動作の時間的連続性を考慮した Boosting アルゴリズムに基づくオンライン日常動作認識

下 坂 正 倫* 柊次金 佑* 森 武 俊* 佐 藤 知 正*

Online Action Recognition Based on Boosted Sequential Classification

Masamichi Shimosaka*, Yu Nejigane*, Taketoshi Mori* and Tomomasa Sato*

In this paper, we propose a robust online action recognition method based on boosted sequential classification. Our method utilizes boosting algorithm that is one of ensemble learning algorithms. This algorithm is also known as a feature selector and has been utilized in the fields of image processing and natural language processing in recent years. Based on the boosting scheme, our method can automatically and efficiently select significant features for action recognition. Additionally, the method leverages temporal dependency of actions based on Ising model to improve recognition performance. We evaluated our method to action recognition, such as walking and running, using motion capture data only with posture features. In the result, our method can classify the actions more robustly than the method that does not utilize temporal dependency of actions.

Key Words: Boosting, AdaBoost, Online Action Recognition, Motion Capture, Ising Model

1. 緒 論

動作認識は将来、ロボットによる人間への支援や、ロボットと人とのスムーズなコミュニケーション、不審者に対するサーベイランスシステムなどを実現するうえで重要な技術である。動作認識に関する研究としては、Yamato ら [1] や Starner ら [2]、Cao ら [3] をはじめ、hidden Markov model (HMM) [4] や support vector machine (SVM) [5] といった統計学習法を導入し認識性能の頑健化を狙う研究 [6]~[8] が近年数多く行われてきた。

上記のような従来の動作認識手法では、入力される運動情報のうち、どのような特徴量を認識処理に利用するかは、人があらかじめ決定し与えることが多い。このような方法は、直感的かつ、認識精度の向上に大きく役立てられることが多い一方で、どのような特徴量をどれだけ利用すればよいのかについての判断が難しい。用いる特徴量が不十分であれば、高い認識精度を得られないおそれがあり、逆に過剰であれば、構築される認識器が不要に複雑になり、認識処理コストが高くなるという問題がある。これはオンラインでの動作認識を実現するうえでは大きな障害となる。また、利用する特徴量を人が丁寧かつ慎重に決定したとしても、認識性能の向上において、それが十分適切である保証は得られないし、認識対象とする動作種が多ければそれだけ、特徴量の取舍選択に関わる人への負担は大きくなる。

以上のような問題を受け筆者らは、特徴量選択の機能を有する Boosting アルゴリズム [9] を利用し、特徴量選択における人への負担や、認識処理における計算コストといった観点において実用性に主眼を置いた、オンラインでの動作認識手法 [10] を構築してきた。この手法により、認識に利用される特徴量は学習の過程で自動的に選出され、特徴量選択における人への負担を大きく低減させたうえで、認識処理コストの非常に低い頑健な認識が可能となった。

しかし、上記の手法では、時々刻々に行われる認識処理はそれぞれの時刻において独立に実行され、日常動作の持つ重要な性質の一つである、動作の時間的連続性を考慮していない。ここで言う動作の時間的連続性とは、ある時刻、ある動作が起こっているとき、続く時刻でも同一の動作が起きている可能性が高い、ということの意味する。最もシンプルな対処法としては、Boosting とその出力結果の時系列に対するローパスフィルタを併用することが挙げられるが、現時刻の認識結果が以前の時刻の結果にどの程度強く影響を受けるかは動作種によって異なるはずであり、例えば、「歩いている」動作であれば、頑健な認識のためには以前の時刻に比較的強い影響を受けるべきであるが、逆に「ジャンプする」のような突発的で生起時間の短い動作の場合、受ける影響は低くすべきである。適切なフィルタを実現するためのフィルタ中の遅延時定数の設計は慎重を期す必要がある。一方、本稿では、動作の時間的連続性をモデル化し、従来の Boosting アルゴリズムに基づく認識手法を発展させることで、学習によってその影響の強弱を自動的に獲得することが可能な、より頑健で柔軟な認識手法の構築を狙う。

原稿受付 2007 年 1 月 15 日

*東京大学大学院情報理工学研究所

*Graduate School of Information Science and Technology of the University of Tokyo

2. Boosting アルゴリズムと動作の時間的連続性を考慮した認識方法

本章では、動作の時間的連続性を考慮したうえで、Boosting アルゴリズムをどのように動作認識に適用するかについて説明する。

2.1 提案手法における認識処理の概要

本稿において構築する動作認識手法では、各認識対象動作について、その動作が起こっているか否かの2値分類器が並列的に構成されたものとする (Fig. 1)。つまり、入力を時系列の運動データとした場合、得られる出力は2値の成分を対象動作の数だけ持つ時系列でのベクトルデータとなる。こうすることで、同じ時刻に独立して共起しうる複数の動作に対しても認識処理を行うことができる。おのおの2値分類器を、次節で述べる Boosting アルゴリズムによって実現する。

2.2 Boosting アルゴリズム

Boosting アルゴリズムとは、集団学習アルゴリズム (ensemble learning) [11] の一つであり、学習時、訓練データの重みを変えさせながら逐次的に単純な認識器を多数構築し、それらの重み付き多数決によって精度の高い認識を実現するものである (以降、この「単純な認識器」を要素認識器と呼ぶこととする)。要素認識器そのものは単純であるため計算コストが低く、したがってそれらを必要最低限な数だけ逐次的に構築することで、計算コストを非常に低く抑えたうえで頑健な認識器を実現することができる。近年、自然言語処理における文書分類 [12] や画像処理における歩行者追跡 [13] といったタスクに応用され効果をあげている。

以下では、まず、Boosting アルゴリズムの中でも最も代表的な AdaBoost の学習時での一般的な流れを述べるため、時間的連続性を考慮せずに、毎時刻運動情報のみから動作の2値分類を行う場合を考える。N 個の時系列の訓練データとして $(\mathbf{x}_1^{(n)}, y_1^{(n)}), \dots, (\mathbf{x}_{T_n}^{(n)}, y_{T_n}^{(n)})$ が与えられたとする。ここで、 $n = 1, \dots, N$ であり、 T_n は n 番目の時系列訓練データの長さ、具体的にはフレーム数を指す。いずれも一定の計測周期 (30 [Hz]) で計測したものとす。 i 番目の訓練データ (モーションデータ) の長さが $T_i = 3000$ であった場合、 i 番目の訓練データの長さは 100 秒を意味する。また、 $\mathbf{x}_t^{(n)} \in \mathbf{X}, y_t^{(n)} \in Y = \{-1, +1\}$ であり、本稿においては $\mathbf{X}$ は、身体各関節の姿勢情報、 Y はある動作が起きているか否かを示すラベルとする。例えば、 $y_t = +1$ であれば、時刻 t においてある動作が起きているとい

うことを示す。

まず k 番目に構築される要素認識器 ($h_k: \mathbf{X} \rightarrow Y$) を以下の誤り率を最小化し獲得する。

$$\epsilon_k = \sum_{n, t: h_k(\mathbf{x}_t^{(n)}) \neq y_t^{(n)}} D_k(n, t)$$

ここで $D_k(n, t)$ は n 番目の時系列における時刻 t の訓練データの重みを表す。次に、得られた h_k の信頼度 α_k を以下のように算出する。

$$\alpha_k = \frac{1}{2} \ln \left(\frac{1 - \epsilon_k}{\epsilon_k} \right)$$

そして、信頼度 α_k から次のように訓練データの重みを更新する。

$$D_{k+1}(n, t) = \frac{\exp(-y_t^{(n)} \sum_{i=1}^k \alpha_i h_i(\mathbf{x}_t^{(n)}))}{Z_k} \quad (1)$$

Z_k は正規化項を意味する。以上を繰り返すことで得られた複数の要素認識器により、最終的な認識器は

$$H(\mathbf{x}_t) = \text{sgn} \left(\sum_k \alpha_k h_k(\mathbf{x}_t) \right) \quad (2)$$

のようになる。各 h_k がそれぞれある単一の特徴量に着目して認識を行うよう設計を行うことで、Boosting による特徴量選択を実現することができる。動作認識において、 h_k をどのように設計するかについては 2.5 節で説明する。

なお、以上の学習手順をそのまま利用しただけでは動作の時間的連続性を考慮することはできない。これに次節で説明する修正を加えるものが本稿で提案する学習法である。

2.3 動作の時間的連続性の考慮

動作の時間的な連続性を考慮する方法の一つとして、現時刻での出力が、その近傍の時刻の出力に影響を受けるようなモデルを利用してやる方法が挙げられる。そのような出力を可能とするものとしては、HMM や Conditional Random Fields (CRF) [14] など、グラフィカルモデルに基づく様々な手法が提案されている。例えば、Sminchisescu ら [15] は CRF を時系列のシルエット画像に適用し、「歩いている」などの動作の認識を実現している。しかし CRF においては要素認識器の数は固定であり、この数を十分に大きくすることで高精度な認識を可能にするが、Boosting にあるような特徴量選択の機能を持たず、コンパクトな認識器を構成することは難しい。Boosting の枠組みにおいても、同様の出力を実現する手法 [16] はすでに提案されているが、主に自然言語処理や DNA 配列の解析等に利用するのに適した形となっており、動作認識のように連続値を対象とする分類問題には適用が困難である。

そこで本稿では、近傍の時刻と現時刻との出力の関係性を Ising モデルと同様の考え [17] [18] を用いて表現し、Boosting により得る要素認識器の出力に加味することで、動作の時間的連続性を考慮した認識を実現する。

時間的連続性を考慮する場合、認識処理は式 (2) には従わず、確率的な方法によって処理を行う。長さ T の時系列運動情報 $\mathbf{x}_{1:T} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ が与えられたとき、対応する時系列ラベル $\hat{\mathbf{y}}_{1:T} = \hat{y}_1, \hat{y}_2, \dots, \hat{y}_T$ を推定することは以下の問題を解

Fig. 1 Outline of the proposed method

Fig. 2 Graphical model of the proposed method

くことに等しい。

$$\hat{\mathbf{y}}_{1:T} = \underset{\mathbf{y}_{1:T}}{\operatorname{argmax}} P(\mathbf{y}_{1:T} | \mathbf{x}_{1:T})$$

ここで、時間的連続性を踏まえ、Fig. 2 に示すように、各時刻の動作ラベルの出力が、一つ前の時刻のラベルとその時刻での観測情報に影響を受けるグラフィカルモデルを考え、以下のようにな式を展開する。

$$\hat{\mathbf{y}}_{1:T} = \underset{\mathbf{y}_{1:T}}{\operatorname{argmax}} P(y_1 | \mathbf{x}_1) P(y_2 | y_1, \mathbf{x}_2) \cdots P(y_T | y_{T-1}, \mathbf{x}_T)$$

さらに、運動情報 $\mathbf{x}_t$ と過去の時刻の出力 y_{t-1} が与えられたとき、 y_t を出力する確率を、

$$P(y_t | y_{t-1}, \mathbf{x}_t) = \frac{\exp\{y_t(\sum_k \alpha_k h_k(\mathbf{x}_t) + \beta y_{t-1})\}}{Z_t}$$

$$Z_t = \sum_{y_t=+1, -1} \exp\left\{y_t \left(\sum_k \alpha_k h_k(\mathbf{x}_t) + \beta y_{t-1}\right)\right\}$$

として、この確率式をもとに Viterbi 復号法によって動作ラベル列 $\hat{\mathbf{y}}_{1:T}$ の推定を行う。ただしオンラインでの認識のため、ラベル推定時は毎時刻、その時点で最も尤度の高い動作ラベルをその時刻のラベルとして時々刻々と出力する。 β が時間的連続性の強弱を決めるパラメータであり、訓練データから自動的に最適化されるため、認識対象動作の増加にも、人への負担を増やすことなく対応できる。

2.4 提案手法における認識器の学習方法

パラメータ β の最適化を Boosting アルゴリズム中で同時に実行していくことは非常に難しい。そこで提案手法では、Boosting による要素認識器構築と、パラメータ β の最適化を交互に繰り返す行い、最終的な認識器の構築を実現する。つまり、

- 現時点での訓練データの重みをもとに要素認識器の候補群を最適化、その中から最も成績の良いものを選択し、その信頼度を固定する。
- 現時点で構築済みの要素認識器群をもとに時間的連続性に関するパラメータ β を、以下に述べる規準に従って調整し、その影響を考慮したうえで訓練データの重みを更新。

この2段階の最適化処理を交互に行うサイクル (Fig. 3) を繰り返すことにより、時間的連続性を考慮した認識器の構築を行う。このアルゴリズムは検定用のデータセットに対して識別性能がある一定のレベルに達するまでこれを繰り返すものである。なお、 β は学習の初期段階では 0 として設定する。

Fig. 3 におけるおのおの処理内容について具体的に説明する。まず訓練データの重み更新式である式 (1) を以下のように

Fig. 3 Process of training classifier in the proposed method

改める。

$$D_{k+1}(n, t) = \frac{\exp\{-y_t^{(n)}(\sum_{i=1}^k \alpha_i h_i(\mathbf{x}_t^{(n)}) + \beta y_{t-1}^{(n)})\}}{Z_k}$$

この式を通じて更新される重みをもとに、通常の Boosting と同様の手順に従い要素認識器を構築すればよい。このように隣接する時刻の出力を訓練データの重みの更新に影響させることで、動作の時間的連続性を考慮するように要素認識器が逐次最適化・選出されることが期待できる (訓練データは時系列順に並んでいることを前提とする)。

また、要素認識器の逐次構築に伴う、パラメータ β の最適化は、擬似尤度 [17] [18] と事後確率最大化法 (MAP 推定) を利用し、下式の対数尤度を最大化することで行われる。

$$L(\beta) = \frac{1}{N} \sum_{n,t} \ln P(y_t^{(n)} | y_{t-1}^{(n)}, \mathbf{x}_t^{(n)}) - \frac{C}{2} |\beta|^2 \quad (3)$$

右辺第 2 項はパラメータ β に関する事前分布を考慮した項であり、Gaussian Prior (L2-norm) [19] に相当する。この項は β が訓練データに過適合するのを避ける目的で導入する。 C はハイパーパラメータであり、この値が小さすぎると過適合を防ぐことができず、大きすぎると学習がうまく行われなくなる。本稿ではこの値を経験的に決定した。また、 N は訓練データの総フレーム数である。この対数尤度の最大化は、勾配法に基づく最適化計算法により行う。なお、要素認識器を固定した下での尤度の最大化では、その勾配量のノルムが一定量以下になれば調整を終了するものとした。

2.5 要素認識器の設計

本稿で、Boosting アルゴリズムにおいて利用する要素認識器 h_k は、入力運動情報 $\mathbf{x}_t$ から抽出される、あるスカラ値について閾値判別を行う深さ 1 の決定木とする。このような認識器は Decision Stumps [20] と呼ばれ、式で表すと以下ようになる。

$$h_k(\mathbf{x}_t) = \operatorname{sgn}(\phi_k(\mathbf{x}_t) - \gamma_k) \quad (4)$$

ϕ_k は運動情報 $\mathbf{x}_t$ からあるスカラ値を抽出する関数である。 γ_k はそのスカラ値に対する閾値であり、要素認識器の構築の際に、訓練データの重み D_k に基づき最適化する。 ϕ_k によって抽出されるスカラ値は、ある身体部位におけるある一次元の運動特徴量とする。具体的にどのような運動特徴を抽出するかについては次章で説明する。

3. 評価実験

本章では、提案する認識手法に対して行った評価実験について述べる (なお、実験における式 (3) のハイパーパラメータは $C = 2.0 \times 10^{-3}$ とした)。

3.1 認識対象動作

日常生活においては、「寝ている」などの動きを伴わない動作と「歩いている」などの動きを伴う動作の2種類に大別される。動きを伴わない動作には、その動作を表現する一定のポーズが存在し、逆に動きを伴う動作では、ポーズは時刻と共に変化し常に一定とはならない。時間的連続性を考慮することの有効性を評価するためには、これら2種類のうち動きを伴う動作を対象とすべきであると考え、よって本実験では、分類対象動作を動きを伴う動作である「歩いている」と「走っている」とした。

3.2 実験に用いる運動データ

実験で用いた訓練用・評価用運動データについて説明する。運動データはサンプリングレート 30 [Hz] の磁気式モーションキャプチャシステムにより取得され BVH 形式で記録されたもので、Fig. 4 に示す 36 自由度を持つ人体骨格モデルにより表現される。BVH 形式とは一般的な CG ソフトウェアで読み込み可能なモーションキャプチャデータフォーマットである。

この運動データに含まれる動作は「歩いている」と「走っている」と「立ち止まっている」およびその動作間の遷移動作である。実験では、約 3,000 フレームの運動データを含むデータセットを3セット用意し、2セットを学習用、1セットを評価用とする3分割交差検定により行った。用意する運動データにはあらかじめ人間が正解となる動作名ラベルを付与しておく。

3.3 動作要素認識器の候補

本実験では日常生活における脚動作である「歩いている」と「走っている」を対象とするため、下半身の部位にのみ着目し要素認識器を構築する。実用の際の人体への拘束性を抑える目的から、着目する身体部位の数をできる限り少なくしたいと考える。モーションキャプチャシステムより姿勢情報が取得できる部位は Fig. 4 に示す計 11 箇所であるが、脚動作を対象とした本実験において、提案手法の有効性を適切に評価するため、着目する身体部位は両腿部和両脛部の4箇所とした (Fig. 4 の RightUpLeg, RightLowLeg, LeftUpLeg, LeftLowLeg に該当)。

また、将来的に、3軸地磁気センサや3軸加速度センサなどを利用して、身体部位の姿勢状態をワイヤレスに取得することを想定し、式 (4) において ϕ_k により抽出する各身体部位の運動特徴としては、

- 外界に固定した座標系から見た、ある身体部位の姿勢行列における鉛直成分に当たる値
- ある身体部位以外の着目部位の座標系から見た、その部位の姿勢行列における各成分の値
- 以上のそれぞれの値の時刻ごとの変化量

Fig. 4 Human model utilized in the experiment

といった、姿勢情報に関するもの (ベクトルで表現すれば 240 次元のデータ) に限った。

本稿の Boosting アルゴリズムでは、両腿部和両脛部の4箇所について上記三つの性質の運動特徴の中から選ばれることになる。

3.4 評価指標

認識精度の評価指標には F 値を用いた。F 値とは、Recall 値 (再現率) と Precision 値 (適合率) の調和平均をとったものである。正解データ中で $y_t = +1$ であるフレーム数を N_a 、提案手法が $y_t = +1$ と出力したフレーム数を N_m 、提案手法が正しく $y_t = +1$ と出力したフレーム数を N_c とし、Recall 値を R 、Precision 値を P とすると、 R, P, F はそれぞれ以下のよう表される。

$$R = \frac{N_c}{N_a}, \quad P = \frac{N_c}{N_m}, \quad F = \frac{2RP}{R+P}$$

3.5 実験結果

2 種の対象動作について、提案手法を適用して構築する要素認識器の数を 60, 80, 100 と変えた場合の認識精度の F 値 (%) を Table 1 に示す。また比較として、動作の時間的連続性を考慮しない場合 (式 (1), (2) によって訓練データの重み更新、および認識を行う場合) の F 値も記載する。表中の括弧内の数字が要素認識器の数を表す。

使用した運動特徴量は姿勢情報に関するものに限ったものの、両動作について頑健な認識が達成できている。また、Boosting アルゴリズムをそのまま適用する従来の方法と比較して、提案手法では両対象動作において認識精度が向上していることも分かる。「走っている」に関しては F 値の上昇が比較的小さいが、これは、要素認識器を十分多く構築すれば「走っている」動作は運動情報のみに基づく場合でも比較的高精度な認識が可能のため、時間的連続性の考慮による寄与が小さかったのではないかと考えられる。

学習結果を確認してみたところ、構築する要素認識器の数を 100 とした場合、最適化された β の平均値は「歩いている」が 1.81, 「走っている」が 0.98 であり、それに対し、要素認識器群の出力の和 $\sum_k \alpha_k h_k(x_t)$ は「歩いている」の場合、およそ ± 10.0 の範囲に分布し、「走っている」の場合、およそ ± 15.0 の範囲に分布していた。つまり、「歩いている」のほうが時間的連続性の影響を比較強く受けるように学習が行われたということである。

自明ながら、時間的連続性の影響を過度に強くすれば、続く時刻の出力が悪影響を受けるおそれもある。上記のように、瞬間瞬間の運動情報に基づく出力の信頼性を考慮して、時間的連

Table 1 Experimental results for target actions

Target Action	AdaBoost Only	Proposed Method
Walking(60)	92.4	94.9
Walking(80)	93.2	95.2
Walking(100)	93.8	95.6
Running(60)	95.0	96.3
Running(80)	95.0	96.2
Running(100)	95.4	96.6

Fig. 5 Number of elemental classifiers vs. F-measure for test data in walking classification

Fig. 6 Number of elemental classifiers vs. F-measure for test data in running classification

Fig. 7 Thumbnails of the motion data example

Fig. 8 Recognition result for the motion data example

続性の強弱をうまく設定すべきである。要素認識器の逐次構築を行いつつ、この強弱を自動調整できる点が提案手法の持つ利点の一つであり、これはローパスフィルタのような手法を単純に適用しただけでは実現できない点である。

続いて、構築する要素認識器の数を変えながら評価用データに対し適用した場合について、F 値の変化の仕方の一例を、「歩いている」については Fig. 5 に、「走っている」については Fig. 6 に示す。提案手法による場合を実線で、動作の時間的連続性を考慮しない場合を破線で表示している。

両図から分かるように、両動作において、提案手法を用いた場合の方が F 値が高く、要素認識器が少ない段階から高い認識性能を達成することが確認できる。以上のことから、提案手法を利用することで、従来手法よりも少ない要素認識器で頑健な動作認識を実現することが可能であるということが分かる。つまり、提案手法によって、よりコンパクトな認識モデルでオンライン動作認識が実現できるということである。

3.6 提案手法の適用例

提案する動作認識手法を適用した場合の出力例を示す。例として適用した運動データを 3DCG により視覚化したサムネイルを Fig. 7 に、その運動データに対する認識結果を Fig. 8 に示す。比較のため、提案手法による出力結果と AdaBoost のみによる手法の出力結果の両方を示してある（それぞれの動作に対し正解となる区間を矢印で示す）。なお、この例では両手法、両動作とも構築した要素認識器の数は 100 とした。Fig. 7 に示した運動データは、立って静止している状態から走り出し、やがて歩いている状態に移る、というものであるが、Fig. 8 から確認できるように、より安定した頑健な認識が提案手法により実現できている。

4. 結 論

本稿では、動作の持つ時間的連続性を考慮した、Boosting アルゴリズムに基づくオンライン動作認識手法を提案した。通常の Boosting アルゴリズムに加え、Ising モデルと同様の考えに基づいた、隣接する過去の出力を考慮した項を組み込むことで、動作の時間的連続性をモデル化した。時間的連続性の強弱を示すパラメータは動作ごとに訓練データから決定されるため、認識対象動作の増加にも対応が容易である。モーションキャプチャデータを利用し、「歩いている」「走っている」といった脚動作を対象に評価実験を行った結果、提案手法は従来手法と比較して、より頑健な認識が可能となることを確認した。また、現実のセンサでの取得の容易さを考慮し、利用する運動特徴を姿勢情報に関するものに限った場合でも、頑健な認識器を構築できることが分かった。

将来課題としては、多数の動作に柔軟に対応するための多クラス分類への拡張や、観測運動情報の欠落やノイズに強いより安定した認識処理が可能な手法の構築が挙げられる。また、近年手軽に手に入りやすくなっている小型の姿勢センサを利用し、ワイヤレスかつウェアラブルなセンサを製作し本手法を適用することも、将来的な展望の一つである。

参 考 文 献

- [1] J. Yamato, J. Ohya and K. Ishii: "Recognizing Human Action in Time-sequential Images Using Hidden Markov Model," Proceedings of the 1992 IEEE Conference on Computer Vision and Pattern Recognition, pp.379-385, 1992.
- [2] T. Starner and A. Pentland: "Real-time American Sign Language Recognition from Video Using Hidden Markov Models," Proceedings of the 1995 IEEE International Conference on Computer Vision, pp.265-270, 1995.
- [3] D. Cao, O. Masoud, D. Boley and N. Papanikolopoulos: "Online Motion Classification Using Support Vector Machine," Proceedings of the 2004 IEEE International Conference on Robotics and Automation, vol.3, pp.2291-2296, 2004.
- [4] L. Rabiner: "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol.77, pp.257-285, 1989.
- [5] B. Schölkopf and A. Smola: Learning with Kernels. MIT Press, 2002.
- [6] J. Davis and A. Bobick: "The Representation and Recognition of Human Movement Using Temporal Templates," Proceedings of the 1997 IEEE International Conference on Computer Vision and Pattern Recognition, pp.928-934, 1997.
- [7] M. Shimosaka, T. Mori, T. Harada and T. Sato: "Recognition of Actions in Daily Life and Its Performance Adjustment Based on Support Vector Learning," International Journal of Humanoid Robotics, vol.1, no.4, pp.565-583, 2004.
- [8] T. Inamura, Y. Nakamura, H. Ezaki and I. Toshima: "Imitation and Primitives Symbol Acquisition of Humanoids by Integrated Mimesis Loop," Proceedings of the 2001 IEEE International Conference on Robotics and Automation, pp.4208-4213, 2001.
- [9] R. Schapire and Y. Singer: "Improved Boosting Using Confidence-rated Predictions," Machine Learning, vol.37, no.3, pp.297-336, 1999.
- [10] M. Shimosaka, T. Nishimura, Y. Nejigane, T. Mori and T. Sato: "Fast Online Action Recognition with Boosted Combinational Motion Features," Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.5851-5858, 2006.
- [11] E. Bauer and R. Kohavi: "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants," Machine Learning, vol.36, no.1-2, pp.105-139, 1999.
- [12] T. Kudo and Y. Matsumoto: "A Boosting Algorithm for Classification of Semi-structured Text," Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp.301-308, 2004.
- [13] P. Viola, M. Jones and D. Snow: "Detecting Pedestrians Using Patterns of Motion and Appearance," Proceedings of the 9th IEEE International Conference on Computer Vision, vol.2, pp.734-741, 2003.
- [14] J. Lafferty, A. McCallum and F. Pereira: "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," Proceedings of the Eighteenth International Conference on Machine Learning, pp.282-289, 2001.
- [15] C. Sminchisescu, A. Kanaujia, Z. Li and D. Metaxas: "Conditional Models for Contextual Human Motion Recognition," Proceedings of the Tenth IEEE International Conference on Computer Vision, vol.2, pp.1808-1815, 2005.
- [16] Y. Altun, T. Hofmann and M. Johnson: "Discriminative Learning for Label Sequences via Boosting," Advances in Neural Information Processing Systems, 2003.
- [17] S. Li: Markov Random Field Modeling in Image Analysis. Springer-Verlag, 2001.
- [18] S. Kumar and M. Hebert: "Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification," Proceedings of the 2003 IEEE International Confer-

ence on Computer Vision and Pattern Recognition, pp.1150–1157, 2003.

- [19] S. Chen and R. Rosenfeld: “A Gaussian Prior for Smoothing Maximum Entropy Models,” Technical Report CMUCS-99-108,

Carnegie Mellon University, 1999.

- [20] W. Iba and P. Langley: “Induction of One-Level Decision Trees,” Proceedings of the 9th International Conference on Machine Learning, pp.233–240, 1992.

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。2006年同大学大学院博士課程修了。現在東京大学大学院情報理工学系研究科助教。2004年より2006年まで独立行政法人日本学術振興会特別研究員。機械学習、コンピュータビジョンなどの研究を行っている。博士（情報理工学）。IEEE、電子情報通信学会、人工知能学会等の会員。（日本ロボット学会正会員）

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理工学系研究科准教授。動作認識、行動理解、パースペクティ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）

祢次金佑 (Yu Nejigane)

1981年9月11日生。2005年3月東京大学工学部機械情報工学科卒業。同年4月より東京大学大学院情報理工学系研究科知能機械情報学修士課程。在学中、日常動作のオンライン認識に関する研究を行った。2007年に修士課程を修了後、一般企業に就職し現在に至る。

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）