隠れマルコフモデルと動作の階層構造の木表現による 日常動作認識

森 武 俊* 瀬 川 友 史* 下 坂 正 倫* 佐 藤 知 正*

Recognition of Human Daily Actions Based on Continuous Hidden Markov Models and Hierarchical Structure of Actions as Tree Representation

Taketoshi Mori*, Yushi Segawa*, Masamichi Shimosaka* and Tomomasa Sato*

This paper presents a recognition method of human daily-life action. The method utilizes hierarchical structure of actions and describes it as tree. We modelize actions by Continuous Hidden Markov Models which output time-series feature vectors extracted based on knowledge of human. In this method, recognition starts from the root, competes the likelihoods of child-nodes, chooses the maximum one as recognition result of the level, and goes to deeper level. The advantages of hierarchical recognition are: (1) recognition of various levels of abstraction, (2) simplification of low-level models, (3) response to novel data by decreasing degree of details. Experimental result shows that the method is able to recognize some basic human actions.

Key Words: Action Recognition, Motion Capture, HMM, Novelty Detection, Various Levels of Abstraction

1. は じ め に

ロボットなどの機械による人間の生活支援や、機械と人間とのスムーズな意思疎通を考えた場合、機械が人の動作を認識・理解し、人と動作の概念を共有することが望まれる。従来の動作認識の研究の多くはジェスチャ[1]や手話[2]を対象としているが、人の動作の大部分は歩く、座るなどの日常動作であることを考えると、日常動作の認識も重要である。よって本研究では、日常動作を認識の対象とする。

また、一般的な動作認識の研究では、各動作は個別・並列に扱われ、多くの場合動作の概念間の関係は利用されていない。しかしながら、動作の概念間には、排他性、同時性、階層性などの関係があるはずである。そのような関係を利用することで多様で頑強な認識が可能になると考え、本研究ではそのうちの抽象度の違いから生じる階層構造に着目して認識を行う。階層構造を木構造として表現し利用することで、一つの動作を様々な詳細さでとらえることができる。

抽象度の違いに着目した研究は、ほかにも行われている。例えば Bui ら [3] は Abstract Hidden Markov Model を提案し、ビル内の人の位置を実際の座標、部屋、階、棟、ビル全体などの様々な抽象さで表現しているし、Ogawara ら [4] は詳細さの異なる二つの段階から構成される注視点解析により、人の机上

作業を解析・認識している。また、Oriver ら [5] は HMM を階層状にした Layered Hidden Markov Model を用いて Distant Conversation, Face to Face Conversation, Phone Conversation などのオフィスでの人の活動を認識している。認識は、まず音などのデータが人の声、キーボードの音などに認識され、その認識結果を基に一つ上の階層の HMM がより抽象的な活動を認識し、といったように、抽象度が低い方から高い方へと進む。しかし、これらの研究は人間の日常動作・全身動作を対象としたものではない。本研究では日常動作の概念の階層関係を整理し認識に利用する。

本研究が対象とする日常動作は、(1) 複雑・多様性、(2) 曖昧性、(3) 時系列性、(4) 構造的性という四つの特性を持つものとしてとらえられる。一つ目の複雑・多様性とは、日常動作は全身を用いた動作であり、動作それぞれを特徴付ける運動情報が様々であるということである。また、二つ目の曖昧性とは、日常動作はジェスチャや手話のようにあらかじめ定まった意味を持たないため、個人差や時と場合による変化が大きく同一の運動に対する解釈にも幅があることを指す。三つ目の時系列性とは、日常動作はある瞬間だけを見ても判断できず、時系列の情報として扱う必要があることを示す。最後の構造的性とは、前述した動作間の関係により生じるものである。これを踏まえて、本研究では

- 複雑で多様な日常動作を的確に捉えるため、人の動作に関する知識を特徴量の抽出に利用
- 日常動作の曖昧さ・時系列性を取り扱うため、隠れマルコ

原稿受付 2004 年 8 月 9 日

*東京大学大学院情報理工学系研究科

*Faculty of Information Science and Engineering, Univ. of Tokyo

フモデル (HMM) を利用

- 多様で頑強な認識のため、動作の階層構造を木表現し認識に利用

という三つの特徴を持つ日常動作認識手法を構築することを目的とする。

二つ目の特徴として採用した HMM は不確実性を含む時系列データを扱う枠組みとして一般的なものであり、動作認識にもよく利用される。例えば Yamato ら [6] は、連続画像から背景差分により人領域を抽出し、その 2 値化画像からメッシュ特徴を計算してベクトル量子化により時系列のシンボル列を得て、それを観測列とする離散 HMM によりフォアハンド、サーブなどの 6 種類のテニスの動作を認識している。また、Starnier ら [7] は画像から得られた二次元の両手の座標を基に、HMM を用いて American Sign Language の 40 個の単語をモデル化して文単位の認識をしているし、Brand ら [8] は、二つのマルコフ連鎖が相互作用するモデルである Coupled HMM を提案し、ステレオ画像から得た両手の三次元座標を基に、太極拳における三つの両手を使った動作を認識している。

全身動作を対象とした HMM を用いた研究としては、Yamato らの研究のほかに、Inamura ら [9] の研究が挙げられる。Inamura らは、HMM を用いることで人間の全身の関節角の時系列データから動作のシンボル表現 (原始シンボル) を得て、それを用いてヒューマノイドロボットの動作生成を行うというミメシスループを提唱している。これらと本研究の最も大きな違いは、動作の概念の階層構造を利用していることである。

研究の方法としては、まず上記三つの特徴の一つ目と二つ目のアプローチにより対象動作ごとの動作モデルを構築し、さらに三つ目のアプローチによりそれらを木表現を用いて構造化する。

2. 階層構造による認識手法の構成

2.1 入力：時系列の運動データ

動作認識は全身の運動データの取得という問題と、その取得された運動の解釈・認識という問題との二つに大きく分けられる。本研究ではそのうちの後者に注力する。そのため、計測の簡便性を考慮し入力としてモーションキャプチャにより得られた時系列の人間の全身運動データを用いる。使用したのはモーションファイルフォーマットとして一般的な Biovision 社の BVH 形式である。BVH 形式では、人は剛体の連結モデルとして表現される。本手法が採用した人間の骨格モデルは Fig. 1 のようなもので、11 個の関節により構成され、合計で 36 自由度であ

Fig. 1 Human figure utilized in this research

る。また、運動計測時のサンプリングレートは 30 [Hz] である。

本手法では、認識はそれぞれの BVH ファイル単位で行われ、そのファイルに含まれる一定時間の運動が何であるかが識別される。つまり、人間の連続的な運動を短く分節化する作業はあらかじめ人が行い、本手法はその分節化後の各セグメントに対する認識を行う。

2.2 階層構造による認識の概要

認識対象の動作それぞれは対応するモデル (動作モデル) を持っている。それぞれの動作モデルは独立に、入力されたデータからその動作を特徴付ける運動情報を抽出し、それを基に入力がどれくらいその動作らしいかを表す尤度を計算する。この動作モデルの詳細については、第 3 章で説明する。

本手法ではさらに動作間の階層構造を考慮し、動作モデルを構造化する。具体的には、一つのルートから派生し、個々の動作モデルをノードとする木表現を用いる。認識プロセスはルートから始まり、上の階層から下の階層へと進む。下の階層に進むに従って、認識はより詳細なものになる。

各階層ではその階層に属する動作モデルの尤度を比較し、最も確からしいノードを選んでそれをその階層における認識結果とし、そのノードに遷移して次の階層の認識へと進む。これを木構造のリーフまで繰り返すことで、例えば「Lying - LyingOnSide - LyingOnLeftSide」のような異なる抽象度での複数の階層に渡る認識結果が一つの運動データに対し得られる。

このように競合による認識を行う場合、どの動作にも属さないような新奇な運動の取り扱いが問題となる。これについては、各階層において「その他」という特殊なカテゴリを用意し、「競合結果に対する信頼度」という概念を利用した新奇判別手法により対処する。これについても詳細は後述する。

2.3 階層構造による認識の利点

階層構造を用いて認識を行うことで、以下のような三つの利点を得られる。

(1) 様々な詳細さでの認識

各階層で最も確からしい動作を選ぶことで、一つの運動データに対し様々な詳細さでの複数の認識結果を得ることができる。

(2) 下位のモデルの単純化

詳細な動作の認識には本来多くの特徴量が必要であるが、上の階層の認識結果を利用することで必要な特徴量の数を減らすことができる。例えば、ある運動データが「寝ている」と認識された場合にはそれ以下の階層ではそれを事前知識として利用でき、「横向きに寝ている」などを認識するためには向きだけを見ればよい。

(3) 新奇動作への対応

大雑把から詳細へ至る認識の流れにより、認識システムが動作モデルを持たない新奇な動作が入力された場合にも認識の詳細さを下げることで対応できる。例えば、変わった座り方が入力されても、より大雑把に「座っている」ということは認識できるなどである。

3. 動作モデル

ある動作のモデルは、Fig. 2 に表されるように、特徴量抽出フィルタと HMM という二つの要素から成る。

Table 1 Example of utilized features

Action Name	Feature Expression by Human	Extracted Physical Quantity
Sitting	Upper body is almost straight.	vertical distance from Hips to Head
	Hips are at the level of chair or floor.	vertical distance from Ground to Hips
	Hips are bent.	distance from Hips to the segment that connects Head and Knees
SittingOnFloor	The person is sitting.	recognition result of the upper level
	Hips are at the level of floor.	height of Hips
	Legs are at the level of hips.	vertical distance from Feet to Hips
SittingOnChair	The person is sitting.	recognition result of the upper level
	Hips are at the level of chair.	height of Hips
	Hips are higher than legs.	vertical distance from Feet to Hips
Agiya	The person is sitting on floor.	recognition result of the upper level
	Knees are bent and apart from each other.	inclination of Thighs from vertical direction
	"	inclination of Shins from vertical direction
	Legs are crossed.	distance between Feet
Seiza	The person is sitting on floor.	recognition result of the upper level
	Knees are at the level of floor.	height of Knees
	Knees are bent.	angle of Knees
	Hips are on legs.	vertical distance from Feet to Hips
	"	horizontal distance from Hips to Feet
	Shins are under thighs.	inclination of Shins from vertical direction
Chouza	The person is sitting on floor.	recognition result of the upper level
	Knees are at the level of floor.	height of Knees
	Knees are straight.	angle of Knees
Ryoutatehiza	The person is sitting on floor.	recognition result of the upper level
	Knees are raised.	height of Knees
	Knees are bent.	angle of Knees
	Knees are close to each other.	inclination of Thighs from vertical direction
	"	inclination of Shins from vertical direction

Time Series 3D Motion Data of Human

Fig. 2 Example of action model (Standing)

3.1 特徴量抽出フィルタ

特徴量抽出は、日常動作認識において最も重要な要素の一つである。日常動作は複雑・多様であるが、人は直感的に各動作に固有の特徴を抽出し、その動作の要素を言葉で表現することができる。よって本研究では、各動作の特徴量抽出フィルタを人による動作の条件記述に基づいて作成する。例えば Standing の場合、人による動作の記述は「頭が高い」「腰が高い」「体が

ほぼ真っ直ぐ」のようになり、それに対応して Standing の特徴量抽出フィルタは、「頭の高さ」「腰の高さ」「腰と足先の水平方向のずれ」という物理量を抽出する。各動作の条件記述は、過去の研究 [10] を参考にし、数人にアンケートを取ることで作成した。また、それらに対応し抽出される特徴量については、試行錯誤を繰り返しながら妥当なものを選択した。

いくつかの動作に対する動作の記述と抽出される物理量の例を、Table 1 に示す。

3.2 隠れマルコフモデル

特徴量抽出フィルタが「どの情報に注目すべき動作か」を表現しているのに対し、HMM はそれが「どんな動作か」を時系列データのシンボル表現としてモデル化している。以下で、HMM について説明する。

3.2.1 隠れマルコフモデルの概要

HMM は確率的な時系列データの生成モデルである。時間とともに有限個の状態間を確率的に遷移しながら、各状態において確率的に値を出力する。この HMM から出力される時系列の値を観測列と呼ぶ。出力される値は離散値の場合と連続値の場合とがあるが、本研究では連続値である。具体的には、前述した特徴量抽出フィルタにより時系列の全身の運動データから抽出された時系列の特徴量ベクトルが観測列である。また、HMM において状態遷移はマルコフ性に従う。ここで、マルコフ性とはある時刻における状態が、その直前の状態のみに依存して決まるという性質のことである。

HMM のトポロジーは、あらゆる状態遷移を許す既約なもの、初期状態から終了状態への一方向の遷移のみを許す LR 型など様々であるが、本研究では Fig. 3 のようなリング状のものを採

Fig. 3 HMM utilized in this research

用した。

3.2.2 隠れマルコフモデルの構成

HMM は $\theta = \{S, A, B, \pi, q\}$ という五つのパラメータにより定義される。式で表すと、

$$S = \{S_1, S_2, \dots, S_N\}$$

$$A = \{a_{ij}\}$$

$$B = \{b_i(o)\}$$

$$\pi = \{\pi_i\}$$

$$q = \{q_i\}, \quad q_i = P(\text{stop} | s_T = S_i)$$

のようになる。それぞれ、 S_i は取り得る状態、 a_{ij} は状態 S_i から状態 S_j へ遷移する確率、 $b_i(o)$ は状態 S_i において観測値 o を出力する確率、 π_i は初期時刻において状態 S_i である確率、終端状態確率 q_i は状態遷移が終端時刻 $t = T$ において状態 S_i で終わる確率である。また、 N は状態集合の要素数を示す。

この定義から分かるように、HMM では状態遷移列はマルコフ性に従い、一方観測列は互いに独立でその時刻における状態のみに依存して決まる。つまり、HMM は少数の離散的な状態の確率的な連結で連続信号を表現しているといえる。

出力確率 $b_i(o)$ として、本研究では複数のガウス分布の重み付き和として表現される混合ガウス分布を用いる。つまり、 $b_i(o)$ は

$$b_i(o) = \sum_{k=1}^K C_{ik} \frac{1}{\sqrt{(2\pi)^D |\Sigma_{ik}|}} e^{-\frac{1}{2}(o - \mu_{ik})^T \Sigma_{ik}^{-1} (o - \mu_{ik})}$$

として表現される。ここで、 K は混合するガウス分布の数（混合数）を表し、 μ_{ik} および Σ_{ik} はそれぞれ状態 S_i の k 番目のガウス分布の平均と共分散行列を表す。また、 D は出力 o の次元で、 C_{ik} は状態 S_i の k 番目のガウス分布の重み（混合率）を示す。 C_{ik} は、 $\sum_{k=1}^K C_{ik} = 1$ をすべての i について満たす。

結局、出力確率 $b_i(o)$ は平均 μ_{ik} 、分散 Σ_{ik} 、混合率 C_{ik} の三つのパラメータで定義されることとなる。

3.2.3 HMM による尤度計算

認識の際には、観測列 $O = \{o_1, o_2, \dots, o_T\}$ を得たときに、それがモデル M に属する確率、 $P(M|O)$ を考え比較する必要がある。ここでベイズの定理より

$$P(M|O) = \frac{P(O|M)P(M)}{P(O)}$$

であり、 $P(O)$ は観測列のみによって決まるので、結局は各モ

デルの $P(O|M)P(M)$ を計算し比較する問題となる。さらに、事前分布 $P(M)$ は全モデルで等しいとすれば各モデルの尤度 $P(O|M)$ を比較すればよい。ここで、 $P(O|M)$ は

$$P(O|M) = \sum_s P(O, S|M) \\ = \sum_s \left(\pi_{s_1} b_{s_1}(o_1) \cdot \prod_{t=2}^T a_{s_{t-1}s_t} b_{s_t}(o_t) \cdot q_{s_T} \right)$$

のように展開できる。ここで、 s は考えられる状態遷移列 $\{s_1, s_2, \dots, s_T\}$ 全体の集合を示す。つまりモデル M の尤度 $P(O|M)$ はあらゆる状態遷移列を考え、それらから与えられた観測列 O が出力される確率の総和をとることで計算できる。実際には、あらゆる状態遷移列を考え和をとるという方法は計算量が膨大になり現実的でないため、再帰的な効率の良い方法である、前向きアルゴリズムあるいは後ろ向きアルゴリズム [11] を用いて尤度を計算する。

前向きアルゴリズムは、時刻 t の状態は時刻 $t-1$ のみに依存するということを利用して効率良く尤度の計算を行うアルゴリズムである。まず、時刻 t までの部分観測列が $o_1, o_2, \dots, o_t$ であり、かつ時刻 t での状態が S_i である確率を前向き確率 $\alpha_t(i)$ とし、以下のように定義する。

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, s_t = S_i | M)$$

すると、尤度 $P(O|M)$ は、最終時刻 T までの観測列が $o_1, o_2, \dots, o_T$ であり、かつ時刻 T での状態が S_i である確率 $\alpha_T(i)$ と、状態 S_i での終了確率 q_i を用いて、

$$P(O|M) = \sum_{i=1}^N \alpha_T(i) q_i$$

のように表される。そして、 $\alpha_t(i)$ は以下のように再帰的に求めることができる。

$$\begin{cases} \alpha_1(i) = \pi_i b_i(o_1) & (1 \leq i \leq N) \\ \alpha_{t+1}(i) = \sum_{j=1}^N \alpha_t(j) a_{ji} b_i(o_{t+1}) & (1 \leq t \leq T-1, 1 \leq i \leq N) \end{cases}$$

一方の後ろ向きアルゴリズムでは、時刻 t において状態 S_i であり、それ以降の状態遷移によって $o_{t+1}, o_{t+2}, \dots, o_T$ が出力される確率を考える。それを後ろ向き確率 $\beta_t(i)$ とすると、

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T, s_t = S_i | M)$$

という式で表現できる。これを用いて、前向きアルゴリズムとは逆向きに、

$$\begin{cases} \beta_T(i) = q_i & (1 \leq i \leq N) \\ \beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) & (t = T-1, T-2, \dots, 1, 1 \leq i \leq N) \end{cases}$$

のようにして、 $\beta_t(i)$ を計算することが可能であり、

$$P(\mathbf{O}|\mathbf{M}) = \sum_{i=1}^N \pi_i b_i(\mathbf{o}_1) \beta_1(i)$$

により尤度を求めることができる。

3.2.4 HMM の学習

前節では、HMM の各パラメータが決定した下で、観測列 $\mathbf{O}$ が与えられた場合の尤度の計算について述べた。本項では、そもそもその HMM の各パラメータをどのようにして決定するかについて説明する。

HMM は、与えられた時系列データのデータセットから各パラメータを学習する。学習とは、与えられたパターンを発生しやすいような HMM のパラメータを推定することである。ここでいうパラメータとは $\theta = \{a_{ij}, \pi_i, \mu_{ik}, \Sigma_{ik}, C_{ik}, q_i\}$ のことであり、状態数やガウス分布の混合数、トポロジーは試行錯誤的に決める。

確率モデルのパラメータの推定は、最尤推定により行うのが普通であるが、HMM では観測列を生成した状態遷移列は非観測であるため、直接最尤推定を行うことはできない。このように、観測空間における確率変数と状態空間における確率変数とが存在するようなモデルでは、EM アルゴリズム (Expectation-Maximization Algorithm, 期待値最大化法) [12] に基づく繰り返し計算によりパラメータを徐々に最適化していくことで近似的に推定することになる。

本研究でも、この EM アルゴリズムによって導かれる以下のような更新式を用いて HMM のパラメータを逐次的に推定する。ただし、これらの式は E 個の観測列が学習用に与えられた場合の推定値であり、変数 $X^{(e)}$ は e 番目の観測列から求められる X の値を表している。式中の $\gamma_t(i)$ および $\xi_t(i, j)$ は

$$\begin{aligned} \gamma_t(i) &\equiv P(s_t = S_i | \mathbf{O}) \\ &= \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)} \\ \xi_t(i, j) &\equiv P(s_t = S_i, s_{t+1} = S_j | \mathbf{O}) \\ &= \frac{\alpha_t(i) a_{ij} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_{t-1}(i) a_{ij} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)} \end{aligned}$$

のように定義される変数であり、前述の前向き確率 $\alpha_t(i)$ と後ろ向き確率 $\beta_t(i)$ によって計算できる。また、 $\gamma_t(i, k) = P(s_t = S_i, C_i = k | \mathbf{O})$ は観測列 $\mathbf{O}$ を得たときの、時刻 t における観測値が状態 i の k 番目の分布から得られる確率を示す。 C_i は状態 i における混合分布の番号を表す変数である。

$$\begin{aligned} \hat{\pi}_i &= \frac{\sum_{e=1}^E \gamma_1^{(e)}(i)}{E} \\ \hat{q}_i &= \frac{\sum_{e=1}^E \gamma_T^{(e)}(i)}{E} \\ \hat{a}_{ij} &= \frac{\sum_{e=1}^E \sum_{t=1}^{T_e} \xi_t^{(e)}(i, j)}{\sum_{e=1}^E \sum_{t=1}^{T_e} \gamma_t^{(e)}(i)} \\ \gamma_t^{(e)}(i, k) &= \gamma_t^{(e)}(i) \frac{C_{ik} P(\mathbf{o}_t | s_t = S_i, C_i = k)}{P(\mathbf{o}_t | s_t = S_i)} \\ \hat{C}_{ik} &= \frac{\sum_{e=1}^E \sum_{t=1}^{T_e} \gamma_t^{(e)}(i, k)}{\sum_{e=1}^E \sum_{t=1}^{T_e} \gamma_t^{(e)}(i)} \end{aligned}$$

$$\begin{aligned} \hat{\mu}_{ik} &= \frac{\sum_{e=1}^E \sum_{t=1}^{T_e} \gamma_t^{(e)}(i, k) \mathbf{o}_t^{(e)}}{\sum_{e=1}^E \sum_{t=1}^{T_e} \gamma_t^{(e)}(i, k)} \\ \hat{\Sigma}_{ik} &= \frac{\sum_{e=1}^E \sum_{t=1}^{T_e} \gamma_t^{(e)}(i, k) (\mathbf{o}_t^{(e)} - \hat{\mu}_{ik})(\mathbf{o}_t^{(e)} - \hat{\mu}_{ik})^t}{\sum_{e=1}^E \sum_{t=1}^{T_e} \gamma_t^{(e)}(i, k)} \end{aligned}$$

3.3 動作モデルの出力

時系列の全身の運動データ $\mathbf{Y}$ がある動作 act の動作モデル M_{act} に入力されると、動作 act の特徴量抽出フィルタ Φ_{act} によって動作 act を特徴付ける運動情報 $\Phi_{act}(\mathbf{Y})$ が抽出される。次にこの抽出された時系列の特徴量ベクトルが、あらかじめ動作 act の運動データからパラメータを学習済みの HMM に渡され、動作 act の入力データ $\mathbf{Y}$ に対する尤度 $\mathcal{L}_{act}(\mathbf{Y})$ が計算される。尤度は、計算時のアンダーフロー防止のため、以下のような対数尤度を用いた。

$$\mathcal{L}_{act}(\mathbf{Y}) = \ln P(\Phi_{act}(\mathbf{Y}) | M_{act})$$

これは入力データ $\mathbf{Y}$ がどれくらい動作 act らしいかを示しており、認識のための指標となる。しかし、この尤度の尺度は動作モデル間で異なっているため、動作モデルごとに正規化係数を定めてそれによる正規化を行い、動作モデル間で比較可能な値にする。

例えば Fig. 4 の左に示されるように、StandingStill と Walking の動作モデルでは Walking の動作モデルの方が出力する尤度が全体的に大きいため、ただ比較したのでは StandingStill のデータも Walking であると見なされてしまう。一方、正規化を施せば、Fig. 4 の右に示されるように、StandingStill と Walking の動作モデルの出力する尤度の比較が妥当に行われる。

正規化係数としては、その動作の HMM の学習に用いたデータセットに対する尤度の平均値を用いる。つまり動作 act の正規化係数 C_{act} は、 $\mathbf{X}_i \in act$ の学習データとして、

$$C_{act} = \text{average}_i (\mathcal{L}_{act}(\mathbf{X}_i))$$

のように表される。このような正規化を経て、最終的に、以下のような正規化尤度 $\mathcal{N}_{act}$ が動作モデル M_{act} の入力データ $\mathbf{Y}$ に対する出力として得られる。

$$\begin{aligned} \mathcal{N}_{act}(\mathbf{Y}) &= \mathcal{L}_{act}(\mathbf{Y}) - C_{act} \\ &= \mathcal{L}_{act}(\mathbf{Y}) - \text{average}_i (\mathcal{L}_{act}(\mathbf{X}_i)) \end{aligned}$$

Fig. 4 Example of normalization

認識は、基本的にこの正規化尤度 $\mathcal{N}_{act}$ を競合し、最も確からしい動作を選ぶことで行う。

この正規化は、動作 act のデータに対する動作モデル M_{act} の尤度が正規分布に従うと仮定したもので、単なる尤度をデータとモデルとの近さに変換する意味合いがある。実際、このような正規化により「ある値とモデルの出力する値の分布との近さ」を考える例は多く、例えば Tax ら [13] は確率密度関数 $p(x)$ を用いて外れ値を検出する際の指標として、

$$\log(p(x)) - \mathcal{E}[\log(p(x^{tr}))] < -3.0 * \text{var}[\log(p(x^{tr}))]$$

のように、学習データに対する確率密度の値の平均値からのずれがある閾値以上かどうかを用いている。ただし、上式で $\mathcal{E}[y]$ は y の平均値を、 x^{tr} は学習データを、 $\text{var}[y]$ は y の分散を示している。

4. 階層構造による日常動作認識

4.1 階層構造の木表現

本研究の対象動作とその階層構造の木表現は、Fig. 5 に示されるようなものである。Fig. 5 において、実線で囲まれたノードは対応する動作モデルが存在するノードを示し、破線で囲まれたノードは「その他」という特殊なノードを示している。この「その他」というノードは、各階層においてどんな分類の仕方をしてもし必ずカバーしきれない領域が存在することを考慮したものである。「その他」には対応する動作モデルは存在せず、後述する新奇判別のアルゴリズムによって入力システムにとって新奇、すなわちその階層において確からしい動作が一つもないと判断された場合にこのノードに遷移する。

木構造は、以下の方針に従って人が作成する。

- 同じ親ノードを持ち並列関係にあるノードは、同時に起こりえない動作である。
- ノードが親子関係にある場合、子ノードは親ノードの成立が前提の動作である。

Fig. 5 Tree structure of actions

なお、「床に座っている (SittingOnFloor)」以下のカテゴリについては文献[14]を参考に作成した。

本研究では Fig. 5 のように木構造を定義したが、これは絶対的なものではなく、その他の分類の仕方も当然あり得る。例えば、Fig. 5 の木構造は初めに姿勢的な要素に着目して分けていくものとなっているが、その代わりに人の移動に着目して、まず「移動している (Moving)」, 「移動していない (Staying)」のように分けた後、「移動している」の下に「走っている (Running)」や「歩いている (Walking)」, 「這っている (Crawling)」が来て、一方「移動していない」の下には「立ち止まっている (StandingStill)」や「座っている (Sitting)」, 「寝ている (Lying)」が来る、といった構造を定義することも可能である。つまり、Fig. 5 の木構造は、人が日常動作を分類するときの見方のうちの一つを反映したものであると言える。

また、木構造を定義し、「その他」というカテゴリを用意したことで、拡張も容易である。例として「胡座 (Agura)」などが属する階層を用いて説明すると、座り方には本研究で対象としたもの以外にも、「片足を伸ばして座る」、「横座り」など様々なものがあるが、これらは新たに認識したいのであれば動作モデルを作成して「胡座」などが属する階層に加えればよいし、特に区別して認識する必要がないのであれば動作モデルを作る必要はなく、「その他 (EtcSittingOnFloor)」として扱うことができる。その場合でも、より大雑把に「床に座っている (SittingOnFloor)」とは認識できる。

4.2 階層構造による認識の流れ

階層構造を利用した認識では、ある一つの運動データは様々な詳細さで解釈される。まずは大雑把な認識が行われ、認識プロセスが下の階層に移るにつれてより詳細な認識が行われる。具体的には、認識は以下のような流れで行われる。Fig. 6 は認識の流れを、深さ 2 の階層の例で示したものである。

- (1) 時系列の全身運動データが入力されると、認識プロセスはルートノードから始まる。
- (2) 対応する動作モデルを持たない「その他」を除き、すべての子ノードについて対応する動作モデルの正規化尤度 $\mathcal{N}_{act}$ が計算される。
- (3) 正規化尤度 $\mathcal{N}_{act}$ が最大の子ノードが、その階層における競合結果として選ばれる。
- (4) 選ばれた子ノードの正規化尤度 $\mathcal{N}_{act}$ を基に、その競合結果が信頼できるものかどうかを判断するための新奇判別 (Confidence Check) が行われる。
- (5) 新奇判別により競合結果に対する信頼度 (Confidence Value) が十分大きいと判断された場合、競合結果がその階層における最終的な認識結果となる。一方信頼度が小さい場合には、その階層における認識結果は「その他」となる。
- (6) 認識結果の子ノードに遷移し、一つ下の階層の認識へ移る。
- (7) (2) から (6) を、認識プロセスがリーフノードに到達するまで繰り返す。

このような認識の流れにより、例えば「Lying - LyingOnSide - LyingOnLeftSide」のような階層的な認識結果が得られる。この認識結果は、入力された運動データが三つの異なる抽象度で認識されたことを示しており、各階層での認識結果はそれぞれ、

Fig. 6 Flow of recognition

第1階層で「Lying」、第2階層で「LyingOnSide」、第3階層で「LyingOnLeftSide」である。

4.3 新奇動作の判別

4.3.1 新奇判別の導入

本手法は、尤度の競合による認識を基本としているが、単純に最も尤度の大きいノードを認識結果としてしまうと、システムが動作モデルを持っていない新奇なデータに対処できなくなってしまう。例えば、もし変わった座り方の運動データが入力として与えられた場合、システムはその座り方について何の知識も持っていないが、すでに動作モデルとして知識を持っている動作の中から単純に最も尤度の大きいものを選んでしまう。その結果、仮に値としては小さくとも、その中で正座の類似度が尤度が最大ならば、「Sitting - SittingOnFloor - Seiza」のような認識結果を出してしまう。

しかしこの例の状況において、第3階層において正座の動作モデルの正規化尤度 $\mathcal{N}_{Seiza}$ が最大であるとしても、その変わった座り方のデータに対する $\mathcal{N}_{Seiza}$ は、普通の正座のデータに対する $\mathcal{N}_{Seiza}$ よりも小さい値であると考えられる。したがって、本研究では各階層において、「競合結果に対する信頼度 (Confidence Value) = 「競合により選ばれたノードの正規化尤度 $\mathcal{N}_{act}$ 」」と定義する。そして、その信頼度の値がある階層においてあらかじめ設定した閾値よりも小さい場合には、その階層における競合結果は棄却され、その運動データは新奇なものに見なされ、「その他」に分類される。信頼度の値が閾値よりも大きい場合には、競合結果がそのまま認識結果となる。これが前節の認識の流れの(4)で行われる新奇判別 (Confidence

Check) の詳細である。

このような新奇判別を行うことで、例えば「横座り」というシステムが動作モデルを持っていない新奇な座り方のデータが入ってきた場合には、「Sitting - SittingOnFloor - EtcSittingOnFloor」のような認識結果が得られるはずである。これは「横座り」を知らないシステムが出す認識結果として、妥当なものである。

4.3.2 新奇判別の閾値の決定

問題は、どのようにして信頼度の閾値を決定するかである。ここで考慮しなければならないのは、正規化尤度 $\mathcal{N}_{act}$ の値の傾向は、動作モデルによって異なるということである。尤度の正規化によりそろえるのは、各動作モデルが自分のカテゴリに属する運動データに対して出力する尤度の値のスケールである。一方の、各動作モデルが自分のカテゴリに属さない運動データに対して出力する尤度の値の振る舞いは、各動作モデルに依存したばらばらなものである。これは、各動作モデルが学習するのは自分のカテゴリに属する運動データのみであり、他のカテゴリに属する運動データについての振る舞いについてはどのようなか保障されないことを考えると、当然のことである。

この動作モデルによる $\mathcal{N}_{act}$ の傾向の違いを考慮すると、信頼度の閾値は動作モデルごとに決定するのが妥当である。本研究では、以下のようなアルゴリズムにより動作モデル M_{act} の信頼度の閾値 $\mathcal{K}_{act}$ を決定した。

新奇判別のための閾値決定アルゴリズム

全学習データセットのうち、動作 act の動作モデル M_{act} が属する階層のデータ群 $\{\mathbf{X}_i\}$ を考える。

それらに対し、以下の二つの値 A , B を計算する。

$$A = \min_i (\mathcal{N}_{act}(\mathbf{X}_i)), \quad \text{where } \mathbf{X}_i \in act$$

$$B = \max_i (\mathcal{N}_{act}(\mathbf{X}_i)),$$

$$\text{where } \mathbf{X}_i \notin act \quad \text{and} \quad \mathcal{N}_{act}(\mathbf{X}_i) < A$$

そして、以下のように A , B の平均を動作 act の信頼度の閾値として採用する。

$$\mathcal{K}_{act} = \frac{A + B}{2}$$

以上のアルゴリズムにおいて、 A はその階層における学習データ中の、動作 act に属するデータに対する動作モデル M_{act} の正規化尤度 $\mathcal{N}_{act}$ の最小値を表し、 B は動作 act に属さないデータに対する動作モデル M_{act} の正規化尤度 $\mathcal{N}_{act}$ のうち、 A の値を超えないものの最大値を表す。 A , B の値の算出の過程は、Fig. 7 に示されるようなものである。

B を A を超えない値の中から選ぶ理由は、動作 act に属さないデータに対する $\mathcal{N}_{act}$ の値は、動作 act に属するデータに対する $\mathcal{N}_{act}$ の値よりも信頼できないものであるからである。つまり、動作 act に属さないデータに対しても動作モデル M_{act} が大きな尤度を出してしまうことはあり得るが、そのような値

Fig. 7 Key values for novelty detection threshold

は信頼できないものとして無視するということである。これは、動作モデル M_{act} は動作 act に属するデータのみを学習していることに起因する。

なお、本手法ではある動作のモデルという正の分布のみを用いて閾値を決定したが、多くのこのような外れ値検出の問題では、閾値の決定の際に「その動作である」モデルと「それ以外である」モデルとを用いて、その境界を Support Vector Machine [15] やベイズ判別などの統計的な手法により定めることが多い。本稿では詳細は省くが、このような手法と本研究における新奇判別手法との日常動作を対象とした判別性能の比較を行った結果、本研究の手法が統計的手法と同等の性能を持つことが示された。

5. 評価実験

5.1 実験の条件

前述したように、本手法が入力とするのは BVH 形式のモーションキャプチャファイルである。各ファイルの長さは平均およそ 90 フレーム（約 3 秒）で、それぞれ 6 人の被験者のうちの 1 人が対象動作のうちのある動作を意識的に行ったものであり、ファイル中にはその 1 種類の動作のみが含まれている。このような BVH ファイルを各対象動作ごとに用意した。新奇な運動としては、“EtcSittingOnFloor”には片膝を立てて座っているものや横座りなどの運動データを、“EtcSitting”にはしゃがんでいるデータを、“EtcLying”には寝返りを打っているデータを、“EtcStanding”にはその場で足踏みをしているものやキックをしているデータを、“Etc”には四つん這いになっているデータをそれぞれ用いた。

実験においては、合計で 1,201 個の BVH ファイルを三つのデータセットに分割した。分割は、各動作に属するデータが等分配されるという条件以外には何の制約も課さず、無造作に行った。この三つのデータセットのうちの一つを学習に、残り二つを認識実験に用いた。

動作モデルの学習の際には、前述した通り、階層構造を用いることで下位のモデルを単純化できる。例えば長座の認識の際には、床に座っていることは上の階層の認識結果を利用することで前提とできるため、腰の高さや上体の姿勢を特徴量に含める必要はなく、膝の高さや曲がりだけに注目すればよい。

また、各 BVH ファイルにはあらかじめ人が正解ラベルを付与しておき、認識実験においては認識結果が人による正解ラベルと一致するときを正解とした。正解ラベルは、あらかじめ対象動作を列挙しておき、各 BVH ファイルが各階層においてそのうちのどれに当てはまるか、あるいはどれにも当てはまらないかということを考えて付与したものである。

5.2 評価の指標

本認識手法では、誤認識の原因として「尤度の競合による間違い」、「新奇判別による間違い」の二つがある。一つ目の尤度の競合による間違いとは、尤度の値の逆転によってある動作が別の動作として認識されてしまうことを指す。これを「誤競合 (Incorrect Competition)」と呼ぶことにする。一方、二つ目の新奇判別による間違いとは、信頼度とその閾値との逆転によって新奇なものが既知の動作と認識されたり、逆に既知の動作が新奇なものとして認識されたりすることを指す。これを「誤判別 (Incorrect Detection)」と呼ぶことにする。

さらに本手法では階層構造を用いているため、上の階層での間違いが下の階層に影響を及ぼす。つまり、ある動作 act について「誤競合」であるとは正解ラベルに act を含む運動データに対し、 act が属する階層がそれより上の階層において競合による間違いが起きた結果 act と認識されなかったことを示す。同じように、 act について「誤判別」であるとは正解ラベルに act を含む運動データに対し、 act が属する階層がそれより上の階層において判別による間違いが起きた結果 act と認識されなかったことを示す。

以上を踏まえ、本実験では以下の四つの指標を用いて認識性能の評価を行った。

$$\text{誤競合率} = \frac{\text{act に属するデータで、誤競合が起きたデータ数}}{\text{act に属するデータの総数}} \times 100$$

$$\text{誤判別率} = \frac{\text{act に属するデータで、誤判別が起きたデータ数}}{\text{act に属するデータの総数}} \times 100$$

$$\text{再現率} = \frac{\text{act に属するデータで、正しく認識されたデータ数}}{\text{act に属するデータの総数}} \times 100$$

$$\text{適合率} = \frac{\text{act に属するデータで、正しく認識されたデータ数}}{\text{act と認識されたデータの総数}} \times 100$$

5.3 実験結果

認識性能の評価実験の結果は Table 2 のようになった。既知動作、すなわちシステムが動作モデルを持っている動作の BVH ファイルに対する認識は、合計で再現率が 95.1%、適合率が 97.7%となった。また、システムが動作モデルを持っていない新奇動作の BVH ファイルに対する認識は、合計で再現率が 84.5%であり、適合率が 75.5%であった。これから、新奇判別の枠組みがおおむねうまくいっていることが示された。最終的に、全体での再現率および適合率は 93.2%であり、これらの値から、例えば「Sitting - SittingOnFloor - Chouza」といった複数の抽象度による階層的な動作認識が高い精度で行われていることが示された。

全体での誤競合率と誤判別率を見ると、誤競合率は 1.0%であるのに対し誤判別率は 5.8%となっており、誤認識の主な原因が新奇判別における間違いであることが分かる。また、既知動作と新奇動作とを比べると再現率・適合率ともに既知動作の方が大きいことから、やはり「ある動作である」という正のモデルのみを用いて新奇な「どの動作でもない」データを判別する 1 クラス判別の難しさが分かる。

既知動作では特に、“Agura”の再現率が低い。これは“Agura”のデータが新奇と見なされることが多かったためである。原因としては、胡座のかき方は体の柔らかさなどの要因が絡んでく

Table 2 Experimental result of whole body actions

Action Name	IC Ratio	ID Ratio	Recall Ratio	Precision Ratio
Standing	0.0	0.5	99.5	100.0
Sitting	0.6	1.0	98.4	98.7
Lying	0.8	0.0	99.2	99.2
StandingStill	0.0	3.4	96.6	98.2
Walking	1.5	0.0	98.5	100.0
Running	0.0	3.0	97.0	98.5
SittingOnFloor	1.1	2.1	96.8	92.9
SittingOnChair	1.9	2.9	95.2	99.0
LyingOnSide	2.4	1.2	96.4	100.0
LyingOnBack	0.0	0.0	100.0	98.9
LyingOnFace	0.0	0.0	100.0	100.0
Agura	3.4	27.6	69.0	87.0
Seiza	5.6	11.1	83.3	88.2
Chouza	0.0	0.0	100.0	92.6
Ryoutatehiza	2.7	16.2	81.1	88.2
LyingOnRightSide	2.1	2.1	95.8	100.0
LyingOnLeftSide	2.8	0.0	97.2	100.0
Total of known Actions	1.2	3.7	95.1	97.7
EtcSittingOnFloor	0.0	11.1	88.9	74.2
EtcSitting	0.0	64.3	35.7	55.6
EtcLying	0.0	11.1	88.9	72.7
EtcStanding	0.0	4.0	96.0	88.9
Etc	0.0	15.4	84.6	73.3
Total of novel Actions	0.0	15.5	84.5	75.5
Total of all Actions	1.0	5.8	93.2	93.2

IC Ratio stands for Incorrect Competition Ratio.

ID Ratio stands for Incorrect Detection Ratio.

All values are presented in percentage (%)

るため個人差が大きく、今回の実験ではその多様性を十分に表現できるモデルが得られなかったことが挙げられる。この対策としては、学習データの見直しや、胡座の動作モデルが用いた特徴量の見直しなどが考えられる。また、計測上の問題として、筋肉の歪みが大きい姿勢をとると、センサがずれたり人間の剛体近似が成り立たなかったりということが生じ、誤差が大きくなってしまいうことがあある。胡座などの床に座る動作では床に足が接触して歪んでしまうため、これも認識率を下げる要因の一つになっていると思われる。

一方、新奇動作では“EtcSitting”が再現率・適合率ともに低い。これは“EtcSitting”の運動データの多くが“SittingOnFloor”と認識されてしまったためである。原因としては、“SittingOnFloor”の動作モデルが十分適切でなかったということが考えられる。本研究で“SittingOnFloor”の動作モデルが用いた特徴量は「地面からの腰の高さ」、「腰と足先の高さの差」であるが、それらだけでは床に座っているデータと、“EtcSitting”として用いたしゃがんでいるデータとの区別がつけられなかったと考えられる。床に座っているか、しゃがんでいるかは、一見「地面からの腰の高さ」で区別できそうであるが、正座をしている場合には腰の高さはしゃがんでいる場合とほぼ同じである。しゃがんでいる、とは「膝を折り曲げて腰を落とす」ことであることを考えると、膝に関する何らかの情報が“SittingOnFloor”の動作モデルに必要と思われる。

6. 結 論

本研究では、多様かつ頑強に日常動作を認識する手法の実現を目標に、(1) 複雑で多様な日常動作を的確にとらえるため、人の動作に関する知識を特徴量の抽出に利用、(2) 日常動作の曖昧さ・時系列性を取り扱うため、HMMを利用、(3) 多様で頑強な認識のため、動作の階層構造を木表現し認識に利用、という特徴を持つ日常動作認識手法を実現した。

具体的には、個々の動作のモデルを、人による動作の条件記述を利用した特徴量抽出フィルタと、各特徴量の振る舞いのモデル化である HMM との二つの要素を用いて構築した。また、抽象度の違いから生じる動作の概念の階層構造を各動作モデルをノードとする木構造で表現し利用した。木構造を利用した認識はルートから始まり、各階層において最も確からしい動作を選び、その結果が信頼できるか判別するというのを下の階層へと移りながら繰り返すことで行う。

階層構造を利用した認識を行うことで、(1) 一つの運動データに対し様々な抽象度での複数の認識結果が得られる、(2) 下位の動作のモデルを単純化できる、(3) 新奇データに対し認識の詳細さを下げることで対応できる、という利点が得られた。

評価実験では、全体で 93.2% という再現率・適合率であり、認識が高い精度で実現されていることを示した。

将来課題として、まずは頑強な新奇判別手法の構築や対象動作の拡張が挙げられる。また、本研究では動作の切り出しは人が行ったが、一連の動作を自動的にセグメントし認識するような枠組みが必要である。さらに、本手法では一つのルートから派生する木構造を用いたが、これをネットワーク表現に拡張すること、動作間の関係を確率的に表現することも望まれる。

参 考 文 献

- [1] A.F. Bobick and Y.A. Ivanov: “Action Recognition Using Probabilistic Parsing,” In Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.196-202, 1998.
- [2] T. Starner, J. Weaver and A. Pentland: “Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.1371-1375, December 1998.
- [3] H.H. Bui, S. Venkatesh and G. West: “Tracking and Surveillance in Wide-Area Spatial Environments Using the Abstract Hidden Markov Model,” International Journal of Pattern Recognition and Artificial Intelligence, vol.15, no.1, pp.177-195, 2001.
- [4] K. Ogawara, S. Iba, T. Tanuki, H. Kimura and K. Ikeuchi: “Recognition of Human Task by Attention Point Analysis,” In Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol.3, pp.2121-2126, November 2000.
- [5] N. Oliver, E. Horvitz and A. Garg: “Layered Representations for Human Activity Recognition,” In Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, pp.3-8, 2002.
- [6] J. Yamato, J. Ohya and K. Ishii: “Recognizing Human Action in Time-Sequential Images using Hidden Markov Model,” In Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.379-385, 1992.
- [7] T. Starner and A. Pentland: “Visual Recognition of American

- Sign Language using Hidden Markov Models," In Proceedings of the First IEEE International Workshop on Automatic Face and Gesture Recognition, pp.189-194, 1995.
- [8] M. Brand, N. Oliver and A. Pentland: "Coupled Hidden markov models for complex action recognition," In Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.994-999, 1997.
- [9] T. Inamura, Y. Nakamura, H. Ezaki and I. Toshima: "Imitation and Primitive Symbol Acquisition of Humanoids by the Integrated Mimesis Loop," In Proceedings of the 2001 IEEE International Conference on Robotics and Automation, pp.4208-4213, 2001.
- [10] T. Mori, K. Tsujioka, M. Shimosaka and T. Sato: "Human-like Action Recognition System Using Features Extracted by Human," In Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1214-1220, 2002.
- [11] L. Rabiner: "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol.77, no.2, pp.257-285, 1989.
- [12] A.P. Dempster, N.M. Laird and D.B. Rubin: "Maximum likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society B(Methodological), vol.39, no.1, pp.1-38, 1977.
- [13] D.M.J. Tax and R.P.W. Duin: "Outlier Detection Using Classifier Instability," In Proceedings of the Joint IAPR International Workshops on Structural and Syntactic Pattern Recognition (SSPR 1998) and Statistical Techniques in Pattern Recognition (SPR 1998), pp.593-601, 1998.
- [14] 新エネルギー・産業技術総合開発機構 (NEDO) 人間生活工学研究センター: 身体機能データ・ベースの構築に関する調査研究, 平成 10 年度, 1998.
- [15] V.N. Vapnik: The Nature of Statistical Learning Theory (Statistics for Engineering and Information Science). Springer Verlag, 1995.

森 武俊 (Taketoshi Mori)

1967 年 9 月 8 日生。1990 年東京大学工学部機械工学科卒業。1995 年東京大学大学院工学系研究科博士課程修了 (情報工学専攻)。同年東京大学先端科学技術研究センター助手。1998 年同講師。2001 年米国 MIT 客員研究員。現在東京大学大学院情報学環助教授。動作認識, 行動理解, パーベイシブコンピューティング, ヒューマンコンテンツの研究をしている。博士 (工学)。IEEE, ACM, 電子情報通信学会, 人工知能学会, 計測自動制御学会等の会員。(日本ロボット学会正会員)

下坂正倫 (Masamichi Shimosaka)

1978 年 6 月 5 日生。2001 年東京大学工学部機械情報工学科卒業。2003 年東京大学大学院情報理工学系研究科知能機械情報学専攻修士課程修了, 同博士課程進学。2004 年より独立行政法人日本学術振興会特別研究員。人間の運動データのモデル化, 動作認識とその学習の研究に従事。統計的機械学習に基づくロボットの模倣学習, コンピュータビジョン, 自然言語処理などに興味を持つ。IEEE, 電子情報通信学会, 人工知能学会の会員。(日本ロボット学会学生会員)

瀬川友史 (Yushi Segawa)

1981 年 8 月 28 日生。2004 年 3 月東京大学工学部機械情報工学科卒業。同年 4 月同大学情報理工学系研究科知能機械情報学専攻修士課程進学。現在に至る。確率統計的手法による人間の動作の分節化・認識・解析などの研究を行っている。

(日本ロボット学会学生会員)

佐藤知正 (Tomomasa Sato)

1948 年 9 月 22 日生。1971 年東京大学産業機械工学科卒業。1976 年同大学博士課程修了後, 電子技術総合研究所入所。1991 年東京大学先端科学技術研究センター教授に移籍。1998 年東京大学工学系研究科機械情報工学専攻に移籍。2000 年同大学情報理工学系研究科知能機械情報学専攻に改組。人間共棲ロボットの研究に従事。日本機械学会, IEEE などの会員。

(日本ロボット学会正会員)