Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

ウェアラブルセンサによるモノを用いた行動の認識について

前川卓也[†1] 柳沢豊[†2] 岸野泰恵[†1]
石黒勝彦[†1] 亀井剛次[†3]
櫻井保志[†1] 岡留剛[†4]

本稿では，カメラ，マイク，加速度センサなどのセンサ搭載する手首装着型センサデバイスを用いて人の日常行動を認識する手法および，そのデバイスの設計について述べる．センサデバイスは，ユーザの手のひらの周辺の領域を撮影するように設置されたカメラを備えることを特徴とし，これにより,「コーヒーを作る」,「水やりをする」などのモノを用いた行動の認識が可能となる．既存のウェアラブルセンサを用いた行動認識の研究では，加速度センサやマイクのみしか用いていなかったため，このような行動の認識は困難だった．また，提案デバイスはカメラやマイクを備えるため，ユーザのプライベートな生活を画像や音声として記録してしまうという，プライバシーの問題をもつため，本稿では，原画像や音声を必要としない行動認識手法を提案する．さらに，評価実験において，提案手法およびデバイスの有効性を確かめた．
キーワード ウェアラブルセンサ，行動認識，実験

Recognizing Object-related Activities with Wearable Sensors

Takuya Maekawa ,[†1] Yutaka Yanagisawa ,[†2]
Yasue Kishino ,[†1] Katsuhiko Ishiguro ,[†1]
Koji Kamei ,[†3] Yasushi Sakurai [†1]
and Takeshi Okadome[†4]

This paper describes a method that recognizes activities of daily living (ADLs) by employing a wrist worn sensor device with such various kinds of sensors as a camera, a microphone, and an accelerometer, and also describes the design of the wrist worn device. Specifically, the device captures a space around the user's hand by the camera to recognize ADLs that involve the manual use of objects such as making tea or coffee and watering plant. Existing wearable sensor devices equipped only with a microphone and an accelerometer cannot recognize these ADLs without object embedded sensors. We also propose an ADL recognition method that takes privacy issues into account because the camera and microphone can capture aspects of a user's private life. Furthermore, we experimentally confirmed the effect of our proposed device and method.
Keywords Wearable sensors, Recognizing activities of daily living, Experiment

1. はじめに

行動認識は，コンテキストアウェアサービスやライフログアプリケーションのための基盤となる技術であり，これまでに多くの研究がなされてきた．行動認識手法は，環境に設置したセンサを用いるものと，身体に装着したセンサ（ウェアラブルセンサ）を用いるものの２つのアプローチに大まかに分けられる．コンピュータビジョンの分野では環境に設置したカメラを用いた行動認識がこれまでに多く行われてきたが[17),21)]，近年では，環境に埋め込まれたユビキタスセンサを用いた行動認識が主流となっている．特に，環境内のオブジェクトに設置した RFID タグやスイッチセンサなどを用いて，モノ（オブジェクト）を用いた行動を推定するアプローチが多くとられている[9),18),23)]．例えば，コーヒーメーカーやカップや流し台に設置したセンサにより，それらのモノに関連する日常行動（例えば，‘コーヒーを作る’，‘食器を洗う’ など）を推定している．

ウェアラブルセンサを用いた手法では，加速度センサやマイクにより，身体の周期的な動きや姿勢，行動の際に発せられる音などの特徴を捉えることで，行動を推定する[1),2),10),13),14)]．先行研究では，これらのセンサを用いて，歩行，走行，歯磨き，会話，笑い声，鋸引きや穿孔などの工作作業などの認識が行われてきた．ウェアラブルセンサを用いた手法の利点は，環境にセンサを設置する必要がないことである．また，本稿で提案する手法もウェアラブルセンサを用いたものである．しかし，ほとんどの先行研究では加速度センサやマイクなどのセンサしか用いていなかったため，動きや音に特徴のない行動は認識できなかった．例えば，‘コーヒーを作る’，‘薬を飲む’ などの行動は，加速度センサやマイクのみでは認識が

†1 NTT コミュニケーション科学基礎研究所, NTT Communication Science Laboratories
†2 NTT 西日本, NTT West
†3 国際電気通信基礎技術研究所, ATR
†4 関西学院大学, Kwansei Gakuin University

Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

困難である.（モノに添付したユビキタスセンサを用いれば，これらの行動は認識可能だろう.）本研究では，このようなモノを用いた行動の認識を，カメラや加速度センサやマイクなどの多種のセンサを搭載した単一のウェアラブルセンサデバイス用いて認識することを目指す．特に，ユーザの手首に装着したカメラにより，ユーザが使っているモノの視覚的な特徴を捉え，行動の認識に利用することが本研究で提案するデバイスの特徴である．これは，ユーザが使っているモノは，そのユーザが行っている行動によく関係し，その行動の認識に有用であるというわれわれの考えに基づくものである．

本稿では，まず 2 章において，ウェアラブルセンサデバイスの提案とプロトタイプの実装を行う．提案デバイスは，身体の一箇所にのみ装着され，モノを用いた行動の認識に必要なセンサデータを取得する．3 章において，提案デバイスを用いた行動認識のための教師あり学習手法，および，センサデータからの学習に必要な特徴の抽出について説明する．4 章では，実装したプロトタイプデバイスを用いたセンサデータ収集実験および，それを用いた行動認識手法の性能評価を行う．データ収集実験は，‘緑茶を淹れる’，‘紅茶を淹れる’，‘掃除機掛けをする’，‘皿を洗う’ などのモノを用いた行動を対象とした．5 章では行動認識に関する関連研究を紹介し，最後に 6 章において，本稿の結論を述べる．

2. 提案デバイス

本稿の目的はモノの利用を含む行動の認識である．われわれは，人のほとんどの行動は手を使って行われるという点に着目し，手（手首）にセンサデバイスを装着することで，その行動に特徴的な手の動作などを取得する．また，われわれは身体の一箇所のみ（手首のみ）にセンサデバイスを装着すると想定する．足や腕，腰などの複数の箇所にセンサデバイスを装着することを想定した先行研究は多くあるが，日常生活においてそれらのデバイスを装着し続けることはユーザに多大な負担を強いる．われわれの提案するセンサデバイスは，マイク，加速度センサ，照度センサ，方位センサを搭載する．これらのセンサは多くの先行研究において用いられてきたものである[10),13)]．提案デバイスの特徴は，以上のセンサに加えてカメラを搭載することである．特に，カメラをユーザが手で持っているモノを撮影するように設置する．これにより，行動の際に用いているモノの視覚的な情報を，その行動の認識に用いることができる．

以上の考えを基に設計したセンサデバイスのコンセプト図を図 1 (a) に示す．センサデバイスは，搭載するセンサから取得したデータを無線通信によりホスト PC に送信することを想定している．また，実際に製作したプロトタイプセンサデバイスを図 1 (b) に示す．

(a) (b)

図 1 (a) 理想的なデバイス，(b) 製作したプロトタイプデバイス

カメラのレンズは，ユーザの手のひら周辺の領域を撮影するように手首の内側に設置した．われわれは，USB カメラ，有線マイク，USB 接続のセンサボードをリストバンドに搭載した．また，センサボードは，3 軸加速度センサ，照度センサ，3 軸方位センサを搭載している．カメラは，352 × 288 の 24 ビット JPEG 画像を約 6FPS のサンプリングレートで撮影できる．また，マイクは全指向性であり，サンプリングレートは 44.1 kHz である．センサボードは搭載するセンサから約 30Hz でデータを収集する．また，センサデバイスはプロトタイプであるため，センサ類は有線でラップトップ PC に接続されている．実験の際は，被験者が背負ったバックパック内に入れた PC にデバイスを接続する．

3. 行動認識手法

われわれは，教師あり学習を用いて行動の認識を行う．トレーニングデータは，手で付与されたラベル（行動のクラスとその開始・終了時刻）を含み，トレーニングデータにより学習されたモデルを用いて，テストデータの認識を行う．

3.1 特徴抽出

行動認識のために，まずセンサデータからの特徴抽出を行う．われわれは様々な種類の様々なサンプリングレートをもつセンサから取得された時系列データを対象としている．そこで，まずそれぞれのセンサについて適切なウインドウサイズでセンサデータから特徴を抽出した後，1 秒間のウインドウ内でそれらの特徴の平均を求める．ただし，ウインドウは前後のウインドウと 50%の重複を持たせた．そして，すべての特徴において求めた平均を結合することで特徴ベクトルを求める.（特徴ベクトルの 1 つ要素は 1 つの特徴の平均に対応する.）以上のように求めた特徴ベクトルの時系列を行動の認識に用いる．以下では，センサデータからの特徴抽出について説明する．

Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

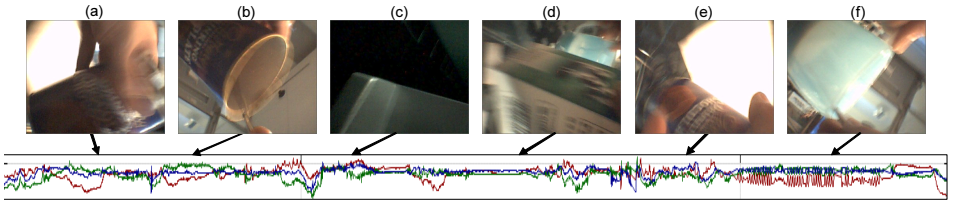

図 2 ココアを作っている際のカメラ画像と加速度データ

3.1.1 画像からの特徴抽出

ユーザが現在使っているモノの情報は，その行動の認識に有用だろう．以下では，われわれのデバイスが搭載するカメラにより撮影される画像の特徴と，撮影された画像のプライバシなどの問題について述べたあと，それらを基に画像からどのような特徴を抽出するかについて述べる．また，リアルタイムに行動推定するためには，画像からの特徴抽出は高速に行われるべきである．

[カメラ画像の特徴]

図 2 は，センサデータ取得実験において被験者がココアを作ったときに取得されたカメラ画像と加速度データを示す．図 2 (a) は，ココア缶を棚から取り出しているときに撮影された画像，(b) はココア粉末をスプーンですくっているときの画像，(c) は冷蔵庫に近付いているときの画像，(d) は牛乳パックを持っているときの画像，(e) はココア缶を棚に戻しているときの画像，(f) はココアを混ぜているときの画像である．このような画像には以下の特徴が見られる．(1) モノは様々な角度から撮影される．(2) ほとんどの画像はモノの一部しか捉えていない．(3) 手の動きにより，ほとんどのモノはブレて撮影されている．(4) 撮影されているモノの明度は，照明，モノ，カメラの位置関係（撮影条件）により容易に変わる．多くの画像認識研究では，オクルージョンや回転，スケールの変化などを考慮した画像からの物体認識が行われている[12),20)]．しかし，そのためにはモノの詳細なモデルをあらかじめ用意する必要があるため，そのコストは高い．

[カメラ画像の問題]

以下では，センサデバイスに搭載したカメラから取得した画像に関する 2 つの問題について議論する．1 つ目の問題はプライバシの問題である．われわれは，センサデバイスによって取得されたデータを無線通信でホスト PC に送信することを想定している．しかし，手首に装着したカメラによって撮影された画像をそのまま無線通信で送信することに，ユーザは抵抗を感じるだろう．2 つ目の問題は通信量の問題である．われわれのデバイスは，生画像をリアルタイムに送信することで，約 90 KB/sec もの通信帯域を占有すると想定される．またそれにより，デバイスのバッテリ消費量も多くなるだろう．以上から，デバイスは生データの代わりに抽象的なデータ量の少ないデータを送るべきと考える．

[画像からの特徴抽出の概要]

以上のカメラ画像の特徴と問題から，われわれはデバイスから送られた抽象的な画像データから，大まかな色情報のみを画像の特徴として抽出こととする．抽象的な画像データは，データ量も少なく，プライバシの問題も少ない．具体的には，画像の色ヒストグラムをデバイスが作成し，送信するとする．(一般的に用いられている色ヒストグラムとは異なる.）そして，ヒストグラムを用いて原画像に，‘行動に特徴的な色’ に似た色を持つピクセルが含まれている数をカウントし，その数をそのヒストグラムの特徴として用いる．例えば，ココア缶の色が紫であるとき，画像中に含まれる紫に似た色のピクセル数の情報は，ココアを作るという行動を認識するために有用であろう．われわれは，トレーニングデータを用いて，あらかじめ行動ごとに特徴的な色を抽出しておく．そして，抽出された特徴的な色ごとに，ヒストグラム中に含まれているその色に似たピクセルの数をカウントする．つまり，1 つのヒストグラムに対して，抽出された特徴的な色ごとに，その色に似たピクセルの数が求められる．これらの数をそのヒストグラム（画像）の特徴として用いる．以下では，行動ごとに特徴的な色を抽出する方法，画像からヒストグラムを作成する方法，および，それらから特徴を計算する方法について説明する．

[行動に特徴的な色の抽出]

ラベリングされたトレーニングデータから，行動に特徴的な色をあらかじめ取得しておく．ある行動のクラスに対して特徴的な色を抽出する方法を以下に説明する．(i) まず，その行動としてラベリングされた全ての画像の全てのピクセルの色を k 平均法を用いて 64 のクラスタに分割する．ただし，クラスタリングは HSB 色空間 (hue, saturation, and brightness color space) 内で行った．そして，クラスタごとにその平均色を求める．これにより，行動ごとに 64 の代表的な色が得られた．ここで，われわれは明度の軸をもつ HSB 色空間に着目した．上で述べたように，撮影されているモノの明度は撮影条件により変わる．そこで，それぞれの色の明度の値に 0.5 を乗算することで，明度の重みを低減させている．(ii) 以上の 64 の代表色から，top-m の色のみを行動に特徴的な色として採用する．われわれは，64

Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

の代表色を情報利得（情報量）の概念を用いてランキングする．情報利得は，分類問題において，インスタンスを精度よく区別可能な属性を発見するために用いられる．ある属性の値を用いて，より多くのインスタンスを精度良く分類できるとき，その属性の情報利得は大きくなる．われわれは，画像（インスタンス）をその行動のクラスとそれ以外のクラスに分類する際の情報利得を，属性ごとに求める．ここで，属性とは，画像に含まれる代表色に似た色のピクセル数である．以上のようにして求めた情報利得の値により代表色をランキングし，その top-m のみをその行動に特徴的な色とする．

ここで，1 つの例を用いて説明する．ココアを作る行動に用いられるココア缶の色が紫であるとし，その他の行動に用いられるモノに紫に似た色をもつものが無いとする．ココアを作る行動の間に撮影された画像内の紫に似た色のピクセル数は多く，その属性（紫に似た色のピクセル数）によりココアを作る行動の画像の多くを区別することができると考えられるため，その属性の情報利得は大きいだろう．情報利得（情報量）の計算については文献 25) を参照して欲しい．以上の手順により，行動ごとに m の特徴的な色を取得できる．

[ヒストグラムの作成]

デバイスが，1 つの画像からヒストグラムを生成する方法について説明する．(i) 画像内の全てのピクセルの色を k 平均法により 64 のクラスタに分ける．それぞれのクラスタの代表色を，そのクラスタに含まれるピクセルの色の平均とする．(ii) 得られた 64 のクラスタから，ビンの数が 64 のヒストグラムを作成する．それぞれのビンは 1 つのクラスタに対応し，ビンの値はクラスタに含まれるピクセルの数である．このヒストグラムは，それぞれのクラスタの HSB データと，それに含まれるピクセルの数の情報のみもつ．以上のような減色処理とピクセルの座標情報の破棄により画像のデータ量を大幅に削減できる．1 つのヒストグラムのデータ量は 448 バイトであり，必要とされる通信速度は 90 KB/sec から 2.7 KB/sec にまで削減される．

[ヒストグラムからの特徴抽出]

1 つのヒストグラムに対して，あらかじめ求めた行動ごとに特徴的な色に似た色のピクセルが含まれている数を求め，それを特徴として用いる．具体的には，それぞれの特徴的な色ごとに，その色と似ている色をもつビンを求め，それらのビンの値の数を合計する．色同士の類似度は，HSB 色空間におけるユークリッド距離を用い，距離が th（実装では 15）以内の色同士を似ているとした．ただし，上記と同様に明度の重みを低減させた色空間を用いた．また，この処理はホスト PC 上で行われると想定している．

3.1.2 音からの特徴抽出

行動が行われている時に発生する音から特徴を抽出する．例えば，掃除機がけをしているときや，蛇口から水を流しているときなどに特徴的な音が得られるだろう．このような持続的な環境音は，その周波数に特徴がある．文献 5) では，Mel-Frequency Cepstral Coefficient (MFCC) が環境音の周波数的特徴を最もよく捉える変換手法であるとされている．また，文献 3) では音を用いたバスルームでの行動認識において，MFCC により高い精度を達成している．以上から，われわれも MFCC を用いた特徴抽出を行う．MFCC は Fast Fourier Transform (FFT) をベースにしているため，高速に計算できる．ここで，マイクで録音された音声は，カメラで撮影された画像と同様にプライバシの問題を孕んでいる．そこで，デバイス上で音声から特徴抽出を行い，ホスト PC に送信することとする．また，高いサンプリングレートでの音声録音は高コストである．そこで，短時間の音声を断続的に記録し，それに対して Hamming 窓を掛けたあと，13 次元の MFCC を計算する．実装では，25 ミリ秒の長さの音声を 1 秒間に 6 回記録するとした．

3.1.3 加速度データからの特徴抽出

加速度データからは，手の姿勢や周期的な動きの特徴を得られる．例えば，図 2 (f) に示すように，ココアを撹拌しているときには，加速度データに周期的な特徴が見られるのが分かるだろう．われわれは，1 軸の加速度データの 64 サンプルウインドウの FFT 成分から，平均，エネルギー，周波数領域エントロピー，主要周波数成分の特徴を抽出する．平均は，手の姿勢を捉えるための特徴である．例えば，歯を磨いているときの手の姿勢は特徴的だろう．平均は FFT の直流成分を用いた．エネルギーは，動きの強度を表す特徴であり，立った状態などと，歩行などの強度の異なる行動を区別できる[24]．エネルギーは，FFT 成分ごとの振幅の二乗を合計し，さらにウインドウサイズで除算することで正規化したものである．ただし，その合計から FFT の直流成分は除いている．周波数領域エントロピと主要周波数成分は，異なる周波数の周期的動作を区別するために用いる．例えば，後述の実験において，ココアを撹拌する動きの主要周波数は 2 から 4Hz だったが，歯を磨いている動きは 4 から 6Hz だった．周波数領域エントロピーは，FFT 成分の情報エントロピーにより求められる[1]．主要周波数成分は，最大の値をもつ FFT 成分の周波数であり，実装では全ての成分の平均より 3 倍以上の値をもつものとしている．主要周波数成分がない場合は，その特徴の値を 0 としている．以上のような特徴抽出を 3 軸の加速度データそれぞれに対して行う．

3.1.4 照度と方位データからの特徴抽出

われわれは，照度センサと 3 軸方位センサからのセンサデータをそのまま特徴として用

Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

いる．方位のデータは，行動における人が向いている方位の特徴を捉える．例えば，ある人が毎朝流し台で歯を磨くことを習慣にしているとき，その人が歯を磨いているときに向いている方位はほぼ一定だろう．

3.2 行動の分類

われわれは，ラベリングされたトレーニングデータから特徴抽出を行い，それにより得られた特徴ベクトルの系列から行動のクラスを学習する．そして，テストデータから抽出した各時刻の特徴ベクトルを対応する行動のクラスに分類する．機械学習を用いたクラス分類手法は，識別モデルと生成モデルに 2 つの手法に分類される．識別モデルはクラス間の境界を学習し，生成モデルはクラスごとの確率密度関数を学習する．一般的に識別モデルの方がクラス分類問題において精度がよいとされているが，生成モデルの方が欠損データの扱いに長けている．機械学習の分野では，それらの長所を生かした識別・生成ハイブリッドモデルに注目が集まっている[8),19)]．近年の行動認識の研究においても，ハイブリッドモデルを用いることで高い認識精度を達成している[7),10),11)]．また，われわれは時系列データを扱うため，行動の時間的パターンのモデル化によく用いられる生成モデルの 1 つである隠れマルコフモデル (HMM) をハイブリッドモデルに組み込むことは効果的であると考えられる．

以上から，HMM を生成モデルとして用いた識別・生成ハイブリッドモデルを行動のモデル化に採用する．われわれの用いるハイブリッドモデルは，図 3 のように，識別型分類器と HMM 分類器の主に 2 つのモジュールから構成される．1 つ目のモジュールへの入力は 3.1 章において抽出された特徴ベクトルの系列である．1 つ目のモジュールは行動ごとに用意した識別分類器であり，特徴ベクトルが対応する行動のクラスに分類されるか，そうでないかを学習・認識する（2 値分類）．つまり，図 3 中の n は，学習する行動のクラスの種類の数に対応する．それぞれの 2 値分類器は，特徴ベクトルごとに，対応するクラスに分類される確率を出力する．つまり，1 つ目のモジュールからは，n 次元の確率のベクトルの系列が出力され，2 つ目のモジュールへの入力となる．2 つ目のモジュールは，行動ごとに用意した right-to-left HMM から構成され，それぞれの HMM は，それぞれの確率のベクトルに対して，対応する行動のクラスに分類される尤度を出力する．ある時刻において，最も高い尤度を出力した HMM に対応するクラスを，その時刻に行われた行動のクラスとする．

4. 評価実験

4.1 データセット

被験者の日常生活から得られるデータが最も自然なデータであり，そのようなデータを用いて実験を行うことが望ましい．しかし，そのためには被験者の日常生活を常にモニタリングしておく必要があるため，そのようなデータを十分な量だけ集めることは困難である．われわれは，センサデータを文献 1) で用いられている手法を用いて収集する．その手法では，被験者は与えられたワークシートに従って行動を行う．ワークシートにより，被験者ごとにランダムに並べられた行動を順に行うように指示する．また，ワークシートによる指示は,「部屋を掃除機がけする」,「ラックの中の CD から曲を 1 曲選んで聞く」など，ある程度あいまいであるため，被験者はある程度自由に行動を行うことができる．つまり，多様なセンサデータを収集することができる．

データは 10 人の被験者によって，われわれの実験環境でプロトタイプデバイスを用いて収集された．被験者は，われわれの研究所の作業員（研究者ではない）である．提案デバイスによって取得されるセンサデータはその環境に大きく依存するため，われわれは 2 つの環境（環境 1 と環境 2）においてそれぞれセンサデータを取得し，それぞれのデータを用いて提案手法を評価する．つまり，環境 1(2) で取得したトレーニングデータを用いて学習したモデルを用いて，環境 1(2) で取得したテストデータの評価を行う．環境 1 は，われわれの研究所における家庭を模した実験環境である[16)]．環境 1 には，キャビネットや机や調理用具などがあらかじめ設置されており，今回の実験ではそれらを用いた．環境 2 は，われわれの研究所の一室であり，今回の実験用に必要な備品やモノを設置した．それぞれのデータ収集実験では，被験者は表 1 に示した 15 種類の行動をランダムな順序で行う．上記のようなセッションをそれぞれの環境で 14 セッションずつ行った．

以上の実験で得られたセンサデータは多様であり現実的な利用に近いデータであると考えている．まず，実験は午前 9 時から午後 6 時までの間に行われたため，画像データはさまざまな照明条件で撮影されている．また，10 人の被験者に行ってもらったため，行動の行い方には幅があるだろう．さらに，カメラにしばしば映り込む被験者の衣服の色も実験ごとに異なる．実験で用いたモノもさまざまな色やテクスチャをもつもの，半透明なものなどが含まれている．また，いくつかのモノ同士は似た色をもっている．

以上のようにして得られたデータに対して，各々の被験者がデバイスのカメラで撮影された画像を見ながらラベリングを行った．（トレーニングデータの取得期間は，デバイスは撮影した原画像を保存すると想定している．）ラベリング実験の詳細については，文献 15) を参照して欲しい．

4.2 評価手順

それぞれの環境で得られたセンサデータごとに，Leave-one-session-out 交差検定を用い

Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

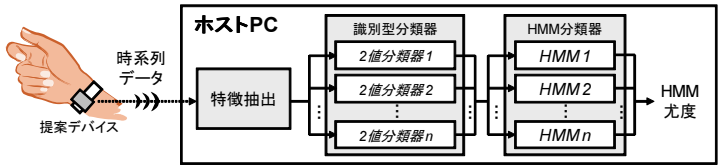

図 3 分類手法の概要

表 1 実験で行った行動とその平均時間

	行動	平均時間 (分)		行動	平均時間 (分)
A	brush teeth	3.65	I	make juice	1.77
B	cook pasta	5.98	J	make tea	1.37
C	cook rice	4.33	K	practice aromatherapy	0.66
D	feed fish	0.40	L	take supplement	0.82
E	listen to music	1.69	M	vacuum	1.26
F	make cocoa	1.37	N	wash dishes	3.68
G	make coffee	1.63	O	water plants	0.27
H	make green tea	1.16			

表 2 認識手法の平均精度 (適合率 / 再現率)，値はパーセント

	AdaBoost+HMM (特徴ベクトル)		C4.5+HMM (特徴ベクトル)		AdaBoost+HMM (インスタンス)		C4.5+HMM (インスタンス)	
	環境 1	環境 2	環境 1	環境 2	環境 1	環境 2	環境 1	環境 2
A: brush teeth	42.1/73.0	75.2/91.4	74.3/79.0	84.3/88.1	27.5/78.6	50.0/92.9	92.9/92.9	77.8/100
B: cook pasta	97.3/86.4	99.2/90.4	97.2/83.7	98.7/84.7	100/92.9	100/100	100/100	100/92.9
C: cook rice	76.2/93.1	79.1/96.0	88.3/85.1	88.3/87.5	54.2/92.9	66.7/100	81.2/92.9	87.5/100
D: feed fish	44.9/3.0	0.0/0.0	60.5/67.7	74.1/58.7	0.0/0.0	0.0/0.0	92.3/85.7	88.9/57.1
E: listen to music	86.7/81.2	50.2/65.3	84.7/90.1	58.4/82.4	80.0/85.7	45.0/64.3	93.3/100	72.2/92.9
F: make cocoa	0.0/0.0	87.9/72.0	74.6/64.4	85.2/76.4	0.0/0.0	84.6/78.6	91.7/78.6	92.9/92.9
G: make coffee	36.4/61.3	49.2/77.8	73.8/66.5	85.2/90.4	24.2/57.1	40.7/78.6	69.2/64.3	93.3/100
H: make green tea	16.4/16.6	69.9/7.0	50.1/13.8	34.5/72.9	18.8/21.4	100/7.1	40.0/14.3	45.8/84.6
I: make juice	86.1/72.9	27.0/53.1	79.7/78.2	76.4/70.4	92.3/85.7	17.9/50.0	93.3/100	92.3/85.7
J: make tea	0.0/0.0	72.1/47.8	24.5/70.3	72.7/42.3	0.0/0.0	60.0/42.9	47.6/71.4	75.0/42.9
K: practice aroma.	66.2/38.7	97.4/57.7	72.8/68.6	77.1/75.4	83.3/35.7	90.9/71.4	100/85.7	100/85.7
L: take supplement	0.0/0.0	0.0/0.0	50.8/69.2	73.7/62.4	0.0/0.0	0.0/0.0	70.6/85.7	90.9/71.4
M: vacuum	96.8/82.0	89.4/80.1	89.0/87.8	93.2/83.1	100/85.7	86.7/92.9	100/100	100/92.9
N: wash dishes	98.3/80.9	97.6/77.5	93.1/82.6	94.3/89.9	100/85.7	100/92.9	93.3/100	93.3/100
O: water plants	100/88.4	0.0/0.0	84.5/92.4	40.5/59.8	100/100	0.0/0.0	100/100	100/71.4
平均	56.5/51.8	59.6/54.4	73.2/73.3	75.8/75.0	52.0/54.8	56.2/58.1	84.4/84.8	87.3/84.7

て行動認識手法の評価を行った．つまり，1 つのセッションで得られたデータをテストデータとし，残りの 13 セッションで得られたデータをトレーニングデータとする．また今回の評価では，識別型分類器として，AdaBoost M1 と C4.5 決定木[25] を比較した．AdaBoost は，弱学習器を組み合わせて性能の良い学習器を構成する手法の 1 つであり，今回は弱学習器として決定株（decision stump）を用いた．

4.3 評 価 結 果

4.3.1 認識手法の性能評価

表 2 に 2 つの認識手法の性能を示す．AdaBoost+HMM (特徴ベクトル) と C4.5+HMM (特徴ベクトル) は，特徴ベクトルを基に計算した適合率と再現率である．つまり，あるクラスにおける適合率は，正しくその行動のクラスに分類された特徴ベクトルの数と，その行動のクラスに分類された特徴ベクトルの数との比である．あるクラスにおける再現率は，正しくその行動のクラスに分類された特徴ベクトルの数と，実際にその行動のクラスに分類される特徴ベクトルの数との比である．C4.5 決定木を識別型分類器として用いた C4.5+HMM は，AdaBoost により構築されたアンサンブル分類器を識別型分類器として用いた AdaBoost+HMM より高い性能を示した．AdaBoost+HMM は，‘feed fish’，‘take supplement’，‘water plants’ などの短い長さの行動の精度が極端に低いことが多かった．これは，トレーニングデータにおいて，長さが短い行動に対応する特徴ベクトルの数が少ないことによる．AdaBoost アルゴリズムは，クラスに対応する特徴ベクトルの数のバランスが取れていない場合，特徴ベクトルの数が少ないクラスを無視して学習器を構成する傾向がある．これは，2 値分類の場合，特徴ベクトルの数が多いクラスに全ての特徴ベクトルを分類するよう学習器を構成すれば十分な性能が出せるためである．

一方 C4.5+HMM においても，‘feed fish’，‘practice aromatherapy’，‘take supplement’，‘water plants’ といった短い長さの行動の精度はそれほど高いとは言えない．これは，ラベルの先頭と末尾の区間の影響によるものであると考えられる．手で付けられたラベルの先頭と末尾の区間にクラスの区別が困難な特徴ベクトルが含まれることは避けられないだろう．例えば，サプリメントを飲むという行動では，被験者が棚までピルケースを取りに行こうとする間の特徴ベクトルは区別が困難である．このような区間の特徴ベクトルは誤って分類されることが多い．長さが短い行動においては，このような区間の長さと行動自体の長さの比

Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

表 3 C4.5+HMM の混合行列

環境 1	A: brush teeth	B: cook pasta	C: cook rice	D: feed fish	E: listen to music	F: make cocoa	G: make coffee	H: make green tea	I: make juice	J: make tea	K: practice aroma.	L: take supplement	M: vacuum	N: wash dishes	O: water plants
A	13	0	0	0	0	0	0	0	0	0	0	0	0	1	0
B	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	0	13	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	1	12	0	0	1	0	0	0	0	0	0	0	0
E	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	11	1	0	1	0	0	1	0	0	0
G	0	0	0	1	0	0	9	1	0	1	0	2	0	0	0
H	0	0	0	0	0	0	2	2	0	9	0	1	0	0	0
I	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0
J	0	0	0	0	0	1	0	2	0	10	0	1	0	0	0
K	0	0	1	0	1	0	0	0	0	0	12	0	0	0	0
L	0	0	1	0	0	0	0	0	0	1	0	12	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14

環境 2	A: brush teeth	B: cook pasta	C: cook rice	D: feed fish	E: listen to music	F: make cocoa	G: make coffee	H: make green tea	I: make juice	J: make tea	K: practice aroma.	L: take supplement	M: vacuum	N: wash dishes	O: water plants
A	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	13	0	0	1	0	0	0	0	0	0	0	0	0	0
C	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0
D	1	0	0	8	2	0	0	3	0	0	0	0	0	0	0
E	0	0	1	0	13	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	13	0	1	0	0	0	0	0	0	0
G	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	11	0	2	0	0	0	0	0
I	1	0	0	0	1	0	0	0	12	0	0	0	0	0	0
J	0	0	0	0	0	1	1	5	0	6	0	1	0	0	0
K	1	0	0	0	1	0	0	0	0	0	12	0	0	0	0
L	1	0	0	0	0	0	0	2	1	0	0	10	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	13	1	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0
O	0	0	1	1	0	0	0	2	0	0	0	0	0	0	10

が大きくなるため，その精度も低くなることが多い．しかし，C4.5+HMM においてはほとんどの特徴ベクトルが精度良く分類されていると考える．それを示すため，表 2 に行動のインスタンスごとに多数決を用いて精度を求めた結果も示している．つまり，ある行動のインスタンスが分類されるクラスは，そのインスタンスに含まれる特徴ベクトルの分類結果の多数決によるものとする．提案デバイスから得られるセンサデータはその環境に大きく依存するにも関わらず，C4.5+HMM は両方の環境で高い精度を達成している．

表 3 に，環境 1 と 2 における C4.5+HMM の混合行列を示す．この行列は行動のインスタンスごとの分類に基づいて求めたものである．表 3 から分かるように，‘make green tea’ と ‘make tea’ の区別は非常に困難であった．これは，これらの行動において被験者の手の動きがほとんど同じであることや，これらの行動に用いられるモノのほとんど（急須や電気ポットなど）が重複しているためである．さらに，環境 1 においては，紅茶を入れる缶と緑茶を入れる缶の色も類似していた．また，‘feed fish’，‘take supplement’，‘water plants’ などの少ないの数のモノしか使われず，音や手の動きに特徴がないような行動の認識も失敗していることがあった．特に，他の行動で似た色のモノが使われている場合は認識が困難だった．例えば環境 2 において，魚の餌袋が急須の色に酷似していたため，‘feed fish’ の精度は低かった．

表 4 C4.5+HMM における，さまざまなセンサの組み合わせによる精度 (適合率 / 再現率)

センサ	条件	環境	精度
カメラ	only	1	76.7/73.2
		2	75.1/71.8
	w/o	1	77.7/75.2
		2	71.8/67.6
マイク	only	1	28.3/32.9
		2	21.8/28.6
	w/o	1	84.9/83.3
		2	83.8/81.0
加速度センサ	only	1	48.5/44.3
		2	47.3/43.8
	w/o	1	82.1/80.5
		2	84.9/79.5

センサ	条件	環境	精度
照度センサ	only	1	0.1/6.7
		2	0.4/6.7
	w/o	1	81.7/82.4
		2	89.6/88.0
方位センサ	only	1	23.1/21.9
		2	10.8/10.0
	w/o	1	85.9/84.8
		2	89.9/87.0

4.3.2 それぞれのセンサの貢献

表 4 は，C4.5+HMM さまざまなセンサの組み合わせによるインスタンスごとの分類に基づいた精度を示したものである．例えば，‘only camera’ は，カメラから抽出した特徴のみを用いて推定を行ったときの精度であり，‘w/o camera’ は，カメラ以外のセンサから抽出した特徴のみを用いて推定を行ったときの精度である．カメラのみを用いただけでも約 75%もの精度を達成していることが分かることから，カメラの行動認識に果たす役割の大きさが分かるだろう．今回の実験では，行動認識に貢献しているセンサは，カメラ，加速度センサ，マイクの順だった．照度センサと方位センサは行動認識精度の向上にはほとんど役立たなかった．また，行動ごとの精度についての評価は，文献 15) を参照して欲しい．

5. 関 連 研 究

本章では，これまでに紹介できなかったウェアラブルカメラを用いたセンシングに関する研究を紹介する．文献 6) では，靴に搭載したカメラと加速度センサにより，歩行の解析やユーザが居るフロアの認識などを行っている．文献 26) では，キッチンのモノに添付した RFID タグと，キッチンを俯瞰するよう設置されたカメラを用いたキッチンでの行動認識を行っている．ユーザが手首に装着した RFID リーダにより，モノの利用を検知できる．文献 4) では，首掛けストラップに装着したカメラとマイクを用いて場所の移動を検出している．文献 22) では，帽子に装着した 2 つのカメラを用いて，実世界ゲームでのユーザのアクション（銃をかまえるなど）を認識している．一方，本研究ではモノを用いた行動に注目し，手に装着したカメラを含むセンサを用いて行動の特徴を捉えている．

Vol.2010-SLDM-144 No.57
Vol.2010-EMB-16 No.57
Vol.2010-MBL-53 No.57
Vol.2010-UBI-25 No.57
2010/3/27

6. おわりに

本稿では，モノを用いた行動認識のための手首装着型センサデバイスのプロトタイプおよび，そのデバイスを用いた行動認識手法について述べた．提案デバイスは，特に，カメラによりユーザが使っているモノの視覚的特徴を捉えることを特徴とする．これにより，ウェアラブルセンサのみでは困難だった複雑な行動の推定を行うことができる．また，実験では，提案デバイスを用いて高い精度の行動推定が可能なことを示した．今後は，さらなる精度向上のため，SIFT 特徴[12] などをブレなどの無い良質な画像から抽出し，認識の特徴に用いることなどを考えている．また，図 1 (a) に示すような単体で動作するデバイスを作成し，より自然なセンサデータを用いた行動認識実験も行う予定である．

参 考 文 献

1) Bao, L. and Intille, S.: Activity recognition from user-annotated acceleration data, *Pervasive 2004*, pp.1–17 (2004).
2) Blum, M., Pentland, A. and Troster, G.: Insense: Interest-based life logging, *IEEE Multimedia*, Vol.13, No.4, pp.40–48 (2006).
3) Chen, J., Kam, A., Zhang, J., Liu, N. and Shue, L.: Bathroom activity monitoring based on sound, *Pervasive 2005*, pp.47–61 (2005).
4) Clarkson, B., Mase, K. and Pentland, A.: Recognizing user context via wearable sensors, *ISWC 2000*, pp.69–75 (2000).
5) Cowling, M.: Non-speech environmental sound recognition system for autonomous surveillance, PhD Thesis, Griffith University (2004).
6) Fitzpatrick, P. and Kemp, C.: Shoes as a platform for vision, *ISWC 2003*, pp. 231–234 (2003).
7) Huynh, T. and Schiele, B.: Towards less supervision in activity recognition from wearable sensors, *ISWC 2006*, pp.3–10 (2006).
8) Jaakkola, T. and Haussler, D.: Exploiting generative models in discriminative classifiers, *NIPS 1999*, pp.487–493 (1999).
9) Kasteren, T., Noulas, A., Englebienne, G. and Krose, B.: Accurate activity recognition in a home setting, *UbiComp 2008*, pp.1–9 (2008).
10) Lester, J., Choudhury, T. and Borriello, G.: A practical approach to recognizing physical activities, *Pervasive 2006*, pp.1–16 (2006).
11) Lester, J., Choudhury, T., Kern, N., Borriello, G. and Hannaford, B.: A hybrid discriminative/generative approach for modeling human activities, *IJCAI 2005*, pp. 766–772 (2005).
12) Lowe, D.: Distinctive image features from scale–invariant keypoints, *Int'l Journal on Computer Vision*, Vol.60, No.2, pp.91–110 (2004).
13) Lukowicz, P., Junker, H., Stager, M., Buren, T.V. and Troster, G.: WearNET: a distributed multi-sensor system for context aware wearables, *UbiComp 2002*, pp. 361–370 (2002).
14) Lukowicz, P., Ward, J., Junker, H., Stager, M., Troster, G., Atrash, A. and Starner, T.: Recognizing workshop activity using body worn microphones and accelerometers, *Pervasive 2004*, pp.18–32 (2004).
15) Maekawa, T., Yanagisawa, Y., Kishino, Y., Ishiguro, K., Kamei, K., Sakurai, Y. and Okadome, T.: Object-based activity recognition with heterogeneous sensors on wrist, *Pervasive 2010* (2010 to appear).
16) Maekawa, T., Yanagisawa, Y., Kishino, Y., Kamei, K., Sakurai, Y. and Okadome, T.: Object-blog system for environment-generated content, *IEEE Pervasive Computing*, Vol.7, No.4, pp.20–27 (2008).
17) Mihailidis, A., Carmichael, B. and Boger, J.: The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home, *IEEE Trans. on Info. Tech. in BioMedicine*, Vol.8, No.3, pp.238–247 (2004).
18) Philipose, M., Fishkin, K. and Perkowitz, M.: Inferring activities from interactions with objects, *IEEE Pervasive Computing*, Vol.3, No.4, pp.50–57 (2004).
19) Raina, R., Shen, Y., Ng, A. and McCallum, A.: Classification with hybrid generative/discriminative models, *Advances in Neural Information Processing Systems 16* (2003).
20) Schiele, B. and James, L.: Object recognition using multidimensional receptive field histograms, *ECCV 1996*, pp.610–619 (1996).
21) Shi, Y., Huang, Y., Minnen, D., Bobick, A. and Essa, I.: Propagation networks for recognition of partially ordered sequential action, *CVPR 2004*, Vol.2, pp.862–869 (2004).
22) Starner, T., Schiele, B. and Pentland, A.: Visual contextual awareness in wearable computing, *ISWC 1998*, pp.50–57 (1998).
23) Tapia, E., Intille, S. and Larson, K.: Portable wireless sensors for object usage sensing in the home: challenges and practicalities, *AmI 2007*, pp.19–37 (2007).
24) Welk, G. and Differding, J.: The utility of the Digi-Walker step counter to assess daily physical activity patterns, *Medicine & Science in Sports & Exercise*, Vol.32, No.9, pp.S481–S488 (2000).
25) Witten, I. and Frank, E.: *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann (2004).
26) Wu, J., Osuntogun, A., Choudhury, T., Philipose, M. and Rehg, J.: A scalable approach to activity recognition based on object use, *ICCV 2007*, pp.1–8 (2007).