

顔認証システムにおける多要素認証の可能性の検討

小 山 洋 太^{*1}

坪 井 瑞 輝^{*1}

中 川 高 晃^{*2}

【要 旨】

本研究では近年様々な場面で活用されている「顔認証」を補完する手法を念頭に、「カメラ像からのデータ入力の可能性」について、「目線」、「手の形」そして「姿勢」に注目して検証を行った。検証の結果、依然として改良・更なる検討の余地はあるものの、各々、手法としては十分に可能性があることを確認した。

1 はじめに

新型コロナウイルス感染症の影響により、接触の機会をできる限り減らしたニューノーマル社会に対応したインフラや機器の需要が高まっている。このような状況では現在、空中ディスプレイを用いたタッチパネル様の制御システム¹⁾、及び顔などの画像を用いた認証（「顔認証」）などが注目されている。この顔認証については、現在、スマートフォンのロック解除から顔認証決済システムの構築まで様々な場面で利活用が進んでいる。

しかしながら、「顔」という情報は、非常に他者からアクセスしやすい、開かれた情報と考えられ、認証のより高い真性を担保するためには、何らかのパスフレーズなど、本人しか知り得ない情報を援用する手法が必要となってくると考えられる。また、これを敷衍し、画像データから被撮像対象の意図（能動的なデータ入力）を捉える手法はデジタル技術を活用した社会において必須の技術になると推察される。

そこで本研究では、画像データを介したデータの入力の可能性を検討するべく、カメラ像から被写体の意図を読み取るシステムの構築可能性につ

いて検討を行った。

2 被写体の解析

2. 1 ランドマークの取得

2. 1. 1 ライブラリの選定

今回、被写体の意図を読み取る手法として、体の各部位の状況を判断するいわゆるジェスチャーによる入力を検討した。その際、体の各パーツを追いかけるため画像から被写体のランドマーク（特徴点）を取得するライブラリを援用した。

このようなランドマーク取得のライブラリとしては、Alphabet 社 (Google) の「MediaPipe」²⁾や Intel 社の「OpenVINO」³⁾など、機械学習ベースのモジュールが現在、注目を集めている。

しかしながら、OpenVINO は公式的には Intel 社のアーキテクチャのみの対応とのことであることなどから、本研究では、AMD などの他のアーキテクチャでも利用可能な「MediaPipe」を援用することとした。

2. 1. 2 MediaPipe でのランドマーク取得

MediaPipe においては、現在、物体の検出や追跡などの様々な機械学習ベースのソリューションが提供されている。中でも人にかかるソリューションが豊かであり、顔の検出から顔のランドマーク

* 1 応用技術課 副主査

* 2 応用技術課 技師

取得や瞳に特化した検出、また、手・体（姿勢）のランドマーク取得、それを一体化した検出や、さらには頭髮部の検出なども準備されている。これらは、事前にマーカーなどを検出部位に設置する必要なく、取得された画像から解析し、ランドマークを割り出すことが可能となっている。なお、ソリューションの提供については、OS やプログラミング言語ごとに異なる状況となっているが、今回はこれらソリューションのうち、「顔（および瞳）のランドマーク取得（FaceMesh）」、「手のランドマーク取得（Hands）」、「体のランドマーク取得（Pose）」について python3 により、意図取得の実装を試みた。

2. 2 目線の取得 (GazeDirection)

2. 2. 1 手法

まず、MediaPipe の「FaceMesh」を用いて、瞳の位置（目線）による上下左右入力について検証を行った。「FaceMesh」においては、468 点に及ぶ顔面上のランドマーク取得が可能であるが、加えて REFINELANDMARKS オプションを使用することにより、顔面の目や口などの各部位についてより詳細にマッピングを行うことが可能で、その場合、瞳の位置や大きさを取得することが可能となるため、ここでは REFINELANDMARKS オプションによって取得したランドマークを利用することとした（図1）。

以上によって得られたランドマークより、瞳のずれ方向を認識し、目線を推定することとしたが、まずこの際、左右を認識することとした（理由については後述）。左右の認識方法は以下で行った。

- ① まず目の上下部のランドマーク及び左右のランドマークから目の中心位置を特定する。
- ② 次に、瞳が中心からずれたことを演算するが、この際、およそヒトの眼球は 23~24mm 程度⁴⁾であり、瞳の大きさは 11~12mm 程度⁵⁾である。

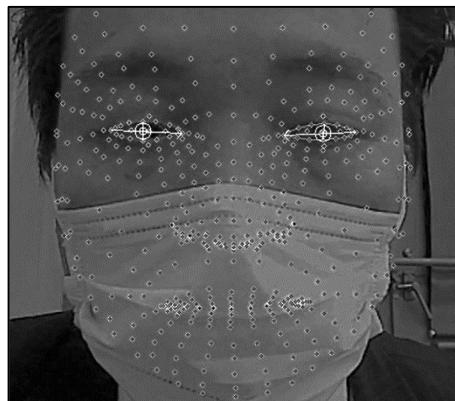


図1 ランドマーク取得例

そこで、眼球上を瞳が 15 度程度移動したとすると、およそ 3mm ずれることとなり、ちょうど瞳の半分となるため、「瞳がその直径の半分程度ずれた場合、目線が動いた」とする。（図2）

- ③ このずれにおいて、瞳がずれた先が左右の範囲内であれば、左右とする。（図3）

以上が左右の判定となるが、上下についてはこの判定では難しいところがあった。

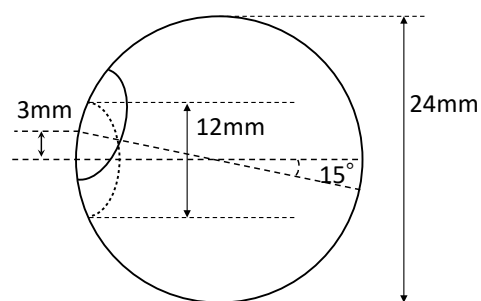


図2 瞳のずれ

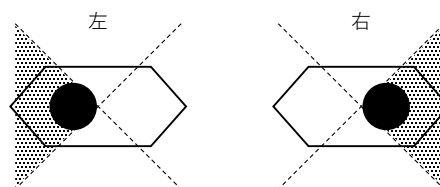


図3 左右の判定

まず、上については、図4に概略図を示すとおり、認識した目の範囲が上に引きずられる（下まぶたをしっかりと認識できない）傾向があり、結

果、想定よりも目が中心からずれていない結果となってしまう。これは MediaPipe の性能の影響によるものと考えられ、現状では対応することが難しく、今回は、上方向についてはずれの判定条件を緩和して対応した。また、下については、そもそも伏し目となる場合、目が開いた像が得られない(図4)。このため、今回は目の範囲の上下左右のアスペクト変化から下を認識することとした。なお、このことから、目をつむった状況でも下と判定される状況となっているが、以下述べる通り、判定は1秒間の撮像結果より実施するため、瞬きに関しても問題と考えられる。

以上の通り、この度上下の判定は現状では厳しく、前述の通り、確定的な左右の判定を先行することにて実証を実施した。

また、判定においては1秒間の撮像において最も判定が多かった方向を目線方向とすることとした。

2. 2. 2 結果

図5にこの度の結果の一例を示す。このように目線については一定取得可能であったものの、上下については改善の必要性が認められるところであった。なお図5に示す通り、マスクを着用した状況であっても目を認識、判定することは可能であった。

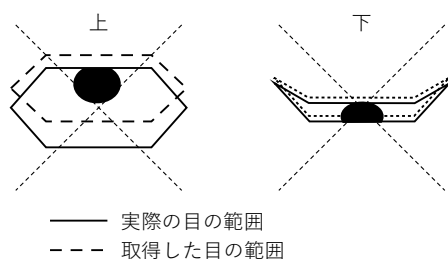


図4 上下の状況(概略図)

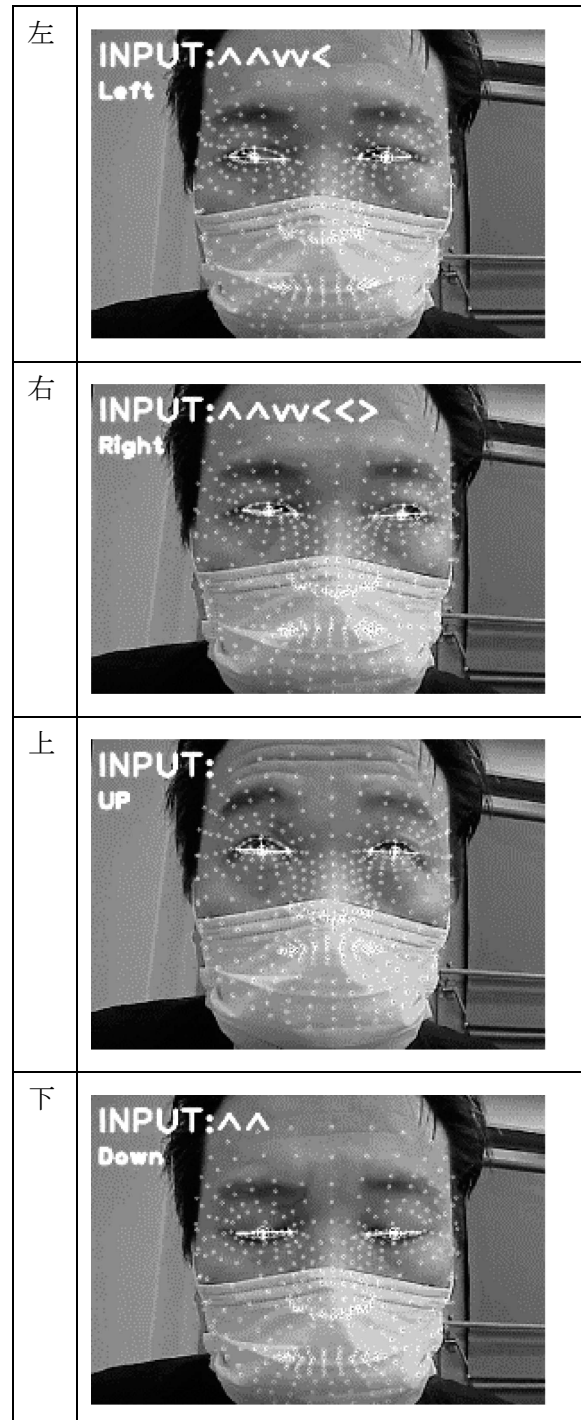


図5 目線の判定

2. 3 手の形の取得 (HandShape)

2. 3. 1 手法

次に、MediaPipe の「Hands」を用いて、手の形状(指折り)による数値入力について検証を行った。「Hands」においては、図6に示すように全21点の手のランドマーク取得が可能となつて

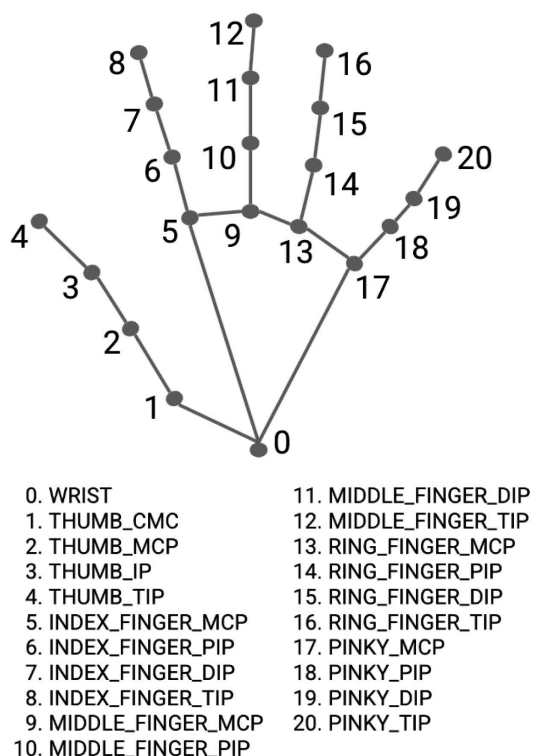


図6 ランドマーク一覧

いる。また、この度は片手だけ（認識する手は1つ）にて行ったが、複数の手を同時に認識することも可能であり、両手での入力も可能である。

以上によって得られたランドマークより、以下により指折りの状況を確認することとした。

- ① 親指については、親指の CM 関節（ランドマーク 1）と MP 関節（ランドマーク 2）を通る直線を基準として、伸ばすと指先（ランドマーク 4）が手首（ランドマーク 0）や小指の MC 関節（ランドマーク 17）と反対側となり、折ると同じ側となる。これにより指の伸ばし折りを判定する。
- ② そのほかの指に関しては、上記の同様の判定を、各指先に関して、PI 関節（たとえば 6 と 14）を通る線に対して、手首側と同じになるか異なるかによって折り曲げを判定する。

なお、判定に際しては、目線と同様、1 秒間の

撮像で最も多かった判定を採用する手法を用いた。

2. 3. 2 結果

図 7 にこの度の結果の一例を示す。指折りに関しては判定に恣意性も少なく非常に簡便で認識性高い結果となった。また、結果では 2 回の入力を合算することで、0～9 の入力ができるようにしているが、先述のとおり、両手の認識を組み合わせることで、一度に 0～9 の入力を実施することも可能と考える。

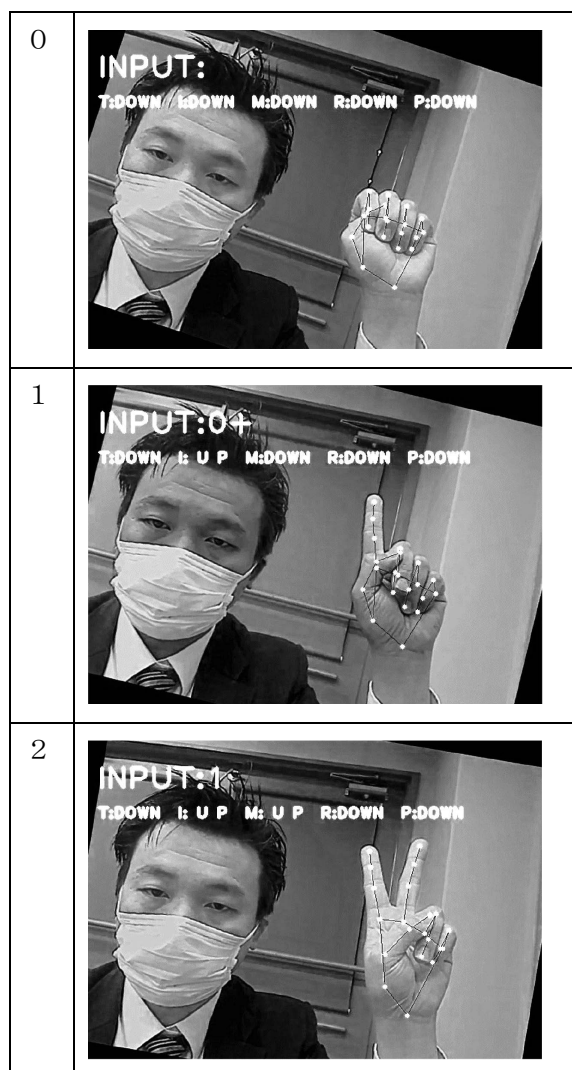




図7 指折りの判定

2. 4 手旗信号 (Semaphore)

2. 4. 1 手法

次に、MediaPipe の「Pose」を用いて、特定の姿勢による文字入力について検証を行った。ここでは特定の姿勢として、和文手旗信号を例とした。

「Pose」においては、図8に示すように一部顔のランドマークを含めた全33点のランドマーク取得が可能となっている。一方、手旗信号の原画（手旗の1つ1つの符号）においては、およそ図9に示す8象限（左右の腕、各々肩を原点とする）にエリアを分け、どの象限に手があるかで、一部を除く原画について認識が可能である⁶⁾。ここではこの考えで手旗信号認識を行った。なお、一部

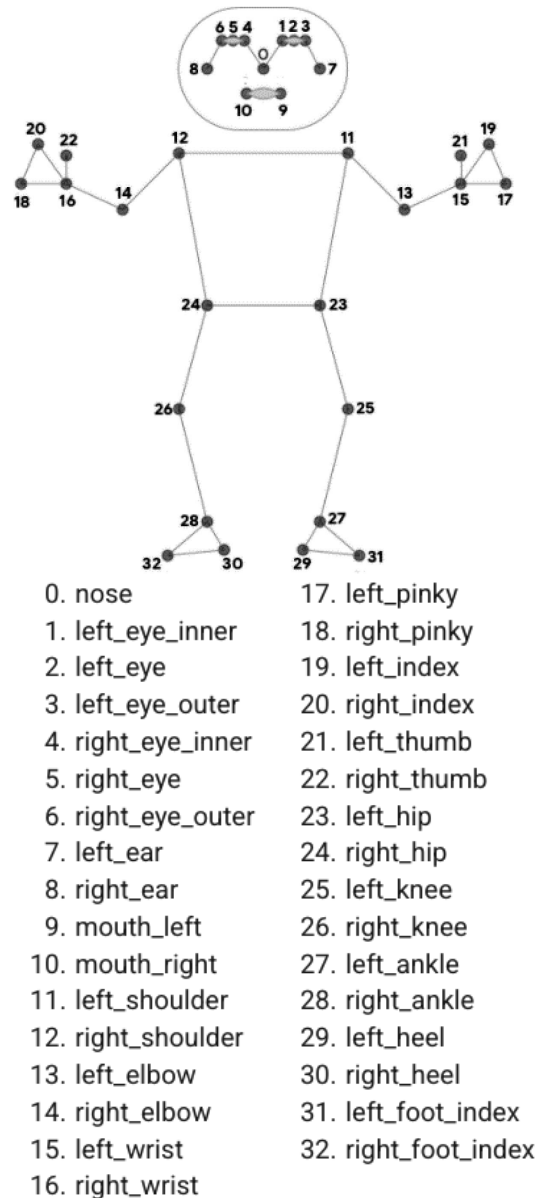


図8 POSE のランドマーク一覧

の原画（第0原画および第11原画）においては、「動き」を認識する必要が出てくる。これらについて対応するため、2秒間の撮影中に手が動いている（ある象限に手が停留していない状態。つまり、撮像解析において認識された象限のうち主となる象限が見当たりにくい。）については、これらの原画を行うこととし、第0原画 については、「すべての象限にて認識された実績がある」こと、第11 厳格については、「象限ではなく、肩と腰で構成される四角形内に手が入ったことがあ

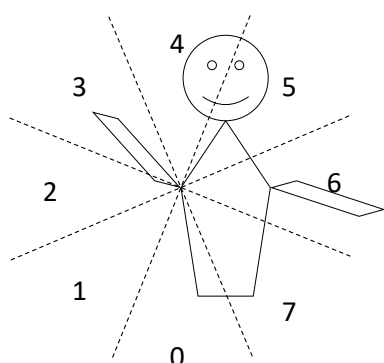


図9 腕の方向の8象限

る」の特例にて対応し、取得することとした。

2. 4. 2 結果

図10に結果の一部を示す。基本的には認識ができたところであるが、像が見切れる（手の先が映らない）などの場合、腕が曲がっているように認識される場合があり、しっかりと全体を取られることが重要であった。また、8象限で認識していることを意識し、しっかりと腕を伸ばすことで、正確な認識を実現することが可能であった。

3 まとめ

本研究では近年様々な場面で活用されている「顔認証」を補完する手法のための画像データからの意図入力に発想を得て、被写体であるひとの挙動から意図を読み解く手法の可能性について調査を行った。その際、挙動の解析には「MediaPipe」を援用した。結果、「目線」、「手の形」そして「姿勢」に注目して検証を行ったところ、それぞれについて改善の余地はあるものの、十分な利活用可能性が見いだされた。「目線」に関しては左右の認識については問題なく、上下が厳しいところであった。また、「手の形」は、この度、数字入力を対象としたが、十分に利用可能と思われるところである上、さらなる複雑な形状の可能性も認めら

原 姿		
チ		
ユ		
ウ		
セ		
ン		

図10 「チュウセン」の入力

れるところである。さらに「姿勢」については、全像が映ることが重要であるものの、認識は可能であり、動きがある動作については更なる考察が必要であるが、静止した姿勢であれば、比較的、簡易に取得し解析することが可能であると思われる。更なる改善方法としては、目や手、姿勢など

について解析の前段で原点位置を指定する撮像を実施することが考えられる。また、この度は閾値を設け、それに沿って判定したところであるが、機械学習的な分類を援用することで、認識率の向上が図られるかとは思われるところであり、これは今後の課題である。

(参考文献)

- 1) 谷口 友一, 他: 京都府中小企業技術センター技報, No47, p31 (2019)
- 2) <https://google.github.io/mediapipe/>
- 3) <https://www.intel.co.jp/content/www/jp/ja/internet-of-things/opencvino-toolkit.html>
- 4) Bekerman I, et al. Variations in eyeball diameters of the healthy adults. J Ophthalmol. 2014;2014. doi:http://dx.doi.org/10.1155/2014/503645
- 5) <https://google.github.io/mediapipe/solutions/iris.html>
- 6) <https://www.mod.go.jp/msdf/special/flag/>