小特集 4.

ディジタルヒューマンのための 3D 表現

3D Representations for Digital Human Technologies

武富貴史

Abstract

本稿では，フォトリアルなディジタルヒューマンを実現するための技術について，最近の研究事例を中心に紹介する．特に，3D での表現に限定し，人物の表現で最も重要な顔及び全身，手，髪といった体の各部位の表現及び再現について研究事例を紹介する．また，フォトリアルなディジタルヒューマンを実現するために近年では 4D キャプチャ（ボリュメトリックキャプチャ）技術の利用が盛んに行われているため，4D キャプチャによって得られたデータを利用したディジタルヒューマン技術についても紹介する．

キーワード：ディジタルヒューマン，4D キャプチャ，3D 顔，3D 頭髪

1. はじめに

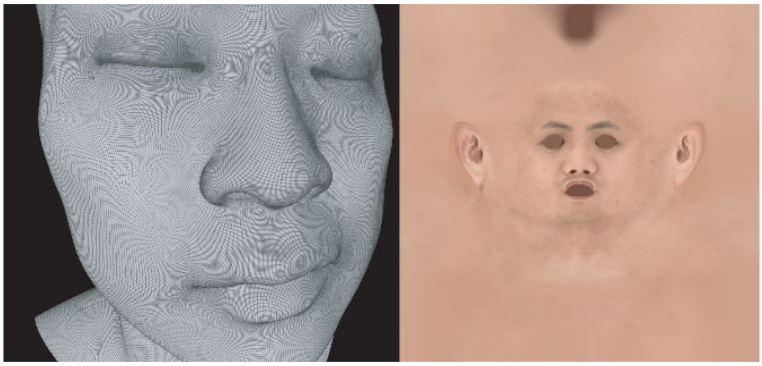
実在する人物をコンピュータ上で忠実に再現したフォトリアルなディジタルヒューマンは映像産業，バーチャルリアリティ，広告制作などの様々な分野において利用が進んでいる．ディジタルヒューマン技術は 2D 画像に基づくものとコンピュータグラフィックス技術などを用いて 3D 空間で表現されるものに大別できる．2D 画像に基づく方法では，1 枚の顔写真に写る人物の表情を入力された音声や参照となる人物の表情を転写することによってあたかも顔写真の人物が発話しているような映像を作り出す Reenactment と呼ばれる技術が盛んに研究されている[(1),(2)]．2D 画像に基づく方法は，多くの場合においてターゲットとなる人物の画像（または少量の動画像）のみが必要であり，その手軽さから映像制作の専門家のみでなく幅広いユーザを獲得している．しかしながら，現状の技術では，生成される映像中の表情と人物胴体部分の動きに違和感が生じる，違和感なく生成可能な顔の向きに制限がある，髪形や服装などの編集が難しい，というような目的に合致した高品質な映像を制作する場合の様々な問題が残されている．一方で，3D 空間で表現されるディジタルヒューマンについては，3D コンテンツの制作に多くのコストがかかるが，高い編集可能性を実現することが可能であり，また，高品質な映像生成を実現することが容易であるという利点がある．そこで，本稿では 3D 空間におけるディジタルヒューマンに焦点を当て，ここ数年の研究について概観する．特に，人物の表現において顔は最も重要な部位であり，多くの取組みがなされているため，2. では顔のみに着目して研究を概観する．その後，全身，手，髪といったその他の身体部位に対するディジタルヒューマン研究について概観する．最後に，まとめと 3D 空間でのディジタルヒューマン技術の今後の展望について述べる．

2. 3D 人物顔の表現と再現

人物の表現において顔は最も重要な部位であり，ディジタルヒューマンの領域においても多くの研究が取り組まれている[(3),(4)]．本章では，3D 空間での人物顔の表現技術として，まず，顔のパラメトリックなモデルである 3D Morphable Face Model（3DMM）について基本技術の解説をしたのち，最近の研究事例について紹介する．次に，3D 顔形状の計測について最近の研究事例を幾つか紹介する．その後，フォトリアルな人物 3D 顔を表現するための計測技術などについて最近の研究事例を幾つか紹介する．

武富貴史　(株)サイバーエージェント AI 事業本部 AI Lab
E-mail　taketomi_takafumi@cyberagent.co.jp
Takafumi TAKETOMI, Nonmember (AI Tech Studio AI Lab, CyberAgent, Inc., Tokyo, 150-6121 Japan).
電子情報通信学会誌　Vol.107 No.10 pp.943-948 2024 年 10 月

(a) スキャン結果



(b) リトポロジー結果

図 1　顔形状の計測とリトポロジー処理　(a)は左からレンダリング結果，テクスチャ画像，3D 顔メッシュ．(b)は左からリトポロジー処理済み 3D 顔メッシュ，テクスチャ画像．

2.1　3D Morphable Face Model（3DMM）

顔の形状や肌色などの見た目をパラメトリックに表現することができる 3D Morphable Face Model（3DMM）は顔認識や画像からの 3D 顔復元など様々な顔関連のタスクにおいて広く用いられている[5]．代表的な 3DMM として Basel Face Model（BFM）[6] や FLAME[7] がある．これらのモデルは形状（個人の顔形状特徴や表情）を表すパラメータと肌色などの見た目を表すパラメータから構成されている．このような 3DMM を構築するためには，一般に，異なる人物の多数の 3D 顔形状を計測する必要がある．顔形状の計測には多視点ステレオ法やレーザ計測機器が用いられるが，計測された 3D 顔モデル（図 1(a)）の 3D 顔メッシュの頂点数やテクスチャ画像のレイアウトは計測ごとに異なる．そのため，計測された 3D 顔形状の人物間での対応関係を得ることが難しく，主成分分析などの方法を適用することによって少数のパラメータを用いた 3D 顔表現（3D 顔のパラメトリック表現）を取得する場合に問題となる．そこで，多くの場合，計測された 3D 顔形状に対して基準となるテンプレートモデルを当てはめることによって，3D 顔メッシュの頂点配置やテクスチャ画像のレイアウトを統一するリトポロジー処理が実施される（図 1(b)）．具体的には，リトポロジー処理を施すことによって，例えば，3D 顔メッシュの目頭の位置の頂点 ID は全てのモデルで統一され，また，テクスチャ画像上の位置も固定される．これにより，3D 顔形状間での対応関係を得ることが可能となり，主成分分析などの方法によってパラメトリックな 3D 顔モデルの構築を行うことができる．

BFM や FLAME の機能拡張として，拡散反射のみでなく鏡面反射もパラメトリックに表現可能にした AlbedoMM[8] やしわなどの高周波な形状特徴も表現可能とした DECA[9]，内部の頭蓋骨もパラメトリックに操作可能な SCULPTOR[10] などのモデルが提案されている．更に，近年では，Neural Network を用いた 3DMM についても研究がなされており，Neural Parametric Head Models（NPHM）[11] では符号付距離場で表現される空間で個人特徴（Identity）を表現している．NPHM は BFM や FLAME などに基づく 3DMM と異なり髪形状も含めたモデルとなっている．

筆者らの研究グループでも 3DMM を拡張することによって地肌の色のみでなく化粧パターンも表現することを可能としている[12]．この研究では，まず，化粧パターンを 3D モデルで用いられる UV テクスチャの状態で抽出することのできる手法[13] を用いて，化粧付き顔画像データセットから化粧パターンの抽出を行っている．これにより，化粧パターンは統一されたレイアウトで抽出が行われることになり，主成分分析などの手法によってパラメータ化することが可能となる（図 2）．文

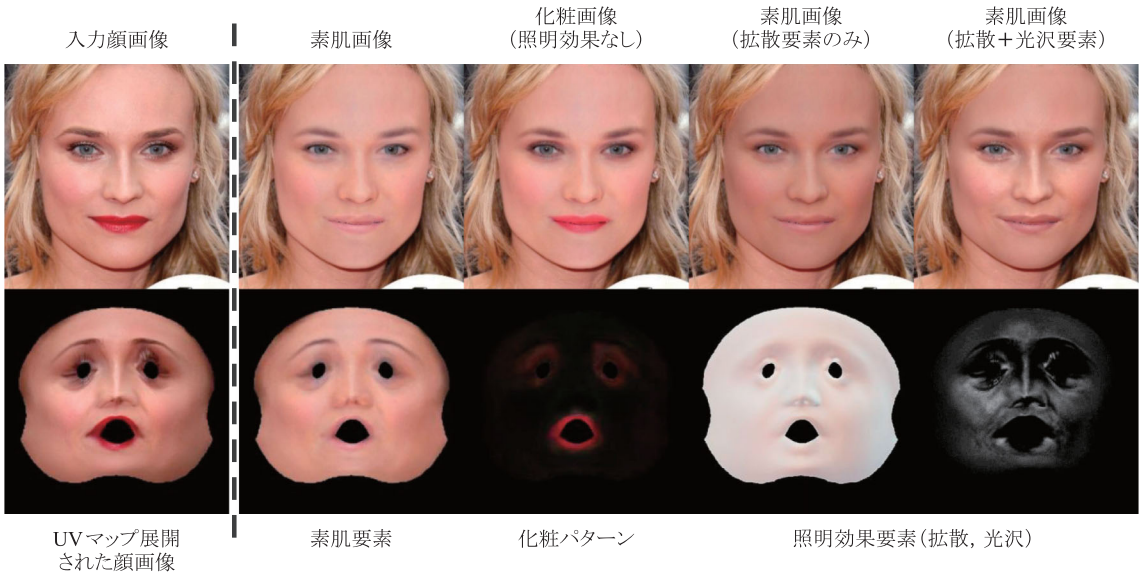


図2　顔画像からのUV化粧パターンの抽出

献(12)では主成分分析に基づくモデルとStyleGAN2に基づくモデルの2種類の化粧モデルを提案しており，それぞれ，軽量な処理が可能だが詳細なテクスチャパターンの表現が難しい，計算コストが高いが詳細なテクスチャパターンの表現が可能，といった異なる特徴を持っている．これらのモデルを用いることによって，化粧の転写や補間といったアプリケーションへの応用が可能となる．

そのほかにも様々な取組みがなされているが，2020年頃までの研究については，文献(5)を参照されたい．

2.2　3D顔形状の計測

高品質な3D顔の計測については，複数のカメラと光源を配置したライトステージなどの計測機器を用いた取組みなどが古くからなされている[14]．一方で，近年では複数カメラから得られる同期映像から動的な顔の動きを計測する取組みが複数なされている．このような撮影は，ボリュメトリックキャプチャや4Dキャプチャと呼ばれている．一般に，時系列の3D顔の計測結果から表情アニメーション可能なモデルを構築する場合には，3DMMの構築と同様に時間方向で一貫した3D顔メッシュを生成する必要がある．これは，各時刻の3D顔計測結果に対してリトポロジー処理を施すことになり，多くの人的コストが必要となる．このような問題に対して，複数視点の画像・映像から一貫した3D顔メッシュを生成する取組みがなされている．ここでは，このような取組みついて，最近の研究事例を幾つか紹介する．

Baiらは，複数視点の顔画像に対して3DMMを当てはめることによって三次元顔形状の復元を行っている[15]．多視点カメラ映像より高品質な時系列3D顔形状を復元する方法として，Liらは多視点ステレオ法を経ずにNeural Networkを用いて直接各時刻に対応した一貫した3D顔メッシュの生成を可能としている[16]．この手法では，学習データとして多視点撮影した画像と対応するリトポロジー処理済みの顔メッシュが必要であり，このような4Dキャプチャデータに対してこのようなデータを準備することは多くの人的コストが必要となる．そこで，Bolkartらは多視点撮影した画像とリトポロジー処理前の3D顔メッシュを用いて，Liらと同様に直接各時刻に対応した一貫した3D顔メッシュの生成を可能とした[17]．この方法では，学習の過程でリトポロジー処理前のメッシュに対してFLAMEモデルをフィッティングする処理を実行することで，多視点ステレオ法などによって得られる3D顔形状の計測結果をそのまま利用可能となっている．しかしながら，Liらの手法もBokartらの手法も多視点ステレオ法などを経ずに一貫した3D顔メッシュの生成が可能であるが，利用する学習データを撮影したカメラ設定に対してのみ利用可能な方法となっており，新たな多視点カメラ設定に適用するためには学習データを再度構築する必要があるという制限がある．

2.3　フォトリアルな3D顔画像の生成

ここでは，3D計測した結果からフォトリアルな顔画像を生成する研究について，最近の事例を幾つか紹介する．NeRSemble[18]では，多重解像度のハッシュグリッドの組合せでシーンを表すことによって人物の顔を表現している．これは，BFMなどで用いられているブレンドシェイプ（頂点数などが同一の幾つかの形状が異なるモデルから新たな形状を作る手法）によって動的な人の表情を表現するアイデアと同様のものとなっている．NeRSembleはNeural Radiance Fields（NeRF）に基づ

く表現を採用することによってフォトリアルな顔画像を生成することを可能としている．また，GaussianAvatar[19]は，Gaussian Splattingの考えに基づいている．まず，多視点画像に対して，NPHMをフィッティングする．その後，NPHMの各三角形メッシュ上に局所的な座標を設定し，その座標上にSpherical Harmonicsや不透明度を表現する3D Gaussianを設定する．各Gaussianのパラメータを最適化することによってフォトリアルな画像の生成を実現している．更に，DiffusionAvatars[20]では，2Dの顔画像生成はフォトリアルな画像の生成が可能だが表情などの詳細なコントロールが難しいという問題に対して，3DMM（NPHM）を組み合わせることによってフォトリアルかつ編集可能性の高い人物顔画像生成を実現している．このように，3DMMなどのパラメトリックなモデルとNeRF，Gaussian Splatting，Diffusion Modelといったフォトリアルな画像生成が可能な方法を組み合わせることによって，表情アニメーションなどの操作性を高めた顔画像生成が可能となっている．

3. 全身，手，髪の再現

ここでは，顔以外の取組みについて，全身，手，髪の再現について最近の研究事例を幾つか紹介する．

3.1 全身の再現

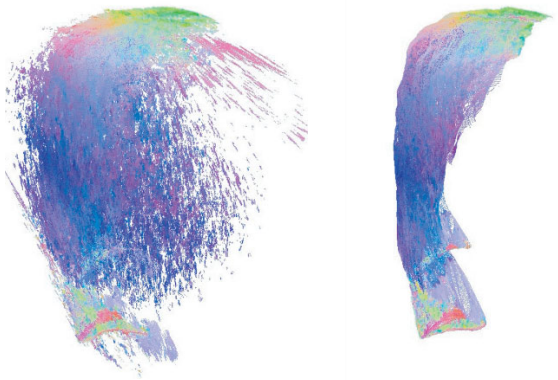
人物の全身に関しても4Dキャプチャで取得した多視点動画像を入力とした研究が行われている．HumanRF[21]では，時空間情報のNeural Radiance Fieldsの分解を低ランク制約によって分解している．また，長いシーケンスに対しては一貫性を保つことができるセグメントに動的に分割することでGPUメモリの消費を抑える工夫がなされている．また，衣服のモデリングも同時に実現するために，AniDress[22]ではリグベースの衣服の物理シミュレーションの結果と多視点映像から得られる観測が一致するように最適化することで，スカートなどのルーズな服装も表現可能としている．更に，PhysAvatar[23]では，観測画像からレンダリングに必要なパラメータを推定するInverse renderingと観測から物理モデルを推測するInverse Physicsを組み合わせることによって，衣服の物理パラメータの推定と物理ベースレンダリングに必要なマテリアルの情報を同時に推定することを可能としている．

3.2 手の再現

手も人物の再現で重要な部位の一つであるため，フォトリアルな手の再現を実現するための研究が幾つかなされている．4Dキャプチャなどによって手の計測を行う際には顔などと異なり，自己接触や自己遮蔽の問題が生じ時系列での追跡が難しいという問題がある．これに対して，Smithらは画像ベースの追跡に加えて物理シミュレーションを組み合わせることで高精度な3D手形状の追跡を実現した[24]．Iwaseらは，ライトステージを用いて計測した手のデータを用いて再照明可能な手のモデリングを実現している[25]．この研究では，教師ニューラルネットワークを用いて，視点位置や光源位置の組合せを変化させてレンダリングした画像を擬似的な正解データとして生徒ニューラルネットワークの学習に用いることによって，一般的な光源環境への対応を実現している．しかしながら，教師ニューラルネットワークの学習に用いるデータの収集も多くの労力が必要となることから，Chenらは光輸送の線形性や物体表面の反射といった物理法則を導入することで一般化とレンダリング品質のトレードオフを考慮した手法を提案している[26]．また，Huらは，物体を把持した手の画像に対して，手のモデルをフィッティングした結果と把持されている物体の3Dモデルをフィッティングした結果を用いて，新たな手の姿勢に対応するフォトリアルな画像を生成する手法を提案している．

3.3 髪の再現

髪の毛の3Dモデル制作は多くの人的コストが必要となるため，多視点撮影した画像などから髪形状の計測や髪のモデリングを行う際に制作されるガイドヘアーを自動生成する方法などが研究されている．Namらは PatchMatch Stereo法を髪の毛のような細長い形状の物体に適応させたLine-based PatchMatch Multi-view Stereo（LPMVS）を提案した[27]．LPMVSを用いることで多視点画像から髪の詳細形状を復元することが可能となる．しかしながら，LPMVS法で復元される結果は多くの雑音を含んでいる．これに対して，筆者らの研究グループでは，LPMVS法で得られる方向付きの三次元点群に対して方向制約を用いた最適化を行うことで推定された形状を修正するStrand Integration法を提案している[28]．Strand Integration法を用いることで，図3に示すように，LPMVSで推定された点群の誤差を大きく減らすことが可能となる．一方で，多視点ステレオ法などによって得られた3D形状に対して頭髪のモデリングで利用されるガイドヘアーを生成する研究が行われている．髪形データセットを用いて学習したVAE（Variational Autoencoder）や拡散モデルをPriorとして用いるNeural Strand[29]やNeural Haircut[30]はNeural Renderingと組み合わせることで高品質な髪の再現を実現している．しかしながら，生成されるガイドヘアーはデータセットに強く依存しており，また，髪の接続方向に曖昧性があるため不自然な復元結果が得られることがある．これに対して，Dr. Hair[31]では，髪の自然な流れの方向を画像から得られた方向場から設定することに



（a）LPMVS　　（b）Strand Integration

図 3　LPMVS と Strand Integration の結果

よって，不自然なガイドヘアーの生成を抑制することを可能としている．また，Dr. Hair は微分可能レンダリングに基づく最適化によってガイドヘアーを生成するため，学習データセットに依存しないという特徴を持っている．しかしながら，髪の流れの方向は滑らかであるという制約があるため，Curly な髪形など複雑な形状を取り扱うことは難しい．現状，Curly な髪形のような複雑な形状も含めて自動でガイドヘアーを生成する手法は提案されておらず，いまだに難しい問題として残っている．

4. ま　と　め

本稿では，フォトリアルなディジタルヒューマンに関して特に 3D 表現に基づく最近の研究事例の紹介を行った．ディジタルヒューマンについては，特定の人物に対して様々な会話や言語をしゃべらせることが可能となることや，実際の撮影にかかる労力やコストを抑えることができることから映像制作の現場などにおいて活用が進んでいる．しかしながら，高精細なディジタルヒューマンの作成に関してはまだ様々な技術課題が残されている．特に，衣服を含む全身の表現については，編集可能性などの点からまだ実用化が難しい状況となっている．今後，人体に関連する様々な部位や衣服などの装飾についての表現が洗練されることで，ディジタルヒューマンの活用の幅がより広がっていくことが期待される．

文　献

（1） Y. Liu, L. Lin, F. Yu, C. Zhou, and Y. Li, “MODA : Mapping-once audio-driven portrait animation with dual attentions,” Proc. ICCV, pp. 22963-22972, 2023.

（2） L. Chen, G. Cui, C. Liu, Z. Li, Z. Kou, Y. Xu, and C. Xu, “Talking-head generation with rhythmic head motion,” Proc. ECCV, 2020.

（3） S. Bi, S. Lombardi, S. Saito, T. Simon, S.-E. Wei, K. McPhail, R. Ramamoorthi, Y. Sheikh, and J. Saragih, “Deep relightable appearance models for animatable faces,” ACM Trans. Graphics, vol. 40, no. 4, 89, 2021.

（4） C. Cao, T. Simon, J.K. Kim, G. Schwartz, M. Zollhoefer, S. Saito, S. Lombardi, S. Wei, D. Belko, S. Yu, Y. Sheikh, and J. Saragih, “Authentic volumetric avatars from a phone scan,” ACM Trans. Graphics, vol. 40, no. 4, 163, 2022.

（5） B. Egger, W.A.P. Smith, A. Tewari, S. Wuhrer, M. Zollhoefer, T. Beeler, F. Bernard, T. Bolkart, A. Kortylewski, S. Romdhani, C. Theobalt, V. Blanz, and T. Vetter, “3D morphable face models-past, present, and future,” ACM Trans. Graphics, vol. 39, no. 5, 157, 2020.

（6） T. Gerig, A.M. Forster, C. Blumer, B. Egger, M. Lüthi, S. Schönborn, and T. Vetter, “Morphable face models-An open framework,” Proc. FG, pp. 75-82, 2018.

（7） T. Li, T. Bolkart, M.J. Black, H. Li, and J. Romero, “Learning a model of facial shape and expression from 4D scans,” ACM Trans. Graphics, vol. 36, no. 6, 194, 2017.

（8） W.A.P. Smith, A. Seck, H. Dee, B. Tiddeman, J. Tenenbaum, and B. Egger, “A morphable face albedo model,” Proc. CVPR, pp. 5011-5020, 2020.

（9） Y. Feng, H. Feng, M.J. Black, and T. Bolkart, “Learning an animatable detailed 3D face model from in-the-wild images,” ACM Trans. Graphics, vol. 40, no. 4, 88, 2021.

（10） Z. Qiu, Y. Li, D. He, Q. Zhang, L. Zhang, Y. Zhang, J. Wang, L. Xu, X. Wang, Y. Zhang, and J. Yu, “SCULPTOR : Skeleton-consistent face creation using a learned parametric generator,” ACM Trans. Graphics, vol. 41, no. 6, 213, 2022.

（11） S. Giebenhain, T. Kirschstein, M. Georgopoulos, M. Rünz, L. Agapito, and M. Nießner, “Learning neural parametric head models,” Proc. CVPR, 2023.

（12） X. Yang, T. Taketomi, Y. Endo, and Y. Kanamori, “Makeup prior models for 3D facial makeup estimation and applications,” Proc. CVPR, 2024.

（13） X. Yang, T. Taketomi, and Y. Kanamori, “Makeup extraction of 3D representation via illumination-aware image decomposition,” CGF, 2023.

（14） P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, “Acquiring the reflectance field of a human face,” Proc. SIGGRAPH, pp. 145-156, 2000.

（15） Z. Bai, Z. Cui, J.A. Rahim, X. Liu, and P. Tan, “Deep facial non-rigid multi-view stereo,” Proc. CVPR, pp. 5850-5860, 2020.

（16） T. Li, S. Liu, T. Bolkart, J. Liu, H. Li, and Y. Zhao, “Topologically consistent multi-view face inference using volumetric sampling,” Proc. ICCV, pp. 3824-3834, 2021.

（17） T. Bolkart, T. Li, and M.J. Black, “Instant multi-view head capture through learnable registration,” Proc. CVPR, pp. 768-779, 2023.

（18） T. Kirschstein, S. Qian, S. Giebenhain, T. Walter, and M. Nießner, “NeRSemble : Multi-view radiance field reconstruction of human heads,” ACM Trans. Graphics, vol. 42, no. 4, 161, 2023.

（19） S. Qian, T. Kirschstein, L. Schoneveld, D. Davoli, S. Giebenhain, and M. Nießner, “GaussianAvatars : Photorealistic head avatars with rigged 3D gaussians,” Proc. CVPR, pp. 20299-20309, 2024.

（20） T. Kirschstein, S. Giebenhain, and M. Nießner, “DiffusionAvatars : Deferred diffusion for high-fidelity 3D head avatars,” Proc. CVPR, 2024.

（21） M. Işık, M. Rünz, M. Georgopoulos, T. Khakhulin, J. Starck, L. Agapito, and M. Nießner, “HumanRF : High-fidelity neural radiance fields for humans in motion,” ACM Trans. Graphics, vol. 42, no. 4, 160, 2023.

（22） B. Chen, Y. Shen, Q. Shuai, X. Zhou, K. Zhou, and Y. Zheng, “AniDress : Animatable loose-dressed avatar from sparse views using garment rigging model,” arXiv preprint arXiv : 2401.15348, 2024.

（23） Y. Zheng, Q. Zhao, G. Yang, W. Yifan, D. Xiang, F. Dubost, D. Lagun, T. Beeler, F. Tombari, L. Guibas, and G. Wetzstein, “PhysAvatar : Learning the physics of dressed 3D avatars from visual observations,” arXiv preprint arXiv : 2404.04421, 2024.

（24） B. Smith, C. Wu, H. Wen, P. Peluse, Y. Sheikh, J. Hodgins, and T. Shiratori, “Constraining dense hand surface tracking with elasticity,” ACM Trans. Graphics, vol. 39, no. 6, 219, 2020.

（25） S. Iwase, S. Saito, T. Simon, S. Lombardi, T. Bagautdinov, R. Joshi, F. Prada, T. Shiratori, Y. Sheikh, and J. Saragih, “RelightableHands : Efficient neural relighting of articulated hand models,” Proc. CVPR,

pp. 16663-16673, 2023.
(26) Z. Chen, G. Moon, K. Guo, C. Cao, S. Pidhorskyi, T. Simon, R. Joshi, Y. Dong, Y. Xu, B. Pires, H. Wen, L. Evans, B. Peng, J. Buffalini, A. Trimble, K. McPhail, M. Schoeller, S.-I Yu, J. Romero, M. Zollhöfer, Y. Sheikh, Z. Liu, and S. Saito, "URHand : Universal relightable hands," Proc. CVPR, pp. 119-129, 2024.
(27) G. Nam, C. Wu, M.H. Kim, and Y. Sheikh, "Strand-accurate multi-view hair capture," Proc. CVPR, pp. 155-164, 2019.
(28) R. Maeda, K. Takayama, and T. Taketomi, "Refinement of hair geometry by strand integration," CGF, 2023.
(29) R.A. Rosu, S. Saito, Z. Wang, C. Wu, S. Behnke, and G. Nam, "Neural strands : Learning hair geometry and appearance from multi-view images," Proc. ECCV, 2022.
(30) V. Sklyarova, J. Chelishev, A. Dogaru, I. Medvedev, V. Lempitsky, and E. Zakharov, "Neural haircut : Prior-guided strand-based hair reconstruction," Proc. ICCV, pp. 19762-19773, 2023.
(31) Y. Takimoto, H. Takehara, H. Sato, Z. Zhu, and B. Zheng, "Dr. Hair : Reconstructing scalp-connected hair strands without pre-training via differentiable rendering of line segments," Proc. CVPR, 2024.

(2024年5月10日受付　2024年7月6日最終受付)

武富(たけとみ)　貴史(たかふみ)

(株)サイバーエージェント AI Labのリサーチサイエンティスト．拡張現実感，バーチャルリアリティ，コンピュータグラフィックス関連の研究に従事．2011-03奈良先端大情報科学研究科にて博士号を取得．2011～2018奈良先端大情報科学研究科助教．2018～2020華為技術日本株式会社東京研究所シニアエンジニア．2020-11から現職．