UNIVERSITY OF TARTU
Institute of Computer Science
Software Engineering Curriculum

Vladyslav Kupriienko

Glyptics Portrait Generator

Master's Thesis (30 ECTS)

Supervisor: Raimond-Hendrik Tunnel, MSc

TARTU 2020

Glyptics Portrait Generator

Abstract:

Glyptics is the art of engraving on precious and semi-precious gems. The resulting engraved gems, also called *intaglios* or *cameos,* were considered a luxury in a variety of ancient civilizations. The University of Tartu Art Museum planned an exhibition, dedicated to that art. The Museum has requested the development of an interactive exhibit for its visitors to create virtual engraved gems with their faces on them.

The main result of this thesis is the development of an interactive exhibit that lets visitors of the University of Tartu Art Museum or any users with web-cameras generate a virtual model of an engraved gem with their face on it. This generator was being developed with a stationary exhibit with a mounted depth camera in mind, but due to the COVID-19 pandemic, it was later repurposed to work with any camera even from home.

The thesis focuses on the development of the algorithm, that produces a 3D rendering of one's face in the glyptic style. The process consists of two main steps: 1) creating a 3D reconstruction of the user's face, 2) rendering it as an engraved gem.

The thesis also covers the research of properties of different materials, used for engraved gems, for their representation by the means of computer graphics. A preliminary user testing took place in January, which demonstrated visitors' interest in the development. This is covered and analysed in this thesis for understanding the main problems users had during the testing and consequently proposing solutions to tackle them.

Keywords:

Computer graphics, glyptics, mesh reconstruction, facial reconstruction, facial landmarks, morphable face fitting, depth camera, RGB camera, rendering, PBR, gems, optical effects

CERCS: P170 Computer science, numerical analysis, systems, control

Gemmipildi generaator

Lühikokkuvõte:

Kivilõikekunstiks nimetatakse gemmide ehk vääris- ja poolvääriskividele graveerimise tehnikat. Tulemuseks saadud gemmid, mida nimetatakse ka kamee või intaljo, olid luksusesemed paljudes vana aja kultuurides. Tartu Ülikooli muuseum planeeris sellele pühendatud näitust. Muuseum soovis näitusele sellist interaktiivset eksponaati, kus külastajad saaksid luua enda nägudest virtuaalseid gemmipilte.

Käesoleva lõputöö põhiline tulemus on sellise interaktiivse eksponaadi arendus, mis laseb Tartu Ülikooli muuseumi külastajatel või ka teistel veebikaamerat omavatel kasutajatel genereerida oma näost vrituaalse graveeritud gemmi. Sede gemmipildi generaatorit arendatati statsionaarse eksponaadina ja sisaldas seinale paigaldatud sügavuskaamerat, kuid COVID-19 pandeemia tõttu tehti rakendus ümber töötamaks kasvõi kodust ja ükskõik millise kaameraga.

Lõputöö keskendub sellise algoritmi arendamisele, mis renderdab inimese nägu kivilõikekunsti stiilis. See protsess koosneb kahest sammust: 1) kasutaja näost virtuaalse 3D mudeli loomine, 2) selle renderdamine graveeritud gemmina.

Lõputöö käsitleb ka erinevate kivilõikekunstis kasutatud gemmide erinevaid materjaliomadusi eesmärgil neid arvutigraafikaga representeerida. Esialgne kasutajatestimine tehti jaanuaris ja see kinnitas külaliste huvi eksponaadi arendusse. Seda testimist on lõputöös käsitletud ja analüüsitud, et välja selgitada põhilised kasutajatel esinenud probleemid ning seejärel pakkuda välja võimalusi nende lahendamiseks.

Võtmesõnad:

Arvutigraafika, kivilõikekunst, võrestiku rekonstrueerimine, näo rekonstrueerimine, näo tähised, morfitava näo sobitamine, sügavuskaamera, RGB kaamera, renderdamine, füüsikal põhinev renderdamine, gemmid, optilised efektid

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimiteooria)

Table of Contents

1 Introduction

Glyptics, or glyptic art, is an ancient art technique that produces engraved (carved) portraits or other images in precious and semi-precious stones. While originally appearing in the Near East, this technique achieved its greatest form in Ancient Greece and the Roman Empire[1]. These engraved gems were mostly considered as counterfeit-resilient signatures in the ancient cultures, as well as just expensive pieces of jewellery. Additionally, in the Middle Ages, such gems were sometimes used in a religious context, also keeping their precious collectable status. The Renaissance period revived interest to glyptic engraving, as modern artists of such technique emerged. Venice is considered to be the centre of engraved gems production in the Renaissance. Although, a noticeable amount of the artists of that period remains unknown due to their works being presented as the actual ancient pieces of glyptic art [1]. Glyptic art was created from a variety of materials, including amethyst, garnet, rock crystal, topaz, beryl, agates, carnelian and sard [2]. Examples of glyptic art pieces are presented in Figure 1.

Figure 1. Engraved portrait of Aurelian carved in amethyst, Roman Empire, 260–280 AD (A); carnelian ring stone with a portrait of Octavian, Roman Empire, mid-1st century BC (B); garnet finger ring with Danae, Ancient Greece, 3rd century BC (C); Gonzaga Cameo (three-layered sardonyx), Hellenistic Egypt, 3rd century BC (D)

Glyptic art is the main topic of the University of Tartu Art Museum's summer exhibition. Therefore, developing a Glyptic Portrait Generator (GPG) exhibit is a perfect possibility to create a modern fusion of this art and computer graphics, as well as to provide an interactive and entertaining way to create something memorable for visitors on their own.

To generate glyptic portraits in real-time, the software developed in the thesis solves different problems from the fields of computer vision and computer graphics. More

[1] https://en.wikipedia.org/wiki/Engraved_gem

precisely, this includes facial landmark detection, 3D reconstruction, and rendering using gem materials. These solutions are also coupled with the overall configuration of the exhibit.

One of the main objectives is to choose the appropriate facial reconstruction method, considering different available techniques, their complexity, implementation costs and produced model quality. Chapter 2 describes the different facial reconstruction techniques that were considered during the development of the Glyptics Portrait Generator. The chapter compares the possibilities and restrictions of these technologies and also justifies the choice of hardware for this exhibit. The implementation of the exhibit, including the algorithm, the physical configuration, the user experience (UX) and user interface (UI) design, and the software architecture, is described in Chapter 3. Chapter 4 provides the assessment of the developed solution in terms of recognisability, similarity to the intended style and usability. To judge the usability of the product, two usability tests were conducted: one during the Delta building opening event (29.01.2020) and the second one as a set of private sessions during the final stages of development. Chapter 5 provides the conclusion to the thesis and discourse about the future development possibilities.

This thesis is also followed by the Appendix, which contains:

1) the list of abbreviations used throughout the thesis;
2) the Glossary with terms and definitions;
3) the results and the screenshots of tests;
4) the overview of the accompanying files;
5) the License.

2 3D Facial Reconstruction

This chapter gives a brief overview of the main techniques used for 3D facial reconstruction. It is a process of recreation of one's face as a 3D mesh (model) through computer vision. This step is essential to the Glyptics Portrait Generator, as it creates the computer graphics representation of a user's face, which can be later rendered, or visualized, to look like an engraved gem. The literature review for this chapter was conducted following the procedure defined by Kitchenham [3].

2.1 Facial Reconstruction Techniques

A vast majority of computer vision problems can be mapped to the corresponding processes happening in the human brain. One of such problems is the ability to reconstruct a 3D representation of the object.

Depth perception happens constantly in the human brain. One can perceive the shape and depth information about the object by looking at a 2D image, even without any prior knowledge about the object. The ability to reproduce such a process in computer vision is a complex problem, still being actively researched. Widanagamaachchi & Dharmaratne in their work [4] categorize existing 3D reconstruction techniques into three broad categories: image-based (from several 2D images to 3D object), hybrid image-based and 3D scanning.

2.1.1 Image-based Techniques

For image-based techniques, three-dimensional facial reconstruction relies on the usage of several 2D images of the face from different angles or with different lighting conditions. The relation between points on the face is calculated from those images, creating a 3D point cloud, which can be itself transformed into a 3D model of the face.

Different setups can be used for this technique. Several approaches, like the ones proposed by Liu et al. [5] and Booth et al. [6], generate the point cloud or fit the morphable 3D model gradually, while the camera or the face itself is in motion. For example, facial features can be detected, and by tracking them while the person is slowly turning their head sideways, their relative positions are determined. However, such a method might produce a defective model due to changes in facial expression, lighting conditions or other factors that affect the scanning. Such problems are highlighted in comparison sections of Liu et al. [5] and Booth et al. [6] to other contemporary methods.

Other approaches rely on several still images being taken almost or completely simultaneously. For instance, Kumar et al. [7] propose the setup with 5 cameras (Figure 2), placed in different angles on the single horizontal plane. One is frontal (0°), others are shifted to -40°, -20°, 20° and 40° relative to the face that is being scanned.

Figure 2. 5-camera setup proposed by Kumar et al.

These cameras take a photo simultaneously. From these photos, the face is located, then facial features are extracted, tracked, their positions triangulated and transformed into a point cloud. Noise is filtered and the cloud is transformed into a surface.

Another method, proposed by Zhang et al. [8], incorporates the usage of four white LEDs, fixed to the camera in a vertically aligned cross shape 330 mm (in each directional) from the lens. This setup is schematically represented in Figure 3.

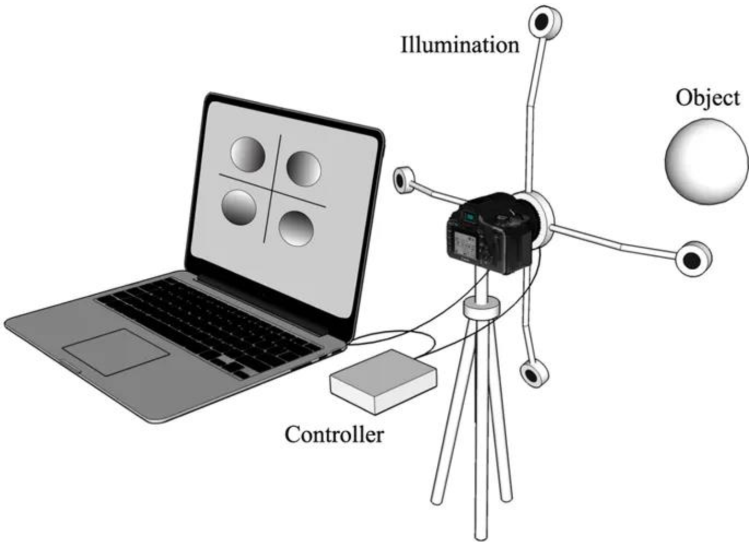


Figure 3. Schematic diagram of the setup proposed by Zhang et al.

The system successively activates each of the four LEDs, taking a photo each time. This results in a set of four photos of the same object (human face, in this case) with different shading. These images go through the segmentation process (to distinguish background from the object itself), then the feature vectors are extracted. They are turned into a normal map, then height map and eventually the texture is applied to the heightmap, producing a 3D rendering of the object.

2.1.2 Hybrid Techniques

Hybrid techniques combine 3D and 2D information to produce a final model. While 2D images can still be obtained with previously described techniques, in hybrid techniques the prior 3D model acts as the 3D information, that is morphed and shaped to represent the desired object. The human brain acts similarly – it has a reference material in terms of shapes of known objects. A human knows what an average human face looks like and is able to imagine it. 2D images (information from the eyes) are projected on that prior to produce an actual representation of the object. Furthermore, these techniques often use deep learning at one or several steps of the algorithm (e.g., for feature detection).

Jiang et al. [9] use such an approach in their method, that relies only on one 2D reference image, and a database of face meshes with different shapes, facial features, and expressions. The method finds the most suitable mesh by fitting the parametric model to the input image, aligning 68 detected facial landmarks. This step produces the model that has the overall general look of the desired face. The second step lies in detecting medium-scale facial features and applying them to the mesh. In the end, illumination parameters and the enhanced mesh are used to determine the finest details of the face, applying them to the model. The final result captures even the most sublime facial features of the original (**Figure 4**).

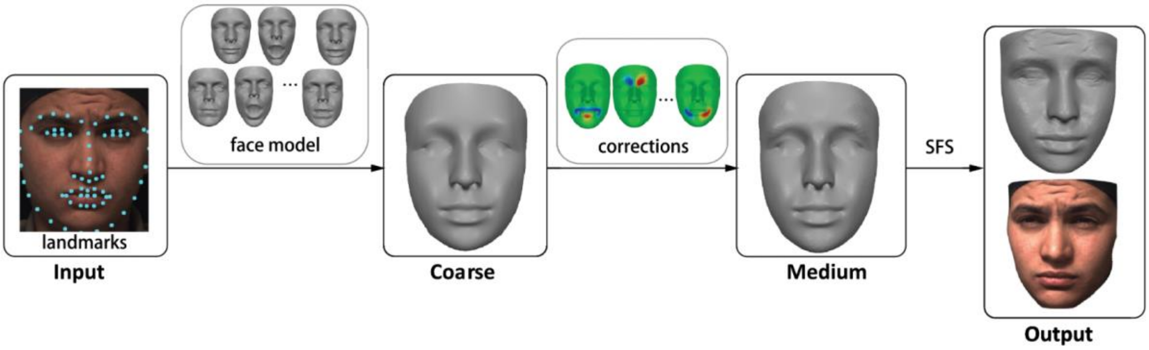


Figure 4. The algorithm of the method, proposed by Jiang et al. [9], with intermediate results

In general, the hybrid technique produces anatomically correct faces because they are based on a prior model. However, the level of detail depends on the implementation, and in the general case, the result might be recognizable, but not completely accurate.

2.1.3 3D Scanning

Techniques of this category explicitly rely on the usage of depth cameras. Such cameras are capable of acquiring depth information in addition to RGB data. However, there are different classes of such devices. The industrial level depth sensors provide high-resolution and high-quality depth maps, but it comes with a drawback of high cost, performance demands, and overall spatial bulkiness. On the other end of this market are much smaller and faster cameras that capture information with much lower precision and resolution [10].

High-end depth sensors usually have the capability and toolkit to directly produce a 3D rendering of an object. Measurements from such devices are often used as "ground truth" in numerous papers researching faster and cheaper approaches to 3D scanning. Actually, in most cases, the main goal of the research and experimentation is to approximate the result to such ground truth with other methods or other equipment.

Figure 5. Microsoft Kinect camera for Windows

As mentioned above, a variety of papers propose different solutions to the problem of accurate 3D face scanning using cheap cameras or different conditions (lack of user cooperation, shading, poor lighting). For instance, Bondi et al. [10] propose a method of face scanning using a Microsoft Kinect[2] camera (**Figure 5**), which acquires a set of low-resolution images over time with 30 FPS (frames per second) frequency. The author calls such a camera 4D, implying that the fourth dimension is time, as the complete picture is formed over time, and not instantly. An example of the frames taken during the process is shown in Figure 6.

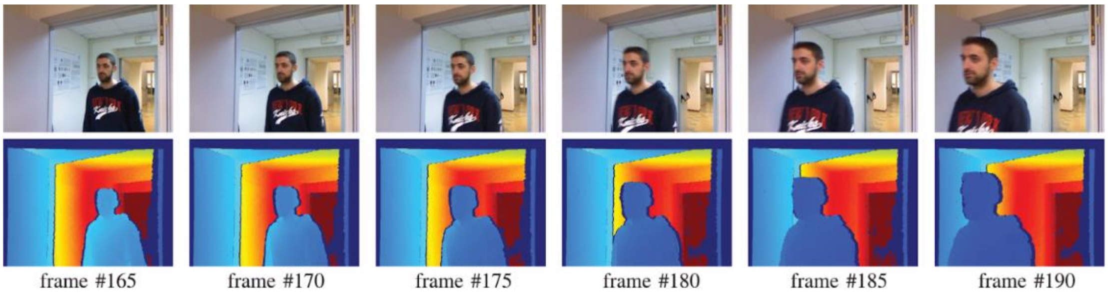


Figure 6. Sample RGB and depth frames from the un-cooperative acquisition process, proposed by Bondi et al.

[2] https://developer.microsoft.com/en-us/windows/kinect/

Such an approach reconstructs a 3D model with high vertex count from a sequence of lower resolution frames, both in cooperative and non-cooperative contexts. The face of the person is detected in every frame, resulting point clouds are joined into a single high-resolution cumulative point cloud, that is then processed to reduce noise.

Hernandez et al. use a similar approach [11], but in their method, each frame is projected to an imaginary cylinder that surrounds the head of the person. This allows for easier processing of each frame and for the endless addition of new frames to the cumulative model. Furthermore, the algorithm of transforming depth frames to 3D mesh is simplified by creating trigonal polygons among the neighbouring pixels on these cylindrical projections. The entire algorithm is represented in Figure 7.

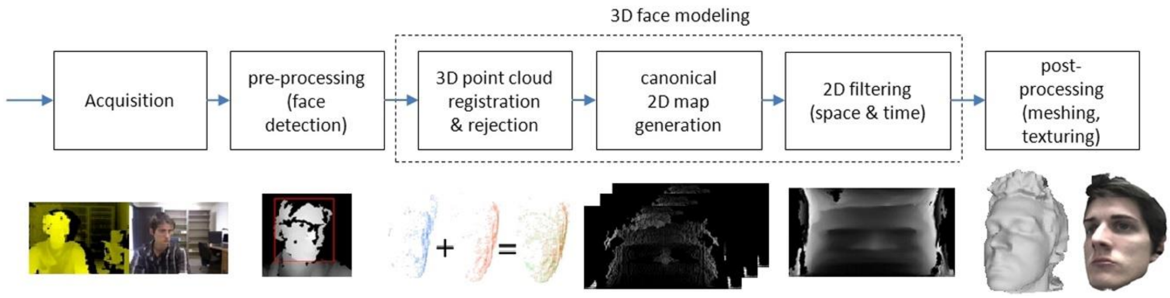


Figure 7. The algorithm of the method, proposed by Hernandez et al. [11]

The main problem with the 3D scanning technique is the topology of the resulting model. It does not follow the topology, expected from the facial mesh[3] (e.g., having more detail around the eyes or the mouth, but less on cheeks and forehead). Bad topology leads to problems in rendering, such as less FPS due to unnecessary high polygon count or lighting problems.

2.2 Comparison of Techniques

To choose the most suitable approach for the implementation of the GPG, a comparison of the reviewed techniques was conducted. More precisely, several methods from each category were selected and evaluated. The characteristics by which the comparison was made are the following:

- price;
- setup complexity;
- resulting model accuracy;

[3] http://wiki.polycount.com/wiki/FaceTopology

- operating area;
- special requirements.

However, it is important to note that the comparison acts as a guideline for help in the selection of a general approach, not the concrete implementation. This is due to these implementations being quite specific for the selected hardware and/or situation. Such a choice only helps with the general direction of the development, without applying unnecessary restrictions or requirements.

From the three aforementioned techniques, six implementations were selected. The 5-camera setup proposed by Kumar et al. [7] and 4-LED setup proposed by Zhang et al. [8] represent the image-based technique. The hybrid technique is illustrated by a single-image to mesh database fitting method by Jiang et al. [9]. Kinect depth sequences method by Bondi et al. [10] and the Low-quality depth stream method by Hernandez et al. [11] fall under the 3D scanning category. Industrial laser scanning also falls into the same category, represented by 3dMDface [11]. The results are reported in Table 1.

The prices are not given directly in the papers: therefore, they are assessed approximately by looking up devices' prices used in said methods. The price column is intended for comparative analysis of different methods rather than a precise measurement of costs. Apart from the specific devices, each method assumes an average PC or laptop, therefore they were not included in the price for comparison sake. The installation costs are also omitted. The overall cost would be about €1000 higher if these exclusions are considered.

Table 1. Comparison of several methods of 3D face reconstruction in terms of price, setup complexity, the resulting model accuracy, and special requirements

Method	Technique	Price	Complexity	Accuracy, RMSE	Requirements
5 cameras	Image-based	€700+[4,5]	High	High, N/A	Precise setup in terms of camera angles
4 LEDs	Image-based	€140[4,6,7]	Medium	Low, 15.6 mm	Specific distance to the object

[4] https://www.amazon.com/Digital-Cameras/b?ie=UTF8&node=281052
[5] Complete Tripods on Amazon
[6] https://store.arduino.cc/arduino-uno-rev3
[7] https://www.luxeonstar.com/white-rebel-leds

Fitting mesh to a 2D image	Hybrid	€50[8]	Low	High, ~1.6 mm	Access to face meshes libraries
Kinect depth sequences	3D scanning	€230[9]	Low	High, 1.5 mm	40-180 cm working area
Low-quality depth stream	3D scanning	€75[10]	Low	High, 1 mm	10s of exposure to the scanner
3dMDface	3D scanning	€1000-20000[11]	Very high	Very high, <0.2 mm	Manual scanning or very spacious setup

The comparison suggests that the best solution in terms of Glyptics Portrait Generator would be the usage of a cheap depth camera in combination with some method that enhances the resolution of the output model, like in methods by Bondi et al. [10] and Hernandez et al. [11]. The complexity of such a method is the lowest – mounted depth camera is the only element responsible for face scanning, the technique does not require the user to do anything specific apart from staying in the working area for some time. The resulting accuracy is high enough for the desired solution, and the costs are the lowest among all reviewed methods.

This approach was chosen initially, and the development was following it as well. However, a new decision had to be made due to the COVID-19 pandemic and the state of emergency in Estonia. As the Art Museum was closed for the duration of the emergency, it was decided that there should be a way to use the GPG for people at home. Given that the vast majority of the users do not possess the depth cameras required for the chosen approach to work, the algorithm had to rely only on data from RGB cameras (i.e. 2D images). Therefore, a solution using a variation of the hybrid technique was developed. Methods of this technique suit the situation the best because they do not require any specific setups or devices (apart from the standard web-camera, which most of the users have[12]). Chapter 4 provides more information about the development and the algorithm.

[8] Webcams on Amazon
[9] https://www.amazon.com/gp/offer-listing/B006UIS53K/ref=dp_olp_all_mbc?ie=UTF8&condition=all
[10] https://store.intelrealsense.com/buy-intel-realsense-depth-camera-sr305.html
[11] https://www.aniwaa.com/product/3d-scanners/3dmd-3dmdface-system/
[12] http://zugara.com/webcam-penetration-rates-adoption

3 Engraved Gem Rendering

To represent engraved gems in computer graphics, their properties have to be researched. Computer graphics heavily employ different approximations to visualize things not only as similar as possible but as efficiently as possible too. The rendering is the second big part of this thesis, which is responsible for the look of the final result. This chapter describes ways of representing different materials in computer graphics, explains the approximations used for simulating the behaviour of light and looks into optical effects, present in gems.

3.1 Material Representation in Computer Graphics

From the inception of computer graphics in their current form, lighting has been constantly researched and different models were developed. Lighting is responsible for how the colour of each pixel is perceived, thus it is essential to computer graphics. The main problem in that area is to achieve a balance between computational complexity and the acceptable degree of approximation to the real world. Modelling the behaviour of light is a difficult task given how many parameters have to be considered to produce a correctly lighted render. This area of research in computer graphics is very tightly coupled with physics.

Photons behave differently when hitting different surfaces because each material has its own distinct set of parameters. To describe material properties in computer graphics, so-called bidirectional distribution functions, or BxDFs are used. It is considered that every material in computer graphics has these functions. Such functions can be measured empirically (with specific devices) or theoretically derived. They describe the behaviour of light when it interacts with the surface they represent. Among these functions are:

- BSSRDF – *bidirectional scattering-surface reflectance distribution function*
- BRDF – *bidirectional reflectance distribution function*
- BTDF – *bidirectional transmittance distribution function*

BSSRDF is used to calculate the amount of reflected light and the point where it exits the object, which may be different from the entrance point due to subsurface scattering – phenomenon, when light bounces inside the material before coming out, thus changing its position and angle. Such behaviour can be seen in materials such as milk, jade, marble, wax or human skin [12].

BRDF is a simplified version of BSSRDF that does not take subsurface scattering into account, assuming that the point where the light hits the surface is the same point that it reflects from. BRDF has arguably the biggest presence on the computer graphics scene. There exist numerous datasets of BRDFs for different materials (e.g., the MERL database, measured in 2003, includes 100 materials). Most of the widely used lighting models (e.g. Lambertian, Phong, Blinn-Phong) are variations of BRDF [13].

BTDF is an extension of BRDF, which describes the light that passes through the surface. It is used much more rarely than BRDF in computer graphics. All three aforementioned functions are visually represented in Figure 8.

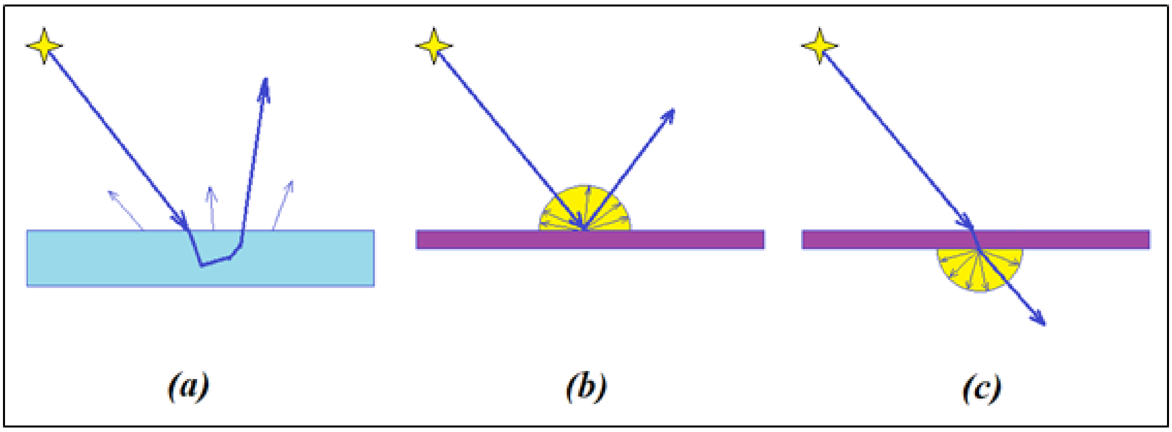


Figure 8. Visual representation of BSSRDF (a), BRDF (b), and BTDF (c)

However, modelling light with its real-world behaviour is an extraordinarily computationally heavy operation. To achieve such effect, light has to be seen as a collection of rays – basically, photon flows. Illumination models, that use such an approach, are called global illumination models. For each ray, all its interactions with objects in the scene have to be simulated: how is it reflected or scattered from the surfaces, what amount of energy is lost, is the refraction happening and how strong is it, etc. As a result, complete accordance with the real world in terms of lighting can be achieved, with all the effects and phenomena caused by it. However, as noted previously, the computational power of modern devices is not sufficient to use such a technique in real-time rendering. Methods that use this approach are still used in several domains, where it is acceptable for the rendering of a single frame to run for hours. There are, however, models that simplify the described process while still keeping the main advantage of being physically correct. Namely, ray tracing, photon mapping, and radiosity are considered optimized global illumination models and are used for real-time rendering coupled with modern high-end graphic processors.

To overcome the problem of insufficient computational power, several lighting models were proposed, that achieve different degrees of approximation to physically correct behaviour, without actually tracing the rays of light. Furthermore, such models usually rely on a much smaller set of parameters. These models are called object-oriented lighting, or local illumination models. Local illumination models only consider the light that comes to the surface from the light sources, without taking into account other objects in the scene. Illumination models, mentioned above – Lambertian, Phong and Blinn-Phong – are instances of the most used local illumination models. Proposed in the 1970s, they are still widely used due to being easy to compute. However, the main drawback of such models is the lack of ability to represent more advanced lighting effects on their own (reflectance, subsurface scattering, Fresnel effect, etc.) [14].

In the context of the Glyptics Portrait Generator, local illumination models may be insufficient to represent materials used for engraved gems. Such materials as marble, jade or amber have partial translucency, and therefore cannot be correctly represented by models designed specifically for opaque surfaces. On the other hand, using global illumination might be too performance-heavy for the exhibition of such a scale with the provided resources. Therefore, to achieve the desired result with the given limitations, a combination of both models might be constructed, mostly relying on local illumination models, but borrowing properties from global, to represent properties, present in precious materials.

3.2 Optical Effects in Gems

Precious and semiprecious gems obtained their status of highly sought after valuables mostly because of their appearance. From the earliest civilization stages, humans were captivated by the looks of gems. Whole professions formed around gems, like jewellers, gemologists, and appraisers.

Gems can display several optical effects, that capture the eye of humans. Their structure causes the light, that falls on the gems, to behave quite unusually. Basic optical effects such as refraction, reflection and diffraction are much more prominent in gems and are often combined in very unique ways. Some gems possess visual effects, that are simply not present in any other material.

The most prominent results of gem production are cut gems like diamonds, rubies or sapphires. They do not have any special effects on their own, but their internal structure combined with proper cuts causes such effects as *dichroism* (multi-colouring), *brilliance* (reflecting light from inside the gem, rather than from the surface) and *fire* (splitting light into colours of the spectrum). However, this is mostly irrelevant to the field of glyptic art, as these effects are not present directly in engraved gems. Despite that, such effects are an interesting computer graphics topic, and implementing them would allow seeing, how the engraved gems would look like with these effects.

Apart from the effects, mentioned earlier, there exist more specific more optical phenomena, observed in different gems. There is no consensus on the classification of these effects, but the most common ones distinguish the following:

- Adularescence
- Chatoyancy
- Asterism
- Aventurescence
- Colour change
- Iridescence
- Play of colour
- Pleochroism

Adularescence is the effect, most notable present in moonstones (Figure 9). A gem with such effect creates the illusion of having an internal light source just below the surface. The colour of such "light" ranges from milky white to blueish. Such an effect is caused by light refracting and reflecting from the lamellar structure of the gem. It means, that the internal structure of such gem consists of very thin planar surfaces, located close to each other. The light partially reflects from the top layer, and partially travels through, refracting in the process. It is then partially reflected from the next layer and so on. The resulting reflected rays interfere, changing the wavelength to the blue part of the spectrum, thus creating the effect. A possible solution to reproduce this effect in computer graphics is to put a scaled-down copy of the mesh inside the original one with a glossy surface of the required colour and make the original mesh transparent, with glossy reflectivity.

Figure 9. Moonstones with adularescence

Chatoyancy, or “cat’s eye” is the effect, when the light forms a single bright, mobile reflective line on the surface of the gem (Figure 10). This phenomenon is most prominent when the gem is cut *en cabochon* (i.e. rounded, not faceted). This is caused by the fibrous structure of a material or fibrous inclusions/cavities. In simpler terms, the internal structure of such gems consists of numerous tiny tubes or fibres, aligned in the same direction. The light is reflected from these fibres across the same angle, creating a line, perpendicular to the structural alignment. One way to represent such an effect in the computer graphics is to simulate the internal structure of the gem, e.g. via the hair particle system. This would result in the desired optical effect.

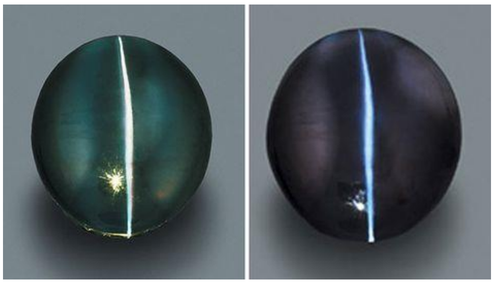


Figure 10. Alexandrite with chatoyancy (colour change effect is also present)

Asterism is an effect when the reflected and refracted light forms a star on the surface of the gem (Figure 11). This star can consist of 4, 6, 8 or even sometimes 12 rays. It also requires the *en cabochon* cut to be noticeable. Basically, this is a more complex version of chatoyancy, where the fibres are aligned not to a single axis, but several. Asterism can be seen as several combined chatoyancy effects, centred around the crystal axis (also called the *c*-axis). The amount of rays depends on the number of axes present inside the gem. As for the computer graphics approximation, a special “light multiplication” texture would possibly suffice. This would be possible because the star’s location is permanent and the only thing that changes is its intensity (based on the amount of incoming light).

Figure 11. Asterism in rose quartz star

Aventurescence is a pattern of brilliant flashes and colour spots inside the gem. In a way, it resembles glitter very much (Figure 12). The physical reasoning behind this is actually identical to the glitter – tiny flat metallic inclusions, that reflect light, each under different angles. If the inclusions are numerous, the whole gem’s colour is affected. This

Figure 12. Aventurescence in goldstone

effect is most prominent in the synthetically created gem called goldstone. For this effect, simulating the structure might be too computationally heavy. A much simpler method was proposed by Weidlich and Wilkie [15], which uses several layers of Voronoi noise textures with different granularity and weights.

Colour change is the ability of the gem to change the colour depending on the nature of the light (e.g. natural sunlight vs electric lighting). For example, alexandrite, depicted in Figure 13, can have green tones in natural light and red tones in electric lighting conditions. The physical reasoning behind this phenomenon lies in the nature of the light, emitted by different sources. Every source emits light, made up of different wavelengths in different proportions. Gems with colour change have several "absorbance windows", meaning that they absorb different wavelengths. Thus, when the light has more of one colour, it becomes the dominant colour of the gem. To represent this in computer graphics, one might add an explicit parameter to the light source, describing its nature. Depending on that value, the colour change gem will alternate its colours, given that some kind of colour change function was provided.

Figure 13. Alexandrite under sunlight (left) and under electric lighting (right)

Iridescence is a name for the rainbow-like effect on the surface. It is not exclusive to gems – any material can be called iridescent if it demonstrates said property. The iridescent gems might have a full spectrum of colours, or only some of them due to interference (Figure 14). This effect is caused by the thin-film-like structure of the gems. Such thin film is the reason for the iridescence in general – e.g., "rainbows" on water from the spilt gas or soap bubbles. This thin film causes different attenuation for different light wavelengths, causing diffraction. Not all iridescent gems have the same structure, but the general reason is the same for all of them. For that particular phenomenon, there exists a wide variety of implementations for Unity, Blender or Unreal Engine. There is no single way to achieve it: for instance, in Blender, it is most frequently achieved by combining the Principled BSDF and a Color Ramp with a rainbow gradient.

Figure 14. Iridescence in blue part of the spectrum in labradorite

Play of colour (not to be confused with colour change) is a property of the gems to produce rainbow-like flashes of colour that change with the angle of observation (Figure 15). Some gemologists put this effect under the iridescence, but the physical nature of this effect is quite different. This phenomenon is found exclusively in precious opals. It happens due to the opals' internal structure. They consist of stacked silica spheres. If these spheres are uniform in size and shape, they will diffract light. This diffraction creates the play of colour. The size of the spheres affects the produced colour – smaller spheres produce blue and violet, bigger – red and orange. Currently, there are no implementations in computer graphics, that approximate this effect well-enough, possibly due to the rarity and specificity of it. Existing solutions use the Voronoi noise, similar to aventurescence. However, because of that, it looks just like a multicoloured aventurescence, while in reality, the effect is much more prominent and rich. Sometimes emissive colour is used (meaning that the material emits the light on its own), which is physically incorrect, but it brings the final result somewhat closer to the real-life look.

Figure 15. Precious opal with play of colour

The last effect from the list, *pleochroism*, is a phenomenon which causes the gems to appear to have different colours when observed from different angles (Figure 16). It is different from colour change, as it depends on the angle, and not the nature of the light source. The only gems that are pleochroic, must have birefringence (i.e. light is split into two separate rays inside the gem during refraction). If the split rays have different wavelengths, the gem demonstrates pleochroism. These gems have different absorbance spectra depending on the light directing relative to the crystal axis. Simulating the internal structure of pleochroic gems is impossible, as it is too granular. However, some approximations take pleochroism into account. One of such methods is the algorithm of rendering faceted gems, proposed by Guy and Soler [16].

Figure 16. Tourmaline with pleochroism

Even though the engraved gems usually do not demonstrate such behaviour, adding it might produce interesting-looking results. The main goal of the thesis is to create something visually fascinating and memorable, and these optical effects might prove very useful in achieving that.

4 Implementation

The main goal of the development process is the creation of an exhibit, that produces the visual This chapter covers the structure of the Glyptics Portrait Generator. It describes the process of its creation, the technologies and algorithms used, decisions made and achieved results.

4.1 Algorithm

The algorithm of the Glyptics Portrait Generator consists of two main parts (or phases), which can, in turn, also be divided into subparts. These main parts are *facial reconstruction* and *gem rendering*.

The start of the user's interaction with the system marks the beginning of the facial reconstruction phase. Its goal is to produce a mesh of the user's face to be later processed and rendered in the second phase. The facial reconstruction algorithm can be subdivided into the following parts:

1) frame acquisition;
2) landmark detection;
3) facial mesh morphing.

The program constantly goes through all these parts during the facial reconstruction phase, because one of the goals of the GPG was to provide continuous real-time preview of the result to the user. Thus the morphed 3D model, rendered as a gem, is always updated for the current user.

During the frame acquisition phase, the program simply captures a frame from the video stream of the connected video device. These frames can come in different formats depending on the device. For example, Intel RealSense camera produces RGB frames as objects of its own `rs2::video_frame` class, a part of the Intel RealSense SDK[13] (**Figure 17**). To solve that problem, the OpenCV library acts as an intermediary between the devices and the rest of the program.

Figure 17. Intel RealSense logo

[13] https://www.intelrealsense.com/sdk-2/

All of the frames are converted into the OpenCV's `cv::Mat` class[14] and passed on. This step is encapsulated in the FrameCapturer module of the application.

The landmark detection part is implemented using dlib library[15]. This library provides an out of the box solution for detecting 68 iBUG landmarks[16] on the images (Figure 18). This functionality is encapsulated in the FacialDetector module of the GPG. This is a widely accepted format, compatible with many other computer vision libraries. The iBUG group also used their solution on several popular facial image datasets, such as 300-W, LFPW and HELEN.

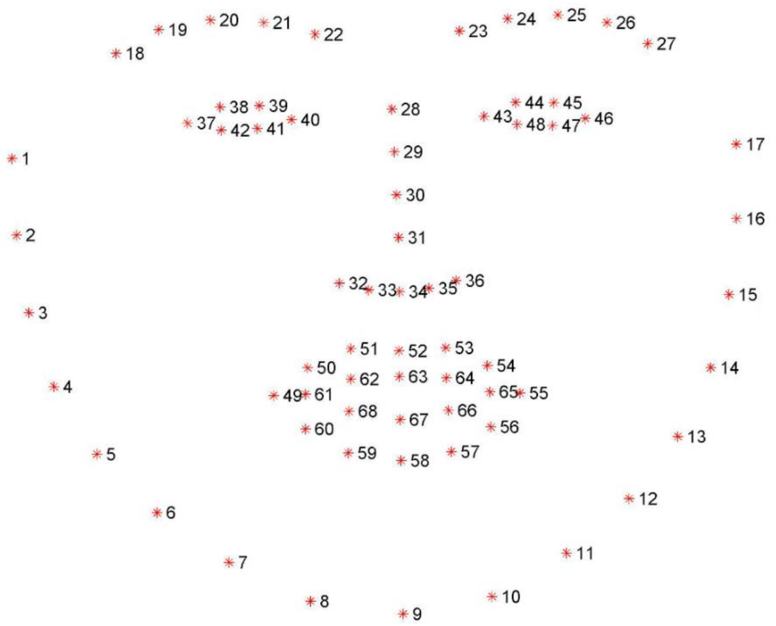


Figure 18. The 68 landmarks used for the iBUG annotations

Per dlib's official documentation, the landmark detection algorithm uses the classic Histogram of Oriented Gradients feature[17] combined with a linear classifier[18], an image pyramid[19] and a sliding window detection scheme[20]. The detection algorithm is quite fast (~60 ms for a 640x480 image) and accurate. The visualization of the detected landmarks (represented as green lines) can be seen in Figure 19.

---

14 https://docs.opencv.org/3.4/d3/d63/classcv_1_1Mat.html
15 http://dlib.net/
16 https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/
17 https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients
18 https://en.wikipedia.org/wiki/Linear_classifier
19 https://en.wikipedia.org/wiki/Pyramid_(image_processing)
20 http://dlib.net/face_landmark_detection.py.html

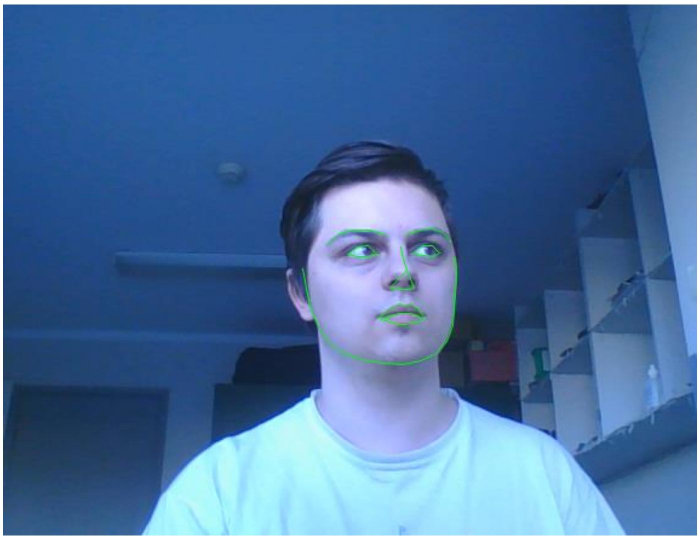


Figure 19. Visualization of facial landmarks detected by dlib

These landmarks are passed to the next step in the algorithm, encapsulated in the FacialMorpher module. The backbone of this part is the eos library[21]. It is a header-only library, designed for fitting different 3D morphable face models. Out-of-the-box it supports the Surrey Face Model (SFM)[22], Basel Face Model (BFM)[23] 2009 and 2017, and 4D Face Model[24] (4DFM) – model, developed by the author of the library. The Glyptics Portrait Generator supports SFM (medium resolution) and BFM2017, however, the latter comes without the blendshapes. Blendshapes are responsible for the representation of different facial expressions. Thus, using the SFM it is possible to represent them, but the BFM assumes that the facial deformations, caused by facial expressions, are just a part of the face and morphs the face accordingly.

Apart from that, SFM represents only the face, while BFM has almost all of the head (except for the top) and a part of a neck. The number of polygons is also different – 6,736 in SFM and 105,694 in BFM. This might seem like a clear advantage of the BFM, but it is not necessarily true. The level of detail, provided by the SFM, is enough for the GPG, and having that much more polygons in the BFM only introduces performance issues. For comparison, the full cycle of facial reconstruction with the SFM takes around 270 ms (170 ms in the morphing phase). The cycle with the BFM takes around 1150 ms (1050 ms in the

[21] https://github.com/patrikhuber/eos
[22] https://cvssp.org/faceweb/3dmm/facemodels/
[23] https://faces.dmi.unibas.ch/bfm/bfm2019.html
[24] https://www.4dface.io/4dfm/

morphing phase). This difference alone introduces a big difference in FPS, and the rendering times also differ, albeit not so drastically.

The BFM also seems to be much more rigid than SFM, meaning that it tends to change much more slightly. This affects the level of recognizability too much – all faces start looking almost the same. However, there is a `lambda` parameter for the fitting phase, which acts as a regularization parameter. It makes the result smoother, more symmetrical and less affected by outliers or errors in the landmark detection phase. Still, changing this parameter seems to affect SFM much more, than BFM, as seen in Figure 20 and Figure 21. The lack of facial expressions in the second model is due to the lack of blendshapes.

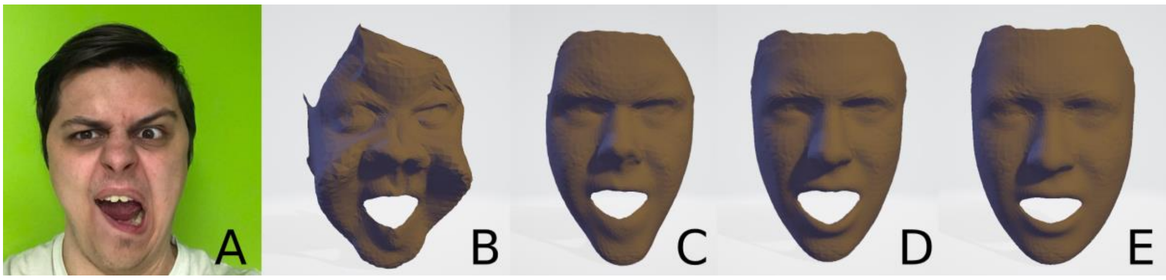


Figure 20. SFM model from single image with different lambda parameters: input photo (A), `lambda = 0.05` (B), `lambda = 0.5` (C), `lambda = 5.0` (D), `lambda = 50.0` (E)

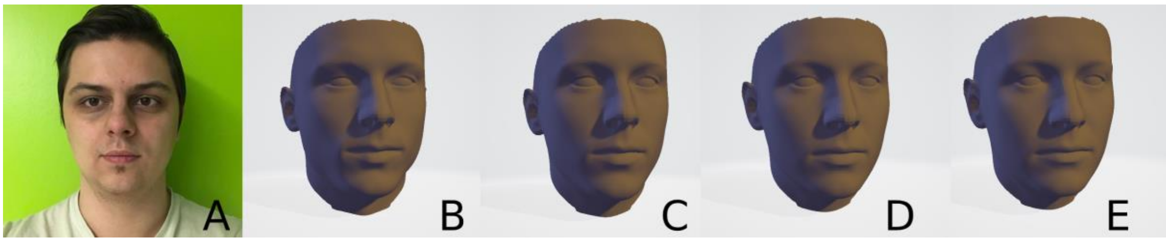


Figure 21. BFM model from single image with different lambda parameters: input photo (A), `lambda = 0.05` (B), `lambda = 0.5` (C), `lambda = 5.0` (D), `lambda = 50.0` (E)

As seen from the figures, the higher the lambda parameter is, the closer the result resembles the default model. However, a too low value causes deformations and unwanted artefacts in the SFM. The BFM barely changes even with a very low value. Because of that, all the facial models, produced by it, look almost the same. Therefore, the SFM model and a `lambda` value of 5.0 were chosen as the best balance between the level of detail and the visual appeal.

The resulting mesh of the user's face is then passed on to the OgreApp module. Its name comes from the graphics engine used for the rendering – OGRE 3D[25] (Figure 22). It is an open-source graphics engine, developed from 2001. The name stands for **O**bject-Oriented **G**raphics **R**endering **E**ngine. Currently, it is divided into two branches – Ogre1 and Ogre Next, or Ogre 2.0+. They are supported in parallel, but Ogre Next does not provide all the features, available in Ogre1, and is considered less stable. Its main advantage over the older branch is the performance boost in the case of having many (order of 10,000) objects per frame. However, this is irrelevant in the context of the GPG – the resulting frame consists of only several objects. Considering all these points, Ogre1 (more specifically, version 1.12.3) was used in the thesis.

Figure 22. Ogre3D graphics engine logo

The OgreApp module receives the mesh in a raw format, with vertices represented as an indexed array, and the faces (triangles) as a matrix of size $n * 3$, where n is the number of faces. The application then checks if an object with such name is already present on the scene. If not, the mesh is added to the scene via `Ogre::ManualObject` class. If it is present, an update function is triggered, which does not re-add the object to the scene, but rather updates the positions of the vertices. To get a good-looking result, a predefined mesh of an ancient Greek statue's head without a face is added to the scene. This solves the problem of the SFM mesh not having anything apart from the face itself. The meshes are then scaled to make them flatter (like on the actual engraved gems) and attached to another additional mesh, that has the general shape of an engraved gem. Finally, a PBS material[26] is applied to all the objects in the scene, creating the look of a glyptic art piece.

4.2 Physical Configuration

The initial plan for the exhibit was to mount an Intel RealSense SR305[27] camera (**Figure 23**) and a touch-screen tablet on two adjoining (at an angle of 90 degrees) walls and connect them. Such a setup would produce a natural feeling for the users, who

Figure 23. Intel RealSense SR305 depth camera

[25] https://www.ogre3d.org/
[26] http://wiki.ogre3d.org/HLMS+Materials
[27] https://www.intelrealsense.com/depth-camera-sr305/

would have their face both scanned and rendered from the side. Nevertheless, this setup can be somewhat confusing for some users, as proven by usability testing (see Chapter 5.1). Therefore, visual hints and intuitive indications of what is required from the user are essential in such a case.

Accessibility is among the concerns when designing the physical exhibit. The Glyptics Portrait Generator should be usable for anyone, therefore a solution for people of different heights is essential. One of the approaches is to have the camera attached to a movable mount, that can be adjusted for visitors of all heights. Another, simpler and cheaper solution, is to use an additional stand or chair in front of the exhibit, which would allow children to reach it.

However, these plans were not put into place due to the pandemic and subsequent closing of the Museum for the duration of the state of emergency. Therefore, a new goal was set – make the GPG as independent on the setup as possible. Because of that, the depth camera was replaced with a regular camera, and the implementation did not require specific positioning anymore. The only requirements that were left in the end is a computer, running a 64-bit version of Windows, a camera, connected to it, and sufficient lighting for the camera to distinguish the face. In the Museum itself, the exhibit will consist of a specifically designed wall with information about the engraved gems and three tablets. The side tablets are just for the pictures of the engraved gems, and the middle one will run the Glyptics Portrait Generator.

4.3 UI/UX Design

The idea of the UI from the very beginning of the development was to make it as minimalistic as possible. Three main rules for the UX for the GPG were outlined as such:

- as quick as possible to use
- intuitive to use
- requires minimal interaction

These rules align with the intended experience the visitor of the Museum should have. The GPG serves as a way to create a personalized memory from the exhibition, so the visitor should not waste time on understanding how the exhibit works or dealing with the problems. The first usability testing session in the Delta building has proven these points (see Chapter

5.1.1). Some testers struggled with the UI of the prototype and could not get the result they desired.

To achieve set goals, the final version of the GPG updates the preview in realtime. The user does not have to switch between the regular camera view and the result to see how it would look. Instead, the result is visible all the time and is updated accordingly to the current's user facial features and facial expression.

4.4 Software Architecture

Initially, the development began in C#. This language was chosen due to its convenience and previous personal experience with it. One of the main libraries, used in the beginning – Intel RealSense SDK – came with natively developed C# wrappers for its functionality. Although the documentation was lacking at times, the first prototype was successfully built. To tackle the rendering part of the thesis, the Helix Toolkit[28] library was used. This library had extensive out-of-the-box functionality and was very easy to use. It also supported custom shaders, which would be definitely needed in the later stages of development.

However, as the development continued, it became apparent, that the development in the field of computer vision is mostly done with C++ or Python. As the thesis progressed, new algorithms and libraries were required, and only a small fraction of them had implementations for C#. While it is possible to use C++ code in C# programs by creating wrappers and/or invokers, this is a tedious work that only halts the progress of the development process. Because of that, a decision was made to rewrite the program in C++. This decision had two sides to it: on the one hand, libraries became much more accessible, the choice was plentiful and the documentation was usually very detailed. On the other hand, however, C++ is known to be a very powerful, but hard to master language. Furthermore, the personal experience in the case of C++ was much, much lower. These factors slowed down the process, but also motivated to personally improve and develop. C++ is seen as the best choice for computer graphics due to being close to the machine code level and generally being very versatile and powerful[29].

---

[28] https://helix-toolkit.github.io/
[29] https://www.gamedesigning.org/career/programming-languages/

The program is divided into submodules, that are connected in the main file (Figure 24). These modules are:

- FrameCapturer – responsible for detecting video devices, starting video streams and capturing frames
- FacialDetector – responsible for detecting faces on captured frames and detecting landmarks on these faces
- FacialMorpher – fits the predefined morphable face model to the provided landmark data
- OgreApp – renders the result as an engraved gem

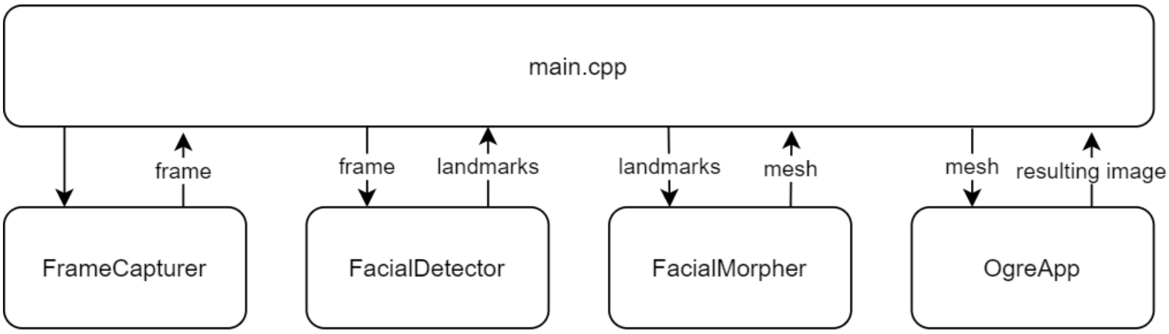


Figure 24. A diagram showing the modules of the Glyptics Portrait Generator

These modules are connected in such a way, that the result of each step serves as an input to the next one. If no faces are detected in the FacialDetector module, it idles and waits for any faces to appear in the view, and only then passes the result forward. Such structure allows for easier changes in the code when a change in one module does not directly affect all the other modules. Additionally, this has sped-up compilation times, which was very useful in this project due to the constant changes in parameters.

5 Assessment of the Results

To judge the success of the development as a whole, several testing sessions were held at different points of the development process. Therefore, this chapter contains the description and conclusions of these testing sessions.

5.1 Usability Testing

The main goal of the thesis is to create an interesting experience for the final users of the Generator. Therefore, usability testing was a vital part of the development process. Usability testing allows to collect feedback from the potential users, observe how they interact with the system and outline the main problems they encounter while doing so[30].

To test the Glyptics Portrait Generator, two usability testing sessions were conducted. The first one was held at the beginning of the development process and with a wide audience, while the second one was held at the end of the development and with a smaller audience.

5.1.1 Delta Centre Opening Event

The first iteration of usability testing was conducted during the opening of the new University of Tartu Delta Centre[31]. It took place on January 29th, 2020. This event was a good opportunity to have the first prototype of the GPG tested on a large and diverse group of people. Visitors included students, teachers, researchers, science enthusiasts, entrepreneurs and some other university workers. Among them were the workers of the University of Tartu Art Museum, whose original request lead to the creation of the thesis. They were most interested to see the results and test the prototype.

An MVP was constructed for the event. This process included both the creation of the software prototype and physical setup for it. Created software prototype implemented both parts of the algorithm (facial scanning and rendering) minimally. The stand with the GPG was constructed in the same way as it was planned to be put in the museum, but all parts and bindings were temporary to allow for the fast construction and deconstruction.

[30] https://www.nngroup.com/articles/usability-testing-101/
[31] https://www.ut.ee/en/news/university-tartu-delta-centre-be-opened-tartu-city-centre

The biggest difference with the final exhibit plan was the number of the stands. The demonstration in the Delta Centre had two identical exhibits, while the plan for the museum was to have only one. Having two stands allowed for two users using the Generator at the same time. This turned out to be a good decision due to the nature of the event: visitors usually came in groups. If several people wanted to use the Generator, it was possible to conduct two sessions at the same time and even compare the results. However, the amount of museum visitors is lower, than that of the Delta opening event. Therefore, having only one exhibit would be sufficient.

The user flow of the Glyptics Portrait Generator was the same as it was planned for the final product. The initial screen is the viewfinder (Figure 25, left), where the user can assess the expected result in real-time. In the prototype, the preview consisted of two overlaid streams: RGB and depth. The RGB stream was a regular video stream one can get of any video recording device. The depth stream was the product of the camera depth sensor, visualized by appointing different colours to different depth values. Blue meant the closest, red meant the farthest points.

From the viewfinder screen, users could go to the result view screen by pressing the “Capture” button (Figure 25, right). This action would capture the current frame and extract the depth data. It was then processed (centred, rotated and thresholded) and then turned into a point cloud. This point cloud was passed to the visualization module of the application, rendering it as-is on a predefined coin model with the background depicting the Blue Hall of the University of Tartu Art Museum. This model could be rotated, panned and scaled using the touch gestures. The button “Try again” took the user back to the viewfinder screen to allow for the new capture.

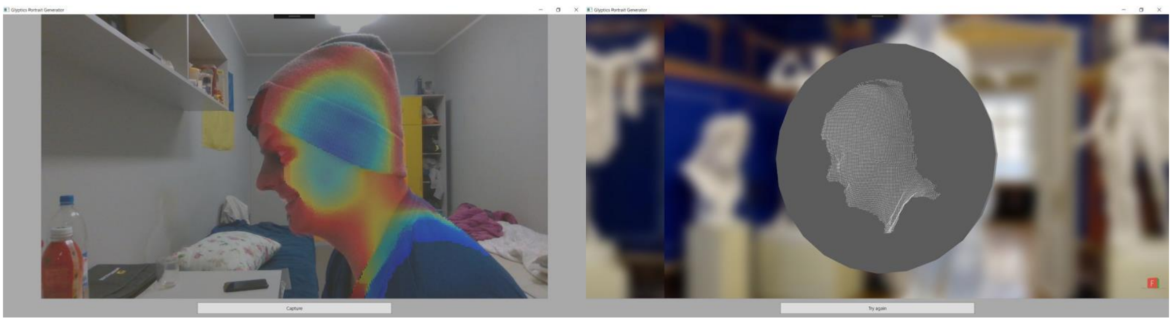


Figure 25. User Interface of the prototype presented at the Delta Centre opening event

The physical configuration of the prototype included two separate areas with the Generator, created by putting poster stands in a T shape (Figure 26). Each of these areas included a touch screen tablet and an Intel RealSense SR305 depth camera connected to it

(Figure 27). These devices were attached to the stand with duct tape and Velcro straps to allow for the easy disassembly. In the left-hand area, the camera was located to the right of a user, and in the right area, it was to the left. Because of that, the first major characteristic of the depth camera became apparent: it could transmit the data only through USB 3.0. This became clear when an attempt was made to connect the camera to the tablet through a USB extension cable in the right area due to the relative position of the devices. However, none of the available cables supported USB 3.0 and the camera did not work that way. To tackle that problem, the camera in the right area was rotated 180°, eliminating the need for the extension, and the software part of that exhibit was changed to account for that as well.

Figure 26. The overall appearance of the exhibit during the Delta Centre opening event

Figure 27. A close-up photo of the right area of the exhibit. The camera rotation issue is not fixed yet

Near the exhibit itself was located a table with a small informational stand (Figure 28). The stand was two-sided, containing identical information in Estonian and English. It described the motivation, goals and the general implementation behind the prototype. To illustrate the expected result better, the University of Tartu Art Museum provided two plaster replicas of engraved gems (Can be seen on the table near the informational stand in Figure 28 and closely in Figure 29). During the exhibition, it was noticed, that it was much easier to explain what engraved gems are and what exactly does the Glyptics Portrait Generator do with the help of these pieces.

Figure 28. A close-up photo of the table contents with the exhibit in the background

Figure 29. Two plaster engraved gem replicas provided by the University of Tartu Art Museum

The testing started at 11:15. All the preparations were conducted on a previous day. At first, there were not many visitors. Officially, the event began at 12:00 and after that time bigger groups of people started appearing. The testing lasted almost 6 hours, until 17:00. In general, during the entire session, 76 visitors tested the Glyptics Portrait Generator.

The testers had no prior knowledge of the GPG whatsoever. They were invited to the stand and offered to create a virtual 3D model of their face, similar to the demonstrated plaster replicas. Some amount of visitors hesitated but were convinced by the promise of the test taking as little time as possible. 12 of the testers explicitly expressed their satisfaction with how quickly it was done. All the testers were presented with an initial screen of the prototype and were prompted to try and create the 3D model. At this point, the opinions divided: quite a lot of the users were confused by the UI. Some of them have not noticed the "Capture" button, some did not understand, what the colours meant. To avoid causing frustration, help was provided after the first failed attempts. Out of the 76 users, around 58% managed to successfully use the prototype without any external aid. Still, sometimes it took several attempts to achieve the desired result. Nevertheless, all 76 users have eventually managed to proceed to the final step and assess the result.

After using the prototype, the users were asked only two questions – what would you like to see added or improved and how easy it was to use the GPG. These interactions were carried out verbally to make the testing experience as quick and as unsophisticated as possible. The results were still fixed in a written form, just without giving the users any kind of questionnaires as to not delay them.

43 testers noticed problems with hair in the resulting model. This was caused by the specifics of the used hardware. The Intel RealSense SR305 camera, used for the prototype, receives the depth information by rapidly projecting different patterns with UV light, capturing the results and getting depth information by assessing the deformations. This light, emitted by the camera, got dispersed and/or absorbed in the hair regions, causing the depth information to be absent or incorrect. Because of that, hair was mostly absent from the resulting model, leaving blank space in its place. This issue was more noticeable for the users with dark hair, as opposed to blonde-haired visitors, possibly since darker materials absorb more light and thus cause the UV-patterns to disappear more frequently.

A vast majority of the users were looking for a way to somehow share the result. These requests could be subdivided into two categories: physical and virtual. Visitors, who wanted to receive a physical representation of the result, were speaking about the ability to 3D print the resulting model. Virtual sharing meant having the ability to get a screenshot or a model file of the result, e.g. via email. To be more precise, 44 users asked about the ability to 3D print the result, and 14 inquired about the way to obtain a screenshot or a file (these numbers are partly overlapping, as some users asked about both). One person liked the result to such extent, that they requested a screenshot to be done manually (by the means of Windows operating system) and to be sent to their email.

11 users mentioned the fact that the result did not quite resemble their expectations. Usually, it meant the material of the resulting model. They expected the result to be rendered with some real engraved gems materials, but the actual model was using the default grey material. However, all of these users understood, that it was simply a demonstration of the work in progress.

11 users either did not understand that the final screen was interactable, or did not want to try. Some of the other users had problems with the controls. These issues were caused by the fact that the development was done on a laptop using a mouse and a keyboard, and the prototype had touch controls. The tablets, used in the demonstration, were received only a day before the event, thus there was not enough time to fix the controls. It is important to mention, that all the functions (rotating, panning, zooming) did actually work, albeit in a strange and non-intuitive fashion.

10 users struggled with positioning themselves correctly within the working area. The prototype was implemented in such a way, that it only captured objects located between 40 and 50 centimetres from the camera. This was visualized in the initial screen by the use of depth data colourizer – any part, which was not covered by the colour overlay, would not get scanned. Despite that, these users did not understand how exactly they had to move to get their face in the desired area. They were given directions on where to move, which they sometimes misinterpreted. For example, “move to the left” often resulted in users *turning* their head left instead of just *moving* it, keeping the rotation. Actually, more users had such a problem, but only 10 of them explicitly expressed it at any point during or after the testing.

Apart from that, both the display and the camera were fixed at one position, and because of that visitors, who were much higher or lower than the average person, had problems with the vertical position of the devices. This problem was expected and therefore an attempt to diminish it was made during the preparation – the left camera was located slightly higher, than the right one. Sadly, this difference was still not enough for some users.

Overall, using the prototype was a good experience for the vast majority of the users, despite any problems they had. They were asked to be honest and to not restrict themselves by politeness or any other factors. As for the main outcomes of the testing, the main problems were highlighted. All of them were expected to some degree, but the number of users mentioning 3D printing was astonishing. Issues with the users positioning themselves could be solved in different ways:

- By attaching the camera (and, possibly, the display) to an adjustable mount, allowing it to move vertically to account for height differences.
- By physically marking the place where the user is supposed to stand in an obvious and intuitive way (e.g. footprints)
- By providing more information on the display: where to look, where to move, how to turn, etc.

Several solutions for the missing hair were considered. One of them is to add the ability to put hair pre-sets on the resulting model, choosing the most suitable or likeable one. Another is to detect the hair regions from the image and add it procedurally. The first method seems less restrictive and allows the users to fantasize, possibly using the hairstyles they do not actually have. The second method, however, requires less interaction with the GPG and is thus quicker.

The ability to share the result was planned from the beginning, but the way to do it was not clear. The users, who mentioned this feature, were mostly meaning the ability to get some kind of link via email, but this would require them entering the email, which is both cumbersome and can be considered unsafe by some. Therefore, the better solution would be to use QR codes, in the same way as the "Magic Mirror"[32] in the Ülikooli 17 hallway did. The system could upload the result, generate a link to it and convert it to a QR code, allowing the user to scan it. The 3D printing request was deemed too complex to implement for the time being, but the feedback was still redirected to the staff of the museum.

The visual part (i.e. actual engraved gem materials) were to be implemented later, and this testing has only shown that users want to see it, although not as much, as expected. A lot of the users found the look of the prototype to be interesting already, but that might be because there was not much to compare it to – users were usually more captivated by the depth-scanning technology itself, and not the graphical aspects behind the project.

[32] https://twitter.com/Ansip_EU/status/968078775039791104

5.1.2 Private Testing Sessions

The initial development plan included the final usability testing of the GPG in the intended environment – during the museum exhibition. However, due to the COVID-19, the Art Museum of the University of Tartu was temporarily closed and such testing became impossible. Furthermore, the restrictions introduced by the government ruled out the possibility of any massive testing session.

Because of such a situation, it was decided to instead conduct several 1-on-1 private sessions with the people, willing to participate in the testing. Due to the changes in the development schedule and major changes regarding the algorithm itself, only the facial reconstruction part was tested. By this point the algorithm has completely changed – instead of capturing the depth information from the depth camera, it now relied on the landmark detection from the 2D image and the morphing of the pre-existing mesh to the results. The test itself was extremely simple – the facial reconstruction would run, using the image of tester's face as an input, and the tester would express their opinion on the result. By the time of these sessions, the GPG did not require any specific setup apart from the laptop with a camera, so no preparations were needed. To avoid personal contact as much as possible, all of the tests were conducted remotely, as they did not require any physical interaction with the system from the tester.

8 people have directly participated in the testing. The tester would send a picture of their face (frontal view) or join a video call. The image (or the screen of the video call) was then passed to the GPG and the result was sent back to the tester. The GPG produces a *.obj* file of the face as an output, and this format is fairly common. A program called Microsoft 3D Viewer is included in all Windows 10 versions starting from April 2017. If the tester did not have the required software, they would be directed to a website, which supports online preview of *.obj* files[33]. The output of the program along with the real photo (both from the lateral view) is shown in Figure 30.

[33] https://www.creators3d.com/online-viewer

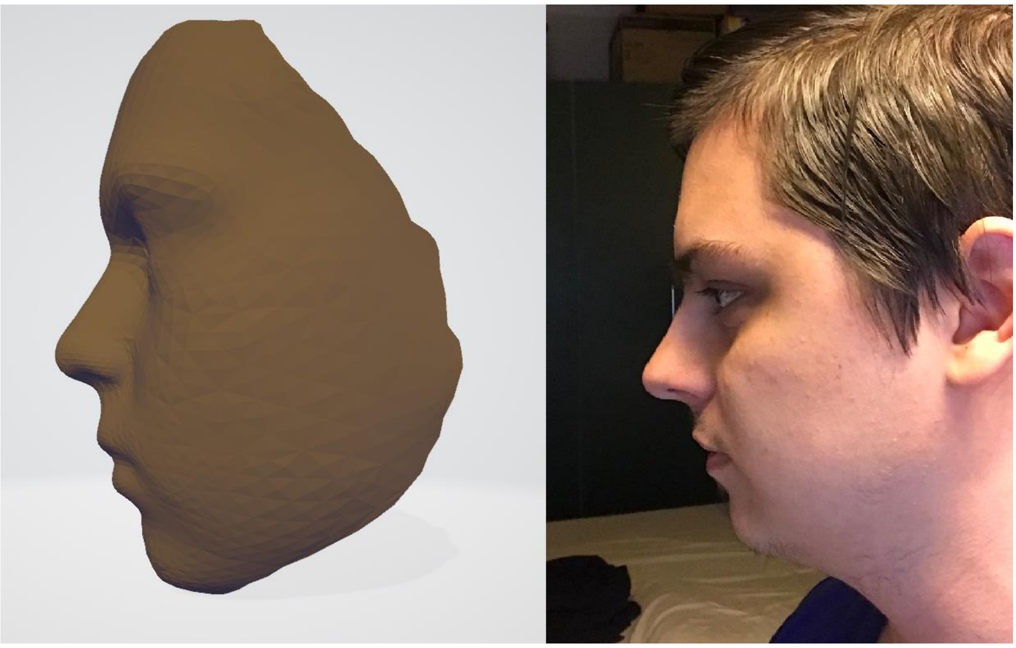


Figure 30. The output of the facial reconstruction algorithm (left) and the actual photo (right)

After the testers received the results, they were asked to assess how closely the output resembles them. They were asked to give an answer in the form of a number between 0 and 10, where 0 means “completely unrecognisable” and 10 means “identical”. The average result was 7.125, which is quite good considering how late the change had to be introduced.

Usually, the testers used the GPG several times (around 3-5). The reasoning for that was different. Some of the testers tried to intentionally obstruct the face or negatively affect the result in any other way to test the robustness of the solution. Others acted in the opposite direction by trying to get a better result. Due to that, several more conclusions were made.

First of all, the results from different unobstructed attempts were fairly similar to each other. The implementation is indifferent towards slight changes in the lighting or face position due to the nature of the landmark detection algorithm. However, the implementation is quite susceptible to intentional “sabotage”. If the tester would hide some part of the face or turn away to the side, landmark detection stopped working and therefore the output was not produced. This is, however, not unexpected, because the algorithm heavily relies on detecting distinctive facial landmarks from the image only and thus fails if it does not manage to do so properly. It is also worth noting that glasses (regular) do not affect the result, as the eye contours and eyebrows are still clearly visible. Putting on dark sunglasses yields an interesting effect – the landmark detection works, but is very “jumpy” with the eyes and the eyebrows, as it is unable to locate them and probably just attempts to "guess" the presumed locations.

The other issue with the system was the fact that it did not incorporate any actual depth measurements into the implementation, working with just RGB images. This caused several discrepancies between actual faces and the received results in such things as the nose length or the eyes depth. However, most of the times the results were quite similar to the expected ones, even though the testers expressed their doubt in the algorithm's ability to correctly estimate the depth of the face. Most likely this is caused by the fact, that some features (e.g., which lip protrudes further), can be detected from the frontal view images.

Apart from these sessions and testers, other images were used in the development process for the testing purpose as well. Images of celebrities of different age, gender and race were used to test the system (see Appendix). This was, however, done just to see how well the implementation works without conducting any numerical assessments, as they would be biased and highly subjective. Of course, the conducted sessions were not completely objective as well, but it is much easier for someone to evaluate the likeliness of something to their face and notice any problems while doing so.

6 Conclusions and Future Work

In this thesis, the Glyptics Portrait Generator was created – a piece of software that creates visual models of people's faces in glyptic art style. The program was initially planned to be exclusively an exhibit for the University of Tartu Art Museum's spring exhibition, dedicated to engraved gems. However, due to the state of emergency, caused by the COVID-19 pandemic, the GPG was repurposed to be accessible to any user with a web-camera. This change caused a significant redesign of the solution in general, starting from the technologies used, to the UI and ways of distribution.

Coming from the nature of the Glyptics Portrait Generator, two extensive scientific areas were researched from the fields of computer vision and computer graphics, namely facial reconstruction and physically-based rendering with the accent made on gems. As a result, several facial reconstruction techniques were outlined and compared. The initial decision was to explicitly use 3D scanning, and several Intel RealSense SR305 cameras were purchased to create a physical exhibit in the Museum. Furthermore, the first usability testing session, held during the University of Tartu Delta Centre opening event, featured a prototype of such an exhibit with the depth cameras. However, later the approach had to be changed and the final algorithm falls under the hybrid technique category, which combines the information from 2D images with prior 3D model. To be more precise, the information from the 2D images is the landmarks, detected on the user's face, which are used to morph the initial 3D model into the resulting mesh.

As for the rendering part of the thesis, the ways to represent gems in computer graphics were researched. Apart from that, the optical effects present in gemstones were described along with their theoretical implementation. These optical phenomena intend to make the result more visually captivating and memorable. The rendering was implemented in the form of shaders for the materials in the OGRE 3D graphics engine. These shaders create the desired look of engraved gems for the previously generated meshes.

This thesis could potentially be split into two parts and thus created in a team of two. One person would focus on the facial reconstruction part, while the other implemented the rendering part of the thesis. This way the work would be much more focused and productive, due to the thesis being so easily dividable. This would also allow for much more extensive research. As such, the future work on the GPG may come from both parts.

The facial reconstruction part can be improved by raising the precision of the algorithm, while still keeping the general hybrid technique approach. For that, the landmark detection algorithm can be extended upon, to work not only from the frontal view but from the lateral as well. This task boils down to deciding on the lateral view landmarks and training the model on the sets of faces from the lateral view. On top of that, a temporal term can be introduced into the scanning process, by capturing several frames over time and tracking the landmarks in them to correctly assess the depth of the facial features. In the rendering part of the thesis, better approximations of materials can be researched and developed. The optical effects, described in Chapter 3.2. can be simulated to assess their influence on the overall look and feel of the result.

As a personal note, I am grateful to the University of Tartu Art Museum for the opportunity to create something this exciting, for constant support and great feedback.

References

[1] Gem Research. The Classical Art Research Centre, Oxford University. [Online]. http://www.beazley.ox.ac.uk/gems/default.htm

[2] F. L. Gray, "Engraved gems: a historical perspective," *Gems and gemology*, no. 9, pp. 191-201, 1983.

[3] B. Kitchenham,. Keele, UK: Keele University, 2004, ch. 33, pp. 1-26.

[4] W. N. Widanagamaachchi and A. T. Dharmaratne, "3D face reconstruction from 2D images," *IEEE, Digital Image Computing: Techniques and Applications (DICTA)*, pp. 365-371, December 2008.

[5] S. Liu, Z. Wang, X. Yang, and J. Zhang, "Realtime dynamic 3D facial reconstruction for monocular video in-the-wild," in *IEEE International Conference on Computer Vision Workshops*, 2017, pp. 777-785.

[6] Booth J. et al., "3D reconstruction of "in-the-wild" faces in images and videos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2638-2652, 2018.

[7] A. V. Kumar et al., "Dense reconstruction of 3D human face using 5 images and no reference model," *4th International Conference on Control, Decision and Information Technologies (CoDIT)*, pp. 1185-1190, April 2017.

[8] Y. Zhang et al., "A fast 3D reconstruction system with a low-cost camera accessory," *Scientific reports*, vol. 5, no. 1, pp. 1-7, 2015.

[9] L. Jiang, J. Zhang, B. Deng, H. Li, and L. Liu, "3D face reconstruction with geometry details from a single image," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 4756-4770, 2018.

[10] E. Bondi, P. Pala, S. Berretti, and A. Del Bimbo, "Reconstructing high-resolution face models from Kinect depth sequences," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2843-2853, 2016.

[11] M. Hernandez, J. Choi, and G. Medioni, "Near laser-scan quality 3-D face reconstruction from a low-quality depth stream," *Image and Vision Computing*, no. 36, pp. 61-69, 2015.

[12] A. Munoz et al., "BSSRDF estimation from single images," *Computer Graphics Forum*, vol. 30, no. 2, pp. 455-464, April 2011.

[13] S. Marschner and P. Shirley, *Fundamentals of computer graphics*. Natick, Massachusetts: CRC Press, 2015.

[14] J. Kautz, "Hardware lighting and shading: a survey," *Computer Graphics Forum*, vol. 23, no. 1, pp. 85-112, March 2004.

[15] Weidlich A. and Wilkie A., "Modeling aventurescent gems with procedural textures," in *Proceedings of the 24th Spring Conference on Computer Graphics*, 2008, pp. 51-58.

[16] Guy S. and Soler C., "Graphics gems revisited: fast and physically-based rendering of gemstones," *ACM SIGGRAPH 2004 Papers*, pp. 231-238, 2004.

Appendix

I. List of Abbreviations

GPG	Glyptics Portrait Generator
2-, 3-, 4D	Two-, three-, four-dimensional
LED	Light-emitting diode
RGB	Red, Green, Blue
FPS	Frames per second
RMSE	Root mean square error
PBR/PBS	Physically-based rendering/shading
BSSRDF	Bidirectional scattering-surface reflectance distribution function
BRDF	Bidirectional reflectance distribution function
BTDF	Bidirectional transmittance distribution function
SFM	Surrey Face Model
BFM	Basel Face Model
4DFM	4D Face Model
UI	User Interface
UX	User Experience

II. Glossary

Rendering	In computer graphics: the process of model visualization from its description and properties.
Vertex (mesh)	A point in 2D or 3D space, usually a part of some model.
Edge (mesh)	Line segment, connecting two vertices.
Face (mesh)	A closed set of edges (3 or 4), representing surface or its part.
Polygon	Usually the same as a face, sometimes a coplanar set of faces; representation of the surface
Mesh	Short for “polygon mesh”; same as “[3D] model”; Representation of 3D objects in computer graphics as a set of vertices, edges and faces.
Morphing	The technology of transforming one object into another by changing the vertices positions in the model.
Gem cut	The shape of the gem produced by cutting it[34].
Blendshape	A set of mesh deformations, representing different predefined shapes; in the context of the GPG – different facial expressions

[34] https://www.gemsociety.org/article/gem-cutting-terms/

III. Testing results

1. Delta building testing results

Total number of participants: **76**

The following numbers are partially overlapping, as most of the testers had several opinions and several feedback points.

Feedback or Result	Number of testers	Percentage of total testers
Managed to get the result without external help	44	57.8%
Liked how quickly the prototype worked	12	15.8%
Had problems with hair visualization (missing or noisy)	43	55.1%
Asked about the possibility of 3D printing	44	57.8%
Asked about a way to share a screenshot	14	18.4%
Noticed that the result did not resemble an actual engraved gem	11	14.5%
Did not realize that the screen was interactable	11	14.5%
Struggled with positioning themselves in the working area	10	13.2%

2. Private testing sessions

Tester #	1	2	3	4	5	6	7	8
Grade	7	8	7	6	8	6	8	7

Average grade: 7.125

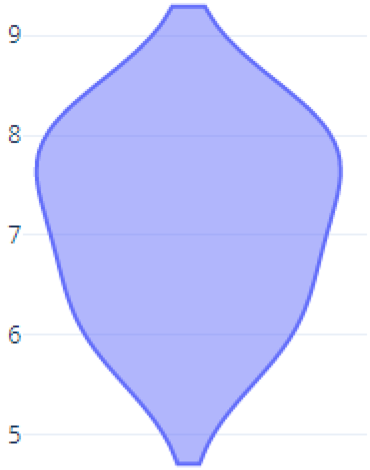
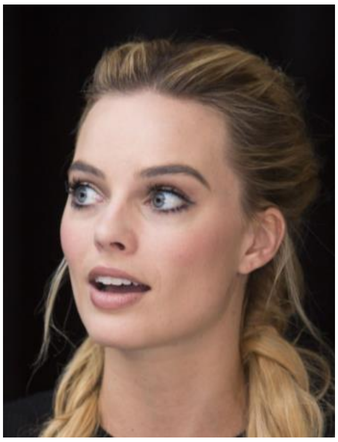
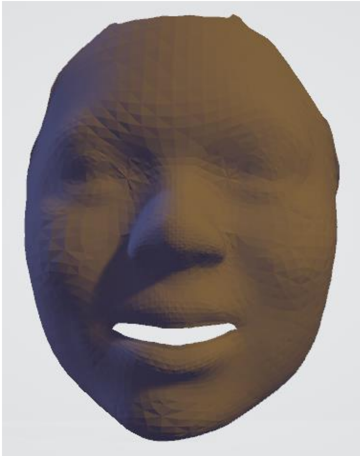
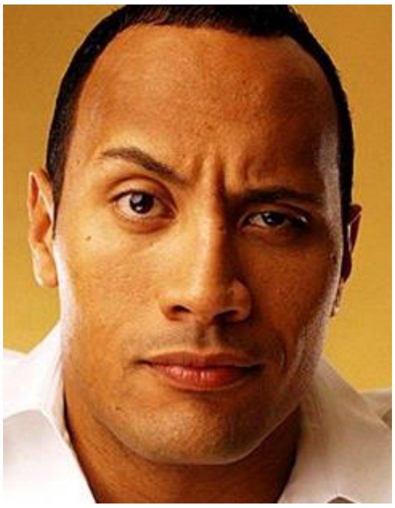
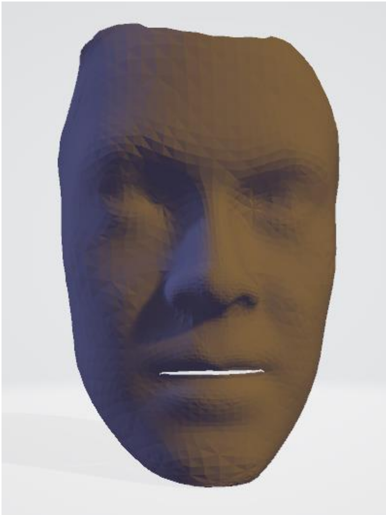


Figure 31. Violin plot of the grades' distribution

3. Celebrity images and results

Name	Celebrity photo	Morphed Face
Hideo Kojima		
Margot Robbie		
Donald Trump		

Oprah Winfrey		
Dwayne "The Rock" Johnson		

IV. Accompanying files

The thesis is supplied with accompanying files, that contain the source code of the application.

The source code can also be found on GitHub with the following URL: https://github.com/AllysanderStark/glypticsgenerator.

Structure of **glypticsgenerator** folder:

- res – supporting media files, required for the program
- src – the source code
- Visual Studio project files

Unfortunately, the build is not provided with the source files, as currently there is an issue with having the ability to run the program directly from the .exe file.

V. License

Non-exclusive licence to reproduce thesis and make thesis public

I, **Vladyslav Kupriienko**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Glyptics Portrait Generator,

supervised by **Raimond-Hendrik Tunnel**.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Vladyslav Kupriienko

15/05/2020