

ウェーブレット変換の位相情報を用いた顔動画からの表情認識に関する検討

5N-9

大塚 尚宏 海老原 一之 大谷 淳
(株) ATR 知能映像通信研究所

1. はじめに

表情認識は、画像通信における知的な符号化^[1]や高度なマン-マシンインタフェースの実現のために重要な基礎技術である。例えば、通信会議システムにおいて表情認識が実現されれば、伝送すべき情報の高圧縮化が実現できる。また、対話型のシステム(例えば教育ソフト)では、ユーザの表情を理解して対話の流れを円滑にする高度な制御が実現できると考えられる。

筆者らは、顔動画からの表情認識を、隠れマルコフモデル(HMM: Hidden Markov Model)を用いる手法を提案した^[2]。動画のフレーム毎に特徴量ベクトルを求め、その特徴量ベクトルの時系列を生成する確率モデルであるHMMを用いて表情認識を行う。HMMは、トレーニングデータからの高速な学習が可能、時間的変化特徴の利用が可能、時間軸方向の伸縮に対する耐性がある等の特徴がある。

筆者らの従来のHMMを用いた表情認識法では、目および口の形状変化により空間周波数の電力が変化する点に着目し、ウェーブレット変換によりそれらを検出していた。しかし、空間周波数領域の電力からは顔の構成要素の移動情報が得られず、正確な認識が実現できないという問題があった。そこで、本稿ではウェーブレット変換の位相情報からエッジの移動量を算出して利用することにより、表情認識性能の向上を目指す手法を提案する。

2. ウェーブレット変換からの移動量の算出

ウェーブレット変換は、基本ウェーブレットから平行移動および拡大・縮小操作により得られる基底を用いて展開(変換)する手法である。直交性があり、基底が空間および周波数ともに局所的なので分解能が高い特徴がある。基本ウェーブレットを拡大(縮小)した基底からは低(高)周波成分が得られる。2次元のウェーブレット変換では、基本ウェーブレットとして x, y 軸方向それぞれのローパスフィルタ、ハイパスフィルタから x, y

軸方向毎にひとつずつ選んだ組みのうちローパスフィルタ同士の組みを除いた3つの組が使われる。3つの基底は、それぞれ x, y 軸および45度方向の周波数成分を検出する。

ウェーブレット変換値は画像中の濃淡パターンの移動に対応して変化する。画像中の濃度値とウェーブレット変換値との関係を示すために、濃度値を実軸、変換値を虚軸とする複素平面を図1のように定義する。いま、階段関数型のエッジ、即ち上(下)側が低濃度のエッジが、画像中の領域Pを上から下に移動して通過する場合を考えると、領域Pの濃度パターンの変化と複素平面における位置との関係は、図1における1→2→3(3→4→1)の軌跡となる。即ち、これらの軌跡は図1で矢印A(B)により示されるように、Pにおけるパターンが2(4)の時にウェーブレット変換値が最大(小)値ともつ円となる。ここで、位相変化をエッジの移動量と解釈すると、ウェーブレット変換からエッジの移動量の算出が可能となる。

階段関数型ではない一般のエッジの移動の場合では、複素面での軌跡の中心は原点とはならない。そこで、動画シーケンスの軌跡から円の中心を平均二乗法により求め、その中心を原点として位相を求めればよい。

3. 実験結果と考察

実験では、画像内での顔の位置が移動しないように、ヘルメットに固定されたカメラにより撮影された動画画像を用いた。1シーケンスは50~60フレーム(30フレーム/秒)で、各フレームからそれぞれ右目と口を含む各256×256ピクセルの2つの領域を切り出して、特徴量抽出のための画像処理を行った。基本ウェーブレットとしては32次のミラーフィルタを用い、4および5段目の変換値(16×16, 8×8行列)に対して本手法を適用した。得られた移動ベクトルから、 x, y 軸方向の移動ベクトルの平均値および移動ベクトルの x, y 軸方向の1次のモーメントを特徴量(合計24個)として算出した。表情の動画画像としては、同一人物から得られた6種類の基本表情(図2参照)それぞれ10シーケンスを用い、そのうち5シーケンスをHMMの学習用に、残り

¹A Study of Facial Expression Recognition from Image Sequence Using Phase Information of Wavelet Transform
Takahiro Otsuka, Kazuyuki Ebihara, and Jun Ohya
ATR Media Integration & Communications Research Lab.

の5シーケンスを評価用に用いた。

筆者らの従来法と、本稿の手法による実験結果をそれぞれ表1と表2に示す。表1(従来法)では、4および5段目の変換値の電力平均を特徴量(合計12個)として用い、表2では、本稿の手法により得られる特徴量を加えた特徴量(合計36個)を用いた。表の要素は、入力表情(列)に対して判定結果として得られた表情(行)の個数である。本稿の手法を用いることにより認識率は80%から90%に向上している。誤判定の原因としては、ウェーブレット変換の段数を大きくして低周波成分に感度を合わせているため、目および口の開閉に伴う微小な移動を検出できないことが考えられる。

4. まとめ

表情認識のための新手法として、ウェーブレット変換の位相情報から顔の構成要素の移動量を求め、それから得られる特徴量を使って表情認識する方法を提案した。実験結果から本手法を用いることにより10%の認識率の向上が得られた。今後は、ウェーブレット変換の段数の最適化を行うとともに、より多くのデータを用いて評価を行い本手法の性能を見極めたい。また、本手法はエッジの近傍のオプティカルフローを算出していると解釈できるので、従来のオプティカルフローアルゴリズムを用いた実験を行って比較したい。

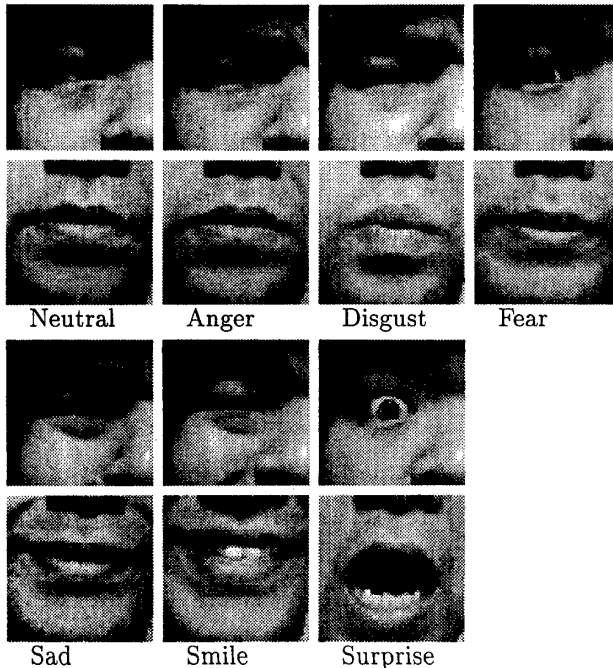


図2 無表情と6種類の基本表情

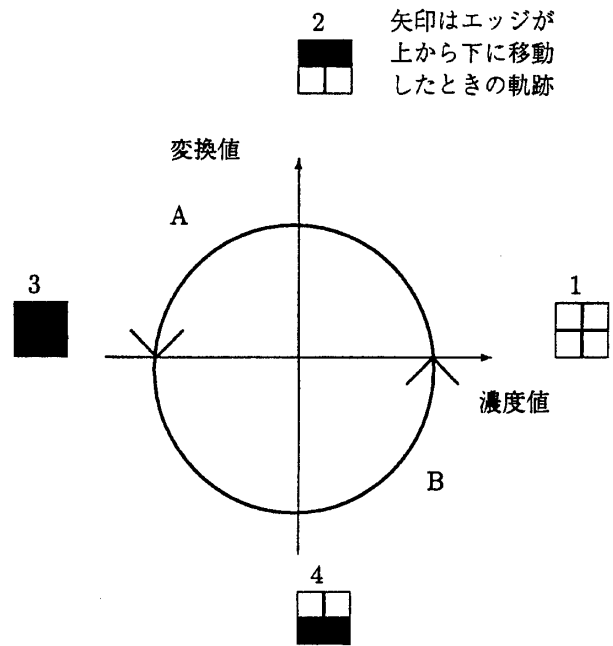


図1 エッジの移動に伴う濃度値(実軸), 変換値(虚軸)の変化

入力表情 結果	An.	Di.	Fe.	Sa.	Sm.	Su.
Anger	4	2	0	0	0	0
Disgust	1	3	0	0	0	0
Fear	0	0	4	0	0	0
Sad	0	0	0	3	0	0
Smile	0	0	0	2	5	0
Surprise	0	0	1	0	0	5

表1 表情認識実験結果(従来法)

入力表情 結果	An.	Di.	Fe.	Sa.	Sm.	Su.
Anger	5	1	0	0	0	0
Disgust	0	4	0	0	0	0
Fear	0	0	4	0	0	0
Sad	0	0	0	4	0	0
Smile	0	0	0	1	5	0
Surprise	0	0	1	0	0	5

表2 表情認識実験結果(従来法 + 本手法)

[参考文献]

- [1] H. Harashima and F. Kishino, "Intelligent image coding and communications with realistic sensations - recent trends-", IEICE Tr., Vol. E74, No. 6, pp. 1582-1592 (1991).
- [2] 坂口、大谷、岸野: 隠れマルコフモデルによる顔動画像からの表情認識, テレビジョン学会誌, Vol. 49, No. 8 (1995).