*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 10 Issue 1 January 2022

Unsupervised Machine Learning based rapid Pose-Invariant Three-Dimensional Facial Reconstruction for Biomedical Extended Reality and Phantom Generation in Oncologic Cranio-Maxillofacial Surgery

Author: Abhishek Pillai

Affiliation: PACE Junior Science College, Andheri, Mumbai

E-mail: abhishekpillai94@gmail.com

DOI: 10.26821/IJSHRE.10.1.2022.91101

ABSTRACT

Cranio-maxillofacial surgery is a surgical specialty that focuses on the reconstructive surgery of the entire cranio-maxillofacial complex: the anatomical area of the mouth, jaws, face, and skull, head and neck as well as associated structures. A highly precise and complex procedure, it demands a well-structured and comprehensive treatment planning process. Biomedical extended reality (XR) has proven to be beneficial in the pre-surgery visualization step as it enables interaction in the three-dimensional (3D) space relative to the patient, however the tedious nature of the construction of unique, patient-specific facial models makes it difficult to implement XR and commercial physical patient phantoms are far too generic and expensive for the task. In the case of cancerous tumors on the head and face, the numbers of which have increased tremendously over the past few years, a suitable oncologic therapy needs to be delivered in the shortest possible period of time, as they spread rapidly and the possible complications can be fatal to patients. While PET/CT scans do help in accurately visualizing such tumors, and are the universally-accepted approach in surgically examining cases of craniofacial trauma and oral cancer, they are time-consuming and only assist in two-dimensional (2D) visualization. They have to be further mapped by a surgical or radiological expert to the 3D anatomy of the patient, a task that can be eliminated in both oncologic cranio-maxillofacial surgery and cosmetic surgery.

In this study, we explore near-instantaneous external 3D facial reconstruction from landmarks obtained specifically from a single 2D image of the human face, which is unchallenging to sample, making the system deployable as an application even on mobile devices. We also aim to eliminate the limitations encountered due to pose variations, through the frontalization of these images, while maintaining focus on the affected area, with the implementation of an illumination-preserving Generative Adversarial Network (GAN), integrated with a light Convolutional Neural Network (CNN) for feature extraction, trained and tested on the Multi-PIE and LFW (Labelled Faces In-the-Wild) datasets. The output is then processed to obtain 2D landmarks (disregarding external features) that are crucial to project the image on a 3D morphable model (3DMM). The final output is a digital, printable 3D mesh in both the OBJ and STL format, which can be used for generating custom patient phantoms as well as be implemented in biomedical XR for pre-surgical planning and evaluation. It can also be integrated with a CT scan for internal skeletal 3D imaging in other cranio-maxillofacial cases.

We ultimately evaluate the performance of the system on images captured from several angles, to show its generalization and robustness to novel facial configurations and unseen data.

Keywords: **Cranio-maxillofacial surgery, biomedical extended reality (XR), custom patient phantoms, cancerous facial tumors, PET/CT scans, cosmetic surgery, three-dimensional reconstruction, 2D**

images, 3D meshes, pose invariant face frontalization, Generative Adversarial Network (GAN), Convolutional Neural Network (CNN), morphable 3D model, 2D facial landmarks, printable model

I. INTRODUCTION

1. *Cranio-Maxillofacial surgery as a reconstructive treatment*

Cranio-maxillofacial surgery is the recommended course of treatment for cases involving cosmetic facial surgery, cranio-maxillofacial trauma and penetrating injuries, pediatric maxillofacial surgery (including cleft -lip, Fronto-Orbital Advancement and Remodeling (FOAR) [1], and total vault remodeling), and maxillofacial regeneration, for the reformation of the facial region. This surgical procedure is also of great significance in the management of cancerous cranio-maxillofacial tumors [2] after diagnosis, the goal of which is to, most commonly, entirely remove the cancerous cells (most often keratinocyte carcinomas (non-melanoma cancer) removed through Mohs surgery). This invasive surgical specialty requires extensive planning and analysis of patient-specific anatomy, to determine the approach to be taken.

Fig.[A]: Face Reconstruction after Mohs Surgery

Biomedical XR is an excellent aid to pre-surgery visualization, specifically in cranio-maxillofacial surgery [3], as indicated by recent studies in the field [4][5][6], as it allows accurate analysis of that particular case, relative to the 3D space of the patient, instead of on a separate workspace (Badiali, G. et al. (2019) [7], Landaeta-Quinones, C. G., Hernandez, N. & Zarroug, N. K (2018) [8]). This relatively new approach supports the clinical diagnostic and oncologic treatment process by allows 3D analysis and visualization of the tumor and nearby anatomical structures, while examining medical data simultaneously (Figure B). However, the approach requires unique, patient-specific 3D models, that need to be accurately generated for its implementation. This is a time-consuming process, that requires frequent interaction with the patient and affected areas, which is neither desirable for the patient nor clinically viable (especially during a pandemic). The use of commercial phantoms (Figure C), is unsuitable, as they are too expensive and not case specific (which is required by precise surgical procedures). A less arduous alternative -PET/CT (Positron Emission Tomography/ Computed Tomography) scans- can also be used [9][10][11], and are currently employed for visualizing skeletal damage and hidden tumors, but the 2D imaging requires further 3D modelling (Figure D), not optimal for external cases involving facial skin cancer.

Fig.[B]

Fig.[C]

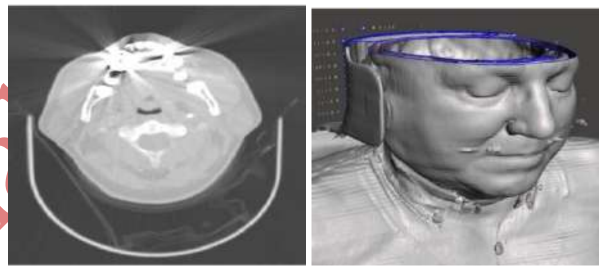


Fig. [D]

Our method offers a solution to three-dimensionally visualize external facial structures with an asymmetric technique, which can also be integrated with a more accurate scan for internal imaging and skeletal reconstruction purposes. It also generates a flexible, printable 3D model [12] that can be used as a physical phantom [13] for interactive experimentation of various invasive surgical procedures.

2. *Previous and Related work in 3D Reconstruction*

The construction of a 3D facial model allows transformation into a known view for the purpose of facial analysis, and improves the extraction of depth and texture information in the alignment of features, informative during the XR examination [14]. It enables a complete understanding of the surrounding area of skin and external skeletal structure [15], ideal for pre-surgery visualization relative and unique to a patient. For the specific purpose of 3D facial reconstruction, several algorithms have been presented in the past. Berar et al. (2006) [16] focused on 3D face reconstruction with a cranio-facial approach, through statistical modelling with a collection of CT scans. Elyan and Ugail (2007) [17], proposed a method for the construction of the 3D geometry of human faces based on the use of Elliptic Partial Differential Equations (PDEs). Here the geometry corresponding to a human face is treated as a set of surface patches, whereby each surface patch is represented using four boundary curves in the 3-space that formulate the appropriate boundary conditions for the chosen PDE. These methods primarily dealt with the combination of extracted 3D scans. Other methods have utilized a single or several 2D images for the process of reconstruction with a geometric approach. Wang et al. (2005) [18] collected several 2D images, captured from different angles, and combined them into a 3D facial model by detecting corresponding points using silhouettes. Isaac Rudomin and Hector Cuevas (2000) [19] applied Principal Component Analysis (PCA) to create a set of related 2D and 3D eigenfaces, and Y. Guan (2007) [20] utilized affine transformation and a 3D statistical face model in the reconstruction process, with the input being only a single 2D image in both cases. More recent methods have utilized autonomous supervised or weakly-supervised deep learning models (including regression-based networks [21][22]), integrated with a 3DMM for facial reconstruction, showcasing promising results (Yu Deng et al- 2019 [23] (leveraged a robust, hybrid loss function for weakly-supervised learning which takes into account both low-level and perception-level information for supervision, and performed multi-image face reconstruction by exploiting complementary information from different images for shape aggregation), Dou et al.- 2019 [24](utilized a DNN (Deep Neural Network) to extract and refine the 3D face in an iterative manner using both an RGB image and an initial 3D facial shape rendering, integrated with a multi-task loss function and a fusion convolutional neural network (CNN) to improve facial expression reconstruction)).

The proposed method takes an unsupervised approach to frontalization, which is desired by the reconstruction task for oncologic or unspecialized cranio-maxillofacial pre-surgery visualization (cases may involve protruding tumors or external growths), and adopts a geometric projection technique, which does not rely on the symmetry of the face, but learns patterns and structures on its own.

3. *Approach and Objectives*

Generative Adversarial Networks (GANs) present a unique solution to this extensively researched reconstruction problem. Generative modelling is an unsupervised machine learning task that involves automatically discovering and learning patterns and regularities in the data to generate output that could have plausibly been drawn from the original dataset. It retains the original characteristics of the input, which is extremely useful in this particular case, that only requires an additional frontalization, while maintaining the initial alignment features, that contribute to depth assessment. While this deep learning method does require significantly more data to train, it is efficient and computationally simpler as it relies on processing primarily 2D images, rather than the 3D space, while working on new inputs, and handles inconsistency in illumination and pose in a similar manner during data generation and discrimination.

The core aims of this study are (1) analysis of the performance of a generative adversarial network on the task of facial frontalization of a 2D image, with feature retention, to generate input suitable for reconstruction, (2) determination of 2D landmarks, to be projected into the 3D space, on a morphable model, with a CNN, (3) generation of a texture map from the processed images, for the enhancement of geometric surgical perception and oncologic analysis, (4) projection and creation of a 3D mesh from the extracted landmark nodes, (5) assessment of the accuracy of the autonomous system and the obtained result, by transforming the morphable model into a known view and comparing it with the original input structure, for determining its use in XR and understanding the potential of the system in both unseen, in-the-wild and cranio-maxillofacial surgical cases.

II. WORKING PRINCIPLE

The proposed solution includes a generative machine learning model that is executed by providing a single 2D facial image as input. The image is tested by the trained GAN model that retains the principal geometric characteristics of the image (cranio-maxillofacial complex and affected areas) and focuses on frontalizing the detected face (features retained and extracted by a light CNN). The image, still in the 2D space, is then processed by a Multi-Task Cascaded Convolutional Neural Network (MTCNN) [28] (or a regression-tree based detector (DLib) depending on feature location), to detect and extract key facial landmarks, specifically from the affected and following areas: the eyes, the nose, mouth, and jaw, that aid in maintaining relative alignment during the reconstruction process.

The location of these landmarks is identified on the images and projected into the 3D space, onto a morphable model, for the generation of a mesh. The GAN processed image is retained and used in addition to the original image, to create a texture map. The texture is then layered on the mesh by locating the corresponding landmark projection, and thus ultimately generating a morphable, highly accurate 3D model (in both the OBJ and STL formats), with the only data provided being a single 2D image. This outputted 3DMM can be converted to a known view with little computational complexity in implementation. The model, with the addition of the textured mesh, retains the geometric characteristics of the patient's face and clearly identifies the affected areas by retaining the original illumination setup, important for pre-surgery examination. This base structure can also be coupled with internal 3D imaging to produce a well-scaled phantom or 3D model for craniofacial surgery and therapy. Finally, the modified 3DMM can be implemented in biomedical XR for patient-specific examination or be 3D printed for pre-surgical, physical phantom interaction, ultimately delivering an assured and highly accurate treatment, that benefits the patient.

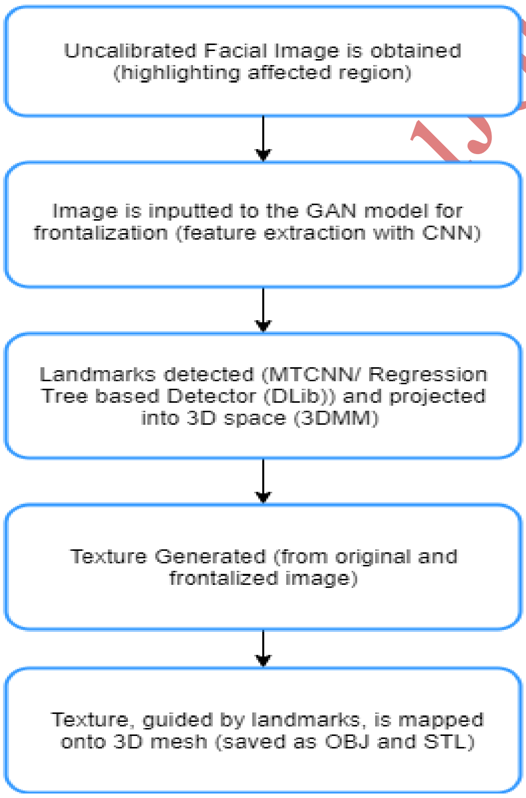


Fig.[E]: Workflow of Architecture

III. DETAILED OVERVIEW

A. Generative modelling for face frontalization (unsupervised-learning based architecture)

1. *Intuition and architecture*

Face frontalization is the process of synthesizing the frontal view of a face, from the given profile (or angular view). This additional step reduces the computational complexity of the overall procedure, as a frontalized face view can be used directly for specific reconstruction methods (or can be further processed, as in this particular case) without integrating any other complex modules.

Fig.[F]: Frontalization of the human face

Traditional methods address this problem through 2D/3D local texture warping and statistical modelling. More recent approaches implement a cascade-regression based frontalization technique. Over the past few years, extensive research has shown the ability of deep learning methods, making them highly suited for this task. This is why we rely on GAN architecture to synthesize the frontal perspective of the face, detected in the 2D image provided. Moreover, the utilization of GAN architecture adds the advantage of an unsupervised learning mechanism, for it assesses and retains features that may be ignored by supervised or weakly-supervised mechanisms (ideal for oncologic and other cranio-maxillofacial cases).

2. *Procedure and Machine-Learning Model*

This initial step is initiated by passing a single, uncalibrated 2D image of the patient's face (including and highlighting affected areas) to the system, that feeds it to the trained and tested generative adversarial network, with a Face Verification Accuracy (ACC) (positive identification) of 90% on the LFW dataset

[30] (5750 test images). For the task of 3D cranio-maxillofacial reconstruction and precise texture generation, which are the primary outcomes of this study, the GAN must produce photo-realistic and illumination preserving frontal images with high accuracy, thus explaining the need for an extensive collection of image data. Oncologic or trauma affected cranio-maxillofacial data is not readily accessible, so we rely on a generalized face dataset to preserve other features, and customize the mapped texture and landmarks with the patient-specific cancer affected areas for clinical identification and surgical analysis. In this study, we have created a Pytorch based Generative Adversarial Network, integrated with a light CNN, fine-tuned on the Multi-PIE dataset [31], for enhanced feature extraction. In addition to this, we have utilized several libraries, including NumPy, OpenCV, and Torch vision for the frontalization task (SciKit-Learn to determine the accuracy of the translation). Several GANs have been introduced in the past for the specific purpose of frontalization and feature translation. After exploring several models and comparing their performance on the synthesis task, a customized FFWM architecture was chosen for this procedure. It out-performs several other highly accurate models, including TP-GAN [32], DR-GAN [33], and DA-GAN [34] , being proposed as recently as in the year 2020 (Yuxiang Wei et al), with a state-of-the-art accuracy of 99.65% on the LFW testing dataset (as mentioned in the original paper) (Figure G).

Method	FaceNet	VGGFace	FF-GAN	CAPG-GAN	DA-GAN	FFWM
ACC(%)	99.63%	98.95%	96.42%	99.37%	99.56%	99.65%
AUC(%)	.	.	99.45%	99.90%	99.91%	99.92%

Fig.[G]: Comparison of FFWM with other GAN frontalization frameworks

The FFWM (Flow- Based Feature Warping Model) architecture is a novel approach to this problem, and due to its proven synthesis ability [35], ideal for the task of 3D reconstruction of the cranio-maxillofacial complex, due to the to the near-perfect warping of facial features, made possible by two aspects of this GAN. The FFWM includes an Illumination Preserving Module (IPM) that learns illumination preserving image synthesis from illumination inconsistent image pairs. The IPM includes two pathways that ensure the detailed, illumination consistent preservation of facial features. The second aspect is a Warp Attention Module (WAM) to reduce the pose discrepancy in the feature level, and hence to synthesize frontal images more effectively and preserve more details of profile and angular images (enabling the system to work even on larger pose-variations and discrepancies).

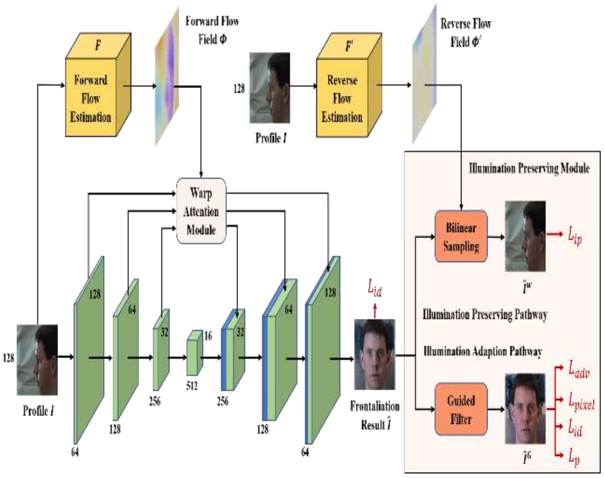


Fig. [H]: FFWM Architecture (integrated IPM and WAM)

The principal component of our generator is a residual network with 9 residual blocks (ResNet-9) between up sampling/down sampling operations (2 encoding and 2 decoding blocks), as proposed by Johnson et al (U-Net is used in the original implementation). It processes images of size 3 × 128 × 128 and returns an image of the same size. Each encoding/ decoding block has a two stride Convolution/Deconvolution-BatchNorm-ReLU structure, and each residual block has a Convolution-BatchNorm-ReLU-Convolution-BatchNorm residual connection structure (skip connections). This residual network is integrated with FlowNet 2.0 (FlowSD), to maintain optical flow (original FlowNet implementation requires two frames (input channels- 6), we use a single image (input channels -3)). Employing a forward and reverse Flow-Field we warp the image to its frontal view (additionally training the IPM (Leaky ReLU activation function) and WAM in the process). During the testing phase, the IPM thus establishes a consistency in illumination between the input and generated synthetic images, acting as a filter to transfer the illumination (Illumination Adaption Pathway). It also serves as a guidance layer, regulating the warping of profile images by the FlowNet module (Illumination Preserving Pathway). A 29-layer Light CNN is integrated to enhance the feature extraction process (trained and fine-tuned on the Multi-PIE dataset- 250 individuals in 300 different pose and illumination conditions). The discriminator is a 70 × 70 PatchGAN classifier, similar to the one used in the original pix2pix architecture (Isola et al. [36]). PatchGAN discriminators determines if the image is fake or real by analyzing a ‘patch’ of 70 × 70 pixels at a time, and generating final matrix with binary values for each patch on the 3 × 128 × 128 image. The image is passed through 4 Convolution/Deconvolution-BatchNorm-Leaky ReLU down sampling/ up sampling layers, to identify and classify the located patches (in a non-linear manner). The optimizer used is ‘Adam’, which is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments. The GAN cross-entropy loss implemented is the LSGAN (Least Squares GAN) objective function [37], which minimizes the Pearson x^2 divergence (as compared to Vanilla GAN loss, used in the original implementation). The objective function can be defined as:

$$\min_D V_{LSGAN}(D) = \frac{1}{2}E_{x\sim p_{data}(x)}\left[(D(x)-b)^2\right] + \frac{1}{2}E_{z\sim p_{data}(z)}\left[\left(D(G(z))-a\right)^2\right]$$

Using these parameters, the specialized model, and determining and setting the hyperparameters to optimum levels, we are able to successfully complete the photo-realistic synthetic frontalization task, which plays a crucial role in the second process involved in the system.

B. Face reconstruction and texture generation through 3D landmark projection

The second aspect of the system is the reconstruction procedure, that maps the images onto a 3D morphable model (BASEL [38] and FLAME [39] models are utilized in this study) to generate the facial mesh in the OBJ and STL format. FLAME is a lightweight and expressive generic head model learned from over 33,000 of accurately aligned 3D scans. It is written in TensorFlow, allowing it to be edited and mapped on in this process. We use the FLAME 2020 model in this study. The second model utilized in this study is the BASEL Face Model, which is a 3DMM that focuses specifically on the frontal face (compared to this, FLAME generates a model of the entire human head).

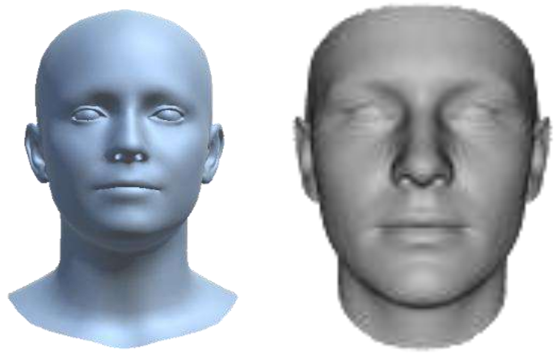


Fig.[I]: FLAME and BASEL Face Models (3DMMs)

The second procedure is further divided into several components:

1) <u>Landmark Identification & Extraction</u>

The pre-processed, frontalized image serves as the input for this step. We make use of the OpenCV (cv2) and MTCNN libraries to identify and extract the essential features of the generated face. The MTCNN model resizes the images to different scales in order to build an image pyramid, which serves as the input for the 3-staged cascaded network (Proposal Network (P-Net), Refine Network (R-Net), Output Network (O-Net)), that localizes and extracts the 5 facial landmarks: the left eye, the right eye, the nose, the left mouth corner and the right mouth corner.

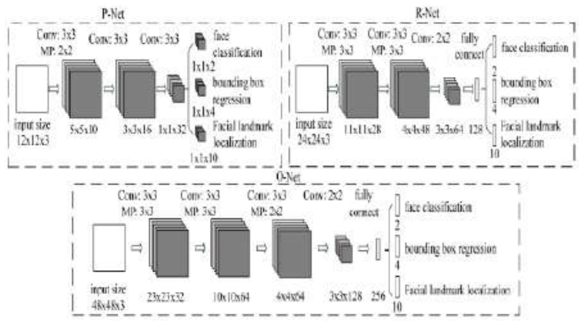


Fig.[J]: MTCNN Architecture

These points are perfect for the BASEL Face Model (utilized for cancer or skeletal injury cases on the central area of the face). For a higher projection accuracy, as desired, and required by other cranio-maxillofacial and cosmetic cases, we employ the DLib Package with OpenCV to extract 68 facial landmarks along with the identified areas (suitable for the more accurate FLAME Face Model). The facial landmark detector utilized by DLib is an implementation of the One Millisecond Face Alignment with an ensemble of Regression Trees paper (Kazemi and Sullivan, 2014). These identified landmarks are stored for landmark projection in the 3D space (Figure K).

ISSN-2347-4890
Volume 10 Issue 1 January 2022

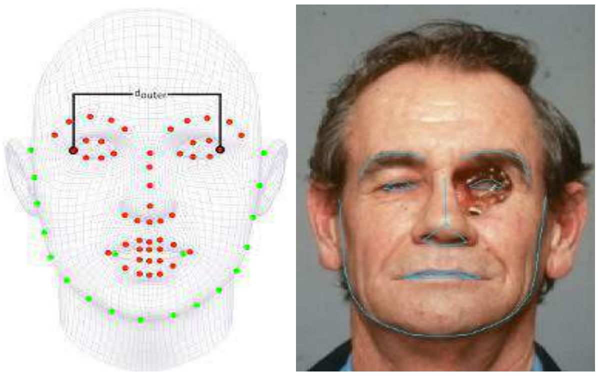


Fig.[K]

2) Texture Generation from the determined Texture Colour Space (FLAME/ BASEL)

The landmark projection and texture generation processes make use of a specialized mesh processing library- 'PyMesh' (version 0.4)- developed by the Max Planck Institute for Intelligent Systems. It enables the generation, editing, viewing, and exporting of 3D meshes and models. We have used PyMesh in Python version 3.7.9 along with libraries like Boost, TensorFlow, OpenCV, Pillow, MatPlotLib, Numpy, SciPy, OpenGL, and PyYaml, to map the landmarks onto the 3DMM (morphable model) and export it the OBJ and STL format. The FLAME and BASEL models are TensorFlow based 3DMM frameworks. We first use the landmarks identified in the previous step to find the corresponding regions on the 3DMMs. These identified regions are traced on the generated frontal image and warped to be transposed on the desired 3DMM (Generic, Male, or Female Model- based on the required outcome). This procedure is guided by the pre-computed FLAME and BASEL landmark embeddings and texture mappings, to generate the texture from the given colour space.

The synthesized texture is then baked and rendered on the computed mesh, as highlighted in the next step. This texture is essentially an edited and combined version of the frontalized and original image, that is eventually transposed on the 3DMM for enhanced accuracy in structure identification and oncologic visualization (geometric and finer details are retained by PyMesh through the embeddings).

3) Projection of landmark embeddings on the 3DMM and the generation of a 3D mesh

The stored landmark embeddings, from the previous step, are projected onto the 3DMM (BASEL/FLAME). They are scaled and regularized to be mirrored, utilizing the barycentric co-ordinates of the landmark embeddings in FLAME/BASEL topology (weight of the three vertices the landmark is embedded in – the two eyes and the mouth to locate the cranio-maxillofacial complex). This step uses transformation matrices (rotation, translation, scale, and shear) to transform the 2D co-ordinates to the 3D projection space. This fitting procedure utilizes the PyMesh and TensorFlow libraries, as the models (Generic, Male, Female- for FLAME) have different characteristic features, thus requiring different points (of the co-ordinates computed) to be projected. The geometric structure of the model is captured by the DLib library and the original landmark embeddings, so after projecting the 2D co-ordinates in the 3D space, based on the structure of the 3DMM, we process and transpose the texture onto the model. This texture mapping follows the Perspective Correct Mapping technique, which is suitable as it accounts for the vertices' positions in the 3D space than just regular simple 2D mapping. To perform perspective correctness, we first calculate the reciprocals at each vertex of our geometry. For vertex n we have $\frac{u_n}{z_n}, \frac{v_n}{z_n}, and\ \frac{1}{z_n}$, where u, v and z are depth components from a fixed point of view. Then we linearly interpolate these reciprocals across the surface, computing corrected values across the surface. These interpolated reciprocals are then used to recover the original co-ordinates, which is mathematically represented as:

$$u_\alpha = \frac{(1-\alpha)\frac{u_0}{z_0} + \alpha\frac{u_1}{z_1}}{(1-\alpha)\frac{1}{z_0} + \alpha\frac{1}{z_1}}$$

This creates a 3D mesh, which is saved as both an OBJ (generated textured mesh) and an STL (raw projected mesh) flexible, printable 3D image file, completing the process of mapping the image and reconstructing the entire face of the given individual. These files can be directly implemented and scaled for biomedical XR experimentation, relative to the patient, or be 3D printed for use as physical phantoms. The patient-specific and time-efficient nature of this rapid process allows it to be implemented in various other facial surgeries, in addition to oncology: trauma and even cosmetic applications, for it provides a relatively accurate 3D visual of the case at hand.

IV. ADVANTAGES

A. Outcome and Executional Improvements

Our system has an accuracy of 90% on the test dataset of in-the-wild images (LFW) with respect to the task of frontalization and cranio-maxillofacial reconstruction. It works effectively on several cases involving varied positional angles, similar to the geometric and computationally supervised procedures, with the added advantages of the unsupervised, generative approach that synthesizes highly similar photo-realistic and intricately extracted feature-centric images (due to strict classification by the discriminator). The generative model does not need manual input or classification of data to produce reasonable outputs, and works by learning facial structure on its own, unlike certain existing methods that reasonably rely on the symmetry of facial features. This enables a detailed feature extraction and translation process, identifying and exploiting previously unseen data without eliminating any features, which may be the case in manual computation, resulting in a loss of details and lower accuracy. As it identifies patterns on its own, without any manual intervention, there is a minimal workload to audit and prepare the training set, as compared to the other supervised and weakly-supervised deep learning approaches, where the images need to be extensively classified, for both training and testing. The integrated IPM and WAM models maintain optical flow, which is not considered in the standard geometric and computational translation processes. The use of Perspective Correct Texture Mapping adds to the accuracy of the created 3DMM, which would not have been the case if the traditional Forward Texture Mapping method, which reads texture maps in a linear manner and traverse texture co-ordinates directly, had been implemented. These detailed enhancements to the procedure improve the identification process (when compared to high pose variance in computational cases), and the use of the FFWM architecture that outperforms existing deep-learning solutions, result in an enhanced 3D-face reconstruction system for implementations in oncologic pre-surgery visualization. It is also executionally more efficient than existing alternatives (3D/PET/CT scans), allowing for instantaneous skin and structure visualization in relatively simpler, more external procedures and providing a scaled base for addition of skeletal scans (CT- for internal imaging).

B. Future Scope and Areas of Implementation

The final result of the entire process is a 3DMM that represents the structure of the cancer affected head/ face (FLAME/ BASEL model) of the individual in the given image, with pose variation. This model/mesh can be transformed into any known view for further surgical or clinical analysis and representation. In addition to the oncologic cases, it is suitable for, this instantaneous reconstruction method can also be potentially useful in cosmetic surgery, to provide a structural base. Such a system, when slightly modified, may also be useful in pediatric, oral and maxillofacial trauma assessments (including cleft-lip and external cell growths, when integrated with scaled skeletal scans). A similar reconstruction method may also be implemented for other areas of the body (neck or foot carcinomas, for example) that demand external examination in surgical or other invasive approaches. Moreover, access to a digital 3D model of the area to be operated on, may allow implementation of a surgical recommendation system for standardized cases. It also has applications in the design of drug-delivery systems for oral medication and aerosol masks, that must be individual-specific. Outside of the field of medicine, it may be applied to 3D face recognition and security analysis, in instances involving large pose variation and partial obstruction. Having been trained and tested on a vast collection of data (Multi-PIE and LFW), the model generates a closely similar potential 3D image, even on wider angles, allowing for it to be used in several, varied fields (can, however, be improved with a more specialized training process).

V. RESULTS AND CONCLUSION

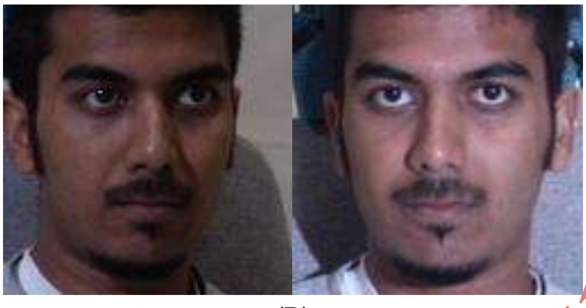
A. Reconstruction Assessment on LFW Dataset

The current system has been trained on images from the Multi-PIE dataset, for the task of frontalization of the cranio-maxillofacial complex. This model has been tested on the LFW dataset to verify its accuracy in facial translation. This data must be photo-realistically accurate (enabled by the WAM and IPM of the FFWM framework), as it aids the final projection and reconstruction procedure. The trained model had an accuracy of 90% on the LFW test dataset, and works on a variety of pose discrepancies (even large pose-variations (including profile images)). Described below are pose-varied detailed examples from the LFW dataset for analysis of the model on the frontalization task (by comparing with original frontalized data) and the eventual reconstruction procedure. The synthesized images on the original training set are also presented for review.

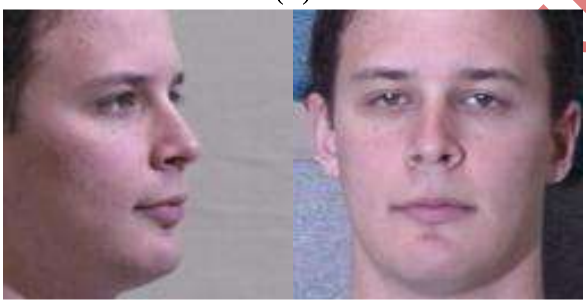
Multi-PIE Dataset

The model successfully synthesized images from the Multi-PIE dataset under varied pose and illumination conditions (Figure L).

(A)



(B)



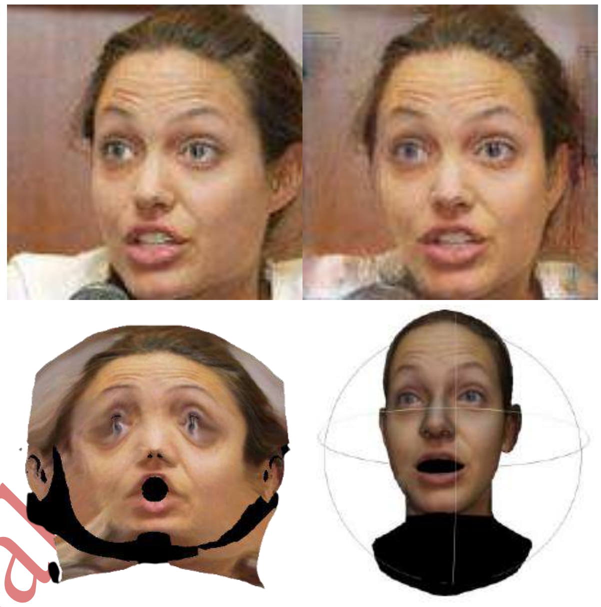
(C)

Fig.[L]

LFW Dataset

The system was tested on the LFW test dataset to determine the quality of generated 3D models. The results are displayed below (Figure M).

(A)



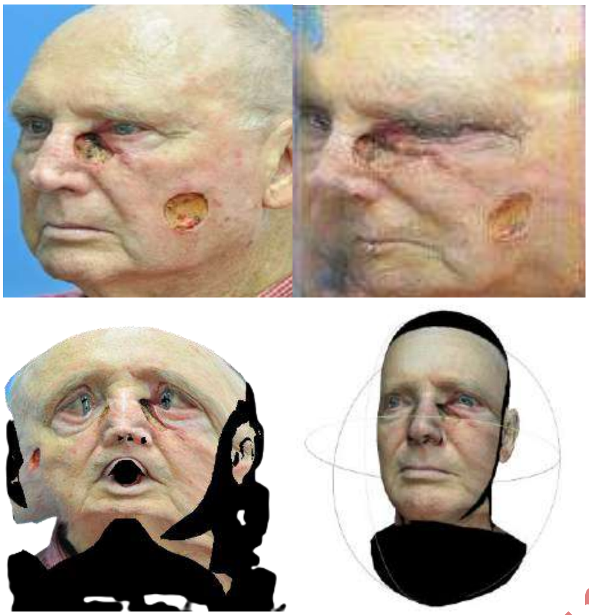
(B)

Fig.[M]

B. Oncologic Patient-Specific Reconstruction

The purpose of this study was to enable rapid reconstruction of cranio-maxillofacial surgical cases to provide a base (or even final) model for pre-surgical visualization. The extensive nature of keratinocyte and other carcinomas can be captured by the system, and here we review the proposed reconstruction procedure, specifically on an unseen, oncological cranio-maxillofacial case (frontal [A] (for procedure) and [B] wide angle) (Figure N).

(A)



(B)

Fig.[N]

The final 3D morphable mesh produced can be easily, digitally implemented in biomedical XR, and the conversion to both STL and OBJ formats, by the system, enables flexible, 3D printing of physical phantoms, that can be varied as per the specific surgical case and hand. To conclude, we believe that the attributes of such a system, will allow for a shorter and cost-efficient treatment planning process, one that will be beneficial to both the patient and the surgical team, It also provides a base structure to examine and add 3D skeletal arrangements, for cases involving cranio-maxillofacial trauma and cosmetic surgery, eliminating the 3D modelling and commercial phantom obtainment step. Moreover, its accuracy and illumination-preserving, photo-realistic results can also help in the discovery of its applications in other medical areas (and external fields).

VI. REFERENCES

1. Robins JMW, Sheikh AJ, Shastin D, Schramm MWJ, Carter P, Russell JL, Liddington M, Chumas PD. Fronto-orbital advancement and reconstruction using reverse frontal bone graft without the use of orbital bar: a technical note. Childs Nerv Syst. 2020 Jun;36(6):1295-1299. doi: 10.1007/s00381-020-04583-w. Epub 2020 Mar 26. PMID: 32219525; PMCID: PMC7250796. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7250796/

2. Chen WL, Yang ZH, Huang ZQ, Fan S, Zhang DM, Wang YY. Craniofacial Resection and Reconstruction in Patients With Recurrent Cancer Involving the Craniomaxillofacial Region. J Oral Maxillofac Surg. 2017 Mar;75(3):622-631. doi: 10.1016/j.joms.2016.08.044. Epub 2016 Sep 13. PMID: 27717818. https://pubmed.ncbi.nlm.nih.gov/27717818/

3. Benmahdjoub M, van Walsum T, van Twisk P, Wolvius EB. Augmented reality in craniomaxillofacial surgery: added value and proposed recommendations through a systematic review of the literature. Int J Oral Maxillofac Surg. 2021 Jul;50(7):969-978. doi: 10.1016/j.ijom.2020.11.015. Epub 2020 Dec 16. PMID: 33339731. https://pubmed.ncbi.nlm.nih.gov/33339731/

4. Xiaojun Chen, Lu Xu, Yiping Wang, Huixiang Wang, Fang Wang, Xiangsen Zeng, Qiugen Wang, Jan Egger, Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display, Journal of Biomedical Informatics, Volume 55, 2015, Pages 124-131, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2015.04.003. (https://www.sciencedirect.com/science/article/pii/S1532046415000702)

5. Hsieh, Chung-Hung, Jiann-Der Lee, and Chieh-Tsai Wu. "A Kinect-based medical augmented reality system for craniofacial applications using image-to-patient registration." Neuropsychiatry 7.6 (2017): 927-939.

6. Alshomer F, Alazzam A, Alturki A, Almeshal O, Alhusainan H. Smartphone-assisted Augmented Reality in Craniofacial Surgery. Plast Reconstr Surg Glob Open. 2021 Aug 13;9(8):e3743. doi: 10.1097/GOX.0000000000003743. PMID: 34414055; PMCID: PMC8367041. https://pubmed.ncbi.nlm.nih.gov/34414055/

7. Giovanni Badiali, Vincenzo Ferrari, Fabrizio Cutolo, Cinzia Freschi, Davide Caramella, Alberto Bianchi, Claudio Marchetti, Augmented reality as an aid in maxillofacial surgery: Validation of a wearable system allowing maxillary repositioning, Journal of Cranio-Maxillofacial Surgery, Volume 42, Issue 8, 2014, Pages 1970-1976, ISSN 1010-5182,

https://doi.org/10.1016/j.jcms.2014.09.001. (https://www.sciencedirect.com/science/article/pii/S1010518214002522)

8. Carlos G. Landaeta-Quinones, Nicole Hernandez, Najy K. Zarroug, Computer-Assisted Surgery: Applications in Dentistry and Oral and Maxillofacial Surgery, Dental Clinics of North America, Volume 62, Issue 3, 2018, Pages 403-420, ISSN 0011-8532, ISBN 9780323610766, https://doi.org/10.1016/j.cden.2018.03.009. (https://www.sciencedirect.com/science/article/pii/S0011853218300247)

9. Beyer, T., Townsend, D.W., Brun, T., Kinahan, P.E., Charron, M., Roddy, R., Jerin, J., Young, J., Byars, L. and Nutt, R., 2000. A combined PET/CT scanner for clinical oncology. Journal of nuclear medicine, 41(8), pp.1369-1379.

10. Modeling and Simulation of 4D PET-CT and PET-MR Images Tsoumpas, Charalampos et al. PET Clinics, Volume 8, Issue 1, 95 - 110

11. Wong WL, Batty V. Role of PET/CT in maxillo-facial surgery. Br J Oral Maxillofac Surg. 2009 Jun;47(4):259-67. doi: 10.1016/j.bjoms.2008.11.011. Epub 2009 Feb 5. PMID: 19200626. https://pubmed.ncbi.nlm.nih.gov/19200626/

12. Bibb, R., Eggbeer, D. & Paterson, A. Medical modelling: the application of advanced design and rapid prototyping techniques in medicine. (Woodhead Publishing, 2014).

13. Filippou, V. and Tsoumpas, C. (2018), Recent advances on the development of phantoms using 3D printing for imaging with CT, MRI, PET, SPECT, and ultrasound. Med. Phys., 45: e740-e760. https://doi.org/10.1002/mp.13058

14. Amorim, P., Moraes, T., Silva, J. & Pedrini, H. Invesalius: An interactive rendering framework for health care support. In Bebis, G. et al. (eds.) Advances in Visual Computing. Lecture Notes in Computer Science, vol. 9474, 45–54 (Springer, Cham, 2015)

15. Clement, J.G. and Marks, M.K., 2005. Computer-graphic facial reconstruction. Elsevier.

16. 3D Semi-Landmarks Based Statistical Face Reconstruction. Maxime Berar, Michel Desvignes, Gerard Bailly, Yohan Payan https://doi.org/10.2498/cit.2006.01.04

17. Elyan, Eyad & Ugail, Hassan. (2007). Reconstruction of 3D Human Facial Images Using Partial Differential Equations. Journal of Computers. 2. 10.4304/jcp.2.8.1-8. https://www.researchgate.net/publication/27313708_Reconstruction_of_3D_Human_Facial_Images_Using_Partial_Differential_Equations

18. Wang, D. 2005. Face Reconstruction Across Different Poses and Arbitrary Illumination Conditions. In Audio- and Video-Based Biometric Person Authentication (pp. 91–101). Springer Berlin Heidelberg. https://doi.org/10.1007/11527923_10

19. Rudomin, Isaac & Cuevas, Hector. (2000). Generating a 3D Facial Model From a Single Image using Principal Components Analysis. https://www.researchgate.net/publication/261954131_Generating_a_3D_Facial_Model_From_a_Single_Image_using_Principal_Components_Analysis

20. Y. Guan, "Automatic 3D Face Reconstruction based on Single 2D Image," 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07), 2007, pp. 1216-1219, doi: 10.1109/MUE.2007.95. https://ieeexplore.ieee.org/abstract/document/4197445

21. Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, & Xi Zhou. (2018). Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network. https://arxiv.org/abs/1803.07835

22. Aaron S. Jackson, Adrian Bulat, Vasileios Argyriou, & Georgios Tzimiropoulos. (2017). Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression. https://arxiv.org/abs/1703.07834

23. Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, & Xin Tong. (2020). Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set. https://arxiv.org/abs/1903.08527

24. Pengfei Dou, Shishir K. Shah, & Ioannis A. Kakadiaris. (2017). End-to-end 3D face reconstruction with deep neural networks. https://arxiv.org/abs/1704.05020

25. Alexandros Lattas, Stylianos Moschoglou, Baris Gecer, Stylianos Ploumpis, Vasileios Triantafyllou, Abhijeet Ghosh, & Stefanos Zafeiriou. (2020). AvatarMe: Realistically Renderable 3D Facial Reconstruction "in-the-wild". https://arxiv.org/abs/2003.13845

26. Gecer, B., Ploumpis, S., Kotsia, I., & Zafeiriou, S. (2019). GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://arxiv.org/abs/1902.05978

27. https://towardsdatascience.com/getting-started-with-gans-using-pytorch-78e7c22a14a5

28. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal

Processing Letters, 23(10), 1499–1503. https://arxiv.org/abs/1604.02878

29. https://medium.com/ai%C2%B3-theory-practice-business/3d-morphable-face-models-past-present-and-future-b3a7f47e1b04

30. Huang, Gary & Mattar, Marwan & Berg, Tamara & Learned-Miller, Eric. (2008). Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. Tech. rep.. https://www.researchgate.net/publication/29622837_Labeled_Faces_in_the_Wild_A_Database_forStudying_Face_Recognition_in_Unconstrained_Environments

31. R. Gross, I. Matthews, J. Cohn, T. Kanade and S. Baker, "Multi-PIE," 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1-8, doi: 10.1109/AFGR.2008.4813399. https://ieeexplore.ieee.org/document/4813399

32. Rui Huang, Shu Zhang, Tianyu Li, & Ran He. (2017). Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis. https://arxiv.org/abs/1704.04086

33. L. Tran, X. Yin and X. Liu, "Disentangled Representation Learning GAN for Pose-Invariant Face Recognition," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1283-1292, doi: 10.1109/CVPR.2017.141. https://ieeexplore.ieee.org/document/8099624

34. Shuang Ma, Jianlong Fu, Chang Wen Chen, & Tao Mei. (2018). DA-GAN: Instance-level Image Translation by Deep Attention Generative Adversarial Networks (with Supplementary Materials). https://arxiv.org/abs/1802.06454

35. Wei, Yuxiang & Liu, Ming & Wang, Haolin & Zhu, Ruifeng & Hu, Guosheng & Zuo, Wangmeng. (2020). Learning Flow-based Feature Warping for Face Frontalization with Illumination Inconsistent Supervision. https://www.researchgate.net/publication/343711482_Learning_Flow-based_Feature_Warping_for_Face_Frontalization_with_Illumination_Inconsistent_Supervision

36. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, & Alexei A. Efros. (2018). Image-to-Image Translation with Conditional Adversarial Networks. https://arxiv.org/abs/1611.07004

37. Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, & Stephen Paul Smolley. (2017). Least Squares Generative Adversarial Networks. https://arxiv.org/abs/1611.04076

38. P. Paysan, R. Knothe, B. Amberg, S. Romdhani and T. Vetter, "A 3D Face Model for Pose and Illumination Invariant Face Recognition," 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 296-301, doi: 10.1109/AVSS.2009.58. https://ieeexplore.ieee.org/document/5279762

39. Li, Tianye & Bolkart, Timo & Black, Michael & Li, Hao & Romero, Javier. (2017). Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics. 36. 1-17. 10.1145/3130800.3130813. https://www.researchgate.net/publication/321234860_Learning_a_model_of_facial_shape_and_expression_from_4D_scans

40. Gsaxner, C., Wallner, J., Chen, X. et al. Facial model collection for medical augmented reality in oncologic cranio-maxillofacial surgery. Sci Data 6, 310 (2019). https://doi.org/10.1038/s41597-019-0327-8

VII. ACKNOWLEDGEMENT

I would like to express my deep gratitude to Dr. Ganesh Mani (Carnegie Mellon University, School of Computer Science), my research mentor, and Dr. Shoou- I Yu (Research Scientist at Oculus Research) for their enthusiastic encouragement and useful critiques of this research work. Their invaluable inputs truly guided the development and execution of this project.

I would also like to thank my parents for their constant support during the course of this study.

VIII. AUTHOR

Abhishek Pillai

Cleared the Indian Certificate Of Secondary Education Examination (2020). Currently pursuing the Higher Secondary School Certificate from PACE Junior Science College, Andheri.

Email Address: abhishekpillai94@gmail.com