Scribble: Auto-generated 2D avatars with diverse and inclusive art direction

LOHIT PETIKAM[1], CHARLIE HEWITT[1], SHIDEH REZAEIFAR[2]

MICROSOFT MESH LABS ([1]CAMBRIDGE, UK, [2]ZURICH, SWITZERLAND)

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Microsoft

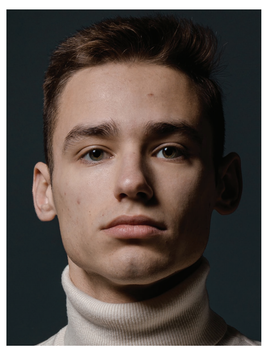
Introduction

Stylized avatars used in hybrid telepresence must support the large diversity in human appearance existing in the world. Most avatar systems assume a generic base rig that morphs into varied face and body shapes. However, retrofitting shapes for under-represented populations as an afterthought is costly. We need new avatar systems, visual languages, and human understanding technology, all co-designed with diversity in mind from the start.

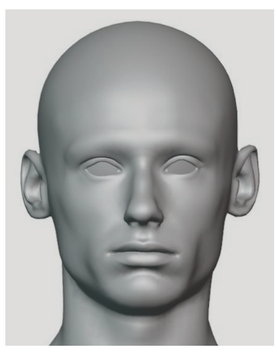
Photoreal avatars [Hu et al. 2017; Nagano et al. 2018] are representative but can fall into the uncanny valley. Existing 3D stylized avatars [Sang et al. 2022; Wang et al. 2023] are auto-generated but have little consideration for diversity. Scribble automatically generates a 2D animatable avatar with jointly-designed art direction and stylization techniques that accommodate diverse human attributes.

Face shape and avatar rig

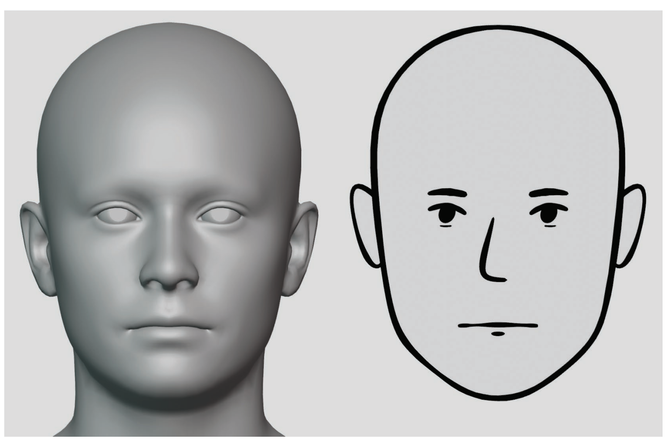
We predict a user's 3D face mesh from a frontal selfie using the method of Wood et al. [2022] and remove pose and expression to estimate their neutral identity face shape.



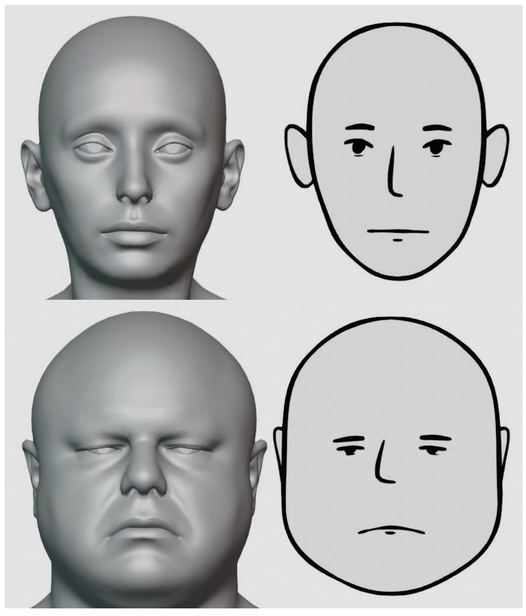
Input selfie



Identity mesh



Template mesh *Base 2D avatar*

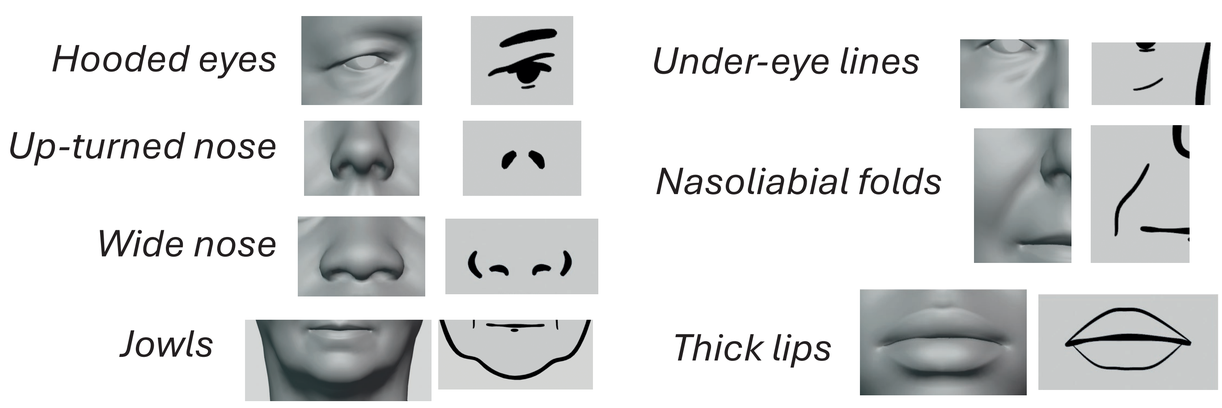


Rig variation with identity

Our avatar rig is a set of 2D curves designed based on the template mesh of Wood et al. [2022]. The curves follow the shape of the predicted mesh depicting users' face shape and relative proportions.

Style mapping

Inspired by how 2D artists draw diverse characters we design a style mapping that depicts diverse facial features in fair and appealing ways.

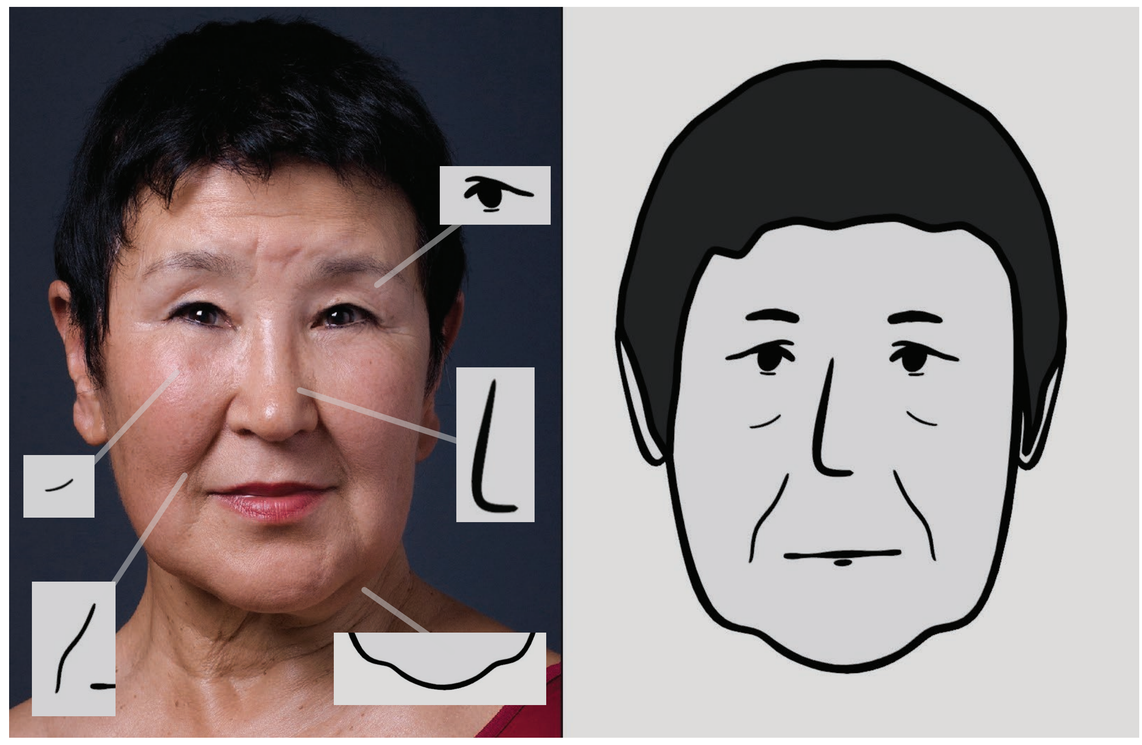


For features in our style mapping, we explicitly detect their presence in the identity mesh using our hand-crafted rules incorporating part-based PCA, surface curvature analysis, and measured distances. For each detected feature we render the corresponding styled curves.

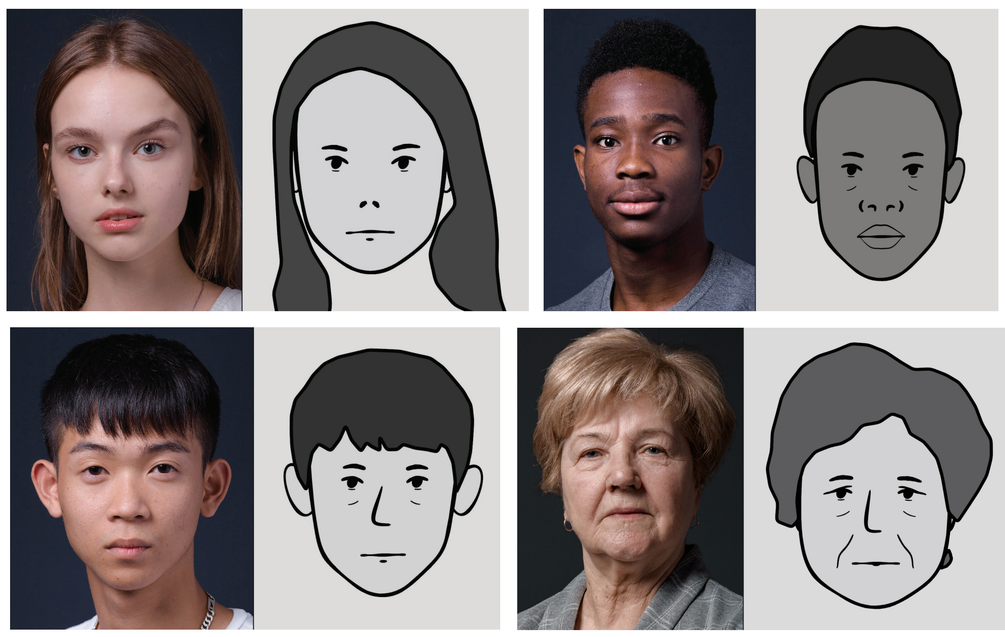
Hair and accessories

We use face segmentation [Wood et al. 2021] to extract skin regions, hair shape and detect accessories. Hair and skin mean colour are extracted and mapped to our grayscale color palette.

Results



Detected: hooded eyes, under-eye lines, long nose, jowls, nasolabial folds



Animation

Our avatars are driven by facial animation supporting the diversity options in our rig, including custom accessory assets.

References

Liwen Hu, Shunsuke Saito, Lingyu Wei, Koki Nagano, Jaewoo Seo, Jens Fursund, Iman Sadeghi, Carrie Sun, Yen-Chun Chen, and Hao Li. 2017. Avatar digitization from a single image for real-time rendering. ACM TOG

Koki Nagano, Jaewoo Seo, Jun Xing, Lingyu Wei, Zimo Li, Shunsuke Saito, Aviral Agarwal, Jens Fursund, Hao Li. 2018. paGAN: real-time avatars using dynamic textures. ACM TOG

Shen Sang, Tiancheng Zhi, Guoxian Song, Minghao Liu, Chunpong Lai, Jing Liu, Xiang Wen, James Davis, and Linjie Luo. 2022. Agileavatar: Stylized 3d avatar creation via cascaded domain bridging. In SIGGRAPH Asia

Shizun Wang, Weihong Zeng, Xu Wang, Hao Yang, Li Chen, Chuang Zhang, Ming Wu, Yi Yuan, et al. 2023. SwiftAvatar: efficient auto-creation of parameterized stylized character on arbitrary avatar engines. In AAAI

Erroll Wood, Tadas Baltrušaitis, Charlie Hewitt, Sebastian Dziadzio, Thomas J Cashman, and Jamie Shotton. 2021. Fake it till you make it: face analysis in the wild using synthetic data alone. In ICCV. Springer

Erroll Wood, Tadas Baltrušaitis, Charlie Hewitt, Matthew Johnson, Jingjing Shen, Nikola Milosavljević, Daniel Wilde, Stephan Garbin, Toby Sharp, et al. 2022. 3D face reconstruction with dense landmarks. In ECCV. Springer