博　士　論　文

人間の内部状態推定のための動作センシング

Sensing Human Actions for Mental State Estimation

東京大学大学院
情報理工学系研究科
電子情報学専攻

48-57416　熊　野　史　朗

指導教員　佐藤 洋一 准教授

平成20年12月

概 要

理想的なヒューマン・マシン・インタラクションシステムの能力の一つとして，単にユーザの指示を待つだけでなく，システムが自律的にユーザの感情や意図といった心的状態を推定し，各場面で望ましいサービスを提供できることが挙げられる．しかし，心的状態はいわば人間の内部状態であり基本的に直接観察することができない．よって，システムは，日常的に発現される観察可能なユーザの動作を手がかりとして，その内部状態を推定するというアプローチをとる必要がある．そこで，本研究では，人間の心的状態に関する多くの情報を含む動作として，特に表情と注視動作の認識に関する研究を行う．

まず，対話において内部状態に関する多くの情報を含む表情を，単眼動画像に基づき，対象人物の頭部姿勢に関わらず正しく推定することが可能な手法を提案する．従来手法は複雑な顔モデルを用いるために，ステレオシステムや事前の膨大な学習データの収集を要するなどの問題があった．そこで，本研究では，その問題の解決を目指し，その場で簡単に個人に特化したものとして作成可能な変動輝度テンプレートと呼ぶ新たな顔モデルを用いた手法を提案する．変動輝度テンプレートは，形状モデル，顔部品の周辺に離散的に配置した注目点の集合，および，それらの注目点の表情変化による輝度変化のモデルからなる．提案手法では，この変動輝度テンプレートを用い，本論文にて提案するパーティクルフィルタと勾配法を組み合わせた推定方法により頑健かつ効率的に表情と頭部姿勢を同時に推定する．

次いで，運転の場面において，コンテクスト情報を用いた，ドライバの意図の異なる注視動作を識別するための手法を提案する．近年の重大事故の主要な原因の一つである前方不注意には，ミラーの死角を確認するといった安全性を高めるための前方不注意と，景色に見とれるといった逆に事故リスクの増大を承知した前方不注意とがある．提案手法の特長は，これらの前方不注意に含まれる意図の違いの重大さに注目している点にある．従来は，これらの 2 種類の前方不注意が区別されずに一つの前方不注意として検出されてきたのに対し，提案手法ではこれら 2 種類の前

方不注意，および，前方注視の計 3 種の注視動作の識別を初めて実現する．観測変数としては，ドライバの身体動作である頭部姿勢に加えて，コンテクスト情報としてペダルやステアリングといった運転操作，および，車速などの運転状況という計 3 種類の情報を用いる．

本論文にて提案する以上 2 つの異なるタイプの動作の認識手法，すなわち，様々な情報が統合された観測データからの複数動作の認識手法，および，コンテクストに依存する動作の認識手法は，様々な情報から人間の内部状態を推定する手法を構築する上でいずれも重要な要素技術となると考える．

目 次

図 目 次

第1章　序論

1.1　本研究の背景

理想的なヒューマン・マシン・インタラクションシステムの能力の一つとして，単にユーザの指示を待つだけでなく，システムが自律的にユーザの感情や意図といった心的状態を推定し，各場面で望ましいサービスを提供できることが挙げられる．しかし，それらの人間の内部状態は，基本的には直接観察することができない．よって，システムは，図 1.1 に示すような日常的に発現される観察可能な様々なユーザの動作を手がかりとして，その内部状態を推定するというアプローチをとる必要がある．

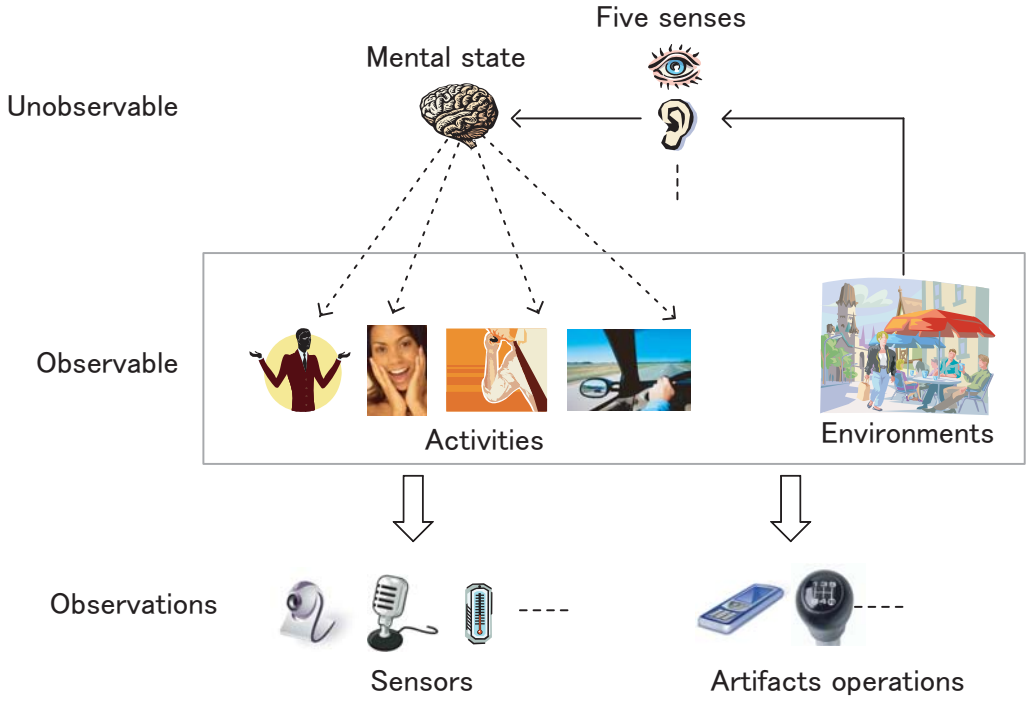


図 1.1: 人間の心的状態，動作および観測量の関係

様々なセンサを日常環境中に多数配置させ，人間行動の計測や支援を行おうとする考えは以前からある．その代表的なものとして，ubiquitous computing [93]，ambient intelligence [19]（環境知能 [110]），human computing [59] などが挙げられる．これらの考えは抽象的であり，実際にどのようなセンサを用いるのか，ユーザのどういった状態を推定するのか，また，そこからどのようなサービスを提供するのかといったことについて千差万別の組み合わせが考えられる．一部実験的に実用化されている部分もあるが，現時点においても，どのようなセンサが人間動作の計測に適しているかといったボトムアップ的なアプローチと，どういったサービスが考えられ，そのためにどのような心的状態の情報が必要かというトップダウン的なアプローチの両面からの研究が行われている．その一方で，用いるセンサ，推定すべき内部状態や提供するサービスが比較的明確な場面がある．例えば，コミュニケーション（対話）科学や ITS(Intelligent Transportation Systems) が挙げられる．まず，対話の場面では，画像や音声からユーザの感情を推定し，対話を円滑にすることが第一の対話支援として考えられる．一方，運転の場面では，ドライバの画像や音声に加えて，車の操作情報や運転状況といった様々な情報を統合的に利用して，覚醒度や安全に対する姿勢といったドライバの内部状態を推定し，それをもとに安全かつ快適な運転を支援することが考えられる．特に，これらの場面では，遠隔対話（たとえば [37]）やインテリジェントカー（たとえば [14]）といったような観測に必要なセンサ環境が整ってきていることから，内部状態推定技術の早期確立が望まれる．以上の背景を踏まえ，本研究ではこのような対話や運転についての内部状態推定につながる人間動作の認識に関する研究を行うこととする．

1.2　本研究の目的

本研究の目的は，人間の内部状態の推定のための人間動作の認識手法を構築することである．そこで，まずは，我々人間が行っている多様な動作を階層的に分類する方法について説明し，本研究で対象とする人間動作の概念のレベルを明確にすることとする．

人間動作についての階層的な分類方法についてはこれまでにいくつか提案されている [7,47,65]．たとえば，動きの機械的知覚に関する先駆的な定義を行った Nagel [65] は，動きをより上位の概念から順に，history, episode, verb, event, change の 5 段階

表 1.1: 人間行動・動作の階層構造の例

Mental states	感情，意図，姿勢など	
Activities	対話 [105]	運転
Actions	役割 [20]	基本運転行動
Action primitives	注視対象の変化 表情の表出 ジェスチャ	注視対象の変化 先行車両との間隔の調整 車両水平位置の調整
Movements （物理レベル）	頭部姿勢の変化 眼球移動 眉毛の移動 発話ピッチ（高低）の変更	頭部姿勢の変化 眼球移動 アクセルペダル踏込量 ハンドルの回転

に分類した．他方，Bobick [7] は actions, activity, movement という 3 段階に分類し，また，Kruger ら [47] は activity, action, action/motor primitives という 3 段階の用語を定義している．Bobick および Kruger らの分類方法では activity についての定義が異なるものの，それ以外の動作については定義が類似しており，それらを併せると activity, actions, action primitives, movements の 4 段階となる．本論文では，この 4 段階の動作分類に従って議論する．

対話および運転を例とした際の人間動作の階層構造の一部を，内部状態を含めて表 1.1 に挙げる．各レベルの概念についての定義は以下のとおりである．まず，最上位の activity とは，主要タスクとして位置付けることのできる程度の大きな概念の行動カテゴリである．表 1.1 の例では，対話および運転そのものがこれに相当する．その次の action とは，activity を遂行するための基本的な機能である．表 1.1 の例では，話し手／聞き手／傍観者といった対話での役割 [20] や，直進／右左折／停止などの基本運転行動がこのレベルに当たる．3 つめの action primitive とは，経験的な知識にもとづく何らかの意味をもつ動作の最小単位である．表 1.1 の例では，表情の表出や前方注視，先行車両との間隔の調整などがこれに相当する．これに対し，最も低いレベルの movement とは，解釈に特定の知識を必要としない，単なる物理現象として記述可能な動作である．このため，movement レベルの動作は単に計測器の出力として扱うことができる場合がある（たとえば，接触型の磁気センサは姿勢情報を出力する）．これらの動作や心的状態を認識するためには，基本的に，観測から低次のレベルの動作，そして高次のレベルの動作や心的状態へと，計測・認識をボトムアップ的に進めていく必要がある．本研究では，内部状態と関わりの深い

action primitive レベルの動作として，特に，対話の場面で感情に関連した様々な情報を含む表情，および，運転の場面で意図や安全に対する姿勢といった情報を含む注視動作を主な認識対象とする．

人間動作を認識するために用いるセンサとしては様々なものが考えられるが，対象動作の特徴をよく抽出可能なセンサを常に使用できるとは限らない．例えば，眉毛の上昇／下降といった表情に関連した動作については，筋電計を顔面に貼り付けて表情筋に与えられている刺激を測定することが，最も確実に表情情報を獲得可能な方法であると考えられる．だが，少なくとも対話の場面ではこのような方法は通常許容されず，したがって非接触型である画像センサを用いるのが妥当な選択肢となる．しかし，画像中には推定したい表情の情報以外にも，対象人物の頭部姿勢の情報などが表情情報と結合した形で含まれてしまっている．このため，画像から表情を正しく認識するためには頭部姿勢も同時に推定する必要がある．しかし，従来の表情と頭部姿勢を同時に推定可能な手法は，複雑な顔モデルを用いるがゆえに，ステレオシステムや事前の膨大な学習データの収集を要するなどの問題があった．そこで，本研究では，その問題の解決を目指し，作成の容易な新たな顔モデルとそれを用いた表情と頭部姿勢を同時に推定する手法を提案する[1]．

一方，動作や内部状態の推定には状況（コンテクスト）情報が重要となることも多い．つまり，ある 1 つの動作の意味（より上位レベルでの解釈）が，その動作が行われた状況により変化する場合である．たとえば運転の場面を考えると，重大事故の主要原因である前方不注意 [113] がそのような動作として挙げられる．この前方不注意には，ミラーの死角を確認するという運転に関わる前方不注意（死角確認）と，景色に見とれるといった運転とは直接関係のない対象への注視による前方不注意（脇見）とがある [114]．これらは身体動作的には類似しているものの，これらの安全に対する姿勢は正反対である．つまり，前者が安全性を高めるために行われるのに対し，後者はむしろ事故リスクの増大を承知で行われる．しかし，従来，これらの 2 種類の前方不注意は区別されずに一つの前方不注意として検出されてきた．そこで，本研究では，それらの前方不注意に含まれる意図の違いの重大さに注目し，これら 2 種類の前方不注意動作，および，前方注視の計 3 種の注視動作を初めて識

[1]本論文では，主眼を表情の認識に置き，頭部姿勢については，図 1.1 での *movement* レベルとして取り扱う．すなわち，物理的な位置および姿勢を計測するに留め，そこから *action primitive* レベルの注視対象や頷き動作の認識は行わない．

別可能な手法を提案する．観測としては，前方不注意を主に特徴づける *movement* レベルのドライバの身体動作に加え，死角確認と脇見を識別しするためのコンテクスト情報として，ペダルやステアリングなどの運転操作情報，および，車速などの運転状況の情報を扱う．

1.3　推定の枠組み

1.2 節で述べた本論文で対象とする 2 つの動作の認識は，2 つの異なる基本的な動作認識のタイプに分類される．それは，様々な情報が統合された観測データからの複数動作の認識手法，および，コンテクストに依存する動作の認識手法である．本研究の大目的である人の内部状態推定を可能とするには，多数の観測，様々な動作および心的状態を統一的に扱える枠組みが必要となる．よって，本論文にて述べるこれらの 2 つの動作の認識においても，そのような統一的な枠組みを適用することとする．以下に，そのような統一的な枠組みに対する望ましい特性を列挙する．

1. 不確定性を表現可能
   観測（計測器の出力など既知の情報）にはノイズが混入している場合がほとんどである．さらに，認識対象である人間行動にも個人間のばらつきはもちろん，個人内のばらつきが多く含まれる．そのため，このような数々の不確定性を扱えることが望ましい．

2. 事前知識を導入可能
   我々は，もちろん，認識の対象としている人間の内部状態や動作についての事前知識を豊富に持っている．よって，それらを簡単にモデルに導入できることが望ましい．

3. 観測の一部欠損への対処が可能
   観測として様々な要素を扱う場合，その一部が観測できないという場面が多々生じ得る．そのとき，その観測欠損による影響を最小限に抑えることができることが望ましい．

4. 要因の追加や削除が容易
   最終的には多様な人間動作や心的状態を組み込んだ複雑なモデルとなること

が考えられる．よって，局所的なモデルを複数構築し，それらを統合することで全体のモデルを作り上げるというアプローチが適していると考えられる．また，観測についても，技術の進展に伴うセンサの取捨選択などが考えられる．

5. 必要な学習サンプル数が少ない
   一般的に統計的なアプローチでは，必要な学習サンプル数が扱う要因の次元数に対して指数的に増加する（次元の呪い (curse of dimensionality)）．このため，問題を回避できることが望ましい．

6. 要素のダイナミクスを表現可能
   動作や心的状態は時間変化するものであるため，同一要素間の時間遷移を含めた要素間の時間を跨ぐ関係を記述できることが望ましい．

本研究では，これらの特性をいずれも満たす推定の枠組みとして，動的ベイジアンネットワーク（DBN: Dynamic Bayesian Networks）を用いる．これは，時間方向を含めた要素間の因果関係を確率的に記述するモデリング手法の一種である．なお，動的ベイジアンネットワークがこれらの特性を満たす理由については付録にて述べる．

1.4　本論文の構成

本論文は，以下，3 つの章から構成される．

まず，2 章にて，単眼動画像に基づき，表情および頭部姿勢を同時にする手法を提案する．本研究では，その場で簡単に個人に特化したモデルを作成可能な変動輝度テンプレートと呼ぶ新たな顔モデルを用いた手法を提案する．変動輝度テンプレートは，形状モデル，顔部品の周辺に配置した離散的な注目点の集合，および，それらの注目点の表情変化による輝度変化をモデル化したものからなる．提案手法は，この変動輝度テンプレートを用い，本論文にて提案するパーティクルフィルタと勾配法を組み合わせた頑健かつ効率的な推定の枠組みで頭部姿勢と表情を同時に推定する．評価実験では，様々な頭部姿勢において表情をどの程度正しく認識できるか，および，表情と頭部姿勢が同時に変化する状況に対処可能かどうかを中心に検証する．

次いで，3 章では，運転中のドライバの死角確認と脇見における安全に対する意図の違いの重大さに注目し，それら 2 つの前方不注意に加えて，前方注視という 3 種類の注視動作を識別可能な手法を初めて提案する．観測変数には，ドライバの身体動

作である頭部姿勢に加え，ペダルやステアリングといった運転操作，および，車速などの運転状況という計3種類の情報を用いる．注視動作と観測変数の間の関係の記述には動的ベイジアンネットワークを用い，観測が与えられたもとでのそれぞれの注視動作の事後確率をリアルタイムにて逐次的に計算する．ドライビングシミュレータを用いた実験により提案手法の有効性を検証する．

最後に，4章にて本研究のまとめを行う．

第2章　表情と頭部姿勢の同時推定

2.1　はじめに

近年，ヒューマンコンピュータインタラクションをはじめとして，様々な分野で人物の顔の表情認識が脚光を浴びており，これまでに画像に基づく表情認識手法が多数提案されてきた．それらの手法の多くはほぼ正面を向いた顔画像を対象としたものである [2,12,15,43] が，対象人物は必ずしも常に正面を向いているとは限らない．例えば，複数人対話中の人物は，しばしば，他の対話参加者に対して顔を向ける [68]．このため，このようなシーン中の人物の表情を正しく認識するためには，頭部姿勢も同時に推定する必要がある．

頭部姿勢を変化させる人物の表情を認識するための従来のアプローチは，頭部姿勢変動および表情変化を分離してモデル化しておき，それを用いて入力画像中の頭部姿勢および表情を推定するというものが一般的である．中でも，顔の形状モデルおよびその変形のモデルを用意し，頭部姿勢変動を形状モデルの大局的な並進・3 次元回転として，表情変化を形状モデルの局所的な変形として表現するものが多い [18,31,58,66,104]．ここでは，形状モデルおよび表情変化のモデルを併せて顔モデルと呼ぶ．このアプローチは精緻な顔モデルを必要とする．なぜなら，表情変化による顔画像の変化は頭部姿勢変動による顔画像の変化に比べて小さく，粗い顔モデルを用いた場合には頭部姿勢変動と表情変化の分離が困難となるためである．

また，表情認識システムには，不特定多数の人物に対して適用できることが求められる．従来，この要求に対して，人物依存の顔モデルを利用の場で作成するというアプローチと，非人物依存の顔モデルをあらかじめ準備するという 2 つのアプローチによる対処が行われてきた．前者のアプローチは，その人物に特化した精緻なモデルを獲得することにより，精度の高い表情認識を実現することを目指すものである [31,66]．しかし，このアプローチは，精緻な顔形状モデルを獲得するためにステレオシステムなどの装置を要し，そのため，適用可能な場面が限定されてしまうと

いう問題がある．一方，後者のアプローチは，複数の人物の顔形状および表情変化についての個人間の変動を含めたモデルを作成することで，人物に依らない表情認識の実現を目指す手法である [18,58,104]．しかし，多数の人物についての学習データの準備が煩雑である上，未学習の人物に対するモデルの精度が学習済みの人物に対して低いといった問題がある [32].

動画像に基づく頭部姿勢変化に対処可能な表情の推定手法としては，中でも，サンプリング手法の一種であるこのパーティクルフィルタ (PF) [41] の有用性が，文献 [18,31] などをはじめとして多く確認されている．複数の仮説を用いて事後確率密度分布を近似的に推定する PF の特長は，局所解への陥りにくさや一時的な追跡の失敗からの回復性といった推定の頑健性にある．だが，仮説を確率的に生成するため，精度の高い推定結果を得るのに概して計算コストを要するという欠点を持つ.

この問題点の改善を目指した PF の改良手法がこれまでにいくつか提案されている．それらの多くは，推定の精度や効率を高めるために各仮説をなるべく尤度の高い領域に配置させることを目指したものである．例えば，proposal distribution に現在の観測を取り入れる方法 [72,89] や，一旦生成した仮説群を勾配法を用いて移動させる方法 [11,81] などがある．しかし，これらは，事後分布を単峰の正規分布で近似する文献 [89] をはじめとして，基本的に尤度がなだらかであることを暗に仮定している．このため，これらの PF の改良法を適用した頭部追跡手法には，たとえば文献 [11,73,81,99] のように，そのような尤度関数を作成しやすいよう，頭部姿勢を本来の 6 自由度（3 次元位置および 3 次元回転）ではなく，回転について面内方向のみを扱うなどより低い自由度で扱うものが多い．しかし，6 自由度の頭部姿勢にさらに表情を加えた状態を正しく分離可能であり，かつ，なだらかな尤度関数を設計することは難しいことが予想される.

また，必要以上に仮説を生成することを避けるために，推定の不確定性に着目して仮説数を適応的に変化させる方法も提案されている [28,81]．この方法は，合計の処理時間を短縮可能なため，オフライン処理には有用である．しかし，運動モデルの崩壊を招くフレーム落ちを避ける必要のあるオンライン処理に対しては根本的な解決とならない.

本研究では，上記の従来手法の問題の解決を目指し，以下の特長を有する新たな表情認識手法を提案する．その特長とは,

1. 単眼システムである
2. 頭部姿勢変動に対する高い頑健性を有する
3. 対象人物に特化した顔モデルを容易に作成できる
4. 頑健かつ効率的な推定を行うことができる

の 4 つである．これらを実現するため，本研究では，表情変化を形状モデルの局所的な変形ではなく顔の輝度変化によって表現するというアプローチを考え，表情認識のための新たな顔モデルを提案する．さらに，頑健かつ効率的な推定を実現するための方法として，パーティクルフィルタと勾配法を組み合わせた推定手法を提案する．

本論文で提案する顔モデルは，形状モデル，注目点集合，および，表情輝度分布モデルからなる．形状モデルには，実際の顔形状を計測したもの，あるいは，顔形状の近似モデルとして円柱 [10] や楕円体 [3] といった単純な幾何形状など，様々な形状を使用することが可能である．注目点とは，目や口といった顔部品において，顔部品と非顔部品の境界（エッジ）から少し離れたところに離散的に配置される点のことである．表情輝度分布モデルとは，各注目点についての，怒り，喜びといった表情のカテゴリに依存した輝度分布からなる混合輝度分布モデルのことである．本論文では，このような顔モデルのことを変動輝度テンプレートと呼ぶ．この変動輝度テンプレートは，単眼システムにおいて対象人物がその場で表出した表情の画像から，直ちにその人物に特化したものとして作成することができる．このため，本手法は，マンマシンインタフェースをはじめとして幅広い場面に対して適用可能である．

この変動輝度テンプレートを用いた表情認識の原理，および，その効果は以下のとおりである．注目点が顔部品の近傍に配置されているため，表情変化に伴う顔部品の移動により注目点の輝度が変化する．本研究ではこの性質に着目し，予め各表情について獲得しておいた輝度と入力の輝度との照合を行うことで表情を認識する．また，提案手法は，構築が容易な近似的な顔の形状モデルを用い，しかも，同時に頭部姿勢変動への頑健性も実現することを目指すものである．これを実現するためには，次の問題に対処する必要がある．それは，近似的な形状モデルを使用した場合，対象人物の顔の向きが注目点を定義したときと同じ方向（本手法では正面）で

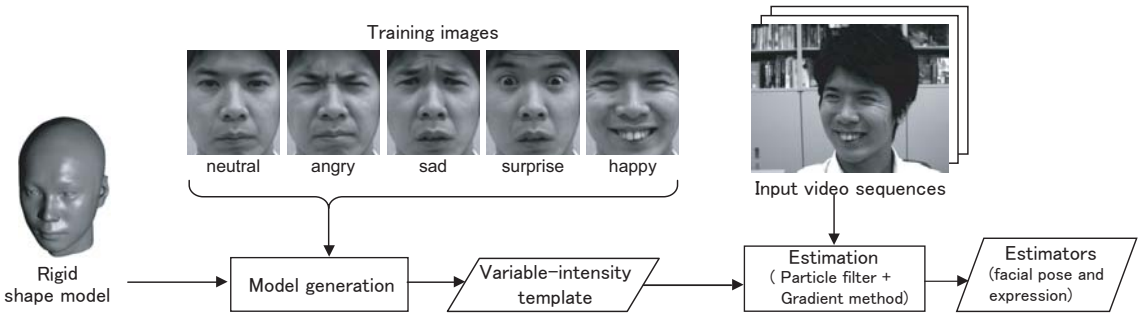


図 2.1: システムフロー

なければ，注目点の算出位置が実際の位置に対してずれてしまうという問題である．そこで，本手法では，注目点の算出位置がずれたとしても，その算出位置における輝度が実際の位置での輝度から大きく外れることがないよう，注目点を空間的な輝度変化の大きい顔部品のエッジから少し離して（マージンを設けて）配置する．

また，本論文では，単眼動画像に基づいた，頑健かつ効率的な推定を実現するための新たな推定手法を提案する．提案手法では，まず，パーティクルフィルタ (PF) により近似的な推定値を頑健に得た後，それを初期値として勾配法により高速に最尤推定値を探索する．(図 2.1 参照)．この方法は，(1) 勾配法の反復計算において，最初の段階でのみ事前情報（予測分布）を正則化項として含む，一種の罰則付き最尤推定法であり，(2) この予測分布を安定に得る手段として PF を用いる手法と解釈できる．確率的探索手法（ここでは PF）に確定的探索（ここでは勾配法）を組み合わせる試みは，著者の知る限り，少なくとも表情認識においては初めてのものである．

提案手法は，事前に準備した変動輝度テンプレートを用い，パーティクルフィルタと勾配法を組み合わせた枠組みにて，頭部姿勢および表情を同時に推定する．この推定の段階では，仮説に頭部姿勢および表情の状態を持たせ，その頭部姿勢に従って形状モデルを並進・回転させた位置での各注目点の輝度と，その表情についての輝度分布との照合を行うことで，それらを逐次的に推定する．

以下，最初に，本表情認識手法について説明し，そして，実験結果，考察と続け，最後に，本論文のまとめを述べる．

2.2　関連研究

これまでに提案されてきた表情認識手法の多くは表面顔，すなわち，対象動画像中の人物が常にカメラに対してほぼ正面を保ったままであることを仮定している（たとえば文献 [2, 12, 26, 39, 46, 52, 84]）．ただし，中には，対象人物の頭部姿勢変動を考慮したものもある．表情と頭部姿勢は基本的にはそれぞれ独立に変動するとみなせるが，それらによる画像の変化は，通常，表情変化による変化の方が，頭部姿勢によるそれよりも小さい．

このため，表情を正しく推定するためには，表情成分と頭部姿勢成分の精緻な分離が必要となる．しかしながら，大きな面外方向の頭部姿勢変動のもとでは，高次に非線形な画像の見えの変化が生じるため，これら成分の分離が困難となる．さらに，非正面顔を含む公開されたデータベースの数少ない（たとえば Face Video Database[1]）．このため，これまで提案されてきた表情認識の性能比較は困難であった[2]．

本節では，表情認識手法の中でも，面外の頭部姿勢変動に対処可能な手法について議論する．ここでは，それらを使用する顔形状のモデルに注目しつつ整理することとする．なお，正面顔を仮定した数多くの手法については，[25, 69, 70, 85] などで広くサーベイされているので，そちらを参照されたい．

ここでは，関連研究を使用する顔形状モデルをもとに 5 つのグループに分類する．その 5 つとは，(1) 計測装置を用いて計測された顔形状，(2) 単眼画像から推定された形状，(3) 事前に準備された汎用的な形状，(4) 単純幾何形状，(5) 顔形状を用いないアプローチである．この分類方法は，以下のように解釈できる．まず，これらの形状モデルは，(2) および (5) を除いて，顔の 3 次元形状の近似精度が高いと思われる順に整列されている．また，(1) から (4) のグループについては，さらに，3 種類の分類が可能である．計測された形状 (1) ／非計測形状 (2-4)，個人特化モデル (1,2) ／汎用モデル (3,4)，および，変形可能モデル (1-3) ／剛体モデル (4) である．

(1) 計測装置を用いて測定された形状

[1] Face Video Database of the Max Planck Institute for Biological Cybernetics in Tuebingen: *http://vdb.kyb.tuebingen.mpg.de/*.

[2] 近年，BU 4D-FE DB [97] というデータベースが公開されている．このデータベースは，意図的な表情ではあるものの，100 人程度の被験者についての，3 次元形状と画像の時系列データからなる．頭部姿勢変動を伴う表情を認識するという問題はまだ未解決の研究分野であるが，この BU 4D-FE DB の登場により表情認識研究の進展の加速が予想される．

頭部姿勢変動に対する頑健性を向上させる最も効果的な方法は，対象人物の 3 次元の顔形状を単眼カメラ以外の何らかの計測装置を用いて実際に計測することである．Gokturk ら [31] は，ステレオカメラを用いた表情認識手法を提案している．彼らは，顔面上の 19 つの特徴点の 3 次元位置を計測し，それらの点の移動をオプティカルフロー推定に類似した手法をもとに算出し，表情を認識している．評価実験では，様々な面外方向の頭部姿勢のもとでも表情を正しく認識できることが示されている．なお，ステレオカメラを用いて，表情による変形と頭部姿勢変化を同時に扱う手法は，頭部追跡の分野でも見られる（例えば，[66,104]）．また，画像からではなく，3 次元形状から表情を認識するという手法も近年提案されている [82,92]．彼らは，まず対象人物の 3 次元形状をアクティブステレオ法により計測し，表情表出による顔形状の変化を特徴量として表情を認識している．これらの手法は，一旦，3 次元の顔形状，および，表情変化によるその変形を獲得できれば，頭部姿勢に頑健な表情認識が可能であることを示唆している．だが，これらの手法は，システムに単眼カメラ以外の計測機器を要するため，適用可能な場面が限定的であるという問題がある．

(2) **単眼画像から推定された形状**

付加的な計測機器を要するという計測された形状を用いた手法の問題点を解決できる可能性のあるアプローチとして，3 次元の顔形状，および，表情変化による変形のモデルを単眼動画像から獲得しようとする試みもある．Xiao ら [94] は，structure-from-motion 技術を用い，2D+3D Active Appearance Models (AAM) を自動的に獲得する手法を提案している．まず，顔面上の特徴的な点を 2 次元の AAM を用いて追跡する．そして，それらの特徴点の 2 次元座標の時系列データから，形状成分および表情変化成分を分離する．Lucey ら [58] は，その復元された 2D+3D AAM を用いて，表情を認識する手法を提案している．しかし，彼らの実験では，復元された AAM が十分な精度でなく，むしろ表情認識率を低下させるという結果が得られている．

(3) **事前に準備された汎用的な形状**

以上の 2 つは対象人物専用のモデルを使用する手法であった．ここでは，1 つのモデルを不特定多数の人物に対して使用するアプローチをとる手法を紹介する．Cohen ら [15] は，まず，汎用的なワイヤフレームモデル上に定義された特徴点を Piecewise Bezier Volume Deformation (PBVD) トラッカを用いて追跡し，その後，それらの特

徴点の運動パラメタを時系列の特徴量として表情を認識する手法を提案している．Dornaika ら [18] は，CANDIDE [1] と呼ばれる既存の変形可能な顔モデルを用いたシステムを提案している．CANDIDE の変形パラメタは，FACS の AU に近似的に対応している．この手法では，各時刻において頭部姿勢と表情を順に推定する．まず，現時刻における頭部姿勢を，変形成分を一時刻前の推定結果に固定した状態で勾配法を用いて推定する．次いで，頭部姿勢をその推定値に固定した状態で，顔面上の特徴点の位置を特徴量として，パーティクルフィルタ [41] を用いて表情のカテゴリと各 AU の表出強度を同時に推定する．しかし，Gross ら [32] は，顔の形状および変形の個人間の変動は大きく十分な汎化性能が得られない，つまり，汎用モデルでは特に未学習の人物に対して高い識別性能が得られにくいという結果を報告している．

(4) **単純幾何形状**

(1-3) のような複雑な形状ではなく，単純な幾何形状を用いた表情認識手法もいくつか提案されている．Black ら [6] および Tong ら [87] は平面を，Liao ら [54] は円柱をそれぞれ用いている．特徴量には，Black ら [6] および Liao ら [54] は顔面内のオプティカルフローを，Tong ら [87] はガボールウェーブレット係数を用いている．しかし，単純な幾何形状では各特徴点のアラインメント誤差，すなわち，形状モデル上に定義された特徴点を画像上に射影したときの位置ずれ量を低く抑えることができない．よって，頭部の面外回転量が大きくなるに従い，表情の認識率が低下することが予想される．Tong らは，AU 間の共起関係を含む動的ベイジアンネットワークを学習し，それを推定に利用する手法を提案している．簡単に言えば，少数の AU の推定の失敗を，それと共起しやすい他の AU の正しい検出により補うというモデルである．彼らの実験では，そのような AU 推定の補償の効果が示唆されている．しかし，頭部の面外回転角が大きくなるにつれ，推定に失敗する AU の割合が増加するとともに，この補償の枠組みが逆に誤認識を助長する方向に進むことも予想される．さらに，オプティカルフロー推定はそれ自体が照明変化や対象物体の変形などに弱く，フロー算出自体に失敗すればそれが表情認識に大きな影響を及ぼすという問題がある．

(5) **顔形状を用いないアプローチ**

Hu ら [38] は，顔形状を用いない手法を提案している．彼らは，Support Vector Machines (SVMs) を用いて，離散的な頭部姿勢（各方向に 15 度以上の間隔），および，表情カテゴリを，順にあるいは同時に推定する手法を提案している．彼らのシステムは正面からほぼ真横方向まで広く対応可能であるが，離散的な頭部姿勢についてのみ学習するため，それ以外の頭部姿勢についてはモデルでは表現しきれない誤差が表情認識率の低下を招く可能性が高い．とはいえ，頭部姿勢の間隔を密にして学習することは，処理速度やメモリ資源の観点から好ましくない．しかも，彼らは顔面領域のみを含む画像を暗に仮定しているが，実際にはセグメンテーションを行わなければならず，その誤差による誤識別も生じるはずである．

まとめると，著者の知る限り本研究の目標を全て達成する手法は存在しないと言える．つまり，

- 単眼動システムであること．他の計測装置を用いて計測された形状 (1) は，この要件を満たさない．
- 個人特化の顔モデルを容易に作成可能なこと．汎用モデル (3) や形状モデルを使用しない手法 (5) は，モデル作成の複雑さからこの要件を満たさない．
- 大きな頭部姿勢変動のもとでも表情を正しく推定できること．推定された形状 (2) や単純幾何形状形状 (4) はこの要件を満たさない．

本論文の提案手法は，以上の要件を満たすことが可能な新たな注目点輝度ベースのアプローチを採る．提案手法は，表情変化を，形状モデルの変形や顔領域のオプティカルフローなどから推定するのではなく，顔の局所領域の輝度変化のみから推定する．さらに，本研究では，面外の頭部姿勢変動に対する表情認識率の変化についての検証を行う．このような評価は，これまで，[38, 92] 以外では検証されていなかった，なお，従来の各文献では，それぞれ異なるデータベースを用いた評価が行われており，手法の定量的な比較は難しい．

2.3　提案手法

本手法は，図 2.1 に示すとおり，まず，変動輝度テンプレートを準備し，それを用い，パーティクルフィルタと勾配法を組み合わせた枠組みにて入力動画像中の人物

の頭部姿勢および表情を同時に推定する．この変動輝度テンプレートの作成に必要な画像は，図 2.1 に示したとおり，認識対象とする表情それぞれにつき 1 枚の正面顔画像（学習画像と呼ぶ）である．

以下では，まず，単眼動画像，表情および頭部姿勢の間の関係を記述する動的ベイジアンネットワークについて述べる．次いで，尤度の算出に用いる顔モデルについて説明する．さらに，その顔モデルに基づき尤度関数を定義する．続いて，表情と頭部姿勢を同時に推定するためのパーティクルフィルタと勾配法を組み合わせた手法について説明する．

2.3.1　表情および頭部姿勢の同時推定のための動的ベイジアンネットワーク

本手法では，頭部姿勢状態および表情状態がそれぞれ独立な 1 次マルコフ過程に従うことを仮定し，時刻 t における顔画像 $\boldsymbol{z}_t$ がそのときの頭部姿勢状態 $\boldsymbol{h}_t$ および表情状態 e_t に依存して観測されているものと考える．これらの関係は，図 2.2 に示す動的ベイジアンネットワーク (Dynamic Bayesian Network: DBN)[3]にて表現される．ここで，頭部姿勢状態 $\boldsymbol{h}$ は，形状モデル中心の入力画像上での位置 (t_x, t_y)，カメラに正対する顔向きを基準とした頷き（上下），首振り（左右），傾げ（面内）に対応する頭部姿勢角 $(\theta_x, \theta_y, \theta_z)$，および，スケール s の 6 連続変数からなる．一方，表情状態 e は，認識対象とする表情のカテゴリを表す離散変数である（たとえば，無表情，怒りや悲しみなどである）．

この動的ベイジアンネットワークのもとでの本研究の目的は，各時刻 t における頭部姿勢 $\boldsymbol{h}_t$ および表情 e_t の同時事後確率密度分布 $p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t})$，すなわち，対象時刻 t までに観測可能な全ての顔画像 $\boldsymbol{z}_{1:t}$ が与えられたもとでの頭部姿勢 $\boldsymbol{h}_t$ および表情 e_t の同時確率密度分布，を推定することである．この同時事後確率密度分布 (joint posterior probability distribution function (pdf)) $p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t})$ は，ベイズフィルタリング（例えば文献 [74] 参照）と呼ばれる方法で逐次的に算出可能である．その算出

[3] より詳しくは factorial DBN と呼ばれるタイプの DBN である [63].

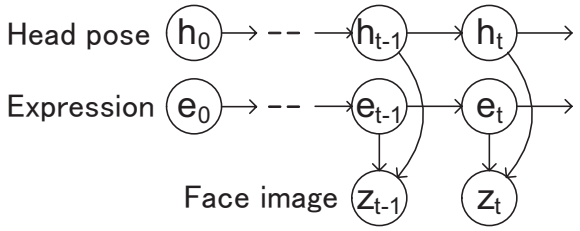


図 2.2: 頭部姿勢，表情および顔画像の関係を表す動的ベイジアンネットワーク

式は，

$$\begin{aligned} p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t}) &= \alpha p(\boldsymbol{h}_t, e_t, \boldsymbol{z}_t|\boldsymbol{z}_{1:t-1}) \\ &= \alpha p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t) p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t-1}) \\ &= \alpha p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t) \int p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}) \sum_{e_{t-1}} P(e_t|e_{t-1}) \\ &\qquad p(\boldsymbol{h}_{t-1}, e_{t-1}|\boldsymbol{z}_{1:t-1}) d\boldsymbol{h}_{t-1} \end{aligned} \tag{2.1}$$

と表わされる．ここで，α は正規化定数であり，$p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t)$ は観測画像 $\boldsymbol{z}_t$ に対する頭部姿勢 $\boldsymbol{h}_t$ かつ表情 e_t の状態の尤度である．以下では特に，この尤度が $\boldsymbol{h}_t$ および e_t の関数であることを明確にするため，この尤度を $L(\boldsymbol{h}_t, e_t|\boldsymbol{z}_t)$ と表す．また，$p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$ および $P(e_t|e_{t-1})$ は頭部姿勢および表情のそれぞれについての状態遷移モデル，$p(\boldsymbol{h}_{t-1}, e_{t-1}|\boldsymbol{z}_{1:t-1})$ は前時刻 $t-1$ で得られた事後分布である．以下では，尤度 $P(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t)$ の定義，頭部姿勢および表情それぞれについての状態遷移モデル $p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$ および $P(e_t|e_{t-1})$，そして，頭部姿勢 $\boldsymbol{h}_t$ および表情 e_t の同時事後確率密度分布 $p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t})$ の具体的な計算方法について順に説明する．

2.3.2　顔モデル

本論文では，表情および頭部姿勢を推定するための新たな顔のモデルを提案する．これを変動輝度テンプレートと呼ぶこととする．変動輝度テンプレートは，剛体の形状モデル $\mathcal{S}$，注目点集合 $\mathcal{P}$，および，表情輝度分布モデル $\mathcal{I}$ の 3 つの要素から構成される．表情輝度分布モデルは，各注目点の輝度が，各表情によってどのように変化するかをモデル化したものである．提案手法で用いる変動輝度テンプレートは，形状モデル $\mathcal{S}$ については汎用モデルであるが，注目点集合 $\mathcal{P}$ および表情輝度分布モ

デル$\mathcal{I}$については個人特化モデルである．なお，この拡張として 2.4.2 節では，いずれの要素も汎用モデルである変動輝度テンプレートについて議論する．

変動輝度テンプレートは学習画像のセット，および，剛体形状モデルから生成される．学習画像は，図 2.1 に示すような，対象ユーザの各対象表情につき 1 枚の正面顔画像$\{\boldsymbol{g}_{e=1}, \cdots, \boldsymbol{g}_{e=N_e}\}$である[4]．ここで，$N_e$は認識対象とする表情のカテゴリ数である．特に，無表情についての学習画像については，$\boldsymbol{g}_{\mathrm{NEU}}$と表す．各ユーザについてのこれらの学習画像$\boldsymbol{g}$中の顔は，画像間で仮想的に静止しているものとする．これらの学習画像の作成方法については 2.7.1 節にて述べる．

形状モデル$\mathcal{S}$

形状モデルは，学習画像中に定義される各注目点の 3 次元座標を獲得するためのものである．さらに言うと，様々な頭部姿勢における各注目点の画像中での位置を計算するためのものである．なお，この注目点の 3 次元座標の算出方法の詳細は 2.3.3 節にて述べる．

形状モデルには，実際の顔形状を計測したもの，あるいは，顔形状の近似モデルとして円柱 [10] や楕円体 [3] といった単純なパラメトリックな幾何形状など，様々な形状を使用することが可能である．本論文では，形状モデルとして，平均顔形状[5]，を用いる．なお，この平均顔形状を各ユーザに対してフィットさせる方法については 2.7.1 節にて述べる．

注目点集合$\mathcal{P}$

注目点集合は，無表情の正面顔画像$\boldsymbol{g}_{\mathrm{NEU}}$において，目や口などの顔部品の近傍に離散的に配置した注目点の集合である．図 2.3 にその一例を示す．このような注目点を用いることで，表情変化に伴い顔部品が移動したときに，注目点の輝度変化からその表情変化を検出することが可能となる．ここでは，注目点集合$\mathcal{P}$を，

$$\mathcal{P} = \{\boldsymbol{p}_1, \cdots, \boldsymbol{p}_N\} \tag{2.2}$$

[4]このように学習画像に対して表情をラベルを付与しているところに表情についての解釈のルールが含まれている．よって，認識対象の表情は表 1.1 でいうところの *action primitive* レベルの動作である．他方，表情と同時に推定する頭部姿勢については，単に物理的な状態として扱うため *movement* レベルの動作である．

[5]産業技術総合研究所デジタルヒューマン研究センターの「日本人青年男性の平均頭部ダミー」（*http://www.dh.aist.go.jp/research/centered/facedummy/*）を用いた．

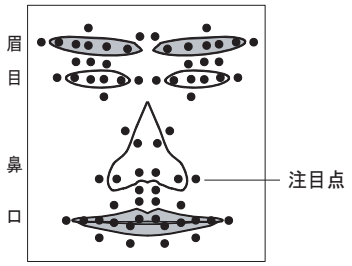
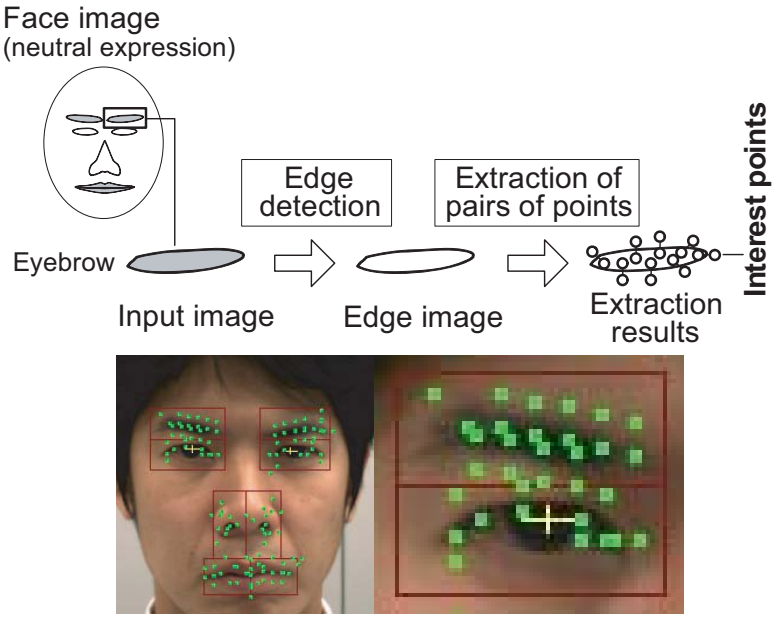


図 2.3: 注目点集合 $\mathcal{P}$ の一例



大きな矩形の枠は各顔部品領域を表す.

図 2.4: 注目点の抽出方法の概要（上）と自動検出された 128 点の注目点集合（下）.

と表す．ここで，$\boldsymbol{p}_i$ は注目点 i の無表情の学習画像 $\boldsymbol{g}_{\mathrm{NEU}}$ における画像座標を，N は注目点の数を表す．この注目点集合 $\mathcal{P}$ は個人毎に異なる．

本論文では，顔部品の近傍に配置されるこのような注目点の一種として，松原・尺長 [108] の境界ダイポール（以下ダイポール注目点と呼ぶ）を用いる．このダイポールとは，エッジ上の 1 点に対して，そのエッジの直交方向に等距離（ここではマージンサイズ）離れた 2 点の組のことである．図 2.4 右上のように，4 つの顔部品（眉，目，鼻および口）の内外にダイポールを複数配置することで，様々な方向への顔部品の移動が検出可能となる（この詳細については，2.7.2 節にて述べる）．顔部

品の検出方法については 2.7.1 節にて述べる．ここでは，選択する注目点の数を 128（眉：20×2，目：12×2，鼻：24，口：40）とする．この数は，計算コストと推定精度を考慮して経験的に決定した．図 2.4 下に，自動抽出された顔部品領域および注目点集合を示す．

このダイポールの抽出方法は，以下の通りである．なお，併せて図 2.4 上図を参照されたい．

(1) 各顔部品領域内でエッジを検出する．本論文では，エッジ検出方法の一つとして，ラプラシアン－ガウシアンフィルタ画像におけるゼロ交差境界をエッジとして検出する．

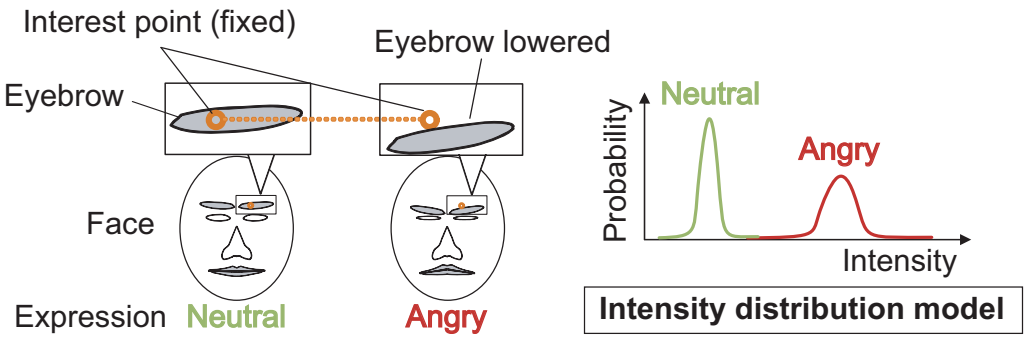
(2) そのエッジを跨ぐ全てのダイポール候補を抽出する．そして，これらの候補をその輝度勾配方向によって 8 つの方向グループ（上，下，左，右，あるいは，斜め 4 方向のいずれか）に分類する．注目点のマージンサイズについては，顔部品毎に経験的に決定した定数を検出された顔領域の幅に乗じた値とする[6]．本論文では，この比例定数の値を，表情変化による移動量の小さい目領域については $1/60$，それ以外の顔部品領域については $1/45$ とした．

(3) 各方向グループについて，ダイポール候補から，ダイポール内の 2 点の輝度差が大きいものから順に 1 つずつ，実際に使用するダイポールを選択する．このとき，選択しようとする候補は，既に選択されているダイポールから，ダイポールの中心（ダイポール内 2 点の中心）間の距離が経験的に決定した閾値以上離れていなければならないとしている．現在の対象の方向グループから 1 つ選択する，あるいは，その方向グループから選択可能な候補がなくなれば，次の方向グループへ移る．初期設定については，経験的に，方向グループを上方向とし，ダイポール間の距離の閾値を対象顔部品の領域の幅の 0.2 倍とした．

(4) 選択したダイポール数が目標数（ここでは $128/2 = 64$）に達し次第，選択を終了する．もし，(3) の条件を満たすダイポール候補が目標数に達しない場合には，ダイポール間の距離の閾値を小さくしたのちに，(3) に戻る．

以上の処理を人物毎に行う．結果として，選択される注目点は人物毎に異なる．なお，(2) のマージンサイズの妥当性ついては，2.7.3 節にて述べる．

[6]実装の上では，これを最近接の整数へと丸めている．



注目点の輝度は表情変化により大きく変化する.

図 2.5: 表情輝度分布モデル $\mathcal{I}$.

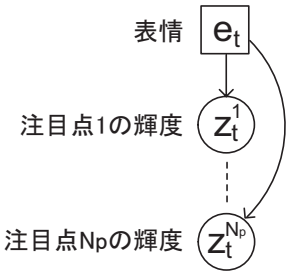


図 2.6: 各時刻 t における表情および注目点輝度の関係を表すベイジアンネットワーク

表情輝度分布モデル $\mathcal{I}$

表情輝度分布モデルは，各注目点の輝度が各表情によってどのように変化するかをモデル化したものである．図 2.5 に示すように，注目点の輝度はその近傍の顔部品の移動により大きく変化する．提案手法は，この輝度分布が各表情によって異なることを利用して，事前に準備した表情輝度分布モデルと入力画像における注目点の輝度とを比較することで，そのときの表情を認識する．

観測される注目点の輝度は表情変化以外の要因によってもばらつく．この主な原因は，注目点のアラインメント誤差，つまり，形状モデルの実際の顔形状に対する誤差による注目点の算出位置のずれや，2.3.4 節にて説明する輝度補正の誤差によるものであると考える．本手法では，これらのばらつきを正規分布にてモデル化する．

さらに，このばらつきが注目点間で独立であるとする．これにより計算コストが低く抑えられる[7]．各時刻 t における表情と注目点の輝度の間の関係をグラフ表示す

[7] ((2.6) 式の計算コストが $O(N_p)$ であるのに対し，注目点の相関を考慮すると，$N_p \times N_p$ の共分

ると，図 2.6 に示すようなナイーブベイズモデルとなる．この表情輝度分布モデル $\mathcal{I}$ を次のように表す.

$$\mathcal{I} = \{\mathcal{I}_i(e)\}_{i=1,\cdots,N_p,\, e=1,\cdots,N_e}, \tag{2.3}$$

$$\mathcal{I}_i(e) = \mathcal{N}\left(\mu_i(e), \sigma_i(e)\right), \tag{2.4}$$

$$\sigma_i(e) = k\mu_i(e). \tag{2.5}$$

ここで，$\mathcal{I}_i(e)$ は表情 e についての注目点 i の表情輝度分布モデル $\mathcal{I}$ を表す．また，$\mathcal{N}(\mu, \sigma)$ は，平均 μ，標準偏差 σ の正規分布を表し，添え字 i および e は，それぞれ注目点および表情を表す．この平均 $\mu_i(e)$ については，表情 e の学習画像 $\boldsymbol{g}_e$ における，座標 $\boldsymbol{p}_i$ での輝度とする．さらに，標準偏差 σ_i については輝度 μ_i に比例するものとする[8]．本論文では，この比例定数を経験的に 0.3 とした.

図 2.7 に，眉上に配置された注目点で実際に観測された，無表情および怒り表情での輝度ヒストグラムの一例[9]を示す．なお，両表情について，それぞれ，最頻値を中心とした大きな分布と，その周辺に存在する小さな分布が見られる．前者の分布が正規分布にて表現される分布である．後者は，大きな注目点の算出位置のずれ（本来眉上にあるはずの注目点が眉外にあると算出されてしまった場合など）によるものであると考えられる．本論文では，2.3.3 節にて詳細を述べるとおり，正規分布では表現しきれないこのノイズによる分布を，正規分布にロバスト関数を導入することで表現する．図 2.8 にその分布形状を示す.

2.3.3　尤度関数

入力画像 $\boldsymbol{z}$ に対する頭部姿勢 $\boldsymbol{h}$ および表情 e の同時尤度 $L(\boldsymbol{h}, e|\boldsymbol{z})$ については，各注目点の輝度が，正規分布で表現されるモデルの輝度とどれだけずれているかに基づき計算する．本論文ではこの尤度を，

$$L(\boldsymbol{h}, e|\boldsymbol{z}) = \prod_{i=1}^{N_p} \frac{1}{\sqrt{2\pi}\sigma_i(e)} \exp\left\{-\frac{1}{2} f_i\left(\boldsymbol{h}, e, \boldsymbol{z}\right)\right\} \tag{2.6}$$

$$f_i\left(\boldsymbol{h}, e, \boldsymbol{z}\right) = \rho\left(d\left(z_i\left(\boldsymbol{h}\right), \mathcal{I}_i\left(e\right)\right)\right) \tag{2.7}$$

散行列を用いた $O(N_p \times N_p)$ の計算が必要となる．

[8]光学的に，照明光が強くなるとそれに比例して顔面上の点の輝度も高くなることから，注目点の輝度のばらつきもそれに伴い大きくなるという考えによる．

[9]図 2.7 左に示す人物がカメラに対して左右様々な方向を向いた動画像に対し，本手法を適用したときに観測される輝度から作成したものである．なお，両表情についての頻度は正規化されている．

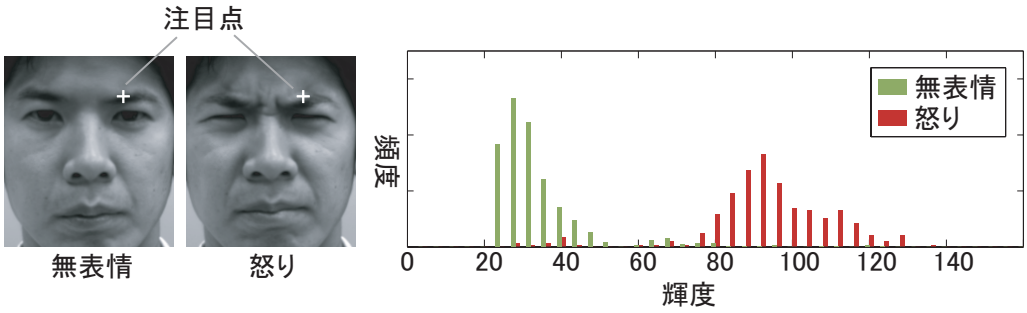


図 2.7: 注目点において観測される輝度ヒストグラムの一例

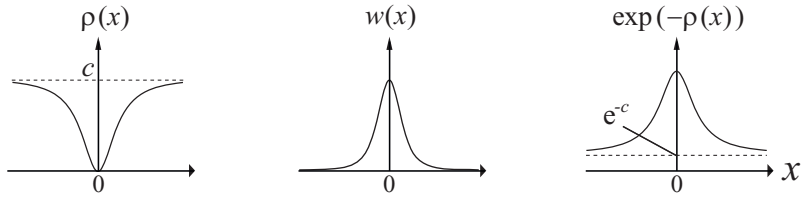


図 2.8: （左）ロバスト関数 ρ,（中）重み関数 w,（右）ロバスト関数を正規分布に組み合わせた分布.

と定義する．ここで，$z_i(\boldsymbol{h})$ は，頭部姿勢が $\boldsymbol{h}$ であるときの注目点 i の画像座標 $\boldsymbol{q}_i(\boldsymbol{h})$ における画像 $\boldsymbol{z}$ で観測される輝度を表す．すなわち，本論文では，各注目点の輝度に対する尤度を正規分布に本節にて後述するロバスト関数 $\rho(\cdot)$ が組み込まれた関数にて表わしている．

距離 $d(\cdot,\cdot)$ については，本論文では次のように定義する．

$$d(z_i, \mathcal{I}_i(e)) = \begin{cases} \dfrac{\tilde{z}_i - \mu_i(e)}{\sigma_i(e)}, & \text{if visible} \\ d_o, & \text{otherwise.} \end{cases} \tag{2.8}$$

ここで，$\tilde{z}_i$ は，観測輝度 z_i を 2.3.4 節で述べる方法により補正した輝度である．もし，注目点における形状モデルの法線ベクトルがカメラの存在する方向を向いていなければ，その点は遮蔽されているとみなして一定の尤度を与えている．

頭部姿勢 $\boldsymbol{h}$ における注目点 i の画像座標 $\boldsymbol{q}_i(\boldsymbol{h})$ については，まず，学習画像の座標 $\boldsymbol{p}_i$ にある注目点 i を，カメラに対して正対させた形状モデル $\mathcal{S}$ 上へ正射影する．次いで，その状態で，形状モデル $\mathcal{S}$ を頭部姿勢 $\boldsymbol{h}$ に従い，並進・回転する．最後に，その状態の形状モデル $\mathcal{S}$ 上にある注目点 i を，入力画像平面へ弱中心投影した座標

を $\boldsymbol{q}_i(\boldsymbol{h})$ とする．以上を数式にて表わすと，

$$\boldsymbol{q}_i(\boldsymbol{h}) = s\boldsymbol{R}_{2\times 3}\boldsymbol{p}_i^{(\text{shape})} + \boldsymbol{t}, \tag{2.9}$$

$$\boldsymbol{p}_i^{(\text{shape})} = g(\mathcal{S}, \boldsymbol{p}_i), \tag{2.10}$$

となる．ここで，$\boldsymbol{p}_i^{(\text{shape})}$ は注目点 i の形状モデル $\mathcal{S}$ 上での 3 次元座標である．これについては，無表情の学習画像において検出された顔領域の中心（2.7.2 節参照）に形状モデル $\mathcal{S}$ の中心を，画像 y 軸に形状モデル $\mathcal{S}$ の回転軸を一致させた状態で，注目点 i の学習画像における画像座標 $\boldsymbol{p}_i$ を形状モデル上へ平行投影する関数 $g(\mathcal{S}, \boldsymbol{p}_i)$ により求める．$\boldsymbol{R}_{2\times 3}$ は頭部姿勢角 $(\theta_x, \theta_y, \theta_z)$ に対応する回転行列 $\boldsymbol{R}$ の上 2 行からなる行列を，$\boldsymbol{t}$ は並進ベクトル $[t_x \;\; t_y]^T$ をそれぞれ表す．

関数 $\rho(\cdot)$ は，

$$\rho(\xi) = c \cdot \frac{\xi^2}{1+\xi^2} \tag{2.11}$$

と表わされるロバスト関数の一種の Geman McClure 関数 [30] である[10]．ロバスト関数を用いることで，2.3.2 節の表情輝度分布の部分で説明した注目点の大きな算出位置ずれなどにより生じる外れ値に対して推定が頑健になる．もし，単純に正規分布の積とした尤度を用いたのでは，ごく一部の注目点の輝度のみが外れ値である場合にも尤度が大きく低下してしまう．

2.3.4　輝度補正

顔面の輝度は，照明変動や上下方向の首振りに伴う相対的な照明方向の変化により変化する．提案手法では，このような外部要因による輝度変動が時間的に激しく起こる状況にも対処できるようにするために，入力画像から観測される各注目点の輝度をフレーム単位で次のように補正する．

$$\tilde{z}_i = \gamma_b \cdot z_i \tag{2.12}$$

ここで，γ_b は，顔のブロック b についての輝度補正係数を表す．本論文では，それらのブロックとして，{左眉＋左目，右眉＋右目，鼻＋口の左部，鼻＋口の右部} の 4 つを用いる．つまり，モデルの輝度に対する観測の輝度の変化の割合が，それぞれ

[10] c は出力を 0 から c にスケーリングする係数である．

(1) 重み行列 $\boldsymbol{W} \in \mathbb{R}^{N_b \times N_b}$ の初期化：
$\boldsymbol{W}^{(0)} = \mathrm{diag}\,[1, \cdots, 1]^{\mathrm{T}}$, ここで，$N_b$ は注目点セット $\mathcal{P}_b$ 中の要素数を表す.

(2) 輝度補正ブロック b の輝度補正係数 $\gamma^{(m)}$ の算出：

$$\gamma_b^{(m)} = (\boldsymbol{z}_b^{\mathrm{T}} \boldsymbol{W}^{(m-1)} \boldsymbol{z}_b)^{-1} \boldsymbol{z}_b^{\mathrm{T}} \boldsymbol{W}^{(m-1)} \boldsymbol{\mu}_b$$

ここで，m は反復回数を，$\boldsymbol{z}_b$ および $\boldsymbol{\mu}_b$ はそれぞれ標準偏差で除した観測輝度および輝度分布モデルの $N_b \times 1$ ベクトル：$\boldsymbol{z}_b = [\cdots, z_i/\sigma_i, \cdots]^{\mathrm{T}}$ and $\boldsymbol{\mu}_b = [\cdots, \mu_i/\sigma_i, \cdots]^{\mathrm{T}}$ である．ここで $i \in \mathcal{P}_b'$ である.

(3) 重み行列 $\boldsymbol{W}$ の更新：

$$W_{ii}^{(m)} = w\left(\frac{\gamma_b^{(m)} z_i - \mu_i}{\sigma_i}\right)$$

ここで，W_{ii} は重み行列 $\boldsymbol{W}$ の i 番目の対角要素を，$w(\cdot)$ はロバスト関数 $\rho(\cdot)$ に関する重み関数：$w(x) = (d\rho(x)\,/dx)/x$ である.

(4) 収束するまでステップ 2 および 3 の繰り返し.

図 2.9: 反復重み付き最小二乗法による輝度補正

の顔のブロックにおいて一様であることを仮定している．この仮定は厳密には成立しないが，若干の輝度補正誤差は表情認識にほとんど影響を及ぼさない．なぜなら，注目点を顔部品の周辺に配置することで，表情輝度分布モデルのもつ輝度パターンは表情毎に大きく異なるためである.

提案手法では，各ブロック b における輝度補正係数 γ_b を，頭部姿勢 $\boldsymbol{h}$ および表情 e が与えられたもとでの，(2.6) 式についての最尤推定値 $\widehat{\gamma}_b$ として算出する.

$$\widehat{\gamma}_b = \arg\min_{\gamma_b} \sum_{i \in \mathcal{P}_b} f_i\,(\boldsymbol{h}, e, \boldsymbol{z}) \tag{2.13}$$

ここで，$\mathcal{P}_b (\subset \mathcal{P})$ は顔ブロック b に属し，かつ，遮蔽されていない注目点の集合を表わす．(2.13) 式は一種のロバスト回帰問題であり反復重み付き最小二乗法 [4] を用いて効率的に算出することができる．そのアルゴリズムを図 2.9 に示す.

(2.13) 式による輝度補正は，新たな頭部姿勢 $\boldsymbol{h}$ または表情 e が与えられる度，すなわち，パーティクルフィルタにおける仮説生成の度（後で述べる (2.15) 式），お

よび，勾配法による頭部姿勢の推定値の更新（(2.25) 式）の度に行われる．

2.3.5　遷移モデル

頭部姿勢の遷移モデル（運動モデル）$p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$ には，各変数がそれぞれ独立なランダムウォークモデルを用いる．本手法では，少ない計算量で仮説を真の状態のより近傍に配置することを目指し，文献 [66] の手法を参考にして，次のようにシステムノイズの大きさを適応的に変化させる[11]．

$$\boldsymbol{h}_t = \boldsymbol{h}_{t-1} + w(\boldsymbol{v}_{t-1}) \tag{2.14}$$

ここで，$\boldsymbol{v}$ は頭部姿勢の速度ベクトル，$w(\boldsymbol{v})$ はゼロ平均ガウス過程のシステムノイズである．システムノイズの共分散行列は対角行列であり，各要素 j は $\boldsymbol{v}$ の j 番目の要素の絶対値に比例する．

表情の状態遷移モデル $P(e_t|e_{t-1})$ については，全ての表情間の遷移が等しい確率とする．なお，提案手法の枠組みでは任意の状態遷移モデルを導入可能であるが，事前に実際の状態遷移モデルを得ることは難しい．そのため，本論文では，事前知識が何も得られていないことを仮定したこのような状態遷移モデルを，妥当な遷移モデルの 1 つとして用いる．

2.3.6　パーティクルフィルおよび勾配法を組み合わせた状態推定

(2.1) 式の頭部姿勢および表情の分布は，遮蔽などにより複雑な形状となるため，解析的に厳密解を得るということができない．このため，本手法はこの頭部姿勢と表情の同時事後分布を，サンプリング手法の一種であるパーティクルフィルタ [41] と勾配法を組み合わせた方法により頑健かつ効率的な推定を実現する．つまり，パーティクルフィルタは頑健だが，目的関数のピークを高精度で得るためには多数の仮説を要するため効率が悪い．一方，勾配法は微分可能な目的関数のピーク探索に適しているが，初期値がピーク値付近になければ局所解に陥るという問題を抱える．提案する推定の方法には両者の利点が組み合わされている．このパーティクルフィルタと勾配法を組み合わせる方法は，勾配法の反復計算において最初の段階でのみ事

[11]つまり，厳密には，頭部姿勢 $\boldsymbol{h}$ については 2 次の過程となっているが，簡単のため図 2.2 や式の上（$p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$）では 1 次の過程として記述している．

前情報（予測分布）を正則化項として含む一種の罰則付き最尤推定法であり，この予測分布を安定に得る手段としてパーティクルフィルタを用いる手法であると解釈することができる．

パーティクルフィルタによる推定

本手法におけるパーティクルフィルタの枠組みでは，頭部姿勢 $\boldsymbol{h}_t$ および表情 e_t の確率分布が，パーティクルと呼ばれる重み付きのサンプル（仮説）集合 $\{\boldsymbol{h}_t^{(l)}, e_t^{(l)}, \omega_t^{(l)}\}_{l=1}^{N_x}$ によって近似的に表現される．ここで，N_x は仮説数を，$\boldsymbol{h}_t^{(l)}$, $e_t^{(l)}$ および $\omega_t^{(l)}$ はそれぞれ時刻 t における l 番目の仮説の保持する頭部姿勢，表情および重みを表す．ここでは要点のみ説明する．パーティクルフィルタの詳細な実行方法および手順については文献 [41] を参照されたい．

頭部姿勢 $\boldsymbol{h}_t^{(l)}$ および表情 $e_t^{(l)}$ は，予測分布 (predictive distribution) と呼ばれる分布 $p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t-1})$ よりサンプリングした 1 つの実現値である．つまり，

$$\{\boldsymbol{h}_t^{(l)}, e_t^{(l)}\} \sim p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t-1}) \tag{2.15}$$

である．ここで，記号 $x \sim X$ は，変数 x を分布 X からサンプリングすることを意味する．この予測分布は，一時刻前の同時事後確率密度分布 $p(\boldsymbol{h}_{t-1}, e_{t-1}|\boldsymbol{z}_{1:t-1})$ をそのときの頭部姿勢 $\boldsymbol{h}_t$ および表情 e_t により周辺化した分布，言い換えると，一時刻前の同時事後確率密度分布 $p(\boldsymbol{h}_{t-1}, e_{t-1}|\boldsymbol{z}_{1:t-1})$ に頭部姿勢および表情それぞれの遷移モデル $p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$ および $P(e_t|e_{t-1})$ を畳み込んだ形となっている．数式にて表すと，

$$p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t-1}) = \int \sum_{e_{t-1}} p(\boldsymbol{h}_t, \boldsymbol{h}_{t-1}, e_t, e_{t-1}|\boldsymbol{z}_{1:t-1}) d\boldsymbol{h}_{t-1} \tag{2.16}$$

$$= \int p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}) \sum_{e_{t-1}} P(e_t|e_{t-1}) p(\boldsymbol{h}_{t-1}, e_{t-1}|\boldsymbol{z}_{1:t-1}) d\boldsymbol{h}_{t-1} \tag{2.17}$$

である．これは，(2.1) 式の第 2 行および第 3 行目の尤度 $p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t)$ よりも右側の成分に対応していることに注意されたい．

重み $\omega_t^{(l)}$ については，

$$\omega_t^{(l)} \propto L(\boldsymbol{h}_t^{(l)}, e_t^{(l)}|\boldsymbol{z}_t) \tag{2.18}$$

とし，$\sum_l \omega_t^{(l)} = 1$ を満たすものとする．顔画像の尤度 $L(\boldsymbol{h}_t^{(l)}, e_t^{(l)}|\boldsymbol{z}_t)$ については，2.3.3 節にて述べた．なお，本手法は，リサンプリング処理（例えば文献 [106]），す

なわち，同じ時刻において 2 回の仮説の生成・重み付けの処理を行うことで，頭部姿勢および表情の事後分布をより効率的に表現している．

各時刻における頭部姿勢および表情の推定値については，仮説集合の期待値をベースに算出する．頭部姿勢の推定値 $\widetilde{\boldsymbol{h}}_t$ については，頭部姿勢についての周辺事後分布 $P(\boldsymbol{h}_t|\boldsymbol{z}_{1:t})$ の期待値とし，表情 e_t については，周辺事後確率 $P(e_t|z_{1:t})$ が最大となる表情を推定表情 $\widetilde{e}_t$ とする．これらを数式にて表わすと，

$$\widetilde{\boldsymbol{h}}_t = \sum_l w_t^{(l)} \boldsymbol{h}_t^{(l)} \tag{2.19}$$

$$\widetilde{e}_t = \arg\max_{e_t} P(e_t|z_{1:t}) \tag{2.20}$$

$$= \arg\max_{e_t} \int p(e_t, \boldsymbol{h}_t|z_{1:t}) d\boldsymbol{h}_t \tag{2.21}$$

$$= \arg\max_{e_t} \sum_l w_t^{(l)} \delta_{e_t}(e_t^{(l)}) \tag{2.22}$$

となる．ここで，$\delta_e(e')$ は，$e = e'$ のとき 1，そうでなければ 0 を返す関数である．

頭部姿勢の最尤推定値の獲得

パーティクルフィルタは，多数の仮説を用いるため追跡の失敗等に頑健である反面，目的関数のピークを得るためには大量の仮説を要する，つまり，計算コストがかかるという欠点をもつ．そこで，本論文では，6 自由度の連続量で扱われる頭部姿勢については，パーティクルフィルタにより得られる各時刻 t の推定値 $\widetilde{\boldsymbol{h}}_t$ からより精度の高い推定値 $\widehat{\boldsymbol{h}}_t$ を得るために，$\widetilde{\boldsymbol{h}}_t$ を初期値の 1 つとした勾配法による最尤推定を行う．なお，表情に関しては，パーティクルフィルタで得られる推定値 $\widetilde{e}_t$ を最終的な推定値 $\widehat{e}_t$ とする．これは，表情状態を離散状態として扱っていることによる．さらにまた，表情変化は個人特化の表情輝度分布モデル $\mathcal{I}$ にて表現されるため，汎用的な形状モデル $\mathcal{S}$ にて表現される頭部姿勢に比べて高精度な推定値を得やすいことにもよる．以上を，数式にて表すと，

$$\widehat{\boldsymbol{h}}_t = \arg\max_{\boldsymbol{h}} L(\boldsymbol{h}|\widetilde{e}_t, \boldsymbol{z}_t) \tag{2.23}$$

となる．

(2.23) 式の頭部姿勢の最尤推定値の定義式は，(2.6) 式を代入することで，次のように簡略化される．

$$\widehat{\boldsymbol{h}}_t = \arg\min_{\boldsymbol{h}} \sum_i f_i\left(\boldsymbol{h}, \widetilde{e}_t, \boldsymbol{z}_t\right). \tag{2.24}$$

for t=1:T

(1) パーティクルフィルタによる同時事後確率密度分布 $p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t})$ の算出

・仮説の生成 $\cdots$ (2.15) 式

$\{\boldsymbol{h}_t^{(l)}, e_t^{(l)}\} \sim p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t-1})$

・輝度補正 $\cdots$ (2.12) 式

$\tilde{z}_i = \gamma_b \cdot z_i$

・仮説の重みの計算（仮説の評価）$\cdots$ (2.18) 式

$w_t^{(l)} \propto p(\boldsymbol{z}_t|\boldsymbol{h}_t^{(l)}, e_t^{(l)})$

・期待値 $\tilde{\boldsymbol{h}}_t, \tilde{e}_t$ の算出

(2) 勾配法による最尤推定値 $p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_t)$ の算出

while $|\widehat{\boldsymbol{h}}^{(m)} - \widehat{\boldsymbol{h}}^{(m-1)}| >$ threshold

・輝度補正 $\cdots$ (2.12) 式

$\tilde{z}_i = \gamma_b \cdot z_i$

・推定値の更新 $\cdots$(2.25) 式

$$\widehat{\boldsymbol{h}}^{(m)} = \widehat{\boldsymbol{h}}^{(m-1)} - \alpha \cdot \nabla \textstyle\sum_i f_i \left(\widehat{\boldsymbol{h}}^{(m-1)}, \tilde{e}, \boldsymbol{z} \right)$$

end

end

図 2.10: 提案手法の推定アルゴリズムのフロー

本論文では，これを勾配法を用いて解く．

この勾配法の初期値には，次の $\widehat{\boldsymbol{h}}_1^{(0)}$ および $\widehat{\boldsymbol{h}}_2^{(0)}$ の 2 つを用いる．一つは，現時刻におけるパーティクルフィルタの推定値，すなわち，$\widehat{\boldsymbol{h}}_1^{(0)} = \tilde{\boldsymbol{h}}_t$ である．もう一つは，一時刻前で得られた最尤推定値 $\widehat{\boldsymbol{h}}_{t-1}$ から (2.14) 式の運動モデルを用いて予測される現時刻の頭部姿勢の分布の期待値，すなわち，$\widehat{\boldsymbol{h}}_2^{(0)} = E_{p(\widehat{\boldsymbol{h}}_t|\widehat{\boldsymbol{h}}_{t-1})}[\widehat{\boldsymbol{h}}_t] = \widehat{\boldsymbol{h}}_{t-1}$ である．これらの 2 つの初期値は，現時刻での頭部姿勢が一時刻前から大きく変化している場合，および，ほとんど変化していない場合にそれぞれ有用である．

推定値の更新については，それぞれの初期値について独立に次式に従い行う．

$$\widehat{\boldsymbol{h}}^{(m)} = \widehat{\boldsymbol{h}}^{(m-1)} - \alpha \cdot \nabla \sum_i f_i \left(\widehat{\boldsymbol{h}}^{(m-1)}, \tilde{e}, \boldsymbol{z} \right) \tag{2.25}$$

ここで，m は反復ステップ数，$\alpha(> 0)$ は学習率である．勾配ベクトル $\nabla \sum_i f_i$ の j 番目の要素は，$\partial / \partial h_j (\sum_i f_i) = \sum_i \partial f_i / \partial h_j$ と変換される．この $\partial f_i / \partial h_j$ は，(2.8) 式

および (2.11) 式より，

$$\frac{\partial f_i}{\partial h_j} = \frac{\partial f_i}{\partial d_i}\frac{\partial d_i}{\partial z_i}\left(\frac{\partial z_i}{\partial X_i}\frac{\partial X_i}{\partial h_j} + \frac{\partial z_i}{\partial Y_i}\frac{\partial Y_i}{\partial h_j}\right) \tag{2.26}$$

と展開される．ここで $\partial z_i/\partial X_i$ および $\partial z_i/\partial Y_i$ は，座標 $\boldsymbol{q}_i(\boldsymbol{h}) = [X_i\ Y_i]^{\mathrm{T}}$ における画像 $\boldsymbol{z}_t$ の輝度勾配である．

各時刻 t において，2 つの初期値 $\widehat{\boldsymbol{h}}_1^{(0)}, \widehat{\boldsymbol{h}}_2^{(0)}$ を (2.25) 式に従いそれぞれ更新して最終的に得えられた $\widehat{\boldsymbol{h}}_{1,t}$ および $\widehat{\boldsymbol{h}}_{2,t}$ のうち，より尤度の高い頭部姿勢を時刻 t における最終的な頭部姿勢の推定値 $\widehat{\boldsymbol{h}}_t$ とする．

$$\widehat{\boldsymbol{h}}_t = \max\left\{\widehat{\boldsymbol{h}}_{1,t}, \widehat{\boldsymbol{h}}_{2,t}\right\}. \tag{2.27}$$

最後に，以上の頭部姿勢および表情の推定フローを図 2.10 にまとめる．

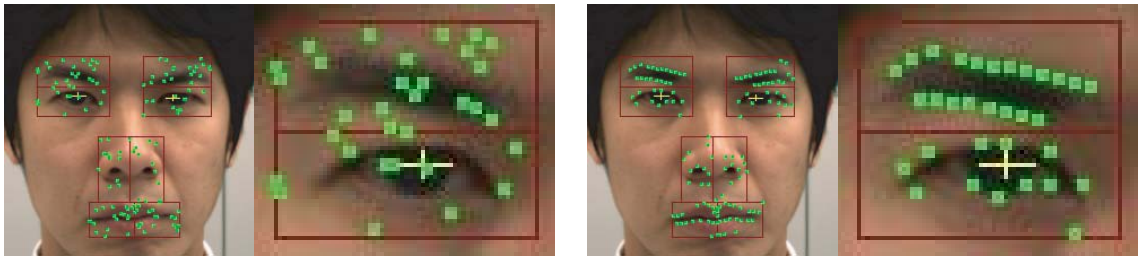
2.4　評価実験

提案手法の有効性を確認するため，本論文では 3 種類の検証実験を行った．1 つめは，被験者が様々な方向に顔を向けた状態において表情をどの程度正しく認識できるかの検証である．2 つめは，表情輝度テンプレートを個人特化でなく汎用的に使用した場合の性能検証である．3 つめは，提案手法がどれだけ頭部姿勢を安定に推定できるかの検証である[12]

またそれと同時に，本論文で用いるダイポール注目点（2.3.2 節）の有効性を評価するために，その他 2 種類の注目点を用いた変動輝度テンプレートを準備した．一つは，注目点をランダムに配置したもの（これをランダム注目点と呼ぶ）であり，もう一つは，注目点を顔内のエッジ上に配置したもの（これをエッジ注目点と呼ぶ）である．どちらの注目点についても，図 2.11 に示す 4 つの顔部品領域内に配置した．

以下の実験では，あらかじめ撮影した動画像に対して本手法をオフラインにて適用し，各フレームに対して算出された推定表情をそのときの表情の正解ラベルと照らし合わせた際の正解率を表情認識率とした．また，いずれの評価においても，頭部姿勢については 6 変数全て（2.3.1 節参照）を推定した．また，パーティクルフィ

[12]本論文での評価実験では，いずれも，本手法における最大の貢献である変動輝度テンプレートの有用性に主眼を置く．このため，2.3.6 節にて述べた推定アルゴリズムについては，表 2.3 での勾配法導入の効果についての検証のみ勾配法を用い，それ以外の検証については最終的な推定値をパーティクルフィルタにて得られる推定値とした．



顔部品領域内にランダムに配置した注目点（左）と，エッジ上に配置した注目点（右）．
これらの注目点は，図 2.4 で示した提案手法で用いるダイポール注目点に対する比較対象として扱う．

図 2.11: ダイポール注目点以外の 2 種類の注目点の例．

ルタの仮説の数を $1,500$ とした．このときの推定処理に要した時間は，Intel Core 2 Extreme 3.00GHz プロセッサ，2.0GB メモリの PC にて，およそ 50[msec/frame] であった．なお，個人認証については，本研究の主眼ではないため手動にて行った．

2.4.1　非正面顔に対する性能評価

真の顔形状を用いない限り，形状の誤差による注目点のアラインメント誤差が生じてしまう．注目点は正面顔画像にて定義されているため，このアラインメント誤差は頭部の面外回転角が大きくなるに従い増加する．よって，システムの頭部姿勢に対する頑健性の評価は重要である．しかしながら，著者の知る限り，表情および頭部姿勢を同時に変化させる表情データベースは公開されていない[13]．よって，本論文では，提案手法を評価するために独自に新たなデータセットを作成した．

ここでは，その独自の表情データセットを用い，提案手法に対する 2 種類の性能評価を行った．1 つめの評価の目的は 3 つある．まず，様々な首振り角（首を左右に振った際の水平方向の頭部姿勢角）に対して，提案手法によりどの程度の表情認識率が得られるかを検証するためのものである．次いで，複数の人物に対して提案手法が有効であるかどうかの検証である．最後は，提案手法で用いるダイポール注目点の有効性についての検証である．一方，2 つめの評価は，頭部姿勢（水平方向）および表情が同時に変化する状況において，表情を正しく推定可能かどうかを検証するためのものである．

[13]．既に 2.2 節にて述べたとおり，本論文の発表の直前に，そのようなデータベース（BU 4D-FE DB [97]）が公開された．

オリジナル表情データセット

本研究用に作成した独自の表情データセットには，2 種類の動画像が含まれている．それは，頭部姿勢固定データセット，および，頭部姿勢変動データセットである．両データセットに含まれる表情のカテゴリは，無表情，および，意図的に表出した怒り，悲しみ，驚き，喜びの計 5 種類である[14]．そして，両データセットは，各被験者について，学習用静止画像およびテスト用動画像から構成されている．いずれも，学習用静止画像はテスト用動画像の撮影の直前に撮影されたものである．これらの画像は，いずれも，PointGrey Research 社製の IEEE 1394 カメラを用いて，15 fps で撮影された XGA サイズ (1024 × 768 ピクセル) のカラー動画像である．本論文では，これらの動画像を，512 × 384 ピクセルのグレイスケール動画像に変換したものを使用した．なお，撮影中の照明は，天井に配置された蛍光灯および窓から拡散的に入射する太陽光のみであり，照明変動については無視できる程度であったと考えられる．

頭部姿勢固定データセットでは，被験者がまず水平，垂直，あるいは面内のいずれかの方向に頭部を回転させ，その方向を向いたままの状態で表情を変化させている．頭部姿勢の範囲は，水平方向には $0/\pm 20/\pm 40$ 度，垂直方向には $0/\pm 20$ 度，そして，面内方向には $0/\pm 20/\pm 40$ 度である (図 2.12-2.14 参照)．水平，垂直，および，面内それぞれの被験者数は，それぞれ，9 名（20〜40 代の男性 7 名および女性 2 名），4 名（水平方向の男性被験者のうちの 4 名），および，1 名（垂直方向の被験者のうちの 1 名）である．それぞれの方向において，各被験者につき 1 つの動画像がある．つまり，動画像数の合計は，$9\times 5+4\times 3+1\times 5=62$ である．各方向についての動画像の詳細は以下のとおりである．被験者は，まず最初に無表情でカメラに正対しており，次いで，無表情状態のまま対象方向に顔を向ける．その後，その姿勢を保ったまま，PC のモニタ上に自動的に映し出されるカテゴリ名の表情を順に表出する．指示された表情の順番は，無表情，怒り，悲しみ，驚き，喜びの順であり，各表情についての指示提示時間，および，指示間の間隔は，共に 4[sec](60[frame]) であった．なお，この指示された表情が，表情の正解ラベルとして保存されている．

一方，頭部姿勢変動データセットは，被験者が頭部姿勢を上下左右に自由に動かす間に表情も変化させたときのデータであり，被験者 1 名（頭部姿勢固定データセッ

[14]FACS [22] で定義される 6 基本表情に含まれる恐れおよび嫌悪表情については，表出の困難さや表出頻度を考慮して除外した

表 2.1: 頭部姿勢固定データセットに対する各頭部姿勢方向での表情認識率.

Point type	Total	Angle [deg]				
		-40	-20	0	20	40
Yaw (Horizontal): nine sequences.						
Pair	92.3	83.3	94.3	95.4	95.9	92.5
Random	90.1	90.2	92.1	95.4	96.1	76.7
Pitch (Vertical): four sequences.						
Pair	94.0	N/A	87.0	97.5	97.6	N/A
Random	86.4	N/A	74.1	98.1	87.0	N/A
Roll (In-plane): one sequence.						
Pair	100.0	100.0	100.0	100.0	100.0	100.0
Random	100.0	100.0	100.0	100.0	100.0	100.0

Pair: ダイポール注目点， Random: ランダム注目点.

トの面内方向の男性被験者）分のデータを含む．表情の正解ラベルについては，被験者自身が撮影後に動画像を見ながら付与した.

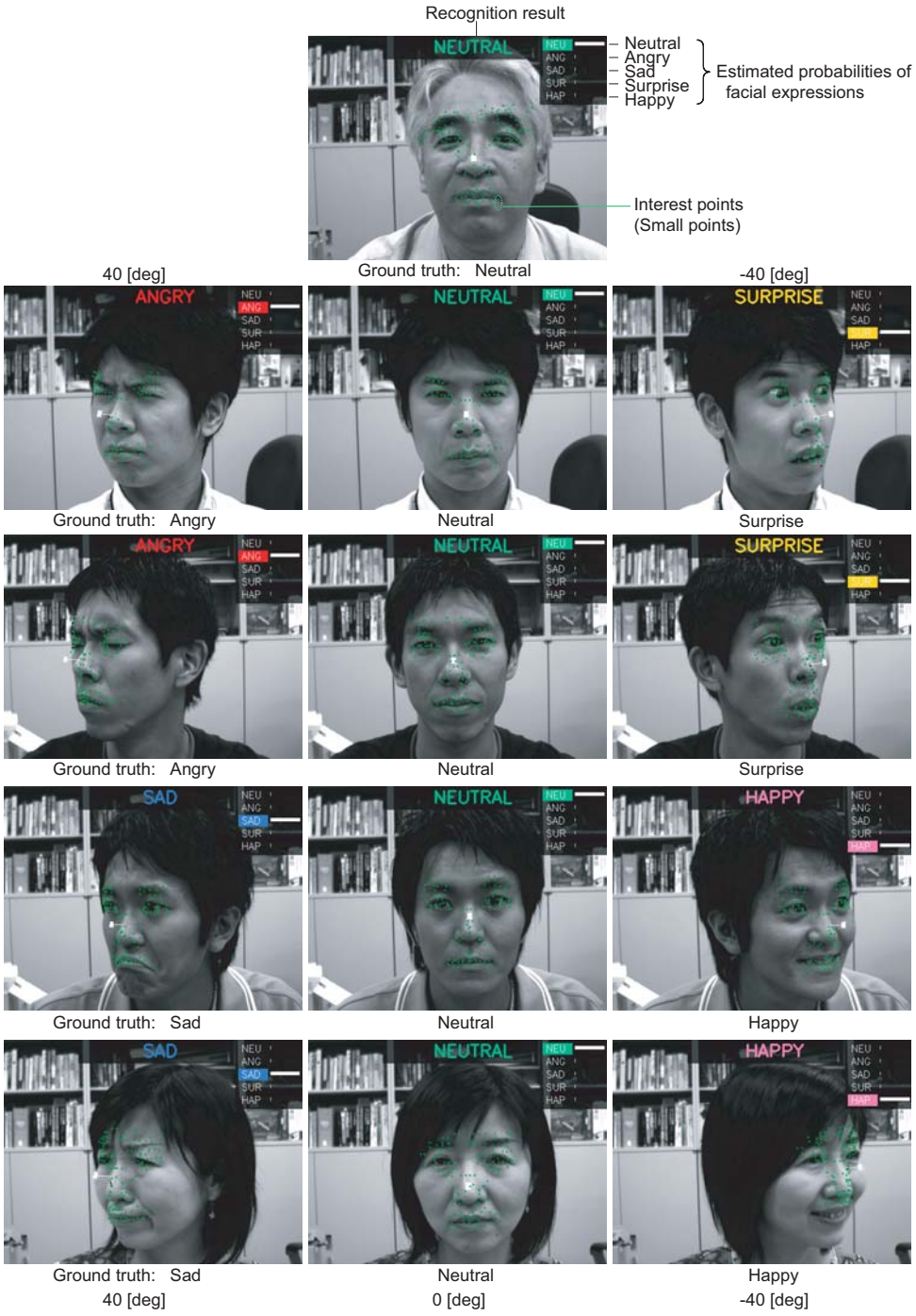
また，両データセットに付与されている学習画像については，各被験者に対して以下の指示を行い作成した．(1) 撮影中常にカメラに対して正対し，(2) PC のモニタ上に自動的に映し出されるカテゴリ名の表情を順に表出する．(3) 各表情を表出できた時点でキーボードを押下する．この瞬間の画像を学習画像として保存した.

頭部姿勢固定データセットを用いた評価

まず，頭部姿勢固定データセットを用いて，様々な面外方向の頭部姿勢に対する提案手法の表情認識性能について検証する．なお，このとき，実験データ作成時での，表情指示開始と実際に表出を行うまでの時間差を考慮し，表情が指示された直後から 20 フレーム分の画像を認識対象から除外した.

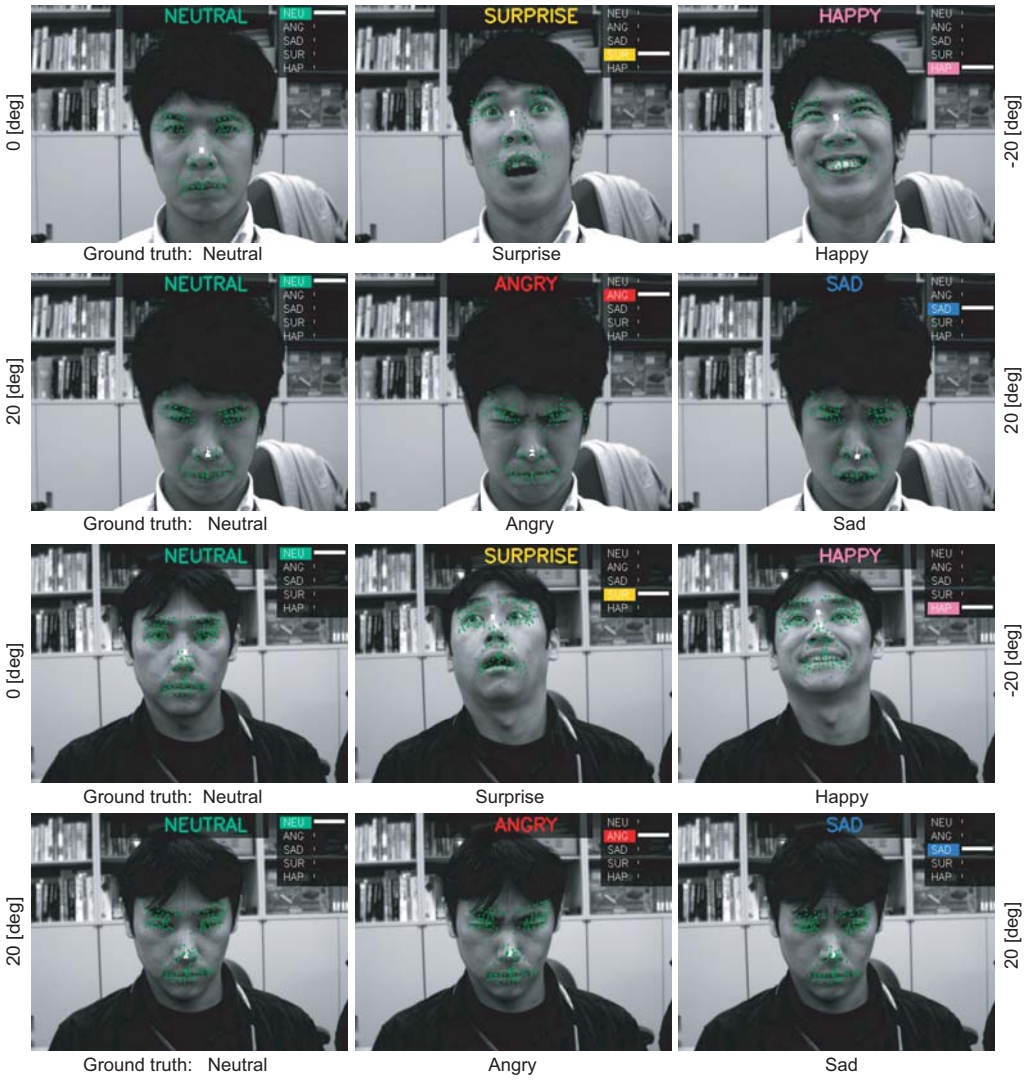
図 2.12，図 2.13，および，図 2.14 に，それぞれ水平，垂直，および，面内の頭部回転のデータセットについてのダイポール注目点を用いた場合の表情および頭部姿勢の推定結果を示す．いずれの被験者に対しても表情および頭部姿勢が正しく推定されている.

表 2.1 に，各頭部姿勢角における，ダイポール注目点とランダム注目点を用いた場合の表情認識率を示す．面外回転については頭部姿勢角の増大に従い認識率が低



各画像の上部に表情の認識結果を示す．また，画像右上部の各バーは，算出した各表情の事後確率 $P(e_t|z_{1:t})$ を表す．顔面上の小さな点群は注目点を表す．

図 2.12: 頭部姿勢固定データセット（水平方向）に対する頭部姿勢および表情の推定結果の例.



各画像の上部に表情の認識結果を示す．また，画像右上部の各バーは，算出した各表情の事後確率 $P(e_t|z_{1:t})$ を表す．顔面上の小さな点群は注目点を表す．

図 2.13: 頭部姿勢固定データセット（垂直方向）に対する頭部姿勢および表情の推定結果の例.

下している．これについては，注目点のアラインメント誤差の増大が主要な原因であると考える．これは正確な顔形状を用いない限り避けることができないが，ダイポール注目点の方がより認識率が高いため，より形状誤差に頑健であることが示唆される．面内方向については，注目点のアラインメント誤差が生じないため，面外方向よりも認識が容易である．よって，1 人の被験者についてのみ実験を行ったが，全ての対象表情を正しく認識できている．ダイポール注目点を用いた場合の全ての

各画像の上部に表情の認識結果を示す.

図 2.14: 頭部姿勢固定データセット（面内方向）に対する頭部姿勢および表情の推定結果の例.

方向を含む全体の認識率は 90%を超えている.

表 2.2 に，ダイポール注目点を用いた場合の表情の混同行列を示す．多くの被験者に対して，悲しみ表情と，無表情および怒り表情の間の誤認識率が高いのは，これらの表情が他の表情対に比べてより類似していたことを示唆する.

なお，エッジ注目点を注目点として用いた場合では，多くの動画像に対して頭部追跡に失敗した．これについては，エッジ注目点が注目点のアラインメント誤差に対して非常に敏感であるためだと考える．また，それ以外にも，学習されている表情から少しでも表情を変えると注目点の輝度が大きく変化するため，推定が不安定となるという現象が確認された．よって，エッジ注目点は提案手法の枠組みに適さないと考える.

さらに，パーティクルフィルタに勾配法を組み合わせる効果について検証する．頭部姿勢固定データセット中の正面方向の動画像に対して，提案手法（パーティクルフィルタと勾配法の組み合わせ）と勾配法を適用しない場合についての，頭部姿勢推定結果の標準偏差の平均値を表 2.3 に示す．これらの動画像には頭部の若干の揺れが含まれるものの，頭部姿勢はほとんど変化していないため，標準偏差が小さいほど安定した推定が行えていると考える．このとき，勾配法を組み合わせた結果の方が標準偏差が最も小さいため，より安定した推定が行えていると言える．この理由については，次のように考える．パーティクルフィルタのみを用いた場合，確率

表 2.2: 頭部姿勢固定データセットのうちの水平方向および垂直方向の頭部姿勢に対する表情の混同行列（平均）[%].

GT \ RCG	Neutral	Angry	Sad	Surprise	Happy
Neutral	88.9	1.5	7.9	1.6	0.1
Angry	0.4	97.6	0.9	1.1	0.0
Sad	3.2	8.0	85.4	3.3	0.1
Surprise	0.0	0.0	4.2	95.8	0.0
Happy	0.2	1.1	0.0	1.1	97.7

GT: 正解ラベル, RCG: 認識結果．全体の認識率は 93.1[%] である．

表 2.3: 頭部姿勢固定データセット（正面方向）での頭部姿勢の推定結果の標準偏差

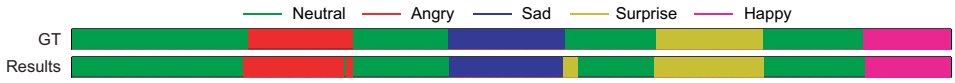
Methods	σ_x	σ_y	σ_{θ_x}	σ_{θ_y}	σ_{θ_z}	σ_s
	[pixel]		[degree]			($\times 10^{-2}$)
Only particle filter	8.7	8.2	4.1	4.5	0.8	1.9
Particle filter + Gradient method	**6.1**	**6.0**	**3.0**	**3.2**	**0.6**	**1.7**

$\sigma_x, \sigma_y, \sigma_{\theta_x}, \sigma_{\theta_y}, \sigma_{\theta_z}, \sigma_s$ は，それぞれ，画像の水平／垂直位置，垂直／水平／面内回転角，および，スケールについての推定結果の標準偏差を表わす.

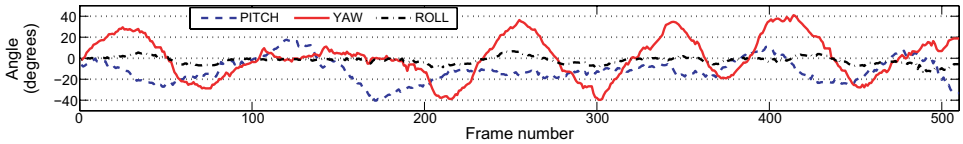
的にのみ生成される仮説数が十分でないことによる推定値の揺らぎが生じる．しかし，提案手法では，勾配法を用いることにより最尤値（少なくともその近傍の局所解）に到達できるため，その推定値の揺らぎを除去できたと考える．

本論文での実験における被験者数 9 名は比較的少数であるものの，提案手法では変動輝度テンプレートを利用の場で各被験者に特化して作成するため，他の多くの被験者に対しても有効であることが予想される．また，経験的に設定したマージンサイズ（2.7.2 節）に関しては，表 2.1 のように 9 名の被験者に対して高い表情認識率が得られていることから妥当であったと考える（定量的な議論は 2.7.3 節にて行う）．また，ここでは，近似的な顔の形状モデルとして，平均顔形状に簡単なスケーリングを施したものを使用した．その結果として，様々な頭部姿勢において高い表情認識率が得られたため，このような近似的な形状モデルでも提案手法の枠組みでは高い表情認識率が得られることが示唆された．もちろん，この平均顔形状の各人物へのフィッティング方法を改良して形状モデルの誤差を低減すれば，注目点のアラインメント誤差が抑えられるため，より頭部姿勢に対して頑健な推定が行えるこ

(a) テスト動画像のキーフレーム
（左から順にフレーム番号 1, 80, 130, 180, 200, 250, 373, 400, 450, 510）



(b) 表情についての正解ラベル（上段）および認識結果 $\widehat{e}_t$（下段）



(c) 頭部姿勢の推定結果 $\widehat{\boldsymbol{h}}_t$（横軸は (b) のものに一致）

図 2.15: 頭部姿勢変動データセットに対する推定結果

とが予想される．

頭部姿勢変動データセットを用いた評価

次いで，被験者が表情および頭部姿勢を同時に変化する頭部姿勢変動データセットを用いた評価を行った．時系列の推定結果を図 2.15 に示す．図 2.15(b) から，表情がほぼ全てのフレームで正しく推定できていることを見てとれる．さらに，正解表情が他の表情よりも突出して高い表情確率を持つことからも提案手法の頑健性が示唆される．頭部姿勢の推定精度に関する定量的な評価については 2.4.3 節にて行う．

この評価結果により，頭部姿勢および表情が同時に変化する場合においても，本手法がそれらを正しく推定可能であることが，1 名についてのみであるが確認された．他の被験者に対しても同様に正しい推定が可能であると予想する．それは，頭部姿勢固定データセットを用いた評価において，本手法が複数に被験者に対しても有用であるという結果が得られているためである．以上から，複数の人物に対して，単眼動画像を用いて，その場で各人物に特化したモデルを作成し，頭部方向変化が生じる状況においても表情および頭部姿勢を頑健に推定できることが示唆された．

2.4.2　汎用モデルについての検証

提案手法の最大の利点は，表情輝度テンプレートが各対象人物に特化した顔モデルとして容易に作成できる点である．2.4.1 節にて述べたとおり，個人特化のモデルを用いることで高い精度で表情を認識できる．この個人特化モデルを用いる方法は，少数の特定ユーザのみが利用するアプリケーションに対して特に有効である．しかし，そのように各ユーザに対する学習を行うことなく，任意の人物に対して直ちに推定を開始可能なシステムが望まれることも多い．そこで，本節では，提案手法の汎用モデルへの拡張可能性について検証する．

検証する汎用システムを既存手法と比較するため，ここでは使用する表情データセットとして，幅広く用いられている Cohn-Kanade DFAT-504 データベース [44] を用いる．Cohn-Kanade データベースは 104 名分の動画像からなり，いずれの動画像も，無表情の状態から始まり，ある 1 つの表情を表出し終えた時点で終わる．平均の長さは 8 フレーム程度である．被験者は，どのような表情を表出すべきかを撮影前に詳しく指示されている．

Cohn-Kanade データベースの各動画像には FACS [22] に基づく表情の Action Unit（視覚的に認識可能な表情を構成する基本単位）のラベルが付けられているものの，怒りや喜びといった表情カテゴリは付与されていない．そこで，本論文では，FACS で定められたルールに従い [23]，各動画像の最終フレームに 6 基本表情（怒り，悲しみ，驚き，喜び，恐れ，嫌悪）[21] のラベルを手動で付けた．このとき，約 1/3 の動画像に対しては基本表情のラベルを割り当てることができなかったため，本研究での評価対象から除外した．次いで，残りの 1/2 程度の動画像を，2.7.1 節で仮定した学習画像中では頭部姿勢が固定されているという条件を満たさないという理由で除外した．以上の過程で最終的に残ったのは，53 名分の 129 つの動画像（怒り，悲しみ，驚き，喜び，恐れ，嫌悪表情について，それぞれ，13, 20, 14, 43, 12, 27 つ）である．また，129 の全ての動画像の初期フレームに対しては，無表情のラベルを割り当てた．以上の処理でラベルづけられた全てのフレーム（選択された動画像の初期および最終フレーム）を学習画像として用いた．つまり，この汎用モデルに対する評価では，無表情に 6 基本表情を加えた 7 表情が認識対象である．

ここでは，平均の表情認識率を以下の方法で算出した．まず，各動画像に対して提案手法を用いた表情および頭部姿勢の推定を行った．次いで，各動画像の初期お

各平均顔は，眉，目，鼻および口それぞれの平均顔部品画像からなる．

図 2.16: Cohn-Kanade DFAT-504 データベースより学習した各表情についての平均顔画像 $\bar{\boldsymbol{g}}_e$.

よび最終フレームでの表情の認識結果を取り出して集めた[15]．最後に，7 種類の各表情カテゴリについての認識率をそれぞれ算出した．

学習段階

ここでは，学習画像が与えられた下で，汎用的な変動輝度テンプレートを完全に自動で学習する枠組みを提案する．なお，学習画像については，2.7.1 節で述べた方法により検出された顔領域のみを切り取り，それを両瞳孔中心が水平になるよう既に回転されているものとする．以下では，被験者 j の表情 e での学習画像を $\boldsymbol{g}_{j,e}$ と表し，また特に，それが無表情の場合には $\boldsymbol{g}_{j,\mathrm{NEU}}$ とも表すこととする．

まず，無表情の学習画像に対して，2.7.1 節で述べた方法により顔部品を検出する．無表情以外の学習画像については，それと対の無表情の学習画像における顔部品の検出結果をそのまま適用する．

次に，全ての学習画像を照明環境や肌色などの変動を除去するための正規化を行う．ここではその正規化を，$\tilde{\boldsymbol{g}}_{j,e} = (\boldsymbol{g}_{j,e} - \alpha^{(j)}\mathbf{1})/\beta^{(j)}$ として行う．ここでの被験者 j に対する正規化係数 $\alpha^{(j)}$ および $\beta^{(j)}$ については，無表情の学習画像 $\boldsymbol{g}_{j,\mathrm{NEU}}$ での，目および口領域に外接する矩形領域内の輝度の平均および分散とする．

続いて，各表情 e の動画像間（被験者間）の平均画像 $\bar{\boldsymbol{g}}_e$ を，正規化画像 $\tilde{\boldsymbol{g}}_{j,e}$ を全ての被験者 j について平均したものとする．この平均化については顔部品 p 毎に行う．この顔部品 p の平均画像を，以下では平均顔部品画像と呼び，$\bar{\boldsymbol{g}}_{e,p}$ と表す．この平均顔部品画像 $\bar{\boldsymbol{g}}_{e,p}$ の一例を図 2.16 に示す．

最後に，無表情の平均顔部品画像 $\bar{\boldsymbol{g}}_{e,p}$ において，2.3.2 節にて説明した方法で注目点集合を選択する．選択された注目点集合の一例を図 2.17 の左に示す．

[15] このとき，無表情以外のテストデータ数はそれぞれのラベルの付いた動画像の数，無表情のテストデータ数は選択した動画像の総和となる．

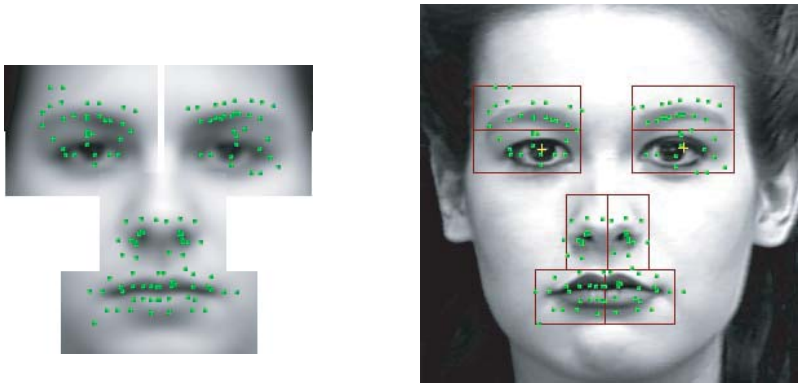


図 2.17: （左）データベースより学習した平均顔部品画像 $\overline{\boldsymbol{g}}_{e,p}$ 上に配置した注目点集合 $\mathcal{P}$.（右）左の注目点 $\mathcal{P}$ を対象テスト動画像の初期フレームに投影した結果.

テスト段階

テスト段階では，各動画像に対して，平均顔形状 $\overline{\boldsymbol{g}}_e$ とそれに対して定義された注目点集合から変動輝度テンプレートを作成する.

まず，テスト動画像の初期フレームに対して，2.7.1 節にて説明した方法により，回転補正を行った後に顔部品領域を検出する．本節では以後，この回転補正を行った画像を単に初期フレーム画像と呼ぶ.

次いで，初期フレーム画像における注目点 i の画像座標 $\boldsymbol{p}_i$ を，平均顔部品画像 $\overline{\boldsymbol{g}}_{e,p}$ 中に既に定義されている注目点集合に対する並進およびスケーリングを施すことで定義する．このマッピングは，平均顔部品画像の位置および大きさを，初期フレーム画像中の対応する顔部品領域のそれらに一致させ，そのとき同時に注目点集合も併せて位置および大きさを変化させることで得る．図 2.17 の右にマッピング結果を示す.

最後に，輝度分布モデルでの平均輝度 $\mu(e)$ を，平均顔部品画像を逆の正規化を行うことで得る．つまり，$\widehat{\boldsymbol{g}}_{e,p} = \beta^{(j)}\,\overline{\boldsymbol{g}}_{e,p} + \alpha^{(j)}\mathbf{1}$, である．ここで，$\widehat{\boldsymbol{g}}_{e,p}$ は逆の正規化を施された学習画像，$\alpha^{(j)}$ および $\beta^{(j)}$ については，学習段階で述べた方法において初期フレーム画像を $\boldsymbol{g}_{j,\mathrm{NEU}}$ として算出した．最後に，平均輝度 $\mu_i(e)$ を，復元された学習画像 $\widehat{\boldsymbol{g}}_{e,p}$ において画像座標 $\boldsymbol{p}_i$ での輝度とする.

注目点集合 $\mathcal{P}$ には図 2.17 右に示したものを用いている．

図 2.18: Cohn-Kanade DFAT-504 データベースに対する認識結果の一例．

表 2.4: Cohn-Kanade DFAT-504 データベースに対する汎用モデル（ダイポール注目点）を用いた提案手法による表情の混同行列 [%].

GT \ RCG	N	A	Sd	Sp	H	F	D
Neutral (N)	82.9	5.7	1.9	1.6	0.3	5.7	1.9
Angry (A)	23.1	30.7	7.7	7.7	7.7	7.7	15.4
Sad (Sd)	20.0	15.0	35.0	5.0	5.0	10.0	10.0
Surprise (Sp)	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Happy (H)	5.0	3.7	0.0	2.5	73.8	2.5	12.5
Fear (F)	16.7	0.0	8.3	8.3	33.4	25.0	8.3
Disgust (D)	3.8	19.2	3.8	0.0	0.0	1.9	71.3

トータルの認識率は 59.8[%] である．

推定結果

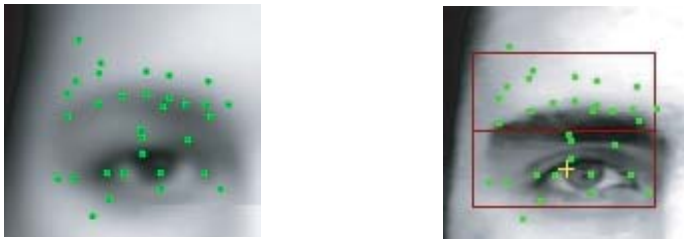
ここでは，leave-one-subject-out 交差検定法に基づきこの汎用モデルを評価する．すなわち，評価対象の被験者以外の全ての被験者の動画像を用いて平均顔画像を学習して汎用変動輝度テンプレートを作成し，それを用いて対象被験者の動画像中の表情の認識率を算出する．ダイポール注目点に加えて，ランダム注目点をそれぞれ用いた結果を表 2.4（併せて図 2.18 を参照）および表 2.5 に示す．全体の平均認識率は，ダイポール注目点で約 60%，ランダム注目点で約 70%である．どちらの注目点を用いた場合でも，恐れ表情の認識率が他の表情に比べて特に低くなっている．ちなみに，この恐れ表情を除外した平均認識率はおよそ 80%である．

この認識率は，90%以上の認識率を達成している個人特化の場合の提案手法（2.4.1

表 2.5: Cohn-Kanade DFAT-504 データベースに対する，汎用モデル（ランダム注目点）を用いた提案手法による表情の混同行列 [%].

GT \ RCG	N	A	Sd	Sp	H	F	D
Neutral (N)	82.1	4.1	4.6	3.8	0.0	1.6	3.8
Angry (A)	7.7	69.2	0.0	0.0	0.0	15.4	7.7
Sad (Sd)	15.0	5.0	70.0	5.0	0.0	5.0	0.0
Surprise (Sp)	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Happy (H)	0.0	3.7	0.0	2.5	83.8	5.0	5.0
Fear (F)	25.0	0.0	8.3	8.3	33.4	25.0	0.0
Disgust (D)	9.6	3.8	3.8	3.8	0.0	17.3	61.7

トータルの認識率は 70.2[%] である．



（左）平均顔画像上に定義したダイポール注目点.（右）それをテスト動画像の初期フレームにマッピングした結果．どちらも無表情の画像である．
特に，平均顔では眉上にある注目点が，初期フレームではそうなっていない．

図 2.19: 注目点のアラインメントの失敗例.

節），あるいは，文献 [46, 55, 77, 96, 101][16]といった他の最先端の手法には及ばない．だが，ここで提案した汎用モデルの作成方法はいささかシンプルなものであり，まだ十分に改良の余地が残されているため，ここでの結果は希望あるものであると考える．なお，そもそも本論文での対象は個人特化モデルであり，ここでの評価の目的は，提案手法の枠組みを汎用モデルに適用可能かどうかを検証すべきであった点に再度注意されたい．

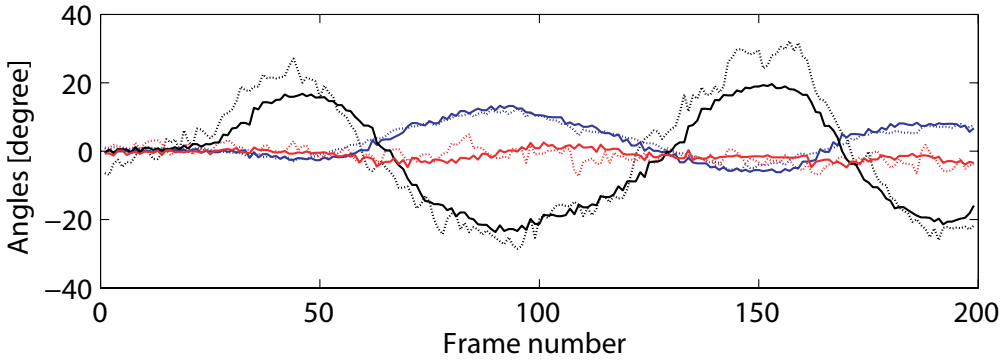
提案手法の枠組みにおける個人特化モデルとここで提案した汎用モデルの大きな違いは，注目点が適切な位置に配置されているかという点である．個人特化モデルでは，図 2.4 に示した通り，対象人物の顔画像に対して注目点を直接配置するため所望の注目点配置が実現される．他方，汎用モデルでは，注目点をデータベースか

[16]文献 [101] では無表情は認識対象に含まれていない．

ら学習した平均顔部品画像上に配置する．このため，正確な顔部品の検出，より正確に言うと，学習画像とテスト画像における顔の変形を含めた位置合わせ（例えば Active Appearance Models [83]）を行わなければ，図 2.19 に示すとおり，注目点の位置が学習画像とテスト画像で大きく異なってしまう．これでは，表情輝度分布モデルを正しく学習できないため，汎用モデルの認識率が個人特化モデルの認識率に比べて大きく低下していると考える．顔部品の位置合わせの精度を高めることで，汎用モデルの認識率を向上させることができるものと考える．また，その場合，2.4.1 節での結果同様，ランダム注目点よりもダイポール注目点の方がより高い認識性能を示すことが予想される．

注目点の配置ミスは，特に，見えの類似した表情間の誤認識を引き起こす．例えば，図 2.16 に示したとおり，怒り，悲しみ，および，嫌悪の表情は，特に眉領域で類似している．このため，本論文では矩形領域として顔部品を検出したが，これでは位置，大きさや形状などの個人差による大きな位置合わせ誤差が生じる．さらに，表 2.4 および表 2.5 における無表情の誤認識も顔部品の位置合わせ誤差に影響を受けていると考える．例えば，図 2.19 右の眉領域だけに着目した場合，怒りや恐れ表情に関係する眉の下降が生じているかのように見える．

恐れ表情の認識率が特に低い理由は，学習画像中の恐れ表情，特に眉領域，のばらつきが大きかったためであると考える．図 2.16 の眉領域は他の表情に比べて特にぼけを生じている．これは，眉毛と周囲の肌の輝度の差が実際の差に対してかなり小さいモデルが復元されていることを意味している．一方，口領域については，歯がよく見えているため比較的喜び表情に類似している．このため，恐れ表情の画像に対して変動輝度テンプレートを頭部姿勢を含めて当てはめる際，実際の頭部姿勢かつ恐れ表情という状態よりも，実際の頭部姿勢からスケールや位置が少し異なる喜び表情の方がより尤度が高くなったためであると考える．この問題については，顔部品の位置合わせ精度を高めるという改良以外にも，Action Unit [22] のようにモデルを顔の上下に分割することで対処可能であると考える．



破線: 提案手法による推定結果，実線：磁気センサによる計測結果.
黒：水平，赤：垂直，青：面内の回転を表す.

図 2.20: 頭部姿勢の推定結果.

2.4.3　頭部姿勢推定精度の検証

最後に，提案手法の頭部姿勢推定精度についての検証を表情認識を切り離して行う．ここでは，このための評価用データとして，Boston 大学の顔追跡データベース[17] [10] を用いる．このデータベースを以下 BU 顔追跡データベースと呼ぶ．BU 顔追跡データベースは 30[fps] の QVGA（320×240 ピクセル）サイズの MPEG フォーマットの動画像のセットである．頭部姿勢については並進および 3 次元の回転が含まれている．これらの頭部姿勢変動については磁気センサのデータが動画像と対で付与されている．照明については，時間的に一定，あるいは，時間変化するというそれぞれの条件で撮影された動画像が含まれるが，ここでは前者データのみを使用する．なお，いずれの動画像も表情変化は基本的には含まれていない．このため，ここでの評価では，表情輝度分布モデルを無表情についてのみ準備し，推定においても常に無表情であるとする．

提案手法による頭部姿勢の推定結果の一例を図 2.20 に示す．また，全動画像に対する平均絶対角度誤差を，最新のサーベイ論文 [64] にて紹介されている，この BU 顔追跡データベースを用いた評価を行っている 2 つの最先端の手法による推定結果と併せて表 2.6 に示す．提案手法は性能的に両者の中間に位置する．また，表 2.6 から，提案手法の推定精度は，水平方向よりも垂直方向の方がより高くなっている．

[17] *http://www.cs.bu.edu/groups/ivc/HeadTracking/*

図 2.21: BU 顔追跡データベースに対する推定結果の例

表 2.6: 提案手法と既存手法に対する平均絶対誤差の比較.

Methods	Errors [deg]		
	yaw	pitch	roll
Pair	7.1	4.2	2.9
Edge	9.5	6.6	5.5
Random	7.1	4.9	3.0
文献 [10]	3.3	6.1	9.8
文献 [95]	3.8	3.2	1.4

既存手法の結果については [64] より引用した.

表 2.6 では，注目点の種類については，ダイポール注目点とランダム注目点が類似した精度であり，エッジ注目点よりも精度が高い．エッジ注目点の精度が低いのは，形状モデルの誤差による注目点のアラインメント誤差，および，表情の若干の変化による影響を大きく受けた結果であると考える.

2.5　考察

本節では，以上の 3 種類の検証を踏まえて議論する．その 3 つとは，個人特化モデルを用いた非正面顔に対する表情認識，汎用モデルを用いた表情認識，および，頭部姿勢の推定精度であった.

本研究でのオリジナルのデータベースを用いた評価では，個人特化モデルの提案手法が幅広い範囲の方向の頭部姿勢変動に対して表情を頑健に認識できることが示唆された．これは，個人特化モデルでは，顔部品の形状や幾何配置や表情変化といった個人差の影響を受けないために，形状モデルの誤差に起因する多少の注目点のアラインメント誤差がある場合にも，表情を正しく推定できるためであると考える．

注目点の選択方法を改良することで，提案手法の性能を向上させることができるものと考える．1 つは，本論文では注目点を無表情の学習画像のみを用いて選択している．しかし，これを，表情変化による輝度変化の大きな注目点を選択するようにすることで，より正しく各表情を識別できると考える．2 つめは，本論文では非常に簡単に形状モデルをフィッティングしたが，これをより優れた方法に切り替えることが考えられる．

本論文での実験により，また，提案手法の汎用モデルへの拡張可能性が示唆された．本論文で提案した汎用モデルは非常にシンプルな方法で作成されるが，active contours [45] や輝度プロファイルベースのマッチング [39] などを参考にして，各注目点を 2 点を通る直線方向に移動させて正しくエッジをまたぐようにするなど，より優れた顔部品検出器を用いることが考えられる．

個人特化モデルおよび汎用モデル 2 つの評価結果から，汎用モデルも個人特化モデルと同様，大きな頭部姿勢変動に対処可能であると考える．3 次元の変形ベースの従来手法と比べ，提案手法は汎用の表情モデルを正面顔画像のみから学習可能であるため，複数の既存の表情の画像データベースを使用できる点で優れている．さらに，頭部姿勢に対する頑健性は形状モデルに大きく依存するが，オプティカルフロー推定などの密な特徴を利用するアプローチでは，形状モデルの精度を犠牲にして，計算コスト的に有利な平面などの単純な幾何形状が用いられることが多い．一方，提案手法では，計算コストを変えることなく，任意の形状モデルを使用することが可能である．よって，将来的に顔形状推定の枠組みを導入することも容易であるため，提案手法の方が拡張性に優れていると言える．

2.6　結論

本論文では，頭部姿勢と表情を同時に推定する手法として，変動輝度テンプレートに基づく手法を提案した．この変動輝度テンプレートは，従来法の煩雑なモデル構

築のための事前準備が不要であるため，その場で簡単に個人特化のモデルを作成可能である．提案手法では，この変動輝度テンプレートを用い，パーティクルフィルタと勾配法を組み合わせた方法により頑健かつ効率的な推定を行う．実験により，様々な方向に頭部を向けた場合でも頑健に表情を実時間で認識できることを確認した．

最後に，本手法の課題，および，それらに対する解決案について簡単に述べる．本論文では，注目点のマージンサイズの妥当性の検証において，最適なマージンサイズの一例を，算出位置のずれ量および表情変化に起因する顔部品の移動量から算出しているものの，それを実際に使用するマージンサイズの決定に導入していない．さらに，注目点の抽出に無表情の顔画像のみを用いているため，無表情以外の表情の認識に特に有用な点，すなわち，顔部品の移動量が大きな点を抽出できていない可能性がある．妥当なマージンサイズの決定と表情識別に有用な注目点の配置を同時に行う手法を構築することが，さらなる表情認識率の向上や頭部姿勢に対する頑健性の向上に有用であると考える．

また，本論文では明確な表情カテゴリを出力することを目指して意図的な表情を認識対象としたが，自発的に表出された微細な表情を認識対象とする必要もある．そのため，オンラインクラスタリングによる表情の自動学習や，オプティカルフロー推定を参考にした注目点の輝度変化からの顔部品の移動量の推定，すなわち，表情の表出強度の推定にも取り組みたい．

2.7　補足

2.7.1　学習画像および顔モデル作成のための前処理

学習画像 $\boldsymbol{g}$ については，以下の方法で撮影した．なお，ここでの学習画像とは，おそよ顔領域のみを含み，両目の中心が平行に位置するように前処理を施された画像とする．なお，これに対して，被験者の顔を含むもともとの画像（必ずしも被験者の両目は水平に位置していない画像）を学習元画像と呼ぶ．

対象人物がカメラに対して正面を向いた状態で頭部姿勢を変化させずに各表情を順に表出したときの顔の静止画像である（図 2.1 参照）．学習元画像は，頭部姿勢をカメラに対して正対させた状態で固定し，それぞれの対象表情を表出したときの各表情の画像とする．学習画像は，同一人物の異なる表情間ではアラインメント（位置合わせ＋スケーリング）が行われている必要がある．だが，個人特化モデルを作

成する場合には，異なる人物間の学習画像のアラインメントは行われている必要はない．

本論文では，学習元画像において，顔のアラインメントが行われていることを前提する．そのような学習元画像の獲得方法については，2.4.1 節にて述べる．回転補正については，まず，両眼の瞳孔中心を検出し，それが水平になるように画像面内の回転を行うこととする．この瞳孔中心の検出についてはどのような方法を用いてもよいが，本論文では Fast Radial Symmetry [56, 99] を用いることとする．学習画像の表情間の顔のアラインメントができていることを前提としているため，無表情の学習画像についてこの回転角を算出し，無表情以外の表情の回転にはこの回転角を用いることとする．

顔領域および顔部品領域の検出

2.3.2 節にて述べた注目点の選択の前に，無表情の学習元画像 $\boldsymbol{g}_{\mathrm{NEU}}$ において，顔領域，および，各顔部品（眉，目，鼻，および，口）を検出する．まず，顔領域については，Haar-like 特徴を用いたカスケード型 AdaBoost 識別機 [91] を用いておよその位置を検出する．次いで，顔部品領域については，図 2.4 に示すような矩形領域として検出する．

目および口領域については，以下の方法で検出する．まず，顔領域検出と同様，それぞれの領域ごとに用意した Haar-like 特徴を用いたカスケード型 AdaBoost 識別機 [91] を用いて顔部品の候補を検出する．各顔部品について候補領域が複数検出された場合には，その位置および大きさが最も尤もらしい候補，すなわち，事前に経験的に定めた尤度を最大化する候補領域を検出結果とする．本論文では，この尤度を，$L(\boldsymbol{y}) = \prod_k \mathcal{N}(y_k; \mu_k, \sigma_k)$ とする．ここで，$L(\boldsymbol{y})$ は，候補領域の特徴ベクトル $\boldsymbol{y}(= [X, Y, W, H]^{\mathrm{T}})$ の尤度である．$[X, Y]^{\mathrm{T}}$，W および H は，それぞれ，候補領域の顔領域内での相対的な中心位置，高さおよび幅である．μ_k および σ_k は，$\boldsymbol{y}$ の各要素 k についての経験的に決定した平均および標準偏差である．

鼻領域については，本論文では以下の方法で検出する．まず，鼻孔の位置を，上述した方法で検出した目と口の間に位置する，水平および垂直方向それぞれの輝度プロファイルの極小点とする．最後に，眉領域については，本論文では，単に目領域の上側に接し，目領域と同じ大きさの領域とする．

図 2.22: 学習画像に対する平均顔形状のフィッティング結果の一例

形状モデルのフィッティング

形状モデル $\mathcal{S}$ の各被験者に対するフィッティングについては以下の方法で行う．まず，形状モデルの正面方向（正面顔方向）を対象被験者の無表情の学習画像平面に対して直交させた状態で，学習画像 $\boldsymbol{g}_{\mathrm{NEU}}$ の顔領域の中心と，形状モデルの中心の位置とを合わせる．次いで，この両者の顔領域の幅，および，目中心ー口中心間の高さを，それぞれ，縦方向および横方向に線形に伸縮することで合わせる．最後に，縦方向および横方向のスケーリング係数の積の平方根をスケーリング係数として，奥行き方向に伸縮する．フィッティング結果の一例を図 2.22 に示す．

2.7.2　注目点の選択基準

本手法は，近似的な形状モデルを用いるとともに，注目点を正面顔画像上にて定義するため，頭部姿勢の増大に従って注目点の算出位置（算出方法の詳細は 2.3.1 節参照）がずれてしまう．このため，この算出位置のずれが生じても表情を正しく認識できる必要がある．以下では，注目点の輝度の変動要因として表情変化による顔部品の移動および注目点の算出位置のずれのみを考慮し[18]，ある 1 つの注目点の輝度に基づき無表情と他のある表情 $e(\neq 0)$ とを識別できるための条件について考える．

本手法では，注目点の輝度のモデルを事前に準備し，それを用いて入力画像における注目点の算出位置での輝度から表情を認識する．このとき，1 つの注目点の輝度から無表情と表情 e が正しく識別されるのは，

[18] 2.3.2 節にて述べたとおり，注目点の輝度に影響を与える要因にはこれら以外にも様々なものがある．これら全ての要因を含んだ表情認識に関する頑健性の検証については，実際の動画像に対して本手法を適用して得られる表情の認識結果から行う（2.4.1 節）．

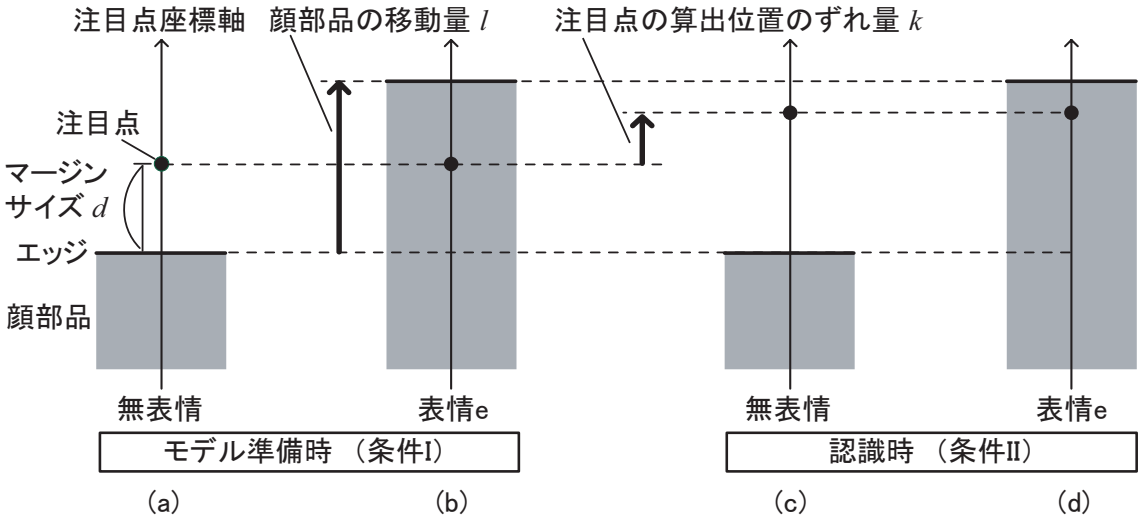


図 2.23: 無表情およびその他の表情 e の条件 I および条件 II における，注目点のマージンサイズ d，算出位置のずれ量 k および顔部品の移動量 l の関係

条件 I モデル準備時において獲得される輝度が，無表情および表情 e で大きく異なっている

条件 II 認識時の入力画像から観測される輝度が，そのときの表情についてのモデルの輝度に類似している

という 2 つの条件を共に満たす場合である．以下では，注目点の近傍において，顔面上での輝度が顔部品のエッジにおいて空間的にステップ状で変化すること，および，各顔部品が十分に大きいことを仮定する．このとき，条件 I および条件 II については，図 2.23(a) のように注目点を通りその近傍の顔部品のエッジに直交する軸（注目点座標軸と呼び，エッジから注目点への方向を正とする）を基に考えることができる．ここで，図 2.23(a) のように，注目点 i と顔部品のエッジとの間にマージン（このサイズを $d_i(> 0)$ にて表す）を設けることで，条件 I および条件 II を共に成立させることを考える．

まず，条件 I は，表情 e のモデル準備時において，顔部品が十分移動し注目点位置を超えている（図 2.23(b)），すなわち，

$$\textbf{条件 I} \qquad d_i \leq l_{i,e} \tag{2.28}$$

であるときに成立する．ここで，$l_{i,e}$ はこのときの顔部品の移動量の注目点座標軸成

分（正方向は注目点座標軸の正方向と一致）である．

次いで，条件 II については，無表情および表情 e のそれぞれに対して，注目点の算出位置のずれを考慮する必要がある．ここでは，注目点 i の算出位置のずれ量の注目点座標軸成分を k_i（正方向は注目点座標軸の正方向と一致）にて表わす．無表情については，注目点 i の算出位置がエッジの方向にずれた ($k_i < 0$) としても，そのずれ量の大きさがマージンサイズ d_i より小さければ，注目点における輝度がほとんど変化しないため条件 II が成立する（図 2.23(c)）．一方，表情 e については，条件 I を満たしているものとすると，認識時に注目点の算出位置がずれていたとしても，この顔部品のエッジがこの注目点の算出位置を超えていれば条件 II が成立する（図 2.23(d)）．以上をまとめると，条件 II は，

$$\textbf{条件 II} \qquad -k_i \leq d_i \leq l_{i,e} - k_i \tag{2.29}$$

と書き表わされる．

これらの条件 I および条件 II は，各注目点に対し，無表情以外の対象表情の数 ($N_e - 1$) だけ存在するが，それらの全てが満たされていなければならないというわけではない．この理由は以下のとおりである．表情には，変化する顔部品およびその変化パターンが表情間でそれぞれ異なるという性質がある．そのため，各注目点についてみてみると，顔部品が無表情時から動く表情もあれば，そうでない表情もある．このうち，顔部品が動いている場合についてのみ，その顔部品の移動を正しく検出可能なこと，すなわち，条件 I および条件 II を共に満たすことが要求される[19]．このような両表情を識別可能な注目点が注目点集合の中に 1 つ以上含まれていることが，無表情と表情 e を識別するための必要条件である．さらに，対象とする全ての表情を識別できるためには，それら全ての表情に対して以上のことを言える必要がある．

以上より，各注目点 i について，算出位置のずれ量 k_i および顔部品の移動量 $l_{i,e}$ が既知であれば，各表情 e に対して条件 I および条件 II がなるべく成立するようにマージンサイズ d_i を設定することが可能である（2.7.4 節にてその一例について述べる）．しかし，本論文の提案手法では，その利用場面において顔形状を精密に計測する装置を用いないことを想定しているため，事前に形状モデルの誤差から算出位置のず

[19]顔部品が動いていない場合，その注目点は頭部姿勢の推定に寄与し得る．本論文での提案手法では頭部姿勢および表情を同時に推定するため，正しい頭部姿勢の推定に寄与する点は，正しい表情識別にも間接的に寄与する．

れ量 k_i を計算し，そこからマージンサイズ d_i を決定するという方法論を取ることができない．そのため，本論文では，その代わりに，注目点の算出位置のずれ量および表情表出時の各顔部品の移動量を勘案して，経験的にマージンサイズを決定するという方法を選択する（具体的には 2.3.2 節を参照）．このように経験的に決定されたマージンサイズの妥当性については，2.7.3 節において，実際に装置を使用して顔形状を計測し，そこから算出される最適なマージンサイズとの比較により評価する．なお，今後，分散情報を含んだ平均顔形状を準備することで，算出位置のずれ量 k_i の期待値を算出し，そこから最適なマージンサイズ d_i を決定するという方法も考えられる．また，顔部品の移動量 $l_{i,e}$ については，オプティカルフロー推定 [57] や注目点座標軸上での輝度プロファイルのマッチング（文献 [16] が参考になる）などシンプルな従来手法により算出可能であろう．

2.7.3　注目点のマージンサイズの妥当性の検証

本節では，2.7.2 節にて述べた表情を正しく識別できるための条件 I および条件 II を基に，経験的に決定したマージンサイズの妥当性を検証する．注目点の算出位置のずれ量を求めるためには実際の顔形状が必要である．そこで，1 名の被験者について，3 次元デジタイザ (KONICA MINOLTA 社製 VIVID910) を用いて実際の顔形状を計測した．この計測顔形状を図 2.24 に示す．

そして，形状モデルとして円柱を用い，頭部姿勢角のうちの首振り角のみが ± 40[deg] の範囲で変動する場合を想定した[20]．注目点 i の算出位置のずれ量 k_i については，以下のようにして算出した．まず，(2.9) 式および (2.10) 式において，形状モデルとして実際に計測機を用いて計測した形状モデルおよび円柱モデルのそれぞれを用いて算出される注目点の画像座標（$\boldsymbol{q}_{i,t}$ および $\tilde{\boldsymbol{q}}_{i,t}$ とする）を算出した．そして，それらの差分ベクトル $(\tilde{\boldsymbol{q}}_{i,t} - \boldsymbol{q}_{i,t})$ を，注目点座標軸への射影したときの成分を算出位置のずれ量 k_i とした．一方，顔部品の移動量 $l_{i,e}$ については，学習画像から目視にて読み取った．このうち，本論文では議論を簡単にするため，各注目点に対し，顔部品の移動量が最大である表情についてのみを考慮した．以上のようにして得られた，算出位置のずれ量 k_i および顔部品の移動量の最大値 $l_{i,\max}$ を用い，各注目点 i において最適なマージンサイズ d_i^* を算出した（算出方法の詳細については 2.7.4 節参照）．表

[20]本論文では形状モデルとして平均顔形状を用いた議論を行ってきたが，ここでは，発表文献 [10] にて行った円柱を用いた場合についての結果を載せている．

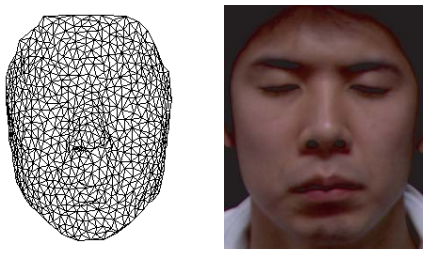


図 2.24: 計測顔形状（左）と計測時のテクスチャ画像（右）

表 2.7: 算出位置のずれ量の最大値, 顔部品の移動量の最大値, および，最適なマージンサイズの平均値

顔部品	$k_{i,\max}$	$l_{i,\max}$	d_i^*
非目領域	2.54	5.11	3.75
目領域	3.39	2.26	3.09

単位：[pixel]

2.7 に，首振り角の対象範囲内における算出位置のずれ量の最大値 $k_{i,\max}$，顔部品の移動量の最大値 $l_{i,\max}$，および，最適なマージンサイズ d_i^* についての全注目点での平均値を示す．経験的に決定したマージンサイズは非目領域および目領域についてそれぞれ 4 および 3[pixel] であったが，これらは表 2.7 に示す最適なマージンサイズ d_i^* の平均値（3.75 および 3.09[pixel]）とほぼ一致することが分かる．

さらに，最適なマージンサイズを用いたときに条件 I および条件 II を満たす注目点のうち，経験的に設定したマージンサイズを用いた場合においても両条件を満たしているものがどの程度の割合で存在するのかを算出した．結果は，非目領域で 92.3[%], 目領域で 100.0[%] と高い割合であった．なお，ここでも，最適なマージンサイズの算出時と同様，各注目点において移動量が最大の表情を対象とした．

以上の 2 種類の検証結果より，経験的に決定したマージンサイズが妥当であることが，一人の被験者についてのみではあるが定量的に明らかにされた．

2.7.4　最適なマージンサイズの算出方法の一例

本付録では，注目点 i において，無表情と表情 e を正しく識別するために最適なマージンサイズ d_i を，2.7.2 節にて述べた条件 I および条件 II を用いて算出する方

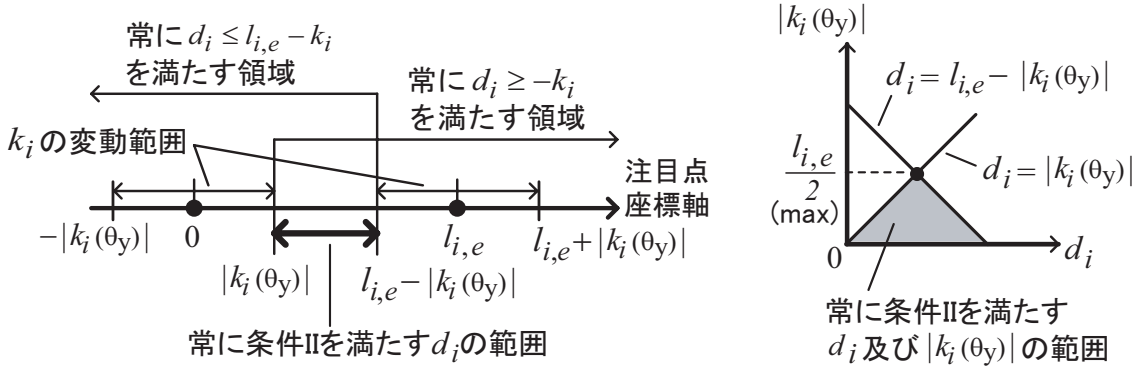


図 2.25: 条件 II を満たすマージンサイズの範囲.

法の一例について述べる．ここでは，頭部姿勢角のうち，首振り角 θ_y のみが正負対称の範囲で変化する場合について考える．そして，そのときに，なるべく広い範囲の首振り角に対して条件 I および条件 II を満たすようなマージンサイズを，最適なマージンサイズ d_i^* と定義する．

注目点の算出位置のずれ量 k_i は首振り角 θ_y の関数であるため，本付録ではこの関数を $k_i(\theta_y)$ と表す．ここで，首振り角 θ_y のみの変動を考えていることから，注目点の算出位置のずれる方向は x 軸（画像水平）方向のみである．よって，注目点座標軸が y 軸（画像鉛直）方向と一致していれば，任意の首振り角 θ_y に対して $k_i(\theta_y) = 0$ となる．一方，そうでなければ，$k_i(\theta_y)$ は正弦関数である．このとき，首振り角が $(-\theta_y \sim \theta_y)$ の範囲で変動すると，注目点の算出位置のずれ量の範囲は $(-|k_i(\theta_y)| \sim |k_i(\theta_y)|)$ となる．

まず，条件 I$(d_i \le l_{i,e})$ を満たし得る $l_{i,e} > 0$ の場合について考える．このとき，$(-\theta_y \sim \theta_y)$ 内の任意の首振り角に対して条件 II を満たすマージンサイズ d_i と，注目点の算出位置のずれ量 k_i および顔部品の移動量 $l_{i,e}$ との関係をグラフ化したものが図 2.25 左である．この図から，に条件 II を満たすマージンサイズの下限値 d_i^{lw} および上限値 d_i^{up} は，それぞれ，$d_i^{\mathrm{lw}} = |k_i(\theta_y)|$, $d_i^{\mathrm{up}} = l_{i,e} - |k_i(\theta_y)|$ であることが分かる．この場合での最適なマージンサイズは，下限値 d_i^{lw} および上限値 d_i^{up} の中点 $(= l_{i,e}/2)$ となる．なぜなら，このとき，条件 II を満たす $|k_i(\theta_y)|$ の値が最大となり（図 2.25 右参照），その結果，そのとき条件 II を満たす頭部姿勢角の範囲 $(-\theta_y \sim \theta_y)$ が最も広くなるためである．なお，条件 I$(d_i \le l_{i,e})$ については，条件 II の上限側の不等式 $(d_i \le l_{i,e} - |k_i(\theta_y)|)$ が成立していれば，明らかに成立する．

一方，条件 I を満たし得ない ($l_{i,e} \leq 0$) 場合には，単に無表情についての条件 II($-k_i \leq d_i$) のみを考慮する．そして，対象とする頭部姿勢内（最大値を $\theta_{y,\max}$ とする．2.7.3 節の議論では 40[deg] である.）において，無表情についての条件 II を常に満たすようなマージンサイズ ($d_i \geq |k_i(\theta_{y,\max})|$) のうちの最小値を最適なマージンサイズとする．

以上のような最適なマージンサイズ d_i^* を，数式にてまとめると，

$$d_i^* = \begin{cases} l_{i,e}/2, & \text{if } \ l_{i,e} > 0 \\ |k_i(\theta_{y,\max})| & \text{otherwise} \end{cases} \tag{2.30}$$

となる．

第3章　コンテクスト情報を用いた運転時の注視動作の識別

3.1　はじめに

近年 Intelligent Transportation Systems(ITS) の進展が目覚ましいが，より高度な ITS を実現するために，次世代の ITS にはドライバの意図や感情などを読み取り，適切なサポートをしてくれる機能が望まれる．それらの心的状態は直接計測することができないため，観測可能なドライバの行動から推定する必要がある．観測可能な情報としては，注視動作といったドライバの身体動作 [13, 67, 102]，ステアリング角やペダル踏み込み量などの運転操作 [48, 67, 75, 88] や，自車両のスピードや先行車両との間隔といった運転状況 [67, 75, 88] などが挙げられる．また，従来手法で推定対象となっている内部状態には，疲労度 [102]，警戒/注意度 [5] およびストレス度 [36] といったものが挙げられる．

一方で，近年の死亡事故件数の主要な原因の一つに前方不注意がある [113]．この前方不注意には，ミラーの死角を確認するという運転に関わる前方不注意（死角確認）と，景色に見とれるといった運転とは直接関係のない対象への注視による前方不注意（脇見）とがある．これらは身体動作的には類似しているが，背後にある意図の違いは重大である．つまり，前者が安全性を高めるために行われるのに対し，後者はむしろ事故リスクの増大を承知で行われる．よって，ドライバの心的状態を知るためにこれらの動作を区別することは重要である．しかし，従来，これらの 2 種類の前方不注意が区別されていなかった [50, 78].

以上の背景を踏まえ，本研究では前方注視，死角確認および脇見の 3 種類の注視動作を初めて識別可能な手法を提案する．提案手法では，これらの 3 種類の注視動作を識別するための観測として，前方不注意の主な特徴であるドライバの身体動作情報に加えて，コンテクスト情報として，前方不注意時のステアリング技術の低下 [9] といった注視動作に影響を受ける運転操作情報，および，それぞれの注視動作をど

の程度引き起こしやすいかに関係する運転状況という計 3 種類の情報を用いる．

これらの注視動作をモデル化には注視動作のダイナミクスが組み込まれた動的ベイジアンネットワークを用いる．この推定の枠組みでは，観測が与えられたもとで，それぞれの注視動作の事後確率が逐次的に推定される．なお，このような統計モデルは，一般に十分な数の学習サンプルを要する．とはいえ，実環境での故意的な脇見は安全上困難であるため，本研究では提案手法の評価においてドライビングシミュレータを用いることとする．

以下，3.2 節では関連研究を取り上げ，次いで 3.3 節では提案手法について説明する．続いて 3.4 節で実験結果について述べる．そして，3.5 節にて議論を行い，最後に 3.6 節で本研究のまとめを行う．

3.2　関連研究

本節では，まず，前方不注意についての用語の定義を説明した後に，それについての既存の行動学的研究を紹介する．次いで，主に前方不注意に関する既存研究を，その動作レベル（表 1.1）の順に取り上げるとともに，ドライバの心的状態の推定手法を取り上げる．なお，これらの代表的な手法で用いられている観測量を図 3.1 にまとめている．

用語の定義

本論文では，事故統計上での「外在的前方不注意」(visual distraction)，すなわち，注意すべき対象以外を注視していたことによる発見の遅れ[1]のうち，さらに水平方向の首振り動作を伴うもののことを単に前方不注意と呼ぶ．この前方不注意は，さらに次の 2 つに大別される [113]．1 つは，景色に見とれる，ナビゲーションの操作といった運転上必ずしも必要でないものである．もう 1 つは，車線変更直前にミラーの死角を確認するといった運転の安全性を維持するために必要な前方不注意[2]である [113]．本論文では，前者を単に「脇見」，後者を「死角確認」と呼ぶ．

[1]．もう 1 つの前方不注意は，「内在的前方不注意」(cognitive distraction) と呼ばれるものである．考え事や会話等による意識や注意力の低下による危険の発見の遅れであり，漫然運転などがこれに該当する [113]．なお，外在的前方不注意と内在的前方不注意は両立しうる．例えば，同乗者に対して視線を向け（内在的前方不注意），さらに，その会話に意識の多くを向ける（外在的前方不注意）ことも頻繁に起こる．

[2]首振りをあまり伴わない運転に関連した前方不注意としては，信号やメータの確認などがある．

3.2.1　脇見についての行動学的研究

前方不注意と視線および頭部方向の関係

前方不注意と視線および頭部方向の関係については次のような知見が得られている．Inuzuka ら [40] によれば，市街地走行での前方注視時の注視点分布はほぼ 20[deg] 以内に収まるとされている．また，中越ら [111] は，注視対象の方向角が小さい場合には眼球回転を主とした注視が行われるが，注視対象の方向角が大きくなるに従って頭部回転角の割合が増加し，注視対象の水平角が 20[deg] を超えたあたりから頭部の水平回転角が眼球回転角を上回るという結果が得られている．これらの 2 つの知見をまとめると，前方注視と前方不注意のおよその境界は注視対象の水平方向角が 20[deg] のところであり，そのときの頭部の水平回転角はその半分の 10[deg] 程度であるということになる．つまり，頭部の水平回転角が 10[deg] を超えているか否かが前方注視か前方不注意かの判断基準となるといえる．なお，本論文での提案手法もこれに従っている（3.3.5 節参照）．

脇見行動に関する行動学的研究

田久保ら [114] は，実際の事故データおよびドライビングシミュレータ (DS) を用いた脇見についての行動学的研究を行っている．彼らの脇見の定義は，運転中に車両進行方向の道路状況および先行車以外の対象に注視点を移動させることである．彼らの脇見行動のモデルは，まず外部要因に起因する脇見に対する要求が入り，そのときの運転状況から脇見可能時間を算出するとともに脇見を実行するかの判断を行い，実行後はいつ脇見を終了するかの判断を行う，というものである．彼らは特に脇見時間と種々の要因との関係を分析している．脇見時間についての主な判断要因は衝突時間であり，それは先行車と自車の運動状態（絶対的／相対的な速度や加速度など）から算出される．たとえば，先行車への相対速度が高い（接近速度が高い）と脇見時間が短くなり，擬似衝突余裕時間が長いと脇見時間が長くなるという結果が得られている．また，副次的な要因としては，個人特性（たとえば女性では脇見時間は長い）や交通環境状況（たとえば晴天や交通閑散時に脇見時間が長い）があるとしている．事故データおよび DS を用いたデータから，いずれも，脇見時間が 2[sec] 付近を最頻値とした分布形状となるという結果が得られている．

3.2.2　前方不注意検出 (*action primitive* レベル)

動作検出の既存研究については, まずは，*action primitive* レベルである前方不注意の検出に関する研究を取り上げる．ここで特に注意すべき点は，いずれの従来手法においても，上で定義した死角確認と脇見の区別がなされていない点である．

注視動作情報のみを用いる手法

脇見検出に関する既存研究では，主にドライバの顔向き（頭部姿勢角）や視線方向を用いて前方不注視を検出する手法が提案されている [79, 107, 111, 112]．しかし，いずれの研究においても安全確認と脇見との区別はなされていない．

鈴木ら [116] は，車線変更前に行われる予備動作としての死角確認動作を認識する手法を提案している．具体的には，頭部姿勢を観測変数とする直進時と車線変更時の 2 クラスを識別する教師ありにて学習したベイズ識別器を用いている．観測変数は，視線の水平成分および垂直成分それぞれについての瞬間値および移動分散の計 4 つである．DS を用いた実験により，彼らの手法が予備動作の検出に有用であることが示唆されている．だが，対象は死角確認のみであり脇見は考慮されておらず, また，彼らのモデルには動作の時間方向の遷移も含まれていない．

身体動作以外の情報を用いる方法

Torkkola ら [88] は，ランダム森 (Random Forest [8]) を用いた死角確認動作の検出手法を提案している．しかし，脇見については考慮していない．観測としては，ハンドル角度，ペダル位置や自車から車線エッジまでの距離（自車両の横位置）などが用いられている（頭部姿勢の情報は用いられていない）．なお，これらの観測の統計量には，各時刻での瞬間値に加えて，対象時刻から数フレーム前までのデータの移動平均や移動分散といったものが用いられている．DS を用いた実験では，自車の横位置およびアクセルペダル踏込み量の瞬間値，および，ステアリング角の移動分散（約 1[sec]）などが死角確認の検出に特に有用であるという結果が得られている．

3.2.3　運転行動の認識 (*action* レベル)

次いで，より上位レベルの *action* レベルの運転行動認識に関する研究について取り上げる．ドライバ行動のモデル化を扱った研究には，車線変更行動 [48,60,71,76,116]，右折行動や停止行動 [35, 49, 61, 67] などがある．ここでは，特に，死角確認と関係の深い車線変更行動の認識に関する従来研究を取り上げる．ドライバの身体動作情報のみからでは認識が困難な動作レベルであるため，運転操作情報や運転状況など様々な観測変数を用いた手法が多い．

車線変更行動の認識

Pentland ら [71] は，Markov Dynamic Model を用いてドライバの車線変更行動を推定した．推定に用いられたデータは，アクセルとブレーキの踏込み量，およびステアリング角である．彼らは，車線変更行動を次の 6 つの動作に分割した．(1) 現在の車線において中心線に車両を合わせる, (2) 周囲を見回し隣の車線が開いているかを確認する，(3) 車線変更へ向けてハンドルをきる (4) 車線変更中，(5) 車線変更を終えるためにハンドルをきる，および，(6) 新しく走る車線に対して中心線に合わせる，である．この中で車線変更の予備動作に当たるのは，(1)〜(3) とし，このデータを学習データとして用いて，運転者が現在何を行動に移そうとしているのかを推定した．DS を用いた評価実験により，行動開始から 1.5 秒後での平均 95[%] の認識率が得られてている．この予備動作 (2) が本論文での死角確認におよそ相当するが，ここでもやはり脇見については考慮されていない．

Mandalia ら [60] は，線形 SVM(Support Vector Machines) を用いた実時間でうごく車線変更検出手法を提案している．特徴量には，運転操作情報としてアクセルの踏込み量およびステアリング角，また，運転状況情報として速度，先行車との距離および車線における横位置が用いられている．これらの観測変数の統計量としては 1.2 秒の時間窓が最も効果的であり，推定開始から 1.2 秒で最大 98%の推定精度が得られている．ただし，カーブ走行中には著しく推定精度が落ちることも報告されている．

3.2.4　ドライバの心的状態の推定 (*mental* レベル)

最後に，ドライバの心的状態を推定する研究を取り上げる．現時点において心的状態の対象として推定されているのは，ドライバの疲労度 [33, 34, 51, 103] および警

Authors			N. Oliver A. Pentland	N. Kuge et al.	D. Salvucci et al.	J. McCall M. Trivedi	K. Torkkola	L. Fletcher et al.	M. Su et al.	M. Kutila	Q. Ji et al.	H. Veeraraghavan et al.	T. Orazio et al.	**Proposed method**
Year			2000	2000	2007	2007	2004	2005	2006	2007	2006	2007	2007	2006
Target			Basic maneuvers (pass/stop/trun etc.)	Lane change	Lane change	Lane change	Inattention	Awareness of trafic signs	Fatigue & visual dist.	Distruction	Fatigue	safe / unsafe	normal / anomalous	no visual dist. / 2 types of visual dist.
Algorithm			DBN (HMM)	DBN (HMM)	Model tracing	Sparse Bayesian Learning	Random forests	Thresholding	Thresholding	SVM	DBN	Bayes classfier	Bayes classifier	DBN
Data collection (Real / Simulation / sTational car)			R	S	S, R	R	S	R	S	R	Y	T	R	S
Observation variables	Driver	head movements	✓			✓		✓	✓	✓	✓			✓
		eye movements						✓		✓	✓			
		blinking							✓		✓		✓	
		FE / AU									✓			
		body actions										✓		
	Operation	steering wheel angle	✓	✓	✓	✓	✓							✓
		gas pedal	✓		✓	✓	✓							✓
		brake pedal	✓			✓								
		gear	✓											
		winker												✓
	Own car	lateral lane position	✓		✓	✓	✓			✓				✓
		speed	✓			✓								✓
		logitudinal acceleration				✓								
		lateral acceleration				✓								
		yaw rate				✓								
	Surroundings	longt. dist. to a leading vehicle			✓									✓
		time headway to a lead vehcile			✓									
		longt dist. to vehcles in adj. lanes			✓									
		following vehicle												
		traffic signs						✓						

各論文への参照は，左から順に [67] [48] [75] [62] [88] [27] [79] [50] [42] [90] [17] である．

図 3.1: 従来研究で用いられている観測変数の一覧

戒/注意度 [5] ストレス度 [36] などである．これらはいずれも，ドライバの表情や視線といった顔に関する *action primitive* レベルあるいは *movement* レベルの動作と心的状態を直接関連付けるものである．

まず，Bergasa ら [5] は，近赤外線撮像システムにより計測した目と頭部の動きを観測変数として，Fuzzy Sysmtem を用いたドライバの不注意レベルの推定手法を提案している．

Ji ら [42] らは，ドライバの疲労度を推定するための動的ベイジアンネットワーク (DBN) を提案している．この DBN は，各時間スライスを表す階層的構造を BN で表し，そのうちの疲労度や表情などについて時間遷移（1 次の Markov 過程）から構成されている．BN については，ドライバの置かれた状況 (contextual information) → 疲労度 → ドライバの表情や視線などの動作情報 （→ 動画像) の階層構造により構成されている [33]．ここで，状況としては，睡眠の質，体調や時間帯などとしている．観測量は，瞼，視線き，頭部姿勢および表情 (Action Unit [22]) についての Fatigue index と呼ばれる指標（たとえば時間窓内の目の閉じている割合）としている．評価の評価に使われているデータは，運転中のデータではなく心理テスト (TOVA) 中に獲得したデータに留まっている．だが，心的状態状態の推定に，種々のコンテクスト情報や身体動作を要素として持つ階層的な構造の DBN を用いる有効性を示した文献として重要である．

3.2.5　従来手法のまとめ

本節では，前方不注意に関する従来研究について議論した．以上をまとめると，

1. 注視動作は，前方注視／死角確認／脇見の 3 種類に大きく分類される．しかし，前方不注意を検出に関する研究は複数存在するが，いずれも脇見と安全確認の区別を行っていない．

2. 運転動作を認識するためには，ドライバの身体動作，運転操作情報および運転状況情報といった様々な観測を用いることが有用であることが多く報告されている．

3. ドライバの心的状態の推定に，心的状態や動作のダイナミクスを取り入れた DBN が有用であることが示されている．だが，多くの注視動作の認識手法の

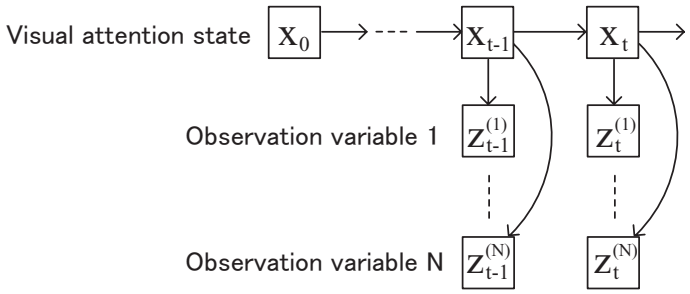


図 3.2: 注視動作と観測変数の間の因果関係を表す動的ベイジアンネットワーク

モデルには，注視行動のダイナミクスが取り入れられていない．

本研究では，これら全てを考慮した手法を提案する．まず，1 の 3 種類の注視動作を認識対象とする．また，提案手法は，2 を参考にして, ドライバの身体動作，運転操作情報および運転状況情報を観測として用いる．そして，これらの観測変数（結果）と認識対象の注視動作（原因）との間の因果関係，および，注視動作のダイナミクス（3）を，全て確率的に記述した動的ベイジアンネットワークを用いる．

続いて 3.3 節では，提案手法の詳細について説明する．

3.3　提案手法

本手法では，前方注視，死角確認および脇見の 3 つの注視動作と，ドライバの身体動作，運転操作および運転状況との間の関係を DBN を用いて表現する．そして，注視動作の時間遷移を扱いながら，観測が与えられたもとでのそれぞれの注視動作との間の尤度を計算することで，各時刻での注視動作の事後確率を逐次的に計算していく．

3.3.1　注視動作モデル

本手法では，注視動作および観測変数の間の確率的な因果関係を動的ベイジアンネットワーク (DBN) を用いてモデル化する．そして，各時刻での注視動作 x の事後確率を，注視動作と観測 z の間の尤度を計算することで推定する．図 3.2 にこの注視動作モデルのグラフィカル表示を示す．本論文では，注視動作 x のとり得る状態

は 3 種の離散的状態，つまり，$x \in \{$ 前方注視, 死角確認, 脇見 $\}$ である．この注視動作の定義については，3.3.5 節にて述べる．

この DBN は 2 つの部分からなる．1 つは，図 3.2 中の縦の矢印によって表わされている，注視動作と観測の間の因果関係である．これは一般に尤度と呼ばれることが多い．本手法では，各時刻 t における M 種類の全ての観測 $\boldsymbol{z}_t = \{z_t^{(i)}\}_{i=1}^{M}$ が，注視動作 x_t が与えられたもとで互いに独立であることを仮定する．すなわち，

$$p(\boldsymbol{z}_t|x_t) = \prod_{i=1}^{M} p(z_t^{(i)}|x_t) \tag{3.1}$$

である[3]．これは，付録にて述べる Naïve Bayes モデルに相当する．観測変数については 3.3.4 節にて述べる．

2 つめは，図 3.2 中の水平方向の矢印によって表わされている，注視動作の時間遷移を表すモデルである．ここでは，注視動作が 1 次のマルコフ過程に従うとする．すなわち，現在の注視動作 x_t は，直前の状態 x_{t-1} のみに依存すると仮定している．これにより，入力ノイズや運転操作上の突発的な乱れなどに対して頑健な推定が可能となる．

なお，注視動作の事前確率 $P(x_0)$，注視動作の遷移 $P(x_t|x_{t-1})$，および，尤度関数 $p(\boldsymbol{z})|x$ が，この DBN の 3 つのパラメタであり，いずれも時間不変であるものとする．これらの学習方法については，3.3.6 節にて述べる．

3.3.2　注視動作の事後確率の逐次的推定式

現時刻 t までに得られた全ての観測 $\boldsymbol{z}_{1:t}$ が与えられたもとでの，その時刻での各注視動作 x_t についての事後確率質量関数 (posterior probability mass function (pmf)) $P(x_t|\boldsymbol{z}_{1:t})$ を以下の逐次的に算出する[4]．

$$\begin{aligned} P(x_t|\boldsymbol{z}_{1:t}) &= \alpha p(x_t, \boldsymbol{z}_t|\boldsymbol{z}_{1:t-1}) \\ &= \alpha p(\boldsymbol{z}_t|x_t) P(x_t|\boldsymbol{z}_{1:t-1}) \\ &= \alpha p(\boldsymbol{z}_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}) P(x_{t-1}|\boldsymbol{z}_{1:t-1}) \end{aligned} \tag{3.2}$$

[3](3.1) 式を，2 章での (2.7) 式に対応させて書くと，$L(x_t|\boldsymbol{z}_t) = \prod_{i=1}^{M} L(x_t|z_t^{(i)})$ となる．

[4]2 章で説明した表情と頭部姿勢の同時認識のモデルとは異なり，推定対象が離散状態のみであるため，この (3.2) 式は厳密に計算可能である．また，計算コストに関しても各時刻において，3 つの注視動作間の遷移，および，それぞれの注視動作についての尤度（(3.1) 式）を計算するだけである．よって，サンプリングレート (約 30Hz．3.4.1 節）に比べてはるかに高速に計算可能である．

表 3.1: 本論文で用いる観測変数の一覧

Physical actions (*movement* level)	Head horizontal orientation (I) Head horizontal orientation (RSD)
Operations	Gas pedal depression (RSD) Steering angle (RSD) Winkers (I)
Context (Own-vehicle state)	Speed (I) Distance to the current lane edge (RSD)

I:Instantaneous value, RSD: Running standard variation.

ここで, α は, 事後確率密度関数が確率であることを満たす, すなわち, $\sum_{x_t} P(x_t|\boldsymbol{z}_{1:t}) = 1$ を保証するための正規化定数である.

3.3.3　推定結果の算出

(3.2) 式から得られるのは，各注視動作についての事後確率であるので，最終的な認識結果として，定量評価のためにも扱いやすい 1 つの状態として出力するようにする．ここでは，各時刻 t での注視動作の推定結果, $\widehat{x}_t$, については，その時刻での事後確率を最大とする状態，すなわち，

$$\widehat{x}_t = \arg\max_{x_t} P(x_t|\boldsymbol{z}_{1:t}) \tag{3.3}$$

とする.

3.3.4　観測変数

観測変数は，ドライバの身体動作[5]に加えて，コンテクスト情報として運転操作および運転状況からなる．これらの観測変数の一覧を表 3.1 に示す．身体動作は主に前方注意か前方不注視かを特徴付ける要素である．本論文では，この身体動作として，計測が比較的困難な視線方向ではなく，視線方向と関連の深い頭部の水平方向角 [111] を代用している．また，瞬間値（対象時刻での観測値）に加えて移動分散も観測変数として扱う．これは，車線変更の予備動作の検出に有用であることが従来研究 [116] にて示されているためである．次いで，運転操作情報は注視動作に影

[5]1.2 節での *movement*（物理）レベルの動作である．これについては，3.4.1 節にて述べる既存システムを用いて計測している.

響を受ける要素である．本論文では，従来知見を参考に，前方不注意時のステアリング技術の低下 [9]（ここでは死角確認の検出に有用な移動分散 [88] を特徴量とする），アクセルペダルの緩め [114]（移動分散）を扱うとともに，車線変更や合流などのよい指標であるウインカ（車線変更の認識に有用であろうことが文献 [75] にて述べられている）を観測として用いる．3 つめの運転状況はそれぞれの注視動作をどの程度引き起こしやすいかに関係する要因である．ここでは，脇見をする余裕に関係する自車のスピード [114]，および，前方不注意時のステアリング技術の低下 [9] を示す自車と現在レーンのエッジまでの水平距離を用いることとする．なお，移動標準偏差については，経験的に，頭部姿勢については対象時刻の 2 秒前〜6 秒前の間，その他の変数については対象時刻〜3 秒前の間の標準偏差とした．これらの観測量の獲得方法については 3.4.1 節にて述べる．

なお，本手法では，各観測変数に対する尤度 $p(z^{(i)}|x)$ をヒストグラムを用いて近似する[6]．ヒストグラムのビンの数については，ウインカのみ 3 つ（なし／左／右），それ以外については経験的に全て 12 とする[7]．このヒストグラム化は，ノイズや学習サンプル数に対して推定を頑健にする．また，(3.2) 式の計算を高速に実行できるというメリットもある．提案手法は複数の観測変数を用いて総合的に注視動作を推定するため，それぞれの観測変数についての尤度分布のヒストグラム化による近似誤差は推定に大した影響を及ぼさないものと考えられる．

なお，学習したヒストグラムに確率（頻度）が 0 となるビンが含まれていることは望ましくない．それは，認識の際の走行データにおいて，ノイズや運転のばらつきによりその確率 0 のビンに割り当てられる観測変数が 1 つでも含まれていれば，その注視動作の尤度（(3.1) 式）が 0 となりその注視動作は認識結果となり得ないためである．この問題を避けるため，各ビンの頻度の算出の際にそれぞれの頻度にあらかじめ 1 を加えて学習を行った．

[6]なお，DBN は各変数について，離散量／連続量のいずれとも扱うことが可能である．

[7]このビンの数とそれらがカバーする値の範囲については，与えられた学習サンプル数のもとで，識別対象の注視動作間の分布をなるべく多くするものが望ましい．つまり，ビンの数が多すぎれば，全てのビンに適切にデータを入れるために膨大な数の学習サンプルが必要となり，逆に，ビンの数が少なすぎれば，識別対象の注視動作間の分布の違いが失われていく．本論文では，このビンの数を単に経験的に決定した．

3.3.5　注視動作ラベル付け方法

注視動作に対する教師あり学習（3.3.6 節）や認識結果の定量評価のため，以下の方法で各サンプルのフレームに対して前方注視，死角確認および脇見の 3 種のいずれかの注視動作のラベルを付ける．

まず，ドライバの頭部姿勢の水平方向角 θ が閾値以下であれば，すなわち，$\theta < \tau$ であれば，前方注視状態であるとする．逆に，$\theta \geq \tau$ であれば，そのとき脇見用ターゲット（3.4.1 節）が表示されていれば脇見状態とし，表示されていなければ死角確認状態とした．なお，本論文では，この頭部姿勢角の閾値 τ については，3.2 節で述べた既存研究から得た知見をもとに 10[deg] とした．

3.3.6　DBN パラメタの学習

DBN の 3 つのパラメタである，注視動作の事前確率 $P(x_0)$，注視動作の遷移 $P(x_t|x_{t-1})$，および，尤度関数 $p(\boldsymbol{z}|x)$ については，教師あり学習を行う[8]，すなわち，各時刻での注視動作 x が予め付与された学習データを用いて学習を行う．なお，これらの学習データに必要な観測の獲得方法については，3.4.1 節にて述べる．

注視動作および各観測量は，ヒストグラム化（3.3.1 節）により，全て離散変数とされているため，DBN の 3 つのパラメタはいずれも対応するフレームをカウントすることにより算出される．

$$P(x_0 = \xi) = \sum_t \{x_t = \xi\}/M \tag{3.4}$$

$$P(x_t = \xi | x_{t-1} = \eta) = \sum_t \{x_t = \xi \wedge x_{t-1} = \eta\} / \sum_{t-1} \{x_{t-1} = \eta\} \tag{3.5}$$

$$p(z_t^{(i)} = \zeta | x_t = \xi) = \sum_t \{z_t^{(i)} = \zeta \wedge x_t = \xi\} / \sum_{t-1} \{x_t = \xi\} \tag{3.6}$$

ここで，M は学習サンプルフレームの総数を，$\sum\{\cdot\}$ は条件 $\{\cdot\}$ を満たすフレーム数をそれぞれ表す．

[8]教師なし学習では，正しくこれらのパラメタを学習することが困難である．それは，注視動作についての観測は，それぞれの動作（クラス）内において，大きなばらつきがある．例えば，前方注視の状態でも，前方不注意状態で比較的よく行われる，アクセルを緩めたりステアリングをほとんど動かさなかったりする動作は稀ではない．よって，これらを教師なしでクラスタリングのような手法で分類すれば，本研究で識別したい状態とは異なる状態を特徴づけるクラスタが形成されてしまう可能性が高い．

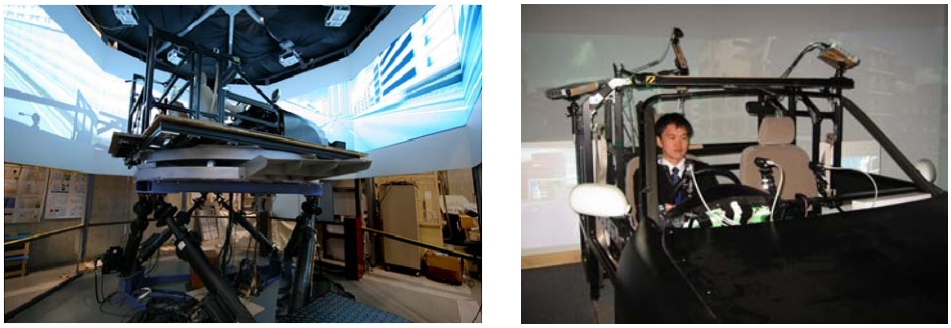


図 3.3: 実験に用いたドライビングシミュレータの外観

図 3.4: 頭部姿勢推定用カメラ（左）および頭部姿勢推定結果の一例（右）

3.4　評価実験

本節では，提案手法の有効性を実験により確認する．

本提案手法で用いる DBN は一種の統計モデルであるため，一般にそのパラメタの学習には十分な学習データが必要となる．これは，実環境では稀にしか発生しない脇見について特に問題となる．とはいえ脇見を故意的に発生させることは安全面から困難である．そこで，本研究では，提案手法の評価にドライビングシミュレータ（DS）を用いることとする．脇見の発生方法については 3.4.2 節にて述べる．

3.4.1　実験システム概要

本研究では，ユニバーサルドライビングシミュレータと呼ばれる DS [80, 109] を評価実験用に用いた．この DS の概要を図 3.3 に示す．この DS は車両を囲う正八角形の形状をした投影面に，プロジェクタを用いて周囲環境および周辺車両を 360 度全方位に投影する．また，左右のサイドミラーに対しては車両を囲う投影面とは別

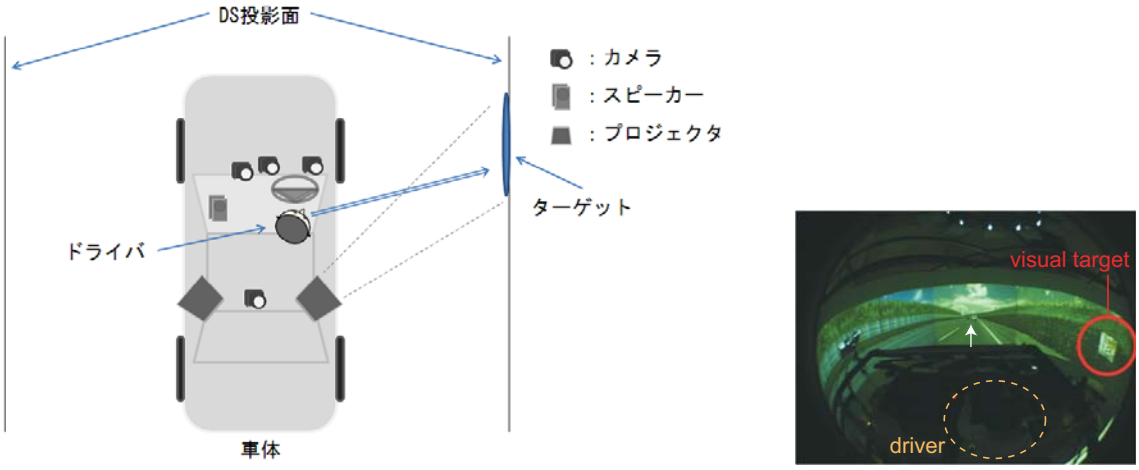


図 3.5: 脇見発生装置の概要（左），および，脇見ターゲット画像表示の一例（右）

に投影面を用意し，より現実感の強い情景がミラーに映りこむように工夫されている．また，周辺車両についても，その挙動をインタラクティブに変化させるよう制御されている．これにより実環境での運転に極めて近い視界環境が構築されている．また，操作環境も実車を可能な限り再現しており，ステアリング操作やペダル類の操作はもちろん，様々な情報を 60[fps] にて取得可能である．

この DS は，ドライバの頭部姿勢を出力する機能を備えていないため，本研究では，この DS に既存の頭部姿勢推定システム [106] を組み込んだ[9]．この追跡システムは，複数台のカメラにより観測される顔の特徴点をパーティクルフィルタを用いて追跡することで，6DOF の頭部姿勢（3 次元の位置および 3 次元の方向）をおよそ 30[fps] で出力する．この頭部姿勢推定システム用に設置した運転席前方 2 台および左方 1 台の計 3 台の VGA 解像度のグレースケールカメラを図 3.4 および図 3.5 左に示す．また，運転席上部後方に，車線変更など各時刻での運転行動やイベントの検証用に運転席および前方を広く撮影する広角レンズを搭載したカラーカメラを 1 台設置した（図 3.5 左）．

さらに，脇見を発生させるためのターゲット画像の投影用として，後部座席位置の左右外側に合計 2 台のプロジェクタ[10]を設置した（図 3.5 左）．これらのプロジェクタの設置位置については，ターゲットがミラーからでは目視できないドライバの

[9] これは，注視動作時のドライバの首振りの範囲が広く，また，表情の情報は注視動作の識別にさほど寄与しないため，この *movement* レベルの動作である頭部姿勢のみを得るためには，2 章にて述べた方法よりもこのシステムの方が優れているとの判断による．

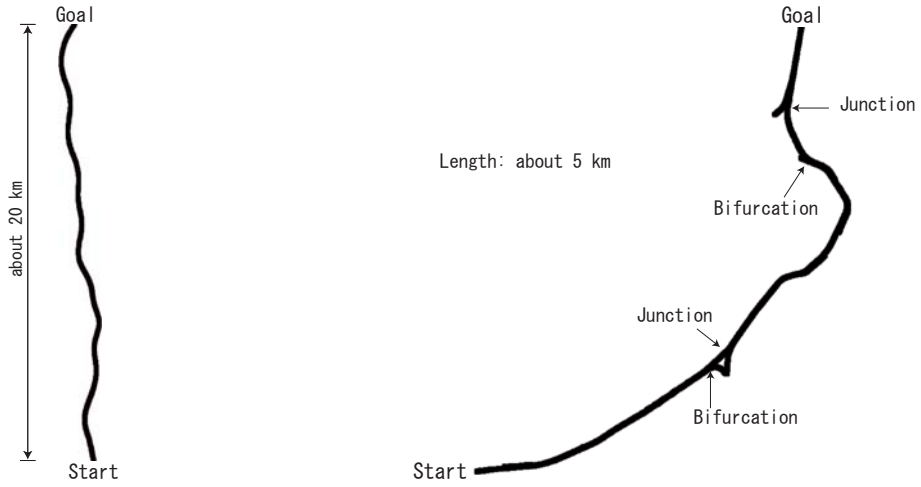
[10] このプロジェクタは最大，上下 30[deg]，左右 40[deg] の範囲に投影可能である．

死角，すなわち，首を振らなければ目視できない位置に投影されるようにした．この脇見用ターゲット画像は表示中静止しており，また，これらのプロジェクタはドライバの搭乗している車体のフレームに取り付けられているため，脇見用ターゲット画像は一回の表示の間ではドライバに対して方向を変えることはない．図 3.5 右にその一例を示す．表示の時間間隔，投影に使用するプロジェクタの選択（投影の左右方向），プロジェクタ投影可能範囲内でのターゲットの提示位置，および，表示するターゲットの選択は，各表示に対していずれもランダムとした．1 ターゲット当たりの提示時間は 5[sec]，各表示の時間間隔は 41.6[sec/回]，選択可能なターゲットの種類は 51 種類とした．表示するターゲットとしては，道路標識やドライバの興味を引く画像，四字熟語などを用いた．また，ドライバに心的負荷を与えずにより自然に近い脇見を発生させるために，次のターゲットが左右どちらの方向に表示されるのかを，その表示の 5 秒前および 1 秒前に助手席の足元に置いたスピーカの音（ターゲットが表示される方向（左／右）で異なる音）により知らせた．

なお，これらそれぞれのシステムはサンプリングレートが異なるため，本研究では最も遅い頭部姿勢のサンプリングレート（約 30[fps]）にて全データを取得した．

3.4.2　ドライバへのタスク

被験者に対しては，以下の 3 つの指示を行った．まずは，実世界のときと同様に，安全かつ自然な運転をすることとした．これがメインタスクである．被験者にはさらに 2 つのサブタスクを与えた．1 つめのサブタスクは，脇見を発生させるためのものである．ドライバの左右方向に現れる脇見用のターゲット画像（3.4.1 節）を，それが表示されている間，安全の範囲内で見続けるよう指示した．なお，自然な脇見を再現するために，ターゲットを注視するタイミングおよび継続時間については自由とした．2 つめのサブタスクは，ときどき可能であれば車線変更をするということである．これは，特に，分岐や合流といった死角確認を必要とする場面のない単調なコース（3.4.3 節参照）において，死角確認を行わせることを意図した指示である．ただし，ターゲットが間もなく表示されるということが音声合図により分かっている（3.4.1 節）場合には，車線変更を行わないように指示した．



(a) 東名大和サグ　　(b) 首都高速 3 号線

図 3.6: テストコースの線形

3.4.3　走行コース

本研究では，特性の異なる 2 つの実在する高速道路を模擬したコースを用いて提案手法の評価を行う．2 つのコースの違いを一言で言うと運転の難易度であり，結果的に脇見のしやすさや安全確認の必要性などが異なる．1 つめのコースは，緩やかなカーブからなる単調なコースである．死角視認を行う必要があるのは基本的には車線変更のみである．よって，前方注視，死角確認および脇見による観測の分布が明確，つまり，観測量にノイズが少なく，これらの注視動作の識別が比較的容易であることが予想される．これに対して 2 つ目のコースは，曲率の小さなカーブ，合流，分岐，トンネルなどを含む変化に富むコースである．つまり，様々な場面で，前方注視，死角確認および脇見を行うこととなる．よって，観測のばらつきが大きく，これらの注視動作の識別はより難しいことが予想される．両コースの線形を図 3.6 に示すとともに，両コースの特徴を表 3.2 にまとめる．両コースに対して 4 名の被験者がそれぞれ走行した．

3.4.4　実験結果

ここでは，観測変数の違い，および，コースによる認識結果の違いを検証する．検証方法としては，leave-one-subject-out 交差検定法を用いる．すなわち，評価対象の

表 3.2: 実験の要約.

Course name	Course A	Course B
Existing road name	Yamato Sag （大和サグ）	Metropolitan Expressway （首都高速 3 号線）
Length [km]	ca. 20	ca. 5
Geometry	slightly curved	winding
#Bifurcations + junctions	0	4
#Lanes	3	2
Traffic	moderate	moderate
#drives	15	32
#subjects	4	4
#frames	306,389	215,033
no VD	264,545	195,166
Dr-VD	21,389	9,503
non-Dr-VD	20,455	10,364

’no VD’，’Dr-VD’ および’non-Dr-VD’ は，前方注視 (no visual distraction),
死角確認（driving-related visual distraction）,
および脇見（non-driving-related visual distraction）をそれぞれ示す.

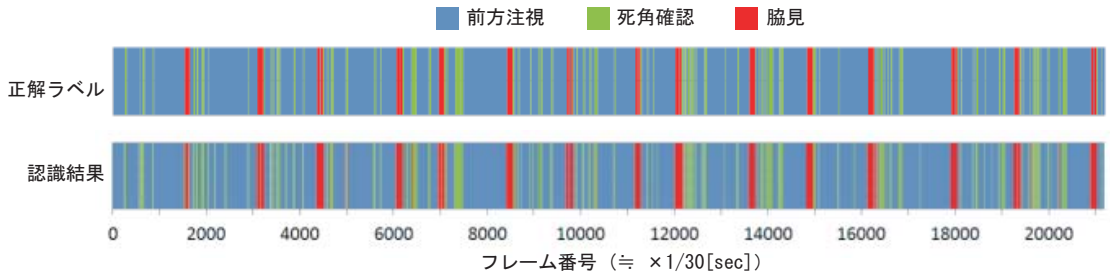
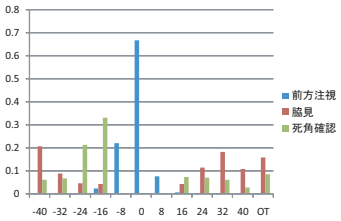


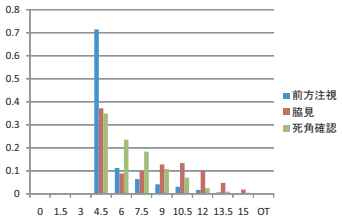
図 3.7: 1 走行データについての注視動作の認識結果例（コース A）

被験者以外の全ての被験者の走行データを用いて DBN パラメタの学習を行い，そのパラメタを用いて対象被験者の各走行データについての認識率を算出する．なお，この DBN パラメタの学習の際にはコースを区別していない．つまり，人物およびコースどちらについても汎用なモデルを作成している．これは，運転の場面では個人毎に注視状態ラベルの付いた学習データを準備することがあまり現実的でないとの考えによる．

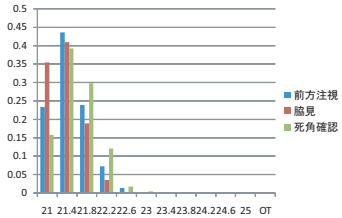
まず，図 3.7 にコース A についての 1 走行データの注視動作の認識結果の時系列



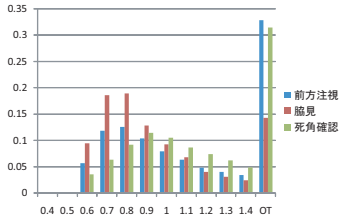
(a) 頭部水平方向角 (I)



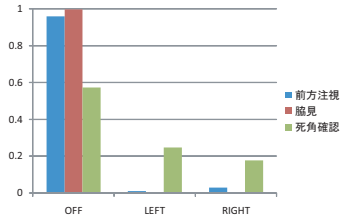
(b) 頭部水平方向角 (RSD)



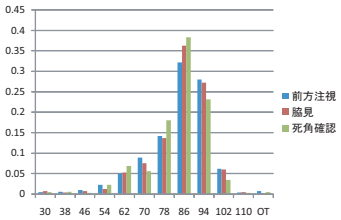
(c) アクセルペダル (RSD)



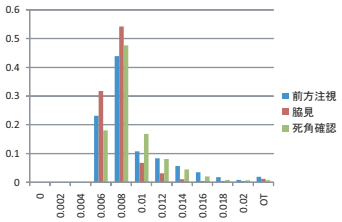
(d) ステアリング角 (RSD)



(e) ウィンカー (I)



(f) 車速 (I)



(g) レーン上の位置 (RSD)

それぞれ横軸はヒストグラムの各ビン，縦軸は確率（頻度）を表す．また，“I” は瞬間値，“RSD” は移動標準偏差をそれぞれ表わす．なお，各ビンの一番右の “OT” には，それ以外のいずれのビンにも該当しなかったものをまとめている．

図 3.8: 各注視動作について学習された観測変数のヒストグラム

データを示す．動作モデルに対象人物の走行データが含まれていないにもかかわらず，3 種類の注視動作がほぼ正確に識別できていることが見て取れる．また，誤認識が生じている部分については，その 1 回の注視動作がまるごと誤認識されてことはほとんどなく，多くは，例えばその注視動作の開始部分といったように部分的に生じていることが見て取れる．

次いで，図 3.8 に各注視動作について学習された観測変数のヒストグラムの例を示す．各観測変数について順に説明していくと，まず，頭部姿勢の水平角の瞬間値

(3.8(a)) については，前方注視と前方不注意（脇見＋死角確認）の間に大きな差がみられる．これは 3.3.5 節の定義上当然の結果である．なお，前方注視と前方不注意の定義における境界である ±10[deg] を跨ぐようにビンが設定されているため，一部両者の確率が共に 0 でないビンが存在している．次いで，頭部姿勢の水平角の移動標準偏差 (3.8(b)) については，前方注視と死角確認については，前方注視の方がやや小さな値へと偏っているものの，どちらも大きな標準偏差になるほど確率が減っている．一方，脇見では 10[deg] あたりにある第二のピークが特徴的であるが，これは，被験者の多くが脇見ターゲットを初めに見た後に，一旦正面方向を向き，その後，再度脇見ターゲットを見るということを行った結果であると考えられる．このような行動は実際の脇見においてもよく行われるものと考えられる．

続いて，運転操作についてのヒストグラムについてみる．まず，アクセルペダル (3.8(c)) については，前方注視と脇見の間に差はあまり見られないが，死角確認については踏み込み量が比較的大きいことが見て取れる．これは，車線変更の開始時に加速するという行動によるものであると考える．続いて，ステアリングの移動標準偏差 (3.8(d)) については，脇見の際にはステアリング動作が貧弱になる [88]，つまり，ステアリングを切る頻度が少なく，切る場合には大きく切るという現象が表れていることが見て取れる．すなわち，標準偏差の小さな部分と大きな部分に 2 つのピークがはっきりと表れている（レンジオーバーの “OT” のビンがステアリングを急に大きく切る場合に相当する）．また，ウインカ (3.8(e)) については，本実験では主に車線変更を行う際に点灯されたものであり，その際に死角確認が行われたことが表れている．このことからウインカ情報が死角確認の識別に非常に有用であることが分かる．

最後に運転状況についてのヒストグラムについてみる．まず，自車速度 (3.8(f)) については 3 つの注視動作間の差がほとんど見られない．よって，今回の実験ではあまり識別に寄与していないと考えられる．一方，レーン上の位置，すなわち，自車と現在レーンのエッジまでの水平距離 (3.8(g)) については，注視動作間の違いが若干みられる．特に，死角確認については，車線変更の予備動作（ミラー確認など）が長く，その間にレーン上の位置が多少ばらついていることが表れているものと考える．

続いて，表 3.3 に，表 3.1 に挙げた観測変数全てを用いた場合と，それとの比較のために頭部姿勢に関する観測変数のみを用いたそれぞれの場合における認識結果の混同行列を示す．この表 3.3 には，コース毎の認識率およびそれらをあわせた全体

表 3.3: 全体の認識結果についての混同行列 (%).

Label \ Recog.	no VD	Dr-VD	non-Dr-VD
Total			
no VD	97.9 / **97.9**	0.2 / **0.4**	1.9 / **1.7**
Dr-VD	1.5 / **2.3**	79.9 / **86.1**	18.6 / **11.6**
non-Dr-VD	1.8 / **4.5**	19.6 / **16.7**	78.6 / **78.8**
Course A			
no VD	97.7 / **97.7**	0.2 / **0.4**	2.1 / **1.9**
Dr-VD	1.4 / **1.4**	81.9 / **92.0**	16.7 / **6.6**
non-Dr-VD	1.7 / **3.4**	19.6 / **18.9**	78.7 / **77.7**
Course B			
no VD	98.2 / **98.4**	0.2 / **0.3**	1.6 / **1.3**
Dr-VD	1.7 / **4.1**	76.1 / **74.6**	22.2 / **21.3**
non-Dr-VD	2.1 / **7.0**	19.8 / **11.6**	78.1 / **81.4**

行および列は，それぞれ，正解ラベルおよび認識結果を表す．つまり，3 つの行列の各対角要素および各対角要素は，正しい認識および誤認識をそれぞれ表している．各セルについては，観測として，ドライバの頭部姿勢情報のみを用いた認識結果を左側に，表 3.1 に示した全ての観測を用いた認識結果を右側に示す．

の認識率を示している．なお，これらの認識率は，全走行サンプルの各フレームについて認識結果と正解ラベルとの一致／不一致をもとに算出した．

上でも述べたとおり前方注視と前方不注意（死角確認と脇見を区別しない場合）の識別については定義上容易であるが，識別率が 100%でないのは前方注視と前方不注意の定義における境界である ±10[deg] を跨ぐようにヒストグラムのビンが設定されているためである．なお，仮にビンの境界が ±10[deg] となるように設定しても，3.3.4 節にて述べたとおり，各ビンの頻度が 0 とならないようバイアスを設けているため，識別率が常に 100%となるとは限らない．

一方，死角確認と脇見の識別については，全ての観測変数を用いた場合のほうが，認識率が死角確認について 6.2 ポイント，脇見について 0.2 ポイント上昇している．これにより，ドライバの頭部姿勢情報に運転操作情報および運転状況情報を加えることが，死角確認と脇見の識別に有用であることが示唆される．

また，2 種類のコースについての識別結果の違いは次のとおりである．まず，使用する観測変数にかかわらず，より難易度の高いコース B の方が単調な線形のコース A よりも死角確認の認識率が低い．また，コース B については，使用する観測変数の違いがコース A ほど顕著に現れていない．これは，コース A での死角確認が行

われるのが基本的に車線変更時のみであるのに対し，コース B ではそれに加えて合流がある．よって，コース B の方が学習される死角確認時および脇見時の観測量のばらつきが大きくなり，その結果誤認識される割合が多かったことが示唆される．

この評価では，DBN のパラメタを対象被験者以外の走行データから学習しているにもかかわらず，高い認識率が得られていることから，提案手法の注視動作の個人差に対する頑健性が示唆される．

3.5　考察

ここでは，死角確認および脇見の識別についての考察を行う．まず，コンテクスト情報，すなわち，運転操作情報および運転状況情報の導入については，単純なコース A では主に死角確認についての脇見への誤認識率が向上している．これについては次のことが考えられる．まず，車線変更での死角確認と脇見が，身体動作（頭部姿勢）的には類似していたが運転操作や運転状況には違いがあった．このため，これらの一部は身体動作情報のみからでは識別困難であったが，運転操作や運転状況を導入することで正しい識別が可能となったものと考える．一方，コース B については，運転操作情報および運転状況情報は，死角確認と脇見の識別にほとんど寄与していない．これは，コース B では死角確認が様々な状況の下で行われたことによるものと考える．つまり，車線変更以外にも合流等で行われるとともに，線形や勾配の変化が激しいため，死角確認と脇見のそれぞれの観測量分布のばらつきが大きくなり，違いが不明確になっているものと思われる．

また，死角確認と脇見それぞれについての結果を見ると，死角確認の方がよりコンテクスト情報を用いることによる識別精度の向上がみられる．これについては，本論文での脇見発生方法に起因すると考える．本論文での実験では，運転状況を考慮せずにランダムに脇見ターゲットを表出させた．このため，被験者が実際にはあまり脇見をする余裕がなくともターゲットを注視してしまうという事態が多々発生していたものと考える．その結果，脇見についてのコンテクスト情報があまり識別に有用でなかったと考える．他方，死角確認については自然な状態で行われており，より安定した動作が行われているものと考える．

なお，曲率の小さなカーブの多いコース B では，前方注視の認識率が単調な線形のコース A の認識率よりも低いことが予想されたが，結果はほとんど変わらなかっ

た．この予想は，カーブ走行時にはドライバはカーブの接点（視野内に移るカーブの曲線の接線方向が鉛直方向に一致する点）に着目することが多いことが明らかにされている [24] が，提案手法では道路線形に関わらず観測される頭部姿勢をそのまま入力していることにもとづく．だがそのような道路線形による注視方向の変化は，文献 [40] あるいは文献 [111] で述べられているように，主に視線方向により対処され，頭部姿勢変化にはさほど現れなかったためであると考える．とはいえ，この道路線形による視線方向の変化の補正は，死角確認と脇見の識別性能をより向上させることができると考える．また，特に視線を観測に導入する際にも必要であると考える．

3.6　結論

本研究では，これまで区別されてこなかった死角確認と脇見という，身体動作的には類似した 2 つの注視動作の意図の違いの重大さに注目した．そして，これら 2 つの注視動作と，さらに前方注視を識別可能な手法を初めて提案した．これらの注視動作を識別するための観測としては，ドライバの身体動作である頭部姿勢に加え，2 種類の前方不注意の識別のためのコンテクスト情報として，ペダルやステアリングといった運転操作，および，車速などの運転状況という計 3 種類の情報を用いた．これらの注視動作と観測変数の間の関係については，それらを確率的な因果関係として表現する動的ベイジアンネットワークを用いてモデル化した．そして，観測が与えられたもとでのそれぞれの注視動作の事後確率を，リアルタイムにて逐次的に計算した．ドライビングシミュレータを用いた実験により，3 種類の注視動作の識別に，ドライバの身体動作に加えて運転操作および運転状況というコンテクスト情報を観測として用いることにより高い認識率を達成可能なことが示唆された．

今後の課題としては，以下のとおりである．まず，観測量の種類については，本論文では従来研究から得られた様々な知見に基づき決定した．だが，より高精度の注視行動の認識，さらには，より多様な運転動作の理解には，この他にも特に様々なコンテクスト情報を導入する必要があるものと思われる．それにはまず道路線形がある．これについては，3.5 節にて述べたとおり，カーブ走行時のドライバの視線に対するバイアス補正に有用である．さらに，現在あるいは少し先の道路線形はドライバが脇見をするかどうかの判断に大きく影響を及ぼすことが予想される．その

他にも，先行車両との間隔が，脇見を行うかどうか，また，脇見の持続時間を決定するのに利用されている [114] という点から重要である．この他にも考えられる観測は多数ある．

ただし，やみくもに多数の観測を加えることは，むしろ推定性能を下げる可能性があるとともに，計算量の削減の観点からも望ましくない．そこで，そのような様々な観測変数を候補として，そこから機械的に識別に最適な観測量集合を選択する枠組みを導入する必要があると考える．そのような特徴選択の手法には，例えば，文献 [88] で用いられている Random Forest [8]，文献 [62] で使用されている Sparse Bayesian Learning [86]，あるいは，Adaboost [29] など様々なものがある（その他の手法についてはたとえば [53] など参照）．

また，本研究では，注視動作の時間的遷移を 1 次マルコフ過程とした．これは，脇見の持続時間の確率が持続時間について単純減少することを仮定していることを意味する．しかし，実際の脇見時間の分布は，3.2 節にて取り上げた田久保らの研究 [113] で得られているおよそ 2 秒を最頻値とした分布に近いものとなっているはずである．このようなより実際に近い遷移モデルの導入も，推定精度の向上に有用であるものと思われる．

提案手法の枠組みは拡張性が高いため，車線変更や右左折などの基本運転行動を扱うとともに，より深いレベルにあるドライバの内部状態の推定に対して拡張可能であると考える．

第4章　結 論

4.1　本研究のまとめ

本論文では，*action primitive* レベルの動作の認識に関する 2 種類の手法を提案した．それは，様々な情報が統合された観測データからの複数動作の認識手法，および，コンテクストに依存する動作の認識手法である．それらのモデル化には，観測変数から人間動作そして内部状態までを統一的に表現可能であると考えられる動的ベイジアンネットワークを用いた．また，最終的にそのような多数の要因を扱うことを前提とし，提案手法ではなるべく要因間の関係をシンプルに表現した．このため，本論文にて行った観測変数から人間動作を認識するモデルに，より複雑な人間行動や心的状態を要因として加えることが容易である．

2 章では，様々な情報が統合された観測データからの複数動作の認識手法を提案した．具体的には，単眼動画像に基づき表情および頭部姿勢を同時にする手法を提案した．表情にはその人物が特に伝達したい多くのメッセージを含むことが明らかにされているが，人はまた，対話のなかで他の対話参与者へ顔を向けるなど様々な方向に顔を向ける動作も行う．この頭部姿勢の変化は観測である画像に大きな見えの変化をもたらすため，表情を認識するためには頭部姿勢も併せて推定する必要がある．だが，これまで提案されてきた多くの表情認識手法では，画像中の顔がカメラに対して正面正立であることを前提としたものが大部分であった．また，頭部姿勢変動に対処可能な従来手法は複雑な顔モデルを用いるため，その顔モデルの作成に，ステレオシステムや事前の膨大な学習データの収集を要するなどの問題があった．そこで，本研究では，その問題の解決を目指し，その場で簡単に個人に特化して作成可能な変動輝度テンプレートと呼ぶ新たな顔モデルを用いた手法を提案した．提案手法は，この変動輝度テンプレートを用い，本論文にて提案したパーティクルフィルタと勾配法を組み合わせた頑健かつ効率的な推定の枠組みにより頭部姿勢と表情を同時に推定する．実験により，面内のみならず面外方向の回転を含む様々な

頭部姿勢において驚きや笑顔などの 5 種類の表情を 90%以上の精度で認識できることが確認された．顔の動画像から表情と頭部姿勢を同時に推定する提案手法は，遠隔対話システムやインテリジェントルーム，エージェントロボットなど，人対人あるいは人対機械の間のインタラクションを行う幅広い場面へ適用することが可能であると考える．

次いで，3 章では，単に身体動作情報のみからでは正しい識別が困難な，コンテクストに依存する動作の認識手法を提案した．近年の重大事故の主要な原因の一つである前方不注意には，ミラーの死角を確認するという運転に関わる前方不注意と，景色に見とれるといった運転とは直接関係のない前方不注意とがある．これらは身体動作的には類似しているが，背後にある意図の違いは重大である．つまり，前者が安全性を高めるために行われるのに対し，後者はむしろ事故リスクの増大を承知で行われる．だが，従来の脇見検出に関する研究では，これらの 2 種類の前方不注意が区別されてこなかった．そこで，本研究ではこの点に注目し，前方注視／死角確認／脇見の 3 種類の注視動作を識別可能な手法を提案した．観測変数としては，ドライバの身体動作である頭部姿勢に加えて，コンテクスト情報としてペダルやステアリングといった運転操作，および，車速などの運転状況という 3 種類を用いた．提案手法では注視動作と観測変数の間の関係を確率的因果関係として扱う動的ベイジアンネットワークにてモデル化し，観測が与えられたもとでのそれぞれの注視動作の事後確率をリアルタイムにて逐次的に計算する．ドライビングシミュレータを用いた実験では，ドライバの身体動作に加えて運転操作および運転状況というコンテクスト情報を観測として用いることで，3 種類の注視動作を高精度に識別可能なことが示唆された．

4.2　今後の課題

今後の課題としてまず考えられるのは，提案手法の汎用性を生かして観測変数および認識対象を追加するという動的ベイジアンネットワーク (DBN) の拡張を行うことである．つまり，多くの観測を入力とし，そこから様々な動作を認識しつつ，最終的に本研究の大目的である内部状態についての推定を実現できるようにすることである．

まず，推定対象を増やすことについては，多くの情報を含む画像から，視線，ジェ

スチャ，姿勢など様々な *action primitive* レベルの動作を認識するということが第一に考えられる．次いで，より上位レベルの動作や心的状態を認識できるようにしたい．これについては，付録で述べているとおり，基本的には，新たに追加する要因とその下位の要因の間の条件付き確率を学習し，それを DBN に追加することで可能となる[1]．

また，本研究では，学習方法として教師あり学習，すなわち，全ての学習サンプルについて，動作の正解ラベルを人手で与えるという方法を選んだ．だが，扱う要因の数が多くなるにつれてラベリングの作業量が増大し，このアプローチは困難になる．このような問題に対処するためには，少数の意味付与された学習サンプルを用いる，すなわち，半教師あり学習が適しているものと考えられる．ただし，教師あり／半教師ありのいずれの場合においても，内部状態についての学習を行う際には問題が生じる．それは，どのようにして内部状態についての正解ラベルを与えるかという問題である．たとえば，比較的簡単に実施可能なアンケート方式では，正確かつ時間解像度の高いラベル付けは困難である．また，脳波計などの生体センサの情報を利用することも考えられるが，それらのセンサデータと推定したい心的状態のラベルへのマッピング技術が確立されているとは言い難い．

一方，観測を増やすことについては，まず，音声情報を取り入れることが考えられる．音声には，発話内容はもちろん，発話のパワーやピッチなど，感情に影響を受けるノンバーバル情報を含まれている．表情と音声はそこから認識しやすい感情カテゴリが異なる [98] ことから，感情の推定には音声情報を取り入れることが重要であると思われる．一方，運転の場面では，ドライバが利用している重要な運転状況として，道路線形，道路勾配あるいは先行車などがある．最終的な高次レベルの動作や内部状態の推定には，人間が利用するあらゆる情報を獲得できる様々なセンサを導入する必要があるものと考えられる．

ただし，種々の観測変数の導入については，やみくもに多くの観測を扱えばよいというわけではない．それは，使用する観測には，ノイズを多く含むなど対象の動作の認識にむしろ悪影響を及ぼすものが含まれうるためである．よって，それぞれの対象動作の識別に有用な観測を見つけるという特徴選択 (feature selection) の枠組みを取り入れる必要がある．これについては，たとえば，文献 [88] で用いられてい

[1]詳しい計算方法をここで述べることは避ける．なお，これについては文献 [42, 63] などが参考になる．

る Random Forest [8]，文献 [62] で使用されている Sparse Bayesian Learning [86]，あるいは，Adaboost [29] など様々なものがある．

最後に，我々人間は誰しも，他人と接する場合と同様，システムと接する場合にも自分の内部状態の全てを暴かれることに対して抵抗があるだろう．よって，人間の内部状態を読み取ることが可能なシステムができたとしても，人は身の内の状態を隠そうとして偽りの動作を行うようになることが予想される．よって，どのような人の内部状態を推定し，それをどのように生かすかについては，慎重に議論しなければならない．多方面で本当に役に立つ人間の内部状態推定システムを実現することは決して容易ではない．

付 録

本研究では，観測と認識対象動作との間の関係を記述するために動的ベイジアンネットワーク (DBN: Dynamic Bayesian Networks) を用いる．そこで，DBN が人間の内部状態推定に適している理由についてここで簡単に説明する．

まず，1.3 節にて述べた人間の内部状態を推定可能のための枠組みとして望ましい特性を再度ここに挙げる．

1. 不確定性を表現可能
2. 事前知識を導入可能
3. 観測の一部欠損への対処が可能
4. 要素の追加や削除が容易
5. 必要な学習サンプル数が少ない
6. 要素のダイナミクスを表現可能

以下順にこれらの特性と DBN の関係について説明する．

まず，特性 1 についての解決策の一つは，各要素について取りうる値および要素間の関係を確率分布で表現することである．

特性 2 についてはベイズ推定が適している．表 1.1 に既に示した $\boldsymbol{A}$(Action) および $\boldsymbol{M}$(Measurement) の関係のみを考えると，

$$\boldsymbol{A} \rightarrow \boldsymbol{M} \tag{4.1}$$

という因果関係がある．そして，$\boldsymbol{A}$ についての事前知識として，$\boldsymbol{A}$ の確率分布 $p(\boldsymbol{A})$，すなわち，他に何も情報がないときに $\boldsymbol{A}$ がどのような値をどのくらいの確率で取り得るかの情報を持っているとする．このとき，ベイズ推定では，$\boldsymbol{M}$ が与えられたも

とでの $\boldsymbol{A}$ の事後確率を，

$$p(\boldsymbol{A}|\boldsymbol{M}) = \frac{p(\boldsymbol{M}|\boldsymbol{A})p(\boldsymbol{A})}{\sum_{\boldsymbol{A}} p(\boldsymbol{M}|\boldsymbol{A})p(\boldsymbol{A})} \tag{4.2}$$

と算出する．ここで，$p(\boldsymbol{M}|\boldsymbol{A})$ は $\boldsymbol{A}$ の $\boldsymbol{M}$ に対する尤度と呼ばれる．また，分子については，$p(\boldsymbol{M}|\boldsymbol{A}) = p(\boldsymbol{M}, \boldsymbol{A})$ とも表せ，これは $\boldsymbol{M}$ と $\boldsymbol{A}$ の同時確率と呼ばれる．

特性 3 については，周辺化（例えば文献 [74] 参照）と呼ばれる方法により対処可能である．たとえば，推定したい $\boldsymbol{A}$ と関係のある d 種類の movement 要素 $\boldsymbol{M} = \{M_1, \cdots, M_d\}$ があるとする．つまり，(4.2) 式での尤度は

$$p(\boldsymbol{M}|\boldsymbol{A}) = p(M_1, \cdots, M_d|\boldsymbol{A}) \tag{4.3}$$

と書き表わされる．このとき，既知であるべき M_1 が与えられていないとすると，この M_1 について取り得る値全てを考慮することで尤度が計算可能である．数式にて表わすと，

$$p(M_2, \cdots, M_d|\boldsymbol{A}) = \sum_{M_1} p(M_1, \cdots, M_d|\boldsymbol{A}) \tag{4.4}$$

となる．

特性 4 についての要素の追加や削除も比較的容易である．例えば，$\boldsymbol{A}$ のより上位の動作 $\boldsymbol{B}$(Behavior) をモデルに追加する，すなわち，

$$\boldsymbol{B} \rightarrow \boldsymbol{A} \rightarrow \boldsymbol{M} \tag{4.5}$$

と拡張したいとする．このとき，$\boldsymbol{B}$, $\boldsymbol{A}$ および $\boldsymbol{B}$ の同時確率は,

$$p(\boldsymbol{B}, \boldsymbol{A}, \boldsymbol{M}) = p(\boldsymbol{M}|\boldsymbol{A})p(\boldsymbol{A}|\boldsymbol{B})p(\boldsymbol{B}) \tag{4.6}$$

と分解される．これは，既に (4.2) 式に示すモデルを学習できていれば，単に新たに $p(\boldsymbol{A}|\boldsymbol{B})$ を追加し，$\boldsymbol{A}$ についての事前知識を $\boldsymbol{B}$ についての事前知識 $p(\boldsymbol{B})$ に変えるだけで済むことを意味している．これは一種のグラフィカルモデリング（DBN もこの一種）の特性である．グラフィカルモデリングでは，扱う要因のうちの各要因のペアが，それぞれ因果関係にあるかそれとも独立であるかという事前知識（人間の経験的知識でもよい）が既に与えられている場合には，それに従って直接的な因果関係のある要因のみをリンク（(4.5) 式での $\rightarrow$）でつなぐことでその因果関係を表現可能という特性，つまり，特性 2 を併せ持つ．

学習サンプル数が少なくすむという特性 5 については，条件付き独立性の導入により達成される．一般的に，過学習を避けることができるだけの学習サンプル数は確率分布の次元数に合わせて増加する．(4.6) 式でみたように，要素間の条件付き独立性を導入することは，それに関係する変数の同時確率（高次元）を，複数の低次元の部分空間の集合へと落とすことと等価である．仮に各変数の次元数が等しく d であるとすれば，(4.6) 式の左辺の次元数は $O\left(d^3\right)$ であるのに対し，右辺の各確率分布の次元数は高々 $O\left(d^2\right)$ と低い．さらに，この条件付き独立性は尤度についても導入することが可能である．(4.6) 式では，2 つの尤度それぞれについて，$\boldsymbol{A}$ が与えられたときの $\boldsymbol{M}$ の各要素の条件付き独立性，$\boldsymbol{B}$ が与えられたときの $\boldsymbol{A}$ の各要素の条件付き独立性をともに仮定すると，

$$p(\boldsymbol{B}, \boldsymbol{A}, \boldsymbol{M}) = \left\{ \prod_{i=1}^{d} p(M_i|\boldsymbol{A}) \right\} \left\{ \prod_{j=1}^{d} p(A_j|\boldsymbol{B}) \right\} p(\boldsymbol{B}) \tag{4.7}$$

となる．右辺の各確率分布の次元は高々 $O(d)$ にまで落ちている．これはごく少量の学習サンプルを用意するだけでよいことを意味する．このように，尤度に対する条件付き独立性を仮定したモデルは一般に Naïve Bayes モデルと呼ばれる．この条件付き独立性の仮定は非常に強いもので一般には成立しないにもかかわらず，多くの研究においてその有効性が示されている[2] [100]．よって，本研究でもこの Naïve Bayes モデルを積極的に利用している．

以上の特性 1〜5 については，ベイジアンネットワーク (BN: Bayesian Networks)（例えば文献 [74] 参照）が持つ特性である．この BN の有効性は人間行動のモデリングにおいても既に示されている（たとえば [115]）．動的ベイジアンネットワーク (DBN: Dynamic BN) は，この BN に特性 6，すなわち，時間を跨ぐ要素間の因果関係（特性 6）を明示的に導入したものである．これについては本節の少し後で説明する．なお，動的ベイジアンネットワークについては文献 [63] に詳しく説明されている．

Naïve Bayes モデルを取り入れた模擬的な行動・動作の階層構造を BN を用いて表現すると，図 4.1 のようになる[3]．この構造は一見複雑に見えるが，図 4.2 に示す 2

[2] この理由は，局所的な従属性による誤識別への影響が，単純に蓄積されるわけではなく，全体としてみると相殺されるものが含まれるため，単純なモデルの割に高い識別性能が得られるということが証明により明らかにされている [100].

[3] 同一階層間の因果関係が存在しない点が Naïve Bayes モデルでの尤度の条件付き独立性の仮定に相当する．

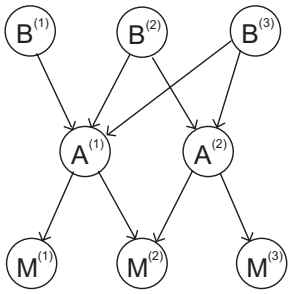


図 4.1: Behavior，Action および Movement の因果関係を表すベイジアンネットワークの例

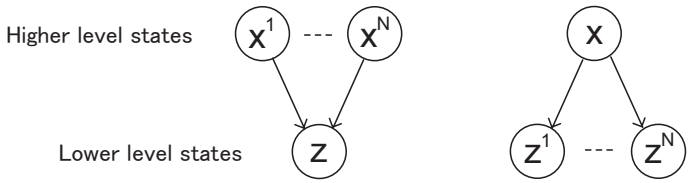


図 4.2: 階層的構造を持つ動作認識のための基本モデル構造

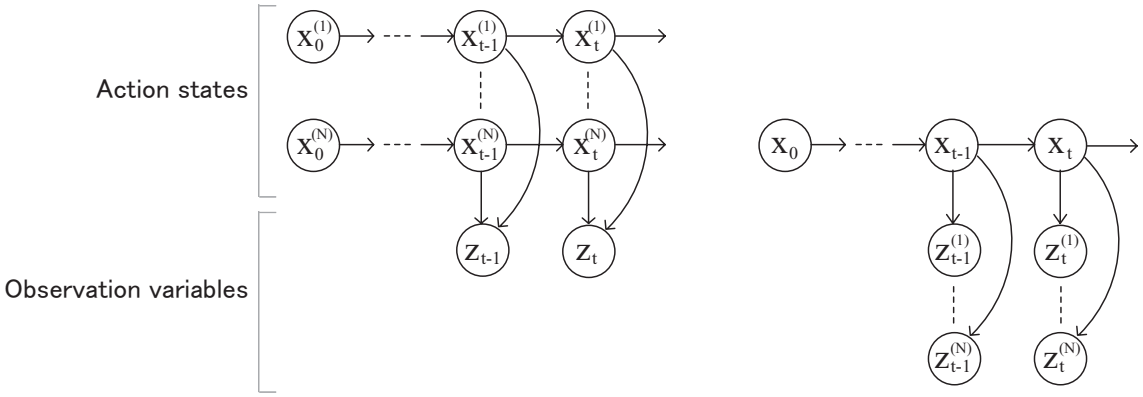


図 4.3: 観測と動作の関係を表す動的ベイジアンネットワーク.

種類の局所構造の組み合わせにより構成されていることが分かる．図 4.2 左のタイプは，1.2 節にて述べた様々な情報が統合された観測データからの複数動作の認識手法，より具体的には 2 章にて説明した単眼動画像にもとづく表情および頭部姿勢の同時推定に相当する．一方，図 4.2 右のタイプは，1.2 節にて述べたコンテクストに依存する動作の認識手法，より具体的には 3 章にて説明したコンテクスト情報を用

いた運転時の注視動作の識別に当たる．このようなシンプルな基本モデルにて個別に構築していけば，それらを組み合わせることにより心的状態，動作および観測間の様々な隣接レベル間の関係を含む最終的な複雑なモデルを構築することが可能である．

図 4.2 において，上位レベルの要因に対して Markov 過程と呼ばれる時間方向の遷移（ここでは 1 次）$p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})$ を導入すると図 4.3 に示す DBN となる．この時間遷移のモデルは時間方向に対して推定を安定化させる働きがある．本論文では，これらの DBN を用いて対象動作を認識している．具体的には，図 4.3 左のモデルを動画像にもとづく表情および頭部姿勢の同時推定に用いている．このとき，変数 X が表情および頭部姿勢，変数 Z が動画像となっている．また，図 4.3 右のモデルをマルチモーダルな観測変数にもとづくドライバの脇見動作の認識に用いている．このときは，変数 X がドライバの注視状態，変数 Z はそれに関係する様々な観測変数となっている．

謝辞

佐藤洋一准教授には，博士課程より入学した私を暖かく迎えてくださり，無事に修了するまで熱心にご指導頂きましたことに心より感謝申し上げます．佐藤准教授より学んだことは本当に数知れません．特に，その御指導のおかげで受賞することができたACCV'07で味わった興奮と感激は今でも私の脳裏に強く焼き付いており，研究を進めるための力の源となっております．大変有意義な研究生活を送らせて頂きましたことへの申し上げる御礼の言葉が尽きません．

また，私のこの博士課程への入学について親身に様々な助言をくださった修士課程時代の恩師である藤野陽三教授に深く感謝致します．

池内克史教授および石塚満教授には，アドバイザ教員として研究の早い段階から，手法の完成度を高めるための多くの貴重な御意見を頂戴しましたことここで御礼申し上げます．須田義大教授および鈴木高宏准教授には，脇見推定に関する多くの御指導を頂きましたこと御礼申し上げます．須田研究室の山口大助特任助教には，実験面をはじめとしてお世話になりましたことに感謝致します．

また，日本電信電話株式会社の前田英作様，大和淳司様，大塚和弘様には，平成18年夏より実習生，また表情認識に関して共同研究として真にお世話になりました．特に大塚様には，理論部分以外にも学会論文の校正等多大な御指導を頂きましたことここに深く御礼申し上げます．国立情報学研究所の杉本晃宏教授には，計測機器などに関し御助力・助言を頂きましたこと御礼申し上げます．また，佐藤いまり准教授には，いつも暖かく接して頂き，研究者としてのみならず人として行うべき多くの事柄を教えて頂きましたこと深く感謝致します．

研究室では，岡部孝弘助教には鋭い御指摘を頂き，研究に深みを与えて頂きましたこと御礼申し上げます．また，卒業生である堀口研一氏には，共同研究者として注視動作識別に関する議論および実験など様々な点においてお世話になりました．同じく卒業生である木谷クリス真実氏や，博士課程の菅野裕介氏や杉村大輔氏には，特に理論の部分で熱い議論を交わさせて頂きました．また，秘書の鈴木咲恵さんおよ

び今川洋子さんには，出張を始めとして様々な場面で大変お世話になりました．苦楽を共にした研究室メンバー，皆様に心より感謝致します．

勤めていた会社を辞めて学生に戻った私を暖かく見守ってくれた両親に感謝します．最後に，仕事と家事を両立させながらも，いつも私を精神的に支えてくれた妻に心より感謝します．

平成 20 年 12 月

熊野 史朗

参考文献

[1] Jrgen Ahlberg. An active model for facial feature tracking. *EURASIP Journal on Applied Signal processing*, Vol. 2002, pp. 566–571, 2002.

[2] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, Vol. 1, No. 6, pp. 22–35, 2006.

[3] S. Basu, I. Essa, and A. Pentland. Motion regularization for model-based head tracking. *Proc. of the International Conference on Pattern Recognition*, pp. 611–616, 1996.

[4] A.E. Beaton and J.W. Tukey. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, Vol. 16, No. 2, pp. 147–185, 1974.

[5] L.M. Bergasa, J. Nuevo, M.A. Sotelo, R. Barea, and M.E. Lopez. ”Real-time system for monitoring driver vigilance”. *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 7, No. 1, pp. 63–77, March 2006.

[6] M.J. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, Vol. Vol. 25, pp. 23–48, 1997.

[7] A. Bobick. Movement, activity and action: The role of knowledge in the perception of motion. *Philosophical Trans. Royal Soc. London*, Vol. 352, pp. 1257–1265, 1997.

[8] L. Breiman. ”Random Forests”. *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.

[9] Oliver Carsten and K. Brookhuis. Issues arising from the haste experiments. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 8, No. 2, pp. 191–196, 2005.

[10] Marco La Cascia, Stan Sclaroff, and Vassilis Athitsos. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 4, pp. 322–336, 2000.

[11] C. Chang and R. Ansari. Kernel particle filter for visual tracking. *IEEE Signal Process. Lett*, Vol. 12, pp. 242–245, 2004.

[12] Ya Chang, Changbo Hu, Rogerio Feris, and Matthew Turk. Manifold based analysis of facial expression. *Image and Vision Computing*, Vol. 24, No. 6, pp. 605–614, 2006.

[13] Shinko Y. Cheng, Sangho Park, and Mohan M. Trivedi. Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis. *Computer Vision and Image Understanding*, Vol. 106, No. 2-3, pp. 245–257, 2007.

[14] Shinko Yuanhsien Cheng and Mohan Manubhai Trivedi. Turn-intent analysis using body pose for intelligent driver assistance. *IEEE Pervasive Computing*, Vol. 5, No. 4, pp. 28–37, 2006.

[15] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S. Chen, and Thomas S. Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, Vol. 91, No. 1-2, pp. 160–187, 2003.

[16] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, Vol. 61, pp. 38–59, 1995.

[17] T. D'Orazio, M. Leo, C. Guaragnella, and A. Distante. A visual approach for driver inattention detection. *Pattern Recognition*, Vol. 40, No. 8, pp. 2341–2355, 2007.

[18] Fadi Dornaika and Franck Davoine. Simultaneous facial action tracking and expression recognition in the presence of head motion. *International Journal of Computer Vision*, Vol. 76, No. 3, pp. 257–281, 2008.

[19] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J-C. Burgelman. Scenarios for ambient intelligence in 2010 final report. *ISTAG*, 2001.

[20] S. Duncan. Toward a grammar for dyadic conversation. *Semiotica*, Vol. 9, pp. 29–46, 1973.

[21] P. Ekman and W. 1975. Friesen. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expresssions*. NJ: Prentice Hall., 1975.

[22] Paul Ekman and Wallace V. Friesen. *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 1978.

[23] Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. *FACS Investigator's Guide. A Human Face*. A Human Face, 2002.

[24] Land M F and Lee D N. Where we look when we steer. *Nature*, Vol. 369, pp. 742–744, 1994.

[25] B. Fasel and J. Luttin. Automatic facial expression analysis: Survey. *Pattern Recognition*, Vol. 36, pp. 259–275, 2003.

[26] Beat Fasel, Florent Monay, and Daniel Gatica-Perez. Latent semantic analysis of facial action codes for automatic facial expression recognition. *In Proceedings of the ACM SIGMM international workshop on Multimedia information retrieval*, pp. 181–188, 2004.

[27] L. Fletcher, G. Loy, N. Barnes, and A. Zelinsky. Correlating driver gaze with the road scene for driver assistance systems. *Robotics and Autonomous Systems*, Vol. 52, pp. 71–84, 2005.

[28] Dieter Fox. Kld-sampling: Adaptive particle filters. *In Advances in Neural Information Processing Systems 14*, pp. 713–720, 2001.

[29] Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. *In Proc. International Conference on Machine Learning*, pp. 148–156, 1996.

[30] S. Geman and D. E. McClure. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, Vol. LII, pp. 5–21, 1987.

[31] Salih Burak Gokturk, Carlo Tomasi, Bernd Girod, and JY Bouguet. Model-based face tracking for view-independent facial expression recognition. *In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 287–293, 2002.

[32] Ralph Gross, Iain Matthews, and Simon Baker. Generic vs. person specific Active Appearance Models. *Image and Vision Computing*, Vol. 23, No. 11, pp. 1080–1093, 2005.

[33] H. Gu and Q. Ji. ”An automated face reader for fatigue detection”. *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pp. 111–116, 17-19 May 2004.

[34] H. Gu, Y. Zhang, and Q. Ji. ”Task oriented facial behavior recognition with selective sensing”. *Computer Vision and Image Understanding*, Vol. 100, No. 3, pp. 385–415, 2005.

[35] K. Hayashi, Y. Kojima, K. Abe, and K. Oguri. ”Prediction of stopping maneuver considering driver’s state”. *Intelligent Transportation Systems Conference, 2006. ITSC ’06. IEEE*, pp. 1191–1196, 2006.

[36] JA Healey and RW Picard. ”Detecting Stress During Real-World Driving Tasks Using Physiological Sensors”. *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 6, No. 2, pp. 156–166, June 2005.

[37] Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemi Aoyagi, Yoshinari Shirai, Naomi Yamashita, and Junji Yamato. The t-room: Toward the future phone. *NTT Technical Review*, Vol. 4, No. 12, pp. 26–33, 2006.

[38] Yuxiao Hu, Zhihong Zeng, Lijun Yin, Xiaozhou Wei, Xi Zhou, and Thomas S. Huang. Multi-view facial expression recognition. *In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.

[39] Chung-Lin Huang and Yu-Ming Huang. Facial expression recognition using model-based feature extraction and action parameters classification. *Journal of Visual Communication and Image Representation*, Vol. 8, No. 3, pp. 278–290, 1997.

[40] Y. Inuzuka, Y. Osumi, and H. Shinkai. ”Visibility of Head Up Display (HUD) for Automobiles”. In *Proceedings of the Human Factors Society 35th Annual Meeting*, pp. 1574–1578, 1991.

[41] Michael Isard and Andrew Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, Vol. 29, No. 1, pp. 5–28, 1998.

[42] Qiang Ji, P. Lan, and C. Looney. A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, Vol. 36, No. 5, pp. 862–875, 2006.

[43] R.E. Kaliouby and P. Robinson. Generalization of a vision-based computational model of mind-reading. *Proc. of the First International Conference on Affective Computing and Intelligent Interaction*, pp. 582–589, 2005.

[44] Takeo Kanade, Jeffrey Cohn, and Ying-Li Tian. Comprehensive database for facial expression analysis. *In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46–53, March 2000.

[45] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, Vol. 1, No. 4, pp. 321–331, 1988.

[46] Irene Kotsia and Ioannis Pitas. Facial expression recognition in image sequences using geometric deformation features and Support Vector Machines. *IEEE Transactions on Image Processing*, Vol. 16, No. 1, pp. 172–187, 2007.

[47] V. Krüger, D. Kragic, A. Ude, and C. Geib. The meaning of action: A review on action recognition and mapping. *Advanced Robotics*, Vol. 21, No. 13, pp. 1473–1501, 2007.

[48] N. Kuge, T. Yamamura, and O. Shimoyama. A driver behavior recognition method based on a driver model framework. *Soc. Automotive Engineers Publication*, Vol. 109, No. 6, pp. 469–476, 2000.

[49] T. KUMAGAI and M. AKAMATSU. ”Prediction of Human Driving Behavior Using Dynamic Bayesian Networks”. *IEICE TRANSACTIONS on Information and Systems*, Vol. E89-D, No. 2, pp. 857–860, 2006.

[50] Matti Kutila, Maria Jokela, Gustav Markkula, and Maria Romera Rue. Driver distraction detection with a camera vision system. *International Conference on Image Processing*, pp. 201–204, 2007.

[51] P. Lan, Q. Ji, and C.G. Looney. ”Information fusion with Bayesian networks for monitoring human fatigue”. *Information Fusion, 2002. Proceedings of the Fifth International Conference on*, Vol. 1, pp. 535–542, 2002.

[52] Andreas Lanitis, Chris J. Taylor, and Timothy F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 743–756, 1997.

[53] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain. Simultaneous feature selection and clustering using mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, pp. 1154–1166, 2004.

[54] Wei-Kai Liao and Isaac Cohen. Belief propagation driven method for facial gestures recognition in presence of occlusions. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop*, pp. 158–163, 2006.

[55] Gwen Littlewort, Marian Stewart Bartlett, Ian R. Fasel, Joshua Susskind, and Javier R. Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, Vol. 24, No. 6, pp. 615–625, 2006.

[56] Gareth Loy and Alexander Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 8, pp. 959–973, 2003.

[57] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. *Proc. of the 7th International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

[58] Simon Lucey, Iain Matthews, Changbo Hu, Zara Ambadar, Fernando Torre, and Jeffrey Cohn. AAM derived face representations for robust facial action recognition. *In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 155–160, 2006.

[59] A. Nijholt M. Pantic, A. Pentland and T. Huang. Human computing and machine understanding of human behavior: A survey. In Proc. of ACM Int ’l Conf. Multimodal Interfaces (ICMI ’06), pp. 239–248, 2006.

[60] H.M. Mandalia, D. Salvucci, College of Arts, Sciences, and Drexel University. *”Pattern Recognition Techniques to Infer Driver Intentions”*. PhD thesis, Drexel University, 2004.

[61] JC Mccall and MM Trivedi. ”Human Behavior Based Predictive Brake Assistance”. *Intelligent Vehicles Symposium, 2006 IEEE*, pp. 8–12, 2006.

[62] J.C. McCall, D.P. Wipf, M.M. Trivedi, and B.D. Rao. Lane change intent analysis using robust operators and sparse bayesian learning. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 3, pp. 431–440, 2007.

[63] Kevin P. Murphy. Dynamic bayesian networks: Representation, inference and learning. *PhD thesis, University of California, Berkeley*, 2002.

[64] Erik Murphy-Chutorian and Mohan M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (To be appeared)*, 2008.

[65] H. H. Nagel. From image sequences towards conceptual descriptions. *Image Vision Computing*, Vol. 6, No. 2, pp. 59–74, 1988.

[66] Kenji Oka and Yoichi Sato. Real-time modeling of face deformation for 3D head pose estimation. *In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 308–320, 2005.

[67] N. Oliver and AP Pentland. Graphical models for driver behavior recognition in a smartcar. *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, pp. 7–12, 2000.

[68] Kazuhiro Otsuka, Hiroshi Sawada, and Junji Yamato. Automatic inference of cross-modal nonverbal interactions in multiparty conversations: ”who responds to whom, when, and how?” from gaze, head gestures, and utterances. *In Proceedings of the International Conference on Multimodal Interfaces*, pp. 255–262, 2007.

[69] M. Pantic and M.S. Bartlett. *Machine Analysis of Facial Expressions*, pp. 377–416. I-Tech Education and Publishing, 2007.

[70] Maja Pantic and Leon J. M. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, pp. 1424–1445, 2000.

[71] A. Pentland and A. Liu. ”Modeling and Prediction of Human Behavior”, 1999.

[72] Michael K. Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, Vol. 94, No. 446, pp. 590–599, 1999.

[73] Yong Rui and Yunqiang Chen. Better proposal distributions: Object tracking using unscented particle filter. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 786–793, 2001.

[74] Stuart Russell and Peter Norvig. *Artificial Intelligence – A Modern Approach*. Pearson Education, 2003.

[75] Dario D. Salvucci, Hiren M. Mandalia, Nobuyuki Kuge, and Tomohiro Yamamura. Lane-change detection using a computational driver model. *Human Factors*, Vol. 49, No. 3, pp. 532–542, 2007.

[76] D.D. Salvucci. ”INFERRING DRIVER INTENT:A CASE STUDY IN LANE-CHANGE DETECTION”. In *Proceedings of the Human Factors Ergonomics Society 48th Annual Meeting*, pp. 2228–2231, 2004.

[77] Nicu Sebe, Michael S. Lew, Yafei Sun, Ira Cohen, Theo Gevers, and Thomas S. Huang. Authentic facial expression analysis. *Image and Vision Computing*, Vol. 25, No. 12, pp. 1856–1863, 2007.

[78] Paul Smith, Student Member, Mubarak Shah, and Niels Da Vitoria Lobo. Determining driver visual attention with one camera. *IEEE Trans. on Intelligent Transportation Systems*, Vol. 4, p. 2003, 2003.

[79] Mu-Chun Su, Chao-Yueh Hsiung, and De-Yuan Huang. A simple approach to implementing a system for monitoring driver inattention. *In Proc. of IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Vol. 1, pp. 429–433, 2006.

[80] Y. Suda, Y. Takahashi, M. Kuwahara, S. Tanaka, K. Ikeuchi, M. Kagesawa, T. Shraishi, M. Onuki, K. Honda, and M. Kano. Development of universal driving simulator with interactive traffic environment. *Intelligent Vehicles Symposium*, pp. 743–746, 2005.

[81] Josephine Sullivan and Jens Rittscher. Guiding random particles by deterministic search. *In Proceedings of the IEEE International Conference on Computer Vision*, Vol. 1, pp. 323–330, 2001.

[82] Hao Tang and Thomas S. Huang. 3D facial expression recognition based on properties of line segments connecting facial feature points. *In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.

[83] T.F.Cootes, G.J. Edwards, and C.J.Taylor. Active appearance models. *In Proc. of European Conference on Computer Vision*, Vol. 2, pp. 484–498, 1998.

[84] Ying L Tian, Takeo Kanade, and Jeffrey F. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 97–115, 2001.

[85] Ying-Li Tian, Takeo Kanade, and Jeffrey Cohn. *Facial expression analysis*. Springer, 2005.

[86] Michael E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, Vol. 1, pp. 211–244, 2001.

[87] Yan Tong, Wenhui Liao, and Qiang Ji. Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 10, pp. 1683–1699, 2007.

[88] K. Torkkola, N. Massey, and C. Wood. Driver inattention detection through intelligent analysis of readily available sensors. *IEEE Conf. Intelligent Transportation Systems*, pp. 326–331, 2004.

[89] Rudolph van der Merwe, Nando de Freitas, Arnaud Doucet, and Eric Wan. The unscented particle filter. *Advances in Neural Information Processing Systems 13*, pp. –, 2001.

[90] Harini Veeraraghavan, Nathaniel Bird, Stefan Atev, and Nikolaos Papanikolopoulos. Classifiers for driver activity monitoring. *Transportation Research Part C*, Vol. 15, No. 1, pp. 51–67, 2007.

[91] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 511–518, 2001.

[92] Jun Wang, Lijun Yin, Xiaozhou Wei, and Yi Sun. 3D facial expression recognition based on primitive surface feature distribution. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1399–1406, 2006.

[93] M Weiser. The computer for the twenty-first century. *Scientific American*, Vol. 265, No. 3, pp. 94–104, 1991.

[94] Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade. Real-time combined 2D+3D Active Appearance Models. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 535 – 542, June 2004.

[95] Jing Xiao, Tsuyoshi Moriyama, Takeo Kanade, and Jeffrey Cohn. Robust full-motion recovery of head by dynamic templates and re-registration techniques. *International Journal of Imaging Systems and Technology*, Vol. 13, pp. 85–94, September 2003.

[96] Peng Yang, Qingshan Liu, Xinyi Cui, and Dimitris N. Metaxas. Facial expression recognition based on dynamic binary patterns. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.

[97] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3d dynamic facial expression database. *The 8th International Conference on Automatic Face and Gesture Recognition (FGR08)*, 2008.

[98] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. A survey of affect recognition methods: audio, visual and spontaneous expressions. *In Proc. of the International Conference on Multimodal Interfaces*, pp. 126–133, 2007.

[99] Bo Zhang, Weifeng Tian, and Zhihua Jin. Head tracking based on the integration of two different particle filters. *Measurement Science and Technology*, Vol. 17, No. 7, pp. 2877–2883, November 2006.

[100] Harry Zhang. Exploring conditions for the optimality of naïve bayes. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 19, No. 2, pp. 183–198, 2005.

[101] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, pp. 915–928, 2007.

[102] Z. Zhu and Q Ji. Real time and non-intrusive driver fatigue monitoring. *In Proc. of the International IEEE Conference on Intelligent Transportation Systems*, pp. 657–662, 2004.

[103] Z. Zhu and Q. Ji. ”Real time and non-intrusive driver fatigue monitoring”. *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, pp. 657–662, 3-6 Oct. 2004.

[104] Zhiwei Zhu and Qiang Ji. Robust real-time face pose and facial expression recovery. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 681–688, 2006.

[105] 大坊郁夫. しぐさのコミュニケーション. サイエンス社, 1998.

[106] 岡兼司, 菅野裕介, 佐藤洋一. 頭部変形モデルの自動構築を伴う実時間頭部姿勢推定. 情報処理学会論文誌：コンピュータビジョンとイメージメディア, Vol. 47, No. SIG10(CVIM15), pp. 185–194, 2006.

[107] 沓名守通, 井東道昌, 山本修身, 中野倫明, 山本新. インパネ位置撮像システムによる顔向き検出と運転支援システムへの応用の試み. Technical report, IEICE PRMU, 2006.

[108] 松原康晴, 尺長健. 疎テンプレートマッチングとその実時間物体追跡への応用. 情報処理学会論文誌 コンピュータビジョンとイメージメディア, Vol. 46, No. SIG9, pp. 17–40, 2005.

[109] 須田義大, 高橋良至, 大貫正明. 研究用ユニバーサルドライビングシミュレータ. 自動車技術, Vol. 59, No. 7, pp. 83–88, 2005.

[110] 前田英作, 南泰浩, 堂坂浩二. 妖精・妖怪の復権 : 新しい「環境知能」像の提案. 情報処理, Vol. 47, No. 6, pp. 624–640, 2006.

[111] 中越聡, 木村賢治, 金森等. ドライバーの顔向きによる前方不注視の推定と警報反応時間の研究. 自動車技術会 学術講演会前刷集, No. 58-06, pp. 17–20, 2006.

[112] 鳥山将司, 井東道昌, 小塚一宏, 中野倫明, 山本新. 画像処理によるドライバの視線推定検出と脇見検知への適用. 自動車技術会 学術講演会前刷集, No. 10-05, pp. 13–16, 2005.

[113] 田久保宣晃. 交通事故データによる運転者のヒューマンエラーと心的負荷の一考察. *IATSS. Review*, Vol. 30, No. 3, pp. 23–32, 2005.

[114] 田久保宣晃, 藤岡健彦. 運転中の脇見行動に関する分析. 自動車技術会論文集, Vol. 34, No. 2, pp. 107–112, 2003.

[115] 本村陽一, 西田佳史. ベイジアンネットワークによるヒューマンモデリング. 人工知能学会誌, Vol. 22, No. 3, pp. 320–327, 2007.

[116] 鈴木正裕, 稲垣伸吉, 鈴木達也, 早川総一郎, 土田縫夫. 視線情報とベイズ推定による運転行動意図の推定. 電気学会 産業計測制御研究会, IIC-07-75, pp. 29–34, March 2007.

発表文献

論文誌

[1] S. Kumano, K. Otsuka, J. Yamato, E. Maeda and Y. Sato, “Pose-Invariant Facial Expression Recognition Using Variable-Intensity Templates”, *International Journal of Computer Vision*, 2008 (Online First).

[2] 熊野史朗, 大塚和弘, 大和淳司, 前田英作, 佐藤洋一, 「変動輝度テンプレートによる頭部姿勢と表情の同時推定」, 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol. 1, No. 2, pp. 50-62,2008.

[3] 熊野史朗, 大塚和弘, 大和淳司, 前田英作, 佐藤洋一, 「変動輝度テンプレートを用いた頭部姿勢変動に頑健な確率的表情認識手法」, 情報科学技術レターズ, pp.215-218, 2007.

国際会議

[4] S. Kumano, K. Otsuka, J. Yamato, E. Maeda and Y. Sato, “Combining Stochastic and Deterministic Search for Pose-Invariant Facial Expression Recognition”, *British Machine Vision Conference (BMVC)*, 2008.

[5] S. Kumano, K. Otsuka, J. Yamato, E. Maeda and Y. Sato, “Pose-Invariant Facial Expression Recognition Using Variable-Intensity Templates”, *Asian Conference on Computer Visoin (ACCV)*, Vol. I, pp.324-334, 2007. （Honorable Mention Award 受賞）

国内会議・研究報告

[6] 熊野史朗, 大塚和弘, 大和淳司, 前田英作, 佐藤洋一, 「パーティクルフィルタと勾配法の組み合わせによる頭部姿勢変動に頑健な表情認識手法」, 画像の認識・理解シンポジウム (MIRU2008), IS4-28, 2008.

[7] 堀口研一, 熊野史朗, 山口大助, 佐藤洋一, 須田義大, 鈴木高宏, 「運転状況を考慮した脇見推定手法」, ITS シンポジウム, O6-1, 2007.

[8] 堀口研一, 熊野史朗, 山口大助, 佐藤洋一, 須田義大, 鈴木高宏, 「ドライバの頭部姿勢及び自車情報を用いた脇見状態推定手法」, 自動車技術会 2007 年秋季大会学術講演会前刷集, no.110-07, pp.1-6, 2007.

[9] 熊野史朗, 大塚和弘, 大和淳司, 前田英作, 佐藤洋一, 「変動輝度テンプレートを用いた頭部姿勢変動に頑健な表情認識手法」, 画像の認識・理解シンポジウム (MIRU2007), IS4-20, 2007.

[10] 熊野史朗, 大塚和弘, 大和淳司, 前田英作, 佐藤洋一, 「表情認識のための変動輝度テンプレートとその頭部姿勢変動に対する頑健性の一検討」, 情報処理学会コンピュータビジョンとイメージメディア研究報告, pp.145-152, 2007.

発明

[11] Toshiyuki Namba, Hiroaki Sekiyama, Keisuke Okamoto, Yoshihiro Oe, Yoichi Sato, Yoshihiro Suda, Takahiro Suzuki, Daisuke Yamaguchi, Shiro Kumano, Kenichi Horiguchi, “Inattentive State Determination Device and Method Of Determining Inattentive State”, PCT/IB2008/002739, 2008.

[12] 難波利行, 関山博昭, 岡本圭介, 大栄義博, 佐藤洋一, 須田義大, 鈴木高宏, 山口大助, 熊野史朗, 堀口研一, 「脇見状態判定装置」, 特願 2007-269455, 2007.

[13] 大塚和弘, 大和淳司, 澤木美奈子, 前田英作, 佐藤洋一, 熊野史朗, 「テンプレート作成装置及び表情認識装置並びにその方法, プログラム及び記録媒体」, 特願 2007-284029, 2007.