

角膜反射を利用したスマートフォンの把持姿勢推定

張 翔^{1,a)} 池松 香^{2,b)} 加藤 邦拓^{3,c)} 杉浦 裕太^{1,d)}

概要：スマートフォンをユーザがどのように把持しているのかを測定することで、把持姿勢に合わせた画面表示や操作ボタン位置の自動切り替えが可能となり、ユーザに合わせたインタラクションが可能になる。本研究では、スマートフォンを操作する際に、角膜にスマートフォンの画面の光が反射して写る像を内蔵インカメラで撮影することで把持姿勢を推定する手法を提案する。角膜に反射したスマートフォンの像ではユーザが画面に指を置いている部分が影となって欠けるため、把持姿勢の推定が可能となる。本稿では、提案手法の精度を検証する実験を行った。

1. はじめに

スマートフォンを持っている手や操作している指などの把持姿勢を測定することで、把持姿勢に合わせた画面表示、操作ボタン位置の自動切り替えや、アプリや機能を自動で立ち上げることが可能になる。特に画面表示や操作ボタン位置に関して、親指や人差し指で操作しやすい領域は Thumb Zone や Index Finger Zone と呼ばれ、研究が行われている [6][9] もの、スマートフォンアプリによっては右手親指による把持姿勢では操作は考慮されていても、その他の把持姿勢では操作が行いづらい場合がある。

把持姿勢を測定する方法として、スマートフォンに静電容量センサなどのセンサを外付けして測定を行う方法 [10][1] や、ジャイロセンサなどのスマートフォンに内蔵されたセンサを用いて測定を行う方法 [2][7] が提案されている。しかし、それらの方法ではスマートフォン単体では計測できなかったり、センサ位置の違いから端末によってパラメータが大きく変化したりするといった課題がある。

本研究ではスマートフォンの内蔵インカメラを用いてユーザの顔写真を撮影することで図 1 に示す 6 種類の把持姿勢を識別する手法を提案する。スマートフォンを操作する際、角膜にスマートフォンの画面の光が反射して像が写るが、ユーザが画面に指を置いている部分は影となって欠ける。そこで、撮影した顔写真からスマートフォンの像が写っている瞳の部分クロップし、畳み込みニューラル



図 1 把持姿勢の推定の流れ

ネットワーク (CNN) を使って作成した学習モデルに瞳の画像を入力することで識別を行う。14 人の実験協力者を対象にして本提案手法の精度を検証する実験を行った結果 84% の精度で識別をすることができた。

2. 関連研究

把持姿勢を測定する手法として、端末に測定用のセンサを外付けする手法が提案されている。Yoon ら [10] は静電容量センサをタブレット端末の裏面に貼り付けて把持姿勢を推定し、画面表示を最適化する手法を提案している。Cheng ら [1] は静電容量センサを用いて把持姿勢を推定し、バーチャルキーボードの表示を最適化させる手法について検討をしている。また、Yu ら [11] はスマートフォンのインカメラにプリズムミラーを装着することでステレオビジョンを実現し、推定された奥行き情報から把持姿勢や操作状態を測定する手法を提案している。

¹ 慶應義塾大学
² ヤフー株式会社
³ 東京工科大学
a) shocho0101@keio.jp
b) k-ikematsu@acm.org
c) kkunihir@acm.org
d) sugiura@keio.jp

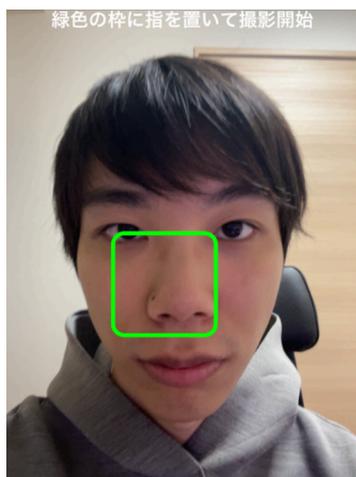


図 2 撮影画面

スマートフォンに内蔵されたセンサを用いて把持姿勢を測定する方法では加速度センサやジャイロセンサを用いる手法とスピーカとマイクを用いる手法が提案されている。Goelら[2]はスマートフォンをバイブレーションさせ、その際の振動を加速度センサを用いて測定する手法を提案している。また、Parkと小川[7]は加速度センサとジャイロセンサで測定されたデータを使用してサポートベクタマシン(SVM)で把持姿勢を測定する手法を提案している。スピーカとマイクを用いて識別を行う手法ではKim[5]らがスピーカから音を鳴らし、マイクで計測された伝搬音の違いを用いて把持姿勢を推定する手法を提案している。これらの方法ではセンサが取得されるデータがスマートフォンの筐体に依存したり、外部環境に影響を受けるといった課題が挙げられる。

また、光の反射によって写る像を利用した研究として、ユーザの正面に置かれたミラーレス一眼カメラを使って角膜に反射したユーザの手の像を撮影することでスマートフォンの周辺領域での操作を検出する手法[8]や、サンガラスに反射したスマートフォンの画面やユーザの手の像を内蔵インカメラを使用して撮影することでスマートフォンの入力を拡張する手法[4]が提案されている。

本研究ではスマートフォンの内蔵インカメラを使用するため、センサを外付けする必要がない他、スマートフォンの機種による影響が小さいというメリットがある。

3. 実装

本研究では、スマートフォンを持った際の顔写真を収集するアプリケーションを作成し、収集した顔写真からCNNで把持姿勢の学習および識別を行った。この際、図1に示す6種類の把持姿勢を対象とした。

3.1 顔写真の収集

把持姿勢をCNNで学習、識別するためのデータを収集す

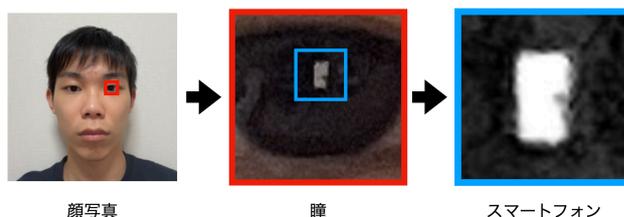


図 3 顔写真のクロップの流れ

るために、それぞれの把持姿勢でスマートフォンを持った際の顔写真を撮影する iPhone 端末向けのアプリケーションを実装した。このアプリケーションでは図1に示す6種類の把持姿勢でスマートフォンを持った際の顔写真を内蔵インカメラを用いて撮影する。各把持姿勢について12枚ずつ、合計72枚撮影の顔写真の撮影を行い、サーバに送信する。撮影時の画面には図2のように緑色の枠が表示され、ユーザはその枠に指をおいた状態で撮影を行う。枠は各把持姿勢の際に、親指または人差し指で無理なく触れられる位置に表示した。また、撮影時には画面が白くフラッシュし、1つの把持姿勢につき画面輝度が100%、85%、70%の状態それぞれ4枚ずつ撮影される。なお、顔写真の撮影の際には眼鏡やコンタクトレンズは外し、パソコンの画面やテレビなどが瞳に写り込まない環境で行うようにする指示を撮影前にアプリ内に表示するようにした。

3.2 把持姿勢の学習

収集された顔写真から把持姿勢を学習、識別するための前処理として、顔写真から瞳の中でスマートフォンが写っている部分をクロップ、画像の明るさやコントラストの調整、データ数を増やすためのデータオーグメンテーションの3つを実施した。

顔写真のクロップは図3に示すように顔写真から瞳の部分、瞳の部分からスマートフォンが写っている部分の2段階でクロップを行い、最終的に80ピクセル×80ピクセルの画像として出力する。顔写真から瞳の部分のクロップにはGoogleが提供するMediaPipe[3]を使用した。また、瞳の部分からスマートフォンが写っている部分のクロップではOpenCVを使用し、画像内の明るい部分をスマートフォンが写っている領域として輪郭抽出を行い、その周辺をクロップした。なお、この際、スマートフォンが写っている領域の面積を画像全体の面積の5-20%であると想定し、画像内で明るい領域の面積がその範囲内に収まらない場合についてはエラーとして、その画像についてはクロップを行わず、CNNでの学習、識別の際にもデータに含めないようにした。次にクロップされた画像の明るさをデータセット全体で揃えた後、スマートフォンの像で指の影となってかけている部分を強調するためにコントラストを向上させた。

このように前処理された画像を学習用データとテスト

表 1 実験結果

	識別精度
全ユーザ学習	84%
他ユーザ学習	43%

表 2 全ユーザ学習, 6 姿勢での個別の識別結果

	適合率	再現率	F1 値
両手横持ち	96%	98%	97%
両手縦持ち	92%	92%	92%
左手親指	82%	77%	80%
左手人差し指	74%	90%	81%
右手親指	89%	68%	77%
右手人差し指	83%	85%	84%

データに分割した後、画像の角度や大きさを少しずつ変更してデータ数を増やすデータオーグメンテーションを実施した。最後に両眼の画像のデータを重ねて $80 \times 80 \times 2$ の 1 つの 3 次元テンソルとして CNN のモデルに入力し、学習と識別を行う。

4. 実験

本研究で提案する手法の精度を検証するための実験を行った。

4.1 概要

実験では 14 人の実験協力者が参加した。参加者は、顔写真収集用のアプリケーションを使用して、それぞれの把持姿勢でスマートフォンを持った際の顔写真を撮影した。本実験では、14 人 \times 6 把持姿勢 \times 各 12 枚 = 1008 枚分のデータを収集した。収集した画像に対して上述の前処理を行い、エラー画像を除いたところ 71% のデータが残った。このデータを使用して、前章で述べた手法を用いて CNN で学習モデルを作成するとともに、その学習モデルの識別精度を検証した。なお、撮影される画像の画素数の条件を揃えるため、インカメラの画素数が 1200 万画素である iPhone11, iPhone12 系の端末を使用した。

CNN の学習に用いるデータと識別精度の検証に用いるテストデータの分割には、全ユーザのデータを学習用データ、テストデータに用いる全ユーザ学習と、Leave-one-subject-out 交差検証での他ユーザ学習の 2 種類の方法で行った。なお、どちらの場合でも学習用データとテストデータは 3:1 の割合となるようにした。

4.2 結果

テストデータの分割方法、及び把持姿勢数ごとの識別結果を表 1 に示す。他ユーザ学習に比べて全ユーザ学習の方が 40% 近く精度が高いことが分かる。また、全ユーザ学習、6 姿勢での学習、識別を行った際の結果を図 4 と表 2 に示す。特に右手親指と右手人差し指、左手親指と左手人差し指について識別が混同してしまっている場合があるこ

両手横持ち	98	0	0	0	0	2
両手縦持ち	2	91	0	4	2	0
左手親指	0	4	76	14	2	2
左手人差し指	0	2	7	89	0	2
右手親指	2	0	9	12	68	10
右手人差し指	0	2	0	9	4	85

図 4 6 姿勢での識別結果の混合行列

とが図 4 からわかる。また、両手横持ちに関しては適合率、再現率がともに最も高く他の把持姿勢と混同される場合が少なかった。

4.3 考察

今回の提案手法では角膜に反射したスマートフォンの画面の像のうち指に隠れて影となって欠けている部分から識別を行っている。そのため、右手で持っている条件と左手で持っている条件間では像の欠け方が大きく変わるのに対し、親指と人差し指のどちらで操作しているかの条件については像の欠け方に大きな影響を及ぼさないため、混同したと考えられる。

また、両手横持ちの識別が他の把持姿勢に比べて精度高く行われた原因としては、この把持姿勢のみスマートフォンを横にして持つため、瞳にも横長の画面が反射し、識別が行いやすかったためであると考えられる。

データの分割方法に着目した場合、他ユーザ学習は全ユーザ学習に比べて精度が低くなった。これについては、瞳の色や光の反射のしやすさなどの生物学的個人差による影響に加えて、スマートフォンを持つ位置や撮影する際の環境光の明るさなどによる影響も考えられる。

5. 今後の課題

本研究での提案手法を使って把持姿勢を推定できる条件を検証したり、精度をさらに向上させるために、より高画質な画像や低画質な画像を使った検証、グレースケール画像ではなく、2 値画像や RGB 画像を使った検証、スマートフォンのサイズごとの識別精度の検証などを行うことが考えられる。また、他ユーザ学習での識別精度が全ユーザ学習での識別精度に比べて低い結果となったが、実際にアプリケーションとして使用する際にユーザが手軽に精度良く把持姿勢を識別できるようにするために転移学習を用い

る方法が考えられる。

今回、角膜に確実にスマートフォンの画面の光を反射させるため、画面をフラッシュさせる条件での実験を行った。しかし、実際に把持姿勢に合わせて画面表示や操作ボタンの自動切り替えなどで使用する際、画面が定期的にフラッシュするとユーザに負担を与えてしまう。今後、角膜にスマートフォンの画面の光が反射するための画面輝度や表示される色などの条件を調査したり、一般的なブラウザやゲームの画面などを使用し、画面をフラッシュさせない場合の本手法の精度を検証する必要がある。

また、本研究では内蔵インカメラを使用したが、親指と人差し指のどちらの指で操作をしているのかを高精度で識別するためには、指の動かし方の違いに着目して画像ではなく、動画を撮影して識別を行う方法や、ジャイロセンサやアクティブ音響センシングなど関連研究で使用されたセンサや手法を併用する方法が考えられる。

今回の研究ではユーザは顔の正面にスマートフォンを持っている前提で行っているが、実際にスマートフォンを使用する際は顔の斜め下などの位置にある場合もある。そのような場合でも識別できるようにするために、瞳の領域をクロップする方法の改善や、角度によらず識別できる学習モデルを作成する手法の検討が今後の課題である。

6. 結論

本研究ではスマートフォン内蔵のインカメラを用いてユーザの顔写真を撮影し、6種類の把持姿勢を識別する手法を提案した。撮影された顔写真から瞳の部分をクロップし、CNNで学習をして識別を行い、精度を検証したところ全ユーザ学習では平均85%で把持姿勢を識別することができた。

参考文献

- [1] Cheng, L.-P., Liang, H.-S., Wu, C.-Y. and Chen, M. Y.: iGrasp: grasp-based adaptive keyboard for mobile devices, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, ACM, p. 3037–3046 (2013).
- [2] Goel, M., Wobbrock, J. and Patel, S.: GripSense: using built-in sensors to detect hand posture and pressure on commodity mobile phones, *Proceedings of the 25th annual ACM Symposium on User Interface Software and Technology - UIST '12*, ACM, p. 545 (2012).
- [3] Google: MediaPipe, available from (<https://google.github.io/mediapipe/>) (accessed 2021-07-26).
- [4] Grubert, J., Ofek, E., Pahud, M., Kranz, M. and Schmalstieg, D.: GlassHands: Interaction Around Unmodified Mobile Devices Using Sunglasses, *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces - ISS '16*, ACM, p. 215–224 (2016).
- [5] Kim, N., Lee, J., Whang, J. J. and Lee, J.: Smart-Grip: grip sensing system for commodity mobile devices through sound signals, *Personal and Ubiquitous Computing*, Vol. 24, No. 5, p. 643–654 (2020).
- [6] Le, H. V., Mayer, S., Bader, P. and Henze, N.: Fingers' Range and Comfortable Area for One-Handed Smartphone Interaction Beyond the Touchscreen, New York, NY, USA, ACM (2018).
- [7] Park, C. and Ogawa, T.: A Study on Grasp Recognition Independent of Users' Situations Using Built-in Sensors of Smartphones, *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology - UIST '15 Adjunct*, ACM, pp. 69–70 (2015).
- [8] Schneider, D. and Grubert, J.: Towards Around-Device Interaction using Corneal Imaging, *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces - ISS '17*, ACM, p. 287–293 (2017).
- [9] Yoo, H., Yoon, J. and Ji, H.: Index Finger Zone: Study on Touchable Area Expandability Using Thumb and Index Finger, *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, ACM, p. 803–810 (2015).
- [10] Yoon, D., Hinckley, K., Benko, H., Guimbretière, F., Irani, P., Pahud, M. and Gavrilu, M.: Sensing Tablet Grasp + Micro-mobility for Active Reading, *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology - UIST '15*, ACM, pp. 477–487 (2015).
- [11] Yu, C., Wei, X., Vachher, S., Qin, Y., Liang, C., Weng, Y., Gu, Y. and Shi, Y.: HandSee: Enabling Full Hand Interaction on Smartphone with Front Camera-Based Stereo Vision, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, New York, NY, USA, ACM, p. 1–13 (2019).