



感情を纏うパビリオン ——ヒトとモノが呼応するコミュニケーション体験

本稿では、2025年日本国際博覧会（大阪・関西万博）におけるNTTパビリオンの未来のコミュニケーション創出のコンセプトの1つである「感情を纏うパビリオン」を実現している、IOWN（Innovative Optical and Wireless Network）光コンピューティングとMediaGnosisの取り組み、および、NTTパビリオンへ訪問されるお客さまへの安心・安全を提供している、IOWN光コンピューティングと身体行動理解技術の取り組みについて紹介します。

キーワード：#IOWN光コンピューティング、#IOWN Data Centric Infrastructure、
#光電融合デバイス

大阪・関西万博のNTTパビリオンにおける演出概要

人の感情や感覚がモノと連携する未来のコミュニケーションの実現に向け、IOWN（Innovative Optical and Wireless Network）時代のコンピューティング基盤である「IOWN光コンピューティング」とNTT独自の次世代メディア処理AI（人工知能）「MediaGnosis」を活用し、音楽ユニットPerfumeの演出が上映されるZone2内部に設置された5台のカメラで撮影した来場者の顔映像を、1秒に1回の頻度で表情分析し、笑顔の人数を算出し、その笑顔の数に応じてNTTパビリオンを覆う幕を複雑に制御しています（図1）。

また、NTTパビリオンへ訪問されるお客さまへの安心・安全を提供するため、「IOWN光コンピューティング」と「身体行動理解技術」を活用し、NTTパビリオン外部に設置された16台のカメラで撮影した映像から人物を検出し、カメラごとの混雑度（個々のカメラ撮影範囲の人物検出数／個々のカメラの撮影範囲面積）の算出結果、および、検出した人物の骨格を推定し当該人物が転倒しているかどうかを判定し、転倒者が検出されたエリア情報をNTTパビリオンのキャストが所有する運用者端末に通知しています。通知を受けたNTTパビリオンのキャストは、混雑度緩和のためのお客さまの誘導、もしくは転倒されているお客さまの介護のため、当該エリアにタ

イムリーに駆けつける運用を実現しています。

IOWN光コンピューティングの特徴

IOWN光コンピューティングは、GPU（Graphics Processing Unit）などのアクセラレータを搭載するサーバと光電融合スイッチから成るハードウェア部と、GPUなどのアクセラレータを柔軟に制御するDCI（Data Centric Infrastructure）コントローラのソフトウェア部から構成される、次世代のコンピューティング基盤です（図2）。

市中においてマーケティングやセキュリティなどの目的で映像処理のAI化が加速



図1 大阪・関西万博NTTパビリオンにおけるDCIシステム概要

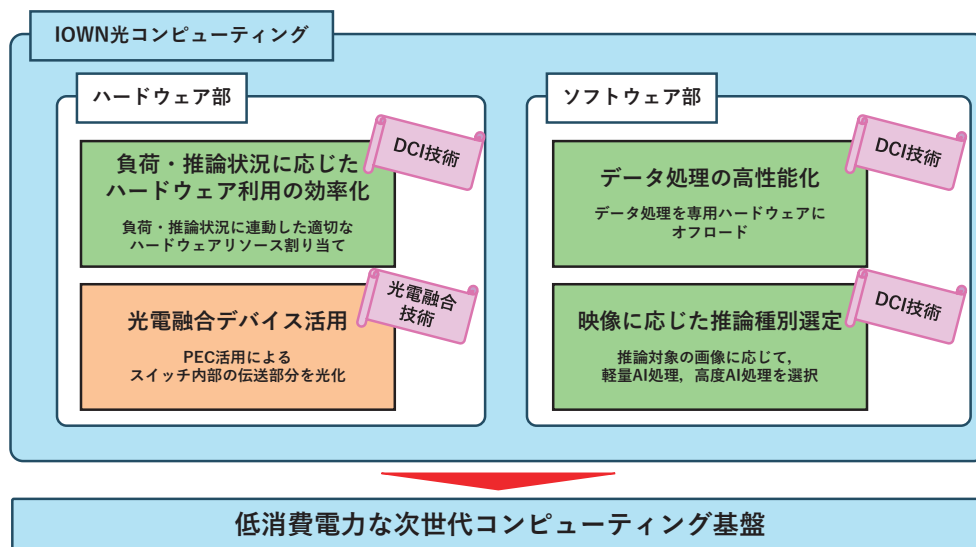


図2 IOWN光コンピューティングの構成・特徴

していますが、一般的なアクセラレータの利用ではGPUなどの物理リソースが単調増加してしまう、通信量の増加によりイーサネットスイッチの処理能力の限界を迎えてしまうなどの課題があり、特に電力消費量の増大に関しては地球環境への影響が問題視されています。その課題解決に向け、GPUなどの物理リソースの効率的な利用、光電融合デバイスを活用したイーサネットスイッチの適用により、標準的なサーバ構成と比較し消費電力を8分の1に抑えています。

DCIコントローラの特徴には大きく3点あります。

1点目は、データ処理の高性能化に向けた、最先端ハードウェアの性能を引き出す、ソフトウェア実装技術です。GPUやスマートNIC（Network Interface Card）などの専用ハードウェアへの処理のオフロードにより、高速なデータ処理、低消費電力化を実現しています。

2点目は、推論を担うアプリケーションの負荷状況や推論状況に応じて、ハードウェア利用を効率化する技術です。具体的な例を挙げると、人物の映像解析の場合、人が多い日中帯には推論に必要なGPUなどの物理リソースを増加させ、人が閑散となる夜間には物理リソースを減少させる技術です。推論処理に必要な物理リソース量

を適正化することで、消費電力を削減しています。

3点目は、映像の内容に適した推論選定技術です。従来技術では、人が映っている画像と人が映っていない画像を区別することなく、同一の推論を実施していますが、本技術では人が映っている画像は高度推論処理、人が映っていない画像は軽量推論処理というように、画像ごとに処理する推論の種別を都度決定し、推論処理の実行回数を最小限に抑えることで低消費電力化を実現しています。

光電融合スイッチの特徴は、従来のイーサネットスイッチ機構とは異なる、スイッチ内部の伝送部分を電気接続から光接続に置き換える、光電融合デバイスを組み入れたイーサネットスイッチであり、従来技術の128ポートイーサネットスイッチを小型筐体1つで構成可能とすることにより、イーサネットスイッチの性能向上、低消費電力化を実現しています。

光電融合スイッチ内部に組み込まれている光電融合デバイスは、従来の400 G光モジュール8個分のデータ処理性能、従来の光モジュールと比較し、2分の1の消費電力性能を有している非常に小型のデバイスです。

MediaGnosisの特徴

来場者の方の感情を分析し、パビリオンを覆う幕の制御につなげるための技術について概略・特徴を説明します。具体的には、DCIの推論選定技術を活用しながら検出を行ったパビリオン内部の来場者の方の顔画像に対して、NTT独自の次世代メディア処理AI「MediaGnosis」を用いて来場者の表情を分析します。この分析結果から特に楽しそうな表情や驚いた表情の感情をしている来場者の人数を算出し、パビリオン内部の盛り上がりとして幕演出の制御に反映しています。大阪・関西万博の実環境で利用するにあたり、MediaGnosisの感情分析AIモデルに対して、以下の2点の課題を解決するように改良を実施しました（図3）。

1点目は、5台のカメラが設置されているNTTパビリオンのZone2では、演出によって照度が大きく変化するモニターと、舞台上に設置されている経時的に点灯・消灯するLEDが光源となります。したがって、来場者の方々の顔の見え方が一定ではなく経時的に大きく変化するという、光学的に非常に不安定な環境での運用が想定されます。さらに、多数の来場者の方がカメラに同時に映りこむ際に、一部の来場者どうしの顔が重なり、顔全体が取得できないケースも

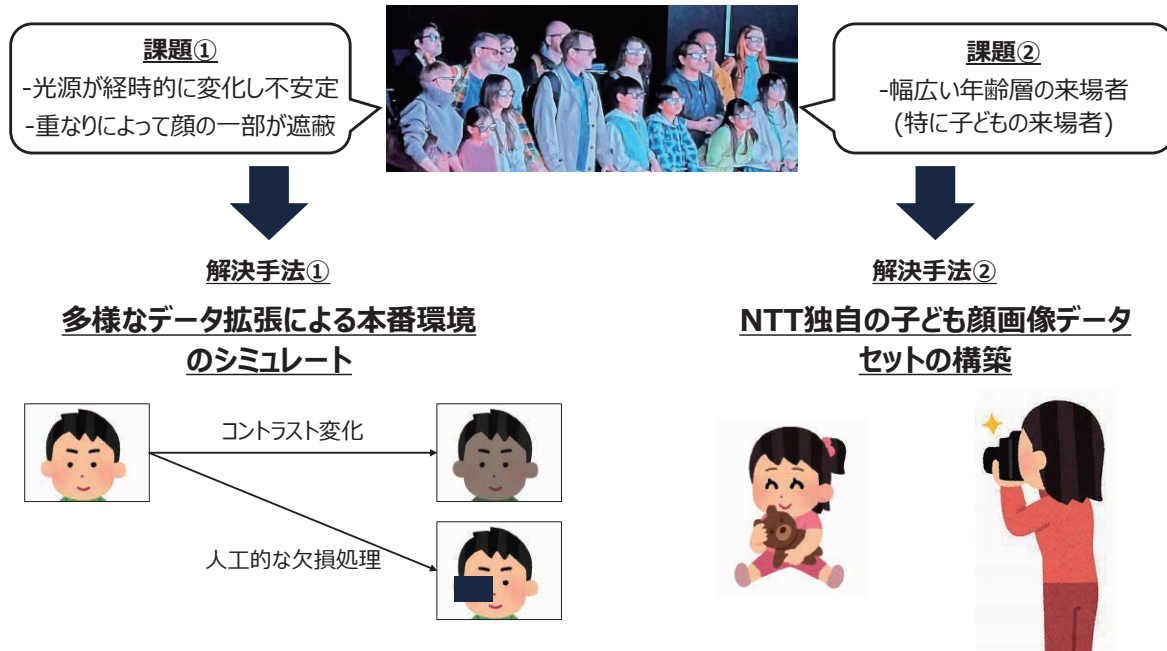


図3 MediaGnosisにおける課題と解決手法

発生すると考えられます。そこで、コントラスト変換や画像の人工的な欠損処理などを含めた10種類以上ものデータ拡張技術をAIの学習に導入しました。これらの拡張は万博での本番環境で発生し得る変動を疑似的にシミュレートし、AIモデルが本番の環境下でも安定した挙動を行うように学習されることが期待できます。結果として、万博で想定される状況下においても、高いロバスト性を持つ顔感情分析AIを実現できました。

2点目は、万博では幅広い年齢層の来場者、とりわけ子どもの来場者の方が多いことが想定されます。しかし、AIを学習するための市中データの多くは15歳以上の方の顔画像から構成されており、15歳未満の年齢層の表情分析において高い性能を実現することが難しいという課題がありました。これは、不特定多数が閲覧できてしまう市中データは、プライバシーの観点で子どもの顔画像を含めにくいという事情があると考えられ、市中のデータをどれだけ集めても、なかなか子どもに対する性能向上にはつながりません。そこで0歳から15歳の年齢層の顔画像を中心とした希少価値の高いデータセットをNTT研究所

内部でだけ閲覧可能なかたちで独自に作成しました。データセットの作成にあたっては、大人の顔画像データセットと同等の人数の被験者を集めることは困難であるため、画角や背景のパターンを多数用意して撮影することで、少ない人数でもできる限り豊富なバリエーションを準備しました。この独自データセットをAIの学習に活用することで、子どもを含む幅広い年齢層の来場者に対して高い分析性能を実現しました。

開幕前に実際のNTTパビリオンZone2においてさまざまな年齢・国籍を含むエキストラの方にご協力いただいた実証では、幕演出に対して十分な性能と頑健性を確認することができました。この成果は、NTT独自のAI技術であるMediaGnosisの強みとして、大阪・関西万博の演出にとどまらず幅広いユースケースで活用できると考えています。

身体行動理解技術の特徴

身体行動理解技術は、監視カメラやロボット搭載カメラから人の身体行動を認識する技術です。大阪・関西万博では、NTTパビリオン敷地内における転倒者をNTT

パビリオン外部に設置されたカメラの映像から検出するAIをDCI上で動作させることで、NTTパビリオンのAIによる見守りを効率的に実現しています。

多くの転倒検出技術は監視カメラ等のエッジ端末にAIモデルを搭載するか、CPU/GPUサーバを設置しAIモデルを搭載しますが、前者のAIモデルは簡略化され精度が低下する傾向にあり、サーバで動作させる場合では多くのカメラに多数の人を想定した効率的なAIモデルにならないことがほとんどです。今回、NTTパビリオン外部に設置された16台のカメラに対してAIによる見守りを実現するため、多カメラ多人数の推論に対応させたAIモデルを具備した転倒者検出技術を確立する必要がありました。

技術のポイントは2点あります(図4)。

1点目は、カメラの高さ、角度の変化に対応した転倒検出モデルの創出です。NTTパビリオンの外部に設置されているカメラは4～8m程度の範囲にさまざまな高さで設置され、カメラの角度も多様です。これまでの転倒検出モデルは屋内を対象としていることがほとんどであり、屋外に設置されたカメラの高さや角度の変化に対応す

ポイント① さまざまな高さ、角度のカメラ映像に適した転倒検出モデルの創出



ポイント② 人物数の変化にロバストな人物トラッキングモデルの構築

図4 身体行動理解技術における技術ポイント

ることが困難でした。そのため、カメラの高さや角度の変化に対応した転倒検出モデルの構築に向け、イベント会場を模した映像データセットを制作しました。50人程度の演者がイベント会場にいる状況下を想定し、演者が転倒を含む行動を演じ、それらの行動にアノテーションを実施して、多カメラ多数の映像データセットを制作しました。そのデータセットを用いることで従来の転倒検出モデルよりもさまざまなカメラの高さ、角度のカメラ映像に対して転倒者を検出可能なモデルを構築しました。

2点目は、人物数に依存しない人物トラッキング技術です。転倒者の検出では、まず人物を検出し、その人物が一定時間動作せず転倒姿勢であることを判別して、転倒者の有無を通知します。そのため、映像中に映る人物1人ひとりをトラッキングしてIDを付与、管理する必要があります。また、人物1人ひとりをトラッキングすると、人物ごとの処理が必要になるため、人物数の増加に応じて処理量が増加、多人数が映るカメラのある大規模なイベント会場への適用が困難な場合があります。今回、人物ごとに生じるトラッキング内部の処理をGPUで並列処理可能とすることで、人物数の変化に対してロバストな人物トラッキングモデルを構築しました。特に従来CPUで処理せざるを得なかったループ処理をGPUで並列処理可能とすることで、複数のカメラ映像に多人数が映る場合においてもCPU処理量の増加を低減したうえで、効率的にGPUで実行可能としました。また、転倒検出モデルはDCI上で動作させ

ており、DCIによる低消費電力化の恩恵を受けています。

今後の展開

DCIは、今回の大阪・関西万博で培った知見を活用し、2026年度完了予定の商用版の開発を進めていきます。さらに、IOWNの実現目標である2030年に向けて、さらなる消費電力の削減、新たな価値の創出に取り組んでいきます。

光電融合デバイスは、さらなる小型化、およびCPUやGPUとの接続などのコンピュータ内部の電気接続を光接続に置き換えを進めるとともに、CPU、GPUなどの同一種別のリソースを束ねるリソースプールの仕組みを導入し、必要な物理リソースを自由につなぐコンピュータの実現、さらなる低消費電力化に取り組んでいきます。

MediaGnosisは、デモサイト(<https://www.rd.ntt/mediagnosis/demo/>)でも試していただけるように、顔感情分析に限らずさまざまな機能を持つメディア処理AIであることから、大阪・関西万博での取り組みを通して積み上げた技術・知見・データを基に、MediaGnosisの多様な画像認識機能に対して、さまざまな実環境でロバストに動作するAIの実現を図っていきます。さらに、MediaGnosisの最大の特徴である「知識集約型アーキテクチャ」と組み合わせ、画像に限らずさまざまなメディア処理の性能向上をめざします。

身体行動理解技術は、大阪・関西万博で得られた多カメラ多数に対して効率的に

行動を認識するための技術を発展させ、転倒だけでなく、人のより高度な行動を多カメラ多数の環境下でリアルタイムに検出する技術の確立をめざします。これにより、さまざまな業界の業務行動への適用拡大に取り組んでいきます。



(上段左から) 岡 順一 / 史 旭 / 水野 伸太郎

(下段左から) 鈴木 聡志 / 中澤 裕一 / 高木 基宏

本稿では、大阪・関西万博における、「感情を纏うパビリオン」を具現化するMediaGnosis、およびDCI、かつ、NTTパビリオンの安心・安全な運用に向けた身体行動理解技術、およびDCIの取り組みを紹介しています。今後これらの技術を発展・普及させ、IOWN構想の実現に貢献していきます。

◆問い合わせ先

NTTソフトウェアイノベーションセンサ
システムソフトウェアプロジェクト