AI 画像認識による服装・所持品推定技術の実用化に向けた取り組み

Efforts toward practical application of clothing and belongings estimation technology according to the AI image recognition

髙橋　光市　Takahashi Koichi
高山　恒一　Takayama Koichi
佐藤　健　Sato Ken
田村　栞里　Tamura Shiori

AIを用いた画像認識は社会のあらゆる分野に広がっている。一方でこの技術の活用，特に人物が画像に写り込む可能性がある場合にはプライバシーへの配慮が欠かせない。それは，その人物の顔の画像が個人情報だからである。私達はAI画像認識の有用性とプライバシーへの配慮などの課題にうまく折り合いをつける技術を考案した。これにより画像認識の応用分野は格段に広がる可能性がある。本稿ではこの技術の特長と活用シーンの展望ならびに試作モデルのテスト結果を説明する。

1. はじめに

画像認識技術はディープラーニングをはじめとするAIの発展とともに進化してきた。その応用は医療や自動運転，製造現場での品質管理，危険区域における安全確保など社会のあらゆる分野に広がっている。一方でこの技術の活用にはプライバシーへの配慮が欠かせない。例えば，カメラ画像に写り込んだ人物の顔は個人情報であり，その人物の権利を損なうことのないよう十分な配慮が求められるからである。

本稿で紹介する技術は画像認識の応用分野を格段に広げる可能性がある。なぜならこの技術が画像認識の有用性とプライバシーへの配慮などの課題にうまく折り合いをつけることができるからである。

これは(株)日立ソリューションズ東日本(以下, HSE)の研究プロジェクトの成果である。研究の目的は，お客様や地域・社会に貢献するためHSEのAI技術力を高めること。具体的なテーマは，カメラで撮影した画像から人物の服装や所持品などを推定することである。これらの情報をエンドユーザに提供することで社会貢献ができると考えた。

例えば人物が着用している服装，具体的にはその種類や色を推定し世の中の流行を推測することで，その情報を販売予測に活用できる可能性がある。また銀行などに設置されているATMの前にいる人物が，スマートフォン（以下，スマホと略す）を使って電話をしている場合，特殊詐欺の誘導によって取引をしている懸念がある。そのようなATM操作者の行動を把握できれば特殊詐欺を未然に防ぐこともできる。

以下，私達が考案した技術と活用シーンならびに今回試作したモデルのテスト結果を説明する。

2. 本技術の特長

先に述べた通り，画像認識を行う際には人物画像に対するプライバシーへの配慮が課題となる。また設置するカメラの数や画像認識処理の頻度によってはコンピュータの処理能力が課題となる。これらの課題に対し独自の処理方式で解決を図るのが本技術である。

本技術は図 1 に示すようにカメラ内蔵コンピュータと分析サーバに処理を分散することで画像認識を実現する。

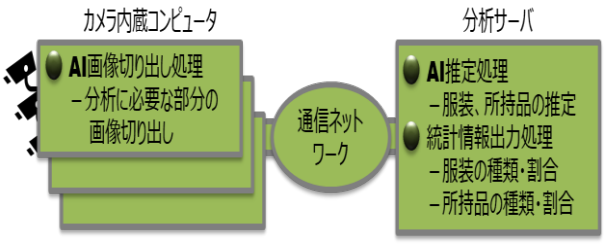


図 1　本技術のシステム構成

まず処理の概略について述べる。カメラ内蔵コンピュータ側では, カメラで撮影した画像を入力データとして, 分析に必要な部分の画像を切り出し，それを出力データとして分析サーバに送る。分析サーバ側では，送られてきたデータを入力データとして服装と所持品などを推定し，服装の種類と割合，所持品などの種類と割合を出力する。

次にカメラ内蔵コンピュータ，分析サーバそれぞれの処理を詳しく述べる。

図 2 はカメラ内蔵コンピュータの処理イメージである。図 2 の左側にあるのが撮影対象の人物である。カメラ内蔵コンピュータは，まずこの人物を含んだ画像から人物の顔の位置を推定する。それから図 2 の右側にあるように人物の顔を除いた部分（網掛けした部分）を出力データとする。その際，人物の顔の画像とともに分析に不要な風景などを除去する。また所持品の推定を行う場合は，人物周辺にある物品の画像を切り出して出力データに加える（図では省略）。

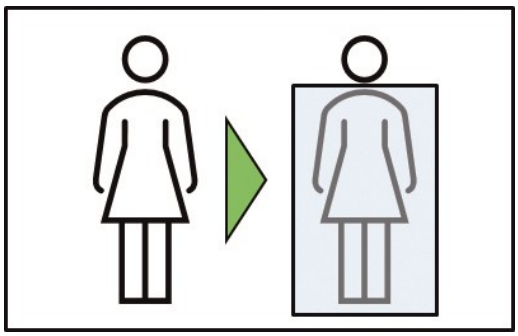


図 2　カメラ内蔵コンピュータの処理

図 3 は分析サーバの処理のイメージである。分析サーバでは，カメラ内蔵コンピュータの出力データをもとに服装と所持品などの種類を推定，またそれらの割合を算出する。

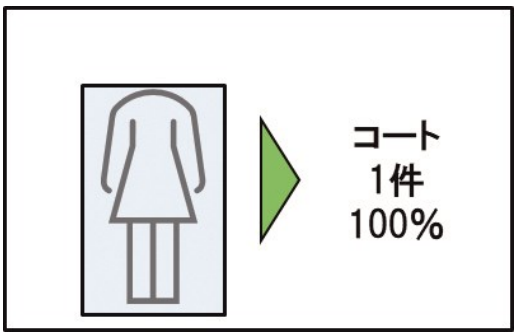


図 3　分析サーバの処理

本技術は以下の特長を持つ。

(1) 取り扱う画像データをコンパクトにできる

カメラで撮影した画像から分析に必要な部分だけを切り出すことで，分析サーバの性能要件を緩和できる。具体的には，より低スペックのコンピュータで分析処理ができる。また同じスペックであればより精度の高い分析やリアルタイム性の高い処理が可能となる。この画像データの削減効果を検証したところ，今回試作したモデルでは，出力データの容量を入力データの 29%に削減できた（テストデータ 12 件による平均値）。

(2) 顔を含む画像の収集を抑制する

分析サーバ側では，顔を含む画像を極力収集しないことでプライバシーに配慮したシステムを実現できる。またカメラ内蔵コンピュータ側は，出力処理のあと入力画像を破棄することでプライバシーへの配慮をより高めることができる。

近年，人物画像の取り扱いは難しくなっている。個人情報保護法では，個人情報を取得・利用する際に通知や公表を求めている。一方で，撮影対象となる人物の理解を得るのは容易ではなく，それがビジネス上の障壁となり得る。本技術は顔画像を含む個人情報の取り扱いをゼロにできるものではない。しかし個人情報の収集を極力抑えることは，撮影対象となる多くの人物の安心感につながる。それによって情報取得・利用に対する理解が得やすくなり，これまで実現が難しかった活用シーンにも対応できる。なお，顔画像除去のための顔認識の検証結果については「試作モデルによる画像認識のテスト」で述べる。

(3) 複数の人物が重なった画像も処理できる

カメラで撮影した画像に複数の人物が重なって写った場合，前景側の顔画像を取り除くと背景側の人物の画像の一部が失われてしまう場合がある。これは例えば図 4 の左側にあるように前景側の人物の顔から下の部分を切り出すと，背景側の人物の上半身の画像が失われることになるからである（この場合，背景側の人物の下半身の画像だけが取得できる)。そこで画像の顔の部分にモザイク化処理を行う。こうすることでプライバシーを保護するとともに図 4 の右側にあるように背景側の人物全体の画像を取得することができる。

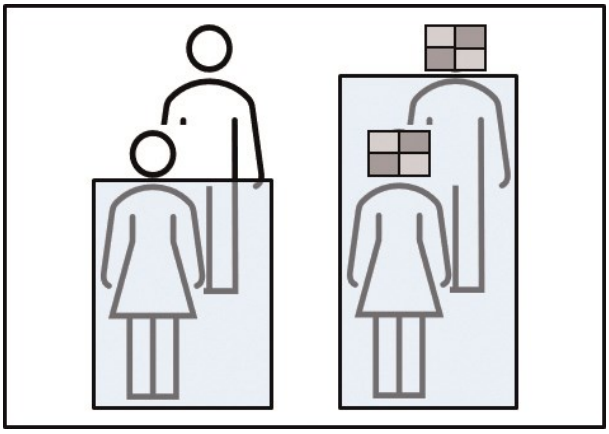


図 4　複数の人物が重なった画像の処理

3. 活用シーンの展望

本技術は，人物画像に対するプライバシーへの配慮が強く求められるケース，また比較的，設置するカメラの数が多い場合や画像認識処理の頻度が高いケースで効果を発揮する。以下，その観点から想定した活用シーンについて展望を述べる。

(1) 服装の流行に関する情報の提供(小売／消費者向け)

図 5 に活用シーンのイメージを掲載する。網掛けの箇所は画像切り出しを行う部分である（以下，共通）。

【活用シーン】

街中で流行する服装のデザインに関する情報を小売業者や消費者向けに提供する画像認識システム

【提供価値】

街中での，人物の服装（トップス，ボトムス，靴など）の種類と色，柄などの割合を推定することにより，流行のデザインに関する情報を小売業者または消費者に提供できる。秋／冬物であれば北の地方から南の地方へ，春／夏物であれば南の地方から北の地方へ，というように情報を先取りして衣類を準備することができる。

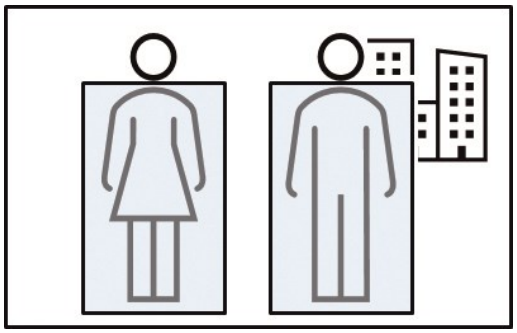


図 5　服装の流行に関する情報の提供(小売/消費者向け)

(2)　行方不明者の捜索支援

図 6 に活用シーンのイメージを掲載する。

【活用シーン】

迷子の子供や認知症罹患者などの行方不明者の捜索を支援するための画像認識システム

【提供価値】

対象となる人物が着用している服装の特徴(種類や色)をもとに，その人物の居場所を特定する。服装の特徴は例えば「水色のブラウスと黄色いスカートを着用」または「緑のジャンバーと薄茶色のズボンを着用」のように指定する。大型商業施設や街中のいたるところにカメラ内蔵コンピュータを設置することにより，短時間で居場所を特定できる可能性がある。人物特定の精度を高めるために，服装とともに，おおよその身長など他の特徴を組み合わせることも有効だと考える。

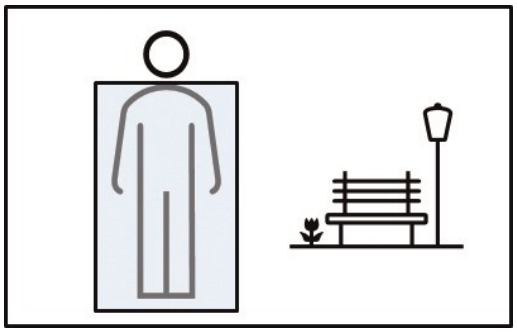


図 6　行方不明者の捜索支援

(3)　ATM でのスマホ利用の監視支援

図 7 に活用シーンのイメージを掲載する。

【活用シーン】

ATM の前にいる人物がスマホを利用している場合，特殊詐欺の誘導によって ATM を操作している可能性がある。そのようなケースによる振り込みなどの取引を未然に防ぐための監視支援システム

【提供価値】

ATM の前にいる人物の，スマホや携帯電話利用の情報を取得することで特殊詐欺に対する安全性をより高めることができる。システムが直接一般利用者に注意喚起することもできるが，システムによる取引への過度な介入を避けるため，情報を銀行やコンビニエンスストアなどの担当者に通知することで，必要に応じ利用者のフォローに活用することも可能である。この活用シーンでは，スマホ・携帯電話の利用を推測するとともに，音声認識や言語理解の技術を用いることで，会話の内容を分析し不適切取引を防止する方法も有効だと考える。

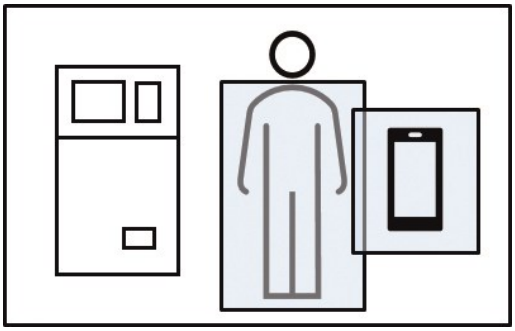


図 7　ATM でのスマホ利用の監視支援

以上，活用シーンについての展望を述べた。これらのシステムは，顔を含む画像を取り扱う一般的な手法によって構築することも技術的には可能である。しかしそれを実社会に適用するには，これまで述べたようにプライバシーへの配慮と性能面での考慮が必要である。それに対し本技術は，この二つの観点で優れ，現代社会に適合した実現性の高いシステムを提供できる。

4. 試作モデルによる画像認識のテスト

本技術を使った試作モデルによる画像認識のテストについて説明する。これは前述した「ATM でのスマホ利用の監視支援」を想定したモデルである。このモデルは，顔画像のモザイク化処理と切り出し処理のため，またスマホ利用の有無を推測するため，入力した画像から顔とスマホを認識する。

5. 試作モデルとテスト内容

表 1 の通り二つのモデルを試作した。このモデル構築には，一般に公開されている dlib[1]，yoloV3-tf2[2]というフレームワークとその学習済モデルを使用した。これらのフレームワークは顔とスマホの画像認識を標準でサポートしている。一般的に機械学習モデルを構築するには，学習データの準備とモデルのリファイン作業が必要であり，それには膨大な時間と費用がかかる。そのモデルの構築を，このようなフレームワークを用いることで比較的容易に実現できる。

表 1　試作モデル

モデル1	フレームワーク／学習済モデル
顔認識	dlib（SVM版） ／get_frontal_face_detector
スマホ認識	yoloV3-tf2 ／yolov3.weights
モデル2	フレームワーク／学習済モデル
顔認識	dlib（CNN版） ／mmod_human_face_detector.dat
スマホ認識	（モデル1と同じ）

これら二つのモデルの違いは，顔認識に dlib の SVM 版を使用しているか，CNN 版を使用しているかである。dlib の SVM 版はサポートベクターマシンを用いた機械学習のモデル，CNN 版は畳み込みニューラルネットワークを用いたディープラーニングのモデルである。この二つのモデルを構築した理由は顔認識の精度の違いを確認するためである。一般的にディープラーニングのモデルは，従来からある機械学習のモデルに比べ認識精度が高く，その代わり処理能力の高いコンピュータを必要とする。

テスト用の入力データはカメラ内蔵コンピュータを使用して撮影した。具体的には，人物の前方から，その人物がスマホを手持ちまたは通話した状態で，撮影する角度を変えながら複数枚の画像を取得した。用意したテストデータはモデル 1，モデル 2 それぞれ 24 枚である。

5.1 テスト結果

各モデルが顔／スマホを認識できたことを示す正解率は表 2 の通りとなった。

表 2 顔とスマホ認識の正解率

モデル1	顔認識	スマホ認識
手持ち	92%	100%
通話	83%	33%
全体	88%	67%
モデル2	顔認識	スマホ認識
手持ち	100%	83%
通話	100%	50%
全体	100%	67%

5.2 考察

まず顔認識について述べる。モデル 1 で顔認識ができなかったのはテストデータ 24 枚中 3 枚であり，手で顔の一部が隠れるケース（2 枚），横向きの顔（1 枚）である。これに対しモデル 2 では 100%顔認識ができている。先にも述べた通り，ディープラーニングを用いたモデル 2 は，機械学習を用いたモデル 1 より，カメラ内蔵コンピュータの処理装置やメモリなどに負荷がかかる。従って，求める顔認識の精度やリアルタイム性に応じてモデルを使い分けるのが良いと考える。

次にスマホ認識についてであるが，全体の正解率は 67%であり，約 1/3 は認識できていないことになる。これについては二つの対策が考えられる。

一つ目の対策は，一定時間ごとに画像を撮影し，複数枚の画像を入力データとすることである。一般に物体の画像認識はその見え方に大きく左右される。人物の動きにより撮影角度の変わった複数の入力データを用いることで正解率の向上が期待できる。

二つ目の対策は，スマホ認識に複数の手法を用いることである。例えば一般に公開されている Media Pipe[3]という，画像に含まれる手の位置情報（座標）を取得できるフレームワークがある。手の位置が分かれば，試作モデルで取得できる顔の位置と比較して「通話の可能性あり」と判断できる。このフレームワークを使って手の位置情報を取得できるかを試したところ，試作モデルのテストで通話時のスマホ認識が NG となった 14 枚の画像のうち，7 枚の画像で取得することができた。これを試作モデルと組み合わせると「通話の可能性あり」と判断できるのは全体で 81%となる。この値は，一つ目の対策として述べた複数の入力データを用いることで，さらに向上すると考える。

6. おわりに

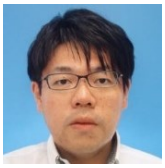
本稿では考案した技術の特長，活用シーンの展望，試作モデルのテスト結果を説明した。今回説明した試作モデルの他に，行方不明者の捜索支援を行うモデルも現在開発を進めている。今後はこれらのモデルの実業務を想定した検証と性能向上を図り実用化を目指す。本技術は特許出願中である（画像認識システムおよびプログラムセット，特願 2022-124221）。

参考文献

1) dlib (2022/11/9 閲覧)
http://dlib.net/
2) yoloV3-tf2 (2022/11/9 閲覧)
https://github.com/zzh8829/yolov3-tf2
3) Media Pipe (2022/11/9 閲覧)
https://google.github.io/mediapipe/

高橋　光市　　1989 年入社
RPA・AI 推進センタ
新事業企画・導入推進

高山　恒一　　2022 年入社
社会基盤ソリューション第三本部
ソリューション第二部
スーパーコンピュータ向け並列アプリケーション開発，数値解析



佐藤　健　　2010 年入社
社会基盤ソリューション第三本部
ソリューション第二部
量子コンピュータ関連システムの設計・開発，データ分析

田村　栞里　　2016 年入社
社会基盤ソリューション第三本部
ソリューション第二部
AI 画像認識