

Selfie WanD: 自撮り棒を動かすことによる 撮影用入力インタフェース

澤田 直春¹ 山本 匠¹ 雨坂 宇宙¹ 杉浦 裕太¹

概要:近年、自撮り棒が普及して、より高い位置からの撮影や広い画角での撮影を実現する。しかし、自撮り棒に取り付けたデバイスの操作性の低さや、自撮り棒と撮影デバイスの接続の煩雑さといった課題もある。そこで、本研究では自撮り棒を用いたジェスチャ入力インタフェースを提案する。自撮り棒全体を動かすジェスチャで操作することにより、スマートフォンに直接触らない操作と複数の異なる操作を可能にする。システムにはスマートフォンに内蔵された加速度センサとジャイロセンサを用いるため、自撮り棒とスマートフォンの接続や自撮り棒の別電源も必要がない。システムを評価するため、8種類のジェスチャを用いて識別精度評価実験を実施した。

1. はじめに

近年、カメラやソーシャルネットワーキングサービス (SNS) の普及によって自撮りの機会が増加している。ほとんどのスマートフォンユーザは自撮りをした経験があり、若いユーザ間では自撮り棒の人気も高まっている [1]。自撮り棒とは先端にスマートフォンやカメラを取り付けて、自分自身や背景を含めて写真・動画を撮影するための道具である。取り付けたデバイスを広い範囲で動かすことができるため、より高い位置からの撮影や広い画角での撮影が可能になる。

しかし、自撮り棒を使った撮影には課題もある。まず、スマートフォンが自撮り棒の先に取り付けられているため、身体の近くで操作することが難しい。カメラのタイマー機能を使ってシャッター制御ができるが、画角調整や、静止画と動画の切り替えなど、撮影に必要な設定は変更できない。ボタンやリモコンでスマートフォンを遠隔で操作できる自撮り棒もあり、イヤホンジャックを利用した有線接続や、Bluetooth を利用した無線接続で実現される [1]。しかし、有線接続ではシャッター制御以外の操作が行えず、撮影に十分な操作性がない。無線接続ではスマートフォンと自撮り棒のペアリングや、自撮り棒に別の電源が必要になってしまう。

そこで、本研究では自撮り棒を動かすことによるジェスチャ入力インタフェースを提案する (図 1)。自撮り棒全体を動かして表現されるジェスチャで、棒の先に取り付けられたスマートフォンを操作する。ジェスチャを入力に利



図 1 自撮り棒を動かすことによるジェスチャ入力インタフェース

用することで、スマートフォンに直接触らずに操作でき、シャッターの制御だけでなく複数の異なる操作を実現できる。また、ジェスチャの識別にはスマートフォンに内蔵された加速度センサとジャイロセンサを用いる。これにより遠隔操作のために自撮り棒とスマートフォンを接続したり、自撮り棒用の別電源を用意する必要がない。システムを評価するため、6人の参加者に8種類のジェスチャを用いて識別精度評価実験を実施したところ、個人内の平均識別精度は97.3%、個人間の平均識別精度は90.8%であった。

2. 関連研究

2.1 自撮りと自撮り棒に関する研究

自撮りと自撮り棒は複数の側面から研究されている。Arifら [1] は市場で入手できる一般的なタイプの自撮り棒についてレビューと分類を行い、市販のほとんどの自撮り

¹ 慶應義塾大学

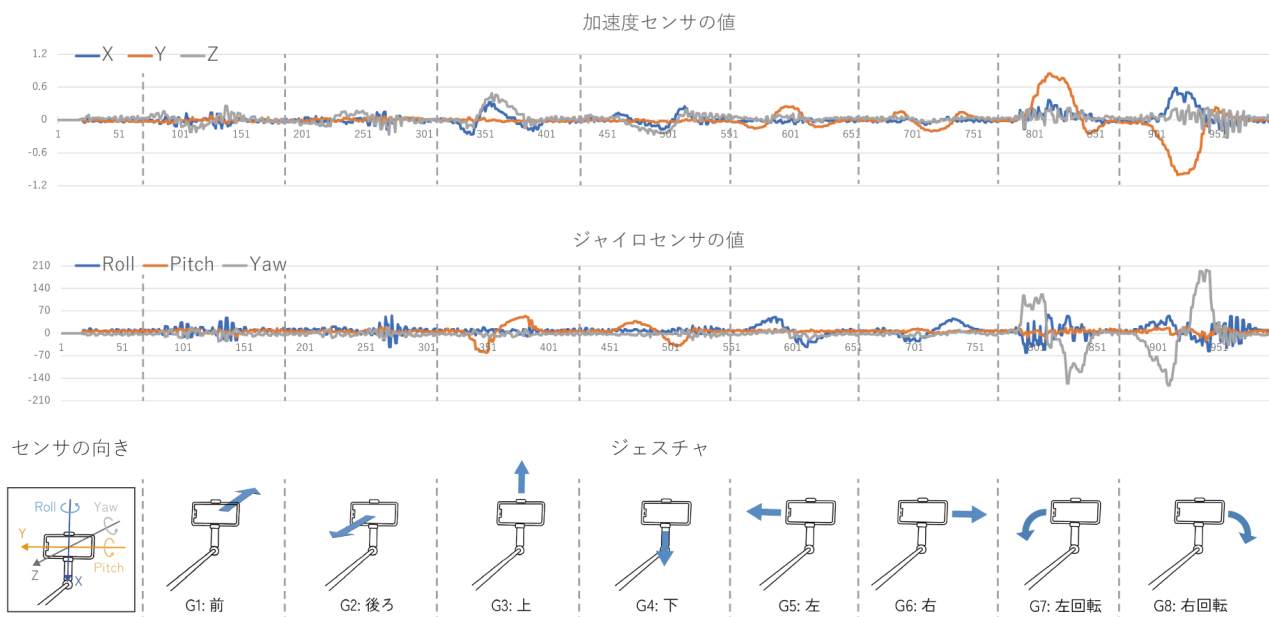


図 2 ジェスチャ時の加速度センサとジャイロセンサの値の変化

棒は人間工学に基づかず、ユーザに短絡的な疲労を引き起こすことを明らかにした。自撮り棒の使用行動、ユーザにとっての長所と短所、既存の自撮り棒の使い勝手を改善する方法についての見解も明らかにしている。Flaherty と Choi [2] は自撮り特有の状況認識の欠落や注意が散漫になることによって、旅行者が危険に晒されることを報告している。Svelander と Wiberg [3] はインタラクション技術、社会的プロセス、心理学の組み合わせで自撮り行為を考察した。

自撮り棒を利用した際の撮影機能を応用した研究もある。Cricito と Stein [4] は体の背面の皮膚の自己検査に自撮り棒を用いたシステムを提案した。自撮り棒を利用することによって、背中など体の見えにくい部分を見えるようにし、クローズアップして検査することを実現した。

本研究では自撮り棒を用いて撮影をする際、自撮り棒自体を入力インターフェースとして利用する。

2.2 カメラ操作やスマートフォン内蔵センサに関するジェスチャ入力

ジェスチャを認識して、入力インターフェースとして利用する研究が多くある。Chu と Tanaka [5] は自撮り時のカメラとの自然なインタラクションを提供するためにハンドジェスチャでカメラを操作する手法を提案した。手を振る、8方向の選択、円の描画の3種類のジェスチャを認識することで、デジタル一眼レフカメラを制御する。システムの評価実験を行ったところ、ジェスチャインターフェースによる入力の方がセルフタイマーより使いやすいことが分かった。Liou ら [6] は手指の動きやジェスチャで自撮りを可能にする手に装着するデバイスを提案した。3つの加速度セ

ンサを用いて、手指の動きやジェスチャを認識し、スマートフォンのカメラを操作した。Patil ら [7] は視覚障害者が既存の白杖でジェスチャを行い、スマートフォンを操作するためのクリップ型デバイスを提案した。ナビゲーションのために白杖を使用している人はスマートフォンなどのデバイスに同時にアクセスすることが困難であるため、タッチフリーの新しいインタラクションを検討した。Hursale ら [8] はスマートフォンに内蔵された加速度センサを利用してジェスチャを認識した。ユーザはスマートフォンをペンのように持ち、図形や文字のジェスチャを行う。

本研究は自撮り棒の先に取り付けたスマートフォンでジェスチャを認識し、撮影に必要なデバイス操作を行う。

3. 提案手法

本システムは自撮り棒に取り付けられたスマートフォンを用いて、撮影時に自撮り棒を動かさずジェスチャ入力を認識する。ジェスチャはスマートフォンに内蔵された加速度センサとジャイロセンサを利用して識別する。図 2 はジェスチャ時の加速度センサの値とジャイロセンサの値である。動きに応じてセンサが異なる反応をしていることがわかる。センサ値の違いを利用して、Python による機械学習でジェスチャ時のセンサ値を分析する。本研究では Python を利用するために取得したセンサ値を PC に転送するが、スマートフォンに内蔵されたセンサで自撮り棒のジェスチャ入力を実現する可能性を示すものである。スマートフォンは Huawei P30 lite, PC は ASUS ExpertBookB9 B9450FA(B9450FA-BM0295TS) を使用した。自撮り棒は全長が 71 cm, 重量が 140 g のものを利用した。自撮り棒にはスマートフォンを左向きに倒し、横向きに取り付けた。

また、ジェスチャ時のセンサ値は Unity で作成したアプリケーションで取得した。

本システムのジェスチャ識別について説明する。まず、各ジェスチャでは約 50 fps で 180 フレーム分のセンサ値を取得する。加速度センサとジャイロセンサからはいずれも 3 次元のセンサ値を取得する。取得した時系列データは 10 個のセグメントに分割し、それぞれのセグメントで統計的特徴量を抽出する。統計的特徴量には平均、分散、最小値、最大値を用いた。最終的に 1 回のジェスチャ毎に 240 次元 (6 次元 × 10 個 × 4 統計的特徴量) の特徴量を抽出した。その後、PC で Python の scikit-learn ライブラリにあるランダムフォレストを用いて識別した。

4. 実験

4.1 実験概要

システムを評価するために、自撮り棒を用いた 8 種類のジェスチャ (図 3) の識別精度を調査した。ジェスチャセットは、デバイス操作で想定される動作を元に、簡単に操作できるジェスチャから決定した。実験参加者は初めに 8 種類のジェスチャを練習した。その際、初期位置は自撮り棒が地面と水平になる状態として、スマートフォンの中心から約 20 cm の範囲でジェスチャするように指定した。すべてのジェスチャを 1 回ずつ順番に取得し、各ジェスチャが 10 回ずつ取得された時点で実験を終了した。1 回のジェスチャ取得では、実験参加者は画面で動作開始が示されたことを確認したタイミングでジェスチャを開始した。実験参加者は 6 人 (男性 : 5 人, 女性 : 1 人), 平均年齢は 23.7 歳 (SD = 2.7 歳) であった。

取得したデータから、個人モデルと Leave One Out Cross Validation (LOOCV) モデルの 2 つのモデルから平均識別精度を求めた。個人モデルは、1 人の実験参加者のデータを学習データとテストデータで 9:1 に分割して 10 分割検証を行う。全ユーザのデータで識別精度を求め、平均識別精度を算出する。LOOCV モデルは 5 人のデータを学習データ、残った 1 人のデータをテストデータとして検証する。全ての人をテストデータとして利用して識別精度を求め、平均識別精度を算出する。

4.2 結果

図 4 に実験結果を示す。個人モデルの平均識別精度は 97.3% (SD=2.0%) であった。LOOCV モデルの平均識別精度は 90.8% (SD=5.0%) であった。

個人識別モデルの結果から、個人内で 8 種類のジェスチャを識別できることが分かった。LOOCV モデルから、学習データに存在しない個人のジェスチャも約 90 % の精度で識別できることが分かった。しかし、LOOCV モデルでは自撮り棒を前に動かすジェスチャ (G1) と後ろに動かすジェスチャ (G2) を誤って認識することが多かった。こ

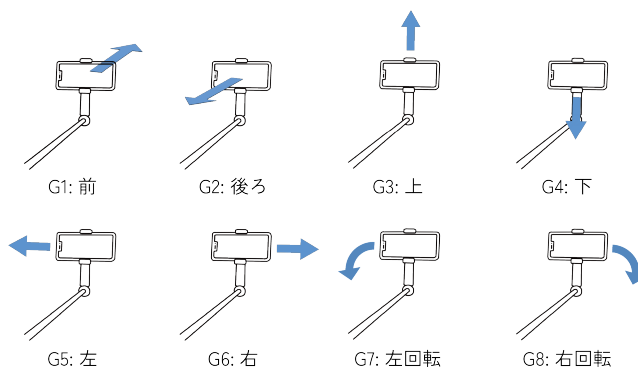


図 3 実験で使用した 8 種類のジェスチャ

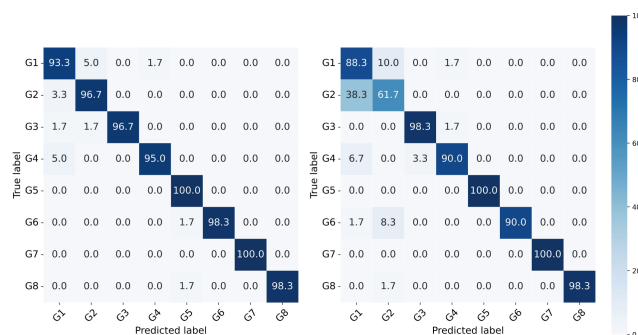


図 4 実験結果 (左: 個人モデル, 右: LOOCV モデル)

れは、図 2 のセンサ値のグラフから分かるように前後する動作は他のジェスチャに比べて値の変化量が少なく、ジェスチャ開始時間やジェスチャ時間の個人差が影響してしまったことが原因として挙げられる。

5. 議論と制約・今後の課題

5.1 制約

本研究では、自撮り棒にスマートフォンを左向きに倒して取り付けた。その状態で取得したセンサ値からジェスチャ識別モデルを作成しているため、スマートフォンが他の角度で固定されている場合はモデルを適用できないという制約がある。

また、今回はスマートフォンに内蔵されたセンサを利用してジェスチャを識別した。しかし、自撮り棒に取り付ける撮影デバイスは一眼レフやアクションカメラなども想定され、センサを内蔵していないデバイスではジェスチャ識別のために追加センサが必要になってしまう。

5.2 ジェスチャ識別のロバスト性

実験では自撮り棒を地面と水平にした状態からジェスチャを開始したため、ジェスチャ入力時の条件が制限されていた。実際に撮影をしている時には、ジェスチャ開始時の自撮り棒の状態は様々である。ジェスチャ時の地平面に対する角度が違くと加速度センサとジャイロセンサの反応も異なる。そのため、ジェスチャ時の自撮り棒の多様な状態に対応できるようなセンサ値処理やモデル構築が必要に

なる。他にも、自撮り棒の形状、ユーザの体格、ジェスチャ時の速度、ユーザの歩行状態など、センサ値に影響がある状況を想定した実装が必要である。

5.3 ジェスチャ入力撮影に与える影響

本研究では、自撮り棒を使って撮影するシーンでのジェスチャ入力を想定している。シャッター制御や静止画・動画の切り替え、画角調整といった設定変更などである。しかし、ジェスチャ入力で自撮り棒を動かすことによって、カメラ映像も変化してしまうため撮影フローを妨害する可能性がある。そのため、撮影時のユーザに適切なジェスチャを設計する必要がある。

そこで、今後は Elicitation Study を実施して、自撮り棒を使ったユーザ定義のジェスチャを決定する。Wobbrockら [9] が実験参加者に対象の操作に適したジェスチャ提案させる Elicitation Study を実施してから、多くの研究が Elicitation Study によってジェスチャを決定している。今後の研究では自撮り棒について、撮影時の操作に適したジェスチャを探索する。

5.4 今後の課題

本研究では PC を使ってセンサ値からジェスチャを識別している。そのため、自撮り棒の先に取り付けられたスマートフォンのみで操作が可能なシステムは構築できていない。今後はスマートフォンで完結する、ジェスチャを識別し操作に反映させるアプリケーションを実装する。また、ジェスチャの識別については、先述した様々な状態に対応できるロバスト性を今後獲得する。それに加え、実際の利用で想定される多様な動きに対して、撮影時の通常の動きなのかジェスチャなのかを検出するシステムも実装する。

6. 結論

本研究では、自撮り棒を動かすことにより撮影操作をする入力インタフェースを提案した。スマートフォンのセンサでジェスチャを識別するシステムを評価するために、8種類のジェスチャを用いて6人の実験参加者に識別精度評価実験を実施した。実験の結果、個人内の平均識別精度は97.3%、個人間の平均識別精度は90.8%であった。今後は、実際の利用を想定した十分なジェスチャ識別能力を実装し、スマートフォンで完結したアプリケーションを作成する。

参考文献

[1] Arif, A. S., Kim, S. and Lee, G.: Usability of Different Types of Commercial Selfie Sticks, *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3098279.3098549 (2017).
[2] Flaherty, G. T. and sung Choi, J.: The 'selfie' phe-

nomenon: reducing the risk of harm while using smartphones during international travel., *Journal of travel medicine*, Vol. 23 2, p. tav026 (online), available from (<https://api.semanticscholar.org/CorpusID:205193256>) (2016).
[3] Svelander, A. and Wiberg, M.: The Practice of Selfies, *Interactions*, Vol. 22, No. 4, p. 34–38 (online), DOI: 10.1145/2770886 (2015).
[4] Criscito, M. C., . S. J. A.: The selfie skin examination, *Journal of the American Academy of Dermatology*, Vol. 74, No. 6, pp. e123–e125 (online), DOI: <https://doi.org/10.1016/j.jaad.2015.12.023> (2016).
[5] Chu, S. and Tanaka, J.: Interacting with a Self-Portrait Camera Using Motion-Based Hand Gestures, *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction*, APCHI '13, New York, NY, USA, Association for Computing Machinery, p. 93–101 (online), DOI: 10.1145/2525194.2525206 (2013).
[6] Liou, J.-C., Lin, W.-C., Wen, W.-J. and Lin, W.-D.: Manipulated cell phone camera with using accelerometer devices, *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 165–166 (online), DOI: 10.1109/ICCE-TW.2016.7520969 (2016).
[7] Patil, S. G., Dennis, D. K., Pabbaraju, C., Shaheer, N., Simhadri, H. V., Seshadri, V., Varma, M. and Jain, P.: GesturePod: Enabling On-Device Gesture-Based Interaction for White Cane Users, *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, New York, NY, USA, Association for Computing Machinery, p. 403–415 (online), DOI: 10.1145/3332165.3347881 (2019).
[8] Hursale, S., Makwana, H., Panchal, A., Bhakkad, C. and Vairagar, S.: Smartphone's Based Gesture Recognition in Air, *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)*, ICARCSET '15, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/2743065.2743114 (2015).
[9] Wobbrock, J. O., Morris, M. R. and Wilson, A. D.: User-Defined Gestures for Surface Computing, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, New York, NY, USA, Association for Computing Machinery, p. 1083–1092 (online), DOI: 10.1145/1518701.1518866 (2009).