AR/VR 技術によるライブ映像演出

Stage Visual Direction with AR/VR Technologies

真鍋大度　花井裕也

Abstract

昨今のディープラーニング技術の発展により，入手コスト，設置コストの高いモーションキャプチャや 3D スキャナなどを使用せず，安価なカメラを用いてオブジェクト認識や人物のシルエット認識などが可能となった．急速なスピードで様々なプロダクトがリリースされ，放送やライブパフォーマンス向けのリアルタイム映像演出が進化しつつある．現時点では映像クリエイターが扱える環境とは言い難いが，ここ 1，2 年で機械学習技術を用いた様々なプロダクトがリリースされて民生化し AR/VR ライブ演出も進化していくだろう．一方でライゾマティクスでは小さなシアターから大型のフェスティバルまで様々な条件に応じたシステムを開発し，Perfume や ELEVENPLAY のステージパフォーマンス，大きなものではリオ 2016 大会の閉会式におけるフラッグハンドオーバーセレモニーなどに実装してきた．今回はライゾマティクスが行った映像演出の裏側にフォーカスしながら映像技術の進化を考察する．

キーワード：AR，VR，機械学習，映像演出，ライブ演出技術

1. Augmented Reality（AR）/Virtual Reality（VR）映像演出の手法について

ステージにおける AR/VR 演出手法は大きく二つに大別できる．一つは多数の観客が肉眼で体感可能なことを想定した手法，もう一つが会場の LED スクリーンやテレビ中継など画面上での体験を想定した手法である．例えば，リオ 2016 大会の閉会式におけるフラッグハンドオーバーセレモニーでは，前者を床面プロジェクションマッピングの映像演出（図 1）として，後者を 33 競技の AR 映像演出（図 2）として実装した．

更に一般的な例を挙げると，前者の具体例としては紗

図 1　リオ 2016 大会の閉会式におけるフラッグハンドオーバーセレモニーのプロジェクションシーンのシミュレータ

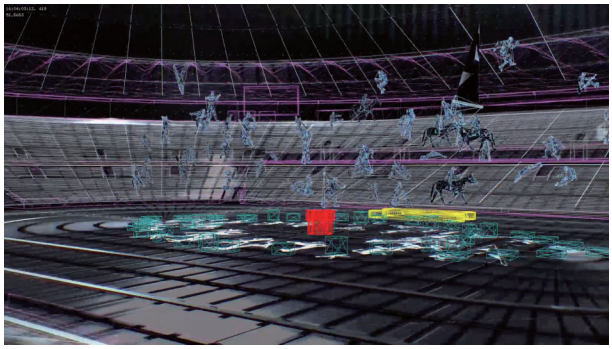

図 2　リオ 2016 大会の閉会式におけるフラッグハンドオーバーセレモニーの AR シーンのプレビズ

本会誌では，用語は①文部省（文部科学省）学術用語集電気工学編，②本会編の改訂電子情報通信用語辞典，③本会編のエンサイクロペディアハンドブック，に基づき統一している．
本稿中の「大型」は，上記①に従うと「大形」であるが，ここでは著者の希望により「大型」で掲載した．

真鍋大度　花井裕也　ライゾマティクス
Daito MANABE and Yuya HANAI, Nonmembers（Rhizomatiks, Tokyo, 150-0011 Japan）.
電子情報通信学会誌　Vol.103 No.6 pp.564-570 2020 年 6 月

本記事の著作権は著者に帰属します．

幕スクリーン[1]や身体へのプロジェクションマッピング[2]，ハーフミラーを用いたペッパーズゴースト[3]，複数プロジェクタを用いたボリュメトリックプロジェクション[4]~[6]などが挙げられる．

後者の具体例としては，カメラ間の自由視点映像生成[7]，モーフィング[8],[9]などが挙げられる．要素技術として顔認識や姿勢推定技術が組み合わされて使用されることが多い．なお，ライゾマティクスではこれらの独自開発した映像技術を総称して Seamless Mixed Reality（Seamless MR）と呼んでいる．

また両者の中間的技術として，CAVE[10]などの体験者視点依存の立体錯視技術が挙げられる．ライゾマティクスでは，CAVE を複数カメラのスイッチングや可動式の LED スクリーンに対応させるなど中継に特化させる形で独自開発した技術を Dynamic VR[11]と呼んでいる．以下，各項目の演出技術について，ライゾマティクスで開発を担当したプロジェクトの事例を交えて紹介する．

2. 多数の観客が肉眼で一度に体験することを想定した演出手法

本手法群は，ステージ上でパフォーマンスを行う演者に対して多数の観客が直接視認可能な映像効果を重畳することを目的とする．特徴としては，映像効果は平面的か，立体であったとしても視認性は高くない．そのため多くの場合，実施可能な会場は 1 方向にのみ客席を持つ劇場型の会場に限られる．また，演者の動きをセンシングして映像効果に反映するといったリアルタイム性が問われる演出の場合，できるだけシステム全体の系の遅延を削減する必要がある．

2.1 プロジェクションマッピング

プロジェクションマッピングというと，イベントなどで建造物に対して映像を投影するというイメージが強いと思われる．これは 2008 年辺りからプロジェクタの解像度や輝度の向上とともにスクリーン以外の物体に映像を投影するということが現実的になってきたためである．一方で，センシング技術の発展とともに，ダンス作品において，身体への動的なプロジェクションマッピングも ARToolKit の生みの親である Hirokazu Kato の技術開発の下，Ars Electronica Future Lab，Klaus Obenmeier[2]などによって 2004 年頃から行われてきた．また紗幕スクリーンに映像を投影することでダンサーとのインタラクションを可視化した Adrien M & Claire B[1]らの作品もある．

ライゾマティクスでは，2012 年の Perfume at Cannes Lions においてデプスセンサ Xtion で再帰性反射材で作られた衣装と身体のパーツ分類を行い，別々の映像を投影するプロジェクションマッピング表現を実現した（図 3）．

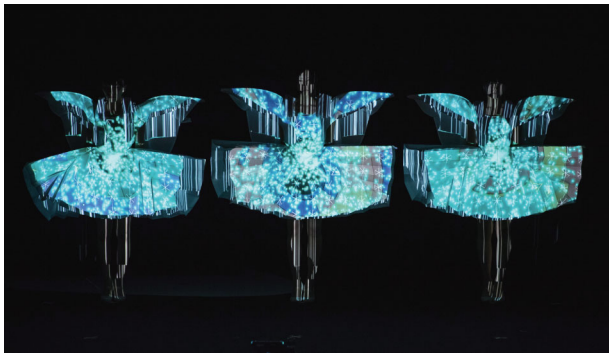

図 3　Perfume at Cannes Lions（©Cannes Lions, Dentsu. inc）

また，後述する Dynamic VR と組み合わせ，移動するスクリーンに対して遅延なく映像を投影する技術も開発した．

2.2 ボリュメトリックプロジェクション

本手法は 3D 表示装置，あるいは空中三次元結像装置の一種である．こういった装置の研究は古くからなされており，様々な手法が存在する．映像を表示する平面を高速に回転させ立体映像表示する装置[12]や，水滴を利用して空間を充填し表示装置として使用する例[13]などが挙げられる．ステージ作品で使用するには，これらの手法は大型化が難しく，表示領域に入り込めないため演者とのインタラクションが難しいといった問題がある．一方，アート作品では Kimchi and Chips の Light Barrier[4]，WOW の Light of Birth[5]など複数のプロジェクタとスモークを用いた作品の先行事例が存在する．これは，複数プロジェクタの光源位置を計算し想定の空間座標に光を照射することと，スモークや霧などの煙霧媒体を空間内に充満させることでチンダル現象を誘発させ，特定空間に像を表出させるというものである．大型化が容易であり，演者とのインタラクションも可能であるが，視認性が得られにくいという問題もある．

2017 年に発表されたライゾマティクスと ELEVEN-PLAY のコラボレーションによるダンスインスタレーション作品 phosphere[6]（図 4）では，本手法を拡張し，独自のキャリブレーションシステムを開発することで光学式モーションキャプチャシステム OptiTrack[14]と連携（図 5）．実際のダンサーと，現実には存在しない光のダンサーとの共演を実現した．

3. スクリーン上で体験することを想定した演出手法

本章では，会場 LED スクリーン，テレビ中継，ライ

図 4　“phosphere at Sónar Barcelona”（Rhizomatiks Research × ELEVENPLAY，2017）

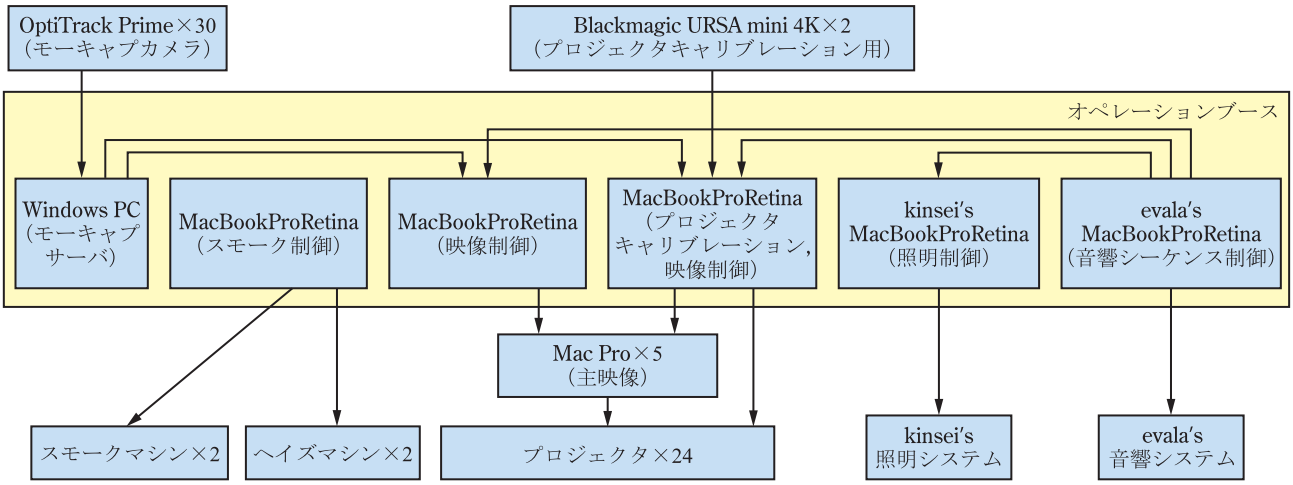

図 5　phosphere システム

ブストリーミングでの映像体験に特化した AR/VR 映像演出の手法について概説する．本手法群は，リアルタイムに映像の処理ができればどのような効果を乗せることもできるため自由度が高い．また，近年ではディープラーニング，画像解析技術の発展により，できることの幅も広がりつつある分野である．

3.1　Seamless Mixed Reality（Seamless MR）

本節ではカメラ間のモーフィング，自由視点映像生成を特徴とする Seamless MR について紹介する．

モーフィングとは，二つの画像の間をスムーズにつなぐ技術のことであり，古くから研究開発されている分野である．1991 年には Michael Jackson の Black or White の MV などで表現自体は既に実用化されていたが，開発を行った Beier，T らの研究論文[8]によれば，当時は画像間の領域の対応付けは手動であった．この位置関係の対応付けを自動的に行うための研究としては Maxime Lhuillier らの研究[9]などがあり，二つの画像間で，テクスチャネスが高い領域について準密な画素の対応付けを行い，対応する平面パッチを逐次作成し，幾何拘束を導入することで破綻の少ない補間画像の生成が可能であった．

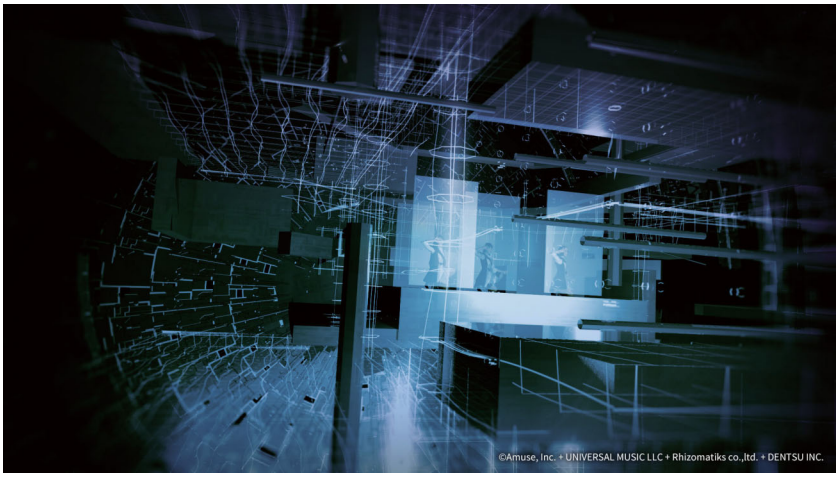

図 6　Perfume at SXSW 2015（©Amuse, Inc. + UNIVERSAL MUSIC LLC + Rhizomatiks co., ltd + DENTSU INC.）

こういったモーフィングでは，画像間の内挿補間はできても，外挿補間は想定していない．つまり二つのカメラの映像をシームレスにつなぐことを考えると，基本的にはカメラ間を直線的につないだパスでしか機能しない．その問題を解決するのが自由視点映像生成である．

自由視点映像とは，空間全体を 3D データ化することで，カメラを実際に置いていない位置・角度からの視点の映像を生成することができる技術のことを指す．リアルタイムのスポーツの中継などでは Replay Technologies 社の FreeD（現 Intel 社の True-View）[7] など，数十台のカメラを会場に設置して，画像情報から三次元データを再構成し，自由視点映像を生成する手法が用いられてきた．こういった手法では，正確な同期が可能な高価なカメラや，3D 再構成を行うための膨大な計算資源が必要である．また，パフォーマンスのステージなどのテクスチャに乏しい環境や，照明変化が大きい環境では画像情報から 3D 再構成を行うことが難しい場合が多い．そのため，こういった技術をそのままステージ作品の分野に適用するのは，コストの面でもクオリティの面でも難しい．そこで，ライゾマティクスではステージ作品に適用可能なモーフィング・自由視点映像生成技術を開発した．

大きな特徴として，事前に空間の 3D モデリングを行い，可能な限りその情報をそのまま使って空間の 3D データ化するようにした．こうすることで，計算コストを抑え，照明変化にも頑健な自由視点映像生成が可能となった．また，必要なカメラ台数も少なくて済み，カメラ自体にも正確な同期を必要としないことから，設置コスト面でも画像特徴ベースのモーフィングに比べてメリットがある．3D モデルのデータを活用したモーフィング技術であることから，本技術をモデルベースビューモーフィングと呼んでいる．

例えば，2015 年に実施した Perfume at SXSW（図 6）では，事前の空間の 3D モデリング化については以下のように実施した．まず建物の構造物については FARO [15] の 3D スキャンで 3D モデルを取得．固定のステージセットについては図面からモデリング．可動の道具類，具体的には 3 人が持つスクリーンについては各アセットの 3D モデルを図面から事前に起こした上で，再帰性反射マーカを装着．光学式モーションキャプチャシステム VICON [16] でリアルタイムにトラッキングを行うことで，その道具が 3D 空間上にどのように位置しているかを計算し，空間の 3D データに反映した．こうすることで画像情報を一切使うことなく空間の 3D データ化を行うことができた．会場には反射マーカを付けたトラッキング可能な手持ちカメラを 3 台，固定のカメラを 3 台設置し，計 6 台のカメラ間でのモーフィングを実現した．

また，3 人のダンスシーケンスを事前に 4DViews [17] で 3D 映像として記録し，この事前記録 3D 映像を用いた CG 映像とリアルタイムのライブ映像をモデルベースビューモーフィングによってシームレスにモーフィングさせることで，Perfume 3 人のダンスの再現性の高さもあいまって，リアルとバーチャルを行き来するような演出を実現した．

またリオ 2016 大会閉会式のフラッグハンドオーバーセレモニーでは，同様の仕組みで会場であるマラカナンスタジアムの 3D スキャンを事前に行い，ステージ上のアセット類も事前にモデリングを行った．そして，モデルベースビューモーフィングと CG 合成を組み合わせて，リアルタイムの AR 演出を実現した（図 2）．

3.2　姿勢推定，セマンティックセグメンテーション

さて，上記の Seamless MR 演出において考慮されていないのが，演者の 3D モデルの扱いである．2015 年の時点では，演者のシルエットを画像からリアルタイム

に抽出することは難しかった．例えば Perfume at SXSW では，演者がスクリーンを持って移動するという演出のために，演者自身のシルエットを抽出できなくても，スクリーンの面でビューモーフィングができれば映像が成立していた．

また 2016 年に実施した Nosaj Thing とのコラボレーションツアー No Reality Tour[18] においては Kinect V2[19] 3 台で DJ（Disc Jockey）/VJ（Visual Jockey）のリアルタイム 3D スキャンを行い，リアルタイムに VJ 映像に反映させた．Kinect V2 で深度情報は最大 5 m 程度に限られるため，3D スキャン可能なボリュームは大きくない．Nosaj Thing ツアーでは DJ/VJ が余り動かないから成立したが，本手法を一般的なダンスパフォーマンスに適用するのは，機材そのものがステージ上で目立ってしまうことや，キャプチャボリュームの問題などから難しい．そんな中，近年では深層学習技術の発展によって，リアルタイムに画像情報のみから人の姿勢を検出[20]したり，物体のシルエットを抽出[21]することが実用的な精度で可能になってきた．また商用利用可能なソフトウェアも多数リリース[22]され，興業案件に姿勢推定などの技術を使用しやすい環境になっている．

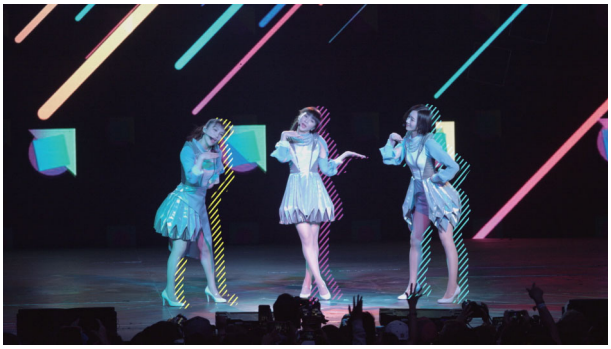

図 7　Perfume at Coachella 2019 におけるリアルタイム Scribble Effect

2019 年の Perfume at Coachella（図 7）では演者自身のリアルタイムモーフィングを高精度に実現するために，これらの技術を応用して，深層学習を用いて画像情報のみから 3 人の姿勢推定と 3D シルエットの検出を行った．

以前までの Seamless MR 同様，会場のセットは事前にモデリングし，3 人の姿勢及び形状のみリアルタイムに検出することで，モデルベースビューモーフィングのクオリティを大幅に改善しつつ，人の輪郭をなぞるような Scribble Effect といった新たなリアルタイム映像効果を加えることに成功した．

3.3　一般物体認識

同様に，深層学習技術の発展により，画像から任意の物体の位置と大きさを検出する一般物体認識と呼ばれる問題の達成可能精度が大幅に向上した[23],[24]．

本技術を応用し，フェンシングに適用したのが，フェンシングビジュアライズド[25]という剣先の動きを可視化するプロジェクトである．ライゾマティクスでは 2013 年から本プロジェクトに関わっている．2013 年当初は，剣先に再帰性反射マーカを付けて，光学式モーションキャプチャシステムで剣先の軌跡を取得．その軌跡に対して AR で映像効果を加え，より分かりやすくフェンシングの魅力を伝えることを目的とし，ポストプロダクションでの映像制作を行った[26]．

また 2016 年から，実試合導入を目標として，剣に一切の加工をせず，選手に負担を掛けずに剣先の位置をカメラ画像のみから検出するためのシステムの開発を開始した．

フェンシングの剣先は非常に早く動き，剣は曲がりやすいので形状変化が大きい．また剣先は 4K カメラで捉えて数ピクセル程度の幅しかなく，既存の画像認識アル

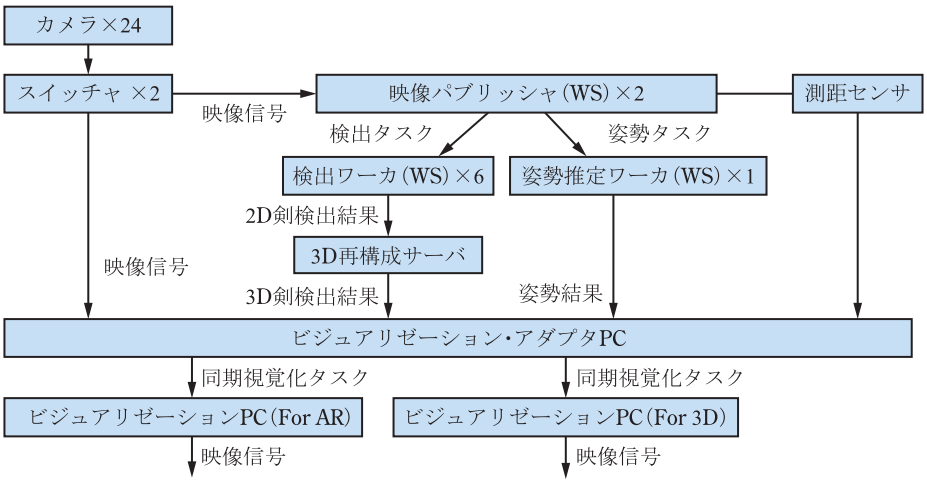

図 8　フェンシングビジュアライズドシステム

ゴリズムで見つけるには細過ぎるといった問題があった．そのため，解析用カメラのハード選定からはじめ，一般物体認識用の深層学習アルゴリズムである YOLO v3[24]を拡張し，多段構成の一般物体認識用の深層学習ネットワークを新規に開発．4K の解像度情報を失うことなく剣先位置を高精度に検出することが可能になった．また，1 台のカメラでカバーできる範囲は 8 m ほどしかないため，最終的にはピストの両サイドに 24 台の 4K カメラを設置し，試合領域全域をカバーするとともに，剣先検出のロバスト性を向上させた．また，複数台の検出結果から 3D の剣先位置を推定するシステムの開発も行った（図 8）．

本システムは 2018 年のフェンシング全日本選手権でデモを披露し，その後実試合導入に向けて精度向上施策を多数実装し，2019 年の全日本選手権で実際の試合で初めて導入を行った．また，前述の姿勢推定技術を組み合わせることでビジュアライズ表現も大幅にアップデートした（図 9）．

4. 中間的な手法

本章では，その映像効果を正しく体験できる体験者が限られる手法について概説する．ライゾマティクスの関わるプロジェクトでは，限られた体験者に向けて提供するというよりは，中継向けに特化させた状態で使用することの方が多い．

図 9　高円宮杯 2019 におけるフェンシングビジュアライズド

4.1　Dynamic Virtual Reality（Dynamic VR）

体験者の視点位置によって適切に描画する 3D オブジェクトをゆがませることで，実際には平面の表示装置にもかかわらず，移動視差によって立体的なものと知覚することができる表示装置が存在する．古くは Colin Ware らの Fish tank virtual reality（1993）[27]に代表される研究例があり，コンピュータビジョンやトラッキング技術の進歩とともに様々な派生系の技術が生まれ，Rekimoto らの研究[28]ではカメラと画像認識を用いたヘッドトラッキングによる手法が提案された．

また CG 技術や映像機材の進歩とともに，部屋全面にプロジェクションしトラッキングに Playstation Move を使用した Memo Akten らの Sony Playstation 3 Video Store（2011）[29]や，ロボットアーム制御の動的なスクリーンとモーションコントロールカメラを複雑に組み合わせた Bot & Dolly らの Box（2013）[30]といった高度な映像表現も生まれてきた．

ライゾマティクスでは，これらの先行技術を更に中継向けに進化させ，2016 年に Dynamic VR Display[11]として発表した．具体的には映像表示に LED を採用し，プロジェクションで発生する物体による影をなくし没入感を向上．また多くのシステムでは視点生成に 1 台のカメラ若しくは人物のみをトラッキングしているが，Dynamic VR システムでは複数のカメラをトラッキングし，グラフィックの描画はスイッチャの制御とも同期している．そのため，スイッチングとともにカメラそれぞれに対して適切なパースで描画を行うことが可能となり，中継時のカット切換の際にも立体感を破綻させずに済むことができる．また光学式モーションキャプチャシステムとの統合がなされており，複数の移動するスクリーンなどに対しても本技術は容易に適用可能である

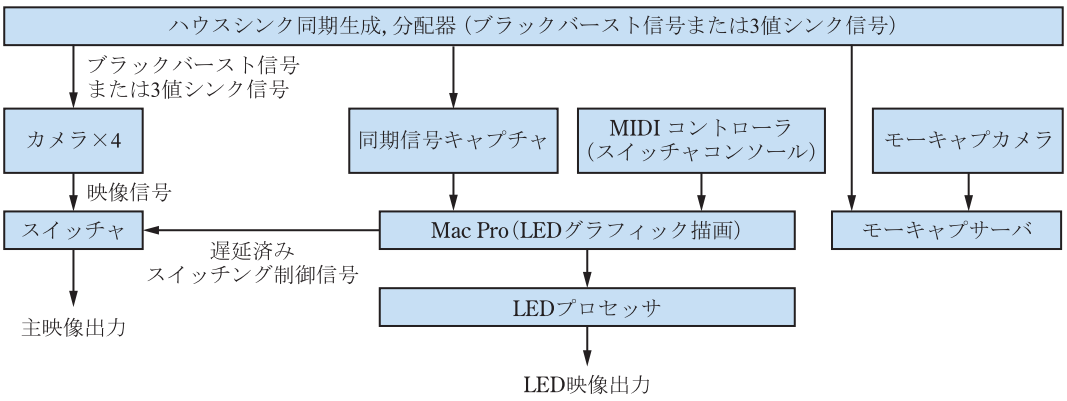

図 10　Dynamic VR Display システム

（図 10）．

本技術の一部は，リオ 2016 大会閉会式のフラッグハンドオーバーセレモニーでも用いられており，LED フレームが床面に反射しているかのような映像効果は，実はプロジェクション映像によるものである．これは使用するカメラに対して適切に映像をゆがめることと，LED フレームの発光パターンと映像を同期させることで実現している（図 1）．

5. 最　後　に

これまで紹介した AR/VR 演出はテレビの生中継やライブパフォーマンスの中でライゾマティクスが実際に使用してきたものであり厳しい条件の中で演出家や制作メンバーとともに創意工夫を凝らすことによって困難を解決し少しずつ進化してきたものであるが，ダンサーのパフォーマンス能力に依存する部分も決して小さくはない．例えばリオ 2016 大会閉会式のフラッグハンドオーバーセレモニーにおける映像効果はパフォーマーが緻密に LED フレームを動かすことができて初めて成立するパフォーマンスであり，Perfume の SXSW のパフォーマンスにおける Seamless Mixed Reality 映像効果はパフォーマーが二次元の立ち位置だけでなく，三次元の腕の位置や角度などを精密に合わせる必要があった．このパフォーマーの能力こそが他者が容易に模倣できない要因の一つになっていることは間違いなく，AR/VR 演出において身体表現のポテンシャルをいかにうまく引き出すかという点は重要である．今後オリンピック・パラリンピックで行われるセレモニーやスポーツの試合など AR/VR 演出を実装するチャンスが数多くあるだろう．技術を用いることによってアスリートやパフォーマーの身体表現の限界に迫り，魅力を伝えることに貢献できれば幸いである．

文　　献

（1） Adrien M & Claire B, https://www.am-cb.net/
（2） Klaus Obenmeier, http://www.exile.at/ko/
（3） Eyeliner, http://www.musion-holotec.com/
（4） Kimchi and Chips, “Light Barrier,” https://www.kimchiandchips.com/works/lightbarrier/
（5） WOW, “Light of Birth,” https://www.w0w.co.jp/art/light_of_birth
（6） Rhizomatiks Research, “phosphere,” https://research.rhizomatiks.com/works/phosphere-sonar-barcelona.html
（7） Intel, “True-View,” https://www.intel.com/content/www/us/en/sports/technology/true-view.html
（8） T. Beier and S. Neely, “Feature-based image metamorphosis,” SIGGRAPH ’92, pp. 35-42, 1992.
（9） M. Lhuillier and L. Quan, “Image interpolation by joint view triangulation,” Proc. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition（Cat. No PR00149）, vol. 2, pp. 139-145, 1999.
（10） C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti, “Surround-screen projection-based virtual reality : the design and implementation of the CAVE,” SIGGRAPH ’93, pp. 135-142, 1993.
（11） Rhizomatiks Research, “Dynamic VR Display,” https://research.rhizomatiks.com/works/prfm_ms_ultra_fes_2016.html
（12） G.E. Favalora, J. Napoli, D.M. Hall, R.K. Dorval, M. Giovinco, M.J. Richmond, and W.S. Chun, “100-million-voxel volumetric display,” Proc. SPIE 4712, Cockpit Displays IX : Displays for Defense Applications, Aug. 2002.
（13） S. Eitoku, T. Tanikawa, and Y. Suzuki, “Display composed of water drops for filling space with materialized virtual three-dimensional objects,” IEEE Virtual Reality Conference（VR 2006）, pp. 159-166, 2006.
（14） OptiTrack, https://optitrack.com/
（15） FARO, https://www.faro.com/ja-jp/
（16） VICON, https://www.vicon.com/
（17） 4DViews, https://www.4dviews.com/
（18） Rhizomatiks Research, “Nosaj Thing No Reality Tour,” https://research.rhizomatiks.com/works/nosaj_thing_coachella.html
（19） Kinect V2, https://developer.microsoft.com/ja-jp/windows/kinect/
（20） Z. Cao, T. Šimon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” 2017 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）, pp. 1302-1310, 2016.
（21） K. He, G. Gkioxari, P. Dollár, and R.B. Girshick, “Mask R-CNN,” 2017 IEEE International Conference on Computer Vision（ICCV）, pp. 2980-2988, 2017.
（22） WrnchAI, https://wrnch.ai/
（23） J. Redmon, S.K. Divvala, R.B. Girshick, and A. Farhadi, “You only look once : Unified, real-time object detection,” 2016 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）, pp. 779-788, 2015.
（24） J. Redmon and A. Farhadi, “YOLOv3 : An incremental improvement,” ArXiv abs/1804.02767, 2018.
（25） Fencing Visualized Rhizomatiks Website, https://research.rhizomatiks.com/s/works/fencing_tracking/
（26） Fencing Visualized, 2013, https://www.youtube.com/watch?v=FCgiOiEO_uQ
（27） C. Ware, K. Arthur, and K.S. Booth, “Fish tank virtual reality,” CHI ’93, 1993.
（28） J. Rekimoto, “A vision-based head tracker for fish tank virtual reality-VR without head gear,” Proc. Virtual Reality Annual International Symposium ’95, pp. 94-100, 1995.
（29） Memo Akten, “Playstation3 video store,” https://www.memo.tv/portfolio/sony-playstation3-video-store/
（30） Bot & Dolly, “BOX,” https://gmunk.com/BOX

（2020 年 3 月 6 日受付）

真鍋（まなべ）　大度（だいと）

東京を拠点に活動するアーティスト，インタラクションデザイナー，プログラマ，DJ．2006 ライゾマティクス設立．注意深く観察することで発見できる現象，身体，プログラミング，コンピュータそのものが持つ本質的な面白さや，アナログとディジタル，リアルとバーチャルの関係性，境界線に着目し，デザイン，アート，エンターテイメントの領域で様々なジャンルのアーティストとコラボレーションプロジェクトを行う．

花井（はない）　裕也（ゆうや）

2010 慶大大学院理工学研究科了．その後，ソニー株式会社で R & D ソフトウェアエンジニアとして AR（拡張現実）に携わり，「Live-Action AR」等の技術を開発．2014 ライゾマティクス入社．Seamless MR，Dynamic VR，インタラクティブレーザなど，カメラやプロジェクタ等を用いた数々のビジュアルシステムを開発．