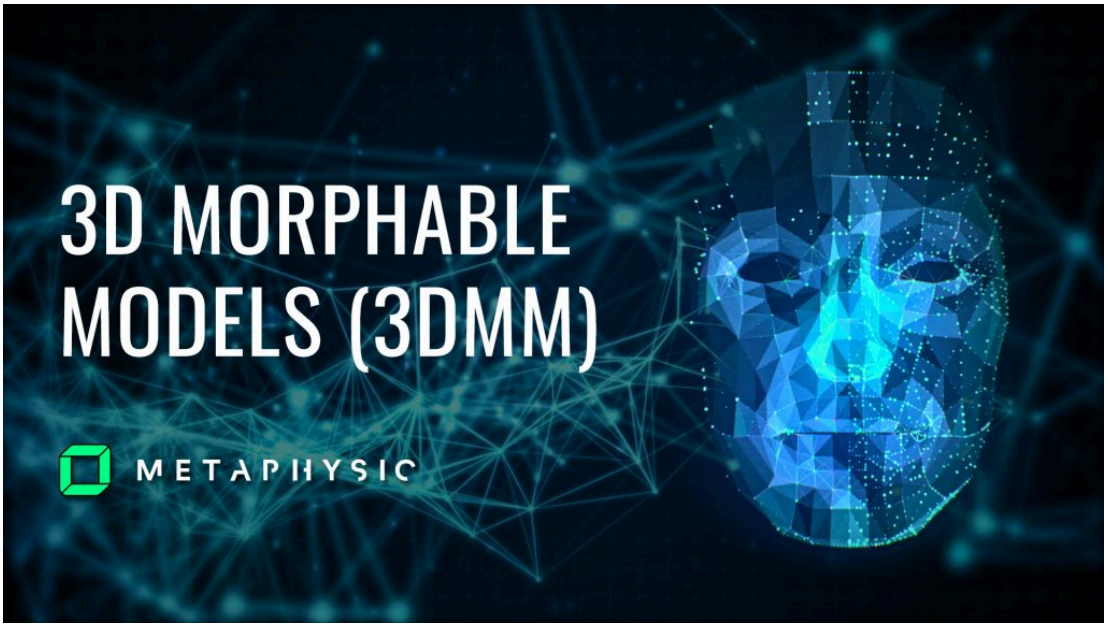
3D Morphable Models (3DMMs)

Originally published at https://blog.metaphysic.ai/3d-morphable-models-3dmms/

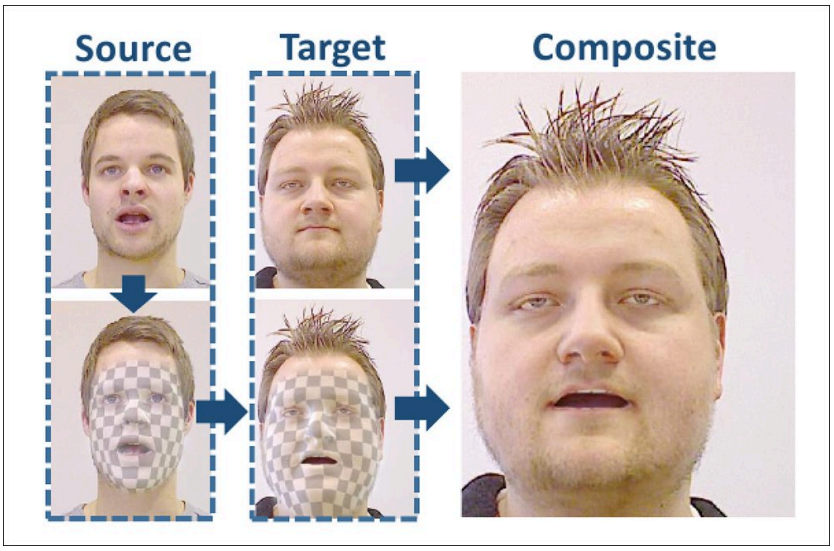
January 30, 2023



3D Morphable Models (3DMMs) are parametric human-focused CGI models that are increasingly being used as a way to interact with the content of the latent space of trained neural image synthesis networks.

In this example, from the StyleRig project, the facial coordinates of a 3DMM mesh are mapped to latent codes in a Generative Adversarial Network (GAN), allowing the user explicit control over viewpoint, appearance, and other aspects. Source: https://www.youtube.com/watch?v=eaW_P85wQ9k

3DMMs were devised and introduced by Volker Blanz and Thomas Vetter of the Max Planck Institute in 1999, in the paper *A Morphable Model For The Synthesis Of 3D Faces.*

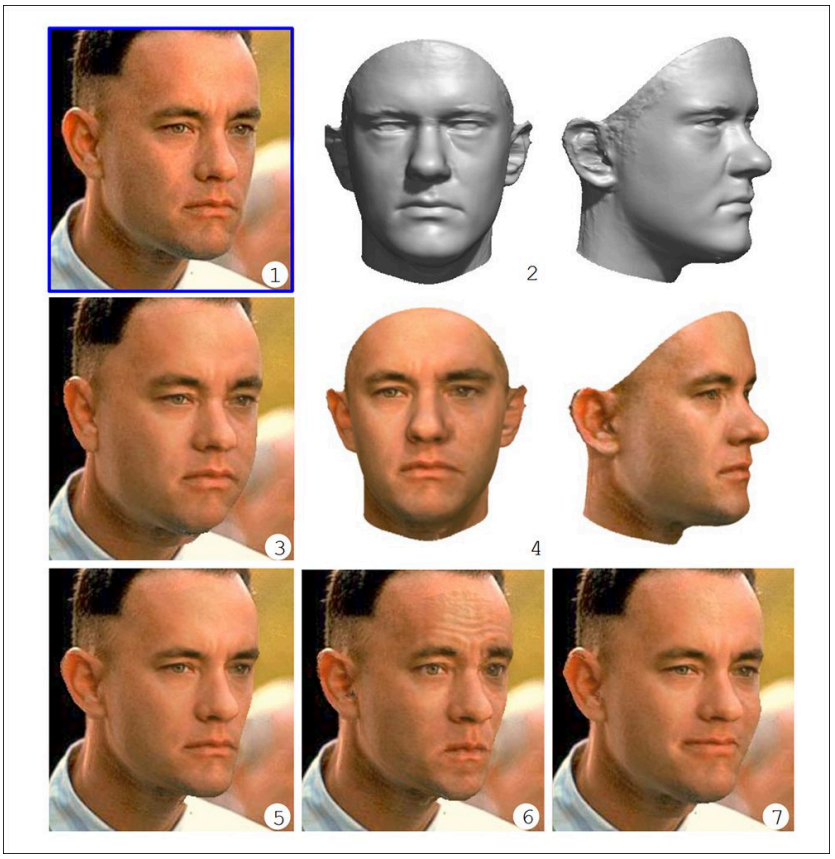


Mapping features from one real face to another real identity, using 3DMM as an intermediary. Source: https://arxiv.org/pdf/1909.01815.pdf

The late 1990s had seen the rise of consumer-level 3D programs, particularly for the Metacreations software stable, which would eventually be divided and sold among a number of companies, including Adobe and Daz.
Among these, the Poser software (which has survived many acquisition rounds) particularly captured the public imagination at the time, for its ability to recreate human faces and bodies, giving rise to various long-lived enthusiast communities, and causing many to believe that convincing recreation of dead movie stars would be a CGI (rather than AI) achievement.

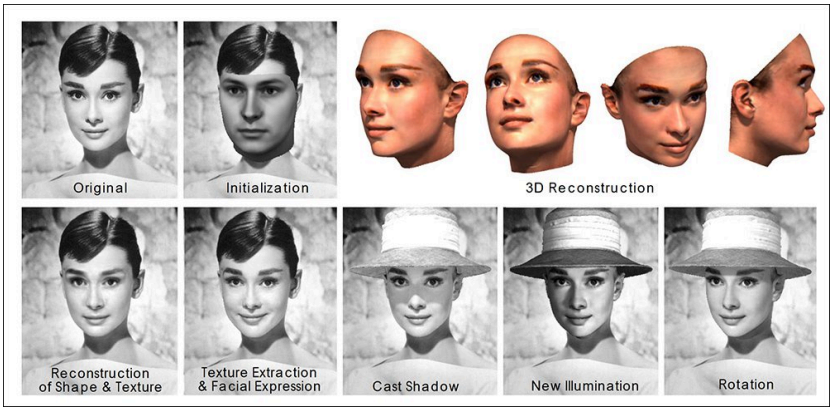
From a 2006 project, Poser is used to create a synthetic dataset. Source: http://lear.inrialpes.fr/people/triggs/pubs/Agarwal-pami05.pdf

Though that was not to be, the Poser-style parametric heads introduced in 1999 would evolve with the computer vision and synthesis research sector, while Poser itself is currently enjoying a revival, along with other pose-synthesis applications, as a template-generator for figurative poses in Stable Diffusion output.



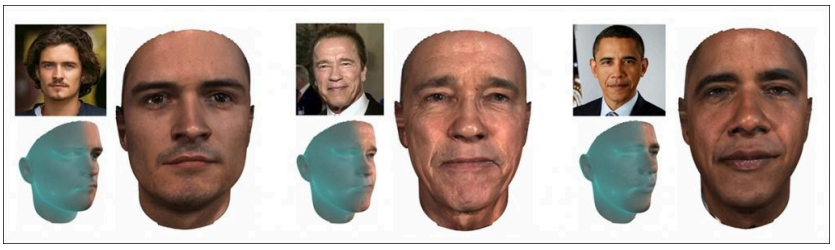
From the original paper for 3DMM in 1999, actor Tom Hanks is recreated parametrically. Source: https://www.face-rec.org/algorithms/3d_morph/morphmod2.pdf

The process of conforming a particular identity to a specific image for a 3DMM involves the initial use of a 'generic' template, and the gradual algorithmic customization of the 'blank' face with the target face.



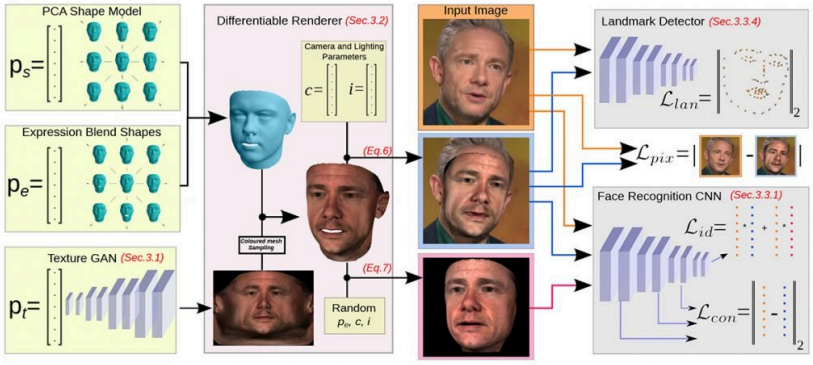
Another illustration from the 1999 paper, showing the processes involved in conforming a generic 3DMM template to a target identity.

Once some correlation is established, the revised 3DMM can be used as a method of projecting control points into the latent space, so that movements and changes in characteristics are (hopefully) reflected in the neural model.



From the 2019 GANFIT paper: deep fitting of 3DMM content to specific identities. Source: https://openaccess.thecvf.com/content_CVPR_2019/papers/Gecer_GANFIT_Generative_Adversarial_Network_Fitting_for_High_Fidelity_3D_Face_CVPR_2019_paper.pdf

The 2019 GANFIT approach, a collaboration led by Imperial College London, uses a GAN to train a UV texture generator in a neural space (two years later the method was adapted into Fast-GANFIT, an optimized version with lower latency).

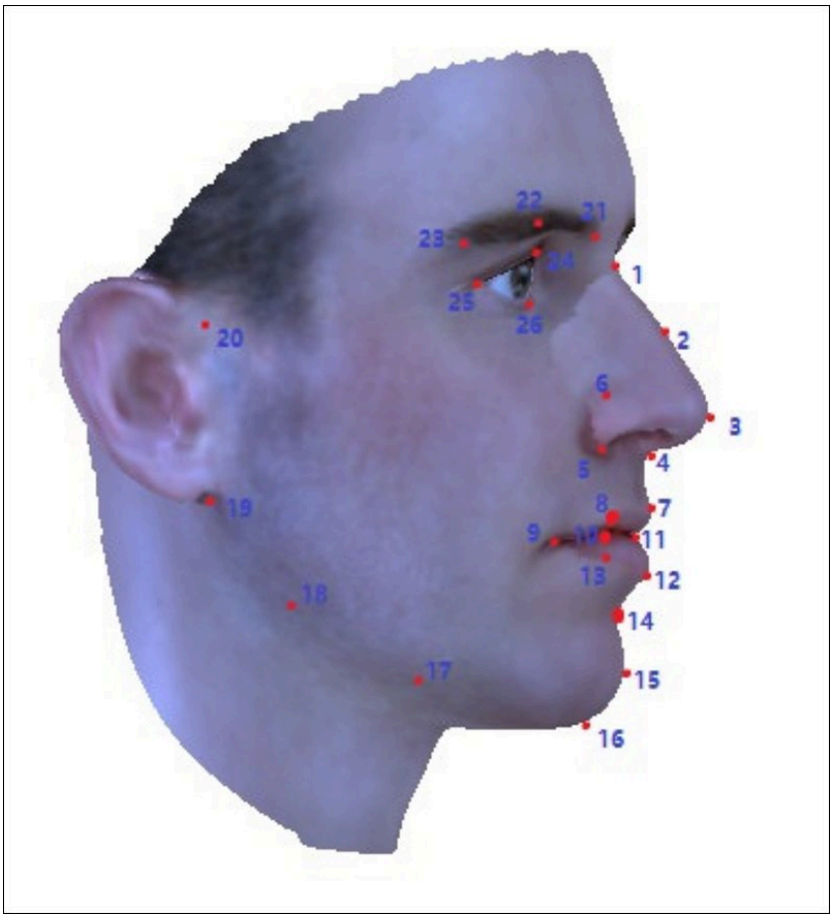


The GANFIT workflow, where UV coordinates become interrelated between the accessible 3DMM model and the more opaque facial representations stored in the latent space of the Generative Adversarial Network. The grey panels on the right indicate the usage of a pretrained face recognition network that's capable of identifying and assigning facial landmarks to the input material.

During the conforming process, the traditional UV coordinates from the model are turned into vectors (mathematical representations) and passed through Principal Component Analysis (PCA). Prior works, including works from the same group of researchers, have used diverse active Appearance Models (AAMs) to create this mapping, such as Scale Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG).

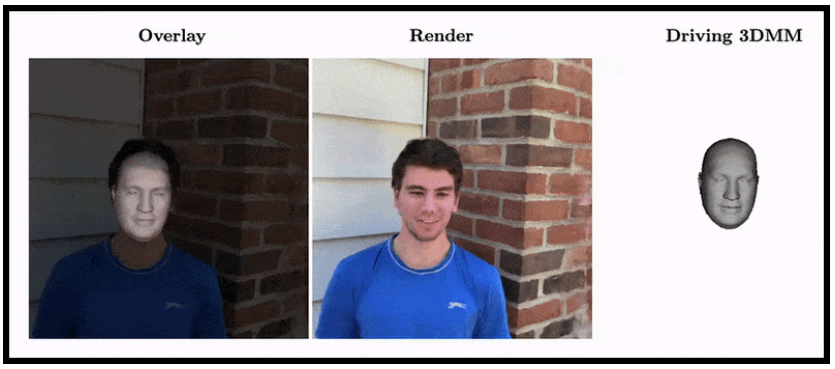
Taming the Latent Space with CGI

A 3DMM or other parametric model has a fixed, definable and controllable set of parameters, in stark contrast to the latent codes inside the latent space of a neural network, where the underlying semantic relationships are still being studied. Fixing these known coordinates to the approximate equivalent latent code enables a kind of crude puppeteering inside the latent space.



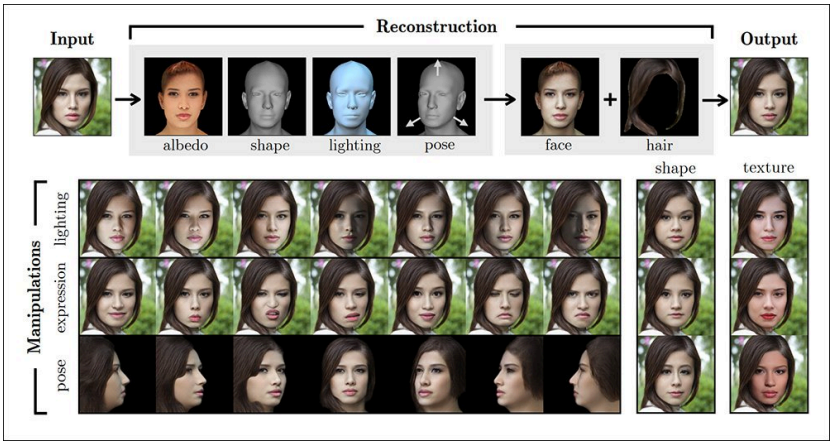
Fitting 3DMM landmarks to a neural space. Source: https://github.com/Yinghao-Li/3DMM-fitting

The features that are extracted during model training of more abstract neural networks such as latent diffusion and Generative Adversarial Networks (GANs) do not come with rational, voxel-style 3D coordinates. In the case of latent diffusion models, the visual aspects are deeply associated with descriptive text content, making using of the labels on the images that were trained, which adds an additional layer of complexity in 'targeting' a specific latent code.
Though Neural Radiance Fields (NeRF) *does* encode geometric data, it is not directly accessible in terms of explicit control over the geometry, which is learned from observing pixels from real and simulated viewpoints. Therefore, though 3DMM usage has been predominant in projects that seek control over a GAN's latent space, a growing number of initiatives are also leveraging parametric approaches as an intermediary for NeRF representations.



The RigNeRF project, from 2021, uses a 3DMM head to perform deepfake functionality in a NeRF environment, enabling a level of editability that is far from native to Neural Radiance Fields. Source: https://shahrukhathar.github.io/2022/06/06/RigNeRF.html

In the case of both NeRF and GAN, the research community turned to parametric methods only after a long and largely fruitless search for implicit methods of control that might be more native to the trained networks, and for less complex approaches to instrumentality. Ultimately the general consensus formed that these architectures are not easily susceptible to external control, and a certain initial resignation and disappointment about this fact has evolved into a more vigorous pursuit of superior 3DMM interfaces for otherwise 'closed' and paradoxical networks.
Most of the concrete examples of 3DMM use as a neural control system are in the GAN space, if only because GAN is among the older of the current crop of facial synthesis architectures, and the sector has been trying to make it more interpretable and governable for quite a long time.
In 2021 Mitsubishi released MOST-GAN, which uses non-linear 3DMMs as a facial control interface, offering a solid but far from definitive attempt at disentanglement (i.e., being able to edit a facet of a face without changing other aspects, currently an equally fervent pursuit in latent diffusion).



MOST-GAN had some success in differentiating and addressing different facets of latent codes trained into a GAN, though with some caveats and limitations. Source: https://arxiv.org/pdf/2111.01048.pdf

Mitsubishi Electric Research Labs (MERL)
2.14K subscribers
[AAAI 2022] MOST-GAN: 3D Morphable StyleGAN for Disentangled Face Image Manipulation
Mitsubishi Electric Research Labs (MERL)
If playback doesn't begin shortly, try restarting your device.

You're signed out

Videos you watch may be added to the TV's watch history and influence TV recommendations. To avoid this, cancel and sign in to YouTube on your computer.

More videos

Switch camera

Share

Include playlist

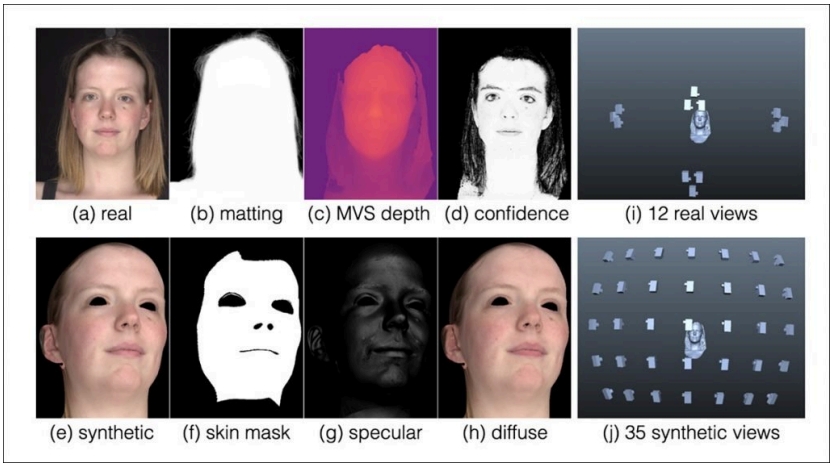
An error occurred while retrieving sharing information. Please try again later.

Watch on

0:00

0:00 / 1:11•

The use of CGI heads as a neural interface has been perhaps most extensively explored by Disney Research, notably in their 2022 offering *MoRF: Morphable Radiance Fields for Multiview Neural Head Modeling*.
Though the Disney Research paper covers strict 3DMMs only in its *Related Work* section, the methodology is very similar, with controllable and parametrized 3D rigs influencing trained facets in a neural network – this time a NeRF, or, as the paper names it, a Morphable Radiance Field (MoRF).



MORF projects have in-built alpha and depth channels, as well as albedo and specular layers, and multiple subject sources on which to draw. Source: https://studios.disneyresearch.com/app/uploads/2022/07/MoRF-Morphable-Radiance-Fields-for-Multiview-Neural-Head-Modeling.pdf

MORF uses a ‘deformation field’ to interact with the neural space, but a number of projects have used traditional 3DMMs to control the much older Signed Distance Function|Field (SDF) approach.

Beyond 3DMM

Now 24 years old, 3DMM is getting to be quite a venerable technology in computer vision, but has enjoyed a recent resurgence as an ‘off the shelf’ approach to facial movement synthesis in otherwise intractable architectures. There has, however, been a certain amount of innovation in 3DMMs themselves.
Besides Disney’s re-conceptualization of the role of CGI in facial synthesis (see above), one 2020 project, titled *i3DMM*, extended the capabilities of a traditional 3DMM model by adding full-head capture and extended features, including hair:

Christian Theobalt
3.18K subscribers
i3DMM: Deep Implicit 3D Morphable Model of Human Heads
Christian Theobalt
If playback doesn't begin shortly, try restarting your device.

You're signed out
Videos you watch may be added to the TV's watch history and influence TV recommendations. To avoid this, cancel and sign in to YouTube on your computer.

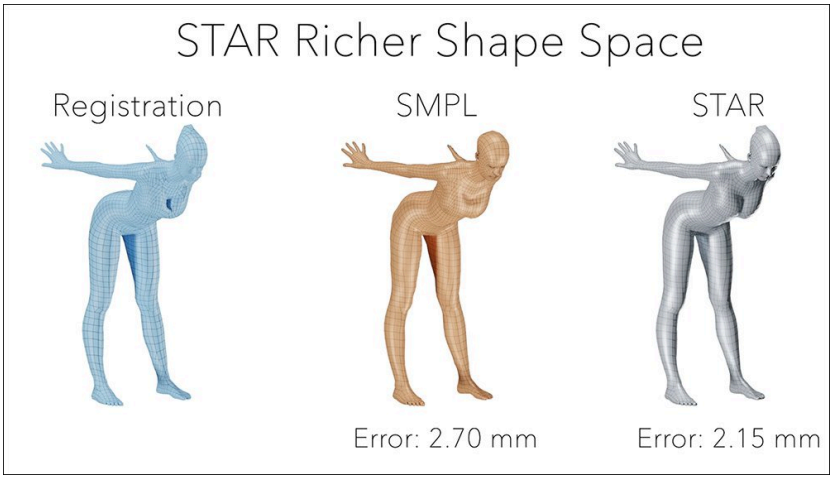
More videos

Switch camera
Share
Include playlist
An error occurred while retrieving sharing information. Please try again later.
Watch on

0:00

0:00 / 5:05•

As the notion of full body synthesis becomes more achievable and gains greater traction in the generative image synthesis field, the need for extended human representations has inspired the Max Planck Institute, the original creator of the 3DMM approach, to develop full-body parametric models such as the Sparse Trained Articulated Human Body Regressor (STAR) system.



The STAR system has a much higher number of parameters than SMPL, and is claimed to generalize better. Source: https://www.youtube.com/watch?v=JchovWRhrBs

However, the prior offering, Skinned Multi-Person Linear Model (SMPL), features more prominently in research papers, perhaps because of the wider body of literature concerning its use in neural synthesis, or because it has a more extensive like-for-like history in testing rounds.

3DMM in Latent Diffusion Models

3DMM has historically represented the 'last chance' to impose instrumentality and composability into neural systems that have been found, after much research, to lack such 'easy' mechanisms. By the time 3DMM interfaces are being investigated, practically every other potential 'native' method of achieving these results with less complicated approaches has been exhausted.

For latent diffusion models such as Stable Diffusion, at the time of writing, the research sector still holds out hope that semantic or other similarly 'in-built' approaches could transform diffusion systems from static image generators to semantically-complete 3D environments. However, a small number of projects are beginning to experiment with the 3DMM 'plan B' approach.

The DiffFace system from Vive Studios and Korea University, for instance, has stated that its diffusion-based face-swapping method is amenable to a 3DMM approach; however, 3DMM may eventually prove useful in diffusion models more as an additional or primary method of obtaining facial landmarks, for the growing number of current projects that are taking an interest in quantifying and governing diffusion-based facial content through semantic segmentation and facial analysis.