



3D SfM as a Measuring Technique for Human Body Transformation

Alessandro Marro^(✉), Stefan Wiesen, Max Langbein, and Hans Hagen

TU Kaiserslautern, Kaiserslautern, Germany
marro@rhrk.uni-kl.de
<https://hci.uni-kl.de/>

Abstract. The tracking of fat loss as well as muscle gain has always been one of the most important steps during a person's fitness journey. It does not only motivate to continue practicing exercises, but also helps to develop specific workout plans to enhance particular body parts of athletes. Structure for Motion (SfM), unlike other reconstruction techniques, produces acceptable results from low-quality inputs. This makes the method applicable for ubiquitous equipment like a smartphone camera, while still being scalable to professional environments with proper equipment. In order to track overall body transformation, we propose a photogrammetry workflow employing SfM, reproducibly generating a model of the human body in different stages of a fitness plan. For visualization, we do a mesh alignment step followed by a comparison between the reconstructed body models of the subject, resulting in color-mapped meshes. Following this workflow the transformation of specific body regions can be analyzed in detail, only using consumer hardware.

Keywords: Body transformation · SfM · Photogrammetry

1 Introduction and Related Work

Simple indicators of health condition and fitness level are changes in human body shape. They motivate people physically and psychologically. Some of the methods tracking body transformations are targeted to internal changes, like fat loss and muscle gain [4]. Others measure body shape, including volumes and dimensions, to track external body transformations. *3D Body Scanners* (3D-BS) were used by the clothing market and *3D Photonic Scanners* (3D-PS) in medical clinics to take volumetric measurements of the human body [6, 8, 9]. They have shown high validity and reliability. However, most of this equipment is too expensive for general consumers, like personal trainers, nutritionists and coaches. Simpler measurements, such as skinfold thickness (SF) and body mass index (BMI) are also used, but their outcome is a crude index of the body shape and can not differentiate fat and lean mass. For further readings, Wells et al. [7] discuss about the different techniques for measuring body composition.

Table 1. Comparison of different techniques from best (green) to worst (red)

| | SF/BMI | 3D-BS | 3D-PS | Our Method |
|-----------------|--------|-------|-------|------------|
| Cost | | | | |
| Accessibility | | | | |
| Accuracy | | | | |
| Level of Detail | | | | |
| Speed | | | | |

The Problems can be summarized in difficulties such as expense, access to hardware and portability, single variable to describe the body and lack of visual outcome to represent the changes to the body. *Our Proposed Solution* uses SfM employed in a photogrammetry pipeline to reconstruct, up to a scale factor, the subject’s body as a 3D model, which is later on rescaled to real world scale and aligned to previous reconstructions. As a result, one can analyze the physical progress made at specific body parts. Figure 1 shows the outcome of the whole process as well as each performed stage. SfM has been used previously as main method of reconstruction by known research projects [2,3,5], and since then it has become widely used by commercial softwares. Due to the method’s concept of taking several photos of a still scene in different time steps, static objects are reliably rebuilt. However, when applied to subjects that change their shape or move relatively to the scene, aberrations occur. One solution to that problem are layers of cameras which are synchronized to shoot at the same time, but this is expensive due to the number of cameras needed. Our solution is to help the

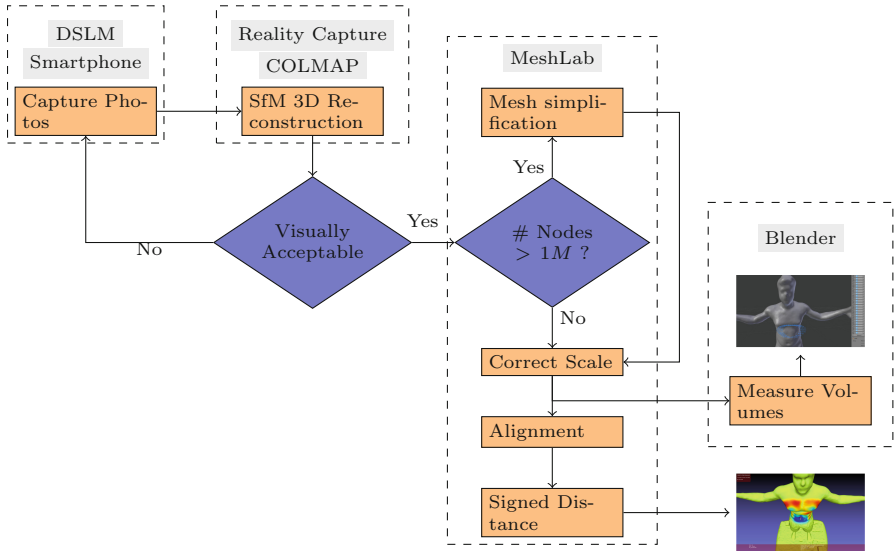


Fig. 1. Data processing pipeline

subject to keep a relaxed pose under the constraint of having all important body parts non-occluded, using two tripods as a handle. According to Wells [9] most of the unsuccessful scans made with 3D-PS were primarily due to body movement, or software inability to reconstruct the desired shape, which also applies to our method. Table 1 shows a basic comparison between the cited methods.

2 Method

2.1 Image Capturing

Location and Situation. Before we got to our final location we tested several areas with varying sizes, heights and light conditions. The best overall results were achieved in a subsection of the gym at TU Kaiserslautern with the dimensions $18 \times 27 \times 9$ m (LxWxH). The area is well illuminated by three tent-shaped skylights and light gray walls, floor (with colored stripes) and ceiling (Fig. 2).



Fig. 2. Dimensions and location

In general it should be considered to have a comparable uniformly lit environment, with enough space to walk around the person. Furthermore a closed, quiet area with no movement in the background and no other people in scene is important.

Base and Reference. The base (Fig. 3) consists of a flat PVC flooring sized 2×1.5 m, with fixed positions for tripods (used as handles) and cubes, size-corresponding footprints and markers composed of facing letters (A & T, U & W) surrounding the footprints. Two 3D-printed 10 cm wide cubes, color-coded and textured with shapes and letters, as shown in Fig. 4, are added as size reference. The base and all its elements are used as reference. The distances between every tripod leg, the height of the handles (1.40 m each) and the cubes are known values used for scaling to have a true-sized reconstruction.

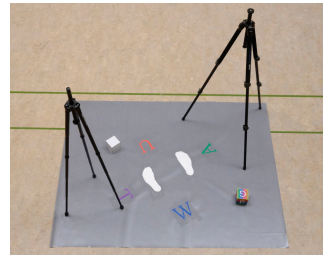


Fig. 3. Base plane, cubes, markers and tripods

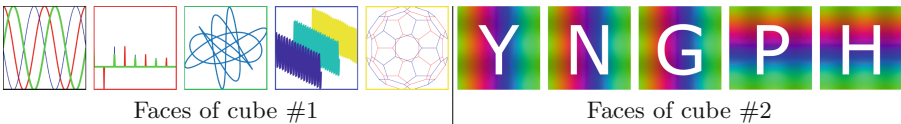


Fig. 4. Faces of the reference cubes

Test Person and Position. The test person stands upright on the fixed marked positions for their feet with stretched arms grabbing the handles (tripods). The tripods define the height (z-coordinate) and the local position (x- and y-coordinate) of the hands. In this position the hands and arms are accurately located in the same way for every capturing process. It is necessary to stretch the arms in a way like this to capture all parts of the upper body and the arms. The T-shaped stand provides a straight back and horizontal stretched arms. This position generates a little tension on the shoulder and arm muscles by rising up the arms and grasping the prepared handles. It is important that the test person is not moving during the capturing process, excluding subconscious body functions like breathing, blinking etc. This relative comfortable and mostly relaxed posture and the controlled environment are necessary.

Camera Position and Movement. The capturing process should be done as fast as possible to have a minimum of movement of the test person. To generalize it: the lower the capturing time and the lower the movement of the person, the higher the reconstruction quality.

Considering this we developed a routine for the capturing process. All images are shot in portrait format. The following numbers are representative for the work with the DSLM camera from Table 2. The first set of images (around 50–60 pictures) need to be captured in a wide circle with a radius of ≈ 6 m around the test person with equal distances from the lower/upper end of the subject to the image border. The focus should be fixed at the center of the upper body. This set is needed to capture the whole setup and to match further detail images to the body. The second set is needed to get more detailed images, starting with a turn in chest level with a ≈ 2 –3 m distance around the person. After completing this turn the camera is risen in an overhead position (≈ 40 –50 cm above) so that the test person and the base are visible. After every full turn around the subject the height decreases so that the camera positions are spherically arranged with the person in center. This second step should produce around 230 to 270 pictures. Now the base plane and the cubes with the lower body of the person are captured, so that the person can start relaxing the upper body. The focus is set to automatic continuous mode and the camera release is

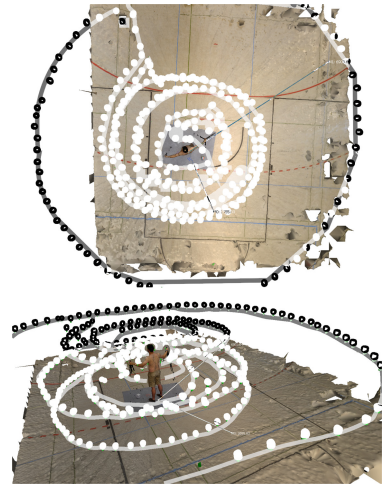


Fig. 5. Camera positions for capturing, top and side view

programmed to shoot every second to have an overlap of $>60\%$ and an angle of $\approx 15^\circ$ between two consecutive images. In total it takes between 360 s and 440 s to take all pictures. The test person is not allowed to move 300 s to 380 s. The instructions defined are the ones that worked best in our condition, but they can vary depending on how much space is available. If the person keeps the T-shaped stand and all angles have been captured with sharp images, the reconstruction will be acceptable.

Doing the same process with a smartphone, the duration of the procedure and the amount of pictures taken will decrease, because of a wider angle of view of 78.3° . In total around 260 photos in about 370 s were produced. The test person needs to stand still for about 300 s. The smartphone was released manually (Fig. 5).

Equipment. For the experiments we need besides the location, the base plane materials and tripods a smartphone and a camera for image capturing. The used devices are technically described in Table 2.

Table 2. Hardware specification smartphone and DSLM

| Type | DSLM | Smartphone |
|-----------------------|-------------------------------------|--|
| Name | Fujifilm X-T1 | Samsung Galaxy S7 |
| Sensor size | 23.6 mm \times 15.6 mm (APS-C) | 5.76 mm \times 4.29 mm (IMX260 1/2.55") |
| Sensor surface | 368.16 mm ² | 24.71 mm ² |
| Resolution | 16.3 MP | 12.2 MP |
| Pixel size | 22.5865 μm^2 | 2.0255 μm^2 |
| Lens (focal length) | 27 mm f2.8 | 4.2 mm f1.7 |
| Aperture | f6.4 | f1.7 (fixed aperture) |
| ISO | 3200 | 50-800 (auto ISO) |
| Full frame equivalent | 41 mm f10 | 26.6 mm f10 |
| Angle of view | 55.7 | 78.3 |

2.2 3D Reconstruction Using SfM

SfM output quality depends on the quality of the images taken in the previous step. Images out of focus, with low resolution and/or high distortion, bad lighting conditions, among other factors will probably lead to low quality reconstructions, filled with strange artifacts, holes, deformations, i.e. unacceptable results. Even if a set of images with acceptable quality is used, other factors can lead to poor outputs, such as occluded body parts and lack of reference points. Therefore, the image capturing workflow explained previously need to be followed in order to achieve acceptable reconstructions. For this step, the paid software Reality Capture was used. For a reliable free alternative COLMAP could also be used. Since the proposal of this work is to present more accessible alternatives to expensive body measuring devices, Fig. 11 shows that it is possible to achieve acceptable results using COLMAP and a smartphone camera. Reality Capture performs the usual SfM pipeline. First, it starts by aligning images using detected features presented in each image and matched in between pairs. An initial model is then calculated as a sparse point cloud by triangulating matched feature points. After that, several optimizations are done by calculating the error of reprojection of each triangulated point. At this moment, the positions of the cameras are recalculated based on the found error and the images are realigned. A dense point cloud can be estimated after the new optimized camera registration. The mesh is then extracted and textured (Fig. 6). To ensure precision, we measured the height of the person's model and the length of edges of the reference cube model in Meshlab [1], both resulting in 180.6 cm and 9.99 cm, respectively. In comparison to the real measurements of 180.4 cm and 10 cm, which were acquired using a measuring tape, the person's height was off by 0.2 cm and the cube's edge by 0.01 cm.

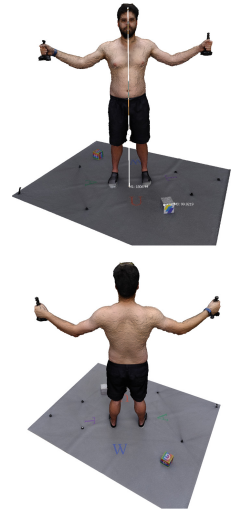


Fig. 6. Results: edge length cube: 99.92 mm
- height of test person: 1806.5 mm

2.3 Mesh Alignment

A mesh alignment step is a mandatory process to prepare the models for comparison. The final result of the reconstruction step is a representative model of the subject on its local coordinate system. Therefore, each mesh is in its own orientation, position and scale. The alignment starts by rescaling the meshes to real world scale, that can be easily done due to our reference cubes. Having all meshes in the same scale, we perform a user-defined point alignment. It works by selecting a set of points in each of the models, which are used to estimate a transformation matrix containing translation and rotation information. The “moving mesh” is transformed to align the defined “fixed mesh”. Figure 7 illustrates the whole procedure done in Meshlab. In our case, we defined the vertices

of our reference cubes as the set of points, totalizing 16 points in each mesh. The alignment of the meshes is a hard task, especially when dealing with human or animal subjects, which are in constant movement. We tried to counteract this by posing the subject as relaxed as possible using the tripods, and always positioning his feet over the markers in every photo section. Otherwise, the alignment would not be practicable without independently modifying parts of the mesh, that would cause unacceptable changes in the body volumes and measurements.

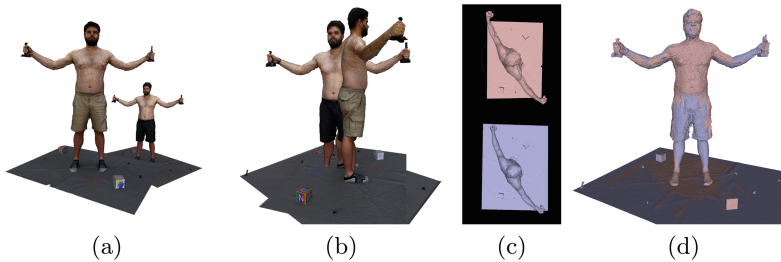


Fig. 7. Alignment process: (a) Before scale (b) After scale (c) Point-wise alignment (d) Aligned models

2.4 Comparison Between Models

In order to visualize in detail the body transformations, a per vertex computation of the signed distance between a mesh and a reference mesh is performed. The pair of vertices is defined by the closest distance. We also use Meshlab to perform this task (Fig. 8).

An *evaluation* of single reconstructions was performed by comparing the height and abdomen measurements of the subject’s digital model to his real measurements, respectively (Figs. 6 and 10).

3 Results

On three different days, a mesh of the subject’s body has been produced. The complete acquisition and processing of one model was done in around four hours. The resulting meshes are of good quality and the difference (shorts should not be considered) between two of those meshes is shown in Fig. 8. To demonstrate the results of the alignment and comparison procedure for larger body transformations, we modified two of the models using a sculpting software. The second model (Fig. 9(b)) was mostly changed on the abdominal regions, adding more muscles and making the waist thinner. The third model (Fig. 9(c)) have even more developed muscles on the abdominal area, the trapezius muscle was enhanced near the neck and the biceps were also reworked. As visible in Figs. 9(d–f), these transformations are well represented in the visualizations.

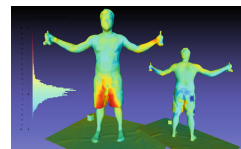


Fig. 8. Comparison

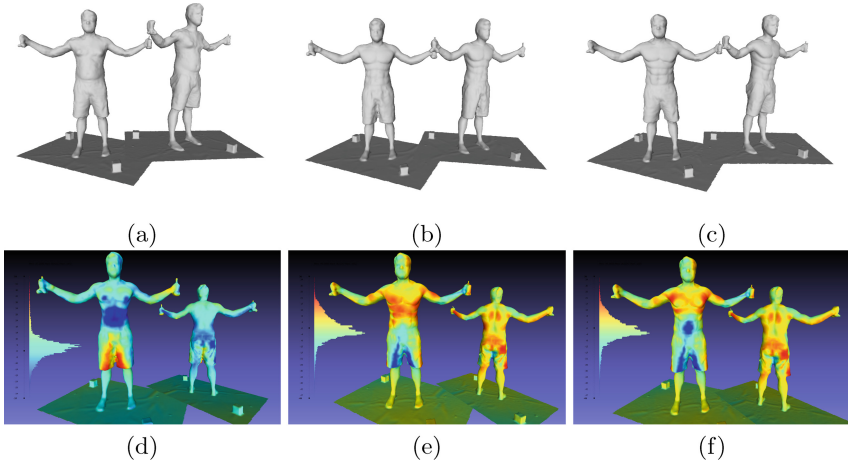


Fig. 9. Models and comparisons in millimeters (histogram). **(a)** Original model of day 1. **(b)** & **(c)** Modified models. Comparison **(d)** of (b) to (a), **(e)** of (c) to (b), **(f)** of (c) to (a)

4 Conclusion

We demonstrated that using SfM employed in a photogrammetry pipeline can be a tool to measure body transformations. As proposed, it is an easily accessible alternative to expensive commercial devices, yet scaling towards precision if using advanced equipment. For future works, we would like to test the concept with other subjects and build a framework to automatically handle all the performed operations. The *Contribution* of this work defines a workflow for measuring body transformation that works with accessible, ubiquitous equipment and free software, yet producing acceptable results (Fig. 11). It also scales in precision when using high-end devices and/or using commercial high-end softwares. The output of mutually aligned 3D meshes allows to employ all mesh-based analyses, including the measurement of volume and other dimensions (Fig. 10), but also more elaborated mesh comparison algorithms, and 3D visualizations. Finally, using these analyses and visualizations, a better understanding of the body transformation can be achieved.

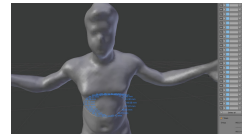


Fig. 10. Blender measurements.

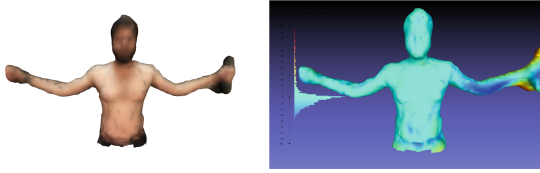


Fig. 11. Smartphone & COLMAP.

References

1. Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G.: MeshLab: an open-source mesh processing tool. In: Eurographics Italian Chapter Conference (2008)
2. Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered image sets, or “How Do I Organize My Holiday Snaps?”. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 414–431. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-47969-4_28
3. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113 (2016)
4. Shaw, M.P., Robinson, J., Peart, D.J.: Comparison of a mobile application to estimate percentage body fat to other non-laboratory based measurements. *Biomed. Hum. Kinet.* **9**(1), 94–98 (2017)
5. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph. (TOG)* **25**, 835–846 (2006)
6. Wang, J., Gallagher, D., Thornton, J.C., Yu, W., Horlick, M., Pi-Sunyer, F.X.: Validation of a 3-dimensional photonic scanner for the measurement of body volumes, dimensions, and percentage body fat—. *Am. J. Clin. Nutr.* **83**(4), 809–816 (2006)
7. Wells, J., Fewtrell, M.: Measuring body composition. *Arch. Dis. Childhood* **91**(7), 612–617 (2006)
8. Wells, J., Ruto, A., Treleaven, P.: Whole-body three-dimensional photonic scanning: a new technique for obesity research and clinical practice. *Int. J. Obes.* **32**(2), 232 (2008)
9. Wells, J.C., et al.: Acceptability, precision and accuracy of 3D photonic scanning for measurement of body shape in a multi-ethnic sample of children aged 5–11 years: the slic study. *PLoS One* **10**(4), e0124193 (2015)