

# 「夢のアリーナ」を実現する 映像解析技術

## Image Analysis Technologies to Realize “Dream Arenas”

● 石井大祐      ● 秋山深一      ● 都市雅彦      ● 山本琢磨

### あらまし

東京2020オリンピック・パラリンピック競技大会の開催を控え、日本ではスポーツの活性化に向けた取り組みが活発化している。富士通では、公益財団法人日本バスケットボール協会(JBA)および公益社団法人ジャパン・プロフェッショナル・バスケットボールリーグ(B.LEAGUE)をICTで支援すべく、スマートアリーナソリューションの実用化に向けた開発を進めている。本ソリューションの目的は、選手とチームを強化する、視聴者を楽しませるという二つである。前者に対しては、プレー中の選手の時々刻々と変化する位置情報を漏れなく捉えるために、特に選手が重なる密集時の対応を強化した高精度なモーショントラッキング技術を開発した。また後者に対しては、プレー中のフィールド内から見た映像や360度あらゆる角度からのリプレイ映像など、物理的な制約を受けることなく自由な視点からの映像を視聴者に届ける自由視点映像生成技術を開発した。

本稿では、これら二つの映像解析技術の特長について述べる。

### Abstract

Gearing up for the Olympic and Paralympic Games Tokyo 2020, various initiatives are gaining impetus to make sports more popular in Japan. Fujitsu pursues the development for practical adaptation of a smart arena solution to make ICT work for the Japan Basketball Association (JBA) and the Japan Professional Basketball League (B.LEAGUE). This solution aims to achieve two objectives: strengthening players and teams, and entertaining spectators. In terms of strengthening players and teams, we have developed a high-precision motion tracking technology with enhanced performance in following players in a game, particularly when they are bunched together. In terms of entertaining spectators, we have developed free-viewpoint image rendering technology that enables the generation of video from any point of view without physical restrictions. For example, it allows for the creation of videos seen from inside the playing field and 360-degree multi-viewpoint replay videos. This paper presents the benefits of these two types of image analysis technology.

## まえがき

東京2020オリンピック・パラリンピック競技大会の開催を控え、日本ではスポーツの活性化に向けた取り組みが活発化している。富士通では、公益財団法人日本バスケットボール協会（JBA）および公益社団法人ジャパン・プロフェッショナル・バスケットボールリーグ（B.LEAGUE）をICTで支援するために、スマートアリーナソリューションの実用化に向けた開発を進めている。<sup>(1)</sup>本ソリューションの目的は、選手とチームを強化することと、視聴者を楽しませることの二つである。

まず、選手とチームの強化に関しては、スター選手の育成に加え、チームの強化による競技力向上を目指している。その実現のために、選手の様々なプレーデータやチームの成績をまとめた統計情報（スタッツ）を活用し、科学的な観点から効果的なコーチングを行う仕組みが必要となる。

スタッツの中でも、選手の位置情報はチームのフォーメーション解析や選手のプレー状況の把握に不可欠であり、最も重要な情報である。プレー中の選手の位置は時々刻々と変化するため、これまでの人手による作業では粗い情報しか得られないという問題があった。そこで、富士通研究所はフィールドを取り囲むように設置した複数のカメラで撮影した映像を用いて、選手の位置情報を漏れなく捉える高精度なモーショントラッキング技術を開発した。

次に、視聴者を楽しませるということに関しては、高い臨場感のある映像を作り出すことができる自由視点映像生成技術の研究開発を行っている。自由視点映像生成技術とは、複数のカメラ映像を入力として空間の3次元構造を復元し、実際にはカメラのない任意の視点からの映像を生成する技術である。本技術をスポーツ会場に適用することで、プレー中のフィールド内から見た映像や、360度あらゆる角度からのリプレイ映像など、物理的な制約を受けることなく自由な視点からの映像を視聴者に届けることが可能となる。

本稿では、モーショントラッキング技術と自由視点映像生成技術について述べる。

## モーショントラッキング技術

本章では、プレーする選手の時々刻々と変化する位置情報を漏れなく捉える高精度モーショントラッキング技術について述べる。

### ● モーショントラッキングシステムの現状と課題

屋内スポーツ選手のトラッキングシステムは、選手に発信機を取り付けて位置を計測するビーコン型と、カメラで撮影した映像を解析して位置を計測する映像解析型の二つのアプローチがある。現在、開発中のトラッキングシステムでは、プレー中の選手への影響を避けるために、選手への機器装着を必要としない映像解析によるアプローチを採用することとした。

映像解析によるトラッキングでは、選手が隠れることなくカメラに写っている必要がある。天井へのカメラの設置は、フィールド全体を真上から見下ろすことができ、映像中の選手が重なることなく撮影できるため、トラッキングには適している。一方、大がかりな設置工事が必要となるため、様々な施設への導入には障壁となる。そこで、選手同士の重なりが発生する問題はあるものの、設置が容易な壁面にカメラを設置し、フィールドを斜め上から撮影する配置とした。このとき、フィールドを取り囲むように8台のカメラを配置し、様々な方向から選手を撮影することで、選手同士の重なりを最小限に抑えるマルチカメラシステムとした。

本方式では、あらかじめ選手が写っていない背景画像を用意しておき、撮影したカメラ映像から背景映像と異なる部分として抽出した体のシルエット形状を処理することで選手を検出する方式を基本とする。体全体から得られる特徴的な形状情報を用いることで、個々の選手の重なりが少ない状況では安定してトラッキングできる。

しかし、プレーの要となるゴールに近いエリアでの攻防など、選手同士の距離が近く映像上で重なりが多発する場合には、マルチカメラで様々な方向から撮影したとしても複数の選手のシルエットが重なり合ってしまう。このような場合、それぞれの選手を正確に分離できず、トラッキング中の選手間の入れ替わりや多重検知、ロストなどが

発生する。そのため、選手が密集する状況でも正確にトラッキングできる高性能なトラッキング技術が必要となる。

### ● 密集状況への対応方針

密集状況に対応するために、選手のシルエットに基づくトラッキング方式を基本として、トラッキングの状況から密集の度合いを推定する。密集と判断された場合には、その部分だけを密集時に特化したトラッキング方式で処理するハイブリッド方式とした。以下では、密集時に特化したトラッキング方式について述べる。

密集時には、体全体ではほかの選手との重なりが大きくなるが、頭部に限れば比較的小さい。そこで、密集時には頭部を対象として選手をトラッキングする方針とした。しかし、頭部は体全体と比較して小さく単純な形状であることから、識別対象としての特徴が少ない。そのため、単純にトラッキング対象を切り替えただけでは安定してトラッキングできない。また、密集時には近くにも別の選手の頭部が存在することから、周囲の影響をうまく避け、トラッキングを安定化させるための工夫が必要となる。

### ● マルチカメラ選択型高精度トラッキング技術

密集時の処理として、マルチカメラ選択型高精度トラッキング技術を開発した。この技術では、複数の方向から撮影するマルチカメラの特性を最大限に活かして、ほかの選手との重なりが少なくトラッキングに適した映像を選び出して処理する。

具体的には、複数の選手に対して安定したトラッキング処理が期待できる順序を決定する「トラッキング順序スケジューリング」と、各選手の頭部トラッキングで、マルチカメラ映像での頭部の見え方の良否をスコア化し、優先度を付けて処理する「マルチカメラ優先度評価」の二つの処理技術を開発した。

#### (1) トラッキング順序スケジューリング

カメラから隠れることなく観測される選手は、安定して位置を特定しやすい。そこで、映像中での選手の隠れ度合いから、安定して頭部位置を検出できそうな順序を定めておき、安定して検出できる選手の頭部から順に処理を進める。既に検出した選手の頭部位置は、以降の処理領域から除外する。これによって、隠れてしまった部分が大きく、

そのままでは検出が困難な選手であっても、頭部の存在領域を大幅に絞り込めるため、安定した検出が可能となる。

検出順序を判断するための指標は、カメラと検出対象との間に生じるほかの選手による遮蔽、および検出対象とその背後や周囲との映像としての類似度の二つの観点に基づいて決定する。選手ごとに、トラッキング結果を用いて移動先を予測することで、ほかの選手および背景との位置関係を推定し、全体の処理順序を決定する。

#### (2) マルチカメラ優先度評価

選手個人に着目した場合も、カメラごとに見え方は異なる。そこで、選手個人の検出に当たり、条件の悪いカメラを除外し、条件の良いカメラのみを用いて3次元的に位置を検出することで、更に検出を安定化させる。

マルチカメラ優先度評価では、(1)と同じ指標を用いて、隠れてしまった部分が大きく条件の悪いカメラを除外し、条件の良いカメラの中で指標に応じた重みを付けて検出する。この動的なカメラの選択と重み付けの概要を図-1に示す。検出時には、使用する各カメラの映像上で頭部らしさを評価するとともに、カメラ間の幾何学的関係からなる空間的な制約のもとで総合的に評価することで、3次元空間上の位置を決定する。

### ● トラッキング性能の検証

前節で紹介したマルチカメラ選択型高精度トラッキング技術をシステムに実装し、性能評価を行った。密集シーンを処理した結果の例を図-2に示す。図左側の2次元投影は、各選手のフィールド上の位置と、その検出に使用するカメラの対応関係である。右側はその時刻での実際のカメラ映像を示しており、各選手の状況に合わせて、検出に使用するカメラが選択されていることが分かる。

今回の評価では、8台のカメラを設置して撮影した試合シーンから、バスケットボールの攻撃制限時間(24秒)で約18回に相当する420秒の区間を抜き出し、テスト映像として使用した。本テスト映像において、密集発生の割合は約60%であった。

検証では、当該シーケンスを10秒ごとの区間に区切り、人物ごとにトラッキング成否を判定している。検証結果を図-3に示す。今回開発したマルチカメラ選択型高精度トラッキング技術では、密

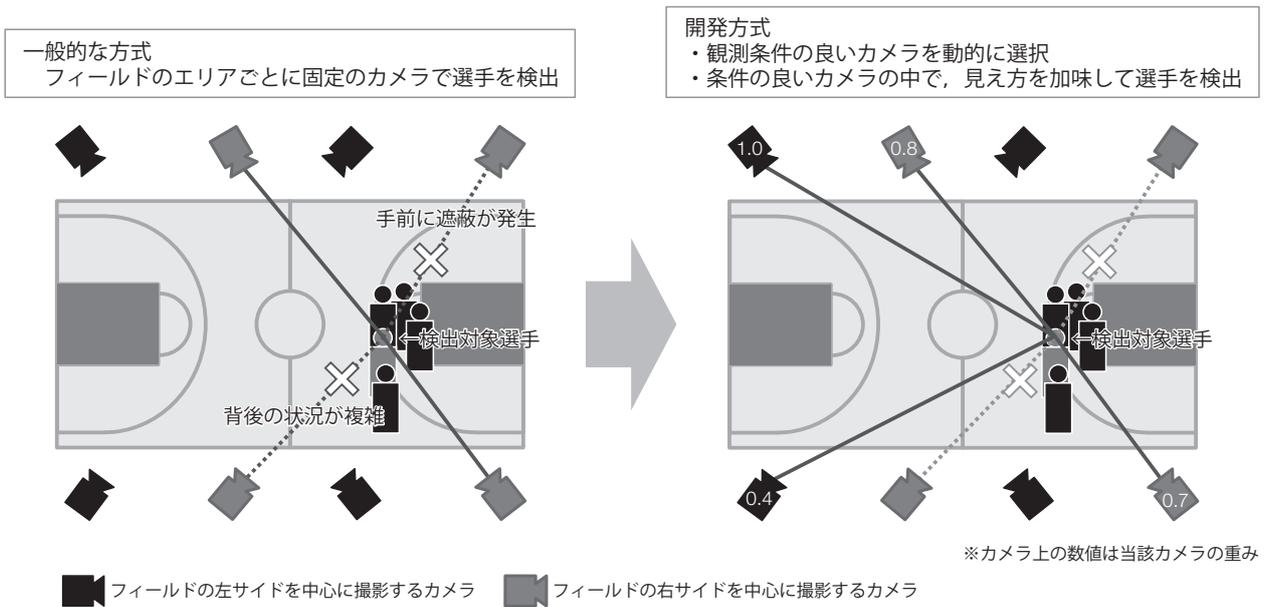


図-1 選手ごとの動的カメラ選択と重み付け

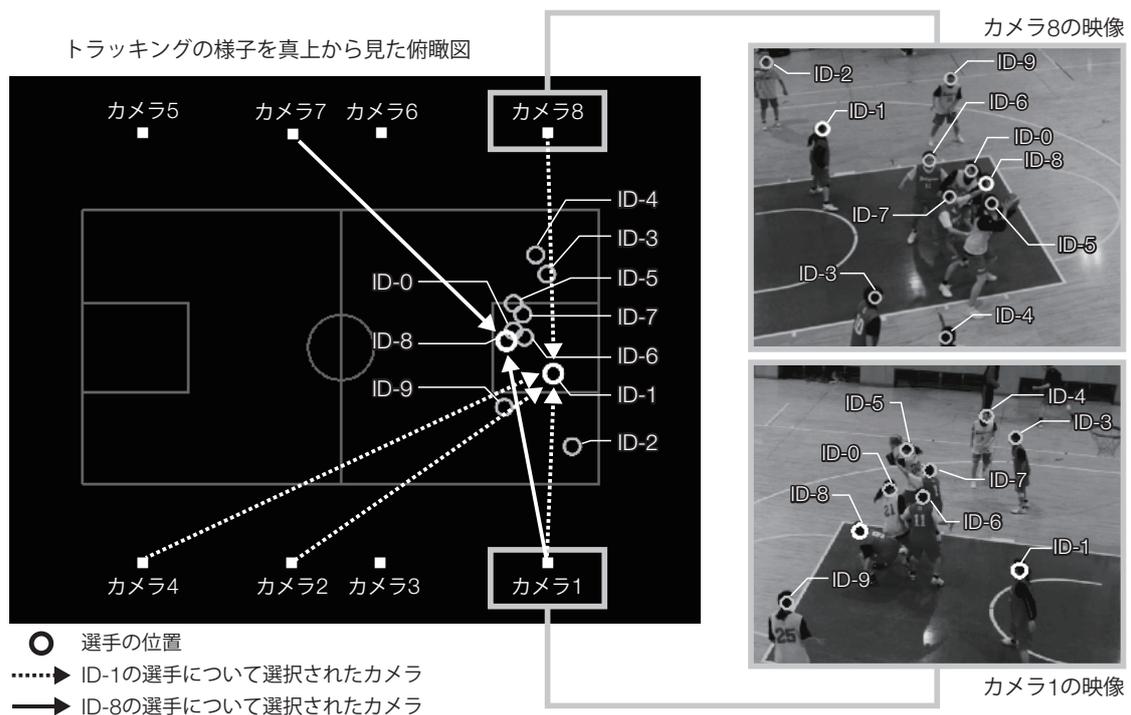


図-2 マルチカメラ選択型高精度トラッキングによる密集時のカメラ選択とトラッキング結果の例

集区間のトラッキング成功率が大きく改善され、全区間平均では31%の向上となる、94%のトラッキング成功率が得られた。これらの結果から、開発方式では、ゴール付近など選手同士が密集するシーンにおいても、安定して選手をトラッキングできることが確認できた。

● まとめ

スポーツシーンからスタッツの自動生成実現に向け、課題となっていた選手の密集時でも安定して位置情報を取得できるマルチカメラ選択型高精度トラッキング技術を開発した。これにより、選手同士が密集したシーンであっても高精度に位置

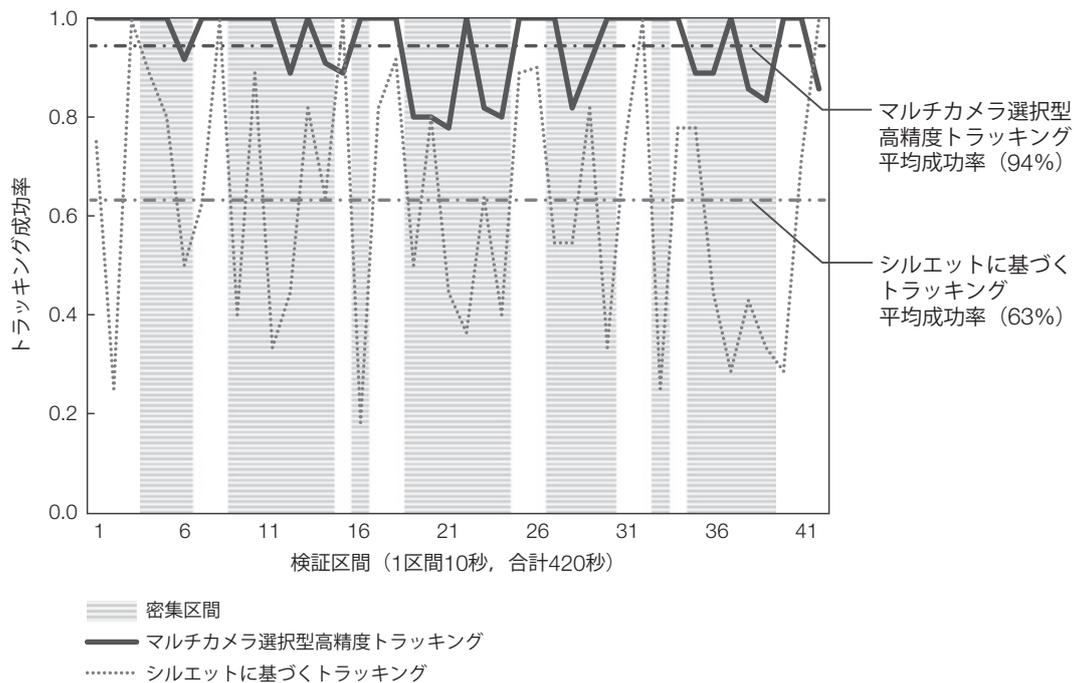


図-3 マルチカメラ選択型高精度トラッキングの精度評価結果

情報が得られることから、入れ替わりのない自動モーショントラッキング実現の見通しを得た。今後は、実運用に向けて、スタッツ自動生成システムの構築と機能面の拡充を図る計画である。

### 自由視点映像生成技術

本章では、フィールド内の任意の視点からの映像を生成する自由視点映像生成技術について述べる。

#### ● 自由視点映像生成技術の動向

自由視点映像生成技術は1990年頃から研究が始められ、映画<sup>(2)</sup>やCMの映像効果として最初に実用化された。近年では、スポーツ中継のリプレイ映像<sup>(3)・(4)</sup>にも適用されている。代表的な実現手法として、多視点のカメラ映像をフレームごとに切り替えて視点移動の映像を作る手法<sup>(5)</sup>や、映像から対象者の3次元構造を復元し任意の視点から見た映像を再構成する手法<sup>(6)・(7)</sup>が知られている。

前者は比較的簡易な処理で高い画質が得られるが、滑らかな視点移動のためには多数のカメラが必要であり、視点移動の自由度が低いという課題がある。後者は視点の自由度が高く、対象の3次元情報を復元するため、CGの合成処理との親和性が

高い。しかし、高い画質を得るためには、人物と背景を明確に分離する必要がある。様々なスポーツ施設に導入するためには、撮影環境や背景に影響を受けない安定した分離手法の確立が不可欠である。

富士通研究所では、3次元情報を解析し様々なデータ利活用サービスへの発展も期待できる後者の手法について検討している。

#### ● 自由視点映像生成技術の処理の流れ

富士通研究所が開発している自由視点映像システムでは、観客席後方や壁面部にコートを取り囲むように複数のカメラを配置している(図-4)。このように設置すると、フィールド内の選手や審判などの人物やボール、ゴールなどの構造物は、ほぼ全てのカメラの視野に入る。一方、観客席などの背景に相当する部分は、対面するカメラからのみ撮影されることとなる。

撮影されるカメラの台数や方向が異なるため、映像を分離しそれぞれに応じた3次元復元処理を適用する方法を採っている。前景となるオブジェクト(選手やボールなどの物体)の3次元構造復元には、比較的安定して精密な形状を求めることができる視体積交差法を適用し、背景部の3次元復元に

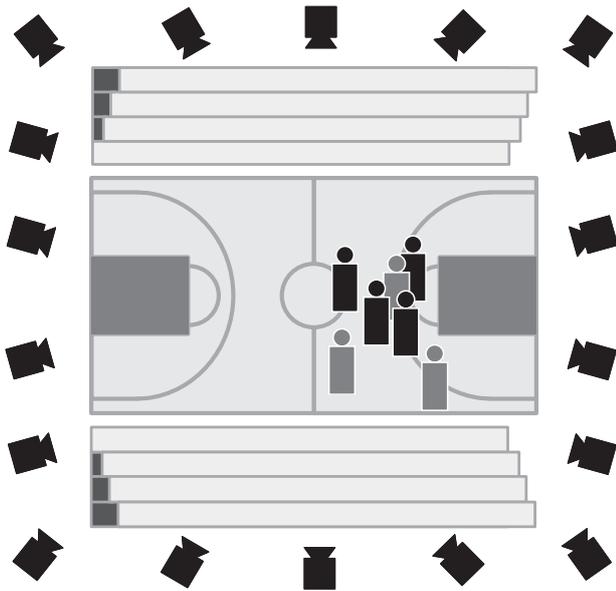


図-4 自由視点映像システムのカメラ配置

は、ステレオ視の技術を応用した方法を適用している。

#### ● 従来の前景背景分離技術の課題

自由視点映像の処理フローでは、前景背景分離で抽出された前景マスク（前景となるオブジェクトの領域）が後段の処理で用いられるため、前景背景分離の精度が最終結果に大きな影響を与える。そのため、前景背景分離の高精度化が、高画質な自由視点映像生成には不可欠である。

前景背景分離は、動画像処理分野において重要な基礎技術であり、様々な従来法がある。例えば、緑や青などの単一色のシートを背景に撮影し、背景色を除外することで前景を分離するクロマキー<sup>(8)</sup>や、あらかじめ撮影した背景画像と現在撮影した対象画像との差分を取ることで前景を分離する背景差分法<sup>(8)</sup>などがある。

クロマキーは、テレビスタジオなどで、実際には存在しない人物や物体を重畳表示した合成映像の生成に用いられる。背景差分法は、映像監視における移動体の検出などで用いられる。前景背景分離の従来法であるクロマキーや背景差分法をスポーツの自由視点映像生成に適用しようとすると、以下の二つの問題がある。

まず、バスケットボールやサッカーなどのスポーツでは所定のフィールドでプレーする必要があるため、テレビスタジオのように単一色のシートを

設置することは難しく、クロマキーは適用できない。また、バスケットボールなどでは、選手のユニフォームと背景になる床や壁などが同一色になる場合が多いため、背景差分法では選手の前景マスクに欠損が生じる。

#### ● Deep Neural Networkを応用した前景背景分離方式の開発

そこで、近年AI（人工知能）技術として目覚ましい成果を上げているDNN（Deep Neural Network）<sup>(9)</sup>を用いて人物を検出することで、高精度な前景背景分離を実現する方式を開発した。具体的には、画像処理向けのDNNであるCNN（Convolutional Neural Network）<sup>(9)</sup>を用いて、各画素それぞれが人物領域であるか否かを判別することで、高精度に人物領域を抽出する。更に、前景マスクが空間的に滑らかになるような制約のもと、グラフカット<sup>(10)</sup>を用いて全体最適化する。

CNNでは、複数の畳み込みフィルターにより、色以外にもエッジや全体の概形などの人物の様々な特徴が抽出できる。このため、背景と類似の色を持つ人物であっても、はっきりと検出できる。更に、グラフカットを用いた全体最適化の効果により、細かなノイズを排除した高品質な前景マスクが抽出できる。

#### ● 前景背景分離性能の評価

バスケットボール会場での運用を想定し、体育館に4Kカメラを18台設置し、評価映像の撮影を行った。カメラはグローバルシャッター方式を採用し、各カメラの撮影タイミングを同期させることで、全てのカメラが同じ瞬間を撮影できるようにしている。カメラとサーバの間は光ファイバーで接続されており、映像はサーバに集約され映像処理が行われる。

実際のB.LEAGUE試合会場では、ゴール前のペイントエリアにチームカラーが配色され、ユニフォームと同色になることがある。今回の実験では、その状況を再現するために、ペイントエリアと類似した色のユニフォームを着用した選手に模擬試合を行ってもらい、その様子を撮影した。

撮影した評価映像を基に、色分布の類似性を用いた背景差分法<sup>(8)</sup>を従来方式として採用し、開発方式と比較した。図-5 (a) を対象画像としたときの前景マスクの比較を、図-5 (b) に示す。選手



図-5 開発した前景背景分離方式の効果

のユニフォームと背景が同色（赤色）であるため、従来方式では選手のユニフォームの一部が欠損している。これに対して、開発方式では欠損が生じていないことが分かる。

この前景マスクを用いて視体積交差法による3次元復元処理を行い、撮影カメラとは異なる視点の映像を生成した結果を図-5 (c) に示す。従来方式では、前景マスクの欠損によって、手前の選手にアーチファクト<sup>(注)</sup>が発生している。また、奥の人物の左足もズボンと背景の色が類似しているため、欠損していることが分かる。これに対して、開発方式ではアーチファクトが少なく高画質である。

更に、前景背景分離により抽出した前景マスクと、精緻な手作業により作成した正解マスクをF値（適合率と再現率の調和平均）により比較した。従

来方式では0.67であったのに対して、開発方式により0.77に改善し、開発方式の効果が定量的に確認できた。開発方式により、撮影カメラとは異なる様々な視点の映像を生成した結果を図-5 (d) に示す。

● まとめ

自由視点映像を生成する際に行う選手映像の切り出しにおいて、DNNによる人物検出を応用することで、分離精度を高精度化する技術を開発した。本技術の開発によって、複数の選手が重なるようなシーンや、チームカラーを配色した会場においても安定した自由視点映像の生成が可能となった。

自由視点映像の生成処理では、空間の3次元情報を復元しているため、CGやVR (Virtual Reality) とも親和性が高い。ヘッドマウントディスプレイを用いることで、あたかもフィールド内にいるかのような視聴体験を提供することも検討している。

(注) 画像処理で生じたエラーやひずみ。

む す び

本稿では、B.LEAGUEが提唱している「夢のアリーナ」の実現を目指して開発した、モーショントラッキング技術と自由視点映像生成技術について述べた。今回紹介したスマートアリーナソリューションは、バスケットボール業界の振興における課題を富士通および富士通研究所の映像解析技術で解決し、時間と空間と動画を自由にスタッツにひも付けられるようにするものである。

今後は、モーショントラッキング技術と自由視点映像生成技術の双方を活用し、実際に選手がプレーするフィールドに入り込んでいく。それとともに、選手個々のプレーデータも確認しながら観戦するといった、俯瞰的視点と臨場的視点の融合による新たな観戦体験の提供を可能にしていく。このように、競技に対する知的探求心を満足するだけでなく、一つひとつのプレーの意味を誰もが容易に理解できるような、夢のアリーナ実現を目指している。

参考文献

- (1) 富士通：日本のバスケットボールの未来をICTでサポート。  
<https://pr.fujitsu.com/jp/news/2016/09/21.html>
- (2) マトリックス。  
<https://www.warnerbros.com/matrix/>
- (3) KDDI：au BLUE CHALLENGE — au Future Stadium。  
<https://www.au.com/pr/soccer/>
- (4) Intel：インテルfreeDテクノロジー。  
<https://www.intel.co.jp/content/www/jp/ja/sports/technology/intel-freed-360-replay-technology.html>
- (5) 4DReplay。  
<http://4dreplay.com/>
- (6) T. Kanade et al.：Virtualized Reality：Constructing Time-Varying Virtual Worlds From Real World Events. Proc. of IEEE Visualization, p.277-283, Oct. 1997.
- (7) W. Matusik et al.：Image Based Visual Hulls. ACM SIGGRAPH2000, p.369-374, 2000.
- (8) 奥富正敏編：デジタル画像処理 [改訂新版]. 画像情報教育振興協会, 2015.

- (9) I. Goodfellow et al.：Deep learning. MIT press, 2016.
- (10) Y. Y. Boykov et al.：Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. ICCV 2001, 2001.

著者紹介



石井大祐 (いしい だいすけ)

(株)富士通研究所  
フロントテクノロジー研究所  
映像の認識・理解に関する研究に従事。



秋山深一 (あきやま しんいち)

富士通(株)  
スポーツ・文化イベントビジネス推進本部  
Athlete Sensing Solutionの企画に従事。



都市雅彦 (といち まさひこ)

(株)富士通研究所  
フロントテクノロジー研究所  
画像処理および3次元空間認識に関する研究開発に従事。



山本琢磨 (やまもと たくま)

(株)富士通研究所  
フロントテクノロジー研究所  
コンピュータビジョンに関する研究に従事。