画像生成 AI による即興プロジェクションマッピングの提案

中島次郎[1,a)]　酒井修二[1,b)]

概要：本研究では，専門的な知識がなくても簡単にセットアップでき，立体的な投影対象に対して位置合わせされた映像をその場で生成する新しいプロジェクションマッピング手法を提案する．提案手法では，体験者が投影対象を自由に動かしたり変形させたり，または，プロジェクタを動かしたりしながら，プロンプトを用いて映像（静止画，動画）を生成・投影する．これにより，これまでにない即興のインタラクティブプロジェクションマッピング体験を提供できる．

1. はじめに

プロジェクションマッピングは，物体や空間に映像を投影し，その形状や特徴に合わせて映像を変形させることで，リアルな立体感や動きを表現する技術である．この技術により，建物の外壁，舞台のセット，さらには不規則な形状の物体に至るまで，様々な表面をスクリーンとして使用することが可能になる．また，映像を投影するだけでなく，鑑賞者が介入して映像が変化するインタラクティブプロジェクションマッピングや，動く物体へ映像を投影するダイナミックプロジェクションマッピングが実用化されている．例えば，塗り絵をした魚の絵をスキャンし，壁に投影されているバーチャル水槽で泳ぐインタラクティブプロジェクションマッピング [1] や，変化する表情や顔の立体的な凹凸に合わせてメイクを投影するダイナミックプロジェクションマッピング [2] が映像表現技術として活用されている．

近年，Latent Diffusion Model[3] をはじめとした画像生成 AI 技術の進展により，即座に高品質な映像を生成することが可能となった．プロジェクションマッピングにおいても，画像生成 AI の導入により映像が即座に生成されることで，空間や鑑賞者の状態に応じた，より魅力的な視覚体験を提供できる可能性がある．また，映像コンテンツの準備が不要になるため，プロジェクションマッピングのプロセスが大幅に簡略化される．

しかしながら，従来の生成 AI を活用したプロジェクションマッピング技術にはいくつかの課題が残っている．プロセスが簡略化された一方，依然として環境のセットアップには専門的な知識や特殊なハードウェア，複雑な手順が必要である．例えば，一般的なプロジェクションマッピングにおいて，立体的な投影対象物に対して映像を投影するためには，簡易的な方法として，映像を投影しながら目視で確認しつつ，手動で映像を位置合わせする方法がある．しかし，この方法では，投影対象の形状が変化したり投影環境が変化したりする場合に，再度目視での位置合わせを行う必要があるため，体験者によるインタラクティブプロジェクションマッピングには向いていない．一方，自動的に位置合わせを行う方法としては，カメラを用いて投影対象物を撮影する方法がある．プロジェクタでグレーコードなどのパターン画像を投影した様子をカメラで撮影し画像処理を行うことで，投影対象の位置姿勢や形状を推定して位置合わせを行うことが可能となる．このとき，投影対象とプロジェクタの位置関係を正確に推定するためには，プロジェクタ・カメラ間のキャリブレーションを事前に行う必要がある．プロジェクタ・カメラ間のキャリブレーション手法としては，チェッカーボードなどを複数回撮影する手法が一般的である．この手法により精度よく位置合わせを行うことは可能であるが，専門的知識が必要でセットアップも煩雑である．また，投影対象の形状計測や追跡のために RGBD カメラなどを用いる手法もあるが，特殊なカメラが必要になる上に，キャリブレーションの必要性は変わらない．このような手順があるため，画像生成 AI により即時的な映像生成が可能になったにもかかわらず，プロジェクションマッピングの実現には依然として高いコストと労力が伴う．

そこで本研究では，専門的な知識がなくても簡単にセットアップでき，立体的な投影対象に対して位置合わせされた映像をその場で生成する即興プロジェクションマッピングシステムを提案する．これにより，プロジェクタ・カメラ間のキャリブレーションが不要で，セットアップを簡素化し，可搬性を向上させたプロジェクションマッピングを

[1] TOPPAN デジタル株式会社
[a)] jiro.nakajima@toppan.co.jp
[b)] shuji.sakai@toppan.co.jp

実現する．

2. 関連研究

ControlNet[4] の登場により，プロジェクションマッピングにおいて画像生成 AI が活用されるようになった．ControlNet は，画像生成モデルに空間的な制御を追加するためのニューラルネットワークであり，エッジ，デプス，セグメンテーション，人体のポーズ，線画など，さまざまな条件画像で画像生成を制御することが可能である．ControlNet の条件画像として投影対象の深度などを用いることで，投影対象の形状に合わせ，プロンプトなどで指定した画像を生成することができる．これにより，従来は投影対象の形状に合わせて人手によって行われていた映像制作作業を省略し投影画像を生成することが可能となった．以下，画像生成 AI を活用したプロジェクションマッピングに関連する研究や事例をいくつか紹介する．

Urban Symphony[5] や Moment Factory[6] による映像作品においては，プロジェクションマッピングのコンセプトデザインを支援するために ControlNet による画像生成を活用している．これにより，アーティストは全体的なクリエイティブの決定に集中でき，デザインプロセスを効率化し，時間と労力を削減している．これらの技術や作品では，画像生成によるプロジェクションマッピングのコンセプトデザインに着目しているため，投影対象の形状計測や映像の位置合わせは一般的なプロジェクションマッピングと同様のプロセスを行う必要がある．

「Unreal Pareidolia -shadows-」[7] は，壁面に投影される日用品や玩具の影に対し，img2img と BLIP-2 を使用しリアルタイムで画像とキャプションを生成する映像作品である．体験者は物体を動かすことで影絵を作り，その後ボタンを押すことで影絵に合わせた画像をインタラクティブに生成することが可能である．一般的なインタラクティブプロジェクションマッピングの自由度は，事前に CG により制作された範囲内に限定されるが，生成 AI により任意の入力に映像を生成可能となり，自由度が飛躍的に向上したことを示す一例の作品である．一方，投影する映像は平面への投影となっており，立体物へのインタラクティブな投影は行っていない．

Casper DPM[8] は動的な手の動きに対して高い精度と低遅延で映像を投影することができる．この技術により，手の形状に応じて生成 AI が生成したコンテンツを投影するアートデモが提案されている．ここで，Caper DPM は手の動きに特化したダイナミックプロジェクションマッピングを目的としているため，LeapMotion や高速プロジェクタといった特殊で大掛かりな機材が必要となっている．

これらの研究や事例では，画像生成によるプロジェクションマッピングを実現しているが，投影を行うまでのセットアップは一般的なプロジェクションマッピングと同様のプロセスが必要となっている．例えば，チェッカーボードと複数のパターン画像を用いて事前にプロジェクタとカメラのキャリブレーションが必要であり，手間と時間を要する作業となっている．また，特殊な機器を必要とする場合，コストが高くなり，容易に実現することは難しい．

ここで，特殊な機材やプロジェクタとカメラのキャリブレーションが不要となり，かつ，画像生成 AI により即座に映像生成が行えるようになれば，投影対象や投影映像の自由度が向上し，様々な場面で応用可能なインタラクティブプロジェクションマッピングが実現可能となる．

3. 提案手法

本研究では，プロジェクションマッピングにおいて環境のセットアップから映像の生成, 投影までを一貫して即座に行えることを目標とする．提案手法では，DenseMatching による画像位置合わせと ControlNet による画像生成を組み合わせることで，3 次元としての投影対象の位置関係を考慮せずに，2 次元画像上での位置合わせで処理が完結するフローを採用する．これにより，投影対象やプロジェクタ・カメラの位置関係が変化しても即座に投影対象に合わせたプロジェクションマッピングを可能とした．

提案手法で必要な機材としては，プロジェクタとカメラ及び画像生成用のマシン（GPU サーバ）である．プロジェクタは映像が鮮明に投影可能であり，カメラは投影画像が視認可能な画像が撮影できればどのようなものでもよい．ここで，提案手法は，プロジェクタ・カメラ間のキャリブレーションを行わないため，ピンホールカメラモデルから外れるようなプロジェクタやカメラを用いてもよい．ただし，カメラとプロジェクタは光軸が一致することが理想的であるため，なるべく近接して設置する．

提案手法のフロー図を図 1 に示す．以下，各処理について詳しく説明する．

3.1 入力

図 1 の Input 部に対応する入力処理を説明する．投影空間における投影対象を撮影するために，Blank Image として黒で塗りつぶされた画像 B を投影し，カメラにより撮影を行い投影空間画像 I_B を取得する．ここで，投影空間が暗い場合は，明るさ調整などのために黒色から白色の階調表現により適宜明るくする．また，画像位置合わせのため，Registration Image としてコントラストが高めで特徴点が検出しやすい画像 R を投影し，カメラにより撮影を行い位置合わせ用画像 I_R を取得する．加えて，画像生成用のプロンプト p を入力し，プロジェクションマッピングにより投影したい映像のイメージをテキストなどで指定する．以上の I_B, I_R, p が本提案手法における基本的な入力となる．

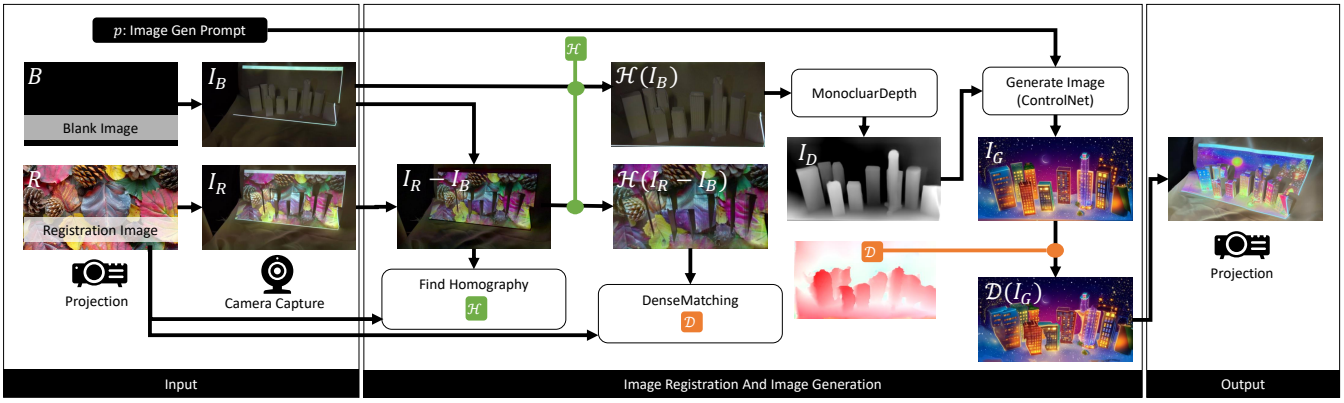

図 1 提案手法のフロー図

3.2 画像位置合わせと画像生成

図 1 の Image Registration And Image Generation 部に対応する画像処理を説明する．初めに，後段の処理の精度向上のために，事前処理として投影対象の色情報を除去した差分画像 $I_R - I_B$ を計算する．後段の処理で十分な精度が出る場合は，高速化のため，この処理は省略してもよい．

次に，カメラとプロジェクタ間の光軸を大まかに合わせる処理として，ホモグラフィ変換 $\mathcal{H} = \text{FindHomography}(R, I_R - I_B)$ を計算する．そして，$\mathcal{H}$ により投影空間画像と差分画像それぞれホモグラフィ変換を行い画像 $\mathcal{H}(I_B), \mathcal{H}(I_R - I_B)$ を計算する．投影対象が立体形状である場合は，この処理だけでは位置合わせが不十分であるため，後段の処理でより詳細な位置合わせを行う．

より詳細な位置合わせ処理として，DenseMatching による位置合わせを行う．提案手法では，RoMa[9] による画像位置合わせを採用する．RoMa は視差や照明変化に頑健に密なマッチングを行うことができる．DenseMatching による変換を $\mathcal{D} = \text{DenseMatching}(R, \mathcal{H}(I_R - I_B))$ として計算する．

並行する処理として，ControlNet による画像生成を行うため，$I_D = \text{MonocularDepth}(\mathcal{H}(I_B))$ として MonocularDepth によりデプス推定を行い I_D を条件画像とする．MonocularDepth は単眼の RGB 画像を入力としてデプス画像を推定することが可能である．提案手法では，ZoeDepth[10] によるデプス推定を採用する．I_D は ControlNet への入力でのみ使用し，大まかな形状が推定できれば良いため，MonocularDepth は低精度で高速な他の手法でもよい．また，ControlNet への入力はデプス画像以外にもエッジ画像や線画画像を用いてもよい．

画像生成としては，SDXL[11] などの画像生成モデルと ControlNet を使い $I_G = \text{GenerateImage}(I_D, p)$ としてデプス画像とプロンプトを入力し画像生成を行う．提案手法では ControlNet が利用可能であれば，別の画像生成モデルや LoRA[12] などの追加学習モデル，IPAdapter[13] による画像参照技術などをアプリケーションに応じて使用することができる．あるいは ControlNet ではなく $\mathcal{H}(I_B)$ を入力として img2img による画像変換で投影対象に合わせた画像を生成してもよい．また，画像だけでなく，Animatediff[14] により形状に合わせた動画像を生成することも可能である．

以上の処理により，生成画像 I_G と，DenseMatching による位置合わせ変換 $\mathcal{D}$ が得られるので，最後に位置合わせ済み生成画像 $\mathcal{D}(I_G)$ を計算する．

3.3 出力

位置合わせ済み生成画像 $\mathcal{D}(I_G)$ をプロジェクタで投影することで，投影対象の形状に即した画像を投影することができる．投影対象の形状や位置姿勢，プロジェクタ・カメラの位置姿勢が変化した場合やプロンプトを変更する場合は，改めて Input からの処理を行う．

4. 実験・考察

本手法の有効性を検証する．検証内容としては，位置合わせ精度の定量的評価・目視による投影品質の確認・投影までの時間の計測を行った．ここで，位置合わせ精度の評価としては，DenseMatching 前の R と $\mathcal{H}(I_R - I_B)$，DenseMatching 後の R と $\mathcal{D}(\mathcal{H}(I_R - I_B))$ の MSE（値が小さいほど位置合わせの誤差が小さく，精度が良い．MSE 画像は誤差が小さいほど暗くなるように表示）をそれぞれ計算した．投影までの時間は，プロンプト入力後に，画像 B, R の投影から生成画像を投影完了するまでの時間を計測した．

実験環境は，プロジェクタとして Anker Nebula Apollo，カメラとして Sony DSC-RX0M2，カメラ入力・プロンプト入力・映像出力用のクライアント PC と画像処理・画像生成用の RTX A6000 による GPU サーバでシステムを構成した．画像生成モデルとしては SDXL ベースのモデルに Hyper-SD[15] を組み合わせサンプリングステップを 8step とし，1280×720 の解像度で画像生成を行った．提案手法を以下の 4 つの主要なインタラクションパターンに分類し，それぞれのパターンごとに一連の実験を行った．

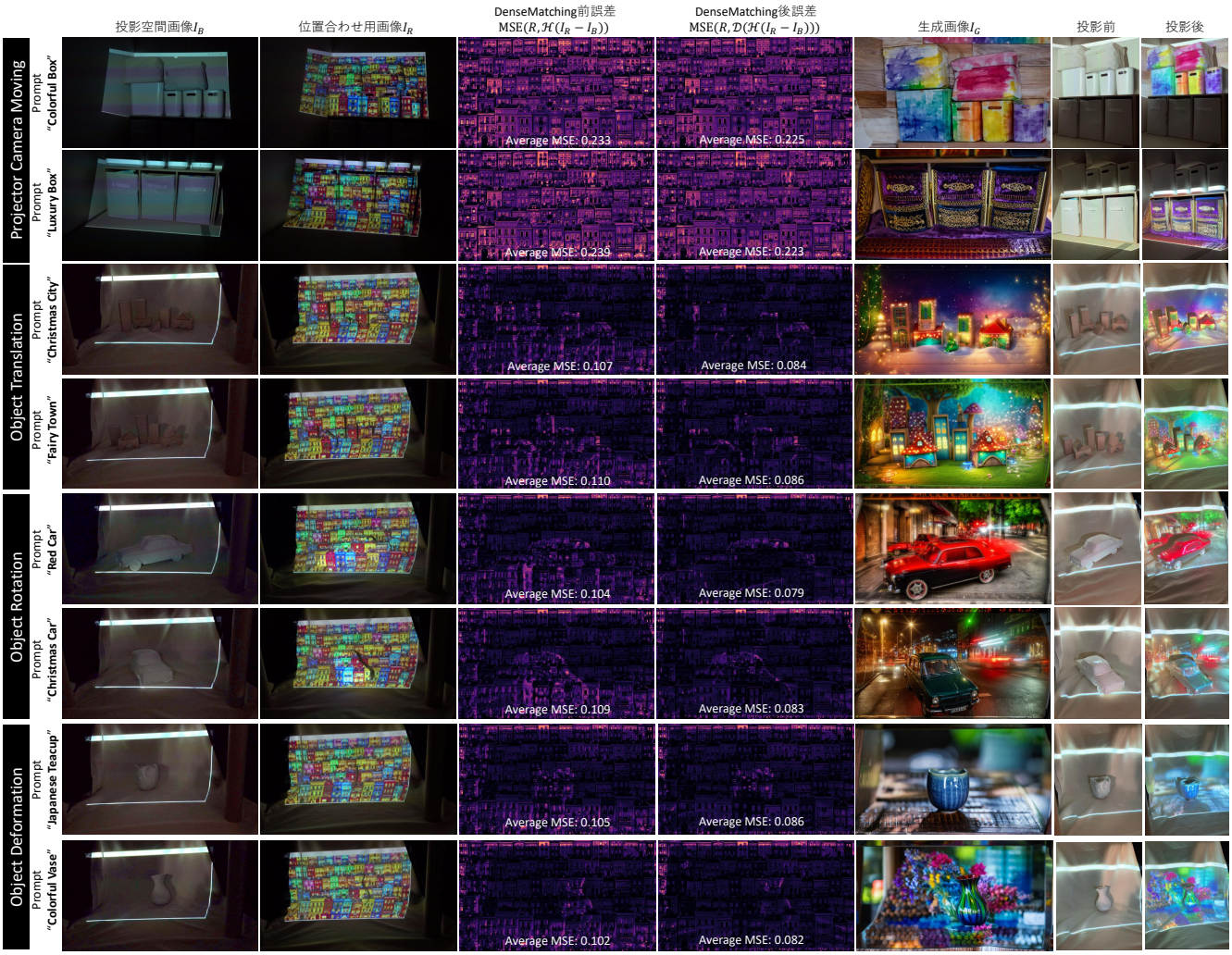

図 2 実験結果

- Projector Camera Moving: プロジェクタ・カメラの移動．実験では，空間に並べた箱に対して，プロジェクタとカメラを動かし，投影する位置の変更を行った．
- Object Translation: 投影対象物の移動．実験では，積み木を組み替えることを行った．
- Object Rotation: 投影対象物の回転．実験では，3D プリントした車形状の物体を回転させて異なる角度から投影を行った．
- Object Deformation: 投影対象物の変形．実験では，粘土を変形させた．

本実験結果のまとめを図 2 に示す．

はじめに，画像位置合わせ精度の評価を行った．いずれのパターンにおいても，DenseMatching を介すことで，画像位置合わせ精度が向上できていることを確認できる．提案手法では RoMa による DenseMatching を行ったが，入力画像サイズと画素の移動範囲の関係から位置合わせ精度は変化するため，投影対象空間に応じた調整によりさらなる精度向上は可能であると考える．Projector Camera Moving パターンにおいては，一般的なプロジェクションマッピングの場合は改めてカメラ・プロジェクタ間のキャリブレーションを行う必要があるが，本実験により位置関係が変わっても画像位置合わせ精度はほとんど変化ないことを確認できた．これは従来のプロジェクションマッピング技術に対して可搬性がある点で優位性があると言える．例えば，ロボットなどに載せてプロジェクタとカメラを動かす場合，微小な揺れで再キャリブレーションの必要があるが，本手法により移動可能なプロジェクションマッピングを提供できる．

次に，目視による投影品質の確認を行った（図 2 の投影後画像参照）．投影対象の境界部で位置ずれが気になるところはあるが，基本的には形状に合わせて投影できていることを確認した．また，MonocularDepth により形状が取得できているため，布状の箱のしわや，粘土の細かい凹凸に合わせて画像生成が行えており，詳細な立体的な構造に対しても対応できることを確認した．

最後に，投影までの時間の評価を行った．いずれの結果も 16 秒から 17 秒程度で画像を投影できることを確認した．インタラクティブな体験としては，1 秒以下の速度で

実施できることが理想だが，現状は待機時間が生じている．画像生成のステップ数の調整や，各種技術の高速化，ローカル PC でカメラ入力から画像生成・投影を完結させるなど，高速化は十分可能であるが，生成品質と生成速度はトレードオフの関係であるため，1 秒以下にするのは難しいと考えられる．アプリケーションを考案する際には，現状は数秒待機することを前提とする必要がある．

以上の実験結果を踏まえ，各インタラクションパターンに応じたユースケースアイデアを下記の通り考案した．

- Projector Camera Moving: 移動を伴うことを特徴とするため，モデルルームや自宅において，任意の壁面に対して移動しつつ壁紙の変更を行うなど，リフォームのシミュレーションで活用
- Object Translation: 直感的な操作により実在感のあるイメージを投影できることから，子供が動かして構築したブロックや積み木に都市空間などを投影する創造体験で活用
- Object Rotation: 任意のオブジェクト・任意の向きにイメージを投影できることから，デザイナーや営業が商材（車や靴など）の向きを変えながら顧客のイメージを投影するデザイン支援で活用
- Object Deformation: 形状をその場で変えてイメージを投影できることから，粘土を使って国宝の壺や土偶などの文化財（追加学習モデルを使用）を模倣し独自の作品を作り上げる文化財創作体験で活用

これらユースケースアイデアの実装とユーザー評価を通じて、本提案手法の有効性を評価することが今後の課題となる．

5. 結論

本研究では，専門的な知識がなくても簡単にセットアップでき，立体的な投影対象に対して位置合わせされた映像をその場で生成する即興プロジェクションマッピングシステムを提案した．これにより，プロジェクタ・カメラ間のキャリブレーションが不要で，セットアップを簡素化し可搬性を向上させたプロジェクションマッピングを実現した．

実験では，Projector Camera Moving，Object Translation，Object Rotation，Object Deformation の 4 つのインタラクションパターンに対し，いずれも大きな位置ずれなく，投影対象の形状に即して生成画像を投影できていることを確認した．

今後の課題として次の点が挙げられる．

- 境界部の位置ずれを軽減させるような位置合わせ手法の検証
- 投影までの時間短縮
- ユースーケースを想定した被験者による本提案手法によるインタラクションの評価

参考文献

[1] teamLab Inc.: Sketch Aquarium, `https://www.teamlab.art/w/aquarium/` (2013-). Accessed: 2024-12-06.

[2] 渡辺義浩: 顔へのプロジェクションマッピングによる化粧体験，システム制御情報学会　研究発表講演会講演論文集， Vol. SCI24, pp. 615–616 (2024).

[3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B.: High-Resolution Image Synthesis With Latent Diffusion Models, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695 (2022).

[4] Zhang, L., Rao, A. and Agrawala, M.: Adding Conditional Control to Text-to-Image Diffusion Models (2023).

[5] Lin, R. R., Ke, Y. and Zhang, K.: Urban Symphony: An AI and Data-Driven Approach to Real-Time Animation for Public Digital Art, *Proceedings of the 16th International Symposium on Visual Information Communication and Interaction*, VINCI '23, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3615522.3615553 (2023).

[6] Borgomano, G.: Prompt to Anything?: Exploring Generative AI's Iterative Potential in Mapping Show Production., *ACM SIGGRAPH 2024 Talks*, SIGGRAPH '24, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3641233.3664729 (2024).

[7] Allen, S.: Unreal Pareidolia -shadows-, `https://scottallen.ws/work/unreal-pareidolia-shadows/` (2023). Accessed: 2024-12-06.

[8] Erel, Y., Kozlovsky-Mordenfeld, O., Iwai, D., Sato, K. and Bermano, A. H.: Casper DPM: Cascaded Perceptual Dynamic Projection Mapping onto Hands, *SIGGRAPH Asia 2024 Conference Papers*, SA '24, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3680528.3687624 (2024).

[9] Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M. and Felsberg, M.: RoMa: Robust Dense Feature Matching, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19790–19800 (2024).

[10] Bhat, S. F., Birkl, R., Wofk, D., Wonka, P. and Müller, M.: ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth (2023).

[11] Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J. and Rombach, R.: SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis (2023).

[12] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models., *ICLR*, OpenReview.net (2022).

[13] Ye, H., Zhang, J., Liu, S., Han, X. and Yang, W.: IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models (2023).

[14] Guo, Y., Yang, C., Rao, A., Liang, Z., Wang, Y., Qiao, Y., Agrawala, M., Lin, D. and Dai, B.: AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning (2023).

[15] Ren, Y., Xia, X., Lu, Y., Zhang, J., Wu, J., Xie, P., Wang, X. and Xiao, X.: Hyper-SD: Trajectory Segmented Consistency Model for Efficient Image Synthesis (2024).