# Boosted Randomized Trees による人物検出と行動の同時認識

山下　隆義[†,††]　　山内　悠嗣[††]　　藤吉　弘亘[††]

† オムロン株式会社　滋賀県草津市西草津 2 丁目 2-1
†† 中部大学 愛知県春日井市松本町 1200

**あらまし**　人体の行動を理解する上で，人体検出，姿勢推定および行動認識は，行動理解の重要な要素技術である．しかしながら，これまでの多くの研究ではこれらを独立した分野としてとらえており，1 つの枠組みで人体検出から行動認識までを行っていない．そこで，我々は人体を検出すると同時に姿勢の推定および行動の認識を 1 つのフレームワークで実現可能な Boosted Randomized Trees を提案する．提案手法は，Randomized Trees のノード生成時に，クラス尤度をもとにした階層的な認識クラスの定義方法および，Joint Boosting による事前特徴選択を導入することで，認識に有効な特徴を効率的に選択することが可能である．これにより，上位ノードは人体検出，下位ノードは姿勢推定または行動認識を階層的に行うことができる．提案手法は，従来の Randomized Trees を用いた場合よりも人体検出率および行動認識率を向上させることができた．

**キーワード**　Randomized Trees，Joint Boosting，人体検出，行動認識

# Simultaneous Human Detection and Action Recognition based on Boosted Randomized Trees

Takayoshi YAMASHITA[†,††], Yuji YAMAUCHI[††], and Hironobu FUJIYOSHI[††]

† OMRON Corporation 2-2-1, Nishikusatsu, Kusatsu-city, SHIGA, JAPAN
†† Chubu University 1200, Matsumoto-cho, Kasugai-city, AICHI, JAPAN

**Abstract**　To understanding human behaviors, human detection, pose estimation and activity recognition are important technologies. However, most of the researches are deal with these topics independently. There has been few works done from human detection to human activity recognition in one framework. In this paper, we have proposed “Boosted Randomized Trees” that can detect human while simultaneously estimate poses and recognizing activities, in one framework. The proposed algorithm can select efficient features for multi class recognition by recognition class set definition using class likelihood and preliminarily feature selection with Joint Boosting in node generation of Randomized Trees. In this approach, it detects human in upper nodes of decision tree and estimates pose and recognizes activity in lower nodes of one. The performance of proposed method is a better detection and recognition rate than the conventional Randomized Trees.

**Key words**　Randomized Trees，Joint Boosting，Human detection，Activity recognition

## 1. はじめに

セキュリティやロボット、ユーザインタフェースなどの分野において，人の行動を理解することは重要な技術である．人の行動を理解することにより，セキュリティにおいては，不審人物の行動を検知することが可能となる．また，ロボットやユーザインタフェースにおいては，より様々なインタラクションを実現することにつながる．

人の行動を理解するための重要な要素技術として，人体検出手法，姿勢推定手法，行動認識手法がある．図 1 のように，1 枚の画像のみ与えられた場合，(a) および (b) は，人が歩いている，または走っていることが認識できる．一方，図 1(c) は人がかがみこんでいることが 1 枚の画像から認識することができるが，しゃがもうとしているのか，起き上がろうとしているのかまでは認識することができない．しかしながら，図 2 のように，3 枚の動画像の場合，起き上がろうとしていることが認識することができる．行動認識の過程において，人は人体の検出から行動の認識を静止画または動画像から得られる情報をもとに，より人の詳細な変化をとらえていると考えられる．すなわち，図 3(b) に示すように，人は人体の検出を行い，検出した人体の姿勢を推定，行動の認識を階層的に行っている．

そこで，我々は，図 3(b) に示すような，人体検出か

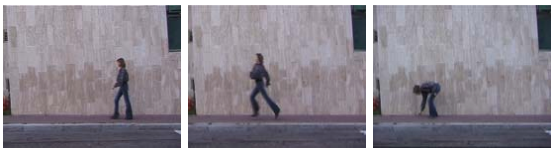


図 1　静止画における人の行動
(a) 歩いている (b) 走っている (c) しゃがんでいる

図 2　動画像における人の行動

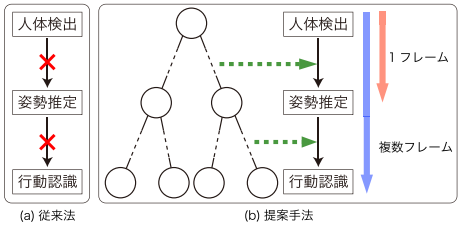


図 3　提案手法と従来法の行動認識フレームワーク

ら行動認識までを 1 つのフレームワークで実現する．本フレームワークを用いることで，静止画を与えた場合は，姿勢推定および単純な行動が認識でき，動画像を与えた場合は，より複雑な行動を認識することができる．我々は，このようなフレームワークを実現するための階層的な構造として，Boosted Randomized Trees を提案する．提案手法は，複数の姿勢の人体を検出するために，マルチクラスの識別手法として注目されている Randomized Trees [1] をベースとしている．Randomized Trees は，各決定木の生成時にランダム選択を用いているため，有効な特徴が選択されない場合が生じる．提案手法は，クラス尤度をもとにした認識クラスの定義方法および Joint Boosting [16] による事前特徴選択を導入することで，効率的な特徴選択を実現する．これにより，従来の Randomized Trees の課題を解決することができる．

本稿では，2 章で関連研究，3 章で提案手法について述べる．4 章で様々な姿勢における人体検出および姿勢推定，行動認識の各性能を Randomized Trees を用いた場合と比較する．また，5 章において，まとめを行う．

## 2. 関 連 手 法

人の行動を理解するための人体検出と行動認識は，ほとんどの場合が独立して研究されている．人体検出は，Dalal らが勾配方向ヒストグラムを利用した HOG 特徴量と SVM による手法 [5]，Zhu らは HOG 特徴量と AdaBoost による高速な手法 [6] を提案している．Hou らは HOG 特徴量を簡略化した EHOG と木構造の Vector Boosting による手法 [7] により，人体の検出と直立した人体の向きを推定している．また，Gall ら [18] は局所特徴量と Randomized Trees を用いて局所特徴と重心の位置関係を抽出し，Hough 変換により人の検出する手法を提案している．これらの手法は，直立した人体を前提としており，直立姿勢以外の姿勢を検出することができない．

姿勢推定は，シルエット画像のサンプルベースによる手法や人体の手や足などのパーツの関連性をもとにした手法などが提案されている．サンプル画像を利用した代表的な手法として，Shakhnarovich [12] の手法がある．これは大量のシルエット画像のサンプルからハッシュを利用した最近傍探索による類似画像探索をベースとした姿勢推定である．Fathi らは，動画像における人の動きや見えの変化から人の姿勢を推定している [11]．これらは，大量のサンプル画像が必要となり，それらはあらかじめ正規化しておかなければならず，その切り出し精度により性能が大きく左右される．一方，Bissacco [13] の見えと動きを利用した手足のパーツベースの手法も同様に，人体の位置情報を必要としており，その位置精度が性能を左右する．このように，姿勢を推定するためには，人体の位置があらかじめ分かっていることが前提である．

行動認識は，姿勢の変化に着目した手法やフレーム間での動きや見えの変化をもとにした手法が提案されている．Blank らは，局所領域での時空間変化をとらえる ST-Patch を用いた手法 [8]，Niebles らは動画像における動きや見えの変化を階層的なモデルに利用した手法 [9] を提案している．提案されている多くの手法は，フレーム間での変化をもとにしている．しかしながら，人の行動によっては，1 フレームにおける姿勢から大まかな行動を推定することができる．Schindler ら [10] は，行動認識を行うために必要なフレーム数について実験しており，単純な行動の場合，1 フレームで認識できている．また，その他の行動については，3 から 5 フレームで行動を認識できると報告されている．

人体の検出から姿勢推定までを 1 つのフレームワークで実現している手法として，[14] [15] がある．[15] は，直立した人体における詳細な姿勢をとらえることを対象とし，サンプル画像に類似する姿勢を推定する．[14] は提案手法と同様に Randomized Trees を利用した姿勢推定を行っているが，直立姿勢における詳細な手足の位置を推定しスケルトンモデルとして姿勢を推定している．人の行動を認識する場合，これらのような詳細な姿勢情報は必要とせず，大まかな姿勢が分かれば行動を認識することができる．これらの手法は姿勢を推定するのみであり，行動を認識することはできない．

一方で，提案手法のように階層的な構造にブースティングを導入した方法 [17] がある．これは，ブースティングで学習した強識別器を木構造のノードとして扱うため，1 つの木で多クラス認識が可能であるが，過学習が生じやすい．提案手法は，過学習を防ぐために，ノードは 1 つ

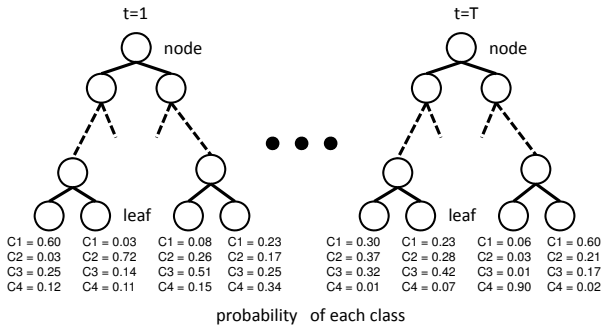


図 4　Randomized Trees の構造

の弱識別器をブースティングにより学習している．これにより，木構造のランダム性を維持したまま，より識別に有効な特徴量をブースティングで選択することが可能である．また，木構造を利用して階層的に局所領域の多クラス識別と物体の位置推定を行う手法がある [18] [19]．これは，物体の局所領域を抽出し，その局所領域が属するクラスを識別した後，複数のクラスの位置関係から物体の位置を推定する．これらの手法は，局所領域から向きや姿勢などを推定した後，その推定結果をもとに物体の位置を推定しているといえる．提案手法は，物体を 1 つの領域として捉えるため，位置の推定をする必要はなく，1 つのフレームワークで位置推定と同時に行動を認識することができる．

## 3. 提 案 手 法

Randomized Trees を構成する決定木は，階層的な構造であり，上位ノードは複数のクラスに共通する特徴，下位ノードは特定のクラスに対応する特徴を選ぶ必要がある．また，人体の検出から行動認識を行うためには，上位ノードは複数の姿勢と背景を識別するのに有効な特徴，下位ノードは特定の姿勢を認識するのに有効な特徴を選択する必要がある．しかしながら，ランダムに特徴候補を用意するため，有効な特徴がその中に含まれているとは限らない．そこで，我々は各ノード生成において，認識対象クラスに適した特徴を効率的に特徴選択する Boosted Randomized Trees を提案する．Boosted Randomized Trees は，ノード生成時に認識対象クラスをクラス尤度をもとに絞り込み，そのクラス集合に適した特徴を Joint Boosintg により選択する．これにより，上位ノードでは複数クラスに対応した特徴，下位ノードでは特定クラスに有効な特徴を選択することが可能である．

本章では，従来の Randomized Trees および Joint Boosting について述べた後，提案する Boosted Randomized Trees について述べる．

### 3.1 Randomized Trees

Randomized Trees [1] は，マルチクラスの認識に対応したアンサンブル学習手法の一つであり，キーポイントの検出 [3] やセグメンテーション [4] など様々な分野に応用されている．Randomized Trees は，学習サンプルに含まれるノイズに対してロバスト性があり，認識対象によっては，Boosting よりも精度が高い場合がある．また，Randomized Trees を構成する各決定木は独立しており，学習および評価の処理を決定木ごとに並列に行うことができるため高速である．

図 4 に Randomized Trees の構成を示す．図 4 のように，複数の決定木 $T$ から構成されている．各決定木には，分岐ノードと木の終端のリーフがある．$C$ 個のクラスを認識する場合，各リーフは，各クラス $c = (1, 2, ..., C)$ に対する確率分布を有する．また，決定木の各ノードは，スプリット関数により分岐処理を行う．スプリット関数は，式 (1) のように，特徴量 $I(x)$ が閾値 $\theta$ より小さい場合は左側の子ノード，大きい場合は右側の子ノードに進む．

$$I(x) = \begin{cases} < \theta & go\ left\ child\ node \\ \geqq \theta & go\ right\ child\ node \end{cases} \quad (1)$$

Randomized Trees の学習は，サブセットの作成，ノードの生成，サブセットの分割の 3 つのプロセスからなる．

まず，各決定木を学習するために，学習サンプル $X = \{x_i, c_j; i \in [1, N], j \in [1, C]\}$ からサブセット $X_s$ を作成する．サブセットは，ランダムに選択された $S$ 枚のサンプル画像からなる．

次に，ノードを生成する．ノードは，スプリット関数とその関数に用いる特徴と閾値から構成される．ノードの生成では，スプリット関数に用いる特徴と閾値を決める．特徴および閾値の選択は，Extremely Randomized Trees のようにランダムに用意された候補の中から最適な組み合わせを選択する方法 [3] とノード生成時に用いる学習サンプルに対して，最もエラー率が小さくなる特徴を選択する方法 [14] がある．ここでは，Extremely Randomized Trees を用いた場合について説明する．

ランダムに用意された特徴 $f_m; m \in [1, M]$ に対して，閾値 $\theta_{m,k}; k \in [1, K]$ をランダムに用意する．すべての特徴に対して，ランダムに閾値を用意し，最も最適な組み合わせとなる特徴と閾値を選択する．最適なものを選択する際の選択基準として，式 (2) のような情報量を用いる．

$$\triangle E = -\frac{|I_l|}{|I|} E(I_l) - \frac{|I_r|}{|I|} E(I_r) \quad (2)$$

ここで，$E(I_l)$ および $E(I_r)$ はある特徴と閾値の組み合わせを用いた場合に，右側 $r$ または左側 $l$ に分岐した各クラスのサンプルに対するシャノン情報量である．シャノン情報量は式 (3) のように算出する．

$$E(I) = -\sum_{j=1}^{C} P(c_j) \log P(c_j) \quad (3)$$

**アルゴリズム 1** Joint Boosting による学習アルゴリズム

**初期化:**

1. 学習サンプルの重み $w_i^c$ を初期化

*For* $i = 1..N$　　//サンプル数

　*For* $c = 1..C$　　//クラス数

　　学習サンプルの重み $w_i^c$ を初期化

**学習:**

2. $m = 1, 2, .., M$　　//選択する弱識別器数

　(a) $n = 1, 2, .., 2^C - 1$　　//クラスの全組み合わせ回数

　　(i) 全弱識別器候補に対するエラーを算出

　(b) 最小エラーとなるクラスの組合せ $n$ と弱識別器候補 $h_m^n(v, c)$ を選択

　(c) 重み $w_i^c$ を更新

3. 選択した弱識別器を組合せ $n$ ごとに統合し，$G^{S(n)}(v)$ を得る

$P(c_j)$ はノードにおけるクラス $c_j$ の確率分布である．

サブセットの分割は，上記の特徴量選択により得られた特徴を用いて，式 (1) のように分割する．閾値以下の場合は左の子ノードを生成するためのサブセット，閾値以上の場合は右側の子ノードを生成するためのサブセットとなる．

上記の処理をそれぞれの子ノードに対して，分割した新たなサブセットを用いて繰り返し行う．到達した学習サンプルが一定枚数より少なくなった場合，1 つのクラスの学習サンプルのみの場合，または一定のノードの深さになった場合にノード生成の処理を打ち切る．終端ノードはリーフ $l$ として，各クラスの確率分布 P(c) を持っている．クラス $c_j$ の確率分布は式 (4) のように求める．

$$P(c_j|l) = \frac{|I_{c_j}|}{|I|} \tag{4}$$

$|I|$ は全クラスのサンプル数，$|I_{c_j}|$ はクラス $c_j$ のサンプル数である．

評価は，生成されたすべての決定木から得られる各クラスの確率を求める．入力された画像は各決定木の 1 つのリーフへ到達する．到達したリーフ $L = \{L_t; t \in [1, T]\}$ が持つ確率分布 $P(C|L_t)$ を式 (5) のようにクラスごとに累積して平均を求める．

$$P(C|L) = \frac{1}{T}\sum_{t=1}^{T} P(C|L_t) \tag{5}$$

式 (5) の平均確率が最大となるクラスを認識結果として出力する．

### 3.2 Joint Boosting

Joint Boosting [16] は，マルチクラスの Boosting であり，複数のクラスに共通する特徴を選択することができる．Joint Boosting は，式 (6) に示すように，全クラスの部分集合 $S(n)$ に対する強識別器を学習する．

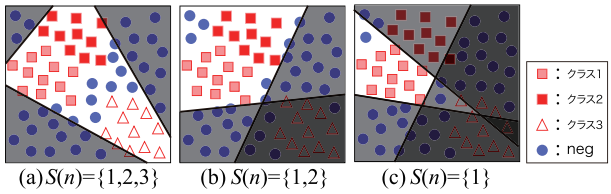


図 5　Joint Boosting による学習例

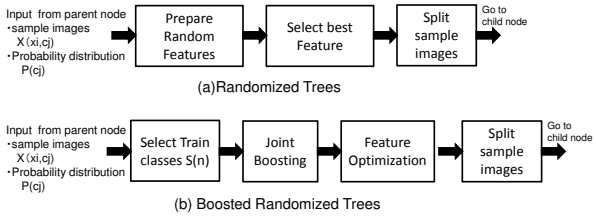


図 6　ノードの学習方法

$$G^{S(n)}(v) = \sum_{m=1}^{M} h_m^n(v) \tag{6}$$

ここで，$h_m^n(v)$ は $m$ 番目の弱識別器であり，$v$ は特徴ベクトルである．アルゴリズム 1 に Joint Boosting の処理手順を示す．Joint Boosting の学習は，正解クラスの組み合せを変えながら最適な弱識別器を選択する．すなわち，$2^C - 1$ の組み合わせの中からエラーが最小となるクラス集合 $S(n)$ に対応した弱識別器 $h_m^n(v)$ を得る．3 クラスの場合，組み合わせ数は 7 通りである．Joint Boosting では，対象とするクラス集合 $S(n)$ に共通する弱識別器を統合し，弱識別器の集合 $G^{S(n)}$ を得る．3 クラスを Joint Boosting で学習する場合の学習例を図 5 に示す．図 5 のクラス 1，クラス 2，クラス 3 は正解クラスであり，$neg$ は非正解クラスである．Joint Boosting では，非正解クラスの学習サンプルをあらかじめ別途用意している．$S(n) = \{1, 2, 3\}$ の場合，図 5(a) のようにすべてのクラスを正解クラスとするように弱識別器を学習することができる．一方，$S(n) = \{1, 2\}$ の場合は，図 5(b) クラス 1 とクラス 2 を正解するように弱識別器を学習する．また，$S(n) = \{1\}$ の場合は，図 5(c) のようにクラス 1 を正解するように弱識別器を学習することができる．

### 3.3 Boosted Randomized Trees

従来の Randomized Trees および Boosted Randomized Trees におけるノードの生成方法を図 6 に示す．図 6(a) に示すように，従来の Randomized Trees のノード生成は，特徴と閾値をランダムに用意する処理，情報利得が最大となる特徴および閾値の選択，サンプル画像の評価と子ノードのサブセット作成の 3 つのプロセスからなる．一方，提案する Boosted Randomized Trees は，クラス集合の定義，Joint Boosting による事前特徴選択，特徴の最適化，子ノードのサブセット作成の 4 つのプロ

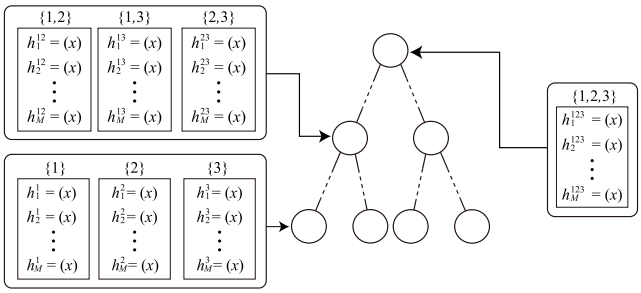


図 7　Joint Boosting による事前特徴選択

セスから構成される．ここでは，提案手法の各プロセスについて述べる．

### 3.3.1　クラス集合の定義

Joint Boosting は，クラスの全組み合わせの中から特定のクラス集合に対する弱識別器を選択する．そのため，クラス集合をあらかじめ決めることができない．Boosted Randomized Trees を構成する決定木は階層的な構造であり，上位ノードは複数のクラス，下位ノードは特定のクラスに対応している．また，人体検出から行動認識までを階層的に行う場合，上位ノードは人体と背景，下位ノードは特定の姿勢または行動を区別する特徴を選択することができれば，効率的な決定木を学習することができる．すなわち，上位ノードは複数のクラスのクラス集合，下位ノードは特定のクラスのクラス集合に対する弱識別器を学習する方がよい．そこで，ノードにおける最適なクラス集合をクラス尤度をもとに定義する．各クラスの尤度は式 (4) で示したように，各クラスに対する確率から求める．あるノードにおいて，クラスの組合せのうち，式 (7) のようにクラス尤度の合計が閾値 $\tau$ より大きくなる組み合わせをクラス集合 $S(n)$ として定義する．

$$S(n) = id \in \mathcal{I} : L(id) > \tau \tag{7}$$

$id$ は組み合わせの番号，$\mathcal{I}$ は全組み合わせ，$L(id)$ はクラス尤度の合計，$\tau$ はしきい値を表わす．その際，クラス数が最も少なくなる組合せを選ぶ．上位ノードは，多数のクラスが存在しており，各クラスの尤度は低くなる．そのため，多くのクラスを組み合わせたクラス集合を定義することができる．一方，下位ノードは，特定のクラスに偏っているため，特定クラスの尤度が高くなる．そのため，尤度の低いクラスを認識対象としてみなす必要がなく，クラス集合は特定クラスのみになる．これにより，特定のクラスの組合せにのみ Joint Boosting により学習することができる．

### 3.3.2　Joint Boosting による事前特徴選択

3.2 節で述べた Joint Boosting は，与えられたクラスの全組合せの中から最適な特徴を選択する．そのため，上位ノードにおいて，複数のクラスに共通した特徴が選択されず，特定のクラスにのみ特化した特徴が選択される場合がある．そこで，提案手法では，3.3 節で述べたクラス尤度をもとにしたクラス集合を利用して，クラスの組合せを限定する．すなわち，アルゴリズム 1 で示す学習手順において，ループ (a) を全組み合わせでなく，限定した組合せにのみ行う．図 7 に示すように，多くのクラスを含む上位ノードでは，クラス集合に多くのクラスが含まれており，それに共通する特徴が選択される ($S(n) = \{1, 2, 3\}$)．一方，下位ノードは，特定のクラスのみクラス集合に含まれているため，特定のクラスにのみ適した特徴が選択される ($S(n) = \{1\}$)．このように，クラス尤度を用いたクラス集合を用いて，Joint Boosting の特徴選択方法を限定することで，階層的な特徴を効率的に選択することが可能である．

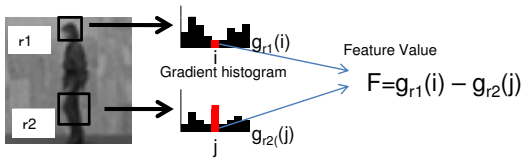


図 8　本手法に用いる特徴量

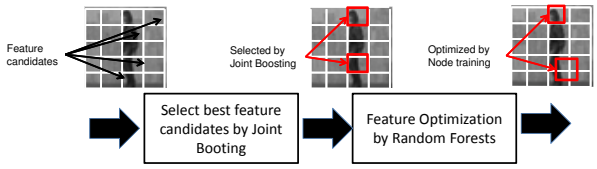


図 9　特徴量の最適化方法

### 3.3.3　特　徴　量

本稿では，人体検出に有効な勾配方向ヒストグラムをベースとした特徴量を用いる．特徴量の概要を図 8 に示す．特徴量は式 (8) に示すように，2 つの局所領域 $r_1$,$R_2$ における勾配方向ヒストグラムの 1 つのビンの値 $g_{r1}(i)$, $g_{r2}(j)$ の差である．本稿では勾配方向数を 9 としている．

$$F = g_{r1}(i) - g_{r2}(j) \tag{8}$$

ここで，$i$ および $j$ は勾配ヒストグラムのビンの位置である．勾配ヒストグラムの差をとらえることで，局所領域におけるエッジの関連性をとらえることができる．

### 3.3.4　特徴量の最適化

本稿で用いる特徴量は，2 つの領域間における差に着目しており，領域の組み合わせは非常に膨大である．そこで，本稿では，図 9 に示すような Joint Boosting での事前特徴選択とノード内における特徴最適化の 2 段階の特徴選択を行う．まず，Joint Boosting で選択する特徴候補は，領域の位置および大きさを一定の間隔でサンプリングされた領域を特徴候補として与える．Joint Boosting では，これらの領域の組み合わせのうち，最適な 2 つの領域の組み合わせを特徴として選択する．次に，各領域をサンプリング間隔の範囲で位置および大きさをずらし，その中で最適な組み合わせを最終的な特徴として選択する．スプリット関数の閾値は，Joint Boosting

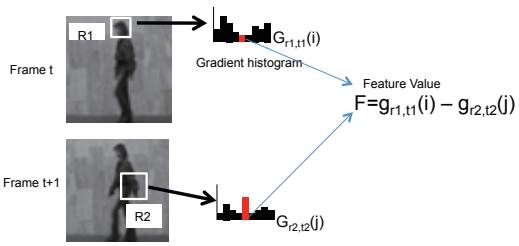


図 10　動画像に拡張した特徴量

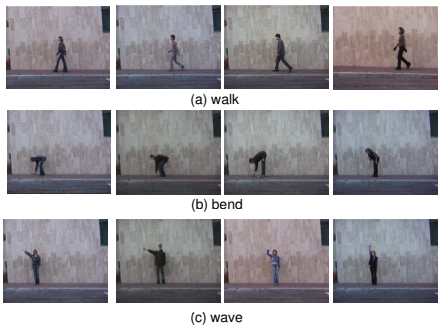


図 11　実験画像例

により得られた閾値をもとに，新たな組み合わせに対して最適な閾値を求める．特徴および閾値は，式 (3) を用いて，情報利得が最大となるものを選択する．

### 3.4　特徴量の行動認識への拡張

フレーム間での特徴の変化は，行動認識において重要な特徴である．そこで，行動認識を行うために，特徴量を図 10 のように，複数フレーム間における勾配方向ヒストグラムの差を捉えることができるように拡張する．

$$F = g_{r1,t1}(i) - g_{r2,t2}(j) \tag{9}$$

ここで，$t_1$ および $t_2$ は着目するフレームである．このように複数フレーム間に拡張することで，同一フレーム内における勾配方向の違いをとらえるだけでなく，異なるフレーム間での勾配方向の変化をとらえることができる．

## 4.　実　　験

### 4.1　実 験 概 要

本提案手法の有効性を示すために，人体検出率と行動認識率の性能を Joint Boosting および Randomized Trees を用いた場合との比較実験を行う．入力として 1 枚の静止画を与えた場合，動き情報はとらえないため，行動認識は大まかな姿勢推定とみなすことができる．本稿では，1 枚の静止画を入力した場合の行動認識を姿勢推定としてみなし，実験を行う．提案手法は，直立した人体だけでなく様々な姿勢の人体検出から行動認識までを 1 つのフレームワークで行うことができる．そこで，姿勢変化を含む評価データセットとして，行動認識の研究で一般的に用いられるデータセット [8] を用いる．評価データセットは，行動ごとに 9 名分の被験者の動画像から構成されている．本稿では，姿勢の異なる行動シーンにおける人体検出および姿勢推定の評価を行うために，図 11 に示すような“ walk ”,“ bend ”,“ wave ”の 3 つの行動シーンを用いる．行動認識は,“ bend ”および“ wave ”を構成する行動を細分化して認識する.“ bend ”はしゃがむ、起き上がるの 2 つのどちらであるかを認識する.“ wave ”は手を挙げるまたは手を下げるのどちらであるかを認識する．評価は 9 人のうち 1 人を評価データとし，残り 8 人を学習データとして用いる．これらの評価をすべての人物に対して行い，平均精度を求めた．各動画像のフレーム数は一定ではないため，評価対象フレームはフレーム内に人物が映っている範囲のうち 30 フレームとした．人体の検出は，スライディングウィンドウ方式により網羅的に探索する．そして，式 (5) から得られるクラスの平均確率が閾値より大きい領域を人体領域として出力する．本実験では，閾値を 7.5 としている．

### 4.2　Boosted Randomized Trees の学習

Randomized Trees および Boosted Randomized Trees で学習する決定木数は 10 とし，決定木の最大深さは 15 階層とする．Randomized Trees のノード生成時に用意する特徴候補はランダムに 100 個用意し，各候補に対して 100 個の閾値を同様にランダムに用意する．これらの組み合わせの中から最適なノードを選択する．一方，Boosted Randomized Trees は Joint Boosting により 10 個の特徴候補を事前学習する．Joint Boosting に与える特徴候補は，位置を 4 ピクセル間隔でサンプリングして限定している．Joint Boositng により得られた各候補に対して，位置を 4 サンプリングで間引かれた 4 ピクセルの範囲で変えて最適な組み合わせを選択する．各決定木の学習には，48x48 ピクセルに正規化した各姿勢のサンプル画像を 240 枚用いる．サンプル画像は全サンプル画像からランダムに選択する．非人体画像は，背景データセットから 1 万枚の正規化した画像を用意してランダムに 2000 枚を選択する．行動認識の学習は，3 フレームを 1 つのサンプル動画像とし，ランダムに動画像を 240 セット選択する．決定木数およびその他の設定は，静止画における実験と同様である．

### 4.3　実 験 結 果

#### 4.3.1　人体検出結果

図 12 に Joint Boosting(JB)，Randomized Trees(RTs) および Boosted Randomized Trees(BRTs) の人体検出率を示す．Boosted Randomized Trees は，Joint Boosting および Randomized Trees を用いた場合に比べて，全姿勢において検出率が向上していることが分かる．walk データセットは直立姿勢の人体であり，体型による個人差はあるものの姿勢のバリエーションが少ないため，検出率は高い．そのため，すべての手法の検出率に大きな

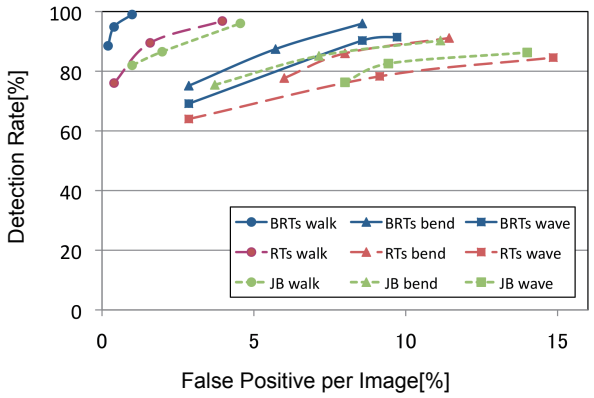


図 12　人体検出率の ROC カーブ

表 1　データセットごとの姿勢推定率 [%]

| dataset | JB | RTs | BRTs |
|---|---|---|---|
| walk | 98.7(95.6) | 98.1(96.4) | 98.4(97.6) |
| bend | 88.3(77.7) | 91.3(71.6) | 94.5(84.9) |
| wave | 68.3(61.5) | 65.3(55.0) | 70.0(63.2) |
| average | 85.1(78.3) | 84.9(74.3) | 87.6(81.9) |

差はない．一方，bend はしゃがみこむ角度により様々な形状があり，姿勢のバリエーションが多い．wave も同様に，手の挙げ方や挙げている行動過程における手の角度が様々であり，バリエーションが多い．Joint Boosting はクラス間またはクラス内に共通するバリエーションが少ない特徴をとらえる．また，クラスの組み合わせは事前に定義できないため，特定のクラスに偏る場合もある．そのため，これらの姿勢における検出率が低下している．Randomized Trees はランダムに特徴を選択するため，これらの姿勢をとらえるのに有効な特徴が選択されない場合が多い．一方，提案手法は，ノードに最適なクラスの組み合わせを定義するため，複数のクラスに共通する有効な特徴を事前に選択することができる．また，木構造のため，同一クラスに対して複数のリーフが存在する．そのため，クラス内のバリエーションに対応するリーフを生成することができる．これにより，提案手法は bend および wave データセットにおいて，Joint Boosting および Randomized Trees より検出率を向上させることができている．

**4.3.2**　静止画における行動認識結果

前節での実験結果に対して，検出結果がどの行動クラスに属していたかを調べた．静止画における行動認識は，大まかな姿勢推定とみなすことができる．表 1 は，検出した人体に対する姿勢推定率であり，かっこ内は，全評価画像に対する姿勢推定率である．walk データセットにおいて，検出された人体に対する姿勢推定率は Joint Boosting および Randomize Trees と差はない．しかし，bend および wave データセットでは，提案手法は Joint Boosting および Randomized Trees に比べて，姿勢推定率は向上している．これより，クラス内でのバリエーションが多い各姿勢に対して有効な特徴が選択することができていることが分かる．

表 2　RTs の行動認識精度

| | walk | bend-up | bend-down | Wave-up | Wave-down | False Negative |
|---|---|---|---|---|---|---|
| walk | 95.5 | | | 2.0 | | 2.5 |
| bend-up | | 91.1 | 1.0 | 3.2 | | 4.7 |
| Bend-down | | 2.1 | 89.5 | | 3.0 | 5.4 |
| wave-up | | 5.5 | | 80.1 | 1.3 | 13.1 |
| wave-down | | | 4.2 | 1.8 | 84.7 | 9.3 |
| average | 89.4% | | | | | |

表 3　BRTs の行動認識精度

| | walk | bend-up | bend-down | Wave-up | Wave-down | False Negative |
|---|---|---|---|---|---|---|
| walk | 99.5 | | | | | 0.5 |
| bend-up | | 95.1 | | | | 4.9 |
| Bend-down | | | 95.6 | | | 4.4 |
| wave-up | | | | 85.1 | | 14.9 |
| wave-down | | | | | 90.3 | 9.7 |
| average | 93.9% | | | | | |

**4.3.3**　動画像における行動認識結果

動画像に対する行動認識性能の比較実験を行った．その結果を表 2 および表 3 に示す．Boosted Random Trees は Randomized Trees を用いた場合に比べて，行動認識率が向上していることがわかる．従来法は，誤った行動として認識するケースがあるが，提案手法は，行動認識ができなかったものは検出できない場合であり，誤認識がない．提案手法は，動画像において，フレーム間の変化をとらえる特徴を事前に効率的に選択できているため，認識精度向上につながったといえる．これより，提案手法は動画像を入力とした場合に，人体検出から行動認識までを 1 つのフレームワークで実現することができている．

## 5.　考　　察

クラス尤度をもとにしたクラス集合の定義と Joint Boosting による事前特徴選択の有効性を示すために，各ノードにおけるクラスの分割結果を図 13 に示す．図 13 のルートノードでは，Joint Boosting により 3 つの姿勢に共通する足元をとらえる特徴が選択されている (a)．ルートノードで選択された特徴により学習サンプルは 2 つのサブセットに分割される (b)．分割された一方は bend と背景クラス，もう一方は walk と wave を主に含んでいる．ルートノードにおいて，形状の類似している walk と wave，それ以外のクラスに分けることができている．1 階層目の左側のノード (c) では，bend と背景クラス，右側のノード (d) では walk と wave を識別しており，2 層目では，各クラスと背景を主に識別している．このように，最初の数層で様々な姿勢の人体と背景クラスを識別することができており，人体検出に相当している．また，各姿勢ごとに分割することもできており，人体検出と同時に大まかな姿勢推定も行うことができている．さらに，学習を数層行うことで，より詳細な行動クラスに分割している．(f) では bend を多く含むノードになっており，bend クラスに共通するしゃがんだ背中付近に相当する特徴が Joint Boosting により事前学習されている (e)．(h) では walk を多く含むノードになっており，walk に共通する手や足の位置に相当する特徴が事前学習されており (g)，walk と wave を識別している．このよ

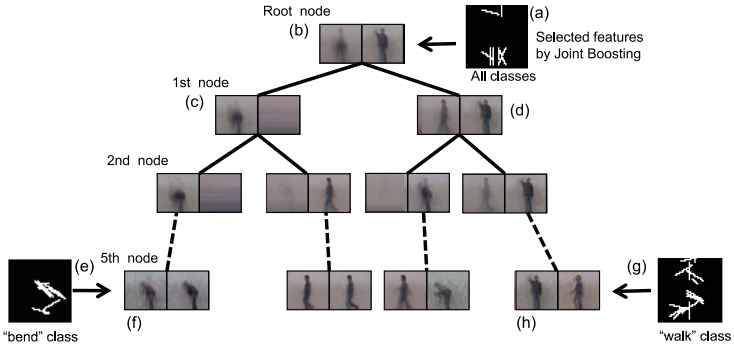


図 13　各ノードにおける事前選択特徴とクラス分割結果例

うに，下位ノードでは各行動クラスを認識するための特徴が多く選択されている．

提案手法で導入した事前特徴選択により，上位ノードは多数のクラスに共通する特徴，下位ノードは特定のクラスに対応する特徴を階層的に選択することができている．これにより，1 つのフレームワークで階層的に人体検出から姿勢推定，行動認識を実現している．

## 6. ま と め

我々は，人の行動を理解するために，人体検出から行動認識までを 1 つのフレームワークで実現する Boosted Randomized Trees を提案した．提案手法は，決定木のノード生成時に各クラスの尤度をもとに認識対象となるクラスを定義し，Joint Boosting により効率的に特徴を選択することで，上位ノードは多数のクラスに対応した特徴，下位ノードは特定のクラスに対応した特徴を階層的に選択することが可能である．今後は，姿勢のバリエーションを増加させていくとともに，一般物体認識などに応用していく．

## 文　献


[1] L. Breiman, “Random forests”, Machine Learning, No.45(1), pp. 5-32, 2001.

[2] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees”, Machine Learning, No.36, Vol.1, pp. 3-42, 2006.

[3] V. Lepetit, P. Lagger, and P. Fua, “Randomized trees for real-time keypoint recognition”, IEEE Conf. on Computer Vision and Pattern Recognition, pp. 775-781, 2005.

[4] J. Shotton, M. Johnson, R. Cipolla, “Semantic Texton Forests for Image Categorization and Segmentation”, IEEE Conf. on Computer Vision and Pattern Recognition, 2008.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection”, IEEE Conf. on Computer Vision and Pattern Recognition, 2005.

[6] Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, “Fast human detection using a cascade of histograms of oriented gradients, IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 1491-1498, 2006.

[7] C. Hou, H. Ai and S. Lao, “Multiview Pedestrian Detection Based on Vector Boosting”, Asian Conf. of Computer Vision, pp. 210-219, 2007.

[8] M. Blank, L. Gorelick, E. Shechtman, M. Irani,and R.Basri. “Actions as space-time shapes”, IEEE Intl. Conf. of Computer Vision, 2005.

[9] J. C. Niebles and L. Fei-Fei. “A hierarchical model of shape and appearance for human action classification”, IEEE Conf. on Computer Vision and Pattern Recognition, 2007.

[10] K. Schindler and L. Van Gool, “Action snippets: How many frames does human action recognition require?”, IEEE Conf. on Computer Vision and Pattern Recognition, 2008.

[11] A. Fathi and G. Mori, “Human pose estimation using motion exemplars”, IEEE Intl. Conf. of Computer Vision, 2007.

[12] G. Shakhnarovich, P. Viola, D. Trevor, “Fast pose estimation with parameter-sensitive hashing”, IEEE Intl. Conf. of Computer Vision, 2003.

[13] A. Bissacco, M.-H. Yang and S. Soatto, “Fast human pose estimation using appearance and motion via multi-dimensional boosting regression”, IEEE Conf. on Computer Vision and Pattern Recognition, 2007.

[14] G. Rogez, J. Rihan, S. Ramalingam, C. Orrite and P. H.S. Torr, “Randomized Trees for Human Pose Detection”, IEEE Conf. on Computer Vision and Pattern Recognition, 2008.

[15] A. Bissacco, M.-H.Yang and S. Soatto, “Detecting humans via their pose”, In NIPS, pp. 169-176, 2006.

[16] A. Torralba, K. P. Murphy and W. T. Freeman, ”Sharing features: efficient boosting procedures for multiclass object detection”, IEEE Conf. on Computer Vision and Pattern Recognition, 2004.

[17] Z. Tu, ”Probabilistic boosting-tree: learning discriminative models for classification, recognition and clustering”, IEEE Intl. Conf. of Computer Vision, 2005.

[18] J. Gall and V.Lempitsky, ”Class-specific hough forests for object detection”, IEEE Conf. on Computer Vision and Pattern Recognition, 2009.

[19] R. Okada, ”Discriminative and Generalized Hough Transform for Object Detection”, IEEE Intl. Conf. of Computer Vision, 2009.