

3 ① ホワイトボックスサーバ技術

データ処理・演算の効率を徹底的に高め、 電力消費の抜本的な削減を目指す

カメラやセンサーからのデータを分析し価値・サービスを提供するようなサイバーフィジカルシステム（以下、CPS）の活用が進んでいる。データ処理の負荷増加に伴う電力消費の増加が今後重要な問題になることが予想される。NTTソフトウェアイノベーションセンタ（以下、SIC）が研究開発に取り組む「ホワイトボックスサーバ技術」は、このような問題の解決策として期待される。

大量のデータ処理・演算に伴う 電力消費の増大が課題に

IOWN（Innovative Optical and Wireless Network）構想ではさまざまなサービスが検討されている。その1つがスマートシティにおいてCPSを活用するサービスだ。たとえば、CPSでは多数のカメラからの映像をリアルタイムで分析する。こうしたユースケースでは今後カメラの台数が増加し、映像の高解像度化も進む。そのため映像データを処理するCPSの負荷は高まる一方であり、電力消費の抑制が重要な課題となっている。

カメラ映像の解析に用いられることが多いAI推論などの演算にはGPU（Graphics Processing Unit）が適しており、CPU（Central Processing Unit）で演算するより高速に演算できる。また演算によってはGPU以外のアクセラレータが適している場合もある。例えば、映像のフィルタリングやリサイズなどの演算はFPGA（Field Programmable Gate

Array）が適している。用途に応じて最適なプロセッサやアクセラレータを組み合わせることが重要だ。

しかし既存のコンピューティング基盤ではアクセラレータとメモリの間でデータをやり取りする際にもCPUが介在する。そのためCPUの電力消費が問題となる。またCPUやGPUなどを備えるサーバを基本として計算資源を拡張するのが一般的であるため、大量のアクセラレータを用いて演算を行おうとするとサーバ間のデータ転送が必要になり、効率が良くない。

これらの課題を解決するのがNTTデバイスイノベーションセンタと連携して研究開発に取り組むホワイトボックスサーバ技術だ。



NTTソフトウェアイノベーションセンタ
システムソフトウェアプロジェクト
（左）横野 智也氏
（右）史 旭氏

ホワイトボックスサーバ技術

ホワイトボックスサーバ技術の目的はデータ処理や演算の効率を徹底的に高め電力消費を抜本的に削減することにある。基本的なコンセプトは、特定の用途や演算に特化した多種多様な専用プロセッサやアクセラレータを効率の良いインターコネクト

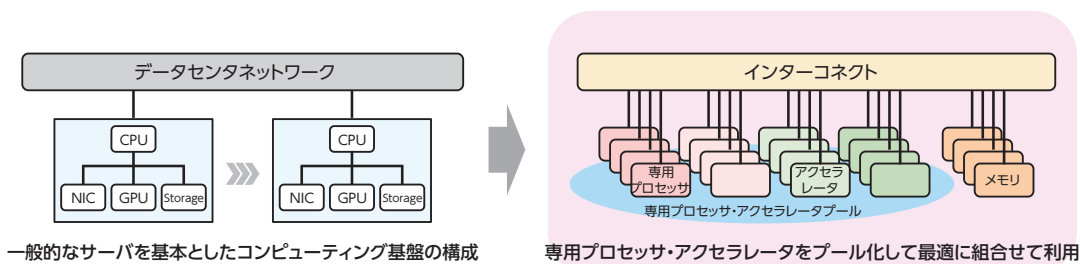


図1 ホワイトボックスサーバ技術のコンセプト

① 専用プロセッサの効率的な活用

専用プロセッサが自律的に通信することで、CPUを介さないデータ移動を可能とし、専用プロセッサ主体のコンピューティングを実現

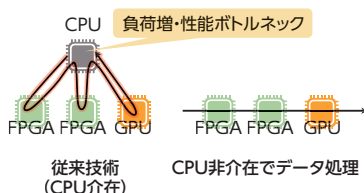


図2 ホワイトボックスサーバ技術：2つの特徴

トで接続し、それらの計算資源をプール化する(図1)というものだ。サーバをまたがるデータ転送を不要にし、サーバの筐体と関係なく拡張可能な計算資源の中から最適なものを適宜組み合わせ効率良く処理・演算できるようにする。

実現に向けSICで研究開発している技術を大きく2つの要素(図2)に分けることができる。1つはインターコネクに接続された専用プロセッサやアクセラレータがCPUを介さず、自律的にデータを転送し合えるようにする技術。もう1つは用途に応じて最適なコンピューティングリソースを必要なだけ組み合わせ処理・演算できるようにする技術だ。

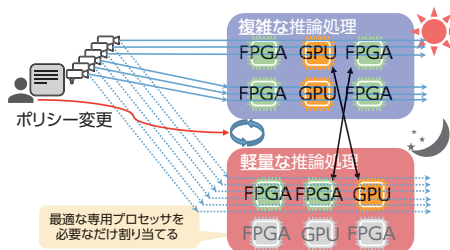
これまでに前者についてはCPUを介さず自律的にデータを転送し合うための技術を実装したFPGAの試作の開発、また、後者については予め定義した順序でデータの転送先を制御する技術を開発した。

イベント駆動型推論技術による電力消費抑制

ホワイトボックスサーバ技術に関する取り組みに加え、AI推論の電力消費を抑える研究開発も進めている。たとえばカメラ映像から交通量を可視化するユースケースでは、カ

② 専用プロセッサの柔軟な構成変更

用途やシーンに応じて専用プロセッサを最適に組み合わせる



メラ映像を解析して自動車や人を検出する。従来の手法ではカメラが自動車や人を捉えていない状態の映像に対しても一律にGPUを必要とするAI推論処理を実行するため、電力を無駄に消費してしまう。

SICでは解析が必要な瞬間のデータだけに対してAI推論の処理を実行することで、電力消費を抑えるコンセプトを考案した。このコンセプトに基づき研究開発しているのが「イベント駆動型推論」だ。映像の各フレームに対しイベントの発生有無を判定する軽量の処理を実行し、イベントが発生しているフレームだけ、上記の例であれば自動車や人が写っているフレームだけを対象にAI推論処理を実行する。

ホワイトボックスサーバ技術とイベント駆動型推論技術組み合わせて効果を確認

2022年4月時点でホワイトボックスサーバ技術とイベント駆動型推論技術を組み合わせ、街中を写した映像から人が写っているフレームだけを対象にAI推論の処理を行う実験を行った。^{※1}

この実験システムではまずカメラ映像の受信とデコードをそれぞれCPUとGPUで行う。その後はCPU

を介さず自律的にデータを転送し合うための技術を実装したFPGAでリサイズ処理を実行し、その画像に対してCPUで解析対象の有無を判定する。高精度な解析が必要な場合、GPUによる検出処理を実行する。

従来手法と実験システムによる解析処理の消費電力を比較した結果から、一般的な街中の映像であれば約40%の消費電力削減効果を見込めるという結論を得た。

計算資源のプール化を実現する試作機を制作

ホワイトボックスサーバ技術の目的の1つである計算資源のプール化に関する研究開発も進んでいる。PCI Express 4.0によるインターコネクで複数のデバイスを接続可能な市販のデバイス拡張ボックスを活用し、複数のFPGAとGPUを1つの筐体に収めたシステムを試作した。このシステムには、プール化されたコンピューティングリソースから用途に応じてFPGAやGPUを組み合わせて処理を実行可能にする仕組みを実装した。

処理に応じた計算資源の利用で電力効率を2.5～4倍に向上

2022年11月に開催された“NTT R&D フォーラム — Road to IOWN 2022”では上記のシステムを使い、街中に設置された8つのカメラの映像を解析し、AI推論によって人物を検出するデモ展示を行った。その概要を図3に示す。

昼間は人が多いのに対して夜は人が少ないことに着目し、昼に撮影された映像と夜に撮影された映像をそ

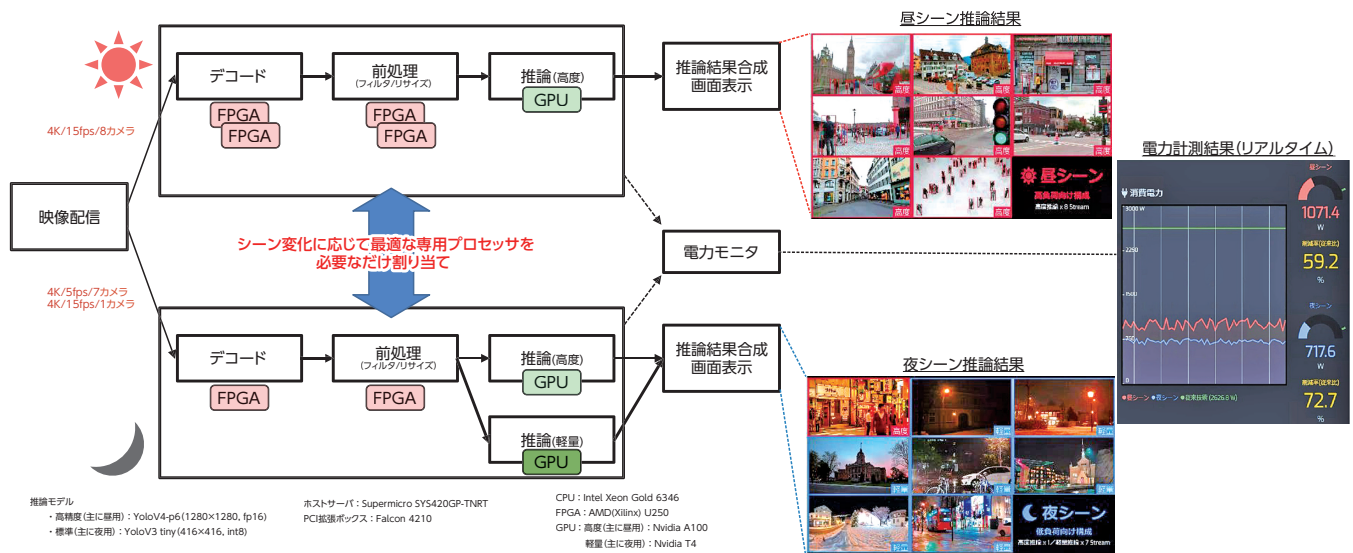


図3 処理に応じて計算資源の構成を変更したデモの概要

それぞれ異なる構成のアクセラレータ群を使って解析している。

「人が多い昼の映像ではFPGAを多く使用し、AI推論には高性能なGPUを使用します。一方で人が少ない夜の映像ではFPGAの数を減らし、AI推論の処理にも性能が比較的安く消費電力が少ないGPUを利用する構成にしました。」(史氏)

こうしてカメラ映像から人の検出を行った際の消費電力を図3の右側「電力計測結果 (リアルタイム)」に示す。緑色の線は2020年時点で一般的なCPUやGPUベースのサーバ構成で測定した消費電力を示している。全カメラに対して同じ処理を行うため、電力消費が一定であることがわかる。

赤と青のジグザグの線は、ホワイトボックスサーバ技術とイベント駆動型推論技術の組み合わせにより昼と夜の映像を解析した際の消費電力を示す。昼の映像に関する処理では59.2%、夜の映像では72.7%の消費電力削減を達成した。昼と夜の映像解析で計算資源の構成を変更した

効果は明らかと言える。本技術により1カメラあたりの電力効率を約2.5～4倍まで向上可能であることを実証した。

「計算資源の構成を制御する『計算資源コントローラ』の開発も進めています。昼と夜の映像解析の例のように予め計算資源の構成を決めておくだけでなく、状況に応じて自動で構成を変更するような制御も実現したいと考えています。」(史氏)

xPUへの対応やユースケースを探る活動に注力

ホワイトボックスサーバ技術はCPUやメモリを光で接続し、物理的に離れた場所の計算資源も活用して高速かつ低消費電力で処理を可能にする次世代のコンピューティング技術「ディスアグリゲータッドコンピューティング」につながる技術の1つだ。NTTデバイスイノベーションセンタによる光電融合デバイスの研究開発が進むにつれ、インターコネクトを電気から光に置き換えることが想定されている。こうしたハー

ドウェアへの対応を含む今後の研究活動について横野氏は次のように述べている。

「インターコネクトが電気から光に置き換わることへの対応以外では、『xPU』と表現している多種多様な専用プロセッサやアクセラレータへの対応を考えています。将来的にはNIC (Network Interface Controller) とのインターコネクトを制御対象とする可能性もあります。ベンダー依存の回避を意識しているのですが、たとえばGPUはベンダーにより実装がさまざまであり、多くの検討課題があると思っています。IOWN Global ForumにおけるDCI (Data-Centric Infrastructure) の議論に参加するベンダーが増えるほど、こうした課題は解決されやすくなるのでは、と期待しています。このほか計算資源コントローラの研究開発や、映像解析以外のユースケースを探る取り組みにも力を入れる考えです。」

※1 参考資料: 「膨大な映像データの解析処理を効率化し都市規模のサイバーフィジカルシステムを支える」