

# 姿勢推定による人間拡張のためのリアルタイム人影生成

吉田匠吾<sup>†</sup> 謝浩然<sup>†</sup> 宮田一乗<sup>†</sup>

**概要:** 影は、インタラクションやメディアアートなどで用いられているように人間およびオブジェクトの存在を暗黙的に表す。従来のアプローチには影を動的に生成する研究はあるが、空間拡張現実に用いるための、ユーザの全身の影の生成は行われていない。本研究では、プロジェクタ・カメラシステムを用いてユーザのインタラクティブな影を作成する「インタラクティブな影生成システム」を提案する。本システムでは、まずユーザの体の事前モデリングと仮想環境の構築というオフラインプロセスを経て、ユーザをスキャンして作成した 3D モデルに物体を動かす仕組みを付与する。その後、システム内で設定した仮想光源を 3D モデルに照射することで影を生成する。最後に、プロジェクタを用いて実環境であるユーザの床面の足元に投影する。また、本研究では提案システムの使い勝手と生成された影の印象を調査するために被験者による評価実験を行った。調査の結果、ユーザの動きの認識の限界や生成された影の動作の遅延などが見られたが、多くのユーザから影の印象に対して高評価を得た。

## 1. はじめに

影は直観的な操作が可能であり、インタラクションやメディアアートにおいて、わかりやすく直観的な応用が可能である。例えば、Temari and Shadow [1] や Silhouette Interactions [2] では、ユーザの手や体全体の影が、投影された仮想物体とのインタラクションや家電製品を操作するためのジェスチャーとして利用されている。また、エンタテインメントのみならず人間拡張としての使用も可能であり、体が不自由な者や、日常生活の支援にも活用できる [3]。加えて、影はアンビエント情報としての使用もでき、寂しさ解消や癒しなどのセラピーの分野に活用することもできる。しかし、影の形状はユーザ自身の身体の形状に依存するため、ユーザは自身の影の伸縮や複製ができない。つまり、従来のシステムでは、インタラクションの自由度が制限されており、ユーザは限られた動作を行うことしかできない。

本研究では、実環境においてインタラクション可能なユーザの影の動的生成を行う影生成システムを提案する。提案システムでは、まずユーザをスキャンし、3D 人体モデルを作成する。そして、作成した 3D モデルを姿勢推定によりユーザの動きと同期させる (図 1a)。仮想環境内で生成された影は、プロジェクタ・カメラシステムを用いてユーザの足元に投影する。本システムを用いることにより、実世界にできる影と同じような影をユーザに提示することができる。また提案システムはユーザの日常生活に応用でき、例えば手の届かない位置を、影を用いて指し示したり、他者との遠隔コミュニケーションとして用いたりできる。さらに、複数の生成された影を用いてユーザの影を複製し、あらかじめ用意されたモーションデータを適用することで、ユーザの動作に依存しないアクションを行うことも可能である。

## 2. 関連研究

本章では、実際の影と人工的な影を用いたインタラクションの関連研究を紹介する。

実際の影について、Shadow Reaching [3] は、ユーザの動作を影に適用するインタラクション技術を提案している。この技術は、光源とユーザの位置に応じてユーザの影のサイズが変化する現象を利用し、インタラクションのために自分の手の影が目標点に到達することを実現したものである。また Shadow Communication [4] は、ユーザの影を利用した遠隔コミュニケーションを提案している。このシステムでは、遠隔地にいるユーザを撮影し、そのユーザの影を半透明の画面に表示する。これにより、参加者は対面しているのと同じようにコミュニケーションすることができる。

人工的な影という点では、Ishii ら [5] は影の色を利用したアバターシステムを提案している。このシステムでは、話し手の気持ちを色で表現することで、ユーザとのインタラクションの活性化を目指したものである。また Morita ら [6] は、投影型バーチャルハンドシステムを提案している。このシステムでは、自身の手の動きを反映させた仮想の手を投影し、対象物を指し示す。加えて、ユーザの頭の動きで投影領域を操作するため、手の届かない物体にも手が届くようになる。この研究は、Shadow Reaching [3] と似ているが、車いす利用者を対象とした支援に主軸を置いている点で異なる。

他には、実際の影と人工的な影の両方を用いている研究もある [7, 8]。しかし、これらの影システムでは、投影される影がユーザの体型に依存するため、影の形状を変更して使用することは困難である。

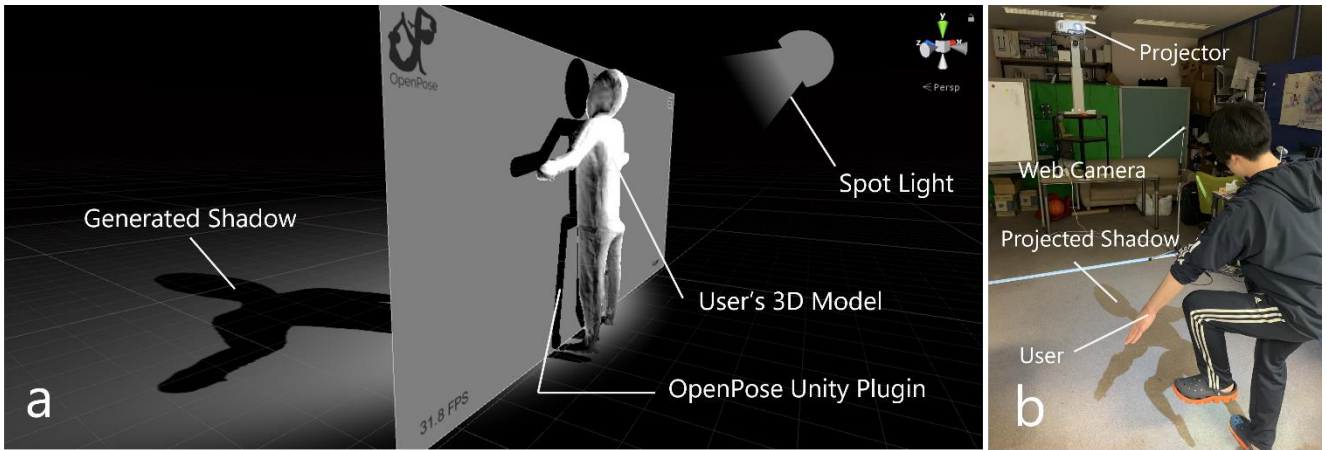


図 1 姿勢推定によるリアルタイム人影生成とユーザ体験の様子。(a) スキャンした 3D モデルに仮想光源を照射することで、影を生成する。Web カメラからユーザのスケルトン情報を取得し、アプリケーション内に設置されたキャンバスに表示させる。スケルトン情報から得たキーポイントを 3D モデルに当てはめることで、ユーザの動きと影の動作を同期させることができる。(b) 生成した影は、プロジェクタを用いてユーザの足元に投影する。

一方、提案システムでは、ユーザの影を生成して、ユーザの動きに動的に対応させる。提案システムを用いることで、ユーザは影の形状とサイズの変更、複製を体験でき、新しい影絵の手法として、エンタテインメントや人間拡張の分野で活用できる。

### 3. 提案システム

提案システムのフレームワークを図 2 に示す。本システムの実装には、ゲームエンジンである Unity と姿勢推定手法を用いている。まず、市販の 3D スキャナを用いてユーザをスキャンし、ユーザの 3D モデルを作成する。スキャンの際、ノイズやスキャン精度の限界により不要なメッシュやモデルに穴が開くことがあるため、メッシュ修正ソフトを用いて処理を行う。さらに、作成した 3D モデルに関節を付与するために、物体を動かす仕組みである「リグ」を作成した 3D モデルに当てる。次に、単眼カメラで撮影した画像から、深層学習を用いた OpenPose [9]により、ユーザのスケルトン情報をリアルタイムに取得する。取得したスケルトン情報を作成した 3D モデルに当てはめ、仮想光源を照射することで、影を生成する。

また、実験のための環境構築として、プロジェクタと単眼カメラを用いたプロジェクタ・カメラシステムを屋内に設置した。まず、部屋の大きさに合わせて投影サイズを目視で調整した。次に、自然な影を表現するために、ゲームエンジンの仮想光源の方向を調整し、実空間の光の方向を模倣した。最後に、ユーザの 3D モデルに基づいて生成された影を、実空間に投影した。ゲームエンジン内に 3D モデルを配置し光源を設定することで、任意の位置や方向から人工的な影を生成できるため、投影された影をユーザの動きと同期させることができる。

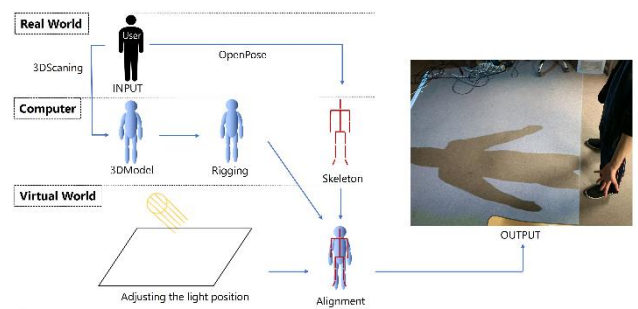


図 2 システムのフレームワーク

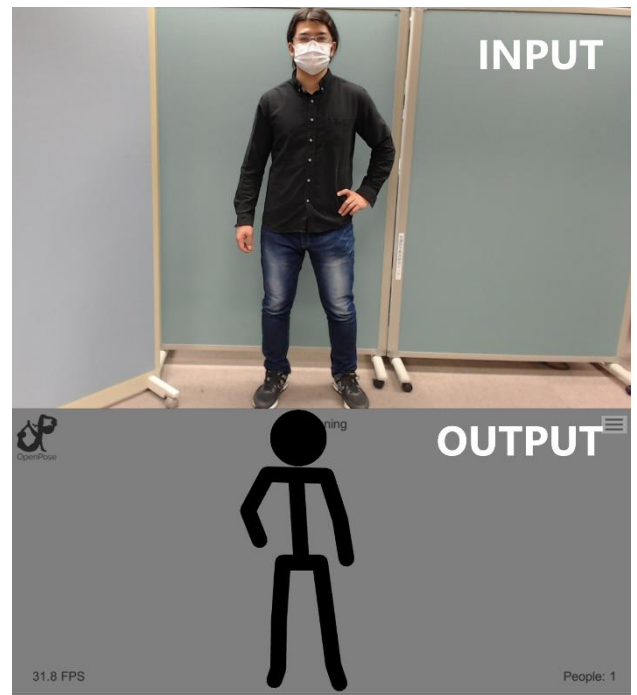


図 3 OpenPose で取得したユーザのスケルトン情報

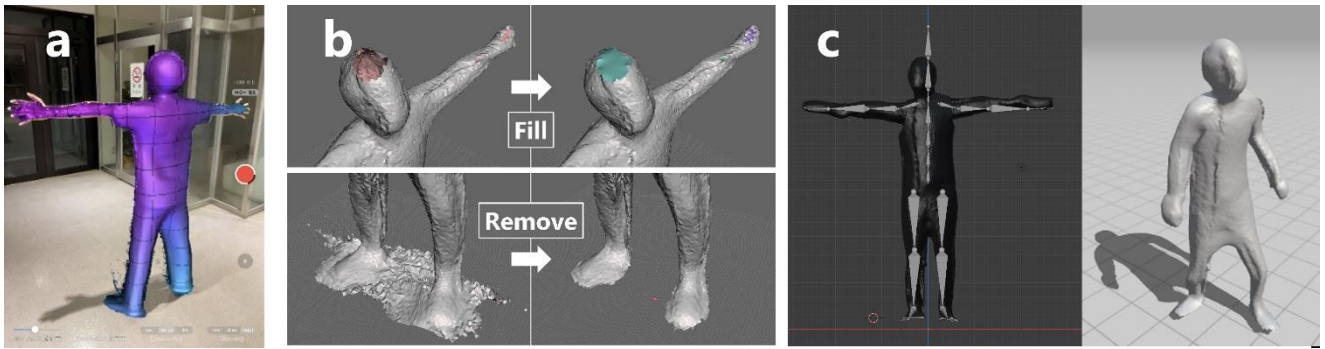


図 4 (a) 3D モデルの生成と修正, LIDAR スキャナを用いてユーザのスキャンを行う. (b) スキャンした 3D モデルには不要なメッシュや穴が開いているため, メッシュ修正ソフトで修正を行う. (c) 修正した 3D モデルには関節が付与されていないので, リグ付けを行う.

### 3.1 システム構成

プロトタイプシステムの構築には, デスクトップ PC (intel i9-10900KF CPU 3.70GHz, RAM 64GB, GeForce RTX 2080 Ti), iPad Pro (Early 2021, 3D スキャンに使用.), プロジェクタ (BENQ TH671ST, 投影映像は高い位置から床に投影), Web カメラ (Logicool C922n) を使用した. また, プロジェクタの上部に鏡を設置し, 投影映像を反射させて床に映像を投影している. ユーザのスケルトン情報の取得には OpenPose と Unity を用いている. 姿勢推定の結果から生成されるボーンモデル (図 3) の各関節には, あらかじめキーポイントが設定されている. このキーポイントを取得し, 対応する 3D モデルの関節に当てはめることで, 3D モデルとユーザの動きを同期させることができる. また, 影の生成には Unity のスポットライト機能を利用している. 3D モデルに向けスポットライトを照射することで影が生成され, その影を仮想カメラで撮影し, ユーザの足元に投影することで, ユーザは生成された影を視認できる.

### 3.2 3D スキャンとリギング

3D スキャンには, iPad Pro に取り付けられた LIDAR スキャナおよび Laan labs 社の 3D スキャナアプリケーションを使用した (図 4a). スキャンの解像度は 5 ミリで, 図 4 の事例では, 身長 165cm の男性をスキャン対象とした. また, 精度上の問題から, スキャンした 3D モデルに小さな穴や不要な断片が発生することがあるため, Autodesk MeshMixer<sup>1)</sup> を使用して, 3D メッシュを修復した (図 4b). スキャンした 3D モデルへのリギングは, Adobe mixamo<sup>2)</sup> を使用した. 3D モデルのリギングにかかる時間は約 2 分を要した (図 4c). なお, 使用した機器では詳細な 3D モデルは作成できず, ユーザごとに 3D モデルを作成しても, その区別は判断し難い. そのため, 本研究ではユーザごとのスキャンは行わず, 上記の 3D モデルによる実験を行った.

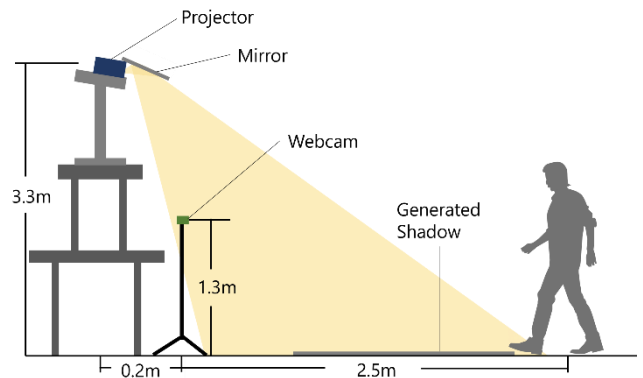


図 5 提案システムの全体像

## 4. 実験と評価

提案したシステムの使いやすさを確認するために, 実際に生成された影とユーザ自身の動きの同期を体験させる実験を行った. 各ユーザは, 3.2 節で生成した男性の 3D モデルにより生成された影に対し, 自身の動きとの同期を視認する. ユーザの中には, 体を大きく動かす者や, 手を振る者, ダンスをする者など, 様々な動作をするユーザがいた.

### 4.1 提案システムを用いた実験

本実験として, 7 人の被験者に対して提案システムの評価実験を行った. なお, 被験者のうち 6 名が男性で, 1 名が女性であり, 年齢層は 20~30 代である. 評価実験は, 提案システムを用いた影とユーザの動きの体験と, アンケート調査からなる. 被験者は Web カメラから約 2.5 メートル離れた場所に立ち, 自身の体を自由に動かすことができる (図 5). 体験中は参加者の様子を観察し, 約 2 分間の体験後, 提案システムに対するフィードバックを求めた (図 6).

実験後は, 本システムのユーザビリティや自身の動きに追従する影の再現性, 遅延具合を確かめるために, 5 段階のリッカート尺度を用いたアンケートによるユーザ調査を行った (1: 最も悪い~5: 最も良い).

1) <https://www.meshmixer.com/>

2) <https://www.mixamo.com/#/>



図 6 提案システムを用いたユーザ体験。手を振る者や、ポーズをとる者など、様々な動きを見せた。

1. 自身の動きと影の動きは同期していましたか。
2. 生成された影は自身の影のように感じましたか。
3. 生成された影は遅延なく動作しましたか。
4. 本システムの満足度を教えてください。

#### 4.2 結果と考察

実験の結果、被験者が急に動いてもその動作に合わせて投影された影が素早く追従するなど、モーション・トラッキングが正常に機能していることが確かめられた。しかし、カメラのフレームレートが 30fps 程度であったため、早い動きに対して生成された影の遅延が発生した。また、生成された影が細かく振動するという意図しない動きが見られた。この問題は、高性能な GPU (グラフィック・プロセッシング・ユニット) の使用や、ボーンモデルの各関節に追従する 3D モデルの動きをなめらかにすることで、解決できると考える。またすべてのユーザが、生成された影が自分の影のように感じたことと評価していることから、提案システムが空間拡張現実のためのインタラクティブな影の生成に有用であることが確かめられた。その一方で、先述したような影の動きの遅延に関する報告もあった。また、姿勢推定アルゴリズムでは、参加者がスカートをはいていると正しく認識できなかった。

アンケート調査の結果を図 7 に示す。項目の内容は、生成された影がユーザと同じ動きをしていたか(同期性)、生成された影は自身の影のように感じたか(親和性)、生成された影には遅延があったか(遅延)、システムの使い勝手(満足度)に関する評価である。評価項目 1~4 は、図 7 の評価項目に対応している。評価の結果、項目 1 の「同期性」が高い評価を得ていることが確かめられた。また、項目 2 の

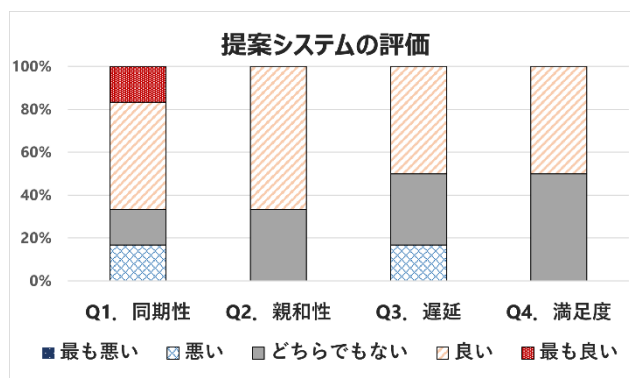


図 7 提案システムの評価結果

「再現性」と項目 4 の「満足感」については、低い評価をした被験者はいなかったが、最高評価をする被験者もいなかった。しかし、両項目において、スキャン精度の限界によりユーザの 3D モデルの再現度が低いことや、ある程度の遅延があったにもかかわらず、半数以上のユーザから高い評価を得た。これは、「同期性」の評価の高さとの関連が示唆され、本システムがユーザを満足させるだけの影を生成できることを意味している。「同期性」と項目 3 の「遅延」では、それぞれ 1 名の被験者から低評価を得た。これは、本物の影と同じ反応速度を再現するのは難しく、ユーザが不可解な姿勢をとったときにシステムの認識が遅れてしまうことが要因だと考えられる。

#### 5. 議論と今後の課題

本論文では、姿勢推定を用いたインタラクティブな影生成システムを提案した。ユーザのスケルトン情報から各関

節の位置に付与されているキーポイントを取得し、Unityにてユーザの3Dモデルに貼り付けることで、リアルタイムに影を生成した。ユーザ体験では、生成した影の同期性や再現性、遅延の程度、システムの満足度を確かめた。その結果、特に同期性の項目で高評価を得られた。実験時においても、被験者は体を自由に動かしており、生成された影がそれに追従していた。しかし、被験者から影の動作の遅延により生成された影が自分のものであると十分に感じる事ができないという意見を得た。これが、本システムに対し最高評価が得られなかった要因であると考えられる。また、OpenPoseのシステムの都合上、参加者の服装には制限がある。つまり、参加者は自分の体の形がわかるような服を着る必要がある。

提案するシステムの応用例として、ユーザが作成した影絵が自律的に動き出すという新しいエンタテインメントを提供できると考えられる。また、孤独感を和らげるセラピーの分野や、Shadow Reaching [3]と同様、ウェアラブルプロジェクトを用いた人間拡張[10,11]の分野にも応用が可能であると考えられる。

今後の課題としては、以下の課題が挙げられる：

**1) 遅延の改善** 現在のシステムでは、影の動作に遅延が生じており、これが影の同期性やシステムの満足度に影響を与えている。この課題は、複数のカメラを用いたモーションキャプチャの高速化手法[12]を応用することで解決できると考えられる。

**2) ユーザの移動範囲の拡大** OpenPoseは、単眼のRGB画像から2次元の姿勢推定を行うシステムであるため、現在のシステムでは奥行き情報を取得することはできない。つまり、ユーザがWebカメラに対して深度方向に移動しても、生成された影のサイズや位置に変化はない。そのため、ユーザが前後に動くと、生成される影とユーザとの位置がずれてしまう。

**3) 3Dモデルの生成方法の改善** 現在のスキャンシステムでは、3Dスキャンには高価な機器を必要とし、かつスキャン解像度に限界がある。これらの問題は、深度センサおよび高性能3Dスキャナを使用することや、Neural Body [13]の手法を応用することで解決できると考えられる。とりわけ、Neural Bodyを用いた手法では、ユーザの指や頭、服の形など、細かい部分をスキャンすることも可能になる。細かい部分のスキャンが可能になることで、より詳細な影が投影できる。例えば、各ユーザに自身の形状の影を提示することや(4章を参照)、指先や服の形状を表す影を利用したインタラクションが可能になる。

**4) 3Dスキャンとリギングの処理時間の改善** 現在のシステムでは、3Dスキャン後にモデルの保存や別アプリケーションでのメッシュ処理を手動で行い、さらに別のアプリケーションにてリギングを行う。つまり、これらの工程を3つのアプリケーションで行っている。この問題も、デブ

スセンサを使ったリアルタイムのメッシュ生成やリギングを行うことで、解決できる。

## 参考文献

- [1] Obushi, N., & Koshino, M. (2018, April). Temari and Shadow: an interactive installation combining virtual handicraft and real shadow. In Proceedings of the Virtual Reality International Conference-Laval Virtual (pp. 1-4).
- [2] Chita, E., Sugiura, Y., Hashimoto, S., Kunze, K., Inami, M., & Ogata, M. (2015, September). Silhouette interactions: using the hand shadow as interaction modality. In Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (pp. 69-72).
- [3] Shoemaker, G., Tang, A., & Booth, K. S. (2007, October). Shadow reaching: a new perspective on interaction for large displays. In Proceedings of the 20th annual ACM symposium on User interface software and technology (pp. 53-56).
- [4] Miwa, Y., & Ishibiki, C. (2004, November). Shadow communication: system for embodied interaction with remote partners. In Proceedings of the 2004 ACM conference on Computer supported cooperative work (pp. 467-476).
- [5] Ishii, Y., Watanabe, T., & Sejima, Y. (2016, October). Development of an Embodied Avatar System using Avatar-Shadow's Color Expressions with an Interaction-activated Communication Model. In Proceedings of the Fourth International Conference on Human Agent Interaction (pp. 337-340).
- [6] Morita, K., Hiraki, T., Matsukura, H., Iwai, D., & Sato, K. (2020, September). Extension of Projection Area using Head Orientation in Projected Virtual Hand Interface for Wheelchair Users. In 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE) (pp. 421-426). IEEE.
- [7] Iwasaki, H., Kondo, M., Ito, R., Sugiura, S., Oba, Y., & Mizuno, S. (2016). Interaction with virtual shadow through real shadow using two projectors. In ACM SIGGRAPH 2016 Posters (pp. 1-2).
- [8] Xu, H., Kanaya, I., Hiura, S., & Sato, K. (2006). User interface by Real and Artificial shadow. In ACM SIGGRAPH 2006 Research posters (pp. 83-es).
- [9] Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. IEEE transactions on pattern analysis and machine intelligence, 43(1), 172-186.
- [10] Yoshida, S., Xie, H., & Miyata, K. (2021, March). A wearable augmented reality system with anamorphosis projection. In International Workshop on Advanced Imaging Technology (WAIT) 2021 (Vol. 11766, p. 117662K). International Society for Optics and Photonics.
- [11] Adachi, Y., Xie, H., Torii, T., Zhang, H., & Sagisaka, R. (2020, March). EgoSpace: Augmenting Egocentric Space by Wearable Projector. In Proceedings of the Augmented Humans International Conference (pp. 1-2).
- [12] Dong, J., Jiang, W., Huang, Q., Bao, H., & Zhou, X. (2019). Fast and robust multi-person 3d pose estimation from multiple views. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7792-7801).
- [13] Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., & Zhou, X. (2021). Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9054-9063).