

CNN を用いた手の形状による個人認証

日大生産工(院) ○江波戸 海 日大生産工 析窪 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。その中でも生体認証には主に署名や音声といった行動的特徴を用いた認証と、指紋や静脈、顔の特徴といった身体的特徴を用いた認証方式に大きく分けられる。しかしながら、生体認証システムにはパスワードなどの記憶を用いた認証やICカードなどの所持物を用いた認証と比較して100%本人だけを認証することが保証できないため、誤って認証すべき人ではない他人を受け入れてしまう他人受入率(False Accept Rates)と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)を限りなく低くしなければいけないという課題がある。さらに、他人受入率を低く抑えようとすれば本人拒否率が上がってしまい、逆に本人拒否率を低く抑えようとすれば他人受入率が上がってしまうというトレードオフの関係性を持つ。従来の身体的特徴を用いた生体認証では、指紋や手相、虹彩などから幾何学的な特徴点を抽出することで登録した時の特徴と認証する時の特徴を照らし合わせて認証を行う。本稿で用いる生体認証では、幾何学的に特徴を決定せずに人の手の形状を撮影した画像を用いて、画像認識に優れているとされる畳み込みニューラルネットワーク(CNN)で処理を行い、その手が誰の手であるかを識別させることで生体認証システムの実現を図る。また、その過程において本人拒否率と他人受入率を低く抑えるために、CNNで学習する画像をそれぞれRGB画像、グレースケール(白黒)画像、エッジを抽出した画像を使い、CNNの入力データ数を示すバッチサイズや、元の画像から解像度を縮小させた入力データといったパラメータを変化させることで現実的な生体認証に近づけるべくその評価を行う。

2. CNN

2.1 局所特徴の抽出

従来の生体認証システムに用いられる幾何学的特徴の具体例として指紋認証では、隆線の分岐や端点の位置、種類、方向を計算する方式

がある。指紋認証にて用いられる特徴をFig.1に表す[1]。

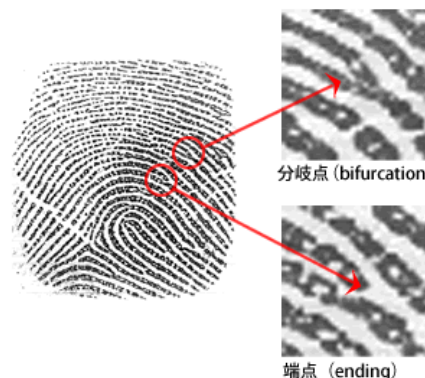


Fig.1 指紋認証に用いられる特徴

本稿での生体認証システムでは従来の認証で用いられる幾何学的な特徴点は抽出せずただの画像の一部として機械が認識する点が大きく異なる。CNNでは学習時にFig.1の分岐点や端点といった特徴を、画像の隣り合う線や点の集まりとして1つの局所特徴とすることで、機械学習時にその特徴を検出する。CNNには畳み込み層があり、カーネルと呼ばれるあらかじめ決められた値の並びを少しずつ入力された画像に対してずらしながら掛け合わせることで入力画像の局所特徴を出力する。機械学習ではこのカーネルを重みとして、画像におけるパターン検出を行っていく過程で調整を行う。入力画像のデータを x 、重みを W とすると出力される局所特徴を $a = W \times x$ として表せる。この a を引数として活性化関数を計算することで次のユニットに送る出力値が求まる。本稿ではこの活性化関数にReLU関数を用いており、次のような性質を持つ。

$$R(a) = \begin{cases} a & (a \geq 0) \\ 0 & (a < 0) \end{cases} \quad (1)$$

$$\frac{dR(a)}{da} = \begin{cases} 1 & (a > 0) \\ \text{undefined} & (a = 0) \\ 0 & (a < 0) \end{cases}$$

機械学習では、重みの更新に活性化関数の微分を繰り返し計算する必要があるために、その計算に時間がかかってしまうとパラメータの更新回数を増やすことが困難となる。そのために活性化関数を単純化する必要がある。

2.2 誤差逆伝播法

誤差逆伝播法は順伝播と逆伝播で構成される。順伝播では、2.1で述べた畳み込み層で訓練データの入力に対してランダムで決められた重みを用いて出力を行っていく。最後の出力ユニットに、本稿ではsoftmax関数を用いており次のような性質を持つ。

$$S_j(\mathbf{u}) = \frac{\exp(u_j)}{\sum_{k=1}^m \exp(u_k)} \quad (2)$$

ここで m は引数となるベクトル $\mathbf{u}$ の次元であり、 $S_j(\mathbf{u})$ を並べることで得られる m 次元ベクトルがsoftmax関数 $\mathbf{S}$ の値となる。softmax関数の出力は総和1の確率分布ベクトルで表されるために入力画像が何の画像であるかを示す正解ラベルと同次元のベクトルで比較できるため、画像認識及び多値分類において有効である。順伝播でこのsoftmax関数の出力された画像の識別予測と実際の正解データを比較してその誤差を補うために逆伝播にてパラメタの更新を行う。その際、順伝播にて行われた各層の出力値は保存されている。逆伝播ではこの保存された値をもっとも出力に近い重み行列から誤差を少なくするために微分を用いて入力に向かって計算を行う。逆伝播の流れをFig.2に表す。

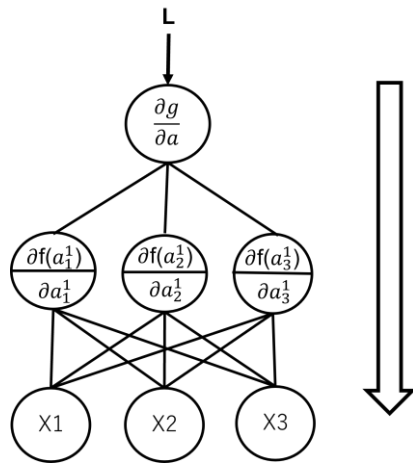


Fig.2 逆伝播の流れ

ここで、 L は正解データと予測データの差を、 $X1, X2, X3$ は入力データを示し、 g はsoftmax関数、 f はReLU関数を示す。この順伝播と逆伝播の一回の操作をepoch数と呼ぶ。機械学習では、入力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させ、それがepoch数となる。ここで、epoch数20、データ数1000、バッチサイズ100と仮定すると、1epochの学習が終わるまでに1000/100=10回

の学習が入力データ数100で行われ、それが20epoch分行われることを表す[2]。

3. 機械学習に与える画像

3.1 RGB画像

RGB画像は撮影した手の動画のフレームレートから抽出した画像をそのまま用いる。

3.2 グレースケール画像

グレースケール画像は、R、G、Bの情報に対してそれぞれ重み0.299、0.587、0.114をかけたものを足し合わせた情報を一つのピクセルに格納している。

3.3 エッジ検出を行った画像

機械学習は与えられる手の画像の背景や影、光の反射光なども学習時に特徴として抽出してしまうため、これらの手の形状とは関係のない情報を抜き取るためにエッジを抽出する。このエッジ画像は入力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。Fig.3はエッジの抽出の過程を表す。

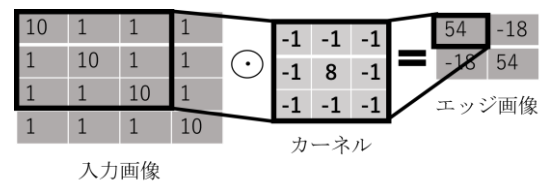


Fig.3 エッジ抽出の過程

Fig.3では、入力画像の右肩下がりの斜め線に対して出力が高い値で、それ以外のピクセルは低い値で出力される。このようにしてエッジを抽出することが可能となる。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[3]。

3.4 入力画像の準備

本稿で用いる入力画像には、三脚と照明を固定し、背景を白くした状態でスマートフォンを用いて撮影したものを用いた。機械学習時のデータ数確保のため、動画にて約5秒間手を撮影し、そのフレームレートから画像を抽出した。その際、手を広げたり閉じたりする動作を行いながら撮影することで様々な形状の手の画像を取得した。ここで取得した手の画像は1920×1080ピクセルである。また、スマートフォン

のフレームレートが30、撮影時間が約5秒、表と裏の手を撮影するために一人当たりのデータ数は $30 \times 5 \times 2 = 300$ 枚となる。本稿ではその入力画像を15人撮影した内の8人分のデータを用いるので約2400枚のデータを用意した。本稿ではこの抽出した画像を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)として画像処理を行い、それぞれを入力画像及びテスト画像として機械学習及び評価を行った。また、学習時の訓練データを9割、正解データを1割として用い、epoch数を20とした。それぞれの画像処理された表の手の画像をFig.4に表す。

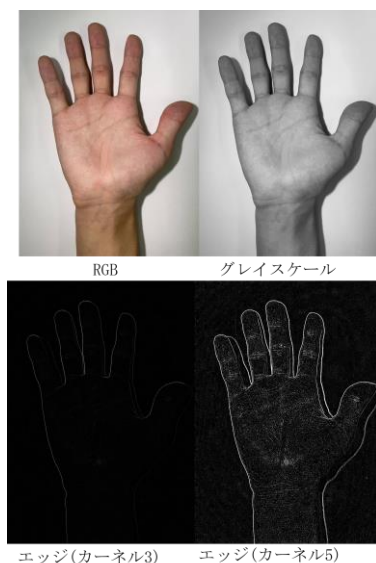


Fig.4 各画像処理を行った入力画像

一人当たりの手の画像データ数約300枚の画像群の内、手の形状が異なる画像を表と裏の手でそれぞれ10枚ずつ抜き取り、それを機械学習の評価時におけるテスト画像として用いる。Python言語のkerasライブラリを用いた機械学習結果にテスト画像を与えると、そのテスト画像が誰の手で表裏のどちらであるかのラベルの中で、確率として一番高い画像が選ばれ、その確率値とラベルが返される。本稿ではその返される確率に対して閾値99.9を設け、閾値以上であればライブラリから返される確率値を元に誰の手で表裏のどちらかのラベルを判定し、閾値未満であればそれに加え「認証できません」と判定させる。本人拒否率の評価時は、8人分の画像群約2400枚の内テスト画像を事前に160枚抜き取った約2240枚を学習させて、テスト画像160枚を学習結果に与えた際に閾値未満で識別されたもの、またはテスト画像と識別された結果が一致しなかったものを、全体のテスト画像160枚で割った割合を本人拒否

率と定義する。他人受入率の評価時は、用意した15人分の画像群の内8人分の画像群約2240枚を学習させて、残りの7人分の未学習の画像群の中から表裏の手のテスト画像140枚を学習結果に与えた際に誤って閾値以上で学習済みの人の手であると識別してしまったものを、全体のテスト画像140枚で割った割合を他人受入率と定義する。Fig.5は閾値未満で認証されなかった時の判定結果を表す。



Fig.5 閾値未満で識別されたテスト画像

4. バッチサイズ可変時の評価

機械学習をする際に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)のそれぞれに対してバッチサイズを32、64、128、256と徐々に増加させていった際の本人拒否率と他人受入率の評価を行った。その際、機械学習に与える画像データの解像度は元の1920ピクセル×1080ピクセルから128ピクセル×128ピクセルにリサイズ処理したものを固定した。Fig.6は各バッチサイズに対する他人受入率を棒グラフで、本人拒否率を折れ線グラフで示したものである。

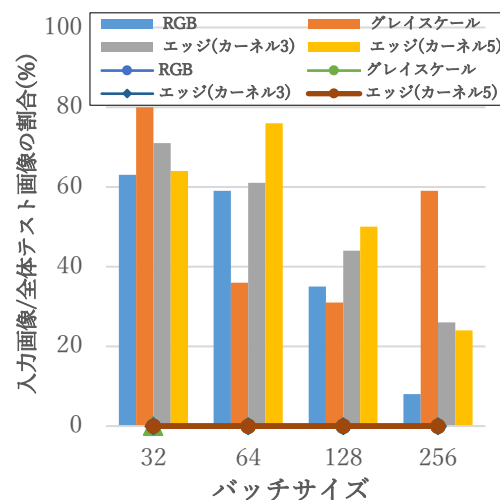


Fig.6 本人拒否率と他人受入率評価

Fig.6の他人受入率の棒グラフよりRGB画像がバッチサイズ256の約8%、グレイスケール画像

がバッチサイズ128の約31%、エッジ画像(カーネル3)がバッチサイズ256の約26%、エッジ画像(カーネル5)がバッチサイズ256の約24%で一番低く良い結果となった。また、Fig.6の本人拒否率の折れ線グラフより、4つの画像項目ともにすべて閾値99.9%以上で本人の手の表裏のどちらであるかを識別できているため、本人拒否率は0%に近い。しかし、各画像項目に対して機械学習ライブラリがテスト画像を本人であると識別して返した確率値は100%が多かったものの、バッチサイズが増加するにつれて確率値100%ではなく本稿で設定した閾値99.9%と100%の間の確率値が散見された。

5. 学習画像の解像度可変時の評価

機械学習をする際にRGB画像、グレースケール画像、エッジ画像(カーネル3)、エッジ画像(カーネル5)のそれぞれに対して画像の解像度の縦×横を32、64、128、256ピクセルと徐々に増加させたものを機械学習の入力データとして本人拒否率と他人受入率の評価を行った。その際、機械学習に与える画像データのバッチサイズは128で固定した。Fig.7は解像度を徐々に増加させたときの他人受入率を棒グラフで、本人拒否率を折れ線グラフで示したものである。

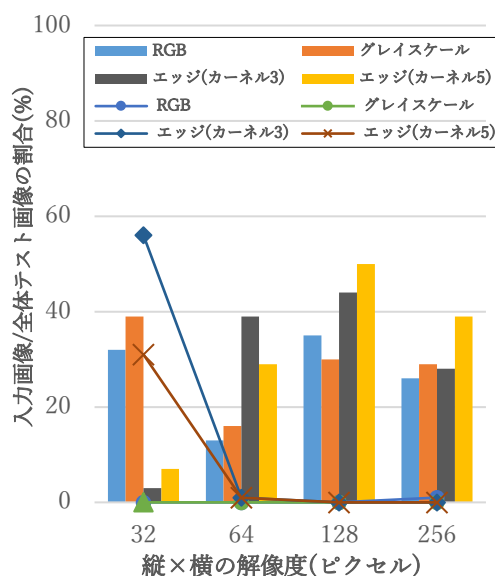


Fig.7 本人拒否率と他人受入率評価

Fig.7の他人受入率の棒グラフよりRGB画像が解像度64で約13%、グレースケール画像が解像度64で約16%、エッジ画像(カーネル3)が解像度32で約3%、エッジ画像(カーネル5)が解像度32で約7%と一番低く良い結果となった。また、Fig.7の本人拒否率の折れ線グラフよりRGB画

像、グレースケール画像ともに本人拒否率0%に近い推移であり、本人の手の表裏どちらかであることを識別できるが、エッジ画像ではカーネル3と5それぞれ解像度32の時約56%、約31%という割合で本人拒否している。また、バッチサイズ可変時と比較して解像度の増加に関係なく全体的に機械学習ライブラリがテスト画像を本人であると識別して返した確率値は、本稿で設定した閾値99.9%と100%の間の確率値が散見された。

6. 他人受入率及び本人拒否率の評価

本人拒否率の評価はバッチサイズ可変評価時ではほとんどが、解像度可変時では縦×横64ピクセル以上でほとんどが閾値99.9%以上で本人の手の裏表を識別できたため本人拒否率は0%に近い結果となった。他人受入率の評価では、バッチサイズ可変時でFig.6よりRGB画像のバッチサイズ256が約8%で他人受入率が一番低い約10回に1回は他人を受入れてしまう。解像度可変時でFig.7よりエッジ画像(カーネル3)の縦×横が32ピクセルの時約3%で他人受入率が一番低い、本人拒否率が約56%という結果となったため、それよりは縦×横が64ピクセルの時のRGB画像の方が他人受入率は劣るものの約13%で、本人拒否率が0%と低いことからこちらの方が実用的と言える。しかしながらこれらの評価は何十回に一回他人を受け入れてしまう点で認証システムの実用性としては程遠いといえるため、今後は他人受入率を、学習時に与える際のリサイズした解像度やバッチサイズといったパラメータを上昇させるなどして改善するとともに、本人拒否率が0%に近い結果が多かったことから、閾値を現状の99.9%と100%の間で調整することでさらなる他人受入率及び本人拒否率の向上に努める。

謝辞 本研究を行うにあたり、画像処理や機械学習に関するご指導を頂きました日本大学生産工学部数理情報工学科 目黒光彦准教授に深く感謝を申し上げます。

参考文献

- [1] 池上美保, 電子パスポートにも採用バイオメトリクス認証3月号, COMZINE, 2004.
- [2] 手塚太郎, しくみがわかる深層学習, 2020.
- [3] 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.