Evaluating End-to-End and Multi-Stage 3D Model Generation with 3D Morphable Models: A Comparative Study on Eyeglasses Frames

Ervinas GISLERIS, Artūras SERACKIS

E-mail: {ervinas.gisleris|arturas.serackis}@vilniustech.lt,
Department of Electronic Systems, Vilnius Gediminas Technical University (VILNIUS TECH)

Introduction

In the domain of product visualization, the ability to generate precise 3D models from 2D images is pivotal to enhancing user experience and engagement. Our investigation conducts an exhaustive comparative analysis of two distinct methodologies in generating 3D models of eyeglasses frames from a pair of product images: the end-to-end 3D model generation and a multi-stage approach utilizing 3D Morphable Models (3DMM). The end-to-end generation approach leverages deep learning algorithms to directly map input images into 3D models, ensuring an uninterrupted and expedited transformation. On the other hand, the multi-stage approach begins with the extraction of pertinent features from the 2D images, followed by segmentation and the application of 3DMM for shape reconstruction.

Aims and Goals

The aim of the presented investigation is to develop a model capable of generating accurate 3D eyeglass frame models from 2D images, enhancing the transition from 2D product visualization to 3D product interaction.

Main objectives:

- Create a multi-stage model to transform 2D images into 3D eyeglass frames, focusing on high fidelity and detailed representation.
- Evaluate the proposed model against existing models on metrics such as accuracy, efficiency, and detail capture.

Methods

In developing our proposed model for the generation of 3D eyeglass frames, we first applied the YOLO object detection algorithm to segment eyeglass frames from 2D images. This was followed by contour extraction to delineate the frame's edges. We then implemented an automatic key-point selection process, relying on a clustering algorithm to pinpoint essential frame features. These key points were aligned with a 3D mesh template, allowing the update of mesh coordinates and culminating in the accurate construction of the 3D eyeglass frame output.

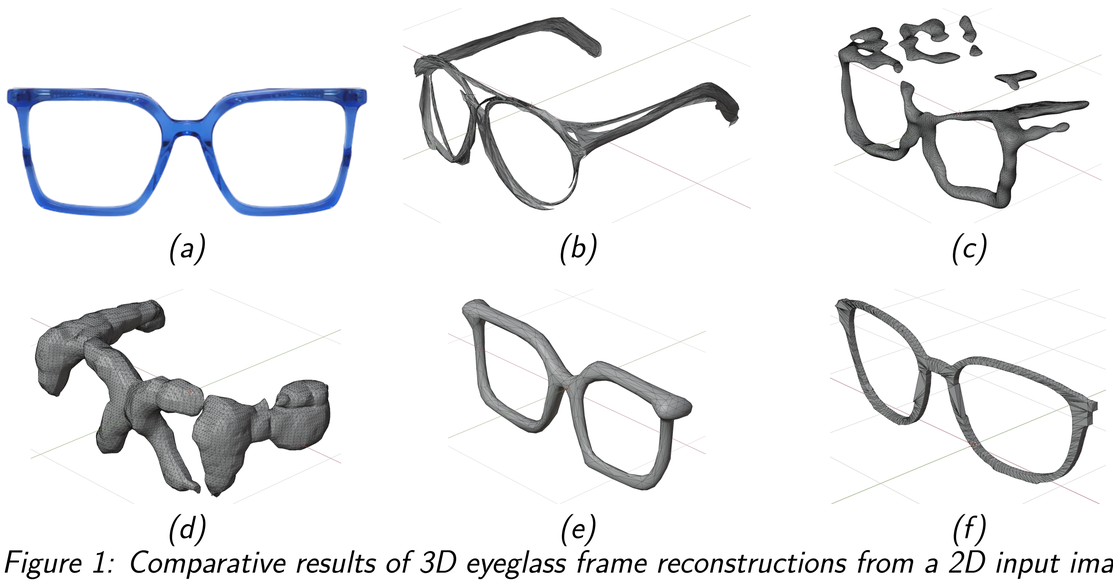

Figure 1: Comparative results of 3D eyeglass frame reconstructions from a 2D input image, showcasing outputs from various generation methods: (a) Front view input image, (b) AtlasNet [1], (c) Wonder3D [2], (d) One-2-3-45 [3], (e)One-2-3-45++ [4], and (f) ours.

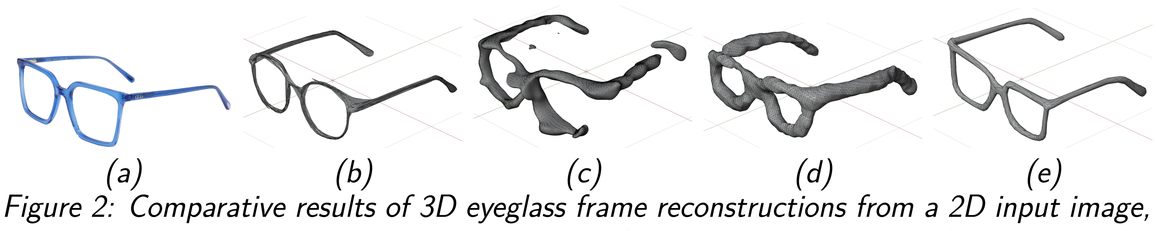

Figure 2: Comparative results of 3D eyeglass frame reconstructions from a 2D input image, showcasing outputs from generation methods: (a) Side view input image, (b) AtlasNet, (c) Wonder3D, (d) One-2-3-45, and (e) One-2-3-45++.

For AtlasNet, we addressed the lack of a pre-trained eyeglass model by creating a tailored dataset of 5,000 synthetic 3D eyeglass frames for training purposes. This approach was necessary to equip AtlasNet with the ability to render eyeglass frames accurately. In contrast, the comparative models Wonder3D, One-2-3-45, and One-2-3-45++ did not require additional data preparation or training adjustments. These models were directly utilized in their standard configurations by assessing their out-of-the-box capabilities to generate 3D models.

Results

The One-2-3-45++ model followed closely, demonstrating its ability to recreate very similar 3D objects, albeit with slightly longer processing times. AtlasNet showed a propensity for bias in its generated 3D mesh, which requires model modification.

Table 1: Processing Time Comparison for 3D Model Generation

Model	Computing Environment	Avg. Time: Front View (s)	Avg. Time: Side View (s)
AtlasNet	Local	11	11
Wonder3D	Local	345	344
One-2-3-45	Cloud	65	63
One-2-3-45++	Cloud	79	81
Ours	Local	33	-

Conclusions

- The highest fidelity is achievable using only multi-stage aproach;
- AtlasNet is the fastest approach, however, no suitable dataset for eyeglasses is available.

References

[1] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[2] Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, and Wenping Wang. Wonder3d: Single image to 3d using cross-domain diffusion, 2023.

[3] Minghua Liu, Chao Xu, Haian Jin, Linghao Chen, Zexiang Xu, Hao Su, et al. One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization. *arXiv preprint arXiv:2306.16928*, 2023.

[4] Minghua Liu, Ruoxi Shi, Linghao Chen, Zhuoyang Zhang, Chao Xu, Xinyue Wei, Hansheng Chen, Chong Zeng, Jiayuan Gu, and Hao Su. One-2-3-45++: Fast single image to 3d objects with consistent multi-view generation and 3d diffusion. *arXiv preprint arXiv:2311.07885*, 2023.