で必要な顔姿勢の検知範囲を明らかにすべく，20～40 才台の被験者 5 人を対象とし，1 人当たり指定の市街地コースを 5 周走行する走行実験を実施した．そして，周回コース内の以下に示す 4 つの交差点における顔姿勢の変化を調査した．

【交差点 1】見通しの良いＴ字路における右左折
【交差点 2】見通しの良い十字路における右左折
【交差点 3】見通しの悪いＴ字路における右左折
【交差点 4】見通しの悪い十字路における右左折

なお運転中の顔姿勢は，予め被験者の顔にマーカーを付けた状態で，複数台のカメラにより同期撮影した後にオフラインで算出することが可能な SmartEye 社 [25]の計測システムを利用した（図 3-2 参照）．

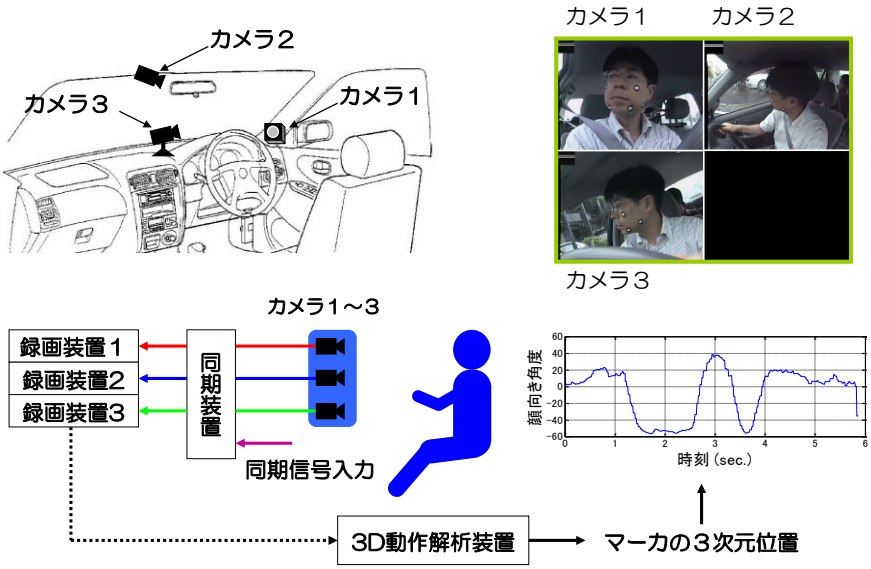


図 3-2　運転中の顔姿勢計測システム

本計測システムを利用して全被験者の全運転試行における顔姿勢の変化を調査した結果を図 3-3 に示す. この結果より, 顔の Yaw 方向の回転角度については左右 80 度程度, 前後方向の移動については，前方向に 300 mm 程度，後ろ方向に 50 mm 程度，左右方向については 200 mm 程度の移動があることがわかった.

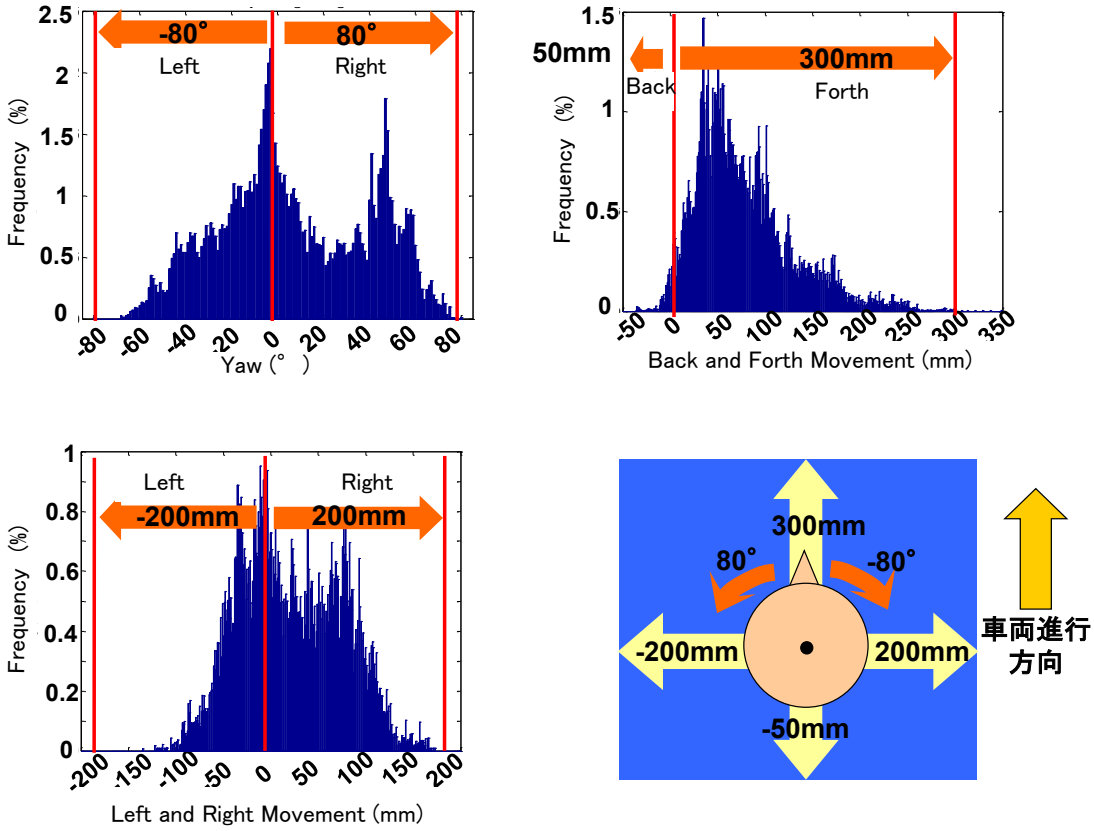


図 3-3　運転中のドライバの顔姿勢変化の調査結果

次に，道路に応じたドライバの安全確認行動を明らかにするため，交差点における安全確認行動と左折時における安全確認行動に着目した．図 3-4 の左に交差点 1 での，右に交差点 4 での顔姿勢変化の調査結果を各々示す．

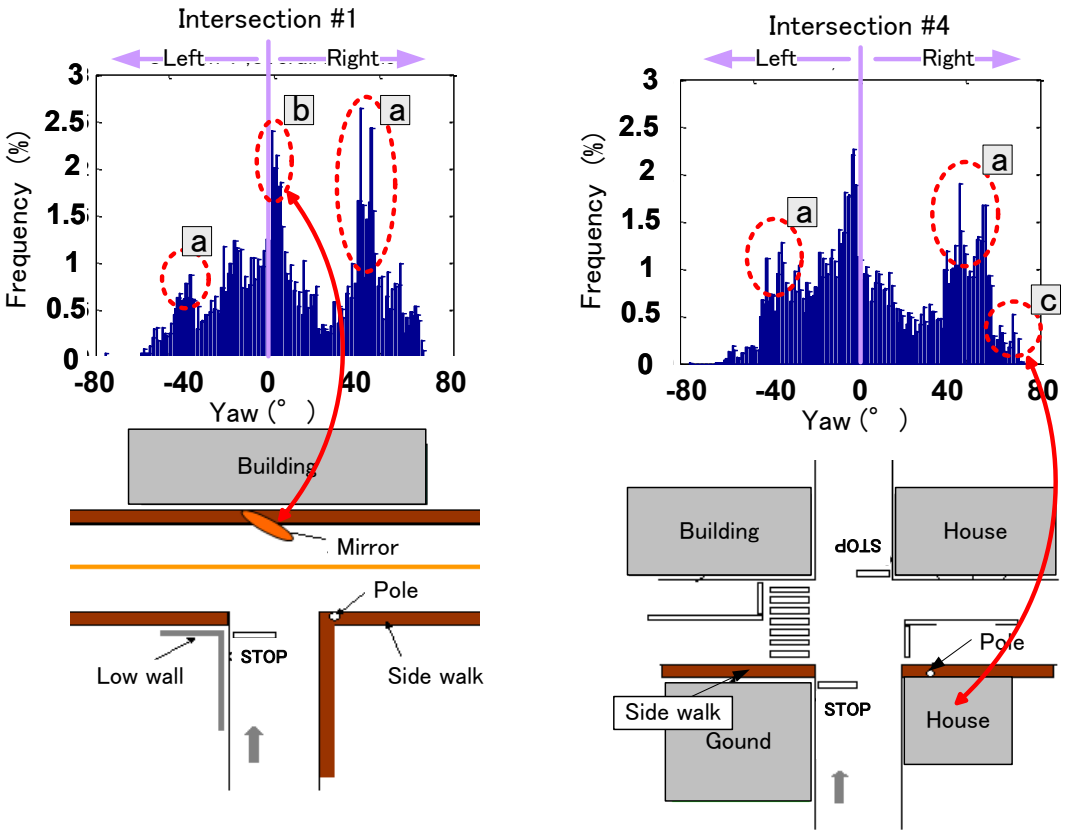


図 3-4　交差点１，４での顔姿勢変化の調査結果

どちらの交差点でも，左右確認に伴って±45 度前後にピークがあることが見て取れる（図中【a】）．また，交差点 1 に着目すると交差点にミラーがあるときに，その方向への顔向き頻度が高くなっていることがわかる（図中【b】）．また，右図より，家や壁などの遮蔽がある場合には，顔向きの最大角が大きくなることが読み取れる（図中【c】）．これらの実験を通して，信号のない交差点での安全確認行動では，Yaw 方向に最大で±80 度程度もの顔姿勢変化が生じることがわかった．

最後に，右左折時の顔姿勢変化を確認するため，左折時の安全確認行動に着目し顔姿勢の変化を調査した結果を図 3-5 に示す．

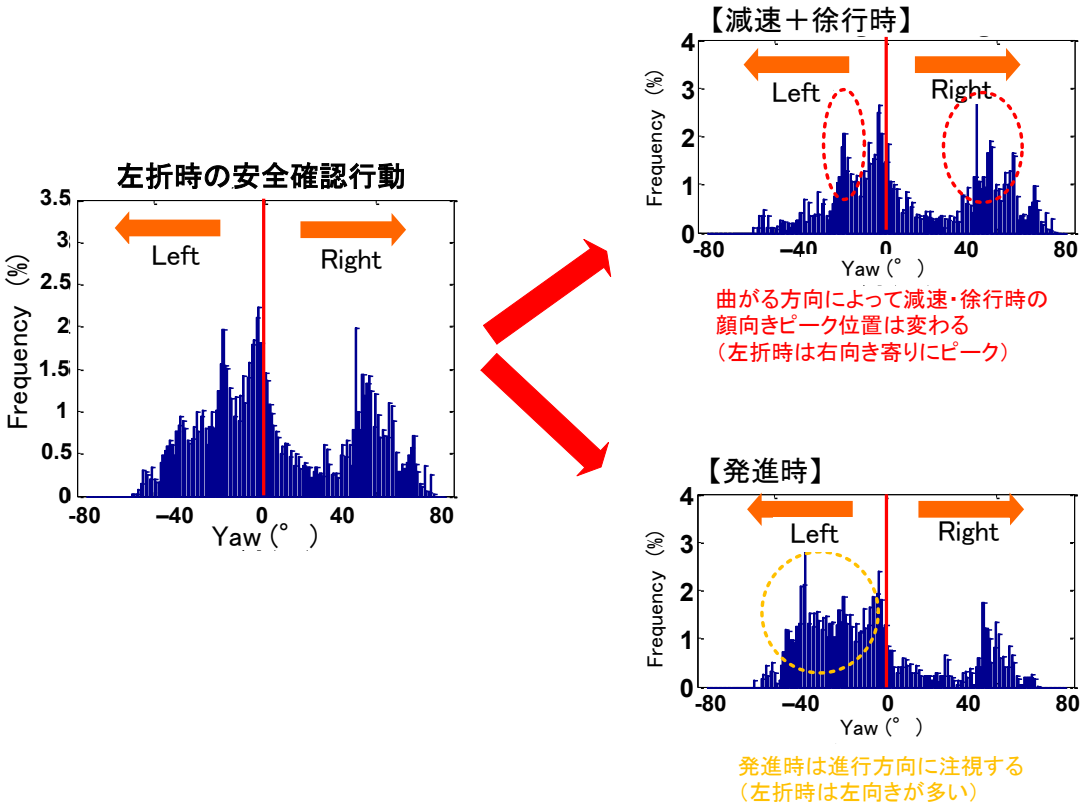


図 3-5　左折時の安全確認行動における顔姿勢変化の調査結果

図 3-5 の左図に着目すると，右左折時には曲がる方向への顔姿勢変化が多くなっていることがわかる．また図 3-5 の右上図及び右下図に着目すると，減速・徐行時には進行方向と逆方向（左折時には右側，右折時には左側）への顔向き変化が大きくなり，逆に発進時には，進行方向と同じ方向（左折時には左側，右折時には右側）への顔向き頻度が多くなることがわかった．

これら運転中のドライバの顔姿勢の調査結果から交差点での左右確認，右左折時には顔角度が大きく変化することが分かり，また右折か左折か，減速中は発進中かによってもドライバの安全確認行動が異なるため，道路状況と共に運転行動にも応じて適切な安全確認行動がなされたか否かを判断する必要性があると考えられる．

3.3　　ドライバの顔及び顔器官の検出

本研究では，ドライバの映像を昼夜問わず安定して撮影するために，車載カメラとして図 3-6 に示すような LED 照明付きの近赤外線カメラを試作し，走行中のドライバの顔映像の撮影に利用した．

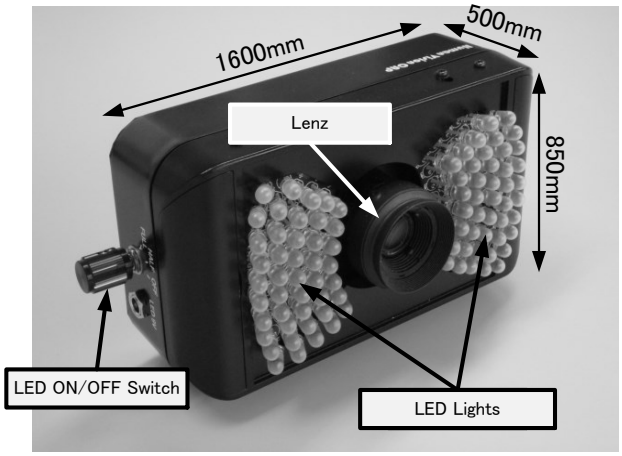


図 3-6　LED 照明付き近赤外線カメラ

赤外線は可視光に比べて波長が長いため散乱しにくい性質があり，煙や薄い布などを透過して向こう側の物体を撮影するために用いることができる．ただし，あくまで光であるため近赤外線光が当たっていない物体は写らず認識できず，また得られる映像はカラーではなくモノクロとなる．一方で，赤外線は目に見えないため，外部に近赤外線光源を持つことで被写体に気付かれることなく夜間などでも撮影することができるメリットがある．　本試作カメラで撮影した運転中のドライバの映像サンプルを図 3-7 に示す．昼間に比べ夜間はやや画像全体に明度が低いものの，近赤外線 LED 照明により昼間，夜間を問わずドライバの顔映像が撮影できることが確認できる．

図 3-7　近赤外線カメラによる撮影映像サンプル

本研究ではこの近赤外線カメラをドライバ撮影用の車載カメラとし，ダッシュボード上に設置した本カメラより取得される 256 階調グレイ・スケールの映像からドライバの顔姿勢を推定すること検討した．

撮影されたカメラ映像からドライバの顔姿勢を推定するためには，はじめに取得された映像の中からドライバの顔を検出し，更に検出された顔の中から姿勢を推定するために必要な顔器官を検出する必要がある．そして，それらの検出された顔器官の位置関係から顔姿勢を推定することを考える．本研究では，顔及び顔器官の検出器として，最も性能の優れた識別器のひとつとして知られている SVM（2.2.1 節）を用いた．

3.3.1 顔検出器

車載カメラ自体は固定設置されているものの，走行路の周辺環境によって取得される画像には背景の変化や激しい照明変動も生じる．そのため，ドライバの顔検出器には，不特定人への対応は勿論のこと，照明変動へのロバスト性も要求される．また，3.2 節で述べたように交差点での安全確認行動では，顔姿勢が Yaw 方向に大きく変化するため，取得される画像内でのドライバの見た目は大きく変化する．そのため，Yaw 方向に±75 度，Pitch 方向に±15 度の顔向きが既知である顔画像(図 3-8 参照)と，実環境での性能を考慮して教習所内での実走行映像とを合わせた広角度範囲の顔画像を学習に用いた．学習する顔画像は，計算量削減のためにサイズを 10×10 画素に縮小した後，背景や髪型の影響を少なくするために左右上段，左右下段の一定領域にマスキング処理を施した後に正規化輝度値（2.1.1 節）を特徴量として抽出する．

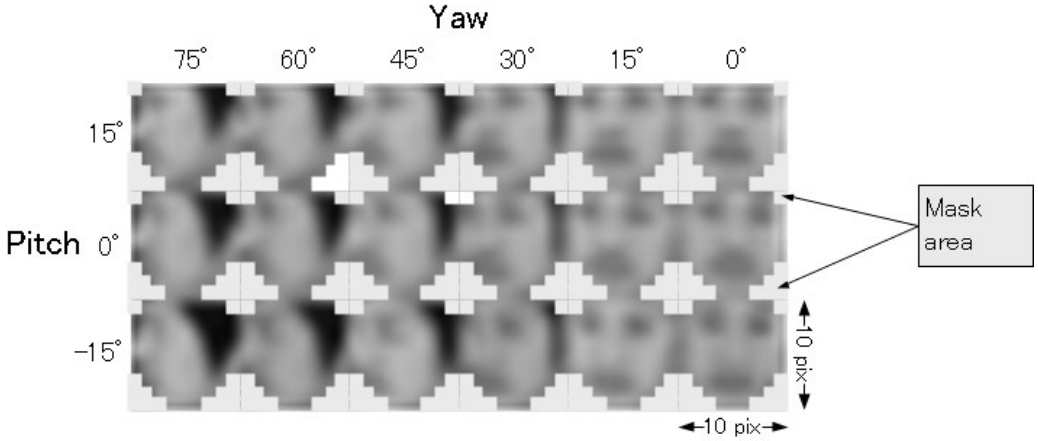


図 3-8　学習顔画像の角度，サイズとマスキング領域の例

3.3.2 顔器官検出器

顔姿勢を精度よく推定するためには，できるだけ多くの顔器官を正確に検出することが望ましい．本研究では，正面向き時の安定性と広角度範囲での検出を実現するため，左右の目，目頭，目尻，鼻孔，口唇端，耳の計 12 点を検出すべき顔器官と定義した．耳については，髪形による隠れの問題があり不特定人に対して安定して検出されるかどうかという問題はあるものの，特に横向き時の顔姿勢推定性能を向上させるために，本研究では積極的に利用することとした．図 3-9 に示す位置とサイズで各顔器官を定義し，顔検出器の学習と同様に Yaw 方向に±75 度，Pitch 方向に±15 度の顔向きが既知の顔画像データと実走行映像データを学習サンプルとし，顔検出と同じく正規化輝度値を特徴量として各顔器官の検出器を学習した．

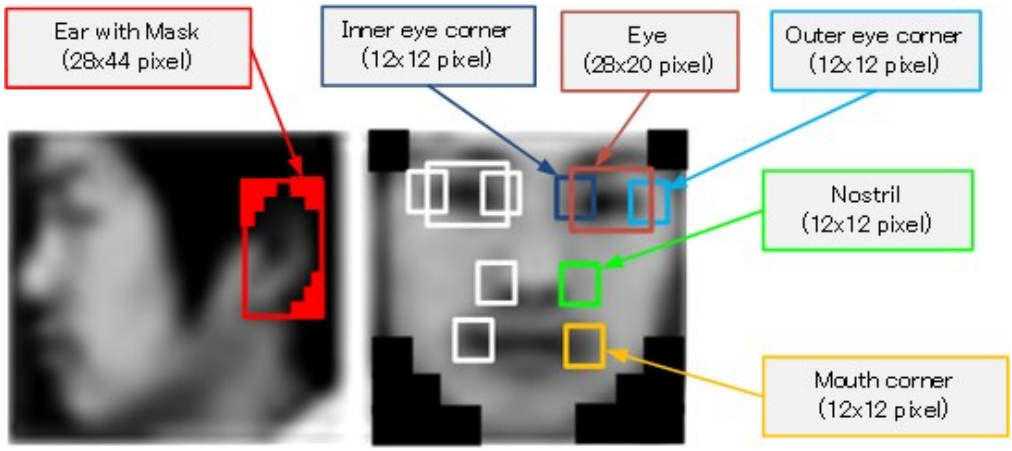


図 3-9　学習した顔器官の位置とサイズ

3.3.3 検出器の高速化検討

SVM は優れた識別性能を示す一方で，学習の結果得られるサポート・ベクタ数が多いと計算コストが非常に大きくなるという欠点がある．そこで本研究では，SVM による顔及び顔器官検出の前処理としての各顔器官の探索エリアの絞込みと，2 段カスケード構造による検出器の高速化を検討した．

顔及び器官探索エリアの絞り込み

顔及び顔器官検出器は，その性能もさることながら実用性まで考慮すると実時間での演算処理の実現は重要な要素の 1 つとなる．特に，適用するアプリケーションによっては，映像のフレームレートとして主流である 30 fps 又はそれ以上での動作を保障しなけ

ればならない場合も考えられる．しかし，一般的に SVM 等のノンパラメトリックな統計的学習手法を用いて構成された識別器は，学習サンプルを増やせば増やすほど，学習によって得られる学習モデル（SVM の場合はサポート・ベクタ数）は巨大化し，その結果，識別処理の演算量が爆発的に増加する．そこで，本研究では，学習結果に依らず検出処理を低減できる方法として，顔器官ごとに探索範囲を必要最小限にするという点に着目し，多方向顔画像 DB の画像とその画像にマニュアルラベリングで付加した顔及び各顔器官（ここでは，左右目と左右口唇端を例とする）の位置情報を元に，ヒストグラムの分布と位置関係による探索エリアの絞込みを検討した．

探索エリア絞込みのためのヒストグラム分布の調査

本研究では，SVM による識別計算の前処理として，より計算量の少ない処理で明らかな非検出対象を予め探索エリアから排除できるような仕組みを検討するため，探索窓内のヒストグラム分布を調査し，ヒストグラム分布の広がり（分布範囲）により対象／非対象の判定を行うことを検討した．ヒストグラムはある一定領域内での輝度値の分布を観測でき，またその領域内での輝度値の配置関係に不変であることから，ある程度顔向きが変化しても得られる値が変化しないことが予想される．このヒストグラムの分布範囲を観測することで，検出対象とかけ離れた分布範囲を持つ領域は予め探索エリアから除外し SVM による検出演算処理を実施しない．例えば，背景における単一色の壁や，顔における前額部や頬のように比較的均一なテクスチャの部位では分布範囲は狭く，一方，木々の葉や目の周辺などはある程度広い分布範囲を持ち，かつそれらの分布は対象によってある程度の共通性があると考えられる．そこで，低解像度化した顔や顔器官のヒストグラムの分布範囲は似通っているという仮説の下，多方向顔画像 DB の顔画像（サブジェクト：300 名）とマニュアルラベリングされた顔と顔器官の位置情報を元に，広角度範囲を想定した水平方向に 0～75 度（15 度ステップ），垂直方向に±45 度（15 度ステップ）の範囲で，各々のヒストグラム分布範囲を調査した．その結果を図 3-10～図 3-13 に示す（横軸はサブジェクト ID，縦軸は分布範囲）．なお，調査実施時の顔，各顔器官の探索窓の大きさは以下の通りである．

- 顔　　：10 × 10 画素
- 目　　：16 × 16 画素
- 口唇端　：8 × 8 画素

図 3-10～図 3-13 に示すように，水平方向に 0 度，垂直方向に-45～45 度の結果を見ると，垂直方向の変化はヒストグラムの分布範囲にあまり影響を与えず，どの顔器官も垂

直方向の角度に依らず似たような分布傾向を持つことがわかる．（なお，分布範囲が 0 となっているサブジェクトはラベリング情報がないか，または対象が髪などで隠蔽しているため調査から除外したものである．）

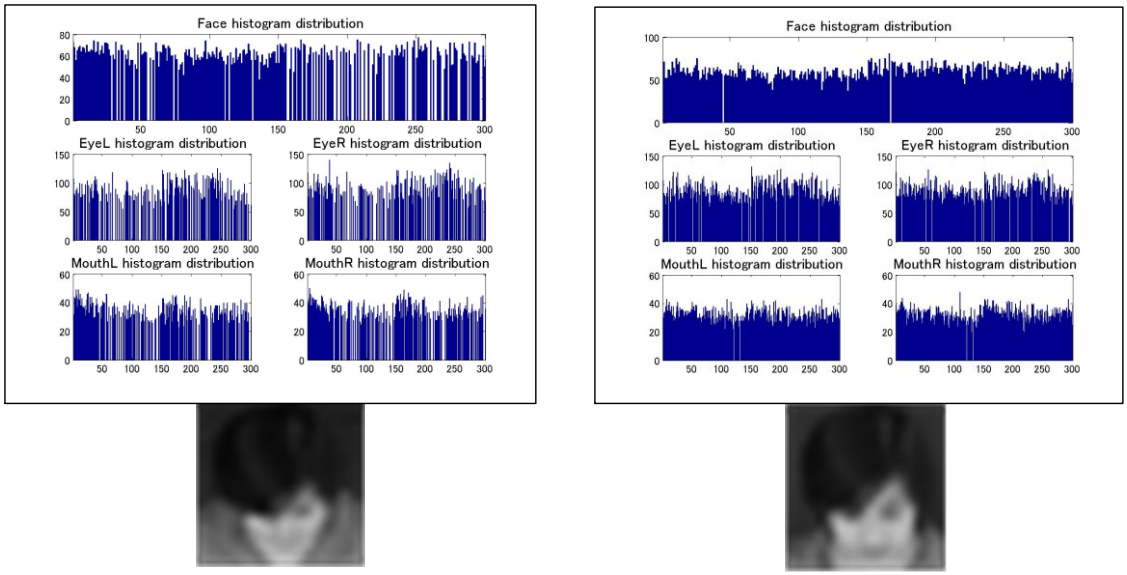


図 3-10　顔及び顔器官のヒストグラム分布範囲（左 : 水平 0 度，垂直-45 度，右 : 水平 0 度，垂直-30 度）

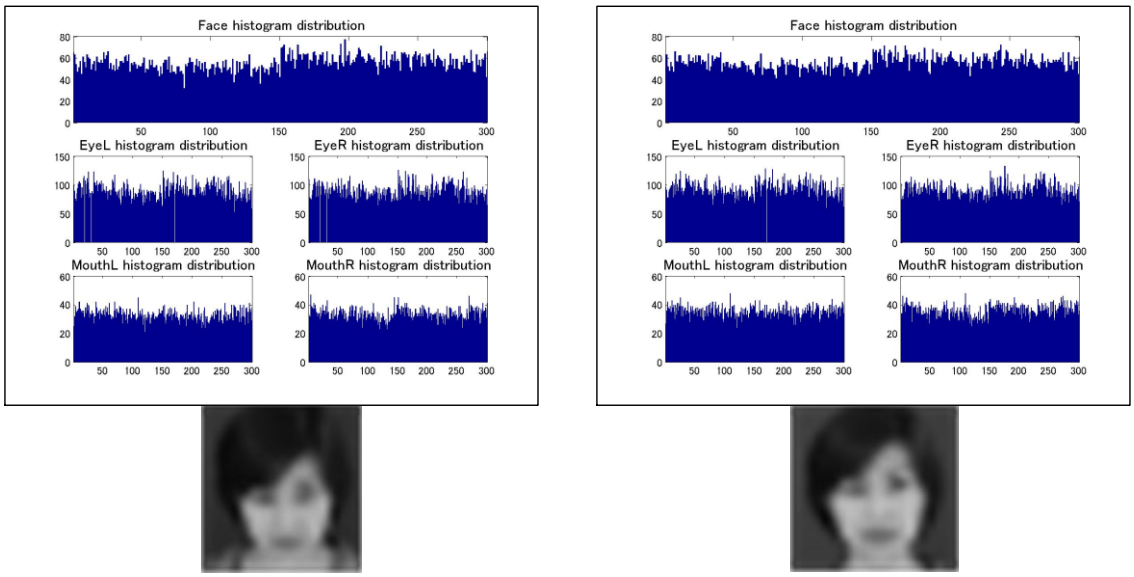


図 3-11　顔及び顔器官のヒストグラム分布範囲（左 : 水平 0 度，垂直-15 度，右 : 水平 0 度，垂直 0 度）

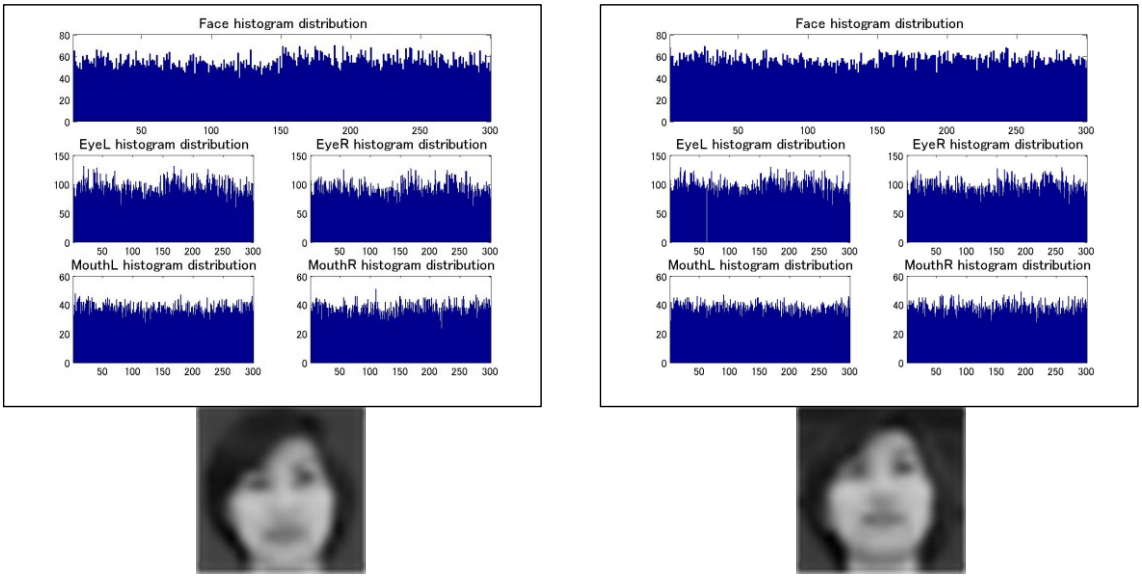


図 3-12　顔及び顔器官のヒストグラム分布範囲（左 : 水平 0 度，垂直 15 度，右 : 水平 0 度，垂直 30 度）

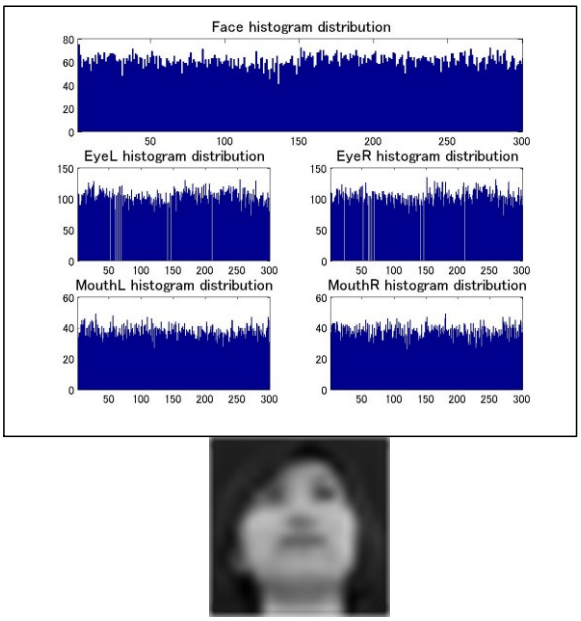


図 3-13　顔及び顔器官のヒストグラム分布範囲（水平 0 度，垂直 45 度）

このようにして，水平方向を順次変化させ，調査範囲を水平方向に 0～75 度に広げた結果の平均，最大・最小値を表 3.1 に示す.

表 3.1　ヒストグラム分布範囲の統計量

	ヒストグラム分布		
	平均	最小	最大
顔	61.77309	32	84
左目	92.48295	47	135
右目	99.69153	56	142
左口端	34.71617	17	53
右口端	37.63275	19	55

ここで顔は左右対称であると仮定し，水平に反対方向の 0～-75 度を考慮すると，結局，水平方向±75 度，垂直方向±45 度の広角度範囲の顔向きにおいては，顔：32～84，目：47～142，口唇端 : 17～55 の範囲を外れるものは SVM による検出演算処理の前段階で探索候補から外すことが可能と考えられる.

顔器官探索エリアの位置と大きさの調査

各顔器官（左右目，左右口唇端）は，顔検出器により入力画像内で顔が検出された場合にのみ探索を実施する．この場合，顔検出の結果を受けて各顔器官をどの範囲で探索すべきかを予め決定しておくことで各顔器官の探索エリアを限定することができる．一般的に考えると，左目や左口唇端は検出された顔の左半分を探索すれば良く，また目は顔の上半分を, 口唇端は顔の下半分を各々探索すれば良いと考えられるが, 本研究では，広角度範囲の顔向きに対してロバストな顔及び顔器官検出器を目指すため，必ずしもこの仮定は成立しない．そこで，多方向顔画像 DB における水平方向に 0～75 度（15 度ステップ），垂直方向に±45 度（15 度ステップ）の顔画像から各顔器官が顔中心に対してどのような配置になりえるかを調査し，各顔器官の探索エリアを定義することとした．例として図 3-14～図 3-17 に水平 0 度における垂直-45～45 度の範囲での 300 人分の各顔器官の分布（緑：右目，赤：左目，紫：右口唇端，黄：左口唇端）を示す．各図は，検出された顔の中心を原点として，各顔器官が人ごとにどのように分布しているかを示している.

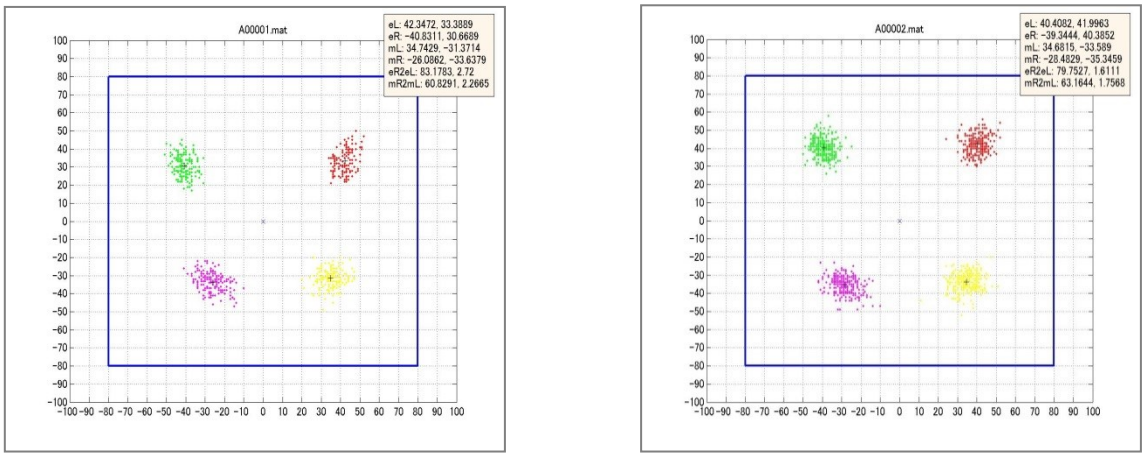


図 3-14　顔器官の分布（左：水平 0 度，垂直-45 度，右：水平 0 度，垂直-30 度）

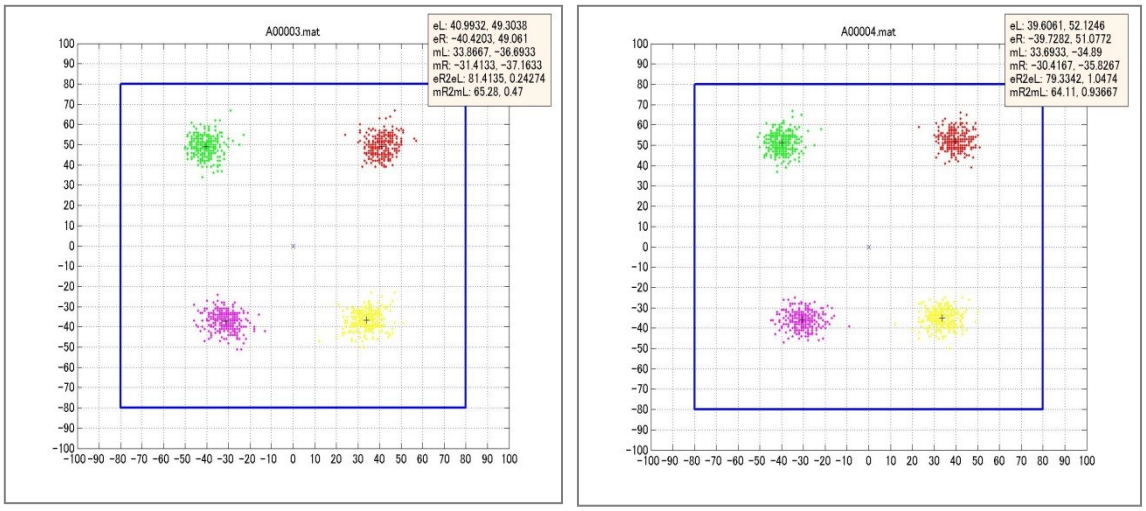


図 3-15　顔器官の分布（左：水平 0 度，垂直-15 度，右：水平 0 度，垂直 0 度）

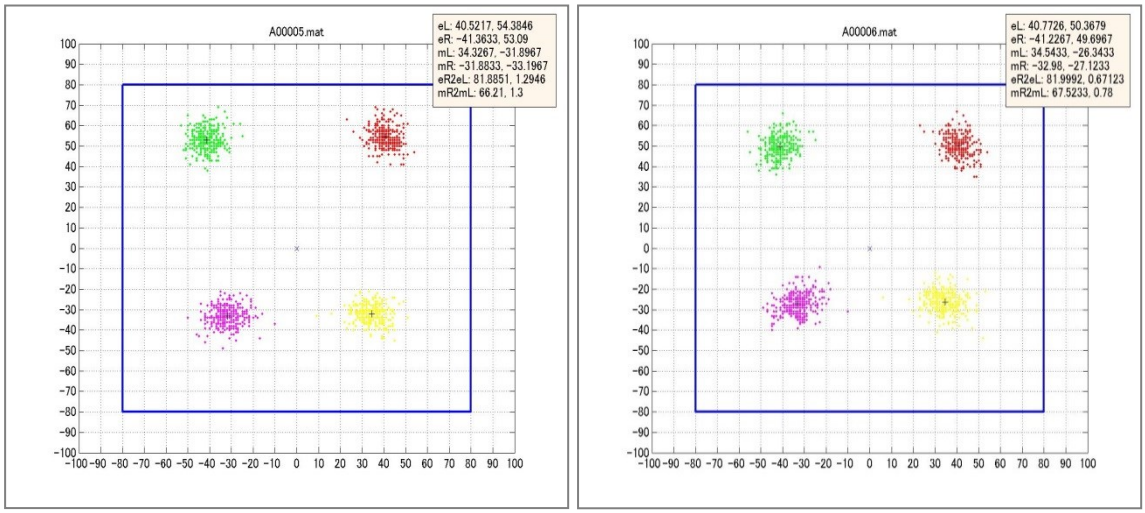


図 3-16　顔器官の分布（左：水平 0 度，垂直 15 度，右：水平 0 度，垂直 30 度）

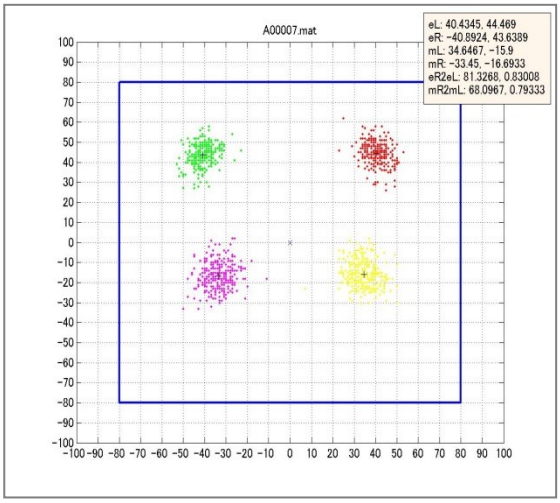
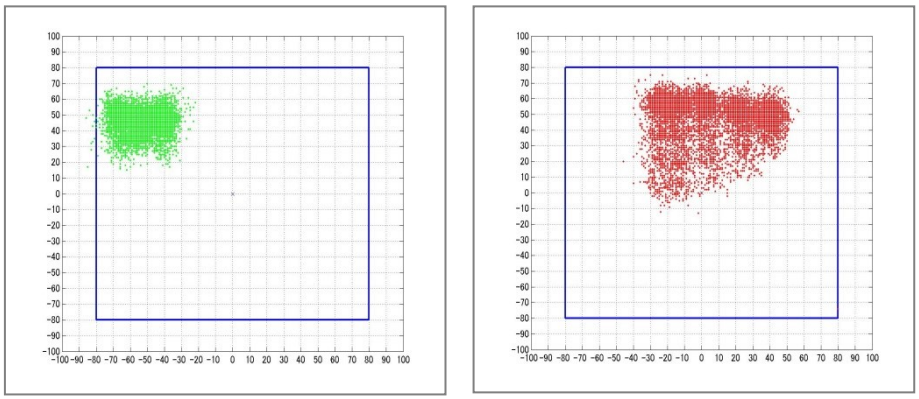


図 3-17　顔器官の分布（水平 0 度，垂直 45 度）

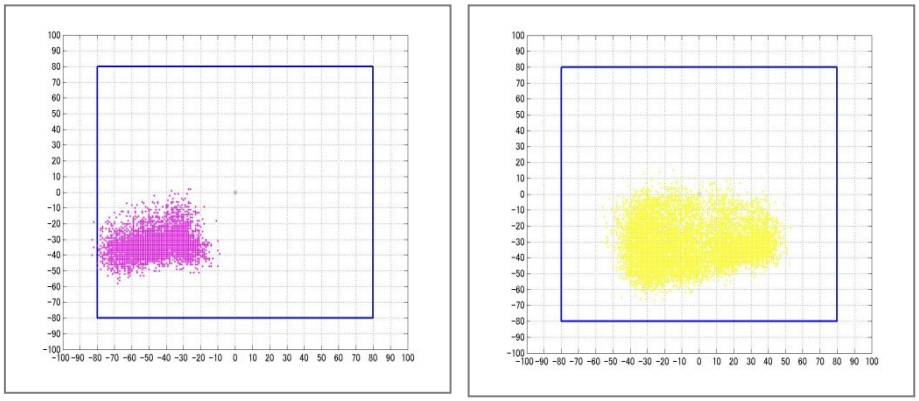
このようにして水平方向を順次変化させ，調査範囲を水平方向に 0～75 度に広げた結局，水平方向に 0～75 度，垂直方向に±45 度の範囲における 300 人の左右目の分布は図 3-18 のようになり，左右口唇端の分布は図 3-19 のようになった.



X:-86 ～-22
Y:+15 ～+70

X:-46 ～+57
Y:-13 ～+75

図 3-18　左右目の分布範囲（左：右目，右：左目）



X:-83 ～-9
Y:-58 ～+2

X:-53 ～+56
Y:-67 ～+15

図 3-19　左右口唇端の分布範囲（左：右口唇端，右：左口唇端）

これらの結果から例えば水平方向に±75 度程度の広角度範囲の顔向きにおいて，各顔器官を検出しようとした場合，必ずしも左目が顔の左半分のみに，右目が右半分のみに存在しないことがわかる．また，左右口唇端についても目と同様の結果が得られた．よって，図 3-18，図 3-19 に示す 0～75 度の調査結果から，水平反対方向の 0～-75 度を考慮すると，図 3-20 に示すように水平方向±75 度，垂直方向±45 度における左目の探索エリアは，X軸方向：-46～83，Y軸方向：-13～75 となり，左口唇端の探索エリアは，X軸方向：-53～83，Y軸方向：-67～15 と定義できる．また，右目，右口唇端については左の各対応器官の探索エリアを左右反転すれば良い．

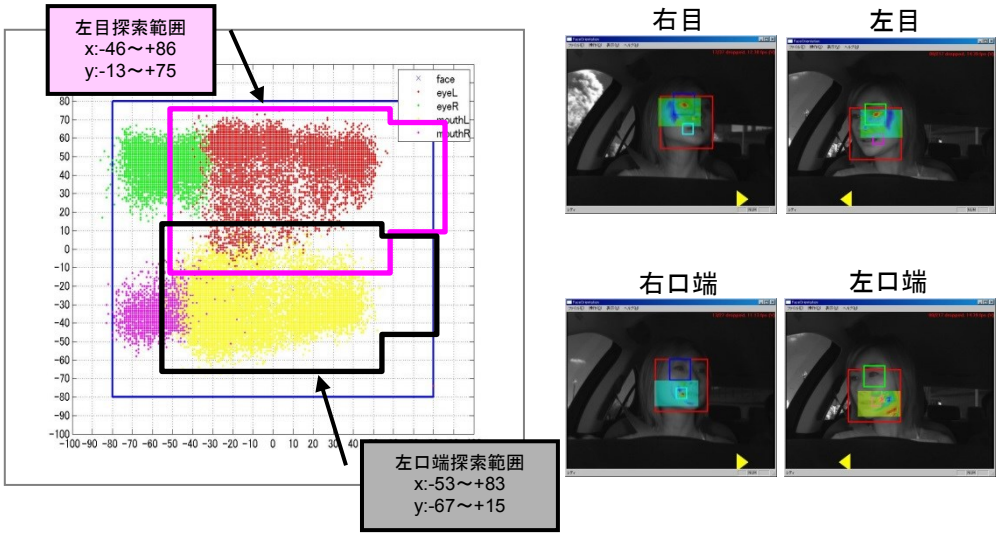


図 3-20　顔器官の探索エリアと探索結果例

カスケード型 SVM 検出器

次にサポート・ベクタ数を削減し演算量を低減することを考える．本研究では検出器を少数のサポート・ベクタからなる副検出器と，学習で得られる全サポート・ベクタからなる主検出器との 2 段カスケード構造とすることで，高検出率と低計算コストを両立することを検討した．以下にカスケード構造化のステップを示す．

1. 全学習サンプルを用いて主検出器を学習
2. 1 で得られた主検出器のサポート・ベクタ数を Reduced Set Method 法 [30] [31]により削減
3. 2 で削減されたサポート・ベクタを新たなサポート・ベクタとする副検出器につい

て,未検出率が主検出器のそれと同程度になるように,副検出器の閾値を調整
4. 3 で得られた副検出器を 1 段目とし, 副検出器により検出された画像についてのみ 2 段目の主検出器による検出処理を実施するように 2 段カスケード構造化

ここで，3 において閾値調整を行うのは副検出器の段階で対象物が検出されず，主検出器に検出処理が渡らないことを極力防ぐためである．なお，この閾値の調整と副検出器を構成するサポート・ベクタ数は試行錯誤により決定した．

3.4 単眼カメラによる顔姿勢の推定と追跡

3.3 節で述べた顔及び顔器官検出器により, 撮影された画像内でのドライバの顔器官位置が取得できる．しかし，本研究では単眼カメラを用いるため，ここで取得される顔器官の位置は画像平面上での 2 次元座標位置である．そのため，この情報だけでは実際の 3 次元空間内での各顔器官の位置はわからず，ドライバの顔姿勢を正確に推定することができない．そこで，顔器官検出器によって検出された画面平面上での顔器官位置情報を元に，予め用意した 3 次元顔モデルとパーティクル・フィルタによりドライバの実 3 次元空間内での顔姿勢を推定・追跡することを考える．パーティクル・フィルタとは，多数のパーティクル（仮説群）を用いて前の状態からの予測情報と現在の観測情報に基づいて，現在の状態推定を行う手法で，近年，物体追跡などによく用いられる手法の一つである．本節では，はじめに顔姿勢推定・追跡の処理概要について説明し，その後，3 次元顔モデルとパーティクル・フィルタによる顔姿勢推定・追跡の方法について各々述べる．

3.4.1 顔姿勢推定・追跡処理の概要

本研究では　主に広範囲の顔及び顔器官の検出を実施するサーチ・モードと顔姿勢の推定・追跡を行うトラッキング・モードの 2 つの処理モードを切り替えてドライバの顔姿勢の推定を行う．図 3-21 にサーチ・モード時の処理フロー概要を示す．サーチ・モードでは 3.3 節で述べた手法を用いて入力画像全体からドライバの顔検出を試み，顔が検出された場合には，その顔の範囲内から顔器官を検出する．サーチ・モードではドライバがまだ検出されていない状態であるため，広範囲の顔向きを対象とした顔及び顔器官の検出と顔姿勢の安定した初期推定を行うことを目指す．そして，顔器官検出結果がトラッキング・モードへの遷移条件を満たした場合には，トラッキング・モードへと遷移し，そうでない場合はサーチ・モードの処理を繰り返す．

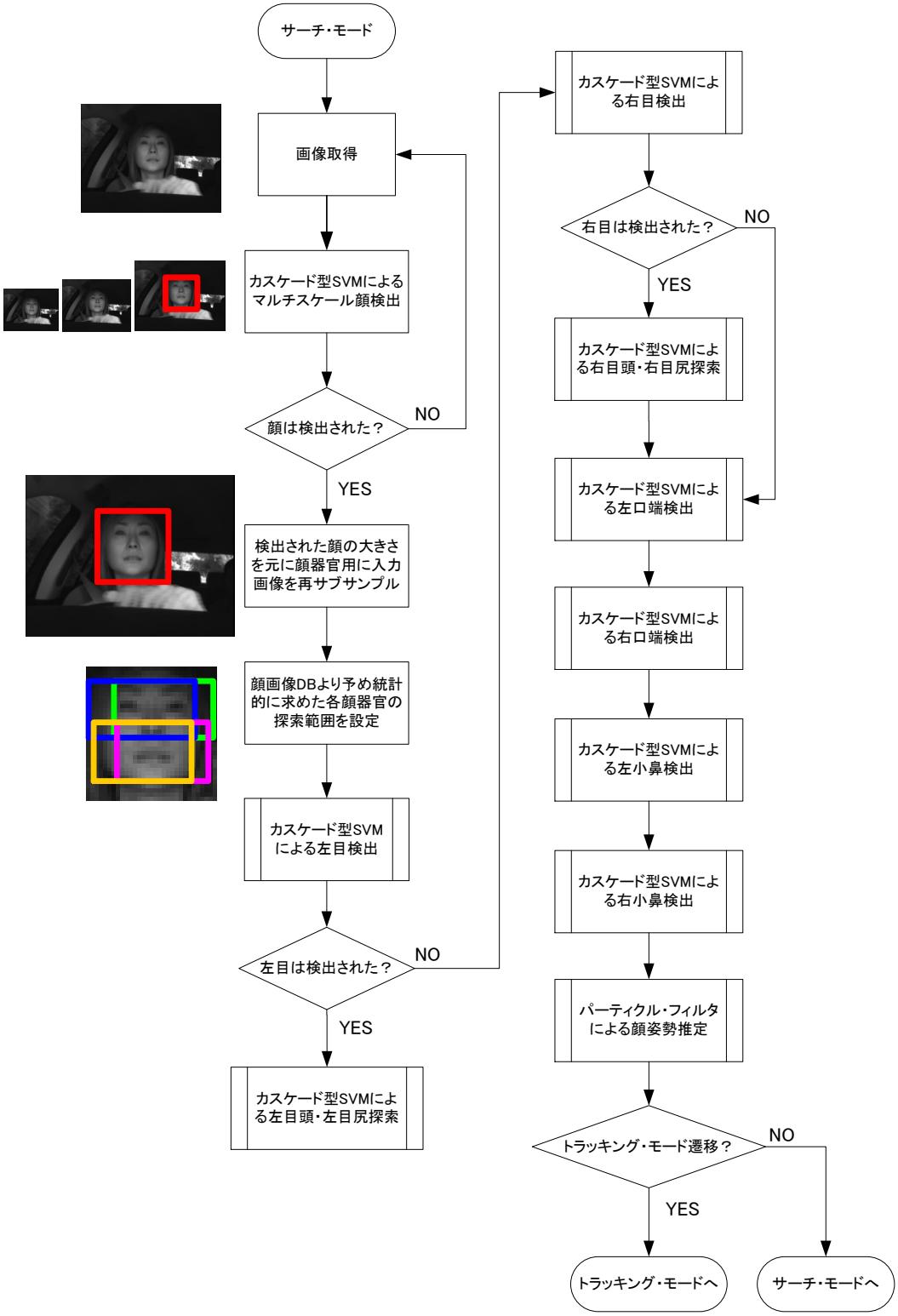


図 3-21　サーチ・モードの処理フロー概要

ここで，本研究ではサーチ・モードからトラッキング・モードへ遷移する条件として以下の 2 つを定義した．

1. 顔器官が 3 点以上検出されており，かつその内最低 1 点は両口唇端，両小鼻であること
   （縦方向の顔器官が検出されていると顔姿勢推定・追跡が安定するため）
2. 条件 1 を満たした上で顔姿勢推定した結果，SVM による顔器官検出位置と顔姿勢推定結果を 2 次元投影した顔器官位置との最大距離が閾値 Th_D以下であること

なお上記 2 つの条件は，トラッキング・モード時の姿勢推定を安定して行うため，ヒューリスティックに決定した条件である．そしてこれら 2 つの条件を満たした場合には，図 3-22 に処理フロー概要を示したトラッキング・モードへと遷移する．

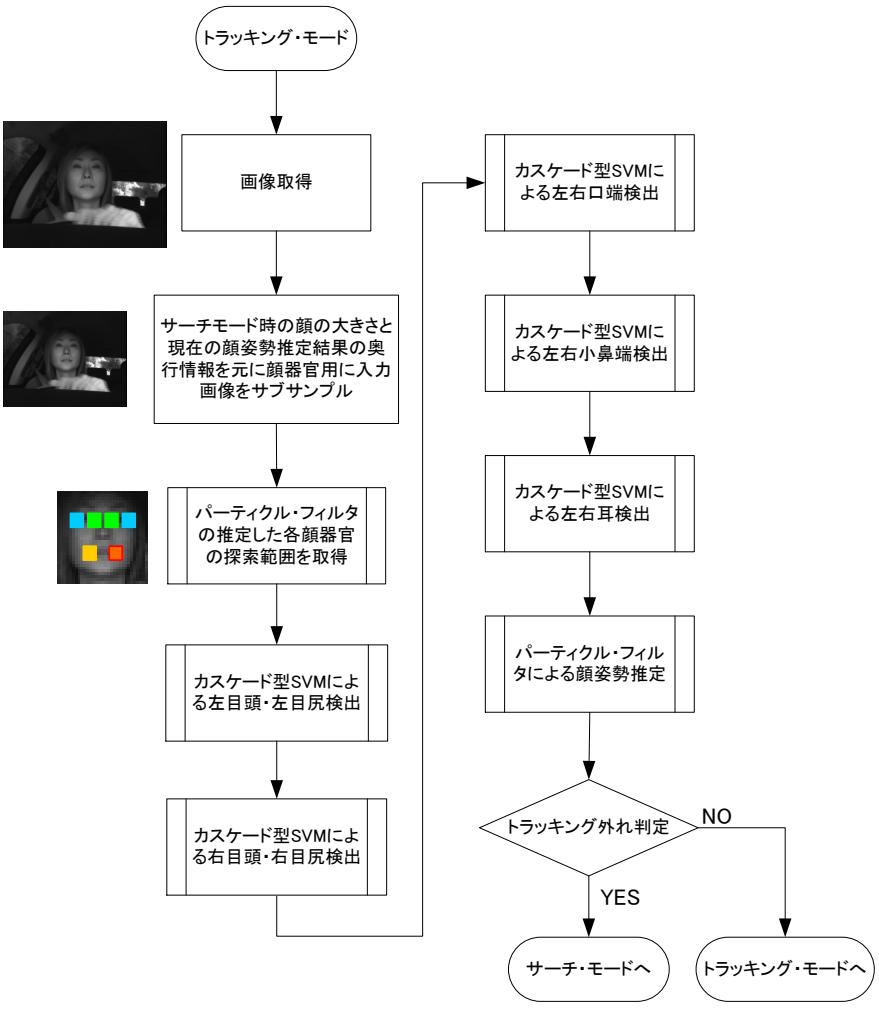


図 3-22　トラッキング・モードの処理フロー概要

トラッキング・モードではサーチ・モードを経て既にドライバの顔が検出された状態であるため，画像全体からの顔検出は行わず，基本的に顔器官のみの検出と，3 次元顔モデルとパーティクル・フィルタによる顔姿勢の推定と追跡を繰り返す．トラッキング・モードではできるだけ演算量を削減し，ドライバの顔姿勢推定の変化に追従することを目指す．そのため，トラッキング・モード時の顔器官検出処理では，探索エリアを絞り込むためパーティクル・フィルタによる顔姿勢推定の予測結果を利用する．そして，下記 2 つの条件を 1 つでも満たす場合にのみ，顔姿勢の推定・追跡に失敗したと判断し，サーチ・モードへと遷移する．

1. パーティクル・フィルタにより推定された各顔器官の探索エリアに関して，最大探索エリアが閾値 Th_A以上の場合
   （顔器官の探索エリアが発散したと考える）
2. 演算処理等による遅延が発生し画像取得間隔が一定時間 Th_T以上空いた場合
   （取得される画像がスキップした場合には，その間に姿勢が大きく変化している可能性があるため，一旦トラッキング・モードを解除する）

以上説明したように，本研究では初期検出を行うサーチ・モードと追従を行うトラッキング・モードとを適宜切り替えながら広範囲に変化するドライバの顔姿勢を推定・追跡する．

3.4.2　　3 次元顔モデル

3 次元顔モデルは，3 次元計測器（NEC エンジニアリング製 Danae 100SP）により実際の人物の顔を計測して取得した．計測後のデータ点数は約 120,000 点あるが，これを計算・データコストを小さくするために，顔器官検出器により検出される顔器官に対応する 3 次元座標点を保持しつつ，約 500 点に削減したものを 3 次元顔モデルとした(図 3-23 参照).

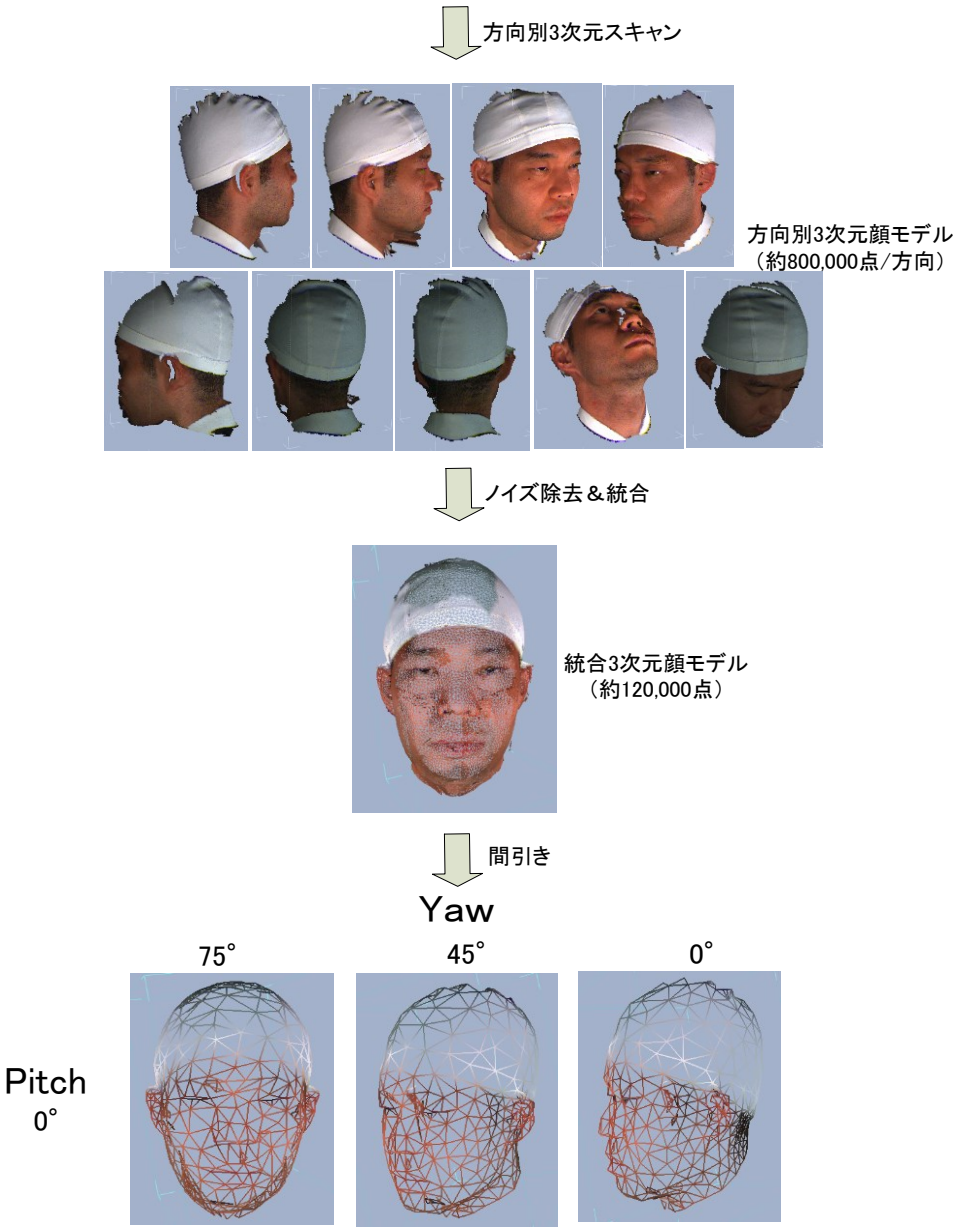


図 3-23　3 次元顔モデルの作成フロー

3.4.3 パーティクル・フィルタによる顔姿勢推定

SVM により検出された顔器官の画面座標位置(u, v)から，現在の顔姿勢（回転+平行移動）を推定する．画面座標は顔器官の 2 次元位置情報しか含まないため，これらのみから 3 次元情報である顔姿勢を復元することはできない．そこで，3 次元顔モデルとその姿勢$(\theta_y, \theta_p, \theta_r, T_X, T_Y, T_Z)$を状態とするパーティクル・フィルタを用いることで，本来（単眼カメラの 1 枚の静止画からは）復元不可能な顔の 3 次元姿勢情報を復元することを試みる．ここで$\theta_y, \theta_p, \theta_r$は各々Yaw 方向，Pitch 方向，Roll 方向の回転角度，T_X, T_Y, T_Zは各々X 軸方向，Y 軸方向，Z 軸方法の移動量を表す．また，本研究では，パーティクル・フィルタの尤度としては，顔器官検出結果と 3 次元顔モデルにより求まる姿勢推定の確からしさである姿勢推定尤度と，1 期前と現在のパーティクルの状態を運動モデルにより推定した時の確からしさである運動尤度の 2 つを併せた統合尤度を用いる．以下にパーティクル・フィルタによる顔姿勢の推定フローを記す．

① 各パーティクルの状態に従い 3 次元顔モデルを回転・平行移動する
② 3 次元顔モデルを画面平面上へ投影し各顔器官の画面座標位置を求める
③ SVM により検出された顔器官の座標位置と②のユークリッド誤差Jを算出し，各パーティクルの姿勢尤度L_pを求める
④ 静止モデルを仮定し，各パーティクルの運動尤度L_kを算出する
⑤ ③で求めた姿勢尤度L_pと④で求めた運動尤度L_kを統合し，各パーティクルの尤度Lと算出する
⑥ 各パーティクルの尤度Lを比較し，尤度最大のパーティクルnを推定結果とする

姿勢推定尤度の定義

本研究で用いるパーティクル・フィルタのパーティクルは前述の顔姿勢（6 次元）を状態とするM個の仮説群により構成される．あるパーティクルmの状態に基づいて，3 次元顔モデルの姿勢を変化させ，その時の各顔器官を画面上へ投影した時の画面座標位置$(\hat{u}_m, \hat{v}_m)$と SVM により検出された顔器官の画面座標位置(u, v)との二乗誤差J_mを求める．今，SVM により検出された顔器官iの画面座標位置を(u_i, v_i)とし，あるパーティクルmの状態に従って求められた顔姿勢から推定された顔器官iの画面座標位置を$(\hat{u}_{mi}, \hat{v}_{mi})$とした場合，パーティクル$m$による姿勢推定の二乗誤差$J_m$は式(3.1)で求まる．

$$J_m = \frac{1}{N}\sum_{i=1}^{N}\left\{(u_i - \hat{u}_{mi})^2 + (v_i - \hat{v}_{mi})^2\right\} \tag{3.1}$$

ここで，Nは検出された顔器官の数，iは顔器官のインデックス，mはパーティクルの個体番号$(1, 2, \ldots, M)$を表す．この誤差J_mを用いてパーティクルmの姿勢推定尤度L_{mp}を式(3.2)のように定義する．

$$L_{mp} = \exp(-J_m) \tag{3.2}$$

姿勢推定尤度の算出方法

今，3 次元顔モデルのある顔器官iにのみ着目し，その 3 次元座標を $P_i = (X_i, Y_i, Z_i)$とする．ここで，あるパーティクルmの状態が$(\theta_y, \theta_p, \theta_r, T_X, T_Y, T_Z)$の時，この状態に従って，顔器官$i$を回転平行移動した座標$P_i' = (X_i', Y_i', Z_i')$は式(3.3)より求まる．

Yaw(*Y*軸周りの回転)　*Pitch*(*X*軸周りの回転)　*Roll*(*Z*軸周りの回転)

$$\begin{aligned}\begin{pmatrix} X_i' \\ Y_i' \\ Z_i' \end{pmatrix} &= \begin{pmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_p & -\sin\theta_p \\ 0 & \sin\theta_p & \cos\theta_p \end{pmatrix}\begin{pmatrix} \cos\theta_r & -\sin\theta_r & 0 \\ \sin\theta_r & \cos\theta_r & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \\ &= \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}\end{aligned}$$

これを同次座標系で表現すると、 (3.3)

$$\begin{aligned}\begin{pmatrix} X_i' \\ Y_i' \\ Z_i' \end{pmatrix} &= \begin{pmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{pmatrix}\begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \\ &= [R \mid T]\begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}\end{aligned}$$

これにより，パーティクルmの状態に従った 3 次元空間上での顔器官iの 3 次元座標位置が求まる．

次に，3 次元顔モデルの各顔器官を画面上へ投影することを考える．3 次元顔モデルの

任意の顔器官iをあるパーティクルmの状態に従って回転平行移動した時の3次元座標位置を(X_i', Y_i', Z_i')とした時，それを画面上へ透視投影した点$(\hat{u}_i, \hat{v}_i)$は，カメラの内部パラメータ行列Aを用いて，式(3.4)で定義される．ここで，fはカメラの焦点距離，δ_uとδ_vは各々横方向と縦方向の画素の物理的な間隔，(C_u, C_v)は画像座標系における光軸と座標面との交点の位置（画像中心）を表す．

$$\begin{pmatrix} \hat{u}_i \\ \hat{v}_i \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{X_i'\left(\frac{f}{\delta_u}\right)+Z_i'C_u}{Z_i'} \\ \frac{Y_i'\left(\frac{f}{\delta_v}\right)+Z_i'C_v}{Z_i'} \\ 1 \end{pmatrix} \tag{3.4}$$

これにより，パーティクルmの状態に従った顔器官$\boldsymbol{i}$の画面上での2次元座標位置$(\hat{u}_i, \hat{v}_i)$が決まる．そして，ここで求まった$(\hat{u}_i, \hat{v}_i)$とSVMによる顔器官iの検出結果を利用して式(3.1)の誤差を計算することにより，式(3.2)で示した姿勢推定尤度が算出できる．

平行移動成分の算出方法

前述の通り，パーティクルの状態により3次元顔モデルを姿勢変化させた後に画面へ投影した顔器官の位置とSVMにより検出された顔器官の画面座標位置とのユークリッド誤差を求めることで，最も尤もらしい顔姿勢が推定できることを示した．しかし，実際には平行移動成分は回転成分に比べて分布範囲が非常に広いため，比較的少数のパーティクルにより構成されるパーティクル・フィルタによる予測では，精度よく推定することが困難である．そこで本研究では，図 3-24 に示すフローに従って，パーティクルの平行移動成分(T_x, T_y, T_z)はSVMによる顔器官検出点との二乗誤差が最小になるように，最適化手法によって解析的に求める．

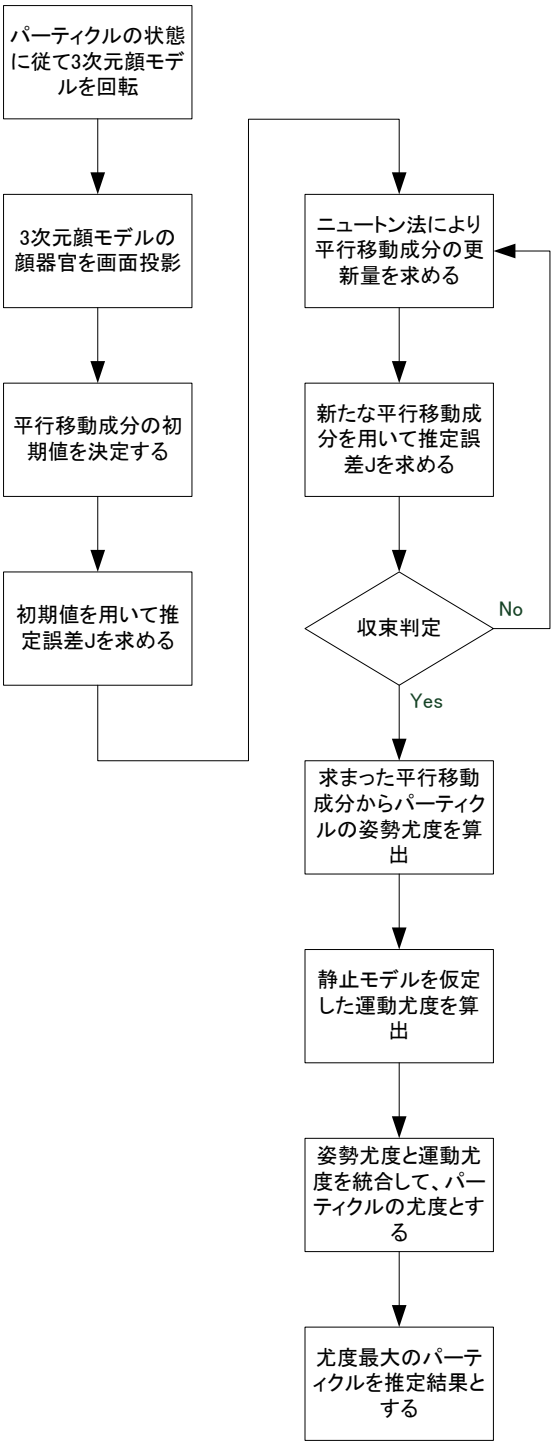


図 3-24　平行移動成分の算出フロー概要

今，あるパーティクルmの回転成分$(\theta_y,\ \theta_p,\ \theta_r)$は決まっていて，平行移動成分$(T_x,\ T_y,\ T_z)$は未知とし，SVM により検出された，ある顔器官$i$の画面座標位置を$(u_i,\ v_i)$とする．ま

た，3 次元顔モデルのある顔器官iの 3 次元座標位置(X_i, Y_i, Z_i)をパーティクルmの回転成分に従って回転を施した座標位置を(X_i', Y_i', Z_i')とし，それを X, Y, Z 軸方向に各々(T_X, T_Y, T_Z)平行移動した後に画面上へ投影した点を$(\hat{u}_{mi}, \hat{v}_{mi})$とする．このとき，そのパーティクル$m$の推定誤差$J_m$は式(3.5)で表される．ここで，$A_{ij}$はカメラ内部パラメータ行列の$i$行$j$列目の要素を表す.

$$J_m = \frac{1}{N}\sum_{i=1}^{N}\left\{\left(u_i - \frac{A_{00}(X_i^{'} + T_x) + A_{02}(Z_i^{'} + T_z)}{(Z_i^{'} + T_z)}\right)^2 + \left(v_i - \frac{A_{11}(Y_i^{'} + T_y) + A_{12}(Z_i^{'} + T_z)}{(Z_i^{'} + T_z)}\right)^2\right\} \tag{3.5}$$

この推定誤差J_mを最小にする平行移動成分(T_x, T_y, T_z)を求める.

はじめに，平行移動成分の初期値(T_x^0, T_y^0, T_z^0)を考える．SVM により検出された顔器官の数をNとし，その平均座標を$(\bar{u}, \bar{v})$，分散を(V_u, V_v)とすると各々式(3.6)(3.7)により算出できる.

$$(\bar{u}, \bar{v}) = \left(\sum_{i=1}^{N} u_i, \sum_{i=1}^{N} v_i\right) \tag{3.6}$$

$$(V_u, V_v) = \left(\frac{1}{N}\sum_{i=1}^{N}(u_i - \bar{u})^2, \frac{1}{N}\sum_{i=1}^{N}(v_i - \bar{v})^2\right) \tag{3.7}$$

また，SVM により検出されたN個の顔器官に対応した 3 次元顔モデルの顔器官をパーティクルの状態に従って回転した後の平均座標を$(\bar{X}, \bar{Y}, \bar{Z})$，分散$(V_X, V_Y, V_Z)$をとすると各々式(3.8)(3.9)で算出できる.

$$(\bar{X}, \bar{Y}, \bar{Z}) = \left(\sum_{i=1}^{N} X_i, \sum_{i=1}^{N} Y_i, \sum_{i=1}^{N} Z_i\right) \tag{3.8}$$

$$(V_X, V_Y, V_z) = \left(\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2, \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2, \frac{1}{N}\sum_{i=1}^{N}(Z_i - \bar{Z})^2\right) \tag{3.9}$$

ここで, 3 次元空間での X, Y 軸方向の分散と, 画面座標での u, v 軸方向の分散を用いて,

Z 軸方向の平行移動成分の初期値T_z^0を式(3.10)のように定義する．

$$T_z^0 = \sqrt{\frac{V_x + V_y}{V_u + V_v}} - \bar{Z} \tag{3.10}$$

次に，X,Y 軸方向の平行移動成分の初期値T_x^0, T_y^0を求める．3 次元座標点$(\bar{X}, \bar{Y}, \bar{Z})$が画面上の点$(\bar{u}, \bar{v})$に投影されたとすると，先に求めた Z 軸方向の平行移動成分の初期値T_z^0を用いて，T_x^0, T_y^0は各々式(3.11)(3.12)のように求められる

$$T_x^0 = \frac{\bar{u}(\bar{Z} + T_z^0) - (A_{00}\bar{X} + A_{02}\bar{Z}) - A_{02}T_z^0}{A_{00}} \tag{3.11}$$

$$T_y^0 = \frac{\bar{v}(\bar{Z} + T_z^0) - (A_{11}\bar{Y} + A_{12}\bar{Z}) - A_{12}T_z^0}{A_{11}} \tag{3.12}$$

次に，求まったT_x^0, T_y^0, T_z^0を初期値として，目的関数J_mの最小化問題をニュートン法により解くことを考える．目的関数J_mを最小にする平行移動成分(T_x, T_y, T_z)を求めるには，式(3.13)の連立方程式から各成分の更新量$(\Delta T_x, \Delta T_y, \Delta T_z)$を求めれば良い．

$$\begin{pmatrix} \frac{\partial^2 J_m}{\partial T_x^2} & \frac{\partial^2 J_m}{\partial T_x \partial T_y} & \frac{\partial^2 J_m}{\partial T_x \partial T_z} \\ \frac{\partial^2 J_m}{\partial T_x \partial T_y} & \frac{\partial^2 J_m}{\partial T_y^2} & \frac{\partial^2 J_m}{\partial T_y \partial T_z} \\ \frac{\partial^2 J_m}{\partial T_x \partial T_z} & \frac{\partial^2 J_m}{\partial T_y \partial T_z} & \frac{\partial^2 J_m}{\partial T_z^2} \end{pmatrix} \begin{pmatrix} \Delta T_x \\ \Delta T_y \\ \Delta T_z \end{pmatrix} = \begin{pmatrix} \frac{\partial J_m}{\partial T_x} \\ \frac{\partial J_m}{\partial T_y} \\ \frac{\partial J_m}{\partial T_z} \end{pmatrix} \tag{3.13}$$

そして，求まった更新量から新たな平行移動成分(T_x', T_y', T_z')を式(3.14)により求め，更新量が予め決定した閾値以下となるまで，誤差の計算と平行移動成分(T_x', T_y', T_z')の更新を繰り返し，収束した時の平行移動成分をパーティクルmの平行移動成分(T_x, T_y, T_z)と決定する．

$$\begin{pmatrix} T_x' \\ T_y' \\ T_z' \end{pmatrix} = \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} + \begin{pmatrix} \Delta T_x \\ \Delta T_y \\ \Delta T_z \end{pmatrix} \tag{3.14}$$

運動尤度の算出方法

顔姿勢推定においては運動モデルとして静止モデルを仮定する．静止モデルでは 1 期前の状態と同じ状態であると尤度が高く，状態に変化があるほど尤度が低くなるモデルである．今，あるパーティクルmの現在の状態をp^t，1 期前のパーティクル状態をp^{t-1}とし，パーティクルの状態を$(\theta_y, \theta_p, \theta_r, T_X, T_Y, T_Z)$，各要素の標準偏差を$(\sigma_y, \sigma_p, \sigma_r, \sigma_X, \sigma_Y, \sigma_Z)$と表す．この時，このパーティクルの運動尤度$L_{mk}(p^t|p^{t-1})$ を以下のように定義する．

$$L_{mk}(p^t \mid p^{t-1}) = \exp\left\{-\left(\frac{(\theta_y^t-\theta_y^{t-1})^2}{\sigma_y^2}+\frac{(\theta_p^t-\theta_p^{t-1})^2}{\sigma_p^2}+\frac{(\theta_r^t-\theta_r^{t-1})^2}{\sigma_r^2}+\frac{(T_X^t-T_X^{t-1})^2}{\sigma_X^2}+\frac{(T_Y^t-T_Y^{t-1})^2}{\sigma_Y^2}+\frac{(T_Z^t-T_Z^{t-1})^2}{\sigma_Z^2}\right)\right\} \tag{3.15}$$

統合尤度の算出方法

あるパーティクルmの姿勢推定尤度がL_{mp}，運動尤度がL_{mk}であったとき，このパーティクルの統合尤度L_mは次のように求める．

$$L_m = L_{mp} \cdot L_{mk}(p^t \mid p^{t-1}) \tag{3.16}$$

結局，式(3.17)で表せるように，この統合尤度L_mを最大にするパーティクルnが最も尤もらしい顔姿勢推定結果を与えるパーティクルとなる．

$$\mathrm{n} = \underset{m}{\mathrm{argmax}}(L_m) \tag{3.17}$$

なお，ドライバの姿勢変化によっては，顔器官が画像上で観測されない場合が存在する（例えば，ドライバが左右確認のために大きく右を向いた場合には，右側面に存在する顔器官の幾つかは画像上に映らない）が，このような場合にもロバストに顔姿勢を推

定するため，姿勢推定された 3 次元顔モデルの各顔器官の法線方向を求め，その顔器官を画像上に投影した場合に実際には映りえないと判定された場合には，顔姿勢の尤度L_{mp}の計算からこれらの顔器官を除外した．またその他に，何らかの遮蔽物等による顔器官の隠れや顔器官検出器の検出ミスにより検出できなかった顔器官が存在した場合についても，同様に尤度計算から除外することで全ての顔器官が検出できなくても顔姿勢が推定できるよう対処した．

3.4.4 パーティクル・フィルタによる顔姿勢追跡

トラッキング・モードでは顔器官の検出を高速に行う必要がある．そのため，各顔器官の探索エリアはできるだけ小さい方が望ましい．サーチ・モードにおいては顔姿勢が推定できていないため，顔検出結果から存在し得る最大範囲において各顔器官を探索する必要があったが，トラッキング・モードではパーティクル・フィルタにより顔姿勢が推定されているため，その状態から顔器官の探索エリアを予測することができる．今 3 次元顔モデルの顔器官iをパーティクルjの状態により画面上へ投影したときの座標位置$(\hat{u}_{ij}, \hat{v}_{ji})$のパーティクル jの尤度L_jを重みとした場合，その加重平均(x_i, y_i)及び分散・共分散(v_x, v_y, v_{xy})は式(3.18)で算出できる．

$$
\begin{aligned}
&(x_i, y_i) = \left(\frac{1}{W}\sum_{j=1}^{M} L_j \hat{u}_{ji}, \frac{1}{W}\sum_{j=1}^{M} L_j \hat{v}_{ji} \right) \\
&(v_x, v_y, v_{xy}) = \left(\frac{1}{W}\sum_{j=1}^{M} L_j (\hat{u}_{ji} - x_i)^2, \frac{1}{W}\sum_{j=1}^{M} L_j (\hat{v}_{ji} - y_i)^2, \frac{1}{W}\sum_{j=1}^{M} L_j (\hat{u}_{ji} - x_i)(\hat{v}_{ji} - y_i) \right) \qquad (3.18) \\
&\text{ここで、} W = \sum_{j=1}^{M} L_j,
\end{aligned}
$$

そして，求まった顔器官iの平均位置(x_i, y_i)，分散(v_x, v_y)をその顔器官の予測分布とし，顔器官ごとに予め設定されている標準偏差σ_iを考慮して，トラッキング・モードにおける次フレームでの各顔器官の探索エリア（左上座標, 幅, 高さ）=(sx_i, sy_i, sw_i, sh_i)を式(3.19)により決定する．

$$
(sx_i, sy_i, sw_i, sh_i) = \left(x_i - \sigma_i\sqrt{v_x}, y_i - \sigma_i\sqrt{v_y}, 2\sigma_i\sqrt{v_x}, 2\sigma_i\sqrt{v_y} \right) \qquad (3.19)
$$

3.5 評価実験

3.5.1 顔及び顔器官検出器の性能評価

学習した顔及び顔器官検出器の性能を評価するため，全サンプルからランダムに選択した半数を学習サンプルとし,残りの半数を評価サンプルとしたオープンな検出評価を 5 回繰り返し，検出率と計算時間の平均値を調査した.

はじめに，主検出器のみを用いた場合とカスケード構造化した場合との検出率の比較を表 3.2 に示す.

表 3.2 顔及び顔器官検出率

Detection Target	Detection Rate (%)		
	Full	32 & Full Cascaded	128 & Full Cascaded
Face	93.87	92.50	93.19
Eye	92.06	91.63	91.88
Inner eye corner	90.25	89.67	90.19
Outer eye corner	86.28	86.17	86.59
Nostril	87.31	86.44	87.20
Mouth corner	95.12	94.53	94.92
Ear	93.96	93.63	93.80

Best case

各検出器とも 86～95 %の検出率を達成しており，主検出器のみを用いた場合（Full)，32 個のサポート・ベクタで構成された副検出器と主検出器をカスケード構造にした場合（32 & Full Cascaded)，128 個のサポート・ベクタで構成された副検出器と主検出器をカスケード構造にした場合（128 & Full Cascaded）を比較すると，何れの検出器においても同程度の検出率が得られていることがわかる.

次に，計算時間の比較を表 3.3 に示す．今回の結果では 1 段目のサポート・ベクタ数に依らず, カスケード構造化により検出処理が 3〜5 倍程度高速化されていることがわかる．

表 3.3　顔及び顔器官検出処理時間

Detection Target	5 times average "CPU Time (sec)"		
	Full	32 & Full Cascaded	128 & Full Cascaded
Face	268.13	141.27	86.29
Eye	21.67	7.78	6.45
Inner eye corner	1.38	0.45	0.49
Outer eye corner	4.39	0.95	1.18
Nostril	5.60	2.37	1.78
Mouth corner	13.80	4.67	4.63
Ear	250.77	75.33	43.01

Best case

これは，計算量の少ない副検出器により検出対象候補が絞り込まれ，少数の検出困難な候補のみが計算量の多い主検出器に渡されていることを示している．ここで，必ずしも副検出器のサポート・ベクタ数が少ない方の処理速度が速くなっていないのは，副検出器の性能が低すぎると計算量の多い 2 段目の主検出器に渡る候補の数が増えるため，全体として計算量が増えるためである．以上 2 つの評価実験より，カスケード構造化により検出率を維持しつつ，計算時間の大幅な削減が実現出来ることが確認された．

次に，耳検出の有無による Yaw 方向の顔姿勢変化に対する効果を調べるため,実験室内での左右往復動作に対する追跡結果の比較を図 3-25 に示す. 2 回の左右往復動作に対して，耳の検出を実施した方が安定して顔姿勢の追跡ができていることがわかる．特に実際に耳が見えてくる広角な顔姿勢（図中【a】）においてその差が顕著に表れていることが読み取れる.

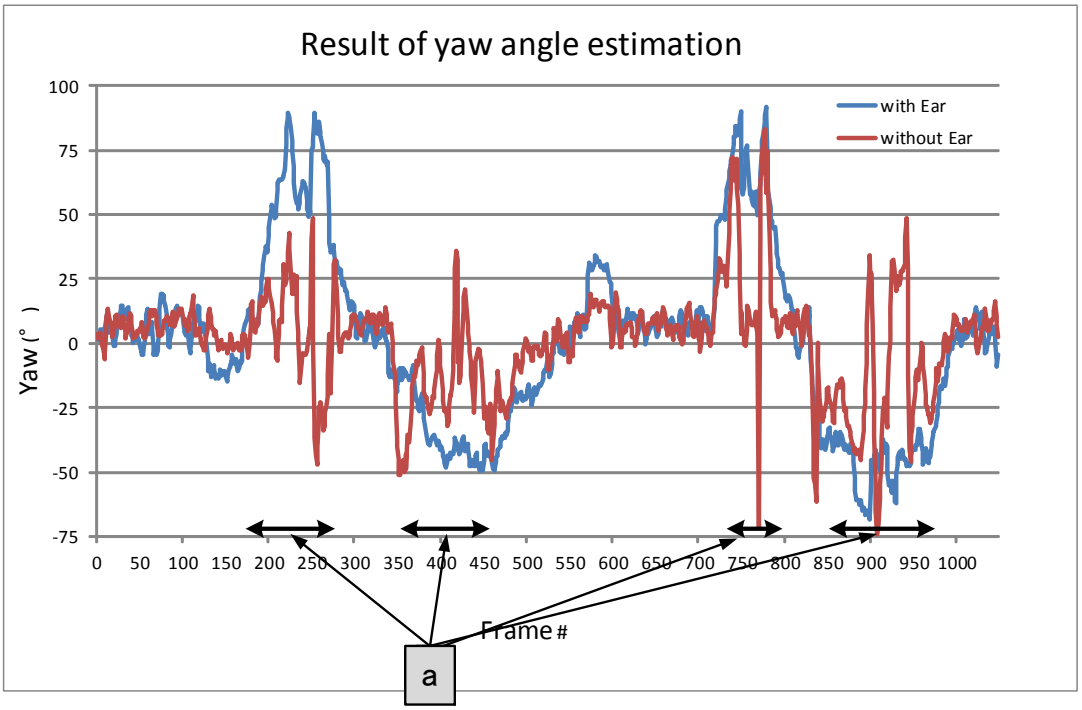


図 3-25　耳の検出有無による顔姿勢（Yaw 角）推定結果の比較

3.5.2　　室内映像による実験

提案したパーティクル・フィルタによる顔姿勢推定手法の検出精度・範囲を調査するため，実験室内にて顔姿勢の評価映像を撮影すると共に，被験者にマーカー付の帽子を被ってもらい顔姿勢の正解値を同時記録した．

はじめに，Yaw 方向に約±80 度顔姿勢変化を変化させた場合の推定結果と正解値を図 3-26 に示す．なお本実験では，パーティクル・フィルタの推定精度に着目するため，画像に映っている顔器官が安定して検出されたと仮定するために各顔器官の検出位置の誤検出/未検出の影響を避けるため，各顔器官の位置は SVM 検出器による検出結果ではなく，人手によりハンドラベルした位置を用いた．

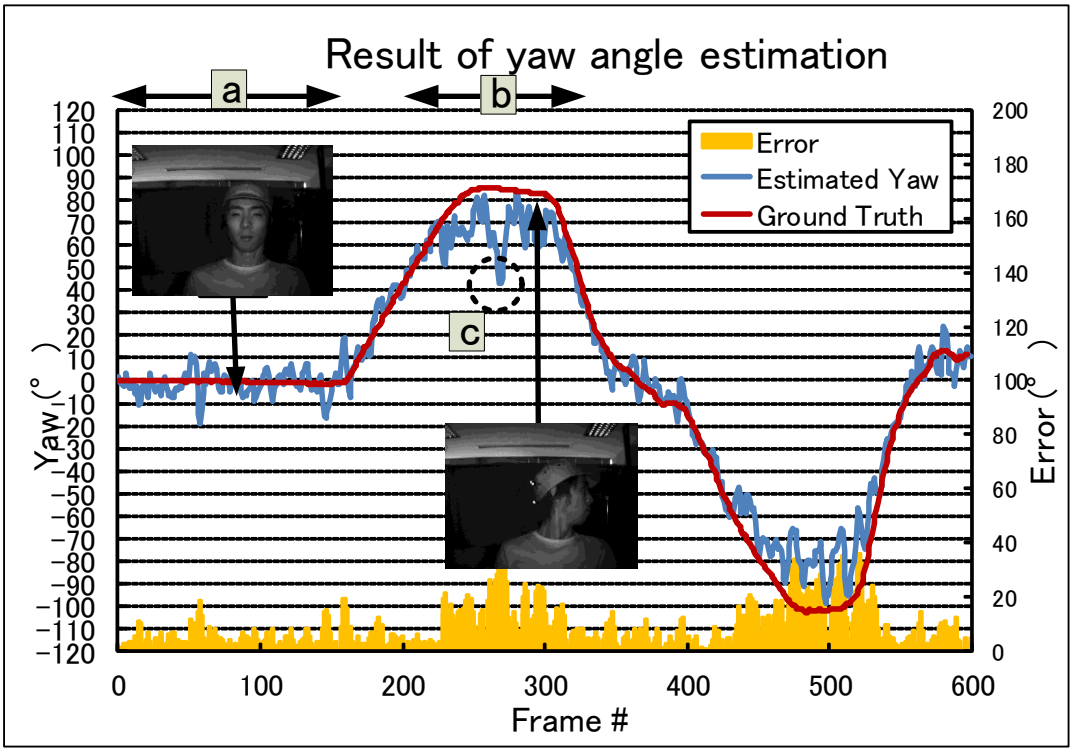


図 3-26　左右首振り動作時の顔姿勢（Yaw 角）推定結果

正面付近（図中【a】）での平均誤差は約 4.3 度と小さいが，横向きになるほど誤差が増加する傾向にあり，Yaw 方向 60～80 度付近（図中【b】）における平均誤差は約 13.5 度であった. 推定誤差の要因としては, 本人以外の 3 次元顔モデルを使用していること，カメラ校正の問題などが考えられる．また，Yaw 方向の顔姿勢変化の実験では，突発的に誤差が大きい場合がある（図 3-26 中【c】，最大誤差：約 42.1 度）が，この様に突発

的に生じる誤差は後処理により縮小できると考えられる.

次に，前後方向に約 500 mm 顔姿勢を変化させた場合の推定結果と正解値を図 3-27 に示す．なお、本実験では顔器官の検出位置は SVM による検出位置を用いた.

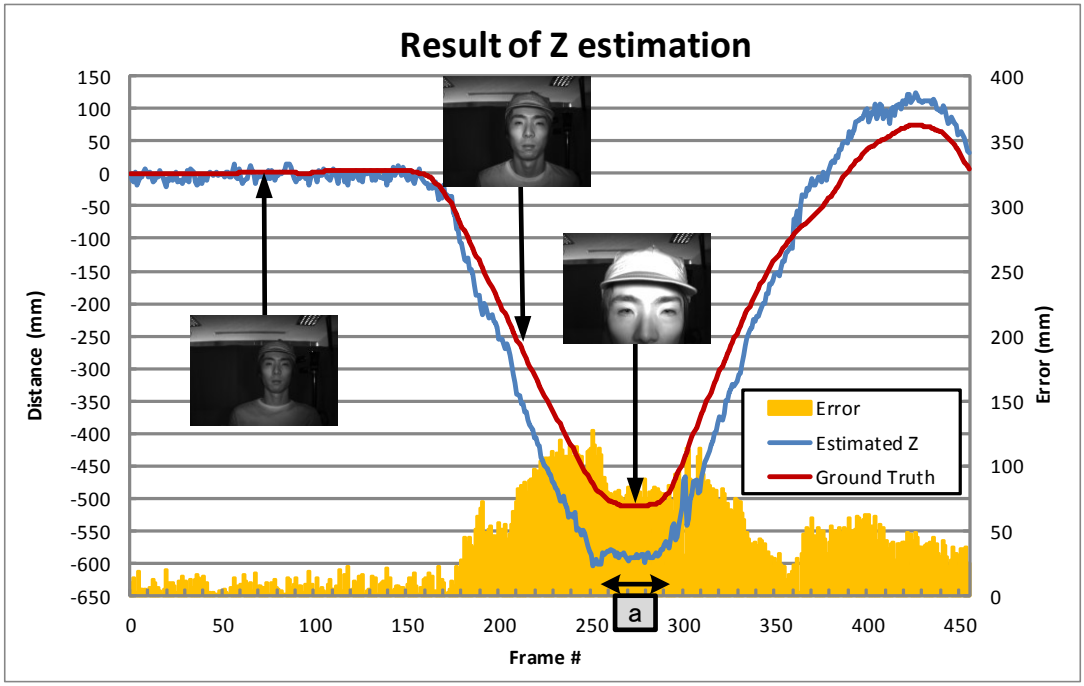


図 3-27　前後動作時の顔位置推定結果

前後動作の起点を 0 mm とし，推定誤差は被験者がカメラに接近する程大きくなり，最接近区間（図 3-27 中【a】付近）に着目すると 75～125 mm 程度の誤差が生じていることがわかる．ここで誤差の要因は，左右首振り動作の実験同様に本人以外の 3 次元顔モデルを使用していることによる誤差，カメラ校正による誤差に加え SVM による顔検出点の誤差などが考えられる. また, 区間【a】で誤差が大きくなった原因の一つは，図 3-27 中に表示した最接近時の顔画像を見てもわかるように鼻から下の顔器官が画角外へはみ出したことにより，これらの顔器官を顔姿勢推定に使えないことなども影響していると考えられる.

これらの 2 つの実験結果から，誤差は含まれるものの提案手法により単眼カメラでもドライバの安全確認時の行動を模擬した，左右首振り及び前後移動動作の広範囲な姿勢変化を推定できることを確認した.

3.5.3　　実走行映像による実験

次に，背景や照明環境変動の影響や実際の走行時の顔姿勢変化に対する顔・顔器官の検出及び顔姿勢推定の性能を調査するため，運転中のドライバ映像に本手法を適用した結果を図 3-28 に示す.

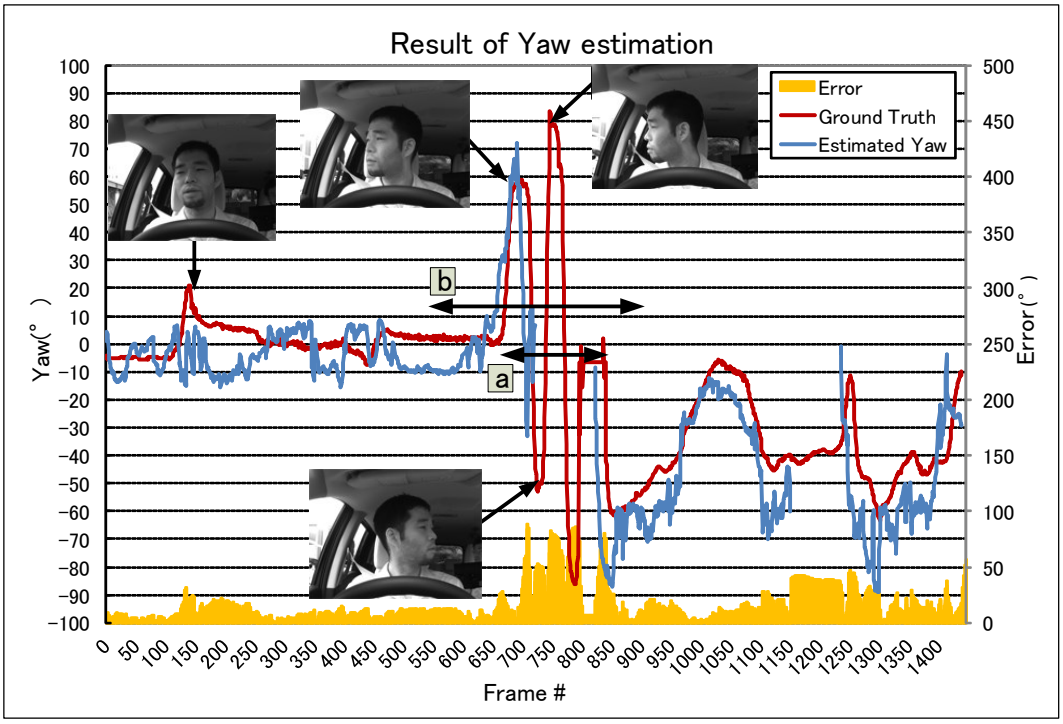


図 3-28　自動車運転時の顔姿勢（Yaw 角）推定結果

Yaw 角の変化が非常に速くて大きい図 3-28 中【a】の区間では，パーティクル・フィルタの仮説群の範囲を超えてドライバの顔姿勢が変化したため，途中から顔器官の検出・追跡に失敗し，顔姿勢の推定ができていない．しかし，その後，顔姿勢の変化速度が緩やかになると，再び顔及び顔器官が検出・追跡され，概ね顔姿勢が推定できていることがわかる.

次に，速い顔姿勢変化に対するトラッキング性能を改善するため，図 3-28 中【b】の区間について, パーティクル数を 500 から 2500 個へ,顔器官の探索エリアを 1.5 倍に拡張した時の結果を図 3-29 に示す.

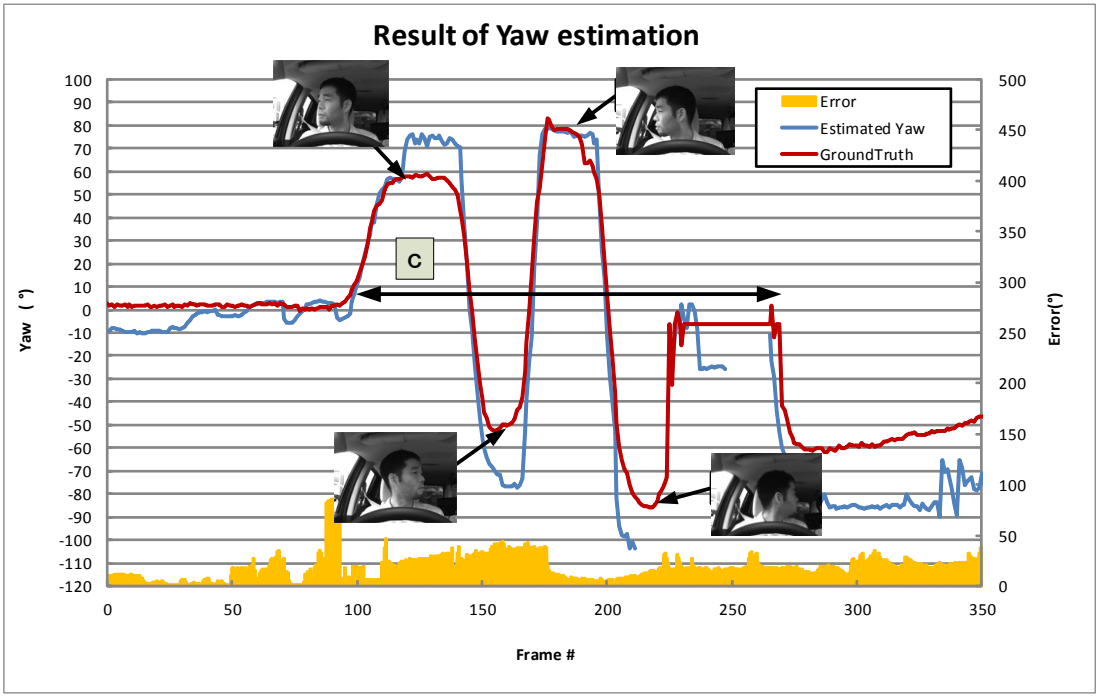


図 3-29　パーティクル数と探索エリアを拡張した時の顔姿勢（Yaw 角）推定結果

図 3-29 中【c】の区間のドライバの左右安全確認動作に着目すると，図 3-28 では失敗していた 150～200 frame 付近まで追跡が出来ていることがわかる. また, 200～275 frame では追跡が外れ，顔姿勢の推定ができていないが，これはドライバが左やや後方を注視したことで，画像上に顔器官がほとんど映らなかったためである.本評価を通して，このような特別な場合を除いて，パーティクル数と顔器官の探索エリアの大きさを調整することで，運転中のドライバの顔姿勢を追跡・推定できることが確認できた.

3.6　　まとめ

ドライバの安全確認行動に応じて適切な支援を行う安全運転支援システムの実現を目指し，映像のみから顔姿勢を推定する技術について検討した．そして，不特定人に対してロバストな SVM による顔・顔器官の検出法，3 次元顔モデルとパーティクル・フィルタを用いた姿勢推定により，単眼車載カメラ映像のみからドライバの顔姿勢を推定する

手法を確立した.

顔及び顔器官検出器の評価実験では，検出器のカスケード構造化により検出率を維持しつつ計算量を大幅に削減可能なことを確認し，また Yaw 方向の顔姿勢変化に対する耳検出器の有効性についても確認した.

顔姿勢の追跡・推定評価実験では，実験室内の評価映像に対し，推定誤差は含まれるものの前後方向に 500 mm，Yaw 方向に±80 度程度の範囲で推定ができることを示した.そして，実走行映像に対しては，パーティクル数や顔器官探索エリアを適切に調整することで,素早い動作を含むドライバの顔姿勢変化を追跡･推定ができることを確認した.

これらの結果を通して，本提案手法により顔姿勢が広範囲に変化するドライバの安全確認行動を検知できる可能性が示せたと考える．今後は，現状未達であるリアルタイムでの動作，走行映像による実験で明らかとなった素早い動作へ対応するための適応的なパーティクル数と探索エリア調整手法の確立，汎用 3 次元顔モデルの導入などの検討を行い，よりロバストで実用的なシステムとして成熟させていきたいと考えている.

第4章　笑顔度推定とその音楽療法効果評価への応用

4.1　はじめに

我が国では，2013 年 10 月 1 日現在，総人口における 65 歳以上の割合が 25.1 %にも達し，今や国民の約 4 人に 1 人が高齢者となり，さらに 75 歳以上人口が全体の 12.3 %を占めるなど，本格的な超高齢社会を迎えている [32]．そして，この少子高齢化傾向が今後も継続すると，2060 年には 2.5 人に 1 人が 65 歳以上，4 人に 1 人が 75 歳以上となり，先進諸国の中でも特に高齢化が急速に進んでいくことが予測されている．そのため，介護・医療・福祉の充実とそれに付随する社会保障費の削減という相反する課題を共に解決することが強く求められている．特に近年は，このような長寿・高齢への社会構造変化に伴い，病院を訪れる患者の疾病内容や治療方法も変化，多様化してきており，生活習慣病のケア，脳卒中後や要介護者の自立を促すためのリハビリテーションなどの重要性が年々増してきている．このような背景から，近年は絶対的な患者数の増加が一因となり，リハビリテーションが必要な患者に対して適切な方法と期間を持って施術が行えないなどの問題も出始めている．この問題を解決するためには，社会的な体制，施設や人員を備えていくことも必要ではあるが，と同時に医療施設等において患者の疾患から完治，ポストケアに至るまでの治療サイクルを最適化していくかが重要である．そしてその為には，個々の患者の回復状態を適切に評価・把握しつつ，必要なリハビリテーションを計画的に実施するため，リハビリテーションの効果を客観的・定量的に評価することが臨床現場での重要な課題の 1 つとなってきている．

そのような病院でのリハビリテーションの一環として行われる取り組みの 1 つに音楽療法がある．音楽療法とは，「音楽のもつ生理的，心理的，社会的働きを用いて，心身の障害の回復，機能の維持改善，生活の質の向上，行動の変容などに向けて，音楽を意図的，計画的に使用すること」と定義されている [33]．本療法においては従来，患者に対する療法の効果を，病院や音楽療法士が各々独自に設けた評価基準と介入内容の記録などを通じて質的・量的に評価することが試みられてきた．しかし，患者の症状や回復状況が様々であり，また評価者としての音楽療法士の判断基準も経験等によって左右される部分も少なからず存在することから，客観的で統一的な評価方法を確立することは非常に困難であった．そこで我々は，人間の心身賦活に伴って広く一般的に見られる表情変化である笑顔に着目した．

そもそも我々の社会生活において，視覚から得られる情報の果たす役割は非常に大き

く，特に人と人とのコミュニケーションにおいては，言語に加え，相手の仕草や顔の表情が視覚的に確認できることで，我々はよりスムーズに，より深く相手の意図を理解することができる．人間の表情の解析については，心理学 [34]を始めとした様々な学術分野からアプローチがなされているが，工学の分野 [35]においても，表情や感情を顔画像，声，生体信号などを用いて如何に機械により解析・推定・理解するかという試みが多くの研究者によってなされてきた．しかし，顔画像のみから表情を認識することは，データ収集の難しさ，表情判定の難しさ，収集データと実環境との乖離，表情と感情とのずれなどの問題から，未だ実環境下で十分な性能を発揮する認識手法が確立されたとは言いがたく，今もなお活発に研究がなされており，また，人間の表情を治療に役立てようとする試みとしては，例えば様々な表情の合成顔画像から脳障害の程度を評価しようとする試み [36]や，人の情動変化に起因する表情の変化を依存症の診断に役立てようという試み [37]などがある．

そのような人間の表情の中でも特に笑顔は，我々の日常生活においてコミュニケーションや社会活動を円滑に進める上で重要な役割を果たす表情の 1 つであり，笑顔が高速・高精度に認識できることは， Human Machine Interface（HMI）やデジタルカメラなどの既存のアプリケーションにとって有用なのは勿論，将来的には医療や福祉など更に多くの応用分野で活かされる可能性がある．事実，笑顔はリハビリテーションの一環である音楽療法中の患者の症状の改善過程に伴っても頻繁に観察されるようになるポジティブな表情変化であることが経験的にわかっている．しかし，このように主観的には笑顔の発生頻度や度合の変化が音楽療法の効果と何らかの関係がありそうなことがわかってはいても，従来はそれを定量化して客観的に評価する術がなかった．そこで我々は，映像データのみから非接触・非拘束に人物の顔を自動的に検出し，その顔画像のみから笑顔度を推定し定量化する技術を応用し，音楽療法セッション記録映像から患者の笑顔度を定量的に測定し，療法経過や介入内容による変化を多重比較することによる音楽療法効果の客観的な評価方法について提案し，その有効性を評価した（図 4-1 参照）．

本章の構成であるが，4.2 節ではリハビリテーションにおける音楽療法の位置付けと表情との関連性ついて述べ，4.3 節では映像中から笑顔度を推定する手法について説明し，その性能評価結果について 4.4 節で述べる．そして，4.5 節で実際の臨床現場で撮影された映像から笑顔度を推定した結果と統計的検定による客観評価結果ついて報告し，4.6 節で本章のまとめを述べる．

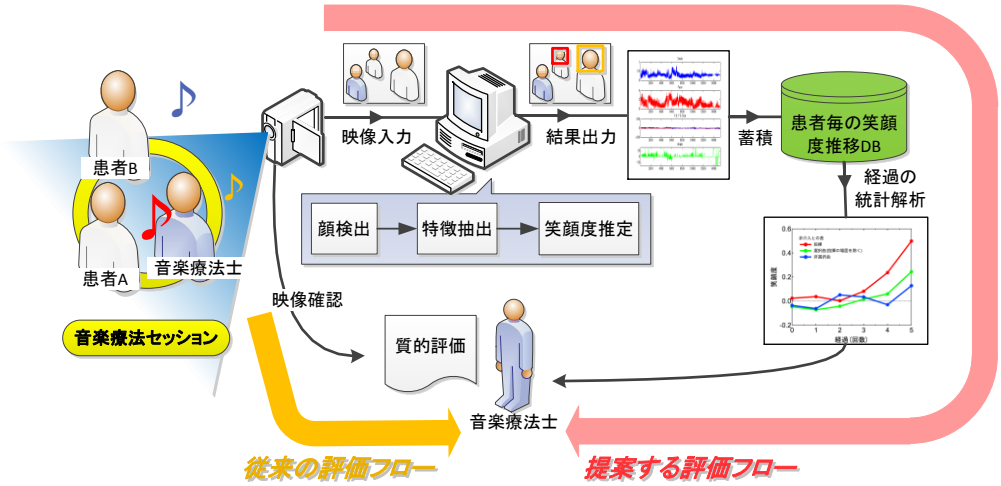


図 4-1　提案する音楽療法効果の評価フロー

4.2　音楽療法と表情の関連性

4.2.1　リハビリテーションにおける音楽療法

図 4-2 にリハビリテーションにおける音楽療法の一般的な役割とその活動構造を示す.

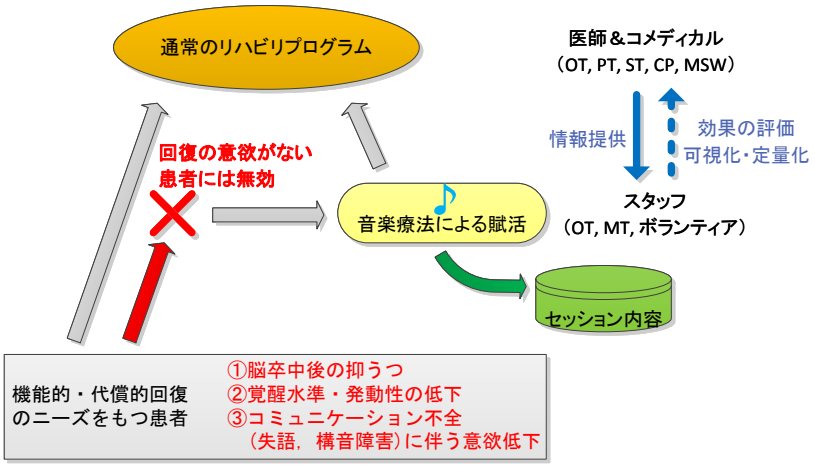


図 4-2　リハビリテーションにおける音楽療法の一般的な活動構造

そもそもリハビリテーションとは，疾病や外傷によって心身の機能が障害され，日常生活や社会参加に支障を来している個人が，複数の医療専門職の支援と連携の下で自身の問題を解決していくプロセスと捉えることができる．そこで個人が直面する問題はじつ

に様々であり，また各専門職が支援する分野や用いる手法も異なる．例えば，理学療法士や作業療法士は歩行や上肢動作の機能回復訓練を支援する．しかし，リハビリテーションの対象のうち脳機能障害患者などは，しばしば覚醒水準や発動性の低下や抑うつを呈し，情動反応も低調である．このため内発的な回復意欲に欠け，リハビリテーションの効果が上がりにくい場合がある．こうした症例に対して患者の覚醒水準や発動性の向上を期待して音楽療法が行われている．一般的な音楽療法は図 4-3 に示すようなプロセスサイクルに従って計画的に実行され [38]，我々は，病院内でこのようなプロセスに従い小集団の音楽療法を行い，患者の情動反応や発動性の向上を促して機能回復訓練へとつなげる取り組みを十数年にわたって行ってきた．

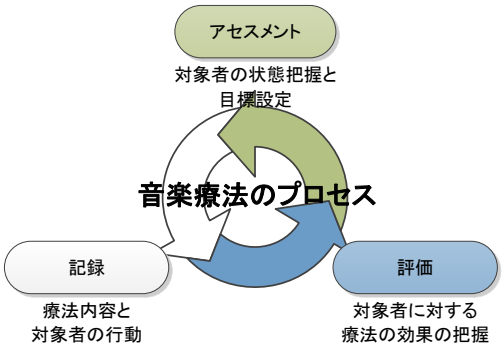


図 4-3　音楽療法の一般的なプロセス

4.2.2　療法効果の評価

このような取り組みを効果的に進めるためには，覚醒水準，発動性，情動反応のような精神機能面での対象者の変化を異なる複数の医療専門職の間で的確に共有することが重要である．そのため，現状では予め用意した主観評価表などを用いた点数化を行っているが，症例によっては評価項目の妥当性や，評価者の熟練度による点数の信頼性などの問題がある．また，作業療法士らが音楽療法の現場に立ち会って担当患者の発動性の改善を直接確認する例も，取り組みの効果が認識されるにつれて増えつつあるが，このような立会いを日常的に行うことは難しい．故に，こうした取り組みの効果を広く一般の医療関係者に伝達するためには，異なる症例でも共通に適用可能な客観的評価手法を通じて，その介入効果を検証することが必要である．

4.2.3　表情の活用

精神機能を客観的に評価する手法は今なお殆ど提案されていない一方で，喜び，悲し

み，怒り，驚き，嫌悪などの情動と強く結びついて表出される表情は人種や文化を超えて広く共通していることが知られている [34]．そこで我々は，表情を通じて覚醒水準や発動性の向上に伴って患者に現れる情動反応を客観的に評価できれば，こうした精神機能を評価する有力な手段になると考えた．

音楽療法では，同じ音に対して状況によって対象者が異なる情動反応を示すこともしばしばあり，療法士はその反応に常に注意深く向き合いつつ，望ましい行動を最大限に引き出すために介入の質と量を調整しなければならない．表情の中でも笑顔は快感情の表出である点で他の表情とは明確に異なり，この点が介入において非常に大きな手がかりを与えている．実際，快感情すなわち笑顔を誘起するように介入を調整することによって，対象者に深刻な悪影響を及ぼす事態を回避できるばかりか，望ましい行動を引き出せる見込みは高まる．そのため，ほとんどの音楽療法は不快感情ではなく快感情を誘起するように，またそれが対象者の達成すべき目標行動の獲得に合致するように設計される [38]．また，怒りや悲しみ等の不快感情の低減を評価尺度に用いる場合に比べると研究デザインが単純で済むため，評価結果の解釈に曖昧さが少ないという利点もある．こうした背景から我々は，音楽療法効果の評価尺度として笑顔度が様々な現場で広く利用できると考えた．

4.3 SVM による笑顔度推定手法

本研究では，対象者に先入観や警戒感を極力抱かせないため，特別な器具や装置の装着を必要とせず非侵襲・非接触・非拘束に笑顔度を推定する方法として，ビデオカメラ等で撮影された映像データのみから対象者の笑顔度を推定することを試みた. 一般的に，笑顔を含む表情を映像情報のみから認識する手法は 2 つに大別される．1 つはアピアランス・ベースの手法 [39] [40] [41] [42]で，もう 1 つは特徴点ベース（又はモデル・ベース）の手法 [43] [44]である．アピアランス・ベースの手法は，特徴点ベースの手法に比べ，顔を検出した後に目，口や鼻といった顔器官を検出する必要がないため処理が単純で高速であるが，検出された顔の位置ずれ，顔向きの変化や個人ごとの顔器官配置の違いの影響を受けやすい．一方，顔検出後に顔器官を検出する特徴点ベースの手法は，処理は複雑になるが特徴点周辺の詳細な解析，顔の検出位置ずれや顔向きの補正が可能なため, 一般的にはアピアランス･ベースの手法に比べて表情の検出性能が頑健であるが，性能が顔器官の検出性能・精度に依存するなどの問題点もある．本研究では，画像認識技術が広く一般的に利用されるためには，プライバシー保護の観点から映像データをストレージ等に保存することなくリアルタイムに処理することが重要な課題の一つの考え，

高速かつロバストな笑顔度の推定を目指し，カラー情報を使わず又顔器官等の特徴点を抽出する必要がないアピアランス・ベースの特徴抽出法をベースに多少の顔の検出位置ずれや顔向き変化にも対応可能な特徴抽出法について検討した．図 4-4 に本提案手法の全体処理フロー概要を示す．はじめに 256 階調グレイ・スケールの入力画像から顔を検出し，検出された顔領域から顔画像を切り出し，スケーリングと照明の影響を低減するための正規化を施した後にアピアランス特徴量を抽出し，笑顔の検出/笑顔度の推定を行う．

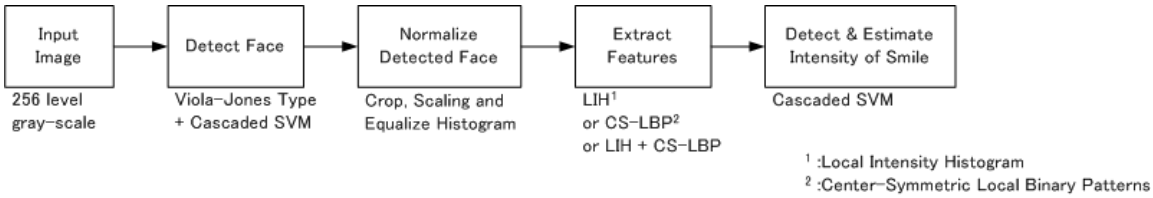


図 4-4　全体処理フロー概要

4.3.1　　顔の検出

映像から笑顔度を推定する第 1 ステップとして，入力された映像から対象者の顔を検出する必要がある．本研究では，図 4-5 に示すように，顔検出器として Haar-like 特徴量と AdaBoost による Viola-Jones 型の顔検出器 [13]とカスケード構造化された 2 つの SVM による顔検出器を組み合わせた，高速かつロバストな顔検出器 [45]を用いた．

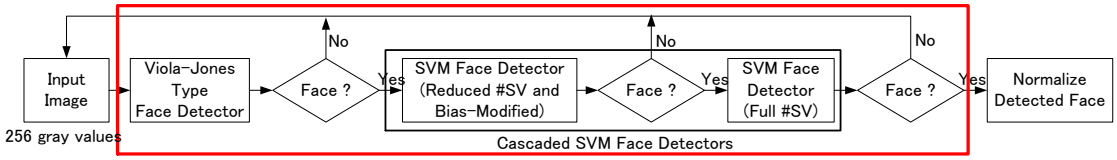


図 4-5　顔検出処理フロー

Viola-Jones 型の顔検出器はカスケードの段数を適当に設定することで，誤検出をある程度許容しつつも高速な顔検出が行える．一方, SVM は識別性能が高いことで知られるが, 顔検出のような非線形問題へ対応するためにカーネル･トリックを用いた場合, 探索数, サポート・ベクタ数，次元数に従って計算コストが膨大となる．そこで，はじめに入力画像全体から Viola-Jones 型の顔検出器により高速に顔候補を絞り込み，検出された高々数個の顔候補に対してのみ SVM 顔検出器により顔検出を実施する．さらに SVM 顔検出

器を，Reduced Set Method 法 [30] [31]を用いて SV 数を削減した少数のサポート・ベクタからなる副顔検出器と全サポート・ベクタを用いた主顔検出器の 2 段カスケード構造とすることで，SVM 顔検出器の性能を維持しつつ更なる計算速度の向上を図った．また本研究では，学習用画像として水平方向へ±30 度程度以内の顔画像を用いることで，真正面以外の多少の顔向き変化に対しても顔検出が可能なように顔検出器を学習した．そして，この顔検出器により検出された顔の位置とサイズを元に，入力画像から顔領域のみを切り出し，サイズを 40×40 画素にスケーリングし，照明環境の影響を抑えるためにヒストグラムの平滑化を施した後に，後述する笑顔度推定のための特徴抽出処理を施す．

4.3.2 笑顔度の推定

顔検出器によって検出された顔画像に対し，笑顔度を推定するための特徴量を抽出する．本研究では，目，鼻や口などの顔器官を抽出する必要がないアピアランス・ベースの特徴を笑顔度推定のための特徴量として抽出し，その特徴量を入力として SVM による笑顔検出器を構築する．そして，その SVM 笑顔検出器からの出力値を事後確率に変換することで笑顔度を算出する．図 4-6 に笑顔検出及び笑顔度推定の処理フローを示す．

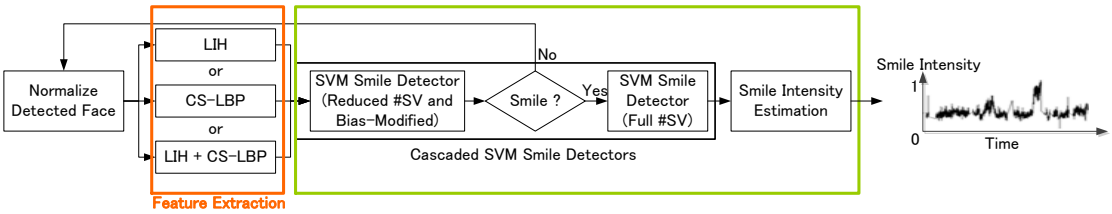


図 4-6 笑顔検出及び笑顔度推定処理フロー

一般的に顔検出器により検出された顔は，目や口などの顔器官の位置が常に一致している保証はない．そのため，顔のアピアランス情報のみから精度よく笑顔を検出するためには，顔の検出位置ずれ等に対してロバストな特徴抽出法が必要である．そこで本研究では，顔画像をある一定の大きさの格子セルに分割し，各格子セル内で局所的な特徴量を算出する手法である LIH 特徴量（2.1.2 節），CS-LBP 特徴量（2.1.4 節）及びこれら 2 つを連結した統合特徴量である LIH + CS-LBP 特徴量（2.1.5 節）を笑顔度推定のための特徴量として検討した．そして特徴量が抽出された後に，その特徴量を用いて笑顔か非笑顔かの識別を行う．本研究では，笑顔検出器としてガウス・カーネルによる非線形 SVM を用い SVM 笑顔検出器の出力yが，$y \geq 0$ときに笑顔，$y < 0$のときに非笑顔と判

定する．ここで SVM 笑顔検出器についても SVM 顔検出器と同様に，性能を維持しつつ計算量を大幅に削減するためにカスケード構造とした．そして最後に，笑顔検出の結果から笑顔度I_sを推定する．本研究では笑顔検出器の出力yから式(4.1)で示されるシグモイド関数を用いて事後確率へと変換することで笑顔度I_sを算出する．

$$I_s = \frac{1}{1+e^{-\lambda y}} \tag{4.1}$$

ここで，ゲインλは笑顔度の推定感度に相当し，本研究では$\lambda = 5$として後述の実験を実施した．

4.4 笑顔検出性能の評価

4.4.1 評価用画像データについて

本研究の評価用画像データとしては，独自に収集し構成した独自画像データベースと研究用に公開されている研究用公開画像データベースを各々評価目的に応じて利用した．独自画像データベースは，Web 上の公開画像，TV 番組の録画映像，顔画像の収録を行い収集した画像群に対し，性能評価の対象を本提案手法で検出可能な顔画像に絞るために，4.3.1 節で述べた手法により顔検出を行い，正しく顔が検出された画像のみを抽出して構築した．ここで，本提案手法の顔検出器は，例えば真横向き，つまり片方の目が完全に隠れてしまっているような“横顔”に対しては，顔を安定して検出することが困難であるが， 4.3.1 節で述べたように，ほぼ真正面を向いている“正面顔”に加えて，水平方向に±30 度程度までの“斜め顔”の顔画像を用いて顔検出器を学習しているため，大凡学習画像と同程度の範囲内での顔向き変化に対しては顔検出が可能である．このような顔検出器によって検出された顔画像群に対し，更に“正面顔”/“斜め顔”の分類を目視で行い，“正面顔”の笑顔/非笑顔画像データベース（笑顔：2,620 枚，非笑顔：3,620 枚）と“斜め顔”の笑顔/非笑顔画像データベース（笑顔：2,260 枚，非笑顔：4,170 枚）を各々構築した．この時，“正面顔”と“斜め顔”との目視分類の基準は，片方の眼の目尻が顔の輪郭に達している場合は“斜め顔”と判定し，それ以外は“正面顔”とした．

このようにして構築した“正面顔”の笑顔/非笑顔画像データベース及び”斜め顔”の笑顔/非笑顔画像データベースのサンプル画像を図 4-7 に示す.

図 4-7　独自画像データベースのサンプル画像（上段：“正面顔”の笑顔/非笑顔画像，下段：“斜め顔”の笑顔/非笑顔画像）

更に，提案した笑顔検出手法の汎化性能等を評価するため，前述の独自画像データベースに加え，評価目的ごとに以下に記す研究用公開画像データベースを併せて利用した.

- CK+データセット [46]

  CK+データセットは，表情画像データベースとして研究用途に広く利用されてきた Cohn-Kanade AU-Coded Facial Expression Database [47]を更に拡張し，18 歳から 50 歳の被験者 123 人に対する表情画像データが，無表情の状態から表情が表出されるまでの連続した画像として 593 シーケンス収められており，またその内 327 シーケンスについては Emotion ラベル（Angry, Contempt, Disgust, Fear, Happy, Sadness, Surprise）も同時に提供されている.

- The MPLab GENKI Database, GENKI-4K Subset [48]

  GENKI-4K データベースは，笑顔／非笑顔ラベルが付与された計 4000 枚（笑顔：2162 枚，非笑顔：1838 枚）の画像から構成される笑顔画像データベースである．画像が Web ベースで収集されているため，人種や年齢といった人物属性のバリエーションが豊富で，かつ様々なカメラや場所で撮影されているため多撮像環境となっている．そのため，笑顔検出器の汎化性能を調査するのに有用である．また，画像内の人物の顔姿勢は，大半が Yaw，Pitch，Roll が±20 度以内とされており，本研究の評価目的に合致している.

- Facial Expression and Emotion Database (FEED) [49]

  FEED データベースは，笑顔を含め，怒りや恐れなど Eckman & Friesen [34] [50] が定義した 6 つの表情の画像がラベル付きで用意されている．さらにこれらの表情

が 18 人の人物に対して無表情からの画像シーケンスとして提供されているため, 正面向きの顔のみではあるが笑顔の検出に加え，笑顔の推移を評価することが可能である.

4.4.2 特徴量及びカスケード型笑顔検出器の評価

はじめに特徴量ごとの笑顔検出性能の評価と，カスケード型笑顔検出器の性能評価を実施した．ここで，検出性能は 5 分割交差確認法によるオープンな評価を実施し，ROC（Receiver Operating Characteristic）曲線の AUC（Area Under the Curve）を算出して性能を比較した.

LIH 特徴量の性能評価

LIH 特徴量の抽出において，入力画像サイズ，格子セル数，ヒストグラムビン数の笑顔検出性能に与える影響を調べるため，“正面顔”の笑顔/非笑顔画像データベースを用いて，以下の評価実験を実施した．はじめに，入力画像サイズの性能に与える影響を調べるため，入力画像サイズを 20×20 画素又は 40×40 画素としたときの性能を比較した（図 4-8 参照）. 格子セル数が 4×4, 5×5 の双方において入力画像サイズ 20×20 画素に比べ 40×40 画素の方が優位な性能を示すことがわかる.

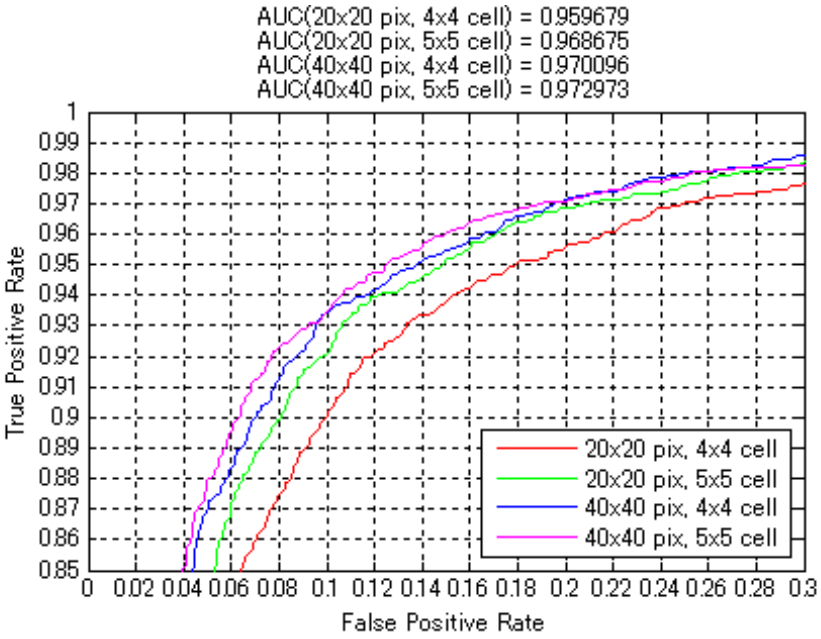


図 4-8 入力画像サイズによる性能比較（格子セル数 :4×4 又は 5×5, ヒストグラムビン数 :8）

次に, 格子セル数の性能に与える影響を調査するため, 入力画像サイズを 40×40 画素, ヒストグラムビン数を 8 に各々固定し, 格子セル数を 4×4, 5×5, 8×8, 10×10 と変化させて場合の各々の性能について評価した結果を図 4-9 に示す.

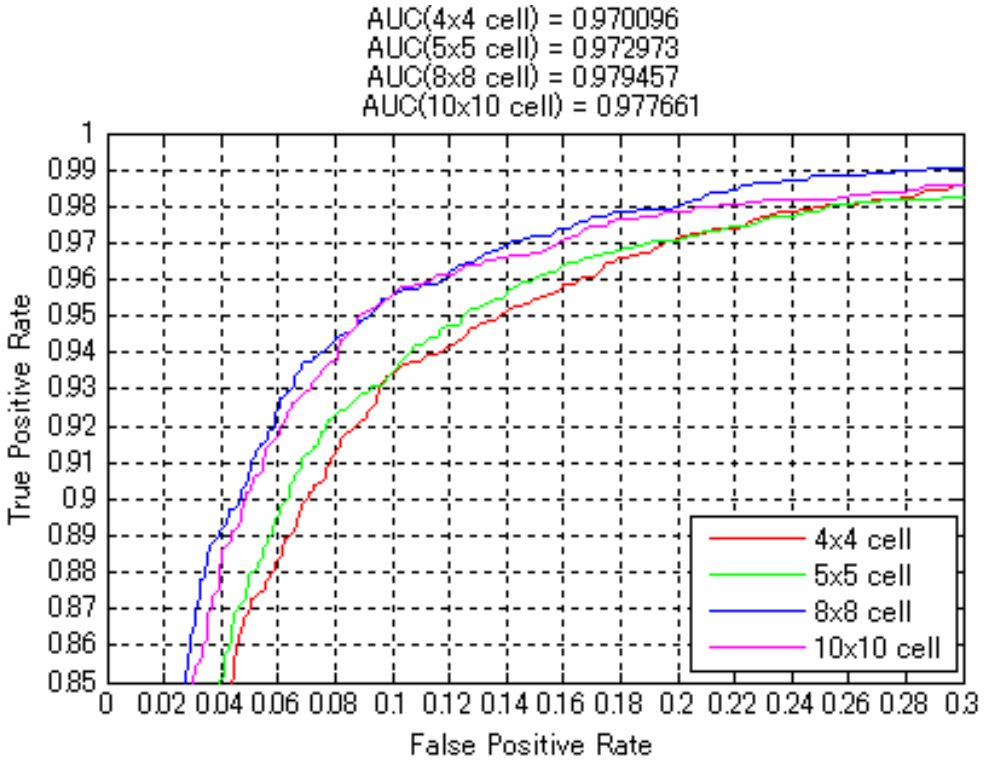


図 4-9 格子セル数による性能比較（入力画像サイズ :40×40, ヒストグラムビン数 :8）

格子セル数が多くなるほど性能が良くなる傾向にあり，格子セル数が 8×8 と 10×10 の場合においてほぼ同程度の検出性能を示したが，格子セル数が 8×8 のときに最も良い性能を示した.

最後に，ヒストグラムビン数の影響を調べるため，入力画像サイズを 40×40 画素，格子セル数を 8×8 に各々固定し，ヒストグラムビン数を 4, 8, 16 と変化させた場合の評価結果を図 4-10 に示す.

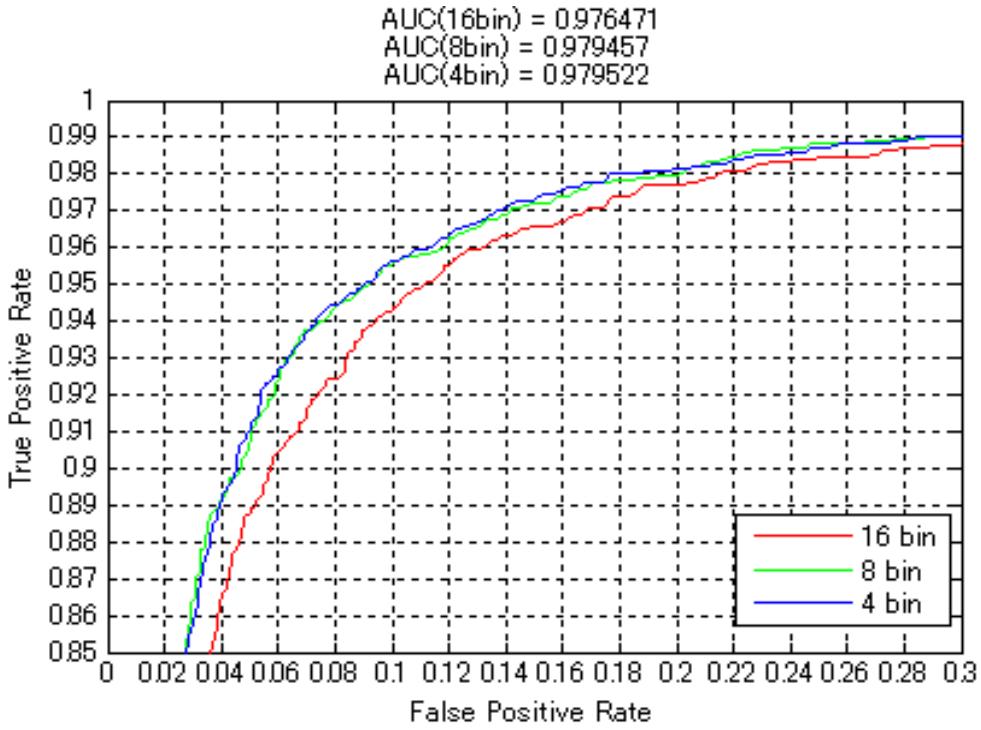


図 4-10　ヒストグラムビン数による性能比較 （入力画像サイズ :40×40, 格子セル数 :8×8）

ビン数が 4 と 8 で同程度の性能を示しており，このことは，笑顔検出のために必要な顔画像の階調は高々4 階調で十分なことを示している．以上の評価実験から，LIH 特徴量の笑顔検出における最適パラメータは,

- 入力画像サイズ　　: 40×40 画素
- 格子セル数　　: 8×8
- ヒストグラムビン数　　: 4

であることがわかり，このときの特徴量の次元数は，256 次元となった.

CS-LBP 特徴量の性能評価

次に，CS-LBP 特徴量についても同様に特徴量抽出時の可変パラメータである，入力画像サイズ，格子セル数，エンコード閾値の笑顔検出性能に与える影響を調べるため，“正面顔”の笑顔/非笑顔画像データベースを用いて，以下の評価実験を実施した．

はじめに，入力画像サイズの性能に与える影響を調べるため，入力画像サイズを 20×20 画素又は 40×40 画素としたときの性能を比較した結果を図 4-11 に示す．LIH 特徴量の場合と同様に，格子セル数が 4×4，5×5 の双方において入力画像サイズ 20×20 画素に対し 40×40 画素の方が優位な検出性能を示した．

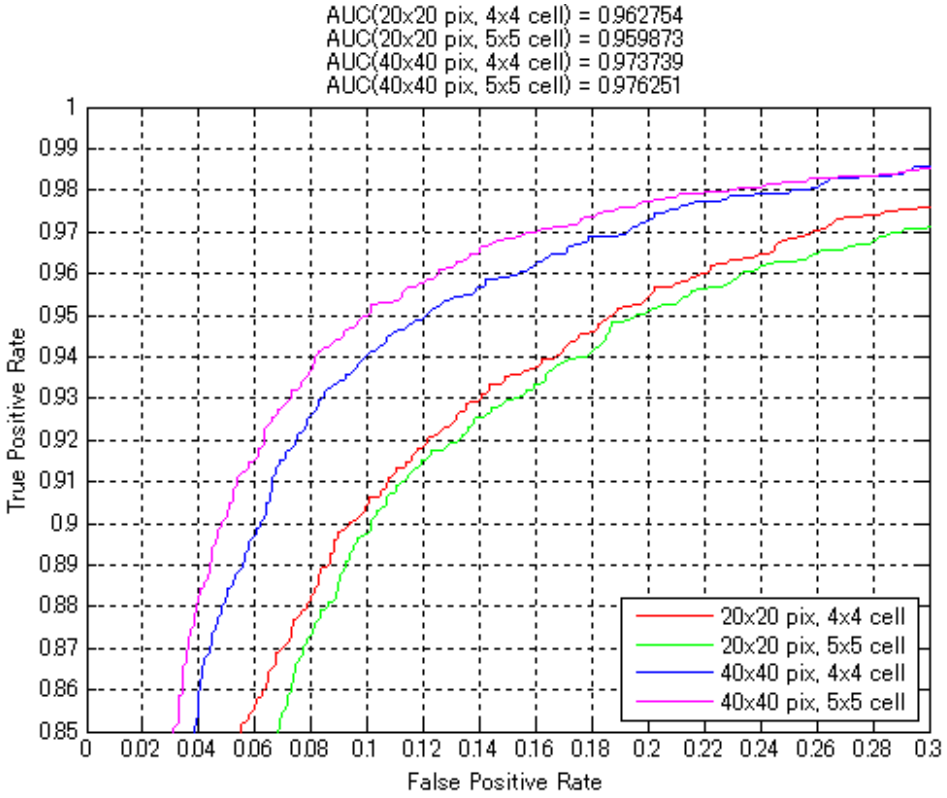


図 4-11　入力画像サイズによる性能比較（格子セル数 :4×4 又は 5×5, 閾値 :0.00）

次に，格子セル数の性能に与える影響を調べるため，格子セル数を 2×2，4×4，5×5，8×8 と変えた時の AUC の比較を図 4-12 に示す．LIH 特徴量と同様に格子セル数を増やすと徐々に性能が良くなる傾向にあるが，分割数を増やしすぎると逆に性能が劣化しており，本実験より，格子セル数 5×5 のときに最も性能が良いことがわかった．

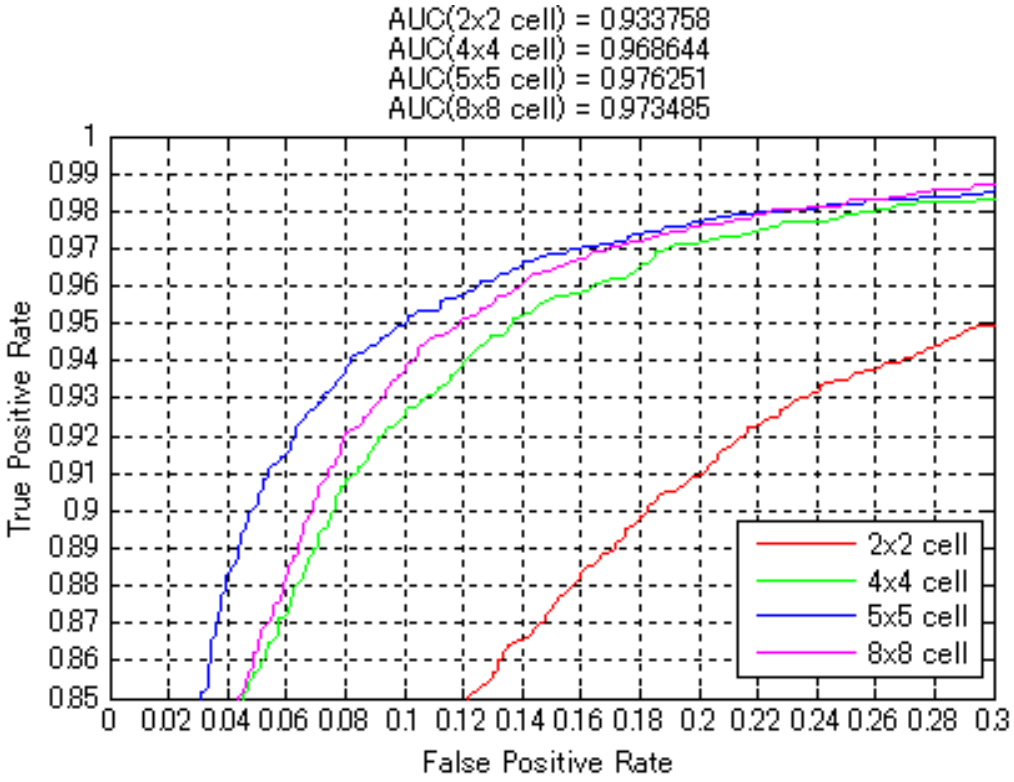


図 4-12　格子セル数による性能比較（入力画像サイズ :40×40，閾値 :0.00）

最後に，入力画像サイズを 40×40 画素，格子セル数を 5×5 に各々固定し，エンコード閾値を 0.00，0.02，0.05 と変化させたときの性能比較を図 4-13 に示す．格子セル数を変化させたときほどの大差はないが，閾値が小さいとノイズに過敏になり，また大きいとテクスチャ特徴をうまく表現できなくなっていると考えられ，本実験においてはエンコード閾値が 0.02 のときに最も良い結果を示した．

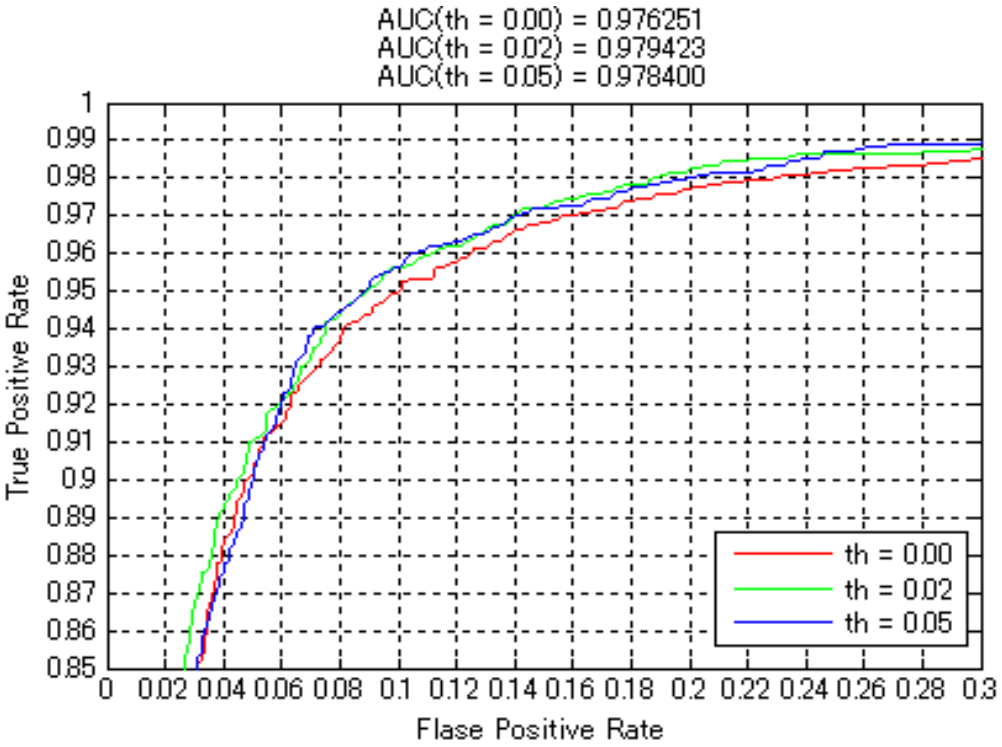


図 4-13　エンコード閾値による性能比較（入力画像サイズ :40×40, 格子セル数 :5×5）

以上の評価実験から，CS-LBP 特徴量の笑顔検出における最適パラメータは，

- 入力画像サイズ　　: 40×40 画素
- 格子セル数　　: 5×5
- エンコード閾値　　: 0.02

であることがわかり，このときの特徴量の次元数は 400 次元である．

LIH + CS-LBP 特徴量の性能評価

次に，LIH と CS-LBP を連結統合した LIH + CS-LBP 特徴量による評価を実施した．このとき，前述の評価実験により，入力画像サイズは 40×40 画素とし，また各特徴量のパラメータは最も性能が良かった以下の値とし，“正面顔”の笑顔/非笑顔画像データベースを用いて評価実験を実施した．その結果を図 4-14 に示す．

- LIH 特徴量
  - 格子セル数　　：8×8
  - ヒストグラムビン数　　：4
- CS-LBP 特徴量
  - 格子セル数　　：5×5
  - エンコード閾値　　：0.02

LIH + CS-LBP 特徴量は，256＋400＝656 次元と次元数は増えるもののAUC = 0.982269と LIH 特徴量，CS-LBP 特徴量各々を単独で用いたときよりも検出性能が向上していることがわかる．

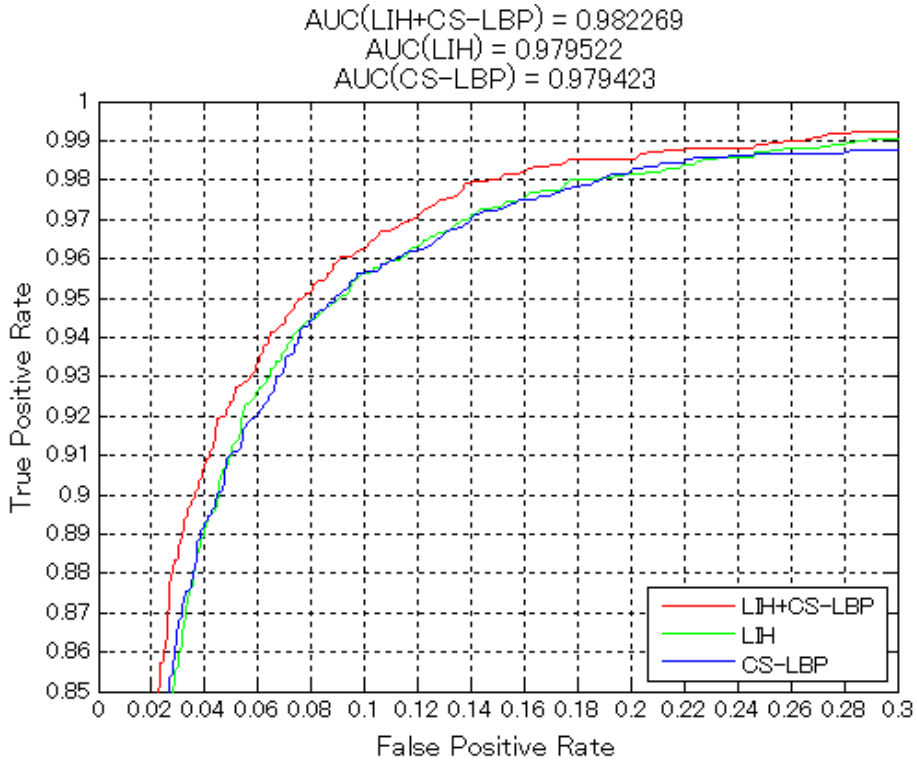


図 4-14　“正面顔”の笑顔/非笑顔画像データベースを用いた各特徴量の性能比較

カスケード型笑顔検出器の性能評価

次に，笑顔検出器を Reduced Set Method 法により SV 数を削減した副笑顔検出器と全サポート・ベクタを用いた主笑顔検出器により構成される 2 段カスケード構造としたときの検出性能と処理速度を調査した．ここで，評価に用いた画像データは，“正面顔”の笑顔/非笑顔画像データベースと GENKI-4K データベースを併せて用い，特徴量は特徴量ごとの性能比較で最も性能の良かった LIH + CS-LBP 特徴量とした．

図 4-15 に全サポート・ベクタによる検出器と，Reduced Set Method 法により SV 数を各々32，64，128，256，512 に削減した検出器との性能比較を示す．当然ながら，SV 数が少なくなるほど性能が劣化していることがわかる．

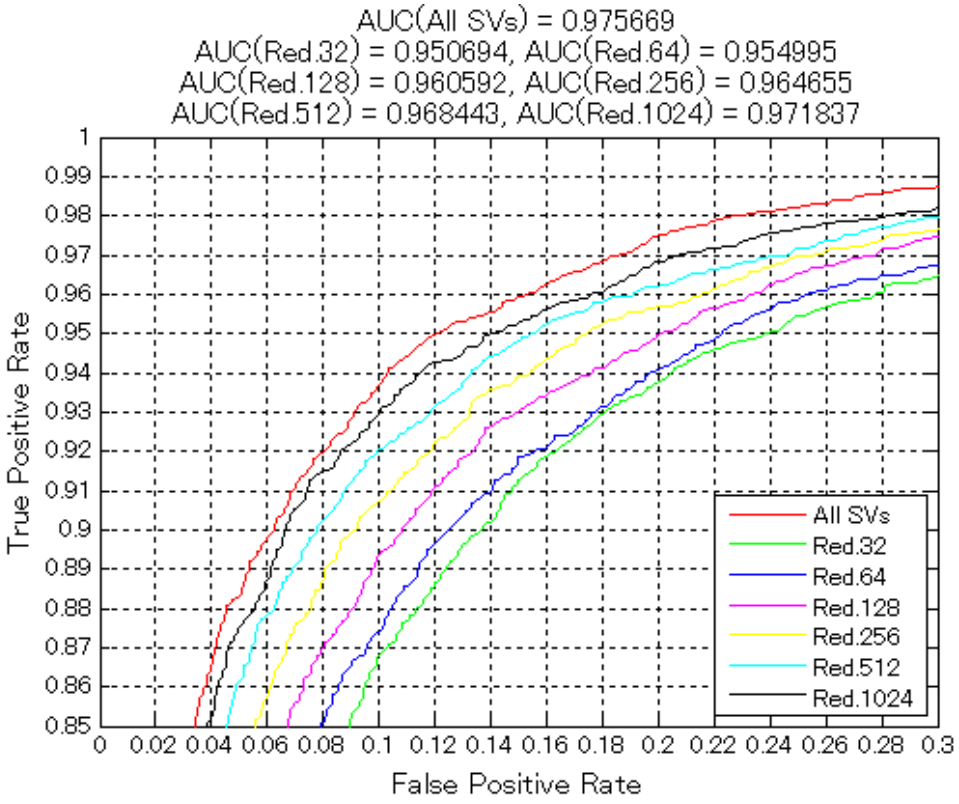


図 4-15　サポート・ベクタ数による性能比較

次に，Reduced Set Method 法により SV 数を 32，64，128，256，512 に削減した検出器を 1 段目に，全サポート・ベクタによる検出器を 2 段目としたカスケード構造としたときの性能比較を図 4-16 に示す．同様に SV 数を削減すると性能が劣化しているが，カスケード構造としたことで，単に SV 数を削減したときよりも検出性能が向上していることがわかる．

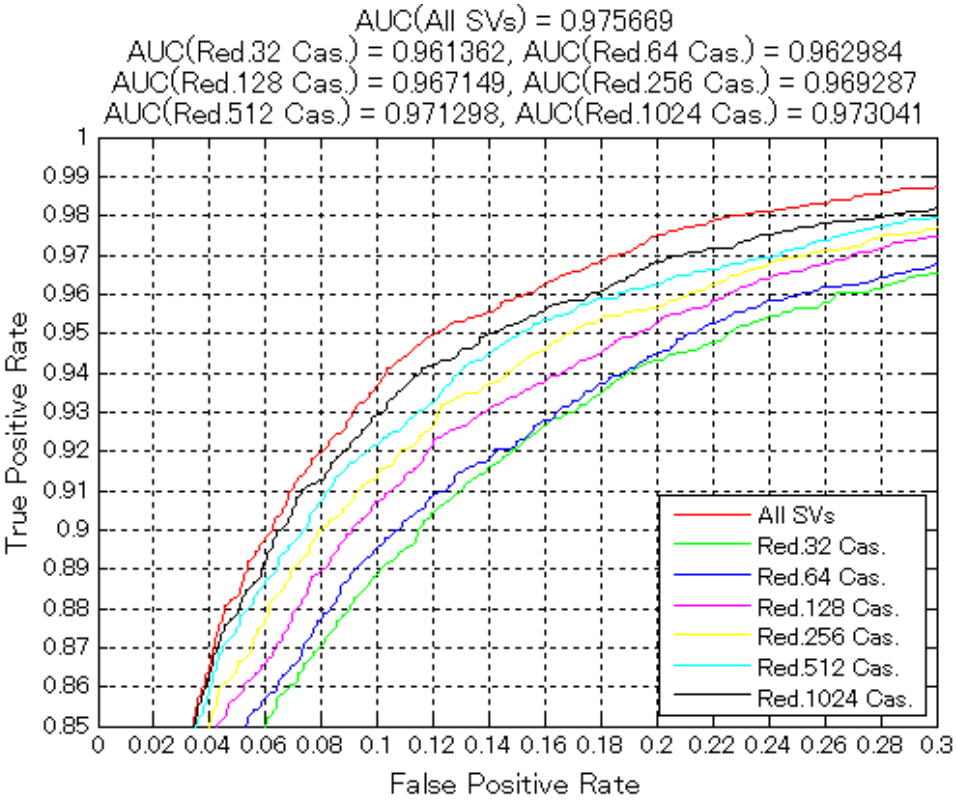


図 4-16　副笑顔検出器のサポート・ベクタ数による性能比較

次に，1 段目で笑顔候補が未検出となるのを極力防ぐため，全サポート・ベクタによる検出器と同じ検出率を達成するように事前に 1 段目の検出器の閾値を調整した上でカスケード構造とした場合の結果を図 4-17 に示す．閾値を調整したことで，1 段目で笑顔候補が未検出となるケースが減ったため，単にカスケード構造にした時よりも若干検出性能が向上していることがわかる．

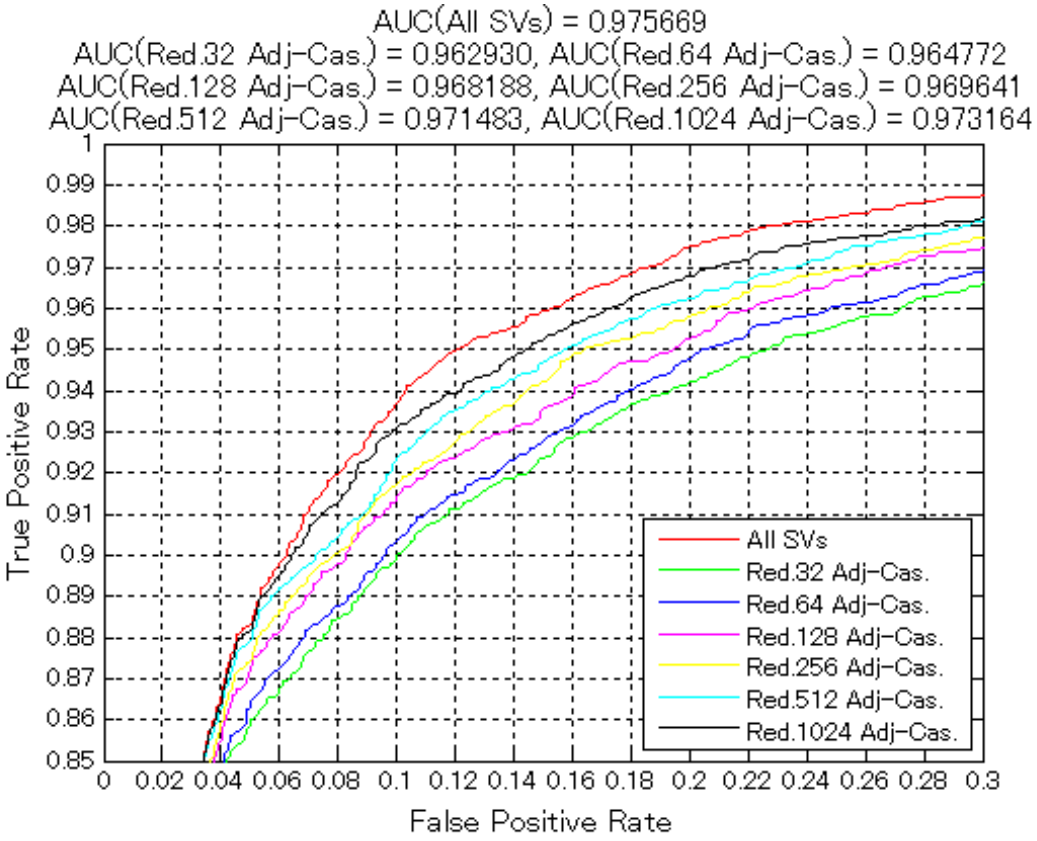


図 4-17　バイアス項を調整した場合の副笑顔検出器のサポート・ベクタ数による性能比較

最後に各実験におけるサポート・ベクタ数，笑顔検出性能（AUC），処理速度の比較を表 4.1 に示す.

表 4.1　笑顔検出性能（AUC）と処理時間のサポート・ベクタ数による比較

Classifier	＃SVs	AUC (LIH+CS-LBP)	CPU Time (msec)
Normal SVM（with all SVs）	3567	0.975669	17.3698
Reduced SVM	32	0.950694	0.1275
	64	0.954995	0.2584
	128	0.960592	0.4116
	256	0.964655	0.8694
	512	0.968443	2.2068
	1024	0.971837	4.8441
Cascaded SVMs	32 (& 3567)	0.961362	7.5823
	64 (& 3567)	0.962984	7.7662
	128 (& 3567)	0.967149	8.0570
	256 (& 3567)	0.969287	8.4609
	512 (& 3567)	0.971298	9.7745
	1024 (& 3567)	0.973041	12.3801
Cascaded SVMs（Bias-adjusted）	32 (& 3567)	0.962930	8.6267
	64 (& 3567)	0.964772	8.5697
	128 (& 3567)	0.968188	8.5719
	256 (& 3567)	0.969641	8.8934
	512 (& 3567)	0.971483	10.1280
	1024 (& 3567)	0.973164	12.7219

on Matlab @ 3.0GHz Core 2 Quad

カスケード構造とした Cascaded SVMs 及び閾値調整した後にカスケード構造とした Cascaded SVMs (Bias-modified) の双方とも，副検出器としてサポート・ベクタ数が 512 以上の時に，$AUC > 0.97$という高性能を維持しつつ，計算量を約 30～40 %程度削減可能なことがわかり，本手法により計算時間を削減しつつ全サポート・ベクタを用いたときに匹敵する性能を維持できることが確認できた.

4.4.3　笑顔検出の汎化性能評価

本研究では，対象者が通常の音楽療法活動が可能なように非接触・非拘束な状態で撮影された映像データを用いる．そのため，対象者の顔向きが撮影環境，活動内容や姿勢に応じて変化することが想定される．また病院や福祉施設などで実際に運用されることを目指す上では不特定多数の対象者に対して笑顔が検出可能な汎化性能が求められる.

顔向きに対する汎化性能評価

本節では顔向きに対する笑顔の検出性能を評価することを考える．ここで評価用の画像データベースとしては，“正面顔”の笑顔/非笑顔画像データベース及び“斜め顔”の笑顔/非笑顔画像データベースを用い，5 分割交差確認法により顔向きごとに笑顔検出性能のオープンな評価を実施した．本評価では特徴量として，LIH 特徴量，CS-LBP 特徴量，そしてそれら 2 つを連結した LIH + CS-LBP 特徴量を用いて，笑顔検出性能を ROC 曲線の AUC にて比較する．ここで，各々の特徴量の特徴抽出パラメータは予め実施した基礎評価実験（4.4.2 節）により最も高い性能を示した以下の値とした．

- LIH 特徴量
  - 格子セル数 :8 × 8
  - ヒストグラムビン数 :4
  - 次元数 :256
- CS-LBP 特徴量
  - 格子セル数 :5 × 5
  - 近傍画素数 :8
  - 半径 :1
  - エンコード閾値 :0.02
  - 次元数 :400
- LIH + CS-LBP 特徴量
  - LIH と CS-LBP 各々のパラメータは上記に従う
  - 次元数 :656

LIH，CS-LBP，LIH + CS-LBP 各々を特徴量とした場合の“正面顔”に対する笑顔検出 ROC 曲線を図 4-18 に“斜め顔”に対する笑顔検出 ROC 曲線を図 4-19 に各々示す．
“正面顔”の笑顔検出性能に着目すると，何れの特徴量を用いても $AUC > 0.97$ と非常に高い検出性能を示しており，中でも 2 つの特徴量を結合した LIH + CS-LBP 特徴量を用いたときに $AUC = 0.981039$ と最も高い性能を示した．一方，“斜め顔”の笑顔検出性能は，“正面顔”と比較すると各特徴量共に若干検出性能が劣るものの，何れの特徴量においても $AUC > 0.96$ を示し，また“正面顔”同様に LIH + CS-LBP 特徴量を用いた時に $AUC = 0.976259$ と最も高い検出性能を示した．ここで特徴量別に観察すると，CS-LBP 特徴量は他の特徴量に比べて“正面顔”と“斜め顔”の検出性能を比較した場合に劣化が大きいことから，顔向きに対しては CS-LBP 特徴量よりも LIH 特徴量の方が有効であると考えられる．これらの結果から，本研究では音楽療法効果の評価に用いる笑顔検出器を

構成する特徴量としては，顔検出が可能な水平方向に±30 度程度以内の顔向き変化において最もロバストな笑顔検出性能を示した LIH + CS-LBP 特徴量を用いることとした．

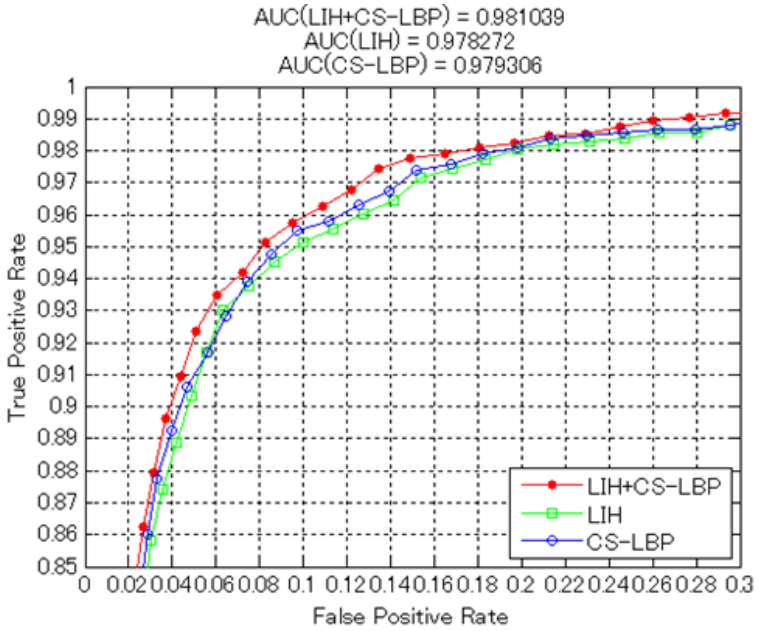


図 4-18 “正面顔”に対する笑顔検出 ROC 曲線の特徴量による比較

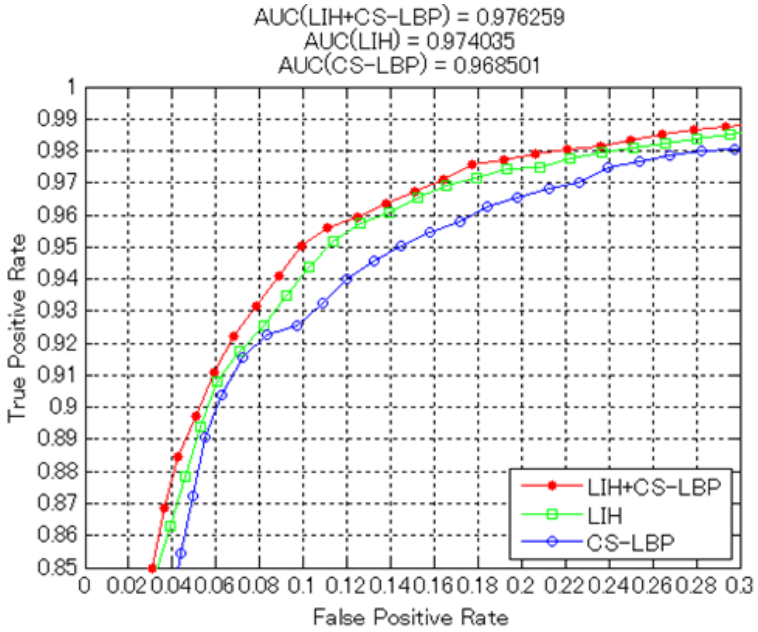


図 4-19 “斜め顔”に対する笑顔検出 ROC 曲線の特徴量による比較

汎化性能の評価と他手法との比較

本研究では, 提案手法が実際の臨床映像データに適用されることから, 不特定人の様々な表情に対して笑顔が検出できるよう高い汎化性能が求められる．ここで，汎化性能を検証するために研究用に公開されている画像データベースを評価用画像として利用し，また併せて本提案手法と他の手法との性能を比較評価することを考える．本評価では評価用画像として，表情画像データベースとして著名な Extended Cohn-Kanade Dataset（以下，CK+データセット） [46]と，大規模な笑顔画像データベースとして知られている GENKI-4K データベース [48]を利用する. また他手法との比較としては, アピアランス・ベースの表情検出手法 [41] [42]等でも用いられ高い表情検出性能を示すことが知られており, 画像認識の分野において代表的な画像特徴量の 1 つである LBP ベースの特徴量（2.1.3 節）を笑顔検出のための特徴量として用い，本研究で採用した LIH + CS-LBP 特徴量と検出性能を ROC 曲線により比較する. ここで, LBP ベースの特徴量としては [41] [42]でも提案手法との比較対象手法として用いられている Uniform LBP 特徴量 [11]を用い，特徴抽出パラメータは予め実施した基礎評価実験により最も高い性能を示した以下の値とした.

- Uniform LBP 特徴量
  - 格子セル数　　:5 × 5
  - 近傍画素数　　:8
  - 半径　　:1
  - 次元数　　:1,475

はじめに，多表情画像データを用いて様々な表情に対する笑顔検出器の汎化性能を評価するため CK+データセットを利用して評価を実施する．本評価では，本データセットに付与された Emotion ラベル（Angry, Contempt, Disgust, Fear, Happy, Sadness, Surprise）の情報を元に笑顔画像として Happy の Emotion ラベルが付与されている画像シーケンスの最終フレームの画像（69 枚）を，また非笑顔画像としては，Happy 以外の Emotion ラベルが付与されている画像シーケンスの最終フレームの画像（258 枚）と，無表情画像として各画像シーケンスの先頭フレームの画像（327 枚）を併せて用い，その結果，評価用画像群は笑顔画像 69 枚，非笑顔画像 585 枚の構成となった.

この評価用画像群に対して，4.4.1 節で説明した独自笑顔/非笑顔画像データベースの全画像を学習用画像として, LIH + CS-LBP 特徴量と Uniform LBP 特徴量を各々抽出して学習した笑顔検出器を用いて笑顔の検出性能を比較評価した結果を図 4-20 に示す.

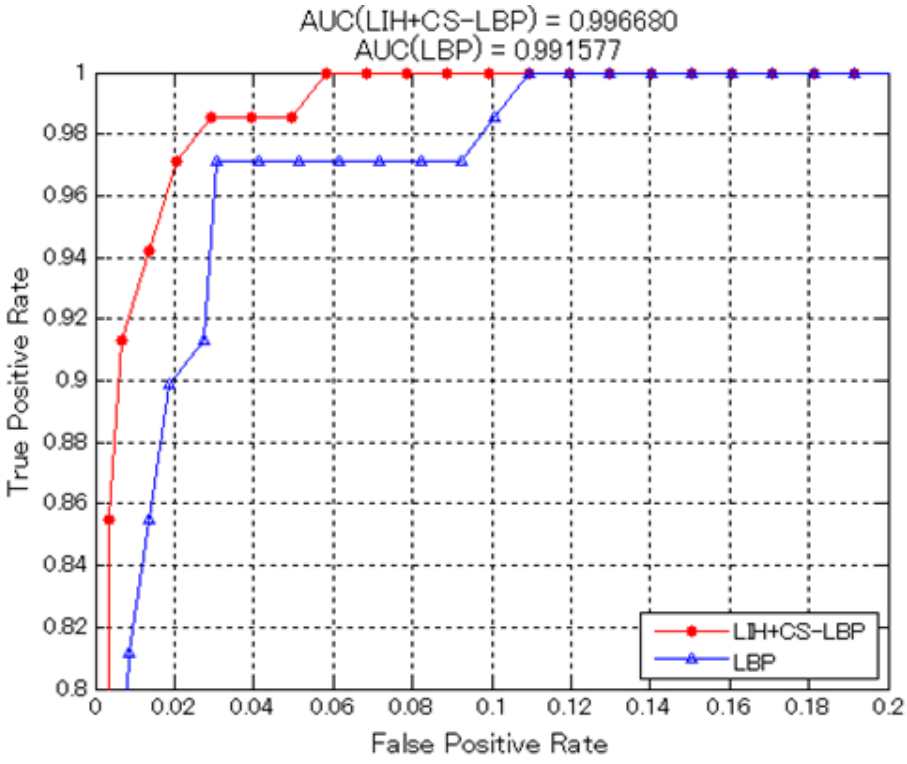


図 4-20　CK+データセットを用いた表情に対する汎化性能比較

何れの特徴量を用いても $AUC > 0.99$ と非常に高い笑顔検出性能が示されており，この結果から本研究で笑顔度の定量化手法として用いる LIH + CS-LBP 特徴量を用いた場合においても，従来手法と同様に笑顔以外の表情を非笑顔として高精度に検出できることを確認した．また例えば，多方向の顔画像に対して高い表情検出性能が報告されている[42]で用いられている LGBP 特徴量と比較すると，16 サブブロックの LGBP 特徴量を用いた場合に $AUC = 0.99556$ と提案手法と同等の検出性能を示すことを確認したが，この時, 特徴量の次元数は 37,760 次元と 656 次元の LIH + CS-LBP 特徴量に比べ 60 倍弱であり，臨床応用を考えた場合に重要な要素の一つである笑顔検出のリアルタイム性の確保という点においては課題があると考える．

次に，不特定人に対する笑顔検出器の汎化性能を評価する目的で GENKI-4K データベースを利用する．本データベースは，CK+データセットのように実験室等の整えられた

撮影環境下で収集された画像ではなく，Web ベースで画像が収集されており多人種，多撮像環境，多解像度の画像かつ自然な表情でデータベースが構成されているため，笑顔検出器の汎化性能を評価するのに適している．この GENKI-4K データベースの全画像データ 4,000 枚の内，本研究の顔検出器で顔が検出できた 3,653 枚（笑顔：2,010 枚，非笑顔：1,643 枚）を，4.4.1 節で説明した独自画像データベースに追加し，より大規模な笑顔/非笑顔画像データベース（笑顔：6,890 枚，非笑顔：9,433 枚）を構築した．そして，この大規模笑顔/非笑顔画像データベースを対象として，LIH + CS-LBP 特徴量，Uniform LBP 特徴量各々について 5 分割交差確認法にてオープンな比較評価を実施した結果を図 4-21 に示す．

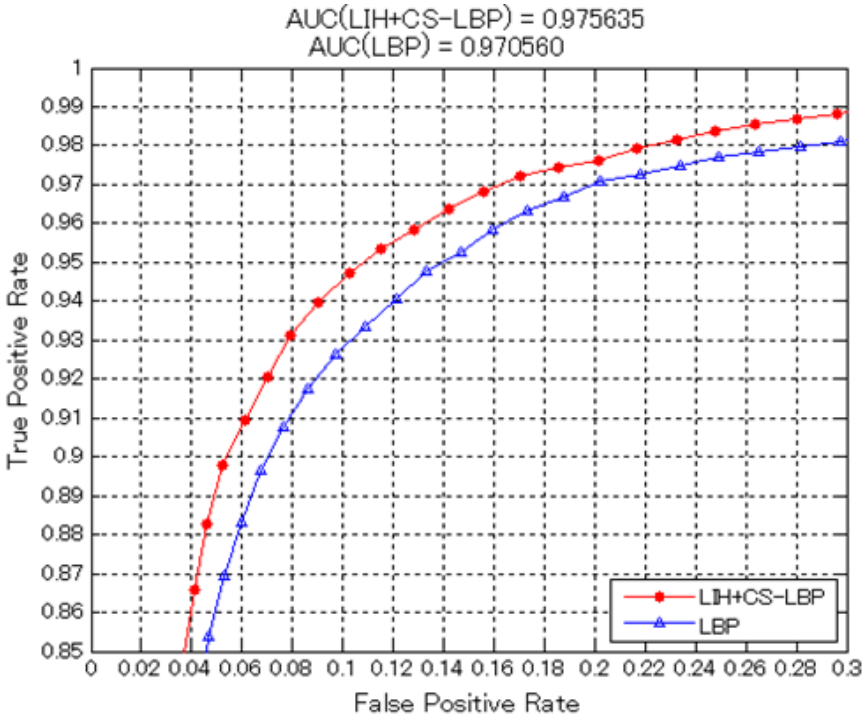


図 4-21　大規模笑顔/非笑顔データベースを用いた不特定多人数に対する汎化性能比較

このように大規模な画像データを対象としても，LIH + CS-LBP 特徴量の場合に，$AUC = 0.975635$，Uniform LBP 特徴量の場合に $AUC = 0.97058$ と両特徴量共に高い笑顔検出性能を示し，不特定人に対してもロバストに笑顔が検出可能なことを確認した．よって本節の 2 つの評価実験を通じ，LIH + CS-LBP 特徴量は Uniform LBP 特徴量の半分以下の次元数で，顔表情及び不特定人に対して同等の汎化性能を示すことを確認した．

4.4.4 笑顔度推定評価

次に，FEED データベース [49]を用いて笑顔度を推定した結果を示す．本データベースは，画像シーケンスが 100 から 150 フレーム程度と比較的長いフレーム数で構成されているため，表情の微妙な変化度合を評価する目的に非常に適している．図 4-22～図 4-24 に提案した笑顔度推定手法を用いて FEED データベースに含まれる，3 名の笑顔画像シーケンス(サブジェクト番号 0001 の第 1 試行, サブジェクト番号 0005 の第 1 試行, サブジェクト番号 0006 の第 1 試行) に対して笑顔度を推定した結果を各々示す．両者共に無表情から笑顔への表情変化が笑顔度としと非常に良く数値化されており，特に図中赤枠で示した付近の微妙な表情変化も推定できていることがわかる．また，図 4-23 に示したサブジェクト番号 005 の第 1 試行の画像シーケンスでは無表情から笑顔，そして笑顔から無表情へ，図 4-24 に示したサブジェクト番号 006 の第 1 試行の画像シーケンスでは，無表情から笑顔，笑顔から無表情，そして最後にまた無表情から笑顔へと各々表情が変化しているが，本提案手法により各表情変化を追従できていることが確認できる．また提案手法を，例で示した 3 名の他の試行及びその他 15 名の全試行についても同様に適用した結果, 実際には笑顔の表情を示していなかったサブジェクト番号 0017 の第 1 試行，サブジェクト番号 0011 の第 2 試行，サブジェクト番号 0014 の第 3 試行の画像シーケンス以外は，図 4-22～図 4-24 に示したのと同様に対象者の笑顔の変化度合に応じて笑顔度が推定できていることを確認した．

図 4-22　サブジェクト番号 001，第 1 試行に対する笑顔度推定結果

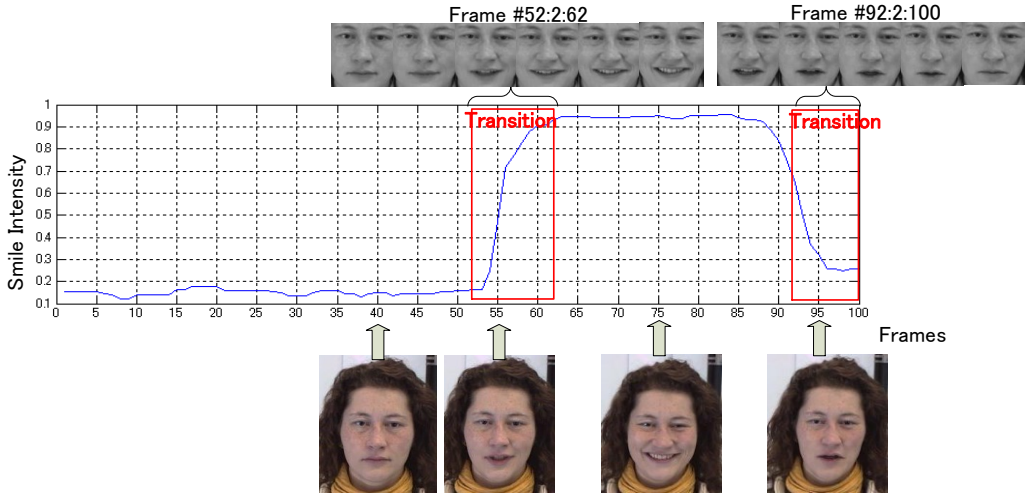


図 4-23　サブジェクト番号 005，第 1 試行に対する笑顔度推定結果

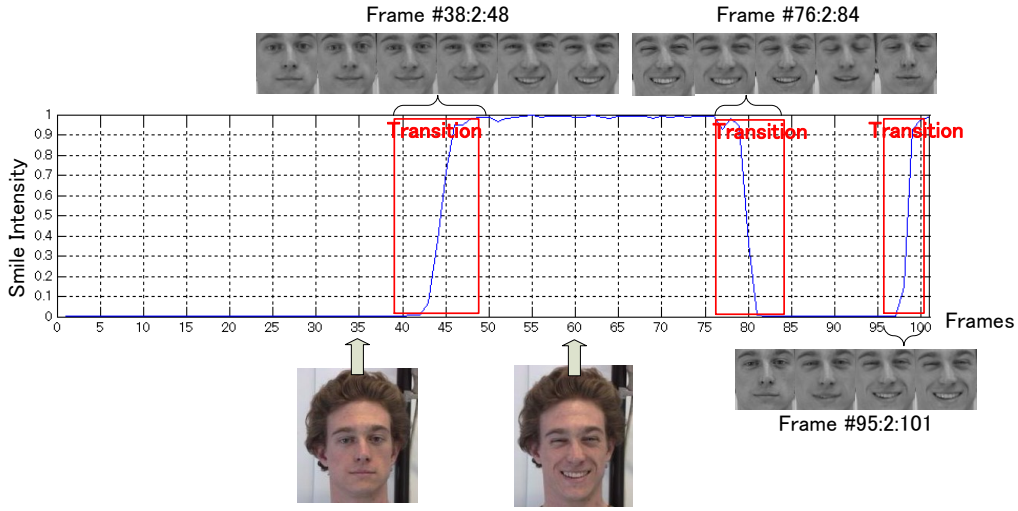


図 4-24　サブジェクト番号 006，第 1 試行に対する笑顔度推定結果

4.5　療法効果評価への応用

次に，提案した笑顔度推定手法を実際の臨床映像に適用し，音楽療法効果の評価へ応用することを考える．本研究は茨城県立医療大学倫理委員会の承認を得て実施され，映像を研究利用するための患者へのインフォームド・コンセントは以下の 2 段階で行われた．1 段階目のインフォームド・コンセントは，茨城県立医療大学付属病院で実施される音楽療法を受ける全ての患者に対して初回導入時に適用されるものである．ここでは，映像記録を行うこと，その映像データを参加者のアセスメントや療法内容の向上のために用いること，非公開で院外には持ち出さないこと，希望者は背面位置や撮影視野外に座席を移動して顔を映さないことも可能であること等を説明して承諾を得る．それに加えて，2012 年 1 月からは本研究への協力の呼びかけを開始し，それに応じた参加者もしくはその家族に対して個別に 2 段階目のインフォームド・コンセントを実施した．ここでは，研究の目的と目的外で映像データを使用しないこと，映像データから抽出された匿名化数値（本研究においては，笑顔度等）が公表される場合があることなどを口頭・書面で説明して承諾を得た．結果，同年 8 月の時点で 3 名の研究協力（つまり，2 段階目のインフォームド・コンセントへの承諾）が得られた．この 3 名の映像データを視聴し，退院までの期間が短かった患者 1 名，座位姿勢が困難な症状で着座中は顔が大きく俯いてしまい，映像データからは殆ど顔及び表情が観測できなかった為，本研究の顔検出器の現状性能では検出が困難であった 1 名を各々対象から除き，残りの 1 名について機器の不具合で記録できなかった回や着座位置の関係で顔が確認できなかった回などを除外した結果，比較的長期に渡る 6 回のセッションで良好な映像が確認できた．そこで本研究では，この 1 名を対象に笑顔度の推定を試みることとした．

ここで解析対象となった患者は，NMDA 受容体関連脳炎を発症した 10 代女性で，初回参加の 3 月時点では覚醒状態が安定せず，四肢麻痺があり，日常生活動作に全介助が必要な状態であった．そこで，覚醒水準の向上とコミュニケーション反応を促す目的で担当作業療法士の要請を受けて音楽療法が導入され，月に隔週 2 回の頻度で 3 月から 8 月までに実施された 12 回の音楽療法セッションのうち 10 回に参加した．しかし前述の通り，撮影機器の不具合や着座位置の影響により，実際に研究利用可能な映像データが得られたセッションは，3，4，5，7 月の各 1 回および 8 月の 2 回の計 6 回であった．

4.5.1　臨床映像からの笑顔度推定

本研究が対象とした音楽療法セッションは各回の参加患者が 4〜7 名，音楽療法士，スタッフらが円陣に着座した状態で行われ，患者の着座位置は特に定めなかった．そのた

め，患者が死角に入るのを極力防ぐため，約 50 度異なる方向から円陣を望むように 2 台のビデオカメラ（Sony 製 HDR-CX560）を設置し，プログレッシブ・ハイビジョン画質（1920 × 1080 画素）で撮影した．円陣の直径は約 3 m，各カメラは円陣中心から距離約 4.5 m，高さ約 1.2 m の位置に設置し，円陣全体が撮影されるように画角を調整し撮影した．また，撮影を患者になるべく意識させない配慮として，患者の着座位置を特に定めないことに加えて，三脚などの撮影機材を用いず，ピアノや音響機器などの室内備品の上に無造作にカメラを設置した．

1 回の音楽療法セッションは，各月ごとに季節にちなんだ唱歌を取り上げる“季節の歌”，簡単に演奏できる打楽器を音楽に合わせて打ち鳴らす“楽器演奏”や，各人がマイクを使って好みの曲を歌う“カラオケ”など幾つかのプログラムから構成され 1 時間程度実施される．プログラム内容によっては対象者の顔角度が頻繁に変わったり，他の患者の影に入ったりして安定して顔検出できない場合があった．今回はその中で最も顔検出が安定して処理できた“季節の歌”プログラムを笑顔度推定の対象プログラムとした．図 4-25 に対象とした患者の全 6 回の音楽療法セッションの“季節の歌”プログラムの映像データに対して，笑顔度推定を行った結果を示す．この内最後の 2 回（8 月 10 日，8 月 24 日）に着目すると，本提案手法では特に患者を拘束していないため，従来同様に自然な形で音楽療法セッションが行える一方で，顔向きの大きな変化などに起因して顔検出に失敗した区間（図 4-25 中④の区間）もあるものの，ほぼ全域にわたって笑顔度の推定ができていることがわかる．ここで“季節の歌”プログラムでは，季節にちなんだ唱歌 2 曲を取り上げ，音楽療法士は患者に，どちらかの曲を選んで皆を指揮しながら歌うように働きかける構成になっている．通常患者は自分が好きな方の曲を選択して，ピアノ伴奏に合わせて歌唱の指揮を行う．このとき音楽療法士やスタッフは障害の状況に応じて患者の指揮を介助する．この介入をすべての患者で順番に行うので，1 人の患者にとっては，自分が好きな曲を選んで指揮して歌う場面【指揮】（図 4-25 中③の区間），自分が選んだのと同じ曲を指揮せず歌う場面【選択曲】（図 4-25 中①の区間），自分が選んだのとは違う曲を歌う場面【非選択曲】（図 4-25 中②の区間），および，それら以外の積極的介入がない場面【非介入】という 4 種類の区間が生じることになる．

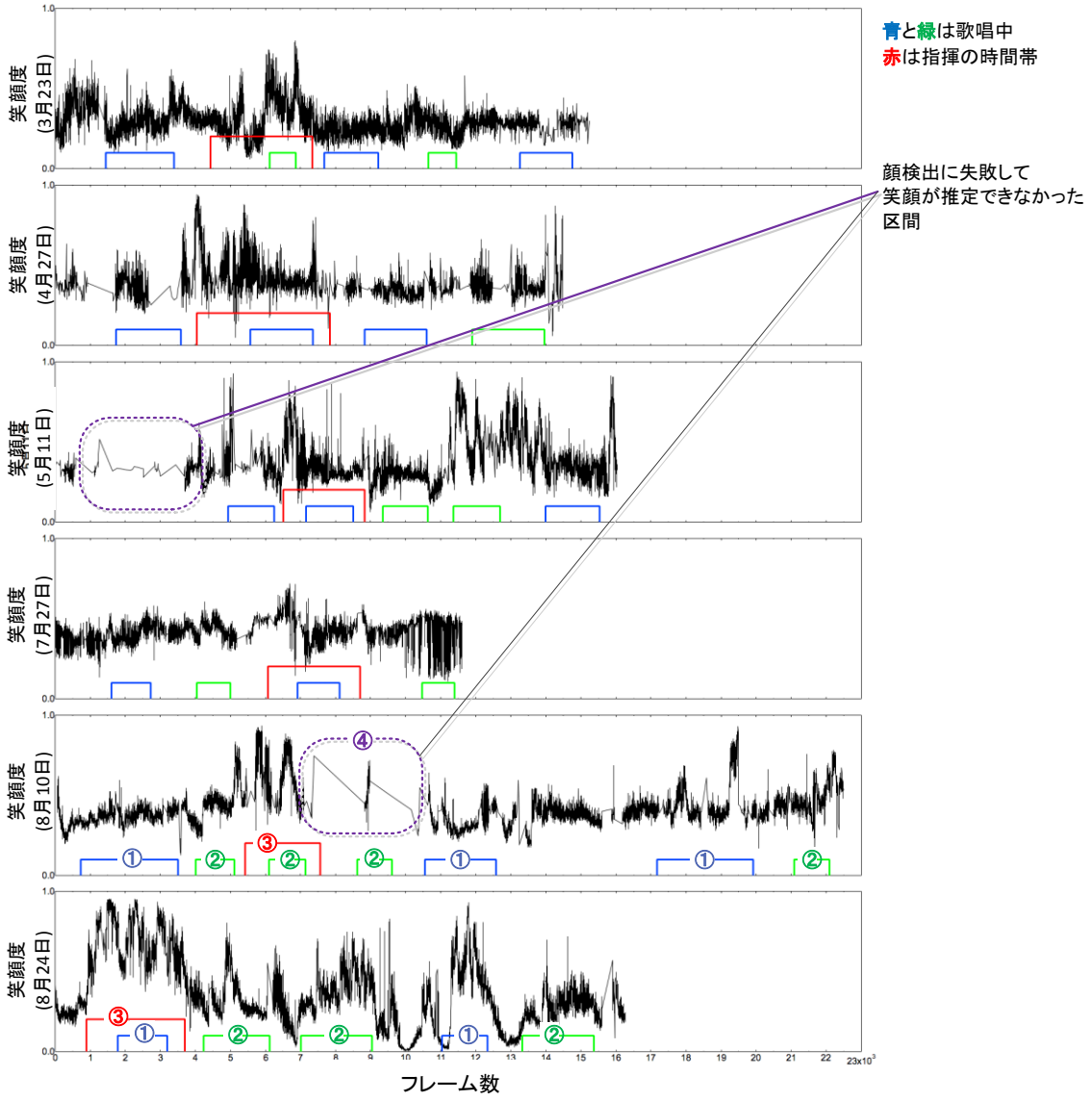


図 4-25　“季節の歌”プログラム中の笑顔度推定結果

4.5.2　　統計的検定による推定結果の評価

はじめに，4 つの介入区間において音楽療法の経過とともに平均笑顔度がどのように変化したかを観察するため，図 4-26 に介入内容ごとの平均笑顔度の音楽療法経過回数による推移を示す.

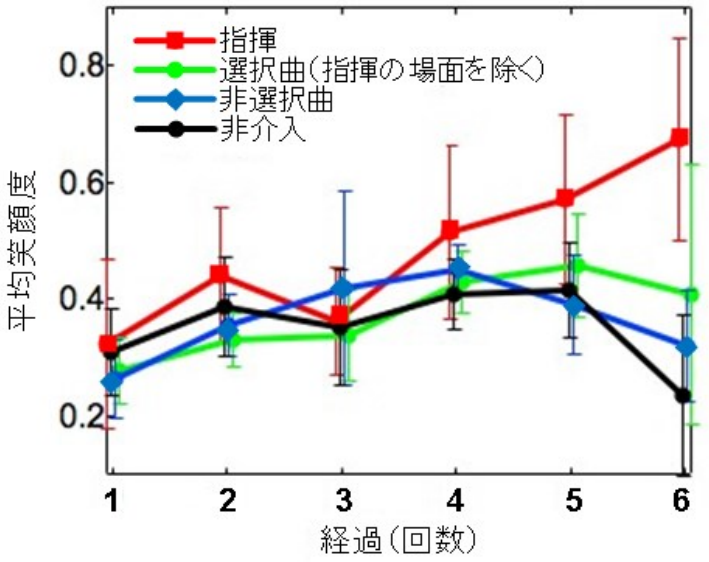


図 4-26　介入内容ごとの平均笑顔度の推移

次に大まかなデータの傾向を知るために，経過回数（1～6）と介入内容（指揮，選択曲，非選択曲，非介入）を要因とする二元配置分散分析を行った結果を表 4.2 に示す.

表 4.2　二元配置分散分析結果

変動要因	偏差平方和	自由度	平均平方	F値	P値	F(0.95)
全変動	11.273	651				
行間変動(介入内容)	1.299	3	0.433	41.898	1.086E-24	2.619
列間変動(経過回数)	1.570	5	0.314	30.380	1.062E-27	2.228
交互作用	1.914	15	0.128	12.347	3.997E-27	1.682
誤差変動	6.490	628	0.010			

その結果，$p < 0.01$ の有意水準で音楽療法の経過回数による平均笑顔度の違い及び介

入内容による平均笑顔度の違いが認められたが, 同時にそれらの交互作用も認められた. そこで更に詳細な分析を行うため，経過回数（1～6）と介入内容（指揮，選択曲，非選択曲, 非介入）に関して Tukey-Kramer 法による多重比較検定を実施した（表 4.3～表 4.6 参照）．すると例えば，表 4.3 に示すように，【指揮】時の平均笑顔度における回数による漸増は有意であることが分かったが，と同時に表 4.6 に示すように，【非介入】の平均笑顔度も有意に変動していることが分かった.

表 4.3 【指揮】時の Tukey-Kramer 法による多重比較結果 (:p < 0.01, * :p < 0.05)**

	#1_指揮	#2_指揮	#3_指揮	#4_指揮	#5_指揮	#6_指揮
#1_指揮						
#2_指揮	*					
#3_指揮						
#4_指揮	**		*			
#5_指揮	**	*	**			
#6_指揮	**	**	**	**		

表 4.4 【選択曲】時の Tukey-Kramer 法による多重比較結果 (:p < 0.01, * :p < 0.05)**

	#1_選択曲	#2_選択曲	#3_選択曲	#4_選択曲	#5_選択曲	#6_選択曲
#1_選択曲						
#2_選択曲						
#3_選択曲						
#4_選択曲						
#5_選択曲	*	**	*			
#6_選択曲						

表 4.5 【非選択曲】時の Tukey-Kramer 法による多重比較結果 (:p < 0.01, * :p < 0.05)**

	#1_非選択曲	#2_非選択曲	#3_非選択曲	#4_非選択曲	#5_非選択曲	#6_非選択曲
#1_非選択曲						
#2_非選択曲						
#3_非選択曲	**					
#4_非選択曲	**					
#5_非選択曲	**					
#6_非選択曲				*		

表 4.6　【非介入】時の Tukey-Kramer 法による多重比較結果 (** :p < 0.01, * :p < 0.05)

	#1_非介入	#2_非介入	#3_非介入	#4_非介入	#5_非介入	#6_非介入
#1_非介入						
#2_非介入						
#3_非介入						
#4_非介入	**					
#5_非介入	**		*			
#6_非介入		**	**	**	**	

そこで，各回での【非介入】時の平均笑顔度をその回の対象者の表情（又は感情）ベースラインと考え，他の介入内容の平均笑顔度からその値を減ずることで平均笑顔度を正規化し，その正規化平均笑顔度に対して Dunnett 法による多重比較検定を実施した．
その結果，図 4-27 に示すように【指揮】時は後半 3 回で継続して $p < 0.01$ の有意水準で正規化平均笑顔度の増加が顕著に認められ，最終回に至っては，【選択曲】【非選択曲】においても各々 $p < 0.01$，$p < 0.05$ の有意水準で笑顔が増えたことが認められた．

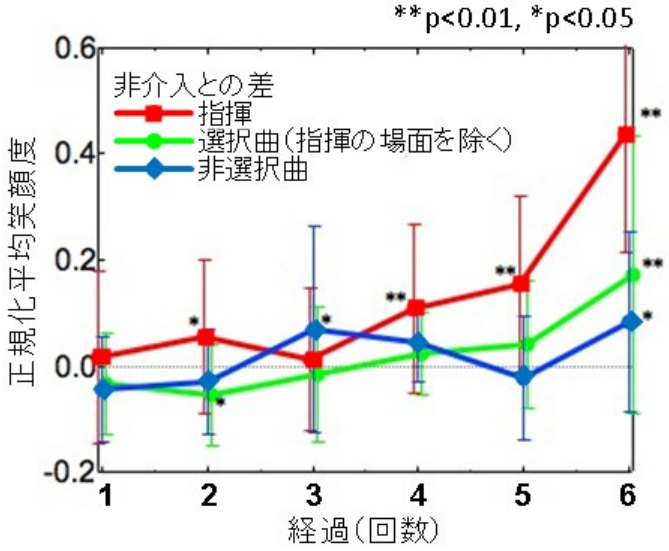


図 4-27　Dunnett 法による多重比較結果

4.5.3　主観評価との比較

最後に，従来の一般的な評価方法と笑顔度推定結果との関連性について調査した結果について述べる．従来の評価では，音楽療法中の患者の様子を記憶や記録映像を確認し

ながら，予め用意された主観評価表に記載の各評価項目に関して段階評価するなどの方法がとられてきた．ここで主観評価表は音楽療法セッションごとにその音楽療法に参加した複数の音楽療法士スタッフによって評価されるため，各評価項目おける平均値を主観評価値として，各介入内容における平均笑顔度との相関を調査した．

表 4.7 に全 6 回の音楽療法における，各評価項目の主観評価値を，表 4.8 にその主観評価値と提案手法により推定された平均笑顔度との相関係数を各々示す．

表 4.7　主観評価項目の評価値

年月日	スタッフ数	積極性	持続性	協調性	情動表出	状況適切性	知的機能	発声発語	歌唱
2012/3/23	3	1.667	2.250	1.250	2.250	2.000	2.000	2.500	1.000
2012/4/27	6	1.250	1.000	1.083	2.000	2.333	2.000	1.000	1.083
2012/5/11	3	2.000	2.000	1.000	2.000	2.000	2.000	1.000	1.000
2012/7/27	4	3.167	3.000	2.375	3.125	3.000	2.625	1.000	1.375
2012/8/10	3	3.667	3.333	3.000	3.667	3.000	2.833	1.833	1.000
2012/8/24	3	3.333	3.000	3.000	3.500	3.333	3.000	1.667	1.000

表 4.8　主観評価値と各セッションの平均笑顔度との相関

	指揮	選択曲	非選択曲	非介入
積極性	0.801	0.883	0.353	0.025
持続性	0.609	0.794	0.193	-0.059
協調性	0.906	0.963	0.141	-0.089
情動表出	0.868	0.942	0.156	-0.041
状況適切性	0.971	0.975	0.221	-0.099
知的機能	0.971	0.975	0.221	-0.099
発声発語	-0.083	0.114	-0.784	-0.403
歌唱	0.096	0.202	0.605	0.470

評価項目の中でも，特に情動に関する情動表出，状況適切性や協調性，積極性，知的機能などと平均笑顔度との相関が非常に高いことがわかる．また，ここで介入の内容に着目して観察してみると【指揮】や【選択曲】など介入レベルが高いほど相関が高くなっていることから，音楽療法士の積極的な介入により，患者の社会性やコミュニケーション力が改善した結果，表情変化が豊かになっていったとも考えらえる．一般的に考えても笑顔は，情動反応や他者とのコミュニケーション場面で表出されることが多いと想定されるが，この結果からもその妥当性が示され，従来の評価方法との比較によっても笑

顔度と患者の療法効果による症状の回復度合（特に情動に関わる）との高い関連性が示されたと考える.

4.6 まとめ

本研究では，映像情報から顔器官を検出することなく顔のアピアランス情報のみからSVM により笑顔を検出する技術を開発し, その技術を応用し音楽療法中の記録映像から患者の笑顔度を推定し定量化することで，その時間的・量的推移から音楽療法の効果を客観的に評価する方法について検討した.

笑顔検出器の性能評価では，はじめに各特徴抽出法の比較検討として独自画像データベースを対象に，アピアランス・ベースの特徴量として， LIH 特徴量，CS-LBP 特徴量及びこれら 2 つの統合特徴量である LIH + CS-LBP 特徴量を用いて，各々の特徴抽出法に最適な特徴抽出パラメータの調査と笑顔の検出性能の評価を実施した. その結果, LIH 特徴量の場合には, 格子セル数 8×8, ヒストグラムビン数 4 のときに $AUC = 0.979522$, また CS-LBP 特徴量の場合には，格子セル数 5×5，エンコード閾値 0.02 のときに $AUC = 0.979423$ を得た. そして, LIH + CS-LBP 特徴量を用いた場合に $AUC = 0.982269$ と最も高い性能を示すことを確認した．また，SVM による笑顔検出器を少数のサポート・ベクタからなる副笑顔検出器と全サポート・ベクタからなる主笑顔検出器のカスケード構造とすることで, $AUC > 0.97$ という性能を維持しつつ計算量を約 40 % 程度削減可能なことを示した．次に，顔向きに対する基本的な笑顔の検出性能を顔向き別の笑顔/非笑顔画像データベースを用いて検証した結果，LIH + CS-LBP 特徴量を用いた場合に，“正面顔”において $AUC = 0.981039$，“斜め顔”において $AUC = 0.976259$ と何れの顔向きにおいても高い笑顔検出性能を達成した．また汎化性能の評価においては，目的別に 2 種類の研究用公開データベースを用い，他手法の比較と併せて評価した結果，CK+データセットを対象とした場合に $AUC = 0.99668$ ，GENKI-4K データベースを対象とした場合に $AUC = 0.975635$ となり，LIH + CS-LBP 特徴量を用いると従来手法の半分以下の次元数で同等の笑顔検出性能を示すことを確認し，また FEED データベースを用いた笑顔度推定の評価では，人物の表情変化画像シーケンスを用いて本手法が無表情から笑顔への微妙な表情変化を笑顔度として定量化可能なことを示した.

次に，開発した笑顔度推定手法を用いた音楽療法効果の評価実験では，半年間計 6 回に渡って継続的に撮影された実際の臨床現場の音楽療法記録映像を用いて，対象患者の顔向きが正面付近の場合には, 顔検出及び笑顔度の推定が安定して行えることを確認し，またその測定結果を多重比較検定することで，音楽療法の経過回数に伴い患者の平均笑

顔度が有意に増加していることを示した．また，これと並行して実施した従来の音楽療法士による主観評価結果と本提案手法による笑顔度の定量化結果が，特に情緒を示す評価項目において高い相関があることが確認された．ただ一方で，実際の臨床映像を対象とした場合，現状の顔検出及び笑顔度推定技術では多種多様な症状を抱えた患者に対し実用上十分に安定して笑顔度を数値化できるほどの性能を達したとは言えず，今後アルゴリズムの更なる改良，撮像方法の工夫などが必要であることも確認した．

第5章　人物属性推定とそのデモグラフィック調査への応用

5.1　はじめに

2008 年 9 月に起きたリーマンショックに端を発した経済危機以降特に，大量生産・大量消費の時代は終焉を迎え，大勢の満足から個々の満足へ，物質による満足から心の満足へと世の中の価値観が変遷し，人々のニーズを掴むことが非常に難しい時代になってきた．このような時代において本当に役立つモノやサービスを生み出すためには，あらゆる市場領域でデモグラフィックと呼ばれる顧客分類やプロフィールのより細やかで迅速な把握を行なうことが重要である．しかし，従来の一般的な調査手法は主として人手によるため，導入コストがかさみ手軽に，また継続的に行なえるものではなく，かつ結果を得るまでにある程度時間を要するため，例え分析ができたとしてもそれを適時的に活用することが難しかった．

その一方で，デジタルサイネージ，自動販売機やマーケティング調査などで画像認識技術を応用しようとする新たな動きがある．そのような用途においては，顔画像等から性別や年齢といった，その人の属性情報を正確に推定することが重要な課題となる．そこで，近年急激にそのニーズが増加しているこれらのマーケット調査の根幹として，画像認識によるカメラ映像からの自動的なデモグラフィック調査技術を開発することで，圧倒的に安価なマーケット調査が可能となり，これまでコスト面で手が出せていなかったようなあらゆる領域に適用できるマーケット調査手法を提供できると考えた．本研究では，定点カメラで商業施設等の来場者を撮影し，上半身画像から人物を検出することで来場者数をカウントし，更に顔画像から性別と年齢を推定することで人物属性構成比率を算出するデモグラフィック調査のための来場者モニタリング・システムについて提案する．

来場者数をカウントするためには映像情報の中から人が存在するか否かを判定する必要があるが，本研究では人物検出に HOG（2.1.7 節）と呼ばれる輝度勾配方向ヒストグラムを特徴量とした識別手法を用いた．一般的な人物検出手法では人物の顔画像や全身画像を用いて検出を行うが，本研究では人物を顔や全身画像からではなく上半身部分から検出することを検討した．これにより，横向きや後ろ向きなどカメラに顔が写らない場合にも人数判定を行なえ，また従来行なわれていた全身画像からではなく，上半身画像からの人物検出を行なうことで，混雑時に人物が重なり合った状況での人物検出の精度を高め，更に人物追跡を行なうことで人物属性判定結果を映像シーケンスの中で同一

人物ごとに総合判定を行なうことで判定性能を向上させた．人物上半身検出の後，その検出領域の中に顔探索領域を設定し顔検出を行なった．顔検出には Haar-Like 特徴量と AdaBoost による Viola-Jones 型の顔検出器 [13]とカスケード構造化された 2 つの SVM による顔検出器を組み合わせた高速でロバストな顔検出器 [45]を用いた．そして，顔が検出された場合，顔画像から性別・年齢の推定を行なう．顔画像から性別・年齢などの人物属性を推定しようとした場合，特徴点ベースの特徴量を用いる手法 [51] [52]とアピアランス・ベースの特徴量を用いる手法 [53] [54] [55] [56]とに大別される．特徴点ベースの手法の多くは顔器官や顔から特徴点を検出し，その位置情報や周辺の特徴抽出により顔の詳細な解析を行うことで精度の良い人物プロフィール推定を目指している反面，認識処理演算量が多く，カメラ画像内の多数の人物を通常の演算処理環境でリアルタイムに属性推定することは困難である．一方で，商業施設などにおいては来場者のプライバシーへの対応から，リアルタイムに認識処理を行ない，撮影映像をその場で破棄することが求められる．そこで本研究では，顔器官の位置情報や特徴点検出を用いず，アピアランス・ベース特徴量を用いた人物属性推定により認識コストを低減したリアルタイム処理を行なうことで，撮影映像のその場での破棄を実現したため，実際の商業施設への導入も可能となった．

5.2　来場者モニタリング

来場者モニタリングの全体処理フローを図 5-1 に示す.

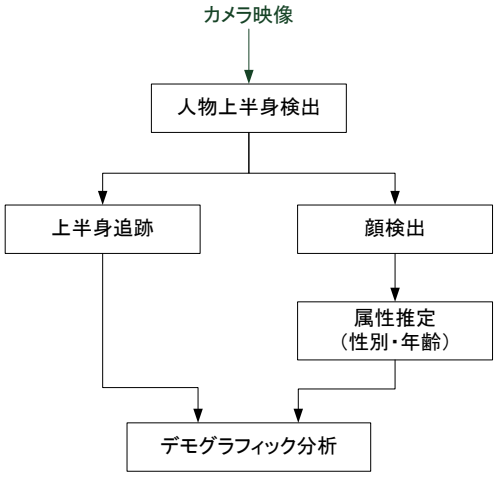


図 5-1　全体処理フロー

はじめにカメラから得られる入力画像に対して，人物の向きに依存しないマルチ・スケールでの上半身検出行なう．そして検出された上半身領域の位置とスケール情報から顔探索領域と顔探索スケール範囲を限定することで高速な顔検出を実施し，顔が検出された場合には，その顔画像から性別と年齢の推定を行なう．一方，カメラより得られるフレーム画像シーケンスにおける上半身検出領域の位置とスケールから，誤検出や未検出も考慮した上半身追跡を行なうことで来場者数を算出する．そして上半身の追跡結果から，同一人物の複数フレーム画像での属性推定結果を総合判断することで安定した人物ごとの属性推定を実現し，来場者数と人物属性推定結果を統計量としてまとめることでモニタリング結果として出力する.

5.2.1　人物検出

来場者モニタリングで想定される混雑時の人が重なりあった状況における人の向きに依存しない人物検出を目的として，全身画像ではなく上半身のみの画像による人物検出を行なった．この上半身画像領域を水平・垂直方向それぞれに 4 等分した 16 個の領域をブロックとし，その各ブロック内を水平・垂直方向にそれぞれ 2 等分した 4 個の領域をセルと定義した HOG 特徴量を上半身検出のための特徴量とした．はじめに，セル内の各画素における近傍 4 画素から水平・垂直方向の輝度勾配を求め，輝度勾配強度と方

向を算出する．そして，輝度勾配方向の 0 から 180 度までの 20 度ごとのビン数 9 の輝度勾配方向に対するセル内の輝度勾配強度累積をヒストグラムとする．次に，ブロック内のセル 4 個の輝度勾配方向ヒストグラムを連結し，正規化することで 36 次元のブロック内輝度勾配方向ヒストグラムを生成する．最後に．このブロックをセル幅ステップでオーバーラップ・スキャンさせ，上半身画像領域内を水平方向に 7 ステップ，垂直方向に 7 ステップの計 49 ステップのスキャンを行なうことで，上半身画像領域として 1,764 次元の HOG 特徴量を得る．この HOG 特徴量を用いて線形 SVM により上半身検出を行なっている．アプリケーションによるが，例えば入力画像の解像度が 640×480 画素の場合，上半身画像領域サイズを 480×480 画素から 48×48 画素までの 16 段階にスケール変化させることでマルチ・スケール検出を行った場合，図 5-2 に示すように通常 1 つの上半身に対して複数の近傍位置・スケールでの上半身領域が検出される．

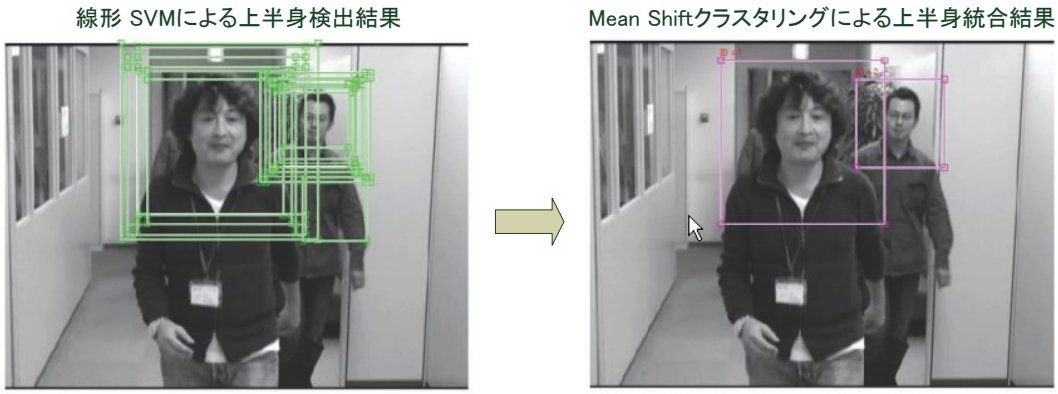


図 5-2　上半身検出結果の統合

そのため，検出された複数の上半身領域に対し，上半身領域の位置 (x_i, y_i)とスケール s_i を式(5.1)のガウシアン・カーネル関数を用いて位置とスケールのシフト値を算出し，それぞれのシフト値から位置とスケールを修正更新することで収束させ，1 つの上半身領域に統合している．

$$K_i = \exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{b_p^{\,2}}\right)\times\exp\left(-\frac{(s-s_i)^2}{b_s^{\,2}}\right) \tag{5.1}$$

ここに，b_pとb_sは各々位置とスケールにおけるそれぞれのクラスタリング探索領域として設定するパラメータである．

5.2.2　　上半身追跡

カメラより得られるフレーム画像ごとに検出された上半身の位置とスケールを用いて，フレーム画像シーケンスにおけるフレーム画像間での上半身追跡を行なう．過去の複数フレーム画像において同一人物の上半身と判定された上半身の位置とスケールから現フレーム画像における人物ごとの上半身の位置とスケールを予測する．そして，過去のフレーム画像から予測された人物ごとの上半身の位置とスケールと，現フレーム画像で検出された上半身の位置とスケールをそれぞれ比較し，同一性を判定することで上半身を追跡する．同一性の判定は，2 つの上半身領域の左上頂点の水平距離 $dh1$と垂直距離 $dv1$，そして右下頂点の水平距離 $dh2$と垂直距離 $dv2$を求め，それらのすべてが 2 つの上半身領域のスケールの小さい方の領域の 1 辺長に重み計数 aを掛けた判定閾値よりも小さい場合に同一であるとする（図 5-3 参照）．ここで，重み計数 aはカメラの画角や設置場所などに応じて，検出人物の平均的な歩行速度などを考慮して決められる．上半身追跡によりフレーム画像シーケンスを通じた人物 ID を検出上半身に付与することで，カメラ映像内を通過する人数の算出が可能となる．

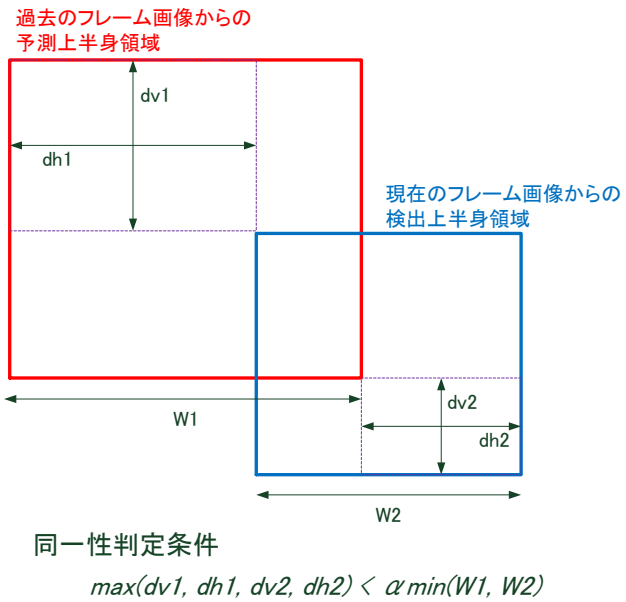


図 5-3　上半身の同一性判定

5.2.3 人物属性（性別・年齢）の推定

人間は，見た目だけである程度その人の性別を推定することが可能である．これは，性別による顔，髪型，体型，姿勢，動作や服装の違いなどを，人間がその成長過程において自然と得られる経験から学習し，その過去の経験と現在観測されている状態を統合的に判断してある程度無意識の内に推定していると考えられる．顔のパーツの配置情報が年齢，性別に関してどの程度明確な指標となるのかを検討した研究 [57]などもなされてはいるが，人が顔のみからその人の性別や年齢を区別できる決定的な要因は，未だはっきりとは解明されていないのが現状である．

性別による違いについては，一般的には，女性の方が男性に比べて目が大きく，長髪の割合が高いと考えられる．また女性の方が，化粧により顔器官のコントラストが強く，服装も男性に比べると明るい暖色系が好まれる傾向にあると言えるかもしれない．また，幼少期には性別を見分けることが困難なことから，幼少期にはみられず，成長するに従って差が表れてくる特徴が見た目から性別を推定する一因になっているとも考えられる

一方，年齢の推定については，性別の推定と同様に，ある程度顔のみから年齢を推定できると考えられるが，見た目だけから正確に年齢を推定することは非常に困難であり，複数の人が1人の年齢を推定した場合には10歳程度のばらつきがあるという報告もある．また，人種が異なると性別よりも一層年齢の推定が困難なことから，性別よりも人種ごとに経験的に学んでいる要因も多いと考えられる．人間が年齢を推定する場合，顔から得られる情報の内，年齢に関わる幾つかの特徴を抽出し，その個々の特徴について何らかの判定基準によって評価した後，その結果を統合し最終的な年齢を推定していると考えられる．このとき，人によって各特徴に対する重み付けが異なるため，同じ対象に対しても，異なる年齢を推定してしまうため推定結果がばらつくと考えられる．この年齢推定のための特徴と判定基準が明確であれば，それらを顔画像から抽出し所望の重み付けを用いて演算することで，年齢の推定が可能であると考えられる．例えば，皺，肌のクスミ，頭髪などが，その特徴の一部を構成すると考えられるが全てではない．このように，人間が見た目から性別・年齢を推定する仕組み自体が未だはっきりとは解明されていないため，顔画像から計算処理により自動的にそれらを推定する手法は，古くから研究されているが未だ決定的な手法が確立されておらず，今もなお活発に研究されている画像認識分野の一つである．

性別・年齢推定のための予備実験

はじめに，性別・年齢により顔画像のアピアランスにどのような差があるのかを調査するため，顔画像の輝度情報のみを用いた予備実験を行った．実験には財団法人ソフトピアジャパンから使用許諾を受けた顔画像データベース[2]の顔画像を用い，予め目の位置と顔の大きさを概ね揃えた後に，頭髪や背景除去のためのマスク処理を施し，輝度値をL2 ノルムで正規化した．性別による違いを確認するため，全男性の顔画像から求めた平均顔（平均男性顔）及び全女性の顔画像から求めた平均顔（平均女性顔）を作成し，各々の差分を算出した結果を図 5-4 に示す．

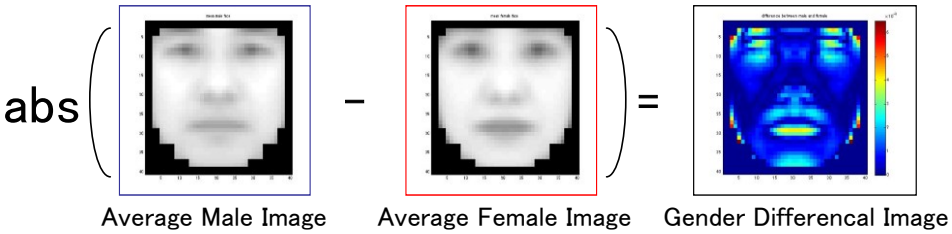


図 5-4　平均男性顔，女性平均顔，及びその絶対差分顔

図 5-4 右の絶対差分顔に着目すると眉，目，口唇の辺りに顕著な差が見られることから男女の見た目の差がそれらの場所に表れていることが確認できる．

[2] 本章に使用した顔画像データの一部は，財団法人ソフトピアジャパンから使用許諾を受けたものです．権利者に無断で複写，利用，配布等を行うことは禁じられています．

次に全顔画像を主成分分析した結果．その第 1～10 主成分の重みを変化させて再構成した顔画像を図 5-5 に，第 1～3 固有顔を図 5-6 に示すが，全画像の主成分分析結果からは性別を示す明らかな軸は見られなかった．

図 5-5　全顔画像の主成分分析結果（第 1～10 主成分）

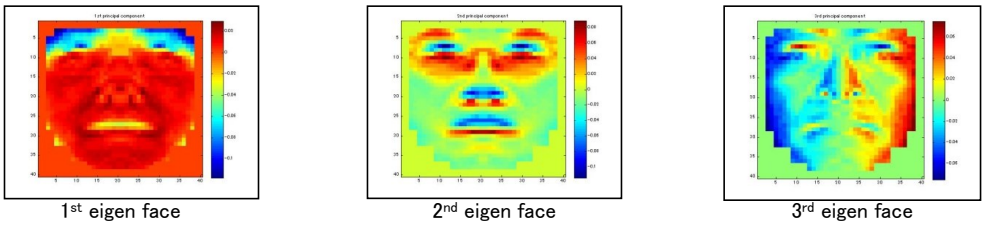


図 5-6　第 1～3 固有顔

次に，年齢による顔画像のアピアランス情報の違いを調査するため，男性と女性各々の顔画像について主成分分析を実施した．また主成分分析に際し，顔画像情報に実際の年齢情報を付加してから主成分分析を適用し，第 1 主成分を意図的に年齢の軸と対応させた．図 5-7，図 5-8 に第 1～5 主成分の重みを変化させて再構成した男性・女性の顔画像を各々示す．男性，女性共に第 1 主成分が年齢の軸に対応していることがわかる．

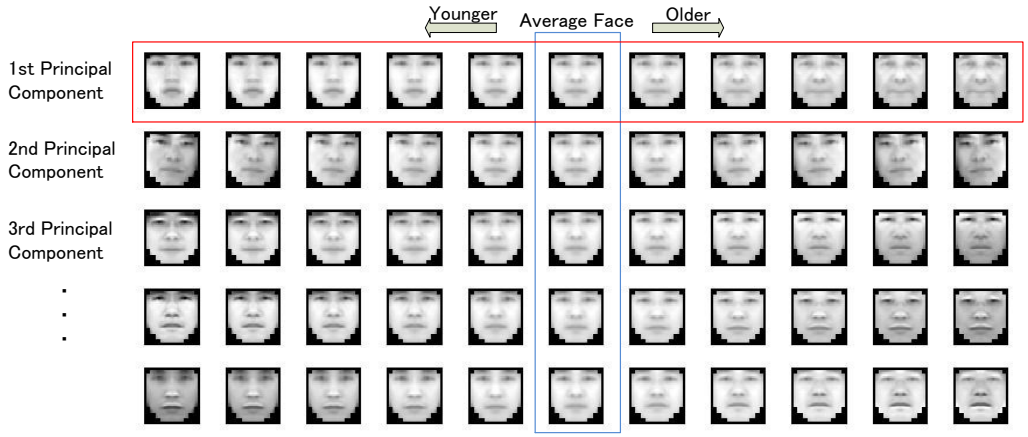


図 5-7　年齢を加味した男性顔画像の主成分分析結果（第 1～5 主成分）

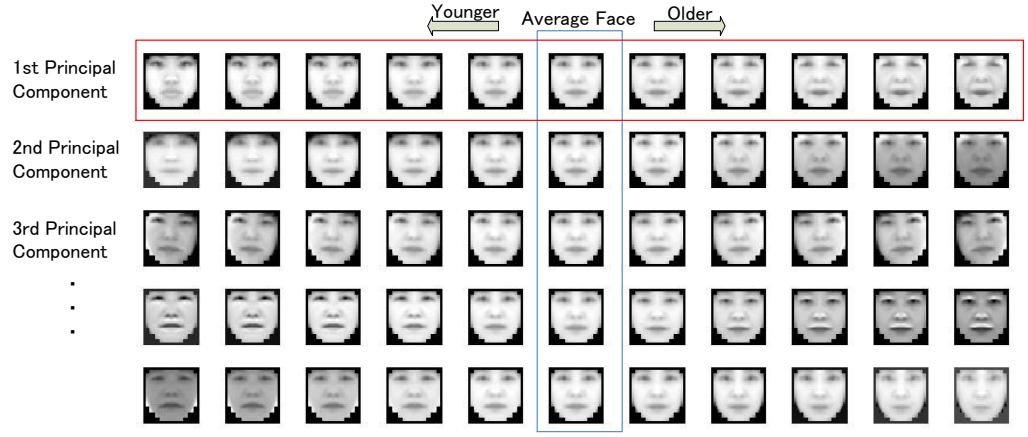


図 5-8　年齢を加味した女性顔画像の主成分分析結果（第 1～5 主成分）

次に，男女各々について第 1 主成分の重みを変化させて若年齢再構成した顔画像と，老年齢再構成した顔画像との差分画像を作成した（図 5-9 及び図 5-10 参照）.

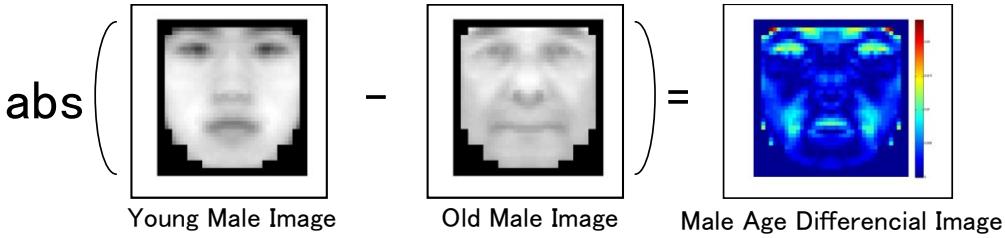


図 5-9　男性の若年・老年顔画像間の差分

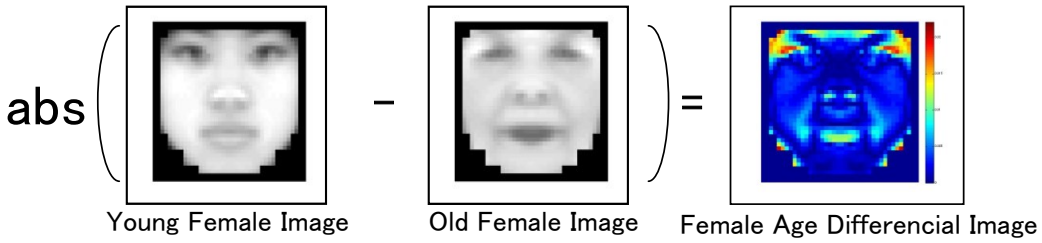


図 5-10　女性の若年・老年顔画像間の差分

その結果，男女の区別なく目元，口元に若年老年の差が大きいことがわかる．また，性別とは異なり鼻唇溝（いわゆる法令線）のあたりに差分が顕著に表れていることがわかる．これらの予備実験から顔画像のアピアランス情報のみからでも性別・年齢による明らかな差異が観測されたため，顔画像のみからそれらの推定が可能であると考えた.

性別・年齢の推定手法

本研究における顔画像からの性別・年齢推定の処理フローを図 5-11 に示す．はじめに検出された上半身領域から顔を検出する．このとき，検出された上半身の位置とスケールを元に，顔の探索スケールと位置を限定することで演算量を削減する．そして，顔探索領域内から計算速度と性能を両立させたカスケード型 SVM 顔検出器 [45]により顔の検出を行い，検出された顔領域を演算量削減のため 40×40 画素にリサイズし，照明変動の影響を抑えるためヒストグラム平滑化を行った後に，性別・年齢推定のための特徴量を抽出する．ここで実環境でのリアルタイム動作を目指し，顔から特徴点を抽出する必要がないアピアランス・ベースの特徴量として LIH ＋ CS-LBP 特徴量（2.1.5 節）

を用いると共に，性別・年齢各々の推定に用いる特徴量を共通化することで特徴抽出に要する時間を削減する．この特徴量を用いて予め SVM により学習しておいた性別推定器により人物の性別を識別した後に，SVR により学習しておいた性別ごとの年齢推定器により，その人物の年齢を推定する．ここで，年齢推定器を性別ごとに構成したのは，一般的に性別に依らず年齢推定器を構築して年齢を推定するよりも，性別ごとに分けて年齢推定器を構築し，性別ごとに年齢を推定した方が高い推定精度が得られるためである [55]．なお，性別推定に用いる SVM 及び年齢推定に用いる SVR ともにカーネル関数にはガウシアン・カーネルを用いた．

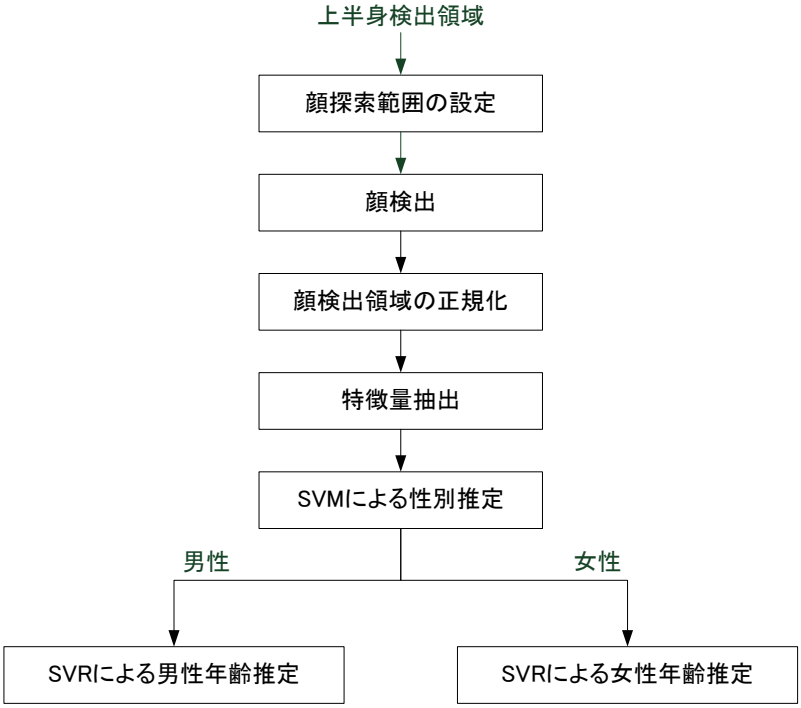


図 5-11　性別・年齢推定処理フロー

5.2.4　　来場者モニタリング分析

上半身追跡により，複数フレーム画像にわたって検出される人物の同定が可能となる．これにより，フレーム画像シーケンスからの来場者数カウントが可能となると同時に，同一人物に関して検出された複数フレーム画像の属性推定結果を総合的に判断することができる．実環境においては，撮像時の照明条件，人物の顔向き，姿勢や大きさなど様々な外的要因により同一人物においてもその時々によって属性推定の結果が変化する場合がある．また人物が向きを変えながら歩いている場合などにおいても，顔画像が撮影さ

れたフレーム画像を集めて判定を行なうことで，フレーム画像シーケンス全体としての人物属性推定を行うことが可能となる．本研究では，外的要因と顔検出の信頼度には負の相関があると仮定し，人物属性推定結果を顔検出器から得られる顔検出信頼度を重みとした加重平均とした．これにより，外的要因により突発的に属性推定値が変化した場合においても，その前後のフレーム画像における属性推定値と顔検出信頼度からそれらを抑制することができ，より信頼性の高いデモグラフィック分析が可能となった．今，人物 ID = i の人物について，顔が検出されたフレーム画像の総数を N_i，各検出フレーム画像での顔検出信頼度を f_j，性別推定値を g_jとしたとき，その総合性別推定値 G_iを式(5.2)のように定義した．

$$G_i = \frac{\sum_{j=1}^{N_i} f_j g_j}{\sum_{j=1}^{N_i} f_j} \tag{5.2}$$

同様に，顔が検出されたフレーム画像ごとの年齢推定値を a_jとしたとき，総合年齢推定値 A_iを式(5.3)のように定義した．

$$A_i = \frac{\sum_{j=1}^{N_i} f_j a_j}{\sum_{j=1}^{N_i} f_j} \tag{5.3}$$

5.3　　属性推定の基礎実験

各識別器の構築には独自に収集した上半身画像と顔画像を学習用画像として用い，SVM 上半身検出器は 5,000 枚，SVM 顔検出器は 10,000 枚，SVM 性別推定器は 10,000 枚，SVR 年齢推定器は 890 枚の画像からそれぞれ構築した．

はじめに，性別と年齢が明らかな顔画像 890 枚（男性 592 名，女性 298 名）を用いて，性別・年齢推定の Leave-One-Out 交差確認法による基本性能評価を行った．その結果，性別推定については男性を女性への誤推定が 24 画像，女性を男性への誤推定が 30 画像発生し，6.07 %の誤推定を生じた．またこのとき，実性別構成比と推定性別構成比のヒストグラム類似度は 0.99 であった（図 5-12 参照）．

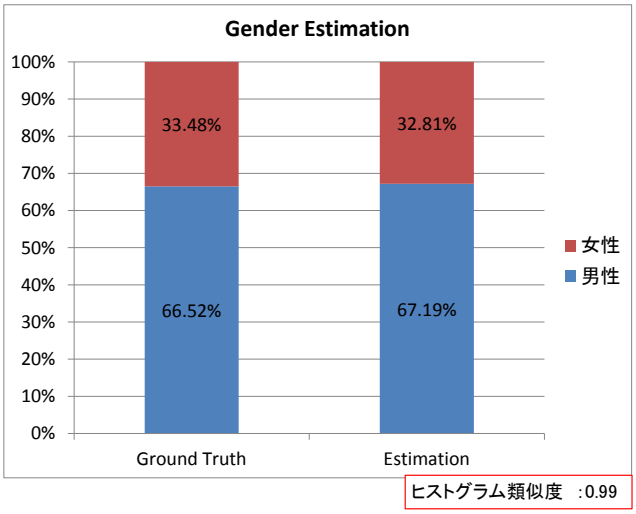


図 5-12　性別推定結果

ここで 2 つのヒストグラムの類似度は，ヒストグラム・インターセクション [58]を用いて算出した．今，ビン数 Nの 2 つのヒストグラム $H1$と $H2$があり，i番目のビンの値を各々 $H1_i$，$H2_i$とすると，そのヒストグラム類似度 Sは式(5.4)で表される．

$$S = \frac{\sum_{i=1}^{N} \min(H1_i, H2_i)}{\sum_{i=1}^{N} H1_i} \tag{5.4}$$

また，年齢推定については図 5-13 に示すような推定分布となり，男性については平均推定誤差が 6.66 歳，実年齢と推定年齢との相関係数が 0.82 であり，女性については平均推定誤差が 7.70 歳，実年齢と推定年齢との相関係数が 0.74 であった．

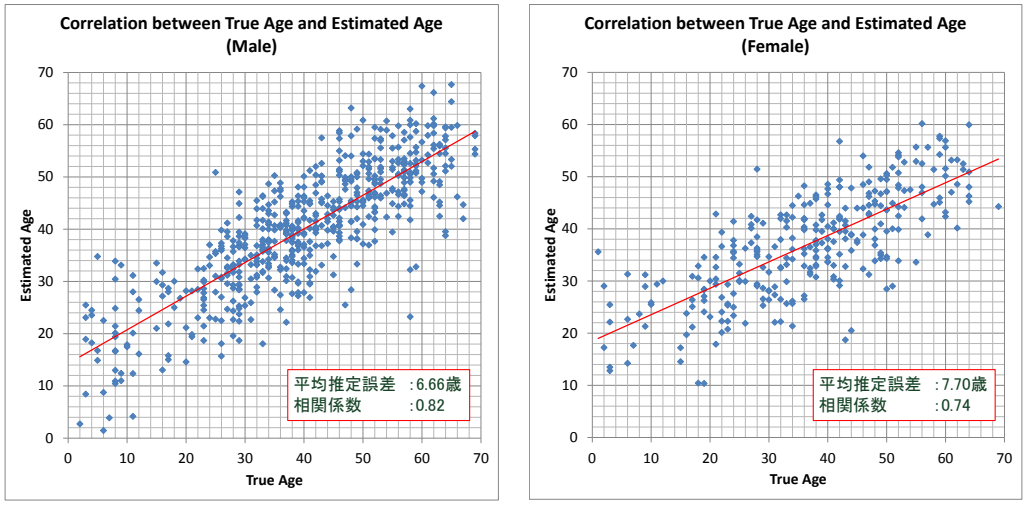


図 5-13　実年齢推定結果（左：男性，右：女性）

次に，デモグラフィック調査を想定し，個々の年齢ではなく年齢推定結果を 10 歳ごとの年代構成という統計量として表現してみる．例えば，顔画像から幼年/青年/壮年/老年といった年齢層を直接推定する研究 [59]もなされているが，本研究では一旦個々の年齢を推定した後に後処理として年代構成を算出する．結果，男女別の推定年代比率は図 5-14 のようになり，このとき男性については実年代比率と推定年代比率とのヒストグラム類似度は 0.86，女性については 0.81 であった．

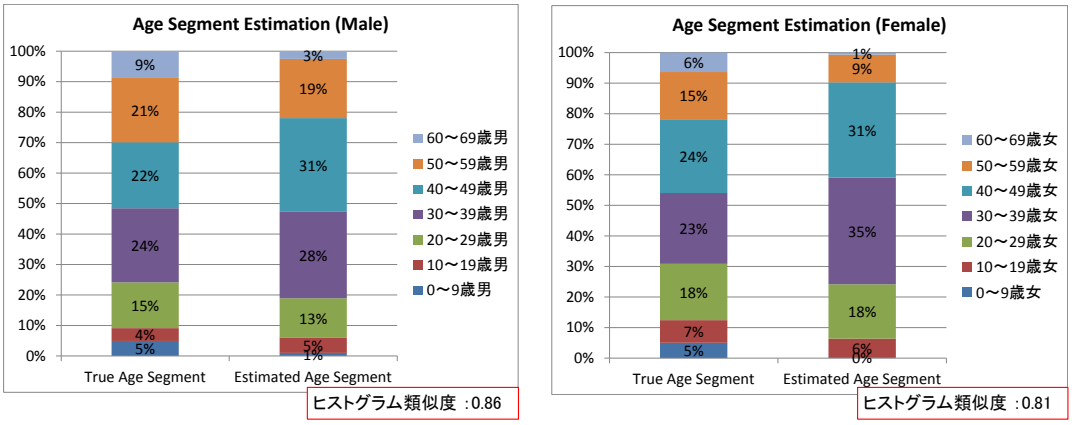


図 5-14　実年代推定結果（左：男性，右：女性）

次に年齢の推定値ではなく正解値について少し視点を変えて考えてみる．これまで，年齢の正解値は実際の年齢である実年齢（True Age）を用いてきたが，本研究では顔のアピアランス情報から年齢を推定するため，正解値も同様に実年齢ではなく顔の見た目のみから判断した見た目年齢（Perceived Age）とすることを考えてみる．

ここで見た目年齢は 5 名の評定者が顔画像のみを見て各々判断した年齢の平均値とした．この見た目年齢を正解値として年齢推定器を学習・評価した結果を図 5-15 に示す．

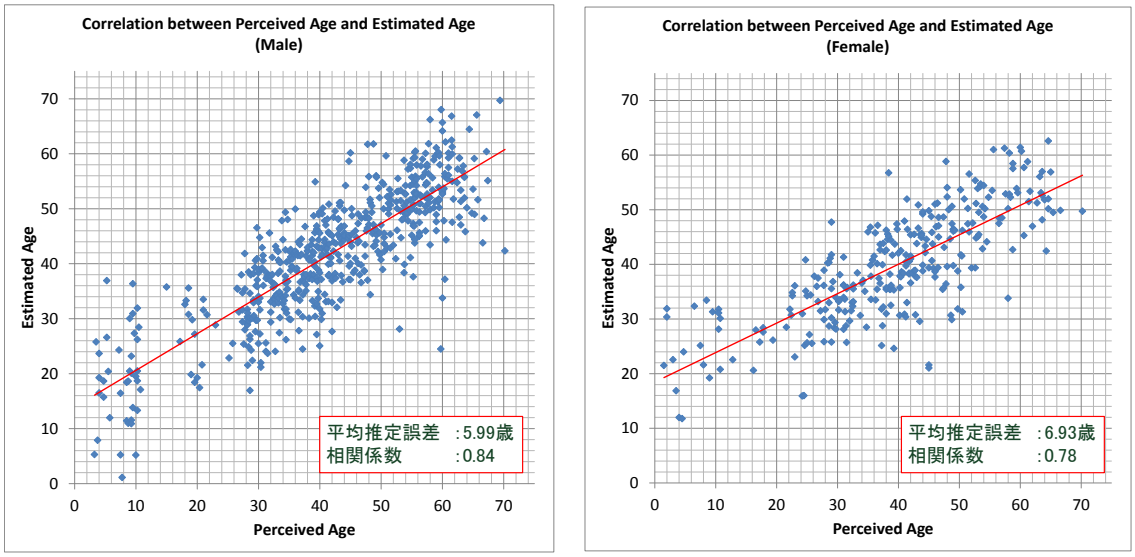


図 5-15　見た目年齢推定結果（左：男性，右：女性）

男性については平均推定誤差が 5.99 歳，見た目年齢と推定年齢との相関係数が 0.84 であり，女性については平均推定誤差が 6.93 歳，見た目年齢と推定年齢との相関係数が 0.78 となり，いずれも実年齢を正解値として用いた場合よりも推定性能が向上した．

次に，実年齢の場合と同様に，個々の見た目年齢ではなくデモグラフィック調査を想定し，年齢推定結果を 10 歳ごとの年代構成という統計量で比較したところ，男女別の推定年代比率は図 5-16 のようになり，このとき男性については見た目年代比率と推定年代比率とのヒストグラム類似度は 0.88，女性については 0.86 となり，いずれも実年齢による年代推定よりも見た目年齢による年代推定の方が高いヒストグラム類似度を示した．

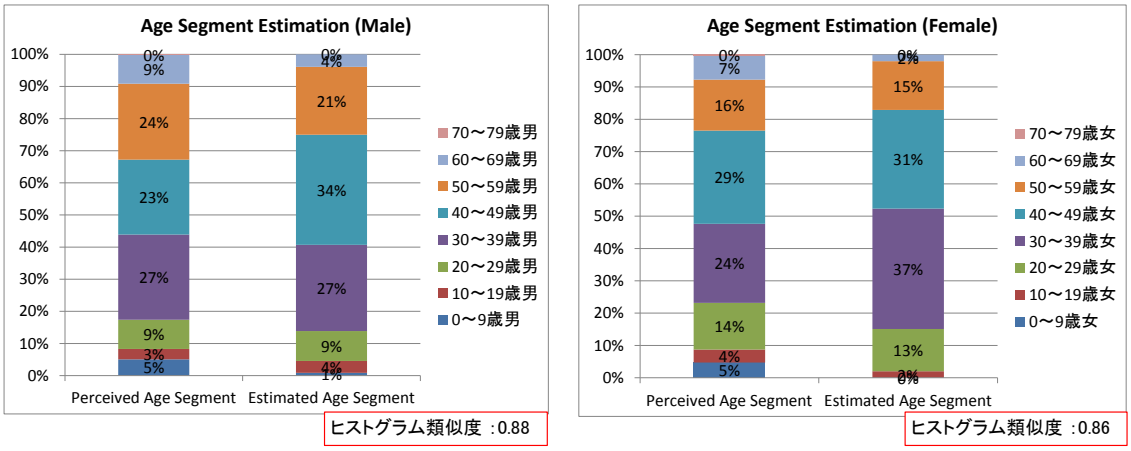


図 5-16　見た目年代推定結果（左：男性，右：女性）

以上の基礎実験より，顔画像のみからでも性別については非常に高い精度で推定でき，また年齢についてもある程度の推定が可能なことが示せた．また，年齢については実年齢よりも見た目年齢を正解値とした場合の方が高い推定性能を示すことが確認された．これは本研究の年齢推定器が顔のアピアランス，つまり見た目情報を用いているため，ある意味人間が見た目で人の年齢を推定するメカニズムに近いからだとも考えられる．

5.4　来場者モニタリングの実環境実験

実環境実験として，ショッピング・モールのエントランス 4 ヶ所に天井埋め込み型ネットワーク・カメラ（Panasonic 製 DG-SC385）を，来場者モニタリング・エリアから水平方向に 20 m 離れた天井高 4 m の天井に設置し，イーサネット経由でサーバー室内の処理用 PC2 台に接続し，1 台の PC でカメラ 2 台分の入力映像を処理するようにシステムを構成した（図 5-17 参照）.

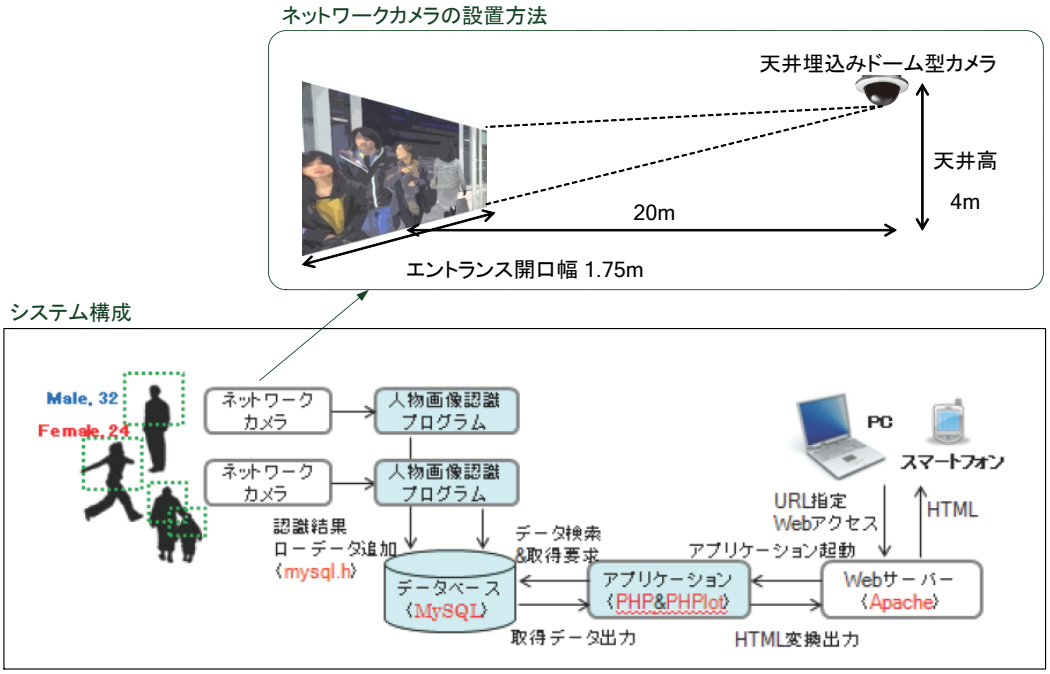


図 5-17　実環境実験システム構成

来場者の撮影画像は，入場者がこちら向き（顔が前向きになり撮像される方向）に，退場者が向こう向き（顔が後ろ向きになり撮像されない方向）に撮影され，その撮影映像がイーサネット経由で Motion JPEG として取得される．照明環境は通常の施設照明のみを使用し，来場者への歩行条件は課していない．実環境実験での混雑度は入場者と退場者を合わせて 30 分間で 331 人であり，このような混雑環境においては来場者が重なり合うため全身が撮影されることは殆どなく，上半身画像による人物検出を行うことで来場者の検出と追跡が可能となった．また 640×480 画素の撮影画像に対して，480×480 画素から 48×48 画素までの 16 段階のスケーリングで上半身検出を行うことで来場者がカメラから遠くにいる場合にも近くにいる場合にも検出を可能とした．動作速度は入力画像内で検出された顔の数に応じて属性推定を行う顔画像数が異なるため処理時間に変

動を生じるが，CPU：X3480(3.07 GHz)，メモリ：4 GB，OS：Windows7 の環境において概ね 3 fps であった．

性能評価として 1 ヶ所のエントランスを選び，3 名の評定者により 30 分間の目視による人数カウントと来場者一人ひとりに対する性別・年代判定を行い，それらの平均をもって実測値とし，本手法による来場者モニタリング結果との比較を行った．上半身検出・追跡による人物の向きによらない来場者と退場者を合わせた人数カウント値は，実測値の 331 人に対して認識結果は 313 人となり，-5.4 %の誤差を生じた．ここで，誤差の主な原因は人物の重なりなどによる上半身未検出であった．

次に来場者と退場者を合わせた 331 人の中で顔が映っている 185 人を対象に，評定者 3 名による性別・年代判定結果を実測値とし，識別結果と比較した．実測値の 185 人に対して顔が検出できたのは 169 人であり，-8.6 %の誤差を生じた．そして，この顔が検出できた 169 人に対して性別を推定した結果を図 5-18 に示す．このとき，性別の実測比率と推定比率のヒストグラム類似度は 0.95 であった．

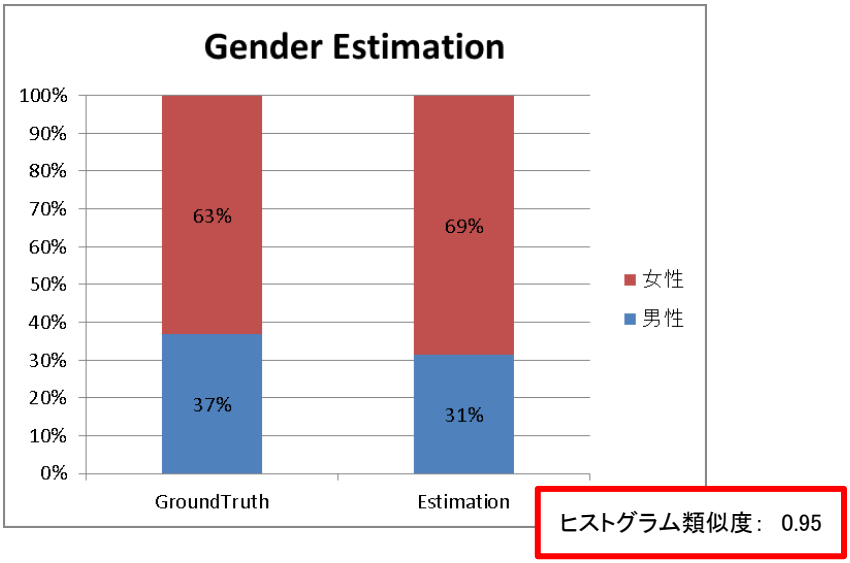


図 5-18　性別推定結果

同様に顔が検出された 169 人に対して年齢を推定し，男女別の 10 歳ごとの年代構成として出力した結果を図 5-19，図 5-20 に各々示す．このとき，男性の年代に対する実測比率と推定比率のヒストグラム類似度は 0.81，女性の年代に対する実測比率と推定比率のヒストグラム類似度は 0.76 であった．

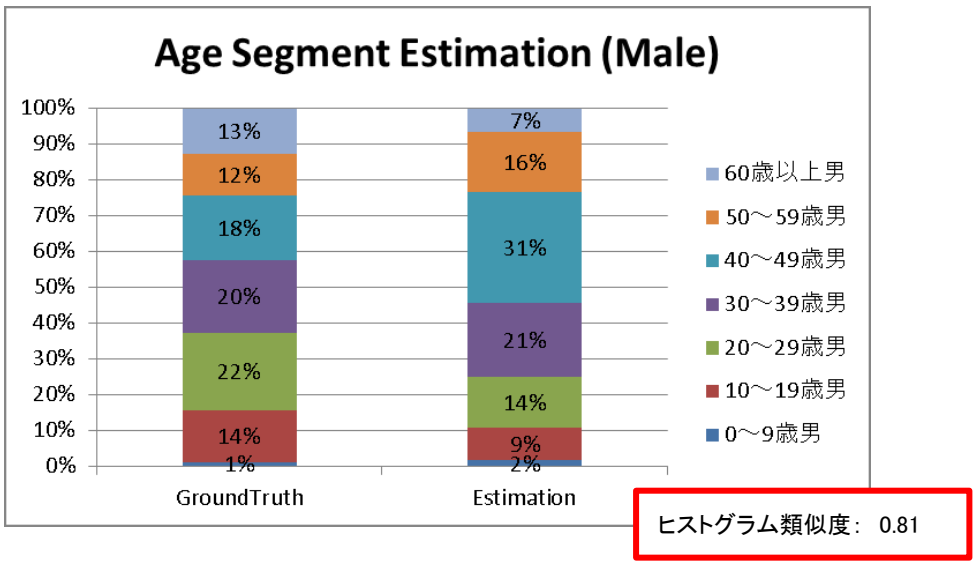


図 5-19　男性の年代推定結果

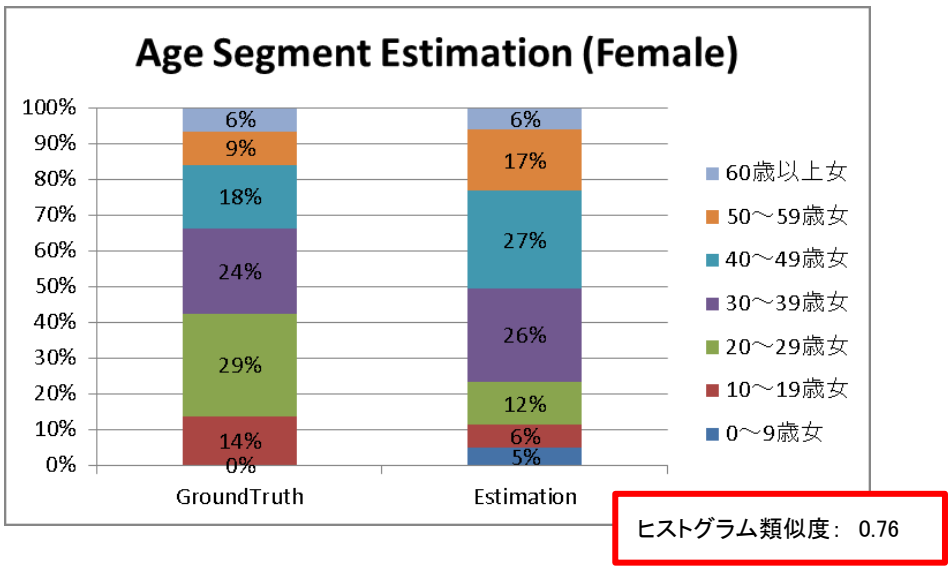


図 5-20　女性の年代推定結果

5.5　まとめ

近年ニーズが増大してきているマーケット調査のためのデモグラフィック分析において，人手によらない安価な来場者モニタリング・システムの実現を目指し，来場者映像のみから人数カウントと人物属性（性別・年齢）を推定する手法について報告を行なった．人物の向きや人物同士の重なりに対してロバストな人物上半身検出とその追跡，アピアランス・ベースの特徴量による顔画像からの人物属性推定を実現し，また人物追跡結果を用いて複数フレーム画像にわたる人物属性推定結果を総合的に判定することで来場者モニタリングの安定性と信頼性の向上を実現した．実験・評価では，はじめに研究用画像データベースを用いた基礎実験によりアピアランス・ベース特徴量での人物属性推定の実現可能性を示した後に，実際の商業施設の出入り口付近に設置したカメラ映像での実環境実験を実施した．その結果，推定精度には改善の余地はあるものの，撮像環境や対象者に対して特に拘束条件を設けない自由度の高い実際の商業施設で撮影された映像においても，人物検出及び顔画像のみからの性別・年齢推定が可能であることを示した．今後は，実環境下での上半身検出，属性推定（特に年齢推定）の精度を更に向上させる検討を行なうと同時に，マルチコア CPU や GPU のアーキテクチャを活用した並列演算処理を検討することで更に処理速度を向上させ，より低価格なプラットフォームでのリアルタイム動作を実現したいと考えている．また産業分野や組み込み分野などの新たな市場ニーズを把握することで，産業の発展，活性化に寄与できることを期待している．

第6章　結論

本論文では，人物の顔映像情報から顔姿勢，表情，性別や年齢といったその人の動作，状態や属性情報を推定する手法とその応用アプリケーションについて提案し，その研究成果について述べた．

第 3 章で述べた顔及び顔器官検出とそのドライバ姿勢推定への応用に関する研究では，夜間でも撮影可能な近赤外線単眼カメラを用いて撮像されたドライバの映像から検出された顔・顔器官の位置情報と 3 次元顔モデルを用いてパーティクル・フィルタによりドライバの顔姿勢を推定し，今後整備が進んでいく道路交通網における路車間協調システムとの協調動作を見据え，自動車運転時における安全性の向上に役立てることを提案した．はじめに道路交通網の中でも特に事故の発生率が高い，市街地の交差点におけるドライバの顔姿勢変化を調査し，映像からの顔姿勢推定に要求されるスペックを明らかにし，それを満たす顔及び顔器官の検出器を構築した．そして実験室内での映像及び実車両での運転映像を用いて，検出された顔器官の位置情報と 3 次元顔モデルの様々な顔姿勢を表現する仮説群から構成されるパーティクル・フィルタにより被験者及びドライバの顔姿勢を広範囲に推定可能なことを示した．ただし，現状ではドライバの顔姿勢変化が顔器官の観測ができなくなるほど大きい場合や姿勢が急速に変化して映像がぼやけてしまう場合などに顔姿勢の追跡･推定が困難である．これらの課題を解決するためには，映像情報処理のみでなくハードウェアも含めた撮像システム全体の構成による解決方法の検討が必要と考えている．また，顔姿勢の変化を伴わない脇見やより高度なドライバの注視行動 [60]の推定にまで応用することを考えた場合，視線のリアルタイム検出とも組み合わせる必要がある．更に，顔姿勢推定の応用範囲としては例えば，対話型ロボットへの適用が検討 [61]されるなど，今後も HMI を支える技術として非常に広範囲に応用されていくことが期待されるが，自動車の特に安全性に関わるシステムの実用化に向けては，より多くの実環境での実験･試験を繰り返し，課題の抽出と解決が必要であり，そのためには今後も道路網との連携も含めた産官学共同で大規模な実証実験を行うなど単なる技術開発に留まらない取組が不可欠だと考える．

第 4 章で述べた笑顔度推定とその音楽療法効果評価への応用に関する研究では，従来は定量的に計測することが困難であった療法効果の評価方法に対し，新たに顔の表情推定結果という数値化された指標を用いることを提案した．独自顔画像 DB と研究用公開画像 DB を用いて映像情報から様々な対象者の微細な笑顔への変化を笑顔度としてロバストに推定できることを示し，その継時変化を多重比較検定することで音楽療法の経過とともに患者の笑顔度が有意に変化していることを示し，表情をリハビリテーション効

果の評価に活用することへの可能性を示した．ただし，本研究で継続してデータが蓄積でき，笑顔度変化の経過を解析できた患者数は 1 名であった．そのため，療法効果に関する今回の結果が多くの患者に共通するかについては，現状では多くを語ることはできず，臨床現場の諸制約のもと粘り強くデータを蓄積して検証を重ねていくことが今後の大きな課題の一つである．その検証過程では，非接触・非拘束の状態で，より安定して顔検出・笑顔度推定ができるように種々の要因（撮影・照明方法，画像解像度，ターゲットとするプログラム等）を体系的に検討し，最適化を図っていくことも必要である．また，4.2.3 節で述べたように，評価尺度として笑顔だけでなく，怒りや悲しみなどの不快感情を利用するための表情検出手法の検討及び不快感情を評価尺度として用いる場合の研究デザインの構築について検討し，新たな療法効果の客観的評価手法についても検討していきたい．また，本研究では脳機能障害に伴い精神機能が当初低調であった症例を取り上げたが，対象を脳卒中や認知症の患者へも広げていくことを考えてみると，例えば脳卒中の後遺症として見られる感情失禁のような病理例に対しては本提案手法を単純にそのまま適用することはできない．従来，感情失禁はその有/無や笑い/泣きの主観的鑑別に基づいて神経心理学的な手法で研究されてきた [62]が，本提案手法の適用を同一刺激下での感情失禁患者と健常者（例えば，患者と共に音楽療法へ参加した家族）とへ拡大し，各々の表情表出の差異比較へと発展させ表情変化の共起性を考慮することができれば，感情失禁の鑑別やその程度をより客観的なものにすることができ，それによって病態機序のより詳細な解明への道が拓ける可能性があると考えている．また，認知症ではその進行とともに感情が平板化して表情表出が乏しくなると云われている．その進行の程度を評価する目的で，古くは Philadelphia Geriatric Center Affect Rating Scale [63]や近年では Psychological Assessment Scale by Facial Expression for Demented People [64]などの評定スケールが提案されているが，何れの手法も評定者の主観に基づく採点を基礎においており，その信頼性は評定者の熟練度などに大きく依存する．本研究で示した笑顔検出技術を改良し，多様な表情で確立できれば，認知症の進行に伴う表情の平板化を何年にも亘って客観的に評価できる可能性がある．例えば，実施プロトコルを適切に設定し，仮に施設利用者を定点観測した映像データを得ることが可能となった場合，そこで得られた映像データは従来と異なって，評定者の熟練度などに依存しないため複数の施設間で相互に比較できる．大規模疫学的調査などでこうした手法を導入することは，効率性・客観性の観点から今後重要性を増していくものと考えられる．

第 5 章で述べた性別・年齢推定とそのデモグラフィック調査への応用に関する研究では，近年その重要性が益々増してきているマーケティング調査の一環として映像情報を活用することを提案した．従来は主に人手によって実施されてきため高コストであった

商業施設等の来場者モニタリングに対し，来場者数とその性別・年代構成比を映像情報から自動的にリアルタイムに解析し，デモグラフィック調査結果として出力する安価に実現可能なシステムを開発した．混雑環境下での人数カウントを考慮した上半身のみによる人物検出・追跡と顔のアピアランス情報のみから性別・年齢を推定する技術を提案し，研究用画像 DB による基礎実験を通じてアピアランス情報のみから性別・年齢が推定可能なことを示すとともに，実環境においても実用的な速度・性能で人物属性推定が可能なことを確認した．今後は，性別・年齢に加え人種 [65] [66]や表情などを加え属性情報を増やしていくことで，より詳細で有用なデモグラフィック調査が可能になると考えられる．また，アピアランス情報に加えて特徴点情報を併用 [67]して推定精度の向上が図られたり，歩容 [68] [69]や全身画像 [70] [71]を利用して顔が見えない状況でも属性を推定する試みがなされたり，今後これらの技術が実用レベルに達すれば，これまで適用が難しかった顔が観測されない状況下でのデモグラフィック調査も可能になることが期待できる．一方で，顔画像と性別や年齢などの人物属性情報は個人情報とも密接な関係にあり，統計的な学習手法を行うためには技術面以外の課題もある．一般に統計的な学習手法を適用する場合，大量の特徴量と正解ラベルとがセットで必要となるが，属性推定においては，この正解ラベルが個人情報にあたる場合もあるため大規模な収集が困難となる．これに対しては，教師なし/半教師学習や見た目属性判定結果を正解ラベルとした検討を行うなど正解ラベル数に捉われず大規模学習が可能な取り組みも行っていく必要がある [72] [73].

最後に本論文では，処理対象として主に実環境で撮影した映像を取り上げた．そのため，顔映像情報のみからでも本論文が対象とした人物状態を推定可能であることを示す一方で，実用化に向けては被写体の撮影に制約が少ない非拘束環境下における，対象者や姿勢変化の多様性，撮像条件の影響に対する課題も明らかとなった．また本論文では対象を人物に限定しているが，昨今では人物のみに留まらず，動物，人工物，風景から細胞に至るまで多岐に渡る映像が画像認識の対象となってきており，映像情報の認識技術とその実社会への応用に関する研究・開発は益々活発になってきている．これら研究が進展しその成果が実用化されれば，我々の日常生活や業務において，従来人間が行ってきた多くの視覚情報処理活動が映像情報・認識処理により自動化され，より安全・安心・快適な社会へと向かっていくことが期待される．本論文で述べた研究成果とその応用事例が，その試みの一つとして将来のそのような社会の実現と，そこで暮らす人々の生活の質の維持・向上のための一助となれば幸いである．

謝辞

本論文の執筆にあたり，主査であり指導教官でもありました筑波大学大学院の福井和広教授におかれましては，在学中の長きにわたり多大なご指導とご鞭撻を賜りました．私の至らない点に対しましても根気強くご指導くださり，社会人学生として休学を挟みながらも本論文執筆までに至ることが叶いました．心より厚く御礼申し上げます．

本論文を審査していただきました牧野昭二教授，栗田多喜夫教授，亀山啓輔教授，伊藤誠教授におかれましては，貴重なお時間を審査に費やしていただき，また貴重なご意見を賜りましたことに御礼申し上げます．

画像認識関連研究テーマの立上げから，産業技術総合研究所での招聘研究員としてのカーブアウト事業化活動に至るまで，研究のご指導とご協力をいただいた旭化成株式会社の野口祥宏氏，顔姿勢推定の共同研究に際し多大なるご協力を頂いた旭化成株式会社の笹原英明氏，マツダ株式会社の山本雅史氏，為貝仁志氏に厚く御礼申し上げます．

笑顔度推定の音楽療法効果評価への応用研究では，共同で研究を進めてくださった産業技術総合研究所の山田亨氏，茨城県立医療大学付属病院の高崎友香氏，国際医療福祉大学大学院の山崎郁子教授，そして研究協力にご同意いただいた茨城県立医療大学附属病院の音楽療法参加者の皆様，及び研究の場を与えてくださった病院関係者の皆様に深く感謝致します．

性別・年齢推定とそのデモグラフィック調査への応用研究では，人物検出処理の実装やデモシステムの構築などにご協力いただいたマノジ・ペレラ氏に深く感謝致します．

画像認識の研究・開発に欠かすことのできない学習・評価データの収集，ラベリング等の画像データベースの整備においては，赤川縁さん，池田詩子さん，佐伯享さん，佐藤亜希子さん，渡邊幸恵さんに日々ご尽力いただき研究活動を支えていただきました．ここに深く感謝致します．

最後に，常に私を支援・応援し温かく見守ってくれた両親と，社会人となってからの学位取得への挑戦に理解を示し，惜しみない協力と励ましで幾度となく私に力を与え助けてくれた家族一同（弘子，理玖）に心より感謝し，本論文の締めくくりとさせて頂きます．

参考文献

[1] 栗田多喜夫, "統計的パターン認識入門－サポートベクターマシン，カーネル学習法の利用と展望－," *第８回画像センシングシンポジウムチュートリアル講演会テキスト,* pp. 11-22, 2002.

[2] 前田英作, "痛快!サポートベクトルマシン －古くて新しいパターン認識手法－," *情報処理,* vol. 42, no. 7, pp. 676-683, 2001.

[3] E. Osuna, R. Freund and F. Girosi, "Training Support Vector Machines: an Application to Face Detection," *In Proceedings of CVPR,* pp. 130-136, 1997.

[4] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing,* vol. 14, no. 3, pp. 199-222, 2004.

[5] T. Matsukawa, A. Hidaka and T. Kurita, "Automatically Evaluation of Degree of Spectators' Satisfaction in Video Sequences Based on Their Facial Expression and Face Directions," *IPSJ,* vol. 50, no. 12, pp. 3222-3232, 2009.

[6] T. Ojala, M. Pietikäinen and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition,* vol. 29, no. 1, pp. 51-59, 1996.

[7] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *e, IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 12, pp. 2037-2041, 2006.

[8] T. Ahonen, A. Hadid and M. Pietikainen, "Face Recognition with Local Binary Patterns," *Computer Vision - ECCV 2004,* vol. 3021, pp. 469-481, 2004.

[9] L. Wolf, T. Hassner and Y. Taigman, "Descriptor based methods in the wild," *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition,* 2008.

[10] H. Jin, Q. Liu, H. Lu and X. Tong, "Face detection using improved LBP under Bayesian framework," *Third International Conference on Image and Graphics (ICIG'04),* pp. 306-309, 2004.

[11] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 7, pp. 971-987, 2002.

[12] M. Heikkilä, M. Pietikäinen and C. Schmid, "Description of Interest Regions with Center-Symmetric Local Binary Patterns," *Computer Vision, Graphics and Image Processing,* vol. 4338, pp. 58-69, 2006.

[13] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* vol. 1, pp. 511-518, 2001.

[14] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* vol. 1, pp. 886-893, 2005.

[15] X. Wang, T. X. Han and S. Yan, "An HOG-LBP human detector with partial occlusion handling," *IEEE 12th International Conference on Computer Vision,* pp. 32-39, 2009.

[16] C. C. Chang and C. J. Lin, "LIBSVM：a library for support vector machines," 2001. [Online]. Available: Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[17] Y. Freund and R. Ｅ. Schapir, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences,* vol. 55, no. 1, pp. 119-139, 1997.

[18] 国土交通省, "ホーム>政策・仕事>道路>ITS," [Online]. Available: http://www.mlit.go.jp/road/ITS/j-html/.

[19] 国土交通省, "ホーム>報道・広報>報道発表資料>ＩＴＳによる安全運転支援システムに係る公開デモンストレーション等の実施について," [Online]. Available: http://www.mlit.go.jp/report/press/jidosha07_hh_000019.html.

[20] 一般社団法人 UTMS 協会, "UTMS ホームページ," [Online]. Available: http://www.utms.or.jp/index.html.

[21] 一般社団法人 UTMS 協会, "安全運転支援システム (DSSS)," [Online]. Available: http://www.utms.or.jp/japanese/system/dsss.html.

[22] 警察庁, "警察庁トップページ＞統計＞交通事故発生状況＞統計表一覧＞交通死亡事故の特徴及び道路交通法違反取締状況について," [Online]. Available: https://www.npa.go.jp/toukei/koutuu48/toukei.htm.

[23] Y. Noguchi, K. Shimada, M. Ohsuga, Y. Kamakura and Y. Inoue, "The Assessment of Driver's Arousal States from the Classification of Eye-Blink Patterns," *Proc. of 12th International Conference, HCI International 2009,* vol. LNAI 5639, pp. 414-423, 2009.

[24] 嶋田 敬士, 野口 祥宏, 大須賀 美恵子, 井上 裕美子, "眼瞼映像を用いたドライバ覚醒状態推定," *第 24 回生体・生理工学シンポジウム論文集, BPES 2009,* Vols. 3A2-5, pp. 325-328, 2009.

[25] SmartEye, "SmartEye 社ホームページ," [Online]. Available: http://smarteye.se/.

[26] 岡兼司, 佐藤洋一, 中西泰人, 小池英樹, "適応的拡散制御を伴うパーティクルフィルタを用いた頭部姿勢推定システム," *電子情報通信学会論文誌 D,* Vols. J88-D2, no. 8, pp. 1601-1613, 2005.

[27] M. Isard and A. Blake, "CONDENSATION—Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision,* vol. 29, no. 1, pp. 5-28, 1998.

[28] 篠原雄介, "パーティクルフィルタの理論と応用～動きまわる物体を追いかけろ！ビジュアルトラッキングの新定番～," *第 14 回画像センシングシンポジウムチュートリアル講演会テキスト,* vol. チュートリアル 2, pp. TU2-1-TU2-30, 2008.

[29] A. Doucet, N. d. Freitas and N. Gordon, Sequential Monte Carlo Methods in Practice (Information Science and Statistics), Springer, 2010.

[30] C. J. C. Burges, "Simplified support vector decision rules," *13th International Conference on Machine Learning,* vol. 96, pp. 71-77, 1996.

[31] B. Schoelkopf, C. J. C. Burges and A. J. Smola, "16. Reducing the Run-time complexoty in Support Vector Machines," in *Advances in Kernel Methods*, 1998, pp. 271-284.

[32] 総務省統計局, "統計局ホームページ/人口推計/人口推計(平成 26 年 10 月 1 日現在)," [Online]. Available: http://www.stat.go.jp/data/jinsui/2013np/index.htm.

[33] 日本音楽療法学会, "日本音楽療法学会ホームページ," [Online]. Available: http://www.jmta.jp/index.html.

[34] P. Ekman and W. V. Friesen, Unmasking the face. A guide to recognizing emotions from facial cues, Prentice Hall, 1975.

[35] B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," *PATTERN RECOGNITION,* vol. 36, no. 1, pp. 259-275, 2003.

[36] K. Yonemura and A. Sugiura, "An assessment of cerebral disease by using face image synthesis," *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE,* vol. 1, pp. 406-414, 2000.

[37] S. Takasu, A. Sugiura and K. Yonemura, "Suggestion of the dependency diagnosis support system focused on the expression change," *World Congress on Medical Physics and Biomedical Engineering 2006,* vol. 14, pp. 400-404, 2007.

[38] 二俣泉, 音楽療法の設計, 春秋社, 1999.

[39] M. N. Dailey, G. W. Cottrell and R. Adolphs, "A Six-Unit Network is All You Need to Discover Happiness," *In TwentySecond Annual Conference of the Cognitive Science Society,* pp. 101-106, 2000.

[40] O. Deniz, M. Castrillon, J. Lorenzo, L. Anton and G. Bueno, "Smile Detection for User Interfaces," *ISVC '08: Proceedings of the 4th International Symposium on Advances in Visual Computing, Part II,* pp. 602-611, 2008.

[41] C. Shan, S. Gong 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.

[42] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer Vision and Image Understanding,* vol. 115, no. 4, pp. 541-558, 2011.

[43] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett and J. Movellan, "Developing a practical smile detector," *In Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition,* 2008.

[44] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett and J. Movellan, "Toward Practical Smile Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, no. 11, pp. 2106-2111, 2009.

[45] K. Shimada, H. Sasahara, Y. Noguchi, M. Yamamoto and H. Tamegai, "Detection of Driver's Face Orientation for Safety Driving Assistance," *Transactions of Society of Automotive Engineers of Japan,* vol. 41, no. 3, pp. 775-780, 2010.

[46] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2010,* pp. 94-101, 2010.

[47] T. Kanade, J. F. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," *Proceedings on Fourth IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 46-53, 2000.

[48] Machine Perception Laboratory, University of California, San Diego, "The MPLab GENKI Database, GENKI-4K Subset," [Online]. Available: http://mplab.ucsd.edu.

[49] Wallhoff, Frank;Technische Universitat Munchen, "Facial Expressions and Emotion Database," [Online]. Available: http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html.

[50] P.エクマン(著), W.V.フリーセン(著), 工藤 力(訳編), 表情分析入門, 誠信書房, 1987.

[51] R. Brunelli and T. Poggio, "Caricatural effects in automated face perception," *Biological Cybernetics,* vol. 69, no. 3, pp. 235-241, 1993.

[52] 瀧川 えりな, 細井 聖, "顔画像による自動性別・年代推定," *Omron technics,* vol. 43, pp. 37-41, 2003.

[53] B. Moghaddam , M. H. Yang, “Gender classification with support vector machines,” *IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 306-311, 2000.

[54] B. Moghaddam and M. H. Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 707-711, 2002.

[55] Y. Fu, Y. Xu and T. S. Huang, "Estimating Human Age by Manifold Analysis of Face Pictures and Regression on Aging Features," *IEEE International Conference on Multimedia and Expo,* pp. 1383-1386, 2007.

[56] 安本 護, 林 純一郎, 輿水 大和, 丹羽 義典, 山本 和彦, "平均顔との距離を用いた性別・年令推定手法の提案," *電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎,* vol. 101, no. 426, pp. 1-6, 2001.

[57] 山口 真美, 加藤 隆, 赤松 茂, "顔の感性情報と物理的特徴との関連について－年令/性の情報を中心に," *電子情報通信学会論文誌 A,* vol. 79, no. 2, pp. 279-287, 1996.

[58] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision,* vol. 7, no. 1, pp. 11-32, 1991.

[59] Y. H. Kwon and N. D. V. Lobo, "Age classification from facial images," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp. 762-767, 1994.

[60] 山城 賢二, 高橋 友和, 井手 一郎, 村瀬 洋, 樋口 和則, 内藤 貴志, "ドライバの注視行動を利用した視線計測システムの自動校正," *電子情報通信学会論文誌 D,* Vols. J92-D, no. 8, pp. 1308-1316, 2009.

[61] T. Vatahska, M. Bennewitz and S. Behnke, "Feature-based head pose estimation from images," *IEEE-RAS International Conference on Humanoid Robots,* pp. 330-335, 2007.

[62] E. C. Lauterbach, J. L. Cummings and P. S. Kuppuswamy, "Toward a more precise, clinically-informed pathophysiology of pathological laughing and crying," *Neuroscience & Biobehavioral Reviews,* vol. 37, no. 8, pp. 1893-1916, 2013.

[63] M. P. Lawton, "Quality of life in Alzheimer disease," *Alzheimer Disease & Associated Disorders,* vol. 8, no. 3, pp. 138-150, 1994.

[64] 矢野 啓明, 高橋 伸佳, 斯波 純子ほか, "重度認知症患者における視線と表情による簡易心理評価スケールの開発," *高次脳機能研究 (旧 失語症研究),* vol. 32, no. 2, pp. 312-319, 2012.

[65] 瀧川 えりな, 細井 聖, 川出 雅人, "顔画像による人種推定技術(顔とジェスチャの認識)," *電子情報通信学会技術研究報告. HIP, ヒューマン情報処理,* vol. 103, no. 454, pp. 19-24, 2003.

[66] W. Gao and H. Ai, "Face Gender Classification on Consumer Images in a Multiethnic Environment," in *Lecture Notes in Computer Science*, vol. 5558, 2009, pp. 169-178.

[67] Z. Xu, L. Lu and P. Shi, "A hybrid approach to gender classification from face images," *International Conference on Pattern Recognition,* pp. 1-4, 2008.

[68] C. Barclay, J. Cutting and L. Kozlowski, "Temporal and spatial factors in gait perception that influence gender recognition," *Perception & psychophysics,* vol. 23, no. 2, pp. 145-152..

[69] X. Li, S. J. Maybank, S. Yan, D. Tao and D. Xu, "Gait components and their application to gender recognition," *Part C: Applications and Reviews, IEEE Transactions on Systems, Man, and Cybernetics,* vol. 38, no. 2, pp. 145-155, 2008.

[70] M. Collins, J. Zhang, P. Miller and W. Hongbin, "Full body image feature representations for gender profiling," *IEEE International Conference on Computer Vision Workshop,* pp. 1235-1242, 2009.

[71] L. Cao, M. Dikmen, Y. Fu and T. S, "Gender recognition from body," *In Proceedings of the 16th ACM international conference on Multimedia,* pp. 725-728, 2008.

[72] K. Ueki, M. Sugiyama and Y. Ihara, "Semi-supervised Estimation of Perceived Age from Face Images," *VISAPP,* pp. 319-324., 2010.

[73] K. Ueki, M. Sugiyama and Y. Ihara, "Perceived Age Estimation from Face Images," in *Advanced Biometric Technologies*, INTECH Open Access Publisher, 2011, pp. 325-342.

研究業績一覧

【本論文に関連する参考論文】

本論文の主要部分は下記参考論文の(1)～(6)として公表済みである.
また各章との関連は以下の通りである.
第 3 章に関連した論文：(1)
第 4 章に関連した論文：(2), (3), (4), (6)
第 5 章に関連した論文：(5)

・公表済み論文

(1) 嶋田 敬士, 野口 祥宏, 笹原 英明, 山本 雅史, 為貝 仁志,「ドライバの安全確認行動検知のための顔向き検出技術の開発」, 自動車技術会論文集, Vol.41, No.3, 2010, pp.775-780

(2) Keiji Shimada, Tetsu Matsukawa, Yoshihiro Noguchi and Takio Kurita, "Appearance-Based Smile Intensity Estimation by Cascaded Support Vector Machines", Proc. of The 2nd International Workshop on Video Event Categorization, Tagging and Retrieval, LNCS 6468, 2010, pp.277-286

(3) Keiji Shimada, and Yoshihiro Noguchi, "Fast and Robust Smile Intensity Estimation by Cascaded Support Vector Machines", Proceedings of 2010 3rd International Conference on Machine Vision, 2010, pp.76-81

(4) Keiji Shimada, Yoshihiro Noguchi and Takio Kurita, "Fast and Robust Smile Intensity Estimation by Cascaded Support Vector Machines", International Journal of Computer Theory and Engineering, Vol.5, No.1, 2013, pp.24-30

(5) 野口 祥宏, 嶋田 敬士, マノジ ペレラ, 栗田 多喜夫,「人物画像認識による来場者モニタリング」, 精密工学会誌, Vol.80, No.1, 2014, pp.89-93

(6) 嶋田 敬士, 山田 亨, 高崎 友香, 野口 祥宏, 山崎 郁子, 福井 和広,「SVM による笑顔度推定技術を用いた音楽療法効果の評価」, 情報処理学会論文誌, Vol.55, No.12, 2014, pp.2569-2581

【その他の論文】

・公表済み論文

(1) Keiji Shimada, Hiroyuki Sasaki and Yoshihiro Noguchi, "The Home Networking System Based on IEEE1394 and Ethernet Technologies", Proc. of 2001 IEEE International Conference on Consumer Electronics, THAM13.3, 2001, pp.234-235

(2) Yoshihiro Noguchi and Keiji Shimada, “A Voice Activity Detection using Driver’s Lip Image for Bimodal Speech Recognition System in Automobile”, Proc. of 11th International Conference on Human-Computer Interaction, Vol.6, Human Factors Issues in Human-Computer Interaction, 2005

(3) Yoshihiro Noguchi, Keiji Shimada, Mieko Ohsuga, Yoshiyuki Kamakura and Yumiko Inoue, “The Assessment of Driver’s Arousal States from the Classification of Eye-Blink Patterns”, Proc. of 12th International Conference, HCI International 2009, LNAI 5639,2009, pp.414-423

(4) Mieko Ohsuga, Yoshiyuki Kamakura, Yumiko Inoue, Yoshihiro Noguchi, Keiji Shimada and Masami Mishiro, “Estimation of Driver’s Arousal State Using Multi-dimensional Physiological Indices”, Proc. of 14th International Conference, HCI International 2011, LNAI 6781, pp.176-185

(5) Yuka Takasaki, Ikuko Yamazaki, Toru Yamada, Keiji Shimada and Yoshihiro Noguchi, “Emotional evaluation using the facial expression detection software during group music therapy sessions. -a secondary report-”, 16th International Congress of the World Federation of Occupational Therapists, RE 2-6-5, 2014

(6) 嶋田 敬士，野口 祥宏，大須賀 美恵子，井上 裕美子，「眼瞼映像を用いたドライバ覚醒状態推定」, BPES 2009 第 24 回生体・生理工学シンポジウム論文集, 3A2-5, 2009，pp.325-328

(7) 松川 徹，嶋田 敬士，野口 祥宏，栗田 多喜夫，「複数特徴量のカスケード型サポートベクターマシンによる猫の顔検出」，第 16 回画像センシングシンポジウムダイジェスト集，IS4-14，2010

・査読のない発表論文

(8) 嶋田敬士，佐藤 宏介，千原 國宏，「マイクロスペース VR システム」，第 39 回自動制御連合講演会，1996，pp.333-334

(9) 嶋田敬士，野口祥宏，「IEEE1394 と Ethernet による統合ホーム・ネットワーク・システムの構築」，情報処理学会第 62 回全国大会，S5-3-7，2001

(10) 野口祥宏，嶋田敬士，佐々木裕之，「IEEE1394 と Ethernet による統合ホーム・ネットワーク・システム」，映像情報メディア学会コンシューマ・エレクトロニクス研究会，ITE Technical Report 25(61)，2001，pp.67-71

(11) 嶋田敬士，野口 祥宏，三代 真己，大須賀 美恵子，鎌倉 快之，井上 裕美子，「ドライバの覚醒状態推定(4) –眼瞼映像からの開口度検出–」，自動車技術会 2009 年春季大会，学術講演会前刷集，No.3-09，覚醒状態推定，2009，pp.13-16

(12) 嶋田敬士，野口 祥宏，笹原 英明，山本 雅史，為貝 仁志，「ドライバの安全確認行動検知のための顔向き検出技術の開発」，自動車技術会 2009 年秋季大会，学術講演会前刷集， No.139-09，生体計測，2009，pp.5-8

(13) 大須賀 美恵子，鎌倉 快之，井上 裕美子，野口 祥宏，嶋田 敬士，三代 真己，「多次元生理指標を用いたドライバの覚醒状態推定」，自動車技術会 2010 年春季大会，学術講演会前刷集，No.65-10，ドライバセンシング I，2010，pp.7-12

(14) 大須賀 美恵子，鎌倉 快之，井上 裕美子，野口 祥宏，嶋田 敬士，三代 真己，「多次元生理指標を用いたドライバの覚醒状態推定(2)」，自動車技術会 2011 年春季大会，学術講演会前刷集，No.74-11，精神負担，2011，pp.21-26

(15) 嶋田 敬士，野口 祥宏，山田 亨，山崎 郁子，高崎 友香，三ツ井 詠子，小堀 英子，丸山 徳子，山崎 克江，長嶋 律子，池田 恭敏，岩井 和子，「SVM による笑顔度推定技術を用いた音楽療法効果の評価方法に関する検討」，情報処理学会第 75 回全国大会，講演論文(分冊 4)，コンピュータと人間社会，バイオインフォマティクスと医療，2013，pp.557-558

(16) 片桐 章宏，加藤 智之，中川 竜太，長濱 克昌，嶋田 敬士，「タブレット端末センサーを用いた個人識別技術の検討」，電子情報通信学会研究会パターン認識・メディア理解研究会(PRMU)，電子情報通信学会技術研究報告，113(75)，2013，pp.35-40

【解説記事等】

(1) 嶋田 敬士，野口 祥宏，マノジ ペレラ，「カメラ映像からのデモグラフィック調査」，画像ラボ，日本工業出版，Vol.23，No.4，2012，pp.8-15

【表彰等】

(1) 野口 祥宏，嶋田 敬士，マノジ ペレラ，「カメラ映像からの来場者デモグラフィック調査」，TX テクノロジー・ショーケース ベスト産業実用化賞，2012 年 1 月