

SVM を用いた動作認識の為の動作状態の自動推定

松永武文[†] 尾下真樹[‡]

[†]九州工業大学大学院 情報工学研究科 情報科学専攻

[‡]九州工業大学 情報工学研究院 システム創成情報工学研究系
〒820-8502 福岡県飯塚市川津 680-4

E-mail: matsunaga@cg.ces.kyutech.ac.jp, oshita@ces.kyutech.ac.jp

カメラやモーションキャプチャなどの機器から取得した手足の位置や速度等の特徴量から、対象者がどんな動作を行っているのかを識別する動作認識の技術には、ゲームのインターフェースや異常行動の監視システムといった様々な応用が考えられる。我々は、これまでに、SVM と状態マシンを組み合わせた動作認識手法を開発した。しかし、実用に足る認識精度を得るには、動作毎に適切な状態遷移図の設計を行う必要があり、手作業では手間が掛かってしまうという問題があった。本論文では、状態遷移図作成の手間を減らす為の、SOM による動作の自動分割手法を提案する。

Automatic estimation of motion state for motion recognition using SVM

Takefumi MATSUNAGA and Masaki OSHITA

Department of Systems Innovation and Informatics, Kyushu Institute of Technology
680-4 Kawazu, Iizuka-Shi, Fukuoka, 820-8502, Japan

E-mail: matsunaga@cg.ces.kyutech.ac.jp, oshita@ces.kyutech.ac.jp

There are many applications of motion recognition technique such as gaming interface, monitoring systems, and controlling an intelligent robot. Motion recognition detects subject's motion based on the feature vector values that are acquired from a camera or a motion capture. We developed a motion recognition technique which combines Support Vector Machine (SVM) and a state machine. However, it was necessary to make a state transition chart manually for each motion. Making a state transition chart that achieves good recognition results takes a great deal of time. In this paper, we propose an automatic division technique of motion by using Self-Organizing Maps (SOM) to make a state transition chart automatically.

1. はじめに

動作認識は、カメラやモーションキャプチャなどの機器の入力データから人体の関節角度や手足の位置、速度といった何らかの特徴量を得て、特徴量をもとに対象人物がどのような動作を行ったかを識別する技術である。この動作認識の技術は、任天堂の Wii のようなユーザの体の動きによる直感的な操作を可能にするゲームのインターフェース、家電機器やロボットのジェスチャ制御[1][2]、あるいはカメラの動画画像から異常行動を感知する監視システム[3]など、様々な分野での応用が考えられる。一般

に、動作認識のような時系列パターンの認識を行う場合、対象を複数の状態に分割し、状態毎に個々の独立した認識器を用意して、それらの情報を統合する事で動的現象の認識を行う。そのため、こういった従来の動作認識手法を用いて、精度の高い認識を得る為には、あらかじめ手作業で、認識を行う動作を幾つかの状態に分割した上で、各状態間の遷移条件を学習する必要があった。しかし、認識する動作の種類が増えれば、分割された動作の状態と状態間の遷移条件を含んだ状態遷移図の作成は大きな手間となる。そこで、本研究では、実用に足る高精

度の動作認識の実現と、複数の様々な動作について手間を掛けず状態遷移の推定を行う事の2つを本研究の目的に定める。

我々は、これまでに、高精度の動作認識を実現する為に、Support Vector Machine[4] (SVM)と状態マシンを組み合わせた認識手法を提案した[5]。SVMによる認識は高い汎用性を持つが、ある瞬間の特徴量だけにもとづいて認識を行うだけでは、動作の前後関係等の時間的要素を考慮出来ないという問題がある。そこで、本手法[5]では、SVMに状態マシンを組み合わせて用いる。

例えば任天堂のWiiリモコンから取得した手の加速度に基づいて、手を上げた後に下げるといった一つの動作を認識する場合、まず、特徴量として用いる手の加速度の時間変化に応じて「手を下げた状態」「手を上げる途中の状態1」「手を上げる途中の状態2」…「手を上げる途中の状態n」「手を上げた状態」「手を下げる途中の状態1」「手を下げる途中の状態2」…「手を下げる途中の状態n」といった具合に、複数の状態に分割する(図1)。次に、分割された各状態に、遷移条件の判定を行う為のSVMの学習モデルを持たせる。SVMは、例えば、図1において「手を下げた状態」から手の加速度が上方に大きくなれば、「手を上げる途中の状態1」へ遷移するように、サンプル動作にもとづいて学習を行う。また、図1の動作の場合、手を上げる途中で一旦止めて、再び上げ始めるような場合を考慮して、「手を下げた状態」に遷移するまでは、いつでも「手を上げる途中1」に遷移出来るように、遷移を学習する。

本手法を用いて動作認識を行う際には、リアルタイムに対象者から取得する特徴量に応じて、現在の状態のSVMを用いて状態遷移の判定を行う。この時、遷移の経路が学習したサンプル動作と同じ経路を辿っていたら、対象者がその動作を行ったと判定できる。

しかし、本手法を用いる場合、事前に、認識したい動作を複数の状態に分割し、各状態間の遷移関係も設定する必要がある。このような推定作業を手作業で行うと、大きな手間が掛かる。また、状態図の設計によって認識精度は異なるため、複数の異なる状態図を作成して比較するなどの試行錯誤が必要となってしまう。

そこで本論文では、推定作業の手間を省き、認識の精度をより向上させる為に、Self-Organizing Maps (SOM)[6]による動作状態の自動推定手法を提案する。サンプル動作に

ついて、動作の全時刻の特徴量を、SOMを用いて分類し、分析結果の各グループを一つの状態とする事で、動作の分割を行う。そして、サンプル動作の時系列による状態の変化から、各状態間に遷移関係を設定する。遷移関係に基づき、従来の手法[5]を用いて、状態間の遷移条件としてSVMのモデルを持たせ、状態遷移図を作成し、動作認識を行う。すなわち、提案手法により、従来は手作業で行う必要があった状態遷移図の作成という処理を、自動的に行う事が出来る。これにより、使用者にとって動作の学習に必要であった手間が大幅に省かれる。

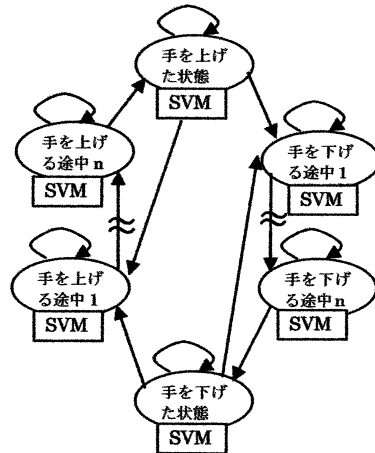


図1：手の上げ下げ動作の状態遷移図

2. 関連研究

事前学習によって動作モデルを生成して動作認識を行う手法は、本研究以外にも幾つかの研究が行われている。

豊倉ら[7]による研究では、HMMを用いた手話の認識を行っている。この手法では手話動作を構成する基本動作の枠組みを定義し、手話の単語を基本動作単位に分類し、基本動作毎にHMMに基づいたモデル化を行っている。基本動作は2次元平面上の動作情報を用いて定義している。動作分類には決定木に基づいた状態クラスタリング法が用いられているが、分類条件は使用者側が与える必要がある。この為、ある程度手話動作について解析した知識を持ち合わせてなければ、適切に基本動作の分類を行う事は難しい。一方、本論文で提案しているSOMによる自動推定手法では、使用者が分類条件を与える必要が無い為、動作についての知識を持たなくとも、自動的に動作の分類を行

えると思われる。但し、手動で分類を行った場合と比べて、認識精度が低下する可能性もある。

3. 動作認識手法

本論文で提案するシステムの処理の概要を説明する。処理の流れは、大きく学習処理と認識処理の2つに分けられる(図2)。

学習処理では、学習元となる複数のサンプル動作の入力から、状態遷移図を出力する。状態遷移図は、動作の各状態、状態間の遷移関係、各状態から遷移関係にある状態への遷移条件を含む。学習処理は、SOMによる動作分割と遷移関係の設定、状態遷移図の生成という2段階の処理に分けられる。

SOMによる動作分割と遷移関係の設定は、認識を行うサンプル動作を10~100回程度行った時の、動作中の全時刻の特徴量を入力とする。例えば、本論文の実験では、手を上げた後に下げるといった一つの動作を10回程度行い、Wiiコントローラから取得した両手の3次元の並進加速度を特徴量として、一回の動作につき約200フレーム程度取得した。入力された全時刻の特徴量をもとに、類似した特徴量同士をまとめて一つの動作状態とする事で、サンプル動作を複数の状態に分割する。また、入力動作における動作状態の時系列による変化から、状態間の遷移関係を設定する。従来の手法[5]では、動作分割と遷移関係の設定を手作業で行っていたため、手間が掛かっていた。本研究では、SOMを用いる、本作業を自動化した。SOMを用いた動作状態の推定については、4節で詳しく説明する。

状態遷移図の生成では、まず、入力された全時刻の特徴量の中から、各状態において、その状態から遷移する可能性にある全ての状態を取得する。その全状態に含まれる全ての特徴量からSVMによる遷移条件の学習を行う[5]。

認識処理では、利用者の動作をリアルタイムに受け取り、学習動作を行ったかどうかを判定する。まず、各時刻の利用者の状態から、特徴量を計算する。そして、その特徴量と、現在の状態の持つSVMの認識モデルをもとに、遷移判定を行う。遷移判定の結果、他の状態に遷移する場合は、現在の状態を更新する。この時、状態遷移図の動作開始状態から動作終了状態まで、あらかじめ定義された一連の状態遷移が行われた場合に、その動作が実行されたと認識される。あらかじめ、判定を行う動作毎に認識モデルを作成しておき、この認識処理を複数の

モデルに対して行うことで、利用者が行っている動作の種類を認識できる。以上の処理を、利用者の動作が入力される限り、各時刻毎に繰り返し行う。

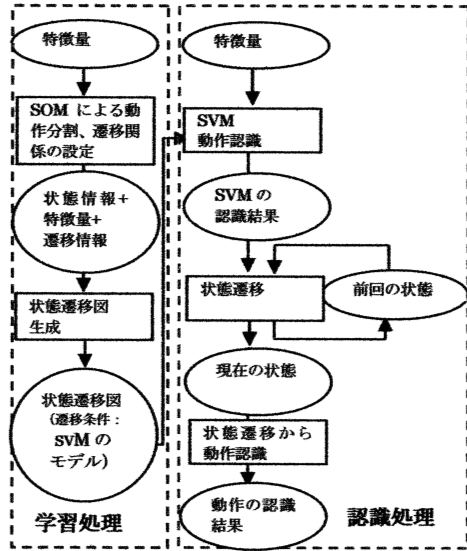


図2：システム概要

4. SOMを用いた動作状態の推定

4.1 SOMの概要

自己組織化マップ(SOM)[6]とは、教師なしニューラルネットワークの一種であり、画像認識や音声認識のみならず、遺伝子解析や医学、土木といった幅広い分野で、データの分類や視覚化などに利用されている。基本的に、高次元である複数の入力データ群を1次元ないし2次元のマップ上のユニットへ写像する機能を持つ。この時、数値の近い入力データ同士がマップ上の同じユニットや近い位置のユニットに配置される為、類似度に応じてデータを分類する事が出来る。入力データの分類は参照ベクトルをもとに行われる。参照ベクトルはユニットが個々に持つ、入力データと同じ次元数のベクトルである。入力データは N 次元の特徴量ベクトル $\mathbf{x} = (x_1, x_2, \dots, x_N)$ が、データの個数分入力される。また、各ユニット $u_{i,j}$ が持つ N 次元の参照ベクトルは、 $\mathbf{w}_{i,j} = (w_{j,1}, w_{j,2}, \dots, w_{j,N})$ で表される。

4.2 SOMの処理の流れ

複数の入力データが、それぞれどのユニットに配置されるかを出力する、一般的なSOMの

処理ステップは、以下の通りである。

1. 各ユニットが持つ参照ベクトルの初期化を行う。特徴量ベクトルの各次元 n 毎に、入力された全特徴量ベクトルの最大値と最小値から、各ユニット $u_{i,j}$ の参照ベクトル $w_{i,j}$ の n 次元目の要素をランダムに初期化する。

2. 各時刻 f の特徴量ベクトル \mathbf{x} と全てのユニットの参照ベクトル $w_{i,j}$ との距離を求め、最も距離の小さいユニットを探索し、探索結果に応じて全ユニットの参照ベクトルを更新するという処理を、あらかじめ設定された回数(T 回)繰り返す。なお、適切な T の値は、マップ上のユニット数に依存し、一般にユニット数の500倍程度の値が必要だと言われている。

2-1 各時刻の入力ベクトル \mathbf{x} と参照ベクトル $w_{i,j}$ のユークリッド距離 L を式(1)で計算する。

$$L = \sum_{n=1}^N \|x_n - w_{i,j,n}\| \quad (1)$$

入力ベクトルと全ユニットの参照ベクトルの距離から、最短距離のユニットのマップ上の位置 (I, J) を決定する(図3)。

2-2 処理2で探索した最近ユニットの位置 (I, J) を元に、全ユニットの参照ベクトルの値を更新する。この処理により、最近ユニットに近い位置のユニットほど入力ベクトルに近い参照ベクトルを持つようになる。各ユニットの参照ベクトルの要素 $w_{i,j,n}$ を式(2)により更新する。 $w'_{i,j,n}$ は更新前の参照ベクトルを表す。

$$w_{i,j,n} = w'_{i,j,n} + h((i, j), (I, J))(x - w'_{i,j,n}) \quad (2)$$

h は近傍関数と呼ばれ、以下の2つの性質を持つ関数である。一つ目の性質として、処理の繰り返し回数 t に関する単調減少関数であり、 t が無限大になると h は0に収束する。また、もう一つの性質として、マップ上の格子点 (i, j) と最短距離ユニットの格子点 (I, J) の距離 $\|u_{i,j} - u_{I,J}\|$ に関して単調減少する。この距離による単調減少の程度は t が増加するほど大きくなる。式(2)により、近傍関数の値が大きい程、参照ベクトルが入力ベクトルに近い値になるように更新される。

3. 処理ステップ2を、全時刻分の入力データに対して順に行う。これにより全ての入力データ \mathbf{x} がどのユニットに属するかという情報 U_f を持つ事になる。

4.3 SOMによる動作分割

本研究では、動作を複数の状態に分割する目的でSOMを使用する。サンプル動作の全時刻の特徴量ベクトル \mathbf{x}_n を入力として、SOMによる分類結果を出力する。この時、マップ上の各ユニット $u_{i,j}$ を、動作を分割した状態の一つとみなす。また、各特徴量ベクトルの最短距離ユニットの情報 U_f から、各特徴量ベクトルがどの状態に属するかを決定する。ただし、SOMによる分類処理の結果、空のユニットが出来た場合、空のユニットは状態としない。

SOMによる分類処理を行う際、ユニットの数 K を指定する必要がある。この時、ユニット数が多すぎると、無駄な遷移やノイズとなる状態が増えるため、認識精度が悪くなる恐れがある。また、ユニット数が少なすぎると、認識対象とは異なる動作を誤認識しやすくなる恐れがある。従って、認識を行う動作に応じて、適切なユニット数を決定する必要がある。本手法では、適切なユニット数を推定するための方法として、単純に幾つかのユニット数で状態遷移図の作成と学習を行い、比較を行って、最も認識精度の高いユニット数を使用する。ユニット数に応じた認識精度の比較については、5.2節で実験結果を示し、考察する。

4.4 状態遷移図の生成

SOMによる分類処理の結果、動作を分割した各状態が作成されたら、次は時系列でみた動作の状態の変遷をもとに、状態間の遷移関係の設定を行う。サンプル動作において、例えば、ある時刻の特徴量ベクトルが状態1に分類され、その次の時刻のベクトルが状態2であった場合、状態1と状態2の間に遷移関係を設定する(図4)。この処理を全時刻の特徴量ベクトルに対して行い、遷移条件を与えて状態遷移図を生成する。

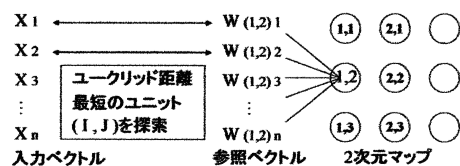


図3：最近ユニットの探索

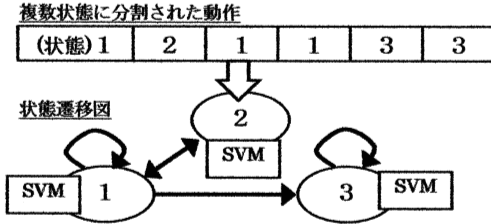


図 4：状態遷移図の作成

5. 実験

5.1 実験概要

本手法の有効性を評価するため、ジェスチャ動作の認識システムを作成し、実験を行った。5.2 節に、認識精度とマップのユニット数の比較実験の結果を示す。5.3 節に、手作業による動作分割と SOM による動作の自動分割との、学習時に費やした時間と認識精度の比較実験を示す。

入力デバイスとして、任天堂の Wii リモコンとヌンチャクを用いた。また、特徴量として、右手を基準とした右手の併進加速度と、左手を基準とした左手の併進加速度について、x 軸方向、y 軸方向、z 軸方向の計 6 次元の値を取得した。特徴量の取得周期は秒間 16 回である。SVM には、Lin らが開発したライブラリである LibSVM[8]を用いた。SOM は、[6]を参考に実装した。

5.2 ユニット数に対する認識精度の評価

SOM のユニット数による認識精度の評価を行う為に、右手を上げた後に下げるという一連の動作について認識実験を行った。手を下げた状態を初期状態として、手を上げてから下げる動作までの動きをサンプル動作とし、10 回動作を行って、その各時刻の特徴量を取得した。SOM のユニット数を 2 個、4 個、6 個と変化させ、それに対応した特徴量の分類とサンプル動作の分割を行った。図 5 にユニット数 4 の場合のサンプル動作の分割の様子を示す。上方の折れ線グラフは取得した 6 次元の特徴量を表している。極大値と極小値の差が最も大きい折れ線が、右手の z 軸方向の加速度である。下方の棒グラフは各時刻の特徴量がどのユニットに分類されたかを表している。点線の四角で囲まれた範囲は、手を一度上げた後に下げる、一連の動作の一周期分を表す。動作を分割した結果を基に、それぞれの状態遷移図を作成し、動作を行った時の状態遷移の経路の決定を行った。例えばユニット数が 2 個の場合、状態 1 → 状態 2 → 状態 1 → 状態 2 → 状態 1 と遷移した

場合に、一回の動作と数える。それぞれのユニット数での動作の経路を図 6 に定義した。

それぞれのユニット数の場合において、各状態に遷移条件として SVM のモデルを与えて認識システムを構築した。認識対象として 50 回動作を繰り返して行い、それぞれのユニット数で定義した遷移経路と一致した状態遷移の数を測定した。また、認識したい動作とは別の、右に一度振って左に戻すという動作についても 50 回繰り返して行い、遷移経路と一致した数を測定した。実験の結果を図 7 の表に示した。図 7 に示した認識結果から、ユニット数が少ないほど認識の精度が高く、多いほど低い事が分かる。しかしながら、状態数が少な過ぎると、目標と違う動作を誤認識する割合も高くなってしまふ。よって、正しい認識と誤認識の釣り合いの取れたユニット数を適切に選択する必要がある。実験結果から、右手を上げた後に下げる動作の認識には、ユニット数 4 個が最も適切である事が示された。

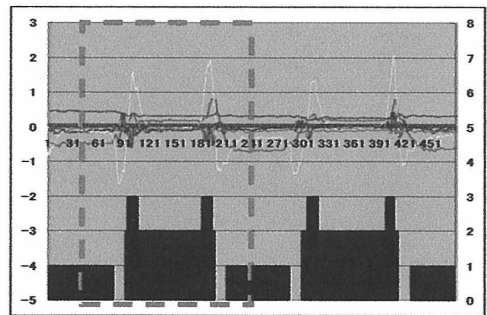


図 5：特徴量グラフ

ユニット数	経路
2	1→2→1→2→1
4	3→4→3→1→2→1→3→4→3
6	4→2→1→2→4→6→5→6→4→2→1→2→4

図 6：ユニット数と動作の状態遷移の経路

ユニット数	正しい動作の認識率(%)	間違った動作の誤認識率 (%)
2	98	100
4	84	0
6	60	0

図 7：ユニット数と認識精度、誤認識率

5.3 手作業と自動生成での状態遷移図による認識の比較

手作業で作成した状態遷移図と SOM を用いて自動生成した状態遷移図との、作成に掛かった時間と認識精度の比較を行った。認識対象

の動作として、前述の5.2節で行った手の上げ下げ動作を動作A、両手を体の中央に揃えた状態を初期状態として中央→右方→左方と動かして最後に両手を左右に広げる(右手を右方に左手を左方に)一連の動作を動作Bと定義する。5.2節と同様に、動作を10回ずつ行いサンプル動作として取得し、SOMによる動作の分割と手作業での分割をそれぞれ行った。ユニット数は、5.2節と同様の実験を行って最も精度の高かった個数を使用した。動作Aについてはユニット数2個、動作Bについてはユニット数4個とした。動作を定義する経路を決定した後、50回の動作を試行し、それぞれのユニット数で定義した経路と一致した状態遷移の数を測定した。動作の分割と遷移の設定に要した時間と認識結果を図8に示す。

結果から分かる通り、手作業に比べて、十分短い時間で提案手法による動作の分割を行う事が出来た。また、認識精度については、手作業と比較すると、やや悪くなってしまふ事が分かった。原因は、学習で定義した経路と異なる状態遷移が、提案手法による認識でより多く発生した為である。この問題は、学習する動作の試行回数を10回より大幅に増やして、定義する経路の数を増やす事で解決出来ると思われる。また、手作業による動作Bの設定に掛かった時間が動作Aの10倍である事から、ユニット数が多くなれば手作業での設定が現実的ではなくなると考えられる。

方法・動作の種類	時間(sec)	認識精度(%)
手作業 動作A	510	100
提案手法 動作A	4	98
手作業 動作B	5100	84
提案手法 動作B	8	80

図8：手作業と提案手法の比較

6. おわりに

本論文では、SVMと状態マシンを組み合わせた動作認識手法と、SOMによる動作の自動分割手法を提案した。SOMによって自動で特徴量を分類し、手作業に比べて十分短い時間で状態遷移図を作成する事が出来た。また、生成された状態遷移図を元に作成した認識モデルを使って動作認識の実験を行い、手作業に比べて認識精度がやや落ちる事を確認した。

更に認識精度を向上させる為のアイデアとして、特徴量の次元削減が考えられる。対象者の体全身の関節角度や全て部位の位置・速度など多くの次元の特徴量ベクトルを取得した場

合、それに応じて多様な動作の認識に対応する事が出来るが、個々の動作の認識精度は低くなってしまふ可能性が高い。また、例えば右手を主に動かす動作に着目すると、特徴量の内、左手や他の部位の要素はあまり意味を持たない為、無駄になるといった場合がある。そこで、それらの特徴量を自動で削減、あるいは一まとめにして精度を向上させる手法を考えている。具体的には、主成分分析を適用する事を検討している。

参考文献

- [1] Kota Irie, Naohiro Wakamura, and Kazunori Umeda: "Construction of an Intelligent Room Based on Gesture Recognition -Operation of Electric Appliances with Hand Gestures-", Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.193-198 (2004)
- [2] Toshiya Yamada, and Kazunori Umeda: "Improvement of the Method of Operating a Mobile Robot by Gesture Recognition", JSME Conference on Robots and Mechatronics, Vol.2001, p.49 (2001)
- [3] Takuya Nanri, and Nobuyuki Otsu: "Anomaly Detection in Motion Images Containing Multiple Persons", PRMU2004-77, Vol.104, No.291, pp.583-588 (2004)
- [4] Nello Cristianini and Jhon Shawe-Taylor, "An Introduction to Support Vector Machines", Cambridge University (2000)
- [5] Takefumi Matsunaga and Oshita Masaki, "Recognition of Walking Motion Using Support Vector Machine", ISICE2007, pp.337-342 (2007)
- [6] T.Kohonen, "The self-organizing map", hoc. IEEE, vol78, pp. 1468-1480 (1990)
- [7] Yukitaka Toyokura and Yoshihiko Nankaku et al. "Approach to Japanese Sign Language Word Recognition using Basic Motion HMM", Proceedings of the Society Conference of IEICE, Vol.2006, pp.72 (2006)
- [8] Chih-Chung Chang and Chih-Jen Lin, LIBSVM:a library for support vector machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (2001)