

階層的隠れマルコフモデルによる居住者の行動推定

○杉野原 和也(九州大学) 諸岡 健一(九州大学)
辻 徳生(九州大学) 倉爪 亮(九州大学)

Estimation of Resident Behavior Using Hierarchical Hidden Markov Model

○ Kazuya SUGINOHARA, Ken'ichi MOROOKA, Tokuo TSUJI, and Ryo KURAZUME

Abstract: This paper presents a new method for estimating human actions in a room by hierarchical hidden Markov model (HMM) using a laser range finder (LRF). The proposed method uses two layers of HMMs. The first layer is the set of the HMMs, each of which estimates its corresponding action. In the second layer, one HMM is generated by the integrating the HMMs in the first layer.

1. はじめに

高齢社会となった日本をはじめ、多くの先進国では高齢者の介護や生活支援に関する諸問題が生じている。この問題の解決策として、高齢者を見守りながら、日常生活を支援するロボットやシステムが期待されている。このシステムは、居住者の日常行動を観測したデータを蓄積、分析し、生活習慣の変化を検出することで、適切な介護サービスの提供や、転倒などの異常行動の早期検出ができる。我々は、環境側に固定センサを配置して、環境の地図と併せて、ロボットや居住者の移動体および物品の情報を取得・管理する環境情報構造化¹⁾という概念を提案している。そして、情報構造化環境下での高齢者支援システムを開発している。本研究では、情報構造化環境下で居住者の行動を認識するシステム構築を目的とし、その要素技術である居住者の動作を推定するシステムを提案する。ここで、本研究では、歩行、静止、床座、横臥、を対象動作とする。

これまで、ステレオカメラや RGB カメラを使用した動作・行動認識の研究が多く行われている²⁾。しかし、画像には、顔などのプライバシー情報も記録される。情報漏洩の可能性を完全に無くすことは現状困難であり、したがって、プライバシー保護の観点から画像情報を記録することは、居住者にとって好ましくない。これに対し、レーザレンジファインダ (Laser Range Finder: LRF) を使って人物追跡し、居住者の行動予測を行う手法^{3,4)}がある。文献³⁾は、人物位置の遷移から、数十秒後内に行われる行動を予測する。しかし、この行動予測は、人物位置の遷移に基づくため、静止時や転倒時など、位置の遷移がない間は推定できない。また文献⁴⁾は、行動推定で用いる閾値や、行動のパターンやパターン間のつながりなど、多くの要素を手動で決定している。そのため、限定された環境でしか適用できず、汎化性が

低い。以上を踏まえ、プライバシー保護を考慮しつつ、人物位置の遷移がなくとも動作・行動を推定できる汎化性の高いシステムが必要である。

そこで、本研究では、情報構造化環境内で LRF によって得られたデータから、隠れマルコフモデル (Hidden Markov Model: HMM) を階層的に組み合わせることで、居住者の動作を推定する手法を提案する。提案手法では、各動作を認識する HMM と、これらの識別結果を統合し最終的な動作認識をする HMM、の 2 種類からなる。後者の HMM は、動作間のあいまいさを考慮した識別を行うために、観測データに対する複数の動作モデルの尤度の組を特徴量として学習する。

2. センサシステム

2.1 フロアセンシングシステム⁴⁾

フロアセンシングシステム (Fig.1 左) は、床上に設置した 1 基の LRF からなる。LRF は、レーザの走査面が床上約 2.7cm で水平になるよう設置されており、毎秒 40 フレームで床上を走査する。物体が床上にあるときはその表面までの距離が、そうでない時は壁や家具までの距離が、それぞれ得られる。この差分により物体を検出し、その位置を計測できる。その結果、Fig.1 右に示すように床上に現れる人の足や物体の輪郭を点群 (クラスタ) として得る。我々の研究グループでは、これまで、フロアセンシングシステムで得られたクラスタから、床上の物体および人の足を識別する手法を提案した⁴⁾。本研究では、室内に居住者が 1 人であると仮定し、文献⁴⁾の手法により居住者の足のデータを得る。

2.2 クラスタの統合による人物位置算出

フロアセンシングシステムでは、歩行時の足のクラスタは交互に計測され、かつ左右の足の計測時間に重なりがある。

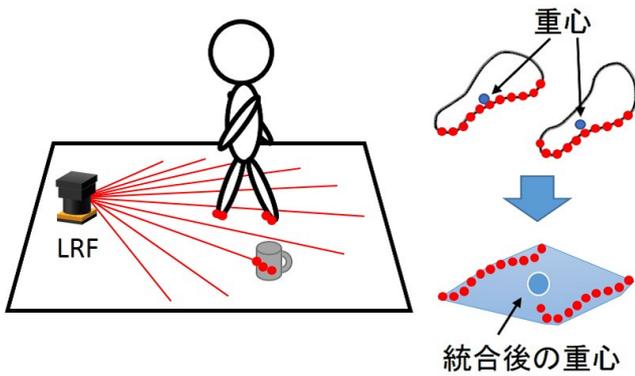


Fig.1 Floor Sensing System

また、床座や横臥時には、体の凹凸により複数のクラスタが観測されることがある。このように、計測時のクラスタ数は動的に変化するため、それに対応した人物位置の特徴量を定める必要がある。本研究では、Fig.1右に示すように、各フレームにおいて観測される全クラスタの重心座標の平均値を計算し、そのフレームでの人の座標とすることで、クラスタを1つに統合する。

3. 提案手法

各動作を認識するHMM（以後、動作モデル）を構築し、それらを統合する。このような複数のHMMを統合する方法として、全動作モデルから観測データの尤度を求め、尤度の高い動作モデルを選択するものがある。しかし、選択されなかった動作モデルの尤度を無視する結果となり、複数のHMMで尤度が類似するようなあいまいな動作を推定できない。動作の遷移過程まで詳細に推定することができれば、より多くの動作を認識可能になると考える。さらに、個人によって異なる動作の遷移過程がわかることで、個人を特定することも期待できる。

そこで、本研究では、学習した複数の動作モデルを統合し、動作間のあいまいさを考慮した1つの大規模なHMM（以後、統合動作モデル）を生成する。Fig.2に、2層の階層的なHMMからなる、本動作認識システムの構造を示す。推定する人の動作は、歩行、静止、床座、横臥とする。第1層では、人の動作特徴から各動作モデルを推定する。第2層では、ある動作データに対する各動作モデルの尤度の組を、新たな動作特徴として学習することで、統合動作モデルを構築する。各HMMは、パラメータとして、各状態間の状態遷移確率行列 A 、各状態における各シンボルの出力確率分布 B 、初期状態分布 π を持ち、パラメータを用いて $\lambda = (A, B, \pi)$ と表される。本研究では、1つのモデルで多くの動作・行動を学習できる ergodic 型 HMM を使用する。

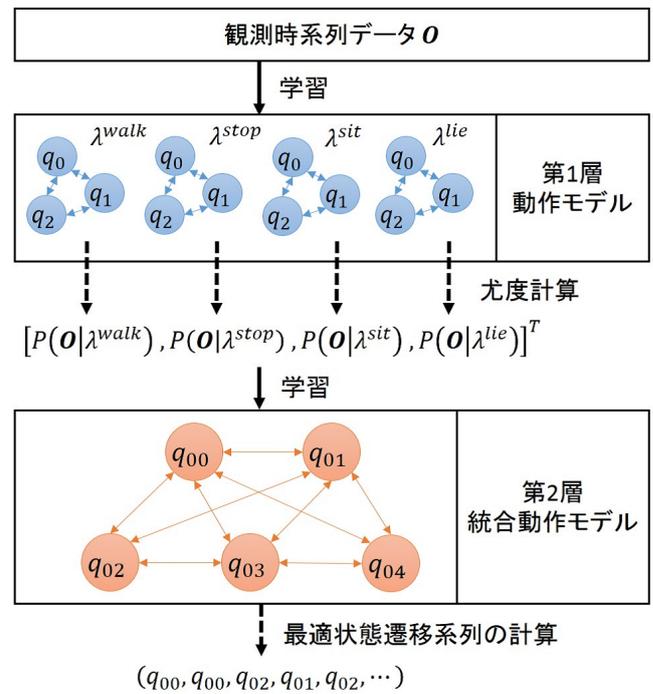


Fig.2 Schema of the H-HMM

3.1 第1層：動作モデルの構築

フロアセンシングシステムにより得られた足のクラスタ情報から、動作の特徴量を抽出する。動作特徴として、フレーム間の速度ベクトルの大きさと、クラスタ内の任意の2点の最大距離の集合である2次元ベクトル \mathbf{o} を用いる。抽出した特徴の時系列データをシンボル系列 $\mathbf{O} = (o_0, o_1, o_2, \dots, o_T)$ とし、Baum-Welch アルゴリズムによって HMM の最適パラメータ値を推定し、動作モデルを生成する。ここで、動作モデル統合処理の際、使用する時系列データ長を長くする必要がある。一方、各動作モデル構築時の学習データの長さを短くすることで、より詳細な動作推定ができると思う。そこで、前処理として、観測データを分割する。歩行を例とすると、まず、歩行データを、2秒(80フレーム)毎のデータに分割し、新たなデータの配列 $\mathbf{O}^{walk} = (o^{walk}(0), o^{walk}(1), \dots, o^{walk}(T'))$ とする。各データ $o^{walk}(i)$ は、長さ80のベクトルである。分割したデータ間の繋がりも学習するために、 $o^{walk}(i)$ と $o^{walk}(i+1)$ は1秒間のデータの重なりがある。他の動作データについても、同様の処理を行い、学習データ $\bar{\mathbf{O}} = (\mathbf{O}^{walk}, \mathbf{O}^{stop}, \mathbf{O}^{sit}, \mathbf{O}^{lie})$ を生成する。その後、得られた $\bar{\mathbf{O}}$ をそれぞれ用いて、各動作モデル $\lambda = (\lambda^{walk}, \lambda^{stop}, \lambda^{sit}, \lambda^{lie})$ を学習する。

3.2 第2層：動作モデルの統合

まず、歩行のデータ $\mathbf{o}^{walk}(t) \in \mathbf{O}^{walk}$ に対し、各動作モデルの尤度を求め、4つの尤度からなるベクトル \mathbf{p}_t^{walk} を得る:

$$\mathbf{p}_t^{walk} = \begin{pmatrix} P(\mathbf{o}^{walk}(t)|\lambda^{walk}) \\ P(\mathbf{o}^{walk}(t)|\lambda^{stop}) \\ P(\mathbf{o}^{walk}(t)|\lambda^{sit}) \\ P(\mathbf{o}^{walk}(t)|\lambda^{lie}) \end{pmatrix}. \quad (1)$$

\mathbf{p}_t^{walk} は歩行時の各動作モデルの尤度の関係を表す。これを \mathbf{O}^{walk} のすべてのデータに対して行い、得られた配列 $\mathbf{P}^{walk} = (\mathbf{p}_0^{walk}, \mathbf{p}_1^{walk}, \dots, \mathbf{p}_{T'}^{walk})$ を新たな特徴データとする。他動作についても同様に処理を行い、新たな時系列特徴データ $\bar{\mathbf{P}} = (\mathbf{P}^{walk}, \mathbf{P}^{stop}, \mathbf{P}^{sit}, \mathbf{P}^{lie})$ を生成する。この $\bar{\mathbf{P}}$ を観測シンボルとして1つのHMMに学習させることで、動作間の関係を考慮した統合動作モデルを生成する。統合動作モデルの各状態における出力確率分布は4次元であり、4つの各動作の尤度に対応している。この尤度の組み合わせによって1つの動作を表しているため、複数の動作が混同するあいまいな状態も1つの状態として認識することができる。そのため、統合動作モデルでは、「歩行」動作を複数の状態遷移によって表すことができる。これにより、より詳細な動作認識が可能となると考える。

統合動作モデルを用いた動作認識では、Viterbi アルゴリズムによってモデルの最適な状態遷移系列を推定する。すなわち、観測された動作時の時系列データに対応する、モデル内部の状態遷移系列 $\mathbf{Q} = (q_0, q_1, \dots, q_T)$ を求める。各状態がある動作に対応しているため、 \mathbf{Q} を動作の時系列データとみなすことができる。ただし、状態がどのような動作を表しているかは、現状では人手によってラベリングする必要がある。

4. 実験

提案手法の有効性を示すために、統合動作モデルを用いた動作認識の実験を行った結果を示す。実験環境を Fig.3 に示す。男性7名、女性3名を被験者とし、各動作について、以下に示すシナリオのように環境内で20~30秒の実験を行い、データを取得した。

- 歩行動作 入出口から入室し、室内を自由に歩き回り退出
- 静止動作 入出口から入室し、テーブルおよびベッドにそれぞれ近づき数秒間静止した後退出
- 床座動作 入出口から入室し、室内中央に移動し数秒間床座した後退出
- 横臥動作 入出口から入室し、室内中央に移動し数秒間横臥した後退出



Fig.3 Environment for experimentation

いずれの動作も室内への進入から開始するため、静止・床座・横臥シナリオには歩行が、床座・横臥シナリオには歩行・静止が、横臥シナリオには床座がそれぞれ含まれている。

1名の被験者の動作データをテストデータとし、残りの9人のデータを学習データとした、1-cross-validation法を用いて実験を行った。実験時に録画した画像を真値とし、各時刻における推定結果と比較して識別率を計算した。

4.1 実験結果

Table 1に、統合動作モデルの識別結果を示す。また、比較のため、各動作モデルを用いた識別結果をTable 2に示す。統合動作モデルの平均識別率は、79%であった。各動作モデルの平均識別率77%であり、提案手法により識別性能が向上している。提案手法では歩行と横臥は85%以上の識別率を得られ、さらに歩行と静止間の誤認識を減らすことができた。また、被験者によっては、個別の動作モデルでは全く認識できなかった動作が、統合することにより認識することが可能になり、統合動作モデルでの認識で誤認識が0になるケースも得られた。しかし、静止と床座の識別率が低く、お互いの動作が誤認識される結果が得られた。

4.2 考察

個別の動作モデルを統合することで、各動作の尤度を考慮した結果、認識率が向上し、明確に特徴が異なる動作間の誤認識を減らすことができた。しかし、統合したことで、似た特徴を持つ動作間の誤認識が増えた。現在の実験環境では、LRFを1つのみ使用して計測しているため、LRFに対する人の向きによっては、オクルージョンより計測されるクラスターの形に大きく影響を与えてしまう。例えば、LRFに対する座る向きによっては、得られるクラスターサイズが、静止時のものに近くなる。逆に、静止時に足を大きく開いた場合、クラスターサイズが床座時に類似してしまう。そのため、静止と床座の誤認識が起きていると考えられる。また、歩行が静

Table 1 Proposed method(Integrated HMM)

		推定結果				識別率
		歩行	静止	床座	横臥	
真値	歩行	<u>359</u>	45	11	6	0.85 ± 0.14
	静止	26	<u>112</u>	31	3	0.65 ± 0.30
	床座	14	29	<u>91</u>	3	0.66 ± 0.44
	横臥	0	1	0	<u>185</u>	0.99 ± 0.02

Table 2 Conventional method

		推定結果				識別率
		歩行	静止	床座	横臥	
真値	歩行	<u>339</u>	59	12	11	0.80 ± 0.09
	静止	37	<u>106</u>	25	4	0.62 ± 0.28
	床座	12	12	<u>95</u>	18	0.69 ± 0.39
	横臥	2	0	1	<u>183</u>	0.98 ± 0.03

止に誤認識されてしまうこともわかる。これは、歩行中に方向転換を行った際の一瞬の静止が原因であると考えられる。

こうした問題に対し、室内に床上 LRF を増やし、多方向から計測を行うことで、オクルージョン領域を減らし計測されるクラスタを増やすことが有効であると考えられる。また、本研究では、分割したデータ $\mathbf{o}(t)$ のみの動作識別を検討しているため、ある瞬間での誤認識が、システム全体の性能に与える影響が大きい。今後、時系列で連続するデータの識別結果の積算を使用し、行動識別をすることで、瞬間の誤認識に対して頑強な識別が期待できる。

5. まとめ

動作間のあいまいさを考慮した動作認識を行うため、複数の動作モデルを統合する手法を提案し、動作認識に使用する特徴として、フロアセンシングシステムで得られる床面の人のクラスタサイズと、クラスタの重心のフレーム間速度の大きさを使用した。統合した動作モデルを用いた動作認識の実験を行った結果、統合前の各動作モデルによる認識より識別率が改善される動作と改悪される動作がわかり、改悪された動作については解決策を示した。今後の課題として、考察で述べた解決策の実装を含めた HMM 統合手法の再検討を行い識別率を向上させることが挙げられる。展望として、動作モデルの識別結果に、動作時の室内の家具との位置関係などの特徴を付加し、居住者の生活行動を推定するようシステムを拡張することが挙げられる。

文献

- [1] 村上剛司, 長谷川勉, 木室義彦, 千田陽介, 家永貴史, 有田大作, 倉爪亮, “情報構造化環境における情報管理の一手法”, 日本ロボット学会誌, Vol.26, No. 2, pp.192-199, 2008
- [2] Pau-Choo Chung, Chin-De Liu, “A daily behavior enabled hidden Markov model for human behavior understanding”, Pattern Recognition, Vol.41, Issue.5, pp.1572-1580, 2008
- [3] Taketoshi Mori, Shoji Tominaga, Hiroshi Noguchi, Masamichi Shimosaka, Rui Fukui, and Tomomasa Sato, “Predicting Behaviors of Residents by Modeling Preceding Action Transition from Trajectories”, Journal of Robotics and Mechatronics, Vol.24, No.2, pp.320-329, 2012
- [4] 長谷川勉, Pyo Yoonseok, 田中真英, 辻徳生, 諸岡健一, 倉爪亮, “床上センシングシステムを用いた生活環境における移動物体の位置計測と居住者の行動推定”, 日本ロボット学会誌, Vol.31, No.8, pp.769-779, 2013