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　50–62 (July 2008)

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

熊野史朗[†1]　大塚和弘[†2]　大和淳司[†2]
前田英作[†2]　佐藤洋一[†1]

本論文では，単眼動画像に基づいた，人物の頭部姿勢変動に頑健な表情認識手法を提案する．複雑な顔モデルを用いる従来の手法には，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，本論文では，その問題の解決を目指し，その場で簡単に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．本手法は，変動輝度テンプレートを用い，パーティクルフィルタの枠組みで頭部姿勢と表情を同時に推定する．実験を行ったところ，カメラ正面に対して水平方向 ±40 [deg] の範囲の頭部姿勢において，90%程度の高い表情認識率が得られた．

Simultaneous Estimation of Facial Pose and Expression Using Variable-intensity Template

Shiro Kumano,[†1] Kazuhiro Otsuka,[†2]
Junji Yamato,[†2] Eisaku Maeda[†2] and Yoichi Sato[†1]

In this paper, we propose a method for pose-invariant facial expression recognition from monocular video sequences. Existing methods require stereo cameras or the preparation of a large amount of learning data to generate an accurate face model. With the aim of solving these problems, we propose a novel face model, called the variable-intensity template, which can be generated with very little time and effort. The variable-intensity template consists of a face shape model, a set of discrete interest points defined in the vicinity of facial parts, and an intensity model which describes how the intensity of interest points vary with different facial expressions. By using this model in the framework of a particle filter, our method is capable of estimating facial poses and expressions simultaneously. Experiments demonstrate the effectiveness of our method. A recognition rate of about 90% was achieved for horizontal facial orientations over the range of ±40 degrees from the frontal view.

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[1),5),6),13)]が，対象人物は必ずしもつねに正面を向いているとは限らない．たとえば，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[21)]．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．なかでも，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[8),9),15),18),19)]．ここでは，形状モデルおよび表情変化のモデルをあわ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[9),19)]．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[8),15),18)]．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うえ，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

†1 東京大学生産技術研究所
Institute of Industrial Science, The University of Tokyo
†2 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

して低いといった問題がある[10]．

我々は，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こと

(2)　頭部姿勢変動に対する高い頑健性を有すること

(3)　対象人物に特化した顔モデルを容易に作成できること

の 3 つである．これらを実現するため，我々は，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．

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[4] や楕円体[3] といった単純なパラメトリック形状など，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．我々は，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，表情変化にともなう顔部品の移動により注目点の輝度が変化する．我々はこの性質に着目し，あらかじ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．本論文では，この注目点のマージンサイズ，算出位置のずれ量，および，表情変化にともなう顔部品の移動量の関係を考察し，頭部姿勢変動に対して頑健に表情認識を行うための条件を導出する．

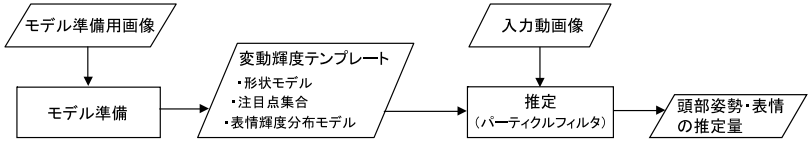


図 1　システムフロー
Fig. 1　System flow chart.

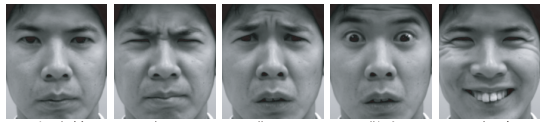


図 2　モデル準備用画像の一例
Fig. 2　Example of images for model preparation.

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の枠組みで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，粒子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．

本手法は，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，ただちにその人物に特化した変動輝度テンプレートを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．

以下，最初に，2 章において本表情認識手法について説明し，続いて，3 章で実験方法およびその結果を示す．最後に 4 章において本論文のまとめを述べる．

2. 提 案 手 法

本手法は，**図 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の枠組みで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，認識対象とする表情それぞれにつき 1 枚の正面顔画像（モデル準備用画像と呼ぶ）である．モデル準備用画像の一例を**図 2** に示す．

2.1　変動輝度テンプレート

変動輝度テンプレート $\mathcal{M}$ は以下の 3 つの要素から構成される．

$$\mathcal{M} = \{\mathcal{S}, \mathcal{P}, \mathcal{L}\}\,. \tag{1}$$

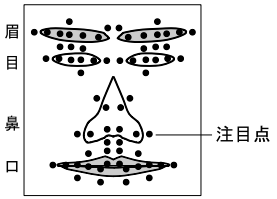


図 3　注目点集合 $\mathcal{P}$ の一例
Fig. 3　Example of a set of interest points $\mathcal{P}$.

ここで，$\mathcal{S}$，$\mathcal{P}$，および，$\mathcal{L}$ は，形状モデル，注目点集合，および，表情輝度分布モデルをそれぞれ示す．以下，これらについて順に述べる．

形状モデル $\mathcal{S}$　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[4] や楕円体[3] といった単純なパラメトリック形状など，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，円柱モデル，および，特定人物 1 名について計測した顔形状を各ユーザに対してスケーリングした形状モデルを用いる（これらの作成方法は 3.1 節で述べる）．

注目点集合 $\mathcal{P}$　注目点集合は，無表情の正面顔画像において，顔部品の近傍に複数の注目点を離散的に配置したものである．**図 3** にその一例を示す．このような注目点を用いることで，表情変化にともな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，

$$\mathcal{P} = \{\boldsymbol{p}_1, \cdots, \boldsymbol{p}_N\} \tag{2}$$

と表す．ここで，$\boldsymbol{p}_i$ は注目点 i の無表情のモデル準備用画像における画像座標を，N は注目点の数を表す．具体的な注目点の選択基準および抽出方法については 2.2 節で述べる．

表情輝度分布モデル $\mathcal{L}$　表情輝度分布モデルは，**図 4** に示すように，表情変化による各注目点の輝度変化をモデル化したものである．ここでは，各注目点の輝度がそれぞれ独立な分布に従うことを仮定し，表情輝度分布モデル $\mathcal{L}$ を，

$$\mathcal{L} = \{L_{1,0}, \cdots, L_{i,e}, \cdots, L_{N,N_e-1}\} \tag{3}$$

と表す．ここで，$e \in \{0, \cdots, N_e - 1\}$ は表情状態（$e = 0$ を無表情とする）を，N_e は対象とする表情の数を表し，$L_{i,e}$ は注目点 i の表情 e についての輝度分布モデル（図 4 右の各々の峰）を表す．この輝度分布モデル $L_{i,e}$ として任意の分布が使用可能であるが，本論文では正規分布と一様分布を組み合わせた分布を用いる（詳細は 2.3 節で述べる）．本手法では，正規分布で表現される輝度分布が表情間で大きく離れていることを利用して，入力画像中の

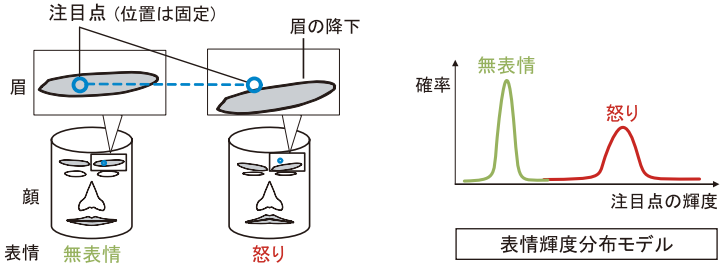


図 4　表情輝度分布モデル $\mathcal{L}$．注目点の輝度は表情変化により大きく変化する
Fig. 4　Intensity distribution model $\mathcal{L}$: The intensity distributions of interest points change strongly with facial expressions.

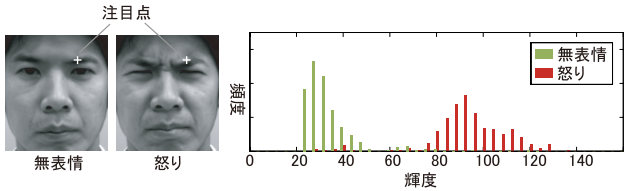


図 5　注目点において観測される輝度ヒストグラムの一例
Fig. 5　Example of intensity histogram observed at an interest point.

注目点集合の輝度からそのときの表情を推定する．さらに，注目点の大きな算出位置のずれ，鏡面反射および撮像ノイズなどによる輝度のばらつきを一様分布で表現することにより，これらの外れ値に対する頑健性を保持する．

図 5 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[⋆1] を示す．このヒストグラムでは，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．これらの輝度のばらつきは，前者については，照明変動（頭部姿勢変動を起因とするものを含む）や若干の注目点の算出位置のずれによるものであり，後者については，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

⋆1 図 5 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，両表情の頻度は正規化されている．

あると考えられる．本論文では，前者を正規分布，後者を一様分布で近似している．

2.2　注目点の選択基準および抽出方法

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で定義するため，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 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[⋆1]，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，

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 つの条件をとも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 6** (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 6 (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）で表す）を設けることで，条件 I および条件 II をとも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越えている（図 6 (b)），すなわち，

$$\textbf{条件 I} \qquad d_i \leq l_{i,e} \tag{4}$$

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）で表す．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 6 (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越えていれば条件 II が成立する（図 6 (d)）．以上をまとめると，条件 II は，

$$\textbf{条件 II} \qquad -k_i \leq d_i \leq l_{i,e} - k_i \tag{5}$$

と書き表される．

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，それらのすべ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，各注目点について見てみると，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 をともに満たすことが要求される[⋆2]．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

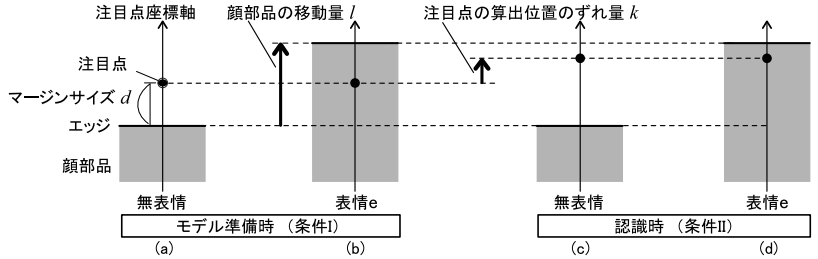


図 6　無表情およびその他の表情 e の条件 I および条件 II における，注目点のマージンサイズ d，算出位置のずれ量 k および顔部品の移動量 l の関係

Fig. 6　The relationship between margin size, d, shift length of calculated position of interest point, k, and displacement of facial part, l under conditions I and II.

⋆1 2.1 節で述べたとおり，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．これらすべ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（3.2 節）．

⋆2 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うる．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．

る．さらに，対象とするすべての表情を識別できるためには，それらすべての表情に対して以上のことがいえる必要がある．

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（付録 A.1 で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とることができない．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（具体的には次項を参照）．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，3.4 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[14] 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 7) 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．

注目点の抽出方法　まず，無表情の正面顔画像に対して，顔検出器を用いて（本論文では Adaboost を用いたカスケード型識別器[17] を使用）検出された顔領域内において，各顔部品領域の水平および垂直方向の境界を以下のように自動的に抽出する．眉，目および口の垂直境界，および，眉・目の水平境界（**図 7** 右の a-f）については，顔領域内の行あるいは列の画素の平均輝度が極大あるいは極小となる位置から決定する[2]．鼻および口の水平境界（図 7 右の g，h）については，顔領域の水平方向の中心（図 7 右の i）から顔領域の幅に定数を乗じた距離だけ離れた位置とする．

次いで，各顔部品領域内において注目点を抽出する．前項での議論のとおり，1 つの注目点から検出可能な顔部品の移動方向は，注目点座標軸の正方向のみである．そのため，図 7 中央のように注目点を顔部品の内外に複数配置することで，様々な方向の顔部品の移動を検出可能とする．ここでは，顔部品のエッジの近傍に配置されるこのような注目点の一種として，松原ら[20] の境界ダイポール（ここでは点対と呼ぶ）を用いる．この点対とは，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．この点対の抽出方法は，縦，横，斜め ×2 の 4 方向について，対中心

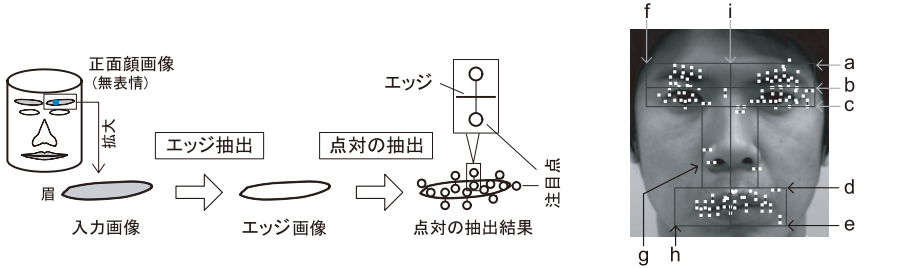


図 7　注目点の抽出方法（左）と自動検出された注目点集合（右）．右図において，矩形の枠は各顔部品領域を，白色の 2 点とそれを結ぶ線分の組は 1 つの点対を表す

Fig. 7　Left: The method to extract interest points P. Right: The extraction result. Small white rectangles represent interest points; point pairs are indicated by the lines. The set of large rectangles represents the regions holding the interest points.

間の距離が閾値以上離れるよう，点対候補のうち点対内の 2 点の輝度差が大きいものから順に，抽出数が上限を超えない限りにおいて可能な限り抽出するというものである[20]．エッジについては，本論文では，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として検出した．注目点のマージンサイズについては，顔部品ごと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した[⋆1]．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．図 7 右に，自動抽出された顔部品領域および注目点集合を示す．このときの顔領域のサイズは 170×192 [pixel]，抽出された注目点の数は 126 であった．

2.3　頭部姿勢および表情の同時推定法

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 8** に示す動的ベイジアンネットワークで表現される．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，2.2 節で述べたとおり各表情を表す離散変数である．なお，4 章で述べるとおり，本論文での提案手法では，大きな上下方向の頭部姿勢変動に対応できない場合がある．

⋆1 実装のうえでは，これを最近接の整数へと丸めている．

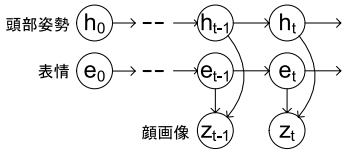


図 8　頭部姿勢，表情および顔画像の動的ベイジアンネットワーク
Fig. 8　Dynamic Bayesian Network of head poses, facial expressions and face images.

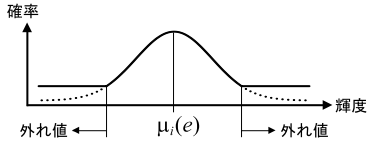


図 9　輝度分布モデル $L_{i,e}$ の分布形状
Fig. 9　Shape of distribution of the intensity distribution model $\mathcal{L}_{i,e}$.

この動的ベイジアンネットワークでは，現在の時刻 t までのすべての顔画像 $\boldsymbol{z}_{1:t}$ が与えられたときの頭部姿勢 $\boldsymbol{h}_t$ および表情 e_t の同時事後確率密度分布 $p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t})$ の逐次的な推定式は，

$$p(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t}) = \alpha\, p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t) \int p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}) \sum_{e_{t-1}} P(e_t|e_{t-1})\, p(\boldsymbol{h}_{t-1}, e_{t-1}|\boldsymbol{z}_{1:t-1})\, d\boldsymbol{h}_{t-1} \tag{6}$$

と表される（この展開方法については，たとえば文献 16) を参照）．ここで，$\alpha = 1/p(\boldsymbol{z}_t)$，$p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t)$ は頭部姿勢 $\boldsymbol{h}_t$ かつ表情 e_t である状態に対する観測画像 $\boldsymbol{z}_t$ の尤度，$p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$ 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．頭部姿勢の状態遷移モデル $p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$ は，各変数がそれぞれ独立なランダムウォークモデルとする．表情の状態遷移モデル $P(e_t|e_{t-1})$ については，すべての表情間の遷移が等しい確率とする．

一般に，式 (6) 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，厳密に解くことはできない．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パーティクルフィルタ[11] を用いて推定する．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（粒子）と呼ばれる重み付きのサンプル集合 $\{\boldsymbol{h}_t^{(l)}, e_t^{(l)}, \omega_t^{(l)}\}_{l=1}^{N_x}$ によって近似的に表現される．ここで，N_x は粒子数を，$\boldsymbol{h}_t^{(l)}$，$e_t^{(l)}$ および $\omega_t^{(l)}$ はそれぞれ時刻 t における l 番目の粒子の保持する頭部姿勢，表情および重みを表す．重み $\omega_t^{(l)}$ は，次項で述べる顔画像の尤度 $p(\boldsymbol{z}_t|\boldsymbol{h}_t^{(l)}, e_t^{(l)})$ に比例するとともに，$\sum_l \omega_t^{(l)} = 1$ を満たす．なお，本手法は，リサンプリング処理（たとえば文献 19)），すなわち，同じ時刻において 2 回の粒子の生成・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．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 11) を参照されたい．

顔画像の尤度　頭部姿勢 $\boldsymbol{h}_t$ かつ表情 e_t での顔画像 $\boldsymbol{z}_t$ の尤度 $p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t)$ は，2.1 節で仮定した各注目点の輝度の独立性を用いて，

$$p(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t) = \prod_i^N p(z_{i,t}|\boldsymbol{h}_t, e_t) \tag{7}$$

と展開される．ここで，$z_{i,t}$ は頭部姿勢 $\boldsymbol{h}_t$ のときに算出される注目点 i の画像座標における画像 $\boldsymbol{z}_t$ の輝度であり，$p(z_{i,t}|\boldsymbol{h}_t, e_t)$ は注目点 i の表情 e_t についての輝度分布モデル L_{i,e_t} に対するこの輝度 $z_{i,t}$ の尤度を表す．

本論文では，2.1 節で述べたとおり，正規分布と一様分布を組み合わせた分布（**図 9**）を輝度分布モデルとして用いる．このとき，輝度 $z_{i,t}$ の尤度 $p(z_{i,t}|\boldsymbol{h}_t, e_t)$ は，正規分布の尤度に外れ値の影響を軽減する一種のロバスト推定の枠組みを取り入れた尤度としてとらえられ，

$$p(z_{i,t}|\boldsymbol{h}_t, e_t) = \frac{1}{\sqrt{2\pi}\sigma_i(e_t)} \exp\left[-\frac{1}{2}\rho\left(d_m\right)\right], \tag{8}$$

$$d_m = \frac{z_{i,t} - \mu_i(e_t)}{\sigma_i(e_t)}, \tag{9}$$

$$\rho(x) = \begin{cases} x^2, & \text{if } x^2 < \epsilon \\ \epsilon, & \text{otherwise} \end{cases} \tag{10}$$

と書き表される．ここで，$\mu_i(e_t)$ および $\sigma_i(e_t)$ は，注目点 i の表情 e_t についての輝度分布モデル $L_{i,e}$ における正規分布項の平均および標準偏差である．また，本論文における $\rho(\cdot)$ は刈り込みを行うロバスト関数（$\epsilon = 9$ とした）である．なお，その他のロバスト関数を用いることも可能である．また，法線ベクトルがカメラ方向を向いていない注目点，すなわち，自己遮蔽されている注目点については，その輝度の値にかかわらず外れ値と見なす．

頭部姿勢 $\boldsymbol{h}_t$ における注目点 i の画像座標 $\boldsymbol{q}_{i,t}$ は，弱中心投影のカメラモデルを用い，

$$\boldsymbol{q}_{i,t} = s_t \boldsymbol{R}_{2\times 3,t} \boldsymbol{p}_i^{(\mathrm{shape})} + \boldsymbol{t}_t, \tag{11}$$

$$\boldsymbol{p}_i^{(\mathrm{shape})} = f(\mathcal{S}, \boldsymbol{p}_i), \tag{12}$$

と算出される．ここで，$\boldsymbol{p}_i^{(\mathrm{shape})}$ は注目点 i の形状モデル $\mathcal{S}$ 上での 3 次元座標である．これは，無表情のモデル準備用画像において検出された顔領域の中心（2.2 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 のモデル準備用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 $f(\mathcal{S}, \boldsymbol{p}_i)$ により求まる．$\boldsymbol{R}_{2\times 3}$ 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．

頭部姿勢および表情の推定量　各時刻において，頭部姿勢の推定量 $\widehat{\boldsymbol{h}}_t$ 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 の推定量（表情 e_t の表情確率と呼ぶ）については，表情についての周辺事後確率 $P(e_t|z_{1:t})$ とする．そして，表情確率が最大となる表情を推定表情 $\widehat{e}_t$ とする．これらの頭部姿勢の推定量 $\widehat{\boldsymbol{h}}_t$ および推定表情 $\widehat{e}_t$ は，それぞれ，

$$\widehat{\boldsymbol{h}}_t = \sum_l \omega_t^{(l)} \boldsymbol{h}_t^{(l)} \tag{13}$$

$$\widehat{e}_t = \arg\max_{e_t} \sum_{l\in\{e_t^{(l)}=e_t\}} \omega_t^{(l)} \tag{14}$$

と書き表される．

3. 実　　験

提案手法の有効性を確認するため，本論文では，被験者がカメラに対して左右様々な方向に首を振ったまま表情のみを変化させるテスト（テスト 1．3.2 節），および，頭部姿勢および表情を同時に変化させるテスト（テスト 2．3.3 節）という 2 種類のテストの動画像を撮影し，そのデータに対しオフラインで本手法を適用する実験を行った．認識対象とする表情は，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 種類とした．テスト 1 については被験者 9 名（20 代〜40 代の男性 7 名および女性 2 名）が 1 回ずつ，テスト 2 については被験者 1 名（テスト 1 の被験者のうちの 20 代の男性 1 名）が 1 回参加した．テスト 1 は，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．本論文では，対象とする首振り角を，カメラ方向を基準として水平方向 0 [deg]（カメラに正対），±20 [deg] および ±40 [deg] の計 5 方向とした．また，テスト 1 には，複数の人物に対して提案手法が有効であるかどうかを検証するという意図も含まれる．さらに，テスト 1 では，近似精度の異なる 2 つの形状モデルを用い，認識率にどのような違いが生じるのかということも検証する．使用した 2 つの形状モデルは，その場で容易に作成可能な円柱モデル，および，より近似精度が高いことが予想される，実際に計測した 1 名の被験者の顔形状（計測顔形状と呼ぶ）を被験者ごとに簡単にスケーリングしたもの（計測形状モデルと呼ぶ）である．一方，テスト 2 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．また，3.4 節では，経験的に設定したマージンサイズ（2.2 節）の妥当性を検証する．

テスト 1 およびテスト 2 ともに，頭部姿勢のうち変化しているのは主に首振り角のみであるが，どちらの場合も頭部姿勢については 6 変数すべて（2.3 節参照）を推定した．粒子数は 1500 とした．IEEE1394 カメラを用いて 15 [fps] で撮影した 512 × 384 [pixel] のグレースケールの動画像を用いた．このときの処理時間は，Intel Pentium D 3.73 GHz プロセッサ，3.0 GB メモリの PC で，入出力処理を除きおよそ 80 [msec/frame] であった．なお，本論文では表情の認識精度の検証のためオフラインで処理を行ったが，我々はオンラインで動作するシステムも構築しており，リアルタイム（10 [fps] 程度）で動作することも確認済みである．

3.1　モデル準備

円柱モデル，注目点集合，および，表情輝度分布モデルについては，モデル準備用画像から作成した．このモデル準備用画像については，被験者が各々その場で，カメラに正対して頭部を固定したまま認識対象の 5 表情を順に表出したときに撮影された顔画像とした．円柱モデルの作成方法は，無表情のモデル準備用画像において顔検出器により検出された顔領域の幅に，あらかじめ定めた定数（本論文では 2.46 とした）を乗じた値を半径とするものとした．注目点集合の抽出方法については 2.2 節で述べた．表情輝度分布モデルの作成方法については以下のとおりとした．各輝度分布モデル $L_{i,e}$ に対し，表情 e のモデル準備用画像において，注目点 i が定義された画像座標 $\boldsymbol{p}_i$ での輝度を平均 $\mu_i(e)$ とした．標準偏差 $\sigma_i(e)$ については，平均値 $\mu_i(e)$ に比例するものとした[⋆1]．なお，本論文ではその比例定数

⋆1 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ともない大きくなるという考えによる．

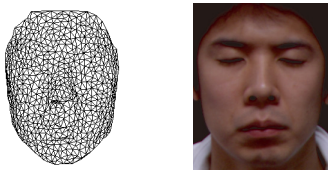


図 10　計測顔形状（左）と計測時のテクスチャ画像（右）
Fig. 10　Left: Measured shape. Right: Texture Image.

を経験的に 1/3 とした．

計測顔形状は，被験者のうちの 1 名に対し，3 次元デジタイザ（KONICA MINOLTA 社製 VIVID910）を用いて作成した．この計測顔形状を**図 10** に示す．各被験者についての計測形状モデルは，この計測顔形状を，対象被験者の無表情のモデル準備用画像へ以下の方法によりフィッティングしたものとした．まず，対象被験者のモデル準備用画像の顔中心とこの計測顔形状の顔中心の位置を合わせた．次いで，この両者の顔領域の幅と目中心—口中心間の高さを合わせるという縦方向および横方向のスケーリングを行った．奥行き方向のスケーリング係数は，縦方向および横方向のスケーリング係数の積の平方根とした．

3.2　表情のみを変化させるテスト（テスト 1）の内容および結果

テスト 1 では，各被験者に対し，各対象方向（5 つの水平方向のうちのいずれか 1 つ）について，まず顔をカメラに正対させ，次いで，対象方向に首を振り，その後はその頭部姿勢を保ったまま，PC のモニタ上に文字として自動的に映し出される 5 表情を順次表出するよう指示した．各被験者は，上記の手順を 5 方向について繰り返した．モニタ上の各表情の指示の表示時間，および，それらの指示の間隔は，ともに 60 [frame] とした．表情認識率については，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とした．表情の正解ラベルについては，そのときモニタ上で指示されている表情とした．なお，各表情が指示された直後からの 20 [frame] は，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，認識率の算出対象から除外した．

円柱モデルを用いた場合　**図 11** に，各被験者のテスト 1 の動画像に対し，円柱モデルを用いたときの推定結果の一例を示す．この図では，すべての被験者に対して画像中の表情が正しく認識されていることが分かる．このときの表情認識率を首振り角ごとに算出した結果を**表 1** に示す．ここでの認識率は，各被験者の 1 回ずつのテスト，すなわち，計 9 回分のテストの平均である．首振り角が 0 [deg] から離れるにつれおおむね認識率が低下しているものの，首振り角 ±40 [deg] においても平均 80.9 [%] という良好な認識率が得られている．

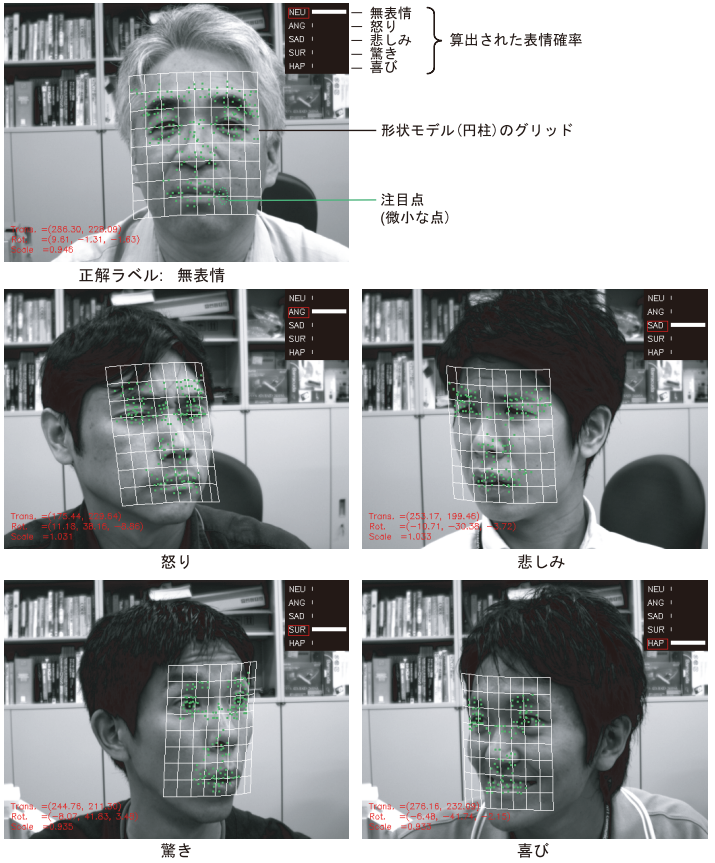


図 11　各被験者の頭部姿勢および表情の推定結果の一例（テスト 1）
Fig. 11　Example of estimation results of facial poses and expressions in Test 1.

これらの結果から，本手法により，計測形状モデルのようなより近似精度の高いモデルを準備しなくても，1 枚の画像から容易に作成可能な円柱モデルを使用することで高い認識率が得られることが複数の人物に対して確認された．

計測形状モデルを用いた場合　テスト 1 の動画像に対して計測形状モデルを用いて推定を行い，円柱の場合と同様にして表情認識率を算出した結果を表 1 に示す．円柱モデルを使

表 1　9 名の被験者の表情の平均認識率（テスト 1）
Table 1　Average recognition rates of facial expressions of nine subjects in Test 1.

首振り角 [deg]		全平均	−40	−20	0	20	40
認識率 [%]	円柱	87.6	83.1	91.9	94.8	89.6	78.6
	計測形状モデル	91.8	87.8	93.6	96.8	93.4	87.2

表 2　表情の混同行列（テスト 1）．行の表情は正解表情を，列の表情は認識結果を，数値は認識率をそれぞれ表す
Table 2　Confusion matrix of facial expressions in Test 1: The expressions in the rows are the ground truths and the expressions in the columns describe the recognition results. The numerals denote average recognition rates.

	無表情	怒り	悲しみ	驚き	喜び
無表情	86.9	0.6	10.8	0.1	1.6
怒り	3.9	90.7	4.7	0.0	0.7
悲しみ	0.9	7.8	83.4	6.6	1.3
驚き	0.2	0.0	0.0	99.8	0.0
喜び	1.7	0.1	0.2	0.1	97.9

単位：[%]

用した場合に比べて認識率が全体で 4.2 [%] 向上し，全体で 91.8 [%] という非常に高い認識率が得られた．また，このときの表情についての混同行列を**表 2** に示す．悲しみ表情と無表情や怒り表情との間の混同率が特に高いことから，多くの被験者においてこれらの表情間の類似度が他の表情間の類似度に比べて特に高かったことが示唆される．

ここでは，円柱形状よりも近似精度が高いことが予想される計測形状モデルとして，1 名の被験者の計測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，円柱モデルを用いた場合よりも大きく向上した非常に高い認識率が得られたことから，このような計測形状モデルが本手法で用いる形状モデルとして十分有用であることが示唆された．なお，1 名の計測顔形状の代わりに平均顔形状を用いることで，多くの人物に対してより近似精度の高い計測形状モデルを作成可能となることが推測される．

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（2.2 節）に関しては，表 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 3.4 節で行う）．

3.3　頭部姿勢および表情をともに変化させるテスト（テスト 2）の内容および結果

テスト 2 では，被験者は首を左右に大きく振りながら各表情を任意のタイミングで表出した．また，形状モデルには円柱モデルを使用し，表情については正解ラベルを目視により付与した．このときの入力動画像，および，各フレームにおける表情確率および頭部姿勢の推定結果を，**図 12** の (a)，(b) および (c) にそれぞれ示す．図 12 (b) では各フレームにおいて表情が正しく認識されていることが見てとれる．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．また，図 12 (c) では，頭部姿勢角の推定結果が，3 回の首振り動作を検出するのに十分な精度を有していることが見てとれる．

このテスト 2 の結果より，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は，テスト 1 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ことから，テスト 2 においても同様に正しい推定が可能であると予想される．また，形状モデルについては，すでに円柱モデルを用いて精度の高い推定結果が得られているため，より高い認識率を与える計測形状モデルを用いたとしても同様の結果となることが予想される．

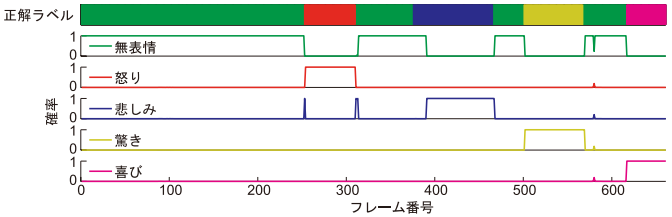
以上の実験により，複数の人物に対して，単眼動画像を用いて，その場で各人物に特化したモデルを作成し，頭部方向変化が生じる状況においても，表情および頭部姿勢を頑健に推定できることが示唆された．

3.4　注目点のマージンサイズの妥当性の検証

本節では，2.2 節で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．対象人物は，注目点の算出位置のずれ量を求めるために必要な実際の顔形状が得られている図 10 に示す被験者とした．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．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，式 (11) および式 (12) において，形状モデルとして計測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，モデル準備用画像から目視で読み取った．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

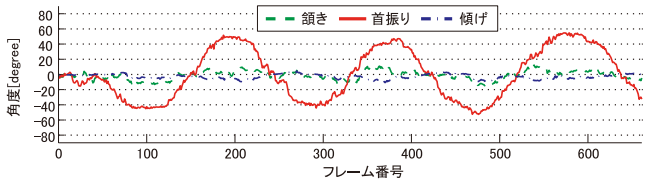
(a) 入力動画像（左から右へフレーム番号 100，290，400，560，660）

(a) Input video sequence (from left to right, frame number 100, 290, 400, 560, 660).



(b) 表情の正解ラベル（最上段）および各表情確率（2 段目以下）．(横軸：フレーム番号)

(b) Ground truth (top) and probabilities of facial expressions (others): The probability of correct expression is remarkably higher than that of other expressions.



(c) 頭部姿勢角の推定結果．(横軸：フレーム番号（(b) の横軸と一致))

(c) Estimation results of facial pose (horizontal axis same as that of (b)): Facial poses (red solid line) are estimated with enough accuracy to detect three head shake cycles.

図 12　入力動画像および推定結果（テスト 2）

Fig. 12　Images and estimation results in Test 2.

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付録 A.1 参照）．**表 3** 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 3 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．

表 3　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^* の平均値

Table 3　Average of maximum shift length of calculated position, $k_{i,\max}$, maximum displacement of facial part, $l_{i,\max}$ and best margin size d_i^*.

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]

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，1 名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．

4. むすび

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．本手法は，従来法の煩雑なモデル構築のための事前準備を不要とし，その場での簡単なモデル準備のみで，ただちに表情認識を実行可能とする．また，首を大きく左右に振った場合でも頑健に表情が認識できることを実験により確認した．

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．

提案手法は，顔面上の輝度自体が大きく変化する事態に対処する機構を備えていない．このため，注目点の大きな輝度変化を生じる激しい照明変動，あるいは，ユーザの上方に照明が存在する環境における頷き動作（上下方向の頭部姿勢変動）には対応できない[⋆1]．後者は一般的な室内環境においても発生することから，水平方向および面内方向の回転が提案手法の頭部姿勢角についての主な適用対象となる．しかし，このような顔面上の輝度自体が変化するという問題は，表情輝度分布モデルを逐次的に更新する（たとえば文献 12) が参考になる），観測される輝度をモデルに対して誤差が最小となるように補正することにより解決

⋆1 頷き動作により顔面上の各位置の法線方向と照明方向とのなす角が大きく変化することで，顔面上の輝度も大きく変化するためである．

可能なものと考える．

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．

参 考 文 献

1) Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I. and Movellan, J.: Automatic Recognition of Facial Actions in Spontaneous Expressions, *Journal of Multimedia*, Vol.1, No.6, pp.22–35 (2006).
2) Baskan, S., Bulut, M. and Atalay, V.: Segmentation of Human Face Using Gradient-Based Approach, *Proc. SPIE*, Vol.4301, No. Machine Vision Applications in Industrial Inspection IX, pp.52–63 (2001).
3) Basu, S., Essa, I. and Pentland, A.: Motion Regularization for Model-Based Head Tracking, *Proc. International Conference on Pattern Recognition*, pp.611–616 (1996).
4) Cascia, M.L., Sclaroff, S. and Athitsos, V.: Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.22, No.4, pp.322–336 (2000).
5) Chang, Y., Hu, C., Feris, R. and Turk, M.: Manifold based analysis of facial expression, *Image and Vision Computing*, Vol.24, No.6, pp.605–614 (2006).
6) Cohen, I., Sebe, N., Chen, L., Garg, A. and Huang, T.: Facial expression recognition from video sequences: Temporal and static modeling, *Computer Vision and Image Understanding*, Vol.91, pp.160–187 (2003).
7) Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J.: Active shape models — Their Training and Application, *Computer Vision and Image Understanding*, Vol.61, pp.38–59 (1995).
8) Dornaika, F. and Davoine, F.: Simultaneous facial action tracking and expression recognition using a particle filter, *Proc. 10h IEEE International Conference on Computer Vision*, Vol.2, pp.1733–1738 (2005).
9) Gokturk, S.B., Tomasi, C., Girod, B. and Bouguet, J.: Model-Based Face Tracking for View-Independent Facial Expression Recognition, *Proc. 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pp.287–293 (2002).
10) Gross, R., Matthews, I. and Baker, S.: Generic vs. person specific Active Appearance Models, *Image and Vision Computing*, Vol.23, No.11, pp.1080–1093 (2005).
11) Isard, M. and Blake, A.: Condensation — conditional density propagation for visual tracking, *International Journal of Computer Vision*, Vol.29, No.1, pp.5–28 (1998).
12) Jepson, A.D., Fleet, D.J. and El-Maraghi, T.F.: Robust Online Appearance Models for Visual Tracking, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.10, pp.1296–1311 (2003).
13) Kaliouby, R. and Robinson, P.: Generalization of a Vision-based Computational Model of Mind-Reading, *Proc. 1st International Conference on Affective Computing and Intelligent Interaction*, pp.582–589 (2005).
14) Lucas, B.D. and Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision, *Proc. 7th International Joint Conference on Artificial Intelligence*, pp.674–679 (1981).
15) Lucey, S., Matthews, I., Hu, C., Ambadar, Z., Torre, F. and Cohn, J.: AAM Derived Face Representations for Robust Facial Action Recognition, *Proc. 7th International Conference on Automatic Face and Gesture Recognition*, pp.155–160 (2006).
16) Rich, E. and Knight, K.: *Artificial Intelligence*, 2nd edition, pp.537–583, McGraw-Hill Book Company (1991).
17) Viola, P.A. and Jones, M.J.: Rapid object detection using a boosted cascade of simple features, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.511–518 (2001).
18) Zhu, Z. and Ji, Q.: Robust Real-Time Face Pose and Facial Expression Recovery, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.681–688 (2006).
19) 岡　兼司，菅野裕介，佐藤洋一：頭部変形モデルの自動構築をともな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).
20) 松原康晴，尺長　健：疎テンプレートマッチングに基づく実時間物体追跡，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 (CVIM11),

pp.60–71 (2004).
21) 大塚和弘，大和淳司，村瀬　洋：複数人物の対面会話シーンを対象とした画像中の人物頭部追跡に基づく会話構造のモデル化と確率的推論，画像の認識・理解シンポジウム，pp.84–91 (2006).

付　　録

A.1　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，2.2 節で述べた条件 I および条件 II を用いて算出する方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（$d_i \leq l_{i,e}$）を満たしうる $l_{i,e} > 0$ 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 13** 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 13 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（$d_i \leq l_{i,e}$）については，条件 II の上限側の不等式（$d_i \leq l_{i,e} - |k_i(\theta_y)|$）が成立していれば，明らかに成立する．

一方，条件 I を満たしえない（$l_{i,e} \leq 0$）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．3.4 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 をつね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．

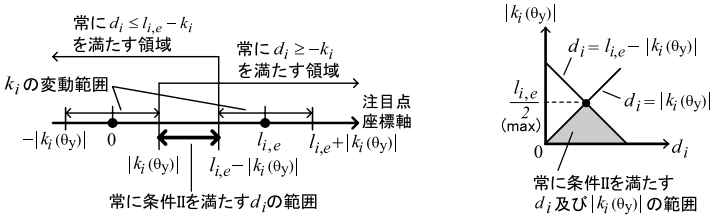


図 13　条件 II を満たすマージンサイズ d_i の範囲．ただし，$l_{i,e} > 0$，かつ，$|k_i(\theta_y)| \leq l_{i,e} - |k_i(\theta_y)|$ とする
Fig. 13　Margin size range satisfying condition II.

以上のような最適なマージンサイズ d_i^* を，数式でまとめると，

$$d_i^* = \begin{cases} l_{i,e}/2, & \text{if } \ l_{i,e} > 0 \\ |k_i(\theta_{y,\max})| & \text{otherwise} \end{cases} \tag{15}$$

となる．

(平成 19 年 9 月 25 日受付)
(平成 20 年 3 月 10 日採録)

(担当編集委員　斎藤 英雄)

熊野　史朗（学生会員）

2004 年東京大学大学院工学系研究科修士課程修了．2005 年同大学院情報理工学系研究科博士課程入学，現在に至る．コンピュータビジョン，ヒューマン・コンピュータ・インタラクションに関する研究に従事．2002 年土木学会年次学術講演会優秀講演者賞，8th Asian Conf. Computer Vision Honorable Mention Award を受賞．自動車技術会会員．

大塚　和弘（正会員）

1993 年横浜国立大学工学部電子情報工学科卒業．1995 年同大学院工学研究科博士課程前期修了．2007 年名古屋大学大学院情報科学研究科博士後期課程修了．1995 年日本電信電話（株）入社．現在，NTT コミュニケーション科学基礎研究所主任研究員．コンピュータビジョン，時系列画像解析，コミュニケーションシーンの分析に興味を持つ．第 55 回全国大会優秀賞，IAPR 10th Int. Conf. Image Analysis and Processing Best Paper Award，ACM 9th Int. Conf. Multimodal Interfaces Outstanding Paper Award ほか各賞受賞．電子情報通信学会，IEEE，各会員．博士（情報科学）．

大和　淳司

1988 年東京大学工学部精密機械工学科卒業．1990 年同大学院工学系研究科精密機械工学専攻修士課程修了．1998 年マサチューセッツ工科大学電気工学およびコンピュータサイエンス学科修士課程修了．1990 年 NTT 入社．画像認識等の研究に従事．NTT ヒューマンインタフェース研究所，MIT 人工知能研究所，NTT 第三部門を経て，現在 NTT コミュニケーション科学基礎研究所メディア情報研究部，メディア認識研究グループ，主幹研究員/グループリーダ．IEEE，ACM，電子情報通信学会各会員．

前田　英作（正会員）

1984 年東京大学理学部卒業．1986 年同大学院理学系研究科修士課程修了．同年日本電信電話（株）入社．現在，NTT コミュニケーション科学基礎研究所人間情報研究部長．工学博士．1995〜1996 年ケンブリッジ大学（英国）客員研究員．主としてパターン認識，統計的機械学習，生物情報処理，環境知能の研究に従事．IEEE，電子情報通信学会，日本バイオインフォマティクス学会各会員．

佐藤　洋一（正会員）

1990 年東京大学工学部機械工学科卒業．1997 年カーネギーメロン大学計算機科学部ロボティクス学科博士課程修了．Ph.D. in Robotics. 同年より東京大学生産技術研究所研究機関研究員，講師，助教授を経て，現在，同大学大学院情報学環准教授．コンピュータビジョン，ヒューマン・コンピュータ・インタラクション，コンピュータグラフィックスに関する研究に従事．2008 年電子情報通信学会論文賞，2006 年電子情報通信学会論文賞，1999 年情報処理学会山下記念研究賞，1999 年日本バーチャルリアリティ学会論文賞等を受賞．電子情報通信学会，日本バーチャルリアリティ学会，ACM，IEEE 各会員．