



超高臨場感通信技術「Kirari!」 —サラウンド映像合成・同期伝送技術—

日本電信電話株式会社 NTT サービスエボリューション研究所 主幹研究員

ふかつ しんじ
深津 真二

日本電信電話株式会社 NTT サービスエボリューション研究所 主任研究員

ほし て たかひで
星出 高秀

日本電信電話株式会社 NTT サービスエボリューション研究所 研究主任

お の まさと
小野 正人

日本電信電話株式会社 NTT サービスエボリューション研究所 研究主任

みやした ひろむ
宮下 広夢

1. はじめに

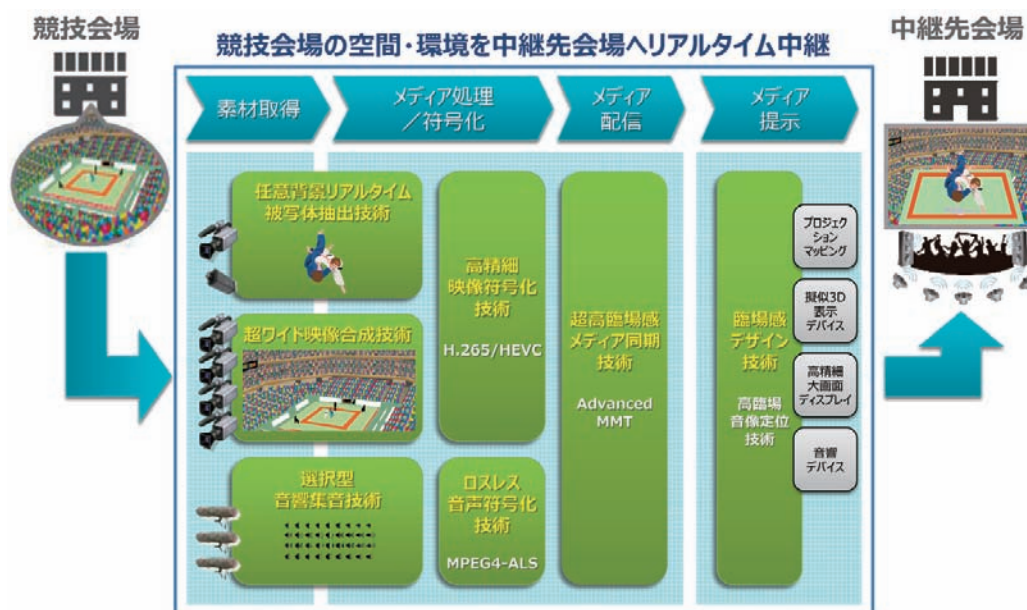
超高臨場感通信技術「Kirari!」は、NTTが2015年から開発を進めている「メディア制御」「メディア処理」「リアルタイム同期伝送」を実現する技術集合で、遠隔地にネットワークを介して、リアルタイムに競技空間やライブ空間を「丸ごと」伝送、再現し、あたかも競技場にいるかのような体験をあらゆる場所で感じることができる超臨場感サービスの実現に向けて取り組んでいる。本稿では、Kirari!を支える中核技術である「サラウンド映像合成・同期伝送技術」の概要と、本技術を活用した取組みについて紹介する。

2. 超高臨場感通信技術「Kirari!」

Kirari!はNTT独自となる最新技術を中心に、標準化技術（高精細映像符号化技術、ロスレス音声符号化技術）や市中の多様な提示系システムとを組み合わせることで、新

しい映像体験を実現している。Kirari!の技術概要を図1に示す。まず、「任意背景リアルタイム被写体抽出技術」は、グリーンバックなどのスタジオ設備を用いずに試合会場や演技している舞台映像からリアルタイムに被写体の領域のみを抽出する技術で、機械学習を導入し、任意背景からの精緻な被写体抽出を実現している。次に、「臨場感デザイン技術（高臨場音像定位技術）」は、超音波の拡散反射を利用することで、映像の任意の位置に音像を定位できる技術で、従来よりも少ないスピーカ数で広範囲な受聴エリアに対応している。残りの「サラウンド映像合成技術」と「超高臨場感メディア同期技術」は、4K/8Kを超える解像度と広い視野角を有する映像をリアルタイムで合成・同期伝送する技術で、あたかも現地で競技を見ているかのような体験を実現する。これら2つの技術は次章以降で詳しく説明する。

また、Kirari!の活動を通じ、高臨場で没入感を深めたラ



■図1. Kirari!の技術概要



イブ体験を伝送する高臨場感ライブ体験 (ILE: Immersive Live Experience) の標準化にもITU-Tにて取り組んでいる。

3. サラウンド映像合成技術

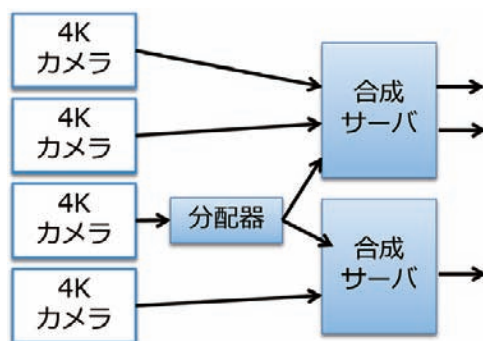
サラウンド映像合成技術は、複数の4Kカメラ映像をリアルタイムに合成することにより、単一のカメラでは撮影できない高精細かつ広視野角のサラウンド映像を生成する技術である。

サラウンド映像合成の処理フローを図2に示す。まず、できるだけ光軸を合わせ、隣り合う映像間で撮影範囲が少し(20~40%)重複するようにカメラを設置した状態で映像を撮影する。合成処理では、事前処理として、基準となる静止画を選択・解析することにより、視差を補正するための射影変換行列を算出する。合成時は、各映像フレームに対して、算出した射影変換行列を使って補正を行いつつ、適切な映像の結合線(シーム線)を見つけること(以下、シーム探索)で、映像フレームを合成する。

本技術の特徴は、「同期分散処理」及び「シーム探索の



■図2. サラウンド映像合成の処理フロー



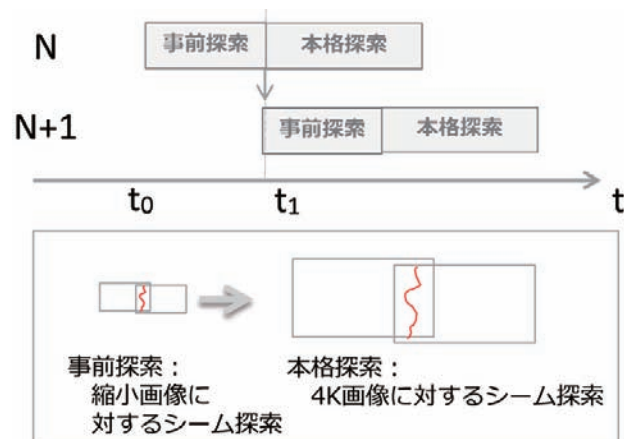
■図3. 複数サーバでの同期分散処理

高速化」、「GPU (Graphics Processing Unit) 活用」の3点である。

同期分散処理では、リアルタイムに膨大な映像処理を実行するため、入力・出力条件により合成処理を複数サーバに分散させる(図3)。この際、各サーバにおいて合成タイミングを同期させるために、入力時に映像フレームに対して同じタイムスタンプを付与し、このタイムコードを後段のエンコーダ等の伝送装置に伝搬させることで、システム全体で同期制御を可能としている。なお、タイムスタンプは、撮影装置等で付与され入力映像に重畳されたタイムコードを利用する方法と、映像が入力された際にタイムコードを生成する方法の2つの方法から選択することができる。

シーム探索の高速化では、シーム探索時に映像フレームを解析して得られる動物体の動きや形を用いることに加え、過去の映像フレームのシーム位置を用いることによりシーム位置のバタつきを防止する。この際、過去の映像フレームのシーム探索が終わってから次のフレームのシーム探索を行う方法ではリアルタイム合成を実現することが難しい。そのため、事前探索として縮小した映像フレームに対してシーム探索を行い、その事前探索結果を次のフレームにおけるシーム探索で利用する2段階シーム探索を行うことによりリアルタイム合成を可能としている(図4)。

GPU活用による高速化では、ほぼ全ての映像処理をGPU上で実行することにより、処理の集約化・高速化を達成している。具体的には、NVIDIA GPUDirect™ for Videoを利用することにより、SDIボードから入力された映像を最小限のCPUコストでGPUに転送し、その後の映像処理をGPUで実行する。その結果、例えば、4台の4Kカメラからの映像より12K (4K×3) のサラウンド映像を合成す



■図4. 2段階シーム探索のイメージ



る場合、GPUを使わない場合に比べて必要となるサーバ台数を1/3 (6台→2台) に削減することができた。また、合成処理の遅延時間は、条件により異なるが、例えば4K 59.94fps (frames per second) 映像3本から8K (4K×2) のサラウンド映像を合成する場合、平均100msec未満である。

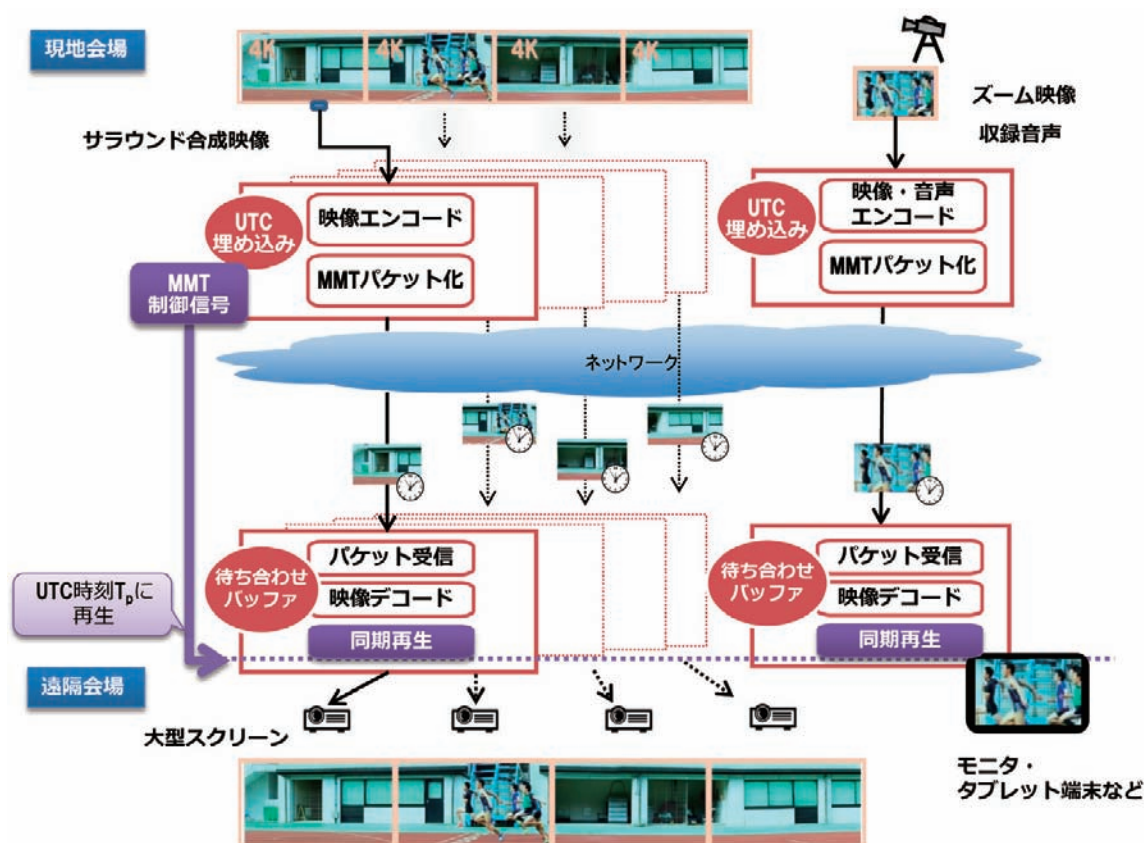
4. 超高臨場感メディア同期技術 (Advanced MMT)

超高臨場感メディア同期技術 (Advanced MMT) は、メディア伝送規格MMT (MPEG Media Transport) に準拠したUTC (Coordinated Universal Time) ベースの同期制御信号の活用によって、複数映像・音声を低遅延に同期伝送する技術である。

競技空間全景をカバーするサラウンド映像は、一般的な表示機器で扱うことができるように、複数の4K映像に分割して、処理・伝送を行う。この際、分割されて処理・伝送される個々の映像・音声の再生タイミングを完全に同期させる手段として、超高臨場感メディア同期技術 (Advanced MMT) を活用している。本技術により、サラウンド映像だ

けでなく、特定の選手を追跡するような複数アングルのカメラ映像 (2K等) を同期して表示したり、各カメラ映像・音声伝送されるネットワーク (専用線やインターネット、無線LAN等) や端末処理による遅延の違いを吸収することが可能となり、複数の映像・音声完全同期した超高臨場感な視聴体験を実現できる。

超高臨場感メディア同期処理の処理フローを、サラウンド映像の伝送を例に図5に示す。まず、4K単位に分割されたサラウンド映像及び音声をHEVC (High Efficiency Video Coding) 等で符号化する。次に、エンコードした映像・音声データをMMTパケット化する。このとき、遠隔会場側での再生時刻TpをUTCベースのタイムスタンプとして埋め込む。他方、遠隔会場側の視聴端末はMMTパケットを受信し、映像・音声データをデコードするとともに、タイムスタンプ情報を取り出す。そして、デコードした映像・音声データをバッファ領域で再生時刻Tpまで待ち合わせたうえで提示する。これにより、伝送ネットワークや端末処理に違いがあるような場合においても、複数の映像・音声の再生タイミングを完全に同期させることが可能となる。



■ 図5. 超高臨場感メディア同期処理の処理フロー



この際、超高臨場感メディア同期処理でも、エンコード・デコードの映像処理をGPU (NVIDIA GPU Direct™ for Videoを利用) 上で実行することにより、処理の集約化・高速化を達成している。その結果、4K映像2本のエンコード処理、4K映像3本のデコード処理をそれぞれ1サーバで実施できる。また、伝送処理の遅延時間は、例えば4K 59.94fps映像を1本処理する場合、エンコード処理は50msec未満、デコード処理は100msec未満と低遅延化を実現している。

さらに、超高臨場感メディア同期技術 (Advanced MMT) では、映像や音だけではなく、空間の再構成に必要な位置や方向などの3次元情報、ステージ演出で広く用いられている照明機器などをコントロールするDMX信号なども同期伝送できる。このことにより、映像・音に他の情報を付加した高臨場な演出に活用することも可能となる。

5. 実装、更なる展開

2018年11月に開催された「NTT R&Dフォーラム2018 (秋)」において、サラウンド映像合成技術・超高臨場メディア同期伝送技術のデモンストレーションを公開した。会場に設置した14m×2mの横長スクリーンに、アスリートが競技する姿を、4Kカメラ5台で撮影してつなぎ合わせた横16Kの超高解像度映像を生成し、人の視野角を超えた形で映像投影される新しい映像視聴の形を提案した(図6)。この際、映像の表示サイズを実寸大・等身大で表示することで、あたかも目の前で競技しているかのような躍動感と迫力がリアルに体感できた。さらに、NTT研究所が研究開発している波面合成音響技術も活用し、迫力あるサラウンド音響を実現することで、臨場感ある新たなスポーツ観戦体験を実現した。

最後に、更なる高臨場感の実現を目指して検討を開始しているサラウンド映像合成技術の8K対応に関して述べる。



■図6. 臨場感ある新たなスポーツ観戦体験

まず、合成の素材となる8K映像を扱うために、市販の映像機器に手を加えた8K同期再生システムを構築した。これは12台の4Kビデオレコーダを搭載し、最大3枚の8K映像 (7680x4320/59.94p) を記録・再生する機能を有しており、すべての映像をフレーム単位で同期再生できる。この同期制御はタイムコードを基準としており、再生開始時にズレが発生した場合はそれを自動的に検知して即座に修正する。

構築したシステムを8K映像の入出力についてハードウェア性能の観点からフィジビリティ検証を実施した。検証用のサーバには、ビデオキャプチャボードであるBlackmagic Design DeckLink 8K Proを3枚と、合成処理のためのNVIDIA Quadro GV100を2台搭載した。検証では2枚の8K画像をメモリからGPUに転送する処理が59.94fpsに間に合わないという課題が表出したが、これは8K画像を1枚ずつ異なるGPUに転送してそのデータをNVLinkで共有する構成に変更したことで解決された。

次に、GPUによる合成処理について性能評価を行った。この合成処理は、入力画像を射影変換するワープ機能、シーム位置を見つけるシーム探索機能、2枚の画像を結合するマージ機能、フォーマットを変換するコンバート機能に分類される。各機能はフレーム単位での並列実行が可能であるが、性能評価において8K画像を入力とした場合にシーム探索機能のみ59.94fpsに間に合わないことが判明した。これはシーム探索アルゴリズムの性質として、画素間のコストが最小となる経路の探索がGPUで十分に並列化できないことに起因している。また、シーム探索機能では過去のシーム位置を参照するため、スレッド間の追越しが発生しないように並列化が制限されていることも影響している。

今回の検証結果から、現状では、シーム位置が固定の場合に限り、サラウンド映像合成技術を8K映像に拡張できることが確認された。現在、今回の検証結果も踏まえ、8K対応に加えて、より精緻かつ低遅延な合成処理の実現に向けてアルゴリズムの改善・最適化を進めている。

6. おわりに

昨今、人気の高いスポーツ競技の試合が開催される際には、パブリックビューイングが行われることが一般的になってきており、2020年に向けては、世界中で多くの視聴者がパブリックビューイングなどにより、スポーツの感動を共有することが予想される。今後は、Kirari!のスポーツ、エンターテインメント分野での社会実装を促進すると同時に、エンタープライズ分野への展開拡大も推進していく。