

Conditional Classification Forests を用いた人体パーツ識別

○ 縣 禎輝†, 三品 陽平†, 弓場 竜‡, 藤吉 弘亘†

○ Yoshiki AGATA †, Yohei MISHINA †, Ryo YUMIBA ‡ and Hironobu FUJIYOSHI †

†: 中部大学, {a_y,mishi} @vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

‡: 株式会社 日立製作所 日立研究所, ryo.yumiba.xp@hitachi.ac.jp

本稿では, Conditional Classification Forests を用いた人体の向きを考慮した人体パーツ識別手法を提案する. 提案する Conditional Classification Forests は, 人体の向きを回帰により求める Regression Forest と人体パーツをマルチクラス識別する Classification Forest から構成される. まず Regression Forest により人体の向きを回帰により推定する. そして回帰推定した人体の向きを条件として, Conditional Classification Forests により人体パーツを識別する. 提案手法は, 全方向の学習サンプルを 1 つの Classification Forest で学習した手法と比較した場合, 精度が 13.2% 向上した. また, 提案手法は 4.96° の誤差で人体の向きを回帰推定できることを確認した.

<キーワード> 人体パーツ識別, Conditional Classification Forests, Classification Forest, Regression Forest

1. はじめに

画像認識技術による人体パーツの識別は, 防犯システムのための動作解析, テレビの操作やゲーム入力のためのジェスチャ・ユーザインタフェースなどの要素技術として重要である. 従来の人体パーツ識別には, Kinect などの距離計測カメラから取得した距離情報を用いた手法が多く提案されている[1]-[4]. 距離情報を利用することで人体パーツの前後関係や形状を捉えることが可能となり, 高精度な人体パーツ識別の実現が可能となる. また, Kinect[5]や TOF 方式の赤外線を利用した距離計測カメラ[6]は, 照明のない暗闇でも距離画像の取得が可能というメリットがある.

距離画像からの人体パーツ識別で最も成功した手法は, Shotton 等が提案した Random Forest を用いたアプローチ[7]である. 本手法は, Xbox360 のジェスチャによるユーザインタフェース技術として発表され, リアルタイム性と高精度な識別, また, Kinect が距離計測カメラとして安価であることから本分野が盛んに研究される発端となった. 本手法で用いられる Random Forest[8]は, 複数の決定木の識別結果を統合するアンサンブル学習手法であり, 大規模なデータセットの学習が容易である. そのため, 大規

模データセットを必要とする人体パーツのようなマルチクラス識別に適した手法である.

一方, Dantone 等は回帰を行う Regression Forest[9]を用いて人の頭部の向きを条件として, 顔特徴点を検出する Conditional Regression Forests[10]を提案した. 文献[10]では, まず Regression Forest により頭部の向きを回帰により推定し, 頭部向きの確率分布を算出する. 次に, 頭部の向き毎に学習した Conditional Regression Forests から複数の Regression Forest により顔特徴点座標の事後確率をそれぞれ算出する. そして, 頭部向きの確率分布と顔特徴点座標の事後確率から, 最終的な顔特徴点座標の条件付き確率を算出する. これにより, 頭部の向きを考慮した高精度な顔特徴点検出を実現した. 2 段階に Regression Forest を構築する本アプローチは, 人体パーツ識別の際に人体の向きを条件とすることで同様の効果が期待できると考えられる.

そこで本稿では, Regression Forest により回帰した人体の向きを条件とした Conditional Classification Forests による高精度な人体パーツ識別手法を提案する. 提案手法では, まず Regression Forest により人体の向きを回帰推定する. 回帰推定した向きを条件として, 向き毎に学習した Conditional Classification Forests から対応する Classification Forest の重みを大

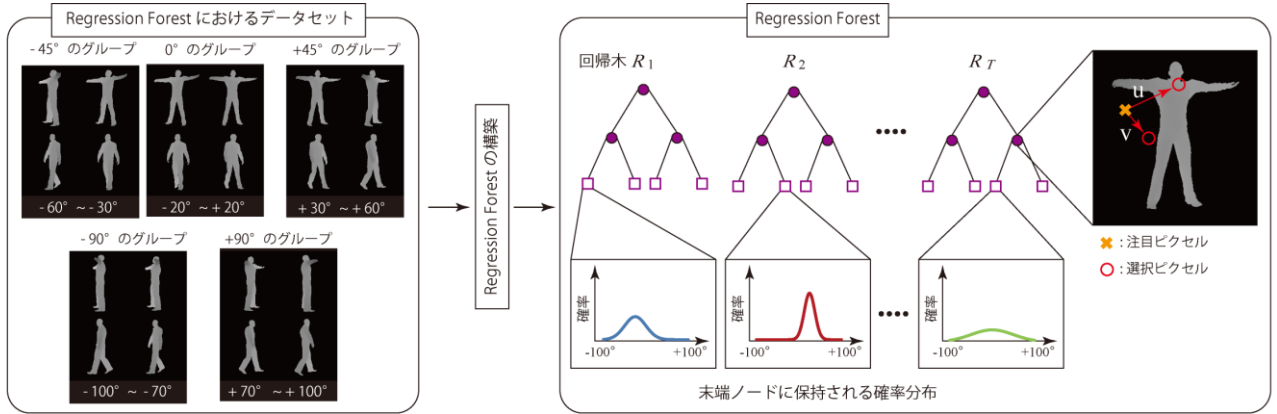


図1 Regression Forest の学習

きくすることで、高精度な人体パーツ識別を実現する。

2. Regression Forest による人体向き推定

人体向き推定のための Regression Forest の学習と回帰推定について述べる。

2.1 学習

人体モデルにより $\{-100 \sim +100\}^\circ$ 方向の範囲で 10° 毎に生成した距離画像データを $\omega \in \{-90, -45, 0, +45, +90\}^\circ$ の5方向にグループ化する。図1に示すように、 -45° のグループには $\{-30, -40, -50, -60\}^\circ$ 方向の距離画像、 -0° のグループには $\{-20, -10, 0, +10, +20\}^\circ$ 方向の距離画像が含まれる。つまり、基準となる角度に最も近い方向の距離画像を1つのクラスにグループ化する。学習サンプルは、人体の距離画像とその向きのセットとなる。

Regression Forest の学習では、Random Forest[8]と同様に学習サンプルをサブセットに分割し、 T 個のサブセット毎に回帰木を作成する。回帰木の構築は、学習サンプルを分岐ノードにより分岐することで行う。各分岐ノード n に対し、入力サンプル集合 S_n 、左右に分岐するサンプル集合をそれぞれ S_l 、 S_r とすると分岐は以下の式により行われる。

$$S_l = \{i \in S_n \mid f_\theta(I) < t\} \quad (1)$$

$$S_r = S_n \setminus S_l \quad (2)$$

ここで、 $f_\theta(I)$ は入力画像 I から抽出された特徴量、 t はしきい値である。回帰木の分岐ノードにおける特徴量は、文献[8]と同様に2点間の距離差を特徴量として用いる。図1に示すように、入力画像 I からピクセル \mathbf{x} とオフセット量 $\theta = (\mathbf{u}, \mathbf{v})$ をランダムに選択し、2点の距離値 d の差分を次式より算出する。

$$f_\theta(I, \mathbf{x}) = d_l\left(\mathbf{x} + \frac{\mathbf{u}}{d_l(\mathbf{x})}\right) - d_l\left(\mathbf{x} + \frac{\mathbf{v}}{d_l(\mathbf{x})}\right) \quad (3)$$

各分岐ノードでは、(4)式に示す情報利得(Information gain) ΔE により、その分岐を評価する。

$$\Delta E = -\frac{|S_l|}{|S_n|} E(S_l) - \frac{|S_r|}{|S_n|} E(S_r) \quad (4)$$

ここで、 $E(S)$ はある特徴量 f_θ としきい値 t による各クラスの情報エントロピーである。分岐ノードにおける分岐関数には、情報利得が最大となる特徴量としきい値を用いる。末端ノードでは、トラバーサルして到達したサンプルの人体の向きを表す角度の平均 $\bar{\omega}_l$ と分散 σ_l^2 を算出し、次式のように正規分布 N を求めることで、人体の向きの確率分布を保持する。

$$p(\omega \in N \mid \bar{\omega}_l, \sigma_l^2) \quad (5)$$

2.2 回帰推定

入力画像 I が与えられたとき、回帰木をトラバーサルして到達した末端ノード l が保持する人体の向きの確率分布 $p(\omega \mid l)$ を(6)式により統合する。

$$p(\omega \mid I) = \frac{1}{T} \sum_t p(\omega \mid l_t) \quad (6)$$

これにより、図2に示すような各末端ノードが持つ正規分布を重ね合わせた確率分布を得ることができる。以上により算出した確率分布を Conditional Classification Forests において人体パーツ識別に利用する条件の確率分布とする。

3. Conditional Classification Forests による人体パーツ識別

提案手法では、まず2章の Regression Forest により入力画像から人体の向きの確率分布を算出する。そして、入力画像の各画素 x に対して、Conditional

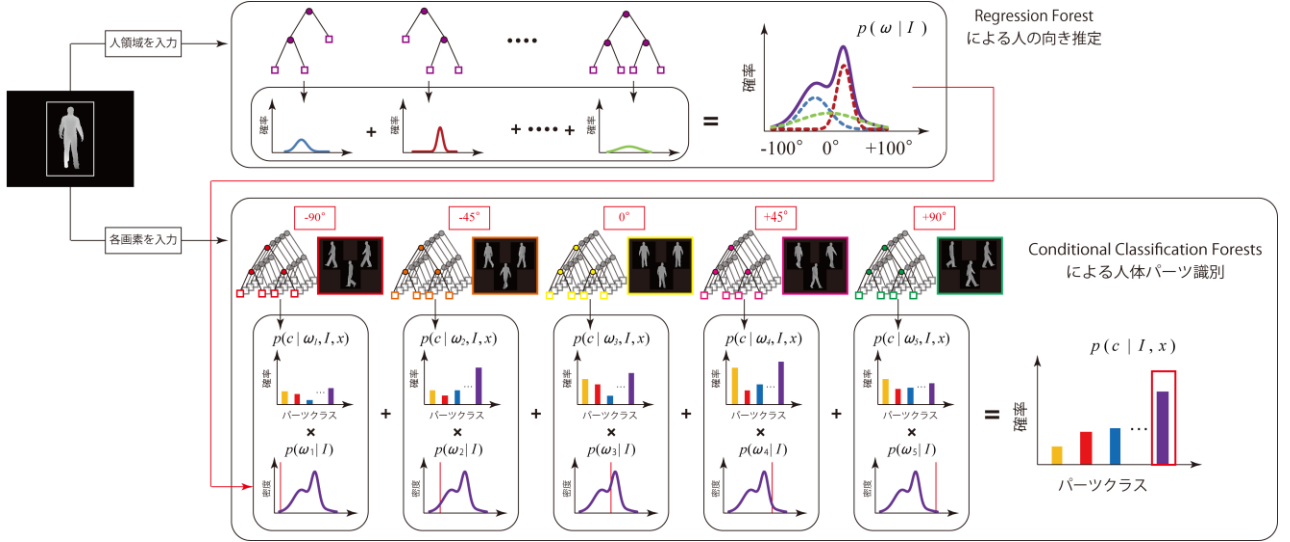


図2 Conditional Classification Forests による人体パーツ識別の流れ

Classification Forests を用いて対応する末端ノードの事後確率 $p(c | \omega_i, I, x)$ と人体の向きの確率分布 $p(\omega_i | I)$ から条件付き確率 $p(c | I, x)$ を求める。これにより、人体の向き推定結果を考慮した人体パーツ識別を行うことができる。人体パーツの事後確率は、人体の向き ω 毎に学習した Classification Forest により算出する。以下に提案する Conditional Classification Forests の学習と識別について述べる。

3.1 学習

Conditional Classification Forests の学習は、人体の向き $\omega \in \{-90, -45, 0, +45, +90\}^\circ$ の5方向毎に Classification Forest を構築することにあたる。人体パーツ識別における学習サンプルは、入力画像 I の各ピクセル x とそのピクセルが属するパーツのクラス番号 c が学習データとなる。

Classification Forest の学習は、基本的に2.1節と同様である。異なる点は、学習サンプルと末端ノードの表現が異なる。

まず、学習サンプルをサブセットに分割し、分岐は(1)(2)式により行う。分類木の分岐ノードにおける特徴量も回帰木と同様に文献[8]で用いられた2点間の距離差を用いる。2点間の距離差を用いることで、各人体パーツの前後関係や形状を捉えることができる。サンプルを分岐させた後、(4)式により情報利得 ΔE を求め、情報利得 ΔE が最大となる特徴量としきい値の組み合わせを分岐ノードのパラメータとして決定する。末端ノード l では、到達したサンプルのクラスラベルから、次式のようにクラス c のサンプル集合 S_c の確率を算出することで、事後確率 $p(c | l)$ を保持する。

$$p(c | l) = \frac{|S_c|}{|S|} \quad (7)$$

ここで S は全クラスのサンプル数である。以上のように、5つの Classification Forest からなる Conditional Classification Forests を構築する。

3.2 識別

学習により構築した Conditional Classification Forests による人体パーツ識別の流れを以下に示す(図2)。

Step1. Regression Forest より入力画像 I の人体向きを回帰推定し、(6)式により確率分布 $p(\omega_i | I)$ を求める。

Step2. 人体の向き $\omega \in \{-90, -45, 0, +45, +90\}^\circ$ 毎の各 Classification Forest より入力画像 I の各画素 x の人体パーツクラスの事後確率 $p(c | \omega_i, I, x)$ を求める。

$$p(c | \omega_i, I, x) = \frac{1}{T} \sum_{t=1}^T p_t(c | l) \quad (8)$$

Step3. 人体向きの確率分布 $p(\omega_i | I)$ と5つの人体パーツクラスの事後確率 $p(c | \omega_i, I, x)$ を用いて、次式より条件付き確率を算出する。

$$p(c | I, x) = \sum_i (p(c | \omega_i, I, x) p(\omega_i | I)) \quad (9)$$

Step4. ナイーブベイズ識別を用いて出力クラス C_i^* を決定する。

$$C_i^* = \arg \max_{c_i} p(c_i | I, x) \quad (10)$$

以上により、Conditional Classification Forests による人体の向きを考慮した人体パーツの識別を行う。

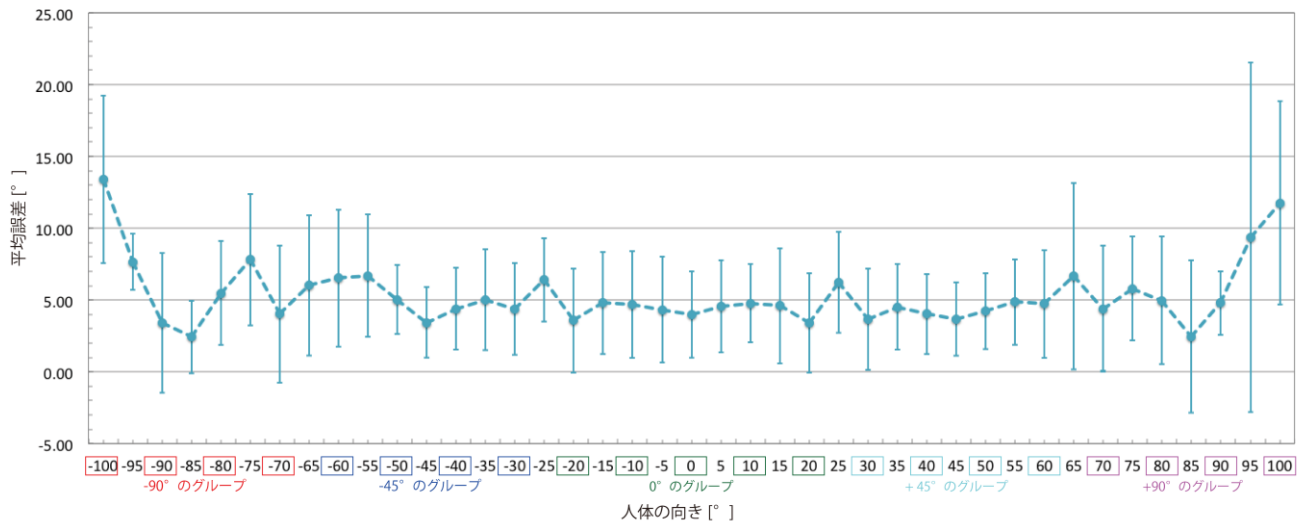
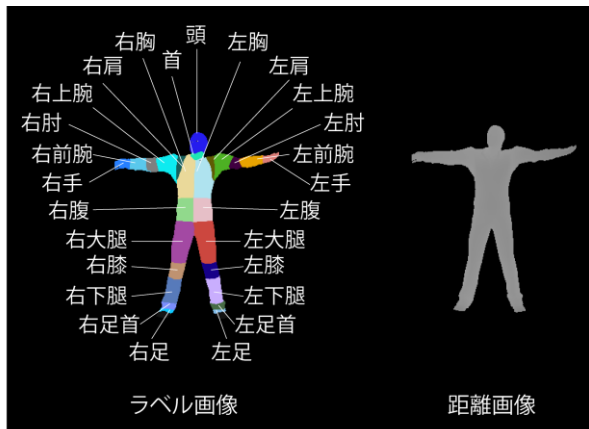
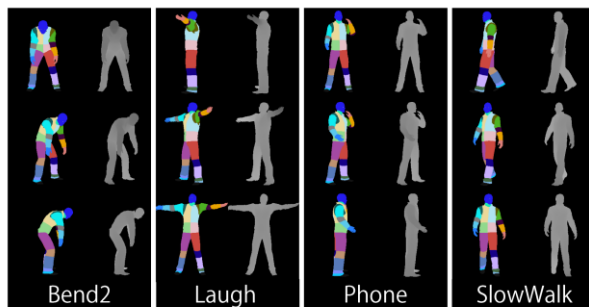


図 4 人体向きの推定精度



(a) 教師用ラベル画像と距離画像



(b) 動作毎のデータセット

図 3 学習データセット

4. 評価実験

提案手法の有効性を示すために、人体モデルより作成したサンプルと実環境で撮影したサンプルにより評価する。

4.1 実験概要

従来法として全ての向きを含む学習データセットを Classification Forest により学習した手法と提案手法 (Conditional Classification Forests) を比較する。Regression Forest と各 Classification Forest の学習

パラメータを表 1 に示す。

表 1 学習パラメータ

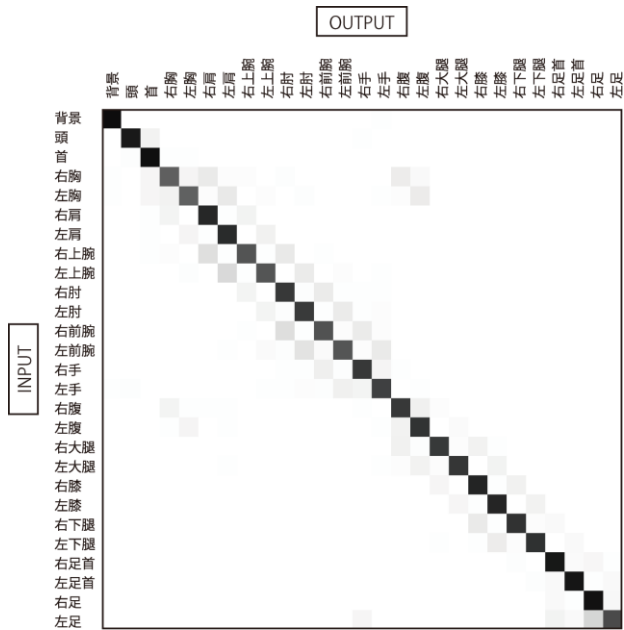
パラメータ	木の数	木の深さ	選択回数		サブセットの割合
			特徴量	しきい値	
Regression	10	20	100	10	0.75
Classification	5	20	30	15	0.25

4.2 学習データセットの生成

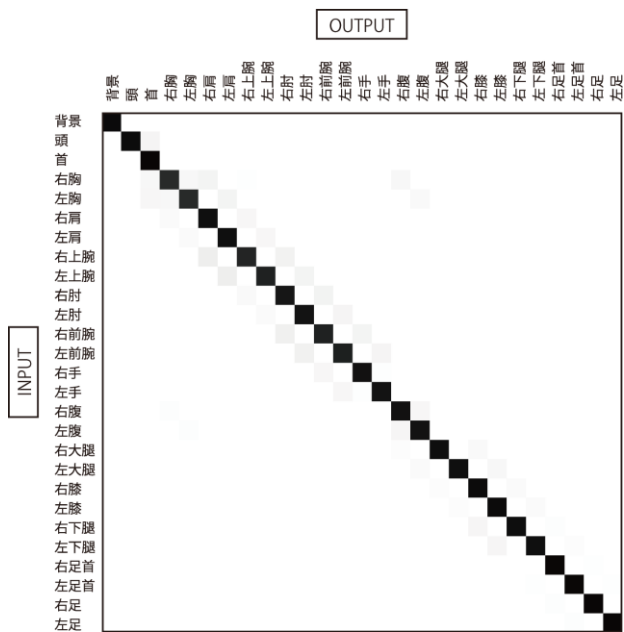
学習データセットの生成には、Autodesk MotionBuilder [11] とカーネギーメロン大学(CMU)が提供するモーションキャプチャデータ[12]を利用する。Autodesk MotionBuilder は、選択した人体モデルにモーションキャプチャデータを入力することで、人体モデルをレンダリングすることができる。本研究では、しゃがむ、笑う、携帯電話を掛ける、歩くの 4 つの動作{Bend2, Laugh, Phone, SlowWalk}を入力し、Autodesk MotionBuilder により人体距離画像と教師データを生成する。また、図 3(a)に示すように人体パーツは 26 クラスとし、背景の 1 クラスを加えた計 27 クラスを教師データとする。図 3(b)に生成したデータセットの一例を示す。学習データと評価データの枚数を表 2 に示す。学習には、5 つのグループに含まれる距離画像から半分使用し、13,293 枚用いる。評価には、5° 刻みに作成した距離画像を 38,653 枚用いる。

表 2 データセットの枚数

	Bend2	Laugh	Phone	SlowWalk	合計
学習	1,911	3,423	5,166	2,793	13,293
評価	3,660	6,520	9,840	5,340	38,653



(a) 従来法



(b) 提案手法

図5 人体パーツの識別精度

4.3 人体向きの推定実験

Regression Forest により回帰推定した人体向きの推定精度を評価する．図4に人体の向きを回帰推定した際の角度毎の平均2乗誤差と分散を示す．実験により平均 4.96° の誤差で人体の向きを推定できることを確認した．入力距離画像の角度が $\pm 25^\circ$ や $\pm 65^\circ$ の周辺で平均誤差が高いことがわかる．これは，学習に用いていない角度のためだと考えられる．しかしながら，誤差 7° 以内で推定可能である．また，入力画像が $+65^\circ$ と推定された場合，

分散が大きい確率分布となった．そのため，分散が大きい $+65^\circ$ のような確率分布を得た場合，Conditional Classification Forests では $+45^\circ$ と $+90^\circ$ の Classification Forest の重みが大きくなる．これにより，グループ化した5方向に含まれない向きの入力画像に対しては，複数の Classification Forest の識別結果を利用することで精度良い人体パーツ識別を可能とする．

4.4 人体パーツの識別実験

Conditional Classification Forests による人体パーツの識別精度を評価する．評価には，部位の識別が正解した画素の総数を TP，他の部位と間違えた画素の総数を FN として， $\text{Precision} = \text{TP} / (\text{TP} + \text{FN})$ により識別率を算出する．図5に人体パーツ識別結果のコンフュージョンマトリクスを示す．対角成分が黒であるほど，そのパーツを正しく識別できたことを表す．図5から提案手法は，従来法と比べ識別率が高いといえる．提案手法の平均識別率は 89.1% であり，従来法の 75.9% と比較して， 13.2% の精度向上を実現した．

提案手法と従来法による人体パーツの識別例を図6に示す．図6より提案手法は，入力距離画像から精度良く人体パーツを識別できていることが確認できる．

4.5 実環境における人体パーツ識別実験

実環境における実験として，TOFカメラを用いて撮影した距離画像から人体パーツ識別を行う．本研究で用いる TOF カメラは，MESA 社の SR-4000 である．SR-4000 は， $0.3\text{m} \sim 5.0\text{m}$ までの距離情報をリアルタイムで取得することができる．図7に TOF カメラにより撮影した距離画像の識別結果を示す．図7より，実環境においても人の向きに関わらず精度良く識別できていることがわかる．また，表3に提案手法を構成する Regression Forest と Conditional Classification Forests，従来法の Classification Forest の処理速度を示す．提案手法は，画像1枚につき平均 $298.8[\text{ms}]$ で人体の向き推定と人体パーツ識別を行うことができる．

表3 処理速度

	画像1枚の処理時間 [ms]
Regression Forest	11.5
Classification Forest	71.4
Conditional Classification Forest	287.3




入力	距離画像											
	角度		-100°	-85°	-60°	-45°	-15°	0°	+25°	+45°	+65°	+80°
出力	提案手法	識別結果										
		角度	-85°	-86°	-49°	-58°	-20°	-6°	+29°	+46°	+71°	+85°
	従来法	識別結果										

図 6 人体パーツの識別例


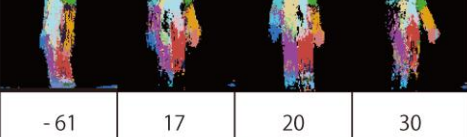
出力	入力距離画像				
	識別結果				
	角度	-61	17	20	30

図 7 実環境における人体パーツの識別例

5. まとめ

本研究では、人体の向きを考慮した **Conditional Classification Forests** を用いた人体パーツ識別手法を提案した。提案手法は、従来法である 1 つの **Classification Forest** を用いた手法と比較した場合、13.2%の精度向上を実現した。また **Regression Forest** により 4.96° の誤差で人体向きを回帰推定できることを確認した。今後は、人体パーツ識別結果を用いた動作認識手法について検討する。人の動作認識において、同じ動作を異なる方向を向いて行った場合、同様の特徴を捉えることは困難である。そのため、提案手法により人体の向きを考慮することで高精度な動作認識が期待できる。

参考文献

- [1] J.Wang, Z.Liu, Y.Wu and J.Yuan, “Mining actionlet ensemble for action recognition with depth cameras”, CVPR, 2012.
- [2] L.Xia, C.Chen and J.Agarwal, “View invariant human action recognition using histograms of 3D joints”, CVPR, 2012.
- [3] J.Taylor, J.Shotton, T.Sharp and A.Fitzgibbon, “The Vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation”, CVPR, 2012.
- [4] M. Ye, X. Wang, R. Yang, L. Ren, and M. Pollefeys, “Accurate 3D pose estimation from a single depth image”, ICCV, pp. 731-738, 2011.
- [5] Microsoft. Kinect camera. <http://www.xbox.com/en-US/kinect/default.htm>, 2010.
- [6] A. Kolb, E. Barth, R. Koch, and R. Larsen. “Time-of-Flight Sensors in Computer Graphics”, Eurographics State of the Art Reports, 2009.
- [7] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time human pose recognition in parts from single depth images”, CVPR, 2011.
- [8] L. Breiman, “Random forests”, Machine Learning, No.45(1), pp. 5-32, 2001.
- [9] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon. “Efficient regression of general-activity human poses from depth images”, ICCV, 2011.
- [10] M. Dantone, J. Gall, G. Fanelli and L. Van Gool, “Real-time facial feature detection using conditional regression forests”, CVPR, pp. 2578-2585, 2012.
- [11] Autodesk MotionBuilder. <http://www.autodesk.co.jp/>
- [12] CMU Mocap Database. <http://mocap.cs.cmu.edu/>.