

齊藤弘紀* 清水尚吾**
若林 慧*
草野勝大**

監視システムでの映像解析

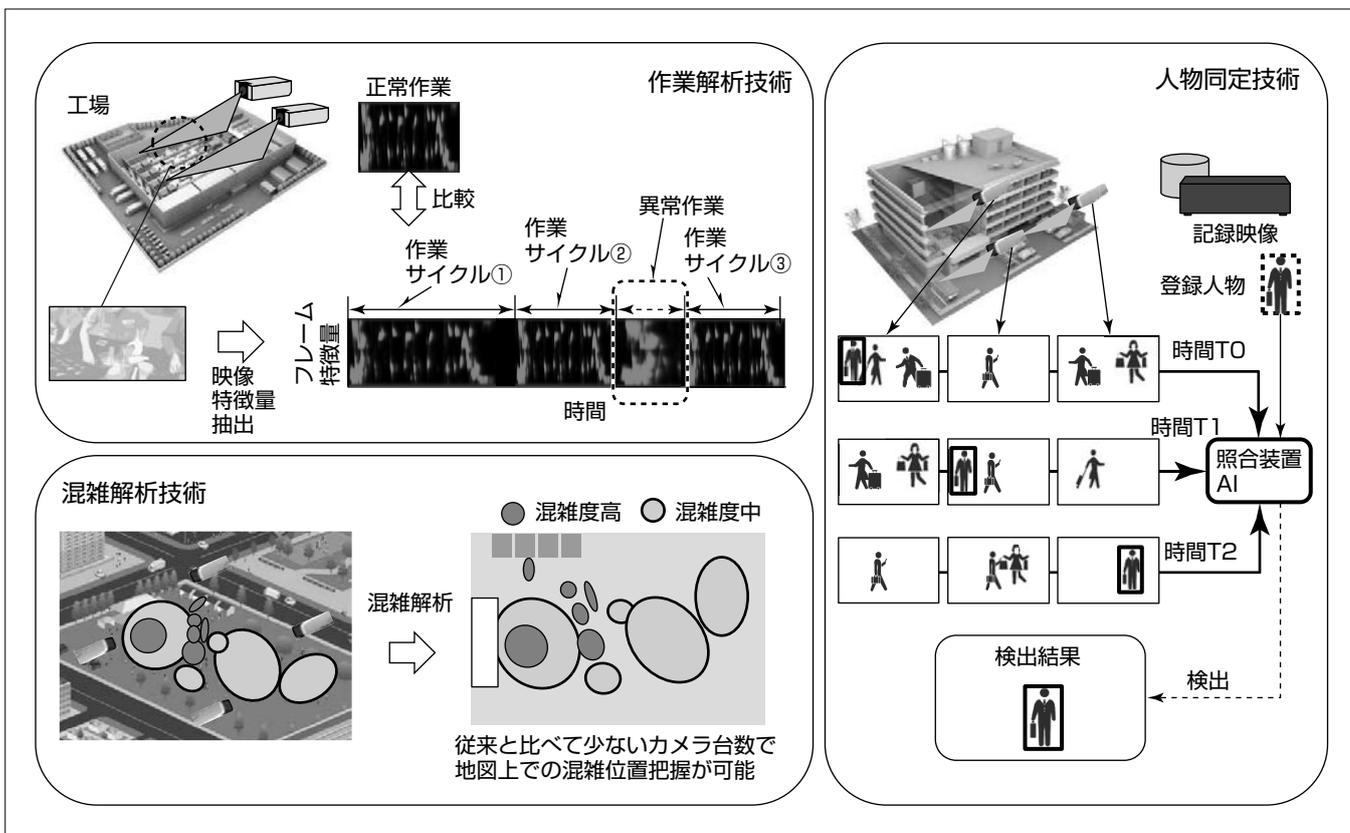
Video Content Analysis on Surveillance Systems

Hironori Saito, Kei Wakabayashi, Katsuhiko Kusano, Shogo Shimizu

要 旨

近年の映像セキュリティ分野は、映像の監視・記録による防犯用途に加え、映像解析(Video Content Analysis : VCA)を活用した安全性向上や業務支援などの付加価値提供によって、適用領域を拡大しつつある。三菱電機は、安心・安全・快適な社会の実現に向け、作業解析技術、混雑解析技術、人物同定技術の各分野で、製品化に向けた技術開発に注力している。作業解析技術は、例えば生産ラインに適用し、目視による作業の異常や無駄の検出を自動化することで、作業負担の軽減と生産性向上を図る。解析に使う映像特徴量の抽出で、計算量を当社従来比65%削減し、リアルタイム解析を実現した。混雑解析技術は、人数カウントや動線解析等へ応用可能で、セキュリティやマー

ケティングの分野でのニーズが高まっている。従来は、混雑による人物間遮蔽回避のため、複数視点での解析が必要であったが、当社は一視点映像での混雑度解析技術を開発し、少ない監視カメラ台数で実空間にマップ可能な混雑解析を可能にした。人物同定技術は、複数カメラ間で同一人物かどうかを照合する技術であり、行方不明者の捜索や不審者の追跡を実現する。当社は、カメラ間での人物の見え方の違いに対応するためAI(Artificial Intelligence)を適用し、解析対象の画像に対して時間的近傍の画像も学習させ、照合精度を当社従来比15%向上させた。今後は、これら技術開発成果の製品化に取り組み、安心・安全・快適な社会の実現に貢献していく。



作業解析技術・混雑解析技術・人物同定技術のイメージ

図は技術要素ごとのイメージを示している。作業解析技術の図は製造現場の作業員を動きの情報から算出される映像特徴量で解析している様子を示す。混雑解析技術の図は、実空間の混雑度合いを地図上に可視化した様子を示す。人物同定技術の図は複数カメラの映像データをAI技術で照合するシステム構成の一例を示す。

1. ま え が き

近年、監視カメラが広く普及したことによって、監視者の負担が増大しており、監視業務の自動化が求められている。加えて、カメラのデジタル化や高画素化の進展に伴い、VCAを組み合わせた高度な付加価値サービスの実現による、適用領域の拡大が見込まれている。市場ニーズとして、セキュリティ分野だけでなく、FA、交通、流通などの分野で、安全性向上と生産性向上に寄与するサービスの提供が期待されている。

このような動向の中、当社は監視システムの付加価値向上を目指し、VCAを用いた作業解析技術、混雑解析技術、人物同定技術の3分野を柱に技術開発を進めている。また、VCA実用化のアプローチとして、エッジコンピューティングの活用による迅速かつ高性能な分散処理の実現に注力し、一般的な映像解析システムとの差異化と、有用性の向上を目指している。

本稿では、当社が重視する解析処理の軽減・高速化技術を軸に、先に挙げた3分野での各解析技術の実現手段と成果・効果を述べる。

2. VCA分野での動向

2.1 VCAの動向

VCAの日本国内での市場規模は、東京2020オリンピック・パラリンピック開催に向けた需要もあって2020年頃まで拡大傾向と予測されている。VCAは顔認証、人数カウント、作業分析、混雑解析、人流検知、物体検知等、多様な分野で人を支援するサービスへと活用されつつある。VCAの解析精度も飛躍的に向上し、人を支援するサービスへのVCAの活用は今後高い成長が見込まれる。

また、近年VCAの分野では装置単体で解析するだけでなく、他システムとデータ連携をすることが増えてきた。解析処理をシステムのどの装置で実施するかによって構成が決定される。大きくは①エッジノード(カメラ、解析サーバ)又は、②クラウド上での解析の二つに分類される傾向にある。

2.2 当社の映像セキュリティ分野での取り組み

当社は、エッジコンピューティングの活用による分散処理に注力している。エッジノードで迅速かつ高性能に分散処理を行い、エッジノードでリアルタイム分析が可能なシステムの構築を目指している。近年ニーズが増加傾向にある他システムとのデータ連携では、エッジノードで一次解析を実施し、クラウドに送るデータ量削減を目指す。クラウドに送るデータ量削減に伴い、クラウド上での解析処理の計算量削減も可能になる。

当社は安心・安全・快適性の提供に向け、①製造現場での作業者の安全確保と生産効率の改善のための作業解析技術、②催事や商業施設での来場者の安全確保と動向分析するための混雑解析技術、③市街地や商業施設での迷子や不審者を捜査するための人物同定技術の開発に取り組んでいる。

3章では、当社が取り組む安心・安全・快適な社会の実現に向けた三つの解析技術について述べる。

3. 安心・安全・快適な社会を実現する映像解析技術

3.1 作業解析技術⁽¹⁾

工場の製造現場で、計画通りの生産性を達成するためには、作業工程ごとに定めた標準時間と実際の作業時間の差異を把握する必要がある。通常、標準時間は“部品を手取る”“ねじを1本締める”といった要素作業ごとに設定されており、現場の作業分析でも要素作業の単位で時間計測が行われる。従来は目視で計測を行い、また同時に異常作業や正常作業を分析していたが、分析者にかかる負担が大きいため、分析作業の自動化が課題であった。

この課題に対し、既設の監視カメラの映像を活用したVCAによる作業解析技術を開発した。開発技術は、実際に作業している映像(入力映像)と、事前に記録した作業映像(基準映像)とを比較分析することで作業時間計測及び異常作業抽出を行う。比較処理には、両映像から抽出した映像特徴量“映紋”を用いる。映紋は、映像を構成する各フレームについて、動きベクトルの方向別発生数を抽出し、時系列に並べた映像特徴量であり(図1)、作業動作に対して固有なパターンを示す性質を持つ。

一般に製造現場の作業は、同じ動作の繰り返しで構成されるため、入力映像の映紋には、基準映像の映紋と同じパターンが繰り返し現れる。この繰り返されるパターンを検出することで作業が行われたことを検知し、パターンの長さから作業時間を計測する。また、工場稼働時で基準映像の映紋と類似しない区間を検出することで、異常作業の発生を検知する。

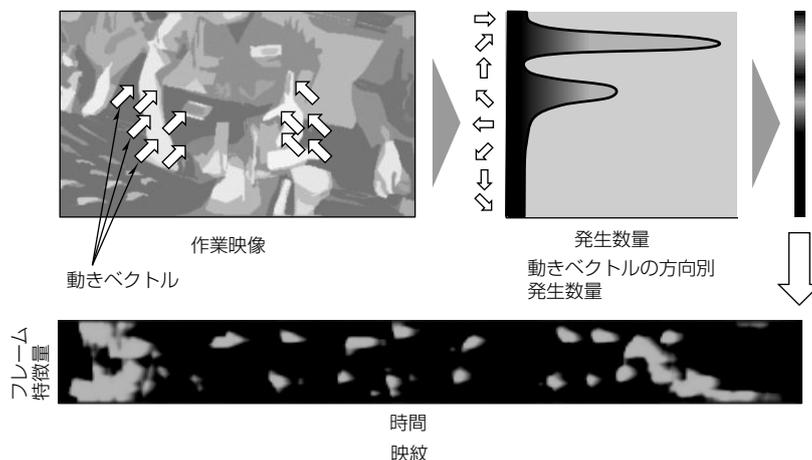


図1. 映紋抽出処理

映紋を抽出する上で重要となる動きベクトル情報を得るためには、通常、圧縮符号化された映像データをデコードで復号した後、時間的に連続する2フレームで動きベクトルを抽出する必要がある。しかし、復号処理と動きベクトル抽出の処理負荷が大きく、作業映像をリアルタイムで解析できないことが課題であった。そこで、圧縮符号化された映像データから直接動きベクトル情報を抽出する仕組みを構築した(図2)。圧縮符号化された映像データはもとより動きベクトル情報を持っており、復号処理の一部を実施することで抽出できるので復号処理の一部と動きベクトル抽出処理を削減でき、映像データから映紋を生成する処理について計算量を約65%削減した。これによって、作業映像のリアルタイム解析が可能になり、24時間稼働する生産ラインに対しても常時、作業を解析することが可能になった。開発技術の解析精度については、当社工場での検証の結果、95%以上の抽出精度で作業サイクルを抽出し、平均誤差0.5秒未満で要素作業の作業時間を計測できることを確認した。異常検知機能については、異常作業を模した作業映像から、異常区間を検知できることを確認した。

今後は、更なる計算量低減によって、同時に解析できる作業映像の数を増やし、工場内の生産性を工程単位でリアルタイムに把握可能なシステムの開発につなげ、製造現場の改善活動を支援する。

3.2 混雑解析技術⁽²⁾

セキュリティやマーケティングの分野で、監視カメラの映像を用いた混雑解析のニーズが高まっている。監視カメラの映像を用いるアプローチは、既設の監視カメラを利用できる点で有効である。しかし単眼の監視カメラで撮影した一視点映像を用いる場合、人物間の遮蔽があると混雑解析精度が低くなってしまう。

人物間の遮蔽が発生する高密度環境での一般的な混雑度

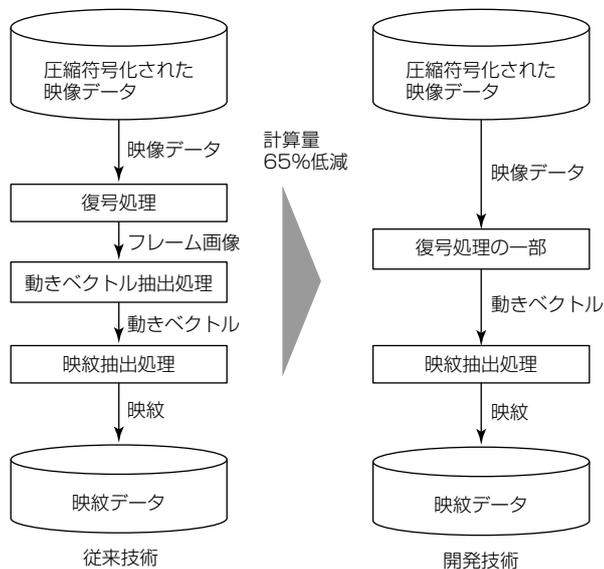


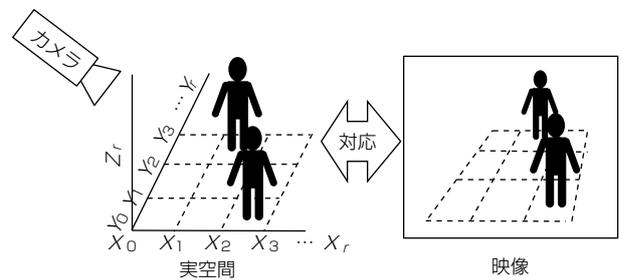
図2. 従来技術と開発技術の比較

解析技術として、前景の面積から密度を推定することで、人数をカウントすることなく混雑度を解析する手法がある。例えば、前景として複数の領域が抽出された際に、前景面積が小さい領域は人物間の遮蔽なし、大きい領域は人物間の遮蔽があると仮定して人数を推定している。しかし、この手法で可能なのは、映像上の混雑度の推定であり、実空間でどこが混雑しているか分からないという課題がある。

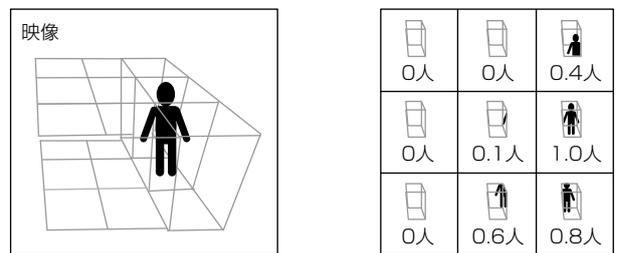
この課題を解決するために、人物間の遮蔽が発生している一視点映像から、実空間上の混雑度を解析する技術を開発した。この技術は、事前に撮影した映像を用いて、実空間と映像上での座標対応を取得し、映像上での人物の映り方から実空間上の位置及び人数の分布への変換式を導出する。その後、監視カメラ映像に対して、映像上の人物領域を抽出し、導出した変換式を適用することで混雑度解析が可能になる。

この技術は、変換式導出処理と混雑度解析処理に分かれている。変換式導出処理の流れは次のとおりである(図3)。

- (1) 実空間と映像上での座標の対応を取る
- (2) 映像を実空間と対応した直方体領域で分割
- (3) 直方体領域中の前景面積から映像上の人数分布を推定
- (4) 映像上の人数分布と実空間上の人数分布から変換式を導出

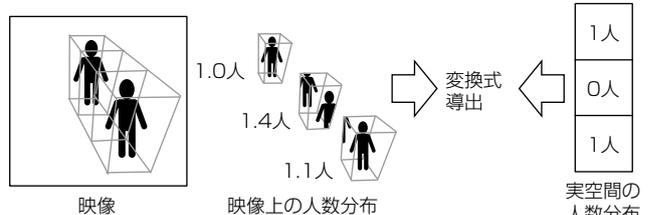


(a) 実空間と映像上の座標対応



(b) 直方体領域分割

(c) 映像上の人数分布推定



(d) 映像と実空間の人数分布から変換式導出

図3. 変換式導出処理

混雑度解析処理の流れは次のとおりである(図4)。(1)から(3)は、変換式導出処理と同様のため図は省略する。

- (1) 実空間と映像上での座標の対応を取る
- (2) 映像を実空間と対応した直方体領域で分割
- (3) 直方体領域中の前表面積から映像上の人数分布を推定
- (4) 映像上の人数分布を実空間上へ変換することによって、実空間上の混雑度を推定

この手法によって、解析対象2.5m四方のエリアを撮影した映像で、混雑度解析を実施した結果を図5に示す。横軸が映像中の解析範囲内の人数、縦軸が提案手法による混雑度推定結果の二乗平均平方根誤差(Root Mean Square Error: RMSE)である。解析範囲内の人数が少ない低密度環境、人数が多い高密度環境でもRMSEは最大でも0.4人と小さく、高精度に混雑度を解析できている。RMSE0.4人は、従来技術では2台のカメラを使用して実現しており、1台のカメラで実現できる提案手法が有用であると言える。

3.3 人物同定技術

イベント会場やショッピングモールでは、エリア全体に100台規模の監視カメラが設置されており、単一カメラ内で検出した人物を、エリア内に設置される複数カメラをまたがって追跡したい、又は過去に蓄積された映像から同一人物を検索したいという要望がある。

複数カメラの映像から同一人物を検出するためには、人物検出技術及び人物同定技術が必要となる。人物検出技術は、カメラ画像から人物領域を特定する技術である。人物検出技術によって、エッジノードで不要な情報量を削減することを可能にしている。人物同定技術は、複数カメラ間で同一人物かどうかを照合する技術である。

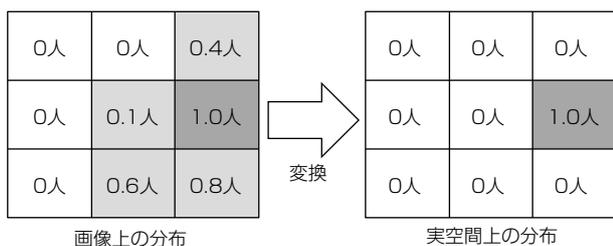


図4. 混雑度解析処理

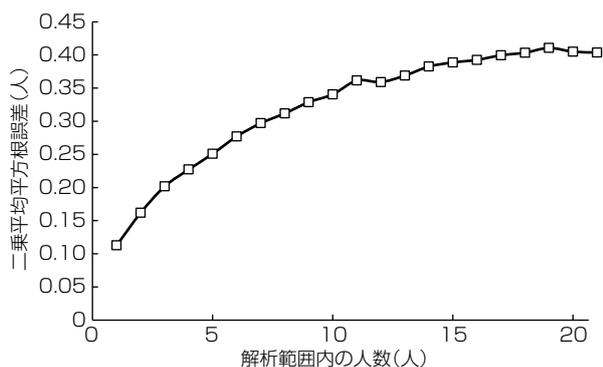


図5. 解析結果

複数カメラ間での人物同定を実現するために、AIを適用した人物同定技術を開発した。複数のカメラで撮影した人物ペアをあらかじめ学習することで、カメラ間での人物の見え方の違いも同時に学習できる。その後、学習した判定モデルを用いて、あるカメラで撮影された人物が、他のカメラで撮影された人物と同一人物か否かを判定できる。

この技術は、映像が時系列順に連続しているという特性を生かし、同定精度を向上させている。具体的には、ある特定フレームで、同定したい人物が映っている場合、従来は特定フレームでだけ特徴量を抽出していたが、この技術では特定フレームだけでなく、過去及び未来のフレーム群からも特徴量を抽出している(図6)。

この技術は、学習処理と人物同定処理に分かれている。学習処理の流れは次のとおりである(図7)。

- (1) 映像上の人物を検出する
- (2) 検出した人物に個人IDを付与
- (3) 検出した人物の特徴量を学習する
- (4) 人物同定を行う全てのカメラについて(1)から(3)を実施する

人物同定処理の流れは次のとおりである(図8)。

- (1) カメラAの映像から同定したい人物を指定する
- (2) 他のカメラBの映像で人物を検出する

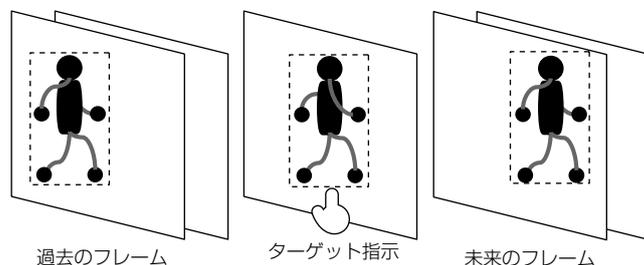


図6. 特定フレームの過去及び未来のフレームを利用

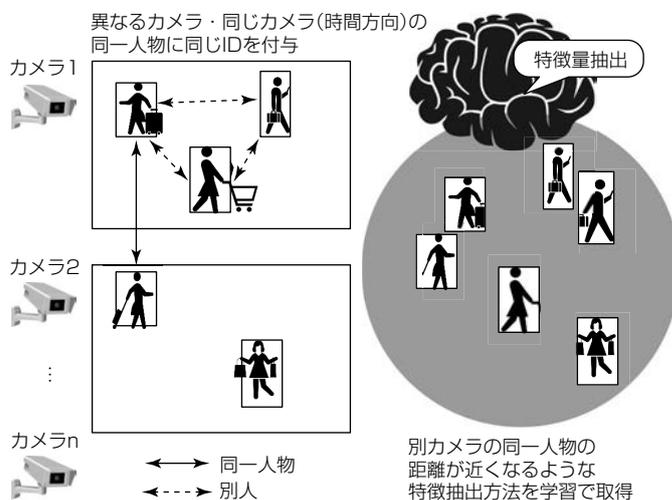


図7. 学習処理

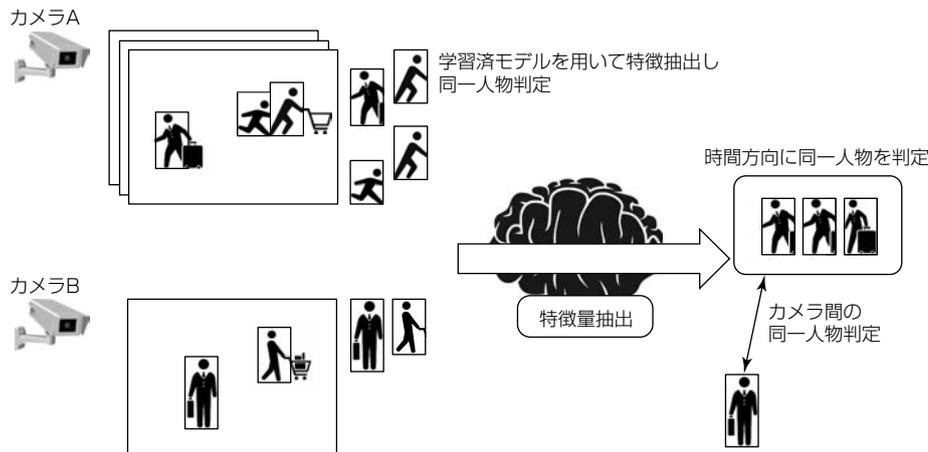


図8. 人物同定処理

(3) 学習した判定モデルによってカメラAで指定した人物の特徴量とカメラBで検出した人物の特徴量を比較し、同一人物か否かを判定する

当社オフィスビルで、エントランスと事務所入り口をカメラで撮影し、エントランスで指定した人物を事務所入り口で同定できるか否かを検証した。その結果、時系列方向の特徴を使わない手法と比べて照合精度が約15%改善した。

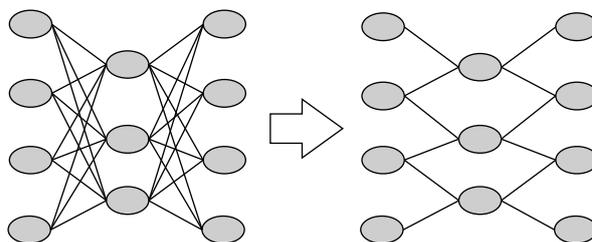


図9. コンパクトAIのイメージ

4. む す び

VCAは、現在目視で行っている、あらゆる事柄を自動化できる可能性があり、様々な分野で広く社会貢献が期待できる技術である。ただし、映像データの解析には膨大な計算が必要となるため、現状ではGPU(Graphics Processing Unit)搭載サーバやクラウド連携による実現が主流である。クラウドベースのVCAは、増加の一途と見込まれる映像データがネットワーク上に流出することが、普及での大きな課題である。また、クラウド処理はデータ転送遅延等によって、リアルタイム性が損なわれるため、VCAを活用するアプリケーションの広がりには制限が生じる。

そこで、当社では、VCAの普及にはエッジコンピューティングが不可欠と考え、エッジデバイスでの実用性確保のため、計算量削減と処理高速化に注力している。

作業解析技術では、先に述べたとおり、計算量を65%削減した。混雑解析技術では、カメラを増やさず1台の映像から解析可能にしたため、データ量が大幅に削減され、解析装置での処理高速化を実現した。人物同定技術では、人物検出技術を用いてエッジデバイス上で解析するデータ量を必要最小限に抑制することで高速化を実現している。

また、本稿では詳述していないが、当社はAIの推論処理の計算量を軽減して高速化する技術“コンパクトAI(図9)”を保有している。コンパクトAIは、推論に用いるニューラルネットワーク構造と計算方法を効率化することで、計算量を削減する技術であり、映像解析など大規模データの高速解析に効果を発揮する。さらには、ソフトウェア実装技術面でも処理高速化に取り組んでおり、VCA処理の演算特性に着目した実装最適化によって、GPUを使用せずに実用速度の確保を目指している。

今後、これら技術開発の更なる進展を図るとともに、取り組みの集大成としてVCA技術を搭載した製品を市場に投入し、安心・安全・快適な社会の実現に貢献していく。

参 考 文 献

- (1) ネットワークカメラの付加価値向上技術“映紋”，三菱電機技報，92，No.1，25（2018）
- (2) 新井士人，ほか：一視点画像を用いた空間的な人数分布の推定，2017年映像情報メディア学会冬季大会，21A-5（2017）