

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

熊野 史朗[†] 大塚 和弘^{††} 大和 淳司^{††} 前田 英作^{††} 佐藤 洋一[†]

[†] 東京大学生産技術研究所

^{††} NTT コミュニケーション科学基礎研究所

あらまし 人物の頭部姿勢変動を許容する従来の画像に基づいた表情認識手法は、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。そのため、我々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している。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。本稿では、新たに、形状モデルの誤差のある場合における本手法の頭部姿勢変動に対する頑健性を理論的に検討した。

Variable-Intensity Templates for Facial Expression Recognition And Theoretical Study of Its Robustness Against Pose Variation

Shiro KUMANO[†], Kazuhiro OTSUKA^{††}, Junji YAMATO^{††}, Eisaku MAEDA^{††}, and Yoichi SATO[†]

[†] Institute of Industrial Science, The University of Tokyo

^{††} NTT Communication Science Laboratories

Abstract Existing methods for simultaneously handling head pose variations and facial expressions are based on tracking features on a face mesh model. However, these feature-based methods require troublesome preparation steps such as manually selecting feature points or complex face models. To avoid such costly preparation steps for better ease of use, we have proposed a novel method using a variable-intensity template that models face shape and facial expressions. A variable-intensity template consists of a set of discrete interest points on a rigid face shape model, and can model changes in facial expressions using the luminance variation for each point. Our proposed method estimates head poses and facial expressions simultaneously utilizing a particle filter framework. In this paper, we theoretically discussed that our method is pose-invariant despite a shape model error.

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] [3]。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 [4]。頭部姿勢と表情はそれぞれ独立に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、無表情時の顔の形状モデル、及び、表情変化のモデルを用意し、頭部姿勢変動及び表情変化を分離して扱う必要がある。ここでは、形状モデル、及び、表情変化モデルを併せて顔モ

デルと呼ぶ。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並進・3次元回転として表現し、表情変化を形状モデルの局所的な変形として表現する。すなわち、従来手法における表情変化モデルとは形状モデルの局所的な変形のことである。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 [5] [6]。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、複数の人物の顔形状及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 [7] [8]。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 [9]。

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している [10]。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、我々は、パラメトリックな形状モデルを用いて頭部姿勢変動を表現すると共に、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している。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、表情変化に伴い位置が変化する顔部品内において、無表情時でのそ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、粗い形状モデルの実際の顔

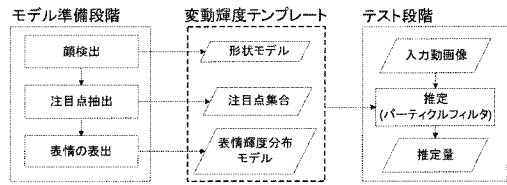


図 1: 本手法のシステムフロー

形状に対する誤差の影響を軽減する [10] [11]。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んでいる。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、粒子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表出した表情の画像から、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。

本稿では、新たに、形状モデルの誤差を含む場合における提案手法の頭部姿勢変動に対する頑健性を、顔部品の移動量、及び、3次元デジタイザを用いて計測した被験者の精緻な顔の形状データを用いて、理論的に検討する。

以下、最初に、2章において本表情認識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、続いて4章で本手法の頭部姿勢変動に対する頑健性の理論的な検討を行う。最後に5章において本稿のまとめを述べる。

2. 提案手法

本手法は、図1に示す通り、モデル準備段階とテスト段階から構成される。モデル準備段階では、変動輝度テンプレートの作成、すなわち、形状モデルの作成、注目点の抽出、及び、表情輝度分布モデルの作成をその場に行う。続いて、テスト段階では、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 M は以下の3つの要素から構成される。

$$M = \{S, P, L\}. \quad (1)$$

ここで、 S 、 P 、及び、 L は、それぞれ、形状モデル、注目点集合、表情輝度分布モデルを示す。以下でこれらに

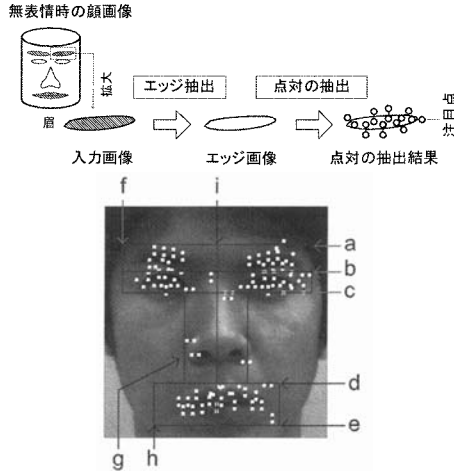


図 2: 注目点の抽出方法 (上) と自動検出された注目点集合 (下) 下図において、矩形の枠は各顔部品領域を、白色の 2 点とそれを結ぶ線分の組は 1 つの点対を表す。

ついて順に述べる。

顔形状モデル S . 本手法は、顔形状モデルとして、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 [12] 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するものとして、自動的に作成される。なお、形状モデル座標系は、検出された顔領域の中心を原点とし、各軸がカメラ座標系の各軸に平行であるような座標系として定義する。

注目点集合 P . 注目点集合は、図 2 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自動的に抽出する。眉、目及び口の垂直境界、及び、眉の水平境界 (図 2 下の 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 [13]。鼻及び口の水平境界 (図 2 下の g,h) については、顔領域の水平方向の中心 (図 2 下の 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 $\times 2$) 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 [11]。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

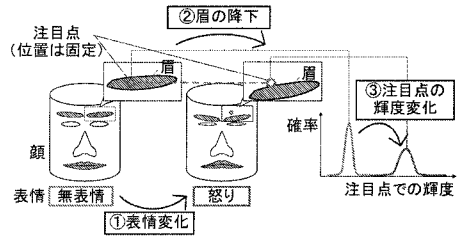


図 3: 注目点の輝度分布モデル

下部の線種及び線の色は上部の各注目点に対応している。

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 2 下に、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 170×192 [pixel]、抽出された注目点の数は 126 である。

このような点対を注目点として用いるのは、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 (詳細は 4. 章参照)。しかし、このような注目点を用いた手法 [11] でも、大きな輝度変化を引き起こす表情の変化 (図 5) 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 L . 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 3)。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 3 右の各々の峰が、1 つの注目点の各表情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 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 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 $z_{1:t}$ が与えられた際の頭部姿勢 h_t 及び表情

e_t の同時事後確率密度分布 $P(\mathbf{h}_t, e_t | \mathbf{z}_{1:t})$ の逐次的な推定式は,

$$P(\mathbf{h}_t, e_t | \mathbf{z}_{1:t}) = \alpha P(\mathbf{z}_t | \mathbf{h}_t, e_t) \int P(\mathbf{h}_t | \mathbf{h}_{t-1}) P(e_t | e_{t-1}) P(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1} \quad (2)$$

と表わされる [10]. ここで, \mathbf{x}_t は \mathbf{h}_t 及び e_t を併せた状態 ($\mathbf{x}_t = [\mathbf{h}_t, e_t]$), $\alpha = 1/P(\mathbf{z}_t)$ であり, $P(\mathbf{x}_0)$ は頭部姿勢及び表情の事前分布である. 頭部姿勢の状態遷移モデル $P(\mathbf{h}_t | \mathbf{h}_{t-1})$ 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. 頭部姿勢 \mathbf{h}_t は, 顔形状モデルの中心の入力画像上での位置 ($t_{x,t}, t_{y,t}$), カメラに正対する顔向きを基準とした傾き (ピッチ), 首振り (ヨー), 傾げ (ロール) に対応する頭部姿勢角 ($r_{x,t}, r_{y,t}, r_{z,t}$), 及び, スケール s_t の 6 連続変数からなる. 一方, 表情状態 e_t は前述した通り各表情を表す離散変数である.

一般に, (2) 式の頭部姿勢及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, 本手法はパーティクルフィルタを用いて頭部姿勢及び表情を同時に推定する [10].

顔画像の尤度. 頭部姿勢 \mathbf{h}_t かつ表情 e_t での顔画像 \mathbf{z}_t の尤度 $P(\mathbf{z}_t | \mathbf{h}_t, e_t)$ は, 2.1 節にて仮定した各注目点の独立性を用いて,

$$P(\mathbf{z}_t | \mathbf{h}_t, e_t) = \prod_{i \in \mathbf{P}} P(z_{i,t} | \mathbf{h}_t, e_t) \quad (3)$$

と展開される. ここで, $z_{i,t}$ は画像 \mathbf{z}_t の注目点 i での輝度を, $P(z_{i,t} | \mathbf{h}_t, e_t)$ は頭部姿勢 \mathbf{h}_t における輝度 $z_{i,t}$ 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, \mathbf{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.

正規分布に従うことを仮定した輝度 $z_{i,t}$ (2.1 節) の尤度 $P(z_{i,t} | \mathbf{h}_t, e_t)$ については, 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,

$$P(z_{i,t} | \mathbf{h}_t, e_t) = \frac{1}{\sqrt{2\pi}\sigma_i(e_t)} \exp \left[-\frac{1}{2} \rho \left(\frac{z_{i,t} - \mu_i(e_t)}{\sigma_i(e_t)} \right) \right], \quad (4)$$

$$\rho(x) = \begin{cases} x^2, & \text{if } x^2 < \epsilon \\ \epsilon, & \text{otherwise} \end{cases} \quad (5)$$

と定義する. ここで, $\mu_i(e_t)$ 及び $\sigma_i(e_t)$ は, 表情輝度分布モデルの注目点 i 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.

注目点の画像座標の算出方法. 頭部姿勢 \mathbf{h}_t における注目点 i の画像座標 $\mathbf{q}_{i,t}$ は, 弱中心投影のカメラモデルを用いて以下のように算出される.

$$\mathbf{q}_{i,t} = s_t \mathbf{R}_{2 \times 3, t} \mathbf{p}_i^{(shape)} + \mathbf{t}_t, \quad (6)$$

$$\mathbf{p}_i^{(shape)} = f(S, \mathbf{p}_i), \quad (7)$$

$$\mathbf{R}_t = \mathbf{R}_x(\theta_{x,t}) \mathbf{R}_y(\theta_{y,t}) \mathbf{R}_z(\theta_{z,t}), \quad (8)$$

$$\mathbf{R}_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix}, \quad (9)$$

$$\mathbf{R}_y(\theta_y) = \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix}, \quad (10)$$

$$\mathbf{R}_z(\theta_z) = \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (11)$$

$$\mathbf{t}_t = [t_{x,t} \ t_{y,t}]^T. \quad (12)$$

ここで, $\mathbf{p}_i^{(shape)}$ は注目点 i の形状モデル S 上での 3 次元座標であり, 関数 $f(S, \mathbf{p}_i)$ は X 座標及び Y 座標が注目点 i の画像座標 \mathbf{p}_i に等しい位置での形状モデル S の 3 次元座標を返す関数であり, $\mathbf{R}_{2 \times 3}$ は回転行列 \mathbf{R} の上 2 行からなる行列である.

頭部姿勢及び表情の推定量. 各時刻における頭部姿勢の推定量, 及び, 表情の推定量 (表情確率と呼ぶ) は, それぞれの分布の期待値として算出する [10]. なお, ここでは事後確率が最大となる表情を推定表情と呼ぶ.

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 $512 \times 384[\text{pixel}]$ のグレースケールの動画画像を用いた. 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. 次に, 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

表 1: 各首振り角における表情認識率の平均 (テスト 1)

首振り角 [deg]	全平均	-40	-20	0	20	40
認識率 [%]	89.9	90.6	97.8	97.5	85.8	77.8

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 60[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。ここでは、テスト 1 を全ての被験者に対して 1 回ずつ、テスト 2 を 1 名の被験者に対して 1 回行った。

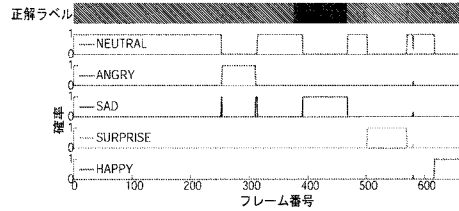
3.3 表情認識結果

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。表情認識率は、テスト 1 の結果を用いて以下の方法により算出された。各フレームでの表情の正解ラベルは、前節の要領にてテスト動画像を作成する際に表出の指示を行った表情とした。なお、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[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。このときの表情認識率を首振り角毎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 5 回のテストの平均である。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 度の平均においても 84.2[%] という良好な結果が得られた。

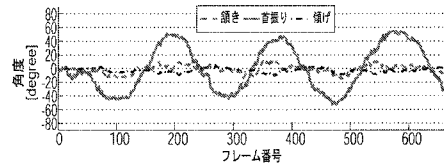
テスト 2 についての入力動画像、及び、各フレームにおける表情確率及び頭部姿勢の推定結果を、それぞれ、図 4 の (a)、(b) 及び (c) に示す。図 4(b) では、各フレームにおいて表情が正しく認識されていることを見て取れる。さらに、正解表情が他の表情よりも突出して高い表



(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) 表情の正解ラベル (最上段) 及び表情推定結果 (それ以外) (横軸: フレーム番号, 縦軸: 各表情確率)



(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), 縦軸: 頭部姿勢角 [deg])

図 4: テスト 2 での入力動画像及び推定結果

(b) では、正解表情の確率が他の表情に比べて顕著に高いことが読み取れる。(c) では、頭部姿勢角の推定値が 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取れる。

情確率を持つことから提案手法の頑健性が示唆される。また、図 4(c) では、頭部姿勢角の推定値が、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を見て取れる。

4. 本手法の頭部姿勢に対する頑健性の理論的検討

入力画像中の各注目点の輝度は、形状モデルの誤差、頭部姿勢変動による顔への照明方向の変化も含めた照明変動、頭部姿勢推定誤差、撮像系ノイズなどによる影響を受ける。このため、注目点の輝度を用いて頑健な推定を行うためには、これらの要因に対処する必要がある。以下では、これらの中で最も大きな影響を及ぼすと考えられる形状モデルの誤差に対する、本手法の頑健性の理論

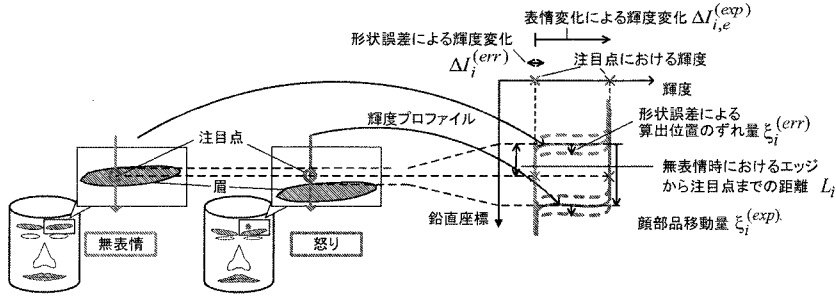


図 5: 誤差及び表情変化時の顔部品の移動による注目点の輝度変化の概念図

的な検討を行う。なお、ここでは、注目点集合として図 2 に示したものをを用いる。

本手法は、注目点の輝度の変化から表情変化を検出するものである。このため、無表情とその他のある 1 つの表情 e の認識について考えたとき、理想的な注目点集合とは、全ての注目点 i について、図 5 に示すように、表情変化による輝度変化 $\Delta I_{i,e}^{(exp)}$ が形状モデルの誤差による輝度変化 $\Delta I_i^{(err)}$ よりも大きい、すなわち、

$$\Delta I_{i,e}^{(exp)} < \Delta I_i^{(err)} \quad (13)$$

を満たすような注目点の集合となる。図 5 中の輝度プロファイルが示すように、注目点付近の輝度が顔部品のエッジにおいて空間的にステップ状に変化することを仮定すると、最も厳しい (13) 式の条件は、

$$\xi_i^{(err)} < L_i < \xi_{i,e}^{(exp)} - \xi_i^{(err)} \quad (14)$$

と等価となる。ここで、 $\xi_i^{(err)}$ は、(6) 式より算出される注目点 i の位置が形状モデルの誤差によりずれる量を、 $\xi_{i,e}^{(exp)}$ は、無表情時において注目点 i の属する点対が跨いでいたエッジが表情 e の表出により移動する量、すなわち、顔部品の移動量をそれぞれ表す。また、 L_i は無表情時における顔部品のエッジからの注目点 i までの距離である。ここでは、図 2 に示した注目点集合における L_i の値は、全ての注目点について、X 軸方向及び Y 軸方向ともに、2.1 節にて述べた点対内の 2 点の間隔の半値、すなわち、3[pixel] である。なお、(14) 式の $\xi_i^{(err)}$ 及び $\xi_{i,e}^{(exp)}$ はそれぞれの元の値の絶対値であるとする。

以下では、各注目点について、 $\xi_i^{(err)}$ 及び $\xi_{i,e}^{(exp)}$ の値を算出することで、(14) 式がどの程度成立しているかどうかを検討する。ここでは、各注目点を、図 6 に示すような、顔部品内の左部、右部、上部、及び、下部のいずれかの領域に分類した。分類方法は、その注目点の属する点対の中心がいずれの領域中に存在するかを手動で判断し、その領域をその注目点の領域とするものとした。そして、(14) 式中の 3 つの変数について、左部及び右部に

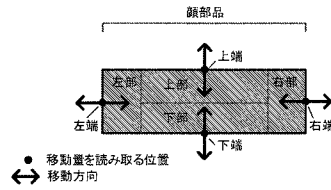


図 6: 顔部品移動量の読み取り位置、及び、注目点の領域区分

属する注目点については X 軸方向成分についてのみ、上部及び下部に属する注目点については Y 軸方向成分についてのみ考慮することとした。

4.1 形状モデルの誤差による注目点の算出位置のずれ量 $\xi_i^{(err)}$

本手法は、(7) 式のように、画像平面上で抽出された注目点 i を形状モデル上に弱中心投影にて投影することでその点の形状モデルにおける 3 次元座標 $p_i^{(shape)}$ を得ている。このため、形状モデル S に誤差がある場合、この 3 次元座標 $p_i^{(shape)}$ の Z 座標値に誤差が加わることとなる。注目点 i における Z 座標値の誤差を $\Delta Z_i^{(err)}$ とすると、(6) 式より算出される注目点の位置 q_i の実際の位置に対するずれ量 $\xi_i^{(err)}$ は、(7) 式において $p_i^{(shape)}$ の Z 座標に $\Delta Z_i^{(err)}$ を加えたものと加えないものとの差として得られる。

(6) 式及び (8) 式より分かるとおり、算出位置 q_i には各軸まわりの回転角が影響するが、ここでの議論では頭部姿勢は各時刻においてある 1 つの回転角のみが変化するものとする。さらに、顔部品の左部あるいは右部 (図 6) に属する注目点については、Y 軸方向の算出位置のずれには影響を受けないものとして、X 軸方向の算出位置のずれのみを考慮することとする。一方、顔部品の上部あるいは下部に属する注目点については、Y 軸方向の算出位置のずれのみを考慮することとする。このとき、左部あるいは右部に属する注目点については、Y 軸まわりの回転により X 軸方向の算出位置のずれが生じ、そのずれ量は、

$$\xi_i^{(err)} = \Delta Z_i^{(err)} \sin \theta_y \quad (15)$$

となる。一方、上部あるいは下部に属する注目点については、X軸まわりの回転によりY軸方向の算出位置のずれが生じ、そのずれ量は、

$$\xi_i^{(err)} = \Delta Z_i^{(err)} \sin \theta_x \quad (16)$$

となる。ここでは、 θ_x 及び θ_y の範囲を、 $-15 < \theta_x < 15$ [deg]、 $-40 < \theta_y < 40$ [deg] とする。以下では、両角度について最大値、すなわち、 $\theta_x = 15, \theta_y = 40$ の場合での議論を行う。

注目点のZ座標の誤差 $\Delta Z_i^{(err)}$ は、以下のようにして算出した。まず、1名の被験者の実際の顔の形状 S' を、カメラに対して正面を向いた状態で、3次元デジタルカメラ (KONICA MINOLTA 社製 VIVID910) を用いて計測した。次いで、各注目点 i のZ座標の誤差 $\Delta Z_i^{(err)}$ を、(7)式において、計測により得られた形状 S' を用いて算出した $p_i^{(shape)}$ のZ座標値に対する、円柱モデル S を用いて算出した $p_i^{(shape)}$ のZ座標値のずれ量として定義した。各注目点における $\Delta Z_i^{(err)}$ をプロットしたものを図7に示す。

4.2 顔部品の移動量 $\xi_{i,e}^{(exp)}$

顔部品の移動量 $\xi_{i,e}^{(exp)}$ は次のようにして取得した。まず、各顔部品について、左部、右部、上部及び下部内の各表情での顔部品の移動量は、それぞれ、左端、右端、上端及び下端における顔部品の移動量に等しいと仮定した(図6)。このとき、例えば、注目点 i が顔部品の左部に属する場合、その点に関する表情 e における顔部品の移動量 $\xi_{i,e}^{(exp)}$ は、その顔部品の左端における移動量に等しい値となる。ここでは、眉、目、鼻、口のそれぞれの各部品についての左端、右端、上端及び下端における各表情での顔部品の移動量を、変動輝度テンプレート作成時に用いた画像から手動にて読み取った。

4.3 各注目点の評価

本節では、各注目点の表情認識に対する有用性を評価する。ここでは、注目点の評価を、 $\xi_i^{(err)}$ 、 L_i 及び $\xi_{i,e}^{(exp)}$ の大小関係を用いて行う。例えばそれらの3変数が(14)式を満たすような注目点については、評価を最も高くする。表2に各場合に対する評価の一覧を示す。表2の評価に従い、ここでは注目点 i の表情 e についての評価値 $c_i(e)$ を、

$$c_i(e) = \begin{cases} \xi_{i,e}^{(exp)} - \xi_i^{(err)}, & \text{if } \xi_i^{(err)} < L_i \\ -\xi_{i,e}^{(exp)} - \xi_i^{(err)}, & \text{otherwise} \end{cases} \quad (17)$$

と定義した。この評価値は、表2の種類の番号が小さい、すなわち、表情 e に対する有用性が高い注目点ほど、評価値 $c_i(e)$ が高いという条件を満たす。

図8に各注目点の各表情についての評価値 $c_i(e)$ をプロットしたものを示す。図8より、驚き表情や喜び表情について評価値 c_i が高くなる注目点が多く、これらの表情の認識が比較的容易であることが推測される。一方、悲しみ表情については、認識に有用な注目点(種類1)の数よりも悪影響を及ぼす点(種類4)の数の方が多い。しかし、ここでの議論は最も厳しい条件におけるものであり、評価値が実際よりも低く算出されている点が多数存在するはずである。また、(5)式のロバスト推定の枠組みにより悪影響を及ぼす点の影響が軽減される。これらより、悲しみ表情の認識は可能であるものと推測される。実際、3章に示した結果では、悲しみ表情の認識が正しく行われている。これらの推測は被験者1名に対する検討によるものであるが、5名分の実験における表情認識結果(詳細は[10]の表2参照)の傾向に矛盾しない。

5. むすび

本稿では、我々が提案している、その場で容易に作成可能な変動輝度テンプレートを用いた表情認識手法の、頭部姿勢に対する頑健性の理論的な検討を行った。

今後は、推定精度を向上させるために、楕円体や粗い顔の近似形状モデルといった作成の容易な円柱よりも近似精度の高いモデルの使用を検討する予定である。また、既に、本手法が横方向の首振り動作に対して頑健であるが、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ことを確認済みである。この課題への対処として、各頭部姿勢変動での注目点の輝度の事前準備や表情輝度分布モデルの逐次更新、あるいは、照明変動に不変な特徴量の使用を検討中である。さらに、オンラインクラスタリングによる表情の自動学習や、Fisher判別法などの識別の概念を取り入れた効率的かつ効果的な注目点の抽出手法の構築を目指す予定である。

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] R. Kallouby 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).
- [3] 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).
- [4] 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).
- [5] S. B. Gokturk, C. Tomasi, B. Girod and J.-Y.

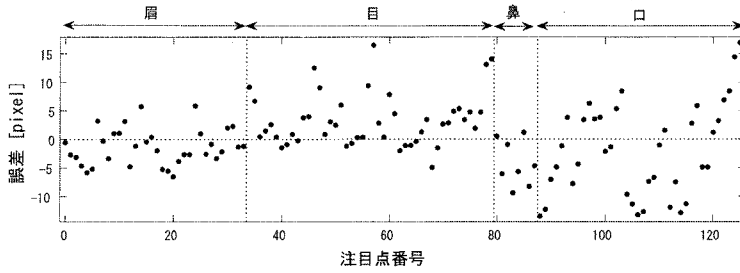


図 7: 各注目点の Z 座標値の誤差

口において誤差が大きくなる事が分かる。なお、眉や目領域において誤差が大きい点は顔の端部周辺に配置された点である。

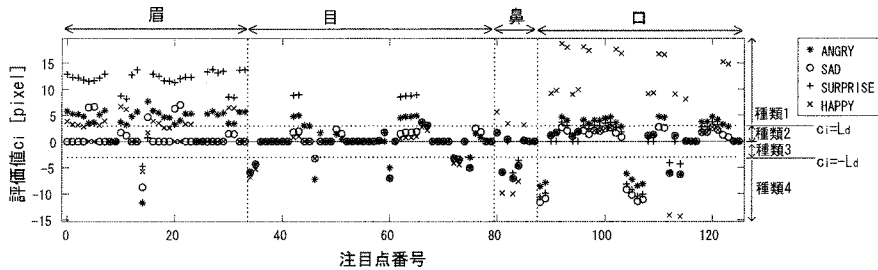


図 8: 注目点の評価値

怒り表情や驚き表情に関しては主に肩部や目部に、喜び表情に関しては主に口部に認識に有用な注目点 (種類 1) が多数存在することが見て取れる。一方、悲しみ表情に関しては認識に有用な注目点 (種類 1) が他の表情に比べて少ない。なお、注目点の種類の詳細は表 2 のとおりである。

表 2: 形状誤差による算出位置のずれ量 $\xi_i^{(err)}$, 注目点のエッジからの距離 L_i , 及び、顔部品の移動量 $\xi_{i,e}^{(exp)}$ の大小関係による注目点の評価

種類	条件	評価	効果
1	$\xi_i^{(err)} < L_i < \xi_{i,e}^{(exp)} - \xi_i^{(err)}$	最高	頭部姿勢及び表情認識共に有用である
2	$\xi_i^{(err)} < L_i \cap \xi_{i,e}^{(exp)} - \xi_i^{(err)} < L_i$	中	頭部姿勢推定には有用であるが、表情認識には大きな影響を及ぼさない
3	$\xi_{i,e}^{(exp)} + \xi_i^{(err)} < L_i < \xi_i^{(err)}$	低	頭部姿勢推定の精度低下を招くが、表情認識には大きな影響は及ぼさない
4	$L_i < \xi_i^{(err)} \cap L_i < \xi_{i,e}^{(exp)} + \xi_i^{(err)}$	最低	頭部姿勢変動と表情変化の分離ができず、両方の推定の精度低下を招く

※種類 1, 2 及び種類 3, 4 の条件で一部符号が異なっているのは、最も厳しい条件で考えたことによる。

- Bouquet: “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).
- [6] 岡, 菅野, 佐藤: “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, **47**, SIG10 (CVIM15), pp. 185–194 (2006).
- [7] F. Dornaika and F. Davoine: “Simultaneous facial action tracking and expression recognition using a particle filter”, Proc. of the Tenth IEEE International Conference on Computer Vision, **2**, pp. 1733–1738 (2005).
- [8] Z. Zhu and Q. Ji: “Robust real-time face pose and facial expression recovery”, Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 681–688 (2006).
- [9] R. Gross, I. Matthews and S. Baker: “Generic vs. person specific active appearance models”, Image and Vision Computing, **23**, 11, pp. 1080–1093 (2005).
- [10] 熊野, 大塚, 大和, 前田, 佐藤: “変動輝度テンプレートを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 (2007).
- [11] 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).
- [12] 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).
- [13] 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).