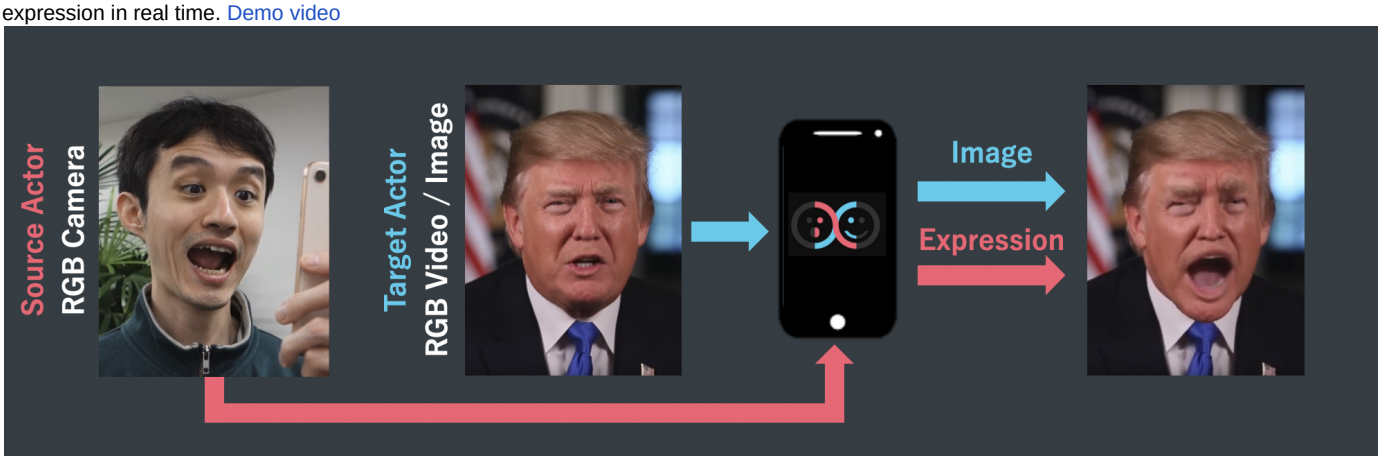
Xpression App Architecture

Xpression is a mobile application which transfers facial expressions in real time. It uses deep learning technology to control someone else's facial expression in real time. Demo video



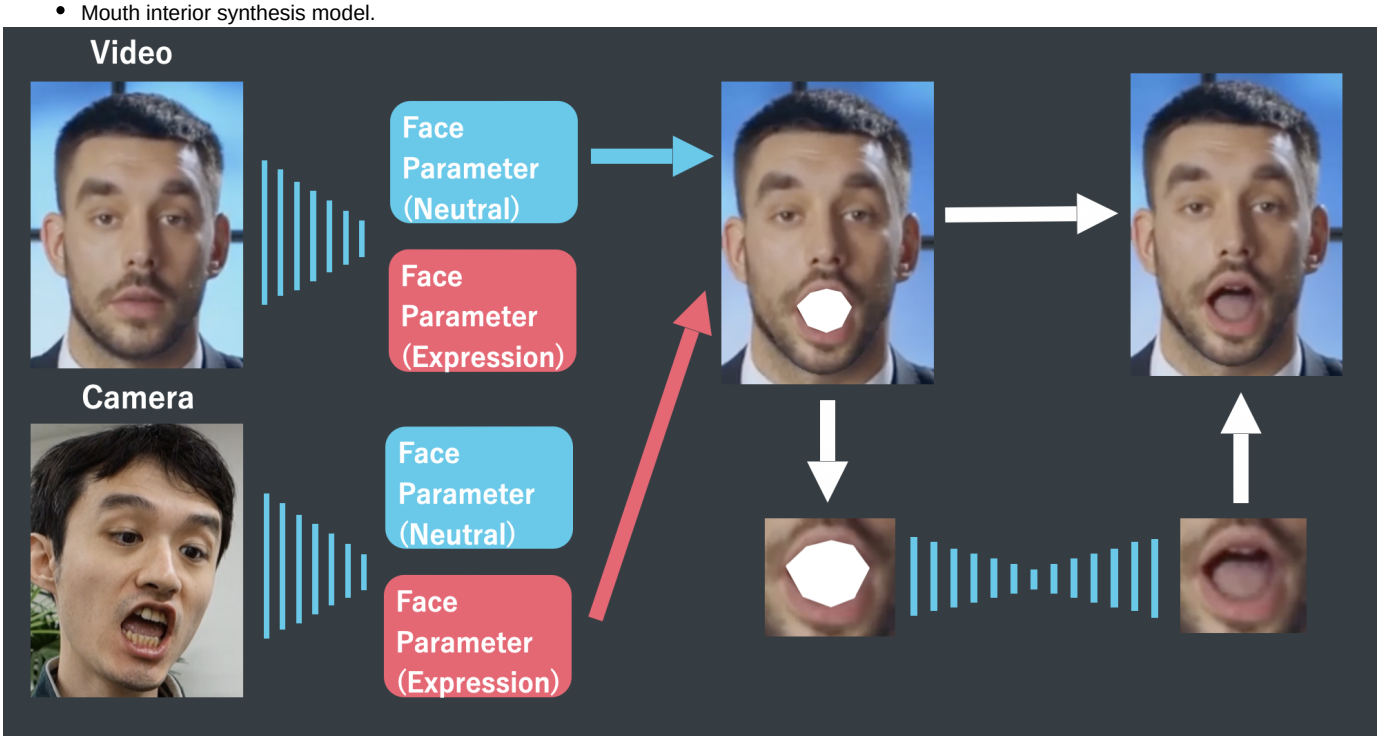
Unique Characteristics

- First real time facial reenactment system on mobile devices and low end PCs. **(Will update with link)**
- No preprocessing time. **(Will update with link)**
- Allow reenactment of still images and short videos. **(Will update with link)**

Brief Architecture

Xpression app works mainly on two deep learning architecture i.e.

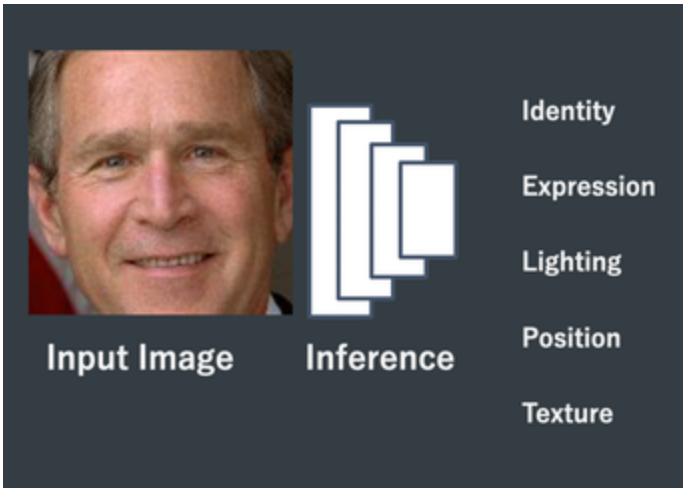
- Face parameter extraction model.
- Mouth interior synthesis model.



Face parameter extraction model

As the name suggest, this model extracts the facial parameters given an input (image or video). Here facial parameters are texture, lighting, position, identity and expression. These parameters are then used with PCA matrix to generate 3D face. **Working of this model will be updated later.** Fundamental idea behind the working can be found below:-

- 3D Morphable Models
- https://embodyme.atlassian.net/wiki/x/RYDgDQ

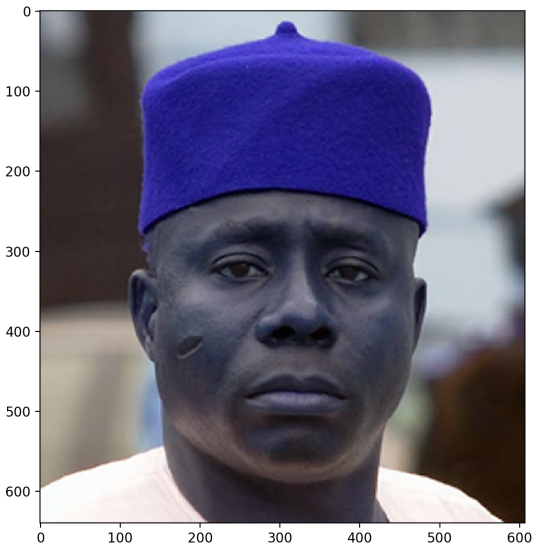


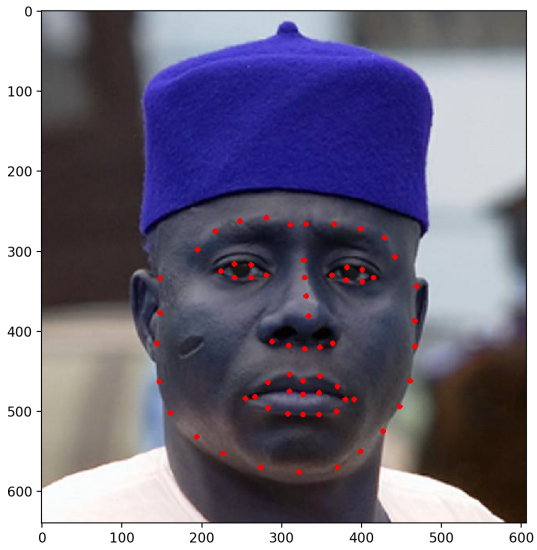
Code for training testing of **Face parameter extraction model** is below (requires permission from administrators to view):

https://github.com/embodyme/UnsupervisedDeepFaceReconstruction

Training data used for training and testing are shown below:-

For each facial image 68 points are being marked on it. Below is an example image and its corresponding data points marked.





Now that it is known what the “Face parameter extraction model” does in the architecture.

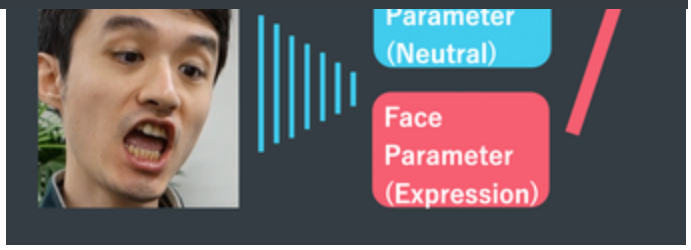
On the image beside, video is the target and camera is source. Both, the source video (or image) and target image (or video) is inferred and facial parameters for both neutral and expression is being extracted. Now the Face Parameter *Expression* (red arrow) from source image (or video) is being rendered with the Face Parameter *Neutral* (blue arrow) of target video (or image) to generate target video (or image) with source expression (Since we want to transfer expression from source to target).

Now after rendering expression parameters from source with neutral parameters from target we generate a new image which will possess the expression of source image on target video.

Mouth interior synthesis came into picture as the expression target image should be consistent with its neutral image. If the initial target image is neutral and final target image is expressive then the mouth of the expressive face should be consistent with the rest of the image.

Mouth Interior Synthesis

This part of model uses Generative Adversarial Network approach to generate the mouth interior region which is consistent with the target face. The generator network used it Mobile-Net based efficient network (similar to U-Net containing skip connections) and discriminator network is used to differentiate between original and generated dataset. First we extract the mouth region from Face parameter extraction model and from we use GAN based model to generate interior of mouth.



Code for training testing of **Mouth Interior Synthesis** is below (requires permission from administrators to view):

https://github.com/embodyme/deep-image-inpainting

Link to details about working of both the models are below

- “Face Parameter Extraction Model” https://embodyme.atlassian.net/wiki/x/JQKmEw
- “Mouth GAN” https://embodyme.atlassian.net/wiki/x/BwA8F