

AIを活用した映像解析ソリューション “kizkia”

中尾堯理* 鈴木宏則*
菊池 領* 山足光義*
北島周平*

kizkia : Video Analysis Solution Using Artificial Intelligence

Takamasa Nakao, Ryou Kikuchi, Shuuhei Kitajima, Hironori Suzuki, Mitsuyoshi Yamatari

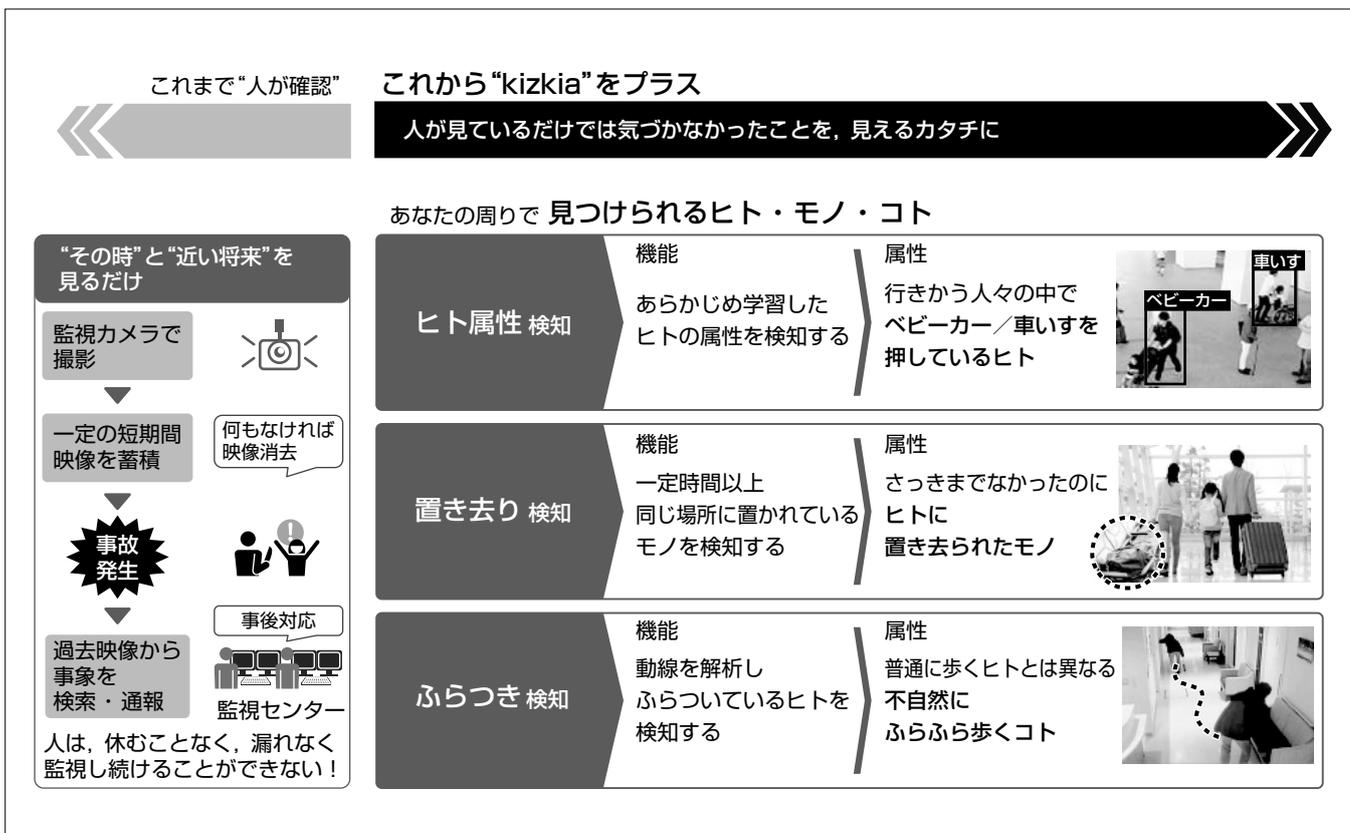
要 旨

街中に監視カメラがあふれているが、従来は事件発生の際に撮り溜(た)めた映像を検証するだけで、貴重な映像情報が有効活用されていなかった。東京2020オリンピック・パラリンピックに向けて安心・安全の向上が求められており、監視カメラの市場は拡大傾向にある。また訪日外国人や高齢者など、サポートが必要な人へのおもてなしが期待されている。

このような期待にこたえるため、三菱電機インフォメーションシステムズ株式会社(MDIS)は、近年急速な発展をとげているAI(Artificial Intelligence)技術を活用した、映像解析ソリューション“kizkia(きづきあ)”を開発した。kizkiaは、ヒト属性検知、置き去られたモノやふらついてい

るヒトの検知機能を備え、従来の機械学習では定義が難しかった曖昧な属性や、人が気づけないことを見つけ出せる。また、新しくモノ属性検知の開発を行っており、これもkizkiaに実装予定である。kizkiaは三菱電機のAI技術ブランド“Maisart(マイサート)”の技術を活用することで、GPU(Graphics Processing Unit)レスの映像処理を実現している。これらの機能によって、サポートが必要な人への“リアルタイムな対応”や、従来は分析・想定がしにくかった“未来の予測”の支援を可能にし、安心・安全・快適な社会の実現に貢献する。

MDISではkizkiaで用いるAIに必要な学習データの作成で、数々の実証実験の経験・知見を活用した効率的な学習プロセスを確立している。今後は映像以外のデータとの連動によって、様々な利用シーンでの活用が期待される。



映像解析ソリューション“kizkia”

kizkiaは、人が見つけたいことを、様々な映像から人工知能によって休むことなく自動的に見つけ出し通知する。これまで人が見ているだけでは気づかなかったことへの対応や未来予測の支援が可能となる。

1. ま え が き

街中の監視カメラやドライブレコーダなどの映像が様々な事件の解決に活用されることが多くなった。しかし事件が発生した後に撮り溜めた膨大な映像から、人手をかけて関係する事象を見つけ出す必要があり、リアルタイムにキーとなる事象を見つけて事件を解決するようなことには活用できていない。

東京2020オリンピック・パラリンピックに向けて安心・安全の向上が求められており、カメラの市場は拡大傾向にある。また訪日外国人や高齢者など、サポートが必要な人へのおもてなしが期待されている。

このような期待に応えるためにMDISは、近年急速に発展しているAI技術を活用し、カメラの映像をリアルタイムに解析するソリューション“kizkia”を開発した。

本稿では、AI技術を活用した映像解析ソリューションkizkiaの特長や学習プロセスの一例について述べる。

2. ディープラーニング

今注目されているAIは、従来のAIと何が違うのか。例えば、従来の機械学習による顔認証では、スキルのある技術者が顔の特徴量である目の中心や唇の端などの位置や距離から計算される値を定義し、それを検出するアルゴリズムを作りこむ必要があった。人手による設計は職人技に近い部分もあり、検出精度にも限界があった。

ところが2012年の画像処理コンテストで、ディープラーニングが従来手法に対して圧倒的な精度で勝利したところから有名になり、現在のAIブームを引き起こした。ディープラーニングで画像認識を行うと、これまで人が定義していた対象の特徴をAIが掴(つか)むことができる。例えば、車いすやベビーカーを認識させるために、それぞれの大量の画像を用意すれば、AIが自分で着目すべき特徴を見つけて、それぞれを識別することができるようになる。

kizkiaは、ディープラーニングで映像を解析することで、従来の機械学習では見つけられなかった曖昧な属性の違いや、人が見ているだけでは気づかなかったことを見つけ出し、サポートが必要な人への“リアルタイムな対応”や、これまでは分析・想定しにくかった“未来の予測”を支援することが可能になる。

3. 映像監視での課題

大規模ビルや公共施設などの監視センターではスタッフが多数のモニターで監視カメラ映像を監視して異常がないかを確認している。しかし人は休むことなく監視し続けることができないため、漏れなく事件や事故を発見することは難しく、また未然に防ぐこともできない。したがってほとんどの場合、事件が起こってからレコーダに蓄積した過去の映像から事象を検索するという使い方しかできていなかった(図1)。

また、カメラの台数が増加するにつれ、大量のカメラ映像をネットワークに流す必要があり、設備コストや配備する監視員数の増加といった問題があった。

4. 映像解析ソリューション“kizkia”

4.1 kizkia

kizkiaはネットワークカメラやレコーダから映像を取得し、AI技術を使ってヒトの属性やふるまい、所有物、モノの種類や属性をリアルタイムに検知する映像解析ソリューションである。

多くのカメラに対応するため、映像処理をカメラに近いところで実施し、解析結果をセンターに送信するエッジコンピューティングの構成で実現している。また、リアルタイム解析を実現するために様々な映像処理の軽量化やマルチCPU(Central Processing Unit)による多重処理を行うソフトウェアアーキテクチャになっている。MDISのシステム化技術によって、三菱電機の“コンパクトな人工知能”

■ 従来は“録画映像を後から確認”するだけ → “人”は休まず漏れなく監視し続けることはできない



図1. 映像解析を実現に向けた現状の課題

を実装して、AI製品で一般的に使用されるGPUなしでの映像処理を実現している。

また機械学習で重要となる学習データの作成では、数々の実証実験の経験・知見を活用した効率的な学習プロセスを確立した。4.6節で、kizkiaの特長と、AIの学習プロセスの一例について述べる。

4.2 kizkiaのアーキテクチャ

kizkiaは、個人情報漏洩(ろうえい)やネットワーク負荷を減らすため、ネットワークカメラの映像の処理はカメラが設置されている拠点(解析サーバ)だけで行い、検知したイベントの情報だけを抽出して監視センター(統合サーバ)に送付するエッジコンピューティングのアーキテクチャになっている。映像を、カメラやレコーダからストリームで受信し、処理が終わると破棄するセキュリティに配慮した構造になっている(図2)。カメラ映像の処理は、カメラごとに独立・並行して実行する分散処理構造になっており、カメラの増減、カメラごとの監視機能のオンオフ、検知対象の変更などきめ細かく柔軟な運用を可能にしている。さらに、リアルタイム性を実現するため、ストリームの映像処理(切り出しや判定などの処理)をフレーム単位で並行して実行するマルチタスク処理に加え、ディープラーニング技術による切り出しや所持品・種別判定の処理のコンパクト化⁽¹⁾によって、計算量を従来の1/10以下にまで削減してGPUレスでの映像処理を実現している⁽²⁾。また、複数の解析サーバの設定変更やカメラごとの監視機能のオンオフを監視センターから遠隔で操作できるため、保守性の高いシステムの構築を可能にしている。

4.3 ヒトの属性・所有物の識別

kizkiaでは、ディープラーニングの利用によって、従来のAIでは難しかったヒトの属性や所有物などの曖昧な違いもリアルタイムに把握できる。これによって、ベビーカーを押すヒト、車いすを押すヒト、杖(つえ)をつくヒトなど、ヒトの特徴を捉える“ヒト属性検知”を行うことができる(図3)。“ベビーカーを押すヒト”などの特徴を属性と呼び、ディープラーニングに属性を持った画像を大量に学習させることで、特徴量の設計やプログラミングを行わずに、画像の中から属性を識別するパラメータを自動算出する。kizkiaは、この属性を最大40種類まで同時識別できる。ヒト属性検知機能によって、公共施設などで、車いすのヒトや白杖(はくじょう)をつく視覚障がいのある人をリアルタイムに検知してスタッフに通知するなどして事故を未然に防ぐことなど、サポートが必要な人への早急な対応が可能になる。

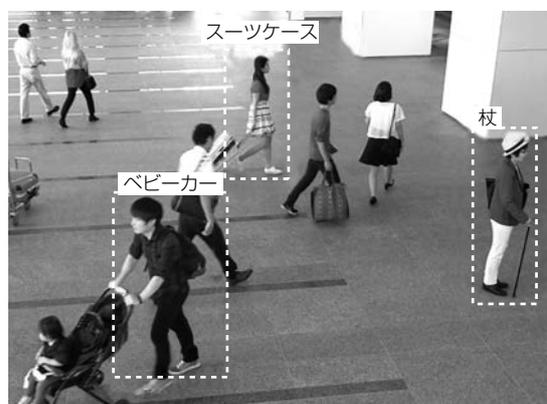


図3. ヒト属性検知

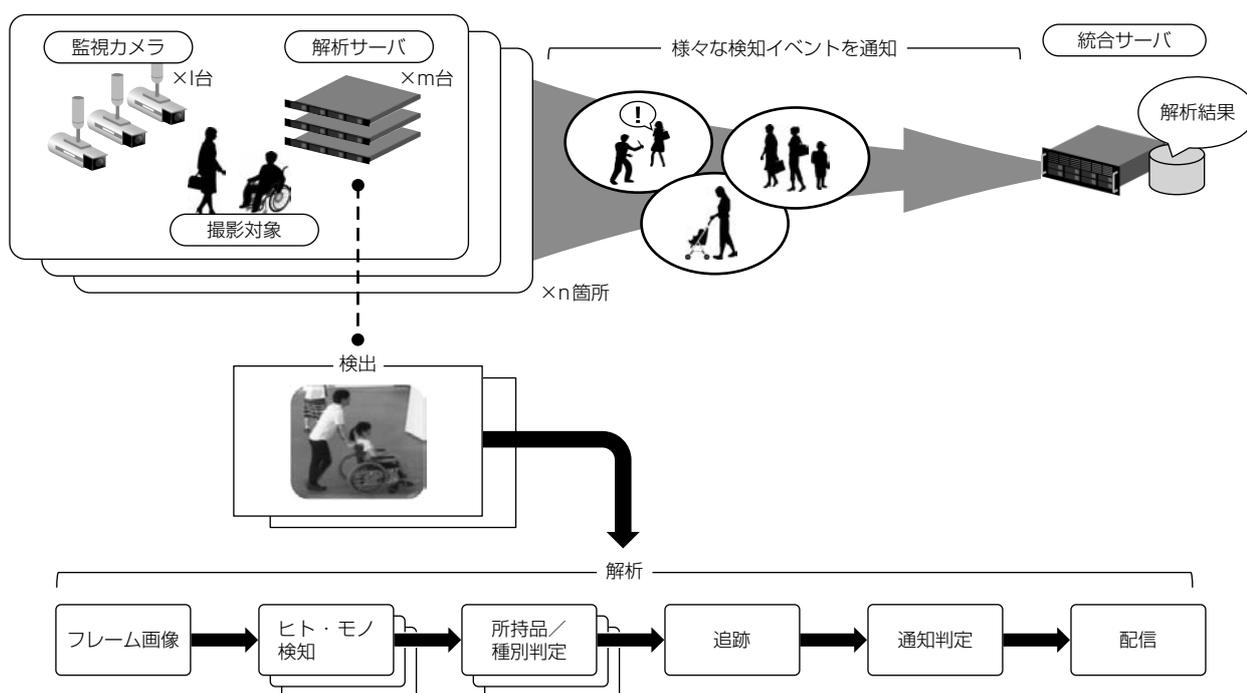


図2. kizkiaのアーキテクチャ

4.4 モノの種類・属性の識別

“ヒト属性検知”に対して、モノをリアルタイムに検知する“モノ属性検知”を新たに開発しており、この機能は今後kizkiaに実装する予定である。モノ属性検知は、2段階のディープラーニングを使って実現している。まず、車、工具、ショッピングカート、ジョッキなどのモノの種類を事前に学習させて検知するモノ検知機能と、次にヒトと同様にモノの属性(トラック、乗用車、救急車など)を別に学習させて属性を識別する属性検知機能、この二つの機能を組み合わせることで様々なモノの種類と属性の検知を実現している。

モノ属性検知機能で、例えばジョッキを検知し、さらにジョッキの中のドリンクの量の違いを画像で学習させることで、図4のようにドリンクが空(実線の枠)のジョッキとそうでないジョッキ(点線の枠)を識別する。これによって、店員が客から呼ばれる前に追加注文が入りそうな客のところに向かうなど、客へのサービス向上、売上げ拡大の応用が考えられる。また、車両判定への応用として、車種別の通行量をカウントさせるといったことや、工場への応用としてラインの不良品を学習・検知させることで、人手作業による漏れの予防や作業員・監視員の削減によるコスト改善等が期待できる。

4.5 ふるまい検知

kizkiaでは、“ふらついているヒト”や“うろついているヒト”など、行動を分析し、特徴的な行動をリアルタイムに



図4. モノ属性検知(ジョッキの状態)

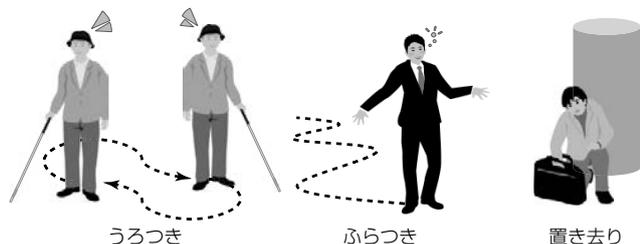


図5. ふるまい検知例

検知することが可能である。また、“置き去られたモノ”を検知することもできる(図5)。

ふらつきやうろつきの検知では、時系列に沿って同一人物を観察し、動線を検出・解析することによって、ふらついているヒトや、特定のエリアに何回も現れる不審者を検知できる。

これらの検知によって、酔っ払いの道路への飛び出しや、線路への転落の予兆を検出し、事故発生前に対応できる。

置き去りの検知では、長時間置き去りにされたモノを検出することで、忘れ物を見つけるだけでなく、放置された危険物があることを察知して事件を未然に防ぐことを支援できる可能性がある。

4.6 学習プロセスと精度

学習では、どのような画像をどれほど用意できるかが重要であり、またその学習後の検知結果を評価してチューニングするといったプロセスが検知結果の精度に大きな影響を与える。ここではベビーカーを例に挙げて学習プロセスを述べる。

AIにベビーカーを検知するよう学習させたい場合は、まずOK画像(ベビーカーを押しているヒト)とNG画像(ベビーカーを押していないヒト)を大量に与える。すると2種類の画像群のどこが異なるかを自動的に計算し、ベビーカーを押しているヒトの画像だけに現れる特徴を抽出し、AIはOK・NGを判定できるようになる。ここで重要なのが、二つの画像の異なる部分がどのようになっているかであり、もしOK画像とNG画像にベビーカー以外の特徴の差(例えばベビーカーを押している人の服の色が赤く、そうでない人の服が青ばかりなど)が多く出現すると、そちらの特徴に反応してしまうこともあり得る。そのため、できるだけ多種多量の画像を集めてOK画像とNG画像の間に想定しない差が現れないようにする必要がある。

さらに、OK画像とよく似た特徴を持つもの(車いす・ショッピングカートなど)が現れるような場合は、それらをOK画像と誤判定しないように、NG画像にOK画像と似ていると思われる画像を含める必要がある。具体的にどのような物がOK画像と似ているかを事前に想定することや、一度学習をした後にkizkiaを適用する現地の映像に対して解析を実施してOKと誤判定したヒトを学習データとして追加するなど、できるだけOK画像と似ているがOK画像の特徴ではないことを学習させていくことが重要となる。このような学習をさせていくことで間違いやすい部分を減らしていけるが、どこまでの想定やデータの追加が必要かの判断が難しい。この判断を行うための一つとして“中間処理の可視化”がある。

AIが具体的にどのように判定を行っているかは、中間処理を可視化することである程度は確認できる。例えば画像のどこを見てOK・NGを区別しているかの確認や、そ

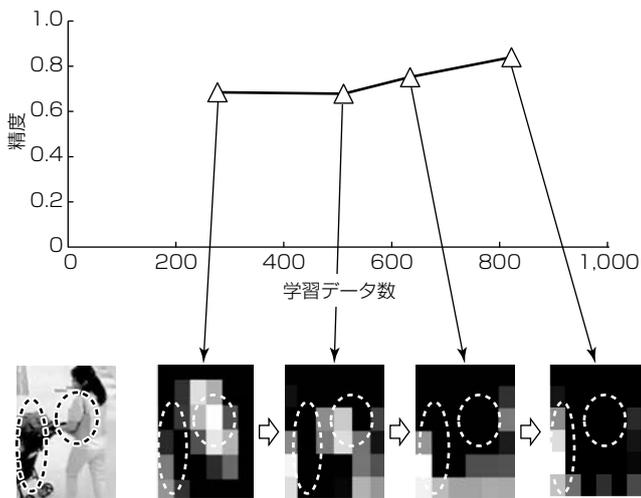


図6. 学習データと判定部位の可視化の例

これらの特徴の大きさ(AIにとっての重要さ)の確認等が可能である。図6の上のグラフから学習データ数が増えるほど精度が上がっていることが分かる。また、下の図は左の人物画像のどこに特徴があるかを可視化したものである。白いほど強い特徴として判定されており、学習データが増加していく中で対象物の特徴を正しく捉えていっていることが分かる。このような可視化情報に加え、様々な情報(OKとNGの比率、学習データと評価データの分布の差、複数属性の場合の影響度等)を考慮しながら学習を進めていくことで、高い精度を持つAIを構築できる。

5. 今後の展開

kizkiaはネットワークカメラの映像データの解析を対象としているが、今後はカメラ以外の映像データや、SNS (Social Networking Service)上のテキスト、街中や工場

などに設置されたセンサ情報(匂い、音なども含めて)、個人のID (IDentification)情報、移動体の情報など、多種類のメディアデータの解析結果と連携させることによって、様々な利用シーンでの活用が期待される。

6. むすび

AI技術を活用した映像解析ソリューションkizkiaの特長や効率的な学習プロセスの一例について述べた。

kizkiaは、公共機関や製造ラインなど様々な場所で、人や業務を支援する仕組みとして広く利用されはじめた。映像データから様々な事象を検知できることから、その応用範囲は広がる可能性が高い。

ただし何に対しても効果が上がるわけではなく、kizkiaを導入するに当たっては、業種・業務ごとに解決したい課題を洗い出して検知したい事象を明確に定義した上で、学習の基となる映像データの収集を行うことが重要である。様々な業種・業務の現場の人々と一緒に目的を共有し、現場で実証実験を十分に行ってからシステムを導入するステップを実践することによって、安心・安全・快適に暮らせる社会の実現に貢献していく。

参考文献

- (1) Matsumoto, W., et.al : A Deep Neural Network Architecture Using Dimensionality Reduction with Sparse Matrices, 23rd International Conference, ICONIP 2016, 397~404 (2016)
- (2) 中尾堯理, ほか: 疎なネットワーク構造を持つDeep Learningを用いた映像分析システム, 情報処理学会第79回全国大会, 6B-07 (2017)