Non-linear 3D Face Geometry and Appearance Modelling

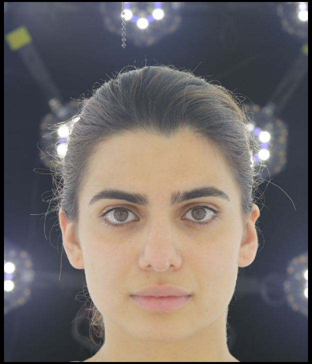
Vikas Thamizharasan
Programming Intern
CTX, Central Tech, Activision

<u>**TEAM**</u>

Mike Sanders (*Manager*)
Xianchun Wu
Sanjit Patel
Kayla O’Neill
Sean Comer

Motivation and Introduction

- Lightstage facilitates acquiring high quality face geometry, reflectance maps, skin microstructure.



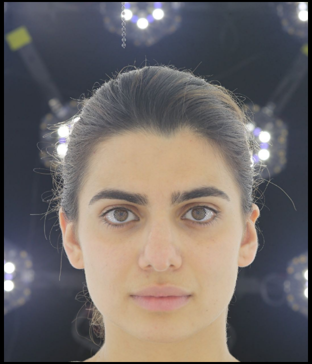
Lightstage capture
Source: lightstage.activision.com

Some of the acquired assets

Production ready

Motivation and Introduction

- Lightstage facilitates acquiring high quality face geometry, reflectance maps, skin microstructure.
- Is an expensive operation - cost of talent, human hours of artists and engineers.



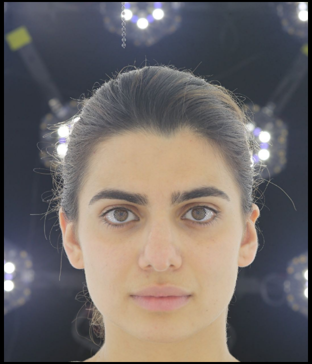
Lightstage capture
Source: lightstage.activision.com

Some of the acquired assets

Production ready

Motivation and Introduction

- Lightstage facilitates acquiring high quality face geometry, reflectance maps, skin microstructure.
- Is an expensive operation - cost of talent, human hours of artists and engineers.
- Has certain bottlenecks - rapid creation of game assets, impractical to acquire the entire human population.



Lightstage capture
Source: lightstage.activision.com

Some of the acquired assets

Production ready

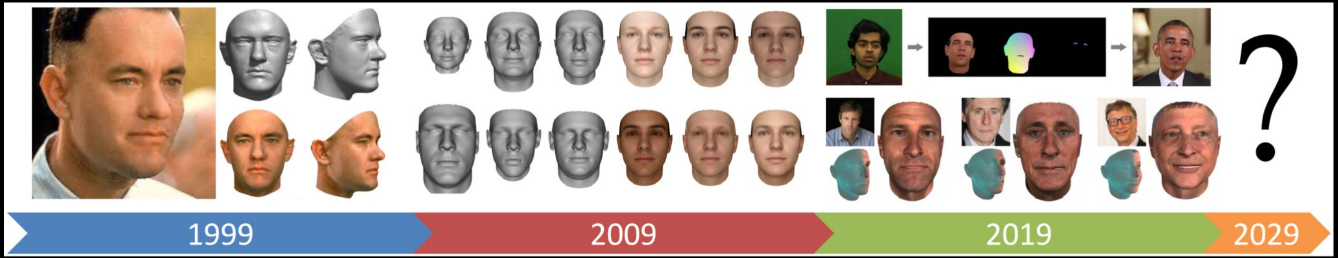
Motivation and Introduction

- The graphics community over the past two decades have parallelly focused on building statistical models.

Motivation and Introduction

- The graphics community over the past two decades have parallelly focused on building statistical models.

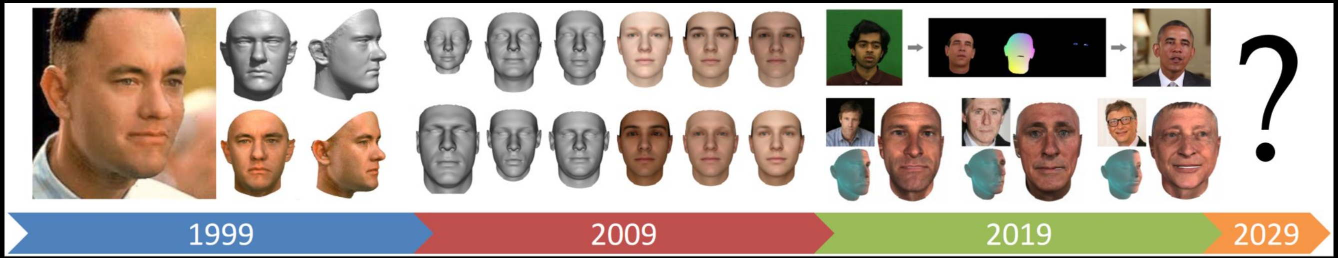
- Since seminal work of 3D morphable face model (3DMM) by Blanz and Vetter, SIGGRAPH 1999 there has been numerous attempts at building such statistical models. See "3D Morphable Face Models - Past, Present and Future" from SIGGRAPH 2021 for more details.



Source: B. Egger et al.

Motivation and Introduction

- The graphics community over the past two decades have parallelly focused on building statistical models.
- Since seminal work of 3D morphable face model (3DMM) by Blanz and Vetter, SIGGRAPH 1999 there has been numerous attempts at building such statistical models. See "3D Morphable Face Models - Past, Present and Future" from SIGGRAPH 2021 for more details.



Source: B. Egger et al.

- Computer Vision practitioners have also been using these models as geometric and appearance priors for various tasks like 3D reconstruction, face tracking, expression transfer i.e real-time puppeteering, 'deepfake' etc.

Motivation and Introduction

- Most of previous research focused on linear models. Limitations include:
  - Low geometry and texture resolution. Does not capture the high fidelity of the light stage scanned data.
  - Limited or no semantic control.
  - Geometry and texture are modelled separately leading to unrealistic faces.
  - Limited expressivity.

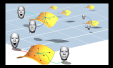
- We attempt to solve some of these challenges while keeping in mind how one would integrate them into a production pipeline and make their usage intuitive and artist friendly.

Motivation and Introduction

- Summer 2021 internship work: built a non-linear face geometry and appearance model.
- Inspired by recent research in deep 3D representation learning.
- Resulting in:
  - generative face model with disentangled identity and expression spaces by design,
  - ensures correlation between appearance and geometry,
  - captures high frequency texture details *(in progress)*,
  - guarantees artistic semantic control,
  - whole process is differentiable

Internship work : Problem formulation

Main inspiration have come from recent works of Chandran et al. from ETH/DisneyResearch and Li, Bladin and Zhao et al. from USC/ICT.



"Manifold of human faces" Jiang et al.

Process:

- ❏ Architecture: VAE consisting of residual blocks. Input: vertex displacements (this is our representation) and one-hot vector blendweights. Output: Deformed vertex displacements in accordance to the blendweights and per-vertex albedo.
- ❏ Training data: 3D face scans from the lightstage [*Fig. 1*] - all processed to the same normalized canonical space and 179 isolated expression basis.
- ❏ Objective function: the L1 term to minimize reconstruction error and KL divergence term to constrain the latent space.

Fig. 1: Captured data

Internship work : Disentangling identity and expression

Results of random geometry only face synthesis:

- At inference time, we sample the identity and expression latent codes from a unit normal distribution.
- The output vertex displacements are added to the mean face to get the final mesh.

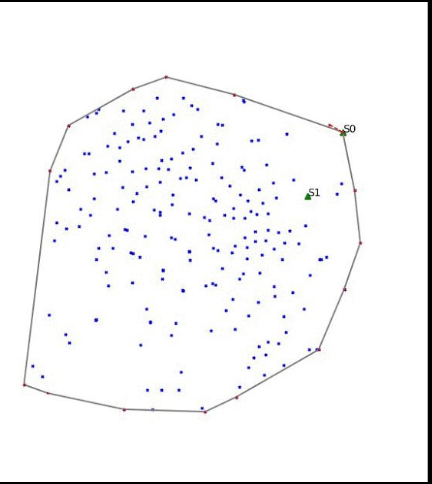
Observations: (i) Synthesised faces are realistic and diverse, (ii) Deformations due to expressions preserve identity.

Disentangling identity and expression : exploring the latent space

Interpolating in the identity latent space

t-SNE (projected to 2D)

Interpolated faces



- ● : 200+ faces in the training set.
- ▲ : S_i is a randomly selected subject in the training set used as a marker for the interpolation.
- ■ : I_j is an interpolated face between S_i and S_{i+1}.

[Random results]

Disentangling identity and expression : generalizing to dynamic expressions

Interpolating in the expression latent space: despite the model only being trained on static expression shapes our expression space generalizes to dynamic expression deformations.

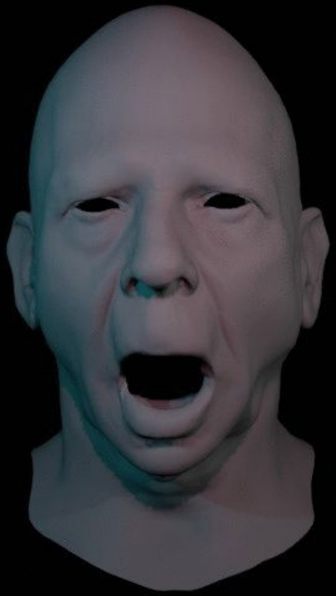
Note: this is a random face generated and has no rigs. All of these deformations are occurring directly on the mesh vertices.

Random face (F) via sampling

Linearly interpolating in exp. latent space
Front View

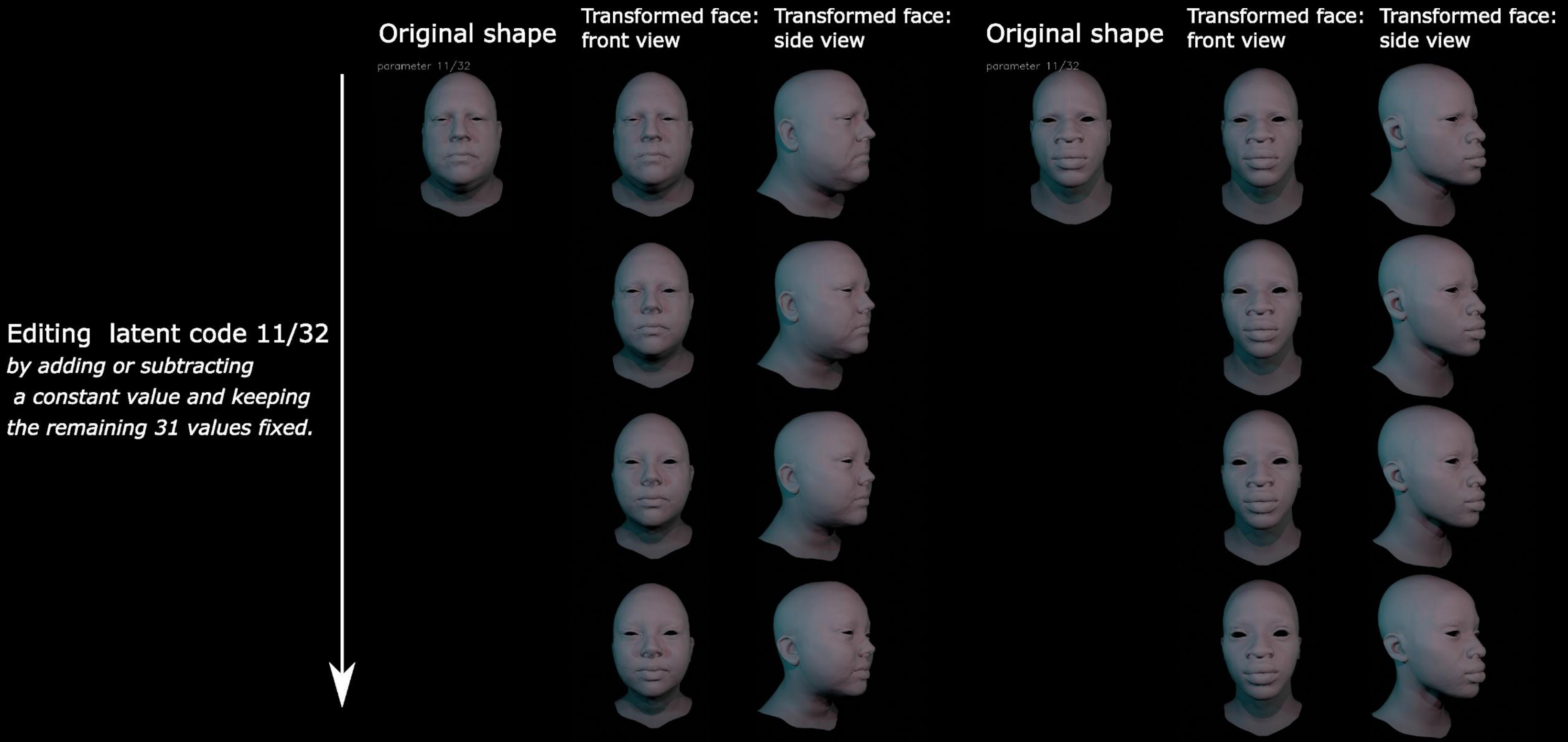
Side view

[Random results]



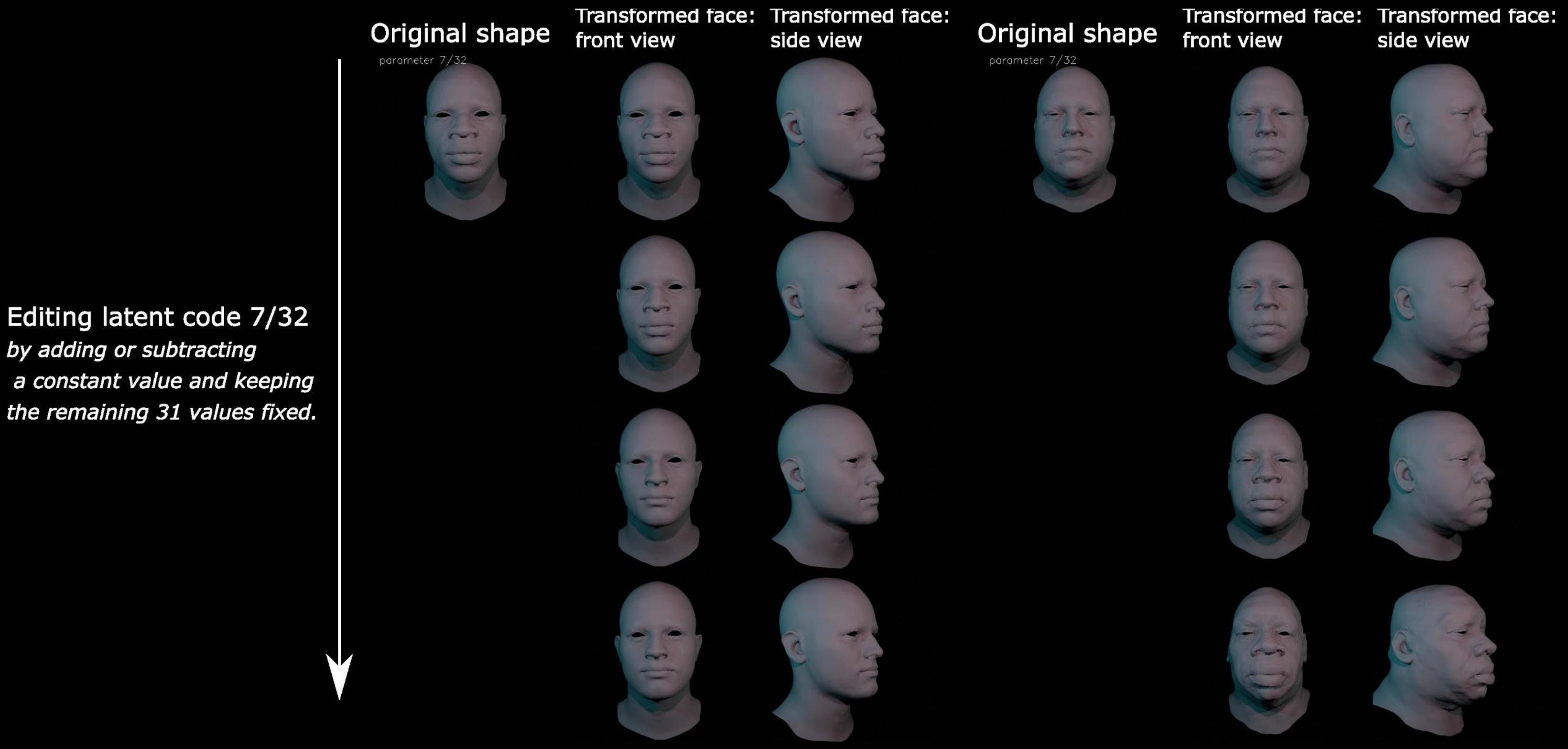
Disentangling identity and expression : editing latent codes [1/3]

Previously we showed you what happens if you generate a latent code via random sampling. Here we instead encode a latent code from a given face and then individually edit each variable...



Disentangling identity and expression : editing latent codes [2/3]

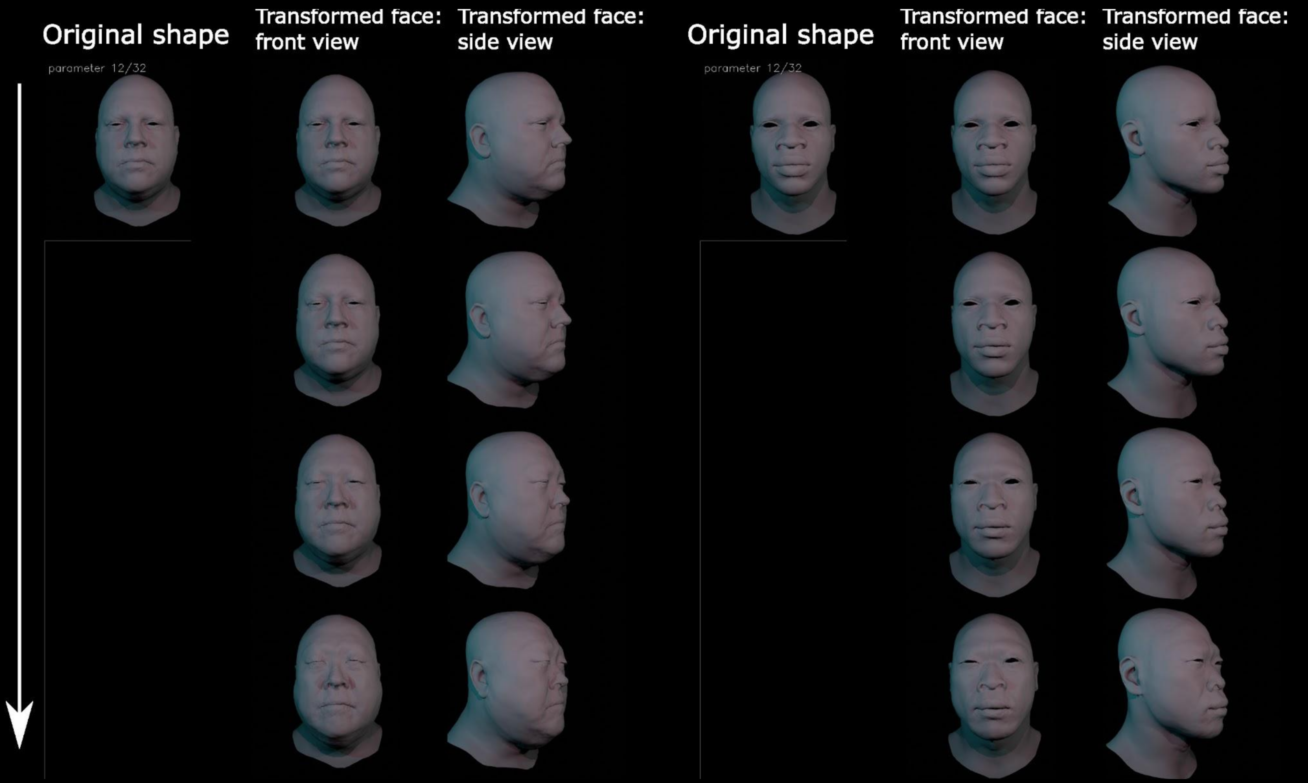
...The results show some fascinating edits that can be performed...



Disentangling identity and expression : editing latent codes [3/3]

Note: We neither provide nor does our model know any notion of the subjects ethnicity, age or gender.

Editing latent code 12/32
by adding or subtracting a constant value and keeping the remaining 31 values fixed.



Internship work : Jointly solving geometry and appearance

Now that you've seen the capabilities of the geometry model, here is the corresponding per-vertex albedo predicted via a separate decoder branch.

These are the 21k per-vertex albedo colours

Internship work : super-resolution

Because of our choice of representation, we run into a limitation where our uv mapped per-vertex albedos are of low resolution. What we have is **A** but we'd ideally like to have **B**.

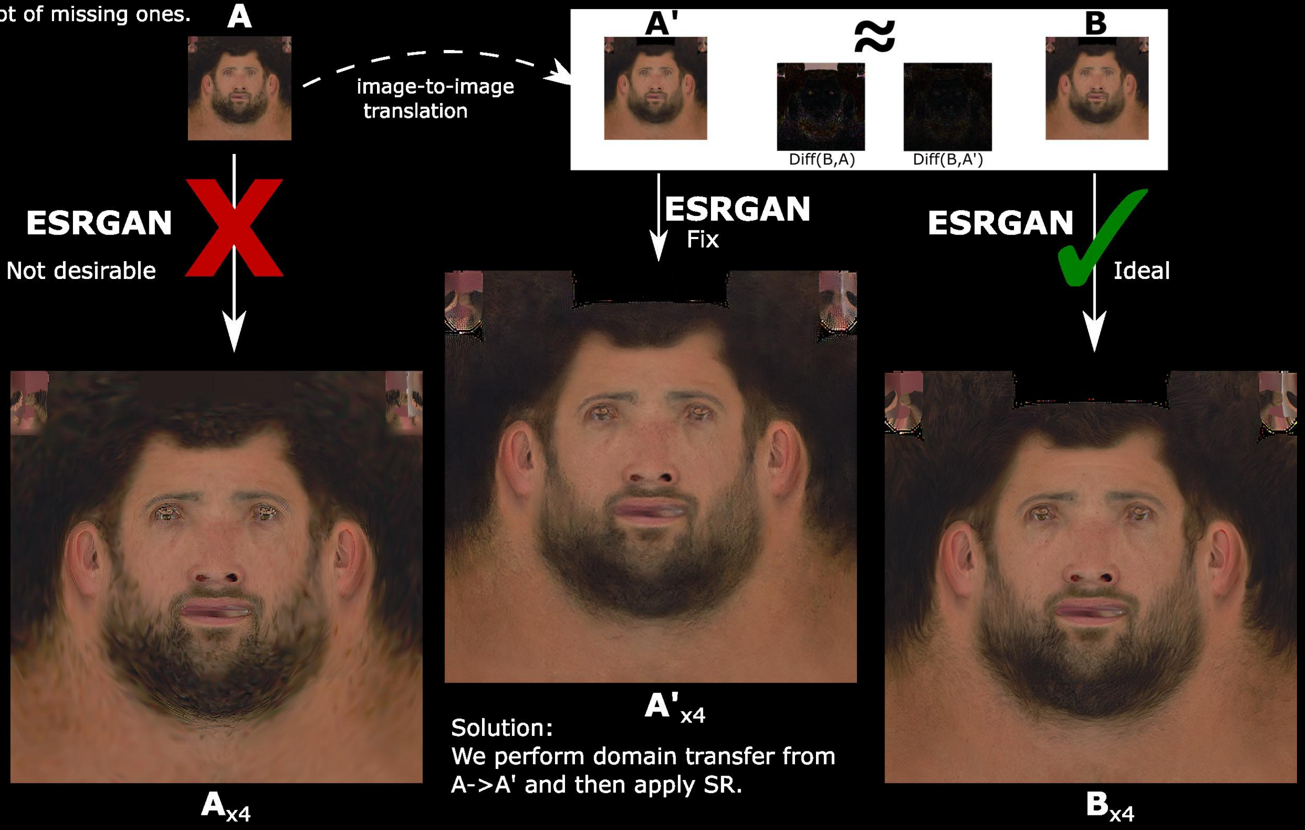
A B Diff(B, A)

One might be tempted to use some off the shelf super-resolution (SR) software but doing so leads to unpleasant artifacts. We are missing high frequency information, especially at regions with facial hair. The solution is to reuse the adversarial loss present in SR techniques like ESRGAN to force hallucination of the desired high frequency details.

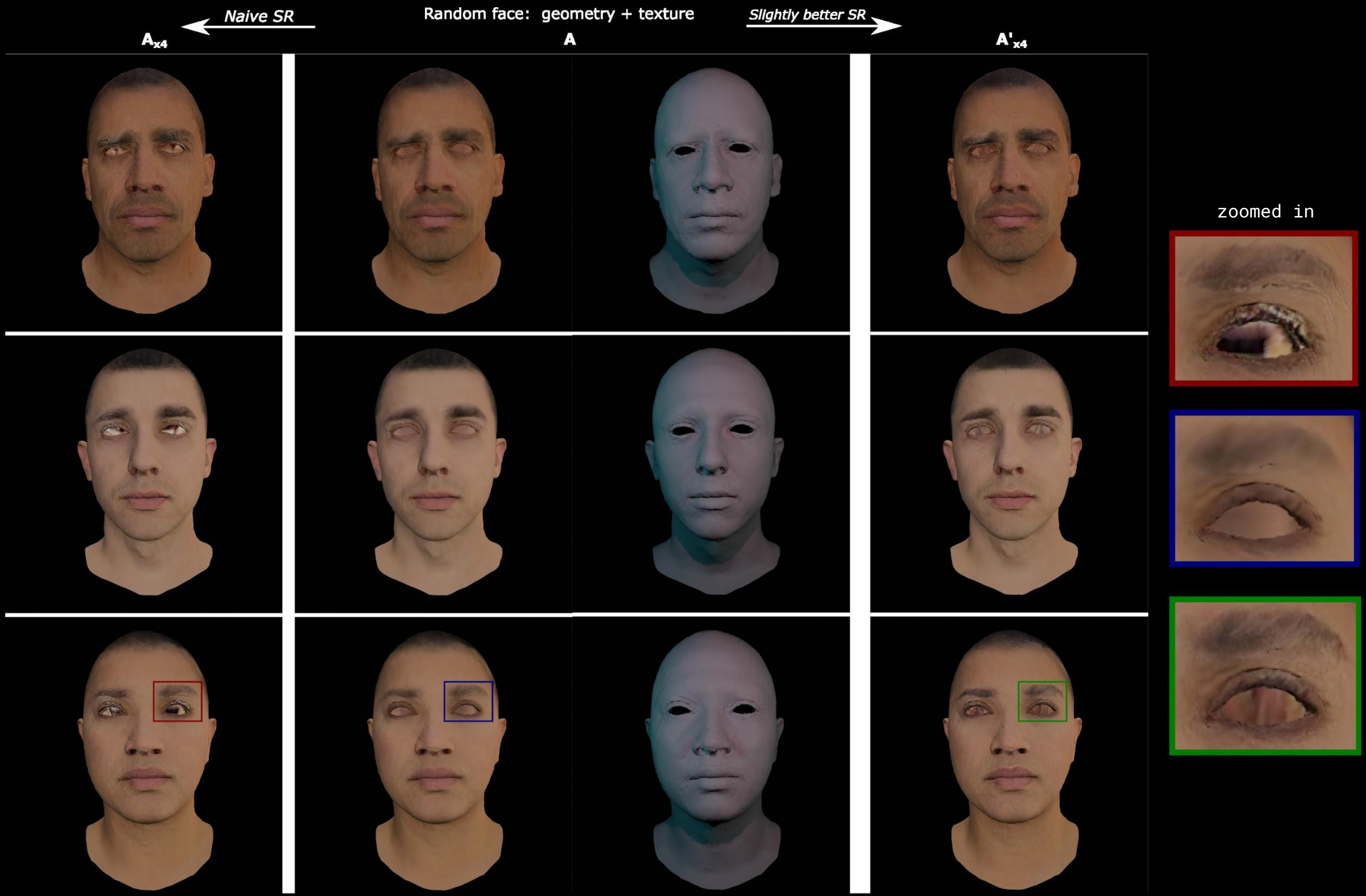
Super-resolution

Expected output of our non-linear geometry and texture model.
Contains some high frequency details but a lot of missing ones.

To use ESRGAN directly, our LR image should belong to the domain of LR images generated as a result of downscaling using a bicubic kernel function.



Super-resolution



Non-linear 3D Face Geometry and Appearance Modelling, Vikas Thamizharasan et al.

[Random results]

[Random results]

Implementation details

Tools and Languages: PyTorch, Numpy, C++, CUDA, bpy (blender python api)

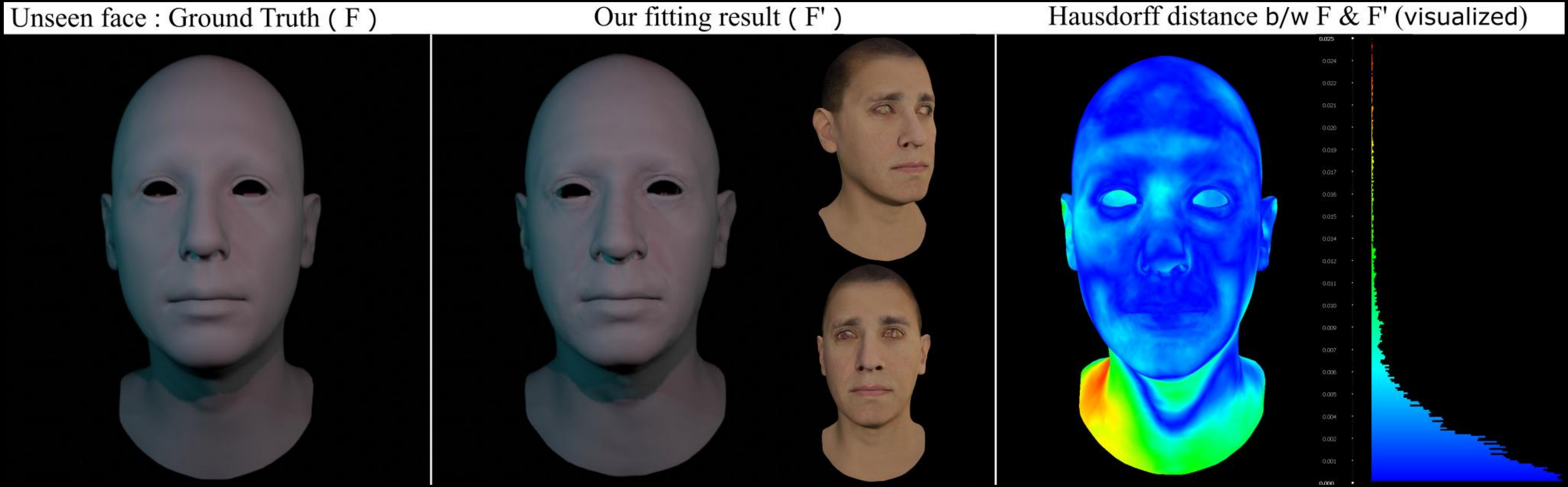
Hardware: NVIDIA GeForce GTX 1070

Execution time: 80ms to synthesize a single face, thus suitable for various interactive real time applications.

Applications

What are the other applications besides random face generation and semantic editing?

- 3D Fitting : Fitting to unseen faces - user provides new geometry, can we use our parametric model to find a suitable fit?



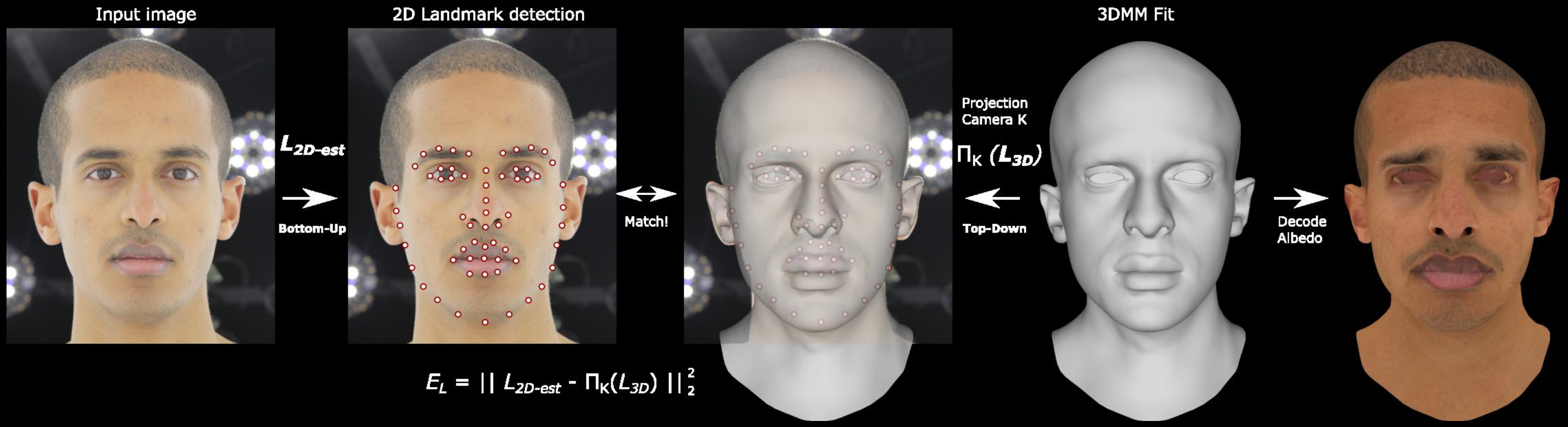
Units: cm

Non-linear 3D Face Geometry and Appearance Modelling, Vikas Thamizharasan et al.

Applications

What are the other applications besides random face generation and semantic editing?

- 2D Fitting : 3D reconstruction from single image. Fitting to 2D landmarks via a classical optimization method by minimizing the projection loss.



Applications

What are the other applications besides random face generation and semantic editing?

- 2D landmark based face tracking a.k.a real time puppeteering:

  Achieved by translating image based landmarks to expression blendweights.

Applications

More random faces synthesized by our system :)

Future

As the creation of high quality digital assets become a staple in modern AAA games, faster turnaround times for the creation of these assets become critical. This requires designing algorithms that can model and replicate objects of the complex world we live in. The next steps from here would be to make these tools accessible to the end users in some form. If these tools can be democratized, games in the future could allow users to captures images of cities / landscapes of their choice and have them transformed to instantly usable game maps, create a digital double from just a selfie and much more. The popularity of battle royal style games provide a perfect testing ground for players to interact with these systems.

"What would I do if I had 3 more months of my internship?"

- Work on improving the texture model - currently the results don't have the desired fidelity.
- Focus on the applications - I believe this is crucial to justify the work done so far and it's integration into the current production pipeline.
- Build a standalone app with a GUI or a Maya plugin to enable real-time interaction.

<- Me

MS in CS @ Brown University | Graduating: December 2021

Thank you!

Please reach out to us if you have any questions or thoughts about our work.

Acknowledgement

I would like to thank my team for giving me the freedom to work on a project of my interest and the folks at IW for sharing their data.