

映像解析ソリューション“kizkia” による実用的なAIシステムの構築

山足光義*
Mitsuyoshi Yamatari
中尾堯理*
Takamasa Nakao
菊池 領*
Ryou Kikuchi

阿部紘和*
Hirokazu Abe

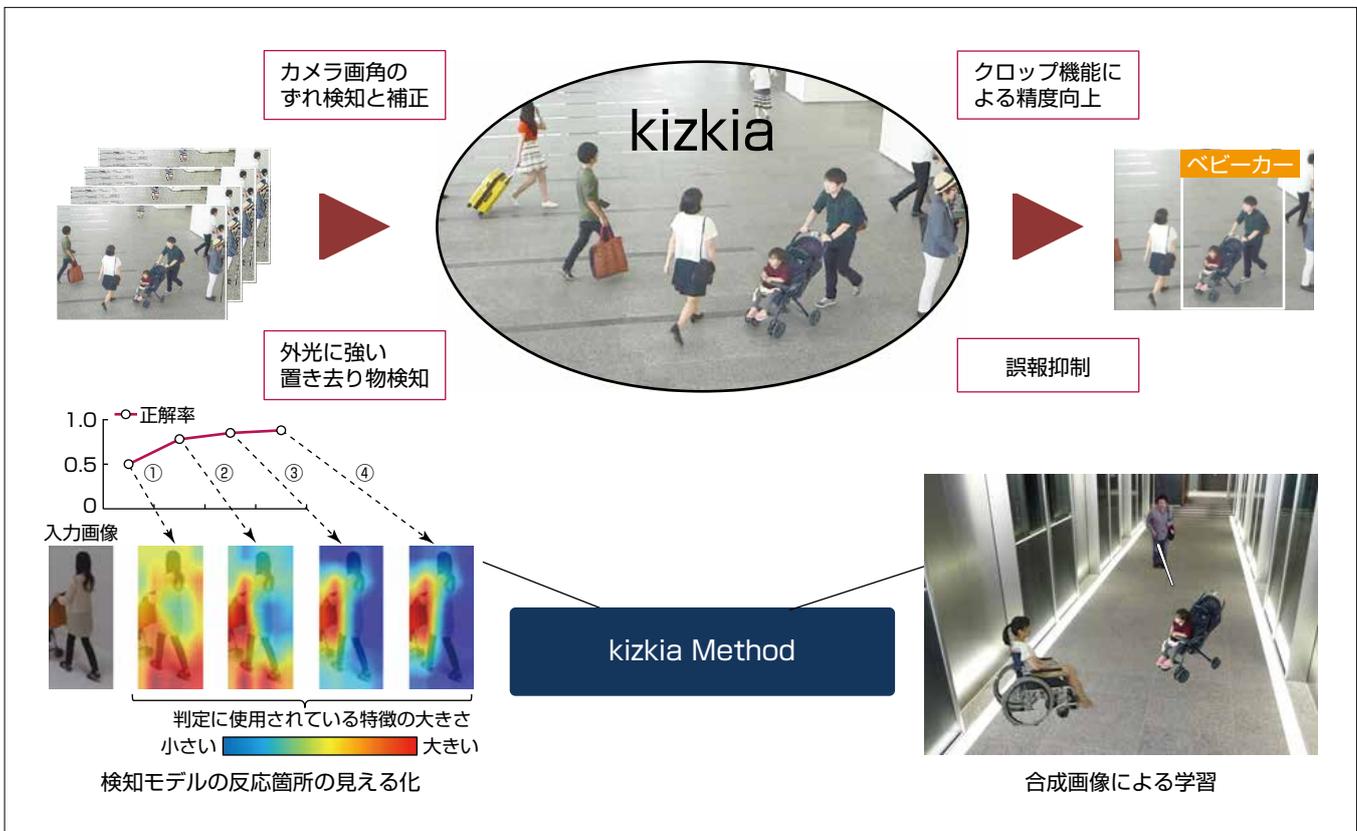
Construction of Practical Artificial Intelligence System
by Video Analysis Solution "kizkia"

要 旨

ネットワークカメラ、AI(Artificial Intelligence)技術の進展によって、AI技術を活用した映像解析の応用システムが広がり始めている。一方で、AI技術はブラックボックス的でどのように動作しているのかが分からず、安定的な実用システムを構築する上での課題が多い。人のレベルを超えた高精度な検知が可能な一方で、100%の検知が難しいAI技術を実システムに組み込むには、いろいろな工夫が必要である。そのため、実証実験のフェーズからなかなか脱却できない事例が多いのも実情である。

三菱電機インフォメーションシステムズ株式会社(MDIS)は、AI技術を活用した実用的なAIシステムの構築を可能にする映像解析ソリューション“kizkia(きづきあ)”の製品を開発し、提供している。

kizkiaは、ヒト属性検知、モノ検知、置き去り物やふらついているヒトなどの動き・状態の検知機能を備え、従来の機械学習では定義することが難しかった曖昧な属性や、人が見ているだけでは気づけないコトを見つけ出すことが可能である。kizkiaを利用した大規模な公共施設向けのシステム構築の実績等で培った知見を基に、実用的なAIシステムを構築する上での考慮点や、AIシステムで重要になる検知モデルについて、高精度な検知モデル開発手法“kizkia Method(kizkiaを用いた検知モデル開発手法)”を確立した。kizkia Methodを活用することで効率的に高い検知精度を実現することが可能になる。今後もAIを活用した安心・安全・快適な社会システム実現に貢献していく。



映像解析ソリューション“kizkia”と検知モデル開発手法“kizkia Method”

MDISは、2014年度から数々の実証実験に参画し、映像解析技術の知見を蓄積してきた。それらの知見に基づいて、三菱電機のAI技術“Maisart”を実装した映像解析ソリューション“kizkia”による実用的なAIシステムの構築を可能にする検知モデル開発手法“kizkia Method”を確立した。

1. ま え が き

街中に監視カメラがあふれてきている中、ディープラーニングの技術の進展によってカメラ画像を活用したAI技術による目視作業の省力化が少しずつ広がり始めている。AI技術は、学習によって人より高精度な検知率の実現が可能とされている。一方で、検知率向上の具体的な学習方法や100%にはならない検知率を実システムでどのように取り入れていけばよいのかなど、実用化に向けて試行錯誤しているところが多く、実証実験のフェーズからなかなか脱却できない事例が多いのが実情である。

MDISは、ディープラーニングが広く注目されはじめた2014年頃からAI技術への取組みを始め、特に“実用性”にフォーカスしながらリアルタイムに映像解析を行えるソリューション“kizkia”の開発に取り組んできた。

本稿では、どのようにして実用的なAIシステムを構築していくのかを、AIシステム構築の実績をベースに述べる。

2. 実用的なAIシステムとは

ディープラーニング技術や、TensorFlow^(注1)、PyTorch^(注2)などのOSS(Open Source Software)の普及によって、AI技術を活用して画像から物体を認識させることが容易になってきている。OSSやサービスが提供している検知モデル(検知するアルゴリズムとアルゴリズムで使用するパラメータのセット)を使うことで画像から物体や状態が判定できる一方、検知モデルが提供されていないものを検知するためには、検知モデルを作成する“学習”という作業を行う必要があり、この作業が従来のシステム開発にない新しいプロセスになっている。また、提供されている検知モデルを使っても、検知精度が悪く、検知精度を向上させたい場合も同様である。この段階に至ると、どれくらいの量を学習させればよいのかといった知見が必要で、さらに学習させる画像の不足、画像収集に起因する開発期間や費用の増大などの課題に直面する。また、外光変化の影響によって朝の日差しが強い時は検知できないなどの時間帯の検知精度の変化やカメラの設置角度や背景の違いなどで検知精度が変わるなどカメラごとの検知精度の変化への対応も必要になってくる。AIというキーワードが普及したこともあり、AIが万能でいろいろなことができると考えがちであるが、現在の技術レベルを考慮した上で、AIが得意とするところ、不得意とするところを見極め、不得意とするところは避け、AIをうまく活用したシステムを構築していく必要がある。

(注1) TensorFlowは、Google LLCの登録商標である。

(注2) PyTorchは、Facebook, Inc.の登録商標である。

3. 映像解析ソリューション“kizkia”

3.1 kizkia

映像解析ソリューションkizkiaは、ネットワークカメラやレコーダから映像を取得し、三菱電機のAI技術“Maisart”を使ってAI処理を軽量化することで、杖(つえ)を持つヒトなどヒトの所持物を条件にした検知(人属性検知)、スーツケースなどモノの検知、置き去り物の検知、酔っ払いなど人のふらつき検知、人の座込み検知などをリアルタイムに行う製品である⁽¹⁾⁽²⁾。図1は、カメラ映像からリアルタイムに特定の所有物を持っている人を検知している例である。

また、ディープラーニングを使って様々な検知ができるように、検知モデルを作成する学習ツールも提供している。図2にツールが提供するAIに学習させるための物体を四角で囲って切り出すGUI(Graphical User Interface)画面の例を示す。このツールを利用することで、ベビーカーやスーツケース、台車など様々な検知したいものを簡単に学習させ、システム化することが可能になっている。

実用的なAIシステムを構築する上で特徴的なkizkiaの機能を3.2節から3.5節で、MDISの検知モデル学習ノウハウを体系化した検知モデル開発手法であるkizkia Methodについては4章で述べる。

3.2 検知精度向上クロップ機能

AIを使って画像の中から物体を検知する場合、小さい

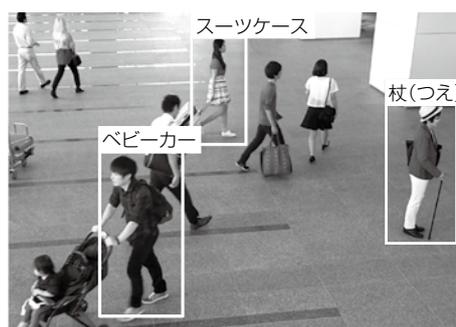


図1. リアルタイムな人属性検知



図2. 検知モデル作成学習ツールのGUI画面

物体の検知が難しいという課題がある。ディープラーニングは、事前に設計した深層学習のネットワークに合わせて入力する画像サイズのリサイズ処理を行っている。しかし、**図3**のように画面遠くに写り込む対象は、リサイズすることによって小さくつぶれてしまい、検知できなくなることが実運用での課題になった。特に白杖(はくじょう)や松葉杖などの元々細い物体はリサイズによって更に細くなってしまい、検知精度低下の一つの要因になっている。

この課題に対応するため、**図4**のように画面奥側の一部領域を拡大したクロップ画像を、元画像と同時に画像解析して結果を統合する2段階での検知機能(クロップ機能)を採用している。この結果、リサイズ後の画像に写る検知対象の大きさを十分確保できるため、画面奥の小さな物体であっても精度よく検知することが可能になった。

3.3 外光に強い置き去り物検知機能

カメラを屋外に設置して利用する場合、例えば、日照変化による見え方や明るさ、色の変化等、実運用するに当たってはこれらの変化を考慮する必要がある。

屋内の置き去り物検知では、背景画像と現在映像との差分を時系列的に判断し、色をベースにして一定時間以上差分が存在する領域を検知していた。しかし、屋外では外光等日照変化によって置き去り物がない状態でも色が変わり、正しく置き去り物が検知できない事象が多発した。そこで、屋外での利用に対応するため、外光の影響を受けない方式

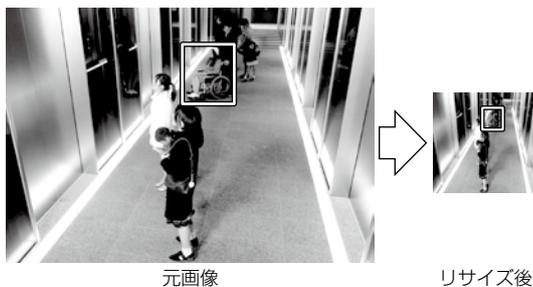


図3. リサイズで車椅子が見えなくなる様子

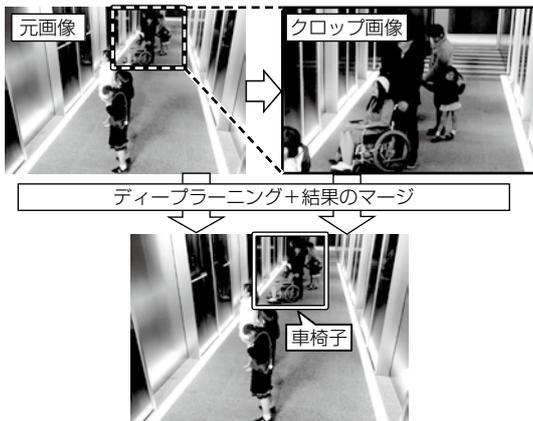


図4. クロップ機能

として画像の色の比較でなく物体の輪郭を比較する方式を採用し、屋外で誤検知を少なくする屋外用置き去り検知機能を実現した。

図5に色を活用した差分と輪郭を活用した背景差分の比較を示す。背景は、朝の人がいない時の背景画像であり、映像は昼のある時点のカメラ映像である。色を活用した差分(**図5(a)**)では昼の日照の強さによって地面等の色が変わり、差分の出る範囲が広がっている。一方、輪郭を活用した差分(**図5(b)**)では日照変化による影響を減らし、誤検知が削減できている。

3.4 誤報抑制機能

3.3節で述べた外光に強い置き去り物検知方式を用いることで多くの誤検知を削減できるが、例えば日照によって発生する柱の影は決まった位置に発生し、背景を人のいない夜間に取得すると背景には影が存在しない。その結果、背景にない影を置き去られたものと誤検知してしまうといった事象が発生する。このような影は毎日同じ場所で発生するため、毎日誤検知が発生する。また、人目でも明らかな変化があるため、誤検知を抑制することもできない。誤報が発生すると、置き去りか誤検知か個別の判別を人手で確認する必要がある。

一般的に、AI技術を活用する上で100%の検知率を目指すのは、技術的にもコスト的にも現実的ではなく、ある程度誤検知が出るという前提でシステムを開発する必要がある。このような誤検知による誤報は設備や一定の場所で定期的に多く発生するケースが多く、次のように誤検知による誤報を抑制することが現実的には有効である。

- (1) 発生した誤報の検知画像を人が目視確認し、誤検知であればその発生情報を登録しておく。
- (2) 以降、新規に検知されたものが登録された発生情報と類似するか判定を行い、登録されている誤検知と類似しない場合のみ通知することで誤報を抑制する。

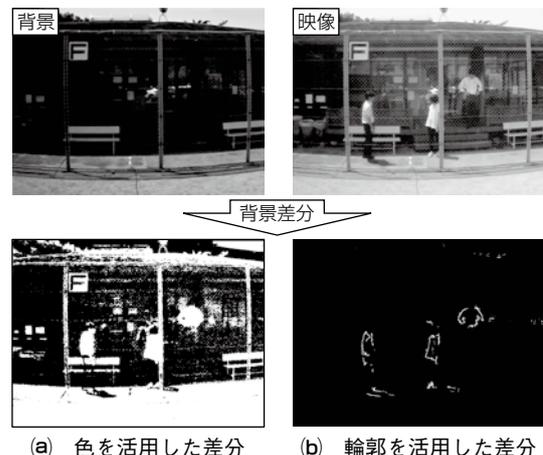


図5. 背景差分の比較

この機能を活用することで、定期的に発生するような誤検知については検知を抑制することが可能になる。

3.5 画角ずれ検知・補正機能

実運用では、風、振動、レンズ清掃等によってカメラの画角が物理的に移動することが多々起こり得る。カメラの画角ずれが発生すると、映像の中で監視領域を設定している場合に、監視領域と違う場所を監視することによる誤報や失報など、目的の検知が行われれないという課題が発生する。また、置き去り物検知では、背景画像と現在映像との差分を利用するが、背景画像と現在映像の画角が異なるため、誤報が発生する原因にもなる。これらに対応するため、kizkiaでは、画角が変化したことを自動的に検出し、ずれの大きさに応じて適切な処理を行う機能を持っている。

自動検出を行う方式として、ディープラーニングで用いられている特徴点抽出と、パノラマ画像の合成で使われる特徴点マッチングに着目した。

図6は、上段に示す画角がずれた2枚の画像からそれぞれ特徴点を抽出し、得られた特徴点同士をマッチングさせた結果を下段に示している。図中で、互いに同じと判定された特徴点同士は線で接続されており、片方の画像にだけ出現する特徴点是对応する特徴点がないため接続されていない。

この結果を利用し、互いに同じと判定された特徴点同士が画像間でどのくらいずれているかを算出して平均を得ることによって、画像間のずれ値(水平移動・垂直移動・回転・ズーム率)を精度よく算出できる。算出した値の大きさに応じて、運用者への通知や監視領域設定の自動補正などを実施する。

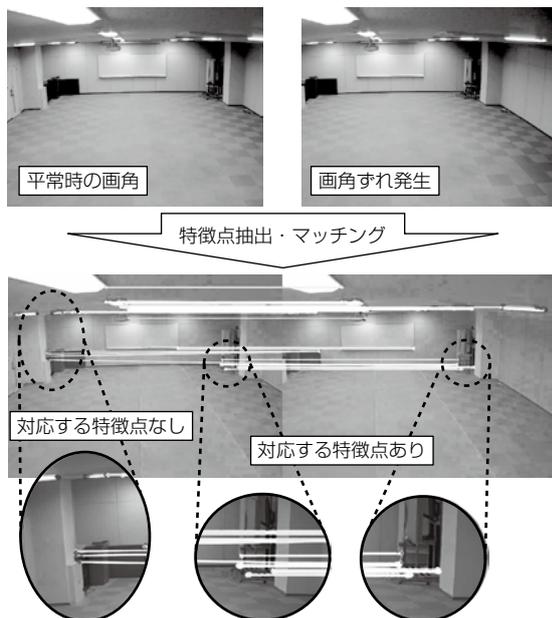


図6. 画像での特徴点抽出とマッチング

4. 検知モデル開発手法“kizkia Method”

多くのカメラ映像で安定した精度を確保する検知モデルの作成手法は、まだ確立されておらず、現在、各社の貴重なノウハウになっている。MDISは、数千台のカメラを対象にして、高い検知精度を安定的に実現するための学習モデル作成のノウハウを検知モデル開発手法“kizkia Method”として体系化し、継続的に整備を行ってきた。

このkizkia Methodの中から“学習するデータをどうやって収集するか”“どうすると検知精度が上がるのか”といった、検知モデルを開発する上でよく突き当たるであろう課題に対する解決方法の幾つかについて述べる。

4.1 学習精度と可視化

検知モデルの精度を向上させるためには、検知モデルがどのような処理をしているかを確認し、処理の内容を良い方向に修正することが効果的である。ディープラーニングの処理内容はブラックボックスと言われるが、近年ディープラーニングを説明するための研究が盛んに行われており、ある程度までは人間が理解できる形で可視化することが可能である。可視化技術はモデルの判定処理を逆算して、画像のどの部分の情報が最終的な計算結果に寄与しているかを数値化することで実現している。検知モデルの可視化技術は手法が幾つか存在するが、検知モデル(特にone-stage型)の可視化はあまり一般的ではない。検知枠を生成する処理の部分で情報の集約が行われており、逆算が困難なためである。MDISでは逆算時に検出結果による重みづけを行うことで集約部分の解釈を行い、これによって検知モデルの可視化を実現している。

図7はベビーカーを識別するモデルが、画像のどの部分に着目しているかを表している。グラフの縦軸はモデル精度、横軸は学習のステップ、一つの点がモデルの精度を表している。図ではモデルが“ベビーカー”と判定した画像に

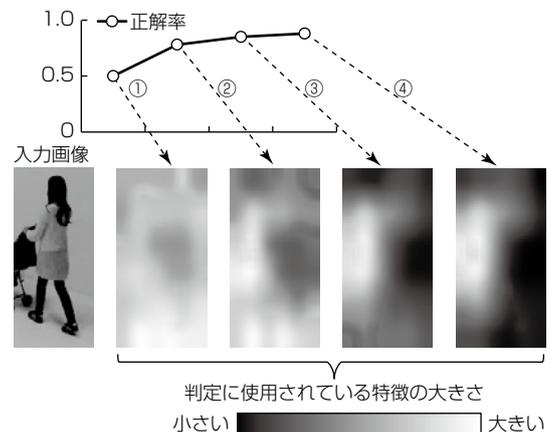


図7. 学習によるモデルの反応箇所の推移

対して、画像のどの部分から判定しているのかをヒートマップで表している。この図では、学習を進めるにしたがって、最初はモデルがベビーカーではないところを判定に用いていたのが、ベビーカーがある領域に強く反応するようになっていく。

可視化技術を精度向上に活用する例としてベビーカー検知モデルで述べる。例えばモデルがベビーカーの判定で幌(ほろ)(日よけ部分)を強く見ているとすると、そのモデルは幌のないタイプのベビーカーは見逃す可能性が高い。これを改善するために幌のないベビーカーの学習データを追加する、といったように精度向上を進めていく。

4.2 モデルの学習方法と評価方法

モデルをある目標精度に到達するまで学習・評価するに当たって重要なことは、学習データを作成する際にモデルがどのように処理しているかを理解して最も適切な形になるようにデータセットを構築することにある。例えば長い棒を持った人を識別するモデルを考えてみる。学習データに図8の(a)(b)のような、棒の見え方が一定でないものを混ぜて入れてしまうと、人の肩の上に棒状のものがあれば検出するモデルができる可能性があり、図8(c)のような、背景に棒状の模様がある場合に誤検出が多くなってしまふ。これを改善するためには(c)のような状況の学習データを大量に集めるか、又は(b)のような学習データを取り除いて学習させないという選択肢もある。後者を選択する場合は、可視化技術を活用して誤検知課題と類似する学習データを探し出して取り除くという手順で実現できる。モデル開発の現場ではこのような精度課題が頻繁に起きるので、課題ごとに学習データを集めることは現実的ではない。そのため、後者を選択することで達成可能な範囲でモデルの精度向上を進めることができる。

4.3 合成画像を活用した学習手法

モデル作成では大量の学習データが必要になるが、出現頻度が低い対象を検知したい場合、どのように学習データを収集するかが課題になる。実用可能な検知精度を達成するためには、実際にカメラを導入する環境で撮影した学

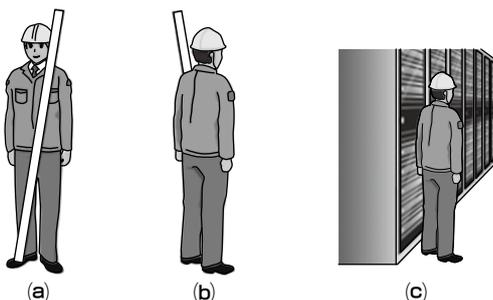


図8. 長い棒を持った人のデータ例



図9. 合成画像による学習のイメージ

習データを用いて学習する必要があることが分かっている。これまでは運用環境で白杖を所持したり車椅子に座ったりする演者(エキストラ)を撮影することで学習データを収集していたが、バリエーションを増やすために人員と時間を要しコストが増大していた。このような背景から、学習データ収集のコストを削減するために、学習データを現場の画像に合成する手法を検討した。

背景画像に切り出した人や物などの前景画像を張り付けるだけの単純な手法では、前景と背景の境界部に輪郭が現れてしまい、AIがこの輪郭部分を学習することで検知精度が悪化することが課題になった。そこで、輪郭部をなじませて違和感なく背景画像に合成することで、学習データの大部分を合成画像に置き換えて学習させても高い検知率を維持できる技術を開発した。

合成画像を活用した学習手法では、スタジオでのクロマキー撮影(背景が均質な色になる撮影)で作成した画像を図9のように現地の画像に合成するだけで学習データを用意できる。この手法の利点は、一度撮影したデータを流用できる上に、合成する座標や向きを変えることで学習データを大量に増やせることにあり、学習データ収集の低コスト化を実現した。

5. む す び

実用的なAIシステムを構築するための機能や検知モデルを開発するための手法について述べた。kizkiaでは、今後より多くの監視カメラを有機的に融合させ、カメラをまたいだ人や物体の高精度な追跡や検索をリアルタイムに実現するより実用的な映像監視ソリューションの実現を目指していく。また、社会インフラ設備の異常点検や製造ラインの外観検査など、適用範囲を拡大し、安心・安全・快適な公共施設や工場などの実現に引き続き貢献していく。

参 考 文 献

- (1) Matsumoto, W., et.al: A Deep Neural Network Architecture Using Dimensionality Reduction with Sparse Matrices, International Conference on Neural Information Processing, 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回全国大会講演論文集, 13~14 (2017)