

ディープラーニングによる 顔のスケッチ画像からの3Dモデル生成

川中 雄太^{1,a)} 佐藤 周平^{1,b)} 櫻井 快勢² 高尚策¹ 唐 政¹

概要: 近年コンピュータグラフィックス (CG) は映画やゲームを含む様々な分野で利用されており, CG を制作するための様々な機能を備えたソフトウェアが広く普及している. しかし CG ソフトを使用するのは敷居が高く, 初心者が自由に扱えるようになるには長い時間が必要となる. そこで本研究では, CG ソフトの機能の一つであるモデリングに着目し, 顔のスケッチ画像を用いて簡単に3次元の顔のモデルを生成する手法の研究を行っている. 本発表では, 学習データの準備の仕方を工夫することで, 生成される3次元モデルの精度が向上したので, その結果について報告する.

キーワード: 深層学習, モデリング, 顔モデル, 線画

1. はじめに

近年, コンピュータグラフィックス (CG) 技術を用いた多くの映像作品やゲームなどが作られている. それに伴い, 高機能の CG 制作用のソフトウェアが開発され, モデリングやシミュレーション, レンダリングなど様々な機能が含まれているソフトや一つの機能に特化したソフトなど様々な種類が存在する. しかし, それら高機能のソフトを使用して映像を作るためには, 専門的な知識が必要であり, また, これらのソフトに特有の複雑なユーザインタフェースに慣れるには長い時間を要する. そのため, 初心者が使いこなせるようになるには敷居が高い.

この問題を解決するために, 簡単な入力でも CG 制作を行うための手法が数多く研究されている. その中でも, ユーザによりスケッチされた線画を入力として, その入力の通りの3次元モデルを得るための手法がいくつか提案されている [1], [2], [3]. 3次元モデルをソフトにより作成する場合, モデルの各頂点を直接調整する必要があるなど, 難易度が高く, 時間のかかる操作が要求される. そういった問題に対し, 上記の手法ではラフなスケッチを描くことで3次元モデルを作成でき, 専門家でなくとも簡単かつ短時間で所望のモデルを得られる.

我々もこの考え方にに基づき, 線画から3次元モデルを生成する深層学習ベースのシステムを提案する. 本研究では人間の顔に着目し, 顔の線画を入力すると, 線画の通りの顔の3次元モデルを生成する. 3DCG ソフトにより作成した顔のモデルとそれを線画風にレンダリングした画像をセットとして学習データを構築する. 本手法では, 深層学習モデルとして, Saito らが提案した, 人物の全身画像からその3次元モデルを再構成する, PIFu (Pixel-Aligned Implicit Function) と呼ばれる手法 [4] を線画へ応用する. しかし, 単に PIFu に学習させるデータを線画へ変更しただけでは, 良い精度が得られなかったため, 損失関数を計算する際のサンプリングをアダプティブにする, 線画の線の太さを変化させることで学習データを増やす, という工夫により学習モデルの精度が向上したので, その結果について本稿では報告する.

2. 関連研究

2.1 線画を入力とした3次元モデル生成

簡単な入力により所望の3次元モデルを作成することを目的として, 線画を入力とする手法がいくつか提案されている. Delanoy らは, 複数視点からの線画を入力することで, 多様な形状の3次元モデルを作成する手法を提案した [3]. しかし, 平らな面に凹凸が見られたり, 角が丸まったりしてしまうなど, 生成されるモデルの精度は十分でない. Li らは, 線画の描画中にインタラクティブに3次元モデルが生成される手法を提案した [2]. この手法は, モデルの大きな形状を正面から線画として描き, 加えてモデ

¹ 富山大学
University of Toyama

² 株式会社ドワンゴ
DWANGO Co., Ltd.

a) m2171116@ems.u-toyama.ac.jp

b) ssato@eng.u-toyama.ac.jp

ルの奥行き方向の凹凸やふくらみなどを表現するために異なる視点からの線画を追加することで、奥行きについても指定可能なシステムを作成した。一方我々は、1枚の正面から描いた顔の線画のみを入力とし、鼻や耳などの複雑な形状まで精度よく生成することを目的としている。

顔のモデルに特化した手法として、Hanらは、線画を入力として顔の3次元モデルを得るための手法 [1] を提案した。この手法では、単一の顔モデルを変形させた様々な表情のモデルをデータベースとして用意し、各データとそれを線画風にレンダリングした画像とを学習データセットとして深層学習モデルを構築する。そして、入力された線画に最も近いデータベース内の顔モデルを、構築した学習モデルを介して出力する。しかし、この手法はデータベース内の顔モデルを検索するようなシステムとなっており、用意したデータ以外のモデルを生成するようなことはできない。そのため、似たような表情の線画では全く同じ顔モデルが選択される。本研究では、この手法のようなデータベースの検索ではなく、入力した線画に対応した顔モデルを生成し、微小な変化でも反映されるようなシステムの構築を目的としている。

2.2 PIFu: Pixel-Aligned Implicit Function

PIFuは人物の全身画像を入力として、その3次元モデルを推定し再構成する手法であり、Saitoらによって提案された [4]。本手法では、このPIFuを深層学習モデルとして利用するため、以下でこの手法について簡単に説明する。

PIFuでは学習データとして、人間の全身を表す3次元モデルと、対応する人物のテクスチャ画像をそのモデルに貼り周囲360度からレンダリングした画像群を用意する。PIFuは、入力画像から特徴量を抽出する畳み込みニューラルネットワークをベースとした画像エンコーダと、その特徴量から画像に対応したモデルを作成するための陰関数を予測する多層パーセプトロンをベースとしたネットワークの2つのネットワークから構成されている。この陰関数は、3次元空間上に定義され、モデルの表面上を0.5とし、モデルの内側は(0.5, 1]、外側は[0, 0.5)の範囲の値をとる。学習の際には画像エンコーダに学習データとして用意したレンダリング画像を入力として与えて、特徴量を抽出する。得られた特徴量および入力画像に対応するモデル(正解データ)の表面付近に配置されたサンプル点の位置座標を、多層パーセプトロンをベースとしたネットワークに入力として与え、サンプル点の位置に対応した陰関数値として予測データを得る。そして、各サンプル点の予測データと正解データの値(モデルの内側の場合は1、それ以外の場合は0)との平均二乗誤差を計算し、その誤差が最小になるように2つのネットワークを学習させる。

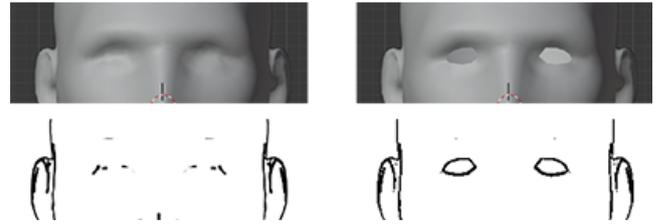


図1: 学習用モデル(左)とレンダリング用モデル(右)。それぞれ上段のモデルを線画風にレンダリングした画像が下段である。レンダリング用では目の輪郭線がはっきりと描画されるよう、目の部分のメッシュを除去している。

3. 提案手法

本手法では、様々な人物の顔の3次元モデルと、それらを線画風にレンダリングした画像をデータセットとし、PIFu [4]と同様のネットワークにそのデータセットを学習させる。顔の3次元モデルは一般的なCG制作ソフトウェアを用いて手動で作成し、また顔のサイズを自動で変化させることでデータを増やした。そして、これらの3次元モデルを同じくソフトを使用して線画風にレンダリングした画像を用意する。この画像は、Saitoらの場合と同様に顔の周囲について1度ずつ異なる角度から見た場合のレンダリング画像を360度分用意する。この時、線の太さの違い複数種類の画像を用意することで入力する線画の線の太さに対してロバストに3次元モデルを生成できる(3.1節)。また、これらのレンダリング画像において、顔の領域とそれ以外の領域を区別するためのマスク画像を合わせて用意する。ただし、学習データとして用意したメッシュが閉じたモデル(学習用モデル)に対し線画風のレンダリングを行った場合、目の部分にはっきりと線が現れなかったため(図1左)、目の部分のメッシュを除去したモデルをレンダリング専用として用いた(図1右)。

続いて、上記の方法で様々な顔のモデルおよび画像を作成してデータセットとし、PIFuのネットワークに学習させる。学習時には、ネットワークからの出力(予測データ)と正解データとの差分を損失関数として、それを最小化するように学習を進める。この損失関数は、正解データとなる3次元モデルの表面の周りに配置されたサンプリング点に基づいて評価されるが、本手法ではこのサンプリング点の数をモデル表面の曲率に応じて適応的に変化させることで学習精度が向上することが分かった(3.2節)。最後に、学習済みのネットワークを用いて線画から3次元モデルを得る。ただし、ネットワークの出力は陰関数として表現されるため、メッシュ化を行う必要があるが、Saitoらの研究 [4] で用いているマーチングキューブ法 [6] では結果にアーティファクトが確認されたため、制限付きドロネー三

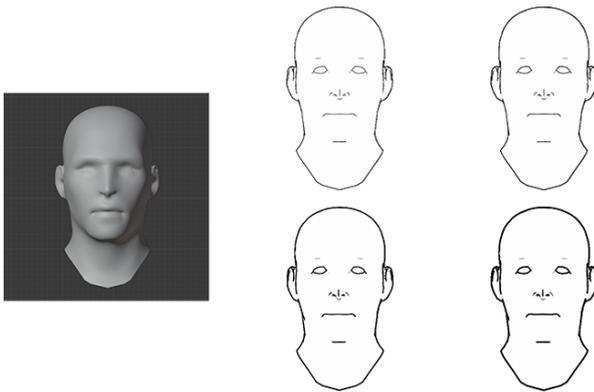


図 2: 正解データとなる顔の 3 次元モデル (左) と線の太さの異なる 4 種の線画風レンダリング画像 (右).

角形分割を用いて滑らかなメッシュを得る (3.3 節). 以下では本手法における 3 つの工夫について詳細を説明する.

3.1 様々な線の太さの線画の利用

Saito らの手法 [4] では, 3 次元モデルからレンダリング画像を作成する際に, 様々な種類の照明を様々な方向から当ててレンダリングした画像を用意することで, 入力がどのような光源下でもロバストに再構成ができる. 一方, 本研究で扱う線画では, そのような光源による変化はないため, 単一のモデル, 単一の視点に対して 1 種類の線画のみが生成される. この場合, 入力する線画の線の太さが学習時のものに制限されてしまい, 異なる太さの線画を入力すると部分的に形状の崩れたモデルが生成されてしまう. そこで本研究では, 線の太さの異なる複数種類の線画風レンダリングの画像を学習データセットとして用意する (図 2). そして, 学習時の 1 回の計算ごとにそれらの中からランダムに使用する画像を選択する. このようにすることで, 線の太さに対してロバストに 3 次元モデルを生成することができる.

3.2 モデル表面の曲率に基づく適応的サンプリング

Saito らの手法 [4] では, 詳細な形状まで正確に表現できるように, 3 次元モデルの表面から一定の範囲内に配置されたサンプル点において損失関数を評価する. サンプル点数はモデル表面全体に一様に分布するように設定され, 表面上のある点を中心とした正規分布に基づいた確率に従って配置される.

これを顔のモデルに適用した場合, 耳や鼻といった凹凸の多い個所で生成されるモデルの精度が低下することを我々は実験により確認した. そこで本手法では, 顔モデルの表面を表すメッシュの曲率を基にサンプル点の数を適応的に変化させる. 具体的には, 三角形メッシュの各頂点毎に 2 つの主曲率から平均曲率を算出し, 3 つの頂点についてその平均値を求める. そしてその平均値の大きさに応じ

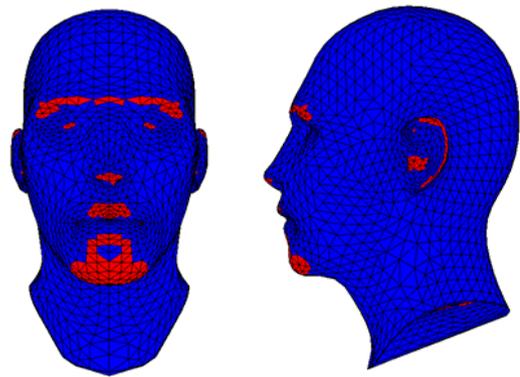


図 3: 曲率の平均値によりメッシュを 2 つのグループに分けた様子. 青が平均値の低いメッシュであり, 赤が高い部分.

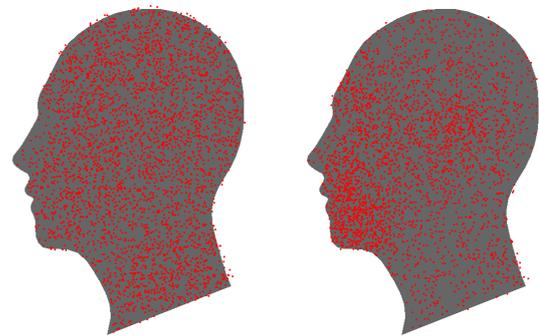


図 4: 一様な場合と適応的な場合における配置されるサンプル点の比較. 左が一様な場合であり, 右が適応的な場合.

て各メッシュを 2 つのグループに分け (図 3), 平均値が高いグループの周辺にはサンプル点を多く, 低いグループには少なく配置する. 2 つのグループに分けるための平均値に対する閾値はユーザが手動で設定する. 図 4 は, 従来の一様にサンプル点を配置した場合と, 提案手法による適応的な配置をした場合との比較を示す. 図 3 において曲率が高いと判定された, 顎や口, 目などの周辺に点が多く集まっているのがわかる.

3.3 制限付きドロネー三角形分割

PIFu のネットワークの出力は陰関数として表現されたボリュームデータであり, 最終的に顔の 3 次元モデルを得るためには, そのボリュームデータを基にメッシュ化を行う必要がある. Saito らの研究 [4] では, このメッシュ化にマーチングキューブ法 [6] を用いているが, 我々の実験では結果のモデルに縦縞模様のようなアーティファクトが確認された. そこで本実験では, マーチングキューブ法 [6] に代えて, 制限付きドロネー三角形分割 [5] を用いた. これにより, マーチングキューブで確認されたようなアーティファクトがなく, 滑らかなメッシュを得ることができた. この結果については次節にて示す.

4. 実験結果

提案手法による実験結果を図 5-8 に示す。本稿の実験では、ネットワークの学習時のエポック数を 8 回、学習データとしてトレーニング用データを 100、テスト用のデータを 4 つ用意し、それに対応したレンダリング画像とマスク画像をそれぞれトレーニング用 144000 枚、テスト用 5760 枚用意して学習を行った。ネットワークの学習には Adam を用いた。

まず学習に用いる線画の種類を変えた場合の結果の比較を示す。線画が 1 種類の場合には図 5 の (a) のように、学習時にない太さの線を含む線画が与えられた際に、出力されるモデルの形状が大きく崩れてしまっている。一方、提案手法では (b) のように形を崩すことなくロバストに 3 次元モデルを生成できた。

次に、適応的なサンプリングを行った場合の結果の比較を図 6 に示す。モデル全体のサンプル点の数はどちらの場合も同一である。一様にサンプル点を配置する場合には、耳の表面の形状が崩れているのがわかる (図 6 左)。一方、我々の曲率に基づく配置では、そのような問題はなく滑らかな形状が得られている。このことから、曲率に基づいてサンプル点の粗密を変化させることで効率的に学習できていることが分かる。

次に、マーチングキューブ法 [6] と制限付きドロネー三角形分割でメッシュ化した場合の比較を図 7 に示す。マーチングキューブ法 [6] を用いるとモデルの表面に縦縞のようなアーティファクトがはっきり確認できる。しかし、制限付きドロネー三角形分割を用いた場合には、そのようなアーティファクトはなく、表面が滑らかになっていることがわかる。

最後に、提案手法により線画から 3 次元モデルを生成した例を図 8 にいくつか示す。それぞれの結果において、左の画像が入力として与えた線画の画像、右の 3 つの画像が出力された 3 次元モデルを正面、側面、斜めから表示した画像となっている。どのモデルに関しても線画の通りのモデルが生成できており、鼻や口、耳など複雑な形状に対しても形が崩れることなく表現できていることが確認できる。

5. まとめと今後の課題

本稿では、ディープラーニングを用いて線画から 3 次元の顔モデルを生成する手法を提案した。1 枚の画像から人間の全身のモデルを作成する手法である PIFu を応用し、顔モデルとそれを線画風にレンダリングした画像をデータセットとして用意し学習させた。この際、線画の線の太さに対してロバストにモデルを生成するために、線の太さの異なる 4 種類の線画を用意した。また、学習精度を効率的に向上させるために、正解データとなる顔モデルをサン

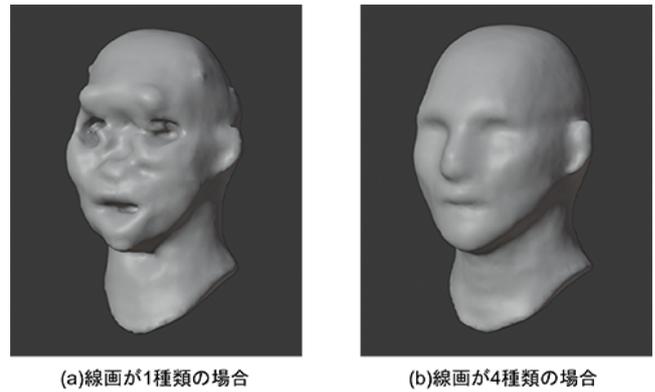


図 5: 学習に用いる線画の種類による結果の比較。左が 1 種類、右が 4 種類を用いた場合の結果。

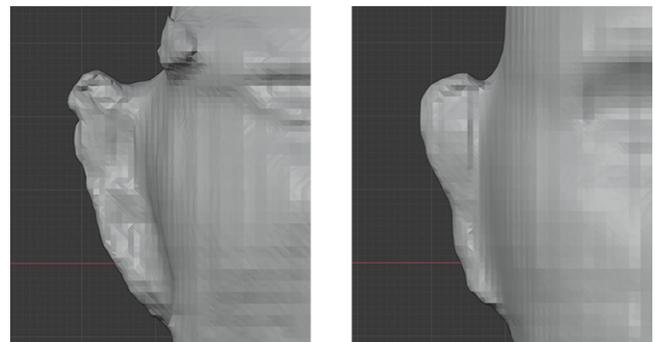


図 6: サンプル点を一様に配置した場合 (左) と曲率に基づいて適応的に配置した場合 (右) の比較結果。

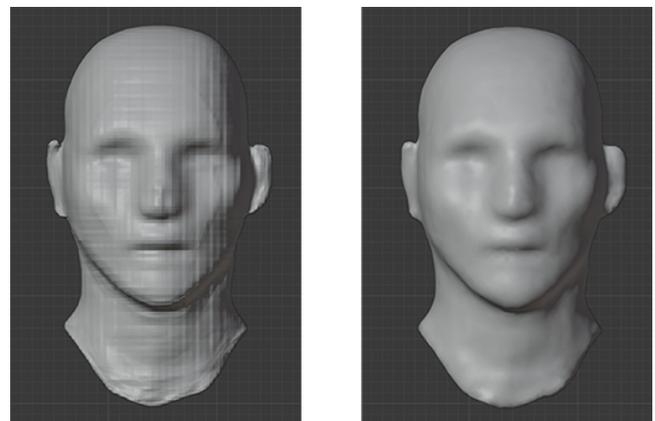


図 7: マーチングキューブ法 (左) と制限付きドロネー三角形分割 (右) によるメッシュ化の比較。

ルする点の位置を、表面を表すメッシュの曲率に応じて適応的に変化させた。また、メッシュ化に制限付きドロネー三角形分割を用いることで、滑らかなメッシュが得られることを確認した。

今後の課題として、現状では頭髪や髭などがない顔のみのモデルを対象としており、それらを含むモデルを生成することはできない。また、頭髪や髭などは顔を表すメッシュとは異なり、曲線や粒子といった別の表現が用いられる場合が多い。そのため、今後は頭髪などのようなメ

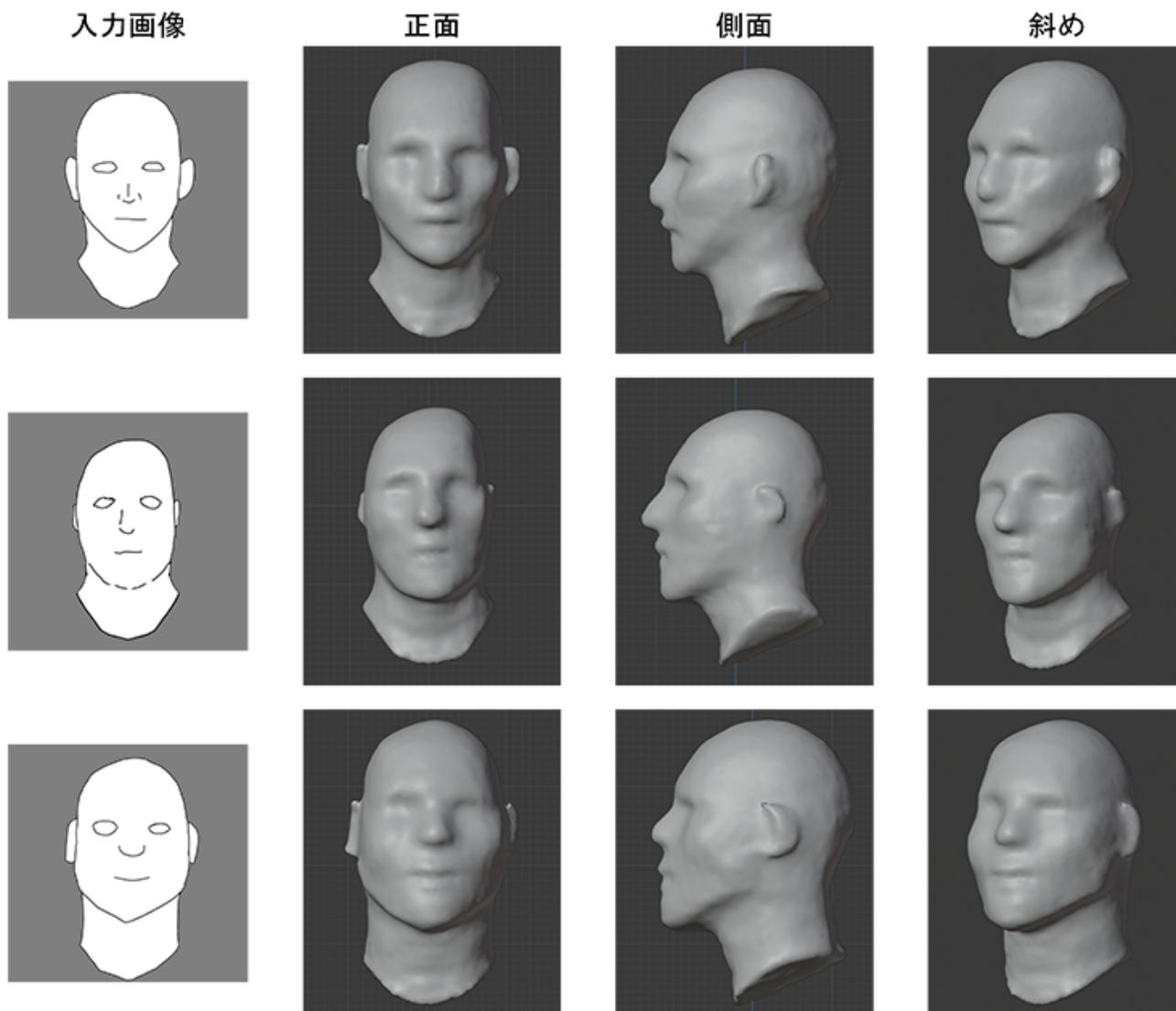


図 8: 線画を入力として与えた際の出力結果

シュとは異なる表現形式の部分を線画から生成できるような手法の確立を目指し研究を進める。

参考文献

- [1] Xiaoguang Han, Chang Gao, Yizhou Yu.: *DeepSketch2Face: A Deep Learning Based Sketching System for 3D Face and Caricature Modeling*, ACM Trans. Graph., vol. 36, no. 4, pp. 126:1–126:12, Aug. 2017.
- [2] Li, Changjian, Pan, Hao, Liu, Yang, Sheffer, Alla, Wang, Wenping: *Robust Flow-Guided Neural Prediction for Sketch-Based Freeform Surface Modeling*, ACM Trans. Graph. (SIGGRAPH ASIA), vol. 37, no. 6, pp. 238:1–238:12, Nov. 2018.
- [3] Delanoj, J., Aubry, M., Isola, P., Efros, A. A., Bousseau, A.: *3d sketching using multi-view deep volumetric prediction*. Proceedings of the ACM on Computer Graphics and Interactive Techniques, vol. 1, no. 1, pp. 21:1–21:22, 2018.
- [4] Saito, Shunsuke, Huang, Zeng, Natsume, Ryota, Morishima, Shigeo, Kanazawa, Angjoo, Li, Hao: *PIFu: Pixel-Aligned Implicit Function for High-Resolution*

Clothed Human Digitization, The IEEE International Conference on Computer Vision (ICCV), vol. 25, no. 3, pp. 624–630, Jul. 2019.

- [5] Sloan, S. W.: *A Fast Algorithm for Generating Constrained Delaunay Triangulations* Computers and Structures, 47, 3, 441-450, 1993.
- [6] William E Lorensen and Harvey E Cline.: *Marching cubes: A high resolution 3d surface construction algorithm*. In ACM siggraph computer graphics, volume 21, pages 163–169.ACM, 1987.