

動画解析に基づく自動くせ検出システムの基礎検討

見澤, 弘樹 / MISAWA, Hiroki

(出版者 / Publisher)

法政大学大学院理工学・工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学・工学研究科編 / 法政大学大学院紀要. 理工学・工学研究科編

(巻 / Volume)

57

(開始ページ / Start Page)

1

(終了ページ / End Page)

4

(発行年 / Year)

2016-03-24

(URL)

<https://doi.org/10.15002/00013291>

動画解析に基づく自動くせ検出システムの基礎検討

Automated habit detection system: A feasibility study

見澤 弘樹

Hiroki MISAWA

指導教員 彌富 仁

法政大学大学院理工学研究科応用情報工学専攻修士課程

In this paper, we propose an automated habit detection system. We define a “habit” in this study as some motion those are significantly different from our common behaviors. Behaviors of two subjects during conversation are tracked by the Kinect sensor and their skeletal and facial conformations are detected. The proposed system detects the motions considered as habits by analyzing them using a principal component analysis (PCA) and wavelet multi-resolution analysis (MRA). In our experiments, we prepare a total of 108 movies containing 5 min of conversation. Of these, 100 movies are used to build the average motion model (AMM), and the remainder are used for the evaluation. The accuracy of habit detection in the proposed system is shown to have a precision of 84.0% and a recall of 81.8%.

Key Words : Kinect, Motion analysis, PCA, MRA

1. はじめに

人は主に言語でコミュニケーションをとるが、表情や身振り手振り、無意識的な行動などの非言語コミュニケーションが意思疎通や対人関係に大きな影響を与えていることが知られている [1]。くせとは人が無意識のうちにいう習慣的な行動であり、自覚症状がない場合が多い。自分のくせが知らない間に相手に不快感を与えていることや、これによりさまざまな機会を喪失している可能性もある。したがって、自分のくせを客観的に認識することは、コミュニティ内の円滑なコミュニケーションの実現のためだけではなく、さまざまな面において意義があると考えられる。

人体の動きを解析し、さまざまな用途に応用する研究が成されてきている。体の動きを捉えるための方法は大きく次の2つの方法に分類される。：(1) 被験者にセンサデバイスを取り付けて、体の動きを取得する方法。(2) 被験者にセンサデバイスなどを付けずに画像処理技術を用いて、体の動きを取得する方法。前者では、勾配や動き、ゆらぎなどの検出に優れている加速度センサが多く用いられている [2]。しかし、これらのセンサは動作解析において有効な情報を取得できるが、コストや重量の課題があり状況によっては利用が難しい。後者では、市販のビデオカメラが広く使用されている。動画像から人のシルエットの変化として binary motion energy image や motion history image を生成し、人の動作を解析する方

法 [3] や、局所特徴量のヒストグラムと support vector machine(SVM) を用いて人の動作を識別する方法 [4] がある。赤外線センサは、照明の影響を受けにくいなどのメリットがあり、単体、もしくは可視光センサと併用して用いられている。しかし、これらの方法において3次元の動作に対しては、複数のカメラを用いる等で奥行きを推定する必要がある。また、高精度な計測を行う場合、被験者の体の各部位にマーカーを配置し、これらを検出するなど、相応の装置や施設を必要としてきた。

Kinect センサは 2010 年に Microsoft から発売された家庭用ゲーム機 Xbox の周辺機器である。Kinect センサは比較的安価であり、精度良く3次元動作解析が可能であるため、これを応用する研究が近年盛んに行われている [5-7]。Xia ら [5] は、HOJ3D(Histograms of 3D joint locations) という方法を用いて3次元ヒストグラムを作成し、線形判別分析とベクトル量子化を行うことで動作を解析した。Evangelidis ら [6] は、フィッシャーベクトルと SVM を用いて動作を識別した。Miranda ら [7] は、SVM と決定木を用いて動作を解析した。しかし、本研究で注目している「人のくせ」についての動作解析の研究は私たちの知る限り行われておらず、またその動きは個々でさまざまであるため、従来の解析方法の適応が難しい。

本研究では、Kinect センサを用いた自動くせ検出システムを提案した。このシステムの更なる適用可能性を考慮し、対話中の人の動作を解析した。

2. くせ検出システム

本研究では、普段の動きから大きく異なる動作をくせとして定義した。提案したくせ検出システムは対話中の被験者2人の動きを Kinect センサで捉え、独特の動きをくせとした。くせ検出システムの概要図を Fig.1 に示す。

提案手法は学習段階と評価段階の2つの処理に分けられる。学習段階では、Kinect センサを用いて多くの対話の動画を録画し、各部位の一般モデルを形成した。評価段階では、被験者の動きを一般モデルと比較し、大きく異なった動作をくせとして検出した。提案手法は、被験者の各部位の時系列速度ベクトルを計算し、主成分分析ならびに多重解像度解析を用いて解析した。

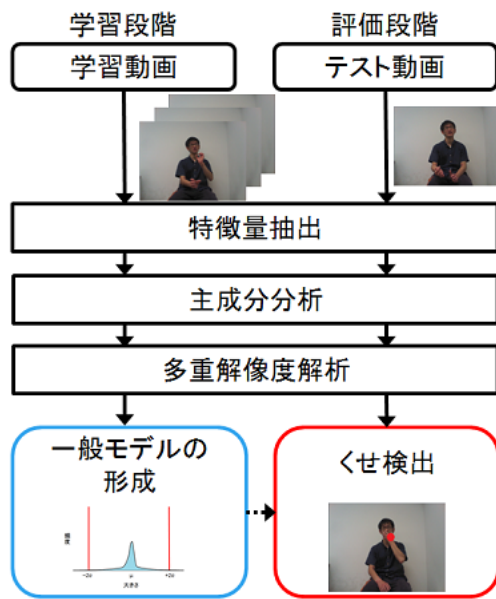


Fig.1 くせ検出システムの概要図

(1) 録画環境

本研究では、以下のような流れで動画撮影・録画を行った。(1) 被験者2人が向き合いながら椅子に着席する。(2) 被験者の上半身を検出するために Kinect センサをそれぞれの被験者の前に設置する。(3) Kinect センサを用いて対話を 300(s) ほど録画する。このとき Fig.2 のように、Kinect センサと被験者の距離は約 1m とし、被験者同士の距離は約 2.5m とした。いくつかの対話の議題 (例: 学校生活、趣味、友人関係) を準備し、これらの中で被験者同士が話しやすいものを録画する前に選択した。

本研究では、被験者男性 36 人 (24.3 歳 $\pm$ 1.3 歳) から 1 人 3 回動画の撮影を行い、合計 108 編、各 300(s) の動画を解析に用いた。Kinect Studio(SDK 1.5) を用いて 3 次元骨格情報を検出し、30fps で会話を録画した。Kinect センサの RGB と Depth の解像度は、それぞれ



Fig.2 実験環境

640 × 408 と 320 × 240(30fps) である。

(2) 特徴量抽出

本研究では、対話中の人物の上半身骨格特徴量ならびに、顔特徴量を用いた。前者では、座った状態の人物の上半身骨格特徴量 10 点を検出した (Fig.3)。後者では、121 点取得できる顔特徴量のうちの 5 点 (額、左目、右目、鼻、口、Fig.4) を検出した。計 15 の特徴点、計 45 の特徴量 (15 × 3 次元) をそれぞれの被験者の画像フレームから抽出した。各部位のフレーム間の差分を速度ベクトルとして計算した。したがって、時刻 t における各部位 $p(p = 1, 2, \dots, 15)$ の 3 次元速度ベクトルは次のように表される:

$$\mathbf{v}^p(t) = [v_x^p(t), v_y^p(t), v_z^p(t)]^T \quad (1)$$

人の動きは身体的、幾何学的に制限があるため、各部位の速度ベクトルに主成分分析を用いた。各部位での第 1 主成分の寄与率を Table 1 に示す。前実験の結果から、多くの部位において第 1 主成分の寄与率が高いことが確認できた。したがって、第 1 主成分方向の $v_{1st}^p(t)$ のみを用いて、3 次元ベクトル $\mathbf{v}^p(t)$ を取得した。本研究では、15 次元の時系列データ

$$\mathbf{V}(t) = [v_{1st}^1(t), v_{1st}^2(t), \dots, v_{1st}^{15}(t)]^T \quad (2)$$

に Daubechies Wavelet(N=2) の多重解像度解析を用いて解析した。前実験の結果、各部位の周波数が 1.25 Hz の Wavelet 係数に着目した。

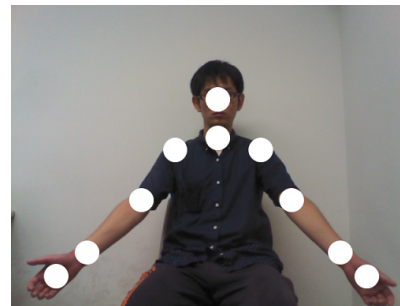


Fig.3 上半身骨格特徴量 10 点

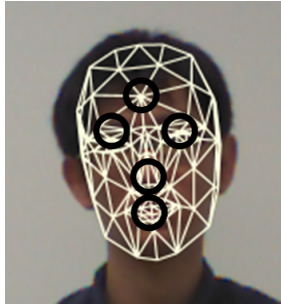


Fig.4 顔特徴量 5 点

Table 1 各部位の第 1 主成分の寄与率

部位	寄与率 (%)	部位	寄与率 (%)
頭	82.6	肩中央	79.6
左肩	81.1	右肩	79.7
左肘	64.7	右肘	65.3
左手首	45.4	右手首	45.9
左手	47.6	右手	50.6
額	88.1	鼻	88.2
左目	89.1	右目	89.4
口	87.9		

(3) くせ検出の定義

学習段階では、学習データセット (動画 100 編) の Wavelet 係数の平均から各部位の一般モデルを形成した。評価段階では、それぞれの被験者の動きから Wavelet 係数を計算した。評価データと一般モデル間の違いが標準偏差の 2 倍以上離れている場合、普通ではない動き、つまりくせとして定義した。

提案システムの評価をするためには、gold standard が必要である。身体の動きの中でくせとして定義するものはないので、著者が評価データからくせとして考えられる動作を選び、それを gold standard とした。性能評価として、precision と recall を用いた。これらは次のように計算される。

$$\text{precision} = \frac{\text{正しく検出できたくせ}}{\text{提案手法によって検出されたくせ}} \quad (3)$$

$$\text{recall} = \frac{\text{正しく検出できたくせ}}{\text{人の主観によって検出されたくせ}} \quad (4)$$

提案システムと gold standard によって検出した時間が 2(s) 以内の場合、正しく検出できたと考える。

3. 結果

本研究では、合計動画 100 編を用いて一般モデルを形成し、残りの動画 8 編を評価データとした。そのうち 77

の動作を gold standard とした。このとき形成した頭と右手の一般モデルのヒストグラムを Fig.5 と Fig.6 に示す。Fig.5 と Fig.6 はいずれもヒストグラムの中心が大きく盛り上がっていた。他の各部位の一般モデルもこれらのヒストグラムとほとんど同じく中心が大きく盛り上がっていた。これはほとんどの動作が中心に集中しており、標準偏差の 2 倍以上離れた動作はかなり特殊といえる。また、Fig.5 のヒストグラムは μ 周辺の勾配に対して、Fig.6 のヒストグラムは勾配が広がっていた。ヒストグラムが広がっているということは、頭より右手のほうが動きが多くあることが言える。

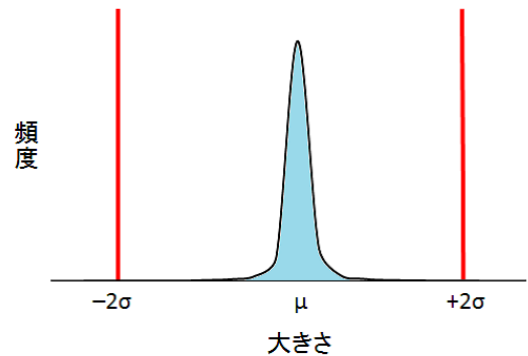


Fig.5 頭の一般モデルのヒストグラム

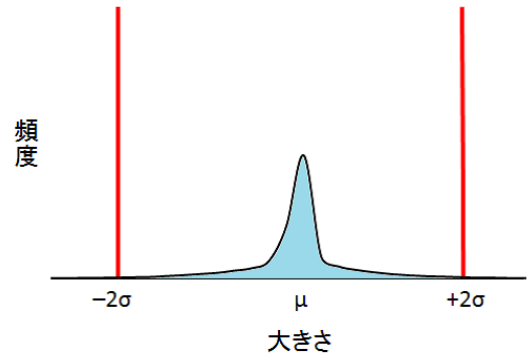


Fig.6 右手の一般モデルのヒストグラム

くせ検出例を以下に示す。Fig.7 は提案システムと gold standard のくせ検出結果のダイアグラムの例を示す。Fig.8 と Fig.9 は true positive の例、Fig.10 は false positive の例、Fig.11 は false negative の例を示す。提案システムでは、くせがあると考えられる部位に明るい丸を表示する。Fig.8 と Fig.9 では、被験者が右手で鼻を掻く動作 (10(s), 6.1SD) と体を後ろにそらす動作 (90(s), 2.4SD) をくせとして検出した。これらの動作は一般モデルから 2SD 以上離れているためくせとして検出した。Fig.10 では、提案システムが普通に見える動作 (60(s), 2.1SD) を間違ってくせとして検出

した例である。反対に、Fig.11 は提案システムがゆっくり前のめりになる動作 (190(s)、0.6SD) をくせとして検出できなかった例である。結果の confusion matrix を Table 2 に示す。提案システムのくせの検出精度は precision=84.0%、recall=81.8% となった。

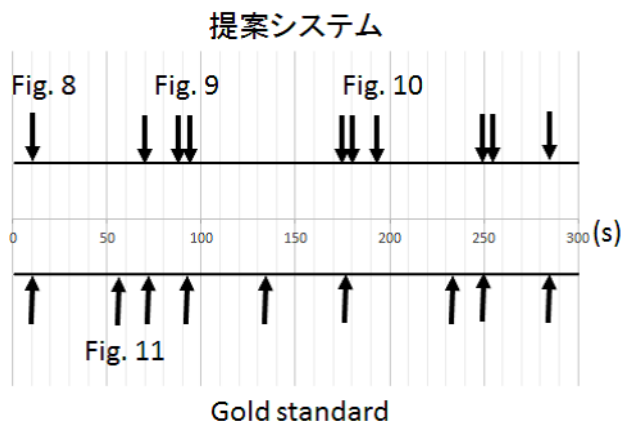


Fig.7 くせ検出結果のダイアグラム



Fig.8 true positive
の例 1 (6.1SD)

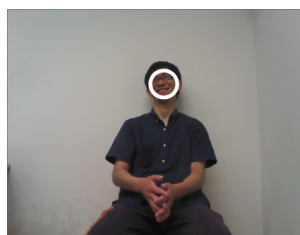


Fig.9 true positive
の例 2 (2.4SD)

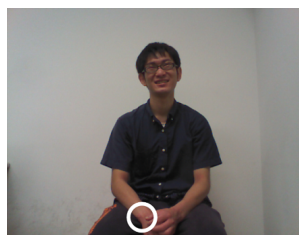


Fig.10 false positive
の例 (2.1SD)

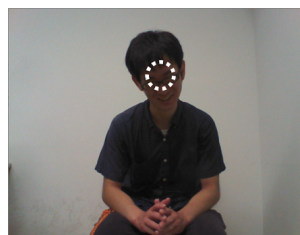


Fig.11 false negative
の例 (0.6SD)

Table 2 くせ検出の confusion matrix

提案システム	Gold standard		
		くせ	非くせ
	くせ	63	12
	非くせ	14	

4. 考察

本研究では、提案システムは比較的良好なくせの検出性能を実現できたことから、人のくせを検出する性能があることが示唆された。しかし、改善すべき問題がいくつかある。本研究では、普通でない動きのみをくせとしたが、今後はくせの速度や周期性、他の人の主観評価も考慮し改善していきたい。

5. 結論

本研究では、自動くせ検出システムのプロトタイプを提案した。対話中に被験者の動画 108 編を用いて、precision=84.0%、recall=81.8% というくせ検出性能を実現した。今後、提案手法の性能と精度の改善に注目していきたい。

参考文献

- [1] Merabian, A.: Communication without words. *Psychological Today* **2** (1968) 53–55
- [2] Sagawa, K., Abo, S., Tsukamoto, T., Kondo, I.: Forearm Trajectory Measurement during Pitching Motion using an Elbow-mounted Sensor. *Journal of Advanced Mechanical Design, Systems and Manufacturing* **3** (2009) 299–311
- [3] Bobick, A.F., Davis, J.W.: The Recognition of Human Movement Using Temporal Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23** (2001) 257–267
- [4] Schuldt, C., Laptev, I., Caputo, B.: Recognizing Human Actions: A Local SVM Approach. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)* **3** (2004) 32–36
- [5] Xia, L., Chen, C.C., Aggarwal, J.K.: View Invariant Human Action Recognition Using Histograms of 3D Joints. *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2012) 20–27
- [6] Evangelidis, G., Singh, G., Horaud, R.: Skeletal Quads: Human Action Recognition Using Joint Quadruples. *2014 22nd International Conference on Pattern Recognition (ICPR)* (2014) 4513–4518
- [7] Miranda, L., Vieira, T., Martinez, D., Lewiner, T., Vieira, A.W., Campos, M.F.M.: Online gesture recognition from pose kernel learning and decision forests. *Pattern Recognition Letters* **39** (2014) 65–73