

HMMを用いた顔動画からの表情認識に関する検討¹

6 Q-3

大塚 尚宏 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。筆者らは、顔動画からの表情認識に隠れマルコフモデル (HMM) を用いる手法 [2] を拡張した手法を提案し [3], 6種類の基本表情に対して高い認識率を得た。そこでは、認識対象が特定の人物に限定されていたのに対して、本稿では複数人物に拡張するために HMM の出力確率を正規分布の荷重平均で近似する手法を提案する。この手法により、従来は特徴量として使用できなかった速度ベクトルの利用が可能となった。

2. HMMを用いた表情認識手法

HMM は非決定的オートマトンであり、状態遷移は遷移先の状態毎に異なった確率 (遷移確率) で起きる。また、状態遷移にともなって出力されるシンボルは確率的な分布を持ち (出力確率)、その分布は遷移先の状態毎に異なる。シンボル列が与えられたとき、HMM がそのシンボル列を生成する確率は遷移確率と出力確率から求められる。すなわち、初期状態のみに確率が付与された初期状態からシンボル列のシンボルを順に用いて各状態の確率を更新する。そして、シンボル列が終了した時点での最終状態の確率をシンボル列を生成する確率とする。HMM を用いた時間パターンの認識では、特徴抽出によって得られた時系列データを用いて識別すべきカテゴリー毎に HMM を予め学習しておき、時系列データを生成する確率が最大となる HMM のカテゴリーを認識結果とする。

連続的な時系列データをコードブックを用いて離散的なシンボルに変換し、それらのシンボルに出力確率値を付与する方法は、量子化誤差が発生するため認識率向上の妨げとなる。そこで、連続量をそのまま用い、出力確率を連続分布に変更することにより量子化誤差の問題点を解消した。さらに、複数人物の認識に対応するため、連続分布の出力確率を以下のように M 個の正規分布の荷重平均により近似した (M を分岐数と呼ぶ)。ここで、 c_k, μ_k, σ_k はそれぞれ k 番目の正規分布の荷

重、平均、標準偏差であり、荷重の総和は 1 である。

$$b(O) = \frac{1}{\sqrt{2\pi}} \sum_{k=1}^M \frac{c_k}{\sigma_k} \exp \left[-(O - \mu_k)^2 / \sigma_k^2 \right] \quad (1)$$

HMM のパラメータを推定する方法としては、学習データを用いてそれらの生成確率の総和が最大となるように推定を繰り返す Baum-Welch アルゴリズムが知られている。しかし、多数の学習データを用いる場合には、このアルゴリズムにより得られたパラメータを使った認識率が最大とはならない場合が生じる。さらに、学習データに使う被験者数が増えると発生頻度が増加することが実験により分かった。そこで、本研究では HMM のパラメータの推定方法として、クラスタ分析を併用する方法を採用した。すなわち、学習データの時系列をクラスタに分解し、それらのクラスタの要素数、平均値、標準偏差を対応する状態の出力確率のパラメータとした。一方、遷移確率は Baum-Welch アルゴリズムを用いて推定した。

3. 特徴抽出手法

上記のクラスタ分析を併用したパラメータ推定法を用いることにより、Baum-Welch アルゴリズムを使った場合には使えなかった特徴量が使えるようになった。すなわち、速度ベクトルは表情の開始および終了時に 0 であるため、Baum-Welch アルゴリズムを用いた学習結果が無表情のシーケンスとなる場合があった。しかし、上記推定法を用いて中間状態のパラメータを推定することにより、速度ベクトルを特徴量として利用することが可能となった。

特徴抽出は、動画の連続する 2 枚の画像から勾配法により求めたオプティカルフローを速度ベクトルとし、右目および口をそれぞれ中心とする 2 つの領域に対して速度ベクトルの各成分にフーリエ変換を施して得た。右目を含む領域からは 7 個、口を含む領域からは水平方向の動きを考慮して 8 個のフーリエ変換係数の低周波成分を選択した。さらに、表情変化の始めと終わりとを識別するために、2 つの領域の特徴量の自乗和の時間積分を特徴量として加えた。したがって、合計 17 個の特徴量を連続する 2 枚の画像から順に求め時系列データとした。

表情の変化にともなって生じる速度ベクトルは、無表

¹A Study of Facial Expression Recognition from Image Sequence Using HMM

Takahiro Otsuka, and Jun Ohya

ATR Media Integration & Communications Research Lab.

情からある表情に変化する始めと無表情に戻る終わりにピークを持つと考えられ、それらのピークの近傍の時系列データを切り出して認識処理を行うことは早い表情変化にも追従できるので有効な方法である。そこで、以下の実験ではフーリエ変換係数である15個の特徴量の自乗和のピークを検出し、その前後の定数個の時系列データを切り出して学習および認識に用いた。

4. 実験結果と考察

実験は4人の被験者(男性3人, 女性1人)から得られた6種類の基本表情(怒り, 嫌悪, 恐れ, 悲しみ, 幸福, 驚き)それぞれ10シーケンスを用いて行った。動画は、被験者が被ったヘルメットに装着された小型CCDカメラにより撮影されており、顔の部位の移動は表情変化のみに起因するとみなせる。図1に示すように右目および口を含む2つの領域(それぞれ256×256ピクセル)を切り出して画像処理を行った。2つの領域の位置は、右目および口がそれぞれの領域の中心に位置するように手動で行った。図1では、画像処理により得られたオプティカルフローを正方格子の交点の位置のずれとして表している。

HMMのパラメータの推定では、各被験者の10個のシーケンスから3個を選んだ合計12個のシーケンスを用いた。図2にクラスタ分析を用いて推定したHMMの出力確率分布を示す。縦軸(横軸)は右目(口)を含む領域の鉛直方向の平均速度である。出力確率分布は2つの正規分布の荷重平均により近似し、速度ベクトルの大きさが最大となる状態の出力確率分布を2つの楕円により表している。楕円の長径・短径は縦軸および横軸の特徴量の標準偏差に対応している。図2より出力確率分布が多峰性であることが分かる。

表情の判定では、HMMのパラメータの推定で用いたシーケンスを除いた合計28シーケンスを用いた。また、認識率の最適条件を求めるために、分岐数と速度ベクトルのピークを中心として切り出されるシーケンスの時間長を変化させて認識率を求めた。図3に認識結果を示す。分岐数の増加に対しては、認識率は単調に増加しているが分岐数2でも高い認識率が得られている。シーケンスの時間長に関しては、0.5秒より短いと急激に認識率が悪化している。したがって、高い認識率を得るためには0.5秒以上にすることが望ましい。しかし、0.5秒の間に無表情から表情が変化し再び無表情に戻るような表情変化が発生することは考えにくく十分に実用的な値であるといえる。

5. むすび

複数人物の表情を認識する手法として、HMMの出力確率を正規分布の荷重平均で近似し、そのパラメータをクラスタ分析とBaum-Welchアルゴリズムを併用して推定する方法を提案し良好な認識結果を得た。さら



図1: 速度ベクトルの一例

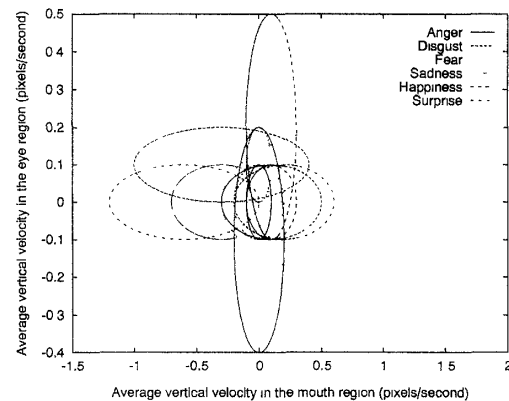


図2: 出力確率分布の一例

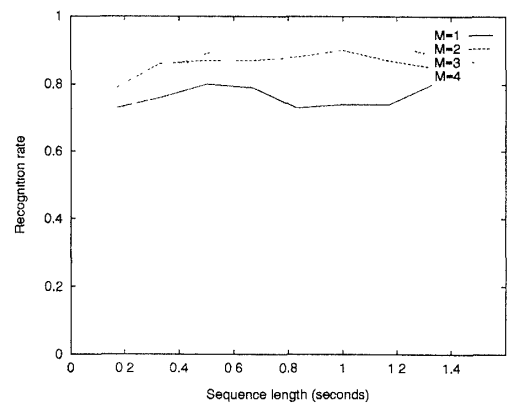


図3: 認識率と分岐数 M, シーケンス時間長との関係

に、速度ベクトルから得た特徴量を認識で使うことが可能となり、早い表情変化にも追従できる見通しを得た。

[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からの表情認識, テレビジョン学会誌, Vol. 49, No. 8 (1995).
- [3] T. Otsuka, and J. Ohya, "Recognition of facial expressions using HMM with continuous output probabilities", Proc. IEEE International workshop on Robot and Human Communication, pp. 323-328 (1996).