

# ウェアラブルカメラに基づく日常生活における注意行動認識

## Recognizing Attentional Action based on using the Wearable camera in Daily life.

大北 拓哉<sup>\*1</sup>  
Takuya Okita

土井 健輔<sup>\*1</sup>  
Kensuke Doi

大井 翔<sup>\*2</sup>  
Sho Ooi

佐野 睦夫<sup>\*1</sup>  
Mutsuo Sano

<sup>\*1</sup> 大阪工業大学情報科学部

Faculty Information Science and Technology, Osaka Institute of Technology

<sup>\*2</sup> 大阪工業大学情報科学研究科

Major in Information Science and Technology, Graduate School of Osaka Institute of Technology

For supporting the elderly and cognitive disabled patients, it is necessary to recognize their attention behavior such as focused/ selective/ divided/ controlled attention, and navigate their daily lives. This paper proposes an algorithm to classify and recognize attention behaviors in typical daily lives by using a wearable camera. We verify the effectiveness of the proposed algorithm through the experiments of detecting the focused attention in a specific situation of listening/ speaking/ reading/ seeking out /eye-tracking/ walking.

### 1. はじめに

近年、認知症患者は増加傾向にあり、厚生労働省によると2012年で462万人と推定され、予備軍とされるMCI(軽度認知障害)400万人を合わせると65歳以上の高齢者の約7人に1人が認知症であると報告されている。また2025年には65歳以上の認知症患者は、高齢者の約5人に1人にあたる約700万人になると推定される[1]。つまり、認知症患者に対する支援の重要性は大きいと考えられ、当研究室ではこれまでに認知リハビリテーションとして効果が高い調理の行動ログ取得と認知評価を行っていたが、リアルタイム性が低かった[2]。

そこで本研究では認知機能評価のうち注意機能に着目した。具体的には、スマートグラスを用いてリアルタイムに取得したセンサ情報から日常生活における行動の不注意行動の認識を行う方式の検討を行った。また、スマートグラスで得られた画像データを使用したリアルタイムの認識を用いてセンサ情報だけよりも認識率が向上するかについて実験を行う。

日常生活の注意行動認識を行うことで、生活の中での認知評価ができ、将来的には不注意になる前に検知し事故を未然に防ぐことや、不注意状態を認識し必要な時に自動でのナビゲートする。また、未然に検知することで危険行動を防ぐことで認知症患者がチャレンジできることが増え、認知リハビリテーションの円滑化などが期待できる。

### 2. 提案方式

本研究で提案する方式はスマートグラスをクライアント、ノートPCをサーバとするクライアントサーバ型のシステムである。今回は注意行動の認識を慣性センサ(加速度・ジャイロ情報)のみを用いた場合と慣性センサ情報に画像情報を付加した場合の2種類の認識方式を提案する。

### 2.1 データ送受信の方法

スマートグラスに搭載されているセンサ情報(ジャイロセンサ、加速度センサ)とカメラから得られる画像(320px×240px)をサーバへ20fpsの速度で送信する。図1にスマートグラス側(クライアント側)、図2にサーバ側の処理フローを示す。

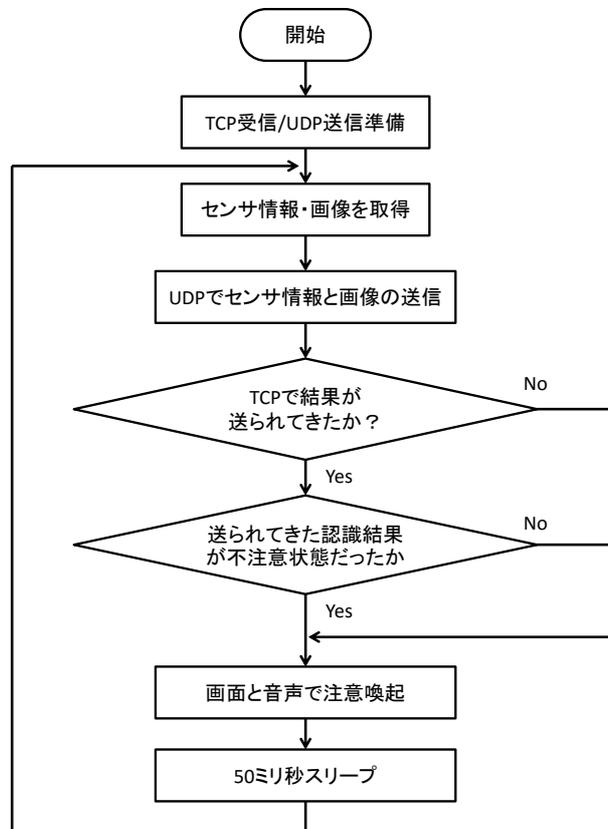


図1 クライアント側フロー図

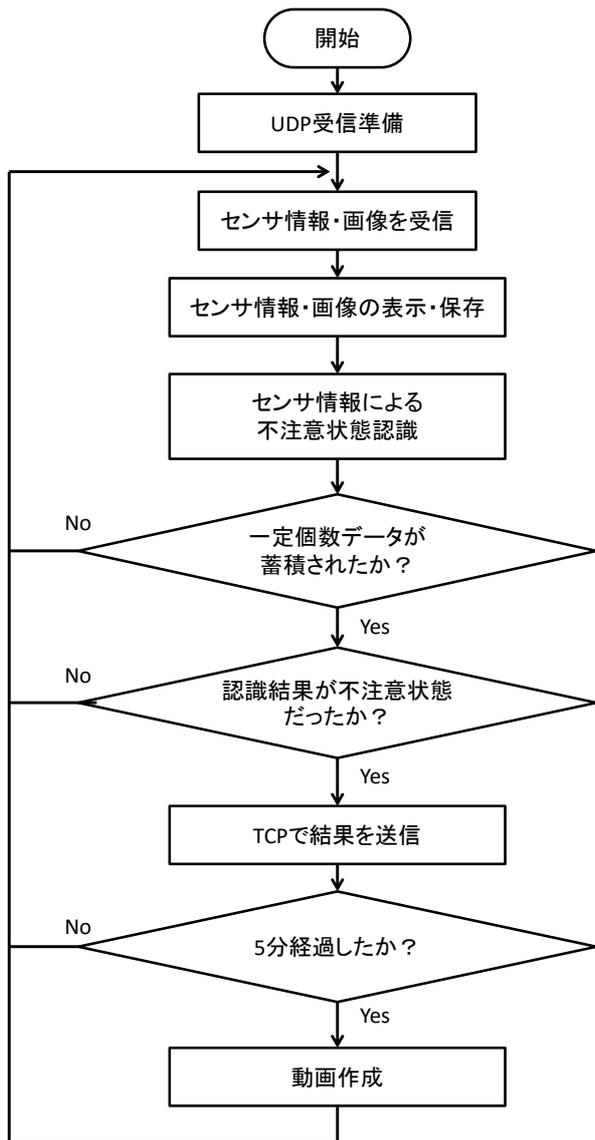


図2 サーバ側フロー図

今回画像サイズを 320px×240pxとしたのは処理時間を短縮するためであり、画像およびセンサ情報は UDP で送信を行った。ここで TCP ではなく UDP を使用した理由は、通信速度が優れており、大量の画像を送信する中で、数枚の画像が正しく順番通りに送られるという冗長性を重視する必要がないからである。

## 2.2 センサ情報を用いた不注意行動認識手法

センサ情報(ジャイロセンサ X, Y, Z 座標, 加速度センサ X, Y, Z 座標の 3 軸ずつ)を用いた行動識別の手法として田淵らの日常行動識別[3]や、濱らの手衛生活動認識[4]があり、本研究でも同様に、平均 $\bar{x}$ , 標準偏差 $\sigma_x$ , エネルギー $E_n$ , 周波数領域エントロピーFDE を 3 軸それぞれに対して計算した 12 特徴量と XY 座標, YZ 座標, XZ 座標それぞれの相関係数  $\text{corr}(x, y)$  を特徴量とした計 15 特徴量を算出する。各特徴量の求め方を式(1)~(6)に示す。なお、ある一軸の FFT の成分を  $F_1, F_2, F_3, \dots, F_n$  とする。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

$$E_n = \frac{1}{n} \sum_{i=2}^n |F_i|^2 \quad (3)$$

$$p(i) = \frac{|F_i|^2}{\sum_{j=2}^n |F_j|^2} \quad (i = 2, 3, \dots, n) \quad (4)$$

$$\text{FDE} = - \sum_{i=2}^n p(i) \log p(i) \quad (5)$$

$$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} \quad (6)$$

ここで、加速度とジャイロの値に大きな差が存在すること、特徴量ごとの値の大きさにばらつきがあることから、今回得られた各特徴量を-1 から+1 の範囲で正規化を式(7)で行った。このとき、特徴量全体の中の最大値を  $max$ , 最小値を  $min$ , 正規化する際の上限を  $upper$ , 下限を  $lower$  であり、今回  $upper$  が 1,  $lower$  が -1 である。

$$x_i^* = lower + (upper - lower) \cdot \frac{x_i - min}{max - min} \quad (7)$$

データの抽出方法として一定時間のスライディングウィンドウに分割し、ウィンドウごとに特徴量を求める手法[5][6]を用いる。今回は 2 回目以降、N 回目のデータ抽出で、あらかじめ決めたデータ数の内、半数を新規データ、残りを N-1 回目の抽出に用いたデータの後半部分を使用する。

また、ウィンドウサイズは FFT を行うためにデータ数が 2 のべき乗である必要がある。これを踏まえ今回の処理に使用するデータ数は 128 とする。これは、クライアント側が 1 秒間で 20 個のデータを送信しており、リアルタイム性を考慮し、1 分間に約 20 回の状態認識を行うことにしたためである。

不注意状態の認識手法として、k-NN(k 近傍)法を用い、距離尺度としてユークリッド距離で行った。

## 2.3 画像情報を付加した不注意行動認識手法

2.2 の方式に加えて、スマートグラスから送られてくる画像を付加し、不注意行動の認識率を向上させる。

画像情報を付加する具体例として、“言う”、“聴く”という動作状態では話す相手の顔が通常正面にあると想定される。つまり、相手の顔が画像中央に存在する場合、今行っている動作が“言う”、“聴く”である確率が高いと言える。今回顔検出の方法として Haar-Like 特徴量を用いた顔検出用の分類器を使用した結果を図 3 に示す。

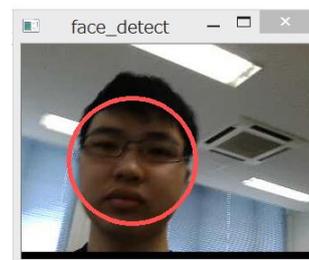


図3 顔検出の結果

また、“読む”動作は通常、本などの物体を支えることから手が視界の下半分にあり、ある程度本によって手が隠されていると考えられるので、手の位置、面積を測定し“読む”動作である確率が高いと判定する。図 4(a)にスマートグラスで取得した画像、

(b)に手を検出した画像を示す。今回、手の領域の検出としてHSVを使用した。

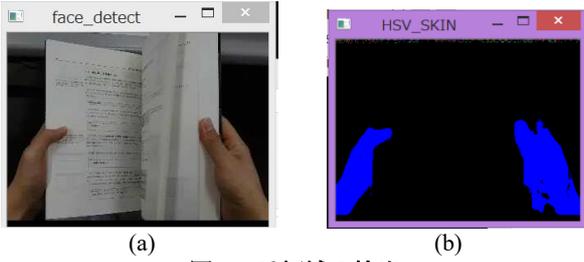


図4 手領域の検出

また、センサ情報のみでは動作の開始時間と終了時間が分からず、人によって時間・行動が決まっていない開始前、終了後の不要な動作行動まで含めて状態認識を行ってしまう可能性があるため、画像を用いて動作の開始と終了を認識できるようにする。

### 3. 実験

本研究では Android4.2.2 を搭載するウエストユニティ社のスマートグラスである InfoLinker を使用した。

#### 3.1 動作の定義

動作として、“読む”、“聴く”、“言う”、“動くものを追いかける”、“探す”、“歩く”の6動作について集中時・非集中時に分けて定義する。具体的な動作の定義を表1に示す。

表1 動作の定義

動作	状態	具体的な行動
(1) 読む	集中	あらかじめ指定した本について、読書を行った。
	非集中	あらかじめ指定した本を読むと同時に、PCかスマートフォンを操作することを指定した。
(2) 聴く	集中	椅子に座った状態かつ、一対一の対話形式で会話する状況を指定した。
	非集中	座った状態かつ実験参加者に対して話を始める側が2人以上での会話、もしくはPCかスマートフォンを操作して一対一で会話する状況を指定した。
(3) 話す	集中	座った状態かつ一対一の対面形式で会話する状況を指定した。
	非集中	座った状態かつ話を聞く側が2人以上、もしくはPC操作かスマートフォンを操作して会話するシチュエーションを指定した。
(4) 動くものを追いかける	集中	65型大型テレビに動画を映し、その正面から首を動かしながら動画内の動く丸を追いかけてもらった。
	非集中	映像は集中時と同じものを使用する。動く丸を追いかける際には、同時並行でPCを操作、あるいはスマートフォンを操作することを指定した。
(5) 探す	集中	テーブル(縦226×横120×高さ70)上の範囲で、事前に隠した物品を探してもらった。
	非集中	テーブル(縦226×横120×高さ70)上の範囲で、事前に隠した物品を探してもらった。探し物をする際には同時にPCを操作、あるいはスマートフォンを操作することを指定した。
(6) 歩く	集中	実験参加者には研究室を歩き回ってもらった。その際、歩くことに集中するように指示をした。
	非集中	実験参加者には研究室を歩き回ってもらった。その際、ふらふらと動き、周りの物に当たってしまうような歩き方かスマートフォンを操作しながら歩くことを指示した。

#### 3.2 実験方法

今回、3.1節で示した6動作2状態の計12動作に対して、大学生の男性4名、女性1名から各動作20個を取得した。つまり、各行動に対して100個となり、計1200個の学習データとなる。

次に、本手法の評価を行う上で大学生の実験参加者10名に対して12動作を20回ずつ行ってもらった。このデータを未知データ(計1200)として非リアルタイムで不注意状態の認識を行った。今回はジャイロセンサのみ、加速度センサのみ、ジャイロセンサと加速度の両方を用いた、3種類についてk-NN法でkを1~11までの奇数値により状態認識を行った。また、画像認識・センサ情報を用いる方式も同様に実験する。

#### 3.3 評価指標

認識結果として以下の表2に示すように4種類の組み合わせがある。未知データが集中行動である場合、それをシステムが集中と判別した場合はTrue Positive(TP)、未知データが非集中、システムが集中と判断した場合はFalse Positive(FP)、未知データが集中、システムが非集中と判断した場合はFalse Negative(FN)、未知データが非集中、システムが非集中と判断した場合はTrue Negative(TN)とする。

表2 未知データと判定結果の集中・非集中の組み合わせ

	実際の行動が集中	実際の行動が非集中
システムが集中と判断	True Positive	False Positive
システムが非集中と判断	True Negative	False Negative

これら4つの場合分けを使用して、集中・非集中時の精度、再現率、F値と正解率の計7項目を算出する。それぞれの算出方法を式(8)~(11)に示す。

$$\text{精度} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{再現率} = \frac{TP}{TP + FN} \quad (9)$$

$$F \text{ 値} = \frac{2 \cdot \text{精度} \cdot \text{再現率}}{\text{精度} + \text{再現率}} \quad (10)$$

$$\text{正解率} = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

### 4. 結果

加速度のみ、ジャイロのみ、加速度とジャイロの両方、それぞれに対して全体の正解率、精度、再現率、F値のそれぞれのkの値を変化させた時の最大値を表3に示す。

表3 センサ別の認識結果

	正解率	精度	再現率	F値
加速度のみ	0.625(k=11)	0.642(k=11)	0.691(k=1)	0.647(k=11)
ジャイロのみ	0.601(k=7)	0.607(k=7)	0.628(k=7)	0.612(k=7)
加速度+ジャイロ	0.669(k=5)	0.717(k=5)	0.779(k=5)	0.702(k=5)

表3よりセンサ別で比較してみると、ジャイロと加速度を組み合わせたものが正解率、精度、再現率、F値全てにおいて他の2つより認識率が高いことがわかった。

また、kごとのセンサ別正解率の最大値を求めたものが表4となる。なお、黄色で囲まれている箇所が最大値である。

表4 センサ別の最大認識率

		k=1	k=3	k=5	k=7	k=9	k=11
加速度のみ	集中	0.860	0.895	0.895	0.905	0.910	0.940
	非集中	0.560	0.585	0.625	0.635	0.640	0.660
ジャイロのみ	集中	0.570	0.595	0.665	0.725	0.705	0.720
	非集中	0.715	0.695	0.68	0.67	0.655	0.635
加速度+ジャイロ	集中	0.875	0.885	0.88	0.92	0.93	0.95
	非集中	0.705	0.68	0.69	0.695	0.7	0.705

表4よりk=11が高い正解率であることが分かる。このことから、k=11に着目する。k=11の場合のセンサ別の正解率に着目し、集中状態と非集中状態の行動別の結果を表5.6および図5.6に示す。

表5 集中時の行動別正解率

集中時(k=11)						
	(1)読む	(2)聴く	(3)話す	(4)追いかける	(5)探す	(6)歩く
加速度のみ	0.940	0.580	0.520	0.835	0.675	0.565
ジャイロのみ	0.720	0.600	0.625	0.465	0.660	0.595
加速度+ジャイロ	0.950	0.805	0.720	0.795	0.715	0.665

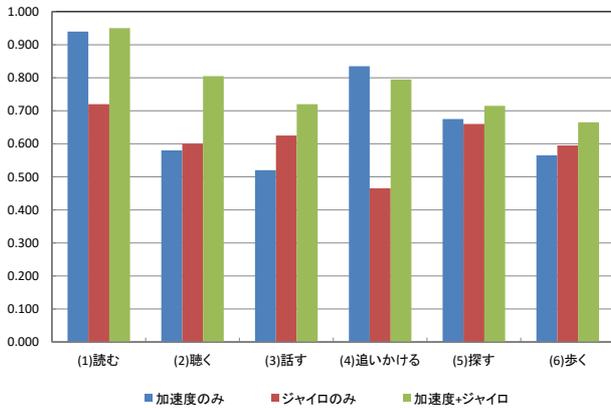


図5 集中時の行動別正解率

集中時の結果を見ると、加速度+ジャイロの組み合わせが全体的に65%以上の認識結果となった。また、追いかける動作以外は2種類のセンサの組み合わせが最も良い結果となった。

表6 非集中時の行動別正解率

非集中時(k=11)						
	(1)読む	(2)聴く	(3)話す	(4)追いかける	(5)探す	(6)歩く
加速度のみ	0.590	0.445	0.500	0.560	0.630	0.660
ジャイロのみ	0.595	0.480	0.585	0.635	0.500	0.620
加速度+ジャイロ	0.610	0.390	0.485	0.560	0.610	0.705

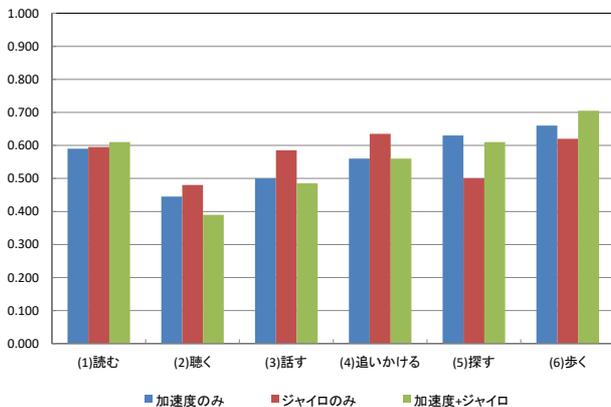


図6 非集中時の行動別正解率

非集中時の結果を見ると、センサによる大きな差は見られなかった。

画像認識を併用する方式は現在実験中であり、今後結果を報告したい。

## 5. 考察

表3で加速度のみだけで比較したときに再現率のみがk=1の時最大値となっている。これは各データが分散しておりkの値を上げた際、本来のデータと違うクラスに分類されやすいためと考えられる。また、加速度のみの正解率・精度・F値は今回

kの範囲の最大値である11の時に最大値を取っている。kの値を増やした場合、値が向上する可能性があるため次回からはkの上限を上げて測定するべきだと考えられる。

表5より、追いかける動作を除いた集中時には加速度・ジャイロセンサの両方を利用した場合が高い正解率であった。これは、追いかける動作が加速度のみが高くなった理由として他の動作と違いほとんど向いている方向が変わっていないことが理由であると考えられる。また、同じ座った状態での動作である”聴く”、”読む”、”話す”の中で”聴く”の認識率は加速度のみと加速度・ジャイロの組み合わせの場合が集中時とかなり高い数字を指している。しかし、同じセンサ情報を使用した残りである2つが0.52や0.625など小さい値が取られていることから本来”聴く”、”話す”と判定されなければならないデータまで”読む”と判断されていると考えられる。

表6の非集中状態ではセンサごとに大きな差が見られなかった理由として、個人ごとに非集中動作(不注意行動)に大きな差があると考えられる。これに対する対処として個人ごとの非集中動作(不注意行動)を網羅するような、非集中行動を行動の種類ごとに分類した多数の学習データが必要と考えられる。

## 6. まとめ

本研究ではスマートグラスに内蔵された加速度センサ・ジャイロセンサにより、6種類の基本的な生活行動時の集中状態を約8割の確率で認識することができた。一方で、非集中状態の正解率があまり高くなかった理由として、動作の開始と終了時間が判別できていなかったことが原因だと考えられる。この問題を解決するために、注意対象物の認識や手の認識などを用いて画像認識を使用し、注意動作の開始・終了のタイミングを認識することで認識率を向上させていきたい。

## 謝辞

本研究の一部は文部科学省研究費補助金(基盤C15K00368)の支援を受けた。

## 参考文献

- [1] 厚生労働省, 認知症施策総合戦略(新オレンジプラン)～認知症高齢者等にやさしい地域づくりに向けて～, 2015年.
- [2] 大井翔, 池ヶ谷剛, 渋谷咲月, 水野翔太, 佐野睦夫, 中山佳代, 大出道子, 萩原摩紀: “認知リハビリテーションのための一人称視点による生活行動の注意評価”, 情報処理学会 第77回全国大会講演論文集, 2015(1), 237-239, 2015.
- [3] 田淵勝宏, 納谷太, 大村廉, 野間春生, 小暮潔, 岸野文郎: 無線加速度センサを用いた人の日常行動識別におけるデータ収集条件の影響評価, 日本バーチャルリアリティ学会論文誌 Vol. 12, No. 1, pp. 25-35, 2007.
- [4] 濱恵美子, 納谷太, 野間春生, 小暮潔, 土田敏恵: 加速度センサを用いた手指衛生行動識別手法の検討, 電子情報通信学会技術研究報告 Vol108, No. 198, pp-235-240, 2008.
- [5] L.Bao and S.S.Intile: Activity recognition from user-annotated acceleration data, in Proc. of PER-VASIVE 2004, vol.LNCS 3001, A. Ferscha and F.Mattern (Eds.), Springer-Verlag, Berlin Heidelberg, pp.1-17, 2004.
- [6] N.Ravi, N.Dandekar, P.Mysore, and M.L.Littman: Activity recognition from accelerometer data, American association for artificial intelligence(www.aaai.org), 2005.