

深層学習による手の平と甲の両方を用いた個人認証

日大生産工(院) ○江波戸 海 日大生産工 柄窪 孝也

1. まえがき

個人認証はモノや場所、情報へアクセスする過程において、その権限がアクセス者にあるかどうかを識別する際に用いられている。近年では人が生まれながらに持つ特徴を用いた生体認証が注目されており、署名や音声といった行動的特徴を用いた認証と、指紋、静脈や顔の特徴といった身体的特徴を用いた認証に大きくわけられる。しかしながら、生体認証は記憶を用いたパスワード認証や所有物を用いたICカードの認証と比較して100%本人だけを認証することが保証できないため、誤って認証すべき本人を拒否してしまう本人拒否率 (False Reject Rates) と誤って認証すべき人ではない他人を受入れてしまう他人受入率 (False Accept Rates) を限りなく低くしなければいけないという課題がある。従来の手の平を用いた認証手法として掌紋認証があり、非接触かつ非拘束な状態で撮影された手の平画像を用いて認証ができることから利便性や受容性に優れる。一方、撮影の度に周囲の照明や手の平の方向・位置・姿勢などの撮影環境が変化するため、認証精度が安定した掌紋認証を実現するためには、撮影された手の平画像から掌紋領域を幾何学的に正しく推定する前処理が必要となる。一方、[1]では深層学習を用いて手の平画像を幾何学的な特徴量を用いずに手の平と甲が誰の手であるかを識別できるかを評価している。手の平と甲の画像を機械学習させて、手の平と甲の評価用画像を片方ずつ与えて、一定の閾値を超えた手を本人として判定し、FRRとFARを評価している。しかしながら、評価結果にてグレイスケール画像ではFRR2%と最も良い結果が得られた一方でFARが14%となり、エッジ加工を施した画像ではFARを2%まで抑えることができたがFRRが19%となり、FRRとFARの両方を低く抑えることが課題となった。本稿では、手の平と甲の評価用画像を片方ずつ判定した[1]の課題点を踏まえて、FRRとFAR両方を低く抑えるために、手の平と甲を様々なパターンで機械学習させて、手の平と甲の評価用画像両方をペアとして与えて、両方の手の平と甲の画像が一定の条件を満たしたときに本人として認証することで、精度の改善を図る。

2. 実験手法

2.1 機械学習時のパラメタ

機械学習は学習をする際に与える入力画像データすべてをランダムにシャッフルし、バッチサイズと呼ばれる画像の入力データ数で全データ数を割り、その商の回数分を学習し、それが1 epoch分に相当する。ここで、epoch数20、データ数1000、バッチサイズ100と仮定すると、1 epochの学習が終わるまでに1000/100=10回の機械学習が入力データ数100で行われ、それが20 epoch分行われることを表す。

2.2 画像加工の種類

RGB画像は撮影した手の平と甲の画像をそのまま用いる。グレイスケール画像は、R, G, Bの情報に対してそれぞれ重み0.299, 0.587, 0.114を掛け合わせた情報を一つのピクセルに格納したものをを用いる。RGB画像、グレイスケール画像を機械学習に与えたときに、手の画像の背景や影、光の反射光などが機械学習時に影響することを避けるため、エッジ画像を用意する。このエッジ画像は画像のそれぞれの画素に対して、注目画素及びその画素を中心とした周辺の画素値に対して重みであるカーネルと呼ばれる行列を掛け合わせ、その総和を出力画像の画素値として決定する。このカーネルのサイズは3×3や5×5が主に用いられるため、本稿ではこの2つのカーネルを用いてエッジ画像(カーネル3)とエッジ画像(カーネル5)として機械学習および評価を行う。カーネルのサイズが大きいと大局的な変化を捉えることができ、サイズが小さいと局所的に変化を捉えることが可能となる[2]。

2.3 機械学習用画像の準備

機械学習用画像には、箱型の撮影ボックスと三脚、照明を用いてスマートフォンで撮影したものを用いる。撮影時には機械学習時のデータ数確保のため、動画にて右手の平と甲をそれぞれ約10秒撮影し、その動画から各フレームを取り出した。撮影する際は、手の指を広げたり閉じたりする動作を繰り返しながら撮影することで様々な姿勢の手の平と甲の画像を取得した。ここで取得した画像は1920×1080ピクセルである。また撮影時に使用したスマート

フォンのフレームレートが30、撮影時間が約10秒、手の平と甲を撮影したため一人当たりの手の画像枚数は約 $30 \times 10 \times 2 = 600$ 枚となり、評価用画像のために手の平と甲の画像を10枚ずつ合計20枚抜き取るので機械学習には約580枚を使用する。この機械学習用画像約580枚を20人分、計約12000枚用意した。評価用画像を機械学習結果に与えると、その評価用画像が誰の手であるか、手の平と甲のどちらかのラベルの中で、確率が一番高い画像が判定され、その確率とラベルが返される。機械学習用画像と評価用画像には手の平と甲の画像をRGB、グレイスケール、エッジ(カーネル3)、エッジ(カーネル5)に画像処理を行い、それぞれ機械学習と評価を行った。また、一人当たりの機械学習時の訓練データを9割、正解データを1割として、epoch数を20とした。それぞれの画像処理された手の平の画像処理結果をFig.1に示す。

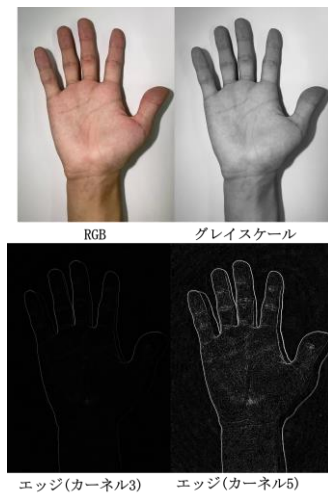


Fig.1 各画像処理を行った手の平画像

3 手の平または甲の片方を用いた方式

[1]では約12000枚の機械学習用画像を一緒に機械学習した結果に評価用画像の手の平と甲を片方ずつ与えて、事前に設けた閾値以上であれば本人であると判定し、閾値未満であれば他人であると判定して、本稿の実験と比較して評価を行う。Fig.2は一枚ずつ評価用画像を機械学習結果に与えた時の流れを示す。

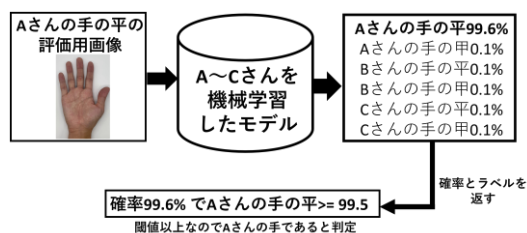


Fig.2 手の平の評価用画像を判定する例

Fig.2の例では、A, B, Cさんの手の平と甲を一緒に機械学習した結果にAさんの評価用画像の手の平を与えた結果として、一番高い確率99.6%とAさんの手の平というラベルを機械が返す。Fig.2の例は閾値99.5と設けているので、返された確率が閾値以上なので本人であると判定する。[1]ではFig.2の判定方法と実際に与えた評価用画像の人物が一致しているかをもとにFRRおよびFARの評価をしている。まずFRRの評価をするために、機械学習用画像の準備にて抜き取っておいた一人当たりの手の平と甲の画像10枚ずつの20枚×20人分=400枚の評価用画像を用いる。この手の平と甲の評価用画像400枚を片方ずつ機械学習結果に与えたときのFRRを次のように定義した。

$$FRR = \frac{\text{閾値未満で本人と判定しなかった枚数}}{\text{学習済みの人物の評価用画像400枚}} \quad (1)$$

次に、機械学習していない人によるFARを評価するために、撮影ボックスを用いた撮影方法①による4人分の手の平と甲の評価用画像4人×20枚=80枚と、照明等の環境が異なる撮影方法②による6人分の手の平と甲の評価用画像6人×20枚=120枚、撮影方法①と②の合計200枚の評価用画像を用意した。また、撮影方法②の撮影機器、フレームレートや画像サイズは撮影方法①および機械学習用画像を撮影するものと同じであり、撮影時も同様に姿勢の異なる手の平と甲の画像を取得した。[1]では手の平と甲の評価用画像200枚を、機械学習結果に与えたときのFARを次のように定義した。

$$FAR = \frac{\text{閾値以上で本人以外の人と判定した枚数}}{\text{未学習の人の評価用画像200枚}} \quad (2)$$

また、FRR評価時に学習済みの評価用画像を機械学習結果に与えて、本人であると判定した場合でも必ずしもその判定が正しいとは限らない。例えばAさんとBさんの手の平を機械学習した結果にAさんの評価用画像を与えたとき、Bさんであると判定してしまった場合は他人受入である。このように判定された評価用画像の枚数をカウントし、他人受入率(FAR')とする。[1]では機械学習済みの手の平と甲の評価用画像400枚を片方ずつ機械学習結果に与えたときのFAR'を次のように定義した。

$$FAR' = \frac{\text{閾値以上で本人以外の人と判定した枚数}}{\text{学習済みの人物の評価用画像400枚}} \quad (3)$$

機械学習をする際は以下に示す画像処理およびパラメタの組み合わせを用いる。

- RGB 画像、グレイスケール画像、エッジ画像(カーネル 3)、エッジ画像(カーネル 5)の画像処理を行った 4 種類の画像項目
- バッチサイズを32, 64, 128, 256と増加させた4種類のパターン
- 機械学習に与える画像の画像サイズを元の 1920×1080ピクセルからリサイズ処理した 32×32, 64×64, 128×128, 256×256ピクセルと増加させた4種類のパターン

また、手の平と甲の評価用画像を機械学習結果に与えた際に返される確率に対する閾値は適切なものを設けている。[1]ではこれら画像処理項目4通り×画像サイズ4通り×バッチサイズ4通りの64通りのパターンでFRRとFARの評価を行った。Fig.3はその評価結果を示す。ここでFig.3は64通りの中で最も良かった組み合わせを、4通りの画像サイズを横軸に固定した各画像処理項目のFRRとFARを示す。ここで、バッチサイズは画像サイズ32では128、画像サイズ64では256、画像サイズ128では256、画像サイズ256では256を用いており、縦軸は100%を最大値としている。

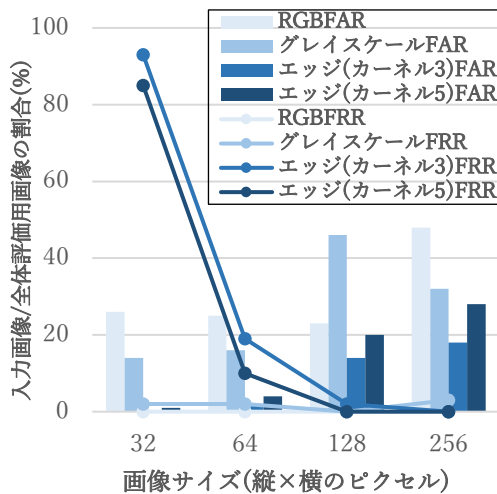


Fig.3 手の平または甲の FRR と FAR
また、Fig.3におけるFARは0%であった。

4 手の平と甲の両方を用いた方式

本稿では手の平または甲の片方を用いた方式の評価と比較して手の平と甲の評価用画像をペアで用いた場合、どれほど認証精度が向上するかを評価する。[1]では約 12000 枚の機械学習用画像を用いていたのに対して機械学習用画像が少ない場合の評価も行うために一人当たり約 580 枚の手の平と甲の画像 2 種類をそれぞれ 100 枚に制限した 20 人分の画像、100 枚×2種類×20人分=計 4000 枚の機械学習用画像を用意した。本稿では[1]で用いた約

12000 枚と 4000 枚の機械学習用画像を用いて Fig.4 に示す、機械学習用画像枚数と機械学習方法の組み合わせで実験 1~4 を行う。

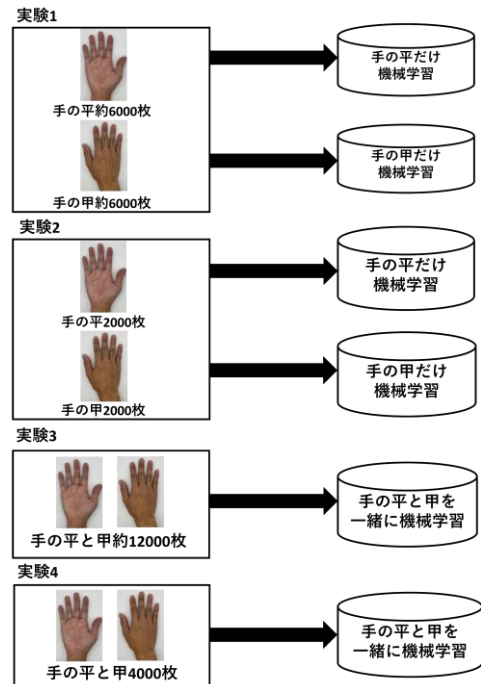


Fig.4 各実験の機械学習方法

Fig.4の実験1~4では、機械学習用画像が約12000枚または枚数を制限した4000枚を、機械学習するときには手の平と甲を別々または一緒に機械学習するかの組み合わせで異なっている。実験1~4の機械学習結果に評価用画像の手の平と甲を一ペアとして手の平、甲の順番に与えて一番高い確率として判定されて返された値に対して、手の平と甲に別々の閾値を設けて、どちらも閾値以上かつ返される手の平と甲が誰の手であるかのラベルが一致して、手の平と甲の順番が正しいとき本人と判定する認証条件を定義し、その条件を満たしたときに返されたラベルの人物を本人と判定して、満たさなければ本人以外の人物と判定する。Fig.5は評価用画像を機械学習結果に与えた時の流れを示す。

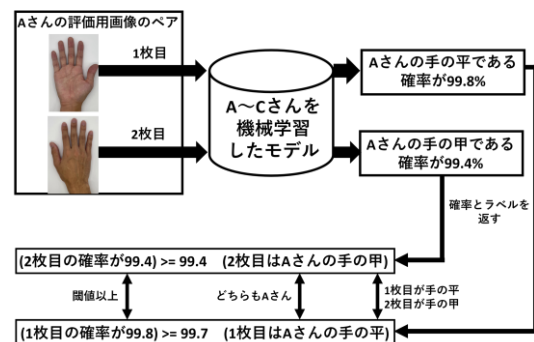


Fig.5 ペアの評価用画像の判定の例

Fig.5の例ではA, B, Cさんの手の平と甲を一緒に機械学習した結果にAさんの評価用画像の手の平と甲をペアとして順番に与えた結果として、Aさんの手の平である確率が99.8%、Aさんの手の甲である確率が99.4%と一番高い確率で判定され、そのラベルであるAさんの手の平と甲をそれぞれ機械が返す。Fig.5の例は返された確率に対する閾値を手の平99.7%、手の甲99.4%と設けており、1枚目が手の平、2枚目が手の甲と一致しており、両方Aさんと判定していることで認証条件を満たしていることから本人であると判定する。本稿ではFig.5の認証条件による判定方法から、実際に与えた評価用画像の人物が一致しているかをもとにFRRおよびFARの評価を行う。また、FRR, FARおよびFAR'に用いる評価用画像は[1]と同じである。まずFRRの評価は、手の平と甲の評価用画像400枚を手の平と甲のペア200組として機械学習結果に与えたときにFig.5の例のように閾値以上かつ手の平と甲の判定された人物のラベルが一致しているかつ手の平と甲の順番で識別しているときを認証条件として次のように定義する。

$$FRR = \frac{\text{手の平と甲のどちらか一方が閾値未満で本人と判定しなかったペア数}}{\text{学習済みの人物の評価用画像ペア200組}} \quad (4)$$

次に機械学習していない人によるFARの評価は、手の平と甲の評価用画像200枚を手の平と甲のペア100組として機械学習結果に与えたときにFig.5の閾値以上かつ手の平と甲の判定された人物のラベルが一致しているかつ手の平と甲の順番で識別しているときを認証条件として次のように定義する。

$$FAR = \frac{\text{手の平と甲の両方が閾値以上で本人以外の人と判定したペア数}}{\text{未学習の人物の評価用画像ペア100組}} \quad (5)$$

また、FRRの評価時に学習済みの人の評価用画像400枚を手の平と甲のペア200組として機械学習結果に与えたとき、評価用画像の人物とは異なる人が判定されたときのFAR'を次のように定義する。

$$FAR' = \frac{\text{手の平と甲の両方が閾値以上で本人以外の人と判定したペア数}}{\text{学習済みの人物の評価用画像ペア200組}} \quad (6)$$

機械学習をする際は、[1]と同じ画像処理項目4通り、画像サイズ4通り、バッチサイズ4通りの64通りを用いる。また、手の平と甲の評価用画像を機械学習結果に与えた際に返される確率に対する閾値は適切なものを設ける。Fig.6に実験1~4を横軸とした64通りのパターンの中

で最も良かった組み合わせのFRRとFARを示す。また、縦軸は20%を最大値としている。

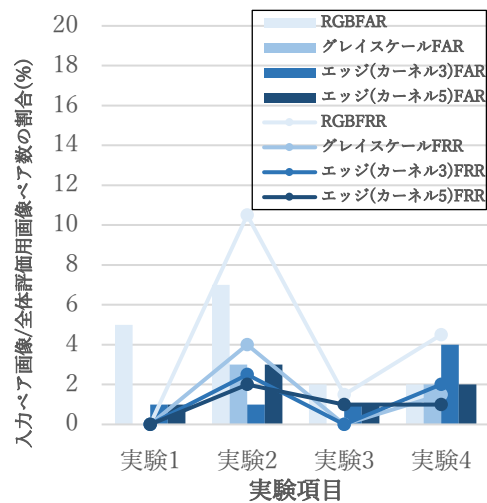


Fig.6 実験 1~4 の FRR と FAR

表1にはFig.6の各画像処理項目に対するバッチサイズと画像サイズの組み合わせを示す。

表 1. Fig.6 のパラメタの組み合わせ

	バッチサイズ	画像サイズ (ピクセル)
実験 1	64	128×128
実験 2	64	128×128
実験 3	128	64×64
実験 4	64	128×128

また、Fig.6におけるFAR'は0%であった。

5 おわりに

[1]のFig.3と比較して本稿の評価結果のFig.6ではパラメタの組み合わせによっては機械学習の画像枚数が多い実験1と実験3でFRR0%、FAR0%という結果が得られた。このことから判定の条件を厳しくした手の平と甲両方のペアを用いた手法は手の表裏の片方だけを用いた手法よりFRRとFARどちらも改善することができた。今後の課題として本稿では20人分の手の平と甲の機械学習用画像を用いたが、より多くの人数の手の平と甲の画像を用いることで評価結果の信頼性向上を図りたい。

参考文献

- [1] 江波戸海, 柘窪孝也, 目黒光彦, “深層学習を使った特徴量を用いない手のひら画像による個人認証,” 電子情報通信学会バイオメトリクス研究会 (BioX), 信学技報 (IEICE), vol 122, no 394, 2023.
- [2] 永田雅人, 豊沢聡, 実践 OpenCV 4 for Python 画像映像情報処理と機械学習, 株式会社カットシステム, 2021.