Vol.0 No.0, 2002

原著論文

表情譜: 顔パーツ間のタイミング構造に基づく表情の記述

平山 高嗣 *1 川嶋 宏彰 *1 西山 正紘 *1 松山 隆司 *1

Facial Expression Representation based on Timing Structures in Faces

Takatsugu Hirayama,*1 Hiroaki Kawashima,*1 Masahiro Nishiyama,*1 and Takashi Matsuyama*1

Abstract – **This paper presents a method for interpreting facial expressions based on temporal structures among partial movements in facial image sequences. To extract the structures, we propose a novel facial expression representation, which we call a facial score, that is similar to a musical score. The facial score enables us to describe facial expressions as spatio-temporal combinations of temporal intervals; each interval represents a simple motion pattern with the beginning and ending times of the motion. Therefore, we can classify fine-grained expressions from multivariate distributions of temporal differences between the intervals in the score. In this paper, we provide a method to obtain the score automatically from input images using bottom-up clustering of dynamics. Our experiments show the effectiveness of the method by separating smiling expressions into intentional and spontaneous categories using the obtained scores.**

Keywords : **facial expression representation, timing structure, linear dynamical system, bottom-up clustering, intentional smile, spontaneous smile**

1. はじめに

1.1 研究背景

人間とコンピュータの間のより高次な協調関係を実現するために，ヒューマンインタフェースへの応用の観点から，顔表情に対する研究が活発に行われている．具体的には，顔表情を対象とする人間の視覚機能をコンピュータに持たせるための表情認識技術や，コンピュータに表現力豊かな顔表情の表出機能を持たせるための表情生成技術である．

その高次な協調関係を実現するためには，表情生成の入力，もしくは表情認識の出力として，どのような表情を扱うかという問題は重要である．従来の研究では，主に静止画をもとにして基本的なカテゴリ（喜び・驚き・恐怖・怒り・嫌悪・悲しみ・軽蔑）に分類された表情が扱われている [1]．しかし，実際の表情は，相手に何を伝えたいかという目的に応じて意図的に作られるものもあれば，情動・反射によって自発的に表出されるものもあり，同じカテゴリ内でも多様な種類に分類できる．さらに，表情の持つ意味が個人ごとに異なる場合もあり，表情における個性の扱いについても考慮する必要がある．

このような表情の多様性や微妙さに注目した研究としては，実験心理学的な分析研究がいくつか行われている．Schmidt らは，いろいろな状況下で起こる笑いにおける顔器官の動きを分析した [2]．この研究で明らかとなった知見は，一人でビデオ鑑賞をするような場面と面接などの社交的な場面での自発的な笑いには，口端の動きの笑いの頂点を示す状態の持続長に差異が表れるということである．また，内田らは，高速度カメラによって表情を撮影することで，顔の非線形な変化と主な顔器官の動き始めの時間差を捕らえることに成功し，さらに，意図的な表情と自発的な表情における動的変化（顔特徴点の変位）パターンの類似点，及び相違点を示した [3]．

これらの研究は，表情の表出者を対象にして，顔器官の動きやその時間関係がどのように変化するかを分析した研究であるが，表情観察者を対象にして，微妙な表情変化に対する認知の影響を分析した研究もある．蒲池らは，表情の表出時間を変化させた CG 映像を作成し，変化速度が表情認知に影響を与えることを示した [4]．また，Krumhuber らも同様の分析を自然な笑いと偽りの笑いを対象に行った [5]．一方，Nishio らは表出時間だけでなく，より詳細に目と口の動きの時間関係を制御した様々な笑いの CG 映像を用いることで，笑い開始時の目と口の動きのタイミングの違いに基づいて表情観察者が快の笑い・不快の笑い・社交の笑いを分類していることを明らかにした [6]．

このように，心理学の立場から，顔の動きに関わる要因，特に時間的要因が表情理解に重要な役割を果たし，顔の器官別にその要因を考察することが有効であると示されている．一方，コンピュータビジョン分野の立場での表情認識・生成システムの研究においても，顔器官の動きを扱う研究は近年になって行われてきて

*1: 京都大学大学院 情報学研究科
*1: Graduate School of Informatics, Kyoto University

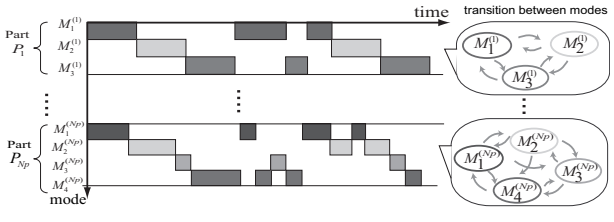


図 1　　表情譜
Fig. 1　Facial score.

いる[7]~[11]．しかし，それらの研究は以下の問題点を持つ．

- 顔器官の動きのタイミングや時間長をモデル化しておらず，表情の動的側面を十分に用いていない
- 表情の分類が感情に基づく基本カテゴリーに留まり，微妙な表情の変化を扱えていない

そこで，本研究では，顔の構成要素（顔パーツ）それぞれの動きに関するタイミング構造に基づいて，表情の動的側面をより詳細に表現する枠組を提案する．

1.2　提案手法

タイミング構造とは，ある 2 つの区間がどのような時間関係で発生し終了するのかといった構造を表すものと定義する．また，区間とは，静止状態や収束性の運動のような単純な変化を行う事象の時間範囲を表すものとし，開始時刻（始点），終了時刻（終点），及び運動パターン（モード）のラベルを属性として持つものとする．本論文では，顔パーツの動きを区間を単位として表し，表情におけるタイミング構造を記述する表現形式を，音符と音符のタイミングの芸術である音楽を記述する楽譜になぞらえて「表情譜」と呼ぶ．表情譜の概念図を図 1 に示す．

このような区間を単位とした記述では，区間のモードをどのように定義するかが重要である．モードに対応する表情の記述形式としては，Ekman らが開発した FACS （Facial Action Coding System）における AU（Action Unit）が挙げられる[12],[13]．AU は，解剖学的に独立し，視覚的に識別可能な表情動作の最小単位として設定されているが，多くの表情を人間が観察することで主観的に分類したものであるので，それによって表現しきれないような表情動作も実際には存在するのではないかと考察される．例えば，目を閉じるという動きは AU43 という単位で表現されるが，実際の動きは速度や強度に関して幅広く変化する．そこで，本研究ではモードの抽出を，顔表情の特徴を表す特徴ベクトル系列からボトムアップに行う．これによって，AU では表現しきれない表情動作も表現可能となる．

以上をまとめると，本論文で提案する表情譜は以下の特徴を持つ．

- 区間を単位とした表情のタイミング構造の記述が可能
- 区間のモードとして学習データからボトムアップに抽出された運動パターンの利用が可能

表情から表情譜を獲得し，表情譜で記述されるタイミング構造から，表情変化の時間的要因を詳細に理解する流れを以下に示す（図 2 参照）．

1. 表情から顔の特徴を表す特徴ベクトル系列を抽出
2. 特徴ベクトル系列を用いて顔の運動をモードに分節化し，表情譜を獲得
3. 表情譜から，表情を理解する上で有用なタイミング構造を抽出

以上の処理を自動化することができれば，表情認識システム等に応用でき，コンピュータがより詳細に人間の表情を読み取ることができるようになると考える．

そこで，本論文では，まず，表情譜の自動獲得手法を提案する．次に，表情変化の微妙な差異を時間的要因に基づいて詳細に理解するための，表情譜からのタイミング構造の抽出手法を提案する．そして，これらの手法の有効性を検証し，表情譜が表情を詳細に表現する枠組みとして有効であるかを評価する．評価実験で対象とする表情は，人間同士の日常コミュニケーションにおいて重要な役割を持ち，かつ，動きの時間的構造が微妙に異なるという知見が得られている意図的な笑いと自発的な笑いとし，両者のタイミング構造の比較を行う．また，個人間のタイミング構造の比較も行い，この構造から個性を抽出できるかどうかを検証し，表情譜が備える表情の表現能力を評価する．さらには，意図的な笑いと自発的な笑いの識別実験を行い，タイミング構造に基づいた表情識別の可能性を検証する．

次章では，表情におけるタイミング構造を記述する表現形式として提案する表情譜を設計する．3. 章では，入力として顔画像系列を与えた時に，出力として表情譜を自動獲得する手法について述べる．4. 章では，実際に撮影した顔画像系列から表情譜を自動獲得し，笑顔を対象として表情譜の有効性の評価を行う．最後に，5. 章では本論文の結論を述べる．

2.　表情譜の設計

2.1　表情譜の定義

表情譜とは，顔の各構成要素がどのようなパターンで，どのような時間関係で運動するかを記述する表現形式である．ここで，以下の用語を定義する．

顔パーツと顔パーツ集合:　顔パーツとは，空間的に分離可能な顔の構成要素のことを表す．表情譜で記述する顔パーツの個数を N_p とした時，顔パーツ集合を $\mathcal{P} = \{P_1, ..., P_{N_p}\}$ で定義する．

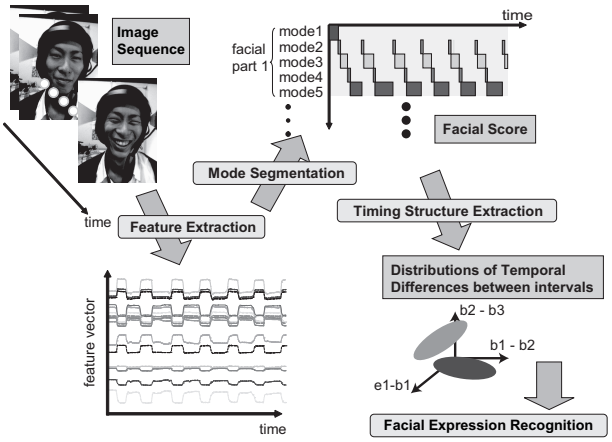


図 2　表情譜を用いた表情理解の流れ
Fig. 2　The flow of facial expression recognition using the facial score.

モードとモード集合:　モードとは，単調な変化を行う事象を表す．顔パーツ $P_a\,(a \in \{1,...,N_p\})$ におけるモードの個数を N_{m_a} とした時，顔パーツ P_a におけるモード集合を $\mathcal{M}^{(a)} = \left\{M_1^{(a)},...,M_{N_{m_a}}^{(a)}\right\}$ で定義する．

区間と区間集合:　区間とは，単調な変化を行う事象の時間範囲を表す．顔パーツ P_a における時系列データが T 個あり，その時系列データが N_{k_a} 個の区間で表される時，顔パーツ P_a における区間集合を $\mathcal{I}^{(a)} = \left\{I_1^{(a)},...,I_{N_{k_a}}^{(a)}\right\}$ で定義する．また，区間 $I_k^{(a)}\,(k \in \{1,...,N_{k_a}\})$ は始点 $b_k^{(a)} \in \{1,...,T\}$，終点 $e_k^{(a)} \in \{1,...,T\}$，及びその区間を表現するモードのラベル $m_k^{(a)} \in \mathcal{M}^{(a)}$ を属性として持つ．

表情譜:　表情譜とは，全ての顔パーツにおける区間集合の集合である．つまり，表情譜を $\{\mathcal{I}^{(1)},...,\mathcal{I}^{(N_p)}\}$ で定義する．

2.2　表情譜における顔パーツ

タイミング構造から得られる情報を利用して表情を表現，理解するという観点から考えると，動きのタイミングの差異が表れる領域を別の顔パーツとして扱うべきである．従って，動きに常に共起性があり，他の領域との独立性が観測される領域を顔パーツとする．

Ekman らは，基本的感情（喜び・驚き・恐怖・怒り・嫌悪・悲しみ）が表情に現れる様子の違いを，appearance-base で独立した動きが観測される顔の 3 領域（眉の周辺部分・目の周辺部分・口の周辺部分）の組み合わせにより解明した [12]．本論文では，この 3 領域に着目するとともに，眉と目の周辺部分に関しては左右のパーツを別のパーツとして扱う．これは，実際の表情において，眉と目に非対称な動きが観察されることがあるので，別々に扱うことでより微妙な表情の変化を記述することができると考察されるためである．

次に，どのような特徴量（特徴ベクトル）を設定すれば微妙な表情変化を表現できるかという問題も重要である．本論文では，顔パーツを代表する特徴点の座標値を特徴ベクトルとし，動きの情報を直接扱うことにする．微妙な表情を表現するために有効な情報を提供すると考えられる特徴として，皺が挙げられるが，その動きは顔パーツの動きと関連しているため，特徴点の動きによって間接的に表現することが可能である．例えば，笑う時にできる頬皺は鼻の特徴点の動きと共起関係にある．

ゆえに，顔パーツ集合 $\mathcal{P}$ の要素は，右眉，左眉，右目，左目，鼻，口とする．また，顔パーツ P_a の特徴ベクトル $z^{(a)}$ は，顔パーツ P_a における特徴点の数を n_{p_a}，p 番目 $(p \in \{1,...,n_{p_a}\})$ の特徴点における座標値を $\left(x_p^{(a)},\ y_p^{(a)}\right)$ とすると，

$$z^{(a)} = \left(x_1^{(a)},\ y_1^{(a)},...,x_{n_{p_a}}^{(a)},\ y_{n_{p_a}}^{(a)}\right)^{\mathsf{T}} \tag{1}$$

という $2n_{p_a}$ 次元列ベクトルとして表せる．

なお，顔の向きも表情を理解する上では重要であるが，顔の向きを同定することにより正面顔を復元 [14] することができると仮定すれば，顔の向きも含めた表情の記述は，顔の向きを同定する問題と，正面顔における顔パーツの動きを記述する問題に分離して扱うことができる．本論文では，正面顔における顔パーツの動きの記述に焦点を絞って論旨を展開する．

2.3　表情譜におけるモード

顔パーツの動きは，ある静止状態から運動状態に，そして，運動状態からは別の静止状態，もしくは別の運動状態に遷移するという形の状態遷移の繰返しで記述できると考え，それぞれの変化（静止の場合は直流成分）をモードと呼ぶ．ここで，運動状態とは，周期的な動きや特徴ベクトルの値が急激に増加していくような動きではなく，速度の符号が変わらず，かつ一定の値に収束していくような動きのみが行われる状態を指すことにする．例えば，口の開閉の運動は，閉じている（静止状態），開く（運動状態），開いている（静止状態），閉じる（運動状態），閉じている（静止状態）というように記述できる．そして，これら個々のモードは，それぞれが線形動的システムで表現できると仮定する．3.2 節では，特徴ベクトル系列からこれらのモードをボトムアップに抽出する方法について述べる．

一般に線形動的システムは，状態の時間変化を表す状態方程式と，状態から観測への写像を表す観測方程式からなるが，ここでは状態が特徴ベクトルそのもの

であると仮定する[1]．したがって，顔パーツ P_a におけるモード $M_i^{(a)}$ $(i \in \{1, ..., N_{m_a}\})$ は，次の状態方程式で表される．

$$z_t^{(a)} = F^{(a,\,i)} z_{t-1}^{(a)} + f^{(a,\,i)} + \omega_t^{(a,\,i)} \qquad (2)$$

ここで，$z_t^{(a)}$ は時刻 t における特徴ベクトルである．$F^{(a,\,i)}$ は遷移行列であり，モード毎に異なる．$f^{(a,\,i)}$ はバイアス項である．$\omega^{(a,\,i)}$ はプロセスノイズであり，平均ベクトル $\mathbf{0}$，共分散行列 $Q^{(a,\,i)}$ の正規分布に従うとする．

線形動的システムは周期的，振動的な動きも表現可能である．しかし，先に述べたように，我々は静止状態もしくは収束性の運動状態のみをモードとして抽出したい．そこで，3.2 節では，式 (2) における遷移行列 F の固有値に制約を加え，静止状態もしくは収束性の運動状態のみをモードとして抽出する手法について述べる．

2.4　表情譜におけるタイミング構造

前節までに定義した表情譜から，顔パーツ間の動きの時間関係，すなわちタイミング構造を抽出することが可能となる．ここでは，表情変化の微妙な差異を抽出するためのタイミング構造の表現方法について考察を行う．

2.4.1　分布によるタイミング構造の表現

一般に，2 つの区間 I_i, I_j の時間関係は，区間の始点 b_i, b_j，終点 e_i, e_j の前後関係 {前，後，同時} に注目すれば，図 3 に示すように 13 通りに分類可能である [15], [16]．しかし，実際に表情変化の微妙な差異が理解できるほど詳細に表情を表現する上では，単なる前後関係だけでは不十分であり，区間がどの程度の時間差で開始，終了するのかといったずれの程度が重要となる．従って，本論文では図 3 の 13 通りの関係を拡張し，区間の始点・終点の時間差の分布を用いたタイミング構造の表現方法を提案する．

まずはじめに，2 つの区間 I_i, I_j のタイミング構造を 1 次元空間の分布で表現すると，$H(b_j - b_i), H(e_j - e_i), H(b_j - e_i), H(e_j - b_i)$ の 4 個の分布で表現できる．ここで，$H(r)$ は r を変数とする 1 次元空間の分布とする．同様に，2 次元空間の分布で表現すると，$H(b_j - b_i, e_j - e_i), H(b_j - b_i, b_j - e_i), H(b_j - b_i, e_j - b_i) H(e_j - e_i, b_j - e_i), H(e_j - e_i, e_j - b_i), H(b_j - e_i, e_j - b_i)$ の 6 個の分布で表現できる．ここで，$H(r_1,\, r_2)$ は r_1, r_2 を変数とする 2 次元空間の分布とする．その例として，横軸を始点の差，縦軸を終点の差とする分布 $H(b_j - b_i,\, e_j - e_i)$ を図 4 に示す．同様に 3 次元以上の空間の分布も表現できる．

1: この場合，線形動的システムは 1 次の多変量自己回帰モデルとなる．経験的には，1 次もしくは 2 次までの自己回帰を用いることで，人の表情の顔パーツの動きをうまく分節化できる．

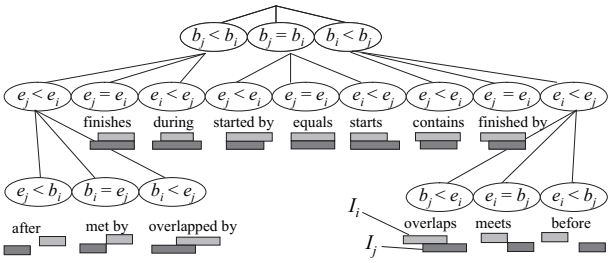


図 3　2 つの区間の時間関係
Fig. 3　Temporal relations between two intervals.

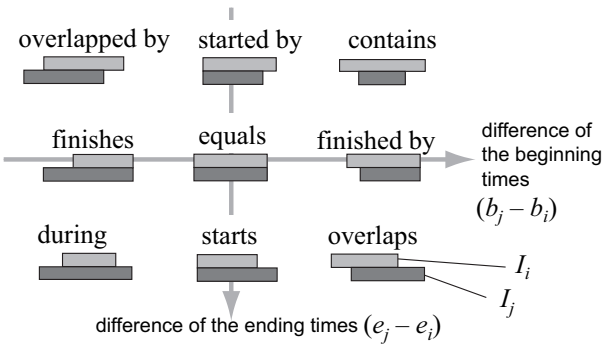


図 4　2 つの区間の始点の差・終点の差で表される 2 次元空間の分布の一例
Fig. 4　An example of two-dimensional distributions of temporal differences between two intervals.

2.4.2　表情譜から抽出するタイミング構造

実際にこれらの分布を考える際には，どの区間の組み合わせを扱うかが重要となる．本論文では，その組み合わせを以下のように定める．まず，顔パーツ P_a における区間 $I_k^{(a)}$ と，P_a 以外の全ての顔パーツ $P_{\bar{a}}$ $(\bar{a} \neq a, \bar{a} \in \{1, ..., N_p\})$ における区間 $I_l^{(\bar{a})}$ のうち，時間的に最も近い区間の組み合わせに注目する．これらの区間は $I_{l^*}^{(\bar{a})}$ $\left(l^* = \arg\min_l \mathsf{IntervalDist}\left(I_k^{(a)},\, I_l^{(\bar{a})}\right)\right)$ で求められる．ここで，区間同士の時間的な近さを表す距離 IntervalDist は次式で表されるものとする．

$$\mathsf{IntervalDist}\ \left(I_k^{(a)},\, I_l^{(\bar{a})}\right) = \min\left(\left|b_k^{(a)} - b_l^{(\bar{a})}\right|, \left|e_k^{(a)} - e_l^{(\bar{a})}\right|\right) \qquad (3)$$

求められた区間のタイミング構造を 2 次元空間の分布で表現するならば，1 次元の分布の総数が ${}_{N_p}C_2 (= N_h)$ 個であるので，${}_{N_h}C_2$ 個の分布が構成される．

3.　表情譜の自動獲得

3.1　自動獲得手法の概要

まず，顔画像系列を与えた時に，Active Appearance Model（AAM）[17] を用いて顔パーツの特徴点の座標値を抽出する．AAM とは，shape（特徴点の座標値）

と grey-level（輝度値）の相関をパラメタとして持つ統計学的なモデルであり，高速かつ安定に画像とのマッチングを行うことが可能である．また，モデルを画像にマッチングさせた時のモデルパラメタから簡単に特徴点の座標値を得ることができる．次に，得られた特徴点の座標値系列を，大きさ・並進・回転に関して正規化する [18]．最後に，正規化後の特徴点の座標値系列を特徴ベクトル系列とし，Hybrid Dynamical System [19] で用いられている線形動的システムのパラメタ推定法を利用して，特徴ベクトル系列をモードへと分節化する．モードへの分節化は，各顔パーツ毎に適用し，そしてその結果から表情譜を獲得する．次節ではモードへの分節化の手法について詳細に述べる．

3.2 モードへの分節化

本論文では，階層的クラスタリングを用いて，系列を構成する線形動的システム集合の各パラメタを，各システムで表される区間への分節化と同時に推定する手法を提案する．

各モードは，それぞれ式 (2) の線形動的システムによって表現される．ここで，特徴量の値がほとんど変化しないような静止状態や,「はじめは早い変化であり，次に静止していく」ような収束性の運動を単位として分節化を行うために，式 (2) における遷移行列 F に制約を加え，全ての固有値の絶対値が 1 より小さい線形動的システムを考える．以下では，まず最大固有値の絶対値に制約を加える方法について述べ，次に線形動的システムのクラスタリング手法について述べる．なお，説明を簡便にするために顔パーツ P_a に属することを示す添字 a を省略して述べる．

3.2.1 制約付き線形動的システム同定

特徴ベクトル系列 $z_1^{(i)},..,z_T^{(i)}$ から遷移行列 $F^{(i)}$ を計算するためにまず，$Z_0^{(i)} = \left[z_1^{(i)},...,z_{T-1}^{(i)}\right]$，$Z_1^{(i)} = \left[z_2^{(i)},...,z_T^{(i)}\right]$ と置く．このとき，$F^{(i)}$ の同定は，各時刻における自乗予測誤差を最小にする，以下の問題と考えることができる．

$$F^{(i)*} = \arg\min_{F^{(i)}} \left\| F^{(i)} Z_0^{(i)} - Z_1^{(i)} \right\|^2 \qquad (4)$$

これを行列方程式として微分法を用いて解くことで $F^{(i)}$ は，

$$F^{(i)*} = \lim_{\delta^2 \to 0} Z_1^{(i)} Z_0^{(i)\mathsf{T}} \left(Z_0^{(i)} Z_0^{(i)\mathsf{T}} + \delta^2 I \right)^{-1} \qquad (5)$$

と求められる．ここで，I は $2n_p$ 次元の単位行列であり，δ は正の実数値である．δ を 0 に収束させずに適当な正の実数値にすることで $F^{(i)}$ の固有値を 1 より小さくする．

3.2.2 モードの自己組織化

2.3 節で述べたように，顔パーツの動きを，静止状態と収束性の運動状態に分ける．そのために，まずは特徴ベクトルの時間差分のノルム，つまり動きの速度に対応する指標が零付近か否かを閾値処理で判断し，おおまかに分節化を行う（初期分節化）[2]．これによって，静止区間と運動区間が交互に表れるような分節化が行われるが，実際には特徴点追跡時などのノイズによって，比較的短い区間が多く得られる．

次に，分節化された区間の間に距離を定義することで，階層的クラスタリングに基づいて区間を併合していく．併合処理はまず，それぞれの区間が別のモードに従うものとして，モードを表す線形動的システムのモデルパラメタを同定する．そして，全てのモード間の距離 Dist を計算し，その中で最も近い 2 つのモードを 1 つのモードとして併合し，新たにそのモードのモデルパラメタ及び他のモードとの距離 Dist を計算する（距離 Dist の定義は次節で行う）．この併合処理を繰り返し，モードの数を減らしていく．適切なモード数は，モードから復元した特徴ベクトル系列と元の系列の誤差を併合処理の各段階で計算し，その変化を見ることで決定する [19]．付録に階層的クラスタリングのアルゴリズムを示す．

3.2.3 モード間の距離

モード間の距離尺度としては，次の式で表される予測誤差 E を距離尺度として用いる．

$$E\left(M_i||M_j\right) = \frac{1}{C} \sum_{I_k \in \mathcal{I}_i} \sum_{t=b_k}^{e_k} \left({E_t^{(i|j)}}^2 - {E_t^{(i|i)}}^2 \right) \qquad (6)$$

ここで，C は区間集合 $\mathcal{I}_i$ に含まれる区間 I_k の区間長の総和であり，これによって時間的な正規化を行う．また，$E_t^{(i|j)}$ は，

$$E_t^{(i|j)} = F^{(j)} z_{t-1}^{(i)} + f^{(j)} - z_t^{(i)} \qquad (7)$$

で表されるものとする．式 (6) は，モード M_i と M_j に関して非対称であるため，これを相互に評価することで以下のような対称な距離 Dist を定義する．

$$\mathrm{Dist}\ \ (M_i,\ M_j) = \left\{ E\left(M_i||M_j\right) + E\left(M_j||M_i\right) \right\} / 2 \qquad (8)$$

4. 評価実験

4.1 表情映像の獲得

表情譜が，表情変化の時間的要因と，表情間におけるその微妙な差異を詳細に理解するための表現として有効であるかを評価するため，意図的に作った笑いと自発的に表出された笑いを対象に実験を行った．顔画像の撮影には，顔の変化を詳細に捕らえるために高速

2: 適切な閾値を求める手法を検討する余地がある．現状では，静止状態から運動状態へ変化する時点を映像から目視によって数箇所抽出し，それらの時点の特徴ベクトルの時間差分のノルムを平均し，それを閾値として設定している．

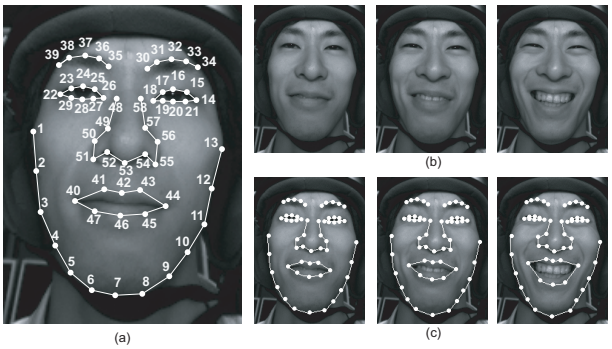


図 5　(a) 特徴点の定義．(b) 評価実験で撮影した画像系列．(c)AAM を用いて追跡した特徴点を表示した画像系列

Fig. 5　(a)Feature points, (b)input image sequence, (c)tracked feature points using AAM.

撮影が可能な小型 IEEE1394 カメラ（Point Grey Research 社製 Flea）を使用した．入力顔画像系列は，6 人（被験者 A〜F）の人物のそれぞれの笑いを，解像度 240 × 320pixel，フレームレート 60fps で撮影したものを使用した [3]．撮影は，頭部の動きが生じた場合でも正面顔の撮影を行うために，ヘルメット前方にカメラを固定したカメラシステムを用いた．これは 2.2 節で述べたように，本研究では表情の記述を，顔向きを同定する問題と，正面顔における顔パーツの動きを記述する問題に分離して考えているためである．意図的な笑いは，被験者に偽りの笑いを表出させることを狙いとして，Gross らによって標準化された嫌悪感喚起映像 [20]（10 分間）を提示し，その視聴中に適当なタイミングで笑顔を表出するように指示して撮影した．自発的な笑いは，被験者に漫才の映像（22 分 33 秒間）を視聴させることで表出されたものを撮影した．各表情の表出数は，意図的な笑いが 50，自発的な笑いが各被験者で異なり，被験者 A が 37，B が 39，C が 30，D が 38，E が 31，F が 29 であった [4]．

4.2　表情譜の獲得

まず，撮影された顔画像系列に対して，AAM を用いて各顔パーツの特徴点の追跡を行った [5]．特徴点は，図 5 (a) に示す点の位置として定義し，各眉に 5 点，各目に 8 点，鼻に 11 点，唇に 8 点とした．また，それらの追跡精度を向上させるために，顎部に特徴点を配置した．画像上の数字は特徴点のラベルを表す．撮影された顔画像系列（図 5 (b)）に対して AAM を適用して，特徴点を追跡した結果を図 5 (c) に示す．これらの図を比較すると，表情の変化に伴う特徴点の変化が精度良く追跡されていることが読み取れる．

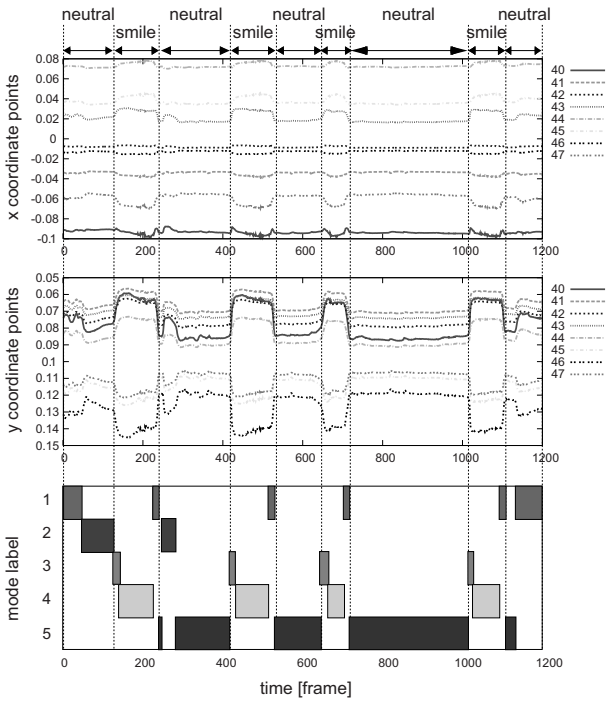


図 6　自発的な笑いの口パーツの特徴ベクトル系列と分節化の結果（凡例の数字は 図 5 (a) の特徴点の番号に対応している）

Fig. 6　Feature vector sequences of the mouth part during spontaneous smiles, and its segmentation result (the numbers in the legends correspond to the feature point numbers in Fig.5 (a)).

次に，得られた特徴ベクトル系列に対して，各パーツ毎に 3.2 節で述べた方法を用いてモードへの分節化を行った．分節化の結果の一例として，自発的な笑いの口パーツに関するものを図 6 に示す．図の上段の縦軸は特徴点の x 座標の値，中段は特徴点の y 座標の値，下段はモードの種類を表している．横軸は時間軸である．この図より，無表情の状態（モード 2 と 5），笑いの頂点の状態を示す最大表出状態（モード 4），笑いの開始時の動き（モード 3）や笑いの終了時の動き（モード 1）がそれぞれ異なるモードとして適切に分節化されていることが確認できる．

上記の分節化によって各パーツ毎に得られた区間集合から，2.1 節で定義した表情譜を構成した．意図的な笑いと自発的な笑いの表情譜の一部をそれぞれ図 7 と図 8 に示す．これらの図より，各パーツ間の区間の開始と終了のタイミングにずれが存在していることを確認できる．特徴的な例としては，図 8 の 320 フレームに注目すると，目と鼻は静止状態に入っているが，口は閉じた後に一旦下がり，静止状態の位置まで上がるという動作状態が続いている．以上の結果から，表情譜が表情変化に伴う顔パーツそれぞれの動きのタイ

3: いくつかの民生用デジタルカメラで 60fps の映像を撮影することができる．
4: 著者でも被験者でもない観察者によってカウントされた．
5: 特徴点の追跡には，AAM-API を用いた [21]．

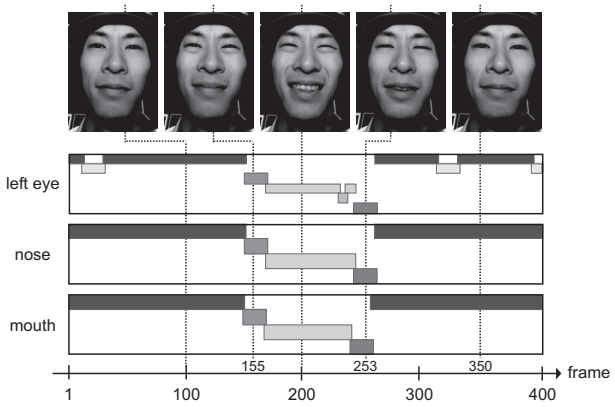


図 7　意図的な笑いの表情譜
Fig. 7　The facial score of an intentional smile.

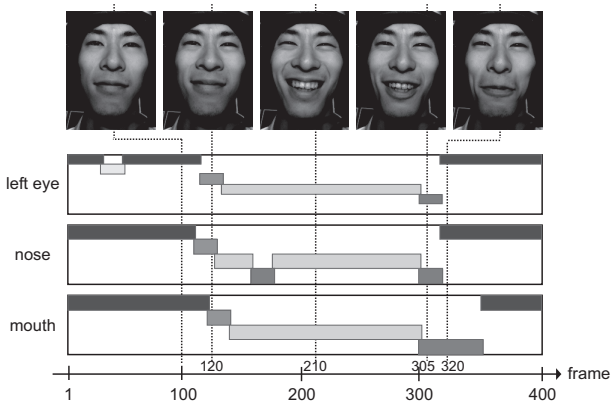


図 8　自発的な笑いの表情譜
Fig. 8　The facial score of a spontaneous smile.

ミング構造を詳細に表現することができるといえる．

ここで，図 7 と図 8 を比較すると，意図的な笑いに伴う顔パーツ間の動きのタイミングはほぼ同期し，自発的な笑いのタイミング構造はより複雑であることが見て取れる．この一例からではあるが，それらの笑いのタイミング構造に差異が存在することが推測される．次節では，この差異について検証を行う．

なお，カメラシステムの容量，および提示映像の長さの限界から，一度のセッションで全ての表情表出を撮影するのではなく，数回の表出を含んだ映像を何度か撮影する方法をとった．表情譜は，これらの映像のそれぞれに対して自動で求め，得られた複数の表情譜間でのモードの対応付けは手動で行った．また，ノイズなどの影響により，1 つのモードに対して複数の区間が対応する場合があり，これは人手で映像を確認しながら，複数の区間を 1 つの区間に統合した．

以上の手動操作は，文献 [19] の Expectation-Maximization アルゴリズムを用いた学習方法を利用することで自動化できるが，得られる表情譜のタイミング構造は，学習アルゴリズム自体の精度に影響を

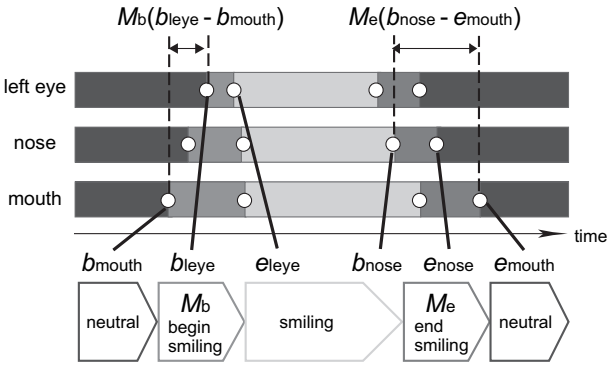


図 9　タイミング構造を抽出するための表情譜の要素
Fig. 9　Elements of the facial score to extract timing structure.

受ける可能性がある．本論文は，表情においてタイミング構造を表現することの有効性を検証することが目的であるため，自動化は，表情譜における分節化の候補を抽出するものであると位置づけ，最後に，人手で映像を確認することによって，候補に対して対応付けや統合を行うという形をとった [6]．

4.3　意図的な笑いと自発的な笑いにおけるタイミング構造の比較

得られた表情譜を用いて，意図的な笑いと自発的な笑いを分離できるタイミング構造が存在するか，それがどのような構造であるかを検証すると共に，個人差についても検討を行った．具体的には，注目するモードを笑い開始時の動き M_b と笑い終了時の動き M_e として，左目・鼻・口のそれぞれのパーツにおけるモード M_b と M_e が，互いにどのような時間関係で開始し終了するかを調べた．なお，無表情状態と笑いの最大表出状態のモードは，その時間長が漫才の文脈の特性 (笑いを誘う区間と誘わない区間の長さ) によって直接影響を受けるため，本論文では注目しなかった．眉パーツについては，笑いによる変化がほとんど観察されないため，分析の対象外とした．また，右目を分析の対象としなかったが，これは従来研究 [2], [3], [5], [6] において笑いの種類の違いによって左右の目の動きの時間的要素に差異があるという知見が得られていないことを参考にしたためである．

図 9 に示すように，左目，鼻，口に関するモードの始点と終点をそれぞれ b_{leye} , e_{leye} , b_{nose} , e_{nose} , b_{mouth} , e_{mouth} と表し，例えば，笑いの開始時のモード M_b に関する鼻の始点と口の始点の差を $M_b(b_{\mathrm{nose}} - b_{\mathrm{mouth}})$ と表すことにする．

6: この場合，候補を自動抽出する際に，閾値処理の要素（初期分節化の閾値および 3.2.1 節の固有値の制約）が入っている．したがって，人手による作業のあいまい性はほとんどなく，分節化結果（タイミング構造）が細かくずれる恐れはないと考えられる．

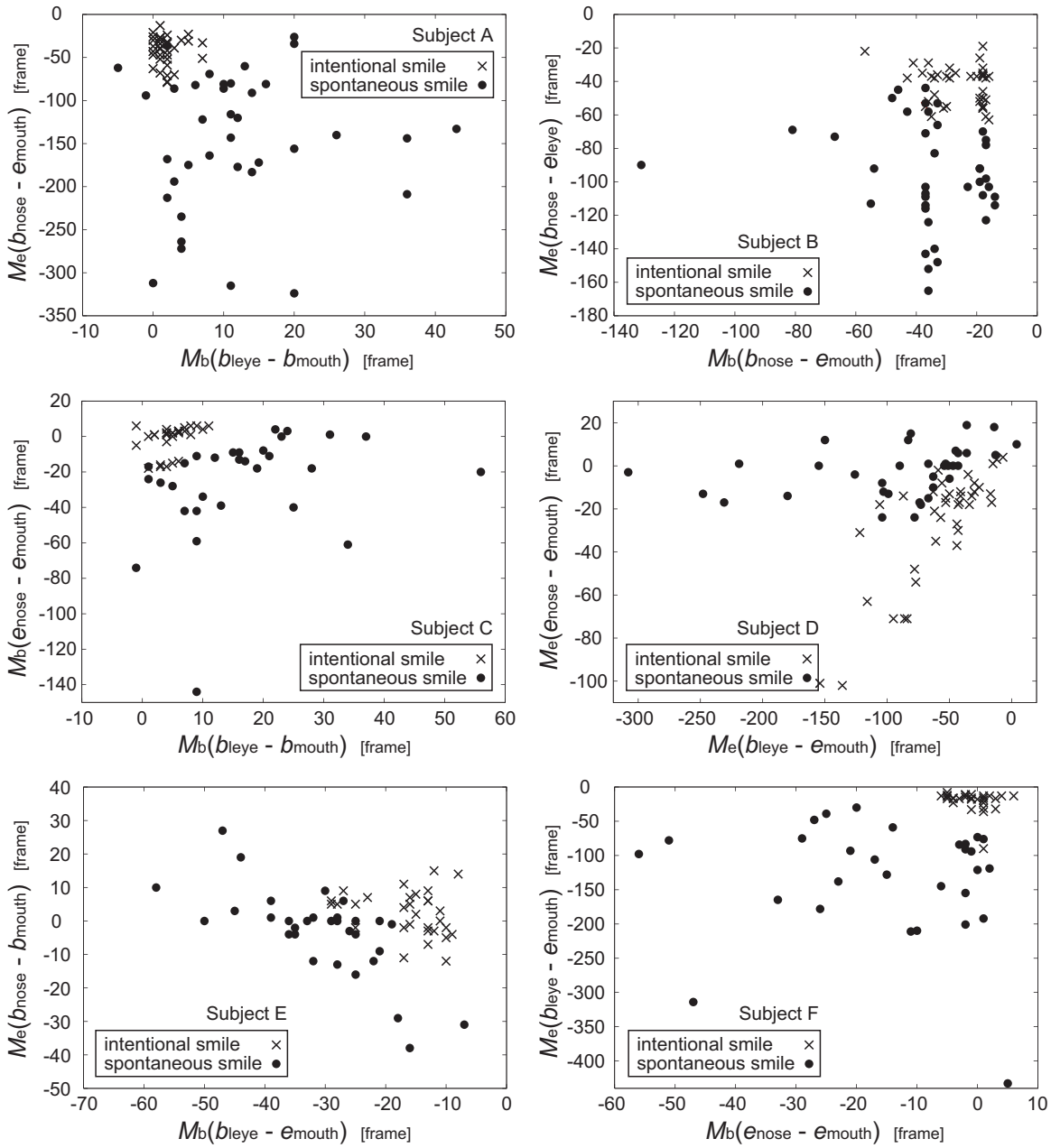


図 10　意図的・自発的な笑いの分布間の距離が最大となったタイミング構造
Fig. 10　Timing structure with the longest distance between intentional and spontaneous smiles distribution.

まず，2.4 節に従い，2 つの異なる顔パーツのタイミング構造の 1 次元空間分布での分析を行ったが，全ての被験者で 2 つの笑いの表情を明確に分離できる空間は存在しなかった．そこで，空間を拡張し，2 次元空間の分布での分析を行った．2 つの異なる顔パーツの区間の始点と終点の時間関係から構成し得る全ての 2 次元空間に対して，各表情の分布間のマハラノビス汎距離 [22] を求め，その距離が最大となった空間を図 10 に被験者毎に示す．これらの図より，各被験者で各表情に対応したクラスタが形成される空間が存在することを確認できる．

図 10 からは，被験者毎に空間を構成する軸が異なることも確認できる．この結果から，意図的な笑いと自発的な笑いを分離できるタイミング構造に個人差が検出されたと考えられる．従来の表情理解や認知に関する研究では，主として被験者間で共通性のある要素が探し求められてきた．しかし，本実験の結果からは，表情によっては個性を考慮する必要性が示唆される．なお，被験者 A と C では，笑い開始時の左目と口パーツのモードの始点の差 $M_b(b_{\mathrm{leye}} - b_{\mathrm{mouth}})$ が各表情のタイミング構造の差をもたらす軸となっている．これは，Nishio ら [6] によって明らかにされた表情認知の

知見と一致する．ゆえに，被験者 A と C は意図的な笑いと自発的な笑いの違いを認知され易いという可能性がある．

4.4 タイミング構造に基づく意図的な笑いと自発的な笑いの識別

図 10 の空間上で意図的な笑いと自発的な笑いの識別実験を行った．実験は，各被験者毎に，図 10 の空間上の全てのデータに対して leave-one-out 法 [22] を適用し，Support Vector Machine [23] [7]によって識別境界の学習とテストデータの識別を行った．この実験で得られた各被験者毎の識別率と全被験者の平均識別率を表 1 に示す．この表から，全ての被験者で高い識別率が得られていることが確認できる．ゆえに，図 10 に示す空間が意図的な笑いと自発的な笑いを良く分離する空間であると評価できる．

さらに，ここで，従来の表情認識手法との比較評価を行う．比較手法として，顔認識に広く用いられている部分空間法を拡張させた相互部分空間法 [25] を用いる．この手法は時系列データの照合に適用することができる．この手法を特徴ベクトル z の系列に適用することで得られた全被験者の平均識別率は，意図的な笑いに対して 97.9%，自発的な笑いに対して 83.0% であった．意図的な笑いに比べ，自発的な笑いの識別精度が大きく低下した．表 10 と比較すると，意図的な笑いは提案手法より相互部分空間法によって精度良く識別され，自発的な笑いは提案手法によって相互部分空間法より精度良く識別されたことがわかる．これは，以下のように考察できる．本実験では，意図的な笑いは表出毎の変動が小さく，入力系列全体を通じてほとんど同じ動きのパターンとなった．統計的手法である相互部分空間法はこのような分散が小さなデータ系列を精度良く表現することができる．これに対し，自発的な笑いは，表出毎の顔パーツ間の時間的構造のばらつきが大きかったため，相互部分空間法で表現することが困難であったと考えられる．一方，提案手法はその時間的構造を詳細に表現するため，自発的な笑いも精度良く識別することができる．

以上のように，表情譜から抽出される時間差の分布に基づくタイミング構造によって，意図的な笑いと自発的な笑いの微妙な差異が抽出されることから，表情譜は表情の時間的変化の詳細な表現形式として有効であると言える．

5. 結論

本論文では，表情変化を顔パーツそれぞれの時間的な運動によって生じるものと考え，そのタイミング構造を記述する表現形式として「表情譜」を定義し，表情の動的な側面を表現・理解する枠組を提案した．

そして，実際に撮影した顔画像系列から表情譜を自動獲得する手法を提案し，顔パーツの運動が，線形動的システムで表される区間を単位として記述可能であることを示した．さらに，この表情譜を用いることで，意図的な笑いと自発的な笑いが分離されることを確認し，表情変化における時間的要因の差異を理解するために，タイミング構造を用いることが有効であることを示した．

本論文では，2 つの区間のタイミング構造を 2 次元空間での分布から考察したが，表情をより詳細に分析・理解するために，3 つ以上の区間の 3 次元以上の空間での分析へと拡張することが可能である．また，個人差についての検討を行ったが，被験者間で共通の分布を持つタイミング構造が存在するのであれば，その検出に取り組むことも重要である．共通性のある構造が存在すれば，それを未知人物の表情の認識に適用することができる．さらに，本論文ではタイミング構造のみに注目して表情の分類を試みたが，実際には顔パーツの動きの大きさや加速度など顔形状の変化に関わる要因にも重要な情報が存在すると考えられる．これは個性を表現する上でも重要な要素である．

今後はこれらの問題に取り組むとともに，より多様な表情から得られる表情譜を分析することでより詳細な表情の理解を目指す．特に，ヒューマンコミュニケーションにおける，二者以上の間での表情の関係や，表情・ジェスチャ・会話といった様々な要素が複雑に絡む時間関係のモデル化に，表情譜とそのタイミング構造を応用したいと考えている．

表 1　図 10 のタイミング構造に基づく意図的・自発的な笑いの識別の精度
Table 1　Accuracy of discrimination between intentional and spontaneous smiles based on timing structures in fig. 10.

subject	**intentional** (%)	**spontaneous** (%)
A	100	83.8
B	100	79.4
C	82.4	96.4
D	85.1	79.7
E	85.3	90.3
F	96.6	93.1
ave.	91.6	87.1

謝辞

本研究の一部は，科学研究費補助金 18049046 の補助を受けて行った．

7: SVM light [24] を用いて，線形識別境界面を学習した．

参考文献

[1] Essa, I. A., Pentland, A. P.: Facial Expression Recognition using a Dynamic Model and Motion Energy; Proceedings of the 5th IEEE International Conference on Computer Vision, pp. 360–367 (1995).

[2] Schmidt, K. L., Cohn, J. F., Tian, Y.: Signal Characteristics of Spontaneous Facial Expressions: Automatic Movement in Solitary and Social Smiles; Biological Psychology, Vol. 65, pp.49-66 (2003).

[3] 内田英子, 四倉達夫, 森島繁生, 山田寛, 大谷淳, 赤松茂: 高速度カメラを用いた顔面表情の動的変化に関する分析; 電子情報通信学会技術研究報告 HIP99-76, pp. 1–6 (2000).

[4] 蒲池みゆき, 吉川左紀子, 赤松茂: 変化速度は表情認知に影響するか? - 動画刺激を用いた顔表情認知の時間特性の解明; 電子情報通信学会技術研究報告 HCS98-34, pp. 17–24 (1998).

[5] Krumhuber, E., Kapps, A.: Moving Smiles: The Role of Dynamic Components for the Perception of the Genuineness of Smiles; Journal of Nonverbal Behavior, Vol. 29, No. 1, pp. 3–24 (2005).

[6] Nishio, S., Koyama, K., Nakamura, T.: Temporal Differences in Eye and Mouth Movements Classifying Facial Expressions of Smiles; Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 206–211 (1998).

[7] Fasel, B., Luettin, J.: Automatic Facial Expression Analysis: A Survey; Pattern Recognition, Vol. 36, pp. 259–275 (2003).

[8] Black, M. J., Yacoob, Y.: Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion; International Journal of Computer Vision, Vol. 25, No. 1, pp. 23–48 (1997).

[9] Essa, I. A., Pentland, A. P.: Coding, Analysis, Interpretation, and Recognition of Facial Expressions; IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 757–763 (1997).

[10] Otsuka, T., Ohya, J.: Recognizing Abruptly Changing Facial Expression from Time-Sequential Face Images; Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 808–813 (1998).

[11] Cohen, I., Sebe, N., Garg, A., Chen, L. S., Huang, T. S.: Facial Expression Recognition from Video Sequences: Temporal and Static Modeling; Computer Vision and Image Understanding, Vol. 91, No. 1-2, pp. 160–187 (2003).

[12] Ekman, P., Friesen, W. V.: Unmasking the Face; Prentice Hall (1975).

[13] Tian, Y.-L., Kanade, T., Cohn, J. F.: Recognizing Action Units for Facial Expression Analysis; IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp. 97–115 (2001).

[14] Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces; Proceedings of ACM SIGGRAPH'99, pp. 187–194 (1999).

[15] Allen, J. F.: Maintaining Knowledge about Temporal Intervals; Communications of the ACM, Vol. 26, No. 11, pp. 832–843 (1983).

[16] Pinhanez, C., Bobick, A.: Human Action Detection using PNF Propagation of Temporal Constraints; Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 898–904 (1998).

[17] Cootes, T. F., Edwards, G. J., Taylor, C. J.: Active Appearance Model; Proceedings of European Conference on Computer Vision, Vol. 2, pp. 484–498 (1998).

[18] Stegmann, M. B., Gomez, D. D.: Brief Introduction to Statistical Shape Analysis; Informatics and Mathematical Modelling, Technical University of Denmark (2002).

[19] Kawashima, H., Matsuyama, T.: Multiphase Learning for an Interval-based Hybrid Dynamical System; IEICE Transactions on Fundamentals, Vol. E88-A, No. 11, pp. 3022–3035 (2005).

[20] Gross, J. J., Levenson, R. W.: Emotion Elicitation Using Films; Cognition and Emotion, Vol. 9, pp. 89–108 (1995).

[21] Stegmann, M. B., Ersboll, B. K., Larsen, R.: FAME - A Flexible Appearance Modelling Environment; Informatics and Mathematical Modelling, Technical University of Denmark (2003).

[22] 石井 健一郎, 上田 修功, 前田 英作, 村瀬 洋: わかりやすいパターン認識; オーム社 (1998).

[23] Vapnik, V.: The Nature of Statistical Learning Theory; Springer Verlag (1995).

[24] Joachims, T.: Making Large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning; MIT-Press(1999).

[25] 西山 正志, 山口 修, 福井 和広: 制約相互部分空間法を用いたジェスチャー認識; 第 10 回画像センシングシンポジウム講演論文集, pp. 439–444 (2004).

付録

1. 動的システムの階層的クラスタリング

Algorithm 1 Agglomerative Hierarchical Clustering of Dynamical Systems

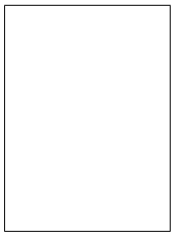
for $i \leftarrow 1$ to N **do**
  $M_i^{(a)} \leftarrow \mathsf{Identify}\left(I_i^{(a)}\right)$
end for
for all pair$\left(M_i^{(a)},\ M_j^{(a)}\right)$ where $M_i^{(a)},\ M_j^{(a)} \in \mathcal{M}^{(a)}$ **do**
  $\mathsf{D}\,(i,\ j) \leftarrow \mathsf{CalcDistance}\left(M_i^{(a)},\ M_j^{(a)}\right)$
end for
while $N \geqq 2$ **do**
  $(i^*,\ j^*) \leftarrow \arg\min_{(i,\ j)} \mathsf{D}\,(i,\ j)$
  $\mathcal{I}_{i^*}^{(a)} \leftarrow \mathsf{MergeIntervals}\left(\mathcal{I}_{i^*}^{(a)},\ \mathcal{I}_{j^*}^{(a)}\right)$
  $M_{i^*}^{(a)} \leftarrow \mathsf{Identify}\left(\mathcal{I}_{i^*}^{(a)}\right)$
  erase $M_{j^*}^{(a)}$ from $\mathcal{M}^{(a)}$
  $N \leftarrow N - 1$
  for all pair$\left(M_{i^*}^{(a)},\ M_j^{(a)}\right)$ where $M_j^{(a)} \in \mathcal{M}^{(a)}$ **do**
    $\mathsf{D}(i^*,\ j) \leftarrow \mathsf{CalcDistance}\left(M_{i^*}^{(a)},\ M_j^{(a)}\right)$
  end for
end while

動的システムの階層的クラスタリングを Algorithm 1 に示す．$M^{(a)}$ や $I^{(a)}$ 等に見られる添字 a は顔パーツ P_a のものであることを示す．Identify は，3.2.1 節で述べたシステム同定法を表し，区間内にある特徴ベクトル系列を用いて，モードのモデルパラメタ $\theta_i^{(a)} = \left\{F^{(a,\,i)},\ f^{(a,\,i)},\ Q^{(a,\,i)},\ z_{init}^{(a,\,i)}\right\}$ を同定する．階層的クラスタリングでは，時間的に離れた位置にある（互いに重なりを持たない）区間であっても，同じモードで表現されることがある．そこで，モード $M_i^{(a)}$ によって表現される区間の集合を，集合 $\mathcal{I}_i^{(a)}$ で表している．CalcDistance は，モード間の距離を求める処理であり，その距離は 3.2.3 節で定義している．MergeIntervals によって 2 つの区間集合は併合され，得られた区間集合からモードのモデルパラメタを再同定する．

（2002 年 1 月 1 日受付，1 月 1 日再受付）

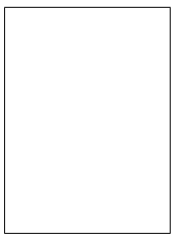
著者紹介

平山　高嗣　（正会員）



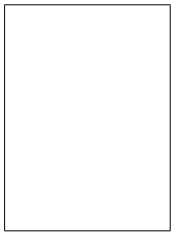
2000 年金沢大学工学部電気情報工学科卒業．2002 年大阪大学大学院基礎工学研究科修士課程修了．2005 年同大大学院博士課程修了．2005 年より京都大学大学院情報学研究科特任助手．博士（工学）．顔画像認識，ヒューマンコンピュータインタラクションの研究に従事．

川嶋　宏彰　（正会員）



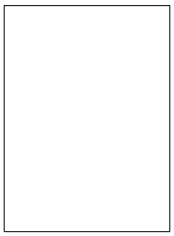
2001 年京都大学大学院情報学研究科修士課程修了．2002 年同大大学院博士課程中退．2002 年より同大大学院助手（現，助教）．博士（情報学）．時系列パターン認識，メディア統合，ハイブリッド・ダイナミカル・システム，実世界インタラクションの研究に従事．2004 年 FIT 論文賞，2005 年 FIT 船井ベストペーパー賞．

西山　正紘



2005 年京都大学工学部電気電子工学科卒業．2007 年同大大学院情報学研究科修士課程修了．メディア情報処理に興味を持つ．2005 年情報処理学会 CVIM 卒論セッション優秀賞．

松山　隆司



1976 年京大大学院修士課程修了．京大助手，東北大助教授，岡山大教授を経て 1995 年より京大大学院電子通信工学専攻教授．現在同大学院情報学研究科知能情報学専攻教授．2002 年学術情報メディアセンター長，京大評議員．2004 年情報環境機構長．工博．画像理解，分散協調視覚，3 次元ビデオの研究に従事．最近は「人間と共生する情報システム」の実現に興味を持っている．1980 年情報処理学会創立 20 周年記念論文賞，1990 年人工知能学会論文賞，1993 年情報処理学会論文賞，1994 年電子情報通信学会論文賞，1995 年第 5 回国際コンピュータビジョン会議 Marr Prize，1999 年電子情報通信学会論文賞，2000 年画像センシングシンポジウム優秀論文賞，2004 年，2005 年 FIT 優秀論文賞．IAPR，情報処理学会，電子情報通信学会フェロー．日本学術会議連携会員．