Anonymization of 3D face models for GDPR compliant outsourcing to 3rd party companies

Pietro Rustici[1], Patrizia Marti[2] and Klaus Hildebrandt[3]
[1] Delft University of Technology, [2] (Local supervisor) University of Siena, [3] (Remote supervisor) Delft University of Technology

ABSTRACT This study investigates whether an automatic anonymization algorithm that takes as input a 3D model of a human face can produce an output model exempt from General Data Protection Regulation (GDPR) biometric data definition. The algorithm first uses Random Sample Consensus (RANSAC) for registering the source point cloud globally to an oriented template. Next, the alignment is refined using an Iterative Closest Point (ICP) technique. Secondly, a subset of the source point cloud is created using a fixed radius vector on each template point resulting in the corresponding face contour. Finally, the algorithm uses a point cloud template to remove the unnecessary facial features and converts the point cloud to a mesh. The quality of the anonymization has been evaluated using a survey assessment of 100 participants. The latter resulted in half of the participants failing to recognize any of the anonymized models, one-fifth scoring one out of four correct. Only 2% correctly associated all the models to the right individual.

KEYWORDS Face anonymization; GDPR; Point clouds; Pose-orientation; Pose-registration;

1 INTRODUCTON

Healthcare facilities are increasingly dependent on third parties to provide mission-critical services. However, storing and sharing medical data between the two raise severe concerns around the matter of individual privacy. A recent design example that is affected by the above- mentioned issue is SuperPowerMe, an ongoing research project aiming at designing customized face-masks for the early treatment of Class III malocclusion in children. The project developed a concept of an innovative augmented face-mask for the orthopedic correction of this maxillofacial disorder in children with the objective to overcome the limitations of commercial face-masks (poor ergonomics, skin irritation, poor aesthetics) and to improve the acceptance by the young patients [1].

In order to create the customized face-mask, a mesh of the child's face is captured using a structured light 3D scanner or photogrammetry. The latter is then used as a reference to design the parts of the mask in contact with the patient's skin. Under GDPR regulations, this type of data falls under the "personal" data category [2] which requires the data subject, or its legal guardian, to give "explicit consent" for the use of the latter. In order to ease this bureaucratic procedure or in case the consent is not given, the 3D model can be anonymized by extracting some of the features not needed to create the face-mask devices.

The literature review shows that face features anonymization is commonly performed on 2D images but rarely on 3D models. Current 2D anonymization methods include quality reduction techniques (blurring, pixelation, deterioration), k-Same-Net [3][4], and face-swapping [5].

1.1 GDPR RESTRICTIONS AND DEFINITIONS

For the context of this research, some specific features of the face need to remain unaffected by the anonymizing algorithm because they are used to create the customized face-mask. With the aid of the experts in charge of developing the mask, we identified the minimal set of requirements that includes the forehead, the chin, the profile silhouette of the face, and the face's external outline. The remaining portion of the face can be removed entirely. GDPR defines 'biometric data' as *personal data resulting from specific technical processing*

relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data and 'pseudonymisation' as the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information (Article 4, EU GDPR "Definitions") [19]. The regulation does not provide a clear framework of the facial features to be removed in order to comply. Instead, it exposes a gray area that leaves room for personal interpretations.

1.2 OBJECTIVE

The objective of this research paper is to understand if given the physiological reference points needed to design the customized face-mask, it is possible to automatically anonymize the face model by removing the remaining features and granting the GDPR compliance. This research is intended as a starting point to further research.

1.3 RELATED WORK

Several approaches for 3D face registration and pose estimation exist in literacy [6][7]. They can be categorized as landmark-based, holistic or a combination of the two. Since the spatial orientation of the point clouds examined in this research is unknown, facial-landmarks can probably not be extracted without first aligning the point cloud. Holistic point cloud registration algorithms can be divided into two main categories: global registration and fine registration. The first does not assume any prior alignment. An example can be the RANSAC approach [8] that takes random point groups from the source cloud and detects the corresponding points in the target cloud by querying the nearest neighbor in the FPFH feature space. Fine registration is mainly used when the two point clouds are already close to the alignment position. Numerous solutions exist in the literature. However, the majority revolves around the Iterative Closest Point (ICP) algorithm [9], a pillar of geometric registration in both research and industry.

Regarding the evaluation of the anonymized output model's quality, it is not straightforward to categorize and understand what has been done in literacy. As discussed in Section 1.1, GDPR provides an imprecise definition of what can be considered anonymized. However, we can identify two trends in the industry when it comes to asserting the anonymization efficacy. The most popular and straightforward method is to have the internal Data Protection Officer (DPO) certificating compliance. Alternatively, the data subject itself is asked to verify and acknowledge the anonymization.

Concerning the acquisition of 3D models, the technology used is closely related to the object's properties [11]. For example, photogrammetry, which utilizes a set of images taken at different positions to reconstruct the 3D geometry, is commonly used by land surveyors to create relief maps or perform measurements. Instead, total stations (TS) are widely used by civil engineers, archeologists, and police officers to precisely measure the boundaries of the surrounding environment. A more compact solution is structured light, which produces exact measurements for small objects.

3 METHODOLOGY

The methodology used during this research is composed of four main steps and their respective sub-steps. The first step is data acquisition and preprocessing, which collects point cloud data from a 3D scanner and preprocesses them. Next, the point cloud is oriented by registering it to a target template. In the third step, the facial contour is extracted, anonymized, and exported as mesh. The anonymization is then validated through a survey. submitted to a broad audience. The block diagram of the proposed methodology is shown in Fig. 1.

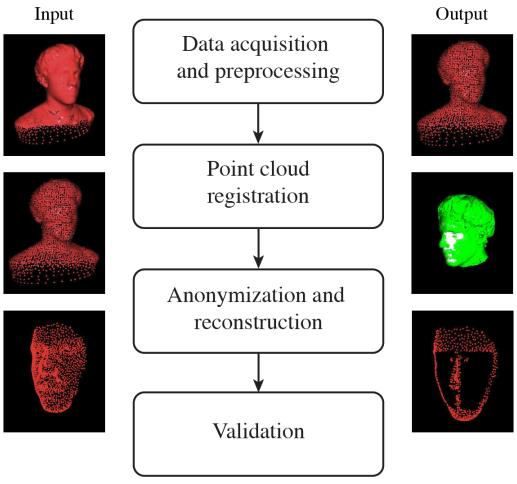


Figure 1 - Algorithm steps diagram

4 OVERVIEW OF THE PROPOSED ALGORITHM

In the following sections, the algorithm used to register and anonymize all the face models present in this research will be explained in detail.

4.1 DATA ACQUISITION AND PREPROCESSING

The model is acquired using an Artec Leo scanner, a wireless handheld 3D scanner using structured light scanning technology (Vertical Cavity Surface Emitting Laser light technology). The scanner excels in its ability to digitize hard to scan textures, and it's able to acquire objects in high resolution [12]. The Artec Leo was selected due to its advanced capabilities and flexibility to capture objects of different sizes and varying detail. The scanned area comprises the face of the patient and also part of the torso. The resulting clouds contain between 20.000 and 30.000 points. The point cloud is then exported in the standard .ply format.

Since the point cloud produced from the scanner contains a large number of points, a copy of it is downsampled to about 2.000 / 3.000 points using the function provided by the open3d python library [13]. The downsampling is performed with a voxel dimension of 10. Next, the normals are recomputed, and the Fast Point Feature Histograms (FPFH) is generated. The process is depicted in Fig. 2.

(a) Out of the scan pointcloud

(b) Voxel downsampled

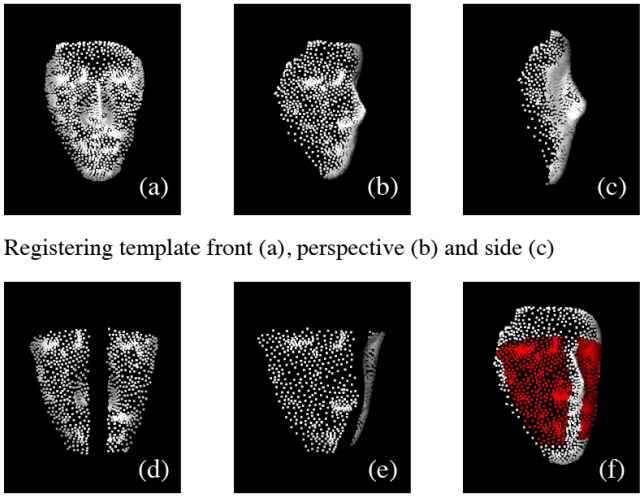
(c) Normals computation

Figure 2 - Preprocessing steps and outcomes.

4.2 TEMPLATE GENERATION

Depending on the initial scanning position, the resulting model might not be oriented according to one of the three conventional axes. The proposed algorithm uses a universal reference template point cloud to calculate the source object translation and rotation matrices. Needless to say, that human face cannot be uniquely reconducted to a strict geometrical structure. However, some particular features can be identified, and they can be exploited to create a universal template.

The template used for this study has been generated using the average points of four models of relatively different face traits. From the previous template, another one has been created. The latter does not contain the features to be anonymized and will be used later in the anonymization process to extract the final point cloud. Fig. 3 (a,b,c) depict the registration template, (d,e) corresponds to the anonymization template and (f) is the union of the two templates.



Registering template front (a), perspective (b) and side (c)

Anonymization template front (a), perspective (b) and union (c)

Figure 3 - Overview of the two generated templates.

4.3 POINT CLOUD REGISTRATION AND PERFORMANCE EVALUATION

In this part of the algorithm, the word template refers to the registration template and not the anonymization template, unless specified.

The registration of the source point cloud w.r.t. the template consists of two steps. The first step is a coarse alignment and does not rely on previous knowledge about the source cloud position. Next, a fine alignment is performed. The initial alignment algorithm makes use of the previously computed FPFH to identify the point-to-point correspondence between the source point cloud and the target one. The algorithm iterates several times until it satisfies a stopping criterium. For the point clouds used in this study, two estimation methods are used:

- **Distance-based checker:** tests whether all randomly sorted aligned points are closer than a specified threshold to their corresponding points in the target cloud.
- **Edge lenght-based checker:** Given the correspondence set calculated by confronting the FPFH of the source and target point cloud, this checker verifies whether each edge couple's length differs at most to a given threshold.

The coarse registration has been performed on the point clouds captured from four subjects.. As shown in Fig. 4, the resulting pose is not highly accurate but provides a strong starting point for the fine alignment algorithm. Table 1 shows the fitness and the Root-Mean-Square Error (RMSE) between the source cloud and the target template.

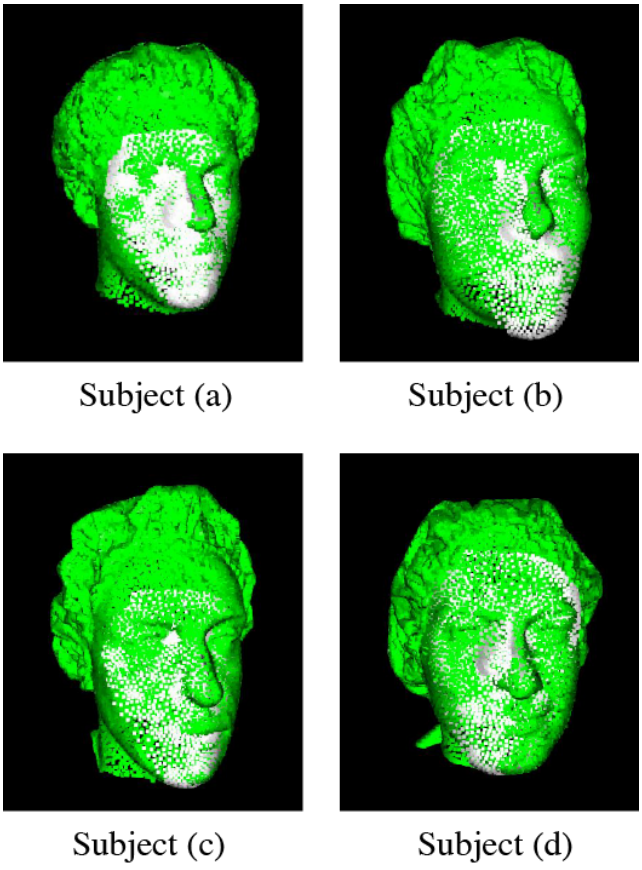


Figure 4 - Coarse registration result on the four subjects.

Subsequently to the coarse alignment, the original point cloud is transformed accordingly to the transformation matrix generated by the algorithm. Next, the source point cloud is transformed using an ICP algorithm, which keeps the target cloud fixed, and iteratively tries to move the target cloud to best match the reference. For each iteration, a correspondence set of points from the source and target cloud is identified. The current transformation is updated by minimizing an objective function, as shown in [10]. The algorithm's output on the subject's point cloud is shown in Fig. 5, and a fitness/ RMSE comparison between the coarse and fine alignment is illustrated in Table 1.

subject	**coarse alignment** fitness / RMSE	**ICP alignment** fitness / RMSE
(a)	9,903E-02 / 5.054	1.163E-01 / 3.371
(b)	9,645E-02 / 5.608	1.094E-01 / 4.230
(c)	10,046E-02 / 5.381	1.143E-01 / 3.933
(d)	12,710E-02 / 5.201	1.288E-01 / 4.112

Table 1 Alignment result and comparison of the two algorithms

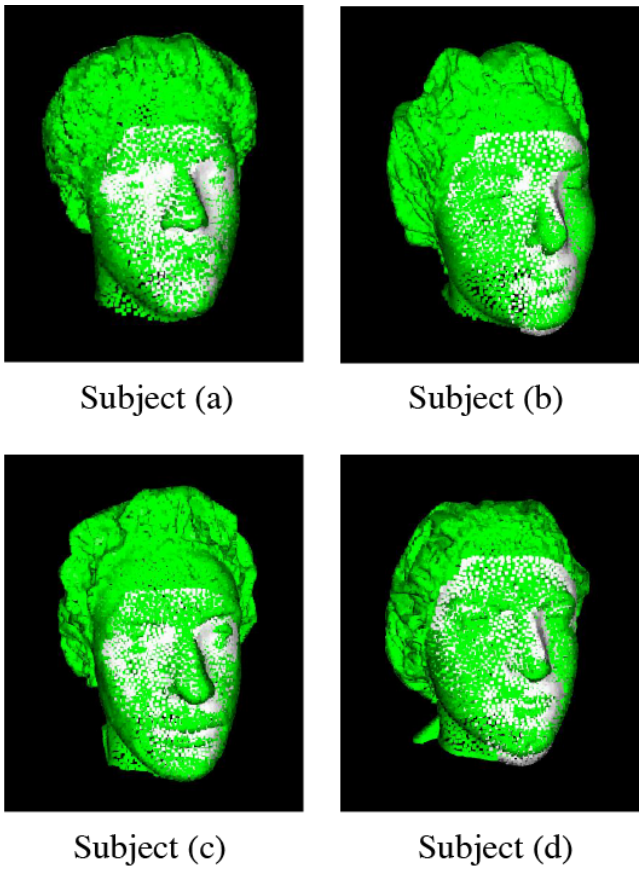


Figure 5 - ICP registration result on the four subjects.

4.4 POINT CLOUD ANONYMIZATION AND MESH RECONSTRUCTION

Removing as many points as possible from the point cloud indeed increases the anonymization strength. However, it's clear that the majority of facial features

that are unique for a human being reside in the facial part of the head. To extract the contour of the face, the same template applied for registering it's used. The technique adopted it's somewhat naive yet powerful for the specific application. First, a kD-Tree spacial locator is generated from the source point cloud to increase the search speed. The tree is generated using the Fast Library For Approximate Nearest Neighbour (FLANN), as described in [14]. Next, the algorithm iterates through all the points of the template cloud and, for each of them, calculates the points from the target cloud that are closer - 3D space - from a specified distance. By increasing or decreasing the distance value, more or fewer points are captured from the algorithm. All the points collected from the search are saved in a new cloud. The outcome is the facial area shown in Fig. 6 for each subject point cloud.

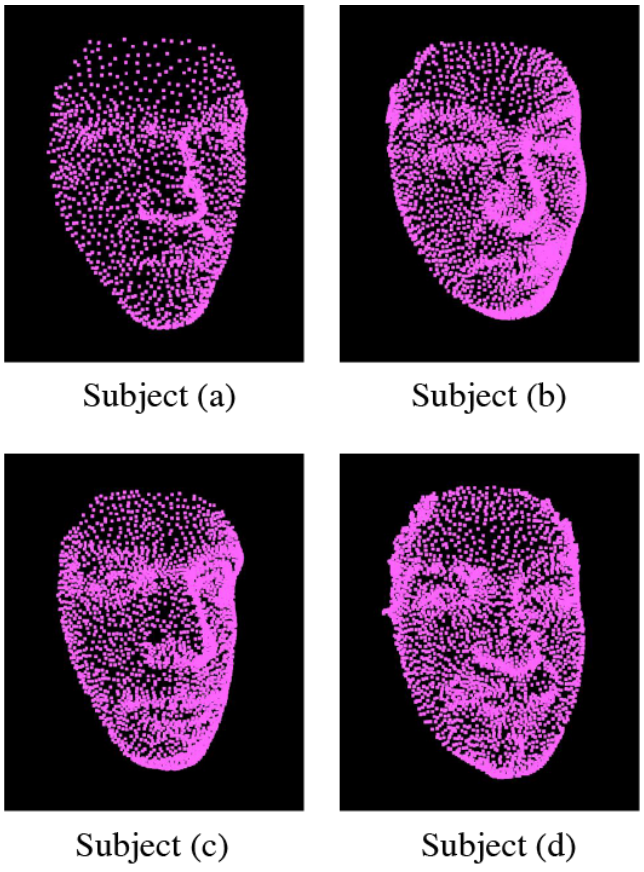


Figure 6 - Contour extraction result on the four subjects.

As briefly mentioned in section 1.2, some facial features are required to be kept intact in order to develop the customized face-mask. In particular, the algorithm is expected to keep the forehead area, the chin, the profile silhouette of the face, and the face's external outline. Similarly to the technique used for removing everything except the frontal face area, a template it's used. However, this time the template corresponds to the area to be removed/anonymized Fig. 3 (d,e,f). A new kD-Tree is generated from the facial contour point cloud. The algorithm iterates over the template points and detects the points to be removed. The contour extraction and anonymization algorithm could be easily embedded into a single solution, e.g., using a colored point cloud. However, for the sake of clarity - in this report - it's presented as two separate steps. Finally, the point cloud is converted into a mesh using the standard Ball-Pivoting Algorithm (BPA) [15], which utilizes a virtual ball of a chosen radius to roll on the cloud's points, interpolating the mesh. The outcome is the anonymized point cloud shown in Fig. 7 for each subject point cloud.

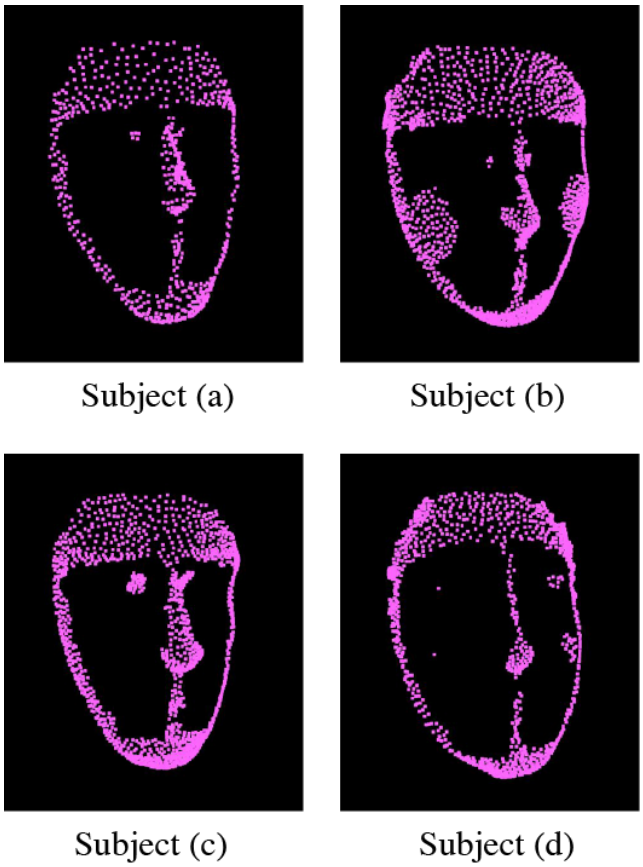


Figure 7 - Anonymization result on the four subjects.

5 EXPERIMENTAL RESULTS

The anonymization algorithm's efficacy has been evaluated with the aid of a survey. The survey is built

upon the GDPR's definition, which states that biodata is considered anonymized if it can not be attributed to a specific individual without the use of additional information. The experiment is considered successful if the participants will not associate the anonymized model to the correct individual among six possible pictures of different people.

5.1 EXPERIMENTAL DESIGN

This study was performed using an online questionnaire for several reasons. Firstly allow the participants to navigate the model using a 3D viewer, resulting in a more accurate and realistic understanding of the facial characteristics. Secondly, to increase the number of participants and reach a varied audience.

The questionnaire consisted of two pages:

- A landing page containing the informed consent and the research context and goal.
- The survey page, containing the four anonymized models and their corresponding solutions as shown in Figure 8.

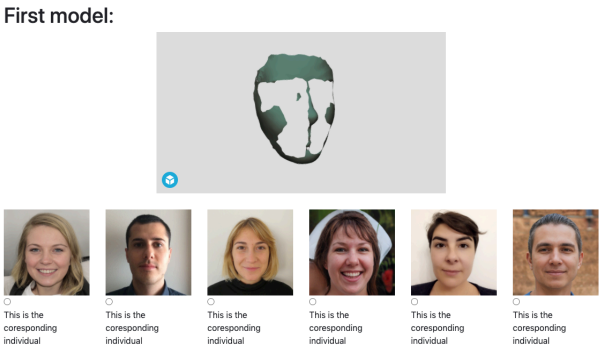


Figure 8 - An example of the survey interface.

The survey remained online for 15 days and was shared through social media until it reached 100 answers. Participants' ages ranged from 25 to 60 years, with an equal share between females and males. All the participants knew at least one of the subjects of the anonymization procedure. The images used as answers for the survey are taken from an online database of AI-generated photos of humans [20], except for the correct one, which is a patient's real photo. To prevent bias in the judgments, no background on the creation of the anonymized model was provided to the participants. However, considering that each participant knew at least one of the subjects might have introduced a small bias that can only negatively impact the survey outcome w.r.t. the algorithm's success.

5.2 EVALUATION OF RESULTS

Figure 9 describes the distribution of correct answers plotted as a bar chart. The x-axis represents the number of correct answers, while the y-axis indicates the percentage of participants. The questionnaire's outcome shows that more than half of the participants' share did not manage to associate the anonymized model to the right individual correctly. One-fifth correctly detected only one out of four models, and only 2% recognized all subjects. Although the data was collected from samples that abundantly differ in face size and morphology, it does not mean that the results apply to the whole population. However, this research's findings contribute to filling a gap in the anonymization of 3D medical face images and provide a starting point for further investigation.

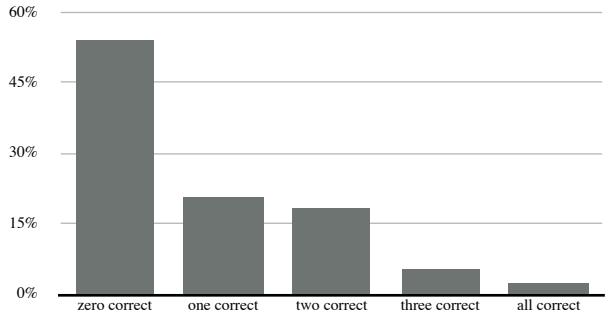


Figure 9 - Percentage share of correct answers among 100 participants.

6 RESPONSIBLE RESEARCH

In the last decade, we have witnessed to a massive growth in the interdisciplinary connection between health care systems and computing, which have led to a significant increase in volume and variety of individual's health and biometric data. Data science, data mining, and machine learning advancements allowed researchers to examine and discover new trends, develop new possibilities, produce more efficient services. However, the massive usage of personal data raised a severe concern about the privacy of the individual. The research illustrated in

this paper tries to combine a particular need from the medical community and the GDPR privacy restriction - w.r.t. the individual - that all European citizens are subjected to. As discussed in section 1.1, the current regulation does not explicitly provide a framework to assess an anonymization algorithm's efficacy. The survey conducted in this study tries to give a qualitative and quantitative answer to the research question. However, it does not exempt the service provider from seeking bureaucratic consent from the data subject or an internal DPO. Although, it can increase the individual's acquiescence or provide a useful tool for the person in charge of certifying the anonymization. The algorithm proposed exploits the face-mask design requirement to anonymize facial features, which are particularly convenient since it does not include facial traits commonly used - from a human point of view - to identify an individual uniquely; for example, the eyes, nose shape, and mouth. Depending on the medical treatment requirements might or might not be possible to apply the same procedure described in this paper and achieve an equal anonymization strength.

7 CONCLUSION AND FUTURE WORK

The goal of this research was to provide a methodology to answer whether is it possible to anonymize - in a GDPR compliant way - a 3D scan of a human face while maintaining a pre-defined set of facial features that are later utilized by 3rd party companies to develop a face-mask for the treatment of Class III malocclusion. The research question has been answered by creating an algorithm that takes as input a 3D model and outputs its anonymized version. The subjects have been locally 3D acquired using a structured-light scanner. The algorithm aligns the model to a universal template in two steps. The alignment utilizes a coarse registration technique to provide a global positioning of the source face w.r.t. the target template. Next, the registration is refined using an ICP algorithm. After the alignment is completed, the head's frontal face area is isolated using the upper-mentioned template as a reference. Finally, the model is anonymized using a modified template and recomputed as a mesh using a ball-pivoting algorithm. The validation of the algorithm is achieved using a survey conducted over an audience of 100 participants, resulting in only 2% being able to identify all four subjects correctly and more than half of the participants not recognizing any.

The study has shown that it's indeed possible to anonymize the facial model for this research's medical case subject. However, it does not replace the data subject formal consent when sharing the data with third-parties but provides a tool that can enforce DPO's decisions or soften the patient apprehension.

This research did not account for any possible statistical correspondence that might occur when confronting the anonymized model's morphological structure with any additional reference. However, as stated in Article 4 of EU GDPR "Definitions" data is considered biometric when "*allows or confirm the unique identification of that natural person*". The intrinsic noise derived from the 3D scanner accuracy, the subsequent point cloud decimation, and the information deleted from the anonymization vanish or enormously decrease the possibility of any statistical identification. A question that remains open is the risk of re-identification [16][17]. GDPR defines data as anonymized if it "*can no longer be attributed to a specific data subject without the use of additional information*" however it has been proved that using generative adversarial networks (GANs), it is possible to restore the facial features of anonymized 2D images [18] without the aid of additional information. It's easy to imagine that - considering the additional information in a 3D model w.r.t. a 2D image - the same methodology applies to 3D facial data. However, depending on the anonymization strength, it's not a foregone conclusion that the individual can be uniquely identified after the re-identification. All the above opens a concreate possibility for future research in the field of 3D facial anonymization.

8 REFERENCES

1. Marti P., Goracci, C. Lampus, F. Franchi L., Children as superheroes: designing playful 3D-printed facemasks for maxillofacial disorders. In Francesca Tosi, Antonella Serra, Alessia Brischetto, Ester Iacono "Design for inclusion, gamification and learning experience", Franco Angeli, 2020. ISBN-13: 9788891797780.
2. J. Rohatgi. (2018). GDPR and healthcare understanding health data and consent, [Online]. Available: https://www.pega.com/insights/articles/gdpr-and-healthcare-understanding-health-data-and-consent
3. S. Chhabra, R. Singh, M. Vatsa, and G. Gupta, Anonymizing k-facial attributes via adversarial perturbations, 2018. arXiv: 1805.09380 [cs.CV]
4. B. Meden, Ž. Emeršič, V. Štruc, and P. Peer, "K-same-net: K-anonymity with generative deep neural networks for face deidentification", Entropy, vol. 20, p. 60, Jan. 2018. doi: 10.3390/e20010060
5. S. Mahajan, L. Chen, and T. Tsai, "Swapitup: A face swap application for privacy protection", English, in Proceedings - 31st IEEE International Conference on Advanced Information Networking and Applications, AINA 2017 , T. Enokido, H.-H. Hsu, C.-Y. Lin, M. Takizawa, and L. Barolli, Eds., ser. Proceedings - International Conference on Advanced Information Networking and Applications, AINA, 31st IEEE International Conference on Advanced Information Networking and Applications, AINA 2017 ; Conference date: 27-03-2017 Through 29-03-2017, Institute of Electrical and Electronics Engineers Inc., May 2017, pp. 46–50. doi: 10.1109/AINA.2017.53
6. P. Bagchi, D. Bhattacharjee, and M. Nasipuri, "Reg3dfaceptcd: Registration of 3d point clouds using a common set of landmarks for alignment of human face images", KI - KunstlicheIntelligenz, Apr. 2019. doi: 10.1007/s13218-019-00593-2
7. Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration", vol. 9906, Oct. 2016. doi: 10.1007/978-3-319-46475-6_47
8. D. Fontanelli, L. Ricciato, and S. Soatto, "A fast ransac-based registration algorithm for accurate localization in unknown environments using lidar measurements", in 2007 IEEE International Conference on Automation Science and Engineering, 2007, pp. 597–602. doi: 10.1109/COASE.2007.4341827
9. S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm", in Proceedings Third International Conference on 3-D Digital Imaging and Modeling, 2001, pp. 145–152. doi: 10.1109/IM.2001.924423
10. P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239–256, 1992. doi: 10.1109/34.121791
11. M. Daneshmand, A. Helmi, E. Avots, F. Noroozi, F. Alisinanoglu, H. Arslan, J. Gorbova, R. Haamer, C. Ozcinar, and G. Anbarjafari, "3d scanning: A comprehensive survey", Jan. 2018.
12. Professional 3d scanners: Artec 3d: Best 3d scanning solutions. [Online]. Available: https://www.artec3d.com/
13. Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing", arXiv: 1801.09847, 2018.
14. M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration", in VISAPP, 2009.
15. F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, "The ball-pivoting algorithm for surface reconstruction", IEEE Transactions on Visualization and Computer Graphics, vol. 5, no. 4, pp. 349–359, 1999. doi: 10.1109/2945.817351
16. B. Lubarsky. (2019). Re-identification of anonymized data, [Online]. Available: https://georgetownlawtechreview.org/wp-content/uploads/2017/04/Lubarsky-1-GEO.-L.-TECH.-REV.-202.pdf
17. L. Rocher, J. Hendrickx, and Y.-A. Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models", Nature Communications, vol. 10, Dec. 2019. doi: 10.1038/s41467019-10933-3
18. A. E. David Abramian. (2019). Refacing: Reconstructing anonymized facial features using gans, [Online]. Available: https://www.biorxiv.org/content/10.1101/447102v1.full.pdf
19. May 25, 2018). 2018 reform of eu data protection rules, European Commission, [Online]. Available: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf
20. Wang, P. (2019, February). This Person Does Not Exist. Retrieved September 13, 2020, from https://www.thispersondoesnotexist.com/