

# 放送番組と同期する自由視点 AR のためのリアルタイム伝送システム

河村 侑輝 大槻 一博

NHK 放送技術研究所 〒157-8510 東京都世田谷区砧 1-10-11

E-mail: {kawamura.y-du, otsuki.k-ek}@nhk.or.jp

**あらまし** 2018 年 12 月に開始される新 4K/8K 衛星放送では、IP (Internet Protocol) ベースのメディアトランスポート方式 MMT (MPEG Media Transport) が採用され、インターネットを活用した放送通信連携サービスの高度化が期待される。本稿では、放送番組と同期する自由視点 AR (Augmented Reality) コンテンツのためのリアルタイム伝送システムのモデルを提案する。提案モデルでは、MMT で放送される映像と同期した AR 提示のため、3 次元モデルのモーションに UTC (Coordinated Universal Time) による提示時刻タイムスタンプを付与してリアルタイムに伝送する。スポーツ中継を模擬した試作コンテンツでは、8K テレビに競技場全体を俯瞰する映像を提示し、タブレットのカメラで映された実空間のテーブル上にミニチュアの競技場を AR 合成して提示する。視聴者は、自分がカメラマンになった感覚でタブレット画面を通してテレビ映像では映らない場所に注目して見たり、好みの選手だけを自由に追いかけて見たりといった、従来のテレビと異なる新たな視聴体験が可能となる。

**キーワード** MPEG Media Transport, 放送通信連携, 拡張現実, 自由視点

## Real-time Distribution for Free-Viewpoint Augmented Reality Content Synchronized with a Broadcast Program

Yuki KAWAMURA and Kazuhiro OTSUKI

Japan Broadcasting Corporation Science and Technology Research Laboratories

1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8510 Japan

E-mail: {kawamura.y-du, otsuki.k-ek}@nhk.or.jp

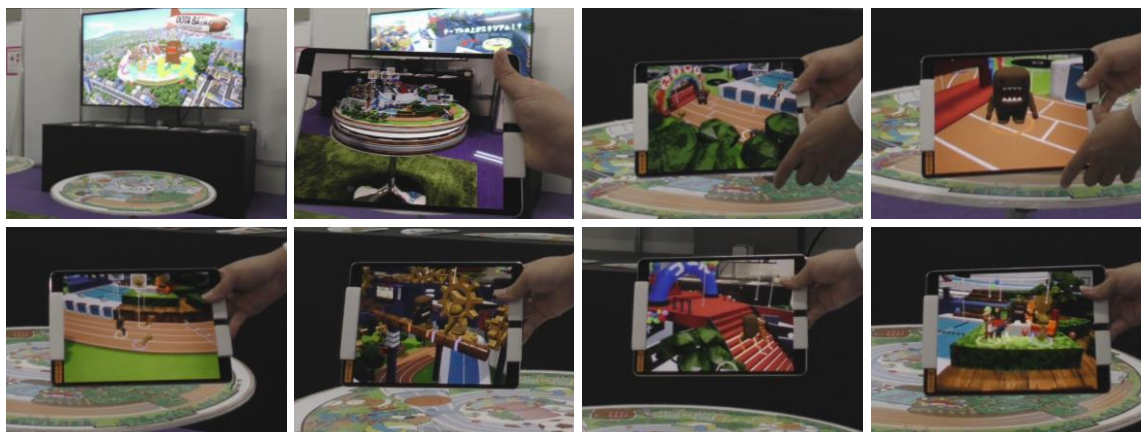
**Abstract** The advanced satellite broadcasting for 4K/8K will be launched in Japan in December 2018. MPEG media transport (MMT) will be used as the Internet-protocol-based media transport scheme to realize advanced hybrid services over the Internet. In this paper, we describe a real-time distribution model of 3D free-viewpoint augmented reality (AR) content for second-display devices via the Internet. For synchronized presentation with the video delivered by MMT, motion data of 3D models are transmitted along with coordinated universal time (UTC)-based presentation timestamps. In a prototype content which is like live sports broadcasting, an 8K television shows overview of a stadium and a tablet device shows the stadium overlaid on top of a real table with AR. The viewer can enjoy a new viewing experiment in which they always can follow their favorite player from any angle as if they were the camera operator.

**Keyword** MPEG Media Transport, Hybrid Delivery, Augmented Reality, Free-Viewpoint

### 1. はじめに

2018 年 12 月に開始される新 4K/8K 衛星放送では、IP (Internet Protocol) ベースのメディアトランスポート方式である MMT (MPEG Media Transport) [1,2] が採用され、インターネットなどの通信網を活用した放送通信連携サービスの高度化が期待される。従来のデジタル放送で使用されている TS (Transport Stream) では、それぞれの符号化装置が独自に生成するクロックを基準とした提示タイムスタンプ (PTS, Presentation Time Stamp) によってコンテンツ提示の制御を行う。このため、伝送遅延が異なる伝送路を経由した複数の映像の精密な同期提示を行うために、各映像の PTS の基準

となるクロックの相対関係を記述した付加情報の伝送が必要になるなど、システムの複雑化が課題となる。一方、MMT では符号化装置の個体差や動作条件に依存しない協定世界時 (UTC, Coordinated Universal Time) を基準とした PTS によってコンテンツ提示の制御を行う。これにより、MMT では放送とインターネットなど伝送遅延の異なる伝送路を経由した映像の同期や差し替えなどを行う放送通信連携サービスが、TS の場合よりもシンプルなシステムで実現できる。筆者らはこれまで、MMT を用いた放送通信連携サービスの例として、放送映像と通信経由の別アングルの映像を同時視聴するマルチビューサービス、視聴者の属性



Domo ©NHK・TYO

図 1 どーもくんのどたばたレース

に応じて放送映像の一部を通信経由の映像で差し替えるターゲット広告サービス[3], 放送電波が受信できなくなったときにシームレスに通信経由の映像に乗り替わる通信補完サービス[4]のシステム検討を進めてきた。

他方, 新たなコンテンツ技術として, 現実世界に付加的な情報を合成して提示する拡張現実 (AR, Augmented Reality) 技術への注目が高まっている。スマートフォンやタブレット向けの基本ソフトである iOS や Android には高性能な AR 支援機能が実装され, 低コストに高機能な AR 活用アプリケーションの提供が可能になるなど, AR 技術の発展・普及が進んでおり, 放送通信連携コンテンツにおいても AR 技術の活用への期待が高い。AR コンテンツの配信については, 事前にダウンロードされたコンテンツデータをユーザ操作のタイミングで再生するものが多く, データのリアルタイムな伝送と受信は現状一般的ではない。また, 映像と AR コンテンツとの同期提示手法としては, 映像に電子透かしやマーカーによるタイムコード情報を重畳し, これを画像解析して同期を行う手法[5]が報告されているが, AR コンテンツについても映像・音声と同列のコンポーネントとして扱い PTS を用いた提示制御を行う例はない。

そこで本稿では, 放送番組と同期する自由視点 AR コンテンツを映像・音声と同列のコンポーネントとしてリアルタイムに伝送し, スマートフォンやタブレットで放送番組と同期した視聴が可能なセカンドディスプレイ AR サービスのシステムモデルを提案する。映像と同期した AR 提示のため, コンテンツの 3 次元シーンを構成する情報のうち, 動的データである 3 次元モデルの動き情報に UTC による PTS を付与してリアルタイムに伝送する。受信端末では, 受信した動き情報に付与された PTS に従って 3 次元シーンをリアルタ

イムレンダリングして, AR 提示を行う。

スポーツ中継を模擬した試作コンテンツでは, 8K テレビに競技場全体を俯瞰する高精細映像を提示し, 一方, タブレットのカメラで映された実空間のテーブル上にはミニチュアの競技場を AR 合成して提示する。視聴者は, タブレットを自由に動かして, 例えば, テレビには映し出されない場所に注目して見ることや, 好みの選手だけを追いかけて見ることが可能となる。

## 2. セカンドディスプレイ AR のサービスイメージ

セカンドディスプレイ AR サービスは, テレビ画面に提示される放送映像とセカンドディスプレイに AR 提示される 3 次元シーンとが高精度に同期することで, 従来のテレビ視聴とは異なる新たな視聴体験を実現する。試作コンテンツ「どーもくんのどたばたレース」を図 1 に示す。この試作コンテンツは, NHK のマスコットキャラクターである「どーもくん」を含む 4 体のキャラクターが, 空想世界の競技場の中で様々な障害物を乗り越えてレースを繰り広げるスポーツ中継を模擬した内容である。8K テレビ画面では 3 次元シーンから事前レンダリングした競技場の全景映像を大画面で提示し, 手元のタブレット画面にはカメラを通した現実世界のテーブル上に競技場のミニチュアを AR 合成して提示する。視聴者は, 自らがカメラマンになったような感覚でタブレット画面を通して競技場を覗き込み, テレビ画面には映し出されない部分を見たり, 応援したい好みの選手を追いかけてたりして楽しむことができる。

上記の様にコンテンツの 3 次元シーンをミニチュアにして AR 合成するのは, あくまで提示スタイルの一つである。視聴環境の空間に十分な広さがあれば, 例えば, 3 次元シーンを実寸大で床面に AR 合成することも考えられる。この場合には, 視聴者は 3 次元シーンの中をウォークスルーする体験が可能となる。

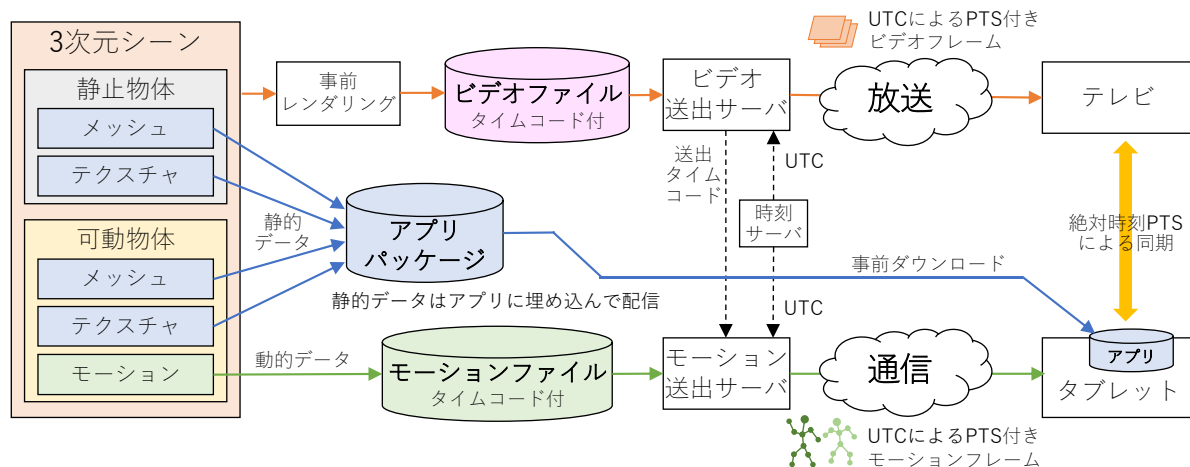


図2 セカンドディスプレイ AR サービスの放送通信連携配信モデル

### 3. 放送通信連携配信

#### 3.1. システムモデル

提案するセカンドディスプレイ AR サービスの放送通信連携配信システムモデルを図2に示す。このモデルでは、コンテンツの3次元シーンから事前レンダリングした2次元映像を、放送伝送路を介して配信してテレビで受信する。これに加え、3次元シーンを構成する3次元モデルの動きの情報を、インターネットを介して配信してAR受信端末で受信する。AR受信端末は、例えば、AR受信アプリをインストールしたスマートフォンやタブレットである。

3次元シーン内の物体の3次元モデルの表現手法について図3に示す。一般に、3次元シーン内の物体は、静止物体または可動物体である。各物体の3次元モデルは、静止物体か可動物体かによらず、物体の3次元形状を多面体で近似する頂点情報であるメッシュと、メッシュの面に張り付ける2次元画像であるテクスチャをもつ。テクスチャをメッシュの面にマッピングするため、メッシュの頂点にはテクスチャの切り出し位置を指定するUV座標が付与される。加えて、可動物体の場合は、その動きを記述する情報であるモーションをもつ。モーションは、3次元モデルに骨格を与え、その骨格にある各関節の6自由度の座標および回転をサンプル化したデータであり、映像フレームと同じように60FPS (Frames Per Second) などのフレーム周期でサンプル化する。

本システムモデルでは、メッシュとテクスチャは時間経過で変化しない静的データであることを前提とし、アプリパッケージに埋め込んで配信する。このため、AR受信端末には番組の放送前に、このアプリケーションをインストールしておく必要がある。一方、動的データであるモーションは放送の映像や音声と同様に、

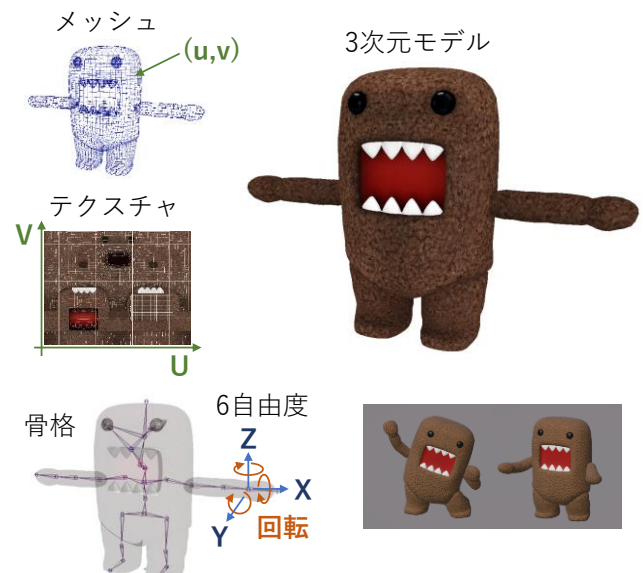


図3 3次元モデルの表現

各フレームにUTCベースのPTSを付与してリアルタイムに配信する。AR受信端末は、PTSが付与されたモーションを受信し、UTC時刻と比較して3次元シーンのレンダリングを行い、カメラ入力映像とAR合成した映像を画面に出力する。

テレビとAR受信端末は、それぞれUTCベースのPTSが付与された映像フレームとモーションを受信する。ただし、PTSの時刻は受信時には過去の時刻となり得るため、受信端末間でコンテンツを同期させるためにはPTSに加算するオフセット時間を受信側の端末間で共有する必要がある。オフセット時間の共有の概念図を図4に示す。AR受信端末では、レンダリングの主要時間を1フレームと推定し、PTSにオフセット時間を加算した時刻から1フレーム分手前の時刻にモーションをレンダラに投入する。

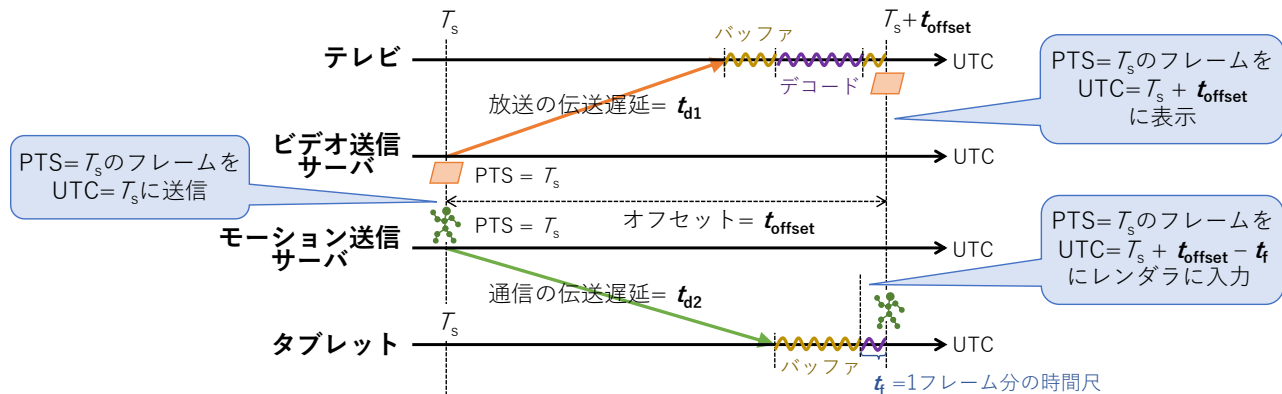


図 4 PTS オフセットによる端末間同期

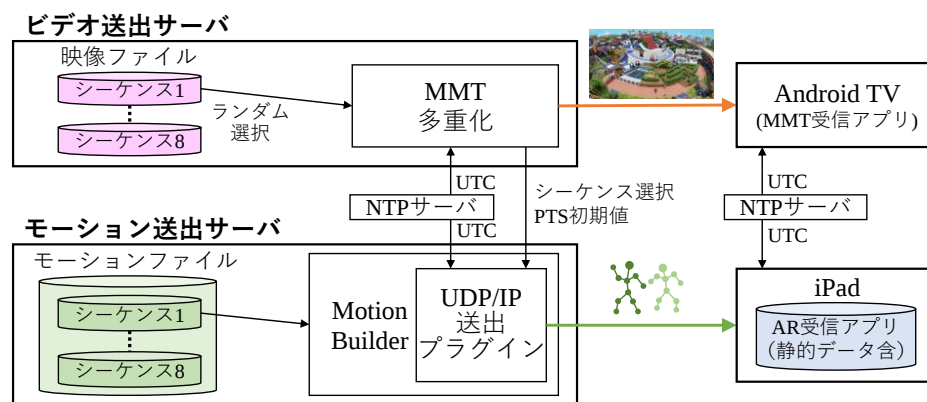


図 5 試作コンテンツのシステム実装

### 3.2. 実装

前述の試作コンテンツ「どーもくんのどたばたレース」のシステム実装を図5に示す。現時点の実装は、実際のサービスを想定したシステムではなく、主に体験型のデモ展示を目的とした実装である。

3次元シーンのデザインおよび、AR受信アプリの機能実装とパッケージ化には、Epic Game社製のゲームエンジンとして知られるUE4 (Unreal Engine 4)を用いた。UE4は、iOSのAR支援機能であるARKitを活用したAR応用アプリの開発をサポートしている。3次元シーンの静的データであるメッシュ（骨格、UV座標を含む）とテクスチャについては、アプリパッケージに内包した状態として、AR受信端末として用いるiPadに事前にインストールする。

一方、モーションについては、アプリパッケージには含めず、FBX (FileBox) ファイルにパッケージ化してモーション送信サーバの入力ファイルとする。モーション送信サーバについては、Autodesk社のモーション編集ソフトであるMotion Builderのプラグインとして新規に開発し、FBXファイルから読み出した1フレーム分のモーションを構造化して、UDP (User Datagram Protocol) /IP パケットに多重化し、PTSを付

与してリアルタイムに送信する機能を実装した。

ビデオ送出サーバとモーション送出サーバは、同一のネットワークに接続して送出タイムコードを共有する。さらに、同一ネットワーク内にNTPサーバを接続し、UTC時刻の参照を可能として、UTC時刻の経過に同期したコンテンツ送出を行う。送出開始タイミングの同期は、ビデオ送出サーバをマスタ、モーション送出サーバをスレーブとして、送出開始時にビデオ送出サーバからモーション送出サーバに対して、シーケンスの先頭フレームに付与するPTS初期値として使用するUTC時刻を記述したトリガパケットを送出する。

また、体験型のデモ展示のため、映像とモーションについてレース展開が異なる複数パターンのシーケンスを用意し、送出するシーケンスの選択もトリガパケットで同期できるようにした。毎回同じレース展開とならないように、8シーケンスの中から送出シーケンスをランダムに選択して送出する。

ビデオ送出サーバは、3次元シーンから事前レンダリングした映像をHEVC (High Efficiency Video Coding) で符号化したシーケンスからフレーム単位のデータを読み出してMMT/UDP/IPパケットに多重化し、UTCベースのPTSを付与して送信する。モーション送

出サーバの出力については、現状実装は独自のプロトコルであるが、MMTと同様にUTCベースのPTSを付与してモーションフレームを送出することにより、MMTによって送信される映像フレームとの同期を可能とする。

### 3.3. データ量の実績

「どーもくんのどたばたレース」の実績値として、選手のキャラクタ4体と背景の可動物体37個で合計関節数が382個、60FPSのモーション伝送で伝送容量はおよそ4Mbpsであった。一方、放送前に事前ダウンロードが必要なアプリパッケージのサイズはおよそ860メガバイトであった。これは、例えば20Mbpsの伝送速度でダウンロードに約6分を要する。

## 4. 提案モデルの拡張

### 4.1. リアルタイムに生成されるモーションのリアルタイム伝送への対応

これまで説明したモデルおよび実装では、モーションを含めて全て事前に制作された3次元シーンを対象としていた。これを拡張し、リアルタイムに生成されるモーションのリアルタイムな伝送に対応できれば、ライブ制作コンテンツを対象としたセカンドディスプレイARサービスが可能になると考えられる。例えば、特殊スタジオを用いるモーションキャプチャシステムをモーション送信サーバの入力として接続すれば、伝送やAR受信アプリの実装は現状のままでライブ制作コンテンツに対応できる可能性がある。また、将来的にはAI(Artificial Intelligence)技術の発達により、特殊なマーカ等を使用せず自然映像から被写体の骨格を抽出し、モーションを生成することが可能になることも期待される。一例として、ディープラーニングを用いたOpenPose[6]は、現状は2次元画像が対象ではあるが、カメラ映像入力から同時に複数人の骨格検出をリアルタイムに行うことが可能である。

例えばサッカー中継を想定すると、3次元スキャンにより選手の3次元モデルを事前に生成し、骨格を入れたメッシュとテクスチャを静的データとして中継の放送前に配信することが考えられる。これに加えて、モーションのみを試合中にリアルタイムに生成することであれば、サッカー中継のセカンドディスプレイARサービスが実現できる。AR提示だけでは選手の表情など細かい部分の再現まではできないが、テレビの実写映像と合わせて、選手が動き回るサッカーコートやARで俯瞰的に視聴できれば、サッカー観戦の楽しみ方が増すことが期待できる。

### 4.2. 将来の3次元空間情報伝送に向けた機能追加

本稿で提案するセカンドディスプレイARサービスのモデルは従来のテレビ放送と合わせて視聴する付加

コンテンツとしての位置づけであるが、将来的には放送サービスの本線として3次元空間情報含むコンテンツを伝送し、AR/VR(Virtual Reality)や3次元テレビを含む様々な視聴スタイルに対応する新たな放送サービスの実現を目指す。放送サービスの本線となり得るためには、従来のテレビ視聴と同様に多チャンネルで伝送される複数コンテンツをザッピングするユーザ体験を、現実的な伝送路の伝送容量という制約下で実現することが要求条件と言える。

すぐにこれを実現することは難しいため、現状の実装をベースとした次段階の機能拡張として、下記に示すAR受信アプリの共通化と、メッシュとテクスチャのリアルタイム伝送の対応を行う。

#### ・受信アプリの共通化

現状の提案モデルでは、事前に3次元シーンの静的データを含むアプリパッケージをダウンロードしてAR受信端末にインストールする手間を要する。コンテンツアクセスに要する手間が大きいと、ユーザの視聴意欲の低下要因となる得るため、これらの手間を削減することが望ましい。そこで、端末にインストールが必要なAR受信アプリはコンテンツによらず共通のものとし、コンテンツごとの静的データをパッケージ化したデータをダウンロードする機能を追加する。

#### ・メッシュとテクスチャのリアルタイム伝送

試作コンテンツ「どーもくんのどたばたレース」は人の手でデザインされた3次元シーンであるため、メッシュとテクスチャを静的データ、モーションのみを動的データとして完全に分離することが容易である。しかし、現実空間をもとに画像認識技術等で生成した3次元シーンを想定する場合、メッシュとテクスチャにも時間経過による変化が生じることが考えられる。そこで、メッシュとテクスチャについても、リアルタイム伝送の本線に多重して伝送する機能と、AR受信アプリでこれらを受信して3次元モデルのレンダリングに利用する機能を追加する。

メッシュとテクスチャをリアルタイム伝送の本線に多重することは、それらが動的に変化する場合に限らず、静的な場合にも有用である。なぜなら、本線からメッシュとテクスチャを受信できれば、事前に別線で静的データのパッケージをダウンロードする手間を省略できる可能性がある。事前に別線から静的データをダウンロードしていない状態でリアルタイム伝送の本線を受信開始した場合、AR受信アプリは本線に多重されたメッシュとテクスチャを受信できたタイミングから3次元モデルのレンダリングを開始できる。つまり、本線で周期的にメッシュとテクスチャを伝送することで、AR受信アプリがそれらを受信できるタイミングが、映像符号化のイントラピクチャと同様にラ

ンダムアクセスポイントとして機能する。

なお、MPEG (Moving Picture Experts Group) による MPEG-I (Immersive) 標準化において、3次元モデル情報を伝送する G-PPC (Geometry-based Point Cloud Coding) の議論が始まっており、今後、3次元空間情報伝送技術の技術開発が活発化することが予想される。

## 5. まとめ

本稿では、放送番組と同期する自由視点 AR コンテンツをスマートフォンやタブレット向けに配信するセカンドディスプレイ AR サービスのシステムモデルを提案し、試作コンテンツの実装について報告した。AR コンテンツのモーションを MMT と同様に UTC ベースの PTS を付与してリアルタイム伝送することで、テレビの映像提示とセカンドディスプレイの AR 提示が高精度に同期する新たな視聴体験を実現した。

今後は、現実空間をもとに制作される3次元シーンの対応も考慮し、動的なメッシュやテクスチャも含めた3次元空間情報伝送技術の検討を進める。

## 文 献

- [1] ISO/IEC 23008-1, "High efficiency coding and media delivery in heterogeneous environments: MPEG media transport", (March 2014)
- [2] ARIB STD-B60, “デジタル放送における MMT によるメディアトランスポート方式”, (July 2014)
- [3] Y. Kawamura, K. Otsuki, A. Hashimoto and Y. Endo : “Functional Evaluation of Hybrid Content Delivery Using MPEG Media Transport”, Proc. IEEE International Conference on Consumer Electronics 2016, pp.267-268 (January 2016)
- [4] Y. Kawamura and K. Otsuki : “Field Experiment of Hybrid Video Delivery Using Next-Generation Terrestrial Broadcasting and a Cellular Network”, Proc. IEEE International Conference on Consumer Electronics 2018, pp.173-174 (January 2018)
- [5] 山本 奏, 田中 秀典, 安藤 慎吾, 片山 淳, 筒口 拳 : “Visual SyncAR : 映像に同期して情報を重畳表示する映像同期型 AR 技術”, 画像電子学会誌, 43 巻 3 号, pp.397-403 (July 2014)
- [6] Zhe Cao, Tomas Simon, Shih-En Wei and Yaser Sheikh : “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, Proc. IEEE International Conference on Computer Vision and Pattern Recognition 2017, pp.7291-7299 (July 2017)