《論　説》

AIで飛躍的に進化する顔画像識別技術

澤　田　雅　之
（澤田雅之技術士事務所所長
元警察大学校警察情報
通信研究センター所長）

要　　旨

今世紀初頭に実用化が始まった顔画像識別技術は、近年ではAI（ニューラルネットワークのディープラーニング）の活用により識別の精度や速度が飛躍的に向上し、「人の目」を遥かに凌駕する驚異的な識別性能の実現に至っている。

例えば、顔の長期経年変化、あるいは、顔の表情や撮影角度の大きな違いなどが影響して、「人の目」には別人としか見えない顔画像であったとしても、数百万人分の中から本人の顔画像を高精度かつ瞬時に選択することができる。また、低解像度や低コントラストなどの低品質顔画像に対する識別精度についても、大幅な改善が見られる。このような識別性能の向上は、警察における被疑者写真検索や指名手配犯の発見にも大いに資するところである。

そこで、米国立標準技術研究所が実施しているFace Recognition Vendor Testの結果等に基づき、社会の多方面における実用化が進展している顔画像識別技術について、その最先端の機能・性能を論説する。

キーワード：顔画像識別技術、AI、ディープラーニング

はじめに

顔画像識別技術は、スマートフォンの顔認証によるロック解除手段として、あるいは、フェイスブックの顔画像自動探知機能などとして、既に身近な存在である。社会インフラとしても、非接触で本人確認ができるという顔画像識別技術の優れた特性を生かして、ICパスポートに記録されている顔画像とビデオカメラで捉えた顔画像との自動照合による出帰国審査の自動化ゲートなどが既に実現している。

このような顔画像識別技術の実用化が始まったのは今世紀の初頭、つまり、わずか20年程前のことである。当初は、アナログビデオカメラで捉えた顔画像の品質が高くはなかったことと、数学モデルに基づいて設計しプログラミングした「顔を識別するアルゴリズム」の性能が高くはなかったことが相乗して、顔画像自動探知機能や出帰国審査の自動化などは全くの夢物語であった。

しかし、10年ほど前から、ビデオカメラのデジタル化に伴う画質の高精細化と機器の小型化が加速度的に進み、高品質な顔画像を身近で捉えることが難しくはなくなった。また、6年ほど前から、「顔を識別するアルゴリズム」を抜本的に改善する切り札として、AI（ニューラルネットワークのディープラーニング）を活用するようになった。つまり、「顔を識別するアルゴリズム」については、人の手でプログラミングすることにより明示的に作成するのではなく、ディープラーニングによりニューラルネットワーク内に暗示的に生成するようになったのである。その結果、識別性能が飛躍的に向上したことから、高精細デジタルビデオカメラを活用することによる相乗効果も発揮され、フェイスブックの顔画像自動探知機能や出帰国審査の自動化ゲートなどの実現に至っている。

ここで、『ディープラーニングによりニューラルネットワーク内で「顔を識別するアルゴリズム」を暗示的に生成する』というところが、最も重要なキーテクノロジーである。しかし、なかなかイメージし難いところで

あるため、このキーテクノロジーについて次章で詳しく解説する。

1　顔画像識別に活用されるAI

(1)　ニューラルネットワークのディープラーニング

AI（Artificial Intelligence）の実現手法は様々であるが、顔画像識別に関して特筆すべきは、ディープラーニングと言われるDNN（Deep Neural Network）を用いた手法である。ディープラーニング以外の手法では、人がAIのアルゴリズムを、つまり、入力された情報を処理して結果を出力するまでの具体的な仕組みと働き方（情報の処理方法）を、数学モデルに基づき明示的に設計してプログラミングする必要がある。他方、ディープラーニングでは、入力された情報とそれに対する望ましい出力結果の組合せをDNNに反復して学習させることによって、DNNのネットワークを次第に鍛え上げていき、実現したいAIのアルゴリズムをネットワーク全体にわたって暗示的に生成する。このようなアルゴリズムの生成は、人が経験を積み重ね、学習や訓練を反復することによって、頭脳内部で識別・判断・対処の能力を次第に高めていくことと同じである。なぜならば、ニューラルネットワークとは、人の頭脳内部での神経回路網の仕組みと働きを、「入力層―隠れ層―出力層」としてコンピュータ上で数学的に模したものだからである。また、DNNのDeepとは、識別・判断・対処の能力向上を図るために、隠れ層を多層化して深く学習（ディープラーニング）できるようにしたことを意味する。それゆえ、人には得意・不得意の分野や能力の高低があるように、DNNにも、その構成の仕方の違いにより得意・不得意の分野が生じるのであり、また、その鍛え方（学習の方法や学習用教材）の違いにより識別・判断・対処の能力に高低が生じるのである。

ここで、DNNの学習についてであるが、学習に用いた教材を十分にマスターすることにより、教材以外にも応用が利くようにすること（汎用化、汎化）が最終目的である。例えば、インターネットに投稿された映像や画像の中から「ありとあらゆる猫」を自動的に見つけ出すことを最終目的としてみよう。この場合の学習としては、インターネットから抽出した数万枚の猫の画像をDNNの入力層への教材として用いて、猫の画像の1枚ごとに、DNNの出力層における「分類結果」を「猫」に少しずつ近づけていく作業を反復するのである。具体的な作業内容としては、「分類結果が猫ではないとした誤差の度合い」をDNNの隠れ層内に逆伝播させることにより、隠れ層のノード（ネットワークの結節点）の間における信号の伝わりやすさ（重み付け値）やノードが信号を発する閾（バイアス値）を、「分類結果の誤差の度合い」が小さくなる方向に少しずつ変化させる。このような学習を多数の教材を用いて反復していくうちに、隠れ層内には「普遍的な猫の特徴」を抽出して「猫という概念に昇華」するアルゴリズムが、ノードへの信号入力時の重み付け値やノードからの信号出力時のバイアス値の集合体として、暗示的に生成される（これがDNNの「学習フェーズ」）ことになる。この学習フェーズで手間暇をかけて「汎化したアルゴリズム」を生成すれば、その後はこのアルゴリズムを「猫を見つけ出す画像フィルタ」として用いて、インターネットに投稿された映像や画像の中から「ありとあらゆる猫」を自動的に見つけ出す（これがDNNの「推論フェーズ」）ことができる。

ちなみに、学習フェーズで気を付けなければならないのは、「過学習」である。多数の教材を何度も反復して学習させていくうちに、いつの間にか最終目的である「汎化」を通り越してしまい、教材以外には応用が利かなくなる現象がDNNで生じることがあるが、これが「過学習」である。人に例えてみれば、下手なガリ勉のやり過ぎで、丸暗記したこと以外には応用が利かなくなることと同じと言える。

ところで、この「ありとあらゆる猫を見つけ出す画像フィルタ」のような「汎化したアルゴリズム」を、人の手で明示的に設計してプログラミングすることはおそらく不可能である。このような困難さこそ、過去の2度（1950～1960年代と1980～1990年代）にわたるAIブームが、いずれも一過性のブームで終わってしまった最大の原因である。しかし、ディープラ

ーニングであれば、「ありとあらゆる猫を見つけ出す画像フィルタ」のような「汎化したアルゴリズム」を生成することができる。つまり、人の手で設計してプログラミングする従来の手法では解決が困難であった様々な技術的課題や問題について、ディープラーニングの適用と工夫により解決できる道筋が開けたと言えるのである。

2012年に開催された国際的な画像認識コンペティション（ILSVRC2012）において、初登場したディープラーニングによる認識手法は、飛躍的な認識性能を発揮して、人の手で明示的に設計してプログラミングした認識手法を全て圧倒した。これを契機として、大いに注目を集めるようになったディープラーニングであるが、その有用性は瞬く間に多方面で認められ、顔画像識別への活用を始め、多言語間の自動翻訳や車の自動運転など、今日では広範な活用が進展し社会に広く深く浸透しつつある。このことから、ディープラーニングによる目下のAIブームは、この先も一過性のブームで終わることはないのであり、ディープラーニングこそ、これからのSociety5.0（超スマート社会）を支えていく最も重要なキーテクノロジーであると言っても過言ではない。

なお、ディープラーニングは、DNNの構成の仕方によって、2次元静止画像を入力として事物の分類などができるCNN（Convolutional Neural Network：畳み込みニューラルネットワーク）を用いる場合と、時系列データを入力として予測などができるRNN（Recurrent Neural Network：再帰型ニューラルネットワーク）を用いる場合に大別される。顔画像識別にはCNNが適するので、その構成や具体的な用い方について次節に記載する。

(2) 顔画像識別に適するCNN（畳み込みニューラルネットワーク）

CNN（Convolutional Neural Network：畳み込みニューラルネットワーク）は、2次元静止画像を入力として、画像内の事物をカテゴリー分け（あらかじめ決められた項目に分類）したり、クラスター分け（類似の特徴を備えたグループに配属）したりすることができる。

CNNの最もシンプルな構成は、「入力層―隠れ層（畳み込み層―プーリング層―全結合層）―出力層」であり、入力された信号は、隠れ層内の各層を含めて、入力層から出力層までのネットワーク全体をフィードフォワードで伝播し、フィードバックされることはない。隠れ層内の各層にはそれぞれ役割があり、畳み込み層は、入力層のノード全体を細分化して部分的な特徴パターンで一斉にスキャンすることにより、部分的な特徴を抽出してその場所をマッピングする画像フィルタとして働くのであり、抽出しようとした部分的な特徴の数に応じたチャネル数のマップが出力される。プーリング層は、マッピングした部分的な特徴を損なわないようにマップ全体を縮める（つまり、重要な情報を失わないように圧縮する）ことにより、マップ内における部分的な特徴の位置変動の影響を減ずるものであり、出力されるチャネル数は変わらない。全結合層は、この層に入力された全てのチャネルの特徴を統合するように働くのであり、カテゴリー分けの場合には、SoftMax層（SoftMax関数を用いてカテゴリーごとの信頼度〈出力確率〉の総和を1とする層）を介して出力され、クラスター分けの場合には、128次元などの数値ベクトルに集約して出力される。このため、カテゴリー分けの場合には、出力層に分類結果が信頼度の数値とともに出力されるが、クラスター分けの場合には、クラスター分布のベクトル空間において、出力層に出力される数値ベクトルを用いたベクトル間の距離計算を行うことにより、クラスター分け作業を行う必要がある。

ちなみに、CNNの入力層は、静止画像の画素数に応じた2次元空間（例えば、ハイビジョンのフレーム画像であれば1920×1080ノード）であり、モノクロ画像ではチャネル数は1で、RGB画像ではチャネル数は3となる。高精細画像を扱う場合には、隠れ層内の「畳み込み層―プーリング層―全結合層」の構成は多層化され、「[複数の畳み込み層―プーリング層]―[複数の畳み込み層―プーリング層]―【多層（略）】―[複数の畳み込み層―プーリング層]―[複数の全結合層]」といった具合に、100層以上で構成されることもある。

ところで、CNNは、RPN（Region Proposal Network：領域提案ネッ

トワーク）を組み込むことにより、R-CNN（Regions with Convolutional Neural Network：領域提案できる畳み込みニューラルネットワーク）に機能を拡張できる。R-CNNでは、あらかじめ学習した事物やパターンを２次元静止画像の中から全て見つけ出し、それらが写っている領域を四角形の枠で個々に特定し、それらが何であるかを個々に分類して、信頼度の数値とともに分類結果（カテゴリー）を個々に表示することができる。ちなみに、映像（動画）は、２次元静止画像であるフレーム画像が１秒間に数十枚連続したものであるため、フレーム画像を１枚ずつ順番にR-CNNで高速処理すれば、映像（動画）をリアルタイムに解析することが可能となる。

そこで、顔画像識別の一連のプロセスでは、R-CNNとCNNがそれぞれ次のように用いられる。

ア　顔画像の検出

顔画像の検出とは、フレーム画像などの２次元静止画像の中から「人の顔」を見つけ出して、その写っている領域を特定することであるが、これにはR-CNNを用いる。多数の「人の顔」が写っている場合であっても、R-CNNであれば、瞬時に「人の顔」を全て見つけ出して、それらが写っている領域を四角形の枠で個々に特定し、カテゴリーとしての「人の顔」らしさを表す信頼度の数値を枠の近傍に表示することができる。

このようにして「人の顔」を検出するR-CNNの学習フェーズとしては、様々な条件下（性別、年齢、人種、表情、撮影角度、髪型、ひげの有無、帽子・眼鏡・マスク等の有無、ピントやコントラストなど）で撮影された多数の「人の顔」をR-CNNの入力層への教材として用いて、「人の顔」の１枚ごとに、R-CNNの出力層における「分類結果」を「人の顔」に少しずつ近づけていく（つまり、カテゴリーとしての「人の顔」らしさを表す信頼度の数値を少しずつ高めていく）作業を反復する。具体的な作業内容としては、『分類結果が「人の顔」ではないとした誤差の度合い』をR-CNNのネットワーク内に逆伝播させることにより、

ネットワークのノード（結節点）の間における信号の伝わりやすさ（重み付け値）やノードが信号を発する閾（バイアス値）を、『分類結果が「人の顔」ではないとした誤差の度合い』が小さくなる方向に少しずつ変化させる。このような学習を多数の教材を用いて反復していくうちに、R-CNNのネットワーク内には『普遍的な「人の顔」の特徴』を抽出して『「人の顔」という概念に昇華』するアルゴリズムが、ノードへの信号入力時の重み付け値やノードからの信号出力時のバイアス値の集合体として、暗示的に生成されるのである。

イ　顔画像の識別

顔画像の識別とは、前記のR-CNNで検出して切り出した顔画像と識別対象顔画像との「類似度」を計算により求めることである。この「類似度」は、顔の特徴を数値ベクトルで表現した「特徴ベクトル」を用いて、ベクトル間の距離計算により算出するのであるが、この「特徴ベクトル」の生成にはCNNが用いられる。つまり、顔画像のクラスター分けにCNNを用いて、入力した顔画像の様々な特徴を抽出して、それらを128次元などの数値ベクトルに集約した「特徴ベクトル」として出力するのである。このため、顔画像の識別とは、具体的には、識別対象顔画像からCNNで生成した「特徴ベクトル」が分布するベクトル空間において、R-CNNで検出して切り出した顔画像からCNNで生成した「特徴ベクトル」との距離計算を行うことにより、識別対象顔画像との類似度をベクトル間距離という数値で求めることに等しい。

このようにして「特徴ベクトル」を生成するCNNの学習フェーズとしては、多様性（性別、年齢、人種など）のある多くの人々の顔画像を教材として準備するとともに、同一人物についても様々な条件下（年齢、表情、撮影角度、ひげの有無、眼鏡・マスク等の有無、解像度やコントラストなど）で撮影された多数の顔画像を準備することから始める。次に、任意の同一人物の任意の２枚の顔画像を選択するとともに、任意の別人の任意の１枚の顔画像を選択する。その次に、これらの３枚の顔画像を１つのセットとしてCNNの入力層への教材として用いて、出力さ

れる3つの「特徴ベクトル」を比較する。そして、同一人物の2つの「特徴ベクトル」の間の距離を少し狭めるとともに、別人の「特徴ベクトル」との距離を少し広げるよう、「距離の開き具合」をCNNのネットワーク内に逆伝播させることにより、ネットワークのノード（結節点）の間における信号の伝わりやすさ（重み付け値）やノードが信号を発する閾（バイアス値）を少しずつ変化させる。このような学習を多数の教材を用いて反復していくうちに、CNNのネットワーク内には、様々に異なる条件下（性別、年齢、人種、表情、撮影角度、ひげの有無、眼鏡・マスク等の有無、解像度やコントラストなど）であっても、別人の顔画像の「特徴ベクトル」からは離間するが、同一人物の顔画像であれば近接する「特徴ベクトル」を生成するアルゴリズムが、ノードへの信号入力時の重み付け値やノードからの信号出力時のバイアス値の集合体として、暗示的に生成されるのである。

2　米国立標準技術研究所の顔認識技術ベンダーテスト

米国立標準技術研究所（NIST：National Institute of Standards and Technology）では、顔認識技術に係る各種の評価試験を2000年以来実施している。中でも、2013年と2018年に実施した顔認識技術ベンダーテスト（FRVT：Face Recognition Vendor Test）は、「顔を識別するアルゴリズム」の生成にAI（CNNのディープラーニング）を活用することによる識別性能の改善効果を理解する上で重要である。なぜならば、2013年のFRVTではディープラーニングを活用したベンダーは皆無であるが、2018年のFRVTでは多くのベンダーがディープラーニングを活用し、2013年と類似したテスト環境における識別精度が桁違いに向上しているからである。ちなみに、2013年のFRVTの結果は、「NIST Interagency Report 8009」[1]として2014年5月に公表され、2018年のFRVTの結果は、「NIST Interagency Report 8271」[2]として2019年9月に公表されている。そこで、「顔を識別するアルゴリズム」へのディープラーニングの活用効果を見極めるため、公表されたReportに基づき、2013年と2018年のFRVTの結果について以下に記載する。

(1)　2013年と2018年のFRVTに参加したベンダー

2013年のFRVTに参加したベンダーは、我が国の3社、米国の3社、中国の3社と4研究機関、フランスの1社、ドイツの1社、リトアニアの1社である。

2018年のFRVTに参加したベンダーは、我が国の4社、米国の13社、中国の9社と1研究機関、台湾の1社、韓国の2社、ロシアの3社、フランスの2社、ドイツの3社、チェコの2社、リトアニアの1社、英国の1社、オーストラリアの1社、スロバキアの1社、クロアチアの1社、フィンランドの1社、オランダの1社、インドの1社、国籍不詳の1社である。

この中で、我が国の3社、ドイツの1社、フランスの2社（内の1社は、2013年の時点では米国籍）、リトアニアの1社、中国の1社は、2013年と2018年のFRVTのいずれにも参加している。それゆえ、2013年と2018年のFRVTでの類似したテスト環境下における、これらのベンダーごとの識別精度を見比べることにより、「顔を識別するアルゴリズム」の生成へのディープラーニングの活用効果が見えてくる。そこで、この点の詳細について、後記の(4)項に記載する。

(2)　2013年と2018年のFRVTで用いられた顔画像

2013年と2018年のFRVTでは、警察のMugshot画像（数百万人分）と国境警備隊のWebカメラ画像（数十万人分）が主に用いられた。この中には、Mugshot画像を複数回撮影された累犯者が多数含まれており、また、MugshotとWebカメラの双方で顔画像を撮影された者も多数含まれている。このため、Mugshot画像とWebカメラ画像の品質面（緻密さ、鮮明さ、顔の撮影角度）の違いなどが、1対多数の顔画像照合時における識別精度に及ぼす影響について、定量的な評価が可能となる。そこで、Mugshot画像とWebカメラ画像の品質などについて、以下に記載する。

ア　Mugshot 画像

米国の警察の実際の現場でデジタルスチルカメラにより撮影された高品質な顔画像であり、我が国の被疑者写真に相当する。顔画像の緻密さの指標となる目間画素数（顔画像上の両目の中心を結ぶ線上に並ぶ画素の総数）は、平均で107画素であるため緻密である。適切な照明の下で、顔を正面から鮮明に捉えた画像が大半であり、左右を向いている顔画像であっても、その撮影角度は顔の真正面から見て５度から10度ほどにすぎない。

なお、Mugshot 画像は、正面から捉えた顔画像（正面顔）と真横から捉えた顔画像（横顔）がセットになっているので、2018年の FRVT では、横顔で正面顔を検索した場合の識別精度についても試験されている。

イ　Web カメラ画像

米国の国境警備隊が勾留した中南米からの不法入国者を、国境警備隊執務室卓上の Web カメラにより撮影した顔画像である。隊員の指示に基づき、Web カメラに顔を向けた瞬間を捉えている。目間画素数は平均で45画素であり、Mugshot 画像に比べて緻密さに欠けている。室内照明の下で撮影されているため、顔に影ができて低コントラストである顔画像が大半であり、Mugshot 画像に比べて鮮明さにも欠けている。顔を真正面から捉えた画像は少なく、５度から10度ほど左右を向いている顔画像や、下からやや見上げるように撮影された顔画像が大半である。

ちなみに、2013年と2018年の FRVT で用いられた Web カメラ画像は、前記のとおり緻密さと鮮明さの点で高品質とは言えないが、今日の防犯・監視カメラ市場で流通している高精細デジタルビデオカメラを用いれば、ワイドダイナミックレンジ機能や３次元ノイズリダクション機能の活用により、室内照明下であっても緻密で鮮明な顔画像を撮影することが可能である。

(3)　2013年と2018年の FRVT で用いられた２種類の識別精度評価方法

2013年と2018年の FRVT では、１対１の顔画像認証（つまり、同一人物であるか否かを顔画像により確認すること）の識別精度ではなく、１対多数の顔画像照合（つまり、同一人物の顔画像を多数の顔画像の中から探し出すこと）の識別精度を試験により評価している。１対多数の顔画像照合では、前記 **1** の(2)のイ項に記載のとおり、探し出そうとする人物の顔画像（検索用顔画像）と、探す対象となる多数の顔画像（被検索顔画像）の各々との間の「類似度」を算出することが基本となる。この算出は、具体的には、検索用顔画像から生成した「特徴ベクトル」と、各被検索顔画像から生成した各「特徴ベクトル」との間の距離を計算することにより行う。このため、１対多数の顔画像照合の結果は、検索用顔画像との「類似度」が高い順（つまり、特徴ベクトル間の距離が短い順）に、被検索顔画像がスコア値（特徴ベクトル間距離の短さに応じた１～０の正数）とともにリストアップされる。そこで、１対多数の顔画像照合における識別精度の評価方法としては、スコア値に対する「閾値」を設けるか否かにより、次の２種類に大別される。

ア　スコア値に対する「閾値」を設けない場合

１対多数の顔画像照合を行った結果として、検索用顔画像との「類似度」のスコア値が高い順に被検索顔画像がリストアップされる。これらのスコア値に対する「閾値」を設けない場合の識別精度を評価するには、多数の検索用顔画像を用いて、その都度に出力された前記のリストにおいて、検索用顔画像と同一人物の被検索顔画像が、TOP１（つまり、リストの最上位）、あるいは、TOP50（つまり、リストの上位50位以内）にリストアップできなかった割合（つまり、本人見逃し率）を調べるのである。ここで、スコア値に対する「閾値」を設けなければ、検索用顔画像と同一人物の被検索顔画像が存在しない場合の TOP１における「他人誤認率」が100％となることに注意を要する。

スコア値に対する「閾値」を設けない１対多数の顔画像照合としては、防犯カメラ等の録画映像に遺留された犯人の顔画像に基づき、被疑者写

真データベースを検索・照合し、出力されたTOP5,000の被疑者写真を上位から順番に一枚ずつ「人の目」により犯人の顔画像と対照して、犯人の身元を割り出すといったことが挙げられる。

イ　スコア値に対する「閾値」を設ける場合

1対多数の顔画像照合を行った結果として、検索用顔画像との「類似度」のスコア値が高い順に被検索顔画像がリストアップされる。スコア値に対する「閾値」を設ける場合の識別精度を評価するには、多数の検索用顔画像を用いて、その都度に出力された前記のリストにおいて、検索用顔画像と同一人物の被検索顔画像が、設けた「閾値」を上回ってTOP1（つまり、リストの最上位）にリストアップできなかった割合（本人見逃し率）を調べるのである。このため、本人の顔画像とよく似た他人の顔画像とをどの程度まで区別することができるのかといった性能面（顔画像の「類似度」のスコア値における本人と他人との隔たりが大きいほど、本人と他人とを「閾値」で区別する性能が高いと言える。）について、具体的に調べることができる。

ここで、「本人見逃し率」と「他人誤認率」は、トレードオフの関係にあることに注意を要する。つまり、「閾値」を高く設定して「他人誤認率」を小さくするほど、「本人見逃し率」は大きくなるのであり、逆に、「閾値」を低く設定して「他人誤認率」を大きくするほど、「本人見逃し率」は小さくなるのである。そこで、スコア値に対する「閾値」を設ける場合の識別精度の評価は、「他人誤認率」が1/10、1/100、1/1000となるように「閾値」をパラメータとして設定して、それぞれの「閾値」における「本人見逃し率」を調べることにより行う。

スコア値に対する「閾値」を設ける1対多数の顔画像照合としては、スマートグラスのウェアラブルカメラで捉えた不審人物の顔画像に基づき、指名手配写真データベースを検索・照合し、「他人誤認率」が極力小さくなるように「閾値」を設定してスマートグラスに出力されたTOP1の指名手配写真を、目前にいる不審人物の顔とスマートグラスを通して対照して、指名手配犯を発見するといったことが挙げられる。

ちなみに、ここに例示した活用方法は、近年の中国における指名手配犯の発見に実際に用いられているものである。

(4) 2013年と2018年のFRVTの対比で見えてくるAIの活用効果

下記の①～⑥は、2013年と2018年のFRVTのいずれにも参加したベンダーについて、160万人分（160万枚）の顔画像データベースに対して、高品質なMugshot画像により検索した場合の「本人見逃し率」と、品質の劣るWebカメラ画像により検索した場合の「本人見逃し率」を示したものである。この「本人見逃し率」とは、顔画像データベースの検索を多数回行った結果として、「閾値」を設けずにTOP1（つまり、検索結果リストの最上位）に、検索用顔画像と同一人物のデータベース内顔画像がリストアップされた割合の補数である。例えば、「本人見逃し率」が0.003であれば、顔画像データベースの検索を1,000回行った内の997回において、検索用顔画像と同一人物のデータベース内顔画像がTOP1にリストアップされたことを意味する。

ここで、2013年のFRVTでは、160万人分（160万枚）の顔画像データベース（内訳：150万3,115人分のMugshot画像と9万6,885人分のWebカメラ画像）に対して、5万人分（5万枚）のMugshot画像（データベース内に同一人物の画像が存在するが、データベース内とは別の画像）を用いて5万回の検索を行い、次いで、1万660人分（1万660枚）のWebカメラ画像（データベース内に同一人物の画像が存在するが、データベース内とは別の画像）を用いて1万660回の検索を行っている。

次に、2018年のFRVTでは、160万人分（160万枚）の顔画像データベース（全てMugshot画像）に対して、15万4,549人分（15万4,549枚）のMugshot画像（データベース内に同一人物の画像が存在するが、データベース内とは別の画像）を用いて15万4,549回の検索を行い、次いで、8万2,106人分（8万2,106枚）のWebカメラ画像（データベース内に同一人物の画像が存在するが、データベース内とは別の画像）を用いて8万2,106回の検索を行っている。

なお、2013年と2018年のFRVTで用いられた顔画像データベースはいずれも160万人分（160万枚）であるが、2018年の顔画像データベースは全て高品質なMugshot画像であるのに対して、2013年の顔画像データベースには品質の劣るWebカメラ画像が全体の６％ほど（160万枚中の９万6,885枚）含まれている。このような品質の劣る顔画像の存在は、「本人見逃し率」の悪化に繋がる。そこで、悪化の度合いを推測するための目安として、2013年のFRVTにおいて、２万人分（２万枚）の顔画像データベース（全てMugshot画像）に対して、２万人分（２万枚）のMugshot画像（データベース内に同一人物の画像が存在するが、データベース内とは別の画像）を用いて、２万回の検索を行った結果としての「本人見逃し率」を、下記の①～⑥の「Mugshot画像による検索時」における括弧内に示す。

① 我が国のＮ社（2013年と2018年のFRVTでの識別性能がトップ）
Mugshot画像による検索時：2013年のFRVTでは0.041（0.028）であったが、2018年のFRVTでは0.003に改善
Webカメラ画像による検索時：2013年のFRVTでは0.113であったが、2018年のFRVTでは0.010に改善

② フランスのＴ社
Mugshot画像による検索時：2013年のFRVTでは0.172（0.105）であったが、2018年のFRVTでは0.006に改善
Webカメラ画像による検索時：2013年のFRVTでは0.364であったが、2018年のFRVTでは0.020に改善

③ 我が国のＴ社
Mugshot画像による検索時：2013年のFRVTでは0.107（0.060）であったが、2018年のFRVTでは0.007に改善
Webカメラ画像による検索時：2013年のFRVTでは0.237であったが、2018年のFRVTでは0.022に改善

④ リトアニアのＮ社
Mugshot画像による検索時：2013年のFRVTでは0.205（0.142）であったが、2018年のFRVTでは0.007に改善
Webカメラ画像による検索時：2013年のFRVTでは0.702であったが、2018年のFRVTでは0.024に改善

⑤ ドイツのＣ社
Mugshot画像による検索時：2013年のFRVTでは0.136（0.085）であったが、2018年のFRVTでは0.008に改善
Webカメラ画像による検索時：2013年のFRVTでは0.576であったが、2018年のFRVTでは0.025に改善

⑥ フランスのＩ社
Mugshot画像による検索時：2013年のFRVTでは0.091（0.068）であったが、2018年のFRVTでは0.009に改善
Webカメラ画像による検索時：2013年のFRVTでは0.307であったが、2018年のFRVTでは0.032に改善

さて、上記の①～⑥から、全てのベンダーにおいて、高品質なMugshot画像による検索時と品質の劣るWebカメラ画像による検索時のいずれについても、2013年（「Mugshot画像による検索時」の括弧内に示した数値を含めて）と比べて2018年のFRVTでは、「本人見逃し率」の数値に桁違いの改善が見られる。ディープラーニングは、2013年のFRVTではまだ活用されておらず、2018年のFRVTでは活用が一気に進んだことから、識別性能が飛躍的に向上した主因はディープラーニングの活用にあると言える。そこで、2018年のFRVTからうかがえる最先端の顔画像識別技術の識別性能について、次章に多角的な視点から記載する。

3　2018年のFRVTからうかがえる最先端の顔画像識別技術

(1) 品質の劣る顔画像（Webカメラ画像）に対する識別精度

下記の①～④は、2018年のFRVTにおいて、品質の劣る顔画像（Web

カメラ画像）に対する識別精度（TOP１における「本人見逃し率」）が、参加ベンダーの中で最も優れていた上位４社の結果である。

ここでのTOP１における「他人誤認率」や「本人見逃し率」の求め方については、次のとおりである。

160万人分（160万枚）の顔画像データベース（全て高品質なMugshot画像）に対して、33万1,254人分（33万1,254枚）のWebカメラ画像（いずれもデータベース内に同一人物のMugshot画像が含まれていない。）を用いて33万1,254回の検索を行い、「他人誤認率」が1/10、1/100、1/1000となる「閾値」を求める。

次に、前記の顔画像データベースに対して、８万2,106人分（８万2,106枚）のWebカメラ画像（いずれもデータベース内に同一人物のMugshot画像が含まれている。）を用いて８万2,106回の検索を行い、「閾値」を設けない場合のTOP１における「本人見逃し率」や、「他人誤認率」が1/10、1/100、1/1000となる「閾値」を設けた場合のTOP１における「本人見逃し率」を求める。

① 我が国のＮ社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP１）：0.017
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP１）：0.013
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP１）：0.011
閾値を設けない場合の本人見逃し率（TOP１）：0.010

② 中国のＹ社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP１）：0.027
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP１）：0.017
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP１）：0.011
閾値を設けない場合の本人見逃し率（TOP１）：0.008

③ 米国のＭ社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP１）：0.037
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP１）：0.024
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP１）：0.016
閾値を設けない場合の本人見逃し率（TOP１）：0.011

④ 中国のＳ社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP１）：0.063
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP１）：0.040
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP１）：0.025
閾値を設けない場合の本人見逃し率（TOP１）：0.016

さて、上記の４社の「閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP１）」は、0.017～0.063である。これらの数値は、160万人分（160万枚）の顔画像データベース（全て高品質なMugshot画像）に対して、品質の劣るWebカメラ画像で検索した結果であり、しかも、「他人誤認率」が1/1000となる「閾値」を設けた上での検索結果である。ところで、前記**２**の⑷の①に記載した、我が国のＮ社の「2013年のFRVTにおける、高品質なMugshot画像による検索時での、閾値を設けない場合の本人見逃し率（TOP１）」は、0.041（0.028）であった。特に、括弧内の0.028という「本人見逃し率」は、２万人分（２万枚）の顔画像データベース（全て高品質なMugshot画像）に対して、高品質なMugshot画像による検索を行った結果であり、しかも、「閾値」を設けてはいない検索結果である。このことから、上記の①～④で示した４社の「品質の劣る顔画像（Webカメラ画像）に対する識別精度」は、2013年のFRVTにおけるトップベンダー（我が国のＮ社）による「高品質な顔画像（Mugshot画像）に対する識別精度」を凌駕する水準に達していると言える。

⑵ 顔の経年変化に対する識別精度

下記の①～④は、2018年のFRVTにおいて、最長で18年に及ぶ顔の経年変化に対する識別精度（TOP１における「本人見逃し率」）が、参加ベンダーの中で最も優れていた上位４社の結果である。

ここでのTOP１における「他人誤認率」や「本人見逃し率」の求め方

については、次のとおりである。

300万人分（300万枚）の顔画像データベース（全てMugshot画像）に対して、33万1,254人分（33万1,254枚）のMugshot画像（いずれもデータベース内に同一人物のMugshot画像が含まれていない。）を用いて33万1,254回の検索を行い、「他人誤認率」が1/1000となる「閾値」を求める。

次に、１人につき複数枚のMugshot画像がある306万8,801人の累犯者について、最も古いMugshot画像のみを306万8,801枚集めたものを顔画像データベースとして、それ以外のMugshot画像（285万3,221人分の1095万1,064枚）を用いて顔画像データベースを1095万1,064回検索する。そして、最も古いMugshot画像と検索に用いたMugshot画像との間の経過年数（例えば、【２～４年の隔たり】や【10～12年の隔たり】など）をパラメータとして、「他人誤認率」が1/1000となる「閾値」を設けた場合のTOP１における「本人見逃し率」を求める。

以下の①～④の各欄に記載した数値は、「他人誤認率」が1/1000となる「閾値」を設けた場合のTOP１における「本人見逃し率」である。

① 我が国のＮ社

【０～２年の隔たり】における「本人見逃し率」：0.007
【２～４年の隔たり】における「本人見逃し率」：0.009
【４～６年の隔たり】における「本人見逃し率」：0.011
【６～８年の隔たり】における「本人見逃し率」：0.013
【８～10年の隔たり】における「本人見逃し率」：0.015
【10～12年の隔たり】における「本人見逃し率」：0.017
【12～14年の隔たり】における「本人見逃し率」：0.021
【14～18年の隔たり】における「本人見逃し率」：0.027

② 中国のＹ社

【０～２年の隔たり】における「本人見逃し率」：0.012
【２～４年の隔たり】における「本人見逃し率」：0.020
【４～６年の隔たり】における「本人見逃し率」：0.031

【６～８年の隔たり】における「本人見逃し率」：0.047
【８～10年の隔たり】における「本人見逃し率」：0.067
【10～12年の隔たり】における「本人見逃し率」：0.096
【12～14年の隔たり】における「本人見逃し率」：0.14
【14～18年の隔たり】における「本人見逃し率」：0.20

③ 米国のＭ社

【０～２年の隔たり】における「本人見逃し率」：0.027
【２～４年の隔たり】における「本人見逃し率」：0.047
【４～６年の隔たり】における「本人見逃し率」：0.072
【６～８年の隔たり】における「本人見逃し率」：0.10
【８～10年の隔たり】における「本人見逃し率」：0.13
【10～12年の隔たり】における「本人見逃し率」：0.16
【12～14年の隔たり】における「本人見逃し率」：0.20
【14～18年の隔たり】における「本人見逃し率」：0.26

④ オランダのＶ社

【０～２年の隔たり】における「本人見逃し率」：0.048
【２～４年の隔たり】における「本人見逃し率」：0.080
【４～６年の隔たり】における「本人見逃し率」：0.13
【６～８年の隔たり】における「本人見逃し率」：0.17
【８～10年の隔たり】における「本人見逃し率」：0.21
【10～12年の隔たり】における「本人見逃し率」：0.24
【12～14年の隔たり】における「本人見逃し率」：0.30
【14～18年の隔たり】における「本人見逃し率」：0.36

さて、我が国のＮ社の『【14～18年の隔たり】における「本人見逃し率」』の数値（0.027）は、他の３社の数値（0.20～0.36）と比べて桁違いに優れている。このことは、前記**２**の(3)のア項に記載した被疑者写真データベースの検索・照合において大きな意味がある。なぜならば、被疑者写真データベースには、数十年前に撮影した被疑者写真も多いからである。

なお、数十年に及ぶ顔の経年変化に対する識別精度を向上させるには、「顔を識別するアルゴリズム」をディープラーニングにより生成する「学習フェーズ」において、より多くの「同一人物の昔の顔写真と今の顔写真のセット」を教材として用いて反復学習することが効果的である。

⑶　真横顔に対する識別精度

下記の①～④は、2018年のFRVTにおいて、Mugshot画像（正面顔）に対して、別の時点で撮影したMugshot画像（真横顔）で検索した場合の識別精度（TOP 1における「本人見逃し率」）が、参加ベンダーの中で最も優れていた上位4社の結果である。

ここでのTOP 1における「他人誤認率」や「本人見逃し率」の求め方については、次のとおりである。

160万人分（160万枚）の顔画像データベース（全て正面顔のMugshot画像）に対して、10万人分（10万枚）の真横顔のMugshot画像（いずれもデータベース内に同一人物のMugshot画像が含まれていない。）を用いて10万回の検索を行い、「他人誤認率」が1/10、1/100、1/1000となる「閾値」を求める。

次に、前記の顔画像データベースに対して、10万人分（10万枚）の真横顔のMugshot画像（いずれもデータベース内に、撮影時点が異なる同一人物のMugshot画像が含まれている。）を用いて10万回の検索を行い、「閾値」を設けない場合のTOP 1における「本人見逃し率」や、「他人誤認率」が1/10、1/100、1/1000となる「閾値」を設けた場合のTOP 1における「本人見逃し率」を求める。

① 米国のM社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP 1）：0.203
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP 1）：0.148
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP 1）：0.109
閾値を設けない場合の本人見逃し率（TOP 1）：0.089

② オランダのV社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP 1）：0.461
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP 1）：0.322
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP 1）：0.198
閾値を設けない場合の本人見逃し率（TOP 1）：0.130

③ ロシアのN社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP 1）：0.566
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP 1）：0.443
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP 1）：0.317
閾値を設けない場合の本人見逃し率（TOP 1）：0.208

④ 我が国のN社

閾値（他人誤認率が1/1000）を設けた本人見逃し率（TOP 1）：0.664
閾値（他人誤認率が1/100）を設けた本人見逃し率（TOP 1）：0.479
閾値（他人誤認率が1/10）を設けた本人見逃し率（TOP 1）：0.340
閾値を設けない場合の本人見逃し率（TOP 1）：0.272

さて、上記の4社の「閾値を設けない場合の本人見逃し率（TOP 1）」は、0.089～0.272である。これらの数値は、160万人分（160万枚）の顔画像データベース（全て高品質な正面顔のMugshot画像）に対して、高品質な真横顔のMugshot画像での検索を行った結果であることから、前記**2**の⑶のア項に記載した被疑者写真データベースの検索・照合において大きな意味がある。なぜならば、犯人の遺留顔画像の撮影角度が真横に近かったとしても高品質（顔画像が緻密で鮮明）であれば、被疑者写真データベースを検索して得られるTOP5,000の中に、犯人の被疑者写真をリストアップし得ることを示しているからである。

なお、真横顔に対する識別精度を向上させるには、「顔を識別するアルゴリズム」をディープラーニングにより生成する「学習フェーズ」において、より多くの「真横顔と正面顔のセット」を教材として用いて反復学習することが効果的である。

(4) 同一人物の複数ショット検索による識別精度

被疑者写真データベースでは、累犯者には複数の被疑者写真が存在する。このため、防犯カメラの録画映像などに遺留された犯人の顔画像に基づき、被疑者写真データベースを検索する際には、累犯者の被疑者写真の扱い方について２通りの方法がある。つまり、最新の被疑者写真のみを検索対象とする方法と、全ての被疑者写真を検索対象とする方法である。どちらの方法が識別精度に優れるのか、興味深いところである。

下記の①～④は、2018年のFRVTにおいて、「同一人物の複数ショットを含まない顔画像データベースを検索した場合の本人見逃し率（TOP１）」と、「同一人物の複数ショットを含む顔画像データベースを検索した場合の本人見逃し率（TOP１）」について、参加ベンダーの中で最も優れていた上位４社の結果である。

ここでのTOP１における「他人誤認率」や「本人見逃し率」の求め方については、次のとおりである。

160万人分（160万枚）の顔画像データベース（全てMugshot画像）に対して、33万1,254人分（33万1,254枚）のMugshot画像（いずれもデータベース内に同一人物のMugshot画像が含まれていない。）を用いて33万1,254回の検索を行い、「他人誤認率」が1/1000となる「閾値」を求める。

次に、上記の160万人分（160万枚）の顔画像データベース（全てMugshot画像で、同一人物の複数ショットを含まない。）に対して、15万4,549人分（15万4,549枚）のMugshot画像（いずれもデータベース内に同一人物のMugshot画像が含まれている。）を用いて15万4,549回の検索を行い、「他人誤認率」が1/1000となる「閾値」を設けた場合のTOP１における「本人見逃し率」を求める。これが、下記の①～④における「同一人物の複数ショットを含まない顔画像データベースを検索した場合の本人見逃し率（TOP１）」である。

また、160万人分（335万1,206枚）の顔画像データベース（全てMugshot画像で、同一人物の複数ショットを全て含む。）を別に用意して、15万4,549人分（15万4,549枚）のMugshot画像（いずれもデータベース内に同一人物のMugshot画像が含まれている。）を用いて15万4,549回の検索を行い、「他人誤認率」が1/1000となる「閾値」を設けた場合のTOP１における「本人見逃し率」を求める。これが、下記の①～④における「同一人物の複数ショットを含む顔画像データベースを検索した場合の本人見逃し率（TOP１）」である。

なお、160万人分（335万1,206枚）の内訳についてであるが、１人１枚のみが160万人中の80.1％を占めており、１人２枚が13.4％、１人３枚が3.7％、１人４枚が1.4％、１人５枚が0.6％、１人６枚が0.3％、１人７枚以上で最大33枚が0.2％を、それぞれ占めている。

① 我が国のＮ社

同一人物の複数ショットを含まない顔画像データベースを検索した場合の本人見逃し率（TOP１）：0.0044

同一人物の複数ショットを含む顔画像データベースを検索した場合の本人見逃し率（TOP１）：0.0021

② 中国のＹ社

同一人物の複数ショットを含まない顔画像データベースを検索した場合の本人見逃し率（TOP１）：0.0123

同一人物の複数ショットを含む顔画像データベースを検索した場合の本人見逃し率（TOP１）：0.0074

③ 米国のＭ社

同一人物の複数ショットを含まない顔画像データベースを検索した場合の本人見逃し率（TOP１）：0.0141

同一人物の複数ショットを含む顔画像データベースを検索した場合の本人見逃し率（TOP１）：0.0080

④ 中国のＳ社

同一人物の複数ショットを含まない顔画像データベースを検索した場合の本人見逃し率（TOP１）：0.0234

同一人物の複数ショットを含む顔画像データベースを検索した場合の

本人見逃し率（TOP１）：0.0165

さて、160万人分（335万1,206枚）の顔画像データベースについては、160万人中の約32万人（全体の約２割）が、撮影期日の異なる複数枚のMugshot画像を有するため、160万人分（160万枚）の顔画像データベースと比べてその規模が著しく増加（160万枚から約335万枚に倍増）している。しかし、上記の①～④に示すとおり、160万人分（335万1,206枚）の顔画像データベースを検索した場合（つまり、同一人物の複数ショットを含む顔画像データベースを検索した場合）には、検索対象枚数の倍増にも関わらず、４社の全てにおいて識別精度（「他人誤認率」が1/1000となる「閾値」を設けた場合のTOP１における「本人見逃し率」）に顕著な改善が見られる。

このことから、１人当たり１枚限りの本人顔画像に対して検索するよりも、１人当たり何枚もの本人顔画像に対して検索する方が、本人見逃し率をより低減できると言える。

また、視点を変えて見れば、ビデオカメラのライブ映像などの中から目的とする人物を見つけ出そうとする場合に、１人当たり１枚限りの検索用顔画像を用いて顔画像データベースを検索（つまり、映像内の１ショットで検索）するよりも、１人当たり何枚もの検索用顔画像を用いて顔画像データベースを検索（つまり、映像内の複数ショットで検索）する方が、一定の閾値の下で他人誤認率を悪化させることなく、本人見逃し率を確実に低減できると言える。

おわりに

2013年の時点では、真横顔画像を正面顔画像と照合して同一人物であるか否かを見分けられる顔画像識別技術はどこにも存在せず、顔の正面から見て30度から45度ほどの斜め横方向から撮影した顔画像が、正面顔画像との照合により同一人物であるか否かを見分けられる限界であった。しかし、

2018年の時点では、識別精度が高いとはまだ言えないが、真横顔画像を正面顔画像と照合して同一人物であるか否かを見分けられる顔画像識別技術が出現している。

このように識別性能が飛躍的に向上した主因は、顔画像識別技術にAI（CNNのディープラーニング）を活用したことにある。それゆえ、真横顔画像と正面顔画像との照合時の識別精度を更に向上させるには、「顔を識別するアルゴリズム」をディープラーニングにより生成する「学習フェーズ」において、より多くの「真横顔と正面顔のセット」を教材として用いて、効果的かつ効率的な反復学習を行うことに尽きるのである。

ところで、米国では、監視カメラによる指名手配犯の発見といった犯罪捜査に顔画像識別技術を用いた場合における人種バイアス、つまり、白人の顔に比べてアフリカ系やアジア系の有色人種の顔に対する他人誤認率がかなり高いのではないかということが、人種的偏見の助長に繋がりかねないとして大きな問題となっている。前記１の(1)に記載のとおり、最先端の顔画像識別技術に用いられているAI（CNNのディープラーニング）は、人の頭脳内部での神経回路網の仕組みと働きを、「入力層―隠れ層―出力層」としてコンピュータ上で数学的に模したものである。このため、上記の人種バイアスの原因や対策を検討する上で、「人の目」における人種バイアスについて検討することが有益であると考えられる。そこで、筆者が『警察政策第17巻（2015）』に投稿した論説「顔画像識別における人の目の特性と機械の目の特性」の中から、「人の目」における人種バイアスについて言及した一文を次に引用したい。

『人は、相手の顔を見て美醜などの特徴を瞬時に感じ取っている（見分けている）。この見分ける基となっているものは、誕生以来、見たり接したりしてきた無数の顔から生み出された「頭の中の平均顔」と考えられる。見たり接したりしてきた顔の大半が日本人であれば、日本人の平均顔が頭の中に生み出されるであろう。日本人には、白人の顔、あるいは黒人の顔が、日本人の顔ほどには見分けが付けづらいが、日本人の平均顔を基に顔の特徴を見分けているとすれば、白人の顔、あるいは黒人の顔が、日本人

には皆同じように見えてしまうのも無理はない。』[3)]

このことから類推すれば、大半を日本人の顔で学習した顔画像識別技術では、白人や黒人の顔に対する識別精度が日本人の顔ほどには期待できないのは道理である。要するに、米国で犯罪捜査に顔画像識別技術を用いた場合における上記の人種バイアスの問題は、白人の顔に比べてアフリカ系やアジア系の有色人種の顔に対する学習不足が最も大きな原因であると考えられる。

この問題を技術的な観点から深掘りすれば、ディープラーニングにより「顔を識別するアルゴリズム」を生成する「学習フェーズ」で用いた教材用顔画像の数量について、人種によりかなりの多寡があれば、そのまま人種による他人誤認率の高低に繋がるということである。つまり、米国における多数派である白人の顔画像が、少数派であるアフリカ系アメリカ人やアジア系アメリカ人の顔画像よりも、学習用教材として格段に多く用いられた場合に、このようなディープラーニングで生成した「顔を識別するアルゴリズム」を「推論フェーズ」で用いたときに、本人と他人とを区別する唯一の手掛かりとなる「類似度を示すスコア値の差」が、白人の顔ほどにはアフリカ系アメリカ人やアジア系アメリカ人の顔では開かない（差があまり付かない）のである。これでは、他人誤認率が例えば1/1000となる閾値（つまり、1000回の顔画像照合を行った場合に１回の誤認が発生する閾値）を白人の顔画像を基本に設定したとしても、アフリカ系アメリカ人やアジア系アメリカ人の顔画像に対しては、このような閾値では他人誤認率がかなり悪化する。要するに、アフリカ系アメリカ人やアジア系アメリカ人の顔画像に対する識別性能が良くないのである。

この問題への対策は、「顔を識別するアルゴリズム」をディープラーニングにより生成する「学習フェーズ」における学習用教材として、アフリカ系アメリカ人やアジア系アメリカ人の顔画像を更に多く用いて、白人の顔画像と同等の識別精度が達成できるまで、効果的かつ効率的な反復学習を行うことに尽きる。つまり、先に述べた真横顔画像と正面顔画像との照合時の識別精度向上方策と同じく、AI（CNN のディープラーニング）を

活用した顔画像識別技術における識別精度の向上には、反復学習の充実強化が最も重要なファクターとなるのである。

引用文献

1）NIST「Face Recognition Vendor Test」（NIST Interagency Report 8009（2014年））
2）NIST「Face Recognition Vendor Test」（NIST Interagency Report 8271（2019年））
3）澤田雅之「顔画像識別における人の目の特性と機械の目の特性」（『警察政策第17巻』188頁（2015年））

参考文献

澤田雅之「警察情報通信の発注者エンジニアリング～ターゲット発見システムの実現に向けて」（『警察政策第19巻』245頁（2017年））
澤田雅之「顔画像識別技術と監視カメラが産み出す機械の目の特性」（月刊技術士３月号12頁（2016年））
堀内雄人・羽田拓朗「顔画像自動識別技術の大規模データベースに対する適用に向けて」（『警察政策第16巻』163頁（2014年））
堀内雄人「顔画像自動識別技術の動向」（『警察政策第14巻』67頁（2012年））
NIST「Face In Video Evaluation ～ Face Recognition of Non-Cooperative Subjects」（NIST Interagency Report 8173（2017年））
NIST「The IJB-A Face Identification Challenge ～ Performance Report」（2017年）
法務省「日本人出帰国審査における顔認証技術に係る実証実験結果報告」（2014年）

（さわだ　まさゆき）