

[研究助成 (A)]

単眼ビデオによる物体操作を含む人間の 4D 再構成 およびアクション編集への応用

2251023



研究代表者
(助成金受領者)

東京大学
大学院情報理工学系研究科

教授

鄭 銀 強

共同研究者

東京大学
大学院情報理工学系研究科
知能機械情報学専攻

博士課程

ZHAN Yifan

[研究の目的]

本研究は、特別なマルチカメラシステムや専用の撮影スタジオといった高コストな設備を必要とせず、スマートフォンなどの単一カメラで撮影された「単眼ビデオ」のみから、道具や小物の操作 (Hand-held Objects) およびゆったりとした衣服 (Loose-fitting Garments) を着用した人間の複雑な 3 次元ジオメトリと動的なモーション (4D アバター) を高精度かつ高速に再構成・デジタル化する革新的な技術の確立を目的とする (図 1)。

従来の 3D 人間再構成技術 (標準的な SMPL モデルなど) は、人間の皮膚表面にのみ適合するように数理的に設計されている。そのため、人間が手に持っている物体 (スマートフォン、バッグ、スポーツ用具等) や、身体の運動と完全には同期しない衣服の「はためき・ゆとり」

を同一の統一フレームワーク内で表現することが極めて困難であった。これらの要素の運動は、人体の骨格運動から物理的にデカップリング (分離) しているため、従来のトポロジー固定型のモデルでは形状の破綻やテクスチャの深刻な歪み、位置の不整合が生じる原因となっていた。また、これらを解決するための従来のアプローチは、事例ごとに膨大な時間をかけて個別に最適化計算を行う必要があり、実用性の面で大きな課題を残していた。

本研究は、当財団の趣意である「人間と機械の調和の促進」に深く根ざしている。コンピュータが人間の外観だけでなく、着用している衣服や操作する道具の物理的・幾何学的特性までシームレスに理解・デジタル化できる基盤を構築することを目指す。これにより、バーチャルリアリティ (VR/AR)、映像制作、アスリートの動作解析、高度な作業現場におけるパフォーマンス向上など、人間と機械がサイバー・フィジカル空間で協働する多様な分野において、リアリティに富んだ自由視点合成と直感的なアクション編集の基盤を提供する。

[研究の内容, 成果]

1. ToMiE といった技術提案

本研究では、従来の SMPL モデルの表現能力の限界を打破し、時間のかかる事例ごとの最

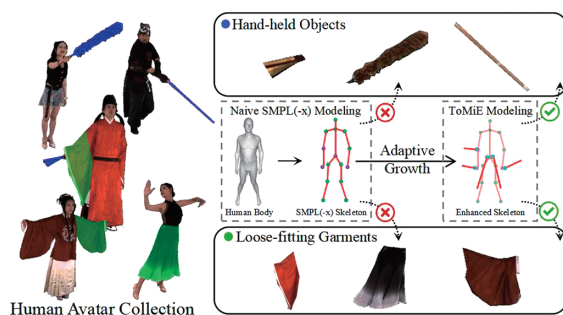


図 1 本研究が対象とする道具・小物の操作と、ゆったりとした衣服を着用した人物の例

適化を排除して、複雑な動的シナリオに適応可能な新しい4D アバター再構成アーキテクチャ ToMiE を開発した。図2には、ToMiE の処理フローを示している。ToMiE の中心的な考え方は、標準的な人体骨格木を固定したまま利用するのではなく、観測データに応じて外部関節を適応的に成長させる点にある。これにより、身体表面から外れた物体や衣服の動きを、単なる非剛体変形としてではなく、明示的な運動構造として扱うことができる。

ToMiE の処理パイプラインは、標準 SMPL 骨格を基盤としながら、手持ち物体やゆったりとした衣服など、SMPL のみでは表現が困難な外部要素を適応的にモデル化する構成となっている。全体の流れは以下の5段階にまとめられる。

1.1. Canonical 空間における Gaussian の初期化

まず、標準 SMPL メッシュの頂点を用いて canonical space に 3D Gaussian を初期化する。

これにより、人体本体の幾何構造を SMPL に基づいて安定的に表現し、その後の姿勢変形やレンダリングの基準となる canonical representation を構築する。初期段階では、Gaussian は基本的に人体表面に対応して配置され、従来の SMPL ベースの人体アバター表現と同様に、人体の姿勢変化に追従できるように設計されている。

1.2. LBS による Gaussian の変形と Adaptive LBS

次に、Linear Blend Skinning (LBS) を用いて、canonical space 上の Gaussian の位置を各フレームの観測空間へ変換する。Gaussian の

位置は SMPL 骨格の姿勢に従って変形されるが、単純な LBS だけでは回転やスケールのずれを十分に補正できないため、ネットワークを用いて Gaussian の回転およびスケールを補正する。学習初期の warmup phase では、Adaptive LBS は標準 SMPL 骨格のみを利用する。その後、adaptive growth によって外部関節が新たに生成されると、LBS の対象骨格は標準 SMPL 骨格に加えて、成長した外部関節も含む拡張骨格へと更新される。これにより、手持ち物体やゆったりとした衣服のように、人体本体とは異なる運動をする領域も、より柔軟に表現できるようになる。

1.3. Gaussian Rasterization と勾配の逆伝播

変形された Gaussian は Gaussian rasterization によって画像としてレンダリングされ、入力画像との誤差に基づいて学習が行われる。このとき、レンダリング誤差から得られる勾配は各 Gaussian に逆伝播される。ToMiE では、この勾配情報を単なるパラメータ更新だけでなく、「どの関節周辺で表現能力が不足しているか」を判断するための重要な手がかりとして利用する。つまり、ある関節に関連する Gaussian に大きな勾配が集中している場合、その領域では標準 SMPL 骨格だけでは十分に観測を説明できておらず、新たな外部関節を成長させる必要がある可能性が高いと考える。

1.4. 勾配に基づく親関節の局在化と Motion Kernel

外部関節をどこに追加すべきかを決定するため、ToMiE は gradient-based parent joints localization を行う。具体的には、各 Gaussian の勾配を関節ごとに集計し、大きな勾配が蓄積された関節を、新しい外部関節を成長させる候補、すなわち parent joint として選択する。ただし、SMPL の LBS weight のみで Gaussian と関節の対応関係を決めると、手持ち物体や衣服のように人体表面から離れた領域では誤った割り当てが生じやすい。そこで ToMiE では、LBS weight に加えて Motion Kernel を導入する。Motion Kernel は観測空間における運動の

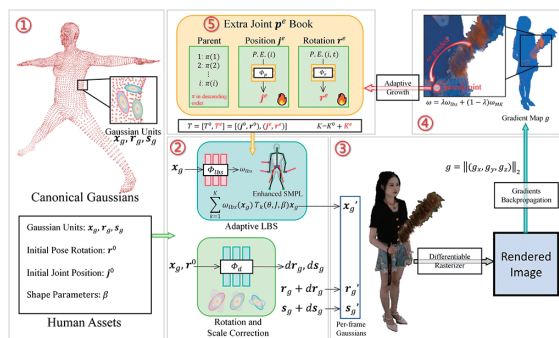


図2 提案の ToMiE の処理フロー

類似性に基づいて Gaussian と関節の対応を補正する仕組みであり、canonical space 上の誤分類に過度に依存せず、より頑健に勾配を適切な関節へ割り当てることを可能にする。

1.5. Extra Joint Book と MLP による外部関節の最適化

親関節が決定されると、その子関節として外部関節を生成し、これらを extra joint book として管理する。extra joint book には、親関節の情報、canonical space における外部関節位置、および親関節座標系における外部関節の回転が保存される。さらに、浅い MLP によって外部関節の位置および回転を最適化し、フレームごとの SE(3) 変換を明示的に推定する。これにより、手持ち物体やゆったりとした衣服を人体本体の SMPL 姿勢から切り離して表現できる。すなわち、身体の姿勢を固定したまま、手持ち物体や衣服のみを独立に動かすといった、部分ごとの明示的なアニメーション編集が可能となる。

このように、ToMiE は標準 SMPL に基づく人体表現を出発点としつつ、Gaussian のレンダリング誤差から得られる勾配情報を用いて、表現能力が不足している部位を自動的に検出する。そして、Motion Kernel によって関節と Gaussian の対応付けを高精度化し、必要な箇所外部関節を適応的に成長させる。最終的には、extra joint book によって外部関節を明示的に管理・最適化することで、手持ち物体や loose-fitting clothing を人体本体から分離して再構成・アニメーション編集できる点が、本研究の大きな特徴である。

2. 実験セットアップおよび評価

提案手法の有効性を実証するため、日常的な手持ちオブジェクト（バッグ、ガジェット、スポーツ道具など）を操作するシーンや、ルーズな衣服（コート、ゆったりとしたスカート、ドレスなど）をまとった人物を撮影した単眼ビデオデータセットを用いて、新規視点合成（Novel

View Synthesis）の品質および未知のポーズに対するアニメーション制御（Animation）の精度を検証した。比較対象（Baseline）には、当時の世界最高水準（State-of-the-Art）である GauHuman、およびレンダリング特化型の 4D 表現手法である Im4D を選定した。

新規視点合成の品質評価として、ピーク信号対雑音比（PSNR）、構造的類似性（SSIM）、および知覚的類似性（LPIPS）を測定した。具体的な値は、表 1 に示している。

表 1 の示すように、提案手法（ToMiE）は PSNR と SSIM の指標において、衣服や手持ち物体のトポロジー変化を表現できなかった従来のベースライン手法を圧倒した。特に、激しい運動時に物体の周囲に発生がちであったゴーストノイズ（残像）や、衣服の境界部分におけるカラーの歪みが劇的に低減され、極めて高い再現精度を達成した。

図 3 の示すように、ルーズな衣服の細かなシワや、手持ちの小物の細部が、視点を 360 度

表 1 定量実験の結果

Dataset Method Metric	Animatable		DNA-Rendering			PSNR(masked) ↑
	B	G	PSNR(full) ↑	SSIM ↑	LPIPS ↓	
3DGS-Avatar	✓	✗	28.93	0.953	0.0459	17.69
GART	✓	✓	29.25	0.958	0.0480	17.12
Im4D	✗	✗	26.28	0.965	0.0308	14.90
GauHuman	✓	✗	30.22	0.962	0.0405	18.26
ToMiE(ours)	✓	✓	31.28	0.966	0.0374	19.75

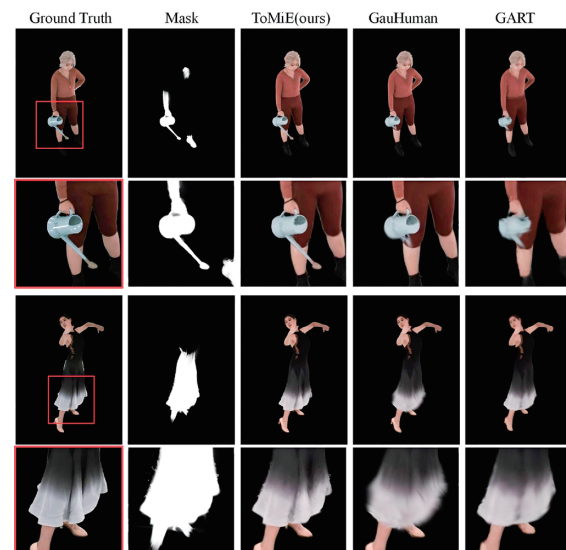


図 3 自由視点レンダリング画像の比較

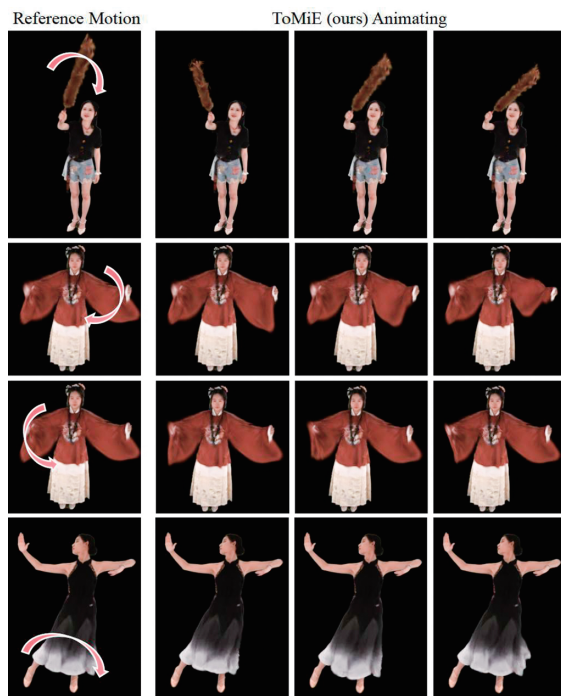


図4 3D アクション編集・モーション操作の実施例

シームレスに変更した場合でもブレることなく、極めて高い視覚的忠実度を持ってレンダリングできることを確認できた。

図4では、標準的な人体骨格のポーズを固定したまま、ToMiEによって自動生成された「拡

張関節 (Extra Joints)」のみをアニメーション制御することで、「人物の身体を静止させた状態で、手に持っているオブジェクトのみを独立して動かす」、あるいは「衣服のみを風になびかせるように独立してはためかせる」という、従来の SMPL ベースの手法では不可能であった高度な 3D アクション編集・モーション操作を実証した。

以上により、当初の研究目的であった「単眼ビデオによる物体操作を含む人間の 4D 再構成」と「アクション編集への応用」に向けた中核技術を確立した。手作業のリギングを必要とせず、標準 SMPL に含まれない外部要素を適応的にモデル化できる点は、日常環境での 4D 人間理解および応用展開に有用である。

[成果の発表、論文など]

Yifan Zhan, Qingtian Zhu, Muyao Niu, Mingze Ma, Jiancheng Zhao, Zhihang Zhong, Xiao Sun, Yu Qiao, Yinqiang Zheng: ToMiE: Towards Explicit Exoskeleton for the Reconstruction of Complicated 3D Human Avatars, IEEE/CVF International Conference on Computer Vision (ICCV), pp.14259-14269, 2025.