WHITE PAPER

3D FACE RECONSTRUCTION:

a journey to accurate, fast, and accessible mobile solution

Authors: Pavlo Menshykh, Taras Khapko

tsukat

2023

CONTENTS

Navigate further:
Click on a preferred chapter

Introduction

01

Businesses need to stay ahead of the curve in order to remain competitive, in any era and in any industry. In our case, accurate, fast and accessible simple 3D face reconstruction is an innovative technology that will give companies who have it massive advantage over their competitors. Utilizing advanced scanning technologies and algorithms, businesses can leverage 3D face reconstruction to unlock numerous game-changing advantages like customization and personalization, virtual try-on and fitting, improved security measures and more. In today's competitive landscape, where personalization, efficiency, and customer experiences are paramount, achieving accurate, fast, and simple 3D face reconstructions holds the key to unlocking new opportunities and driving business success.

At **tsukat**, we recognize the crucial role reliable technology products play in our clients' operations. To ensure that we provide only the most exceptional and trustworthy solutions, we conduct thorough internal testing on all new technology products before making them available to our clients. This meticulous testing and evaluation process instills confidence in our clients, allowing them to focus on their business operations, knowing they can rely on the quality and performance of the products we offer.

This whitepaper documents our journey of trial and error toward enhancing the 3D face scanning process. Our focus lies on the essential aspects of 3D face reconstruction, namely accuracy, speed, and accessibility, which we briefly discuss throughout this document. What sets our approach apart is the ability to scan a dynamic human face using a mobile device.

At the end of the day, we believe that by leveraging advanced scanning techniques, intelligent algorithms, and streamlined workflows, businesses have the potential to revolutionize their operations, products, and services.

> **3D face reconstruction**
> **is an innovative technology that will give companies who have it massive advantage over their competitors.**

02

CHALLENGE:

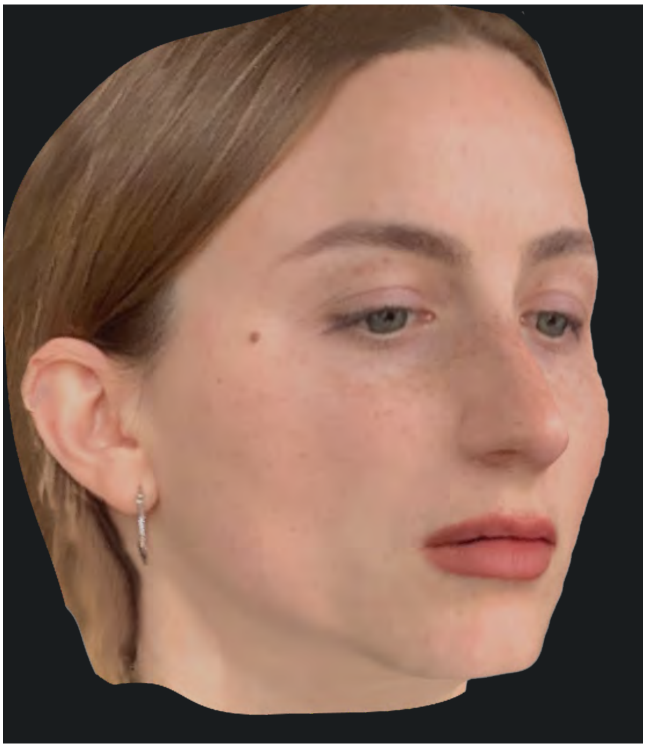
Obtaining accurate 3D face reconstruction on a mobile device.

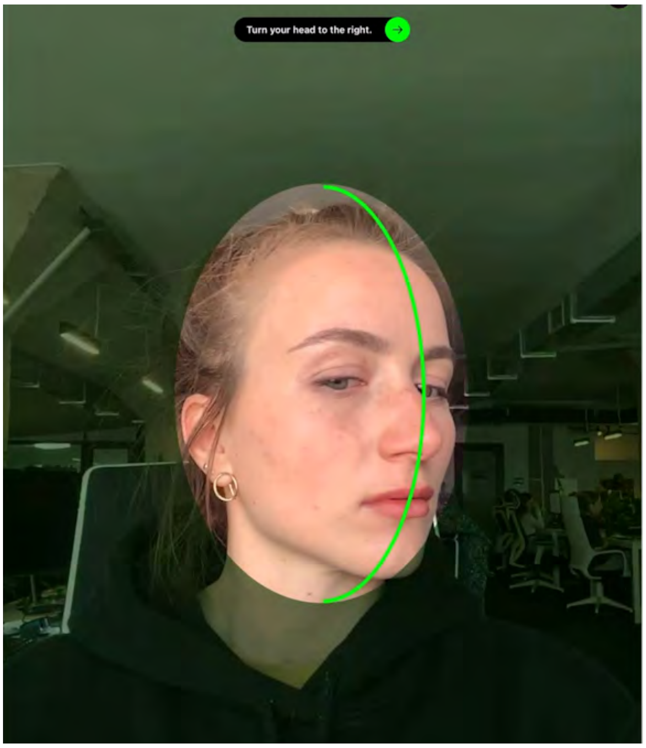
3D object reconstruction is a challenging task that involves capturing the spatial structure and appearance of real-world objects in order to obtain an accurate three-dimensional digital representation, which can be later altered, viewed and shared. With ongoing digitalization trends, this problem garnered significant attention due to its wide range of applications across various domains, including design and engineering, architecture, art, healthcare, education, etc.

Among the various applications of 3D object reconstruction, scanning the human face stands out in terms of its complexity and types of use cases. Accurate 3D face reconstruction enables a range of applications, including personalized avatar creation and animations, surgical planning and prosthetics design, as well as virtual apparel and makeup try-on.

Modern phones with depth-sensing capabilities have transformed the 3D scanning field by combining portability, advanced hardware, sophisticated software algorithms and artificial intelligence advances. They made the process more accessible to a broader audience while also consistently raising the quality of the produced results. Still, when it comes to face scanning, there are specific challenges and considerations that need to be accounted for.

Scanning procedures rely on matching pictures/point sets between different views. For the matching to be accurate the object has to remain stationary in different perspectives. It is, therefore, typical for the scanned object to be locked in place, and the scanning device to be moved around it. Unlike inanimate objects, however, humans are always in motion consciously or unconsciously. Facial expressions, movements and articulations introduce complexities that render regular approaches impractical. In addition, faces are often occluded by hair or accessories, while variations in lighting conditions introduce noise in the captured data and deteriorate texture quality.





Nevertheless, there are a number of existing solutions for tackling the aforementioned issues to create realistic face scans. The one that stood out was the Bellus3D face scanning solution, a powerful and easy-to-use application for iPhones and iPads with TrueDepth cameras. When Bellus3D exited the face scanning market, it created a void that impacted numerous use cases in which easy-to-create but also accurate facial scans were required. Existing alternatives are designed as general usage scanning apps, and are neither made for making human scans without external help, nor providing the necessary quality of texture and geometry accuracy of the scanned object.

Thus we set out on a mission to develop a custom solution that would address the specific needs of partners who relied on Bellus3D for face scanning functionality.

Our primary goals in this endeavor were:

DETECTION OF KEY FACIAL FEATURE POSITIONS IN 3D, INCLUDING EARS

Beyond the overall face reconstruction, we aimed to develop algorithms capable of accurately detecting and positioning key facial features, including the ears. This feature detection would enable customization and further applications based on specific facial attributes.

ON-DEVICE PROCESSING WITHIN A REASONABLE TIME FRAME

To ensure the best user experience, where users can gain real-time or near-real-time results, we focus on performing processing tasks directly on the scanning devices. In cases with limited data transmission speeds or low-quality internet, this approach aimed to ensure timely feedback and minimize delays associated with transferring data to external processing resources.

INCLUSION OF THE FRONTAL PART OF THE FACE, HAIR, AND EARS

Our solution aimed to capture not only the frontal region of the face but also the surrounding elements such as hair and ears. This comprehensive scanning approach would enable a more holistic representation of the individual's facial structure.

USER-FRIENDLY SCANNING PROCEDURE

We strived to design an intuitive and simple scanning procedure for users to perform independently. By holding the device in front of them and rotating their heads, users would be able to conduct the scanning process without requiring external assistance, ensuring ease of use and convenience.

ACHIEVING THE HIGHEST POSSIBLE ACCURACY IN SCAN GEOMETRY

We aimed to optimize the scanning process utilizing Apple devices equipped with TrueDepth cameras, while ensuring the resulting scans exhibited exceptional precision in capturing facial features. Additionally, we aimed to maintain high-quality texture representation for a comprehensive and detailed 3D reconstruction.

By pursuing these goals, we aimed to create a robust and reliable solution that met the requirements of our partners and delivered exceptional results in terms of accuracy, accessibility, and efficiency. Through our research and development efforts, we sought to push the boundaries of existing 3D face scanning technology and empower businesses with advanced capabilities for their respective industries.

> **“ Thus we set out on a mission to develop a custom solution that would address the specific needs of partners who relied on Bellus3D for face scanning functionality.**

03

TRIAL AND ERROR

With a range of possible approaches and ideas at our disposal, we started rapid prototyping in order to fail-fast irrelevant ones and focus on the most promising ideas.

PHOTOGRAMMETRY

The first act was to explore the potential of utilizing photogrammetry. Adopting photogrammetry would eliminate the requirement for the TrueDepth camera, thereby expanding the range of supported devices and increasing accessibility.

Photogrammetry offers several advantages. It enables the creation of detailed 3D reconstructions by capturing multiple images from different angles. However, one of the challenges in using photogrammetry for human face scanning is the need for consistent object positioning between frames. This becomes particularly difficult when users rotate their heads, resulting in subtle facial movements and the displacement of loose hair.

As a result, the industry has devised workarounds to overcome these challenges. In production environments where photogrammetry is utilized for capturing human faces, professionals often employ hair net caps and multiple cameras to capture simultaneous images. This approach helps to maintain consistency and mitigate the impact of head rotation and hair movement, ensuring more accurate and reliable results.

Our particular scenario presented computational challenges for this approach when operating within the limitations of a mobile device. Furthermore, despite utilizing more powerful computational resources, we were unable to achieve the desired level of quality while adhering to our preferred scanning methodology.

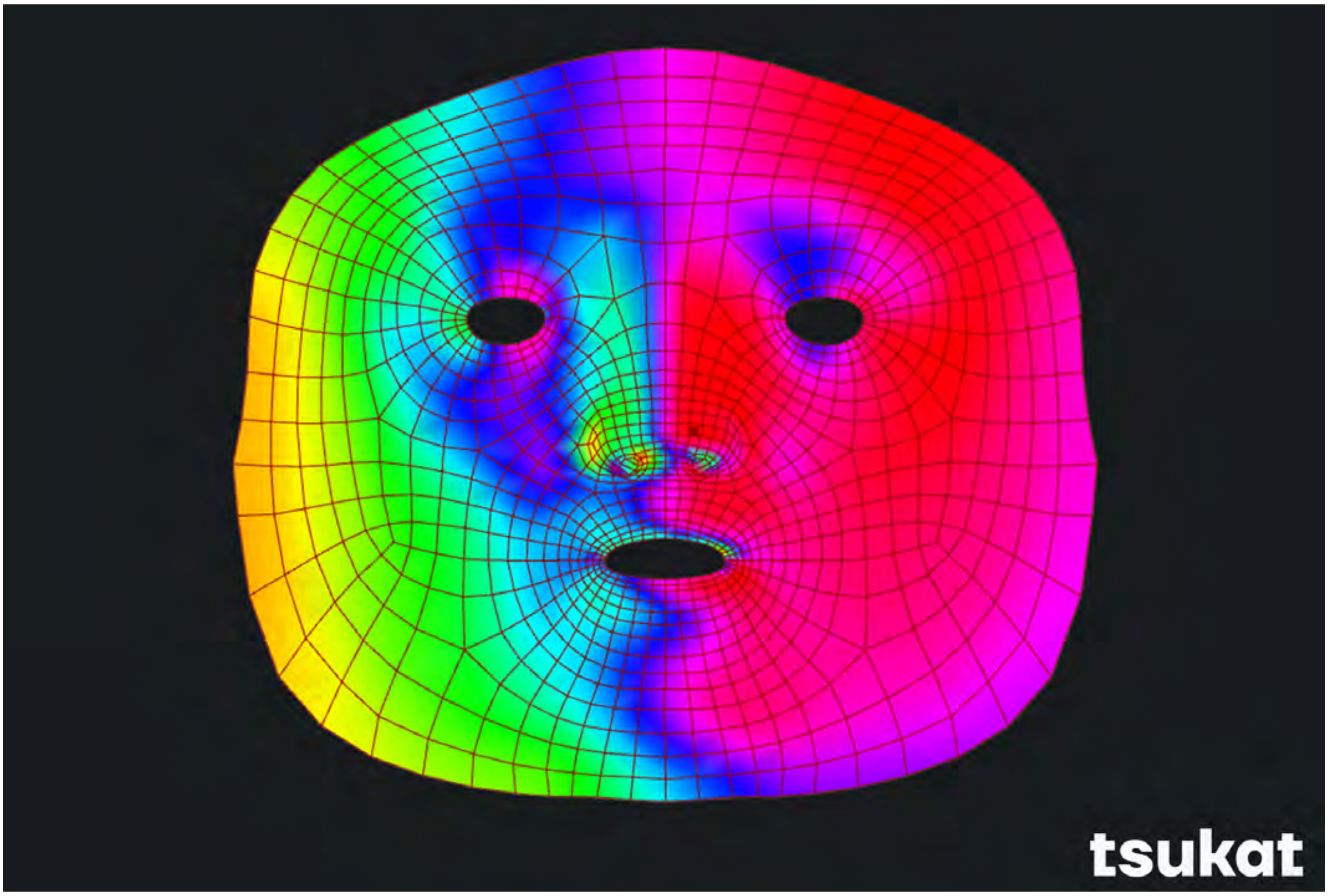
TEMPLATED FACE MODEL

Our next bet was to return to Apple devices with TrueDepth and utilize the capabilities of the ARKit. The latter has the functionality to track human faces and produces a templated version of the face mesh (in the form of a hockey-like mask) in each frame.

Texture for this mesh was produced by blending the color information between several frames at a set of different rotation angles. Since the user is performing the scan by rotating the head, and the ARKit alignment does not provide pixel-level accuracy, the resulting color entry points are not consistent in position and lightness.

To overcome the texture quality issue a different blending technique was designed, with a blend map generated for each frame.

This allowed us to achieve better uniformity of the texture, albeit lacking in sharpness.

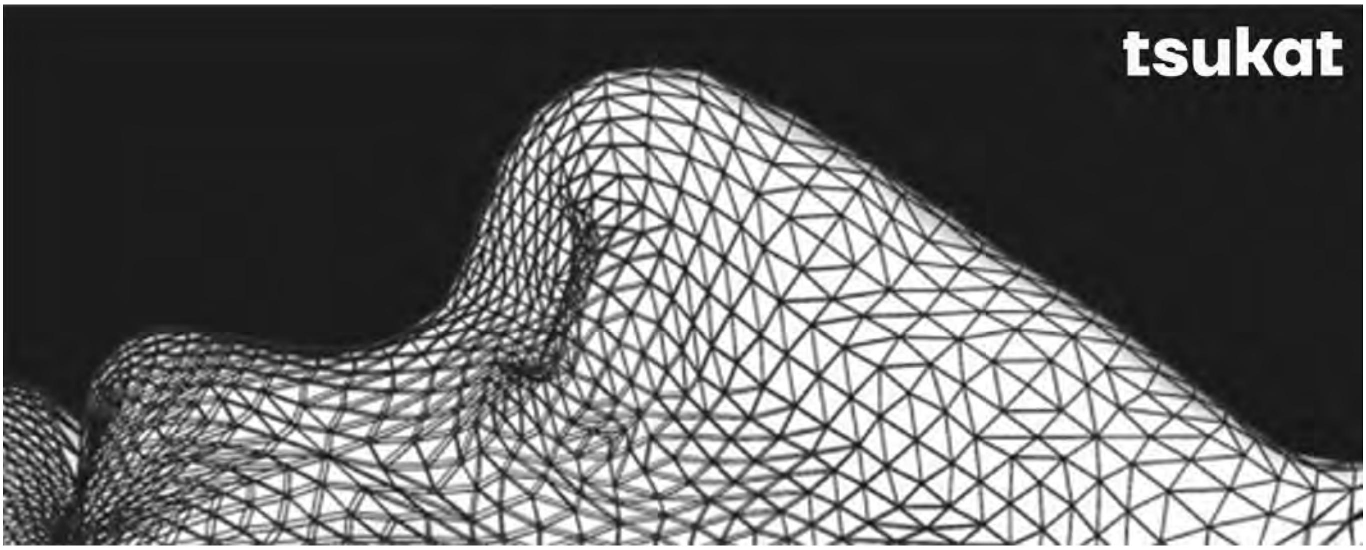


Another shortcoming of this approach was the fact that the face mesh was limited to a hockey-like mask, which would severely limit possible use cases. To address this, the model used for reconstruction was expanded with procedural ears and more face space.

> **The face mesh was limited to a hockey-like mask, which would severely limit possible use cases.**

In order to leverage the extended model, more frames had to be used for the reconstruction. The enhanced texturing approach mentioned above did not scale well with an increase in the number of frames and the face at different angles with respect to the camera. The texture approach was thus re-worked with use of a custom alignment method between the frames.

In addition, the eyes had to be treated separately because the person tends to focus their gaze on the camera instead of following the movement of the head. Given the role of the human eye in potentially conveying or revealing deep emotional states, and the opportunities this suggests, we wanted to ensure the eyes are being reconstructed accurately.



Additionally, the templated mesh in its original form is a little over 1000 vertices. This is more than enough for applications where the face scan is needed in the case of visual representation only, but is insufficient for cases requiring millimeter-level accuracy.

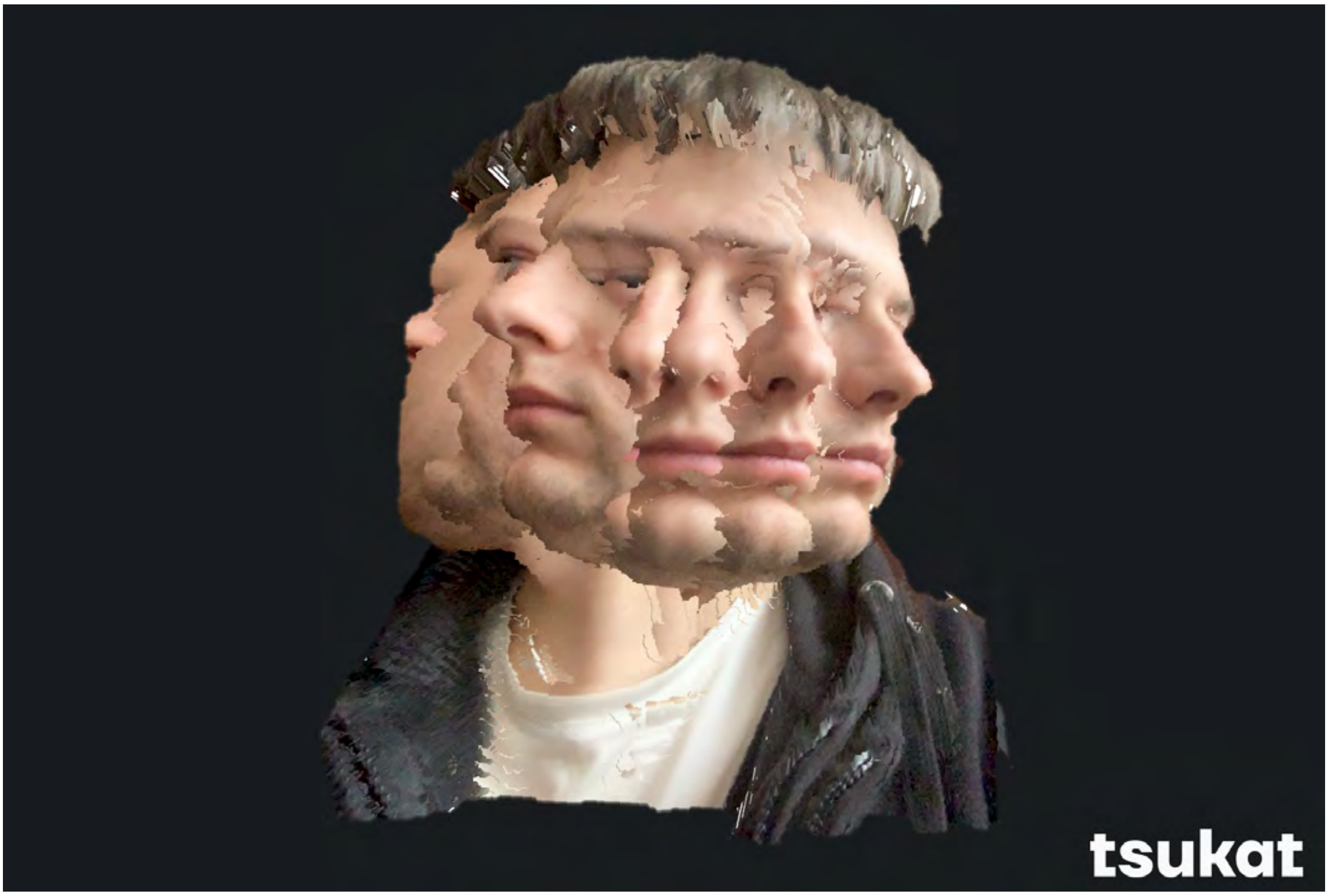
To improve this aspect, a custom subdivision approach was developed. It adds new vertices not by classical mesh sub-division (which would change the existing geometry and smooth everything out), but instead by adding new vertices that continue the existing curvature of the already-present vertices.

While achieving positive results, ultimately the outcomes of this method fall short of the desired scan quality. The resulting reconstruction is a templated mesh fitted to the individual human face. The fitting itself captures the overall looks and achieves similar appearance to a specific person, however completely neglects irregularities and asymmetries of the face, ultimately falling short in terms of point-by-point accuracy.

“ **While achieving positive results, ultimately the outcomes of this method fall short of the desired scan quality.**

RECONSTRUCTED MESH

In order to capture the intricate details and unique features of each face, we had to take a step back and create our own reconstructed mesh. This was accomplished by leveraging raw point cloud data from the device, and combining several frames from different perspectives to form a single point cloud of the frontal part of the head.



It is important to note that Apple devices use a laser that projects 30000 dots in a known pattern captured by the camera. Some of the dots will fall on object edges (hair is especially problematic in this regard), while others will fail to be registered at all. The resulting depth buffer has a resolution of 640x480, which is 307200 data points. This is already heavily interpolated compared to the projected 30000 points pattern, which means some degree of inaccuracy will always exist due to the scanning hardware itself. After additional filtering of outliers, however, we were able to run the geometry reconstruction and obtain the initial mesh.

Industry standard approaches for all stages of model creation from point clouds were developed to be universal and work for most if not all objects that could be scanned (buildings, figures, etc.). Typically they are also not targeted to mobile device platforms and require extensive computational power. By using these more standard approaches we could get a reconstruction of the face model with a good level of detail, but containing large amounts of noise and artifacts. Moreover, the computation time was well over a minute, which did not fit our initial plans and constraints.

The accuracy level in the model's color representation is determined by the texture resolution when utilizing a texture, or the vertex density when employing vertex coloring. In the former case, it is necessary to construct a texture atlas and project the texture onto the model, whereas vertex coloring does not require this step. The latter approach is commonly adopted in the context of mobile device scanning applications.

Considering that a high-resolution mesh of a human face typically consists of approximately 50,000 points, the level of detail achieved with vertex coloring would be equivalent to that of a 256x256 texture.

The computation of the texture atlas is performed iteratively, which leads to slower processing times and greater resource demands.

Furthermore, the resulting texture atlas contains fragmented regions and visible seams. In many cases, manual intervention by skilled artists is required to conceal these seams in inconspicuous areas. However, our approach requires this process to be automated, calling for a solution that handles the texture atlas generation automatically while mitigating the presence of visible seams.

In order to achieve desired quality and performance, we had to create both our own custom meshing and texturing approaches, which finally produced the outcomes that satisfied all of the initial constraints on the scan quality. We have provided the outcome of our approach in the results section below.

“ **We had to create both our own custom meshing and texturing approaches, which finally produced the outcomes that satisfied all of the initial constraints on the scan quality.**

LANDMARKS DETECTION

After achieving the required level of realism and accuracy of the face scan, we still had to incorporate key facial landmark positions into the output. While there are numerous viable options for face landmark detection, none of them includes ears, which are of great importance in certain applications.

In general, ear landmarks can be extrapolated from the ones on the face, which yields lower accuracy and larger deviations from person to person. Instead, we chose to train our own neural network to accurately predict ear positions in the image space, which would then be unprojected to the 3D scan to obtain their 3D positions.

There is a lack of available training data for such a problem, leading us to generate our own ground truth. Manual labeling of captured images was considered to be far too tedious a process, and we resorted to synthetic data generation. Creating a synthetics pipeline allowed us to obtain the necessary quality and variety of data in required large amounts, ensuring successful training and ultimately great accuracy of ear landmarks detection.

04

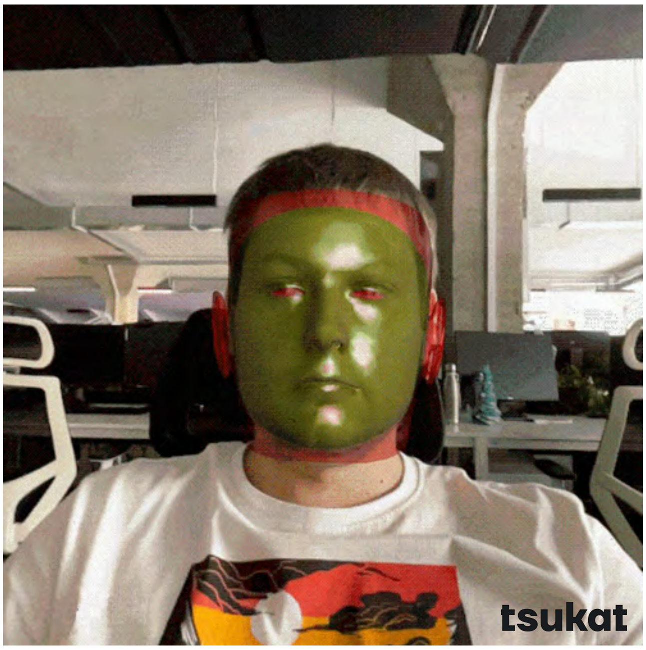
RESULT

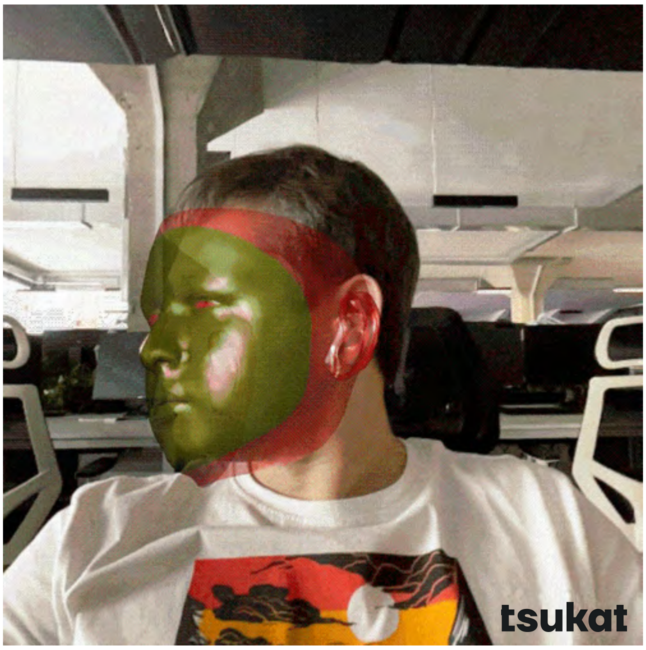
We believe that every
face tells a unique story

The path to obtaining the required quality of the face scan and the performance of the computations led us to developing our own approaches for model reconstruction from point cloud, as well as texture unwrapping and texture computation.

Those approaches are tailored specifically for face scanning and obtaining maximum speed while maintaining high quality.

SCANNING PROCEDURE



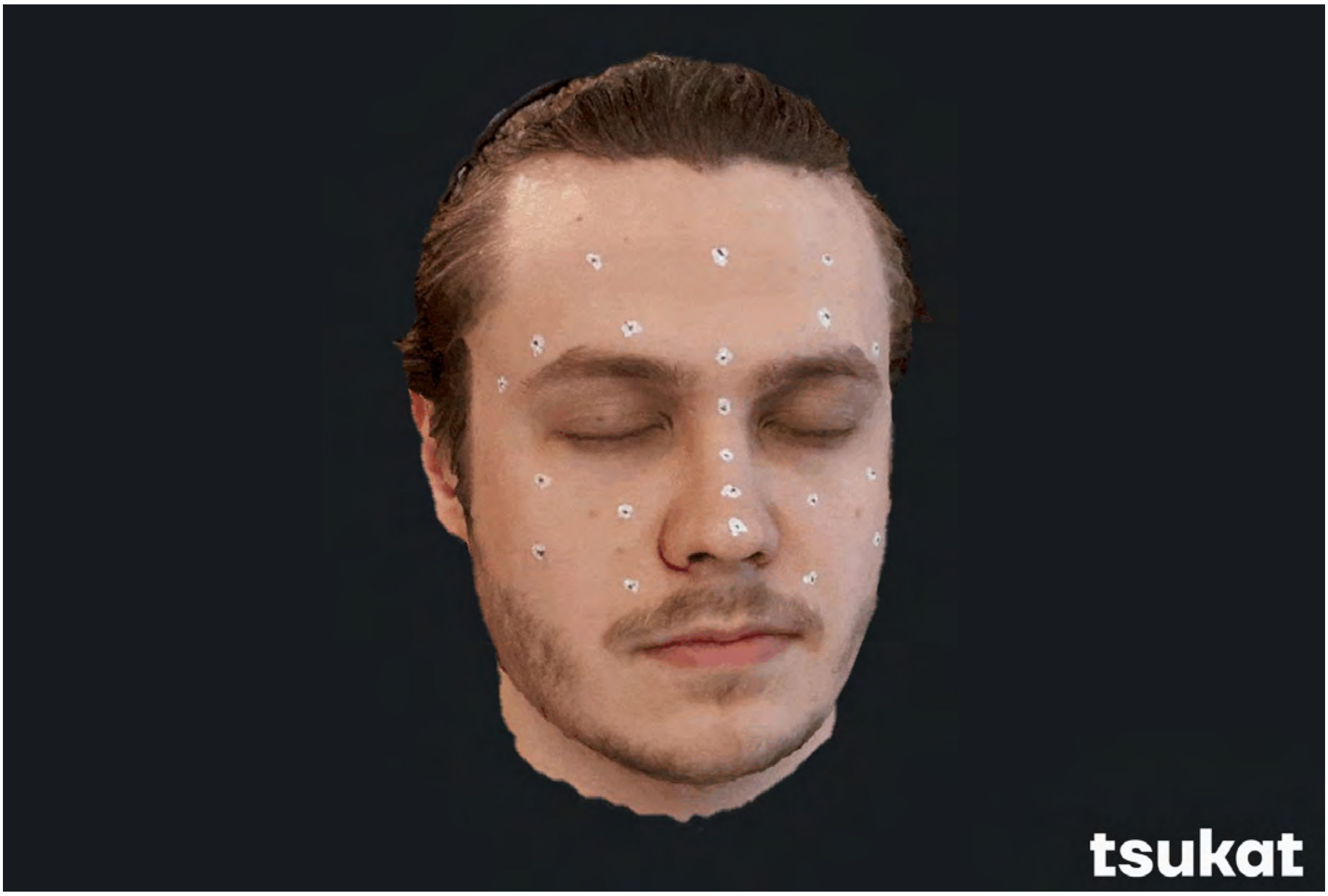


The scanning procedure is designed to be user-friendly, allowing the person being scanned to perform the process themselves. During the scanning, the phone is positioned at an approximate arm's distance from the head, and the individual rotates their head following the on-screen instructions. To ensure optimal quality, certain recommendations should be followed. These include maintaining a well-lit environment with minimal shadows or reflections to ensure high-quality texture capture.

It is important to ensure that the face is unobstructed by excessive hair or head accessories that may disrupt the alignment of point clouds. Additionally, tilting the phone slightly relative to the head can improve the capture of areas beneath the nose and chin. By adhering to these recommendations, the scanning process can yield superior results in terms of accuracy and visual fidelity.

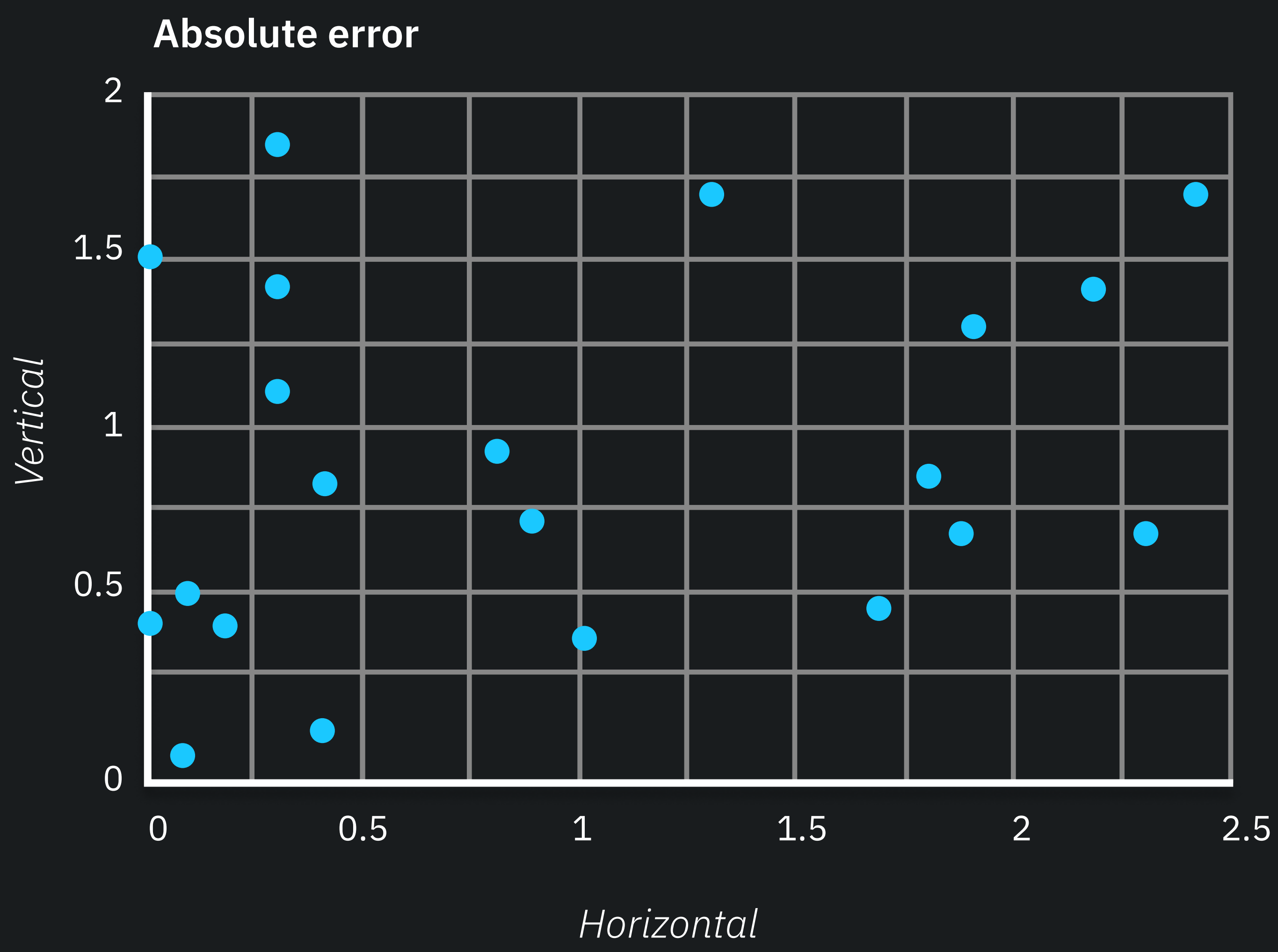
RECONSTRUCTION ACCURACY

A comprehensive assessment was conducted to evaluate the accuracy of the acquired mesh, specifically emphasizing the diversity of face shapes and environmental conditions. This evaluation involved the application of reference marks on individuals' heads with precise caliper measurements and photogrammetry as the ground truth. Several measurements were taken for each individual in order to assess the error in both horizontal and vertical directions.



In both instances, the absolute error falls within the range of 2.5mm, with slightly larger deviations in the horizontal direction. The latter is particularly susceptible to the reconstruction method, as it aggregates errors from the fusion of individual point clouds.

Still, it is noteworthy that the error remains relatively small, comparable in magnitude to the error exhibited by a single point cloud generated by the mobile device, thereby demonstrating the effectiveness of the reconstruction method in achieving accurate results.



MODEL ADVANTAGES

The obtained model is structured in a way which makes it possible to reference points or areas by using model space coordinates.

This makes designing perfectly fitting accessories much easier compared to using a typical unstructured mesh.

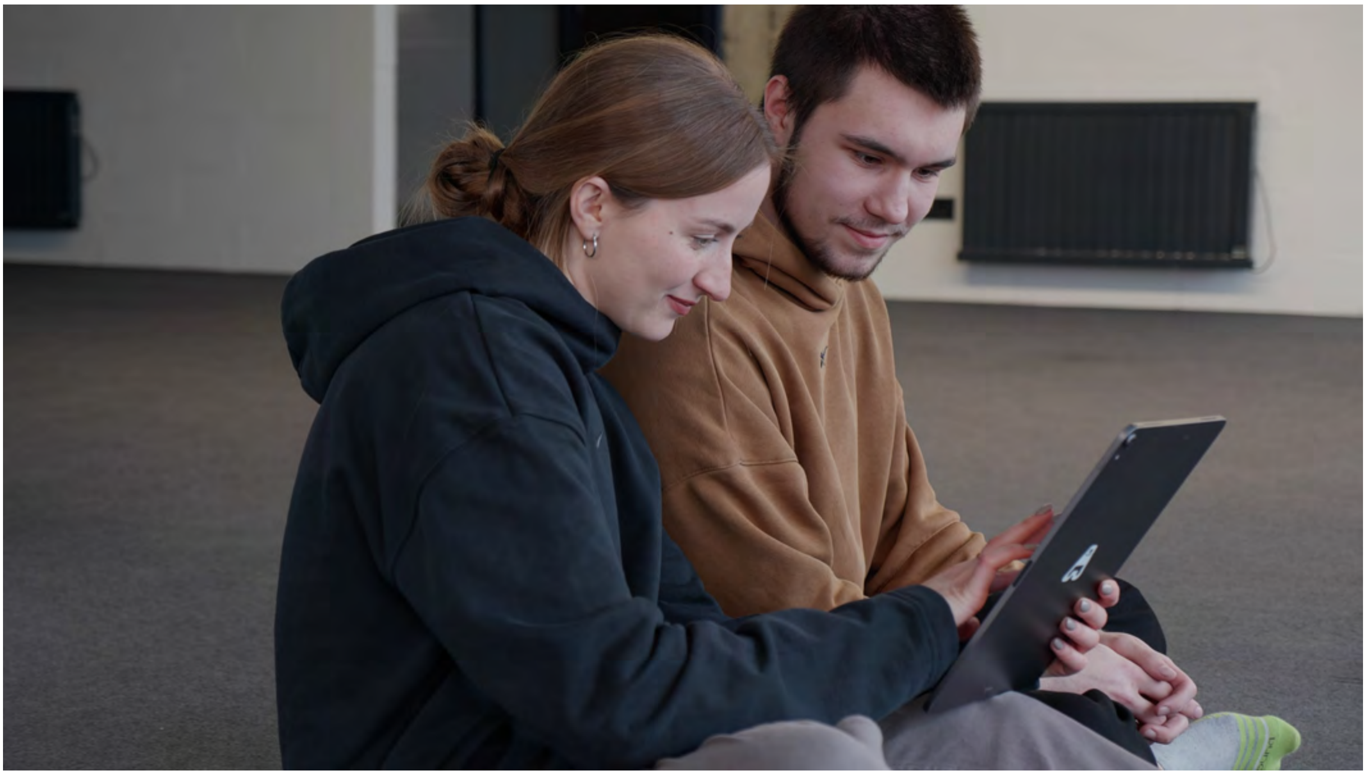
In addition, our approach can also build the model at a desired level of detail from the raw data. We do not rely on model subdivision or reduction to fit a particular vertex budget.

Final Thoughts

05

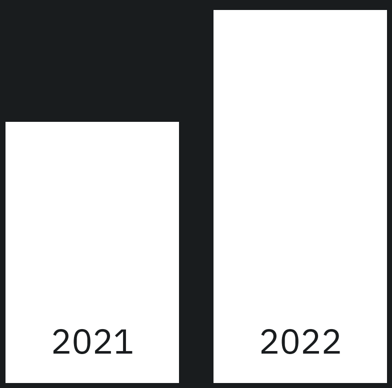
To ensure our place as innovators in the industry, and to provide the most advanced technology to our partners, we have embarked on a mission to create an accurate, fast, and accessible 3D face reconstruction technology using a mobile device. This activity, and the accompanying whitepaper that tracks it, demonstrate our commitment to delivering reliable technology solutions through rigorous research, experimentation and testing. We have shared our experiences, challenges, results and innovative approach in the realm of 3D face reconstruction.

Not only do we fully appreciate the diversity of our clients and respond with advanced solutions, we go out of our way to ensure their expectations are met, and more so. We focus all our expertise on helping them materialize their ideas no matter how forward thinking they may be, or how roughly sketched out they are at the time of first contact. By utilizing the agilest of methodologies combined with the most diverse of XR specialists, our engineering process is fully transparent, including crystal clear visibility from the release roadmap down to daily activities.



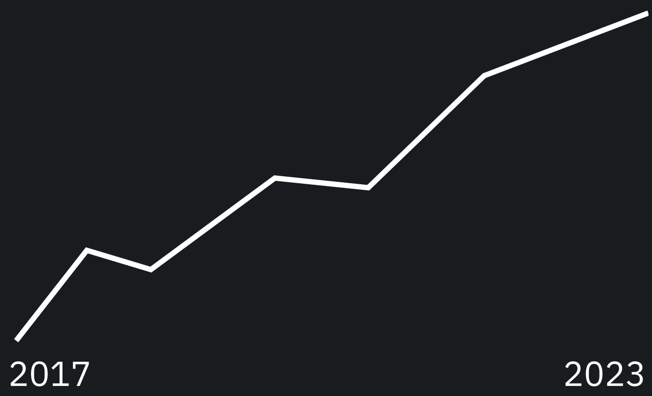
About tsukat

We are a multidisciplinary team of XR experts, passionate about cutting-edge interactive, VR, AR, AI technologies and their massive impact on the modern world. The practical experience of our tech leads in this field exceeds 10 years, and during this time, they have been diligently working with projects of various complexity and duration. This has ultimately allowed us to permanently involve and use next-generation technologies for the needs of our clients.



60+

Projects Completed in 2022



80+

In-house talents

REVIEWED ON

Clutch

4.9/5

Star Rating Based on
Clutch reviews

How we different:

IDEA IMPLEMENTATION	No matter if you have just an idea or a detailed plan we can implement it for you
NARROW FOCUS	Focused solely on VR, AR, AI technologies
TOP 10 ON CLUTCH	Top 10 Virtual Reality Companies and Augmented Reality Developers by Clutch
KEEP DEADLINES	Proven accuracy in estimation
PROFICIENCY AND TRANSPARENCY	By utilizing agile methodologies, the engineering process is fully transparent. This includes visibility from the release roadmap down to daily activities.

READY TO TURN YOUR IDEAS INTO POWERFUL TOOLS THAT ENHANCE YOUR BUSINESS?

Connect with tsukat to unlock the potential of immersive technologies and discover how they can transform your business today.

CONTACT US

hello@tsukat.com

+17866001713

4300 Biscayne blvd, Suite 203, Miami, Florida 33137, USA.

www.tsukat.com

tskt