

機械学習を活用した映像解析技術による 社会課題解決ソリューション

Solutions to Social Problems Leveraging Image Analysis Technology Based on Machine Learning

● 長部 大 ● 飯塚祐介 ● 東 明浩

あらまし

Deep Learningに代表される機械学習を映像解析に応用することが注目されている。筆者らは、交通・セキュリティ分野における都市が抱える社会問題を解決することを目的に、駐車場における満空検知ソリューションFUJITSU Technical Computing Solution GREENAGES Parking Analysis(以下、パーキングアナリシス)、および人物・車両を認識するソリューションFUJITSU Technical Computing Solution GREENAGES Citywide Surveillance(以下、シティワイドサーベイランス)を開発した。パーキングアナリシスでは、リアルタイムに駐車場の満空状況を検知することで、利用者は目的地周辺における駐車場の空き状況を把握することができ、駐車場運営側もデータ解析に基づいた駐車場の有効活用が可能になる。一方、シティワイドサーベイランスでは、リアルタイムに監視映像中の異常事象を自動検知することに加えて、大量なアーカイブ映像データから検索したい条件に関連する映像を精度良く抽出することが可能になる。これにより、監視や検索といった業務負荷の軽減を価値として提供する。

本稿では、提供するソリューションに適用した技術、インフラからサービスまで垂直統合したソリューション構成、および活用シーンについて述べる。

Abstract

An increasing amount of attention is being placed on the application of machine learning such as deep learning to image analysis. Aiming to provide a solution to social issues in urban areas, we focused on traffic and security and developed systems for parking analysis (detecting the availability of parking spaces) and citywide surveillance (human/vehicle recognition). The former is called FUJITSU Technical Computing Solution GREENAGES Parking Analysis (hereinafter "Parking Analysis"), and the latter is called FUJITSU Technical Computing Solution GREENAGES Citywide Surveillance (hereinafter "Citywide Surveillance"). Parking Analysis detects vacant parking spaces, providing users with real-time information on available parking spaces in their destination areas. The data also enable owners of parking lots to enhance operational efficiency based on data analysis. Citywide Surveillance automatically detects unusual events by monitoring images in real time. In addition, the image archive can be searched to accurately extract images related to given keywords. This functionality serves as an added value for reducing the workload required for surveillance and investigation. This paper explains the technology applied in these solutions, and describes their vertically integrated configuration including infrastructure and services, as well as the contexts in which they are leveraged.

ま え が き

IoT (Internet of Things) により、様々なデバイスから取得したデータを活用することで、生活レベルの向上が期待される。特に、都市サービスを効率的に運営するスマートシティが注目されている。スマートシティ市場は2020年までに、1.6兆ドルにも上るビジネス規模が期待される中、セキュリティ (13.5%)、交通 (8.7%) 分野が注目を浴びている。⁽¹⁾

セキュリティ・交通分野で活用されるIoTデバイスと言えば、監視カメラである。都市への人口集中に起因する交通問題や住民生活の安全への関心の高まり、更には昨今のネットワークカメラの低コスト化などにより、カメラ市場規模は拡大傾向にある。世界の監視カメラ出荷台数は2018年では4,320万台にもなると予測されている。⁽²⁾

近年、VCA (Video Content Analytics) 技術の進展により、映像解析を異常事態検知や犯罪対策で活用することが期待されている。このように、監視カメラの普及とその映像データを利活用することにより、交通・セキュリティ分野におけるスマート化に向けた新たな価値創出の取り組みが広がっている。

更に、2012年にDeep Learningの出現により脚光を浴びている機械学習を画像解析に応用することで、今までにない映像データの活用が可能になった。Deep Learningとは、脳内の神経回路網とそのプロセスをコンピュータ上で模倣した「ニューラルネットワーク」という機械学習ロジックであり、これを多層にして組み合わせたものである。⁽³⁾ 富士通では、中国語の手書き帳票の処理効率化にDeep Learningを活用して、96.7%の認識率を達成している。⁽⁴⁾ 筆者らは、このDeep Learningを含む機械学習の適用範囲を、画像や音声といった文字以外のメディア認識への拡大を図っている。更に、統計的学習手法Boostingの一つであるAdaBoostも加えて、監視カメラで取得した映像の解析を統合ソリューションサービスとして構築することを目指している。

機械学習のアルゴリズムは多数存在するが、今回対象としているのは教師あり学習と呼ばれているアルゴリズムであり、以下の二つのプロセスに

大別される。

(1) 学習フェーズ

識別フェーズで「教師」として使用される「学習済みモデル」を作成する。ここでは、多数の学習サンプルが必要になる。

(2) 識別 (検知・分類) フェーズ

作成した「学習済みモデル」を識別機能として、監視カメラからの映像を自動識別する。

本稿では、映像解析における機械学習の活用事例として、AdaBoostによる駐車場の満空状態の自動検知、Deep Learningによる人物・車両の自動分類の事例を紹介する。また、これら技術を活用した実際のソリューションFUJITSU Technical Computing Solution GREENAGES Parking Analysis (以下、パーキングアナリシス)、およびFUJITSU Technical Computing Solution GREENAGES Citywide Surveillance (以下、シティワイドサーベイランス) を紹介する。

AdaBoostによる駐車場満空検知

都市部、特に移動手段が自動車に依存している新興国の都市においては、慢性的に駐車場が不足している状態であり、空き駐車場を探す車両が増えている。このことが、交通渋滞を引き起こす原因 (都市部で発生する交通渋滞の約3割) となっている。⁽⁵⁾ ドライバーに目的地周辺の駐車場状態 (空き情報) をリアルタイムに提供することができれば、駐車場を探す時間や無駄な走行の削減、更には駐車場起因の渋滞緩和も期待できる。筆者らは、汎用的なネットワークカメラで撮影した映像を機械学習で解析することにより、駐車場の満空状態をリアルタイムに検知することを実践した。⁽⁶⁾

駐車場の満空状態の自動検知は、古くから取り組まれており、超音波センサーといったデバイスが活用されてきた。近年、カメラで撮影された静止画や動画の映像解析処理も実現されているが、依然として課題や制約が残っている。例えば、カメラを用いた方式では、1台のカメラでモニタリング可能なスロット (駐車スペース) 数が2～4に限られる。このほかに、車両がカメラに正対する必要がある、初期データとして全スロットが空車の状態を作る必要がある、といった課題が残っている。

筆者らは、1台のカメラで多くのスロットをモニタリングすること、および車両がカメラに正対するといった制限を外すこと（縦・横・斜めに対応）を目的としたソフトウェアを開発した。AdaBoost方式を用いた駐車場満空検知の処理フローを図-1に示す。

(1) リアルタイム映像収集

汎用的なネットワークカメラ（解像度1080p、フレームレート10 fps）を使用して映像を取得する。取得した映像はサーバ側に転送され、サーバ上でリアルタイムに解析を実施する。

(2) 映像前処理

夜間や悪天候時の明るさ対策として、暗所補正、ノイズ除去、および輪郭強調を実行する。

(3) 状態変化検知

各スロットで変化が発生しているか否かを判断する。これにより、初期処理で満空を判断したあと変化がなければ、次の満空判定をスキップすることが可能になり、これは処理負荷軽減につながる。

(4) 満空判定

満空判定する識別機能は事前に学習フェーズにて準備した。ここでの学習フェーズとは、収集した学習サンプル（ポジティブサンプル：車両あり、ネガティブサンプル：車両なし）の特徴量を認識させることにより、学習済みモデルを作成することである。⁽⁷⁾ 学習サンプルとして、車両の正面に加えて背面・側面・斜めの画像を収集する。作成された学習済みモデルが識別機能となり、実際の駐車場を撮影した映像の各スロットに対して車両の

有無を判定する。

実際の駐車場において、1台のカメラで数十スロットをモニタリングするためには、カメラの設置は屋外に限られ、高い位置に設置する必要がある。専用に高いポールを設置することはコストの増大につながるため、隣接するビル壁面や屋上などに設置することが望ましい。筆者らは、20 m程度の高さにカメラを設置した場合、1台のカメラで80～100スロットのモニタリングを可能にした。

Deep Learningによる車両・人物の自動分類

Deep Learningの進展により、2020年までには静止画像・動画からの一般物体認識が人間レベルに到達すると言われている。⁽⁸⁾ これによりセキュリティ分野では、従来大型マルチスクリーン上に多数のカメラ映像を並べて人手で映像を監視していたことが、無人状態でリアルタイムかつ高精度に異常事態の検知が可能になる。更にアーカイブ映像から目的とするオブジェクトを高精度・高速に検索することも可能になる。また交通分野においては、物体認識技術によって取得したデータが、あらかじめ定められたルールの範囲内であることを監視し、交通違反や事故を削減することが期待されている。

筆者らは、監視カメラから取得した車両・人物などの映像データをDeep Learningで解析し、セキュリティ監視、交通量把握などへの活用を図った。Deep Learning方式を用いた処理フローを図-2に示す。



図-1 AdaBoost方式による駐車場満空検知処理フロー



図-2 Deep Learning方式による車両・人物分類処理フロー

(1) リアルタイム映像収集

前章の駐車場満空検知と同様に、汎用的なネットワークカメラ（解像度1080p, フレームレート5 fps）を使用して映像を取得する。取得した映像はサーバ側に転送され、サーバ上でリアルタイムに解析を行う。

(2) 映像前処理

屋内、夜間、悪天候時の明るさ対策として、暗所補正を実行する。

(3) 動体検知～車両・人物分類

映像フレーム内で動体を検知し、普通車（car）、バス（bus）、ワゴン（van）、トラック（truck）、オートバイ（motorbike）、人（person）に分類する。更に、車両については色の判別も行う。ここで、動体検知から車両や人物の分類を実現するために、学習フェーズにおいてFaster R-CNN⁽⁹⁾をベースにした学習済みモデルを独自に開発した。作成された学習済みモデルが識別機能となり、車載カメラや交差点に設置された固定カメラからの映像に対して、検知・分類を行うことが可能になる（図-3）。

(4) 自動追尾

各映像フレームの検知・分類をDeep Learningで行うと処理負荷が増大する。処理負荷を軽減するため、全フレームを処理するのではなく、一定間隔ごとのフレームのみに対してDeep Learningを実施し、検知したオブジェクトを次映像フレーム以降にわたって追尾を行う。自動追尾の組み合わせにより、効率的に分類された情報については、映像に対するタグとして保持する。



図-3 車載カメラ(上段)、固定カメラ(下段)で取得した映像の車両・人物分類

ソリューションの型化と提供価値

前述したAdaBoost方式による駐車場満空検知とDeep Learning方式による車両や人物の自動分類を実現するために、他社技術とOSS（Open Source Software）を組み合わせたオープンプラットフォームをコンセプトとして、インフラからサービスまでを垂直統合したソリューションを提供する。そのアーキテクチャーの概要を図-4に示す⁽¹⁰⁾。

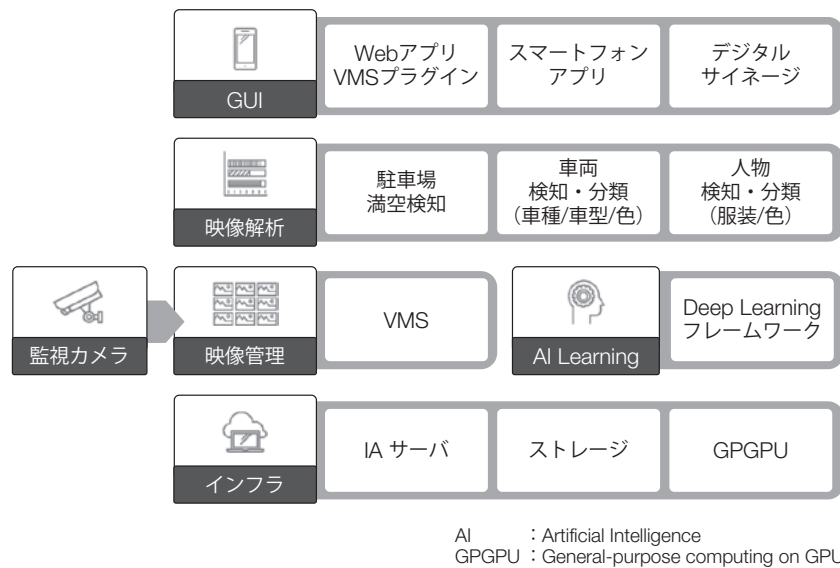


図-4 ソリューションアーキテクチャ概要

● ソリューション構成

(1) インフラストラクチャー

映像管理および解析用のサーバ・ストレージに加え、Deep Learning環境としてGPU(Graphics Processing Unit)をサーバに搭載する。GPUについては、NVIDIA社のTITAN X、Tesla K40/K80など、GPUあたり12 Gバイトのメモリを搭載したグラフィックボードを搭載する。また、Deep Learningフレームワークについては、Caffe・Chainer・Torch・Theano・Tensor-Flowといった数多くのOSS⁽¹¹⁾が存在し、日進月歩のスピードで技術が進化している。その中で画像処理に最適化され、畳み込みニューラルネットワーク「CNN(Convolutional Neural Net)」, GPU, Python言語・C++言語をサポートしているなどの理由からCaffe⁽¹²⁾を採用している。

(2) VMS (Video Management System)

映像解析とともにキーとなるビデオ管理システム(VMS)には、デファクトスタンダードであるパッケージを採用する。本VMSは、120社・4,000種類を超える監視カメラに対応し、ソリューションを展開する際のカメラ依存性をなくし、顧客が既に設置済みのカメラの有効活用を可能とする。

(3) GUI (Graphic User Interface)

解析結果を顧客に提供するGUIでは、ドライバーや警備員などの現場向けにスマートフォンアプリ

やデジタルサイネージを提供する。また、中央監視センターでの運用を想定し、管理者向けにWebアプリやVMSへのプラグインGUIを提供する。

● ユーザーへの提供価値

(1) パーキングアナリシス

リアルタイムに検知した駐車場の満空状況を、スマートフォンアプリやデジタルサイネージで可視化する。これにより、イベント会場などの目的地周辺における駐車場の空き状況を把握することが可能になり、駐車場サービスの利便性が高まることで、交通渋滞の軽減も期待できる。

また、駐車場運営側にとっては、需要に基づく最適プランニングによる駐車場の利用率やアセットの最大化も重要である。本ソリューションでは、駐車場の利用時刻と滞留時間に関する週次や日次、更にはイベント開催時などのタイムトレンド情報(稼働率変動の統計)、駅や基幹道路との位置関係などを提供する。これらの情報に基づき運営者は、例えば駅や会場に近い駐車場や大型駐車場の駐車ゾーンごとの料金設定を検討できる。加えて、予約システムやゲート管理システムと連携させることで、VIP向けの事前予約やバレーパーキングなどのサービス提供も可能となる。

(2) シティワイドサーベイランス

顧客の監視業務や搜索業務などの負荷軽減を価値として提供する。監視業務においては、多くの

監視カメラが街中や施設内に設置されているが、それらの映像をリアルタイムにオペレーターが目視監視するには限界がある。本ソリューションでは、あらかじめ指定した重点監視イベントを監視カメラ映像からリアルタイムに検知し、オペレーターに通知する。例えば一定時間以上ある場所に滞留している車や人を不審な事象として検知することや、車や人の混在度合いから重点監視エリアを識別するといった使い方が想定できる。更には、検知・分類に用いるDeep Learningモデルを追加学習することで、より特徴的な監視に対応することが可能であり、顧客ニーズに合わせて柔軟にユースケースを拡充できる。

搜索業務においては、ナンバープレートや顔画像など個を特定する情報がないことも多く、通報された限られた情報から人手で候補を探す作業は膨大な労力を必要とする。本ソリューションでは、車型・車種や色、人物の服装や色といった特徴により、大量なアーカイブ映像データから対象を絞り込むことで、搜索にかかる負荷の軽減に大きく貢献する。

筆者らは、ソリューション名称にパーキングとサーベイランスという言葉を用いているが、様々な分野に適用が可能だと考えている。例えば、警察や警備会社などのセキュリティ分野、空港や鉄道などの交通分野、工業団地や展示場などの施設管理分野、ガソリンスタンドなどのリテール分野、公的機関による都市計画など、多岐にわたる分野への適用が提案できる。

今後の方向性

本章では、映像解析における機械学習活用の今後の方向性について述べる。

(1) 解析機能の更なる拡充

映像解析に機械学習を用いる利点は、学習サンプルを増やすことで既存識別（検知・分類）機能の精度向上が可能になる点や、同じ映像を使って新たに学習済みモデル・識別機能を追加しやすい点である。筆者らは学習フェーズをサービス（学習サービスと呼ぶ）として顧客に提供することを目指している。学習サービスにおいては、顧客視点で課題を深耕し、検知したい行動の内容・検知基準・目標精度などの要件を定義することが重要

である。また学習サンプルは、通常、人手による画像データへのラベリングにより作成している。将来的には、シミュレーションを用いたデータ生成までをソリューションスコープに含めるべきと考える。学習済みモデルの実装、目標精度の達成に向けたチューニングとその評価には、高い専門性が必要である。

(2) データ伝送負荷の低減

映像解析の精度は、解像度やフレームレートなどの映像品質に依存する。高品質な映像はデータ量が大きく、映像を伝送するネットワーク帯域の確保が課題となる。これに対しては、監視カメラにより近い場所でデータ処理用のデバイスを配置し、1次解析を行うことで伝送するデータ量を抑える「エッジコンピューティング」が有効と考える。

(3) クラウドによる学習サービス提供

Deep Learningの場合、前述の学習サービスでは、多くの計算リソースを要する。顧客ニーズに合わせて柔軟、かつ迅速に解析機能を拡充するために、GPGPUを含むHPC（High Performance Computing）クラウドを学習用基盤として利用していく。また、エッジコンピューティングにより、インターネット回線を用いたエッジデバイスへのデータ伝送も可能となる。一連のソリューション構成をクラウドサービスとして提供することで、顧客へのシステム導入を容易にすることが今後の重要施策と考える。

む す び

本稿では、映像解析における機械学習の事例として、駐車場の満空検知と車両・人物の自動分類の技術とソリューションについて述べた。富士通は、Deep Learningを始めとする機械学習を中核としたインフラからサービスまでを垂直統合したソリューションを提供することにより、スマートシティ市場と機械学習・Deep Learning市場が合わさる領域を狙い、交通・セキュリティ分野における社会課題解決に取り組んでいく。

参考文献

- (1) FROST and SULLIVAN : Strategic Opportunity Analysis of the Global Smart City Market, Smart

City Market is Likely to be Worth a Cumulative \$1.565 Trillion by 2020.

<http://www.egr.msu.edu/~aesc310-web/resources/SmartCities/Smart%20City%20Market%20Report%202.pdf>

- (2) 矢野経済研究所：監視カメラ世界市場に関する調査結果2015.

<http://www.yano.co.jp/press/pdf/1420.pdf>

- (3) Atmarkit：Deep Learning超入門.

<http://www.atmarkit.co.jp/ait/articles/1510/27/news008.html>

- (4) 富士通：FUJITSU JOURNAL, 人工知能の能力が人間を超えた!? 世界初、中国語の手書き文字認識率で96.7%を達成.

<http://journal.jp.fujitsu.com/2015/10/02/01/>

- (5) Cisco：Smart + Connected City Parking.

http://www.cisco.com/c/dam/en_us/solutions/industries/docs/parking_aag_final.pdf

- (6) Fujitsu Research & Development Center：Vision-based Parking Lots Management Technology for Smart Parking.

<http://www.fujitsu.com/cn/en/about/resources/news/press-releases/2015/frdc-0818.html>

- (7) R. Fusek et al.：AdaBoost for Parking Lot Occupation Detection. Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013 (Advances in Intelligent Systems and Computing .226), p.681-690, 226 (2013).

- (8) 新エネルギー・産業技術総合開発機構 次世代人工知能技術社会実装ビジョン作成検討会資料：次世代人工知能技術社会実装ビジョン.

<http://www.nedo.go.jp/content/100782828.pdf>

- (9) S. Ren et al.：Faster R-CNN：Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems 28 (NIPS 2015).

- (10) 富士通：AI技術を活用した都市監視や駐車管理のソリューションを販売開始.

<http://pr.fujitsu.com/jp/news/2016/10/3-1.html>

- (11) 村上真奈：ディープラーニング最新動向と技術情報.
<https://images.nvidia.com/content/APAC/events/deep-learning-day-2016-jp/NVIDIA-DeepLearning-Intro.pdf>

- (12) Y. Jia et al.：Caffe：Convolutional Architecture for Fast Feature Embedding. MM '14 Proceedings of the 22nd ACM international conference on Multimedia, p.675-678 (2014).

著者紹介



長部 大 (おさべ だい)

テクニカルコンピューティング・ソリューション事業本部
TCフロンティアセンター
社会課題解決に向けたソリューションの企画・開発に従事。



飯塚 祐介 (いづか ゆうすけ)

テクニカルコンピューティング・ソリューション事業本部
TCフロンティアセンター
社会課題解決に向けたソリューションの企画・拡販に従事。



東 明浩 (ひがし あきひろ)

富士通研究開発中心有限公司
メディアプラットフォーム技術研究部
Deep Learningなど機械学習を活用した映像解析技術の研究開発に従事。