

5Gを使った画像AIエッジコンピューティング構想

5G Edge Computing for Imaging AI

細野 真央* 村上 昌弘** 濱 健朗*
Masao HOSONO Masahiro MURAKAMI Kenro HAMA

要旨

コニカミノルタは、画像IoT技術を用いて問題の解決や改善を行う分析ソリューションの提供を目指している。

我々は、クラウドとエッジIoTデバイス及びセンサーから成るエッジコンピューティング技術を導入し、より高度なソリューションを実現させようとしている。

近年、高度で多彩なデータ収集が可能になるにつれて、よりセンサーに近いところで高速にAI処理や判断する分散化処理が求められるようになってきた。5Gは大容量、多接続、低遅延の特徴を持っており、5GとエッジIoTデバイスを組み合わせることが分散化の実現に必要な要素である。また、画像IoTをより進化させるためにはエッジでの処理性能を向上させる必要がある。

本稿では、これからのAI機能を持ったエッジIoTデバイスに必要な高速な映像入力インターフェース、低遅延スループット、画像処理、AI機能、高速な通信機能、高度なセキュリティを持ったハードウェアと、柔軟にデバイス間の連携が可能でハードウェアの性能を活かしきるソフトウェアについて紹介する。

画像IoT技術とエッジコンピューティングと5Gを組み合わせることでオフィスや工場、医療、監視用途において高画質な画像を用いた高精度の分析が行えるだけでなく、様々なセンサーを使って複合的に分析が行えるようになり、これまでないリアルタイムで精度の高い分析や提案が行えるようになる。

今後は、「オープンラボ」環境を構築することで、開発パートナーとの共創を一層深め、新たな製品やサービスを生み出し、社会価値を創造していく。

Abstract

Konica Minolta is committed to provide analytics solutions to solve customer needs by utilizing imaging IoT technologies. The solutions will be even more advanced by implementing edge computing consists of cloud, edge IoT devices, and sensors.

In recent years, as collecting varied and high-quality data becoming possible, distributed processing is required which achieve faster decision making, faster AI acceleration, near sensors. To implement distributed computing, it is necessary to combine 5G networks and edge IoT devices, as 5G has the characteristics of high capacity, multiple connection, and low latency. Moreover, improving processing speed is necessary for advanced imaging IoT.

In this paper, we introduce the hardware with required components for future AI based edge IoT devices, such as, high-speed video input interface, low latency throughput, image processing, AI accelerator, high-speed communication function, and advanced security. We also refer to the software which enables flexible inter-device integration and takes advantage of the hardware performance.

By combining imaging IoT, edge computing, and 5G, not only the high accuracy analysis using high quality images, but also the complex analysis using various sensors, will be possible at office, factory, medical and surveillance field. Consequently, we can perform accurate analysis and make recommendation in real time than ever before.

Konica Minolta will continue contribution to society by offering new products and solutions collaborating with partners in our brand new "open-lab".

* IoTサービスPF開発統括部 エッジコントローラ開発部 第1グループ
** IoTサービスPF開発統括部 エッジコントローラ開発部

1 はじめに

近年、様々な分野で各種センサーを利用してデータの分析を行い、問題の解決や改善を行う分析ソリューションが発展している。これはディープラーニング (DL) をはじめとした人工知能 (AI) の発展が大きい。コニカミノルタでも、デジタルマニュファクチャリングや介護、監視カメラ事業においてAIを活用した分析、改善ソリューションを提供している。

こうした分析ソリューションにおいてはこれまでクラウドを用いたクラウドコンピューティングが主流であった。データの分析を行うためには強力なマシンパワーが必要であり、サーバーの強力な演算能力やデータサーバー機能を用いて分析を行う必要があったためである。

しかしながら、IoT (モノのインターネット) が発達し、高度で多彩なデータ収集が可能になるにつれて、よりエッジ (末端) IoTデバイスに近いところで高速にAI処理や判断することが望まれるようになってきた。そこで「エッジコンピューティング」という考えが提唱され、エッジコンピューティングを完成させるための通信技術として、「5Gネットワーク」が誕生した。

コニカミノルタが進める画像IoT技術にとっても、エッジコンピューティングと5Gネットワーク技術を取り入れることによって画像センシング技術を活かしたより高速で高機能なAI分析を提供することができると考える。

2 エッジコンピューティングとは

エッジコンピューティングとは、これまでクラウドで行っていた処理の一部、又は全部をユーザー環境に近い位置に配置したエッジサーバーで行うことである。これにより、大量のデータをクラウドサーバーへ送る必要がなくなり、処理負荷やネットワークトラフィックの増大を防ぎ、より高速に処理を行うことができるようになる (Fig. 1)。

より高度で高速な画像IoTを実現するためにはエッジコンピューティングを採用すべきと考え、さらには、データ蓄積やAI分析処理をデバイス/エッジサーバー/クラウドサーバーに適切に配置、連携することが必要と考えている。具体的にはエッジIoTデバイスを単体推論層、エッジサーバーを複合推論層、クラウドサーバーを分析/学習層と分類することで、より早く、より正確な推論を行うことを考えている。

この構成を実現するためには、エッジIoTデバイスと通信の高速化が欠かせない。処理の高速化はもちろん、エッジサーバーとの高速、低遅延な通信が必要で、これは5G通信を利用することで達成することができるため、5Gに最適化されたエッジIoTデバイスを開発することが必要となってくる。

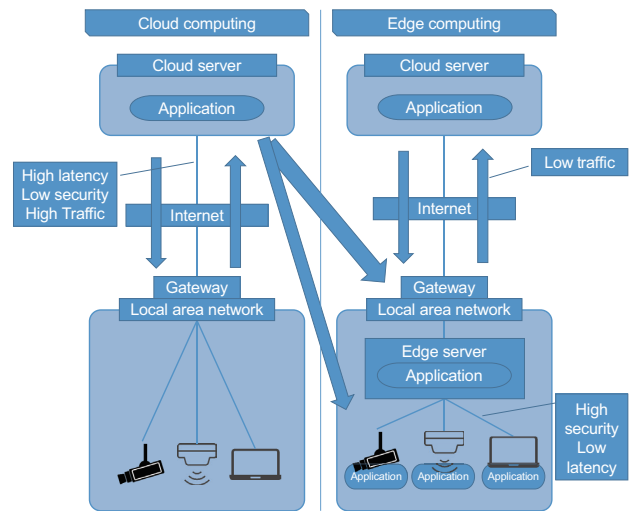


Fig. 1 Difference between cloud computing and edge computing

Edge computing is to process a part of, or all computation with edge servers located close to customers, which used to be done by cloud servers.

3 エッジIoTデバイス開発について

エッジIoTデバイスとして必要な機能としては、以下が挙げられる (Fig. 2)。

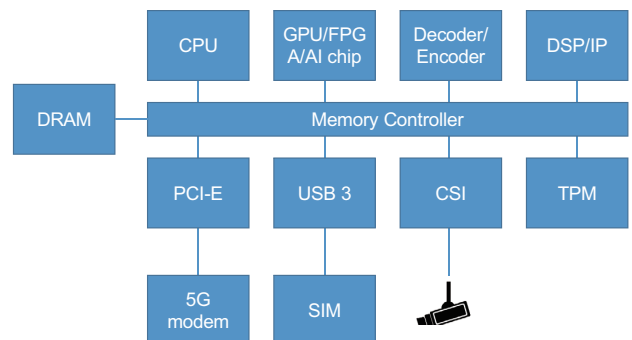


Fig. 2 Configuration of edge IoT device

Functions required for edge IoT device are listed below.

- High resolution image input (CSI)
- High speed image processing (DSP/IP, Decoder/Encoder)
- AI accelerator (GPU/FPGA/AI chip)
- Low latency, and high capacity communication using 5G (5G modem)
- Security (SIM)

3.1 高解像度の映像入力

高解像度な映像入力インターフェースとその映像を蓄積する高転送レートメモリ、その転送経路である内部バス構造でQoS (Quality of Service) サポートが必要になる。カメラからの入力は低遅延で高速なMIPI CSI-2を採用している。USBインターフェースと比較すると、低レイテンシ且つ高速なデータ転送が可能である。また、カメラからの入力はより低遅延で画像転送を行うためにエンコードを行わない非圧縮の画像転送をサポートする。4K非圧縮データ6Gbpsを処理するためには、38Gbpsの帯域を持つDDR4メモリを中心としたシステム設計を行うことで、高速な映像入力を可能としている。

3.2 高速な画像処理

入力した画像をAI処理するために画像サイズの変更や、エッジの強調や明るさ調整などの画像処理が必要となる。より高速で低電力な画像処理を行うために専用の画像処理IP (Intellectual Property) やDSP (Digital Signal Processor) を使用する。また、映像をクライアントアプリに送信する場合は、その要件に合わせてハードウェアまたはソフトウェアの動画エンコーダを搭載し、データを圧縮して送信する動画データの転送レートを抑えることでネットワークの負荷を低減させる。

3.3 高速なAI機能

リアルタイムな判断を行うためには、高速なAI処理が欠かせない。エッジIoTデバイス上で最小限の電力で推論を行うために、用途に応じてAIチップやFPGA (Field Programmable Gate Array), GPU (Graphics Processing Unit) を使い分けることでデバイスの最適化をはかることが重要であり、またそれにより幅広い用途に活用できるようになる。最適化を行うための選択肢の幅を広げるために、PCI-E/USB 3.0/SOM (System On Module) の各インターフェースを搭載する。

FPGA活用では、FPGAに一度実装した機能だけが実行されるのではなく、システム停止することなく必要に応じて学習モデルの更新や違った機能に更新する機能をサポートすることでデバイスの機能拡張性を持たせている。

3.4 5G通信を利用した低遅延、大容量通信

5Gはレイテンシ1ms以下を保障しており、エッジコンピューティングに5Gを導入することで無線通信でありながらリアルタイム性を実現することができるが、システムとして低レイテンシを保障するためにはデバイス側での遅延をできる限り少なくする必要がある。そのために低遅延通信プロトコルや大容量の packets 送信技術をエッジIoTデバイスに搭載する開発を進めている。また、5Gの特徴としては大容量、低遅延、多接続を同時に実現できるわけではなく、様々な環境に適應できるセッティングを見つけ出していく必要がある。もう一つの5Gの特徴であるネットワークスライシング技術は、5Gネットワークを分離して最適な設定で動作させることを可能とする技術で、大容量の動画データと低遅延の指示を同時に運用することが可能になる。これらの技術を適用することでエッジコンピューティングを無線ネットワークで実現することが可能となる。しかしながら5Gネットワークの細かなセッティングはローカル5Gでこそ可能なものであり、キャリア5Gを利用した場合の影響は検証していく必要がある。

3.5 セキュリティ

IoTデバイスについては外部からのファームウェアの改ざんや不正アクセスによって攻撃の際の踏み台にされたり、通信データの盗聴による情報が漏洩するといった

問題が発生しておりセキュリティの重要性が高まってきている。エッジIoTデバイスを攻撃から守るため、まずハードウェアには外部からの攻撃を防ぐためにセキュリティチップを搭載し、ファームウェアの不正な書き換えを防止する。ソフトウェアに関しては、通信の安全性を保障するために、最新の暗号化技術を搭載するのはもちろんのこと、5Gが持つSIM (Subscriber Identity Module) をベースとした強固なセキュリティを利用することでなりすましや盗聴を防ぐこともできる。ハードウェア、ソフトウェアのセキュリティと合わせてシステムを構成することで弱点のないセキュリティを提供できるようになると考える。

3.6 入力から出力までをコントロール可能なソフトウェア

エッジIoTデバイスに搭載するソフトウェアは、ハードウェアの性能を十分に発揮できる必要がある。様々な業種、シーンに対応するために、各種センサーを大きな変更なく取り込める抽象化されたインターフェースを設計することで、デバイスに取り付けられたセンサーに対応することを可能としている (Fig. 3)。

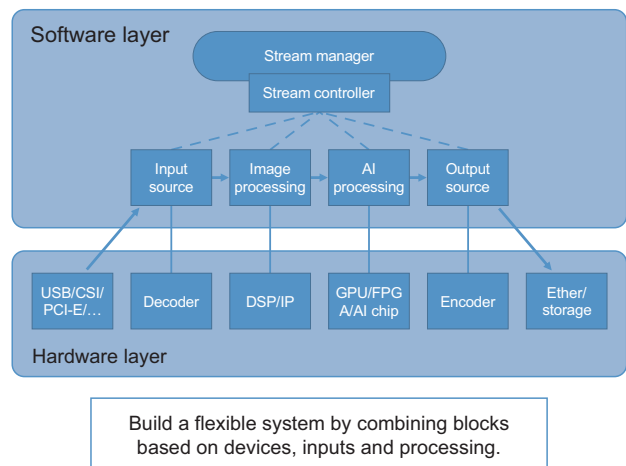


Fig. 3 Configuration of edge IoT device software.

Abstract interface design that imports sensors of every kind without a big change enables to use sensors attached to devices in varied industry and scenes.

4Kや8Kなど、今後より高画質になるカメラの画像をまとめて処理を行うために、複数のセンサーデータの同時入力を可能とし、マルチコアプロセッサ、画像処理チップを有効に利用することで遅延なく処理を行えるようにする。

処理負荷の偏りやパフォーマンスを最大限発揮するため、ストリーム処理の途中で他エッジIoTデバイスへ送信し結果を受け取ることで複数のデバイス間で協調して動作を行う分散AI処理を実現する (Fig. 4)。分散処理やデータ転送を安全に行うために、更新可能な暗号化ライブラリによる通信を可能にする。AIモデルや各種ソフトウェア更新をシステム停止することなく更新することを可能にするUpdaterを搭載する。

このプラットフォームを搭載することで、規模や要求性能に応じたエッジコンピューティングを実現する。

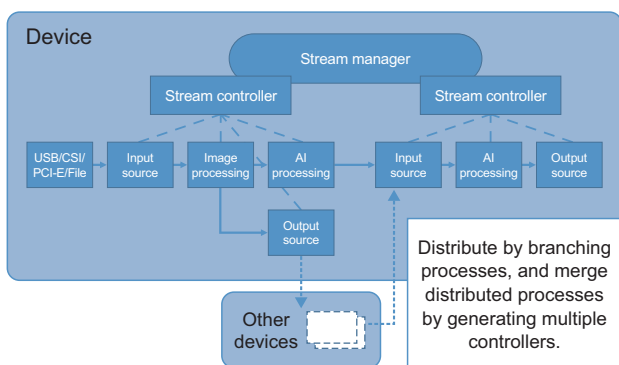


Fig. 4 Outline of cooperative processing.

Distributed AI which cooperatively process among multi devices by sending computation to other edge IoT devices during stream processing and receive the output, can resolve the unevenness of processing load and maximize the performance.

4 5Gエッジコンピューティング適用による価値

エッジコンピューティングと5Gの組み合わせはこれまで適用が難しかったリアルタイム性、信頼性が重要な顧客に対して画像IoT+AIサービスを提供することができる価値があると考えている。今後適用が期待される業種としては以下が考えられる。

4.1 オフィス用途への適用

これまでオフィスにはあまりセンサー技術は導入されなかったが、5Gがオフィスに浸透することで、オフィスの計測も個人へと広がっていくと考えられる。個人の健康やメンタルをチェックし、状態に応じたケアが行われるようになっていく。あらゆる機器の故障予測が行われるだけでなく、オフィスの管理も高度化し、室温や照度だけでなく人の密集や施設の利用率のような様々なデータから快適なオフィスが提供されるようになる (Fig. 5)。

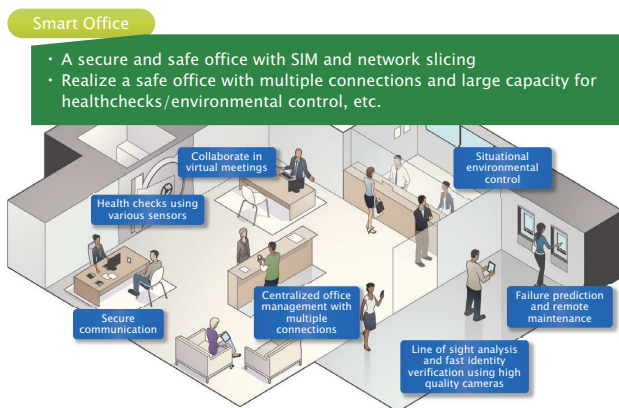


Fig. 5 Edge computing for office.

Comfortable office becomes possible by predicted maintenance of any devices along with advanced office management from the data of room temperatures, illuminance, and varied data such as crowd density, utilization rate.

4.2 工場機器監視用途への適用

生産現場では稼働している多数の機器を監視し、異常があった場合は速やかに機器を停止する必要がある。このような場合は5Gエッジコンピューティングを適用することでリアルタイムな監視が可能となる。また、工場のレイアウト変更にも容易に対応が可能である (Fig. 6)。

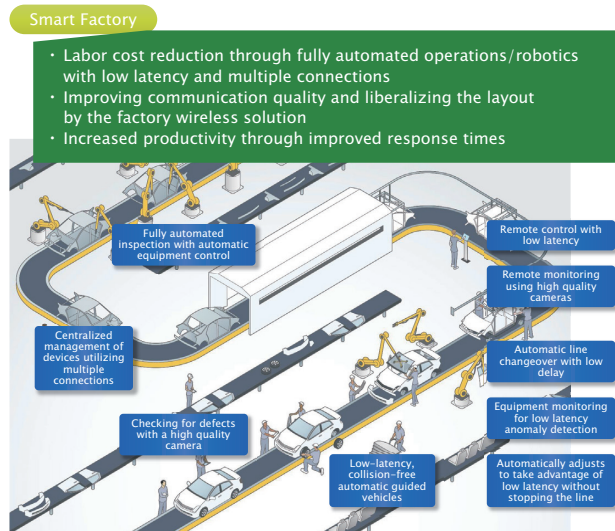


Fig. 6 Edge computing for factory.

Real time monitoring which can easily change layouts becomes possible by 5G edge computing.

4.3 医療、看護用途への適用

医療、看護用途へは異常を発見した場合に確実に通知される高信頼性、異常を即座に通知できるリアルタイム性が求められており、エッジコンピューティングの導入により、患者や機器のモニタリング用途への適用が期待される (Fig. 7)。

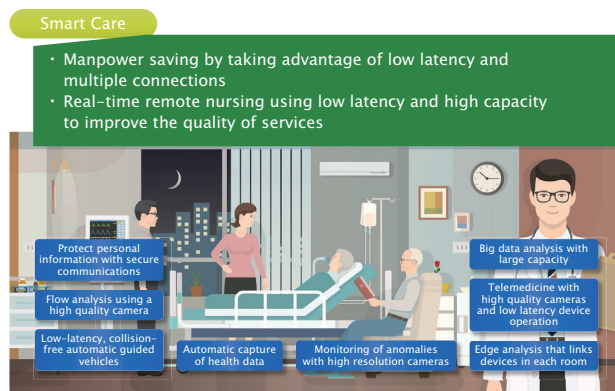


Fig. 7 Edge computing for healthcare.

Monitoring patients and devices by utilizing edge computing is expected, as reliable and real time emergency alerts have been required for medical use.

4.4 監視用途への適用

監視カメラ単体での異常検知が可能になることで、異常発生時に即座に近くの警備員への通報が可能となる (Fig. 8)。

Surveillance Camera

- Individual cameras react in real time to anomalies, enabling notification and building control
- Multiple cameras work together to track a suspicious person

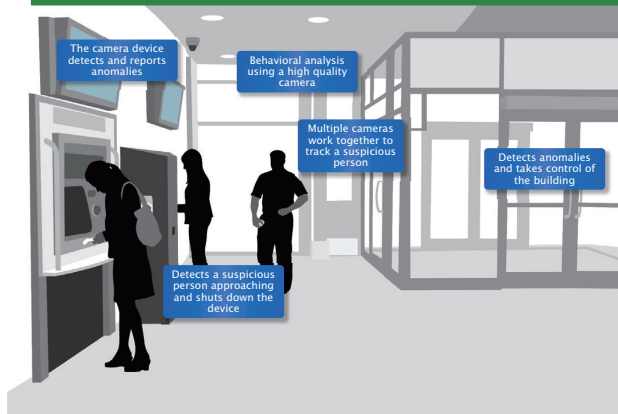


Fig. 8 Edge computing for surveillance application.

Immediate alert to the security guards near by become possible by detecting emergency using only a single surveillance camera.

5 最後に

コニカミノルタはInnovation Garden OSAKA Centerにローカル5G, キャリア5Gの設備を導入し, 日本でも稀な「ハイブリッドの5Gオープンラボ」としての開発環境を備え, パートナーとの共創を一層深め, 新たな製品やサービスを生み出し, 社会価値を創造していく。同時に画像IoTとエッジコンピューティング, 5Gを組み合わせたオープンイノベーションを推進し, 新たな価値創造の実現に向けて取り組んでいく。