

表情認識のための感情空間の構成

5P-6

松野 勝弘* 李 七雨* 辻 三郎**

*(財)イメージ情報科学研究所

**大阪大学基礎工学部

1. はじめに

人間同士のコミュニケーションにおいて、人間の顔に含まれる情報は非常に重要な役割を演じている。自分の対面している相手を誰であるかを知り、その感情や意図を理解することはコミュニケーションの成否を決定する上で重要である。従って、顔に含まれるこれらの情報を計算機で認識することができれば、ユーザフレンドリなヒューマンインターフェースを実現することが可能となり、顔認識に関する研究は、ますますその重要性を増している。本稿では、一枚の正面顔画像から表情を認識する手法について述べる。

2. ポテンシャルネットによる特徴抽出⁽¹⁾⁽²⁾

2.1 ポテンシャルネットの構造

ポテンシャルネットは、2次元グリッド状に各ノードの分布がある系のポテンシャルを表す物理モデルである。その構造を図1に示す。最外郭のノードは枠に固定され、その他のノードは4近傍のノードにバネを通して結合されている。従って、あるノードの位置変化はネット全体の構造に影響を及ぼす。このネットを微分顔画像上に置き、ネットを変形させ、特徴抽出を行う。ポテンシャルネットの変形手法については2.2で詳しく述べる。

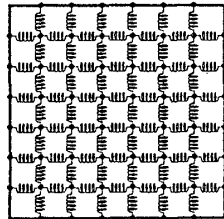


図1 ポテンシャルネットの構造

2.2 ポテンシャルネットの変形

ポテンシャルネット上に設けられた有限個のノードにノード番号($i=1\sim L, j=1\sim M$)を付け、各ノードを $N(i,j)$ と表し、その座標を $\mathbf{n}=(x_{i,j}, y_{i,j})^T$ とする。適当な外力が加えられた場合の各ノードの運動方程式は以下の式で表される。

$$m_{i,j} \frac{d^2 \mathbf{n}_{i,j}}{dt^2} + \gamma \frac{d\mathbf{n}_{i,j}}{dt} + \mathbf{F}_{spring} = \mathbf{F}_{ext} \quad (1)$$

ここで、 $m_{i,j}$ はノードの質量、 γ は減衰係数、 \mathbf{F}_{spring} は内力(以後弾性力と呼ぶ)、 \mathbf{F}_{ext} は外力(以後画像力と呼ぶ)である。

この画像力を用いて、運動方程式(1)を解くことになるが、運動方程式(1)を動的問題として解を求めようとすると各項は時間的に変化するため非常に計算コストがかかる。そこで、静的問題として考え、運動方程式(1)の平衡状態

Emotional Space for Recognizing Facial Expressions
Katsuhiro MATSUNO*, Chil-Woo LEE*, Saburo TSUJI**
*Laboratories of Image Information Science and Technology
**Osaka University

を考える。

$$\mathbf{F}_{spring} = \mathbf{F}_{ext} \quad (2)$$

式(2)は、弾性力と画像力のある瞬間の平衡状態を表している。

ノード $N(i,j)$ に働くバネの弾性力は式(3)で表される。

$$\mathbf{F}_{spring} = k \sum_a (|l_{i,j}^a| - l_0) \frac{l_{i,j}^a}{|l_{i,j}^a|} \quad (3)$$

$$l_{i,j}^1 = \mathbf{n}_{i,j} - \mathbf{n}_{i+1,j}, \quad l_{i,j}^2 = \mathbf{n}_{i,j} - \mathbf{n}_{i-1,j}$$

$$l_{i,j}^3 = \mathbf{n}_{i,j} - \mathbf{n}_{i,j+1}, \quad l_{i,j}^4 = \mathbf{n}_{i,j} - \mathbf{n}_{i,j-1}$$

ここで、 k はバネの弾性定数、 l_0 はノード間に連結されているバネの自然長である。式(3)は、1つのノードに働く4方向に連結されたバネの弾性力の合力を求めている。

ノード $N(i,j)$ に働く画像力は式(4)で表す。

$$\mathbf{F}_{ext} = \alpha (\nabla(G_\sigma * I(x_{i,j}, y_{i,j}))) \quad (4)$$

ここで、 α は重み係数、 $I(x_{i,j}, y_{i,j})$ はノード $N(i,j)$ の位置にある微分画像の濃度値である。また、 G_σ は、幅 σ を持つ2次元ガウシアンフィルタによるコンボリューションを表す。従って、式(4)は、微分画像 I を標準偏差 σ のガウシアンフィルタでぼけ変換を行い、ポテンシャル場を作成した後、その傾斜を計算することによって画像力を求めている。

次に、式(2)を変形して、式(5)により合力 \mathbf{F}_{total} を定義する。

$$\mathbf{F}_{total} = \mathbf{F}_{ext} - \mathbf{F}_{spring} \quad (5)$$

ノードの合力が閾値以下であればそのノードは釣り合っているとみなし動かさない。一方、ノードの合力が閾値以上であれば8方向に近似してノードの合力が小さくなる方向にノードを移動させる。この処理を全ノードに対して繰り返し実行し、各ノードの安定な位置を求める。初めは各ノードの力のバランスが崩れて合力の平均値は増加するが、繰り返し処理を実行することによりその値が小さくなりネットは収束する。収束したネットの各ノードの初期位置 $\mathbf{n}_{i,j}^0$ から安定位置 $\mathbf{n}_{i,j}^c$ までの各ノードの移動量を変位ベクトル $\mathbf{V}_{i,j}$ として表す。

$$\begin{aligned} \mathbf{V}_{i,j} &= \mathbf{n}_{i,j}^c - \mathbf{n}_{i,j}^0 \\ &= (x_{i,j}^c, y_{i,j}^c)^T - (x_{i,j}^0, y_{i,j}^0)^T \\ &= (\Delta x_{i,j}, \Delta y_{i,j})^T \end{aligned} \quad (6)$$

変形したネットの形状には顔の特徴がよく反映されているため、顔の特徴を示すネットの各ノードの変位ベクトルを認識特徴として使用する。

3. KL展開による感情空間の構成

3.1 本方式の概要

本稿で提案する認識手法の手順を以下に示す。

(1) 顔領域の抽出と正規化

入力画像からマウスで、顔領域を指定する。顔領域は、眉の両端、眉の上端、口の下端により位置決めを行って切り出される。次に、切り出された顔領域を同じ大きさに正規化する。

(2) 前処理

画像内に存在する様々な雑音を除去するために3×3のメディアンフィルタを適用し、顔画像を平滑する。その顔画像から輪郭を抽出するために、微分フィルタを用いて各点の微分値を計算する。

(3) 変位ベクトルの抽出

図1の初期ネットの最外郭より1つ内側のノードの枠と正規化された顔領域を基準にしてネットを置き、2.2の手法によりネットを変形させ、各ノードの変位ベクトルを求める。

(4) KL展開⁽³⁾

予め作成しておいた表情照合ネットの各ノードの変位ベクトルを標本集合としてKL展開を行い、固有ベクトルを求める。上位数個の固有ベクトルが標本集合をKL展開することにより得られた低次元の正規直交基底となる。この正規直交基底を用いて低次元のカテゴリ空間（以後感情空間と呼ぶ）を構成する。入力顔画像より求めたネットの各ノードの変位ベクトルをこの空間に写像し、表情照合ネットの特徴ベクトルとの距離を求め、最小距離判別法により入力顔画像の表情認識を行う。

3. 2 変位ベクトルのKL展開

認識に有効な低次元の特徴ベクトルを獲得するために、ポテンシャルネットにより抽出した各ノードの変位ベクトルにKL展開を適用する。Q個の表情照合ネットの標本集合を{V_d}とする。V_dは表情照合ネットdの各ノードの変位ベクトルの成分である。ここで、dは表情の種類を表す。表情照合ネットは、ノード間の距離l₀=3、バネの弾性定数k=2.5、重み係数α=0.5のパラメータ値を用いて、10人の表情ネットより作成した平均表情ネット(幸福、怒り、驚き、悲しみ)である。これらのパラメータ値は予備実験に基づいて決定した⁽⁴⁾。以下にKL展開の手順を示す。

(1) Q個の表情照合ネットのV_dの平均μを計算する。

(2) 表情照合ネットの集合{V_d}の共分散行列Cを次式より計算する。

$$C = (1/Q) \sum_{d=1}^Q (V_d - \mu)(V_d - \mu)^T \quad (7)$$

(3) 共分散行列Cの固有値λ_p、固有ベクトルU_r(以降固有ネットと呼ぶ)を計算する。

(4) 固有ネットの集合{U_r}を用いて次式によりG次元ベクトルR=(r₁, r₂, ..., r_G)が表情照合ネットV_dに対する特徴ベクトルとなる。

$$r_f = U_r^T (V_d - \mu) \quad f=1, \dots, G \quad (8)$$

3. 3 感情空間の構成

3.2のKL展開手順により、固有ネットを求め、この固有ネットの中で、固有値の大きい順に上位G個を用いて式(8)により、表情照合ネットに対するG次元の特徴ベクトルを求める。そして、G個の正規直交基底で構成される感情空間を作成し、この空間に表情照合ネットと入力ネットを写像し、この空間内で認識処理を行う。図2に表情照合ネットを写像した感情空間を示す。次元数は、感情空間に

写像された表情照合ネット間の分離性より、G=3とした。

認識処理は、入力画像から求めたネットの特徴ベクトルR_r=(r₁, r₂, ..., r_G)とクラスdの表情照合ネットの特徴ベクトルR_{df}=(r_{d1}, r_{d2}, ..., r_{dG})を全クラスに対して、次式により距離Dを求め、その最小値を与えるクラスを認識された表情とする。

$$D_d = \sum_{f=1}^G |R_{df} - R_r| \quad (9)$$

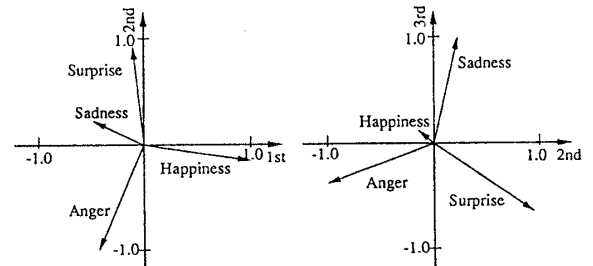


図2 感情空間

4. 実験結果

本実験では、表情照合ネットの作成に用いなかった21人の44枚の顔データ(幸福、怒り、驚き、悲しみ)について式(9)を用いて認識実験を行った。その実験結果を表1に示す。

表1 未知の顔画像に対する認識実験(単位:人)

(a) 本手法の出力 (b) 入力顔画像

(a) (b)	幸福	怒り	驚き	悲しみ
幸福	11	0	0	1
怒り	0	10	0	0
驚き	0	0	11	2
悲しみ	0	1	0	8
合計	11	11	11	11

5. おわりに

本稿では、ポテンシャルネットとKL展開を組み合わせた表情認識手法について述べた。幸福、怒り、驚き、悲しみの4種類の表情の認識実験で、90%以上の認識率を得ることができ、本手法の有効性を示した。

参考文献

(1) 松野勝弘, 李七雨, 辻三郎: "ポテンシャルネットを用いた顔表情の認識", 信学技報, PRU93-38 (1993)
 (2) Matsuno.K., Lee.C. and Tsuji.S. "Recognition of Facial Expression with Potential Net", Proc.ACCV, pp.504-507 (1993)
 (3) Turk.M.A. and Pentland.A.P.: "Face Recognition Using Eigenface", Proc.CVPR, pp.586-591 (1991)
 (4) 松野勝弘, 李七雨, 辻三郎: "ポテンシャルネットとKL展開を用いた顔表情の認識", 信学論(D-II)に投稿中