8章 8K/4Kハイブリッドフリーナビゲーション視聴のための映像処理技術

正会員 野中 敬介[†], 正会員 内藤 整[†]

キーワード：自由視点映像，フリーナビゲーション，空間再現，MV-HEVC，人物抽出，人物追跡

1. まえがき

2020年の東京オリンピックを控え、またヘッドマウントディスプレイに代表されるような映像表示デバイスの多様化や8K/4Kといった映像の高解像度化を受けて、近年では高い臨場感の得られる映像技術に対するニーズがますます高まっている。

こういったニーズに対する一つの方向性として、疎に配置された複数のカメラで同一被写体を撮影し、その映像から被写体の空間を3Dモデルとして再現することで、ユーザが任意のアングルからの映像を楽しめる「自由視点映像生成技術」が提案されている^{1)~4)}。筆者らはこれまでこの自由視点映像の高品質化・高速化に取り組んできた。さらに、この技術に加えて、複数のカメラの映像を高効率で圧縮し伝送する符号化標準 Multiview High Efficiency Video Coding (MV-HEVC⁵⁾) を利用することで、カメラ映像とともにオフライン制作された自由視点用データを送信し、受信側(テレビやスマートフォン)のディスプレイ上にリアルタイムで自由視点映像を表示する「8K/4Kフリーナビゲーションシステム」を開発した。本稿では、本システムの全体の構成に触れるとともに、選手の追跡技術を含む自由視点映像生成におけるいくつかの要素技術の概要について紹介する。

2. フリーナビゲーションシステム概要

図1に本システムの概要を示す。まず本システムは8Kまたは4Kカメラで同一被写体を円周上に囲うように撮影した映像を入力とし(図1ではサッカーグラウンドおよび選手が被写体)、これらの映像はあらかじめ時刻同期が取られているものとする。次に、これら複数のカメラの映像を用いて自由視点映像生成に必要なデータ(以下、合成用データ)をあらかじめオフラインで計算しておく(図1：自由視点映像生成・編集)。同時に、複数カメラ間の画像相関を利用する符号化方式MV-HEVCを用いて、映像符号化ストリームを生成する。当該ストリームに前述の合成用データを付与したものを送信する。図1に示す実際のシステム例では、固定カメラで撮影したサッカー(ズームアウト映像で動体が少ない)映像を対象として、8K×1本、4K×2本の映像を入力した上記ストリームを約35Mbpsで送信している。

受信側では、このストリームを受けMV-HEVCの規格に則り映像をデコードするとともに、その映像と受信した合成用データとを用いて、リアルタイムで自由視点映像の生成を行う(図1：フリーナビゲーションデコード)。このように受信側で自由視点映像を生成するため、カメラ映像を鑑賞するモードと、自由視点映像を鑑賞するモードとを

図1 8K/4Kフリーナビゲーションシステムの一例

[†] 株式会社KDDI研究所

"Image Processing for 8K/4K Hybrid Free-navigation System" by Keisuke Nonaka and Sei Naito (KDDI R&D Laboratories, Inc., Saitama)

図2 仮想視点からの映像の一例

ユーザが任意に切替えることが可能となる。

自由視点映像については、被写体の3D空間をリアルタイムで生成しているため、動画を再生しながらゲームパッドなどの入力デバイスを用いることで任意の視点における映像を視聴可能となる。以降の節では、特に自由視点映像生成の技術特徴について述べる。

3. 自由視点映像生成の技術特徴

近年、自由視点映像生成技術にはいくつかの方式が存在するが、著者らはカメラ映像から得られる人物のテクスチャをそのまま利用する「ビルボード方式³⁾」を採用している。

この方式では、まずカメラごとの映像から選手のテクスチャを抽出し3Dの板(ビルボード)とする。そして、図2のようにユーザが選択した仮想視点において、最近傍カメラに正対するようにビルボードを3D空間上に配置することで、任意のアングルからの映像を生成する。この方式の利点としては、「合成用データ算出の計算コストが低い」、「合成用データのサイズが小さい」、「表示デバイスの処理負荷が軽い」といった点が挙げられる。

図1のとおり、本システムではオフライン処理にて合成用データを算出しておく。この処理フローは図3のとおり

図3 合成用データ生成および表示の処理フロー

であり、主にカメラパラメータ算出、選手領域抽出、選手追跡および移動軌跡の平滑化処理からなる。選手追跡は最終的な映像品質を高める処理として利用される。

また、受信側での映像表示における主な処理としては、選手以外の3D空間(背景モデル)の構築、選手領域の背景モデルへの配置である。このうち、背景モデルについては、あらかじめ作成されたモデルデータに選手以外の映像をマッピングすることで生成される。

8K/4Kといった高解像度映像の恩恵により、選手領域抽出の処理における精度も飛躍的に向上し、より自然な選手の動きが実現される。また、表示の際の選手ビルボードの品質も向上するため、自由視点映像を拡大し選手に近い視点で鑑賞するといった用途にも利用可能となる。

以降では、合成用データを生成する各処理の詳細について述べる。

3.1 カメラパラメータ算出

各カメラ映像と実空間との幾何学的対応関係に基づき、カメラごとの位置・姿勢などをパラメータとして算出する。例えば、対象がサッカーである場合(図2)、実空間上のフィールドサイズが規格化されているため、フィールド上の白線の交点座標から上記パラメータを算出することが可能である。また、本処理は固定カメラで撮影された映像の入力を前提としているため、撮影中に一度行えば充分である。

本システムでは、映像送信元が映像内の特徴的な点(図2の例では白線の交点)をマウスなどで選択することで正確なカメラパラメータを算出しているが、前述のとおり一度の事前処理で充分であるため、送信元への負担は比較的軽微である。

3.2 選手領域抽出

次に、映像内の選手の領域を抽出する処理を行う。選手領域の抽出は、最終的な3D空間上の選手の座標などの見え方に大きく影響する処理である。前述のとおり入力映像

の高解像度化に伴い本処理の精度は向上するため、8K/4K映像では、サッカーのような遠距離撮影映像でも選手を精度よく抽出し自然な映像を生成することができる。

ここで、本システムはスポーツ映像を対象としているため、選手は映像内においてほぼ常時動き続けていることを前提としている。その一方で、選手のいない時刻の映像を利用することで動物体のない静的な映像(背景映像)を作成しておくことが可能である。

この背景映像と実際のスポーツ映像との差分をとり、閾値処理することで選手を抽出することが可能となる。実際の映像では照明環境に依存して、選手の足元に影などのノイズが発生し悪影響を及ぼす可能性があるが、これに対しては、影の領域の輝度のヒストグラムなどをチェックすることで除去する方法を提案している⁷⁾。最終的にこれら選手領域を2値マスク画像として得る。

また、この抽出された選手領域に対して、前節で算出したカメラパラメータを利用することで、当該領域がフィールド上のどの3D座標にマッピングされるかを算出し、選手座標とする。

3.3 選手追跡および移動軌跡の平滑化

前節までに算出された選手座標について、同一選手の座標位置の時間変化の平滑化を行うことで、より自然な選手の移動を表現する。このためには、同一選手に同一IDを付与する人物追跡技術が肝要となる。また、この際に特に課題となるのは、選手同士の映像内での重なりあい(オクルージョン)が発生した場合である。

従来さまざまな人物追跡技術が提案されているが、著者らは処理コストの観点から、前節で得られた各カメラの選手領域画像ベースの方法をとっている。一例としては、ある時刻の単カメラ映像における選手領域の近傍に対してパーティクルフィルタ(ランダムサンプリングによるモンテカルロ近似)を用いて、次時刻の選手の位置を推定する方法を提案している⁸⁾。オクルージョンが発生した場合は、他カメラの追跡結果を参照して、結果の補正を行っている。その他、選手が比較的大きく映っている映像に対しては、あらかじめフィールド上へ投影したすべてのカメラの選手領域により3D上の選手の座標の候補を算出し、この候補に対してラベリングによる同一選手の識別を行う方法もとっている⁹⁾。

このように得られた同一IDの選手座標に対して、近傍数フレームの重み付き和をとることで移動軌跡の平滑化を行う。しかしながら、選手が急に逆方向に移動する映像(例:反復横跳び)の場合には、平滑化の効果が悪影響を及ぼす場合も考えられるため、映像の内容に応じて当該重みを変更するなどの対応を想定している。

以上の処理を通して得られた、カメラパラメータや選手領域の2値マスク画像、選手座標を合成用データとして送信し、これらを用いて前述のとおり受信側での表示を実現している。

4. むすび

本稿では、8K/4K映像を利用し実映像と自由視点映像の双方を同時に視聴可能とするハイブリッドナビゲーションシステムを紹介した。

当該システムでは、8K/4KをMV-HEVCによって圧縮したストリームを生成するとともに、あらかじめオフライン処理で算出しておいた自由視点映像生成に必要な合成用データを付与し送信する。受信側では、映像をリアルタイムでデコードするとともに付与された合成データを基に、自由視点映像生成をリアルタイムで行う。

自由視点映像合成に関する要素技術としては、「映像からの選手抽出」、「選手の移動軌跡平滑化のための追跡」が挙げられ、その概要と一例を説明し、最終的な映像品質に大きく寄与するものであることを述べた。また、8K/4K映像を用いることで、映像品質が飛躍的に向上することを述べた。

今後の課題としては、自由視点映像の完全リアルタイム合成に向けた各要素技術のさらなる高精度化やパン・チルトを伴う動的なカメラ映像での合成技術の確立が挙げられる。

(2016年5月31日受付)

〔文 献〕

- 1) 谷本正幸: “自由視点テレビ”, 映情学誌, 55, 12 (2001)
- 2) 古山孝好, 北原 格, 大田友一: “スタジアムの自由視点ライブ中継が可能な3次元映像システム”, 3次元画像コンファレンス講演論文集, pp.225-228 (2003)
- 3) K. Hayashi and H. Saito: "Synthesizing Free-Viewpoint Images from Multiple View Videos in Soccer Stadium", International Conference on CGIV, pp.220-225 (2006)
- 4) P.E. Debevec, G. Borshukov and Y. Yizhou: "Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping", Proc. of Eurographics Workshop (1998)
- 5) 河村 圭, 内藤 整: “MV-HEVC準拠4K映像リアルタイムソフトウェアコーデック”, 画像符号化シンポジウム, P-3-04 (2014)
- 7) Q. Yao, H. Sankoh, H. Sabirin and S. Naito: "Accurate Silhouette Extraction of Multiple Moving Objects for Free Viewpoint Sports Video Synthesis", International Workshop on MMSP, pp.1-6 (2015)
- 8) 三功浩嗣, 内藤 整: “選手領域の抽出と追跡によるサッカーの自由視点映像生成”, 映情学誌, 68, 3, pp.J125-J134 (2014)
- 9) 野中敬介, 内藤 整: “近距離撮影型自由視点映像のためのオクルージョンに頑健な人物追跡手法の一検討”, 映像メディア処理シンポジウム, I-3-11 (2015)

野中 敬介 2010年、東京工業大学大学院理工学研究科集積システム専攻修士課程修了。2014年、同大学大学院総合理工学研究科物理情報システム専攻博士後期課程修了。同年、KDDI(株)に入社。以来、同社研究所にて自由視点映像生成、コンピュータビジョンの研究に従事。現在、(株)KDDI研究所超臨場感通信グループ研究員、博士(工学)、正会員。

内藤 整 1994年、早稲田大学理工学部電子通信学科卒業。1996年、同大学大学院修士課程修了。同年、国際電信電話(株)に入社。以来、同研究所にて、主にHDTV、SDTVの動画画像符号化および伝送方式の研究・開発に従事。現在、(株)KDDI研究所超臨場感通信グループリーダー、工学博士、正会員。