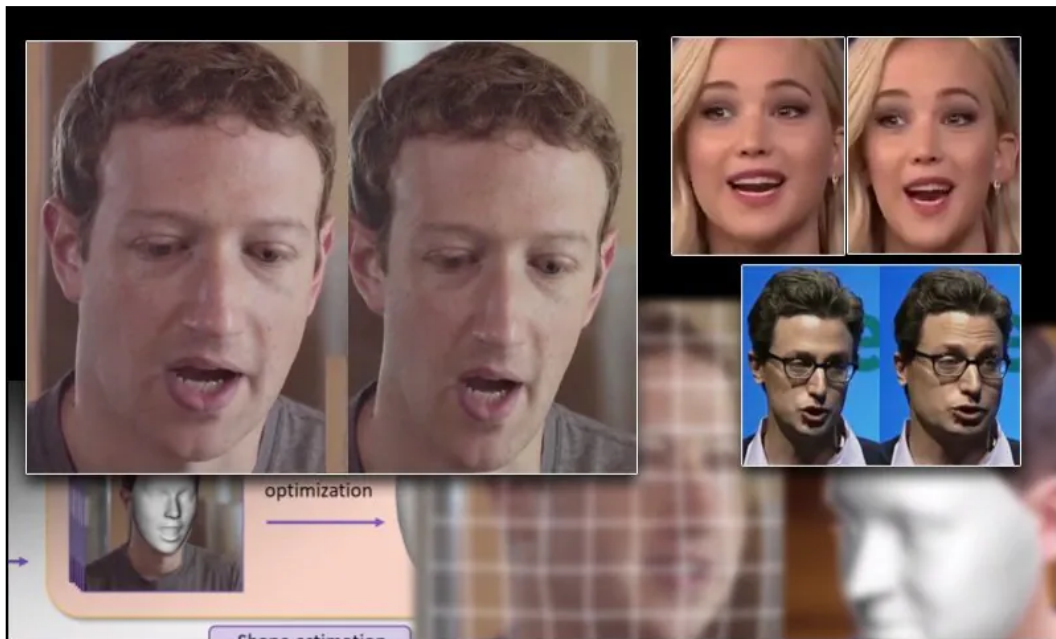


Restructuring Faces in Videos With Machine Learning

Originally published at <https://www.unite.ai/restructuring-faces-in-videos-with-machine-learning/>



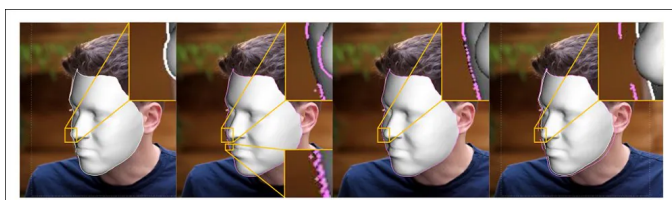
A research collaboration between China and the UK has devised a new method to reshape faces in video. The technique allows for convincing broadening and narrowing of facial structure, with high consistency and an absence of artifacts.



From a YouTube video used as source material by the researchers, actress Jennifer Lawrence appears as a more gaunt personality (right). See the accompanying video embedded at the bottom of the article for many more examples at better resolution. Source: <https://www.youtube.com/watch?v=tA2BxvrKvjE>

This kind of transformation is usually only possible through traditional CGI methods that would need to entirely recreate the face via detailed and expensive motion-capping, rigging and texturing procedures.

Instead, what CGI there is in the technique is integrated into a neural pipeline as parametric 3D face information that's subsequently used as a basis for a machine learning workflow.



Traditional parametric faces are increasingly being used as guidelines for transformative processes which use AI instead of CGI. Source: <https://arxiv.org/pdf/2205.02538.pdf>

The authors state:

'Our aim is to generate high-quality portrait video reshaping [results] by editing the overall shape of the portrait faces according to natural face deformation in real world. This can be used for applications such as shapely face generation for beatification, and face exaggeration for visual effects.'

Though 2D face-warping and distortion has been available to consumers since the advent of Photoshop (and has led to strange and often unacceptable [subcultures](#) around face distortion and body dysmorphia), it's a tough trick to pull off in video without using CGI.



CLICK TO VIEW ANIMATED GIF

Mark Zuckerberg's facial dimensions expanded and narrowed by the new Chinese/British technique.

Body reshaping is currently a field of [intense interest](#) in the computer vision sector, mainly due to its potential in fashion ecommerce, though making someone appear taller or skeletally diverse is currently a [notable challenge](#).

Likewise, changing the shape of a head in video footage in a consistent and convincing manner has been the subject of [prior work](#) from the new paper's researchers, though that implementation suffered from artifacts and other limitations. The new offering extends the capability of that prior research from static to video output.

The new system was trained on a desktop PC with an AMD Ryzen 9 3950X with 32GB of memory, and uses an optical flow algorithm from [OpenCV](#) for motion maps, smoothed by the [StructureFlow](#) framework; the Facial Alignment Network ([FAN](#)) component for landmark estimation, which is also used in the popular deepfakes packages; and the [Ceres Solver](#) to resolve optimization challenges.



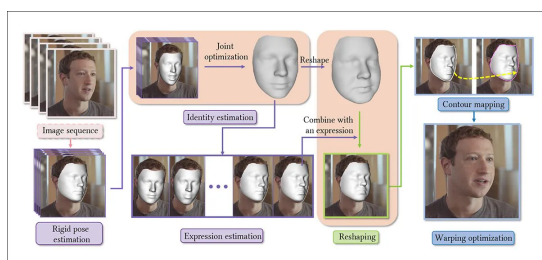
CLICK TO VIEW ANIMATED GIF

An extreme example of facial widening with the new system.

The [paper](#) is titled *Parametric Reshaping of Portraits in Videos*, and comes from three researchers at Zhejiang University, and one from the University of Bath.

About Face

Under the new system, the video is extracted out into an image sequence, and a rigid pose is first estimated for each face. Then a representative number of subsequent frames are jointly estimated to construct consistent identity parameters along the entire run of images (i.e. the frames of the video).



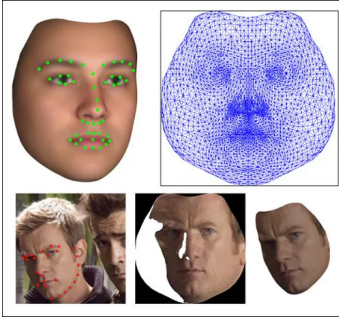
Architectural flow of the face warping system.

After this, the expression is evaluated, yielding a reshaping parameter that's implemented by linear regression. Next a novel signed distance function ([SDF](#)) approach constructs a dense 2D mapping of the facial lineaments prior to and after reshaping.

Finally, a content-aware warping optimization is performed on the output video.

Parametric Faces

The process makes use of a 3D Morphable Face Model (3DMM), an increasingly [popular adjunct](#) to neural and GAN-based face synthesis systems, as well as being [applicable](#) for deepfake detection systems.



Not from the new paper, but an example of a 3D Morphable face Model (3DMM) – a parametric prototype face used in the new project. Top left, landmark application on a 3DMM face. Top right, the 3D mesh vertices of an isomap. Bottom left shows landmark fitting; bottom-middle, an isomap of the extracted face texture; and bottom right, a resultant fitting and shape. Source: <http://www.ee.surrey.ac.uk/CVSSP/Publications/papers/Huber-VISAPP-2016.pdf>

The workflow of the new system must consider cases of occlusion, such as an instance where the subject looks away. This is one of the biggest challenges in deepfake software, since FAN landmarks have little capacity to account for these cases, and tend to erode in quality as the face averts or is occluded.

The new system is able to avoid this trap by defining a *contour energy* that's capable of matching the boundary between the 3D face (3DMM) and the 2D face (as defined by FAN landmarks).

Optimization

A useful deployment for such a system would be to implement real-time deformation, for instance in video-chat filters. The current framework does not enable this, and the computing resources necessary would make 'live' deformation a notable challenge.

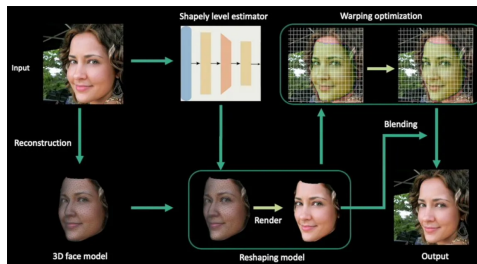
According to the paper, and assuming a 24fps video target, per-frame operations in the pipeline represent latency of 16.344 seconds for each second of footage, with additional one-time hits for identity estimation and 3D face deformation (321ms and 160ms, respectively).

Therefore optimization is key to making progress towards lowering latency. Since joint optimization across all frames would add severe overhead to the process, and init-style optimization (presuming on the consistent subsequent identity of the speaker from the first frame) could lead to anomalies, the authors have adopted a sparse schema to calculate the coefficients of frames sampled at practical intervals.

Joint optimization is then performed on this subset of frames, leading to a leaner process of reconstruction.

Face Warping

The warping technique used in the project is an adaptation of the authors' 2020 work [Deep Shapely Portraits](#) (DSP).

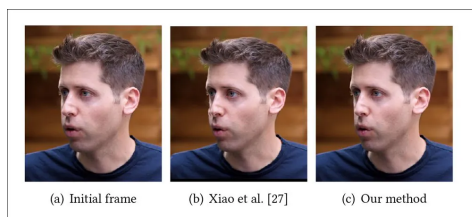


Deep Shapely Portraits, a 2020 submission to ACM Multimedia. The paper is led by researchers from the ZJU-Tencent Game and Intelligent Graphics Innovation Technology Joint Lab. Source: <http://www.cad.zju.edu.cn/home/jin/mm2020/demo.mp4>

The authors observe 'We extend this method from reshaping one monocular image to reshaping the whole image sequence.'

Tests

The paper observes that there was no comparable prior material against which to evaluate the new method. Therefore the authors compared frames of their warped video output against static DSP output.



Testing the new system against static images from *Deep Shapely Portraits*.

The authors note that artifacts result from the DSP method, due to its use of sparse mapping – a problem that the new framework solves with dense mapping. Additionally, video produced by DSP, the paper contends, [demonstrates](#) lack of smoothness and visual coherence.

The authors state:

'The results show that our approach can robustly produce coherent reshaped portrait videos while the image-based method can easily lead to noticeable flickering artifacts.'

Check out the accompanying video below, for more examples:

Parametric Reshaping of Portraits in Videos - ACM MM 2021



First published 9th May 2022. Amended 6pm EET, replaced 'field' with 'function' for SDF.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More