

## HMMを用いた動画像からの動作認識の高精度化と評価

5 J-9

大和淳司、倉掛正治、伴野明、石井健一郎

NTT ヒューマンインタフェース研究所

## 1 はじめに

筆者らは、シーン中の人物の動作を認識するため、HMM (隠れマルコフモデル) を適用した、特徴量ベースの学習によって認識系を構成する手法を提案した [1]。これまで、テニス動作などを対象にして、学習と認識を同一人物で行なった場合において、90%以上の認識率を達成している [2]。

しかし学習に用いた人物と異なる人物の認識においては、認識率が70~80%程度とやや低くなる傾向が見られ、さらなる高精度化を図る必要があった。

そこで今回、誤認識要因の一つとして考えられるベクトル量子化部分の改善を行い、その効果を評価した。

## 2 HMMによる動画像認識の概要と問題点

HMMによる動画像認識のフローをFig.1に示す。詳細については既報を参照されたい。HMMの学習は、HMMのパラメタを推定することで行う。認識の際は、ある認識対象の時系列画像に対して、各カテゴリのHMMが認識対象パターンを出力する確率(尤度)を計算し、尤度最大のHMMに対応するカテゴリを認識結果とする。

以上の手法の問題点を明らかにするために、テニス動作の画像に対する実験で誤認識を起こした事例に関して検討した結果、VQによるシンボル変換で既に問題があることが明らかになった [3]。すなわち、誤認識は、VQにより得られた動作のシンボル列の中に、学習パターンで得られたシンボル以外のものが頻繁に出現する場合に多く見られ、時系列的なパターンの違い以前の段階で、シンボルそのものの違いが影響しているものが多かった。

従って、VQ段階の不安定性が、全体の認識率の低下の一つの原因であるといえる。そこで認識率の向上を目指して、学習ベクトル量子化(Learning Vector Quantization:LVQ) [4]を導入し、VQの高精度化を図った。

## 3 LVQの導入

VQコードブックは、各クラスの代表点の特徴ベクトルと、各クラスの名前であるラベルの集合からなり、各

クラスはHMMのシンボルに対応する。ラベルの種類数がHMMのシンボル数に等しい。既報の実験においてはVQコードブックは学習パターンの特徴ベクトル列を時間等間隔にサンプリングして作成したが、この部分をより高精度化するためLVQを導入した。LVQはVQコードブックの学習法の1つで、以下のような教師付き学習を行う。クラス既知の、学習用特徴ベクトル $x$ に関して、コードブック $\{m_i\}$ でもっとも近いクラス $c$ すなわち、

$$c = \arg \min_i \{length(x - m_i)\}$$

が、正しいクラスならば次式、

$$m_c(t+1) = m_c(t) + \alpha(t)[x(t) - m_c(t)]$$

によって、そのクラスの代表点特徴ベクトルを $x$ に近づけ、誤ったクラスならば次式、

$$m_c(t+1) = m_c(t) - \alpha(t)[x(t) - m_c(t)]$$

によって遠ざける。他の代表点特徴ベクトルに対してはなにもしない。

$$m_i(t+1) = m_i(t)。$$

また、 $\alpha$ の値は時間 $t$ とともに単調に減少させてこの手続きを収束させる。

## 4 実験

実験は6種類のテニス動作の画像を使用した。6つの動作カテゴリは、バックボレー、バックストローク、フォアボレー、フォアストローク、スマッシュ、サービスである。3人の被験者の各動作カテゴリの10試行の画像データを対象に認識実験をおこなった。画像サイズは200画素×200画素、白黒で、背景画像との差分により人物領域を抽出し、位置と大きさの正規化をした後、人物領域のメッシュ特徴を抽出し、特徴ベクトル列を得る。メッシュサイズは8画素×8画素であるので、各特徴ベクトルは625次元となる。これをベクトル量子化し、シンボル列を得る。

LVQにおける学習の効果は、学習後のシンボル変換の際に、学習に使用した特徴ベクトルのラベルと同じクラスにシンボル変換されたものの割合(以下シンボル変換

Human Action Recognition using HMM:improvement and evaluation

Junji YAMATO,Shoji KURAKAKE, Akira TOMONO and Kenichiro ISHII

NTT Human Interface Laboratories

の正答率と呼ぶ)で評価した。このシンボル変換の正答率が学習回数によってどのように向上するかを調べた。実験条件は学習率  $\alpha = 0.05$ , シンボル数 72 である。

また、学習後のコードブックを使用した結果、最終的な動作認識の認識率がどのように向上するかを調べた。実験条件は、HMM の状態数 12、学習の際の Baum-Welch Algorithm の繰り返しは 100 回とした。認識率の評価では、学習データとテストデータが同一人のものであるため、両者の分離のために、LVQ によるコードブックの学習、HMM の学習ともに、10 試行中 9 試行を学習に用いて残り 1 試行を認識に使用した (Jack-knife 法)。この 9 試行の選び方を変えて 10 通りの実験を行った。

5 結果と考察

Fig.2 に、被験者 A に関して、LVQ によるシンボルアサインの正答率を示す。この図から LVQ によってシンボル変換正答率がかなり向上したことが確認できる。学習が 10000 回を越えると頭打ちになるが、クラス数による性能限界と考えられる。次に LVQ でトレーニング ( $\alpha = 0.05$  学習回数 10000 回) したコードブックを使用した場合の HMM による認識実験の結果を Table.1 に示す。

いずれの被験者に関しても認識率の向上が確認できた。認識率の向上が、LVQ の正答率向上に比べて小幅であるのは VQ 段階での正答率があるまま認識率に跳ね返るのではないことによる。つまり、学習データとテストデータで、類似した特徴ベクトルが同じシンボルに安定して変換されればよいのであって、LVQ の教師データと同じクラスに変換されなくても正しく認識できていたと考えられる。従って、LVQ 導入によって認識率が向上する効果は、シンボル変換正答率が向上したことにより、シンボル変換の安定度も向上したことに起因するものであり、シンボル変換正答率の向上が全て認識率に反映するわけではないことになる。

6 おわりに

LVQ を導入したことによる、シンボル変換の正答率、学習と認識が同一人の場合における動作認識率の向上を確認した。この手法により、従来法で認識率が比較的低かった、学習と認識が別人物の場合についても改善が期待できる。

Table.1 LVQ+HMM による動作認識率 (%)

被験者	VQ	
	前回手法	LVQ1
A	98.3	99.7
B	93.3	95.0
C	90.0	95.0

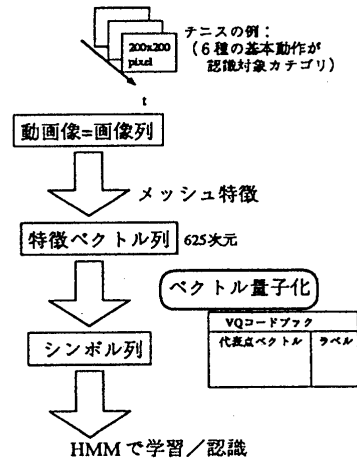


Fig.1 HMM による動作認識フロー

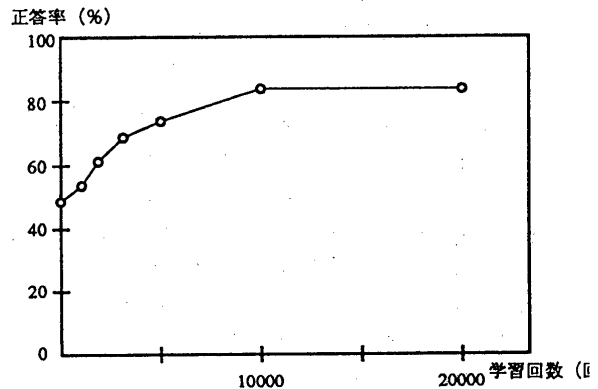


Fig.2 LVQ によるシンボル変換正答率の向上

謝辞

御指導御協力をいただいた、ATR 通信システム研究所大谷淳主幹研究員、および、ヒューマンインタフェース研究所第 6 プロジェクトの皆様へ感謝致します。

参考文献

- [1] 大和淳司, 大谷淳, 石井健一郎. “HMM を用いた行動認識法の検討”, 第 43 回情報処理学会全国大会, 10,1991.
- [2] J.Yamato, J.Ohya and K.Ishii. “Recognizing Human Action in Time-Sequential Images using Hidden Markov Models”. In *Proc. Computer vision and Pattern recognition (CVPR92)*, pp. 379-387, 1992.
- [3] 大和淳司, 大谷淳, 石井健一郎. “HMM を用いた動画像からの人物の行動認識”, 信学技報 PRU, 9,1991.
- [4] T.Kohonen. “The self-organizing map”. *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1464-1480, 1990.