

Development of an Easy Mobile Smartphone-Based Image Capturing System for Avatar Production for Garment Manufacturing

Verena ZIEGLER, Felix HINDERER, Frauke LINK
OpenDress GmbH, Konstanz, Germany

<https://doi.org/10.15221/21.52>

Abstract:

Starting with an innovative approach for generating 4D sewing patterns, OpenDress was faced with the task of developing a pipeline that leads from an autonomous user scanning process to the development of a personalized avatar to a production procedure for 4D sewing patterns. We were looking for an approach that makes it possible to reconstruct a 3D scan from a 2D photo (see also: Ballester et. al.(2018), Saito et. al. (2019), Saito et. al. (2020), Zhang et. al. (2020)) apply AI trained movement to the generated avatar and construct the sewing pattern directly on the movement-trained avatar. The goal was to develop a pipeline, i.e. as platform-independent as possible, without special hardware for broad and easy user access. This paper presents partial results of a larger validation study that compared data results from different low-threshold and pseudo-monocular technologies to investigate the development of an ai-driven approach. The study conducted body measurements of more than 200 participants, using the Lidar sensor scan function from mobile devices and photos generated via mobile cameras in comparison to scans from the structure scanner via tablet and measurements by hand. Shape and pose parameters from the systems tested were analyzed in order to create a customized, parametric 3D model (avatar). Here the importance of physical realistic shape-, pose, and measurement parameters of the individual human body was emphasized, in order to further process the 4D sewing patterns.

Keywords: 3D, body, scanning, shape, measurement, data-driven, personalization, size allocation, low-cost, smartphone, machine learning, lidar, garment manufacturing

1. Introduction

Our project is about the disruptive expansion of fashion tech software development towards 4D patterns for mass customized clothing. The current basis is our 3D scan app and API solution for capturing individual 3D body scan data via mobile devices like smartphones, tablets or laptops. The IT interface to the eCommerce sector is used for digital fit testing of clothing in standard ready-to-wear sizes. The business model addresses the threedimensional fit testing of clothing in the eCommerce sector in order to reduce the problem of high returns rates of over 40%¹ in online retail. OpenDress aims to transform the fashion industry towards a circular, inclusive and sustainable fashion industry through various software packages and coordinated micro-factory concepts.

However, we are aware that standard clothing sizes only cover about 30% of the population in terms of fit. 70% of Germans do not fit optimally in standard sizes (SizeGermany, 2020²). It can be assumed that those who do not fit optimally in standard sizes repeatedly have difficulty finding clothing. The fit is, after all, only one parameter after a garment is picked out. So it makes sense to think about alternative ways in which mass customization for apparel can be profitable and help solve the overproduction, fit, and waste problem. We have developed an innovative 4D sewing pattern approach for digitizing the supply chain and process of mass tailoring, including motion profiles (see also: Wolff, Ziegler, Link, et. al., 2021³).

¹https://www.ehi-shop.de/image/data/PDF_Leseproben/Studie_Versand-und%20Retourenmanagement-19_Leseprobe.pdf

²SizeGermany. 2020. SizeGermany. <https://portal.sizegermany.de>. (accessed 19.09.2021)

³https://www.researchgate.net/publication/349195500_3D_Custom_Fit_Garment_Design_with_Body_Movement

By digitizing the value chain, we believe on-demand manufacturing at competitive prices will become possible within Europe. Our uniqueness compared to existing competitors worldwide is the creation of 4D cutting patterns generated directly on a digital twin (avatar), based on a 3D body scan. Once the shape of the garment is finalized for production, established techniques can be used to translate the pattern lines into a 2D sewing pattern or to translate them into a knitting pattern.

This paper focuses on the evaluation of different low-threshold and pseudo-monocular technologies, as independent as possible and without a special hardware to use, but also accessible for a broad and easy user base. This paper presents partial results of a larger validation study of over 200 participants, that compared data results from three different low-threshold and pseudo-monocular technologies against a hand measuring method with measuring harness to investigate the further development of an ai-driven sewing pattern approach based on an individual avatar. The generation of a digital twin (avatar), is the prerequisite for the application of our 4D pattern technology. We were especially interested in 3d reconstruction based on photo or video data sets and time-of-flight sensors of different designs.

2. Data collection

OpenDress was faced with the task of developing a pipeline for capturing accurate body measurements with a smartphone, tablet or laptop for the creation of avatars for use in apparel eCommerce and garment manufacturing pipeline that leads from an autonomous user scanning process to the development of a personalized avatar to a production procedure for the 4D sewing patterns. The goal is to develop a pipeline, i.e. as platform-independent as possible, without special hardware for a broad and easy user base access. In a larger validation study we compared data results from different low-threshold and pseudo-monocular technologies to investigate the development of an AI-driven approach. The study conducted body measurements of more than 200 participants, and this paper focuses on four different approaches we used throughout this study: The first approach tested was the Lidar sensor scan function from mobile devices; The second approach was photos generated via mobile cameras; The third approach was using the structure scanner via tablet; The fourth approach was taking measurements by hand with a special measuring utensil, which was created by Inge Szoltysik-Sparrer, a German master-tailor⁴. Shape and pose parameters from the systems tested were analyzed in order to create a customized, parametric 3D model (avatar). Here the importance of physical realistic shape, pose, and measurement parameters of the individual human body was measured, in order to further process the 4D sewing patterns.

2.1 Data-driven reconstruction procedures of 3D bodies

In particular, this paper presents three 3d data-driven reconstruction procedures of low-threshold and pseudo-monocular technologies compared against hand measured data with measuring harness:

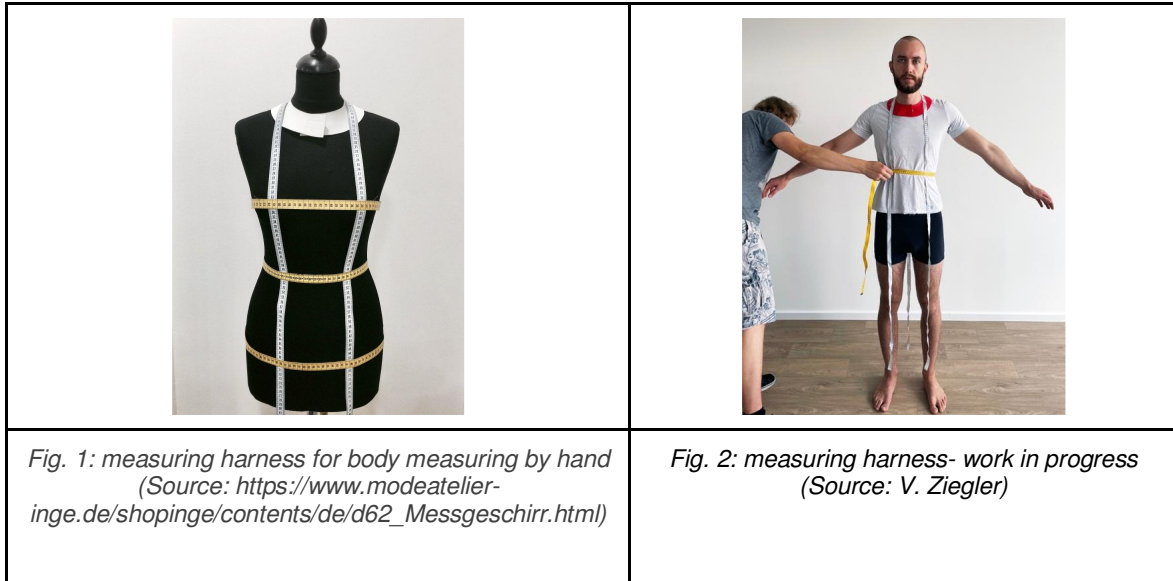
2.1.1 *Measurement by hand (with measuring harness)*

The measuring utensil we used is a collar with 3 tape measurements attached. Two of them laying over the breast points and the third at the back alongside the spine (cf. Fig. 1).

A measuring harness is a neck brace to which 3 measuring tapes are attached, two that run down both sides of the chest at the front and one that runs down the spine (see Fig. 1 and 2). The chest circumference, waist and hip circumference as well as the back height, back length and hip depth are measured.

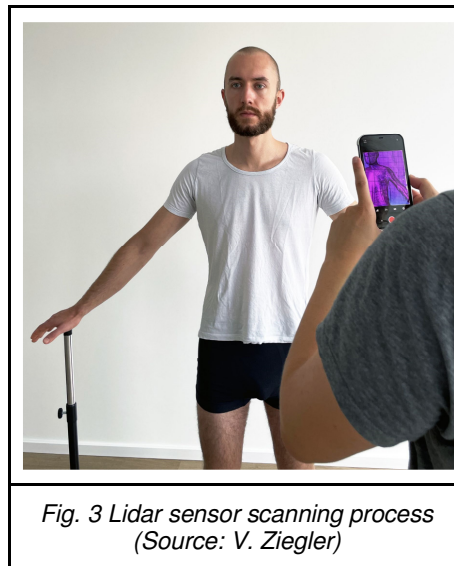
These hand-measured values are used in the following as ground truth, since these measurements are very accurate due to the manual procedure used to document them.

⁴ https://www.modeatelier-inge.de/html/inge_szoltysik-sparrer.html (accessed 18.09.2021)



2.1.2 Lidar sensor scan function from mobile devices

As a new, low-threshold and pseudo-monocular technology, the lidar sensor installed on the iPhone 12 is a promising technology for measuring space. It uses infrared light and time-of-flight (ToF), i.e. the time it takes for the light to travel from the phone to the object and back again, to generate a 3D depth image. Point clouds can then be further added into surfaces using algorithmic processes.



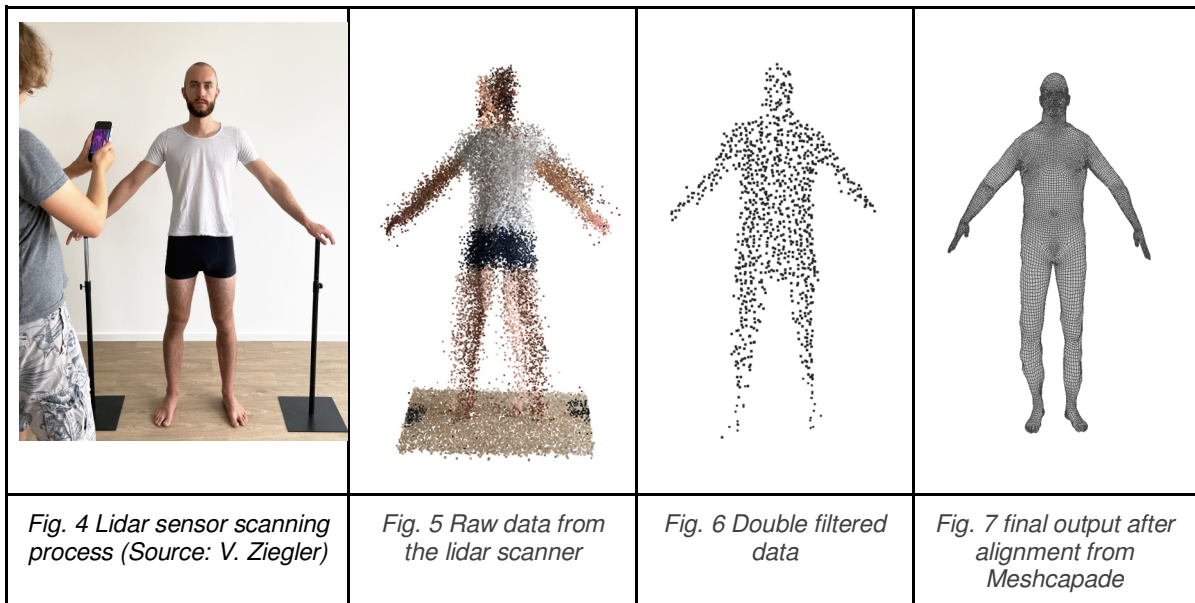
The aim of the investigation testing the Lidar sensor in the iPhone 12 was to estimate the precision in comparison to the Structure Sensor, which already provides very good 3D body meshes in combination with the ready built in software.

Since it was also of interest to see the difference between the lidar and the structure sensor, since the lidar is already used in mobile phones and the structure is an expensive piece of hardware, the lidar models were positioned and rotated by hand to the structure data as best as possible.

Afterwards the scans of the Lidar were double filtered in a standard way.

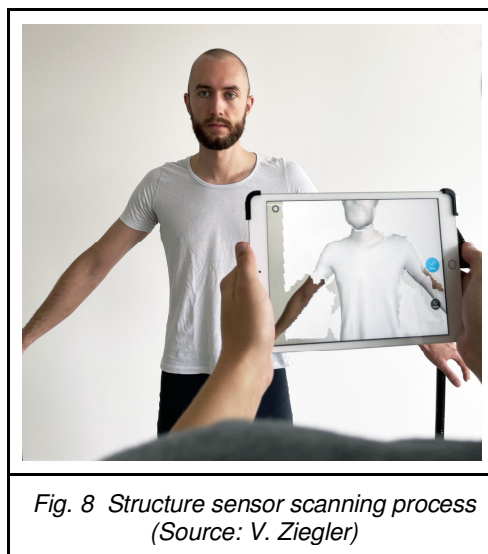
1. Discarding points whose n-neighbors exceed a certain distance, i.e. all points flying around in the air are discarded. These come from the inaccuracy already described. This is probably caused by the hardware, because the sensor is only meant as supporting hardware to the camera.
2. Clustering the points where the n-neighbors are averaged and combined into one point. This is done to smooth the Implicit Surface of the model, since this surface is very inaccurate.
3. Finally, the filtered scans were passed to Meshcapade's ganymede service, which then converted the point cloud into a correctly positioned and aligned SMPL model, where one can then read out the relevant sizes (chest circumference, waist, hips) with little effort.

In Fig. 4 - 7 is an overview of the lidar data generation process.




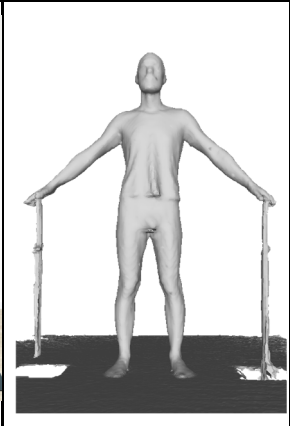
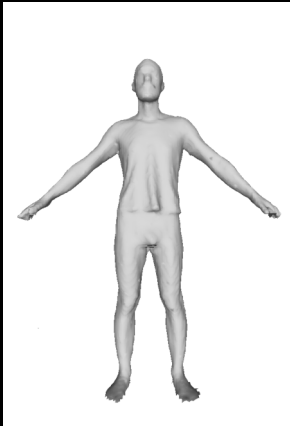
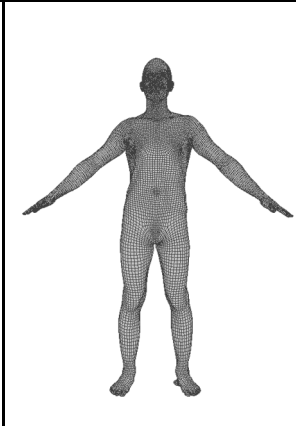
2.1.3 Scans from Structure Sensor via tablet

Similar to the Lidar Sensor, the Structure Sensor also uses infrared light. However, a special pattern of points is generated via their position and distortion to find out how the observed surface is shaped. This information is then combined to form a mesh.



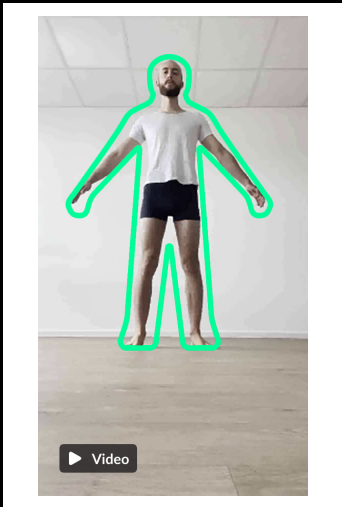
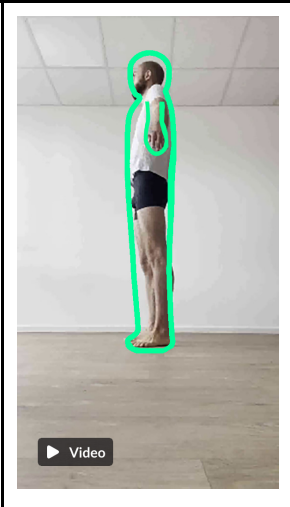
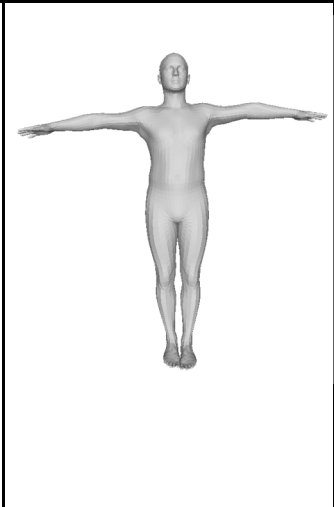

However, since there are still various foreign objects in this scan, such as the hand supports and in some cases the bottom as well as artifacts that have formed in the body, these were all cleaned again by hand.

Again, this further processed data was sent on to Meshcapade who position, align and convert an SMPL model from which the data can then be easily read.

			
<p><i>Fig.9 Structure sensor scanning process (Source: V. Ziegler)</i></p>	<p><i>Fig.10 Input Scan</i></p>	<p><i>Fig.11 cleaned up Mesh</i></p>	<p><i>Fig. 12 Alignment from Meshcapade</i></p>

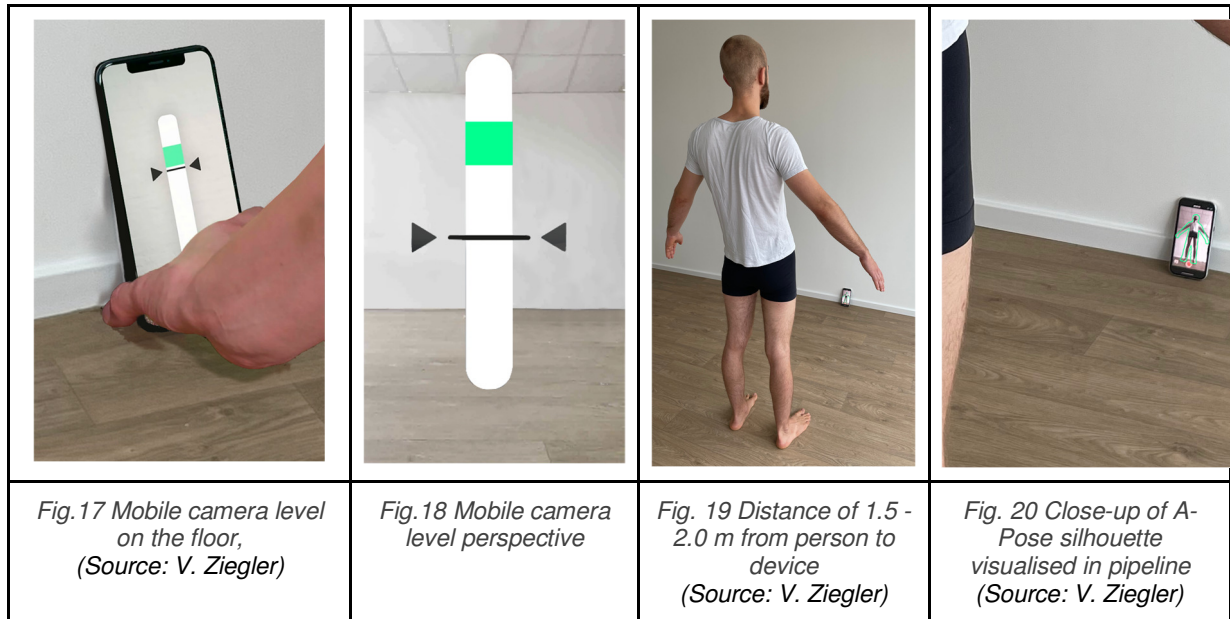
2.1.4 Developed Pipeline

To reconstruct an autonomous user scanning process without specialized hardware for a broad and simple user base, we developed an AI pipeline that reconstructs a 3D object from two frames via a photogrammetry process obtained from a video recording, possible with any mobile device and based on a single person.

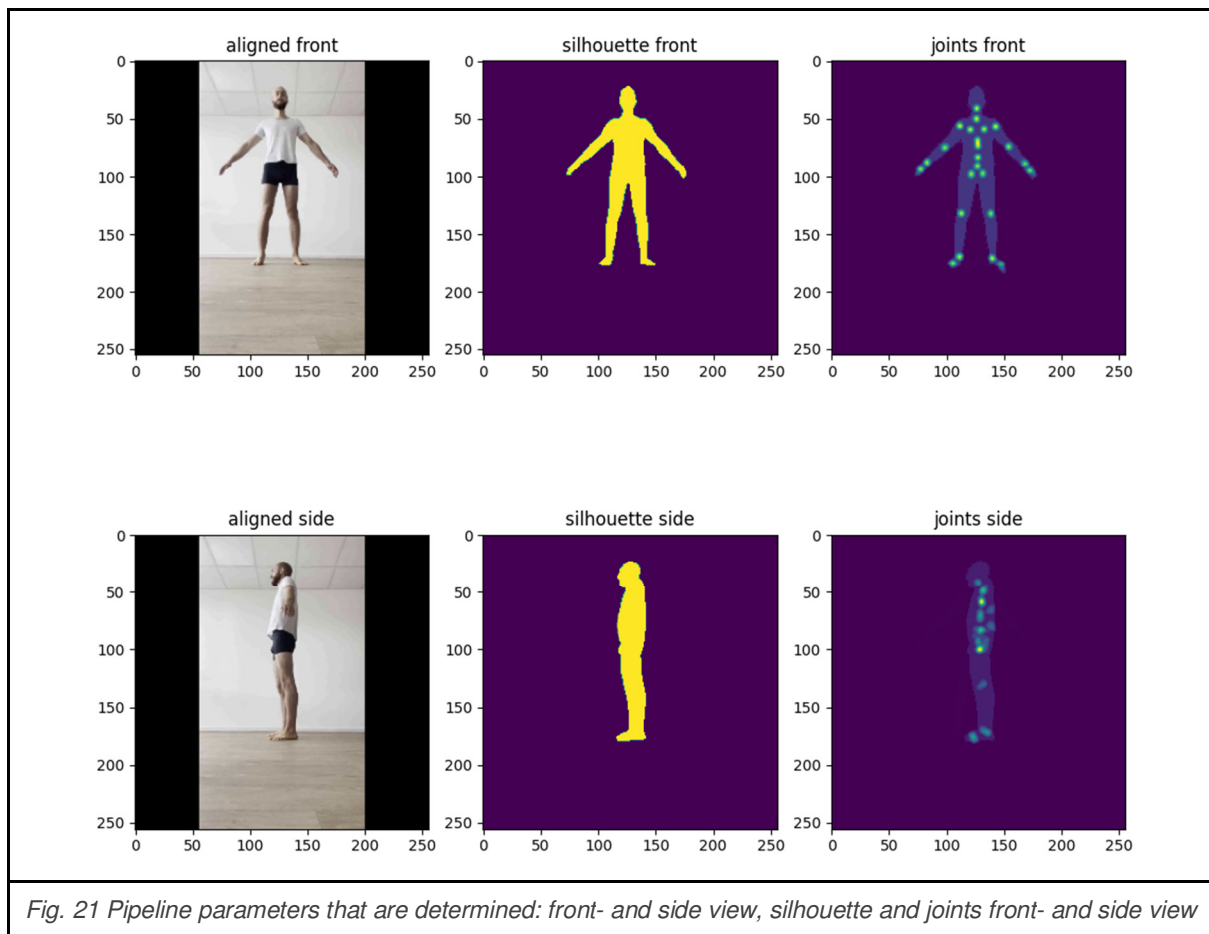
			
<p><i>Fig.13 Camera perspective front view</i></p>	<p><i>Fig.14 Camera perspective side view</i></p>	<p><i>Fig.15 Reconstructed avatar model in T-pose</i></p>	<p><i>Fig. 16 Reconstructed avatar model in T-pose</i></p>

First, multiple views are extracted from a video using an advanced pose detection algorithm. Each of these RGB views is then passed through a segmentation network to generate the silhouette of the person as a binary mask.

Then, for each of these silhouette images, the joints of the person's skeleton are detected.



Based on the detected joints and the generated silhouettes per view, a neural network estimates the pose and shape parameters of the captured person. With these estimated parameters, we can create a 3D body mesh and finally measure all body distances.



2.2 Quality of comparison

For us, the basic data were the hand measurements, the results of the structure sensor scan, the results of the lidar scan, and the 3d data that came from our pipeline. The hand-measured values were used as ground truth. While these are relatively prone to error, they best reflect the real data even under clothing.

2.2.1 Ground truth & distance function

For the comparison, we decided to collect the body height on the one hand, and the bust, waist and hip measurements on the other. Distance measures are therefore differences in lengths, i.e. circumferences.

2.2.2 Comparability

In order to establish comparability between the digital datasets, we decided to create an alignment for lidar and structure sensor data using a SMPL tool⁵. This is a reliable, simple, and open way to generate meshes from a wide variety of human 3d data. Using the intelligence of SMPL did not work well for our data to the same degree. 3d scans, which were already available in (very) good quality, were reconstructed accordingly. For the lidar data there were often difficulties despite cleaning. Therefore, the alignments that were optically human-shaped were compared in the measurements.

2.2.3 Reliability

The measurement of living beings is essentially different from the measurement of static 3d objects. People are constantly in motion: Not only does active movement change body height and circumferences, but simply breathing in and out can also change the circumference of the chest by several centimeters. The question of how upright you stand also depends on your daily constitution. A comparison of the measured values at a later time is therefore impossible. The measurement is a snapshot. In our case, all measurements were taken in a time window of up to 15 minutes. The diversity of the procedures ensured that "outliers" in the data were detected. However, the reliability of the measured values also differs with the measurement procedure.

For the hand-held measurement method, the objectivity of the measurement was increased *a priori* by using the measurement harness and conducting a training session in which classic difficulties (deciding where to take the measurement, tightening the measuring tape, the person's posture) were reduced in order to produce uniform solutions.

The Structure Sensor and the Lidar Sensor were used in parallel to capture the same body posture.

3. Evaluation of data-driven reconstruction

More than 200 people participated in the study. However, to ensure greater measurement accuracy, we decided that we would always evaluate all methods on the data from participants wearing tight clothing (about 90-100 data points in total).

In the first section of the results analysis, we discuss the body size data across all methods.

⁵ www.meshcapade.com

3.1 Body height

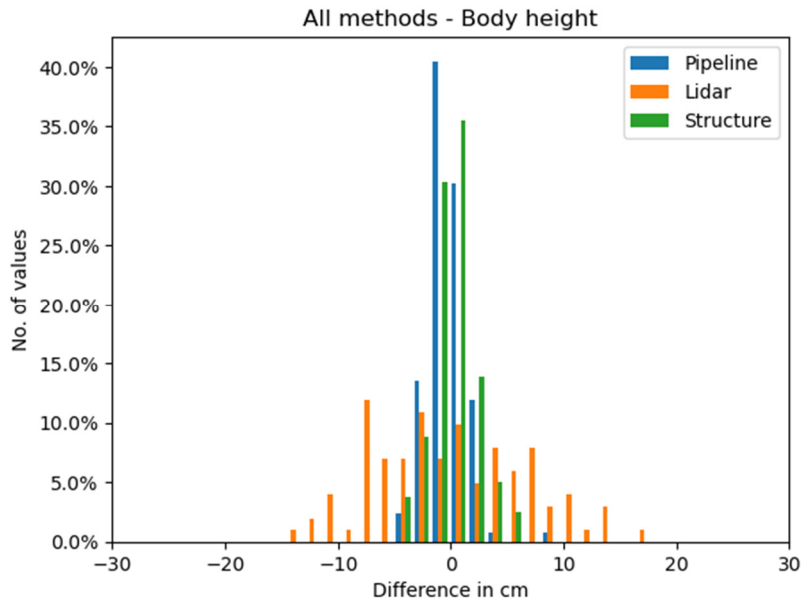


Figure 22- All methods - Body height - Difference to ground truth in cm

The structure sensor and output of our pipeline each form a cluster around 0 with a Gaussian distribution, while the height predicted by lidar varies greatly.

The large difference in height with lidar can be explained by the ground truth. The lidar reconstructs the point depth in a cloudy way. The points lie distributed on the ground like a “flokati carpet”. Automatic removal is difficult because the feet are woven into this “carpet” and individual points even lie outside the carpet cloud. We assume that either the upper or the lower limit of the “carpet” height is taken as a basis for the automatic cleaning of the lidar data. Outliers sometimes appear on data which maybe wasn’t trained by the AI yet, for example very small or very large people.

3.2 Circumferences

In the second section we compared the circumferences of the 3d-scans at the chest, waist and hip levels. In figure 23 we see the difference to the ground truth. We see a Gaussian distribution for all three methods with regard to the ground truth, but with different positions. Lidar and pipeline are centered around zero, the structure data are shifted to the right.

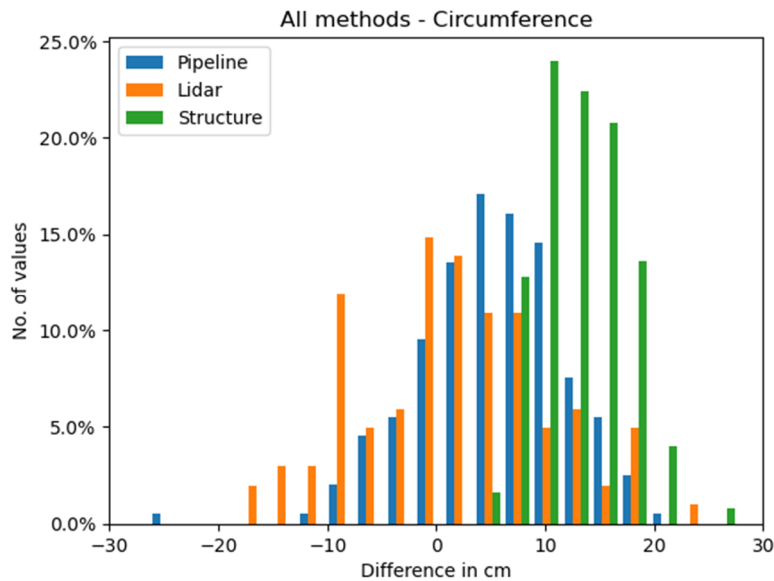


Figure 23- All methods - circumference in cm

The corresponding values can be seen in table 1. The average of the structure sensor is 12.94 cm, for our pipeline it is 5.18 cm. This corresponds to an average body spacing of 20.6 and 8.2 mm. Structure data show the distance of the (still) existing clothing above the ground truth. Lidar data is difficult to interpret. Basically, measurements are also made here on the surface of the body (such as structure sensor meshes). The fact that the result here seems to be better compared to the structure data can, for example, be due to the fact that the filtering reduces the spread too aggressively and thus the overall scope is reduced. In fact Lidar data only seems to be better than structure data or equal to pipeline data. In reality, the lidar sensor data method varies a lot in quality. Our pipeline is set up in such a way that it has “learned away” clothing, meaning it accounts for the additional bulk apparel adds to the body. Because of that it fits better to the ground truth than the structure sensor data and the lidar sensor.

Table 1 - All methods - circumference in cm

Circumference - All	68% (1σ)	95% (2σ)	99% (3σ)	Average
Lidar	±8.57	±17.14	±25.71	1.58
Structure	±4.02	±8.04	±12.06	12.94
Pipeline	±6.68	±13.36	±20.04	5.18

The graphical differences and comparisons in male and female data sets can be found in the appendix. A look at the tables (table 2 and 3) shows that the mean value (average) for the male data sets are slightly better in our developed pipeline (table 2 to table 1) and for the female the mean value lies in the structure data method (table 3 to table 1).

Table 2 - All methods - circumference in cm (male only)

Circ - Male	68% (1σ)	95% (2σ)	99% (3σ)	Average
Lidar	±9.21	±18.42	±27.63	1.69
Structure	±3.80	±7.60	±11.40	13.99
Pipeline	±7.28	±14.56	±21.84	4.10

Table 3 - All methods - circumference in cm (female only)

Circ - Female	68% (1σ)	95% (2σ)	99% (3σ)	Average
Lidar	±7.97	±15.94	±23.91	1.48
Structure	±3.99	±7.98	±11.97	12.15
Pipeline	±6.05	±12.10	±18.15	6.00

In order to get a better insight into how much the deviations in the scope improve, taking into account the deviation from the height, we have determined the total error in figure 24 (and following).

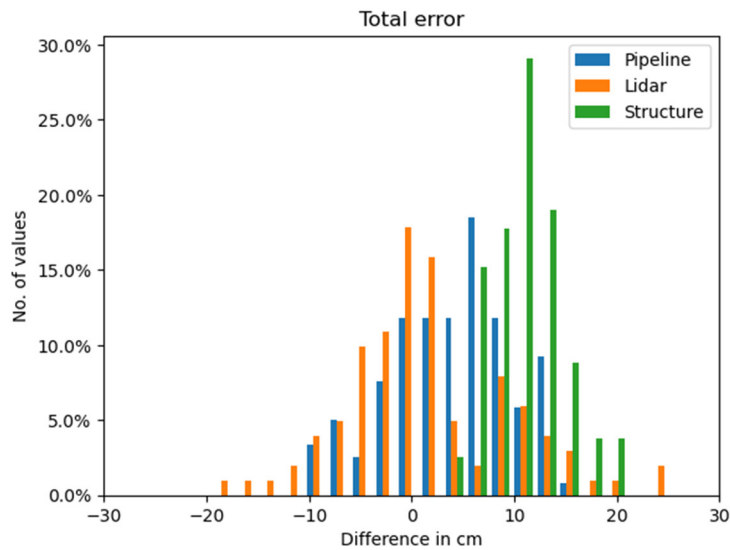


Figure 24- All methods - Total error

The relationship between the hand-measured values and the estimated values of the three different low threshold procedures were determined for each case. These values were compared to the respective scope values.

Table 4 shows a slight improvement (compared with table 1) in the mean value for our pipeline and the structure sensor data.

Table 4 - Total error

Total Error	68% (1σ)	95% (2σ)	99% (3σ)	Average
Lidar	± 7.97	± 15.94	± 23.91	1.64
Structure	± 3.45	± 6.90	± 10.35	10.96
Pipeline	± 5.92	± 11.84	± 17.76	3.92

3.2.2 Closer look at our pipeline

With regard to our pipeline, we're examining the data in this section even more closely.

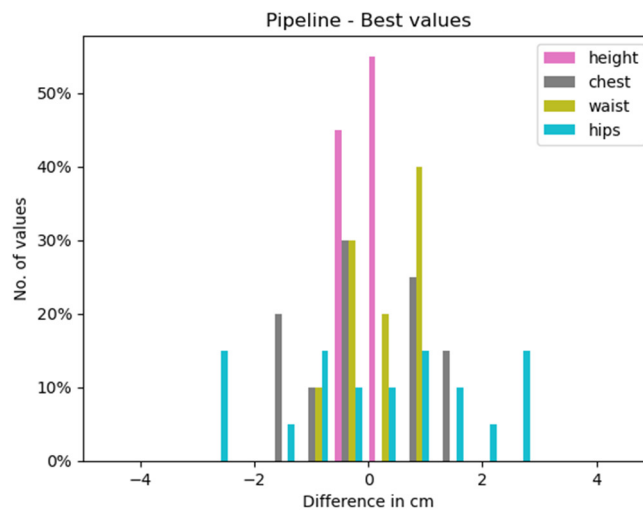


Figure 25- Best values pipeline

In figure 25 we see the 20 best values in the pipeline for height, chest, waist and hips. It can be seen that the deviations are only small in this ideal case. Reasons for deviations cannot be clarified in detail here. Typical sources of error for photogrammetric methods like ours are the variation of the distance to the camera, alignment of the camera, posture while moving, ground truth problems and much more.

The Pipeline performance is split into male and female data: When dividing further into the male and female data sets, there are a few interesting connections to see. The male data set is more evenly distributed, while the female data set is more Gaussian.

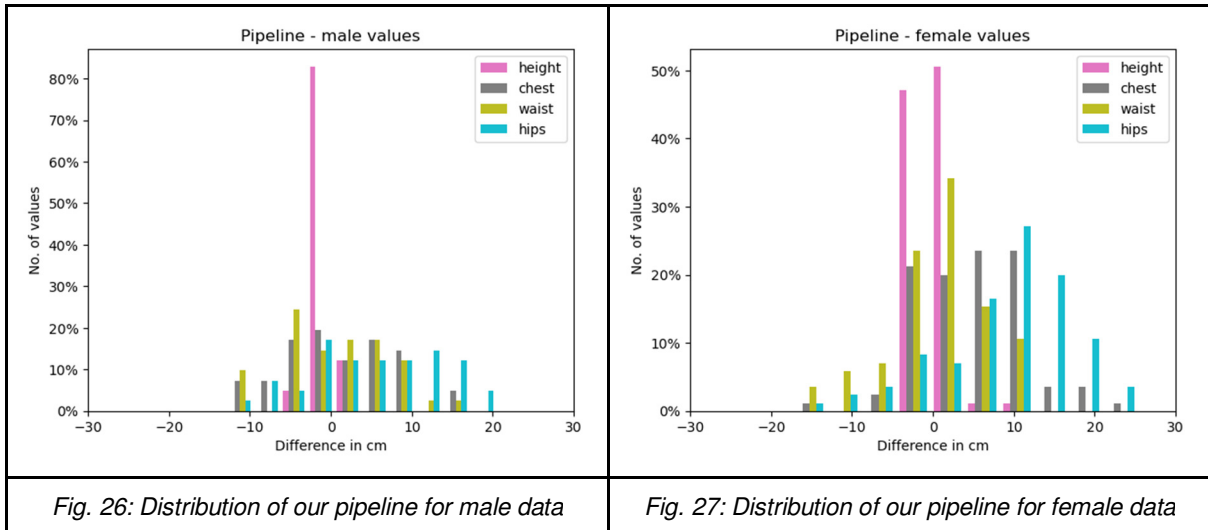


Fig. 26: Distribution of our pipeline for male data

Fig. 27: Distribution of our pipeline for female data

In the female data set, it is noticeable that only the hip data is shifted in the mean. We assume that this either has to do with the perspective of the smartphone or with aspects of movement that might be different for women than for men.

3.2.3 Closer look at structure sensor vs. pipeline data

As expected, the structure data set performance was good. Corresponding figures can be found in the appendix. All deviations can be explained by tight clothing compared to the ground truth, be it for example tight t-shirts or boxer shorts. We were all the more interested in looking at the discrepancy between the structure sensor data and the pipeline data, as shown in the following figure 29.

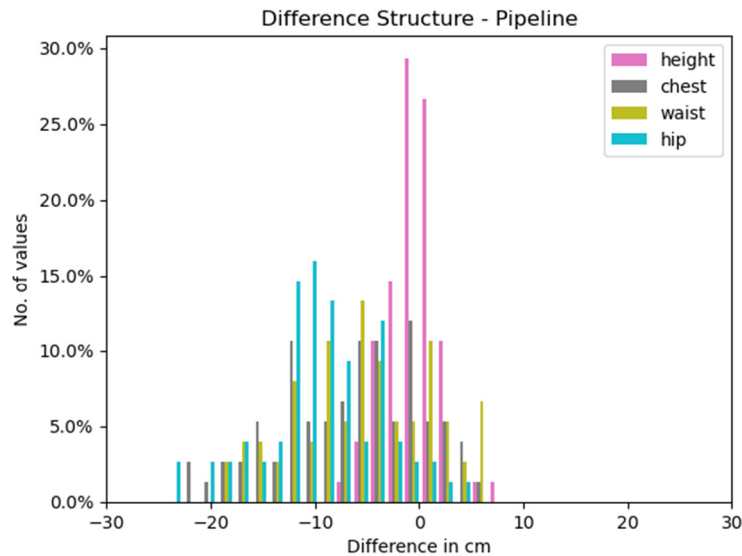


Fig 29 - pipeline vs. structure data

Both methods (structure and pipeline) determine the height very precisely, accordingly the inaccuracy is also very little and centered around 0. Here, the structure sensor data was taken as ground truth and the pipeline was compared against it. The general shift into the negative realm can be explained by the fact that our method is a little more varying, but is generally more precise (more centered around 0 when compared to the ground truth). The shift of the mean value to the left can be explained by the people wearing clothes, as the structure sensor cannot simply get rid of them. In fact our pipeline detects clothing and has learned to subtract the clothes, but the structure sensor is not able to do so.

Large deviations are only additive (largely due to the structure sensor), as the pipeline underestimates the body heights in these cases and the structure sensor in means has more bodyheight than the ground truth.

The two methods both perform very similarly and also relatively accurately, our developed pipeline method even seems to perform slightly better because the center of the distribution is closer to 0 than that of the structure sensor (5cm - 7cm vs. 10 + cm). Of course, one can argue that one simply subtracts a constant value which then makes the results of the structure sensor better because it has a lower variance, but this offset does not necessarily apply to all groups of people.

3.3 Evaluation of lidar sensor data

We were also very interested in a more detailed analysis of the lidar sensor data, since iPhones pro max 12 and later have a lidar sensor built in, but public perception is that it is insufficient for most 3D scanning purposes.

3.3.1 Lidar data vs. ground truth

The following figure shows the lidar data in relation to the ground truth (measured values by hand).

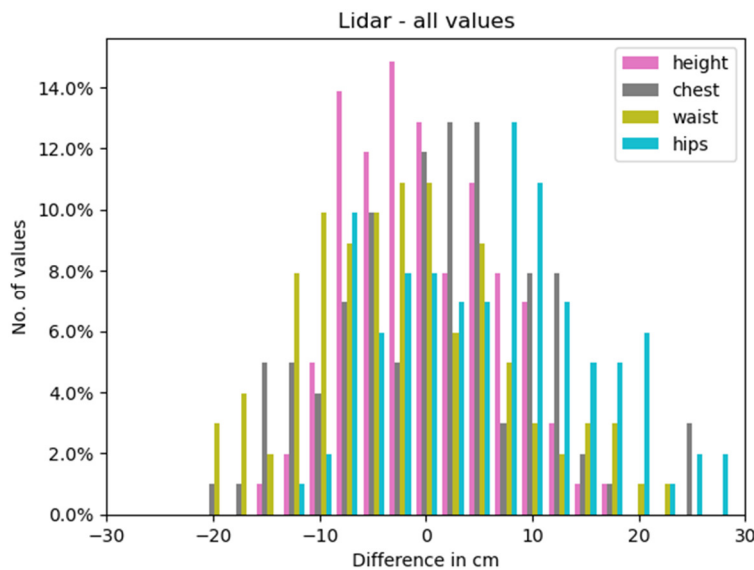


Fig 30 - lidar sensor data

We see a comparatively confusing behavior here, which is not centered and is relatively broadly spread. This is related to the points already listed: Disturbances arise through light, through movement, in the cleaning process, through irregularities on the floor and through the subsequent unsafe application in the SMPL generation. One factor is the ratio of tight clothing to hand-measured body measurements. To rule this out, we decided to change the ground truth and compare the lidar data with the structure data, as we will show in the following.

3.3.2 Lidar data vs. structure data

We collected the lidar sensor data parallel to the structure sensor data, i.e. at the same time but with different devices. This forced us to manually superimpose the body scans. *A priori* accuracy was lost here, up to 0.5 degrees of rotation and up to 5 mm of displacement.

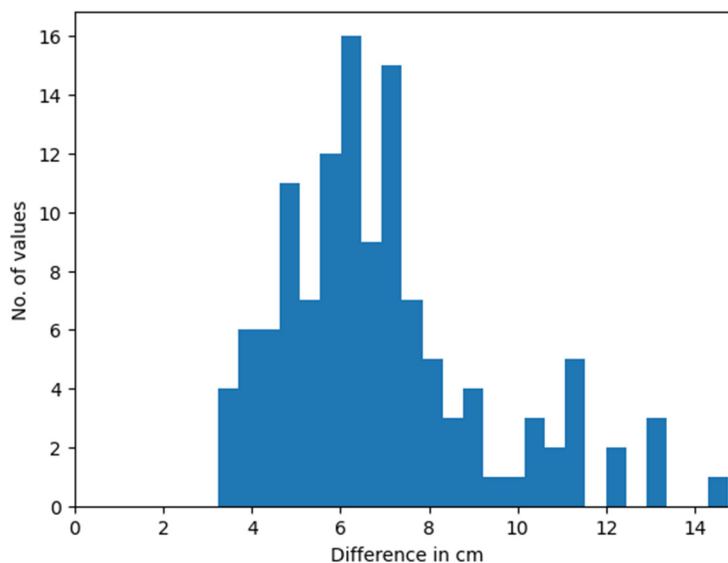


Fig 31 - Hausdorff distance between lidar and structure data

In figure 31 you can see the Hausdorff distance of the superimposed 3d scans. Since the Hausdorff distance reflects the maximum distance, this explains the absence of values smaller than 3. Larger deviations were examined in detail by looking back at the specific data. There were some systematic deviations here, which are explained below.

Existence of arm rests: As explained at the beginning of this article, we used hand rests to avoid movement of the arms during the scanning process. The armrests were metal tubes anchored in the floor, which inherently had two properties that were problematic for scanners: shape and color. The narrow shape can lead to data loss during scanning if the scanning width is wider than the object itself. The reflective surface interferes with the scan signal, just as it does with glasses or window panes.

Handling the armrests: In the data processing following the data collection, the structure data was cleaned manually. Among other things, the armrests were also removed here. The lidar data instead was cleaned using an algorithm. Here it was not always possible to remove the armrests completely.

In conclusion, one can say that the lidar sensor can sometimes achieve good results, but also bad results—the distribution of the results is more reminiscent of a uniform distribution that is also quite wide. The main reasons for this are the "cloudy" outputs generated by the lidar sensor as well as the unclear way in which the 3D clouds are put together (it sometimes seems as if these are only registered by odometry vs. photogrammetry, where the position and orientation of the camera is relatively precise (this phenomenon manifests itself mainly in degenerate point clouds where the body has collapsed or has two layers) which may even be caused by the software used. But since this can not be the only factor, other variables certainly also play a role, such as the lighting of the room, disruptive external light (in fact, sunlight—the time of day—may even play a role). Even small movements can be a problem.

4. Conclusions and future work

Generating data is a lot of work, but it also creates new insights and helps to relate things to one another and to reveal practical gaps in theoretical approaches. In addition to the results of this article, we were able to gain experience and get well acquainted with participants during our data collection and discuss with them fears and reservations about data processing of body data, which helps us a lot for future developments. It was foreseeable that the body height could be easily determined. Our pipeline is working well. The Lidar sensor delivers data that is currently no better than our developed pipeline, but it is a technology that should not be neglected. In any case, we see great potential in the lidar scanning technology - although this is not yet fully realized due to unreliable hardware and unreleased software updates.

Even if the lidar sensor technology in smartphones is quite imprecise right now, with more updates it has a great potential to become the future body scanning technology. For our liking, the Structure sensor is too expensive for everyday consumers, but has high accuracy and overall good results. For our use-case the pipeline in comparison to the Structure Sensor performs better, because the values are closer to the ground truth (hand measurement data with measuring harness). The quality of our pipeline so far is a work in progress and improves continuously, but already performs better (is more precise) than the Lidar sensor overall.

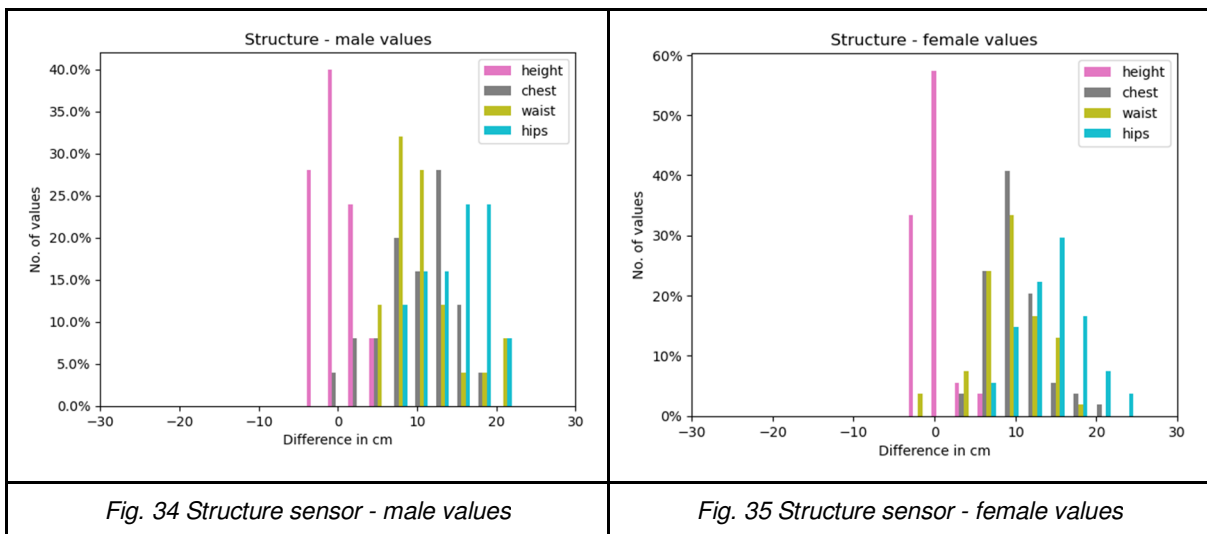
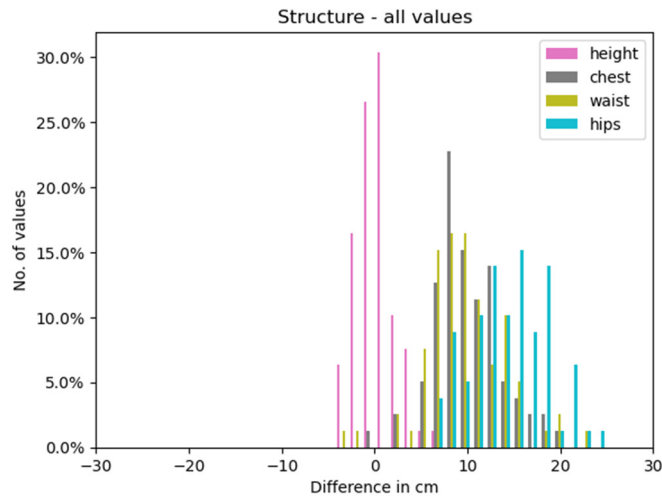
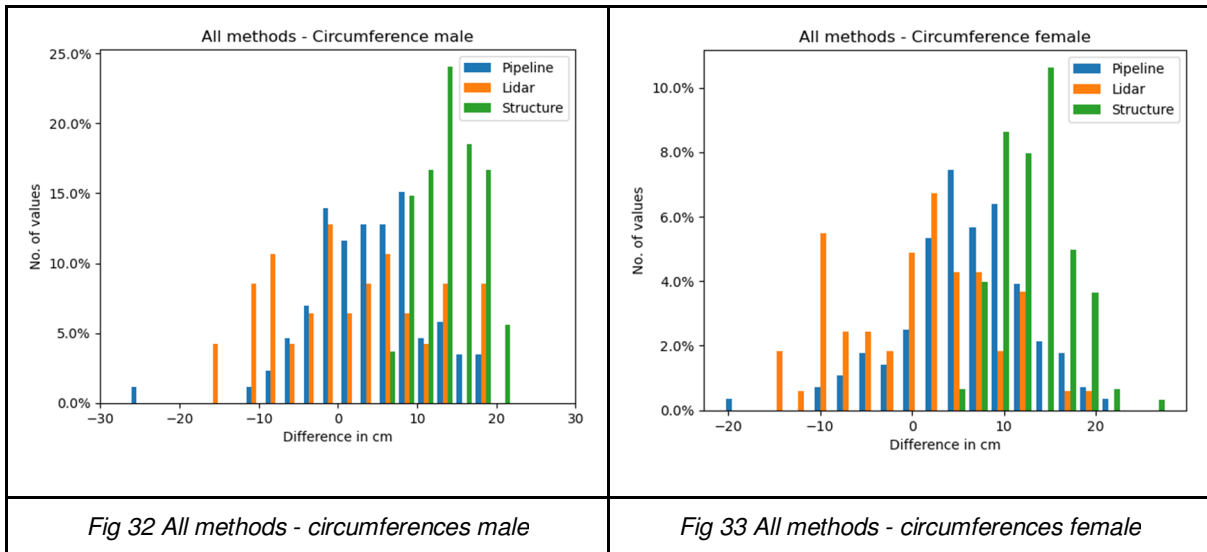
5. Acknowledgements

We would like to thank all the willing participants who took part in our Scan Study 2021 in Konstanz. We want to thank the HTWG Konstanz (HTWG - University of applied sciences Konstanz) for the support, especially a big thank you to Prof. Dr. Georg Umlauf and his team from the Institute for Optical Systems (IOS). We would like to thank the rest of our team, especially Henning Krause and Nico Brügel for their support. Last but not least we want to thank Leonard Schwob for being our 3d-scanning expert and male photo model for this paper.

6. References

- Alfredo BALLESTER*, Ana PIÉROLA, Eduardo PARRILLA, Jordi URIEL, Ana V. RUESCAS, Cristina PÉREZ, Juan V. DURÁ, Sandra ALEMANY, *3D Human Models from 1D, 2D & 3D Inputs: Reliability and Compatibility of Body Measurements*. In Proceedings of 3DBODY.TECH 2018 9th Int. Conference and Exhibition on 3D Body Scanning and Processing Technologies, Lugano, Switzerland, 16-17 Oct. 2018, Instituto de Biomecánica de Valencia (IBV), Universitat Politècnica de València, Valencia, Spain
- Shunsuke SAITO, Zeng HUANG, Ryota NATSUME, Shigeo MORISHIMA, Angjoo KANAZAWA, and Hao LI. 2019. *PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization*. In Proceedings of the IEEE International Conference on ComputerVision. 2304–2314.
- Shunsuke SAITO, Tomas SIMON, Jason SARAGIH, and Hanbyul JOO. 2020. *PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.84–93
- Ligang Liu, Lei ZHANG, Yin XU, Craig GOTSMAN, and Steven J. GORTLER. 2008. *A local/global approach to mesh parameterization*. In Proceedings of the Symposium on Geometry Processing (SGP '08). Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 1495-1504. <https://dl.acm.org/citation.cfm?id=1731336>
- Rohan SAWHNEY and Keenan CRANE. 2017. *Boundary First Flattening*. ACM Trans. Graph. 37, 1, Article 5 (December 2017), 14 pages. DOI: <https://doi.org/10.1145/3132705>
- Yousuf SOLIMAN, Dejan SLEPCEV, and Keenan CRANE. 2018. *Optimal cone singularities for conformal flattening*. ACM Trans. Graph. 37, 4, Article 105 (July 2018), 17 pages. DOI: <https://doi.org/10.1145/3197517.3201367>
- Olga Sorkine, Daniel Cohen-Or, Rony Goldenthal, and Dani Lischinski. 2002. *Bounded-distortion piecewise mesh parameterization*. In Proceedings of the conference on Visualization '02 (VIS '02). IEEE Computer Society, Washington, DC, USA, 355-362.

7. Appendix - Pictures



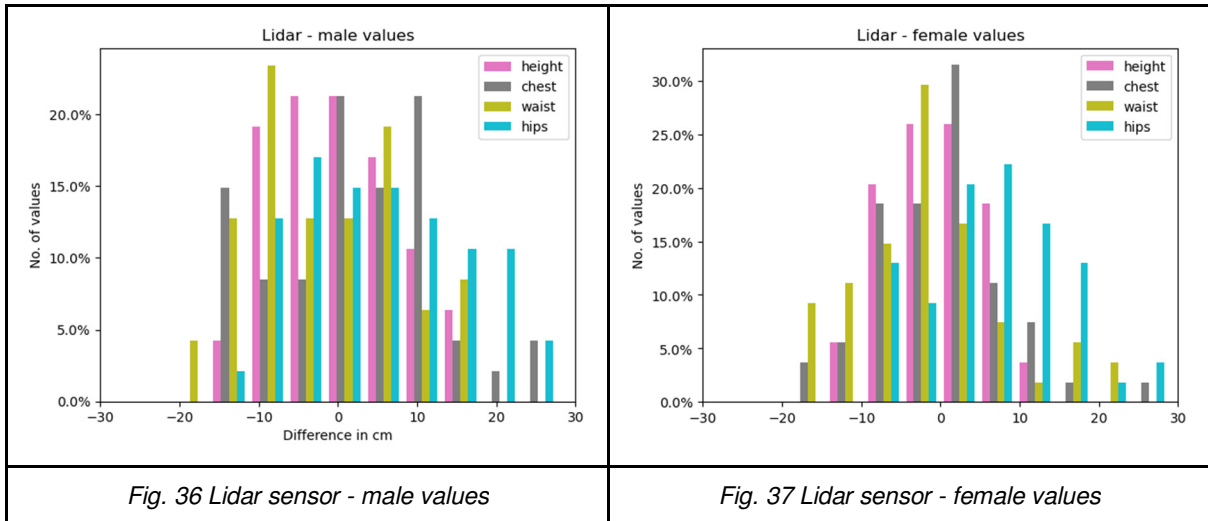


Fig. 36 Lidar sensor - male values

Fig. 37 Lidar sensor - female values