課題は残るも、着実に進歩する顔認識技術

ジェフ・ヘクト

最高レベルのラボシステムは、人間とほぼ同等レベルの顔認識能力を備えるまでに進歩している。しかしその能力は、セキュリティや警察業務といった要件の厳しい分野への適用や、ソーシャルメディアでの広範な使用に耐えるだけの水準に達しているのだろうか。

私たちは、目にしたものを解釈して人間の顔を認識するように仕組まれた脳を生まれながらにして備えている。機械による顔認識は、光センサや赤外線センサによって収集されたデータに対するコンピュータ解析を拠り所とする。しかし、半世紀にもわたって取り組みが続けられているにもかかわらず、人間と同じように顔を認識する方法をいかにしてコンピュータに教えるかという問題は、未だ完全には解明されていない。

顔認識は一見、パターン認識というシンプルな問題のように見えるかもしれないが、実際には顔は、コンピュータ認識に対して特に複雑な問題を呈する。顔は三次元であるため、二次元画像や検出器に現れる外観は、視野角、照明角度、照明の性質といった要因に依存する。また顔は、人間の表情によって変化する。笑っているとき、顔をしかめたとき、舌を出したときでは見え方が異なる。髪型、眼鏡、化粧、体重によっても外観は異なって見える。

それでも顔を簡単に確認できることから、顔認識は生体認証において非常に魅力的な手段となっている。米連邦捜査局（FBI：Federal Bureau of Investigation）や米国土安全保障省（DHS：Department of Homeland Security）は、警察業務やセキュリティを目的とした顔認識システムを開発している。民間企業も、民間セキュリティシステムや、ソーシャルネットワーク上の写真のラベル付けを対象とした顔認識に取り組んでいる。Android搭載端末や米アップル社（Apple）製スマートフォンには、起動してカメラに顔を向けることによって端末のロックを解除できるアプリケーションが提供されており、紛失や盗難時に端末をロックできるようになっている。

公的用途と民間用途

警察機関は何年も前に、標準的なフォーマットで写真を撮影すれば顔が認識しやすいことに気が付いた。正面と横向きの顔写真が昔から変わらず警察で使用されるのは、そのためである。政府機関は、正式な本人確認用写真に対する厳格な標準フォーマットを定めており、眼鏡を外すこと、自然な表情をとること、カメラを直視することを求めている。眼鏡を外したことでカメラに焦点を合わせられなければ、指示に従っていないとみなされる可能性もあるが、このような標準的な写真が本人確認用に最適であると考えられている。

FBIが10億ドルを投じて開発したNGI（Next Generation Identification：次世代認証）プログラムの顔認識に使用される写真リポジトリに含まれるのは、すべて標準フォーマットの犯罪者顔写真である。2012年2月に開始されたパイロットプログラムでは、警察機関によって提供された、およそ1300万枚もの正面向きの犯罪者顔写真からなるリポジトリが使用されている。同システムは、提示された写真をリポジトリの中の写真と電子的に比較し、コンピュータによって順位付けされた一致候補のリストを生成する。ここでの課題は、監視カメラやソーシャルメディアといった多様な情報源から収集された、標準フォーマットではない写真に対して、一致する候補を検索することである。順位付けは、警察が捜査で参考にするためのものであり、本人であることを証明するものではない。

このパイロットプログラムの目標は、「顔認識プロセスを試験し、方針や処理の問題を解決し、プライバシー保護対策を確立し、ユーザの懸念に対処すること」だと、FBI高官のジェローム・ペンダー氏（Jerome Pender）は2012年、米上院司法小委員会に対して語っている[1]。提示された写真は、写真リポジトリには追加されなかった。その写真の中の人物は、はっきりと身元が確認されたわけではないためである。この慣行は、同システムがより大規模なデータベースを装備して稼働を開始する、2014年夏の終わりまで継続される予定である。

DHSのBOSS（Biometric Optical Surveillance System：生体光学監視システム）は、群衆を録画したビデオの中に映る人々をコンピュータで本人確

認するといった、さらに野心的な目標を掲げている[2]。しかし、確実な確認はきわめて困難な作業であり、群衆が写る写真は、視野角や照明を超える難題をもたらす。カメラは一般的に、群衆からかなり離れた場所に設置されており、個々の顔は低い解像度で録画される。また、人々は帽子やフードをかぶっていたり、サングラスをかけていたりする場合もある。

民間の顔認識アプリケーションは一般的に要件が異なり、通常は写真ではなく人間を確認することが求められる。アクセス制御用の生体認証システムは、限られた数の人物を正確に認識することによって、それ以外のユーザにはアクセスを許可しないようにする。スマートフォンのロック解除では、1人の人物を高い信頼性で確認することが求められる。米マイクロソフト社（Microsoft）のKinectは、モーションセンサによってユーザの顔の3Dプロファイルを作製して、ゲームのプレーヤーを識別するため、正確さが極めて重要というわけではない。

ソーシャルメディア上の顔認識は、人間ではなく写真を対象とする必要があるが、Facebookの友達ステータスといった画像以外のデータを利用して、写真に写っている可能性の高い人物の数を限定することができる。誤りがあっても、通常は重大な問題とはみなされないが、誤認識が決まりの悪い事態につながることはあり得る。

多様な技術

コンピュータによる二次元の顔認識ではまず、写真の中から顔を検出することが行われる。続いて**図1**に示すように、鼻の頭、目や口の端、あごの先といった主要な形状特徴の位置を特定してデジタル化する必要がある。この

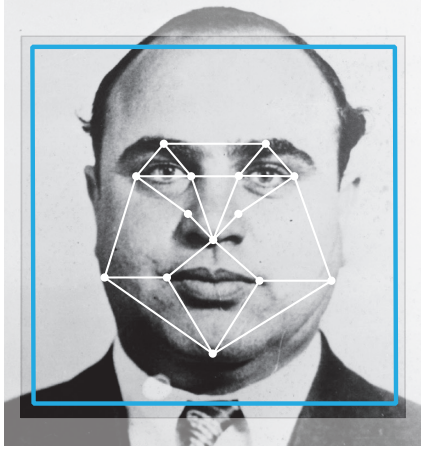


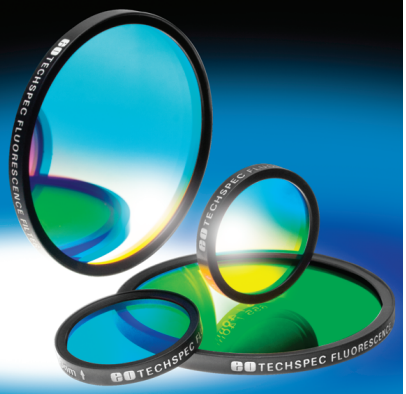
図1　顔の主要ポイントが示されたアル・カポネ（Al Capone）の古い顔写真（出典：米司法省およびウィキメディア[Wikimedia]）。顔認識は、目と口の両端、鼻と口の先端、あごと眉毛上の点といった主要ポイントの識別に基づいている。これらの識別ポイントの中には、広く使用されるものもあれば、特定のアルゴリズムにおいてのみ使用されるものもある。3Dセンサを利用すれば、顔に対する同様のマッピングを三次元で行うことができる。

情報を解析することにより、コンピュータは写真が正面から撮影されたものであるかどうかを判断する。そうでない場合は、視野角を推定して、対象とする顔のデジタル表現においてその角度を補正する必要がある。その後、顔の形状特徴のデジタル化された位置を、犯罪者顔写真のように制御された条件下で撮影された身元確認済みの顔写真ライブラリ内の写真と比較する。

安価な光学式三次元スキャナの進歩によって、その場にいる人物の顔の3D形状に関するデータの直接収集が実用化され、従来の3D画像処理を補うか、それに代わる手段となっている[3]。この追加情報によって、一部の用途における二次元顔認識の限界を克服することができる。また、セキュリティの確保や電子機器のロック解除といったアプリケーションにおいて、権限を持つ人物の写真を提示することによって犯罪者がアクセスを得ようと試みるケースに対して、権限を持つ人物が実際

2,950品目の
在庫用フィルター
只今販売中

設計から試作、そして
量産まで



エッジフィルター
OD >6.0, カットオン/
オフスロープ <1%までの
標準規格

バンドパスフィルター
バンド幅別に多数。
OD >6.0まで、193~1650nmの
波長域内で高透過率

蛍光用フィルター
顕微鏡メーカーのフィルターキューブに装着したプリマウント型と、装着しないフィルターセット。市販の蛍光試薬を用いた像の明るさとコントラストを最大化

ご注文は今すぐ！

www.edmundoptics.jp/
filters

more optics | more technology | more service

Edmund optics | japan

エドモンド・オプティクス・ジャパン株式会社
〒112-0001 東京都文京区白山5-36-9 白山麻の実ビル
TEL：03-5800-4751　FAX：03-5800-4733 | www.edmundoptics.jp

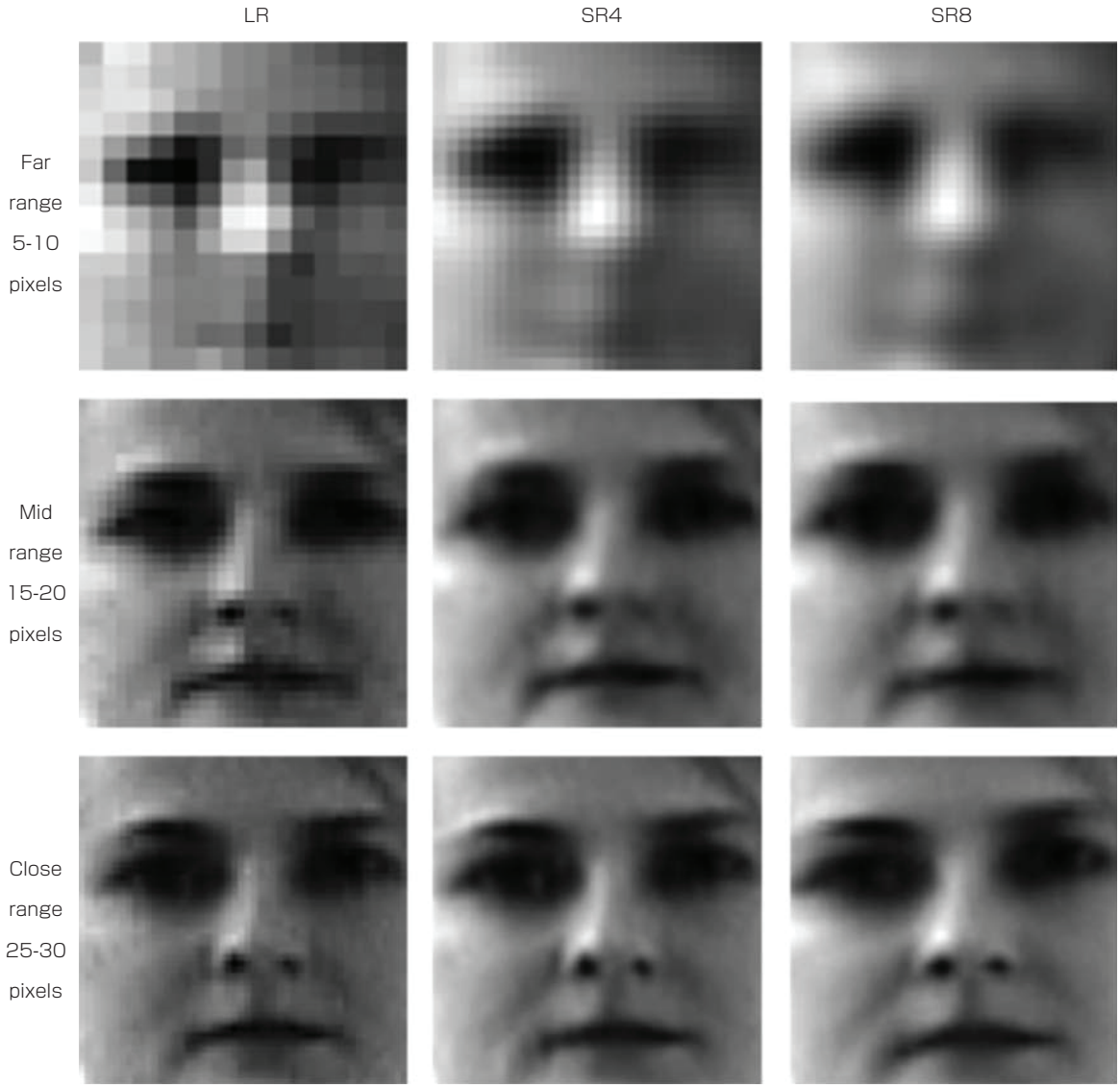


図2 複数の低解像度フレームを組み合わせることによって、見た目がはっきりとした「超解像度」画像が得られる。解像度が低いほど、その効果は顕著である。各画像の上部の数字は、その画像を得るために組み合わせたフレーム数を示す（資料提供：米陸軍研究所[7]）。

にそこに居ることを確認することができる。顔認識の性能を高めるための手段として、3Dステレオ画像処理技術を取り入れる開発も行われている[4]。

コンピュータ処理技術は、パフォーマンスの観点から不可欠である。顔認識用にニューロネットワークなどの自己学習コンピュータシステムを開発しようという取り組みは、1990年代にさかのぼる[5]。強力なコンピュータによって、標準的な顔の特徴化よりも有効なパターンを認識できるのではないかという期待が寄せられていたが、成功にはいまだ至っていない。米フェイスブック社の人工知能研究所（Facebook AI Research）のヤニフ・タイグマン氏（Yaniv Taigman）とそのグループは現在、同氏らが開発した「DeepFace」プログラムによって、「ディープラーニング」と呼ばれる一連のアルゴリズムを用いて、身元確認済みだが標準的な条件下で撮影されてはいない400万枚の写真を対象とする顔認識を行ったと主張している[6]。

同氏らは、明示的な3Dモデリングを用いて、写真の中の顔の向きを調整した。これに、データベース内の写真を解析するディープニューラルネットワークアーキテクチャを組み合わせて、他の写真集合にも適用可能な独自の顔表現手法を開発した。顔の向きを調整しない場合、「一般収集した」（つまり、意図的な調整が行われていない）テスト写真集合に対する識別精度は87.9%だった。二次元の調整を加えると精度は94.3%に上がり、三次元調整ではさらに97.35%にまで高まって、人間の識別率である97.5%に近い精度が得られたという。同システムの処理時間は

1画像あたり約0.33秒である。正確さに対してセキュリティシステムほどの厳しい条件は課されないソーシャルネットワーク上での人物特定に向けて、一歩前進したといえるだろう。

新しい光学技術

新しい光学技術によっても顔認識精度が向上する可能性がある。監視カメラは、多くの犯罪者の割り出しに貢献してきたが、シングルフレームの画質は通常、センサ解像度や視野角、遠くに映っているために顔のサイズが小さいといった要因によって制約を受ける。顔画像内のピクセル数（一般的には両目の間のピクセル数で測定される）が少ないほど、顔認識精度は低くなる。しかし、同じ姿勢の一連の低解像度顔画像を組み合わせることによって、どのシングルフレームよりも解像度の高い超解像度画像を再現することができる。米陸軍研究所（Army Research Laboratory）における試験からは、超解像度画像処理によって解像度が大幅に向上し、顔認識精度も高まることが示された[7]。一連の4フレームおよび8フレームに対するこの様子を**図2**に示す。

赤外線観測は、肌の色や、照明が不適切であったり不均一であったりすることに起因する、視覚的な顔認識の問題を回避する手段となる可能性がある。長波長の赤外線センサは、反射光ではなく、人体から10マイクロメートルの範囲の熱放射を記録することができる。初期の研究からは、単純な光度に基づく画像は、正確な顔認識に必要なコントラストと空間分解能に欠けている場合が多いことが明らかになっている。米陸軍研究所における最近の実験では、赤外光の極性を記録することによって、顔の解像度が高まることが示されている[8]。

カナダのラヴァル大学（Laval University）のレザ・ギアス氏（Reza Ghiass）とそのグループは最近の調査の中で、熱赤外線顔認識は、眼鏡などの障害物や体温変化による制約を受ける可能性があると警告している。しかし同氏らは、熱画像は新しい識別基準をもたらす可能性もあるとしている。例えば血管網など、可視情報では得られないデータを視覚化することができる[9]。

今後の展望

顔認識技術に課される要件の進化に伴い、開発者らは、アプリケーションによって性能要件がどのように異なるかを慎重に検討する必要がある。アクセス制御においては、誤認識は大問題である。ひげをそり忘れたとか、アレルギーで目が腫れているといった理由で、職場や個人所有の電子機器でシステムからの締め出しを食らうというのでは困る。警察機関においても、誤認識は問題となる可能性がある。高いセキュリティが求められる金融システムの場合は、なりすましや侵入が可能であるとすれば危険だ。ソーシャルメディアでは、不名誉な写真の中で集団訴訟の弁護士を誤認するといった問題が発生しない限り、被害はないかもしれない。

顔認識技術の台頭には、米電子フロンティア財団（Electronic Frontier Foundation）といった団体から、プライバシー侵害に関する懸念も寄せられている。プライバシー擁護団体は、ハイテクを駆使したセキュリティ体制が個人の自由を侵害し、ストーカー行為や嫌がらせから、過度に強引なマーケティングにいたるまでの濫用を招くのではないかと不安視している。米グーグル社（Google）が、Google Glassに対する顔認識アプリを承認しないと決断したことは、このような懸念の緩和につながったが、非公式の顔認識アプリが、ジェイルブレイクされたGoogle Glassに対して提供されている。

問題は複雑であり、技術はまだ発展途上にある。2013年に起きたボストンマラソンの爆破事件の容疑者を特定したのは、コンピュータではなく人間の目であったことを覚えておきたい。

参考文献

(1) J.M. Pender, "What facial recognition technology means for privacy and civil liberties," http://www.fbi.gov/news/testimony/what-facial-recognition-technology-means-for-privacy-and-civil-liberties.

(2) C. Savage, "Facial scanning is making gains in surveillance," New York Times, Aug. 21, 2013, http://www.nytimes.com/2013/08/21/us/facial-scanning-is-making-gains-in-surveillance.html.

(3) L. Zhang et al., "3D face recognition based on multiple keypoint descriptors and sparse representation," PLoS ONE V. 9, e100120, (June 2014); doi:10.1371/journal.pone.0100120.

(4) Y. Zheng and E. Blasch, "The advantages of stereo vision in a face recognition system," Proc. SPIE 9091, 90910Y; doi: 10.1117/12.2049985.

(5) S. Lawrence et al., "Face recognition: A convolutional neural-network approach," IEEE Trans. Neural Netw. 8, 98 (Jan. 1997).

(6) Y. Taigman, "DeepFace: Closing the gap to human-level performance in face verification," https://www.facebook.com/publications/546316888800776/.

(7) S. Hu et al., "Face recognition performance with superresolution," Appl. Opt. 51, 4250 (June 20, 2012).

(8) K. Gurton et al., "Advanced facial recognition for thermal imagery using polarimetric imaging," Opt. Lett. 39, 3857?3859 (July 1, 2014); doi.org/10.1364/OL.39.003857.

(9) R. S. Ghiass et al., "Infrared face recognition: a comprehensive review of methodologies and databases," http://arxiv.org/abs/1401.8261.

LFWJ