Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

A Review of Body Measurement using 3D Scanning

KRISTIJAN BARTOL, DAVID BOJANIĆ, TOMISLAV PETKOVIĆ and TOMISLAV PRIBANIĆ

University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Electronic Systems and Information Processing (e-mail: name.surname@fer.hr)

Corresponding author: Kristijan Bartol (e-mail: kristijan.bartol@fer.hr)

This work has been supported by the Croatian Science Foundation under the project IP-2018-01-8118.

ABSTRACT The understanding of body measurements and of body shapes in and between populations is important and has many applications in medicine, surveying, fashion industry, fitness, and entertainment. Body measurement using 3D surface scanning technologies is faster and more convenient than traditional measurement methods and at the same time provides much more data which requires automatic processing. A multitude of 3D scanning methods and of processing pipelines have been described in the literature and the advent of deep learning based processing methods has generated an increased interest in the topic. Also, over the last decade larger public 3D human scanning datasets have been released. This paper gives a comprehensive survey of body measurement techniques with an emphasis on 3D scanning technologies and on automatic data processing pipelines. An introduction to three most common 3D scanning technologies in body measurement, passive stereo, structured light, and time-of-flight, is provided and their merits w.r.t. body measurement are discussed. Methods described in the literature are discussed within the newly proposed framework of five common processing stages: preparation, scanning, feature extraction, model fitting, and measurement extraction. Synthesising the analysed prior works, recommendations on specific 3D body scanning technologies and accompanying processing pipelines for the most common applications are given. Finally, an overview of about 80 currently available 3D scanners manufactured by about 50 companies as well as their taxonomy regarding several key characteristics is provided in the Appendix.

INDEX TERMS body measurement, 3D surface scanning, body shape, anthropometry, deep learning

I. INTRODUCTION

Anthropometry, a subfield of applied metrology, is the study of how to measure humans. In general anthropometry includes the complete process of data collection, documentation, summarization, and analysis [174]. In a more narrow sense anthropometry can be defined as the science of *body measurement* where lengths, breadths, heights, and circumferences are used to numerically describe body segments and overall body shape [11]. Body measurement is essential in quantifying the variations in population and between populations of different countries, ethnicities, cultures, and ages [26], [137], and it strongly impacts medicine [46], [68], surveying [56], [174], fashion industry [174], fitness [32], and entertainment [38].

Body dimensions may be measured in various ways, e.g. they can be obtained manually using traditional tools such as calipers and tape measure [174], or automatically using 3D scanners and then extracting the measurements from the obtained data. To ensure both the comparability and the repeatability, body measurements are standardized via definition of measurement postures and of body landmarks [74], [75]. Although manual measurement is the gold standard, several reports suggested that human expert measurers and 3D scanners achieve comparable accuracy and that the repeatability is generally better for 3D scanners [82], [85], [109]. Another advantage of using 3D scanners over expert measurers is the measurement speed [91]: duration of an automatic scan is often under few seconds and may go up to half-a-minute for high-quality scans[1]. Therefore, although first commercial 3D body scanners appeared around 1990s [39] and were expensive, required trained personnel and ex-

[1]See Appendix A for more details

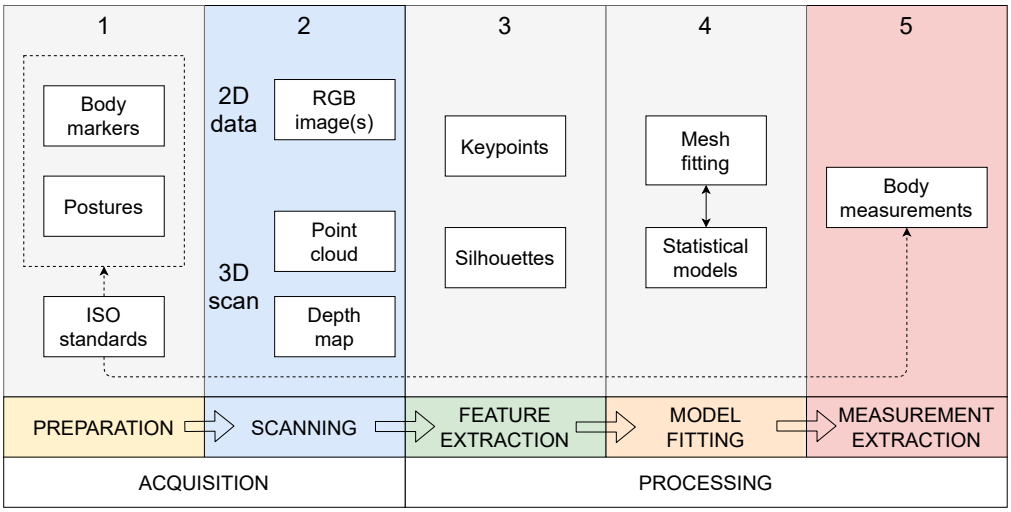


FIGURE 1: The proposed body measurement framework. All measurements methods include scanning and measurement extraction stages and one or more of the remaining stages.

tensive manual postprocessing [114], the scanning technology is currently mature and is comparable in performance to human measurers [24], [64], [175].

There are three commonly used scanning technologies for human body data acquisition: (a) passive stereo (PS); (b) structured light (SL); and (c) time-of-flight imaging (ToF). PS uses images from multiple viewpoints to reconstruct the 3D body surface using the triangulation principle [63]; it fails in the case of low or no texture. SL extends the PS approach by projecting known light patterns, which mitigates the main drawback of PS. In SL 3D body surface is reconstructed from the deformations of the projected light pattern [54]. Regarding SL, we distinguish between projector- and laser-based methods. In ToF modulated light is projected onto a person and the 3D body surface is directly obtained by measuring the travel time of modulated light [70]. Considering the multitude of data acquisition and processing methods which have been described in the literature there exists an increased interest in the topic which is also substantiated by many large public 3D human body datasets [3], [8], [24], [64], [175] released over the last decade.

In this work we provide a comprehensive review of body measurement based on 3D scanning starting from a review of 3D scanning technologies and ending by describing the most recent advances in pose and shape estimation. We propose to divide the body measurement processing pipeline into five stages, (1) preparation, (2) scanning, (3) feature extraction, (4) model fitting, and (5) measurement extraction (Fig. 1). In the preparation (stage 1), markers which identify standard body landmarks may be placed on the body [3], [24], [166]. The person is asked to take a pre-defined pose [75] and to hold still until the scan ends. Scanning (stage 2) produces a 3D point cloud or depth map(s), along with the set of images, if RGB cameras are used. In stage 3, features such as keypoints and silhouettes are extracted from a 3D scan and images. Based on the features or raw image data [84], in stage 4, the optimal human 3D template mesh[2] is estimated. The primary advantage of fitting the template mesh (a model) to the 3D scan is that any measurement may be easily and conveniently determined from the semantics of the model. Mesh fitting techniques enable the creation of statistical body models, as described in Sec. IV-D. The statistical models enable template mesh regression directly from images and image features. Finally, body measurements are extracted from the processed data (3D scan, images, features, and template mesh) in stage 5. Note that stages 2 and 5 are mandatory, while stages 1, 3, and 4 are optional.

The remainder of this paper is structured as follows: Prior review works on body measurements and on 3D scanning are briefly listed in Sec. II. Three most common scanning technologies (PS, SL and ToF) are described in Sec. III. Proposed body measurement framework and the five processing stages are introduced and discussed in more detailed in Sec. IV. In Sec. V a methodology for comparison of reviewed methods is described and methods are discussed w.r.t. their limitations and achieved measurement errors. We also recommend on specific scanning technology and on the most suitable measurement pipeline for selected anthropometric applications. Finally, Appendix A provides an overview of currently available commercial body scanners and Appendix B lists currently available mobile applications for body scanning.

[2]A template mesh is a graphical model comprised of vertices and surfaces which depicts a standard human, usually in a T-pose, having a known number of parameters which control the appearance of the mesh (sex included).

II. PRIOR REVIEWS

We briefly describe prior reviews in a chronological order. The review covers 3D scanning technology and body measurement.

3D scanning technology. One of the first reviews on 3D scanning technology for anthropometry is done by Daanen and Van de Water [39] in 1998, covering 8 commercially available full-body scanners. The most developed scanning technology at that time was laser line-based scanners with vertically moving scanning heads, projecting a horizontal line over the human body. A review on 3D body scanners for the apparel industry [76] (2001) distinguishes laser, LED SL, and white-light SL scanners. Based on their analysis, the scanning time of laser scanners is usually higher than the latter two, but SL scanners have longer data processing times. Olds and Honey [114] claimed in 2005 that structured light 3D scanners using white light are generally cheaper and faster than their laser counterparts, but produce lower quality scans. A review by D'Apuzzo from 2007 [43] focuses on 3D body scanning technology and its application in fashion and apparel industry. The paper distinguishes SL and photogrammetry (passive stereo) approach. ToF sensors were still not commercially used for 3D scanning in 2007. Even though scanning systems were becoming smaller in size, there were no commercial handheld or mobile scanners dedicated to anthropometry. Another review from 2007, by Treleaven and Wells [156], analyze 3D scanning technology and methodology for various medical applications, like skin analysis and burn treatment, deformity detection, and obesity.

The updated review by Daanen and Van de Water, from 2013 [40], points out that the 3D scanning technology improved in terms of transportation (mobility), speed, price, and accuracy, especially regarding SL scanners. Around that time, ToF scanners appeared on the market. The review focuses on stationary 3D scanners. A book on 3D cameras [56], from 2018, describes and provides in-depth comparisons of ToF, SL and photogrammetry-based (PS) 3D cameras. Finally, a survey by Haleem and Javaid [61] from 2020, similar to the one by Treleaven and Wells [156], is focused on 3D scanning technology in medicine. The difference is that they also take into account X-ray, CT, MRI, and ultrasound, analyze strengths and limitations, and discuss the specific applications of each technology.

Body measurement. A body measurement review by Wang et al. [164] from 2000 is focused on the measurement and analysis of body length, width, circumference, and skinfold thickness to predict body fat percentage. The main issue in their survey that still has not been completely solved is the lack of standardization in body measurement. A review by Lescay et al. [91] compares different anthropometric measurement techniques; traditional anthropometry, structured light, photogrammetry, and mobile applications, based on precision, number of measurements, speed, and price. Another review by Heymsfield et al. from 2019 [68] describes the process of acquiring 3D human body scans, creating and processing meshes, validating the acquired data, and the applications of the obtained data in anthropometry and medicine. It also distinguish between SL and ToF scanners in terms of data acquisition techniques and mention several stationary scanner models. A review by Dianat et al. [44] focuses on the methodology and applications of anthropometry in ergonomics. Their paper mostly mentions measurement methods in terms of traditional anthropometry and covers the existing 3D scanning technology on a high level only.

Taking into account prior work on 3D scanning technology, we detect the lack in the review of existing handheld and mobile scanners, as well as the review of existing mobile applications for 3D scanning and especially anthropometry. To the best of our knowledge, we are the first to provide a complete and modern overview of body measurement based on 3D scans and RGB images.

III. 3D SCANNING TECHNOLOGIES

Several 3D scanning technologies have been proposed over the years. As mentioned in the Introduction, we distinguish between three common approaches, passive stereo, structured light, and time-of-flight imaging, which we now describe in more detail.

A. PASSIVE STEREO

Passive stereo is a measuring technique for 3D reconstruction from multiple camera views. Photogrammetry is the science of measuring objects from photographs. Passive stereo and photogrammetry are sometimes used interchangeably in the context of 3D scanning [41], [48], [135]. For clarity, we use the passive stereo in the remainder of the paper. PS-based 3D scanners use RGB cameras to obtain color images. The PS assumes that multiple cameras are pointing to a person. Under passive stereo, in this Section, we describe stereo and monocular reconstruction principles, as well as motion capture systems.

Stereo reconstruction. The simplest PS configuration is a binocular stereo, a camera configuration of two horizontally or vertically aligned RGB cameras (see Fig. 2). The reconstruction is based on the correspondences between the images, and triangulation [63]. The point P in the 3D scene projects to pixels p_1 in the first and p_2 in the second image (for example, as in Fig. 2). However, for a fixed pixel location p_1, the corresponding pixel location p_2 is not known a priori. The location p_2 is determined by matching an image block around p_1 with the most similar block along the epipolar line l [63]. The difference between the corresponding pixel coordinates[3] $|p_1 - p_2|$ (disparity) is used to triangulate the depth of a point P [63]. The stereo approach can be extended to more than two cameras by coupling pairs of cameras [147] or by using multi-view-stereo techniques [52].

[3]Note that the images first need to be rectified [101].

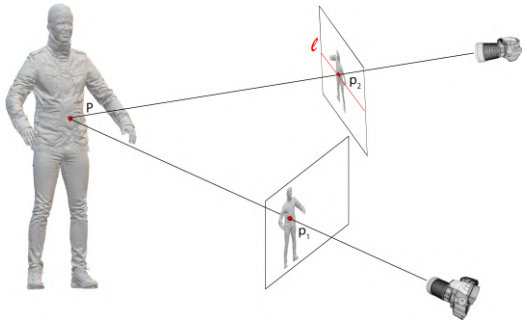


FIGURE 2: Passive stereo approach. Point p_2 is the most similar image pixel to point p_1 along the epipolar line l, as described in Sec. III-A

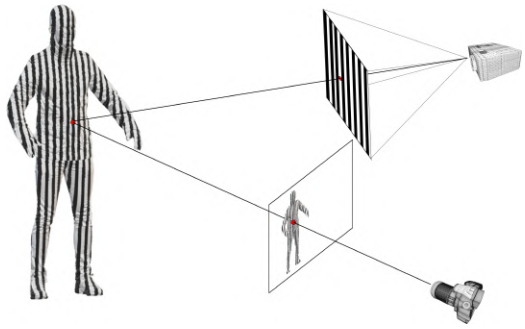


FIGURE 3: Structured light (projector based) approach.

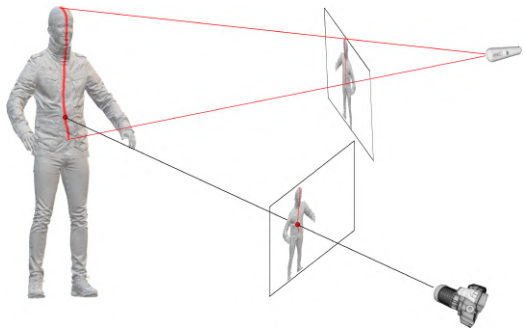


FIGURE 4: Structured light (laser based) approach.

Monocular reconstruction. A monocular moving-camera-based 3D reconstruction is a special case of stereo reconstruction, where each viewpoint (frame) is considered as a separate camera [52]. The general monocular approaches [21], [146], [179] jointly reconstruct 3D scene and estimate camera locations in every frame. First, the keypoints are detected [10], [104], [138], [139] and matched between the images [36] to find the correspondences. The correspondences are then used for the initial 3D reconstruction and camera parameter estimation, usually followed by bundle adjustment (BA) refinement [157]. Human 3D scanning is usually simpler as camera locations can be obtained prior to the reconstruction. This is implemented in a way that either the camera is rotating around the person or the person is standing on a rotating platform, mimicking camera rotation. Note that the person needs to stay still during the quasi-static scanning. The relative camera positions with respect to the subject are extracted based on timestamps. To acquire dense 3D reconstruction, the principles of stereo reconstruction described above can be used.

Motion capture. MoCap is a (semi-) passive stereo technique that uses body markers visible under the standard or near-infra-red light. The MoCap markers are usually small, round objects with reflective surfaces. MoCap produces sparse 3D reconstructions and is usually used for motion tracking. The number of body markers is between 30 [72] and 300 [121]. Multiple markers are often used to estimate the location of a single keypoint (joint), as markers can only be placed on the surface of the body.

Human body scanners using PS. Commercial 3D scanners either use rotating monocular system or multiple fixed cameras. For example, Texel Portal MX, Fit3D, and BodyGee Orbiter rotate a person that is standing on a platform, while Texel Portal BX circles around a static body. A few examples of fixed-camera scanning systems are Bootscan Neo, TC^2-21B, and 3IOSK by Mantis Vision that uses from several to more than 50 RGB cameras to obtain the reconstruction. There are several advantages of fixed multi-camera over single-camera scanners. The first advantage is reduced scanning time because neither the cameras nor the person need to move. The second one is the ability to scan people in motion over a period of time, also called 4D scanning (Move4D scanner by IBV). Thirdly, it is possible to reconstruct multiple people at once, if the scanning area is large enough to avoid occlusions, for example, as in Panoptic Studio [78].

Based on the images and the reconstruction described in this Section, a mobile device camera can be considered a special case of a monocular PS-based scanner, where a camera is moved around a person to record a video or take individual images. For a comprehensive overview of the commercial 3D scanners, please refer to Appendix A and B.

B. STRUCTURED LIGHT

To address a poor 3D reconstruction quality of PS in the case of low or repeating texture, the usual approach is to project a textured pattern over the scene. Active stereo (AS) [59], [60], [71], [100] upgrades PS by projecting a light pattern onto the body to improve the correspondence search between views. Structured light approaches [17], [89], [142], [160], on the other hand, search for the camera-to-light-pattern correspondences. In the remainder of this Section, we focus on SL technology and methods.

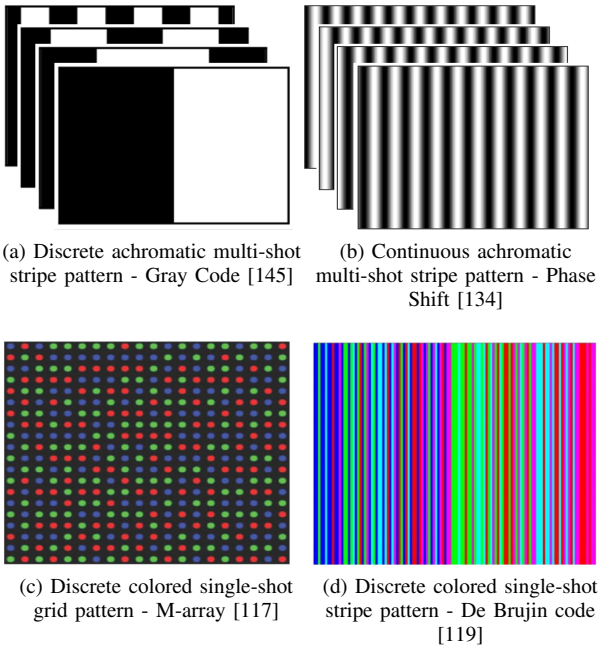
Technology. We distinguish two scanner types based on SL technology - laser and projector scanners. Laser scanners [9], [47], [100], seen in Fig. 4, use a laser to project dot or stripe patterns over the scene. Lasers scanners present sub-millimeter accuracy [22], [161], [174] and simpler decoding

procedure with respect to projector-based scanners [39]. However, laser scanners usually suffer from slow scanning time since the laser line needs to sweep the whole body [43]. Projector-based scanners are usually faster compared to laser scanners [160], since more complex 2D patterns can be projected and the whole body can be scanned at once from one view. Additionally, projector-based scanners present less safety constraints compared to laser scanners [136]. Even though projected-based scanners are not as accurate as laser scanners, their accuracy range (μm - mm) is sufficient for high-quality body measurement (see Sec. V).

In general, many classifications of the projected light patterns have been proposed: based on the number of projected patterns (single- or multi-shot), color (achromatic or colored), transitions (discrete or continuous), and structured form (stripes, grids, dot arrays, gradients, etc.) [54], [118], [143], [144], [160], as seen in Fig. 5. For (quasi-) static human 3D scanning, short-duration achromatic multi-shot patterns are usually used, presenting a trade-off between acquisition speed and reconstruction accuracy [160]. For dynamic scenes where fast acquisition is needed (see Sec. IV-B), single-shot patterns are more suitable [81].

Reconstruction. The camera-to-light-source correspondences are found depending on the light and pattern projected. Laser-based approaches mostly use pattern detection algorithms to find the (monochromatic) light projections in the image [50], [53]. Visible-light scanners, on the other hand, have more complex pattern decoding mechanisms [115], [129], especially in case of multiple projectors and light interference [115], [154]. For more details, we refer readers to relevant survey papers [144]. After the correspondences have been obtained, ray-to-ray or ray-to-plane triangulation can be applied [54], [56], [89], [100] to reconstruct the 3D human body.

Human body scanners using SL. Commercial SL scanners either rotate around a person or have a fixed multi-sensor configuration that surrounds them. Stationary scanners such as the HP Pro S3, 4DDynamics EOS, TC^2-105 or Hexagon Aicon Primescan, rotate around the body to obtain a whole 3D scan. Another way to move around the body is to use handheld scanners such as the Artec Eva, TechMed3D BodyScan Scanner, Mantis Vision F6 Smart or ScanTech Axe B17. Stationary scanners with fixed sensor position, such as the Artec Shapify Booth, botscan Neo, botscan OptaONE+, TC^2-105, 4D Dynamics IIID Body Scan, showcase a camera and projector filled booth in fixed positions that surround the scanned subject. Solutions to avoid light interferences [163] from multiple projectors have been proposed, but in practice, every projector illuminates the subject in its designated time interval. Hence, acquisition time is prolonged and proportional with the number of scanners. For a comprehensive overview of the commercial 3D scanners, please refer to Appendix A.



(a) Discrete achromatic multi-shot stripe pattern - Gray Code [145]

(b) Continuous achromatic multi-shot stripe pattern - Phase Shift [134]

(c) Discrete colored single-shot grid pattern - M-array [117]

(d) Discrete colored single-shot stripe pattern - De Brujin code [119]

FIGURE 5: Structured light pattern examples.

C. TIME-OF-FLIGHT

ToF scanners, shown in Fig. 6, measure the time needed for an emitted light signal to travel from the illumination source to the 3D scene and back to the sensor. The distance information is directly proportional to the time of flight of the light signal [51], [56], [70], [92].

Technology. The main components of a ToF scanner are the light emitter and the photodetector [56]. The light emitter uses a laser or an LED to send a modulated beam of light, typically in the NIR range [92]. The lens is used to spread the light from the emitter over the whole scene. The photodetector usually uses a matrix of point-wise sensors [70]. For human 3D scanning, CCD/CMOS matrix sensors are usually used.

Reconstruction. Two reconstruction methods can be distinguished: pulsed-light (direct) and continuous-wave (indirect) [56], [70]. Continuous-wave (CW) methods indirectly measure the round-trip time of an emitted light pulse and collecting the time-dependent intensity information of the signal [51], [126]. The distance of a point is then retrieved (demodulated) from the phase shift of the emitted and received light signal by their cross-correlation [70], [130]. The emitted illumination signal amplitude is usually modulated using a sine or a square wave [18]. The periodicity of the waves implies a maximum scanning range, at half of the modulation wavelength, after which an ambiguity problem arises [65]. Increasing the modulation frequency increases the measurement accuracy, but shortens the maximum range [70]. The range can be extended using multiple modulation frequencies [57], [125]. Fortunately, this does not usually present a problem in the anthropometry application, since

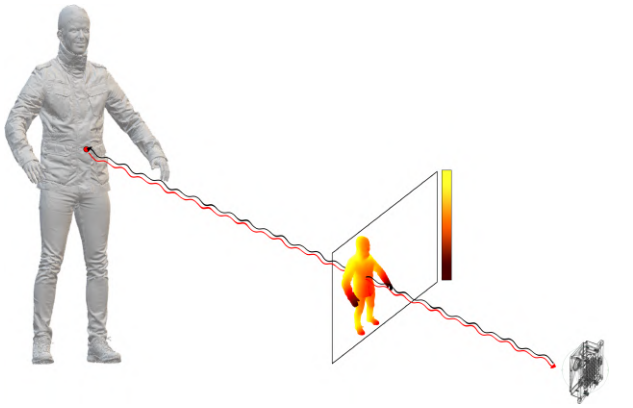


FIGURE 6: Time-of-flight approach. The black arrow indicates the emanated light signal path. The red arrow indicates the received light signal path.

human bodies are scanned from close range. Pulsed-light (PL) methods directly measure the round-trip time of an emitted light pulse using time-to-digital (TDC) or time-to-amplitude (TAC) circuitry [70], [126]. Since the speed of light is very fast, PL methods require extremely precise timing information, in the order of picoseconds, to obtain millimetric distance range [56], [92], [126]. Hence, PL is not usually used for 3D body scanning.

ToF cameras present a low cost, compact size, accurate and reliable sensors with lower power consumption [51], [65], [67]. Compared to SL, ToF does not have a spatially separate light source and camera, avoiding occlusions problems between views. Additionally, it is texture-independent with minimal post-processing time and lower-light capabilities [151]. Even though fast frame rates can be achieved, which is suitable for dynamic scanning [51], [151], the biggest problem of a single ToF camera scanners is a low scanning resolution [51]. It is possible to increase the resolution by using multiple ToF cameras [167], but complex light interference issues then need to be addressed [130]. Therefore, ToF is still less applicable for (quasi-) static scanning and body measurement.

Human body scanners using ToF. Most of the commercial human body scanners, such as the SizeStream SS20, Styku S100 and TC^2-30R are based on indirect ToF methods. In general, ToF as a standalone solution is unable to provide high-quality 3D human body scans due to its lower resolution. Hence, they are usually used in combination with RGB cameras. Noticeably, a bigger percentage of stationary scanners, such as the TC^2-19R, Naked scanner, and BodyGee Orbiter, come with a turntable on which subjects take a standard scanning position. This alleviates solving light interferences from having multiple cameras. Note that all mini scanners are ToF-based and therefore used for 3D data acquisition in mobile applications (see Appendix B). For a comprehensive overview of the commercial 3D scanners, please refer to Appendix A.

D. SUMMARY OF SCANNING TECHNOLOGIES

The comparison of the three scanning technologies is provided in Table 1. Regarding methodology, PS and SL rely on finding the correspondences between the views to triangulate 3D points in space, while ToF uses time-to-distance conversion and thus avoids the correspondence search problems. The common challenge for the triangulation approaches are the potential (self-) occlusions between the views, which might result with holes in the 3D point cloud [52]. A way to cope with the occlusions is to use more cameras or viewpoints (achieved by rotating the subject or the scanner) and to use the T-pose where self-occlusions are mitigated.

SL and ToF use light sources. In one way, it helps SL in low-textured body areas, but it also limits its applications to specific indoor lightning conditions. For multi-ToF scanners, light causes the interference problems. Regarding the scanning ranges, SL and ToF are limited by the illumination source. PS scanning range is, in theory, limited by the optics, but in practice it is several meters. All the scanning ranges are suitable for human body scanning.

With regards to scanning of moving subjects (dynamic scanning), PS is the most suitable because of its fast acquisition time, good overall reconstruction performance, and no light interference issues. ToF has a high reconstruction frame rate, making it applicable for dynamic applications [62], [152]. SL can also be used for dynamic scanning with single-shot patterns, but for scanning slower movement only. Moreover, single-shot offers lower reconstruction accuracy compared to multi-shot patterns.

Finally, SL offers the best accuracy[4] and resolution[5], making it the method of choice for quasi-static scanning and body measurement. This can also be seen in the number of commercial SL scanners[6]. PS and ToF have similar accuracy and resolution range (see Table 1), but ToF generally has a lower resolution.

IV. BODY MEASUREMENT

We describe our proposed body measurement framework (Fig. 1), dedicating Subsections to each of the five processing stages: preparation, scanning, feature extraction, model fitting, and measurement extraction. The first two stages are the acquisition stages, and the latter three are the processing stages (see Fig. 1). In the acquisition stage, the subjects are prepared and the data in form of 3D points clouds, depth maps, or 2D images are obtained. There are two acquisition protocols - static and dynamic. In the processing stage, the collected input is used for body measurement. Body measurement can be done directly on the given inputs, but usually the features are first extracted and the body model is fitted based on these features or the inputs.

[4]Accuracy is the distance between the reconstructed location and true true location of a 3D point in space.

[5]Resolution is the minimal distance of two points in space that can be differentiated and reconstructed.

[6]See Appendix A for more details.

TABLE 1: Main properties of the three 3D scanning technologies with respect to human body scanning.

	Passive stereo	**Structured light**	**Time-of-flight**
Method	triangulation	triangulation	time-to-distance conversion
Illumination	passive (ambient)	active (visible, IR)	
Scanning range	several meters	< 5 m (illumination source limited)	
Dynamic scanning	yes	yes (slower movement only)	yes
Accuracy range	mm - cm	μm - cm	mm - cm
Resolution range	mm	μm - mm	mm
Main issues	textureless body parts	light interference	lower resolution, multi-camera interference

A. PREPARATION

Standardization of body landmarks, measurements, and postures is the first step to ensure the comparability of measurements between the body measurement surveys [174] and to compare the scientific results. Body landmarks represent the same semantics for every measured subject (Fig. 7A) and some of the body measurements can be directly derived from landmarks (see Sec. IV-E). The landmarks are defined on the skin to reduce the ambiguity in their locations between the subjects. In practice, markers that represent the landmarks are manually placed on the human body. The markers are useful in the feature extraction (stage 3), however, marker placing is a tedious and error-prone process so successful markerless systems have been proposed [79], [106].

Standardization. ISO standard 7250-1:2017 [75] specifies a list of body landmarks and measurements. The complete list of body landmarks is given in Table 2 and the corresponding points are shown in Fig. 7 (left). The complete list of body measurements is given in Table 3 and the corresponding Fig. 11. There are two standard standing poses recommended by the ISO 20685-1:2018 [74] (Fig. 7). The person is asked to take one of the two poses, hold its breath during the scanning, and try to keep as calm as possible [106]. In the first pose (I-pose), the subject stands upright with the shoulders relaxed and arms hanging down naturally, feet together. In the second pose (A-pose), feet are 20 cm apart, arms form a 20° angle with the torso, the elbows are straight and the palms face backward [174]. Using the standard or fixed body postures is not always required for body measurement, but usually, it is when creating datasets that capture shape variations [3], [8], [12], [24], [64], [72], [175]. The pose that is also often used for scanning is a T-pose, as seen on a neutral template mesh in Fig. 9.

TABLE 2: The list of human body landmarks according to ISO 7250-1:2017 standard [75]. The numbers correspond to the numbers in Fig. 7 (left). The letters *R* and *L* abbreviate right and left.

Human body landmarks (ISO 7250)		
1 tragion	12 axilla pnt. post. RL	23 supratarsale fib. R
2 orbitale	13 iliocristale RL	24 metatarsale tib. R
3 glabella	14 iliospinale ant. R	25 metatarsale fib. R
4 sellion	15 acromiale RL	26 waist level
5 gnathion	16 radiale R	27 stylion ulnare R
6 cervicale	17 stylion R	28 stylion ulnare R
7 suprasternale	19 stylion ulnare R	29 trochanterion RL
8 front neck	20 trochanterion RL	30 abdom. ext. level
9 side neck	21 tibiale R	31 buttock pnt. R level
10 mesosternale	22 sphyrion R	32 gluteale R level
11 axilla pnt. ant. RL	22 sphyrion fib. R	

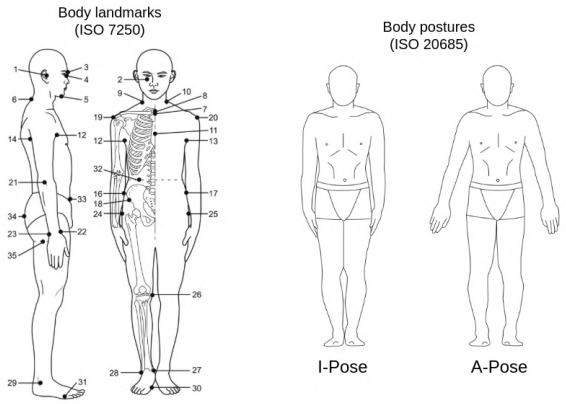


FIGURE 7: Body landmarks according to ISO 7250-1:2017 [75] standard and two standing postures according to 20685-1:2018 [74]. Left axilla point posterior (#12) is not shown. Images adapted from: [85].

B. SCANNING

Regarding the acquisition protocol when using 3D scanners human body may be measured in a stationary position [3], [6], [8] or in motion [2], [108], [158]. In static scanning a person is asked to take a pre-defined pose and to hold still until the scan ends. For 3D scanners which have longer acquisition times, e.g. scanners with rotating heads or handheld scanners, subjects may unintentionally move during acquisition which introduces errors so we may distinguish such situations as quasi-static scanning. Static scanning is the method of choice to obtain most precise body measurements and is routinely used in production of relatively large and diverse public 3D human body datasets [3], [8], [24], [64], [175]. Scanning in motion usually limits the technology to either PS or ToF. Most common are motion capture (MoCap) systems [12], [103], [107], [108], [158] which are PS based and which use markers attached to the body to track movement. Other dynamic 3D scanning systems [2] record a person in motion to analyze soft-tissue

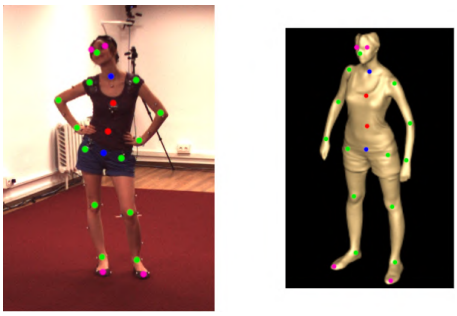


FIGURE 8: An example of 2D keypoints and their corresponding 3D scan keypoints. The typical keypoint extraction algorithm finds between 13 (green) and 21 keypoints (other colors). The blue keypoints represent the neck and pelvis, the red ones two spinal points, and the pink ones details on the face and feet. The images are adapted from Human3.6M dataset [72]. Note that the keypoints are manually annotated and do not necessarily reflect the H36M dataset ground truth locations.

deformations over time [128].

Scanning usually produces a 3D point cloud, one or more depth maps, or a set of RGB images. In case of dynamic scanning, a so-called 4D scans are obtained [2]. In the processing stage, some or all of these data is used to extract the measurements.

C. FEATURE EXTRACTION

Two types of features that are usually extracted from 3D scans and images are keypoints and silhouettes. The location of keypoints[7] can be determined based on markers or can be estimated automatically from a 3D scan [79], [106]. The keypoints usually represent a selected subset of human joints (see Fig. 8). Silhouettes either represent the points or pixels for the whole human body, or the body segments.

Keypoint extraction. Most of the keypoint estimation algorithms detect human joints from single or multiple images. The joints can be represented by 2D pixel coordinates in an image or 3D points in the scene. If having a moving person, the time component can be exploited and temporal smoothness can be applied to improve the estimation accuracy [124]. Therefore, keypoint estimation methods can be divided into: single-image [14], [30], [31], [111], [176], multi-frame [124], and multi-view methods [66], [73], [133] for 2D [14], [30], [31], [153], [176] or 3D [66], [73], [111], [124], [133] keypoint estimation. The keypoint estimation algorithms usually find between 14 and 21 keypoints, as shown in the Fig. 8. Most of the state-of-the-art keypoint estimation methods are deep learning-based, due to the availability of large annotated datasets [72], [110], [132], [148], [162]. In practice, the extracted 2D and 3D keypoints are used for mesh fitting [23], [83], [116] (see the next Subsection) and are typically not combined with landmark extraction from 3D scans.

Motion capture [158] is a movement tracking technique that enables the direct acquisition of precise ground truth 2D and especially[8] 3D keypoint locations. Most of the previously mentioned keypoint estimation algorithms take advantage of the ground truth data obtained using motion capture. The examples of MoCap datasets are Human3.6M dataset [72] (Fig. 8), HumanEva [148] (Fig. 10), and TotalCapture [158]. The disadvantage of motion capture systems is that they are impractical for in-the-wild scenarios.

Regarding keypoints from 3D scan, Lu and Wang [106] proposed a system for markerless 3D scan keypoint detection. A body scan is firstly cleaned by removing the outlier points and then segmented into five parts: head and torso, left arm, right arm, left leg, and right leg. The initial keypoint locations are derived from the anthropometric database [165] and then refined using four algorithms: silhouette analysis, minimum circumference determination, grayscale detection, and human body contour plots. The results of the four algorithms are combined to determine the final keypoint locations and body measurements.

Silhouette extraction. Silhouette extraction methods separate pixels that represent an object of interest (the human body) from other pixels in an image [15]. There are three approaches to silhouette extraction: background subtraction [19], semantic segmentation [168], and multi-view segmentation (visual-hull) [90]. In the work by Lin and Wang [99], two silhouettes are extracted using background subtraction, from front and side input images and 60 feature points in total are detected on the edge of the silhouette, based on the curve distance between them [98]. The extracted feature points are directly used for approximate body measurement extraction (see Subsec. IV-E).

State-of-the-art semantic segmentation methods [33], [49], [93], [96], [171], similar to human pose estimation, are also deep learning-based. Except for whole-body segmentation [97], [112], there are also body-part-segmented datasets [95], [178]. Both whole body and body part segmentation problems are particularly interesting in terms of silhouette and body measurement extraction, as they achieve relatively high accuracy[9], even on difficult examples. A visual hull is reconstructed by applying background subtraction or semantic segmentation for multiple images of a fixed object from different views [52]. A visual hull can be used as an initial solution for mesh fitting.

D. MODEL FITTING

Model fitting is a set of techniques for finding a 3D template mesh that best represents a given input. The given input can be a 3D scan, 2D or 3D keypoints, or silhouette(s). The advantage of using template meshes in the context of body

[7]Note that keypoints are called landmarks if they refer to standardized body locations [75].

[8]Another way to obtain 2D pose estimation data is to manually label human joints on a large number of images. However, this is impractical and unreliable in case of 3D data.

[9]The accuracy is measured as a mean IoU (intersection over union).

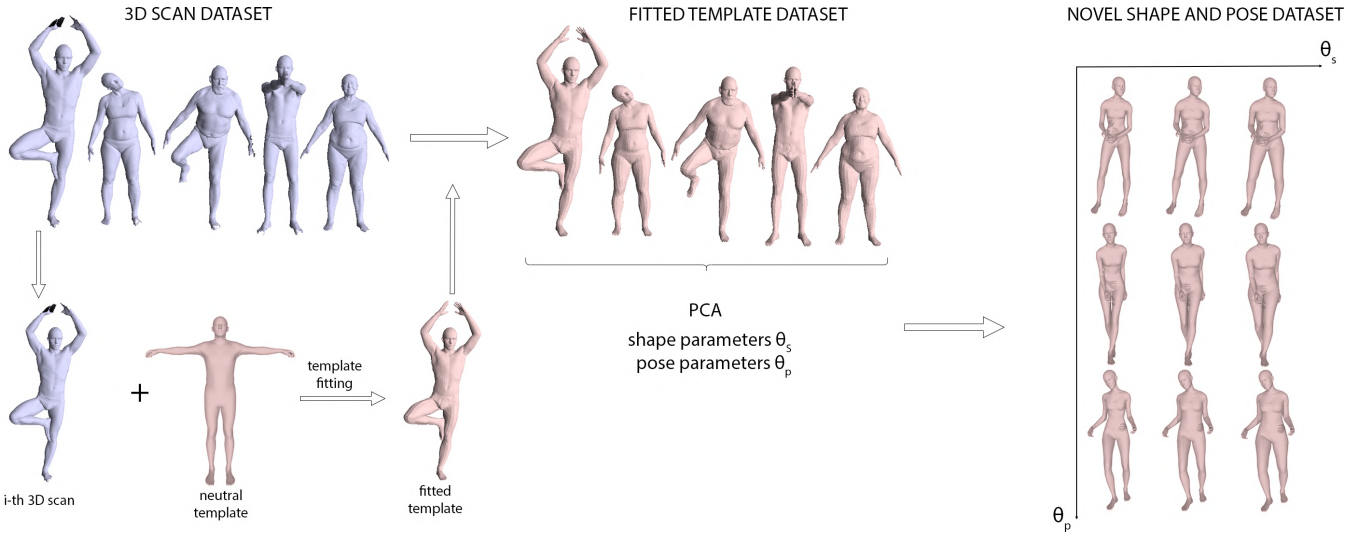


FIGURE 9: An overview of a mesh fitting process for the creation of statistical models. For every 3D scan in the scanning dataset, a neutral template mesh is registered to the scan, producing a dataset of registered template meshes. Based on the shape and pose variations of the registered templates, PCA can be applied to create the statistical model. The principal components can be used to generate novel 3D meshes from the pose-shape space. The 3D scans and template meshes are retrieved from the FAUST dataset [24]. The novel 3D meshes are generated using SMPL-X [123].

measurement estimation is that the number of vertices is fixed and corresponding vertices have the same semantics for all the registered meshes in the dataset. Once body measurements are obtained for a single mesh, they can be obtained in the same way for all the meshes. We distinguish two model fitting techniques - mesh fitting (registration or deformation) and mesh regression using statistical models. In this Subsection, we describe mesh fitting, statistical model creation, and mesh regression from 3D scans and images.

Mesh fitting. Mesh fitting is an optimization process of deforming an initial, template mesh to the 3D scan[10]. Mesh fitting consists of pose and shape fitting [12], [69], [102], [116], [123]. Before the optimization, a 3D scan is usually subsampled so that the number of points is the same or larger than the number of vertices in the template mesh [12], [172]. First, the landmarks are used to roughly align 3D scan and mesh [127]. Next, pose fitting is done by rigging [16] the body skeleton parts of the template mesh and then skinning of surface points, using linear blend (LBS) [12], [102] or dual quaternion skinning (DQS) [116]. Once the pose satisfies the convergence criterion, shape fitting is done using non-rigid registration, minimizing the loss function that usually consists of three components: landmark term, smoothness term, and data term. The landmark term minimizes the distance between the corresponding landmarks of the template mesh and the 3D scan. The smoothness term minimizes the difference between the spatial transformations of the neighbouring vertices. Finally, the data term minimizes the distances between the corresponding vertices. Note that the correspondence is determined at the beginning of a shape fitting phase. Pose and shape fitting is usually done alternately multiple times, until final convergence [127]. Some works also take texture into account [24], which improves fitting. The described fitting is a method of choice for almost-complete 3D scans, obtained using high-quality scanners. For partial 3D scan fitting, a method based on implicit functions [37] has shown promising results on SHARP 2020 (SHApe Recovery from Partial textured 3D scans) challenge [141]. The result of fitting is a clean mesh that fills up the holes in the original, noisy 3D scan.

Statistical models (SMs). Statistical models represent the population of human bodies with respect to pose and shape variations, usually represented by the principal components (PCs). To create a statistical model, mesh fitting procedure needs to be applied to each scan in a dataset. The work by Hirshberg et al. proposed to simultaneously fit meshes, while creating the body model [69]. One of the advantages of simultaneous fitting and creating the model is that the occluded 3D scan regions are properly fitted based on the scans in different poses where these regions are not occluded. To describe pose and shape variations in the set of fitted template meshes, PCA is used. The purpose of PCA is to compress the dataset of registered meshes by finding pose and shape principal components that explain the maximal variance of the dataset. An important advantage of the PCA is that the PCs can be used to generate novel template meshes [107], [122], [149], [177] from a pose-shape parameter space. The datasets commonly used for building SMs are CAESAR [3], Size-UK [8], ScanDB [64], and possibly other datasets containing 3D scans [12], [24],

[10]For simplicity, we only describe mesh fitting on 3D scans, but similar techniques can be applied to features or images [84].

[72], [148], [175].

SCAPE [12] is the first SM for both pose and shape deformations, as well as pose-dependent shape changes (for example, muscle contractions in different poses). They use a set of initial, physical markers and the correlated correspondence algorithm [13] to generate around 150 additional markers. Then they apply non-rigid registration to obtain the articulated human model. One of the main disadvantages of SCAPE is that each body part is independently rotated, which introduces artifacts near joints. To that end, BlendSCAPE [69] smooth SCAPE body part segmentations across part boundaries, which solves the artifacts problem. A disadvantage of both BlendSCAPE and SCAPE is that they use triangle deformations for the PCA. One of the most popular statistical models, SMPL [102], showed that using vertex instead of triangle transformations improves the final SM. SMPL also enforces body symmetry to produce models that are visually more pleasing for animation. Enforcing the symmetry, however, sacrifices realism in particular poses. An improvement over SMPL is a STAR [116] model that enforces spatially local and sparse pose corrective blend shapes and is independent of the symmetry optimization component. STAR is the most expressive SM, partly due to the fact that it is built using the largest database, a combination of CAESAR [3] (4000 scans) and SizeUSA dataset [6] (9000 scans).

Mesh regression from 3D scans using SMs. Once a statistical model is built, it can be used for mesh regression. The idea of mesh regression is to find pose and shape parameters of the SM that best fit a given input. An example of such an approach is done by Kwok et al. [88], consisting of iteratively selecting the mesh from the statistical pose-shape space and fitting the clothes to match the input 3D scan. Prokudin et al. [131] propose a deep learning model for template fitting, supervised by SMPL templates fitted to the dataset before learning. The learning is based on the distances between the set of 3D scan features, called the basis point set and the ground truth template mesh. The advantage of using the features to find optimal parameters is that the (slow) rendering part that is needed to verify the parameters is avoided.

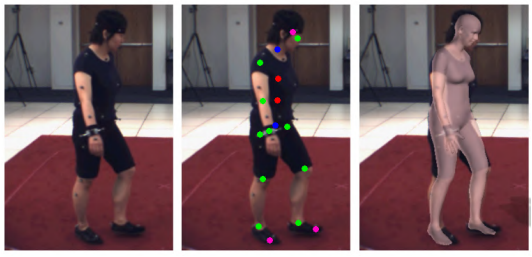


FIGURE 10: An example of a SMPLify [23] mesh fitting approach, based on 2D keypoints, on HumanEva dataset [148]. The image is adapted from: [23].

Mesh regression from images using SMs. There is a group of methods that use extracted image features (body pose or silhouette) or RGB images direcly and exploit the SMs for mesh regression. A large body of these methods are based on SMPL statistical model [23], [84], [150]. For example, SMPLify [23] is a deep learning model for 3D shape and pose estimation from 2D keypoints. The keypoints are detected using a 2D pose estimation algorithm [31]. Using sex-specific SMPL models, SMPLify simultaneously estimates 3D pose and shape parameters and produces a template mesh (see Fig. 10). The main disadvantage of SMPLify approach is that it does not exploit image information. A multi-task learning approach by Smith et al. [149] uses front and side-view silhouettes and feed them into a convolutional model to estimate 3D joints, mesh volume, shape parameters, and pose angles (the angles between the adjacent joints), using SMPL as an underlying statistical model. The results, as seen in Sec. V, show that silhouette-based approaches can be used to obtain accurate body measurement. However, the major issue of silhouette-based approaches is the clothed-people scenario, where it is difficult to estimate the underlying body shape. A recent method by Kolotouros et al. [84] uses raw pixels and deforms an initial mesh based on graph CNN [169]. The most similar mesh from the SMPL pose-shape space can then be matched to the deformed mesh. Note that graph CNN approaches can also be interesting for mesh-from-3D-scan regression.

E. MEASUREMENT EXTRACTION

Body measurements can be extracted from: 3D scan, template mesh, or image features. We focus on two measurement types - distances (lengths, breadths, depths, and heights) and circumferences. For other measurement types, such as surface area measurement, we refer readers to [155]. A subset of the standardized body measurements [75] is listed in the Table 3.

Measurements from template mesh. In case the fitted or regressed template mesh is obtained, the number of the vertices is known and their semantics are the same across all samples [102]. To calculate distance measures, for example, elbow-wrist, hip breadth, or chest depth, the distance between the semantically corresponding vertices can be used. The circumferences, for example, waist, thigh or calf circumference (see Fig. 11, left), can be calculated as the extent of an intersection between the mesh and a plane.

Measurements from the 3D scan. The measurements can also be extracted directly from a 3D scan. The landmarks can help to obtain distances and some of the circumferences [166]. In the work by Lu and Wang [106], the circumferences are calculated from a point cloud using a convex hull polygonal approximation method. The circumference points are obtained by slicing the point cloud with a perpendicular plane. The algorithm starts in the point with the highest X coordinate (Fig. 12b). The next point is selected as the point with the minimal angle between the Y-axis and the line connecting the current point X and the

TABLE 3: An example list of 44 standardized human body measurements [75]. The measurements consist of distances (lengths, breadths, depths, and heights), circumferences, and soft biometrics (weight, height, BMI).

Human body measurements			
1 eye	12 forearm circum. L	23 weight	34 bicep circum. R
2 cervicale	13 forearm circum. R	24 height	35 shoulder breadth
3 shoulder-elbow L	14 neckbase breadth	25 BMI	36 elbow circum. L
4 shoulder-elbow R	15 thigh clearance	26 neck circum.	37 elbow circum. R
5 crotch height	16 wall-acromion distance	27 chest circum.	38 knee circum. L
6 tibial height	17 grip and forward reach	28 waist circum.	39 knee circum. R
7 chest depth	18 elbow-wrist L	29 thigh circum. L	40 neck base circum.
8 body depth	19 elbow-wrist R	30 thigh circum. R	41 neck circum.
9 thorax depth	20 hip circum.	31 calf circum. R	42 head circum.
10 chest breadth	21 buttock-popliteal	32 calf circum. R	43 trouser waist circum.
11 hip breadth	22 buttock-knee	33 bicep circum. L	44 iliac spine breadth

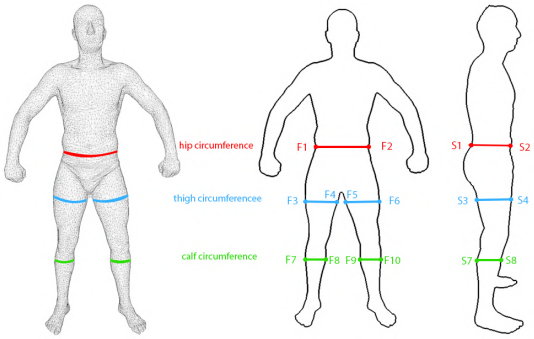


FIGURE 11: Body measurements on a 3D mesh (left) and corresponding feature points on front and side-view silhouettes (right). The feature points can be used to approximate the measurements. The mesh is generated using the SCAPE model [12].

next point, in the counter-clockwise direction (Fig. 12b). The process is continued until the polygon is closed. The circumference is approximated as the sum of the line lengths between the selected points.

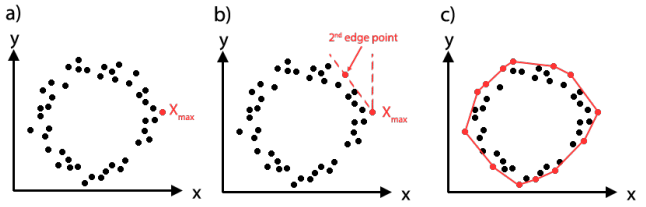


FIGURE 12: The convex hull polygonal approximation method.

Measurements from image features. If the front- and side-view silhouettes are extracted [77], [98], [99], the measurements can be approximated using the distances between the feature points on the silhouette (see Fig. 11). The circumference can be approximated by a circle or an ellipse. For example, waist breadth is the distance between the F1 and F2 and waist depth is the distance between S1 and S2. The distance between F1 and F2 and the distance between S1 and S2 can be used as a major and a minor axis to approximate the hip circumference.

V. DISCUSSION

We discuss the main limitations and issues of current scanning technology and body measurement framework, as well as gathered state-of-the-art results from Table 4. Based on the presented framework and the scanner types introduced in this Section, we recommended pipelines for particular body measurement applications (see Fig. 15).

A. LIMITATIONS AND CHALLENGES

Absolute scale. An important practical challenge for some body measurement approaches, in particular, monocular and self-calibrated [63] PS methods, is the unknown absolute scale. The simplest way to obtain the scale is to use a calibrated 3D scanner data as input. Selected body measurement methods [7], [64], [159], [172] compared in Table 4 use 3D scans on the absolute scale as input. Another way to recover the scale is to place an object of known size (the calibration object) next to the subject since, to recover absolute body measurements, it is sufficient to retrieve a scale of a single measurement. Usually, the body height is the most convenient body measure. Selected approaches presented in Table 4 use either the height [45], [149], [173] or camera parameters [25], [170] to scale images or silhouettes in order to extract anthropometric measurements on the absolute scale. Finally, some approaches [34], [80] presented in Table 4 do not know the height prior to body measurement. Hence, they estimate the camera parameters as a part of the learning procedure to infer the absolute scale. While [80] uses an encoder and regression approach, [34] uses a Gaussian process latent variable model to estimate the camera parameters.

Evaluation measures. There are multiple evaluation measures concerning accuracy, precision and reliability [113] that are usually reported [29], [87], [172]. This lack of standardized evaluation measures complicates straightforward comparisons of different body measurement methods, since the different error measures cannot be converted one to the other. To compare selected methods, we focus on a single reported measure, the mean absolute error (MAE), since it is mostly reported for the published body measurement methods. The MAE is a measure of accuracy, and is calculated between the body measurement method

TABLE 4: MAE in millimetres for different measurement methods for measurements shown in Fig. 13. The measurements are grouped into circumferences, lengths and breadths. All the methods are evaluated on some sample of the CAESAR [137] dataset with the exception of Yan et al. [172] (denoted with †). The results of each method were extracted from the paper denoted in the "From" column. The table is split into three parts: 2D-based methods, 3D-based methods (further split into published and commercial methods) and the allowable error AE [58] for some of the body parts. The best results are bolded for both the 2D and 3D categories. The mean MAE for every method is provided.

			Circumference											Length			Breadth	Height	
		From	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Mean
2D	Smith et al. [149]	[149]	14.2	**11.4**	**16.2**	25.0	15.2	**5.5**	**10.4**	7.9	**11.1**	**10.4**	**6.3**	**11.0**	**6.0**	**8.0**	8.4	**7.9**	**10.9**
	Yan et al. [173]	[173]	11.6	12.3	26.1	28.7	22.6	6.9	13.0	**7.8**	18.2	11.7	7.8	13.9	9.5	11.2	7.6	20.1	14.3
	Dibra et al. 17 [45]	[173]	10.8	13.1	28.3	38.6	26.0	6.5	13.4	8.0	18.5	11.8	7.9	13.4	6.9	8.7	7.7	11.8	14.5
	Boisvert et al. [25]	[25]	**11.0**	27.0	21.0	**14.0**	42.0	21.0	23.0	13.0	33.0	12.0	14.0	20.0	20.0	34.0	30.0	9.0	21.5
	Chen et al. [34]	[149]	23.0	27.0	18.0	37.0	**15.0**	24.0	59.0	76.0	19.0	16.0	28.0	52.0	53.0	9.0	12.0	21.0	30.6
	Kanazawa et al. [80]	[173]	16.3	27.2	68.3	85.3	62.8	14.3	35.6	16.7	39.3	21.4	13.6	28.6	45.3	37.2	21.8	96.5	39.4
	Xi et al. [170]	[149]	50.0	59.0	36.0	55.0	23.0	56.0	146.0	182.0	35.0	33.0	61.0	119.0	109.0	19.0	24.0	49.0	66.0
	Bogo et al. [23]*	[173]	28.1	24.4	74.5	72.8	99.1	11.9	28.4	25.9	51.3	28.4	28.8	57.8	150.2	219.1	51.9	398.5	84.4
3D	Yan et al. [172]†	[172]	-	**9.1**	14.3	**12.4**	**8.9**	**4.5**	**5.5**	-	**7.9**	**3.0**	10.6	-	13.2	-	-	-	**8.9**
	Tsoli et al. [159]	[159]	**5.9**	15.8	**12.7**	-	12.4	-	-	-	-	-	**6.2**	-	**10.1**	-	-	**7.5**	10.1
	Hasler et al. [64]	[159]	7.5	17.0	13.0	-	16.2	-	-	-	-	-	6.6	-	10.4	-	-	10.2	11.5
	Anthroscan [7]	[159]	7.4	21.1	12.4	-	7.5	-	-	-	-	-	7.6	-	11.7	-	-	5.6	10.4
	AE [58]	[58]	± 5	± 11	± 15	± 12	± 12	-	-	-	± 6	-	± 4	-	-	-	± 8	± 10	± 9.2

estimation E_{est} and the ground truth E_{gt} as follows:

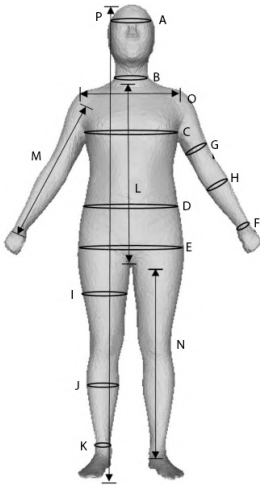
$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} \mid E_{est} - E_{gt} \mid \qquad (1)$$

for every subject i from the dataset. The ground truth is usually obtained by manual measurement as noted in the ISO 20685-1:2018 [74] standard.

Allowable error. The allowable error (AE), based on the ANSUR study [58], defines an upper bound on the acceptable MAE for a measurement method. The study reports the median absolute deviation between measurements made by human experts, which are considered to be the golden standard in anthropometry, and the ground truth for the body measurements. This indicates the limitation of anthropometric measurement methods since the ground truth is never exactly known. Additionally, note that the AE measure represents the median, while the MAE represents the mean absolute deviation from the ground truth. Hence, the MAE is affected by possible outlier measurements and can present higher or lower values than AE.

Datasets. While most methods are evaluated on the CAESAR dataset, there is a fair amount of methods that evaluate their results using their own data [27], [35], [85], [94], [99], [105], [172]. Additionally, methods evaluated on the CAESAR dataset, tend to use its random subsets [149] or gender specific subsets [159]. This lack of a standardized benchmarking dataset presents problems for the direct comparison of measurement methods. We present and compare results of different body measurement methods evaluated on different datasets in Sec. V-B, assuming that the quality and variability of each dataset is similair enough.

Body measurements. Different body measurement methods can be compared on standardized measurements defined in the ISO 7250-1:2017 [75] standard. However, different methods tend to report their evaluations for different measurements, not always equal, as can bee seen by the missing values in Table 4. This hinders a comprehensive comparison of each method.



	Measurements
A	Head C
B	Neck base C
C	Chest C
D	Waist C
E	Hip C
F	Wrist C
G	Bicep C
H	Forearm C
I	Thigh C
J	Calf C
K	Ankle C
L	Shoulder-crotch L
M	Shoulder-wrist L
N	Inside leg L
O	Shoulder B
p	Height

FIGURE 13: Body measurements reported in Table 4 abbreviated accordingly: C stands for circumference, L for length and B for breadth. The image is adapted from: [149]

B. METHODOLOGY COMPARISON

In Table 4 we gather the MAEs of different anthropometric measurement methods for measurements from Fig. 13, addressing the lack of literature on body measurement method evaluations and comparisons. We classify each method depending on their input (2D or 3D) and highlight the best obtained results. Additionally, we group the measurements into three categories, namely circumferences, lengths, and

breadths, to compare their performance on different measurement tasks.

The 2D methods presented in Tab 4 can be divided into two groups. One group of methods [23], [45], [80], [149], [173] uses images or silhouettes to learn the shape and pose parameters of a SMPL model, or a 3D point cloud. As can be seen, these methods outperform the second group [25], [34], [170], that tries to map a 2D PCA space into a 3D PCA space using a Gaussian process latent variable model [25], [34], or simply a linear regression [170]. Additionally, methods that try to estimate the absolute scale [34], [80], in parallel to estimating the body measurements, seem to perform worse in their appropriate groups.

The 3D methods presented in Table 4 are based on fitting the SCAPE model on a 3D scan, and extracting body measurements from the fitted model. In general, there are fewer 3D methods compared to 2D methods, probably as image data is more accessible that the 3D scanning datasets.

Based on the presented results, 3D-based methods are generally better, but do not outperform 2D-based methods by a large margin. Intuitively, 3D scans hold more information about the shape of the human body than 2D images, and hence obtain better circumference measures. On the other hand, 2D methods slightly outperform 3D methods in length estimations, as seen from the shoulder-wrist measure (measure M in Table 4), which may be easier to estimate in 2D. Breadth measurements are unfortunately not comparable, since measurements from 3D methods are not provided, confirming the limitations noted in Sec. V-A. The best 2D [149] and 3D [172] method are both based on the SMPL [102] model. While Smith et al. [149] (2D) use a deep learning model to predict the shape and pose parameters of a SMPL model, Yan et al. [172] (3D) fit an initial SMPL template to a 3D scan using ICP [20].

Compared to the commercially available anthropometry software Anthroscan [7], 3D methods present slightly better results. Anthroscan predicts measurements directly from a 3D scan in the standing pose and is frequently used as a body measurement approach [42], [87], [159]. It achieves an average MAE 1.5mm worse than the 3D method from Yan et al. [172].

In the third part of Table 4 we show the allowable error (AE) for measurements for which AE was measured [58]. While we can observe that the MAE is decreasing with more recent measurement methods, none of the presented methods are within the allowable error, indicating that the automatic body measurement methods are still lagging behind human anthropometers. However, this does not indicate that the assessment methods are insufficient for real-world applications [28]. Additionally, there are commercially available 3D scanners with 3D anthropometry software [86] that claim to obtain results lower than the AE, and can hence be used in applications that require greater accuracies, such as medical and surveying applications.

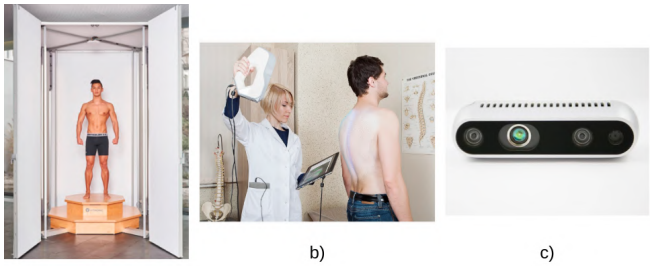


FIGURE 14: Three types of 3D scanners in terms of mobility and size: stationary (a), handheld (b) and mini-scanners (c).

C. RECOMMENDATIONS

Based on the presented technologies, the proposed measurement framework, and the previous discussion, we finally provide practical recommendations for body measurement, as shown in Fig. 15. First, the scanner classification is introduced. Next, specific pipelines are proposed with respect to their input. Finally, the requirements for the applications are described along with the introduced scanner types and pipeline recommendations.

Scanner types. We classify scanners based on their mobility/size, into: (a) stationary; (b) handheld; (c) mini; and (d) mobile camera[11]. Stationary scanners (see Fig. 14a) are usually installed in a fixed location, e.g. a lab or a medical facility. They are usually SL or PS based. Compared to other scanner types they are the most accurate and reliable and are therefore typically used to obtain ground truth data, e.g. stationary scanners were used to create 3D body scanning datasets like CAESAR [3], SIZE-UK [8], Scan DB [64], and FAUST [24]. Handheld scanners (see Fig. 14b) are designed to be moved around the imaged body area by hand. Most of the existing handheld 3D scanners are SL based. Mini-scanners (see Fig. 14c) are embedded in or attached to mobile devices like smartphones and tablets to enable 3D data acquisition. Most mini-scanners are ToF or SL based. Finally, we distinguish mobile RGB cameras as a separate scanner type, because they are wide-spread and convenient for non-demanding users, and usually rely simply on monocular measurement estimation techniques[12]. The four scanner types represent the data acquisition techniques for body measurement, as shown in Fig. 15.

Pipelines. We propose and distinguish three possible pipelines for body measurement, as shown in the right part of Fig. 15. The first pipeline, sufficient for majority of applications, consists of: preparation, 3D scanning, and measurement extraction. The second pipeline is more flexible and consists of: 3D scanning (without prior subject preparation), feature extraction along with or without mesh

[11]For more details on the currently available scanners on the market see Appendix A.

[12]For more details on mobile devices and applications for body measurement assessment, see Appendix B.

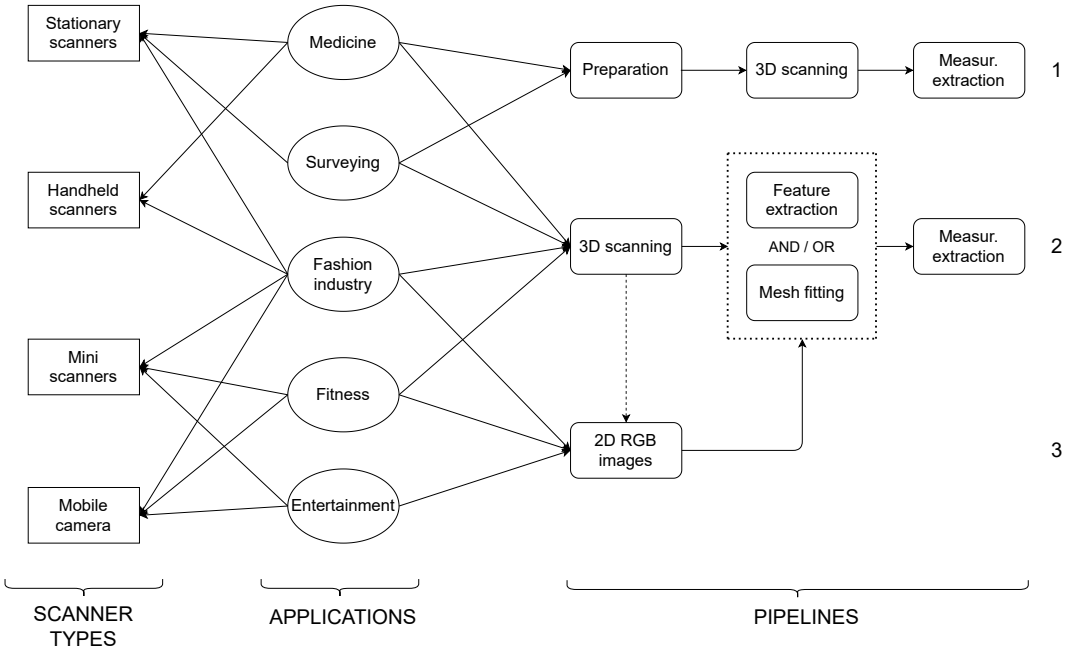


FIGURE 15: The diagram of practical body measurement recommendations.

fitting, and measurement extraction. In both pipelines 2D images acquired using RGB cameras are often useful for improving the reconstruction [120]. Finally, the third and usually the least precise pipeline only takes 2D RGB images as input. These images are then used for feature extraction, mesh fitting, and measurement extraction.

Applications. We recommend specific measurement pipelines and scanner types for different anthropometric applications: medicine, surveying, fashion industry, fitness, and entertainment.

For medical applications [46], [68], it is usually desirable that high-quality body measurements are obtained. Therefore, 3D scanning using stationary or handheld scanners, along with the preparation stage (marker placement), is recommended (see the first pipeline in Fig. 15). The measurements can then be directly extracted from the 3D scan (see Sec. IV).

The second application is surveying, a systematical measurement of a population sample for the purpose of analyzing and tracking the properties of human bodies over time [56], [174]. High-quality surveys sometimes release their data publicly [3], [8], which serves for the creation and improvement of statistical models [12], [69], [79], [102], [116], [127]. Surveying is usually done using stationary scanners and the markers are sometimes placed on the body to improve and simplify the measurement [56].

For fashion industry applications (garment and clothing design), all of the four data acquisition techniques are used. For individually designed garments, stationary scanners are preferable [174]. For less reliable measurements and mass-produced clothes, other data acquisition techniques are sufficient.

For fitness and entertainment applications (gaming, AR, VR, etc.), low-budget solutions using mini scanners and mobile cameras are ideal for individual users. For fitness applications, the body measurements are used for tracking physical progress over time. As seen in the Appendix B, there are a few fitness-based mobile applications that estimate body measurements. Most of them use one or two RGB images from different views. For gyms or fitness centers, stationary 3D scanners might be more convenient. Regarding entertainment, 3D human pose [73], [124] in an AR setup allows the creation of a rigged character [16]; therefore only a rough estimate of body measurements is needed.

VI. CONCLUSION

Anthropometry is a very important, interdisciplinary area of research, still strongly entangled with 3D scanning technology for the purpose of body measurements. We have concisely reviewed the fast developing and improving scanning technologies, becoming therefore more applicable for the automatic body measurements. As a consequence of this development, larger and more diverse body scanning datasets became publicly available. This work has also proposed and discussed different processing stages of the body measurement framework. It was pointed out that a particularly important processing stage is model fitting, which includes mesh fitting and mesh regression, since it allows the development of the expressive statistical body models that describe pose and shape variances of a human population sample. The 3D and 2D measurement methodologies and published works have been compared, the main challenges and limitations have been identified, based on

which several measurement pipelines have been proposed for various applications. Reflecting on the future we recall that pose and shape estimation from images is increasingly becoming a very active area of research. Consequently, it is now possible to estimate human pose and shape from an RGB image only, in a large extent due to the advances in deep learning research and in optimization. Combining those advances with improvements of scanning technology, primarily scanners becoming smaller and more convenient while maintaining the high reconstruction accuracy, we conclude that accuracy and reliability of body measurements from 3D as well as from 2D data will be significantly improved in the near future.

ACKNOWLEDGMENT

This work has been supported by the Croatian Science Foundation under the project IP-2018-01-8118.

REFERENCES

[1] 3d avatar body - institute of bbomechanics, valencia. https://www.ibv.org/en/3d-avatar-body-3/. Accessed: 2020-02-09.
[2] 4d human body motion scanning. https://www.ibv.org/en/technologies/4d-human-body-motion-scanning/. Accessed: 2020-03-12.
[3] Caesar i, the most comprehensive source for body measurement data. http://store.sae.org/caesar/. Accessed: 2020-11-20.
[4] Revopoint tanso s1. https://www.revopoint3d.com/portable-3d-scanner-tanso-s1/. Accessed: 2021-03-16.
[5] Scandy time-of-flight smartphone technology. https://www.scandy.co/. Accessed: 2021-01-01.
[6] Sizeusa dataset. https://www.tc2.com/size-usa.html. Accessed: 2021-03-17.
[7] Uhuman solutions gmbh. anthroscan. http://www. https://www.human-solutions.com/. Accessed: 2020-11-20.
[8] Uk national sizing survey. http://www.size.org/. Accessed: 2020-11-20.
[9] Ahmed Abdelhafiz. Integrating Digital Photogrammetry and Terrestrial Laser Scanning. PhD thesis, 02 2009.
[10] P. Alcantarilla, J. Nuevo, and A. Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In BMVC, 2013.
[11] Brett Allen, Brian Curless, and Z. Popovic. The space of human body shapes: reconstruction and parameterization from range scans. ACM Trans. Graph., 22:587–594, 2003.
[12] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: Shape completion and animation of people. ACM Trans. Graph., 24(3):408–416, July 2005.
[13] Dragomir Anguelov, Praveen Srinivasan, Hoi-Cheung Pang, Daphne Koller, Sebastian Thrun, and James Davis. The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In Proceedings of the 17th International Conference on Neural Information Processing Systems, NIPS'04, page 33–40, Cambridge, MA, USA, 2004. MIT Press.
[14] Bruno Artacho and Andreas Savakis. Unipose: Unified human pose estimation in single images and videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.
[15] Guido Ascenso, Moi Hoon Yap, Thomas Allen, Simon S. Choppin, and Carl Payton. A review of silhouette extraction algorithms for use within visual hull pipelines. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 8(6):649–670, 2020.
[16] Ilya Baran and Jovan Popović. Automatic rigging and animation of 3d characters. ACM Trans. Graph., 26(3):72–es, July 2007.
[17] Tyler Bell, Beiwen Li, and Song Zhang. Structured Light Techniques and Applications, pages 1–24. American Cancer Society, 2016.
[18] S. Bellisai, F. Villa, S. Tisa, D. Bronzi, and F. Zappa. Indirect time-of-flight 3D ranging based on SPADs. In Manijeh Razeghi, Eric Tournie, and Gail J. Brown, editors, Quantum Sensing and Nanophotonic Devices IX, volume 8268, pages 282 – 289. International Society for Optics and Photonics, SPIE, 2012.
[19] Yannick Benezeth, Pierre-Marc Jodoin, Bruno Emile, Hélène Laurent, and Christophe Rosenberger. Comparative study of background subtraction algorithms. Journal of Electronic Imaging, 19:033003–033003, 07 2010.
[20] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):239–256, 1992.
[21] Simone Bianco, Gianluigi Ciocca, and Davide Marelli. Evaluating the performance of structure from motion pipelines. Journal of Imaging, 4:98, 08 2018.
[22] W. Boehler, M. Vicent, and A. Marbs. Investigating laser scanner accuracy. Proc. CIPA XIXth Int. Symposium, 34, 01 2003.
[23] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, P. Gehler, J. Romero, and Michael J. Black. Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In ECCV, 2016.
[24] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Piscataway, NJ, USA, June 2014. IEEE.
[25] Jonathan Boisvert, Chang Shu, Stefanie Wuhrer, and Pengcheng Xi. Three-dimensional human shape inference from silhouettes: reconstruction and validation. Machine Vision and Applications, 24(1):145–157, July 2011.
[26] Jennifer Bougourd and Philip Treleaven. Uk national sizing survey - sizeuk. pages 327–337, 10 2010.
[27] B. Bradtmiller and M. E. Gross. 3d whole-body scans: Measurement extraction software validation. SAE transactions, 108:394–400, 1999.
[28] B. Bradtmiller and M. E. Gross. 3d whole-body scans: Measurement extraction software validation. SAE transactions, 108:394–400, 1999.
[29] Sara Bragança, Pedro Arezes, Miguel Carvalho, Susan P. Ashdown, Ignacio Castellucci, and Celina Leão. A comparison of manual anthropometric measurements with kinect-based scanned measurements in terms of precision and reliability. Work, 59(3):325–339, April 2018.
[30] Adrian Bulat, Jean Kossaifi, Georgios Tzimiropoulos, and M. Pantic. Toward fast and accurate human pose estimation via soft-gated skip connections. ArXiv, abs/2002.11098, 2020.
[31] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
[32] Kyle Casadei and John Kiel. Anthropometric Measurement. StatPearls Publishing, Treasure Island (FL), 2020.
[33] Liang-Chieh Chen, Maxwell D. Collins, Y. Zhu, G. Papandreou, Barret Zoph, Florian Schroff, H. Adam, and Jonathon Shlens. Searching for efficient multi-scale architectures for dense image prediction. In NeurIPS, 2018.
[34] Yu Chen, Tae-Kyun Kim, and Roberto Cipolla. Inferring 3d shapes and deformations from single views. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, Computer Vision – ECCV 2010, pages 300–313, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.
[35] Yu Chen, Duncan Robertson, and Roberto Cipolla. A practical system for modelling body shapes from single view measurements. BMVC 2011 - Proceedings of the British Machine Vision Conference 2011, 01 2011.
[36] J. Cheng, C. Leng, J. Wu, H. Cui, and H. Lu. Fast and accurate image matching with cascade hashing for 3d reconstruction. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, 2014.
[37] Julian Chibane and Gerard Pons-Moll. Implicit feature networks for texture completion from partial 3d data. In ECCV Workshops, 2020.
[38] Jorge D. Camba, Alejandro Leon, Jorge Cantero, José Saorín, and Manuel Contero. Application of low-cost 3d scanning technologies to the development of educational augmented reality content. pages 1–6, 10 2016.
[39] H. Daanen and G. Jeroen van de Water. Whole body scanners. Displays, 19(3):111 – 120, 1998.
[40] H. A. Daanen and F. T. Haar. 3d whole body scanners revisited. Displays, 34:270–275, 2013.
[41] Morteza Daneshmand, A. Helmi, E. Avots, F. Noroozi, Fatih Alisinanoglu, H. Arslan, Jelena Gorbova, R. E. Haamer, C. Ozcinar, and G. Anbarjafari. 3d scanning: A comprehensive survey. ArXiv, abs/1801.08863, 2018.
[42] N. D'Apuzzo. Recent advances in 3d full body scanning with applications to fashion and apparel. 2009.

[43] Nicola D'Apuzzo. 3D body scanning technology for fashion and apparel industry. In J.-Angelo Beraldin, Fabio Remondino, and Mark R. Shortis, editors, Videometrics IX, volume 6491, pages 203 – 214. International Society for Optics and Photonics, SPIE, 2007.

[44] Iman Dianat, Johan Molenbroek, and Héctor Ignacio Castellucci. A review of the methodology and applications of anthropometry in ergonomics and product design. Ergonomics, 61(12):1696–1720, 2018. PMID: 30022717.

[45] Endri Dibra, Himanshu Jain, A. Cengiz Öztireli, Remo Ziegler, and Markus H. Gross. Human shape from silhouettes using generative hks descriptors and cross-modal neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21-26, 2017, 2017.

[46] M. Đ onlić. Three-dimensional analysis of back surface under dynamic conditions in scoliosis diagnostics. PhD thesis, 2019.

[47] Mostafa Ebrahim. 3d laser scanners: History, applications, and future. 10 2014.

[48] Sabry El-Hakim, J.-A Beraldin, and Francois Blais. A comparative evaluation of the performance of passive and active 3-d vision systems. Proceedings of SPIE - The International Society for Optical Engineering, 05 2003.

[49] Haoshu Fang, Guansong Lu, Xiaolin Fang, Jianwen Xie, Yu-Wing Tai, and Cewu Lu. Weakly and semi supervised human body part parsing via pose-guided knowledge transfer. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 70–78, 2018.

[50] David Fofi, Tadeusz Sliwa, and Yvon Voisin. A comparative survey on invisible structured light. SPIE Electronic Imaging-Machine Vision Applications in Industrial Inspection XII, San José, USA, 5303:90–97, 05 2004.

[51] S. Foix, G. Alenya, and C. Torras. Lock-in time-of-flight (tof) cameras: A survey. IEEE Sensors Journal, 11(9):1917–1926, 2011.

[52] Y. Furukawa and C. Hernández. Multi-View Stereo: A Tutorial. 2015.

[53] Zhongzue. Gan. Visual sensing and its applications intergration of laser sensors to industrial robots. Advanced topics in science and technology in China. Zhejiang University PressHeidelberg, Hangzhou, 2011. Includes bibliographical references and index.

[54] Jason Geng. Structured-light 3d surface imaging: a tutorial. Adv. Opt. Photon., 3(2):128–160, Jun 2011.

[55] I. A. Ghaffar and M. N. H. Mohd. Selection of android smartphones with built-in dual lens camera for stereo vision android app development. In 2018 7th International Conference on Computer and Communication Engineering (ICCCE), pages 74–78, 2018.

[56] Silvio Giancola, M. Valenti, and R. Sala. A survey on 3d cameras: Metrological comparison of time-of-flight, structured-light and active stereoscopy technologies. In SpringerBriefs in Computer Science, 2018.

[57] S. B. Gokturk, H. Yalcin, and C. Bamji. A time-of-flight depth sensor - system description, issues and solutions. In 2004 Conference on Computer Vision and Pattern Recognition Workshop, pages 35–35, 2004.

[58] Claire Gordon, Thomas Churchill, Charles Clauser, Bruce Bradtmiller, John McConville, Ilse Tebbetts, and Robert Walker. Anthropometric survey of u.s. army personnel: Summary statistics, interim report for 1988. 01 1989.

[59] J. Hafeez, A. Hamacher, S. Kwon, and S. Lee. Performance evaluation of patterns for image-based 3d model reconstruction of textureless objects. In 2017 International Conference on 3D Immersion (IC3D), pages 1–5, 2017.

[60] J. Hafeez, S. Kwon, S. Lee, and A. Hamacher. 3d surface reconstruction of smooth and textureless objects. In 2017 International Conference on Emerging Trends Innovation in ICT (ICEI), pages 145–149, 2017.

[61] Abid Haleem and Mohd. Javaid. 3d scanning applications in medical field: A literature-based review. Clinical Epidemiology and Global Health, 7(2):199–210, 2019.

[62] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Horaud. Time-of-Flight Cameras. Springer London, 2013.

[63] Richard Hartley and Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2 edition, 2004.

[64] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel. A Statistical Model of Human Pose and Body Shape. Computer Graphics Forum, 2009.

[65] Y. He and S. Chen. Recent advances in 3d data acquisition and processing by time-of-flight camera. IEEE Access, 7:12495–12510, 2019.

[66] Yihui He, Rui Yan, K. Fragkiadaki, and Shoou-I Yu. Epipolar transformers. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 7776–7785, 2020.

[67] Ying He, Bin Liang, Yu Zou, Jin He, and Jun Yang. Depth errors analysis and correction for time-of-flight (tof) cameras. Sensors, 17:92, 01 2017.

[68] Steven B Heymsfield, Brianna Bourgeois, Bennett K Ng, Markus J Sommer, Xin Li, and John A Shepherd. Digital anthropometry: a critical review. European journal of clinical nutrition, 72(5):680–687, 2018.

[69] D. Hirshberg, M. Loper, E. Rachlin, and M.J. Black. Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In European Conf. on Computer Vision (ECCV), LNCS 7577, Part IV, pages 242–255. Springer-Verlag, October 2012.

[70] R. Horaud, Miles E. Hansard, Georgios D. Evangelidis, and C. Ménier. An overview of depth cameras and range scanners based on time-of-flight technologies. Machine Vision and Applications, 27:1005–1020, 2016.

[71] Ali Hosseininaveh Ahmadabadian, Ali Karami, and Rouhallah Yazdan. An automatic 3d reconstruction system for texture-less objects. Robotics and Autonomous Systems, 117:29 – 39, 2019.

[72] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7):1325–1339, jul 2014.

[73] K. Iskakov, Egor Burkov, V. Lempitsky, and Yury Malkov. Learnable triangulation of human pose. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 7717–7726, 2019.

[74] 3-D scanning methodologies for internationally compatible anthropometric databases - Part 1: Evaluation protocol for body dimensions extracted from 3-D body scans. Standard, International Organization for Standardization, 2018.

[75] Basic human body measurements for technological design - Part 1: Body measurement definitions and landmarks. Standard, International Organization for Standardization, 2017.

[76] Cynthia Istook and Su-Jeong Shin. 3d scanning systems with application to the apparel industry. Journal of Fashion Marketing and Management, 5:120–132, 06 2001.

[77] Lingyan Jiang, Jian Yao, Baopu Li, Fei Fang, Qi Zhang, and Max Q.-H. Meng. Automatic body feature extraction from front and side images. Journal of Software Engineering and Applications, 05:94–100, 2012.

[78] Hanbyul Joo, Tomas Simon, Xulong Li, H. Liu, L. Tan, L. Gui, Sean Banerjee, Timothy Godisart, Bart C. Nabbe, I. Matthews, T. Kanade, S. Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for social interaction capture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41:190–204, 2019.

[79] Hanbyul Joo, Tomas Simon, and Yaser Sheikh. Total capture: A 3d deformation model for tracking faces, hands, and bodies. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8320–8329, 2018.

[80] A. Kanazawa, Michael J. Black, D. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7122–7131, 2018.

[81] Hiroshi Kawasaki, Ryo Furukawa, Ryusuke Sagawa, and Yasushi Yagi. Dynamic scene shape reconstruction using a single structured light pattern. 06 2008.

[82] Nikola Koepke, M. Zwahlen, J. Wells, N. Bender, M. Henneberg, F. Rühli, and K. Staub. Comparison of 3d laser-based photonic scans and manual anthropometric measurements of body size and shape in a validation study of 123 young swiss men. PeerJ, 5, 2017.

[83] Nikos Kolotouros, Georgios Pavlakos, Michael J Black, and Kostas Daniilidis. Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In ICCV, 2019.

[84] Nikos Kolotouros, Georgios Pavlakos, and Kostas Daniilidis. Convolutional mesh regression for single-image human shape reconstruction. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4496–4505, 2019.

[85] Makiko Kouchi and Masaaki Mochimaru. Errors in landmarking and the evaluation of the accuracy of traditional and 3d anthropometry. Applied Ergonomics, 42(3):518 – 527, 2011.

[86] Andrei KOVAL. Quantitative comparison of manual vs. 3d scanner human body measurements. In Proceedings of 3DBODY.TECH 2020 - 11th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, Online/Virtual, 17-18 November 2020. Hometrica Consulting - Dr. Nicola D'Apuzzo, November 2020.

[87] Andreas Kuehnapfel, Peter Ahnert, Markus Loeffler, Anja Broda, and Markus Scholz. Reliability of 3d laser-based anthropometry and comparison with classical anthropometry. Scientific Reports, 6(1), May 2016.

[88] Tsz-Ho Kwok, Kwok-Yun Yeung, and C. Wang. Volumetric template fitting for human body reconstruction from incomplete data. Journal of Manufacturing Systems, 33:678–689, 2014.

[89] Douglas Lanman and Gabriel Taubin. Build your own 3d scanner: 3d photography for beginners. ACM SIGGRAPH 2009 Courses, SIGGRAPH '09, page 8, 01 2009.

[90] A. Laurentini. The visual hull concept for silhouette-based image understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(2):150–162, 1994.

[91] Rosmery Nariño Lescay, A. Becerra, and Anaisa Hernández González. Anthropometry. comparative analysis of technologies for the capture of anthropometric dimensions. 2017.

[92] L. Li. Time-of-flight camera - an introduction. 2014.

[93] Peike Li, Y. Xu, Yunchao Wei, and Yezhou Yang. Self-correction for human parsing. IEEE transactions on pattern analysis and machine intelligence, PP, 2020.

[94] Zhaoxin Li, Wenyan Jia, Zhi-Hong Mao, Jie Li, Hsin-Chen Chen, Wangmeng Zuo, Kuanquan Wang, and Mingui Sun. Anthropometric body measurements based on multi-view stereo image reconstruction. In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, July 2013.

[95] X. Liang, K. Gong, X. Shen, and L. Lin. Look into person: Joint body parsing & pose estimation network and a new benchmark. IEEE Transactions on Pattern Analysis & Machine Intelligence, 41(04):871–885, apr 2019.

[96] Kevin Lin, Lijuan Wang, Kun Luo, Yinpeng Chen, Zicheng Liu, and Ming-Ting Sun. Cross-domain complementary learning using pose for multi-person part segmentation. IEEE Transactions on Circuits and Systems for Video Technology, PP:1–1, 05 2020.

[97] Tsung-Yi Lin, M. Maire, Serge J. Belongie, James Hays, P. Perona, D. Ramanan, Piotr Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. ArXiv, abs/1405.0312, 2014.

[98] Yueh-Ling Lin and Mao-Jiun J. Wang. Automated body feature extraction from 2d images. Expert Systems with Applications, 38(3):2585 – 2591, 2011.

[99] Yueh-Ling Lin and Mao-Jiun J. Wang. Constructing 3d human model from front and side images. Expert Systems with Applications, 39(5):5012 – 5018, 2012.

[100] Yonghuai Liu, Nick Pears, Paul L. Rosin, and Patrik Huber, editors. 3D Imaging, Analysis and Applications. Springer International Publishing, 2020.

[101] Charles Loop and Zhengyou Zhang. Computing rectifying homographies for stereo vision. volume 1, pages –131 Vol. 1, 02 1999.

[102] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. ACM Trans. Graphics (Proc. SIGGRAPH Asia), 34(6):248:1–248:16, October 2015.

[103] Matthew M. Loper, Naureen Mahmood, and Michael J. Black. MoSh: Motion and shape capture from sparse markers. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia), 33(6):220:1–220:13, November 2014.

[104] D. G. Lowe. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, volume 2, pages 1150–1157 vol.2, 1999.

[105] J. Lu and M. J. Wang. The evaluation of scan-derived anthropometric measurements. IEEE Transactions on Instrumentation and Measurement, 59(8):2048–2054, 2010.

[106] Jun-Ming Lu and Mao-Jiun J. Wang. Automated anthropometric data collection using 3d whole body scanners. Expert Syst. Appl., 35(1–2):407–414, July 2008.

[107] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In International Conference on Computer Vision, pages 5442–5451, October 2019.

[108] Charles Malleson, Marco Volino, Andrew Gilbert, Matthew Trumble, John Collomosse, and Adrian Hilton. Real-time full-body motion capture from video and imus. In 2017 Fifth International Conference on 3D Vision (3DV), 2017.

[109] J. Medina-Inojosa, V. Somers, T. Ngwa, Ling Hinshaw, and F. López-Jiménez. Reliability of a 3d body scanner for anthropometric measurements of central obesity. Obesity, open access, 2 3, 2016.

[110] Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. Monocular 3d human pose estimation in the wild using improved cnn supervision. In 3D Vision (3DV), 2017 Fifth International Conference on. IEEE, 2017.

[111] Dushyant Mehta, Oleksandr Sotnychenko, Franziska Mueller, Weipeng Xu, Srinath Sridhar, Gerard Pons-Moll, and Christian Theobalt. Single-shot multi-person 3d pose estimation from monocular rgb. In 3D Vision (3DV), 2018 Sixth International Conference on. IEEE, sep 2018.

[112] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[113] K. Norton and T. Olds. Anthropometrica : a textbook of body measurement for sports and health courses. 1996.

[114] Tim Olds and F. Honey. The use of 3d whole-body scanners in anthropometry. Kinanthropometry IX, pages 1–12, 01 2005.

[115] M. Đonlić, T. Petković, and T. Pribanić. 3d surface profilometry using phase shifting of de bruijn pattern. In 2015 IEEE International Conference on Computer Vision (ICCV), pages 963–971, 2015.

[116] Ahmed A. A. Osman, Timo Bolkart, and Michael J. Black. Star: Sparse trained articulated human body regressor. In ECCV, 2020.

[117] J. Pages, C. Collewet, F. Chaumette, and J. Salvi. An approach to visual servoing based on coded light. In Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006., pages 4118–4123, 2006.

[118] J. Pages, J. Salvi, R. Garcia, and C. Matabosch. Overview of coded light projection techniques for automatic 3d profiling. In 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422), volume 1, pages 133–138 vol.1, 2003.

[119] Jordi Pagès, Joaquim Salvi, Christophe Collewet, and Josep Forest. Optimised de bruijn patterns for one-shot shape acquisition. Image and Vision Computing, 23(8):707 – 720, 2005.

[120] J. Park, Q. Zhou, and V. Koltun. Colored point cloud registration revisited. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 143–152, 2017.

[121] Sang Il Park and Jessica K. Hodgins. Capturing and animating skin deformation in human motion. ACM Trans. Graph., 25(3):881–889, July 2006.

[122] Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, jun 2020.

[123] Georgios Pavlakos, Vasileios Choutas, N. Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 10967–10977, 2019.

[124] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli. 3d human pose estimation in video with temporal convolutions and semi-supervised training. In Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[125] Andrew D. Payne, Adrian P.P. Jongenelen, Adrian A. Dorrington, Michael J. Cree, and Dale A. Carnegie. Multiple frequency range imaging to remove measurement ambiguity, Jul 2009. Conference Contribution.

[126] F. Piron, D. Morrison, M. R. Yuce, and J. M. Redouté. A review of single-photon avalanche diode time-of-flight imaging sensor arrays. IEEE Sensors Journal, pages 1–1, 2020.

[127] L. Pishchulin, Stefanie Wuhrer, Thomas Helten, C. Theobalt, and B. Schiele. Building statistical shape spaces for 3d human modeling. Pattern Recognit., 67:276–286, 2017.

[128] Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J. Black. Dyna: A model of dynamic human shape in motion. ACM Transactions on Graphics, (Proc. SIGGRAPH), 34(4):120:1–120:14, August 2015.

[129] Tomislav Pribanić., T. Petković, David Bojanić., K. Bartol, and Mohit Gupta. Scene adaptive structured light 3d imaging. In Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP,, pages 576–582. INSTICC, SciTePress, 2020.

[130] Tomislav Pribanić, Tomislav Petković, David Bojanić, Kristijan Bartol, and Mohit Gupta. Smart time-multiplexing of quads solves the multi-camera interference problem. In Proc. 3DV, 2020.

[131] S. Prokudin, Christoph Lassner, and J. Romero. Efficient learning on point clouds with basis point sets. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 4331–4340, 2019.
[132] Albert Pumarola, Jordi Sanchez, Gary Choi, Alberto Sanfeliu, and Francesc Moreno-Noguer. 3DPeople: Modeling the Geometry of Dressed Humans. In International Conference in Computer Vision (ICCV), 2019.
[133] Haibo Qiu, Chunyu Wang, Jingdong Wang, Naiyan Wang, and Wenjun Zeng. Cross view fusion for 3d human pose estimation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 4341–4350, 2019.
[134] C. Quan, X.Y. He, C.F. Wang, C.J. Tay, and H.M. Shang. Shape measurement of small objects using lcd fringe projection with phase shifting. Optics Communications, 189(1):21 – 29, 2001.
[135] Fabio Remondino, Alberto Guarnieri, and Antonio Vettore. 3d modeling of close-range objects: Photogrammetry or laser scanning. Proc SPIE, 5665:216–225, 12 2004.
[136] M. Ribo and M. Brandner. State of the art on vision-based structured light systems for 3d measurements. In International Workshop on Robotic Sensors: Robotic and Sensor Environments, 2005., pages 2–6, 2005.
[137] K. M. Robinette, H. Daanen, and E. Paquet. The caesar project: a 3-d surface anthropometry survey. In Second International Conference on 3-D Digital Imaging and Modeling (Cat. No.PR00062), pages 380–386, 1999.
[138] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, Computer Vision – ECCV 2006, pages 430–443, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
[139] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In 2011 International Conference on Computer Vision, pages 2564–2571, 2011.
[140] A. Saint, E. Ahmed, A. E. R. Shabayek, K. Cherenkova, G. Gusev, D. Aouada, and B. Ottersten. 3dbodytex: Textured 3d body dataset. In 2018 International Conference on 3D Vision (3DV), pages 495–504, 2018.
[141] Alexandre Saint, A. Kacem, K. Cherenkova, Konstantinos Papadopoulos, Julian Chibane, Gerard Pons-Moll, G. Gusev, D. Fofi, Djamila Aouada, and B. Ottersten. Sharp 2020: The 1st shape recovery from partial textured 3d scans challenge results. In ECCV Workshops, 2020.
[142] J. Salvi, E. Mouaddib, and J. Batile. An overview of the advantages and constraints of coded pattern projection techniques for autonomous navigation. In Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS '97, volume 3, pages 1264–1271 vol.3, 1997.
[143] J. Salvi, J. Pagès, and J. Batlle. Pattern codification strategies in structured light systems. Pattern Recognit., 37:827–849, 2004.
[144] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanić, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. Pattern Recognition, 43(8):2666 – 2680, 2010.
[145] Kosuke Sato and Seiji Inokuchi. Three-dimensional surface measurement by space encoding range imaging. Journal of Robotic Systems, 2(1):27–39, 1985. cited By 105.
[146] Johannes L. Schönberger and J. Frahm. Structure-from-motion revisited. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 4104–4113, 2016.
[147] T. Schöps, J. L. Schönberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, and A. Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2538–2547, 2017.
[148] L. Sigal, A. Balan, and M. J. Black. HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International Journal of Computer Vision, 87(1):4–27, March 2010.
[149] B. M. Smith, V. Chari, A. Agrawal, J. M. Rehg, and R. Sever. Towards accurate 3d human body reconstruction from silhouettes. In 2019 International Conference on 3D Vision (3DV), pages 279–288, 2019.
[150] B. M. Smith, V. Chari, A. Agrawal, J. M. Rehg, and R. Sever. Towards accurate 3d human body reconstruction from silhouettes. In 2019 International Conference on 3D Vision (3DV), pages 279–288, 2019.
[151] E. Stoykova, A. A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis. 3-d time-varying scene capture technologies—a survey. IEEE Transactions on Circuits and Systems for Video Technology, 17(11):1568–1586, 2007.
[152] L. Streeter and Y. C. Kuang. Metrological aspects of time-of-flight range imaging. IEEE Instrumentation Measurement Magazine, 22(2):21–26, 2019.
[153] K. Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 5686–5696, 2019.
[154] Z. Sun, G. Qian, Z. Peng, W. Dai, D. Sun, G. Zhang, N. Zhang, J. Xu, R. Wang, and C. Li. Orthogonal coded multi-view structured light for inter-view interference elimination. In 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP), pages 181–184, 2020.
[155] P. Tikuisis, P. Meunier, and C. Jubenville. Human body surface area: measurement and prediction using three dimensional body scans. European Journal of Applied Physiology, 85:264–271, 2001.
[156] Philip Treleaven and Jonathan Wells. 3d body scanning and healthcare applications. Computer, 40:28 – 34, 08 2007.
[157] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment — a modern synthesis. In Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, Vision Algorithms: Theory and Practice, pages 298–372, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.
[158] Matthew Trumble, A. Gilbert, Charles Malleson, A. Hilton, and J. Collomosse. Total capture: 3d human pose estimation fusing video and inertial sensors. In BMVC, 2017.
[159] A. Tsoli, M. Loper, and M. J. Black. Model-based anthropometry: Predicting measurements from 3d human scans in multiple poses. In IEEE Winter Conference on Applications of Computer Vision, pages 83–90, 2014.
[160] Sam Van der Jeught and Joris J.J. Dirckx. Real-time structured light profilometry: a review. Optics and Lasers in Engineering, 87:18 – 31, 2016. Digital optical & Imaging methods in structural mechanics.
[161] Nick Van Gestel, Steven Cuypers, Philip Bleys, and Jean-Pierre Kruth. A performance evaluation test for laser line scanners on cmms. Optics and Lasers in Engineering, 47(3):336 – 342, 2009. Optical Measurements.
[162] Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In CVPR, 2017.
[163] Jianfeng Wang, Cha Zhang, Wenwu Zhu, Zhengyou Zhang, Zixiang Xiong, and Philip Chou. 3d scene reconstruction by multiple structured-light based commodity depth cameras. pages 5429–5432, 03 2012.
[164] Jiantian Wang, John Thornton, S Kolesnik, and Richard Pierson. Anthropometry in body composition: An overview. Annals of the New York Academy of Sciences, 904:317–26, 06 2000.
[165] M. Wang, M. Wang, and Y. Lin. Hsinchu: The Ergonomics Society of Taiwan, 2000.
[166] Mao-Jiun J. Wang, Wen-Yen Wu, Kuo-Chao Lin, S. Yang, and Jun-Ming Lu. Automated anthropometric data collection from three-dimensional digital human models. The International Journal of Advanced Manufacturing Technology, 32:109–115, 2007.
[167] Felix Wermke, Thorben Wübbenhorst, and B. Meffert. Interference avoidance for two time-of-flight cameras using autonomous optical synchronization. 2020 6th International Conference on Control, Automation and Robotics (ICCAR), pages 586–595, 2020.
[168] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.
[169] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, C. Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32:4–24, 2021.
[170] Pengcheng Xi, Won-Sook Lee, and Chang Shu. A data-driven approach to human-body cloning using a segmented body database. pages 139–147, 10 2007.
[171] Fangting Xia, P. Wang, X. Chen, and A. Yuille. Joint multi-person pose estimation and semantic part segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 6080–6089, 2017.
[172] S. Yan, Johan Wirta, and Joni-Kristian Kämäräinen. Anthropometric clothing measurements from 3d body scans. Machine Vision and Applications, 31:1–11, 2020.
[173] Song Yan and Joni-Kristian Kämäräinen. Learning anthropometry from rendered humans, 01 2021.
[174] N. Zakaria and D. Gupta. Anthropometry, Apparel Sizing and Design. The Textile Institute Book Series. Elsevier Science, 2019.

[175] Chao Zhang, Sergi Pujades, Michael J. Black, and Gerard Pons-Moll. Detailed, accurate, human shape estimation from clothed 3d scan sequences. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[176] F. Zhang, Xiatian Zhu, Hanbin Dai, M. Ye, and C. Zhu. Distribution-aware coordinate representation for human pose estimation. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 7091–7100, 2020.

[177] Yan Zhang, M. Hassan, Heiko Neumann, Michael J. Black, and Siyu Tang. Generating 3d people in scenes without people. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 6193–6203, 2020.

[178] Jian Zhao, Jianshu Li, Yu Cheng, Terence Sim, Shuicheng Yan, and Jiashi Feng. Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing. In Proceedings of the 26th ACM International Conference on Multimedia, MM '18, page 792–800, New York, NY, USA, 2018. Association for Computing Machinery.

[179] Onur Özyeşil, Vladislav Voroninski, Ronen Basri, and Amit Singer. A survey of structure from motion. Acta Numerica, 26:305–364, 2017.

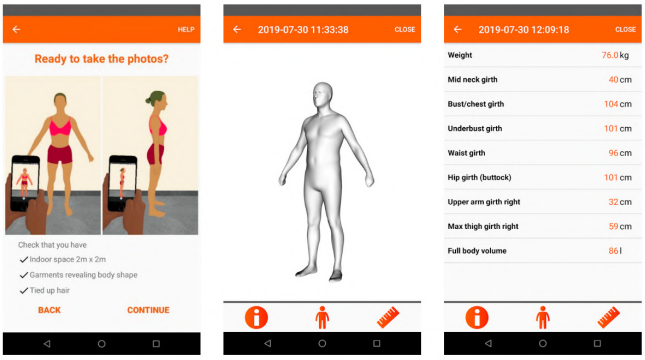


FIGURE 16: A common mobile application body measurement pipeline is to take front and side images, estimate 3D human mesh, and assess the body measurements from the mesh. Image credits: [1].

APPENDIX A - 3D SCANNERS

Table VI presents an overview of the commercial 3D scanners that have the ability to scan human bodies, excluding scanners that are not fit to the task, such as the Revopoint Tanso S1 [4], used to reconstruct smaller objects. We provide more than 80 currently available 3D scanners manufactured by more than 50 companies, as well as its taxonomy regarding several key characteristics: their mobility, method of reconstruction, price, resolution, accuracy, number of sensors, dimensions, provided texture, scanning time and provided anthropometric software. Additionally, we comment on their affect on human body scanning.

We observe an equal amount of stationary (booth-like) scanners as handheld ones, whereas only a few mini scanners on the market. While handheld scanners offer a quicker scanning setup time in new environments, stationary scanners are more ideal for fixed scenarios, omitting (almost) entirely the setup process. Naturally, the mobility of a scanner is correlated with its dimensions. Stationary scanners are large and bulky, while mini scanners are compact and portable. Hence, mini and handheld scanners offer better applicability to the task of distributed data collection process [174] since they present higher portability. On the other hand, stationary scanners offer faster scanning times in the seconds' range, while handheld scanners offer scanning times in the minutes' range; presenting a trade-off between their dimensions and applicability. Since breathing and fidgeting causes human bodies to move during the scanning process, faster scanning times are more desirable. Nevertheless, the performance of handheld scanners does not seem to lag behind the stationary 3D scanners, as seen by their accuracy.

The mobility and scanning time of a scanner, seem to mostly drive its price. Smaller scanners tend to be cheaper, while scanners offering faster scanning times tend to be pricier, indicating that the market is still more appreciative towards stationary scanners. Most of the scanners use structured light (SL) to reconstruct the human body since it offers the best reconstruction accuracy within the methods presented in Sec. III. Additionally, they present the lowest resolution, followed by passive stereo (PS) and time-of-flight (ToF), respectively. Hence, they allow dense 3D human body reconstructions, appropriate for the anthropometric application. To this end, we additionally report if the scanner comes with an anthropometry software that can automatically extract body measurements from a 3D scan. While texture does not directly impact the scanning process, arguments have been made in favour of greater usecase for textured 3D human body models [140].

The market is moving towards handheld and mini scanners. Mini scanners are particularly important for the future of tablet and smartphone scanning because they can be attached or even embedded into devices. For example, the Occipital sensors can be attached to a smartphone device, while the Apple iPhone 12 has a LIDAR sensor embedded (see Appendix B). Mini scanners are usually ToF-based, which can be seen from Table VI. As the computing capabilities of mobile devices further improve and ToF-based mini scanners increase the resolution, we expect that mobile devices will more reliable and accurate 3D scanners.

APPENDIX B - MOBILE APPLICATIONS

Mobile phones have become an emerging market for making 3D scanners more approachable. Multiple cameras [55], new ToF sensors [5], and general development of said phones, have made the implementation of 3D scanning techniques easier. Additionally, stationary scanners are relatively bulky and pricey. Hence, mobile 3D scanning technology has become important, particularly for the distributed data collection process [174]. In this Subsection, we present a comprehensive list of available applications for 3D human body measurement estimation (see Table VI).

The majority of the existing applications use a single RGB camera for computing body measurements. The most common approach (as seen in MeThreeSixty, Meepl, 3DAvatarBody, and many others) is to fit a tempalte mesh to a front and side image of the subject (see Fig. 16). The

measurements can then be extracted from the template mesh, as described in Sec. IV. Some of the applications extract the measurements from a single image (Nettelo) and some take multiple images from different angles and rely on photogrammetry for 3D reconstruction and measurements (3DCreator, Qlone, Scann3D, 3DAvatarBody, Two Pictures 3D BODYSCAN, Mobile Scanner, SizeYou, 3D Scanner Pro, BodyGee Coach App). A few applications use the 3D Occipital scanner attached on the smartphone device (ItSeez3D, TechMed3D, Occipital original app), while one (Scandy Pro) uses Apple's LIDAR embedded sensor to directly retrieve 3D human scans.

TABLE 6: The existing 3D scanners on the market capable of scanning the human body in its entirety. Each scanner is described by their respective mobility gradation: mini, handheld and stationary, with mini being the most portable ones. "Method" describes the 3D reconstruction approach as discussed in Sec. III. Additionally, we denote laser-based SL approaches with "SL*" and projector-based SL approaches with "SL". The "Acc." and "Res." define the reconstruction capabilities of the scanners in terms of accuracy and resolution. The measures are given in millimetres. "No. Sens." reports the number of cameras and lighting sensors in the given scanner. The "Dims." column reports the dimension of the scanner in centimetres. The dimension can be given as a product of the height, width and depth, or as a product of the diameter and the height of the scanner. The "Scan. Time" column reports the scanning duration in seconds of one single working volume. "Antrhopo." reports if anthropometric measurement extraction is available in the given software. The remaining "Link", "Price" and "Texture" columns are self-explanatory.

Manufact.	Product	Link	Mobility / Size	Method	Price ($)	Res. (mm)	Acc. (mm)	No. Sens.	Dims. (cm)	Texture	Scan. Time	Anthropo.
SizeStream	SS20	sizestream.com	Stationary	ToF	from 15k	-	-	20	145x188x203	Yes	-	-
Vitronic	Vitus	vitronic.com	Stationary	SL*	5k-10k	-	1	8	-	Yes	-	Yes
Texel	Portal BX	texel.graphics	Stationary	SL*	31k	1	-	-	225×258	Yes	-	Yes
	Portal MX		Stationary	SL*	26k	1	1	-	260×60×60	Yes	-	Yes
IBV	Move4D	ibv.org	Stationary	PS	-	1	-	12+	200x200x300	Yes	-	-
Artec	ArtecLeo	artec3d.com	Handheld	SL	29.8k	0.2	0.1	3	23.1x16.2x23	Yes	-	Yes
	Artec Eva		Handheld	SL	19.8k	0.2	0.1	3	26.2x15.8x6.3	Yes	-	Yes
	Artec Eva Lite		Handheld	SL	9800	0.5	0.1	2	26.2x15.8x6.3	No	-	Yes
	Artec Space Spider		Handheld	SL	24.8k	0.1	0.05	5	19x14x13	Yes	-	Yes
	Shapify Booth		Stationary	SL	180k	1.5	0.25	12	330x330x280	Yes	12 s	Yes
Thor3D	Calibry	thor3dscanner.com	Handheld	SL	5790	0.6	0.1	5	16.5x8.5x27.3	Yes	60 s	-
Fit3D	Fit3D	fit3d.com	Stationary	PS	10k	-	-	3	-	-	-	Yes
Styku	Styku S100	styku.com	Stationary	ToF	10k	-	-	-	254x254x117	No	-	Yes
Revopoint3D	Handysense	revopoint3d.com	Handheld	SL	3000	0.3	0.1	3	21.5x12x33.6	Yes	300 s	-
	Acusense A1		Mini	SL	1000	-	0.1-1	3	15x25x38	Yes	-	No
Apple	iPhone 12 Pro Lidar		Mini	ToF	999	-	-				-	No
HP	HP Pro S3	hp.com	Stationary	SL	3400	0.05	-	2	-	Yes	-	No
DexaFit	DexaFit	dexafit.com	Stationary	SL	-	-	-	-	-	No	-	Yes
botspot	botscan Neo	botspot.de	Stationary	SL	-	0.1	-	120	305×246	Yes	-	No
	botscan Pro S		Stationary	SL	-	0.2	-	300	355x2600	Yes	-	No
	botscan Cargo		Stationary	SL	-	0.2	-	70	605x243x259	Yes	-	No
	OptaOne+		Stationary	SL	10k- 50k	0.2	-	68	314x254	Yes	-	No
3dMD	3dMDBody	3dmd.com	Stationary	PS	-	0.7	-	78+	-	Yes	-	No
	3dMDflex		Stationary	PS	20k- 50k	-	0.2	27	-	Yes	-	No
TechMed3D	BodyScan Scanner	techmed3d.com	Handheld	SL	-	-	-	2	-	Yes	-	Yes
4DDynamics	IIID Body	4ddynamics.com	Stationary	SL	-	-	0.5	10	170x170x210	-	-	Yes
	IIID Trailer		Stationary	SL	-	-	-	-	-	-	-	Yes
	IIID ScanBooth		Stationary	SL	-	-	-	-	170x170x170	-	-	Yes
	Memphisto EOS		Stationary	SL	-	-	-	2	-	-	-	-

Manufact.	Product	Link	Mobility / Size	Method	Price ($)	Res. (mm)	Acc. (mm)	No. Sens.	Dims. (cm)	Texture	Scan. Time	Anthropo.
	Memphisto EX		Stationary	SL	-	-	-	3	-	-	-	-
	Pico		Stationary	SL	-	-	-	2	-	-	-	-
	Pico Pro		Stationary	SL	-	-	-	2	-	-	-	-
	Gotcha		Handheld	SL	-	-	-	3	-	-	-	-
	Gotcha Pro		Handheld	SL	-	-	-	3	-	-	-	-
Shape Labs	ShapeScale	shapescale.com	Stationary	SL	499	3.1	1.58	3	120x145	Yes	60 s	Yes
Naked Labs	Naked	nakedlabs.com	Stationary	ToF	1395	-	5	3	158x30x30	No	20 s	Yes
mPort Ltd.	mPod	mport.com	Stationary	SL	-	-	10	-	-	No	-	Yes
Telemat Industrie	Symcad II ST	telmat.fr	Stationary	SL	44k	-	-	4	402x160x235	-	-	-
	Symcad II HD		Stationary	SL	31.3k	-	-	4	358x134x230	-	-	-
	Symcad III		Stationary	SL	15.9k	-	-	16	190x173x210	-	-	-
$[TC]^2$ Labs	TC2-105	tc2.com	Stationary	SL	100k- 250k	0.7	0.1	-	-	Yes	-	No
	TC2-30R		Stationary	ToF	-	2	-	-	177x102	Yes	-	Yes
	TC2-19M		Stationary	ToF	-	1	-	-	-	-	-	Yes
	TC2-21B		Stationary	PS	30k	-	-	-	-	-	-	-
	TC2-19R		Stationary	ToF	1k- 10k	-	-	-	-	-	-	-
	TC2-19B		Stationary	ToF	-	-	-	-	-	-	-	-
Spacevision	SCUVEG4-Portable	spacevision.tokyo	Stationary	-	50k- 100k	-	2	-	205x60x80	-	-	-
	SCUVEG4-Flex		Stationary	-	50k- 100k	-	2	-	205x60x80	-	-	-
QuantaCorp	Shapewatch	quantacorp.io	Stationary	PS	-	-	-	-	-	-	-	Yes
Mantis Vision	Studio 3iosk	mantis-vision.com	Stationary	PS	30k	-	-	15	250x247	Yes	-	-
	Studio 3iosk XT		Stationary	PS	-	-	-	16	250x247	Yes	-	-
	F6 SR		Handheld	SL	10k- 50k	0.4	0.1	3	-	Yes	-	No
	F6 Smart		Handheld	SL	10k- 50k	0.4	0.5	3	-	Yes	-	No
TG3D Studio	TG 2000-F	tg3ds.com	Stationary	SL*	from 15k	-	-	-	152x132x202	-	-	Yes
GOM	ATOS Q	gom.com	Handheld	SL	from 60k	0.03	-	3	34x24x8.3	No	-	-
ScanTech	KScan 20	3d-scantech.com	Handheld	SL	from 40k	0.01	0.02	6	-	No	-	-
	KScan Magic II		Handheld	SL	from 40k	0.01	0.02	-	-	No	-	-
	Prince 775		Handheld	SL	from 40k	0.02	0.03	-	31x16x10	No	-	-
	HScan 771		Handheld	SL	from 40k	0.05	0.03	-	31x16x10	No	-	-
	Axe B17		Handheld	SL	from 40k	0.025	0.02	-	-	No	-	-
	IReal 2E		Handheld	SL	4980	3	0.1	9	14x9.4x25.8	Yes	-	No
Hexagon	Aicon Primescan	hexagonmi.com	Stationary	SL	35k	0.016	0.016	3	300x210x175	No	-	No
EvixScan3D	Heavy Duty Basic	evixscan3d.com	Stationary	SL	10k- 50k	-	0.02	3	430x220x65	Yes	-	-

Manufact.	Product	Link	Mobility / Size	Method	Price ($)	Res. (mm)	Acc. (mm)	No. Sens.	Dims. (cm)	Texture	Scan. Time	Anthropo.
	Heavy Duty Optima		Stationary	SL	10k- 50k	-	0.0183	3	430x220x65	Yes	-	-
	Heavy Duty Quadro		Stationary	SL	10k- 50k	-	0.013	5	520x280x95	Yes	-	-
Polyga	Carbon	polyga.com	Stationary	SL	14.9k	0.362	0.05	2	13x41x12	Yes	1.2 s	-
	Compact L6		Stationary	SL	11.9k	0.18	0.08	2	5.5x12.9x40	Yes	1.2 s	No
	Polyga H3		Handheld	SL	9990	0.5	0.08	2	28x20x6	Yes	-	No
Shining 3D	Freescan X7	shining3d.com	Handheld	SL*	-	0.05	0.03	2	13x9x31	No	-	-
	EinscanPro 2X		Handheld	SL	5499	0.2	0.04	3	37x36.5x13.5	Yes	-	-
	EinscanPro 2X Plus		Handheld	SL	6899	0.2	0.04	3	37x36.5x13.5	Yes	-	-
	Einscan H		Handheld	SL		3	0.05	-	10.8x11x23.7	Yes	-	-
Peel 3D	Peel 1	peel-3d.com	Handheld	SL	5990	0.5	0.25	2	9.6x14x25.8	No	90 s	Yes
	Peel 2		Handheld	SL	7690	0.5	0.25	4	15x17.1x25.1	Yes	90 s	Yes
Faro	Freestyle3D	faro.com	Handheld	SL*	10k- 20k	0.2	1.5	-	260x310x105	Yes	-	-
	Freestyle3D X		Handheld	SL*	10k- 50k	0.2	1	-	26x31x10.5	Yes	-	-
Creaform	GO!Scan Spark	creaform3d.com	Handheld	SL	39.9k	0.2	0.05	-	8.9x11.4x34.6	Yes	-	-
	HandyScan Black		Handheld	SL*	50k- 100k	0.1	0.035	9	14.2x7.9x28.8	No	-	-
	HandyScan Black Elite		Handheld	SL*	50k- 100k	0.025	0.025	11	14.2x7.9x28.8	No	-	-
Occipital	Structure Sensor	structure.io	Mini	SL	379	-	-	3	-	Yes	-	No
BodyGee	Orbiter	bodygee.com	Stationary	ToF	-	-	-	-	-	Yes	90 s	Yes
	Boxx		Stationary	ToF	-	-	-	-	-	Yes	-	Yes
Intel	RealSense LiDAR Camera L515	intelrealsense.com	Mini	ToF	349	-	14	2	6.1x2.6	-	-	No
	RealSense Depth Camera D455		Mini	PS	239	-	-	3	12.4x2.6x2.9	No	-	No
	RealSense Depth Camera D435i		Mini	PS	179	-	-	4	9x2.5x2.5	No	-	No
	RealSense Depth Camera D415		Mini	PS	149	-	-	3	9.9x2x2.3	No	-	No

TABLE 7: The existing 3D body measurement mobile applications for iOS and Android. We distinguish between the apps based on the company, main application, scanner and the OS. Note that the main applications specified in the Table are retrieved based on the app description and the actual purpose might differ in reality. The applications that use 2D input data only are listed above the double line; the ones that also use 3D data are given below the line. The scanner that is used for data acquisition is specified within the "Scanner" column. The values denoted with "-" are not available.

Company	App	Manufacturer link	Main application	Scanner	OS
Sony	3DCreator	sony.com	Entertainment	RGB camera	Android
Standard Cyborg	Capture: 3D Scan Anything	standardcyborg.com	Entertainment	RGB camera	iOS
EyeCue	Qlone	qlone.pro	Entertainment	RGB camera	iOS+Android
SmartMobileVision	Scann3D	smartmobilevision.com	Entertainment	RGB camera	Android
IBV	3DAvatarBody	www.ibv.org	Fitness / Fashion	RGB camera	Android
3DLook	Mobile Tailor	3dlook.me	Fashion	RGB camera	-
	Your Fit		Fashion	RGB camera	-
	Uniform Pro		Fashion	RGB camera	-
QuantaCorp	Two Pictures 3D BODYSCAN	quantacorp.io	Fitness / Fashion	RGB camera	iOS+Android
Size Stream LLC	Mobile Scanner	www.sizestream.com	Fashion	RGB camera	iOS+Android*
	MeThreeSixty		Fitness	RGB camera	iOS+Android
Fision AG	Meepl	meepl.com	Fashion	RGB camera	iOS+Android
NetVirta	NetVirta	netvirta.com	Fashion	-	-
SizeYou	SizeYou	sizeyou.it	Fashion	RGB camera	iOS+Android
Nettelo	Nettelo	nettelo.com	Fashion	RGB camera	Android
Xplorazzy Tech	3D Scanner Pro	xplorazzi.com	Entertainment	RGB camera	Android
BodyGee	BodyGee Coach App	bodygee.com	Fitness	RGB camera	Android
Scandy	Scandy Pro	scandy.co	Entertainment	LIDAR	iOS
Itseez3D, Inc.	ItSeez3D	itseez3d.com	Entertainment	Occipital	iOS
TechMed3D	3DSizeMe	techmed3d.com	Entertainment	Occipital	iOS

BARTOL KRISTIJAN received the B.Sc. degree in computing (2016) and the M.Sc. degree in computer science (2019). He is currently pursuing his Ph.D. in Computing at the University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia. His current research interests include computer vision, human pose and shape estimation and deep learning. He is a Student Member of IEEE.

BOJANIĆ DAVID received his B.Sc. in Mathematics (magna cum laude, 2016) and M.Sc. in Mathematics (2019) at the University of Zagreb, Croatia. He is currently pursuing his Ph.D. in computer science at the University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia. His current research interests lie in computer vision, focusing on shape analysis and 3D reconstruction of human bodies. He is a Student Member of IEEE.

PETKOVIĆ TOMISLAV is Assistant Professor at the University of Zagreb Faculty of Electrical Engineering and Computing. He teaches several graduate courses in the field of digital signal and image processing. His main fields of interest are digital image processing and analysis, depth sensing and 3D imaging. He received the engineer's degree, magister degree and Ph.D. in electrical engineering all from the University of Zagreb, in 2002, 2006 and 2010 respectively. He is a member of IEEE, ACM and HDBIMF.

PRIBANIĆ TOMISLAV is Professor at the University of Zagreb Faculty of Electrical Engineering and Computing. He teaches several undergraduate and graduate courses in the field of algorithms and data structures, image processing, sensors and human motion analysis. His main research interests include computer vision and biomedical signal measurement and analysis. He has led a number of domestic and international scientific projects, collaborating with researchers from within and outside the EU. He was a visiting researcher at INRIA Rhone-Alpes, Grenoble, France and Fraunhofer IGD, Darmstadt, Germany, and Fulbright Visiting Scholar at the University of Wisconstin, Madison, USA. The results of his research have been implemented in technological projects and he has also received recognition for innovations. He is a senior member of IEEE and IFMBE and a collaborating member of the Croatian Academy of Engineering.