

天井設置型深度カメラを用いたテーブルトップシステム 利用者の個人認証手法の提案

増田 彬^{1,a)} 前川 卓也^{1,b)} 浪岡 保男^{2,c)}

概要: 日常的に用いられる家具の一つであるテーブルに情報を提示して、ユーザの作業をインタラクティブに支援するテーブルトップシステムが教育や商業分野などにおいて普及しつつある。複数人で利用されることが多いテーブルトップシステムにおいて各ユーザにパーソナライズされた情報を提示するため、テーブルトップシステムのための個人認証技術に関する研究が広く行われている。多くの既存研究では全てのユーザがデバイスを装着する必要があったり、設置したカメラがユーザの行動を妨げるなどの課題が存在した。そこで本研究では天井に設置した深度カメラを用いてテーブルトップシステムを操作するユーザを認証する手法を提案する。提案手法では天井の視点から取得した深度画像のみを用いてユーザの肩幅、頭の形状、背中姿勢といったソフトバイオメトリクス情報を抽出して個人認証に用いる。抽出したソフトバイオメトリクス情報はそれぞれ異なる数値の分布を持つため、そのようなデータの扱いに長けた Multiview Learning を適用することでより高精度な個人認証を実現する。天井から個人認証を行う手法はユーザを正面から認証する従来の手法と比べて、天井から複数人を同時に捕捉できる点や人体や障害物に遮られることがない点、天井に設置された照明と同じ電源から容易に電気を供給できる点、そしてユーザの作業の邪魔にならないといった利点がある。本稿では天井に設置した深度カメラを用いた個人認証手法を提案し、その性能を検証した。

1. はじめに

日常的な生活の中でテーブルは勉強や食事、ディスカッションなど様々な作業の場として頻りに用いられている。また近年の情報通信技術の進展に伴い、マルチタッチディスプレイの大画面化や安価なセンサを用いて人のジェスチャを認識するインタラクティブプロジェクトの実現により、テーブルに情報を提示してユーザの作業を支援するテーブルトップシステムがオフィスや商業施設、教育機関などで普及しつつある。テーブルトップシステムは、複数人の利用者による協調作業に用いられることが多く、各利用者にパーソナライズされた情報を提示するためには個人を識別する手法が必要となる。そのため、いくつかの既存研究では、RFID タグを組み込んだ指輪を利用者が装着して認証する手法 [1] のようにデバイスを装着して個人認証

する手法が提案されている。しかし、これらの手法はテーブルトップシステムを日常的に用いるユーザにとってデバイスの装着コストやデバイスの電池の維持コストなどがかかる課題が存在した。ユーザの負担にならない個人認証手法として、顔認証といったビジョンを用いた認証手法が挙げられるが、これらの手法をテーブルトップシステムに適用するには以下のような課題がある。

- 認証するユーザの正面にカメラを設置する必要があり、テーブル上にカメラを設置した場合はテーブルで作業を行うユーザの行動の妨げになる。ユーザの妨げにならないように壁面に設置する場合は、壁に取り付け器具を設置するコストがかかる。

- 複数人で利用されるテーブルトップシステムの場合は、壁面に取り付けられたカメラとユーザの間に他のユーザが映り込む場合がある。またユーザは様々な方向を向いて作業等を行うため、一つのカメラのみですべてのユーザの正面を捉えることは困難である。

- 壁やテーブルに取り付けられたカメラに電源を供給するためには電源の延長ケーブルなどを壁や床に這わせる必要がある、設置に手間がかかる。

これらの課題を解決するため、本研究では深度カメラを図 1 のように天井に設置することを想定する。天井に深度

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University

² 株式会社 東芝 生産技術センター 設計生産システム変革推進部
Design & Manufacturing Innovation Department, Corporate Manufacturing Engineering Center, Toshiba Corporation

a) masuda.akira@ist.osaka-u.ac.jp

b) maekawa@ist.osaka-u.ac.jp

c) yasuo.namioka@toshiba.co.jp

カメラを設置する方法は以下のような利点がある。

- 天井に設置するため、テーブル上に設置したカメラと異なり、テーブルトップシステム利用者の行動を妨げない。
- ユーザとカメラの間に障害物が入り込みにくく、より確実にユーザを捕捉できる。また複数人のユーザを同時に撮影することができる。
- 天井に設置された電灯やプロジェクタの近くに深度カメラを設置することでそれらと同じ電源から容易に電気を供給することができる。

一方、天井からの視点では個人認証に有用な顔や指紋のような特徴を取得するのが困難になる問題がある。そこで本研究では天井に設置した深度カメラで取得可能な利用者の肩幅や頭の形状、姿勢の癖といったソフトバイオメトリクス情報を深度画像から抽出して組み合わせることで、テーブルトップシステム利用者の行動を阻害しないで高精度に個人を認証する手法の実現を目指す。

本稿では住宅やオフィス、研究室などに導入されたテーブルトップシステムの真上の天井に深度カメラが設置されている環境を想定する。利用者がテーブルで Web ブラウジングや食事を始めると、提案手法によりユーザの行動を阻害することなく深度画像から抽出されたソフトバイオメトリクス情報を用いて自動的に個人認証を行う。本研究では天井に設置された深度センサから取得できる以下のソフトバイオメトリクス情報に着目した。

- 骨格情報：深度カメラから取得した深度画像を用いて骨格認識を行い、肩や肘、手などの関節の位置を検出する。そして関節間の距離を求めることで例えば肩幅や上腕の長さといった個人を特定するのに役立つ骨格に関する特徴を抽出する。
- 部位の形状：深度データから頭や肩といった身体の部位を推定し、その形状を取得する。頭の形状は天井からの視点では容易に取得することができ、ユーザによって大きさや形が異なるため、個人の特정에役立つと考えられる。
- 部位の面積：肩や背中面積を取得することでユーザの身体の大きさを見積もることができるため、個人の特정에有用であると考えられる。
- 姿勢の癖：本研究で想定している環境ではユーザがテーブル上で食事などの行動を行う。行動の際の背中の曲がり具合などの姿勢の癖は個人の特정에役立つと考えられる。

ただし、これらの情報にはそれぞれ欠点が存在し、例えば頭の形状は髪型によって変化したり、部位の体積は服の種類によって変化する。そこで本研究では Multiview Learning を用いてそれらの特徴を組み合わせることでそれぞれの欠点を補い、これらのノイズに頑健な個人認証手法を目指す。

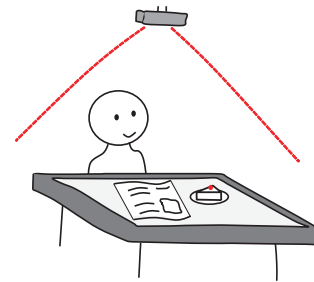


図 1 想定するテーブルトップシステムの利用環境

2. 関連研究

2.1 日常生活やオフィス環境におけるセンシング

日用品に組み込まれた RFID タグやウェアラブルデバイスを用いて、日常行動の補助や記録を行うための行動認識に関する研究が数多く行われるようになってきている。机上行動認識の研究も多数行われ、例えばテーブル上で行われた食事や食事準備、工作など様々な作業をテーブルを見下ろすカメラを用いて認識する研究も行われている [2], [3]。このような日常的環境下でパーソナライズされたサービスを提供するためには設置や運用が容易な個人認証手法が求められている。

2.2 デバイスフリー個人認証

個人認証手法として RFID による認証や慣性センサを用いたジェスチャ認証などのデバイスを用いた認証手法 [1], [4] が研究されているが、利用者がデバイスを常に持ち歩く必要があるなどの課題があった。そこで利用者の負担を減らすため、デバイスを必要としない個人認証手法が研究・開発されている。例えば、Schmidt らはテーブル内に組み込んだカメラで利用者の手の形を撮影し、指の長さなどのソフトバイオメトリクス情報を抽出して個人認証を行った [5]。Schmidt らの手法は高い精度の個人認証を実現しているが、テーブル内に手を映すためのカメラを組み込んだり、テーブルに手をかざすという特定の動作をユーザに要求するといった利便性の課題があった。

特別な動作を必要とせずユーザの行動を阻害しない認証手法として、顔や骨格を用いた認証手法が挙げられる [6] が、複数人が同時に利用するテーブルトップシステムでは、他のユーザが認証に用いるカメラの妨げになる可能性がある。そこで [7] では他のユーザに遮られることのないテーブルの下から利用者の靴を RGB-D カメラで撮影して、靴の画像で認証を行う手法を提案した。

天井からソフトバイオメトリクス情報を用いて認証を行う手法として Augsten らの研究 [8] が挙げられる。[8] では、ドアの上部に設置した超音波距離センサを用いてユーザの身長を取得して認証を行うのに対して、本研究ではいくつかのソフトバイオメトリクス情報を Multiview Learning で学習することにより、テーブルトップシステム環境下で

ユーザの姿勢に依存しない個人認証手法を実現する。

3. 提案手法

3.1 概要

提案手法ではテーブルトップシステムを利用する各ユーザからあらかじめ学習データを取得して個人認証モデルを学習する。想定する利用環境は図1に示している。個人認証モデルを準備した後、新たにユーザ（これをテストユーザとする）がテーブルトップシステムを利用し始めると深度データの変化を検出して深度画像の記録と個人認証を開始する。時系列的に得られる深度画像ごとに個人認証を行い、深度画像の記録開始から N_{test} 枚の画像の認証結果の多数決を用いて最終的な個人認証の結果を出力する。

提案手法の概要を図2に示す。まずランダムフォレストを用いた身体の部位判定アルゴリズムにより、深度画像からユーザの各身体部位に対応するピクセルを推定する。推定した身体の部位から1節で示したソフトバイオメトリクス情報を抽出し、Multiple Kernel Learning(MKL)を用いてユーザの個人認証を行う。



図2 提案手法の概要

3.2 身体の部位判定

提案手法では身体の部位判定アルゴリズムを用いて、図3のように深度画像に含まれる各ピクセルを身体の部位あるいは背景のいずれかに分類する。Microsoft社が提供しているKinect APIではユーザを正面から撮影したときのみ部位の判定が可能であるため、本研究では独自に天井か撮影した画像の部位判定を行うモデルを学習する必要がある。部位判定アルゴリズムは基本的にKinect APIで用いられている手法[9]と同じであるが、机（上から撮影する深度センサ）に対して任意の方向で座るユーザを認識するためには回転不変である必要がある。

部位判定は各ピクセルごとに特徴量を算出してそれらを結合することで特徴ベクトルを作成する。その後、特徴ベクトルを身体の部位クラスに分類する。本研究では特徴ベクトルを頭、左手、右手、左腕、右腕、左肘、右肘、左肩、右肩、その他の部位、背景の11クラスに分類する。

3.2.1 特徴ベクトルの抽出

部位判定は深度画像に含まれる各ピクセル周辺の形状情報を用いて行う。形状情報は[9]と同様に着目したピクセルとその周辺のピクセルとの深度の差を計算してオフセット特徴とする。画像 I において着目したピクセルを x とするとその特徴の値は以下のようにして求められる。

$$f_{\mathbf{u}}(I, x) = d_I\left(x + \frac{\mathbf{u}}{d_I(x)}\right) - d_I(x)$$

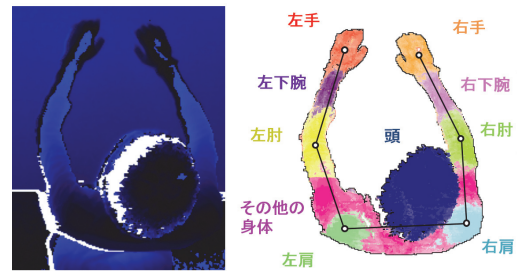


図3 深度画像と身体部位判定結果の例。

ここで $d_I(x)$ は画像 I におけるピクセル x の深度の値、 \mathbf{u} は x を始点としたオフセットベクトルである。オフセットベクトルに $\frac{1}{d_I(x)}$ を乗算することで、この特徴は深度不変となる。本研究ではオフセットベクトルをあらかじめランダムに500個生成して特徴を計算する。

3.2.2 ラベル付けされた深度画像の事前準備

特徴ベクトルを学習する際、その特徴ベクトルがどの部位であるかを教師データとして与える必要がある。深度画像に対して手で部位のラベル付けを行うことは非常にコストがかかるため、本研究では深度画像と同時に撮影したRGB画像を用いて各部位を自動的にラベリングした。具体的には被験者の身体に特徴的な色のマーカを装着してもらい、RGB画像から各部位に対応するピクセルを検出したあと、同時に撮影された深度画像ピクセルを求めた。また、回転不変の部位判別モデルを作成するため、深度画像を z 軸中心にランダムに回転させて学習を行う。認識時には事前に学習した部位判別モデルを用いるため、マーカの装着は必要としない。

3.2.3 ランダムフォレストによる部位判別

高速で頑健な部位判別を行うため、本研究では部位判別モデルとしてランダムフォレストを用いた。ランダムフォレストではランダムに選択された特徴ベクトルおよびその属性の部分集合を用いて、決定木が学習される。ランダムな選択・学習を T 回行うことで T 個の決定木が作成される。

特徴ベクトル x' に対する部位判定は各決定木の多数決によって決まる。本研究では x' がどの部位 C_n に分類されるかを

$$p(C_n|x') = \frac{n(x', C_n)}{T}$$

によって求める。ここで $n(x', C_n)$ は x' を C_n と分類した決定木の数である。

3.2.4 骨格認識

深度画像には複数人のユーザが含まれる可能性があるため、複数人の骨格を同時に認識する必要がある。そこで本研究ではまず深度画像から頭に分類されるピクセルを求め、x-means法を用いてクラスタリングする。クラスタのうち閾値よりも広い面積を持つものにユニークなIDを割り当て、そのクラスタが1人のユーザに対応するものとす

る。その後、頭のクラスタに最も近い左肩、右肩、次に左肘、右肘に順次最近傍の隣接クラスタと同じ ID を割り当てていき、判定された全ての部位に ID が割り当てられるようにする。

ID ごとに求められたそれぞれの部位において重心を求め、それを各部位の関節とする。これらの関節を結ぶ直線から肩幅や腕の長さなど個人を特定するのに役立つ特徴を抽出することができる。図 3 に一人のユーザにおいて骨格認識を行った結果の例を示している。

3.3 ソフトバイオメトリクス情報の抽出と距離計算

認証時には時系列の深度画像を用いて認証を行うため、時系列順にそれぞれの深度画像からソフトバイオメトリクス情報を抽出する。抽出されたソフトバイオメトリクス情報には深度画像の測定誤差や部位判定時の推定誤差が含まれるため移動平均法を用いて外れ値を持った深度画像を取り除く。例えば、図 4 では左から 3 番目のテスト画像から抽出されたソフトバイオメトリクス情報に外れ値が含まれている（肩幅が 0.17[m]）ため、認識には用いない。同様にユーザ 1 の学習用画像も最初の画像において外れ値が含まれている（左上腕の長さが 0.06[m]）ため、学習に用いない。

外れ値を除いた後、ランダムに選んだテスト画像と学習用画像からそれぞれ抽出されたソフトバイオメトリクス間の距離を計算し、それらを連結して距離ベクトルを作成する。提案手法で抽出したソフトバイオメトリクス情報の詳細とそれぞれの情報における距離計算の方法を以下に示す。

1. 骨格情報

腕や肩などの部位の長さはユーザの身体の大きさを表す指標の一つとなる。肩幅や上腕の長さ、下腕の長さなどは部位判定によって求められた関節間の距離から求めることが

できる。骨格情報の距離計算は値の差の絶対値によって求める。例えばテスト画像から求めた肩幅が 0.34[m] でランダムに選ばれた学習時の画像から求められた肩幅が 0.38[m] である場合、それらの距離は $|0.34 - 0.38| = 0.04$ となる。

2. 部位の形状

部位判定の結果から頭や肩などの部位に対応する 3 次元点群を求めることができるため、テストユーザの部位（頭、肩）の形状と登録されているユーザの部位の形状を比較することができる。二つの点群の距離は Iterative Closest Point(ICP)[10] によって求められる。ICP は点群間の位置合わせを行う代表的なアルゴリズムの一つであり、二つの点群間の距離が最も小さくなるように一方の点群を移動・回転する。二つの点群間の距離は対応点（最も距離の近い二点）同士の距離の平均によって求められる。提案手法ではテストユーザの部位の点群と登録されているユーザの部位の点群を ICP で位置合わせしてその距離を部位の形状の距離と定義する。同一ユーザの頭の形状を ICP で位置合わせした結果と異なるユーザの頭の形状を位置合わせした結果を図 5 に示す。同一ユーザの場合は頭の輪郭や大きさが類似しているため二点群間の距離が小さいのに比べて、別ユーザの形状を比較した場合は点群の位置合わせに失敗していることが分かる。

ここで深度カメラの測定方法による制約から深度画像に欠損値が含まれることがある。例えば図 3 において深度画像が白くなっている箇所（頭の左側など）が欠損値であり、頭の形状が正しく計測されていないことが分かる。正しく計測された場合とそうでない場合の ICP の距離は大きくなってしまいうため、個人認証の精度を低下させる要因となる。そのため、提案手法では、テストユーザ（もしくは登録されているユーザ）の時系列データに階層的クラスタリングを適用して外れ値検出を行うことで欠損値が多い深度データを除く。また、ICP は計算コストが大きく、本研究で扱ったユーザの頭の点群を並列処理しても 1 回の比較で 0.05 秒ほどかかってしまう。そこで認証の際にはテスト用の深度画像において最初の N_{icp} 枚のみの点群における距離を求め、 $N_{icp} + 1$ 枚目の画像からは過去に求めた ICP の値をランダムに選択して用いる。

3. 部位の面積

部位の表面積もユーザの身体の大きさを示す指標の一つである。提案手法では 3 次元空間の点群から GreedyTriangle 法を用いて肩や胴体の表面積を計算する。

4. 姿勢の癖

Web ブラウジングなどの作業を行っているときのユーザの姿勢も個人を特定するのに役立つと考えられる。事前検証により背中曲がり具合が人の姿勢をよく反映することが分かった。二人の異なる被験者の上半身の点群の例を図 6 に示す。図 6 において左の被験者は背中曲がり具合が小さく姿勢が良いのに比べて、右の被験者は背中が大きく

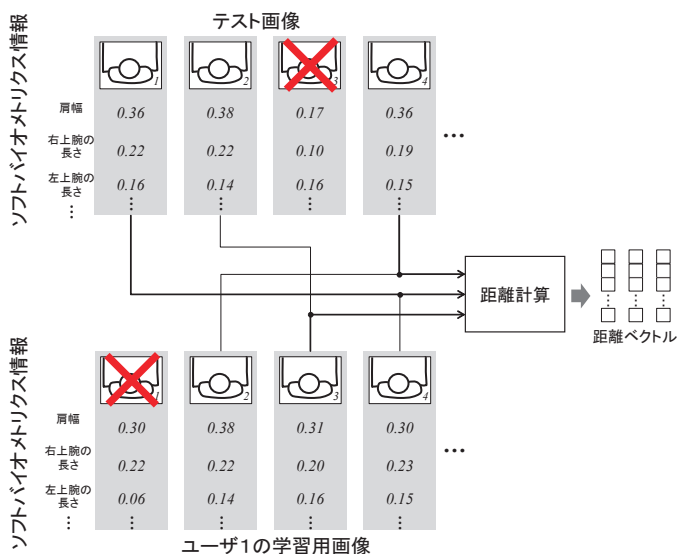


図 4 2 ユーザ間におけるソフトバイオメトリクス情報の距離計算

曲がっており猫背になっている。天井に設置した深度カメラでは座っている椅子の背もたれやユーザの位置によって背中を捉えることが困難な場合があるため、本研究では首や肩の角度を求めて姿勢の癖として用いる。首の角度は部位判定で求めた両肩の中心点（これを首の付け根とする）と頭の頂点を結ぶベクトルと地面との角度を計算することで求められる。首は背骨と直接つながっているため、背中の曲がり具合をある程度反映するものと考えられる。肩の角度は両肩を結ぶベクトルと地面との角度から求める。首と肩の角度の距離は骨格情報と同様に差の絶対値によって求める。ここで、ユーザの姿勢は行動の最中に変化する。例えば食事において、器を持っている時の姿勢と箸を伸ばしている時の姿勢は異なる。そこで本研究では最も頻りに観測された姿勢を骨格情報と同様に移動平均法を用いて求め、それ以外は外れ値とする。

3.4 Multiple Kernel Learning を用いた分類

3.4.1 概要

提案手法ではテストユーザのソフトバイオメトリクス情報を登録されたユーザのものとそれぞれ比較して距離ベクトルを求める。距離ベクトルは図4で示したように各ソフトバイオメトリクス情報の距離を結合して構築され、登録されたユーザごとに学習した2クラス Support Vector Machine(SVM) に入力される。入力した距離ベクトルとSVMの超平面から距離を求めることができ、その値からSVMを学習したユーザへ分類される確率が求められる。図7に距離ベクトルから登録済みのユーザへ分類する手順の概要を示す。これによりある時刻の深度画像におけるあるユーザへの分類確率が求められるが、最終的にどのユーザとして認証するかは時系列データ全ての分類確率を求めてその総和により決定する。

ユーザ1の認証用SVMの学習に用いる距離ベクトルを求めるとき、ユーザ1の深度画像群の中で互いに異なる画像同士で距離を求め、これを正例の距離ベクトルとする。ユーザ1と異なるユーザの場合もランダムなペアで距離を

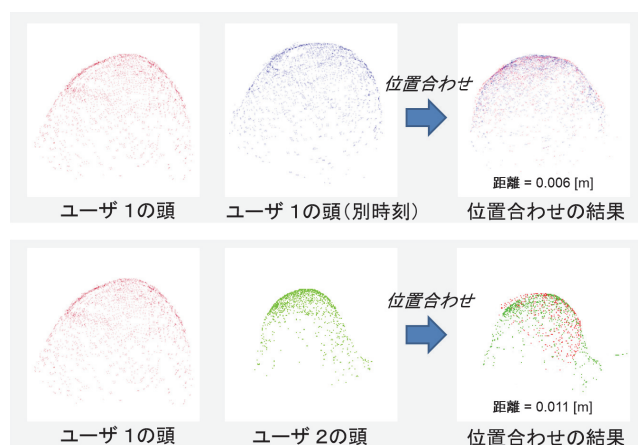


図5 2ユーザ間における頭の位置合わせの結果の例

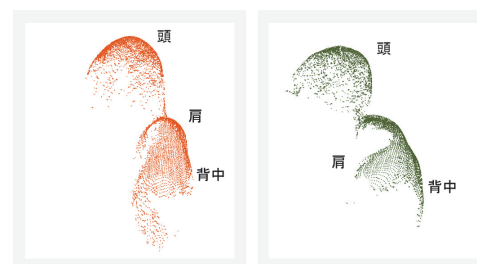


図6 異なるユーザの上半身の点群

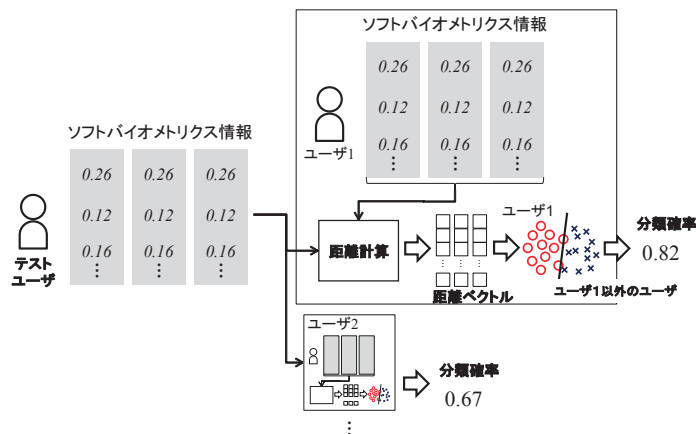


図7 登録したユーザごとに学習した2クラスSVMを用いた認証の概要

求め、これを負例の距離ベクトルとする。ユーザ1と異なるユーザは複数人存在するため、負例の数が正例の数よりも多い不均衡データになる。そこで提案手法では正例と同数の負例距離ベクトルの集合を学習用データからランダムに複数回選択してSVMを複数個学習する。そして、学習したSVMのうち学習用データに対して分類精度が最も高いものを選んで認識に用いる。

距離ベクトルはいくつかの異なるソフトバイオメトリクス情報によって構成されており、例えば骨格情報の中でも肩幅だけでなく腕の長さや首の長さの距離が含まれている。このように異なる数値の分布を持つ情報によって構成された距離ベクトルをより高精度に分類するため、提案手法ではMultiview Learningの一つであるMultiple Kernel Learning(MKL)を2クラスSVMに用いる。MKLはいくつかの異なる数値の分布に対応した基底カーネルの線形結合によって構成されるため、提案手法では骨格情報や形状情報に応じた基底カーネルを用意して結合する。

3.4.2 基底カーネルの線形結合と個人認証

カーネル関数は特徴空間においてインスタンス間の距離を求める際に用いられる。 N 個の学習インスタンス $\{x_i \in \mathcal{X}\}_{i=1}^N$ があるとき、テストインスタンス x_* の推定を行う決定関数は

$$f(x_*) = \mathbf{a}^T \mathbf{k}_* + b,$$

と表される。ここで \mathbf{a} は各学習インスタンスに対する重みであり、 b はバイアスである。また $\mathbf{k}_* =$

$[k(x_1, \mathbf{x}_*) \dots k(x_N, \mathbf{x}_*)]^T$ であり, $k(\cdot, \cdot)$ は二つのインスタンス間の距離を求めるカーネル関数である.

基底カーネルを線形結合して構成される MKL において, sk, sh, d, h がそれぞれ骨格情報, 部位の形状, 部位の表面積, 姿勢の癖とすると決定関数は

$$f(\mathbf{x}_*) = \mathbf{a}^T (e_{sk} \mathbf{k}_{sk,*} + e_{sh} \mathbf{k}_{sh,*} + e_d \mathbf{k}_{d,*} + e_h \mathbf{k}_{h,*}) + b, (1)$$

となる. ここで e_m は m つ目のカーネルの重みであり, $m \in \{sk, sh, d, h\}$ である. また $\mathbf{k}_{m,*}$ は $[k_m(x_1, \mathbf{x}_*) \dots k_m(x_N, \mathbf{x}_*)]^T$ である. 実際の実装ではガウシアンカーネル, シグモイドカーネル, 多項式カーネルを各ソフトバイオメトリクス情報 m ごとに用意した. 異なるカーネルを用いることで異なる数値分布を持つ特徴同士の非線形な関係を学習させることができる. 式 1 の決定関数に基づき, 提案手法では 2 クラス SVM を登録したユーザごとに作成し, その分類結果は $y = \text{sign}(f(\mathbf{x}_*))$, $y \in \{-1, +1\}$ のように求められる. $f(\mathbf{x}_*)$ が $+1$ に近い値をもつとその SVM に対応したユーザである確率が高いとし, -1 に近いほど SVM に対応したユーザとは異なることを示す.

3.4.3 パラメータ推定

式 1 のパラメータを推定するために本研究では Bayesian Efficient Multiple Kernel Learning (BEMKL) を用いた. 図 8 に二値分類器のグラフィカルモデルを示す. ここで, e_{sk}, e_{sh}, e_d と e_h はまとめて e としている. \mathbf{K}_m は $N \times N$ のカーネル行列で二つのインスタンス間の距離を定義する. 補助変数のベクトルは \mathbf{f} で表され, 中間出力の行列は \mathbf{G} で表される. ハイパパラメータ λ, γ と ω はそれぞれ \mathbf{a}, b と e の事前確率を決定する. パラメータ推定の詳細は [11] を参照されたい.

4. 評価

4.1 実験データ

実験に用いるデータはテーブルから約 1.5[m] 離れた天井付近に設置した Microsoft 社の Kinect V2 を用いて, 2Hz のサンプリングレートで取得した. テーブルの席に座った被験者は「テーブルトップディスプレイを用いた Web ブラウジング」と「テーブルでの食事」の二つの行動をそれぞれ 140 秒ほど行った. Web ブラウジングに関しては, 被験者にテーブルトップディスプレイに表示した Web ブラ

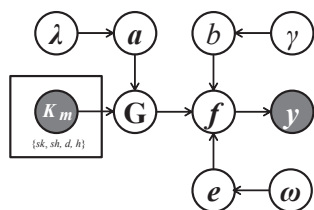


図 8 2 クラス分類に用いる MKL のグラフィカルモデル

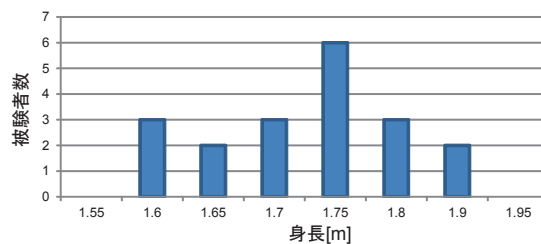


図 9 被験者の身長分布

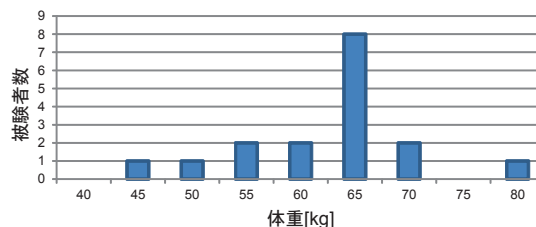


図 10 被験者の体重分布

ウザを用いて興味のあるニュース記事を調べてもらった. この行動を選んだ理由は, パーソナライズされた情報検索や情報提示などのアプリケーションを想定しているためである. 食事に関しては, 研究所やオフィスに導入された共用のダイニングテーブルにおいて個人向けの情報提示や食事に関する情報の提示, ライフログを行うアプリケーションを想定している.

被験者は 19 名 (19 クラス分類問題) で実験期間は約 1 か月である. 長期間に渡る提案手法の精度を調べるため, 1 名の被験者は 60 日以上に渡ってデータの収集を行った. 被験者は男性 16 名, 女性 3 名であり, 平均年齢は 23.5 歳である. 図 9 と図 10 に被験者の身長と体重の分布を示す. 研究室に所属する学生を被験者としたため, 身長や体重の分布が偏っていることが図 9 と 10 からわかる. 各被験者は異なる日に異なる服を着て, 2 つの行動をそれぞれ 3 セッションずつ行った.

4.2 評価手法

各行動において個人認証モデルを作成し, 1 つのセッションをテスト, 2 つのセッションを学習に用いる leave-one-session-out 交差検定によって提案手法および以下に示す比較手法を評価した. テストセッションでは最初の N_{test} 枚の深度画像を用いて評価した. このとき点群の位置合わせ (ICP) の計算には最初の N_{icp} 枚のみを用いた.

- 提案手法: 全てのソフトバイオメトリクス情報を用いて MKL を学習する手法.
- SVM: 全てのソフトバイオメトリクス情報を用いるが, 線形カーネルを用いた多クラス SVM を学習する手法.
- 提案手法 (骨格特徴抜き): 肩幅といった骨格特徴を除いて MKL を学習する手法.
- 提案手法 (形状特徴抜き): 頭の形状特徴を除いて MKL を学習する手法.
- 提案手法 (表面積特徴抜き): 体の表面積特徴を除いて

表 1 ユーザ分類の精度 [%]

	Web	食事	平均
提案手法	86.0	91.2	88.6
<i>SVM</i>	78.9	86.0	82.5
提案手法 (骨格特徴抜き)	82.5	84.2	83.4
提案手法 (形状特徴抜き)	77.2	77.2	77.2
提案手法 (表面積特徴抜き)	82.5	82.5	82.5
提案手法 (姿勢特徴抜き)	77.2	75.4	76.3
提案手法 (骨格特徴のみ)	36.8	42.1	39.5
提案手法 (形状特徴のみ)	43.9	43.9	43.9
提案手法 (表面積特徴のみ)	57.9	49.1	53.5
提案手法 (姿勢特徴のみ)	50.9	47.4	49.2

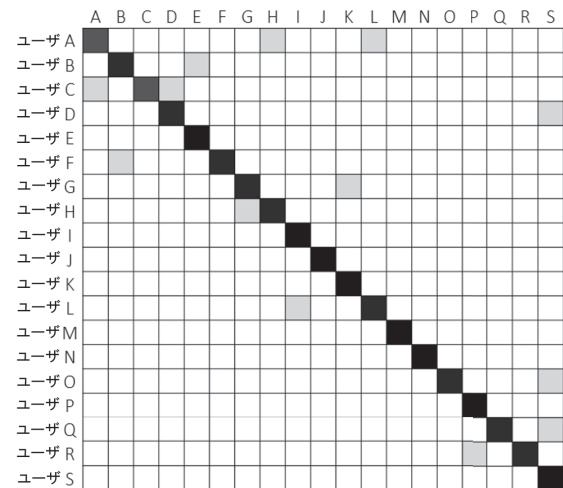


図 11 提案手法における Web ブラウジングと食事の混同行列

MKL を学習する手法.

- 提案手法 (姿勢特徴抜き) : 姿勢の癖の特徴を除いて MKL を学習する手法.

4.3 結果

4.3.1 分類精度

提案手法による個人認証の精度を表 1 に示す. 提案手法では 9 割に近い推定精度を達成することができた. 特に食事行動においては図 9, 10 に示したように偏った体型の分布の中で 9 割を超える推定精度を達成することができた. この推定精度は RGB-D カメラを用いて靴の画像を取得して認証を行う既存研究 [7] と比べて精度は劣る (靴を用いた認証では 18 名の被験者に対して 95.8% の認証精度であった) が, この研究では靴を履き替えたり, 別のユーザが同じ靴を履いていると認証ができない課題があった. 一方で提案手法では色情報を用いず, 天井に設置した深度カメラから得られた深度画像のみを利用するため, 異なる服装であったとしても形状や骨格が取得できれば認証できる利点がある.

また提案手法では線形カーネルを用いた SVM で分類する場合と比べて精度が 6% ほど向上した. これは各ソフトバイオメトリクス情報ごとにいくつかのカーネルを組み合わせさせて学習を行ったためと考えられる.

Web ブラウジングの行動は食事よりも動きが少なく, 被験者によってはテーブルに肘をついて姿勢が前かがみのまま作業を続けていることがあった. 天井からの視点ではユーザが前かがみになると腕を補足することが困難になり, その結果 Web ブラウジングの認証精度に悪影響を与えていると考えられる.

図 11 に Web ブラウジングと食事の結果を総合した混同行列を示す. 混同行列から分かるようにユーザ A はユーザ H とユーザ L に誤分類されている. またユーザ C はユーザ A とユーザ D に誤分類されている. ユーザ A とユーザ H, ユーザ C とユーザ D の身長差はそれぞれ 0.01 メートルと 0.03 メートルであった. これらの誤分類は体型が

似ているために生じたものと考えられる. 一方でユーザ A とユーザ L, ユーザ C とユーザ A の身長差はそれぞれ 0.1 メートルと 0.09 メートルであった. これらのユーザは姿勢が椅子に座った際の姿勢が類似しており, 姿勢の違いで識別することができなかったことが原因だと考えられる.

提案手法の個人認証精度はソフトバイオメトリクス情報を用いているため完璧ではない. そのため, なりすましが許容されないアカウント管理のようなアプリケーションでは提案手法を用いることは適切ではない. しかし, パーソナライズした情報を提示したり, ライフログのような高度なセキュリティを要求しないサービスで利用するための十分な精度は達成できていると考えられる.

4.3.2 ソフトバイオメトリクス情報の有用性

表 1 に特定のソフトバイオメトリクス情報を用いなかった場合の推定精度を示した. 形状や姿勢情報を用いなかった場合, 分類精度は 10% ほど低下した. 特に形状情報を用いなかった場合, 提案手法では身長が類似した被験者を分類することが困難であった. そのため, これらの特徴は体型が似ている被験者を分類するのに役立つと考えられる.

また同表より表面積の情報は骨格情報と比べて有用であることがわかる. これは部位の表面積の方が骨格情報よりも被験者の体型を反映しやすいためだと考えられる. 例えば太った体型の被験者は細い体型の被験者よりも特定の部位の表面積が大きくなるが, 身長が類似している場合は肩幅や腕の長さはあまり差がない. 本研究の被験者は図 9 のように身長に差があまりないことから身長に比例するであろう骨格情報は表面積と比べて個人を特定するのに有用ではなかったと考えられる.

骨格情報や形状情報, 表面積や姿勢の癖のみを用いて分類を行った場合, 分類精度は大きく低下した (表 1 の下部分). このことからいくつかの特徴を組み合わせることにより高い精度を達成できることがわかる. 特に頭の形状は体型が類似している被験者の分類に役立つが, 単体では体

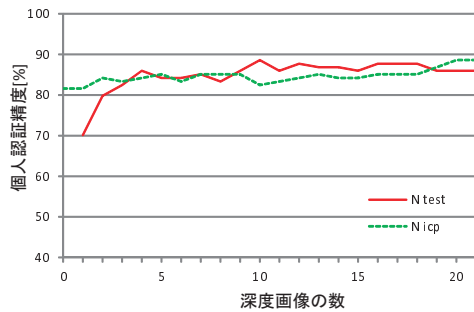


図 12 N_{test} と N_{icp} の数を変化させたときの認証精度の変化

型の違いが反映されないため、分類が困難になる。

4.3.3 N_{test} の影響

個人認証は深度画像の時系列テストデータの最初の N_{test} 枚を用いて行う。テストに用いる N_{test} 枚の数を変化させて分類精度を求めた結果を図 12 に示す。 N_{test} が小さいほど認証に必要な画像数が少ないため、早く個人を認証することができるが、精度は低下することがわかる。 N_{test} が 10 枚に達したとき、提案手法では精度が約 90% になった。深度カメラのサンプリングレートが 2Hz であったため、 $N_{test} = 10$ は 5 秒間に対応する。

4.3.4 個人認証にかかる計算時間

形状情報を求める際に用いる ICP アルゴリズムは計算コストが大きいため、ICP を行う深度画像の数を減少させた場合の計算時間を求めた。 $N_{test} = N_{icp} = 10$ のとき、提案手法では 10 枚目の画像を処理するのに 10.5 秒かかった。ICP を行う画像の数を減らすと $N_{test} = 10$, $N_{icp} = 5$ で 7.6 秒、 $N_{test} = 10$, $N_{icp} = 1$ で 5.6 秒かかることが分かった。形状情報を求める枚数を減らすと図 12 の N_{icp} のとおり分類精度はなだらかに低下する。本研究では将来的に計算時間の短い骨格情報や姿勢情報を用いて明らかに異なるユーザをフィルタリングした後、分類が難しいユーザのみ形状情報を用いて分類することで計算時間の短縮を実現する予定である。

4.3.5 長期間の分類精度

時間の経過による分類精度の変化を調べるため、1 人の被験者が 63 日にわたってデータを収集した。図 13 にこの被験者が正しいクラスに分類される確率の推移を示す。形状情報への影響を調べるため、被験者は 13 日目に散髪を行った。散髪を行った後の 14 日目は分類確率が下がっているが、それ以降は分類確率が戻っている。これより散髪によって形状情報が変化して被験者の分類確率に影響を与えるが時間の経過とともに分類精度が改善することがわかる。実際のアプリケーションでは散髪後に取得したデータを用いてユーザの SVM を再学習させることができるため、散髪による影響を抑えることができると考えられる。

5. おわりに

本稿ではテーブルトップシステムの利用者の行動を障害

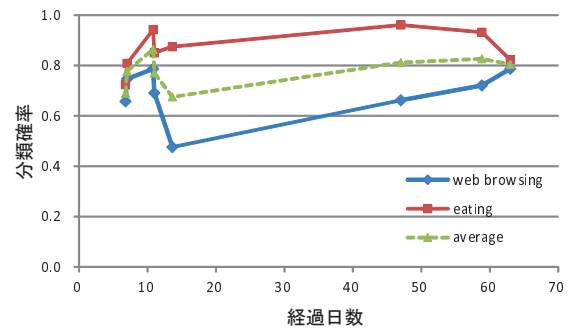


図 13 被験者が正しいクラスと分類される確率の変化

せずに認証を行う手法を提案した。提案手法では天井に設置した深度カメラから得られた深度画像から個人の特定するのに有用なソフトバイオメトリクス情報を抽出して MKL で学習を行った。評価実験では 19 名の被験者から得られた実データを用いて提案手法の有効性を検証し、約 90% の高精度で認証を行うことができた。

参考文献

- [1] V. Roth, P. Schmidt, and B. Güldenring, “The ir ring: authenticating users’ touches on a multi-touch display,” UIST 2010, pp.259–262, 2010.
- [2] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. Rehg, “A scalable approach to activity recognition based on object use,” ICCV 2007, pp.1-8, 2007.
- [3] Y. Shi, Y. Huang, D. Minnen, A. Bobick, and I. Essa, “Propagation networks for recognition of partially ordered sequential action,” CVPR 2004, vol.2, pp.862-869, 2004.
- [4] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, “uWave: Accelerometer-based personalized gesture recognition and its applications,” PerCom 2009, pp.1–9, 2009.
- [5] D. Schmidt, M.K. Chong, and H. Gellersen, “Hands-down: hand-contour-based user identification for interactive surfaces,” The 6th Nordic Conference on Human-Computer Interaction, pp.432–441, 2010.
- [6] R.M. Araujo, G. Graña, and V. Andersson, “Towards skeleton biometric identification using the Microsoft Kinect sensor,” ACM SAC 2013, pp.21–26, 2013.
- [7] S. Richter, C. Holz, and P. Baudisch, “Bootstrapper: recognizing tabletop users by their shoes,” CHI 2012, pp.1249–1252, 2012.
- [8] T. Augsten, K. Kaefer, R. Meusel, C. Fetzer, D. Kanitz, T. Stoff, T. Becker, C. Holz, and P. Baudisch, “Multi-toe: high-precision interaction with back-projected floors based on high-resolution multi-touch input,” UIST 2010, pp.209–218, 2010.
- [9] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, “Real-time human pose recognition in parts from single depth images,” Communications of the ACM, vol.56, no.1, pp.116–124, 2013.
- [10] A. Segal, D. Haehnel, and S. Thrun, “Generalized-icp,” Robotics: Science and Systems, vol.2, no.4, 2009.
- [11] M. Gönen, “Bayesian efficient multiple kernel learning,” ICML 2012, 2012.