LH-003

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

Stochastical Facial Expression Recognition Method Using Variable-Intensity Template for Handling Head Pose Variation

熊野 史朗 † 大塚 和弘 ‡ 大和 淳司 § 前田 英作 ‡ 佐藤 洋一 ¶
Shiro Kumano Kazuhiro Otsuka Junji Yamato Eisaku Maeda Yoichi Sato

あらまし　本稿では，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，本稿では，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を用いた手法を提案する．変動輝度テンプレートとは，形状モデルと，離散的な注目点の集合，及び，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である．本手法は，変動輝度テンプレートを用いて，パーティクルフィルタの枠組みにて，頭部姿勢と表情を同時に推定する．実験により横方向の首振り運動に対する本手法の頑健性を確認した．

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いる．従来の表情認識手法は，ほぼ正面を向いた顔画像を対象としたものが多い [1] [2]．しかし，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 [3]．頭部姿勢と表情はそれぞれ独立かつ同時に顔画像に影響を与えるため，このようなシーン中の人物の表情を認識するためには，頭部姿勢も同時に推定しなければならない．

頭部姿勢及び表情を認識するためには，それぞれのモデルを用意する必要がある．従来手法の多くは，頭部姿勢変動を形状モデルの大局的な 3 次元の並進・回転として表現し，表情変化を形状モデルの局所的な変形として表現する．つまり，形状モデルの局所的な変形が表情モデルに相当する．本稿では，形状モデル，及び，表情モデルを併せて顔モデルと呼ぶ．しかし，このアプローチは精緻な顔モデルを必要とする．それは，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及び表情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モデルを獲得できるため，精度の高い表情の推定を行うことが可能である [4]．しかし，このアプローチには，ステレオシステムなどの特殊な装置が必要であり，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，複数の人物の個人間変動を含めた顔形状及び表情変化についての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ものである [5]．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 [6]．

そこで，本稿では，これら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. 人物依存の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，本手法では，頭部姿勢変動をパラメトリックな形状モデルの 3 次元の並進・回転により表現するとともに，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取り，そのための新たな顔モデルを提案する．

この顔モデルは，形状モデル，目や口といった顔部品の周辺に配置した離散的な注目点集合，及び，注目点の輝度分布モデルからなる．特に，本手法は，各注目点の輝度分布を，各表情に依存した輝度分布からなる混合分布にて定義する．そして，この輝度分布モデルを用い，各注目点の輝度からそのときの表情を推定する．本手法は，表情変化を検出するために，無表情時の顔部品のエッジを跨いだ点対として注目点を配置する．さらに，この点対の間にマージンを設けることで，形状モデルの誤差に起因する，算出される注目点の画像上での位置とその点の実際の画像位置とのずれの影響を軽減する [7]．本稿では，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．

本手法は，パーティクルフィルタの枠組みにおいて，変動輝度テンプレートを用いて頭部姿勢及び表情を同時に推定する．ここでは，各パーティクルに頭部姿勢及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，対象人物が単眼システムにおいてその場で各表情を表出した画像から，直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 章において本手法の概要を述べる．次いで，3 章にて実験方法及びその結果を示し，最後

†東京大学, IPSJ
‡日本電信電話 (株), IPSJ, IEICE
§日本電信電話 (株), IEICE
¶東京大学, IPSJ, IEICE

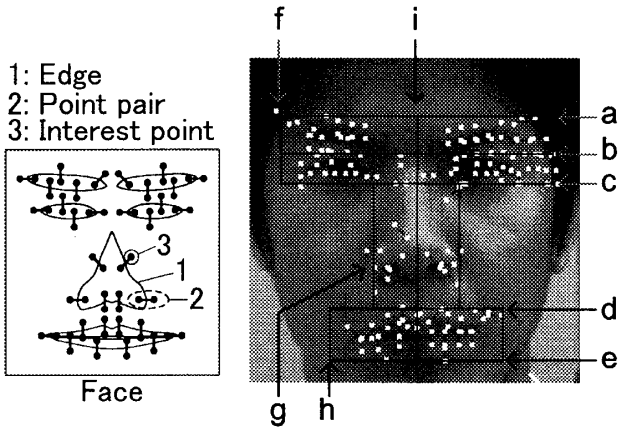


図 1: 注目点配置の概念図，及び，注目点の抽出結果
右図において，矩形の枠は各顔部品領域を，白色の 2 点とそれを結ぶ黒色の線分の組は 1 つの点対を表す．

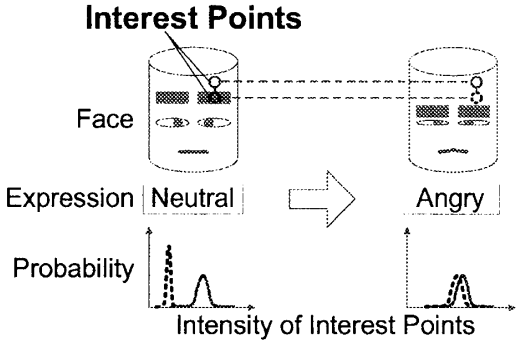


図 2: 注目点の表情輝度分布モデル
図下部の各輝度分布の色及び線種は，図上部の各注目点の色及び線種に対応している．

に 4 章において本稿のまとめを述べる．

2. 提案手法

本手法は，まず，変動輝度テンプレートをその場にて準備し．続いて，その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にて，入力画像における頭部姿勢及び表情を同時に推定する．

2.1 変動輝度テンプレート

変動輝度テンプレートは，形状モデル，注目点集合，及び，表情輝度分布モデルから構成される．

形状モデル． 顔の形状モデルには，おおまかに顔形状を近似する円柱モデルを用いる．この円柱モデルは，正面かつ無表情の顔画像において，文献 [8] の手法を用いて検出された顔領域の幅に，予め定めた定数を乗じた値を半径として，自動的に作成される．

注目点集合． 注目点集合は，図 1 左に示すような，眉や口といった顔部品のエッジを跨ぐ点対により構成される．本手法は，これらの点対を，正面を向いた一枚の無表情の顔画像から抽出する．まず，前述した方法により検出された顔領域内において，各顔部品の領域の水平及び垂直方向の境界を以下のように決定する．眉，目及び口の垂直境界，及び，眉の水平境界（図 1 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 [9]．鼻及び口の水平境界（図 1 右の g,h）については，顔領域の水平方向の中心（図 1 右の i）から顔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，点対を，以下の条件を満たす 4 方向（縦，横，斜め ×2）の点対候補のうち，点対内の 2 点の輝度差が大きいものから順に，抽出した数が一定数に達する，あるいは，条件を満たす候補が存在しなくなるまで抽出する [7]．以下の 3 つがそれらの条件である．(1) エッジを跨いでいる．(2) 中心が顔部品のエッジ上にある．(3) 中心が既に選択された全ての点対中心から規定の距離以上離れている．点対内の 2 点の間隔は，顔領域の幅に経験的に決定した定数を乗じた値とする．図 1 右に，自動抽出された顔部品領域，及び，注目点集合を示す．このときの点対内の 2 点の間隔は 6（縦及び横方向）あるいは $6\sqrt{2}$（斜め方向）[pixel]，顔領域のサイズは 151×170[pixel]，抽出された注目点の数は 158 である．

このような点対を注目点として用いるのは以下の理由による．円柱のような粗い形状モデルを用いた場合，その誤差により，算出される注目点の画像上での位置（方法は次節にて述べる）が，その点の実際の画像位置に対してずれてしまう．このずれは面外方向の頭部姿勢角の増大従って大きくなるため，一般にこれらの姿勢角が大きくなるほど正しい推定が困難となる．本手法のように注目点を輝度変化の激しいエッジから離れたところに定義することが，この問題の効果的な回避策の 1 つとなる．なぜなら，たとえ形状モデルの誤差による注目点の算出位置のずれが発生しても，このような注目点の輝度の変化は抑えられるからである．しかし，このような注目点を用いた手法 [7] でも，大きな輝度変化を引き起こす表情の変化（図 2）が生じた場合には，頭部姿勢を正しく推定することはできない．この問題は，以下の表情輝度分布モデルを用意することで解決される．

表情輝度分布モデル． 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）．ここでは，各注目点の輝度がそれぞれ独立な正規分布に従うことを仮定する．さらに，それらの分布が表情によって変化するものとする．輝度を正規分布にて表現するのは，輝度が形状モデル誤差や撮像系ノイズなどによる複合的な要因による広がりを持つという考えによる．図 2 下の各々の峰が，1 つの注目点についての輝度分布を表している．本手法は，この表情輝度分布モデルを用意することで，表情変化に対する頭部姿勢推定の頑健性を実現するとともに，輝度分布が各表情によって異なることを利用して，入力画像における注目点の輝度のパターンからそのときの表情の推定を行う．

2.2 頭部姿勢及び表情の同時推定法

本手法では，頭部姿勢状態 $\boldsymbol{h}_t$ 及び表情状態 e_t 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，各時刻におけるそれらの状態がそのときの顔画像 $\boldsymbol{z}_t$ に同時に影響を及ぼしているとする．このとき，現在の時刻 t までの顔画像 $\boldsymbol{z}_{1:t}$ が与えられた際の頭部姿勢 $\boldsymbol{h}_t$ 及び表情 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，

$$P(\boldsymbol{h}_t, e_t|\boldsymbol{z}_{1:t}) = \alpha P(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t) \iint P(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}) P(e_t|e_{t-1}) P(\boldsymbol{h}_{t-1}, e_{t-1}|\boldsymbol{z}_{1:t-1}) d\boldsymbol{h}_{t-1} de_{t-1} \quad (1)$$

と表される．ここで，$\alpha = 1/P(\boldsymbol{z}_t)$ であり，$P(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t)$ は以下で述べる頭部姿勢 $\boldsymbol{h}_t$ 及び表情 e_t における画像 $\boldsymbol{z}_t$ の尤度を，$P(\boldsymbol{h}_0)$ 及び $P(e_0)$ は頭部姿勢及び表情の事前分布をそれぞれ表す．また，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，全ての表情間の遷移が等しい確率，すなわち，定常過程とする．なお，(1) 式の左辺から右辺への変換には，ベイズ則，頭部姿勢 $\boldsymbol{h}_t$ 及び表情 e_t の画像 $\boldsymbol{z}_t$ に対する条件付き独立性を用いている．頭部姿勢 $\boldsymbol{h}_t$ は，形状モデル中心の入力画像上での位置，カメラに正対する顔向きを基準とした頷き（ピッチ），首振り（ヨー），傾げ（ロール）に対応する頭部姿勢角，及び，スケールの6連続変数からなる．一方，表情 e_t は各表情を表す離散変数である．

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節にて仮定した各注目点の独立性を用い，

$$P(\boldsymbol{z}_t|\boldsymbol{h}_t, e_t) = \prod_{p \in \boldsymbol{P}} P(z_{p,t}|\boldsymbol{h}_t, e_t) \quad (2)$$

と展開される．ここで，$z_{p,t}$ は画像 $\boldsymbol{z}_t$ の注目点 p での輝度を，$P(z_{p,t}|\boldsymbol{h}_t, e_t)$ は頭部姿勢 $\boldsymbol{h}_t$ における輝度 $z_{p,t}$ の表情 e_t についての表情輝度分布モデルに対する尤度を，$\boldsymbol{P}$ は自己遮蔽されていない（法線ベクトルがカメラ方向を向いている）注目点の番号の集合を表す．ここでは，正規分布に従うことを仮定した輝度 $z_{p,t}$（2.1節）の尤度 $P(z_{p,t}|\boldsymbol{h}_t, e_t)$ の定義を，外れ値の影響を軽減するロバスト推定の枠組みを取り入れ，

$$P(z_{p,t}|\boldsymbol{h}_t, e_t) = \frac{1}{\sqrt{2\pi}\sigma_p(e_t)} \exp\left[-\frac{1}{2}\rho\left(\frac{z_{p,t} - \mu_p(e_t)}{\sigma_p(e_t)}\right)\right], \quad (3)$$

$$\rho(x) = \begin{cases} x^2, & \text{if } x^2 < \epsilon \\ \epsilon, & \text{if } x^2 \geq \epsilon \end{cases} \quad (4)$$

とする．ここで，$\mu_p(e_t)$ 及び $\sigma_p(e_t)$ は，表情輝度分布モデルの注目点 p の表情 e_t での輝度の平均及び標準偏差である．また，ここでは $\epsilon = 9$ とする．輝度 $z_{p,t}$ は，画像 $\boldsymbol{z}_t$ の，以下の3つの処理を行うことにより得られる座標における輝度とする．(1) 画像上で抽出された注目点を形状モデル上へ投影する．(2) 頭部姿勢 $\boldsymbol{h}_t$ に従ってその形状モデルを3次元的に並進・回転する．(3) 形状モデル上の注目点を弱中心投影により入力画像平面に投影する．

パーティクルフィルタによる状態推定． 一般に，(1) 式の頭部姿勢及び表情の状態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．そこで，本手法はこれらをパーティクルと呼ばれる重み付きのサンプル集合で近似的に表現するパーティクルフィルタ [10] を用いて推定する．ここでは，頭部姿勢及び表情の推定量を，各パーティクルの保持するそれぞれの値の加重平均，すなわち，期待値として算出する．

3. 実験

提案手法の有効性を確認するため，様々な頭部姿勢において表情を変化させる被験者の動画像を撮影し，そのデータに対して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，及び，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．なお，恐れ及び嫌悪表情は，表出の困難さや表出頻度を考慮して除外した．ここでは，IEEE1394 カメラを用いて 15[fps] にて撮影した 512×384[pixel] のグレースケールの動画像を用いた．今回の被験者は男性5名であった．なお，パーティクル数は 1500 とした．このときの処理時間は，Intel Pentium D 3.73GHz プロセッサ，3.0GB メモリの PC にて，およそ 80[msec/frame] であった．

3.1 モデル準備

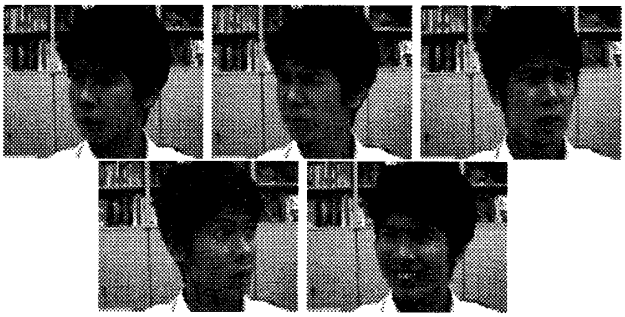
まず，各被験者に顔をカメラに正対させ，その頭部姿勢を固定したまま各表情を順に表出させた顔画像を撮影した．そして，それらの画像から，本手法により変動輝度テンプレートを作成した．まず，無表情の画像から 2.1 節にて述べた方法により形状モデルの作成，及び，注目点の抽出を行った．次いで，各表情の画像から，注目点が定義された各画素の各表情における輝度をモデルの平均 μ とし，さらに，標準偏差については $\sigma = \mu/6$ として表情輝度分布モデルを作成した．

3.2 テスト動画像

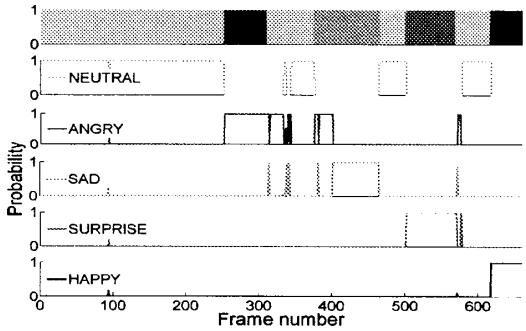
被験者が表情を表出している間に頭部姿勢を変化させないテスト（テスト1），及び，表情を表出している間に頭部姿勢を変化させるテスト（テスト2）の2種類のテストを行い，そのときの動画像を撮影した．テスト1では，頭部姿勢角として，カメラ方向を基準として0度（カメラ正対），±20度及び±40度の計5方向の首振り角を対象とした．被験者に対する指示は，各対象方向でのテストについて，まず，無表情で顔をカメラに正対させ，次いで無表情のまま対象方向に顔を向け，最後にその姿勢を保ったまま，PC のモニタ上に文字として自動的に映し出される表情を順次表出する，というものであった．各被験者は，上記の手順を5方向について繰り返した．モニタ上の各表情の指示の表示時間，及び，それらの指示の間隔は，共に 30[frame] とした．一方，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順に表出した．テスト1は全ての被験者に対して1回ずつ，テスト2は1名の被験者に対して1回行われた．

3.3 表情認識結果

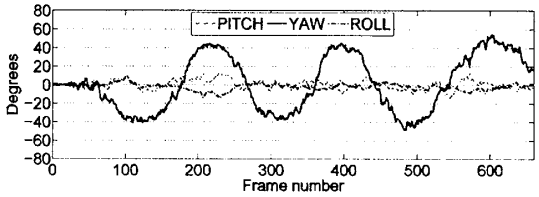
前節の要領で作成したテスト動画像セットに対して，本手法により表情認識を行った．表情認識率については，テスト1の動画像を対象として以下の方法により算出した．各フレームでの表情の正解ラベルは，そのときにモニタで表出を指示していた表情とした．なお，表情の指示開始と実際の表出までの時間差を考慮し，表情が指示された直後から 10[frame] 分の画像は認識率算出の対象から除外した．これらの各フレームについての表情推定結果を正解ラベルと照らし合わせた際の正解率を，表情認識率として算出した．このときの首振り角毎の全ての被験者についての平均の表情認識率を表1に示す．首振り角が大きくなるにつれ認識率が低下しているものの，首振り角± 40 度の平均において



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



(b) 表情（最上段：正解ラベル，2 段以降：推定結果，横軸：フレーム番号，縦軸：確率．正解ラベルと推定結果の色は一致している）



(c) 頭部姿勢角（横軸：(b) と一致，縦軸：角度）

図 3: 入力動画像及び推定結果（テスト 2）

も 84.2[%] という良好な結果が得られた．左右方向で認識率に違いが見られるのは，今回の実験を行った場所が一般的な照明環境を持つ室内であり，カメラに対して正面を向いた場合に対する顔の輝度変化が，負の首振り角の方向よりも正の首振り角の方向においてより大きかったためであると考えられる．

テスト 2 についての入力動画像，及び，各フレームにおける表情確率及び頭部姿勢の推定結果を図 3 に示す．図 3(b) では各フレームにおいて表情がほぼ正しく認識されていることが，図 3(c) では 3 回の首振り動作を検出するのに十分な精度で頭部姿勢が推定されていることが見て取れる．さらに，図 3(b) において，正解表情が他の表情よりも突出して高い確率を持つことからも提案手法の頑健性が示唆される．なお，テスト 2 では，被験者自らが手動にて正解ラベルを付けた．

4. むすび

本稿では，変動輝度テンプレートを用いた頭部姿勢と表情を同時に推定する手法として，パーティクルフィルタ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，直ちに表情認識を実行可能とする．また，実験により，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確認した．

本手法はまだ，大きな頷き動作のような注目点の激しい輝度変化を生じる頭部姿勢変動には対応できていない．今回，首振り動作に対する頑健性を優先したのは，我々が表情の認識を目指す複数人会話の場面では，他の会話参加者への注視に関係する首振り動作が重要な役割を果たすと考えているためである [3]．今後は，この課題への対処として，照明変動に不変な特徴量の使用や表情輝度分布モデルの逐次更新などを検討したい．また，被験者数を増やした統計的な検証も行う予定である．

表 1: 表情の首振り角毎の平均認識率（テスト 1）

Yaw (deg)	-40	-20	0	20	40
Recog. (%)	90.6	97.8	97.5	85.8	77.8

参考文献

[1] I. Cohen, N. Sebe, L. Chen, A. Garg and T. Huang: "Facial expression recognition from video sequences: Temporal and static modeling", Computer Vision and Image Understanding, **91**, pp. 160–187 (2003).

[2] Y. Chang, C. Hu, R. Feris and M. Turk: "Manifold based analysis of facial expression", Image and Vision Computing, **24**, 6, pp. 605–614 (2006).

[3] 大塚, 大和, 村瀬："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).

[4] S. B. Gokturk, C. Tomasi, B. Girod and J.-Y. Bouguet: "Model-based face tracking for view-independent facial expression recognition", Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 287–293 (2002).

[5] F. Dornaika and F. Davoine: "Simultaneous facial action traking and expression recognition using a particle filter", Proc. of the Tenth IEEE International Conference on Computer Vision, **2**, pp. 1733–1738 (2005).

[6] R. Gross, I. Matthews and S. Baker: "Generic vs. person specific active appearance models", Image and Vision Computing, **23**, 11, pp. 1080–1093 (2005).

[7] 松原, 尺長："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, **46**, SIG9 (CVIM11), pp. 60–71 (2004).

[8] P. Viola and M. Jones: "Robust real-time face detection", Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **57(2)**, pp. 137–154 (2004).

[9] S. Baskan, M. Bulut and V. Atalay: "Segmentation of human face using gradient-based approach", Proc. of SPIE, **4301**, Machine Vision Applications in Industrial Inspection IX, pp. 52–63 (2001).

[10] M. Isard and A. Blake: "Condensation – conditional density propagation for visual tracking", International Journal of Computer Vision, **29**, 1, pp. 5–28 (1998).