

一枚の顔画像とテキストから 360 度の 3D 顔モデル自動生成

Automatic 3D face whole model generation from a single face image and text

平岩 篤信*¹
Atsunobu Hiraiwa

大西 一貫*¹
Kazuhiro Onishi

*¹ 株式会社博報堂テクノロジーズ
Hakuhodo Technologies inc.

一枚の顔画像とテキストを条件として与えることにより、360 度の 3D 顔モデルを自動生成するシステムを提案する。一枚の顔画像のみでは、十分な画像品質がえられない。またテキストを併用より、顔の表情を変える、3D 顔モデル修正についても実験結果を報告する。さらに技術的課題点と今後の AI 活用に向けた研究の方向性を示唆する。

1. はじめに

テキストから 3D モデルを自動生成するシステムは DreamFusion [Poole 2022] が提案され、Stable-dreamfusion [Tang 2022]により実装が公開されている。さらに Threestudio [Guo 2023]により 3D コンテンツ生成のフレームワークが公開されている。

一枚の顔画像とテキストを条件として与えることにより 360 度の 3D 顔モデルを自動生成するシステムを提案する。一枚の顔画像のみでは、十分な画像品質がえられない。またテキストを併用より、顔の表情を変える、3D 顔モデル修正についても実験結果を報告する。

2. ではシステムフローと構成およびデータ要件について紹介し、3. で実験に使用する試験画像および定性評価について述べる。4. ではまとめと課題について考察する。

2. Zero-1-to-3 を利用した 3D モデル生成

全体の処理の流れを図1に示す。開始点ではユーザはシステムに 3D モデル化したい画像を入力し、Depth 情報による背景情報を削除し、透明化する。図2に適用例を示す。BLIP-2 [Li 2023]を呼び出し、画像から TEXT を自動抽出する。AI 技術者が自動抽出された TEXT に情報を追加し、プロンプトとする。Zero-1-to-3 [Liu 2023] を呼び出し、学習を行い、検証により 3D モデルの 360 度動画を得る。

3. 試験画像を用いた 3D モデル検証

試験画像として stable-diffusion を用いた生成画像と写真を用いた顔の 3D モデルの検証とテキストを加え表情を笑顔に変換する検証例を示す。

3.1 画像のみを入力とする検証

図3に、360度動画の8方向からの 3D モデルを示す。図3で使用した画像では、stable-diffusion V1.5 で作成した生成画像を使用し、背景情報を透明化した図2-1を用いた。図2-2では、写真1枚から背景情報を透明化した画像を使用する。図4-1では3方向からの 3D モデルをする。背景の削除で顔との境界線でボケが発生し 3D 顔モデルの側面では、髪の毛の白色化する不都合が確認されている。

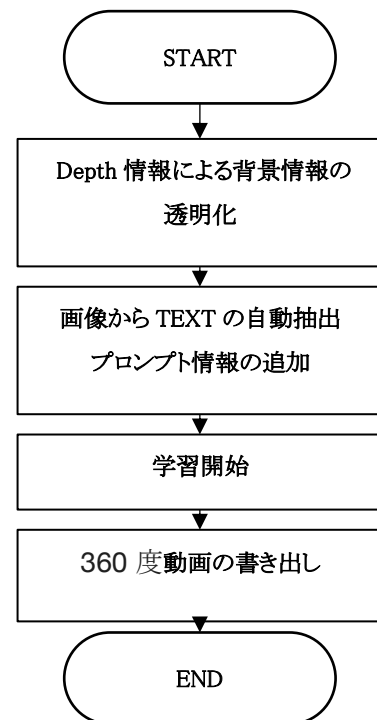


図1 360度動画書き出しまでの処理の流れ

3.2 画像とテキストを入力とする検証

テキストは、BLIP-2-flan-t5-xl 用いて図 2-2 の画像から “a young man in a blue shirt” を自動抽出した。さらに表情を笑顔に変換するためにプロンプトを “--text A DSLR 3D photo of a young man is smiling in a vietnamese --negative Nude, naked, logo, (low quality)1.4, (middle quality)1.4, (gray scale)1.4, (monochrome)1.4, unhappy, sad, mustache” に修正した。図4—2は、プロンプトによる笑顔への変換例を示す。なお、この学習には、RTX3090 で2時間必要であった。



図2-1 画像から背景削除



図2-2 画像から背景削除



図3 8方向からの3Dモデル



図4-1 3方向からの3Dモデル



図4-2 プロンプトによる笑顔への変換

4. まとめと課題

Zero-1-to-3 を用いての生成画像と写真画像から、3D 顔モデルが生成できた。背景の削除で顔との境界線でボケが発生し、3D 顔モデルの側面での不都合が確認されており、対策が必要である。またテキストによる顔の修正についても、詳細なプロンプトにより画像の品質を改善することを確認している。画像からテキスト自動抽出技術の向上に期待する。

本研究では、動画広告を生成するシステムを実現することを想定して一部機能である Zero-1-to-3 を利用した 3D モデル生成の結果から、技術的課題点を明確にし、今後の AI 活用に向けた研究の方向性を示唆するものである。

参考文献

- [Poole 2022] Ben Poole, Ajay Jain, Jonathan T. Barron, Ben Mildenhall, “DreamFusion: Text-to-3D using 2D Diffusion”, <https://arxiv.org/pdf/2209.14988.pdf>, 2022.
- [Tang 2022] Jiayang Tang, “Stable-dreamfusion: Text-to-3D with Stable-diffusion”, <https://github.com/ashawkey/stable-dreamfusion>, 2022.
- [Liu 2023] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, Carl Vondrick, “Zero-1-to-3: Zero-shot One Image to 3D Object”, <https://arxiv.org/pdf/2303.11328.pdf>, 2023.
- [Guo 2023] Yuan-Chen Guo, Ying-Tian Liu, Chen Wang, Zi-Xin Zou, Guan Luo, Chia-Hao Chen, Yan-Pei Cao, Song-Hai Zhang, “threestudio: A unified framework for 3D content generation”, <https://github.com/threestudio-project/threestudio>, 2023.
- [Li 2023] Junnan Li, Dongxu Li, Silvio Savarese, Steven Hoi, “BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models”, ICML2023.