H-049

多重分解処理を用いたジェスチャからの感情認識

A multi-factorized method for emotion recognition from human gestures

苗村 昌秀† 高橋 正樹† 藤井 真人† 八木 伸行†

Masahide Naemura Masaki Takahashi Mahito Fujii Nobuyuki Yagi

1．まえがき

近年，メディア処理のあり方として，コンテンツの意味的要素までを特定できる高次なセマンティック処理への要求が高まっている．とくに，我々が推進する番組映像からのメタデータ抽出の研究においては，人に感動を与えるようなシーンの抽出には，セマンティック性の高いメタデータの抽出が必要となってくる．そこで，このような高次メタデータ抽出を可能にする手段として，人間のジェスチャからの感情認識の研究に取り組んでいる．

人間のジェスチャ認識の研究は比較的広く行われている[1][2]が，ジェスチャからの感情認識の研究についてはまだ十分な検討が行われておらず，未開発の分野である．ただし，いろいろな応用分野で多種多様な方法で人間の内面的状態を探るといった研究は重要になってきており，すでに顔表情からの感情認識の研究が高い注目を浴びている[3][4]ことからもジェスチャからの感情認識の研究も同じように盛んになってくることが予想される．

顔表現からの感情認識に比して，ジェスチャからの感情認識の場合，その動き特徴量の自由度が高く，さらにそれらが 3 次元空間上を動き回るため，効果的な特徴量の抽出，認識処理が困難である．そこで，我々は，時系列の動き特徴量を多重分解処理で 3 次元テンソル時系列信号に変換後に HMM を用いる，新しい感情認識手法を考案したので，報告する．

2．関連研究

現状では，ジェスチャからの感情認識の研究については心理的な側面からの研究が主流で，感情を識別するような物理的特徴量を抽出するといった，工学的な見地からの研究はほとんど見当たらない．これは，ジェスチャにおける動作パラメータの自由度が高く，それらパラメータのどの部分が感情に関連しているかの基本的な知見が不足しているためである．ただし，近年，感情までの感性的な情報抽出まではいかなくても，人間の動作姿勢の識別の研究は盛んになりつつある[5][6]．これらの研究では，入力映像から識別精度の高い特徴量を生成し，その時系列信号を HMM に入力して人物特定や姿勢種別などを検出するものが大半である．今後は，これらの研究を発展させ，より高次な身体動作からの感情認識の研究へと進んでいくものと思われる．

コンピュータビジョン以外の分野では，人間の感情動作を活用する研究として CG におけるアニメーション生成の研究で先進的なものが多く見受けられる．[7][8]の研究では，CG キャラクタの動きを事前に定義した感情に基づいて変形させる内容である．ここでの感情による動きは，CG キャラクターの動きパラメータに摂動を与えて実現するものであり，統計的には保障されていないが一定の感情らしき動きを実現している．また，最近の CG 研究では，動きの微妙なスタイルの違いを数学的に記述し，その数学的パラメータを調整することで微妙な動きの違いを与えるといった試みが盛んに行われている[9][10]．[9]の研究では，歩き方の違いを統計的に学習し，その結果を HMM パラメータで記述することにより，所望の歩き方の時系列モーションデータを生成している．また，[10]の研究では，いろいろなポーズからの Inverse Kinematics を学習して，学習結果を確率モデルで記憶させることにより，学習されていないポーズについても確率的な内挿処理でより自然なアニメーション生成を実現している．これらの CG 生成の技術では，異なる動きの違いを数学的に記述してアニメーション生成するものであり，感情の違いでアニメーション生成するといったものでない．しかし，ジェスチャからの感情認識の研究により感情表現を決定付ける動作特徴量が明らかになれば，CG 研究の成果と組み合わせて CG キャラクタに感情動作を与えるような仕組みを実現できることが期待できる．

そこで，我々は単純にジェスチャからの感情を分類して認識するだけでなく，感情認識によって得られた特徴量の再利用を考慮して，以下の章で述べるように入力の動き特徴量を解析して感情認識に適した特徴量へ変換し，学習・認識する手法を考案した．

3. 提案手法

感情認識過程で得られた特徴量を再利用して新しく感情表現ジェスチャを再現するためには，認識アルゴリズム自体がモデル生成できる動的なものでなければならない．このことを考慮して，提案手法ではパラメータからモデル生成可能な HMM をベースにした．以下に提案手法について述べる．

3.1 概要

図 1 に提案手法の処理の全体の流れを示す．入力は，モーションキャプチャにより取得した人体の各関節部の動き特徴量で，具体的には関節ごとに角度情報より求めたクオータニオンとそのクオータニオンを時間微分した 8 次元の時系列ベクトルである．入力された動き特徴量 M[176]は，

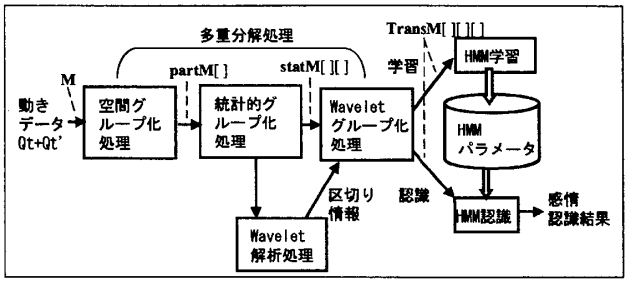

図 1 提案手法の全体の流れ

† 日本放送協会 放送技術研究所

空間グループ化処理，統計的グループ化処理，および，Wavelet グループ化処理で順に **partM[]**, **statM[][]**, **TransM[][][]**に空間的，統計的，および，周波数空間的に多重分解処理され，HMM 学習・認識用の 3 次のテンソル時系列信号に変換される．また，Wavelet グループ化処理と並行した Wavelet 解析処理により HMM 認識に適した時間区切り情報を自動検出して，学習データの蓄積を行っているのも提案手法の特徴である．このようにして生成された 3 次テンソル時系列信号を HMM 認識することによりジェスチャからの感情認識を実現している．

3.2 多重分解処理

(1)空間(部位)グループ化処理

空間グループ化処理では，人体の各関節から抽出した動き特徴量を複数の主要部位でまとめてグループ化する．図 2 に提案手法でのグループ化部位を示す．図に示されているように動き特徴量は，以下の 5 つの部位グループにグループ化される．

BB={Hips,Spine,Spine1,Spine2,Neck,Head}
RA={RightShoulder,RightArm,RightForeArm,RightHand}
LA={LeftShoulder,LeftArm,LeftFroeArm,LeftHand}
RL={RightUpLeg,RightLeg,RightFoot,RightToes}
LL={LeftUpLeg,LeftLeg,LeftFoot,LeftToes}

この結果，入力の 22×8=176 次元の動き特徴量 **M** は，BB については 48 次元，RA,LA,RL,LL については 32 次元の動き特徴量 **partM**[5]として個別に扱うことが可能となる．

(2)統計的グループ化処理

空間グループ化処理でグループ化された 5 つの動き特徴量を部分空間法[11]で統計的に次元変換処理しグループ化する．部分空間法で用いる部分空間への射影ベクトルは，全ての感情について集めた学習シーケンスから動き特徴量を収集してマトリックスに格納して，そのマトリクスを SVD することにより求めた．次元圧縮するため，有効な射影ベクトルの集合の寄与率を基準に設定して，主要な N_D 個の部分空間への射影成分だけを選択する．

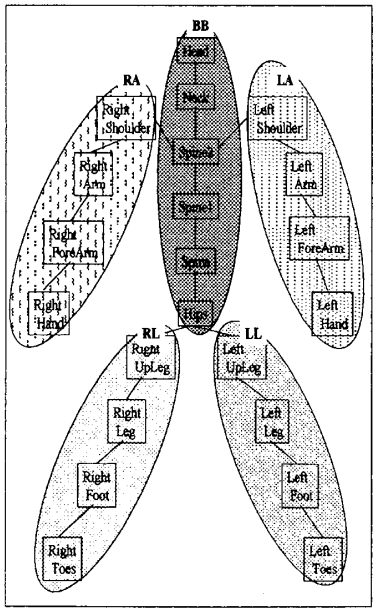

図 2　部位のグループ化

(3)Wavelet グループ化処理

Wavelet 解析が動作解析に有効である，という先行研究[12]の結果を参考に，空間グループ処理，統計的グループ処理後の特徴量を Wavelet 分解処理して，周波数空間でグループ化した．Wavelet 分解には Harr フィルターを用い,分解数 7 まで分解して，**TransM**[5][7][N_D]の 3 次テンソル特徴量にデータを格納する．

以上の 3 つのグループ化処理で構成される多重分解処理により，入力の 176 次元の時系列の動き特徴量 **M** は，3 次のテンソル時系列信号 **TransM**[5][7][N_D]に変換される．ここで，**TransM** の 1 次インデクスは 5 箇所の部位グループを，2 次インデクスは 7 階層の Wavelet 分解数を，3 次インデクスは，統計的グループ化処理で次元圧縮した N_D 次元の次元数を表す．提案手法では，このように生成した 3 次テンソル時系列から適切な部分空間を選んで感情認識を行うことになる．

3.3 Wavelet 解析による時間セグメンテーション

提案手法は HMM を基本としているので，感情ごとの効果的な学習データの収集がその性能を大きく左右する．しかし，入力の動きデータから感情をよく表現している区間を人手で区切ることは相当な労力を必要とし，効率的でない．そこで，Wavelet 解析により自動的に感情を表現している区間を時間セグメンテーションする手法を新たに開発した．図 3 にアルゴリズムのフロー図を示す．入力信号は，認識対象のジェスチャをよく表している部位グループの第 1 主成分にあたる要素の時系列信号である．今回の評価実験で用いた「歩く」ジェスチャでは，**statM**[RL][0] の信号となる．また，基本周期 T は，Wavelet 解析で周期性が最も高いレベル階層の信号を FFT 処理して求められる．そして，図 3 のフローに示されているように，求めた基本周期を基準に探索エリアを段階的に変えて，

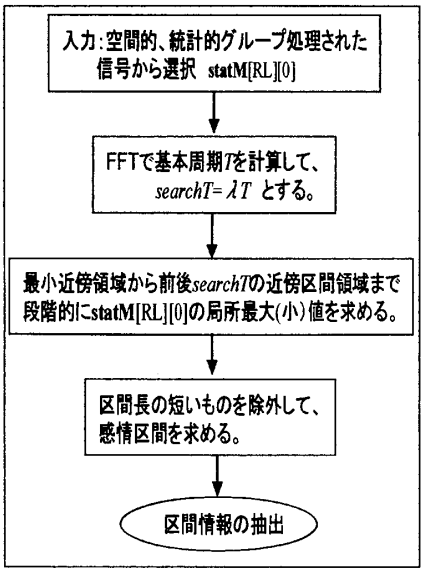

図 3　Wavelet 解析による時間セグメンテーションのフローチャート

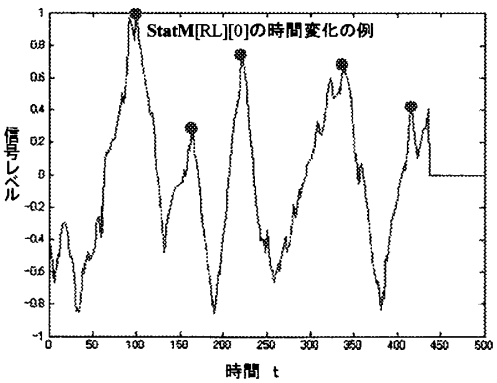

図 4　時間セグメンテーションの結果例

局所最大値を求めてその隣接区間を感情区間として出力していく．この手法では，信号に含まれている基本周波数に近い期間に存在する2つの局所最大(小)値間の区間はその一連の動作を良く表しているという仮定に基づいている．図4にこの処理を実際の信号に適用した場合の区間位置の検出例を示す．図から近接の局所最大値に影響されることなく，時間セグメンテーションができているのがわかる．

3.4 HMMによる感情認識

対象となる感情ごとに求められた3次テンソル時系列信号 **TransM** のシーケンス集合を HMM 学習して，それぞれの感情の HMM パラメータを抽出する．そして，未知のジェスチャからの動き特徴量に対し学習したパラメータで HMM 認識することにより，感情認識を実現する．認識過程は次式に従って決定される．

$$emotionID = \arg\max_{k}\{\lambda_k\}$$

λ_k：感情kの*HMM*認識における尤度

4. 実験と評価結果

4.1 モーションキャプチャ実験

オプティカルモーションキャプチャ装置（Vicon)を使って，演技者に感情をこめて演じてもらった演技からモーションデータを取得した．そのときの実験条件を以下に列挙する．

・演技者　　：女性3名，平均的サイズ
　　(160cm前後中肉中背)
・演技内容　：歩行
・演技回数　：1人の演技者が2回繰返す
・感情の種類：怒り，喜び，普通，悲しみ
・マーカー数：22箇所

取得した動き特徴量はAutoDesk社のMotionBuilder™で簡単な加工処理を行い，Fbxフォーマットでシーケンスごとにファイル化した．1シーケンスは約600フレーム長でこのシーケンスから前述したWavelet解析による時間セグメンテーション処理で歩行の基本セグメントを自動的に切り出した．このようにして得られた学習用セグメントの数は，感情ごとに以下のようになった．

怒り：40，喜び：40，普通：32，悲しみ：32

4.2 評価結果

入力の動き特徴量 **M**[176]を多重分解した3次テンソル **TransM**[][][]から認識用の特徴量を生成することが有効であることを確かめるために，条件の異なる特徴量で認識精度の比較評価を行った．HMM は連続 HMM を用い，その条件は，全ての感情について状態数 7 の Left-Right モデルで，Gaussian Mixture 数は 2 とした．また，HMM の学習には，怒り 40，喜び 40，普通 32，悲しみ 32 の 108 シーケンスに対し，k-fold Cross Validation(k=4)で評価を行った．

なお，今回の評価実験では，部位グループとして LL, RL，BB に限定して行った．これは，予備実験の段階で LA, RA を含めると全体として学習効果が低下することが確認され，現状では両手部分のジェスチャによる感情表現の多様性を吸収できるだけの学習データが不足している，との判断からである．

まず，部位グループとして，{LL, RR}を選択した場合の特徴量について以下に記す．(1-a)が提案手法の特徴量で(1－b)と(1－c)は比較用の特徴量である．

(1-a) 提案手法：**TransM** からの 8 次元データ
　　(部位 G∈{LL,RL})×(統計的 G∈{0, 1})×(Wavelet G∈{レベル 5, 6})
(1-b) 比較手法：入力動き特徴量 **M** からの 32 次元データ
　　(quaternion+微分 quaternion) × (LeftLeg, LeftFoot, RightLeg, RightFoot)
(1-c) 比較手法： **statM** からの 4 次元データ
　　(部位 G∈{LL,RL})×(統計的 G∈{0, 1})

この場合の感情認識の評価結果を表 1 に示す．表中の total rate は，全体の認識率を表し，全サンプル数に対する正解数の割合を示す．表から明らかなように入力信号を多重分解処理した特徴量の方が性能がよくなっている．とくに，Wavelet 分解で周波数的に分解することによる効果が顕著である．これは，感情固有の成分が周波数的なものであり，それが Wavelet 分解で分離できるため，認識結果の向上に表れているのではないか，と考えられる．

次に，部位グループに BB を追加して，{LL, RL, BB}とした場合の特徴量を列挙する．{LL, RL}の時と同じように(2-a)が提案手法，(2-b)，(2-c)が比較のための特徴量である．

(2-a) 提案手法：**TransM** からの 12 次元データ
　　(部位 G∈{LL,RL,BB})×(統計的 G∈{0, 1})×(Wavelet G∈{レベル 5, 6})
(2-b) 比較手法：一般的な次元圧縮データ
　　{LL,RL,BB}の 104 次元の特徴量 M[104]を直接，12 次元に次元圧縮したデータ
(2-c) 比較手法：(2-b)を Wavelet 分解した 12 次元データ
　　Wavelet 分解のレベル 5, 6 からのデータを選択して 12 次元化したデータ

この場合の感情認識の評価結果を表 2 に示す．(2-b)の特徴量は，一般によく用いられる手法で，部位グループ化せずに直接，高次特徴量を部分空間法で次元圧縮している．(2-a)と(2-b)の比較より，{LL, RL, BB} の場合でも提案手法の優位性が確認できる．また，(2-c)の手法は空間的に分解せずに，Wavelet グループ化処理を施した特徴量で，提案手法と同等の認識率が達成できている．ただし，(2-a)と(2-c)の認識率の標準偏差を求めたところ(2-a)が 0.03，(2-b)が 0.11 で提案手法の認識率の方が小さい値を示していることがわかった．これは，提案手法の方が学習セットに影響されずに安定して認識を行えたことを示すもので，空間的にグループ化することの効果を示すものと考えられる．

以上の結果より，感情認識ではデータを提案手法のように多重分解処理することが有効である，ことが確認された．

表1 {LL,RL}での評価結果

(1-a)の結果

	Precision	Recall
Angry	0.59	0.58
Happy	0.52	0.58
Neutral	0.64	0.66
Sad	0.86	0.75
Total Rate = 0.632		

(1-b)の結果

	Precision	Recall
Angry	0.47	0.60
Happy	0.30	0.48
Neutral	0.56	0.16
Sad	0.62	0.41
Total Rate = 0.406		

(1-c)の結果

	Precision	Recall
Angry	0.48	0.78
Happy	0.23	0.18
Neutral	0.47	0.66
Sad	1.00	0.13
Total Rate = 0.437		

表2 {LL,RL,BB}での評価結果

(2-a)の結果

	Precision	Recall
Angry	0.65	0.55
Happy	0.55	0.78
Neutral	0.88	0.72
Sad	0.96	0.84
Total Rate = 0.715		

(2-b)の結果

	Precision	Recall
Angry	0.47	0.40
Happy	0.34	0.50
Neutral	0.64	0.56
Sad	1.00	0.72
Total Rate = 0.535		

(2-c)の結果

	Precision	Recall
Angry	0.64	0.68
Happy	0.62	0.65
Neutral	0.80	0.75
Sad	0.87	0.82
Total Rate = 0.715		

5. まとめ

ジェスチャに伴う感情の認識を行う，といった工学的な研究があまり進んでいない研究テーマを取り上げ，その有効な認識手法の提案を行った．提案手法は，時系列の動き特徴量データを多重分解処理で 3 次テンソル時系列信号に変換して HMM で感情を認識することを特徴としたものである．多重分解処理は，動き特徴量を空間的に配置された人体グループ，それぞれの人体グループごとの統計的グループ，および，時系列データに含まれている周波数成分グループの性質の類似したグループでグループ化する処理である．実験結果より，多重分解処理を用いない場合に比べ，多重分解処理した場合，認識率が改善することが確認され，ジェスチャからの感情認識の研究に対する方向性を示すことができた．

今回の実験は，心理評価を実施せずに演者の感情演技をそのまま正解の感情として採用しているため，第 3 者のジェスチャからの感情認識判断と異なる可能性がある．より精度の高い感情認識手法の確立のためには，評価手法などの更なる検討が必要である，と考えている．

また，実験のデータ数が限られていたため，LA，RA の手の部分も含めた体全体の動き特徴量を用いた場合うまく改善効果を得ることができなかった．提案手法では，空間的に部位グループごとに別々に特徴量を管理しているので，部位グループ間のインタラクションを効果的に捉えることができる可能性が高い，と考えられる．今後の検討課題としたい．

[参考文献]

[1] C. Bregler, "Learning and Recognition Human Dynamics in Video Sequences," CVPR 1997, pp. 568-574,1997.

[2] A.D. Wilson and A.F. Bobick, "Parametric hidden Markov models for gesture recognition, " IEEE Trans. PAMI, vol.21, no.9, pp.884-900, 1999.

[3] Y. Yacoob and L.S. Davis, "Recognizing Human Facial Expressions from Long Image Sequences Using Optical-Flow," IEEE Trans. PAMI, vol.18, no.6, pp.636-642, 1996.

[4] T. Otsuka and J. Ohya, "Recognizing Abruptly Changing Facial Expressions from Time-Sequential Face Images," CVPR 1998, pp.808-813, 1998.

[5] L. Lee and W.E.L. Grimson, "Gait analysis for recognition and classification," Proceedings of the IEEE Conference and Face and Gesture Recognition, pp. 155–161, 2002.

[6] A.F. Bobick and A. Johnson, "Gait recognition using static activity-specific parameters," CVPR2001.

[7] M. Unuma, et.al., "Fourier Principles for emotion-based human figure animation," SIGGRAPH 1995, pp.91-96(1995).

[8] K. Amaya and et.al., "Emotion from motion," Graphic Interface 1996, pp. 222-229, 1996.

[9] M. Brand, et. al., "Style Machines," SIGGRAPH 2000, pp. 183-192, 2000.

[10] K. Grochow, et.al., "Style-Based Inverse Kinematics," SIGGRAPH 2004.

[11] T. Katayama, "Subspace Methods for System Identification," Springer-Verlag,2005.

[12] M. Tada and M. Naemura, " Dance Evaluation System based on Motion Analysis," GRAPP2006, pp.243-250, 2006.