

Kinect による姿勢認識を用いた食事姿勢の計測

史 宇華^{*1} 松村 耕平^{*2} Roberto Lopez-Gulliver^{*2} 野間 春生^{*2}

Abstract — 本論文では、マイクロソフト社の Kinect の深度画像から得られるユーザーのスケルトン情報に対して、Random Forest 分類器と AdaBoost を使用してユーザの姿勢をリアルタイムで自動認識するシステムを開発した。ここでは特に食事時の姿勢を対象とし、Kinect で深度データが含まれる悪い姿勢の学習データを録画する。これらの学習データから、ユーザーの Joint の特徴量を抽出し、悪い食事姿勢のデータベースを構築して学習し、五つの悪い姿勢を動画から自動認識できるシステムを開発した。悪い姿勢を食事時の映像から学習し、50名の被験者に対して行った食事の映像から93%の認識機能を実現した。本システムを応用し、将来はユーザーの悪い食事姿勢を検出した後、ユーザーにフィードバックを与え、食育トレーニングすることを目指している。

Keywords : Kinect, 姿勢認識, 食事姿勢, 食育

1 はじめに

食事マナーは豊かな人間形成に資するための食に関する基本所作として重要である。近年、様々なところでその乱れが指摘されている。細川らの研究 [1] によれば、調理実習での試食の際の学生が様々な食事マナーの問題を示し、例えば「肘をついて食べる」、「椅子の上で胡坐をかいて食べる」、「立て膝をして食べる」ような食事姿勢を日常的に取っている。平成17年に制定された食育基本法によれば、“食育を、生きる上での基本であって、知育、徳育及び体育の基礎となるべきものと位置付けるとともに、様々な経験を通して「食」に関する知識と「食」を選択する力を習得し、健全な食生活を実践することができる人間を育てる”とし、食育の重要性を謳っている。食育は従来は家庭で主に行われていたものであるが、親世代の食に関する知識や技術の不足も指摘され、さらに、家族の形態そのものが多様化しており、家庭の食卓でマナーを学ぶ機会が少なくなっていることは否めない。

また、外国を旅行する時に、事前にマナーの講習を受けることも現地での儀礼を尊重するために必要である。例えば、中国人と日本人は同じアジア人であるが、文化や習慣は違う。食事のとき、日本では皆で「いただきます」と言い、それから一緒に食事を始める。そして、食事の終わりに「ごちそうさまでした」と言う。しかし、中国ではどちらの言葉も言わない。麺類を食べているとき、中国では音を立ててはいけない。しかし、日本では音を立てて食べることが美しいと考えられている。その音には「おいしい」の意味があるからである。まだ、日本では、食器を手を持って食べること

正しいマナーであるケースがあるが、中国ではそれがきれいに見えないことがある [2,3]。

さらに悪い姿勢により様々な悪い影響もある。例えば、手に顎を載せた状態である姿勢（図1）や、ひじをついて食べた状態である姿勢では、万国共通に見た目が美しくないうえに、食べ物の消化吸収が悪くなるとも言われている [1]。

本研究では食事時における姿勢の矯正を目的としたシステムを作成した。マイクロソフト社の Kinect の深度画像から得られるユーザーのスケルトン情報に対して、Random Forest 分類器と AdaBoost を使用してユーザの姿勢を自動認識するシステムを構築する。ここでは特に食事時の姿勢を対象とし、まず Kinect で録画した姿勢学習データからユーザーの関節特徴値を抽出して悪い食事姿勢のデータベースを構築して学習し、五つの悪い姿勢をリアルタイムで自動認識できるシステムを開発した。



図1: 手に顎を載せた状態
Fig.1 Hand in jaw posture

^{*1}立命館大学大学院情報理工学研究所

^{*2}立命館大学情報理工学部

2 関連研究

手軽に3次元深度情報が得られるマイクロソフト社のKinectの登場に伴い、姿勢認識分野で現在まで様々な姿勢の認識手法が提案されている。

Suma [4] がFAAST(flexible action and articulated skeleton toolkit)を提案した。このツールキットは27種類の事前定義された姿勢を含んでいる。例えば,LEFT_ARM_UP,LEFT_

FOOT_UP など。ユーザーの骨格情報が取得される。

Kangら [5] は新しい3Dアプリケーションを制御するため手法を提案した。この手法はユーザーの距離情報と関節位置の情報を抽出して,3Dアプリケーションを制御する。

Zhangら [6] はKinectに基づいてのSVM分類システムを提案した,このシステムは高い認識率で22種類の人体姿勢を認識できる。SVMは,教師あり学習を用いるパターン認識モデルの一つであり,分類や回帰へ適用できる。このSVMによる姿勢の認識手法は姿勢学習と姿勢認識この二つ技術から構成される(図2)。

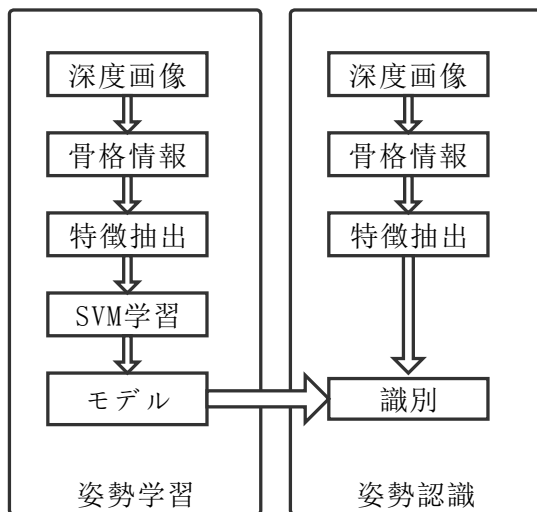


図2: SVMによる姿勢の認識手法のフローチャート
 Fig.2 Flowchart of SVM-based posture recognition

一般に,学習のためのある姿勢を定義するためには,ユーザーがKinectセンサーの前に立って,学習のための十分な画像データを生成するために対象となる姿勢を数秒保持するの必要がある。その後,画像データからスケルトン情報を得らて姿勢学習および認識ための関節情報の特徴値(座標,角度など)を抽出する。そして,抽出された特徴値がSVM分類システムに送り込まれて姿勢モデルを構築する。姿勢を認識する時は,ユーザーの姿勢を同じように処理して,SVMモデルによって

認識する。

ZhangらのSVMによる手法では,各姿勢について少なくとも100個以上の学習データを収集しなければならない。さらに,このSVMに基づく手法では腕の姿勢の認識は困難である。異なる腕の姿勢(例えば,腕を上げると腕を下げる)間の違いが小さいため,いくつかの姿勢が誤って特徴空間に分類される危険性が高い。

Kinectを利用した,現在の姿勢認識技術が非常に進歩しているが,依然としてこのような問題がある。これらの方法は全て限られた姿勢を対象とし,比較的シンプルな姿勢しか処理しておらず,複雑な姿勢の認識は対象としていない。また,本論文で対象とする食事姿勢に対して,腕部分の認識は重要であるが,従来の研究は腕の姿勢を精度高く認識できていない。本研究ではこれらの課題を解決するために,Random forestとAdaBoostを用いた識別法を提案した。

3 提案手法

第1章で述べたように,食事の悪い食事姿勢を矯正するために,自動的に悪い姿勢を検出したシステムが必要である。本研究では以下5種類の悪い姿勢を認識するためにシステムを作成した。

- 姿勢1 顎を乗せて食べる姿勢
- 姿勢2 ひじをついて食べる姿勢
- 姿勢3 背中を丸めて食べる姿勢
- 姿勢4 髪を触る姿勢(図3)
- 姿勢5 携帯を見て食べる姿勢

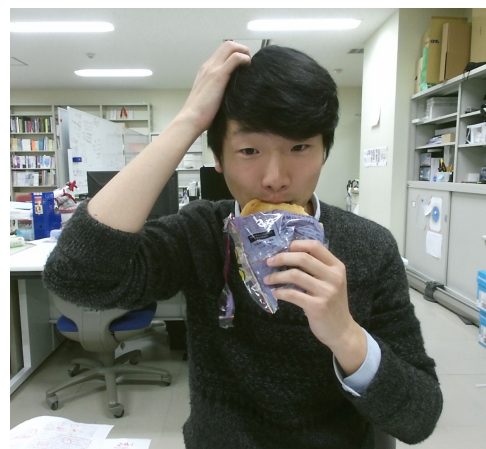


図3: 髪を触る姿勢
 Fig.3 Touching hair posture

Zhangらの姿勢認識手法で高認識率を達成するために,各姿勢が100個以上のタグ付き静止画学習デー

タ(深度情報を含めた画像)が必要である。また、食事姿勢に対する極めて重要な腕部分の姿勢は良く認識できない。そこで本研究ではSVMを代わりにRandom Forest 分類器と AdaBoost アルゴリズムを用いて、姿勢を認識する手法を採用した。提案システムの流れは以下の三つのステップである。

Step1 学習のための姿勢データの収集

Step2 姿勢データの学習

Step3 姿勢の認識

ここでは,AdaBoost アルゴリズムを用いて収集したデータを学習する。以下,姿勢データの収集と学習を個別に説明する。

3.1 学習のための姿勢データの収集

姿勢認識ためにユーザーのRGB情報と深度情報が必要なので,RGBカメラのRGBストリームと赤外線カメラの深度データを記録する。本研究では各姿勢につきサンプルの動画を学習データとして記録する。事前実験でサンプルの数は35個から40個まで決めた。各サンプルは,ある姿勢を表現している時の1名の体の深度情報とJointデータを含めた約5秒の動画である。Jointデータとはユーザーの関節の三次元座標と速度などの情報である。

本研究ではKinect Studioを用いて動作を録画した。Kinect StudioはKinect for Windows SDKに付属しており,深度データを含めた動画を録画するためのツールである。

本研究ではKinectで録画したの深度情報と関節情報を含めた動画学習データを事前処理する。まずは,各姿勢が必要ないJointの情報を除去する。例えば,本研究が対象とする姿勢は全て上半身のみの姿勢であるため,下半身のJoint情報を除去している。そして,学習データを図4のようにミラー操作する。これにより,データの量を仮想的に二倍にえる。これによって,例えば左肘をついた姿勢を記録するだけで,右肘をついた姿勢も認識対象とできる。人には左利きと右利き両種があるので,この操作は必要と考えられる。5姿勢かける2で10カテゴリーを識別している。最後はジェスチャーの時間的位置を設定する。得られた学習データの中で,対象となるジェスチャーを取っている部分の開始点と終了点を設定する。

処理した後の動画はこの提案手法の学習データである。

3.2 姿勢の学習

本研究はAdaBoostアルゴリズムを姿勢の学習のために使用する。AdaBoost(Adaptive Boosting)は,Yoav FreundとRobert Schapireによって考案された弱分

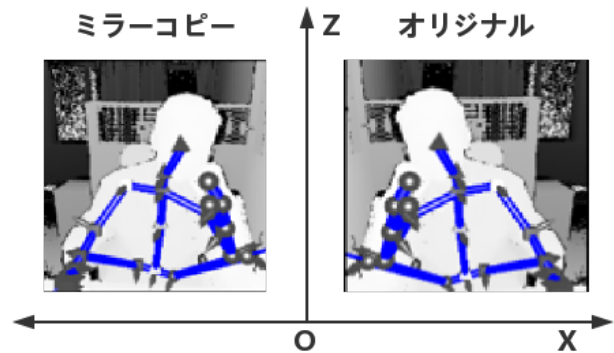


図4: ミラー操作
Fig.4 Mirroring Operation

類器を組み合わせることで強分類器を構成の機械学習アルゴリズムである[7]。

AdaBoostは学習データに対して複数個の弱学習器で順番に処理を行う。それぞれの弱学習器の認識によって重みを逐次的に更新する。最後に,重みをつけて組み合わせて強分類器をつくる。

提案した手法では弱分類器がRandom Forest(RF)分類器を選択する。Random Forestとはランダムに特徴値と組み合わせて分類器を多数作って決定木を効率よく学習させるという分類器である。SVMなどの手法に比べて,いくつかの利点がある。まずは特徴量の重み(重要度)が学習と共に計算できる。学習のための必要なデータが少なく学習のスピードが早い。本提案では関節ジョイントの38種類の特徴量(表2)をランダムで組み合わせて,弱分類器を構成した。

AdaBoostの学習ステップは:

1. 学習データの重みは1/40(40は学習データ数)で初期化する。
2. 姿勢によって複数個弱分類器を作る。対象の姿勢によるが8000個から12000個作った。
3. 一番目の弱分類器を作って,弱分類器の認識率を求める。
4. 認識率から弱認識器の確信度を求める。認識率が高いなら確信度が小さくなるようにする。
5. 学習データの重みを更新して,重みの和が1になるように正規化する。
6. ステップ3に戻って,次の弱分類器を作る。全ての弱分類器を作ったら終了する。
7. 最終は,全ての弱分類器を確信度によって強分類器を作る。

表 1: 特徴量
Table 1 Features

Label	Label
<i>DiffPositionX</i>	<i>DiffMuscleForceX</i>
<i>DiffPositionY</i>	<i>DiffMuscleForceY</i>
<i>DiffPositionZ</i>	<i>DiffMuscleForceZ</i>
<i>Angels</i>	<i>VelocityX</i> \wedge 2
<i>TmeSpaceAngles</i>	<i>VelocityY</i> \wedge 2
<i>Speed</i>	<i>VelocityZ</i> \wedge 2
<i>Speed</i> \wedge 2	<i>Acceleration</i>
<i>VelocityX</i>	<i>AccelerationX</i>
<i>VelocityY</i>	<i>AccelerationY</i>
<i>VelocityZ</i>	<i>AccelerationZ</i>
<i>AngleVelocity</i>	<i>BoneLengthChanges</i>
<i>AngleAcceleration</i>	<i>HandValueRaw</i>
<i>MuscleForceX</i>	<i>handValueMultiClass</i>
<i>MuscleForceY</i>	<i>HandDifferenceRawBest</i>
<i>MuscleForceZ</i>	<i>HandDifferenceMultiClass</i>
<i>Musclepower</i>	<i>HandDifferenceRaw</i>
<i>MuscleTorequeX</i>	<i>HandDifferenceMultiClass</i>
<i>MuscleTorequeY</i>	<i>RefinementWrist</i>
<i>MuscleTorequeZ</i>	<i>RefinementHand</i>

AdaBoost アルゴリズムでジョイントの位置, 角度などの情報を特徴値として, 姿勢データの分類器を生成し, 姿勢データベースを構成した. データベースには五つの悪い姿勢の学習データと分類器を格納する. 学習されたモデルを利用して, リアルタイムで自動姿勢認識を実現する.

4 システム実装

4.1 システム構成

提案システムは, ユーザの姿勢を認識する「姿勢認識部」, ユーザの姿勢状況を通知する「通知部」で構成される. 姿勢認識部では, ユーザの姿勢状態を Kinect で深度情報を計測し, ジョイント情報に変換した後に, 学習した姿勢モデルを用いて, ユーザの姿勢を判定する. 通知部では, システムが, 悪い姿勢を検知したときに, その結果をユーザに警告として通知する. 以下, システムの実装の詳細を分別に説明する.

実装環境を表 1 に示す. 主要なデバイスは Kinect v2 である.

表 2: 実装環境
Table 2 Environment of system

OS	Windows 8.1
機材	Mac Mini, Kinect v2
開発言語	C#

システムの全体を図 5 に示す. 40 個の悪い姿勢の動画を教師データとして姿勢のデータベースを作成する. ユーザーが使う時にリアルタイムにユーザーの悪い姿勢を識別する. ユーザーの悪い食事姿勢を検出し

た後, ユーザーにフィードバックを与える.

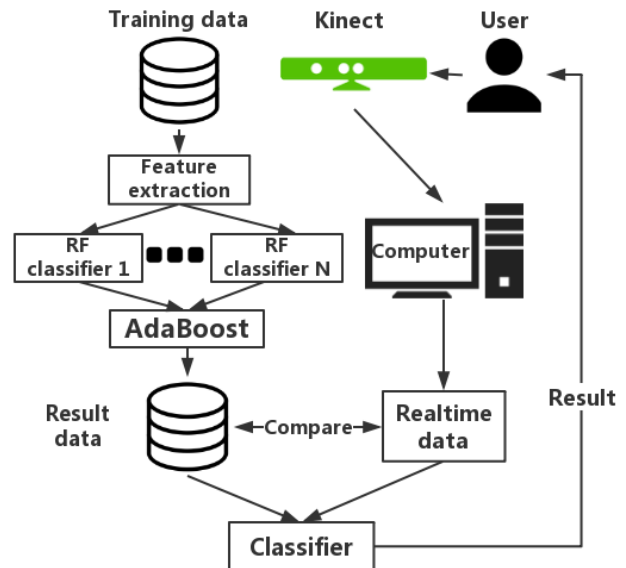


図 5: システムの全体
Fig. 5 Overview of system

4.2 姿勢の認識

プログラムを実行する時は, まず姿勢データベースから姿勢モデルを読み込んで, 使用する姿勢データをデータベースから取り出す. その後, ユーザーの関節情報を取得して, 姿勢を自動的に識別する. もし悪い姿勢があれば, ユーザーに警告を与える.

4.3 通知部

通知部はシステムがユーザーの悪い姿勢を認識した際に, ユーザーにフィードバックを提示する. 現状の通知手法は悪い姿勢を検出した時に, ユーザーに警告を与える. 図 6 に示すように悪い姿勢を検出した時, その悪い姿勢に対応な警告をユーザーに提示する. 将来は, この手法でユーザーのマナーをトレーニングすることを想定している.

5 評価実験

提案手法の有効性を確認し, さらに食事時の悪い姿勢を改善するため識別性能の評価実験を行った. 被験者は, 大学 4 年生の生徒 35 名, および 3 年生 10 名と 2 年生 5 名である. 被験者の内訳は, 男子生徒 40 名と女子生徒が 10 名である. 実験では, 各被験者に事前に特に説明をせず, Kinect の前で自由に食事をしてもらい, その間の姿勢を計測した. 提案手法の認識率を評価するため, 全員に同じ料理を提供して, 一人ずつ 10 分食事した. 計測した生徒 50 名食事動作データの食事動作をテストデータとし, 提案手法によって悪い姿勢の検出を行った.

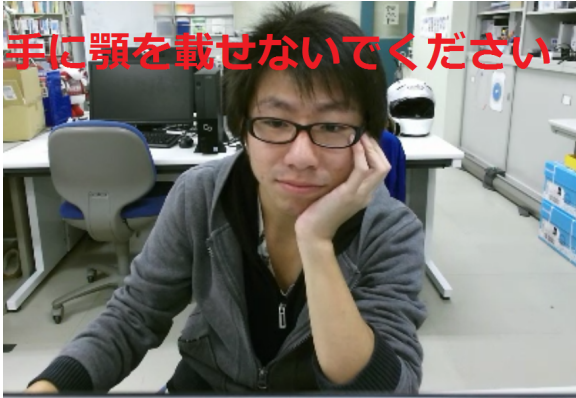


図 6: 画面の上に警告を提示する
Fig.6 Give a warning

食事時間に実験者が主観的に悪い姿勢の種類と回数を記録した。同時に、システムが自動認識した悪い姿勢も記録した。この二つの記録を比較して、認識率を評価した結果を表 3 と表 4 と表 5 に示す。表では実験者による悪い姿勢の観測した回数、システムが悪い姿勢と自動認識した回数、システムが検出できなかった姿勢の回数をそれぞれ示している。また、表 4 では 3 章で示した五つの姿勢を姿勢 1-5 と略称する。

表 3: 評価実験の結果 (性別)
Table 3 Experimental result(Gender)

	人数	主観観測数 (回)	自動認識数 (回)	識別率 (%)
男性	40	25	24	96%
女性	10	6	5	83%
総計	50	31	29	93%

表 4: 評価実験の結果 (年齢)
Table 4 Experimental result(Age)

	人数	主観観測数 (回)	自動認識数 (回)	識別率 (%)
20 歳以下	15	9	9	100%
20 歳以上	35	22	20	90%
総計	50	31	29	93%

以下は実験の三つの例を説明する。

被験者 2 号 (図 7) は食事中に右手に顎を載せる姿勢を表現した時、システムがその姿勢を検出した。

表 5: 姿勢認識の結果
Table 5 Result of posture recognition

姿勢	主観観測数 (回)	自動認識数 (回)	識別率 (%)
姿勢 1	5	5	100%
姿勢 2	6	6	100%
姿勢 3	7	6	85%
姿勢 4	8	8	100%
姿勢 5	5	4	80%
総計	31	29	93%



図 7: 被験者 2 号 : 左手に顎を載せる姿勢を検出
Fig.7 Subject No.2:Hand in jaw posture detection

被験者 5 号 (図 8) は食事に髪を触る姿勢を表現した時、システムがその姿勢を検出した。



図 8: 被験者 5 号 : 髪を触る姿勢を検出
Fig.8 Subject No.5:Touching hair posture detection

被験者 17 号 (図 9) は食事にひじをついて食べる姿勢を表現した時、システムがその姿勢を検出した。



図 9: 被験者 17 号 : ひじをついて食べる姿勢を検出
Fig.9 Subject No.17:Elbows on the table posture detection

以上の結果をまとめると識別率で 93 % の結果を得た。学習を使用した学習データは全て男性の姿勢だけであるが、今回の被験者の女性の姿勢の識別率が男性の識別に比べて低下した。これは女性被験者数が少ないことが原因で、結果的に識別率は男性より低下したと考える。また姿勢ごとには、姿勢 1,2,4 の認識率は非常に高いが、姿勢 3,5 の識別率は低下している。これは姿勢 3, 5 学習時の学習データ数 (35 個) が姿勢 1,2,4 の学習データ数 (40 個) より少なかったことが原因と考えられるため、この二つの姿勢の学習データを追加し、追加学習してすることで、識別率を向上できると考えられる。

以上の分析より、食事の悪い姿勢を検出する目的においては、Random Forest 分類器と Adaboost に基づいた学習手法は有効であるといえる。

6 おわりに

本研究では食育の観点から、食事における癖の矯正を目的としたシステムについての研究を行った。特に食事の姿勢を対象として、Random Forest 分類器と Adaboost を利用して、五つの悪い姿勢を学習して 50 名の被験者に対して 93 % の認識機能を実現した。

この自動認識システムを応用し、悪い姿勢が検出されるとその悪い癖を改善するための手段としてもちいる。今後はユーザーへ通知の部分が悪い姿勢の発生の仕方によって通知を分けるよう通知システムを改善する。

参考文献

- [1] 細川裕子 and 福田綾子, “短大生の食事マナーに対する意識と実態—食教育における一考察として—,” 白大学短期大学部研究紀要, vol. 50, pp. 31–42, 2014.

- [2] 木村春子, “食事マナーの変遷—日本と中国 (特集 食事マナーの変遷と比較),” ヴェスタ, no. 41, pp. 16–21, 2001.
- [3] 陳弘美, “中式, 西式, 日式餐礼用知,” 2005.
- [4] E. A. Suma, B. Lange, A. S. Rizzo, D. M. Krum, and M. Bolas, “Faast: The flexible action and articulated skeleton toolkit,” in *Virtual Reality Conference (VR), 2011 IEEE*, pp. 247–248, IEEE, 2011.
- [5] J.-w. Kang, D.-j. Seo, and D.-s. Jung, “A study on the control method of 3-dimensional space application using kinect system,” *International Journal of Computer Science and Network Security*, vol. 11, no. 9, pp. 55–59, 2011.
- [6] Z. Zhang, Y. Liu, A. Li, and M. Wang, “A novel method for user-defined human posture recognition using kinect,” in *Image and Signal Processing (CISP), 2014 7th International Congress on*, pp. 736–740, IEEE, 2014.
- [7] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” 1995.