

個別衣服の学習によるリアルタイム仮想試着

Zaiqiang Wu[†] Jingyuan Liu[†] Yechen Li[†] Toby Chong[†] I-Chao Shen[†] 五十嵐 健夫[‡]
金 孟佳[‡], 久保 涉[‡], 柴田 勇希[‡], 堀 隆之[‡], 田之上 隼人[‡], 西原 大輝[‡]

概要：個別衣服の学習による仮想試着法は、個別の対象衣服に対してきめ細かなディテールを持つ高品質な結果を生成できる対象衣服専用のネットワークを学習する。しかし、既存の個別衣服の学習による仮想試着法では、物理的な測定衣服の使用が必要であり、適用性が制限されている。本発表では、様々なポーズ下での高品質で時間的に一貫性のある衣服画像の合成するために、物理的な測定衣服が不要な仮想測定衣服を活用する、個別衣服の学習による仮想試着法を紹介する。

1. はじめに

仮想試着は、ユーザが実際に衣服を着用することなく、衣服の外観をプレビューできるため、近年人気が高まっている。しかし、一般的な仮想試着のための既存の画像ベースの手法は、2つの主な困難のために、異なるポーズの下で現実的で一貫性のある衣服画像を合成する際に問題を抱えている。1) これらの手法のトレーニングに使用されるデータセットには、膨大な衣服のコレクションが含まれているが、各衣服の細かいディテールが欠けている。2) これらの手法は、静止ポーズにおけるターゲット衣服の正面画像を変形することによって結果を合成するため、他の視点やポーズでは品質やディテールが低くなる。これらの欠点を克服するために、個別衣装の学習による試着法は、特定の対象衣装に対してきめ細かなディテールを持つ高品質な結果を生成できる対象衣装専用のネットワークを学習する[1]。しかし、個別衣服の学習による仮想試着法では、物理的な測定衣服の使用が必要であり、適用性が制限されている。

本発表では、様々なポーズ下での高品質で時間的に一貫性のある衣服画像を合成するために、物理的な測定衣服が不要な仮想測定衣服を活用する、新しい個別衣服の学習による仮想試着法を紹介する [2]。提案手法では、さらに、合成された衣服と身体パーツの間のギャップを効果的に埋めるギャップフィリングモジュールを導入している。

2. 関連研究

3D モデルベースのバーチャル試着：3D モデルベースの手法[3]は、3D 測定データ（例えば、3D の人間のポーズや形状）をキャプチャし、測定データに適合するドレープのある衣服を生成する。いくつかの初期の手法[4] は、物理ベースのシミュレーションを適用して、身体と接触する衣服

のアニメーションを生成している。このようなアプローチは、詳細な皺を含む視覚的に妥当な結果を生成することができるが、計算コストが高いという問題があり、広く適用することができない。シミュレーションの計算コストを削減するために、最近の研究 [5]では、衣服のアニメーションと変形を生成するためにデータ駆動型の手法を利用している。しかし、これらの方法は、分布外の入力に対して正確な結果を生成することが保証されていない。さらに、3D モデルベースのアプローチでは、高いリアリズムと忠実度を示す画像を生成するために、衣服の物理的および材料的特性をキャプチャする必要がある、これは大きな課題となり、多くの労力を必要とする。

画像ベースのバーチャル試着：画像ベースのアプローチ[6]は、3D 測定データを必要としないため、3D モデルベースの手法よりも広く適用できる。CAGAN [7]の先駆的な研究は、学習ベースの方法を用いて、人間の画像上でファッションアイテムを交換するタスクに最初に取り組んだ。VITON[8]は、出力の質を高めるために粗いものから細かいものへの合成戦略を導入し、ターゲット衣服に対応する身体領域と位置合わせするために薄板スプライン (TPS) 変換を適用した。CP-VTON [9]は、ターゲット衣服のディテールを保持するために、幾何学的マッチングモジュールを導入し、TPS 変換パラメータを明示的に学習する。VTNFP[10]と ACGPN[11]は、セマンティックレイアウトを予測するために、ボディセグメンテーションマップ予測モジュールを利用している。VITON-HD[12]は、高解像度の仮想試着画像を合成するために、アライメントを考慮したセグメンテーション正規化とジェネレーターを提案し、ずれた領域を処理し、ターゲット衣服の詳細を保持する。HR-VITON[13]は、衣服のワーピングとセグメンテーションの生成ステージを統合することで、VITONHD で発生するピクセルのスクワイーズアーチファクトをさらに排除している。しかし、既存

[†] 東京大学

[‡] ソフトバンク (株)

の画像ベースの仮想試着手法では、個別の衣服に特化したデータセットではなく、一般的なデータセットを用いて衣服合成ネットワークを学習するため、きめ細かなディテールを持つ対象衣服を合成できないことが多い。さらに、これらの方法は、トレーニング中に見ることができない対象衣服に対して満足いく結果をもたらさない傾向がある。

ビデオ仮想試着：ビデオ仮想試着の目的は、画像ベースの仮想試着手法に基づき、対象となる衣服を着用した人物のリアルで時間的に一貫性のあるビデオを合成することである。FW-GAN [14]や FashionMirror [15]などのこれまでの手法は、隣接するフレーム間のフリッカーを滑らかにする後処理としてオプティカルフローを利用している。しかし、フリッカーアーチファクトは平滑化後も存在する。時間的一貫性をさらに改善するために、ClothFormer [16]は、改善された時空間的一貫性を持つターゲット衣服シーケンスを合成するために、密なフローマッピングを予測する2段階の変形モジュールを使用することを提案した。しかし、これらの方法は、訓練のためのビデオ仮想試着の大規模なデータセットに依存しており、その収集は困難である。現在までのところ、公にアクセス可能なビデオ仮想試着のデータセットはVVT [17]のみである。ビデオデータセットの不足は、ビデオ仮想試着技術の発展を妨げている。

3. 提案手法

本論文では、図1下に示すように、物理的な測定用衣服の必要性を排除する仮想測定用衣服を活用した、新しい個別衣服の学習による仮想試着フレームワークを紹介する。仮想計測衣服は、推定された人間の3Dポーズに応じて変形可能なスキンメッシュであり、ターゲット衣服画像を合成するためのガイドとして機能する。まず、2D画像から人間の3Dポーズを推定する。次に、推定された人間のポーズに従って3D仮想計測衣服を変形させ、2D画像にレンダリングする。レンダリングされた画像は、衣服ごとのネットワークに入力され、入力画像の人物に適合したポーズを持つターゲット衣服画像を合成する。最後に、合成された衣服画像を入力画像と合成し、ギャップ充填ネットワークを用いてギャップを埋めることで、最終的な結果を生成する。

本手法は、各フレームを独立に処理し、時間的平滑化を行わない画像ベースのアプローチであるにもかかわらず、動画に適用した場合に優れた時間的一貫性を達成する。一般的な仮想試着手法と比較すると、我々の個別衣装の学習による仮想試着手法は、より複雑なデータセットの取得とネットワークの学習を必要とする。しかし、我々のターゲットユーザーは、製品のプロモーションを目的とし、そのような処理を行うリソースを持っている衣服小売業者である

と仮定している。さらに、我々の衣服に特化した手法は、専門的な3Dモデリングを必要とする3Dモデルベースの仮想試着技術よりも労力がかからず、一般的な画像ベースの手法よりも優れた結果を得ることができる。

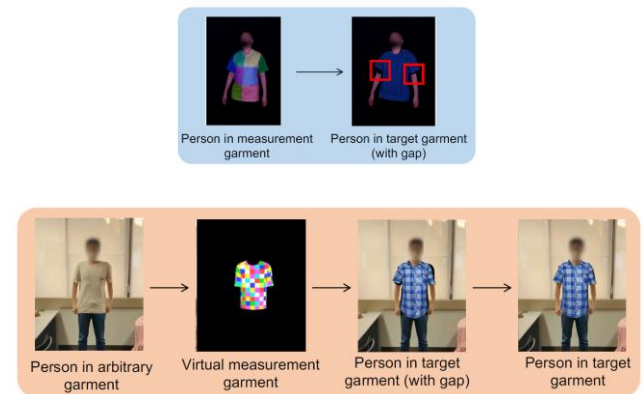


図1：上) 測定衣服を利用した仮想試着手法（既存手法）。下) 提案している仮想測定衣服による仮想試着手法。

参考文献

- [1] T. Chong, I.-C. Shen, N. Umetani, and T. Igarashi. Per garment capture and synthesis for real-time virtual try-on. In The 34th Annual ACM Symposium on User Interface Software and Technology, pp. 457–469, 2021.
- [2] Zaiqiang Wu*, Jingyuan Liu*, Long Hin Toby Chong, I-Chao Shen, Takeo Igarashi (*: joint first authors). Virtual Measurement Garment for Per-Garment Virtual Try-On. Graphics Interface 2024.
- [3] X. Pan, J. Mai, X. Jiang, D. Tang, J. Li, T. Shao, K. Zhou, X. Jin, and D. Manocha. Predicting loose-fitting garment deformations using bone-driven motion networks. In ACM SIGGRAPH 2022 Conference Proceedings, pp. 1–10, 2022.
- [4] A. Selle, J. Su, G. Irving, and R. Fedkiw. Robust high-resolution cloth using parallelism, history-based collisions, and accurate friction. IEEE transactions on visualization and computer graphics, 15(2):339–350, 2008.
- [5] D. Xiang, F. Prada, T. Bagautdinov, W. Xu, Y. Dong, H. Wen, J. Hodgins, and C. Wu. Modeling clothing as a separate layer for an animatable human avatar. ACM Transactions on Graphics (TOG), 40(6):1–15, 2021.
- [6] S. Lee, G. Gu, S. Park, S. Choi, and J. Choo. High-resolution virtual tryon with misalignment and occlusion-handled conditions. In European Conference on Computer Vision, pp. 204–219. Springer, 2022.
- [7] N. Jetchev and U. Bergmann. The conditional analogy gan: Swapping fashion articles on people images. In Proceedings of the IEEE international conference on computer vision workshops, pp. 2287–2292, 2017.
- [8] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis. Viton: An image-based virtual try-on network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7543–7552, 2018.
- [9] B. Wang, H. Zheng, X. Liang, Y. Chen, L. Lin, and M. Yang. Toward characteristic-preserving image-based virtual try-on network. In Proceedings of the European conference on computer vision (ECCV), pp. 589–604, 2018.
- [10] H. Yang, R. Zhang, X. Guo, W. Liu, W. Zuo, and P. Luo. Towards photo-realistic virtual try-on by adaptively generating-preserving image content. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7850–7859, 2020.

- [11] R. Yu, X. Wang, and X. Xie. Vtnfp: An image-based virtual try-on network with body and clothing feature preservation. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 10511–10520, 2019.
- [12] S. Choi, S. Park, M. Lee, and J. Choo. Viton-hd: High-resolution virtual try-on via misalignment-aware normalization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 14131–14140, 2021.
- [13] S. Lee, G. Gu, S. Park, S. Choi, and J. Choo. High-resolution virtual tryon with misalignment and occlusion-handled conditions. In European Conference on Computer Vision, pp. 204–219. Springer, 2022.
- [14] H. Dong, X. Liang, X. Shen, B. Wu, B.-C. Chen, and J. Yin. Fw-gan: Flow-navigated warping gan for video virtual try-on. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 1161–1170, 2019.
- [15] C.-Y. Chen, L. Lo, P.-J. Huang, H.-H. Shuai, and W.-H. Cheng. Fashionmirror: Co-attention feature-remapping virtual try-on with sequential template poses. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 13809–13818, 2021.
- [16] J. Jiang, T. Wang, H. Yan, and J. Liu. Clothformer: Taming video virtual try-on in all module. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10799–10808, 2022.
- [17] H. Dong, X. Liang, X. Shen, B. Wu, B.-C. Chen, and J. Yin. Fw-gan: Flow-navigated warping gan for video virtual try-on. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 1161–1170, 2019.